Control and Cybernetics

vol. **31** (2002) No. 4

Axiomatization of utility, outranking and decision rule preference models for multiple-criteria classification problems under partial inconsistency with the dominance principle

by

Roman Słowiński[1], Salvatore Greco[2] and Benedetto Matarazzo[2]

[1] Institute of Computing Science, Poznań University of Technology, Piotrowo 3a, 60-965 Poznań, Poland
E-mail: slowinsk@sol.put.poznan.pl

[2] Faculty of Economics, University of Catania, Corso Italia, 55, 95129 Catania, Italy
E-mail: {salgreco, matarazz}@mbox.unict.it

Abstract: Multiple-criteria classification (sorting) problem concerns assignment of actions (objects) to some pre-defined and preference-ordered decision classes. The actions are described by a finite set of criteria, i.e. attributes, with preference-ordered scales. To perform the classification, criteria have to be aggregated into a preference model which can be: utility (discriminant) function, or outranking relation, or "if..., then..." decision rules. Decision rules involve partial profiles on subsets of criteria and dominance relation on these profiles. A challenging problem in multiple-criteria decision making is the aggregation of criteria with ordinal scales. We show that the decision rule model we propose has advantages over a general utility function, over the integral of Sugeno, conceived for ordinal criteria, and over an outranking relation. This is shown by basic axioms characterizing these models. Moreover, we consider a more general decision rule model based on the rough set theory. The advantage of the rough set approach compared to competitive methodologies is the possibility of handling partially inconsistent data that are often encountered in preferential information, due to hesitation of decision makers, unstable character of their preferences, imprecise or incomplete knowledge and the like. We show that these inconsistencies can be represented in a meaningful way by "if..., then..." decision rules induced from rough approximations. The theoretical results reported in this paper show that the decision rule model is the most general aggregation model among all the considered models.

Keywords: multiple-criteria classification, preference modeling, decision rules, conjoint measurement, ordinal criteria, rough sets, axiomatization.

1. Introduction

Making decisions with respect to a set of actions (objects) requires, usually, taking into account multiple criteria representing different and conflicting points of view on evaluation of the actions. Multiple-criteria decision making (MCDM) may concern either a choice of the best action, or a ranking, or a classification of the set of actions (see Roy, 1985; Gal, Stewart and Hanne, 1999). It is well known that a MCDM problem is mathematically ill posed, because the only objective information that follows from its mathematical formulation is the dominance relation. Action x dominates action y if x is at least as good as y on all considered criteria. If it is the case, x is objectively better than y. Apart from trivial problems, this information does not permit, however, to solve a particular MCDM problem because the most interesting actions, that are non-dominated, are non-comparable to each other, unless one gives additional information permitting to aggregate multiple criteria into a single preference model. The preference model establishes a preference structure in the set of actions; a suitable exploitation of this structure yields a recommendation of the "best compromise solution" for either choice, or ranking, or multi-criteria classification.

The information permitting to move forward the solution process is called preferential information; its acquisition, construction of the preference model and, finally, exploitation of the preference structure involve a single or multiple decision makers (DM) in the solution process – for this reason this is not an automatic solution procedure but a decision aiding method.

A much-desired feature of decision aiding is transparency of the methodology that should be intelligible, at least intuitively, for the users. The transparency may enhance the confidence to recommendations and facilitate their robustness analysis. There are two elements of crucial importance for the transparency: the type of a preference model and the type of DM's preferential information used for its construction. Very often the model adopted requires this information to be given in terms of preference model parameters, such as importance weights, substitution ratios and various thresholds. Giving such information requires a great cognitive effort of the DM.

It is generally acknowledged, however, that people prefer to make exemplary decisions rather than to explain them in terms of specific parameters. For this reason, the idea of inferring preference models from exemplary decisions provided by the DM is compatible with the aim of transparency. Artificial intelligence and, particularly, inductive learning approach, submits a simple idea of inferring the preference model in terms of decision rules being logical statements of the type "if..., then...". Such preference model is comprehensible for the users, because it speaks the language of examples, and its distributed form is able to represent local trade-offs and dependencies among criteria that are hidden by more synthetic models, like a utility function.

The exemplary decisions may, however, be inconsistent because of limited discriminatory power of criteria and due to hesitation of DMs, unstable character of their preferences, imprecise or incomplete information, etc. (see Roy, 1989). Inconsistent examples concern pairs of actions (x, y) such that x dominates y, however, x has been assigned to a worse class than y. These inconsistencies cannot be considered as simple error or noise. They can convey important information that should be taken into account in the construction of the DM's preference model. To deal with this problem, the authors have adapted the rough set theory (Pawlak, 1982, 1991; Pawlak and Słowiński, 1994) that, in its classic form, is not able to use the information about preference order in attribute domains, i.e. about scales of criteria (Greco, Matarazzo and Słowiński, 1999, 2000, 2001a, 2002a). The authors' extension of the rough set concept permits a separation of certain and doubtful knowledge about the DM's preferences by distinction of different kinds of decision rules, depending whether they are induced from lower approximations of decision classes or from the boundaries of these classes composed of inconsistent examples that do not observe the dominance principle.

The above two elements, preference information (possibly inconsistent) in terms of exemplary decisions and preference model in terms of rules, are the main features of our methodology characterized in this paper with respect to multiple-criteria classification problem (also called multiple-criteria sorting problem). Let us recall that classification concerns an assignment of a set of actions to a set of pre-defined decision classes. The actions are described by a set of criteria, i.e. attributes with preference-ordered domains (scales). The decision classes are preference-ordered.

Although multi-attribute classification is the most popular problem considered also in Artificial Intelligence and its derivative – Knowledge Discovery and Data Mining – this methodology ignores the preference scale that gives to regular attributes the meaning of criteria. On the other hand, the usual MCDM methodology assumes that all attributes are criteria. Our methodology, based on the extended rough set approach, permits to take into account both regular attributes and criteria in multiple-criteria classification, which makes sense in many real decision problems.

Summing up this introduction, one can remark that the central problem of any decision-aiding methodology proposed for multiple-criteria and/or multiple-attribute classification is the aggregation of the multiple criteria and attributes into a single preference model. In this paper, we propose to compare different paradigms used to solve this central problem by different theories (see Table 1.1). This comparison will be made at the level of axiomatic foundations, which has no precedence in the theoretical research concerning multi-criteria classification. The axiomatic approach is interesting for at least three reasons:

- it exhibits differences between preference models and methods,
- it permits to interpret methods conceived for one model in terms of another model,
- knowing the basic axioms, one can pass from one method to another with different preference models.

Moreover, we will consider aggregation of ordinal criteria that has been studied much less than that of cardinal criteria (see Roberts, 1979). Among several aggregation models, a particular interest has been paid recently for an integral proposed by Sugeno (1974), able to deal with ordinal data; it has been considered the most general ordinal aggregation operator of the max-min average type. It appears, however, that this operator has some unpleasant limitations: the most important is the so-called commensurability (Modave and Grabisch, 1998), i.e. the evaluations with respect to each considered criterion should be defined on the same scale. Comparison of the Sugeno integral with the decision rule model at the axiomatic level permits to show other limitations of the former.

Table 1.1. Different paradigms of aggregation and preference representation

Theory (paradigm)	Main preoccupation (axiomatic basis)	The aggregation model shows
Decision Theory	Definition of preference structures	Relation in X
Measurement Theory	Cancellation property	Function, like in conjoint measurement
Measure Theory & Fuzzy sets	Capacity or fuzzy measure	Weights or interactions among criteria, like in Choquet integral or Sugeno integral
Artificial Intelligence, Logical Analysis of Data & Rough Sets	Boolean or pseudo-Boolean function, decision rules or decision trees	Knowledge, like in knowledge discovery or data mining

This article is extending a preliminary version (Greco, Matarazzo and Słowiński, 2001b) that did not include all the results related here. For the reason of space limitation, we are omitting formal proofs. The article is organized as follows. In Section 2, an axiomatic characterization of multiple-criteria classification is presented. The main result is a theorem proving equivalence of four elements: a simple cancellation condition, a utility function with a set of thresholds which works as a discriminant function, an outranking function which, together with a set of reference actions, operates as a decision model of ELECTRE TRI-like methods (Roy and Bouyssou, 1993), and a set of "*if...*, *then...*" decision rules. In Section 2, we give, moreover, a theorem characterizing multiple-criteria classification based on Sugeno integral utility function, in terms of a simple cancellation condition or, equivalently, in terms of a set of "*if...*, *then...*" decision rules having a specific syntax. Sections 3 presents main steps of the rough set approach to multiple-criteria classification. Using the concepts recalled in Section 3, we give in Section 4 a result on representation of multiple-criteria classification in case of inconsistent data. Section 5 groups conclusions.

2. Axiomatic foundations of multiple-criteria classification and associated preference models

2.1. A representation theorem

$X = \prod_{i=1}^{n} X_i$, where X_i is an evaluation scale of *criterion* $i = 1, \ldots, n$. With appropriate topological conditions we can also work with infinite non-denumerable space, but in this paper, for the sake of simplicity, we will skip this possibility. When aggregating multi-criteria evaluations within a preference model, we will exploit the *ordinal* character of the criteria scales only. This means that evaluations on particular criteria are considered as if they were words (linguistic qualifiers, like *bad, medium, good, very good*) even if an original scale was numerical. Let $(x_i z_{-i})$, $x_i \in X_i$ and $z_{-i} \in X_{-i} = \prod_{j=1, j\neq i}^{n} X_j$, denote an element of X equal to z except for its i-th coordinate being equal to x_i.

Moreover, let $\mathbf{Cl} = \{Cl_t, t \in T\}$, $T = \{1, \ldots, m\}$, be a set of classes of X, such that each $x \in X$ belongs to one and only one class $Cl_t \in \mathbf{Cl}$ and no class Cl_t is empty. We suppose, moreover, that the classes of $\mathbf{Cl}$ are increasingly ordered, i.e. for all $r, s \in T$, such that $r > s$, the elements of Cl_r have a better comprehensive evaluation than the elements of Cl_s. In consequence, the classes of $\mathbf{Cl}$ are equivalence classes of a weak preference relation $\succeq$ being a complete preorder, and we say that $\mathbf{Cl}$ is a classification in X.

Let us also consider the following *upward and downward unions of classes*, respectively,

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

Observe that $Cl_1^{\geq} = Cl_m^{\leq} = X$, $Cl_m^{\leq} = Cl_m$ and $Cl_1^{\leq} = Cl_1$. We will use these unions in the syntax of decision rules in order to handle the preference order of classes and to respect the *dominance principle*. It requires that actions having not-worse evaluation with respect to a set of considered criteria than a referent action cannot be assigned to a worse class than the referent action.

