Design of Fuzzy Cash Flows Applying Most Typical Values to a Case-Based Reasoner Outcome

Rosina Weber-Lee

Alejandro Martins

Roberto Pacheco

Ricardo Miranda Barcia

rosina@eps.ufsc.br

Federal University of Santa Catarina, Industrial Engineering,

Florianópolis, SC, BRAZIL.

Abstract. When dealing with economic decision making, (e.g., financial decision making, budgeting, business feasibility evaluation), one always needs to model cash flows that are uncertain by nature. Due to the lack of information, one has to rely on expert's knowledge to perform such task. Experts use their expertise that combines knowledge and experiences within the context. We propose a system that builds a fuzzy cash flow from the outcome of a Case-Based Reasoning (CBR) system . This outcome is a set of numeric values where we calculate the Most Typical Values (MTV). The CBR system suggests a set of estimated values, appraising cash flow accounts. The system selects the values that better represent the given set using MTV approach, automatically creating Most Typical Fuzzy Sets describing values such as "around $500.00". The content of the fuzzy cash flow consists of actual numbers (provided by certain liabilities and receivables), stated values (such as production targets and sales forecasts) and fuzzy constraints. The actual and stated values are combined with the fuzzy constraints with the purpose of building fuzzy cash flows to support financial decision making.

1. Introduction

Cash flow modeling is vital for financial decision making. Due to the lack of information, cash flows are estimated by experts that use their experience. Case-Based Reasoning (CBR) parallels humans in searching for past experiences that are similar to the current one. In the predict task, the reasoner parallels the human act of searching for a past similar experience to infer for the future. One recognizes that a similar situation has occurred in the past and then adapts the consequence of this experience to the current problem. Consequently CBR is the appropriate choice to estimate cash flows automatically.

Uncertainty present in cash flows is not random, and, like Bellman & Zadeh pointed out, (1970) the use of probabilistic techniques implies the presence of randomness. The cash flow accounts are present in the human discourse through imprecise expressions such as "around two thousand". This type of imprecision derives from ill-defined boundaries to the value indicated, that is fuzziness. We find in Fuzzy Set Theory the appropriate means to model the imprecision and uncertainty present in cash flows.

Weber-Lee, R.; Martins, A.; Pacheco, R.; Barcia, R.. (1996). Design of Fuzzy Cash Flows Applying Most Typical Values to a Case-Based Reasoner Outcome. Third Congress of the International Association for Fuzzy-Set Management and Economics, November 11-13, Buenos Aires, Argentina, 1996.

Financial decision making has been object of concern of the researchers of Fuzzy Set Theory for some time. In 1970, Bellman & Zadeh published "Decision-Making in a Fuzzy Environment" and this was one of the first lights indicating the use of Fuzzy Set Theory on financial decision making. Buckley, (1987), proposed fuzzy analogues for compound interest problems, the fuzzy future value and the fuzzy present value. In Ward, (1989), it is presented a fuzzy discounted cash flow analysis. In 1992, Buckley solved fuzzy equations for the economic and financial environments. Recently Chiu & Park, (1994) proposed an approach to analyze fuzzy cash flows using the present worth criterion. All these publications develop mechanisms to deal properly with fuzzy variables in the financial decision making. However, it is necessary to bridge these works and propose tools capable of solving the whole problem in a single system using the solutions already proposed.

When representing fuzzy concepts such as "around 5", Chiu & Park, (1994) use triangular fuzzy numbers while Ward (1989) use flat fuzzy numbers. The latter author points out to the need of the development and validation of techniques for estimating membership functions for fuzzy cash flows. In this paper, we introduce our proposal for creating membership functions for fuzzy cash flows.

Presenting an extension of the works mentioned above and aiming to bridge the gaps left, we propose a system that starting from past data, estimate cash flow accounts and fit them into the fuzzy cash flow. The main advantage in our approach is that the impreciseness and uncertainty embedded in the process is maintained and it is conveyed to the fuzzy cash flow, with the consequent avoidance of losing relevant information. As an extension to this contribution, we propose the construction of Most Typical Fuzzy Sets (MTFSs) to model cash flow accounts.

We use the CBR paradigm to build the reasoner that estimates cash flow accounts. The result of the reasoning is evaluated using the Theory of Typicality (TT) and we calculate the Most Typical Value (MTV) for this outcome. From these values we construct MTFSs to represent the outcome of the reasoner. These sets can be directly used as elements of the Fuzzy cash flow as they represent fuzzy constraints.

These days, the continuity of the business plays a much more important role than in the past. The concern with maximizing profits has given place to the concern of maximizing and maintaining profits. This places the design of cash flows in an even more relevant position as its contribution becomes more decisive to the health of the business. The availability of an automatic tool to design fuzzy cash flows that uses all the uncertainty without losing relevant data increases value to the cash flow information and its capability of helping the success of the business.

In next section there is an explanation about why CBR is an appropriate tool to estimate accurate cash flow accounts. Next, we present how the reasoner works with examples. Then, we introduce the TT and the reasons to use it within our application.. We illustrate the whole methodology with examples. Interpreting the results we describe the construction of the MTFSs. We conclude indicating how these sets can be used in fuzzy cash flows in fuzzy decision making environments.

2. Assessing Cash Flow Accounts

Building cash flows for financial decision making is a task usually required in several situations. One is a regular task performed in firms on a daily basis and it is an attempt to foresee the actual balance within some time to properly make decisions that maximize profit and minimize risk. Capital investment decisions have to be grounded on cash flows. The decisions related to capital investments can be oriented to feasibility evaluation, capital expenditures for projects or expansion strategies.

There are many reasons to explain the difficulty and the importance of estimating accurate cash flows. Regardless the segment the firm does its business (e.g., industrial, commercial, services) or the political and economic context (clients in multiple countries) there are several conditions that induce to fluctuations that are sometimes unexpected or unobserved. Seasonal factors of the firm, its clients or suppliers; months with different number of weeks and days; terms of some payments differ from the accrual to the actual disbursements; sudden purchase of items from suppliers offering great discounts; necessity of financing of funds, etc. Besides these issues, corporate management must ensure daily liquidity so the firm is capable of paying its bills, keeping away risks of bankruptcy. Accurate cash flows represent the only tool able to support such tasks.

Building the cash flow is then a task of estimating inflows and outflows. Most frequently, experts build cash flows based on experience and their expertise. Considering that assessing cash flows is indeed a task performed using past experience, we choose the CBR paradigm to implement an automatic tool to estimate cash flow accounts.

3. The outcome of the Case-Based Reasoner

The case-based reasoner estimates cash flow accounts (Weber-Lee et. al., 1995). Our system performs the tasks of identification and prediction. The cases are cash flow accounts and they are represented through descriptors related to present and previous amounts, nature and period in time. The system receives as the target problem one cash flow account with the attributes properly assigned. The system searches throughout the case base to identify the most similar cases using a similarity metric that assigns scores representing the degree of matching between the target case and the case being evaluated. There are different levels of importance for the descriptors. The solutions to the target case are the amounts that occurred for these accounts in the past, after an equal period of time, as the period described in the target case. The solution is adapted by a *parameter adjustment* (Kolodner,1993) operation to fit to the target case. The solution adapted is the amount estimated for the account in the cash flow. The system suggests a solution that is an amount that is likely to happen the following period. The solutions of the cases in this set are candidate solutions. Conventionally, case-based reasoners use some technique to select one best match to provide the outcome or a combination of the solutions of most similar cases -- that are retrieved after being compared to the target case. In our approach, we offer an automatic technique that reaches the solution to solve the target problem evaluating this set of solutions to find out what values represent this set (the typical values) calculating the MTV of the set provided by the most similar cases.

