Kernel Width Selection for SVM Classification: A Meta-Learning Approach

Shawkat Ali, Monash University, Australia
Kate A. Smith, Monash University, Australia

ABSTRACT

The most critical component of kernel-based learning algorithms is the choice of an appropriate kernel and its optimal parameters. In this paper, we propose a rule-based meta-learning approach for automatic radial basis function (RBF) kernel and its parameter selection for Support Vector Machine (SVM) classification. First, the best parameter selection is considered on the basis of prior information of the data with the help of Maximum Likelihood (ML) method and Nelder-Mead (N-M) simplex method. Then, the new rule-based meta-learning approach is constructed and tested on different sizes of 112 datasets with binary class as well as multi-class classification problems. We observe that our rule-based methodology provides significant improvement of computational time as well as accuracy in some specific cases.

Keywords: automatic RBF width selection; Maximum Likelihood; Nelder-Mead; Support Vector Machine; RBF kernel

INTRODUCTION

Support Vector Machines (SVMs) (Boser et al., 1992; Cortes & Vapnik, 1995; Vapnik, 1995) are a relatively new statistical supervised learning method first introduced by Vapnik and his co-worker for binary classification problems. After that, it has been extended to multi-class problems, regression tasks, and novelty detection. These statistical learning algorithms are gaining rapid popularity due to quite a large number of attractive performance results in areas including bioinformatics (Guyon et al. 2002), text mining (Paab et al., 2002), fraud detection (Hyun-Chul et al. 2002), speaker identification (Wan & Renals, 2002), and database marketing (Bennett et al., 1999), among many others.

SVMs adopt the Structural Risk Minimization (SRM) principle, as opposed to the Empirical Risk Minimization (ERM) approach most commonly employed within statistical, neural, and rule-based learning.
methods. This SRM principle has made SVMs an excellent tool for improved generalization. A kernel transforms the data points from input space to higher dimensional feature space by generating the dot product. The feature space theoretically could be of infinite dimension, where linear discrimination is possible by constructing the optimal hyperplane. This is another significant speciality of SVMs compared to other traditional learning algorithms.

The polynomial and radial basis function (RBF) kernels are the most popular classical SVM kernel. According to Ou et al., RBF kernel is more suitable than others SVM kernels (Ou et al., 2003). Hsu et al. suggested that in general RBF is a reasonable first choice for SVM classification. Up to now a good number of kernels have been proposed by researchers, but there is no any unique kernel that performs best for all problems. The performance of the SVM method depends on the suitable selection of a kernel. The most common procedure for SVM best kernel selection is the trial-and-error approach. Joachims argues SVMs are universal learners with a simple “plug-in” of an appropriate kernel function to learn the problems (Joachims, 1998). This is a very lengthy procedure due to a vast range of kernel function available. Onoda et al. argued selection of the suitable kernel for SVM is an important research issue for real world applications (Onoda, et al. 2002). A priori kernel selection for SVM is a difficult task for the user though (Amari, and Wu 1999; Parrado-Hernandez, et al. 2003). Clearly, automatic kernel selection is a key issue for SVM given the number of kernels available rather than the current trial-and-error nature of selecting the best kernel for a given problem. We found in SVM literature (Joachims, 1998; Morik et al., 1999), manually feeding the parametric kernel parameter is a traditional approach for SVM user. The RBF parameter (width) could be selected by optimization approach (Chapelle et al., 2002). Muller et al. (Muller et al., 2001) suggest RBF width could be selected by following cross-validation procedure. This is the most common way of the RBF width selection method. Carlos et al. argue both optimization and cross-validation methods are computationally very expensive and they suggest selecting the RBF width within a range for regression using meta-learning (Carlos et al., 2004). Schölkopf (Schölkopf, 2003) suggested searching the RBF width between 0.2 and 1. Up to still this is the most popular way to feed RBF kernel parameter for SVM. Therefore, it is a research issue how to choose automatically the most suitable RBF kernel function and its optimum width for SVM classification.

Our methodology seeks to understand the characteristics of the data (classification problem), understand RBF kernel perform well on which types of problems, and generate rules to assist in the automatic selection of RBF kernel for SVMs. First we classify a wide range of classification problems with different kernels and then identify the dataset characteristics matrix by statistical measures as we have done in some previous related work (Smith et al. 2002; Smith et al. 2001). We then build models for 112 classification problems (see Appendix A) from the UCI Repository (Blake and Merz, 2002) and Knowledge Discovery Central (Lim, 2002) database using SVM with six different kernels. Finally we use the induction algorithm C5.0 (Windows version See5, http://www.rulequest.com/see5-info.html) to generate the rules to describe RBF kernel is suitable for which type of problem, given the dataset characteristics and the performance of RBF kernel on each dataset. We also examine the rules by 10 Fold Cross
Related Content

A Perturbation Method Based on Singular Value Decomposition and Feature Selection for Privacy Preserving Data Mining

Early Warning System for SMEs as a Financial Risk Detector
[www.irma-international.org/chapter/early-warning-system-smes-financial/7554/](www.irma-international.org/chapter/early-warning-system-smes-financial/7554/)

E-Government Knowledge Management (KM) and Data Mining Challenges: Past, Present, and Future
[www.irma-international.org/chapter/government-knowledge-management-data-mining/29063/](www.irma-international.org/chapter/government-knowledge-management-data-mining/29063/)

Multi-Documents Summarization Based on TextRank and its Application in Online Argumentation Platform
Caiquan Xiong, Xuan Li, Yuan Li and Gang Liu (2018). *International Journal of Data Warehousing and Mining* (pp. 69-89).

GeoCache: A Cache for GML Geographical Data
[www.irma-international.org/article/geocache-cache-gml-geographical-data/1779/](www.irma-international.org/article/geocache-cache-gml-geographical-data/1779/)