

Application of Twin Objective Function SVM in Sentiment Analysis

Qiaoman YANG^{a1}, Chunyu LIU^b

^a*Shaanxi College of Communication Technology, China*

^b*Beijing Normal University, China*

Abstract. Classification modeling is one of the key issues in sentiment analysis. Support vector machine (SVM) has been widely used in classification as an effective machine learning method. Generally, a common SVM is only for decision-making that sacrifices the distribution of data. In practice, sentiment data are big and mazy, which results in the deficiency of accuracy and stability when common SVM is used. The study investigates sentiment analysis by applying the twin objective function SVM, including nonparallel SVM(NPSVM) and twin SVM (TWSVM). From the experiments, we concluded that twin objective function SVMs are superior to NB and single objective function SVM in accuracy and stability.

Keywords. sentiment analysis, twin objective functions, SVM, NB

1. Introduction

Because of the flourishing development of the internet and Web2.0, an increasing number of people are depending on the network. They express their feelings or evaluate on things via discussion forums, blogs, twitter, etc., which results in massive and big data explosion. However, it is difficult to collect and process vast amounts of information online by artificial methods. Sentiment analysis technology came into being in this environment. Sentiment analysis technology is an emerging field that mine unstructured information. Its aim is to take advantage of automation and intelligent technology analyzing mining, reasoning and learning the subjectivity text messages that have some emotional colors, obtaining the potential and valuable hidden information. Currently, sentiment analysis technology has many widely used applications in fields of business intelligence application, recommender systems, message filtering and so on [1][2]. Therefore sentiment analysis technology is a hot research field. Sentiment information classification, which classifies the binary appraisal data, that have subjective information, is one of the main tasks in the sentiments analysis. Sentiment analysis, which is based on supervised learning, emotional research is a hot research area based on machine learning. The machine learning technique of emotional classification is of because it can model many functions and capture context [3] in the process. And they are relatively easy to adapt to

¹ Corresponding Author. Assistant, Shaanxi College of Communication Technology, 19 Wenjing Road, Xi'an, Shaanxi China; E-mail: yangqman@163.com.

changing inputs and the possibility of measuring the uncertainty of the classification. The method of monitoring training from the manual classification is the most popular.[4],[5],[6],[7].

In addition to the Non-negative Matrix Tri-factorization[8],Genetic Algorithms[9], the most supervised learning algorithm which used on the sentiment analysis are still Naïve Bayes (NB),Maximum Entropy (ME)and Support Vector Machine (SVM) [10],[11]. A SVM [12] Operate by constructing a hyper plan with the closest training distance with the largest Euclidean distance. This can regard as the range between the separating hyper plane and the two parallel hyper planes on each side, indicate the border of an example of a class in the feature space. It is known that SVM is robust in the case of many features, not destined by the curse of dimensions, and produces the highest accuracy in sentiment classification [13]. However, a single objective classification, that aims at structuring the decision-surface of data, does not fit the data well, and it has some disadvantages, such as, running slowly, and being applicable only to small data.

Twin objective function SVM are now a hot in the field of studying support vector machine. The method has two objective functions, aiming to produce two separating hyper planes, thus, having more advantages than single objective function SVM. Such as, considering more data characteristics, more amenable to parallel computing, making the classification more accurate, faster, broadening the scope of data, and so on. This study applies two general and representative twin objective function SVM to the sentiment analysis. These two methods are Twin Support Vector Machine (TWSVM) introduced by R.K.jayaadeva. et. al. and Non-parallel Support Vector Machine introduced by Tian [14]. In our experiment, we compare Naïve Bayes (NB), the classical SVM algorithm Lib SVM and SMO with the two methods that are used in this study. The results show that twin objective function SVM are prior to other machine learning methods in accuracy and stability on the sentiment analysis data sets.

