

Politechnika Poznańska

Wydział Informatyki i Zarządzania

Instytut Informatyki

Wielokryterialna ocena atrakcyjności reguł decyzyjnych i asocjacyjnych

Streszczenie rozprawy doktorskiej

Izabela Szczęch

Promotor: prof. dr hab. inż. Roman Słowiński

Poznań 2007

Spis treści

Spis treści.....	2
1. Wstęp.....	3
1.1. Uzasadnienie podjęcia tematyki badań.....	3
1.2. Cel i zakres pracy.....	4
2. Podstawy ilościowej oceny reguł.....	6
2.1. Miary atrakcyjności reguł decyzyjnych i asocjacyjnych.....	7
2.2. Pożądane własności miar atrakcyjności.....	9
2.2.1. Własność M.....	9
2.2.2. Własność Bayesowskiej confirmacji.....	10
2.2.3. Własność symetrii hipotetycznej.....	10
3. Analiza własności miar atrakcyjności.....	11
3.1. Analiza miar atrakcyjności względem własności M.....	11
3.2. Analiza miar atrakcyjności względem własności Bayesowskiej confirmacji ...	13
3.3. Analiza miar atrakcyjności względem własności symetrii hipotetycznej.....	14
4. Wielokryterialna ocena atrakcyjności reguł.....	16
4.1. Porządek częściowy na zbiorze reguł i zbiór Pareto-optymalny.....	16
4.2. Ocena reguł względem miar wsparcia i ufności.....	17
4.3. Ocena reguł względem miary wsparcia i miary f	18
4.3.1. Nałożenie semantyki confirmacji na przestrzeń wsparcie–ufność.....	19
4.4. Ocena reguł względem miary wsparcia i miary s	20
4.5. Ocena reguł względem miary wsparcia i miar confirmacji z własnością M.....	21
5. Ocena reguł względem miar wsparcia i anty-wsparcia.....	22
5.1. Nałożenie semantyki confirmacji na przestrzeń wsparcie–anty-wsparcie.....	23
6. Przykłady zastosowania miar atrakcyjności w wielokryterialnej ocenie reguł.....	25
6.1. Koncepcja systemu wielokryterialnej oceny reguł.....	25
6.2. Przykłady zastosowania systemu.....	27
6.2.1. Zbiór danych <i>census</i>	27
6.2.2. Zbiór danych MSweb.....	29
6.2.3. Zbiór danych HSV.....	31
7. Podsumowanie.....	32
Bibliografia.....	34

1. Wstęp

Stosowane od lat w bardzo różnych obszarach systemy informatyczne oraz systemy zautomatyzowanego pozyskiwania informacji pozwalają na gromadzenie ogromnych wolumenów danych, których efektywne przetwarzanie i analiza wciąż stanowią problem. W szczególności, ważnym i aktualnym obszarem badań jest poszukiwanie nowych, nietrywialnych i potencjalnie użytecznych wzorców z danych [21], [57].

Jak wiadomo, powszechnie stosowaną formą reprezentacji wzorców odkrytych z danych są reguły decyzyjne bądź asocjacyjne. Duża liczba potencjalnie wyindukowanych reguł rodzi potrzebę oceny ich użyteczności. W tym celu wprowadza się pewne ilościowe kryteria oceny reguł zwane miarami atrakcyjności. Każda z miar kładzie nacisk na uwzględnienie innych cech pojedynczej reguły, czy całego ich zbioru. Jedne mają charakter bardziej probabilistyczny, inne zaś określają siłę zależności między przesłanką a konkluzją reguły.

Przy wyborze miary użytkownicy kierują się też pewnymi własnościami miar, które odzwierciedlają ich oczekiwania w stosunku do zachowania się tych miar w określonych sytuacjach. Przykładowo, intuicyjnie zrozumiałe jest oczekiwanie użytkownika, by wraz ze wzrostem liczby przykładów potwierdzających regułę, wartość miary poprawiła się, a przynajmniej nie uległa pogorszeniu. W ogólności, ze względu na mnogość miar, ich dobór dla konkretnego zastosowania, konkretnego zbioru danych i dziedziny problemu jest zadaniem nietrywialnym.

Oceniając reguły za pomocą miar, dąży się oczywiście do poszukiwania reguł optymalnych. W praktyce jednak pojedyncza miara nie jest na ogół dostatecznym wskaźnikiem jakości reguł i dlatego niezbędna jest ocena wielokryterialna, czyli z użyciem różnych miar atrakcyjności jednocześnie. W przypadku wielokryterialnej oceny reguł poszukiwany jest zbiór reguł Pareto-optymalnych (tj. niezdominowanych).

1.1. Uzasadnienie podjęcia tematyki badań

Dotychczasowe prace w literaturze dotyczącej wielokryterialnej oceny atrakcyjności reguł są z jednej strony nieliczne, a z drugiej, pozostawiły wiele otwartych problemów. W szczególności dotyczą one analizy własności miar, relacji między różnymi miarami i zależności między zbiorami reguł Pareto-optymalnych.

W tym kontekście uzasadnione było podjęcie prac zmierzających do wykazania, które z popularnych miar oceny atrakcyjności reguł posiadają pożądane własności: własność M, własność konfirmacji oraz własność symetrii hipotetycznej. Ponadto, istnieje potrzeba analizy zależności pomiędzy wybranymi miarami, a co za tym idzie, wykazania, czy między zbiorami reguł niezdominowanych względem wybranych miar zachodzą związki zawierania. Wyniki takich badań pozwolą na efektywniejsze wyznaczanie zawężonego zbioru reguł, gdyż zamiast wielokrotnej oceny reguł w różnych przestrzeniach, można by dokonać tej oceny raz, znajdując najbardziej ogólny zbiór Pareto, który dzięki odkrytym zależnościom pomiędzy miarami, na pewno zawierać będzie wszystkie inne interesujące reguły Pareto-optymalne.

1.2. Cel i zakres pracy

W pracy rozważa się zależności pomiędzy miarami atrakcyjności dla reguł prowadzących do tego samego wniosku. Badania dotyczą miar, które posiadają pożądaną własność monotonicznej zależności od liczby obserwacji wspierających lub nie przesłankę i konkluzję (własność M).

W kontekście powyższych założeń celem ogólnym pracy jest:

Zbadanie występowania zależności pomiędzy popularnymi miarami atrakcyjności reguł decyzyjnych i asocjacyjnych oraz zaproponowanie wielokryterialnej oceny reguł, według której zbiór reguł niezdominowanych zawierałby wszystkie reguły optymalne ze względu na dowolną miarę z własnością M.

Wyniki tych badań pozwolą na efektywniejsze wyznaczenie zbioru reguł, zawierającego reguły optymalne względem wielu miar.

Cele szczegółowe pracy obejmują:

1. Analiza własności następujących miar atrakcyjności reguł: wsparcia (rule support), anty-wsparcia (rule anti-support), ufności (confidence), funkcji atrakcyjności reguł (rule interest function), zysku (gain), współczynnika zależności (dependency factor), miar f oraz s . Analiza dotyczy własności M , własności konfirmacji i własności symetrii hipotetycznej.
2. Analiza zależności zachodzących pomiędzy rozważanymi miarami atrakcyjności oraz zbadanie zachodzenia związków zawierania się między zbiorami reguł niezdominowanych względem tych miar.
3. Zaproponowanie wielokryterialnej oceny reguł, w której zbiór reguł niezdominowanych zawiera reguły optymalne ze względu na dowolną miarę z pożądaną własnością M .
4. Określenie interesującego obszaru w zaproponowanej przestrzeni wielokryterialnej, zawierającego reguły z dodatnią konfirmacją.
5. Dostosowanie algorytmu Apriori do przypadku generowania reguł podlegających ocenie za pomocą miar o pożądanych właściwościach oraz prezentacja zastosowania uzyskanych wyników do analizy zbiorów reguł wyindukowanych z przykładowych zbiorów danych.

2. Podstawy ilościowej oceny reguł

W pracy rozważać będziemy odkrytą wiedzę reprezentowaną w formie reguł. Punktem wyjściowym indukcji reguł jest tablica informacyjna stanowiąca próbkę (podzbiór) większej rzeczywistości. Formalnie, *tablica informacyjna* to para:

$$S = (U, A). \quad (2.1)$$

gdzie U jest niepustym i skończonym zbiorem obiektów, a A jest niepustym, skończonym zbiorem atrybutów opisujących obiekty. Z każdym atrybutem $a \in A$ związana jest dziedzina jego wartości oznaczana jako V_a . Wartości obiektu $x \in U$ na atrybucie $a \in A$ oznaczają będziemy przez $a(x)$. Atrybuty opisujące obiekty mogą być różnego rodzaju. Klasyczna hierarchia [83], [5], [54], [13], [82] dzieli atrybuty na nominalne, porządkowe, atrybuty ze skalą przedziałową lub interwałową. W literaturze wyróżnia się ponadto *kryteria* będące atrybutami o dziedzinie uporządkowanej ze względu na preferencje (np. od najmniej do najbardziej pożądanej wartości) [77]. Wielu autorów dostrzega także złożone (strukturalne) rodzaje atrybutów [31], [51].

Reguły asocjacyjne i decyzyjne

Reguły indukowane z tabeli S mają postać $\phi \rightarrow \psi$ ("jeżeli ϕ , to ψ "), gdzie ϕ (przesłanka) oraz ψ (konkluzja, decyzja, hipoteza, klasa) są zbudowane z koniunkcji warunków elementarnych. *Warunek elementarny* reguły to para atrybut-wartość atrybutu połączona relacją $=$, $<$, \leq , \geq lub $>$. Użytkownik może wyróżnić pewien atrybut jako *atrybut decyzyjny*, przez co wymuszone będzie występowanie tylko tego atrybutu po stronie konkluzji w regule, oraz nie wchodzenie tego atrybutu w skład przesłanki reguły. Jeśli podział na atrybuty decyzyjne i warunkowe (nie-decyzyjne) jest ustalony i niezmienny dla całego procesu generowania reguł, to będziemy wówczas mówili o indukcji *reguł decyzyjnych*, w przeciwnym zaś razie - o indukcji *reguł asocjacyjnych*.

2.1. Miary atrakcyjności reguł decyzyjnych i asocjacyjnych

Liczba reguł potencjalnie wyindukowanych z danych często ogranicza lub wręcz wyklucza możliwość ręcznej oceny i analizy ich przez eksperta. Ponadto, zwykle tylko pewna część reguł jest nietrywialna i użyteczna. W tym kontekście naturalnie zrodziła się potrzeba wprowadzenia pewnych ilościowych kryteriów oceny reguł zwanych miarami atrakcyjności. Obecnie literatura stanowi bogaty zbiór miar wprowadzanych przez autorów, by uwypuklić szczególnie ważne ich zdaniem cechy pojedynczej reguły lub całego ich zbioru [1], [43], [48], [86]. Poniżej zdefiniowano miary atrakcyjności analizowane w pracy.

Wsparcie reguły (Rule support)

Wsparcie jest jedną z najpopularniejszych miar atrakcyjności reguł [2], [35]. *Wsparcie przesłanki (konkluzji)* oznaczane jako $sup(\phi)$ ($sup(\psi)$) jest liczbą obiektów w U spełniających przesłankę (konkluzję). *Wsparcie reguły* $\phi \rightarrow \psi$ oznaczane jako $sup(\phi \rightarrow \psi)$, jest liczbą obiektów w U , dla których zarówno przesłanka ϕ jak i konkluzja ψ są spełnione.

Anty-wsparcie reguły (Rule anti-support)

Anty-wsparcie reguły $\phi \rightarrow \psi$ oznaczane jako $anti-sup(\phi \rightarrow \psi)$, to liczba obiektów ze zbioru U , które spełniają przesłankę ϕ , ale konkluzję mają inną niż ψ . Miara anty-wsparcia może być zatem postrzegana jako liczba anty-przykładów czyli $sup(\phi \rightarrow \neg\psi)$.

Zarówno wsparcie jak i anty-wsparcie reguły są czasem w literaturze definiowane jako wartości względne w stosunku do liczebności zbioru U . W pracy nie posługiwano się tą definicją, aczkolwiek nie wpłynęłaby ona na ogólność wyników i wniosków. Dla prostoty będziemy stosować nazwy *wsparcie* i *anty-wsparcie* jako synonimy wsparcia i anty-wsparcia reguły.

