Engineering Applications of Artificial Intelligence 74 (2018) 43–53

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Integration of an improved dynamic ensemble selection approach to enhance one-vs-one scheme

Zhong-Liang Zhang [a,b], Xing-Gang Luo [a,b,*], Yang Yu [b], Bo-Wen Yuan [b], Jia-Fu Tang [b]

[a] School of Management, Hangzhou Dianzi University, 310018 Hangzhou, China

[b] School of Information Science and Engineering, Northeastern University, 110819 Shenyang, China

ARTICLE INFO

Keywords:
Dynamic selection
Heterogeneous ensemble
One-vs-one
Decomposition strategy
Multi-class classification

ABSTRACT

The One-vs-One (OVO) scheme that decomposes the original more complicated problem into as many as possible pairs of easier-to-solve binary sub-problems is one of the most popular techniques for handling multi-class classification problems. In this paper, we propose an improved Dynamic Ensemble Selection (DES) procedure, which aims to enhance the OVO scheme via dynamically selecting a group of appropriate heterogeneous classifiers in each sub-problem for each query example. To do so, twenty heterogeneous classification algorithms are selected to obtain a set of candidate classifiers for each sub-problem derived from the OVO decomposition. Then, a simple yet efficient DES procedure is developed to execute the dynamic selection for each query example in each sub-problem. Finally, all the selected binary heterogeneous ensembles are combined by using majority voting to obtain the final output class. To evaluate the proposed method, we carry out a series of experiments on twenty datasets selected from the KEEL repository. The results supported by proper statistical tests demonstrate the validity and effectiveness of our proposed method, compared with state-of-the-art methods for OVO-based multi-class classification.

1. Introduction

Many real-world applications involve classification tasks, such as face recognition (Juefei-Xu and Savvides, 2016), document categorization (Tang et al., 2016), and medical diagnosis (Peker, 2016). According to the number of classes involved in the task, the classification problem can be roughly categorized into two groups: binary classification problem and multi-class classification problem. In the former one, there are only two classes needed to be distinguished, while in the latter case the problem refers to more than two classes.

It is obvious that multi-class classification tasks are much harder to handle, since there are more classes involved to be identified. One of the most popular strategies to deal with multi-class classification problems is considered as binarization decomposition, which has recently attracted much attention in the research community of data mining and machine learning (Galar et al., 2011). Several alternatives for the decomposition of multi-class problems into a set of two-class classification problems can be found in the specialized literature (Lorena et al., 2008). Among them, One-vs-One (OVO) (Knerr et al., 1990) and One-vs-All (OVA) (Clark and Boswell, 1991) are the most popular techniques, which can be included within Error Correcting Output Codes (ECOC) framework (Allwein et al., 2000; Dietterich and Bakiri, 1995). In this study, we focus on the OVO scheme, which is widely applied in the well-known software tools to obtain multi-class Support Vector Machines (SVMs) classifier, such as KEEL (Alcalá-Fdez et al., 2009), LIBSVM (Chang and Lin, 2011) and WEKA (Hall et al., 2009).

In the OVO scheme, a divide-and-conquer strategy is used to simplify the initial multi-class problems. Concretely, the original multi-class problem is first decomposed into as many as possible pairs of binary sub-problems. Then, a base learner is employed to obtain the binary classifier for each sub-problem. Finally, all the binary classifiers are combined by an appropriate aggregation method to predict the output of the query example which is needed to be classified. The previous studies (Galar et al., 2011) have shown the effectiveness of the OVO scheme to address the multi-class problems despite its simplicity. Regarding to the aggregation method for combining binary classifiers in the OVO scheme, many approaches have been developed, such as majority voting (MV) (Friedman, 1996), weighted voting (WV) (Hüllermeier and Vanderlooy, 2010), learning valued preference for classification (LVPC) (Hühn and Hüllermeier, 2009; Hüllermeier and Brinker, 2008), and preference relations solved by non-dominance criterion (ND) (Orlovsky, 1978; Fernández et al., 2010).

* Corresponding author at: School of Management, Hangzhou Dianzi University, 310018 Hangzhou, China.
E-mail address: xgluo@mail.neu.edu.cn (X.-G. Luo).

https://doi.org/10.1016/j.engappai.2018.06.002
Received 16 September 2017; Received in revised form 20 April 2018; Accepted 7 June 2018

In the classic OVO scheme, each sub-problem is handled by the same single classification algorithm, which could lead to a problem of non-competent classifiers in some sub-problems. In order to fix this issue, a diversified OVO is developed to select the best classification algorithm for each class pair (Kang et al., 2015). Following this line of research, a strategy based on Dynamic Classifier Selection (DCS) is applied in the OVO scheme, which dynamically select the best base classifier for each query instance in each sub-problem (Mendialdua et al., 2015). More recently, we have proposed to improve the performance of the OVO decomposition scheme using the Dynamic Ensemble Selection (DES) procedure based on a competence measure for each sub-problem, where the pairwise classes are distinguished by selecting the most appropriate ensemble for each query example (Zhang et al., 2017).

The previous works have indicated that the boundaries distinguishing the different sub-problems vary depending on the classes. That is, the difficulty in each sub-problem derived from the OVO decomposition can be different. Therefore, the performance of an OVO system can be improved through selecting the most appropriate classifiers for each sub-problem. On this account, we develop in this paper a novel approach named as OVO-DHES, which aims to enhance the performance of the OVO scheme using a new dynamic selection of heterogeneous ensemble procedure for each sub-problem. Specifically, the initial multi-class classification problem is firstly decomposed into as many as possible pairs of sub-problems. Then, the selected learners are independently employed to generate candidate binary classifiers pool for each sub-problem, and a group of appropriate heterogeneous classifiers are dynamically selected from the corresponding pool to distinguish each pair class for each query instance. Finally, the majority voting method is employed to combine all the binary ensembles to obtain the final output for the query instance.

