UNIVERSITAT POLITÈCNICA DE CATALUNYA
Departament de Llenguatges i Sistemes Informàtics
Programa de Doctorat en Intel.ligència Artificial

TESI DOCTORAL

ClusDM: A Multiple Criteria Decision Making Method for Heterogeneous Data Sets

Aïda Valls Mateu

Director: Dr. Vicenç Torra Reventós
Supervisor: Dr. Ulises Cortés García

Dipòsit: Setembre 2002
Lectura: 13 de desembre de 2002

Pel meu estimat Toni

Acknowledgements

I would like to thank all those people who have made this thesis possible. First of all, to my Ph.D. director, Dr. Vicenç Torra, who introduced me in the research field a long time ago, and who has always been there to give me precious comments. I also acknowledge the advice of my supervisor, Dr. Ulises Cortés, who has helped me in all the steps that lead to this dissertation.

Fruitful discussions with other researchers have influenced this thesis, like the ideas provided by Dr. Karina Gibert or the comments received in conferences or reviews.

But this work has been possible with the help of many other people. Especially my husband, Toni, who has always believed that I could achieve those goals that seemed impossible to me. My parents, who let me study in Tarragona and Barcelona and encouraged me to continue with the Ph.D. studies. My sister, who has always listened to my discussions.

Finally, this work would not have been possible without the help of the people in the Software Department of Universitat Politècnica de Catalunya, who have always provided me assistance. As well as, the people of the Department of Computer Science and Mathematics of Universitat Rovira i Virgili, where this research has been done. In particular, the members of the Banzai and Crises groups who have always helped me.

ABSTRACT

This thesis presents a new methodology for decision making. In particular, we have studied the problems that consider more than one criterion, which is known as Multiple Criteria Decision Making (MDCM) or Multiple Criteria Decision Aid (MCDA). The difference relies on the fact of imitating the behaviour of the decision maker (i.e. develop a method that makes decisions) or giving to the decision maker some additional information that allows him to understand the mechanism of solving decisions (i.e. the decision maker can learn from the use of the method). Our proposal fits better in the MCDA approach, but has also similarities with the MCDM perspective. On one hand, the method we have designed is independent enough to not require a deep understanding of the process by the decision maker. On the other hand, we have carefully studied the process and the method is able to extract knowledge about the decision problem, which is given to the user to let him know any special characteristics of the data analysed.

ClusDM is a new method to solve multicriteria decision problems. It is able to find a ranking of alternatives or to select the best ones. Some extensions to the classical numerical approach have been studied, such as, fuzzy or ordinal values. However, we have noticed that they require having a common scale for all criteria. This thesis faces the problem of managing different types of criteria at the same time. ClusDM follows the utility approach, which considers two steps to sort a decision problem out: the aggregation and the ranking. We have included some additional steps in order to improve the process: (i) the explanation phase and (ii) the quality measurement phase. With these additional stages, ClusDM is able to build a qualitative preference ranking, which can be easily understood by the decision maker. As well as, some quality measures have been defined in order to give an idea of the trustworthiness of the ranking.

CONTENTS

List of Figures

List of Tables

CHAPTER 1. Introduction

In this introductory chapter, we will briefly situate the research area where we have been working: multicriteria decision aid. We will also present the goals formulated at the beginning of this thesis and we will summarise our main contributions to the field. Finally, we will describe the structure of this document.

1.1 What is multicriteria decision aid

The study of decision problems has a long history, and in the last decades has been one of the major research fields in decision sciences. People have to make lots of decisions during their life. Moreover, some of these decisions are directly related to the main worries of humankind, such as survival, security or perpetuation [Yu,1990]. The mathematical modelisation of these decision making problems started in the 19th century with economists and applied mathematicians like Pareto, VonNeumann, Morgenstern and many more. The first approaches considered monocriterion decision problems. In 1951 two research teams introduced the multicriteria problem: Koopmans [Koopmans,1951] and Kuhn & Tucker [Kuhn&Tucker,1951]. That is, the problem of finding the best alternative (or a ranking of all of them) considering multiple conflicting criteria or goals.

It was in late 60s that multicriteria decision making research experimented an explosive growth. In 1972 the first international conference in Multiple Criteria Decision Making was done. From then on, this is an active area of research as can be observed in the many conferences organised every year and the specialized journals (e.g. *Journal of Multi-Criteria Decision Analysis*). Some associations gather the researchers in the field (e.g. the *International Society on Multiple Criteria Decision Makin*g [MCDA,2002] or the *European Working Group on MCDA* [EWGMCDA,2002]).

Although we have been talking about Multiple Criteria Decision Making (MCDM), this is not the only denomination that can be found. In Europe, it is most common the use of MCDA, which stands for Multiple Criteria Decision Aid. This differentiation is not only in the name but also in the underneath philosophy of how to help decision makers to make better decisions (called the French School). MCDA researchers devote their main efforts to develop methods to help the user to understand the preference model behind his/her decisions. This approach includes for instance: graphical tools to represent the data or interactive methods to build a model of the preferences. On the other hand, the American School of MCDM takes a more descriptive approach. The goal is to build a model of the behaviour of the decision makers and let them to apply the model to solve new problems.

The method that we will explain in this thesis fits better with the MCDA approach. As it will be introduced in the next section, we have developed a methodology for decision making but having into account that the decision maker must understand the solution and the degree of confidence he should attribute to it, as well as, he must also be aware of how the solution has been obtained in order to let him to modify the elements that take part in the process.

1.2 Motivation and goals

The main difficulty in MCDA[1] problems lies in the fact that usually there is no objective or optimal solution for all the criteria. Thus, some trade-off must be done among the different points of view to determine an acceptable solution. Therefore, it is not an easy problem at all, which explains the large amount of publications in the area in the last decades.

Although MCDA problems have been studied in the Operational Research area for a long time, recently there is an increasing interest in including Artificial Intelligence techniques to the classical numerical methods [Bana e Costa,1990]. Sometimes, the knowledge available about the alternatives cannot be expressed numerically (or it is difficult to use numbers instead of other types of values) [Wang,2001]. For instance, assume that we need to have into consideration the height of each person in a team; if we do not have any tool to measure the height, it is difficult to give a numerical value for each person. On the contrary, it is very natural to say: “tall”, “short”, “very tall”, etc. Therefore, different approaches to the use of non-numerical values in MCDA have been developed (see Chapter 2).

