Using Typicality Theory to Select the Best Match

Rosina Weber-Lee, Ricardo Miranda Barcia, Alejandro Martins, and Roberto C. Pacheco
rolee@eps.ufsc.br
Federal University of Santa Catarina, Industrial Engineering,
Florianópolis, SC, BRAZIL.

Abstract. This paper focuses on the problem of choosing the best match among a set of retrieved cases. The *Select* step is subtask of case retrieval that produces the case that suggests the solution for the input case. There are many different ways to accomplish this task and we propose an automatic means for it. Following the original motivation of paralleling the human similarity heuristic we argue that the selection of the best match is performed by humans choosing the solution that best represents the set of candidate solutions retrieved. The solution that best represent a given data set is the "most typical" solution. Therefore, we describe an application in a Case-Based Reasoning system using the Theory of Typicality to calculate the Most Typical Value of a given set to automatically perform the *Select* task. An example illustrates the application.

1. Introduction

In a Case-Based Reasoner, the input case is compared with the cases in the memory in order to retrieve the set of most similar cases. The case retrieval step is usually referred to as consisting of four subtasks: Identify Features, Initially Match, Search, and *Select* (Aamodt & Plaza, 1994). Here, we address the subtask *Select* that chooses a best match from a set of most similar cases to suggest the solution to the input problem. Several approaches for selecting one best match have been proposed, from developing a separate module to leaving the selection to the user (see section 2). However, one considers that some important piece of information and knowledge may be lost when only one case is used. To overcome such possible loss it might be necessary to select more than one match to produce a good and useful solution (Shinn, 1988). That is the situation we have faced when trying to produce the solution in our case-based reasoner developed to predict cash flow accounts (Weber-Lee, Barcia, & Khator, 1995). Hence, we propose to perform the *Select* subtask by means of the Theory of Typicality. It is an attempt to capture all the knowledge embedded in the set of the most similar cases avoiding loosing important information. Building this last subtask of the case retrieval step places us closer to achieving the goal of developing an automatic tool. Whenever knowledge can be manipulated without depending upon humans, we are evolving towards a better Artificial Intelligence tool. Section 2 discusses the efficient automatization of the *Select* phase in a CBR system.

The motivation for this work has its origin in observing that the process starting from identifying a current problem and followed by the search for similar experiences does not stop there. The humans very often retrieve more than one similar case and they go through a selection phase before choosing the solution for the current problem. Section 3

Weber-Lee, R.; Pacheco, R.; Martins, A.; Barcia, R. (1996). Using Typicality Theory to Select the Best Match. Ian Smith; Boi Faltings (eds.) Advances in Case-Based Reasoning (LNAI 1168), pp. 445-459. Berlin: Springer.

illustrates what we claim to be the human process with examples that show that the human approach actually goes up to the point of selecting the best match.

For the purpose of modeling the *Select* phase, we propose the use of Theory of Typicality (Friedman, M, Ming, M., & Kandel, A., 1995), we comment on the motivation for this choice in section 4. Next, in Section 5 we review this theory that provides a measure of central tendency -- the Most Typical Value of a data set. Then, in Section 6 we demonstrate an example where the Theory of Typicality is applied to select the best match of a set of retrieved cases. We comment on this proposal and indicate future research in Section 7.

2. The Best Match

A CBR system has five major topics to be managed, studied, and developed: case representation, case retrieval, case adaptation, revision and learning. All these five topics consist of knowledge engineering tasks that have to be carefully considered within the domain of the application. The system works better with better design of these issues. This is about maximizing the potential of CBR as a tool.

The goal of a CBR system is effectively solving the input problem. The knowledge engineer attempts to *optimize* the potential of all processes that embody a case-based reasoner. We strive to maximize the efficiency of case-based systems towards continuous improvement. Above all, we seek for ways to improve the CBR as an automatic technique. The automatic selection we are introducing chooses the best match using more cases without increasing computing time. Using more cases increase the chances that no important information is rejected.

