Rough Set Approach to Ordinal Classification with Monotonicity Constraints

Marcin Szeląd

Institute of Computing Science, Poznań University of Technology

10.1.2013
Outline

1. Introduction
   - Problem Statement
   - Motivations for Research

2. Dominance-based Rough Set Approach

3. Decision Rules

4. VCDomLEM Algorithm

5. Computational Experiment

6. Conclusions
Introduction
There is given a finite set $U$ of objects described on attributes from finite set $A = C \cup D$, where $C$ denotes a set of condition attributes, $D$ denotes a set of decision attributes, and $C \cap D = \emptyset$.

Decision attributes make a partition of set $U$ into a finite number of $n$ disjoint sets of objects, called decision classes. We denote this partition, also called classification, by $\mathcal{X} = \{X_1, \ldots, X_n\}$. 
Problem Statement

Ordinal classification problem with monotonicity constraints

- Relationship between description of objects by some of condition attributes and values of decision attribute.
- Knowledge about orders (of preferences) on the value sets of the attributes – criteria.
- **Semantic correlation**: a better evaluation of an object on a condition attribute with other evaluations being fixed should not worsen its evaluation on decision attribute.
- **Inconsistency**: violation of monotonicity constraints (expressed by semantic correlation).

Examples of monotonicity constraints

- “The lower the price and the higher the quality, the higher the customer’s satisfaction”.
- “The higher the mass and the lower the distance, the higher the gravity”.
The goal is to build a classifier which, given the evaluations of object $y \in X_i$ on condition attributes, suggests assignment of $y$ to class $X_i$.

A good classifier is characterized by:
- high accuracy,
- interpretability (i.e., it preserves monotonicity constraints),
- traceability (glass box).

We estimate accuracy of prediction of the classifier by:
- the percentage of correctly classified objects (PCC),
- the mean absolute difference between index of the class suggested by a classifier and index of the class to which an object belongs; this measure is called mean absolute error (MAE).
Multi-attribute-multi-criteria classification is an important, non-trivial and practical problem.

Main difficulty consists in aggregation of different and usually conflicting attributes/criteria; usually such aggregation is performed arbitrary, using weights or aggregation operators like sum, average or distance metrics.

Need for modelling method that allows to include domain knowledge, can handle possible inconsistencies in data, and avoids any aggregation operators.
Motivations for Application of DRSA

Ordinal classification problem with monotonicity constraints can be effectively solved using Dominance-based Rough Set Approach (DRSA), which:

- can handle inconsistencies in data (preprocessing), resulting e.g. from imprecise of incomplete information,
- takes into account domain knowledge:
  - domains of attributes, i.e. sets of values that an attribute may take while being meaningful for user’s perception,
  - division of attributes into condition and decision attributes,
  - preference order in domains of attributes and semantic correlation between attributes, both addressed by the dominance principle,
- works with heterogeneous attributes – nominal, ordinal and cardinal (no need of discretization),
- enables to infer decision rule model from decision table (disaggregation-aggregation paradigm).
Motivations for Using Decision Rule Model

Advantages of decision rules:
- comprehensible form of knowledge representation,
- can represent any function (more general than utility functions or binary relations),
- resistant to irrelevant attributes,
- do not require aggregation operators,
- support “backtracing”,
- can explain past decisions and predict future decisions.
Dominance-based Rough Set Approach
In rough set approaches, learning of a classifier is preceded by rough set analysis of data presented as decision table. It consists in checking the data for possible inconsistencies by calculation of lower approximations of considered sets of objects.

Due to this type of data structuring, one may restrict a priori the set of objects on which the classifier is learned to a subset of sufficiently consistent 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 (sufficiently) certain.
The original Rough Set Approach proposed by Pawlak, called Indiscernibility-based Rough Set Approach (IRSA):
- concerns non-ordinal classification,
- employs indiscernibility relation,
- involves approximations of decision classes $X_i$.

Dominance-based Rough Set Approach (DRSA) proposed by Greco, Matarazzo and Słowinski:
- concerns ordinal classification with monotonicity constraints,
- employs dominance relation,
- involves approximations of unions of ordered decision classes:
  - upward unions $X_i^\geq = \bigcup_{t \geq i} X_t$, where $i = 2, 3, \ldots, n$, and
  - downward unions $X_i^\leq = \bigcup_{t \leq i} X_t$, where $i = 1, 2, \ldots, n - 1$. 
Basic notions and definitions

In case of non-ordinal classification handled by IRSA, set of attributes $A$ is composed of regular attributes only. Indiscernibility relation makes a partition of $U$ into disjoint blocks of objects called granules. Moreover, $I(y)$ denotes a set of objects indiscernible with object $y \in U$.