Ufność (Confidence)

Ufność reguły $\phi \rightarrow \psi$, oznaczaną przez $conf(\phi \rightarrow \psi)$, definiuje się jako [2]:

$$conf(\phi \rightarrow \psi) = \frac{sup(\phi \rightarrow \psi)}{sup(\phi)}. \quad (2.2)$$

Mnożona przez 100 ufność może zatem być interpretowana jako procent obiektów spełniających całą regułę w zbiorze obiektów spełniających jej przesłankę.

Funkcja atrakcyjności reguły (Rule interest function)

Funkcja atrakcyjności reguły RI wprowadzona przez Piatetsky-Shapiro [67] mierzy korelację pomiędzy przesłanką a konkluzją reguły. Definiowana jest jako:

$$RI(\phi \rightarrow \psi) = sup(\phi \rightarrow \psi) - \frac{sup(\phi)sup(\psi)}{|U|}. \quad (2.3)$$

Dla reguł $\phi \rightarrow \psi$, wartość $RI=0$, oznacza statystyczną niezależność ϕ i ψ . Jeśli $RI > 0$ ($RI < 0$), wówczas mówimy o pozytywnej (negatywnej) korelacji pomiędzy ϕ i ψ [35].

Funkcja zysku (Gain function)

Funkcja zysku gain została zdefiniowana przez Fukudę i innych [25] jako:

$$gain(\phi \rightarrow \psi) = sup(\phi \rightarrow \psi) - \Theta sup(\phi) \quad (2.4)$$

gdzie Θ jest ułamkiem z zakresu 0 - 1.

Można zaobserwować, że dla wartości $\Theta = sup(\psi)/|U|$, miara *gain* sprowadza się do miary *RI*.

Współczynnik zależności (Dependency factor)

Współczynnik zależności η stosowany przez Pawlaka [65], a wcześniej także w pracach Poppera [69] jest definiowany jako:

$$\eta(\phi \rightarrow \psi) = \frac{\frac{sup(\phi \rightarrow \psi)}{sup(\phi)} - \frac{sup(\psi)}{|U|}}{\frac{sup(\phi \rightarrow \psi)}{sup(\phi)} + \frac{sup(\psi)}{|U|}}. \quad (2.5)$$

W sytuacji niezależności ϕ i ψ miara η przyjmuje wartość 0. Gdy zaś występowanie ϕ wpływa pozytywnie na prawdopodobieństwo wystąpienia ψ , η przyjmuje wartość z przedziału (0, 1), a gdy wpływ ten jest negatywny - to z przedziału (-1, 0).

Miary f oraz s

Wśród miar konfirmacji (szerzej omawianych w rozdziale 2.2.2) ważną rolę odgrywają miary f oraz s , definiowane w następujący sposób:

$$f(\phi \rightarrow \psi) = \frac{\Pr(\phi | \psi) - \Pr(\phi | \neg\psi)}{\Pr(\phi | \psi) + \Pr(\phi | \neg\psi)}, \quad (2.6)$$

$$s(\phi \rightarrow \psi) = \Pr(\psi | \phi) - \Pr(\psi | \neg\phi). \quad (2.7)$$

W literaturze, miara f była poruszana przez takich autorów jak Kemeny i Oppenheim [44], Good [27], Heckerman [33], Horvitz i Heckerman [37], Pearl [66], Schum [75] oraz Fitelson [22]. Miara s została zaproponowana przez Christensena [11] oraz Joyce'a [40].

2.2. Pożądane własności miar atrakcyjności

W ogólności, ze względu na mnogość miar ich dobór dla konkretnego zastosowania, konkretnego zbioru danych i dziedziny problemu jest zadaniem nietrywialnym. Dlatego też, przy wyborze miary użytkownicy kierują się też pewnymi własnościami miar, które odzwierciedlają ich oczekiwania w stosunku do zachowania się tych miar w określonych sytuacjach. Własności miar grupują je względem podobieństwa ich cech, ułatwiając użytkownikowi dobór miary.

2.2.1. Własność M

Greco, Pawlak and Słowiński [28] zdefiniowali własność M monotonicznej zależności miar atrakcyjności od liczby obiektów potwierdzających bądź nie przesłankę lub konkluzję reguły. Formalnie, miara atrakcyjności

$$I(\phi \rightarrow \psi) = F[\sup(\phi \rightarrow \psi), \sup(\neg\phi \rightarrow \psi), \sup(\phi \rightarrow \neg\psi), \sup(\neg\phi \rightarrow \neg\psi)] \quad (2.8)$$

będąca kryterium typu zysk (tzn. im wyższa wartość miary tym bardziej pożądana), posiada własność M wtedy i tylko wtedy, gdy jest funkcją:

- niemalejącą względem $\sup(\phi \rightarrow \psi)$, oraz
- nierosnącą względem $\sup(\neg\phi \rightarrow \psi)$, oraz
- nierosnącą względem $\sup(\phi \rightarrow \neg\psi)$, oraz
- niemalejącą względem $\sup(\neg\phi \rightarrow \neg\psi)$.

Odpowiednio, miara atrakcyjności

$$I(\phi \rightarrow \psi) = F[\sup(\phi \rightarrow \psi), \sup(\neg\phi \rightarrow \psi), \sup(\phi \rightarrow \neg\psi), \sup(\neg\phi \rightarrow \neg\psi)] \quad (2.9)$$

będąca kryterium typu koszt (tzn. im wyższa wartość miary tym gorzej), posiada własność M wtedy i tylko wtedy, gdy jest funkcją:

- nierosnącą względem $\sup(\phi \rightarrow \psi)$, oraz
- niemalejącą względem $\sup(\neg\phi \rightarrow \psi)$, oraz
- niemalejącą względem $\sup(\phi \rightarrow \neg\psi)$, oraz
- nierosnącą względem $\sup(\neg\phi \rightarrow \neg\psi)$.

2.2.2. Własność Bayesowskiej confirmacji

Miara atrakcyjności $c(\phi \rightarrow \psi)$ posiada własność Bayesowskiej confirmacji wtedy i tylko wtedy, gdy spełnia następujące warunki:

$$c(\phi \rightarrow \psi) \begin{cases} > 0 \text{ if } \Pr(\psi | \phi) > \Pr(\psi), \\ = 0 \text{ if } \Pr(\psi | \phi) = \Pr(\psi), \\ < 0 \text{ if } \Pr(\psi | \phi) < \Pr(\psi). \end{cases} \quad (2.10)$$

W ogólności zatem, dla reguły $\phi \rightarrow \psi$, miara atrakcyjności, która posiada własność confirmacji (nazywana również *miarą confirmacji*), wyraża prawdziwość stwierdzenia, że konkluzja ψ zachodzi częściej, gdy zaszło również ϕ niż, gdy ϕ nie zaszło.

2.2.3. Własność symetrii hipotetycznej

Eells and Fitelson analizowali [19] miary confirmacji pod względem czterech własności symetrii zaproponowanych przez Carnapa [10]. Wyniki ich prac stanowią argument za badaniem symetrii względem konkluzji zwanej również *symetrią hipotetyczną*, jako jedynej pożądanej symetrii:

Miara atrakcyjności I posiada własność symetrii hipotetycznej, jeśli jej wartość dla reguł $\phi \rightarrow \psi$ oraz $\phi \rightarrow \neg\psi$ różni się jedynie znakiem, czyli: $I(\phi \rightarrow \psi) = -I(\phi \rightarrow \neg\psi)$.

3. Analiza własności miar atrakcyjności

Analiza popularnych miar atrakcyjności reguł pod względem pożądanych własności rozszerza wiedzę na temat miar i zakresu ich stosowania. Ponadto, umożliwia określenie zależności zachodzących pomiędzy różnymi miarami. Analiza własności miar jest również bardzo użyteczna z praktycznego punktu widzenia, gdyż jej wyniki ułatwiają użytkownikowi wybór miary o pożądanych cechach, potrzebnych w konkretnym zastosowaniu. W literaturze wiele jest prac analizujących miary atrakcyjności pod kątem ich własności ([19], [22], [28]), jednakże analiza wielu powszechnie stosowanych miar pod względem własności M, własności confirmacji, czy własności symetrii hipotetycznej nadal pozostaje otwartym problemem.

Dla prostoty i jasności prezentacji, w dalszych częściach pracy stosowane będą następujące oznaczenia:

$$\begin{aligned} a &= \sup(\phi \rightarrow \psi), \quad b = \sup(\neg\phi \rightarrow \psi), \quad c = \sup(\phi \rightarrow \neg\psi), \quad d = \sup(\neg\phi \rightarrow \neg\psi), \\ a + c &= \sup(\phi), \quad a + b = \sup(\psi), \quad b + d = \sup(\neg\phi), \quad c + d = \sup(\neg\psi), \\ a + b + c + d &= |U|. \end{aligned} \tag{3.1}$$

Zakłada się ponadto, że U jest zbiorem niepustym, a zatem przynajmniej jeden z wyrazów a, b, c, d jest większy od zera.

3.1. Analiza miar atrakcyjności względem własności M

Analiza miary atrakcyjności $I(\phi \rightarrow \psi)$ typu zysk pod kątem posiadania przez nią własności M opierała się na weryfikacji następujących warunków:

1. wzrost wartości a nie implikuje spadku wartości I ,
 2. wzrost wartości b nie implikuje wzrostu wartości I ,
 3. wzrost wartości c nie implikuje wzrostu wartości I ,
 4. wzrost wartości d nie implikuje spadku wartości I .
- (3.2)

Odpowiednio, dla miar I typu koszt analiza sprawdzała spełnienie następujących warunków:

1. wzrost wartości a nie implikuje wzrostu wartości I ,
 2. wzrost wartości b nie implikuje spadku wartości I ,
 3. wzrost wartości c nie implikuje spadku wartości I ,
 4. wzrost wartości d nie implikuje wzrostu wartości I .
- (3.3)

Wyniki analizy

Wsparcie i anty-wsparcie reguły

Przeprowadzona analiza wykazała, że wsparcie i anty-wsparcie reguły są miarami posiadającymi własność M.

Ufność

Analiza wykazała monotoniczną zależność miary ufności od a , anty-monotoniczną zależności od c oraz niezależność od b i d , co pozwoliło sformułować następujące twierdzenie:

Twierdzenie 3.1.

Miara ufności posiada własność M.

Funkcja atrakcyjności reguły

Spełnienie przez miarę RI warunków (3.2) pozwoliło udowodnić następujące twierdzenie:

Twierdzenie 3.2. [29]

Miara RI posiada własność M.

Funkcja zysku

Analiza wykazała monotoniczną zależność miary $gain$ od a , anty-monotoniczną zależności od c oraz niezależność od b i d , z czego wyniknęło następujące twierdzenie:

Twierdzenie 3.3.

Miara $gain$ posiada własność M.

Współczynnik zależności

Analiza współczynnika zależności pod względem spełnienia warunków (3.2) wykazała istnienie kontrprzykładu, co bezpośrednio implikuje następujące twierdzenie:

Twierdzenie 3.4. [29]

Miara η nie spełnia własności M.

Miary f oraz s

W pracy Greco i inni [28] udowodniono, że miary f oraz s posiadają własność M.

3.2. Analiza miar atrakcyjności względem własności Bayesowskiej confirmacji

Bayesowskiej confirmacji

Analiza miary atrakcyjności pod względem własności Bayesowskiej confirmacji polegała na sprawdzeniu czy następujące warunki są spełnione:

1. miara przyjmuje wartości dodatnie $\Leftrightarrow \text{conf}(\phi \rightarrow \psi) > \text{sup}(\psi)/|U|$
 2. miara przyjmuje wartość 0 $\Leftrightarrow \text{conf}(\phi \rightarrow \psi) = \text{sup}(\psi)/|U|$
 3. miara przyjmuje wartości ujemne $\Leftrightarrow \text{conf}(\phi \rightarrow \psi) < \text{sup}(\psi)/|U|$
- (3.4)

Wyniki analizy

Wsparcie i anty-wsparcie reguły oraz ufność

Można łatwo zauważyć, że z racji, że dziedziny miar wsparcia, anty-wsparcia i ufności są zawężone jedynie do wartości nieujemnych, miary te nie mogą spełnić wszystkich warunków (3.4), a zatem nie posiadają własności confirmacji.

Funkcja atrakcyjności reguły

Przeprowadzona analiza wykazała spełnienie przez miarę *RI* wszystkich warunków (3.4), czym umożliwiła sformułowanie następującego twierdzenia:

Twierdzenie 3.5.

