International Journal of Electrical, Electronics and Data Communication, ISSN(p): 2320-2084, ISSN(e): 2321-2950
Volume-6, Issue-9, Sep.-2018, http://iraj.in

SENTIMENT ANALYSIS OF THE SMARTPHONE PRODUCT REVIEWS

[1]MAY THANDA THEINT AUNG, [2]AYE AYE KYAW

Faculty of Information and Communication Technology (ICT), University of Technology (Yatanarpon Cyber City), Pyin Oo Lwin, Myanmar
E-mail: [1]maythandar1994@gmail.com, [2]ayeayekyaw2009@gmail.com

Abstract- Nowadays, social media has become a platform of online communication for people to convey their opinions to the public. Today users can not only obtain information but also actively generate content using web. Blogs, reports, forums, etc. are the main sources of public opinion information called reviews. Good responses from people can awake desire for a product, create brand awareness, encourage a positive attitude toward the product for buying the products. Using natural language processing the text that contains both cases and opinion could be extracted to get some opinionated views. To analysis such texts and reviews sentiment analysis is used. Sentiment analysis is a sub domain of Natural Language Processing and is used to automatically mine the opinions and emotions from text such as comments and posts. In this paper sentiment analysis is carried out by two approaches: machine learning approach and dictionary based approach. And then discuss accuracies for more accurate sentiment analysis and the issues in detecting correct polarity of the texts.

Keywords - Natural Language Processing, Sentiment Analysis, Social media.

I. INTRODUCTION

The Web has dramatically changed the way that people express their reviews and opinions. They can now express their views of products at merchant sites by posting as reviews and on almost anything in Internet forums, discussion groups, and blogs. These are collectively called the user-generated content. Now if one wants to buy a product, he/she can see many product reviews on the Web which give opinions of existing users of the product. The company may no longer be necessary to conduct surveys by organizing focus groups or employing external consultants in order to find consumer opinions about its products and those of its competitors. Because the user-generated content on the Web can already give them such information. However, it is difficult for a human to organize them into usable forms by finding relevant sources, extracting related sentences with opinions, and summarizing. Thus, automated opinion discovery and summarization systems are needed.

Sentiment analysis is used to automatically mine the opinions and emotions from text, speech, and database sources with the help of Natural Language Processing (NLP). Sentiment analysis indicates the classification of opinions in the text into categories like “positive”, “negative” or “neutral”. It’s often referred to as subjectivity analysis or opinion mining. Many customers want to see the opinion of others about the product before buying any product. Customer reviews are very important for business process because business organizations should know to make future better decision. It will provide important functionality for voice of customer and brand reputation management. Thus it is helpful in business process and also for the customers.

II. SENTIMENT ANALYSIS

Major areas of research in Sentiment analysis are Subjectivity Detection, Sentiment Prediction, Aspect Based Sentiment Summarization, and Text Summarization for Opinions, Contrastive Viewpoint Summarization, Product Feature Extraction, and Detecting Opinion Spam. Subjectivity Detection is a task of finding whether text is opinionated or not. Sentiment Prediction is about predicting the polarity of text whether it is positive or negative. Aspect Based Sentiment Summarization generates sentiment summary in the form of star ratings or scores of features of the product. Text Summarization generates a few sentences that summarize the reviews of a product. Contrastive Viewpoint Summarization puts an emphasis on contradicting opinions. Product Feature Extraction is a work that extract product feature from its review. Detecting Opinion Spam is concern with identifying fake or bogus opinion from reviews [2].
Sentiment classification is carried out at three levels- Document level, Sentence level and Aspect or feature level. In Document level the task is to classify complete documents into positive or negative class. Sentence level sentiment classification classifies sentence into positive, negative, neutral class based on each sentence level. First the polarity of each word of a sentence is calculated and then the overall sentiment of the sentence is calculated. Aspect or Feature level sentiment classification identifies and extract product

features from the source data and do the classification [1].

Sentiment analysis process can be divided into four steps: text collection, preprocessing, analysis and validation.

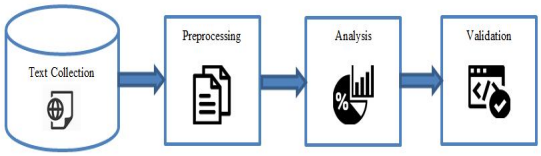

Figure1. Sentiment Analysis Process

Text collection is the first step of the process to obtain the raw data by gathering text data from desired source such as blogs, news, social media. Preprocessing is the most importment step in which collection of raw data need to be trsnsformed and prepared by applying required operation such as tokenization, stemming, lemmatization. In analysis step, sentiment analysis is generally carried out by many approaches: rule based, machine learning based and dictionary based etc.

