11-2018

Comparing ELM with SVM in the field of sentiment classification of social media text data

Zhihuan CHEN

Zhaoxia WANG
Singapore Management University, zxwang@smu.edu.sg

Zhiping LIN

Ting YANG

Comparing ELM with SVM in the field of sentiment classification of social media text data

Zhihuan Chen[1,2], Zhaoxia Wang[1*], Zhiping Lin[2] and Ting Yang[3]

[1]Institution of High Performance Computing (IHPC)
Agency for Science, Technology and Research (A*STAR)
Singapore, 138632
*Corresponding author: wangz@ihpc.a-star.edu.sg
[2]School of Electrical and Electronic Engineering (EEE)
Nanyang Technological University
Singapore, 639798
chen1056@e.ntu.edu.sg, ezplin@ntu.edu.sg
[3]School of Electrical Automation and Information Engineering,
Tianjin University
Tianjin, 300072
yangting@tju.edu.cn

Abstract. Machine learning has been used in various fields with thousands of applications. Extreme learning machine (ELM), which is the most recently developed machine learning algorithm, has become increasingly popular for its good generalization ability. However, it has been relatively less applied to the domain of social media. Support Vector Machine (SVM), another popular learning-based algorithm, has been applied for sentiment classification of social media text data and has obtained good results. This paper investigates and compares the capabilities of these two learning-based methods in the field of sentiment classification of social media. The results indicate that SVM can obtain good performance for analyzing small datasets, while for large datasets, ELM performs better than SVM. This research also indicates that ELM has the potential application in the domain of social media analysis.

Keywords: ELM, SVM, Sentiment Classification, Social Media, Learning-based Method

1 Introduction

With the development of information technology, machine learning has become one of the most important tools to solve data related problems. Without being explicitly programmed, it gives the computers learning capabilities. There are 3 types of machine learning algorithms:

- Supervised learning: machines learn general rules based on labeled data.
- Unsupervised learning: machines learn general patterns of a set of unlabeled data.

adfa, p. 1, 2017.

- Reinforcement learning: machines learn based on rewards and punishments of their actions.

With enough labeled or marked ground truth data available, supervised learning methods expressed outstanding advantages for solving classification and prediction problems.

Sentiment classification is a subfield of sentiment analysis. It refers to classifying a comment or document into positive, negative or neutral class. With the advent of social media, social media data have played a critical role because such huge social media data may hide large amounts of useful information, such as opinions, attitudes or sentiments towards different brands or topics [1] [2]. Such valuable information can be used to produce insights for private or public organizations. Thus, opinion or sentiment classification on social media has attracted more and more researchers' interests. Two general methods are usually applied to this problem: learning based methods and non-learning based methods. Compared to non-learning based methods, the supervised learning based methods are preferred since they can utilize large amount of labeled training data to automatically produce classifier models for efficient sentiment classification.

SVM and ELM are both very popular supervised learning methods. Both of them evolved from the neural network model and different researchers have applied them to different research areas and given different opinions on the pros and cons of these algorithms. In this work, we compare the capabilities of these two algorithms for sentiment classification on social media text data. The results of these two algorithms are evaluated and their performances are compared in terms of accuracy, precision, recall, f-measure and resource consumption.

The rest of the paper is organized as follows: related works are discussed in Section 2; Section 3 focuses on the datasets and methodology; Section 4 discusses the results; and lastly, conclusion and future work are presented in Section 5.

2 Related work

Support Vector Machine (SVM) has become one of the standard tools in the data mining fields due to its good generalization performance [3]. It can be described by a neural network with a kernel function. It has been well applied to many fields and has various applications [4] [5] [6]. Huang et al. has built a prediction model for stock market prices using SVM and the accuracy hit 73% in terms of hit ratio [4]. Heisele et al. has applied SVM to face recognition in which an SVM model is trained using feature vectors with 10 facial components extracted from different images [5]. Additionally, a medical decision making system based on least square SVM has been constructed for the diagnosis of breast cancer in [6].

