CHAPTER 7. Applications

The methodology developed in this thesis is general enough to be used in different kinds of problems. As we have already said, ClusDM can be used as a decision making tool or as an aggregation operator. In this chapter, we will show how to use the methodology in different frameworks. In particular we will explain the following applications:

- *Journal Review*. This is a selection problem in which the decision maker has to distinguish the best papers to be published. We have used real data provided by the editor of a special issue of a scientific journal. Each of the 22 papers submitted was reviewed by 3 experts according to 22 preference criteria. The experts evaluated the papers using a form that included quantitative and qualitative preferences as well as non-ordered categorical criteria.
- *Organ Transplant Receiver*. This decision-making problem consists in assisting the coordinator of organ transplants of a hospital in the determination of the most suitable receivers for a given organ. The goal is to obtain a ranking of the list of waiting patients according to their matching with the characteristics of the organ that is available at a particular time. Although not all the criteria can be analysed by ClusDM, the preference list obtained can be used by the medical specialists in order to make a better selection and increase the transplant success. In this case, at the moment, we have only been able to make a simplified test of ClusDM with artificial data.
- *Statistical Disclosure Control*. This is an application of our method as an aggregation operator for heterogeneous criteria. The goal is to re-identify the real values of a set of records that have been masked using different techniques. Statistical Offices must protect the data published (using masking methods) in order to preserve personal confidentiality. The degree of re-identification achieved by ClusDM gives an idea of the risk of publishing those data sets. In this application we have used the public data of the *American Housing Survey 1993* provided by the *U.S. Census Bureau*.

The following sections are devoted to these three application examples. The results obtained show that our methodology is able to give good results in many different frameworks. The first problem will be used to explain in detail the use of our methodology. For the rest of application examples, we will devote more time to comment the results than to the process itself.

7.1 Journal Review

Research publications are usually reviewed by a group of experts who give their opinion about the quality of different aspects of a set of papers. The evaluations of the experts are collected by a committee who is in charge of the selection of the best papers to be published. This problem is known as Multiple Expert - Multiple Criteria Decision Making (ME-MCDM) [Yager, 1993] because we have multiple experts that provide multiple criteria for evaluating the papers.

In the following section we will explain how we can sort out these ME-MCDM problems. The solution consists of aggregating the information of the experts and criteria at two different stages. Then, in the next section we will see with great detail the use of ClusDM in the selection of the best papers of a real journal.

7.1.1 Making Multiple Experts - Multiple Criteria Decisions

To sort out a ME-MCDM problem we have to deal with the information provided by each expert about a set of criteria. Thus, we have a data matrix for each expert, as it is shown in Figure 34. The ranking or selection of the best alternatives must be done using all this information. A two-stage process can be designed in order to aggregate the data at two different levels. In [Yager,1993] the author proposes to find an overall evaluation function for each individual expert and, in a second stage, a MCDM method is applied to aggregate these evaluations to obtain an overall value for each alternative.

We propose to interchange these two processes. In Yager's proposal, the aggregation of the data matrix provided by an expert gives us the global opinion of the expert. However, the criteria that are aggregated can refer to very different aspects of the problem (i.e. different properties, qualities, preference evaluations, etc.), so the result is putting together a huge variety of questions. Moreover, depending on the aggregation operator and the number of criteria in the matrix, the result may not reflect some important evaluations given by the expert. Our proposal consists of starting by making an aggregation of the information about each criterion given by the different experts. The result will be the consensus of the experts' opinions about a specific aspect of the problem. Then, the second stage consists of applying a MCDM method to the consensued criteria in order to find the overall evaluation for the alternatives. With this approach we pretend to reduce the loss of valuable information during the process. The consensus of the opinions about a single criterion is also interesting to detect the aspects of the problem in which the experts do not agree, or to study the ranking of the alternatives considering only a single criterion. Therefore, with our approach we are able to offer more information about the data to the decision maker. This will be illustrated with an example in the next section.

e_1	...	c_k
a_1		v_{11}
a_2		v_{11}
a_3		v_{11}
a_4		v_{11}
a_5		v_{12}
a_6		v_{12}

Expert e_1's data matrix

e_2	...	c_k
a_1		v_{21}
a_2		v_{21}
a_3		v_{22}
a_4		v_{22}
a_5		v_{21}
a_6		v_{22}

Expert e_2's data matrix

Figure 34. Data matrices about the same domain provided by e_1 and e_2

c_i	e_e	e_2
a_1	v_{11}	v_{21}
a_2	v_{11}	v_{21}
a_3	v_{11}	v_{22}
a_4	v_{11}	v_{22}
a_5	v_{12}	v_{21}
a_6	v_{12}	v_{22}

Figure 35. Data matrix to build the consensus of the i-th criterion

With our approach, in the first stage we use all the information provided by the experts to aggregate each criterion separately (see Figure 35). The preferentially independence of the 0criteria is assumed.

Notice that now the problem of synthesising this data matrix corresponds to the same problem we solve in MCDM. Thus, the same methods can be used. Nevertheless, some difficulties may arise: (i) not all the criteria are used by all the experts, and (ii) the alternatives analysed by the group of experts are not the same for all of them.

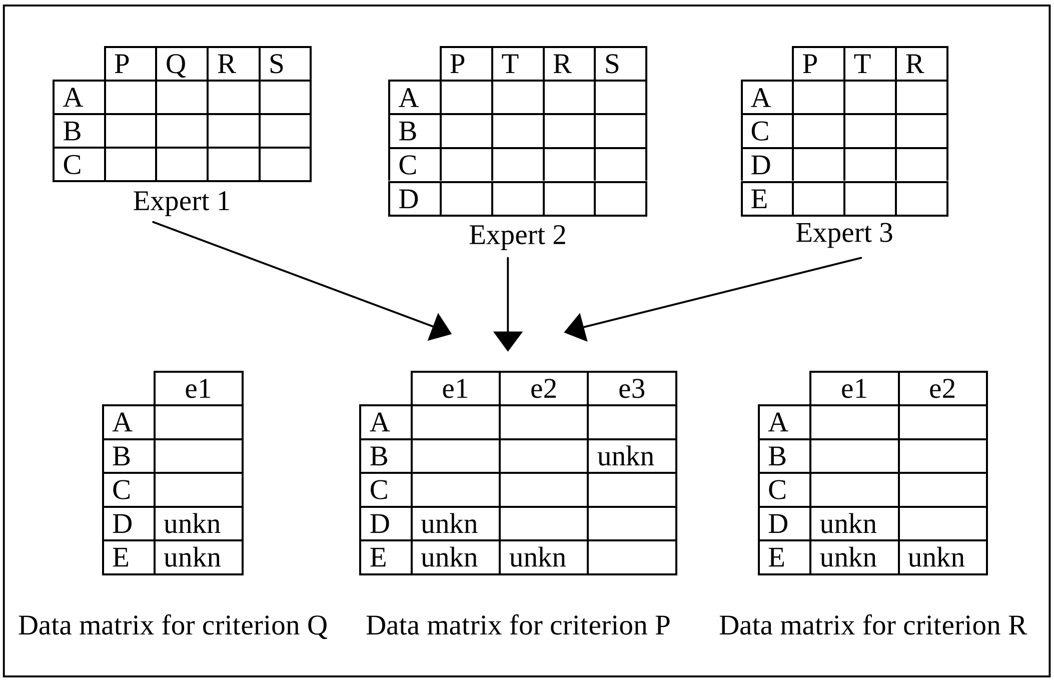

Figure 36. The decision matrix of each criterion (with missing values)

The first case is easily solved because we only put a column in the data matrix of criteria c_i if there is an expert that can fill it (see data matrix of the P and R criteria, in Figure 36). If a criterion is only provided by a single expert, there is no consensus process to be done (step 1 is not applicable, as for the Q criterion in Figure 36).

The second problem is solved using missing values, denoted as “unknown”. Therefore, the process of building the matrices is as follows. First, we put in the data matrix of c_i all the alternatives considered by the experts that use c_i. When an expert does not have a value for an alternative (because he does not know it or is not able to give his opinion about it, etc.) we introduce a special value that indicates that it is not known. Figure 36 also illustrates this procedure ("unkn” denotes a missing or unknown value). Note that this construction requires the aggregation method to be able to deal with this kind of values (as our method based on the classifier *Sedàs* does [Valls et al.,1997]).

The aggregation method to be used depends on the type of criterion (i.e. numerical, qualitative, Boolean, ...). In case of having heterogeneous criteria we can use the ClusDM methodology to find a new qualitative preference criterion. In addition, we will obtain a goodness value that can be used to weight this criterion in the next step of the process.

Once we have obtained the synthesis of each criterion, we proceed to build a data matrix with these new social criteria (Figure 37). Then, an appropriate MCDM method is used again to aggregate and rank the alternatives, and solve the decision problem.

	"P"	"Q"	"R"	"S"	"T"
A					
B					
C					
D					
E					

Figure 37. Data matrix with the consensued criteria (indicated with "")

7.1.2 Selecting the best papers for the journal with ClusDM

The call for papers for this special issue of the journal had two steps. Firstly, the authors sent an extended abstract to the editors. These submissions were numbered from 1 to 33. After a period of time, the authors were required to send the complete paper. Some of the authors did not send their papers, so finally only 22 papers were received. The papers submitted were the ones with the following identifiers: 1, 3, 4, 5, 6, 8, 9, 10, 13, 14, 16, 17, 18, 20, 21, 22, 24, 26, 29, 31, 32 and 33.