Rule based approach is employed by shaping various rules for obtaining the opinion, created by tokenizing every sentence in each document then testing every token, or word, for its presence. If the word is there and has a positive sentiment, a +1 rating was applied to that. Every post starts with a neutral score of zero, and was considered positive. If the ultimate polarity score was bigger than zero, or negative if the score was less than zero once the output of rule based approach it will check or raise whether the output is correct or not. If the input sentence contains any word that isn't present within the database which can facilitate within the analysis of moving picture review, then such words are to be added to the database. This is often supervised learning within which the system is trained to learn if any new input is given [3].

Machine learning based approach applies classification technique to classify text. This classification requires two sets of documents: training set and test set. A training set is to learn the difference signs documents used by classifier. A test set is to validate the performance of classifier and is used for machine learning techniques to classify a number of review adopted. Machine learning approaches include Naïve Bayes (NB), Maximum Entropy (ME), and Support Vector Machine (SVM) and they have achieved great success in text classification. Naïve Bayesian Classification is based on Naïve-Bayes theorem and uses the concepts of maximum likelihood and Bayesian probability. The ME is based on the Principle of Maximum Entropy and from all the models that fit our training data, selects the one which has the largest entropy. SVM is a supervised learning method used to analyze the data and recognize data patterns that can be used for classification and regression analysis. These methods are that the model needs to be trained with a large data volume before testing.

Dictionary based method first finds the opinion word from review text then finds their synonyms and antonyms from dictionary. This method uses sentiment dictionary with opinion words and match them with the data to determine polarity. They assign sentiment scores to the opinion words describing the Positive, Negative and Objective score of the words contained in the dictionary. The dictionary used may be WordNet or SentiWordNet or other. SentiWordNet 3.0 is most useful dictionary used. It is a lexical resources publically available made up of “synsets” in which each is associated with a positive, negative numerical score range from 0 to 1. This score is automatically allotted from the WordNet. It uses a semi-supervised learning method and an iterative random walk algorithm.

TABLE I: PART of SPEECH TYPE

POS Name	POS Abbreviation	SentiWordNet Abbreviation
Noun	NN	n
Nouns	NNS	n
Adjective	JJ	a
Adjectives	JJS	a
Verb	VB	v
Adverb	RB	r

Validation is the final step of sentiment analysis process. In order to validate in the analysis steps are preformed, quantitative measures and qualitative measures are necessary to measure.

III. PROPOSED SYSTEM

The followings show step by step detail implementation of the system.

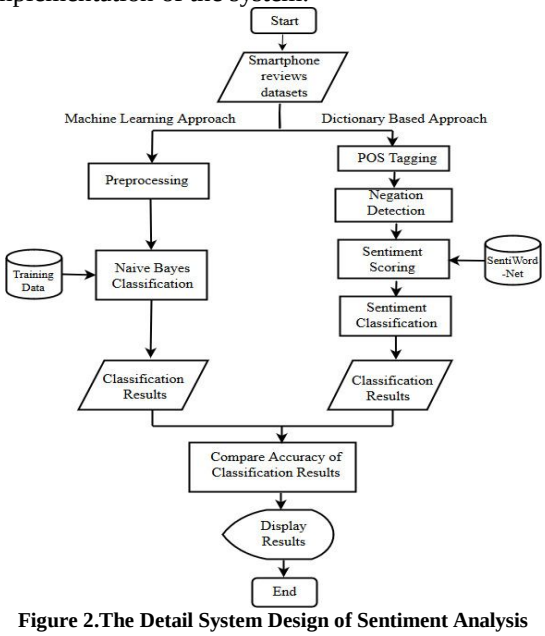

Figure 2.The Detail System Design of Sentiment Analysis

A. Data Description

This system is proposed to classify the polarity of the smartphone product reviews written in English from the web. Five kinds of smartphone (Apple, Samsung, HTC, Sony, LG) released in 2016 is emphasized to apply as testing datasets for this system. Data used in this study are online product reviews collected from Amazon.com. We used total 5000 reviews as training data sets by 1000 reviews from each smartphone brand (500 reviews each from classes positive and negative) to get the balanced dataset of each 2500 reviews for both classes: positive and negative.

B. Machine Learning Approach

Machine learning approach includes the followings steps.

