Neurocomputing 171 (2016) 252–261

Contents lists available at ScienceDirect

Neurocomputing

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

Combining multiple biometric traits with an order-preserving score fusion algorithm

Yicong Liang [a,b,*], Xiaoqing Ding [a], Changsong Liu [a], Jing-Hao Xue [b]

[a] Department of Electronic Engineering, Tsinghua University, Beijing 100084, China
[b] Department of Statistical Science, University College London, London WC1E 6BT, UK

ARTICLE INFO

Article history:
Received 27 August 2014
Received in revised form
25 April 2015
Accepted 21 June 2015
Communicated by Shiguang Shan
Available online 30 June 2015

Keywords:
Score fusion
Multibiometric system
Order-preserving algorithm
Tree-structured ensemble

ABSTRACT

Multibiometric systems based on score fusion can effectively combine the discriminative power of multiple biometric traits and overcome the limitations of individual trait, leading to a better performance of biometric authentication. To tackle multiple adverse issues with the established classifier-based or probability-based algorithms, in this paper we propose a novel order-preserving probabilistic score fusion algorithm, Order-Preserving Tree (OPT), by casting the score fusion problem into an optimisation problem with the natural order-preserving constraint. OPT is an algorithm fully non-parametric and widely applicable, not assuming any parametric forms of probabilities or independence among sources, directly estimating the posterior probabilities from maximum likelihood estimation, and exploiting the power of tree-structured ensembles. We demonstrate the effectiveness of our OPT algorithm by comparing it with many widely used score fusion algorithms on two prevalent multibiometric databases.

1. Introduction

Biometric systems have found an increasingly wide range of applications in both science and industry. However in many of these applications, unibiometric systems, which exercise only a single biometric trait, often cannot fulfil the biometric authentication tasks. This is mainly due to their limited abilities to represent subjects and prevent spoofs. To overcome such limitations, multibiometric systems have been developed. From fusing several different types of complementary biometric traits together, multibiometric systems can benefit substantially in representing and discriminating subjects, as well as in preventing spoofs since it is much more difficult to cheat simultaneously in all the information sources than to deceive a unibiometric system.

In a multibiometric system, there are four stages in which information fusion can be implemented, namely the sensor stage, feature stage, score stage and decision stage, listed from the earliest to the latest. To fuse information in a later stage means the ease of implementation, at a cost of more information loss. To date, fusion in the score stage is generally considered to provide an appropriate trade-off and preferred by many researchers [1–3].

Existing score fusion algorithms can be divided into two main categories: classifier-based algorithms and probability-based algorithms.

Classifier-based algorithms tackle the fusion problem as a pattern classification task. In this framework, source scores of a sample are used as the input features of a classifier to obtain the predicted class. The classifier is trained on the training samples to minimise the training error by using traditional pattern recognition algorithms. Traditional classifier-based approaches include linear classifiers with minimised least squares error [4] or L_1-norm soft margin error [5,6], reduced multivariate polynomial classifier [7], support vector machine [8] and single hidden layer feedforward neural network [9]. Other advanced techniques in the pattern recognition society, such as semi-supervised learning [10,11], ensemble learning [12] and kernel tricks [6], can also be transferred without difficulty to solve the fusion problem. Two recent examples of classifier-based algorithms are FWOT [13] and minCq [14,15]. FWOT optimises an objective function that is a combination of the squared hinge loss and a 2-norm regulariser. The classifier structure of FWOT is a close resemblance of a single-layer neural network. The algorithm minCq, on the other hand, tailors the score fusion problem into a PAC-learning framework and obtains an optimal linear fusion algorithm by minimising an upper bound of the true error risk. Both algorithms hold good ability of generalisation.

* Corresponding author at: Department of Electronic Engineering, Tsinghua University, Beijing 100084, China. Tel.: +86 15010266579.
E-mail address: liangyc04@gmail.com (Y. Liang).

http://dx.doi.org/10.1016/j.neucom.2015.06.039
0925-2312/

Classifier-based algorithms can directly exploit the great progress made in the pattern recognition realm. However, an issue with these algorithms is that they are suboptimal in score fusion, because, different from the usual features found in a pattern recognition task, a source score is by nature that a higher score implies a higher probability of the sample belonging to the genuine class. This intrinsic characteristic, as informative prior knowledge, has been largely neglected by the classifier-based algorithms. Particularly, when the training samples are insufficient, the neglect of this prior knowledge may worsen overfitting and thus the performance of biometric authentication. Another issue with these algorithms is that in the design of a classifier-based fusion algorithm there are often tuning parameters, the optimisation of which may not be a trivial problem and issues such as local optimum or overfitting may exist.