A group of 26 experts on the subject evaluated a subset of papers according to a predefined form with 22 questions. In Appendix A we can see the questions of the form. In Table 15 we have a brief description of the criteria. Ten of the questions receive a numerical mark, two are non-ordered qualitative properties (i.e. categories) and the rest, 10, are qualitative preferences over many different aspects of the paper. For some of the qualitative preferences we assumed a non-classical negation semantics. In Table 16 we show those criteria with the corresponding negation function.

Name	subject	research	relevance	are agents?	MAS descript.	originality	soundness	technical limits
Type	QN	QN	QO	QO	QO	QO	QO	QO
Domain	2 terms	3 terms	4 terms	4 terms	3 terms	5 terms	3 terms	4 terms

Name	approach	applicat. descript.	applicat. method	method. descript.	method. applicab.	abstract	introduction	conclusion
Type	QO	N	N	N	N	N	N	N
Domain	4 terms	[1,5]	[1,5]	[1,5]	[1,5]	[1,7]	[1,7]	[1,7]

Name	organis.	readable	figures	English	references	overall
Type	N	N	N	QO	QO	QO
Domain	[1,7]	[1,7]	[1,7]	5 terms	4 terms	5 terms

Table 15. Summary of the criteria for evaluating the papers of the journal

Criterion	**Vocabulary**	**Negation function for the semantics**[10]

`relevance`	`no`
	`somewhat`
	`quite`
	`very`
		`(lambda (et)`
		`  (case et`
		`      (no '(quite very))`
		`      (somewhat '(somewhat))`
		`      (quite '(no))`
		`      (very '(no))))`

`agents?`	`no`
	`doubts`
	`arguable`
	`yes`
		`(lambda (et)`
		`  (case et`
		`      (no '(yes))`
		`      (doubts '(yes))`
		`      (arguable '(arguable))`
		`      (yes '(no doubts))))`

`tech-limits`	`not-discussed`
	`poorly`
	`briefly`
	`adequately`
		`(lambda (et)`
		`  (case et`
		`    (not-discussed '(adequately))`
		`    (poorly '(adequately))`
		`    (briefly '(briefly))`
		`    (adequately '(not-discussed poorly))))`

`approach`	`not-discussed`
	`poorly`
	`briefly`
	`adequately`
		`(lambda (et)`
		`  (case et`
		`    (not-discussed '(adequately))`
		`    (poorly '(adequately))`
		`    (briefly '(briefly))`
		`    (adequately '(not-discussed poorly))))`

`references`	`poor`
	`basic`
	`old`
	`complete`
		`(lambda (et)`
		`   (case et`
		`      (poor '(complete))`
		`      (basic '(complete))`
		`      (old '(old))`
		`      (complete '(poor basic))))`

Table 16. Non-classical negation based criteria for evaluating journal papers

[10] The functions are written in Lisp. This is the language used to implement *Radames*, which is a system that follows the ClusDM methodology in qualitative and heterogeneous data sets.

Consensus of the opinions of the different judges for each criterion

Using the preferences given by the 3 experts who had evaluated each of those papers, we built a matrix for each criterion. Some papers only received two evaluations, thus they had an unknown value in the third column (the one corresponding to the 3rd evaluation). It is important to note that we put together the first evaluation of each paper in the first column of the matrix although it was not provided by the same person. That is, we were assuming that all the experts had the same interpretation of the vocabularies and semantics of the criteria. This seems a hard assumption but, in fact, when the editors analyse the evaluations given by the experts, they are using their own interpretation of the values, which is the same for all the expert's questionnaires.

The first step was the execution of a decision-making operation for each of the 22 data matrices according to the nature of the values. Using *Radames* we consensued the values using the following operators:

- The Arithmetic Average operator for numerical values
- ClusDM for each of the qualitative criteria (ordered and non-ordered ones)

Let us now follow the ClusDM execution and analyse the results obtained. For the non-ordered criteria, the process consists only of performing the aggregation stage and produces a partition. In relation to the number of clusters obtained for the qualitative criteria, the *subject* criterion generated two clusters, which received two artificial identifiers to distinguish them. The number of clusters obtained from the matrix corresponding to the *research* criterion was so big (about 10). This is an indicator of the disagreement among the experts about the status of the research (preliminary, mature or completed). So, we decided to remove this criterion from the analysis. For the rest of the criteria, the number of clusters was approximately the same than the number of values in the initial domain.

These partitions were obtained using the clustering tool called *Sedàs* (included in *Radames*). The clusters were generated using the Manhattan similarity function to compare the values of the different alternatives, and the Centroid method to build the hierarchical classification.

The next step was to apply the Principal Components Analysis to each criterion. The prototypes of the partitions were ranked using the first component of each data set. Table 17 shows the number of clusters obtained for each criterion and the quality of the PCA ranking.

Criterion	**Num. of clusters**	**Degree of agreement**	G_{PCA}	**Comments**
Relevance	4	62 %	0.627	
Agents?	4	70 %	0.678	
MAS-desc	3	55 %	0.532	High Disagreement We need more than 1 component
Originality	5	58 %	0.660	Expert 3 disagrees with the result
Soundness	4	47 %	0.568	
Tech-limits	4	48 %	0.538	High Disagreement We need more than 1 component
Approach	4	41 %	0.468	High Disagreement We need more than 1 component
English	5	59 %	0.504	High Disagreement We need more than 1 component
References	4	63%	0.556	
Overall	5	62%	0.526	Expert 3 disagrees with the result

Table 17. PCA results for each criterion of the paper's evaluation

We can observe that the quality is low for the majority of the criteria. Moreover, in 4 of them, we cannot use the result obtained because the ranking using only the first principal component may be wrong. Using the quality of the representation of each of the clusters in the first component (Eq.4.18) we could notice that these clusters were really small (see Table 18). Analysing their elements, we can distinguish the conflicting alternatives: 10, 13, 16, 21, 24 and 33.

Relevan	Agent	Original	Sound	Tec.Lim	Appro	English	References	Overall
6,9,16	1,10	10,24	13	10	4,24	21,29,33	13,16,18,21,33	24

Table 18. Conflicting papers of each criterion

According to the editors, the papers with numbers 10, 16, 24, 29, 31 and 33 had received very different marks. After a more exhaustive review, they considered them of poor quality. We want to stress that, in this test, the analysis and selection of papers made for the editors was not influenced by our results because our study was posterior.

We can see that the majority of alternatives that ClusDM discards are the ones that needed a deepest reviewing process by the editors. This shows that this methodology can also help decision makers to identify the problematic alternatives.

At the light of the low quality of the results at this stage, we decided to repeat the process removing the conflicting alternatives from the decision matrices of the criteria. Table 19 shows the new results with only 16 papers.

Criterion	Num. of clusters	Degree of agreement	G_{PCA}	Comments
Relevance	4	78 %	0.67	
Agents?	4	90 %	0.89	Highest Agreement and Quality
MAS-desc	3	76 %	0.69	
Originality	5	85 %	0.84	Good Agreement and Quality
Soundness	4	68 %	0.70	
Tech-limits	4	49 %	0.40	High Disagreement We need more than 1 component
Approach	4	64 %	0.58	
English	5	62 %	0.59	
References	5	79 %	0.65	
Overall	4	81 %	0.76	Good Agreement and Quality

Table 19. Aggregation and ranking with PCA for each journal preference criterion

Notice that the degree of agreement and the overall quality of the ranking has significantly increased when the conflicting alternatives where not disturbing the clustering and ranking processes. The single criterion whose result is not acceptable enough is the one referring to whether the *Technical Limits* of the work explained in the paper are well established or not. In this case, following what we have proposed in Chapter 4, we use the Similarity-based Ranking to compare the prototypes with the ideal alternative. The quality of the ranking obtained with this method is 0.82.

The first results of the explanation stage of the ClusDM process are shown in Table 20. The values of the third column are the positions in the unit interval of the ordered clusters of papers. The second column shows the vocabulary and intervals of the terms that where used by the experts to judge the papers. With these intervals the `Explain_Result` algorithm selects the most appropriate term for each cluster or generates new ones. The term attached to each cluster can be seen in the last column. The clusters that receive the *unknown* value are the ones that have a representation quality lower than 0.4 (Eq.4.18) or the ones with a dispersion higher than 0.2 (Eq.5.11).

