Neural Networks 71 (2015) 45–54

Contents lists available at ScienceDirect

Neural Networks

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

A novel multivariate performance optimization method based on sparse coding and hyper-predictor learning

Jiachen Yang [a], Zhiyong Ding [a], Fei Guo [a,*], Huogen Wang [a], Nick Hughes [b]

[a] School of Electronic Information Engineering, Tianjin University, Tianjin 300072, China

[b] College of Information Science and Technology, University of Nebraska Omaha, Omaha, NE 68182, United States

ARTICLE INFO

Article history:
Received 25 March 2015
Received in revised form 5 June 2015
Accepted 23 July 2015
Available online 4 August 2015

Keywords:
Pattern classification
Loss function
Multivariate performance measures
Sparse coding
Joint learning
Alternate optimization

ABSTRACT

In this paper, we investigate the problem of optimization of multivariate performance measures, and propose a novel algorithm for it. Different from traditional machine learning methods which optimize simple loss functions to learn prediction function, the problem studied in this paper is how to learn effective hyper-predictor for a tuple of data points, so that a complex loss function corresponding to a multivariate performance measure can be minimized. We propose to present the tuple of data points to a tuple of sparse codes via a dictionary, and then apply a linear function to compare a sparse code against a given candidate class label. To learn the dictionary, sparse codes, and parameter of the linear function, we propose a joint optimization problem. In this problem, the both the reconstruction error and sparsity of sparse code, and the upper bound of the complex loss function are minimized. Moreover, the upper bound of the loss function is approximated by the sparse codes and the linear function parameter. To optimize this problem, we develop an iterative algorithm based on descent gradient methods to learn the sparse codes and hyper-predictor parameter alternately. Experiment results on some benchmark data sets show the advantage of the proposed methods over other state-of-the-art algorithms.

1. Introduction

In traditional machine learning methods, we usually use a loss function to compare the true class label of a data point against its predicted class label. By optimizing the loss functions over all the training set, we seek an optimal prediction function, named a classifier (Bhuyan & Gao, 2011; Kang & Choi, 2014; Micheloni, Rani, Kumar, & Foresti, 2012; Roy, Mackin, & Mukhopadhyay, 2013; Wang, Bensmail, & Gao, 2012). For example, in support vector machine (SVM), a hinge loss function is minimized, and in linear regression (LR), a logistic loss function is used (Couellan, Jan, Jorquera, & Georgé, 2015; Patil, Fatangare, & Kulkarni, 2015; Pragidis, Gogas, Plakandaras, & Papadimitriou, 2015; Şiray, Toker, & Kaçiranlar, 2015). However, when we evaluate the performance of a class label predictor, we usually consider a tuple of data points, and use a complex multivariate performance measure over the considered tuple of data points, which is different from the loss functions used in the training procedure significantly (Joachims, 2005; Mao & Tsang, 2013; Walker et al., 2011; Zhang, Saha, & Vishwanathan, 2011, 2012). For example, we may use area under receiver operating characteristic curve (AUC) as a multivariate performance measure to evaluate the classification performance of SVM. Because SVM class label predictor is trained by minimizing the loss functions over training data points, it cannot be guaranteed to minimize the loss function corresponding to AUC. Many other multivariate performance measures are also defined to compare a true class label tuple of a data point tuple against its predicted class label tuple, and they can also be used for different machine learning applications. Some examples of the multivariate performance measures are F-score (Gao et al., 2014; Zemmoudj, Kemmouche, & Chibani, 2014), precision–recall curve eleven point (PRBEP) (Boyd, Eng, & Page, 2013; Lopes & Bontempi, 2014), and Matthews correlation coefficient (MCC) (Kumari, Nath, & Chaube, 2015; Shepperd, 2015). To seek the optimal multivariate performance measures on a given tuple of data points, recently, the problem of multivariate performance measure optimization is proposed. This problem is defined as a problem of learning a hyper-predictor for a tuple of data points to predict a tuple of class labels. The hyper-predictor is learned so that a multivariate performance measure used to compare the true class label tuple and the predicted class label tuple can be optimized directly.

* Corresponding author. Tel.: +86 13821820218.
E-mail addresses: gfjy001@yahoo.com (F. Guo), Nick.Hughes1@yahoo.com (N. Hughes).

http://dx.doi.org/10.1016/j.neunet.2015.07.011
0893-6080/

1.1. Related works

Some methods have been proposed to solve the problem of multivariate performance measures. For example,

- Joachims (2005) proposed a SVM method to optimize multivariate nonlinear performance measures, including F-score, AUC etc. This method takes a multivariate predictor, and gives an algorithm to train the multivariate SVM in polynomial time for large classes so that the potentially non-linear performance measures can be optimized. Moreover, the translational SVM with hinge loss function can be treated as a special case of this method.
- Zhang et al. (2011) proposed a smoothing strategy for multivariate performance score optimization, in particular PRBEP and AUC. The proposed method combines Nesterov's accelerated gradient algorithm and the smoothing strategy, and obtains an optimization algorithm. This algorithm converges to a given accurate solution in a limited number of iterations corresponding to the accurate.
- Mao and Tsang (2013) proposed a generalized sparse regularizer for multivariate performance measure optimization. Based on this regularizer, a unified feature selection and general loss function optimization are developed. The formulation of the problem is solved by a two-layer cutting plane algorithm, and the convergence is presented. Moreover, it can also be used to optimize the multivariate measures of multiple-instance learning problems.
- Li, Tsang, and Zhou (2013) proposed to learn a nonlinear classifier for optimization of nonlinear and nonsmooth performance measures by novel two-step approach. Firstly, a nonlinear auxiliary classifier with existing learning methods is trained, and then it is adapted for specific performance measures. The classifier adaptation can be reduced to a quadratic program problem, similar to the method introduced in Joachims (2005).