However, very few methods consider the possibility to have matrices with heterogeneous criteria (different types and/or different scales). This limitation to a common scale for all criteria forces the data suppliers to use values that could be different to the ones they would normally use. Other approaches let the user to provide heterogeneous data, which is automatically translated into a unified scale before their processing [Herrera&Herrera-Viedma,2000]. In this case, the transformation obtained does not contain all the information that the person has initially provided. For this reason, sometimes it is argued that is better to allow only a unique scale for providing the data. We agree with the authors (e.g. [Delgado et al., 1998]) who argue that, in spite of the increase of the ambiguity,

[1] In this section, we will use the acronym MCDA but what is explained is also valid for MCDM.

uncertainty or contradiction in the data, the more information we have the better understanding of the alternatives.

So, after detecting the problem of heterogeneous criteria, we became interested in studying this situation and **we concentrated our efforts in developing a different approach that is able to handle different types of criteria without making an explicit translation into a common domain a priori**.

In the MCDA field, three kinds of problems are distinguished [Vincke,1992]: *choice problems*, *ranking problems* and *sorting problems*. The goal of the decision maker in each type of problem is different: in choice problems the aim is to find the best alternative, in ranking problems we want to know the *goodness* of all alternatives, which is usually presented as a ranking from the best to the worst, and in sorting problems we want to know which alternatives belong to each class of a predefined set of classes.

As a natural extension of our previous research in unsupervised learning methods, particularly in clustering, **we have focused on *choice* and *ranking* problems** instead of the *sorting* one, which is classically solved with a supervised approach [Zopounidis,2002]. As it will be seen, this unsupervised approach has been considered during all the stages of the decision process. In spite of losing some power, we have intended to not require the user to know technical details about the methods in order to provide an easy tool that does not need a learning stage.

The necessity to give a qualitatively described result has been argued by different authors. The rationale behind this belief is that human decision makers understand better a linguistic statement characterising the selected alternative (or ranking of alternatives) than a numerical result or even a membership function. In [Bana e Costa, 1990] different components of the ideal solution are identified: not only the position of each alternative in the ranking is important, but also the intensity of preference of each one and the degree of truth of the result. The reliability of the system can change depending on the degree of consensus of the different criteria. Thus, if the decision maker's confidence in the system makes him follow the recommendation without doubt and the alternative chosen is not good enough, the result will be disastrous, especially in critical situations. The decision maker will also welcome the addition of other information about the reasons of having obtained a bad or good result. **We have devoted special attention to the definition of quality measures and linguistic descriptions of the decision making process**.
In particular, we have studied a new semantics for qualitative criteria based on the concept of antonym. The key idea is that we can infer the meaning of a term knowing the terms that express an opposite value [de Soto&Trillas, 1999]. This relation among the qualitative terms can be represented with a negation function. In [Torra,1996] an extended negation function is proposed. In [Valls&Torra,1999c] we studied in detail the use of the negation functions and their induced semantics, observing that it is adequate to capture more information than with ordinal qualitative criteria. Nowadays, the fuzzy representation of the

semantics of qualitative terms is the most widespread, however, from the expert's point of view, it is easier to give the information needed to build a negation function than the information required to build a fuzzy set. Therefore, **we have focused on the use of the negation based semantics during the decision analysis and, especially, in the description of the ranking of alternatives.**

In summary, the goal of this doctoral dissertation is ***the development of a multi-criteria decision method for choice and ranking problems that allows different types of values and domains without making an explicit translation into a common domain. In addition, some quality measures and linguistic explanations will be part of the result.***

1.3 Contributions

The main contributions of this research work in the MCDA field are the following:

- ♦ We have designed a new methodology for multi-criteria decision problems with heterogeneous data, called ClusDM. In particular, we have developed a system that considers three different types of values: numerical, qualitative and Boolean. Moreover, each qualitative criterion can have its own set of available linguistic terms.

- ♦ We have studied the use of a new kind of qualitative description, based on negation functions. We have used this simple representation of the semantics for the linguistic terms during all the stages of the process. At the end, the ranking of alternatives (or the selected one) is explained to the decision maker using this type of values.

- ♦ We have developed a method to adapt one of the preference vocabularies provided by the experts to describe the overall preferences of the alternatives, that is, to explain the ranking obtained. The use of terms that are already known for the user can make things easier to him. In particular, we provide a way of selecting one of the vocabularies depending on the similarities among the meaning of its terms and the meaning of the groups of alternatives we want to characterise. Moreover, we have developed some algorithms to select the most appropriate terms among the ones in the vocabulary, to produce new terms and to adapt the meaning of these terms according to the characteristics of the result.

- ♦ We have identified different key points in any MCDA process where the quality of the partial results generated should be evaluated. In particular, we have defined different quality measures for the different stages of our method. With these measures we can give an overall value of the trustworthiness of the final result. This kind of information is very useful for the decision maker in order to pay more or less attention to the recommendations of the system.

- We have developed a methodology that is able to detect conflicting elements. In particular, the decision maker is notified about alternatives that have received opposite preference evaluations for different criteria and about criteria that do not agree with the majority. With this additional information, the user is able to modify the data set, for instance, dropping alternatives that may not have been considered, including additional criteria or modifying their weights.

1.4 Structure of the Thesis

This document is divided into 8 chapters. The first one has given an introduction to the problem we have studied and the goals and contributions of the thesis.

Chapter 2 presents a brief survey of the MCDA research field. From a general view, we go through the classical approaches to multicriteria decisions until focusing on uncertainty in utility-based models using ordinal qualitative criteria.

Chapters 3, 4 and 5 explain in detail the methodology we have developed, called ClusDM (Clustering for Decision Making). Four phases can be distinguished. The first one, the aggregation phase is explained in Chapter 2; the second, the ranking phase, is detailed in Chapter 3; finally, Chapter 4 includes the explanation and quality measurement phases.

ClusDM can be seen as an aggregation operator for qualitative preference criteria. The properties of this new decision operator are defined and proved in Chapter 6. We have studied the usual properties required to this type of operators. We begin with the basic properties: symmetry, idempotence and monotonicity. Then, we continue with more elaborate properties. We will see that ClusDM does not satisfy some of these properties.

