퍼지 및 지능시스템학회 논문지 2001, Vol. 11, No. 4, pp. 307-310

랜덤 샘플에 의한 퍼지 집합의 특징성의 측도

A Measure of typicality of a fuzzy set with respect to a random sample

홍덕헌
Dug Hun Hong

대구가톨릭대학교 자동차 공학부

요 약

확률 공간에서 퍼지 집합의 특징성의 측도에 관하여 소개한다. 이에 관한 Yager의 결과들을 개선 및 일반화하고 랜덤샘플에 의한 퍼지 집합의 특징성의 퍼지측도에 대한 극한에 대하여 연구하고자 한다.

Abstract

A measure of typicality associated with any fuzzy subset on a probability space is introduced. We will improve some results of Yager and consider limit of the measure of typicality of a fuzzy set with respect to a random sample.

Key Word : Random sample, fuzzy set, typicality, fuzzy measure.

1. Introduction

In many applications of the emerging discipline of intelligent information analysis the issue of determining a typical (prototypical) object from a collection of objects arises. This question is of particular importance in development of knowledge discovering systems. The idea of a typical value also plays an important role in common sense reasoning systems where typical values are often used as default values. The usual context in which we desire to determine a typical value involves a situation in which we have a collection of observations and are using these observations as a basis to try find if there is some typical value characterizing this collection of observations. Informally speaking a typical value is a value that is the same or very similar to most the observations in the data we are trying to typify. The issue of typicality has been investigated by other researchers such as Yager [1,2,3], Zadeh [5], Dubois and prade [6], and Kandel [7,8]. In this note, we concentrate on Yager's results [4].

We will improve some results of Yager [4] and consider limit of the measure of typicality of a fuzzy set with respect to a random sample.

접수일자 : 2001년 3월 22일
완료일자 : 2001년 6월 28일

2. Definitions

Assume V is some attribute variable taking its value in the universe of discourse X is some subset of the real line, thus the values of V are numbers. If A is a fuzzy subset corresponding to linguistic value over the space X then for each $x \in X$, the degree of membership of x in A,$A(x)$, denotes the compatibility of the value x with the concept being represented by A. In addition we shall assume that the fuzzy A is normal; there exists at least one element in A having membership grade equal to one. A fuzzy subset A over the space X is said to be unimodal if there exists two values $u \le v$ such that

1. $A(x) \le A(y)$ if $x < y < u$
2. $A(x) = 1$ $u \le x \le v$
3. $A(x) \le A(y)$ $x > y > v$.

We shall call the interval $[u, v]$ the focus of A.

Again assume V is a variable whose domain of discourse X is a subset of the real line which we shall consider to be the interval $[a, b]$. Furthermore, we shall assume that D is a collection of n observations of V, drawn from X. We conjecture that a typical value of V based upon the data D shall be a unimodel fuzzy subset A of X. Furthermore, for A to be a typical value of D we require it to satisfy two criteria. The first condition is a requirement that most of the elements in the collection D

be compatible with the concept A suggested as the typical value. This condition assures us that typical value is indeed representative of the data set D. The second condition associated with an acceptable typical value is that it should be narrow in bandwidth, specific in space. In essence although we are allowing fuzzy sets to typical values we desire these fuzzy sets to be "pointlike," in particular, a typical value should have a small granularity with respect to the possible range of values X.

As we shall see the need to satisfy these two conditions will lead to a measure of typicality that assigns a notion of degree of typicality to any subset A.

We now shall suggest formulations for measuring the degree to which a proposed fuzzy subset A satisfies the idea of being a typical value for a collection of data D. Assume A is a proposed typical value associated with collection D. We shall define the compatibility of A with respect to D as

$$\text{Comp}(A/D) = \sum_{d_i \in D} \frac{A(d_i)}{n}$$

where n is the cardinality of D. We note that this evaluates to a number in the unit interval. We also note the monotonicity with respect to the $A(d_i)$, the larger these values the larger the compatibility. We see that this measure is a generalization of the ratio we introduced earlier in determining the model. In particular if we consider the set A to be the crisp point $\{x\}$ where $A(x)=1$ and $A(y)=0$ for all $y \neq x$ then

