A Journal of Software Engineering and Applications, **2012, 5, 55-58**
doi:10.4236/jsea.2012.512b012 Published Online December 2012 (http://www.scirp.org/journal/jsea)

Text Classification Using Support Vector Machine with Mixture of Kernel

Liwei Wei[1], Bo Wei[2], Bin Wang[1]

[1]China National Institute of Standardization, Beijing 100088, China; [2]Department of Ideological and Political Theory Course Teaching Research, Wuhan University of Technology, Wuhan, China.
Email: weilw@cnis.gov.cn, willweixiaobo_1001@163.com, wangbin@cnis.gov.cn

Received 2012.

ABSTRACT

Recent studies have revealed that emerging modern machine learning techniques are advantageous to statistical models for text classification, such as SVM. In this study, we discuss the applications of the support vector machine with mixture of kernel (SVM-MK) to design a text classification system. Differing from the standard SVM, the SVM-MK uses the 1-norm based object function and adopts the convex combinations of single feature basic kernels. Only a linear programming problem needs to be resolved and it greatly reduces the computational costs. More important, it is a transparent model and the optimal feature subset can be obtained automatically. A real Chinese corpus from Fudan University is used to demonstrate the good performance of the SVM- MK.

Keywords: Text Classification; SVM-MK; Feature selection; Classification model; SVM

1. Introduction

With the arrival of "information explosion" era, the infinite growth information resources put forward a great challenge to the information processing. On the one hand, the digital information resources increase at a high speed. On the other hand, People obtain valuable information demand also continuously improve. So, how search out the effective information in the vast and complex information, which has been the goal of information processing field.

Text classification is the research focus of information processing areas. A text file is the object of study in this method. And a lot of files set are mapped to a predefined text attribute class. And its task will be to divide hypertext files into several categories according to predefined contents. Almost all areas are involved in this kind of problems. For example, email filtering, web search, office automation, subject indexing and classification of news stories.

In text classification system, the classifier is a key part, the classifier performance quality directly related to the effect of text classification and efficiency. At present, most of the classifier reference to the methods from information retrieval and machine learning. And almost all the important machine learning algorithms is introduced in text classification. Such as, support vector machine (SVM), least square support vector machine(LS-SVM).

Support vector machine (SVM) is first proposed by Vapnik[1]. Now it has been proved to be a powerful and promising data classification and function estimation tool. In the first part, Joachims introduced SVM method into the text classification [2]. He compared the traditional method to the SVM method in text classification, which shows that SVM in text categorization has the excellent properties than other algorithms. Reference [3] and [4] applied SVM to text classification. They have obtained some valuable results. But SVM is sensitive to outliers and noises in the training sample and has limited interpretability due to its kernel theory. Another problem is that SVM has a high computational complexity because of the solving of large scale quadratic programming in parameter iterative learning procedure. And as we know feature selection is a very import method to the high vector space of text. The last problem of SVM is that SVM can not realize the feature selection.

Motivated by above questions and ideas, we propose a new method named support vector machines with mixture of kernel (SVM-MK) to classify the text. In this method the kernel is a convex combination of many finitely basic kernels. Each basic kernel has a kernel coefficient and is provided with a single feature. The 1-norm is utilized in SVM-MK. As a result, its objective function

turns into a linear programming parameter iterative learning procedure and greatly reduces the computational complexity. Furthermore, we can select the optimal feature subset automatically and get an interpretable model.

The rest of this paper is organized as follows: section 2 gives a brief outline of SVM-MK. To evaluate the performance of SVM-MK for the text classification, we use a real life Chinese corpus from Fudan University in this test in section 3. Finally, section 4 draws the conclusion and gives an outlook of possible future research areas.

2. Support Vector Machine with Mixture of Kernel

Considering a training data set $G=\{(\vec{x}_i, y_i)\}_{i=1}^{n}$, $\vec{x}_i \in R^m$ is the i^{th} input pattern and y_i is its corresponding observed result $y_i \in \{+1,-1\}$. In test classification model, x_i denotes the attributes of text vector, y_i is class label.