Hence, the outcome of the case-based reasoner consists of numbers that represent estimates of what the expert believes will happen. From this point, we illustrate our proposed approach with examples ran in the case-based reasoner. One complete iteration of our case-based reasoner follows:

Given one target case (Figure 1):

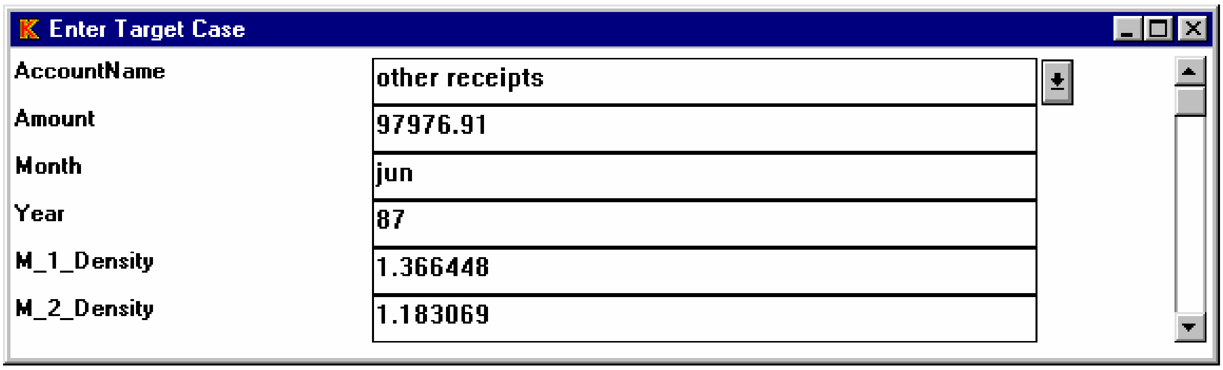

Figure 1. Part of the dialog box with the information on the target case.

The target case is the problem the reasoner is supposed to solve. In this predict application, the solution is an estimate value for the next period. The target case shows an account other receipts from June 1987, see Figure 1. The solution expected is an estimate value for the following period, July 1987.

The second step of the iteration is the retrieval. The reasoner retrieves a set of cases based on a similarity metric, assigning a score for each case on the case-base. The cases retrieved are presented in order of best scores, as it can be seen in the Retrieved Case List in Figure 2.

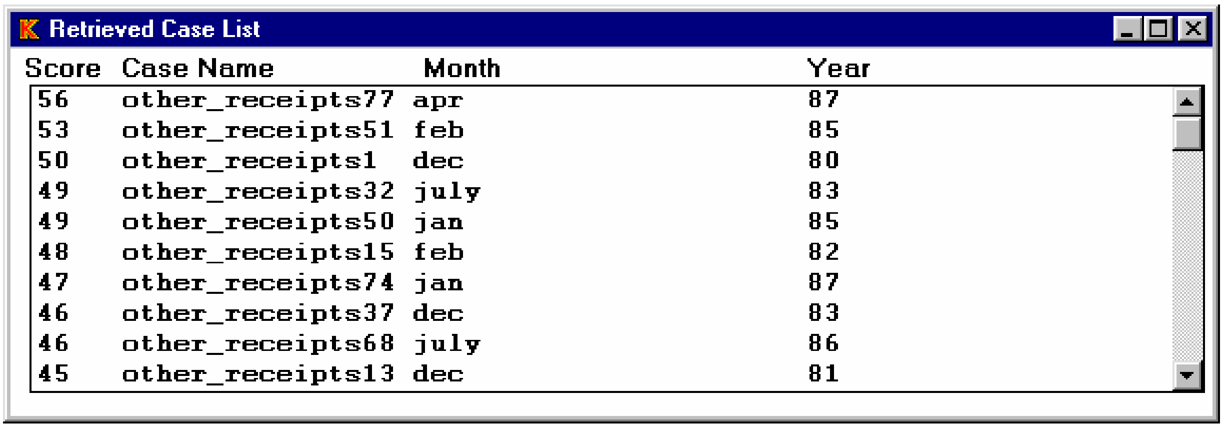

Figure 2. Part of the dialog box with the list of retrieved cases.

This list may contain as many cases as the threshold allows. In this example, the threshold is set for 30%. Very similar cases may have a score around 60%. As an average, for this application, the scores range around 45%.

The next step is the Select phase (Aamot & Plaza, 1994), when one or more cases are selected to provide the solution to the target case. For a discussion of these procedures and our approach see (Weber-Lee et. al, 1996). In our approach, we select the 25 cases with the highest score and use them to calculate the outcome, what is demonstrated in section 4.

The iteration continues with the 25 cases selected going through adaptation phase. Adaptation is the step when the necessary adjustments are made to make the selected case absorb the target cases's functions or parameters. The adaptation employed uses the parameter adjustment method. After adaptation, the retrieved cases provide, each, one estimate to be the solution to the target case. Table 1 shows the 25 cases retrieved for the example above and the estimates.

This is the case-based reasoner outcome. Henceforth we will demonstrate our approach with three examples. I is shown in Table 2 the target cases, retrieved cases, scores, and estimates for the three examples.

#	nature	score	estimate

1	othrecp	56	113163
2	othrecp	53	177474
3	othrecp	50	44282
4	othrecp	49	110574
5	othrecp	49	79266
6	othrecp	48	109003
7	othrecp	47	108144
8	othrecp	46	322234
9	othrecp	46	113563
10	othrecp	45	66725
11	othrecp	44	108125
12	othrecp	44	102093
13	othrecp	43	99256
14	othrecp	43	128909
15	othdis	41	150964
16	othrecp	41	139318
17	othrecp	41	228258
18	othrecp	40	140481
19	othrecp	40	57072
20	othrecp	39	105032
21	othrecp	39	67751
22	othrecp	39	118193
23	othrecp	38	156232
24	othrecp	38	85429
25	othrecp	38	110396

Table 1. Estimates provided by the 25 cases retrieved.

	other	disb			repair &	maint.			other	receipts	
#	nature	score	estimate	#	nature	score	estimat	#	nature	score	estimate
1	othdis	71	44747	1	repmain	67	76557	1	othrecp	56	113163
2	othdis	69	48552	2	repmain	67	66149	2	othrecp	53	177474
3	othdis	68	48733	3	repmain	61	77973	3	othrecp	50	44282
4	othdis	66	32102	4	repmain	60	23551	4	othrecp	49	110574
5	othdis	65	87454	5	repmain	57	53027	5	othrecp	49	79266
6	othdis	63	55037	6	repmain	57	77460	6	othrecp	48	109003
7	othdis	62	43121	7	repmain	56	60366	7	othrecp	47	108144
8	othdis	61	54362	8	repmain	54	68007	8	othrecp	46	322234
9	othdis	61	45668	9	repmain	54	56774	9	othrecp	46	113563
10	othdis	61	51898	10	repmain	53	97839	10	othrecp	45	66725
11	othdis	61	38138	11	repmain	53	70627	11	othrecp	44	108125
12	othdis	61	51115	12	repmain	52	85766	12	othrecp	44	102093
13	othdis	59	49471	13	repmain	51	53105	13	othrecp	43	99256
14	othdis	59	44359	14	repmain	51	64322	14	othrecp	43	128909
15	othdis	58	40471	15	repmain	50	31875	15	othdis	41	150964
16	othdis	58	47250	16	repmain	49	80030	16	othrecp	41	139318
17	othdis	57	52390	17	admexp	48	75203	17	othrecp	41	228258
18	othdis	57	39333	18	admexp	48	87444	18	othrecp	40	140481
19	othdis	55	80211	19	repmain	48	206627	19	othrecp	40	57072
20	othdis	54	43426	20	repmain	48	140305	20	othrecp	39	105032
21	othdis	54	45734	21	repmain	48	72555	21	othrecp	39	67751
22	othdis	53	55481	22	repmain	47	68178	22	othrecp	39	118193
23	othdis	53	38566	23	repmain	47	87554	23	othrecp	38	156232
24	othdis	53	48007	24	repmain	47	55036	24	othrecp	38	85429
25	othdis	52	45735	25	othdis	46	66381	25	othrecp	38	110396

Table 2. Sets of retrieved cases for the three sets.

One may notice that among the retrieved cases for the accounts repair & maintenance and other receipts there are accounts from a different nature. This being a controversial position requires an explanation. The idea is to capture with the retrieved cases, situations that are similar to the one occurred in the target case. Some accounts, within a cash flow, keep similar behaviors since they are subject to the same movements and external facts. We consider that, with a very high weight on the feature account, if even though a case with a different account is retrieved with a fairly good score, then this case is eligible to be part of the construction of the solution.

