International Journal of Engineering & Technology, 7 (4.7) (2018) 164-166

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper



A Novel Sentimental Analysis using Optimized Relevance Vector Machine Classifier

N. Saranya[1], R. Gunavathi[2]

[1] *Assistant Professor, PG Department of Computer Science, Sree Saraswathi Thyagaraja College, Pollachi.*
[2] *Head of the Department, Department of Computer Application, Sree Saraswathi Thyagaraja College, Pollachi.*

Abstract

Sentimental analysis is the process of identifying the human's thoughts or feelings. So Many methods have been developed for the sentimental analysis. Machine learning is one of the widely used approaches towards sentiment classification. In this work, Sentimental analysis is done by using Relevance Vector Machine Classifier with Cuckoo Search Optimization. Here Relevance Vector Machine Classifier (RVMC) is combined with Cuckoo Search Optimization (CSO) for better accuracy and performance. Experiment is made with movie and twitter datasets. Accuracy, precision and recall of all other techniques are evaluated. Here the comparison is made among other algorithms. The result shows that RVMC-CSO algorithm gives accuracy and good performance than other algorithm like SVM, ELM and RVM.

Keywords: *Optimization, Vector Machine Classifier, Cuckoo Search Optimization*

1. Introduction

Sentiment analysis or opinion mining is the computational study of people's opinions, appraisals, attitudes, and emotions toward entities, individuals, issues, events, topics and their attributes. It plays a vital role in the research area of text mining. The aim behind the sentiment classification is to analyze the core (subjective) information in the text and then categorizes into one of the three categories positive, negative and neutral. So many approaches have been introduced for the sentimental analysis. These approaches were using Support Vector Machine (SVM), Twin Support Vector Machine (TSVM), Twin Extreme Learning Machines (TELM), Naive Bayes, Neural Networks and Random Forest.

Monitoring and diagnosis in mechanical systems is of particular importance in preventing serious accidents in equipping and preventing unintended stops. The use of vibrations and telephonic playback signals is very common in the context of the condition of rotating machines. By comparing the signals of a single engine in normal operation and defective conditions, there can be some defects such as non-balancing, rotor wear, axial unevenness, gearbox failure and bearing defects.

The remainder of this paper is organized as follows. Section2 contains the review of some related work. Section3 presents sentimental analysis using RVMC-CSO based. Section 4 contains simulation results. Finally, Section 5 concludes this paper.

2. Related Works

Contrary to neural networks, it does not respond to local maximum levels, it responds to good-sized data for a relatively good fit, but the main difference between the popular neural networks and the backup machine is that in the neural network the parameters of this model are set with an error minimization, but in a risk-based vector machine The result of the lack of proper classification is defined as a function function and the parameters are adjusted and optimized. For some issues, the error criterion may be zero, but among all models with zero error, there is only one case that has the least operational risk. Thus, in some cases, the output of the backup machine, along with better performance, will show a greater resistance of 1 to noise and changeover. Because it is essentially designed and trained to be able to withstand such uncertainties and have its own optimal performance [4]. The use of a backup vector machine to troubleshoot rotating machines with vibration data in recent years has been of interest to researchers.

Tipping (2000) has proposed the relevance vector machine. Absence of probability output is the disadvantage of SVM. So, relevance vector machine is introduced to avoid disadvantage of output. This work demonstrated that the RVM machine can attain a comparable level of generalization accuracy as the well-established support vector approach. Advantage of the RVM classifier is its standard formulation as a probabilistic generalized linear model. Taboada (2011) have introduced Lexicon-based methods for sentiment analysis.

Bishop and Tipping (2000) have proposed variational relevance vector machines. This paper showed, through the use of variation inference, how the RVM can be formulated and solved within a completely Bayesian paradigm thereby giving a posterior distribution over both parameters and hyper parameters. The performance of the variational RVM is demonstrated practically.

Zainuddin and Selamat (2014) have proposed sentimental analysis using SVM. The two datasets Pang Corpus and Taboada Corpus were used. N-grams and different weighting scheme were used as an input to the sentiment classifier.

