CSCanada
Management Science and Engineering
Vol. 6, No. 1, 2012, pp. 64-68
DOI:10.3968/j.mse.1913035X20120601.2850

ISSN 1913-0341 [Print]
ISSN 1913-035X [Online]
www.cscanada.net
www.cscanada.org

Method for Customer Review Category Based MKL-SVM

ZHAO Dan[1,*]

[1] Southwest University for Nationalities, Chengdu, China.
[*] Corresponding author.

Received 27 December 2011; accepted 9 March 2012.

Abstract

Website accumulates a large number of customer reviews for goods and website services. Support vector machine (SVM) is an effective text categorization method, it has strong generalization ability and high classification accuracy which can be used to track and manage customer reviews. But SVM has some weaknesses which slow training convergence speed and difficult to raise the classification accuracy. The paper use heterogeneous kernel functions which have different characteristics to resolve the problem of SVM weak generalization ability to learn and improve the SVM classification accuracy. Through classify customer reviews, online shopping websites resolve issues of critical analysis about mass customers reviews and effectively improve website service standard.

Key words: Customer Review; Text Categorization; SVM; Multiple Kernel Learning

ZHAO Dan (2012). Method for Customer Review Category Based MKL-SVM. *Management Science and Engineering*, *6*(1), 64-68. Available from: URL: http://www.cscanada.net/index.php/mse/article/view/j.mse.1913035X20120601.2850 DOI: http://dx.doi.org/10.3968/j.mse.1913035X20120601.2850

INTRODUCTION

Online shopping website established a customer review system to collect the customer reviews for experience feelings about goods and website services. It is difficult for website to effectively track and manage customer reviews because the reviews expression is complex, content free, sentence variety. Text categorization applied classification function or classification model to map text to a class in those multiple category, so which make retrieval and query faster, accuracy higher. Text categorization applied in many field such as natural language processing, information management, content filtering. The main methods of text categories include Bayesian, decision trees, support vector machine (SVM), neural networks, and genetic algorithms. SVM has good generalization ability to learn, which through the surface separating model to overcome the sample distribution, redundancy features and the impact of factors such as over-fitting[1]. But SVM has slow rate of training convergence and difficult to improve the classification accuracy.

Kernel function is the key method about SVM to solve nonlinear problems. Because the performance of SVM is limited by a single kernel function, SVM generalization learning ability is limited and difficult to improve the classification accuracy. Multiple Kernel Learning (MKL) use of homogeneous or heterogeneous kernel function to optimize, integrate and improve learning ability and generalization of SVM performance. MKL has good flexibility, clear classification results and easy to solve application problems[2], when working with a large number of heterogeneous data, MKL involve many ascertainment and optimization of related parameters, current research about MKL focused on the field of pattern recognition[3, 4]. The paper build the calibration algorithm to determine MKL optimization of heterogeneouskernel power function coefficients and kernel parameters, establish MKL-SVM text categorization model to enhance the application result of SVM in reviews categorization.

1. RELATE WORK

1.1 Kernel Function of SVM

According to statistics, SVM is a learning algorithm based on structural risk minimization, which has high generalization performance of the universal learning machine also. Set split surface $x \cdot \omega + b = 0$, the sample set $\{(x_i, y_i)\}_{i=1}^{l}$ $i = 1, \cdots l, x \in R^d, y \in \{1, -1\}$.

Linear time-sharing SVM split by a hyper plane, the training sample points classification, the two types of training points to the split surface and the minimum distance classification margin maximum[5]. Class interval margin=2/||ω||, so that the maximum interval is equivalent to that $\|\omega\|^2$ min.

