

Intelligent Opinion Polarity and Analysis in Discovering Product Related Customer Reviews

Karuna Sahay^{1*}, Kaptan Singh², Amit Saxena³

¹ Department of CSE, Truba Institute of Engineering & Information Technology, Bhopal, M.P.

² Department of CSE, Truba Institute of Engineering & Information Technology, Bhopal, M.P.

³ Department of CSE, Truba Institute of Engineering & Information Technology, Bhopal, M.P.

Abstract-- Opinion Mining is the computational detailed investigation of people's attitudes, opinions, and emotions concerning of issues, events, topics or individuals. The micro-blogging and social network sites are considered as one of the best source of information for the reason that people discuss and freely share their sentiments about a certain subject freely. Posts are generally composed of poorly structured, incomplete, and noisy sentences, irregular expressions, non-dictionary terms, and ill-formed words. The problem is some customers given rating contrast with their comments. The other reviewers must read many comments and comprehend the comments that are different from the rating. Machine learning is about forecasting the future based on the past information. The proposed method provides automatically preprocessing of data and extract words from a sentence using machine learning. The proposed algorithm used improved Support Vector Machine method for better opinion mining and accurate prediction. The proposed method resolved semantic analysis and opinion mining problems to solve different sentiment polarity. The experimental outcome demonstrate that system is well suitable for opinion polarity and prediction.

Keywords: Sentiment analysis, Opinion mining, Machine learning, Social Media, Support Vector Machine, Sentiment Polarity

I. INTRODUCTION

Opinion mining (OM) is the technique which automatically discovering whether a consumer post information contains emotional content, sentiment, polarity or opinionated. The main aims of opinion mining classification is to recognize the opinion polarity of a consumer tweet as neutral negative, or positive for products. Opinion Mining can also be termed as Sentiment Analysis (SA) is the analysis of people's opinions, attitudes, sentiments or opinions concerning online product. The opinion entity can be product reviews, any educational events, individual's comments, or historical topics. These opinion are likely to be hidden from analyses. The two expressions OM or SA are interchangeable and express a common meaning. Opinion Mining is an unending field of research in document mining field. Opinion mining will review different post of users and mine their opinion about related subjects. The Clustering and natural language processing procedure will be applied for opinion mining. A portion of opinion mining denotes using of natural language processing (NLP)[2] by suggested dissimilar method of dictionary for sentimentality analysis of text data as lexicon, corpus, and specific language dictionary.

The problem is some customers given rating contrast with their comments. The other reviewers must read many comments and comprehensive the comments that are different from the rating. The challenge is in process of opinion mining or sentiment analysis that is unstructured and noisy data on website. The datasets used in OM are an essential problem in opinion review field. The foremost issues that interfere with the user comfort and security are confidentiality breach, groups lacking opt-in options, disorder created out of various groups in which a consumer is a member of and struggle in managing group ideologies. The purpose of utmost of the groups in widely held

social networking societies such as Twitter is misleading by uproars like spams and advertisements, and other clatters that obstruct with a group members interest.

The rest of the paper is organized as follows.

Section 2 represents the related work. Section 3 provides proposed architecture. Section 4 represents the implementation and result analysis. Section 5 provides conclusion of the paper.

II. RELATED WORK

Anh Vo et. el.[3] Suggests a technique to summarize and extract opinion and aspect of product form post of consumer in large dataset. The experimental outcome represents that the proposed method achieves F1-scores of 0.774 for laptop product review and 0.714 for camera product reviews. From results the author examined the usefulness and precision of proposed method. Consumer opinion analysis on online sold product is not an easy task, due to numerous assessments on Internet. Consumer opinion analysis problem can be addressed by applying syntax of reviews and exploring the appearance. It can also be solved by looking for expressed feeling in the post.

The datasets used in OM are an essential problem in opinion review field. The foremost main sources of data are from the consumer product reviews. The people's reviews are significant to the industry holders as they are takings business decisions agreeing to the analysis outcomes of users' thoughts about their products. The assessments sources are primarily review sites. OM can similarly be applicable on news articles, political debates or stock markets. For example in political debates, the investigators could figure out public's opinions on a certain political parties or election candidates. The election outcomes can also be forecast from user's political posts. The micro-blogging sites and social network sites are considered a good source of customer information because people discuss and share their opinions nearly a certain topic freely. They are similarly used as data sources in the OM process.

