Journal of the Society of Korea Industrial and Systems Engineering
Vol, 25, No. 6, pp.54－59, December 2002.

Two-Phased Fuzzy Partitions with Fuzzy Equalization

Kyeongtaek Kim, Chongsu Kim

Division of Industrial & Systems and Mechanical Engineering, Hannam University, Taejon, Korea

퍼지 균등화조건을 갖는 2단 퍼지분할

김경택 · 김종수

한남대학교 산업시스템 · 기계공학부

퍼지 균등화는 어의론적으로 의미있고, 실험적으로 의미있는 언어레이블을 붙이도록 하는 조건이다. 지금까지 발표된 퍼지 균등화조건을 갖는 퍼지분할을 생성하는 알고리듬은 주어진 데이터에 대하여, 오직 하나의 퍼지분할만을 생성할 수 있었다. 만일 생성된 퍼지 분할이 더 이상 유용하지 못한 것으로 판명되면, 이 알고리듬은 주어진 데이터에 대한 퍼지 균등화조건을 갖는 또 다른 퍼지분할을 생성할 수 없다. 이는 생성된 퍼지분할을 사용하여 탐색적 발견을 수행하는 데이터마이닝의 경우 더 이상 프로세스가 진행되지 못함을 의미한다. 본 연구에서는 주어진 데이터에 대한 퍼지 균등화조건을 갖는 서로 다른 두 퍼지분할이 존재한다면, 어떠한 관계가 있는지를 증명하고, 이를 위치적 특성으로 서술한다. 또한 이 특성을 이용하여 퍼지 균등화조건을 갖는 퍼지분할을 원하는 만큼 생성할 수 있는 알고리듬을 제시하고, 예를 들어 설명한다.

Keywords : Two-phased fuzzy partitions; Data mining; Fuzzy equalization; Probability density function;

1. Introduction

Data mining involves the process of identifying interesting patterns and describing them in a concise and meaningful manner. For fully exploiting all the attributes of an object presented in the database, one must use the qualitative encapsulation. The concept of information granule provides such encapsulation. Information granulation is related to partitioning a class of objects into granules, with a granule being a collection of data that by consequence of their similarity, resemblance, indistinguishability, functionality, or operational cohesion are assembled or associated meaningfully [2, 12]. Such information granules reveal underlying relationships more comprehensively and are processed more efficiently. Modes of information granulation in which granules are crisp play important roles in a wide variety of methods, approaches, and techniques. Important though it is, crisp information granulation has a major limitation. It is well known that it fails to reflect the fact that in much of human reasoning and concept formation, granules are fuzzy rather than crisp [12].

The process of granulation that constructs fuzzy information granules from numeric data needs to be done on the sound ground. Information granules constructed from the process should be both semantically meaningful and experimentally meaningful [10]. Semantical meaningfulness of individual fuzzy sets and their families has been already thoroughly discussed [5, 8, 11, 13]. Experimental meaningfulness of a fuzzy set depends on its associated numeric data. If a fuzzy set represents a sufficient number of the associated numeric data, it is said to be experimentally meaningful. When criteria for placing an object into one or another fuzzy sets are indisputable and easily verifiable, fuzzy partitioning creates no problem of its own [3].

When each fuzzy set resulting from a fuzzy partition carries the same expected value, it is called fuzzy equalization. Fuzzy equalization is regarded as a basic vehicle of constructing fuzzy sets that are both semantically and experimentally meaningful [10]. Pedrycz [10] proposed an algorithm for constructing a fuzzy partition with fuzzy equalization condition. If it is applied to the data with the same probability density function, the algorithm always produces the same fuzzy partition. If the constructed fuzzy partition, however, proved not to be useful any more, one or more new fuzzy partitions should be generated for testing their usefulness in further exploring process in data mining. In this paper, we propose a new algorithm for construction of fuzzy partitions with fuzzy equalization condition. For a given peak of a fuzzy set, the proposed algorithm generates a fuzzy partition with fuzzy equalization which includes the fuzzy set. Thus, user can generate as many fuzzy partitions with fuzzy equalization as he changes the peak of the fuzzy set.

The paper is organized as follows. In Section 2, we present the preliminaries on fuzzy partition and fuzzy equalization. In Section 3, a few topological characteristics of two distinct fuzzy partitions with fuzzy equalization are explored for a given probability density function. Then, a new algorithm is proposed from them, which is followed by an example in Section 4. The concluding remarks are given in Section 5.