In case of ordinal classification with monotonicity constraints handled by DRSA, among condition attributes from $C$ there is at least one criterion, decision attribute $d$ has preference-ordered value set, and there exists a monotonic relationship between evaluation of objects on criteria and their values on the decision attribute.

The positive dominance cone $D^+(y)$ is composed of objects that for each $q_i \in C$ are not worse than object $y$.

The negative dominance cone $D^-(y)$ is composed of objects that for each $q_i \in C$ are not better than object $y$. 
In the following, we are going to consider DRSA only.

In order to simplify notation, we use:
- symbol $X$ to denote a set of objects belonging to union of classes $X_i^{\geq}$ or $X_i^{\leq}$,
- symbol $E(y)$ to denote any set $D^+(y)$ or $D^-(y)$, $y \in U$.

Moreover, if $X$ and $E(y)$ are used in the same equation, then for $X$ representing union of ordered classes $X_i^{\geq}$ (resp. $X_i^{\leq}$), $E(y)$ stands for dominance cone $D^+(y)$ (resp. $D^-(y)$).
The lower approximation of set $X$ is defined as:

$$X = \{ y \in X : E(y) \subseteq X \}. \quad (1)$$

This definition of the lower approximation appears to be too restrictive in practical applications. Therefore, various probabilistic rough set approaches were proposed which extend the lower approximation of set $X$ by inclusion of objects with sufficient evidence for membership to $X$. This evidence is quantified by different object consistency measures.
In the following, we consider Variable-Consistency DRSA (VC-DRSA), where the definition of probabilistic lower approximation of set $X$ involves object consistency measure $\Theta_X : U \rightarrow \mathbb{R}^+ \cup \{0\}$:

- given a **gain-type** measure $\Theta_X$ and a **gain**-threshold $\theta_X$:
  \[ X = \{ y \in X : \Theta_X(y) \geq \theta_X \}. \]  
  \[ (2) \]

- given a **cost-type** measure $\Theta_X$ and a **cost**-threshold $\theta_X$:
  \[ X = \{ y \in X : \Theta_X(y) \leq \theta_X \}. \]  
  \[ (3) \]
Basic notions and definitions

**Required monotonicity properties of object consistency measures**

- (m1): Monotonicity w.r.t. set of attributes $P \subseteq C$.
- (m2): Monotonicity w.r.t. set of objects $X \subseteq U$, when set $X$ is augmented by new objects.
- (m3): Monotonicity w.r.t. union of classes $X^\geq_i \subseteq U$ and $X^\leq_i \subseteq U$.
- (m4): Monotonicity w.r.t. dominance relation $D$. 

---

Let $X, \neg X \subseteq U$, where $\neg X = U - X$, $y \in U$. We consider the following object consistency measures:

- cost-type measure $\epsilon_X$ that has property (m1), (m2), and (m4):
  \[
  \epsilon_X(y) = \frac{|E(y) \cap \neg X|}{|\neg X|},
  \]
  (4)

- cost-type measure $\epsilon_X^*$ that has properties (m1)–(m4):
  \[
  \epsilon_{X_i}^{\geq}(y) = \max_{j \leq i} \epsilon_{X_j}^{\geq}(y),
  \]
  (5)

  \[
  \epsilon_{X_i}^{\leq}(y) = \max_{j \geq i} \epsilon_{X_j}^{\leq}(y),
  \]
  (6)

- cost-type measure $\epsilon'_X$ that has properties (m1)–(m4):
  \[
  \epsilon'_X(y) = \frac{|E(y) \cap \neg X|}{|X|}.
  \]
  (7)
\[ \epsilon_{X_3^\geq} (y_5) = \frac{2}{4} \]
Basic notions and definitions

We define the **positive region** of $X$ as:

$$POS(X) = \bigcup_{y \in X} E(y).$$

(8)

One can observe that $POS(X) \supseteq X$.

Basing on the definition of the positive region of set $X$, we also define **negative** and **boundary regions** of an approximated set as follows:

$$NEG(X) = POS(\neg X) - POS(X),$$

(9)

$$BND(X) = (U - POS(X)) - NEG(X).$$

(10)
Decision Rules
Probabilistic lower approximations are basis for induction of (probabilistic) decision rules which are a simple and comprehensive representation of knowledge:

if elementary conditions then decision (prediction).