Henceforth the estimates for other disbursements will be called as set A, estimates for repair & maintenance, set B, and estimates for other receipts, set C.

On the account of seeking for a value that is representative of the set of retrieved cases we propose an approach that uses the content of the sets to produce a reasonable solution. The nature of the data is not random, since it is strongly influenced by the economic conditions and top governmental decisions. Under these circumstances, our analysis leads us to prefer a fuzzy approach. We need to identify the typical behavior of the set of solutions of the retrieved cases. This is clearly a problem of representing a data set with a measure of general tendency.

The research on measures of general tendency in treating fuzzy sets led to the development of the Most Typical Value (MTV), the Most Typical Deviation (MTD), and the Definite Typical Value (DTV), (Friedman, Ming, & Kandel, 1995). At this point, considering the fuzzy nature of the data and the necessity of a measure of central tendency, our analysis shows clear advantages on choosing the TT.

An alternative to this approach is the probabilistic clustering (Cheeseman & Stutz, 1996). We found that this approach requires several assumptions that are not appropriate in financial environments. Particularly, in CBR applications, the dynamic nature along with small and variable data sets contradict the main assumptions of probabilistic approaches: the requirements for defining the shape of the functions, and the hypothesis of the probabilistic behavior of the attributes, and the need of the definition of the parameters of the probabilistic functions.

4. The Theory of Typicality

The advantage of obtaining a typical value of a data set grounds on the possibility of having a measure that is representative of this set regardless of central tendency. According to Friedman, Ming, & Kandel, (1995) traditional statistical measures fail on representing data sets containing more than one cluster since these measures are not always able to represent a typical feature of a given set. The TT calculates a MTV for the data set after grouping the data into clusters using a geometrical fuzzy clustering algorithm.

The expressions MTV, MTD, and DTV are given in Friedman, Ming, & Kandel, (1995). First, one may consider that one typical value of a data set might not exist, what is observed by the failure of the iterative procedure to provide a unique solution. This indicates the existence of multivalue MTV. For obtaining the MTV of a given n-dimensional fuzzy set, the set first undergoes a clustering process using a geometrical fuzzy clustering algorithm (Windham, 1983).

4.1 Geometrical Fuzzy Clustering Algorithms

In a fuzzy clustering, the requirements for clusters, classes and blocks are weaker than in classical ones and they generate fuzzy partitions. Hence, in a fuzzy clustering, the clusters are fuzzy subsets of a collection of elements. The clusters are membership functions of the elements in the cluster. Elements in the data set that are similar to each other are identified by the fact that they have high memberships in the same cluster. The memberships are chosen so that the sum for each element is one.

The choice for this type of clustering is embedded in the approach, nevertheless it has augmented our decision to use it, as the fuzzy clustering is able to describe ambiguities that often occur in real data, such as bridging objects and outliers (Rousseeuw, 1995). Besides, the fact that fuzzy clusterings are non-exclusive was also relevant (i.e., each element is assigned to one or more clusters with a degree of inclusion to each cluster in the partition), (Kim & Novick, 1993).

The elements to be clustered are represented by vectors in some d-dimensional Euclidean space. We have a set X={x_1,…x_N} $\subset R^p$ where the components of each vector are measurements of one of p features of a particular element. The measure of similarity between the elements can be characterized by a differentiable measure of distance between their corresponding data vectors, i.e. $|x_k - x_l|^2_M = (x_k - x_l)^T M(x_k - v_l)$, for some positive semidefinite matrix, M. Under these assumptions a cluster can be viewed geometrically as a region where the data points are highly concentrated or close together as determined by the metric.

The basis for constructing geometrical fuzzy clustering algorithms that are, in fact, an iterative procedure for choosing membership grades that minimize (Bezdek, 1981 and Kandel, 1982).

$$\sum_i \sum_k (u_{ik})^m |x_k - v_i|^2_M . \qquad (1)$$

4.1.1 Fuzzy c-Means Algorithm

Fuzzy c-means clustering is a geometrical fuzzy clustering algorithm and it differs from the fuzzy equivalence relation-based (the second type of fuzzy clustering) algorithm in the requirement of having the following three values defined beforehand:

1. number of clusters, c
2. a real number $m \in [1, \infty)$
3. a small positive stopping criterion number, ε

See (Windham, 1983), (Bezdek, 1981) and (Klir & Yuan, 1995) for further reading.

4.1.2 Cluster Validity and Supervised Tracking

Cluster validity refers to the appropriateness of a partition $U^{(t)}$ resultant from the clustering algorithm. According to Pal & Bezdek, (1995), it depends on what we mean by a good partition. Kim & Novick, (1993), understand that the partition should reflect the inherent organization of the data. They suggest to validate a clustering by verifying that for given c clusters, this organization yields a statistically significant improvement over $c-1$ and it is only marginally worse than c+1. Gath & Geva, (1989), present three requirements to define an *optimal partition*:

" 1) Clear separation between the resulting clusters.

2) Minimal volume of the clusters.

3) Maximal number of data points concentrated in the vicinity of the cluster centroid."

Unsupervised tracking is required when there is no a priori knowledge of the location of the centers for the initial partition. This subject demands proper attention as different initial partitions converge to different local optima. See Gath & Geva, (1989) for a scheme to select the initial clusters' centers. Under supervised clustering, Rousseeuw (1995) illustrates an example presenting an unidimensional data set graphically showing that this "visual system" can easily reveal the number of clusters.

Getting back to our examples, in trying to apply the fuzzy c-means, we face the first difficulty. We need to decide a number of clusters that might work for all our sets, as we cannot evaluate each result during the application. On this account, we have applied the fuzzy c-means algorithm to our three sets A, B, and C, using 2, 3 and 4 clusters to decide what would be the best number. We tried to ground our decision with the guidelines proposed by Gath & Geva (1989) and in trying to identify in the chosen number, an improvement over the others following the Kim & Novick's (1993) suggestion. The results are illustrated next on Figure 3, Figure 4, and Figure 5.