This paper is organized as follows. Section 2 briefly dwells on Naïve Bayes, the single objective function support vector machine. Section 3 describes two twin objective function support vector machines (TWSVM and NPSVM). Section 4 presents the experimental results and section 5 contains the conclusion.

2. Methods

2.1 Naïve Bayes

One approach to text classification is to assign to a given document a class $b^* = \arg \max_b p(b|a)$. We derive the NB classifier by first observing that by Bayes' rule:

$$p(b|a) = \frac{p(b)p(a|b)}{p(a)},$$

where $p(a)$ plays no role in selecting b^* . To estimate the term $p(a|b)$, NB decomposes it by assuming the f_i 's are conditionally independent given a 's class:

$$p_{NB}(b|a) = \frac{p(b)(\prod_{i=1}^m p(f_i|b)^{y_i(a)})}{p(a)}$$

That is to say, assuming that each independent function is an indication of the allocation class, independent of each other, we chose this classifier because of its

simple implementation, high computational efficiency, and straightforward features in learning. [15], [16]. We used the NB classifier of WEKA.

2.2 SVM

SVM, which was introduced by Vapnik and his co-workers in the early 1990s, is a classifier for the two types of the optimal classification that are linear inseparable. The goal of the SVM is to find an optimal separating hyper plane by constructing a hyper plane separating two classes with no error and maximizing the margin of the two types.

For classification about the training data

$$T = \{(x_i, y_i) | i = 1, \dots, l\}, \quad (1)$$

where $x_i \in \mathfrak{R}^n, y_i \in Y = \{+1, -1\}, i = 1, \dots, l$. Linear SVM is applied to solve the following primal quadratic programming problem (QPP):

$$\min_{w, b, \xi} \frac{1}{2} \|w\|_2^2 + C \sum_{i=1}^l \xi_i, \quad (2)$$

$$s.t. y_i((w \cdot x_i) + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, i = 1, 2, \dots, l,$$

where C is a penalty parameter and ξ_i are the slack variables. The Wolf Dual of

(2) can be expressed as:

$$\max_{\alpha} \sum_{j=1}^l \alpha_j - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j (x_i \cdot x_j) \alpha_i \alpha_j \quad (3)$$

$$s.t. \sum_{i=1}^l y_i \alpha_i = 0, \quad 0 \leq \alpha_i \leq C, i = 1, \dots, l,$$

where $\alpha \in \mathfrak{R}^l$ are Lagrangian multipliers. The optimal separating hyper plane of (3) is expressed as:

$$w = \sum_{i=1}^l \alpha_i^* y_i x_i, \quad b = \frac{1}{N_{sv}} (y_j - \sum_{i=1}^{N_{sv}} \alpha_i^* y_i (x_i \cdot x_j)) \quad (4)$$

A new sample is classified as +1 or -1 according to the final decision function $f(x) = \text{sgn}((w \cdot x) + b)$.

Lib SVM and SMO are the effective algorithms to solve SVM. Lib SVM is developed and designed by Zhiren Lin and others of the Taiwan University. It is a simple, easy to use fast, and efficient packages to solve the SVM recognition and regression. It involves few adjustment parameters, and provides cross-validation functions. SMO is a fast and efficient algorithm for the SVM. It transfers the large convex programming problem into a small convex programming problem, and then it can train the larger data sets. The study described in this paper mainly uses the following two methods to solve the problem of the SVM.

2.3 Twin Objective Function SVM

Single objective SVM has some of the following problems when it is applied in the sentiment analysis. A single objective function is aimed at constructing the decision-surface of the data, but it is not fit well to the distribution of data with the different emotional categories. Twin objective function SVM can achieve optimal data classification data while approaching data effectively.

2.3.1 TWSVM

TWSVM is proposed by Jayadeva based on SVM. Its goal is to find two hyper planes for binary classification, with each super plane approaching one of the categories and being far away from other categories as soon as possible.