Solving the optimal separating surface can be transformed into optimization problems:

$$\min \phi(\omega) = \frac{1}{2}\|\omega\|^2 \quad \text{s.t.}$$

$$y_i[(\omega x_i) + b] - 1 \geq 0, i = 1, \cdots n$$

Optimize the use of its Lagrange dual problem is non-negative Lagrange multipliers, to solve the following maximum function:

$$\max: Q(a) = \sum_{i=1}^{l} a_i - \frac{1}{2}\sum_{i,j=1}^{l} a_i a_j y_i y_j (x_i \bullet x_j) \quad \text{s.t.}$$

$$\sum_{i=1}^{l} y_i a_i = 0 \text{ and } a_i \geq 0, i = 1, \cdots l$$

The optimal classification function is

$$f(x) = \operatorname{sgn}\{(w \bullet x) + b\} = \operatorname{sgn}\{\sum_{i=1}^{l} a_i^* y_i (x_i \bullet x) + b^*\}$$

The function of input data is mapped from a low dimensional input space to a high dimensional space by nonlinear mapping function. Non-linear problems input space can be converted into linear problems in attribute space. This non-linear mapping function is called kernel function [6]. Let x be a map to the high latitudes in the corresponding space c, kernel mapping function $\Phi(x)$.

Kernel function K, $\Phi(x)^T \Phi(x') = K(x, x')$, SVM find a hyper plane $w^T \Phi(x) + b$.

Objective function becomes:

$$Q(a) = \sum_{i=1}^{l} a_i - \frac{1}{2}\sum_{i,j=1}^{l} a_i a_j y_i y_j K(x_i, x_j)$$

$$f(x) = \operatorname{sgn}\{\sum_{i=1}^{l} a_i^* y_i K(x_i, x) + b^*\}$$

Considered the largest division and the training error and optimize the formula:

$$\min_{w,b,e} \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{l} \varepsilon \text{ s.t.}$$

$$y_i(\omega^T \Phi(x_i) + b) \geq 1 - \varepsilon_i, \varepsilon_i \geq 0, i = 1, 2, \ldots l$$

1.2 Multiple Kernel Learning

Let the function set M by a number of kernel functions $K_1 \ldots K_m$ form, the kernel function corresponding to the mapping function is $\Phi_1 \ldots \Phi_M$

MKL[7] formula:

$$\min_{w,b,e} \frac{1}{2}(\sum_{k=1}^{M}\|\omega\|^2)^2 + C\sum_{i=1}^{l}\varepsilon_i \text{ s.t.}$$

$$y_i(\sum_{k=1}^{M}\omega_k^T \Phi_k(x_i) + b) \geq 1 - \varepsilon_i, \varepsilon_i \geq 0, i = 1, 2, \ldots l$$

which ω_k means Φ_k that the entire weight of machine learning.

$$\omega_k = \mu_k \sum_{i=1}^{l} a_i y_i \Phi_k(x_i) \quad 0<ai<C, uk>0, k=1, 2\ldots.M$$

$$\sum_{k=1}^{M} u_k = 1$$

Multiple kernel learning K_{MKL} is a linear combination of convex K_i:

$$K_{MKL} = \sum_{k=1}^{M} \eta_k K_k$$

Classification function:

$$f(x) = \sum_{k:u_k 0} \mu_k \sum_{i:a_i 0} a_i y_i K_k(x, x_i) + b$$

1.3 MKL-SVM Calculation and Optimization of Parameters

Choosing MKL-SVM kernel function depends on the requirement for data processing requirements. With the overall function and partial kernel function are mutually complementary in the classification performance, using different kernel functions can compose a multi-kernel kernel function[8], but if too many types of MKL heterogeneous kernel function, SVM training would be too complicated.

Therefore, the paper used M = 2, the kernel function selected:

Gaussian RBF (Radial Basis Function) kernel function has a good local ability to learn, but weak ability to promote generalization.

$$K(x, y) = \exp\left[-\frac{\|x - y\|^2}{2\sigma^2}\right]\sigma > 0$$

Polynomial Function (PF) is a global kernel function, has a better ability to generalization, but weaker ability to learn.

$$K(x, y) = (\langle x, y\rangle + 1)^d, d \in N$$

Sigmoid kernel function applied to neural network, which has a good overall classification performance.

$$K(x, y) = \tanh(b_0 x \bullet y + b_1)$$

To the $K_{MKL} = \sum_{k=1}^{M} \mu_k K_k$ following two kinds of forms:

$$K_{MKL\text{-GP}} = \lambda * K_{\text{Gaussian}} + (1-\lambda) * K_{Poly}$$

$$K_{MKL\text{-GS}} = \lambda * K_{\text{Gaussian}} + (1-\lambda) * K_{\text{Sigmoid}}$$

MKL-SVM calculation kernel parameters σ 、d 、β_0 、β need to find a suitable value to minimize SVM test error rate is minimum. λ Weights of the MKL-SVM also need to play a key role in optimization. Kernel parameters and weights determine the MKL-SVM. The paper establish the optimization solution steps, through use the relationship between the kernel function is equivalent to the relationship between the kernel matrix, and combine with LOO cross-validation techniques and kernel alignment [9].

