Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 108C (2017) 1261–1270

Procedia Computer Science

International Conference on Computational Science, ICCS 2017, 12-14 June 2017, Zurich, Switzerland

Large-scale Nonparallel Support Vector Ordinal Regression Solver

Huadong Wang[1,2,3], Jianyu Miao[1,2,3], Seyed Mojtaba Hosseini Bamakan[1,2,3], Lingfeng Niu[1,2,3], and Yong Shi[1,2,4]

[1] Research Center on Fictitious Economy & Data Science, Chinese Academy of Sciences, Beijing 100190, China
[2] Key Laboratory of Big Data Mining and Knowledge Management, Chinese Academy of Sciences, Beijing 100190, China
[3] School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing, 100049, China.
[4] College of Information Science and Technology, University of Nebraska at Omaha, Omaha, NE 68118, USA
{wanghaudong14,miaojianyu15}@mails.ucas.ac.cn, s_mojtabahossini@yahoo.com, {niulf,yshi}@ucas.ac.cn

Abstract

Large-scale linear classification is widely used in many areas. Although SVM-based models for ordinal regression problem are proven to be powerful techniques, the performance with nonlinear kernels are often suffering from time consuming. Recently, linear SVC not only is shown to obtain competitive performance in most of the cases, but also it is considerably fast during the process of training and testing. However, few studies focused on linear SVM-based ordinal regression models. In this paper, we propose a new approach, called linear Nonparallel Support Vector Ordinal Regression (NPSVOR), which can deal with large-scale problems. An efficient algorithm based on Alternating Direction Method of Multipliers (ADMM) is designed to solve the proposed model. Our experiments are performed on large document data sets to demonstrate the effectiveness of the proposed method.

Peer-review under responsibility of the scientific committee of the International Conference on Computational Science

Keywords: Ordinal Regression, Support Vector Machine, ADMM

1 Introduction

Ordinal regression (OR) is a supervised learning problem where each instance is associated with an ordinal label. Since labels of OR can also be called ranks, it sometimes refers to ranking learning. This situation is ubiquitous in the field of text mining. A notable example is the sentiment analysis, where each text is assigned with a label from some levels, i.e.{ *very negative, negative, neutral, positive, very positive*}. Obviously, an order relation exists among these labels. For Amazon website and other online retailers, consumers usually give their ratings and comments after they bought the products. The rating scheme generally includes five levels, which are represented by different number of stars. One star means the most dissatisfaction, five stars means the most satisfaction, and so on. These comments are very useful for mining the quality and consumer preferences about the products.

1877-0509
Peer-review under responsibility of the scientific committee of the International Conference on Computational Science
10.1016/j.procs.2017.05.104

Gutiérrez et al. [9] carried out a thorough experimental study and compared sixteen state-of-the-arts approaches for ordinal regression. The numerical results demonstrate that SVM-based methods, such as SVOR [3], RedSVM [13], OPBE [15] and SVMOP [17], have shown promising results and achieved relatively better performance. However, these approaches are nonlinear models, which map the data points into a high-dimensional space via a nonlinear feature mapping. Due to the high dimensionality, the computation of the inner product between vectors in the high-dimensional space is memory and time consuming. Therefore Nonlinear SVM-based methods are only suitable for small-scale and low-dimensional OR problems. In some applications, data appears in a high-dimensional feature space, such as word frequency vectors or TF-IDF features extracted from text data by using n-gram words. In this case, the performances are similar with/without nonlinear mapping. So, to reduce the cost of time and memory, linear SVM-based methods are preferred to deal with large-scale data sets. In this paper, we aim to solve large-scale OR problems. Lin et al. [11] studied the algorithms of the standard linear SVC and SVR, and proposed several large-scale algorithms, such as dual coordinate descent algorithm (DCD) and trust region algorithm. Although Lin's algorithms have been widely used in text mining, these algorithms are mainly for general multi-class classification and numerical regression.

In [19], we have proposed a model called Nonparallel Support Vector Ordinal Regression (NPSVOR), which is solved by the Alternating Direction Method of Multipliers (ADMM) in nonlinear case. Numerical experiments have shown that NPSVOR is superior to the other SVM-based methods. Therefore, it is necessary to further study its linear model so that it can be well applied to text mining. Inspired by the good performance of NPSVOR in nonlinear case, we propose NPSVOR in linear case for large-scale problems in this paper. To the best of our knowledge, it is the first work considering large-scale algorithm in ordinal regression (OR) problems. We design a efficient algorithm based on ADMM [2] to solve NPSVOR in linear case. Numeral experiments reveal that our model has a better performance compared with the existing state-of-the-arts methods.

The rest of this paper is organized as follows: In Section 2, we describe the nonparallel support vector ordinal regression model. In Section 3, ADMM algorithm and the preconditioned conjugate gradient procedure will be introduced for each quadratic optimization in NPSVOR. We conduct experiments in Section 4 and the corresponding results are studied and discussed. Some conclusions are given in Section 5.

2 Nonparallel Support Vector Ordinal Regression

In ordinal regression, each sample in the training data set is composed of an input vector and an ordinal label (i.e., rank). Suppose there are p different ordered categories, without loss of generality, we use consecutive integers $1, 2, \cdots, p$ to denote the ranks in this paper. Let n be the number of samples. Then the training set can be represented in the following way

$$\mathcal{S} = \{(\boldsymbol{x}_i, y_i)\}_{i=1,\cdots,n}$$

where $\boldsymbol{x}_i \in \Re^n$ is the input vector and $y_i \in \{1, 2, \cdots, p\}$ is the label of $\boldsymbol{x}_i$. Given $k \in \{1, 2, \cdots, p\}$, we define three index sets for each rank k,

$$\mathcal{L}_k = \{i | y_i < k\},\ \mathcal{I}_k = \{i | y_i = k\} \text{ and } \mathcal{R}_k = \{i | y_i > k\}.$$