

Twitter sentiment analysis using different method -A Survey

Vaishnavi D. Deshmukh, Dr. G. K. Pakle

Dept. of Information Technology

Shri Guru Gobind Singhaji Institute of Engineering and Technology 431606, Nanded
(MH)

Abstract— Sentiment analysis is the most leading data mining fields that get used for the recognition and analysis of sentimental content commonly accessible on social media. Twitter is a most used social networks many users use in the form of tweets on certain topics. By using clustering-based methods, these tweets get be analyzed with classification to find the feelings and viewpoints of the users. Met heuristic-based clustering approaches, outperform conventional methods for sentiment analysis due to the subjective nature of the Twitter datasets. A novel met heuristic method i.e. CSK based on cuckoo search and K-means are therefore proposed. The proposed method was used to identify the best cluster heads from the nostalgic content of the Twitter dataset. The effectiveness of the suggested approach was tested on various Twitter databases and contrasted with particle swarm optimization, differential evolution, cuckoo search, improved cuckoo search, gauss-based cuckoo search, and two n-grams methods. To classify the experimental and mathematical findings of the new approaches. The suggested

approach has theoretical implications for analyzing the data generated by social networks / media for future analysis. In developing a framework that can include definitive feedback on certain social problems, this approach also has very generalized functional information.

Index Terms-

Sentiment analysis, K-mean Cuckoo search, SVM, Twitter Data pre-processing

I. INTRODUCTION

Sentiment analysis is a process of categorizing and identifying opinion behind in a text or data. It is a part of Natural language processing. In which machine learning techniques and natural language processing techniques gets combine to assign polarity or sentiments to a topics or text data. Analysis of sentiment is the most important fields of mining the data which is used for analysis, classification, identification of sentimental content and data which are present and available on social media. It is mostly useful for social media classification and for helping

businesses to collect useful information from its unstructured text by automatically tagging.



Figure 1: Sentiment analysis

Methods of sentiment analysis can be narrowly categorized into methods based on lexicons, methods based on machine learning, and mixed methods that can be further categorized into sub-categories, as seen in Figure 2. To evaluate the polarity of any text, lexicon-based methods include predefined sentiment lexicons. In the case of emoticons and shorthand notes, the precision of the lexicon-based approach is significantly diminished, whereas they are not part of the predefined expression lexicon. In Machine based learning is a technique that train computer to do automatic tasks on coming data. Machines automatically learn how to identify sentiment without human input by teaching machine learning tools with examples of emotions in the text. It performs machine learning techniques to get output from data in positive, negative or neutral and it get considered as classification of supervised tasks. It has various classification algorithms as Navie Bayes, Linear regression, support vector machines and deep learning. For polarity identification, the Hybrid approach uses both mathematical approaches and knowledge-based approaches. It

inherits the high precision of machine learning or computational approaches and the stability of the method based on the lexicon. It is based on both rule-based and automatic approach for sentiment or polarity classification because of that the results are more accurate.

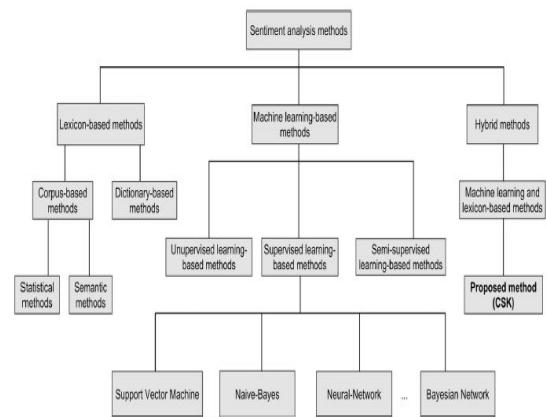


Figure 2: Sentiment analysis methods

Twitter is one of the most used online media with more than 1.3 billion followers and 500 million tweets daily. Also Twitter has the biggest network of 331 million+ active users a day. Its data will help many companies or business to understand how their users are reacting or talking about their services or brand. By using it twitter allows businesses and companies to identify audience opinion and understand them. By using Sentiment analysis we will find out emotions behind the tweet in the form of positive, negative and neutral by using the sentiment analysis methods. This sentiment analysis can process text data automatically by using machine learning, twitter API and display results in the form of positive, negative and neutral sentiments. We can process sentiment analysis on twitter dataset by comparing

sentiments and on real time data by giving hash tag or keywords.

II.PRIVIOUS WORK

This exploits both statistical approaches[1] and information-based approaches of polarity detection. It inherits the high precision of machine learning or computational approaches and the stability of the method based on the lexicon. As the findings are more accurate, it is based on both the rule-based and automatic approach to classification of emotion or polarity.

