Applied Soft Computing Journal 73 (2018) 795–815

Contents lists available at ScienceDirect

Applied Soft Computing Journal

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

Combining heterogeneous classifiers via granular prototypes

Tien Thanh Nguyen [a,*], Mai Phuong Nguyen [b], Xuan Cuong Pham [c], Alan Wee-Chung Liew [c], Witold Pedrycz [d]

[a] School of Computing Science and Digital Media, Robert Gordon University, Aberdeen, Scotland, United Kingdom
[b] FPT Company, Hanoi, Viet Nam
[c] School of Information and Communication Technology, Griffith University, Australia
[d] Department of Electrical & Computer Engineering, University of Alberta, Edmonton, AB, T6R 2V4, Canada

HIGHLIGHTS

- We modelled the base classifiers' output by using a granular prototype.
- We quantified the distance between base classifiers' output and a granular prototype.
- We proposed a novel framework to combine multiple classifiers in an ensemble system.
- The proposed method is highly competitive to some state-of-the-art ensemble methods.

ARTICLE INFO

Article history:
Received 16 May 2017
Received in revised form 6 September 2018
Accepted 13 September 2018
Available online xxxx

Keywords:
Ensemble method
Multiple classifiers system
Information granule
Information uncertainty
Supervised learning

ABSTRACT

In this study, a novel framework to combine multiple classifiers in an ensemble system is introduced. Here we exploit the concept of information granule to construct granular prototypes for each class on the outputs of an ensemble of base classifiers. In the proposed method, uncertainty in the outputs of the base classifiers on training observations is captured by an interval-based representation. To predict the class label for a new observation, we first determine the distances between the output of the base classifiers for this observation and the class prototypes, then the predicted class label is obtained by choosing the label associated with the shortest distance. In the experimental study, we combine several learning algorithms to build the ensemble system and conduct experiments on the UCI, colon cancer, and selected CLEF2009 datasets. The experimental results demonstrate that the proposed framework outperforms several benchmarked algorithms including two trainable combining methods, i.e., Decision Template and Two Stages Ensemble System, AdaBoost, Random Forest, L2-loss Linear Support Vector Machine, and Decision Tree.

1. Introduction

Supervised learning is an active research area in the machine learning community. Many algorithms resulting from different learning methodologies have been introduced to learn the relationship between feature vectors and class labels with the aim of generating discriminative decision model. Experiments have shown that there is no single learning algorithm that performs well on all datasets. A learner can achieve high accuracy on some datasets but high error rate on others. Ensemble learning, where multiple learning algorithms are combined into a single framework to obtain a better discriminative decision model, offers a viable solution [1].

Dietterich [2] showed the benefit of combining multiple classifiers from three aspects: statistical, computational, and representational. When a classifier is learned on a given training set, it gives a hypothesis about the relationship between the feature vectors and the class labels. With a small number of training data, different hypotheses (classifiers) can produce the same error rate on the training data. It might happen that a poor hypothesis is chosen to predict the label of an unseen sample. By combining several hypotheses, we can reduce the risk of choosing a wrong hypothesis. From the computational aspect, many

* Corresponding author.
E-mail address: t.nguyen11@rgu.ac.uk (T.T. Nguyen).

https://doi.org/10.1016/j.asoc.2018.09.021
1568-4946/

algorithms perform local search to obtain locally optimum solution. In ensemble methods, by changing the starting point of algorithms, we can have a better approximation of the unknown relationship than that of a single learning algorithm. Finally, the unknown relationship in some cases cannot be modeled by a single hypothesis. By using a combination of multiple hypotheses, a better approximation for the relationship can be achieved.

In ensemble method, different "models" could refer to the different learning algorithms or to a set of generic classifiers generated by learning a unique learning algorithm on many different training sets [3]. Each learning algorithm learns a classifier on a given training set to describe the relationship between the feature vector and the class label of the training observations. The generated classifier returns the posterior probabilities, i.e., numerical class memberships that an observation belongs to different classes. A combination method is then used to aggregate the outputs of all classifiers to generate the discriminative model. As each classifier may output different results on each observation, uncertainty is introduced.

A combiner which can capture the facet of uncertainty when combining the base classifiers' outputs would be desirable. In the literature, several combiners have been introduced based on this consideration, such as fuzzy IF-THEN rule-based combiner [4] and Decision Template method [5]. In this study, we propose an ensemble framework based on modeling the uncertainty in the base classifiers' output using interval-based representations [6,7]. Here interval-based representations are generated by the notion of information granularity. Starting from the pioneering work of Zadeh [8–10], the concept of information granules have been used to model human cognitive and decision-making activities [11–13], and have been applied to many real-world applications [14].

