Computers & Industrial Engineering 66 (2013) 383–388

Contents lists available at SciVerse ScienceDirect

Computers & Industrial Engineering

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

Improving the efficiency of a mixed integer linear programming based approach for multi-class classification problem

Alaleh Maskooki *

Department of Applied Mathematics, Ferdowsi University of Mashhad (FUM), Pardis Campus, Azadi Square, Mashhad, Iran

ARTICLE INFO

Article history:
Received 19 June 2012
Received in revised form 4 May 2013
Accepted 8 July 2013
Available online 17 July 2013

Keywords:
Data classification
Mixed integer linear programming
Hyper-boxes
Iterative algorithm

ABSTRACT

Data classification is one of the fundamental issues in data mining and machine learning. A great deal of effort has been done for reducing the time required to learn a classification model. In this research, a new model and algorithm is proposed to improve the work of Xu and Papageorgiou (2009). Computational comparisons on real and simulated patterns with different characteristics (including dimension, high overlap or heterogeneity in the attributes) confirm that, the improved method considerably reduces the training time in comparison to the primary model, whereas it generally maintains the accuracy. Particularly, this speed-increase is significant in the case of high overlap. In addition, the rate of increase in training time of the proposed model is much less than that of the primary model, as the set-size or the number of overlapping samples is increased.

1. Introduction

Data classification is generally known as a method for constructing a set of classification rules (classifier), based on data attributes, in order to predict class membership of unknown data. This task is accomplished in two phases. In the training phase, a set of data with known membership (training dataset), is used to estimate the parameters of boundaries between classes. After these boundaries are determined, in the second phase, some new data (test data), which are not used for training, is assigned to a class according to the boundaries obtained, in order to evaluate the accuracy of the model. The latter phase is called testing procedure. For instance, a dataset can be a set of vectors representing chemical attributes of patients' blood and each class can be a disease.

Extensive applications of data classification in different fields of engineering, medicine, industry and finance, result in developing a variety of methods. Decision tree induction (Safavian & Landgrebe, 1991), neural networks (Hertz, Palmer, & Krogh, 1991), Bayesian networks (Russell & Norvig, 2010) and genetic algorithm are some examples.

A simple classifier, which is constructed by a linear programming (LP) formulation, can be in the form of a hyper-plane separating two sets (Bennett & Mangasarian, 1992a; Freed & Glover, 1986; Glover, 1990; Glover, Keene, & Duea, 1988; Lam, Choo, & Moy, 1996; Lam & Moy, 2002). Erenguc and Koehler (1990) studied different types of LP models according to their objective functions. In addition, Koehler (1990) studied the problems that may appear in formulating an LP model, such as choosing an objective function, unbounded or unacceptable solutions, side constraints, data translation and transformation. A survey on mathematical programming models can be found in Pai (2009) and Lee and Wu (2009).

Since hyper-planes are not sufficiently accurate in many real cases, piecewise linear classifiers (as an approximation for nonlinear boundaries) have extensively been studied. Multi Surface Method (MSM) (Mangasarian, 1968) is a piecewise LP-based classifier that forms a hyper-plane at each stage until all data points become completely separated. Ryoo (2006) has proposed a Piecewise Linear and Convex (PLC) discriminant function and used it to develop a mixed integer linear programming (MILP) model. He has compared the PLC method with two common methods including MSM. Results show that, MSM constructs more complex boundaries than the ideal shape when using on datasets with high overlap. On the other hand, PLC is appropriate for data with convex boundaries and may fail to perform accurately in other cases. A piecewise linear function is introduced by Astorino and Gaudioso (2002) that constructs multiple hyper-planes to form a convex polyhedron. Comparing to other linear methods, this model requires high numerical computations.

All the above models are introduced for solving the two-class classification problem. There are fewer mathematical programming models for the multi-class case. Bennett and Mangasarian (1992b) have proposed a piecewise linear classifier which is an extension of their former model for the two-class case. An LP model is proposed by Bal and Orkcu (2011) which is a combination of three LP models proposed by Gehrlein (1986), Gochet, Stam, Srinivasan, and Chen (1995) and Sueyoshi (2006).

* Tel.: +989151249299; fax: +98 5118439924.
E-mail address: a.maskooki@yahoo.com

0360-8352/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cie.2013.07.005

Freed and Glover (1981) presented a linear programming model that assigns intervals to classes. In this model, the data are classified into intervals according to their discriminant scores. A drawback of this method is that an appropriate order for classes should be determined in advance; otherwise, it yields low accuracy. Gehrlein (1986) suggested an MILP method that overcomes the problem of sorting, by introducing binary variables for each pair of classes. If the intervals in Gehrlein's model are defined for each direction (attribute) separately, and data coordinates in $\mathbb{R}^m$ are used instead of discriminant scores in $\mathbb{R}$, then hyper-boxes are produced instead of a single hyper-plane. Uney and Turkay (2006) presented a new data classification method based on the use of hyper-boxes for determining the class boundaries. They developed an MILP model by converting the relationship among discrete variables to their equivalent integer constraints using Boolean algebra. They evaluated the performance of their proposed method on Iris dataset (Fisher, 1936). Xu and Papageorgiou (2009) formulated an MILP model in a similar concept for enclosing the training data using hyper-boxes. They suggested an algorithm that applies the MILP model iteratively in order to improve the accuracy. Kone and Karwan (2011) extended this model for predicting cost-to-serve (CTS) values of new customers in industrial gas business.

