DOI 10.1007/s10958-016-2929-4
Journal of Mathematical Sciences, Vol. 216, No. 5, August, 2016

FUZZY INFERENCE AS A GENERALIZATION OF THE BAYESIAN INFERENCE

M. V. Koroteev, P. V. Terelyanskii, and V. A. Ivanyuk UDC 51-77

Abstract. In this paper, we describe an approach for extending the Bayesian inference for a more general case. For this purpose, we change the algorithm of reducing factors. It is also necessary for the new algorithm to stay completely compatible with the classical Bayesian inference.

1. An alternative definition of Bayesian inference. Consider the modified algorithm of Bayesian inference using indicator functions of equality. In classical logic, two values can be either equal each to other or not. However, there exist multi-valued logics (for example, fuzzy logic) in which the law of excluded middle is not always obeyed.

We define the indicator function of equality

$$I(a,b) = \begin{cases} 0, & \text{if } a \neq b \\ 1, & \text{if } a = b \end{cases} \tag{1}$$

Further, we redefine the operation of reducing factors as follows: instead of deleting cells that are incompatible with the value of reduced variable, we multiply the corresponding values of the factor by the indicator function of equality of the variable in the given row of the table to the value of reduced variable and then marginalize this variable from the factor. In the tables below (see (2)), we show the reduction of the variable $B = 1$ from the factor (left) with the subsequent marginalization of the variable B from the factor obtained (right).

A	B	C	H	$I(B=1)$	$P'(A,B,C)$
1	1	1	0.25	1	0.25
1	1	2	0.35	1	0.35
1	2	1	0.08	0	0
1	2	2	0.16	0	0
2	1	1	0.05	1	0.05
2	1	2	0.07	1	0.07
2	2	1	0	0	0
2	2	2	0	0	0
3	1	1	0.15	1	0.15
3	1	2	0.21	1	0.21
3	2	1	0.09	0	0
3	2	2	0.18	0	0

A	B	C	K
1	1	1	0.25
1	1	2	0.35
2	1	1	0.05
2	1	2	0.07
3	1	1	0.15
3	1	2	0.21

(2)

Obviously, this result is completely identical to the factor that can be obtained by the usual reduction of the variable $B = 1$. Prove this assertion.

Theorem 1. *The traditional and modified algorithms of reduction of variables always lead to the same results for identical factors and identical reduced variables.*

Translated from Sovremennaya Matematika i Ee Prilozheniya (Contemporary Mathematics and Its Applications), Vol. 95, Models of Mathematical Economics, 2015.

Proof. Assume that there exists a factor $H(X)$, where X is a set of variables. The cardinality of this factor is $|H| = \prod\limits_{x \in X} |x|$. After reduction of the variable $x_i = v$, $x_i \in X$, by the traditional method we obtain the factor $K(X/x_i)$ whose cardinality is

$$|K| = \prod_{x \in X/x_i} |x| = \frac{\prod\limits_{x \in X} |x|}{|x_i|} = \frac{|H|}{|x_i|}.$$

As a result of reduction of the variable $x_i = v$, the factor L has been obtained by the modified method. The alternative reduction consists of two steps: multiplication of a factor by the value of indicator function and marginalization. In the first step, the domain of the factor does not change, whereas in the second step the variable x_i is excluded from the domain. Thus, the domains of the factors K and L and hence their cardinalities and appointment sets coincide. We prove that the values of the factors K and L coincide for all appointments of these factors. Introduce the preimage $\mathrm{im}(a)$ of the appointment a as the set of appointments of the initial factor H corresponding to the given appointment a of the factor K or L. Since the sets of appointments of these factors coincide, the preimages of the corresponding appointments also coincide. Based on the descriptions of algorithms, we see that the values of the factors K and L for each appointment depend only on the values of the preimage of this appointment of the factor H. Moreover, based on the completeness of the set of appointments of the factor, we can assert that the preimage of each appointment of the factors K or L contains appointments of all possible values of the variable x_i, and each of them occurs only once. Consider the core core of the preimage of an appointment (it always exists and is unique) in which the variable x_i has the value v. The indicator function of equality for appointments attains the following values in classical logic:

$$I(x, v) = \begin{cases} 0 & \text{if } x \neq \mathrm{core}(\mathrm{im}(a)), \\ 1 & \text{if } x = \mathrm{core}(\mathrm{im}(a)). \end{cases}$$