Consider the binary classification problem with the training set

$$T=\{(x_1,+1),\cdots,(x_p,+1),(x_{p+1},-1),\cdots,(x_{p+q},-1)\}, \quad (5)$$

TWSVM seeks two hyper planes

$$(w_+ \cdot x) + b_+ = 0 \text{ and } (w_- \cdot x) + b_- = 0 \quad (6)$$

pass through seek two small QPPS

$$\min_{w_+, b_+, \xi_-} \frac{1}{2} \sum_{i=1}^p ((w_+ \cdot x_i) + b_+)^2 + d_1 \sum_{j=p+1}^{p+q} \xi_j \quad (7)$$

$$\begin{aligned} \text{s.t. } & (w_+ \cdot x_i) + b_+ \leq -1 + \xi_j \\ & \xi_j \geq 0, j = p+1, \cdots, p+q \end{aligned}$$

and

$$\min_{w_-, b_+, \xi_+} \frac{1}{2} \sum_{i=p+1}^{p+q} ((w_- \cdot x_i) + b_-)^2 + d_2 \sum_{j=1}^p \xi_j \quad (8)$$

$$\begin{aligned} \text{s.t. } & (w_- \cdot x_j) + b_- \leq 1 - \xi_j \\ & \xi_j \geq 0, j = 1, \cdots, p \end{aligned}$$

where $d_i \geq 0, i=1,2$ are the penalty parameters .First, we obtain the dual problems of (7) and (8) by using the Lagrangian equation and KKT conditions:

$$\max_{\alpha} e_-^T \alpha - \frac{1}{2} \alpha^T G (H^T H)^{-1} G^T \alpha \quad (9)$$

$$\text{s.t. } 0 \leq \alpha \leq d_1 e_-$$

and

$$\max_{\beta} e_+^T \beta - \frac{1}{2} \beta^T P (Q^T Q)^{-1} P^T \beta \quad (10)$$

$$\text{s.t. } 0 \leq \beta \leq d_2 e_+$$

where

$$\begin{aligned} G &= [B e_-], H = [A e_-], P = [A e_+], Q = [B e_-] \\ e_+ &= (1, \cdots, 1)^T \in R^p, e_- = (1, \cdots, 1)^T \in R^q \\ A &= (x_1, \cdots, x_p)^T \in R^{p \times n}, B = (x_{p+1}, \cdots, x_{p+q})^T \in R^{q \times n} \end{aligned} \quad (11)$$

We can obtain the solutions (w_+, b_+) and (w_-, b_-) of the problems (9) and (10) from the above calculation. A new data $x \in \mathcal{R}^n$ can be predicted to belong to emotional category by

$$\text{Class} = \arg \min_{k=-,+} |(w_k \cdot x) + b_k| \quad (12)$$

where $|\cdot|$ is the perpendicular distance of point x from the planes

$$(w_k \cdot x) + b_k = 0, k = -, +$$

TWSVM improves the accuracy of the SVM, while increasing the speed of SVM, but TWSVM needs to calculate the inverse matrix during computing process, thus is still very difficult for large-scale data processing.

2.3.2 NPSVM

NPSVM is based on sustain vector machine and double support vector machine, its main goal is also to find two hyper planes fitting each category of data. However, due to its form, it is very similar to the standard SVM, we can solve the NPSVM by using the technology for solving the standard SVM.

We seek two nonparallel hyper planes $(w_+ \cdot x) + b_+ = 0$ and $(w_- \cdot x) + b_- = 0$ for the two types of data (such as (5)) by solving two convex QPPs

$$\begin{aligned} \min_{w_+, b_+, \eta_i^*, \xi_j} & \frac{1}{2} (\|w_+\|^2) + C_1 \sum_{i=1}^p (\eta_i + \eta_i^*) + C_2 \sum_{j=p+1}^{p+q} \xi_j, \\ \text{s.t.} & (w_+ \cdot x_i) + b_+ \leq \varepsilon + \eta_i, i=1, \dots, p, \\ & -(w_+ \cdot x_i) - b_+ \leq \varepsilon + \eta_i^*, i=1, \dots, p, \\ & (w_+ \cdot x_j) + b_+ \leq -1 + \xi_j, j=p+1, \dots, p+q, \\ & \eta_i, \eta_i^* \geq 0, i=1, \dots, p, \\ & \xi_j \geq 0, j=p+1, \dots, p+q, \end{aligned} \quad (13)$$

