jMAF

Dominance-based Rough Set Approach Data Analysis Framework

User's guide

jMAF - Dominance-based Rough Set Data Analysis Framework

http://www.cs.put.poznan.pl/jblaszczynski/Site/jRS.html

Jerzy Błaszczyński, Salvatore Greco, Benedetto Matarazzo, Roman Słowiński, Marcin Szeląg
jurek.blaszczynski@cs.put.poznan.pl

March 18, 2009

Contents

1 Introduction

jMAF is a Rough Set Data Analysis Framework written in Java language. It is based on java Rough Set (jRS) library. jMAF and jRS library implement methods of analysis provided by the Dominance-based Rough Set Approach and Variable Consistency Dominance-based Rough Set Approach. In the following, we give basics of these two approaches and we provide an example of jMAF usage that is meant to instruct novice users.

2 Basic Concepts of Dominance-based Rough Set Approach

Dominance-based Rough Set Approach (DRSA) is defined for problems with background knowledge about ordinal evaluations of objects from a universe, and about monotonic relationships between these evaluations, e.g. “the larger the mass and the smaller the distance, the larger the gravity” or “the greater the debt of a firm, the greater its risk of failure”. Precisely, the monotonic relationships are assumed between evaluation of objects on condition attributes and their assignment to decision classes. The monotonic relationships are also interpreted as monotonicity constraints, because the better the evaluation of an object, the better should be the decision class the object is assigned to. For this reason, classification problems of this kind are called ordinal classification problems with monotonicity constraints. Many real-world classification problems fall into this category. Typical examples are multiple criteria sorting and decision under uncertainty, where the order of value sets of attributes corresponds to increasing or decreasing order of preference of a decision maker. In these decision problems, the condition attributes are called *criteria*.

Although DRSA is a general methodology for reasoning about data describing ordinal classification problems with monotonicity constraints, in this manual, we shall use the vocabulary typical for multiple criteria sorting problems.

2.1 Decision Table

Let us consider a decision table including a finite universe of objects (solutions, alternatives, actions) U evaluated on a finite set of condition attributes $F = \{f_1, \dots, f_n\}$, and on a single decision attribute d.

Table 1: Exemplary decision table with evaluations of students

Student	f_1 - Mathematics	f_2 - Physics	f_3 - Literature	d - Overall Evaluation
$S1$	good	medium	bad	bad
$S2$	medium	medium	bad	medium
$S3$	medium	medium	medium	medium
$S4$	good	good	medium	good
$S5$	good	medium	good	good
$S6$	good	good	good	good
$S7$	bad	bad	bad	bad
$S8$	bad	bad	medium	bad

The set of the indices of attributes is denoted by $I = \{1, \dots, n\}$. Without loss of generality, $f_i : U \rightarrow \Re$ for each $i = 1, \dots, n$, and, for all objects $x, y \in U$, $f_i(x) \geq f_i(y)$ means that "x is at least as good as y with respect to attribute i", which is denoted by $x \succeq_i y$. Therefore, it is supposed that $\succeq_i$ is a complete preorder, i.e. a strongly complete and transitive binary relation, defined on U on the basis of quantitative and qualitative evaluations $f_i(\cdot)$. Furthermore, decision attribute d makes a partition of U into a finite number of decision classes, $\boldsymbol{Cl}=\{Cl_1, \dots, Cl_m\}$, such that each $x \in U$ belongs to one and only one class Cl_t, $t = 1, \dots, m$. It is assumed that the classes are preference ordered, i.e. for all $r, s = 1, \dots, m$, such that $r > s$, the objects from Cl_r are preferred to the objects from Cl_s. More formally, if $\succeq$ is a *comprehensive weak preference relation* on U, i.e. if for all $x, y \in U$, $x \succeq y$ reads "x is at least as good as y", then it is supposed that

$$[x \in Cl_r,\ y \in Cl_s,\ r>s] \Rightarrow x \succ y,$$

where $x \succ y$ means $x \succeq y$ and *not* $y \succeq x$.

The above assumptions are typical for consideration of an ordinal classification (or multiple criteria sorting) problem. Indeed, the decision table characterized above, includes examples of ordinal classification which constitute an input *preference information* to be analyzed using DRSA.

The sets to be approximated are called *upward union* and *downward union* of decision classes, respectively:

$$Cl_t^{\geq} = \bigcup_{s \geq t} Cl_s, \quad Cl_t^{\leq} = \bigcup_{s \leq t} Cl_s, \quad t = 1, ..., m.$$

The statement $x \in Cl_t^{\geq}$ reads "x belongs to at least class Cl_t", while $x \in Cl_t^{\leq}$ reads "x belongs to at most class Cl_t". Let us remark that $Cl_1^{\geq} = Cl_m^{\leq} = U$, $Cl_m^{\geq} = Cl_m$ and $Cl_1^{\leq} = Cl_1$. Furthermore, for t=2,...,m,

$$Cl_{t-1}^{\leq} = U - Cl_t^{\geq} \quad \text{and} \quad Cl_t^{\geq} = U - Cl_{t-1}^{\leq}.$$

2.2 Dominance cones as granules of knowledge

The key idea of DRSA is representation (approximation) of upward and downward unions of decision classes, by *granules of knowledge* generated by attributes being criteria. These granules are *dominance cones* in the attribute values space.

x *dominates* y with respect to set of attributes $P \subseteq F$ (shortly, x P-*dominates* y), denoted by xD_Py, if for every attribute $f_i \in P$, $f_i(x) \geq f_i(y)$. The relation of P-dominance is reflexive and transitive, i.e., it is a partial preorder.

Given a set of attributes $P \subseteq I$ and $x \in U$, the granules of knowledge used for approximation in DRSA are:

- a set of objects dominating x, called P-*dominating set*, $D_P^+(x)=\{y \in U\colon\ yD_Px\}$,
- a set of objects dominated by x, called P-*dominated set*, $D_P^-(x)=\{y \in U\colon\ xD_Py\}$.

Let us recall that the dominance principle requires that an object x dominating object y on all considered attributes (i.e. x having evaluations at least as good as y on all considered attributes) should also dominate y on the decision (i.e. x should be assigned to at least as good decision class as y). Objects satisfying the dominance principle are called *consistent*, and those which are violating the dominance principle are called *inconsistent*.

2.3 Approximation of ordered decision classes

The P-*lower approximation* of $Cl_t^{\geq}$, denoted by $\underline{P}(Cl_t^{\geq})$, and the P-*upper approximation* of $Cl_t^{\geq}$, denoted by $\overline{P}(Cl_t^{\geq})$, are defined as follows ($t = 2, ..., m$):

$$\begin{aligned}\underline{P}(Cl_t^{\geq}) &= \{x \in U : D_P^+(x) \subseteq Cl_t^{\geq}\},\\ \overline{P}(Cl_t^{\geq}) &= \{x \in U : D_P^-(x) \cap Cl_t^{\geq} \neq \emptyset\}.\end{aligned}$$

Analogously, one can define the P-*lower approximation* and the P-*upper approximation* of $Cl_t^{\leq}$ as follows ($t = 1, ..., m-1$):

$$\begin{aligned}\underline{P}(Cl_t^{\leq}) &= \{x \in U : D_P^-(x) \subseteq Cl_t^{\leq}\},\\ \overline{P}(Cl_t^{\leq}) &= \{x \in U : D_P^+(x) \cap Cl_t^{\leq} \neq \emptyset\}.\end{aligned}$$

The P-lower and P-upper approximations so defined satisfy the following *inclusion property*, for all $P \subseteq F$:

$$\underline{P}(Cl_t^{\geq}) \subseteq Cl_t^{\geq} \subseteq \overline{P}(Cl_t^{\geq}), \quad t = 2, \ldots, m,$$

$$\underline{P}(Cl_t^{\leq}) \subseteq Cl_t^{\leq} \subseteq \overline{P}(Cl_t^{\leq}), \quad t = 1, \ldots, m-1.$$

The P-lower and P-upper approximations of $Cl_t^{\geq}$ and $Cl_t^{\leq}$ have an important *complementarity property*, according to which,

$$\underline{P}(Cl_t^{\geq}) = U - \overline{P}(Cl_{t-1}^{\leq}) \text{ and } \overline{P}(Cl_t^{\geq}) = U - \underline{P}(Cl_{t-1}^{\leq}), \ t=2,...,m,$$

$$\underline{P}(Cl_t^{\leq}) = U - \overline{P}(Cl_{t+1}^{\geq}) \text{ and } \overline{P}(Cl_t^{\leq}) = U - \underline{P}(Cl_{t+1}^{\geq}), \ t=1,...,m-1.$$

The P-*boundary* of $Cl_t^{\geq}$ and $Cl_t^{\leq}$, denoted by $Bn_P(Cl_t^{\geq})$ and $Bn_P(Cl_t^{\leq})$, respectively, are defined as follows:

$$Bn_P(Cl_t^{\geq}) = \overline{P}(Cl_t^{\geq}) - \underline{P}(Cl_t^{\geq}), \quad t = 2, \ldots, m,$$

$$Bn_P(Cl_t^{\leq}) = \overline{P}(Cl_t^{\leq}) - \underline{P}(Cl_t^{\leq}), \quad t = 1, \ldots, m-1.$$

Due to the above complementarity property, $Bn_P(Cl_t^{\geq}) = Bn_P(Cl_{t-1}^{\leq})$, for $t = 2, ..., m$.