Condition part of a rule is a conjunction of elementary conditions concerning individual attributes/criteria.

Decision part of a rule suggests an assignment to a union of decision classes.

Rules are characterized by rule consistency measures.

Rules explain decisions observed in data and can be used to classify objects to the predefined decision classes.
Set $X$ is the basis for induction of a set $R_X$ of decision rules that suggest assignment to $X$.

Each rule from $R_X$ is supported by at least one object from $X$, and it covers object(s) from $POS(X)$.

The elementary conditions (selectors) that form a rule $r_X \in R_X$ are built using evaluations of objects belonging to $X$ only.
Decision Rules

Induced rules have the following syntax:

\[\text{if } q_{i_1}(y) \geq t_{i_1} \land \ldots \land q_{i_p}(y) \geq t_{i_p} \land q_{i_p+1}(y) = t_{i_p+1} \land \ldots \land q_{i_z}(y) = t_{i_z} \]
\[\text{then } y \in X_{i}^{\geq},\]

\[\text{if } q_{i_1}(y) \leq t_{i_1} \land \ldots \land q_{i_p}(y) \leq t_{i_p} \land q_{i_p+1}(y) = t_{i_p+1} \land \ldots \land q_{i_z}(y) = t_{i_z} \]
\[\text{then } y \in X_{i}^{\leq},\]

where \(q_{i_1}, \ldots, q_{i_p}\) denote criteria, and \(q_{i_p+1}, \ldots, q_{i_z}\) denote regular attributes; moreover, \(t_{i_j}\) denotes a value taken from the value set of attribute \(q_{i_j}, i_j \in \{i_1, \ldots, i_z\} \subseteq \{1, \ldots, |C|\}\). Symbols \(\geq\) and \(\leq\) indicate weak preference and inverse weak preference w.r.t. single criterion, respectively. If \(q_{i_j} \in C\) is a gain (cost) criterion, then elementary condition \(q_{i_j}(y) \geq t_{i_j}\) means that the evaluation of object \(y \in U\) on criterion \(q_{i_j}\) is not worse than \(t_{i_j}\), \(i_j \in \{i_1, \ldots, i_p\}\). Elementary conditions for regular attributes are of the type \(q_{i_j}(y) = t_{i_j}, i_j \in \{i_{p+1}, \ldots, i_z\}\).
Decision Rules

Decision rules can be characterized by many attractiveness measures, like support, confidence, etc.

A decision rule that suggests assignment to set $X$ is denoted by $r_X$. Condition part of rule $r_X$ is denoted by $\Phi(r_X)$, while its decision part is denoted by $\Psi(r_X)$. Moreover, $|\Phi(r_X)|$ denotes the set of objects satisfying condition part of the rule.

Rule consistency measure is any function $\hat{\Theta}_X : R_X \rightarrow \mathbb{R}^+ \cup \{0\}$, where $R_X$ is a set of rules suggesting assignment to $X$.

We consider the following two rule consistency measures:

- $\epsilon$-consistency of $r_X$:
  \[
  \hat{\epsilon}_X(r_X) = \frac{|\|\Phi(r_X)\| \cap \neg X|}{|\neg X|},
  \]  
  \[
  (13)
  \]

- $\epsilon'$-consistency of $r_X$:
  \[
  \hat{\epsilon'}_X(r_X) = \frac{|\|\Phi(r_X)\| \cap \neg X|}{|X|}.
  \]  
  \[
  (14)\]
Induced rules must satisfy similar constraints on consistency as objects from the lower approximation which serves as a base for rule induction. In particular, each rule is required to satisfy threshold $\hat{\theta}_X$ defined w.r.t. a given rule consistency measure $\hat{\Theta}_X$. Such a threshold can be calculated as follows: $\hat{\theta}_X = \frac{|\neg X|}{|\neg X|} \theta_X$ in case of definition (13) and $\hat{\theta}_X = \frac{|X|}{|X|} \theta_X$ in case of definition (14).

$\epsilon$-consistency measure is related to object consistency measure $\epsilon_X$. $\epsilon'$-consistency measure is related to object consistency measure $\epsilon'_X$.

Both rule consistency measures derive monotonicity properties from the corresponding object consistency measures.

$\epsilon$-consistency measure can be used to induce decision rules from positive regions computed using object consistency measure $\epsilon^*_X$. 
Key concepts concerning rules induced from probabilistic lower approximations:

- A decision rule suggesting assignment to set $X$ is **discriminant** if it covers only objects belonging to positive region $POS(X)$.

