Poznań University of Technology
Faculty of Computing
Institute of Computing Science

Doctoral dissertation

APPLICATION OF THE DOMINANCE-BASED ROUGH SET APPROACH TO RANKING AND SIMILARITY-BASED CLASSIFICATION PROBLEMS

Marcin Szeląg

Supervisor
Professor Roman Słowiński

Poznań, 2015

Politechnika Poznańska
Wydział Informatyki
Instytut Informatyki

Rozprawa doktorska

ZASTOSOWANIE TEORII ZBIORÓW PRZYBLIŻONYCH Z RELACJĄ DOMINACJI DO PROBLEMÓW PORZĄDKOWANIA I KLASYFIKACJI NA PODSTAWIE PODOBIEŃSTWA

Marcin Szeląg

Promotor
Prof. dr hab. inż. Roman Słowiński

Poznań, 2015

Abstract

The thesis concerns *decision aiding* methods for *multicriteria ranking problem* and *similarity-based classification problem* (*case-based reasoning*, CBR). For both problems, we present methods that adapt the *Dominance-based Rough Set Approach* (DRSA).

In the multicriteria ranking problem, there is given a finite set of *objects* (also called alternatives, actions, solutions, etc.), described by a set of *criteria* (cardinal and/or ordinal ones). The aim of decision aiding is to *recommend* to a *decision maker* (DM) a *ranking* of the given objects that reflects her/his preferences. This ranking is supposed to be a *complete preorder* or a *partial preorder*.

The only conclusion that stems from the formulation of a multicriteria decision problem is the *dominance relation* over the set of objects, however, leaving many objects incomparable, it is too poor to make a useful recommendation. Therefore, in Multicriteria Decision Aiding (MCDA), it is necessary to acquire from the DM an additional information, called *preference information*. In our approach, this information consists of *decision examples* in the form of *pairwise comparisons* of some *reference objects*. It is used to build a *preference model* of the DM. This model induces a *preference structure* on the set of objects to be ranked. A proper *exploitation* of this structure yields a ranking of objects that is presented to the DM. Acceptance of this ranking by the DM depends, on one hand, on the accuracy of prediction of pairwise comparisons of new objects, and on the other hand, on the DM's conviction that the preference model represents her/his preferences in a comprehensible and accurate way.

Existing decision aiding methods for multicriteria ranking differ by the type of acquired preference information, and by the type of preference model built from this information. A preference model can take the form of a function (e.g., additive value function), binary relation (e.g., outranking relation), or a set of *if-then decision rules* of the type "if a conjunction of elementary conditions on chosen criteria holds, then it is suggested to take a given decision". Traditional preference models applied in MCDA, i.e., value functions and outranking relations, are hard to understand by the DM, require complex preference information, and are based on rather strong assumptions, that are often not met in practice. Therefore, it is currently strongly postulated that the preference model should be *induced* (learned) from decision examples supplied by the DM. Expression of preferences by making some exemplary decisions does not require from the DM the specification of

parameter values (like thresholds, weights, substitution rates, etc.) of the applied preference model. Another important postulate of MCDA is the *readability* of the constructed preference model which relates to the concept of a *glass box* (as the opposite of a black box).

To fulfill both above postulates, we employ a logical preference model in the form of a set of if-then decision rules induced from decision examples. The rules are induced based on DM's pairwise comparisons of reference objects of the type "object a is *at least as good as* object b" and "object a *is not at least as good as* object b". Such holistic preference information, in terms of *outranking relation* S and *non-outranking relation* S^c, is relatively easy to obtain from the DM. Moreover, rule preference model is the most general preference model as it gives account of most complex interactions among criteria, accepts ordinal evaluation scales, and does not convert ordinal evaluations into cardinal ones. An additional advantage of this model is its readability, as well as its ability to explain past decisions in terms of multicriteria evaluations of objects, and to predict future decisions.

In practice, decision examples given by a DM are often *inconsistent* with the *dominance principle*, which is the basic principle of rational decision making in the presence of multiple criteria. Such an inconsistency is observed, e.g., if object a is preferred to object b at least as much as object c is preferred to object d with respect to each considered criterion, but the DM assigned pair (a, b) to relation S and pair (c, d) to relation S^c. The inconsistency of decision examples may come, e.g., from the lack of some important criteria, hesitation of the DM, or from unstable character of her/his preferences.

In this thesis, we propose a multicriteria ranking method called VC-DRSA^{rank}, which controls consistency of the pairwise comparisons of reference objects given by the DM. This method adapts a variant of DRSA called *Variable Consistency Dominance-based Rough Set Approach* (VC-DRSA). An advantage of this variant is especially visible when the number of inconsistent pairs of objects is significant. The idea of VC-DRSA^{rank} consists in preceding induction of rules with *structuring* pairs of objects from outranking and non-outranking relations, by identification, in both relations, of sufficiently consistent pairs of objects (in the sense of an adopted consistency measure). These pairs form so-called *lower approximations* of the considered relations. Next, a rule induction algorithm is applied to induce (sufficiently certain) decision rules based on pairs of objects from the lower approximations. The set of rules constructed in this way is highly reliable since it represents only strong relationships observed in decision examples. An exemplary decision rule is: "if price difference between disks a and b is no more than 100$, and capacity difference between disks a and b is no less than 500 GB, then disk a is at least as good as disk b".

Induced rules are applied on a set of objects to be ranked. This yields a preference structure on this set of objects that needs to be properly exploited in order to get a final ranking. In the thesis, we analyze several *exploitation procedures*. In particular, we focus our attention on a two-phase exploitation procedure that generalizes previously applied approach based on net flow scores. This new procedure consists in transformation of the preference structure to a valued relation, and in subsequent exploitation of this relation using a *ranking method*. We analyze five rankings methods known from the literature, with respect to several *desirable properties*, such as *monotonicity*, *independence of non-discriminating objects*, *independence of circuits*, etc. We present a series of mathematical proofs concerning these properties. In result, we identify the *Net Flow Rule* as the best ranking method.

We also introduce a new measure for measuring concordance between a final ranking and the initial pairwise comparisons of reference objects chosen from the set of ranked objects. We prove that this new measure is a generalization of the *Kendal rank correlation coefficient*.

In order to verify the proposed method for the multicriteria ranking problem, we performed a computational experiment in which six variants of this method were compared with a reference method SVM^{rank}, considered by the machine learning community to be one of the best learning approaches for the considered problem. The variants of VC-DRSA^{rank} result, first, from different perception of the set of criteria (i.e., whether it is considered to be a consistent set or a not necessarily consistent set of criteria), secondly, from different interpretation of rule matching (qualitative interpretation – is there a rule covering a given pair of objects; quantitative interpretation – what is the strength of the strongest rule covering a given pair of objects), and thirdly, from different definitions of rule strength in case of quantitative interpretation of rule matching. The experiment showed that VC-DRSA^{rank} is strongly competitive to SVM^{rank}. It also allowed to formulate a number of conclusions concerning mutual comparisons of the six variants of VC-DRSA^{rank}.

In the similarity-based classification problem, there is given a finite set of training objects, described by a set of *features* (nominal and/or numerical ones), a set of *marginal similarity functions* (one for each feature), and a set of predefined *decison classes*. Each decision class is considered to be a *fuzzy set* – each object belongs with a certain degree, between 0 and 1, to each of the classes. The aim of decision aiding is to present to a DM a recommendation concerning a new object, in terms of a degree of membership of this object to particular classes. This recommendation is obtained using CBR, taking into account similarity with respect to some *reference objects* indicated by the DM.

In case-based reasoning, one needs a *similarity model*. Traditionally, this model has the form of a real-valued aggregation function (e.g., Euclidean norm) or binary relation (e.g., fuzzy relation). In this thesis, we present a method based on DRSA, using a new similarity model in terms of a set of if-then decision rules employing dominance relation in the space created by marginal similarity functions. Such a model makes it possible to avoid an arbitrary aggregation of marginal similarity functions. An exemplary rule is: "if similarity of patient y to patient x with respect to *temperature* is at least 0.8, and similarity of y to x with respect to *muscle pain* is at least 1.0, then the membership of y to class *influenza* is in the interval $[0.8, 1.0]$", where x is a reference object characterized by the following feature values: *temperature*=39, *muscle pain*=yes, and *headache*=yes. Decision rules are induced separately for each decision class X and each reference object x, for lower or upper approximations of sets of objects whose membership to class X is in the interval $[\alpha, \beta]$ and outside the interval (α, β), where α and β are membership values such that $[\alpha, \beta]$ contains the membership of x. These rules underline the monotonic relationship "the more similar is object y to object x w.r.t. the considered features, the closer is y to x in terms of the membership to a given decision class X". Violation of this principle causes an inconsistency in the set of objects, which is handled using DRSA. An important characteristic of the proposed approach is that induced rules employ only ordinal properties of marginal similarity functions. Thus, this approach is invariant to ordinally equivalent marginal similarity functions.

We propose a rule-based classification scheme for determining a precise value of membership of a new object to each considered class. This scheme involves application of induced certain/possible decision rules. It improves the classification scheme often applied in VC-DRSA.

It is worth underlining that our method for CBR provides additional useful information that other similarity-based classification methods do not provide. In particular, when determining membership of a new object z to class X, the DM can see the rules matching z, and also view the training objects that support the matching rules. Moreover, the rules clearly show the conditions for similarity.

Streszczenie

Rozprawa doktorska dotyczy problematyki *wspomagania decyzji* w *problemie porządkowania* (*rankingu*) *obiektów* z uwzględnieniem preferencji względem wielokryterialnych ocen obiektów oraz w problemie *wieloatrybutowej klasyfikacji obiektów na podstawie podobieństwa do znanych przypadków* (ang. case-based reasoning, CBR). Dla każdego z tych problemów w rozprawie zaproponowano metodykę wspomagania decyzji wykorzystującą *teorię zbiorów przybliżonych opartą na dominacji* (ang. Dominance-based Rough Set Approach, DRSA).

W pierwszym z analizowanych problemów, dany jest skończony zbiór *obiektów* (wariantów decyzyjnych, akcji, rozwiązań itp.), opisanych za pomocą zbioru *kryteriów* (ilościowych i/lub porządkowych). Celem wspomagania decyzji jest przedstawienie *decydentowi* (użytkownikowi) *rekomendacji* w postaci rankingu rozważanych obiektów, który byłby zgodny z preferencjami tego decydenta. Ranking ten może mieć postać *preporządku zupełnego* lub *częściowego*. Ponieważ *relacja dominacji* na zbiorze obiektów jest, poza trywialnymi przypadkami, niewystarczająca do uporządkowania obiektów, w celu wypracowania rekomendacji konieczne jest pozyskanie od decydenta dodatkowej informacji, tzw. *informacji preferencyjnej*, w formie przykładów decyzji w postaci *porównań parami* wybranych obiektów. Na podstawie tej informacji tworzy się tzw. *model preferencji* decydenta, który indukuje strukturę preferencji w zbiorze rozważanych obiektów. Odpowiednia *eksploatacja* tej struktury prowadzi do rekomendowanego rankingu obiektów. Akceptacja tego rankingu przez decydenta jako zgodnego z jego preferencjami zależy, z jednej strony, od jakości predykcji porównań parami nowych obiektów, a z drugiej strony, od przekonania decydenta, że model preferencji reprezentuje jego preferencje w sposób zrozumiały i trafny.

Istniejące metody wspomagania decyzji różnią się rodzajem pozyskiwanej informacji preferencyjnej i postacią modelu preferencji. Model ten może mieć postać funkcji rzeczywistej (np. addytywnej funkcji użyteczności), relacji binarnej (np. relacji przewyższania) lub zbioru reguł decyzyjnych typu "jeżeli zachodzi koniunkcja warunków elementarnych na wybranych kryteriach/atrybutach, to sugerowane jest podjęcie określonej decyzji". Tradycyjne modele preferencji stosowane we wspomaganiu decyzji, mające postać funkcji użyteczności lub relacji przewyższania, są mało zrozumiałe dla decydenta w procesie wspomagania decyzji, a ponadto wymagają od decydenta wielu trudnych informacji

preferencyjnych i są oparte na stosunkowo silnych założeniach, które często nie są spełnione w praktyce. Stąd też obecnie silnie postuluje się tworzenie modelu preferencji na podstawie przykładów decyzji, czyli przez *indukcję* (uogólnianie). Takie podejście nie wymaga od decydenta specyfikacji wartości parametrów modelu preferencji (progów, wag, itp.). Drugim ważnym postulatem jest *czytelność* i *zrozumiałość* tworzonego modelu preferencji, nawiązująca do koncepcji tzw. "przejrzystej skrzynki" (ang. glass box), która jest przeciwieństwem "czarnej skrzynki" (ang. black box).

W celu spełnienia obydwu powyższych postulatów, w rozprawie przyjęto podejście polegające na tworzeniu modelu preferencji w postaci *zbioru reguł decyzyjnych*. Reguły te indukowane są na podstawie podanych przez decydenta przykładów decyzji w postaci porównań parami *obiektów referencyjnych* typu "obiekt a jest *co najmniej tak dobry jak* obiekt b" oraz "obiekt a *nie jest co najmniej tak dobry* jak obiekt b". Tego typu holistyczna informacja preferencyjna, odpowiadająca *relacji przewyższania* S (ang. outranking) i *relacji braku przewyższania* S^c (ang. non-outranking), jest względnie prosta do pozyskania od decydenta. Z kolei model regułowy jest najogólniejszym znanym modelem prefe-rencji, gdyż jest zdolny do reprezentowania ogólniejszych interakcji między atrybutami niż modele funkcyjne i relacyjne. Dodatkową zaletą tego modelu jest jego czytelność, interpretowalność, zdolność wyjaśniania związków między wielokryterialną oceną obiektów a decyzją, występujących w przykładach decyzji, oraz możliwość zastosowania do predykcji decyzji przyszłych.

W praktyce, przykładowe decyzje podjęte przez decydenta często okazują się *niespójne z zasadą dominacji*, która jest podstawową zasadą racjonalnego podejmowania decyzji w obecności wielu kryteriów. Niespójność taka obserwowana jest przykładowo w sytuacji, gdy, z jednej strony, dla każdego rozważanego kryterium obiekt a jest co najmniej tak silnie preferowany nad obiekt b jak obiekt c jest preferowany nad obiekt d (inaczej mówiąc, para (a, b) dominuje parę (c, d)), z drugiej zaś strony, para obiektów (a, b) według decydenta należy do relacji S^c a para obiektów (c, d) należy do relacji S. Przyczyn niespójności upatruje się w pominięciu pewnych kryteriów istotnych dla decydenta, w niepewności towarzyszącej jego decyzjom i w niestabilności jego preferencji.

Zaproponowana w rozprawie metoda porządkowania obiektów VC-DRSA^{rank} kontroluje spójność podanej przez decydenta informacji preferencyjnej mającej postać porównań wybranych obiektów parami. Metoda ta opiera się na wariancie DRSA ze zmienną spójnością (ang. Variable Consistency Dominance-based Rough Set Approach, VC-DRSA). Przewaga tego wariantu na klasycznym DRSA jest widoczna szczególnie w sytuacji, gdy liczba niespójnych porównań jest znacząca. Istotą zaproponowanego w rozprawie podejścia jest poprzedzenie indukcji reguł *strukturalizacją* zbioru par obiektów należących do relacji S i S^c poprzez identyfikację w obu relacjach par dostatecznie spójnych (w sensie przyjętej miary spójności par obiektów). Pary te tworzą tzw. *dolne przybliżenia* odpowiednich relacji. Następnie, stosuje się algorytm indukcji (dostatecznie pewnych) reguł decyzyjnych w oparciu o pary obiektów z dolnych przybliżeń. Utworzony w ten sposób zbiór reguł decyzyjnych cechuje się dużą wiarygodnością, gdyż reprezentuje jedynie silne zależności logiczne obserwowane w przykładach decyzji. Przykładowa reguła decyzyjna ma postać: "jeżeli różnica cen dysków a i b nie przekracza 100 zł oraz różnica pojemności tych dysków jest nie mniejsza 500 GB, to dysk a jest co najmniej tak dobry jak dysk b".

Wyindukowane reguły decyzyjne stosuje się na zbiorze obiektów podlegających porządkowaniu (rangowaniu). Powstała w ten sposób struktura preferencji eksploatowana jest za pomocą *procedury rangującej*. W rozprawie przeanalizowano szereg takich procedur. Szczególną uwagę poświęcono procedurze dwufazowej, uogólniającej stosowane dotychczas podejście do eksploatacji relacyjnej struktury preferencji. Procedura ta polega na transformacji struktury preferencji do relacji wartościowanej, a następnie na eksploatacji tej relacji za pomocą *metody rangującej*. W rozprawie przeanalizowano pięć metod rangujących znanych z literatury, pod kątem kilkunastu pożądanych własności, takich jak *monotoniczność*, *niezależność od niedyskryminujących obiektów*, *niezależność od cykli*, itp. Przeprowadzono szereg dowodów matematycznych dotyczących tych własności. W wyniku tej analizy ustalono, że metodą rangującą o najlepszych własnościach jest metoda *Net Flow Rule*.

W rozprawie zaproponowano również nową miarę zgodności rankingu końcowego z informacją preferencyjną decydenta w postaci porównań parami obiektów referencyjnych wybranych ze zbioru porządkowanych obiektów. Wykazano, że miara te jest uogólnieniem *współczynnika korelacji rangowej Kendalla*.

W celu weryfikacji zaproponowanej metody porządkowania obiektów, przeprowadzono eksperyment obliczeniowy, w którym sześć wariantów tej metody porównano z metodą SVM^{rank}, uznawaną przez specjalistów od *uczenia maszynowego* (ang. machine learning) za jedną z najlepszych metod uczenia się dla problemu rankingu. Wspomniane warianty metody VC-DRSA^{rank} wynikały, po pierwsze, z rożnego spojrzenia na zbiór kryteriów (spójny zbiór kryteriów; niekoniecznie spójny zbiór kryteriów), po drugie, z różnej interpretacji pokrycia par obiektów wyindukowanymi regułami (interpretacja jakościowa – czy para obiektów pokrywana jest przez co najmniej jedną regułę; interpretacja ilościowa – jaka jest siła najsilniejszej reguły pokrywającej parę obiektów), a po trzecie, z różnej definicji siły reguły w przypadku ilościowej interpretacji pokrycia regułami. Eksperyment wykazał, m.in., że metoda VC-DRSA^{rank} jest silnie konkurencyjna w stosunku do metody SVM^{rank}. Pozwolił również sformułować szereg wniosków odnośnie do porównywanych ze sobą wariantów metody VC-DRSA^{rank}.

W drugim z analizowanych w rozprawie problemów, dany jest skończony zbiór obiektów uczących, opisanych za pomocą zbioru *cech* (atrybutów nominalnych bądź numerycznych), zbiór *funkcji podobieństwa*, z których każda określa podobieństwo obiektów na jednej z rozważanych cech oraz zbiór *klas decyzyjnych* będących w ogólności zbiorami rozmytymi – każdy obiekt przynależy w pewnym stopniu, od 0 do 1, do każdej z klas decyzyjnych. Celem wspomagania decyzji jest przedstawienie decydentowi rekomendacji dla nowego obiektu, w postaci jego *stopnia przynależności* do poszczególnych klas decyzyjnych, zaś sposobem dojścia do tej rekomendacji jest wnioskowanie na podstawie podobieństwa do określonych przez decydenta *obiektów referencyjnych*.

Do wnioskowania na podstawie podobieństwa do znanych przypadków konieczny jest *model podobieństwa*. Tradycyjnie, modelem podobieństwa w tym podejściu jest określona funkcja rzeczywista (np. norma euklidesowa) lub relacja binarna (np. relacja rozmyta). W rozprawie przedstawiona została metoda oparta na DRSA, wykorzystująca nowy model podobieństwa w postaci zbioru reguł decyzyjnych wykorzystujących relację dominacji. Jest to model najmniej obciążony arbitralnymi założeniami odnośnie do agregacji podobieństw na poszczególnych cechach. Przykładowa reguła decyzyjna ma postać: "jeżeli podobieństwo pacjenta y do pacjenta referencyjnego x na cesze *temperatura* jest ≥ 0.8 oraz podobieństwo y do x na cesze *ból mięśni* jest ≥ 1.0, to przynależność pacjenta y do klasy decyzyjnej *grypa* zawiera się w przedziale $[0.8, 1.0]$", gdzie obiekt referencyjny x posiada następujące wartości cech: *temperatura*=39, *ból mięśni*=tak, *ból głowy*=tak. Reguły decyzyjne indukowane są osobno dla każdej klasy decyzyjnej X i każdego obiektu referencyjnego x, dla dolnych lub górnych przybliżeń zbiorów takich obiektów, których przynależność do klasy X zawiera się w przedziale $[\alpha, \beta]$ oraz poza przedziałem (α, β), gdzie α i β są wartościami funkcji przynależności do klasy X takimi, że przedział $[\alpha, \beta]$ zawiera przynależność referenta x. Reguły decyzyjne podkreślają następującą zależność monotoniczną: "im bardziej obiekt y jest podobny do obiektu x na rozważanych cechach, tym bardziej podobna jest przynależność y do klasy X do przynależności x do klasy X". Odstępstwa od tej zasady powodują niespójności w zbiorze obiektów, analizowane z wykorzystaniem DRSA. Ważną cechą zaproponowanego podejścia jest to, że reguły decyzyjne wykorzystują jedynie porządkowe własności funkcji podobieństwa zdefiniowanych dla poszczególnych cech. Stąd też, proponowane w rozprawie podejście jest niewrażliwe na dobór tych funkcji, pod warunkiem, iż są one porządkowo równoważne.

W celu zastosowania wyindukowanych pewnych/możliwych reguł decyzyjnych do nowego obiektu z, w rozprawie zaproponowano ulepszony schemat klasyfikacji regułowej znany z klasyfikacji porządkowej, który umożliwia dokonywanie predykcji punktowych, w tym sensie, że dla każdej klasy decyzyjnej X decydent otrzymuje precyzyjną predykcję stopnia przynależności obiektu z do tej klasy (zamiast np. przedziału, do którego ten stopień przynależności należy).

Warto podkreślić, że zaproponowana w rozprawie metoda dla problemu klasyfikacji na podstawie podobieństwa do znanych przypadków dostarcza decydentowi szeregu użytecznych informacji, których inne metody stosowane dla tego problemu nie udostępniają. W szczególności, w związku z określeniem stopnia przynależności nowego obiektu z do klasy X, decydent może uzyskać informację o regułach pokrywających ten obiekt, a także, o obiektach uczących, które wspierają te reguły. Ponadto, reguły są czytelnym interpretatorem warunków podobieństwa.

Preface

Human activity in real-life is inherently connected with solving *decision problems*. Usually, these problems concern a set of *objects* (also called variants, actions, alternatives, cases, observations, options, candidates) described by a set of *attributes* (also called features, characteristics, criteria, variables). The attributes can be of different nature – nominal, ordinal or numerical, and some of them may have ordered value sets, e.g., according to user's preferences. One considers mainly three types of decision problems: *classification* of objects to pre-defined decision classes (also called categories, or simply classes), *ranking* of objects from the best to the worst, and *choice* of a subset of the best objects. Some typical decision problems are: assessment of bankruptcy risk of companies, ranking of universities, or choice of a car to buy.

Due to complexity of many real-life decision problems, there is a need for scientific decision aiding. This need has been answered by many decision aiding methodologies developed withing the domain called *Multiple Criteria Decision Aiding* (MCDA). These methodologies address mainly decision problems concerning a finite set of objects described by a finite set of attributes whose value sets are totally ordered according to the preferences of a *decision maker* (DM) (also called user, expert).

In MCDA, in case of classification problems, it is assumed that decision classes are *ordered* according to the preferences of a DM. Therefore, classification problems addressed by MCDA are called *multicriteria sorting* or *ordinal classification with monotonicity constraints*. In the domain of *Machine Learning* (ML), on the other hand, usually one does not assume a priori any order in value sets of attributes. As to decision classes, there exist two interesting versions of the classification problem, apart from the "standard" classification with unordered crisp (hard) classes. The first version is so-called *soft label classification*, where each object from the training set can belong to a certain degree to each of the considered classes. An example of soft label classification is a problem of the recognition of emotions in recordings of spoken sentences. The second version, called *ordinal classification*, concerns the case where decision classes are ordered. An example of this version is a prediction problem when given values of several medical tests one needs to decide to which illness severity class – Benign, Medium, or Severe – a patient belongs.

In the past, we have studied extensively application of the *Dominance-based Rough Set Approach* (DRSA), and its variable-consistency generalizations, to ordinal classification with monotonicity constraints. We have observed that it has several merits comparing to other methods from the fields of MCDA and ML. First, DRSA can handle inconsistencies present in decision examples (training set) using the concept of lower and upper approximations of considered unions of ordered decision classes; these approximations are the basis for induction of decision rules. Second, DRSA can be used for heterogeneous data, containing simultaneously nominal, ordinal, and cardinal (numerical) attributes – no prior discretization of numerical attributes nor prior conversion of nominal and ordinal attributes into numerical ones is required. Third, the basic idea of DRSA is concordant with an important trend in MCDA consisting in induction of *preference model* from *decision examples* (also known as use of *indirect preference information*). Fourth, application of DRSA enables to induce from decision examples an intelligible preference model in the form of a set of *monotonic if-then decision rules*.

Considering the success of the application of DRSA to ordinal classification with monotonicity constraints, we found it worthwhile to study applications of DRSA to other decision problems. This thesis is a result of this research. It concerns two decision problems. The first one is the ranking problem as considered in MCDA, i.e., with attributes whose value sets are ordered according to DM's preferences. The second problem is the *similarity-based classification*. In this thesis, the ranking problem is handled using a proper adaptation of the *Variable Consistency Dominance-based Rough Set Approach* (VC-DRSA), and the similarity-based classification is handled using a proper adaptation of DRSA to *similarity-based reasoning*. We present both theoretical and experimental results concerning the proposed approaches.

Acknowledgment

This thesis was prepared in the Laboratory of Intelligent Decision Support Systems at the Faculty of Computing at Poznań University of Technology. Numerous people helped me to bring this thesis to life and I would like to thank them here.

I deeply thank my supervisor Professor Roman Słowiński for his inspiration, continuous support, and dedication to guiding my work. I also appreciate invaluable help of Salvatore Greco and of Jurek Błaszczyński, Krzysztof Dembczyński, Wojciech Kotłowski, Miłosz Kadziński and other colleagues I learned so much from.

Finally, I would like to thank my belowed wife Karolina and my parents for their continuous support, encouragement, and enormous patience they showed for my endless work.

Contents

1 Introduction **1**

1.1 Problem Setting . . . 1

1.1.1 Multicriteria Ranking Problem . . . 3

1.1.2 Similarity-based Classification Problem . . . 7

1.2 Review of Existing Approaches to Multicriteria Ranking . . . 8

1.2.1 MCDA Approaches to Multicriteria Ranking . . . 8

Multiple Attribute Utility Theory (MAUT). . . . 9

Outranking Methods . . . 12

Decision Rule-based Methods . . . 18

Other Methods . . . 23

1.2.2 PL Approaches to Multicriteria Ranking . . . 24

SVM^{rank} . . . 24

RankBoost . . . 25

Ensembles of Decision Rules . . . 25

Critical Remarks . . . 26

1.3 Review of Existing Approaches to Similarity-based Classification . . . 26

1.3.1 Lazy Learning Methods . . . 26

1.3.2 Eager Learning Methods . . . 27

1.3.3 Measuring Similarity . . . 29

1.4 Motivation for Dominance-based Rough Set Approaches to Multicriteria Ranking and Similarity-based Classification . . . 29

1.4.1 Multicriteria Ranking Problem . . . 30

1.4.2 Procedures for Exploitation of Preference Graph . . . 31

1.4.3 Similarity-based Classification Problem . . . 32

1.5 Goal and Scope of the Thesis . . . 34

1.6 Thesis Outline . . . 34

2 Application of VC-DRSA to Multicriteria Ranking Problem **37**

2.1 Introduction . . . 37

2.2 Preliminaries . . . 39

2.3 Problem Setting . . . 40

2.4 Pairwise Comparison Table . . . 42
2.5 Rough Approximation of Outranking and Non-outranking Relations . . . 43
2.6 Induction of Decision Rules . . . 47
2.7 Application of Decision Rules . . . 50
2.7.1 Definition of Relations $\mathbb{S}$ and $\mathbb{S}^c$ – Consistent Set of Criteria, Crisp Relations . . . 51
2.7.2 Definition of Relations $\mathbb{S}$ and $\mathbb{S}^c$ – Consistent Set of Criteria, Valued Relations . . . 52
2.7.3 Definition of Relations $\mathbb{S}$ and $\mathbb{S}^c$ – Not Necessarily Consistent Set of Criteria, Crisp Relations . . . 52
2.7.4 Definition of relations $\mathbb{S}$ and $\mathbb{S}^c$ – Not Necessarily Consistent Set of Criteria, Valued Relations . . . 53
2.7.5 Four-valued Outranking – Crisp Relations . . . 53
2.7.6 Four-valued Outranking – Valued Relations . . . 53
2.7.7 Preference Graph . . . 54
2.8 Exploitation of Preference Graph . . . 54
2.8.1 Review of Possible Exploitation Procedures . . . 55
2.8.2 Fusion of Relations $\mathbb{S}$ and $\mathbb{S}^c$ in Order to Exploit a Single Relation . 56
2.8.3 Review of Ranking Methods . . . 57
2.9 Analysis of Final Ranking . . . 60
2.9.1 Kendall Rank Correlation Coefficient τ . . . 61
2.9.2 New Concordance Measure τ' . . . 61
2.10 Illustrative Example . . . 62

3 Analysis of Desirable Properties of Ranking Methods 69
3.1 Desirable Properties of Ranking Methods . . . 69
3.2 Priority Orders of Desirable Properties . . . 75
3.3 Verification of Properties of Ranking Methods and Choice of the Best Method 75

4 Similarity-based Classification using Monotonic Rules 79
4.1 Introduction . . . 79
4.2 Basic Notions and Assumptions . . . 80
4.3 Similarity Learning . . . 82
4.4 Similarity Tables . . . 83
4.5 Comprehensive Closeness of Objects . . . 84
4.6 Rough Approximation by Dominance Relation . . . 85
4.7 Induction of Decision Rules . . . 87
4.8 Application of Decision Rules . . . 88
4.9 Illustrative Example . . . 91

5 Computational Evaluation of the Proposed Methodology 97

5.1 Setup of Computational Experiment for Multicriteria Ranking 98
5.2 Analysis of Results 101

6 Summary and Conclusions **109**
6.1 Main Contributions of the Thesis 109
6.2 Directions of Future Research 115

Appendix **117**

Bibliography **125**

Chapter 1

Introduction

1.1 Problem Setting

Multiple Criteria Decision Aiding (MCDA). The research field of *Multicriteria Decision Aiding* (MCDA) [13, 53, 133] addresses different types of *decision problems* involving a finite set of *objects* (also called variants, actions, alternatives, solutions, options, candidates, etc.) evaluated by a finite set of preference-ordered attributes called *criteria*. The criteria are equipped with monotonic preference scales which specify total preference orders in their value sets. For example, in a decision regarding selection of a notebook, its price and disk capacity are criteria because, obviously, a low price is better than a high price, and high disk capacity is better that a low one. A typical assumption of MCDA is the presence of a consistent set (family) of criteria [146], i.e., a set of criteria that satisfies the properties of:

- *completeness* (all relevant criteria are considered),
- *monotonicity* (the better the evaluation of an object on considered criteria, the more it is preferable to another object),
- and *non-redundancy* (there is no criterion which could be removed without violating one of the previous two properties).

According to Roy [143], the main types of multicriteria decision problems are: *description*, *ranking*, *choice*, and *sorting* (also called ordinal classification). The emergence of MCDA results from the need of scientific decision aiding for many real-life decision problems in which considered criteria represent conflicting viewpoints, preventing from finding an objectively optimal solution of a problem at hand. MCDA offers a wide range of methods aiding *decision makers* (DMs) in making a decision by putting forward a problem-specific *recommendation*. MCDA helps to structure decision problems, incorporate in the analysis the preferences of a DM, develop a model of a decision situation, and exploit this model to work out a final recommendation [148].

Machine Learning (ML). The research field of Machine Learning (ML) [19, 67, 99, 128, 129] addresses different kind of decision problems. Ones of the most often considered are *classification problems*. Problems involving classification of objects are considered in the subfield of *supervised learning*, where the task is to *learn* a *classifier* based on so-called *training objects* for which the correct classification is known. The aim of the learning is to construct a classifier that predicts as accurately as possible the class for the objects from another set of so-called *testing* (*unseen*) *objects*.

Preference Learning (PL). Recently, one can observe *Preference Learning* (PL) [71] emerging as an important subfield of ML. Most of the research in PL concerns "learning to rank" types of problems [43, 46, 123, 140] which often boil down to a series of ordinal classification problems (see, e.g., [3]). PL concerns new challenging applications related to Internet, in particular, recommender systems and information retrieval. In the former application, the task is to recommend to a user a new item (movie, book, etc.) that fits her/his preferences. The recommendation is computed on the basis of the information describing the past behavior of the user. In the latter application, the task is to rank the documents retrieved by a search engine according to user's preferences. There are several algorithms that are tailored for these kinds of problems. The learning is traditionally achieved by minimizing an empirical estimate of an assumed *loss function* [50]. The most popular approaches are based on *rank loss minimization*. These include variants of support vector machines [101, 113] and boosting [46, 65].

In ML, and in particular in PL, it is assumed that there exists a stochastic process generating the observed data (the "ground truth"). Thus, training data are considered to be a sample from some unknown multidimensional probability distribution. Moreover, these data are considered to be noisy. Under these circumstances, the goal of the learning is to induce a preference model that generalizes the training data (i.e., refers to an entire *population of individuals*), and thus, allows for making good predictions *on average*. The main focus of PL is on predictive accuracy of the learned model and on the scalability of proposed algorithms (in view of massive data analysis).

As pointed out in [44, 46], MCDA and PL share some goals, concepts and methodological issues. The main difference between them consists in the way of building a DM's *preference model*. In PL, the preference model results from statistical analysis of data (training objects). In MCDA, the preference model is built from *preference information* elicited from the DM, often interactively.

1.1.1 Multicriteria Ranking Problem

Multicriteria ranking. Among different types of multicriteria decision problems, in this thesis we consider the ranking problem, where the goal is to rank order a given set of objects described by multiple criteria. The ranking to be obtained may be a *total preorder* (i.e., a linear order with possible ties, also called *complete preorder* or *weak order*) or a *partial preorder* (which allows to conclude that some objects are incomparable). Multicriteria ranking problems are frequent quests in such fields as finance, economy, management, and engineering [169, 175]. Some typical examples include ranking of universities, hospitals, cities, countries, smartphones, notebooks, airlines, etc.

Ranking problem is, obviously, not a new problem and it has been investigated in various fields such as decision theory, social sciences, information retrieval, mathematical economics, MCDA and PL. In this thesis, we adopt the MCDA perspective, accounting also for some approaches proposed in the field of PL.

In MCDA, the construction of evaluation criteria with explicit monotonic preference scales is an important step in the procedure of decision aiding. The criteria are functions with ordinal or cardinal (i.e., interval or ratio) scales, built on elementary features of objects to permit a meaningful distinction of objects, i.e., objects which are indiscernible with respect to (w.r.t.) a given set of criteria are considered indifferent. In PL, on the other hand, the relationships between value sets of attributes and DM's preferences are discovered from data for a direct use in decision making. This means that in PL, the monotonic preference scales converting elementary features to criteria are neither used nor revealed explicitly.

Remark that given the taxonomy of "learning to rank" problems considered in PL [72], the ranking problem considered in this thesis belongs to the category of *object ranking* problems [43], where the task is to learn a "good" *ranker* (i.e., a model that predicts a ranking), given some pairwise preference information.

Dominance relation. Given a finite set of objects A, and a finite set $G = \{g_1, \ldots, g_n\}$ of criteria giving evaluations $g_i(a)$ to all $a \in A$, $i = 1, \ldots, n$, the *dominance relation* D over set A is defined as follows. Given $a, b \in A$, object a *dominates* object b, which is denoted by aDb, if and only if $g_i(a) \succeq g_i(b)$ for each $i = 1, \ldots, n$, where $\succeq$ means "is at least as good as". The dominance relation D is a partial preorder, i.e., a reflexive and transitive binary relation defined over A on the basis of evaluations $g_i(\cdot)$, $i = 1, \ldots, n$.

Preference information and preference model. In MCDA, conclusions in terms of the dominance relation, resulting only from the analysis of evaluations of objects on multiple criteria, are usually too weak to make a useful recommendation. This is due to the fact that the dominance relation usually leaves many objects incomparable.

In order to increase comparability of objects, it is necessary to obtain (elicit) from a DM an additional information about the objects at hand. This information is called *preference information.* Different decision aiding methods for *multicriteria ranking* differ by the adopted type of preference information and by the type of *preference model* created using this information. This model is used to aggregate vector evaluations of objects in the way that is consistent with the value system of a DM. A preference model can take the form of a function (e.g., additive utility function) as considered in the *Multiple Attribute Utility Theory* (MAUT) [117], binary relation (e.g., outranking relation) as considered in the *outranking approach* [144], or a set of *if-then decision rules* [83, 84, 86, 159]. The preference model induces a *preference structure* on the set of objects. Proper exploitation of this structure yields a ranking of objects that is presented to the DM.

In PL, preference model is a result of statistical analysis of the training set of objects. Thus, training data are the equivalent of preference information in MCDA.

Decision examples. An important recent trend in MCDA is the use of *indirect preference information.* One of the most significant manifestations of this trend is the postulate that the preference model should be created (learned) based on *decision examples* supplied by a DM. The decision examples may be provided by the DM either on a set of real or hypothetical objects, or may come from the observation of DM's past decisions. Expression of preferences by making some exemplary decisions does not require from the DM the knowledge of the details of applied decision aiding method, or the specification of parameter values (like thresholds, weights, substitution rates, etc.) of the applied preference model.

Learning a preference model from decision examples follows the paradigm of *inductive learning* used in artificial intelligence [126]. It is also concordant with the *disaggregation-aggregation paradigm* [110], and with the principle of posterior rationality postulated by March [124], since it emphasizes the discovery of DM's intentions as an interpretation of her/his actions rather than as a priori position. Inductive learning has been used to construct various preference models from decision examples, e.g., general additive utility functions [56, 94], outranking relations [96, 130], monotonic decision trees [75], and sets of if-then decision rules [86].

In the considered multicriteria ranking problem, decision examples most often take the form of *pairwise comparisons* of objects. This is quite natural since the position of an object in a ranking depends of its relation with other objects to be ranked. For two objects a, b, the simplest result of comparison of a against b is the information concerning weak preference relation between a and b. In a more complex case, one could also consider the intensity of preference of a over b.

Readability of a preference model. The second important postulate of MCDA is the *readability* of the created preference model which relates to the concept of a *glass box* [90]. This means that a good preference model should be comprehensible to the DM so (s)he can understand this model and decide whether it is acceptable from the viewpoint of her/his value system.

Rule preference model. One of the preference models that fulfills the above postulates is a logical preference model in the form of a set of if-then decision rules induced from decision examples. This model has been introduced to decision analysis by Greco, Matarazzo, and Słowiński [82, 84, 160]. A popular saying attributed to Slovic [157] is that "people make decisions and then search for rules that justify their choices". The rules can explain the preferential attitude of a DM and enable understanding of the reasons of her/his decisions taken in the past. The recognition of the rules by the DM [121] justifies their use for decision aiding. Thus, the preference model in the form of rules derived from decision examples fulfills both explanation and recommendation goals of decision aiding.

An advantage of representing preferences by decision rules is the possibility of taking into account, at the same time, attributes of different nature – nominal, ordinal, and cardinal (numerical) ones. There is no need of discretization of numerical attributes. Rule preference model is also attractive according to the results of axiomatic analysis of all three preference model types considered in MCDA (i.e., value function, outranking relation, and decision rules) [85, 87]. According to this analysis, a set of decision rules is the only model that gives account of most complex interactions among criteria, is non-compensatory, accepts ordinal evaluation scales, and does not convert ordinal evaluations into cardinal ones. Moreover, decision rules are easily interpretable by users who trust more proposed recommendation [86].

Domain knowledge. An important aspect in the analysis of decision problems is the *domain knowledge*. Exploitation of this knowledge enables to increase quality of the created preference model and to ensure compatibility of this model with the value system of a DM. In case of the multicriteria ranking problem, domain knowledge concerns *value sets* of criteria, their *preference scales*, and *monotonic relationship* of the following type: improvement of object evaluation on one or more criteria should not result in deterioration of position of this object in relation to other objects. Thus, a DM expects from a decision aiding method for multicriteria ranking that the final ranking recommended by this method will preserve dominance relation over the set of ranked objects.

Consistency of decision examples. In practice, decision examples given by a DM are often inconsistent due to hesitation of the DM, unstable character of her/his preferences, or incomplete determination of the set of criteria [145]. These inconsistencies should not be considered as a simple error or as noise. They can convey important information that should be taken into account when constructing the preference model of a DM. In case of preference information in terms of pairwise comparisons of reference objects, used in multicriteria ranking, an inconsistency is a violation of the following general monotonic relationship: "if object a is preferred to object b at least as much as object c is preferred to object d w.r.t. each considered criterion, then the comprehensive preference of a over b is not weaker than the comprehensive preference of c over d".

Dominance-based Rough Set Approach (DRSA). In order to handle inconsistency of decision examples w.r.t. dominance relation, Greco, Matarazzo and Słowiński proposed the Dominance-based Rough Set Approach (DRSA) [78, 79, 82–84, 159], which has been successfully applied to multicriteria sorting, choice, and ranking problems. In DRSA, inconsistent decision examples are not simply corrected or ignored. Instead, all decision examples are *structured* by calculation of *lower* and *upper approximations* of considered sets (i.e., so-called *unions of decision classes* in case of sorting, or *preference relations* in case of ranking), which enables to distinguish certain and possible knowledge, respectively. In this way, it is possible to restrict a priori the set of decision examples that serve as a basis for induction of a preference model to a subset of consistent decision examples belonging to lower approximations. This restriction corresponds to the concept of learning from consistent data, so that the knowledge gained from this learning is relatively certain (or, in other words, the induced preference model is reliable).

As accurately observed in [46], "the usefulness of DRSA goes beyond the frame of MCDA as the type of monotonic relationships handled by DRSA is also meaningful for problems where preferences are not considered but a kind of monotonicity relating ordered attribute values is meaningful for the analysis of data at hand. Indeed, monotonicity concerns, in general, mutual trends existing between different variables, like distance and gravity in physics, or inflation rate and interest rate in economics. Whenever a relationship between different aspects of a phenomenon is discovered, this relationship can be represented by a monotonicity w.r.t. some specific measures or perception of the considered aspects, e.g., 'the colder the weather, the higher the energy consumption' or 'the more a tomato is red, the more it is ripe'. The qualifiers, like 'cold weather', 'high energy consumption', 'red' and 'ripe', may be expressed either in terms of some measurement units, or in terms of degrees of membership to fuzzy sets representing these concepts."

Variable Consistency DRSA (VC-DRSA). When the number of inconsistent decision examples is relatively large, the lower approximations calculated in DRSA are often relatively small of even empty, which makes further analysis difficult. For this reason, different generalizations of DRSA have been considered in the literature, namely: different Variable Consistency Dominance-based Rough Set Approaches (VC-DRSA) [22–24, 97] and Variable Precision Dominance-based Rough Set Approach (VP-DRSA) [108]. These generalizations relax the definition of the lower approximation of a union of decision classes which leads to inclusion of objects which are "sufficiently consistent". The consistency of an object is measured using a *consistency measure*, which can be, e.g., rough membership [97, 136, 177, 178], Bayes factor [156], one of many well-known confirmation measures [60], or measure ϵ [23] (see [24] for a systematic review of different consistency measures). In [24], the authors compared different generalizations of DRSA w.r.t. four desirable monotonicity properties denoted by (m1)–(m4). This comparison and the results of subsequent computational experiments described in [20, 25, 26], concerning multicriteria sorting problems, showed the advantage of VC-DRSA with consistency measure ϵ [20, 23, 24], denoted by ϵ-VC-DRSA.

1.1.2 Similarity-based Classification Problem

People tend to solve new problems using the solutions of *similar* problems encountered in the past. This process if often referred to as *similarity-based reasoning* or *case-based reasoning* (CBR). As observed by Gilboa and Schmeidler [74], the basic idea of CBR can be found in the following sentence of Hume [104]: "From causes which appear similar we expect similar effects. This is the sum of all our experimental conclusions." We can rephrase this sentence by saying: "The more similar are the causes, the more similar one expects the effects".

In this thesis, we consider classification performed according to the (broadly construed) CBR paradigm, i.e., a similarity-based classification. Precisely, we consider the following classification problem setting. There is given a finite set of objects U (called *universe of discourse*, or *case base*) and a finite family of pre-defined decision classes $\mathcal{D}$. An object $y \in U$ (a "case") is described in terms of features $f_1, \ldots, f_n \in F$. For each feature $f_i \in F$, there is given a *marginal similarity function* $\sigma_{f_i} : U \times U \rightarrow [0, 1]$, such that the value $\sigma_{f_i}(y, x)$ expresses the similarity of object $y \in U$ to object $x \in U$ w.r.t. feature f_i; the minimal requirement that function σ_{f_i} must satisfy is the following: for all $x, y \in U$, $\sigma_{f_i}(y, x) = 1 \Leftrightarrow y$ and x have the same value of feature f_i. Moreover, for each object $y \in U$ there is given an information concerning (normalized) *credibility* of its membership to each of the considered classes. To admit graded credibilities, each decision class $X \in \mathcal{D}$ is modeled as a fuzzy set in U [181], characterized by membership function $\mu_X : U \rightarrow [0, 1]$. Thus, each object $y \in U$ can belong to different decision classes with different degrees of membership. The aim of decision aiding is to present to a DM a recommendation concerning a new object z, in terms of a degree of membership of this object to particular classes.

In ML, problems involving graded membership to decision classes are often referred to as *soft label classification* problems (see, e.g., [174]). Obviously, when $\mu_X \in \{0, 1\}$ for each $X \in \mathcal{D}$, then a soft classification problem boils down to a regular classification problem with crisp classes.

Some exemplary soft label classification problems can be found in [132] and [167]. In the first paper, soft labels (membership degrees to two considered decision classes) were given by doctors who judged, looking at patient data, how strongly they feel a patient is at risk of developing heparin-induced thrombocytopenia (HIT). The second paper concerns detection and classification of emotions in recordings of spoken sentences, where it is very natural to have multiple emotions to varying degrees at the same time.

Soft label classification relates to the analysis of so-called *compositional data* concerning *mixtures* [172]. An exemplary problem given in this reference concerns analysis of soil samples w.r.t. fraction of sand, silt, and clay, where each sample is characterized by some description attributes, e.g., the depth at which the sample was taken. One of the other problems considered in there concerns fatty acid composition of milk obtained from lactating cows subjected to different diets. In compositional data analysis, however, the outputs (reflecting fractions of components in a mixture) are obviously dependent, and the sum of fractions over all outputs is equal to one. On the other hand, in practical soft label classification problems, the outputs may be independent and the input credibilities of membership to different classes (which are often elicited by humans) need not to sum up to one. As an example of such a case, one could consider a situation where a doctor has to express the credibility that a patient has disease X, Y, or Z. Obviously, the patient can suffer from several diseases simultaneously, so the sum of credibilities may exceed one.

1.2 Review of Existing Approaches to Multicriteria Ranking

Ranking problem has been considered both in MCDA and PL. Below, we review approaches proposed in these research fields to handle this problem. Considered methods differ mainly by the type of input preference information and by the form of employed preference model. For the multicriteria nature of the considered ranking problem, we focus mainly on MCDA methods since they make explicit use of the domain knowledge concerning order of evaluations on the scales of criteria. In this way, they "produce" preference models that are concordant with the DM's value system and domain knowledge concerning ordinal nature of considered criteria.

1.2.1 MCDA Approaches to Multicriteria Ranking

In the field of MCDA, multicriteria ranking problem has been handled by many decision aiding methods. Some of these methods require from a DM a direct specification of preference model parameters, and others are based on induction of a preference model from decision examples. There are also several "hybrid" methods in which some preference model parameters are elicited directly from a DM, whereas other parameters are calculated based on decision examples provided by the DM. Below, we review different MCDA methods that recommend to a DM a total preorder (i.e., a linear ranking of objects, with possible ties) or a partial preorder (i.e., a ranking of objects, with possible ties and incomparabilities) on the set of objects, underlining the form of applied preference model (such as additive utility function, outranking relation, or a set of if-then decision rules), the type of preference information required to build this model, and some other aspects related to application of these methods. We also give several notes concerning limitations and shortcomings of the reviewed methods.

Multiple Attribute Utility Theory (MAUT).

The goal of MAUT [117] is to represent the preferences of a DM on a set of objects A by a *value* (*utility*) *function* $U(g_1(\cdot), \ldots, g_n(\cdot))\colon \mathbb{R}^n \to \mathbb{R}$, written in short as $U(\cdot)$, such that $a \succ b$ (a is preferred to b) if $U(a) > U(b)$, whereas $a \sim b$ (a is indifferent to b) if $U(a) = U(b)$. The main preference model of MAUT is the *additive value function*

$$U(a) = \sum_{i=1}^{n} u_i(a), \text{ with } a \in A, \tag{1.1}$$

where $U(a)$ is a short form of $U(g_1(a), \ldots, g_n(a))$, $u_i(a)$ is a short form of $u_i(g_i(a))$, and u_i, $i \in \{1, \ldots, n\}$, are monotone *marginal value functions*.