2.4 Quality of approximation

For every $P \subseteq F$, the *quality of approximation* of the ordinal classification $\boldsymbol{Cl}$ by a set of attributes P is defined as the ratio of the number of objects P-consistent with the dominance principle and the number of all the objects in U. Since the P-consistent objects are those which do not belong to any P-boundary $Bn_P(Cl_t^{\geq})$, $t = 2, \ldots, m$, or $Bn_P(Cl_t^{\leq})$, $t = 1, \ldots, m-1$, the quality of approximation of the ordinal classification $\boldsymbol{Cl}$ by a set of attributes P, can be written as

$$\gamma_P(\boldsymbol{Cl}) = \frac{\left|U - \left(\bigcup_{t=2,\ldots,m} Bn_P(Cl_t^{\geq})\right)\right|}{|U|} = \frac{\left|U - \left(\bigcup_{t=1,\ldots,m-1} Bn_P(Cl_t^{\leq})\right)\right|}{|U|}.$$

$\gamma_P(\boldsymbol{Cl})$ can be seen as a degree of consistency of the objects from U, where P is the set of attributes being criteria and $\boldsymbol{Cl}$ is the considered ordinal classification.

Moreover, for every $P \subseteq F$, the *accuracy of approximation* of union of ordered classes $Cl_t^{\geq}$, $Cl_t^{\leq}$ by a set of attributes P is defined as the ratio of the number of objects belonging to P-lower approximation and P-upper approximation of the union. Accuracy of approximation $\alpha_P(Cl_t^{\geq})$, $\alpha_P(Cl_t^{\leq})$ can be written as

$$\alpha_P(Cl_t^{\geq}) = \frac{\left|\underline{P}(Cl_t^{\geq})\right|}{|\overline{P}(Cl_t^{\geq})|}, \qquad \alpha_P(Cl_t^{\leq}) = \frac{\left|\underline{P}(Cl_t^{\leq})\right|}{|\overline{P}(Cl_t^{\leq})|}.$$

2.5 Reduction of attributes

Each minimal (with respect to inclusion) subset $P \subseteq F$ such that $\gamma_P(\boldsymbol{Cl}) = \gamma_F(\boldsymbol{Cl})$ is called a *reduct* of $\boldsymbol{Cl}$, and is denoted by $RED_{\boldsymbol{Cl}}$. Let us remark that for a given set U one can have more than one reduct. The intersection of all reducts is called the *core*, and is denoted by $CORE_{\boldsymbol{Cl}}$. Attributes in $CORE_{\boldsymbol{Cl}}$ cannot be removed from consideration without deteriorating the quality of approximation. This means that, in set F, there are three categories of attributes:

- *indispensable* attributes included in the core,
- *exchangeable* attributes included in some reducts, but not in the core,
- *redundant* attributes, neither indispensable nor exchangeable, and thus not included in any reduct.

2.6 Decision Rules

The dominance-based rough approximations of upward and downward unions of decision classes can serve to induce a generalized description of objects in terms of "*if* ..., *then* ..." decision rules. For a given upward or downward union of classes, $Cl_t^{\geq}$ or $Cl_s^{\leq}$, the decision rules induced under a hypothesis that objects belonging to $\underline{P}(Cl_t^{\geq})$ or $\underline{P}(Cl_s^{\leq})$ are positive examples, and all the others are negative, suggest a *certain* assignment to "class Cl_t or better", or to "class Cl_s or worse", respectively. On the other hand, the decision rules induced under a hypothesis that objects belonging to $\overline{P}(Cl_t^{\geq})$ or $\overline{P}(Cl_s^{\leq})$ are positive examples, and all the others are negative, suggest a *possible* assignment to "class Cl_t or better", or to "class Cl_s or worse", respectively. Finally, the decision rules induced under a hypothesis that objects belonging to the intersection $\overline{P}(Cl_s^{\leq}) \cap \overline{P}(Cl_t^{\geq})$ are positive examples, and all the others are negative, suggest an *approximate* assignment to some classes between Cl_s and Cl_t $(s < t)$.

In the case of preference ordered description of objects, set U is composed of examples of ordinal classification. Then, it is meaningful to consider the following five types of decision rules:

1) *certain* D$_{\geq}$-*decision rules*, providing lower profile descriptions for objects belonging to $\underline{P}(Cl_t^{\geq})$:
if $f_{i_1}(x) \geq r_{i_1}$ *and* ... *and* $f_{i_p}(x) \geq r_{i_p}$, *then* $x \in Cl_t^{\geq}$,
$\{i_1, \ldots, i_p\} \subseteq I$, $t = 2, \ldots, m$, $r_{i_1}, \ldots, r_{i_p} \in \Re$;

2) *possible* D$_{\geq}$-*decision rules*, providing lower profile descriptions for objects belonging to $\overline{P}(Cl_t^{\geq})$:
if $f_{i_1}(x) \geq r_{i_1}$ *and* ... *and* $f_{i_p}(x) \geq r_{i_p}$, *then* x *possibly belongs to* $Cl_t^{\geq}$,
$\{i_1, \ldots, i_p\} \subseteq I$, $t = 2, \ldots, m$, $r_{i_1}, \ldots, r_{i_p} \in \Re$;

3) *certain* D$_{\leq}$-*decision rules*, providing upper profile descriptions for objects belonging to $\underline{P}(Cl_t^{\leq})$:
if $f_{i_1}(x) \leq r_{i_1}$ *and* ... *and* $f_{i_p}(x) \leq r_{i_p}$, *then* $x \in Cl_t^{\leq}$,
$\{i_1, \ldots, i_p\} \subseteq I$, $t = 1, \ldots, m-1$, $r_{i_1}, \ldots, r_{i_p} \in \Re$;

4) *possible* $D_{\leq}$*-decision rules*, providing upper profile descriptions for objects belonging to $\overline{P}(Cl_t^{\leq})$:
if $f_{i_1}(x) \leq r_{i_1}$ *and* ... *and* $f_{i_p}(x) \leq r_{i_p}$, *then x possibly belongs to* $Cl_t^{\leq}$,
$\{i_1, \ldots, i_p\} \subseteq I$, $t = 1, \ldots, m-1$, $r_{i_1}, \ldots, r_{i_p} \in \Re$;

5) *approximate* $D_{\geq\leq}$*-decision rules*, providing simultaneously lower and upper profile descriptions for objects belonging to $Cl_s \cup Cl_{s+1} \cup \ldots \cup Cl_t$, without possibility of discerning to which class:
if $f_{i_1}(x) \geq r_{i_1}$ *and* ... *and* $f_{i_k}(x) \geq r_{i_k}$ *and* $f_{i_{k+1}}(x) \leq r_{i_{k+1}}$ *and* ... *and* $f_{i_p}(x) \leq r_{i_p}$, *then* $x \in Cl_s \cup Cl_{s+1} \cup \ldots \cup Cl_t$,
$\{i_1, \ldots, i_p\} \subseteq I$, $s, t \in \{1, \ldots, m\}$, $s < t$, $r_{i_1}, \ldots, r_{i_p} \in \Re$.

In the premise of a $D_{\geq\leq}$-decision rule, there can be "$f_i(x) \geq r_i$" and "$f_i(x) \leq r_i'$", where $r_i \leq r_i'$, for the same $i \in I$. Moreover, if $r_i = r_i'$, the two conditions boil down to "$f_i(x) = r_i$".

Since a decision rule is a kind of implication, a *minimal* rule is understood as an implication such that there is no other implication with the premise of at least the same weakness (in other words, a rule using a subset of elementary conditions and/or weaker elementary conditions) and the conclusion of at least the same strength (in other words, a $D_{\geq}$- or a $D_{\leq}$-decision rule assigning objects to the same union or sub-union of classes, or a $D_{\geq\leq}$-decision rule assigning objects to the same or smaller set of classes).