2. Fuzzy Equalization

The concept of fuzzy equalization is based on the fuzzy partition. Several concepts for the definition of a fuzzy partition have been proposed in the literature [4]. For further discussion, we adopt the definition of a fuzzy partition and its related terms as follows :

Definition 1. A set of fuzzy sets $A = \{A_1, A_2, ..., A_n\}$ is said to be a *fuzzy partition* of X [1, 7] if

$$\sum_{i=1}^{n} A_i(x) = 1 \quad \textit{for all } x \in X.$$

Definition 2. A fuzzy partition is called two-phased [6] if for any $x \in X$ there are at most two adjacent fuzzy sets $A_i, A_{i+1} \in A$ such that

$$\mu_{A_i}(x) \neq 0, \quad \mu_{A_{i+1}}(x) \neq 0.$$

Corollary 1. If $A = \{A_1, A_2, ..., A_n\}$ is a fuzzy partition on the closed interval $[a, b]$ and A_i (i=1, 2, ⌝, n) is a fuzzy number, then A must be two-phased [6].

Definition 3. Any fuzzy set can be defined over some universe of discourse X. In this universe, numeric data can be described by a certain probability density function (p.d.f.) $f(x)$. Then, the *probability of fuzzy event A* is

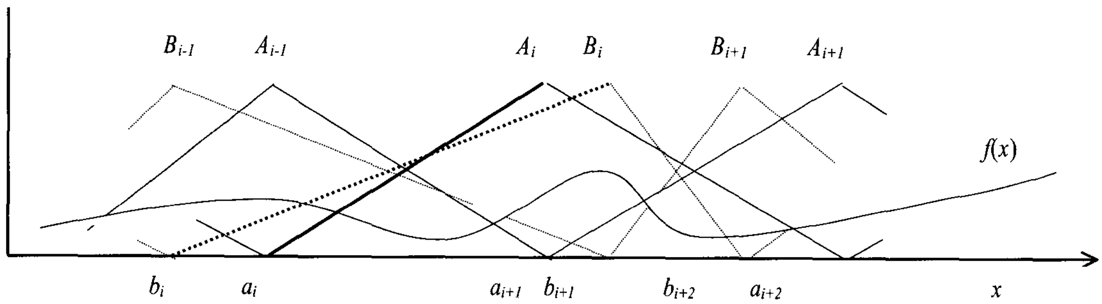
Definition 4. A fuzzy set can be viewed as meaningful if its probability $P(A)$ is equal to or exceeds a certain critical value . If it holds, we say A is *experimentally* justified [9].

Definition 5. A fuzzy partition $A = \{A_1, A_2, ..., A_n\}$ satisfies *fuzzy equalization* condition if [10]

$$P(A_1) = P(A_2) = \mathrm{L} = P(A_n) = \frac{1}{n}.$$

3. An Algorithm to find new fuzzy partitions

Based on the above definitions, lets consider a fuzzy partition $A = \{A_1, A_2, ..., A_n\}$ on a probability density function where both A_1 and A_n are trapezoidal fuzzy sets and A_i for $2 \le i \le (n\text{-}1)$ is a triangular fuzzy set $<a_i, a_{i+1}, a_{i+2}>$. Let a new fuzzy partition be $B = \{B_1, B_2, ..., B_n\}$ on the same probability density function which is distinct from A. Assume that both B_1 and B_n are trapezoidal fuzzy sets and B_i for $2 \le i \le (n\text{-}1)$ is a triangular fuzzy set $<b_i, b_{i+1}, b_{i+2}>$ and that $b_1 = a_1$ and $b_{n+2} = a_{n+2}$. By the Corollary 1, fuzzy partitions A and B, are two-phased. We will take a close look into two re-



<Figure 1> Two i-th fuzzy sets on the same probability density function.

lationships on the basis of both the relative position of b_{i+1} to a_{i+1} and the relative position of b_2 to a_2 : relationships of b_{i+2} and a_{i+2}, and the relationship of b_i and a. The difference of expected values for the increasing part (shown in dark lines in <Figure 1>) of membership function in two i-th fuzzy sets where $1 < i < n$ is as follows :