Kamps et. el. [3] applied the WordNet database to determine the polarity of tokenized words. They related a target word to two key words (usually 'bad' and 'good') to search the shortest path distance amongst the pivot words and target word the in the WordNet hierarchy. The minimum path distance was transformed to an incremental total and this assessment was stored with the tokenizer word in the words dictionary. The described accuracy level of this approach was 64.1%.

Littman and Turney[12] mapped the semantic relationship between the search target word and to each word from the designated set of negative and positive words to a real number. By means of subtracting a word's relationship strength to a set of negative words as of its association strength to a group of positive words, an accuracy rate of 83% was accomplished.

Q. Su et. el. Considers the probable of semantic expertise to address these enquiries. Afterward summarizing approaches to disambiguate and extract context information, the author present visualization procedures to discover the geospatial, lexical, and relational background of entities and topics referenced in these sources. The specimens stem as of the, The Climate Resilience Toolkit Media Watch on Climate Change, and the NOAA Media Watch—3 applications that combined environmental resources as of a widespread range of online sources. These schemes not only represents the value of as long as wide-ranging of information the public, but then again also have assisted to improve an innovative communication success metric that goes out there bipolar calculations of sentiment.

SVMs were applied by Li and A. Li[15] as a sentiment polarity classifier. Dissimilar the binary classification difficulty, authors debated that expresser credibility and opinion subjectivity should also be considered into consideration. Authors suggested a framework that make available a condensed numeric summarization of sentiments on micro-blogs platforms. Authors extracted and identified the subjects mentioned in the opinions connected with the requests of users, and then categorized the opinions by using SVM. Authors worked on Twitter posts data for experiment. Author found out that the concern of user opinion subjectivity and credibility is necessary for accumulating micro-blog opinions. The proposed method proved that mechanism can effectually determine market intelligence (MI) for assistant decision-makers by instituting a monitoring method to track exterior opinions on dissimilar aspects of a business in actual time.

Wararat Songpan [1] Suggests the prediction rating and analysis from customer examinations who mentioned as open opinion with the help of probability's classifier model. The suggested classifier models are applied in case study of consumer review's hotel in open posts comments for training data to classify consumer comments as negative or positive called opinion mining. In further addition, this classifier model has computed probability that represents value of style to provide the rating by using naive bayes procedures, which gives appropriately classifier to 93.47% related with decision tree Methods. The proposed methodology used Thai customer review's hotels from a website of hotel agent service, which service in hotel reservation directly. The target of classify customer review from this website because the comment is posted from customer who is serviced checked-in and checked-out from hotel. The process is started from collected data and preprocessing is cleaned data by removal stop words and using the high frequency of word which will be selected into attribute for using classifier model. The classifier model will be solve the text of customer review that is positive of negative from training data and test data

which are train from behavior posting from customer of hotel service group.

III. PROPOSED WORK

The proposed method has following steps.

Data collection from Social media: The collection of product related data from social media is the first step in sentiment analysis and polarity analysis. The product related post is collected from Twitter. The post consists of product related reviews of Twitter user.

Generation of theme, emotion and sentiment post related to products

The large amount of post is collected related to product reviews. These reviews can be used to generate themes, emotions and sentiment related to products.

Preprocessing of post: The Twitter post may contain many irreverent data which may be not useful for sentiment analysis and polarity analysis. In preprocessing of customer reviews related to products all the unwanted data will removed. This process helps data cleansing which may decreased size of data and errors and improved the accuracy of sentiment analysis.

Reviews of customer opinion: This step reviews all the post related to customer products.

Identification of the opinion: The next step is to identify the opinion of the users.

Opinionative words and phrases: In this step we opinionative all the words and phrases.

Feature selection and extraction

Some of the existing features are terms frequency and presence. These features are distinct word n-grams or words and their frequency sum total. It either gives the words binary weighting. In binary weighting zero if the word appears, or one if otherwise. It also uses term frequency weights to indicate the relative importance of features.

Negations: The presence of negative words could change the opinion meaning like not good is comparable to bad.

Opinion classification: Classification of opinion

The next step is to classify the opinion according to requirement.

Finding opinion polarity: The next step is to polarize the opinion.

Installing dependencies: All the dependencies will be installed in this step. It consists of Python libraries for specific used.

