WE1–3

A proposal of FFD based on Conditional Probability

Rolly Intan Masao Mukaidono

Departement of Computer Science,
Meiji University,
Kanagawa, Japan.

Abstract

This paper introduces a proposal of fuzzy functional dependency (FFD) based on conditional probability. Some properties of conditional probability and its relation with fuzzy sets have been studied. Conditional probability of two fuzzy sets was proposed as a basis of getting similarity relation of two fuzzy sets. By using this property, a proposal of FFD is introduced and proved by example that it satisfies classical/crisp relational database.

1 Introduction

In this paper, we have studied some properties of conditional probability and its relation with fuzzy sets. Even when we realize that interpretation of numerical value between fuzzy sets and probability measures are philosophically distinct, basic operations such as intersection and union of two fuzzy values can be interpreted as maximum intersection and minimum union of two probabilities. Considering this reason, we define three approximate conditional probabilities of two fuzzy sets which are based on minimum, independent and maximum probability intersection between two events. Moreover, conditional probability of two fuzzy sets can be interpreted as probabilistic matching of two fuzzy sets (Baldwin J.F. and Martin T.P., 1996) [2] and as a basis of getting similarity of two fuzzy sets and constructing equivalence classes inside their domain attribute.

By using this property, we construct fuzzy functional dependency (FFD) and prove that it satisfies classical/crisp relational database by example. Considering the concept of FFD, various definitions and the notation have been devised since 1988. Among them are, Raju and Majumdar(1988) [12], defined FFD base on the membership function of the fuzzy relation. Tripathy (1990) [13] proposed definition of the FFD in terms of fuzzy Hamming weight. A. Kiss (1991) [8], constructed FFD by using weighted tuple. G.Chen(1995) [5], Cubero (1994) [6] and W.Liu (1992,1993) [11], introduced definition of the FFD based on the equality of two possibility distributions, nevertheless they used a different type of implication and different expression of cut off. S.Liao(1997) [10], gave design of the FFD by introducing semantic proximity. However, from the technical point of view, we realized that our constructed FFD is different from most FFD, which generally start with the definition of classical functional dependency and weaken the equality relation into a (gradual) resemblance relation (and then choose an appropriate implication) (Bosc, P., Dubois, D., and Prade, H., 1998) [4].

2 Preliminary

2.1 Conditional Probability

Conditional Probability of an event is the probability of the event occurring given that another event has already occurred.

Definition 2.1 Given H and D are two events over a sample space U. $\mathrm{P}(H|D)$ is defined as conditional probability for H given D. Relation between conditional and unconditional probability satisfy the following equation:

$$\mathrm{P}(H|D) = \frac{\mathrm{P}(H \cap D)}{\mathrm{P}(D)}, \qquad (1)$$

where suppose D is an event such that $\mathrm{P}(D) \neq 0$.

In particular, conditional probability satisfies some axioms as follows:

1. $\mathrm{P}(A|B) = 0$ if A and B are disjoint,
2. $\mathrm{P}(\bar{A}|B) + \mathrm{P}(A|B) = 1$,
3. $\mathrm{P}(B|B) = 1$.

2.2 Functional Dependency (FD)

Functional dependency(FD) as one type of integrity constraints has been known and used widely in the design of database system.

Definition 2.2 Given U is the set of attributes and R is a relation over U. The functional dependencies $X \to Y$ holds over $R(U)$ iff:

$$\forall t_i, t_j \in R, (t_i[X] = t_j[X] \Rightarrow t_i[Y] = t_j[Y]), \qquad (2)$$

where $X, Y \subseteq U$ and $t_i[X]$ denotes the restriction of the tuple t_i to the attributes belonging to X.

FDs satisfy some properties, such as (Armstrong's Axioms) [1]:

1. Reflexivity: $Y \subseteq X \Rightarrow X \to Y$,
2. Augmentation: ($X \to Y$ and $Z \subseteq U$, where U: set of attributes) $\Rightarrow X \cup Z \to Y$,
3. Transitivity: ($X \to Y$ and $Y \to Z$) $\Rightarrow X \to Z$.