The rules of type 1) and 3) represent certain knowledge extracted from data (examples of ordinal classification), while the rules of type 2) and 4) represent possible knowledge; the rules of type 5) represent doubtful knowledge, because they are supported by inconsistent objects only.

Given a certain or possible $D_{\geq}$-decision rule $r \equiv$ "*if* $f_{i_1}(x) \geq r_{i_1}$ *and* ... *and* $f_{i_p}(x) \geq r_{i_p}$, *then* $x \in Cl_t^{\geq}$", an object $y \in U$ *supports* r if $f_{i_1}(y) \geq r_{i_1}$ *and* ... *and* $f_{i_p}(y) \geq r_{i_p}$. Moreover, object $y \in U$ supporting decision rule r is a *base* of r if $f_{i_1}(y) = r_{i_1}$ *and* ... *and* $f_{i_p}(y) = r_{i_p}$. Similar definitions hold for certain or possible $D_{\leq}$-decision rules and approximate $D_{\geq\leq}$-decision rules. A decision rule having at least one base is called *robust*. Identification of supporting objects and bases of robust rules is important for interpretation of the rules in multiple criteria decision analysis. The ratio of the number of objects supporting a rule and the number of all considered objects is called *relative support* of a rule. The relative support and the confidence ratio are basic characteristics of a rule, however, some *Bayesian confirmation measures* reflect much better the attractiveness of a rule [21].

A set of decision rules is *complete* if it covers all considered objects (examples of ordinal classification) in such a way that consistent objects are re-assigned to their original classes, and inconsistent objects are assigned to clusters of classes referring to this inconsistency. A set of decision rules is *minimal* if it is complete and non-redundant i.e., exclusion of any rule from this set makes it incomplete.

Note that the syntax of decision rules induced from rough approximations defined using dominance cones, is using consistently this type of granules. Each condition profile defines a dominance cone in n-dimensional condition space $\Re^n$, and each decision profile defines a dominance cone in one-dimensional decision space $\{1, \ldots, m\}$. In both cases, the cones are positive for $D_{\geq}$-rules and negative for $D_{\leq}$-rules.

Let us also remark that dominance cones corresponding to condition profiles can originate in any point of $\Re^n$, without the risk of their being too specific. Thus, contrary to traditional granular computing, the condition space $\Re^n$ need not be discretized.

Procedures for rule induction from dominance-based rough approximations have been proposed in [19]. The following section 3 presents example of use of jMAF for sorting problem (called also multi-criteria classification or ordered classification with monotonicity constraints). The surveys [13, 14, 15, 31, 32] include applications of DRSA.

2.7 Variable Consistency Dominance-based Rough Set Approaches

In DRSA, lower approximation of a union of ordered classes contains only consistent objects. Such a lower approximation is defined as a sum of dominance cones that are subsets of the approximated union. In practical applications, however, such a strong requirement may result in relatively small lower approximations. Therefore, several extensions of DRSA have been proposed. These extensions relax the condition for inclusion of an object to the lower approximation. Variable Consistency Dominance-based Rough Set Approaches (VC-DRSA) include to lower approximations objects which are sufficiently consistent. Different measures of consistency may be applied in VC-DRSA. Given a user-defined threshold value, extended lower approximation of a union of classes is defined as a set of objects for which the consistency measure satisfies that threshold.

Several definitions of VC-DRSA have been considered in the literature so far. In the first papers concerning VC-DRSA [11, 20], consistency of objects have been calculated using rough membership measure [26, 36]. Then, in order to ensure monotonicity of lower approximation with respect to the dominance relation, the idea of the first papers have been extended in the work [2]. Recently, it has been pointed out that it is reasonable to require that consistency measure used in the definition of the lower approximation satisfies some properties of monotonicity [4]. Resulting variable consistency approaches, employing monotonic consistency measures, are called Monotonic Variable Consistency Dominance-based Rough Set Approaches [3, 4].

3 Example of Use

We present a didactic example which illustrates application of jMAF in data analysis.

3.1 Running jMAF

You may find jMAF executable file in the location where you have unpacked the zip file that can be downloaded from `http://www.cs.put.poznan.pl/jblaszczynski/Site/jRS.html`. Please launch this file. A moment later you will see main jMAF window on your desktop. It should resemble the one presented in Figure 1.

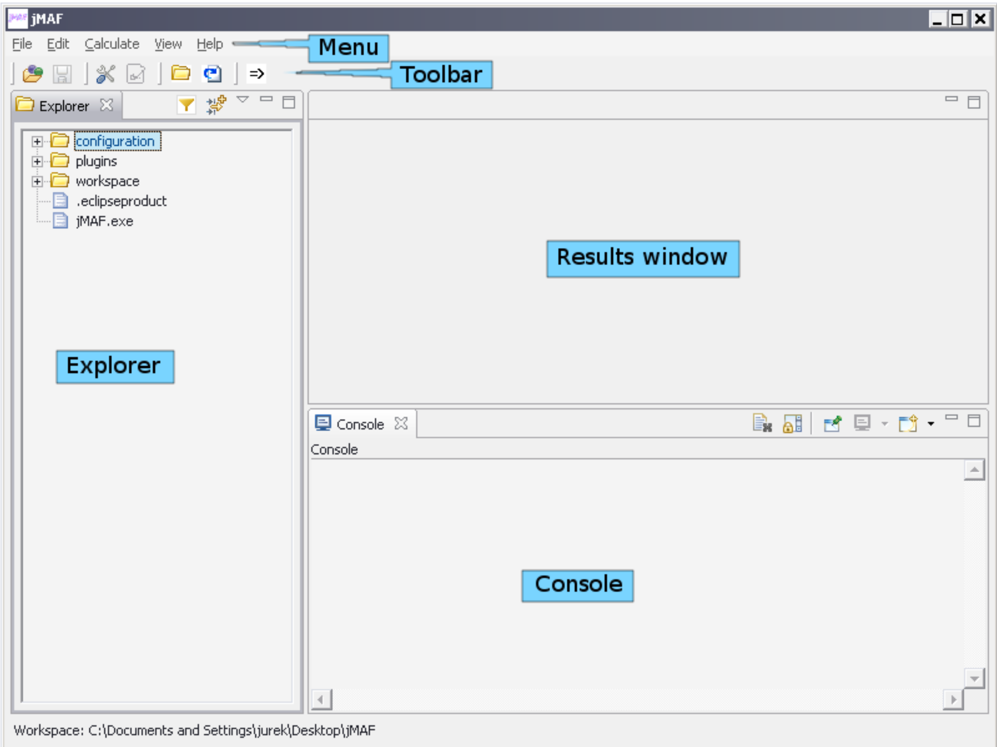

Figure 1: jMAF main window

Now you have jAMM running in workspace folder located in the folder where it was launched from. You can check the content of workspace folder by examining the explorer window. The main jMAF window is divided into 4 sub windows: topmost menubar and toolbar, middle explorer and results window and bottom console window. There is also a status line at the bottom.

3.2 Decision Table

Let us consider the following ordinal classification problem. Students of a college must obtain an overall evaluation on the basis of their achievements in Mathematics, Physics and Literature. These three subjects are clearly criteria (condition attributes) and the comprehensive evaluation is a decision attribute. For simplicity, the value sets of the attributes and of the decision attribute are the same, and they are composed of three values: bad, medium and good. The preference order of these values is obvious. Thus, there are three preference ordered decision classes, so the problem belongs to the category of ordinal classification. In order to build a preference model of the jury, DRSA is used to analyze a set of exemplary evaluations of students provided by the jury. These examples of ordinal classification constitute an input preference information presented as decision table in Table 2.

Note that the dominance principle obviously applies to the examples of ordinal classification, since an improvement of a student's score on one of three attributes, with other scores unchanged, should not worsen the student's overall evaluation, but rather improve it.

Table 2: Exemplary decision table with evaluations of students (examples of ordinal classification)

Student	Mathematics	Physics	Literature	Overall Evaluation
$S1$	good	medium	bad	bad
$S2$	medium	medium	bad	medium
$S3$	medium	medium	medium	medium
$S4$	good	good	medium	good
$S5$	good	medium	good	good
$S6$	good	good	good	good
$S7$	bad	bad	bad	bad
$S8$	bad	bad	medium	bad

Observe that student $S1$ has not worse evaluations than student $S2$ on all the considered condition attributes, however, the overall evaluation of $S1$ is worse than the overall evaluation of $S2$. This violates the dominance principle, so the two examples of ordinal classification, and only those, are inconsistent. One can expect that the quality of approximation of the ordinal classification represented by examples in Table 2 will be equal to 0.75.