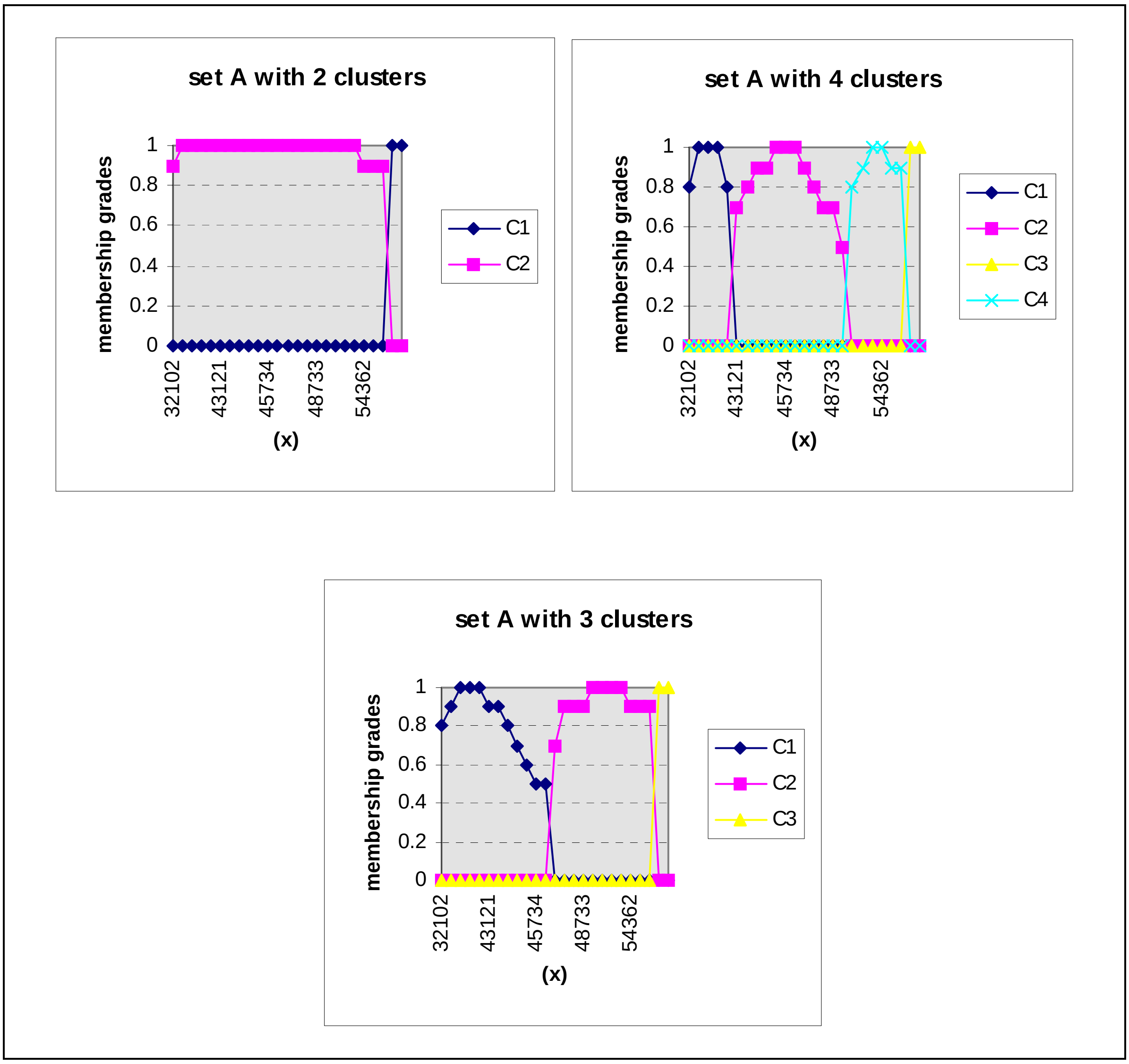

Figure 3. Resulting clusters for set A.

For set A (see Figure 3) it is clear the improvement obtained with three clusters over the result with two clusters. Increasing the number to four did not provide a significant improvement and the choice for the minimal volume of the clusters suggested three as a good option.

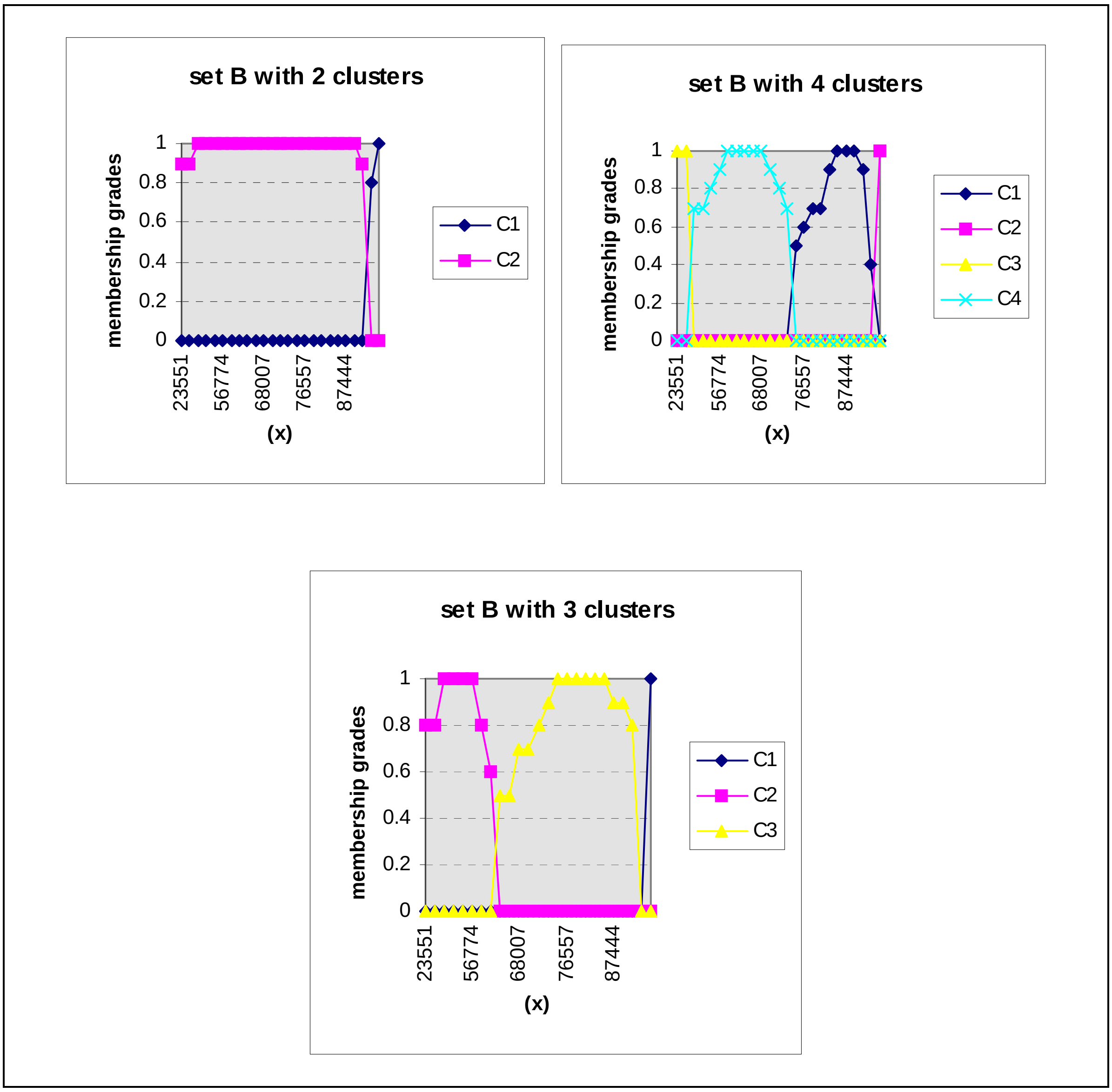

Figure 4. Resulting clusters for set B.

The graphics with the results on set B (Figure 4) seem somehow similar to example A. An improvement over two clusters is clear and again, the inclusion of another cluster did not reveal any improvement. Actually the separation between the clusters looks clearer in the graphic with three clusters, confirming the first hint obtained in set A.

Although three does not represent a probabilistic sample, for this first attempt to demonstrate our approach, we consider that this third set C (Figure 5) shows enough consistency in indicating the choice for the use of three clusters. And this is the number we will be using.

The complete data generated by the optimized clusters using two, three and four clusters for the sets A, B, and C, the resulting centers, membership grades and sizes are shown in the Appendix A.

