Information Sciences 340–341 (2016) 250–261

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

An improved method to construct basic probability assignment based on the confusion matrix for classification problem

Xinyang Deng[a,b], Qi Liu[b,c], Yong Deng[a,d,*], Sankaran Mahadevan[d]

[a] School of Computer and Information Science, Southwest University, Chongqing 400715, China
[b] Center for Quantitative Sciences, Vanderbilt University School of Medicine, Nashville, TN 37232, USA
[c] Department of Biomedical Informatics, Vanderbilt University School of Medicine,Nashville, TN 37232, USA
[d] School of Engineering, Vanderbilt University, Nashville, TN 37235, USA

ARTICLE INFO

Article history:
Received 26 May 2015
Revised 15 September 2015
Accepted 11 January 2016
Available online 15 January 2016

Keywords:
Dempster–Shafer evidence theory
Belief function
Basic probability assignment
Confusion matrix
Classification
Transmembrane proteins

ABSTRACT

The determination of basic probability assignment (BPA) is a crucial issue in the application of Dempster–Shafer evidence theory. Classification is a process of determining the class label that a sample belongs to. In classification problem, the construction of BPA based on the confusion matrix has been studied. However, the existing methods do not make full use of the available information provided by the confusion matrix. In this paper, an improved method to construct the BPA is proposed based on the confusion matrix. The proposed method takes into account both the precision rate and the recall rate of each class. An illustrative case regarding the prediction of transmembrane protein topology is given to demonstrate the effectiveness of the proposed method.

1. Introduction

Dempster–Shafer evidence theory [9,45], also called Dempster–Shafer theory, has been widely applied in many fields, such as information fusion [50], classification [5,37,38], and others [4,10–13,16,20,25,42,49,57,68]. Tabassian et al. [51,52] used Dempster–Shafer theory to handle data with imperfect labels in ensemble learning, and addressed the situation that the class memberships of the training data are subject to ambiguity. Deng [19] proposed a generalized evidence theory (GET) to address conflict management in an open world environment. Thanks to its flexibility, Dempster–Shafer theory has been combined with other theories like fuzzy set theory [8,31,69] and genetic algorithm [22], and many useful tools have been developed to handle various types of uncertainty, which further extends the application of the Dempster–Shafer theory. For instance, in [32], Kang et al. proposed an uncertain-graph structure, called evidential cognitive map (ECM), to represent causal reasoning by combining the cognitive maps and Dempster–Shafer theory. Recently, in the fields of evolutionary game theory [6,7,15,58–64] and game theory [53], Dempster–Shafer evidence theory has also attracted some interests [17,18,35].

The determination of basic probability assignment (BPA) is one of the most important problems in evidential systems. The construction of BPA based on the confusion matrix is a practical and effective method [1,2,24,40,43,54,65]. In a previous related study, Xu et al. [65] presented an elegant method for the construction of BPA based on recognition rate, substitution

* Corresponding author at: School of Computer and Information Science, Southwest University, Chongqing 400715, China. Tel.: +86 23 68254555.
E-mail address: prof.deng@hotmail.com (Y. Deng).

http://dx.doi.org/10.1016/j.ins.2016.01.033
0020-0255/

rate, and rejection rate of the confusion matrix. However, Xu et al.'s method does not consider the difference of the classifier's recognition ability for different classes. To overcome the shortcoming, Parikh et al. [40] proposed a modified method, which is more effective and has been successfully used in condition monitoring. The improvement proposed by Parhikh et al. is on the basis of the prior knowledge provided by the confusion matrix. Specifically, that method utilized the precision rate of each actual class according to the confusion matrix. However, the prior knowledge contained in the confusion matrix is not only the precision rate, but also the recall rate of each class which is another important aspect to reflect the classifier's recognition ability for each class.

Based on this idea, an improved BPA construction method is proposed based on the confusion matrix in this paper. Section 2 introduces some basic concepts and related previous work. Section 3 presents the proposed method. Section 4 gives an illustrative case to demonstrate the effectiveness of the proposed method. Section 5 concludes the paper.

2. Preliminaries

2.1. Basic concepts

The Dempster–Shafer evidence theory [9,45], first proposed by Dempster and further developed by Shafer, is widely used to handle uncertain information. In this theory, basic probability assignment (BPA) is used to represent the uncertain information, and Dempster's rule of combination is used to combine multiple BPAs.

In Dempster–Shafer theory, a problem domain denoted by a finite nonempty set Ω of mutually exclusive and exhaustive hypotheses is called the frame of discernment. Let 2^{Ω} denote the power set of Ω. A BPA is a mapping $m : 2^{\Omega} \to [0, 1]$, satisfying

$$m(\emptyset) = 0 \quad and \quad \sum_{A \in 2^{\Omega}} m(A) = 1 \tag{1}$$

Dempster's rule of combination, also called orthogonal sum, is defined as follows

$$m(A) = \begin{cases} \dfrac{1}{1-K} \displaystyle\sum_{B \cap C = A} m_1(B) m_2(C), & A \neq \emptyset; \\ 0, & A = \emptyset. \end{cases} \tag{2}$$

$$K = \sum_{B \cap C = \emptyset} m_1(B) m_2(C) \tag{3}$$

where K is a normalization constant, called conflict coefficient of two BPAs. The Dempster's rule satisfies commutative and associative properties. Thus if there exist multiple BPAs, the combination of them can be carried out in a pairwise way with any order. However, this rule may give counterintuitive results when there is conflict among BPAs. For this issue, many other combination rules have been proposed, for example Yager's conflict redistribution rule [66], Dubois and Prade's disjunctive rule [26], Murphy's simple averaging rule [39], Deng's weighted averaging rule [23], proportional conflict redistribution PCR5 and PCR6 rules [29,46,47].

2.2. Confusion matrix

Confusion matrix [44] is a concept from machine learning, which contains information about actual and predicted classifications done by a classification system. A confusion matrix has two-dimensions, one dimension is indexed by the actual class of an object, the other is indexed by the class that the classifier predicts. Fig. 1 presents the basic form of confusion matrix for a multi-class classification task, with the classes A_1, A_2, and A_n. In the confusion matrix, N_{ij} represents the number of samples actually belonging to class A_i but classified as class A_j.

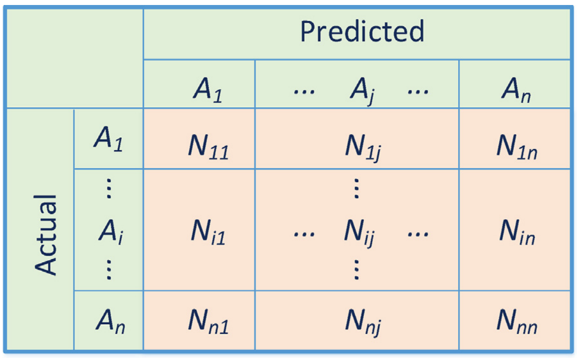

Fig. 1. Confusion matrix.