FACULTÉ POLYTECHNIQUE DE MONS

Two Aspects of User Centered Data Analysis and Decision Aid

Quality Measures for Association Rules and Multiple Criteria Sorting

DEA Thesis
Patrick MEYER

July 2003

First of all, I am very thankful to my advisor Marc Roubens for the fructuous discussions we had on the subject of the second article of this thesis and for supporting my work during the whole period of our collaboration.

I would also like to thank Philippe Lenca for suggesting me to collaborate on the subject of the first article of the present work.

Finally, I would like to thank all the people who helped me consciously or not to achieve the present work.

Foreword

This DEA thesis is the result of two years of research on various subjects involving Multiple Criteria Decision Aid (MCDA) and Data Mining. On the one hand, the close collaboration with Marc Roubens has lead to an MCDA tool for classification in the presence of interacting criteria. On the other hand, discussions on knowledge discovery in databases with Philippe Lenca and his research team have resulted in a rigorous analysis of quality measures for association rules and a detailed decision aiding procedure.

The present work is an attempt to summarise and link both aspects of my work. Two articles represent its structure: first of all, the results of a collaboration on quality measures of association rules with Philippe Lenca, Philippe Picouet and Benoît Vaillant from the IASC Department of the ENST Bretagne (F) and Stéphane Lallich of the Laboratoire ERIC of the University Lumière in Lyon (F); and secondly the work on multiple criteria sorting in collaboration with Marc Roubens from the Department of Mathematics of the University of Liège (B).

The following general introductive section, as well as the preliminary considerations before each of both articles are meant to create a link between these different subjects. They should show that the role of the user in any of these approaches is one of our main concerns.

Introduction

With the permanently increasing speed of modern computers and their high and cheap storage capacity, it has become very affordable to store large amounts of data on any possible subject. Nowadays, nearly any transaction is recorded somewhere. The local supermarket as well as any phone company stores our habits and actions. More and more hospitals record and keep detailed descriptions of their patients, their symptoms and their diseases. All this with one main objective: mine this huge amount of data for valuable information. Supermarket managers are willing to know what we buy and how we combine the goods. Phone companies search for typical patterns of bad payers. And hospitals for example search for a more secure and reliable way to detect diseases. These are only a few examples of where and how data analyses can be used. These type of processes clearly belong to the data mining field.

Similarly, when it comes to making complex decisions in various fields, the economic, ecological or social impact of decisions has become a major concern for many experts. These aspects have to be taken into account among many others. The processes which allow to aid making decisions require to be simultaneously rigourous and flexible. Furthermore, the resulting models should be easily interpretable. Hence, the determination of potential alternatives and discriminant points of view is a complex task which requires a constant dialogue between the experts of the field and the analysts. The modelisation of the preferences of the decision maker as well as the specific structure of the problem are therefore two major concerns of any decision aiding procedure.

These two apparently very different research fields nevertheless meet on the subject of user interaction. In Data Mining, the human aspect should not be neglected, even if a great part of the work is done by computers.

Much time needs to be invested in calibrating the methods so that they fit to the specificities of the analysed problems. This involves more than the determination of the right ranges for the various parameters. The correct interpretation of the results and the extracted information is a determining and fastidious task. The analyst who has to deal with a specific problem should ideally check and validate the relevance of the results in collaboration with a human expert of the analysed field. In Multiple Criteria Decision Aiding, the human factor is even more important. Usually, the whole process of the building of the points of view cannot be done without a permanent and intensive discussion between the analyst and the expert of the field. Furthermore, many methods require preferential independance among the criteria. This shows for example that the final data matrix needs to be built according to certain specificities.

This work deals with two particular and quite different problematics. Nevertheless it permanently focuses on the user. In the first part, we present our research on the evaluation of quality measures for association rules. Association rules clearly belong to the field of data mining. But our work more specifically focuses on the choice of a good measure of quality of the extracted knowledge. The role of the user is very important here, as his objectives and preferences have to be reflected in the choice of the measure. In a first step, the different existing quality measures are analysed and described by a set of properties. After a discussion with the expert of the mined field, criteria are built, and the problem becomes a multiple criteria decision aiding problem. Some analyses are performed with the Promethee-Gaia method. The second part of this work deals with a multiple criteria sorting procedure. It is called TOMASO and deals with data evaluated on ordinal scales. Its main specificity is to allow criteria to interact. This is done through the use of the Choquet integral as an aggregator. The role of the user is again important. In order to obtain an ordered sorting, he has to provide a set of prototypes. Prototypes are some alternatives which are well-known by the Decision Maker. He has an a priori knowledge on the belonging of the alternatives to the predefined classes.

Chapter 1

Data Mining

Nowadays we are overwhelmed with data. The amount of data in every parts of our lives is increasing faster and faster. The personal computers allow to store data which previously were deleted or thrown away. Everyday actions and facts are recorded on cheap storage supports. Many of our decisions are stored: our choices in supermarkets, our financial habits, our diseases, our World Wide Web surfing habits. But this huge quantity of information is useless if nobody can understand its structure and some underlying patterns. In all these data, information is hidden, that is useful and that is rarely made explicit or taken advantage of.

Data mining is about solving problems by analysing data already present in databases [Witten]. The search for patterns or information is automated, or at least augmented, by computers. Those considerations lead to define data mining as the process of discovering patterns in data, in an automatic or semi-automatic way.

These discovered patterns must be meaningful and lead to some advantage. Furthermore, they should allow to make non trivial predictions on new data. Two primitive types of information can be extracted from data: "black box"-like information or structural patterns. The first type of patterns is effectively incomprehensible. However, the structural patterns are represented in a way that allows a further examination and more detailed investigations. In both cases, it is assumed that the extracted information allows to makes good predictions (on the analysed data, or on new data).

Next to these general considerations on data mining, one should also mention that the machine learning techniques used for data mining can be divided into two groups (related to the type of problem which is considered): the supervised learning methods, and the unsupervised learning algorithms.

The first category includes methods like decision trees (structural patterns) or neural networks (black box). Their goal is to search for patterns or rules in the data which explain a certain output variable. As an example one can consider the classical soybean disease problem. Data are taken from questionnaires describing plant diseases. There are about 680 examples, each representing a diseased plant. These were measured on 35 attributes (or variables). Furthermore, the examples are labeled with the diagnosis of an expert in plant biology (17 possible disease categories). A supervised learning algorithm like a decision tree would now search for rules which permit to reproduce the output diagnosis categories by means of the input attributes.

The second category involves clustering techniques like k-means or incremental clustering which divide the instances into natural groups. Here, no "output" variable has to be predicted, and the goal is merely to search for strong patterns in the data. These methods are interesting in case no "global" information on the data is known, and the analyst seeks for unknown categories of objects or examples.

The first chapter of this work deals with a particular type of algorithms. The extracted knowledge is called "association rules". The main idea is that association rules can predict any attribute and combinations of attributes. This is best illustrated with the "supermarket basket" concept. Consider a classical list of consumers with their respective supermarket basket. This information can be represented as a data matrix, where each line corresponds to a customer, and the columns represent all the possible items which can be bought in the shop. If consumer i has bought item j, then element (i, j) of this matrix takes the boolean value "true". Otherwise it takes the value "false". Such a data table is exploited by association rules algorithms to find strong rules, or patterns in the behaviour of consumers. For example, a possible and probable output of such a analysis would be a rule similar to "80% of consumers who bought sugar also bought eggs and flour". If the algorithm mines the data without any restrictions, this type of method falls into the category of unsupervised learning. Some methods allow to restrict the search to a certain set of attributes (if some a priori knowledge on the structure exists). In that case, this type of algorithm belongs to the set of supervised learning methods.

Different association rules express different regularities that underlie the dataset. Because so many different association rules can be derived from a dataset, it is commonly accepted that the interest must be restricted to those applying to a reasonably large number of instances and have a reasonably high accuracy. The *support* of an association rule is the number of instances for which it predicts correctly. The *confidence* is the number of instances that it predicts correctly, expressed as a proportion of all instances it applies to [Witten]. But both quality indexes of association rules have weak points. Therefore, many other measures of the quality of association rules have been suggested. Often they satisfy some local requirements, but they tend to have disadvantages too. In order to chose the appropriate quality measure which corresponds to the preferences of an expert, we suggest to analyse each proposed measure according to a set of attributes, and perform a Multiple Criteria Decision Aiding analysis by the Promethee-Gaia method.

The following article is a translation of a more detailed article in french which has been presented on the 35th Statistical Days in Lyon (2003) [Lenca]. It shows that it is interesting to compare the different quality measures on the basis of some attributes. Besides, it presents an approach to chose the measure which best corresponds to the type of rules it should filter.

Useful characterization and multicriteria decision aid: A two step approach to interestingness measure selection

Translation of a paper presented on the 35th Statistical Days in Lyon (2003) [Lenca].
Authors:

- Stéphane Lallich
  Laboratoire ERIC
  University Lumière - Lyon
  5 avenue Pierre Mendès-France
  69676 Bron Cedex, France
  `lallich@univ-lyon2.fr`

- Philippe Lenca, Philippe Picouet, Benoit Vaillant
  IASC Department
  ENST Bretagne BP 832
  29285 Brest Cedex, France
  `firstname.name@enst-bretagne.fr`

- Patrick Meyer
  Department of Mathematics
  University of Liège
  Bat. B37 Grande Traverse, 12
  B-4000 Liège, Belgium
  `patrick.meyer@internet.lu`

Abstract

Datamining algorithms, especially those used for unsupervised learning, generate a large number of rules. It is hence impossible for an expert of the field being mined to evaluate these rules. To help carrying out the task, many interestingness measures for rules have been developed in order to filter and automatically sort a set of rules with respect to given goals. Since every measure gives a different result, and every expert has a different understanding of what a "good" rule is, we propose in this article a double-step solution to the choice problem of a user-adapted interestingness measure: First, a characterisation of various interestingness measures, based on meaningful classical properties, is provided. Then a multiple criteria decision assistance process is applied to this characterisation and illustrates the benefit for a user, who is not an expert in data mining.

1.1 Introduction

One of the main objectives of Knowledge Discovery in Databases (KDD) is to produce interesting rules with respect to some user's goals and preferences. This user is not supposed to be a data mining expert, but rather an expert of the field being mined. Moreover, it is well known that the interestingness of a rule is difficult to evaluate with objectivity. Indeed, this estimation greatly depends on our expert user's interests [18, 17]. Ideally, a rule should be valid, new and comprehensive [9] but these generic terms cover a large and various number of situations when examined in a precise context. In [33], the interestingness of a measure is analyzed on a two fold basis: the concordance with the expert's beliefs, and its capacity to become an action. Such models are, however, difficult to apply since they need a long study of the beliefs of experts.

In this context, evaluation measures of rule interest play an essential role in KDD processes in order to find the "best rules". Depending on the user's goals, data mining experts may choose the required interestingness measure, but this task cannot be solved by an expert user on his own. Indeed, this choice is hard since rule interestingness measures have many different and contradictory qualities or flaws [36, 20]. A well-known example of a such controversial measure is the support (or covering rate). On the one hand, it is heavily used for filtering purposes in KDD algorithms [1, 27], for its antimonotonicity property simplifies the large lattice that has to be explored. On the other hand, it has almost all the flaws a user would certainly like to avoid: variability of the value under the independence hypothesis or for a logical rule, and so on. Different works [35, 17, 19] have formally extracted several specificities of measures.

The importance of objective evaluation criteria of interestingness measures has already been focused on in [30] and [10]. However, the relevant aspects of these criteria to help the user choose the right measure is still difficult to establish. In [36], the authors provide a comparative study according to these properties and provide an original approach to measure selection by an expert. However, this approach does not exploit the above-mentioned comparative table: from the set of rules resulting from a data mining algorithm, authors propose to extract a small subset of rules where measures give very different results. The authors experimentally establish that the diversity of results on the rule subset enable the user to efficiently select the best adapted measure.

Our article might be seen as an alternative contribution to [36]. We promote a double-step process.

- first, we provide a comparative description of measures thanks to a list of properties that are partially different from the properties evaluated in [36], since some of them do not apply efficiently to association rule interestingness, or others do not make any distinction between the different kinds of interestingness measures we studied. In addition, we introduce and study new properties, such as the easiness to fix a threshold, for filtering purposes. We think that users are quite interested in this property.
- second, we propose to use a Multi-Criteria Decision Aid (MCDA) method on the previous identified set of properties. MCDA [31] methods have already proved their utility in different fields. We argue in this paper that an MCDA could be profitable for the specific problem of a user's choice of a measure.

This paper is organized as follows. In the next Section, we introduce a representative list of existing measures, frequently used in the scientific context of association rules. In Section 1.3, we report some experimental results that underline the diversity of measure evaluation. We propose in Section 1.4 a list of 8 meaningful properties (from the user's point of view) and evaluate the previous list of measures according to them. Section 1.5 is dedicated to the use of the Decision Aid method PROMETHEE. Finally, we conclude in Section 1.6.

1.2 Selected measures

In this section, we list 20 measures which evaluate the interestingness of association rules as defined in [1]: given a typical market-basket (transactional) database, the association rule $\mathtt{A} \rightarrow \mathtt{B}$ means "if someone buys the set of items $\mathtt{A}$, then he/she probably also buys item $\mathtt{B}$". It is very important to differentiate between the association rule $\mathtt{A} \rightarrow \mathtt{B}$, which focuses on cooccurence

A\B	O	1	total
0	$p_{\bar{a}\bar{b}}$	$p_{\bar{a}b}$	$p_{\bar{a}}$
1	$p_{a\bar{b}}$	p_{ab}	p_a
total	$p_{\bar{b}}$	p_b	1

Table 1.1: Notations

and gives asymmetric meaning to A and B, from logical implication A $\Rightarrow$ B or equivalence A $\Leftrightarrow$ B (see [19]).

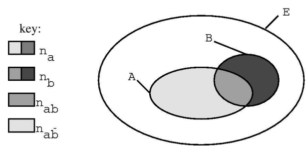

Figure 1.1: Notations.

These measures are usually defined using frequency counts presented in table 1.1 and figure 1.1 where, given a rule A $\rightarrow$ B, we note:

- $n = |\mathtt{E}|$ is the total number of records
- $n_a = |\mathtt{A}|$, the number of records satisfying A
- $n_b = |\mathtt{B}|$, the number of records satisfying B
- $n_{ab} = |\mathtt{A} \cap \mathtt{B}|$, the number of records satisfying the rule
- $n_{a\bar{b}} = |\mathtt{A} \cap \bar{\mathtt{B}}|$, the number of counter examples to the rule
- for X $\subset$ E, we note p_x instead of n_x/n when we consider relative frequency rather than absolute frequencies as in table 1.1.

It is clear that, given n, n_a and n_b or p_a and p_b, knowing one cell of table 1.1 is enough to deduce the other ones.

We have restricted the list of measures evaluated in this paper to decreasing ones, with respect to $n_{a\bar{b}}$, all marginal frequencies being fixed. This choice reflects the common assertion that the fewer counter-examples (A true and B false) to the rule there are, the higher the interestingness of the rule. For a given decreasing monotonic measure μ, it is then possible to select interesting rules by positioning a threshold α and keeping only the rules satisfying $\mu(A \rightarrow B) \geq \alpha$.

Some measures like χ^2, Pearson's r^2, Goodman and Smyth's J-measure [12] or Pearl's measure [28] had to be excluded from this list, because they are not monotonically decreasing. For a discussion about the interest of some of these measures see [35, 5, 34].

	Name	References
SUP	support	[1]
CONF	confidence	[1]
R	Pearson's correlation coefficient	[29]
CENCONF	centered confidence	
PS	Piatetsky-Shapiro	[30]
LOE	Loevinger	[25]
ZHANG	Zhang	[37]
- IMPIND	implication index	[14]
LIFT	Lift	[26]
SURP	statistical surprise	[2]
SEB	Sebag and Schoenauer	[32]
OM	odd multiplier	[19]
CONV	conviction	[6]
ECR	examples and counter-examples rate	
KAPPA	Kappa coefficient	[8]
IG	information gain	[7]
INTIMP	intensity of implication	[14]
EII	entropic intensity of implication	[16]
PDI	probabilistic discriminant index	[22]
LAP	Laplace	[11]

Table 1.2: List of selected measures

From previous works, we selected a reasonable set of such measures of interest: they are listed (with their bibliographical reference) in table 1.3.

Of course, we kept the well-known support (p_{ab}) and confidence ($p_{b|a}$): these are the two most frequently used measures in association rule extraction algorithms based on the selection of frequent itemsets.

Many other measures are linear transformations of the confidence, enhancing it, as they enable comparisons with p_b. This transformation is generally achieved by a centering of the confidence on p_b, using different scale coefficients (centered confidence, Piatetsky-Shapiro's measure, Loevinger's measure, Zhang's measure, correlation, implication index, surprise). It is also possible to divide the confidence by p_b (lift).

