Information Sciences xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Information Sciences

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

Multi-label Lagrangian support vector machine with random block coordinate descent method

Q1 Jianhua Xu*

School of Computer Science and Technology, Nanjing Normal University, Nanjing, Jiangsu 210023, China

ARTICLE INFO

Article history:
Received 27 November 2014
Revised 15 August 2015
Accepted 11 September 2015
Available online xxx

Keywords:
Multi-label classification
Support vector machine
Kernel function
Block coordinate descent method
Quadratic programming

ABSTRACT

When all training instances and labels are considered all together in a single optimization problem, multi-label support and core vector machines (i.e., Rank-SVM and Rank-CVM) are formulated as quadratic programming (QP) problems with equality and bounded constraints, whose training procedures have a sub-linear convergence rate. Therefore it is highly desirable to design and implement a novel efficient SVM-type multi-label algorithm. In this paper, through applying pairwise constraints between relevant and irrelevant labels, and defining an approximate ranking loss, we generalize binary Lagrangian support vector machine (LSVM) to construct its multi-label form (Rank-LSVM), resulting into a strictly convex QP problem with non-negative constraints only. Particularly, each training instance is associated with a block of variables and all variables are divided naturally into manageable blocks. Consequently we build an efficient training procedure for Rank-LSVM using random block coordinate descent method with a linear convergence rate. Moreover a heuristic strategy is applied to reduce the number of support vectors. Experimental results on twelve data sets demonstrate that our method works better according to five performance measures, and averagely runs 15 and 107 times faster and has 9 and 15% fewer support vectors, compared with Rank-CVM and Rank-SVM.

1. Introduction

Multi-label classification is a special supervised learning task, where any single instance possibly belongs to several classes simultaneously, and thus the classes are not mutually exclusive [2,36,37,48]. In the past ten years, such a classification issue has received a lot of attention because of many real world applications, e.g., text categorization [29,46], scene annotation [1,47], bioinformatics [43,46], and music emotion categorization [33]. Currently, there mainly are four strategies to design various multi-label classification algorithms: data decomposition, algorithm extension, hybrid, and ensemble strategies.

Data decomposition strategy divides a multi-label data set into either one or more single-label (binary or multi-class) subsets, learns a sub-classifier for each subset using an existing classifier, and then integrates all sub-classifiers into an entire multi-label classifier. There exist two popular decomposition ways: one-versus-rest (OVR) or binary relevance (BR), and label powerset (LP) or label combination (CM) [2,36,37,48]. It is rapid to construct a data decomposition method since many popular single-label classifiers and their free software are available. But the label correlations are not characterized explicitly in OVR-type methods, and lots of new classes with a few training instances and no new predicting label combination are created in LP-type methods.

* Tel.: +86 25 96306209; fax: +86 25 85891990.
E-mail address: xujianhua@njnu.edu.cn, xujianhua99@tsinghua.org.cn

http://dx.doi.org/10.1016/j.ins.2015.09.023
0020-0255/© 2015 Published by Elsevier Inc.

Algorithm extension strategy generalizes some specific multi-class algorithm to consider all training instances and classes (or labels) of training set all together. But this strategy possibly induces some complicated optimization problems, e.g., a large-scale unconstrained problem in multi-label back-propagation neural networks (BP-MLL) [46] and two large-scale quadratic programming (QP) ones in multi-label support and core vector machines (Rank-SVM [9] and Rank-CVM [43]). Inspiringly, the label correlations of individual instance are depicted sufficiently via pairwise relations between relevant labels and irrelevant ones.

Hybrid strategy not only extends or modifies an existing single-label method but also splits a multi-label data set into a series of subsets implicitly or explicitly. After the OVR trick is embedded, the famous k-nearest neighbor algorithm (kNN) is cascaded with discrete Bayesian rule in ML-kNN [47], logistic regression in IBLR-ML [5], fuzzy similarity in FSKNN [19] and case-based reasoning in MlCBR [25]. After executing feature extraction with principal component analysis and feature selection with genetic algorithm, multi-label naive Bayes (MLNB) utilizes OVR trick to estimate prior and conditional probabilities for each label [45]. Besides a relatively low computational cost, such a strategy weakly characterizes the label correlations either explicitly or implicitly.

