A NEW CASE-BASED APPROXIMATE REASONING BASED ON SPMF IN LINGUISTIC APPROXIMATION

Dae-Young Choi
Dept. of MIS, Yuhan University, Koean-Dong, Sosa-Ku, Puchon City, Kyungki-Do, South Korea

Ilkyeun Ra
Dept. of Computer Science & Engineering, Univ. of Colorado Denver, CO 80204 U.S.A.

Keywords: Case-based approximate reasoning (CBAR), Linguistic approximation.

Abstract: A new case-based approximate reasoning (CBAR) based on SPMF in linguistic approximation is proposed. It provides an efficient mechanism for linguistic approximation within linear time complexity.

1 INTRODUCTION

Case-based reasoning (CBR) is a problem-solving technique that reuses past experiences to find a solution. It is quite simple to implement in general, but it often handles complex and unstructured decision making problems very effectively. Moreover, it is maintained in an up-to-date state because the case-base is revised in real time, which is a very important feature for the real world applications. Due to its strength, CBR has been applied to various problem-solving areas including manufacturing, finance and marketing, intelligent product catalogs for Internet shopping malls, conflict resolution in air traffic control, semiconductors design, medical diagnosis (Ahn et al., 2007, Chiu, 2002, Chiu et al., 2003). While other major artificial intelligence techniques depend on generalized relationships between problem descriptors and conclusions, CBR utilizes specific knowledge of previously experienced and concrete problem situations, so it is effective for complex and unstructured problems and it is easy to update (Ahn et al., 2007). In recent years, CBR has received a great deal of attention and has been used successfully in diverse application areas. As a general problem solving methodology intended to cover a wide range of real-world applications, CBR must face the challenge to deal with uncertain, incomplete, and vague information. Building hybrid approaches by combining CBR with methods of uncertain and approximate reasoning (Zadeh, 1973, Mizumoto et al., 1982) plays an important role in many real-world applications. In this connection, a new case-based approximate reasoning based on SPMF in LA is proposed.

2 SPMF

Technology standards help ensure that packages and application services do not become piecemeal solutions so that you can leverage them across other initiatives. Additionally, enterprises with standards can respond more quickly to changing conditions than those without standards because creating information systems from compatible components is easier and less costly (Tanrikorur, 2001). As Mamdani (Mamdani, 2001) pointed out in 2001 BISC (Berkeley Initiative in Soft Computing) workshop on fuzzy logic and the Internet, it is time to think about 'standardization on fuzzy sets'.

Let A be a fuzzy set for a linguistic term and be a subset of the universal set X, then, for $x \in X$, a triangular-type membership function can be represented by using 3 points $\mu_A(x_L, x_M, x_H)$, where $x_L < x_M < x_H$, and if the result of this membership function is normalized to [0, 1] then $\mu_A(x_L, x_M, x_H) = 0$ for every $x \in [-\infty, x_L] \cup [x_H, \infty]$ and $\mu_A(x_L, x_M, x_H) = 1$ at x_M. A trapezoidal-type can be represented by using 4 points $\mu_A(x_L, x_{I1}, x_{I2}, x_H)$, where $x_L < x_{I1} < x_{I2} < x_H$, and if the result of this membership

Choi D. and Ra I. (2009).
A NEW CASE-BASED APPROXIMATE REASONING BASED ON SPMF IN LINGUISTIC APPROXIMATION.
In *Proceedings of the 11th International Conference on Enterprise Information Systems - Artificial Intelligence and Decision Support Systems*, pages 259-262
DOI: 10.5220/0001816002590262

function is normalized to [0, 1] then $\mu_A(x_L, x_{I1}, x_{I2}, x_H) = 0$ for every $x\in[-\infty, x_L]\cup[x_H, \infty]$ and $\mu_A(x_L, x_{I1}, x_{I2}, x_H) = 1$ at $[x_{I1}, x_{I2}]$. A more comprehensive study of standardized parametric membership functions (SPMF) can be found in (Chang et al., 1991).