The classification decision is generally modeled by one of three models: utility function (scoring methods, Capon, 1982, discriminant analysis, Altman, 1968, UTADIS, Zopounidis and Doumpos, 1998, etc.), outranking relation (as in ELECTRE TRI, Roy and Bouyssou, 1993) or decision rules (as in Dominance-Based Rough Set Approach, Greco, Matarazzo and Słowiński, 1999, 2001a, 2002a):

- ▷ *Utility function* $f(\cdot)$ gives a real value $f(x)$ to each $x \in X$ and assigns x to $Cl_t^{\geq}$ if $f(x) \geq z_t$, where z_t, $t = 2, \ldots, m$, are $m-1$ ordered thresholds satisfying

  $$z_2 < z_3 < \ldots < z_m.$$

- ▷ *Outranking relation* S is a binary relation on X such that for each $x, y \in X$, xSy means "x is (comprehensively) at least as good as y". An outranking relation S on X assigns x to $Cl_t^{\geq}$ if xSp^t, where p^t, $t = 2, \ldots, m$, are $m-1$ reference profiles p^t, such that p^{t+1} dominates p^t (i.e. p^{t+1} is at least as good as p^t with respect to each criterion i and there is at least one criterion for which p^{t+1} is strictly preferred to p^t), $t = 2, \ldots, m-1$.
- ▷ *A set of "if..., then..." decision rules* is a set of logical implications whose syntax is "if x_{i1} is at least as good as r_{i1} *and* x_{i2} is at least as good as

and ... and x_ih is at least as good as r_{ih}, *then* $x \in Cl_r^{\geq}$", where $x_{i1}, r_{i1} \in X_{i1}$, $x_{i2}, r_{i2} \in X_{i2}$, ..., $x_{ih}, r_{ih} \in X_{ih}$, with $\{i1, i2, \ldots, ih\} \subseteq \{1, \ldots, n\}$ and $r = 2, \ldots, m$. These decision rules are called "*at least*" decision rules. Let us consider the case where for each criterion $i = 1, \ldots, n$ there exists a function $g_i : X_i \to R$ such that for each $x, y \in X$: $g_i(x_i) \geq g_i(y_i) \Leftrightarrow x$ is at least a good as y with respect to criterion i (i.e. x_i is at least as good as y_i). In this case, an "at least" decision rule can also be written as

"*if* $g_{i1}(x_{i1}) \geq g_{i1}(r_{i1})$ *and* $g_{i2}(x_{i2}) \geq g_{i2}(r_{i2})$ *and* ...

and $g_{ih}(x_{ih}) \geq g_{ih}(r_{ih})$, then $x \in Cl_r^{\geq}$"

with $\{i1, i2, \ldots, ih\} \subseteq \{1, \ldots, n\}$ and $r = 2, \ldots, m$. The classification of $x \in X$ with "at least" decision rules is done according to the following procedure:

- $x \in Cl_t$ if and only if there exists a rule matching x that assigns x to $Cl_t^{\geq}$, and there exists no rule matching x that assigns x to $Cl_s^{\geq}$, where $s > t$;
- $x \in Cl_1$ if and only if there exists no rule matching x.

The following result is a representation theorem for the multiple-criteria classification problem, stating the equivalence between a very simple cancellation property, a general utility function, a very general outranking relation and a set of decision rules. Let us mention that equivalence of the considered cancellation property and the utility function was already noted by Goldstein (1991), within the conjoint measurement approach, for the special case of three classes.

THEOREM 2.1 *The following four propositions are equivalent:*

1) (cancellation property) for each $i = 1, \ldots, n$, *for each* $x_i, y_i \in X_i$ *and* $a_{-i}, b_{-i} \in X_{-i}$, *and for each* $r, s \in \{1, \ldots, m\}$:

$$\{(x_i a_{-i}) \in Cl_r \text{ and } (y_i b_{-i}) \in Cl_s\}$$
$$\Rightarrow \{(y_i a_{-i}) \in Cl_r^{\geq} \text{ or } (x_i b_{-i}) \in Cl_s^{\geq}\},$$

2) (utility function) there exist

- *functions* $g_i : X_i \to R$ *for each* $i = 1, \ldots, n$, *called criteria,*
- *function* $f : \mathbf{R}^n \to \mathbf{R}$, *non-decreasing in each argument, called utility function,*
- $m - 1$ *ordered thresholds* z_t, $t = 2, \ldots, m$, *satisfying*

  $$z_2 < z_3 < \ldots < z_m$$

  such that for each $x \in X$ *and each* $t = 2, \ldots, m$,

  $$f[g_1(x_1), g_2(x_2), \ldots, g_n(x_n)] \geq z_t \Leftrightarrow x \in Cl_t^{\geq},$$

3) (outranking function and relation) there exist

- *functions* $g_i : X_i \to \mathbf{R}$, $i = 1, \ldots, n$, *called criteria,*

- *function* $s : \mathbf{R}^{2n} \to \mathbf{R}$, *non-decreasing in each odd argument and non-increasing in each even argument, called outranking function,*
- $m-1$ *reference profiles* p^t, $t = 2, \ldots, m$, *satisfying*

  $$g_i(p^2) \leq g_i(p^3) \leq \ldots \leq g_i(p^m), \quad \text{for } i = 1, \ldots, n,$$

  such that for each $x \in X$ *and each* $t = 2, \ldots, m$

  $$s[g_1(x_1), g_1(p^t), g_2(x_2), g_2(p^t), \ldots, g_n(x_n), g_n(p^t)] \geq 0 \Leftrightarrow x \in Cl_t^{\geq},$$

  (N.B. $s[g_1(x_1), g_1(p^t), g_2(x_2), g_2(p^t), \ldots, g_n(x_n), g_n(p^t)] \geq 0 \Leftrightarrow xSp^t$, *where* S *is a binary outranking relation),*

4) ("at least" decision rules) there exist

- *functions* $g_i : X_i \to \mathbf{R}$ *for each* $i = 1, \ldots, n$, *called criteria,*
- *a set of "at least" decision rules whose syntax is*

  $$\text{"if } g_{i1}(x_{i1}) \geq r_{i1} \text{ and } g_{i2}(x_{i2}) \geq r_{i2} \text{ and } \ldots$$

  $$\text{and } g_{ih}(x_{ih}) \geq r_{ih}, \text{ then } x \in Cl_t^{\geq}\text{"},$$

  with $\{i1, i2, \ldots, ih\} \subseteq \{1, \ldots, n\}$, $t = 2, \ldots, m$, *such that for each* $y \in Cl_t$, $t = 2, \ldots, m$, *there is at least one rule implying* $y \in Cl_t^{\geq}$ *and there is no rule implying* $y \in Cl_r^{\geq}$, *with* $r > t$.

Let us remark that the above representation theorem for multiple-criteria classification problem starts with a very weak axiomatic condition called cancellation property. Indeed, this property does not require existence of criterion functions g_i, $i = 1, \ldots, n$, or a dominance relation D on X in order to characterize the three preference models. Instead, for $i = 1, \ldots, n$, it permits to define a binary weak preference relation $\succeq_i$ on X_i which is a complete preorder. Consequently, there exists a function $g_i : X_i \to \mathbf{R}$ such that for each $x_i, y_i \in X_i$

$$x_i \succeq_i y_i \Leftrightarrow g_i(x_i) \geq g_i(y_i).$$

On the basis of relations $\succeq_i$, $i = 1, \ldots, n$, one can also define a dominance relation D on X as follows: for each $x, y \in X$

$$xDy \Leftrightarrow x_i \succeq_i y_i \quad \text{for all } i = 1, \ldots, n.$$

This is of course equivalent to

$$xDy \Leftrightarrow g_i(x_i) \geq g_i(y_i) \quad \text{for all } i = 1, \ldots, n.$$

Cancellation property 1) of Theorem 2.1. permits to state the following *condition of coherence between dominance relation* D *and classification* **Cl**, for each $x, y \in X$

$$xDy \Rightarrow x \in Cl_r \text{ and } y \in Cl_s, \text{ with } r \geq s.$$

For any subset of criteria $P \subseteq \{1,\ldots,n\}$ and for each pair $x, y \in X$ one can also define a dominance relation D_P on X:

$$xD_Py \Leftrightarrow x_i \succeq_i y_i \quad \text{for all } i \in P,$$

which is equivalent to

$$xD_Py \Leftrightarrow g_i(x_i) \geq g_i(y_i) \quad \text{for all } i \in P.$$

Dominance relations D_P, $P \subseteq \{1,\ldots,n\}$, are used in the condition part of decision rules. Being an intersection of complete preorders, binary relations D_P are partial preorders, i.e. they are reflexive and transitive.

Observe, moreover, that Theorem 2.1 regards a representation of classification **Cl** in terms of "lower bounds". Theorem 2.1 can be reformulated in terms of "upper bounds" in such a way that

- ▷ condition of proposition 2) is expressed as
  $$f[g_1(x_1), g_2(x_2), \ldots, g_n(x_n)] \leq w_t \Leftrightarrow x \in Cl_t^{\leq},$$
  where w_t, $t = 1,\ldots,m-1$, are $m-1$ suitably ordered thresholds,
- ▷ condition of proposition 3) is expressed as
  $$s[g_1(x_1), g_1(q^t), g_2(x_2), g_2(q^t), \ldots, g_n(x_n), g_n(q^t)] < 0 \Leftrightarrow x \in Cl_t^{\leq}$$
  where q^t, $t = 1,\ldots,m-1$, are $m-1$ reference profiles q^t, such that q^{t+1} dominates q^t (i.e. q^{t+1} is at least as good as q^t with respect to each criterion $i = 1,\ldots,n$, and there is at least one criterion $j \in \{1,\ldots,n\}$ for which q^{t+1} is strictly preferred to q^t), $t = 1,\ldots,m-2$.
- ▷ condition of proposition 4) considers a set of decision rules whose syntax is

  "*if* $g_{i1}(x_{i1}) \leq r_{i1}$ *and* $g_{i2}(x_{i2}) \leq r_{i2}$ *and* ...
  and $g_{ih}(x_{ih}) \leq r_{ih}$, then $x \in Cl_r^{\leq}$"

  with $\{i1, i2, \ldots, ih\} \subseteq \{1,\ldots,n\}$. These decision rules are called "*at most*" decision rules. The classification of $x \in X$ with "at most" decision rules is done according to the following procedure:

  - $x \in Cl_t$ if and only if there exists a rule matching x that assigns x to $Cl_t^{\leq}$, and there exists no rule matching x that assigns x to $Cl_s^{\leq}$, where $s < t$;
  - $x \in Cl_m$ if and only if there exists no rule matching x.

The reformulation of Theorem 2.1 in terms of "upper bounds" is as follows.

THEOREM 2.2 *The following four propositions are equivalent:*

1) *(cancellation property) for each* $i = 1,\ldots,n$, *for each* $x_i, y_i \in X_i$ *and* $a_{-i}, b_{-i} \in X_{-i}$, *and for each* $r, s \in \{1,\ldots,m\}$:

$$\{(x_i a_{-i}) \in Cl_r \text{ and } (y_i b_{-i}) \in Cl_s\}$$
$$\Rightarrow \{(y_i a_{-i}) \in Cl_r^{\leq} \text{ or } (x_i b_{-i}) \in Cl_s^{\leq}\},$$

2) *(utility function) there exist*

- *functions* $g_i : X_i \to \mathbf{R}$, $i = 1,\ldots,n$, *called criteria*,

- *function* $f : \mathbf{R}^n \to \mathbf{R}$, *non-decreasing in each argument, called utility function,*
- $m-1$ *ordered thresholds* $w_t,\ t=1,\ldots,m-1$, *satisfying*

  $$w_1 < w_2 < \ldots < w_{m-1}$$

  such that for each $x \in X$ *and each* $t=1,\ldots,m-1$

  $$f[g_1(x_1), g_2(x_2), \ldots, g_n(x_n)] \le w_t \Leftrightarrow x \in Cl_t^{\leq};$$

3) (outranking function and relation) there exist

- *functions* $g_i : X_i \to \mathbf{R}$ *for each* $i=1,\ldots,n$, *called criteria,*
- *function* $s : \mathbf{R}^{2n} \to \mathbf{R}$, *non-decreasing in each odd argument and non-increasing in each even argument, called outranking function,*
- $m-1$ *reference profiles* $q^t,\ t=1,\ldots,m-1$, *satisfying*

  $$g_i(q1) \le g_i(q2) \le \ldots \le g_i(q^{m-1}), \quad for\ i=1,\ldots,n$$

  such that for each $x \in X$ *and each* $t=1,\ldots,m-1$

  $$s[g_1(x_1), g_1(q^t), g_2(x_2), g_2(q^t), \ldots, g_n(x_n), g_n(q^t)] < 0 \Leftrightarrow x \in Cl_t^{\leq},$$

(N.B. $s[g_1(x_1), g_1(q^t), g_2(x_2), g_2(q^t), \ldots, g_n(x_n), g_n(q^t)] < 0 \Leftrightarrow q^t S x$, *where* S *is a binary outranking relation),*

4) ("at most" decision rules) there exist

- *functions* $g_i : X_i \to \mathbf{R}$ *for each* $i=1,\ldots,n$, *called criteria,*
- *a set of decision rules whose syntax is*

  "if $g_{i1}(x_{i1}) \le r_{i1}$ *and* $g_{i2}(x_{i2}) \le r_{i2}$ *and* ...

  and $g_{ih}(x_{ih}) \le r_{ih}$, *then* $x \in Cl_t^{\leq}$*",*

  with $\{i1, i2, \ldots, ih\} \subseteq \{1,\ldots,n\}$, $t=1,\ldots,m-1$,
  such that for each $y \in Cl_t$, $t=1,\ldots,m-1$, *there is at least one rule implying* $y \in Cl_t^{\leq}$ *and there is no rule implying* $y \in Cl_r^{\leq}$, *with* $r<t$.