It is important to note that the use of the additive value function involves *compensation* between criteria (i.e., a poor evaluation of an object on some criterion can be compensated by its better evaluations on some other criteria). It also requires a rather strong assumption concerning mutual independence of criteria in the sense of preference.

Over the years, several methods have been proposed within the framework of MAUT for the multicriteria ranking problem, as discussed below.

- **ASSESS** method [52, 117]. The ASSESS method is based on interactive elicitation of preference information by a DM. The DM is asked a series of questions concerning the choice between two options: participation in a given lottery (with assumed probability of win and loss), and obtainment of some certain value offered without any risk. Questions of this type are used to construct piecewise linear marginal utility functions u_i as well as to determine criteria weights k_i, $i = 1, \ldots, n$. The assessed value function U is defined as

  $$KU(a) + 1 = \prod_{i=1}^{n} \left(k_i K u_i(a) + 1\right),$$

  where K is a scaling coefficient. Once value function is constructed, it directly implies the ranking of all objects from set A.

  It is important to note that assessing piecewise linear marginal value function is problematic in case of a purely ordinal criterion since it requires an arbitrary conversion of the ordinal scale to a cardinal scale. Moreover, practice shows that the way of gathering preference information adopted in the ASSESS method may require much effort on the part of the DM as the number of questions may be significant.

- **UTA** (UTilités Additives) method [109, 154]. The UTA method, and its improved version UTASTAR [155], introduce the disaggregation-aggregation paradigm. They employ input preference information in terms of a ranking (total preorder) on a subset $A^R \subseteq A$ of so-called *reference objects*, i.e., objects from set A which are relatively well-known to the DM. This preference information is used to formulate constraints of a linear programming (LP) problem solved to calculate coordinates of characteristic points of monotone piecewise linear marginal value functions that when plugged into (1.1), can reproduce given ranking of reference objects either exactly or approximately.

  When the preferences of a DM can be modeled using (1.1), there are usually many value functions compatible with the ranking of reference objects supplied by the DM. Obviously, the rankings of objects from set A yielded by these value functions can be quite different. In such a case, there is no just single recommendation to be presented to the DM. Different approaches have been proposed to avoid an arbitrary choice of a compatible value function. The first approach consists in leaving this choice to the DM, possibly assisted by a software designed to interactively modify graphically presented marginal value functions; these functions can be changed within allowed limits that result from the solution of appropriate ordinal regression problems. Such a software assistance has been provided for the first time in UTA+ [120]. The second approach involves some predefined rules to choose a so-called *representative value function* (see, e.g., [17, 27, 114, 115, 155]). The third approach aims at drawing so-called *robust conclusions* by considering simultaneously all compatible value functions. This technique is called *robust ordinal regression* (ROR) [44, 94, 96]. Two utility-based methods that follow the robust ordinal regression paradigm, UTA$^{\text{GMS}}$ and GRIP, are discussed next.

  It is useful to remind, that in case of the original UTA method, using piecewise linear marginal value functions raises a concern when such a function is calculated for an ordinal criterion. In such a case, to avoid not meaningful interpolation of marginal utility, it is necessary to calculate this utility for each value in the value set of the ordinal criterion.

- **UTA$^{\text{GMS}}$** method [94]. The UTA$^{\text{GMS}}$ method employs the ROR paradigm. It extends the UTA method in several ways, e.g., by dropping the assumption that marginal value functions are piecewise linear and by considering preference information also in the form of pairwise comparisons of reference objects in terms of preference and indifference relations. When the set of value functions compatible with the preference information given by a DM is nonempty, the method considers simultaneously all compatible value functions in order to determine two weak preference relations – *necessary weak preference relation* $\succsim^N$ being a partial preorder, and *possible weak preference relation* $\succsim^P$ which is strongly complete and negatively transitive. The former relation holds for a pair of objects $(a, b) \in A \times A$ if $U(a) \geq U(b)$ for *all* compatible value functions. The latter relation holds for a pair of objects $(a, b) \in A \times A$ if $U(a) \geq U(b)$ for *at least one* compatible value function. The UTA$^{\text{GMS}}$ method is meant to be used interactively. As the DM enriches her/his preference information, new pairs of objects are added to the necessary relation.

Applying the UTA$^{\text{GMS}}$ method, one does not get a ranking which is a total preorder of objects from set A. To obtain such a ranking, one can choose a representative compatible value function and apply this function on set A. Procedures for calculating representative value functions have been proposed, e.g., in [114, 115]. Obviously, if one accepts the final ranking to be a partial preorder, one can simply use relation $\succsim^N$.

It is worth noting that it is possible that given the preference information of a DM, there is no value function compatible with this information (the case of incompatibility). The reasons for this situation and different ways of handling it are discussed in [94].

- **GRIP** (Generalized Regression with Intensities of Preference) method [56]. The GRIP method is a generalization of the UTA$^{\text{GMS}}$ method that accounts for ordinal intensities of preference between some pairs of reference objects, either in partial or comprehensive comparisons. For instance, a DM may state that "a is preferred to b at least as much as c is preferred to d", referring to their comprehensive evaluation or just to a single criterion $g_i \in G$. The ways of obtaining a recommendation in terms of a total preorder of all objects from set A are analogous to those used for the UTA$^{\text{GMS}}$ method.

- **AHP** (Analytic Hierarchy Process) method [150, 151]. The AHP method employs preference information in terms of pairwise comparisons of all elements present at each level of hierarchy used to model considered decision problem. Thus, it involves pairwise comparisons of the first level criteria, pairwise comparisons of the second level sub-criteria, etc., and pairwise comparisons of objects from set A w.r.t. each sub-criterion from the lowest level. For any two elements, say e_1, e_2, compared pairwise, the DM has to quantify the intensity of preference of e_1 over e_2, by choosing, in fact, a value from the available set $\{\frac{1}{9}, \frac{1}{7}, \frac{1}{5}, \frac{1}{3}, 1, 3, 5, 7, 9\}$. This information is used to calculate weight (priority) of each node of the hierarchy. Then, utility $U(a)$ of each object $a \in A$ is calculated as a sum of products of weights along each path going from the root node to one of the leaf nodes corresponding to a.

  An advantage of AHP is the ability to decompose complex problems into subproblems. However, this method has also relatively many weak points, such as: the need of performing relatively many pairwise comparisons of elements of considered hierarchy, known cases of rank-reversals, difficulty of ensuring global consistency of pairwise comparisons, treatment of ordinal criteria in a "cardinal fashion", unstable character of absolute zero on the employed ratio scale of the intensity of preference, and failure to satisfy the "Condition of Order Preservation (COP)" [6].

- **PAPRIKA** (Potentially All Pairwise RanKings of all possible Alternatives) method [98]. The PAPRIKA method uses additive value function to assess overall utility of each considered object. The overall utility of an object $a \in A$ is calculated as a sum of its *points* over all criteria. The number of points that object a gets for its evaluation on criterion $g_i \in G$ depends on the fact to which so-called *performance category* of g_i the evaluation $g_i(a)$ belongs. In case g_i is an ordinal criterion, its performance categories usually correspond to particular evaluations in the value set of g_i. In case g_i is a cardinal criterion, its performance categories, usually, correspond to intervals defined on the scale of criterion g_i (which requires prior discretization). In order to derive points for each performance category of each criterion, the PAPRIKA method employs preference information in terms of pairwise comparisons of objects from considered set A. A DM is asked a series of questions concerning her/his preference about pairs of mutually non-dominated objects (i.e., such pairs of $a, b \in A$ that neither a dominates b nor b dominates a). For each such pair the DM may decide that (s)he prefers a over b, or b over a, or that a and b are equally good, or that (s)he wants to skip a given pair. The objective of the method is to require from the DM the fewest pairwise comparisons possible (thereby minimizing her/his cognitive effort). Therefore, during the iterative process of comparing objects in a pairwise fashion, some pairwise comparisons are implicitly assessed by the method to ensure transitivity of calculated additive value function and to significantly decrease the number of pairwise comparisons that the DM needs to make. Once the DM is done with pairwise comparisons, points of considered performance categories are calculated by linear programming so as the calculated value function reproduces pairwise comparisons supplied by the DM (multiple solutions are possible at this point).

  It is worth noting that when using PAPRIKA, the number of pairwise comparisons that need to be supplied by the DM may be substantial, especially in case of many criteria with multiple performance categories. The second concern is the potential non-uniqueness of obtained solutions. The third issue is the discretization of scales of cardinal criteria, which often has to be performed to reduce the number of required pairwise comparisons. Finally, the calculated preference model depends on the order in which pairs of nondominated objects are presented to the DM (due to implicit assessments of pairwise comparisons performed by the method).

Outranking Methods

Outranking methods construct the preference model in the form of a so-called *outranking relation*, usually denoted by S. This is a binary relation over the considered set of objects A. If aSb, $a, b \in A$, then a is considered to be (comprehensively) at least as good as b. Outranking relation is reflexive, but not transitive nor complete, in general.

There exist various outranking methods that differ by the way of constructing the outranking relation and by the way of subsequent *exploitation* of this relation to give final recommendation. These methods mainly belong to one of the two families of outranking-based methods developed by the MCDA community: ELECTRE [54, 55, 57] and PROMETHEE [10, 37, 38]. Below, we give a general overview of these two families, and we review some particular methods that are designed to deal with the multicriteria ranking problem.

ELECTRE (ELimination Et Choix Traduisant la REalité – ELimination and Choice Expressing the REality) family methods [54, 55, 57].

In the ELECTRE family methods, the outranking relation S is considered to hold for $a, b \in A$ if there are significant arguments that support the assertion aSb, while the arguments to reject this assertion are not essential. The consideration of the arguments for and against the outranking relation is concordant with the intuitive reasoning of most DMs, which contributes to the popularity of outranking methods. On the other hand, due to the general properties of the outranking relation, i.e., non-transitivity and non-completeness, which allow incomparability of objects, the results that follow from using outranking methods are often less conclusive than the results obtained using MAUT (e.g., one may get a ranking in form of a partial preorder rather than a total preorder).

All methods from the ELECTRE family are based on the same principle. For each ordered pair of objects, they perform a *concordance test* and a *non-discordance* test to check whether the outranking relation holds for this pair. Once determined, the outranking relation is then exploited in a proper way, depending on the considered problem.

ELECTRE family methods make use of the concept of a *pseudo-criterion* [146]. A criterion g_i is a pseudo-criterion if it is associated with an *indifference threshold* q_i and a *preference threshold* p_i, $i \in \{1, \ldots, n\}$. The idea behind these thresholds is to take into account the imperfection of evaluations.

ELECTRE family methods involve moreover two sets of parameters in the construction of an outranking relation for pairs of objects: *importance coefficients* (*weights*) of criteria k_i and *veto thresholds* v_i, $i \in \{1, \ldots, n\}$. One other technical parameter that is used is the *concordance cutting level* $\lambda \in [0.5, 1]$. Weight k_i, $i \in \{1, \ldots, n\}$, reflects the voting power of criterion g_i when this criterion is in the coalition of criteria voting for the assertion aSb. The role of a veto threshold v_i is to block assertion aSb in case when there is a relatively large difference between evaluations $g_i(a)$ and $g_i(b)$, in disfavor of object a.

In the concordance test for aSb, one calculates *concordance index* $C(a, b)$ as the weighted average of *marginal concordance indices* $c_i(a,b)$:

$$C(a,b) = \frac{\sum_{i=1}^{n} k_i c_i(a,b)}{\sum_{i=1}^{n} k_i},$$

where $c_i(a,b)$ is:

- equal to 1, if $g_i(a)$ is either not worse than $g_i(b)$ or just insignificantly worse, i.e., $g_i(b)$ is better than $g_i(a)$ but the difference between these evaluations is not more than q_i,
- equal to 0, if $g_i(a)$ is clearly worse than $g_i(b)$, i.e., $g_i(b)$ is better than $g_i(a)$ and the difference between these evaluations is p_i or more,
- in the interval $[0,1]$ and calculated from linear interpolation, if $g_i(b)$ is better than $g_i(a)$ and the difference between these evaluations is between q_i and p_i.

The result of the concordance test for pair $(a,b) \in A \times A$ is positive if $C(a,b) \geq \lambda$; otherwise the result is negative, in which case the hypothesis aSb is refuted.

If the result of the concordance test is positive for $(a,b) \in A \times A$, the non-discordance test is performed. The result of this test is "by default" positive, unless a veto occurs for some criterion $g_i \in G$, in which case one refutes the hypothesis aSb.

Some methods from the ELECTRE family, in particular those designed for dealing with the multicriteria ranking problem, involve computation of a *credibility degree* of the outranking relation rather than considering this relation to be crisp. The value of the credibility degree reflects both the strength of the coalition of criteria supporting conclusion aSb, and the strength of the criteria that oppose this conclusion. The credibility degree, denoted by $\sigma(a,b)$, is usually defined as follows (see [180]):

$$\sigma(a,b) = C(a,b) \cdot \prod_{j \in F} \frac{1 - d_j(a,b)}{1 - C(a,b)}, \tag{1.2}$$

where $F = \{i : 1 \leq i \leq n,\ d_i(a,b) > C(a,b)\}$ and d_j, $j \in \{1,\ldots,n\}$, denotes *marginal discordance index* that is:

- equal to 0, if $g_j(a)$ is not worse than $g_j(b)$ by more than p_j,
- equal to 1, if $g_j(a)$ is worse than $g_j(b)$ by at least v_j,
- in the interval $[0,1]$ and calculated from linear interpolation, if $g_j(b)$ is better than $g_j(a)$ and the difference between these evaluations belongs to the interval $[p_j, v_j]$.

The credibility degree makes S a valued outranking relation.

- **ELECTRE II-IV** methods. The first method in the ELECTRE family designed to deal with the multicriteria ranking problem was ELECTRE II [141]. Next method was ELECTRE III that improved upon ELECTRE II by introducing pseudo-criteria and valued outranking relation [142]. In this method, first, two total preorders are built according to two variants of the *distillation procedure*, ascending and descending. Second, a partial preorder is created as the intersection of the two total preorders. Third, a so-called *median order* (being a total preorder) is build using *ranks* of objects in the partial preorder; a rank of object $a \in A$ is the length of the path in the graph of partial preorder from the best object to object a. Yet another method is ELECTRE IV. It is based on the construction of five embedded outranking relations of different credibility [103]. It does not requires from a DM the specification of weights k_i. The ELECTRE IV exploitation procedure is the same as in ELECTRE III.

  It is important to note that due to the amount of parameters required by the above methods to built an outranking relation, and due to the need of fixing precise numerical values of these parameters, these methods are rather difficult to use and require from a DM the knowledge of many technical details.

- **ELECTRE$^{\mathbf{GKMS}}$** method [76, 114]. The ELECTRE$^{\text{GKMS}}$ method implements robust ordinal regression in multicriteria ranking and choice ELECTRE methods. The method reduces cognitive effort of a DM as it accepts preference information in terms of pairwise comparisons of reference objects. For a pair of reference objects $(a, b) \in A \times A$, the DM can state the truth or the falsity of the outranking relation. Apart from pairwise comparisons, the DM should supply also the intra-criteria preference information (i.e., indifference and preference thresholds), however, the method admits that the DM can give intervals instead of precise numerical values. Alternatively, the DM can compare some pairs of reference objects w.r.t. particular criteria which conditions the comparison thresholds. The above preference information is used to define a set of constraints on *outranking models* that are compatible with the preference information. A single outranking model is a set of precise values of model parameters, i.e., thresholds q_i, p_i, v_i, and weights k_i, $i = 1, \ldots, n$, as well as concordance cutting level λ. It defines also the shape of each marginal concordance index $c_i(a, b)$, $i \in \{1, \ldots, n\}$, in the interval corresponding to the situation when $g_i(b)$ is better than $g_i(a)$ and the difference of these evaluations is between q_i and p_i. Contrary to the "traditional" methods from ELECTRE family (e.g., ELECTRE III), in ELECTRE$^{\text{GKMS}}$ method, no linear characteristic of $c_i(a, b)$ is assumed in this interval (only monotonicity of $c_i(a, b)$ is required).

  ELECTRE$^{\text{GKMS}}$ method involves calculation of *necessary outranking relation* S^N and *possible outranking relation* S^P. The first relation holds for $(a, b) \in A \times A$ such that aSb for every outranking model compatible with the preference information given by the DM. The second relation holds for $(a, b) \in A \times A$ such that aSb for at least one outranking model compatible with the preference information given by the DM. Calculation of relations S^N and S^P requires solving a series of mixed-integer linear programming (MILP) problems. Remark that $S^P \supseteq S^N$, and both these relations are reflexive, but, in general, they are neither transitive nor complete.

It is worth noting that it is possible that there is no outranking model compatible with the preference information (the case of so-called incompatibility). Then, it is necessary to identify the troublesome pieces of preference information in order to remove this incompatibility [114].

In order to get a total preorder of objects from set A, one needs to exploit the necessary and possible outranking relations. This can be done, e.g., in the way proposed in [114], by calculating a net flow score (NFS) for each object $a \in A$. Namely, $NFS(a)$ is the number of objects $b \in A, b \neq a$, such that aS^Nb or not(bS^Pa) diminished by the number of objects $b \in A, b \neq a$, such that bS^Na or not(aS^Pb). Another approach consists in applying a procedure for selection of a single so-called *representative set of parameters* (representative outranking model), as proposed in [114].

PROMETHEE (Preference Ranking Organization METHod for Enrichment of Evaluations) family methods [10, 37, 38].

PROMETHEE is a family of outranking-based methods that require from a DM to specify for each criterion $g_i \in G$:

- weight k_i of this criterion,
- *preference function* (also called *marginal preference index*) $\pi_i(a, b)$ being monotonically non-decreasing function w.r.t. the difference $d_i(a, b) = g_i(a) - g_i(b)$.

Six types of preference functions have been proposed. Thus, specification of a preference function consists in choosing one type among these six types and, if necessary, giving concrete values of parameters characteristic for the chosen type, e.g., indifference threshold q_i and preference thresholds p_i for the "V-criterion with indifference area" type. In general, preference functions are normalized, and defined so that $\pi_i(a, b) = 0$ if $d_i(a, b) \leq q_i$, and $\pi_i(a, b) = 1$ if $d_i(a, b) \geq p_i$, where $p_i \geq q_i \geq 0, i \in \{1, \ldots, n\}$.

It is worth noting, that some approaches for the PROMETHEE family methods aim at calculation of weights k_i of criteria using indirect preference information in terms of pairwise comparisons of reference objects [70, 182].

To quantify the overall preference of object a over object b, one calculates the *preference index* $\pi(a, b)$ defined as

$$\pi(a, b) = \sum_{i=1}^{n} k_i \cdot \pi_i(a, b).$$

PROMETHEE family methods involve calculation of a *positive outranking flow* $\Phi^+(a)$ and a *negative outranking flow* $\Phi^-(a)$ for every $a \in A$. These flows are defined as:

$$\Phi^+(a) = \frac{1}{n-1} \sum_{b \in A} \pi(a, b),$$
$$\Phi^-(a) = \frac{1}{n-1} \sum_{b \in A} \pi(b, a).$$

The balance between the above flows is reflected by the net outranking flow

$$\Phi(a) = \Phi^+(a) - \Phi^-(a).$$

- **PROMETHEE I** method. In the PROMETHEE I method, the final recommendation is a partial preorder resulting from the intersection of the ranking implied by positive outranking flows $\Phi^+(\cdot)$ with the ranking implied by negative outranking flows $\Phi^-(\cdot)$.

- **PROMETHEE II** method. In the PROMETHEE II method, the final recommendation is a total preorder implied by net outranking flows $\Phi(\cdot)$.

- **PROMETHEE$^{\mathbf{GKS}}$** method [114]. The PROMETHEE$^{\mathrm{GKS}}$ method implements robust ordinal regression in multicriteria ranking PROMETHEE methods. The method accepts preference information in terms of pairwise comparisons of reference objects. For a pair of reference objects $(a, b) \in A \times A$, the DM can state the truth of the weak preference relation ($a \succsim b$), strict preference relation ($a \succ b$), or indifference relation ($a \sim b$). These relations can be defined at two levels: (1) at the *level of construction of the outranking relation*, and (2) at the *level of exploitation of the outranking relation.* In the first case, the statements of the DM are translated into relations between preference indices $\pi(a, b)$ and $\pi(b, a)$. In the second case, they are translated into relations between net outranking flows $\Phi(a)$ and $\Phi(b)$. It is important to underline, that the above distinction of two levels of relations is unnecessary in approaches based on MAUT, where constructed value function $U(a)$ directly implies a total preorder.

  Apart from pairwise comparisons, the DM should also supply the intra-criteria preference information (i.e., indifference and preference thresholds), however, the method allows the DM to give intervals instead of precise numerical values. Alternatively, the DM can compare some pairs of reference objects w.r.t. particular criteria.

The above preference information is used to define a set of constraints on outranking models that are compatible with this preference information. In this case, a single outranking model is a set of precise values of considered parameters, i.e., thresholds q_i, p_i, and weights k_i, $i = 1, \ldots, n$. It defines also the shape of each preference function $\pi_i(a, b)$, $i \in \{1, \ldots, n\}$.

PROMETHEE$^{\text{GKS}}$ method takes into account all outranking models compatible with the preference information to calculate *necessary outranking relation* $\succsim^N$ and *possible outranking relation* $\succsim^P$, either at construction level or exploitation level. These relations are obtained by solving a series of linear programming problems. The first relation holds for $(a, b) \in A \times A$ such that a outranks b for every outranking model compatible with the preference information given by the DM. The second relation holds for $(a, b) \in A \times A$ such that a outranks b for at least one compatible outranking model.

It is worth noting that it is possible that there is no outranking model compatible with the preference information (the case of so-called incompatibility). Then, it is necessary to identify the troublesome pieces of preference information in order to remove this incompatibility, analogously to ELECTRE^{GKMS} method.

To get a total preorder of objects from set A, one needs to exploit the necessary and possible outranking relations. This can be done in the same way as in case of ELECTRE^{GKMS} method.

Decision Rule-based Methods

Decision rule-based methods construct a logical preference model in terms of a set of if-then decision rules. As results from the discussion in Section 1.1.1, rule model is the most general preference model. Below, we review several rule-based MCDA approaches proposed in the literature to deal with the multicriteria ranking problem. They employ indirect preference information in terms of pairwise comparisons of some reference objects, given by a DM. Such decision examples are represented in a so-called *pairwise comparison table* (PCT) [77, 80] and processed using DRSA or VC-DRSA to handle possible inconsistencies and to generate a set of decision rules representing preferences of the DM that gave the pairwise comparisons. The induced rules are then applied on a set A of objects to be ranked and the resulting preference structure is exploited using some ranking procedure to get a final ranking (total preorder). Thus, we can distinguish several key steps, which are:

(s_1) elicitation of preference information in terms of pairwise comparisons of some reference objects,

(s_2) rough approximation of comprehensive relations implied by the pairwise comparisons,

(s_3) induction of decision rules from rough approximations of considered comprehensive relations,

(s_4) application of induced decision rules on set $M \subseteq A$, and

(s_5) exploitation of the resulting preference structure on set M to get a ranking of objects.

The approaches reviewed below differ by the way of performing one or several of these steps.

One common feature of the rule-based approaches to be reviewed concerns the assumption that for each cardinal criterion $g_i \in G$ (i.e., criterion with a cardinal scale, for which it is meaningful to consider intensity of preference) there is given a set of *graded preference relations* $T_i = \{P_i^h, h \in H_i\}$, where H_i is a finite set of integer numbers ("grades of intensity of preference") (see, e.g., [84]). Relations P_i^h are binary relations over A, such that

- if $aP_i^h b$ and $h > 0$, then object a is preferred to object b by degree h w.r.t. criterion g_i,
- if $aP_i^h b$ and $h < 0$, then object a is not preferred to object b by degree h w.r.t. criterion g_i,
- if $aP_i^h b$ and $h = 0$, then object a is similar (asymmetrically indifferent) to object b w.r.t. criterion g_i.

According to [84], the modeling of binary relations P_i^h involves: (i) choosing a function $k_i : \mathbb{R}^2 \rightarrow \mathbb{R}$ which measures the *strength of preference* (positive or negative) of object a over object b, taking into account evaluations $g_i(a), g_i(b)$, $a, b \in A$, (ii) dividing the codomain of function k_i using a suitable set of thresholds, and (iii) numbering the resulting intervals by considered grades $h \in H_i$.

It is important to note that the modeling of binary relations P_i^h, involving determination of several thresholds for each cardinal criterion, may be considered impractical. After all, the thresholds are parameters of the constructed rule-preference model. Thus, the need of defining these thresholds is in fact in the opposition to the idea of employing decision examples.

- The first rule-based MCDA approach to multicriteria ranking problem considered in [78, 79, 81, 82, 93] and reminded in [83, 159], denoted by α, is characterized by the following steps:

  (s_1^α) the pairwise comparisons of reference objects are expressed in terms of outranking and non-outranking relations; given a pair of objects $(a, b) \in A \times A$, a DM may: (i) state that object a is comprehensively at least as good as object b (or, in other words, a outranks b), denoted by aSb, (ii) state that object a

is comprehensively not at least as good as object b (or, in other words, a does not outrank b), denoted by aS^cb, or (iii) abstain from any judgment;

(s_2^α) relations S and S^c are approximated using *graded dominance relations* (called in the following *single-graded dominance relations*) w.r.t. the set of criteria G;

(s_3^α) the approximations of S and S^c are used to generate four types of single-graded decision rules (i.e., concerning the same grade of preference w.r.t. each criterion present in the rule condition part), denoted by D_{++}, D_{-+}, D_{+-}, D_{--}; if a pair of objects $(a, b) \in A \times A$ is covered by a rule of the first two types, it is concluded that aSb, while if it is covered by a rule of the last two types, the conclusion is aS^cb;

(s_4^α) the application of induced rules on set $M \subseteq A$ yields four outranking relations called *true outranking relation*, *false outranking relation*, *contradictory outranking relation*, and *unknown outranking relation*, which together constitute so-called *four-valued outranking* [170, 171];

(s_5^α) the final ranking of objects from set $M \subseteq A$ is obtained using their so-called *net flow scores*; the net flow score of an object $a \in M$, denoted by $S^{NF}(a)$, is calculated as the sum of:

(i) the number of objects $b \in M$ such that the induced rules yield conclusion aSb, and

(ii) the number of objects $b \in M$ such that the induced rules yield conclusion bS^ca,

diminished by the sum of:

(iii) the number of objects $b \in M$ such that the induced rules yield conclusion bSa, and

(iv) the number of objects $b \in M$ such that the induced rules yield conclusion aS^cb.

It is worth noting that the first approach presented in [78, 79, 81, 82] does not account for ordinal criteria (i.e., criteria with ordinal scale, for which consideration of intensity of preference is not meaningful). Moreover, the single-graded dominance relation is too restrictive as it assumes a common grade of intensity of preference for all considered criteria.

- The second rule-based MCDA approach to multicriteria ranking problem, introduced and characterized in [83, 84, 158–160], denoted by β, comprises of the following steps:

(s_1^β) $\equiv$ (s_1^α);

(s_2^β) relations S and S^c are approximated using the dominance relation that accounts for both cardinal and ordinal criteria; w.r.t. cardinal criteria, the *multi-graded dominance relation* is considered;

(s_3^β) the approximations of S and S^c are used to generate three types of decision rules (that can use different grades of preference w.r.t. each cardinal criterion present in the rule condition part), denoted by $D_\geq$, $D_\leq$, and $D_{\geq\leq}$; if a pair of objects $(a, b) \in A \times A$ is covered by a rule of the first type, it is concluded that aSb, while if it is covered by a rule of the second type, the conclusion is aS^cb;

(s_4^β) $\equiv$ (s_4^α);

(s_5^β) $\equiv$ (s_5^α).

It is worth noting that definitions of lower approximations applied in approaches α and β appear to be too restrictive in practical applications. In consequence, lower approximations of S and S^c are often small or even empty, preventing a good generalization of pairwise comparisons in terms of decision rules.

- The third rule-based MCDA approach to multicriteria ranking problem, presented in [86, 158], denoted by γ, is characterized by the following steps:

(s_1^γ) $\equiv$ (s_1^β);

(s_2^γ) relations S and S^c are approximated using the dominance relation that accounts for both cardinal and ordinal criteria; w.r.t. cardinal criteria, the multi-graded dominance relation is considered; contrary to step (s_2^β), the approximations of S and S^c are calculated using a PCT-oriented adaptation of the VC-DRSA proposed originally in [97] w.r.t. the multicriteria classification problems; as this VC-DRSA measures consistency of decision examples using rough membership measure μ, it will be denoted by μ-VC-DRSA;

(s_3^γ) the lower approximations of S and S^c are used to generate two types of probabilistic decision rules (that can use different grades of preference w.r.t. each cardinal criterion present in the rule condition part), denoted by $D_\geq$ and $D_\leq$; if a pair of objects $(a, b) \in A \times A$ is covered by a rule of the first type, it is concluded that aSb, while if it is covered by a rule of the second type, the conclusion is aS^cb;

(s_4^γ) $\equiv$ (s_4^β);

(s_5^γ) $\equiv$ (s_5^β).

- The fourth rule-based MCDA approach to multicriteria ranking problem, denoted by δ, was introduced in [63]. It is characterized by the following steps:

(s_1^δ) the pairwise comparisons of reference objects are expressed in terms of *comprehensive graded preference relations* $\succ^h$, $h \subset [-1, 1]$; given a pair of objects

$(a, b) \in A \times A$, a DM may: (i) state that object a is comprehensively preferred to object b in grade h, i.e., $a \succ^h b$ with $h > 0$, (ii) state that object a is comprehensively *not* preferred to object b in grade h, i.e., $a \succ^h b$ with $h < 0$, (iii) state that object a is comprehensively indifferent to object b, i.e., $a \succ^0 b$, or (iv) abstain from any judgment;

(s_2^δ) *upward cumulated preference relations* (upward unions of comprehensive graded preference relations) $\succ^{\geq h}$ and *downward cumulated preference relations* (downward unions of comprehensive graded preference relations) $\succ^{\leq h}$ are approximated using the dominance relation that accounts for both cardinal and ordinal criteria; w.r.t. cardinal criteria, the multigraded dominance relation is considered; analogously to step (s_2^γ), the approximations of $\succ^{\geq h}$ and $\succ^{\leq h}$ are calculated using a PCT-oriented adaptation of μ-VC-DRSA proposed in [97];

(s_3^δ) the lower approximations of $\succ^{\geq h}$ and $\succ^{\leq h}$ are used to generate two types of *probabilistic decision rules* (that can use different grades of preference w.r.t. each cardinal criterion present in the rule condition part), denoted by $D_{\geq}$ and $D_{\leq}$; each induced rule is additionally characterized by the attained *confidence level*; if a pair of objects $(a, b) \in A \times A$ is covered by a rule of the first type, it is concluded that $a \succ^{\geq h} b$, while if it is covered by a rule of the second type, the conclusion is $a \succ^{\leq h} b$;

(s_4^δ) the application of induced rules on set $M \subseteq A$ yields a graded fuzzy preference relation (of level 2) over M; this relation is graded because of different grades of preference, but it is also fuzzy because of different confidence levels of rules matching pairs of objects from $M \times M$;

(s_5^δ) the final ranking of objects from set $M \subseteq A$ is obtained by exploitation of the preference structure on M using either the Weighted Fuzzy Net Flow Score (WFNFS) procedure or a Lexicographic-fuzzy Net Flow Score procedure.

Remark that approach δ was partially considered also in [86], although only up to step (s_2^δ). Moreover, an early version of approach δ can be found in [62]. This early version, however, features an exploitation procedure that may produce a ranking that does not respect the dominance relation over set $M \subseteq A$ of ranked objects.

It is worth noting that elicitation of preferences in terms of comprehensive graded preference relations $\succ^h$ requires, in general, greater cognitive effort of a DM. Moreover, it complicates exploitation of the preference structure resulting from application of induced decision rules.

It is also important to note that the application of variable consistency model of DRSA considered in approaches γ and δ, relying on rough membership consistency measure μ, leads to the situation when calculated lower approximations of considered comprehensive relations lack several desirable monotonicity properties, as proved in [24].

Other Methods

Below, we review some other relatively simple MCDA methods to multicriteria ranking that do not exactly fit to the three categories of methods considered above.

Weighted sum model (WSM) [58, 168]. WSM is a very simple model where each criterion $g_i \in G$ is assigned a weight w_i (substitution rate) and the comprehensive utility of an object $a \in A$ is calculated as the weighted sum $\sum_{i=1}^{n} w_i \cdot g_i(a)$.

It is important to note that this model admits compensation between criteria. Moreover, it can be meaningfully applied only when all the criteria are expressed in exactly the same unit. As to a DM, to construct the preference model, (s)he needs to give criteria weights, which is not an easy task, in general.

Weighted product model (WPM) [39, 127, 168]. When using WPM, each object is compared with the others by multiplying a number of ratios, one for each criterion. Each ratio is raised to the power equivalent to the weight of the corresponding criterion. Assuming that all criteria need to be maximized, in order to compare pairwise objects $a, b \in A$, the following product has to be calculated:

$$P(a,b) = \prod_{i=1}^{n} \left(\frac{g_i(a)}{g_i(b)} \right)^{w_i}.$$

If $P(a,b)$ is greater than 1, then object a is more desirable than object b (in the maximization case).

The WPM is often called *dimensionless analysis* because its mathematical structure eliminates any units of measure. Thus, contrary to the WSM, this model can be used also in case of incommensurable scales of criteria. However, it still admits full compensation between criteria (which is not desirable in many practical applications) and requires a set of weights to be given a priori by a DM. Moreover, there are two issues concerning the above equation, i.e., evaluations equal to zero (problem with division) and presence of criteria that need to be minimized. Finally, the sole values of “pairwise indicators” $P(a,b)$, $a, b \in A$, are not sufficient to rank the objects from set A. For this purpose one needs to apply some ranking procedure, e.g., based on calculating net flow scores.

TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) [105, 106]. TOPSIS method bases on the idea that the “utility” of an object $a \in A$ depends on its geometric (euclidean) distance from the positive ideal object (i.e., the object, likely fictitious, that has the best possible evaluation on every criterion) and from the negative ideal object (i.e., the object, likely fictitious, that has the worst possible evaluation on every criterion). Let denote the former distance by $d_{a,^*}$ and the latter distance by $d_{a,*}$. Then, given objects $a, b \in A$, a is ranked higher than b if the ratio $d_{a,*}/(d_{a,*} + d_{a,^*})$ is smaller than the ratio $d_{b,*}/(d_{b,*} + d_{b,^*})$.

It is important to note that TOPSIS requires prior normalization of multicriteria evaluations. Therefore, this method is in general not meaningful, as demonstrated in [125, 183]. Application of this method is obviously problematic in case of ordinal criteria. The method requires also direct specification of weights of considered criteria. Finally, it is yet another compensatory MCDA method. All this restricts its correct applications.

1.2.2 PL Approaches to Multicriteria Ranking

In PL approaches to considered multicriteria ranking problem (called object ranking problem), the preference model is build by induction, using available training decision examples. Typically, this model has the form of a *utility function* (or scoring, or value function) or a so-called *preference function*, i.e., a function defined for pairs of objects that induces a comprehensive weak preference relation over set A. In the former case (called a *score-based setting*), one directly gets a total preorder over A. Learning a utility function involves solving an (ordinal) regression problem (e.g., [101]). In the latter case (called a *preference-based setting*), however, after learning a preference function (which can be achieved by learning a binary classifier using pairwise training data), one needs to exploit the resulting preference structure using some ranking procedure to obtain the final ranking (as the induced comprehensive weak preference relation is non-transitive, in general). Such two-stage (relational) approach was considered, e.g., in [3, 42, 43, 46].

Below, we remind a few well-known PL methods that, given some pairwise comparisons of reference objects, learn a DM's preference model which is subsequently used to calculate a total preorder over test set A. More methods are considered, e.g., in [71, 123].

In our short survey, we underline the form of applied preference model, the type of preference information required to build this model, and some other aspects related to application of the considered methods. Next, we give some general notes concerning limitations and shortcomings of the reviewed methods.

SVM^{rank}

Ranking SVM [102, 112], was proposed to solve the problem of ordinal regression. It learns a utility function (ranking function) such that the disagreement between the ranking implied by this function and the observed pairwise comparisons is minimized. The idea is to minimize a convex upper bound on the empirical ranking error (measured by a rank loss function) over a class of (kernelized) ranking functions. The input preference information consists of a total preorder of some reference objects.

Fast implementation of the ranking SVM is described in [113]. This implementation was used in the ranking experiments described in Chapter 5.

RankBoost

According to [65], RankBoost is an approach to the ranking problem based on a machine learning method called *boosting*, in particular, on Freund and Schapire's AdaBoost algorithm [66] and its successor developed by Schapire and Singer [153]. Boosting is a ML approach aimed at obtaining accurate prediction rules by combining many so-called *weak learners* that can be only moderately accurate. Such combination is often referred to as an *ensemble of base learners*. In RankBoost, boosting technique is used to learn a function $H : A \to \mathbb{R}$ whose induced ranking on A is a total preorder that approximates the (weighted) pairwise comparisons given by a DM. This function is defined as

$$H(a) = \sum_{t=1}^{T} \alpha_t h_t(a),$$

where $a \in A$, T is the number of combined weak learners, and $\alpha_t \in \mathbb{R}$ is the weight of weak learner $h_t : A \to \mathbb{R}$ inducing a total preorder over A.

RankBoost performs T so-called *rounds*. In each t-th round, $t \in \{1, \ldots, T\}$, weak learner h_t is trained and added to the current ensemble. During training, h_t has access to evaluations of considered objects and to current square matrix of weights of pairwise comparisons D_t. This matrix is modified after each round so that the training pairwise comparisons not respected by the current model (i.e., the ensemble composed of the weak learners added up to round t) get higher weights (higher priority) for the next round, and thus, will have more influence on the minimized (exponential) ranking loss of h_{t+1}.

Ensembles of Decision Rules

In [46], two ways of learning an ensemble of decision rules are considered. They differ by the type of considered decision rules. The rules of the first type specify conditions concerning differences of evaluations of two objects on particular attributes, implying comprehensive preference of one object over the other object. The resulting approach is called $PrefRules$. The rules of the second type specify conditions concerning evaluations of a single object on particular attributes, implying a given increase or decrease of a comprehensive utility (score) of this object. The resulting approach is called $RankRules$.

In $PrefRules$, the constructed ensemble of rules is applied on set A and the resulting preference structure on A is exploited using the *Net Flow Rule* (NFR) ranking method [30, 36] to get a final ranking. In $RankRules$, the learned ensemble of rules is applied on set A and the resulting comprehensive utility of individual objects induces a final ranking.

In both cases, the input preference information is a set of pairwise comparisons of training objects in terms of strict preference relation $\succ$, and the final ranking is a total preorder over A.

Critical Remarks

The methods reviewed above, although quite effective in learning utility or preference functions and then in ranking with these functions, have several drawbacks from the MCDA perspective. First, they tend to build preference models that are hard to interpret. An extreme case among the above methods is the SVM^{rank} method which is a black-box procedure with little interpretative value. Second, the methods that learn a functional model have problems handling criteria with ordinal scales – such scales tend to be arbitrarily converted to cardinal ones. Third, the considered methods do not take directly into account the domain knowledge concerning ordinal nature of considered criteria, and thus, the induced models may confuse a DM as they may be not concordant with her/his value system.

1.3 Review of Existing Approaches to Similarity-based Classification

In this section, we review some approaches to classification problem that are based on similarity-based reasoning. In our survey, we give several notes concerning limitations and shortcomings of the reviewed methods.

Remark that we use terms similarity-based reasoning and case-based reasoning interchangeably as we understand the latter term in a broad sense, as reflecting a generic reasoning methodology (whose idea was given in Section 1.1.2). In the literature (e.g., in [118]), however, the term case-based reasoning is often identified with a particular "lazy" approach (in ML sense), involving maintenance of a memory of objects (cases) M and reasoning by comparing a new case with the stored cases from M.

The approaches to CBR considered in the literature can be basically divided into two groups – *lazy learning methods* (*instance-based learning* methods or *memory-based learning* methods) and *eager learning methods*. The approaches from the first group employ similarity-based reasoning only to classify new (test) objects, while the methods from the second group use the information concerning mutual similarity of objects (marginal and/or comprehensive) to induce a similarity-based classifier.

1.3.1 Lazy Learning Methods

Lazy learning methods delay generalization beyond the training data until a new object is presented to the system. Thus, instead of learning a classification model (classifier), these methods just use the training objects to compile a *memory of previous cases*, so-called *case base*. Then, when a new object needs to be classified, it is compared with the objects from the case base and the object(s) most similar w.r.t. the considered features are used to work out the classification of the new object.

Among lazy learning classification methods based on the idea of similarity-based reasoning one can distinguish:

- *k-nearest-neighbors* classification method [61, 128];
- instance-based learning algorithms IB1, IB2, and IB3 [2];
- formalized CBR approaches [1, 118], where CBR is considered as a (cyclic) four-step process composed of the *retrieve*, *reuse*, *revise*, and *retain* steps; these approaches trace to the work of Robert Schank concerning dynamic memory model [152];
- the approach to case-based decisions proposed by Gilboa and Schmeidler [73, 74];
- approaches employing fuzzy set modeling (e.g., [51]).

It is important to note that the above methods, due to lack of generalization beyond the training set of objects, may be susceptible to noise observed in the training data, both with respect to irrelevant features, and with respect to outliers. Moreover, due to the possibly large size of maintained case base, the classification of a new object may be relatively slow when the similarity between this new object and each object in the case base needs to be calculated. This issue often calls for different techniques of case base reduction (see, e.g., [122]). An advantage of the above lazy learning classification methods is the natural ability of incremental learning by including newly classified objects in the case base.

1.3.2 Eager Learning Methods

Eager learning classification methods perform an explicit generalization of the training data to build a classifier that will predict classification of new objects presented to the system.

Among eager learning classification methods concerning mutual similarity of objects (marginal and/or comprehensive) in the process of learning a classifier one can distinguish:

- radial-basis function (RBF) artificial neural networks [40, 134];
- dominance-based rough set approach to case-based reasoning [88, 89, 91];
- approaches operating on the comprehensive similarity matrix [41];
- approach employing the Similarity Based Classification (SBC) algorithm proposed in [16].

In the first approach, training of a neural network involves learning of parameters of radial-basis (usually Gaussian) activation functions for hidden layer neurons. Then, given a new (test) object, output signal of each hidden layer neuron depends on the similarity between learned "neuron center" and this new object. This similarity is usually considered to be the inverse of Euclidean or Mahalanobis distance between the "neuron center" and the new object. Thus, the employed notion of similarity requires that object features are expressed on a (commensurable) numerical scale, which limits potential applications.

In the second approach, the training objects are compared pairwise and marginal similarities are aggregated (or rather combined) using decision rules based on the general monotonic relationship "the more similar is object y to object x w.r.t. the considered features, the greater the membership of y to a given decision class X". Although the idea of expressing similarity by decision rules is in itself very attractive, it is important to note two shortcomings of the approach proposed in [88, 89, 91]. First, the assumed monotonic relationship seems reasonable only if the membership of reference object x to class X takes a maximum value. If, e.g., the membership of x to X is just 0.5, then there is no reason to expect that the membership of y to X would increase up to maximum value of 1.0 when y becomes more and more similar to x w.r.t. the considered features. Second, the respective papers focus mainly on the properties of considered rough approximations of α-cuts of decision classes, and thus, they lack the proposal of, both, the way o inducing decision rules, and the way of resolving conflicts arising during application of induced rules to a new case.

In the third approach, the comprehensive similarity matrix (which is considered to be given a priori, although, in general, it may be calculated using marginal similarities) is transformed into a positive semidefinite matrix corresponding to a *kernel function*, which enables further application of a kernel-based classifier, e.g., the support vector machines (SVM). The transformation involves decomposition of the similarity matrix, identification of its negative eigenvalues, and modification of these negative eigenvalues (e.g., by setting them to zeros). This arbitrary modification is, obviously, the main concern of the considered approach.

The fourth approach introduces a measure of similarity between an unclassified object y and a crisp decision class $X \in \mathcal{D}$, denoted by $s_X(y)$, defined as $s_X(y) = \sum_{x \in X} \alpha(x)s(y,x)$, where $s(y,x)$ denotes similarity of object y to object x and $\alpha(x) \geq 0$, adjusted during training of the classifier, reflects the relative importance given to object $x \in U$ with respect to $\mathcal{D}$. Then, the class predicted for object y is calculated using the decision function $\arg\max_X\{s_X(y)\}$ or $\arg\max_X\{s_X(y) : s_X(y) > \sum_{Y \in \mathcal{D} \setminus X} s_Y(y)\}$. The main difficulty of SBC is the calculation of coefficients $\alpha(x)$. As noted by the authors, this method is also sensitive to noisy data. Moreover, since $s(y,x)$ is usually defined as $s(y,x) = e^{||y-x||^2/2\sigma^2}$, the evaluation vectors of y and x need to be expressed on a (commensurable) numerical scale, which limits potential applications.

1.3.3 Measuring Similarity

Measuring similarity is the essential point of all approaches to CBR. Questions related to measuring similarity are encountered at two levels:

- at the level of single features: how to define a meaningful similarity measure w.r.t. a single feature?
- at the level of all features: how to properly aggregate the similarity measures w.r.t. single features in order to obtain a *comprehensive similarity measure*?

Obviously, the task of choosing proper marginal similarity functions is not trivial and should depend on the characteristic of considered features. Nevertheless, as described in Section 1.1.2, instead of focusing on this issue, which is common for all similarity-based classification approaches, we assume in this thesis that these functions are given, and we concentrate on the second of the above levels, i.e., on the way of *aggregating marginal similarities into a comprehensive similarity.*

1.4 Motivation for Dominance-based Rough Set Approaches to Multicriteria Ranking and Similarity-based Classification

The two problems considered in this thesis, i.e., multicriteria ranking and similarity-based classification, are two problems of great practical importance. Exemplary ranking problems concern ranking of universities, ranking of investment funds, ranking of web pages, ranking of sale offers, etc. Similarity-based classification problems arise in clinical practice (classification of patients into risk groups based of previous diagnoses), in document categorization systems (classification of documents using a corpus), etc.

As we show in Sections 1.2 and 1.3, despite the importance of the two considered problems, many methods applied to solve these problems are hard to use (i.e., require too much cognitive effort on the part of a DM), and/or are not always appropriate (e.g., in case of nominal/ordinal attributes), and/or produce preference/classification models that are not meaningful to a DM.

Advantages of DRSA. In this thesis, we propose to handle both problems using adaptations of DRSA which are able to capture and model monotonic relationships typical for these problems. DRSA is a very attractive approach as it requires very weak assumptions concerning the data. DRSA can be used for heterogeneous data, containing (even simultaneously) nominal, ordinal, and cardinal (numerical) attributes – neither prior discretization of numerical attributes nor prior conversion of nominal and ordinal attributes into numerical ones is required. It can naturally handle inconsistency of available decision examples (training data) by calculating rough approximations of considered sets; these approximations are the basis for induction of an intelligible preference/classification model in the form of a set of *monotonic decision rules*. The rules clearly show logical patterns observed in decision examples. In the era of databases, when many companies store historical data, the availability of methods that can induce (learn) useful models from these data is obviously of great interest.

Advantages of decision rules. Decision rules induced in DRSA are relatively easy to read and understand by a DM. They have the ability to explain given decision examples as well as to predict decisions for new cases; a recommendation given by rules is fully *traceable*, i.e., one can see not only the rules matching a new case, but also decision examples supporting these rules. As shown in [87], and already mentioned in Section 1.1.1, axiomatic analysis of all three preference model types considered in MCDA (i.e., value function, outranking relation, and decision rules) leads to the conclusion that decision rules, as they are defined in DRSA, are the only model that gives account of most complex interactions among criteria – it is non-compensatory, accepts ordinal evaluation scales, and does not convert ordinal evaluations into cardinal ones.

Use of domain knowledge. One of the very important aspects when dealing with decision problems is *domain knowledge*. Exploitation of this knowledge can help to improve the quality of the induced model and ensure its compatibility with the value system of the DM. Due to application of DRSA, it is possible to take into account domain knowledge concerning value sets of attributes, division of attributes into condition and decision ones, preference scales of attributes and *monotonic relationships* between attributes.