3 THE PROPOSED METHOD

Based on their behavioral experiment, they recommended the five good distance measure (DM) (i.e., S_4, q_∞, q_*, Δ_∞, Δ_*) between fuzzy subset A and B of a universe of discourse U (Zwick et al, 1987). We note that the five good DM concentrate their attention on a single value rather than performing some sort of averaging or integration. In the case of S_4, attention focuses on the particular x-value where the membership function of $A\cap B$ is largest; in q_∞ and Δ_∞, attention focuses on the α-level set where the x-distance is largest; in q_* and Δ_*, attention focuses on the x-distance at the highest membership grade. Considering the result of their behavioral experiment, we know that the reduction of complicated membership functions to a single 'slice' may be the intuitively natural way for human beings to combine and process fuzzy concepts. Moreover, we know that the DM between two fuzzy subsets can be efficiently represented by a limited number of features. From this idea, a new case-based approximate reasoning (CBAR) based on SPMF in LA is proposed. In the case-based reasoning process, case indexing and retrieval are the most important steps because the performance of CBR systems usually depends on them (Ahn, 2007).

In this paper, we suggest linguistic case indexing and retrieval based on SPMF. We try to find efficiently the fuzzy subset of the linguistic value that is the most similar to the fuzzy subset resulting from observation (A′) related to a linguistic variable in a rule-base. We assume that there is the pre-defined set of linguistic variables (PSLV) sorted by alphabetically. Each linguistic variable in the PSLV has the pointer to indicate its own table with the relevant linguistic values. For example, the table regarding the linguistic variable '*age*' may be consisted of linguistic values such as '*young*', '*very young*' represented by fuzzy subsets. We assume that fuzzy subsets for linguistic values are defined by the SPMF.

The performance of CBR systems usually depends on case indexing and retrieval (Ahn, 2007). In the proposed linguistic case indexing and retrieval, we utilize the partitioning concept that disjoints the linguistic variables used in the process of CBAR. It can be used to avoid exploring the irrelevant linguistic values in the process of CBAR. Since the resulting fuzzy subset obtained from observation (A′) is related to its linguistic variable, we can find the related linguistic variable easily by referencing the linguistic variable matched in a rule-base. After the related linguistic variable (LV) is determined in the PSLV, its corresponding table (TBL) is obtained by using the pointer (P_i, i = 1, 2, …, n) associated with the linguistic variable. Based on the relevant fuzzy subsets represented by the SPMF in the table, the distances between all relevant fuzzy subsets in the table and an observation (A′) represented in the SPMF are computed by using Euclidean distance. Thus, the rule with the most similar linguistic value (i.e., linguistic value with the minimum distance) could be retrieved for the CBAR. An approximate transformation method (ATM) based on the SPMF was proposed in (Choi, 2006). The ATM transforms the non-parametric membership functions into the SPMF. The detail algorithms and their properties for the ATM were described in (Choi, 2006). We note that each table (TBL_i, i = 1, 2, …, n) is consisted of the relevant linguistic values represented by the SPMF. For example, the triangular-type or the trapezoidal-type will be defined as (x_L, x_M, x_H) or $(x_L, x_{I1}, x_{I2}, x_H)$, respectively in Section 2.

Example 1. We consider the linguistic variable '*age*'. For simplicity, we assume that the table regarding the linguistic variable '*age*' is consisted of 2 linguistic values such as '*young*' and '*very young*'. They are defined as $(x_L, x_{I1}, x_{I2}, x_H)_1 = (15, 20, 30, 35)_1 = A_1$ ('*young*') and $(x_L, x_{I1}, x_{I2}, x_H)_2 = (17, 20, 27, 30)_2 = A_2$ ('*very young*') respectively, by using the trapezoidal-type. We assume that an observation (A′) is parameterized to $(x_L', x_{I1}', x_{I2}', x_H') = (16, 20, 25, 29) = A'$ as in Figure 1. The detail algorithms for transforming the non-parametric membership functions into the SPMF were described in (Choi, 2006). The distances among fuzzy subsets are achieved by using Euclidean distance as follows :

$$d_1(A_1,A') = \sqrt{(15-16)^2 + (20-20)^2 + (30-25)^2 + (35-29)^2} = \sqrt{62}\,.$$

$$d_2(A_2,A') = \sqrt{(17-16)^2 + (20-20)^2 + (27-25)^2 + (30-29)^2} = \sqrt{6}\,.$$