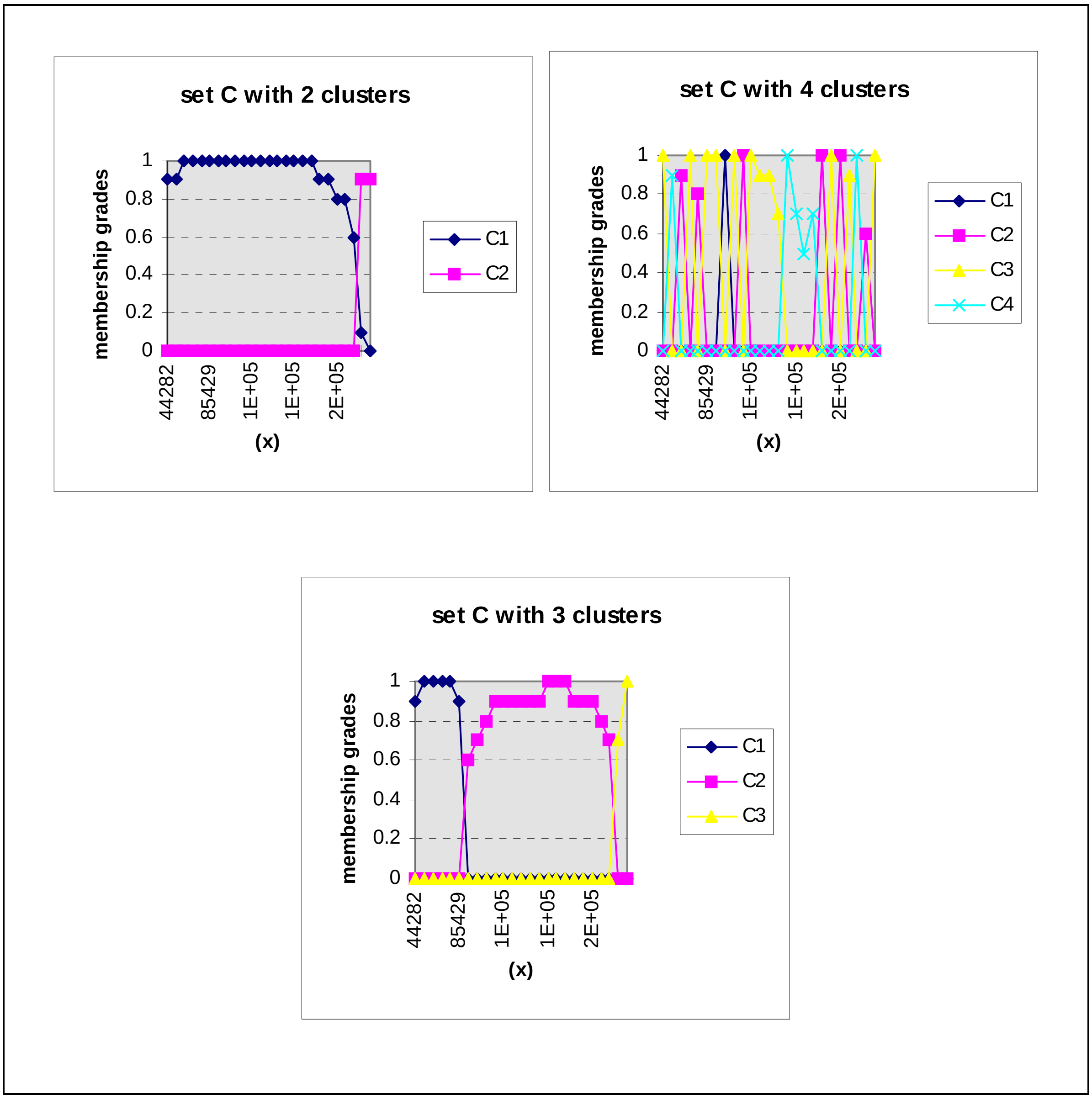

Figure 5. Resulting clusters for set C.

At this point we have gathered all necessary information to start the MTV process.

4.2 MTV

Calculating the MTV is based on two principles:

1. Population effect: Let X consists of two clusters C_1 and C_2 with populations k_1 and k_2 and centers v_1 and v_2 respectively. If $k_1 >> k_2$, then the MTV should be "much closer" to v_1 than to v_2.
2. Distance effect: consider C_i with a center v_i. Then the effect of C_i on the MTV should be a "strong" decreasing monotonic function of $|v_i - MTV|$.

Definition 1: Let X denote a clustered set $\{ C_i = (k_i, v_i)\}^c_{i=1}$, let $\{\gamma_i (u)\}^c_{i=1}$ be nonnegative monotonically decreasing functions defined over the interval $[0, \infty)$ and let λ denote a real number greater than one. A solution **s** in R^p to the implicit vector equation

$$s = \frac{v_1\gamma_1\left(\left|v_1 - s\right|\right)k_1^\lambda + v_2\gamma_2\left(\left|v_2 - s\right|\right)k_2^\lambda + \ldots\ldots + v_c\gamma_c\left(\left|v_c - s\right|\right)k_c^\lambda}{\gamma_1\left(\left|v_1 - s\right|\right)k_1^\lambda + \gamma_2\left(\left|v_2 - s\right|\right)k_2^\lambda + \ldots\ldots + \gamma_c\left(\left|v_c - s\right|\right)k_c^\lambda} \quad (2)$$

is called a most typical value of order λ with the associated weight functions $\{\gamma_i\}^c_{i=1}$, and is denoted by MTV($\gamma_1,\ldots,\gamma_c,\lambda$).

The population effect is guaranteed by the request $\lambda > 1$. The functions $\gamma_1(u),\ldots,\gamma_c(u)$ are strong decreasing monotonic functions in order to assure the distance effect. To each cluster we may attach a different weight function.

The next values are as follows:

$$\lambda = 2; \gamma_i(u) = e^{-\beta u}, 1 \le i \le c \quad (3)$$

where β is a tuning constant that determines the decreasing rate of the weight functions.

A solution **s** to Equation (2) always exists (Friedman, Ming & Kandel) and can be found by using the standard iteration method. However, this solution is not unique unless one of the clusters is dominant.

Using the information gathered beforehand, we calculate the MTV, using β = 3 and λ= 2. We have found a unique solution for all our example sets, what enables us to declare the MTV for the sets:

SETS	MTV
set A cluster 1	45700
set A cluster 2	45700
set A cluster 3	45700
set B cluster 1	75390
set B cluster 2	75390
set B cluster 3	75390
set C cluster 1	121680
set C cluster 2	121680
set C cluster 3	121680

Table 3. Resulting values for MTV for clusters of the sets.

4.3 MTD

A solution **s** to Equation (2) is accepted as a MTV if it is not “too far” from “too many” elements of X. To accept and evaluate this measure, another quantity is defined. The Most Typical Deviation measures the ‘grade of typicality’ of the MTV: a small MTD indicates a unique solution for Equation (2) and that this solution can be accepted as the MTV of the given set.

Definition 2: Let **s** be the solution of Equation (2). The scalar

$$t = \frac{|v_1 - s|^2 \gamma_1(|v_1 - s|) k_1^{\lambda} + |v_2 - s|^2 \gamma_2(|v_2 - s|) k_2^{\lambda} + \ldots\ldots + |v_c - s|^2 \gamma_c(|v_c - s|) k_c^{\lambda}}{\gamma_1(|v_1 - s|) k_1^{\lambda} + \gamma_2(|v_2 - s|) k_2^{\lambda} + \ldots\ldots + \gamma_c(|v_c - s|) k_c^{\lambda}} \quad (4)$$

is called the most typical deviation (MTD) of X, associated with **s**.

If a particular cluster, say C_k, is dominant in the process of determining the MTV, then **s** = MTV is close to v_k, and due to the choice of λ and $\gamma_1(u),\ldots, \gamma_c(u)$ one gets

$$MTD \sim |v_k - s| \quad (5)$$

In this case, the MTD is small and represents a "typical deviation" from the MTV, considering the population of C_k.

The resulting values for MTV and MTD for our example sets are:

	MTV	MTD
set A	45700	310.9
set B	75390	851.2
set C	121680	761.8

Table 4. Values of MTV and MTD for sets A, B and C.

4.4 DTV

The Definite Typical Value (DTV) is the unique solution resulted from Equation (2), after choosing a "appropriate" β. To verify the existence of a DTV, one has to apply the standard iteration method with the initial conditions:

$\mathbf{s}_0 = v_i$, $1 \le i \le c$ and see if it converges to the same solution.

Hence, we have sets A, B and C with its respective values for DTV.

set	DTV
A	45700
B	75390
C	121680

Table 5. DTV for sets A, B and C.

The existence of a DTV, as one can infer from *Table 1* , indicates the values for MTV are a representative measure for the sets. This condition being satisfied, we can use the MTV values ahead.

5. Interpreting the Results

The case-based reasoner parallels the reasoning performed by experts when assessing cash flow accounts. As it happens to the experts, one uses an estimate to direct reasoning and decision-making always taking into account that an estimate is an approximate value that contains imprecision and uncertainty. We have as one of our goals maintaining such imprecision and uncertainty and conveying it to the cash flow trying not to lose any information that may be embedded in those estimates. Using the CBR paradigm we have developed a system able to suggest such an estimate. The regular outcome of a case-based reasoner is, however, one solution chosen by one of the techniques proposed in the literature. In the last section we have proposed to deal differently with the case-based reasoner outcome. From the set of retrieved cases, we calculate its MTV. According to Friedman, Ming & Kandel, (1995) the MTV of a set is an attempt to find a value that is 'very close' to the one that the MTV represents. Consequently, we were able to convey the imprecision and uncertainty contained in the set of retrieved solutions to the result given by the MTV. At this point the estimate is modeled by the values of the MTV, the MTD and their meanings. Suppose we have the resulting MTV of 30,000 in a given set. This means that the value that better represents this set is 'very close' to 30,000. This is clearly a statement that can be represented by a normal fuzzy set, with center at 30,000. The MTD can be interpreted as an analog to the standard deviation. Hence, we can think of creating a fuzzy set where the values of the MTD orient the shape of this set increasing the degree of membership of the numbers around the MTV, as it increases. Henceforth we call this set the "Most Typical Fuzzy Set" (MTFS).