1.4.1 Multicriteria Ranking Problem

The practical importance of multicriteria ranking problem, the observed shortcomings of the methods reviewed in Section 1.2, as well as the above advantages of DRSA and rule-based preference modeling, motivate us to propose an approach to multicriteria ranking based on DRSA but improving and extending previous MCDA rule-based approaches. Thus, our motivation is to propose a method that:

- is concordant with the current trend in MCDA which consists in induction of preference model from decision examples; this leads to reducing cognitive effort on the part of the DM who is not required to provide values of some difficult parameters, like weights of criteria or comparison thresholds (which is the case, e.g., for the methods from ELECTRE and PROMETHEE families described in Section 1.2.1);
- involves simple decision examples in the form of pairwise comparisons of objects

in terms of outranking and non-outranking relations; this information is easier definable by a DM than the pairwise comparisons in terms of comprehensive graded preference relations proposed in [63];

- can handle ordinal and cardinal criteria simultaneously, without prior discretization of numerical attributes or prior conversion of ordinal attributes into numerical ones;

- does not require from a DM to define graded preference relations for particular cardinal criteria, as considered in all rule-based MCDA approaches reviewed in Section 1.2.1; instead it uses difference of evaluations as a simple measure of the strength of preference;

- takes into account domain knowledge concerning ordinal character of criteria and the general monotonic relationship "if object a is preferred to object b at least as much as object c is preferred to object d with respect to each considered criterion, then the comprehensive preference of a over b is not weaker than the comprehensive preference of c over d";

- is better oriented towards solving real-life multicriteria ranking problems by using an adaptation of VC-DRSA [24], in particular VC-DRSA with consistency measure ϵ (called ϵ-VC-DRSA) that was found to be promising in prior experimental studies [25, 26];

- uses the rule preference model that is the most general preference model, easy to understand by a DM; rules induced from a PCT should be of the following type:

  "if car a has maximum speed at least 25 km/h greater than car b, and car a has comfort at least high while car b has comfort at most medium, then car a is at least as good as car b";

- when applying decision rules, takes into account not only the existence of rules concluding outranking or non-outranking but also the credibility and strength of these rules;

- employs proper (i.e., satisfying desirable properties) exploitation procedure for exploitation of the preference graph resulting from the application of induced rules on set A of objects to be ranked;

1.4.2 Procedures for Exploitation of Preference Graph

When using an MCDA rule-based preference model, application of rules on a set A of objects to be ranked yields a preference structure on this set. The induced preference structure, represented by a so-called *preference graph*, denoted by $\mathbb{G}$, needs to be further exploited to get a final ranking of all objects from A. In the four MCDA rule-based approaches reviewed in Section 1.2.1, only the procedures based on net flow scores of objects have been considered in exploitation step (s_5), which may be considered arbitrary. This motivates us to investigate other exploitation procedures.

In our research, we have noticed rich literature concerning different so-called *ranking methods* for exploitation of a (*single*) valued outranking relation over a set of objects [7, 28–31, 34–36, 49, 137, 139]. Preference graphs representing a valued outranking relation are obtained, e.g., in several outranking methods like ELECTRE III [57, 142] as well as PROMETHEE I and II [37, 38].

When decision examples given by a DM concern assignment of pairs of reference objects to outranking relation S and non-outranking relation S^c, which is the case of the first three MCDA rule-based approaches reviewed in Section 1.2.1 (denoted by α, β, and γ), and also of the approach considered in this thesis, then the preference graph resulting from application of rules induced from lower approximations of S and S^c represents *two relations* over set A: $\mathbb{S}$ and $\mathbb{S}^c$. These relations may be crisp or valued, in general. In the previous MCDA rule-based approaches (α, β, and γ), only crisp relations have been considered. They were defined as: $a\mathbb{S}b$ ($a\mathbb{S}^cb$) if there exists a rule that covers pair (a, b) and concludes aSb (aS^cb), $a, b \in A$. In this thesis, we consider both crisp and valued relations. In the latter case, $\mathbb{S}(a, b)$ ($\mathbb{S}^c(a, b)$) can be understood as the credibility of outranking (non-outranking) between a and b, $a, b \in A$. The valued relations can be constructed using different statistics of induced decision rules.

The above duality, i.e., the existence of well-known ranking methods concerning exploitation of a (single) valued relation and “our” case of exploitation of two relations resulting from application of decision rules, motivated us to look for an exploitation procedure that employs a suitable transformation of “our” preference graph $\mathbb{G}$ (representing two relations) to a preference graph $\mathbb{G}'$ representing a valued relation (that can be further exploited using one of the existing ranking methods). As we have managed to find such transformation, we faced a problem of choosing a proper ranking method for exploitation of the resulting preference graph $\mathbb{G}'$. This problem motivates us to analyze and compare several existing ranking methods w.r.t. a number of desirable properties. Some of these properties have been already studied in the literature, and some of them are introduced for the first time in this thesis.

1.4.3 Similarity-based Classification Problem

In similarity-based classification, the key issue is the aggregation of marginal similarities of objects into their comprehensive similarity. Typically, this aggregation is performed using some real-valued aggregation function (involving operators, like weighted L_p norm, min, etc.) (see, e.g., [51]) which is always arbitrary to some extent. This motivated us to look for an approach that measures comprehensive similarity in a (more) meaningful way, avoiding the use of an aggregation function. An approach of this type, employing an adaptation of DRSA, was proposed for the first time in [88], and improved in [89, 91]. In this approach, comprehensive similarity is represented by decision rules induced from classification examples. This enables to obtain a meaningful similarity measure, which is, moreover, resistant to irrelevant (or noisy) features because each decision rule, being a partial dominance cone in a similarity space, may involve conditions concerning only a subset of features. As the induced rules employ only ordinal properties of marginal similarity functions, the considered approach is also invariant to ordinally equivalent marginal similarity functions. A rule-based approach to similarity learning was considered also in [111], although in the context of unsupervised learning (clustering text documents).

Although attractive for the above reasons, the approach proposed in [88, 89, 91] has two main shortcomings, as discussed in Section 1.3.2.

The practical importance of the similarity-based classification problem, observed shortcomings of the methods reviewed in Section 1.3, as well as the above advantages of DRSA and rule-based preference modeling, motivate us to propose an approach to similarity-based classification based on DRSA, improving and extending previous DRSA-based approaches introduced in [88, 89, 91]. Thus, our motivation is to propose a method that:

- uses an adaptation of DRSA to case-based reasoning;

- avoids aggregation of marginal similarities using a real-valued aggregation function but rather combines these similarities in a meaningful way using monotonic decision rules;

- exploits only ordinal properties of marginal similarity functions and membership functions, and thus, it is invariant to ordinally equivalent marginal similarity functions;

- employs decision rules based on the monotonic relationship "the more similar is object y to object x w.r.t. the considered features, the closer is y to x in terms of the membership to a given decision class X"; we believe that this relationship truly reflects the monotonicity characteristic for CBR, i.e., monotonic relationship between comprehensive similarity of objects and their similarities w.r.t. single features; induced decision rules should be of the following type:

  "if similarity of flower y to flower x w.r.t. petal length is at least 0.7, and similarity of flower y to flower x with respect to sepal width is at least 0.8, then the membership of y to class setosa is between 0.7 and 0.9",

  where reference flower x has petal length 1.4, sepal width 3.0, and belongs to class setosa in degree 0.8;

- uses an adaptation of the rule classification scheme proposed in [21] to determine membership of a new object to a considered decision class.

1.5 Goal and Scope of the Thesis

The overall goal of this thesis is to develop adaptations of the Dominance-based Rough Set Approach (DRSA) to multicriteria ranking problem and to similarity-based classification problem (case-based reasoning). This general goal is divided into the following specific objectives:

(o1) develop the methodology for multicriteria ranking using VC-DRSA;

(o2) analyze alternative exploitation procedures that can be used to exploit the preference graph obtained in multicriteria ranking problem after application of a set of induced decision rules on a set of objects to be ranked; in particular, analyze properties of several well-known ranking methods that can be applied in one of the exploitation procedures involving a suitable transformation of the preference graph;

(o3) develop the methodology for similarity-based classification using DRSA;

(o4) verify experimentally the proposed methodology for multicriteria ranking.

The main focus of the thesis is on the methodology for multicriteria ranking. To verify learning potential of the proposed method for multicriteria ranking, we compare this method experimentally with another state-of-the-art method from the field of PL – SVM^{rank}.

The method for similarity-based classification using DRSA is illustrated by an example concerning application of this newly proposed method. Its comparison with other similarity-based classification methods in a simulated computational experiment is not meaningful because the other methods do not account for the same input classification information as our method. Precisely, this is the information concerning credibility of membership to each of the considered classes. Moreover, most of the other methods do not account for the same output classification information as produced by our method. This is the information concerning credibility of membership to each of the considered classes, which results from application of induced decision rules. Moreover, putting aside the final classification result, our method also provides additional useful information that other similarity-based classification methods do not provide.

1.6 Thesis Outline

The remainder of this thesis is organized as follows.

Chapter 2 presents our methodology for dealing with multicriteria ranking problems using VC-DRSA. This methodology is illustrated by an example given in Section 2.10.

In Chapter 3, we define several desirable properties of ranking methods, i.e., methods that exploit a preference graph representing a valued relation over a set of objects to be ranked, and yield a ranking (total or partial preorder) of all objects from this set. The ranking method is a core component of a generic exploitation procedure that we propose for exploitation of preference graph $\mathbb{G}$, representing two relations (crisp or valued), that is obtained when using our method for multicriteria ranking. In this chapter, we also examine the properties of several existing ranking methods, like Net Flow Rule, Min in Favor, etc. The respective proofs can be found in the Appendix. We conclude this chapter by indicating a ranking method that enjoys the best properties.

Chapter 4 concerns adaptation of DRSA to similarity-based classification (case-based reasoning). The proposed methodology is illustrated by an example presented in Section 4.9.

In Chapter 5, we describe the setup and results of a computational experiment in which we compared our method for multicriteria ranking with another approach from the field of PL – SVM^{rank}.

Chapter 6 summarizes the contribution of this thesis and presents possible directions of future research.

Chapter 2

Application of VC-DRSA to Multicriteria Ranking Problem

2.1 Introduction

In this chapter, we present a rule-based methodology for dealing with multicriteria ranking problem. This methodology employs an adaptation of the ϵ-VC-DRSA (i.e., VC-DRSA with consistency measure ϵ) [23, 24] to the multicriteria ranking problem. VC-DRSA is a probabilistic version of DRSA, however, it is not a statistical preference learning methodology in the sense of statistical machine learning, where the preference model is learned so as to minimize a loss function admitted for parameter estimation over a training set. In the current adaptation, decision examples have the form of *pairwise comparisons* of some reference objects, i.e., objects relatively well known to a DM. These pairwise comparisons, presented in a so-called *pairwise comparison table* (PCT), specify if a weak preference relation, called an *outranking relation* S, holds for the considered pairs of reference objects or not. When weak preference relation does not hold, such a relation is called a *non-outranking relation* S^c. As the pairwise comparisons given by a DM are prone to inconsistencies, we approximate *comprehensive preference relations* S and S^c by calculating their lower and upper approximations according to ϵ-VC-DRSA. Then, we employ VC-DomLEM rule induction algorithm [26] to induce probabilistic decision rules from the lower approximations of S and S^c. Thus, induced decision rules involve pairs of objects. They constitute the preference model of the DM who gave the pairwise comparisons. Application of these rules on a set A of objects to be ranked yields a preference structure on A, composed of relations $\mathbb{S}$ and $\mathbb{S}^c$ over A, and represented by a directed multigraph $\mathbb{G}$ called *preference graph*. In order to pass from the preference structure (preference graph) to the recommended ranking of objects, one has to apply an exploitation procedure. In this chapter, we briefly discuss four suitable exploitation procedures and then, we focus on one of these procedures which, in our opinion, deserves the most attention. This procedure consists in two steps: (i) suitable transformation of the preference graph $\mathbb{G}$ (representing two relations over set A) to a preference graph $\mathbb{G}'$ representing a (single) valued relation over A; (ii) exploitation of the resulting graph $\mathbb{G}'$ using a *ranking method*. We review several ranking methods known from the literature. The properties of these methods are analyzed in detail in Chapter 3.

This chapter comprises a synthesis and extension of the research results published in [163, 164]. In [163], the relations $\mathbb{S}$ and $\mathbb{S}^c$ induced by decision rules were both crisp, and thus, the corresponding preference structure was called *crisp preference structure*. Moreover, we adopted a typical assumption of MCDA concerning presence of a *consistent set of criteria* (as defined in Section 1.1). In [164], we did not make any assumption concerning the set of criteria, which is rather typical for PL. Moreover, the relations $\mathbb{S}$ and $\mathbb{S}^c$ induced by decision rules were both valued, and thus, the corresponding preference structure was called *valued preference structure*. In this thesis, we investigate all four combinations resulting from considering consistent or not necessarily consistent set of criteria on one hand, and crisp or valued preference structure on the other hand. The perception of the set of criteria (i.e., whether it is assumed to be a consistent set or a not necessarily consistent set) influences the way of constructing PCT and preference graph; the adopted type of preference structure (i.e., whether it is crisp or valued) influences the way of constructing preference graph and the composition of the considered set of desirable properties of a ranking method. Besides, when constructing valued preference structures, we consider two alternative ways of defining the strength of a probabilistic decision rule. The first way employs only credibility (consistency) of the rule, analogously to [63]. The second way (adopted in [164]) employs the product of credibility and coverage factor of the rule. Thus, altogether, we consider the following six versions of a rule-based approach to multicriteria ranking employing ϵ-VC-DRSA:

- VC-DRSA$^{rank}_{c\,0|1}$ – approach assuming that criteria make up a consistent set, and employing crisp preference structure,

- VC-DRSA$^{rank}_{c\,0\text{-}1_{cr}}$ – approach assuming that criteria make up a consistent set, employing valued preference structure, and measuring rule's strength by its credibility,

- VC-DRSA$^{rank}_{c\,0\text{-}1_{\times}}$ – approach assuming that criteria make up consistent set, employing valued preference structure, and measuring rule's strength by the product of its credibility and coverage factor,

- VC-DRSA$^{rank}_{nc\,0|1}$ – approach allowing not necessarily consistent set of criteria, and employing crisp preference structure,

- VC-DRSA$^{rank}_{nc\,0\text{-}1_{cr}}$ – approach allowing not necessarily consistent set of criteria, employing valued preference structure, and measuring rule's strength by its credibility,

- VC-DRSA$^{rank}_{nc\,0\text{-}1_\times}$ – approach allowing not necessarily consistent set of criteria, employing valued preference structure, and measuring rule's strength by the product of its credibility and coverage factor.

All the above versions were considered and compared in the computational experiment described in Chapter 5. In the following, when addressing all these versions simultaneously, we use a generic denotation VC-DRSA^{rank}. Moreover, we use denotation VC-DRSA$^{rank}_{c}$ when referring to the versions assuming that criteria make up a consistent set, and VC-DRSA$^{rank}_{nc}$ when referring to the versions without this assumption.

2.2 Preliminaries

A *valued relation* R over a set of objects A is a function from $A \times A$ into $[0, 1]$. It is said to be *reflexive* if $R(a, a) = 1$, for all $a \in A$. It is said to be *irreflexive* if $R(a, a) = 0$, for all $a \in A$. We denote by $\mathbf{R}_A$ the set of all valued relations over A. Moreover, we denote by R/A' the *restriction* of valued relation R over A to set $A' \subseteq A$, i.e., valued relation over A' such that for all $a, b \in A'$, $R/A'(a, b) = R(a, b)$. A valued relation R over A such that $R(a, b) \in \{0, 1\}$, for all $a, b \in A$, is said to be *crisp*. In such case:

- if $R(a, b) = 1$, we say that pair (a, b) belongs to relation R, and we write aRb or $(a, b) \in R$,
- if $R(a, b) = 0$, we say that pair (a, b) does not belong to relation R, and we write not aRb or $(a, b) \notin R$.

Let R be a crisp relation over A. This relation is said to be:

- *transitive* if (aRb and $bRc \Rightarrow aRc$),
- *complete* if (aRb or bRa),

for all $a, b, c \in A$.

A *total preorder* (also called *complete preorder* or *weak order*) over A is a crisp binary relation which is reflexive, transitive, and complete. A *partial preorder* over A, often called simply *preorder*, is a crisp binary relation which is reflexive, and transitive. The symmetric part of a total preorder relation R yields equivalence classes ordered by the asymmetric part of R.

Let R be a crisp relation over A. We denote by $G(A, R)$ the set of *greatest elements* of A given R, i.e.,

$$G(A, R) = \{a \in A : aRb \text{ for all } b \in A \setminus \{a\}\}. \tag{2.1}$$

It should be noticed that $G(A, R)$ may well be empty. When R is a total preorder, it is easy to see that set $G(A, R)$ is non-empty and equal to the first equivalence class of R.

A *ranking method* (RM) $\succeq$ is a function assigning a total or partial preorder $\succeq(A,R)$ over A to any finite set A and any valued relation R over this set. Remark that this is an extended definition w.r.t. the one given in [35, 36, 93], where $\succeq(A,R)$ was supposed to be a total preorder. Moreover, in [35, 36], $\succeq$ was called a ranking rule. However, in this thesis, we call $\succeq$ a ranking method to avoid confusion with decision rules.

We, respectively, denote by $=(A,R)$ and $\succ(A,R)$ the *symmetric* and *asymmetric parts* of $\succeq(A,R)$, i.e., the relations such that, for all $a,b \in A$,

$$a =(A,R)\ b \Leftrightarrow a \succeq(A,R)\ b \text{ and } b \succeq(A,R)\ a, \tag{2.2}$$

$$a \succ(A,R)\ b \Leftrightarrow a \succeq(A,R)\ b \text{ and not } b \succeq(A,R)\ a. \tag{2.3}$$

2.3 Problem Setting

In this section, we describe in detail the setting of the considered multicriteria ranking problem introduced in Section 1.1.1. We formalize several aforementioned concepts and give respective notation that is used in the remaining part of this thesis.

We consider a multicriteria ranking problem in which objects belonging to a finite set A have to be ranked, either completely or partially. In the former case, one aims at obtaining a total preorder over A. In the latter case, one accepts a partial preorder over A. The objects from set A are evaluated by set $G = \{g_1, \ldots, g_n\}$ of n criteria. Each *criterion* $g_i \in G, i = 1, \ldots, n$, is a real-valued function $g_i : A \to \mathbb{R}$, with *cardinal* (i.e., interval or ratio) *scale* or *ordinal scale* (which is given a priori or results from an order-preserving number-coding of non-numerical ordinal evaluations). Thus, value $g_i(a)$, $a \in A$, represents the evaluation of object a with respect to (w.r.t.) criterion g_i. A criterion with the cardinal scale is called a *cardinal criterion*; the set of all cardinal criteria is denoted by $G^N \subseteq G$. A criterion with the ordinal scale is called an *ordinal criterion*; the set of all ordinal criteria is denoted by $G^O \subseteq G$. Moreover, $G^N \cup G^O = G$ and $G^N \cap G^O = \emptyset$. The meaning of the two scales is such that in the case of a criterion $g_i \in G^N$ with a cardinal scale, one can define a function $k_i : \mathbb{R}^2 \to \mathbb{R}$ which measures the *intensity of preference* (positive or negative) of object a over object b, taking into account evaluations $g_i(a), g_i(b)$, $a, b \in A$. For properties of function k_i, and different ways of defining it, the reader is referred to [84]. Basically, k_i is non-decreasing w.r.t. the first argument, and non-increasing w.r.t. the second argument. For the sake of simplicity, we assume in this thesis that for each cardinal criterion $g_i \in G^N$, intensity of preference is defined as: $k_i[g_i(a), g_i(b)] = \Delta_i(a,b) = g_i(a) - g_i(b)$. In the case of a criterion $g_i \in G^O$ with an ordinal scale, this is not possible (as differences of evaluations are not meaningful) and one can only establish an order of evaluations $g_i(a)$, $a \in A$.

We assume, moreover, without loss of generality, that all the criteria are of *gain-type*, i.e., the greater the criterion value the better.

Let us denote by $V_{g_i} = \mathbb{R}$ the value set (domain) of criterion $g_i \in G$. Then, set $V_G = \prod_{i=1,\ldots,n} V_{g_i} = \mathbb{R}^n$ is called *G-evaluation space*.

Given the statement of the multicriteria ranking problem, the only objective information one can get is the *dominance relation* D over set of objects A, defined in the G-evaluation space. Let us consider objects $a, b \in A$; object a is said to dominate object b, denoted by aDb, if and only if (iff) for all $g_i \in G : g_i(a) \geq g_i(b)$. The dominance relation D is, however, too poor because it leaves many objects incomparable. In order to make the objects more comparable, a DM must supply some *preference information* revealing her/his value system w.r.t. multicriteria evaluations. We consider a frequent case, when the preference information has the form of *pairwise comparisons* of some objects relatively well known to the DM, called *reference objects*. This information is thus composed of some decision examples on the reference objects. Remark that pairwise comparisons may be given by a DM directly or they may be calculated using some other type of preference information given by a DM, like a *reference ranking* (i.e., a linear ranking of reference objects, possibly with ties) or an ordinal classification of reference objects. We clarify this point below, after we adopt a particular form of pairwise comparisons.

Let us denote by A^R the set of all reference objects. Set A^R can be a subset of A, however, it is not required by the presented methodology. If $A^R \nsubseteq A$, then we just need to define each criterion $g_i \in G, i = 1, \ldots, n$, as function $A \cup A^R \rightarrow \mathbb{R}$, and dominance relation D over set $A \cup A^R$. In any case, $A \setminus A^R$ is a set of objects *unseen* during preference model learning.

Following [82, 84, 160], we consider that for each ordered pair (a, b) of different reference objects, i.e., $(a, b) \in A^R \times A^R, a \neq b$, the DM can state either "object a is at least as good as object b" (in other words – "object a outranks object b") or "object a is not at least as good as object b" (in other words – "object a does not outrank object b"), or abstain from any judgment. The first situation is denoted by aSb (or $(a, b) \in S$), while the second one is denoted by aS^cb (or $(a, b) \in S^c$). Moreover, we fix aSb for pairs $(a, b) \in A^R \times A^R$ such that aDb (in case when set G is assumed to be a consistent set of criteria), or we only fix aSa for all $a \in A^R$ (in case when set G is not assumed to be a consistent set of criteria). Thus, from a formal point of view, the DM can reveal her/his preferences by assigning pairs of objects to any of the two considered disjoint comprehensive preference relations: outranking relation S or non-outranking relation S^c. Obviously, relation S is a weak preference relation which, in general, is only reflexive. It is not symmetric, not transitive, and not complete. Moreover, non-outranking relation S^c is irreflexive, and, in general, it is not symmetric, not transitive, and not complete. This is to say that the preference information coming from the DM is relatively weak and non-exhaustive.

As mentioned above, relations S and S^c may be also calculated using some other type of preference information given by a DM, like a reference ranking or an ordinal classification of reference objects. In the first case, calculated relation S is composed of pairs $(a, b) \in A^R \times A^R$ such that object a is ranked not lower than object b, while calculated relation S^c is composed of pairs $(a, b) \in A^R \times A^R$ such that object a is ranked lower than object b. In the second case, calculated relation S is composed of pairs $(a, b) \in A^R \times A^R$

such that object a is classified to a decision class not worse than that of object b, while calculated relation S^c is composed of pairs $(a, b) \in A^R \times A^R$ such that object a is classified to a decision class worse than that of object b. Observe, however, that the calculated relations S and S^c obtained in the above two cases are always transitive. Thus, the pairwise comparisons adopted in this thesis are a more general type of preference information, as the resulting relations S and S^c are not necessarily transitive. Moreover, the pairwise comparisons provided by the DM need not to be complete, in the sense that (s)he may assign pair $(a, b) \in A^R \times A^R$ to relation S or S^c but abstain from assigning inverted pair (b, a) to any of these relations.

By expressing her/his preferences using statements concerning outranking or non-outranking, for each pair of objects $(a, b) \in A^R \times A^R$, $a \neq b$, the DM can easily specify any of the four situations typically considered in MCDA, i.e.:

- *strict preference* P:
$$aPb \Leftrightarrow aSb \wedge bS^c a, \tag{2.4}$$
- *inverse strict preference* P^{-1}:
$$aP^{-1}b \Leftrightarrow aS^c b \wedge bSa, \tag{2.5}$$
- *indifference* I:
$$aIb \Leftrightarrow aSb \wedge bSa, \tag{2.6}$$
- *incomparability* J:
$$aJb \Leftrightarrow aS^c b \wedge bS^c a. \tag{2.7}$$

It is worth stressing that expressing decision examples on the reference objects is cognitively relatively easy for a DM. In our approach, instead of requiring that the DM provides values of some difficult parameters like weights of criteria or different thresholds (see, e.g., the outranking methods reviewed in Section 1.2.1), and then using this information in a complex preference model, we treat the decision examples supplied by the DM as training data, and then follow with learning of a preference model of the DM in easy rule terms.

To simplify the notation, in the following, we will use unique symbol T to refer to any of the comprehensive preference relations S and S^c when these relations are considered jointly, unless this may cause misunderstanding. Moreover, we denote by $\mathcal{I}_G, \mathcal{I}_{G^N}, \mathcal{I}_{G^O} \subseteq \{1, \ldots, n\}$ the sets of indices of criteria belonging to G, G^N, G^O, respectively, where $\mathcal{I}_{G^N} \cap \mathcal{I}_{G^O} = \emptyset$ and $\mathcal{I}_{G^N} \cup \mathcal{I}_{G^O} = \mathcal{I}_G$.

2.4 Pairwise Comparison Table

The preference information of a DM in the form of pairwise comparisons of reference objects is used to create a pairwise comparison table (PCT), first introduced in [77, 80]. Let us denote by $B \subseteq A^R \times A^R$ the set of pairs of reference objects for which the comprehensive preference of the DM is known. The exact composition of this set depends on the considered version of VC-DRSA^{rank}, as specified below:

- in case when set G is assumed to be a consistent set of criteria, set B is composed of pairs $(a, b) \in A^R \times A^R$, such that not aDb, for which the DM expressed her/his preferences by declaring aSb or aS^cb, as well as of other pairs $(c, d) \in A^R \times A^R$, such that cDd, for which we assume relation S;

- in case when set G is not assumed to be a consistent set of criteria, set B is composed of pairs $(a, b) \in A^R \times A^R$, $a \neq b$, for which the DM expressed her/his preferences by declaring aSb or aS^cb, as well as of pairs (a, a), $a \in A^R$, which are assigned to S.

Thus, $B = S \cup S^c$.

Intuitively, a PCT created on the basis of preference information supplied by the DM is an $m \times (n+1)$ data table, denoted by S_{PCT}, where m is the cardinality of set B of pairs. First n columns of this table correspond to criteria from set G. The last, $(n+1)$-th, column represents the comprehensive preference relation S or S^c. Each row of S_{PCT} corresponds to a pair of reference objects from B. As announced in Section 2.3, when comparing two objects $a, b \in A^R$ on a cardinal criterion $g_i \in G^N$, one puts in the corresponding column of S_{PCT} the difference $g_i(a) - g_i(b)$. When comparing two objects $a, b \in A^R$ on an ordinal criterion $g_i \in G^O$, one puts in the corresponding column of S_{PCT} an ordered pair of ordinal evaluations $(g_i(a), g_i(b))$.

Describing the PCT more formally, one can say that each pair of objects $(a, b) \in B$ is evaluated on set G of criteria, such that:

- for criterion $g_i \in G^N$, the evaluation of $(a, b) \in B$ is defined as $q_i(a, b) = g_i(a) - g_i(b) \in V_{q_i} = \mathbb{R}$,

- for criterion $g_i \in G^O$, the evaluation of $(a, b) \in B$ is defined as $q_i(a, b) = (g_i(a), g_i(b)) \in V_{q_i} = \mathbb{R} \times \mathbb{R}$.

Then, set $V_Q = \prod_{i \in \mathcal{I}_G} V_{q_i}$ is called *Q-evaluation space.*

2.5 Rough Approximation of Outranking and Non-outranking Relations

In Section 2.3, we considered dominance relation D over set of objects A, defined in the G-evaluation space. Here, we introduce another type of dominance relation, denoted by D_2. This binary relation is defined over set B of pairs of objects, in the Q-evaluation space. However, as it is more convenient, below we introduce dominance relation D_2 using only the evaluations of objects from set A^R on the criteria from set G.

First, let us consider a case when set G is composed of cardinal criteria only, i.e., $G^N = G, G^O = \emptyset$. Then, given pairs of objects $(a,b),(c,d) \in B$, pair (a,b) is said to dominate pair (c,d) w.r.t. criteria from G (denoted by $(a,b)D_2(c,d)$) iff $\Delta_i(a,b) \geq \Delta_i(c,d)$ for each $g_i \in G$, where $\Delta_i(a,b)$ denotes $g_i(a) - g_i(b)$. Let D_2^i be the dominance relation over B confined to single criterion $g_i \in G$. This relation is reflexive, transitive and complete. Therefore, D_2^i is a total preorder over B. Since an intersection of total preorders is a partial preorder, and relation D_2 over B is the intersection of relations D_2^i, $i \in \mathcal{I}_G$, then relation D_2 is a partial preorder over B.

Secondly, let us consider a case when set G is composed of ordinal criteria only, i.e., $G^O = G, G^N = \emptyset$. Then, given pairs of objects $(a,b),(c,d) \in B$, pair (a,b) is said to dominate pair (c,d) w.r.t. criteria from G iff $g_i(a) \geq g_i(c)$ and $g_i(b) \leq g_i(d)$ for each $g_i \in G$. In other words, pair (a,b) is said to dominate pair (c,d) w.r.t. criteria from G iff aDc and dDb. Let D_2^i be the dominance relation over B confined to single criterion $g_i \in G$. This relation is reflexive, transitive but non-complete (i.e., it is possible that neither $(a,b)D_2^i(c,d)$ nor $(c,d)D_2^i(a,b)$ for some $(a,b),(c,d) \in B$ and $g_i \in G$). Therefore, D_2^i is a partial preorder over B. Since an intersection of partial preorders is also a partial preorder, and relation D_2 over B is the intersection of relations $D_2^i, i \in \mathcal{I}_G$, then relation D_2 is a partial preorder over B.

Finally, when set G is composed of both cardinal and ordinal criteria, i.e., when $G^N \neq \emptyset$ and $G^O \neq \emptyset$, then given pairs of objects $(a,b),(c,d) \in B$, pair (a,b) is said to dominate pair (c,d) w.r.t. criteria from G iff (a,b) dominates (c,d) w.r.t. both G^N and G^O. Since the dominance w.r.t. G^N is a partial preorder over B and the dominance w.r.t. G^O is a partial preorder over B, then the dominance D_2, being the intersection of these two dominance relations, is also a partial preorder over B.

Let $G' \subseteq G$ and pairs $(a,b),(c,d) \in B$. Then, if (a,b) dominates (c,d) w.r.t. set G of criteria, then (a,b) dominates (c,d) w.r.t. set G'.

Given a pair of objects $(a,b) \in B$ we define the following:

- a set of pairs of objects dominating (a,b), called the *dominating set* or *positive dominance cone* in the Q-evaluation space:

$$D_2^+(a,b) = \{(c,d) \in B : (c,d)D_2(a,b)\}, \tag{2.8}$$

- a set of pairs of objects dominated by (a,b), called the *dominated set* or *negative dominance cone* in the Q-evaluation space:

$$D_2^-(a,b) = \{(c,d) \in B : (a,b)D_2(c,d)\}. \tag{2.9}$$

In equations (2.8) and (2.9), pair of objects (a,b) is called an *origin* of the dominance cone. Dominating and dominated sets of objects are “granules of knowledge” used to approximate outranking and non-outranking relation, respectively.

We formulate the following *dominance principle* w.r.t. pairwise comparisons of a DM: "if a is preferred to b at least as much as c is preferred to d with respect to each $g_i \in G$, then the comprehensive preference of a over b is not weaker than the comprehensive preference of c over d". This means that if $(a, b)D_2(c, d)$, one expects that:

(i) if $aS^c b$, then $cS^c d$,

(ii) if cSd, then aSb.

Violation of this dominance principle is considered as an *inconsistency* w.r.t. dominance relation D_2 over B. Let us observe that, thanks to the presence in S_{PCT} of pairs $(a, a) \in S, a \in A^R$, an inconsistency w.r.t. D_2 appears also when given two objects $a, b \in A^R$, the DM states that $aS^c b$, while aDb. This is related to the reflexivity of S and the irreflexivity of S^c. In fact, aDb implies $(a, b)D_2(a, a)$, and together with aSa, this implies that aSb. Thus, the opposite statement $aS^c b$ is inconsistent with the expectation (ii) listed above.

In practice, decision examples given by a DM are often inconsistent due to hesitation of the DM, unstable character of her/his preferences, or incomplete determination of the set of criteria [e.g., 145]. These inconsistencies cannot be considered as a simple error or as noise. They can convey important information that should be taken into account when constructing a preference model of the DM. Rather than correct or ignore these inconsistencies, we propose to handle them using a dominance-based rough set approach. Before learning of a preference model of the DM, we structure pairs of objects contained in S_{PCT} by calculation of lower approximations of the two comprehensive preference relations. In this way, we restrict a priori the set of pairs of objects on which the preference model is built to a subset of sufficiently consistent pairs of objects belonging to lower approximations. This restriction is motivated by a postulate for learning from (sufficiently) consistent data, so that the knowledge gained from this learning is relatively certain (or, in other words, the induced preference model is reliable). Analogous restriction proved to be effective in case of ordinal classification problems with monotonicity constraints [25, 26]. It is worth underlining that, although only sufficiently consistent pairs of objects from S_{PCT} are used to construct a preference model of the DM, the remaining pairs of objects are not removed from S_{PCT}. In other words, the approach proposed in this thesis does not boil down to a simple pre-processing performed to remove inconsistent decision examples. In fact, inconsistent pairs of objects play the role of "counterexamples", helping this way to induce a preference model.

In some previous PCT-oriented adaptations of DRSA to multicriteria ranking [e.g., 82, 84, 160], outranking and non-outranking relations were approximated using strict inclusion relation between dominance cones originating in pairs of objects $(a, b) \in B$ and the comprehensive preference relations. Precisely, lower approximations of relations

S and S^c were defined as:

$$\underline{S} = \{(a,b) \in B : D_2^+(a,b) \subseteq S\}, \tag{2.10}$$

$$\underline{S^c} = \{(a,b) \in B : D_2^-(a,b) \subseteq S^c\}. \tag{2.11}$$

These definitions of lower approximations appear to be too restrictive in practical applications. In consequence, lower approximations of S and S^c are often empty, preventing generalization of pairwise comparisons in terms of sufficiently certain decision rules. Therefore, in this thesis, we rely on the Variable Consistency Dominance-based Rough Set Approach (VC-DRSA) [24, 97] which is a probabilistic extension of the classical DRSA. Since originally VC-DRSA was introduced for multicriteria classification problems, here we adapt its definitions of variable-consistency (v-c) lower approximations to the case of approximating outranking and non-outranking relations. In the adapted definitions of v-c lower approximations of S and S^c, we use consistency measure $\epsilon_T : B \rightarrow [0,1]$ (whose prototype was introduced in [23, 24]) defined as:

$$\epsilon_S(a,b) = \frac{|D_2^+(a,b) \cap S^c|}{|S^c|}, \tag{2.12}$$

$$\epsilon_{S^c}(a,b) = \frac{|D_2^-(a,b) \cap S|}{|S|}. \tag{2.13}$$

Given pair of objects $(a,b) \in B$ and relation T, value $\epsilon_T(a,b)$ reflects consistency of pair (a,b) w.r.t. T. ϵ_T is a cost-type consistency measure, which means that value zero denotes full consistency and the greater the value, the less consistent is a given pair of objects. The definitions of v-c lower approximations adapted to the case of approximating outranking and non-outranking relations are the following:

$$\underline{S} = \{(a,b) \in S : \epsilon_S(a,b) \leq \theta_S\}, \tag{2.14}$$

$$\underline{S^c} = \{(a,b) \in S^c : \epsilon_{S^c}(a,b) \leq \theta_{S^c}\}, \tag{2.15}$$

where consistency thresholds $\theta_S, \theta_{S^c} \in [0,1)$. The values of these thresholds can be given by a DM or fixed using a simple wrapper-like cross validation procedure. Note that in case $\theta_S = \theta_{S^c} = 0$, the v-c lower approximations (2.14) and (2.15) are equal to the lower approximations (2.10) and (2.11), respectively. In the following, unless this may cause misunderstanding, we drop "v-c" and call sets of pairs of objects defined by (2.14) and (2.15) just lower approximations of relations S and S^c, respectively.

In [24], several consistency measures were defined. The choice of particular consistency measure ϵ_T is dictated by several factors. The first one is that value $\epsilon_T(a,b)$, where $(a,b) \in B$, features an easy interpretation – it can be interpreted as an estimate of conditional probability that a pair of objects $(c,d) \in B$ belongs to the dominance cone originating in pair (a,b), given that pair (c,d) does not belong to comprehensive preference relation T. The second factor is a good performance of measure ϵ_T in prior computational experiments [25, 26], comparing to other consistency measures. The third factor is the fact that measure ϵ_T has all monotonicity properties [24] relevant to the case of a PCT with just two possible decisions for each pair of objects, i.e., assignment to relation S or S^c. Precisely, measure ϵ_T has the following monotonicity properties: (m1) – monotonicity w.r.t. the set of criteria, (m2) – monotonicity w.r.t. relation T, and (m4) – monotonicity w.r.t. dominance relation D_2 over B. Definitions of these properties, for the case of a cost-type consistency measure, can be found in the Appendix (Definitions 14, 15, and 16).

Using definitions (2.14) and (2.15), one can define v-c upper approximations and v-c boundaries of sets S and S^c as in [24].

The coefficient

$$\gamma(S, S^c) = \frac{|\underline{S} \cup \underline{S^c}|}{|B|} \tag{2.16}$$

defines *quality of approximation* of S and S^c by set G. Obviously, $\gamma(S, S^c) \in [0, 1]$, and $\gamma(S, S^c) = 1$ indicates that for given values of consistency thresholds $\theta_S, \theta_{S^c} \in [0, 1)$, the lower approximations given by (2.14) and (2.15) contain all the pairs of objects from relations S and S^c, respectively.

We define *positive regions* of relations S and S^c as follows:

$$POS(S) = \bigcup_{(a,b) \in \underline{S}} D_2^+(a, b), \tag{2.17}$$

$$POS(S^c) = \bigcup_{(a,b) \in \underline{S^c}} D_2^-(a, b). \tag{2.18}$$

Positive region of relation S (respectively, S^c) contains pairs of objects sufficiently consistent, i.e., belonging to lower approximation of relation S (2.14) (respectively, S^c (2.15)), and can also contain some inconsistent pairs of objects which fall into dominance cones $D_2^+(\cdot, \cdot)$ (respectively, $D_2^-(\cdot, \cdot)$) originating in pairs of objects from lower approximation of relation S (respectively, S^c). Moreover, one can define boundary and negative regions of relations S and S^c analogously to [22, 26]. It is also possible to perform further DRSA-like analysis by calculating reducts and the core [e.g., 83, 84, 86, 159, 160].

2.6 Induction of Decision Rules

After structuring decision examples supplied by a DM into lower approximations of comprehensive preference relations, we induce a generalized description of sufficiently consistent pairs of objects from S_{PCT} in terms of a set of *minimal decision rules*. An induced set of rules is considered to be a *preference model* of the DM who gave the pairwise comparisons of reference objects. Each rule is a statement of the type:

if Φ, *then* Ψ,

where Φ and Ψ denote *condition* and *decision* part of the rule, called also *premise* and *conclusion*, respectively. The condition part of the rule is a conjunction of elementary conditions concerning individual criteria, and the decision part of the rule suggests assignment of pairs of objects covered by the rule to outranking relation S or to non-outranking relation S^c. The rule is said to *cover* a pair of objects $(a,b) \in A \times A$ if this pair satisfies all the elementary conditions of the rule. A pair of objects $(a,b) \in B$ is said to *support* the rule if this pair satisfies all the elementary conditions and the conclusion of the rule. Rule r_T, suggesting assignment of covered pairs of objects to relation T, is called *minimal* if there is no other rule r'_T having premise at least as general as that of r_T (i.e., employing subset of elementary conditions of r_T and/or more general elementary conditions than r_T) and consistency not worse than that of r_T (where by consistency of rule r_T we understand the value of a rule consistency measure defined later in this section). In the following, a minimal decision rule is denoted by *m-rule*. The interest in minimal decision rules comes, obviously, from the fact that they generalize decision examples better than non-minimal rules. Thus, generation of minimal decision rules may be seen as a way to avoid overfitting.

Decision rules are induced so as to cover pairs of objects from lower approximations (2.14) and (2.15). However, in some cases it is impossible for a rule to cover only pairs of objects from a lower approximation. To handle these cases, the positive region of the considered comprehensive preference relation is computed according to (2.17) or (2.18).

Set $\underline{T}$ of pairs of objects belonging to the lower approximation of comprehensive preference relation T is the basis for induction of a set of minimal decision rules that suggest assignment to T. A rule from this set is supported by at least one pair of objects from $\underline{T}$, and it covers pair(s) of objects from $POS(T)$. The elementary conditions (selectors) that form this rule are built using only evaluations of objects present in the pairs of objects that belong to $\underline{T}$.

Below, we define the syntax of decision rules that generalize description of sufficiently consistent pairs of objects present in a PCT:

$$\begin{gathered} \textit{if } \left(\Delta_{i_1}(a,b) \geq \delta_{i_1}\right) \wedge \ldots \wedge \left(\Delta_{i_p}(a,b) \geq \delta_{i_p}\right) \wedge \\ \left(g_{i_{p+1}}(a) \geq r_{i_{p+1}} \wedge g_{i_{p+1}}(b) \leq s_{i_{p+1}}\right) \wedge \ldots \wedge \left(g_{i_z}(a) \geq r_{i_z} \wedge g_{i_z}(b) \leq s_{i_z}\right), \\ \textit{then } aSb, \end{gathered} \tag{2.19}$$

$$\begin{gathered} \textit{if } \left(\Delta_{i_1}(a,b) \leq \delta_{i_1}\right) \wedge \ldots \wedge \left(\Delta_{i_p}(a,b) \leq \delta_{i_p}\right) \wedge \\ \left(g_{i_{p+1}}(a) \leq r_{i_{p+1}} \wedge g_{i_{p+1}}(b) \geq s_{i_{p+1}}\right) \wedge \ldots \wedge \left(g_{i_z}(a) \leq r_{i_z} \wedge g_{i_z}(b) \geq s_{i_z}\right), \\ \textit{then } aS^cb, \end{gathered} \tag{2.20}$$

where: $\Delta_{i_j}(a,b)$ denotes $g_{i_j}(a) - g_{i_j}(b)$, $\delta_{i_j} \in \{g_{i_j}(c) - g_{i_j}(d) : (c,d) \in B\} \subseteq \mathbb{R}$, for $i_j \in \{i_1, \ldots, i_p\} \subseteq \mathcal{I}_{G^N}$; $(r_{i_j}, s_{i_j}) \in \{(g_{i_j}(c), g_{i_j}(d)) : (c,d) \in B\} \subseteq \mathbb{R} \times \mathbb{R}$, for $i_j \in \{i_{p+1}, \ldots, i_z\} \subseteq \mathcal{I}_{G^O}$. For instance, considering ranking of cars, a decision rule could be "if car a has maximum speed at least 25 km/h greater than car b (cardinal criterion) and car a has comfort at least 3 while car b has comfort at most 2 (ordinal criterion), then car a is at least as good as car b", where values 2 and 3 code ordinal evaluations 'medium' and 'good', respectively.

The rules with syntax (2.19) are called *at least rules*, while the rules with syntax (2.20) are called *at most rules*. Let us observe that the above syntax of decision rules is concordant with the definition of dominance relation D_2 over B in the sense that the premise of each decision rule corresponds to a positive or negative dominance cone in (a subspace of) the Q-evaluation space. Moreover, as we work with variable-consistency lower approximations, in order to cover by rules all pairs of objects from $\underline{S}$ and $\underline{S^c}$, we have to agree that not all the rules will be fully consistent. For example, it is inevitable that a rule suggesting assignment to relation S covers pairs of objects that do not belong to S but dominate in the Q-evaluation space at least one pair of objects from $\underline{S}$ that supports the considered rule. Therefore, we speak about *probabilistic decision rules* to underline the fact that not all pairs of objects from S_{PCT} that are covered by a rule have to support this rule.

Decision rules can be characterized by many attractiveness measures (see [95] for a study of some properties of these measures).

Since we are working with probabilistic decision rules, it is important to control consistency of these rules. To this end, we define a cost-type *rule consistency measure* [25, 26] denoted by $\widehat{\epsilon}_T$. This measure is a function $\widehat{\epsilon}_T : R_T \rightarrow [0, 1]$, where R_T is the set of rules suggesting assignment to relation T. Let us denote by $\Phi(r_T)$, $\Psi(r_T)$, and $\|\Phi(r_T)\|$, the condition part of rule r_T, its decision part, and the set of pairs of objects covered by the rule, respectively. Then, measure $\widehat{\epsilon}_T$ is defined as:

$$\widehat{\epsilon}_T(r_T) = \frac{\big|\|\Phi(r_T)\| \cap \neg T\big|}{|\neg T|}, \tag{2.21}$$

where $\neg T = B \setminus T$ is the complement of relation T w.r.t. set B (obviously, $\neg S = S^c$ and $\neg S^c = S$).

Induced rules have to satisfy the same constraints on consistency as pairs of objects from the lower approximation which serves as a base for rule induction. In particular, each rule r_T is required to satisfy threshold θ_T, i.e., $\widehat{\epsilon}_T(r_T)$ has to be not greater than θ_T. In the following, rule r_T satisfying threshold θ_T is called *sufficiently consistent* and denoted by *sc-rule*. Since rule consistency measure $\widehat{\epsilon}_T$ is a counterpart of consistency measure ϵ_T defined as (2.12) and (2.13), it can be shown that $\widehat{\epsilon}_T$ derives monotonicity properties from ϵ_T.

Let us now remind some useful definitions concerning probabilistic decision rules, introduced in [26].

A probabilistic decision rule r_T suggesting assignment to relation T is *discriminant* if it covers only pairs of objects belonging to positive region $POS(T)$. In the following, a discriminant decision rule is denoted by *d-rule*. Moreover, rule r_T is *robust* if there exists a pair of objects $(a, b) \in \underline{T}$ which is a *base* of r_T. Considering for example definition (2.19), it means that $q_{i_1}(a, b) = \delta_{i_1} \wedge \ldots \wedge q_{i_p}(a, b) = \delta_{i_p} \wedge q_{i_{p+1}}(a, b) = (r_{i_{p+1}}, s_{i_{p+1}}) \wedge \ldots \wedge q_{i_z}(a, b) = (r_{i_z}, s_{i_z})$. In the following, a robust decision rule is denoted by *r-rule*. Set R_T of rules suggesting assignment to relation T is *minimal* if each pair of objects $(a, b) \in \underline{T}$ is covered by at least one rule $r_T \in R_T$ and elimination of any rule from R_T makes that not all pairs of objects $(a, b) \in \underline{T}$ are covered by the remaining rules. In the following, a minimal set of decision rules is denoted by *m-set of rules*.

Induction of decision rules is a complex problem and many algorithms have been introduced to deal with it. Main rule induction algorithms defined for DRSA can be found in [25, 26, 47, 92, 161]. In general, these rule induction algorithms can be divided into three categories that reflect different induction strategies:

(i) generation of a minimal set of decision rules,

(ii) generation of an exhaustive set of decision rules,

(iii) generation of a satisfactory set of decision rules.

When applied to a PCT, algorithms from category (i) focus on describing all pairs of objects from lower approximations of S and S^c by an m-set of m-rules. Algorithms from category (ii) generate all m-rules. Category (iii) includes algorithms that generate all m-rules that satisfy some a priori defined requirements (concerning, e.g., maximum rule length or minimum support). One should also mention another category of rule induction algorithms based on calculating reducts of a considered set of attributes (see, e.g., [8, 9]), although they concern the indiscernibility-based rough set approach.

In this thesis, we apply VC-DomLEM algorithm [25, 26] which belongs to category (i). Each of the sets R_S and R_{S^c} of decision rules induced by VC-DomLEM for comprehensive preference relations S and S^c, respectively, is an m-set of m-sc-rules (i.e., is a minimal set composed of minimal and sufficiently consistent decision rules). During induction of a single rule, we employ ϵ-consistency measure defined in [26]. Moreover, we parameterize the algorithm in such a way, that it induces d-rules (technically, this is achieved by choosing covering option $s = 1$, which means that each induced rule $r_T \in R_T$ is allowed to cover only pairs of objects belonging to set $POS(T)$). It is important to note that the rules generated by VC-DomLEM do not have to be robust, which means that each rule r_T can employ elementary conditions created using evaluations in Q-evaluation space of different pairs of objects from $\underline{\mathrm{T}}$.

2.7 Application of Decision Rules

After induction of decision rules, the next step of the proposed methodology for multicriteria ranking is the application of induced rules on set A. Each pair of objects $(a, b) \in A \times A$ can be covered by some probabilistic decision rule(s) suggesting assignment to relation S and/or to relation S^c. It can also be not covered by any rule. In order to address these possibilities, we introduce two relations over set A, denoted by $\mathbb{S}$ and $\mathbb{S}^c$, which result from application of decision rules on A. The precise definitions of these relations depend on two determinants. The first determinant is the admitted perception of the set of criteria (i.e., whether this set is assumed to be a consistent set or not). The second determinant is the way of perceiving decision rules, which may be, in a sense, qualitative or quantitative. Precisely, when considering a pair $(a, b) \in A \times A$ and comprehensive preference relation T, a DM may be interested to know:

(i) if there exists at least one rule suggesting assignment of (a, b) to T (qualitative information), or

(ii) what is the strength of the strongest rule suggesting assignment of (a, b) to T (quantitative information).

