Pattern Recognition 44 (2011) 1023–1030

Contents lists available at ScienceDirect

Pattern Recognition

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

Fuzzy posterior-probabilistic fusion

Tuan D. Pham

School of Engineering and Information Technology, University of New South Wales, Canberra, ACT 2600, Australia

ARTICLE INFO

Article history:
Received 23 November 2009
Received in revised form
1 August 2010
Accepted 18 November 2010

Keywords:
Information fusion
Permanence of ratios
Probabilistic measures
Fuzzy measures
Fuzzy integrals

ABSTRACT

The paradigm of the permanence of updating ratios, which is a well-proven concept in experimental engineering approximation, has recently been utilized to construct a probabilistic fusion approach for combining knowledge from multiple sources. This ratio-based probabilistic fusion, however, assumes the equal contribution of attributes of diverse evidences. This paper introduces a new framework of a fuzzy probabilistic data fusion using the principles of the permanence of ratios paradigm, and the theories of fuzzy measures and fuzzy integrals. The fuzzy sub-fusion of the proposed approach allows an effective model for incorporating evidence importance and interaction.

1. Introduction

Combining information from diverse sources is a challenging and important area of research in pattern classification. There are a number of different approaches for information fusion and the literature is vast due to many developments and applications in this broad field, which pervades many scientific disciplines since the early work over many years [1–6]. Although information fusion has been reported to be useful in many practical applications of various science and engineering disciplines [7–18], it is still a challenging task which continues to be explored for effective handling of various kinds of mathematical models for measuring event uncertainty [19]. Particularly on application of fuzzy-set theory to information fusion, recent developments and applications can be found in many areas such as pattern classification, image analysis, decision making, man-made structures, medicine, biology and humanity [20–29].

Mathematical properties of various popular information fusion operators were investigated by Bloch [9]. These operators, including naive Bayesian fusion (product rule), fuzzy fusion (T-norm, T-conorm), certainty-factor fusion (MYCIN) and evidence-based fusion (Dempster–Shafer theory of evidence); can be classified into three behaviors: conjunctive (if the sources of information have low conflicting evidences), disjunctive (if the sources of information have high conflicting evidences), or compromising (if the sources of information have partial conflicting evidences). Fuzzy fusion operators can be suitable for dealing with all the above three behaviors in both context dependence and context independence, while the naive Bayesian fusion and Dempster–Shafer fusion operators were suggested to be only suitable for the conjunctive behavior in the case of context dependence. Although the naive Bayesian fusion is relatively easier to implement and often performs well in many applications, this approach has its own disadvantage on the assumption of the independence of classifiers when conflicting evidence arises. Therefore, good knowledge about the mathematical properties of different information fusion operators allows one to select an appropriate operator for effective application to a particular decision fusion problem.

Altincay [31] investigated the application of the naive Bayesian fusion (NBF) of dependent and independent classifiers and found that the combination of dependent classifiers performed better than that of independent classifiers when both using the same probabilistic approach. In the same study, the author found that the NBF may not give a better result than the best individual classifier when the individual accuracies of individual classifiers are not comparable. The finding may suggest the importance of the dependency of different classifiers in the Bayesian fusion approach. Hence, a probabilistic fusion which can address the property of classifier dependence is expected to perform better than the NBF. In the study of feature subset selection using filter and wrapper methods, Kohavi and John [32] used the naive Bayesian classifier to compute the probability of each class, assuming the features are conditionally independent. The authors found that the naive Bayesian algorithm could hardly improve the results using feature selection as this low performance was partly due to the fact that the naive Bayes was adversely affected by the conditional dependence between the features.

From the standpoint of statistics, information fusion is the study of integrating prior and preposterior probabilities into a posterior

E-mail address: t.pham@adfa.edu.au

0031-3203/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2010.11.014

probability [30]. To alleviate the mathematical derivation of a data fusion model, the assumption of conditional independence of the data events is often imposed to conveniently calculate the posterior probabilities. This assumption may lead to inaccurate or inconsistent results when evidences about the event under study are obtained can relate to each other, and have been widely realized in data analysis in geoscience [19,33]. For example, there is no rigorous support for assuming data independence or conditional independence in various features of background and objects in images; because different features relate to either background or objects events share the same image property. As another example, a particular spatial feature extracted from one region often relates to that from another nearby location: this is due to the nature of continuity of spatial data. This realization gives rise to the challenge of the calculation of posterior probabilities without the hypothesis of any form of data independence.

It has been known that the assumption of data independence is usually employed in statistical theories, and the independence assumption is prevailed by introducing linear dependence between the data values. A simple and practical probabilistic solution for event combination is by means of weighted linear averaging of prior probabilities so that the posterior conditional probability can be conveniently calculated. For example, the calculation of probability $P(A|B,C)$ conditioned to the two data events B and C can be obtained as [19]

$$P(A|B,C) = w_1 P(A|B) + w_2 P(A|C) \tag{1}$$

where w_1 and w_2 are the positive weights associating with $P(A|B)$ and $P(A|C)$, respectively; and $w_1 + w_2 = 1$.