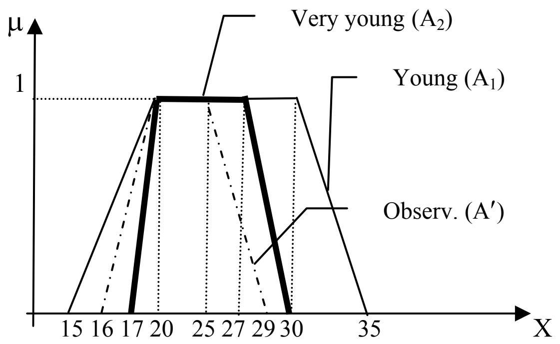

Figure 1: Fuzzy subsets on the linguistic variable '*age*'.

In this case, for an observation A′, the linguistic value '*very young*' is retrieved because $d_1 > d_2$. In a similar way, it may be extended to i, (i =1, 2, …, m), relevant linguistic values in the table. Thus, the linguistic value with min (d_1, d_2, …, d_m) is retrieved for CBAR in LA.

In this paper, we explain the proposed method by using trapezoidal-type membership function. It can be similarly extended to other SPMF such as Π-type, S-type, etc. The proposed method is achieved by using only the small number of parameters in SPMF. In addition, the proposed linguistic case indexing and retrieval utilize the partitioning concept that disjoints the linguistic variables (Zadeh, 1987) used in the process of CBAR. It can be used to avoid exploring the irrelevant linguistic values in the process of CBAR in LA. One of the crucial problems in real world applications is the computational speed of the applied method. Acceptable speed is generally achieved only if the time complexity is at most polynomial. In this respect, the proposed method is valuable because its time complexity is linear as shown in Example 1. It provides an efficient mechanism for LA within linear time complexity.

4 COMPARISONS

Wenstop (1976) suggested the linguistic computation that were almost entirely problem dependent. He may specify only two primary subsets in a universe of discourse composed of perhaps 25 elements. Obviously, this would not be very conducive if a close match was required in such a sparse space. Thus, some care should be taken to ensure a reasonable density of subsets within the primary space. Eshragh & Mamdani (1979) proposed another linguistic computation. But the computational complexity of their method for linguistic processing is very high. The reason is that the search procedure has two main phases. The first phase is exhaustive and the second phase is heuristic. The exhaustive phase takes care of trivial cases. That is, if a given subset shows characteristics similar to those of primary or negated primary subsets, then it will be tested against appropriate types of primary subsets for perfect match. If the exhaustive phase proves unsuccessful, the heuristic phase is applied. In this phase, the input is appropriately processed and its segments are separated. This search process is time-consuming. Degani & Bortolan (1988) proposed another linguistic computation. It is mainly useful for its use of clinically recognized linguistic terms whose meaning is rather well established in the medical community. Batyrshin & Wagbnknbcht (2002) described the problem of a linguistic description of dependencies in data by a set of rules R_k: "If X is T_k then Y is S_k" where T_k's are linguistic terms like *SMALL, BETWEEN 5 AND 7* describing some fuzzy intervals A_k, and S_k's are linguistic terms like *DECREASING* and *QUICKLY INCREASING* describing the slopes p_k of linear functions $y_k = p_k x + q_k$ approximating data on A_k. Their linguistic approach can be used for the calculation of granular derivatives of functional and statistical dependencies between linguistic variables in rules with aforementioned constraints. Their search approach for linguistic terms is time-consuming when merging fuzzy intervals of the partition obtained by the genetic algorithm and retranslation for generating rules from fuzzy partitions and linear approximation.

A key problem in the application of fuzzy set theory to real time control, expert systems, natural language understanding, etc., is devising relatively fast methods. The proposed linguistic case indexing and retrieval is efficiently obtained by using the parameters of the SPMF. Moreover, in order to avoid exploring the irrelevant linguistic values in the process of CBAR in LA, we use the partitioning concept that disjoints the linguistic variables used in the process of CBAR in LA. These features enable the proposed linguistic case indexing and retrieval to be processed relatively fast compared to the previous linguistic approaches (Batyrshin et al., 2002, Degani & Bortolan, 1988, Eshragh & Mamdani, 1979, Kowalczyk, 1998, Wenstop, 1976). From the engineering viewpoint, it may be a valuable advantage.

