IOP Conference Series: Materials Science and Engineering

PAPER • OPEN ACCESS

Aspect level sentiment analysis using machine learning

To cite this article: D Shubham *et al* 2017 *IOP Conf. Ser.: Mater. Sci. Eng.* **263** 042009

View the article online for updates and enhancements.

Related content

- A Framework for Sentiment Analysis Implementation of Indonesian Language Tweet on Twitter
  Asniar and B R Aditya
- Sentiment analysis in twitter data using data analytic techniques for predictive modelling
  A Razia Sulthana, A K Jaithunbi and L Sai Ramesh
- A Peculiar Sentiment Analysis Advancement in Big Data
  Manisha Valera and Yash Patel

This content was downloaded from IP address 128.111.121.42 on 29/05/2018 at 00:31

Aspect level sentiment analysis using machine learning

Shubham D, Mithil P, Meesala Shobharani, Sumathy S*
School of Information Technology and Engineering, VIT University, Vellore-632014, India.

*Email : ssumathy@vit.ac.in

Abstract. In modern world the development of web and smartphones increases the usage of online shopping. The overall feedback about product is generated with the help of sentiment analysis using text processing.Opinion mining or sentiment analysis is used to collect and categorized the reviews of product. The proposed system uses aspect leveldetection in which features are extracted from the datasets. The system performs pre-processing operation such as tokenization, part of speech and limitization on the data tofinds meaningful information which is used to detect the polarity level and assigns rating to product. The proposed model focuses on aspects to produces accurate result by avoiding the spam reviews.

1. Introduction

In modern world the users prefer online shopping because of various options and many facilities such as discounts, cash backs etc. The user reads reviews to decide whether product meets customer expectation in terms of quality or not. Reading thousands of reviews wastes lots of time of user. Sentiment analysis collects the data and classifies into positive and negative reviews. The pre-processing operation is performed on data after collection of data which includes tokenization process. The reviews are analyzed in terms of words such as good, bad, poor, etc which is used classify the reviews into positive and negative reviews. The mining results assign overall rating to product which is taken into consideration by users while purchasing any products. Aspect based opinion mining is most widely used in which the features is extracted from text. Supervised training mechanism is used to train the system for better accuracy. It uses machine learning techniques to train the system about patterns of reviews. Supervised learning algorithm such as SVM, naive bayes is used to classify the reviews into positive and negative classes based on training data.

The major challenge in opinion mining is detection of spam reviews which affects the accuracy level of mining results. The accuracy level depends on the training data which is passed to SVM model. The elimination of spam reviews increases the accuracy level of system. The meta-data is used to detect the fake reviews about product. The meta data includes the information about users and orders. The historical information is used to analyze the behaviour of user for detecting spam reviews.

The paper is organized as follows: section 2 presents the literature review, section 3 describes the proposed methodology, with results and discussion in section 4 and conclusion in section 5.

2. Literature review

Abd.Samad Hasan Basari et al. hybrid method based on support vector for opinion mining.SVM uses supervised learning method to analyze the patterns which are used for classification. The opinion is classified into two classes: positive and negative reviews. Dual optimization problem is solved by using hybrid particle swarm optimization. Sentient analysis classification is done using SVM and result is compared with n-grams and feature weighting. The result shows that in proposed solution accuracy is improved upto 77%.Lei, X., Qian et al. proposed a paper on for sentimental analysis

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

according to rating provided to the product by the user. The system works firstly on the user sentiments provided socially for any product, secondly with the product features that relate with the user sentiments and thirdly with the product status for collecting the user sentiments after proper use. Finally the paper merges all the three aspects to get an accurate prediction for the product using ratings. The result after evaluation over the dataset (Yelp) is increases accuracy.

DikshaSahni surveys the dataset which identifies the human sentiments by categorizing them as the text. The survey contains different types of emotion model with different dataset having different accuracy. Such as survey for vector space model which worked on chatting dataset and had 44.70 % of lowest accuracy, naïve bayes model which worked on blogs, tweets dataset and had 84.36% highest of accuracy. The research also includes other model like decision tree model, kbann model for detection of the sentiments. Rudy Prabowo et al. proposed system which combines three approach first the instrument learning which uses self-made programs for recognizing human sentiments, second is controlled learning that recognizes the human emotions by passing concluding training data sets and the third is the instruction based classification which specifies some rules for calculating human sentiments. The combination of the tree approach shows great improvement for small and average amount of dataset.