$$\int_{a_i}^{a_{i+1}} A_i(x)\,f(x)\,dx - \int_{b_i}^{b_{i+1}} B_i(x)\,f(x)\,dx$$
$$= \int_{a_i}^{a_{i+1}} [1 - A_{i-1}(x)]\,f(x)\,dx - \int_{b_i}^{b_{i+1}} [1 - B_{i-1}(x)]\,f(x)\,dx$$
$$= \int_{a_i}^{a_{i+1}} f(x)\,dx - \int_{b_i}^{b_{i+1}} f(x)\,dx - \int_{a_i}^{a_{i+1}} A_{i-1}(x)f(x)\,dx + \int_{b_i}^{b_{i+1}} B_{i-1}(x)f(x)\,dx$$
$$= \int_{a_i}^{a_{i+1}} f(x)\,dx - \int_{b_i}^{b_{i+1}} f(x)\,dx + \int_{a_{i-1}}^{a_i} A_{i-1}(x)f(x)\,dx - \int_{b_{i-1}}^{b_i} B_{i-1}(x)f(x)\,dx$$

$$\int_{a_{i-1}}^{a_i} A_{i-1}(x)f(x)\,dx - \int_{b_{i-1}}^{b_i} B_{i-1}(x)f(x)\,dx$$
$$= \int_{a_{i-1}}^{a_i} f(x)\,dx - \int_{b_{i-1}}^{b_i} f(x)\,dx + \int_{a_{i-2}}^{a_{i-1}} A_{i-2}(x)f(x)\,dx - \int_{b_{i-2}}^{b_{i-1}} B_{i-2}(x)f(x)\,dx$$

Since

$$\int_{a_1}^{a_2} A_1(x)f(x)\,dx - \int_{b_1}^{b_2} B_1(x)f(x)\,dx = \int_{b_2}^{a_2} f(x)\,dx,$$

without regard to the relative location of b_2 to a_2, the following equation holds.

$$\int_{a_i}^{a_{i+1}} A_i(x)\,f(x)\,dx - \int_{b_i}^{b_{i+1}} B_i(x)\,f(x)\,dx = \int_{b_{i+1}}^{a_{i+1}} f(x)\,dx. \quad \text{........ (1)}$$

By fuzzy equalization condition, the above equation is equivalent to the following equation.

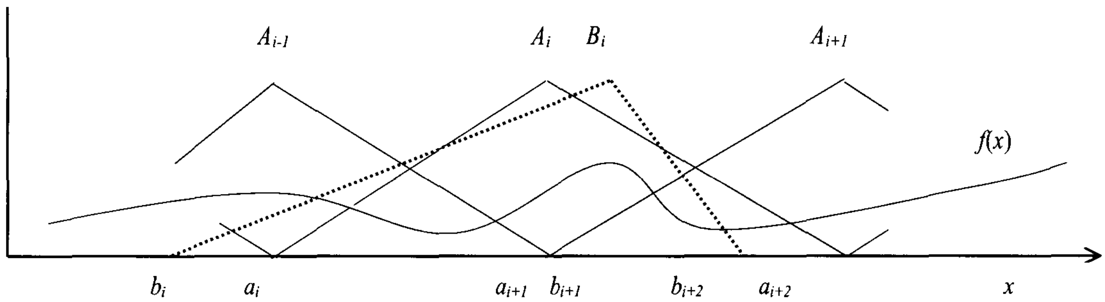
$$\int_{b_{i+1}}^{b_{i+2}} B_i(x)\,f(x)\,dx = \int_{a_{i+1}}^{a_{i+2}} A_i(x)\,f(x)\,dx + \int_{b_{i+1}}^{a_{i+1}} f(x)\,dx. \quad \text{........ (2)}$$

In summary, for two distinct fuzzy partitions with fuzzy equalization condition that have the same number of fuzzy sets, $\boldsymbol{A} = \{A_1, A_2, ..., A_n\}$ and $\boldsymbol{B} = \{B_1, B_2, ..., B_n\}$, the following topological characteristics hold from the above equations (1) and (2).

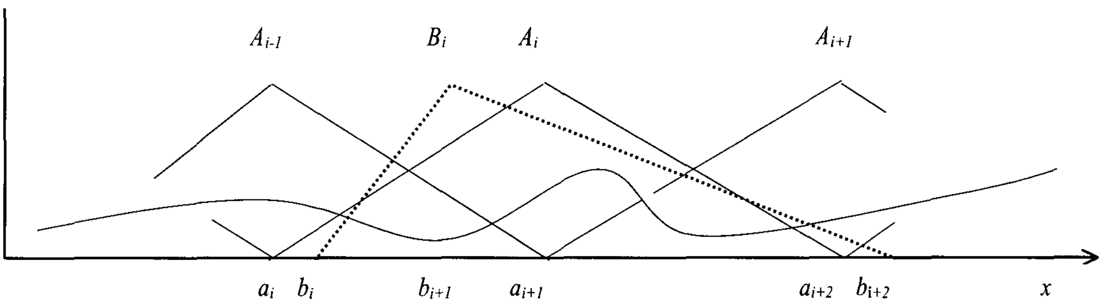
Topological characteristics of corresponding fuzzy sets in two distinct fuzzy partitions

If the lower bound of the support of B_{i+1} is greater than the lower bound of the support of the A_{i+1}, the fuzzy partition B has the following characteristics by fuzzy equalization as shown in <Figure 2>.