3.3 Data File

As the first step you should create a file containing data from the data table. You have now two choices - you may use spreadsheet-like editor or any plain text editor. For this example, we will focus on the second option.

Run any text editor that is available on your system installation. Enter the text shown below.

```
**ATTRIBUTES
+ Mathematics : [bad, medium, good]
+ Physics : [bad, medium, good]
+ Literature : [bad, medium, good]
+ Overall : [bad, medium, good]
decision: Overall

**PREFERENCES
Mathematics : gain
Physics : gain
Literature : gain
Overall : gain

**EXAMPLES
good medium bad bad
medium medium bad medium
medium medium medium medium
good good medium good
good medium good good
good good good good
bad bad bad bad
bad bad medium bad

**END
```

Now, save the file as students.isf (for example in the jMAF folder). At this moment you are able to open this file in jMAF.

3.4 Opening isf File

Use **File** | **Open** to open isf file. You will see a typical file open dialog. Please select your newly created file. Alternatively, you can double click file in the explorer window if you have saved it in the workspace folder. If the file is not visible in explorer window, try right clicking on the explorer window and select from the context menu **Refresh** or **Switch workspace** to choose different workspace folder.

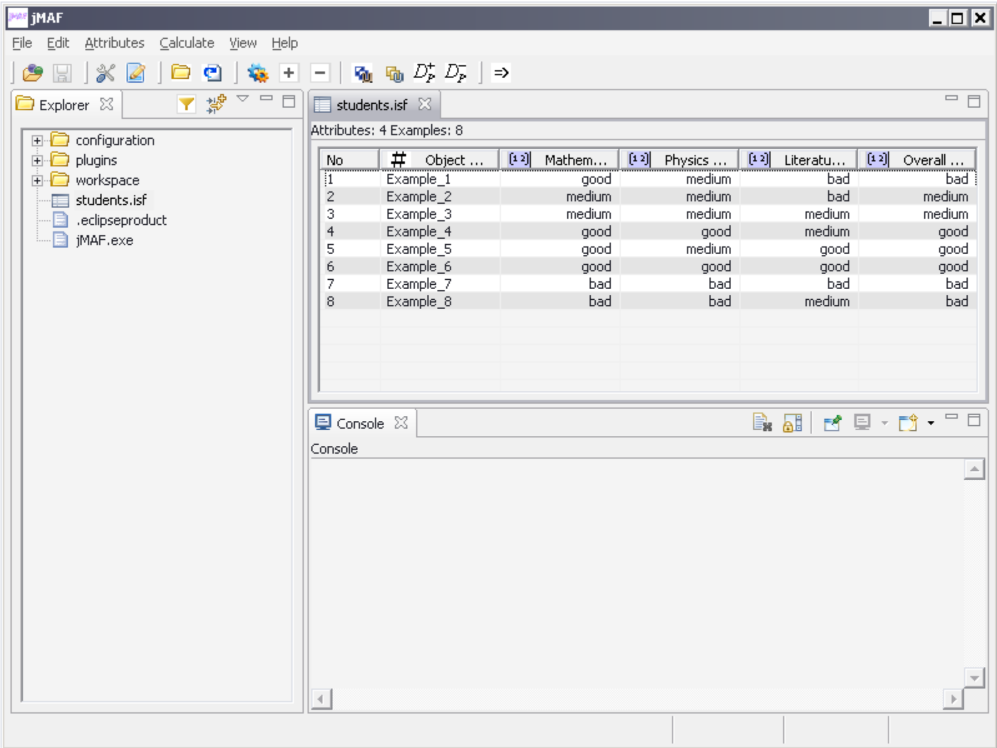

Figure 2: File students.isf opened in jMAF

3.5 Calculation of Dominance Cones

One of the first steps of data analysis using rough set theory is calculation of dominance cones (P-dominating sets and P-dominated sets). To perform this step, you can select an example from the isf file in results window and use **Calculate** | **P-Dominance Sets** | **Calculate dominating set** or **Calculate** | **P-Dominance Sets** | **Calculate dominated set**. You can also use these options from the toolbar menu. The resulting dominance cones for student $S1$ are visible in Figures 3 and 4.

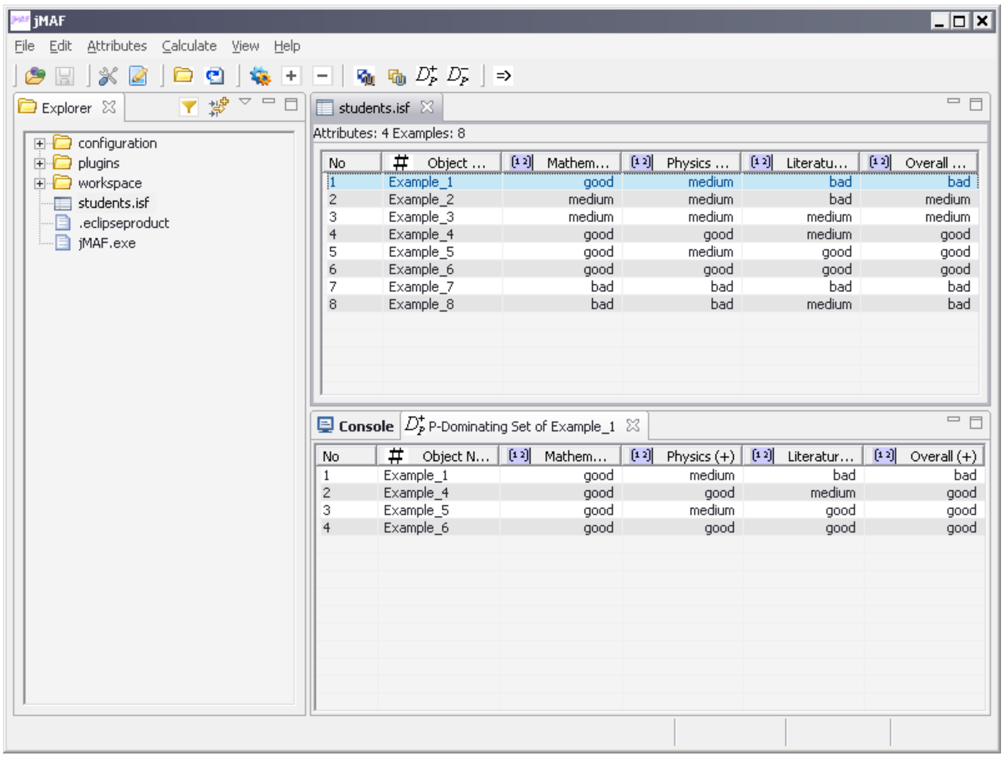

Figure 3: P-dominating cone of Example 1

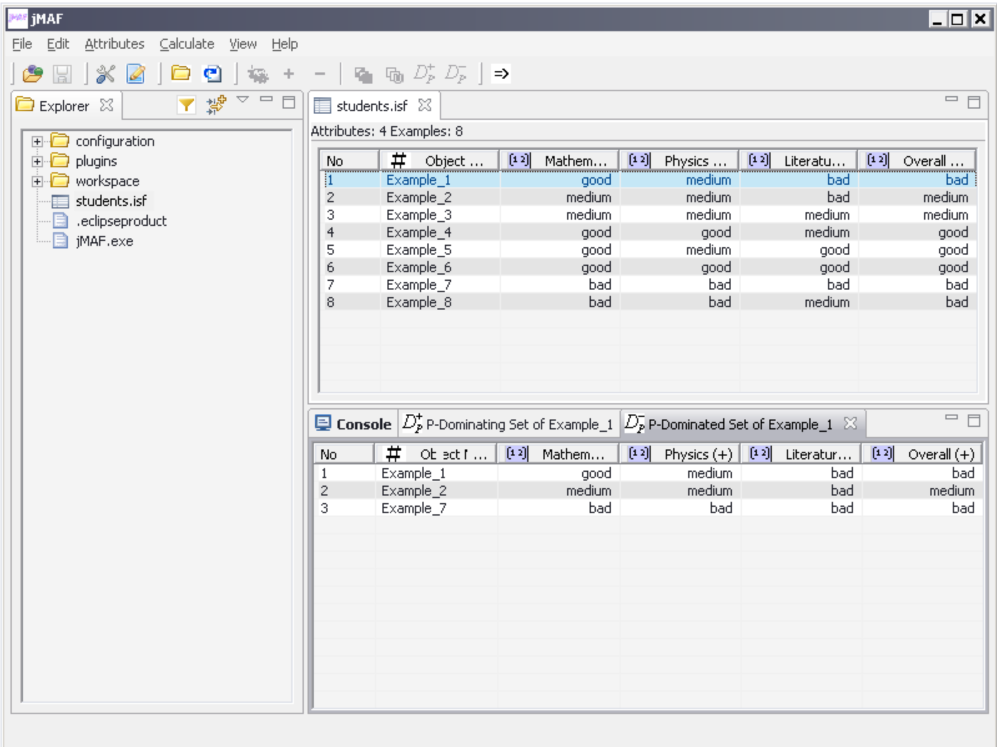

Figure 4: P-dominated cone of Example 1

3.6 Calculation of Approximations

The next step in rough set analysis is calculation of approximations. Use **Calculate** | **Unions of classes** | **Standard unions of classes** to calculate DRSA unions and their approximations. Now, you should see an input dialog for calculation of approximations. It should look like the one presented in Figure 5.