1.2. Contributions

In this paper, we try to investigate the usage of sparse coding in the problem of multivariate performance optimization. Sparse coding is an important and popular data representation method, and it represents a given data point by reconstructing it with regard to a dictionary (Al-Shedivat, Wang, Alzahrani, Huang, & Gao, 2014; Li, 2015; Wang, Bensmail, & Gao, 2013; Wang & Gao, 2014). The reconstruction coefficients are imposed to be sparse, and used as a new representation of the data point. Sparse coding has been used widely in both machine learning and computer vision communities for pattern classification problems. For example, Mairal, Bach, Ponce, Sapiro, and Zisserman (2009) proposed to learn the sparse codes and a classifier jointly on a training set. However, the loss function used in this method is a traditional logistic loss. In this paper, we ask the following question: how can we learn the sparse codes and its corresponding class prediction function to optimize a multivariate performance measure? To answer this question, we propose a novel multivariate performance optimization method. In this method, we try to learn sparse codes from the tuple of training data points, and apply a linear function to match the sparse code tuple against a candidate class label. Based on the linear function, we design a hyper-predictor to predict the optimal class label tuple. Moreover, the loss function of the desired multivariate performance measure is used to compare the prediction of the hyper-predictor and the true class label tuple, and minimized to optimize the multivariate performance measure. The contributions of this paper are of two fold:

1. We proposed a joint model of sparse coding and multivariate performance measure optimization. We learn both the sparse codes and the hyper-predictor to optimize the desired multivariate performance measure. The input of the hyper-prediction function is the tuple of the sparse codes, and the output is a class label tuple, which is further compared to the true class label tuple by a multivariate performance measure. A joint optimization problem is constructed for this problem. In the objective function of the optimization problem, both the reconstruction error and the sparsity of the sparse code are considered. Simultaneously, the multivariate loss function of the multivariate performance function is also included in the objective. The multivariate loss function may be very complex, and even does not have a close form, thus it is difficult to optimize it directly. We seek its upper bound, and approximate it as a linear function of the hyper-predictor function.
2. We proposed a novel iterative algorithm to optimize the proposed problem. We adapt the alternate optimization strategy, and optimize the sparse code, dictionary and the hyper-predictor function alternately in an iterative algorithm. Both sparse codes and hyper-predictor parameters are learned by gradient descent methods, and the dictionary is learned by Lagrange multiplier method.

1.3. Paper organization

This paper is organized as follows. In Section 2, we introduce the proposed multivariate performance measure optimization method. In Section 3, the proposed method is evaluated experimentally and compared to state-of-the-art multivariate performance measure optimization methods. In Section 4, the paper is concluded with future works.

2. Proposed method

In this section, we introduce the proposed method. We first model the problem with an optimization problem, then solve it with an iterative optimization strategy, and finally develop an iterative algorithm based on the optimization results.

2.1. Problem formulation

Suppose we have a tuple of n training data points, $\bar{\mathbf{x}} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$, and its corresponding class label tuple is denoted as $\bar{y} = (y_1, \ldots, y_n)$, where $\mathbf{x}_i \in \mathbb{R}^d$ is the d-dimensional feature vector of the ith training data point, and $y_i \in \{+1, -1\}$ is the binary label of the ith training data point. We can use a machine learning method to predict the class label tuple, $\bar{y}^* = (y_1^*, \ldots, y_n^*)$, where y_i^* is the predicted class label of the ith data point. A multivariate performance measure, $\Delta(\bar{y}, \bar{y}^*)$, is defined to compare a predicted class label tuple $\bar{y}^*$ of a data point tuple against its true class label tuple $\bar{y}$. To learn a hyper-predictor to map a data point tuple $\bar{\mathbf{x}}$ to an optimal class label tuple $\bar{y}^*$, we should learn it to minimize a desired pre-defined multivariate performance measure, $\Delta(\bar{y}, \bar{y}^*)$. The proposed learning framework is shown in the flowchart in Fig. 1.

We propose to present the data points to their sparse codes by sparse coding method, and then use a linear hyper-predictor to predict the class label tuple. We consider the following problems in the learning procedure,

- **Sparse coding of data tuple**: To represent the data points in the data tuple, we propose to reconstruct each data point in the data tuple by using a dictionary,

$$\mathbf{x}_i \approx \sum_{j=1}^{m} s_{ij}\mathbf{d}_j = D\mathbf{s}_i, \quad i = 1, \ldots, n, \tag{1}$$

where $\mathbf{d}_j \in \mathbb{R}^d$ is the jth dictionary element of the dictionary, and $D = [\mathbf{d}_1, \ldots, \mathbf{d}_m]$ is the dictionary matrix with its jth