Susan Thomas describes mythology for hotel suggestion to the customer as per the requirements. The proposed system suggests the hostel to customer based on same search history. Suggestion is provided to user on the bases of the keywords used, if the same keywords are used by another customer, then the customer is recommended by the hotel. But while providing suggestion of hotel the keywords are checked weather the review given by customer for the hotel is good or bad.Pavlopoulos et al. describes system for sentimental analysis which works on the certain set of documents or particular feature of the product. The proposed system aspect base analysis system is divided into three task first is the withdrawal of features, second is making cluster according to aspect and third is to rate the aspect. After combining all the three aspects the performance of the system over the dataset they trained.

Bin Lu et al. proposed a weakly monitoring system that uses of the dataset based with the less previous information to be rate it. Based on this rating the system combines the sentence to get the better performance than the strongly monitoring system. The system can also handle the many more aspect based features. The proposed work produces result as compared to other model likes SVR. Kalyani D. et al. proposed paper is a survey that works on opinion mining that aims to divide the text as per sentiments. The paper also proposed a model that classifies the tweet dataset as positive or negative. The system takes input as opinions from various social websites, extract its features and apply sentence opinion to get result. The system works on RSS that is on the documents from various popular sites.

Deepali Virmani et al. proposed method for sentimental analysis to identify the level of data. The algorithm combines sentimental analysis with opinion mining to get more accurate result in bottom-up manner for a student data set. The proposed system apply algorithm on dataset, which are the remarks given by the teacher and tries to allot rating for each student. The algorithm works on each word of which is taken as opinion and tries to allot rating to archive better correctness. Saprativa Bhattacharjee et al. describe idea for rating the comments over the social web sites. The system is uses Cosine Relationship degree which provides -1 and -2 as negative rating, 0 as neutral and 1 and 2 as positive rating. The results are compared with many sentimental analysis algorithms to check accuracy and get the accuracy of 82.09 %. The system takes the dataset from the telecom service supplier which consists of 8 thousand comments and some web documents were also taken.

Sakshi Gupta et al. describe a method for sentiment analysis which uses facebook comments using vector technique. The proposed work on database generation takes two steps, first is passing a single comment which takes more time. Second is passing a bulk of data were single data is uploaded at a time which takes less time. The comments are rated as positive (+), negative (-) and neutral after passing it through two stage of data pre-processing were data is filtered and data testing were data is rated.

Deepak Kumar Gupta et al. proposed a model to detect the spam review on social web sites like twitter. The three major goals for the paper are clustering the spam review, second is the clustering of sentiments and third is calculate the sentiments and spam review on the bases of clusters made. R programming code is use for detection of fake review over the dataset and log ratio is use for sentimental analysis. Fangtao li et al. proposed method to detect spam reviews. The system uses its own created spam review dataset. The system keeps on monitoring the spammer, to detect fake review by having eye on the authors and check whether the author is a spammer or not. System uses two-view co-training method to monitor the huge unknown or untagged dataset. The result shows great outcome for detecting the fake review.

Poobana et al. describes system that gathers the product review from the various social websites for rating those review. The system uses Naive base and SVM algorithm for rating review. The system collect user review on product perform data processing to remove noise in two stage, first is removal of stop words and second is stemming words to get unigram result. The processed data is analyzed and classified based on opinion mining algorithm to get the final result. Jayashri Khairnar et al. describe the machine learning algorithm through explaining the working principle and its derivations. The result are calculate using the factors like accuracy which is calculated using the number of true values beside all assumption, precision is calculate using true value besides positive assumption, recall is calculated using positive values besides all real values and harmonic value is calculated using precision and recall. The true positive values are taken from confusion matrix.

Ahmad Kamal proposed system is based on various machine learning algorithm to produce accurate opinion mining result. The proposed system collect data over internet, pre-processing of data is done using tokenization; the pre-processed data is later moved to subjective classification using rule-base and supervised machine learning system. After classification the data is rated using opinion mining algorithms like SVM, naive base. Later probability analysis is done to get the final result. A comparative analysis of different methods is given in Table 1.