The optimal separating hyper-plane is found by solving the following regularized optimization problem [1]:

$$\min J(\vec{\omega},\xi)=\frac{1}{2}\|\omega\|^2+c\sum_{i=1}^{n}\xi_i \tag{1}$$

$$s.t.\begin{cases} y_i\left(\vec{\omega}^T\phi(\vec{x}_i)+b\right)\geq 1-\xi_i & i=1,\cdots n \\ \xi_i\geq 0 \end{cases} \tag{2}$$

\where c is a constant denoting a trade-off between the margin and the sum of total errors. $\phi(\vec{x})$ is a nonlinear function that maps the input space into a higher dimensional feature space. The margin between the two parts is $2/\|\omega\|$.

The quadratic optimization problem can be solved by transforming (1) and (2) into the saddle point of the Lagrange dual function:

$$\max\left\{\sum_{i=1}^{n}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j k(\vec{x}_i,\vec{x}_j)\right\} \tag{3}$$

$$s.t.\begin{cases} \sum_{i=1}^{n} y_i\alpha_i=0 \\ 0\leq\alpha_i\leq c, i=1,\cdots,n \end{cases} \tag{4}$$

where $k(\vec{x}_i,\vec{x}_j)=\phi(\vec{x}_i)\bullet\phi(\vec{x}_j)$ is called the kernel function, α_i are the Lagrange multipliers.

Recently, how to learn the kernel from data draws many researchers' attention[5]. And some formulations [6-7] have been proposed to perform the optimization in manner of convex combinations of basic kernels. In practice, a simple and efficient method is that the kernel function being illustrated as the convex of combinations of the basic kernel:

$$k(\vec{x}_i,\vec{x}_j)=\sum_{i=1}^{n}\beta_d k(x_{i,d},x_{j,d}) \tag{5}$$

where $x_{i,d}$ denotes the d^{th} component of the input vector $\vec{x}_i$.

Substituting Equation (5) into Equation (3), and multiplying Equation (3) and (4) by β_d, suppose $\gamma_{i,d}=\alpha_i\cdot\beta_d$, then the Lagrange dual problem change into:

$$\max\left\{\sum_{i,d}\gamma_{i,d}-\frac{1}{2}\sum_{i,j,d=1}^{n,m}\gamma_{i,d}\gamma_{j,d}y_i y_j k(x_{i,d},x_{j,d})\right\} \tag{6}$$

$$s.t.\begin{cases} \sum_{i,d} y_i\gamma_{i,d}=0 \\ 0\leq\sum_{d=1}^{m}\gamma_{i,d}\leq c, i=1,\cdots,n \\ \gamma_{i,d}\geq 0, d=1,\cdots,m \end{cases} \tag{7}$$

The new coefficient $\gamma_{i,d}$ replaces the Lagrange coefficient α_i. The number of coefficient that needs to be optimized is increased from n to $n\times m$. It increases the computational cost especially when the number of the attributes in the dataset is large. The linear programming implementation of SVM is a promising approach to reduce the computational cost of SVM and attracts some scholars' attention. Based on above idea, a 1-norm based linear programming is proposed:

$$\min J(\vec{\gamma},\vec{\xi})=\sum_{i,d}\gamma_{i,d}+\lambda\sum_{i=1}^{n}\xi_i \tag{8}$$

$$s.t.\begin{cases} y_i\left(\sum_{j,d}\gamma_{j,d}y_j k(x_{i,d},x_{j,d})+b\right)\geq 1-\xi_i \\ \xi_i\geq 0, i=1,\cdots,n \\ \gamma_{i,d}\geq 0, d=1,\cdots,m \end{cases} \tag{9}$$

In equation (8), the regularized parameter λ controls the sparse of the coefficient $\gamma_{i,d}$.