Applying Python script: The next step is to implement Python program. The environment will developed to implement the Python program. Result testing: The last step is to test the result and analysis the output.

Flow diagram of proposed work

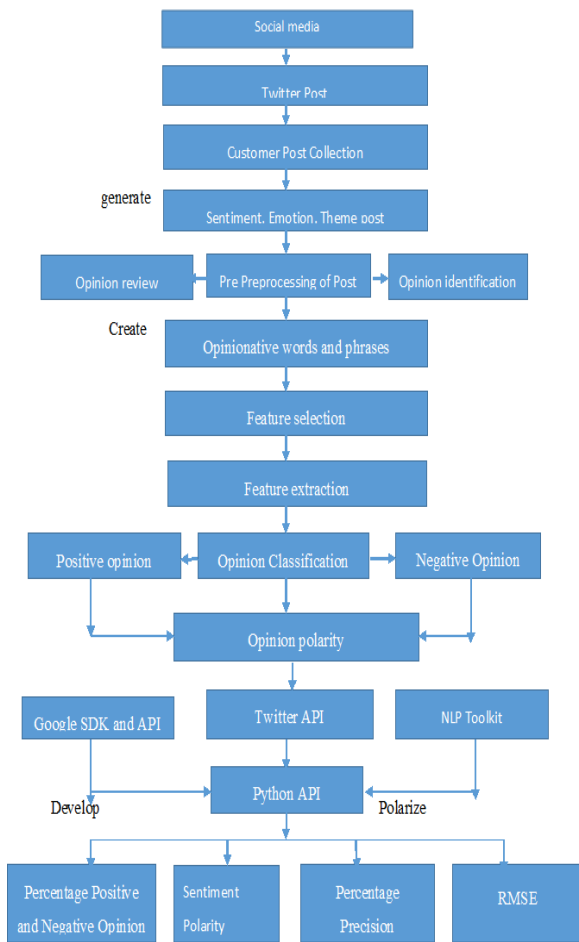


Figure 1: Flow diagram of proposed work

IV. IMPLEMENTATION AND RESULT ANALYSIS

For implementation environment used i3 3.0 GHZ machine with 4GB RAM. Social media data used for opinion mining and polarity finding. The social media post is collected data from social web sites like Twitter. The different post of users are reviewed and mine their opinion about related subjects.

Dataset The dataset has been collected from real-world environment of online social networks. The different posts from customers related to different products are collected. Five different groups have been chosen for performing the testing of the proposed work.

Dataset properties

Table 1: Dataset properties

Group	Members	Post
G1	2365	20321
G2	1388	18000
G3	8000	26000
G4	6002	12000
G5	2000	5987

For accessing dataset from Twitter registration is necessary. After registration new project have to be created from <https://app.twitter.com>. Application management is used to create test app for twitter. For registration some basic information is provided like application name, organization detail, website detail. For authentication Twitter provides keys

and access tokens. After getting access tokens and keys we can access the Twitter dataset.

The feature selection is to be characteristics in model that will be take out words from these consumer reviews as words occurred often to 14, 24 and 40 words. There are positive and negative in Table below, which are well-ordered by descendant frequent.

Table 2: Feature selection from frequent words

No.	Words (Positive)	#Frequent	Words (Negative)	#Frequent
1.	Excellent	400	Old	80
2.	Best	380	Inconvenient	80
3.	Better	350	Costly	70
4.	Very Good	320	Not delicious	65
5.	Good	290	Slow	60
6.	Beautiful	280	Expensive	55
7.	Luxurious	270	Troublesome	50
8.	Convenient	250	Problematic	40
9.	Attractive	150	Improve	30
10.	Nice	140	Not beautiful	25
11.	Delicious	130	Immoral	25
12.	Special	120	Not worth	20
13.	Comfortable	100	Uncomfortable	20
14.	Popular	80	Rare	20
15.	Safe	70	Unpleasant	15
16.	Cheap	70	Risky	15
17.	Not expensive	65	Unsafe	15
18.	Thanks	60	Unfriendly	10
19.	new	60	Not good	10
20.	Enjoy	50	Bad	5

The frequent words of positive are analyzed for attribute transformation individual text of customer review. The training and test data are separated into 3 sets: set 1 is composed 5 positive and 5 negative words; set is 2 is composed of 10 positive and 10 negative words and set 3 is composed of all positive and negative words.