1 $\langle K_1, K_2 \rangle = \sum_{i,j=1}^{l} K_1(x_i, x_j), K_2(x_i, x_j)$ the two kernel matrix inner product.

k_1 and k_2 measure kernel calibration sample set S in the differences in the kernel alignment $\hat{A}(S, k_1, k_2) = \frac{\langle K_1, K_2 \rangle}{\sqrt{\langle K_1, K_1 \rangle \langle K_2, K_2 \rangle}}$ is a scalar value, reflecting the differences between different kernel function relationship.

2 k_1, k_2 use the LOO method to find the kernel parameters $\hat{A}(S, k_1, k_2)$ and achieve the largest kernel parameters.

3 defined functions $f(\lambda) = \frac{\langle K, YY^T \rangle_F}{l\sqrt{\langle K, K^T \rangle_F}}$

$c(\lambda) = [\langle K, K \rangle - 1 \le 0; -K \le 0]$, $Y = (y_1 \dots y_l)^T$

4 construct a power parameter λ and the Lagrange multipliers αi equation construct quadratic programming subproblems:

$$\min_{\Delta s} \frac{1}{2} \triangle\lambda^T \lambda + \nabla f(\lambda)^T \triangle\lambda s.t.\ \ \nabla c_i(\lambda)^T \Delta\lambda + c_i(\lambda) \le 0,\ i = 1 \cdots m$$

5Repeat step 4until the error is minimal, the algorithm converges to the optimal value of λ.

2. EXPERIMENTS

2.1 Evaluation Criteria

For the classification of n states, ce is the number of the correct classification about i state, te is the number of not classified, fe is the number error message. Precision (P) $P = \frac{ce}{ce + fe}$ shows the proportion of system correctly classified information to all classified information. Recall (R) $R = \frac{ce}{ce + te}$ shows the possible proportion of the sy stem correctly classified information to all the correct information. Reviews need to consider performance of model P and R, the introduction $F = \frac{(\beta^2 + 1)PR}{\beta^2 P + R}$. β is the relative weight of P and R which decide to focus on the P or the R , usually set to 1. F value is greater which in dicating better classification performance.

2.2 Dimensions of Classification

According to the web service processes and common theme in reviews, the paper summed up the classification dimensions of the reviews (Table 1).

Table 1
Reviews Categories Shopping Dimension

Object	Code	Dimensions	Explain
Pre-purchase	A1	Counseling services	available to resolve customer questions online Q & A
	A2	Product information	provides a comprehensive website product information to meet customer demand information
Commodity	B1	Price	pricing fluctuations impact to customers
	B2	Product Features	for goods using the experience
Web Services	C1	Facilitate payment	Web site variety, safety
	C2	Internal rationing	transfer cargo internal order processing speed
	C3	Communication information	process information timely delivery of goods to customers
Logistics services	D1	Timely	logistics time is reasonable and timely
	D2	Attitude	courier service delivery approach
	D3	Quality	packaging wear, safety
For service	E1	Return goods	return goods handling problems

2.3 Experimental Results

From online shopping websites such as Amazon, DangDang etl, the paper had downloaded 3000 different customer reviews. Randomly selected 2000 as the training set, and the remaining as test set, compared with the F value in several ways. It can be seen from Figure 1, the Gaussian RBF SVM monocytes better than Polynomial. The classification performance Of MKL-GP and MKL-GS is superior to single-kernel SVM. MKL-GP is better than MKL-SG, because sigmoid function only in certain conditions can meet the symmetric and positive semi-definite requirement of kernel functions, which is affecting its performance of classification.