Criterion	Vocabulary		z_1^{01}	Terms selected
Relevance	no somewhat quite very	[0.0,0.4] [0.4,0,6] [0.6,0.8] [0.8,1.0]	C1 = 0.76 C2 = 0.69 C3 = 0.56 C4 = 0.29	C1 - very C2 - quite C3 - *unknown* C4 - no
Agents?	no doubts arguable yes	[0.0,0.2] [0.2,0.4] [0.4,0.6] [0.6,1.0]	C1 = 0.78 C2 = 0.52 C3 = 0.34 C4 = 0.25	C1 - yes C2 - arguable C3 - doubts C4 - no
MAS-desc	bad normal well	[0.0,0.33] [0.33,0.67] [0.67,1.0]	C1 = 0.56 C2 = 0.40 C3 = 0.26	C1 - normal C2 - *unknown* C3 - bad
Originality	not mostly-not somewhat mostly very	[0.0,0.2] [0.2,0.4] [0.4,0.6] [0.6,0.8] [0.8,1.0]	C1 = 0.80 C2 = 0.61 C3 = 0.54 C4 = 0.27 C5 = 0.10	C1 - very C2 - mostly C3 - *unknown* C4 - mostly-not C5 - not
Soundness	no somewhat yes	[0.0,0.33] [0.33,0.67] [0.67,1.0]	C1 = 0.70 C2 = 0.60 C3 = 0.41 C4 = 0.17	C1 - very-yes C2 - yes C3 - no C4 - very-no
Tech-limits	not-discussed poorly briefly adequately	[0.0,0.2] [0.2,0.4] [0.4,0.6] [0.6,1.0]	C1 = 0.60 C2 = 0.57 C3 = 0.44 C4 = 0.37	C1 - very-adequately C2 - adequately C3 - briefly C4 - *unknown*
Approach	not-discussed poorly briefly adequately	[0.0,0.2] [0.2,0.4] [0.4,0.6] [0.6,1.0]	C1 = 0.62 C2 = 0.57 C3 = 0.38 C4 = 0.30	C1 - *unknown* C2 - briefly C3 - poorly C4 - not-discussed
English	deficient typo&gramm gramm typo correct	[0.0,0.2] [0.2,0.4] [0.4,0.6] [0.6,0.8] [0.8,1.0]	C1 = 0.86 C2 = 0.70 C3 = 0.69 C4 = 0.49 C5 = 0.42	C1 - correct C2 - typo C3 - very-typo C4 - *unknown* C5 - gramm
References	poor basic old complete	[0.0,0.2] [0.2,0.4] [0.4,0.6] [0.6,1.0]	C1 = 0.76 C2 = 0.64 C3 = 0.55 C4 = 0.43 C5 = 0.29	C1 - very-complete C2 - complete C3 - *unknown* C4 - old C5 - basic
Overall	not-accepted doubts accept-with-modif accept-few-modif def-accepted	[0.0,0.2] [0.2,0.4] [0.4,0.6] [0.6,0.8] [0.8,1.0]	C1 = 0.73 C2 = 0.53 C3 = 0.40 C4 = 0.16	C1 - accept-few-modif C2 - accept-with-modif C3 - doubts C4 - not-accepted

Table 20. Explanation of the clusters using the terms in the vocabulary. Green: neutral term; Blue: term generated using the negation function; Red: term generated by splitting the term that should be used by more than one class.

Criteria	Initial Vocab.		New Vocab.		Negation
Relevant	no	[0.0,0.4]	no	[0.0,0.4]	quite,very
	somewhat	[0.4,0,6]	somewhat	[0.4,0.6]	somewhat
	quite	[0.6,0.8]	quite	[0.6,0.73]	no
	very	[0.8,1.0]	very	[0.73,1.0]	no
Agents?	no	[0.0,0.2]	no	[0.0,0.3]	yes
	doubts	[0.2,0.4]	doubts	[0.3,0.42]	yes
	arguable	[0.4,0.6]	arguable	[0.42,0.58]	arguable
	yes	[0.6,1.0]	yes	[0.58,1.0]	no,doubts
MAS descript.	bad	[0.0,0.33]	bad	[0.0,0.38]	well
	normal	[0.33,0.67]	normal	[0.38,0.62]	normal
	well	[0.67,1.0]	well	[0.62,1.0]	bad
Original	not	[0.0,0.2]	not	[0.0,0.25]	very
	mostly-not	[0.2,0.4]	mostly-not	[0.25,0.45]	mostly
	somewhat	[0.4,0.6]	somewhat	[0.45,0.55]	somewhat
	mostly	[0.6,0.8]	mostly	[0.55,0.7]	mostly-not
	very	[0.8,1.0]	very	[0.7,1.0]	not
Sound.	no	[0.0,0.33]	very-no	[0.0,0.29]	very-yes
	somewhat	[0.33,0.67]	no	[0.29,0.45]	yes
	yes	[0.67,1.0]	somewhat	[0.45,0.55]	somewhat
			yes	[0.55,0.64]	no
			very-yes	[0.64,1.0]	very-no
Technic. limits	not-discussed	[0.0,0.2]	not-discussed	[0.0,0.27]	very-adequately
	poorly	[0.2,0.4]	poorly	[0.27,0.47]	adequately
	briefly	[0.4,0.6]	briefly	[0.47,0.53]	briefly
	adequately	[0.6,1.0]	adequately	[0.53,0.58]	poorly
			very-adequately	[0.58,1.0]	not-discussed
Appro-ach	not-discussed	[0.0,0.2]	not-discussed	[0.0,0.34]	adequately
	poorly	[0.2,0.4]	poorly	[0.34,0.44]	adequately
	briefly	[0.4,0.6]	briefly	[0.44.0.56]	briefly
	adequately	[0.6,1.0]	adequately	[0.46,1.0]	not-discus.,poorly
English	deficient	[0.0,0.2]	deficient	[0.0,0.2]	correct
	typo&gramm	[0.2,0.4]	typo&gramm	[0.2,0.4]	typo,very-typo
	gramm	[0.4,0.6]	gramm	[0.4,0.6]	gramm
	typo	[0.6,0.8]	very-typo	[0.6,0.7]	typo&gramm
	correct	[0.8,1.0]	typo	[0.7,0.78]	typo&gramm
			correct	[0.78,1.0]	deficient
Referen.	poor	[0.0,0.2]	poor	[0.0,0.23]	very-complete
	basic	[0.2,0.4]	basic	[0.23,0.43]	complete
	old	[0.4,0.6]	old	[0.43,0.57]	old
	complete	[0.6,1.0]	complete	[0.57,0.70]	basic
			very-complete	[0.70,1.0]	poor
Overall	not-accepted	[0.0,0.2]	not-accepted	[0.0,0.28]	def-accepted
	doubts	[0.2,0.4]	doubts	[0.28,0.45]	accept-few-mod
	accept-with-mod	[0.4,0.6]	accept-with-mod	[0.45,0.55]	accept-with-mod
	accept-few-mod	[0.6,0.8]	accept-few-mod	[0.55,0.75]	doubts
	def-accepted	[0.8,1.0]	def-accepted	[0.75,1.0]	not-accepted

Table 21. Old and new vocabulary and semantics of the qualitative criteria

After the ranking and selection of the terms that describe each of the clusters, we build the new vocabularies and their semantics. As it was explained in sections 5.1.2, the new vocabulary has all the terms of the vocabulary selected as more appropriate (which in our case is the same for all the experts) and also the new terms generated during the explanation process. Comparing the first and second columns of Table 21, we can see the changes in the vocabulary and the semantics of the terms, which is expressed with their corresponding numerical intervals. Using the intervals of the new vocabulary (given by the fuzzy sets built with the positions of the prototypes of the clusters), we defined the negation function of each term, which are shown in the last column of Table 21.

Using the new semantics we know the numerical value that would correspond to each paper according to the interval assigned to each cluster (which can be calculated using the negation function). In Table 22 we can see the numerical value for each cluster before and after defining the negation function.

The results show that the order is kept in the new semantics although there are some small differences in the position of the cluster in the unit interval. These variations are due to the adaptations of the intervals when the new negation function is defined, in order to fulfill the conditions of a negation-based semantics (definition in section 3.1).

	Relevance		Agent?		MAS desc.		Originality		Soundness	
	bef	*aft*	*bef*	*aft*	*bef*	*aft*	*bef*	*aft*	*bef*	*aft*
C1	0.76	0.9	0.78	0.8	0.56	0.5	0.80	0.9	0.70	0.9
C2	0.69	0.7	0.52	0.5	0.40	unkn.	0.61	0.7	0.60	0.7
C3	0.56	unkn.	0.34	0.3	0.26	0.17	0.54	unkn.	0.41	0.3
C4	0.29	0.2	0.25	0.1			0.27	0.3	0.17	0.1
C5							0.10	0.1		

	Techn. Limits		Approach		English		References		Overall	
	bef	*aft*	*bef*	*aft*	*bef*	*aft*	*bef*	*aft*	*bef*	*aft*
C1	0.60	0.9	0.62	unkn.	0.86	0.93	0.76	0.9	0.73	0.7
C2	0.57	0.7	0.57	0.5	0.70	0.78	0.64	0.7	0.53	0.5
C3	0.44	0.5	0.38	0.3	0.69	0.64	0.55	unkn.	0.40	0.3
C4	0.37	unkn.	0.30	0.1	0.49	unkn.	0.43	0.5	0.16	0.1
C5					0.42	0.5	0.29	0.3		

Table 22. Numerical values that represent each cluster

At this moment, we have a new consensus criterion for each of the aspects evaluated by the different experts. Thus, the papers can be studied comparing all these criteria to find out which are the ones that should be selected to be published in the journal.

Before starting the second stage of this ME-MCDM process, we must have a look at the goodness of the new qualitative criteria. To calculate the global goodness we have given the same weight to each step of the ClusDM process, since we had no extra information from the user.