6. The Most Typical Fuzzy Set

The MTFS is an attempt to convey the fuzziness embedded in the result of the calculation of the MTV to the set of retrieved cases of the case-based reasoner.

As it has been stated before, the available parameters for constructing this type of sets are the MTV, the MTD and their meanings. MTV means that there is a region around the MTV value that represents the set if it is not too far from too many elements of the given set. MTD means how far the MTV is from the centers of the clusters that were calculated to classify the elements of the set. Actually, the smaller the MTD, the better the MTV represents the given set.

However, we cannot simply try to employ the MTD as a deviation parameter and create the membership functions. The MTFS aims to represent the possible outcomes coming from the implementation of the TT. In our point of view, there are three types of outcomes and they can be modeled by MTFSs that represent the concept of being close to the MTV, very close to the MTV and fairly close (around) the MTV. For this reason we understand that we would better use the theory behind *hedges* to have more reliable sets. To create the MTFSs in compliance with the theory grounding hedges, we have to try to fit the meaning of the MTD into the same paradigm.

The MTD is defined (Friedman, Ming, & Kandel, 1995) as a measurement of the 'grade of typicality' of the MTV and predicates such as 'small' and 'large' are used to label it. Using our three examples A, B, and C (see Table 4) we do not have enough information to conclude what is a 'small' MTD. We have calculated the MTV and the MTD for different sets trying to establish a range of MTD. Within the samples used that provided a unique solution for the MTV, we had values of MTD ranging from 78 to over 4,000. Note that this range may not be appropriate for other sets as we are working on a specific application where we have settled values that look proper within the scope of this application, such as β and the number of clusters.

Resulting from the analysis of the samples we have tested, we view the MTD as a base variable, which we can label with the linguistic terms *small*, *regular* and *large* . With such implementation, we intend to model the meaning of the MTD according to the way its authors have described it (Friedman, Ming, & Kandel, 1995). However, we have adapted their interpretation to our application and we have chosen to use a *large* MTD only to the extent that it does not indicate several solutions to the MTV values. We feel it is plausible because, according to the authors, "*a large MTD will indicate possible existance .of several solutions to…*" (Friedman, Ming, & Kandel, 1995). Hence, if a large MTD indicates only the possibility of several solutions, it seems clear that it is possible to have a large MTD in a set with one single MTV.

Let us define the linguistic variable (MTD, T, X). This variable expresses the MTD, the base variable. Due to the goals we have already discussed, we define T with three basic linguistic terms: *small*, *regular* and *large*. The range of X is an estimated support that looks consistent to the samples we have tested, the upper bound is 5,000 while the lower is zero as there is no reasonable representation for an MTD outside the positive interval. The constraints expressed by the linguistic values were intuitively chosen with the comparison of the values of MTD and the sets that originated it. Figure 6 shows the linguistic variable.

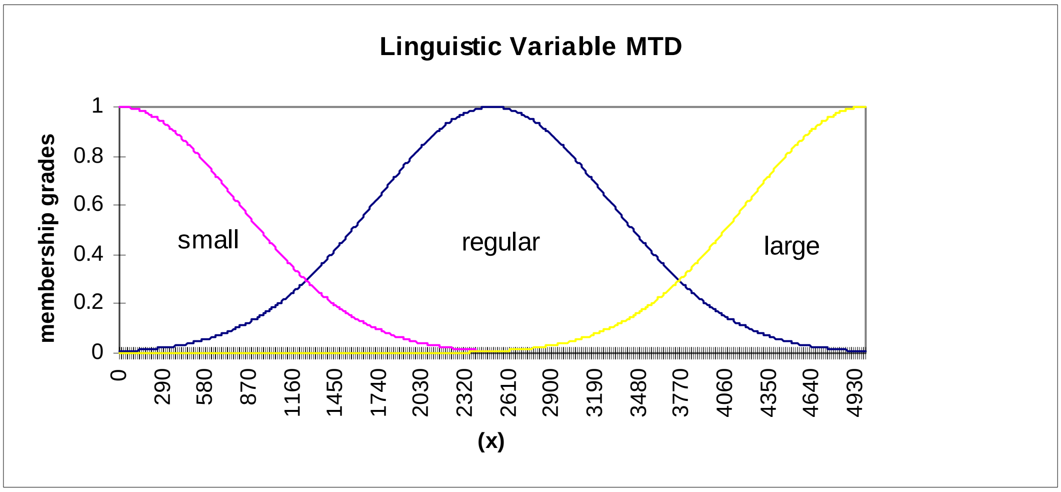

Figure 6 The linguistic variable MTD.

With the parameter MTD represented as a linguistic variable it can be expressed as *small*, *regular* or *large* MTD. These linguistic terms will be the actual parameters in the construction of the MTFSs.

At this point we are able to create the MTFS *close*, *very close* and *around* (fairly close) based on the principles used in hedges. The rules of creation of the MTFSs are:

If the MTD is *small* then MTFS = *very close to MTV.*

If the MTD is *regular* then MTFS = *close to MTV.*

If the MTD is *large* then MTFS = *around the MTV.*

The first rule assigns to the MTFS lower grades of membership to the values *around* the MTV, given that the MTD is *small* what indicates a better grade of typicality of the MTV. The third rule does the opposite, as a *large* MTD indicates a lower grade of typicality, increasing the grades of membership for the numbers around the MTV. The second rule is self explanatory.

Let us see how we construct the MTFSs. Getting the values from our examples, we have on set A, the MTV value of 45,700 and MTD equals 310,9, which indicates a *small* MTD (see Table 4). Then, the MTFS that should be built is the type that represents the concept ‘very close to the MTV’. Effectively, all our examples resulted in *small* MTDs.

The three basic types of MTFSs representing the concepts *very close*, *close* and *around* are illustrated in Figure 7, Figure 8 and Figure 9.

These three MTFS are plotted using 45,7 as the center. Let us use the element x = 40 and compare the grades of membership in each set to demonstrate the accordance with the hedge principles.

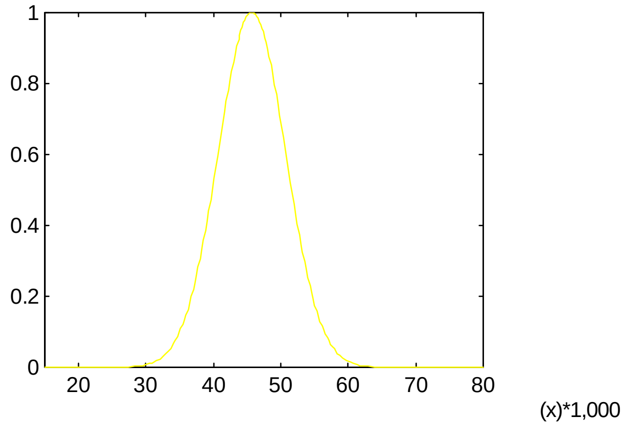

Figure 7. MTFS very close to MTV.

$$\mu_{\text{very close to MTV}}(x) = e^{\left(-(x-MTV)^2 / (.22\,MTV)^2\right)}$$

$$\mu_{\text{very close to MTV}}(45.7) = 1$$

$$\mu_{\text{very close to MTV}}(40) = .52$$

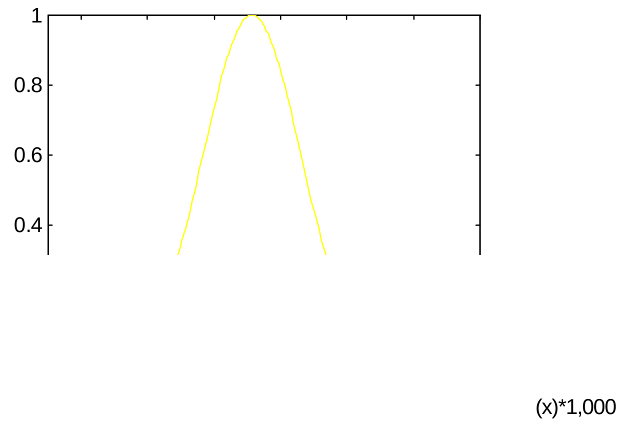

Figure 8. MTFS close to MTV.