The above linear combination imposes convexity of the result. Such convexity is undesirable because the combined probability is bounded by the two prior probabilities $P(A|B)$ and $P(A|C)$. Thus, it prevents the possibility of data integration. Recently, Journal [33] introduced a probabilistic method, which is based on the engineering paradigm of permanence of ratios, for combining information from diverse sources where the conventional assumption of data independence is relaxed. This ratio-based probabilistic fusion method satisfies the mathematical properties of all limit conditions in the presence of complex data independence, and only requires the knowledge of the prior probability and the elementary single event-conditioned probabilities, which can be computed independently from each other.

However, the ratio-based fusion being mentioned above still does not consider the different degrees of influence of the diverse sources when trying to combine the information. Such influence factor is inherent in many practical applications and needs to be considered in the design of any information fusion methods. For example, in pattern recognition, different features of an object may give different effects in classification; and individual classifiers, which are based on different approaches, may give different solutions using the same features. It is the motivation of this paper as an attempt to incorporate the interaction between different attributes of multiple information sources into the ratio-based probabilistic fusion. The proposed fusion model is based on the mathematical frameworks of fuzzy measures and fuzzy integrals, and can be viewed as a generalized version of the ratio-based fusion.

To be self-contained in the development of the proposed approach, Section 2 presents a brief overview of the ratio-based fusion. Section 3 discusses the formulation of the ratio-based fuzzy probabilistic fusion model. Section 4 illustrates the performance of the new approach. Finally, conclusion of the finding is addressed in Section 5.

2. Updating information with ratio-based probabilistic fusion

Let $P(A)$ be the prior probability of the occurrence of data event A; $P(A|B)$ and $P(A|C)$ be the probabilities of occurrence of event A given the knowledge of events B and C, respectively; $P(B|A)$ and $P(C|A)$ the probabilities of observing events B and C given A, respectively. Using Bayes' law, the posterior probability of A given B and C is

$$P(A|B,C) = \frac{P(A,B,C)}{P(B,C)} = \frac{P(A)P(B|A)P(C|A,B)}{P(B,C)} \tag{2}$$

The simplest way for computing the two probabilistic models is to assume the model independence, giving

$$P(C|A,B) = P(C|A) \tag{3}$$

and

$$P(B,C) = P(B)P(C) \tag{4}$$

Thus, (2) can be rewritten as

$$P(A|B,C) = \frac{P(A)P(B|A)P(C|A)}{P(B)P(C)} \tag{5}$$

or

$$\frac{P(A|B,C)}{P(A)} = \frac{P(A|B)}{P(A)}\frac{P(A|C)}{P(A)} \tag{6}$$

However, the assumption of conditional independence between the data events usually does not statistically perform well and leads to inconsistencies in many real applications [33]. Therefore, an alternative to the hypothesis of conventional data event independence should be considered. The permanence of ratios based approach allows data events B and C to be incrementally conditionally dependent and its fusion scheme gives

$$P(A|B,C) = \frac{1}{1+x} = \frac{a}{a+bc} \in [0,1] \tag{7}$$

where a, b, c, and x are taken as the logistic-type ratio of marginal probabilities of A, B after A occurring, C after A occurring, and B and C after A occurring; respectively:

$$a = \frac{1-P(A)}{P(A)}, \quad b = \frac{1-P(A|B)}{P(A|B)}, \quad c = \frac{1-P(A|C)}{P(A|C)}, \quad x = \frac{1-P(A|B,C)}{P(A|B,C)}$$

An interpretation of the fusion expressed in (7) is as follows. Let A is the target event which is to be updated by events B and C. The term a is considered as a measure of prior uncertainty about the target event A or a distance to the occurrence of A without any updated evidence. We have $a=0$ for $P(A)=1$ if target event A is certain to occur; and $a=\infty$ for $P(A)=0$ if A is an impossible event. Likewise, b and c are measures of uncertainty or the distances to A knowing about its occurrence after observing evidences given by B and C, respectively. The term x is the distance to the target event A occurring after observing evidences given by both events B and C. The ratio c/a is then the incremental (increasing or decreasing) information of C to that distance starting from the prior distance a. Similarly, the ration x/b is the incremental information of C starting from the distance b. Thus, the permanence of ratios provides the following approximate relation:

$$\frac{x}{b} \approx \frac{c}{a} \tag{8}$$

which says that the incremental information about C to the knowledge of A is the same after or before knowing B. In other words, the incremental contribution of information from C about A is independent of B. Hence, this mathematical expression relaxes the restriction of the assumption of full independence of B and C.

Expression (7) verifies the limit properties of exact decomposition [34] of the joint probability of A and the two events B and C as follows [33].

1. Consistent probability: $P(A|B,C) \in [0,1]$.
2. Closure condition: $P(\overline{A}|B,C) = xP(A|B,C) = x/(1+x) \in [0,1]$, where $P(\overline{A}) = 1-P(A)$.