[1] $ImpInd^{RC/B}$ is the IMPIND measure, reduced and centered (RC) for a given rule

	Absolute definitions	Relative definitions
Sup	$\frac{n_a - n_{a\bar{b}}}{n}$	p_{ab}
Conf	$1 - \frac{n_{a\bar{b}}}{n_a}$	$p_{b/a}$
r	$\frac{nn_{ab} - n_a n_b}{\sqrt{nn_a n_b n_{\bar{a}} . n_{\bar{b}}}}$	$\frac{p_{ab} - p_a p_b}{\sqrt{p_a p_{\bar{a}} p_b p_{\bar{b}}}}$
CenConf	$\frac{nn_{ab} - n_a n_b}{nn_a}$	$p_{b/a} - p_b$
PS	$\frac{1}{n}(\frac{n_a n_{\bar{b}}}{n} - n_{a\bar{b}})$	$np_a \left(p_{b/a} - p_b\right) = np_a p_b \,(\text{Lift} - 1)$
Loe	$1 - \frac{nn_{a\bar{b}}}{n_a n_{\bar{b}}}$	$\frac{p_{b/a} - p_b}{p_{\bar{b}}} = \frac{1}{p_{\bar{b}}}\text{CenConf} = 1 - \frac{1}{\text{Conv}}$
Zhang	$\frac{nn_{ab} - n_a n_b}{\max\{n_{ab} n_{\bar{b}}, n_b n_{a\bar{b}}\}}$	$\frac{p_{ab} - p_a p_b}{\max\{p_{ab} p_{\bar{b}}; p_b (p_a - p_{ab}\}}$
- ImpInd	$\frac{nn_{a\bar{b}} - n_a n_{\bar{b}}}{\sqrt{nn_a n_{\bar{b}}}}$	$\sqrt{n}\frac{p_{a\bar{b}} - p_a p_{\bar{b}}}{\sqrt{p_a p_{\bar{b}}}}$
Lift	$\frac{nn_{ab}}{n_a n_b}$	$\frac{p_{ab}}{p_a p_b}$
Surp	$\frac{n_{ab} - n_{a\bar{b}}}{nn_b}$	$\frac{p_{ab} - p_{a\bar{b}}}{p_b} = 2\frac{p_a}{p_b}\,(\text{Conf} - 0.5)$
Seb	$\frac{n_a - n_{a\bar{b}}}{n_{a\bar{b}}}$	$\frac{p_{ab}}{p_{a\bar{b}}} = \frac{\text{Conf}}{1 - \text{Conf}}$
OM	$\frac{(n_a - n_{a\bar{b}}) n_{\bar{b}}}{n_b n_{a\bar{b}}}$	$\frac{\frac{p_{b/a}}{p_{\bar{b}/a}}}{\frac{p_b}{p_{\bar{b}}}} = \frac{p_{ab}}{p_b}\frac{p_{\bar{b}}}{p_{a\bar{b}}} = \text{Lift} \cdot \text{Conv}$
Conv	$\frac{n_a n_{\bar{b}}}{nn_{a\bar{b}}}$	$\frac{p_a p_{\bar{b}}}{p_{a\bar{b}}}$
ECR	$\frac{n_a - 2n_{a\bar{b}}}{n_a - n_{a\bar{b}}} = 1 - \frac{1}{\frac{n_a}{n_{a\bar{b}}} - 1}$	$1 - \frac{p_{a\bar{b}}}{p_{ab}} = 1 - \frac{1}{Seb}$
Kappa	$2\frac{nn_a - nn_{a\bar{b}} - n_a n_b}{nn_a + nn_b - 2n_a n_b}$	$2\frac{p_{ab} - p_a p_b}{p_a + p_b - 2p_a p_b}$
IG	$\log(\frac{nn_{ab}}{n_a n_b})$	$\log\frac{p_{ab}}{p_a p_b} = \log\,(\text{Lift})$
IntImp	$\varphi = P\Big[poisson(\frac{n_a n_{\bar{b}}}{n}) \geq n_{a\bar{b}}\Big]$	$P\,[Poisson\,(np_a p_{\bar{b}}) \geq np_{a\bar{b}}]$
EII	$\{[(1 - h_1(\frac{n_{a\bar{b}}}{n})^2)(1 - h_2(\frac{n_{a\bar{b}}}{n})^2)]^{1/4}\varphi\}^{1/2}$	$\{[(1 - h_1(p_{ab})^2)(1 - h_2(p_{ab})^2)]^{1/4}\varphi\}^{1/2}$
PDI [1]	$P\Big[\mathcal{N}(0,1) > ImpInd^{CR/B}\Big]$	
Lap	$\frac{n_{ab} + 1}{n_a + 2}$	$\frac{p_{b/a} + \frac{n}{pa}}{1 + \frac{2n}{pa}}$

Table 1.3: List of selected measures

Other measures, like Sebag-Schoenauer's or the example and counter example rate, are monotonically increasing transformations of confidence, while Information gain is a monotonically increasing transformation of the lift. Some measures focus on counter examples, like conviction $\frac{p_{\overline{b}|a}}{p_{\overline{b}}}$ or the above cited implication index. This latter measure is the basis of several different probabilistic measures like the probabilistic discriminant index, the intensity of implication or its entropic version, which takes into account an entropic coefficient, enhancing the small discriminant power of the intensity of implication. Finally, the odd multiplier is a kind of odd-ratio, based on the comparison of the odd of `A` and `B` on `B` rather than the odd of `A` and $\overline{\texttt{A}}$ on `B`; and Laplace's measure is a variation of the confidence, taking the total number of records n into account.

1.3 Total pre-order comparison

In order to get an idea of the difficulty of selecting the subset of the n best rules, we studied the total pre-orders induced by the measures' values on data sets. We calculate an objective comparison which tells us how two pre-orders are different.

This comparison is based on countings over all the possible couples of rules. For a couple of rules (r_1, r_2)[2], and two measures μ_1 and μ_2, we define six possibilities:

- there is strict agreement when $\mu_1(r_1) < \mu_1(r_2) \Leftrightarrow \mu_2(r_1) < \mu_2(r_2)$.
- there is co-agreement when $\mu_1(r_1) > \mu_1(r_2) \Leftrightarrow \mu_2(r_1) > \mu_2(r_2)$.
- there is large agreement if the two rules are equivalent for both measures.
- there is semi-agreement if, for one of the measures, the value taken for r_1 is greater than the value taken for r_2, the values being equal for the other measure.

database B.

[2]The order of the rules is important, and r_1 and r_2 may represent the same rule.

- there is semi-disagreement if, for one of the measures, the value taken for r_2 is greater than the value taken for r_1, the values being equal for the other measure.
- there is strict disagreement if the value taken by one of the measures is greater for r_1 than for r_2, the opposite being true for the other measure.

[24] defined the τ_1 coefficient, derived from Kendall's τ coefficient, whose values are between -1 and 1. The maximum value is obtained when both preorders are equal. In this case, there are only strict agreements, co-agreements and large agreements. The minimum value is obtained if for any couple of different rules, there is either strict disagreement, semi-disagreement or semi-agreement.

In the first case, both measures would sort the rules in the same way and the subset of the n better rules would not be affected. On the contrary, in the second case, the ordering of the rules would be reversed, and no couple of distinct rules that are equivalent for one of the measures would also be equivalent for the other. Hence, no rule is in the first half of both rankings.

Using the HERBS simulator [38], we computed the values of τ_1 for 18 measures. The results are presented in table 1.4. The database is `cmc` (contraceptive method choice, [23], which is a subset of the 1987 National Indonesia Contraceptive Prevalence Survey), the rule database is composed of 444 rules which were generated by the `apriori` algorithm [3] with a support threshold of 10% and a confidence threshold of 80%.

We have only 4 negative values, the lowest being -0.14 for (SURP, OM). The average value of τ_1 is $0,52$ (we did not take into account the diagonal values), and the variance is 0.08.

Some of the values are equal to 1, and this could have been predicted as the measures are monotonical increasing transformations of one another. This is the case for (SEB, ECR, CONF), (-IMPIND, PDI), (CONV, LOE) and (IG, LIFT), as is confirmed by the equations given in table 1.3.

This means that although some measures do generate the same rankings, there are some significant differences, and selecting a subset of n better rules will depend on the measure used.

	Kappa	Conf	CenConf	Conv	IG	-ImpInd	PDI	Lap	Lift	Loe	OM	PS	r	Seb	Sup	Surp	ECR	Zhang
Kappa	+1																	
Conf	+0.28	+1																
CenConf	+0.78	+0.41	+1															
Conv	+0.64	+0.61	+0.77	+1														
IG	+0.78	+0.39	+0.98	+0.75	+1													
-ImpInd	+0.84	+0.43	+0.84	+0.80	+0.82	+1												
PDI	+0.84	+0.43	+0.84	+0.80	+0.82	+1	+1											
Lap	+0.29	+0.98	+0.40	+0.60	+0.38	+0.43	+0.43	+1										
Lift	+0.78	+0.39	+0.98	+0.75	+1	+0.82	+0.82	+0.38	+1									
Loe	+0.64	+0.61	+0.77	+1	+0.75	+0.80	+0.80	+0.60	+0.75	+1								
OM	+0.13	+0.72	+0.24	+0.44	+0.23	+0.26	+0.26	+0.70	+0.23	+0.44	+1							
PS	+0.96	+0.31	+0.81	+0.67	+0.80	+0.87	+0.87	+0.32	+0.80	+0.67	+0.16	+1						
r	+0.87	+0.40	+0.83	+0.76	+0.81	+0.97	+0.97	+0.40	+0.81	+0.76	+0.24	+0.90	+1					
Seb	+0.28	+1	+0.41	+0.61	+0.39	+0.43	+0.43	+0.98	+0.39	+0.61	+0.72	+0.31	+0.40	+1				
Sup	+0.14	+0.10	-0.03	+0.03	-0.03	+0.09	+0.09	+0.11	-0.03	+0.03	+0.16	+0.13	+0.10	+0.10	+1			
Surp	+0.38	+0.15	+0.22	+0.22	+0.21	+0.33	+0.33	+0.16	+0.21	+0.22	-0.14	+0.37	+0.34	+0.15	+0.75	+1		
ECR	+0.28	+1	+0.41	+0.61	+0.39	+0.43	+0.43	+0.98	+0.39	+0.61	+0.72	+0.31	+0.40	+1	+0.10	+0.15	+1	
Zhang	+0.66	+0.59	+0.81	+0.97	+0.79	+0.82	+0.82	+0.58	+0.79	+0.97	+0.42	+0.70	+0.79	+0.59	+0.02	+0.23	+0.59	+1

Table 1.4: Comparison of total pre-orders between 18 measures

Moreover, as some measures do generate the same rankings, the user may freely take the one that best fits his preferences, without loss of interesting rules.

These two remarks led us to develop a characterization of interestingness measures, based on user preferences, in order to assist him/her in selecting the best measure, from his/her point of view.

1.4 Evaluation criteria

In this section, we propose a list of eligible properties to evaluate the previous list of measures. We first present each property, explaining its interest and its possible values on an ordinal scale. The results of the evaluations are then presented in table 1.5.

Theoretically, for each property, a user may express any preference, and our particular assignment of the values (0, 1 and sometimes 2) is not to be considered as representative of users' preferences.

g_1 **: asymmetric processing of** `A` **and** `B` **[10].** Since the head and body of a rule may have a very different signification, it is desirable to distinguish measures that give different evaluations of rules $\texttt{A} \rightarrow \texttt{B}$ and $\texttt{B} \rightarrow \texttt{A}$ from those which do not. We note 0 if the measure is symmetric, 1 otherwise.

g_2 **: decrease with** n_b **[30].** Given n_{ab}, $n_{a\bar{b}}$ and $n_{\overline{a}\bar{b}}$, it is of interest to relate the interestingness of a rule to the size of `B`. Indeed, in this situation the number of records verifying `B` but not `A` increases, hence, the interestingness of the rule should decrease. We note 1 if the measure is a decreasing function with n_b, 0 otherwise.

g_3 **: reference situations: independence [30].** to avoid keeping rules that contain no information, it is necessary to eliminate the $\texttt{A} \rightarrow \texttt{B}$ rule when `A` and `B` are independent, which means when the probability of obtaining `B` is independent of the fact that `A` is true or not. A comfortable way of dealing with this is to require that a measure's value at independence should be constant. Otherwise, fixing a threshold is difficult as the value at independence may rise over this threshold, and the associated rule would be of no use. We note 1 if the measure value is constant at independence and 0 otherwise.

The 5 properties we now introduce arise from discussions within the CNRS group GafoQualite.

g_4 **: reference situations: logical rule.** In the same way, the second reference situation we consider is related to the value of the measure when there is no counter example. It is desirable that the value should be constant or possibly infinite $+\infty$. We note 1 if the value of the measure is constant or infinite and 0 otherwise.

We do not take into account the value for the incompatibility situation. The latter reference situation is obtained when $\mathtt{A} \cap \mathtt{B} = \emptyset$, and expresses the fact that $\mathtt{B}$ cannot be realized if $\mathtt{A}$ already is. Our choice is based on the fact that incompatibility is related to the rule $\mathtt{A} \rightarrow \overline{\mathtt{B}}$ and not $\mathtt{A} \rightarrow \mathtt{B}$.

g_5 **: linearity with $p_{a\bar{b}}$ around 0^+.** Several works [15] express the desire to have a weak decrease in the neighborhood of a logic rule rather than a fast or even linear decrease (as with confidence or its linear transformations). This reflects the fact that the user may tolerate a few counter examples without significant loss of interest, but will definitely not tolerate too many. However, the opposite choice may be preferred as a convex decrease with $n_{a\bar{b}}$ around the logic rule increases the sensitivity to a false positive. We hence note 0 if the measure is convex with $n_{a\bar{b}}$ near 0, 1 if it is linear and 2 if it is concave.

g_6 **: sensitivity to n (total number of records).** Intuitively, if the rates of presence of $\mathtt{A}$, $\mathtt{A} \rightarrow \mathtt{B}$, $\mathtt{B}$ are constant, it might be interesting for the measure to react to a global extension of the database (with no evolution of rates). The preference of the user might be indifferent to having a measure which is invariant or not with the dilatation of data. If the measure increases with n and has a maximum value, then there is a risk that all the evaluations might come close to this maximum. The measure would then lose its discrimination power. We note 0 if the measure is invariant and 1 if it increases with n.

g_7 **: easiness to fix a threshold.** Even if properties g_3 and g_4 are valid, it is still difficult to decide the best threshold value that separates interesting from uninteresting rules. This property allows us to identify measures whose threshold is more or less difficult to locate. To establish this property, we propose to proceed in the following (and very conventional) way by providing a sense of the strength of the evidence against the null hypothesis, that is the p-value. Due to the high number of tests, this probability should not be interpreted as a statistical risk, but rather as a control parameter ([19]). In some cases, the measure is defined as such a probability: the intensity of implication of Gras ([13]) or the IPD measure of Lerman and Azé ([22]) illustrate that case. More generally, we can define such a threshold from one of the three types of models proposed by Lerman ([21]) to establish the law followed by n_{ab} under the hypothesis of link absence (H_0). We note 1 if the measure easily supports such an evaluation, and 0 otherwise.

g_8 **: intelligibility.** Intelligibility denotes the ability of the measure to express a comprehensive idea of the interestingness of a rule. We will consider that a measure is intelligible if its semantics can be expressed in one concrete sentence. We affect the value 2 to this property if the measure can be expressed in that way, 1 if the measure can be estimated with common quantities, and 0 if it seems impossible to give any concrete explanation of the measure. The extension of this list is currently being studied, and in particular the discrimination and antimonotonicity characters of a measure. Discrimination is quite interesting since it might be related to criteria g_6 (sensitivity to the cardinality of the total space), which generally occurs simultaneously with a loss of discrimination. Antimonotonicity is a very interesting property from the computing point of view, both for Apriori algorithms and Galois lattice based methods [27].

We evaluated the measures described in the previous section with respect to these criteria and we obtain the following decision matrix (table 1.5).

1.5 Evaluation of the interestingness measures

In this section, we will analyze and evaluate the measures described above and resumed in table 1.3. This analysis is done by a Multi-criteria Decision Aiding procedure called Promethee [4]. Its general objectives are to build partial and complete rankings on alternatives (in this case, the measures) and to visualize the structure of the problem in a plane called the Gaia plane, similarly to a principal component analysis. The Promethee method requires information about the criteria given by a set of weights. Several tools

	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8
Sup	0	0	0	0	1	0	1	2
Conf	1	0	0	1	1	0	1	2
r	0	1	1	0	1	0	1	1
CenConf	1	1	1	0	1	0	1	2
PS	0	1	1	0	1	1	1	1
Loe	1	1	1	1	1	0	1	1
Zhang	1	1	1	1	2	0	0	0
- ImpInd	1	1	1	0	1	1	1	0
Lift	0	1	1	0	1	0	1	1
Surp	1	1	0	0	1	0	1	1
Seb	1	0	0	1	0	0	1	1
OM	1	1	1	1	0	0	1	2
Conv	1	1	1	1	0	0	1	1
ECR	1	0	0	1	2	0	1	1
Kappa	0	1	1	0	1	0	1	0
IG	0	1	1	0	2	0	1	0
IntImp	1	1	1	1	2	1	1	0
EII	1	1	1	1	2	1	0	0
PDI	1	1	1	0	1	1	1	0
Lap	1	0	0	0	1	0	1	0

Table 1.5: Decision matrix

allow these weights to be fixed in order to represent the decision maker's (DM) preferences. The first step of the method is to make pairwise comparisons on the alternatives within each criterion. This means that for small (resp. large) deviations, the DM will allocate a small (resp. large) preference to the best index. This is done through preference functions. Then, each alternative is confronted with the other alternatives in order to define outranking flows (the positive (resp. negative) outranking flow expresses how an alternative a is outranking (resp. outranked by) the others). Finally, partial and complete rankings are built out of these outrankings. The Gaia plane provides information on the conflicting character of the criteria and on the impact of the weights on the final decision. It is a projection, based on a net flow derived from the outranking flows, of the alternatives and the criteria in a common plane.

We first present some details of the method. Then we show the analysis of the measures using this method.

1.5.1 A brief look at the PROMETHEE method

Let A be a set of possible alternatives, $A = \{a_1, \ldots, a_n\}$. Let $\{f_j(\cdot), j = 1, \ldots, k\}$ be a set of evaluation criteria (to be maximized or minimized). Each of the possible alternatives of A are evaluated on each of the criteria. Intuitively, the PROMETHEE procedure is based on pairwise comparisons of the alternatives for each criterion. This comparison leads to an aggregated preference index $\pi(a, b)$ for each pair of alternatives $(a, b) \in A \times A$. Then, outranking flows are computed in order to create a partial ranking on the alternatives.