Table 1. Comparative analysis of different methods

Ref no.	Methodology	Advantages
1.	• Hybrid base opinion mining. • SVM to analyze pattern.	Accuracy is improved up to 77%.
2.	• Sentimental analysis for product review.	Accuracy of dataset (Yelp) is improved.
3	• Algorithm for recognizing human sentiments. • Instruction based classification.	Tree approach shows great improvement for small and average amount of dataset.
4	• Mythology for hotel suggestion using keywords.	Customer is provided with hotel suggestion by matching the keywords.
5	• Aspect base analysis system for a specific feature of product.	Clustering of dataset improves the performance.
6	• Weakly monitoring system using rating	Works better for small dataset.

	for small dataset.	
7	• Opinion mining using sentiment analyses techniques.	Paper has good survey as well as implemented opinion mining system.
8	• Combines sentimental analysis with opinion mining in bottom-up manner.	System works for each work and provide better rating.
9	• Cosine Relationship degree for rating review. • sentimental analysis algorithms to classify	System provides 82.09% of accuracy.
10	• Sentiment analysis using vector technique.	Data preprocessing and filtered data provide good data testing for rating.
11	• Detect the spam using R tool. • Log ratio is use for sentimental analysis.	Spam review is detected.
12	• Detect fake review by having eye on the authors.	Helps for detecting spammer.
13	• Naive base and SVM algorithm for rating review.	System collect user review on product performs data processing to remove noise.
14	• Machine learning algorithm	Confusion matrix is used for rating purpose.
15	• Opinion mining algorithms like SVM, naive base tokenization,and data preprocessing method.	Review is rated using machine learning algorithm.

3. Proposed architecture

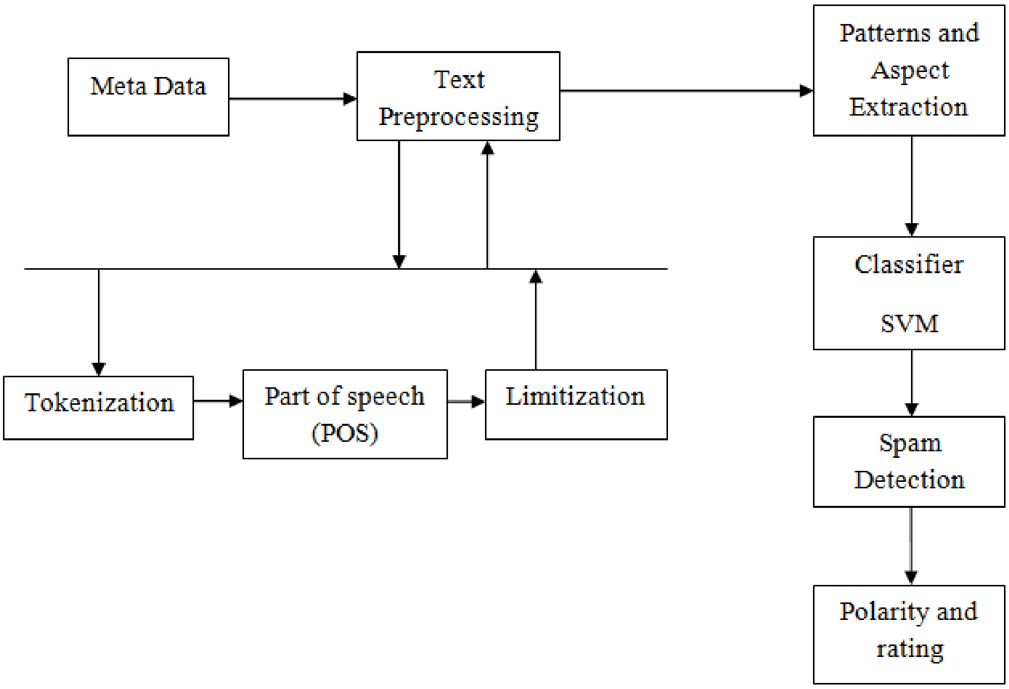


Fig.1: Proposed Architecture

Proposed architecture is given in Fig.1. The description of each mode is given below.