Miara *RI* posiada własność confirmacji.

Funkcja zysku

Analiza miary *gain* pod kątem posiadania przez nią własności confirmacji pozwoliła na wyznaczenie warunków, przy których ma ona tę własność.

Twierdzenie 3.6.

Miara *gain* posiada własność confirmacji wtedy i tylko wtedy, gdy $\Theta = \text{sup}(\psi)/|U|$.

Współczynnik zależności

Spełnienie przez miarę η wszystkich warunków (3.4) implikuje twierdzenie:

Twierdzenie 3.7.

Miara η posiada własność confirmacji.

Miary f oraz s

Miary f oraz s należą do powszechnie stosowanych miar konfirmacji. Już od ich wprowadzenia w literaturze podkreślano posiadanie przez nie własności konfirmacji ([9], [22], [47]).

3.3. Analiza miar atrakcyjności względem własności symetrii hipotetycznej

Analiza miary atrakcyjności pod kątem posiadania przez nią własności symetrii hipotetycznej opiera się na weryfikacji, czy wartości miary dla reguł $\phi \rightarrow \psi$ oraz $\phi \rightarrow \neg\psi$ są takie same, ale o przeciwnych znakach.

Wyniki analizy

Wsparcie i anty-wsparcie reguły oraz ufność

Podobnie, jak w przypadku własności konfirmacji, ograniczona dziedzina miar wsparcia, anty-wsparcia i ufności uniemożliwia im posiadanie własności symetrii hipotetycznej.

Funkcja atrakcyjności reguły

Analiza wartości miary RI dla reguły $\phi \rightarrow \psi$ oraz $\phi \rightarrow \neg\psi$ wykazała poprawność następującego twierdzenia:

Twierdzenie 3.8. [29]

Miara RI posiada własność symetrii hipotetycznej.

Funkcja zysku

Analiza miary $gain$ pod kątem posiadania przez nią własności symetrii hipotetycznej pozwoliła na wyznaczenie warunku, przy którym ma ona tę własność.

Twierdzenie 3.9. [29]

Miara $gain$ posiada własność symetrii hipotetycznej wtedy i tylko wtedy, gdy $\Theta=1/2$.

Współczynnik zależności

Znalezienie kontrprzykładu, dla którego $\eta(\phi \rightarrow \psi) \neq -\eta(\phi \rightarrow \neg\psi)$ implikuje, że miara ta nie posiada własności symetrii hipotetycznej.

Twierdzenie 3.10. [29]

Miara η nie posiada własności symetrii hipotetycznej.

Miary f oraz s

Eells i inni w [19] analizowali liczne miary confirmacji pod kątem własności symetrii.

Udowodnili oni, że miary f oraz s posiadają własność symetrii hipotetycznej.

4. Wielokryterialna ocena atrakcyjności reguł

W praktycznych zastosowaniach często pojedyncza miara atrakcyjności nie jest dostatecznym wskaźnikiem jakości reguł i dlatego ocena reguł jest robiona wielokryterialnie czyli z użyciem kilku miar atrakcyjności jednocześnie. Podejście wielokryterialne jest powszechnie poruszane w literaturze [3], [9], [23], [32]. Jednakże badanie związków pomiędzy częściowym preporządkiem na zbiorze reguł tworzonym przy wielokryterialnej ocenie, a preporządkiem zupełnym generowanym przy ocenie za pomocą pojedynczej miary, nadal pozostaje ciekawym i niewyczerpanym zagadnieniem badawczym.

4.1. Porządek częściowy na zbiorze reguł i zbiór

Pareto-optimalny

Porządek częściowy \preceq_{qt} na zbiorze reguł względem dwóch różnych miar atrakcyjności q oraz t jest zdefiniowany następująco:

dla zbioru reguł X i dwóch reguł $r_1, r_2 \in X$, $r_1 \prec_{qt} r_2$ wtedy i tylko wtedy gdy:

$$\begin{aligned} q(r_1) \leq q(r_2) \wedge t(r_1) < t(r_2) \text{ lub} \\ q(r_1) < q(r_2) \wedge t(r_1) \leq t(r_2), \end{aligned} \quad (4.1)$$

a ponadto $r_1 \sim_{qt} r_2$ wtedy i tylko wtedy gdy:

$$q(r_1) = q(r_2) \wedge t(r_1) = t(r_2). \quad (4.2)$$

Zbiór Pareto-optimalny

Jeśli dla reguły $r \in X$ nie istnieje żadna reguła $r' \in X$, taka że $r \prec_{qt} r'$, to regułę r określamy jako regułę *niezdominowaną* (czyli *Pareto-optymalną*) względem miar q oraz t . Zbiór wszystkich reguł niezdominowanych względem tych miar nazywamy *zbiorem Pareto-optymalnym $q-t$* .

Monotoniczność funkcji

Funkcja $g(x)$ jest *monotoniczna* względem x , jeśli $x_1 > x_2$ implikuje $g(x_1) \geq g(x_2)$. Analogicznie, funkcja $g(x)$ jest *anty-monotoniczna* względem x , jeśli $x_1 > x_2$ implikuje $g(x_1) \leq g(x_2)$.

4.2. Ocena reguł względem miar wsparcia i ufności

Bayardo i Agrawal [3] zaproponowali ocenianie reguł względem miar wsparcia i ufności jednocześnie. Wykazali oni, że dla reguł z tą samą konkluzją, reguły optymalne względem takich miar atrakcyjności jak *gain*, Laplace [12], lift [38], conviction [6], *RI*, etc. znajdują się w zbiorze reguł Pareto-optymalnych względem wsparcia i ufności. Ten praktyczny wynik oznacza, że znalezienie reguł niezdominowanych względem wsparcia i ufności gwarantuje jednoczesne znalezienie reguł optymalnych względem wielu innych popularnie stosowanych miar. Ponadto udowodnili oni, że następujące warunki są wystarczające, by reguły optymalne względem dowolnej innej miary $g(r)$ również były w zbiorze reguł Pareto-optymalnych względem wsparcia i ufności:

- $g(r)$ jest monotoniczna względem wsparcia przy stałej wartości miary ufności i
 - $g(r)$ jest monotoniczna względem ufności przy stałej wartości miary wsparcia.
- (4.3)

Powyższe warunki można uogólnić na dowolną miarę w dowolnej przestrzeni ocen, co oznacza, że wykazanie zależności monotonicznych analogicznych do (4.3) pomiędzy badaną miarą v a miarami tworzącymi przestrzeń oceny, implikuje, że reguły optymalne względem v będą na brzegu Pareto-optymalnym w tej przestrzeni ocen.

Monotoniczna zależność pomiędzy miarą f a miarami wsparcia i ufności

Z uwagi na posiadanie przez miarę f wartościowych własności M, konfirmacji i symetrii hipotetycznej, podjęto w pracy problem analizy czy reguły optymalne ze względu na miarę f należą do zbioru reguł niezdominowanych względem miar wsparcia i ufności. W tym celu wykazano, że miara f jest monotoniczna względem wsparcia, przy stałej wartości ufności i monotoniczna względem ufności przy stałej wartości wsparcia.

Twierdzenie 4.1. [7]

Miara f jest niezależna od wsparcia, a co za tym idzie, monotoniczna względem wsparcia, gdy wartość ufności jest stała.

Twierdzenie 4.2. [7]

Miara f jest rosnąca względem miary ufności, a zatem monotoniczna względem niej.

W kontekście powyższych wyników oraz z uwagi na to, że semantyczna użyteczność miary f przewyższa semantyczną użyteczność miary ufności ([8], [67]), proponujemy nową przestrzeń oceny reguł opartą na mierze wsparcia i mierze f [79].

4.3. Ocena reguł względem miary wsparcia i miary f

Zastosowanie miary wsparcia i miary f w jednej przestrzeni oceny reguł jest uzasadnione obserwacją, że często reguły z wysoką wartością miary f charakteryzują się niskim wsparciem.

Wprowadzenie nowej przestrzeni oceny w naturalny sposób rodzi pytanie o porównanie jej z przestrzenią wsparcie–ufność. Przeprowadzona analiza [8] wykazała, że dla reguł z tą samą konkluzją zbiór reguł niezdominowanych w przestrzeni wsparcie–ufność zawiera dokładnie te same reguły, co zbiór reguł niezdominowanych w przestrzeni wsparcie– f [8], [9]. Przestrzeń wsparcie– f ma jednak nad przestrzenią wsparcie–ufność istotną przewagę: skala miary f pozwala na bezpośrednie odrzucenie reguł, dla których przesłanka nie potwierdza konkluzji. Liczba takich nieinteresujących reguł może być duża, co stanowi dodatkowy argument za stosowaniem przestrzeni wsparcie– f [81]. Informacje o procencie reguł nieinteresujących dla kilku przykładowych konkluzji reguł wygenerowanych ze zbioru *census* przy progu minimalnego wsparcia 0,15 przedstawia Tabela 4.1. Dalsze doświadczenia na tym zbiorze pokazały, że reguły zaprzeczające konkluzji, często znajdują się nawet na brzegu Pareto-optymalnym.

Tabela 4.1. Reguły z niedodatnią wartością miary confirmacji f dla różnych konkluzji reguł ze zbioru *census* (min. wsparcie=0,15)

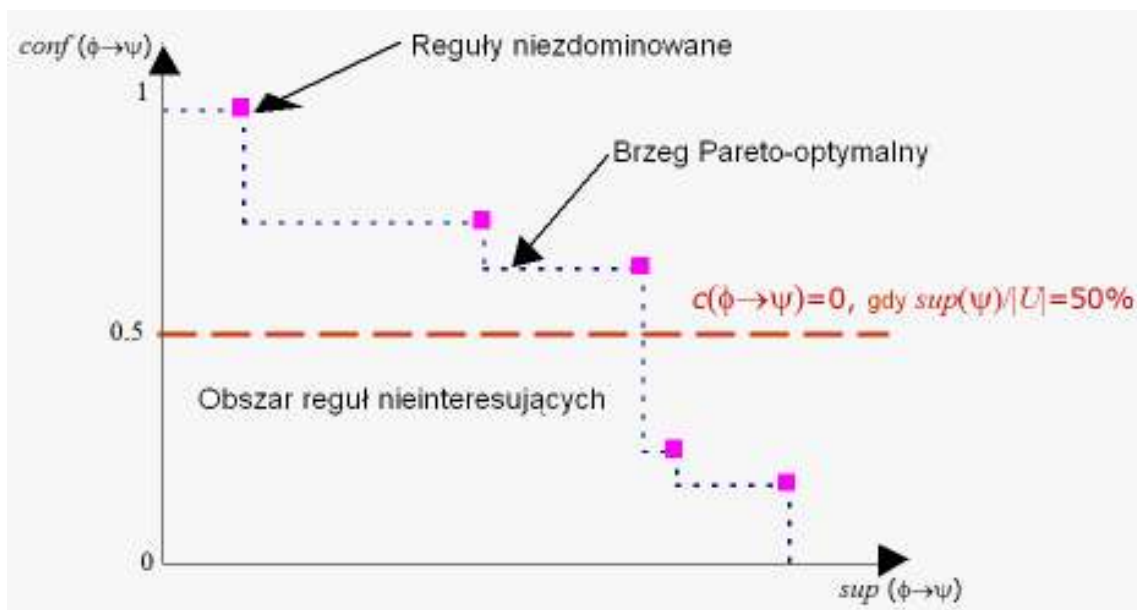
Badana konkluzja	Liczba wszystkich reguł	Lb. reguł z niedodatnią confirmacją	Wielkość redukcji
workclass=Private	84	42	50%
sex=Male	84	23	27%
income<=50kUSD	85	43	51%
race=White	105	27	26%
native_country=USA	111	30	27%

4.3.1. Nałożenie semantyki confirmacji na przestrzeń wsparcie–ufność

Zainspirowani wartością semantyki jaką niesie za sobą własność confirmacji przeprowadziliśmy próbę wyodrębnienia z przestrzeni wsparcie–ufność reguł z ujemną bądź niską wartością miary confirmacji f .