The dual of this linear programming is:

$$\max\sum_{i=1}^{n}u_i \tag{10}$$

$$s.t.\begin{cases} \sum_{i=1}^{n}u_i y_i y_j k(x_{i,d},x_{j,d})\leq 1, j=1,\cdots,n.d=1,\cdots,m \\ \sum_{i=1}^{n}u_i y_i=0 \\ 0\leq u_i\leq\lambda, i=1,\cdots,n \end{cases} \tag{11}$$

The choice of kernel function includes the linear kernel, polynomial kernel or RBF kernel. Thus, the SVM-MK classifier can be represented as:

$$f(\vec{x})=sign\left(\sum_{j,d}\gamma_{j,d}y_j k(x_{i,d},x_{j,d})+b\right) \tag{12}$$

It can be found that above linear programming formulation and its dual description is equivalent to that of the approach called "mixture of kernel" [7-8]. So the new

coefficient $\gamma_{i,d}$ is called the mixture coefficient. Thus this approach is named "support vector machine with mixture of kernel" (SVM-MK). And we can obtain the sparse solution of coefficient $\gamma_{i,d}$ using 1-norm. So the SVM-MK model greatly reduces the computational complexity. It is more important that the sparse coefficients $\gamma_{i,d}$ give us more choices to extract the satisfied features in the whole space spanned by all the attributes.

1. Experiment analysis

In this section, we use the typical experimental data: Chinese corpus that is collected by Fudan University Dr. Li Ronglu. The corpus including the training set and testing set. There are 1882 documents in the training set. The test set contains 934 documents which have no classification label. And the test set is divided into 10 classes. The proportion of training set text number and test set text number is two-to-one.

In automatic text classification system, will be used in the experiment data is usually divided into two parts: the training set and testing set. The so-called training set is composed of a set of have finished classification (namely has a given category label) text, which used for summed up the characteristics of each category in structure classifier. According to the classification system settings, each class should contain a certain amount of training text.

The test set is the collection of documents that used to test the effect of the classification. Each one of these texts was through the classifier classification, and then the classification results contrast to the correct decision. Thus we can evaluate the effect of classifier. But the test set is not participated in constructing the classifier.

Table 1 experimental data

Text Class	Number of experimental set	
	Training set	*Test set*
Art	166	82
Computer	134	66
Economy	217	108
Education	147	73
Environment	134	67
medicine	136	68
Policy	338	167
Sports	301	149
Transportation	143	71
Military	166	83

In addition, three evaluation criteria measure the efficiency of text classification:

$$recall_i = \frac{a}{a+c} \tag{13}$$

$$precision_i = \frac{a}{a+b} \tag{14}$$

$$F1 = \frac{2 \times precision \times recall}{precision + recall} \tag{15}$$

Where a is the positive example test documentations that are correctly classified as belongs to the number of this kind. b is the negative example test documentations that are be error classified for belong to the number of this kind. c is the positive example test documentations that are be error classified for does not belong to the number of this kind.

Firstly, the data is pre-processed by the ICTCLAS (Institute of Computing Technology, Chinese Lexical Analysis System)[9]. In this method the Gaussian kernel is used, and the kernel parameter needs to be chosen. Thus the method has two parameters to be prepared set: the kernel parameter σ^2 and the regularized parameter λ. The recall, precision and F1 for each category text using SVM-MK approach are shown in **Table 2**.

From **Table 2** and **Figure 1**, we can conclude that the SVM-MK model has better text classification capability in term of the recall, precision and the F1 in comparison with the traditional improved SVM models[10].This method is of better text classification performance in kinds of art, computer, politics, environment, sports, and transportation. Compared to the traditional improved SVM, MK-SVM is not very good in

Table 2. Experimental results using SVM-MK.