Retrieval is the essence of the CBR paradigm. Case retrieval consists of *Identify Features, Initial Match*, *Search*, and *Select.* All these subtasks are equally important in order to perform an efficient case retrieval. We address the subtask *Select* that chooses the solution for the input case. *Select* step may combine the cases returned by *Initial Match* and produce the solution or simply choose the best match. Besides, when case retrieval is oriented by scores, one needs to set a threshold to limit the amount of cases to be returned. This threshold determines the domain of the *Select* subtask.

According to (Aamodt, & 1994) the selection of the best match may happen during the Initial Match step but usually the Initial Match returns a set of cases. The selection between this set may depend upon a threshold that determines the amount of cases to retrieve. However, there are other types of retrieval that searches for matching indexes and the outcome of this retrieval can be of any size. The retrieval may yield either a score meaning the degree of match or assign a yes or no to determine whether the case matches sufficiently or not (Kolodner, 1993).

When retrieving cases one searches for the most similar ones. Among these, the case that should be used to solve the input problem must be chosen. There are several ways to determine which one better represents the set of best cases retrieved. In (Kopeikina et al., 1988), it is presented as a separated module, called *Selector,* to perform the choice of the best match. Candidate cases are retrieved and *Selector* determines the most relevant of them identifying more subtle distinctions between them. According to Aamodt and Plaza (1994), *Select* is usually more elaborate than the retrieval itself. They even point out a system where the selection methods generate explanations and the case with the best explanation for the similarity is then selected. The selection may be performed by a second evaluation of the same retrieval function that then examines closer the first set of retrieved cases. Some systems ask the user either for the choice or for some additional information to perform the choice. The use of heuristics is also employed, (Kolodner, 1989). In Clavier (Hennessy and Hinkle,1992) the system presents candidate layouts to the user to choose the best match. In (Hurley, 1993) the retrieval is divided into two stages. First, *the base filtering* determines a set of broadly similar cases. Second, the *detailed matching* performs the matching trying to verify whether or not the target case satisfies the constraints of the indices of the case base. Another interesting approach is presented in Macchion & Vo, (1993) in which the retrieval phase is split into up to six steps. In searching for the perfect match the system uses *exclusion* and *necessity* indexes to reduce the set.

Most systems tend to perform selection based on the highest score or on the user. Others choose for *ad hoc* procedures while some do not use any method at all. Next, we present a list of methods for implementing the *Select* subtask followed by the research systems that has adopted it, (Kolodner, 1993). High score - ABBY, ACBARR, ARCHIE and ARCHIE-2, CABINS, The Compaq SMART System, SCAVENGER; user - ASK systems, Battle Planner, CLAVIER, CASCADE, REMIND; sort cases in a claim lattice - CABARET, HYPO; no method needed - Parse-O-Matic, PRISM; preference method - CELIA, MEDIC; domain specific metrics - KRITIK, NETTRAC; distance in similarity hierarchy - PERSUADER; plausibility of suggested explanation - SWALE; cases must share surface features and abstract problem types - ORCA; cases that deviates the least along goals and preconditions - Internal Analogy; multiple cases used - IVY; exclusion, then weighted similarity metric - MEDIATOR.

No matter the type of retrieval or case representation used, there is a dilemma regarding efficiency, accuracy, and processing time; that is a cost-effective question. In the system presented in (Kopeikina et al., 1988), they point out that a simple *selector* would choose quickly and imprecisely while a best case would be chosen with the use of more resources. Therefore, a more sophisticated approach should be used to avoid the selection of a sub-optimal result and to avoid wasting processing time with more iterations of less complex selection.

In our prediction application we have performed tests with different levels of threshold that resulted in different amounts of cases retrieved. Searching for an appropriate threshold turned out to be a difficult task. We have performed tests setting the threshold to higher levels (decreasing the amount of cases retrieved) trying to ensure that very similar cases were not rejected. This is an important cost-effectiveness matter has to be considered. One may seek for efficiency reducing the cases retrieved and increase the difficulty in choosing the best match or, on the other hand, seek for accuracy and loose efficiency. The solution seems to be keeping a large amount of cases and improve the *Select* task, therefore one may consider all relevant cases without sacrificing efficiency. In order to improve the *Select* task we propose an approach that was motivated by the human behavior as well as the whole philosophy behind CBR: paralleling humans. The choice of the solution to reuse is very often the most frequent and most typical solution in the set, that is the one that better represents the set.