Ensemble strategy either generalizes an existing multi-class ensemble classifier, or realizes a new ensemble of the aforementioned three kinds of multi-label techniques. Two boosting-type multi-label classifiers (Adaboost.MH and Adaboost.MR) are derived from famous Adaboosting method [29] via minimizing Hamming loss and ranking one, respectively. Random k-labelsets (RAkEL) method splits an entire label set into several subsets of size k, trains LP classifiers and then constructs an ensemble multi-label algorithm [38]. Classifier chair (CC) builds an OVR classifier in a cascade way rather than a parallel one, and then its ensemble form (ECC) alleviates the possible effect of classifier order [27]. Variable pairwise constraint projection for multi-label ensemble (VPCME) combines feature extraction based on variable pairwise constraint projection with boosting-type ensemble strategy [20]. In [21], random forest of predictive cluster trees (RF-PCT) is strongly recommended due to its good performance in an extensive experimental comparison, including ECC, RAkEL, ML-kNN and etc. Generally, these ensemble methods spend more training and testing time to improve their classification performance.

Now it is widely recognized that algorithm extension strategy considers as many label correlations as possible, which is an optimal way to enhance multi-label classification performance further [6]. But, its corresponding methods have a relatively high computational cost, which limits their usability for many real world applications. Consequently, it is still imperative to design and implement some novel efficient multi-label classifiers. In this paper, we focus on SVM-type multi-label techniques.

Generally, SVM-type multi-label methods are formulated as QP problems with some different constraint conditions which further are associated with distinct optimization procedures. Rank-SVM [9] includes several equality and many box constraints. When Rank-SVM is solved by Frank-Wolfe method (FWM) [10,15], a large-scale linear programming is dealt with at each iteration. As a variant of Rank-SVM , Rank-CVM [43] involves a unit simplex and many non-negative constraints. Because of its special constraints, at each iteration in FWM, there exist a closed solution and several efficient recursive formulae for Rank-CVM. Although FWM has a sub-linear convergence rate, the experimental results demonstrate that Rank-CVM has a lower computational cost than Rank-SVM. In [44], random block coordinate descent method (RBCDM) with a sub-linear convergence rate is used for Rank-SVM, where at each iteration a small-scale QP sub-problem with equality and box constraints is still solved by FWM due to equality constraint limit. The experimental results show that RCBDM runs three times faster than FWM for Rank-SVM. The success of these methods inspires us to design a special QP problem with some particular constraints and then to construct its efficient optimization solution procedure.

In the past twenty years, it is widely accepted that binary SVM [39] is one of the most successful classification techniques [42]. The success of SVM attracts many researchers to develop its variants with special optimization forms to reduce as many computational costs as possible, e.g., Lagrangian support vector machine (LSVM) [22], proximal SVM (PSVM)[11] and its privacy preserving version (P3SVM) [31], twin SVM (TWSVM) [18] and its sparse and Laplacian smooth forms [4,26], support tensor machine (STM) [32] and its multiple rank multi-linear kernel version [13].

Particularly, binary LSVM [22] is formulated as a QP problem with non-negative constraints only, and then is solved by an iterative procedure with inverse matrix, which has a linear convergence rate. In this paper, we generalize LSVM to construct its multi-label version: Rank-LSVM, which has the same non-negative constraints as those in LSVM and the same number of variables to be solved as that in Rank-SVM and Rank-CVM. Since each training instance is associated with a block of variables and thus all variables are split naturally into many manageable blocks in Rank-LSVM, we implement an efficient training procedure using random block coordinate descent method (RBCDM) with a linear convergence rate [24,28]. Further, to reduce the number of support vectors to speed up the training and testing procedures, we apply a heuristic strategy to avoid updating some training instances which result in no ranking loss. Experimental results on 12 benchmark data sets illustrate that our method is a competitive candidate for multi-label classification according to five performance measures, compared with four existing techniques: Rank-CVM [43], Rank-SVM [9], BP-MLL [46], and RF-PCT [21]. Moreover, our Rank-CVM runs averagely 15 and 107 times faster, has 17 and 9% fewer support vectors than Rank-SVM and Rank-CVM in the training phase.

The rest of this paper is organized as follows. Multi-label classification setting is introduced in Section 2. In Sections 3 and 4, our Rank-LSVM is proposed after binary LSVM is reviewed briefly and then an efficient training algorithm is constructed and analyzed. In Section 5, we formally analyze the relation between our Rank-LSVM and two previous SVM-type algorithms (i.e., Rank-SVM and Rank-CVM). Section 6 is devoted to experiments with 12 benchmark data sets. This paper ends with some conclusions finally.