The value of an appointment a of the factor K for each a is the core of the preimage of the appointment a: $K(a) = H(\mathrm{core}(\mathrm{im}(a)))$. The value of the appointment of the factor L for each a is

$$L(a) = H(\mathrm{core}(\mathrm{im}(a))) \cdot 1 + \sum_{b \notin \mathrm{core}(\mathrm{im}(a))} H(a) \cdot 0 = H(\mathrm{core}(\mathrm{im}(a))),$$

i.e.,

$$\forall a: \; K(a) = L(a),$$

which was required. □

2. Fuzzy logical inference. In mathematical modeling, there exists a problem on the description of variables representing qualitative indicators that cannot be formalized by a discrete set of values, for example, the quality of a product, the efficiency of the institution, the qualification of employees, and many others. At the same time, traditionally levels of these indicators are measured qualitatively, by using expert assessments formulated with using linguistic concepts “low,” “high,” and “very high.” Operating with linguistic concepts represents a certain complexity, overcoming of which requires the involvement of a certain mathematical techniques.

In our studies, we use fuzzy logic, which provides flexibility in calculation in linguistic terms and the possibility of handling uncertainty in terms of lack of information. Linguistic variables allow one to formalize inaccurate, indefinite, and multi-valued notions.

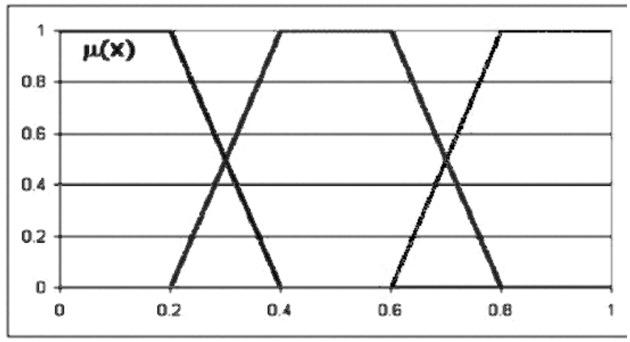

Fig. 1. A simple fuzzy classificator

This property is very useful for expert systems since it provides a methodology that enables the experts to express their knowledge in familiar linguistic form and operate by them as by rigorous mathematical objects. Further, we adapt the algorithm of fuzzy inference to the application in Bayesian nets.

The central notion of fuzzy inference is a *linguistic variable*, i.e., a variable that possesses a certain set of linguistic values (terms) and is defined on a certain domain (usually, a real interval). For example, consider the linguistic variable "quality." We can define a certain integral indicator of quality that allows one to assess quality in some scale. By normalization we can transform this scale to the segment $[0; 1]$. In the sequel, we use this segment as an illustration of the support of a variable.

Each level of quality can be characterized as low, medium, or high, but to a different extent. This set is a set of values of a linguistic variable. Thus, to each value of the linguistic variable, the *membership function* $\mu(x)$ corresponds, where x is an element of the domain. The membership function shows to what grade this value is applicable at a given point of the domain. The membership function attains its values in the segment $[0, 1]$, where 0 corresponds to the situation in which the given value is absolutely inapplicable at this point and 1 corresponds to the absolute applicability of the given value. The set of these functions is called the *fuzzy classificator*. In the case of a usual crisp variable, each point of the domain can belong only to one value. In fuzzy logic, each point belongs to all values, but to different grades.

Note that each term of a linguistic variable is a fuzzy subset (or a fuzzy number, under a certain conditions). This means that all operations defined on fuzzy subsets (FS) can be used for it. In the sequel, we consider the operations of multiplication of a fuzzy subset by a number, addition of fuzzy subsets, and defuzzification of a fuzzy subset. A more detailed description of operators, norms, and measures of fuzzy logic goes far beyond the scope of this paper.

A fuzzy classificator with three terms ("low level," "medium level," and "high level") is shown in Fig. 1. The support of this linguistic variable is the segment $[0, 1]$ (the horizontal axis). The range of the membership function is also the segment $[0, 1]$ (the vertical axis). We see, for example, that the point 0.3 belongs to the term "low level" with "grade of membership" 0.5, to the term "medium level" with "grade of membership" also 0.5, and to the term "high level" with "grade of membership" 0. Informally, we can say that this point does not belong to the term "high level."