Datasets: The reviews or opinion is collected from online shopping website or social networking website to mine the data and classifies the overall review into positive or negative categories. The datasets plays important role in training of system.

Text pre-processing: The text pre-processing module performs certain operations on text to remove the noise and arranges the content to detect the aspect from the text in order to classify it into positive and negative reviews. The module is divided into three sub parts part of speech, limitization and tokenization. After performing operation on text the data is passed to spam detection module.

Tokenization: Tokenization process divides stream or complete sentence into symbols or meaningful words called tokens. These meaningful elements are passes to next block to extract the patterns.

Part of speech: It is the process marking up text according to part of speech. It uses noun, verb, adverb, adjectives for identification of words.

Limitization: Limitization is used to understand the abbreviations of word which are used in reviews.

Spam Detection: It collects the data from text processing block and meta data and performs certain operations to detect the spam reviews. The elimination o spam reviews increases the accuracy of classification results. Meta data is used to track the behaviour of the user.

Classification: The system uses SVM or naïve based algorithm to classify the reviews. Opinion mining uses three types of approach sentence level, document level and aspect level. The proposed system uses aspect level approach for opinion mining because of high accuracy. Sentence level and document level analyze the data and generates rating and classifies the reviews. Aspect level method gives better accuracy because its focuses on words such as good, bad, worst etc.

Validation: The block assigns polarity level to product based on rating. The level of polarity is assigned based on rating mentioned below:

Rating 1 and 2 – Low
Rating 3 –Medium
Rating 4 and 5- High

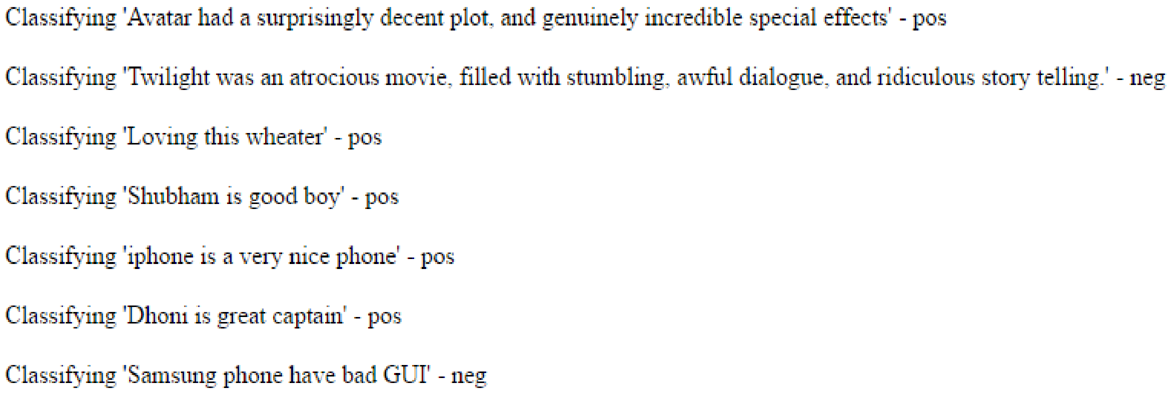
Meta data: Meta data contains user's information which is used to detect the spam reviews. The spam reviews are detected by considering order history and time interval of posting reviews online. It checks whether the users posted any reviews without purchasing product and how many reviews are posted about the product.

Patterns and Aspect Extraction: The aspect and patterns of user from the review are extracted in order to detect the spam classify the review.

Spam Detection Methodology: The detection of spam is done by tracking the user behaviour. The fake account also identified based on the activity log. SVM model are used to train the system to identify the spam and non-spam reviews. In this mechanism the features are identified from the reviews in order to detect whether the reviews are genuine or not. The patterns are extracted from the reviews such as short burst reviews, group behavioural indication which consists of rating deviation and review posting timings.