Let us consider a preference function for a particular criterion j, giving the preferences of a over b $(a, b \in A)$, which is built on the difference of the evaluations on j of a and b:

$$P_j(a, b) = P_j(f_j(a) - f_j(b)), \quad \text{for which} \quad 0 \leq P_j(a, b) \leq 1.$$

In the case of a criterion to be maximized, the preference function should look similar to figure 1.2.

The pair $\{f_j(\cdot), P_j(\cdot, \cdot)\}$ is called the generalized criterion, associated with $f_j(\cdot)$.

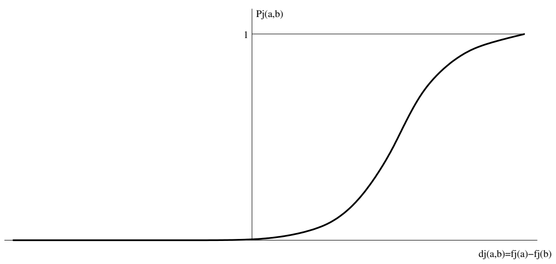

Figure 1.2: Typical preference function

The PROMETHEE approach proposes six different types of preference functions which need between 0 and 3 parameters to be fixed by the user (preference and indifference thresholds).

For each pair (a, b) of alternatives of $A \times A$, the aggregated preference index π should express to which degree a is preferred to b over all the criteria.

This index is defined as follows:

$$\pi(a,b) = \sum_{j=1}^{k} P_j(a,b) \cdot w_j$$

where w_j is the weight associated with criterion j such that $\sum_{j=1}^{k} w_j = 1$.

We see that $\pi(a, a) = 0$ and $0 \leq \pi(a, b) \leq 1$, $\forall a, b \in A$. It is clear that for a given pair (a, b) of alternatives, a preference index close to 0 implies a week global preference of a over b. Similarly, a preference index close to 1 implies a strong global preference of a over b.

In order to build a ranking on A, each alternative has to be compared with the $n - 1$ other ones of A. This is done with the computation of two outranking flows defined as follows:

- The positive outranking flow $\phi^+(a) = \frac{1}{n-1} \sum_{x \in A} \pi(a, x)$
- The negative outranking flow $\phi^-(a) = \frac{1}{n-1} \sum_{x \in A} \pi(x, a)$

The positive outranking flow expresses the power of the considered alternative, whereas the negative outranking flow gives an indication about its weakness.

Two possible rankings can be obtained in the PROMETHEE approach: a partial and a complete ranking. The partial ranking is an intersection of the rankings which can be deduced from both the positive and the negative outranking flow. It is built as follows:

$$\begin{cases} aP_pb \iff \begin{cases} \phi^+(a) > \phi^+(b) \text{ and } \phi^-(a) < \phi^-(b) \\ M\phi^+(a) = \phi^+(b) \text{ and } \phi^-(a) < \phi^-(b) \\ \phi^+(a) > \phi^+(bla) \text{ and } \phi^-(a) = \phi^-(b) \end{cases} \\ aI_pb \iff \phi^+(a) = \phi^+(b) \text{ and } \phi^-(a) = \phi^-(b) \\ aR_pb \text{ otherwise} \end{cases}$$

where P_p, I_p and R_p stand for preference, indifference and incomparability respectively.

The complete ranking is built out of the net outranking flow $\phi(a) = \phi^+(a) - \phi^-(a)$ as follows:

$$\begin{cases} aP_cb \iff \phi(a) > \phi(b) \\ aI_cb \iff \phi(a) = \phi(b) \end{cases}$$

After the computation of the rankings, the PROMETHEE procedure allows the user to view the attributes and the criteria in a common plane called the Gaia plane. It provides a clearer view of the conflicting character of the criteria, and the impact of the weights on the final decision. We will not explain the details of the Gaia plane, but only mention that it is a special principal component analysis of the data. A practical interpretation is given in the following section.

1.5.2 Analysis of two scenarios

We decide to consider the following two scenarios for our analysis:

- **Sc1**: The expert tolerates the appearance of a certain number of counter examples to a decision rule. In this case, the updating of a rule is postponed. The shape of the curve representing the value of the index versus the number of counter examples should ideally be concave (at least in the neighborhood of the maximum); the order on the values of criterion g_6 (non-linearity with respect to the number of counter-examples) is therefore (concave $\succ$ linear $\succ$ convex).

- **Sc2**: The expert refuses the appearance of too many counter-examples to a decision rule. In this case, the updating of the rule must be done more rapidly. The shape of the curve is therefore ideally convex (in the neighborhood of the maximum at least) and the order on the values of criterion g_6 is (convex $\succ$ linear $\succ$ concave).

We first analyze the problem with equal weights for the criteria. The total rankings for both scenarios are given in table 1.6.

Rank	**Sc1**	**Sc2**
1	EII	OM
2	IntImp	Conv
3	Loe	Loe
4	OM	CenConf
5	CenConf	EII
6	Conv	-ImpInd, PDI
7	-ImpInd,PDI	
8		IntImp
9	Zhang	Zhang
10	PS	PS
11	ECR	r, Lift
12	Conf	
13	IG	Seb
14	r	Conf
15	Lift	Kappa
16	Surp	Surp
17	Seb	IG
18	Kappa	ECR
19	Sup	Lap
20	Lap	Sup

Table 1.6: Rankings of the measures for both scenarios

First, we notice that both scenarios reflect the preferences of the DM on the shape of the curve. Therefore, we conclude that it is useful to make the distinction between both cases.

1.5.3 Details of Sc1

Hereafter we present the details of **Sc1**. Most of the conclusions are also valid for **Sc2** (except specific differences due to criterion g_6).

First of all, the analysis of the stability intervals of the weights give an indication about the stability of the complete rankings. We observe that they are quite narrow around the values of the weights. This indicates that a slight variation of the weight of a criterion (and an adjustment of the other weights so that their sum equals 1) will alter the complete ranking of the measures. But fortunately, these changes are only local modifications (two neighboring measures which switch positions). The general structure of the complete ranking remains even with greater changes in the weights (within certain boundaries).

The Gaia plane is given in figure 1.3.

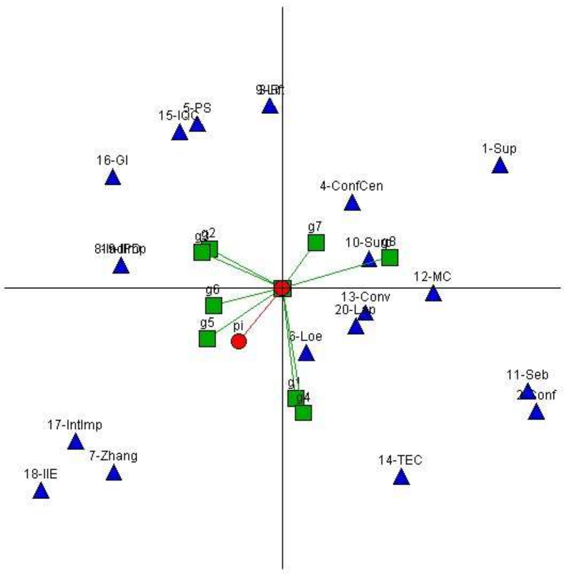

Figure 1.3: The Gaia plane for **Sc1**

1.3 is a synthetic representation of the problem in a common plane (criteria, measures). Criteria representing similar preferences on the set of measures are represented by axes with the same direction and the same orientation (e.g. criteria g_3 and g_4 or g_1 and g_5). On the other hand, criteria expressing conflicting preferences on the set of measures are represented by axes with the same direction and opposite orientations (e.g. criteria g_7 and g_9 or g_6 and g_8). Measures being good on a particular criterion are represented by points located in the direction and orientation of the axis of this criterion (e.g. SUP and criterion g_9 or ZHANG, EII, INTIMP and criteria g_6 and g_7). Similar measures are represented by points located close to each other (e.g. SEB and CONF). In most of the rankings, these similar measures usually remain close to each other.

This visualization tool clearly allows us to better understand the structure of the problem. A further important concept is the PROMETHEE decision axis (labelled pi in figure 1.3). For details, the interested reader should refer to [4]. Its orientation and direction indicate which measures will be assigned to the first positions in the complete ranking. In this particular case, it shows that for **Sc1** ZHANG, EII and INTIMP are well ranked, whereas SUP and SURP are more likely to be in worse positions. But we should like to point out that the information contained in the Gaia plane is only part of the original information. Therefore certain observations can lead to erroneous conclusions.

The decision axis (pi) can be moved around by varying the relative weights of the criteria. For example, by progressively increasing the relative weight of criterion g_9, we observe that the measure SUP moves to the top positions in the complete ranking (and similarly, less intelligible measures migrate to worse positions). This shows that one can deduce the DM's preferences on the criteria by asking him/her simple questions about the measures[3]. For example, if he/she claims to prefer measures like EII and INTIMP this clearly indicates that the intelligibility of a measure is not important for him/her.

To finish this analysis on **Sc1**, we would like to make a few concluding remarks. If the DM has no specific preferences on the weights of the criteria, EII and INTIMP seem to be a good choice for evaluating decision rules. This procedure could be used to make explicit the preferences of a DM and to

[3]In fact, by asking what group of measures should be ranked best according to the DM's preferences, one turns the problem around. By doing so, one can detect inconsistencies in the DM's reasoning, or help him to better understand the implications of his preferences for the final solution.

help him/her understand them (by asking appropriate questions about the measures). In this way, the DM could review his judgments, and improve the efficiency of the updating of decision rules. Besides, the complete rankings shown in table 1.6 are only informative and should not be seen as the only solutions. This study merely shows that a discussion with the DM leads to rankings which correspond to certain preferences which depend on the underlying problem. The PROMETHEE tool provides a useful aid for this analysis.

1.6 Conclusion

In this article, we have proposed an initial array of 20 eligible measures evaluated on 8 properties. Given this array, we have shown how to use an MCDA method, and help expert users choose an adapted interest measure in the domain of association rules.

In addition to the interest of having such a list of evaluation criteria for a large number of measures, the use of the PROMETHEE method has confirmed the fact that the expert's preferences have some influence on the ordering of the interest measures, and that there are similarities between different measures. Moreover, the PROMETHEE method allows us to make a better analysis of user preferences (the Gaia plane makes it easy to identify different clusters of criteria very close to each other).

Of course, the set of criteria has to be extended in order to reinforce the validity of the approach. Our set of criteria covers a large diversity of the user's preferences, but it is clearly not exhaustive. New criteria could also lead to a better distinction between measures similar at the present time. We are confident that some important criteria may also arise from experimental evaluation (for example, the discrimination strength).

1.7 Acknowledgments

Benot Vaillant would like to thanks the CUB (Urban Communauty of Brest) for the financial support of his Ph.D. thesis.The authors would like to thank members of the CNRS group GafoQualite for productive discussions about *interestingness measure*. Last but not least, the auhors would like to thank Janet Ormrod, english teacher at the ENST Bretagne, for reading and correcting this paper.

Bibliography

[1] R. Agrawal, T. Imielinski, and A. N. Swami. Mining association rules between sets of items in large databases. In P. Buneman and S. Jajodia, editors, *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, pages 207–216, Washington, D.C., 26–28 1993.

[2] J. Azé and Y. Kodratoff. Evaluation de la résistance au bruit de quelques mesures d'extraction de règles d'assocation. *Extraction des connaissances et apprentissage (EGC 2002)*, 1(4):143–154, 2001.

[3] C. Borgelt and R. Kruse. Induction of association rules: Apriori implementation. In *Proceedings of the 15th Conference on Computational Statistics (Compstat 2002, Berlin, Germany)*, Heidelberg, Germany, 2002. Physika Verlag.

[4] J. Brans and B. Mareschal. The promethee-gaia decision support system for multicriteria investigations. *Investigation Operativa*, 4(2):102–117, 1994.

[5] S. Brin, R. Motwani, and C. Silverstein. Beyond market baskets: generalizing association rules to correlations. In *ACM SIGMOD/PODS '97 Joint Conference*, pages 265–276, 1997.

[6] S. Brin, R. Motwani, J. D. Ullman, and S. Tsur. Dynamic itemset counting and implication rules for market basket data. In J. Peckham, editor, *SIGMOD 1997, Proceedings ACM SIGMOD International Conference on Management of Data, Tucson, Arizona, USA*, pages 255–264. ACM Press, 05 1997.

[7] K. W. Church and P. Hanks. Word association norms, mutual information an lexicography. *Computational Linguistics*, 16(1):22–29, march 1990.

[8] J. Cohen. A coefficient of agreement for nominal scale. *Educational and Psychological Measurement*, 20:37–46, 1960.

[9] U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, editors. *Advances in Knowledge Discovery and Data Mining*. AAAI/MIT Press, 1996.

[10] A. Freitas. On rule interestingness measures. *Knowledge-Based Systems journal*, pages 309–315, 1999.

[11] I. J. Good. The estimation of probabilities: An essay on modern bayesian methods. The MIT Press, Cambridge, MA, 1965.

[12] R. M. Goodman and P. Smyth. Information-theoretic rule induction. In *Proceedings of the 1988 European Conference on Artificial Intelligence*. Pitman, 1988.

[13] R. Gras. Contribution á l'étude expérimentale et á l'analyse de certaines acquisitions cognitives et de certains objectifs didactiques en mathématiques. Master's thesis, Université de Rennes I, 1979.

[14] R. Gras. *L'implication Statistique*. La Pensée Sauvage, 1996.

[15] R. Gras, R. Couturier, M. Bernadet, J. Blanchard, H. Briand, F. Guillet, P. Kuntz, R. Lehn, and P. Peter. Quelques critères pour une mesure de qualités de règles d'association - un exemple : l'intensité d'implication. Rapport de recherche pour le groupe de travail GAFO-QUALITÉE de l'action spécifique STIC fouille de bases de données (en cours de rédaction), Ecole Polytechnique de l'Université de Nantes, 2002.

[16] R. Gras, P. Kuntz, R. Couturier, and F. Guillet. Une version entropique de l'intensité d'implication pour les corpus volumineux. *Extraction des connaissances et apprentissage (EGC 2001)*, 1(1-2):69–80, 2001.

[17] R. J. Hilderman and H. J. Hamilton. Evaluation of interestingness measures for ranking discovered knowledge. *Lecture Notes in Computer Science*, 2035:247–259, 2001.

[18] M. Klemettinen, H. Mannila, P. Ronkainen, H. Toivonen, and A. I. Verkamo. Finding interesting rules from large sets of discovered association rules. In N. R. Adam, B. K. Bhargava, and Y. Yesha, editors, *Third International Conference on Information and Knowledge Management (CIKM'94)*, pages 401–407. ACM Press, 1994.

[19] S. Lallich and O. Teytaud. Évaluation et validation de l'intérêt des règles d'association. Rapport de recherche pour le groupe de travail GAFOQUALITÉE de l'action spécifique STIC fouille de bases de données, E.R.I.C., Université Lyon 2, 2002.

[20] P. Lenca, P. Meyer, B. Vaillant, and P. Picouet. Aide multicritère á la décision pour évaluer les indices de qualité des connaissances - modélisation des préférences de l'utilisateur. *RSTI-RIA (EGC 2003)*, 1(17):271–282, 2003.

[21] I. Lerman. *Classification et analyse ordinale des données*. Dunod, 1970.

[22] I. Lerman and J. Azé. Une mesure probabiliste contextuelle discriminante de qualité des règles d'association. *RSTI-RIA (EGC 2003)*, 1(17):247–262, 2003.

[23] T. Lim, W. Loh, and Y. Shih. A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms. *Machine Learning*, 40:203–228, 2000.

[24] J. Lingoes. *Geometric Representations of Relational Data*, chapter Indices of configural similarity, pages 675–679. Mathesis Press, 1979.

[25] J. Loevinger. A systemic approach to the construction and evaluation of tests of ability. *Psychological monographs*, 61(4), 1947.

[26] I. B. Machines. *IBM Intelligent Miner User's Guide, Version 1 Release 1, SH12-6213-00*. 1996.

[27] N. Pasquier, Y. Bastide, R. Taouil, and L. Lakhal. Discovering frequent closed itemsets for association rules. In C. Beeri and P. Buneman, editors, *Database Theory - ICDT '99, 7th International Conference, Jerusalem, Israel, January 10-12, 1999, Proceedings*, volume 1540 of *Lecture Notes in Computer Science*, pages 398–416. Springer, 1999.

[28] J. Pearl. *Probabilistic reasoning in intelligent systems*. Morgan Kaufmann, 1988.

[29] K. Pearson. Mathematical contributions to the theory of evolution. iii. regression, heredity and panmixia. *Philosophical Transactions of the Royal Society*, A, 1896.

[30] G. Piatetsky-Shapiro. Discovery, analysis and presentation of strong rules. In G. Piatetsky-Shapiro and W. Frawley, editors, *Knowledge Discovery in Databases*, pages 229–248. AAAI/MIT Press, 1991.

[31] B. Roy. *Multicriteria Methodology for Decision Aiding*. Ed. Kluwer Academic Publishers, 1996.

[32] M. Sebag and M. Schoenauer. Generation of rules with certainty and confidence factors from incomplete and incoherent learning bases. In J. Boose, B. Gaines, and M. Linster, editors, *Proc. of the European Knowledge Acquisition Workshop (EKAW'88)*, pages 28–1 – 28–20. Gesellschaft für Mathematik und Datenverarbeitung mbH, 1988.

[33] A. Silberschatz and A. Tuzhilin. On subjective measures of interestingness in knowledge discovery. In *Knowledge Discovery and Data Mining*, pages 275–281, 1995.

[34] C. Silverstein, S. Brin, and R. Motwani. Beyond market baskets: Generalizing association rules to dependence rules. *Data Mining and Knowledge Discovery*, 2(1):39–68, 1998.

