Chapter 2
Kernel Parameter Selection for SVM Classification: Adaboost Approach

Manju Bala
Jawaharlal Nehru University, India

R. K. Agrawal
Jawaharlal Nehru University, India

ABSTRACT

The choice of kernel function and its parameter is very important for better performance of support vector machine. In this chapter, the authors proposed few new kernel functions which satisfy the Mercer's conditions and a robust algorithm to automatically determine the suitable kernel function and its parameters based on AdaBoost to improve the performance of support vector machine. The performance of proposed algorithm is evaluated on several benchmark datasets from UCI repository. The experimental results for different datasets show that the Gaussian kernel is not always the best choice to achieve high generalization of support vector machine classifier. However, with the proper choice of kernel function and its parameters using proposed algorithm, it is possible to achieve maximum classification accuracy for all datasets.

INTRODUCTION

In the past two decades valuable work has been carried out in the area of text categorization (Joachims, Thorsten, 1998; Thorsten et al., 2001), optical character recognition (Mori et al., 1992), intrusion detection (Mukkamala et al., 2002), speech recognition (Schmidt, 1996), handwritten digit recognition (Weston & Watkins, 1999) etc. All such real-world applications are essentially multi-class classification problems. Multi-class classification is intrinsically harder than binary classification problem because the classification has to learn to construct a greater number of separation boundaries or relations. Classification error rate is greater in multi-class problem than that of binary as there can be error in determination of any one of the decision boundaries or relations.

There are basically two types of multi-class classification algorithms. The first type deals directly with multiple values in the target field *i.e.* K- Nearest Neighbor, Naive Bayes, classification trees in the class etc. Intuitively, these methods can

DOI: 10.4018/978-1-61520-741-1.ch002

be interpreted as trying to construct a conditional probability density for each class, then classifying by selecting the class with maximum a posteriori probability. For data with high dimensional input space and very few samples per class, it is very difficult to construct accurate densities. While the other approaches decompose the multi -class problem into a set of binary problems and then combining them to make a final multi-class classifier. This group contains support vector machines, boosting and more generally, any binary classifier. In certain settings the later approach results in better performance then the multiple target approaches.

Support Vector Machine (SVM) is most commonly and popular classification technique used in literature. Support Vector Machines (SVMs) are based on statistical learning theory developed by Vapnik (Vapnik, 1998; Corts & Vapnik, 1995; Corts & Vapnik, 1995; Burges, 1998). It transforms the training vectors into a high-dimensional feature space, labeling each vector by its class. It classifies data by determining a set of support vectors, which are members of the set of training inputs that outline a hyperplane in feature space (Kittler & Hojjatoleslami, 1998). It is based on the idea of Structural risk minimization, which minimizes the generalization error. The number of free parameters used in the SVM depends on the margin that separates the data points and not on the number of input features. SVM provides a generic technique to fit the surface of the hyperplane to the data by employing an appropriate kernel function. Use of a kernel function enables the curse of dimensionality to be addressed, and the solution implicitly contains support vectors that provide a description of the significant data for classification (Scholkopf & smola, 2002).

In literature, the most commonly used kernel functions are linear, polynomial and Gaussian. However, it has been suggested that the choice of these kernels may not be appropriate for many real time datasets. It is shown in literature (Corts & Vapnik, 1995) that the classifier accuracy of SVM is being improved upon with the use of new kernels other than using linear, polynomial and Gaussian. In training a SVM, we need to select an appropriate kernel with its parameter and suitable choice of the margin parameter C for achieving better classification accuracy. The suitable kernel function can be used in SVM provided the distribution of the data is known. However, finding a suitable kernel function and its parameter becomes more challenging when the underlying distribution is unknown. In such situation, there is need to identify kernel function and its parameters for which the classifier provides maximum classification accuracy. Since for multi-class data, the classification is carried out by constructing and combining several binary SVM classifiers, it is possible that single choice of kernel function and its parameter may not provide better classification accuracy. To improve the performance of SVM classifier, there is need to select appropriate kernel function and its parameters for each individual classifier involved in a multi-class classification problem.

We have proposed few new kernel functions which satisfy Mercer's condition and may be more suitable for classification. Also, based on previous work (Tian et al., 2004; Amores et al., 2006; Yu et al., 2006), we propose a new boosted kernel parameters selection framework to automatically learn the best kernel function and its parameters to achieve maximum classification accuracy for a given multi-class classification problem.

SUPPORT VECTOR MACHINE CLASSIFIER

Theory of SVM

This section briefly introduces the theory of SVM. Let $\{(x_1, y_1), \ldots, (x_m, y_m)\} \in \mathrm{R}^n \times \{+1, -1\}$ be a training set. The SVM classifier finds a canonical hyperplane $\{x \in R^n: w^T x + b = 0, w \in \mathcal{R}^n, b \in \mathcal{R}\}$

Related Content

Factors Influencing Satisfaction with Mobile Portals
Daisy Seng, Carla Wilkinand Ly-Fie Sugianto (2011). *Pervasive Computing and Communications Design and Deployment: Technologies, Trends and Applications (pp. 279-295).*
www.irma-international.org/chapter/factors-influencing-satisfaction-mobile-portals/53794

Humans and Emerging RFID Systems: Evaluating Data Protection Law on the User Scenario Basis
Olli Pitkänenand Marketta Niemelä (2010). *Ubiquitous and Pervasive Computing: Concepts, Methodologies, Tools, and Applications (pp. 1396-1407).*
www.irma-international.org/chapter/humans-emerging-rfid-systems/37858

Recent Trends in Pervasive and Ubiquitous Computing: A Survey
Ramesh Singh (2010). *Strategic Pervasive Computing Applications: Emerging Trends (pp. 1-43).*
www.irma-international.org/chapter/recent-trends-pervasive-ubiquitous-computing/41578

Inscribing Interpretive Flexibility of Context Data in Ubiquitous Computing Environments: An Action Research Study of Vertical Standard Development
Magnus Anderssonand Rikard Lindgren (2009). *International Journal of Advanced Pervasive and Ubiquitous Computing (pp. 1-18).*
www.irma-international.org/article/inscribing-interpretive-flexibility-context-data/3864

The Design and Implement of Electrical Operator Monitoring System
Yao Wan-Ye, Sun Teng-Zhongand Jiang Xue-Li (2013). *International Journal of Advanced Pervasive and Ubiquitous Computing (pp. 60-65).*
www.irma-international.org/article/the-design-and-implement-of-electrical-operator-monitoring-system/100439