In order to verify the usefulness and effectiveness of our proposal, we carry out a thorough experimental study. To do so, twenty real-world applications selected from KEEL dataset repository (Alcalá-Fdez et al., 2011) are used in our experiments. Two assessment metrics, both classification accuracy and Cohen's kappa (Cohen, 1960), are considered to evaluate the performance of the tested methods. In addition, the proper statistical tests recommended in Derrac et al. (2011) and García et al. (2010) are employed to show the significance of the obtained results. The following main contributions of this paper with respect to previous works can be summarized:

- We investigate the inner mechanism of the OVO scheme and point out the potential improvement for the OVO scheme using the dynamic selection.
- We propose a novel methodology which aims to enhance the OVO scheme through employing dynamic heterogeneous ensemble selection for each sub-problem.
- With respect to the dynamic selection of the appropriate ensemble of classifiers, we develop a new procedure which aims to obtain the most appropriate classifiers for each query example in each sub-problem.
- We also extend several well-known DES-based approaches into the OVO scheme for multi-class classification problems and carry out the comparison of the extension and our OVO-DHES.

The rest of this paper is organized as follows. Firstly, some related works with respect to this study are recalled in Section 2. Then, in Section 3, we present the details of our proposal. Next, Section 4 describes the experimental framework. The experimental study is carried out in Section 5, while the discussions and conclusions are presented in the final section.

2. Background

In this section, we describe some works related to this paper. Firstly, we present the OVO decomposition scheme and the MV aggregation strategy. Then, we introduce the DES procedure and recall several well-known DES-based approaches.

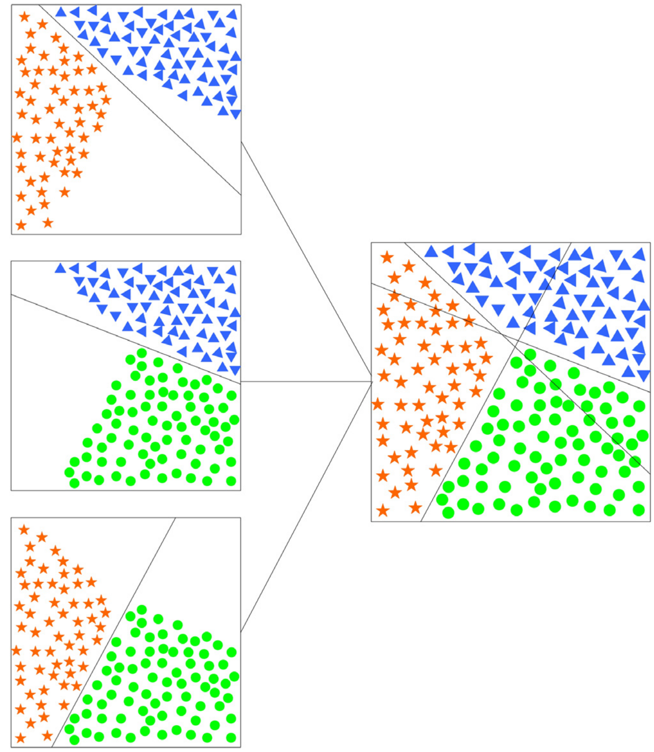

Fig. 1. An example of the OVO decomposition scheme for a three-class problem.

2.1. One-versus-one scheme

In the OVO decomposition scheme, an m-class problem is divided into $m(m-1)/2$ binary class sub-problems. Then, a predefined classification algorithm is employed to learn the hypothesis to distinguish between each pair of classes. In this way, the binary classifiers are constructed by the usage of a subset of instances from the original training dataset, which contains only two corresponding class labels, while ignoring the instances with other different class labels. An example of the OVO decomposition scheme for a three-class classification problem is shown in Fig. 1.

In the prediction phase, a query example is classified as one of the m classes depending on the outputs given by the binary classifiers. In order to do so, a score-matrix R containing the outputs of binary classifiers is usually employed to decide the final class:

$$R = \begin{pmatrix} - & r_{12} & \cdots & r_{1m} \\ r_{21} & - & \cdots & r_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & \cdots & \cdots & - \end{pmatrix} \tag{1}$$

where $r_{ij} \in [0,1]$ is the confidence of the binary classifier distinguishing class i from j, whereas, the confidence in favor of class j is computed as $r_{ji} = 1 - r_{ij}$ (if it is not provided by the classifier). Once the score-matrix is produced, any of the aggregation methods (Galar et al., 2011) can be employed to infer the final class. In this work, we use the well-known aggregation method, i.e. the MV strategy, to combine the binary classifiers. The MV aggregation method is also called binary voting and Max-Wins rule. In this method, each binary classifier trained from the dataset with a pair of labels gives a vote to the corresponding class. The class with the largest number of votes received by each class is considered as the predicted label:

$$class = \arg \max_{i=1,\ldots,m} \sum_{1 \le j \ne i \le m} s_{ij} \tag{2}$$

where $s_{ij} = \begin{cases} 1, & r_{ij} > r_{ji} \\ 0, & \text{otherwise} \end{cases}$.