We briefly summarize the difference between the proposed method and existing linguistic approximation methods in Table 1.

Table 1: Comparisons.

Attributes	Existing methods	Proposed method
Membership f^{n}	Ad-hoc	SPMF
Method	Ad-hoc	CBAR
Complexity	Complex	Simple

5 CONCLUSIONS

A new CBAR based on SPMF in LA is proposed. Compared to existing linguistic approximation methods, the proposed LA is achieved by using only the small number of parameters in SPMF. In addition, the proposed linguistic case indexing and retrieval utilize the partitioning concept that disjoints the linguistic variables used in the process of CBAR in LA. It can be used to avoid exploring the irrelevant linguistic values in the process of CBAR in LA. These features enable the proposed linguistic case indexing and retrieval to be processed relatively fast compared to the previous linguistic approaches. It provides an efficient mechanism for LA within linear time complexity. Thus, the proposed method can be used to improve the speed of LA. In the meantime, a key problem in the application of fuzzy set theory to real time control, expert systems, natural language understanding, etc., is devising relatively fast methods. So, we propose a new CBAR based on SPMF in LA. From the engineering viewpoint, it may be a valuable advantage.

ACKNOWLEDGEMENTS

The author wishes to thank Prof. L. A. Zadeh, University of California, Berkeley, for his inspirational address on the perceptual aspects of humans in the BISC (Berkeley Initiative in Soft Computing) seminars and group meetings, and also thank Prof. Il Kyeun RA, University of Colorado, Denver, for his support.

REFERENCES

Ahn, H., Kim, K. J., Han, I., 2007, A Case-Based Reasoning System with the Two-Dimensional Reduction Technique for Customer Classification, Expert Systems with Application 32, 1011-1019

Batyrshin, I., Wagbnknbcht, M., 2002, Towards a Linguistic Description of Dependencies in Data, Int. Journal of Appl. Math. Compt. Sci., 12(3), 391-401.

Chang, T. C., Hasegawa, K., Ibbs, C. W., 1991, The Effects of Membership Function on Fuzzy Reasoning, Fuzzy Sets and Systems 44, 169-186.

Chiu, C., 2002, A Case-Based Customer Classification Approach for Direct Marketing, Expert Systems with Application 22, 163-168.

Chiu, C., Chang, P. C., Chiu, N. H., 2003, A Case-Based Expert Support System for Due-Date Assignment in a Water Fabrication Factory, Journal of Intelligent Manufacturing 14, 287-296.

Choi, D. Y., 2006, ATM Based on SPMF, Lecture Notes in Artificial Intelligence 4251, 490-497.

Degani, R., Bortolan, G., 1988, The Problem of Linguistic Approximation in Clinical Decision Making, International Journal of Approx. Reasoning 2(2), 143-162.

Eshragh, F., Mamdani, E. H., 1979, A General Approach to Linguistic Approximation, International Journal of Man-Machine Studies 11, 501-519.

Kowalczyk, R., 1998, On Linguistic Approximation with Genetic Programming, Lecture Notes in Computer Science 1415, 200-209.

Mamdani, E., 2001, Soft knowledge as key enabler of future services, Proceedings of the 2001 BISC international workshop on fuzzy logic and the Internet, 145-148.

Mizumoto, M., Zimmermann, H.-J., 1982, Comparison of Fuzzy Reasoning Methods, Fuzzy Sets and Systems 8, 253-283.

Tanrikorur, T., 2001, Great Expectations, Intelligent Enterprise 4(12), 35-38.

Wenstop, F., 1976, Deductive Verbal Models of Organization, International Journal of Man-Machine Studies 8, 293-311.

Zadeh, L. A., 1973, Outline of a New Approach to the Analysis of Complex Systems and Decision Processes, IEEE Transactions on SMC 3, 28-44.

Zadeh, L. A., 1987, The Concept of a Linguistic Variables and Its Application to Approximate Reasoning 1, 2, 3, In R. R. Yager, et al. (Eds.), Fuzzy Sets and Applications (John Wiley & Sons) 219-366.

Zwick, R., Carlstein, E., Budescu, D. V., 1987, Measures of Similarity among Fuzzy Concepts: A Comparative Analysis, International Journal of Approximate Reasoning 1, 221-242.