3. CBR Paralleling Humans

Case-Based Reasoning paradigm has been introduced as the artificial intelligence technique capable to simulate the similarity heuristic (Whitaker et al., 1990) in a computer program. This behaviour refers to the use of the outcome of a prior similar experience to predict the outcome of a current situation. Such simulation is focused on the use of features to ground the comparison of a current problem in the search for similar ones. The search is for a successful past experience that may provide a proper solution to be reused. In fact, many applications present very good simulations of such human behavior. Most introductory papers, as well as application papers present the CBR philosophy and mechanism with examples that start from the identification of a current problem and the search in memory for a similar situation. These examples illustrate how the past experiences can help to provide either a solution, a diagnosis or a prediction for the current problem. Selecting *the* past experience is not highlighted in examples and it may seem as if this final selection is not part of the human approach. However, as the philosophy is originated in simulating the human approach, what actually happens as a result of the search for similar problems? It is not necessarily one similar case that is retrieved. Usually, a number of cases are retrieved and one has to choose one in the set. As we can see through some real life examples, the approach tends to be choosing the most frequent or most typical case. Let us describe some examples to illustrate why we understand that the selection of the best match can also be paralleled like the identification and retrieval of similar cases.

When a crime is committed, police search for all possible characteristics of the crime in order to find enough similarity to other crimes committed in the past. The investigation procedure will point to the criminal whose typical behavior is the most similar to the one under research.

Finding a parking space or the best route to drive to someplace is the type of problem faced by everyone on a daily basis. Humans approach such problem always retrieving past experiences. Sometimes there are more than one choice that has already been successful though the solution chosen is always the route (or the combination of routes) that is typically clearer; or the parking lot which most usually has vacancy.

All these real world situations show that the human approach indeed goes up to the point of selecting the best match and they also provide a strong appeal in considering the choice of the past experience to reuse as one experience that seems to be the most usual, frequent, or typical. Hence, the choice for the best match is, intuitively, a matter of evaluating a set of cases and finding which one(s) better represent this set.

These examples also indicate that the solution to be used in the current problem does not have to be necessarily one solution (adapted or not) from one best match. Conversely, they indicate that the best solution may be one obtained by the typical behavior of the set of solutions presented by the most similar cases. In this sense we now search for the best way to automatically calculate this solution.

4. Searching for a Representative Value

A value that is representative of the set of retrieved cases is the solution we want to identify automatically. On this account, we propose an approach that uses the real content of the best matches retrieved to produce a reasonable solution. At this point we have to consider the approaches available to calculate this value and we will have a CBR system that automatically parallels the whole human reasoning when facing a prediction problem. We have to choose an approach that will enable us to implement this automatic selection for the prediction of cash flow accounts. The solution part of the cases in this application is a single numerical attribute. Therefore we have to calculate one value that better represents this set of numbers. The nature of the data is not random, since it is strongly influenced by the economical conditions and top governmental decisions. Under these circumstances, our analysis leads us to prefer a fuzzy approach. There are measures of central tendency that represent attempts to choose one particular event that represents the behavior of a set (numerical averaging). That is not exactly what we are searching for, we want to calculate the most typical behavior of the set of solutions of the retrieved cases, if there is one. 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 leads to the development of the Fuzzy Expected Value (FEV), (Kandel, 1982). FEV was meant to present a quantity that would not only replace the arithmetic mean and the median but also be accepted as the typical grade of membership of a given fuzzy set. However, the FEV is not always effective on representing typicality. Two other measures were developed, the Weighted Fuzzy Expected Value (WFEV), (Friedman, Schneider, & Kandel, 1989) and the Clustering Fuzzy Expected Value (CFEV), (Vassilliadis et al., 1994). As a conclusion of these developments, the Most Typical Value (MTV), the Most Typical Deviation (MTD), and the Definite Typical Value (DTV) are presented in the paper *On the Theory of Typicality*, by 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 Theory of Typicality.