$$\text{Comp}(A/D) = \frac{n_x}{n}.$$

The requirement of narrowness of the typical value requires the introduction of the concept of specificity. Assume A is fuzzy subset of the space X, the interval $[a, b]$ and a probability measure P on X, the specificity of A with respect to X is defined as

$$\text{Spec}[A/X] = 1 - \int A(x)dP(x).$$

We also note that this is a number in the unit interval. We see that this measure takes its maximum value of one when A is a singleton fuzzy subset, there exits one element $x^* \in X$ such $A(x^*)=1$ and $A(x)=0$ for all others. Furthermore, this takes its minimal value of zero when $A(x)=1$ for all $x \in X$. In addition we note that if $A \subset B$, $A(x) \leq B(x)$ for all x, then $\text{Spec}(A/X) \geq \text{Spec}(B/X)$.

Thus we see this gives bigger values the smaller the subset A. In essence this provides some measure of pointness of A in regard to the space X. It is interesting to note the role the size of the underlying domain, $b-a$, plays in the above measure. In particular we see the large $b-a$ the wider the fuzzy subset A can be and still be considered pointlike.

Having introduced these two measures we suggest that the measure of typicality of fuzzy set A with respect to a data collection D is obtained by considering the conjunction of these two requirements:

$$\text{Typ}(A/D) = \text{Spec}(A/X) \wedge \text{Comp}(A/D)$$

where $\wedge$ is the minimum operator.

We note that this measure does indeed satisfy our original intuition of a typical value, a value for which most of the element in D are tightly clustered about. We see that the condition of specificity forces us to have a narrow value and the compatibility condition requires that most of the elements in D fall in A.

We notice that a kind of tension exists regarding the size of the fuzzy subset A in the formulation of Typ(A/D). On one hand we can increase the compatibility by increasing the bandwidth of A but in doing so we are causing the specifcity to diminish. Thus if A and B are two fuzzy subsets such that $A \subset B$, for all x, then

$$\text{Comp}(A/D) \leq \text{Comp}(B/D)$$

$$\text{Spec}(A/X) \geq \text{Spec}(B/X).$$

The existence of this tension makes the problem of finding an optimal typical value complex.

3. Results

In the following we generalize and provide a better lower bound on the measure of typicality than that in Yager's paper [4].

Theorem 1. Assume D is a data set from the domain X with a probability measure P. Assume the rang of elements in D span the interval $[u, v]$, then there exits a fuzzy set B associated with this data such that

$$\text{Typ}(B/D) \geq (1-(f(v)-f(u))) \vee \frac{1}{2},$$

where $f(u) = \int_X 1_{(-\infty, u]}(x)dP(x)$, and $\vee$ is the maximum operator.

Proof. First, we assume $f(v)-f(u) \leq \frac{1}{2}$ and define a fuzzy number B such that $B(x)=1$ for $x \in [u, v]$ and 0, otherwise. Then clearly $Comp(B/D)=1$ and $\text{Spec}[B/X] = 1-(f(v)-f(u))$.

Hence

$$\mathrm{Typ}(B/D) = \mathrm{Spec}(B/X)$$
$$= 1-(f(v)-f(u)) \geq \frac{1}{2}.$$

Let $(v)-f(u)>\frac{1}{2}$. Find $w \in [u, v]$ such that $P\{w\} = f(w+) - f(w-) = 0$. Define B such that $B(w) = 1$ and $\frac{1}{2}$, otherwise. Then clear B is a fuzzy subset and $\mathrm{Spec}(B/X) = \mathrm{Comp}(B/D) = \frac{1}{2}$, and hence

$$\mathrm{Typ}(B/D) \geq (1-(f(v)-f(u)) \vee \frac{1}{2}.$$

If we put

$X=[a,b]$ and $P(A)=\int_A \frac{1}{b-a}dx$, we have the following result as a corollary.

