Designing a Multi-label Kernel Machine with Two-Objective Optimization

Hua Xu and Jianhua Xu

School of Computer Sceince and Technology
Nanjing Normal University
Nanjing, Jiangsu 210097, China
xuhuayg@163.com, xujianhua@njnu.edu.cn

Abstract. In multi-label classification problems, some samples belong to multiple classes simultaneously and thus the classes are not mutually exclusive. How to characterize this kind of correlations between labels has been a key issue for designing a new multi-label classification approach. In this paper, we define two objective functions, i.e., the number of relevant and irrelevant label pairs which are ranked incorrectly, and the model regularization term, which depict the correlations between labels and the model complexity respectively. Then a new kernel machine for multi-label classification is constructed using two-objective minimization and solved by fast and elitist multi-objective genetic algorithm, i.e., NSGA-II. Experiments on the benchmark data set Yeast illustrate that our multi-label method is a competitive candidate for multi-label classification, compared with several state-of-the-art methods.

Keywords: multi-label classification, kernel, regularization, multi-objective optimization, NSGA-II.

1 Introduction

Multi-label classification is a particular learning task, where some samples are associated with multiple classes at the same time, and thus the classes are not mutually exclusive. Currently there mainly exist two types of discriminative methods for multi-label classification: problem transform and algorithm adaptation [1]. The former converts training data sets to fit a large number of existing classification techniques, while the latter extends some existing multi-class algorithms to satisfy multi-label training data sets.

Problem transform methods split a multi-label training set into either one or more single label (binary or multi-class) subsets, train a sub-classifier for each subset using some popular approaches, and integrate all sub-classifier into an entire multi-label classifier. Now there are three widely used data decomposition strategies: label powerset (LP), one-versus-rest (OVR), and one-versus-one (OVO). LP strategy considers each possible label combination of more than one class in a multi-label data set as a new single class, and then builds a standard multi-class training data set. Its main disadvantage is that many classes have very few samples [1, 2]. OVR strategy divides a k-label training set into k binary subsets, in which the positive and negative

F.L. Wang et al. (Eds.): AICI 2010, Part I, LNAI 6319, pp. 282–291, 2010.

samples in the *i*th subset come from the *i*th class and all other classes respectively. This strategy has been validated using many popular binary methods, e.g., support vector machines (SVM) [1-3], C4.5 [1, 4], naïve Bayes [1, 5], and *k*NN [1, 4]. OVO strategy splits a *k*-label set into $k(k\text{-}1)/2$ subsets, which results into some special subsets including a small proportion of double label samples. It seems reasonable to locate these double label samples between positive and negative samples. Two parallel hyper-planes are trained to separate three classes in [6], whereas in [7] these double label samples are forced to reside at the marginal region of binary SVM. The main limitation of problem transform methods is that they do not directly take the correlations between labels into account.

In contrast to problem transform methods, algorithm adaptation methods extend some specific multi-class approaches to deal with entire multi-label training sets directly, usually resulting into some optimization problems with very high computational complexity. Their main advantage is that the correlations between labels are considered explicitly. So far, many famous multi-class classifiers have been generalized to build their corresponding multi-label versions.

Through modifying the formula of entropy calculation and permitting multiple labels at the leave of the tree, a C4.5-type multi-label algorithm was proposed in [8]. BoosTexter [9] was derived from the well-known Adaboost algorithm, which include two slightly different versions: AdaBoost.MH and Adaboost.MR. The former is to predict the relevant set of labels of a sample, whereas the latter to rank labels of a sample in descending order. ADTboost [10] combines alternative tree with Adaboost.MH. It is hard to control the complexity of such boosting or tree based multi-label methods [3].

Nearest neighbor (*k*NN) or instance-based (IB) method has been extended to build two slightly different algorithms: ML-*k*NN [11, 12] and IBLR-ML [4], cascading standard *k*NN and Bayesian inference. They utilize different ways to estimate the posterior probability for each label, in which ML-*k*NN uses Bayesian rule and IBLR-ML logistic regression.

In principle, traditional back-propagation (BP) neural networks can deal with multi-label classification directly via assigning many ones at their output layer. To improve the performance of BP, its multi-label version (BP-MLL) [13] utilizes a new empirical loss function derived from the ranking loss and a regularization term. Rank-SVM [3] is a typical multi-label support vector machine, whose empirical loss function also is based on the ranking loss. It is noted that for BP-MLL and Rank-SVM, their objective functions combine an empirical loss function with a regularization term through weighted sum method. This is a classical solution method for multi-objective optimization [14, 15].

In this paper, we regard multi-label classification as a multi-objective optimization problem rather than a single-objective one. On the basis of kernel based decision functions, two objective functions: the number of relevant and irrelevant label pairs which are ranked incorrectly, and the model regularization term, are minimized at the same time using fast and elitist multi-objective genetic algorithm, i.e., NSGA-II [16, 17]. This setting can avoid determining a regularization constant in advance. Experimental results demonstrate that our new method is a highly competitive candidate for multi-label classification, compared with many state-of-the-art methods.