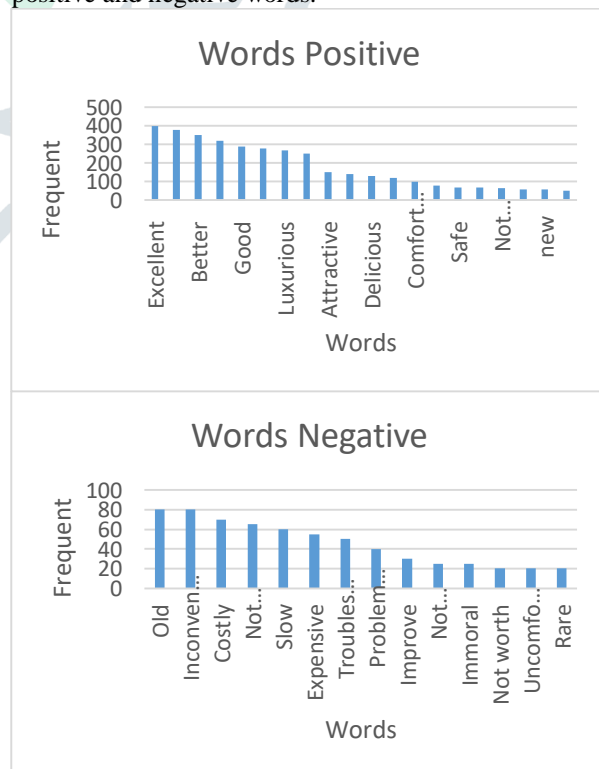


Figure 2: Positive and negative words

Accuracy Analysis

Table 3: Accuracy analysis

Attributes	% Accuracy	
	Aspect Based	Proposed
14 words	92.33	93.09
24 words	94.37	95.32
40 words	94.12	95.51
Average	93.61	93.97

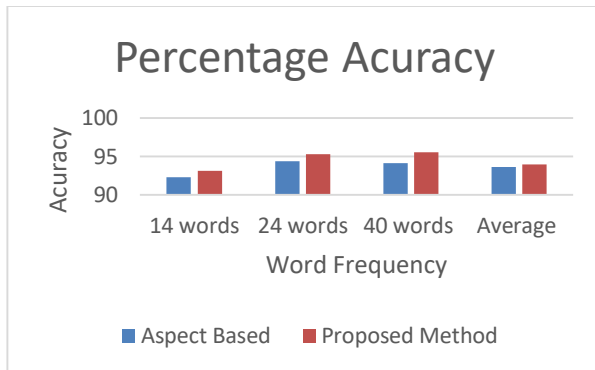


Figure 4: Percentage Accuracy

The experimental outcomes are tested with open opinions customer reviews of 560 from a twitter users posts. The accuracy of proposed work is given values that are higher RMS error

Table 4 : RMSE analysis

Attributes	RMSE	
	Aspect Based	Proposed
14 words	0.3660	0.3532
24 words	0.2390	0.2295
40 words	0.2326	0.2113
Average	0.2792	0.2324

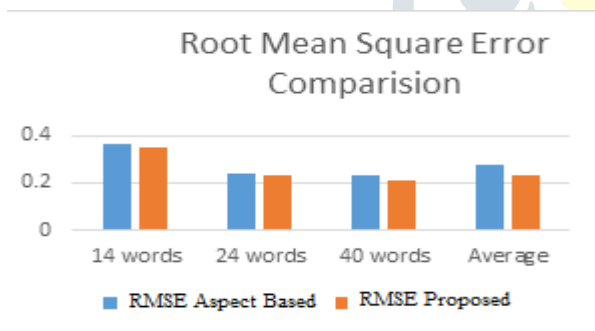


Figure 5: RMSE evaluation

The RMSE values of naïve bayes and proposed method is represented. The table above represents RMSE of different data sets. The lowest of RMSE is 40 words testing data that provide rating that are like to actual score from customer review to 0.2113. The rating of 40 words and 14 words are slightly higher value than 40 words to 0.2295 and 0.3532 respectively. The average of proposed method generates rating value that is similar actual rating as 0.2324 and median as 0.2295.