Przeprowadzona analiza wykazała, że w ramach reguł z tą samą konkluzją, reguły charakteryzujące się ujemną wartością dowolnej miary $c(\phi \rightarrow \psi)$ z własnością confirmacji znajdują się pod prostą $sup(\phi \rightarrow \psi)/|U|$, co formalnie opisuje wzór (4.4), a schematycznie przedstawia Rysunek 4.1 ([80], [81]).

$$c(\phi \rightarrow \psi) > 0 \Leftrightarrow conf(\phi \rightarrow \psi) > \frac{sup(\psi)}{|U|}. \quad (4.4)$$



Rysunek 4.1. Przykład prostej, dla której dowolna miara confirmacji $c(\phi \rightarrow \psi)=0$ w przestrzeni wsparcie–ufność

Uogólnienie wyniku (4.4) dla konkretnej miary confirmacji (w tym wypadku f) nie większej niż pewne $k \geq 0$ przedstawia Twierdzenie 4.3.

Twierdzenie 4.3. [81]

$$f(\phi \rightarrow \psi) \geq k \Leftrightarrow conf(\phi \rightarrow \psi) \geq \frac{sup(\psi)(k+1)}{|U| - k(|U| - 2sup(\psi))}. \quad (4.5)$$

4.4. Ocena reguł względem miary wsparcia i miary s

Występowanie monotonicznych powiązań pomiędzy miarą ufności a miarą f zainspirowało nas do analizy związków pomiędzy miarami wsparcia i ufności a inną miarą confirmacji – miarą s , a w konsekwencji także relacji reguł optymalnych względem s ze zbiorem reguł niezdominowanych względem wsparcia i ufności. Monotoniczne związki omawianych miar dla reguł z tą samą konkluzją określają Twierdzenie 4.4 i Twierdzenie 4.5.

Twierdzenie 4.4. [8]

Dla stałej wartości ufności:

1. miara $s(\phi \rightarrow \psi)$ jest rosnącą względem wsparcia $\Leftrightarrow s(\phi \rightarrow \psi) > 0$,
2. miara $s(\phi \rightarrow \psi)$ jest stała względem wsparcia $\Leftrightarrow s(\phi \rightarrow \psi) = 0$,
3. miara $s(\phi \rightarrow \psi)$ jest malejąca $\Leftrightarrow s(\phi \rightarrow \psi) < 0$.

Twierdzenie 4.5. [8]

Przy stałej wartości wsparcia, miara s jest rosnąco zależna, a zatem monotoniczna względem ufności.

Powyższe twierdzenia gwarantują, że reguły charakteryzujące się nieujemną (a zatem jedyną akceptowalną wartością miar confirmacji) wartością miary s będą znajdowały się na brzegu Pareto przestrzeni wsparcie–ufność.

Przestrzeń oceny reguł wykorzystująca miarę wsparcia i miarę s jest ciekawą alternatywą dla przestrzeni wsparcie–ufność ze względu na własność confirmacji miary s . Zależności pomiędzy zbiorami reguł niezdominowanych w tych przestrzeniach określa Twierdzenie 4.6. Wynika z niego, że zbiór Pareto w przestrzeni wsparcie–ufność jest nadzbiorem zbioru Pareto w przestrzeni wsparcie– s (dla $s \geq 0$).

Twierdzenie 4.6. [9]

Jeśli reguła należy do zbioru reguł niezdominowanych względem wsparcia i miary s (dla nieujemnych wartości s), to również znajduje się w zbiorze reguł niezdominowanych względem wsparcia i ufności, podczas gdy odwrotna sytuacja nie musi zachodzić.

Twierdzenie 4.6 można naturalnie uogólnić na dowolną miarę atrakcyjności i :

Twierdzenie 4.7. [9]

Dla miary atrakcyjności i , monotonicznej względem wsparcia i ufności, zbiór reguł niezdominowanych względem wsparcia i ufności jest nadzbiorem zbioru reguł niezdominowanych w przestrzeni wsparcie– i .

4.5. Ocena reguł względem miary wsparcia i miar konfirmacji z własnością M

Analiza związków monotoniczności z miarami wsparcia i ufności została także rozszerzona na całą klasę miar atrakcyjności z własnością M (jak poprzednio, rozważane są reguły z tą samą konkluzją). Poniższe twierdzenia pokazują, że dla stałej wartości miary ufności miara F z własnością M jest monotoniczna względem wsparcia tylko przy określonych warunkach, natomiast monotoniczność F względem ufności jest bezwarunkowa przy stałej wartości wsparcia.

Twierdzenie 4.8. [9]

Dla stałej wartości ufności, miara atrakcyjności $F(a, b, c, d)$ posiadająca własność M jest monotoniczna względem wsparcia jeśli:

$$\frac{\partial F}{\partial c} = \frac{\partial F}{\partial d} = 0 \text{ lub } \frac{\frac{\partial F}{\partial a} - \frac{\partial F}{\partial b}}{\frac{\partial F}{\partial d} - \frac{\partial F}{\partial c}} \geq \frac{1}{\text{conf}(\phi \rightarrow \psi)} - 1. \quad (4.6)$$

Twierdzenie 4.9. [9]

Dla stałej wartości miary wsparcia, miara atrakcyjności $F(a, b, c, d)$ posiadająca własność M jest monotoniczna względem ufności.

Miary, które spełniają warunki z powyższych twierdzeń i dodatkowo są miarami konfirmacji, są w naszym odczuciu dobrym kandydatem na zastąpienie miary ufności w przestrzeni wsparcie-ufność.

5. Ocena reguł względem miar wsparcia i anty-wsparcia

Twierdzenie 4.8 ujawnia, że wśród miar posiadających własność M , mogą istnieć takie, których reguły optymalne nie będą znajdowały się w zbiorze reguł niezdominowanych względem wsparcia i ufności. Z uwagi na praktyczną wartość własności M , istotnym problemem badawczym jest poszukiwanie przestrzeni oceny reguł, takiej że zbiór Pareto-optymalny względem miar ją tworzących, będzie zawierał wszystkie reguły optymalne względem dowolnej miary atrakcyjności posiadającej cechę M . Odkrycie monotonicznych oraz anty-monotonicznych zależności pomiędzy miarami wsparcia i anty-wsparcia, a miarami z własnością M (*Twierdzenie 5.1*, *Twierdzenie 5.2*), pozwoliło udowodnić, że nowo proponowana przestrzeń oceny reguł względem miar wsparcia i anty-wsparcia będzie rozwiązaniem powyższego problemu.

Twierdzenie 5.1. [9]

Dla stałej wartości miary anty-wsparcia, miara atrakcyjności $F(a, b, c, d)$ posiadająca własność M jest monotoniczna (nie-malejąca) względem wsparcia.

Twierdzenie 5.2. [9]

Dla stałej wartości miary wsparcia, miara atrakcyjności $F(a, b, c, d)$ posiadająca własność M jest anty-monotoniczna (nie-rosnąca) względem anty-wsparcia.

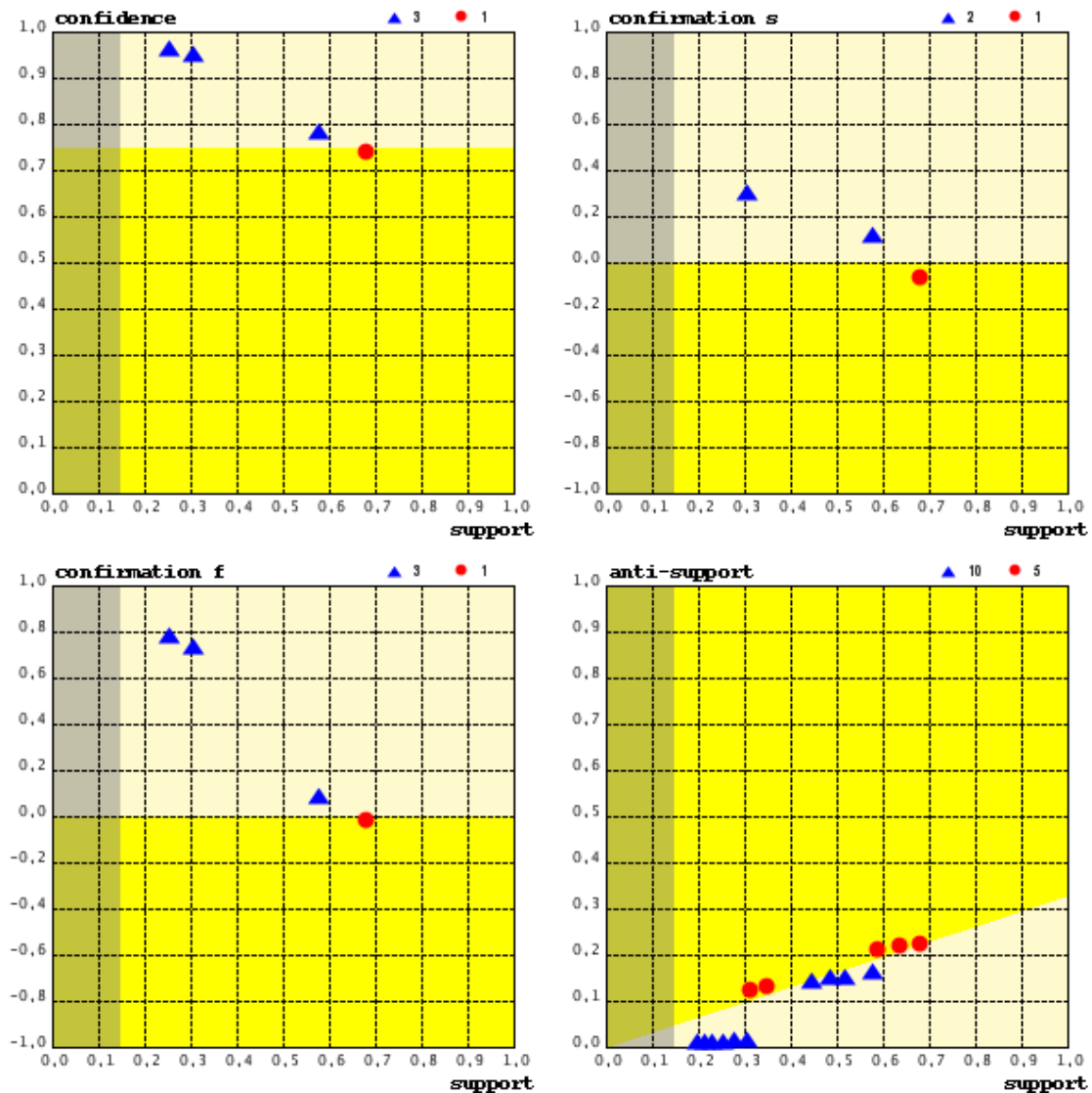
Twierdzenie 5.1 oraz *Twierdzenie 5.2* implikuje, że reguły optymalne względem dowolnej miary atrakcyjności $F(a, b, c, d)$ z własnością M będą znajdowały się w zbiorze reguł niezdominowanych względem wsparcia i anty-wsparcia.

Analiza zależności występujących pomiędzy brzegami Pareto-optymalnymi w przestrzeni wsparcie–ufność i wsparcie–anty-wsparcie pozwoliła wykazać, że zbiór reguł niezdominowanych względem wsparcia i anty-wsparcia jest nadzbiorem zbioru reguł niezdominowanych względem wsparcia i ufności.

Twierdzenie 5.3. [9]

Reguły leżące na brzegu Pareto-optymalnym w przestrzeni wsparcie–ufność, znajdują się również na brzegu Pareto-optymalnym w przestrzeni wsparcie–anty-wsparcie, podczas gdy sytuacja odwrotna nie musi mieć miejsca.

