Fuzzy Partitions with Fuzzy Equalization
- 퍼지 균등화 조건을 갖는 퍼지분할 -

Kyeongtaek Kim*
김 경 택
Chongsu Kim*
김 종 수

요 지

퍼지 균등화(fuzzy equalization)는 어의론적으로(semantically) 의미있고, 실험적으로(experimentally) 의미있는 언어라벨(linguistic labels)을 붙이도록 하는 조건이다. 지금까지 발표된 퍼지 균등화조건을 갖는 퍼지분할을 생성하는 알고리듬은 주어진 데이터에 대하여, 오직 하나의 퍼지분할만을 생성할 수 있다. 만일 생성된 퍼지 분할이 더 이상 유용하지 못한 것으로 판명되면, 이 알고리듬은 주어진 데이터에 대한 퍼지 균등화조건을 갖는 퍼지분할을 생성할 수 없다. 이는 생성된 퍼지분할을 사용하여 탐색적 발견을 수행하는 데이터마이닝인 경우 더 이상 프로세스가 진행되지 못함을 의미한다. 본 연구에서는 주어진 데이터에 대한 퍼지 균등화조건을 갖는 서로 다른 두 퍼지분할이 존재한다면, 어떠한 관계가 있는지를 증명하고, 위치적 특성을 서술하였다. 이 특성은 추후 퍼지 균등화조건을 갖는 퍼지분할을 원하는 만큼 생성할 수 있는 알고리듬을 만드는데 유용하게 사용 될 수 있다.

1. Introduction

Data mining (DM) 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 relates to partitioning a class of objects into granules, with a granule being a collection of data that by consequence of their similarity,

* Department of Industrial and Systems Engineering, Hannam University

resemblance, indistinguishability, functionality, or operational cohesion are assembled or associated meaningfully (Zadeh, 1979, 1997; Hirota and Pedrycz, 1999). 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 (Zadeh, 1997).

The process of granulation that constructs fuzzy information granules from numeric data need to be done on sound ground. Information granules constructed from the process should be both semantically meaningful and experimentally meaningful (Pedrycz, 2001). Semantically meaningfulness of individual fuzzy sets and their families has been already thoroughly discussed (Klir, et al., 1995; Pedrycz, 1998; Zadeh, 1979; Zimmermann, 1991). Experimental meaningfulness of a fuzzy set depends on its associated numeric data. If the fuzzy set represents 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 problems of its own.

When each fuzzy set resulting from a fuzzy partition carries 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 (Pedrycz, 2001). Pedrycz(2001) proposed an algorithm for constructing fuzzy partition with fuzzy equalization condition. However, for the same probability density function, the algorithm always produces the same fuzzy partition. If the constructed fuzzy partition is proved not to be useful any more, one or more new fuzzy partitions require to be generated for testing their usefulness in further exploring process in data mining. In this paper, we extract a few topological characteristics of two distinct fuzzy partitions with fuzzy equalization condition for a given probability density function.

2. Fuzzy Equalization

The concept of fuzzy equalization based on the fuzzy partition. Several concepts for the definition of a fuzzy partition have been proposed in the literature (Klement and Moser, 1979). For further discussion, we adopt the definition of fuzzy partition as follows:

Definition 1. A set of fuzzy sets $A = \{A_1, A_2, \dots, A_n\}$ is said to be a *fuzzy partition* of X (Bezdek, 1981; Negoita and Ralescu, 1975) if

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

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

$$P(A) = \int_x A(x) f(x) dx$$

Definition 2. Fuzzy set can be viewed as meaningful if its probability $P(A)$ is equal or exceeds a certain critical value. If it is holds, we say A is *experimentally* justified (Pedrycz, et al., 2000).

Definition 3. A fuzzy partition $A = \{A_1, A_2, \dots, A_n\}$ satisfies *fuzzy equalization* condition if (Pedrycz, 2001)

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

3. Topological Characteristics of Two Distinct Fuzzy Partitions

Based on the above definitions, let us consider a fuzzy partition, $A = \{A_1, A_2, \dots, A_n\}$ on a distribution where both A_1 and A_n are trapezoidal fuzzy sets and A_i for $1 < i < n$ is a triangular fuzzy set $<a_i, a_{i+1}, a_{i+2}>$. Let a new fuzzy partition be $B = \{B_1, B_2, \dots, B_n\}$, which is distinct from A, on the same distribution where both B_1 and B_n are trapezoidal fuzzy sets and B_i for $1 < i < n$ is a triangular fuzzy set $<b_i, b_{i+1}, b_{i+2}>$. We assume that $b_1 = a_1$ and $b_{n+2} = a_{n+2}$. We will look into the relationship of a_{i+2} and b_{i+2} on the basis of the two relationships: the relationship of a_2 and b_2 and the relationship of a_{i+1} and b_{i+1}. The difference of expected value for the increasing part of membership function in two i-th fuzzy sets where $1 < i < n$ is as follows.