Figure 5: Input dialog for calculation of approximations

Leave default value of the consistency level parameter if you are looking for standard DRSA analysis. You can also set consistency level lower than one, to perform VC-DRSA analysis. This part is however not covered by this guide. You should see the result as presented in Figure 6.

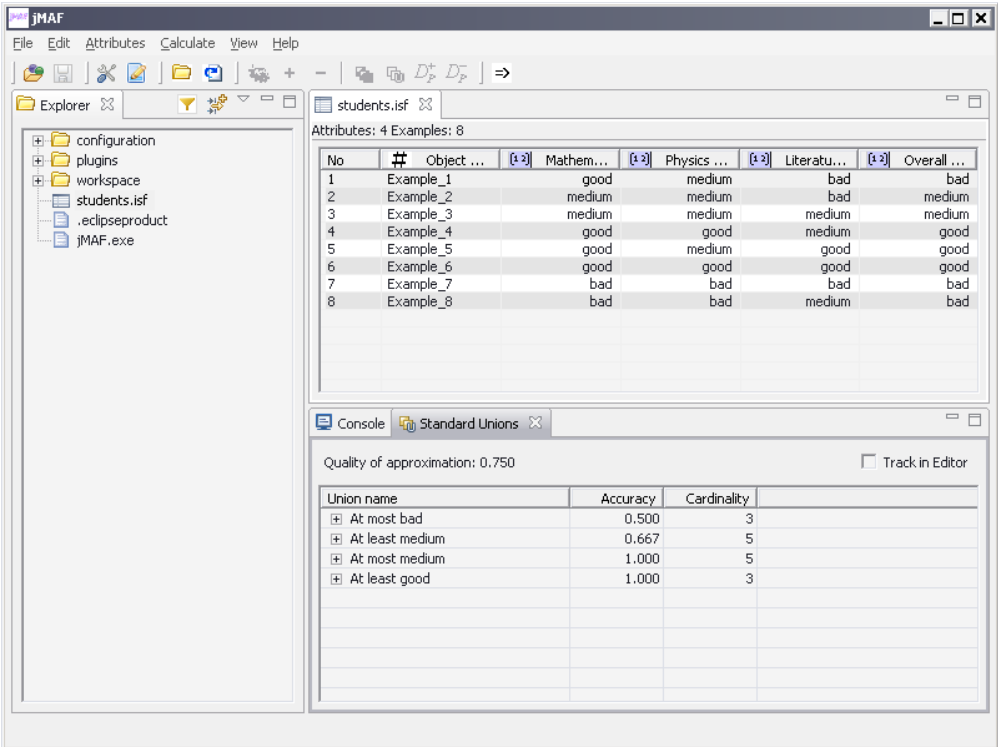

Figure 6: Approximations of unions of classes

You can navigate in Standard Unions window to see more details concerning calculated approximations (they are presented in Figure 7).

As you can see, quality of approximation equals 0.75, and accuracy of approximation in unions of classes ranges from 0.5 to 1.0. Lower approximation of union "at most" bad includes $S7$ and $S8$. Please select **Track in Editor** option to track your selection from Standard Unions window in the results window.

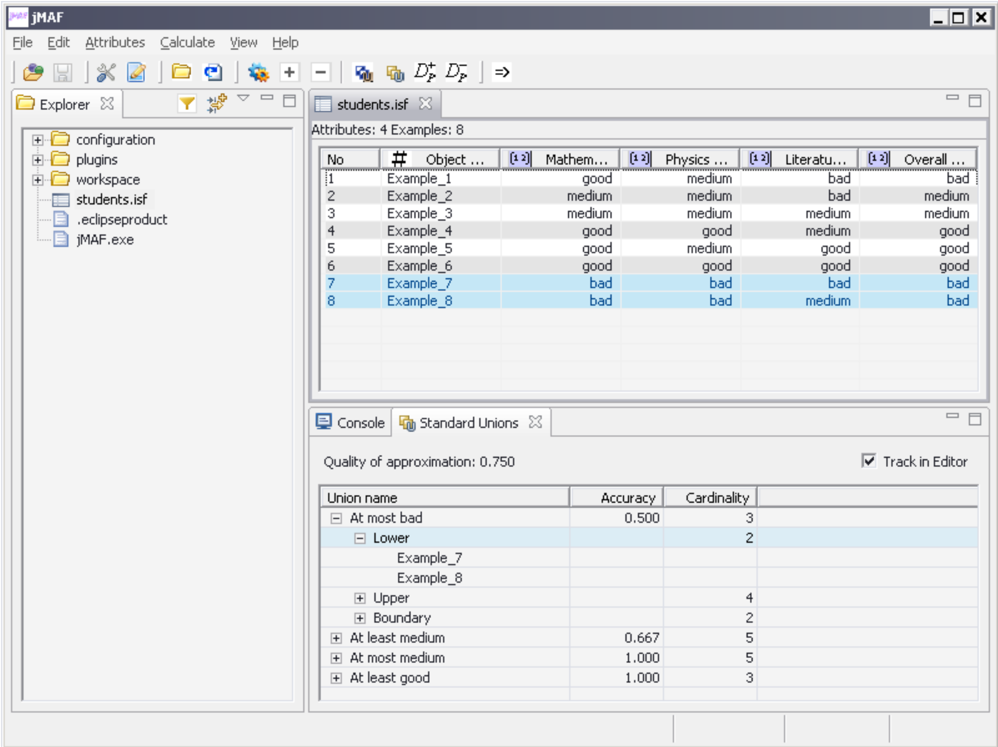

Figure 7: Details of approximations of unions of classes

3.7 Induction of Decision Rules

Given the above rough approximations, one can induce a set of decision rules representing the preferences of the jury. We will use one of the available methods - minimal covering rules (VC-DOMLEM algorithm).The idea is that evaluation profiles of students belonging to the lower approximations can serve as a base for some certain rules, while evaluation profiles of students belonging to the boundaries can serve as a base for some approximate rules. In the example we will consider, however, only certain rules.

To induce rules use **Calculate | VC-DOMLEM algorithm**. You will see a dialog with parameters of rule induction that is presented in Figure 8. Leave default values of these parameters to perform standard rule induction for DRSA analysis.

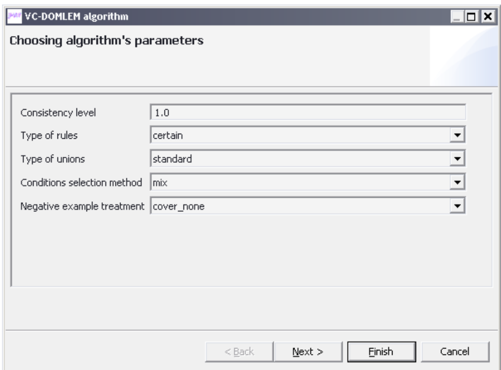

Figure 8: Dialog with parameters of rule induction

To select where the result file with rules will be stored please edit output file in the following dialog (presented in Figure 9).

The resulting rules are presented in results window (see Figure 10).

Statistics of a rule selected in results window can be show by selecting **Open Statistics View associated with selected rule** from toolbar or from the context menu (right click on a rule). Statistics of the first rule are presented in Figure 11.