The methodology we propose has been implemented using Lisp and Java. The Lisp code is included in a system called *Radames*, which performs aggregation of numerical data, qualitative data, heterogeneous data, data matrices and trees. In particular, the ClusDM methodology is the one used in aggregation of qualitative and heterogeneous data sets. The Java code is integrated in an agent developed in Jade, called *ClusDMA*. Chapter 7 gives some details about these systems and explains the results obtained in three different application domains.

Finally, an overview of the thesis can be found in Chapter 8. This chapter also gives some interesting directions to continue this work.

CHAPTER 2. Review of Multicriteria Decision Aid techniques

In this chapter we explain the different approaches for solving a Multiple Criteria Decision Aid problem. We begin with the definition of the problem and the concepts we will deal with during this dissertation. Then, the main characteristics of the most used approaches are given: Multi-attribute Utility Theory and Outranking Relations. Then, the Rough Sets approach is described due to the similarities with our research. The chapter finishes reviewing the field of multicriteria decision in the case of imprecision and uncertainty.

2.1 Formalisation

In this section, we formalize the multi-criteria decision problem. We define the concepts and nomenclature used in this document. This section is included because there is not a common standard for the denomination and nomenclature of the elements that participate in this kind of decision-making frameworks.

2.1.1 Concepts

These definitions are adapted from the ones provided in [Roy,2000]. In the rest of the document we will follow the notation introduced here.

Actor: Any individual, group of individuals or entity which can play a role, directly or not, in the decision process.

Decision-maker: Actor for whom the decision-aid tools are developed and implemented.

Analyst: Actor who is responsible for the decision-aid process.

Action: A generic term used to designate the object of the decision. In practice, the term action may be replaced by such terms as scenario, operation, investment or solution, depending on the situation. We will follow the notation: $A = a_1, a_2, \ldots , a_m$. This is, m different actions in the set of possible actions A.

Alternative: Action that can be implemented independently of the other actions. This term can be used instead of Action when this independence condition is fulfilled.

Potential Action: Action which could be implemented or which is interesting for the analysis during the decision process.

Point of view: A class of effects or attributes which share the same goal or the same type of concerns, thought pertinent by at least one of the actors for evaluating and comparing actions.

Scale: Set of elements, S, (called “degrees”) ranked according to a complete order, reflecting the preferences of the decision-maker for a particular point of view.
Different scales can be considered according to the allowed operators on the set of elements. Some of the most common scales are: numerical, ordinal and categorical.

Vocabulary: Set of elements (or degrees) expressed using linguistic terms. We will refer to the vocabulary of a particular qualitative criterion as T.

Criterion: Application g from the set of actions to a scale, such that it appears meaningful to compare two actions a_1 and a_2 according to a particular point of view, on the sole basis of $g(a_1)$ and $g(a_2)$. We will follow $C = \{c_1, c_2, \ldots, c_p\}$ to denote the criteria, being g_j the function attached to c_j. Also, with v_{ij} we denote the value of $g_j(a_i)$. That is $v_{ij}=g_j(a_i)$.

Threshold: Value that is used to take into account the imprecision on the result of certain comparisons. It permits to establish the equivalence between two alternatives evaluated different in a given scale. For example, we can define an Indifference threshold, a Preference threshold or a Dispersion threshold.

Weight: Value that indicates the relative importance of one criterion in a particular decision process. It models the different roles that an actor would like the different criteria to play in the elaboration and argumentation of comprehensible preferences.

The concrete interpretation of this concept depends on the method, as it will be illustrated later.

Ideal Point: Point in the criterion space that has the maximum value for each dimension.

Nadir Point: Point in the criterion space that has the minimum value for each dimension.

Preference relation: Binary relation that expresses how much an action is preferred over another one. Several scales can be used for expressing these preferences. A preference relation is an application $R: A \times A \rightarrow S$.

Preferential Independence: A subset of criteria C is preferentially independent of C^c (the complement of C) iff any conditional preference among elements of C, holding all elements of C^c fixed, remain the same regardless of the levels at which C^c are held.
A classical example of non preferential independence is the following one: we have two criteria g_1 and g_2, being g_1= {red wine, white wine}, g_2= {meat, fish}. Most people prefer red wine to white wine with meat, but white wine to red wine with fish. Therefore, preference in g_1 depends on g_2.

Cluster: Considering alternatives as points in a p-dimensional space, clusters may be described as continuous regions of this space containing a relatively high density of points, separated from other such regions by regions containing a relatively low density of points [Everitt,1977]. Clusters described in this way are sometimes referred to as *natural clusters*. Other definitions can be found in the literature, however, the advantage of considering clusters in this way is that it does not restrict the shape of clusters as rigidly as do other proposed definitions. For example, definitions suggesting that objects within a cluster should be closer to each other than to objects in other clusters restrict one to the consideration of spherical shapes.

2.1.2 Multiple criteria decision problem

Having a defined set *A* of actions and a consistent[2] family *C* of criteria on *A*, a multiple criteria decision problem is the one that, with respect to *G*, either aims to find:

a) a subset of *A* that contains the best actions,
b) an assignment of the actions into predefined categories, or
c) a rank of the actions in *A* from best to worst.

Each of these objectives defines a different multicriteria decision problem, called: (a) choice problem, (b) classification or sorting problem (depending on if the categories are preferentially ordered or not) and (c) ranking problem.

The main difficulty lies in the fact that it is an ill-defined mathematical problem because there is no objective or optimal solution for all the criteria. Thus, some trade-off must be done among the different points of view to determine an acceptable solution for the decision problem.

2.1.3 MCDM versus MCDA

Multicriteria decision making (or multiple criteria decision making, MCDM) can be understood a part of the more general area of research: Multicriteria decision aid (MCDA). MCDA develops tools to help decision-makers in solving a decision problem with several points of view that have to be taken into account. This is not an easy task because often these points of view are contradictory, consequently, it is not always possible to find a unique solution that is the best for all the points of view.

MCDA intends to give tools that allow the decision-maker to capture, analyse and understand these points of view, in order to be able to find the way in which the decision process must be handled. This is called a *constructivist* approach.