ELM is another machine learning algorithm based on neural network. It is first proposed by Huang et al. as a novel machine learning algorithm based on single-hidden layer feedforward neural networks [7]. Although this algorithm is most recently proposed, it has attracted a lot of research interests because it has better generalization results and faster learning speeds compared to traditional gradient-based machine learning algorithms [3]. Many applications have been developed based on this algorithm [8] [9] [10]. Wong et al. built a real-time fault diagnostic system for gas turbine based on ELM [8]. Xu et al. and Shin et al. have applied ELM in image processing. They have used the model for object recognition and image classification [9] [10].

In the field of sentiment classification, SVM is often used by researchers [11] [12] [13] while ELM is seldom used [14]. Wang et al. combined some pre-processing techniques with machine learning algorithms to improve the performance of the machine learning algorithms [11]. Kolchyna et al. used a hybrid method which combined learning based method and non-learning based method. Lexicon scores of each sentence were fed into an SVM classifier as an additional feature during the training step [12]. Wang et al. has used both ELM and SVM for sentiment classification, and the results showed that ELM outperforms SVM in terms of accuracy with the dataset they used [14]. However, they did not mention anything about the computing time or resource consumption which is another important measurement for the performance.

In this paper, SVM and ELM methods are used for the sentiment classification of social media text data. Chi Square scoring method is used to score each word feature. Different score thresholds are set to select the top n features for the entire set of features. Different sizes of datasets are used for better comparison studies.

3 Datasets and Methodology

3.1 Datasets

With the development of technology, there are different types of sources available online and they are open to the public. The social media text data used in sentiment classification can either be extracted using different application programming interfaces (APIs), such as the Twitter API, or be downloaded directly from third party websites.

The data used in this paper is directly downloaded from the website "twitter-sentiment-analyzer" [15]. The file contains 1.6 million pre-classified tweets. We have extracted different datasets with different sizes for this study. According to the size of datasets used, the datasets are named as ds_10k, ds_20k and ds_40k which consist of 10k, 20k and 40k tweets respectively. Each of these datasets is further divided into two parts: a training set for training the classifier and a testing set to evaluate the performance. The training set derives from 75% of the tweets and the remaining 25% of the tweets is used as the test set.

3.2 Feature Selection

Feature selection is the process that selects a subset from the entire set of features in order to improve the performance of the learning based algorithms [11]. Before performing the feature selection, pre-processing is performed to remove the noise in the text data [16], such as removing the stop word, and converting or replacing the non-word characters to words. The feature selection method used in this paper is Chi Square method as it is one of the top performers for this purpose [17]. The lack of independence between term *t* and class *c* can be measured and calculated using the following equation according to this method [11]:

$$\kappa^2\ (f,c)=N\times(AD\text{-}BC)^2/[(A+C)\times(B+D)\times(A+B)\times(C+D)] \tag{1}$$

where *N* is the total number of the samples. *A* is the number of times that *t* occurs in *c*. *B* is the number of times that *t* does not occur in *c*. *C* is the number of the samples without *t* and *D* is the number of samples that do not contain *t* nor do they occur in *c*. In order to do the comparison analysis, the above equivalent feature selection processes are implemented for both the SVM and ELM methods.

3.3 Implementation of SVM and ELM

SVM is first proposed by Cortes and Vapnik as a kind of support-vector neural network. The idea of this algorithm is: the input vectors are mapped onto some high dimensional feature space so that these inputs can be separated by a linear decision surface named hyperplane in that space [18]. The SVM tool used in this paper is the latest standard software package downloaded from libsvm website https://www.csie.ntu.edu.tw/~cjlin/libsvm/ [19].

In 2004, Huang proposed a new machine learning algorithm named Extreme Learning Machine [7]. It is based on single layer feedforward neural network. Unlike traditional neural network, it is not necessary for ELM to tune the input weights and the hidden layer bias as these parameters are generated randomly. This feedforward neural network can be described as a linear system. By using inversed operation on hidden layer matrices, the output weights can be determined. The ELM tool used in this paper is the latest standard software package downloaded from python website https://pypi.python.org/pypi/hpelm [20].