2.3 Imprecise Data and Fuzzy Sets

In the real-world application, data are often imprecise. The level of precise data is assigned from *total ignorance*(TI) to *crisp* data, where total ignorance and crisp express the most imprecise and the most precise representation of data respectively.

Fuzzy set can be used as a connector to represent imprecise data from *total ignorance* to *crisp* data. The following definition shows how fuzzy set can be used to represent total ignorance and crisp data.

Definition 2.3 Let U be universal set, where $U = \{u_1, u_2, ..., u_n\}$. *Total ignorance*(TI) over U and *crisp* of $u_i \in U$ are defined as

$$\text{TI over } U = \{1/u_1, ..., 1/u_n\}, \quad (3)$$

$$\text{Crisp}(u_i) = \{0/u_1, ..., 0/u_{i-1}, 1/u_i, 0/u_{i+1}, ..., 0/u_n\}, \quad (4)$$

respectively.

3 Conditional Probability of Two Fuzzy Sets

Due to Definition 2.1, conditional probability can be represented by the following formula:

$$\mathrm{P}(H|D) = \frac{\mathrm{P}(H \cap D)}{\mathrm{P}(D)}.$$

The problem is how can we interpret $\mathrm{P}(H \cap D)$, if H and D are two events. We realize that numerical values of both fuzzy and probability philosophically have different meaning. In the fuzzy, numerical value can be interpreted as the level of preference or similarity. Therefore, intersection of two membership functions is overlapping between them and it can be treated by using minimum function.

On the other hand, numerical value of an event in probability proportionally express the number of ways that the event might occur. Every way is assumed having the same value and provided by a function called *basic probability assignment* as follows:

$$\mathrm{P}(u) = \frac{1}{|U|}, \quad \forall u \in U,$$

where U is set of all ways. Intersection of two events can be interpreted as the number of the same ways belong to both of them that the two events might occur. In the situation of lack of information, intersection of two events can be interpreted into three interpretations as follows:

1. *minimum probability of intersection*,
$$\mathrm{P}(H \cap D)_{\min} = \max(0, \mathrm{P}(H) + \mathrm{P}(D) - 1),$$
2. *independent probability of intersection*,
$$\mathrm{P}(H \cap D)_{\text{ind}} = \mathrm{P}(H) \cdot \mathrm{P}(D),$$
3. *maximum probability of intersection*,
$$\mathrm{P}(H \cap D)_{\max} = \min(\mathrm{P}(H), \mathrm{P}(D)).$$

Relation among them is shown as follows:

$$\mathrm{P}(H \cap D)_{\min} \leq \mathrm{P}(H \cap D)_{\text{ind}} \leq \mathrm{P}(H \cap D)_{\max}.$$

Now, we define *conditional probability* between two fuzzy sets based on the three interpretations above as follows.

Definition 3.1 Let $f = \{\chi_1^f/u_1, ..., \chi_n^f/u_n\}$ and $g = \{\chi_1^g/u_1, ..., \chi_n^g/u_n\}$ are two fuzzy sets over $U = \{u_1, u_2, ..., u_n\}$. $\mathrm{P}(f|g)$ is defined as conditional probability for f given g.

1. Based on minimum probability of intersection :
$$\mathrm{P}(f|g) = \frac{\sum_{i=1}^n \max(0, \chi_i^f + \chi_i^g - 1)}{\sum_{i=1}^n \chi_i^g}. \quad (5)$$
It can be proved that,
$$\mathrm{P}(f, f) \leq 1,$$
$$\mathrm{P}(\bar{f}|g) + \mathrm{P}(f|g) \leq 1.$$
2. Based on independent probability intersection :
$$\mathrm{P}(f|g) = \frac{\sum_{i=1}^n \chi_i^f \cdot \chi_i^g}{\sum_{i=1}^n \chi_i^g}. \quad (6)$$
It can be proved that,
$$\mathrm{P}(f, f) \leq 1,$$
$$\mathrm{P}(\bar{f}|g) + \mathrm{P}(f|g) = 1.$$
3. Based on maximum probability intersection :
$$\mathrm{P}(f|g) = \frac{\sum_{i=1}^n \min\{\chi_i^f, \chi_i^g\}}{\sum_{i=1}^n \chi_i^g}. \quad (7)$$
It can be proved that,
$$\mathrm{P}(f, f) = 1,$$
$$\mathrm{P}(\bar{f}|g) + \mathrm{P}(f|g) \geq 1.$$