Corollary 1. Assume D is a data set from the domain X with a probability measure P. Assume the range of elements in D span the interval $[u,v]$, then there exits a fuzzy set B associated with this data such that

$$\mathrm{Typ}(B/D) \geq (1-\frac{v-u}{b-a}) \vee \frac{1}{2}.$$

We note that Corollary 1 has better lower bound than the following result of Yager [4].

Theorem [4]. Assume D is a data set drawn from the domain $X = [a, b]$. Assume the range of elements in D span the interval $[u, v]$, then there exists a typical value B associated with this data such that

$$\mathrm{Typ}(B/D) = 1-\frac{v-u}{b-a}$$
$$= \frac{(b-v)+(u-a)}{b-a}$$

Remark. In Theorem 1, $\frac{1}{2}$ is the critical value. If it is bigger than $\frac{1}{2}$, the theorem does not work. For this, let $\delta > \frac{1}{2}$, then we can construct a sets D such that $\mathrm{Typ}(B/D) < \delta$ for any B.
Actually, let $D = \{1,2\}$

and $P\{1\} = P\{2\} = \frac{1}{2}$. Then

$$\mathrm{Comp}(B/D) = \frac{1}{2}(B(1)+B(2))$$
and
$$\mathrm{Spec}(B/X) = 1-\frac{1}{2}(B(1)+B(2))$$

Therefore,

$$\mathrm{Typ}(B/D) = \frac{1}{2}(B(1)+B(2))$$
$$\wedge (1-\frac{1}{2}(B(1)+B(2))$$
$$\leq \frac{1}{2} < \delta$$

Theorem 2. Assume D_n is a n data set of random sample from the domain X. Assume the range of elements is D_n span the interval $[u, v]$. Then for any fuzzy set A, with probability 1.

$$\lim_{n\to\infty} \mathrm{Typ}(A/D_n)$$
$$= \min(EA, 1-EA) \leq \frac{1}{2},$$
where $\quad EA = \int A(x)dP(x).$

Proof. It is immediate from the facts that

$$\lim_{n\to\infty} \mathrm{Comp}(A/D) = \lim_{n\to\infty} \frac{1}{n} \sum_{d_i \in D_n} A(d_i)$$
$$= E(A)$$
$$= 1 - \mathrm{Spec}[A/X],$$

where the second equality comes from *SLLN*.

Theorem 3. Assume a fuzzy subset A has a degree of typicality $\beta > 0$ associated with a given data set. Any fuzzy subset contained within the negation of A has a degree of typicality less than $1-\beta$.

Proof. Assume

$$\mathrm{Typ}(A/D)$$
$$= \mathrm{Spec}(A/X) \wedge \mathrm{Comp}(A/D)$$
$$= \beta > 0,$$

This implies that

$$\mathrm{Comp}(A/D) = \frac{1}{n} \sum_{i=1}^{n} A(d_i) \geq \beta.$$

Consider the negation of A, $\overline{A}$ where $\overline{A}(x) = 1 - A(x)$. In this case

$$\mathrm{Comp}(\overline{A}/D) = \frac{1}{n} \sum_{i=1}^{n} (\overline{A}(d_i))$$
$$= \frac{1}{n} \sum_{i=1}^{n} (1-A(d_i))$$
$$= 1 - \frac{1}{n} \sum_{i=1}^{n} A(d_i)$$
$$\leq 1-\beta.$$

Let B be any fuzzy subset contained in $\overline{A}$, $B(x) \leq \overline{A}(x)$ for all x. Then

$$\mathrm{Comp}(B/D) \leq \mathrm{Comp}(\overline{A}/D) \leq 1-\beta$$

and hence

$$\mathrm{Typ}(B/D) \leq \mathrm{Comp}(B/D) \leq 1-\beta.$$

The following result, which is a special case of Theorem 3, is due to Yager [4].

Theorem [4]. Assume a fuzzy subset A has a degree of typicality $\alpha \geq 0.5$ associated with a given data set. Any fuzzy subset contained within the negation of A has a degree of typicality less than 0.5.