Another interesting question concerning Theorem 2.1 is that proposition 1) can be reformulated as follows:

$$\{(x_i a_{-i}) \in Cl_r \text{ and } (y_i b_{-i}) \in Cl_s\} \to \{(y_i a_{-i}) \in Cl_r^{\leq} \text{ or } (x_i b_{-i}) \in Cl_s^{\leq}\}.$$

This is formally stated by the following result.

THEOREM 2.3 *The following two propositions are equivalent for each* $i=1,\ldots,n$
1) for each $x_i, y_i \in X_i$ *and* $a_{-i}, b_{-i} \in X_{-i}$, *and for each* $r,s \in \{1,\ldots,m\}$*:*

$$\{(x_i a_{-i}) \in Cl_r \textit{ and } (y_i a_{-i}) \in Cl_s\} \Rightarrow \{(y_i a_{-i}) \in Cl_r^{\leq} \textit{ or } (x_i b_{-i}) \in Cl_r^{\leq}\},$$

2) for each $x_i, y_i \in X_i$ *and* $a_{-i}, b_{-i} \in X_{-i}$, *and for each* $r,s \in \{1,\ldots,m\}$*:*

$$\{(x_i a_{-i}) \in Cl_r \textit{ and } (y_i b_{-i}) \in Cl_s\} \Rightarrow \{(y_i a_{-i}) \in Cl_r^{\leq} \textit{ or } (x_i b_{-i}) \in Cl_s^{\leq}\}.$$

2.2. An example

Let us consider a multiple-criteria classification problem inspired by the example of evaluation in a high school proposed by Grabisch (1995). Suppose that a high school director wants to assign students to different classes of merits on the basis of their scores in Mathematics and Literature. The ordinal scale of evaluation in Mathematics and Literature has been composed of three following grades: "bad", "medium" and "good", while the comprehensive evaluation scale has been composed of two grades: "bad" and "good". The evaluations of student x in Mathematics and Literature are denoted by x_1 and x_2, respectively. To be criteria, functions $g_1(\cdot)$ and $g_2(\cdot)$ must respect monotonicity, i.e. $g_i(bad) < g_i(medium) < g_i(good)$, $i = 1, 2$. For example, a simple way to define $g_1(\cdot)$ and $g_2(\cdot)$ is to set $g_i(bad) = 1$, $g_i(medium) = 2$, $g_i(good) = 3$, $i = 1, 2$.

Table 2.1. Classification of all nine profiles of possible evaluations

Student	Mathematics	Literature	Decision	$f[g_1(x), g_2(x)]$	Matching rules
S1	bad	bad	bad	0	#3
S2	medium	bad	bad	1	#3
S3	good	bad	good	4	#1
S4	bad	medium	bad	1	#3
S5	medium	medium	good	4	#2
S6	good	medium	good	6	#1,2
S7	bad	good	bad	2	#3
S8	medium	good	good	5	#2
S9	good	good	good	8	#1,2

Table 2.1 presents all the possible profiles of students with respect to the two considered criteria, and a classification decision made by the director. Let us observe that the classification of students presented in Table 2.1 satisfies proposition 1) of Theorem 2.1. In fact, each time a student x dominates a student y, student x belongs to the same or higher class than student y.

This can also be seen on the Hasse diagram in Fig. 2.1 where each node corresponds to a profile of evaluations. Profile x corresponding to node α dominates over the profile y corresponding to node β if α is over β and there is a path from α to β.

The diagram in Fig. 2.2 represents the binary relation R defined on the set of all possible profiles of evaluations $X = \{S1, S2, \ldots, S9\}$ as follows: for each $x, y \in X$

$$xRy \Leftrightarrow xDy \text{ or } x \in Cl_r \text{ and } y \in Cl_s, \text{ with } r > s.$$

Binary relation R is used in the proof of Theorem 2.1. In fact, R is reflexive and transitive and, therefore, it is a partial preorder. Thus, there is a function $h : X \to R$ such that for each $x, y \in X$

$$xRy \Rightarrow h(x) \geq h(y).$$

According to the definition and to the condition of coherence between dominance relation D and classification $\mathbf{Cl}$, we have

$$x \in Cl_r \text{ and } y \in Cl_s, \text{ with } r \geq s \Leftrightarrow h(x) \geq h(y). \tag{i}$$

On the basis of property (i), it is possible to build function $f : \mathbf{R}^n \to \mathbf{R}$ and set of thresholds z_t, $t = 2, \ldots, m$, used in Theorem 2.1.

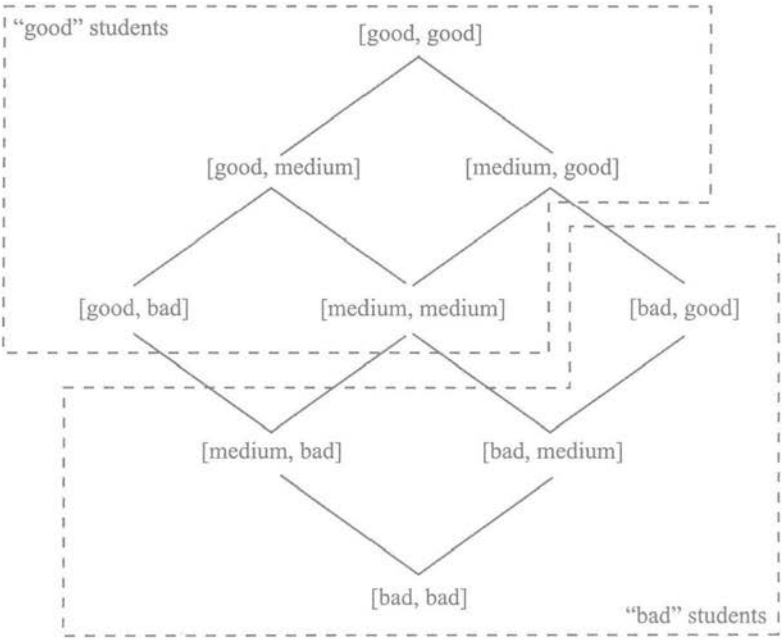

Figure 2.1. Hasse diagram profiles $x = [x_1, x_2]$ where x_1 = score in Mathematics, x_2 = score in Literature

In the diagram presented in Fig. 2.2, the arcs representing relation R are drawn from profile x to profile y if and only if x dominates y or x belongs to the class of "good" students and y belong to the class of "bad" students. Since relation R is transitive, we are not drawing the arcs between profiles that are already connected by a path; e.g. profiles "good-good" and "bad-good" are in relation R, however, there exist a path between them through the profile "medium-good", so the direct arc would be redundant.

Comparing Figs. 2.1 and 2.2, one can see that in the latter there are two additional arcs. The arc from profile "medium-medium" to profile "bad-good" does not exist in Fig. 2.1 because profile "medium-medium" does not dominate profile "bad-good", however, it exists in Fig. 2.2 because, due to the director, the student with profile "medium-medium" is classified better than the student with profile "bad-good" (class "good" vs. class "bad"). For the same reason, in Fig. 2.2 there is an arc from profile "good-bad" to profile "bad-good", while this arc does not exist in Fig. 2.1.

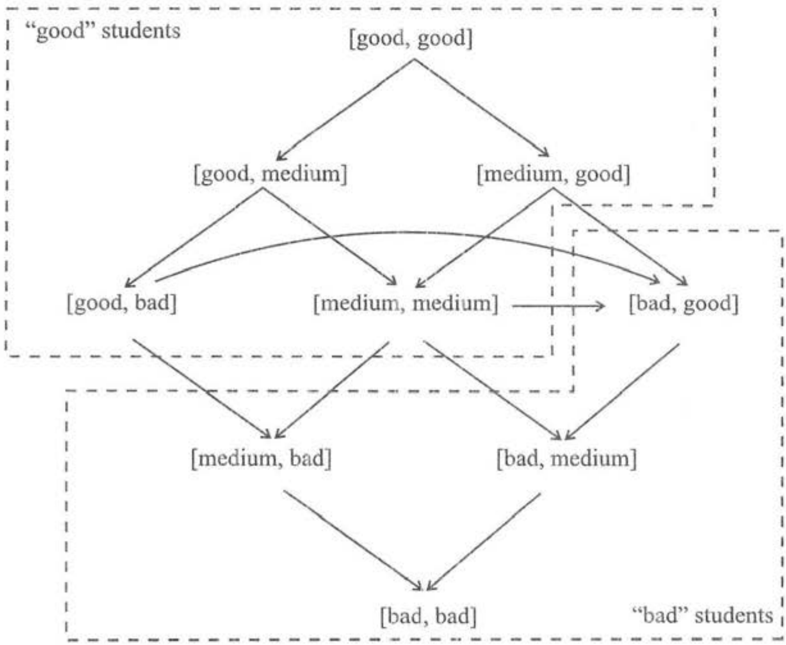

Figure 2.2. The binary relation in the set of student profiles

On the basis of relation R one can build the function $f[g_1(x_1), g_2(x_2)]$, used in proposition 2) of Theorem 2.1, in a very simple way: for each profile $x = [x_1, x_2]$ represented by a node α, $f[g_1(x_1), g_2(x_2)]$ is equal to the number of nodes to which there is a directed path starting from α. The function $f[g_1(x_1), g_2(x_2)]$ is presented in Fig. 2.3. It has the property that all the students classified as "good" have a greater value of this function than the students classified as "bad". Among the "good" students, the minimum value of function $f[g_1(x_1), g_2(x_2)]$, equal to 4, is obtained by students with profiles "good-bad" or "medium-medium". Therefore, the only threshold z_2 is equal to 4; $f[g_1(x_1), g_2(x_2)] \geq 4$ if and only if x is "good", and $f[g_1(x_1), g_2(x_2)] < 4$ if