4. Results and Discussion

The proposed system classifies the review or comments into positive and negative categories in very less time. The accuracy of system is more due to large datasets. The system contains the positive and negative reviews. Aspects plays important role to classify as positive and negative dataset for classification. The classification of reviews used assigns rating to the product. The polarity is also assigned to the product as level low, medium and high.



```
Classifying 'Avatar had a surprisingly decent plot, and genuinely incredible special effects' - pos

Classifying 'Twilight was an atrocious movie, filled with stumbling, awful dialogue, and ridiculous story telling.' - neg

Classifying 'Loving this wheater' - pos

Classifying 'Shubham is good boy' - pos

Classifying 'iphone is a very nice phone' - pos

Classifying 'Dhoni is great captain' - pos

Classifying 'Samsung phone have bad GUI' - neg
```

5. Conclusion

Sentiment analysis is a widely explored research area and lot of applications are associated with it. The accuracy is still a major issue which affects the classification of reviews and rating. In this paper the various sentiment methods are analyzed. The proposed architecture uses part of speech, tokenization and limitization which are used to classify the reviews into positive or negative category and detect the spam reviews. The proposed system classifies the reviews into positive and negative category correctly.

References

[1] 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* Vol. 53 pp. 453-462.

[2] Lei, X., Qian, X., & Zhao, G. 2016, Rating prediction based on social sentiment from textual reviews. *IEEE Transactions on Multimedia* Vol. 18 Issue 90 pp. 1910-1921.

[3] Sahni, D., & Aggarwal, G. 2015, Recognizing Emotions and Sentiments in Text: A Survey, *International Journal of Advanced Research in Computer Science and Software Engineering* Vol. 5 Issue 5 pp. 201-205.

[4] Prabowo, R., & Thelwall, M. 2009, Sentiment Analysis: A Combined Approach, *Journal of Informetrics* Vol. 3 Issue 2 pp. 143-157.

[5] Susan Thomas., Jayalekshmi, 2015, Recommendation System with Sentimental Analysis Using Keyword Search, *International Journal for Advance Research In Engineering And Technology* Vol. 3 Issue 9 pp.73-78.

[6] Pavlopoulos, I. 2014, Aspect Based Sentiment Analysis, *Athens University of Economics and Business.*

[7] Lu, B., Ott, M., Cardie, C., & Tsou, B. K. 2011, Multi-aspect sentiment analysis with topic models. *IEEE 11th International Conference on Data Mining Workshops (ICDMW)*, pp. 81-88.

[8] Gaikwad, M. K. D., Sonawane, V. R. 2016, Opinion Mining and Sentiment Analysis Techniques: A Recent Survey, *International Journal of Engineering Sciences & Research Technology* Vol. 5 Issue 12 pp.1003-1006.

[9] Virmani, D., Malhotra, V., & Tyagi, R. 2014, Sentiment analysis using collaborated opinion mining. *arXiv preprint arXiv:* pp.1401.2618.

[10] Bhattacharjee, S., Das, A., Bhattacharya, U., Parui, S. K., & Roy, S. 2015, Sentiment analysis using cosine similarity measure, In *Recent Trends in Information Systems (ReTIS), 2015 IEEE 2nd International Conference on* pp. 27-32.

[11] Gupta, S., Rasbir Singh 2016, Extraction of Sentimental Analysis Using Vector Techniques and Feed Forward Neural Network, *International Journal of Research in Electronics and Computer Engineering.*

[12] Gupta, D. K., Kumar A. 2016, Spam And Sentiment Analysis Model For Twitter Data Using Statistical Learning, *International Symposium On Computer Vision And The Internet* pp. 54-58.

[13] Li, F., Huang, M., Yang, Y., Zhu, X. 2011, Learning To Identify Review Spam, In *IJCAI Proceedings-International Joint Conference on Artificial Intelligence* Vol. 22 Issue 3 pp. 2488-2493.

[14] SashiRekha k., Poobana S 2015, Opinion Mining from Text Reviews Using Machine Learning Algorithm, *International Journal of Innovative Research in Computer and Communication Engineering* Vol. 3 Issue 3 pp.1567-1570.

[15] Khairnar, J., &Kinikar, M. 2013, Machine Learning Algorithms for Opinion Mining and Sentiment Classification, *International Journal of Scientific and Research Publications* Vol. 3 Issue 3, pp. 1-6.

[16] Kamal, A. 2013, Subjectivity Classification Using Machine Learning Techniques For Mining Feature-Opinion Pairs From Web Opinion Sources, *International Journal of Computer Science Issues* (IJCSI) Vol. 10 Issue 5, No.1, pp 191-200.