In homogeneous ensemble methods like AdaBoost [15], Bagging [16], and Random Forest [17], the focus is on the generation of new training schemes from the original training set. Meanwhile, in the heterogeneous ensemble systems, a fixed set of different learning algorithms learns on the same training set to generate the different base classifiers. The outputs of these classifiers (called meta-data of Level1 data) are then combined to make the final prediction [3–5,18]. In this type of ensembles, the approach is focused on designing algorithms that combine the meta-data to achieve higher accuracy than that using a single classifier. In this work, we use the principle of justifiable information granularity to generate granular prototypes resulting from the outputs, i.e. the meta-data, of a set of base classifiers of heterogeneous ensemble obtained from the training observations. By defining a distance function between a feature vector and a granular prototype, we propose a novel combining algorithm for the heterogeneous ensemble systems via a shortest distance-based mechanism.

The novelty of our work lies in the following:

(i) To the best of our knowledge, this is the first approach that models the uncertainty in the meta-data of training observations by using the granular prototype formalized as a vector of intervals.
(ii) We define a way to quantify the distance between the meta-data (a numerical vector) of an observation and a granular prototype (a vector of intervals).
(iii) We propose a novel combining algorithms for heterogeneous ensemble system via a shortest distance-based mechanism.

The paper is organized as follows. In Section 2, heterogeneous ensemble method and the concept of justifiable granularity in the design of information granules are introduced. In Section 3, the novel combining method based on the idea of justifiable granularity is proposed. Experimental results are presented in Section 4; here the results of the proposed method are compared with the results produced by a number of benchmark algorithms when using 26 datasets. Finally, the conclusions are presented in Section 5.

2. Related work

2.1. Ensemble method

Over the past years, many approaches related to ensemble methods have been proposed, and there are different taxonomies of ensemble methods [1,18–22]. We follow the taxonomy in [22] in which ensemble methods are divided into two types:

- Homogeneous ensemble: A set of classifiers are generated on different training sets obtained from an original one by using the same learning algorithm. The outputs of these classifiers are combined to give the final decision. Several state-of-the-art ensemble methods in the literature are AdaBoost [15], Bagging [16], and Random Forest [17].
- Heterogeneous ensemble: Several different learning algorithms are learned on the same training set to generate the different base classifiers. The heterogeneous ensemble focuses more on the combining strategies on the meta-data [3, 18,23–26]) to achieve higher accuracy than a single classifier.

In the literature, besides the practical applications of ensemble methods in many areas, research on ensemble methods can be divided into three aspects:

- **Design of new ensemble systems**: Several recent research efforts have focused on designing new ensemble systems. Rodriguez et al. [27] proposed the Rotation Forest in which principal component analysis (PCA) is applied to each of the K subsets randomly selected from a feature set. The K axis rotations form the new features for a base classifier. Blaser and Fryzlewicz [28] designed a novel ensemble system by generating random rotation matrices to rotate the feature space before generating the base classifiers. Wu [29] proposed a new ensemble learning paradigm with the consideration of implicit supplementary information about the performance orderings for the trained base classifiers in previous literature. By measuring the similarity between the two learning tasks, the supplementary ordering information for the trained classifiers of a given learning task can be inferred so as to obtain the optimal combining weights of the trained classifiers. Moreover, several ensemble systems were developed for different learning paradigms such as incremental learning [30–32], semi-supervised learning [33], and multi-label learning [34,35]. For instance, Pham et al. [31] combined random projections and Hoeffding tree to construct an incremental online ensemble learning system. Krawczyk and Cano [32] incrementally learnt a threshold for each arrived instance in the online heterogeneous ensemble system. Classifier are selected for the prediction if their support on each instance exceeds the threshold. Wu et al. [35] proposed ML-FOREST algorithm to learn an ensemble of hierarchical multi-label classifier trees to reveal the intrinsic label dependencies. Finally, besides the two popular combiners i.e. Sum and Majority Vote [4,36], novel combining algorithms were introduced to enhance the task of combining on classifiers' outputs. For example, Kuncheva et al. [18] used the Ordered Weighted Averaging (OWA) operators to aggregate the classifiers' outputs. Wang et al. [37] proposed a new fusion scheme based on the upper integrals. Costa et al. [38] used the generalized mixture functions as a combining algorithm in which the weight each classifier put on a class was set dynamically in the combination process.