	G_{ClusDM}	G_{Agg1}	G_{Agg2}	G_{Agg}	G_{PCA}	G_{Sim}	G_{Terms}	G_{Neg}
Relevance	0.83	0.97	0.77	0.71	0.67		1.0	0.92
Agents?	0.93	0.96	0.96	0.88	0.89		1.0	0.94
MAS-desc	0.85	0.96	0.89	0.80	0.69		1.0	0.92
Originality	0.92	0.95	0.88	0.91	0.84		1.0	0.94
Soundness	0.73	0.96	0.75	0.78	0.70		0.56	0.88
TechLimits	0.80	0.91	0.75	0.83		0.82	0.71	0.84
Approach	0.84	0.96	0.82	0.89	0.58		1.0	0.88
English	0.79	0.97	0.97	0.97	0.59		0.69	0.93
References	0.77	0.95	0.86	0.81	0.65		0.71	0.93
Overall	0.90	0.95	0.85	0.90	0.76		1.0	0.94

Table 23. Goodness of ClusDM in the consensus of the criteria

The first column of Table 23 shows that we have achieved very encouraging quality values for all the new social criteria (the smallest is 0.73 and most of them are over 0.8). Although the data were provided by 26 different experts, it seems that we have been able to summarise their opinions for each criterion separately.

Joint analysis of the social criteria

The second stage of the ME-MCDM process consists of aggregating and ranking the consensued data of the new decision matrix. This matrix is built with the new social criteria obtained in the previous stage. In our case, the new matrix has 21 columns, since one of the criteria (the *research* status) have been removed because the system was not able to find a coherent result. Moreover, the number of alternatives has been reduced to 16 after dropping out those that had conflicting evaluations. To aggregate this data we will use again the ClusDM methodology because we must deal with a wide range of data types with different domains.

Before starting the ClusDM decision-making process, we established a predefined vocabulary to explain the result. As it has been explained in section 5.1.1, when the vocabularies of the criteria are not appropriate to describe the overall preference of the alternatives, we must define a suitable vocabulary. In this case the set of terms chosen are: *terrible*, *bad*, *poor*, *borderline*, *acceptable*, *good* and *excellent*. The semantics of them is the classical negation, that is, *borderline* is the neutral term, and we have 3 labels for giving negative values and 3 labels for positive qualifications.

The aggregation of the decision matrix using clustering produces a partition of the papers in 6 groups. At the next step, the Principal Components Analysis builds an axis that is able to explain the 68.5 % of the information of the matrix. The global goodness of the ranking is only 0.54 over 1.0. Moreover, the stopping criterion is saying that we need 4 axes to have a good view of the data, although the first one is pretty better than the others. For this reason, we perform another ranking using the similarity to an ideal paper. With this method, we achieve a quality value of 0.82, which is acceptable enough to consider this ranking as good.

In the explanation step, we select the terms of the vocabulary to describe each cluster (Table 24). The clusters with a variance greater than 0.2 are considered as conflicting ones, because they have significantly different preference values for the criteria. In this case, the cluster with conflicting value has only a paper, number 31, which is one of the papers that were deeply reviewed for the editors, as it has been previously said.

Initial Vocabulary		z_1^{01}	**Terms selected**	**Paper's id.**
terrible	[0.0, 0.14]	C1 = 0.78	C1 - excellent	4 - 14 - 26
bad	[0.14, 0.29]	C2 = 0.67	C2 - good	18
poor	[0.29, 0.43]	C3 = 0.63	C3 - acceptable	3 - 29 - 32
borderline	[0.43, 0.57]	C4 = 0.58	C4 - borderline	6 - 8 - 9 - 17 - 20 - 22
acceptable	[0.57, 0.71]	C5 = 0.38	C5 - poor	1 - 5
good	[0.71, 0.86]	C6 = 0.36	C6 - *unknown*	31
excellent	[0.86, 1.0]			

Table 24. Qualitative description of the papers at the end of the process

To finish the ClusDM process we must obtain the new semantics of the terms. Using the similarities of the clusters to the ideal alternative in [0,1], we build the new negation function that will give meaning to the terms. In Table 25 we can see the intervals corresponding to the classical negation function, which are the original ones of the vocabulary given by the decision maker. The following columns show the intervals generated by the fuzzy sets attached to the terms, which are the ones used to determine the negations given in the last column.

Initial Vocab.		New Vocab.		Negation
terrible	[0.0, 0.14]	terrible	[0.0,0.14]	excellent
bad	[0.14, 0.29]	bad	[0.14,0.29]	excellent
poor	[0.29, 0.43]	poor	[0.29,0.44]	acceptable, good
borderline	[0.43, 0.57]	borderline	[0.44,0.56]	borderline
acceptable	[0.57, 0.71]	acceptable	[0.56, 0.65]	poor
good	[0.71, 0.86]	good	[0.65,0.72]	poor
excellent	[0.86, 1.0]	excellent	[0.72, 1.0]	terrible, bad

Table 25. Vocabulary and negation values of the papers selection criterion

Comparing the intervals corresponding to the terms before and after the process, we can see that the meaning of the positive terms of the vocabulary has changed. The coverage of term new "acceptable" is smaller than the initial one, while the term excellent has now a broader meaning.

Evaluation of the Results

The papers selected by the editors of this special issue of the journal were: 4, 14, 22 and 26. In addition, two other papers, 18 and 21, were recommended to be included in other numbers of the same journal due to the lack of space in this issue. Regarding to the last positions of the expert's preference, 1 and 5 were the worst papers.

If we analyse the results obtained with ClusDM, we can see that the papers greatly recommended for inclusion in the journal were indeed selected by the editors. The paper number 22, which was also included in the journal, was selected after another careful review of the paper by the editors, who considered that the marks given by one the referees were too low. Moreover, this work was about a subject of great interest for the research community. Those additional factors determined the final inclusion of this paper.

Concerning the low positions of the ranking, the worst papers according to ClusDM are the same than the ones indicated by the editor, number 1 and 5. Our method gives them a value of "poor" while the experts qualify them as "bad" and "terrible". This is due to the bad impression of the marks of these papers in comparison to the other ones. However, these marks are not too close to 0 as the editors thought. In spite of not obtaining such a bad qualification, we can see that the method is able to separate them and give them a low quality value.

After this rough analysis, let us pay our attention to the quality of this result. Remember that the confidence on the result is subject to the goodness values obtained in the different stages of the process. For this reason, we have detailed the calculation of these quality values in Table 26.

Measure	Value	Partial values for each element
G_{Agg}	0.90	$G_{Agg1} = 1 - \frac{0.128 + 0.139 + 0.124 + 0.137 + 0.0 + 0.0}{6} = 0.91$ $G_{Agg2} = -\frac{-0.314*2 - 0.173*2 - 0.26 - 0.368}{\ln 6} = 0.89$
G_{Rank}	0.82	$G_{Sim} = 1 - \frac{0.166 + 0.222 + 0.128 + 0.128 + 0.177 + 0.230}{6}$
G_{Terms}	0.81	$G_{Terms} = 1 - (0.0469/0.25)$
G_{Neg}	0.96	$G_{Neg} = 1 - \frac{0.08 + 0.015 + 0.025 + 0.08 + 0.015}{5}$
G_{ClusDM}	**0.87**	$G_{Neg} = 0.25*0.90 + 0.25*0.82 + 0.25*0.81 + 0.25*0.96$

Table 26. Goodness study of the second stage of the ME-MCDM

At the end of the ME-MCDM process, we have been able to rank the papers according to their global preference for being included in the journal. Moreover, the measure of confidence on the result is 0.87 (an 87%).

Nevertheless, as this is a selection problem we are really interested in knowing if the first class is good enough. For this reason we have a look at its size and we can see that, considering the rough parameter of our method, it is the ideal one (i.e. 16 papers in 6 clusters: 16/6, that makes a rounded value of 3). In this particular application, the number of papers that should be selected was 4. With this extra information we can see that we have no problem in presenting to the user the 4 best alternatives: 4, 14, 26 and 18.

With this application, we have seen that the selection made by ClusDM is very similar to the one done by the editors. The only exception is number 22, which is a special case, as mentioned above. Other interesting results have been obtained during the process. For example, the detection of the papers that receive very different evaluations for the different experts.

Having into account that this was a complex problem because we were dealing with very different types of criteria as well as multiple experts and multiple criteria at the same time, the results are very encouraging for the use of this methodology in decision making.

7.2 Organ Transplant Receiver

The Research Group on Artificial Intelligence of *Universitat Rovira i Virgili* (to which I belong) is working on a prototype to support the communication and negotiation layers of the organ transplant co-ordination process [Aldea et al., 2001]. Organ transplants have an increasing importance in Medicine. Nowadays, surgery techniques and medical treatments allow to make transplant operations to many people. However, before an operation is performed a very complex co-ordination process takes place, which could be improved with the help of an intelligent computer system.

The process starts when an organ is available to be transplanted. Then, the most appropriate receiver for this organ must be found and this search must be done in a very brief period of time (in hours). Different organisations co-ordinate all the stages of the donation and transplant process according to the local, regional, national, and international norms and laws. There is a complex co-ordination model that must be followed. For the moment, all the tasks are made by people from different hospitals, who get in touch using the telephone and fax machines. However, many hospitals are interested in automating this process using Internet. This automation could reduce the time needed to find a receiver for an organ, which is important because the organs degrade through time.