- Rule is **free of redundant conditions** if removing any of its elementary conditions causes that it is no more discriminant.

- Rule is **minimal** if there is no other rule with not less general conditions and not less specific decision, i.e., $r_X$ is minimal if there does not exist other rule $r_Y$, $Y \subseteq U$, such that $\|\Phi(r_Y)\| \supseteq \|\Phi(r_X)\|$ and $Y \subseteq X$.

- Set of rules suggesting assignment to $X$ is **complete** iff each object $y \in X$ is covered by at least one rule $r_X \in R_X$.

- Rule $r_X \in R_X$ is **non-redundant** in $R_X$, if removing $r_X$ causes that $R_X$ ceases to be complete.
VCDomLEM Algorithm
VC-DomLEM Algorithm

VC-DomLEM is a sequential covering rule induction algorithm that induces a minimal set of rules satisfying constraints on consistency.

VC-DomLEM algorithm is composed of two parts:
- Algorithm 1 – the main routine,
- Algorithm 2 – $VC$-$SequentialCovering^{mix}$ subroutine.
Algorithm 1: VC-DomLEM

**Input**: set $X$ of upward unions of classes $X_i^\geq \in U$, or downward unions of classes $X_i^\leq \in U$, rule consistency measure $\hat{\Theta}_X$, set $\{\hat{\theta}_X : X \in X\}$ of rule consistency measure thresholds, object covering option $s$.

**Output**: set of rules $R$.

1. $R := \emptyset$
2. foreach $X \in X$ do
   3. $AO(X) := AllowedObjects(X, s)$;
   4. $R_X := VC-SequentialCovering^{mix}(X, AO(X), \hat{\Theta}_X, \hat{\theta}_X)$;
   5. $R := R \cup R_X$;
   6. RemoveNonMinimalRules($R$);
3. end
Each rule $r_X$ belonging to set $R_X$ is allowed to cover only objects from set $AO(X)$, calculated according to chosen option $s \in \{1, 2, 3\}$ (line 3).

We consider three reasonable options:

- (1): $AO(X) = POS(X)$,
- (2): $AO(X) = POS(X) \cup BND(X)$,
- (3): $AO(X) = U$.

Minimality check performed in line 6 can be simplified if in line 2 upward or downward unions are considered from the most specific (i.e., containing the smallest number of objects) to the most general (i.e., containing the largest number of objects). In such a case, only rules from set $R_X$ can be non-minimal.
Algorithm 2: VC-SequentialCovering$^{mix}$

**Input**: set $X \subseteq U$ of positive objects, 
set $AO(X) \supseteq X$ of objects that can be covered, 
rule consistency measure $\hat{\Theta}_X$, 
rule consistency measure threshold $\hat{\theta}_X$.

**Output**: set $R_X$ of rules suggesting assignment to $X$.

1. $B := X$;
2. $R_X := \emptyset$;
3. while $B \neq \emptyset$ do
4.   $r_X := \emptyset$;
5.   $EC := \text{ElementaryConditions}(B)$;
6.   while ($\hat{\Theta}_X(r_X)$ does not satisfy $\hat{\theta}_X$) or ($\|\Phi(r_X)\| \not\subseteq AO(X)$) do
7.     $ec := \text{BestElementaryCondition}(EC, r_X, \hat{\Theta}_X, X)$;
8.     $r_X := r_X \cup ec$;
9.     $EC := \text{ElementaryConditions}(B \cap \|\Phi(r_X)\|)$;
10. end
11. RemoveRedundantElementaryConditions($r_X, \hat{\Theta}_X, \hat{\theta}_X, AO(X)$);
12. $R_X := R_X \cup r_X$;
13. $B := B \setminus \|\Phi(r_X)\|$;
14. end
15. RemoveRedundantRules($R_X, \hat{\Theta}_X, X$);
The best elementary condition $ec$ is chosen according to the following criteria considered lexicographically:

1. the best value of chosen rule consistency measure $\hat{\Theta}_X$ (e.g., $\epsilon$-consistency or $\epsilon'$-consistency) for rule $r_X \cup ec$,
2. the best value of $|\left|\Phi(r_X \cup ec)\right| \cap X|$, where $r_X \cup ec$ denotes a rule resulting from extension of rule $r_X$ by new elementary condition $ec$.