Multicriteria decision making (MCDM) has a more *descriptive* approach. In MCDM it is supposed that there exists "*something*" that will allow the decision-maker to determine which are the best alternatives. This is done using a utility function - if it can be discovered and described in mathematical terms - or using mechanisms based

---

[2] A set of criteria that is **exhaustive** ($g_j(a) = g_j(b), \forall j \Rightarrow$ no preference between a and b), **cohesive** ($g_j(a) = g_j(b), \forall j \neq k$ and a preferred to b for $g_k \Rightarrow a$ preferred to b) and **nonredundant** (leaving out one criterion leads to the violation of one of the previous requirements) is said to be **consistent**.

on comparisons among the different alternatives or options. Thus, the main goal is to observe the behaviour of decision-makers and try to help them to understand the mechanisms intrinsic into the decision process, as well as, to be aware of all the factors that influence the result.

MCDM is mainly developed in United States of America (known as American School), while the constructivist approach of MCDA is the one adopted by most of the European researchers (French School) [Roy&Vanderpooten,1996].

Inside the MCDA research area we can distinguish: multiple objective decision making (MODM) and multiple attribute decision making (MADM). The former deals with problems where the decision space is continuous. MODM has been widely studied with mathematical programming methods, which have a well-formulated theoretical frame in which this optimisation problem can be studied making different assumptions on the variables as well as on functions that define the model and constraints. More information on MODM can be found in [Hwang&Masud, 1979], [Slowinski&Teghem, 1990], [Lai&Hwang, 1996] and [Ehrgott&Gandibleux, 2002]. Recently, evolutionary algorithms have been applied to MODM and they seem more appropriate to deal with problems with multiple solutions than conventional optimisation techniques [Fonseca&Fleming, 1995]. The second type, MADM, concentrates on problems with discrete decision spaces, in which the alternatives have been predetermined in advance. In the literature, it is usual to use MCDM or MCDA to refer only to the second class of problems, MADM, which is the one we are working on.

Since the beginning of the MCDA research field, many different methods have been proposed. Each method has its own characteristics and there are many ways one can classify them. For example, we can separate methods with a *single* decision maker and methods with a *group* of decision makers. The methods involving more than one decision maker are included in the research field of Group Decision Making and Negotiation (an introduction to the field can be found in [DeSanctis&Gallupe,1987] or [Jelassi et. al.,1990]). Another classification distinguishes *deterministic, stochastic* and *fuzzy* methods. The deterministic approach considers that the decision making problem (i.e. the alternatives, criteria, etc.) are perfectly described before applying the decision method. The stochastic or probabilistic case corresponds to a type of modelling in which the criteria are viewed as random variables. Finally, fuzzy methods consider different types of uncertainty and imprecision in some of the elements of the decision making problem. In section 2.5, we will give some details about uncertainty in MADM.

MADM or MCDM methods are also classified into two distinct families: *Aggregation* approaches (based on the Multi-Attribute Utility Theory) and *Order-focussed* approaches (based on Outranking relations). In the following sections we will explain the main ideas of these approaches for the deterministic case. Last section is devoted to the review of methods that allow imprecision and uncertainty, because our work is focused on the case of single decision maker MADM problems with uncertainty in the values of the criteria.

2.2 Multiattribute utility theory

Multiple Attribute Utility Theory (MAUT) has its bases in the philosophical doctrine of the Anglo-Saxon culture called Utilitarism. It was introduced in Economy by Von-Neumann and Morsgenstern to model the behaviour of economic agents. In the 60's these concepts were introduced to the decision making field (two interesting references are [Fishburn,1970] and [Keeney&Raiffa,1976]). MAUT is based on the idea that any decision-maker attempts unconsciously to maximise some function that aggregates the utility of each different criterion.

$$U = U(c_1, c_2, ..., c_p) \qquad \text{Eq. 2.1}$$

In MAUT, data is usually provided through a decision matrix, with alternatives as rows and criteria as columns (see Table 1). The values in this decision matrix can be provided by a single expert (i.e. an actor) or by different ones.

	c_1	c_2	...	c_p
a_1	v_{11}	v_{12}	...	v_{1N}
a_2				
...				
a_m				

Table 1. Decision matrix

Different models exist according to different expressions for function *U* in Eq. 2.1. The simplest model considered in MAUT is the additive one. Here, *U* is an additive combination of utility of the criteria. This is, the function *U* is expressed as:

$$U(a) = \sum_{j=1}^{n} U_j(c_j(a))$$

where U_j (the utility function of criterion c_j) is a strictly increasing function that returns values in a common scale, in order to allow the criteria to be compared and added without problems with different units of measurement. Moreover, additional conditions must be fulfilled to use this model [Vincke,1992]: each criterion must be a preference relation that induces a complete preorder, and any subset of criteria must be preferentially independent.

It has to be noted that in the additive model, other combination functions than the addition can be used to combine the utility function U_j. In particular, U can be defined as the arithmetic mean or the weighted mean of the U_j.

Apart from the use of additive utility functions, it is also possible to use other utility functions, such as the multiplicative utility one. The multiplicative model enables the consideration of the interactions among the different criteria. This model is expressed as:

$$U(a) = \prod_{j=1}^{n} U_j\left(c_j(a)\right)$$

where, as before, U_j is the utility function for criterion c_j.

A key issue in utility-based approaches is the determination of the marginal utility functions, U_j. These functions transform the scale of the corresponding criterion into numerical utility values. The construction of these functions is a difficult issue. The usual case is to build them from the information provided by some domain experts. In this case, the process of eliciting the parameters of the model is usually done through an interactive interrogation procedure (known as *Direct Methods*). However, this is a non non-easy and time-consuming process. For this reason, research on alternative methods not requiring the intensive participation of the experts have also been considered in the literature (*Indirect Methods*), where the utility functions are estimated on the basis of the global judgements made by the decision-maker on the alternatives. [Fishburn,1967] and [Vincke,1992] describe several methods for function estimation.

Once the U_j are known, the MAUT methods consider two steps to be followed [Chen&Klein, 1997], [Henig&Buchanan, 1996]:

- *Aggregation (rating)*: a global value for each alternative is computed, *U(a),* which gives a general idea of the utility of the alternative considering all the criteria at the same time;
- *Ranking or sorting*: the utility values obtained in the first step are used to find the best alternative, to rank them or to classify the alternative into some predefined groups.

When possible, different measures of interpersonal agreement or individual consistency are applied in order to give more information to the decision maker about the characteristics of the decision problem.