Below, as fuzzy classificators we will use so-called *fuzzy partitions*, i.e., classificators that satisfy the following conditions:

1. For each point of the domain, the sum of its grade of membership to all terms is equal to 1;
2. For each point of the domain, there exist at least one, but no more than two terms for which the grade of membership is positive;

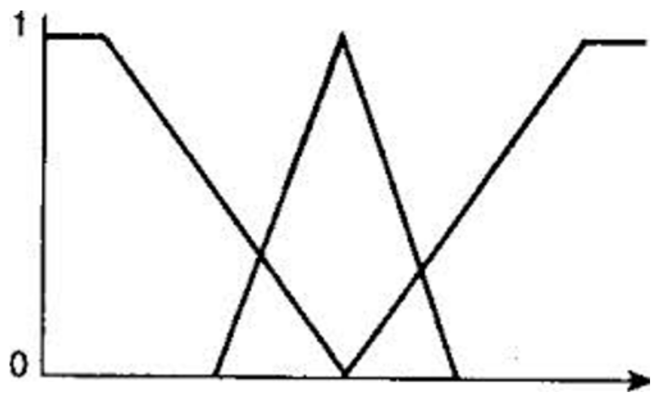

Fig. 2. A fuzzy classificator, which is not a fuzzy partition

3. For each term of a linguistic variable, there exists at least one point for which the grade of membership is equal to 1.

For example, the fuzzy classificator shown in Fig. 2 is not a fuzzy partition.

Describe the algorithm of fuzzy inference using the example of the Mamdani algorithm. Let A and B be two linguistic variables defined on the segment $[0, 1]$ whose possible values are "low," "middle," and "high." The values of B fuzzily depend on the values of A by the following rules of logical inference (they are similar to the rules of ordinary crisp logical inference):

A	B
low	high
middle	middle
high	low

(3)

Assume that we observe the value of A equal to 0.3. Using the fuzzy classificator of the variable A, we find grades of membership of a given point to each term:

A	$\mu_T(x)$
low	0.7
middle	0.3
high	0

(4)

Based on these data, we assign a weight to each inference rule, which shows in what grade this rule can be applied for a given observation:

A	B	α
low	high	0.7
middle	middle	0.3
high	low	0

(5)

In this simple example we use the values of the membership functions as weights of rules. Based on the results obtained, we see that the variable B takes the value $|0.7\cdot$ "high" $+0.3\cdot$ "middle"$|$. Considering each terms as a fuzzy subset, we can calculate the value of a given expression and guarantee that it will be an element of the domain of the linguistic variable B. Except for a numerical value, as a result of inference we can take the fuzzy-set representation in the form of a fuzzy subset $C = 0.7 \cdot$ "high" $+ 0.3 \cdot$ "middle". In the general case, for calculation of the result of fuzzy logical inference, it suffices to calculate weights of all terms of the target variable.

Consider another example of fuzzy inference. Let A, B, and C be three variables and the value of C depends on the values of A and B by the following rules:

IF		**THEN**
A	B	C
low	low	low
low	middle	low
low	high	middle
middle	low	low
middle	middle	middle
middle	high	high
high	low	middle
high	middle	high
high	high	high

(6)

This table shows that the system of rules of fuzzy inference is based on a similar mechanism, where all possible appointments of conditional variables are listed in the IF-part and the corresponding appointment of the sub-conditional variable is indicated in the THEN-part.

