Expert Systems with Applications 42 (2015) 8134–8145

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Hierarchical multi-class LAD based on OvA-binary tree using genetic algorithm

Hwang Ho Kim [a], Jin Young Choi [b,*]

[a] Department of Industrial Engineering, Seoul National University, Seoul, Republic of Korea
[b] Department of Industrial Engineering, Ajou University, Suwon, Republic of Korea

ARTICLE INFO

Article history:
Available online 29 June 2015

Keywords:
Hierarchical multi-class classification
Logical analysis of data
One versus all-binary tree
Genetic algorithm
Classification accuracy

ABSTRACT

Recently, logical analysis of data (LAD) using a classifier based on a linear combination of patterns has been introduced, providing high classification accuracy and pattern-based interpretability on classification results. However, it is known that most of LAD-based multi-classification algorithms have conflicts between classification accuracy and computational complexity because they are based on class decomposition method such as one versus all or one versus one. Furthermore, it is difficult to explain the decision rule in the classification procedure because they only use the final scores calculated by classifiers. To overcome this issue, in this paper, we propose a hierarchical multi-class classification method using LAD based on a one versus all (OvA)-binary tree, called hierarchical multi-class LAD (HMC-LAD). It constructs an OvA-binary tree by partitioning a node with $K(\geqslant 2)$ classes into two sub-nodes by identifying one distinct class from the remaining $(K-1)$ classes repeatedly. Specifically, we suggest a node partition method for constructing an efficient OvA-binary tree, genetic algorithm for generating patterns for a node under consideration, and OvA-binary tree exploration method for performing multi-class classification. Through a numerical experiment using benchmark datasets from the UCI machine-learning repository, we confirm that (i) the suggested node partition method is efficient compared to a random partition method, and (ii) the classification performance of HMC-LAD is superior to existing multi-class LAD algorithms and other supervised learning approaches. The proposed HMC-LAD can be applied to expert and intelligent systems to effectively categorize large amount of data in knowledge base and perform inference for decision making.

1. Introduction

Classification is a machine-learning task that infers a classifier from labeled training data and uses it to classify unseen test data x as class $y = \{1, 2, \ldots, K\}$. A problem with two classes ($K = 2$) is called a binary classification problem (BCP); a problem with more than two classes ($K \geqslant 3$) is called a multi-class classification problem (MCCP) (Fayyad, Piatetsky-Shapiro, & Smyth, 1996; Lugosi, 2002; Mohri, Rostamizadeh, & Talwalkar, 2012).

Because the number of classes in a BCP is less than that in an MCCP, a BCP is easier to solve than an MCCP. Therefore, classification algorithms with high classification accuracy and low computational complexity to solve BCPs have been addressed with several techniques. Decision trees (DT) (Rokach, 2008), Bayesian networks (Friedman, Geiger, & Goldszmidt, 1997), support vector machines (SVM) (Scholkopf & Smola, 2001), and neural networks (NN) (Hopfield, 1988) are widely used to solve BCPs. Recently, logical analysis of data (LAD) using a classifier based on a linear combination of patterns has been introduced and adopted in fields such as medical and manufacturing and has yielded high classification accuracy (Alexe, Alexe, Bonates, & Kogan, 2007; Boros, Hammer, Ibaraki, & Kogan, 1997; Boros et al., 2000; Hammer & Bonates, 2006).

An MCCP can be applied to a wider variety of real-world applications and its importance has been emphasized in recent years. Specifically, in expert and intelligent systems, accumulating knowledge in a simple and clear way is the key point for implementation, so that knowledge categorization is required to effectively save large amount of data in knowledge base and perform inference for decision making. MCCP can play an important role in knowledge categorization and knowledge inference by providing a systematic and computationally efficient procedure for data classification with multi-class. Wafer pattern recognition (Hsu & Chien, 2007) and grade prediction of semiconductors via log dataset (Chien, Wang, & Cheng, 2007), and fault diagnosis (Mortada, Yacout, & Lakis, 2014) in power transformers are representative

* Corresponding author.
E-mail address: choijy@ajou.ac.kr (J.Y. Choi).

http://dx.doi.org/10.1016/j.eswa.2015.06.037
0957-4174/

examples of MCCP applications for expert and intelligent systems. However, because an MCCP must consider a relatively greater number of classes than a BCP, the computational complexity of the classification algorithm is significantly higher and limited methods for MCCP have been studied (Bishop, 2006).

The basic approach to solve an MCCP is decomposition where one MCCP is divided into several BCPs and each of these is solved using binary classifiers (Masulli & Valentini, 2000). Various decomposition methods have been suggested; the most common approaches are one versus all (OvA)-type and one versus one (OvO)-type decomposition (Aly, 2005). OvA-type decomposition divides an MCCP with K classes into two classes, class i as the positive set and the remaining $(K-1)$ classes as the negative set, eventually building K different BCPs. Then, for each BCP, it determines one classifier f_i. By combining all classifiers $f_i(i = 1, 2, \ldots, K)$, it establishes a function:

$$f(x) = \arg\max_i f_i(x), \tag{1}$$

that can classify test data x as class i with the index giving the greatest value of $f_i, \forall i$.

Given a dataset with K classes, OvO-type decomposition builds $K(K-1)$ BCPs by considering all possible pairs of classes, and determines one classifier f_{ij} for each BCP. Then, by combining all classifiers $f_{ij}(i = 1, 2, \ldots, K, j \neq i)$, it constructs a function:

$$f(x) = \arg\max_i \left(\sum_{j \neq i} f_{ij}(x) \right), \tag{2}$$

that can classify test data x as class i corresponding to the index giving the greatest value of $\sum_{j \neq i} f_{ij}, \forall i$.