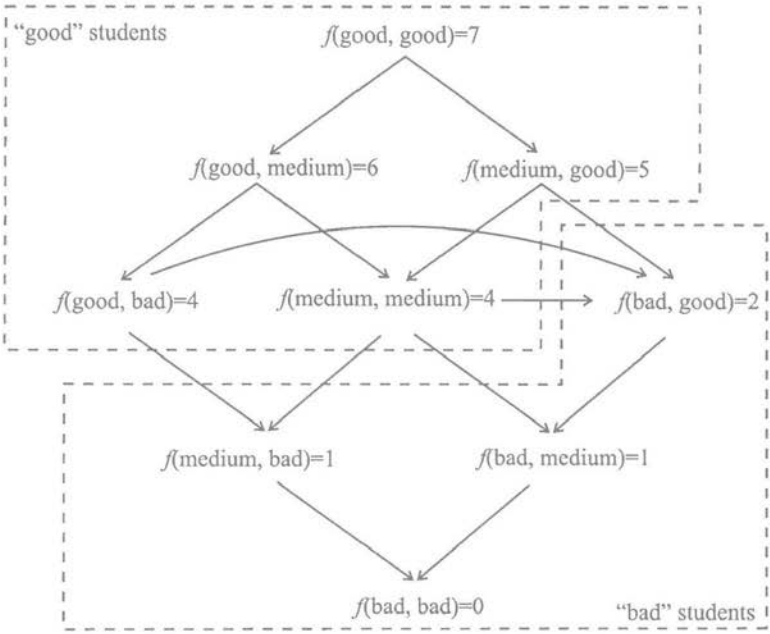

Figure 2.3. The function $f[g(x), g(x)]$ for student profiles $x = [x, x]$

An outranking function satisfying conditions present in proposition 3) of Theorem 2.1 can be built as follows: for each $x, y \in X$ set

$$s[g_1(x_1), g_1(y1), g_2(x_2), g_2(y2)] = f[g_1(x_1), g_2(x_2)] - f[g_1(y1), g_2(y2)].$$

The only reference profile $p^2 = (p_1^2, p_2^2) \in X$ defined in Theorem 2.1 can be chosen such that:

$$f[g_1(p_1^2), g_2(p_2^2)] = \min_{x \in \text{Class "good"}} \{f[g_1(x_1), g_2(x_2)]\}.$$

It is easy to see that the reference profile p^2 is again one of the profiles "good-bad" or "medium-medium".

The *"if..., then..."* decision rules specified in proposition 4) of Theorem 2.1 can be easily built on the basis of the Hasse diagram presented in Fig. 2.1. Let us observe that among the profiles classified as "good", there are two profiles that do not dominate any other profile from this class. These are the profiles "good-bad" and "medium-medium". Starting from these profiles we can induce the two following "at least" decision rules representing the classification of the director:

#1) "*if* Mathematics $\geq$ good *and Literature* $\geq$ *bad*, then student $\geq$ good";
#2) "*if* Mathematics $\geq$ medium *and* Literature $\geq$ medium, *then* student $\geq$ good";
#3) all uncovered students are bad.

Since all the students are at least bad on any criterion, then the rule #1) can be simplified to:

#1*) "*if* Mathematics $\geq$ good, *then* student $\geq$ good".

Let us observe, moreover, that among the profiles classified as "bad", there are two profiles that are not dominated by any other profile from this class. These are the profiles "bad-good" and "medium-bad". Starting from these profiles we can induce the two following "at most" decision rules representing the classification of the director:

#1′) "*if* Mathematics $\leq$ bad *and* Literature $\leq$ good, *then* student $\leq$ bad";
#2′) "*if* Mathematics $\leq$ medium *and* Literature $\leq$ bad, *then* student $\leq$ bad";
#3′) all uncovered students are good.

Since all the students are at most good on any criterion, then the rule #1′) can be simplified to:

#1′*) "*if* Mathematics $\leq$ bad, *then* student $\leq$ bad".

2.3. Ordinal criteria, max-min average and Sugeno integral

Handling ordinal criteria has recently received much attention from researchers considering the multiple-criteria classification. To deal with this problem some max-min aggregation operators have been used, with the most general one – the fuzzy integral proposed by Sugeno (1974). To apply the Sugeno integral, an identical finite ordinal scale $V = \{1, \ldots, m\}$ must be assumed for all criteria, for classes of classification **Cl** and for a fuzzy measure defined on the set of criteria. Let $X = V^n$ denote an evaluation space involving n criteria. Each $x \in X$ is called a *profile*. The scale value of $x \in X$ on criterion g_i is denoted by $g_i(x_i)$ and belongs to V. A *fuzzy measure* on $C = \{g_1, g_2, \ldots, g_n\}$ is a set function $\mu : P(C) \to V$, where $P(C)$ is the power set of C, satisfying the following axioms:

1) $\mu(\emptyset) = 1$, $\mu(C) = m$,
2) $A \subseteq B$ implies $\mu(A) \leq \mu(B)$, for all $A, B \in P(C)$.

For each $x \in X$, the criteria are ordered according to increasing values of $g_i(x_i)$:

$$g_{(1)}, g_{(2)}, \ldots, g_{(n)}, \text{ such that } g_{(1)}(x(1)) \leq g_{(2)}(x(2)) \leq \ldots \leq g_{(n)}(x(n)).$$

The Sugeno integral of $[g_1(x_1), g_2(x_2), \ldots, g_n(x_n)]$ with respect to fuzzy measure μ is defined as follows:

$$\widehat{S}[g_1(x_1), g_2(x_2), \ldots, g_n(x_n); \mu] = \max_{i=1,\ldots,n} \{min\{g_{(i)}(x_{(i)}), \mu(I_{(i)})\}\},$$

where $I_{(i)} = \{g_{(i)}, \ldots, g_{(n)}\}$.

An alternative equivalent definition of the Sugeno integral is the following:

$$\widehat{S}[g_1(x_1), g_2(x_2), \ldots, g_n(x_n); \mu] = \max_{I \subseteq \{g_1, \ldots, g_n\}} \{\min\{g_i(x_i), i \in I, \mu(I)\}\}.$$

The following result is the counterpart of Theorem 2.1 with respect to Sugeno integral, stating equivalence between a simple cancellation property, a utility function represented by a Sugeno integral and a set of specific decision rules. Let $r \in V$ be a grade of scale V and an identifier of class Cl_r corresponding to this grade.

THEOREM 2.4 *The following three propositions are equivalent:*

1) *(cancellation property) for each $i = 1, \ldots, n$, for each $x_i, y_i, z_i \in X_i$ and $a_{-i}, b_{-i} \in X_{-i}$, and for each $r, s \in \{1, \ldots, m\}$:*

$$\{(x_i a_{-i}) \in Cl_r^{\geq} \text{ and } (z_i b_{-i}) \in Cl_s^{\geq} \text{ and } r \geq s\}$$
$$\Rightarrow \{(y_i a_{-i}) \in Cl_r^{\geq} \text{ or } (x_i b_{-i}) \in Cl_s^{\geq}\},$$

2) *(utility function) there exist*
   - *functions $g_i : X_i \to V$ for each $i = 1, \ldots, n$, called criteria,*
   - *a fuzzy measure μ on $C = \{g_1, g_2, \ldots, g_n\}$ having values in V, such that for each $x \in X$ and each $t = 1, \ldots, m$,*

$$\widehat{S}[g_1(x_1), g_2(x_2), \ldots, g_n(x_n); \mu] \geq t \Leftrightarrow x \in Cl_t^{\geq},$$

3) *("at least" decision rules) there exist*
   - *functions $g_i : X_i \to V$ for each $i = 1, \ldots, n$, called criteria,*
   - *a set of "at least" decision rules, called single-graded, whose syntax is*

"if $g_{i1}(xi1) \geq r$ and $g_{i2}(xi2) \geq r$ and ...

$g_{ih}(x_{ih}) \geq r$, then $x \in Cl_r^{\geq}$",

with $\{i1, i2, \ldots, ih\} \subseteq \{1, \ldots, n\}$, $r = 2, \ldots, m$, satisfying the following properties:

(♣) *given the rule: "if $g_{i1}(x_{i1}) \geq r$ and $g_{i2}(x_{i2}) \geq r$ and ... $g_{ih}(x_{ih}) \geq r$, then $x \in Cl_t^{\geq}$", the following rules are also true for each $s < r$:* "if $g_{i1}(x_{i1}) \geq s$ and $g_{i2}(x_{i2}) \geq s$ and ... $g_{ih}(x_{ih}) \geq s$, then $x \in Cl_s^{\geq}$",

(♠) *for each $y \in Cl_r$, $r = 2, \ldots, m$, there is at least one rule implying $y \in Cl_r^{\geq}$" and there is no rule implying $y \in Cl_t^{\geq}$, with $t > r$.*

Comparison of points 2) and 3) of Theorem 2.4 has a positive and a negative interpretation. Positive interpretation says that any preference model expressed in terms of the Sugeno integral can be represented by a set of specific decision rules, i.e. single-graded decision rules satisfying property (♣). Negative interpretation of Theorem 2.4 says that not all preference models represented by a set of decision rules can be represented also in terms of the Sugeno integral. In the next section we present an example of preference model representable by a set of decision rules but not representable by the Sugeno integral. Therefore, the decision rule model has a larger applicability than the Sugeno integral. This seems to be an advantage of the decision rule model compared to the Sugeno integral. In our opinion there is also another advantage of the decision rule model, which is perhaps more important for multiple-criteria decision aiding: the decision rule model expresses the preferences in much more intelligible terms than the Sugeno integral.

2.4. An example

Let us consider an augmented example of evaluation in a high school. Suppose that a high school director wants to assign students to different classes of merits on the basis of their scores in Mathematics, Physics and Literature. The ordinal scales of the evaluation in Mathematics, Physics and Literature, as well as the comprehensive evaluation scale have been composed of three following grades: "bad", "medium", "good".

Table 2.2 presents all possible profiles of the students with respect to the three considered criteria, and a classification decision made by the director. Let us observe that the classification of students presented in Table 2.2 satisfies proposition 1) of Theorem 2.1. In fact, it can be seen in the table that each time a student x dominates a student y, student x belongs to the same or higher class than student y. Furthermore, it is possible to build a utility function $f[g_1(x_1), g_2(x_2), g_3(x_3)]$ where the evaluations of student x in Mathematics, Physics and Literature are denoted by x_1, x_2 and x_3, respectively and functions $g_1(\cdot), g_2(\cdot)$ and $g_3(\cdot)$ respect monotonicity; for example, $g_i(bad) = 1$, $g_i(medium) = 2$, $g_i(good) = 3$, $i = 1,2,3$. The utility function $f[g_1(x_1), g_2(x_2), g_3(x_3)]$ satisfies conditions present in proposition 2) of Theorem 2.1. The values of this function are presented in Table 2.2. The ordered thresholds z_t, $t = 2,3$, defined in Theorem 2.1, are set on the values $z_2 = 20$, $z_3 = 24$, and the classification is performed by checking the following conditions:

$$f[g_1(x_1), g_2(x_2), g_3(x_3)] < 20 \Leftrightarrow \text{“}x \text{ is bad”},$$
$$20 \leq f[g_1(x_1), g_2(x_2), g_3(x_3)] < 24 \Leftrightarrow \text{“}x \text{ is medium”},$$
$$f[g_1(x_1), g_2(x_2), g_3(x_3)] \geq 24 \Leftrightarrow \text{“}x \text{ is good”}.$$

An outranking function, satisfying conditions present in proposition 3) of Theorem 2.1 can be built as follows: for each $x, y \in X$, set

$$s[g_1(x_1), g_1(y1), g_2(x_2), g_2(y2), g_3(x_3), g_3(y3)]$$
$$= f[g_1(x_1), g_2(x_2), g_3(x_3)] - f[g_1(y1), g_2(y2), g_3(y3)].$$

The reference profiles $p^t \in X$ defined in Theorem 2.1 can be chosen as follows:

$$f[g_1(p_1^t), g_2(p_2^t), g_3(p_3^t)] = \min_{x \in Cl_t} \{f[g_1(x_1), g_2(x_2), g_3(x_3)]\}, \text{ for } t=2,\ldots,m.$$

The resulting reference profiles are: p^2 = {good, medium, bad} or {medium, good, bad} or {medium, medium, medium}; p^3 = {good, medium, medium} or {medium, good, medium}.