Another model based on the MAUT principles is the Analytic Hierarchy Process (AHP) developed by Saaty in 1980 [Saaty,1980]. In this model, the MCDM problem is decomposed into a system of hierarchies from which a $m \times p$ matrix is built. The matrix is constructed by using the relative importance of the alternatives in terms of each criterion separately. Each row of this matrix is the principal vector of an $p \times p$ reciprocal matrix determined by pairwise comparisons of the impact of the *m* alternatives on the i-th criterion.

A comparison of these three models (the additive, the multiplicative and the AHP) can be found in [Triantaphyllou,2000].

2.3 Outranking methods

The outranking approach was introduced in the 60s by Roy based on his work on real-world applications. The intention was to overcome some of the difficulties of the aggregation approaches of those days, such as the use of qualitative criteria.

This approach focuses the attention to the fact that in MCDA problems one tries to establish preference orderings of alternatives ([Roy,1991], [Perny&Roy,1992]). As each criterion usually leads to different ranking of the alternatives, the problem is to find a consensued ranking. The outranking methods perform pairwise comparisons of alternatives to determine the preferability of each alternative over the other ones for each particular criterion. Then, a concordance relation is established by aggregating the relative preferences. Moreover, a discordance relation is also established, which is used to determine veto values against the dominance of one alternative over the others. Finally the aggregation of the concordance relation yields the final dominance relation.

The basis of these methods is the definition of an outranking relation *S*. By definition, *S* is a binary relation: *a*'*Sa* holds if we can find sufficiently strong reasons for considering the following statement as being true in the decision maker's model of preferences:

“ *a*' is at least as good as *a* “

The reasons for validating this assertion have to be found in the criterion space. Two conditions must be fulfilled in order to accept that *a'Sa* holds:

1[st]. A *concordance* condition: a majority of criteria must support *a'Sa* (classical majority principle)

2[nd]. A non *discordance* condition: among the non concordant criteria, none of them strongly refutes *a'Sa* (respect of minorities principle)

There are different ways of implementing these conditions and different levels of requirement. Let us explain them in more detail.

Concordance is measured in two steps. Firstly, we measure the contribution of each criterion, c_j, to the outranking relation *a'Sa*. We define the partial concordance of one criterion so that it follows these two conditions: concordance is 1 when the j[th] criterion fully supports *a'Sa* and concordance is 0 when the criterion does not support *a'Sa* at all.

$$concordance_j(a',a) = \begin{cases} 1 & \text{if } c_j(a') \geq c_j(a) - q_j \\ 0 & \text{if } c_j(a') \leq c_j(a) - p_j \\ \dfrac{p_j - (c_j(a) - c_j(a'))}{p_j - q_j} & \text{if } c_j(a) - p_j \leq c_j(a') \leq c_j(a) - q_j \end{cases}$$

where p_j is the preference threshold and q_j is the indifference threshold of the j[th] criterion. These thresholds define 5 different intervals in the domain of preference of the criterion, as it is shown in Figure 1: P_j means "strict preference", Q_j is "weak preference" and I_j corresponds to "indifference".

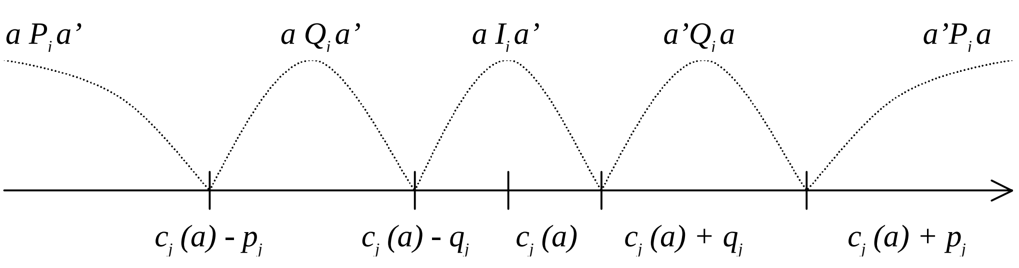

Figure 1. Thresholds in a criterion

Secondly, the overall concordance value is obtained using the partial concordances. We can use the weights associated to each criterion, w_j, to adjust the influence of each of them.

$$Concordance(a',a) = \sum_{j=1}^{p} w_j \cdot concordance_j(a',a)$$

With respect to the discordance condition, outranking methods use the discordance measurement to introduce the opportunity of the non concordant criteria to express their strong opposition, a *veto*, denoted v_j.

If $c_j(a') < c_j(a) - v_j$, for some criterion c_j, then *a'Sa* is rejected.

In these methodologies, the decision process has three steps:

α: isolate the smallest subset of alternatives $A_0 \subset A$ liable to justify the elimination of all actions belonging to $A \setminus A_0$;

β: assign each alternative to an appropriate pre-defined category according to what we want it to become afterwards;

γ: build a partial (or complete) pre-order on the subset A_0 of those among the alternatives of *A*.

Different methods implement this process using different algorithms. Some of the most well-known outranking models are ELECTRE, PROMETHEE, MAPPAC and PRAGMA (see e.g. [Bana e Costa,1990]).

2.4 Another approach: the Rough Sets Theory

Despite that the two major models used in MCDA are the ones based on Utility functions and Outranking relations, there are other approaches that face up the problem from other perspectives. In this section, we will give some details about the Rough Sets approach, because there are some similarities between their ideas and goals and the ones of the methodology we propose, ClusDM.

The rough sets theory was formulated by Pawlak [Pawlak,1982] to deal with inconsistency and vague description of objects. The theory is based on the concept of indiscernibility relation, which induces a partition of the objects into blocks of indiscernible (i.e. indistinguishable) objects, called elementary sets. Being *X* the universe of discourse, any subset *Y* of *X* can be expressed in terms of these blocks either precisely or approximately. In the second case, the subset may be represented by two sets called the *lower* and *upper approximations* of *Y*. A rough set is then defined using these approximation sets.

The lower and upper approximation sets are built from a data matrix of examples. In decision making, an example is formed by a description of an alternative in terms of different criteria and the final decision value given to the alternative by the decision maker after solving the problem. That is, if we use the concepts of machine learning, the rough sets approach is a *supervised method*, because we require the knowledge of some solved problems in order to build a model to solve new ones. In fact, the rough sets methodology was introduced as a method to infer decision rules from a set of examples.