[35] P. Tan and V. Kumar. Interestingness measures for association patterns: A perspective. Technical Report TR00-036, University of Minnesota, Department of Computer Science, 2000.

[36] P.-N. Tan, V. Kumar, and J. Srivastava. Selecting the right interestingness measure for association patterns. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 32–41, 2002.

[37] T. Terano, H. Liu, and A. L. P. Chen, editors. *Association Rules*, volume 1805 of *Lecture Notes in Computer Science*. Springer, April 2000.

[38] B. Vaillant, P. Picouet, and Lenca. An extensible platform for rule quality measure benchmarking. In *Human Centered Processes (HCP'2003)*, Luxembourg, May 2003. To be published.

Chapter 2

Multiple Criteria Decision Aiding

The data mining techniques usually treat huge amounts of existing data in order to detect some regular, interesting or surprising patterns. The field we are talking about in this introductive chapter, and in the following paper is that of Multiple Criteria Decision Aiding (MCDA). Roy [Roy 90] roughly states that the objective of multiple criteria approaches is to help us to make "better" decisions. He states that the purpose of decision-aid is to help us make our way in the presence of ambiguity, uncertainty and an abundance of bifurcations. This clearly differenciates MCDA from Data Mining.

In the context of decision making, Roy [Roy 85] differenciates 4 problematics:

- The choice problematic (α): the goal is to present a set of alternative(s) or example(s) to the decision maker (DM) which contains the "best" one(s).
- The sorting problematic (β): the goal is to assign each alternative to a predefined category.
- The ranking problematic (γ): the goal is to rank the alternatives from the best to the worst.
- The description problematic (δ): the goal is to describe the alternatives in a formal way which is adapted to the specificities of the problem.

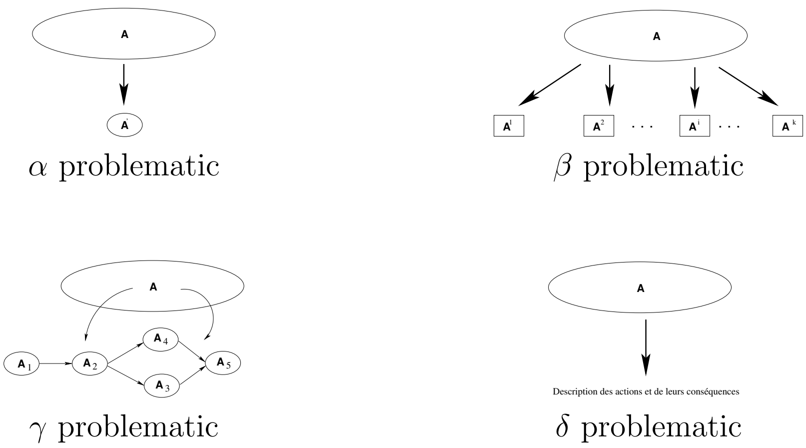

Figure 2.1: Reference problematics in multiple criteria decision aiding

The following article presents a general overview of Multiple Criteria Decision Aid and the different types of data that can be used. It then focuses on the aggregation procedures, and describes a sorting procedure called TOMASO, which deals with purely ordinal data in further details. The TOMASO software is briefly presented and applied to some case studies. Finally some considerations on a choice procedure are made.

This article will be published in [Figueira].

Choice, Ranking and Sorting in Fuzzy Multiple Criteria Decision Aid

Authors:

- Patrick Meyer, Marc Roubens
  Department of Mathematics
  University of Liège
  Grande Traverse, 12
  4000 Liège
  Belgium
  patrick.meyer@internet.lu, m.roubens@ulg.ac.be

Abstract

In this chapter we survey several approaches to derive a recommendation from some preference models for multiple criteria decision aid. Depending on the specificities of the decision problem, the recommendation can be a selection of the best alternatives, a ranking of these alternatives or a sorting. We detail a sorting procedure for the assignment of alternatives to graded classes when the available information is given by interacting points of view and a subset of prototypic alternatives whose assignment is given beforehand. A software dedicated to that approach (Tomaso) is briefly presented. Finally we define the concepts of good and bad choices based on dominant and absorbant kernels in the valued digraph that corresponds to an ordinal valued outranking relation.

2.1 Introduction

Let $A = \{\ldots, x, y, \ldots\}$ be a finite set of potential alternatives, and $\mathcal{J}$ be a set of n points of view. The Multiple Criteria Decision Problem can often be formulated as comparing and/or discriminating between the alternatives on the basis of several points of view. A first step in the Decision Aiding Process consists in the evaluation of the alternatives on each of the points of view and is possibly followed by the definition of a valued preference relation R_j on A for each dimension $j \in \mathcal{J}$.

A second step consists in either determining a global ranking on the alternatives, a sorting into different classes, or a choice function which results in a subset of alternatives of A. Two different procedures can be used: the pre-ranking methods and the pre-aggregation methods.

The *pre-ranking methods* first determine a score $S(x, R_j) = S_j(x)$ for each alternative $x \in A$ and each point of view $j \in \mathcal{J}$. An aggregation rule M_v then transforms those partial scores into a global score $S(x; R_1, \ldots, R_n)$, where v represents weights linked to the points of view. v is either a vector $(v(1), \ldots, v(n))$ or a monotone set function $v : 2^{\mathcal{J}} \to [0, 1]$ fulfilling $v(\emptyset) = 0$ and $v(\mathcal{J}) = 1$. This procedure will be used in the TOMASO method which deals with ordinal data and interacting points of view. A total crisp ordering or a sorting is obtained and all alternatives are comparable.

The *pre-aggregation methods* first determine a global binary relation R on A using an aggregation rule $M_v : R = M_v(R_1, \ldots, R_n)$. Comparisons of partial evaluations are performed dimension by dimension and their results are then aggregated. Usually this relation is constructed so as to reflect the majoritarian preference among the set of points of view. This approach allows a fine and flexible description of preferences without forcing arbitrarily alternatives to be comparable and allows to consider not only concordance between pairs of alternatives but also discordance. A global score $S'(x, R)$ transforms the global information on each pair of alternatives into a global rating related to each alternative. However a global partial order on the alternatives might be obtained if top-down or bottom-up procedures are considered (as the combination of in and out-flows in Promethee [3] and the intersection of direct and inverse complete preorders in Electre II [28]). The global valued binary relation R can also be exploited to deliver a choice set, a sorting or a ranking.

This chapter is built around three main subjects. First of all, a general description of the different ways to deal with a multiple criteria decision problem. Then we present a multiple criteria sorting procedure for the assignment of alternatives to ordered classes. The alternatives are evaluated on different interacting points of view. The objective is to aggregate these partial evaluations by the Choquet integral. The basic technique we present is due to Roubens [26]. An evolution to this method is explicited, in case the basic procedure has no solution. The fuzzy measures associated to the Choquet integral can be learnt from a subset of alternatives (called prototypes) which are assigned beforehand to the classes by the decision maker (DM). This leads in a first stage to solving a linear constraint satisfaction problem whose unknown variables are the coefficients of the fuzzy measure.

If a fuzzy measure is found, the boundaries of the classes are calculated, and the alternatives are classified. If no solution is found to this problem, an alternate way is suggested, which can lead to ambiguous assignments of the prototypes.

Both results can be analysed by means of the importance indices and the interaction indices of the assessed fuzzy measure. These two parameters give the following indications on the fuzzy measure:

- the importance indices make it possible to appraise the overall importance of each point of view and each combination of points of view;
- the interaction indices measure the extent to which the points of view interact (positively or negatively).

Finally, we focus on a choice procedure for the selection of a set of “good” alternatives.

The chapter is organised as follows. In Section 2.2 we present the different types of data which may be encountered in a multiple criteria decision problem. Then, Section 2.3 deals with valued preference relations and outranking relations. Section 2.4 presents possible ways to aggregate the information. In Section 2.5 some general concepts on sorting are introduced. In Section 2.6 we present the classical TOMASO method for sorting. In Section 2.7 an alternate method is presented, in case the classical procedure fails. The next section briefly presents the TOMASO software. Section 2.9 shows the application of the TOMASO method to some case studies. Finally, Section 2.10 deals with the choice problem. The chapter finishes on some conclusions and perspectives.

2.2 The Data Set

Without any loss of generality, we will suppose hereafter that the higher an evaluation of an alternative on a point of view is, the better the alternative is in the eyes of the decision maker.

For each point of view $j \in \mathcal{J}$, the evaluation related to each alternative is possibly given under one of the following forms:

- an *ordinal value* g_j defined on a s_j-point performance scale, that is a totally ordered set $X_j := \{g_1^j \prec_j \ldots \prec_j g_{s_j}^j\}$. It usually corresponds to linguistic ordered data.
- a *fuzzy ordinal value*, i.e. a membership function $\mu_j(u) \in [0,1]$, $\forall u \in X_j$. The degree of membership can be interpreted as the degree of compatibility of the evaluation with u. The fuzzy set is supposed to be normal ($\sup_u \mu_j(u) = 1$) and convex ($\forall u, v, w \in X_j, v \in [u,w], \mu_j(v) \leq \min\{\mu_j(u), \mu_j(w)\}$).
- a *cardinal value* g_j that associates the alternative with a real number indicating its performance. This is the most conventional way of building a preference model and in that case we are talking about a true-criterion.
- a *fuzzy interval*, i.e. a membership function $\mu_j(u) \in [0,1], \forall u \in \mathbb{R}$ that is supposed to be normal and convex. Every λ-cut is a closed interval $I_j^\lambda = \{u : \mu_j(u) \geq \lambda\}$.

  A particular example of a fuzzy interval corresponds to a *trapezoidal* *fuzzy number* defined by the parameters $(g_j^-, g_j^+, \sigma_j^-, \sigma_j^+)$:

$$\mu_j(u) = \begin{cases} 1 - \frac{g_j^- - u}{\sigma_j^-} \text{ if } g_j^- - \sigma_j^- \leq u \leq g_j^- \\ 1 \text{ if } g_j^- \leq u \leq g_j^+ \\ 1 - \frac{u - g_j^+}{\sigma_j^+} \text{ if } g_j^+ \leq u \leq g_j^+ + \sigma_j^+ \end{cases}$$

This may correspond to imprecise information on the evaluation of a given alternative: it lies possibly in the support $(g_j^- - \sigma_j^- \leq u \leq g_j^+ + \sigma_j^+)$ and belongs certainly to the kernel $(g_j^- \leq u \leq g_j^+)$.

A symmetric trapezoidal fuzzy number is such that $\sigma_j^- = \sigma_j^+$ and may translate the indifferences and preferences that might exist between values that are assessed to an alternative. In that situation we call

$$\begin{cases} g_j^+ - g_j^- = q_j \text{ (indifference threshold)} \\ g_j^+ + \sigma_j^+ - (g_j^- - \sigma_j^-) = p_j \text{ (preference threshold)} \\ \frac{g_j^+ - g_j^-}{2} = g_j \end{cases}$$

These definitions are interesting as a help to understand the concepts of indifference and preference thresholds. All the values between $g_j - \frac{1}{2}q_j$ and $g_j + \frac{1}{2}q_j$ are considered as indifferent. Values greater than $g_j + \frac{1}{2}p_j$ are better than g_j and those lower than $g_j - \frac{1}{2}p_j$ are worse than g_j. Even in the case of complete and precise information, a small positive difference does not always justify the preference.

2.3 Valued preference relation and outranking relation

Now that we have described the different possible evaluations in Section 2.2 we can define the degree to which an alternative x is not worse than y for point of view j. Let $R_j(x, y)$ be this degree, for each ordered pair (x, y) of alternatives. We use the same notations as in Section 2.2 for the different possible evaluations.

Similarly to the different possibilities described in Section 2.2, the degree $R_j(x, y)$ has different definitions and properties:

- for an *ordinal* or *cardinal value* g_j:

$$R_j(x,y) = \begin{cases} 1 \text{ if } g_j(x) \geq g_j(y) \\ 0 \text{ otherwise.} \end{cases}$$

  This crisp binary relation is a linear quasiorder.

- for a *fuzzy ordinal value*, $R_j(x,y)$ defines the degree of the preference of x over y and is considered as the possibility that x is not worse than y:

$$\begin{aligned} R_j(x,y) \;=\; \Pi_j(x \geq y) &= \max\nolimits_{u \geq v} \min(\mu_j^x(u), \mu_j^x(v)), u, v \in X_j \\ &= \max\nolimits_u \min(\mu_j^x(u), \mu_j^y(u)), u \in X_j. \end{aligned}$$

  Π_j is a valued binary relation such that $\max(\Pi_j(x,y), \Pi_j(y,x)) = 1, \forall x, y \in A$. Roubens and Vincke [27] have proved that Π_j is a fuzzy interval order and every λ-cut is a crisp interval order.

- for *fuzzy intervals*, $R_j(x,y)$ is also defined as the possibility that x is not worse than y:

$$R_j(x,y) = \Pi_j(x \geq y) = \max_{u \geq v} \min(\mu_j^x(u), \mu_j^y(v)), u, v \in \mathbb{R}.$$

  If the kernel of μ_j^x is located to the right of the kernel of μ_j^y, then $\Pi_j(x \geq y) = 1$ and $\Pi_j(y \geq x)$ equals the height of the intersection of μ_j^x and μ_j^y, $h_j(x,y)$ (see figure 2.1). This valued binary relation presents the same properties as the fuzzy ordinal value.

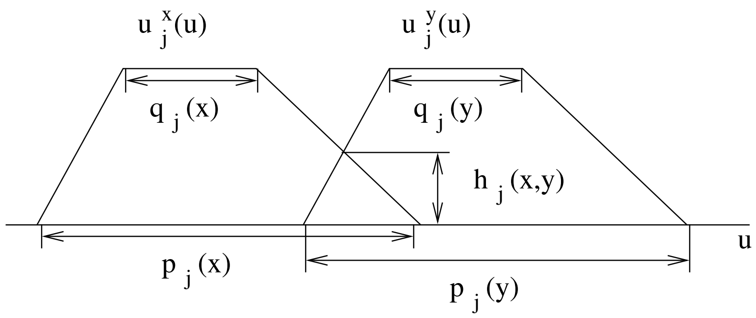

Figure 2.1: Comparing two fuzzy intervals

Starting from the credibility of the preference of x over y it is possible to define [6] [7]:

– the degree of strict preference of x over y as the necessity that x is strictly better than y:

$$P_j(x,y) = 1 - \Pi_j(y \geq x) = 1 - R_j(y,x)$$

– the degree of indifference between x and y as:

$$I_j(x,y) = \min(R_j(x,y), R_j(y,x))$$

- for of a *symmetric trapezoidal number*:

$$R_j(x,y) = \frac{\frac{p_j(x)+p_j(y)}{2} - \min\{g_j(y) - g_j(x), \frac{p_j(x)+p_j(y)}{2}\}}{\frac{p_j(x)+p_j(y)}{2} - \min\{g_j(y) - g_j(x), \frac{q_j(x)+q_j(y)}{2}\}},$$

where $\frac{0}{0}$ should be taken as 0.

If p_j and q_j are linear functions of g_j, then R_j is a fuzzy semiorder and every λ-cut is a crisp semiorder [6] [7].

If $p_j = q_j$, then we obtain a crisp interval order. Let us define $g'_j(x) = g_j(x) - \frac{q_j(x)}{2}$. We then have:

– for the strict preference:

$$\begin{aligned} xP_jy &\iff R_j(x,y) = 1, R_j(y,x) = 0 \\ &\iff g'_j(x) - g'_j(y) > q_j(y) \end{aligned}$$

– for the indifference:

$$\begin{aligned} xI_jy &\iff R_j(x,y) = R_j(y,x) = 1 \\ &\iff |g'_j(x) - g'_j(y)| \leq \min(q_j(x), q_j(y)) \end{aligned}$$

An extra condition of local consistency should be added [30]:

$$g'_j(x) > g'_j(y) \Rightarrow g'_j(x) + q_j(x) \geq g'_j(y) + q_j(y)$$

The criterion function g'_j and the threshold function q_j define a semi-criterion and the structure (P_j, I_j) is a semiorder.

The classical procedures ELECTRE III [29] and PROMETHEE [3] are using this approach based on the intersection of fuzzy sets.

According to Perny [25], the degree of preference of x over y may be considered in very general terms as

$$R_j(x,y) = f_j[g_j(x), \mathcal{N} g_j(y)]$$

where f_j is a non-decreasing function of both arguments, $\mathcal{N}$ is a strong negation and $R_j(x,x) = 1$. Perny proved that such a valued preference relation is a fuzzy semiorder and every λ-cut constitutes a crisp semiorder [25]. As a particular case, we have the concordance index defined by Roy [29]:

$$R_j(x,y) = \frac{p_j(g_j(x)) - \min\{g_j(y) - g_j(x), p_j(g_j(x))\}}{p_j(g_j(x)) - \min\{g_j(y) - g_j(x), q_j(g_j(x))}$$

where p_j and q_j are non-decreasing functions of g_j and correspond respectively to a preference threshold and an indifference threshold. For consistency reasons, $p_j(g_j(x)) \geq q_j(g_j(x))$. The concordance index R_j is meaningful (i.e. is invariant under admissible transformations of g_j) if g_j is defined on an interval scale (admissible transformations are $h_j = r \cdot g_j + s, r > 0$). q_j corresponds to a constant or a proportion of g_j and p_j is expressed as a proportion of g_j [29].

Similarly, according to Perny, we may also define a degree of discredit as

$$D_j(x,y) = h_j[g_j(y), \mathcal{N} g_j(x)]$$

where h_j is a non decreasing function of both arguments, $D_j(x,x) = 0$ and $\min\{R_j(x,y), D_j(x,y)\} = 0$. Under these conditions, D_j is a fuzzy partial order and every λ-cut represents a crisp partial order. As previously, we can consider the particular case of the discordance index defined by Roy [29]:

$$D_j(x,y) = \min\{1, \max\{0, \frac{g_j(y) - g_j(x) - p_j(g_j(x))}{v_j(g_j(x) - p_j(g_j(x))}\}\},$$

where v_j corresponds to a veto threshold which expresses the existence of a discordant point of view that prohibits to accept the idea that x is globally preferred to y.