The Theory of Typicality calculates a MTV for the data set after grouping the data into clusters using a geometrical fuzzy clustering algorithm (see section 5.1). The choice for this type of clustering is embedded in the approach, nevertheless it has augmented our decision, 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.

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 behaviour of the attributes, and the need of the definition of the parameters of the probabilistic functions.

4.1. Using Typicality to Choose the Best Match

The Implementation of the chosen approach in our prediction system follows. A new case is input and the case retrieval returns a set of cases. In the selection of the best match, similarity is no longer considered; conversely we look at the solutions that the retrieved cases offer. These solutions compose a set of data. From this set we employ a fuzzy c-means algorithm (see section 5.1) and next calculate the MTV. If a unique solution exists, we then calculate the DTV. If exists one DTV for this set, this value corresponds to the best match, i.e., the value that becomes the solution of the reasoner. If there is not a unique solution to the MTV, we assume the cluster centers as possible solutions. At this point this second *Select* step is user-driven.

5. 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 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. The MTD measures the grade of typicality of the MTV calculated.

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). We present the algorithm briefly.

5.1. Geometrical Fuzzy Clustering Algorithms

The clustering problem is in essence a task of finding natural groupings in a given data set (Bezdek, 1981). In a fuzzy clustering, the requirements for clusters, classes and blocks are weaker 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 which 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 fuzzy clustering is a non-exclusive method, 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 $X=\{x_1,\dots x_N\} \subset R^p$ 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. $|x_k - x_1|^2_M = (x_k - x_l)^T M(x_k - v_l)$, 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 (Windham, 1983)

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

5.1.1. 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.

5.1.2. Cluster Validity and Unsupervised 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 cluster 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.

5.2. MTV

The definition of the MTV is motivated by the following 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 $|v_i - \text{MTV}|$.

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 1. 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 $\text{MTV}(\gamma_1,\ldots, \gamma_c, \lambda)$.

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:

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

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

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

5.3. MTD

A solution **s** to Equation (2) is accepted as an 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 that Equation (2) is likely to have a unique solution and that this solution can be accepted as an MTV of the given set. While a large MTD indicates the existence of several typical values.

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

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

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

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

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

5.4. DTV

The Definite Typical Value (DTV) is the unique solution resulted from Equation (2), after choosing a reasonable 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.

6. Example

We illustrate the approach proposed in our case-based reasoner developed to predict cash flow accounts (Weber-Lee, Barcia, & Khator, 1995). This system performs the tasks of identification and prediction. The basic case is represented by a list of attributes that identifies the cases that are cash flow accounts. The cash flow accounts are represented through descriptors related to present and previous amounts, nature and period in time. The system receives as the input problem one cash flow account with the attributes properly assigned. The system searches throughout the memory to identify the most similar cases using a similarity metric that assigns scores representing the degree of matching between the input case and the case being evaluated. There are different levels of importance for the descriptors. The solutions are the amounts that occurred for these accounts in the past after an equal period of time as the period described in the input case. The solution is adapted by a Parameter Adjustment (Kolodner,1993) operation to fit to the input case. The solution adapted is the amount forecasted 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. In the proposed approach, we evaluate this set of solutions to find out what values represent this set (the typical values) calculating the MTV of the set of cases retrieved.

Suppose we need to estimate an amount that is likely to be spent during next month in administrative expenses. The account named *administrative expenses* is the input case. We want the case-based reasoner to provide this estimate. A sketch of the input case is shown in Table 1.

The reasoner retrieved 26 cases as the set of most similar cases with a threshold defining that only cases yielding a similarity over 33% are retrieved. The search was done in a case base of 624 cases. Table 2 presents the 26 cases, the scores representing the degree of similarity, the period, and the amount estimated for the period. The last column named *estimation* shows the amount provided by each case if it would be chosen. The parameter we are using for measuring the quality of the solutions is the actual amount that was spent, for this input case of *administrative expenses* of June, 1987, it was $ 96,745.