An interesting characteristic of the rough set approach is that it is possible to deal with heterogeneous data sets without having to use a unified domain. The rules are generated from the analysis of the elements in the lower, upper and boundary approximations of the different solutions. That is, the values of the elements in these sets (in spite of the type and domain) define the conditions of the rules for the different conclusions (i.e. decision results).

Until now, we have introduced the classical rough set approach using an indiscernibility relation. However, there are some generalisations of the method to deal with fuzzy sets, fuzzy indiscernibility relations [Dubois&Prade,1990], or to substitute indiscernibility by a weaker binary similarity relation or even a fuzzy similarity relation [Greco et. al.,1998]. Another generalisation refers to the treatment of missing values. The classical approach requires a complete data matrix, while other works have relaxed this condition to allow the presence of missing values [Greco et. al.,2000].

The application of rough sets to multiple criteria decision making began in the 90's [Slowinski,1993]. The original rough set approach is not able, however, to deal with preference-ordered criteria and decision classes. Moreover, the rough sets theory is devoted to classification problems, while MCDA also deals with choice, ranking and sorting situations. In fact, initially it was only used in MCDA classification applications [Pawlak,1997].

In [Greco et. al.,2001] there is a good explanation of how rough sets theory can be adapted to deal with the particular characteristics of sorting, choice and ranking decisions. The main modification is the substitution of the indiscernibility relation by a dominance relation. Indiscernibility is not able to deal with ordinal properties. In particular, it can not detect some inconsistencies. For example, alternative a_1 is better than a_2 with respect to all the considered criteria, but the decision maker considers than a_1 is worse than a_2 as a solution to the decision problem. In order to detect this inconsistency, the rough approximation must handle the ordinal properties of the criteria. This can be naturally achieved with dominance relations[3]. In the case of

[3] *x dominates y* if *x* is at least as good as *y* for all criteria.

multicriteria choice and ranking problems, other extensions are needed because the data matrices used in the classical rough sets theory do not allow the representation of preferences between alternatives. Here, Greco et al. propose to operate on pairwise comparison tables, where rows represent pairs of alternatives for which multicriteria evaluations and the global preference relation are known.

Therefore, we can see that the rough set theory is also useful in MCDA. Moreover, some of its characteristics are also the goals that we will face up in this thesis: the integration of heterogeneous criteria, the possibility of having missing values and the explanation of the result in a language that is easy for the decision maker.

The main differences between the rough set approach and the one presented in this thesis are: (i) the type of information required to the decision maker and (ii) the type of result obtained. The rough set case needs to have a set of solved decision examples in order to build the rules that explain how to make decisions in the future. In our proposal, we directly deal with the data of an unsolved decision making problem and find the solution for this particular case. Using AI terminology: rough sets is a supervised method while ClusDM is unsupervised. Regarding the type of result, rough sets build a set of decision rules that can help the decision maker to solve future problems, however, the new problems must be similar to the ones used to build the model in order to be able to apply the same rules. In our case, the method is applied to a particular decision case. The solution is also expressed in a language that is familiar to the decision maker, together with additional information that can help him to understand the problem (e.g. alternatives with conflicting values or criteria that do not agree with the majority).

2.5 MCDM with imprecision and uncertainty

Traditionally, MCDA was concerned about decisions under certainty. That is, the parameters and values needed are known with certainty. Later on, the MCDA community became aware of the necessity to develop methods that were able to handle uncertain information. Here, we have to distinguish the study of decisions under risk from the study of decisions under uncertainty, understood as vagueness.

In decision making under risk, the lack of information is about the occurrence of the "state". These problems are usually handled with stochastic programming or Bayesian analysis, because the probability of occurrence of the states is known. On the other hand, we have situations with components that are intrinsically vague, called uncertain. In particular, the management of uncertainty is important in: (1) the evaluation of the alternatives with respect to the criteria, that is, v_{ij} are not known with certainty, and (2) assessing the relative importance of criteria (i.e. the weights w_j).

In this work we are not concerned with risky situations, but rather with the uncertainties as vagueness. This uncertainty arises due to different situations:

- unquantifiable information: some properties can not easily be described using numbers, then linguistic terms are usually used. For example, the comfort of a car can be evaluated with terms as good, fair, poor, etc. This type of criteria is called qualitative.

- incomplete information: obtaining a precise numerical value for some measurements is sometimes a difficult task, because the measurement equipment is not precise enough, such as the velocity of a car.

- non obtainable information: when the methodology involved in a measurement is complex and time consuming approximations of the value are used.

- partial ignorance: the experts that provide the data do not always know all the details of all criteria for all alternatives. This natural ignorance about some criteria or alternatives introduces imprecision in the global process.

The research that attempts to model imprecision into decision analysis is done basically with probability theory or fuzzy set theory [Lai&Hwang,1996]. Probability theory is claimed not to capture the human behaviour, because it models the imprecision considering random or stochastic processes in a statistical way, as if the lack of precision was a matter of randomness. On the contrary, fuzzy set theory [Zadeh,1968] models imprecision in a more human-like way, taking into account the subjectivity of the expert rather than employing only objective probability measures [Zadeh,1978].

In [Bana e Costa, 1990] an introduction to the problem of uncertainty in decision making is done. The basic ideas and problems of the management of uncertainty in the outranking utility approaches are explained. In this section, we will concentrate on the utility approach. More information about fuzzy preference relations and its aggregation operators can be found in [Orlovski,1978], [Kackprzyk&Fedrizzi,1990], [Fodor&Roubens,1994], [Chiclana et. al.,1998] and [Zapico,2000].

As it has been said in the definition of an MCDA problem, the decision maker faced one of 3 type of goals: (i) choice, (ii) classification or sorting and (iii) ranking. For each case, different methodologies are required. [Zopounidis,2002] makes a good review of methods for the classification and sorting tasks. In the rest of the chapter, we will focus on the ranking and choice problems.

When the utility approach is taken, uncertainty is usually understood as fuzzy evaluations of the alternatives with regard to the criteria. A survey of the fuzzy MCDM methods is done in [Ribeiro,1996]. In Table 2 there is a summary of different approaches to deal with fuzziness. In the column named 'Phase' it is distinguished whether the method deals with the aggregation stage (I) or the ranking phase (II). In the next columns, the nature of both the criteria, the weights is indicated and the solution obtained.