1) Data Preprocessing

The dataset is needed to perform these preprocessing steps:

- Punctuation & Special characters Removal: remove any punctuation or symbols that are not letter.
- Lowercase: convert all the text in the dataset to lowercase.
- Tokenization*:* the task of chopping a character sequence or a defined document unit up into pieces, called tokens.
- Stop words Removal*:* deletion of the words that can be ignored in classification to improve processing speed.
- Stemming: the technique to remove affixes from a word, ending up with the stem.

2) Naïve Bayes Classification

After pre-processing steps, each review in testing dataset is classified as follows:

$$P(positive|\,review) = \frac{P(review|\,positive)P(positive)}{P(review)} \quad (1)$$

And then, also calculate P (negative | review). Then, compare the probabilities for each class and use the highest rank as the class for the review.

C. Dictionary Based Approach

Dictionary based approach includes the followings steps:

- Lemmatization: It aims to remove inflectional endings only and to return the base or dictionary form of a word, which is known as the lemma.
- POS-tagging: This is the process of classifying words into their parts of speech and labeling them accordingly.
- Negation Detection: Negation words are the words which reverse the polarity of sentence. This is one of the natural language processing issues. So, we need to detect negation words to improve accuracy.

1) Proposed Algorithm

The followings are step by step of proposed algorithm:

- Read the input data file with a file reader object.
- Parse each sentence token by token with the help of POS tagger. POS tagger will assign a tag to each token.
- Check the tag of each token if the tag is “JJ”, “JJS” and so on (i.e. the tagged token is an adjective/opinion word) then pass this word in SentiWordnet to check the score as well as polarity of that particular word.
- SentiWordnet will return the sentiment-type of that word based on score.
- Calculate the positive score and negative score for each sentence. Then, compare the total positive and negative score.
- If the value of positive score is greater than negative, then the sentence is considered “positive” as a whole.
- If it is not, then the sentence is considered “negative” as a whole.

RESULTS AND DISCUSSION

Performance of each classification model is evaluated by using four indexes: Accuracy, Precision, Recall, and F-measure. The followings are equations for indexes:

$$accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (2)$$

$$recall = \frac{TP}{(TP + FN)} \quad (3)$$

$$precision = \frac{TP}{(TP + FP)} \quad (4)$$

$$f - measure = \frac{(2 * precision * recall)}{(precision + recall)} \quad (5)$$

TABLE II: COMPARING PERFORMANCE OF PRECISION, RECALL AND F-MEASURE FOR EACH BRAND

Brand Name	Method	Precision	Recall	F-measure
Apple	NB	95.58%	98.35%	96.95%
	DB	93.68%	98.34%	95.95%
Samsung	NB	97.01%	94.98%	95.97%
	DB	93.90%	96.65%	95.26%
LG	NB	98.26%	99.12%	98.69%
	DB	91.43%	98.25%	94.71%
HTC	NB	97.18%	100%	98.57%
	DB	92.86%	94.20%	94.00%
Sony	NB	98.26%	99.12%	98.69%
	DB	91.43%	98.25%	94.71%

Table II shows the comparison of Precision, Recall,

and F-measure that are calculated by using above equations.

TABLE III: COMPARING OVERALL PRECISION, RECALL AND F-MEASURE

Approach	Precision	Recall	F-measure
Naïve Bayes	97.26%	98.31 %	97.77%
Dictionary-based	92.66%	97.14 %	76.31%

TABLE IV: ACCURACY COMPARISION for MACHINE LEARNING and DICTIONARY BASED APPROACH

Brand Name	Naïve Bayes (NB)'s Accuracy	Dictionary Based (DB)'s Accuracy
Apple	95.00%	93.33%
Samsung	93.67%	92.33%
LG	98.00%	91.67%
HTC	98.00%	91.00%
Sony	98.00%	91.67%
Total Accuracy	96.53%	92.00%

By observing results from Table IV, we can see Naïve Bayes classifier of machine learning approach give more accurate results than Dictionary-based approach.