In case (i) above, relations $\mathbb{S}$ and $\mathbb{S}^c$ are crisp and thus, the preference structure composed of these relations is called the crisp preference structure. In case (ii) above, these relations are valued and thus, the preference structure composed of these relations is called the valued preference structure; precise values of $\mathbb{S}(a, b)$ and $\mathbb{S}^c(a, b)$ depend, obviously, on the adopted definition of rule strength.

We consider two alternative definitions of the strength σ of rule r_T:

$$\sigma(r_T) = \big(1 - \widehat{\epsilon}_T(r_T)\big), \tag{2.22}$$

$$\sigma(r_T) = \big(1 - \widehat{\epsilon}_T(r_T)\big) cf(r_T), \tag{2.23}$$

where $cf(r_T)$ denotes *coverage factor* of rule r_T, defined as the ratio of the number of pairs of objects supporting r_T and the cardinality of relation T. In this way, the higher the consistency of a rule (i.e., the lower the value $\widehat{\epsilon}_T(r_T)$), and the greater the number of pairs of objects supporting the rule, the stronger the rule is. Measuring rule strength using definition (2.22) employing cost-type rule consistency measure $\widehat{\epsilon}_T$, is analogous to measuring rule strength using gain-type rule confidence (credibility) measure (defined as the ratio of the number of pairs of objects supporting r_T and the number of pairs of objects covered by r_T) applied in [63]. Measuring rule strength using definition (2.23), introduced in [164], allows to take into account not only consistency of the rule, but also the relative number of pairs of objects supporting this rule.

In the following, we give four definitions of relations $\mathbb{S}$ and $\mathbb{S}^c$, corresponding to all combinations of the two determinants introduced in this section: consistent set of criteria or not necessarily consistent set of criteria on one hand, and qualitative or quantitative interpretation of rule matching on the other hand.

2.7.1 Definition of Relations $\mathbb{S}$ and $\mathbb{S}^c$ – Consistent Set of Criteria, Crisp Relations

When assuming that set G is a consistent set of criteria, and using only qualitative information concerning rule matching, relations $\mathbb{S}$ and $\mathbb{S}^c$ over A, induced by the sets of decision rules R_S and R_{S^c}, respectively, are defined as in [163]:

$$\mathbb{S} = \{(a,b) \in A \times A : (\exists\, r_S \in R_S \text{ such that } r_S \text{ covers } (a,b)) \text{ or } (aDb)\}, \quad (2.24)$$

$$\mathbb{S}^c = \{(a,b) \in A \times A : (\exists\, r_{S^c} \in R_{S^c} \text{ such that } r_{S^c} \text{ covers } (a,b)) \text{ and not } (aDb)\}, \quad (2.25)$$

where $\exists\, r_S \in R_S$ is read as "there exists a rule $r_S \in R_S$". Let us observe that relation $\mathbb{S}$ is reflexive and relation $\mathbb{S}^c$ is irreflexive. Moreover, relations $\mathbb{S}$ and $\mathbb{S}^c$ are, in general, neither transitive nor complete.

Considering the six versions of the rule-based approach to multicriteria ranking listed in Section 2.1, defining relations $\mathbb{S}$ and $\mathbb{S}^c$ by (2.24) and (2.25), respectively, corresponds to version VC-DRSA$^{rank}_{c\,0|1}$.

2.7.2 Definition of Relations $\mathbb{S}$ and $\mathbb{S}^c$ – Consistent Set of Criteria, Valued Relations

When assuming that set G is a consistent set of criteria, and using quantitative information concerning rule matching, relations $\mathbb{S}$ and $\mathbb{S}^c$ over A are defined as:

$$\mathbb{S}(a,b) = \begin{cases} \max\{\sigma(r_S) : r_S \in R_S, r_S \text{ covers } (a,b)\}, & \text{if not } aDb \\ 1, & \text{if } aDb \end{cases} \quad (2.26)$$

$$\mathbb{S}^c(a,b) = \begin{cases} \max\{\sigma(r_{S^c}) : r_{S^c} \in R_{S^c}, r_{S^c} \text{ covers } (a,b)\}, & \text{if not } aDb \\ 0, & \text{if } aDb \end{cases} \quad (2.27)$$

where $\sigma(r_S)$ denotes the strength of rule r_S. Let us observe that relation $\mathbb{S}$ is reflexive and relation $\mathbb{S}^c$ is irreflexive.

Considering the six versions of the rule-based approach to multicriteria ranking listed in Section 2.1, defining relations $\mathbb{S}$ and $\mathbb{S}^c$ by (2.26) and (2.27), respectively, corresponds to versions VC-DRSA$^{rank}_{c\,0-1_{cr}}$ and VC-DRSA$^{rank}_{c\,0-1_{\times}}$.

2.7.3 Definition of Relations $\mathbb{S}$ and $\mathbb{S}^c$ – Not Necessarily Consistent Set of Criteria, Crisp Relations

When considering set G to be a not necessarily consistent set of criteria, and using only qualitative information concerning rule matching, relations $\mathbb{S}$ and $\mathbb{S}^c$ over A, induced by the sets of decision rules R_S and R_{S^c}, respectively, are defined as:

$$\mathbb{S} = \{(a,b) \in A \times A : (\exists\, r_S \in R_S : r_S \text{ covers } (a,b)) \text{ or } (a=b)\}, \quad (2.28)$$

$$\mathbb{S}^c = \{(a,b) \in A \times A : (\exists\, r_{S^c} \in R_{S^c} : r_{S^c} \text{ covers } (a,b)) \text{ and not } (a=b)\}. \quad (2.29)$$

where $\exists\, r_S \in R_S$ is read as "there exists a rule $r_S \in R_S$". Let us observe that relation $\mathbb{S}$ is reflexive and relation $\mathbb{S}^c$ is irreflexive. Moreover, relations $\mathbb{S}$ and $\mathbb{S}^c$ are, in general, neither transitive nor complete.

Considering the six versions of the rule-based approach to multicriteria ranking listed in Section 2.1, defining relations $\mathbb{S}$ and $\mathbb{S}^c$ by (2.28) and (2.29), respectively, corresponds to version VC-DRSA$^{rank}_{nc\,0|1}$.

2.7.4 Definition of relations $\mathbb{S}$ and $\mathbb{S}^c$ – Not Necessarily Consistent Set of Criteria, Valued Relations

When considering set G to be a not necessarily consistent set of criteria, and using quantitative information concerning rule matching, relations $\mathbb{S}$ and $\mathbb{S}^c$ over A are defined as in [164]:

$$\mathbb{S}(a,b) = \begin{cases} \max\{\sigma(r_S) : r_S \in R_S, r_S \text{ covers } (a,b)\}, \text{ if } a \neq b \\ 1, \text{ if } a = b \end{cases} \tag{2.30}$$

$$\mathbb{S}^c(a,b) = \begin{cases} \max\{\sigma(r_{S^c}) : r_{S^c} \in R_{S^c}, r_{S^c} \text{ covers } (a,b)\}, \text{ if } a \neq b \\ 0, \text{ if } a = b \end{cases} \tag{2.31}$$

where $\sigma(r_S)$ denotes the strength of rule r_S. Let us observe that relation $\mathbb{S}$ is reflexive and relation $\mathbb{S}^c$ is irreflexive.

Considering the six versions of the rule-based approach to multicriteria ranking listed in Section 2.1, defining relations $\mathbb{S}$ and $\mathbb{S}^c$ by (2.30) and (2.31), respectively, corresponds to versions VC-DRSA$^{rank}_{nc\,0-1_{cr}}$ and VC-DRSA$^{rank}_{nc\,0-1_{\times}}$.

2.7.5 Four-valued Outranking – Crisp Relations

It is worth noting that the information contained in crisp relations $\mathbb{S}$ and $\mathbb{S}^c$ over set A can be represented using the *four-valued outranking* model of preferences, introduced in [170, 171] (see also [93]). In this model, given a pair of objects $(a,b) \in A \times A$, one considers four possible situations of outranking:

- *true outranking*, denoted by $a\mathbb{S}^T b$, iff $a\mathbb{S}b$ and not $a\mathbb{S}^c b$,
- *false outranking*, denoted by $a\mathbb{S}^F b$, iff not $a\mathbb{S}b$ and $a\mathbb{S}^c b$,
- *unknown outranking*, denoted by $a\mathbb{S}^U b$, iff not $a\mathbb{S}b$ and not $a\mathbb{S}^c b$,
- *contradictory outranking*, denoted by $a\mathbb{S}^K b$, iff $a\mathbb{S}b$ and $a\mathbb{S}^c b$.

The relations $\mathbb{S}^T$, $\mathbb{S}^F$, $\mathbb{S}^U$, $\mathbb{S}^K$, defined over A, correspond to the four truth values of Belnap logic [11, 12]: T (true), F (false), U (unknown), and K (contradictory).

2.7.6 Four-valued Outranking – Valued Relations

The information contained in valued relations $\mathbb{S}$ and $\mathbb{S}^c$ over set A can be also represented using four graded (valued) outranking relations $\mathbb{S}^T, \mathbb{S}^F, \mathbb{S}^U, \mathbb{S}^K \in \mathbf{R}_A$, considered, e.g., in [64, 138, 170]. These binary relations also correspond to the four truth values of Belnap logic [11, 12]: T (true), F (false), U (unknown), and K (contradictory). Relations $\mathbb{S}^T$, $\mathbb{S}^F$, $\mathbb{S}^U$, and $\mathbb{S}^K$, are defined in the following way:

$$\mathbb{S}^T(a,b) = min(\mathbb{S}(a,b), 1 - \mathbb{S}^c(a,b)), \tag{2.32}$$

$$\mathbb{S}^F(a,b) = min(1 - \mathbb{S}(a,b), \mathbb{S}^c(a,b)), \tag{2.33}$$

$$\mathbb{S}^K(a,b) = min(\mathbb{S}(a,b), \mathbb{S}^c(a,b)), \tag{2.34}$$

$$\mathbb{S}^U(a,b) = min(1 - \mathbb{S}(a,b), 1 - \mathbb{S}^c(a,b)), \tag{2.35}$$

where $a, b \in A$. It is worth noting that the above definitions relate to preference, indifference, and incomparability indices introduced by Bisdorff [18]. Moreover, the above definitions are generalizations of the corresponding definitions given in Section 2.7.5.

2.7.7 Preference Graph

The preference structure on A, composed of $\mathbb{S}$ and $\mathbb{S}^c$, can be represented by a *preference graph*. It is a directed multigraph $\mathbb{G}$. Each vertex (node) v_a of the preference graph corresponds to exactly one object $a \in A$. One can distinguish in $\mathbb{G}$ two types of arcs: $\mathbb{S}$-arcs and $\mathbb{S}^c$-arcs. In case of a crisp preference structure, an $\mathbb{S}$-arc ($\mathbb{S}^c$-arc) from vertex v_a to vertex v_b indicates that $a\mathbb{S}b$ (respectively, $a\mathbb{S}^c b$). In case of a valued preference structure, each $\mathbb{S}$-arc between vertices v_a and v_b is weighted by value $\mathbb{S}(a,b)$. Analogously, each $\mathbb{S}^c$-arc between vertices v_a and v_b is weighted by value $\mathbb{S}^c(a,b)$. $\mathbb{G}$ is a multigraph since there may be one $\mathbb{S}$-arc and one $\mathbb{S}^c$-arc for each pair of objects $(a,b) \in A \times A$. A *final recommendation* for the multicriteria ranking problem at hand, in terms of a total or partial preorder of all objects belonging to set A, can be obtained upon a suitable exploitation of the preference graph.

2.8 Exploitation of Preference Graph

The exploitation of preference graph $\mathbb{G}$ resulting from application of induced decision rules on set A is not an easy task, especially because this graph represents two crisp/valued relations $\mathbb{S}$ and $\mathbb{S}^c$. This task is more complex than the exploitation of a preference graph representing a crisp or valued relation, well studied in the literature [7, 28–31, 34–36, 49, 137, 139, 173].

Preference graphs representing a crisp relation are obtained, e.g., in some decision aiding methods proposed in the field of MCDA, in which preferences of a DM are modeled in terms of binary relations. Among these methods, one can mention, e.g., ELECTRE IS [57, 147]. Preference graphs representing a valued relation are obtained, e.g., when using ELECTRE III [57, 142], or PROMETHEE I, or PROMETHEE II [37, 38] methods. When preferences are modeled in terms of binary relations, the key question is the existence of evidence in favor of the considered relation. For example, in case of outranking relation S concerned in the methods from ELECTRE family, the evidence concerns the sentence aSb and/or bSa, for any pair of objects $a, b \in A$. In reality, the evidence is often incomplete, thus inducing a graded (valued) relation aSb, i.e., "a is at least as good as b, up to a certain degree of certainty".

It is reasonable to claim that considering only evidence in favor of the considered binary relation does not allow to catch the reality of some decision problems. In fact, such an approach leads to the situation where the evidence in disfavor of a sentence is semantically considered – and thus modeled – as the evidence in favor of the opposite sentence. This mental restriction may induce not only misunderstandings but, which is even more important, it may also imply some loss of information (a good example clarifying this point, concerning government composition, is presented in [64]). Therefore, in this thesis, given a pair of objects $(a, b) \in A \times A$, we consider not only the decision rules supporting conclusion aSb, but also the rules supporting the opposite conclusion, i.e., conclusion aS^cb. In this way, we take into account the arguments in favor of preference of a over b and in disfavor of it; in the following, they will also be called positive and negative arguments, respectively. As described in Section 2.7, in case of a valued preference structure, the accumulated strength of the positive arguments is reflected by the value $\mathbb{S}(a, b)$, while the accumulated strength of the negative arguments is reflected by the value $\mathbb{S}^c(a, b)$.

2.8.1 Review of Possible Exploitation Procedures

Given a preference graph $\mathbb{G}$, one can propose several exploitation techniques that lead to final recommendation for the multicriteria ranking problem at hand, in terms of a total or partial preorder of all objects from set A. We distinguish the following approaches:

(i) direct exploitation of relations $\mathbb{S}$ and $\mathbb{S}^c$ by the *Net Flow Score* (NFS) procedure [93] (when $\mathbb{S}$ and $\mathbb{S}^c$ are crisp) or *Fuzzy Net Flow Score* (FNFS) procedure [63] (when $\mathbb{S}$ and $\mathbb{S}^c$ are valued),

(ii) exploitation of the four graded outranking relations $\mathbb{S}^T$, $\mathbb{S}^F$, $\mathbb{S}^U$, and $\mathbb{S}^K$ (see Section 2.7.6) in the way proposed in [64] (when $\mathbb{S}$ and $\mathbb{S}^c$ are valued),

(iii) independent exploitation of relations $\mathbb{S}$ and $\mathbb{S}^c$,

(iv) suitable transformation of preference graph $\mathbb{G}$ to another graph $\mathbb{G}'$ representing a valued relation, then exploitation of this relation leading to a total or partial preorder over A.

Approach (i) is based on scoring function $S^{NF} : A \to \mathbb{R}$ defined as:

$$S^{NF}(a) = \sum_{b \in A \setminus \{a\}} \mathbb{S}(a, b) - \mathbb{S}(b, a) - \mathbb{S}^c(a, b) + \mathbb{S}^c(b, a). \tag{2.36}$$

Function S^{NF} induces a total preorder over A, which is the solution of the considered multicriteria ranking problem.

In approach (ii), one associates with each object $a \in A$ a vector $\bar{a}$, defined as:

$$\bar{a} = (t_b, t_c, \ldots, f_b, f_c, \ldots),$$

where t_b (respectively, f_b) is equal to $\mathbb{S}^T(a, b)$ (respectively, $\mathbb{S}^F(b, a)$). Then, all vectors corresponding to objects from set A are sorted in the non-decreasing order and compared lexicographically (to resolve ties, one can also take into account for each object $a \in A$ additional vectors composed of values $\mathbb{S}^K(a, \cdot)$ or $\mathbb{S}^U(a, \cdot)$). Such leximin-scoring procedure yields a partial preorder over A.

The general idea of approach (iii) is to exploit relations $\mathbb{S}$ and $\mathbb{S}^c$ independently, obtaining two separate total or partial preorders, and then to conjunct these preorders in the same way as in the ELECTRE III method [57, 142]. This leads to obtaining a partial preorder over A.

In this thesis, we focus on approach (iv), which is the most generic one. The first step of this approach, i.e., the transformation of preference graph $\mathbb{G}$ to graph $\mathbb{G}'$, is presented in the following Section 2.8.2. The second step of this approach consists in application of a so-called *ranking method* to exploit the valued relation obtained in step one. Several existing ranking methods that can be employed in the second step are reviewed in Section 2.8.3.

We concentrate on approach (iv) mainly for the following reasons.

- First, the exploitation of a valued relation over a set of objects has been widely studied in the literature [7, 28–31, 34–36, 49, 137, 139]. These prior studies supply us with several potentially useful ranking methods.
- Second, the diversity of ranking methods proposed in the literature calls for a systematic comparison of their formal properties, which is, however, missing.
- Third, as we explain in the last paragraph of Section 2.8.3, using a suitable transformation of preference graph $\mathbb{G}$ and a particular ranking method to exploit the transformed graph $\mathbb{G}'$, it is possible to obtain the same final ranking as in approach (i). Thus, approach (iv) can be seen as a framework that encompasses approach (i).
- Fourth, when applied to set A, most of the ranking methods considered in the literature yield a total preorder over A, which is generally acknowledged to be more operational for a DM than a partial preorder that can be obtained in approaches (ii) and (iii).

2.8.2 Fusion of Relations $\mathbb{S}$ and $\mathbb{S}^c$ in Order to Exploit a Single Relation

The suitable transformation of preference graph $\mathbb{G}$ representing two relations $\mathbb{S}$ and $\mathbb{S}^c$ to graph $\mathbb{G}'$ representing one valued relation $\mathcal{R} \in \mathbf{R}_A$ (that can be further exploited using a ranking method) consists in defining relation $\mathcal{R}$ in the following way:

$$\mathcal{R}(a,b) = \frac{\mathbb{S}(a,b) + (1 - \mathbb{S}^c(a,b))}{2}, \tag{2.37}$$

where $a, b \in A$. Let us observe that scoring function S^{NF} defined by (2.36) can be expressed in terms of $\mathcal{R}$ as: $S^{NF}(a) = 2[\sum_{b \in A \setminus \{a\}} \mathcal{R}(a,b) - \mathcal{R}(b,a)]$. Moreover, relation $\mathcal{R}$ is reflexive.

Note that when relations $\mathbb{S}$ and $\mathbb{S}^c$ are crisp, relation $\mathcal{R}$ is identical to the valued relation over A introduced in [93], originally denoted by R_{4v}, defined as:

$$R_{4v}(a,b) = \begin{cases} 0 & \text{if } a\mathbb{S}^F b \\ \frac{1}{2} & \text{if } a\mathbb{S}^U b \text{ or } a\mathbb{S}^K b \\ 1 & \text{if } a\mathbb{S}^T b \end{cases},$$

where $a, b \in A$, and relations $\mathbb{S}^T$, $\mathbb{S}^F$, $\mathbb{S}^U$, and $\mathbb{S}^K$, are defined as in Section 2.7.5. When $\mathcal{R}(a,b) \in \{0, \frac{1}{2}, 1\}$ for any $(a,b) \in A \times A$, we call relation $\mathcal{R}$ a *three-valued relation.*

In the following, considering exploitation of relation $\mathcal{R}$, we assume that this relation has no “structural properties” [32], i.e., we assume (what seems to be the case) that:

- given $\mathbb{S}$ and $\mathbb{S}^c$ are crisp, $\mathcal{R}$ may be *any* three-valued relation over A,
- given $\mathbb{S}$ and $\mathbb{S}^c$ are valued, $\mathcal{R}$ may be *any* valued relation over A.

The rationale for this assumption is that relation $\mathcal{R}$ is determined by a considered set of decision rules, and, in general, this set of rules does not depend on A.

2.8.3 Review of Ranking Methods

In this section, we review different ranking methods that can be applied to exploit valued relation $\mathcal{R}$ defined by (2.37).

In the literature, one can find many ranking methods “dedicated” to exploitation of a valued relation over a set of objects [28, 30, 34–36, 139]. On the other hand, as argued by Arrow and Raynaud [5], one can also be interested in another approach to rank objects which consists in (downward) iterative application of a *choice function.* Let us denote by $\mathcal{P}_A$ the set of all nonempty subsets of a finite set of objects A. Then, choice function cf is a function

$$cf : \mathcal{P}_A \times \mathbf{R}_A \to \mathcal{P}_A. \tag{2.38}$$

A choice function associates with each nonempty set $A' \subseteq A$ and each valued relation R over A, a nonempty *choice set* $cf(A', R) \subseteq A'$, which may be interpreted as the set of the “best” objects in A' given relation R. Iterative application of a choice function on a finite set A was considered, e.g., in [33, 35, 36]. It leads to obtaining a total preorder over A. Let us denote by $A^i \subseteq A$ the set of objects considered in i-th iteration and by $|A|$ the cardinality of set A. Obviously, $A^1 = A$. In i-th iteration, $i \in \{1, 2, \ldots, |A|\}$, given choice function cf is applied to set A^i. Then, the objects belonging to choice set $cf(A^i, R)$ are put in i-th rank of the constructed ranking and removed from set A^i. Thus, $A^{i+1} = A^i \setminus cf(A^i, R)$. The construction of a final ranking is finished when this ranking contains all objects from set A.

Most of the proposed "dedicated" ranking methods as well as ranking methods based on iterative application of a choice function employ a *scoring function*. Given a finite set of objects A and a valued relation R over A, scoring function is used to evaluate relative performance of each object $a \in A$ w.r.t. the objects in nonempty set $A' \subseteq A$, taking into account relation R. Thus, scoring function sf is a function

$$sf : A \times \mathcal{P}_A \times \mathbf{R}_A \rightarrow \mathbb{R}. \tag{2.39}$$

Value $sf(a, A', R)$ denotes the *score* of object $a \in A$ calculated w.r.t. the objects in $A' \subseteq A$, given valued relation R.

We define two generic score-based ranking methods: *single-stage ranking method* ($\succeq^1$) and *multi-stage ranking method* ($\succeq^i$). These ranking methods are parameterized by a set of objects A, a valued relation R over A, and a scoring function sf. Moreover, they yield a total preorder over A, by performing the following steps:

- $\succeq^1(A, R, sf)$:
  (1) assign score $sf(a, A, R)$ to each object $a \in A$;
  (2) rank all the objects from set A according to their scores, in such a way that the higher the score of an object, the lower its rank (objects with the same score have the same rank);
- $\succeq^i(A, R, sf)$:
  (1) define choice function cf as follows:
  $cf(A', R) = \{a \in A' : sf(a, A', R) \geq sf(b, A', R)$ for all $b \in A'\}$, where $A' \subseteq A$, i.e., in such a way that it chooses subset of A' composed of objects with the highest score;
  (2) perform (downward) iterative application of the above choice function cf on set A.

Clearly, the aforementioned "dedicated" ranking methods are *instances* of $\succeq^1$, differing only by the definition of function sf. Analogously, ranking methods based on iterative choice considered in [35, 36] are instances of $\succeq^i$, differing only by the definition of function sf.

Let us consider a finite set of objects A and a valued relation R over A. Then, according to Barrett et al. [7], the score of any object $a \in A$ w.r.t. the objects in any set $A' \subseteq A$

can be calculated using one of the following scoring functions:

$$\begin{aligned}
&\textit{max in favor}: && MF(a,A',R) &&= \max_{b\in A'\setminus\{a\}} R(a,b), && (2.40)\\
&\textit{min in favor}: && mF(a,A',R) &&= \min_{b\in A'\setminus\{a\}} R(a,b), && (2.41)\\
&\textit{sum in favor}: && SF(a,A',R) &&= \sum_{b\in A'\setminus\{a\}} R(a,b), && (2.42)\\
&\textit{–max against}: && -MA(a,A',R) &&= -\max_{b\in A'\setminus\{a\}} R(b,a), && (2.43)\\
&\textit{–min against}: && -mA(a,A',R) &&= -\min_{b\in A'\setminus\{a\}} R(b,a), && (2.44)\\
&\textit{–sum against}: && -SA(a,A',R) &&= -\sum_{b\in A'\setminus\{a\}} R(b,a), && (2.45)\\
&\textit{max difference}: && MD(a,A',R) &&= \max_{b\in A'\setminus\{a\}} R(a,b)-R(b,a), && (2.46)\\
&\textit{min difference}: && mD(a,A',R) &&= \min_{b\in A'\setminus\{a\}} R(a,b)-R(b,a), && (2.47)\\
&\textit{sum of differences}: && SD(a,A',R) &&= \sum_{b\in A'\setminus\{a\}} R(a,b)-R(b,a). && (2.48)
\end{aligned}$$

It is worth noting that $SD(a,A',R)$ is a sum of $SF(a,A',R)$ and $-SA(a,A',R)$.

In this thesis, given a finite set of objects A and a valued relation R over A, we consider exploitation of relation R using one of the following ranking methods, well studied in the literature:

(1) *Net Flow Rule* [30, 36], defined as:

$$NFR(A,R) = \succeq^1(A,R,SD), \tag{2.49}$$

(2) *Iterative Net Flow Rule* [36], defined as:

$$It.NFR(A,R) = \succeq^i(A,R,SD), \tag{2.50}$$

(3) *Min in Favor* [28, 35, 36, 139], defined as:

$$MiF(A,R) = \succeq^1(A,R,mF), \tag{2.51}$$

(4) *Iterative Min in Favor* [35], defined as:

$$It.MiF(A,R) = \succeq^i(A,R,mF), \tag{2.52}$$

(5) *Leaving and Entering Flows* [34], defined as:

$$L/E(A,R) = \succeq^1(A,R,SF) \cap \succeq^1(A,R,-SA). \tag{2.53}$$

As can be seen, considered ranking methods employ only some of the defined scoring functions, namely: mF (2.41), SF (2.42), $-SA$ (2.45), and SD (2.48).

NFR orders objects according to their net flow scores. It has a long history in *social choice theory* [4, 59]. It coincides with the rule of Copeland [cf. 59, 100, 149] when R is crisp. When $R(a, b)$ is interpreted as a percentage of voters considering that a is preferred or indifferent to b ($a, b \in A$), it corresponds to the well-known rule of Borda [cf. 59, 179]. Moreover, NFR is used in the PROMETHEE II outranking method [37, 38].

$It.NFR$ consists in iterative application of a choice function that chooses objects with the highest value of scoring function SD (2.48). This ranking method was originally called the *Repeated Net Flow Rule* and denoted by $RNFR$ [36].

L/E is used in the PROMETHEE I method [37, 38]. This ranking method allows any two objects $a, b \in A$ to be declared incomparable. This happens when two conclusions concerning ranking of these objects, one conclusion resulting only from the comparison of their *leaving flows*, i.e., values $SF(\cdot, A, R)$, and the other one resulting only from the comparison of their *entering flows*, i.e., values $-(-SA(\cdot, A, R))$, are contradictory. Such contradiction occurs, e.g., when $SF(a, A, R) > SF(b, A, R)$, while $-SA(a, A, R) < -SA(b, A, R)$.

It should be noted that NFR and L/E make use of the "cardinal" properties of values $R(a, b)$, with $a, b \in A$. On the other hand, MiF represents a prudent approach as it is purely "ordinal" – it uses values $R(a, b)$ as if they were a numerical representation of a credibility of a crisp relation between a and b. Thus, from the fact that $R(a, b) \geq R(c, d)$ it concludes only that the relation between a and b is not less credible than the relation between c and d, with $a, b, c, d \in A$.

The number of conceivable ranking methods calls for a systematic comparison of their properties, which is, however, missing in the literature. Therefore, in Chapter 3, we define several desirable properties of a ranking method, and we compare w.r.t. these properties the well-known ranking methods described in (2.49), (2.50), (2.51), (2.52), and (2.53).

Now, let us come back and explain the sentence "approach (iv) can be seen as a framework that encompasses approach (i)", which appeared in the context of the four approaches for exploitation of preference graph $\mathbb{G}$ (see Section 2.8.1). By saying this, we meant that the total preorder over A obtained using scoring function S^{NF} (2.36) is the same as the total preorder over A obtained using $NFR(A, \mathcal{R})$ (2.49).

2.9 Analysis of Final Ranking

In this section, we propose a way of measuring concordance between a final ranking $\succeq(A, \mathcal{R})$ being a partial preorder (in particular a total preorder) over A, and the initial pairwise comparisons of reference objects from set $A^R \subseteq A$. Remark that the issue of measuring this concordance is only relevant for the case when $|A^R \cap A| \geq 2$ (i.e., when at least two reference objects appear in the final ranking). Below, for the sake of simplicity, we focus on a particular sub-case when $A^R \subseteq A$, and $|A^R| \geq 2$, which is particularly interesting from the MCDA perspective. Moreover, we use the notation introduced in Sections 2.2 and 2.4. First, in Section 2.9.1, we remind the definition of the *Kendall rank correlation coefficient* τ, which is a classical measure used to measure concordance of two total preorders. Second, in Section 2.9.2, we define a new concordance measure τ', which is a generalization of measure τ, suited to the aforementioned task of measuring the concordance between a partial preorder (in particular a total preorder) and pairwise comparisons. In both sections, we denote by $\succeq_{A'}$ a partial preorder over set $A' \subseteq A$. Moreover, we use the following representation of $\succeq_{A'}$ in terms of strict preference, inverse strict preference, indifference, and incomparability relations over set A':

$$aP_{\succeq_{A'}}b \Leftrightarrow a \succeq_{A'} b \text{ and not } b \succeq_{A'} a, \tag{2.54}$$

$$aP^{-1}_{\succeq_{A'}}b \Leftrightarrow \text{not } a \succeq_{A'} b \text{ and } b \succeq_{A'} a, \tag{2.55}$$

$$aI_{\succeq_{A'}}b \Leftrightarrow a \succeq_{A'} b \text{ and } b \succeq_{A'} a, \tag{2.56}$$

$$aJ_{\succeq_{A'}}b \Leftrightarrow \text{not } a \succeq_{A'} b \text{ and not } b \succeq_{A'} a, \tag{2.57}$$

where $a, b \in A'$. Thus, $aP_{\succeq_{A'}}b$ iff object a is ranked higher than object b, $aP^{-1}_{\succeq_{A'}}b$ iff object a is ranked lower than object b, $aI_{\succeq_{A'}}b$ iff the ranks of objects a and b are equal, and $aJ_{\succeq_{A'}}b$ iff objects a and b are incomparable in partial preorder $\succeq_{A'}$.

2.9.1 Kendall Rank Correlation Coefficient τ

Let $\succeq_{A^R}$ and $\succeq_A$ be two total preorders, and let $|A^R| \geq 2$. Then

$$\tau(\succeq_{A^R}, \succeq_A) = 1 - 2\frac{\sum\limits_{(a,b)\in A^R\times A^R, a\neq b} err(a,b)}{|\{(a,b) \in A^R \times A^R : a \neq b\}|}, \tag{2.58}$$

where $err(a, b)$ denotes an error accounted for a pair of objects $(a, b) \in A^R \times A^R$, $a \neq b$. This error is defined as:

$$err(a,b) = \begin{cases} 0, & \text{if } (aP_{\succeq_{A^R}}b \text{ and } aP_{\succeq_A}b) \text{ or } (aP^{-1}_{\succeq_{A^R}}b \text{ and } aP^{-1}_{\succeq_A}b) \text{ or } (aI_{\succeq_{A^R}}b \text{ and } aI_{\succeq_A}b) \\ \frac{1}{2}, & \text{if } \Big((aP_{\succeq_{A^R}}b \text{ or } aP^{-1}_{\succeq_{A^R}}b) \text{ and } aI_{\succeq_A}b\Big) \text{ or } \Big(aI_{\succeq_{A^R}}b \text{ and } (aP_{\succeq_A}b \text{ or } aP^{-1}_{\succeq_A}b)\Big) \\ 1, & \text{if } (aP_{\succeq_{A^R}}b \text{ and } aP^{-1}_{\succeq_A}b) \text{ or } (aP^{-1}_{\succeq_{A^R}}b \text{ and } aP_{\succeq_A}b) \end{cases}. \tag{2.59}$$

Thus, values of coefficient τ belong to the interval $[-1, 1]$. The best possible value of τ is 1 (which corresponds to the maximum concordance of the two total preorders), and the worst possible value is -1 (which, in turn, corresponds to the maximum discordance of the two total preorders).

2.9.2 New Concordance Measure τ'

Let $\succeq_A$ be a partial preorder. Then the concordance between $\succeq_A$ and pairwise comparisons of reference objects from set $A^R \subseteq A$, determining two disjoint crisp relations S

and S^c over A^R, can be measured as

$$\tau'(S, S^c, \succeq_A) = 1 - 2\frac{\sum_{(a,b)\in B, a\neq b} err'(a,b)}{|\{(a,b) \in B : a \neq b\}|}, \tag{2.60}$$

where $B = S \cup S^c$ and $err'(a,b)$ denotes an error accounted for a pair of objects $(a,b) \in B$, $a \neq b$. This error is defined as:

$$err'(a,b) = \begin{cases} 0, & \text{if } (aSb \text{ and } aP_{\succeq_A}b) \text{ or } (aSb \text{ and } aI_{\succeq_A}b) \text{ or } (aS^cb \text{ and } aP^{-1}_{\succeq_A}b) \\ & \text{or } (aS^cb \text{ and } aJ_{\succeq_A}b) \\ 1, & \text{if } (aSb \text{ and } aP^{-1}_{\succeq_A}b) \text{ or } (aSb \text{ and } aJ_{\succeq_A}b) \text{ or } (aS^cb \text{ and } aP_{\succeq_A}b) \\ & \text{or } (aS^cb \text{ and } aI_{\succeq_A}b) \end{cases}. \tag{2.61}$$

Thus, values of coefficient τ' belong to the interval $[-1, 1]$. The best possible value of τ' is 1 (maximum concordance), and the worst possible value is -1 (maximum discordance).

Now, let us show that measure τ' (2.60) is a generalization of measure τ (2.58). First, remind that the pairwise comparisons of reference objects considered in this thesis are a more general type of preference information than a reference ranking or an ordinal classification of reference objects (see Section 2.3). Second, observe that a total preorder $\succeq_{A^R}$ can be represented in terms of relations $P_{\succeq_{A^R}}$, $P^{-1}_{\succeq_{A^R}}$, and $I_{\succeq_{A^R}}$, using properties (2.54), (2.55), and (2.56), respectively. Moreover, the resulting relations $P_{\succeq_{A^R}} \equiv P$, $P^{-1}_{\succeq_{A^R}} \equiv P^{-1}$, and $I_{\succeq_{A^R}} \equiv I$, can be represented in terms of relations S and S^c using properties (2.4), (2.5), and (2.6), respectively. Thus, outranking and non-outranking relations implied by total preorder $\succeq_{A^R}$ are the following:

- aSb if $a \succeq_{A^R} b$,
- aS^cb if not $a \succeq_{A^R} b$,

as already observed in Section 2.3. In such case, $B = S \cup S^c$ is equal to $A^R \times A^R$, and thus, the right hand sides of (2.58) and (2.60) differ only by the applied error measure. Third, let $\succeq_{A^R}$ and $\succeq_A$ be two total preorders, $|A^R| \geq 2$, and let S and S^c be the outranking and non-outranking relations implied by total preorder $\succeq_{A^R}$. Then

$$err(a,b) + err(b,a) = err'(a,b) + err'(b,a)$$

and thus,

$$\tau(\succeq_{A^R}, \succeq_A) = \tau'(S, S^c, \succeq_A).$$

2.10 Illustrative Example

In this section, we present an illustrative example concerning application of the proposed rule-based methodology for multicriteria ranking. In this example, for the purpose of obtaining the final ranking of considered objects, we apply scoring function S^{NF} (2.36), i.e., we exploit the preference graph, yielded by application of induced decision rules, using approach (i) of Section 2.8.1.

Let us consider a hypothetical DM who is a scientist and wants to buy a notebook for personal use. The DM would like to spend no more than €1700. The DM is going to use the notebook for writing scientific papers, programming, performing some computational experiments, and watching movies in her/his free time. For these reasons, the DM considers only 22 high-end notebooks, that have Intel Core i7 processor with four cores, at least 4 MB of RAM (DDR3, 1333MHz), and at least a 15 in. monitor with full high-definition resolution (1920 x 1080 pixels). The DM evaluates the notebooks by three cardinal criteria: price in €(g_1, to be minimized), diagonal of a monitor in inches (g_2, to be maximized), and weight in kilograms (g_3, to be minimized). The weight is important because of the work-related travels (e.g., attending conferences). The evaluations of all 22 notebooks using the three considered criteria are given in Table 2.1. The data come from the Internet store www.komputronik.pl.

TABLE 2.1: Multicriteria evaluations of the considered notebooks

Id	Model	Price (g_1)	Diagonal (g_2)	Weight (g_3)
n_1	Asus N75SF-V2G-TZ025V	865	17.3	3.4
n_2	Asus N75SF-V2G-TZ149V	877	17.3	3.4
n_3	Asus N75SL-V2G-TZ043V	1066	17.3	3.4
n_4	DELL XPS L502X	1031	15.6	2.7
n_5	Asus N55SL-S1072V	1042	15.6	2.84
n_6	Asus X93SM-YZ071V	971	18.4	4.11
n_7	DELL XPS 15	1372	15.6	2.51
n_8	DELL XPS L702X	1254	17.3	3.43
n_9	Samsung NP700G7A-S01PL	1656	17.3	3.2
n_{10}	Samsung NP700G7A-S02PL	1656	17.3	3.81
n_{11}	Asus G53SW-SZ141	1161	15.6	3.8
n_{12}	Asus G53SX-IX059V	1372	15.6	3.92
n_{13}	Asus G53SX-S1163V	1348	15.6	3.92
n_{14}	Asus G73SW-91037V	1538	17.3	3.9
n_{15}	Asus G74SX-TZ055V	1372	17.3	4.28
n_{16}	Asus G74SX-TZ210V	1419	17.3	4.28
n_{17}	Asus VX7SX-S1090V	1538	15.6	3.82
n_{18}	Lenovo ThinkPad T520	1467	15.6	2.5
n_{19}	Lenovo ThinkPad W520	1538	15.6	2.61
n_{20}	Sony VAIO VPC-F21Z1E	1467	16.0	3.1
n_{21}	Sony VAIO VPC-F23S1E	1419	16.4	3.1
n_{22}	Sony VAIO VPC-SE2V9E	1419	15.5	1.96

The set of objects from Table 2.1 constitutes set A of objects to be ranked. In the past, the DM has tested notebooks n_1, n_4, n_{10}, n_{12}, n_{14}, and n_{18} personally. These six objects constitute set A^R of reference objects. Based on personal experience, the DM is able to rank the six reference objects as follows: $n_4 \succ n_1 \succ n_{12} \succ n_{14} \succ n_{10} \succ n_{18}$ (i.e., object n_4

is the best, object n_1 is second best, ..., object n_{18} is the worst). Let us observe that the ranking of reference objects can be used as a source of pairwise preference information. Therefore, given any two notebooks $a, b \in A^R$, we fix aSb whenever notebook a is ranked by the DM not lower than notebook b. Moreover, we fix aS^cb whenever notebook a is ranked lower than notebook b (in all such cases we have "not aDb"). In this way, we get $B = A^R \times A^R$. Remark that relations S and S^c are the same in case of set $G = \{g_1, g_2, g_3\}$ being either consistent set of criteria or not necessarily consistent set of criteria.

Given the preference information, the following calculations are performed using jRank[1] software [165].

The preference information in the form of pairwise comparisons of six reference objects yields a PCT composed of 36 pairs of objects. This PCT is shown in Table 2.2. Let us note that the cardinality of relation S is 21, and the cardinality of relation S^c is 15.

One can observe in Table 2.2 several pairs of objects inconsistent w.r.t. dominance relation D_2 over set B. Such inconsistency occurs when a pair of objects $(a, b) \in S$ is dominated by a pair of objects $(c, d) \in S^c$. Inconsistent pairs of objects appearing in Table 2.2 are marked in this table by an asterisk. All inconsistencies w.r.t. dominance relation D_2 over B are also presented in Table 2.3, where an asterisk indicates that pair $(a, b) \in S$ from the corresponding row is inconsistent with pair $(c, d) \in S^c$ from the corresponding column.

In order to show an advantage of the proposed PCT-oriented adaptation of ϵ-VC-DRSA to multicriteria ranking over previous PCT-oriented adaptations of DRSA to multicriteria ranking, we consider two independent calculation paths, taking the PCT shown in Table 2.2 as a "point of departure". Considering the steps (s_1)-(s_5) of decision rule-based methods presented in Section 1.2.1, this corresponds to the situation when step (s_1) has already been performed and we start calculations from step (s_2). Then, the two considered calculation paths are composed of the following steps:

- (s_2) calculation of lower approximations of relations S and S^c, according to definitions (2.14) and (2.15), respectively,
- (s_3) calculation of a minimal set of decision rules by VC-DomLEM algorithm,
- (s_4) application of the induced rules on set A,
- (s_5) exploitation of the preference graph, resulting from application of induced decision rules, using scoring function S^{NF} (2.36), i.e., using approach (i) of Section 2.8.1,
- (s_6) evaluation of the obtained final ranking on set A using measure τ' (2.60).

In step (s_4) above, for simplicity, we consider the case of a crisp preference structure, composed of crisp relations $\mathbb{S}$ and $\mathbb{S}^c$ calculated using definitions (2.28) and (2.29), respectively.

[1] See http://www.cs.put.poznan.pl/mszelag/Software/jRank/jRank.html.

Table 2.2: The PCT yielded by pairwise comparisons of six reference objects

(a,b)	Δ_1	Δ_2	Δ_3	Relation
(n_4,n_4)	0	0.0	0.0	S
$(n_4,n_1)^*$	166	-1.7	-0.7	S
(n_4,n_{12})	-341	0.0	-1.22	S
(n_4,n_{14})	-507	-1.7	-1.2	S
(n_4,n_{10})	-625	-1.7	-1.11	S
(n_4,n_{18})	-436	0.0	0.2	S
$(n_1,n_4)^*$	-166	1.7	0.7	S^c
(n_1,n_1)	0	0.0	0.0	S
(n_1,n_{12})	-507	1.7	-0.52	S
(n_1,n_{14})	-673	0.0	-0.5	S
(n_1,n_{10})	-791	0.0	-0.41	S
(n_1,n_{18})	-602	1.7	0.9	S
(n_{12},n_4)	341	0.0	1.22	S^c
(n_{12},n_1)	507	-1.7	0.52	S^c
(n_{12},n_{12})	0	0.0	0.0	S
$(n_{12},n_{14})^*$	-166	-1.7	0.02	S
(n_{12},n_{10})	-284	-1.7	0.11	S
$(n_{12},n_{18})^*$	-95	0.0	1.42	S
(n_{14},n_4)	507	1.7	1.2	S^c
(n_{14},n_1)	673	0.0	0.5	S^c
$(n_{14},n_{12})^*$	166	1.7	-0.02	S^c
(n_{14},n_{14})	0	0.0	0.0	S
(n_{14},n_{10})	-118	0.0	0.09	S
$(n_{14},n_{18})^*$	71	1.7	1.4	S
(n_{10},n_4)	625	1.7	1.11	S^c
(n_{10},n_1)	791	0.0	0.41	S^c
(n_{10},n_{12})	284	1.7	-0.11	S^c
(n_{10},n_{14})	118	0.0	-0.09	S^c
(n_{10},n_{10})	0	0.0	0.0	S
$(n_{10},n_{18})^*$	189	1.7	1.31	S
(n_{18},n_4)	436	0.0	-0.2	S^c
(n_{18},n_1)	602	-1.7	-0.9	S^c
$(n_{18},n_{12})^*$	95	0.0	-1.42	S^c
$(n_{18},n_{14})^*$	-71	-1.7	-1.4	S^c
$(n_{18},n_{10})^*$	-189	-1.7	-1.31	S^c
(n_{18},n_{18})	0	0.0	0.0	S

The above steps (s_2)-(s_6) "produce" the following results: lower approximations of outranking and non-outranking relations obtained in step (s_2), sets R_S and R_{S^c} of minimal decision rules obtained in step (s_3), crisp preference structure on A obtained in step (s_4), final ranking (total preorder) on A obtained in step (s_5), and value of concordance measure τ' obtained in step (s_6). These results differ, however, in both calculation paths only due to decisions made in step (s_2), concerning consistency thresholds θ_S and θ_{S^c} used to calculate lower approximations (2.14) and (2.15). In the first path, denoted by $cp_{\theta=0}$, we assume that both consistency thresholds are equal to zero. Thus, calculated lower ap-

TABLE 2.3: Inconsistencies in the PCT yielded by pairwise comparisons of six reference objects

$(a,b) \in S \downarrow \mid\mid (c,d) \in S^c \rightarrow$	(n_1, n_4)	(n_{14}, n_{12})	(n_{18}, n_{12})	(n_{18}, n_{14})	(n_{18}, n_{10})
(n_4, n_1)			*	*	*
(n_{12}, n_{14})					*
(n_{12}, n_{18})	*				
(n_{14}, n_{18})	*				
(n_{10}, n_{18})	*	*			

proximations are the same as the ones obtained using definitions (2.10) and (2.11). In the second path, denoted by $cp_{\theta>0}$, we choose $\theta_S = \theta_{S^c} = 0.1$. In this way, we relax a little bit the conditions for inclusion of pairs of objects to lower approximations (2.14) and (2.15). In particular, a pair of objects $(a, b) \in S$ is considered to be sufficiently consistent (and thus included in $\underline{S}$) if it is dominated by at most one pair of objects belonging to relation S^c (this can be verified using definition (2.12): $1/15 = 0.067 < \theta_S = 0.1 < 2/15 = 0.133$). Moreover, a pair of objects $(a, b) \in S^c$ is considered to be sufficiently consistent (and thus included in $\underline{S^c}$) if it dominates at most two pairs of objects belonging to relation S (this can be verified using definition (2.13): $2/21 = 0.095 < \theta_{S^c} = 0.1 < 3/21 = 0.143$).

Table 2.4 summarizes the results obtained in subsequent steps (s_2)-(s_6), along both calculations paths.

TABLE 2.4: Summary of results obtained in steps (s_2)-(s_6), for calculations paths $cp_{\theta=0}$, $cp_{\theta>0}$

Calculation path	$\lvert\underline{S}\rvert$	$\lvert\underline{S^c}\rvert$	$\lvert R_S\rvert$	$\lvert R_{S^c}\rvert$	τ'
$cp_{\theta=0}$	16	10	3	2	0.400
$cp_{\theta>0}$	19	14	3	2	0.733

Looking at Table 2.4, it is clear that the results obtained along calculation path $cp_{\theta>0}$ are better (greater lower approximations and a higher value of τ'). Thus, for the considered illustrative example, the PCT-oriented adaptation of ϵ-VC-DRSA to multicriteria ranking, proposed in this thesis, proved to be more useful than previous PCT-oriented adaptations of DRSA to multicriteria ranking. In view of this conclusion, in the following, we present only the results obtained along calculation path $cp_{\theta>0}$.

The set of minimal decision rules induced by VC-DomLEM algorithm is presented in Table 2.5, where "Supp" denotes the number of pairs of objects that support given rule r_T. Remark that the third rule covers all pairs of objects $(a, b) \in A \times A$ such that aDb, and that none of the rules suggesting assignment to relation S^c covers any such pair of objects. Thus, relations $\mathbb{S}$ and $\mathbb{S}^c$ implied by induced decision rules are the same independently of whether set $G = \{g_1, g_2, g_3\}$ is a consistent set of criteria or a not necessarily consistent set of criteria. It is worth noting that the induced rules are relatively short and the number

Table 2.5: Minimal decision rules induced by VC-DomLEM algorithm

Decision rule r_T	Supp	$\widehat{\epsilon}_T(r_T)$
if $(\Delta_1(a,b) \leq -284)$, *then* aSb	9	0
if $(\Delta_1(a,b) \leq -166) \wedge (\Delta_3(a,b) \leq 0.02)$, *then* aSb	7	0.067
if $(\Delta_1(a,b) \leq 71) \wedge (\Delta_2(a,b) \geq 0)$, *then* aSb	15	0.067
if $(\Delta_1(a,b) \geq 95)$, *then* aS^cb	12	0.095
if $(\Delta_1(a,b) \geq -189) \wedge (\Delta_2(a,b) \leq -1.7)$, *then* aS^cb	4	0.095

of rules is small w.r.t. the size of the PCT. Moreover, the rules are easy to interpret by the DM.

The final ranking of all objects from set A, obtained using scoring function S^{NF} (2.36), is presented in Table 2.6, where the six reference objects are marked in bold, and for each rank we give respective net flow score, i.e., the value of scoring function S^{NF}.