Table 2.2. Classification of all 27 cases of possible evaluations ($f(x) = f[g_1(x_1), g_2(x_2), g_3(x_3)]$)

Student	Mathematics	Physics	Literature	Decision	$f(x)$	Matching rules
S1	bad	bad	bad	bad	1	#6
S2	medium	bad	bad	bad	10	#6
S3	good	bad	bad	bad	14	#6
S4	bad	medium	bad	bad	10	#6
S5	medium	medium	bad	bad	16	#6
S6	good	medium	bad	medium	20	#2
S7	bad	good	bad	bad	14	#6
S8	medium	good	bad	medium	20	#3
S9	good	good	bad	medium	21	#2,3
S10	bad	bad	medium	bad	7	#6
S11	medium	bad	medium	bad	13	#6
S12	good	bad	medium	bad	15	#6
S13	bad	medium	medium	bad	13	#6
S14	medium	medium	medium	medium	20	#1
S15	good	medium	medium	good	24	#1,2,4
S16	bad	good	medium	bad	15	#6
S17	medium	good	medium	good	24	#1,3,5
S18	good	good	medium	good	26	#1,2,3,4,5
S19	bad	bad	good	bad	12	#6
S20	medium	bad	good	bad	15	#6
S21	good	bad	good	bad	16	#6
S22	bad	medium	good	bad	15	#6
S23	medium	medium	good	medium	21	#1
S24	good	medium	good	good	26	#1,2,4
S25	bad	good	good	bad	16	#6
S26	medium	good	good	good	26	#1,3,5
S27	good	good	good	good	27	#1,2,3,4,5

Finally, according to proposition 4) of Theorem 2.1, the classification of Table 2.2 can be represented by means of the following set of "at least" decision rules:

#1) "*if* Mathematics $\geq$ medium *and* Physics $\geq$ medium *and* Literature $\geq$ medium, *then* student $\geq$ medium";

#2) "*if* Mathematics $\geq$ good *and* Physics $\geq$ medium, *then* student $\geq$ medium";

#3) "*if* Mathematics $\geq$ medium *and* Physics $\geq$ good, *then* student $\geq$ medium";

#4) "*if* Mathematics $\geq$ good *and* Physics $\geq$ medium *and* Literature $\geq$ medium, *then* student $\geq$ good";

#5) "*if* Mathematics $\geq$ medium *and* Physics $\geq$ good *and* Literature $\geq$ medium, *then* student $\geq$ good";

#6) all uncovered students are bad.

The numbers of rules matching a student's profile are indicated in the column "Decision" of Table 2.2. Let us recall that when more than one "at least" rule is matching a student, he/she is assigned to the highest "at least" class indicated by the matching rules.

The classification presented in Table 2.2 can also be represented by a set of "at most" decision rules:

#1′) "*if* Mathematics $\leq$ bad, *then* student $\leq$ bad";

#2′) "*if* Physics $\leq$ bad, *then* student $\leq$ bad";

#3′) "*if* Mathematics $\leq$ medium *and* Physics $\leq$ medium *and* Literature $\leq$ bad, *then* student $\leq$ bad";

#4′) "*if* Literature $\leq$ bad, *then* student $\leq$ medium";

#5′) "*if* Mathematics$\leq$medium *and* Physics$\leq$medium, *then* student$\leq$medium";

#6′) all uncovered students are good.

Let us observe that all the 27 classification decisions made by "at least" or "at most" decision rules in Table 2.2 cannot be represented by the most general max-min aggregation operator permitting ordinal aggregation, i.e. the fuzzy integral proposed by Sugeno (1974). Why? This can be understood intuitively from Theorem 2.4: in fact, many of the rules applied for the classification of the 27 cases are not single-graded, i.e. they use more than one grade of the evaluation scale in conditions and decision.

The answer can also be more direct: consider the decision rule #2):

"*if* Mathematics $\geq$ good *and* Physics $\geq$ medium, *then* student $\geq$ medium".

There are the following possible values of fuzzy measure μ permitting to obtain with Sugeno integral the same classification as with rule #2), without misclassification:

1) either $\mu(\{\text{Mathematics, Physics}\})$ = medium,

2) or $\mu(\{\text{Mathematics}\})$ = good,

3) or $\mu(\{\text{Physics}\})$ = medium.

Case 1) corresponds to the rule:

"*if* Mathematics $\geq$ medium *and* Physics $\geq$ medium, *then* student medium", but it has the condition part weaker than rule #2);

case 2) corresponds to the rule:

"*if* Mathematics $\geq$ good, *then* student $\geq$ medium", but it has the condition part weaker than rule #2);

case 3) corresponds to the rule:

"*if* Physics $\geq$ medium, *then* student $\geq$ medium", but it has the condition part weaker than rule #2).

In conclusion, there is no possibility of representing the classification made

3. Rough set approach to multiple-criteria classification problems

This section summarizes main steps of the methodology presented in Greco, Matarazzo and Słowiński (1999, 2001a, 2002a).

Let $\succeq_i$ be a weak preference relation on $U \subseteq X$ with reference to criterion $g_i \in C = \{g_1, g_2, \ldots, g_n\}$, such that $x \succeq_i y$ means "x is at least as good as y with respect to criterion g_i". Suppose that $\succeq_i$ is a complete preorder, i.e. a strongly complete and transitive binary relation.

It is said that x *dominates* y with respect to $P \subseteq C$ (denoted by xD_Py) if $x \succeq_i y$ for each $g_i \in P$. Since the intersection of complete preorders is a partial preorder and $\succeq_i$ is a complete preorder for each $g_i \in P$, and $D_P = \bigcap_{g_i \in P} \succeq_i$, then the dominance relation D_P is a partial preorder. Given $P \subseteq C$ and $x \in U$, let

$$D_P^+(x) = \{y \in U : yD_Px\},$$
$$D_P^-(x) = \{y \in U : xD_Py\}.$$

We define the *P-lower approximation* and the *P-upper approximation* of $Cl_t^{\geq}$, $t \in T$, with respect to $P \subseteq C$ (denoted by $\underline{P}(Cl_t^{\geq})$ and $\overline{P}(Cl_t^{\geq})$, respectively), as:

$$\underline{P}(Cl_t^{\geq}) = \{x \in U : D_P^+ \subseteq Cl_t^{\geq}\},$$
$$\overline{P}(Cl_t^{\geq}) = \bigcup_{x \in Cl_t^{\geq}} D_P^+(x).$$

Analogously, we define the P-lower approximation and the P-upper approximation of $Cl_t^{\geq}$, $t \in T$, with respect to $P \subseteq C$ (denoted by $\underline{P}(Cl_t^{\geq})$ and $\overline{P}(Cl_t^{\geq})$, respectively), as:

$$\underline{P}(Cl_t^{\geq}) = \{x \in U : D_P^- \subseteq Cl_t^{\leq}\},$$
$$\overline{P}(Cl_t^{\leq}) = \bigcup_{x \in Cl_t^{\leq}} D_P^-(x).$$

The pairs $\underline{P}(Cl_t^{\geq})$, $\overline{P}(Cl_t^{\geq})$, and $\underline{P}(Cl_t^{\leq})$, $\overline{P}(Cl_t^{\leq})$ are called *rough approximations* of $Cl_t^{\geq}$ and $Cl_t^{\leq}$, respectively. Intuitively, $\underline{P}(Cl_t^{\geq})$ represents the set of actions that, according to the information given by criteria of P, can be assigned to $Cl_t^{\geq}$ with certainty. Analogously, $\underline{P}(Cl_t^{\leq})$ represents the set of actions that, according to the information given by criteria of P, can be assigned to $Cl_t^{\leq}$ with certainty. Instead, $\overline{P}(Cl_t^{\geq})$ and $\overline{P}(Cl_t^{\leq})$ represent the sets of actions which, according to the information given by criteria of P, could be assigned to $Cl_t^{\geq}$ and $Cl_t^{\leq}$, respectively, however, there are some doubts due to inconsistency of the available information.

The P-lower and P-upper approximations so obtained satisfy the following properties for each $t \in T$ and for each $P \subseteq C$:

$$\underline{P}(Cl_t^{\geq}) \subseteq Cl_t^{\geq} \subseteq \overline{P}(Cl_t^{\geq}), \quad \underline{P}(Cl_t^{\leq}) \subseteq Cl_t^{\leq} \subseteq \overline{P}(Cl_t^{\leq}).$$

Furthermore, the following specific complementarity properties hold:

$$\begin{aligned}
&\underline{P}(Cl_t^{\geq}) = U - \overline{P}(Cl_{t-1}^{\leq}),\ t = 2,\dots,m,\\
&\underline{P}(Cl_t^{\leq}) = U - \overline{P}(Cl_{t+1}^{\geq}),\ t = 1,\dots,m-1,\\
&\overline{P}(Cl_t^{\geq}) = U - \underline{P}(Cl_{t-1}^{\leq}),\ t = 2,\dots,m,\\
&\overline{P}(Cl_t^{\leq}) = U - \underline{P}(Cl_{t+1}^{\geq}),\ t = 1,\dots,m-1.
\end{aligned}$$

The P-boundaries (P-doubtful regions) of $Cl_t^{\geq}$ and $Cl_t^{\leq}$ are defined as

$$Bn_P(Cl_t^{\geq}) = \overline{P}(Cl_t^{\geq}) - \underline{P}(Cl_t^{\geq}),$$

$$Bn_P(Cl_t^{\leq}) = \overline{P}(Cl_t^{\leq}) - \underline{P}(Cl_t^{\leq}), \quad \text{for each } t \in T.$$

$Bn_P(Cl_t^{\geq})$ and $Bn_P(Cl_t^{\leq})$ represent the set of all the actions which, according to the information given by criteria from P, are assigned to $Cl_t^{\geq}$ and $Cl_t^{\leq}$, respectively, in a way being inconsistent with the dominance principle. The ability of handling inconsistency in classification data is a key feature of the rough set approach (Słowiński et al., 2000).

We define the *accuracy of approximation* of $Cl_t^{\geq}$ and $Cl_t^{\leq}$, for each $t \in T$ and for each $P \subseteq C$, respectively, as:

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

The ratio

$$\gamma_P(\mathbf{Cl}) = \frac{|U - ((\bigcup_{t\in T} Bn_P(Cl_t^{\geq})) \cup (\bigcup_{t\in T} Bn_P(Cl_t^{\leq})))|}{|U|}$$

defines the *quality of approximation of the partition* **Cl** by means of the set of criteria P, or, briefly, *quality of classification*. It expresses the ratio of all P-correctly classified actions to all the actions in the set U. Let us stress that this quality measure does not concern the "prediction quality" of the set of criteria P, but its capacity of non-ambiguous reclassification of the actions from the set U.

Every minimal subset $P \subseteq C$ such that $\gamma_P(\mathbf{Cl}) = \gamma_C(\mathbf{Cl})$ is called a reduct of C with respect to **Cl** and is denoted by $RED_{\mathbf{Cl}}(C)$. For a given classification there may exist more than one reduct. The intersection of all the reducts is known as the core, denoted by $CORE_{\mathbf{Cl}}$.