Text Class	Evaluation index		
	Precision	*Recall*	*F1*
Art	100.00	97.02	98.49
Computer	98.62	98.79	98.7
Economy	95.10	94.65	94.87
Education	98.64	94.67	96.61
Environment	100.00	93.17	96.46
medicine	97.73	95.48	96.59
Policy	93.23	98.15	95.63
Sports	96.17	99.06	97.59
Transportation	98.57	95.06	96.78
Military	90.21	88.72	89.46

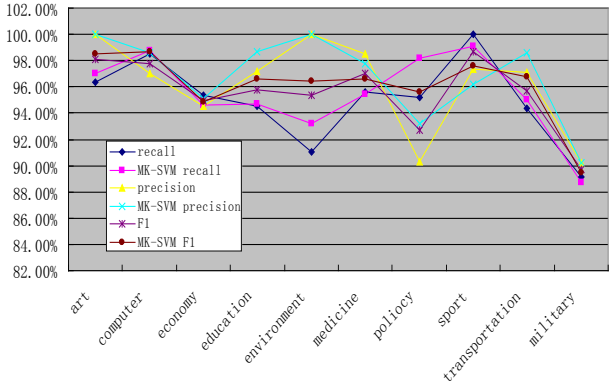

Figure 1. Comparison of results of traditional SVM and MK-SVM.

classifying the kinds of economy, military. This may be because in removing relevant features of test results, and lost some information. So that the recall rate index is affected. This is also need to further improve. Consequently, the proposed SVM-MK model can provide efficient alternatives in conducting text classification tasks.

2. Conclusions

This paper presents a novel SVM-MK text classification model. By using the 1-norm and a convex combination of basic kernels, the object function which is a quadratic programming problem in the standard SVM becomes a linear programming parameter iterative learning problem so that greatly reducing the computational costs. In practice, it is not difficult to adjust kernel parameter and regularized parameter to obtain a satisfied classification result. Through the practical data experiment, we have obtained good classification results and meanwhile demonstrated that SVM-MK model is of good performance in text classification system. Thus the SVM-MK is a transparent model, and it provides efficient alternatives in conducting text classification tasks. Future studies will aim at finding the law existing in the parameters' setting. Generalizing the rules by the features that have been selected is another further work.

3. Acknowledgements

This research has been supported by a public benefit special fund from Quality inspection industry of China (#201210011).

REFERENCES

[1] V. Vapnik, "The nature of statistic learning theory. *Springer*, New York, 1995.

[2] T. Joachims, "Text Categorization with Support Vector Machines Learning with Many Relevant Features," *In European Conference on Machine Learning (ECML).* Chemnitz, Germany: [s.n.], 1998, pp. 137-142.

[3] T. Gartner, P. A. Flach, "WBCSVM: Weighted Bayesian Classification based on support vector machine," *18th Int. Conf. on Machine Learning.* Willianstown, Carla E. Brodley, Andrea Pohoreckyj Danyluk, (eds.), 2001, pp. 207–209.

[4] ChengHua Li, JuCheng Yang, S. C. Park, "Text categorization algorithms using semantic approaches, corpus-based thesaurus and WordNet," *Expert Syst. Appl.* 39(1), pp. 765-772, 2012.

[5] A. Ch. Micchelli, M. Pontil, "Learning the kernel function via regularization," *Journal of Machine Learning Research*, 6, 2005, pp. 1099-1125.

[6] G. R.G. Lanckrient, N. Cristianini, P. Bartlett, L. El Ghaoui, M.I. Jordan. Learning the kernel matrix with semidefinite programming. *Journal of Machine Learning Research*, 5, 2004, pp. 27-72.

[7] F.R. Bach, G. R.G. Lanckrient, M.I. Jordan. Multiple kernel learning, conic duality and the SMO algorithm. *Twenty First International Conference on Machine Learning*, 2004, pp. 41-48.

[8] L.W. Wei, J.P. Li, Z.Y. Chen. Credit Risk Evaluation Using Support Vector Machine with Mixture of Kernel, The 7th International Conference on Computational Science 2007, Lecture Notes in Computer Science 4488, 2007, pp. 431-438.

[9] Institute of Computing Technology, Chinese Lexical Analysis System: http://www.nlp.org.cn/project/project.php?proj_id=6.

[10] F. Jiang, "Research on Chinese Text Categorization based on Support Vector Machine," Degree of Master paper, Chongqing University, 2009.