Table 2.6: Final ranking of all objects from set A, obtained using scoring function S^{NF} (2.36)

Rank	Net flow score	Object(s)
1	39.0	$\mathbf{n_1}$
2	38.0	n_2
3	37.0	n_6
4	30.0	n_3
5	24.0	$\mathbf{n_4}$, n_5
6	17.0	n_8
7	14.0	n_{11}
8	7.0	n_{15}
9	6.0	n_{16}
10	-2.0	n_{21}
11	-7.0	n_{13}
12	-8.0	n_7, $\mathbf{n_{12}}$
13	-14.0	$\mathbf{n_{14}}$
14	-16.0	n_{20}
15	-20.0	n_{22}
16	-27.0	$\mathbf{n_{18}}$
17	-32.0	n_{17}, n_{19}
18	-35.0	n_9, $\mathbf{n_{10}}$

Now, let us analyze the final ranking of reference objects, i.e., $n_1 \succ n_4 \succ n_{12} \succ n_{14} \succ n_{18} \succ n_{10}$, by comparing it to the ranking of these objects given as the preference information, i.e., $n_4 \succ n_1 \succ n_{12} \succ n_{14} \succ n_{10} \succ n_{18}$. The two rankings are quite similar (i.e., the final ranking respects 26 out of 30 preference relations specified in the preference information), with only two “inversions” (n_1 exchanged with n_4, n_{18} exchanged with n_{10}). The value of concordance measure $\tau' = 0.733$, which is a good result concerning that 10 out of 36 pairs of objects given in Table 2.2 are inconsistent. It is worth underlining that the pairs of objects that got “inverted” (i.e., (n_4, n_1) and (n_{10}, n_{18})) were not sufficiently consistent (see the first and the last row of Table 2.3). For example, pair $(n_4, n_1) \in S$ was not included in the lower approximation of relation S since it was dominated by three pairs of objects belonging to relation S^c: (n_{18}, n_{12}), (n_{18}, n_{14}), and (n_{18}, n_{10}). Taking this into account, we can say that the induced set of rules presented in Table 2.5 is a good preference model of the DM.

Chapter 3

Analysis of Desirable Properties of Ranking Methods

When considering properties of a ranking method, it is important to take into account the characteristics of an exploited valued relation R over set A. Concerning definition (2.37), and the discussion from Section 2.8.2, we distinguish the following two cases:

- exploitation of a general valued relation R over A, when $R(a, b) \in \mathbb{R}$ for any $a, b \in A$,
- exploitation of a three-valued relation R over A, when $R(a, b) \in \{0, \frac{1}{2}, 1\}$ for any $a, b \in A$.

In Section 3.1, we define several desirable properties of a ranking method, which exploits a valued relation R over set A of objects in view of constructing a ranking (total or partial preorder) of all objects from this set.

In Section 3.2, we give two supposed *priority orders* of considered desirable properties (which, from our point of view, reflect relative importance of these properties), depending on the characteristic of the exploited valued relation.

Section 3.3 presents the results of examination of the properties of the five well-known ranking methods reviewed in Section 2.8.3. The respective proofs can be found in the Appendix. We conclude this section by indicating a ranking method that enjoys the best properties.

3.1 Desirable Properties of Ranking Methods

In the literature, one can find many properties considered in the context of ranking methods exploiting valued relations. These properties concern the result of application of a ranking method to a general valued relation, or to a valued relation with particular features, e.g., a relation which is crisp and transitive. It should be noticed, however, that these properties concern only the dependencies between the exploited valued relation and obtained final ranking. Thus, they do not concern the dependencies between comprehensive preference relations S, S^c and the final ranking.

The properties of ranking methods can be basically divided into two non-disjoint groups [28, 36]: desirable properties and "characterizing" properties. The former reflect some expectations of a DM w.r.t. the final ranking produced by a ranking method. The latter reflect intrinsic characteristics of a ranking method; given a ranking method, the research concerning "characterizing" properties aims at defining minimal sets of properties that a given ranking method is the only one to satisfy [28, 30, 34, 36, 139]. Since our goal is to obtain the ranking that "best" represents the DM's preferences, we compare different ranking methods w.r.t. desirable properties only. The same way was adopted, e.g., in [173], in the context of exploitation of a crisp relation.

In general, different properties can be considered desirable depending on a particular multicriteria ranking problem (see [36]). In this section, we present the properties that seem to be of interest for most multicriteria ranking problems.

We find it reasonable to consider the following desirable properties of a ranking method to be applied to exploitation of valued relation $\mathcal{R}$ (2.37):

(1) *Neutrality* (property N)

This property was considered, e.g., in [28, 30, 34, 36, 139].

Definition 1 (Neutrality) *A ranking method $\succeq$ is* neutral *if, for any finite set of objects A and any valued relation R over A:*
$\Big(\sigma$ *is a permutation on* $A\Big) \Rightarrow \Big(a \succeq (A,R)\ b \Leftrightarrow \sigma(a) \succeq (A,R^{\sigma})\ \sigma(b)$*, for all* $a,b \in A\Big)$,
where R^{σ} is defined by $R^{\sigma}(\sigma(a),\sigma(b)) = R(a,b)$, for all $a,b \in A$.

Thus, neutrality expresses the fact that a ranking method does not discriminate between objects just because of their labels (or, in other words, their order in the considered set A). It is a classical property in this context (see, e.g., [100, 149]).

(2) *Monotonicity* (property M)

Property of this name was considered, e.g., in [34, 36, 139], although the proposed definitions were semantically slightly different. In this thesis, we adopt the definition of monotonicity property given in [34]. Intuitively, monotonicity says that improving an object cannot decrease its position in the ranking and, moreover, deteriorating an object cannot improve its position in the ranking. In our opinion, the other two definitions, considered in [36, 139], miss at least one aspect of this intuitive formulation. Thus, we propose the following formulation of the monotonicity property.

Definition 2 (Monotonicity) *A ranking method $\succeq$ is* monotonic *if, for any finite set of objects A, any valued relation R over A, and any $a,b \in A$:*

$\Big(a \succeq(A,R)\ b \Rightarrow a \succeq(A,R')\ b\Big)$,
where R' is identical to R except that
$\Big(R'(a,c) > R(a,c)$ *or* $R'(c,a) < R(c,a),$ *for some* $c \in A \setminus \{a\}\Big)$ *or*
$\Big(R'(b,d) < R(b,d)$ *or* $R'(d,b) > R(d,b),$ *for some* $d \in A \setminus \{b\}\Big)$.

Precisely, the definition given in [139] w.r.t. the difference between R' and R concerns only that
$\Big(R'(a,c) > R(a,c),$ for some $c \in A \setminus \{a\}\Big)$ or
$\Big(R'(b,d) < R(b,d),$ for some $d \in A \setminus \{b\}\Big)$.
Moreover, the definition given in [36] lacks the second part of the above disjunction, i.e., the part concerning object b: $\Big(R'(b,d) < R(b,d)$ or $R'(d,b) > R(d,b)$, for some $d \in A \setminus \{b\}\Big)$.

(3) Covering Compatibility (property CC)
This property was considered, e.g., in [36] and [173] (where it was called *respect for the covering relation*).

Definition 3 (Covering Compatibility) *A ranking method* $\succeq$ *is* covering compatible *if, for any finite set of objects A, any valued relation R over A, and any $a, b \in A$:*
$\Big(R(a,b) \geq R(b,a),$ *and for all* $c \in A \setminus \{a,b\}$, $R(a,c) \geq R(b,c)$ *and* $R(c,a) \leq R(c,b)\Big) \Rightarrow a \succeq(A,R)\ b$.

Thus, property CC expresses the intuition that when a "covers" b, b should not be ranked before a. Our interest in this property results also from a very important fact – in case of exploitation of valued relation $\mathcal{R}$ defined by (2.37), property CC of applied ranking method guaranties that the final ranking produced by this method respects dominance relation D over set A. Formally, this can be expressed by:

Corollary 1 *Given any two objects $a, b \in A$, such that aDb, property CC of ranking method $\succeq$ applied to exploitation of relation $\mathcal{R}$* (2.37) *guaranties that* $a \succeq(A,\mathcal{R})\ b$.

The proof of the above corollary can be found in the Appendix.

(4) Independence of Non-Discriminating Objects (property $INDO$)
This property was considered, e.g., in [36] (where it was called *independence of non-discriminating alternatives*) and in [173] (where it was called *independence of non-discriminating elements: weak version*).

Definition 4 (Independence of Non-Discriminating Objects) *A ranking method* $\succeq$ *is* independent of non-discriminating objects *if, for any finite set of*

objects A and any valued relation R over A:
$\Big(R(a,b) = k$ *and* $R(b,a) = k'$, *for all* $a \in A'$ *and all* $b \in A \setminus A'$, *with* $A' \subset A\Big) \Rightarrow$
$\Big(\succeq(A', R/A') = \succeq(A,R)/A'\Big)$.

In the above definition, set $A \setminus A'$ is composed of non-discriminating objects. Thus, independence of non-discriminating objects says that when there is a subset of objects that compare in the same way to all other objects, the ranking of the other objects is not affected by the presence of this subset.

(5) *Independence of Circuits* (property IC)
This property was considered, e.g., in [30, 36]. It reflects the way in which a ranking method deals with circuits (cycles) in the considered valued relation. It uses the concept of *circuit equivalency* of two valued relations.

Definition 5 (Circuit Equivalency) *Let us consider a finite set of objects A. Two valued relations R and R' over A are* circuit-equivalent *if R' is identical to R except that, for some distinct $a, b, c \in A$ and some $\epsilon \in [-1, 1]$:*
$\Big(R'(a,b) = R(a,b) + \epsilon$ *and* $R'(b,a) = R(b,a) + \epsilon\Big)$ *or*
$\Big(R'(a,b) = R(a,b) + \epsilon, R'(b,c) = R(b,c) + \epsilon,$ *and* $R'(c,a) = R(c,a) + \epsilon\Big)$.

Thus, R' and R are circuit-equivalent if they are identical except for a circuit of length 2 or 3 on which a positive or negative value has been added.

Definition 6 (Independence of Circuits) *A ranking method $\succeq$ is* independent of circuits *if, for any finite set of objects A and any two valued relations R and R' over A:*
$\Big(R'$ *and* R *are circuit-equivalent*$\Big) \Rightarrow \Big(\succeq(A, R') = \succeq(A,R)\Big)$.

According to [36], property IC has a straightforward interpretation. When R' and R are circuit-equivalent via a circuit of length 2, independence of circuits implies that the ranking is only influenced by the differences $R(a,b) - R(b,a)$. When R' and R are circuit-equivalent via a circuit of length 3, independence of circuits implies that intransitivities of the kind $R(a,b) > 0$, $R(b,c) > 0$ and $R(c,a) > 0$ can be "wiped out". It is important to notice that property IC makes an explicit use of the "cardinal" properties of values $R(a,b)$, with $a, b \in A$ (except for the particular case in which both R and R' are crisp).

(6) *Ordinality* (property O)
This property was considered, e.g., in [28, 35, 36, 139].

Definition 7 (Ordinality) *A ranking method $\succeq$ is* ordinal *if, for any finite set of objects A, any valued relation R over A, and any strictly increasing and one-to-one*

transformation $\phi : [0,1] \to [0,1]$*:*
$\succeq(A, \phi[R]) = \succeq(A, R)$,
where $\phi[R]$ *is the valued relation over* A *such that* $\phi[R](a,b) = \phi(R(a,b))$*, for all* $a, b \in A$.

Thus, ordinality implies that a ranking method should not make use of the "cardinal" properties of values $R(a, b)$, with $a, b \in A$.

(7) Continuity (property C)
This property was considered, e.g., in [28, 35, 36]. It is meaningful only when the exploited relation R over A is a general valued relation. It uses the concept of *convergence* of a sequence of valued relations to a given valued relation.

Definition 8 (Convergence) *Let us consider a finite set of objects* A *and a sequence of valued relations* $(R^i,\ i = 1, \ldots)$ *that are defined over* A*. We say that this sequence converges to valued relation* R *if for any (arbitrarily small)* $\epsilon > 0$ *there is an integer* k*, such that for all* $j > k$ *and all* $a, b \in A$*, we have* $|R^j(a,b) - R(a,b)| < \epsilon$.

Definition 9 (Continuity) *A ranking method* $\succeq$ *is* continuous *if, for any finite set of objects* A*, any valued relation* R *over* A*, any sequence of valued relations* $(R^i,\ i = 1, \ldots)$ *converging to* R*, and any* $a, b \in A$*:*
$\Big(a \succeq(A, R^i)\ b \text{ for all } R^i \text{ in the sequence}\Big) \Rightarrow \Big(a \succeq(A, R)\ b\Big)$.

Thus, continuity says that "small" changes in an exploited valued relation should not lead to radical changes in the final ranking produced by a ranking method.

(8) Faithfulness (property F)
This property was considered, e.g., in [36] and [173] (where it was called *respect for the data 1.1*).

Definition 10 (Faithfulness) *A ranking method* $\succeq$ *is* faithful *if, for any finite set of objects* A *and any relation* R *over* A*:*
$\Big(R \text{ is a total preorder over } A\Big) \Rightarrow \Big(\succeq(A, R) = R\Big)$.

Thus, faithfulness concerns behavior of a ranking method in a special case when considered relation R is crisp and, moreover, it is a total preorder over A. This property says that a ranking method applied to a total preorder should preserve it.

(9) Data-Preservation (property DP)
This property was considered, e.g., in [36] (where it was called *data-preservation 1*) and in [173] (where it was called *respect for the data 1.3*).

Definition 11 (Data-Preservation) *A ranking method* $\succeq$ *is* data-preserving *if, for any finite set of objects* A *and any relation* R *over* A*:*
$\Big(R$ *is a transitive crisp relation over* $A\Big) \Rightarrow \Big(R \subseteq \succeq(A, R)\Big)$.

Thus, data-preservation says that when it is possible to obtain a partial preorder on the basis of R without deleting information contained in this relation, a ranking method should do so. It is important to note that property DP is not implied by property F and vice versa.

(10) *Greatest-Faithfulness* (property GF)
This property was considered, e.g., in [35, 36].

Definition 12 (Greatest-Faithfulness) *A ranking method* $\succeq$ *is* greatest-faithful *if, for any finite set of objects* A *and any relation* R *over* A*:*
$\Big(R$ *is a crisp relation and* $G(A, R) \neq \emptyset\Big) \Rightarrow \Big(G(A, \succeq(A, R)) \subseteq G(A, R)\Big)$.

Greatest-faithfulness says that if there are some greatest elements of a given set A, then the top-ranked objects should be chosen among them (observe that in case of a ranking method that yields a partial preorder over A, there may be no top-ranked objects, i.e., set $G(A, \succeq(A, R))$ may be empty). Let us note, however, that some authors (see, e.g., [35]) do not find greatest-faithfulness as a particularly intuitive requirement for a ranking method, as this property concerns only the first equivalence class of the obtained ranking (they rather consider this property in the context of choice methods). Moreover, in spite of names, it should be noted that a faithful ranking method is not necessarily greatest-faithful, and vice versa.

(11) *Discrimination* (property D)
This property was not considered in the literature concerning exploitation of a valued relation, although it is relevant in the case of exploitation of a three-valued relation (like relation $\mathcal{R}$ (2.37), when $\mathbb{S}$ and $\mathbb{S}^c$ are crisp).

Definition 13 (Discrimination) *A ranking method* $\succeq$ *is* discriminatory *if for any finite set of objects* A*, there exists a valued relation* R *over* A*, such that the number of ranks in* $\succeq(A, R)$ *is equal to the number of objects in set* A*.*

Thus, discrimination says that for each set of objects A, there exists at least one valued relation R over A, such that the ranking obtained by a considered ranking method is a *total order* over A (i.e., a relation over A that is transitive, antisymmetric, and complete).

3.2 Priority Orders of Desirable Properties

In order to avoid a situation where all considered ranking methods become incomparable (non-dominated), we suppose two *priority orders* of considered desirable properties (which, from our point of view, reflect relative importance of these properties). The first priority order concerns the case of exploitation of a three-valued relation R over A, and the second priority order concerns exploitation of a general valued relation R over A. The supposed priority orders are presented in Table 3.1. These orders are to be used only to resolve situations where two or more ranking methods satisfy the same maximum number of properties.

TABLE 3.1: Priority orders of desirable properties, depending on the characteristic of exploited valued relation R over A

three-valued relation R	**general valued relation** R
neutrality (N)	neutrality (N)
monotonicity (M)	monotonicity (M)
covering compatibility (CC)	covering compatibility (CC)
discrimination (D)	independence of non-discriminating objects ($INDO$)
faithfulness (F)	independence of circuits (IC)
data-preservation (DP)	ordinality (O)
independence of non-discriminating objects ($INDO$)	**continuity** (C)
independence of circuits (IC)	faithfulness (F)
ordinality (O)	data-preservation (DP)
greatest-faithfulness (GF)	greatest-faithfulness (GF)

3.3 Verification of Properties of Ranking Methods and Choice of the Best Method

Before verifying properties of the five ranking methods reviewed in Section 2.8.3, let us make a note concerning reflexivity of an exploited valued relation R over set A. In [139, 173], R was assumed to be irreflexive. In [28, 34], it was assumed that $R(a, b)$ is defined only for pairs of objects $(a, b) \in A \times A$ such that $a \neq b$. Finally, in [35, 36], R was assumed to be reflexive. In this thesis, exploited relation $\mathcal{R}$ (2.37) is reflexive. However, since each of the five ranking methods analyzed here makes use of a scoring function that, for any finite set of objects A and any valued relation R over A, does not take into account values $R(a, a)$, with $a \in A$, previous results concerning properties of the five ranking methods hold.

Table 3.2 and Table 3.3 present properties of the five considered ranking methods, for the case of exploitation of a three-valued relation or a general valued relation, respectively.

In these tables, the properties are ordered according to the respective priority orders defined in Section 3.2. Symbols T and F denote presence and absence of a given property, respectively. Moreover, bold font is used in case when a given pair (Property, RM) was already considered in the literature (where a proof or a counterexample was given), while italics is used otherwise, in which case a proof (showing that the respective property is satisfied for a general valued relation) or a counterexample (showing that the respective property is not satisfied for some three-valued relation) is given in the Appendix. Note that in the row corresponding to property M, some symbols T and F are in italics due to adoption of a particular definition of this property (see Definition 2).

TABLE 3.2: Desirable properties of ranking methods – exploitation of a three-valued relation

Property / **RM**	NFR	$It.NFR$	MiF	$It.MiF$	L/E
N	**T**	**T**	**T**	*T*	*T*
M	*T*	*F*	*T*	*F*	**T**
CC	**T**	**T**	*T*	*T*	*T*
D	*T*	*T*	*F*	*T*	*T*
F	**T**	**T**	**F**	**T**	*T*
DP	**T**	**T**	*T*	*T*	*T*
$INDO$	**T**	**T**	*F*	*F*	*T*
IC	**T**	*F*	**F**	*F*	*F*
O	**F**	*F*	**T**	**T**	*F*
GF	**F**	*F*	**T**	**T**	*T*

TABLE 3.3: Desirable properties of ranking methods – exploitation of a general valued relation

Property / **RM**	NFR	$It.NFR$	MiF	$It.MiF$	L/E
N	**T**	**T**	**T**	*T*	*T*
M	*T*	*F*	*T*	*F*	**T**
CC	**T**	**T**	*T*	*T*	*T*
$INDO$	**T**	**T**	*F*	*F*	*T*
IC	**T**	*F*	**F**	*F*	*F*
O	**F**	*F*	**T**	**T**	*F*
C	**T**	*F*	**T**	**F**	*T*
F	**T**	**T**	**F**	**T**	*T*
DP	**T**	**T**	*T*	*T*	*T*
GF	**F**	*F*	**T**	**T**	*T*

Looking at Tables 3.2 and 3.3, one can observe that the two ranking methods based on iterative application of a choice function, namely $It.NFR$ and $It.MiF$, lack monotonicity property. This observation is concordant with [33]. Moreover, all ranking methods have property CC, which guaranties that when they are applied to exploitation of valued relation $\mathcal{R}$ (2.37), they produce final rankings respecting dominance relation D over set A.

Further analysis of Tables 3.2 and 3.3 leads to the conclusion that, in view of the considered lists of desirable properties, the best ranking method for exploitation of both, a three-valued relation and a general valued relation, is the NFR method. This is because it satisfies most of the properties, i.e., eight out of ten (which is, however, true also for the L/E ranking method), and, moreover:

- in the case of exploitation of a three-valued relation, NFR satisfies the first eight properties (i.e., N, M, CC, D, F, DP, $INDO$, and IC),
- in the case of exploitation of a general valued relation, NFR satisfies the first five properties (i.e., N, M, CC, $INDO$, and IC).

In case of exploitation of a general valued relation, the lack of property O is alleviated by the fact that values $\mathcal{R}(a, b)$, $a, b \in A$, may be interpreted in "cardinal" terms. This is due to the definition of relation $\mathcal{R}$ (2.37), the way of constructing relations $\mathbb{S}$ and $\mathbb{S}^c$, and the semantics of values $\widehat{\epsilon}_T(r_T)$ (2.21) (i.e., relative number of "negative pairs of objects" covered by rule r_T) and $cf(r_T)$ (i.e., relative number of "positive pairs of objects" covered by rule r_T).

It is worth pointing out that the NFR ranking method is attractive not only because of the desirable properties it possesses. It represents an intuitive way of reasoning about relative worth of objects in set A, as it takes into account both positive and negative arguments concerning each object (i.e., strength and weakness of each object), as advocated in [64].

Chapter 4

Similarity-based Classification using Monotonic Rules

4.1 Introduction

In this chapter, we propose an eager learning method for the similarity-based classification introduced in Section 1.1.2. This method employs an adaptation of the Dominance-based Rough Set Approach (DRSA) to case-based reasoning (CBR). In CBR, a conclusion about a new case is inferred by the analysis of conclusions taken for similar cases in the past. In the considered classification context, the new case corresponds to an object z, described by the features from set F, and the conclusion to be made corresponds to the (credibility of) membership of z to a given decision class $X \in \mathcal{D}$, where $\mathcal{D}$ is a finite family of pre-defined decision classes. In the proposed adaptation of DRSA to CBR, we consider the following monotonic relationship:

(mr_2) "the more similar is object y to object x with respect to the considered features, the closer is y to x in terms of the membership to a given decision class X".

It is important to underline that the monotonic relationship (mr_2) is meaningful regardless of the credibility of membership of reference object x to class X. Thus, this monotonic relationship can be seen as a generalization of the monotonic relationship considered in prior adaptations of DRSA to CBR [88, 89, 91], where it was considered that:

(mr_1) "the more similar is object y to object x w.r.t. the considered features, the greater the membership of y to a given decision class X".

Such monotonic relationship is reasonable only if the membership of reference object x to class X takes a maximum value, as explained in Section 1.3.2. Note that in such a case, (mr_2) works in the same way as (mr_1).

The main goal of the proposed methodology is to be possibly "neutral" and "objective" in addressing the questions related to measuring similarity at the two levels mentioned in

Section 1.3.3. At the level of single features, we consider marginal similarities in ordinal terms only. At the level of all features, the marginal similarities are aggregated within decision rules underlying the general monotonic relationship between comprehensive closeness of objects and their marginal similarities.

In this chapter, we extend prior adaptations of DRSA to CBR also by proposing the way o inducing decision rules, and the way of resolving conflicts arising during application of induced rules to a new object (new case).

This chapter is organized as follows. In Section 4.2, we further formalize the setting of the considered similarity-based classification problem introduced in Section 1.1.2; in particular, we define the so-called *similarity space*. We also present assumptions specific for our approach. Section 4.3 particularizes the similarity learning task that we perform with our method. In Section 4.4, we discuss the construction of so-called *similarity tables*. A similarity table stores information that results from comparing each object $y \in U$ with a given reference object $x \in U$. Section 4.5 introduces two *comprehensive closeness relations*. These are binary relations concerning closeness of objects $y, x \in U$ in terms of their membership to a considered decision class $X \in \mathcal{D}$. In Section 4.6, we define rough approximations of the sets of objects being in either kind of comprehensive closeness relation with a reference object $x \in U$. These sets are approximated using dominance cones in the similarity space. Section 4.7 describes induction of monotonic decision rules from the considered rough approximations. Section 4.8 shows how the induced rules can be applied to a new case. In Section 4.9, we present an example illustrating the proposed methodology.

This chapter comprises a revision and extension of the research results originally published in [162].

4.2 Basic Notions and Assumptions

Pairwise fuzzy information base. Given the setting of the similarity-based classification problem introduced in Section 1.1.2, a *pairwise fuzzy information base* $\mathbf{B}$ is the 3-tuple

$$\mathbf{B} =< U, F, \Sigma >, \tag{4.1}$$

where U is a finite set of objects (a case base), $F = \{f_1, f_2, \ldots, f_n\}$ is a finite set of *features*, and $\Sigma = \{\sigma_{f_1}, \sigma_{f_2}, \ldots, \sigma_{f_n}\}$ is a finite set of marginal similarity functions such that $\sigma_{f_i} : U \times U \to [0, 1]$, $f_i \in F$.

Marginal similarity functions. Different marginal similarity functions can be used, depending on the value set V_{f_i} of feature $f_i \in F$. The minimal requirement function σ_{f_i} has to satisfy is that for all $x, y \in U$, $\sigma_{f_i}(y, x) = 1$ if and only if (iff) y and x have the same value of feature f_i. In case of a *numeric feature* f_i, with values on interval or ratio scale, similarity can be defined using a function, e.g.:

- $\sigma_{f_i} = 1 - \frac{|f_i(x)-f_i(y)|}{\max_{v_i \in V_{f_i}} - \min_{v_i \in V_{f_i}}}$,
- $\sigma_{f_i} = \frac{1}{|f_i(x)-f_i(y)|+1}$,
- $\sigma_{f_i} = \frac{1}{(f_i(x)-f_i(y))^2+1}$.

In case of a *nominal feature* f_i, similarity can be defined using a table, like Table 4.1.

TABLE 4.1: Exemplary definition of similarity for a nominal feature

	very low	low	medium	high	very high
very low	1.0	0.8	0.5	0.1	0.0
low	0.8	1.0	0.6	0.3	0.0
medium	0.5	0.6	1.0	0.5	0.3
high	0.1	0.3	0.5	1.0	0.6
very high	0.0	0.0	0.3	0.6	1.0

The marginal similarity functions from set Σ create an n-dimensional *similarity space*. Each pair of objects $(y, x) \in U \times U$ is described in this space by a vector

$$Des_F(y, x) = [\sigma_{f_1}(y, x), \ldots, \sigma_{f_n}(y, x)] \tag{4.2}$$

called *description* of (y, x). This vector represents the available information about similarity between y and x on particular features.

Problem decomposition. In the following, we consider the decision classes belonging to family $\mathcal{D}$ to be mutually independent in the sense of membership function values. Then, we decompose the original multi-class problem π to a set of *single-class subproblems* π_X, where $X \in \mathcal{D}$. Thus, each subproblem concerns a single decision class $X \in \mathcal{D}$ modeled as a fuzzy set in U, characterized by membership function $\mu_X : U \to [0, 1]$.

Aggregation of membership values. For the sake of decreasing calculation time and the number of induced decision rules, one can reduce the set of considered values of μ_X, $X \in \mathcal{D}$. This can be done, e.g., in the following steps:

- define $\mu_X^o(y) \equiv \mu_X(y)$ (i.e., function μ_X^o, identical to function μ_X),
- choose a finite set of k characteristic values $V_X^c = \{v_1, \ldots, v_k\}$, such that $v_1 \in [0, \min_{y \in U} \mu_X^o(y)]$ and $v_k \in [\max_{y \in U} \mu_X^o(y), 1]$, e.g., $V_X^c = \{0.0, 0.1, \ldots, 1.0\}$ or $V_X^c = \{0.0, 0.2, \ldots, 1.0\}$,

- redefine function μ_X as:

$$\mu_X(y) = round(\mu_X^o(y)), \tag{4.3}$$

where function $round : [0,1] \rightarrow V_X^c$ replaces a given original membership value by the nearest characteristic value from set V_X^c.

Regardless of the nature of function μ_X (i.e., whether this function is the original membership function or it results from (4.3)), in the following, we consider set V_{μ_X} defined as:

$$V_{\mu_X} = \{\mu_X(y) : y \in U\}. \tag{4.4}$$

Reference objects. We assume moreover that for each subproblem π_X, there is given a set of so-called *reference objects* $U_X^R \subseteq U$. These are objects to which objects from set U are going to be compared. Typically, the reference objects are indicated by a DM and thus, the set of reference objects is usually relatively small. If such information is not available, it is possible to employ some clustering techniques to choose a suitable set of reference objects, to sample set U, or to treat all the objects from U as the reference ones.

4.3 Similarity Learning

The method proposed in this chapter is designed for the following learning task. Given:

- the pairwise fuzzy information base $\mathbf{B}$,
- the family $\mathcal{D}$ of decision classes, implying subproblems π_X, $X \in \mathcal{D}$,
- the membership functions $\mu_X : U \rightarrow [0,1]$, $X \in \mathcal{D}$,
- the sets of reference objects $U_X^R \subseteq U$, $X \in \mathcal{D}$,

learn, for each subproblem π_X, a similarity-based classification model in terms of a set of decision rules

$$R_X = \bigcup_{x \in U_X^R} R_X(x), \tag{4.5}$$

where $R_X(x)$ is the set of rules describing membership of an object $y \in U$ to class $X \in \mathcal{D}$ based on similarity of y to reference object $x \in U_X^R$.

Notes.

(n1) In the proposed similarity-based classification method, set U is a training set of objects; values $\mu_X(y)$, where $X \in \mathcal{D}$ and $y \in U$, are decision examples.

(n2) The rules from set R_X (4.5) can be used to determine membership $\mu_X(z)$ of a new object (new case) z described by the same features $f_1, \ldots, f_n$.

(n3) The entire method proposed in this chapter is meaningful also when there is only a single decision class $X \in \mathcal{D}$ being a fuzzy set in U. For example, in a medical diagnostic problem, class X may correspond to influenza, with $\mu_X(y)$ expressing the credibility that patient y suffers from this disease. In such a case, a doctor could be asked to select several patients to be the reference objects. The doctor may then indicate, e.g., a "typical" profile x_h of a patient with high credibility of being ill, a "typical" profile x_l of a patient with low credibility of being ill, and a "typical" profile x_m of a patient with credibility about 50%. Then, the set R_X (4.5) would be composed of three subsets of induced decision rules – $R_X(x_h)$, $R_X(x_m)$, and $R_X(x_l)$. These rules would constitute a classification model that could be used to judge what is the credibility that a new patient z suffers from influenza.

(n4) As results from note (n3), and due to treating values $\mu_X(y)$, where $X \in \mathcal{D}$ and $y \in U$, in ordinal terms only, the similarity-based classification method proposed in this chapter may also be applied to ordinal classification problems.

4.4 Similarity Tables

Given a decision class $X \in \mathcal{D}$, and a set of reference objects $U_X^R \subseteq U$, we build for each reference object $x \in U_X^R$ a so-called *similarity table* $ST_X(x)$. This table stores the information that results from comparing each object $y \in U$ with the reference object x.

Formally, a similarity table created for a reference object $x \in U_X^R$ is an $u \times (n+1)$ data table, denoted by $ST_X(x)$, where u is the cardinality of set U, and n is the number of features. First n columns of this table correspond to features from set F. The last, $(n+1)$-th, column relates to the membership function μ_X. Each row of $ST_X(x)$ corresponds to an object $y \in U$ and is composed of the following values:

- $\sigma_{f_1}(y, x)$ in the first column,
- $\sigma_{f_2}(y, x)$ in the second column,
- $\ldots$,
- $\sigma_{f_n}(y, x)$ in the n-th column,
- $\mu_X(y)$ in the last, $(n+1)$-th, column.

4.5 Comprehensive Closeness of Objects

Given a decision class X being a fuzzy set in U, characterized by membership function $\mu_X : U \to [0,1]$, we define two kinds of binary *comprehensive closeness relations* on U:

$$y \succsim^X_{\alpha,\beta} x \Leftrightarrow \mu_X(x) \in [\alpha, \beta] \text{ and } \mu_X(y) \in [\alpha, \beta], \tag{4.6}$$

$$y \precsim^X_{\alpha,\beta} x \Leftrightarrow \mu_X(x) \in [\alpha, \beta] \text{ and } \mu_X(y) \notin (\alpha, \beta), \tag{4.7}$$

where $y, x \in U$ and parameters α, β satisfy:

- in (4.6):
$$0 \le \alpha \le \beta \le 1, \tag{4.8}$$
- in (4.7):
$$0 - \delta \le \alpha \le \beta \le 1 + \delta, \tag{4.9}$$
where $\delta \in \mathbb{R}_+$.

When $y \succsim^X_{\alpha,\beta} x$, then $\alpha \le \mu_X(y) \le \mu_X(x) \le \beta$ or $\alpha \le \mu_X(x) \le \mu_X(y) \le \beta$, i.e., $\mu_X(y)$ is on the left side of $\mu_X(x)$ but not farther than α, or $\mu_X(y)$ is on the right side of $\mu_X(x)$ but not farther than β. On the other hand, when $y \precsim^X_{\alpha,\beta} x$, then $\mu_X(y)$ is on the left side of $\mu_X(x)$ but not closer than α, or $\mu_X(y)$ is on the right side of $\mu_X(x)$ but not closer than β. Thus, α and β play roles of limiting levels of membership to X.

The "special" value $0 - \delta$, where $\delta \in \mathbb{R}_+$, is considered in (4.9) in order to allow $\mu_X(y) \notin (0 - \delta, \beta) \Leftrightarrow \mu_X(y) \notin [0, \beta)$, which boils down to $\mu_X(y) \ge \beta$. Moreover, the "special" value $1 + \delta$, where $\delta \in \mathbb{R}_+$, is considered in (4.9) in order to allow $\mu_X(y) \notin (\alpha, 1 + \delta) \Leftrightarrow \mu_X(y) \notin (\alpha, 1]$, which boils down to $\mu_X(y) \le \alpha$. The inclusion of values $0 - \delta$ and $1 + \delta$, $\delta \in \mathbb{R}_+$, in (4.9) is important for several reasons. The first reason concerns the case when class X is crisp, i.e., $\mu_X(y) \in \{0, 1\}$ for all $y \in U$. Then, one can consider meaningful relations $\precsim^X_{\alpha,\beta}$. Namely, it is meaningful to consider relation $\precsim^X_{0,1+\delta}$, composed of pairs $(y, x) \in U \times U$ such that $\mu_X(y) \le 0$, as well as relation $\precsim^X_{0-\delta,1}$, composed of pairs $(y, x) \in U \times U$ such that $\mu_X(y) \ge 1$. Note that relation $\precsim^X_{0,1}$ is equal to $U \times U$ (and thus, it is not meaningful). Moreover, relation $\precsim^X_{0,0}$ (or $\precsim^X_{1,1}$) is composed of all possible pairs of objects (y, x) such that $y \in U$ and $\mu_X(x) = 0$ (or $\mu_X(x) = 1$, respectively), and thus, it does not convey any useful information. The other reasons are given later in this chapter.

Let us observe that $\succsim^X_{\alpha,\beta}$ is reflexive, symmetric and transitive and thus it is an equivalence relation. Moreover, $\precsim^X_{\alpha,\beta}$ is only transitive.

Given a decision class X and a reference object $x \in U^R_X$, we are interested in characterizing, in terms of similarity-based decision rules, the sets of objects $y \in U$ being in:

- $\succsim^X_{\alpha,\beta}$ relation with x, where $\alpha, \beta \in V_{\mu_X}$,
- $\precsim^X_{\alpha,\beta}$ relation with x, where $\alpha, \beta \in V_{\mu_X} \cup \{0 - \delta\} \cup \{1 + \delta\})$, and $\alpha < \mu_X(x) < \beta$.

To this end, we define sets $S(\succsim^X_{\alpha,\beta}, x)$ and $S(\precsim^X_{\alpha,\beta}, x)$ as:

$$S(\succsim^X_{\alpha,\beta}, x) = \{y \in U : y \succsim^X_{\alpha,\beta} x\}, \tag{4.10}$$

$$S(\precsim^X_{\alpha,\beta}, x) = \{y \in U : y \precsim^X_{\alpha,\beta} x\}, \tag{4.11}$$

where parameters α, β satisfy:

- in (4.10):

$$\begin{cases} 0 \le \alpha \le \mu_X(x) \le \beta \le 1 \\ \alpha, \beta \in V_{\mu_X} \end{cases} \tag{4.12}$$

- in (4.11):

$$\begin{cases} 0 - \delta \le \alpha < \mu_X(x) < \beta \le 1 + \delta \\ \alpha, \beta \in V_{\mu_X} \cup \{0 - \delta\} \cup \{1 + \delta\} \end{cases} \tag{4.13}$$

  where $\delta \in \mathbb{R}_+$.

The constraint $\alpha < \mu_X(x) < \beta$ present in (4.13) prevents from considering not meaningful sets $S(\precsim^X_{\alpha,\beta}, x)$, as will be explained in Section 4.6. From this point of view, it is crucial that when $\mu_X(x) = 0$ (or $\mu_X(x) = 1$), one can take $\alpha = 0 - \delta$ (or $\beta = 1 + \delta$, respectively), where $\delta \in \mathbb{R}_+$.

The sets of objects defined by (4.10) and (4.11) are to be approximated using dominance cones in the similarity space created by functions $\sigma_{f_1}, \ldots, \sigma_{f_n}$, as defined in the next section. Let us note that some of the sets $S(\succsim^X_{\alpha,\beta}, x)$ and $S(\precsim^X_{\alpha,\beta}, x)$ can be disregarded in further considerations since they are equal to U. This is the case for any set $S(\succsim^X_{\alpha,\beta}, x)$ such that $\alpha \le \min_{y \in U} \mu_X(y)$ and $\beta \ge \max_{y \in U} \mu_X(y)$. Moreover, set $S(\precsim^X_{0-\delta,1+\delta}, x)$, with $\delta \in \mathbb{R}_+$, is empty for any $x \in U^R_X$.

4.6 Rough Approximation by Dominance Relation

Now, let us define the dominance relation w.r.t. the similarity to an object $x \in U$, called in short *x-dominance relation*, defined over U, and denoted by D_x. For any $x, y, w \in U$, y is said to *x-dominate* w (denotation yD_xw) if for every $f_i \in F$,

$$\sigma_{f_i}(y, x) \ge \sigma_{f_i}(w, x). \tag{4.14}$$

Thus, object y is said to x-dominate object w iff for every feature $f_i \in F$, y is at least as similar to x as w is.

Given an object $y \in U$, *x-positive* and *x-negative dominance cones* of y in the similarity space are defined as follows:

$$D^+_x(y) = \{w \in U : wD_xy\}, \tag{4.15}$$

$$D^-_x(y) = \{w \in U : yD_xw\}. \tag{4.16}$$

Note that y can be called a *limit object*, because it conditions the membership of w in $D_x^+(y)$ or $D_x^-(y)$.

The above x-dominance cones D_x^+ and D_x^- in the similarity space are used to approximate sets $S(\succsim_{\alpha,\beta}^X, x)$ and $S(\precsim_{\alpha,\beta}^X, x)$, defined, respectively, by (4.10) and (4.11).

In order to induce meaningful certain and possible decision rules concerning similarity to a reference object $x \in U_X^R$, we structure the objects $y \in U$ by calculation of lower and upper approximations of sets $S(\succsim_{\alpha,\beta}^X, x)$ and $S(\precsim_{\alpha,\beta}^X, x)$ as follows.

The *lower approximations* of sets $S(\succsim_{\alpha,\beta}^X, x)$ and $S(\precsim_{\alpha,\beta}^X, x)$ are defined as:

$$\underline{S(\succsim_{\alpha,\beta}^X, x)} = \{y \in U : D_x^+(y) \subseteq S(\succsim_{\alpha,\beta}^X, x)\}, \quad (4.17)$$

$$\underline{S(\precsim_{\alpha,\beta}^X, x)} = \{y \in U : D_x^-(y) \subseteq S(\precsim_{\alpha,\beta}^X, x)\}, \quad (4.18)$$

and the *upper approximations* of sets $S(\succsim_{\alpha,\beta}^X, x)$ and $S(\precsim_{\alpha,\beta}^X, x)$ are defined as:

$$\overline{S(\succsim_{\alpha,\beta}^X, x)} = \{y \in U : D_x^-(y) \cap S(\succsim_{\alpha,\beta}^X, x) \neq \emptyset\}, \quad (4.19)$$

$$\overline{S(\precsim_{\alpha,\beta}^X, x)} = \{y \in U : D_x^+(y) \cap S(\precsim_{\alpha,\beta}^X, x) \neq \emptyset\}. \quad (4.20)$$

With respect to the three basic properties of set approximations defined for rough sets in [135], it follows from definitions (4.10), (4.11), (4.17), (4.18), (4.19), and (4.20), that lower and upper approximations defined above fulfill properties of *rough inclusion* and *monotonicity of the accuracy of approximation*. Moreover, these approximations enjoy also *complementarity* property, as shown in [162].

Using definitions (4.17), (4.18), (4.19), and (4.20), one can define the *boundary* of set $S(\succsim_{\alpha,\beta}^X, x)$ (or set $S(\precsim_{\alpha,\beta}^X, x)$), as the difference between the upper and the lower approximation of this set. It is also possible to perform further DRSA-like analysis by calculating the quality of approximation, reducts, and the core (see, e.g., [83, 84, 86, 160]).

In case of real data sets, the lower approximations defined according to (4.17) and (4.18) may contain relatively small number of objects, due to *inconsistencies w.r.t. x-dominance relation* in the similarity space. Such inconsistency occurs, e.g., when an object $y \in U$ is not less similar to reference object x on every feature $f_i \in F$ than object $w \in U$ is (i.e., $y \in D_x^+(w)$) and, for given values of α and β, there is $w \succsim_{\alpha,\beta}^X x$ and not $y \succsim_{\alpha,\beta}^X x$. Thus, for (highly) inconsistent data sets, it is reasonable to consider some relaxations of definitions (4.17) and (4.18). For this purpose, one could adapt, for instance, the ϵ-VC-DRSA [23, 24].

In prior adaptations of DRSA to CBR [88, 89, 91], the authors considered the following sets of objects:

- *weak upward cut* of decision class X: $X^{\geq t} = \{y \in U : \mu(y) \geq t\}$,
- *weak downward cut* of decision class X: $X^{\leq t} = \{y \in U : \mu(y) \leq t\}$.

Both sets were approximated using x-dominance cones in the similarity space. Set $X^{\geq t}$ was approximated analogously to set $S(\succsim_{\alpha,\beta}^X, x)$. Moreover, set $X^{\leq t}$ was approximated analogously to set $S(\precsim^X_{\alpha,\beta}, x)$. Observe that set $X^{\geq t}$ can be expressed in terms of $S(\succsim^X_{\alpha,\beta}, x)$ by setting: $\alpha = t$, $\beta = 1$, and by taking as reference object x any object $y \in U$ such that $\mu_X(y) \geq t$. Moreover, set $X^{\leq t}$ can be expressed in terms of $S(\precsim^X_{\alpha,\beta}, x)$ by setting: $\alpha = t$, $\beta = 1 + \delta$, and by taking as reference object x any object $y \in U$ such that $\mu_X(y) > t$, where $\delta \in \mathbb{R}_+$. This is the second reason for allowing $\beta > 1$ in (4.9).

In Section 4.5, we pointed out that by requiring in (4.13) that $\alpha < \mu_X(x) < \beta$, we prevent from considering not meaningful sets $S(\precsim^X_{\alpha,\beta}, x)$. Now, let us support this claim by showing an example of what could happen if we would assume $\alpha \leq \mu_X(x) \leq \beta$. First, suppose that $\mu_X(x) = 0.5$. Second, consider set $S(\precsim^X_{0.5,0.7}, x)$, i.e., let $\alpha = \mu_X(x) = 0.5$. Third, consider two objects $y, w \in U$ such that $\mu_X(y) = 0.5$, $\mu_X(w) = 0.6$, and $\sigma_{f_i}(y, x) \geq \sigma_{f_i}(w, x)$, for every $f_i \in F$. In this way, we have a situation where $y \in S(\precsim^X_{0.5,0.7}, x)$, $w \notin S(\precsim^X_{0.5,0.7}, x)$, and $w \in D^-_x(y)$, and thus, $y \notin \underline{S(\precsim^X_{0.5,0.7}, x)}$. Remark that the assumed monotonic relationship is: "the more similar is object y to object x with respect to the considered features, the closer is y to x in terms of the membership to a given decision class X". From this point of view, it is natural that object y, being more similar to reference object x, belongs to class X in degree 0.5, while object w, being less similar to x, belongs to class X in degree 0.6. In other words, values $\mu_X(y)$ and $\mu_X(w)$ obey the assumed monotonic relationship. This shows that set $S(\precsim^X_{0.5,0.7}, x)$ is lacking a meaning, as its rough approximation results in excluding objects obeying the assumed monotonic relationship.

4.7 Induction of Decision Rules

Lower (or upper) approximations of considered sets $S(\succsim^X_{\alpha,\beta}, x)$ and $S(\precsim^X_{\alpha,\beta}, x)$ are the basis for induction of *certain* (or *possible*) *decision rules* belonging to set $R_X(x)$, $x \in U^R_X$. Rules from this set generalize the descriptions of objects from similarity table $ST_X(x)$. Below, we distinguish two types of rules and give their formal syntax:

(1) *at least rules*:
"if $\sigma_{f_{i1}}(y, x) \geq h_{i1}$ and ... and $\sigma_{f_{ip}}(y, x) \geq h_{ip}$, then certainly (or possibly) $y \succsim^X_{\alpha,\beta} x$",

(2) *at most rules*:
"if $\sigma_{f_{i1}}(y, x) \leq h_{i1}$ and ... and $\sigma_{f_{ip}}(y, x) \leq h_{ip}$, then certainly (or possibly) $y \precsim^X_{\alpha,\beta} x$",

where $\{f_{i1}, \ldots, f_{ip}\} \subseteq F$, marginal similarity thresholds $h_{i1}, \ldots, h_{ip} \in [0, 1]$, and limiting levels of membership α, β satisfy $0 \leq \alpha \leq \beta \leq 1$ in case of at least rules, and $0 - \delta \leq \alpha < \beta \leq 1 + \delta$ in case of at most rules, where $\delta \in \mathbb{R}_+$.

For example, a certain rule of type (1) is read as: "if similarity of object y to reference object x w.r.t. feature f_{i1} is at least h_{i1} and ... and similarity of object y to reference object x w.r.t. feature f_{ip} is at least h_{ip}, then certainly object y is in $\succsim^X_{\alpha,\beta}$ comprehensive closeness relation with reference object x.

Remark that according to definitions (4.6) and (4.7), the decision part of the rule of type (1) and (2) can be rewritten, respectively, as:

(1) "then certainly (or possibly) $\mu_X(y) \in [\alpha, \beta]$", i.e., the conclusion is that the membership of object y to decision class X is inside the interval $[\alpha, \beta]$,

(2) "then certainly (or possibly) $\mu_X(y) \notin (\alpha, \beta)$", i.e., the conclusion is that the membership of object y to decision class X is outside the interval (α, β).

Decision rules of type (1) and (2) can be induced using the VC-DomLEM algorithm [26]. On one hand, these rules reveal similarity-based patterns present in the training data. On the other hand, set $R_X = \bigcup_{x \in U_X^R} R_X(x)$ of induced certain/possible rules can be applied to *classify* a new object (new case), i.e., to suggest its degree of membership to considered decision class X, as discussed in the next section.

4.8 Application of Decision Rules

As a result of rule induction, one gets a set of certain (or possible) decision rules R_X for each decision class $X \in \mathcal{D}$. Set R_X is a union of sets of rules $R_X(x)$, where $x \in U_X^R$. The rules from R_X can be applied to a new object (new case) z, described in terms of features $f_1, \ldots, f_n \in F$, to predict its degree of membership to class X. Then, the rules covering the new case may give an ambiguous classification suggestion (intervals of μ_X instead of a crisp value). In order to resolve this ambiguity, we propose to adapt and revise the rule classification scheme described in [21]. In this way, for each new object z classified using rules, one can obtain a precise (crisp) value of membership $\mu_X(z)$. The proposed approach is described below, where Cov_z denotes the set of decision rules covering a given object z and $|\cdot|$ denotes cardinality of a set.

In general, exactly one of the following three situations takes place:

(i) no rule from R_X covers object z (i.e., $Cov_z = \emptyset$),

(ii) exactly one rule $\rho \in R_X(x) \subseteq R_X$, $x \in U_X^R$, covers object z (i.e., $|Cov_z| = 1$),

(iii) several rules from R_X cover object z (i.e., $|Cov_z| > 1$).

Next, we address the above situations one by one, using the following notation:

- $R_X^{\succsim}$ – subset of R_X composed of certain (or possible) rules of type "at least";
- $R_X^{\precsim}$ – subset of R_X composed of certain (or possible) rules of type "at most";
- $V_X^{\succsim}(z)$ – subset of V_{μ_X} containing membership values that are covered by the decision part of at least one rule of type "at least" matching object z, i.e.,

$$V_X^{\succsim}(z) = \{t \in V_{\mu_X} : \text{ there exists a rule } \rho \in R_X^{\succsim} \cap Cov_z \text{ concluding } \mu_X(y) \in [\alpha, \beta] \text{ such that the interval } [\alpha, \beta] \text{ contains } t\}; \quad (4.21)$$

- $R^{\succsim}_X(t)$ – subset of $R^{\succsim}_X$ containing rules whose decision part covers t, i.e.,

$$R^{\succsim}_X(t) = \{\rho \in R^{\succsim}_X : \rho \text{ concludes } \mu_X(y) \in [\alpha, \beta] \text{ and the interval } [\alpha, \beta] \text{ contains } t\}; \tag{4.22}$$

- $V^{\precsim}_X(z)$ – subset of V_{μ_X} containing membership values that are covered by the decision part of at least one rule of type "at most" matching object z, i.e.,

$$V^{\precsim}_X(z) = \{t \in V_{\mu_X} : \text{ there exists a rule } \rho \in R^{\precsim}_X \cap Cov_z \text{ concluding } \mu_X(y) \notin (\alpha, \beta) \text{ such that the interval } (\alpha, \beta) \text{ does not contain } t\}; \tag{4.23}$$

- $R^{\precsim}_X(t)$ – subset of $R^{\precsim}_X$ containing rules whose decision part covers t, i.e.,

$$R^{\precsim}_X(t) = \{\rho \in R^{\precsim}_X : \rho \text{ concludes } \mu_X(y) \notin (\alpha, \beta) \text{ and the interval } (\alpha, \beta) \text{ does not contain } t\}; \tag{4.24}$$

- U^t_X – subset of U containing objects whose membership to class X is equal to t, i.e.,

$$U^t_X = \{y \in U : \mu_X(y) = t\}; \tag{4.25}$$

- U^ρ_X, where $\rho \in R_X$ – subset of U containing objects whose membership to class X is covered by the decision part of rule ρ, i.e.,

$$U^\rho_X = \{y \in U : \mu_X(y) = t \text{ and } t \text{ makes true the statement } \rho \in R^{\succsim}_X(t) \cup R^{\precsim}_X(t)\}. \tag{4.26}$$

Situation (i). As there is no information, there is no reliable suggestion concerning $\mu_X(z)$. However, if a concrete answer is expected, then one can suggest, e.g., that $\mu_X(z)$ is equal to the most frequent value $\mu_X(y)$, where $y \in U$.