Considering the choices for selection of the best match already proposed in the CBR literature, we could, at this point, ask the user for a selection, apply the similarity metrics with more detail, exclude some of the cases or even choose the highest score. Also, we could combine these cases in order to produce one single solution. However, we have chosen to group these cases into a set and find out what is the typical value for this set of solutions.

dimensions	values
AccountName	adm. expenses
Amount	119486.99
Month	jun
Year	87
M_1_Density	0.850396
M_2_Density	1.108768
M_3_Density	1.512038
M_4_Density	1.659253
M_5_Density	1.398817
M_6_Density	1.344115
M_7_Density	1.333065
M_8_Density	2.046216

M_9_Density	1.893387
M_10_Density	1.419327
M_11_Density	1.430320
FuzzyLabel	Low

Table 1. Problem description.

Before developing the MTV approach, let us think of applying one of the alternative methods above mentioned. Let us select according to the highest score and asking the user.

The selection according to the highest score is *adm.exp. aug82*; the user asked opted for *adm.exp. may87*. Table 4 presents the input case and these two selections for comparison. An explanation for the descriptors that are not self-explanatory follows: Next_Amount, this value refers to the actual amount occurred the following month; M_X_Density, these are seasonality factors, an attempt to represent the behavior of the account for the past year.

It is still important to point out that although the highest score has indeed chosen the best possible forecast for the input problem, one cannot rely on one single example to ground validation. This example has only explanatory purposes.

According to the theory of typicality, in the search for a MTV the data must be first clustered by a geometric fuzzy clustering algorithm and replaced by a finite set of ordered pairs, where each pair consists of a cluster's center and the cluster's size. We want to determine whether there is a DTV to represent the set. The data set is

$X=\{x_1,\ldots,x_{26}\} \subset \Re^1$. We use the fuzzy c-means algorithm to determine the clusters' center and size. The fuzzy c-means algorithm requires that a number of clusters c is given and, also, a particular distance, a real number $m \in (1,\infty)$, and a small positive number ε for a stopping criterion. Let us use $c = 4$, and $m = 1.1$, and $\varepsilon = 0.001$. The choice for ε and m was made after the suggested values in the references Klir & Yuan, (1995), and the value for m worked fine in our tests although Pal & Bezdek (1995) suggest the interval of m to be [1.5,2.5]. The choice for $c = 4$ was based on the visual observation of the data. The resultant partition fit the validity guidelines (see section 5.1.2).

	nature	scores	period	estimative
1	adm exp	55.697	aug82	92444
2	adm exp	55.0982	nov 82	140474
3	adm exp	51.7517	nov81	141733
4	adm exp	50.505	jun83	184256
5	adm exp	50.2867	mar81	172925
6	adm exp	50.0313	jan85	158821
7	adm exp	49.487	mar81	121272
8	adm exp	49.31	dec 84	127859
9	adm exp	48.8606	abr81	334044
10	adm exp	48.0239	feb 81	95496
11	adm exp	47.4461	oct82	200932
12	oth recp	46.6733	aug83	134633

13	adm exp	45.4544	jan81	116865
14	adm exp	45.1228	jan 82	168655
15	adm exp	45.0171	may87	158623
16	oth dis	45.0043	jan 87	142776
17	adm exp	44.7688	feb 82	109315
18	adm exp	44.4445	may87	105113
19	adm exp	44.161	aug83	111196
20	adm exp	44.0656	mar 82	140507
21	adm exp	43.28	sep 82	136687
22	adm exp	41.7466	may81	123876
23	oth recp	39.3639	feb 86	131863
24	adm exp	37.229	feb 85	154815
25	suppl	34.7698	dec 82	143533
26	oth disb	33.9213	nov85	120708

Table 2. Cases retrieved.

The algorithm starts with an initial fuzzy pseudopartition $U^{(0)}$, which is a matrix of u_{ik}, the membership grades of the kth element in the ith cluster. Our initial fuzzy pseudopartition was chosen randomly. The outcome of the algorithm is the optimized fuzzy partition $U^{(t)}$ with the finite set of ordered pairs: the cluster's centers v={v1,v2,v3,v4}and sizes k={k1,k2,k3,k4}:
v={v1,v2,v3,v4}= {176.950; 139.410 ; 108.820; 333.930}, and
k={k1,k2,k3,k4}= {1, 4, 8, 10}.