Although hyper-planes are efficient in classifying data into two sets, they can be inaccurate and inefficient when applied to solve multi-class problems (Uney & Turkay, 2006). Fig. 1 shows the schematic representation of classifying three datasets using hyper-planes and hyper-boxes. As can be seen in Fig. 1, hyper-boxes have more flexibility for estimating ideal class boundaries of a pattern in comparison to rigid hyper-planes, particularly for classifying multiple classes. For instance, if square or circle data points are omitted from the pattern in Fig. 1, then the other two sets can be completely separated by a single hyper-plane, but there are misclassified points when using hyper-planes for separating three sets. In addition, the boundaries, which are constructed using hyper-boxes, look closer to the actual class boundaries in comparison to the borders obtained on the same dataset using hyper-planes. However, integer linear programming (ILP) problems are generally NP-complete (Schrijver, 1998). This confirms that, large-scale ILP problems are impractical due to being extremely time-consuming for real-world applications.

This research is focused on Xu and Papageorgiou's MILP-based method. This method can be inefficient when the size of the training set or the overlapping area is increased. Modifications are suggested to improve the efficiency of their proposed model and algorithm.

In the next section, the MILP model and the related iterative algorithm, which are suggested by Xu and Papageorgiou (2009) is briefly stated. The new approach is proposed in Section 3. Numerical computations and comparisons of two methods are illustrated in Section 4. Section 5 concludes the discussion.

Notations are as follows: Consider a classification problem with G classes and n samples ($i = 1,\ldots,n$). The class membership of samples is supposed to be known. Each sample is a vector in $\mathbb{R}^m$ where m is the number of attributes. The parameter a_{ij} represents the value of sample i on attribute j ($j = 1,\ldots,m$).

2. Xu and Papageorgiou's MILP model

In this section, Xu and Papageorgiou (2009)'s model for the multi-class classification problem is summarized. The training process is performed in two stages. In the first stage, an MILP model produces boundaries that form one hyper-box for each class. A hyper-box r is recognized by its central coordinate (B_{rj}) and length (LE_{rj}) on each attribute $j(j = 1,\ldots,m)$ and is assigned to one of the classes $1,\ldots,G$. r_i exists if sample i is assigned to hyper-box r. The objective is to minimize the total number of misclassifications by numerating non-zero variables E_i. The binary variable E_i is equal to 1 if sample i is included in the corresponding hyper-box and is zero otherwise. In addition, the binary variable Y_{rsj} is introduced to prevent boxes with different class labels overlapping each other. Y_{rsj} is zero if box r and s do not overlap on attribute j; otherwise it is equal to 1. U and ε are respectively large and small positive constants with arbitrary values. The parameter A is the initial number of hyper-boxes. The complete MILP model for Multi-Class data classification Problem (MCP) is formulated as follows:

$$\min \sum_{i=1}^{n}(1 - E_i)$$

Subject to:

$$a_{ij} \geqslant B_{rj} - \frac{LE_{rj}}{2} - U(1 - E_i) \quad \forall i, r_i, j \tag{1}$$

$$a_{ij} \leqslant B_{rj} + \frac{LE_{rj}}{2} + U(1 - E_i) \quad \forall i, r_i, j \tag{2}$$

$$B_{rj} - B_{sj} + U \cdot Y_{rsj} \geqslant \frac{LE_{rj} + LE_{sj}}{2} + \varepsilon \quad \forall j\, \forall r, s = 1,\ldots,A; \quad s \neq r \tag{3}$$

$$\sum_{j=1}^{m}(Y_{rsj} + Y_{srj}) \leqslant 2m - 1 \quad \forall r = 1,\ldots,A-1, \quad s = r+1,\ldots,A \tag{4}$$

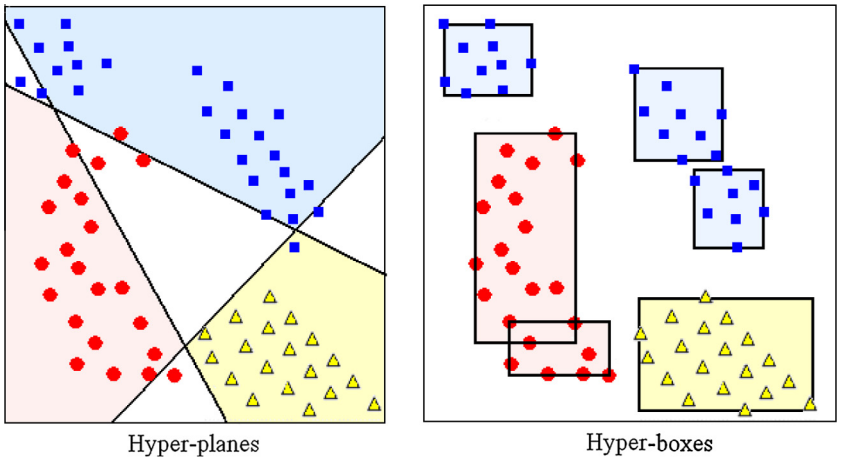

Fig. 1. Linear classifiers: hyper-boxes have more flexibility for estimating the ideal class boundaries of a pattern in comparison to rigid hyper-planes.