$$\mu_{\text{close to MTV}}(x) = e^{\left(-(x-MTV)^2 / (.32\,MTV)^2\right)}$$

$$\mu_{\text{close to MTV}}(40) = .74$$

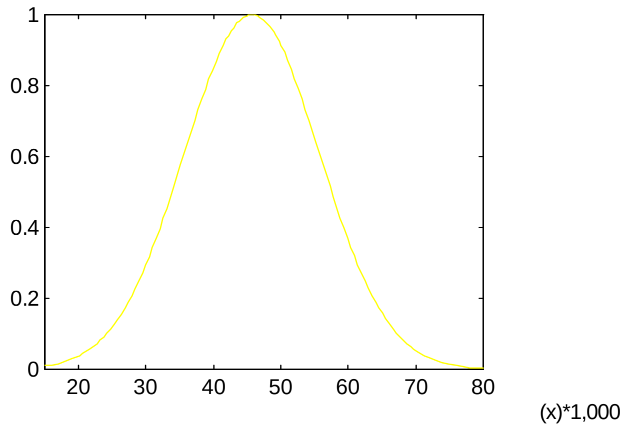

Figure 9. MTFS around the MTV.

$$\mu_{\text{around the MTV}}(x) = e^{\left(-(x-MTV)^2 / (.44\,MTV)^2\right)}$$

$$\mu_{\text{around the MTV}}(40) = .85$$

7. The Fuzzy Cash Flow

The fuzzy cash flow is a decision support device. The fuzzy cash flow consists essentially of data of three natures. The first sketch of the cash flow is built with actual liabilities and receivables. Although these values are usually actual and certain, some firms have some uncertainty in the receivables, but this is outside the scope of the present work. The second nature of data is the one that strongly orients the flow of money, that is a production or a sales target. At times these targets are established beforehand by the headquarters or they are bounded on production capacity or even on laws. Even these targets may also be incorporated to some uncertainties. Again, it is not our intention to discuss the modeling of such variables. The third nature of data is uncertain and imprecise and it is very usually estimated by experts. It consists of accounts that cause the flow of the money. For example, taxes, salaries, suppliers and some receipts stemming from sources other than the main revenue. These accounts must be estimated in order to avoid losses and maximize profits. This third nature of data is the scope of the present paper.

Along the sections above, we have demonstrated our proposal to model this fuzzy data. In last section, we have reached the construction of the MTFSs. There are different approaches to deal with fuzzy cash flow. Bellman & Zadeh (1970) propose the decision making in a fuzzy environment. In a cash flow, each estimate can be modeled by a MTFS and it becomes a fuzzy constraint within a fuzzy decision making environment. These fuzzy constraints are combined with other kinds of data in the design of the fuzzy cash flow. The decisions arise from goals and these goals can be modeled as fuzzy goals. The fuzzy cash flow provides fuzzy decisions and they are found at the intersection of fuzzy goals and constraints. An approach to calculate the Fuzzy Present Worth and the Fuzzy Rate of Return is presented in Ward (1989). Chiu & Park, (1994) present another alternative for calculating the present worth in a fuzzy cash flow.

8. Concluding Remarks

The essence in a designing a cash flow is the assessment of its accounts. We have described an approach to estimate and represent cash flow accounts. The uncertainty and impreciseness embedded in these accounts are maintained with the purpose of avoid loosing relevant information and, consequently, providing better results with a more accurate cash flow.

The decision making concerning the fuzzy cash flow can be dealt by different approaches as it was introduced on section 7. In a forthcoming work, the authors intend to propose the more appropriate approach considering the modeling of all types of fuzzy data that may be part of the process.

Chiu & Park, (1994) state that the membership function may be viewed as an opinion poll of experts thought. The membership functions in the present work are originated from the outcome of the reasoner that uses past experience. Being experience what generates the expert opinion, we can consider our membership functions as legitimate.

The MTDs calculated for the sets A, B, and C (outcomes of the CBReasoner) are all inside the range of *small* MTD. Intuitively, this indicates that our reasoner is able to provide an outcome that is consistent and coherent. Further experiments must be done, although we understand that this is a very promising approach.

The development of MTFSs seems to be another very promising issue. These sets may be created from different types of data that undergo the MTV method as the meanings of MTV and MTD do not vary. Some caution is required in determining the range of the MTD for each application.

9. References

Aamodt, A. & Plaza, E. (1994). Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. *Artificial Intelligence Communications*,7 (1), 39-59.

Bellman, R.E. & Zadeh, L. A. (1970) Decision-Making in a Fuzzy Environment. *Management Science, 17*, 4, 141-164.

Buckley, J.J. (1987). The Fuzzy Mathematics of Finance. *Fuzzy Sets and Systems, 21*, 257-273.

Buckley, J. J. (1992). Solving Fuzzy Equations in Economics and Finance. *Fuzzy Sets and Systems, 48*, 289-296.

Cheeseman, P. and Stutz, J. (1996). Bayesian Classification (AutoClass): Theory and Results, chapter 6, *Advances in Knowledge Discovery and Data Mining*, AAAI Press.

Chiu, Chiu-Yu & Park, Chan (1994). Fuzzy Cash Flow Analysis Using Present Worth Criterion. *The Engineering Economist, 39*, 2, 113-138.

Dubois, D. & Prade, H. (1983). Ranking Fuzzy Numbers in the Setting of Possibility Theory. *Information Sciences, 30*, 183-224.

Friedman, M, Ming, M., & Kandel, A. (1995). On the Theory of Typicality. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 3, 2, 127-142.

Jain, R. (1976). Decision Making in the Presence of Fuzzy Variables. *IEEE Transactions on Systems, Man and Cybernetics*, october, 698-702.

Kandel, A. (1982). *Fuzzy Techniques in Pattern Recognition*. John Wiley & Sons, Inc.

Klir, G. & Yuan, B. (1995). *Fuzzy Sets and Fuzzy Logic: theory and applications*. Prentice-Hall Inc., Upper Saddle River, NJ.

Kim, Steven H. & Novick, Mark B.(1993). Using clustering techniques to support case reasoning. *International Journal of Computer Applications in Technology*, 6, 2, 57-73.

Kolodner, J. (1989). Judging Which is the Best Case for a Case-Based Reasoner. *Proceedings of a Workshop on Case-Based Reasoning*, 77-81.

Pohlman, R., Santiago, E. S., & Markel, F. L. (1988). Cash Flow Estimations Practices of Large Firms. *Financial Management*, 17, 2, 71-79.

Rousseeuw, P. J. (1995). Discussion: Fuzzy Clustering at the Intersection. *Technometrics*, 37, 3, 283-286.

Stottler, R. H.(1994). CBR for Cost and Sales Prediction. *AI Expert, (Aug.)*, 25-33.

Ward, T. L. (1989). Fuzzy Discounted Cash Flow Analysis. Applications of Fuzzy Set Methodologies in Industrial Engineering, G. W. Evans, W. Karwowski, and M. R. Wilhem, 92-102.

Weber-Lee, R., Barcia, R. & Khator, S. (1995). Case-based reasoning for cash flow forecasting using fuzzy retrieval. *Case-Based Reasoning Research And Development: First International Conference*; proceedings/ICCBR-95, Sesimbra, Portugal, October 23-26. Manuela Veloso & Agnar Aamodt (ed.) Springer-Verlag, 510-519.

Weber-Lee, R., Martins, A., Pacheco, R. & Barcia, R. (1996). Using Typicality Theory to Select the Best Match. Submitted to 3rd European Workshop on Case-Based Reasoning (EWCBR'96), Lausanne, November 14-16.

Windham, Michael P. (1983). Geometrical Fuzzy Clustering Algorithms. *Fuzzy Sets and Systems*, 10, 271-279.

APPENDIX A

Optimized clusters from sets A, B, and C with 2,3 and 4 clusters.