2.4 Aggregation procedures

2.4.1 Pre-aggregation methods

Let us first consider the methods that propose to merge the marginal information about each pair of alternatives (x, y) in terms of concordance (and possibly discordance) indexes into a global relation that expresses the overall importance of the consensus on the fact that "x is globally not worse than y".

Roy [30] introduces an outranking relation $\mathcal{O}(x, y)$ that corresponds to the "agreement versus discordance" measure linked to the proposition that x is globaly not worse than y. It indicates the importance of the coalition of the points of view that agree with the proposition by taking also into account the discordance.

In general, if v_j represents the relative importance of each point of view j, $(j \in \mathcal{J}, |\mathcal{J}| = n)$, we may consider two aggregation operators M_R and M_D such that

$$R = M_R(R_1, \ldots, R_n; v_1, \ldots, v_n)$$

$$D = M_D(D_1, \ldots, D_n; v_1, \ldots, v_n)$$

M_R is a monotonic function of the first arguments such that $M_R(0, \ldots, 0; v_1, \ldots, v_n) = 0$ and $M_R(1, \ldots, 1; v_1, \ldots, v_n) = 1$. M_D is a monotonic function of the first arguments that should satisfy:

$$(\exists j, j \in \mathcal{J} : D_j(x, y) = 1) \Rightarrow D(x, y) = 1$$

stating that if at least one point of view is totally discordant with the proposition that x is not worse than y, the global discordance should be maximal for that specific pair of alternatives.

Roy considered more specifically in Electre III

- for R the compensative idempotent operator (weighted sum)

$$R(x, y) = \sum_{j \in \mathcal{J}} v_j R_j(x, y), \sum_{j \in \mathcal{J}} v_j = 1$$

- for $1 - D$ the non-discordance index (geometric mean)

$$1 - D(x, y) = \Pi_{j \in \mathcal{J}}(1 - D_j(x, y))^{v_j}.$$

R measures the overall importance of the agreement and D allows to give a bad rating as soon as one important partial evaluation of the discordance is achieved.

Finally the outranking relation is obtained as a combination of concordant and discordant aspects as:

$$\mathcal{O}(x, y) = R(x, y) \cdot [1 - D(x, y)].$$

Most of the existing proposals linked to pre-aggregation methods simply merge the marginal information related to the agreement on the proposal that x is globally not worse than y. They are thus directly liked to the concordance measures $R_j(x, y)$. The subjectivity of the decision maker with respect to the importance of each of the points of view can be used in different ways to obtain a global compromise. We consider here three of these approaches.

- the weighted sum (good items compensate bad ones with respect to different points of view):

$$R = \sum_{j \in \mathcal{J}} v_j R_j, \sum_{j \in \mathcal{J}} v_j = 1$$

- the weighted minimum (the outranking value is high if the partial evaluations are favorable on each of the points of view)

$$R = \min_{j \in \mathcal{J}} \max(1 - v_j, R_j), \max_{j \in \mathcal{J}} v_j = 1$$

- the weighted maximum (the outranking value is high if at least one of the points of view presents a good evaluation)

$$R = \max_{j \in \mathcal{J}} \min(v_j, R_j), \max_{j \in \mathcal{J}} v_j = 1$$

.

Weighted maximum and minimum can be interpreted as weighted medians (see [5]). The interested reader can refer to [12], [6] and [7] for a more elaborate list of aggregators.

2.4.2 Pre-scoring methods

In this type of approach, the implicit assumption that there exists a complete and transitive comparability of the alternatives is made. The most typical example of such methods corresponds to an ordering or a sorting that is based on the weighted sum of the partial scores. The additive representation of the utilities (expressed in terms of the partial scores) however implies preferential independance of the utilities.

One way to avoid this independance condition is to use the Choquet integral [4] as an aggregator.

Let us consider an alternative x:

$$\mathcal{C}_v(S(x)) := \sum_{i=1}^{n} S_{(i)}(x)[v(A_{(i)}) - v(A_{(i+1)})]$$

where v represents a fuzzy measure on $\mathcal{J}$, that is a monotone set function $v : 2^{\mathcal{J}} \to [0,1]$ fulfilling $v(\emptyset) = 0$ and $v(\mathcal{J}) = 1$. This fuzzy measure merely expresses the importance of each subset of points of view. The parentheses used for indices represent a permutation on $\mathcal{J}$ such that

$$S_{(1)}(x) \leq \ldots \leq S_{(n)}(x),$$

and $A_{(i)}$ represents the subset $\{(i), \ldots, (n)\}$.

We note that for additive measures ($v(S \cup T) = v(S) + v(T)$, whenever $S \cap T = \emptyset$) the Choquet integral coincides with the usual discrete Lebesgue integral and the set function v is simply determined by the importance of each point of view: $v(1), \ldots, v(n)$. In this particular case

$$\mathcal{C}_v(S(x)) = \sum_{i=1}^{n} v(i) S_i(x) \;\; (x \in A),$$

which is the natural extension of the Borda score as defined in voting theory if alternatives play the role of candidates and points of view represent voters.

If points of view cannot be considered as being independent, importance of combinations $S \subseteq \mathcal{J}$, $v(S)$ has to be taken into account.

Some combinations of points of view might present a positive interaction or *synergy*. Although the importance of some points of view, members of a combination S, might be low, the importance of a pair, a triple, ..., might be substantially larger and $v(S) > \sum_{i\in S} v(i)$.

In other situations, points of view might exhibit negative interaction or *redundancy*. The union of some points of view do not have much impact on the decision and for such combinations S, $v(S) < \sum_{i\in S} v(i)$. In this perspective the use of the Choquet integral is recommended.

The Choquet integral presents standard properties for aggregation [15],[34],[19]: it is continuous, non decreasing, located between min and max.

The major advantage linked to the use of the Choquet integral derives from the large number of parameters $(2^n - 2)$ associated with a fuzzy measure. On the other hand, this flexibility can also be considered as a serious drawback when assessing real values to the importance of all possible combinations. We will come back to this important question in Section 2.6.

Let v be a fuzzy measure on $\mathcal{J}$. The Möbius transform of v is a set function $m : 2^{\mathcal{J}} \to \mathbb{R}$ defined by

$$m(S) = \sum_{T\subseteq S} (-1)^{|S|-|T|} v(T) \ \ (S \subseteq \mathcal{J}).$$

This transformation is invertible and thus constitutes an equivalent form of a fuzzy measure and v can be recovered from m by using

$$v(S) = \sum_{T\subseteq S} m(T) \ \ (S \subseteq N).$$

This transformation can be used to redefine the Choquet integral without reordering the partial scores:

$$\mathcal{C}_v(S^N(x)) = \sum_{T\subseteq \mathcal{J}} m(T) \bigwedge_{i\in T} S_i^N(x).$$

A fuzzy measure is k-additive ([10]) if its Möbius transform m satisfies $m(S) = 0$ for S such that $|S| > k$ and there exists at least one subest S such that $|S| = k$ and $m(S) \neq \emptyset$.

Thus, k-additive fuzzy measures can be represented by at most $\sum_{j=1}^{k} \binom{n}{j}$ coefficients.

For a k-additive fuzzy measure,

$$\mathcal{C}_v(S^N(x)) = \sum_{\substack{T\subseteq \mathcal{J} \\ |T|\leq k}} m(T) \bigwedge_{j\in T} S_j^N(x).$$

In order to assure boundary and monotonicity conditions imposed on v, the Möbius transform of a k-additive fuzzy measure must satisfy:

$$\begin{cases} m(\emptyset) = 0, \quad \sum\limits_{\substack{T \subseteq \mathcal{J} \\ |T| \leq k}} m(T) = 1 \\ \sum\limits_{\substack{T\,:\, i \in T \subseteq S \\ |T| \leq k}} m(T) \geq 0, \quad \forall S \subseteq \mathcal{J},\ \forall j \in S \end{cases}$$

In Section 2.5 we present a sorting method using the Choquet integral and based on supervised learning.

2.5 The sorting problem

Let A be a set of potential alternatives which are to be assigned to disjoint ordered classes. Let $F = \{g_1, \ldots, g_n\}$ be a set of points of view. For each index of point of view $j \in \mathcal{J} = \{1, \ldots, n\}$, the alternatives are evaluated according to a s_j-point ordinal performance scale represented by a totally ordered set

$$X_j := \{g_1^j \prec_j \ldots \prec_j g_{s_j}^j\}.$$

Therefore, an alternative $a \in A$ can be identified with its corresponding profile

$$(x_1, \ldots, x_n) \in \prod_{j=1}^{n} X_j =: X,$$

where for any $j \in \mathcal{J}$, x_j is the partial evaluation of x on point of view j.

Let us consider a partition of X into m nonempty increasingly ordered classes $\{Cl_t\}_{t=1}^m$. This means that for any $r, s \in \{1, \ldots, m\}$, with $r > s$, the elements of Cl_r are considered as better than the elements of Cl_s.

The sorting problem we are dealing with consists in partitioning the alternatives of A into the classes $\{Cl_t\}_{t=1}^m$.

In Greco *et al.* [11], a very general theorem states that, under a simple condition of monotonicity, a discriminant function can be found which strictly separates the classes $\{Cl_t\}_{t=1}^m$ by thresholds. In Roubens [26] a restriction to the class of n-place Choquet integrals and normalised scores as criteria functions is made. Hereafter we briefly present the sorting procedure derived from this particular case. For details, the interested reader should refer to [26] or [18].

2.6 The classical Tomaso method

The different stages of the Tomaso (**T**echnique for **O**rdinal **M**ulti**a**ttribute **S**orting and **O**rdering) are listed below:

1. Modification of the criteria evaluations into normalised scores;
2. Use of a Choquet integral as a discriminant function;
3. Assessment of fuzzy measures by questionning the DM and by solving a linear constraint satisfaction problem;
4. Calculation of the borders of the classes and assignment of the alternatives to the classes;
5. Analysis of the results (interaction, importance, leave one out, visualisation).

In this Section we roughly present these different elements.

For each point of view $j \in \mathcal{J}$, the order on X_j ($\preceq_j$) can be caracterised by a valuation $R_j : A \times A \to \{0,1\}$ defined by $R_j(x,y) = 1$ if $x_j \succeq y_j$, 0 otherwise. A partial *net score* $S_j : A \to \mathbb{R}$ is defined by

$$S_j(x) := \sum_{y \in A} [R_j(x,y) - R_j(y,x)] \quad (x \in A).$$

If $A = \prod_{j=1}^{n} X_i$ then the net score becomes

$$S_j(x) = q(\frac{2\mathrm{ord}_j(x) - 1}{s_j} - 1) \quad (j \in \mathcal{J}),$$

where $\mathrm{ord}_j : A \to \{1, \ldots, s_j\}$ is a mapping defined by $\mathrm{ord}_j(x) = r \iff x_j = g_r^j$. This formula is used in the software Tomaso . For justifications refer to Section 2.8.

The integer $S_j(x)$ represents the number of times that x is preferred to any other alternative minus the number of times that any other alternative is preferred to x for point of view j. One can show that the partial net scores identify the corresponding partial evaluations. A further step is to normalise the scores so that they range in the unit interval. We define normalised partial scores S_j^N by

$$S_j^N(x) := \frac{S_j(x) + (q-1)}{2(q-1)} \in [0,1] \quad (j \in \mathcal{J}).$$

On contrary of the partial evaluations, the partial scores and the normalised partial scores are commensurable. Throughout the chapter we will use the notation $S^N(x) := (S_1^N(x), \ldots, S_n^N(x))$.

The next step concerns the aggregation of the normalised partial scores of a given alternative x by means of a Choquet integral [4] which allows to deal with interacting (depending) points of view, according to what has been said in Section 2.4:

$$\mathcal{C}_v(S^N(x)) := \sum_{j=1}^{n} S_{(j)}^N(x)[v(A_{(j)}) - v(A_{(j+1)})]$$

where v is a fuzzy measure on $\mathcal{J}$; that is a monotone set function $v : 2^{\mathcal{J}} \to [0,1]$ fulfilling $v(\emptyset) = 0$ and $v(\mathcal{J}) = 1$. The parentheses used for indices stand for a permutation on $\mathcal{J}$ such that

$$S_{(1)}^N(x) \leq \ldots \leq S_{(n)}^N(x),$$

and for any $j \in \mathcal{J}$, $A_{(j)}$ represents the subset $\{(j), \ldots, (n)\}$. The characterisation of the Choquet integral by Marichal ([13], [15]) clearly justifies the way the partial scores are aggregated.

The objective of this method is to assess the fuzzy measures in order to classify the alternatives of A. Practically, the assessment of the fuzzy measures is done by asking the DM to provide a set of prototypes $P \subseteq A$ and their assignments to the given classes; that is a partition of P into prototypic classes $\{P_t\}_t^m = 1$ where $P_t := P \cap Cl_t$ for $t \in \{1, \ldots, m\}$. As the Choquet integral is supposed to strictly separate the classes Cl_t, the following necessary condition is imposed

$$\mathcal{C}_v(S^N(x)) - \mathcal{C}_v(S^N(x')) \geq \varepsilon \qquad (2.1)$$

for each ordered pair $(x, x') \in P_t \times P_{t-1}$ and each $t \in \{2, \dots, m\}$, where ε is a given strictly positive thershold.

Due to the increasing monotonicity of the Choquet integral, the number of separation constraints 2.1 can be reduced significantly. Thus, it is enough to consider *border elements* of the classes. To formalise this concept, we first define a dominance relation D (partial order) on X by

$$xDy \iff x_j \succeq_j y_j, \text{ for all } j \in \mathcal{J}.$$

As *upper border* of the prototypic class P_t we use the the set of non-dominated alternatives of P_t defined by

$$ND_t := \{x \in P_t \text{ such that } \nexists x' \in P_t \setminus \{x\} : x'Dx\}.$$

Similarly, the *lower border* of the prototypic class is given by the set of non-dominating alternatives of P_t which is defined by

$$Nd_t := \{x \in P_t \text{ such that } \nexists x' \in P_t \setminus \{x\} : xDx'\}.$$

The separation conditions restricted to the prototypes of the subsets $ND_t \cup Nd_t$, $t \in \{1, \dots, m\}$ put together with the monotonicity constraints on the fuzzy measure, form a linear program [17] whose unknowns are the capacities $v(S)$, $S \subset \mathcal{J}$ and where ε is a non-negative variable to be maximised in order to deliver well separated classes.

We use the principle of parsimony for the resolution of this problem. If there exists a k-additive fuzzy measure v^*, k being kept as low as possible, then we determine the boundaries of the classes as follows:

- lowr boundary of Cl_t: $z(t) := \min_{x \in Nd_t} \mathcal{C}_{v^*}(S^N(x))$;
- upper boundary of Cl_t: $Z(t) := \max_{x \in ND_t} \mathcal{C}_{v^*}(S^N(x))$.

At this point, any alternative $x \in A$ can be classified in the following way:

- x is assigned to class Cl_t if $z_t \le \mathcal{C}_{v^*}(S^N(x)) \le Z_t$;
- x is assigned to class $Cl_t \cup Cl_{t-1}$ if $Z_{t-1} < \mathcal{C}_{v^*}(S^N(x)) < z_t$.

A final step of the classical TOMASO method concerns the evaluation of the results and the interpretation of the behaviour of the Choquet integral. The meaning of the values $v(T)$ is not clear to the DM. They don't immediatly indicate the global importance of the points of view, nor their degree of interaction. It is possible to derive some indices from the fuzzy measure which are helpful to interpret its behaviour. Among them, the TOMASO method proposes to have a closer look at the importance indices [32] and the interaction indices [21]. We present the calculation of these indices in Section 2.7.1.

2.7 An evolution to the basic TOMASO method

It may happen that the linear program described in Section 2.6 has no solution. This occurs when the prototypic elements violate the axioms that are imposed to produce a discriminant function of Choquet type ([15] [34]), in particular the triple cancellation axiom. In such a case, and in order to present a solution to the DM, we suggest to find a fuzzy measure by solving the following quadratic program

$$\min_{x \in \cup_{t \in \{1,\ldots,m\}}\{ND_t \cup ND_t\}} \sum_x [\mathcal{C}_v(S^N(x)) - y(x)]^2,$$

where the unkowns are

- the capacities $v(S)$ which determine the fuzzy measure;
- some global evaluations $y(x)$ for each $x \in \cup_{t \in \{1,\ldots,m\}}\{ND_t \cup ND_t\}$.

The capacities $v(S)$ are constrained by the monotonicity conditions (as previously shown in Section 2.6). The global evaluations $y(x)$ must verify the classification imposed by the DM. In other words, for every ordered pair $(x, x') \in Nd_t \times ND_{t-1}$, $t \in \{2, \ldots, m\}$ the condition $y(x) - y(x') \geq \varepsilon'$, $\varepsilon' > 0$ must be satisfied.

Unlike the method described in Section 2.6, in this case, ε' plays the role of a parameter, which needs to be fixed by the DM. As previously, we use the principle of parsimony when searching for a solution (keep k as low as possible; at worst k equals the number of points of view). We show in

Section 2.9 that the choice of the pair (k, ε) can have a non negligible influence on the final solution.