and

$$\begin{aligned} \min_{w_-, b_-, \eta_i^*, \xi_j} & \frac{1}{2} (\|w_-\|^2) + C_3 \sum_{i=p+1}^{p+q} (\eta_i + \eta_i^*) + C_4 \sum_{j=1}^p \xi_j, \\ \text{s.t.} & (w_- \cdot x_i) + b_- \leq \varepsilon + \eta_i, i=p+1, L, p+q, \\ & -(w_- \cdot x_i) - b_- \leq \varepsilon + \eta_i^*, i=p+1, L, p+q, \\ & (w_- \cdot x_j) + b_- \geq -1 + \xi_j, j=1, L, p, \\ & \eta_i, \eta_i^* \geq 0, i=p+1, L, p+q, \\ & \xi_j \geq 0, j=1, L, p, \end{aligned} \quad (14)$$

where $C_i \geq 0, i=1, \dots, 4$ are the penalty parameters. First, we obtain the dual problems of (13) and (14) by using the Lagrangian equation and KKT conditions:

$$\begin{aligned} \min_{\alpha_+^*, \beta_-} & \frac{1}{2} (\alpha_+^* - \alpha_+)^T A A^T (\alpha_+^* - \alpha_+) \\ & - (\alpha_+^* - \alpha_+)^T A B^T \beta_- + \frac{1}{2} \beta_-^T B B^T \beta_- \\ & + \varepsilon e_+^T (\alpha_+^* + \alpha_+) - e_-^T \beta_-, \\ \text{s.t.} & e_+^T (\alpha_+ - \alpha_+^*) + e_-^T \beta_- = 0 \\ & 0 \leq \alpha_+, \alpha_+^* \leq C_1 e_+, \quad 0 \leq \beta_- \leq C_2 e_-. \end{aligned} \quad (15)$$

and

$$\begin{aligned} \min_{\alpha_-^*, \beta_+} & \frac{1}{2} (\alpha_- - \alpha_-^*)^T B B^T (\alpha_- - \alpha_-^*) \\ & + (\alpha_- - \alpha_-^*)^T B A^T \beta_+ + \frac{1}{2} \beta_+^T A A^T \beta_+ \\ & + \varepsilon e_+^T (\alpha_-^* + \alpha_-) - e_+^T \beta_+, \\ \text{s.t.} & e_-^T (\alpha_- - \alpha_-^*) + e_+^T \beta_+ = 0 \\ & 0 \leq \alpha_-, \alpha_-^* \leq C_3 e_-, \quad 0 \leq \beta_+ \leq C_4 e_+. \end{aligned} \quad (16)$$

where

$$\begin{aligned} \alpha_+^* &= (\alpha_+^T, \alpha_+^{*T}) = (\alpha_1, \dots, \alpha_p, \alpha_1^*, \dots, \alpha_p^*)^T, \\ \alpha_-^* &= (\alpha_-^T, \alpha_-^{*T}) = (\alpha_p, \dots, \alpha_{p+q}, \alpha_p^*, \dots, \alpha_{p+q}^*)^T, \\ \beta_+ &= (\beta_1, \dots, \beta_p), \beta_- = (\beta_{p+1}, \dots, \beta_{p+q}). \end{aligned}$$

We can obtain the solutions (w_+, b_+) and (w_-, b_-) of the problems (13) and (14) from the above calculation. A new data $x \in \mathcal{R}^n$ can be predicted to be positive or negative by (12)

$$\text{Class} = \arg \min_{k=-, +} |(w_k \cdot x) + b_k|$$

NPSVM is much more easily solved by computing than TWSVM, Therefore, it runs faster than the latter, and easier to implement as well. Since the main functions of these two SVMs appear in pairs, we call them the dual objective function support vector machine.