If set $R_X$ contains redundant rules, an iterative procedure eliminating redundancy is adopted (line 15). In each step of this procedure, the rule to be removed is chosen according to the following measures considered lexicographically:

1. the worst (i.e., the smallest) value of $|\left|\Phi(r_X)\right| \cap X|$, 
2. the worst value of $\hat{\Theta}_X(r_X)$,
3. the smallest index of $r_X$ on the constructed list of rules.
Computational Experiment
The aim of the experiment was to evaluate the usefulness of VC-DomLEM algorithm in terms of its predictive accuracy (i.e., PCC and MAE).

VC-DomLEM algorithm, as implemented in the java Rough Set (jRS) library, was compared to other methods on 12 ordinal data sets. VC-DomLEM was used to induce rules satisfying $\epsilon$-consistency or $\epsilon'$-consistency condition from lower approximations calculated with object consistency measure $\epsilon_X$ or $\epsilon'_X$, respectively.

In order to classify objects using induced rules, VC-DRSA classification scheme\(^a\) was used.

The other methods compared to VC-DomLEM were: two ordinal classifiers that preserve monotonicity constraints, namely: Ordinal Learning Model (OLM) and Ordinal Stochastic Dominance Learner (OSDL), and four non-ordinal classifiers: Naive Bayes, Support Vector Machine (SVM) with linear kernel, decision rule classifier RIPPER, and decision tree classifier C4.5.
### Table: Characteristics of data sets

<table>
<thead>
<tr>
<th>Id</th>
<th>Data set</th>
<th>Objects</th>
<th>Attributes</th>
<th>Classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>breast-c</td>
<td>286</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>breast-w</td>
<td>699</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>car</td>
<td>1296</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>cpu</td>
<td>209</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>bank-g</td>
<td>1411</td>
<td>16</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>fame</td>
<td>1328</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>denbosch</td>
<td>119</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>ERA</td>
<td>1000</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>9</td>
<td>ESL</td>
<td>488</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td>LEV</td>
<td>1000</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>11</td>
<td>SWD</td>
<td>1000</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>12</td>
<td>windsor</td>
<td>546</td>
<td>10</td>
<td>4</td>
</tr>
</tbody>
</table>
The predictive accuracy (PCC and MAE) was estimated by stratified 10-fold cross-validation, repeated several times.

The tables with results contain the value of measure and its standard deviation for each data set and each classifier.

For each data set we calculated a rank (given in brackets) of the result of a classifier when compared to the other classifiers.

Last row of each table shows the average rank obtained by a given classifier.

Moreover, for each data set, the best value of the predictive accuracy measure, and those values which are within standard deviation of the best value, are marked in bold.
### Computational Experiment