Let us observe that we can write the definitions of P-lower approximations of $Cl_t^{\geq}$ and $Cl_t^{\leq}$ alternatively, as follows:

$$\begin{aligned}
&\underline{P}(Cl_t^{\geq}) = \{x \in U : yD_Px \Rightarrow y \in Cl_t^{\geq}\},\\
&\underline{P}(Cl_t^{\leq}) = \{x \in U : xD_Py \Rightarrow y \in Cl_t^{\leq}\}.
\end{aligned}$$

On the basis of this observation, it is possible to induce from the above approximations a generalized description of the preferential information contained in the given set of classification examples, in terms of "at least" and "at most" decision rules (Greco, Matarazzo and Słowiński, 2002c).

4. Conjoint measurement for multiple-criteria classification problems with inconsistencies

4.1. Representation theorems for multiple-criteria classification with inconsistencies

The conjoint measurement model presented in Section 2 cannot represent the inconsistency with the dominance principle considered within the rough set approach. In this section we present a more general model of conjoint measurement that permits representation of this inconsistency. This model is based on the concepts of rough approximation of upward and downward unions of classes $Cl_t^{\geq}$ and $Cl_t^{\leq}$.

The following concepts will be useful: for each $x \in X$, the *lower class* and the *upper class* of x, denoted by $r_*(x)$ and $r^*(x)$, respectively, are defined as follows

$$r_*(x) = \max\{s \in T : x \in \underline{C}(Cl_t^{\geq})\},$$
$$r^*(x) = \min\{s \in T : x \in \underline{C}(Cl_t^{\leq})\},$$

where $\underline{C}(Cl_t^{\geq})$ and $\underline{C}(Cl_t^{\leq})$ are C-lower approximations of $Cl_t^{\geq}$ and $Cl_t^{\leq}$, respectively.

THEOREM 4.1 *For each set of binary relations $\succeq_i$, $i = 1, \ldots, n$, being complete preorders, and for each classification* **Cl** *there exist*

- *functions $g_i : X_i \to \mathbf{R}$, such that $x_i \succeq_i y_i \Leftrightarrow g_i(x_i) \geq g_i(y_i)$, $i = 1, \ldots, n$,*
- *functions $f^{\geq} : \mathbf{R}^n \to \mathbf{R}$ and $f^{\leq} : \mathbf{R}^n \to \mathbf{R}$, non-decreasing in each argument, such that*

  $$f^{\geq}[g_1(x_1), g_2(x_2), \ldots, g_n(x_n)] \leq f^{\leq}[g_1(x_1), g_2(x_2), \ldots, g_n(x_n)]$$
- *$m-1$ ordered thresholds z_t, $t = 2, \ldots, m$,*

  $$z_2 < z_3 < \ldots < z_m$$

  such that for each action $x \in X$, functions $f^{\geq}$ and $f^{\leq}$ assign x to a lower and an upper class, respectively:

  $$f^{\geq}[g_1(x_1), g_2(x_2), \ldots, g_n(x_n)] \geq z_t \Leftrightarrow x \in \underline{C}(Cl_t^{\geq}),$$
  $$f^{\leq}[g_1(x_1), g_2(x_2), \ldots, g_n(x_n)] < z_t \Leftrightarrow x \in \underline{C}(Cl_{t-1}^{\leq}),$$

  where $C = \{g_1, g_2, \ldots, g_n\}$.

Inconsistency with the dominance principle can also be represented in terms of a set of "at least" and "at most" decision rules considered together. More formally, a set of "at least" and "at most" decision rules does not contradict the classification **Cl** if for each $x \in Cl_t$ there exists no "at least" decision rule for which $x \in Cl_s^{\geq}$, with $s > t$, and there exists no "at most" decision rule for which $x \in Cl_s^{\geq}$, with $s < t$. A set of decision rules is complete if for each $x \in \underline{C}(Cl_t^{\geq})$ there exists a decision rule for which $x \in Cl_s^{\geq}$, with $s \geq t$, and for each $x \in \underline{C}(Cl_t^{\leq})$ there exists a decision rule for which $x \in Cl_s^{\geq}$, with $s \leq t$. A set of decision rules represents the classification **Cl** if it does not contradict **Cl** and it is complete.

THEOREM 4.2 *For each set of binary relations $\succeq_i$, $i = 1, \ldots, n$, being complete preorders, and for each classification* **Cl**, *there exists a set of decision rules representing the classification Cl.*

4.2. Interval classification and its representation

As stated above, one of the main reasons of inconsistency in the sense of dominance is the hesitation of the DM assigning some action $x \in X$ to a class of **Cl**. It is thus worth considering a more relaxed definition of the classification problem, which we call *interval classification*. In the interval classification, **Cl** is not always a partition of X, and Cl_t, $t = 1, \ldots, m$, are preference-ordered and non-necessarily disjoint subsets of X, such that the evaluation corresponding to Cl_r is better than the evaluation corresponding to Cl_s when $r > s$. More precisely, in interval classification for each $x \in X$ we have $x \cup Cl_r \cup Cl_{r+1} \cup \ldots \cup Cl_s$, where $r, s \in \{1, \ldots, m\}$, $r \leq s$. Of course, if $r = s$ for each $x \in X$, the interval classification boils down to the usual classification. For the interval classification, the membership of x to the upward and downward unions $Cl_t^{\geq}$ and $Cl_t^{\leq}$ is defined as follows:

$$x \in Cl_r \cup Cl_{r+1} \cup \ldots \cup Cl_s \Leftrightarrow \begin{cases} x \in Cl_a^{\geq} & \text{for any } a \leq r, \\ x \in Cl_b^{\leq} & \text{for any } b \geq s. \end{cases}$$

For example, if in a school a student x is classified as "between medium and good", we have $x \in Cl_{medium} \cup Cl_{good}$, which is equivalent to: $x \in Cl_{medium}^{\geq}$ and $x \in Cl_{good}^{\leq}$. Each classification with some inconsistency in the sense of dominance can be represented in terms of interval classification by stating for each $x \in X$: $x \cup Cl_t \cup Cl_{t+1} \cup \ldots \cup Cl_u$, where $t = r_*(x)$ and $u = r^*(x)$.

Let us introduce some cancellation conditions having the same nature as proposition 1) in Theorem 2.1:

Condition C1): for each $i = 1, \ldots, n$, for each $x_i, y_i \in X_i$ and $a_{-i}, b_{-i} \in X_{-i}$

$$\{(x_i a_{-i}) \in Cl_r^{\geq} \text{ and } (y_i b_{-i}) \in Cl_s^{\geq}\} \Rightarrow \{(y_i a_{-i}) \in Cl_r^{\geq} \text{ or } (x_i b_{-i}) \in Cl_s^{\geq}\},$$

Condition C2): for each $i = 1, \ldots, n$, for each $x_i, y_i \in X_i$ and $a_{-i}, b_{-i} \in X_{-i}$

$$\{(x_i a_{-i}) \in Cl_r^{\leq} \text{ and } (y_i b_{-i}) \in Cl_s^{\leq}\} \Rightarrow \{(y_i a_{-i}) \in Cl_r^{\leq} \text{ or } (x_i b_{-i}) \in Cl_s^{\geq}\},$$

Condition C3): for each $i = 1, \ldots, n$, for each $x_i, y_i \in X_i$ and $a_{-i}, b_{-i} \in X_{-i}$

$$\{(x_i a_{-i}) \in Cl_r^{\geq} \text{ and } (x_i b_{-i}) \in Cl_s^{\leq}\} \Rightarrow \{(y_i a_{-i}) \in Cl_r^{\geq} \text{ or } (y_i b_{-i}) \in Cl_s^{\leq}\}.$$

Conditions C1)–C3) can be interpreted as follows:

– Condition C1) says that the upward unions of ordered classes induce a represented by the binary relation $\succeq_i^+$ defined on X_i as follows:

$x_i \succeq_i^+ y_i \Leftrightarrow [(y_i a_{-i}) \in Cl_t^{\geq} \Rightarrow (x_i a_{-i}) \in Cl_t^{\geq}$, for each $a_{-i} \in X_{-i}$ and for each $t = 1, \ldots, m]$.

On the basis of condition C1), one can prove that $\succeq_i^+$ is a complete preorder. This is formally stated by Theorem 4.3.

- Condition C2) says that the downward unions of ordered classes induce a preference order on the domain of each criterion. This preference order is represented by the binary relation $\succeq_i^-$ defined on X_i as follows:

  $x_i \succeq_i^- y_i \Leftrightarrow [(x_i a_{-i}) \in Cl_t^{\leq} \Rightarrow (y_i a_{-i}) \in Cl_t^{\leq}$, for each $a_{-i} \in X_{-i}$ and for each $t = 1, \ldots, m]$.

  On the basis of condition C2), one can prove that $\succeq_i^-$ is a complete preorder. This is formally stated by Theorem 4.4.
- Condition C3) says that the two previous preference orders are not contradictory in the sense that there cannot exist $x_i, y_i \in X_i$ such that $x_i \succeq_i^+ y_i$ and *not* $y_i \succeq_i^+ x_i$ (i.e. x_i is preferred to y_i according to $\succeq_i^+$), and $y_i \succeq_i^- x_i$ and *not* $x_i \succeq_i^- y_i$ (i.e. y_i is preferred to x_i according to $\succeq_i^-$). This is formally stated by Theorem 4.5.

Therefore, conditions C1)–C3) are useful to state the following representation theorems for interval classification.

THEOREM 4.3 *The following three propositions are equivalent:*

1) (cancellation property) condition C1) holds,

2) (utility function) there exist

- *functions* $g_i^{\geq} : X_i \to \mathbf{R}$ *for each* $i = 1, \ldots, n$,
- *function* $f^{\geq} : \mathbf{R}^n \to \mathbf{R}$, *increasing in each argument, called upward classification function,*
- $m - 1$ *ordered thresholds* z_t, $t = 2, \ldots, m$, *satisfying*

  $$z_2^{\geq} < z_3^{\geq} < \ldots < z_m^{\geq}$$

  such that for each $x \in X$

  $$f^{\geq}[g_1^{\geq}(x_1), g_1^{\geq}(x_2), \ldots, g_n^{\geq}(x_n)] \geq z_t^{\geq} \Leftrightarrow x \in Cl_t^{\geq}.$$

3) ("at least" decision rules) there exist

- *functions* $g_1^{\geq} : X_i \to \mathbf{R}$ *for each* $i = 1, \ldots, n$, *called criteria,*
- *a set of "at least" decision rules whose syntax is*

  "if $g_{i1}^{\geq}(x_{i1}) \geq r_{i1}$ *and* $g_{i2}^{\geq}(x_{i2}) \geq r_{i2}$ *and* $\ldots$

  and $g_{ih}^{\geq}(x_{ih}) \geq r_{ih}$, *then* $x \in Cl_r^{\geq}$*",*

  with $\{i1, i2, \ldots, ih\} \subseteq \{1, \ldots, n\}$, $r = 2, \ldots, m$, *such that for each* $y \in Cl_r$, $r = 2, \ldots, m$, *there is at least one rule implying* $y \in Cl_r^{\geq}$ *and there is no rule implying* $y \in Cl_t^{\geq}$, *with* $t > r$.