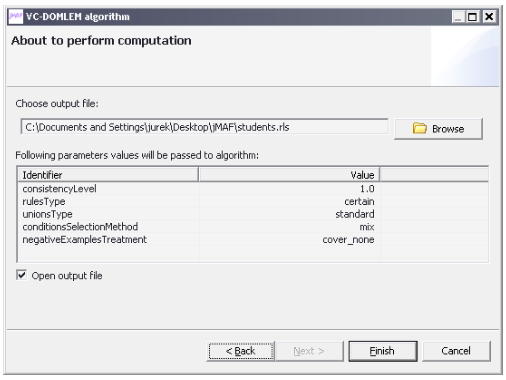

Figure 9: Dialog with parameters of rule induction

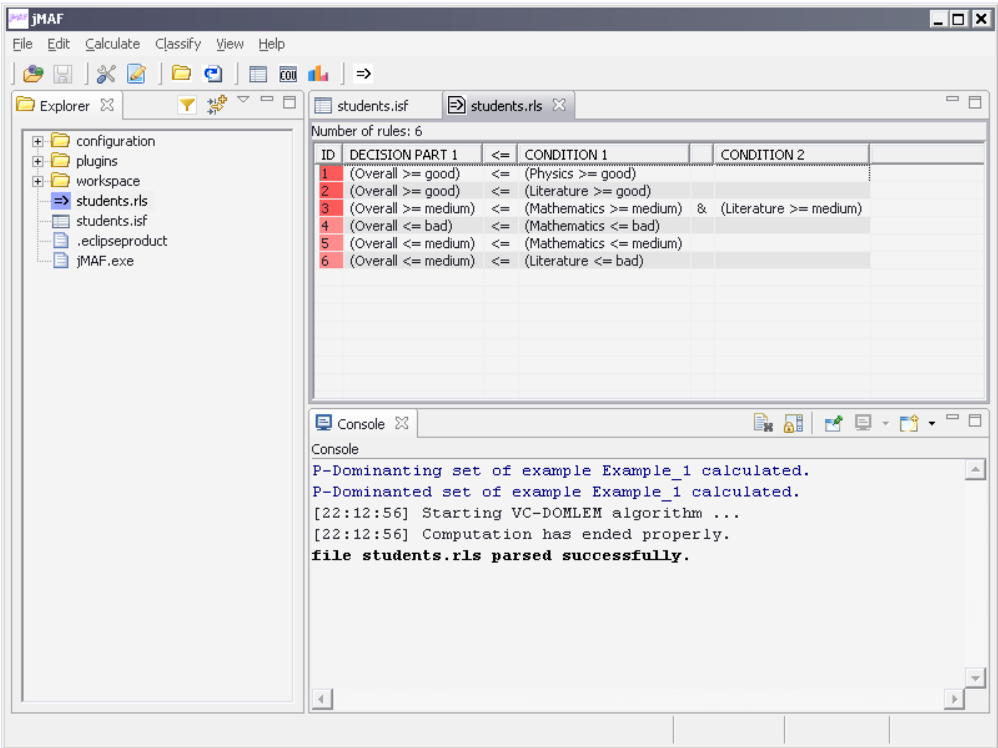

Figure 10: Decision rules

One can also see coverage of a rule (see Figure 12).

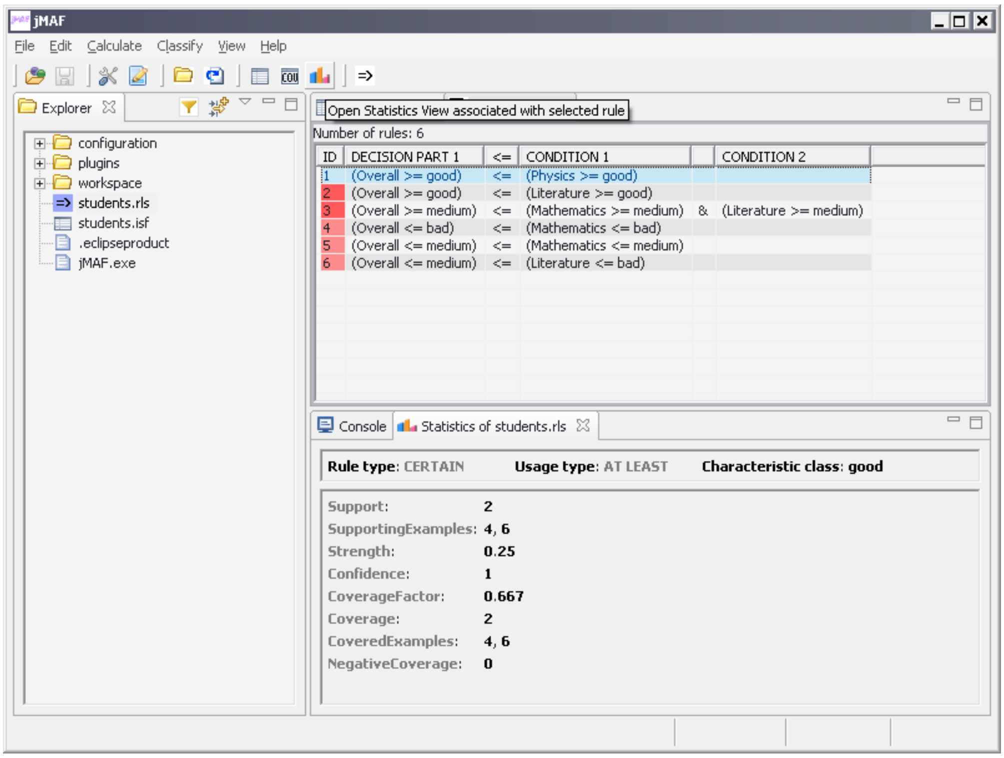

Figure 11: Statistics of the first decision rule

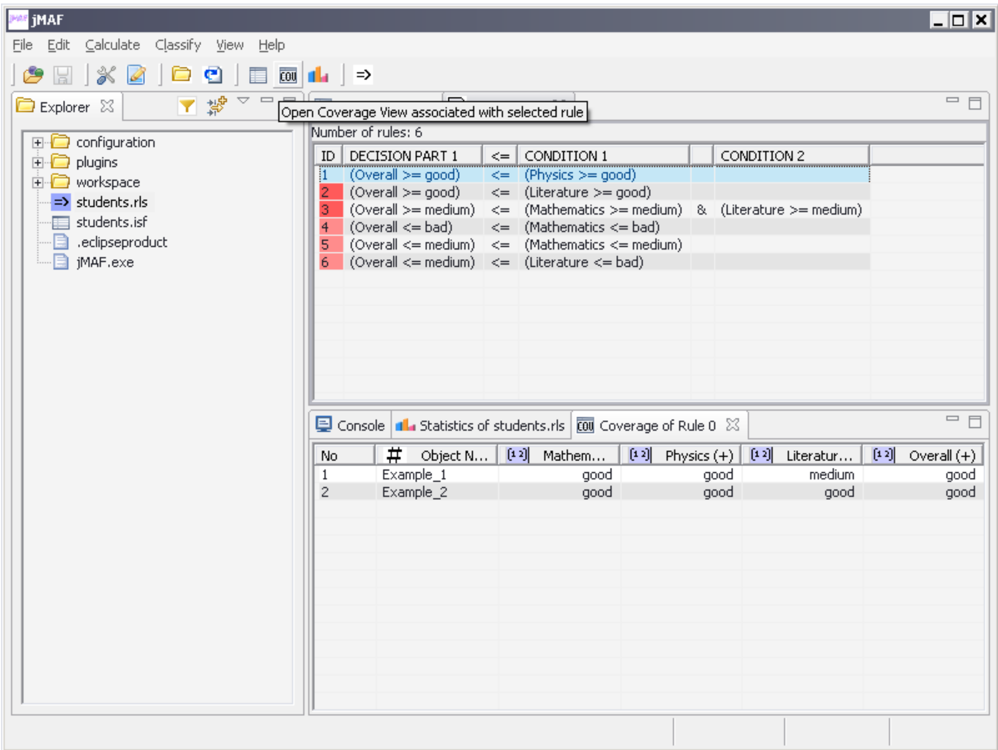

Figure 12: Coverage of the first decision rule

3.8 Classification

Usually data analyst wants to know what is the value of induced rules, i.e., how good they can classify objects. Thus, we proceed with an example of reclassification of learning data table for which rules were induced. To perform reclassification use **Classify** | **Reclassify learning examples**. You will see a dialog with classification options. Select VCDRSA classification method as it is presented in Figure 13. Should you want to know more about VC-DRSA method, please see [1].

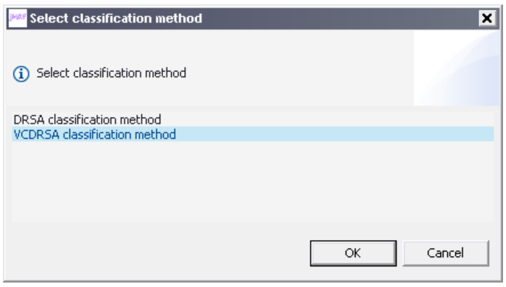

Figure 13: Dialog with classification method

The results of classification are presented in a summary window as it is shown in Figure 14. Use **Details** button to see how particular objects were classified. The resulting window is presented in Figure 15. In this window, it is possible to see rules covering each of the classified examples and their classification.