The basic theory of RVM for regression and classification were explained here with two examples illustrated the application of RVM for object detection and classification. Examples were target detection and microcalcifications detection and classification .RVM used a regression context and classification framework.

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Zou et al (2018) have proposed a new method combining social context and topic context to analyze microblog sentiment. Here structure similarity matrix built by introducing structure similarity context into social contexts and topic context matrix built by topic context. Social context and topic context are combined by the Laplacian matrix of the graph. This method has a better performance and accuracy. Huang et al (2014) have described Asymmetric least squares support vector machine (aLS-SVM) classifiers.

Khan et al (2015) have proposed P2P (Peer-to-Peer) RVM for Distributed Classification. Here a distributed classification technique based on RVM and local model exchange among neighboringpeers in a P2P network is introduced. Jadav and Vaghela (2016) have described support vector machine based on feature selection and semantic analysis. Here unstructured data is converted into structured form. Then lexicon based approach is used to convert the structured review into numerical score value. Finally SVM algorithm is applied for the classification of reviews where RBF kernel SVM is modified by its hyper parameters.

BholaneSavita and Gore (2016) used SVM for twitter data. Here for tracking of sentiment variation, latent dirichlet based approach is used. This approach proved that SVM increased accuracy 23.24% than others. Hoang and Tien Bui (2016) have proposed a novel relevance vector machine classifier with cuckoo search optimization for spatial prediction of landslides. Saranya and Gunavathi (2017) have proposed least squares twin support vector machine for sentiment analysis. Pestian et al (2012) have introduced Sentiment analysis of suicide notes.

4. Simulation and Results

Here Polarity movie review dataset is used [14]. Separate text file is maintained for each review. Twitter dataset is also taken to show effect of proposed method on different dataset. Twitter dataset is taken from twitter API [14].

Table 2 and 3 shows the confusion matrix for small movie and twitter dataset after implementing SVM algorithm.

Table 2:. Confusion Matrix for Movie Dataset

	Correct Labels	
	Positive	Negative
Positive	5236	0
Negative	1564	0

Table 3:. Confusion Matrix for Twitter Dataset

	Correct Labels	
	Positive	Negative
Positive	3210	0
Negative	950	0

Table 4 and 5 shows confusion matrix for small movie and twitter dataset after implementing ELM algorithm.

Table 4:. Confusion Matrix for Movie Dataset

	Correct Labels	
	Positive	Negative
Positive	5687	34
Negative	1745	16

Table 5:. Confusion Matrix for Twitter Dataset

	Correct Labels	
	Positive	Negative
Positive	2897	45
Negative	876	36

Table 6 and 7 shows confusion matrix for small movie and twitter dataset after implementing RVM algorithm.

Table 6:. Confusion Matrix for Movie Dataset

	Correct Labels	
	Positive	Negative
Positive	5489	0
Negative	1765	0

Table 7:. Confusion Matrix for Twitter Dataset

	Correct Labels	
	Positive	Negative
Positive	2574	26
Negative	879	76

Table 8 and 9 shows confusion matrix for small movie and twitter dataset after implementing RVMC-CSO algorithm.

Table 8:. Confusion Matrix for Movie Dataset

	Correct Labels	
	Positive	Negative
Positive	6489	0
Negative	2765	0

Table 9:. Confusion Matrix for Twitter Dataset

	Correct Labels	
	Positive	Negative
Positive	2874	26
Negative	979	66

4.1 Graphical Representation

This section represents comparison of proposed RVMC-SCO, RVM, ELM and SVM. Table 10 shows that RVMC-SCO has good accuracy, precision and recall than others.