The evaluation of the list of possible receivers and their ranking according to their compatibility with the donated organ is a very important task. The transplant co-ordinator must take into account many different criteria: time in the waiting list, physical characteristics, emotional state, etc. Moreover, the decision must be made under critical time constraints. For these reasons we have proposed the use of MCDA methods to help the co-ordinator [Valls et al., 2001]. In particular, the ClusDM methodology is interesting for two reasons: (1) the information is of heterogeneous nature and (2) we can give quality measures about the degree of trust on the preference ranking obtained.

In the initial prototype that we are developing, we have considered 6 criteria: the time the patient has been in the waiting list (the patients that have waited longer, have a high priority), the distance between the hospital of the donor and the hospital of the receiver (because the farther, the more difficult will be the carriage and transplant before the life deadline of the organ), the rest of the criteria are related to some physical characteristics of the person: weight, size of the organ needed/donated, antigens typology and age of the person. With all this information, the donor is compared to the possible receivers[11] and we obtain a decision matrix with the following preference criteria:

- difference between the weights of the donor and the receiver
- difference between the size of the donor's organ and the size of the organ needed by the receiver
- number of different antigens
- difference of age between the donor and the receiver
- distance to cover to bring the organ to the receiver
- amount of time that the receiver has been waiting for this organ

The preferences are expressed with linguistic values in the vocabulary chosen by the hospital transplant co-ordinator (see Table 27).

	worst value . . . *best value*
Weight	inadequate, feasible, good, optimum
Size	inadequate, feasible, good, optimum
Antigens	different, similar, identical
Age	more_thn_20,more_thn_17,more_thn_14, more_thn_11,more_thn_8,more_thn_5,the_same
Distance	country, zone, region, city, hospital
Waiting time	very_short, short, acceptable, long, very_long

Table 27. Vocabularies of the criteria for comparing the receivers of an organ

Although all the criteria are qualitative, they refer to very different aspects that are important in a transplant. Therefore, their vocabularies are not appropriate to describe the global suitability of the receivers. Instead of using the vocabularies of those criteria, we have build a new preference vocabulary. To avoid having to generate new terms, the vocabulary has 15 terms (with the classical negation semantics), indicating different degrees of compatibility with the donated organ, which are: *the_worst, terrible, very_bad, bad, not_recommendable, inappropriate, borderline, acceptable, adequate, recommendable, fairly_good, good, very_good, excellent, optimum*.

As the process of searching the best receiver is done hierarchically, we have fixed the desired number of groups of patients to 7. After ranking these groups using the similarity method, each group receives one of the 15 terms of the vocabulary. The Principal Components Analysis is not appropriate in this case because the criteria talk about concepts that are not correlated.

[11] The list of possible receivers is obtained from the hospitals of the rest of the country following a complex hierarchical procedure that is detailed in [Aldea et al.,2001].

The transplants co-ordination prototype. Agentifying ClusDM.

A prototype of the system that could be used for the transplant co-ordinator is being developed. The system must deal with different distributed data sources, different knowledge levels and a complex set of rules and norms. For this reason, the multi-agent systems technology is particularly appropriate to solve this problem. Before giving more details about the prototype, let me introduce the concept of agent and multi-agent systems:

An *agent* [Wooldridge,2002] is a computer system capable of flexible autonomous action in some environment. An agent has its own goals and the tools to be able to achieve them. The main properties of agents are:

- *Social ability*: an agent must be able to communicate with other agents, and co-operate with them to solve complex tasks.
- *Reactivity*: an agent is aware of the changes in the environment and responds to them in a timely fashion.
- *Autonomy*: the agent may decide whether to fulfil a given request or not, and may decide which is the best way to achieve his goals.

There are particular problems that cannot be solved by a single agent because different resources, knowledge or tools are needed. In this case, agents must cooperate, co-ordinate or negotiate with other agents to achieve their goals. This is a Multi-Agent System (MAS) [Weiss,1999]. MAS are interesting for large and complex systems in several senses: (i) with geographically distributed data, (ii) with many components or entities, possibly with particular interests, (iii) with a broad scope and huge amounts of information to consider. The use of intelligent, distributed agents is a suitable approach for this type of problems.

In the case of organ transplants, the selection of the best receiver must be done in a very short period of time (only some hours) because they cannot be frozen as we do with tissues and bones. For this reason, there is a great interest in having a tool to help in this process. In collaboration with some hospitals we are designing a Multi-Agent System that will follow the national and international rules established for transplants and will try to model the process that is done at the moment, known as the Spanish co-ordination model [Matesanz&Miranda, 1995]. This model is centred in the figure of the Hospital Transplant Co-ordinator (HTC) of each hospital. In each hospital a list of people waiting for an organ is maintained. When a donor is recognised, the searching of the most appropriate receiver starts. This process involves different organisations that collaborate at different levels (see Figure 38). In our MAS we will respect this hierarchical organisation in order to find a receiver.

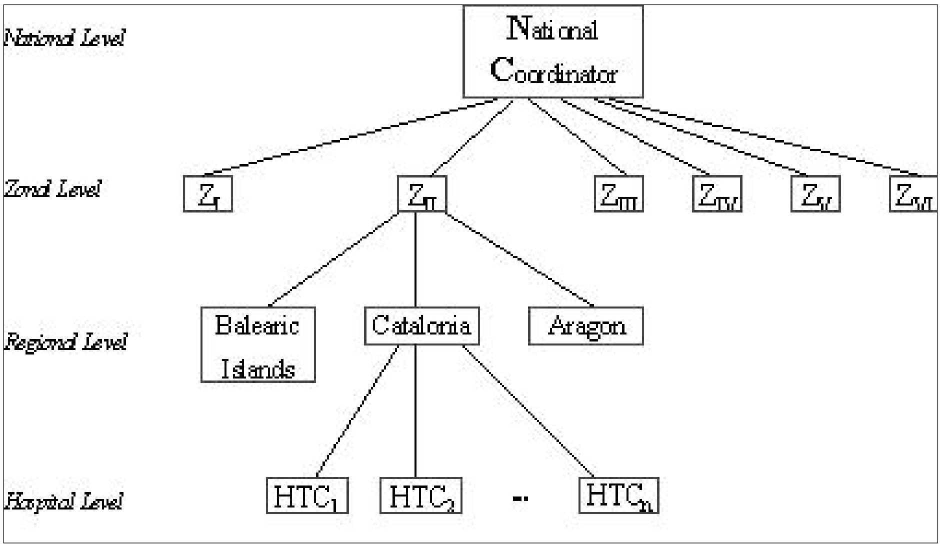

Figure 38. Hierarchical dependencies among transplant co-ordinators in Spain

Concerning the use of MCDM techniques, we must concentrate on the agents working inside a particular hospital. In Figure 39 we can see the architecture of the internal MAS of a hospital. The doted square contains the agents that belong to the same hospital.

The agent that is in contact with the medical personnel (and, in particular, with the hospital transplant co-ordinator) is called TCA (Transplant Co-ordinator Agent). This agent receives the characteristics of the donor and the organ that can be transplanted and starts the process to recommend the best possible receivers. First of all, he searches in the local database for potential receivers at the same hospital. This search is made with the help of the Medical Database Wrapper, which is the agent that is in charge of the access to the hospital database. Then, TCA sends a request for other candidates to the other TCAs in the same region (via the regional co-ordinator) or to the same zone (via the zonal co-ordinator). If no adequate receiver is found, he continues the search to other regions or zones, following some fixed rules.

When TCA has obtained a list of candidates, he sends them to the Transplant Specialist (TS) agent, which has knowledge about the field of organ transplants. This agent discards the patients that do not fulfil some basic compatibility conditions (e.g. the blood types of the donor and the potential receiver are not compatible). After this initial filtering, TS compares the attributes of the donor with the ones of the candidates and builds a preference matrix using qualitative criteria, which is sent to an agent that is able to execute a MCDA method.

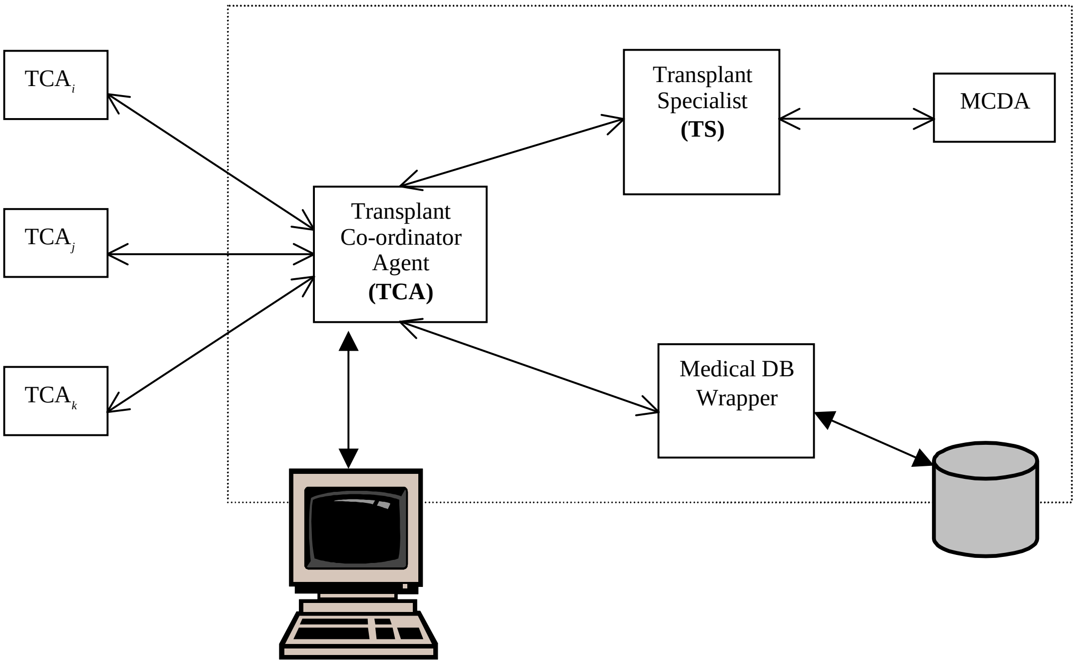

Figure 39. Intra-hospital multi-agent system architecture

We have already implemented an agent called *ClusDMA (ClusDM Agent)*. It is an agent that offers a very specific service: ranking a set of alternatives using multiple criteria. This service may be requested by agents that have to solve a decision problem or agent that have to aggregate heterogeneous data.