In this paper, we have used a voting based ELM instead of original ELM [21]. Multiple ELM classifiers are trained in parallel and used for the classification. These classifiers are trained in parallel so that the multiple training processes can run at the same time. The outcome of each classifier serve as a vote and the final result depends on these votes [20]. We also applied this voting system on SVM. However, it did not help to improve the accuracy of SVM. Both SVM and ELM are implemented by using the corresponding latest source codes for fair comparison.

4 Compared Results and Discovery

We used a virtual machine (16 Cores and 64 GB RAM) to test these two methods on datasets ds_10k, ds_20k and ds_40k, respectively. Figs. 1(a)-(c) show the plots of log(no_of_features) vs. accuracy for datasets ds_10k, ds_20k and ds_40k for both SVM and ELM. Figs.2 (a)-(c) are the plots of log(no_of_features) vs. total_time for datasets ds_10k, ds_20k and ds_40k. In each plot, the red dots represent the results using SVM and blue triangles are for ELM. Table 1 and Table 2 show the best results for both algorithms.

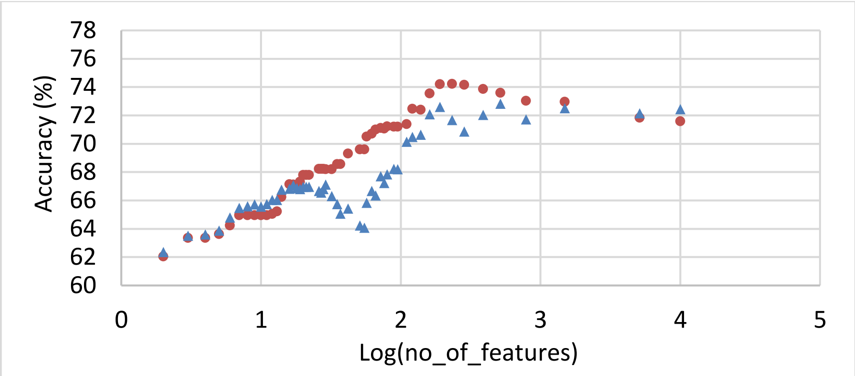

Fig. 1a. Plot of log(no_of_features) vs. accuracy for ds_10k

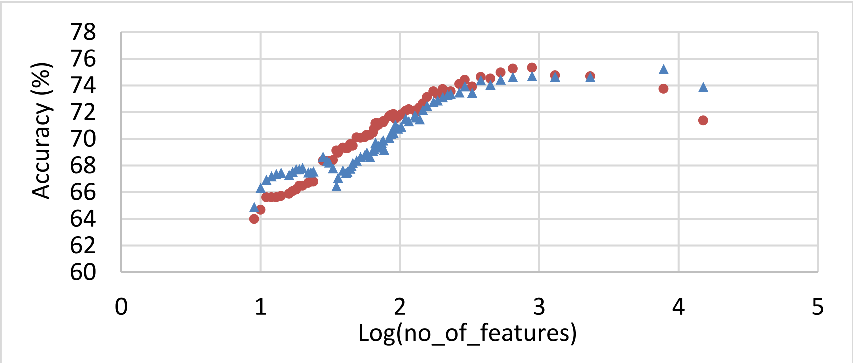

Fig. 1b. Plot of log(no_of_features) vs. accuracy for ds_20k

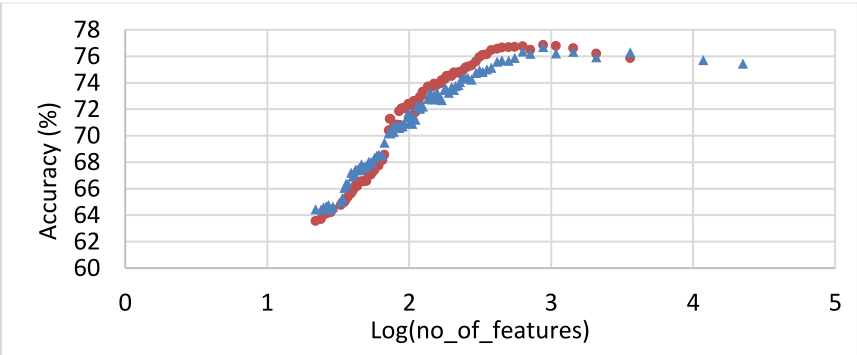

Fig. 1c. Plot of log(no_of_features) vs. accuracy for ds_40k

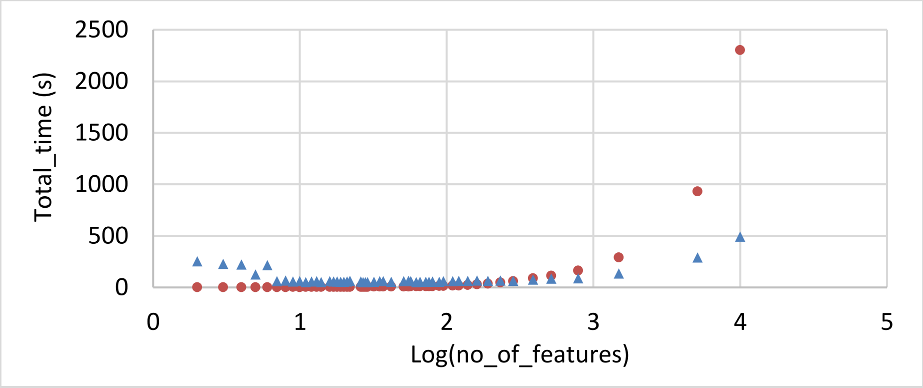

Fig. 2a. Plot of log(no_of_features) vs. total_time for ds_10k

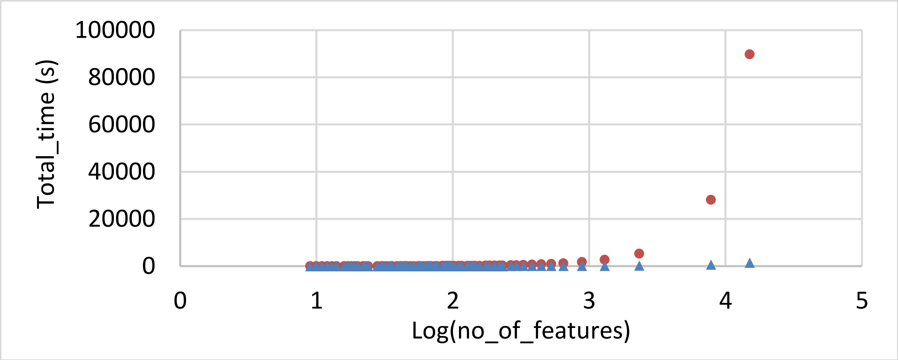

Fig. 2b. Plot of log(no_of_features) vs. total_time for ds_20k

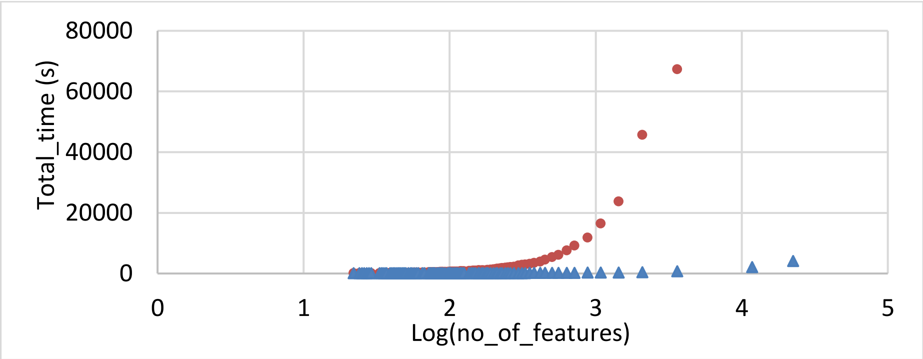

Fig. 2c. Plot of log(no_of_features) vs. total_time for ds_40k

Table 1. Best result for ELM on different datasets

	ELM (5 classifiers)					
Dataset	Number of Features	Accuracy (%)	Precision (%)	Recall (%)	F-score (%)	Total time (s)
ds_10k	517	72.84	71.68	75.52	73.55	85.92
ds_20k	648	74.64	74.27	75.40	74.83	128.73
ds_40k	878	76.69	75.13	79.78	77.39	279.27