In introducing the idea of a typical value we generalized the concept of a mode by allowing the typical value to be a linguistic value, represented as a "narrow" fuzzy subset and we also required that most of observations are compatible with this proposed typical value. Informally we shall say an element is prototypical if this element can be used an exemplifier of the data in our observations.

Definition. Assume X is the underlying space from which our observations are drawn. The degree to which an element $x \in X$ is a prototypical value of a data set D, denoted $PT(x)$, is expressed as follows:

$$PT(x) = \text{Max}_{A \subseteq X}[A(x) \wedge \text{Typ}(A/D)].$$

Essentially we see that a prototypical value is an element which has a high membership grade in a typical value of the data. It is important to emphasize that a prototypical value is not a fuzzy subset but is specific element from X.

Theorem 4. Assume D_n is a n data set of random sample from the domain $X = [a, b]$ with probability P which has density function. Assume the range of elements in D_n span the interval $[a, b]$. Then we have, for all $x \in X$ with probability 1,

$$\lim_{n \to \infty} \text{PT}_n(x)$$
$$= \lim_{n \to \infty} \text{Max}_{A \subseteq X}[A(x) \wedge \text{Typ}(A/\ D_n)]$$
$$= \frac{1}{2}.$$

Proof. By Theorem 2 we have $\lim_{n \to \infty} \text{Typ}(A/\ D_n) = E(A) \wedge (1 - E(A))$. If we construct A such that $A(x) \geq \frac{1}{2}$ and $E(A) = \frac{1}{2}$, then we are done. Since P has density function, we can find $\varepsilon > 0$ such that

$$\int 1_{[x-\varepsilon(x-a),\, x+\varepsilon(b-x)]}(y)\, dP(y) = \frac{1}{2}.$$

Now define $A(x) = 1$ on

$$[x - \varepsilon(x-a), x + \varepsilon(b-x)]$$

and 0, otherwise. Then clearly $A(x) = 1$ and $E(A) = \frac{1}{2}$, which completes the proof.

4. Conclusion

We have improved some results of Yager [4] about fuzzy measure of typicality. We have also considered limit of the measure of typicality of a fuzzy set with respect to a random sample.

References

[1] R. R. Yager and T. C. Rubinson, "Linguistic of data bases," *Proc. IEEE Conf.* on Decision and Control, San Diego, pp. 1094-1097, 1981.

[2] R. R. Yager, "A new approach to the summarization of data," Inform. Sci., vol. 28, pp. 69-86, 1982.

[3] R. R. Yager, "On linguistic summaries of data," in *Knowledge Discovery in Databases,* G. Piatetsky-Shapiro, & B. Frawley, Eds., Cambridge, MA, MIT Press, pp. 347-363, 1991.

[4] R. R. Yager, A note on a fuzzy measure of Typicality, *International Journal of Intelligent Systems*, vol. 12, pp. 233-249, 1997.

[5] L. A. Zadeh, "A note on prototype theory and fuzzy sets," *Cognition*, vol. 12, pp. 291-292, 1882.

[6] D. Dubois and H. Prade, "On data summarization with fuzzy sets," *Proceedings of the Fifth International Fuzzy Systems Association World Congress*, pp. 465-468, 1993.

[7] M. Schneider and A. Kandel, "Properties of the fuzzy expected value and the fuzzy expected interval in fuzzy environment," Fuzzy sets syst., vol. 28, pp. 55-68, 1988.

[8] A. Kandel and M. Friedman, "On weighted expectations," *Proceedings of Joint Conference of NAFIPS, IFIPS and NASA*, San Antonio, pp.13-17. 1994.

저 자 소 개

홍 덕 헌 (Dug Hun Hong)
1974~1981년 : 경북대 수학과
1981~1983년 : 경북대 수학과 대학원
1985~1990년 : Univ. of Minnesota 수학과, MS, Ph. D
1991~1995년 : 대구효성가톨릭대학교 자동차 공학부 교수
1996~현재 : 대구가톨릭대학교 자동차공학부 교수
1997~1998년 : North Carolina State Univ. 방문연구
관심분야 : 퍼지이론 및 응용전 분야
E-mail : dhhong@cuth.cataegu.ac.kr