Table 10:. Comparison of RVMC-SCO, RVM, ELM and SVM

Dataset	RVMC-CSO			RVM			ELM			SVM		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall	Accuracy	Precision	Recall	Accuracy	Precision	Recall
Movie	86.87	89.56	88.37	82.33	86.45	84.67	77.33	79.44	78.12	75.68	77.27	76.65
Twitter	84.65	88.76	87.54	81.43	85.88	84.29	76.82	76.45	76.42	74.56	74.83	74.81

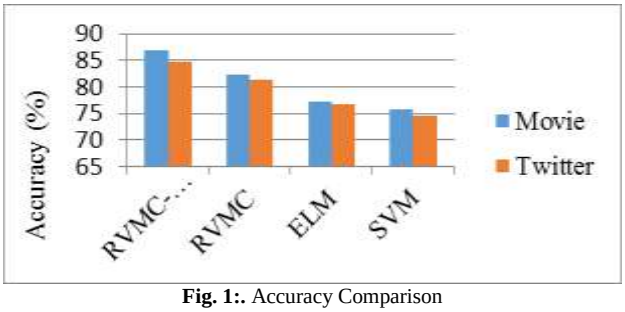


Fig. 1:. Accuracy Comparison

By using the confusion matrix dataset, accuracy of RVMC-CSO, RVM, ELM and SVM were calculated for both movie and twitter. The comparison of accuracy for RVMC-CSO, RVM, ELM and SVM are plotted in the chart. Above Fig.1 shows that RVMC-CSO has the accuracy of 86.87% for movie and 84.65% for twitter .Hence it is proved that RVMC-CSO has good accuracy than others.

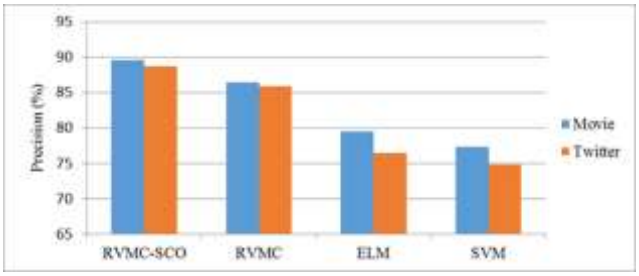


Fig. 2:. Precision Comparison

By using the confusion matrix dataset, precision of RVMC-CSO, RVM, ELM and SVM were calculated for both movie and twitter. The comparison of precision for RVMC-CSO, RVM, ELM and SVM are plotted in the chart. Fig.2 shows that RVMC-CSO has the precision of 89.56% for movie and 88.76% for twitter. Hence it is proved that RVMC-CSO has better precision than others.

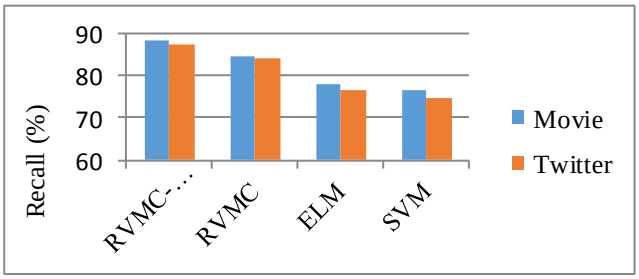


Fig. 3:. Recall Comparison

By using the confusion matrix dataset, recall of RVMC-CSO, RVM, ELM and SVM were calculated for both movie and twitter. The comparison of recall for RVMC-CSO, RVM, ELM and SVM are plotted in the chart. Fig.3 shows that RVMC-CSO has the recall of 88.37% for movie and 87.54% for twitter. Hence it is proved that RVMC-CSO has good recall than others.

5. Conclusion

New approach is invented for sentimental analysis. Relevance vector machine classifier with cuckoo search optimization (RVMC-CSO) is used. RVMC is a machine learning approach used to classify the input and CSO is incorporated into the model used to fine tune the RVMC's free factor that is the RBF. Due to combination of two algorithms it results better than other techniques. The experiment is done by using movie and twitter datasets. The comparison is made among RVMC-CSO, RVM, ELM and SVM. Feature extraction is an important step in defect classification process, because it can prevent redundancy of information and problems related to incremental dimensions. After extracting features, the feature selection process was used to eliminate unrelated and exaggerated features. The distance criterion technique has been used as a simple and reliable method for this purpose. Multi-classification of backup vector machine was used for defect classification process. Accuracy, precision and recall of all algorithms are calculated. Result shown that it gives good performance and accuracy when compare to other traditional techniques.

References

[1] Hwang, S. and Jeong, M.K., 2018. Robust relevance vector machine for classification with variational inference. Annals of Operations Research, 263(1-2), pp.21-43.

