Neural Networks 63 (2015) 87–93

Contents lists available at ScienceDirect

Neural Networks

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

Performance improvement of classifier fusion for batch samples based on upper integral

Hui-Min Feng, Xi-Zhao Wang *

Key Lab. of Machine Learning and Computational Intelligence, College of Mathematics and Information Science, Hebei University, Baoding 071002, China

ARTICLE INFO

Article history:
Received 2 March 2014
Received in revised form 12 October 2014
Accepted 14 November 2014
Available online 28 November 2014

Keywords:
Extreme learning machine
Upper integral
Fuzzy measure
Fuzzy integral
Multiple classifier fusion

ABSTRACT

The generalization ability of ELM can be improved by fusing a number of individual ELMs. This paper proposes a new scheme of fusing ELMs based on upper integrals, which differs from all the existing fuzzy integral models of classifier fusion. The new scheme uses the upper integral to reasonably assign tested samples to different ELMs for maximizing the classification efficiency. By solving an optimization problem of upper integrals, we obtain the proportions of assigning samples to different ELMs and their combinations. The definition of upper integral guarantees such a conclusion that the classification accuracy of the fused ELM is not less than that of any individual ELM theoretically. Numerical simulations demonstrate that most existing fusion methodologies such as Bagging and Boosting can be improved by our upper integral model.

1. Introduction

Huang, Zhu, and Siew (2004, 2006) proposed a new learning algorithm for single-hidden layer feedforward networks (SLFNs) called Extreme Learning Machine (ELM) which overcomes the problems caused by gradient descent based algorithms such as Back propagation applied in artificial neural networks. ELM can significantly reduce the amount of time needed to train a neural network and preserve the universal approximation ability (Huang, Chen, & Siew, 2006). It randomly chooses the input weights and hidden node biases, and analytically determines the output weights of SLFN. It has much better generalization performance with much faster learning speed (Huang et al., 2006). It automatically determines all the network parameters analytically, which avoids trivial human intervention and makes it efficient in online and realtime applications (Huang et al., 2006; Lan, Soh, & Huang, 2009). ELM has several advantages such as ease of use, faster learning speed, higher generalization performance, suitable for many nonlinear activation function and kernel functions (Liu, He, & Shi, 2008; Wang, Chen, & Feng, 2011).

To achieve good generalization performance, ELM minimizes training error on the entire training data set, therefore it might suffer from overfitting as the learning model will approximate all training samples well (Liu & Wang, 2010). Hansen and Salamon (1990) have showed that the generalization ability of a neural network system can be significantly improved through ensembling a number of neural networks. Combining multiple classifiers to solve a given classification problem is an efficient approach to improve the performance of classification and avoid overfitting (Jain, Duin, & Mao, 2000).

When outputs of a base classifier are real-valued vectors (most often posterior probabilities or possibilities (Kuncheva, 2003), sometimes evidences), a fusion operator such as maximum/minimum, median, average, weighted average, ordered weighted average, Dempster–Shafer approach or fuzzy integral, can be selected to aggregate the outputs from all individual base classifiers (Kuncheva, 2003; Schmitt, Bombardier, & Wendling, 2008; Zhai, Xu, & Li, 2013; Zhai, Xu, & Wang, 2012). The fusion based on maximum/minimum, median or average is suitable for the case that in a combination the importance of base classifier is identical (Kuncheva, 2003; Verikas, Lipnickas, Malmqvist, Bacauskiene, & Gelzinis, 1999). If the importance of a base classifier is different from another, weighted average and ordered weighted average can be chosen (Kuncheva, 2003; Yager, 1988). The importance of a single classifier is emphasized in weighted average while the magnitude of output from a base classifier is particularly considered in ordered weighted average (Kuncheva, 2003; Yager, 1988). But the two methods are under an assumption that interaction does not exist among the individual classifiers. However, this assumption may not be true in many real problems. If the interaction is involved, the fuzzy integral (Schmitt et al., 2008; Wang et al., 2011) or Dempster–Shafer approach (Shafer, 1976) is considered as one of the most appropriate choices. Fuzzy integrals are more