Opinion mining model[2] for movie review by using hybrid method of support vector machine (SVM) and particle swarm optimization(PSO). SVM is a supervised learning system used to evaluate data and to distinguish patterns in two groups, positive and negative, and the PSO is used to improve the choice of the appropriate parameter to solve the problem of optimization. 71.8% to 77% consistency is given.

A Twitter opinion lexicon model[3] from automatically linked tweets in which three types of probability distribution of feelings collect tweets and suggested a new way of computing word level attributes with soft labels shape data. Based on the morphological information get conveyed by POS tags and the associations expressed in the tweets containing them between words and feelings. They shows that 3-dimensional word level sentiment classification.

For the stock market, Twitter mood predicts model[4], examining the measurements of

collected mood states extracted from large-scale Twitter feeds associated with the value over time of the Dow Jones Industrial average. Two moods for daily content are measured by the OpinionFinder twitter monitoring tool for positive and negative moods and Google Profile mood for six-dimensional measurements. (Calm, Alert, Sure, Vital, Kind, and Happy). A Granger causality analysis and a self-organizing Fuzzy Neural Network are then used to analyze the hypothesis, as determined by the two moods, that state public mood. In which 85% accuracy in prediction is given. A ontology based sentiment analysis[5] for twitter posts, they used Micro-blogging application of web 2.0 and its related services for twitter data accessing data sources for mining and sentiments. It also use ontology for modelling terms in domain of internet, its relations among services and terms and for extracting information from text, It serve primary means of knowledge representation in web 2.0 by creating domain ontology.

Sarcastic sentiment detection[6] used in tweets for real time for accessing real time tweets. Sarcasm identification and its study offers useful insight into occurrences and developments in social media for public real-time perception. Six algorithms are used to detect sarcasm in tweets: PBLGA, IWS, PSWAP, TCUF, TCTDF, and LDC. Three algorithms run with and without the Hadoop system and the timing of the method decreases by up to 66%.

Cuckoo quest algorithms[7] to solve optimization problems by preserving a balance

between global and local random walks by using swapping parameters and resolving it at 25% due to its usefulness in global optimization. These By distinct probability distribution, algorithms increase performance and are tested on certain empirical test functions and results are found to be more effective than other algorithms relative to their constant and dynamically growing switching parameters.

Cuckoo search method on natural and biological influenced computation through levy flight[10]. This algorithm is based on the obligatory brood parasite behavior of some cuckoo species, in conjunction with the Levy flight behavior of some birds and fruit flies. The proposed algorithm is tested against test functions, and its performance is then compared with that of genetic algorithms and particle swarm optimization. It can effectively extend its theoretically successful optimization approach to explore multi-objective optimization applications with multiple restrictions, including NP-hard problems.

A domain of sentiment lexicon [11] through double propagation. They suggested an approach to domain sentiment word extraction based on the propagation of both the established lexicon of sentiment and extracted product characteristics, which we call double propagation. In order to catch the correlation between emotional characteristics and concepts, as well as concepts and feeling characteristics, the algorithm uses dependence relations. To

remove feeling words and features offered by current feeling words, many methodological rules are organized. They also suggested a new approach to assigning polarities to words of feeling that were derived.

III TECHNIQUES FOR ANALYZING FUNDUS IMAGE PATTERNS

In this section, We present a description of the various machine learning approaches used for the classification of emotion and sentiment analysis. In Cuckoo search methods we perform hybrid cuckoo search method on dataset for feature extraction after that we perform mean, SVM, probability etc. We apply some sentiment analysis and classifications methods for finding sentiments behind different data. For classifying movie review opinion mining get used with hybrid methods of support vector machine and the particle swarm optimization.

Techniques	Tasks	Accuracy
ConceptNet based ontology	It is used for domain polarity which in domain specific important features for polarity detection.	69.2% , 71.3% and 68.3% For different domain
Support vector machine	It is used for analysis dada for pattern classification.	71.8% to 77%
Twitter opinion lexicon	Automatically get connected with tweets and	79%

model	for words probability.	
Fuzzy neural network	For investigating hypothesis that state the public mood and calculate mood.	85%
PBLGA, IWS,PSWP ,TCUF, TCTDF, and LDC	For detecting sarcasm in tweets for sentiments	66%
Naive Bayes	For words probability and sentiment classification	86%

Figure 3: Summary of various Techniques for Analysis

IV CONCLUSION

This study has successfully developed A novel hybrid clustering approach (CSK) was presented in this paper to examine the feelings of tweets using CS and K-means methods. The proposed method modifies the process of CS random initialization through the solutions obtained from K-means that enhance its performance. SVM is a supervised learning system used to evaluate data and to distinguish patterns in two groups, positive and negative, and the PSO is used to improve the choice of the appropriate parameter to solve the problem of optimization. Additionally, the process was compared to CS, ICS, GCS, PSO, DE, SVM-tri, and NB-tri. For a clearer reference, for all the considered datasets, student t-test, box diagram, and convergence diagram analysis have also been carried out. The

viability of the proposed solution was discovered in terms of experimental and mathematical data. The suggested approach indicates increased accuracy compared to current methods, and accuracy enhancements are also needed. Further studies will also involve investigating the possibilities of optimizing accuracy by incorporating certain methods of feature selection and adding various variants of methods of optimization. We use machine learning to increase accuracy and maximize the potency of emotions.