Situation (ii). It is necessary to calculate value $Score^\rho_X(t, z)$ for each $t \in V^{\succsim}_X(z)$ (if $\rho \in R^{\succsim}_X$), or for each $t \in V^{\precsim}_X(z)$ (if $\rho \in R^{\precsim}_X$). The value $Score^\rho_X(t, z)$ is defined as:

$$Score^\rho_X(t, z) = \frac{|Cond_\rho \cap U^t_X|^2}{|Cond_\rho||U^t_X|}, \tag{4.27}$$

where $Cond_\rho \subseteq U$ denotes the set of objects verifying the condition part of rule ρ. After calculating $Score^\rho_X(t, z)$ for all considered values of t, the suggested membership of object z to class X is chosen as $\mu_X(z) = \max_t Score^\rho_X(t, z)$. Let us observe that $Score^\rho_X(t, z)$ belongs to the interval $[0, 1]$. It can be interpreted as the degree of certainty of the suggestion that $\mu_X(z)$ is equal to t.

It is important to note that we calculate $Score^\rho_X(t, z)$ only for membership values t belonging to $V^{\succsim}_X(z)$ or $V^{\precsim}_X(z)$. In this way, we not only adapt but also improve the approach proposed in [21], which, when applied to our context directly, would involve calculating $Score^\rho_X(t, z)$ for each $t \in V_{\mu_X}$. This could lead to the situation when value

$\max_t Score^{\rho}_X(t,z)$ could be obtained for some t not belonging to $V^{\succsim}_X(z)$ or $V^{\precsim}_X(z)$. Such situation could occur in case of employing set R_X composed of possible rules, and also in case of employing set R_X composed of probabilistic rules (i.e., rules induced from lower approximations of sets $S(\succsim^X_{\alpha,\beta},x)$ and $S(\precsim^X_{\alpha,\beta},x)$, calculated when using a variable consistency dominance-based rough set approach to CBR).

Situation (iii). To take into account multiple suggestions of the rules covering object z, it is necessary to calculate value $Score_X(t,z)$ for each $t \in V^{\succsim}_X(z) \cup V^{\precsim}_X(z)$. Value $Score_X(t,z)$ is defined as:

$$Score_X(t,z) = Score^+_X(t,z) - Score^-_X(t,z), \tag{4.28}$$

where $Score^+_X(t,z)$ and $Score^-_X(t,z)$ represent the positive and negative part of $Score_X(t,z)$, respectively.

$Score^+_X(t,z)$ takes into account the decision rules from R_X that cover object z and are concordant with the suggestion $\mu_X(z) = t$, i.e., rules $\rho_i \in Cov_z$ $(i = 1,\ldots,k)$ that belong to $R^{\succsim}_X(t)$ or $R^{\precsim}_X(t)$. We define $Score^+_X(t,z)$ as:

$$Score^+_X(t,z) = \frac{|(Cond_{\rho_1} \cap U^t_X) \cup \ldots \cup (Cond_{\rho_k} \cap U^t_X)|^2}{|Cond_{\rho_1} \cup \ldots \cup Cond_{\rho_k}||U^t_X|}, \tag{4.29}$$

where $Cond_{\rho_1},\ldots,Cond_{\rho_k} \subseteq U$ denote the sets of objects verifying condition parts of rules $\rho_1,\ldots,\rho_k$, respectively. Let us observe that $Score^+_X(t,z) \in [0,1]$.

$Score^-_X(t,z)$ takes into account the decision rules from R_X that cover object z but are discordant with the suggestion $\mu_X(z) = t$, i.e., rules $\rho_i \in Cov_z$ $(i = k+1,\ldots,h)$ that belong to $R_X \setminus \left(R^{\succsim}_X(t) \cup R^{\precsim}_X(t)\right)$. We define $Score^-_X(t,z)$ as:

if $Cov_z \cap \Big(R_X \setminus \left(R^{\succsim}_X(t) \cup R^{\precsim}_X(t)\right)\Big) \neq \emptyset$, then

$$Score^-_X(t,z) = \frac{|(Cond_{\rho_{k+1}} \cap U^{\rho_{k+1}}_X) \cup \ldots \cup (Cond_{\rho_h} \cap U^{\rho_h}_X)|^2}{|Cond_{\rho_{k+1}} \cup \ldots \cup Cond_{\rho_h}||U^{\rho_{k+1}}_X \cup \ldots \cup U^{\rho_h}_X|},$$

otherwise,

$$Score^-_X(t,z) = 0, \tag{4.30}$$

where $Cond_{\rho_{k+1}},\ldots,Cond_{\rho_h} \subseteq U$ denote the sets of objects verifying condition parts of rules $\rho_{k+1},\ldots,\rho_h$, respectively. Let us observe that $Score^-_X(t,z) \in [0,1]$.

After calculating $Score_X(t,z)$ for all considered values of t, the suggested membership of object z to class X is chosen as $\mu_X(z) = \max_t Score_X(t,z)$. It can be interpreted as a net balance of the arguments in favor and the arguments against the suggestion "the membership of object z to class X is equal to t".

It is important to note that we calculate $Score_X(t,z)$ only for membership values t belonging to $V^{\succsim}_X(z) \cup V^{\precsim}_X(z)$. In this way, we not only adapt but also improve the approach proposed in [21], which, when applied to our context directly, would involve calculating $Score_X(t,z)$ for each $t \in V_{\mu_X}$. This could lead to the same problem as in the case of situation (ii).

In Section 1.5, we have mentioned that our method for CBR provides additional useful information that other similarity-based classification methods do not provide. Given a set R_X of rules induced for class $X \in \mathcal{D}$, and a new object z whose membership to X needs to be predicted, this information consists of:

- value $Score_X^\rho(t,z)$ representing degree of certainty of the suggestion $\mu_X(z) = t$ (when z is covered by one rule $\rho \in R_X$ only) or value $Score_X(t,z)$ representing net balance of the arguments in favor and the arguments against the suggestion $\mu_X(z) = t$ (when z is covered by more than one rule from R_X),
- set $Cov_z \subseteq R_X$ of decision rules that cover given object z,
- set of objects $y \in U$ that support each rule $\rho \in Cov_z$.

Thus, for a new object z classified using induced decision rules, it is possible to perform full "backtracking", i.e., to see what were the rules covering this object and, in turn, what were the objects supporting these rules. This is an important feature of our approach, comparing to other approaches to CBR.

4.9 Illustrative Example

Let us consider set U composed of five objects described by two features: f_1, with value set $[0, 8]$, and f_2, with value set $[0, 1]$. Moreover, let us consider decision class X being a fuzzy set in U, characterized by membership function $\mu_X : U \to [0, 1]$. The five objects and their membership function values are presented in Table 4.2. Note that according to (4.4), $V_{\mu_X} = \{0.3, 0.4, 0.5, 0.6, 0.7\}$.

TABLE 4.2: Set of objects considered in the illustrative example

$y \in U$	$f_1(y)$	$f_2(y)$	$\mu_X(y)$
y_1	2	0	0.3
y_2	1	0	0.4
y_3	5	1	0.6
y_4	8	0	0.7
x	4	0	0.5

The five objects are also shown in Fig. 4.1, where the number below an object id denotes the value of membership function μ_X for this object, the hatched area corresponds to dominance cone $D_x^+(y_3)$, and the two dotted areas (one for $f_1(y) \leq 2$, and the other for $f_1(y) \geq 6$) correspond to dominance cone $D_x^-(y_1)$.

We assume that object x is a reference object, and that there are given two marginal similarity functions σ_{f_1}, σ_{f_2} defined as:

$$\sigma_{f_i}(y, x) = 1 - \frac{|f_i(y) - f_i(x)|}{f_i^{max} - f_i^{min}},$$

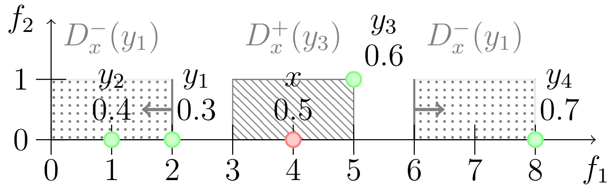

FIGURE 4.1: Set of objects considered in the illustrative example; the number below an object id denotes the value of function μ_X for this object; the hatched area corresponds to dominance cone $D_x^+(y_3)$, and the two dotted areas (one for $f_1(y) \leq 2$, and the other for $f_1(y) \geq 6$) correspond to dominance cone $D_x^-(y_1)$

where $i = 1, 2$, and f_i^{max}, f_i^{min} denote max and min value in the value set of feature f_i, respectively.

Functions σ_{f_1} and σ_{f_2} create a 2-dimensional similarity space. Fig. 4.2 shows pairs of objects $(\cdot, x)$ in this space. The number below a pair of object ids denotes the value of membership function μ_X for the object whose id is the first in the pair.

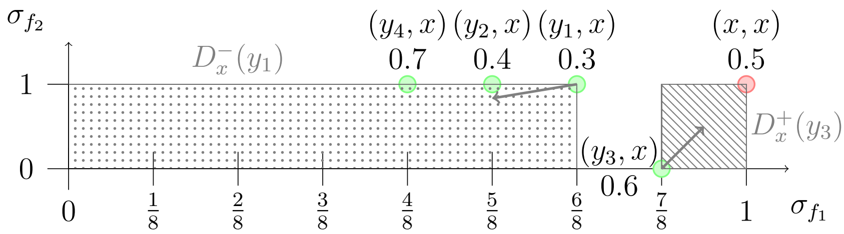

FIGURE 4.2: Pairs of objects $(\cdot, x)$ in the similarity space created by σ_{f_1} and σ_{f_2}; the number below a pair of object ids denotes the value of function μ_X for the object whose id is the first in the pair; the hatched area corresponds to dominance cone $D_x^+(y_3)$, and the dotted area corresponds to dominance cone $D_x^-(y_1)$

First, using (4.15) and (4.16), we calculate the following x-positive and x-negative dominance cones in the similarity space:

$$
\begin{aligned}
D_x^+(y_1) &= \{y_1, x\}, & D_x^-(y_1) &= \{y_1, y_2, y_4\},\\
D_x^+(y_2) &= \{y_1, y_2, x\}, & D_x^-(y_2) &= \{y_2, y_4\},\\
D_x^+(y_3) &= \{y_3, x\}, & D_x^-(y_3) &= \{y_3\},\\
D_x^+(y_4) &= \{y_1, y_2, y_4, x\}, & D_x^-(y_4) &= \{y_4\},\\
D_x^+(x) &= \{x\}, & D_x^-(x) &= \{y_1, y_2, y_3, y_4, x\}.
\end{aligned}
$$

Two of the above calculated cones are shown in Fig. 4.1 (in the feature space) and in Fig. 4.2 (in the similarity space). Note that in Fig. 4.1, the positive cone $D_x^+(y_3)$ corresponds to one area, while the negative cone $D_x^-(y_1)$ corresponds to two areas.

Second, we calculate sets of objects $S(\succsim_{\alpha,\beta}^X, x)$ according to (4.10), for $\alpha \in \{0.3, 0.4, 0.5\}$ and $\beta \in \{0.5, 0.6, 0.7\}$, i.e., for all values of α and β satisfying (4.12). Moreover, we calculate $S(\precsim_{\alpha,\beta}^X, x)$ according to (4.11), for $\alpha \in \{0 - \delta, 0.3, 0.4\}$ and $\beta \in \{0.6, 0.7, 1 + \delta\}$, where $\delta \in \mathbb{R}_+$, i.e., for all values of α and β satisfying (4.13). Calculated sets $S(\succsim^X_{\alpha,\beta}, x)$ and $S(\precsim^X_{\alpha,\beta}, x)$ are shown in Table 4.3.

TABLE 4.3: Considered sets of objects $S(\succsim^X_{\alpha,\beta}, x)$ and $S(\precsim^X_{\alpha,\beta}, x)$; $\delta \in \mathbb{R}_+$

$S(\succsim^X_{\alpha,\beta}, x)$	$\beta = 0.5$	$\beta = 0.6$	$\beta = 0.7$
$\alpha = 0.3$	$\{y_1, y_2, x\}$	$\{y_1, y_2, y_3, x\}$	U
$\alpha = 0.4$	$\{y_2, x\}$	$\{y_2, y_3, x\}$	$\{y_2, y_3, y_4, x\}$
$\alpha = 0.5$	$\{x\}$	$\{y_3, x\}$	$\{y_3, y_4, x\}$

$S(\precsim^X_{\alpha,\beta}, x)$	$\beta = 0.6$	$\beta = 0.7$	$\beta = 1 + \delta$
$\alpha = 0 - \delta$	$\{y_3, y_4\}$	$\{y_4\}$	$\emptyset$
$\alpha = 0.3$	$\{y_1, y_3, y_4\}$	$\{y_1, y_4\}$	$\{y_1\}$
$\alpha = 0.4$	$\{y_1, y_2, y_3, y_4\}$	$\{y_1, y_2, y_4\}$	$\{y_1, y_2\}$

It is important to note that in order to calculate the sets given in Table 4.3, one needs to take into account only the values of membership function μ_X. Moreover, one can discard from the further analysis one of these sets equal to U and one of these sets that is empty.

Third, sets $S(\succsim^X_{\alpha,\beta}, x)$ and $S(\precsim^X_{\alpha,\beta}, x)$ are approximated using the above x-positive and x-negative dominance cones in the similarity space. The lower approximations of these sets, calculated according to (4.17) and (4.18), as well as upper approximations of these sets, calculated according to (4.19) and (4.20), are presented in Table 4.4.

Looking at Table 4.4, one can notice the complementarity of lower and upper approximations w.r.t. U. For example,

$$\underline{S(\succsim^X_{0.3,0.5}, x)} = U \setminus \overline{S(\precsim^X_{0-\delta,0.6}, x)}$$

and

$$\overline{S(\succsim^X_{0.5,0.6}, x)} = U \setminus \underline{S(\precsim^X_{0.4,0.7}, x)}.$$

One can also find inconsistencies w.r.t. the x-dominance relation in the similarity space. Let us give two examples:

- objects $y_2, y_4 \in S(\succsim^X_{0.4,0.7}, x)$ are inconsistent since they are x-dominated by object y_1, and $y_1 \notin S(\succsim^X_{0.4,0.7}, x)$ (because $\mu_X(y_1) = 0.3$),
- object $y_1 \in S(\precsim^X_{0.3,0.6}, x)$ is inconsistent since it x-dominates object y_2 and $y_2 \notin S(\precsim^X_{0.3,0.6}, x)$ (because $\mu_X(y_2) = 0.4$).

The calculated approximations become a basis for induction of decision rules. For example, lower approximation $\underline{S(\succsim^X_{0.3,0.5}, x)} = \{y_1, y_2, x\}$ yields the following certain "at least" rule:

Table 4.4: Lower and upper approximations of considered sets of objects $S(\succsim^X_{\alpha,\beta}, x)$ and $S(\precsim^X_{\alpha,\beta}, x)$; $\delta \in \mathbb{R}_+$

$\underline{S}(\succsim^X_{\alpha,\beta}, x)$	$\beta = 0.5$	$\beta = 0.6$	$\beta = 0.7$
$\alpha = 0.3$	$\{y_1, y_2, x\}$	$\{y_1, y_2, y_3, x\}$	U
$\alpha = 0.4$	$\{x\}$	$\{y_3, x\}$	$\{y_3, x\}$
$\alpha = 0.5$	$\{x\}$	$\{y_3, x\}$	$\{y_3, x\}$
$\overline{S}(\succsim^X_{\alpha,\beta}, x)$	$\beta = 0.5$	$\beta = 0.6$	$\beta = 0.7$
$\alpha = 0.3$	$\{y_1, y_2, x\}$	$\{y_1, y_2, y_3, x\}$	U
$\alpha = 0.4$	$\{y_1, y_2, x\}$	$\{y_1, y_2, y_3, x\}$	U
$\alpha = 0.5$	$\{x\}$	$\{y_3, x\}$	U

$\underline{S}(\precsim^X_{\alpha,\beta}, x)$	$\beta = 0.6$	$\beta = 0.7$	$\beta = 1 + \delta$
$\alpha = 0 - \delta$	$\{y_3, y_4\}$	$\{y_4\}$	$\emptyset$
$\alpha = 0.3$	$\{y_3, y_4\}$	$\{y_4\}$	$\emptyset$
$\alpha = 0.4$	$\{y_1, y_2, y_3, y_4\}$	$\{y_1, y_2, y_4\}$	$\emptyset$
$\overline{S}(\precsim^X_{\alpha,\beta}, x)$	$\beta = 0.6$	$\beta = 0.7$	$\beta = 1 + \delta$
$\alpha = 0 - \delta$	$\{y_3, y_4\}$	$\{y_4\}$	$\emptyset$
$\alpha = 0.3$	$\{y_1, y_2, y_3, y_4\}$	$\{y_1, y_2, y_4\}$	$\{y_1, y_2, y_4\}$
$\alpha = 0.4$	$\{y_1, y_2, y_3, y_4\}$	$\{y_1, y_2, y_4\}$	$\{y_1, y_2, y_4\}$

"if $\sigma_{f_1}(y, x) \geq \frac{5}{8}$ and $\sigma_{f_2}(y, x) \geq 1$, then certainly $\mu_X(y) \in [0.3, 0.5]$",

covering objects y_1, y_2, x, where marginal similarity threshold of $\frac{5}{8}$ results from $1 - \frac{|f_1(y_2) - f_1(x)|}{8-0} = 1 - \frac{|1-4|}{8} = 1 - \frac{3}{8}$. Moreover, upper approximation $\overline{S}(\precsim^X_{0.4,0.7}, x) = \{y_1, y_2, y_4\}$ yields the following possible at "most rule":

"if $\sigma_{f_1}(y, x) \leq \frac{6}{8}$, then possibly $\mu_X(y) \notin (0.4, 0.7)$",

covering objects y_1, y_2, y_4, where marginal similarity threshold of $\frac{6}{8}$ results from $1 - \frac{|f_1(y_1) - f_1(x)|}{8-0} = 1 - \frac{|2-4|}{8} = 1 - \frac{2}{8}$.

Table 4.5 presents all certain and possible minimal decision rules induced by VC-DomLEM algorithm from the non-empty lower and upper approximations shown in Table 4.4 (other than approximations of set $S(\succsim^X_{0.3,0.7}, x) = U$).

Example of application of induced decision rules. Let us now consider application of induced certain decision rules to a new object z described in terms of the same features f_1 and f_2. This application aims at predicting membership degree $\mu_X(z)$.

Let $f_1(z) = 5.5$ and $f_2(z) = 0.5$. These values yield $\sigma_{f_1}(z, x) = 6.5/8$ and $\sigma_{f_2}(z, x) = 1/2$. Looking at certain rules from Table 4.5, we can observe that object z is covered by

Table 4.5: Certain and possible minimal decision rules induced by VC-DomLEM algorithm; for each rule, column 'Supp.' presents ids of objects supporting this rule, while column '¬ Supp.' presents ids of objects covered by this rule but not supporting it

Id	Decision rule					Supp.	¬ Supp.
ρ_1	if $\sigma_{f_1}(y,x) \geq \frac{5}{8}$ and $\sigma_{f_2}(y,x) \geq 1$,	then certainly	$\mu_X(y)$	$\in$	$[0.3, 0.5]$	$\{y_1, y_2, x\}$	
ρ_2	if $\sigma_{f_1}(y,x) \geq \frac{5}{8}$,	then certainly	$\mu_X(y)$	$\in$	$[0.3, 0.6]$	$\{y_1, y_2, y_3, x\}$	
ρ_3	if $\sigma_{f_1}(y,x) \geq 1$,	then certainly	$\mu_X(y)$	$\in$	$[0.5, 0.5]$	$\{x\}$	
ρ_4	if $\sigma_{f_1}(y,x) \geq \frac{7}{8}$,	then certainly	$\mu_X(y)$	$\in$	$[0.5, 0.6]$	$\{y_3, x\}$	
ρ_5	if $\sigma_{f_1}(y,x) \geq \frac{5}{8}$ and $\sigma_{f_2}(y,x) \geq 1$,	then possibly	$\mu_X(y)$	$\in$	$[0.3, 0.5]$	$\{y_1, y_2, x\}$	
ρ_6	if $\sigma_{f_1}(y,x) \geq \frac{5}{8}$,	then possibly	$\mu_X(y)$	$\in$	$[0.3, 0.6]$	$\{y_1, y_2, y_3, x\}$	
ρ_7	if $\sigma_{f_1}(y,x) \geq \frac{5}{8}$ and $\sigma_{f_2}(y,x) \geq 1$,	then possibly	$\mu_X(y)$	$\in$	$[0.4, 0.5]$	$\{y_2, x\}$	$\{y_1\}$
ρ_8	if $\sigma_{f_1}(y,x) \geq \frac{5}{8}$,	then possibly	$\mu_X(y)$	$\in$	$[0.4, 0.6]$	$\{y_2, y_3, x\}$	$\{y_1\}$
ρ_9	if $\sigma_{f_1}(y,x) \geq \frac{4}{8}$,	then possibly	$\mu_X(y)$	$\in$	$[0.4, 0.7]$	$\{y_2, y_3, y_4, x\}$	$\{y_1\}$
ρ_{10}	if $\sigma_{f_1}(y,x) \geq 1$,	then possibly	$\mu_X(y)$	$\in$	$[0.5, 0.5]$	$\{x\}$	
ρ_{11}	if $\sigma_{f_1}(y,x) \geq \frac{7}{8}$,	then possibly	$\mu_X(y)$	$\in$	$[0.5, 0.6]$	$\{y_3, x\}$	
ρ_{12}	if $\sigma_{f_1}(y,x) \geq \frac{4}{8}$,	then possibly	$\mu_X(y)$	$\in$	$[0.5, 0.7]$	$\{y_3, y_4, x\}$	$\{y_1, y_2\}$
ρ_{13}	if $\sigma_{f_2}(y,x) \leq 0$,	then certainly	$\mu_X(y)$	$\geq$	0.6	$\{y_3\}$	
ρ_{14}	if $\sigma_{f_1}(y,x) \leq \frac{4}{8}$,	then certainly	$\mu_X(y)$	$\geq$	0.7	$\{y_4\}$	
ρ_{15}	if $\sigma_{f_1}(y,x) \leq \frac{7}{8}$,	then certainly	$\mu_X(y)$	$\notin$	$(0.4, 0.6)$	$\{y_1, y_2, y_3, y_4\}$	
ρ_{16}	if $\sigma_{f_1}(y,x) \leq \frac{6}{8}$,	then certainly	$\mu_X(y)$	$\notin$	$(0.4, 0.7)$	$\{y_1, y_2, y_4\}$	
ρ_{17}	if $\sigma_{f_2}(y,x) \leq 0$,	then possibly	$\mu_X(y)$	$\geq$	0.6	$\{y_3\}$	
ρ_{18}	if $\sigma_{f_1}(y,x) \leq \frac{4}{8}$,	then possibly	$\mu_X(y)$	$\geq$	0.7	$\{y_4\}$	
ρ_{19}	if $\sigma_{f_1}(y,x) \leq \frac{7}{8}$,	then possibly	$\mu_X(y)$	$\notin$	$(0.3, 0.6)$	$\{y_1, y_3, y_4\}$	$\{y_2\}$
ρ_{20}	if $\sigma_{f_1}(y,x) \leq \frac{6}{8}$,	then possibly	$\mu_X(y)$	$\notin$	$(0.3, 0.7)$	$\{y_1, y_4\}$	$\{y_2\}$
ρ_{21}	if $\sigma_{f_1}(y,x) \leq \frac{6}{8}$,	then possibly	$\mu_X(y)$	$\leq$	0.3	$\{y_1\}$	$\{y_2, y_4\}$
ρ_{22}	if $\sigma_{f_1}(y,x) \leq \frac{7}{8}$,	then possibly	$\mu_X(y)$	$\notin$	$(0.4, 0.6)$	$\{y_1, y_2, y_3, y_4\}$	
ρ_{23}	if $\sigma_{f_1}(y,x) \leq \frac{6}{8}$,	then possibly	$\mu_X(y)$	$\notin$	$(0.4, 0.7)$	$\{y_1, y_2, y_4\}$	
ρ_{24}	if $\sigma_{f_1}(y,x) \leq \frac{6}{8}$,	then possibly	$\mu_X(y)$	$\leq$	0.4	$\{y_1, y_2\}$	$\{y_4\}$

rules σ_2 and σ_{15}. The first rule, of type "at least" suggests that $\mu_X(z) \in [0.3, 0.6]$. The second rule, of type "at most", suggests that $\mu_X(z) \notin (0.4, 0.6)$, which can be also expressed as $\mu_X(z) \leq 0.4$ or $\mu_X(z) \geq 0.6$. Thus, according to (4.21), $V_X^{\succsim}(z) = \{0.3, 0.4, 0.5, 0.6\}$, whereas according to (4.23), $V_X^{\precsim}(z) = \{0.3, 0.4, 0.6, 0.7\}$. As there is more than one rule covering object z, we apply (4.29) and (4.30), and then (4.28), to calculate, respectively, $Score_X^{+}(t,z)$, $Score_X^{-}(t,z)$, and $Score_X(t,z)$, for each membership degree $t \in V_X^{\succsim}(z) \cup V_X^{\precsim}(z) = \{0.3, 0.4, 0.5, 0.6, 0.7\}$. The result of these calculations is shown in Table 4.6. Using the scores from Table 4.6, one can conclude that $\mu_X(z)$ is equal to 0.3, 0.4, or 0.6.

TABLE 4.6: Scores of object z considered in the example of application of induced decision rules

$t \in V_{\mu_X}$	$Score^{+}_{X}(t,z)$	$Score^{-}_{X}(t,z)$	$Score_X(t,z)$
0.3	$\frac{1}{5}$	0	$\mathbf{\frac{1}{5}}$
0.4	$\frac{1}{5}$	0	$\mathbf{\frac{1}{5}}$
0.5	$\frac{1}{4}$	1	$-\frac{3}{4}$
0.6	$\frac{1}{5}$	0	$\mathbf{\frac{1}{5}}$
0.7	$\frac{1}{4}$	1	$-\frac{3}{4}$

Chapter 5

Computational Evaluation of the Proposed Methodology

In this chapter, we present the setup and results of a computational experiment performed to compare six different versions of our method for multicriteria ranking, proposed in Chapter 2, with another state-of-the-art method from the field of Preference Learning – SVM^{rank} (see Section 1.2.2). This method uses the same type of input preference information as our method, i.e., pairwise comparisons of objects, and yields the same information at the output, i.e., a total preorder of objects.

For our method, we used the implementation available in jRank[1] software [165, 166]. For SVM^{rank}, we used fast implementation[2] of the ranking SVM described in [113].

We have also considered the possibility of comparing our method with the UTA method, accepting the same type of input preference information, and producing a total preorder at the output, but we abandoned this idea due to:

- lack of convenient implementations of the UTA method (i.e., software libraries rather that GUI-based programs),

- size of data sets considered in the experiment, implying large number (around 100.000) of pairwise comparisons (and thus large number of constraints of LP problems),

- high number of inconsistent pairwise comparisons for some of the data sets considered in the experiment.

We describe the setup of our experiment in Section 5.1. In Section 5.2, we present and discuss the results of this experiment.

[1] `http://www.cs.put.poznan.pl/mszelag/Software/jRank/jRank.html`
[2] `http://www.cs.cornell.edu/People/tj/svm_light/svm_rank.html`

5.1 Setup of Computational Experiment for Multicriteria Ranking

In the experiment, we compared SVM^{rank} and all six versions of the proposed rule-based approach to multicriteria ranking, i.e.: VC-DRSA$^{rank}_{c\,0|1}$, VC-DRSA$^{rank}_{c\,0\text{-}1_{cr}}$, VC-DRSA$^{rank}_{c\,0\text{-}1_{\times}}$, VC-DRSA$^{rank}_{nc\,0|1}$, VC-DRSA$^{rank}_{nc\,0\text{-}1_{cr}}$, and VC-DRSA$^{rank}_{nc\,0\text{-}1_{\times}}$ (see Section 2.1). All experiments were carried out on a computer cluster in order to speed up computations and to allow a sufficient number of repetitions for observing statistically significant differences between algorithms.

The experiment was performed using 14 publicly available data sets concerning ordinal classification problems. Some of these data sets were taken "as is", whereas others were shrinked before calculations by a random sampling of objects (preserving distribution of decision classes) so as to obtain maximum around 350 objects, and thus, decrease computation time. This number of objects was chosen to limit the number of pairs of objects in a PCT up to around 100.000. The considered data sets are presented in Table 5.1, where #Obj. denotes the number of objects, symbol '*' in column #Obj. indicates that given number of objects results from shrinking original data set, #Crit. denotes the number of criteria, #Class. denotes the number of decision classes, and $\overline{\gamma}(S, S^c)$ denotes average value of the quality of approximation $\gamma(S, S^c)$ defined by (2.16). The values in the last column are averages over 30 PCTs (as explained below), calculated for $\theta_S = \theta_{S^c} = 0$ and for the case when set G is considered to be a not necessarily consistent set of criteria. These average values define the order of data sets – from the (on average) most consistent data set, to the less consistent one. For 3 out of 14 data sets we were not able to obtain any results for SVM^{rank} – these data sets are marked in Table 5.1 by (-). The reason for this situation is the following: (i) `car` and `breast-c` data sets contain purely ordinal criteria that are not accepted by SVM^{rank}; (ii) the results for `windsor` data set could not be obtained as the calculations (for all tested parameter settings, even if increasing the tolerance for termination criterion) went on for many days without progress, and finally had to be terminated.

Data sets: `ERA` (employee rejection/acceptance), `ESL` (employee selection), `LEV` (lecturers evaluation), and `SWD` (social workers decisions) come from [14, 15]. Data sets: `denbosch` and `windsor`, concerning housing prices, were taken from [45] and [119], respectively. Data sets: `car`, `housing`, `cpu`, `breast-w`, `balance-scale`, and `breast-c` were taken from the UCI repository[3] (`breast-c` refers to the Breast Cancer data set while `breast-w` refers to the Breast Cancer Wisconsin (Original) data set). Data sets `bank-g` (Bank of Greece) and `fame` (Financial Analysis Made Easy) come from other public repositories. For `windsor`, `cpu`, and `housing` data sets, original continuous decision attribute was discretized into four levels, containing equal number of objects. For `car` data set, only cars having less

[3] `http://archive.ics.uci.edu/ml`

than 5 doors have been taken into account. For `breast-c` data set, original groups of values of `tumor-size` criterion were aggregated (+) in the following way: {0-4, 5-9 + 55-59, 10-14 + 50-54, 15-19 + 45-49, 20-24 + 40-44, 25-29 + 35-39, 30-34}. In case of data sets for which preference directions of criteria (i.e., information about which criteria are of gain-type, and which are of cost-type) were not given directly, we obtained these directions using domain knowledge about respective problems.

TABLE 5.1: Characteristics of data sets and average values of measure $\gamma(S, S^c)$ for $\theta_S = \theta_{S^c} = 0$ and not necessarily consistent set of criteria

Id	Data set	#Obj.	#Crit.	#Class.	$\overline{\gamma}(S, S^c)$
1	(-) `car`	324*	6	4	0.9732
2	`housing`	253*	13	4	0.9703
3	`cpu`	209	6	4	0.7545
4	`denbosch`	119	8	2	0.7291
5	`bank-g`	353*	16	2	0.7210
6	`fame`	332*	10	5	0.6454
7	(-) `windsor`	273*	10	4	0.6084
8	`breast-w`	350*	9	2	0.6048
9	`balance-scale`	313*	4	3	0.4886
10	`ESL`	244*	4	9	0.3360
11	(-) `breast-c`	286	7	2	0.2494
12	`SWD`	334*	10	4	0.1844
13	`LEV`	334*	4	5	0.1219
14	`ERA`	334*	4	9	0.0087

For each data set, we performed 3 times (each time with a different random seed) a ten-fold cross-validation, which amounts to 30 runs (folds) for each data set. In each cross-validation, we divided the data set in ten parts. Then, in each fold, we took one of these parts as set A^R of reference objects (training set), and the rest of the objects as set A of objects to be ranked (test set). In the following, we describe the calculations performed in a single cross-validation fold by SVM^{rank} and by each of the six versions of VC-DRSA^{rank}.

SVM^{rank}. In a given fold, SVM^{rank} learned a model using training set A^R and value of regularization parameter c which reflects the trade-off between training error and margin (default value was 0.01). We considered $c \in \{0.001, 0.01, 0.1, 1, 10\}$ and we employed (default) loss function minimizing the total number of swapped pairs of objects. The other parameters were set to default values. Remark that the model learned by SVM^{rank} is implied by the pairwise preference constraints concerning objects from different classes only. Precisely, if object $a \in A^R$ is classified to a better class than object $b \in A^R$, then the method incorporates the pairwise preference constraint that a should be ranked higher than b.

After learning, the obtained model was applied on test set A, producing a total preorder of objects from A defined by predicted ranking scores. The quality of this ranking was assessed in the same way as described below for VC-DRSA^{rank}.

VC-DRSA^{rank}. In a given fold, the ordinal classification of objects from set A^R was used to calculate relations S and S^c over A^R in the way described in Section 2.3. In the versions VC-DRSA$_c^{rank}$, if ordinal classification implied aS^cb but object a dominated object b, then we "corrected" the preference information by assuming aSb instead of aS^cb. Lower approximations of relations S and S^c were calculated using (2.14) and (2.15), respectively, with thresholds $\theta_S, \theta_{S^c} \in \{0, 0.01, 0.05, 0.1, 0.15\}$, $\theta_S = \theta_{S^c}$. We also calculated coefficient $\gamma(S, S^c)$ for $\theta_S = \theta_{S^c} = 0$ and relations S, S^c implied directly (i.e., without any "correction") by the ordinal classification on A^R. This value contributed to the average given in the last column of Table 5.1. Lower approximation of S and S^c were the basis for induction of probabilistic decision rules using VC-DomLEM algorithm [25, 26]. The preference structure (crisp or valued) on set A, resulting from application of induced decision rules on A, was exploited using the two-step exploitation procedure (iv) introduced in Section 2.8.1. In the second step of this procedure, we used the NFR ranking method to exploit valued relation $\mathcal{R}$ (2.37) and get a total preorder over set A.

The final total preorder over A, equal to $NFR(A, \mathcal{R})$, and denoted by $\succeq_A^f$, was compared with the initial total preorder over A, denoted by $\succeq_A^i$, resulting from available ordinal classification of all objects from this set. We used two concordance measures for this comparison. First, we calculated $\tau(\succeq_A^i, \succeq_A^f)$, i.e., the Kendall rank correlation coefficient given by (2.58). Second, we calculated $\tau^{\neg I}(\succeq_A^i, \succeq_A^f)$, which reflects concordance of both total preorders but does not take into account the pairs of objects $(a, b) \in A \times A$ such that a and b are considered indifferent according to the input preference information (they belong to the same decision class). Thus, there is no error if objects from a given decision class do not have the same rank in the final ranking on A, but instead are ranked one after another, without "interference" of objects from other classes. Given two total preorders $\succeq_{A^R}$ and $\succeq_A$, such that $A^R \subseteq A$, $|A^R| \geq 2$, value $\tau^{\neg I}(\succeq_{A^R}, \succeq_A)$ is calculated similarly to $\tau(\succeq_{A^R}, \succeq_A)$ defined by (2.58), with the following two differences:

- the sum of $err(a, b)$ in the nominator is calculated over pairs of objects $(a, b) \in P_{\succeq_{A^R}} \cup P^{-1}_{\succeq_{A^R}}$,
- the denominator is equal to $|P_{\succeq_{A^R}} \cup P^{-1}_{\succeq_{A^R}}|$,

where relations $P_{\succeq_{A^R}}$ and $P^{-1}_{\succeq_{A^R}}$ are defined by (2.54) and (2.55), respectively. Thus:

$$\tau^{\neg I}(\succeq_{A^R}, \succeq_A) = 1 - 2 \frac{\sum_{(a,b) \in P_{\succeq_{A^R}} \cup P^{-1}_{\succeq_{A^R}}} err(a, b)}{|P_{\succeq_{A^R}} \cup P^{-1}_{\succeq_{A^R}}|}, \tag{5.1}$$

where $err(a, b)$ is defined by (2.59). Obviously, values of coefficient $\tau^{\neg I}$ also belong to the interval $[-1, 1]$.

5.2 Analysis of Results

Tables 5.2 and 5.3 present performance of the six versions of VC-DRSA^{rank} on all 14 data sets, as well as performance of SVM^{rank} on the 11 data sets to which we were able to apply this method. Each cell of Table 5.2 (Table 5.3), corresponding to data set *dat* and method *meth*, presents – as the first number – the largest average[4] value of τ (2.58) ($\tau^{\neg I}$ (5.1), respectively) in the set of 5 average values of τ ($\tau^{\neg I}$, respectively) calculated for $\theta_S = \theta_{S^c} \in \{0, 0.01, 0.05, 0.1, 0.15\}$ – when *meth* is a version of VC-DRSA^{rank} – or for $c \in \{0.001, 0.01, 0.1, 1, 10\}$ – when $meth =$ SVM^{rank}.

For each average shown in Table 5.2 and Table 5.3, we also present (in parentheses) its rank in the corresponding table row (the lower the rank, the better), and we give (after symbol $\pm$) respective standard deviation (reflecting 30 folds). The averages and standard deviations are presented using four decimal digits. Moreover, for each data set *dat*, we use bold font to indicate the best average(s) avg^*_{dat} in the respective table row, while italics are used to indicate the averages that are smaller than $avg^*_{dat} - dev^*_{dat}$, where dev^*_{dat} denotes the (smallest) standard deviation associated with avg^*_{dat}. The last two rows of Tables 5.2 and 5.3 show average ranks of the compared methods (taking into account all 14 data sets, or only 11 data sets common to all methods); best ranks are given in bold, and for each rank we give its position w.r.t. the other ranks. Obviously, one is interested in obtaining the lowest average ranks.

Table 5.4 and Table 5.5 present the best parameter values (i.e., values yielding the largest averages of τ and $\tau^{\neg I}$, shown in Table 5.2 and Table 5.3, respectively) for the six versions of VC-DRSA^{rank} and for SVM^{rank}. Remind that for VC-DRSA^{rank}, we considered $\theta_S = \theta_{S^c} \in \{0, 0.01, 0.05, 0.1, 0.15\}$, while for SVM^{rank}, we took into account $c \in \{0.001, 0.01, 0.1, 1, 10\}$.

Table 5.6 shows averages over 30 test sets of: (i) percentage of pairs of objects $(a, b) \in A \times A$ assigned by the ordinal classification on test set A to relation $I^{\neq}_{\succeq^i_A} = I_{\succeq^i_A} \setminus \{(a, a) : a \in A\}$, (ii) percentage of pairs of objects $(a, b) \in A \times A$ assigned by the ordinal classification on test set A to sum of relations $P_{\succeq^i_A} \cup P^{-1}_{\succeq^i_A}$, (iii) percentage of pairs of objects $(a, b) \in A \times A$ such that $(a, b) \in I^{\neq}_{\succeq^i_A}$ and $(a, b) \in I^{\neq}_{\succeq^f_A} = I_{\succeq^f_A} \setminus \{(a, a) : a \in A\}$, (iv) percentage of pairs of objects $(a, b) \in A \times A$ such that (a, b) belongs simultaneously to $P_{\succeq^i_A}$ and $P_{\succeq^f_A}$, or to $P^{-1}_{\succeq^i_A}$ and $P^{-1}_{\succeq^f_A}$. Bold font is used to indicate situations where, on average, more than 10% of pairs of objects from relation $I^{\neq}_{\succeq^i_A}$ was assigned by the final ranking on A to relation $I^{\neq}_{\succeq^f_A}$.

Considering the results reported in Tables 5.2, 5.3, 5.4, 5.5, 5.6, and using the values in the last column of Table 5.1, we draw out the following conclusions.

- The experiment shows that the proposed approach to preference learning in multicriteria ranking is *competitive* to SVM^{rank}. Taking into account also wider applicability

[4] averaging was done over 30 folds

Table 5.2: Performance of the six versions of VC-DRSA^{rank} (in short V^{rank}) and SVM^{rank} in terms of **measure** τ (largest average of $\tau^{(rank)}$ $\pm$ standard deviation)

Data set	$V^{rank}_{c0\|1}$	$V^{rank}_{c0\text{-}1_{cr}}$	$V^{rank}_{c0\text{-}1_{\times}}$	$V^{rank}_{nc0\|1}$	$V^{rank}_{nc0\text{-}1_{cr}}$	$V^{rank}_{nc0\text{-}1_{\times}}$	SVM^{rank}
(-) `car`	0.6750$^{(2)}$ ±0.1133	0.6255$^{(3.5)}$ ±0.0964	*0.4485*$^{(6)}$ ±0.0189	**0.6956**$^{(1)}$ ±0.1224	0.6255$^{(3.5)}$ ±0.0966	*0.4950*$^{(5)}$ ±0.0485	–
`housing`	**0.6727**$^{(2.5)}$ ±0.0433	**0.6727**$^{(2.5)}$ ±0.0433	0.6562$^{(6)}$ ±0.0560	**0.6727**$^{(2.5)}$ ±0.0433	**0.6727**$^{(2.5)}$ ±0.0433	0.6607$^{(5)}$ ±0.0567	0.6534$^{(7)}$ ±0.0523
`cpu`	**0.7873**$^{(1.5)}$ ±0.0155	0.7786$^{(6)}$ ±0.0147	0.7735$^{(7)}$ ±0.0154	**0.7873**$^{(1.5)}$ ±0.0155	0.7788$^{(5)}$ ±0.0147	0.7796$^{(4)}$ ±0.0114	0.7858$^{(3)}$ ±0.0061
`denbosch`	**0.5125**$^{(1.5)}$ ±0.1102	0.4774$^{(4)}$ ±0.0937	0.4570$^{(7)}$ ±0.0861	**0.5125**$^{(1.5)}$ ±0.1100	0.4792$^{(3)}$ ±0.0915	0.4754$^{(5)}$ ±0.0925	0.4747$^{(6)}$ ±0.0843
`bank-g`	**0.2696**$^{(1)}$ ±0.0344	0.2543$^{(4)}$ ±0.0286	0.2500$^{(6)}$ ±0.0293	0.2691$^{(2)}$ ±0.0342	0.2494$^{(7)}$ ±0.0318	0.2505$^{(5)}$ ±0.0289	0.2688$^{(3)}$ ±0.0191
`fame`	0.7097$^{(4)}$ ±0.0306	0.7070$^{(6)}$ ±0.0315	0.7030$^{(7)}$ ±0.0286	0.7097$^{(3)}$ ±0.0307	0.7072$^{(5)}$ ±0.0312	**0.7132**$^{(1)}$ ±0.0270	0.7131$^{(2)}$ ±0.0317
(-) `windsor`	0.5944$^{(2)}$ ±0.0607	0.5890$^{(4)}$ ±0.0527	0.5806$^{(6)}$ ±0.0544	**0.5952**$^{(1)}$ ±0.0586	0.5899$^{(3)}$ ±0.0545	0.5841$^{(5)}$ ±0.0577	–
`breast-w`	**0.5387**$^{(1)}$ ±0.0458	*0.4839*$^{(4)}$ ±0.0097	*0.4696*$^{(6)}$ ±0.0062	0.5385$^{(2)}$ ±0.0458	0.5078$^{(3)}$ ±0.0219	*0.4819*$^{(5)}$ ±0.0178	*0.4678*$^{(7)}$ ±0.0078
`balance-scale`	**0.5787**$^{(1.5)}$ ±0.0210	0.5772$^{(3.5)}$ ±0.0224	0.5659$^{(7)}$ ±0.0206	**0.5787**$^{(1.5)}$ ±0.0210	0.5772$^{(3.5)}$ ±0.0224	0.5665$^{(6)}$ ±0.0200	0.5670$^{(5)}$ ±0.0226
`ESL`	**0.7650**$^{(1)}$ ±0.0446	0.7607$^{(3)}$ ±0.0416	0.7556$^{(7)}$ ±0.0351	0.7648$^{(2)}$ ±0.0370	0.7599$^{(4)}$ ±0.0374	0.7592$^{(5)}$ ±0.0374	0.7574$^{(6)}$ ±0.0403
(-) `breast-c`	*0.2208*$^{(4)}$ ±0.0928	*0.2182*$^{(5)}$ ±0.0955	*0.1996*$^{(6)}$ ±0.0873	**0.4415**$^{(1.5)}$ ±0.1002	**0.4415**$^{(1.5)}$ ±0.1002	0.4193$^{(3)}$ ±0.1082	–
`SWD`	0.4074$^{(3)}$ ±0.0934	0.4045$^{(6)}$ ±0.0938	0.4132$^{(2)}$ ±0.0965	0.4054$^{(4)}$ ±0.0954	0.4020$^{(7)}$ ±0.0945	**0.4157**$^{(1)}$ ±0.0967	0.4046$^{(5)}$ ±0.0986
`LEV`	0.5452$^{(5)}$ ±0.0717	0.5424$^{(7)}$ ±0.0713	0.5573$^{(3)}$ ±0.0734	0.5474$^{(4)}$ ±0.0719	0.5424$^{(6)}$ ±0.0751	**0.5634**$^{(1)}$ ±0.0789	0.5615$^{(2)}$ ±0.0753
`ERA`	0.3658$^{(6)}$ ±0.0946	0.3656$^{(7)}$ ±0.0936	0.3837$^{(3)}$ ±0.0901	0.3685$^{(4)}$ ±0.0919	0.3671$^{(5)}$ ±0.0934	0.3876$^{(2)}$ ±0.0892	**0.3976**$^{(1)}$ ±0.0871
average rank (14)	2.57 (2^{nd})	4.68 (5^{th})	5.64 (6^{th})	**2.25** (1^{st})	4.21 (4^{th})	3.79 (3^{rd})	–
average rank (11)	**2.55** (1^{st})	4.82 (5^{th})	5.55 (6^{th})	**2.55** (1^{st})	4.64 (4^{th})	3.64 (2^{nd})	4.27 (3^{rd})

of our approach (which we were able to apply to all 14 data sets), and interpretability of the decision rule preference model, our approach appears to be more attractive for a DM.

- From Tables 5.2 and 5.3, one can discover that the situation (marked by italics) when for a given data set *dat*, the average value of any method is smaller than $avg^*_{dat} - dev^*_{dat}$ is relatively rare (it occurs for 3 data sets in case of measure τ, and for 2 data sets when using measure $\tau^{\neg I}$). This shows that the methods perform similar for most of the data sets.