The *agentification* of a method consists of building an agent that is basically specialised in using this method properly. The agentification of MCDA methods is useful because multi-criteria decision aid is not a simple task. In the real world, it is usually done by experienced analysts who know how to apply the methodology and how to interpret the obtained results. Moreover, not all the MCDA methods can be successfully applied to the same kind of problems, it depends on the properties of the method and the characteristics of the problem. Thus, the first question that the analyst has to solve is the selection of which MCDA technique to apply. In a multi-agent system we could have different MCDA agents, where each of them was an expert in a particular technique. In [Valls&Torra, 2002a] we argue that it may be useful to generate agents that are experts in solving MCDA problems. These agents would receive requests of any other agent that has to face a decision problem, regardless of the particular application or multi-agent system to which it belongs.

In this case, ClusDMA, has the three properties that define an agent:

- ClusDMA is able to communicate with the other agents engaged in the solution of a more complex process.
- ClusDMA aborts the clustering process if the time at his disposal is near to expire, in which case the result is the ordering of the classes built up to that moment. On the other hand, if it detects that the result will not be good (quality measurement stage), *ClusDMA* aborts the process and communicates it to the requester.
- At the reception of a request, ClusDMA can decide if he will make the ranking or not, depending on the characteristics of the message received (the data matrix is correct, the information about the semantics of the criteria is correct and the amount of data is tractable); thus, it also shows a certain degree of autonomy.

ClusDMA has been implemented using Jade, which is a collection of Java libraries that ease the implementation of FIPA-compliant[12] multi-agent systems. We are running this prototype on Windows in standard PCs, although it could be used in any other platform that supports Java. In Figure 40 we can see the interface that the Transplant Co-ordinator Agent shows to the hospital transplant expert. Before requesting the ranking of the set of patients, the user must assign a weight to each criterion. After processing the patients' data (which is stored in the file indicated at the top of the window) the system will display the result at the bottom. Then, the user can express his agreement with this result using the vocabulary listed at the bottom-right side. This information will be interesting to evaluate the ClusDM methodology when the system works with real data. Moreover, we have prepared the system to easily include other agentifications of MCDA methods in order to compare their performances.

This MCDA agent will be working with the multi-agent transplants co-ordination prototype (shown in Figure 38). At the moment, we have done some local tests with artificial data.

[12] FIPA (*Foundation for Intelligent Physical Agents*) is a non-profit association that provides internationally agreed specifications for developing agent-based applications.

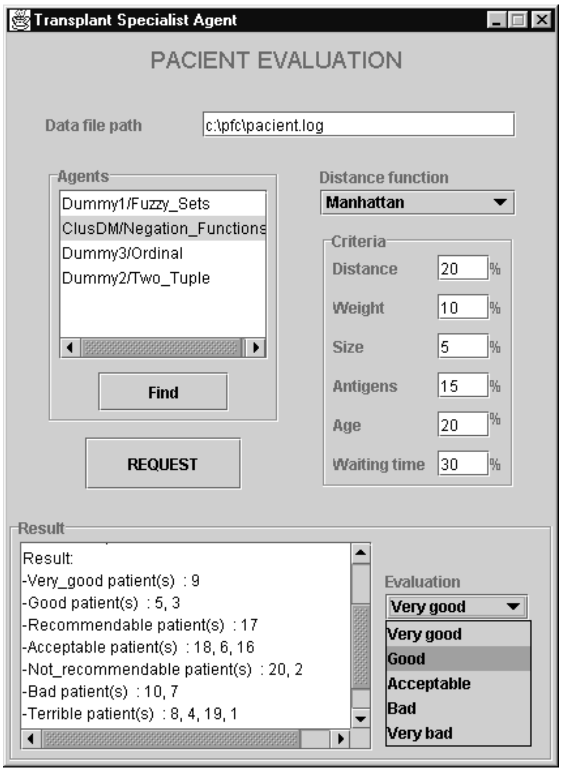

Figure 40. Transplant Co-ordinator Agent Interface

ClusDMA has different parameters (user's preference vocabulary for the description of the result, number of desired groups), which can be fixed in advance, in order to facilitate the use of the system by the transplant co-ordinator. Moreover, we do not show to the user the degree of quality of the result, because we do not want to influence his evaluation of the list of patients. Remember that this is only a prototype to test the possibility of obtaining automatic recommendations in such a critical medical problem. Therefore, we want to be careful in evaluating the appropriateness of the use of MCDA methods.

7.3 Statistical Disclosure Control

In recent years, the so-called information explosion has caused the development of new techniques for data analysis and information management. One class of techniques where this improvement can be found is the one related with information fusion and knowledge integration. As the number of available information sources and the amounts of information increase, the need of these techniques also increases. Applications of these techniques are now as diverse as scientific fields. One of the particular applications of information fusion techniques is Statistical Disclosure Control [Doyle et al., 2001].

National Statistical Offices (NSOs) are devoted to collect information from respondents and to their posterior publication. In fact, data dissemination is a requirement for National Statistical Offices as is the main justification for the resources spent and of their existence. However, data dissemination is usually a sensitive task because of re-identification risk. National Statistical Offices should process data prior to publication so that published data ensures that particular individuals or organizations cannot be re-identified. This is, no sensible data is published in a way that can be re-identified with a particular respondent (see [Torra, 2000] for a review of re-identification methods). Thus, data has to be protected (this is the so-called disclosure control problem) to avoid possible re-identification. Failure of protection can cause major problems due to legal norms and because respondents would refuse to new collaborations with the NSOs.

To avoid disclosure, masking methods are applied (see [Domingo&Torra,2002b] for a comparative study on masking methods performance). Masking methods introduce distortion to the data prior to its publication so that the information is not disclosured. Distortion should be kept small so that published data is valid for researchers and users (they can infer the same conclusions that would be inferred from the original data) but on the other hand should be protected enough so that disclosure is not possible. Statistical Disclosure Control (SDC) studies methods that attempt to perform such a nontrivial distortion.

When different microdata methods are applied to the same original file, different masked files are generated. In some cases, multiple protected versions of the same confidential data set are released, each one protected to minimize information loss for a particular use. In this case, an additional thread for re-identification risk appears due to the formation of coalitions of users. This is so because data fusion techniques can integrate the information contained in n different distorted versions of the data set. Thus, compromising statistical confidentiality. Note that, the better the reconstruction, the larger the disclosure risk. This suggests that data fusion tools can be applied to multiple masked data files to evaluate to what extent the original data file can be reconstructed.

We have studied the problem of fusing categorical data and evaluating the reidentification achieved [Valls&Torra, 2002b], [Domingo et al., 2002b]. We propose the use of the ClusDM as fusion (i.e. aggregation) method for categorical values in order to evaluate the degree of reconstruction achievable in this kind of data.

ClusDM as a fusion operator in Statistical Disclosure Control

We have considered a situation in which several masked versions of a single variable have been published. Our goal is to know if we can re-construct the original values of the alternatives from the fusion of the different releases.

For the application of the system to the SDC problem, we assume that each masking method corresponds to one criteria and that the aggregated criterion obtained by our system corresponds to an approximation of the original values.

To test the behaviour of ClusDM with heterogeneous qualitative variables, we have used 20 records extracted from the *American Housing Survey* of 1993 [Census Bureau, 1993]. Seven releases of the *Degree* variable have been generated using the most common masking methods for categorical data: Top coding (*Tp*), Bottom coding (*Bp*), Global recoding (*Gp*), Rank swapping (*Rp*) and Post-Randomization method (*Pp*). Different parameterisations have been considered, whose value is indicated by *p*. Parameterisations are based on the study of performance comparison for different masking methods with respect to information loss and re-identification risk [Domingo&Torra, 2002a].

Let us now briefly explain the masking methods we have included in this study:

- **Top Coding**. This method, applicable only to variables in ordinal scales, consists of the recoding of the highest *p* values of the variable into a new category. We have used the symbol '&' to denote the new term of the vocabulary, which substitutes the *p* values fusionated. A recoding of 4 categories has been considered (T4 in Table 28). Top coding is applied to avoid the re-identification of largest values as they are frequently easy to re-identify.