[2] Yu, H., Wu, Z., Chen, D. and Ma, X., 2017. Probabilistic prediction of bus headway using relevance vector machine regression. IEEE Transactions on Intelligent Transportation Systems, 18(7), pp.1772-1781.

[3] Tipping, M.E., 2000. The relevance vector machine. In Advances in neural information processing systems (pp. 652-658).

[4] Foody, G.M., 2008. RVM-based multi-class classification of remotely sensed data. International Journal of Remote Sensing, 29(6), pp.1817-1823.

[5] Rafi, M. and Shaikh, M.S., 2013. A comparison of SVM and RVM for Document Classification. arXiv preprint arXiv:1301.2785.

[6] P. Byuvol, L. Gabsalikhova, I. Makarova, E. Mukhametdinov, G. Sadygova, "Improving the Branded Service Network Efficiency based on its Functioning Evaluation", Astra Salvensis, Supplement No. 2, p. 373, 2017.

[7] Azad N., Ghandvar P., Rahimi Z., "Online Search Behaviour of Customers in Shoe Market", Astra Salvensis, Supplement No. 2, p. 793, 2017.

[8] Tzikas, D.G., Wei, L., Likas, A., Yang, Y. and Galatsanos, N.P., 2006. A tutorial on relevance vector machines for regression and classification with applications. EURASIP News Letter, 17(2), p.4.

[9] Zou, X., Yang, J. and Zhang, J., 2018. Microblog sentiment analysis using social and topic context. PloS one, 13(2), p.e0191163.

[10] Khan, M.U., Nanopoulos, A. and Schmidt-Thieme, L., 2015. P2P RVM for Distributed Classification. In Data Science, Learning by Latent Structures, and Knowledge Discovery (pp. 145-155). Springer, Berlin, Heidelberg.

[11] Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., Mohammad, A.S., Al-Ayyoub, M., Zhao, Y., Qin, B., De Clercq, O. and Hoste, V., 2016. SemEval-2016 task 5: Aspect based sentiment analysis. In Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016) (pp. 19-30).

[12] Bowd, C., Medeiros, F.A., Zhang, Z., Zangwill, L.M., Hao, J., Lee, T.W., Sejnowski, T.J., Weinreb, R.N. and Goldbaum, M.H., 2005. Relevance vector machine and support vector machine classifier analysis of scanning laser polarimetry retinal nerve fiber layer measurements. Investigative ophthalmology & visual science, 46(4), pp.1322-1329.

[13] Jadav, B.M. and Vaghela, V.B., 2016. Sentiment analysis using support vector machine based on feature selection and semantic analysis. International Journal of Computer Applications, 146(13).

[14] Huang, X., Shi, L. and Suykens, J.A., 2014. Asymmetric least squares support vector machine classifiers. Computational Statistics & Data Analysis, 70, pp.395-405.

[15] BholaneSavita, D. and Gore, D., 2016.Sentiment Analysis on Twitter Data Using Support Vector Machine. International Journal of Computer Science Trends and Technology (IJCST)–Volume, 4, p.365.

[16] Saranya, N. and Gunavathi, R., 2017. Sentimental analysis using least squares twin support vector machine. International Journal of Advanced Research in Computer Science, 8(7).

[17] Hoang, N.D. and Tien Bui, D., 2016. A novel relevance vector machine classifier with cuckoo search optimization for spatial prediction of landslides. Journal of Computing in Civil Engineering, 30(5), p.04016001.

[18] Taboada, M., Brooke, J., Tofiloski, M., Voll, K. and Stede, M., 2011. Lexicon-based methods for sentiment analysis. Computational linguistics, 37(2), pp.267-307.

[19] Ceron, A., Curini, L., Iacus, S.M. and Porro, G., 2014. Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France. New Media & Society, 16(2), pp.340-358.

[20] Pestian, J.P., Matykiewicz, P., Linn-Gust, M., South, B., Uzuner, O., Wiebe, J., Cohen, K.B., Hurdle, J. and Brew, C., 2012. Sentiment analysis of suicide notes: A shared task. Biomedical informatics insights, 5, pp.BII-S9042.