1) The lower bound of the support of the B_{i+2} should be less than the lower bound of the support of the A_{i+2} .



<Figure 2> Relative locations of b_i and b_{i+2} in case that $b_{i+1} > a_{i+1}$



<Figure 3> Relative locations of b_i and b_{i+2} in case that $b_{i+1} < a_{i+1}$

2) The lower bound of the support of the B_i should be less than the lower bound of the support of the A_i .
If the lower bound of the support of B_{i+1} is less than the lower bound of the support of the A_{i+1}, the fuzzy partition B has the following characteristics by fuzzy equalization as shown in <Figure 3>.
3) The lower bound of the support of the B_{i+2} should be greater than the lower bound of the support of the A_{i+2}.
4) The lower bound of the support of the B_i should be greater than the lower bound of the support of the A_i .

The above properties lead to the algorithm outlined below, which describes how to find a two-phased fuzzy partition with fuzzy equalization when the peak of one of fuzzy sets is given.

Suppose that the peak of a fuzzy set B_i is set to b^* and is distinct from the peak of A_i. Let the term of error be δ. (It means that the lower bound of the $(i+1)$_th fuzzy set in the fuzzy partition should be $(b^* - \delta)$, $(b^* + \delta)$ or some value between them.)

1. Find a fuzzy partition using the algorithm proposed by Pedrycz [10], say $\boldsymbol{A} = \{A_1, A_2, ..., A_n\}$.
2 Find a set of a fuzzy sets, $\boldsymbol{B} = \{B_1, B_2, ..., B_n\}$, as follows.
  1) Set the lower bound of the support of the B_1 to the lower bound of the support of the A_1 .
  2) Set the upper bound of the support of the B_n to the upper bound of the support of the A_n .
  3) Determine b_2
   3-1) (calculation of the ratio r)
     If $a_{i+1} < b^*$,
       $r = (b^* - a_{i+1}) / (a_{i+2} - a_{i+1})$
     else if $b^* < a_{i+1}$,
       $r = (a_{i+1} - b^*) / (a_{i+1} - a_i)$
     end if
   3-2) (calculation of b_2)
     If i is an even number
       if $a_{i+1} < b^*$,
         $b_2 = a_2 + r \times (a_3 - a_2)$
     else if $b^* < a_{i+1}$,
         $b_2 = a_2 - r \times (a_2 - a_1)$
     else if i is an odd number
       if $a_{i+1} < b^*$,
         $b_2 = a_2 - r \times (a_2 - a_1)$
       else if $b^* < a_{i+1}$,
         $b_2 = a_2 + r \times (a_3 - a_2)$
     end if
  4) For k=2 to i
     Set ε to
     $$\int_{b_{k-1}}^{b_k} B_{k-1}(x)\, f(x)\, dx .$$
     Determine b_{k+1} so that it may satisfy the condition
     $$\int_{b_k}^{b_{k+1}} B_{k-1}(x)\, f(x)\, dx = \frac{1}{n} - \varepsilon .$$
   Next
3. If $(b^* - \delta) \le b_{i+1} \le (b^* + \delta)$
   then go to step 4
  else
   replace $A_1, A_2, ..., A_{i-1}$ with $B_1, B_2, ..., B_{i-1}$ respectively and go to step 2
4. Find remaining fuzzy sets $\{B_i, B_{i+1}, ..., B_n\}$ as follows.
  For $k = i+1$ to $n-1$
     Set ε to
     $$\int_{b_{k-1}}^{b_k} B_{k-1}(x)\, f(x)\, dx .$$
     Determine b_{k+1} so that it may satisfy the condition
     $$\int_{b_k}^{b_{k+1}} B_{k-1}(x)\, f(x)\, dx = \frac{1}{n} - \varepsilon .$$
   Next
5. $\boldsymbol{B} = \{B_1, B_2, ..., B_n\}$ is a new fuzzy partition with fuzzy equalization.