* Corresponding author.
E-mail address: xizhaowang@ieee.org (X.-Z. Wang).

http://dx.doi.org/10.1016/j.neunet.2014.11.004

computationally efficient than a strict Dempster–Shafer approach (Keller, Gader, Tahani, Chiang, & Mohamed, 1994). The fuzzy integral as a fusion tool, in which the non-additive measure can clearly express the interaction among classifiers and the importance of each individual classifier, has its particular advantages. Additionally the average, weighted average and ordered weighted average can be regarded as special cases of fuzzy integrals. For a tested sample, each base classifier outputs a vector in which the ith component is the degree of the sample belonging the ith class. The fuzzy integral integrates these degrees with respect to a fuzzy measure for each class. One difficulty of applying fuzzy integrals in classifier fusion is how to determine the fuzzy measures. The training process of fuzzy integral fusion method contains training base classifiers and learning the fuzzy measure from training samples. From references one can find a number of methods to determine fuzzy measures such as linear programming, quadratic programming (Yeung, Wang, & Tsang, 2004), genetic algorithm (Yang, Wang, Heng, & Leung, 2008), neural network (Wang & Wang, 1997), and pseudo-gradient (Wang, Leung, & Klir, 2005).

This paper proposes a new approach to multiple classifier fusion based on the upper integral which is a type of fuzzy integrals proposed by Wang, Li, and Leung (2008). Motivated by the definition of upper integrals which can be considered as a mechanism of maximizing potential efficiency of classifier combination, the new approach is devoted to improve the classification performance of a fusion operator based on upper integrals. It is worth noting that, in our approach, the upper integral itself is not considered as a tool of classifier-fusion but it is considered as a tool to improve any existing classifier-fusion operator. In other words, our approach (in which the upper integral is no longer a fusion operator) differs from all existing fuzzy integral based fusion schemes (which consider the fuzzy integrals as fusion operators). Specifically, given a group of individual classifiers trained from a set of samples and a fusion operator, we regard the classification accuracies of individual classifiers and their combinations as the efficiency measure, which avoids almost the difficulty of determining fuzzy measures. The upper integral plays a role of assigning suitable proportion of samples to different individual classifiers and their combinations to obtain maximum the classification efficiency. It computes how many samples will be allocated to some of individual classifiers and their combinations by solving an optimization problem derived from the upper integral. This implies a proportion of sample-allocation for a given set of samples. Based on this proportion, some oracles are used to determine which samples will be allocated to those individual classifiers and their combinations. Given a sample, the oracle of a combination of classifiers first predicts the possibility with which the combination can correctly classify the sample. Then the sample is allocated to the combination with maximum possibility. When the number of samples allocated to a combination attains the proportion, the allocation to this combination stops, and the allocations to other combinations continue until all samples are allocated. After the allocation, those classifiers perform the classification of the set of samples, which is our final classification result.

The rest of this paper is arranged as follows. In Section 2, the existing multiple classifier fusion schemes are reviewed. Section 3 is devoted to the efficiency measures, fuzzy integrals and upper integrals. Our proposed new fusion scheme based on the upper integral is given in Section 4. Section 5 presents a number of numerical experiments to verify advantages of the new approach, and finally Section 6 concludes this paper.