Probability-based algorithms tackle the fusion problem in a probabilistic manner. These algorithms usually consist of two steps. In the first step, they normalise all source scores to make their values between $[0, 1]$, and treat the normalised score as the posterior probability of the genuine class given the corresponding source as evidence. In the second step, they merge multiple posterior probabilities into a single posterior probability given all sources as evidence. To make the merge work, these algorithms usually need extra assumptions, made a priori or with the help of training samples or as a combination of the former two. For example, Kittler et al. [16] show that, with the assumption that all source scores are mutually conditionally independent, the commonly used Product rule can be derived. On the other hand, given the assumption that posterior probabilities of each classifier do not deviate dramatically from the priors, the Sum rule can be induced. Other off-the-shelf fusion rules, such as the Max, Min, Median and Majority Vote rules, can all be derived as the midway rules of the Sum and Product rules. Terrades et al. [17] also assume that source scores are mutually independent. Prabhakar and Jain [18] and Nandakumar et al. [1] estimate the probability density function of the training samples and use this function to assist the determination of the merged posterior probability. Ma et al. [19,20] assume parametric forms of the merged posterior probability, which take the dependency between scores into consideration, and use the training samples to estimate the parameters. As a recent example, Cheema et al. [21] assume that the merged posterior probability is a linear combination of source probabilities. Their method solves a constrained quadratic optimisation problem to decide the optimal weights and can deal with the cases with more than two classes.

There are also several issues of concern to these probability-based algorithms. Firstly, except the density estimation algorithms [18,1], these probability-based algorithms all make some assumptions about the merged posterior probability, which may not be fulfilled in practice and thus may limit the generalisability of them. Secondly, as for [18,1], the density estimation procedures will induce hyper-parameters, such as the number of Gaussian mixtures in [1] and the Parzen window width in [18], and the estimation results may be unreliable when the sample dimension is high. Thirdly, in the score normalisation step, most of these algorithms apply heuristic techniques such as the min–max, z-score and tanh algorithms [22], and it is in question how well the normalised result reflects the true posterior probability given the score.

To tackle these adverse issues, in this paper we propose a novel probability-based score fusion algorithm, termed Order-Preserving Tree (OPT). The advantages of OPT are threefold. Firstly, OPT treats both the score normalisation and the posterior probability merging procedure as an constrained optimisation problem. The only constraint in optimisation, which is also the only assumption that OPT makes, is the intrinsic characteristic of order preserving: For any two samples A and B, if every source score suggests that the A is no less likely than B to belong to the genuine class, then the fusion result should also give the same suggestion. Secondly, OPT does not assume any parametric form of probabilities, making itself enjoy widely applicability. Moreover, being fully non-parametric, OPT has no hyper-parameters that need to be tuned, which makes the training procedure efficient. Thirdly, OPT bypasses the procedure of probability density estimation of samples; it instead directly estimates the posterior probabilities themselves. This not only can avoid the issues with density estimation, but also according to Occam's Razor can be more suitable for a task like score fusion. To avoid the problem of the curse of dimensionality, we adopt a tree-structured ensemble to hierarchically merge multiple source scores.

To demonstrate the effectiveness of our OPT algorithm, we conduct extensive experiments on the NIST-BSSR1 and XM2VTS databases, two public-domain databases specially designed to evaluate score fusion algorithm in biometric authentication. Our algorithm demonstrates superior performance compared with many off-the-shelf, classifier-based and probability-based score fusion techniques.

The remainder of this paper is organised as follows. In Sections 2 and 3 we give the basic framework and implementation details of our OPT algorithm, respectively. The experimental results are summarised in Section 4. The work is concluded in Section5.

2. Algorithmic framework

In this paper, we will focus on the two-class problem for simplicity. This is the typical case for the multibiometric verification system, where the target is to predict whether a pair of samples belong to the same subject, given a set of biometric similarity scores. We start by establishing notation.

2.1. Notation

The genuine class and the imposter class are denoted by ω_+ and ω_-, respectively. Suppose that there are N training samples, denoted by $x_1, \ldots, x_N$, with corresponding class labels $y_1, \ldots, y_N$, where $y_i = 1$ if $x_i \in \omega_+$ and $y_i = 0$ otherwise. Suppose for each sample x there are K source scores and use $S_i(x)$ to denote the ith source score. We assume that a higher score indicates a higher posterior probability (suggested by the score) of belonging to ω_+. We use $P_{i_1 \ldots i_k}(x)$ to denote the posterior probability $Pr(x \in \omega_+ \mid S_{i_1}(x), \ldots, S_{i_k}(x))$ for short.

2.2. Overall structure

The overall structure of our OPT algorithm is illustrated in Fig. 1. It is divided in two stages: a normalisation stage and a merge stage. In the normalisation stage, we transform each source score into a posterior probability suggested by the score, i.e. we calculate P_i for every i given S_i. In the merge stage, we merge the information given by all P_is together and obtain the final posterior probability $P_{1,\ldots,K}$, i.e. we calculate the conditional probability $Pr(x \in \omega_+ \mid P_1(x), \ldots, P_K(x))$.

The support of the conditional probability $Pr(x \in \omega_+ \mid P_1(x), \ldots, P_K(x))$ is K-dimensional. Since we do not give the functional any parametric form, it will encounter the curse of dimensionality to directly estimate the probability. We apply a tree-like hierarchical structure to circumvent this problem, as illustrated in Fig. 1. Using this hierarchical structure, we only need to calculate a two-dimensional conditional probability function in each node.

We propose the methodology of a score normalisation algorithm and a two-dimensional merge algorithm in Sections 2.3 and