Assume that we have observations of conditional variables $A = x$ and $B = y$. The values of membership function are listed in the following tables:

A	$\mu_T(x)$
low	0.7
middle	0.3
high	0

B	$\mu_T(x)$
low	0
middle	0.4
high	0.6

(7)

We calculate the values of rule weights as the products of the corresponding membership values. In the calculation of rule weights in fuzzy inference, one can use various triangular norms as combination operators, but we use the simplest function:

A	B	C	α
low	low	low	$0.7 \cdot 0 = 0$
low	middle	low	$0.7 \cdot 0.4 = 0.28$
low	high	middle	$0.7 \cdot 0.6 = 0.42$
middle	low	low	$0.3 \cdot 0 = 0$
middle	middle	middle	$0.3 \cdot 0.4 = 0.12$
middle	high	high	$0.3 \cdot 0.6 = 0.18$
high	low	middle	$0 \cdot 0 = 0$
high	middle	high	$0 \cdot 0.4 = 0$
high	high	high	$0 \cdot 0.6 = 0$

(8)

We calculate weights of terms of the target variable:

C	α
low	0.28
middle	$0.42 + 0.12 = 0.54$
high	0.18

(9)

One of advantages of fuzzy logical inference is its flexibility. There is a huge number of forms of membership functions, classificator, fuzzy arithmetics, triangular norms and conorms that can be used in calculation of results of inference. In practical problems, fuzzy logical inference has demonstrated its main advantage—orientation to human, the ability of operating with linguistic concepts and, as a consequence, laconicism of rules of inference and brevity of the calculation process compared to Bayesian inference on continuous variables.

Fuzzy logical inference allows one to perform precise inference on continuous variable without using multidimensional analysisand calculation of functions of several variables. This can be useful for modeling of numerical parameters and performing of precise calculation for a foreseeable computing time.

3. Fuzzy inference as an extension of the Bayesian algorithm. We apply the same approach as in the case of deterministic logical inference to fuzzy logical inference. As an example, we consider the system of inference rules from Table (2). Transform this system to the factor of two linguistic variables:

A	B	$P(B\|A)$
low	low	0
low	middle	0
low	high	1
middle	low	0
middle	middle	1
middle	high	0
high	lo	1
high	middle	0
high	high	0

(10)

For the same value of x as in the previous example, we calculate the weight of each rule. In this case, the weight plays the role of indicator functions of equality in the modified algorithm of reducing factors:

A	B	$P(B\|A)$	$\alpha = I(A = x)$	$P'(A, B)$	
low	low	0	0.7	$0.7 \cdot 0 =$	0
low	middle	0	0.7	$0.7 \cdot 0 =$	0
low	high	1	0.7	$0.7 \cdot 1 =$	0.7
middle	low	0	0.3	$0.3 \cdot 0 =$	0
middle	middle	1	0.3	$0.3 \cdot 1 =$	0.3
middle	high	0	0.3	$0.3 \cdot 0 =$	0
high	low	1	0	$0 \cdot 1 =$	0
high	middle	0	0	$0 \cdot 0 =$	0
high	high	0	0	$0 \cdot 0 =$	0

(11)

Further, as in the modified algorithm of reducing, we marginalize the variable A from the factor:

	$P(B\|A = x)$	
low	0+0+0	0
middle	0+0.3+0	0.3
high	0.7+0+0	0.7

(12)

We see that the result completely coincides with the result of inference obtained by the classical Mamdani algorithm. The proof of the identity of these two algorithm is similar to the proof for the method of alternative reducing presented above.

Acknowledgment. This work was partially supported by the Russian Humanitarian Scientific Foundation (Project No. 14-12-34010).

REFERENCES

1. C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer-Verlag (2006).
2. D. Koller and N. Friedman, *Probabilistic Graphical Models*, MIT Press, Massachusetts (2009).
3. M. V. Koroteev, "Forms of membership function of linguistic variables for economical indicators," *Audit Fin. Anal.*, **2**, 239–244 (2012).
4. M. V. Koroteev, "Arithmetics of fuzzy numbers in general form," *Izv. Volgogr. Tekh. Univ., Ser. Probl. Upravl. Vychisl. Tekh.*, **13**, No. 4 (91), 122–127 (2012).
5. M. V. Koroteev, "Analytical defuzzification of fuzzy numbers," *Izv. Volgogr. Tekh. Univ., Ser. Probl. Upravl. Vychisl. Tekh.*, **14**, No. 10 (97), 32–35 (2012).

M. V. Koroteev

Volgograd State Technical University, Volgograd, Russia

P. V. Terelyanskii

Volgograd State Technical University, Volgograd, Russia

V. A. Ivanyuk

Financial University under the Government of the Russian Federation, Moscow, Russia

E-mail: `ivaver6@gmail.com`