- **Bottom Coding**. This masking method is similar to the previous case but now the lowest *p* categories are recoded into a new one. As before, we have selected *p*=4 and the new category is codified by '&'. As in the case of Top coding, this masking method is applied to avoid the re-identification of the smallest values when the availability of this information allows the re-identification of the individuals.

- **Global Recoding**. Global recoding consists of the recodification of some categories by some other ones. Selection of categories is done on the basis of increasing the number of individuals that match a particular category. For example, if there is a record with "Marital status = Widow/er" and "Age = 17", global recoding could be applied to "Marital status"' to create a broader category "Widow/er or divorced", so that the probability of the above record being unique would diminish. In our experimentation, the following parameterisation has been considered: recode the p lowest frequency categories into a single one. We have used p=4.

- **Rank Swapping**. This method is better explained from their operational point of view. First, values of variable c_i are ranked in ascending order; then each ranked value of c_i is swapped with another ranked value randomly chosen within a restricted range (e.g. the rank of two swapped values cannot differ by more than p% of the total number of records). We have used p=10%.

- **Post-RAndomization Method or PRAM**. This is a perturbative probabilistic method in which the value of a given individual is changed according to a prescribed probability mechanism (a Markov Matrix). This method reduces the number of matching for all categories (reduction depends on the Markov matrix). The selected Markov matrix is based on the approach described in [Kooiman et. al., 1998]. This approach is as follows: Let $T_V=(T_V(1), ..., T_V(K))^t$ be the vector of frequencies of the K categories of variable V in the original file (assume without loss of generality that $T_V(k) \geq T_V(K) > 0$ for $k < K$) and let θ be such that $0 < \theta < 1$. Then the PRAM matrix for variable V is defined as:

$$p_{kl} = \begin{cases} 1-\theta T_V(K)/T_V(k) & \text{if } l = k \\ \theta T_V(K)/((K-1)T_V(k)) & \text{if } l \neq k \end{cases}$$

  In our example we have considered different parameterisations p = 4, 8, 9 and $p=10\theta$.

The original variable *Degree* have a negation-based linguistic vocabulary with 7 terms: L={*coldest*, *cold*, *cool*, *mixed*, *mild*, *hot*}. In Table 28, the original values are replaced by the position of the category in the set L, for the sake of clarity. Thus, value 1 stands for *coldest*, 2 for *cold*, 3 for *cool* and so on. The first column corresponds to the identifier of the record (i.e. alternative), the second column is the original value of the variable (o.v.), columns 3-9 are masked variables, column 10 is the aggregated value (a.v.).

	o.v.	B4	T4	G4	R10	P8	P9	P4	a.v.			o.v.	B4	T4	G4	R10	P8	P9	P4	a.v.
a	3	&	&	3	3	3	3	3	3		k	3	&	&	3	4	3	3	3	3
b	3	&	&	3	2	3	3	3	3		l	3	&	&	3	2	3	3	3	3
c	3	&	&	3	3	3	3	3	3		m	3	&	&	3	3	3	3	3	3
d	3	&	&	3	3	3	3	3	3		n	2	&	2	2	2	2	2	2	2
e	4	4	&	4	4	4	4	4	4		o	3	&	&	3	3	3	3	3	3
f	4	4	&	4	4	4	4	3	4		p	2	&	2	2	2	2	2	2	2
g	4	4	&	4	3	4	4	4	4		q	3	&	&	3	3	3	3	3	3
h	4	4	&	4	4	4	4	4	4		r	5	&	&	n	5	3	3	4	3
i	4	4	&	4	4	4	4	4	4		s	2	&	2	2	2	2	2	2	2
j	1	&	1	n	1	1	1	1	1		t	2	&	2	2	2	3	4	2	2

Table 28. Records used in the re-identification test

Using the classifier *Sedàs* we obtain the dendrogram in Figure 41. Then, a cut level has to be selected in the tree to obtain a partition of the elements. The cut is done so that the number of clusters is equal to 4 because this is the average number of linguistic labels used in columns B4-P4. This cut is also displayed in Figure 41. The obtained partition is defined by 4 sets (named α, β, δ and γ) as follows: α={n,t}, β={a,b,k,r}, δ={j}, γ={e,f,g}. This partition satisfies the conditions required in [Domingo et al., 2002] for a correct partition selection in this context:

- records with all the variables with the same value should correspond to different clusters (e.g. record a and e),
- clusters should be defined according to the dendrogram.

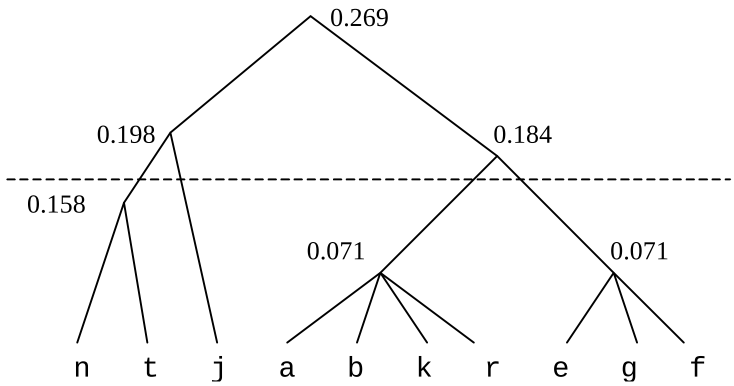

Figure 41. Dendrogram for the clustering of the statistical data

Note that for the sake of simplicity, we only include in the dendrogram and in the partition one of those elements that are indistinguishable (i.e., it appears the element a but does not appear c because it has the same values for all columns).

The 4 clusters obtained have been ranked using the similarity to the best possible alternative, the one that has the largest value for all categories. The ranking with PCA was not possible because the stopping criterion selects two axes instead of a single one.

Following the explanation stage process, each class is given a category from the original vocabulary, L: δ is *coldest* (1), α is *cold* (2), β is *cool* (3) and γ is *mixed* (4). This result can be seen in Table 28. The goodness values for each step are the following: 0.8 for the aggregation, 0.93 for the ranking, 0.63 for the vocabulary and 0.95 for the representation of the clusters by the new semantics. The overall confidence on the result is 0.83. The lowest quality is for the stage of building the new vocabulary and semantics, this indicates that the meaning of the terms has been changed with respect to the original one.

In this example, we can see that we have re-identified the original value of each record, except one, record `r`. Therefore, the publication of these 7 versions of the same data is very dangerous due to the proved ability of discovering the original value of the variable. However, although we re-construct the “labels” attached to the records, we have seen that the semantics of the terms is not completely re-identified.

We can see that the ClusDM methodology can be successfully used as an aggregation or fusion operator for the re-identification of statistical data. However, we must mention that other tests have not been so good. If the masking methods produce some outliers, the clustering method would create a class for those “different” records, which will make that the rest of records ought to be put in a smaller number of clusters than it should. In this case, the reconstruction is more difficult. Nevertheless, we can detect this situation using the information provided by ClusDM about the conflicting records (the outliers), and we can repeat the process increasing the number of clusters in the partition or removing the outliers from the analysis (as we have shown in section 7.1).

CHAPTER 8. Summary and Future Work

After having explained in detail the difficulties of multiple criteria decisions, the different approaches to facilitate the work of the decision maker and having presented a new methodology called ClusDM, this chapter is devoted to review the main characteristics of our approach and to give some future research lines to improve it.

8.1 Summary and conclusions

This thesis proposes a new methodology for Multiple Criteria Decision Aid. This work is the result of some years of research in order to develop a method to deal with complex multicriteria decisions. The difficulties that we have faced up are the management of: criteria of different nature (numerical, qualitative and Boolean), different scales (different vocabularies or categories) and missing values. We considered a new approach to this heterogeneous data that does not require the transformation of all the values into a unified scale. The use of clustering to perform the aggregation of the values has been proved to be a good solution for the integration of heterogeneous data. The approach based on similarities (inherent in clustering techniques) allows us to compare the alternatives and understand the relationships among them. These global preference relationships cannot be found if we assume the independence of the alternatives, as classical MCDA methods do.

The second aim of our research is the development of tools that the user can understand and apply easily. The negation-based semantics seems quite appropriate for this purpose because it is based in the *antonym* concept, which is nothing new for people. During chapter 5, we have seen how to use this semantics representation to attach a suitable linguistic label to each alternative. This is a crucial point, because the decision maker will base their final decision on these values, having into account that the relative preferences over these values are expressed by the negation function attached to its vocabulary.

Nevertheless, we have gone a step further. We have analysed in detail the process in order to extract useful knowledge about the elements of the decision framework. Conflicting alternatives and criteria are detected and presented to the decision maker. This additional information together with a quality evaluation of the process is of great value for understanding and successfully applying the solution obtained.

To end this overview, we would like to mention some drawbacks of ClusDM. The first one is that a minimum number of alternatives are required to obtain sufficiently good results. As alternatives are compared with each other during the first stage of the process, if the number of alternatives is small (e.g. less than 7) the clusters will not be very significant because their number of elements will be low. Therefore, ClusDM is a good method to be used in decision problems that involve a large set of alternatives. However, a second drawback is related to the first: if we study the temporal complexity of the ClusDM process, we can see that it is $O(m,p) = m^2 p$, being m the number of alternatives and p the number of criteria. That is, the number of alternatives in the set has great influence in the time consumed by the process.