3. Experiments

The study reported in this paper mainly uses the standard database that is commonly used in the sentiment analysis ,the first data set is the Intel Movie Database (IMDB),the second dataset is the Pang[13] Database. The third is the NLPDVD database. The features of these three data sets shown in table 1.

Table 1. Database features

Database	Size	Positive	Negative	Feature
Pang2002	2000	1000	1000	8614
IMDB	4000	2000	2000	5706
NLPDVD	4000	2000	2000	2000

In the experiments, in order to obtain the precision execution efficiency, we use the data preprocessing. That is, we select the emotional word with the document frequency greater than two and with the document sets greater than ten as the feature. Then, we use TF-IDF to represent the text. Each experimental data set is randomly and equally divided into a training set and test set.

In this section, we compare NB, SVM (Lib SVM and SMO) and dual objective function on the three standard data sets. In a large number of experiments, the study found using the linear kernel on the three data sets is better than radial basis function. Therefore, all the methods in the experiments use the linear kernel. For all the methods, the optimal parameters $d_i, i=1,2$ in TWSVM and, $C_i, i=1,L,4$ in NPSVM are tuned for the best classification accuracy in the range 2^{-5} and 2^5 . Get the best parameters ε in this range $[0, 0.5]$ with the step 0.05. Lib SVM, SMO and NB taken from WEKA which is an open source collection tools. Classification accuracy of each method is measured by the standard fivefold cross-validation methodology. All methods are implemented on a PC with an Intel Core processor and 4GB RAM. The experimental results are shown in Table2.

Table 2. Mean value of benchmark date set

Data sets	NB Accuracy%	Lib SM Accuracy%	SMO Accuracy%	TWSM Accuracy%	NPSM Accuracy%
Pang2002	70.5	77.9	84.8	92.74	81.4
IMDB	79.94	83.9	86.1	90	92.69
NLPDVD	66.5	75.65	69.4	87.64	81.15

In order to illustrate the stability and accuracy of the algorithm in sentiment analysis, we performed some experiments on IMDB data sets and pang2002 data sets, to account for various percentages of the training set for testing. The results are shown in figures 1 and 2.

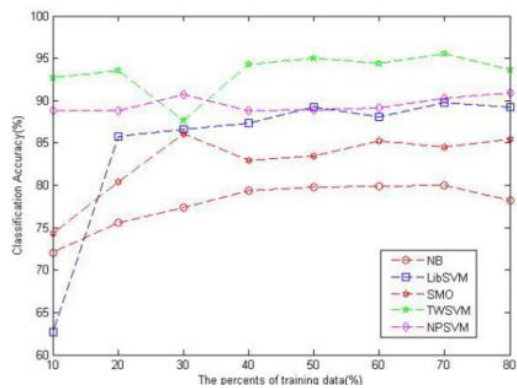


Figure 1. Average results of IMDB

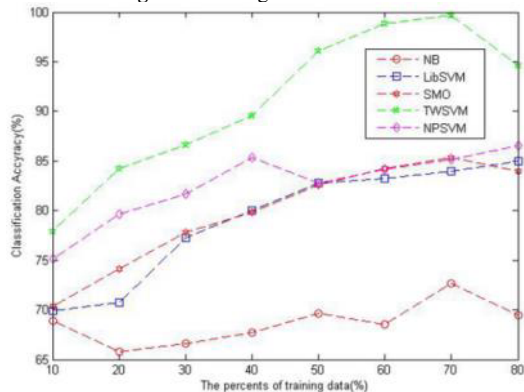


Figure 2. The average results on pang2002

The analysis of the above results according to Table 2 shows that Twin objective function SVMs are higher by approximately 1% to 10% than NB and Single objective function SVM in classification accuracy, when they run on IMDB data sets and pang2002 data sets. As clearly seen from figures 1 and 2, NPSVM and TWSVM are better than NB, Lib SVM, and SVM classifiers in stability and accuracy.