Zbiorcze porównanie relacji zawierania się zbiorów reguł niezdominowanych względem różnych analizowanych przestrzeni oceny, przedstawia Rysunek 5.1



Rysunek 5.1 Porównanie brzegów Pareto- optymalnych w różnych przestrzeniach oceny reguł (reguły wygenerowane dla konkluzji $income \leq 50$ k USD, dla zbioru *census*, przy progu $min. sup=0,15$)

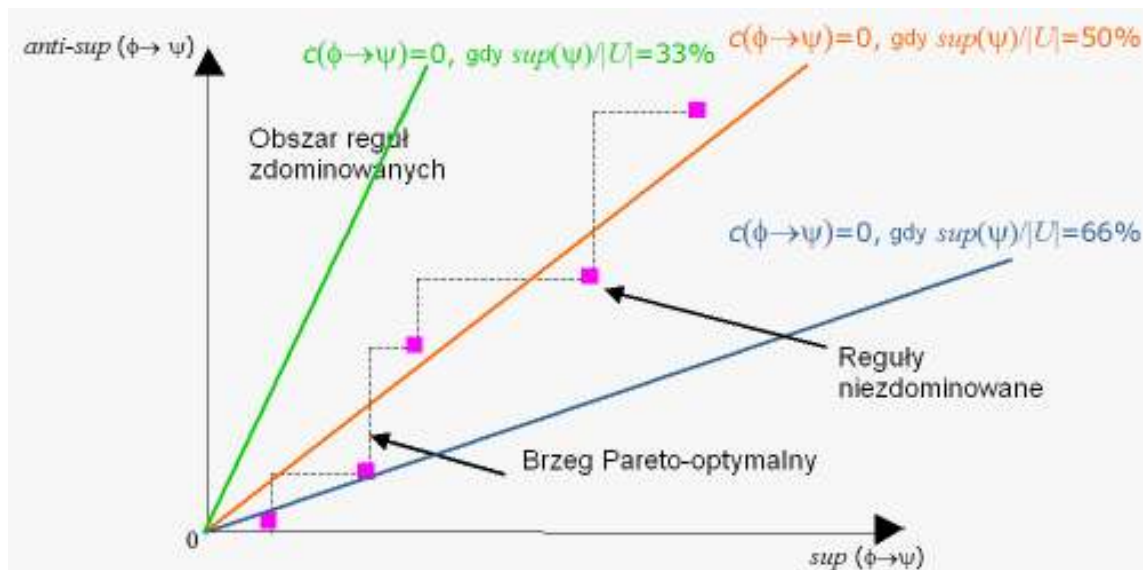
5.1. Nałożenie semantyki confirmacji na przestrzeń wsparcie–anty-wsparcie

Zainspirowani wartością semantyki jaką niesie za sobą własność confirmacji przeprowadziliśmy próbę wyodrębnienia z przestrzeni wsparcie–anty-wsparcie reguł z ujemną bądź niską wartością miary confirmacji f . Przeprowadzona analiza wykazała, że w ramach reguł z tą samą konkluzją, reguły charakteryzujące się ujemną wartością

dowolnej miary $c(\phi \rightarrow \psi)$ z własnością konfirmacji znajdują się pod następującą krzywą [80], [81]:

$$c(\phi \rightarrow \psi) \geq 0 \Leftrightarrow \text{anti-sup}(\phi \rightarrow \psi) \leq \text{sup}(\phi \rightarrow \psi) \left[\frac{|U|}{\text{sup}(\psi)} - 1 \right]. \quad (5.1)$$

Rysunek 5.2 jest ilustracją powyższego wyniku analitycznego, a Tabela 5.1 zbiera informacje o liczbie reguł z niedodatnią konfirmacją (tzn. reguł nieinteresujących) w zbiorach reguł niezdominowanych dla różnych konkluzji.



Rysunek 5.2 Przykłady trzech funkcji liniowych, dla których dowolna miara konfirmacji $c(\phi \rightarrow \psi) = 0$. Każda z prostych reprezentuje przebieg funkcji dla zbioru o innej liczności

Tabela 5.1 Reguły z niedodatnią wartością miary f dla zbiorów Pareto dla różnych konkluzji reguł ze zbioru *census* (min. wsparcie=0,15)

Badana konkluzja	Lb. reguł w zbiorze Pareto	Lb. reguł z niedodatnią konfirmacją	Wielkość redukcji
workclass='Private'	17	4	24%
sex=Male	8	3	38%
income<=50 kUSD	15	5	33%
race=White	17	1	0.6%
native_country=USA	15	0	0%

Uogólnienie wyniku (5.1) dla wartości konkretnej miary konfirmacji (w tym wypadku f) nie większej niż pewne k przedstawia Twierdzenie 5.4.

Twierdzenie 5.4. [81]

$$f(\phi \rightarrow \psi) \geq k \Leftrightarrow \text{anti-sup}(\phi \rightarrow \psi) \leq \text{sup}(\phi \rightarrow \psi) \frac{1-k}{(1+k)\text{sup}(\psi)}. \quad (5.2)$$

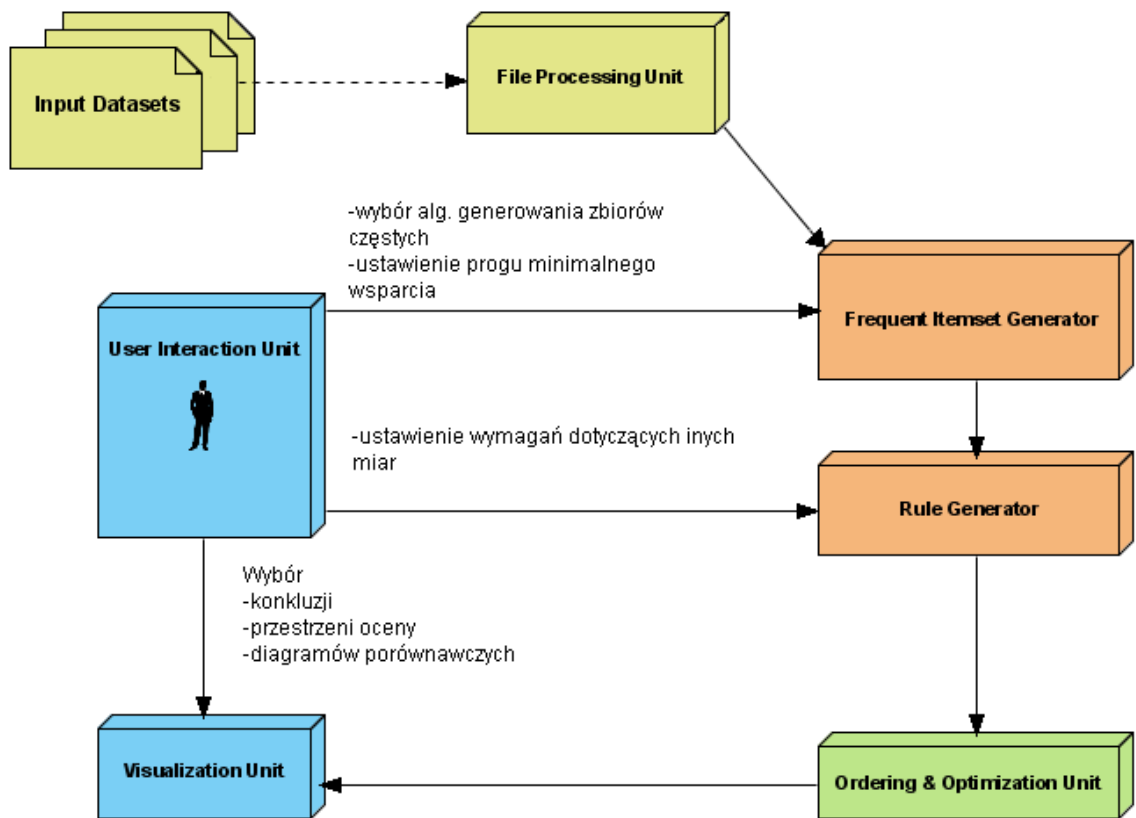
6. Przykłady zastosowania miar atrakcyjności w wielokryterialnej ocenie reguł

6.1. Koncepcja systemu wielokryterialnej oceny reguł

System składa się z następujących modułów:

- *Moduł przetwarzania danych* (File Processing Unit) – odpowiedzialny za wczytywanie i przetwarzanie danych z plików wejściowych,
- *Generator zbiorów częstych* (Frequent Itemset Generator) – wyszukujący wszystkie zbiory częste (tj. z wartością wsparcia nie mniejszą niż zadany przez użytkownika próg minimalnego wsparcia) z danych przy użyciu algorytmu Apriori [2] lub FP Growth [30],
- *Generator reguł* (Rule Generator) – tworzy ze zbiorów częstych reguły spełniające opcjonalnie podane przez użytkownika warunki dotyczące np. maksymalnego dopuszczalnego anty-wsparcia. Przykład reguł tworzonych przez system przedstawia Rysunek 6.2 ,
- *Moduł sortujący i optymalizujący* (Ordering and Optimization Unit) – odpowiadający za uporządkowanie reguł, umożliwiające zoptymalizowanie procesu poszukiwania zbiorów reguł niezdominowanych względem wsparcia i anty-wsparcia oraz wykorzystanie Twierdzenie 5.1 oraz Twierdzenie 5.2 do podniesienia efektywności znajdowania reguł Pareto-optymalnych bądź optymalnych w innych przestrzeniach oceny.
- *Moduł wizualizacji* (Visualization Unit) – prezentujący użytkownikowi reguły w wybranej przestrzeni ocen. Użytkownik ma możliwość obserwowania reguł z ujemną bądź dodatnią wartością confirmacji, może zawęzać zbiór reguł jedynie do reguł Pareto-optymalnych, oraz manipulować progami minimalnych bądź maksymalnych wartości miar. Moduł wizualizacji prezentuje reguły na pojedynczych wykresach oraz na porównawczych zestawieniach wszystkich przestrzeni oceny.
- *Moduł interakcji z użytkownikiem* (User Interaction Unit) – odpowiada za przekazywanie pozostałym modułom parametrów zadawanych przez użytkownika.

Schemat współpracy modułów przedstawia Rysunek 6.1



Rysunek 6.1 Diagram komponentów systemu

premise	conclusion	supp	conf	s	f	a-supp
native-country is United-States	education is Bachelors	0,15	0,17	0,01	0,00	0,76
education is Bachelors	native-country is United-States	0,15	0,92	0,00	0,02	0,01
workclass is Private	education is Some-college	0,17	0,23	0,01	0,01	0,57
education is Some-college	workclass is Private	0,17	0,75	0,02	0,03	0,06
native-country is United-States	workclass is Private and education is Some-college	0,16	0,17	0,04	0,01	0,76
native-country is United-States and workclass is Private	education is Some-college	0,16	0,23	0,03	0,03	0,51