Table 2. Best result for SVM on different datasets

	SVM					
Dataset	Number of Features	Accuracy (%)	Precision (%)	Recall (%)	F-score (%)	Total time (s)
ds_10k	233	74.24	72.02	79.28	75.48	48.84
ds_20k	886	75.34	74.06	78.00	75.98	1798.49
ds_40k	1081	76.77	75.09	80.12	77.52	16488.80

In Figs. 1(a)-(c), the x-axis is log(no_of_features) and y-axis is the accuracy of the machine learning algorithm. From these figures, the accuracy of SVM is slightly higher than ELM. However, the accuracy difference between these two algorithms becomes smaller as larger datasets are used. In Figs. 2 (a)-(c), the x-axis is log(no_of_features) and y-axis is the total computing time. With the number of features increased, the computing times for both algorithms also increased.

From Figs. 2 (a)-(c), it is observed that at the starting phase, the computing time for ELM is slightly higher than that for SVM, but SVM spends much more time than ELM when more features are considered or when larger datasets are analyzed. In our experiments, it takes SVM about 16488s while it takes ELM 279s for analyzing the ds_40k dataset. According to the time used, ELM is about 29 times faster than SVM, while they both have a similar performance accuracy of 76.69% for ELM and 76.77% for SVM for analyzing ds_40k dataset as compared in Table 1 and 2.

Moreover, for larger datasets, the computing time for SVM increases significantly as the number of features increases. For SVM on ds_40k, the computing time is so long that we could not get the results for the last two points: we waited for a few days before stopping the program. The accuracy of these two points should not be higher than the points near the starting phase according to the results of the other datasets.

For generalization ability (accuracy), SVM is better for analyzing small datasets. However, the generalization ability of ELM becomes better when larger datasets are used. Thus, from Table 1 and Table 2, we can infer that the difference between the best results of these two methods gets smaller when larger datasets are used. However, the difference between the computing time becomes extremely large: comparing SVM and ELM, when the size of the datasets is large, the computing time for SVM is significantly more than that for ELM – e.g., Fig. 2c, Table 1 and 2, showing about 29 (=16488.80/128.73) times faster for analyzing ds_40k dataset by using ELM.

Moreover, it is also found that there is a linear relationship between log(no_of_features) and accuracy. This will be discussed in a companion research paper.

5 Conclusion

In this paper, the performances of ELM and SVM in the field of sentiment classification are compared in terms of accuracy, precision, recall, f_score as well as computing time. It is found that for the three sets of data used, SVM has better accuracy overall while its computing time is extremely large when more features are considered, especially when the dataset is large. The accuracy of ELM gets better and closer to SVM when larger datasets are used and the computing time maintains relatively short even as the number of features and the size of the dataset are increased significantly. Thus, we conclude that ELM is more suitable than SVM for analyzing large datasets.

In future work, we will compare these two methods using other datasets with different sizes. In addition, we would also like to compare the performance of the various improved ELM algorithms, such as kernel-based ELM, optimization method-based ELM, and regularized ELM.

Acknowledgment

We wish to acknowledge the funding support for this project from Nanyang Technological University under the Undergraduate Research Experience on CAmpus (URECA) programme. The authors would like to thank the support from A*STAR, IHPC for the project "SentiMo - Advanced Social Media Analytics", and project "Advanced Sentiment Analysis of Social Media Data".