Aggregation rule	Phase	Criteria	Weights	Solution	Authors
OWA operators	I + II	fuzzy	crisp	crisp	Yager [Yager,1988]
Evidential logic rule	I	fuzzy	crisp	crisp	Baldwin [Baldwin,1994]
Choquet integral	I	fuzzy	fuzzy	crisp	Choquet [Choquet,1968]
Sugeno integral	I	fuzzy	fuzzy	crisp	Sugeno [Sugeno,1974]
Hierarchical aggregation	I	crisp	crisp	fuzzy	Laarhoven& Pedrycz [Laarhoven& Pedrycz,1983]
Max min	I + II	fuzzy	crisp	crisp	Bellman & Zadeh [Bellman&Zadeh,1970]
Max min	I + II	fuzzy	crisp or fuzzy	crisp	Yager [Yager,1978; Yager,1981]
Weighted average (WA)	I + II	fuzzy	fuzzy	fuzzy	Baas & Kwakernaak [Baas&Kwakernaak,1977]
WA. Extension principle + α-cuts + intervals	I	fuzzy	fuzzy	fuzzy	Dong, Shah & Wong [Dong&Shah,1985; Dong&Wong,1987]
WA. Approximate extension principle		fuzzy	fuzzy	fuzzy	Dubois & Prade [Dubois&Prade,1980]
WA. Extension principle		fuzzy	fuzzy	fuzzy	Schmucker [Dong,1985]
Weighted average	I	fuzzy	fuzzy	crisp	Tseng & Klein [Tseng&Klein,1992]

Table 2. Aggregation operators used MCDM (adapted from [Ribeiro,1996])

Regarding the criteria, we must distinguish two approaches to fuzziness. We may consider a criterion c_j as a fuzzy set, so that the values v_{ij} indicate the membership degree of the alternative a_i to this fuzzy criterion. For example, the comfort of a car can be evaluated using fuzzy set, and each car has a degree of comfort expressed in the interval [0,1]. On the other hand, we may consider the possible values of a criterion c_j as being uncertain, that is, v_{ij} are linguistic terms. For example, the comfort of a car can have a domain with the values “good”, “not-bad” and “uncomfortable”. In this case, each linguistic term is a fuzzy set.

Before coming into details of the methods, it is important to notice that the *complexity* of the ranking phase depends on the type of the result of the rating phase. That is, if the result of the aggregation is a crisp value, the ranking is straightforward (just select the alternative with the highest value), on the other hand, for other types of results this process can be difficult. For instance, when the result is a fuzzy set, a ranking method to order them must be used. However, there is not a unified methodology for ordering fuzzy sets (see [Klir&Yuan,1995] for details).

The first four methods in Table 2 consider the first approach, so v_{ij} are membership degrees (i.e. crisp numbers), while the rest of the methods use the second approximation, in which the values v_{ij} are linguistic terms with a fuzzy set that gives us its semantics.

Yager uses the OWA operator [Yager,1988], which is an aggregation operator that averages the values giving different weights to the values rather than weights to the criteria. Baldwin proposes to do a simple weighted average and then use a linguistic filter to obtain the level of satisfaction of the criteria. The filters are fuzzy sets as “most”, “all”, “few”, etc. Choquet and Sugeno define fuzzy integrals to make the consensus of the fuzzy values; the weights are given by fuzzy measures of the form: m: $\wp(C) \rightarrow [0,1]$, which define the importance of any subset of C. Several characterizations of Choquet integrals are available, see e.g. [Narukawa&Murofushi, 2002]. In addition, T-conorm fuzzy integral [Murofushi&Sugeno, 1991] generalize Choquet and Sugeno integrals.

The method proposed by Laarhoven and Pedrycz in 1983 is a variation of Saaty’s method AHP for dealing with uncertainty. Saaty used the classification trees to deal with intermediate values like “about three”. Laarhoven and Pedrycz fuzzify the crisp values obtained from pairwise comparisons, as in Saaty’s approach, and use the approximate algorithms of Dubois and Prade to perform the algebraic operations on the fuzzy numbers.

Bellman and Zadeh gave, in 1970, a max-min approach to the aggregation process. The final result is a fuzzy set whose membership function is the degree to which an alternative is a solution. This membership is obtained from the following aggregation function:

$$\mu_D(a_i) = min\left(w_1 * \mu_{c1}(a_i), w_2 * \mu_{c2}(a_i), \ldots, w_n * \mu_{cn}(a_i)\right)$$

$$\text{with } \sum_j \mathrm{w_j} = 1$$

The ranking phase will choose the alternative a_i with the maximum membership to the decision fulfilment. That is the reason why the process is called Max-Min.

Yager assumed the Bellman and Zadeh's max-min principle, but the importance of the criteria is represented as exponential scalars. This is based on the idea of linguistic hedges of Zadeh [Zadeh,1983], which are assigned according to linguistic variables (e.g. μ^2 corresponds to "very"). Formally,

$$\mu_D(a_i) = min\left(\mu_{c1}(a_i)^{\alpha 1}, \mu_{c2}(a_i)^{\alpha 2}, ..., \mu_{cn}(a_i)^{\alpha n}\right) \text{ for } \alpha > 0$$

The rest of the approaches (Baas and Kwakernaak, Dong et al., Dubois and Prade, Schmucker, and Tseng and Klein) propose different methods to compute a weighted average. They deal with fuzzy values and fuzzy weights, so the arithmetic operations needed to calculate the average must be defined for fuzzy numbers. Baas and Kwakernaak formalise the problem as a continuous differentiable function, whose largest maximum is found through the calculus of derivatives. These derivatives are used to calculate the weighted average. Dong, Shah and Wong calculate the weighted average for some α-cuts of the fuzzy sets, using interval operations and the extension principle. From the results obtained for each α-cut the final fuzzy set is built. Dubois and Prade proposed an approach based on the L-R approximation (triangular fuzzy sets are represented with three numbers: *l*, *m* and *u*, corresponding to the lower, medium and upper bounds, respectively; accordingly, the arithmetic operations are redefined using the tuples (*l,m,u*)). Schmucker discretizes the fuzzy numbers into a finite set of points, then calculates their discrete weighted average, and finally approximates the resulting fuzzy set. However, this method has problems since not all algebraic operations result in convex fuzzy numbers. Tseng and Klein gave, in 1992, an approach based on the idea of transforming the fuzzy linguistic values into numeric values by means of a defuzzification process (they use the centre of the area covered by the fuzzy number). When crisp numbers have been obtained the aggregation process belongs to the numerical case.