Using (3), we calculate MTV by Equation (2) resulting MTV = 132.570, as the unique solution. This means that there is one most typical value of order 2 to represent the given set. To measure the grade of typicality, using the result of (2), we calculate the MTD using (4). The result is MTD = 2.558, which represents a small number confirming that Equation (2) has one unique solution. At this time we verify that a DTV exists for the given data set. The resultant MTV when each center is used as the initial "s" in Equation (2) is presented in the following table:

Center	MTV
176,950	132,570
139,410	132,570
108,820	132,570
333,930	132,570

Table 3. The centers and the MTV.

Table 3 above demonstrates that there exists one unique MTV to the referred problem and this solution is called the *definite typical value* (DTV), (Friedman, Ming, & Kandel, 1995).

dimensions	input	highest score	user
AccountName	adm. expenses	adm. expenses	adm. expenses
Amount	119486.99	51277.49	101611.26

Month	jun	aug	may
Year	87	82	87
Next_Amount	96745.42	40981.98	119486.99
M_1_Density	0.850396	0.865198	1.303826
M_2_Density	1.108768	1.094545	1.778040
M_3_Density	1.512038	0.971196	1.951153
M_4_Density	1.659253	1.014571	1.644901
M_5_Density	1.398817	0.764248	1.580575
M_6_Density	1.344115	0.939003	1.567582
M_7_Density	1.333065	0.798711	2.406192
M_8_Density	2.046216	0.590103	2.226477
M_9_Density	1.893387	0.670797	1.669020
M_10_Density	1.419327	1.222284	1.681946
M_11_Density	1.430320	1.120506	1.876473
estimative value after adaptation		95496	140507

Table 4. Comparison of two matches.

From our previous assumptions, the MTV represents (is the most typical of) the set of solutions from the retrieved cases. This is the outcome the reasoner produces. The suggested forecast for the account *administrative expenses,* for July 87 is: $ 132,570. This is the result of the run of the automatic *Select* subtask, that combines the solutions of all retrieved cases to solve the input problem. For this example the estimate resulted is worse than the one chosen based on the highest score and a bit better than the one suggested by the user. However this example has only illustrative purposes.

7. Concluding Remarks and Future Work

In this paper, we proposed the use of the Theory of Typicality to perform automatically the *Select* subtask of case retrieval in a CBR system to predict cash flow accounts. The approach demonstrated in section Example has produced a solution using all the retrieved cases preventing loss of relevant information.

An important issue to be discussed is about systems that may result in a set of retrieved cases with multivalue MTV. In the example presented, a DTV exists, and consequently it becomes the outcome of the system. On the other hand, if there is not one unique solution to the MTV, the *Select* task is not ready. In some domains, the existence of multivalue MTV may indicate that a CBR system is not able to generate a solution. Striving to achieve the goal of implementing a completely automatic reasoner, the authors are working in an approach to deal with this alternative.

The work presented has been designed to a reasoner in which the solution features of the cases are numeric attributes, enabling us to employ the numerical approach easily. In order to provide a greater contribution to the CBR community, we still have to formulate a similar symbolic method. One alternative is to use an approach similar to the one implemented in PROTOS, making use of the prototypicality (Kolodner, 1993), where the solutions are previously categorized. The idea of discovering typicality over a set of symbolic cases might be possible with the use of hierarchical clusterings (Fisher, 1996).

Again, the directions should be determined by the specific applications, expectations and purposes.

The proposal of selecting the best match through the Theory of Typicality seems to be a very good choice in the development of automated CBR systems. The improvement in accuracy is still to be tested and it may vary depending on the domain. The successful implementation of this approach requires that the solution to the input problem is indeed the most typical of the set in the cases retrieved.

8. References

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

Bezdek, J.C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, New York.

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

Fisher, D. (1996). Iterative Optimization and Simplification of Hierarchical Clusterings. *Journal of Artificial Intelligence Research*, 4, 147-180.

Friedman, M., Schneider, M., & Kandel, A. (1989). The use of weighted fuzzy expected value (WFEV) in fuzzy expert systems, *Fuzzy Sets and Systems*, 31, 37-45.

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.