- One can notice that the standard deviations given in Tables 5.2 and 5.3 are usually similar for different methods with only few exceptions among which the most notable one is the relatively very low standard deviation of SVM^{rank} for `bank-g` data set, considering measure $\tau^{\neg I}$. For 88% of pairs (data set, method), a standard deviation concerning measure τ is lower than the respective standard deviation concerning measure $\tau^{\neg I}$. The highest standard deviations, around 0.2, can be observed for measure $\tau^{\neg I}$ and highly inconsistent `breast-c` data set. There are also several other

TABLE 5.3: Performance of the six versions of VC-DRSA^{rank} (in short V^{rank}) and SVM^{rank} in terms of **measure** $\tau^{\neg\mathbf{I}}$ (largest average of $\tau^{\neg I(rank)}$ ± standard deviation)

Data set	$V^{rank}_{c0\|1}$	$V^{rank}_{c0\text{-}1_{cr}}$	$V^{rank}_{c0\text{-}1_{\times}}$	$V^{rank}_{nc0\|1}$	$V^{rank}_{nc0\text{-}1_{cr}}$	$V^{rank}_{nc0\text{-}1_{\times}}$	SVM^{rank}
(-) car	0.9632$^{(3)}$ ±0.0412	0.9620$^{(5)}$ ±0.0456	**0.9685**$^{(1)}$ ±0.0306	0.9623$^{(4)}$ ±0.0421	0.9613$^{(6)}$ ±0.0464	0.9658$^{(2)}$ ±0.0399	–
housing	**0.8566**$^{(2.5)}$ ±0.0538	**0.8566**$^{(2.5)}$ ±0.0538	0.8418$^{(6)}$ ±0.0721	**0.8566**$^{(2.5)}$ ±0.0538	**0.8566**$^{(2.5)}$ ±0.0538	0.8475$^{(5)}$ ±0.0729	0.8382$^{(7)}$ ±0.0673
cpu	*0.9866*$^{(5.5)}$ ±0.0211	*0.9888*$^{(3.5)}$ ±0.0184	*0.9823*$^{(7)}$ ±0.0187	*0.9866*$^{(5.5)}$ ±0.0211	*0.9888*$^{(3.5)}$ ±0.0184	*0.9897*$^{(2)}$ ±0.0139	**0.9980**$^{(1)}$ ±0.0064
denbosch	0.8485$^{(6)}$ ±0.1701	0.8533$^{(3)}$ ±0.1262	0.8378$^{(7)}$ ±0.1579	0.8494$^{(5)}$ ±0.1687	0.8500$^{(4)}$ ±0.1695	**0.8715**$^{(1)}$ ±0.1697	0.8704$^{(2)}$ ±0.1546
bank-g	*0.9064*$^{(4)}$ ±0.0989	*0.9055*$^{(5.5)}$ ±0.0986	*0.9256*$^{(3)}$ ±0.0908	*0.9047*$^{(7)}$ ±0.1042	*0.9055*$^{(5.5)}$ ±0.1015	*0.9272*$^{(2)}$ ±0.0893	**0.9970**$^{(1)}$ ±0.0142
fame	0.8769$^{(6)}$ ±0.0381	0.8778$^{(4)}$ ±0.0392	0.8728$^{(7)}$ ±0.0362	0.8772$^{(5)}$ ±0.0382	0.8780$^{(3)}$ ±0.0388	**0.8855**$^{(1)}$ ±0.0338	0.8850$^{(2)}$ ±0.0394
(-) windsor	0.7567$^{(4)}$ ±0.0758	0.7569$^{(3)}$ ±0.0759	0.7465$^{(6)}$ ±0.0700	0.7580$^{(2)}$ ±0.0722	**0.7583**$^{(1)}$ ±0.0701	0.7510$^{(5)}$ ±0.0742	–
breast-w	0.9952$^{(4.5)}$ ±0.0095	0.9952$^{(4.5)}$ ±0.0096	**0.9957**$^{(1)}$ ±0.0090	0.9952$^{(4.5)}$ ±0.0095	0.9952$^{(4.5)}$ ±0.0094	0.9954$^{(2)}$ ±0.0086	0.9923$^{(7)}$ ±0.0141
balance-scale	**0.9637**$^{(1.5)}$ ±0.0319	0.9635$^{(3)}$ ±0.0313	0.9614$^{(7)}$ ±0.0318	**0.9637**$^{(1.5)}$ ±0.0319	0.9631$^{(4)}$ ±0.0318	0.9624$^{(6)}$ ±0.0304	0.9630$^{(5)}$ ±0.0299
ESL	0.9089$^{(3)}$ ±0.0446	**0.9101**$^{(1)}$ ±0.0443	0.9041$^{(7)}$ ±0.0366	0.9086$^{(4)}$ ±0.0447	0.9093$^{(2)}$ ±0.0398	0.9085$^{(5)}$ ±0.0396	0.9062$^{(6)}$ ±0.0436
(-) breast-c	0.4625$^{(4)}$ ±0.2096	**0.4661**$^{(1)}$ ±0.2138	0.4498$^{(6)}$ ±0.2049	0.4635$^{(3)}$ ±0.2158	0.4649$^{(2)}$ ±0.2159	0.4617$^{(5)}$ ±0.1988	–
SWD	0.5805$^{(5)}$ ±0.1359	0.5807$^{(4)}$ ±0.1359	0.5933$^{(2)}$ ±0.1397	0.5770$^{(7)}$ ±0.1367	0.5772$^{(6)}$ ±0.1369	**0.5970**$^{(1)}$ ±0.1400	0.5810$^{(3)}$ ±0.1426
LEV	0.7317$^{(6)}$ ±0.0951	0.7322$^{(5)}$ ±0.0955	0.7526$^{(3)}$ ±0.0983	0.7289$^{(7)}$ ±0.0952	0.7323$^{(4)}$ ±0.1009	**0.7609**$^{(1)}$ ±0.1059	0.7583$^{(2)}$ ±0.1011
ERA	0.4075$^{(7)}$ ±0.1057	0.4084$^{(6)}$ ±0.1046	0.4288$^{(3)}$ ±0.1005	0.4108$^{(4)}$ ±0.1030	0.4101$^{(5)}$ ±0.1045	0.4332$^{(2)}$ ±0.1000	**0.4445**$^{(1)}$ ±0.0969
average rank (14)	4.43 (4^{th})	3.61 (2^{nd})	4.71 (5^{th})	4.43 (4^{th})	3.82 (3^{rd})	**2.86** (1^{st})	–
average rank (11)	4.64 (5^{th})	3.77 (3^{rd})	4.82 (6^{th})	4.82 (6^{th})	4.05 (4^{th})	**2.55** (1^{st})	3.36 (2^{nd})

TABLE 5.4: Best parameter values for the six versions of VC-DRSA^{rank} (in short V^{rank}) and for SVM^{rank} – performance measured using **measure** τ

Data set	$V^{rank}_{c0\|1}$	$V^{rank}_{c0\text{-}1_{cr}}$	$V^{rank}_{c0\text{-}1_{\times}}$	$V^{rank}_{nc0\|1}$	$V^{rank}_{nc0\text{-}1_{cr}}$	$V^{rank}_{nc0\text{-}1_{\times}}$	SVM^{rank}
(-) car	0.1	0	0.1	0.1	0	0.1	–
housing	0	0	0.01	0	0	0.01	0.1
cpu	0.05	0.05	0.05	0.05	0.05	0.01	0.1
denbosch	0.01	0	0.05	0.01	0	0.01	0.01
bank-g	0.01	0	0.01	0.01	0	0.01	0.001
fame	0.01	0.01	0.01	0.01	0.01	0.01	0.001
(-) windsor	0.01	0	0.05	0.01	0.01	0.01	–
breast-w	0.01	0	0.1	0.01	0	0	0.001
balance-scale	0.05	0	0.15	0.05	0	0	1
ESL	0.01	0.01	0.15	0.15	0.15	0.15	1
(-) breast-c	0.1	0	0.15	0	0	0	–
SWD	0.01	0.01	0.1	0.01	0.01	0.01	0.001
LEV	0.01	0.01	0.1	0.15	0.15	0.1	10
ERA	0.01	0.01	0.1	0.01	0.01	0.1	0.01

TABLE 5.5: Best parameter values for the six versions of VC-DRSA^{rank} (in short V^{rank}) and for SVM^{rank} – performance measured using **measure** $\tau^{\neg \mathbf{I}}$

Data set	$\mathrm{V}^{rank}_{c\,0\|1}$	$\mathrm{V}^{rank}_{c\,0\text{-}1_{cr}}$	$\mathrm{V}^{rank}_{c\,0\text{-}1_{\times}}$	$\mathrm{V}^{rank}_{nc\,0\|1}$	$\mathrm{V}^{rank}_{nc\,0\text{-}1_{cr}}$	$\mathrm{V}^{rank}_{nc\,0\text{-}1_{\times}}$	SVM^{rank}
(-) car	0.01	0.01	0.1	0.01	0.01	0.01	–
housing	0	0	0.01	0	0	0.01	0.1
cpu	0.05	0.05	0.05	0.05	0.05	0.01	0.1
denbosch	0.01	0.05	0.05	0.01	0.01	0.01	0.01
bank-g	0.05	0.05	0.01	0.01	0.01	0.01	0.1
fame	0.01	0.01	0.01	0.01	0.01	0.01	0.001
(-) windsor	0.01	0.01	0.05	0.01	0.01	0.01	–
breast-w	0	0	0.1	0.1	0.1	0.1	0.001
balance-scale	0.05	0.1	0.15	0.05	0.1	0	1
ESL	0.01	0.01	0.15	0.01	0.15	0.15	1
(-) breast-c	0.1	0.1	0.15	0.15	0.15	0.15	–
SWD	0.01	0.01	0.1	0.01	0.01	0.01	0.001
LEV	0.01	0.01	0.1	0.01	0.15	0.1	10
ERA	0.01	0.01	0.1	0.01	0.01	0.05	0.01

TABLE 5.6: Average percentage of pairs of objects $(a,b) \in A \times A$ such that: (i) $(a,b) \in I^{\neq}_{\succeq^i_A}$, (ii) $(a,b) \in P_{\succeq^i_A} \cup P^{-1}_{\succeq^i_A}$, (iii) $(a,b) \in I^{\neq}_{\succeq^i_A}$ and $(a,b) \in I^{\neq}_{\succeq^f_A}$ (columns denoted by '$I^{\neq}_{\surd}$'), (iv) $(a,b) \in P_{\succeq^i_A}$ and $(a,b) \in P_{\succeq^f_A}$, or $(a,b) \in P^{-1}_{\succeq^i_A}$ and $(a,b) \in P^{-1}_{\succeq^f_A}$ (columns denoted by '$P_{\surd}\|P^{-1}_{\surd}$')

Column:	1A	1B	2A	2B	3A	3B	4A	4B	5A	5B	6A	6B	7A	7B	8A	8B
Data set			$\mathrm{V}^{rank}_{c\,0\|1}$		$\mathrm{V}^{rank}_{c\,0\text{-}1_{cr}}$		$\mathrm{V}^{rank}_{c\,0\text{-}1_{\times}}$		$\mathrm{V}^{rank}_{nc\,0\|1}$		$\mathrm{V}^{rank}_{nc\,0\text{-}1_{cr}}$		$\mathrm{V}^{rank}_{nc\,0\text{-}1_{\times}}$		SVM^{rank}	
	$I^{\neq}_{\succeq^i_A}$	$P_{\succeq^i_A} \cup P^{-1}_{\succeq^i_A}$	$I^{\neq}_{\surd}$	$P_{\surd}\|P^{-1}_{\surd}$	$I^{\neq}_{\surd}$	$P_{\surd}\|P^{-1}_{\surd}$	$I^{\neq}_{\surd}$	$P_{\surd}\|P^{-1}_{\surd}$	$I^{\neq}_{\surd}$	$P_{\surd}\|P^{-1}_{\surd}$	$I^{\neq}_{\surd}$	$P_{\surd}\|P^{-1}_{\surd}$	$I^{\neq}_{\surd}$	$P_{\surd}\|P^{-1}_{\surd}$	$I^{\neq}_{\surd}$	$P_{\surd}\|P^{-1}_{\surd}$
(-) car	52%	45%	**22%**	44%	**18%**	44%	0%	44%	**24%**	44%	**18%**	44%	5%	44%	–	–
housing	21%	75%	0%	70%	0%	70%	0%	69%	0%	70%	0%	70%	0%	69%	0%	69%
cpu	20%	75%	1%	74%	0%	75%	0%	74%	1%	74%	0%	75%	0%	75%	0%	75%
denbosch	42%	50%	**5%**	46%	1%	46%	0%	46%	**4%**	46%	1%	46%	0%	47%	0%	47%
bank-g	71%	26%	2%	25%	1%	25%	0%	25%	2%	25%	0%	25%	0%	25%	0%	26%
fame	19%	78%	0%	73%	0%	73%	0%	73%	0%	73%	0%	73%	0%	74%	0%	74%
(-) windsor	21%	75%	1%	66%	0%	66%	0%	65%	0%	66%	0%	66%	0%	66%	–	–
breast-w	52%	45%	**7%**	45%	2%	45%	1%	45%	**7%**	45%	4%	45%	2%	45%	1%	45%
balance-scale	47%	50%	**8%**	49%	**8%**	49%	**7%**	49%	**8%**	49%	**8%**	49%	**7%**	49%	**7%**	49%
ESL	16%	80%	1%	76%	0%	76%	0%	76%	1%	76%	0%	76%	0%	76%	0%	76%
(-) breast-c	55%	42%	2%	31%	2%	31%	0%	30%	**23%**	31%	**23%**	31%	**21%**	31%	–	–
SWD	30%	67%	1%	53%	0%	53%	0%	53%	1%	53%	0%	53%	0%	53%	0%	53%
LEV	26%	71%	1%	62%	1%	62%	1%	62%	1%	61%	1%	62%	1%	63%	1%	63%
ERA	11%	86%	0%	61%	0%	61%	0%	62%	0%	61%	0%	61%	0%	62%	0%	62%

data sets with standard deviations around 0.1 and above – `bank-g` (see measure $\tau^{\neg I}$, and methods other than SVM^{rank}) and `denbosch`, as well as the following highly inconsistent data sets: `SWD`, `LEV` (see measure $\tau^{\neg I}$), and `ERA`.

- The values of measure $\tau^{\neg I}$ given in Table 5.3 are all higher than the respective values of measure τ given in Table 5.2. This proves that performance of all the methods was better when considering preference and inverse preference relations only. This is a result that one could expect. In our approach, two objects $a, b \in A$ have the same rank in the final ranking $NFR(A, \mathcal{R})$ only in a rather rare case, i.e., when they have exactly the same value of scoring function SD (2.48), which occurs if $SD(a, A, \mathcal{R}) = SD(b, A, \mathcal{R})$. Moreover, SVM^{rank} employs pairwise preference constraints concerning objects from different classes only. In this way, it does not learn indifference relations. Therefore, one can argue that in our experiment, values of measure $\tau^{\neg I}$ should be treated with more care as they address directly the more important aspect, i.e., correct prediction of preference and inverse preference relations.

- We tried to identify reasons why for particular data sets we obtained generally lower or higher values of measures τ (see Table 5.2) and $\tau^{\neg I}$ (see Table 5.3). In our opinion, this can be roughly explained using two factors: (i) average value of the quality of approximation $\gamma(S, S^c)$ (2.16) (see Table 5.1), reflecting data set consistency, and (ii) average percentages reported in Table 5.6. We observed the following general trends:

  - the lower the value $\overline{\gamma}(S, S^c)$ (presented in the last column of Table 5.1), the lower the values of τ and $\tau^{\neg I}$,
  - the higher the average percentage of pairs of objects $(a, b) \in A \times A$ such that $(a, b) \in I^{\neq}_{\succeq^i_A}$ (presented in column 1A of Table 5.6), the lower the values of τ.

  The latter trend underlines generally poor prediction of indifference relations by all the methods, although VC-DRSA$^{rank}_{c\,0|1}$ and VC-DRSA$^{rank}_{nc\,0|1}$ are relatively the best at this task, as results from the comparison of values in columns 2A, 3A, ..., 8A of Table 5.6.

- The choice of the best version of VC-DRSA^{rank} depends on the chosen performance measure (τ or $\tau^{\neg I}$).

  Looking at the average ranks given in Table 5.2, one can observe that versions VC-DRSA$^{rank}_{c\,0|1}$ and VC-DRSA$^{rank}_{nc\,0|1}$ obtained the best result (i.e., the lowest average rank) according to measure τ, with a slight advantage of the latter method due to its lower average rank for all 14 data sets. Moreover, VC-DRSA$^{rank}_{c\,0|1}$ and VC-DRSA$^{rank}_{nc\,0|1}$ were better than SVM^{rank} for 8 out of 11 data sets, with the greatest difference of τ in favor of both our methods approximately equal to 0.071, and the greatest difference of τ in disfavor of VC-DRSA$^{rank}_{c\,0|1}$ equal to 0.0319.

On the other hand, from Table 5.3, one can discover that version VC-DRSA$^{rank}_{nc\,0\text{-}1_\times}$ performed best with respect to measure $\tau^{\neg I}$. This version was better than SVM^{rank} for 7 out of 11 data sets, with the greatest difference of $\tau^{\neg I}$ in favor of our method equal to 0.016, and the greatest difference of $\tau^{\neg I}$ in disfavor of VC-DRSA$^{rank}_{nc\,0\text{-}1_\times}$ equal to 0.0698.

The above observations, and the percentages given in columns 2A, 2B, 5A, 5B, 7A, and 7B of Table 5.6, lead to the conclusion that the "crisp" versions of VC-DRSA^{rank} (i.e., VC-DRSA$^{rank}_{c\,0|1}$, VC-DRSA$^{rank}_{nc\,0|1}$) better predict indifference relations, while the "valued" version VC-DRSA$^{rank}_{nc\,0\text{-}1_\times}$ is better at predicting preference and inverse preference relations.

- According to performance measure τ, VC-DRSA$^{rank}_{nc\,0\text{-}1_\times}$ was better than VC-DRSA$^{rank}_{c\,0\text{-}1_\times}$ for all 14 data sets.
- The version VC-DRSA$^{rank}_{c\,0\text{-}1_\times}$ is systematically (i.e., for both performance measures) the worst version of VC-DRSA^{rank} and thus, it is not recommended.
- From Tables 5.4 and 5.5, one can confirm that employing ϵ-VC-DRSA improves performance, especially in case of measure $\tau^{\neg I}$ – in most of the cases the largest average value of considered performance measure was obtained for $\theta_S = \theta_{S^c} > 0$. An interesting observation is that a decrease of data set consistency (which occurs when moving down the tables) does not involve, in general, an increase of thresholds θ_S, θ_{S^c}.

Following the guidelines from [48], we have applied two non-parametric statistical tests to conduct analysis of the results presented in Tables 5.2 and 5.3, concerning the 11 data sets for which we were able to obtain results using all 7 methods. The aim of this analysis was to verify if the compared methods are statistically significantly different, and which are the particular methods that differ in performance.

First, we performed Friedman's test [68, 69] for each performance measure, with the null-hypothesis that all 7 methods are equivalent (they perform equally well) and so their average ranks are equal. In particular, we used less conservative Iman and Davenport's statistic F_F [107], derived from Friedman's statistic, and distributed according to F-distribution. The p-value corresponding to calculated value of statistic F_F was around 0.0022 for measure τ, and around 0.1097 for measure $\tau^{\neg I}$. Assuming significance level $\alpha = 0.01$, we could reject the null-hypothesis only for measure τ. Using the Nemenyi post-hoc test [131], with significance level $\alpha = 0.01$, we obtained critical difference of average ranks CD equal to 3.179748951. As the greatest difference of average ranks for the 11 data sets and measure τ was equal to 3.0 (e.g., between average ranks of VC-DRSA$^{rank}_{c\,0|1}$ and VC-DRSA$^{rank}_{c\,0\text{-}1_\times}$), applied post-hoc test did not reveal any significant difference in performance of any two methods.

As the Friedman's test and the Nemenyi post-hoc analysis are based only on average ranks of the methods (in our case, they were based on the average ranks given in the last rows of Tables 5.2 and 5.3), we decided to apply also Wilcoxon signed-rank paired test [176] to compare all the methods pairwise. The null-hypothesis in each paired test was that both compared methods perform equally well. The p-values obtained in the tests carried out for both performance measures are presented in Tables 5.7 and 5.8. In these tables, bold font is used to denote situations where p-value obtained in the test is lower than the assumed significance threshold $\alpha = 0.01$, and, moreover, the method from the corresponding row has lower average rank that the method from the corresponding column.

TABLE 5.7: p-values in Wilcoxon signed-rank paired tests involving the six versions of VC-DRSA^{rank} (in short V^{rank}) and SVM^{rank} – performance measured using **measure** τ

	$V^{rank}_{c\,0\|1}$	$V^{rank}_{c\,0\text{-}1_{cr}}$	$V^{rank}_{c\,0\text{-}1_{\times}}$	$V^{rank}_{nc\,0\|1}$	$V^{rank}_{nc\,0\text{-}1_{cr}}$	$V^{rank}_{nc\,0\text{-}1_{\times}}$	SVM^{rank}
$V^{rank}_{c\,0\|1}$	–	**0.00592**	0.08301	1.00000	**0.00805**	0.32031	0.27832
$V^{rank}_{c\,0\text{-}1_{cr}}$	0.00592	–	0.41309	0.00592	0.72228	0.83105	0.76465
$V^{rank}_{c\,0\text{-}1_{\times}}$	0.08301	0.41309	–	0.08301	0.36523	0.00098	0.06738
$V^{rank}_{nc\,0\|1}$	1.00000	**0.00592**	0.08301	–	**0.00592**	0.32031	0.27832
$V^{rank}_{nc\,0\text{-}1_{cr}}$	0.00805	0.72228	0.36523	0.00592	–	0.76465	0.83105
$V^{rank}_{nc\,0\text{-}1_{\times}}$	0.32031	0.83105	**0.00098**	0.32031	0.76465	–	0.63770
SVM^{rank}	0.27832	0.76465	0.06738	0.27832	0.83105	0.63770	–

TABLE 5.8: p-values in Wilcoxon signed-rank paired tests involving the six versions of VC-DRSA^{rank} (in short V^{rank}) and SVM^{rank} – performance measured using **measure** $\tau^{\neg\mathbf{I}}$

	$V^{rank}_{c\,0\|1}$	$V^{rank}_{c\,0\text{-}1_{cr}}$	$V^{rank}_{c\,0\text{-}1_{\times}}$	$V^{rank}_{nc\,0\|1}$	$V^{rank}_{nc\,0\text{-}1_{cr}}$	$V^{rank}_{nc\,0\text{-}1_{\times}}$	SVM^{rank}
$V^{rank}_{c\,0\|1}$	–	0.05802	0.70020	0.67260	0.47720	0.05371	0.14746
$V^{rank}_{c\,0\text{-}1_{cr}}$	0.05802	–	0.76465	0.07556	0.36273	0.08301	0.17480
$V^{rank}_{c\,0\text{-}1_{\times}}$	0.70020	0.76465	–	0.63770	0.63770	0.00195	0.10156
$V^{rank}_{nc\,0\|1}$	0.67260	0.07556	0.63770	–	0.07556	0.04199	0.08301
$V^{rank}_{nc\,0\text{-}1_{cr}}$	0.47720	0.36273	0.63770	0.07556	–	0.06738	0.08301
$V^{rank}_{nc\,0\text{-}1_{\times}}$	0.05371	0.08301	**0.00195**	0.04199	0.06738	–	0.76465
SVM^{rank}	0.14746	0.17480	0.10156	0.08301	0.08301	0.76465	–

Taking into account Table 5.7, we draw the following conclusion w.r.t. measure τ:

- VC-DRSA$^{rank}_{c\,0|1}$ performs statistically significantly better than VC-DRSA$^{rank}_{c\,0\text{-}1_{cr}}$ and VC-DRSA$^{rank}_{nc\,0\text{-}1_{cr}}$,

- VC-DRSA$^{rank}_{nc\,0|1}$ performs statistically significantly better than VC-DRSA$^{rank}_{c\,0\text{-}1_{cr}}$ and VC-DRSA$^{rank}_{nc\,0\text{-}1_{cr}}$,
- VC-DRSA$^{rank}_{nc\,0\text{-}1_{\times}}$ performs statistically significantly better than VC-DRSA$^{rank}_{c\,0\text{-}1_{\times}}$.

Moreover, using Table 5.8, we infer that VC-DRSA$^{rank}_{nc\,0\text{-}1_{\times}}$ performs statistically significantly better than VC-DRSA$^{rank}_{c\,0\text{-}1_{\times}}$ also w.r.t. measure $\tau^{\neg I}$.

Chapter 6

Summary and Conclusions

6.1 Main Contributions of the Thesis

This thesis was focused on new adaptations and improvements of the Dominance-based Rough Set Approach (DRSA) to multicriteria ranking and similarity-based classification (case-based reasoning) problems. This goal was decomposed into four specific objectives (o1)-(o4) presented in Section 1.5. In our opinion, we managed to achieve all these objectives. In what follows, we summarize respective contribution of our research reported in this thesis.

Methodology for multicriteria ranking using VC-DRSA.

In Chapter 2, we proposed a rule-based methodology for multicriteria ranking, denoted by VC-DRSA^{rank}. Our approach can be described by the following general features:

- it employs preference model in terms of a set of monotonic decision rules; rule preference model is the most general preference model, relatively easy to understand by a DM;
- it is concordant with the current trend in MCDA which consists in induction of a preference model from decision examples;
- it involves simple decision examples in the form of pairwise comparisons of objects in terms of outranking relation S and non-outranking relation S^c;
- it can handle cardinal and ordinal criteria simultaneously, without prior discretization of numerical criteria or prior conversion of ordinal criteria into numerical ones (it is sufficient that values of each criterion are expressed on an ordinal scale);
- it can handle (using the dominance-based rough set concept) inconsistency of decision examples, i.e., violations of the monotonic relationship "if object a is preferred to object b at least as much as object c is preferred to object d with respect to each

considered criterion, then the comprehensive preference of a over b is not weaker than the comprehensive preference of c over d"; thanks to this ability, it does not require that the set of criteria is a consistent one.

Introducing our approach, we proposed the following extensions and improvements of the previous MCDA rule-based approaches presented in Section 1.2.1.

- VC-DRSA^{rank} does not require from a DM to define graded preference relations for particular cardinal criteria, as considered in all rule-based MCDA approaches reviewed in Section 1.2.1. Instead it uses difference of evaluations as a simple measure of the strength of preference.

- VC-DRSA^{rank} is suited for solving real-life multicriteria ranking problems by using an adaptation of ϵ-VC-DRSA. Thanks to employing this adaptation, lower approximations of outranking and non-outranking relations are allowed to contain not only consistent pairs of objects, but also pairs of objects that are "sufficiently consistent" according to cost-type consistency measures ϵ_S, ϵ_{S^c}, defined in Section 2.5. Consequently, VC-DRSA^{rank} employs probabilistic decision rules whose consistency is measured using a cost-type rule consistency measure $\widehat{\epsilon}_T$ (2.21).

- We considered two view points concerning the nature of the set G of criteria describing considered objects. The first one, typical for MCDA, and reflected by generic version VC-DRSA$_c^{rank}$, consists in assuming that set G is a consistent set of criteria (as defined in Section 1.1). The second one, typical for PL, and reflected by generic version VC-DRSA$_{nc}^{rank}$, does not involve any assumptions concerning set G (i.e., G is considered to be a not necessarily consistent set of criteria, so it can miss some important criteria and/or contain some redundant ones). For each of the two considered view points, we proposed an appropriate way of constructing PCT from decision examples. We also showed how each view point should be taken into account during application of decision rules.

- We employed VC-DomLEM rule induction algorithm to induce probabilistic decision rules from lower approximations of S and S^c. Previous MCDA rule-based approaches to multicriteria ranking relied on the DomLEM algorithm [92, 161].

- We considered, in total, six ways of constructing relations $\mathbb{S}$ and $\mathbb{S}^c$ resulting from application of induced decision rules on a set A of objects to be ranked. These ways are reflected by three versions of VC-DRSA$_c^{rank}$ and three versions of VC-DRSA$_{nc}^{rank}$, all introduced in Section 2.1, and defined in Section 2.7. Thus, we considered:

  - versions VC-DRSA$_{c\,0|1}^{rank}$ and VC-DRSA$_{nc\,0|1}^{rank}$, where relations $\mathbb{S}$ and $\mathbb{S}^c$ are crisp (they reflect existence/non-existence of covering rules);

– versions VC-DRSA$^{rank}_{c\,0\text{-}1_{cr}}$ and VC-DRSA$^{rank}_{nc\,0\text{-}1_{cr}}$, with relations $\mathbb{S}$ and $\mathbb{S}^c$ being valued and reflecting maximum strength of covering rules, where the strength of a rule r_T is measured as $1 - \widehat{\epsilon}_T(r_T)$ ("credibility");

– versions VC-DRSA$^{rank}_{c\,0\text{-}1_{\times}}$ and VC-DRSA$^{rank}_{nc\,0\text{-}1_{\times}}$, with relations $\mathbb{S}$ and $\mathbb{S}^c$ being valued and reflecting maximum strength of covering rules, where the strength of a rule r_T is measured as $\big(1 - \widehat{\epsilon}_T(r_T)\big)cf(r_T)$ (product of "credibility" and coverage factor).

- We proposed a generic two-step procedure for exploitation of the crisp/valued preference structure on set A, composed of crisp/valued relations $\mathbb{S}$ and $\mathbb{S}^c$, and represented by preference graph $\mathbb{G}$. This procedure was introduced in Section 2.8.1 – approach (*iv*). It consists in a suitable transformation (described in Section 2.8.2) of preference graph $\mathbb{G}$ to another graph $\mathbb{G}'$, representing valued relation $\mathcal{R}$ (2.37), and in subsequent exploitation of this relation using a ranking method, leading to a total or partial preorder over A. The proposed two-step exploitation procedure generalizes the one-step procedure based on net flow scores (approach (i) of Section 2.8.1), applied in the previous MCDA rule-based approaches presented in Section 1.2.1. This is because the two-step exploitation procedure yields the same total preorder over A when *Net Flow Rule* ranking method is applied in step two. Moreover, the proposed two-step exploitation procedure enables to use any ranking method known from the literature.

- In Section 2.8.3, we introduced a useful taxonomy consisting of two generic score-based ranking methods: single-stage ranking method $\succeq^1$ (which ranks all objects in a single run), and multi-stage ranking method $\succeq^i$ (which involves iterative application of a choice function). These ranking methods are parameterized by a scoring function sf. We reviewed nine scoring functions, and we showed how five popular ranking methods can be expressed in terms of $\succeq^1$ or $\succeq^i$, using particular scoring functions.

- In Section 2.9.2, we proposed a new concordance measure τ' that generalizes Kendall rank correlation coefficient τ. This measure is suited for the most general case of measuring concordance between a ranking being a partial preorder over set A and pairwise comparisons (in terms of outranking and non-outranking relations) of objects from a subset of A.

Analysis of preference graph exploitation procedures.

Our contribution concerning procedures for exploitation of a preference graph is the following.

- In Section 2.8.1, we analyzed four alternative strategies of exploitation of preference graph $\mathbb{G}$ resulting from application of induced decision rules on a set A of objects to be ranked.

- This analysis motivated us to investigate the newly proposed two-step exploitation procedure consisting in a suitable transformation of preference graph $\mathbb{G}$ to another graph $\mathbb{G}'$, representing a valued relation over A, and in subsequent exploitation of this relation using a ranking method, leading to a total or partial preorder over A. As described above, in step one of this procedure, we employed the transformation defined by (2.37). In case when relations $\mathbb{S}$ and $\mathbb{S}^c$ are crisp, this transformation yields a three-valued relation over A; when $\mathbb{S}$ and $\mathbb{S}^c$ are valued, this transformation yields a general valued relation over A. In step two, we considered application of one of the five ranking methods known from the literature: *Net Flow Rule*, *Iterative Net Flow Rule*, *Min in Favor*, *Iterative Min in Favor*, and *Leaving and Entering Flows.*

- We searched for "the best" ranking method. For this purpose, we considered in Section 3.1 eleven desirable properties of a ranking method. Some of these properties have been already studied in the literature, and some of them were introduced for the first time in this thesis. We presented a large number of proofs concerning the properties of the five ranking methods. They can be found in the Appendix.

- In Section 3.2, we proposed two priority orders of desirable properties of a ranking method, one for the case of a three-valued relation, and the other for the case of a general valued relation. These orders reflect relative importance of desirable properties.

- We found out that the best ranking method for exploitation of, both, a three-valued relation, and a general valued relation, is the *Net Flow Rule* method.

Methodology for similarity-based classification using DRSA.

In Chapter 4, we proposed a rule-based methodology for the similarity-based classification problem introduced in Section 1.1.2. Our approach can be described by the following general features:

- it is based on an adaptation of DRSA to case-based reasoning;

- it avoids using any real-valued aggregation function (involving operators, like weighted L_p norm, min, etc.) to aggregate marginal similarities of objects into their comprehensive similarity; instead, comprehensive similarity is represented by decision rules induced from classification examples;

- it is an eager learning method – it involves learning of similarity in terms of decision rules; comparing to lazy learning methods of CBR, our approach is thus less susceptible to noise observed in the training data, both with respect to irrelevant features (because each decision rule, being a partial dominance cone in a similarity space, may involve conditions concerning only a subset of features), and with respect to outliers;

- it is invariant to ordinally equivalent marginal similarity functions as induced rules employ only ordinal properties of marginal similarity functions;

- it assumes that each decision class X is a fuzzy set in U, characterized by membership function $\mu_X : U \rightarrow [0, 1]$;

- it exploits only ordinal properties of membership functions of considered decision classes.

Let us emphasize that the definitions of rough approximations presented in Section 4.6, and the syntax of decision rules given in Section 4.7, employ only ordinal properties of marginal similarity functions and membership functions. No algebraic operation, such as sum or product, involving cardinal properties of marginal similarity functions, is considered.

Introducing our approach, we proposed the following extensions and improvements of the previous dominance-based rough set approach to case-based reasoning presented in Section 1.3.2.

- We employed a new monotonic relationship "the more similar is object y to object x w.r.t. the considered features, the closer is y to x in terms of the membership to a given decision class X". In our opinion, this relationship truly reflects the monotonicity characteristic for CBR, i.e., monotonic relationship between comprehensive similarity of objects and their similarities w.r.t. single features.

- We revised the definitions of comprehensive closeness relations, appearing in early versions of our methodology, by allowing, in one of these definitions, parameter α to be less than zero, and parameter β to be greater than one.

- We introduced the concept of a similarity table.

- We introduced the concept of x-dominance relation and x-positive and x-negative dominance cones in the similarity space, where x is a reference object.

- We proposed the way of inducing decision rules. The rules are induced independently for each reference object and each decision class, using VC-DomLEM algorithm.

- We proposed a way of application of induced decision rules to a new object z (new case), in order to predict its membership to a given decision class X. This way consists in calculating a score for each membership degree that is covered by the decision part of at least one rule matching z. Then, the predicted membership degree is the one with the highest score.

Experimental verification of the methodology for multicriteria ranking.

In Chapter 5, we presented the results of a computational experiment performed to compare six different versions of VC-DRSA^{rank}, proposed in Chapter 2, with another state-of-the-art method from the field of Preference Learning – SVM^{rank}.

The experiment showed that the proposed approach to preference learning in multicriteria ranking is clearly competitive to SVM^{rank}. Taking into account wider applicability of our approach (which works also for ordinal criteria), and interpretability of the decision rule preference model, our approach appears to be more attractive for a DM.

In order to compare performance of all seven methods, we used Kendall rank correlation coefficient τ (see Section 2.9.1), as well as its variant $\tau^{\neg I}$, introduced in Section 5.1. The latter coefficient does not take into account pairs of objects $(a, b) \in A \times A$ such that according to the true ranking on A, a is considered indifferent to b (both objects have the same rank). According to measure τ, the "crisp" versions of VC-DRSA^{rank}, i.e., VC-DRSA$^{rank}_{c\,0|1}$ and VC-DRSA$^{rank}_{nc\,0|1}$, obtained in the experiment the best (i.e., the lowest) average ranks over 11 data sets. On the other hand, version VC-DRSA$^{rank}_{nc\,0\text{-}1_\times}$ obtained the lowest average rank with respect to measure $\tau^{\neg I}$.

The experiment showed that by adaptation of ϵ-VC-DRSA, it was possible to obtain better values of both performance measures than in case of adapting classical DRSA. This is visible especially for measure $\tau^{\neg I}$. In most of the cases, the largest average value of considered performance measure (τ or $\tau^{\neg I}$) was obtained for thresholds $\theta_S = \theta_{S^c} > 0$, rather than for thresholds $\theta_S = \theta_{S^c} = 0$ typical for the classical DRSA.

We have also conducted a statistical analysis of average values of measures τ and $\tau^{\neg I}$ obtained in the experiment. The aim of this analysis was to verify if the compared methods are statistically significantly different, and which are the particular methods that differ in performance. We assumed significance level $\alpha = 0.01$. Using the Friedman's test, we were able to reject the null-hypothesis that all 7 methods perform equally well only in case of measure τ. However, the Nemenyi post-hoc test did not reveal any significant difference in performance of any two methods. We performed also a series of Wilcoxon signed-rank paired tests to compare all the methods pairwise. The null-hypothesis in each paired test was that both compared methods perform equally well. These tests let us to draw the following statistically significant conclusions:

- VC-DRSA$^{rank}_{c\,0|1}$ performs w.r.t. measure τ better than VC-DRSA$^{rank}_{c\,0\text{-}1_{cr}}$ and better than VC-DRSA$^{rank}_{nc\,0\text{-}1_{cr}}$,

- VC-DRSA$^{rank}_{nc\,0|1}$ performs w.r.t. measure τ better than VC-DRSA$^{rank}_{c\,0\text{-}1_{cr}}$ and better than VC-DRSA$^{rank}_{nc\,0\text{-}1_{cr}}$,
- VC-DRSA$^{rank}_{nc\,0\text{-}1_{\times}}$ performs w.r.t. measures τ and $\tau^{\neg I}$ better than VC-DRSA$^{rank}_{c\,0\text{-}1_{\times}}$.

6.2 Directions of Future Research

Below, we list a few possible directions of future research.

- Extension of the computational evaluation of the proposed methodology for multicriteria ranking by considering also other methods than SVM^{rank}, and more data sets, including real-world decision problems.
- Development of other promising adaptations of DRSA. One possible adaptation concerns problems with hierarchical structure of criteria. In this type of problems, a ranking of objects concerning criteria considered at a given level of hierarchy can be treated as a criterion of the upper level. Concerning the disadvantages of the AHP method discussed in this thesis, an approach adapting DRSA to problems with hierarchical structure of criteria could be a strong alternative to AHP. Another possible adaptation of DRSA concerns a multicriteria ranking problem with preference information given in terms of ordinal classification of objects, and processed as purely ordinal using the concept of the ordinal intensity of preference. The first version of this adaptation was already presented in [116].
- Further development of the methodology for similarity-based classification using DRSA, in particular development of methods of selection of reference objects (cases), and adaptation of ϵ-VC-DRSA.
- Experimental comparison of the proposed methodology for similarity-based classification using DRSA with other methods, admitting some "relaxation" of the presented problem setting, e.g., by:
  - assuming only crisp decision classes – in such case we could compare our method with some machine learning approaches to CBR, e.g., with k-NN,
  - assuming that the classification does not need to be based on similarity to reference objects – in such case we could compare our method with some machine learning approaches to soft label classification (like fuzzy-input fuzzy-output SVM [167]).

  It would also be interesting to verify performance of our similarity-based classification method using DRSA when applied to ordinal classification problems. In particular, it would be interesting to compare its performance with that of the classical DRSA.

Appendix

Definition 14 (Monotonicity property (m1)) *A cost-type consistency measure Θ_T, $T \in \{S, S^c\}$, has monotonicity property (m1) iff it is monotonically non-increasing w.r.t. the considered set of criteria, i.e., iff for all $P \subseteq R \subseteq G$, and for all $(a,b) \in B$*

$$\Theta_T^P(a,b) \geq \Theta_T^R(a,b),$$

where $\Theta_T^P(a,b)$ denotes the value of measure Θ_T calculated for pair of objects (a,b) taking into account only criteria from set $P \subseteq G$.

Definition 15 (Monotonicity property (m2)) *A cost-type consistency measure Θ_T, $T \in \{S, S^c\}$, has monotonicity property (m2) iff it is monotonically non-increasing w.r.t. the considered comprehensive preference relation T, i.e., iff for all $T' = T \cup T^{\Delta}$, $T^{\Delta} \cap B = \emptyset$, and for all $(a,b) \in B$*

$$\Theta_T(a,b) \geq \Theta_{T'}(a,b).$$

Definition 16 (Monotonicity property (m4)) *A cost-type consistency measure Θ_T, $T \in \{S, S^c\}$, has monotonicity property (m4) iff it is monotonically non-increasing w.r.t. dominance relation D_2 over B, i.e., iff*

$$\forall (a,b),(c,d) \in B : (a,b)D_2(c,d) \Rightarrow \Theta_T(a,b) \leq \Theta_T(c,d).$$

Proof (Corollary 1). Let us consider any two objects $a, b \in A$, such that aDb, and let us denote by D_2' the dominance relation over set $A \times A$, defined in the same way as the dominance relation D_2 over set B, with the only difference that B (appearing in the definition of D_2) is replaced by $A \times A$. First, let us observe that aDb implies that $(a,b)D_2'(b,a)$, and, moreover, given any object $c \in A \setminus \{a,b\}$, it is true that $(a,c)D_2'(b,c)$ and $(c,b)D_2'(c,a)$. Secondly, note that every decision rule $r_S \in R_S$ that covers the dominated (w.r.t. D_2') pair of objects (b,a) (respectively, (b,c), (c,a)), covers also the dominating (w.r.t. D_2') pair of objects (a,b) (respectively, (a,c), (c,b)). Analogously, every decision rule $r_{S^c} \in R_{S^c}$ that covers the dominating pair of objects (a,b) (respectively, (a,c), (c,b)), covers also the dominated pair of objects (b,a) (respectively, (b,c), (c,a)). Therefore, after application of decision rules on set A, according to definitions (2.24) and (2.25), or (2.26) and (2.27), or (2.28) and (2.29), or (2.30) and (2.31), we get:

- $\mathbb{S}(a,b) \geq \mathbb{S}(b,a)$ and $\mathbb{S}^c(b,a) \geq \mathbb{S}^c(a,b)$,
- $\mathbb{S}(a,c) \geq \mathbb{S}(b,c)$ and $\mathbb{S}^c(c,a) \geq \mathbb{S}^c(c,b)$,
- $\mathbb{S}(c,a) \leq \mathbb{S}(c,b)$ and $\mathbb{S}^c(a,c) \leq \mathbb{S}^c(b,c)$.

Thirdly, from (2.37), we get $\mathcal{R}(a,b) \geq \mathcal{R}(b,a)$, and, moreover, $\mathcal{R}(a,c) \geq \mathcal{R}(b,c)$, and $\mathcal{R}(c,a) \leq \mathcal{R}(c,b)$. This set of inequalities is the antecedent of the implication given in Definition 3, concerning property CC. Thus, from this definition, we have $a \succeq(A, \mathcal{R})\ b$. □

Proof (Properties of NFR).

(N) Satisfied according to [30, 36].

(M) Satisfied due to the definition of NFR, given by (2.49).

(CC) According to [36], this property is satisfied in case of exploitation of a crisp relation. However, it is evident that this property is also satisfied in general, i.e., when an exploited relation is valued.

($INDO$) According to [36], this property is satisfied in case of exploitation of a crisp relation. However, it is evident that this property is also satisfied in general, i.e., when an exploited relation is valued.

(IC) Satisfied according to [30, 36].

(O) Not satisfied since for a given finite set of objects A and for a valued relation R over A, NFR makes use of the "cardinal" properties of values $R(a,b)$, with $a,b \in A$ [36].

(C) Satisfied according to [36].

(F) Satisfied according to [36].

(DP) Satisfied according to [36].

(GF) Not satisfied according to [36].

(D) Satisfied. Due to the fact that NFR has property F, a total order relation R over A (with $|A|$ ranks) is not going to change after application of NFR, i.e., $NFR(A,R) = R$. □

Proof (Properties of $It.NFR$).

(N) According to [36], this property is satisfied in case of exploitation of a crisp relation. However, it is evident that this property is also satisfied in general, i.e., when an exploited relation is valued.

(M) Not satisfied, as shown by the following example. Consider set $A = \{a, b, c, d, e, f\}$, and valued relation R over A defined as: $R(a, d) = 0.5$, $R(b, c) = 0.5$, $R(c, a) = 1$, $R(c, e) = 1$, $R(d, b) = 1$, $R(d, f) = 0.5$, $R(x, x) = 1$ for all $x \in A$, and $R(x, y) = 0$ for the remaining pairs $(x, y) \in A \times A$. The ranking (total preorder) obtained for relation R is: $c \succ d \succ a, b, e, f$ (i.e., object c is the best, object d is second best, and the remaining objects are in the third equivalence class). Observe that we have $a \succeq(A, R)\ b$. Now, consider relation R' which is identical to R except that $R'(a, c) = 1$. Thus, object a is improved. However, the ranking (total preorder) obtained for relation R' is: $d \succ b, c \succ a, e, f$, i.e., it is not true that $a \succeq(A, R')\ b$.

(CC) According to [36], this property is satisfied in case of exploitation of a crisp relation. However, it is evident that this property is also satisfied in general, i.e., when an exploited relation is valued.

($INDO$) According to [36], this property is satisfied in case of exploitation of a crisp relation. However, it is evident that this property is also satisfied in general, i.e., when an exploited relation is valued.

(IC) Not satisfied, as shown by the following example. Consider set $A = \{a, b, c, d, e\}$, and two valued relations R, R' over A defined as:

- $R(a, b) = 1$, $R(b, c) = 1$, $R(c, a) = 1$, $R(b, e) = 0.5$, $R(c, d) = 1$, $R(x, x) = 1$ for all $x \in A$, and $R(x, y) = 0$ for the remaining pairs $(x, y) \in A \times A$,
- $R'(b, e) = 0.5$, $R'(c, d) = 1$, $R'(x, x) = 1$ for all $x \in A$, and $R'(x, y) = 0$ for the remaining pairs $(x, y) \in A \times A$.

Thus, R and R' are circuit-equivalent (R' is identical to R except for the circuit a-b-c-a of length 3, on which value $\epsilon = 1$ has been subtracted). If property IC would be satisfied, we would have $\succeq(A, R') = \succeq(A, R)$. However, the ranking (total preorder) obtained for relation R is: $c \succ a \succ b \succ e, d$ (i.e., object c is the best, object a is second best, object b is third best, and the remaining objects are in the fourth equivalence class), while the ranking (total preorder) obtained for relation R' is: $c \succ b \succ a, e, d$. Thus, we obtain that $\succeq(A, R') \neq \succeq(A, R)$.

(O) Not satisfied since for a given finite set of objects A and for a valued relation R over A, $It.NFR$ makes use of the "cardinal" properties of values $R(a, b)$, $a, b \in A$.

(C) Not satisfied, as shown by the following example. Consider set $A = \{a, b, c, d\}$, and the family of valued relations R^ϵ over A, with $\epsilon \in (0, 1]$, defined by Table A.1. For any $\epsilon \in (0, 1]$, we have $c \succeq(A, R^\epsilon)\ d$ (as object a is always chosen in the first

TABLE A.1: Considered family of valued relations

R^ϵ	a	b	c	d
a	–	1	1	0
b	1	–	0	0
c	0	1	–	1
d	0	$1-\epsilon$	1	–

iteration, and object c in the second iteration), while for relation R (obtained when $\epsilon = 0$), we have $d \succ(A, R)\ c$ (as objects a and d are chosen in the first iteration), which violates continuity.

(F) Satisfied according to [36].

(DP) Satisfied according to [36].

(GF) Not satisfied since NFR does not satisfy property GF, and the first equivalence classes of the total preorders produced by NFR and $It.NFR$ are the same.

(D) Satisfied. Due to the fact that $It.NFR$ has property F, a total order relation R over A (with $|A|$ ranks) is not going to change after application of $It.NFR$, i.e., $It.NFR(A, R) = R$. □

Proof (Properties of MiF).

(N) Satisfied according to [28, 36, 139].

(M) Satisfied. Note that given a finite set of objects A and a valued relation R over this set, objects from A are ranked according to their scores calculated by function mF (2.41). Thus, we have $a \succeq(A, R)\ b \Leftrightarrow mF(a, A, R) \geq mF(b, A, R)$. Then, if the value $R(a, c)$, for some $c \in A \setminus \{a\}$, is improved, the score of object a cannot decrease; the change of value $R(c, a)$, for some $c \in A \setminus \{a\}$ does not affect the score of object a. Moreover, if the value $R(b, d)$, for some $d \in A \setminus \{b\}$, is decreased, the score of object b cannot increase; the change of value $R(d, b)$, for some $d \in A \setminus \{a\}$, does not affect the score of object b. Thus, for any of the four considered changes of relation R, reflected by relation R', we have $\Big(a \succeq(A, R)\ b \Rightarrow a \succeq(A, R')\ b\Big)$.

(CC) Satisfied. Note that given a finite set of objects A and a valued relation R over this set, objects from A are ranked according to their scores calculated by function mF (2.41). If $R(a, b) \geq R(b, a)$ and for all $c \in A \setminus \{a, b\}$ there is $R(a, c) \geq R(b, c)$, then according to definition (2.41), we have $mF(a, A, R) \geq mF(b, A, R)$. It implies that $a \succeq(A, R)\ b$.

($INDO$) Not satisfied, as shown by the following example. Consider set $A = \{a, b, c, d\}$, and valued relation R over A defined as: $R(a, b) = 1$, $R(a, c) = 1$, $R(a, d) = 0.5$, $R(b, a) = 1$, $R(b, c) = 0.5$, $R(b, d) = 0.5$, $R(c, d) = 0.5$, $R(x, x) = 1$ for all $x \in A$, and $R(x, y) = 0$ for the remaining pairs $(x, y) \in A \times A$. Observe that object d is a non-discriminating object since $R(x, d) = 0.5$ and $R(d, x) = 0$, for $x \in A'$, where $A' = \{a, b, c\}$. We obtain $\succeq(A', R/A') = a \succ b \succ c$ (i.e., object a is the best, and object b is better than object c). This ranking is different than $\succeq(A, R)/A' = a, b, c$ (i.e., all three objects are in the first equivalence class).

(IC) Not satisfied when R is considered to be a general valued relation over set A [36]. Moreover, not satisfied also when R is considered to be a three-valued relation over A, as shown by the following example. Consider set $A = \{a, b, c\}$, and two valued relations R, R' over A defined as:

- $R(a, b) = 1$, $R(b, a) = 1$, $R(b, c) = 1$, $R(x, x) = 1$ for all $x \in A$, and $R(x, y) = 0$ for the remaining pairs $(x, y) \in A \times A$,
- $R'(b, c) = 1$, $R'(x, x) = 1$ for all $x \in A$, and $R'(x, y) = 0$ for the remaining pairs $(x, y) \in A \times A$.