Another approach is to use the order among the linguistic values in a fuzzy criterion, instead of using fuzzy sets or probability theory. Methods that use this approach are given in Table 3.

In the previous techniques (the ones in Table 2), when the original values are linguistic labels in a certain set (each corresponding to a fuzzy set), the fuzzy set obtained may not correspond to any of the linguistic terms in the original term set. Thus, a linguistic approximation process is needed to find the most suitable linguistic term [Herrera&Herrera-Viedma,1997]. This process consists of finding a label whose meaning is the same or the closest (according to some metric) to the meaning of the membership function obtained after the aggregation. In order to avoid this problem, the methods in Table 3 combine the values by direct computation on labels.

Aggregation rule	Phase	Criteria	Weights	Solution	Authors
Plurality rule	I	ordinal linguistic	crisp	set of linguistic labels	axiomatization in [Roberts,1991]
Median	I	ordinal linguistic	crisp	linguistic label	median based operators in [Domingo&Torra,2002c]
LOWA / LWD, LWC, LWA	I + II	ordinal linguistic	crisp/ ordinal linguistic	linguistic label	Herrera et al. [Herrera&Herrera-Viedma,1997]
WM ordinal OWA	I + II	ordinal linguistic	crisp	linguistic label	Yager [Yager,1998]
Sugeno integral	I+II	ordinal linguistic	ordinal linguistic	ordinal linguistic	Sugeno [Sugeno,1974; Marichal&Roubens,1999]
QWM,QOWA, QWOWA, QChoquet Integral	I + II	ordinal linguistic	ordinal linguistic	linguistic label	Godo & Torra [Godo&Torra,2000; Godo&Torra,2001]
2-tuple WA 2-tuple OWA	I	ordinal linguistic	crisp	linguistic label	Martínez&Herrera [Herrera&Martínez,2000b]
antonym-based aggregation	I	ordinal linguistic	-	linguistic label	Torra [Torra,2001]

Table 3. Aggregation operators for ordinal linguistic values

The first approach is known as Plurality rule or Plurality function, and corresponds to the selection of the most frequent label. In fact, the definition does not return a single label but a set of labels that appear more often.

The LOWA operator is a Linguistic version of the OWA operator [Yager,1988]. This method assumes an implicit numerical scale underlying the ordinal linguistic one. Then, if all the criteria take values in an ordinal qualitative scale $L=\{l_1,...,l_r\}$, the Linguistic OWA of the linguistic values of an alternative a_i, with respect to a weighting vector W is recursively defined as:

$$C^m(W,v_i) = C^2\left((w_1, 1-w_1),\left(a_{i,\sigma(j)}, C^{m-1}(W', v_i')\right)\right) \text{ for } m > 2$$

with $w_i \in [0,1]$ and $\sum_j w_j = 1$.

Where $v_i' = \left(v_{i,\sigma(2)},...,v_{i,\sigma(n)}\right)$ and $W' = \left(w_2/(1-w_1),...,w_n/(1-w_1)\right)$, and

$C^2\left((w_1,w_2),(v_{i,1},v_{i,2})\right) = l_k$,

where $k = min\left(r, height\left(v_{i,\sigma(2)}\right) + round\left(w_1 \cdot \left(height\left(v_{i,\sigma(1)}\right) - height\left(v_{i,\sigma(2)}\right)\right)\right)\right)$.

In these expressions, σ is a permutation of v_i such that $v_{i,\sigma(j)} >= v_{i,\sigma(j+1)}$, and *height*$(v_{i,j})$ returns the position of the label within the scale L.

This method carries out an implicit conversion of the labels into the natural numbers corresponding to their position.

The LWD (linguistic weighted disjunction), LWC (linguistic weighted conjunction) and LWA (linguistic weighted averaging) are operators that consider linguistic weights for each criterion. The formulation follows Yager's Min and Max operators based on T-conorms and T-norms, respectively (see [Herrera&Herrera-Viedma,1997] for more details).

Yager's operators [Yager,1998] for qualitative values are based on the idea of the median, that is, the result is the value which is in the median position among the other values to aggregate. The Weighted Median (WM) corresponds somehow to a weighted average, and the Ordinal OWA operator replaces the classical arithmetic weighted mean by the weighted median in the OWA definition. However, as both operators are based on the median, they force the result to be one of the values that are combined, which is not always desirable.

In [Marichal&Roubens,1999] the use of the Sugeno integral as an aggregation operator for multiple criteria is analysed. It is proved that this measure has some desirable properties of aggregation operators (it is a fuzzy measure, is continuous, is idempotent in the first n arguments and is comparison meaningful for ordinal scales).

In [Godo&Torra,2000], a set of qualitative weighted mean-like operators are defined. Their main characteristic is that it is not necessary to use any kind of numerical interpretation of the qualitative (i.e.linguistic) values. They re-define all the arithmetic operations needed to apply some numerical aggregation operators (WM, OWA and WOWA) to handle linguistic terms in an ordered domain. In [Godo&Torra,2001] the extension of the Choquet integral to ordinal values is done. The Choquet measure allows the user to express the interactions between the sources, which cannot be done with the WM, OWA and WOWA operators.

A different approach is the one presented by Martínez and Herrera. A different representation of ordinal linguistic vocabularies is given. In this approach, the semantics is implicit, that is, it is encoded in the aggregation operators. They define a 2-tuple as pair (s_i, α_i), where s_i is a linguistic label and α_i is a number in [-0.5, 0.5], which indicates the distance to the closest label. Some functions to translate 2-tuples into numerical values and viceversa are given. With this functions, some classical numerical operators are redefined for the case of 2-tuples.

In 1996 another way of giving semantics to a qualitative vocabulary was defined in [Torra,1996]. It is based on the concept of antonyms: we can infer the meaning of a term if we know the terms that express an opposite value. In [Torra,2001] an operator to aggregate data described with different vocabularies is explained. It is based on building a unified vocabulary and putting the original values to this common one. In this thesis, we will explain a new methodology that uses this concept of antonyms but avoiding the necessity to work with a common vocabulary.