Conditional probability in (7), principally is the same as *fuzzy relative cardinality*(Dubois and Prade, 1982 [7]) as shown in the following equation:

$$I(F, G) = \frac{|F \cap G|}{|F|}, \quad (8)$$

where $|F| = \sum_u \mu_F(u)$ and intersection is defined as minimum. Kosko [9] has pointed out the analogy between $I(F, G)$ and a conditional probability $\mathrm{P}(A|B)$, where B and F play the same role.

4 Fuzzy Functional Dependency

This section is the main concern in our paper. In this section, we introduce the construction of fuzzy functional dependency (FFD) by applying the concepts of conditional probability of two fuzzy sets as discussed in Section 3. From the technical point of view, we realize that our FFD is different from most FFDs as mentioned in Section 1. Generally, most of FFD start with the definition of classical functional dependency (see Definition 2.2) and weaken the equality relation into a (gradual) resemblance relation (and then choose an appropriate implication) [4].

Definition 4.1 Given U is the set of attributes and R is a relation over U. The fuzzy functional dependencies $X \rightsquigarrow Y$ holds over $R(U)$ iff:

$$\forall x \subseteq X,\ \forall y \subseteq Y,$$

$$\text{if } t(x,y) \in R \text{ then } \mathrm{PR}(x|y) \leq \mathrm{PR}(y|x). \quad (9)$$

Here $X, Y \subseteq U$, t denotes the tuple in relation R and $\mathrm{PR}(x|y)$ is called the *conditional probability relation* for x given y . If there are n tuples in R, then

$$\mathrm{PR}(x|y) = \frac{\sum_{i=1}^{n} \min(\mathrm{P}(x|t_i(X)), \mathrm{P}(y|t_i(Y)))}{\sum_{i=1}^{n} \mathrm{P}(y|t_i(Y))}, \quad (10)$$

where $t_i(X)$ and $t_i(Y)$ denote the restriction of the tuple t_i to the attributes belonging to X and Y, respectively.

Let $X = \{a_1, a_2, ..., a_m\}$, $x_i = \{\chi_1^i/a_1, ..., \chi_m^i/a_m\}$, and $x_k = \{\chi_1^k/a_1, ..., \chi_m^k/a_m\}$, then

$$\mathrm{P}(x_i|x_k) = \frac{\sum_{l=1}^{m} \min\{\chi_l^i, \chi_l^k\}}{\sum_{l=1}^{m} \chi_l^k}, \quad (11)$$

where x_i and x_k are two fuzzy sets over X.

Our consideration to apply minimum function for calculation of conditional probability of two fuzzy sets as shown in (11) such as $\mathrm{P}(x|t_i(X))$ and $\mathrm{P}(y|t_i(Y))$ in (10), with the objective of getting similarity of two fuzzy sets and developing their equivalence classes inside attribute X and Y, respectively.

It can be proved that the FFD satisfies some basic inference rules such as *reflexivity*, *augmentation*, and *transitivity* which are similar to Armstrong's Axioms as follows.

1. Reflexivity:$Y \subseteq X \Rightarrow X \rightsquigarrow Y$,
2. Augmentation:$(X \rightsquigarrow Y$ and $Z \subseteq U) \Rightarrow X \cup Z \rightsquigarrow Y$,
3. Transitivity:$(X \rightsquigarrow Y$ and $Y \rightsquigarrow Z) \Rightarrow (X \rightsquigarrow Z)$.

