Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 91 (2016) 556 – 565

Procedia Computer Science

Information Technology and Quantitative Management (ITQM 2016)

Multi-metrics classification machine

Dewei Li[a,b], Wei Zhang[c], Dongkuan Xu[a,b], Yingjie Tian[b,d,*]

[a] School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China
[b] Research Center on Fictitious Economy and Data Science, Chinese Academy of Sciences, Beijing 100190, China
[c] School of Information, Renmin University of China, Beijing 100872, China
[d] Key Laboratory of Big Data Mining and Knowledge management, Chinese Academy of Sciences, Beijing 100190, China

Abstract

Distance and its related decision rules are important in classification problems. *k*NN classifies a data point by the labels of its *k*-nearest neighbors and can be ameliorated by metric learning. For SVM, representative hyperplanes are found to refer to the location of every class and any point can be labeled by its perpendicular distance from the hyperplanes. Inspired by metric learning and SVM, a multi-metric classification machine, called MMCM, with a new prediction mechanism is proposed based on a novel distance relationship discerned by multi-metrics learning of the specificity information of each class. MMCM aims to find multi-metrics, namely the multiple local linear transformations for each class, to map data points into a new feature space, in which the distance between a point and its corresponding class centroid is minimized and data points of other classes are far from the centroid. An example with unknown label is classified according to the label of its nearest centroid. The primary problem is slacked as a linear optimization problem and kernel is introduced to make a nonlinear transformation. Enormous experiments verify MMCM's competitive performance both on binary classification and multi-class classification compared to state-of-the-art classification methods.

Peer-review under responsibility of the Organizing Committee of ITQM 2016

Keywords: metric learning; classification; distance; multi-metrics; nonparallel SVM

1. Introduction

Classification task is one of the most widely studied problems, which aims to predict the labels of unknown patterns based on knowledge extracted from datasets[1, 2]. *k*-Nearest Neighbor(*k*NN)[3] and support vector machine(SVM) are two of the most classical methods in dealing with classification learning. The so-called knowledge is derived from datasets that consist of different classes distributing in different locations and forming various statistical structures. An unlabeled pattern is classified by its location relative to the location of global class or local region. This kind of relativity is often defined by distance, calculated in multifarious ways in the models of *k*NN and SVM. In *k*NN, the distance is only considered in neighbourhood level, on which the label of an example is

*Corresponding author. Tel.: +86 10 82680997 ; fax: +86 10 82680997 .
E-mail address: lidewei15@mails.ucas.ac.cn(Dewei Li); zw-info@ruc.edu.cn(Wei Zhang); xudongkuan14@mails.ucas.ac.cn(Dongkuan Xu); tyj@ucas.ac.cn(Yingjie Tian).

Peer-review under responsibility of the Organizing Committee of ITQM 2016
doi:10.1016/j.procs.2016.07.142

determined by the labels of k nearest neighbors. But in most implementations, kNN simply takes advantage of Euclidean distance, treating all the classes and features indiscriminately. To extract more potential information from attributes, the research on metric learning is extensively developed to find a data-dependent metric to compute distance in a more reasonable way[4, 5, 6, 7, 8]. The purpose of metric learning is to learn a proper metric M, computing the distance by $d_M(x_i, x_j) = (x_i - x_j)^\top M(x_i - x_j)$, to increase inter-class distance and decrease intra-class distance as much as possible. Popular methods include metric learning with side information[9], with large margin nearest neighbor(LMNN)[10], with information-theory[11], using boosting-like technique[12], by collapsing classes[13], and neighbourhood component analysis[14], large margin component analysis[15], sparse metric learning[16, 17, 18], SVM related metric learning[19, 20, 21, 22], learning metric from network[23, 24]. Metric learning has been applied to many applications, such as face identification[25], image annotation[26], and text classification[27].

But all the above methods are learning a global unique metric without considering the class related information. A appropriate metric should satisfy the property of semi-definite positive, which can make the target metric M be decomposed as $L^\top L$. The new defined distance in the original space is actually Euclidean distance after the linear transformation Lx. So a global metric is accompied with a single global linear transformation. A naturally extension for metric learning is to learn multi-metrics, i.e., each class corresponding to a local linear transformation. This idea takes local data structure into account to mine more statistical information. There are few research on metric learning with multi-metrics up to now. Multi-metric LMNN(mmLMNN)[10] is proposed to define class-dependent distance to realize the mechanism of multiple local linear transformations. It shows better performance in improving the accuracy of kNN classification than the original LMNN.

In traditional SVM, two parallel support hyperplanes are sought to meet the principle of margin maximization[28, 29, 30, 31, 32]. SVM looks for boundary-like hyperplanes to discriminate the location of every class. The label of an unlabeled example is determined by its perpendicular distance to the boundaries. The distance is considered in global level, in contrast to the local level in kNN. In fact, margin maximization does not attend to class specificity, but only feature information, since the two hyperplanes are parallel to meet global distribution. However, nonparallel SVM(NPSVM)[33, 34, 35, 36] for binary classification has made improvements on the issue, which finds two nonparallel hyperplanes to make each hyperplane proximal to its corresponding class and far from the other class as much as possible. The nonparallel hyperplanes can be regarded as two 'centroid' lines, the functions of which are similar as multiple metrics(or multiple linear local linear transformations) in metric learning, to indicate the locations of classes. The better classification performance of NPSVM compared to traditional SVM verifies that the information of class uniqueness is beneficial to improve classification accuracy.

In this paper, we propose a new metric learning approach called MMCM, multi-metrics classification machine. Each metric corresponds to a class. We establish one optimization problem for each class to learn multiple metrics independently. The desired metric is to minimize within-class distance by enforcing the distance between the points and their class centroid as small as possible, with constraints that points of other classes should be far from the centroid unit distance away. Every metric can be decomposed with respect to linear transformation due to its property of semi-definite positive. The optimization problems can be constructed in terms of linear transformation and solved in linear programming after being slacked. Extensive experiments demonstrate MMCM's high performance in both binary and multi-class classification.

We organize our paper as follows. In section 2, multi-metrics related methods, NPSVM, and multi-metric LMNN are introduced. The relation between metric learning and NPSVM is expounded in Section 3. Section 4 proposes and discusses MMCM in detail. Experimental results and conclusions are summarized in Section 5 and 6 respectively.

2. Background

For a classification problem:

$$T = \{(x_1, y_1), (x_2, y_2), \cdots, (x_m, y_m)\}, \tag{1}$$

where $x_i \in R^n, y_i \in \{1, 2, \cdots, c\}, i = 1, \cdots, m(c \geq 2)$. Define $A_k = (x_{k_1}, x_{k_2}, \cdots, x_{k_m}), k = 1, 2, \cdots, c$, and $m = \sum_{k=1}^{c} k_m$.