Based on these decomposition methods, several multi-class classification methods using SVM, NN, and LAD have been suggested (Aly, 2005; Hsu & Lin, 2002). Specifically, various extensions of LAD using decomposition approaches (called multi-class LAD (MC-LAD)) to solve an MCCP have been proposed recently. According to the decomposition type, they can be categorized into OvO-type and OvA-type MC-LAD. Research on OvO-type MC-LAD was conducted by Moreira (2000), Mortada et al. (2014), and Kim and Choi (2015). Moreira (2000) proposed two methods to decompose one MCCP into BCPs. The first method, called a multi-layered 2-class LAD, used the typical OvO-type approach that does not require any modification of the structure of the original LAD algorithm. The second method modified the architecture of the pattern generation and theory formation steps in the original LAD algorithm and introduced the decomposition matrix used for recording multi-class patterns generated through iterations of LAD. Mortada et al. (2014) proposed an MC-LAD algorithm integrating the ideas from the second approach of Moreira (2000) with the pattern generation method using a mixed integer and linear programming (MILP) model by Ryoo and Jang (2009). They showed that the classification accuracy of the proposed method is higher than the LAD models proposed by Moreira (2000). However, the MILP-based approach has the disadvantage that it may require excessive computational time and effort to develop the MILP model. Kim and Choi (2015) proposed an iterative genetic algorithm (GA) with flexible chromosomes and multiple populations (IGA-FCMP) to improve the shortcomings of the existing LAD pattern generation methods including the MILP-based approach.

OvA-type MC-LAD algorithms were suggested by Herrera and Subasi (2013) and Kim and Choi (2015). Herrera and Subasi (2013) proposed an algorithmic approach using MILP to efficiently generate OvA-type LAD patterns in a multi-class dataset. Kim and Choi (2015) applied the IGA-FCMP to generate OvA-type patterns for MC-LAD. Through a numerical experiment, they illustrated that the accuracy of this approach was inferior to that of OvO-type MC-LAD using IGA-FCMP. However, it was more efficient than OvO-type MC-LAD using IGA-FCMP because the number of generated pattern sets was smaller than that using OvO-type.

These two decomposition approaches have conflicts between classification accuracy and computational complexity. OvA-type decomposition approaches require computational complexity of $O(K)$ to construct a classifier; their classification accuracy is relatively low. OvO-type decomposition approaches have higher classification accuracy; however, they require a computational complexity of $O(K^2)$ to generate a classifier, resulting in higher computational complexity as the number of classes K increases. Furthermore, decomposition approaches cannot explain the decision rule in the classification procedure because they only use the final scores calculated by classifiers. Therefore, they cannot be applied to expert and intelligent systems effectively where multi-class classification can play an important role in knowledge categorization and knowledge inference.

To overcome this issue, a new approach, called a hierarchical classification approach has been studied in recent years (Hernández, Sucar, & Morales, 2014; Lorena, De Carvalho, & Gama, 2008; Silla & Freitas, 2011). Since the classification performance can depend on the tree structure, tree building approach, and tree exploration method, the state-of-the-art in hierarchical classification approach for MCCP can be classified by using three criteria as follows (Bi & Kwok, 2012; Hernández et al., 2014).

- The type of hierarchical structure used: Directed Acyclic Graph-based (Cheong, Oh, & Lee, 2004), Decision Tree-based (Weston & Watkins, 1998), Binary Tree-based (Hsu & Lin, 2002)
- How deep the classification in the hierarchy is performed (Bi & Kwok, 2012): mandatory leaf-node prediction, non-mandatory leaf-node prediction
- How the hierarchical structure is explored (Hernández et al., 2014): Local (also known as top–down), Global (also known as Big-Bang), or Flat (only predicts the leaf nodes)

The outstanding characteristic of these is to use a tree structure, where the root node includes all K classes, and each sub-node contains the associated class information and classifier. Classification for test data x is performed by tree exploration, beginning at the root node until x reaches a leaf node corresponding to a certain class.

In particular, binary tree-based hierarchical multi-class classification using top-down approaches have been widely accepted due to high classification accuracy and low computational complexity (Hernández et al., 2014). They can be classified by using two criteria such as a class decomposition method, and a binary classification algorithm used for generating a binary tree. First, there are two types for class decomposition; Half versus Half (HvH)-type and OvA-type. By using a BCP algorithm, HvH-type method divides a node with K classes into two sub-nodes with $\lceil K/2 \rceil$ classes and $\lceil (K+1)/2 \rceil$ classes, respectively. This method had disadvantages such that it has excessive computational time compared to OvA-type method and the correlation between classes becomes large as the number of classes increases (Chen, Wang, & Wang, 2009; Fei & Liu, 2006; Lei & Govindaraju, 2005; Madzarov, Gjorgjevikj, & Chorbev, 2009). To compensate these limitations, OvA-type method has been introduced recently, where it divides a node with K classes into two sub-nodes with 1 and $(K-1)$ classes, respectively. It showed high computational efficiency and classification accuracy compared to HvH-type method (Ramanan, Suppharangsan, & Niranjan, 2007; Sidaoui & Sadouni, 2014; Wu, Lee, & Yang, 2008; Yang, Yu, He, & Guo, 2013).

Second, the majority of the studies for binary classification algorithm used for hierarchical classification proposed SVM-based approaches. SVMs are used as the binary classifier at