Proof. :

1. $Y \subseteq X$, in the sense of number of conditions or constraints of X is greater or equal to number of conditions or constraints of Y, implies that $\mathrm{P}(Y) \geq \mathrm{P}(X)$. Related to Definition 4.1, we have $\mathrm{PR}(x|y) \leq \mathrm{PR}(y|x)$ that means $X \rightsquigarrow Y$.
2. Since from Definition 4.1, if $X \rightsquigarrow Y \Rightarrow \mathrm{PR}(x|y) \leq \mathrm{PR}(y|x)$ is true, then it must be also true for $\mathrm{PR}(x$ and $z|y) \leq \mathrm{PR}(y|x$ *and* $z) \Rightarrow X \cup Z \rightsquigarrow Y$. The reason is $\mathrm{P}(X) \geq \mathrm{P}(X \cup Z)$, where $X \cup Z$ is union of conditions or constraints of X and Z.
3. Since relation R satisfies $X \rightsquigarrow Y$ and $Y \rightsquigarrow Z$, $\mathrm{PR}(x|y) \leq \mathrm{PR}(y|x)$ and $\mathrm{PR}(y|z) \leq \mathrm{PR}(z|y)$, we have $\mathrm{PR}(x|z) \leq \mathrm{PR}(z|x)$. Thus, R satisfies $X \rightsquigarrow Z$.

Example 4.1 Given a relation $R(X,Y)$, is shown in Table 4.1.

Table 4.1 Relation $R(X,Y)$

Rec	X	Y
1	x_1	y_1
2	x_2	y_2
3	x_3	y_1
4	x_1	y_1
5	x_2	y_2
6	x_4	y_2

From Table 4.1, we try to find comparison between $\mathrm{PR}(x_1|y_1)$ and $\mathrm{PR}(y_1|x_1)$ as follows.

Table 4.2 Relation $R(X = x_1, Y = y_1)$

Rec	$X = x_1$	$Y = y_1$	$X = x_1$ and $Y = y_1$
1	$\mathrm{P}(x_1|x_1) = 1$	$\mathrm{P}(y_1|y_1) = 1$	min(1,1)=1
2	$\mathrm{P}(x_1|x_2) = 0$	$\mathrm{P}(y_1|y_2) = 0$	min(0,0)=0
3	$\mathrm{P}(x_1|x_3) = 0$	$\mathrm{P}(y_1|y_1) = 1$	min(0,1)=0
4	$\mathrm{P}(x_1|x_1) = 1$	$\mathrm{P}(y_1|y_1) = 1$	min(1,1)=1
5	$\mathrm{P}(x_1|x_2) = 0$	$\mathrm{P}(y_1|y_2) = 0$	min(1,0)=0
6	$\mathrm{P}(x_1|x_4) = 0$	$\mathrm{P}(y_1|y_2) = 0$	min(0,0)=0
$\sum$	2	3	2

From Table 4.2,
$\mathrm{PR}(x_1|y_1) = 2/3 < \mathrm{PR}(y_1|x_1) = 1$.
The result leads to conclusion that by knowing $X = x_1$, it certainly give $Y = y_1$, otherwise by knowing $Y = y_1$, the probability to give $X = x_1$ is equal to 2/3. By using the same way, we find that,
$\mathrm{PR}(x_2|y_2) = 2/3 < \mathrm{PR}(y_2|x_2) = 1$,
$\mathrm{PR}(x_3|y_1) = 1/3 < \mathrm{PR}(y_1|x_3) = 1$,
$\mathrm{PR}(x_4|y_2) = 1/3 < \mathrm{PR}(y_2|x_4) = 1$.
Finally, related to (9), we conclude that $X \rightsquigarrow Y$.

Example 4.2 Given Table 4.1, relation between two attributes, X and Y and suppose that we want to know relation between a given partial area of X and a given partial area of Y which are represented by two fuzzy sets f and g, respectively. Let suppose that the membership function of f is,

$$\mu(f) = \{1/x_1, 0.5/x_2, 0.1/x_3\},$$

and membership function of g is,

$$\mu(g) = \{1/y_1, 0.2/y_2\}.$$

By using both membership functions above, Table 4.1 is transformed into Table 4.3 as follows.