**Table:** Mean absolute error (MAE)

<table>
<thead>
<tr>
<th>Id</th>
<th>VC-DomLEM</th>
<th>Naive Bayes</th>
<th>SVM</th>
<th>RIPPER</th>
<th>C4.5</th>
<th>OLM</th>
<th>OSDL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><strong>0.2331</strong> (1)</td>
<td>0.2564 (3)</td>
<td>0.3217 (6)</td>
<td>0.2960 (4)</td>
<td>0.2424 (2)</td>
<td>0.324 (7)</td>
<td>0.3065 (5)</td>
</tr>
<tr>
<td></td>
<td>+0.003297</td>
<td>+0.005943</td>
<td>+0.01244</td>
<td>+0.01154</td>
<td>+0.003297</td>
<td>+0.01835</td>
<td>+0.001648</td>
</tr>
<tr>
<td>2</td>
<td>0.03720 (2)</td>
<td>0.03958 (3)</td>
<td><strong>0.03243</strong> (1)</td>
<td>0.04483 (5)</td>
<td>0.05532 (6)</td>
<td>0.1764 (7)</td>
<td>0.04149 (4)</td>
</tr>
<tr>
<td></td>
<td>+0.002023</td>
<td>+0.0006744</td>
<td>+0.006744</td>
<td>+0.004721</td>
<td>+0.00751</td>
<td>+0.00552</td>
<td>+0.001168</td>
</tr>
<tr>
<td>3</td>
<td><strong>0.03421</strong> (1)</td>
<td>0.1757 (6)</td>
<td>0.08668 (3)</td>
<td>0.2029 (7)</td>
<td>0.1168 (5)</td>
<td>0.09156 (4)</td>
<td>0.04141 (2)</td>
</tr>
<tr>
<td></td>
<td>+0.0007275</td>
<td>+0.0002025</td>
<td>+0.0002025</td>
<td>+0.001302</td>
<td>+0.003108</td>
<td>+0.005358</td>
<td>+0.0009624</td>
</tr>
<tr>
<td>4</td>
<td><strong>0.08293</strong> (1)</td>
<td>0.1707 (4)</td>
<td>0.4386 (7)</td>
<td>0.1611 (3)</td>
<td>0.1196 (2)</td>
<td>0.3461 (6)</td>
<td>0.3158 (5)</td>
</tr>
<tr>
<td></td>
<td>+0.01479</td>
<td>+0.009832</td>
<td>+0.01579</td>
<td>+0.01372</td>
<td>+0.01790</td>
<td>+0.02743</td>
<td>+0.01034</td>
</tr>
<tr>
<td>5</td>
<td><strong>0.04559</strong> (1)</td>
<td>0.1146 (5)</td>
<td>0.1280 (6)</td>
<td>0.0489 (2)</td>
<td>0.0515 (3)</td>
<td>0.05528 (4)</td>
<td>0.1545 (7)</td>
</tr>
<tr>
<td></td>
<td>+0.001456</td>
<td>+0.01371</td>
<td>+0.001205</td>
<td>+0.00352</td>
<td>+0.005251</td>
<td>+0.001736</td>
<td>+0</td>
</tr>
<tr>
<td>6</td>
<td><strong>0.3406</strong> (1.5)</td>
<td>0.4829 (5)</td>
<td><strong>0.3406</strong> (1.5)</td>
<td>0.3991 (4)</td>
<td>0.3863 (3)</td>
<td>1.577 (6)</td>
<td>1.592 (7)</td>
</tr>
<tr>
<td></td>
<td>+0.001878</td>
<td>+0.002906</td>
<td>+0.001775</td>
<td>+0.003195</td>
<td>+0.005253</td>
<td>+0.03791</td>
<td>+0.007555</td>
</tr>
<tr>
<td>7</td>
<td><strong>0.1232</strong> (1)</td>
<td>0.1289 (2)</td>
<td>0.2129 (6)</td>
<td>0.1737 (5)</td>
<td>0.1653 (4)</td>
<td>0.2633 (7)</td>
<td>0.1541 (3)</td>
</tr>
<tr>
<td></td>
<td>+0.01048</td>
<td>+0.01428</td>
<td>+0.003961</td>
<td>+0.02598</td>
<td>+0.01048</td>
<td>+0.02206</td>
<td>+0.003961</td>
</tr>
<tr>
<td>8</td>
<td>1.307 (2)</td>
<td>1.325 (5)</td>
<td>1.318 (3)</td>
<td>1.681 (7)</td>
<td>1.326 (6)</td>
<td>1.321 (4)</td>
<td><strong>1.280</strong> (1)</td>
</tr>
<tr>
<td></td>
<td>+0.002055</td>
<td>+0.003771</td>
<td>+0.007257</td>
<td>+0.01558</td>
<td>+0.006018</td>
<td>+0.01027</td>
<td>+0.00704</td>
</tr>
<tr>
<td>9</td>
<td>0.3702 (3)</td>
<td><strong>0.3456</strong> (2)</td>
<td>0.4262 (5)</td>
<td>0.4296 (6)</td>
<td>0.3736 (4)</td>
<td>0.474 (7)</td>
<td><strong>0.3422</strong> (1)</td>
</tr>
<tr>
<td></td>
<td>+0.01352</td>
<td>+0.003864</td>
<td>+0.01004</td>
<td>+0.01608</td>
<td>+0.01089</td>
<td>+0.01114</td>
<td>+0.005019</td>
</tr>
<tr>
<td>10</td>
<td>0.4813 (6)</td>
<td>0.475 (5)</td>
<td>0.4457 (4)</td>
<td>0.4277 (3)</td>
<td>0.426 (2)</td>
<td>0.615 (7)</td>
<td><strong>0.4033</strong> (1)</td>
</tr>
<tr>
<td></td>
<td>+0.004028</td>
<td>+0.004320</td>
<td>+0.003399</td>
<td>+0.00838</td>
<td>+0.01476</td>
<td>+0.00999</td>
<td>+0.003091</td>
</tr>
<tr>
<td>11</td>
<td>0.454 (4)</td>
<td>0.475 (6)</td>
<td>0.4503 (2)</td>
<td>0.452 (3)</td>
<td>0.4603 (5)</td>
<td>0.5707 (7)</td>
<td><strong>0.433</strong> (1)</td>
</tr>
<tr>
<td></td>
<td>+0.004320</td>
<td>+0.004320</td>
<td>+0.002867</td>
<td>+0.006481</td>
<td>+0.004497</td>
<td>+0.007717</td>
<td>+0.002160</td>
</tr>
<tr>
<td>12</td>
<td><strong>0.5024</strong> (1)</td>
<td>0.5488 (3)</td>
<td>0.5891 (5)</td>
<td>0.6825 (7)</td>
<td>0.652 (6)</td>
<td>0.5757 (4)</td>
<td>0.5153 (2)</td>
</tr>
<tr>
<td></td>
<td>+0.006226</td>
<td>+0.005662</td>
<td>+0.02101</td>
<td>+0.03332</td>
<td>+0.03721</td>
<td>+0.006044</td>
<td>+0.006044</td>
</tr>
<tr>
<td></td>
<td><strong>2.04</strong></td>
<td><strong>4.08</strong></td>
<td><strong>4.13</strong></td>
<td><strong>4.67</strong></td>
<td><strong>4.00</strong></td>
<td><strong>5.83</strong></td>
<td><strong>3.25</strong></td>
</tr>
</tbody>
</table>