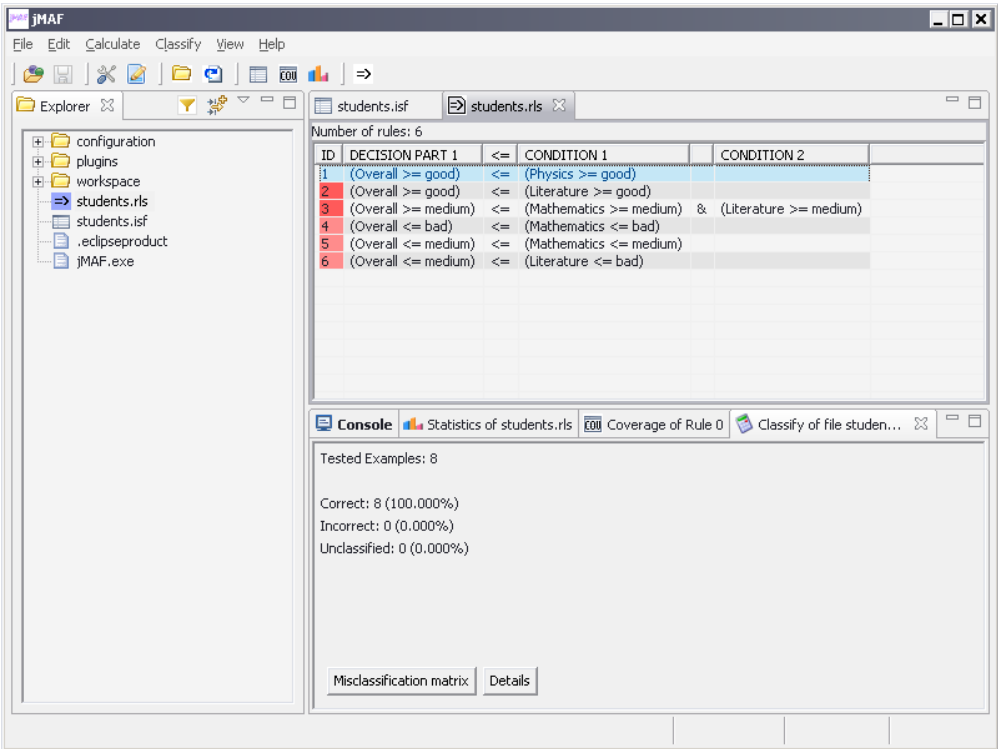

Figure 14: Results of classification

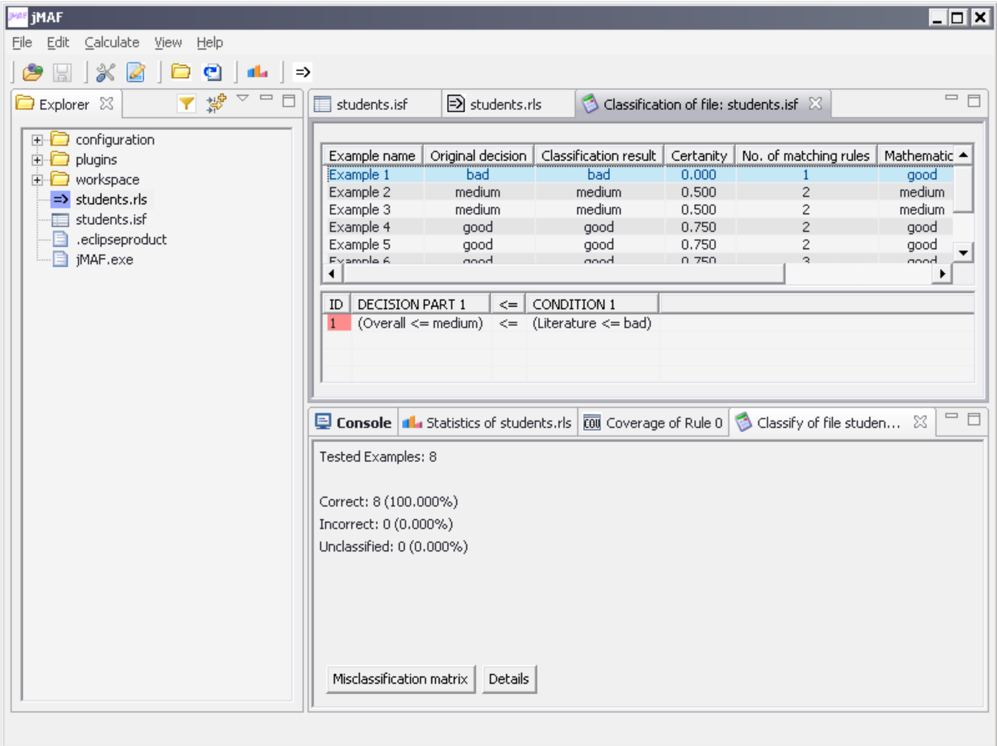

Figure 15: Details of classification

Column "Certainty" in Fig. 15 refers to classification certainty score calculated in a way presented in [1].

4 Exemplary Applications of Dominance-based Rough Set Approach

There are many possibilities of applying DRSA to real life problems. The non-exhaustive list of potential applications includes:

- decision support in medicine: in this area there are already many interesting applications (see, e.g., [27, 22, 23, 35]), however, they exploit the classical rough set approach; applications requiring DRSA, which handle ordered value sets of medical signs, as well as monotonic relationships between the values of signs and the degree of a disease, are in progress;
- customer satisfaction survey: theoretical foundations for application of DRSA in this field are available in [16], however, a fully documented application is still missing;
- bankruptcy risk evaluation: this is a field of many potential applications, as can be seen from promising results reported e.g. in [33, 34, 8], however, a wider comparative study involving real data sets is needed;
- operational research problems, such as location, routing, scheduling or inventory management: these are problems formulated either in terms of classification of feasible solutions (see, e.g., [7]), or in terms of interactive multiobjective optimization, for which there is a suitable IMO-DRSA [18] procedure;
- finance: this is a domain where DRSA for decision under uncertainty has to be combined with interactive multiobjective optimization using IMO-DRSA; some promising results in this direction have been presented in [17];
- ecology: assessment of the impact of human activity on the ecosystem is a challenging problem for which the presented methodology is suitable; the up to date applications are based on the classical rough set concept (see, e.g., [29, 6]), however, it seems that DRSA handling ordinal data has a greater potential in this field.

5 Glossary

Multiple attribute (or multiple criteria) decision support aims at giving the decision maker (DM) a recommendation concerning a set of *objects U* (also called alternatives, actions, acts, solutions, options, candidates,...) evaluated from multiple points of view called *attributes* (also called features, variables, criteria,...).

Main categories of *multiple attribute (or multiple criteria) decision problems* are:

- *classification*, when the decision aims at assigning objects to predefined classes,
- *choice*, when the decision aims at selecting the best object,
- *ranking*, when the decision aims at ordering objects from the best to the worst.

Two kinds of *classification problems* are distinguished:

- *taxonomy*, when the value sets of attributes and the predefined classes are not preference ordered,
- *ordinal classification with monotonicity constraints* (also called *multiple criteria sorting*), when the value sets of attributes and the predefined classes are preference ordered, and there exist monotonic relationships between condition and decision attributes.

Two kinds of *choice problems* are distinguished:

- *discrete choice*, when the set of objects is finite and reasonably small to be listed,
- *multiple objective optimization*, when the set of objects is infinite and defined by constraints of a mathematical program.

If value sets of attributes are preference-ordered, they are called *criteria* or *objectives*, otherwise they keep the name of attributes.

Criterion is a real-valued function f_i defined on U, reflecting a worth of objects from a particular point of view, such that in order to compare any two objects $a, b \in U$ from this point of view it is sufficient to compare two values: $f_i(a)$ and $f_i(b)$.

Dominance: object a is non-dominated in set U (Pareto-optimal) if and only if there is no other object b in U such that b is not worse than a on all considered criteria, and strictly better on at least one criterion.

Preference model is a representation of a value system of the decision maker on the set of objects with vector evaluations.

Rough set in universe U is an approximation of a set based on available information about objects of U. The rough approximation is composed of two ordinary sets, called *lower and upper approximation*. Lower approximation is a maximal subset of objects which, according to the available information, certainly belong to the approximated set, and upper approximation is a minimal subset of objects which, according to the available information, possibly belong to the approximated set. The difference between upper and lower approximation is called *boundary*.

Decision rule is a logical statement of the type "*if..., then...*", where the premise (condition part) specifies values assumed by one or more condition attributes and the conclusion (decision part) specifies an overall judgment.