THEOREM 4.4 *The following three propositions are equivalent:*

1) (cancellation property) condition C2) holds,

2) (utility function) there exist

- *functions* $g_i^{\leq} : X_i \to \mathbf{R}$ *for each* $i = 1, \ldots, n$,
- *function* $f^{\leq} : \mathbf{R}^n \to \mathbf{R}$, *increasing in each argument, called downward classification function,*
- $m-1$ *ordered thresholds* $z_t^{\leq}$, $t = 1, \ldots, m-1$, *satisfying*

  $$z_1^{\leq} < z_2^{\leq} < \ldots < z_{m-1}^{\leq}$$

  such that for each $x \in X$

  $$f^{\leq}[g_1^{\leq}(x_1), g_2^{\leq}(x_2), \ldots, g_n^{\leq}(x_n)] \leq z_t^{\leq} \Leftrightarrow x \in Cl_t^{\leq}.$$

3) ("at most" decision rules) there exist

- *functions* $g_i^{\leq} : X_i \to \mathbf{R}$ *for each* $i = 1, \ldots, n$,
- *a set of decision rules whose syntax is*

  "*if* $g_{i1}^{\leq}(x_{i1}) \leq r_{i1}$ *and* $g_{i2}^{\leq}(x_{i2}) \leq r_{i2}$ *and* …

  and $g_{ih}^{\leq}(x_{ih}) \leq r_{ih}$, *then* $x \in Cl_r^{\leq}$",

  with $\{i1, i2, \ldots, ih\} \subseteq \{1, \ldots, n\}$, $r = 1, \ldots, m-1$, *such that for each* $y \in Cl_r$, $r = 1, \ldots, m-1$, *there is at least one rule implying* $y \in Cl_r^{\leq}$ *and there is no rule implying* $y \in Cl_t^{\leq}$, *with* $t < r$.

THEOREM 4.5 *The following three propositions are equivalent:*

1) (cancellation properties) conditions C1), C2), C3) hold,

2) (utility functions) there exist

- *functions* $g_i : X_i \to \mathbf{R}$ *for each* $i = 1, \ldots, n$,
- *two functions* $f^{\geq} : \mathbf{R}^n \to \mathbf{R}$ *and* $f^{\leq} : \mathbf{R}^n \to \mathbf{R}$, *non-decreasing in each argument, called* upward and downward classification functions, *respectively, such that for each* $x \in X$

  $$f^{\leq}[g_1(x_1), g_2(x_2), \ldots, g_n(x_n)] \geq f^{\geq}[g_1(x_1), g_2(x_2), \ldots, g_n(x_n)],$$

- $m-1$ *ordered thresholds* z_t, $t = 2, \ldots, m$, *satisfying*

  $$z_2 < z_3 < \ldots < z_m$$

  such that for each action $x \in X$

  $$f^{\geq}[g_1(x_1), g_2(x_2), \ldots, g_n(x_n)] \geq z_t \Leftrightarrow x \in Cl_t^{\geq},$$

  $$f^{\leq}[g_1(x_1), g_2(x_2), \ldots, g_n(x_n)] < z_t \Leftrightarrow x \in Cl_{t-1}^{\leq},$$

3) ("at least" and "at most" decision rules) there exist

- *functions* $g_i : X_i \to \mathbf{R}$ *for each* $i = 1, \ldots, n$,

- *a set of "at least" decision rules whose syntax is*

  "if $g_{i1}(x_{i1}) \geq r_{i1}$ and $g_{i2}(x_{i2}) \geq r_{i2}$ and ...

  and $g_{ih}(x_{ih}) \geq r_{ih}$, then $x \in Clr^{\geq}$",

  with $\{i1, i2, \ldots, ih\} \subseteq \{1, \ldots, n\}$, $r = 2, \ldots, m$, such that for each $y \in Cl_r$, $r = 2, \ldots, m$, there is at least one rule implying $y \in Cl_r^{\geq}$ and there is no rule implying $y \in Cl_t^{\geq}$, with $t > r$,

- *a set of "at most" decision rules whose syntax is*

  "if $g_{i1}(x_{i1}) \leq r_{i1}$ and $g_{i2}(x_{i2}) \leq r_{i2}$ and ...

  and $g_{ih}(x_{ih}) \leq r_{ih}$, then $x \in Cl_r^{\leq}$",

  with $\{i1, i2, \ldots, ih\} \subseteq \{1, \ldots, n\}$, $r = 1, \ldots, m-1$, such that for each $y \in Cl_r$, $r = 1, \ldots, m-1$, there is at least one rule implying $y \in Cl_r^{\leq}$ and there is no rule implying $y \in Cl_t^{\leq}$, with $t < r$.

The following Theorem 4.6 gives conditions for utility functions $f^{\leq}[g_1(x_1), g_2(x_2), \ldots, g_n(x_n)]$ and $f^{\geq}[g_1(x_1), g_2(x_2), \ldots, g_n(x_n)]$ of Theorem 4.5 being Sugeno integrals. Theorem 4.6 gives also a syntax of the corresponding "at least" and "at most" decision rules.

THEOREM 4.6 *The following three propositions are equivalent:*

1) (cancellation properties) condition C3) and the following conditions C1′), C2′) hold:

Condition C1'): *for each $i = 1, \ldots, n$, for each $x_i, y_i, z_i \in X_i$ and $a_{-i}, b_{-i} \in X_{-i}$, $t = 1, 2, \ldots, m$*

$$[(x_i a_{-i}) \in Cl_t^{\geq}] \text{ and } (y_i b_{-i}) \in Cl_t^{\geq}]$$
$$\Rightarrow [(z_i a_{-i}) \in Cl_t^{\geq}] \text{ or } (x_i b_{-i}) \in Cl_t^{\geq}],$$

Condition C2'): *for each $i = 1, \ldots, n$, for each $x_i, y_i, z_i \in X_i$ and $a_{-i}, b_{-i} \in X_{-i}$, $t = 1, 2, \ldots, m$*

$$[(x_i a_{-i}) \in Cl_t^{\leq}] \text{ and } (y_i b_{-i}) \in Cl_t^{\leq}]$$
$$\Rightarrow [(z_i a_{-i}) \in Cl_t^{\leq}] \text{ or } (x_i b_{-i}) \in Cl_t^{\leq}],$$

2) (utility functions being Sugeno integrals) there exist

- *functions $g_i : X_i \to V$, $V = \{1, \ldots, m\}$, for each $i = 1, \ldots, n$,*
- *two fuzzy measures μ^1 and μ^2 on $C = \{g_1, g_2, \ldots, g_n\}$, having values in V and satisfying for each $A \subseteq \{g_1, g_2, \ldots, g_n\}$ the inequality $\mu^2(A) \geq \mu^1(A)$, such that for each action $x \in X$*

$$\widehat{S}[g(x_1), g(x_2), \ldots, g(x_n); \mu^1] \geq t \Leftrightarrow x \in Cl_t^{\geq},\ t = 1, \ldots, m,$$
$$\widehat{S}[g(x_1), g(x_2), \ldots, g(x_n); \mu^2] \leq t \Leftrightarrow x \in Cl_t^{\leq},\ t = 1, \ldots, m.$$

3) ("at least" and "at most" decision rules) there exist

- *functions $g_i : X_i \to V$, $V = \{1, \ldots, m\}$, for each $i = 1, \ldots, n$.*

- *a set of decision rules whose syntax is*

  $$\text{“if } g(x_{i1}) \geq r \text{ and } \dots \text{ and } g(x_{ih}) \geq r, \text{ then } x \in Cl_r \geq\text{”},$$

  with $\{i1, i2, \dots, ih\} \subseteq \{1, \dots, n\}$, $r = 2, \dots, m$, *satisfying the following properties:*

  (♣) *given the rule: “if* $g_{i1}(x_{i1}) \geq r$ *and* $g_{i2}(x_{i2}) \geq r$ *and* $\dots$ $g_{ih}(x_{ih}) \geq r$, *then* $x \in Cl_s^{\geq}$*”, the following rules are also true for each* $s < r$*: “if* $g_{i1}(x_{i1}) \geq s$ *and* $g_{i2}(x_{i2}) \geq s$ *and* $\dots$ $g_{ih}(x_{ih}) \geq s$, *then* $x \in Cl_s^{\geq}$*”,*

  (♠) *for each* $y \in Cl_r$, $r = 2, \dots, m$, *there is at least one rule implying* $y \in Cl_r^{\geq}$ *and there is no rule implying* $y \in Cl_t^{\geq}$, *with* $t > r$,

- *a set of decision rules whose syntax is*

  $$\text{“if } g_{i1}(x_{i1}) \leq r \text{ and } \dots \text{ and } g_{ih}(x_{ih}) \leq r, \text{ then } x \in Cl_r^{\leq}\text{”},$$

  with $\{i1, i2, \dots, ih\} \subseteq \{1, \dots, n\}$, $r = 1, \dots, m-1$, *satisfying the following properties:*

  (♣) *given the rule: “if* $g_{i1}(x_{i1}) \leq r$ *and* $g_{i2}(x_{i2}) \leq r$ *and* $\dots$ $g_{ih}(x_{ih}) \leq r$, *then* $x \in Cl_r^{\leq}$*”, the following rules are also true for each* $s > r$*: “if* $g_{i1}(x_{i1}) \leq s$ *and* $g_{i2}(x_{i2}) \leq s$ *and* $\dots$ $g_{ih}(x_{ih}) \leq s$, *then* $x \in Cl_s^{\leq}$*”,*

  (♠) *for each* $y \in Cl_r$, $r = 1, \dots, m-1$, *there is at least one rule implying* $y \in Cl_r^{\leq}$ *and there is no rule implying* $y \in Cl_t^{\leq}$, *with* $t < r$.

4.3. An example

Let us consider a new version of the example presented in Section 2.4. As before, the director of the school wants to assign students to different classes of merits on the basis of their scores in Mathematics, Physics and Literature, but now in his classification decisions there is some inconsistency in the sense of dominance. All the 27 possible cases are presented in Table 4.1. As can be seen in the table, several pairs of students are inconsistent in the sense of dominance. For example, student S3 is dominated by student S6 (in Mathematics S3 and S6 are both good, in Physics S3 is bad and S6 is medium, in Literature S3 and S6 are both bad). The pairs of students inconsistent with the dominance principle are: (S3,S6), (S3,S15), (S12,S15), (S7,S8), (S7,S17), (S16,S17).