Note: Bold values indicate the bestperforming model for each dataset.
Table: Percentage of correctly classified objects (PCC)

<table>
<thead>
<tr>
<th>Id</th>
<th>VC-DomLEM</th>
<th>Naive Bayes</th>
<th>SVM</th>
<th>RIPPER</th>
<th>C4.5</th>
<th>OLM</th>
<th>OSDL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>76.69 (1)</td>
<td>74.36 (3)</td>
<td>67.83 (6)</td>
<td>70.4 (4)</td>
<td>75.76 (2)</td>
<td>67.6 (7)</td>
<td>69.35 (5)</td>
</tr>
<tr>
<td></td>
<td>+0.3297</td>
<td>+0.5943</td>
<td>+1.244</td>
<td>+1.154</td>
<td>+0.3297</td>
<td>+1.835</td>
<td>+0.1648</td>
</tr>
<tr>
<td>2</td>
<td>96.28 (2)</td>
<td>96.04 (3)</td>
<td><strong>96.76</strong> (1)</td>
<td>95.52 (5)</td>
<td>94.47 (6)</td>
<td>82.36 (7)</td>
<td>95.85 (4)</td>
</tr>
<tr>
<td></td>
<td>+0.2023</td>
<td>+0.06744</td>
<td>+0.06744</td>
<td>+0.4721</td>
<td>+0.751</td>
<td>+0.552</td>
<td>+0.1168</td>
</tr>
<tr>
<td>3</td>
<td><strong>97.15</strong> (1)</td>
<td>84.72 (6)</td>
<td>92.18 (3)</td>
<td>84.41 (7)</td>
<td>89.84 (5)</td>
<td>91.72 (4)</td>
<td>96.53 (2)</td>
</tr>
<tr>
<td></td>
<td>+0.063</td>
<td>+0.1667</td>
<td>+0.2025</td>
<td>+1.309</td>
<td>+0.1819</td>
<td>+0.4425</td>
<td>+0.063</td>
</tr>
<tr>
<td>4</td>
<td>91.7 (1)</td>
<td>83.41 (4)</td>
<td>56.62 (7)</td>
<td>84.69 (3)</td>
<td>88.52 (2)</td>
<td>68.58 (6)</td>
<td>72.41 (5)</td>
</tr>
<tr>
<td></td>
<td>+1.479</td>
<td>+0.9832</td>
<td>+1.579</td>
<td>+1.409</td>
<td>+1.409</td>
<td>+2.772</td>
<td>+1.479</td>
</tr>
<tr>
<td>5</td>
<td>95.44 (1)</td>
<td>88.54 (5)</td>
<td>87.2 (6)</td>
<td>95.11 (2)</td>
<td>94.85 (3)</td>
<td>94.47 (4)</td>
<td>84.55 (7)</td>
</tr>
<tr>
<td></td>
<td>+0.1456</td>
<td>+1.371</td>
<td>+0.1205</td>
<td>+0.352</td>
<td>+0.5251</td>
<td>+0.1736</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td><strong>67.55</strong> (1)</td>
<td>56.22 (5)</td>
<td><strong>67.1</strong> (2)</td>
<td>63.55 (4)</td>
<td>64.33 (3)</td>
<td>27.43 (6)</td>
<td>22.04 (7)</td>
</tr>
<tr>
<td></td>
<td>+0.4642</td>
<td>+0.2328</td>
<td>+0.2217</td>
<td>+0.5635</td>
<td>+0.5844</td>
<td>+0.7179</td>
<td>+0.128</td>
</tr>
<tr>
<td>7</td>
<td>87.68 (1)</td>
<td>87.11 (2)</td>
<td>78.71 (6)</td>
<td>82.63 (5)</td>
<td>83.47 (4)</td>
<td>73.67 (7)</td>
<td>84.6 (3)</td>
</tr>
<tr>
<td></td>
<td>+1.048</td>
<td>+1.428</td>
<td>+0.3961</td>
<td>+2.598</td>
<td>+1.048</td>
<td>+2.206</td>
<td>+0.3961</td>
</tr>
<tr>
<td>8</td>
<td>26.9 (2)</td>
<td>25.03 (3)</td>