SET	A	# of	clusters	2	
	samples	C1	C2		
1	32102	0	0.9		
2	38138	0	1		
3	38566	0	1		
4	39333	0	1		
5	40471	0	1		
6	43121	0	1		
7	43426	0	1		
8	44359	0	1		
9	44747	0	1		
10	45668	0	1		
11	45734	0	1		
12	45735	0	1		
13	47250	0	1		
14	48007	0	1		
15	48552	0	1		
16	48733	0	1		

17	49471	0	1		
18	51115	0	1		
19	51898	0	1		
20	52390	0	1		
21	54362	0	0.9		
22	55037	0	0.9		
23	55481	0	0.9		
24	80211	1	0		
25	87454	1	0		
centers	-*-	83230	46150		
sizes	-*-	2	23		
SET	A	# of	clusters	3	
	samples	C1	C2	C3	
1	32102	0.8	0	0	
2	38138	0.9	0	0	
3	38566	1	0	0	
4	39333	1	0	0	
5	40471	1	0	0	
6	43121	0.9	0	0	
7	43426	0.9	0	0	
8	44359	0.8	0	0	
9	44747	0.7	0	0	
10	45668	0.6	0	0	
11	45734	0.5	0	0	
12	45735	0.5	0	0	
13	47250	0	0.7	0	
14	48007	0	0.9	0	
15	48552	0	0.9	0	
16	48733	0	0.9	0	
17	49471	0	1	0	
18	51115	0	1	0	
19	51898	0	1	0	
20	52390	0	1	0	
21	54362	0	0.9	0	
22	55037	0	0.9	0	
23	55481	0	0.9	0	
24	80211	0	0	1	
25	87454	0	0	1	
centers	-*-	41110	50770	83820	
sizes	-*-	12	11	2	
SET	A	# of	clusters	4	
	samples	C1	C2	C3	C4
1	32102	0.8	0	0	0
2	38138	1	0	0	0
3	38566	1	0	0	0
4	39333	1	0	0	0
5	40471	0.8	0	0	0
6	43121	0	0.7	0	0
7	43426	0	0.8	0	0
8	44359	0	0.9	0	0

9	44747	0	0.9	0	0
10	45668	0	1	0	0
11	45734	0	1	0	0
12	45735	0	1	0	0
13	47250	0	0.9	0	0
14	48007	0	0.8	0	0
15	48552	0	0.7	0	0
16	48733	0	0.7	0	0
17	49471	0	0.5	0	0
18	51115	0	0	0	0.8
19	51898	0	0	0	0.9
20	52390	0	0	0	1
21	54362	0	0	0	1
22	55037	0	0	0	0.9
23	55481	0	0	0	0.9
24	80211	0	0	1	0
25	87454	0	0	1	0
centers	-*-	38150	45940	83870	53150
sizes	-*-	5	12	2	6
SET	B	# of	clusters	2	
	samples	C1	C2		
1	23551	0	0.9		
2	31875	0	0.9		
3	53027	0	1		
4	53105	0	1		
5	55036	0	1		
6	56774	0	1		
7	60366	0	1		
8	64322	0	1		
9	66149	0	1		
10	66381	0	1		
11	68007	0	1		
12	68178	0	1		
13	70627	0	1		
14	72555	0	1		
15	75203	0	1		
16	76557	0	1		
17	77460	0	1		
18	77973	0	1		
19	80030	0	1		
20	85766	0	1		
21	87444	0	1		
22	87554	0	1		
23	97839	0	0.9		
24	140305	0.8	0		
25	206627	1	0		
centers	-*-	177550	67740		
sizes	-*-	2	23		
SET	B	# of	clusters	3	
	samples	C1	C2	C3	

1	23551	0	0.8	0	
2	31875	0	0.8	0	
3	53027	0	1	0	
4	53105	0	1	0	
5	55036	0	1	0	
6	56774	0	1	0	
7	60366	0	0.8	0	
8	64322	0	0.6	0	
9	66149	0	0	0.5	
10	66381	0	0	0.5	
11	68007	0	0	0.7	
12	68178	0	0	0.7	
13	70627	0	0	0.8	
14	72555	0	0	0.9	
15	75203	0	0	1	
16	76557	0	0	1	
17	77460	0	0	1	
18	77973	0	0	1	
19	80030	0	0	1	
20	85766	0	0	1	
21	87444	0	0	0.9	
22	87554	0	0	0.9	
23	97839	0	0	0.8	
24	140305	0	0	0	
25	206627	1	0	0	
centers	-*-	194300	52530	79030	
sizes	-*-	1	8	15	
SET	B	# of	clusters	4	
	samples	C1	C2	C3	C4
1	23551	0	0	1	0
2	31875	0	0	1	0
3	53027	0	0	0	0.7
4	53105	0	0	0	0.7
5	55036	0	0	0	0.8
6	56774	0	0	0	0.9
7	60366	0	0	0	1
8	64322	0	0	0	1
9	66149	0	0	0	1
10	66381	0	0	0	1
11	68007	0	0	0	1
12	68178	0	0	0	0.9
13	70627	0	0	0	0.8
14	72555	0	0	0	0.7
15	75203	0.5	0	0	0
16	76557	0.6	0	0	0
17	77460	0.7	0	0	0
18	77973	0.7	0	0	0
19	80030	0.9	0	0	0
20	85766	1	0	0	0
21	87444	1	0	0	0
22	87554	1	0	0	0

23	97839	0.9	0	0	0
24	140305	0.4	0	0	0
25	206627	0	1	0	0
centers	-*-	85760	200440	29370	64480
sizes	-*-	10	1	2	12
SET	C	# of	clusters	2	
	samples	C1	C2		
1	44282	0.9	0		
2	57072	0.9	0		
3	66725	1	0		
4	67751	1	0		
5	79266	1	0		
6	85429	1	0		
7	99256	1	0		
8	102093	1	0		
9	105032	1	0		
10	108125	1	0		
11	108144	1	0		
12	109003	1	0		
13	110396	1	0		
14	110574	1	0		
15	113163	1	0		
16	113563	1	0		
17	118193	1	0		
18	128909	1	0		
19	139318	0.9	0		
20	140481	0.9	0		
21	150964	0.8	0		
22	156232	0.8	0		
23	177474	0.6	0		
24	228258	0.1	0.9		
25	322234	0	0.9		
centers	-*-	105290	259140		
sizes	-*-	23	2		
SET	C	# of	clusters	3	
	samples	C1	C2	C3	
1	44282	0.9	0	0	
2	57072	1	0	0	
3	66725	1	0	0	
4	67751	1	0	0	
5	79266	1	0	0	
6	85429	0.9	0	0	
7	99256	0	0.6	0	
8	102093	0	0.7	0	
9	105032	0	0.8	0	
10	108125	0	0.9	0	
11	108144	0	0.9	0	
12	109003	0	0.9	0	
13	110396	0	0.9	0	
14	110574	0	0.9	0	

15	113163	0	0.9	0	
16	113563	0	1	0	
17	118193	0	1	0	
18	128909	0	1	0	
19	139318	0	0.9	0	
20	140481	0	0.9	0	
21	150964	0	0.9	0	
22	156232	0	0.8	0	
23	177474	0	0.7	0	
24	228258	0	0	0.7	
25	322234	0	0	1	
centers	-*-	70110	123090	287400	
sizes	-*-	6	17	2	
SET	C	# of	clusters	4	
	samples	C1	C2	C3	C4
1	44282	0	0	1	0
2	57072	0	0	0	0.9
3	66725	0	0.9	0	0
4	67751	0	0	1	0
5	79266	0	0.8	0	0
6	85429	0	0	1	0
7	99256	0	0	1	0
8	102093	1	0	0	0
9	105032	0	0	1	0
10	108125	0	1	0	0
11	108144	0	0	1	0
12	109003	0	0	0.9	0
13	110396	0	0	0.9	0
14	110574	0	0	0.7	0
15	113163	0	0	0	1
16	113563	0	0	0	0.7
17	118193	0	0	0	0.5
18	128909	0	0	0	0.7
19	139318	0	1	0	0
20	140481	0	0	1	0
21	150964	0	1	0	0
22	156232	0	0	0.9	0
23	177474	0	0	0	1
24	228258	0	0.6	0	0
25	322234	0	0	1	0
centers	-*-	314280	64540	110050	158580
sizes	-*-	1	6	12	6