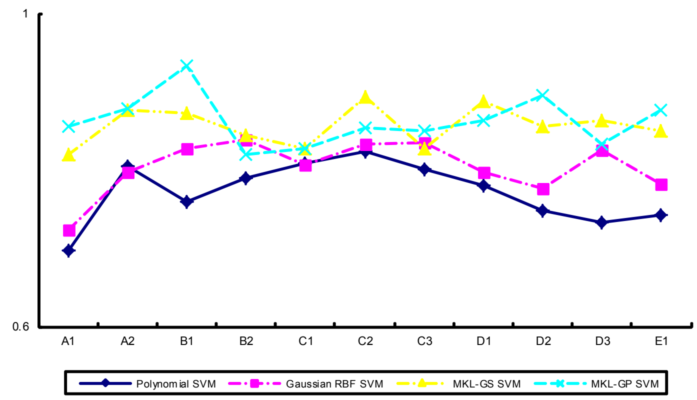

Figure 1
F-Value of Different Methods of Classification

The paper uses a different sample size to compare the two algorithms were compared, F values using the mean. As can be seen from Table 2, MKL-GS shows poor performance characteristics due to the influence of Sigmoid function when the sample size is not large enough. As the number of samples increases, MKL gradually shows obvious advantages and exhibit better classification performance than single-kernel SVM. When the sample size reach 2000, the improvement of SVM classification performance is not obvious.

Table 2
F-value of Different Sample

Sample	Polynomial SVM	Gaussian RBF SVM	MKL-GS SVM	MKL-GP SVM
300	0.6573	0.6639	0.5081	0.6047
600	0.6894	0.7429	0.7026	0.7493
1000	0.7047	0.7579	0.8081	0.8169
2000	0.7715	0.8040	0.8559	0.8621
3000	0.7745	0.8037	0.8551	0.8676

CONCLUSION

Critical analysis of online shopping needs of a variety of technologies. Review classification of critical analysis is the first step in shopping, but also requires a combination of text mining, sentiment analysis, marketing analysis and other methods to help companies master the consumer experience and provide policy recommendations. MKL-SVM ensemble learning to play different characteristics of heterogeneous kerne l function and achieve improvement of classification performance about shopping reviews. There are various kinds of praise on the network, such as comments, news, blog, micro blogging, the analysis for these types of text classification. Through optimize the kernel function and related parameters, other researchers can achieve better classification performance by reference of MKL-SVM method.

REFERENCES

[1] SHU, Jinshu, ZHANG, Bofeng & XU, Xin (2006). Text Classification based on Machine Learning Technology Research. *Journal of Software, 17*(9), 1848-1859.

[2] F. R. Bach, G. R. G. Lanckriet & M. I. Jordan (2004). Multiple Kernel Learning, Conic Duality, and the SMO Algorithm. *Process 21st International Conference Machine*

Learn (pp. 6-14).

[3] S. Sonnenburg, G. Ratsch, C. Schafer (2006). Large Scale Multiple Kernel Learning. *Machine Learning Research, 7*(12), 1531-1565.

[4] Koji Tsuda, Gunnar Rätsch, *et al.* (2001). Learning to Predict the Leave One out Error of Kernel Based Classifiers. *Process International Conference Artificial Neural Networks, 21*(3), 331-338.

[5] Smits, G. F. & Jordaan, E. M. (2005). Improved SVM Regression Using Mixtures of Kernels Neural Networks. *Proceedings of the 2002 International Joint Conference*, *5*, 2785- 2790

[6] A. Rakotomamonjy, F. Bach, S. Canu, Y. Grandvalet (2007). More Efficiency in Multiple-Kernel Learning. *Proc. 24th Int. Conf. Mach. Learn., Corvallis, 6,* 775-782.

[7] HU, Mingqing, CHEN, Yiqiang & James Tin-Yau Kwok (2009). Building Sparse Multiple-Kernel SVM Classifiers. *IEEE Transactions on Neural Networks, 20*(5), 1-12.

[8] LIU, Xiangdong, LU, Bin & CHEN, Zhaoqian (2005). Support Vector Machine Model Selection of Optimal. *Computer Research and Development, 42*(4), 576-581.

[9] N. Cristianini, J. Shawe-Taylor & J. Kandola (2002). On Kernel Target Alignment. *Neural Information Processing Systems Cambridge*, 367-373.