Rysunek 6.2 Przykładowe reguły wygenerowane ze zbioru *census*

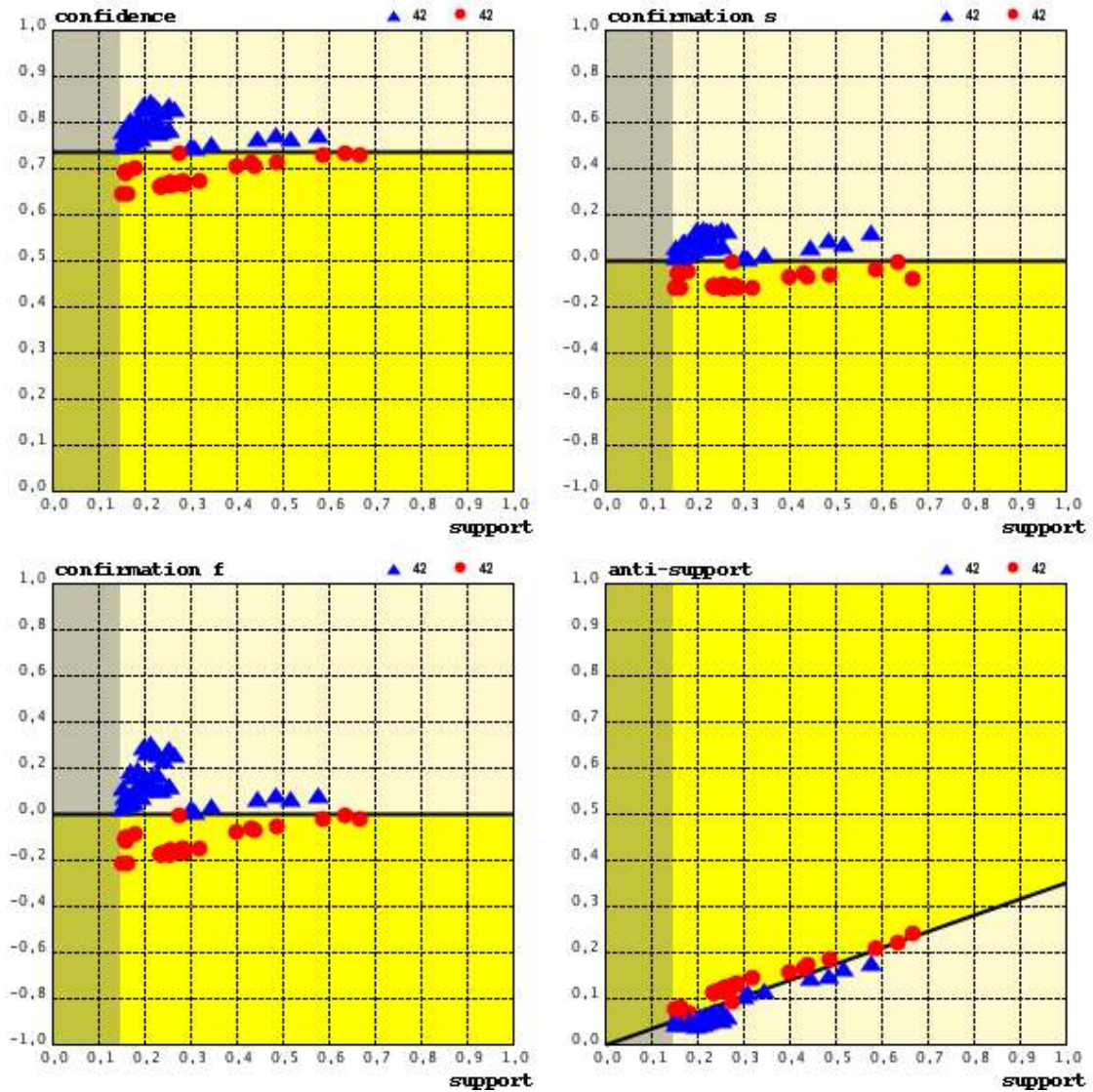
6.2. Przykłady zastosowania systemu

6.2.1. Zbiór danych *census*

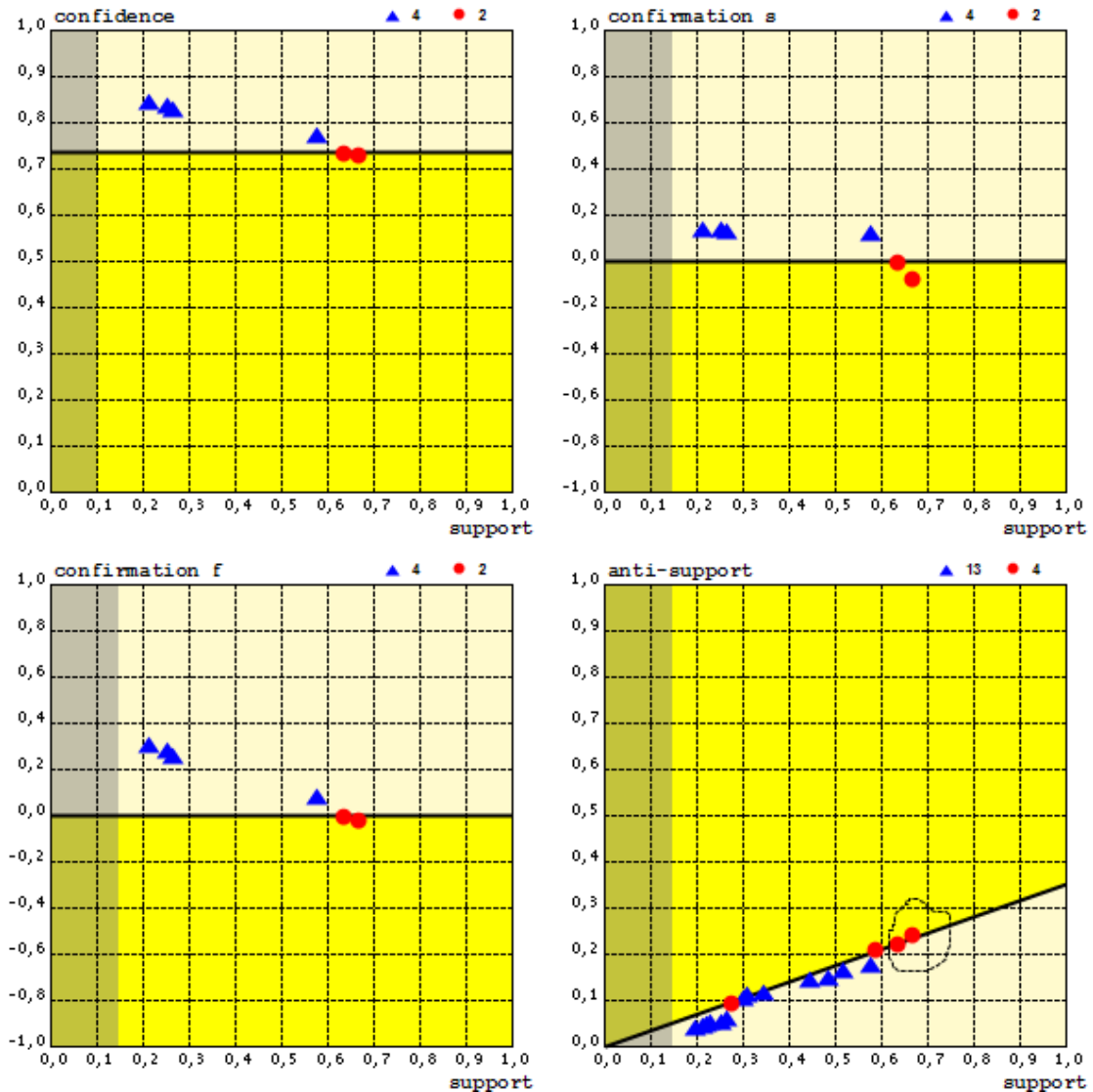
Zbiór *census* jest podzbiorem zbioru przygotowanego przez Kohavi i innych [45], zawierającym 32 561 obiektów, opisanych na 9 atrybutach nominalnych (m.in. wykształcenie, zawód, etc).

Przykładowe zastosowanie systemu do analizy zbioru *census* przedstawiają poniższe diagramy. Opredstawia w różnych przestrzeniach oceny wszystkich (ponad 2200) reguły wygenerowane przy minimalnym progu wsparcia 0,15 dla konkluzji `workclass='Private'`. Semantyka konfirmacji nałożona jest na wszystkie przestrzenie w postaci czarnych prostych oddzielających reguły z niedodatnią konfirmacją (czerwone koła) od reguły z dodatnią konfirmacją (niebieskie trójkąty). Można zaobserwować, że co druga reguła dla tej konkluzji powinna być wyeliminowana, gdyż jej przesłanka zaprzecza konkluzji.

System umożliwia także porównawcze analizowanie reguły z brzegów Pareto- optymalnych w różnych przestrzeniach, jak pokazuje Rysunek 6.4 . Użytkownik może porównać liczności zbiorów reguły niezdominowanych oraz analizować różnice między regułami wchodzącymi w skład tych zbiorów. Przykładowo, Rysunek 6.4 wyodrębnia zakreśleniem reguły, które są na brzegu Pareto w przestrzeni wsparcie–anty-wsparcie, a których nie ma na innych diagramach.



Rysunek 6.3 Diagram porównawczy różnych przestrzeni ocen dla wszystkich reguł z konkluzją `workclass="Private"` wygenerowanych przy minimalnym wsparciu 0,15

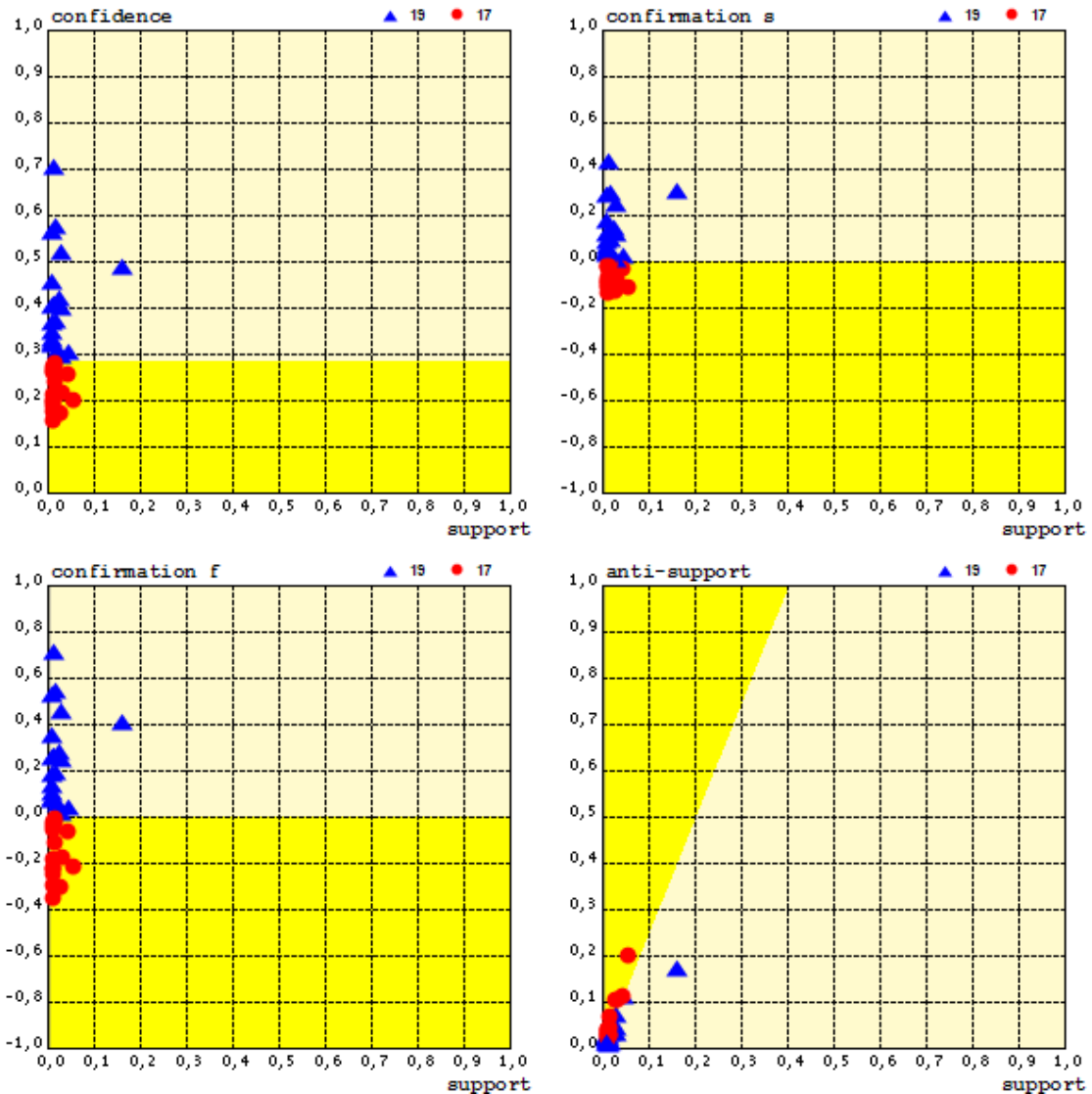


Rysunek 6.4 Zestawienie porównawcze zbiorów reguł niezdominowanych w różnych przestrzeniach wygenerowanych dla konkluzji `workclass="Private"` przy minimalnym wsparciu 0,15

6.2.2. Zbiór danych MSweb

Zbiór danych *msweb* [61] jest rzeczywistym anonimowym dziennikiem serwera WWW Microsoft, opisującym dostępy użytkowników wykonane w jednym tygodniu lutego '98. Obejmuje on 294 różnych adresów URL zorganizowanych w 32 711 transakcji (średnio trzejelementowych).

Msweb jest zbiorem rzadkim i dlatego próg minimalnego wsparcia musiał być bardzo zaniżony (nawet do 0,1).