4. Conclusion

In this study, the Twin objective functions TWSVM and NPSVM were applied to sentiment analysis and good results were obtained. Experiments show that in most cases, the double objective function SVM has better classification accuracy than the other three methods. At the same time, the training speed of NPSVM is significantly faster than Lib SVM and SVM. In addition, the dual objective function SVM generates two separate hyper planes. Therefore, its application range is much wider. In sentiment analysis, the life-long objective function SVM, as a classifier, has a stronger generalization performance than a general support vector machine.

References

- [1] Loren Terveen, Will Hill et al, "PHOAKS: A system for sharing recommendations," *Communications of the ACM*, vol, 40,no.3,pp.59–62,1997.
- [2] Ellen Spertus, "Smokey: Automatic recognition of hostile messages," In *Proc. of Innovative Applications of Artificial Intelligence*, pp.1058–1065, 1997.
- [3] Polanyi L and Zaenen A, "Contextual valence shifters. In: *Computing attitude and affect in text: Theory and applications*," Springer, pp. 1–10,2006.
- [4] Thomas, M, Pang, B, and Lee, L, "Get out the vote: Determining support or opposition from Congressional debate transcripts," In *Empirical Methods in Natural Language Processing*,2006.
- [5] Finn A and Kushmerick N, "Learning to classify documents according to genre.,"*Journal of the American Society for Information Science*, vol,57,pp.1506–1518, 2003.
- [6] Bai X, Padman R and Airolidi E, "On learning parsimonious models for extracting consumer opinions," In: *Proceedings of HICSS-05, 38th Annual Hawaii International Conference on System Sciences*, IEEE Computer Society, Washington, DC, pp. 75–82, 2005.
- [7] Aue A and Gamon M, "Customizing sentiment classifiers to new domains," a case study, Technical report, Microsoft Research,2005.
- [8] Tao Li, Ti Zhang, Vikas Sindhwani, "A Non-negative Matrix Tri-factorization Approach to Sentiment Classification with Lexical Prior Knowledge" *Proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP*, pp.244–252, Suntec, Singapore, August .2009.
- [9] Ahmed Abbasi, Hsinchun Chen, Arab Salem, "Sentiment Analysis in Multiple Languages: Feature Selection for Opinion Classification in Web Forums," *ACM Tranon IS*, Vol,26, no.3, pp. 12:1-12:34,2008.
- [10] Bo Pang and Lillian Lee, "Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts," In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics*, Barcelona, Spain.2004.
- [11] Berger A L, Pietra S D and Pietra V J D, "A maximum entropy approach to natural language processing," *Computational Linguistics*, vol, 22, pp.39–71, 1996.
- [12] V. N. Vapnik, "An overview of statistical learning theory," *IEEE Trans. Neural Networks*, vol ,10, no.1, pp.988–999, 1999.
- [13] Pang B, Lee L, et.al,"Thumbs up? Sentiment classification using machine learning technique," *Proc.of the EMLP* vol,7, pp.79-86, 2002.
- [14] Yingjie Tian, Xuchan Ju, Zhiqian Qi, "Nonparallel support vector for pattern classification,"*IEEE Trans. Cybernetics*, Vol,44,no.7,pp.1067-1076, 2014.
- [15] Salvetti F, Lewis S and Reichenbach C, "Impact of lexical filtering on overall opinion polarity identification," *Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect inText: Theories and Applications*, Stanford, CA.2004.
- [16] Pedro Domingos and Michael J. Pazzani, "On the optimality of the simple Bayesian classifier under zero-one loss," *Machine Learning*, vol, 29, no. (2-3): pp.103–130,1997.