2. Multiple classifier fusion based on fuzzy integrals

Suppose that $X = \{x_1, x_2, \ldots, x_n\}$ is a set of classifiers. The output of classifier x_i is a c-dimensional nonnegative vector $[d_{j,1}, d_{j,2}, \ldots, d_{j,c}]$ where c is the number of classes. Without loss of generality, let $d_{j,i} \in [0, 1]$ denote the support from classifier x_j to the hypothesis that the sample submitted for classification comes from the ith class C_i for $j = 1, 2, \ldots, n$, $i = 1, 2, \ldots, c$. The larger the support, the more likely the class label C_i. All outputs of classifiers for a particular sample can be organized in a matrix

$$DP = \begin{bmatrix} d_{1,1} & d_{1,2} & \cdots & d_{1,i} & \cdots & d_{1,c} \\ d_{2,1} & d_{2,2} & \cdots & d_{2,i} & \cdots & d_{2,c} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ d_{n,1} & d_{n,2} & \cdots & d_{n,i} & \cdots & d_{n,c} \end{bmatrix}.$$

Each column of DP matrix can be regarded as a function defined on the classifier set X, $f_i : X \to [0, 1]$, $f_i(x_j) = d_{j,i}$, $i = 1, 2, \ldots, c$, $j = 1, 2, \ldots, n$. For each class C_i, we need to determine a nonnegative set function μ_i on the power set $P(X)$ of X. μ_i can represent not only the importance of individual classifiers but also the interaction among classifiers towards samples from C_i class. Set functions have some special cases.

Definition 1 (*Wang et al., 2008*)**.** Let X be a nonempty and finite set and $P(X)$ be the power set of X, i.e., the group of all subsets of X. Then $(X; P(X))$ is a measurable space. A set function $\mu : P(X) \to (-\infty, +\infty)$ is called a fuzzy measure or a monotone measure, if

(F1) $\mu(\emptyset) = 0$, (vanishing at the empty set)
(F2) $\mu(A) \geq 0$, for any $A \subset X$, (non-negativity)
(F3) $\mu(A) \leq \mu(B)$, if $A \subset B, A \subset X, B \subset X$, (monotonicity).

Set function μ is called an efficiency measure if it satisfies (F1) and (F2); μ is called a signed efficiency measure if it satisfies (F1) only. Any fuzzy measure is a special case of the efficiency measure; and any efficiency measure is a nonnegative set function. Fuzzy measures have a monotone constraint but efficiency measures have not, so fuzzy measures are sometimes called nonnegative monotone set functions. In multiple classifier fusion, nonnegative set functions are used to describe the importance of classifiers and the interaction among classifiers. The value of set function at a single-point-set $\mu(\{x_i\})$ presents the contribution of the single classifier x_i towards classification, and the value of set function at other sets, such as $\mu(\{x_i, x_j, x_k\})$, presents the joint contribution of classifiers towards classification. Mainly the methods to determine the nonnegative set functions have two types. One is to learn from the history data (Wang et al., 2005; Wang & Wang, 1997; Yang et al., 2008; Yeung et al., 2004) and the other is to specify by experts.

Once the set functions are available, we can use the fuzzy integral to aggregate the outputs from all classifiers. The ith column of DP matrix can be regarded as a function f_i defined on classifier set X, $f_i(x_j) = d_{j,i}$. The integral of function f_i with respect to nonnegative set function μ_i is the degree of fusion system classifying a sample to class C_i. If necessary, we can obtain the crisp class label through $C_t = \arg\max_{1 \leq i \leq c}(\int f_i d\mu_i)$.

Usually the type of fuzzy integral is chosen in advance. Choquet fuzzy integral and Sugeno fuzzy integral are often selected in fusion process. Noting that the addition and the multiplication operators are used in Choquet integrals while the maximum and the minimum operators are used in Sugeno integral, most researchers prefer now to use the Choquet integral in classifier fusion models (Wang et al., 2005). The classification process of a sample by a fused system based on fuzzy integral is illustrated in Fig. 1.

Fig. 1 shows that a sample is first submitted to all classifiers and the results from all classifiers are stored in a DP matrix. Each column of the matrix is a function defined on set X. Then the final classification result can be obtained by calculating the integral of each column of the DP matrix. The crisp class label can be finally obtained through the maximum if necessary.