The above cases of inconsistency represent situations of hesitation in assigning a student to a given class of merit. Therefore, we can represent the inconsistent classification of Table 4.1 in terms of an equivalent interval classification as shown in Table 4.2. In Table 4.2, the values assigned by functions $f^{\geq}[g_1(x_1), g_2(x_2), g_3(x_3)]$ and $f^{\leq}[g_1(x_1), g_2(x_2), g_3(x_3)]$, respectively are presented. Functions $f^{\geq}[g_1(x_1), g_2(x_2), g_3(x_3)]$ and $f^{\leq}[g_1(x_1), g_2(x_2), g_3(x_3)]$ were built as explained in the comment to Theorem 4.1. The ordered thresholds

Table 4.1. An inconsistent classification of the cases of possible evaluations

Student	Mathematics	Physics	Literature	Decision
S1	bad	bad	bad	bad
S2	medium	bad	bad	bad
S3	good	bad	bad	good
S4	bad	medium	bad	bad
S5	medium	medium	bad	medium
S6	good	medium	bad	medium
S7	bad	good	bad	good
S8	medium	good	bad	medium
S9	good	good	bad	good
S10	bad	bad	medium	bad
S11	medium	bad	medium	bad
S12	good	bad	medium	good
S13	bad	medium	medium	bad
S14	medium	medium	medium	medium
S15	good	medium	medium	medium
S16	bad	good	medium	good
S17	medium	good	medium	medium
S18	good	good	medium	good
S19	bad	bad	good	bad
S20	medium	bad	good	bad
S21	good	bad	good	good
S22	bad	medium	good	bad
S23	medium	medium	good	medium
S24	good	medium	good	good
S25	bad	good	good	good
S26	medium	good	good	good
S27	good	good	good	good

z_t, $t = 2, 3$, considered in proposition 3) of Theorem 4.3, are set to the values $z_2 = 10$, $z_3 = 30$, and the interval classification is performed by checking the following conditions:

$$f^{\geq}[g_1(x_1), g_2(x_2), g_3(x_3)] \geq 10 \Leftrightarrow \text{"}x \text{ is at least medium"}$$

$$f^{\geq}[g_1(x_1), g_2(x_2), g_3(x_3)] \geq 30 \Leftrightarrow \text{"}x \text{ is (at least) good"}$$

$$f^{\leq}[g_1(x_1), g_2(x_2), g_3(x_3)] \leq 10 \Leftrightarrow \text{"}x \text{ is (at most) bad"}$$

$$f^{\leq}[g_1(x_1), g_2(x_2), g_3(x_3)] \geq 30 \Leftrightarrow \text{"}x \text{ is at most medium"}.$$

The inconsistent classification of Table 4.1, or the equivalent interval classification of Table 4.2, can also be represented by the following set of "at least" and "at most" decision rules:

1) "*if* Mathematics $\geq$ medium *and* Physics $\geq$ medium, *then* student $\geq$ medium"
2) "*if* Mathematics $\geq$ good, *then* student $\geq$ medium",
3) "*if* Physics $\geq$ good, *then* student $\geq$ medium",
4) "*if* Mathematics $\geq$ good *and* Physics $\geq$ good, *then* student $\geq$ good",
5) "*if* Mathematics $\geq$ good *and* Literature $\geq$ good, *then* student $\geq$ good",
6) "*if* Physics $\geq$ good *and* Literature $\geq$ good, *then* student $\geq$ good",
7) "*if* Mathematics $\leq$ bad *and* Physics $\leq$ medium, *then* student $\leq$ bad",
8) "*if* Mathematics $\leq$ medium *and* Physics $\leq$ bad, *then* student $\leq$ bad",
9) "*if* Mathematics $\leq$ medium *and* Physics $\leq$ medium, *then* student $\leq$ medium".

Table 4.2. Interval classification by rules and by functions $f^{\geq}[g_1(x_1), g_2(x_2), g_3(x_3)]$ and $f^{\leq}[g_1(x_1), g_2(x_2), g_3(x_3)]$ ($f^{\geq}(x) = f[g_1(x_1), g_2(x_2), g_3(x_3)]$ and $f^{\leq}(x) = f[g_1(x_1), g_2(x_2), g_3(x_3)]$)

Student	Mathematics	Physics	Literature	Decision	$f^{\geq}(x)$	$f^{\leq}(x)$	Matching rules
S1	bad	bad	bad	bad	0	0	#7,8,9
S2	medium	bad	bad	bad	1	1	#8,9
S3	good	bad	bad	medium-good	18	32	#2
S4	bad	medium	bad	bad	1	1	#7,9
S5	medium	medium	bad	medium	18	18	#1,9
S6	good	medium	bad	medium-good	20	33	#1,2
S7	bad	good	bad	medium-good	18	32	#3
S8	medium	good	bad	medium-good	20	33	#1,3
S9	good	good	bad	good	36	36	#1,2,3,6
S10	bad	bad	medium	bad	1	1	#7,8,9
S11	medium	bad	medium	bad	3	3	#8,9
S12	good	bad	medium	medium-good	19	33	#2
S13	bad	medium	medium	bad	3	3	#7,9
S14	medium	medium	medium	medium	19	19	#1,9
S15	good	medium	medium	medium-good	23	35	#1,2
S16	bad	good	medium	medium-good	19	33	#3
S17	medium	good	medium	medium-good	23	35	#1,3
S18	good	good	medium	good	41	41	#1,2,3,4
S19	bad	bad	good	bad	2	2	#7,8,9
S20	medium	bad	good	bad	5	5	#8,9
S21	good	bad	good	good	34	34	#2,5
S22	bad	medium	good	bad	5	5	#7,9
S23	medium	medium	good	medium	20	20	#1,9
S24	good	medium	good	good	37	37	#1,2,5
S25	bad	good	good	good	34	34	#3,6
S26	medium	good	good	good	37	37	#1,3,6
S27	good	good	good	good	46	46	#1,2,3,4,5,6

In the last column of Table 4.2 the numbers of rules matching the particular

5. Conclusions

We considered a class of multiple-criteria decision problems, called multiple-criteria classification, concerning an assignment of some actions to some predefined and preference-ordered decision classes. The actions are described by a finite set of criteria with ordinal scales. We characterized the multiple-criteria classification problem by a representation theorem stating equivalence of a very simple cancellation property, a general utility function and a specific outranking relation, on the one hand, and a decision rule model on the other hand. When considering the decision rule model, we focused our attention on the rough set theory, offering an original methodology for extraction of knowledge from data that, in this case, are examples of classification provided by a decision maker. The advantage of the rough set approach with respect to competitive methodologies is the possibility of handling partially inconsistent data that are often encountered in preferential information, due to hesitation of decision makers, unstable character of their preferences, imprecise or incomplete information and the like. Therefore, we proposed a general model of conjoint measurement that, using the basic concepts of the rough set approach (lower and upper approximations), is able to represent these inconsistencies by a specific utility function. We showed that these inconsistencies can also be represented in a meaningful way by "*if...*, *then...*" decision rules induced from rough approximations.

As the rough set approach to multiple-criteria classification problems and the underlying decision rules exploit only the ordinal properties of the scales of criteria, they are appropriate for aggregation of ordinal criteria. This challenging problem of multiple-criteria decision making has been solved until now by using some max-min aggregation operators, with the most general one - the fuzzy integral proposed by Sugeno. We showed that the decision rule model following from the rough set approach has advantages over the integral of Sugeno, in particular, it can represent some (even consistent) preferences that the Sugeno integral cannot.

The characterization of the decision rule preference model performed in this paper shows clearly its extraordinary capacity of criteria aggregation in multiple-criteria classification problems. The decision rule preference model, apart from its capacity of representation, fulfils the postulate of transparency and interpretability of preference models in decision aiding. The characterization shows that the decision rule preference model is a strong alternative to functional and relational preference models to which it is formally equivalent. Recently, similar benefits of the decision rule model have been proved with respect to multiple-criteria choice and ranking problems (Greco, Matarazzo and Słowiński, 2002b).

Acknowledgements

The first author wishes to acknowledge the financial support from the State Committee for Scientific Research, grant KBN 8T11F 006 19, and from the

Foundation for Polish Science, subsidy no. 11/2001. The research of the two other authors has been supported by the Italian Ministry of University and Scientific Research (MURST).

References

ALTMAN, E.I. (1968) Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance*, 589–609.

CAPON, N. (1982) Cedit scoring systems: a critical analysis. *Journal of Marketing*, **46**, 32–91.

GAL, T., STEWART, T. and HANNE, T. (eds.), (1999) *Advances in Multiple Criteria Decision Making.* Kluwer Academic Publishers, Dordrecht, Boston.

GOLDSTEIN, W.M. (1991) Decomposable threshold models. *Journal of Mathematical Psychology*, **35**, 64–79.

GRABISCH, M. (1995) Fuzzy integral in multiple-criteria decision making. *Fuzzy Sets and Systems*, **69**, 279–298.

GRECO, S., MATARAZZO, B. and SŁOWIŃSKI, R. (1999) The use of rough sets and fuzzy sets in MCDM. Chapter 14. In: T. Gal, T. Stewart and T. Hanne, eds., *Advances in Multiple Criteria Decision Making*, Kluwer Academic Publishers, Dordrecht, Boston, 14.1–14.59.

GRECO, S., MATARAZZO, B. and SŁOWIŃSKI, R. (2000) Extension of the rough set approach to multicriteria decision support. *INFOR*, **38**, 3, 161–196.

GRECO, S., MATARAZZO, B. and SŁOWIŃSKI, R. (2001a) Rough sets theory for multicriteria decision analysis. *EJOR*, **129**, 1, 1–47.

GRECO, S., MATARAZZO, B. and SŁOWIŃSKI, R. (2001b) Conjoint measurement and rough set approach for multicriteria sorting problems in presence of ordinal criteria. In: A. Colorni, M. Paruccini, B. Roy, eds., *A-MCD-A: Aide Multi Critere a la Décision – Multiple Criteria Decision Aiding*, European Commission Report, EUR 19808 EN, Ispra, 117–144.

GRECO, S., MATARAZZO, B. and SŁOWIŃSKI, R. (2002a) Rough sets methodology for sorting problems in presence of multiple attributes and criteria. *EJOR*, **138**, 2, 247–259.

GRECO, S., MATARAZZO, B. and SŁOWIŃSKI, R. (2002b) Preference representation by means of conjoint measurement and decision rule model. In: D. Bouyssou, E. Jacquet-Lagrèze, P. Perny, R. Słowiński, D. Vanderpooten, Ph. Vincke, eds., *Aiding Decisions with Multiple Criteria - Essays in Honor of Bernard Roy.* Kluwer Academic Publishers, Boston, 263–313.

GRECO, S., MATARAZZO, B. and SŁOWIŃSKI, R. (2002c) Multicriteria classification. Chapter 16.1.9. In: W. Kloesgen and J. Zytkow, eds., *Handbook of Data Mining and Knowledge Discovery*, Oxford University Press, New York, 318–328.

MODAVE, F. and GRABISCH, M. (1998) Preference representation by the Choquet Integral: the commensurability hypothesis. In: Proc. 7th Int. Conference on *Information Processing and Management of Uncertainty in Knowledge Based Systems*, Paris, La Sorbonne, 164–171.

PAWLAK, Z. (1982) Rough sets. *International Journal of Information & Computer Sciences*, **11**, 341–356.

PAWLAK, Z. (1991) Rough Sets. *Theoretical Aspects of Reasoning about Data.* Kluwer, Dordrecht.

PAWLAK, Z. and SŁOWIŃSKI, R. (1994) Rough set approach to multi-attribute decision analysis. *EJOR*, **72**, 443–459.

ROBERTS, F.S. (1979) *Measurement theory with applications to decision-making, utility and the social science.* Addison-Wesley Publ., Reading, MA.

ROY, B. (1985) *Méthodologie Multicritère d'Aide à la Décision.* Economica, Paris.

ROY, B. (1989) Main sources of inaccurate determination, uncertainty and imprecision in decision models. *Mathematical and Computer Modelling*, **12**, 10/11 1245–1254.

ROY, B. and BOUYSSOU, D. (1993) *Aide Multicritère à la Décision: Méthodes et Cas.* Economica, Paris.

SŁOWIŃSKI, R. (ed.), (1992) *Intelligent Decision Support. Handbook of Applications and Advances of the Rough Sets Theory.* Kluwer Academic Publishers, Dordrecht.

SŁOWIŃSKI, R., STEFANOWSKI, J., GRECO, S. and MATARAZZO, B. (2000) Rough sets based processing of inconsistent information in decision analysis. *Control and Cybernetics*, **29** 1, 379–404.

SUGENO, M. (1974) *Theory of fuzzy integrals and its applications.* Doctoral Thesis, Tokyo Institute of Technology.

ZOPOUNIDIS, C. and DOUMPOS M. (1998) Developing a multicriteria decision support system for financial classification problems: the FINCLAS system. *Optimization Methods and Software* **8**, 277–304.