Similarly to the classical method, the next step is to determine the boundaries of the classes. The temporary intervals $I_s^t := \{z_s^t, Z_s^t\}$ determine an interval order where $I_k \cap I_l,\ k \neq l$ is not necessarily verified. Besides, the property to provide a semiorder $(Z_m \geq \ldots \geq Z_1; z_m \geq \ldots \geq z_1)$ might also be violated. In order to obtain a valid semiorder, the prototypic elements which induce the violation are considered as *ill-classified* with respect to the Choquet methodology and removed for the final determination of the class boundaries. The procedure to eliminate the ill-classified alternatives can be summarised as follows:

- consider the upper boundaries of the intervals I_s^t, $Z_m, \ldots, Z_1$, ordered according to the indexes of the classes, from the best to the worst class.
- for $i = m$ to 2 do
  - if $Z_{i-1} > Z_i$ then
    - mark all $a \in P_{i-1}$ such that $\mathcal{C}_v(a) > Z_i$ as ill-classified;
    - $Z_{i-1} = Z_i$.

A similar procedure is applied to mark the ill-classified prototypes in relation with the lower bounds of the temporary intervals. Next, the marked alternatives are not considered for the determination of the boundaries $I_s := \{z_s, Z_s\}$ of the classes which induce a semiorder. A schematic representation of this procedure is provided on figure 2.2.

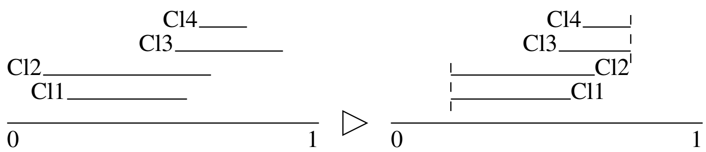

Figure 2.2: Schematic representation of the cutting procedure

Finally, the remaining prototypes are assigned to the classes according to the previously determined boundaries. Four scenarios may occur to such a prototype a:

- a is assigned to a single class and is correctly classified (according to the DM's assignment) (well defined and correct assignment) (wdc);
- a is assigned to a single class but is not correctly classified (according to the DM's assignment) (well defined and erronous assignment) (wde);
- a is assigned to the union of contiguous classes and one of these classes corresponds to the DM's assignment (ambiguously defined and correct assignment) (adc);
- a is assigned to the union of contiguous classes and none of these classes corresponds to the DM's assignment (ambiguously defined and wrong assignment) (adw).

We call the *degree* of an assignment the number of classes to which an alternative belongs. The degree of ill-classified prototypes equals 0. Let us remark that the assignment of the ill-classified prototypes is necessarly erronous. Similarly, the assignment of the non-prototypic elements of A is either well or ambiguously defined.

The quality of a model (classifier) depends on different ratios. A good model has the following *natural* properties:

- a simple model according to parsimony (low k);
- a high wdc over $|P|$ ratio;
- a low wde over $|P|$ ratio;
- a low adc over $|P|$ ratio OR adcs with low degrees;
- a low adw over $|P|$ ratio;
- a low ill-classified over $|P|$ ratio.

In order to obtain such a model the DM has to deal with the pair of parameters (k, ε'). We will show in Section 2.9 how the values of the pair (k, ε') can influence the quality of the model. Finally, the analysis of the interaction and the importance indices can give some indications on the behaviour of the fuzzy measure.

2.7.1 Behavioral analysis of aggregation

Now that we have a sorting model for assigning alternatives to classes (based on the linear program or the quadratic program), an important question arises: How can we interpret the behavior of the Choquet integral or that of its associated fuzzy measure? Of course the meaning of the values $v(T)$ is not always clear for the DM. These values do not give immediately the global importance of the points of view, nor the degree of interaction among them.

In fact, from a given fuzzy measure, it is possible to derive some indices or parameters that will enable us to interpret the behavior of the fuzzy measure. These indices constitute a kind of *id card* of the fuzzy measure. In this Section, we present two types of indices: importance and interaction. Other indices, such as tolerance and dispersion, were proposed and studied by Marichal [13, 16].

2.7.1.1 Importance indices

The overall importance of a point of view $j \in \mathcal{J}$ in a decision problem is not solely determined by the value of $v(\{j\})$, but also by all $v(T)$ such that $j \in T$. Indeed, we may have $v(\{j\}) = 0$, suggesting a priori that element j is unimportant, but it may happen that for many subsets $T \subseteq \mathcal{J}$, $v(T \cup \{j\})$ is much greater than $v(T)$, suggesting that j is actually an important element in the decision.

Shapley [32] proposed in 1953 a definition of a coefficient of importance, based on a set of reasonable axioms. The *importance index* or *Shapley value* of point of view j with respect to v is defined by:

$$\phi(v, \{j\}) := \sum_{T \subseteq \mathcal{J} \setminus \{j\}} \frac{(n - |T| - 1)!\, |T|!}{n!} \left[v(T \cup \{j\}) - v(T)\right]. \tag{2.2}$$

This index is a fundamental concept in game theory and it expresses a power index. It can be interpreted as a weighted average value of the marginal contribution $v(T \cup \{j\}) - v(T)$ of element j alone in all combinations. To make this clearer, we rewrite the index as follows:

$$\phi(v,\{j\}) = \frac{1}{n}\sum_{t=0}^{n-1}\frac{1}{\binom{n-1}{t}}\sum_{\substack{T\subseteq\mathcal{J}\setminus\{j\}\\|T|=t}}[v(T\cup\{j\})-v(T)].$$

Thus, the average value of $v(T\cup\{j\})-v(T)$ is computed first over the subsets of same size t and then over all the possible sizes. Consequently, the subsets containing about $n/2$ points of view are the less important in the average, since they are numerous and a same point of view j is very often involved into them.

The use of the Shapley value in multicriteria decision making was proposed in 1992 by Murofushi [21]. It is worth noting that a basic property of the Shapley value is

$$\sum_{j=1}^{n}\phi(v,\{j\}) = 1.$$

Note also that, when v is additive, we clearly have $v(T\cup\{j\})-v(T) = v(\{j\})$ for all $j\in\mathcal{J}$ and all $T\subseteq\mathcal{J}\setminus\{j\}$, and hence

$$\phi(v,\{j\}) = v(\{j\}),\quad j\in\mathcal{J}. \tag{2.3}$$

If v is non-additive then some points of view are dependent and (2.3) generally does not hold anymore. This shows that it is useful to search for a coefficient of overall importance for each point of view.

2.7.1.2 Interaction indices

A further interesting concept is that of *interaction* among points of view. We have seen that when the fuzzy measure is not additive then some points of view interact. Of course, it would be interesting to appraise the degree of interaction among any subset of points of view.

Consider first a pair $\{i,j\}\subseteq\mathcal{J}$ of points of view. It may happen that $v(\{i\})$ and $v(\{j\})$ are small and at the same time $v(\{i,j\})$ is large. The Shapley index $\phi(v,\{j\})$ merely measures the average contribution that point of view j brings to all possible combinations, but it gives no information on the phenomena of interaction existing among points of view.

Clearly, if the marginal contribution of j to every combination of points of view that contains i is greater (resp. less) than the marginal contribution of j to the same combination when i is excluded, the expression

$$[v(T \cup \{i,j\}) - v(T \cup \{i\})] - [v(T \cup \{j\}) - v(T)]$$

is positive (resp. negative) for any $T \subseteq \mathcal{J} \setminus \{i,j\}$. We then say that i and j positively (resp. negatively) interact.

This latter expression is called the *marginal interaction* between i and j, conditioned to the presence of elements of the combination $T \subseteq \mathcal{J} \setminus \{i,j\}$. Now, an interaction index for $\{i,j\}$ is given by an average value of this marginal interaction. Murofushi and Soneda [21] proposed in 1993 to calculate this average value as for the Shapley value. Setting

$$(\Delta_{ij}\, v)(T) := v(T \cup \{i,j\}) - v(T \cup \{i\}) - v(T \cup \{j\}) + v(T),$$

the *interaction index* of points of view i and j related to v is then defined by

$$I(v,\{i,j\}) := \sum_{T \subseteq N \setminus \{i,j\}} \frac{(n-|T|-2)!\,|T|!}{(n-1)!}\,(\Delta_{ij}\, v)(T). \tag{2.4}$$

It should be mentioned that, historically, the interaction index (2.4) was first introduced in 1972 by Owen (see Eq. (28) in [24]) in game theory to express a degree of complementarity or competitiveness between elements i and j.

2.7.2 Interpretation of the behaviour of the fuzzy measure

In this Section we briefly show the main advantage to use a Choquet integral rather than the weighted sum as a discriminant function. We therefore take the simple case of two points of view, which can be represented in a plane. Figure 2.3 presents 5 possible ranges of values for the weights v and the corresponding structures of the limits of the classes. One can see that the main difference between the classical weighted sum and the Choquet integral is the greater flexibility of the borders of the classes. The Choquet integral creates piecewise linear borders, which allows to build more precise classes. The different possibilities are summarised by the following list:

- I: $v(1) + v(2) < v(12)$: synergy
- II: $v(1) + v(2) > v(12)$: redundancy
- III: $v(1) + v(2) = v(12) = 1$: additivity
- IV: $v(1) = v(2) = 0$: limit case
- V: $v(1) = v(2) = 1$: limit case

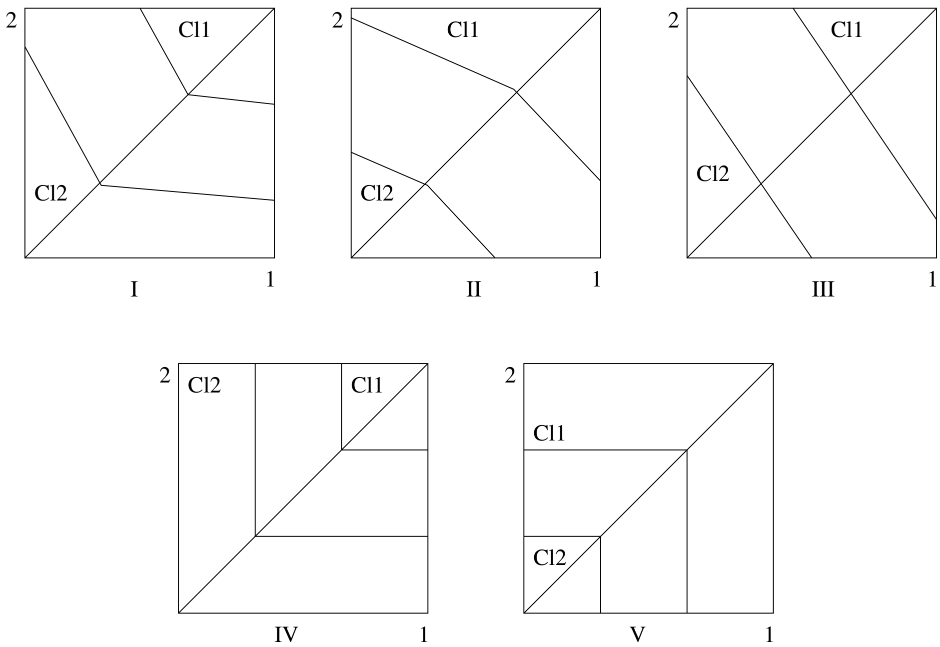

Figure 2.3: Interpretation of the discriminant functions

In [8] the authors give an interpretation to the first two cases. In case of synergy, although the importance of a single criterion for the decision is rather low, the importance of the pair is large. The criteria are said to be *complementary*. In case of redundance, or negative synergy, the union of criteria does not bring much, and the importance of the pair might be roughly the same as the importance of a single criterion.

The limit case (IV) occurs for maximal synergy. In that case, the Choquet integral corresponds to the aggregation by the min function. Maximal redundancy occurs for case (V), where the Choquet integral is the max function.

In case the number of points of view is larger than two, it becomes quite hard to represent the problem. Nevertheless, the previous short example helps to understand how the borders of the classes are built in such more general examples.

2.8 The software TOMASO

This part of the chapter is devoted to the software TOMASO . The software can be downloaded on http://patrickmeyer.tripod.com. It is an implementation of the algorithms which were presented previously. Its name stands for "Tool for Ordinal MultiAttribute Sorting and Ordering". It is written in Visual Basic. We briefly describe the various aspects of the software.

The software TOMASO can be used for supervised classification of ordinal multi-criteria data. In order to work properly, it requires some information about the structure of the data. This information is given by the protoypes, which are a set of objects well known by the DM and used to build the classifier.

When using the software, at first the user has to load the file with the data he wants to analyze (see tutorial on http://patrickmeyer.tripod.com for a detailed description of the data file formats). These data have to be ordinal, and can be composed of the prototype and some alternatives that need to be classified.

In order to be consistent, the method calculates the partial net score (as defined in Section 2.6) of each alternative for each criterion relatively to all possible alternatives. This way, we avoid that the score of an alternative for a particular criterion changes if we add some new alternatives to the set of prototypes. Furthermore, the method gains in stability. To help the software calculate these scores, the user has to define the range of each criterion. We suppose that the minimum value for each criterion is 1 and that the maximum value n_j (for each criterion j) has to be fixed by the user. This means that at this level, the software supposes that the range of criterion j goes from 1 to n_j, with n_j different ordinal rating benchmarks.

After the calculation of the partial net scores, the user has to fix the number of classes and their structure. This last point is achieved by adding the alternatives of the prototype to their respective class. It can be done by loading a file, or by adding manually the objects to the different classes. At the end of this step, the alternatives which are not assigned to any class do not belong to the set of prototypes and will either be used as a test subset (to check the validity of the method) or simply as objects to be classified.

The next stage of the software is to check the consistency of the assignments within the set of prototypes. The inconsistencies can be of two types:

- two alternatives x and y with the same partial net scores for each criterion ($S^N(x) = S^N(y)$) belong to two different classes
- two alternatives x and y so that x dominates y (i.e. $S_j^N(x) \geq S_j^N(y), \forall j \in \mathcal{J}$) and y belongs to a better class than x

If no inconsistency is detected, the software goes on. Else, the user has to change the definition of the classes by the prototypes (either manually, or by loading another class file). The next step is the determination of the non-dominating set of alternatives and the non-dominated set of alternatives in each class. These sets are not empty, because the user should add at least one alternative in each class. As already mentioned earlier, this is meant to reduce the number of separation constraints and is justified by the increasing monotonicity of the Choquet integral.

The resolution of the linear constraint satisfaction problem is the next stage. The user can chose between fixing the value of $k < n$ or solving the problem for $k = n$. In both cases, a file of constraints is computed. An external solver (lp_solve 3.0, ftp://ftp.ics.ele.tue.nl/pub/lp_solve/, released under the LGPL license) uses this file as an input and tries to solve the problem.

If the problem has no solution for $k = n$, this is due to an incompatibility between the assignments of the alternatives of the prototype and the assumption that the discriminant function is a Choquet integral. In this case, the user should either revise the definition of the classes or chose the alternate way, by solving the quadratic problem.

On the other hand, if there is a solution for the problem, the following results are displayed:

- the values the fuzzy measure v takes on all subsets of N
- the values of the Möbius transform m of v
- the value of ε
- the borders of the classes (for each class Cl_t, given by the maximal and the minimal value of the Choquet integral of the alternatives of Cl_t)
- the importance indices (or Shapley values) for each point of view i
- the interaction indices (for each subset of points of view)
- the Choquet integral of each alternative of the prototype

An important tool is the assignment of the alternatives which are not in the prototype to the classes, according to the previously built model. During this step, it may happen that the Choquet integral of some of these objects may not be between the limits of any of the classes. In that case, those alternatives are assigned ambiguously to the union of the neighbour classes. If for one particular alternative x the decision maker is not satisfied with such an ambiguous assignment to classes Cl_r or Cl_s, he may revise the definition of the adjacent classes in order to include x or maybe another similar alternative to x. Else, one can say that alternative x belongs to class Cl_r or Cl_s.

To check the structure of the prototype, it is possible to apply a particular leave-one-out procedure to the data. For each alternative x of the prototype, the model is rebuilt without the alternative. Then x is assigned to a class, according to the prototype. This class should then be the original class of x. If not, it is considered as an error. At the end of the whole procedure, a high error ratio, say $e/|P|$, with e close to $|P|$ (e is the number of badly classified alternatives during the leave-one-out procedure) does not necessarily mean that the data of the prototype are badly chosen. It either stands for a "minimal" prototype, where nearly each alternative is important for the building of the model, or for a quite complex data structure. In this latter case, the prototype should perhaps be revised and enriched with new alternatives to increase its diversity.

In case no solution can be found to the linear problem, the alternate option is to solve the quadratic program. In this case, the user has to fix two parameters: the value of ε' and the value of k. We recommend to fix first the value of ε' and to search a solution for $k = n$. If a solution is found, the value of k can then be decreased to search for a k-additive model. For any solution, the following information is provided:

- the values the fuzzy measure v takes on all subsets of N;
- the values of the global evaluations y;
- the values of the Möbius transform m of v;
- the value of ε';
- the temporary borders of the classes, the final borders of the classes, and the borders given by the global evaluations y;
- the importance indices (or Shapley values) for each point of view i;
- the interaction indices (for each subset of points of view);
- the Choquet integral of each of the prototypes.

In a further step, the user can view the quality of the classifier by analysing the values of the different ratios

- the well defined and correct ratio: wdc over $|P|$ (should be high);
- the well defined and erronous ratio: wde over $|P|$ (should be low);
- the ambiguously defined and correct ratio and the assignment degrees: adc over $|P|$ (should be low, and low degrees);
- the ambiguously defined and erronous ratio: adw over $|P|$ ratio (should be low);
- the ill-classified ratio: ill-classified over $|P|$ (should be low).

If these ratios are not satisfying the DM's preferences, the search for a new fuzzy measure can be repeated with different values for the parameters k and ε'.

A final step consists in viewing the assignments of the prototypes and the non-prototypic alternatives.

2.9 Testing the method on two problems

In this Section, we apply the previously presented method on two particular problems. In a first step we will solve a noise annoyance problem by the classical method. As no solution can be found to the original problem, we have to consider a subset of alternatives as prototypes. The alternate way is then presented on two examples. The first example concerns a theoretical problem of 8 students which need to be assigned to 6 graded classes (according to two criteria). The second problem concerns again the real-life example about noise annoyance [33] and [2].