Table 4.3 Relation $R(X=f, Y=g)$

Rec	$X=f$	$Y=g$	$X=f$ and $Y=g$
1	$P(f\|x_1)=1$	$P(g\|y_1)=1$	1
2	$P(f\|x_2)=0.5$	$P(g\|y_2)=0.2$	0.2
3	$P(f\|x_3)=0.1$	$P(g\|y_1)=1$	0.1
4	$P(f\|x_1)=1$	$P(g\|y_1)=1$	1
5	$P(f\|x_2)=0.5$	$P(g\|y_2)=0.2$	0.2
6	$P(f\|x_4)=0$	$P(g\|y_2)=0.2$	0
$\sum$	3.1	3.6	2.5

From Table 4.3, it is shown that

$$\mathrm{PR}(f|g)=0.694<\mathrm{PR}(g|f)=0.806.$$

We conclude that it is more realistic to say that $f \rightsquigarrow g$ than $g \rightsquigarrow f$.

5 Conclusion

In this paper, we introduced a proposal of fuzzy functional dependency(FFD), where relation between two attribute domains is based on the concept of conditional probability. From the technical point of view, the FFD is different from most FFD, which generally start with the definition of classical functional dependency and weaken the equality relation into a (gradual) resemblance relation (and then choose an appropriate implication) [4]. We also proved that inference rules, which are similar to Armstrong's Axioms [1] for the FFD, are both sound and complete.

References

[1] Armstrong, W.W.,'Dependency Structures of Database Relationship',*Information Processing*, (1974), pp. 580-583.

[2] Baldwin J.F., Martin T.P., 'A Fuzzy Data Browser in Fril', *Fuzzy Logic*, John Wiley & Sons Ltd (1996), pp. 101-123.

[3] Bernstein, P.A., Swenson, J.R., & Tsichritzis. D.C., 'A Unified Approach to Functional Dependencies and Relation', *Proceedings of the ACM SIGMOD Conference*, San Jose, May 14-16, (1975), pp. 237-245.

[4] Bosc, P., Dubois, D., & Prade, H., 'Fuzzy Functional Dependencies and Redundancy Elimination', *Journal of The American Society for Information Science 49(3)*, (1998), pp. 217-235.

[5] Chen, G.Q., 'Fuzzy Functional Dependencies and a Series of Design Issues of Fuzzy Relational Database'. *Fuzziness in Database Management Systems*, Heidelberg: Physical Verlag (1995), pp. 166-185.

[6] Cubero, J.C., Medina, J.M., & Vila, M.A.,'Influence of Granularity Level in Fuzzy Functional Dependencies', *Symbolic and Quantitative Approaches to Reasoning and Uncertainty (Proceedings of ECSQARU'93)*, Berlin: Springer-Verlag Vol. 747 (1993), pp. 73-78.

[7] Dubois, D., Prade, H., 'A Unifying View of Comparison Indices in a Fuzzy Set-Theoretic Framework', *Fuzzy Sets and Possibility Theory-Recent Developments*, Pergamon Press, (1982), pp. 1-13.

[8] Kiss, A. 'λ-Decomposition of Fuzzy Relational Database', *Annales Univ. Sci. Budapest., Sect. Comp. 12*,(1991), pp. 133-142.

[9] Kosko, B., 'Fuzziness vs. Probability', *Int. J. of General Systems*, Vol. 17, (1990), pp. 211-240.

[10] Liao, S.Y., Wang, H.Q., Liu, W.Y., 'Functional Dependencies with Null Values, and Crisp Values', *IEEE Transactions on Fuzzy Systems*, Vol. 7, No. 1, (February 1999), pp. 97-103

[11] Liu, W., 'The Fuzzy Functional Dependency in The Basis of the Semantic Distance', *Fuzzy Sets and Systems, 59*, (1993), pp. 173-179.

[12] Raju, K.V.S.V.N., & Majumdar, A.K., 'Fuzzy Functional Dependencies and Lossless Join Decomposition of Fuzzy Relational Database Systems', *ACM Transactions on Database Systems, 13(2)*, (1988), pp. 129-166.

[13] Tripathy, R.C., Saxena, P.C., 'Multi-valued Dependencies in Fuzzy Relational Database', *Fuzzy Sets Systems*, Vol 38, no. 3, (1990), pp. 267-280.