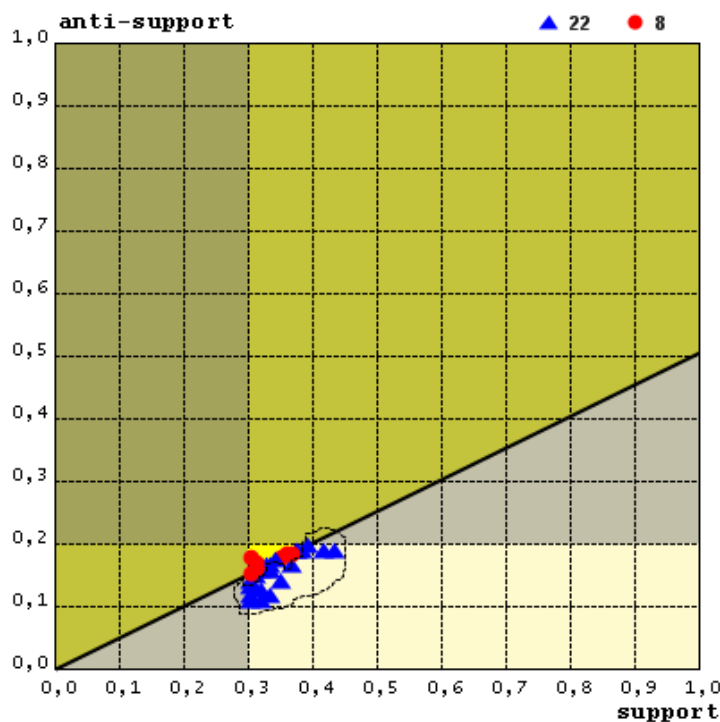


Rysunek 6.5 Diagram porównawczy różnych przestrzeni ocen dla wszystkich reguł z konkluzją `web_site="Internet Explorer"` wygenerowanych przy minimalnym wsparciu 0,005

Rysunek 6.5 przedstawia różne przestrzenie oceny reguł dla jednej z największych klas: `web_site='Internet Explorer'`, która reprezentuje transakcje, w których anonimowi użytkownicy odwiedzili stronę Internet Explorera. Analiza diagramu pozwoliła stwierdzić, że dla reguł z tą konkluzją istnieje tylko jedna reguła o wsparciu większym niż 0,1. Ponadto, w rozważanej klasie ponad 47% wygenerowanych reguł jest mylących, z racji że ich przesłanki przeczą konkluzji (co jest odzwierciedlone ujemną wartością dowolnej miary konfirmacji).

6.2.3. Zbiór danych HSV

Zbiór *HSV* [64] zawiera 122 opisy pacjentów z chorobą wrzodową dwunastnicy leczonych metodą wysoce wybiórczej wagotomii. Dane zostały zebrane przez K. Słowińskiego w jednym z poznańskich szpitali na podstawie historii chorób pacjentów. Każdy z pacjentów opisany jest za pomocą 11 atrybutów warunkowych i przydzielony według atrybutu decyzyjnego do jednej z czterech klas wyrażających skuteczność zabiegu chirurgicznego. Przeprowadzone na zbiorze HSV eksperymenty dotyczyły klasy: skuteczność_leczenia = „doskonała” i miały na celu znalezienie w przestrzeni wsparcie–anty-wsparcie obszaru reguł interesujących ze względu na wsparcie, anty-wsparcie, confirmację i własność M. Ogólnie, możliwych reguł decyzyjnych opisujących badaną klasę jest ponad 2000. Dostosowując progi minimalnego wsparcia i maksymalnego dopuszczalnego anty-wsparcia zakreślony został na Rysunek 6.1 obszar reguł interesujących obejmujący 22 reguły z dodatnią confirmacją. W zbiorze tym znajdują się też reguły optymalne ze względu na miary z własnością M. Obszar reguł interesujących nie został zawężony jedynie do reguł Pareto-optymalnych, gdyż z punktu widzenia pokrycia badanej klasy również reguły zdominowane mogą być interesujące i przydatne dla decydenta.



Rysunek 6.1 Obszar reguł interesujących, z dodatnią confirmacją w przestrzeni wsparcie–anty-wsparcie (skuteczność_leczenia='doskonała')

7. Podsumowanie

Celem pracy była analiza związków zachodzących pomiędzy popularnymi miarami atrakcyjności (wsparcie, anty-wsparcie, ufność, funkcja atrakcyjności reguły, funkcja zysku, współczynnik zależności oraz miary f i s) oraz zaproponowanie takiej wielokryterialnej przestrzeni oceny reguł, że zbiór reguł niezdominowanych w tej przestrzeni zawiera reguły optymalne ze względu na dowolną miarę atrakcyjności posiadającą własność M.

Następujące wyniki szczegółowe umożliwiły osiągnięcie postawionego celu:

1. Przeprowadzono analizę takich miar atrakcyjności jak wsparcie, anty-wsparcie, ufność, funkcja atrakcyjności reguły, funkcja zysku, współczynnik zależności oraz miary f i s , względem własności M, własności Bayesowskiej confirmacji oraz własności symetrii hipotetycznej. Wykazano, że spośród badanych miar jedynie miara dependency factor nie posiada własności M (Twierdzenie 3.1 - Twierdzenie 3.4). Ponadto pokazano, że miary wsparcia, anty-wsparcia i ufności nie mogą być postrzegane jako miary confirmacji, podczas gdy miara *gain* charakteryzuje się własnością confirmacji tylko w pewnych warunkach (Twierdzenie 3.5 - Twierdzenie 3.7). W zakresie analizy własności symetrii hipotetycznej, wykazano, że do grona miar posiadających tę własność dołączyły tylko miary *RI* i *gain* (Twierdzenie 3.8 - Twierdzenie 3.10).
2. Dla reguł z tą samą konkluzją, zanalizowano związki zachodzące pomiędzy badanymi miarami atrakcyjności oraz relacje zawierania się zbiorów reguł niezdominowanych z różnych przestrzeni oceny reguł. Wykazano, że przy stałej wartości miary wsparcia, pomiędzy miarą ufności a miarami f i s występuje monotoniczna zależność (Twierdzenie 4.2, Twierdzenie 4.5), co zostało również uogólnione na dowolną miarę z własnością M (Twierdzenie 4.9). Ponadto, przy stałej wartości ufności wykazano występowanie zależności monotonicznej między miarą wsparcia a miarami f , s (Twierdzenie 4.1, Twierdzenie 4.4) oraz dowolną miarą z własnością M (Twierdzenie 4.8). Dodatkowo wykazano również, że pomiędzy miarami z własnością M a wsparciem (anty-wsparciem) występuje zależność monotoniczna (anty-monotoniczna) (Twierdzenie 5.1, Twierdzenie 5.2). Analiza zawierania się zbiorów reguł niezdominowanych

określonych w różnych przestrzeniach oceny reguł wykazała, że zbiory Pareto-optymalne w przestrzeniach wsparcie–ufność oraz wsparcie– f są identyczne, oraz że są nadzbiorem zbioru Pareto-optimalnego w przestrzeni wsparcie– s . Nadzbiorem wszystkich tych zbiorów jest natomiast zbiór reguł niezdominowanych w przestrzeni wsparcie–anty-wsparcie (Twierdzenie 4.6, Twierdzenie 4.7, Twierdzenie 5.3).

3. Zaproponowano przestrzeń oceny reguł, opartą na miarach wsparcia i anty-wsparcia, w której zbiór reguł niezdominowanych zawiera wszystkie reguły optymalne ze względu na dowolną miarę atrakcyjności posiadającą własność M (Rozdział 5).
4. Nałożono na przestrzenie wsparcie–ufność i wsparcie–anty-wsparcie semantykę konfirmacji. Wykazano, że liniowe funkcje wyodrębniają w tych przestrzeniach reguły z dodatnią wartością dowolnej miary konfirmacji (Twierdzenie 4.3, Twierdzenie 5.4). Wyniki te pozwalają zawęzić zbiór wygenerowanych reguł do zbioru wartościowych reguł, w których przesłanka odpowiednio potwierdza konkluzję.
5. Opracowano i zaimplementowano system wielokryterialnej oceny reguł (Rozdział 6). Oparty jest on na modyfikacji schematu apriori umożliwiającej generowanie reguł z uwzględnieniem różnych miar atrakcyjności. Działanie systemu zostało zaprezentowane na przykładzie trzech zbiorów danych: *census*, *msweb* oraz *hsv*.

Wśród kierunków dalszych badań, w pierwszej kolejności planowane jest:

- rozszerzenie badań na inne miary oceny atrakcyjności i inne własności miar,
- opracowanie algorytmu znajdującego w przestrzeni wsparcie–anty-wsparcie reguł pokrywających w zadanym procencie obiekty ze zbioru danych.

Bibliografia

- [1] Azevedo, P.J., Jorge, A.M., 2007. Comparing rule Measures for Predictive Association Rules, in: Proc of ECML Warsaw 2007, LNAI 4701, Springer-Verlag Berlin Heidelberg, pp.510-517.
- [2] Agrawal, R., Imielinski, T., Swami, A., 1993. Mining associations between sets of items in massive databases, in: Proc. of the 1993 ACM-SIGMOD Int'l Conf. on Management of Data, 207-216.
- [3] Bayardo, R.J., Agrawal, R., 1999. Mining the most interesting rules, in: Proc. of the Fifth ACM-SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining, 145-154. Fayyad U.M., Piatetsky-Shapiro G., Smyth P., Uthurusamy R. (red.), *Advances in Knowledge Discovery and Data Mining*, AAAI/MIT Press, Cambridge Mass. 1996.
- [4] Bramer, M., 2007. *Principles of Data Mining*, Springer-Verlag, New York Inc.
- [5] Briand, L., El Emam, K., Morasc, S., 1995. On the Application of Measurement Theory in Software Engineering, in: *Empirical Software Engineering*, 1, 61-88.
- [6] Brin, S., Motwani, R., Ullman, J. and Tsur, S., 1997. Dynamic itemset counting and implication rules for market basket data, in: Proc. of the 1997 ACM-SIGMOD Int'l Conf. on the Management of Data, 255-264.
- [7] Brzezinska, I., Slowinski, R., 2005. Monotonicity of a Bayesian confirmation measure in rule support and confidence, in: *Recent Developments in Artificial Intelligence Methods*, AI-METH Series, Gliwice, 39-42.
- [8] Brzezinska, I., Greco, S., Slowinski, R., 2005. Investigation of monotone link between confirmation measures and rule support and confidence. Research Report RA-025/05, Institute of Computing Science, Poznan University of Technology, Poznan.
- [9] Brzezińska, I., Greco, S., Słowiński, R., 2007. Mining Pareto-optimal rules with respect to support and anti-support. *Engineering Applications of Artificial Intelligence* Vol 20(5) pp.587-600.
- [10] Carnap, R., 1962. *Logical Foundations of Probability*, second ed, University of Chicago Press, Chicago.
- [11] Christensen, D., 1999. Measuring confirmation. *Journal of Philosophy* XCVI, 437-461.

- [12] Clark, P., Boswell, P., 1991. Rule induction with CN2: some recent improvements, in: *Machine Learning: Proc. of the Fifth European Conference*, 151-163.
- [13] Cichosz, P., 2000. *Systemy uczące się*, Warszawa, WNT.
- [14] Cios, K., Pedrycz, W., Świniarski, R., 1999. *Data mining methods for knowledge discovery*, Kluwer Academic Publishers.
- [15] Crupi, V., Tentori, K., Gonzalez, M., (to appear). On Bayesian Theories of Evidential Support: Theoretical and empirical issues. *Philosophy of Science*.
- [16] Dhar, V., Tuzhilin, A., 1993. Abstract-driven pattern discovery in databases. *IEEE Transactions on Knowledge and Data Engineering*, 5(6).
- [17] Earman, J., 1992. *Bayes or Bust: A Critical Examination of Bayesian Confirmation Theory*. MIT Press, Cambridge, MA.
- [18] Eells, E., 1982. *Rational Decision and Causality*. Cambridge University Press, Cambridge.
- [19] Eells, E., Fitelson, B., 2002. Symmetries and asymmetries in evidential support. *Philosophical Studies*, 107 (2): 129-142.
- [20] Fayyad U.M., Piatetsky-Shapiro G., Smyth P., Uthurusamy R. (Eds.), 1996. *Advances in Knowledge Discovery and Data Mining*, AAAI/MIT Press, Cambridge Mass.
- [21] Fayyad U.M., Piatetsky-Shapiro G., Smyth P., Uthurusamy R., 1999. From data mining to knowledge discovery, in [20], pp.1-36.
- [22] Fitelson, B., 2001. *Studies in Bayesian Confirmation Theory*. Ph.D. Thesis, University of Wisconsin, Madison.
- [23] Francisci, D., Collard, M., 2003. Multi-criteria evaluation of interesting dependencies according to a data mining approach. in: *Congress on Evolutionary Computation*, Canberra, Australia, IEEE Computer Society Press, Los Alamitos, 12 pp.1568-1574.
- [24] Frawley, W.J., Piatetsky-Shapiro, G., Matheus, C.J., 1991. Knowledge discovery in databases: an overview, in: *Knowledge Discovery in Databases*, Eds. Piatetsky-Shapiro G., Frawley W.J., AAAI/MIT Press, Cambridge, MA, 1-27.
- [25] Fukuda, T., Morimoto, Y., Morishita, S., Tokuyama, T., 1996. Data mining using two-dimensional optimized association rules: scheme, algorithms, and visualization, in: *Proc. of the 1996 ACM-SIGMOD Int'l Conf. on the Management of Data*, 13-23.