<td>24.27 (5)</td>
<td>20 (7)</td>
<td><strong>27.83</strong> (1)</td>
<td>23.97 (6)</td>
<td>24.7 (4)</td>
</tr>
<tr>
<td></td>
<td>+0.3742</td>
<td>+0.2494</td>
<td>+0.2494</td>
<td>+0.4243</td>
<td>+0.4028</td>
<td>+0.4643</td>
<td>+0.8165</td>
</tr>
<tr>
<td>9</td>
<td>66.73 (3)</td>
<td>67.49 (2)</td>
<td>62.7 (5)</td>
<td>61.61 (6)</td>
<td>66.33 (4)</td>
<td>55.46 (7)</td>
<td><strong>68.3</strong> (1)</td>
</tr>
<tr>
<td></td>
<td>+1.256</td>
<td>+0.3483</td>
<td>+0.6693</td>
<td>+1.555</td>
<td>+0.6966</td>
<td>+0.7545</td>
<td>+0.3483</td>
</tr>
<tr>
<td>10</td>
<td>55.63 (6)</td>
<td>56.17 (5)</td>
<td>58.87 (4)</td>
<td>60.83 (2)</td>
<td>60.73 (3)</td>
<td>45.43 (7)</td>
<td><strong>63.03</strong> (1)</td>
</tr>
<tr>
<td></td>
<td>+0.3771</td>
<td>+0.3399</td>
<td>+0.3091</td>
<td>+0.6128</td>
<td>+1.271</td>
<td>+0.8179</td>
<td>+0.2625</td>
</tr>
<tr>
<td>11</td>
<td>56.43 (6)</td>
<td>56.57 (5)</td>
<td><strong>58.23</strong> (2)</td>
<td>57.63 (3)</td>
<td>57.1 (4)</td>
<td>47.83 (7)</td>
<td><strong>58.6</strong> (1)</td>
</tr>
<tr>
<td></td>
<td>+0.4643</td>
<td>+0.4784</td>
<td>+0.2055</td>
<td>+0.66</td>
<td>+0.4320</td>
<td>+0.411</td>
<td>+0.4243</td>
</tr>
<tr>
<td>12</td>
<td>54.58 (2)</td>
<td>53.6 (3)</td>
<td>51.83 (4)</td>
<td>44.08 (7)</td>
<td>47.99 (6)</td>
<td>49.15 (5)</td>
<td><strong>55.37</strong> (1)</td>
</tr>
<tr>
<td></td>
<td>+0.7913</td>
<td>+0.2284</td>
<td>+1.813</td>
<td>+0.8236</td>
<td>+2.888</td>
<td>+0.7527</td>
<td>+0.3763</td>
</tr>
</tbody>
</table>

2.25  3.83  4.25  4.58  3.58  6.08  3.42
It follows from the results of the experiment that VC-DomLEM is better than the other compared classifiers – it has the best value of the average rank of both predictive accuracy measures.

More detailed analysis of the results is presented in the literature (see references).
Conclusions
Conclusions

- DRSA is a flexible modeling method that allows to include domain knowledge and can handle possible inconsistencies in data by calculating lower approximations of sets.
- DRSA allows to work with heterogeneous attributes – nominal, ordinal and cardinal (no need of discretization).
- Rule model has many advantages, e.g., comprehensibility, lack of aggregation operators, predictive power, resistance to irrelevant attributes.
- VC-DomLEM performed best in the conducted computational experiment.
- Rule models that preserve monotonicity constraints are more transparent than the other models.