$$\int_{a_i}^{a_{i+1}} A_i(x) f(x)\, dx - \int_{b_i}^{b_{i+1}} B_i(x) f(x)\, dx$$
$$= \int_{a_2}^{a_{i+1}} f(x)\, dx - \int_{b_2}^{b_{i+1}} f(x)\, dx + \int_{a_1}^{a_2} A_1(x) f(x)\, dx - \int_{b_1}^{b_2} B_1(x) f(x)\, dx \text{ ------------ (1)}$$

If $b_2 \geq a_2$, equation (1) is equivalent to the following equation:

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

If $b_2 \leq a_2$, equation (1) is equivalent to the following equation:

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

Therefore without regard to the relative position of a_2 and b_2, the following equation holds.

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

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

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

Let us look into equation (2) in terms the relative location of a_{i+1} and b_{i+1}.

1) $a_{i+1} > b_{i+1}$

In the case of $\int_{b_{i+1}}^{a_{i+2}} f(x)\, dx \neq 0$, b_{i+2} should be greater than a_{i+2}.

If $\int_{b_{i+1}}^{a_{i+2}} f(x)\, dx = 0$, b_{i+2} should be determined by the following equation:

$$\int_{b_{i+1}}^{b_{i+2}} B_i(x)\, f(x)\, dx = 0$$

2) $a_{i+1} < b_{i+1}$

In the case of $\int_{a_{i+1}}^{a_{i+2}} f(x)\, dx \neq 0$, b_{i+2} should be less than a_{i+2}.

If $\int_{a_{i+1}}^{a_{i+2}} f(x)\, dx = 0$, b_{i+2} should be determined by the following equation:

$$\int_{b_{i+1}}^{b_{i+2}} B_i(x)\, f(x)\, dx = 0$$

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

If the lower bound of the support of B_i is greater than the lower bound of the support of the A_i, The fuzzy partition B has the following characteristics by fuzzy equalization.

1) If there exist any interval between the lower bound of the support of the A_i and the lower bound of the support of the A_{i+1} on which the value of probability density function is not zero, the lower bound of the support of the B_{i+1} is less than the lower bound of the support of the A_{i+1}.
2) If there exist any interval between the lower bound of the support of the $A_{i\ 1}$ and the lower bound of the support of the B_i on which the value of probability density function is not zero, the lower bound of the support of the $B_{i\ 1}$ is located between the lower bound of the support of the $A_{i\ 1}$ and the lower bound of the support of the $A_{i\ 2}$.

If the lower bound of the support of B_i is less than the lower bound of the support of the A_i, The fuzzy partition B has the following characteristics by fuzzy equalization.

3) If there exist any interval between the lower bound of the support of the B_i and the lower bound of the support of the A_{i+1} on which the value of probability density function is not zero, the lower bound of the support of the B_{i+1} is located between the lower bound of the support of the A_{i+1} and the lower bound of the support of the A_{i+2}.
4) If there exist any interval between the lower bound of the support of the $A_{i\ 1}$ and the lower bound of the support of the A_i on which the value of probability density function is not zero, the lower bound of the support of the $B_{i\ 1}$ is greater than the lower bound of the support of the $A_{i\ 1}$.

4. Conclusions

As an explorative discovery process, data mining will face great difficulty if only one fixed grouping is allowed for given data. In spite of providing experimentally meaningful encapsulation, 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 extract a few topological characteristics of two distinct fuzzy partitions with fuzzy equalization condition for a given probability density function. We expect that the extracted characteristics lead to an algorithm that generates new distinct fuzzy partitions with fuzzy equalization for the given probability density function as many as the user wants.

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

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

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

[6] Pedrycz, W., and Gomide, F.; *An Introduction to Fuzzy Sets: Analysis and Design*, MIT Press, Cambridge, MA, 1998.

[7] Pedrycz, W.; "A Granular Signature of Data," *Proceedings of the 19th International Conference of the North American Fuzzy Information Processing*, pp. 69-73, 2000.

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

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

[10] 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.

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