Gath, I. & Geva, B. (1989). Unsupervised Optimal Fuzzy Clustering. *IEEE Transactions Pattern Anal. Machine Intelligence*, PAMI-11, 7, 773-781.

Hennessy, D. and Hinkle, D.(1992). Applying Case-Based Reasoning to Autoclave Loading. *IEEE Expert*, 7, 5, 21-26.

Hurley, Neil (1993). A priori Selection of Mesh Densities for Adaptive Finite Element Analysis, using a Case-Based Reasoning Approach. *Topics in Case-Based Reasoning*, (First European Workshop, EWCBR-93). Wess, Stefan, Althoff, Klaus-Dieter & Richter, Michael (editors) Springer-Verlag, 379-391.

Kandel, A. (1982). *Fuzzy Techniques in Pattern Recognition*, John Wiley, New York.

Kaufmann A. (1975). *Introduction to the Theory of Fuzzy Subsets*, vol.1, Academic Press, New York.

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.

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

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

Kolodner, J. (1993). *Case-Based Reasoning*. Morgan Kaufmann, Los Altos, CA,.

Kopeikina, L., Brandau, R. & Lemmon, A. (1988). Case Based Reasoning for Continuous Control. *DARPA Proceedings of a Workshop on Case-Based Reasoning,* Clearwater Beach, Florida, May 10-13, 250-259.

Macchion, D. & Vo, D. P. (1993). Use of case-based reasoning technique in building expert systems. *Future Generation Computer Systems*, 9, 4, 311-319.

Mcintyre, S. H., Achabal, D. D. & Miller, C. M. (1993). Applying Case-Based Reasoning to Forecasting Retail Sales. *Journal of Retailing, 69* (4), 372-398.

Nakatani, Y., Tsukiyama, M. & Fukuda, T. (1991). Case organization in a case-based engineering design support system. *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, 3*, 1789-1794.

Pal, N. R. & Bezdek, J. C. (1995). On Cluster Validity for the Fuzzy c-Means Model. *IEEE Transactions on Fuzzy Systems*, 3, 3, 370-379.

Riesbeck, C. K. & Schank, R.C. (1989). *Inside Case-Based Reasoning*. Erlbaum, Hillsdale, NJ.

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

Shen, Z.L., Lui, H.C. & Ding, L.Y. (1994). Approximate Case-Based Reasoning on Neural Networks. *International Journal of Approximate Reasoning, 10* (1), 75-98.

Shinn, H.S. (1988). Abstractional Analogy: A Model of Analogical Reasoning. DARPA *Proceedings of a Workshop on Case-Based Reasoning*, Florida, May 10-13. Janet Kolodner (ed.) Morgan Kaufmann Publishers, 370-387.

Slade, S. (1991). Case-based reasoning: a research paradigm. *AI Magazine*, 12, 42-55.

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

Sugeno M. (1977). Fuzzy Measures and Fuzzy Integrals - A Survey. In *Fuzzy Automata and Decision Processes*, Gupta Madan M., Saridis, George N. & Gaines, Brian R. (eds.), North-Holland, New York.

Vassilliadis, S., Triantafyllos, G. & Kobrasly, W. (1994). A method for computing the most typical fuzzy expected value, *FUZZ-IEEE'94, Proceedings of the Third IEEE International Conference on Fuzzy Systems* (Orlando, FL), 2040-2045.

Wang, Z. & Klir, G. (1992). *Fuzzy Measure Theory*. Plenum Press, New York.

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.

Whalen, T. & Schott, B. (1985). Goal-Directed Approximate Reasoning in a Fuzzy Production System. *Approximate Reasoning in Expert Systems.* M.M.Gupta, A.Kandel, W.Bandler, J.B.Kiszka (editors) Elsevier Science Publishers B.V.(North-Holland).

Whitaker, Leslie A., Stottler, Richard H., Henke, Andrea, e King, James A. (1990). Case-Based Reasoning: Taming the similarity heuristic. *Proceedings of the Human Factors Society,* 312-315.

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

Zadeh, L.A. (1975). The Concept of a Linguistic Variable and it's Application to Approximate Reasoning-I. *Information Sciences, 8,* 199-249.