References

[1] Błaszczyński, J., Greco, S., Słowiński, R.: Multi-criteria classification - A new scheme for application of dominance-based decision rules, European Journal of Operational Research, 3 (2007) 1030-1044

[2] Błaszczyński, J., Greco, S., Słowiński, R, Szeląg, M.: On Variable Consistency Dominance-based Rough Set Approaches. In: LNAI, vol. 4259, Springler-Verlag, Berlin 2006, pp. 191-202

[3] Błaszczyński, J., Greco, S., Słowiński, R, Szeląg, M.: Monotonic Variable Consistency Rough Set Approaches. In: J. Yao, P. Lingras, W. Wu, M. Szczuka, N. J. Cercone, D. Ślęzak (eds.): Rough Sets and Knowledge Technology, LNAI, vol. 4481, Springer-Verlag, 2007, pp. 126-133

[4] Błaszczyński, J., Greco, S., Słowiński, R, Szeląg, M.: Monotonic Variable Consistency Rough Set Approaches. (submitted)

[5] Figueira, J., Greco, S., Ehrgott, M. (eds.), *Multiple Criteria Decision Analysis: State of the Art Surveys*, Springer, Berlin, 2005

[6] Flinkman, M., Michałowski, W., Nilsson, S., Słowiński, R., Susmaga, R., Wilk, S.: Use of rough sets analysis to classify Siberian forest ecosystem according to net primary production of phytomass, *INFOR*, 38 (2000) 145-161

[7] Gorsevski, P.V., Jankowski, P.: Discerning landslide susceptibility using rough sets. *Computers, Environment and Urban Systems*, 32 (2008) 53-65

[8] Greco, S., Matarazzo, B., Słowiński, R.: A new rough set approach to evaluation of bankruptcy risk, in C. Zopounidis (ed.), *Operational Tools in the Management of Financial Risks*, Kluwer, Dordrecht, 1998, pp. 121-136

[9] Greco, S., Matarazzo, B., Słowiński, R.: The use of rough sets and fuzzy sets in MCDM, chapter 14 in T. Gal, T. Stewart, T. Hanne (eds.), *Advances in Multiple Criteria Decision Making*, Kluwer, Boston, 1999, pp. 14.1-14.59

[10] Greco S., Matarazzo, B., Słowiński, R.: Rough approximation of a preference relation by dominance relations. *European J. Operational Research*, 117 (1999) 63-83

[11] Greco, S., Matarazzo, B., Słowiński R.: Rough sets theory for multicriteria decision analysis, *European J. of Operational Research*, 129 (2001) 1-47

[12] Greco, S., Matarazzo, B., Slowinski, R.: Multicriteria classification. [In]: W. Kloesgen, J. Zytkow (eds.), *Handbook of Data Mining and Knowledge Discovery.* Oxford University Press, 2002, chapter 16.1.9, pp. 318-328

[13] Greco, S., Matarazzo, B., Słowiński, R.: Dominance-based Rough Set Approach to Knowledge Discovery, (I) - General Perspective, (II) - Extensions and Applications, chapters 20 and 21 in N. Zhong, J. Liu, *Intelligent Technologies for Information Analysis*, Springer, Berlin, 2004, pp. 513-612

[14] Greco, S., Matarazzo, B., Słowiński, R.: Decision rule approach, chapter 13 in J. Figueira, S. Greco, M. Ehrgott (eds.), *Multiple Criteria Decision Analysis: State of the Art Surveys*, Springer, Berlin, 2005, pp. 507-563

[15] Greco, S., Matarazzo, B., Słowiński, R.: Dominance-based Rough Set Approach as a proper way of handling graduality in rough set theory. *Transactions on Rough Sets VII*, LNCS 4400, Springer, Berlin, 2007, pp. 36-52

[16] Greco, S., Matarazzo, B., Słowiński, R.: Customer satisfaction analysis based on rough set approach. *Zeitschrift für Betriebswirtschaft*, 16 (2007) no.3, 325-339

[17] Greco, S., Matarazzo, B., Słowiński, R.: Financial portfolio decision analysis using Dominance-based Rough Set Approach. Invited paper at the 22nd European Conference on Operational Research (EURO XXII), Prague, 08-11.07.2007

[18] Greco, S., Matarazzo, B., Słowiński, R.: Dominance-based Rough Set Approach to Interactive Multiobjective Optimization, chapter 5 in J.Branke, K.Deb, K.Miettinen, R.Slowinski (eds.), *Multiobjective Optimization: Interactive and Evolutionary Approaches.* Springer, Berlin, 2008

[19] Greco, S., Matarazzo, B., Słowiński, R., Stefanowski, J.: An algorithm for induction of decision rules consistent with dominance principle, in W. Ziarko, Y. Yao (eds.): *Rough Sets and Current Trends in Computing*, LNAI 2005, Springer, Berlin, 2001, pp. 304-313

[20] Greco, S., Matarazzo, B., Słowiński R., Stefanowski, J.: Variable Consistency Model of Dominance-based Rough Sets Approach. In: W. Ziarko, Y. Yao (eds.): Rough Sets and Current Trends in Computing, LNAI, vol. 2005, Springler-Verlag, Berlin 2001, pp. 170-181

[21] Greco, S., Pawlak, Z., Słowiński, R.: Can Bayesian confirmation measures be useful for rough set decision rules? *Engineering Applications of Artificial Intelligence*, 17 (2004) 345-361

[22] Michałowski, W., Rubin, S., Słowiński, R., Wilk, S.: Mobile clinical support system for pediatric emergencies. *Journal of Decision Support Systems*, 36 (2003) 161-176

[23] Michałowski, W., Wilk, S., Farion, K., Pike, J., Rubin, S., Słowiński, R.: Development of a decision algorithm to support emergency triage of scrotal pain and its implementation in the MET system. *INFOR*, 43 (2005) 287-301

[24] Pawlak, Z.: Rough Sets, *International Journal of Computer and Information Sciences*, 11 (1982) 341-356

[25] Pawlak, Z.: *Rough Sets*, Kluwer, Dordrecht, 1991

[26] Pawlak, Z., Skowron, A., Rough Membership Functions. In: R. R. Yager, M. Fedrizzi and J. Kacprzyk (eds.): Advances in the Dempster-Shafer Theory of Evidence, Wiley, New York 1994, pp. 251-271

[27] Pawlak, Z., Słowiński, K., Słowiński, R.: Rough classification of patients after highly selective vagotomy for duodenal ulcer. *International Journal of Man-Machine Studies*, 24 (1986) 413-433

[28] Pawlak, Z., Słowiński, R.: Rough set approach to multi-attribute decision analysis. *European J. of Operational Research*, 72 (1994) 443-459

[29] Rossi, L., Słowiński, R., Susmaga, R.: Rough set approach to evaluation of stormwater pollution. *International Journal of Environment and Pollution*, 12 (1999) 232-250

[30] Słowiński, R.: Rough set learning of preferential attitude in multi-criteria decision making, in J. Komorowski, Z. W. Ras (eds.), *Methodologies for Intelligent Systems*, LNAI 689, Springer, Berlin, 1993, pp. 642-651

[31] Słowiński, R., Greco, S., Matarazzo, B.: Rough set analysis of preference-ordered data, in J.J. Alpigini, J.F. Peters, A. Skowron, N. Zhong (eds.), *Rough Sets and Current Trends in Computing*, LNAI 2475, Springer, Berlin, 2002, pp. 44-59

[32] Słowiński, R., Greco, S., Matarazzo, B.: Dominance-based rough set approach to reasoning about ordinal data, keynote lecture in M.Kryszkiewicz, J.F.Peters, H.Rybiñski, A.Skowron (eds.), *Rough Sets and Intelligent Systems Paradigms*. LNAI 4585, Springer, Berlin, 2007, pp. 5-11

[33] Słowiński, R., Zopounidis, C.: Application of the rough set approach to evaluation of bankruptcy risk. *International Journal of Intelligent Systems in Accounting, Finance and Management* 4 (1995) 27-41

[34] Słowiński, R., Zopounidis, C., Dimitras, A.I.: Prediction of company acquisition in Greece by means of the rough set approach. *European Journal of Operational Research*, 100 (1997) 1-15

[35] Wilk, S., Słowiński, R., Michałowski, W., Greco, S.: Supporting triage of children with abdominal pain in the emergency room. *European Journal of Operational Research*, 160 (2005) 696-709

[36] Wong, S. K. M., Ziarko, W.: Comparison of the probabilistic approximate classification and the fuzzy set model. Fuzzy Sets and Systems, vol. 21, 1987, pp. 357-362