Thus, R and R' are circuit-equivalent (R' is identical to R except for the circuit a-b-a of length 2, on which value $\epsilon = 1$ has been subtracted). If property IC would be satisfied, we would have $\succeq(A, R') = \succeq(A, R)$. However, the ranking (total preorder) obtained for relation R is: $b \succ a, c$ (i.e., object b is the best, and the remaining objects are in the second equivalence class), while the ranking (total preorder) obtained for relation R' is: a, b, c (i.e., all three objects are in the first and only equivalence class). Thus, we obtain that $\succeq(A, R') \neq \succeq(A, R)$.

(O) Satisfied according to [28, 35, 36, 139].

(C) Satisfied according to [28, 35, 36].

(F) Not satisfied according to [35, 36].

(DP) Satisfied. Due to transitivity of a crisp relation R over a given finite set of objects A, for any pair of objects $(a, b) \in R$ we have $mF(a, A, R) \geq mF(b, A, R)$. This implies that $a \succeq(A, R)\ b$. Thus, $R \subseteq \succeq(A, R)$.

(GF) Satisfied according to [35, 36].

(D)
- Not satisfied for any three-valued relation R over A when $|A| \geq 4$ – in such a case the final ranking is composed of maximum three ranks, as there are only

three possible values of scoring function mF (2.41), i.e., $mF(a, A, R) \in \{0, \frac{1}{2}, 1\}$ for all $a \in A$.

- Satisfied when R is a general valued relation over A. Let $\epsilon = \frac{1}{|A|}$, and let consider an arbitrary order $a_{\{1\}}, a_{\{2\}}, \dots, a_{\{|A|\}}$ of all objects from set A. Then, the following valued relation R over A:

$$R(a_{\{i\}}, a_{\{j\}}) = \begin{cases} 1, & \text{if } i = j \\ i\epsilon, & \text{otherwise} \end{cases}, \tag{1}$$

where $i, j = 1, \dots, |A|$, is such that $MiF(A, R)$ is a total order $(a_{\{|A|\}} \succ \dots \succ a_{\{1\}})$ over A. □

Proof (Properties of $It.MiF$).

(N) Obviously satisfied.

(M) Not satisfied, as shown by the following example. Consider set $A = \{a, b, c\}$, and valued relation R over A defined as: $R(a, b) = 0.5$, $R(b, a) = 0.5$, $R(c, a) = 1$, $R(c, b) = 1$, $R(b, c) = 0.5$, $R(x, x) = 1$ for all $x \in A$, and $R(x, y) = 0$ for the remaining pairs $(x, y) \in A \times A$. The ranking (total preorder) obtained for relation R is: $c \succ a, b$ (i.e., object c is the best, and objects a and b are in the second equivalence class). Observe that we have $a \succeq(A, R)\ b$. Now, consider relation R' which is identical to R except for the pair of objects (c, a), for which we have lower value $R'(c, a) = 0$. This difference between R and R' should not "negatively affect" object a. However, the ranking (total preorder) obtained for relation R' is: $b \succ a, c$. Thus, it is not true that $a \succeq(A, R')\ b$.

(CC) Satisfied. Observe that, given a finite set of objects A and a valued relation R over A, if object a "covers" object b w.r.t. set A, it is also true that object a "covers" object b w.r.t. any subset $A' \subseteq A$. Using the reasoning from the proof of property CC of the MiF ranking method, this implies that for any $A' \subseteq A$, $mF(a, A', R) \geq mF(b, A', R)$. Thus, in any iteration where both objects a and b are considered (i.e., $a, b \in A^i$, with $A^i \subseteq A$), it is impossible (since the choice among objects belonging to set A^i is made based on maximum score, and the score of each object $c \in A^i$ is given by $mF(c, A^i, R)$) that object b will we chosen while object a will not be chosen. Thus, $a \succeq(A, R)\ b$.

($INDO$) Not satisfied. Consider set A and valued relation R given in the proof of property $INDO$ of the MiF ranking method. We obtain $\succeq(A', R/A') = a \succ b \succ c$ (i.e., object a is the best, and object b is better than object c). This ranking is different than $\succeq(A, R)/A' = a, b, c$ (i.e., all three objects are in the first equivalence class).

(IC) Not satisfied since MiF does not satisfy property IC, and the first equivalence classes of the total preorders produced by MiF and $It.MiF$ are the same.

(O) Satisfied according to [35].

(C) Not satisfied according to [35].

(F) Satisfied according to [35].

(DP) Satisfied. Let us consider a finite set of objects A, a transitive and crisp relation R over A, and any pair of objects $(a, b) \in R$. Observe that due to transitivity of R, relation R/A' is also transitive, for any $A' \subseteq A$. This implies that in each i-th iteration, where a choice is made among objects belonging to set $A^i \subseteq A$, if the pair of objects (a, b) belongs to R/A^i, it is true that $mF(a, A^i, R) \geq mF(b, A^i, R)$. Thus, it is impossible (since the choice among objects belonging to set A^i is made based on maximum score, and the score of each object $c \in A^i$ is given by $mF(c, A^i, R)$) that object b will we chosen while object a will not be chosen. Therefore, after all iterations, we have to obtain that $a \succeq(A, R)\ b$. This implies that $R \subseteq \succeq(A, R)$.

(GF) Satisfied according to [35].

(D) Satisfied. Due to the fact that $It.MiF$ has property F, a total order relation R over A (with $|A|$ ranks) is not going to change after application of $It.MiF$, i.e., $It.MiF(A, R) = R$. □

Proof (Properties of L/E).

(N) Obviously satisfied.

(M) Satisfied according to [34].

(CC) Satisfied. Let us consider a finite set of objects A and a valued relation R over A. If $R(a, b) \geq R(b, a)$ and for all $c \in A \setminus \{a, b\}$ there is $R(a, c) \geq R(b, c)$ and $R(c, a) \leq R(c, b)$, then according to definitions (2.42) and (2.45), we have $SF(a, A, R) \geq SF(b, A, R)$ (in other words, object a has not smaller leaving flow than object b) and $-SA(a, A, R) \geq -SA(b, A, R)$ (in other words, object a has not greater entering flow than object b). Due to the definition of L/E (2.53), it implies that $a \succeq(A, R)\ b$.

$(INDO)$ Satisfied. Let us consider a finite set of objects A and a valued relation R over A. Then, each non-discriminating object $b \in A \setminus A'$, with $A' \subset A$, influences leaving and entering flow of each object $a \in A'$ in the same way. Precisely, leaving flow $SF(a, A', R)$ of each object $a \in A'$ increases (or decreases) by k while entering flow $-(-SA(a, A', R))$ of each object $a \in A'$ increases (or decreases) by k'.

(IC) Not satisfied, as shown by the following example. Consider set $A = \{a, b, c, d, e, f\}$, and two valued relations R, R' over A, defined as:

- $R(a,b) = 0.5$, $R(b,c) = 0.5$, $R(c,a) = 0.5$, $R(e,d) = 0.5$, $R(d,f) = 0.5$, $R(x,x) = 1$ for all $x \in A$, and $R(x,y) = 0$ for the remaining pairs $(x,y) \in A \times A$,
- $R'(a,b) = 1$, $R'(b,c) = 1$, $R'(c,a) = 1$, $R'(e,d) = 0.5$, $R'(d,f) = 0.5$, $R'(x,x) = 1$ for all $x \in A$, and $R'(x,y) = 0$ for the remaining pairs $(x,y) \in A \times A$.

Thus, R and R' are circuit-equivalent (R' is identical to R except for the circuit a-b-c-a of length 3, on which value $\epsilon = 0.5$ has been added). If property IC would be satisfied, we would have $\succeq(A,R') = \succeq(A,R)$. However, we have $d \succeq(A,R)\ b$ and $b \succeq(A,R)\ d$ (i.e., objects b and d are in the same equivalence class in case of exploitation of relation R), while not $d \succeq(A,R')\ b$ nor $b \succeq(A,R')\ d$ (i.e., objects b and d are incomparable in case of exploitation of relation R'). Thus, we obtain that $\succeq(A,R') \neq \succeq(A,R)$.

(O) Not satisfied since for a given finite set of objects A and for a valued relation R over A, L/E makes use of the "cardinal" properties of values $R(a,b)$, with $a, b \in A$.

(C) Obviously satisfied.

(F) Satisfied. Let us consider a finite set of objects A and a total preorder relation R over A. First, due to transitivity of R, given any pair of objects $(a,b) \in R$, object a has not smaller leaving flow and not greater entering flow than object b, i.e., $SF(a,A,R) \geq SF(b,A,R)$ and $-SA(a,A,R) \geq -SA(b,A,R)$, respectively. Thus, $a \succeq(A,R)\ b$. This means that $R \subseteq \succeq(A,R)$. Second, due to transitivity and completeness of R, given any pair of objects $(a,b) \notin R$, object a has smaller leaving flow and greater entering flow than object b. Therefore, it is not true that $a \succeq(A,R)\ b$. This means that $\neg R \subseteq \neg \succeq(A,R)$, where $\neg$ denotes complement of a set. Thus, $\succeq(A,R) = R$.

(DP) Satisfied as shown in the first part of the proof of property F above.

(GF) Satisfied. Let us consider a finite set of objects A and a crisp relation R over A. First, assume that the antecedent of the implication in the definition of property GF is true. Thus, $G(A,R) \neq \emptyset$. Second, due to the definition of L/E, given by (2.53), every object $a \in G(A, \succeq(A,R))$ has maximum leaving flow and minimum entering flow among all objects from set A. To have maximum leaving flow, each object $a \in G(A, \succeq(A,R))$ has to belong to set $G(A,R)$.

(D) Satisfied. Due to the fact that L/E has property F, a total order relation R over A (with $|A|$ ranks) is not going to change after application of L/E, i.e., $L/E(A,R) = R$.
$\square$

Bibliography

[1] Aamodt Agnar and Enric Plaza. Case-based reasoning: Foundational issues, methodological variations, and system approaches. *Artificial Intelligence Communications*, 7:39–52, 1994.

[2] David W. Aha, Kibler Dennis, and Albert Marc K. Instance-based learning algorithms. *Machine Learning*, 6:37–66, 1991.

[3] Nir Ailon and Mehryar Mohri. An efficient reduction of ranking to classification. In Rocco Servedio and Tong Zhang, editors, *Proceedings of the 21st Annual Conference on Learning Theory, Helsinki, Finland, 9-12 July 2008*, pages 87–97, 2008.

[4] Kenneth J. Arrow. *Social Choice and Individual Values.* Yale University Press, 1951.

[5] Kenneth J. Arrow and Herve Raynaud. *Social choice and multicriterion decision making.* MIT Press, Cambridge, 1986.

[6] Carlos A. Bana e Costa and Jean-Claude Vansnick. A critical analysis of the eigenvalue method used to derive priorities in AHP. *European Journal of Operational Research*, 187:1422–1428, 2008.

[7] C. Richard Barrett, Prasanta K. Pattanaik, and Maurice Salles. On choosing rationally when preferences are fuzzy. *Fuzzy Sets and Systems*, 34:197–212, 1990.

[8] Jan G. Bazan. A comparison of dynamic and non-dynamic rough set methods for extracting laws from decision tables. In Lech Polkowski and Skowron Andrzej, editors, *Rough Sets in Knowledge Discovery*, volume 1, chapter 17, pages 321–365. Physica-Verlag, Heidelberg, 1998.

[9] Jan G. Bazan, Hung Son Nguyen, Sinh Hoa Nguyen, Piotr Synak, and Jakub Wróblewski. Rough set algorithms in classification problem. In Lech Polkowski, Shusaku Tsumoto, and Tsau Y. Lin, editors, *Rough Set Methods and Applications*, chapter 2, pages 49–88. Physica-Verlag, Heidelberg, 2000.

[10] Majid Behzadian, R.B. Kazemzadeh, A. Albadvi, and M. Aghdasi. PROMETHEE: A comprehensive literature review on methodologies and applications. *European Journal of Operational Research*, 200(1):198–215, 2010.

[11] Nuel Belnap. How a computer should think. In *Proc. of the Oxford International Symposium on Contemporary Aspects of Philosophy*, pages 30–56, 1976.

[12] Nuel Belnap. A useful four-valued logic. In J. M. Dunn and G. Epstein, editors, *Modern Uses of Multiple Valued Logics*, pages 8–37. D. Reidel, Dordrecht, 1977.

[13] Valerie Belton and Theodor Stewart. *Multiple Criteria Decision Analysis: An Integrated Approach*. Kluwer, Dordrecht, 2002.

[14] Arie Ben-David. Automatic generation of symbolic multiattribute ordinal knowledgebased dss: Methodology and applications. *Decision Sciences*, 23:1357–1372, 1992.

[15] Arie Ben-David. Monotonicity maintenance in information-theoretic machine learning algorithms. *Machine Learning*, 19(1):29–43, 1995.

[16] Axel E. Bernal, Karen Hospevian, Tayfun Karadeniz, and Jean-Louis Lassez. Similarity based classification. In M.R. Berthold et al., editors, *IDA 2003*, volume 2810 of *Lecture Notes in Computer Science*, pages 187–197, Berlin, 2003. Springer.

[17] Michel Beuthe and Giuseppe Scannella. Comparative analysis of UTA multicriteria methods. *European Journal of Operational Research*, 130(2):246–262, 2001.

[18] Raymond Bisdorff. Logical foundation of fuzzy preferential systems with application to the electre decision aid methods. *Computers & Operations Research*, 27:673–687, 2000.

[19] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[20] Jerzy Błaszczyński. *Rule Models for Ordinal Classification in Variable Consistency Rough Set Approaches*. PhD thesis, Poznań University of Technology, 2010.

[21] Jerzy Błaszczyński, Salvatore Greco, and Roman Słowiński. Multi-criteria classification – a new scheme for application of dominance-based decision rules. *European Journal of Operational Research*, 181(3):1030–1044, 2007.

[22] Jerzy Błaszczyński, Salvatore Greco, Roman Słowiński, and Marcin Szeląg. On variable consistency dominance-based rough set approaches. In S. Greco, Y. Hata, S. Hirano, M. Inuiguchi, S. Miyamoto, H. S. Nguyen, and R. Słowiński, editors, *Rough Sets and Current Trends in Computing 2006*, volume 4259 of *Lecture Notes in Artificial Intelligence*, pages 191–202, Berlin Heidelberg, 2006. Springer.

[23] Jerzy Błaszczyński, Salvatore Greco, Roman Słowiński, and Marcin Szeląg. Monotonic variable consistency rough set approaches. In J. Yao, P. Lingras, W. Wu,

M. Szczuka, N. J. Cercone, and Dominik Ślęzak, editors, *Rough Sets and Knowledge Technology 2007*, volume 4481 of *Lecture Notes in Artificial Intelligence*, pages 126–133, Berlin Heidelberg, 2007. Springer.

[24] Jerzy Błaszczyński, Salvatore Greco, Roman Słowiński, and Marcin Szeląg. Monotonic variable consistency rough set approaches. *International Journal of Approximate Reasoning*, 50(7):979–999, 2009.

[25] Jerzy Błaszczyński, Roman Słowiński, and Marcin Szeląg. Probabilistic rough set approaches to ordinal classification with monotonicity constraints. In Eyke Hüllermeier, Rudolf Kruse, and F. Hoffmann, editors, *IPMU 2010*, volume 6178 of *Lecture Notes in Artificial Intelligence*, pages 99–108, Berlin Heidelberg, 2010. Springer.

[26] Jerzy Błaszczyński, Roman Słowiński, and Marcin Szeląg. Sequential covering rule induction algorithm for variable consistency rough set approaches. *Information Sciences*, 181:987–1002, 2011.

[27] Geraldine Bous, Philippe Fortemps, Francois Glineur, and Marc Pirlot. ACUTA: A novel method for eliciting additive value functions on the basis of holistic preference statements. *European Journal of Operational Research*, 206(2):435–444, 2010.

[28] Denis Bouyssou. A note on the "min in favor" ranking method for valued preference relations. In M. Cerny, D. Glükaufová, and D. Loula, editors, *Multicriteria Decision Making. Methods – Algorithms – Applications*, pages 16–25. Czechoslovak Academy of Sciences, Prague, 1992.

[29] Denis Bouyssou. A note on the sum of differences choice function for fuzzy preference relations. *Fuzzy Sets and Systems*, 47:197–202, 1992.

[30] Denis Bouyssou. Ranking methods based on valued preference relations: A characterization of the net flow method. *European Journal of Operational Research*, 60:61–67, 1992.

[31] Denis Bouyssou. A note on the 'min in favor' choice procedure for fuzzy preference relations. In P.M. Pardalos, Y. Siskos, and C. Zopounidis, editors, *Advances in Multicriteria Analysis*, pages 9–16. Kluwer, 1995.

[32] Denis Bouyssou. Outranking relations: do they have special properties? *Journal of Multi-Criteria Decision Analysis*, 5:99–111, 1996.

[33] Denis Bouyssou. Monotonicity of 'ranking by choosing': A progress report. *Social Choice and Welfare*, 23:249–273, 2004.

[34] Denis Bouyssou and Patrice Perny. Ranking methods for valued preference relations: A characterization of a method based on leaving and entering flows. *European Journal of Operational Research*, 61:186–194, 1992.

[35] Denis Bouyssou and Marc Pirlot. Choosing and ranking on the basis of fuzzy preference relations with the "min in favour". In G. Fandel and T. Gal, editors, *Multiple Criteria Decision Making - Proceedings of the Twelfth International Conference Hagen (Germany)*, pages 115–127. Springer, 1997.

[36] Denis Bouyssou and Philippe Vincke. Ranking alternatives on the basis of preference relations: A progress report with special emphasis on outranking relations. *Journal of Multi-Criteria Decision Analysis*, 6:77–85, 1997.

[37] Jean Pierre Brans and Bertrand Mareschal. PROMETHEE methods. In Jose Figueira, Salvatore Greco, and Matthias Ehrgott, editors, *Multiple Criteria Decision Analysis: State of the Art Surveys*, Operations Research & Management Science, pages 163–196. Springer, New York, 2005.

[38] Jean Pierre Brans, Bertrand Mareschal, and Philippe Vincke. Promethee: a new family of outranking methods in multicriteria analysis. In J.P. Brans, editor, *Operational Research, IFORS 84*, pages 477–490. North Holland, Amsterdam, 1984.

[39] Percy Williams Bridgman. *Dimensional Analysis*. Yale University Press, 1922.

[40] David S. Broomhead and David Lowe. Radial basis functions, multi-variable functional interpolation and adaptive networks. Technical Report 4148, Royal Signals and Radar Establishment, 1988.

[41] Yihua Chen, Eric K. Garcia, and Maya R. Gupta. Similarity-based classification: Concepts and algorithms. *Journal of Machine Learning Research*, 10:747–776, 2009.

[42] William W. Cohen, Robert E. Schapire, and Yoram Singer. Learning to order things. In *Advances in Neural Information Processing Systems*, volume 10, pages 451–457. MIT, 1998.

[43] William W. Cohen, Robert E. Schapire, and Yoram Singer. Learning to order things. *Journal of Artificial Intelligence Research*, 10:243–270, 1999.

[44] Salvatore Corrente, Salvatore Greco, Miłosz Kadziński, and Roman Słowiński. Robust ordinal regression in preference learning and ranking. *Machine Learning*, 93:381–422, 2013.

[45] Hennie Daniels and B. Kamp. Applications of mlp networks to bond rating and house pricing. *Neural Computation and Applications*, 8:226–234, 1999.

[46] Krzysztof Dembczyński, Wojciech Kotłowski, Roman Słowiński, and Marcin Szeląg. Learning of rule ensembles for multiple attribute ranking problems. In Johannes Fürnkranz and Eyke Hüllermeier, editors, *Preference Learning*, pages 217–247. Springer, Berlin, 2010.

[47] Krzysztof Dembczyński, Roman Pindur, and Robert Susmaga. Generation of exhaustive set of rules within dominance-based rough set approach. *Electronic Notes in Theoretical Computer Science*, 82(4):96–107, 2003.

[48] Janez Demsar. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7:1–30, 2006.

[49] Luis C. Dias and Claude Lamboray. Extensions of the prudence principle to exploit a valued outranking relation. *European Journal of Operational Research*, 201(3):828–837, March 2010.

[50] Michael Doumpos and Constantin Zopounidis. Preference disaggregation and statistical learning for multicriteria decision support: a review. *European Journal of Operational Research*, 209(3):203–214, 2012.

[51] Didier Dubois, Henri Prade, Francesca Esteva, Pere Garcia, Lluis Godo, and Ramon Lopez de Mantaras. Fuzzy set modelling in case-based reasoning. *International Journal of Intelligent Systems*, 13:345–373, 1998.

[52] Peter H. Farquhar. State of the art – utility assessment methods. *Management Science*, 30(11):1283–1300, 1984.

[53] José Figueira, Salvatore Greco, and Matthias Ehrgott, editors. *Multiple Criteria Decision Analysis: State of the Art Surveys.* Operations Research & Management Science. Springer, New York, 2005.

[54] José Figueira, Salvatore Greco, Bernard Roy, and Roman Słowiński. ELECTRE methods: Main features and recent developments. In P. Pardalos, D. Hearn, and C. Zopounidis, editors, *Handbook of Multicriteria Analysis*, volume 103 of *Applied Optimization*, pages 51–89. Springer, 2010.

[55] José Figueira, Salvatore Greco, Bernard Roy, and Roman Słowiński. An overview of ELECTRE methods and their recent extensions. *Journal of Multi-Criteria Decision Analysis*, 20(1-2):61–85, 2013.

[56] José Figueira, Salvatore Greco, and Roman Słowiński. Building a set of additive value functions representing a reference preorder and intensities of preference: Grip method. *European Journal of Operational Research*, 195(2):460–486, 2009.

[57] José Figueira, Vincent Mousseau, and Bernard Roy. ELECTRE methods. In José Figueira, Salvatore Greco, and Matthias Ehrgott, editors, *Multiple Criteria Decision Analysis: State of the Art Surveys*, Operations Research & Management Science, pages 133–162. Springer, New York, 2005.

[58] Peter C. Fishburn. *Additive Utilities with Incomplete Product Set: Applications to Priorities and Assignments*. Operations Research Society of America (ORSA), 1967.

[59] Peter C. Fishburn. *The Theory of Social Choice*. Princeton University Press, 1973.

[60] Branden Fitelson. *Studies in Bayesian Confirmation Theory*. PhD thesis, University of Wisconsin-Madison, 2001.

[61] Evelyn Fix and J. L. Hodges. Discriminatory analysis, nonparametric discrimination: Consistency properties. Technical report 4, USAF School of Aviation Medicine, Randolph Field, Texas, 1951.

[62] Philippe Fortemps, Salvatore Greco, and Roman Słowiński. Multicriteria choice and ranking using decision rules induced from rough approximation of graded preference relations. In S. Tsumoto, Roman Słowiński, J. Komorowski, and J. W. Grzymała-Busse, editors, *Rough Sets and Current Trends in Computing 2004*, volume 3066 of *Lecture Notes in Artificial Intelligence*, pages 510–522, Berlin Heidelberg, 2004. Springer.

[63] Philippe Fortemps, Salvatore Greco, and Roman Słowiński. Multicriteria decision support using rules that represent rough-graded preference relations. *European Journal of Operational Research*, 188(1):206–223, 2008.

[64] Philippe Fortemps and Roman Słowiński. A graded quadrivalent logic for ordinal preference modelling: Loyola-like approach. *Fuzzy Optimization and Decision Making*, 1(1):93–111, 2002.

[65] Yoav Freund, Raj Iyer, Robert E. Schapire, and Yoram Singer. An efficient boosting algorithm for combining preferences. *Journal of Machine Learning Research*, 6(4):933–969, 2003.

[66] Yoav Freund and Robert E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139, 1997.

[67] Jerome Friedman. Recent advances in predictive (machine) learning. *Journal of Classification*, 23:175–197, 2006.

[68] Milton Friedman. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, 32:675–701, 1937.

[69] Milton Friedman. A comparison of alternative tests of significance for the problem of m rankings. *Annals of Mathematical Statistics*, 11:86–92, 1940.

[70] Hela Frikha, Habib Chabchoub, and Jean Marc Martel. Inferring criteria's relative importance coefficients in PROMETHEE II. *International Journal of Operational Research*, 7(2):257–275, 2010.

[71] Johannes Fürnkranz and Eyke Hüllermeier, editors. *Preference Learning*. Springer, Berlin, 2010.

[72] Johannes Fürnkranz and Eyke Hüllermeier, editors. *Preference Learning: An Introduction*, pages 3–8. Springer, 2010.

[73] Itzhak Gilboa and David Schmeidler. Case-based decision theory. *The Quarterly Journal of Economics*, 110(3):605–639, August 1995.

[74] Itzhak Gilboa and David Schmeidler. *A Theory of Case-Based Decisions*. Number 9780521003117 in Cambridge Books. Cambridge University Press, February 2001.

[75] Silvio Giove, Salvatore Greco, Benedetto Matarazzo, and Roman Słowiński. Variable consistency monotonic decision trees. In James J. Alpigini, James F. Peters, Andrzej Skowron, and Ning Zhong, editors, *Rough Sets and Current Trends in Computing (RSCTC 2002)*, volume 2475 of *Lecture Notes in Artificial Intelligence*, pages 247–254. Springer, Malvern, USA, 2002.

[76] Salvatore Greco, Miłosz Kadziński, Vincent Mousseau, and Roman Słowiński. Electre^{GKMS}: Robust ordinal regression for outranking methods. *European Journal of Operational Research*, 214(10):118–135, 2011.

[77] Salvatore Greco, Benedetto Matarazzo, and Roman Słowiński. Rough set approach to multi-attribute choice and ranking problems. Technical Report 38/95, ICS, Warsaw University of Technology, Warsaw, Poland, 1995.

[78] Salvatore Greco, Benedetto Matarazzo, and Roman Słowiński. Rough approximation of a preference relation by dominance relations. In S. Tsumoto, S. Kobayashi, T. Yokomori, H. Tanaka, and A. Nakamura, editors, *Proc. of the Fourth Internat. Workshop on Rough Sets, Fuzzy Sets and Machine Discovery (RSFD'96)*, pages 125–130. Tokyo University Press, 1996.

[79] Salvatore Greco, Benedetto Matarazzo, and Roman Słowiński. Rough approximation of preference relation by dominance relations. ICS Research Report 16/96, Warsaw University of Technology, Warsaw, 1996.

[80] Salvatore Greco, Benedetto Matarazzo, and Roman Słowiński. Rough set approach to multi-attribute choice and ranking problems. In G. Fandel and T. Gal, editors, *Multiple Criteria Decision Making, Proceedings of the Twelfth International Conference, Hagen, Germany*, pages 318–329, Berlin, 1997. Springer.

[81] Salvatore Greco, Benedetto Matarazzo, and Roman Słowiński. Rough approximation of a preference relation in a pairwise comparison table. In Polkowski Lech et al., editors, *Rough Sets in Knowledge Discovery*, volume 2, chapter 2, pages 13–36. Springer, Berlin Heidelberg, 1998.

[82] Salvatore Greco, Benedetto Matarazzo, and Roman Słowiński. Rough approximation of a preference relation by dominance relations. *European Journal of Operational Research*, 117:63–83, 1999.

[83] Salvatore Greco, Benedetto Matarazzo, and Roman Słowiński. The use of rough sets and fuzzy sets in MCDM. In Tomas Gal, Theodor J. Stewart, and Thomas Hanne, editors, *Multicriteria Decision Making: Advances in MCDM models, Algorithms, Theory, and Applications*, volume 21 of *International Series in Operations Research & Management Science*, chapter 14, pages 14.1–14.59. Kluwer, Dordrecht, 1999.

[84] Salvatore Greco, Benedetto Matarazzo, and Roman Słowiński. Rough sets theory for multicriteria decision analysis. *European Journal of Operational Research*, 129(1):1–47, 2001.

[85] Salvatore Greco, Benedetto Matarazzo, and Roman Słowiński. Axiomatic characterization of a general utility function and its particular cases in terms of conjoint measurement and rough-set decision rules. *European Journal of Operational Research*, 158(2):271–292, 2004.

[86] Salvatore Greco, Benedetto Matarazzo, and Roman Słowiński. Decision rule approach. In Jose Figueira, Salvatore Greco, and Matthias Ehrgott, editors, *Multiple Criteria Decision Analysis: State of the Art Surveys*, Operations Research & Management Science, chapter 13, pages 507–562. Springer, New York, 2005.

[87] Salvatore Greco, Benedetto Matarazzo, and Roman Słowiński. Preference representation by means of conjoint measurement and decision rule model. In Denis Bouyssou, Eric Jacquet-Lagrèze, Patrice Perny, Roman Słowiński, Daniel Vanderpooten, and Philippe Vincke, editors, *Aiding Decisions with Multiple Criteria - Essays in Honor of Bernard Roy*, pages 263–313. Kluwer, Boston, 2005.

[88] Salvatore Greco, Benedetto Matarazzo, and Roman Słowiński. Dominance-based rough set approach to case-based reasoning. In V. Torra, Y. Narukawa, A. Valls, and J. Domingo-Ferrer, editors, *Modelling Decisions for Artificial Intelligence*, volume 3885 of *Lecture Notes in Artificial Intelligence*, pages 7–18, Berlin Heidelberg, 2006. Springer.

[89] Salvatore Greco, Benedetto Matarazzo, and Roman Słowiński. Case-based reasoning using gradual rules induced from dominance-based rough approximations. In

G. Wang, T. Li, J. W. Grzymała-Busse, D. Miao, Andrzej Skowron, and Yiyu Y. Yao, editors, *Rough Sets and Knowledge Technology (RSKT 2008)*, volume 5009 of *Lecture Notes in Artificial Intelligence*, pages 268–275, Berlin, 2008. Springer.

[90] Salvatore Greco, Benedetto Matarazzo, and Roman Słowiński. Dominance-based rough set approach to interactive multiobjective optimization. In Juergen Branke, Kalyanmoy Deb, Kaisa Miettinen, and Roman Słowiński, editors, *Multiobjective Optimization: Interactive and Evolutionary Approaches*, volume 5252 of *Lecture Notes in Computer Science*, chapter 5, pages 121–156. Springer, Berlin, 2008.

[91] Salvatore Greco, Benedetto Matarazzo, and Roman Słowiński. Granular computing for reasoning about ordered data: the dominance-based rough set approach. In Witold Pedrycz, Andrzej Skowron, and Vladik Kreinovich, editors, *Handbook of Granular Computing*, chapter 15, pages 347–373. John Wiley and Sons, Chichester, 2008.

[92] Salvatore Greco, Benedetto Matarazzo, Roman Słowiński, and Jerzy Stefanowski. An algorithm for induction of decision rules consistent with the dominance principle. In Wojciech Ziarko and Yiyu Y. Yao, editors, *Rough Sets and Current Trends in Computing 2001*, volume 2005 of *Lecture Notes in Artificial Intelligence*, pages 304–313, Berlin, 2001. Springer.

[93] Salvatore Greco, Benedetto Matarazzo, Roman Słowiński, and Alexis Tsoukiàs. Exploitation of a rough approximation of the outranking relation in multicriteria choice and ranking. In Theodor J. Stewart and Robin C. van den Honert, editors, *Trends in Multicriteria Decision Making*, volume 465 of *Lecture Notes in Economics and Mathematical Systems*, pages 45–60, Berlin, 1998. Springer.

[94] Salvatore Greco, Vincent Mousseau, and Roman Słowiński. Ordinal regression revisited: multiple criteria ranking using a set of additive value functions. *European Journal of Operational Research*, 191(2):415–435, 2008.

[95] Salvatore Greco, Zdzisław Pawlak, and Roman Słowiński. Can Bayesian confirmation measures be useful for rough set decision rules? *Engineering Applications of Artificial Intelligence*, 17:345–361, 2004.

[96] Salvatore Greco, Roman Słowiński, Jose Figueira, and Vincent Mousseau. Robust ordinal regression. In Matthias Ehrgott, Jose Figueira, and Salvatore Greco, editors, *Trends in Multiple Criteria Decision Analysis*, International Series in Operations Research & Management Science, chapter 9, pages 241–283. Springer, New York, 2010.

[97] Salvatore Greco, Roman Słowiński, Benedetto Matarazzo, and Jerzy Stefanowski. Variable consistency model of dominance-based rough sets approach. In Wojciech

Ziarko and Yiyu Y. Yao, editors, *Rough Sets and Current Trends in Computing 2001*, volume 2005 of *Lecture Notes in Artificial Intelligence*, pages 170–181, Berlin, 2001. Springer.

[98] Paul Hansen and Franz Ombler. A new method for scoring additive multi-attribute value models using pairwise rankings of alternatives. *Journal of Multi-Criteria Decision Analysis*, 15:87–107, 2009.

[99] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer, 2001.

[100] Dominique Henriet. The copeland choice function – an axiomatic characterization. *Social Choice and Welfare*, 2:49–64, 1985.

[101] Ralf Herbrich, Thore Graepel, and Klaus Obermayer. Regression models for ordinal data: A machine learning approach. Technical report TR-99/03, TU Berlin, 1999.

[102] Ralf Herbrich, Thore Graepel, and Klaus Obermayer. *Large Margin Rank Boundaries for Ordinal Regression*, chapter 7, pages 115–132. MIT Press, January 2000.

[103] Jean-Christophe Hugonnard and Bernard Roy. Le plan d'extension du métro en banlieue parisienne, un cas type d'application de l'analyse multicritère. *Les Cahiers Scientifiques de la Revue Transports*, 6(6):77–108, 1982. 1er trimestre.

[104] David Hume. *An Enquiry Concerning Human Understanding*. Clarendon Press, Oxford, 1748.

[105] Ching-Lai Hwang, Young-Jou Lai, and Ting-Yun Liu. A new approach for multiple objective decision making. *Computers and Operational Research*, 20:889–899, 1993.

[106] Ching-Lai Hwang and Kwangsun Yoon. *Multiple Attribute Decision Making: Methods and Applications*. Springer, New York, 1981.

[107] Ronald L. Iman and James M. Davenport. Approximations of the critical region of the friedman statistic. *Communications in Statistics*, pages 571–595, 1980.

[108] Masahiro Inuiguchi and Yukihiro Yoshioka. Variable-precision dominance-based rough set approach. In S. Greco, Y. Hata, S. Hirano, M. Inuiguchi, S. Miyamoto, H. S. Nguyen, and R. Słowiński, editors, *Rough Sets and Current Trends in Computing 2006*, volume 4259 of *Lecture Notes in Artificial Intelligence*, pages 203–212, Berlin Heidelberg, 2006. Springer.

[109] Eric Jacquet-Lagrèze and Yannis Siskos. Assessing a set of additive utility functions for multicriteria decision making: The UTA method. *European Journal of Operational Research*, 10:151–164, 1982.

[110] Eric Jacquet-Lagrèze and Yannis Siskos. Preference disaggregation: 20 years of mcda experience. *European Journal of Operational Research*, 130(2):233–245, 2001.

[111] Andrzej Janusz and Dominik Ślęzak. An unsupervised model for rule-based similarity learning from textual data. In Marcin Szczuka et al., editors, *Proceedings of the international Workshop Concurrency, Specification and Programming 2011*, pages 229–237, Pułtusk, Poland, 2011.

[112] Thorsten Joachims. Optimizing search engines using clickthrough data. In *Proceedings of the International Conference on Knowledge Discovery and Data Mining (ACM SIGKDD 2002)*, pages 133–142, Alberta, Canada, 2002. ACM.

[113] Thorsten Joachims. Training linear SVMs in linear time. In *Proceedings of the ACM Conference on Knowledge Discovery and Data Mining (KDD)*, 2006.

[114] Miłosz Kadziński. *New directions in Robustness Analysis and Preference Modeling in Multiple Criteria Decision Aiding.* NAKOM, 2012.

[115] Miłosz Kadziński, Salvatore Greco, and Roman Słowiński. Selection of a representative value function in robust multiple criteria ranking and choice. *European Journal of Operational Research*, 217(3):541–553, 2012.

[116] Miłosz Kadziński, Roman Słowiński, and Marcin Szeląg. Dominance-based rough set approach to multiple criteria ranking with sorting-specific preference information. In Stan Matwin and Jan Mielniczuk, editors, *Challenges in Computational Statistics and Data Mining*, volume 605 of *Studies in Computational Intelligence*. Springer, to appear.

[117] Ralph L. Keeney and Howard Raiffa. *Decisions with multiple objectives: Preferences and value tradeoffs.* Wiley, 1976.

[118] Janet Kolodner. *Case-Based Reasoning.* Morgan Kaufmann, San Mateo, 1993.

[119] Gary Koop. *Analysis of Economic Data.* Wiley, Chichester, 2000.

[120] M. Kostkowski and Roman Słowiński. UTA+ application (v. 1.20) – User's manual. Technical report, Université de Paris-Dauphine, Paris, 1996.

[121] Pat Langley and Herbert A. Simon. Fielded applications of machine learning. In Ryszard S. Michalski, Ivan Bratko, and Miroslav Kubat, editors, *Machine Learning and Data Mining*, pages 113–129. Wiley, New York, 1998.

[122] Yan Li, Simon Chi Keung Shiu, Sankar Kumar Pal, and James N.-K. Liu. A rough set-based case-based reasoner for text categorization. *International Journal of Approximate Reasoning*, 41:229–255, 2006.

[123] Tie-Yan Liu. *Learning to Rank for Information Retrieval.* Springer, Berlin, 2011.

[124] James G. March. Bounded rationality, ambiguity, and the engineering of choice. In David E. Bell, Howard Raiffa, and Amos Tversky, editors, *Decision Making, Descriptive, Normative and Prescriptive Interactions*, pages 33–58. Cambridge University Press, New York, 1988.

[125] Jean-Marc Martel and Bernard Roy. Analyse de la signifiance de diverses procedures d'agregation multicritere. *INFOR*, 44(3):191–214, 2006.

[126] Ryszard S. Michalski. A theory and methodology of inductive learning. In Ryszard S. Michalski, Jaime G. Carbonell, and Tom M. Mitchell, editors, *Machine Learning: An Artificial Intelligence Approach*, pages 83–129. Tioga Publishing, Palo Alto, 1983.

[127] Dawid W. Miller and Martin K. Starr. *Executive Decisions and Operations Research.* Prentice-Hall, Inc, 1969.

[128] Tom Mitchell. *Machine Learning.* McGraw Hill, 1997.

[129] Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of Machine Learning.* Adaptive Computation and Machine Learning. MIT Press, 2012.

[130] Vincent Mousseau and Roman Słowiński. Inferring an ELECTRE TRI model from assignment examples. *Journal of Global Optimization*, 12(2):157–174, 1998.

[131] Peter B. Nemenyi. *Distribution-free multiple comparisons.* PhD thesis, Princeton University, 1963.

[132] Quang Nguyen, Hamed Valizadegan, and Milos Hauskrecht. Learning classification models with soft-label information. *Journal of the American Medical Informatics Association*, 21:501–508, 2014.

[133] Włodzimierz Ogryczak and Rudolf Vetschera, editors. *Methodological foundations of multi-criteria decision making*, volume 158(2) of *European Journal of Operational Research*, pages 267–532. Springer, 2004.

[134] Yen-Jen Oyang, Shien-Ching Hwang, Yu-Yen Ou, Chien-Yu Chen, and Zhi-Wei Chen. Data classification with radial basis function networks based on a novel kernel density estimation algorithm. *IEEE Transactions on Neural Networks*, 16(1):225–236, 2005.

[135] Zdzisław Pawlak. *Rough Sets. Theoretical Aspects of Reasoning about Data.* Kluwer, Dordrecht, Netherlands, 1991.

[136] Zdzisław Pawlak and Andrzej Skowron. Rough membership functions. In Ronald R. Yager, Mario Fedrizzi, and Janusz Kacprzyk, editors, *Advances in the Dampster-Shafer Theory of Evidence*, pages 251–271. John Wiley & Sons, Inc., New York, 1994.

[137] Patrice Perny and Bernard Roy. The use of fuzzy outranking relations in preference modelling. *Fuzzy Sets and Systems*, 49:33–53, 1992.

[138] Patrice Perny and Alexis Tsoukias. On the continuous extension of a four valued logic for preference modelling. In *Proceedings of the IPMU'98 conference*, pages 302–309, 1998.

[139] Marc Pirlot. A characterization of 'min' as a procedure for exploiting valued preference relations and related results. *Journal of Multi-Criteria Decision Analysis*, 4:37–56, 1995.

[140] Filip Radlinski and Thorsten Joachims. Learning to rank from implicit feedback. In *Proceedings of the ACM Conference on Knowledge Discovery and Data Mining (KDD-05)*, pages 239–248, 2005.

[141] Bernard Roy. La méthode ELECTRE II. Technical report, METRA, Direction Scientifique, Note de Travail n. 142, 1971.

[142] Bernard Roy. ELECTRE III : Un algorithme de classements fondé sur une représentation floue des préférences en présence de critères multiples. *Cahiers du CERO*, 20(1):3–24, 1978.

[143] Bernard Roy. *Méthodologie Multicritère d'Aide à la Décision*. Economica, Paris, 1985.

[144] Bernard Roy. The outranking approach and the foundation of ELECTRE methods. *Theory and Decision*, 31:49–73, 1991.

[145] Bernard Roy. *Multicriteria Methodology for Decision Aiding*. Kluwer, Dordrecht, 1996.

[146] Bernard Roy and Denis Bouyssou. *Aide Multicritère à la Décision: Méthodes et Cas*. Economica, Paris, 1993.

[147] Bernard Roy and Jean-Michel Skalka. ELECTRE IS: Aspects méthodologiques et guide d'utilisation. Document du LAMSADE 30, Université Paris-Dauphine, Paris, 1984.

[148] Bernard Roy and Roman Słowiński. Questions guiding the choice of a multicriteria decision aiding method. *EURO Journal on Decision Processes*, 1(1):69–97, 2013.

[149] Ariel Rubinstein. Ranking the participants in a tournament. *SIAM Journal of Applied Mathematics*, 38:108–111, 1980.

[150] Thomas Saaty. *The Analytic Hierarchy Process.* McGraw Hill, New York, 1980.

[151] Thomas L. Saaty. *Fundamentals of Decision Making and Priority Theory.* RWS Publications, Pittsburgh, Pennsylvania, 2001.

[152] Roger Schank. *Dynamic Memory: A Theory of Learning in Computers and People.* Cambridge University Press, 1982.

[153] Robert E. Schapire and Yoram Singer. Improved boosting algorithms using confidence-rated predictions. *Machine Learning*, 37(3):297–336, 1999.

[154] Yannis Siskos, Evangelos Grigoroudis, and Nikolaos F. Matsatsinis. UTA methods. In J. Figueira, S. Greco, and M. Ehrgott, editors, *Multiple Criteria Decision Analysis: State of the Art Surveys*, pages 297–344. Springer, Boston, 2005.

[155] Yannis Siskos and D. Yanacopoulos. UTA STAR - an ordinal regression method for building additive value functions. *Investigacao Operational*, 5:39–53, 1985.

[156] Dominik Ślęzak. Rough Sets and Bayes Factor. In James F. Peters and Andrzej Skowron, editors, *Transactions on Rough Sets III*, volume 3400 of *Lecture Notes in Computer Science*, pages 202–229. Springer, Berlin, 2005.

[157] Paul Slovic. Choice between equally-valued alternatives. *Journal of Experimental Psychology: Human Perception Performance*, 1(3):280–287, 1975.

[158] Roman Słowiński, Salvatore Greco, and Benedetto Matarazzo. Mining decision-rule preference model from rough approximation of preference relation. In *Proc. 26th IEEE Annual Int. Conference on Computer Software and Applications (COMPSAC 2002)*, pages 1129–1134, Los Alamitos, CA, 2002. IEEE Computer Society Press.

[159] Roman Słowiński, Salvatore Greco, and Benedetto Matarazzo. Rough sets in decision making. In R.A. Meyers, editor, *Encyclopedia of Complexity and Systems Science*, pages 7753–7786. Springer, New York, 2009.

[160] Roman Słowiński, Salvatore Greco, and Benedetto Matarazzo. Rough set based decision support. In Edmund K. Burke and Graham Kendall, editors, *Search Methodologies: Introductory Tutorials in Optimization and Decision Support Techniques*, chapter 19, pages 557–609. Springer, New York, 2nd edition, 2014.

[161] Jerzy Stefanowski. *Algorytmy Indukcji Reguł Decyzyjnych w Odkrywaniu Wiedzy*, volume 361 of *Rozprawy.* Wydawnictwo Politechniki Poznańskiej, 2001.

[162] Marcin Szeląg, Salvatore Greco, Jerzy Błaszczyński, and Roman Słowiński. Case-based reasoning using dominance-based decision rules. In J.T. Yao et al., editors, *Rough Sets and Knowledge Technology 2011*, volume 6954 of *Lecture Notes in Computer Science*, pages 404–413, Heidelberg, 2011. Springer.

[163] Marcin Szeląg, Salvatore Greco, and Roman Słowiński. Rule-based approach to multicriteria ranking. In Michael Doumpos and Evangelos Grigoroudis, editors, *Multicriteria Decision Aid and Artificial Intelligence: Links, Theory and Applications*, chapter 6, pages 127–160. Wiley, 2013.

[164] Marcin Szeląg, Salvatore Greco, and Roman Słowiński. Variable Consistency Dominance-Based Rough Set Approach to Preference Learning in Multicriteria Ranking. *Information Sciences*, 277:525–552, September 2014.

[165] Marcin Szeląg, Roman Słowiński, and Jerzy Błaszczyński. jRank – ranking using dominance-based rough set approach. *Newsletter of the European Working Group "Multiple Criteria Decision Aiding"*, 3(22):13–15, 2010.

[166] Marcin Szeląg, Roman Słowiński, Salvatore Greco, Jerzy Błaszczyński, and Szymon Wilk. Ranking using dominance-based rough set approach. Research Report RA-07/10, Poznań University of Technology, Poznań, Poland, 2010.

[167] Christian Thiel, Stefan Scherer, and Friedhelm Schwenker. Fuzzy-input fuzzy-output one-against-all support vector machines. In B. Apolloni et al., editor, *KES 2007/ WIRN 2007, Part III*, volume 4694 of *Lecture Notes in Artificial Intelligence*, pages 156–165, Berlin, 2007. Springer.

[168] Evangelos Triantaphyllou. *Multi-Criteria Decision Making: A Comparative Study.* Kluwer Academic Publishers, Dordrecht, Netherlands, 2000.

[169] Tadeusz Trzaskalik. *Multiple criteria methods in the Polish financial market (in Polish).* Polskie Wydawnictwa Ekonomiczne S.A., 2006.

[170] Alexis Tsoukias and Phil Vincke. A new axiomatic foundation of partial comparability. *Theory and Decision*, 39:79–114, 1995.

[171] Alexis Tsoukias and Philipp Vincke. Extended preference structures in MCDA. In J. Climaco, editor, *Multicriteria Analysis*, pages 37–50. Springer, Berlin, 1997.

[172] Jan Verwaeren. *Mathematical optimization methods for the analysis of compositional data: subset selection, unmixing and prediction.* PhD thesis, Ghent University, 2014.

[173] Philippe Vincke. Exploitation of a crisp relation in a ranking problem. *Theory and Decision*, 32:221–240, 1992.

[174] Willem Waegeman, Jan Verwaeren, Bram Slabbinck, and Bernard De Baets. Supervised learning algorithms for multi-class classification problems with partial class memberships. *Fuzzy S*, 184:106–125, 2011.

[175] Margaret M. Wiecek, Matthias Ehrgott, Georges Fadel, and Jose Rui Figueira. Multiple criteria decision making for engineering. *Omega*, 36(3):337–339, 2008.

[176] Frank Wilcoxon. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83, 1945.

[177] S. K. Michael Wong and Wojciech Ziarko. Comparison of the probabilistic approximate classification and the fuzzy set model. *Fuzzy Sets and Systems*, 21:357–362, 1987.

[178] Yiyu Y. Yao. Probabilistic rough set approximations. *International Journal of Approximate Reasoning*, 49(2):255–271, 2008.

[179] H. Peyton Young. An axiomatization of Borda's rule. *Journal of Economic Theory*, 9:43–52, 1974.

[180] Wei Yu. ELECTRE TRI : Aspects méthodologiques et manuel d'utilisation. Document du LAMSADE no 74, Université Paris-Dauphine, 1992.

[181] Lotfi Zadeh. Fuzzy sets. *Information and Control*, 8(3):338–353, 1965.

[182] Sun Zhaoxu and Han Min. Multi-criteria decision making based on PROMETHEE method. In *2010 International Conference on Computing, Control and Industrial Engineering (CCIE 2010)*, volume 1, pages 416–418, 2010.

[183] Piotr Zielniewicz. A ranking method based on the aggregate distance measure function in the value space. *Expert Systems with Applications*, submitted.

Poznań University of Technology
Faculty of Computing
Institute of Computing Science

Typeset using LaTeX in Computer Modern.

BibTeX:

```
@phdthesis{ key,
    author = "Marcin Szeląg",
    title = "{Application of the Dominance-based Rough Set Approach to
Ranking and Similarity-based Classification Problems}",
    school = "Poznań University of Technology",
    address = "Pozna{\'n}, Poland",
    year = "2015",
}
```