- [26] Geng, L., Hamilton, H.J, 2006. Interestingness measures for data mining: A survey, *ACM Computing Surveys*, 38(3).
- [27] Good, I.J., 1984. The best explicatum for weight of evidence. *Journal of Statistical Computation and Simulation*, 19: 294-299.
- [28] Greco, S., Pawlak, Z., Słowiński, R., 2004. Can Bayesian confirmation measures be useful for rough set decision rules? *Engineering Applications of Artificial Intelligence*, 17: 345-361.
- [29] Greco, S., Słowiński, R., Szczęch, I., 2007. Analysis of monotonicity properties of some rule interestingness measures. in: *Materiały II Krajowej Konferencji nt. Technologie Przetwarzania Danych*. Poznań, 24-26.09.2007, Wydawnictwo Politechniki Poznańskiej, Poznań, 151-161.
- [30] Han, J., Pei, J., Yin, Y., 2000. Mining frequent patterns without candidate generation. in: *Proc. ACM SIGMOD Conference on Management of Data*, 1-12.
- [31] Han, J., Kamber, M., 2000. *Data mining: Concepts and techniques*, San Francisco, Morgan Kaufman Pub.
- [32] Hebert, C., Cremilleux, B., 2007. A Unified View of Objective Interestingness Measures, in: *Proc. of the 2007 MLDM Int'l Conf. on Machine Learning and Data Mining*, LNAI 4571, Springer-Verlag pp. 533-547.
- [33] Heckerman, D, 1988. An axiomatic framework for belief updates, in: Kanal, L., Lemmer, J. (Eds), *Unconfidence in Artificial Intelligence 2*. Elsevier Science Publishers, New York, 11-22.
- [34] Hempel, C.G., 1945. Studies in the logic of confirmation (I). *Mind* 54, 1-26.
- [35] Hilderman, R. and Hamilton, H.: *Knowledge Discovery and Measures of Interest*. Kluwer Academic Publishers (2001).
- [36] Horwich, P., 1982. *Probability and Evidence*. Cambridge University Press, Cambridge.
- [37] Horvitz, E., Heckerman, D., 1986. The inconsistent use of certainty measures in artificial intelligence research, in: Kanal L., Lemmer J. (Eds), *Uncertainty in Artificial Intelligence (1)*. Elsevier Science Publishers, New York, pp. 137-151.
- [38] International Business Machines, 1996. *IBM Intelligent Miner User's Guide*, Version 1, Release 1.
- [39] Jeffrey, H., 1935. Some tests of significance treated by theory of probability, in: *Proceedings of the Cambridge Philosophical Society*, 31: 203-222.

- [40] Joyce, J., 1999. *The Foundations of Causal Decision Theory*. Cambridge University Press, Cambridge.
- [41] Kamber, M., Shingal, R., 1996. Evaluating the interestingness of characteristic rules, in: *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96)*, Portland, Oregon, pp. 263–266.
- [42] Keynes, J., 1921. *A Treatise on Probability*. Macmillan, London.
- [43] Klemettinen, M., Mannila, H., Ronkainen, P., Toivonen, H., Verkamo, A.I., 1994. Finding interesting rules from large sets of discovered association rules, in: *Proc. of Third Int'l Conf. on Information and Knowledge Management (CIKM'94)*. ACM Press, pp. 401-407.
- [44] Kemeny, J., Oppenheim, P., 1952. Degrees of factual support. *Philosophy of Science* 19, 307-324.
- [45] Kohavi, R., 1996. Scaling up the accuracy of Naive–Bayes Classifiers: a decision–tree hybrid. [In]: *Proc. of the 2nd Int'l Conf. on Knowledge Discovery and Data Mining*.
- [46] Kubat, M., Bratko, I., Michalski, R.S., Review of machine learning methods in [51] pp. 3-70.
- [47] Kyburg, H., 1983. Recent work in inductive logic, in: *Recent Work In Philosophy*. Lanham: Rowman & Allanheld, pp. 87–150.
- [48] Lenca, Ph., Meyer, P., Vaillant, B., Lallich, S., (to appear in Jan. 2008). On selecting interestingness measures for association rules: User oriented description and multiple criteria decision aid. *European Journal of Operational Research*, Vol. 184, Issue 2, 610-626.
- [49] Mackie, J.L., 1969. The relevance criterion of confirmation. *The British Journal for the Philosophy of Science* 20, 27-40.
- [50] McGarry, K., 2005. A survey of interestingness measures for knowledge discovery, in: *The Knowledge Engineering Review*, vol. 20:1, Cambridge University Press, pp.39-61.
- [51] Michalski, R.S., 1983. A theory and methodology of inductive learning, in: *Machine Learning: An Artificial Intelligence Approach*, Eds: Michalski R.S., Carbonell J.G., Mitchell T.M., Morgan Kaufman, 83-134.
- [52] Michalski, R.S., 1997. *Machine Learning, Data Mining and Knowledge Discovery. Principles and Applications*, in: *Tutorials of Intelligent Information Systems, IIS'97*, Zakopane, IPI PAN Press.

- [53] Michalski, R.S., Bratko, I., Kubat, M (Eds), 1998. Machine learning and data mining, John Wiley & Sons.
- [54] Michell, J., 1986. Measurement scales and statistics: a clash of paradigms, in: Psychological Bulletin, 3, 398-407.
- [55] Morimoto, Y., Fukuda, T., Matsuzawa, H., Tokuyama, T. and Yoda, K., 1998. Algorithms for mining association rules for binary segmentation of huge categorical databases, in: Proc. of the 24th Very Large Data Bases Conf., 380-391.
- [56] Morishita, S., 1998. On classification and regression, in: Proc. of the First Int'l Conf. on Discovery Science – Lecture Notes in Artificial Intelligence, 1532: 40-57.
- [57] Morzy, T., 2004. *Odkrywanie asocjacji: algorytmy i struktury danych*. Ośrodek Wydawnictw Naukowych, Poznań.
- [58] Morzy, T., Zakrzewicz, M., 2003. Data mining, in: Handbook on data management in information systems, Eds. Błażewicz J., Kubiak W., Morzy T., Rusinkiewicz M.E., Springer-Verlag, 487-565.
- [59] Morzy, M., 2006. Eksploracja danych - przegląd dostępnych metod i dziedzin zastosowań, in: VI edycja Hurtownia danych i business intelligence, Centrum Promocji Informatyki, Warszawa, Poland.
- [60] Morzy, M., 2005. Oracle Data Mining - odkrywanie wiedzy w dużych wolumenach danych, in: XI Krajowa Konferencja PLOUG'2005 "Systemy informatyczne. Projektowanie, implementowanie, eksploatawanie", Zakopane, Poland, 18-21.
- [61] Msweb dataset: <http://kdd.ics.uci.edu/databases/msweb/msweb.html>.
- [62] Nicod, J., 1923. Le probleme de la logique de l'induction. Alcan, Paris.
- [63] Pawlak, Z., 1982. Rough Sets. Int. Journal of Computer and Information Science, 11, 341-356.
- [64] Pawlak, Z., Słowiński, K., Słowiński, R., 1986. Rough classification of patients after highly selected vagotomy for duodenal ulcer, *International J. Man-Machine Studies*, 24, pp. 413-433.
- [65] Pawlak, Z., 2004. Some Issues on Rough Sets. *Transactions on Rough Sets I*, LNCS 3100, 1-58.
- [66] Pearl, J., 1988. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufman, San Francisco.
- [67] Piatetsky-Shapiro, G., 1991. Discovery, analysis, and presentation of strong rules. Chapter 12, in: Knowledge Discovery in Databases, AAAI/MIT Press.
- [68] Pollard, S., 1999. Milne's measure of confirmation. *Analysis* 59, 335-337.

- [69] Popper, K.R., 1959. *The Logic of Scientific Discovery*. Hutchinson, London.
- [70] Quinlan J.R., 1993. *C4.5: Programs for Machine Learning*, San Francisco, Morgan Kaufmann.
- [71] Quinlan J.R., 1986. Induction of decision trees. *Machine Learning* 1(1), pp81-106.
- [72] Reiter, R., 1978. On closed world data bases, in: *Logic and Data Bases*, H. Gallaire and J. Minker (eds), Plenum., New York pp. 119-140.
- [73] Rosenkrantz, R., 1994. Bayesian confirmation: paradise regained. *The British Journal for the Philosophy of Science* 45, 467-476.
- [74] Schlesinger, G., 1995. Measuring degrees of confirmation. *Analysis* 55, 208-212.
- [75] Schum, D., 1994. *The Evidential Foundations of Probabilistic Reasoning*. Wiley, New York.
- [76] Silberschatz, A., Tuzhilin, A., 1996. What Makes Patterns Interesting in Knowledge Discovery Systems, in: *IEEE Transactions on Knowledge and Data Engineering*, vol. 8 no. 6, pp. 970-974.
- [77] Słowiński, R., 2000. Rozszerzenie teorii zbiorów przybliżonych na atrybuty ze skalą preferencji, w: *Materiały konferencji Informatyka teoretyczna: metody analizy informacji niekompletnej i rozproszonej*, Białystok, 114-128.
- [78] Słowiński, R. and Greco, S., 2005. Measuring attractiveness of rules from the viewpoint of knowledge representation, prediction and efficiency of intervention, in: *Szczepaniak, P., Kacprzyk, J., Niewiadomski, A. (Eds.), Advances in Web Intelligence. LNAI, vol 3528, Springer-Verlag, Berlin, 1-12.*
- [79] Słowiński, R., Brzezińska, I., Greco, S., 2006. Application of Bayesian Confirmation Measures for Mining Rules from Support - confidence Pareto-optimal set, in: *Proc. of the 8th International Conference on Artificial Intelligence and Soft Computing. LNCS, Springer-Verlag, Berlin 1018-1026.*
- [80] Słowiński, R., Szczęch, I., Urbanowicz, M., Greco, S., 2006. Experiments with induction of association rules with respect to support and anti-support. *Research Report RA-018/06, Institute of Computing Science, Poznań University of Technology, Poznań.*
- [81] Słowiński, R., Szczęch, I., Urbanowicz, M., Greco, S., 2007. Mining Association Rules with respect to Support and Anti-support - Experimental Results, in: *Proc. of RSEISP2007, The Rough Sets and Emerging Systems Paradigms 2007 International Conference. LNAI, vol. 4585, Springer-Verlag, Berlin, 534-543.*

- [82] Stefanowski J., 2001. *Algorytmy indukcji reguł decyzyjnych w odkrywaniu wiedzy*, Rozprawa habilitacyjna, Politechnika Poznańska, Instytut Informatyki, Poznań.
- [83] Stevens, S.S., 1946. On the theory of scales of measurement, in: *Science*, 103, 677-680.
- [84] Szczepaniak, P.S., Segovia, J., Kacprzyk, J., Zadeh, L.A. (Eds.): *Intelligent Exploration of the Web*. Physica Verlag, c/o Springer-Verlag, Heidelberg, New York, 2003.
- [85] Szczepaniak, P.S., 2004. *Obliczenia inteligentne, szybkie przekształcenia i klasyfikatory*. Akademicka Oficyna Wydawnicza EXIT, Warszawa.
- [86] Tan, P.-N., Kumar, V., Srivastava, J., 2004. Selecting the right objective measure for association analysis. *Inf.Syst.*29(4), pp.293-313.
- [87] Webb, G.I., 1995. OPUS: An efficient admissible algorithm for unordered search. *Journal of Artificial Intelligence Research*, 3: 431-465.
- [88] Yao, Y.Y., Zhong, N., 1999. An analysis of quantitative measures associated with rules, in: *Proceedings of the Third Pacific-Asia Conference on Knowledge Discovery and Data Mining*, LNAI vol. 1574, Springer, pp. 479-488.
- [89] Zakrzewicz M., 1997. *Data Mining i odkrywanie wiedzy w bazach danych*, in: *Materiały konferencyjne III Konferencji Polskiej Grupy Użytkowników Systemu Oracle*, Zakopane.
- [90] Zakrzewicz M., 2004. *Optymalizacja wykonania zapytań eksploracyjnych w systemach baz danych*, Rozprawa habilitacyjna, Politechnika Poznańska, Instytut Informatyki, Poznań.
- [91] Zembrowicz, R., Żytkow, J., From contingency tables to various forms of knowledge in databases, in [20], pp. 329-352.