8.2 Future directions

In chapter 2, some methods that work with uncertain information in MCDA have been presented. Some of them are able to deal with heterogeneous data sets. However, as it has been pointed out before, they perform a transformation of the original values of the criteria into a unified framework, where the decision analysis is done. Some of them define processes to put different linguistic vocabularies into a common one but do not consider the case of including numerical data, others handle the possibility of having qualitative and numerical criteria in the same decision matrix. The two cases are interesting in order to study the behaviour of ClusDM in a qualitative framework with different vocabularies or in the case of mixing numerical and qualitative criteria.

In Table 29 we have made a classification of the methods presented in section 2.5 in terms of the possibility of managing different qualitative vocabularies or mixed types of criteria versus the type of semantics given to the linguistic terms. An exhaustive comparison of these methods would be interesting.

	Many Vocabularies	numerical + Qualitative
Fuzzy Sets	LOWA (after translations explained in [Herrera et al.,2000b])	LOWA (after translations explained in [Delgado et al.,1998])
Negation functions	ClusDM Antonyms-based aggregation	ClusDM
2-tuple	2-tuple Weighted Mean 2-tuple OWA (after translations explained in [Martínez,1999])	2-tuple Weighted Mean 2-tuple OWA (after translations explained in [Herrera&Martínez,2000a])
Ordinal scale	QWM (after translations explained in [Torra&Godo,1999])	

Table 29. Aggregation operators for heterogeneous criteria

Methods for ordinal linguistic values given in section 2.5 can be classified according to the kind of semantics they deal with: explicit semantics (like the use of fuzzy sets or negation functions), implicit semantics (like the 2-tuple linguistic values) or direct computation on the ordinal scale. For applying the methods based on fuzzy sets, we must know the fuzzy membership function for each term. Considering that qualitative criteria have a negation-based semantics, we can use the intervals induced by the negations to build the fuzzy set corresponding to each linguistic value. Following [Yuan&Shaw,1995] (as it has been done in the explanation stage, section 5.1.3), the centre of each interval may be the point of maximum membership to the corresponding term. The rest of the triangular membership function is defined by these points. Obviously, with this approximation, there is a modification of the information given by the terms, which will influence the results when the two semantics are compared.

Once we have a set of criteria described using fuzzy values, negation functions and 2-tuples, we could use the methods in Table 29. In [Zimmermann,1990] some guidelines to compare and classify MCDA methods are given. Zimmermann mentions 5 different dimensions: generality (i.e. the degree of general applicability of the method), discrimination (i.e. the capability of differentiating alternatives with slightly different values), fuzzification (i.e. treatment of uncertainty), information requirements (i.e. if the method needs a standard representation of the inputs) and sophistification (i.e. mathematical complexity). In a recent book ([Triantaphyllou, 2000]), the author compares some classical methods of the utility-based and the outranking approaches. For example, he makes a comparison of the methods in terms of two evaluative criteria: (i) an MCDA method that is accurate in multi-dimensional problems should also be accurate in single-dimensional problems and (ii) an effective MCDM method should not change the indication of the best alternative when an alternative (not the best) is replaced by another worse alternative. Moreover, we should also make comparisons of the goodness of the result obtained for problems with a known solution.

After this comparative analysis, we would like to perform more accurate tests in the application domains presented in chapter 7. For example, we can obtain more statistical public data or we can use ClusDM in other journal or conference reviewing process. In addition, we are considering other application domains. In fact, we are developing a multiagent system to help companies to make personnel selection (an initial prototype is explained in [Batet,2002]).

With the use of our methodology in various domains, we could improve the explanation stage. We pretend to present to the decision maker a more user-friendly view of the quality measures and knowledge extracted during the process. The use of natural language will be of great interest as is argued by people ([Greco et al., 2001], [Bana e Costa, 1990]).

Another future research line is the adaptation of the ClusDM methodology to deal with dynamic environments. As [Olson et al., 2001] pointed out, decision making problems usually deal with changing elements. It would be interesting that ClusDM could include or drop alternatives during the process. In fact, the modification of the alternatives would affect the clusters obtained in the aggregation stage. If the modification of the alternatives set is done after the aggregation, the prototypes should be recalculated before the ranking stage, because both the Principal Components Analysis and the Similarity-based ranking are based on the prototypes values. Once the ranking stage has finished, the inclusion and deletion of alternatives should be carefully studied, for its implication not only in the vocabulary and semantics but also in the ranking itself.

In the same line, it is possible to have to evaluate new alternatives after the analysis of the initial data set. In this case, we would like to study the work on automatic rules generation [Riaño,1998]. Then, the partition induced by the new preference-ordered qualitative criterion given by ClusDM could be explained using if-then rules. These rules will be used to classify (to know the linguistic preference value) corresponding to a new alternative.

The results obtained until now encourage us to continue our work in MCDA. We hope to be able to develop interesting solutions for the open-problems outlined in this section.

APPENDIX A. Review Form of the Journal

The following review form was designed to test the ClusDM methodology. We adapted the model provided by the editors of the journal. We included different types of criteria: numerical, ordinal qualitative and categorical.

The form is divided up in 6 parts; the first one identifies the paper (this information is not used in the reviewing process) and provides some information about the main characteristics of the work, which are two of the criteria included in the analysis. The following two sections (A and B) are devoted to evaluate the content and presentation of the paper. All the questions were considered as qualitative preference criteria in the test, except for question A.4 that is answered in natural language. For the same reason, section C could not be included to the ClusDM analysis. Section D shows the overall evaluation of the decision makers, which was included as another ordinal qualitative criterion. Finally, the last section identifies the reviewer and their confidence on the subject. This additional information was not considered in the test.

AUTHOR(S):

TITLE:

The paper reports on: [ ] A methodology   [ ] Applications

The emphasis of the paper is on:
[ ] Preliminary research
[ ] Mature research, but work still in progress
[ ] Completed research

A. CONTENT

1. How relevant is the content of the paper to the theme of the special issue? (is the described system really composed of "AGENTS" that are applied in an interesting way to any aspect of "HEALTHCARE"?)

[ ] Very Relevant   [ ] Quite Relevant
[ ] Somewhat Relevant   [ ] Not Relevant

2. Is the paper really concerned with "agents" (i.e. autonomous, intelligent, communicative, cooperative, proactive entities)?

[ ] I definitely think so.
[ ] It might be arguable whether "agent" is the best expression for the elements of the described system.
[ ] I have strong doubts regarding the usage of the word "agent" in this work.
[ ] The paper abuses the use of the word "agent", i.e. it tries to "agentify" an otherwise standard AI application.

3. If the paper describes a multi-agent system, are communication, co-ordination and/or negotiation techniques described?

[ ] They are well described.
[ ] They should be explained in more depth.
[ ] They are not clearly explained.

4. What are the main contributions of the paper?

5. How original is the research reported?

[ ] Very Original [ ] Mostly Original
[ ] Somewhat Original [ ] Mostly Unoriginal [ ] Not Original

6. Quality of the Research:

Is the research technically sound? [ ] Yes [ ] Somewhat [ ] No

About the technical limitations/difficulties ...
[ ] They are adequately discussed
[ ] They are briefly discussed
[ ] They are poorly discussed
[ ] They are not discussed

About the approach ...
[ ] It is adequately evaluated
[ ] It is briefly evaluated
[ ] It is poorly evaluated
[ ] It is not evaluated

7. For papers focusing on applications:
(Give a numerical evaluation from 1 to 5, 1 is the worst value)

Is the application domain adequately described? (1..5)
Is the choice of a particular methodology discussed? (1..5)

8. For papers describing a methodology:
(Give a numerical evaluation from 1 to 5, 1 is the worst value)

Is the methodology adequately described? (1..5)
Is the application range of the methodology adequately described, e.g. through clear examples of its usage? (1..5)

B. PRESENTATION

(Give a numerical evaluation from 1 to 7, 1 is the worst value)

1. Are the title and abstract appropriate? (1..7)
2. Does the introduction show the intentions of the paper and presents the rest of the article? (1..7)
3. Does the last section give the conclusions or the most relevant results of the work? (1..7)
4. Is the paper well organized? (1..7)
5. Is the paper easy to read and understand? (1..7)
6. Are figures/tables/illustrations sufficient? (1..7)

7. Is the paper free of typographical/grammatical errors?
[ ] The English is correct
[ ] There are some typographical errors
[ ] There are some grammatical errors
[ ] There are both typographical and grammatical errors
[ ] The English is deficient

8. Is the references section complete?
[ ] Yes
[ ] The basic work is referenced but recent work is not.
[ ] There are missing relevant basic references
[ ] It is very poor

C. SUGGESTED/REQUIRED MODIFICATIONS & ANY OTHER COMMENTS

D. OVERALL

The paper ...
[ ] is definitely recommended for inclusion in the special issue
[ ] is recommended for inclusion in the special issue after a few modifications
[ ] could be recommended for inclusion only after important modifications
[ ] is interesting, but not mature enough to be included in this issue
[ ] is definitely not recommended for inclusion in the special issue

REVIEWER'S NAME:
Reviewer's confidence in the subject area of the paper: [ ] High [ ] Medium [ ] Low