ACKNOWLEDGMENT

I would like to thank to my guides Dr. G. K. Pakle, Associate Professor from Shri Guru Gobind Singhji Institute of Engineering and Technology, Nanded for their encouragement, continuous support, and enhancing me for this work.

REFERENCES

- [1] Agarwal, B. , Mittal, N. , Bansal, P. , & Garg, S. (2015). Sentiment analysis using common-sense and context information. *Computational Intelligence and Neuro- science*, 2015 , 30 .
- [2] Basari, A. S. H. , Hussin, B. , Ananta, I. G. P. , & Zeniarja, J. (2013). Opinion mining of movie review using hybrid method of support vector machine and particle swarm optimization. *Procedia Engineering*, 53 , 453–462 .
- [3] Boiy, E. , Hens, P. , Deschacht, K. , & Moens, M.-F. (2007). Automatic sentiment

analysis in on-line text. In *ELPUB* (pp. 349–360).

[4] Bollen, J. , Mao, H. , & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2 , 1–8 .

[5] Chiang, M. M.-T. , & Mirkin, B. (2010). Intelligent choice of the number of clusters in k- means clustering: An experimental study with different cluster spreads. *Journal of Classification*, 27 (1), 3–40 .

[6] Medhat, W. , Hassan, A. , & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5 (4), 1093–1113 .

[7] Yusof, N. N. , Mohamed, A. , & Abdul Rahman, S. (2015). Reviewing classification approaches in sentiment analysis. In *International conference on soft computing in data science* (pp. 43–53). Springer .

[8] Zalik, K. R. (2008). An efficient k-means clustering algorithm. *Pattern Recognition Letters*, Vol. 29 , 1385–1391 .

[9] Pandarachalil, R. , Sendhilkumar, S. , & Mahalakshmi, G. (2015). Twitter sentiment analysis for large-scale data: An unsupervised approach. *Cognitive Computation*, 7 , 254– 262 .19

[10] Chiang, M. M.-T. , & Mirkin, B. (2010). Intelligent choice of the number of clusters in k-means clustering: An experimental study with different cluster spreads. *Journal of Classification*, 27 (1), 3–40 .

[11] Chen, T. , Xu, R. , He, Y. , Xia, Y. , & Wang, X. (2016). Learning user and product

distributed representations using a sequence model for sentiment analysis. *IEEE Computational Intelligence Magazine*, 11 (3), 34–44 .

[12] Cambria, E. (2016). Affective computing and sentiment analysis. *IEEE Intelligent Systems*, Vol. 31 , 102–107 .

[13] Canuto, S. , Gonçalves, M. A. , & Benevenuto, F. (2016). Exploiting new sentiment-based meta-level features for effective sentiment analysis. In *Proceedings of the ninth ACM international conference on web search and data mining* (pp. 53–62). ACM .

[14] Bravo-Marquez, F. , Frank, E. , & Pfahringer, B. (2016). Building a twitter opinion lexicon from automatically-annotated tweets. *Knowledge-Based Systems* .

[15] Bravo-Marquez, F. , Mendoza, M. , & Poblete, B. (2013). Combining strengths, emotions and polarities for boosting twitter sentiment analysis. In *Proceedings of the second international workshop on issues of sentiment discovery and opinion mining* (p. 2). ACM .

[16] Gupta, D. K. , Reddy, K. S. , Ekbal, A. , et al. (2015). Pso-aset: Feature selection using particle swarm optimization for aspect based sentiment analysis. In *Natural language processing and information systems* (pp. 220–233). Springer .

[17] Bharti, S. , Vachha, B. , Pradhan, R. , Babu, K. , & Jena, S. (2016). Sarcastic sentiment detection in tweets streamed in real time: A big data approach. *Digital Communications and Networks*, 2 (3), 108–121 .

[18] Zhu, J. , Wang, H. , & Mao, J. (2010). Sentiment classification using genetic algorithm and conditional random fields. In *2nd IEEE international conference on information management and engineering (ICIME)* (pp. 193–196). IEEE .20

[19] Zheng, H. , & Zhou, Y. (2012). A novel cuckoo search optimization algorithm based on Gauss distribution. *Journal of Computational Information Systems*, 8 , 4193–4200 .