2.9.1 The noise annoyance problem: A subset

This real-life example concerns noise annoyance caused by different sources. Details on these data can be found in [2] and [33]. It was obtained by a survey performed on 2661 persons (alternatives). They were asked to give an estimation of their annoyance level (not at all annoyed $\preceq$ slightly annoyed $\preceq$ moderately annoyed $\preceq$ very annoyed $\preceq$ extremely annoyed) on 21 different potential noise sources (criteria) (noise annoyance caused by road traffic, by rail traffic, ...). This ordering and the exact wording of the questions is in accordance with international standards. Besides, the questionned persons had to give an overall noise annoyance level on the same scale.

The original dataset contains 2661 alternatives and 21 criteria. But unfortunately, its structure is not proper for the Tomaso method as it contains a lot of inconsistencies (aDb but a is in a worse class than b). Therefore, we restrict ourselves to a consistent subset of 155 alternatives and 6 criteria for this chapter (road traffic (cars, busses,...), air traffic, truck loading and unloading, factories, dance halls, agricultural equipment).

For this first analysis with the classical method (solving the linear problem), no solution exists. We therefore search for a subset of alternatives which does not violate the triple cancellation condition. This subset is clearly not maximal; in fact, we determined it by a simple greedy heuristic.

This reduced dataset now contains 142 alternatives described by 6 criteria. Originally, each alternative belongs to one of 5 graded classes (not at all annoyed $\preceq$ slightly annoyed $\preceq$ moderately annoyed $\preceq$ very annoyed $\preceq$ extremely annoyed) which represent the global evaluations on the general noise annoyance. To simplify the search for a valid subset of alternatives, we simplify these assignments by considering the union of the 3 upper classes: moderately annoyed, very annoyed and extremely annoyed. This means that the goal of this part of the chapter is to find a discriminant function based on a Choquet integral which allows to assign the alternatives to 3 graded classes: not at all annoyed $\preceq$ slightly annoyed $\preceq$ at least moderately annoyed.

A solution to the linear program can be found for $k = 2$ (no solution can be found for the additive model). This means that for a 2-additive model, 100% of the prototypes are assigned correctly. The remaining 13 non prototypic alternatives (but which belong to predefined classes, according to the survey), are assigned erronously in 62% of the cases (8 out of 13). The global correct assignment ratio over the 155 alternatives is 95%.

The importance indexes are given in table 2.1.

Table 2.1: Importance indexes

3 classes, $k = 2$					
street	air	truck	factory	dancing	agriculture
.214	.214	.143	.143	.268	0

The interaction indexes are mostly zero, except for the following couples of criteria: (street, air) = $-.145$; (street, truck) = $-.286$; (air, factory) = $-.286$.

2.9.2 Assigning students to graded classes

This small example clearly illustrates the procedure when no solution can be found to the linear program. We consider a set of 8 students evaluated on 2 courses (C1, C2). For each matter, the evaluation scale has 10 ordered qualitative levels (1-10). In total, this makes 100 possible different ratings. Besides, for each student, the DM has given a global evaluation on a 6-levelled qualitative ordinal scale (the classes): (very good (6)> good (5)> above average (4)> below average (3)> bad (2)> very bad(1)). A summary of the problem is given in table 2.2.

Table 2.2: Profiles of the students

student	profile	net score	class
A	$(7, 5)$	$(.65, .45)$	2
B	$(6, 6)$	$(.55, .55)$	1
C	$(7, 7)$	$(.65, .65)$	3
D	$(6, 8)$	$(.55, .75)$	4
A'	$(10, 7)$	$(.95, .65)$	6
B'	$(8, 8)$	$(.75, .75)$	5
C'	$(10, 5)$	$(.95, .45)$	3
D'	$(8, 6)$	$(.75, .55)$	4

A representation of the 2-dimensional problem is given on figure 2.4. It helps us to understand why the linear problem has no solution. If the triple cancellation property [34] is violated, there exists no Choquet integral which satisfies the constraints imposed by the classification of the prototypes. If triple cancellation was verified in this example, we would have:

$$\left\{\begin{array}{c} \textsc{Class}(B) \leq \textsc{Class}(A) \\ \& \\ \textsc{Class}(D) \geq \textsc{Class}(C) \\ \& \\ \textsc{Class}(D') \geq \textsc{Class}(C') \end{array}\right\} \Rightarrow \textsc{Class}(B') \geq \textsc{Class}(A'),$$

where $\textsc{Class}(X)$ stands for the index of the class to which X belongs (the higher the better). But in this particular example, we clearly have $\textsc{Class}(D') \leq \textsc{Class}(C')$. Therefore, no solution can be found to the linear program. In other words, this problem cannot be described by the classical TOMASO method by means of a Choquet integral. We therefore suggest to go over to the method which tries to find a solution as close as possible to the *ideal* solution.

In this case, a solution can be found for different values of ε and k. Table 2.3 summarises some of these results.

Let us analyse graphically the behaviour of the solution in the particular example where $\varepsilon = .10$ and $k = 1$. In that case, the Choquet integral becomes a classical weighted sum. The Möbius transform of the fuzzy measure v are given by $m(C1) = .45, m(C2) = .55$. We can then draw the limits of the final classes (after dealing with the 3 ill-classified prototypes, which are not

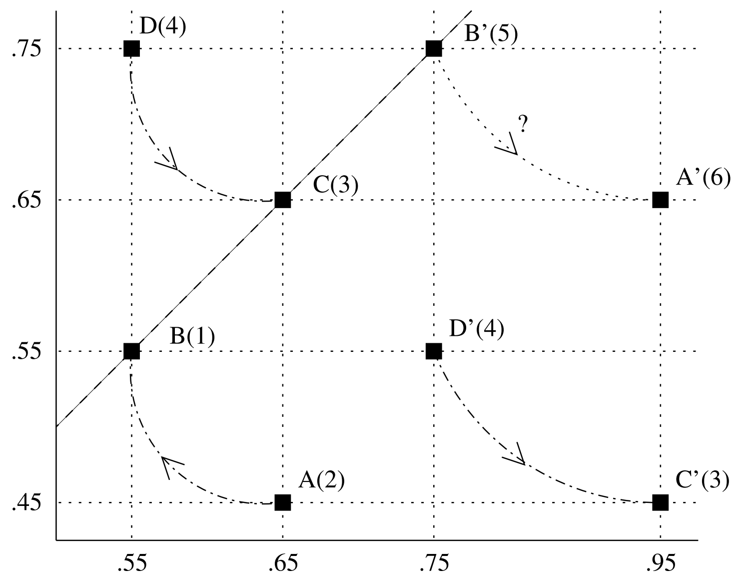

Figure 2.4: Representation of the students problem

considered for the calculation of these boundaries) as shown on figure 2.5. Segment 1 is the lower boundary of class 6. Segment 2 is the lower and the upper boundary of class 5. Segment 3 and 4 are the upper and lower boundaries of the union of classes 3 and 4. Segment 5 is the lower and the upper boundary of class 2, and the upper boundary of class 1. The 3 ill-classified elements are represented by circles. We observe that the 3 ambiguously assigned alternatives are B, C and D.

We observe that both solutions for $\varepsilon = .10, k = 2$ and $\varepsilon = .12, k = 2$

Table 2.3: Profiles of the students

ε	k	ill	wdc	wde	adc2	adw2
.01	2	2	3	0	3	0
.01	1	2	3	0	3	0
.06	2	2	3	0	3	0
.06	1	2	3	0	3	0
.10	2	0	6	0	2	0
.10	1	3	2	0	3	0
.12	2	0	6	0	2	0
.12	1	3	2	0	3	0

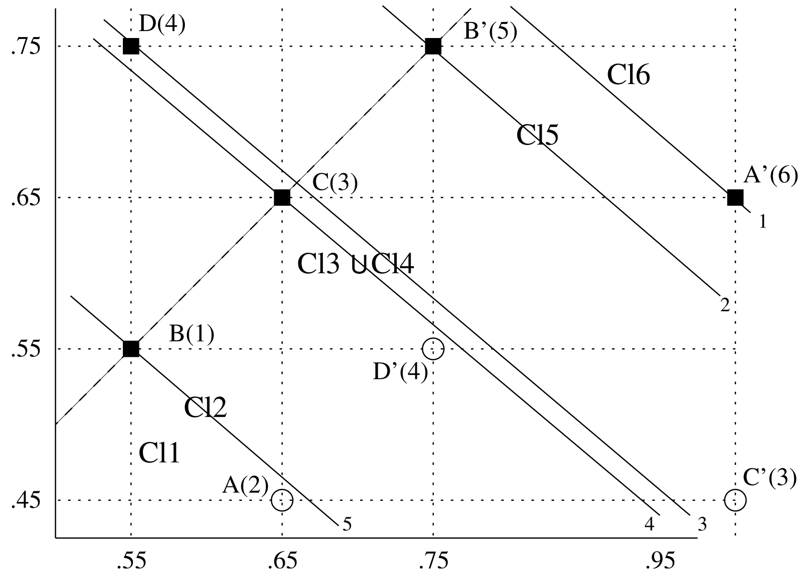

Figure 2.5: Classes for $\varepsilon = .1, k = 1$

are good compromises, according to the criteria defined in Section 2.7. The additional information on these scenarios is given in the following table.

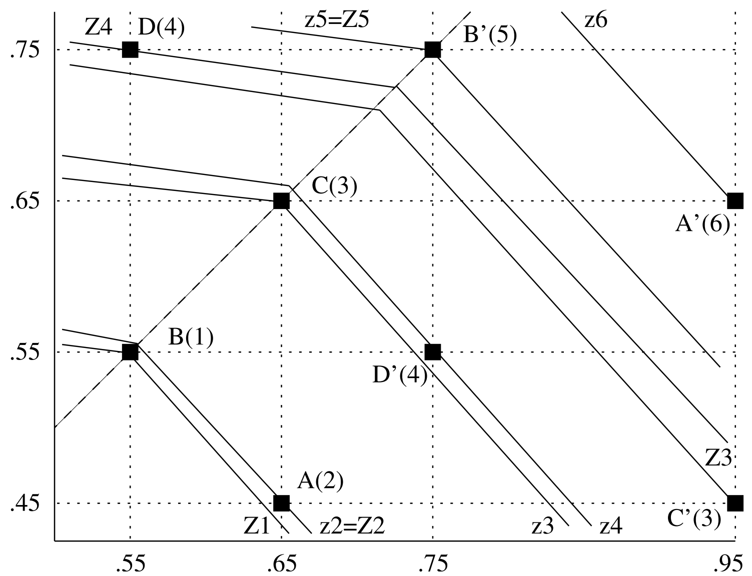

Figure 2.6: Classes for $\varepsilon = .1, k = 1$

Figure 2.6 shows the classes for the particular example where $\varepsilon = .10$ and

$k = 2$. We observe that the borders of the classes are piecewise linear, and that this allows to cope with a larger set of problems. We can also observe the overlapping zone between classes 3 and 4, which induce the ambiguous assignments of C' and D'.

2.9.3 The noise annoyance problem

Let us get back to the problem described in 2.9.1. We will now consider the 155 prototypes, without searching for a subset which does not violate the triple cancellation axiom. Two different analyses are preformed. A first one with the original 5 classes, and a second one with 3 classes.

2.9.3.1 Original 5 classes

The classical TOMASO method provides no solution to the linear problem. We therefore chose the alternate procedure by solving the quadratic program. Here, solutions exist for different values of ε and k. We summarise these results in table 2.4.

Table 2.4: Results for 5 classes

5 classes						
ε	k	ill	wdc	wde	adc2	adc3
.18	6, 5	0	64	0	88	3
.19	6, 5	0	61	0	85	9
.20	6, 5	0	60	0	79	16
.22	6	0	55	0	75	30
.23	6, 5	0	51	0	72	32
.26	6	0	47	0	71	37
.28	6	0	27	0	88	40

We clearly see that the solution for $\varepsilon = .18$ and $k = 5$ is a good compromise according to the criteria defined in Section 2.7. In this particular case, the boundaries of the classes are given in the table 2.5. We observe that there are overlapping zones in this classification (of degree 2 and 3). This can also be clearly seen on figure 2.7.

Table 2.5: Boundaries of the classes

5 classes		
class	min	max
5	.580	.747
4	.482	.728
3	.345	.583
2	.143	.390
1	.100	.241

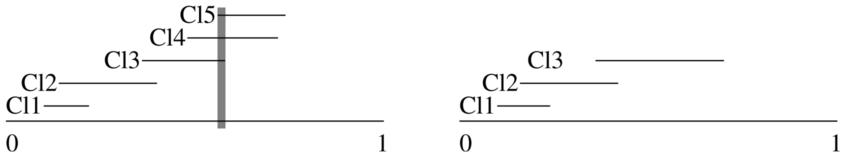

Figure 2.7: 5 classes, $\varepsilon = .18, k = 5$; 3 classes, $\varepsilon = .46, k = 5$

The number of well-defined and correct assignment remains quite low (41.3% of the prototypes), but fortunately we have only 1.9% of ambiguous assignments of degree 3, and no erronous assignment or ill-classified alternatives.

The importance indices of the 6 criteria are given in table 2.6 for a sample of models.

Table 2.6: Importance indexes

5 classes						
model	street	air	truc	factory	dancing	agriculture
$\varepsilon = .18, k = 6$	.296	.189	.070	.131	.262	.079
$\varepsilon = .19, k = 5$	.263	.194	.062	.136	.269	.075
$\varepsilon = .22, k = 6$	.282	.200	.062	.131	.283	.043
$\varepsilon = .28, k = 6$	.316	.217	.047	.110	.298	.013

We can see that these values remain quite identical from one model to another (at least, the order on the importance of the criteria is maintained). An average value of these indices is represented on figure 2.8. We can see that the most important noise annoyance comes from the street noises, followed

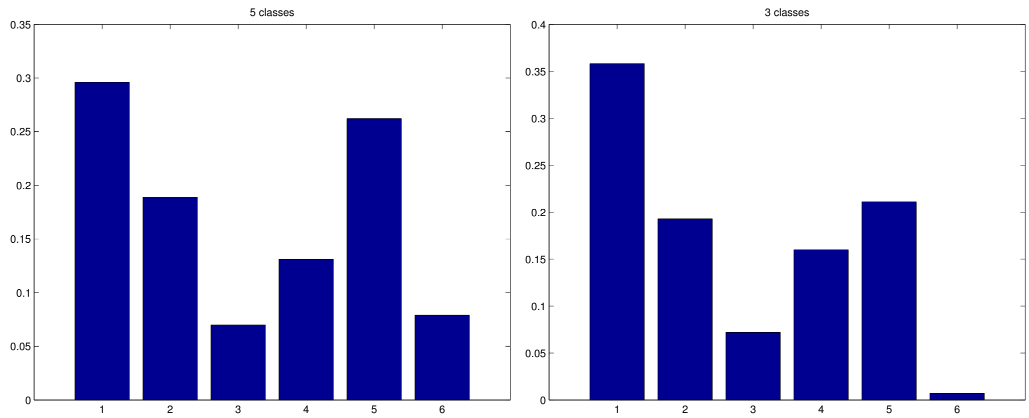

Figure 2.8: Av. values of importance indices: 5 classes; 3 classes (1: street, 2: air, 3: truc, 4: factory, 5: dancing, 6: agriculture)

closely by the dancing halls annoyances, then the airtraffic annoyances, the factory annoyances, the agrucultural annoyances and finally the truck loading annoyances.

In order to improve this classification, we decide to consider the union of the first three classes as a single class. The following Section resumes these considerations.

2.9.3.2 3 classes

The results of different models we tested on this modified problem are summarised in table 2.7.

Here, the solution for $\varepsilon = .47$ and $k = 4$ seems a good compromise. The ratio of well-defined and correct assignments over the number of prototypes has risen to 85.15%, the ambiguous assignments of degree 3 have been removed, and there are no ill-classified alternatives. The boundaries of the classes are given in the table 2.8 and can be visualised on figure 2.7. The importance indices for a sample of models are listed in table 2.9. Once more, we observe no major changes from one model to another. An average value for the importance indices is represented on figure 2.8. We observe that the importance indexes look similar to the 5 classes problem (excepted an inversion on agriculture and truck loading). This is interesting, as it shows a

Table 2.7: Results for 3 classes

3 classes						
ε	k	ill	wdc	wde	adc2	adc3
.4	6	0	131	0	24	0
.43	6	0	129	0	26	0
.44	6, 5	0	129	0	26	0
.45	6	0	129	0	26	0
.46	6	0	132	0	23	0
.46	5,4	0	132	0	23	0
.52	6	0	132	0	23	0
.52	4	0	130	0	25	0
.55	6	0	131	0	24	0
.55	4	0	130	0	25	0

Table 2.8: Boundaries of the classes

3 classes		
class	min	max
3	.355	.700
2	.156	.419
1	.100	.236

certain stability of the method.

2.10 The Choice Problem

In this Section we consider a way to select a subset of alternatives to consider as a good choice.

Consider a binary relation R whose credibility is evaluated as follows:

$R(x,y) = C_v[R_1(x,y), \ldots, R_i(x,y), \ldots, R_n(x,y)] \in [0,1]$, for all $x, y \in A$. In the sequel we will only use the ordering of $R(x,y)$ and not their cardinality and we will obtain a L-valued binary relation R (see [1]).

For all $x, y \in A$, $R(x,y)$ belongs to a finite set $L = \{c_0 = 0, c_1, \ldots, c_m = .5, \ldots, c_{2m} = 1$ that constitutes a $(2m+1)$-element chain $c_0 \prec \ldots \prec c_{2m}$.