4. Examples

Gaussian p.d.f. with a mean 3.0 and a variance 1.0 reads as follows :

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-3)^2}{2} \right)$$

We assume that the domain of the normal distribution is taken as a segment of real numbers spreading 1 to 5. We also assume that the number of fuzzy sets in the family of fuzzy sets is given 5. Then,

$$\int_1^5 f(x) = 0.9544 .$$

By step 1 of the proposed algorithm, the lower bound of the support of the $A2$, $a2$, can be determined from the following equation.

$\int_{1}^{a_2} f(x) = 0.0954$.

By the standard normal distribution table,

a_2 = 1.8160.

This leads to following resulting bound values :

1.0000
1.8160
2.5598
2.9539
3.5361
4.0704
5.0000

This is a fuzzy partition with fuzzy equalization. However, it seems to be natural that the peak of a fuzzy set superimposes on the peak of the normal distribution. Let's find such a fuzzy partition with the peak allowance of 0.0005. In this example, it is assumed that the peak of the third fuzzy set in the fuzzy partition should be located on the interval from 2.9995 to 3.0005. With this information, the first iteration of the algorithm suggested in the previous section gives us a result that the peak of the third fuzzy set in the new fuzzy partition is located on the point 3.0013. Since the peak of the third fuzzy set is not located in the allowed interval, another iteration of the algorithm is required. The second iteration results in the following bound values :

1.0000
1.8732
2.5122
2.9999
3.4881
4.1265
5.0000

Since the peak of the third fuzzy set, 2.9999, in the new fuzzy partition, is located within the allowed interval, this is a solution. Like this, a new fuzzy partition reflecting the users intention can be constructed.

5. Conclusions

As an explorative discovery process, data mining will face great difficulties if only one fixed grouping is allowed for the given data. Even with experimentally meaning encapsulation provided, fuzzy equalization is not widely applied to data mining due to the limitation of its algorithm that generates only one fuzzy partition for the given data. In this paper, we extracted a few topological characteristics of two distinct fuzzy partitions with fuzzy equalization condition for a given probability density function. From the extracted characteristics, we proposed an algorithm that generates new distinct fuzzy partitions with fuzzy equalization for the given probability density function. An example showed that a new fuzzy partition with fuzzy equalization could be generated as the user sets a peak value of a fuzzy set. This implies that another fuzzy partition can be constructed as the user chooses a different peak value of the fuzzy set.

References

[1] Bezdek, J.C.; Pattern Recognition with Fuzzy Objectives, Plenum Press, New York, 1981.

[2] Hirota, K., and Pedrycz, W.; Fuzzy Computing for Data Mining, Proceeding of the IEEE, 87(9) : 1575-1600, 1999.

[3] Kasumov, N.; Metric Properties of Fuzzy Partitions, Fuzzy Sets and Systems, 81 : 365-378, 1996.

[4] Klement, E.P., and Moser, B.; On the redundancy of fuzzy partitions, Fuzzy Sets and Systems, 85 : 195-201, 1997.

[5] Klir, G.J., and Yuan, B.; Fuzzy Sets and Fuzzy Logic and Applications, Prentice-Hall, Englewood Cliff, NJ, 1995.

[6] Ma, M., Turksen, I.B., and Kandel A.; Fuzzy Partition and Fuzzy Rule Base, Journal of Information Sciences, 108, 109-121, 1998.

[7] Negoita, C.V., and Ralescu, D.A.; Applications of Fuzzy Sets to Systems Analysis, Wiley, New York, 1975.

[8] Pedrycz, W., and Gomide, F.; An Introduction to Fuzzy Set s : Analysis and Design, MIT Press, Cambridge, MA, 1998.

[9] Pedrycz, W.; A Granular Signiture of Data, 69-73, 2000.

[10] Pedrycz, W.; Fuzzy Equalization in the Construction of Fuzzy Sets, Fuzzy Sets and Systems, 119 : 329-335, 2001.

[11] Zadeh, L.A.; Fuzzy Sets and Information Granularity, Advances in Fuzzy Set Theory and Applications, (ed.) M.M. Gupta, R.K. Ragade, and R.R. Yager, North Holland, Amsterdam, 3-18, 1979.

[12] Zadeh, L.A.; Toward a Theory of Fuzzy Information Granulation and its Uncertainty in Human Reasoning and Fuzzy Logic, Fuzzy Sets and Systems, 90, 11-127, 1997.

[13] Zimmermann, H.J.; Fuzzy Set Theory and Its Applications, Kluwer, Boston, MA, 1991.