V. CONCLUSIONS

The machine learning technique SVM is applied for sentiment polarity. The proposed method improved prediction for better sentiment polarity outcomes. The method gives automatically preprocessing of data and extract related opinion from a sentence. The database is collected from Twitter product related post. The different post from users revived and mine their opinion about related subjects. The accuracy is improved to 3%

as compared to Aspect based method. The RMSE is also reduced to 4 % as compared to Aspect based method. The proposed novel method can be used for semantic analysis and opinion mining to solve different sentiment polarity for product related reviews. The experimental outcome demonstrate that system is well suitable for accurate sentiment prediction and opinion polarity in customer relationship.

REFERENCES

- [1] Wararat Songpan, The Analysis and Prediction of Customer Review Rating Using Opinion Mining, IEEE SERA 2017, pp. 71-77
- [2] Arno Scharl, David Herring, Walter Rafelsberger, Alexander Hubmann-Haidvogel, Ruslan Kamolov, Daniel Fischl, Michael Föls, and Albert Weichselbraun, "Semantic Systems and Visual Tools to Support Environmental Communication", IEEE SYSTEMS JOURNAL, VOL. 11, NO. 2, JUNE 2017, pp. 762-772
- [3] ANH-DUNG VO , QUANG-PHUOC NGUYEN , AND CHEOL-YOUNG OCK, Opinion-Aspect Relations in Cognizing Customer Feelings via Reviews, IEEE 2018, Vol-6, pp.5414-5426
- [4] Kamps, J., Marx, M., Mokken, R. J. Using WordNet to Measure Semantic Orientation of Adjectives. LREC 2004. Volume IV, pp. 1115-1118.
- [5] Andreevskaia, A., Bergler, S., Urseanu, M. All Blogs Are Not Made Equal: Exploring Genre Differences in Sentiment Tagging of Blogs. International Conference on Weblogs and Social Media (ICWSM-2007), Boulder, CO. 2007.
- [6] Vandana V. Chaudhari*, Chitra A. Dhawale** and Sanjay Misra, "Sentiment Analysis Classification: A Brief Review", I J C T A, 9(23) 2016, pp. 447-454
- [7] ANH-DUNG VO , QUANG-PHUOC NGUYEN , AND CHEOL-YOUNG OCK, "Opinion Aspect Relations in Cognizing Customer Feelings via Reviews", IEEE 2017, pp. 5415-5427
- [8] ATHIRA U, AND SABU M. THAMPI, "Linguistic Feature Based Filtering Mechanism for Recommending Posts in a Social Networking Group", IEEE 2018, pp. 4469-4484
- [9] S. I. Wu, R.D. Chiang and Z.H. Ji, Development of a Chinese opinion mining system for application to Internet online forum, The Journal of Supercomputing, Springer US[Online], 2016.
- [10] Z. Li, L.Liu and C.Li, Analysis of customer satisfaction from Chinese reviews using opinion mining, Proceeding of the 6th IEEE International Conference on Software Engineering and Service Science(ICSESS). 2015, pp.95-99.
- [11] Q.Su, X.Xu, H.Guo, Z.Guo, X. Wu, X. Zhang and B.Swen. Hidden Sentiment association in Chinese web opinion mining. Proceeding of the 17th International Conference on World Wide Web, 2008, pp.959-968.
- [12] R.M. Duwairi and I. Qarqaz, Arabic Sentiment Analysis using Supervised Classification. Proceeding of 2014 International Conference on Future Internet of Things and Cloud. 2014, pp. 579-583.
- [13] H.S. Le, T.V. Le and T.V. Pham, Aspect Analysis for Opinion Mining of Vietnamese Text. Proceeding of International Conference on Advance Computing and Application, 2015, pp.118-123.
- [14] V.B. Raut and D.D. Londhe, "Survey on opinion mining and summarization of user review on web", International Journal of Computer Science and Information Technology, Vol. 5(2), 2014, pp. 1026-1030.
- [15] Fiaidhi, O. Mohammed, S. Mohammed, S. Fong, and T.H. Kim, Opinion Mining over twitterspace: Classifying tweets programmatically using the R approach. Proceeding of the 7th International Conference on Digital Information Management, 2012, pp. 313-319.
- [16] L. Lin, I. Li, R. Zhang, W. Yu and C. Sun, Opinion mInIng and sentiment analysis in social networks: A retweeting structure-aware approach. Proceeding of the 7th International Conferenece on Utility and Cloud Computing, 2014, pp.890-895.