Table 2.9: Importance indexes

3 classes						
model	street	air	truc	factory	dancing	agriculture
$\varepsilon = .40, k = 6$	.358	.193	.072	.160	.211	.007
$\varepsilon = .47, k = 6$	.355	.198	.070	.151	.210	.008
$\varepsilon = .52, k = 6$	.352	.203	.070	.144	.223	.009
$\varepsilon = .52, k = 4$	.358	.195	.110	.150	.168	0
$\varepsilon = .55, k = 4$	.355	.197	.113	.147	.188	0

$R_(x, y)$ may be understood as the credibility that "x is at least as good as y". The set L is built using the values of R taking into consideration an antitone unary contradiction operator $\neg$ such that $\neg c_l = c_{2m-l}$ for $l = 0, \ldots, 2m$.

If $R(x, y)$ is one of the elements of L, then automatically, $\neg R(x, y)$ belongs to L. We call such a relation an L-valued binary relation.

We denote $L^{\succ m} := \{c_{m+1}, \ldots, c_{2m}\}$ and $L^{\prec m} := \{c_0, \ldots, c_{m-1}\}$.

If $R(x, y) \in L^{\succ m}$ we say that the proposition "$(x, y) \in R$" is L-true. If however $R(x, y) \in L^{\prec m}$, we say that the proposition is L-false. If $R(x, y) = c_m$, the median level (a fix point of the negation operator), then the proposition "$(x, y) \in R$" is L-undetermined.

In the classical case, where R is a crisp binary relation we define a digraph $G(A, R)$ with vertex set A and arc family R. A choice in $G(A, R)$ is a non-empty set Y of A.

A (dominant) kernel is a choice that is stable in G, i.e. $\forall x \neq y \in Y, (x, y) \notin R$ and dominant, i.e. $\forall x \notin Y, \exists y \in Y$ such that $(x, y) \in R$.

We now denote $G^L = G^L(A, R)$ a digraph with vertices set A and a valued arc family that corresponds to the L-valued binary relation R.

We define the level of stability qualification of subset Y of X as

$$\Delta^{\text{sta}}(Y) = \begin{cases} c_{2m} \text{ if } Y \text{ is a singleton,} \\ \\ \min_{y \neq x} \max_{x \neq y} \{\neg R(x, y)\} \text{ otherwise;} \end{cases}$$

and the level of dominance qualification of Y as

$$\Delta^{\text{dom}}(Y) = \begin{cases} c_{2m} \text{ if } Y = A, \\ \\ \min_{x \notin Y} \max_{y \in Y} R(x,y) \text{ otherwise.} \end{cases}$$

Y is considered to be an L-good choice, i.e. L-stable and L-dominant, if $\Delta^{\text{sta}}(Y) \in L^{\succ m}$ and $\Delta^{\text{dom}}(Y) \in L^{\succ m}$. Its qualification corresponds to $Q^{\text{good}}(L) = \min\{\Delta^{\text{sta}}(Y), \Delta^{\text{dom}}(Y)\}$.

We denote $C^{\text{good}}(G^L)$ the possibly empty set of L-good choices in G^L.

The determination of this set in an NP-complete problem even if, following a result of Kitainik [12], we do not have to enumerate the elements of the power set of A, but only have to consider the kernels of the corresponding crisp strict median-level cut relation $R^{\succ m}$ associated to R, i.e. $(x,y) \in R^{\succ m}$ if $R(x,y) \in L^{\succ m}$.

As the kernel in $G(X, R^{\succ m})$ is by definition a stable and dominant crip subset of A, we consider the possibly empty set of kernels of $G^{\succ m} = G^{\succ m}(A, R^{\succ m})$ which we denote $C^{\text{good}}(G^{\succ m})$.

Kitiainik proved that

$$C^{\text{good}}(G^L) \subseteq C^{\text{good}}(G^{\succ m}).$$

The determination of crips kernels has been extensively described in the literature (see, for example [31]) and the definition of $C^{\text{good}}(G^L)$ is reduced to the enumeration of the elements of $C^{\text{good}}(G^{\succ m})$ and the calculation of their qualification.

The decision maker might also be interested in bad choices. These choices correspond to absorbent kernels with a qualification greater than c_m. In the classical Boolean framework (see [31]) an (absorbent) kernel is a choice that is stable and absorbent, i.e. $\forall x \notin Y, \exists y \in Y$ such that $(x,y) \in R$. As $(x,y) \in R$ is equivalent to $(y,x) \in R^t$, where matrix R^t represents the transpose of matrix R, all the results obtained for dominant kernels can be immediately transposed for absorbent kernels and definitions like Δ^{bad} and Q^{bad} are obviously and straightforwardly obtained from Δ^{good} and Q^{good}.

Indeed the level of absorbance qualification of Y is defined as

$$\Delta^{\text{abs}}(Y) = \begin{cases} c_{2m} \text{ if } Y = A, \\ \\ \min_{x \notin Y} \max_{y \in Y} R(y, x) \text{ otherwise.} \end{cases}$$

In order to determine a unique rational choice (if any), we first compute dominant kernels in G^L (see [31], [1]) and determine their qualification as not being bad choices, i.e. $\neg Q^{\text{bad}}(Y)$ where $Q^{\text{bad}}(Y) = \min(\Delta^{\text{sta}}(Y), \Delta^{\text{abs}}(Y))$. The selection is based on

$$\min\{Q^{\text{good}}(Y), \neg Q^{\text{bad}}(Y)\}.$$

If more than one candidate remain, other discriminant functions may be added as lowest cardinality, minimum degree of inaccessibility, ...

Another approach was proposed by Orlovski [22] in order to determine the best options. He considered the fuzzy set of maximally non dominated alternatives ND where

$$ND(x) = \min_{y \in A\{x\}} \neg P(y, x)$$

and $P(y, x)$ corresponds to the degree fo strict preference associated to $R(y, x)$ (see [23], [6], [7]). The rational choice corresponds to those alternatives having the maximal value $ND(r^*)$. He also gave conditions to obtain elements that are unfuzzily undominated.

2.11 Conclusion

In this chapter we have presented a few approaches to multiple criteria decision aiding. In particular, we have focussed on fuzzy methods for choice, sorting and ordering. We have also described in details the sorting procedure TOMASO which can deal with interacting criteria. Some tests on examples (theoretical and real-life) have shown the interestingness of this method. Further investigations have to be done on the validation of the models. We intend to implement a cross-validation procedure to make stability tests on the data and the method. Concerning the noise annoyance problem, we would like to treat the whole problem. But some limitations due to the method (consistency of the data) and to the software (memory problems if the number of criteria exceeds 10) will require some further research. The first problem can be solved by searching a maximal subset of alternatives which don't violate the monotonicity condition.

Bibliography

[1] R. Bisdorff, M. Roubens, On defining and computing fuzzy kernels and L-valued simple graphs, Intelligent Systems and Soft Computing, In: D. Ruan, P. D'hondt, P; Govaerts and E. Kerre (eds.): *Intelligent Systems and Soft Computing for Nuclear Science and Industry*, World Scientific, Singapore, (1996) 113-122.

[2] D. Botteldooren, A. Verkeyn, Fuzzy models for accumulation of reported community noise annoyance from combined sources, Journal of the Acoustical Society of America, 112, (2002) 1496 - 1508.

[3] J.P. Brans, P. Vincke, *A preference ranking organization method, Management Sciences*, 31, (1984) 647-656.

[4] G. Choquet, Theory of capacities, *Annales de l'Institut Fourier*, 5, (1953) 131-295.

[5] D. Dubois, H. Prade, Weighted minimum and maximum in fuzzy set theory, *Information Science*, 39, (1986) 205-210.

[6] J. Fodor, S.A. Orlovski, P. Perny, M. Roubens, The use of fuzzy preference models in multiple criteria: choice, ranking and sorting. In: D. Dubois, H. Prade (eds.): chapter 3 of volume 5 (Operations Research and Statistics), *Handbooks of Fuzzy Sets*, Kluwer Academic Publishers, Dordrecht, Boston, London, (1998) 69-101.

[7] J. Fodor, P. Perny, M. Roubens, Decision Making and Optimization, In: E. Ruspini, P. Bonissone and W. Pedrycz (eds.): Chapter 5.1 of *Handbook of Fuzzy Computation*, Institute of Physics Publ. and Oxford University Press, Bristol, (1998) F.5.1:1-14.

[8] M. Grabisch, M. Roubens, Application of the Choquet Integral in Multicriteria Decision Making, In: M. Grabisch, T. Murofushi, M. Sugeno (eds.): *Fuzzy Measures and Integrals*, Physica Verlag, Heidelberg, (2000) 348-374.

[9] M. Grabisch, J.-L. Marichal, and M. Roubens, Equivalent representations of set functions, *Mathematics of Operations Research*, 25, (2000) 157–178.

[10] M. Grabisch, k-order additive discrete fuzzy measure and their representation, *Fuzzy Sets and Systems*, 92, (1997) 167-189.

[11] S. Greco, B. Matarazzo and R. Slowinski, Conjoint measurement and rough set approach for multicriteria sorting problems in presence of ordinal criteria, In: A. Colorni, M. Paruccini, and B. Roy (eds.), *AMCDA: Aide Multicritère à la Décision (Multiple Criteria Decision Aiding)*, European Commission Report, EUR 19808 EN, Joint Research Centre, Ispra, (2001) 117-144.

[12] L. Kitainik, *Fuzzy Decision Procedures with Binary Relations. Towards a Unified Theory*, Kluwer Academic Publishers, Dordrecht, (1993).

[13] J.-L. Marichal, *Aggregation operators for multicriteria decision aid*, Ph.D. thesis, Institute of Mathematics, University of Liège, Liège, Belgium, (1998).

[14] J.-L. Marichal, The influence of variables on pseudo-Boolean functions with applications to game theory and multicriteria decision making, *Discrete Applied Mathematics*, 107, (2000) 139–164.

[15] J-L. Marichal, An axiomatic approach of the discrete Choquet integral as a tool to aggregate interacting criteria, *IEEE Transactions on Fuzzy Systems*, 8, (2000) 800-807.

[16] J.-L. Marichal, Behavioral analysis of aggregation in multicriteria decision aid. In: J. Fodor, B. De Baets, and P. Perny (eds.): *Preferences and Decisions under Incomplete Knowledge*. Series: Studies in Fuzziness and Soft Computing Vol. 51. Physica Verlag, Heidelberg, (2000) 153–178.

[17] J-L. Marichal, M. Roubens, On a sorting procedure in the presence of qualitative interacting points of view. In: J. Chojean, J. Leski

(eds.):*Fuzzy Sets and their Applications*. Silesian University Press, Gliwice, (2001) 217-230.

[18] J-L. Marichal, P. Meyer, M. Roubens, On a sorting procedure in the presence of qualitative interacting points of view, submitted to International Journal of Computers & Operations Research, (2002), http://patrickmeyer.tripod.com.

[19] J-L. Marichal, Aggregation of interacting criteria by means of the discrete Choquet integral. In: T. Calvo, G. Mayor, R. Mesiar, (eds.), *Aggregation operators: new trends and applications*. Series: Studies in Fuzziness and Soft Computing, Vol 97. Physica-Verlag, Heidelberg, (2002) 224-244.

[20] T. Murofushi, A technique for reading fuzzy measures (I): the Shapley value with respect to a fuzzy measure, *2nd Fuzzy Workshop*, Nagaoka, Japan, (1992) 39–48. In Japanese.

[21] T. Murofushi, S. Soneda, Techniques for reading fuzzy measures (III): interaction index, *9th Fuzzy Sytem Symposium*, Sapporo, Japan, (1993) 693-696. In Japanese.

[22] S. Orlovski, Decision Making with a fuzzy preference relation, *Fuzzy Sets and Systems*, 1, (1978) 155-167.

[23] S.V. Ovchinnikov, M. Roubens, On strict preference relations, *Fuzzy Sets and Systems* 43, (1991) 319-326.

[24] G. Owen, Multilinear extensions of games, *Management Sciences*, 18, (1972) 64–79.

[25] P. Perny, Modélisation, Agrégation et exploitation de préférences floues dans une problématique de rangement: bases axiomatiques, procédures et logiciel, PhD thesis, Paris-Dauphine University and LAFORIA document 9 (1992).

[26] M. Roubens, Ordinal multiattribute sorting and ordering in the presence of interacting points of view. In: D. Bouyssou, E. Jacquet-Lagrèze, P. Perny, R. Slowinsky, D. Vanderpooten and P. Vincke (eds.): *Aiding Decisions with Multiple Criteria: Essays in Honour of Bernard Roy*. Kluwer Academic Publishers, Dordrecht, (2001) 229-246.

[27] M. Roubens, P. Vincke, Fuzzy preferences in an optimisation perspective. In: J. Kacprzyk and S.A. Orlovski (eds.): *Optimization Models Using Fuzzy Sets and Possibility Theory*, D. Reidel Publ. Co, Dordrecht, Boston, Lancaster, Tokyo, (1987) 77-90.

[28] B. Roy, P. Bertier, La méthode ELECTRE II. Une application au media planning. In: M. Ross (eds.): *Operational Research*, 72, North Holland Publishing Co., (1973) 291-302.

[29] B. Roy, ELECTRE III: Un algorithme de classement fondé sur une représentation floue des préférences en présence de critères multiples, *Cahiers du CERO*, 20, 1 (1978) 3-24.

[30] B. Roy, *Multicriteria Methodology for Decision Aiding*, Kluwer Academic Publishers, Dordrecht (1996).

[31] G. Schmidt, T. Ströhlein, Relations and Graphs, *Discrete mathematics for Computer Scientists*, Springer Verlag, Berlin, Heidelberg, New York (1999).

[32] L.S. Shapley, A value for n-person games. In: H.W. Kuhn, A.W. Tucker (eds.): *Contributions to the Theory of Games*, Vol. II, Annals of Mathematics Studies, 28, Princeton University Press, Princeton, NJ, (1953) 307-317.

[33] A. Verkeyn, D. Botteldooren, B. De Baets, G. De Tre, Modeling annoyance aggregation with Choquet integrals, *Proceedings of Eurofuse Workshop on Information Systems, Villa Monastero*, Varenna, Italy, (2002) 259 - 264.

[34] P. Wakker, *Additive Representations of Preferences: A new Foundation of Decision Analysis*, Kluwer Academic Publishers, Dordrecht, Boston, London, (1989).

Concluding Remarks

The present work is a collection of two articles on user centered data analysis.

The first one focuses on the selection of a quality measure to evaluate association rules. The presented procedure is used before the beginning of the search for interesting association rules. A discussion with the expert of the mined field is useful to investigate his preferences and to build the criteria correctly. The measure which is chosen after the analysis is then used to filter the association rules. These rules need to be interpreted and validated by the expert. In case the resulting rules are not satisfying the expert, a further step can lead to a redefinition of the objectives, and hence the preferences. This research is far from finished, and we are currently working on other properties of the measures and on different multiple criteria decision aid procedures for the selection of the appropriate measure. Furthermore, an important work needs to be done to permit an easy and intuitive use of the suggested method. For example, the different properties and possible preferences must be explainable to any expert. Besides, a simplified explanation of the meaning of the measures is also one of our major concerns. One of our objectives is to include this methodology of selection into a knowledge discovery procedure called 4ft-Miner [Rauch]. Its main advantage is that the quality measure is directly used when extracting the rules (and not for a later filtering, like in apriori [Aggraval]). This allows us to filter all possible rules according to the selected measure, and not only a restricted subset.

The second article is a general overview of different approaches to multiple criteria decision aid. It focuses more particularly on an ordered sorting procedure called Tomaso. We have shown its advantages and applied it to a particular real life example. Its specificities are its ability to deal with purely ordinal data and interacting points of view. Stability tests have also been performed on the method by means of leave-one-out procedures. They show that the resulting importance indexes remain stable even if the prototype is slightly modified. Currently we are working on different evolutions of the method. First of all, in case the linear program has no solution, to avoid ambiguous assignments, we intend to develop a procedure to force the alternatives to belong to a single class. Secondly, we are working on the determination of quality intervals around the Shapley indexes. Thirdly, the graphical interface of the Tomaso software needs to be improved and simplified. We intend to use graphical tools to represent the information (bar charts, data projections, ...). This visual information should help to better understand the particularities of each problem, and to interprete the different results and parameters more easily.

Bibliography

[Aggraval] R. Aggraval et all., In: U.M. Fayyad et al (eds.): *Fast Discovery of Association rules, Advances in Knowledge Discovery and Data Mining*, AAAI Press / The MIT Press (1996), 307-328.

[Figueira] J. Figueira, S. Greco, M. Ehrgott (eds.): *State of the Art of Multiple Criteria Decision Analysis*, International Series in Operations Research and Management Science, Kluwer Academic Publishers (2003).

[Lenca] P. Lenca, P. Meyer, B. Vaillant, P. Picouet, S. Lallich, *Critères d'évaluation de mesures de qualité des règles d'association*, Sessions Spéciales des XXXVe Journées de Statistique (2003).

[Rauch] J. Rauch, M. Simunek, Alternative Approach to Mining Association Rules, In: T.Y. Lin and S. Ohsuga (eds.): IEEE ICDM02 Workshop Proceedings, The Foundation of Data Mining and Knowledge Discovery, Japan (2002).

[Roy 85] B. Roy, *Méthodologie multicritère d'aide à la décision*, Ed. Economica, Collection Gestion (1985).

[Roy 90] B. Roy, Decision-Aid and Decision-Making, In: C.A. Bana e Costa (eds.): *Readings in Multiple Criteria Decision Aid*, Springer Verlag, Berlin, Heidelberg, New York, London, Paris, Tokyo, Hong Kong, Barcelona (1990), 17-35.

[Witten] I. H. Witten, E. Frank, Data Mining, Practical Machine Learning Tools and Techniques with Java Implementation, Morgan Kaufmann Publishers, Academic Press (2000).