References

1. Z. Wang, J. C. Tong, X. Xin, and H. C. Chin, "Anomaly detection through enhanced sentiment analysis on social media data," in 2014 IEEE 6th International Conference on Cloud Computing Technology and Science, 2014, pp. 917–922.
2. Z. Wang, J. C. Tong, P. Ruan, and F. Li, "Lexicon knowledge extraction with sentiment polarity computation," Data Min. Work. (ICDMW), 2016 IEEE 16th Int. Conf., no. 978–983, 2016.
3. X. Liu, C. Gao, and P. Li, "A comparative analysis of support vector machines and extreme learning machines," Neural Networks, vol. 33, pp. 58–66, 2012.
4. W. Huang, Y. Nakamoria, and S.-Y. Wang, "Forecasting stock market movement direction with support vector machine," Comput. Oper. Res., vol. 32, no. 10, pp. 2513–2522, 2005.

5. B. Heisele, Ý. P. Ho, and Þ. T. Poggio, “Face Recognition with Support Vector Machines : Global versus Component-based Approach,” no. April, 2001.
6. K. Polat and S. Güneş, “Breast cancer diagnosis using least square support vector machine,” Digit. Signal Process., vol. 17, no. 4, pp. 694–701, 2007.
7. G. Huang, Q. Zhu, and C. Siew, “Extreme learning machine: a new learning scheme of feedforward neural networks,” Neural Networks, 2004. Proceedings. 2004 IEEE Int. Jt. Conf., vol. 2, no. August 2004, pp. 985–990 vol.2, 2004.
8. P. K. Wong, Z. Yang, C. M. Vong, and J. Zhong, “Real-time fault diagnosis for gas turbine generator systems using extreme learning machine,” Neurocomputing, vol. 128, pp. 249–257, 2014.
9. J. Xu, H. Zhou, and G. Huang, “Extreme Learning Machine Based Fast Object Recognition,” pp. 1490–1496.
10. J. Shin, H. J. Kim, S. Park, and Y. Kim, “Image classification based on effective extreme learning machine,” Neurocomputing, 2009.
11. Z. Wang, J. C. Tong, and H. C. Chin, “Enhancing machine-learning methods for sentiment classification of web data,” Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 8870, pp. 394–405, 2014.
12. O. Kolchyna, T. T. P. Souza, P. Treleaven, and T. Aste, “Twitter Sentiment Analysis: Lexicon Method, Machine Learning Method and Their Combination,” p. 32, 2015.
13. A. Kaur and V. Gupta, “A survey on sentiment analysis and opinion mining techniques,” J. Emerg. Technol. Web Intell., vol. 5, no. 4, pp. 367–372, Nov. 2013.
14. Z. Wang and Y. Parth, “Extreme Learning Machine for Multi-class Sentiment Classification of Tweets,” Proc. ELM-2015, Springer Int. Publ. 2016, vol. 1, pp. 1–11, 2016.
15. “Twitter-sentiment-analyzer data,” Available from: https://github.com/ravikiranj/twitter-sentiment-analyzer/tree/master/data [Cited 4 Sep. 2013]. .
16. E. Haddi, X. Liu, and Y. Shi, “The role of text pre-processing in sentiment analysis,” Procedia Comput. Sci., vol. 17, pp. 26–32, Jan. 2013.
17. J. D. Prusa, T. M. Khoshgoftaar, and D. J. Dittman, “Impact of Feature Selection Techniques for Tweet Sentiment Classification,” Twenty-Eighth Int. Flairs Conf., pp. 299–304, 2015.
18. C. Cortes and V. Vapnik, “Support-Vector Networks,” Mach. Learn., vol. 20, no. 3, pp. 273–297, 1995.
19. C. Chang and C. Lin, “LIBSVM : a library for support vector machines,” Softw. available http//www.csie.ntu.edu.tw/~cjlin/libsvm, pp. 1–39, 2013.
20. A. Akusok, K.-M. Bj"{o}rk, Y. Miche, and A. Lendasse, “High-Performance Extreme Learning Machines : A Complete Toolbox for Big Data Applications,” IEEE Access, vol. 3, pp. 1011–1025, 2015.
21. J. Cao, Z. Lin, G. Huang, and N. Liu, “Voting based extreme learning machine,” Inf. Sci. (Ny)., vol. 185, no. 1, pp. 66–77, 2012.