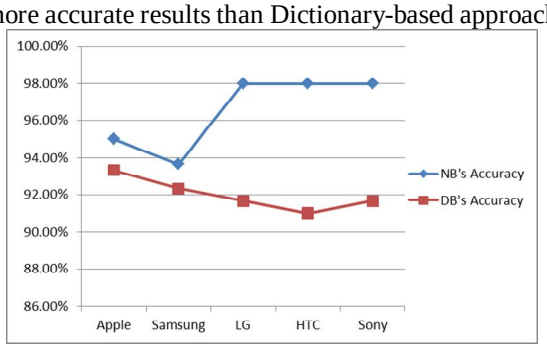

Figure 3.Accuracy Comparison of Sentiment Analysis for Apple, Samsung LG, HTC and, Sony

At first, when we tested the Naïve Bayes classifier by using only 3000 training datasets, the classifier gave less accurate results than another one. But after using more training datasets, Naïve Bayes classifier increase its accuracy. As conclusion, we can see machine learning algorithms change its accuracy according to size of training data used for it.

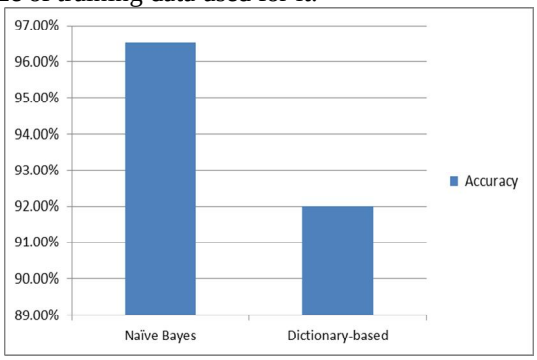

Figure 4.Overall Accuracy Chart for Two Apporaches

CONCLUSIONS

In this paper we have demonstrated how machine learning and dictionary based techniques can be used to infer sentiments over web data. This study is conducted using five kinds of some popular smartphone brands released in year 2016. We observed that how well performed of two approaches on sentiment analysis by comparing the accuracy of algorithms. Our experiments can provide positive and negative opinions of smartphone products from former customers that are useful for both side of customers and companies.

ACKNOWLEDGMENT

This research is supported by Department of Information Science and Technology (IST) from UTYCC. I would like to show gratitude to my supervisor for sharing her knowledge to me during the research period. I also thank all teachers and colleagues from our department for advices and comments that greatly improved the research.

REFERENCES

[1] Varghese,R., Jayasree, M., "A Survey on Sentiment Analysis and Opinion Mining," IJRET: International Journal of Research in Engineering and Technology, Volume: 02 Issue: 11, October 2013.

[2] Vohra, S. M. and Teraiya, J. B., "A Comparative Study of Sentiment Analysis Techniques," Journal of Information, Knowledge and research in Computer Engineering Volume – 02, Issue –02, October 2013.

[3] Anchal Kathuria, Dr. Saurav Upadhyay, "A Novel Review of Various Sentimental Analysis Techniques" International Journal of Computer Science and Mobile Computing IJCSMC, Vol. 6, Issue. 4, April 2017, pg.17 – 22

[4] Tanvi Hardeniya, Dilipkumar A. Borikar "Dictionary Based Approach to Sentiment Analysis- A Review" IJAEMS: International Journal of Advanced Engineering, Management and Science, Volume-02 Issue-5, May 2016.

[5] Deepak Singh Tomor, Pankaj Sharma "A Text Polarity Analysis Using Sentiwordnet Based an Algorithm" IJCSIT: International Journal of Computer Science and Information Technologies, Volume-7(1), 2016.

[6] Reshma Bhonde, Binita Bhagwat, Sayali Ingulkar, Apeksha Pande, "Sentiment Analysis Based on Dictionary Approach", International Journal of Engineering Research and Technology, Volume 3,Issue 1, January 2015.

[7] Haruna Isah, Paul Trundle, Daniel Neagu, "Social Media Anlaysis for Product Safety using Text Mining and Sentiment Analysis ".

[8] Nur Azizah Vidya, Mohamad Ivan Fanany, Indra Budi, "Twitter Sentiment to Analyze Net Band Reputation of Mobile Phone Providers" 3rd Information Systems International Conference, Procedia Computer Science 72, 2015, 519-526.

[9] Mukwazvure, A., K.P. Supreethi, "A Hybrid Approach to Sentiment Analysis of News Comments. Reliability," In 4th International Conference on Infocom Technologies and Optimization (ICRITO)- Trends and Future Directions, IEEE, 2015.

[10] Efthymios kouloumpis, Theresa Wilson, Johanna Moore, "Twitter Sentiment Analysis: The Good the Bad and the OMG" Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media

[11] Bing Liu, "Sentiment Analysis and Subjectivity", Handbook of Natural Language Processing, Second Edition, 2010.
[12] Cambria, E., Speer, R.Havasi, C. and A. Hussain, "SenticNet: A Publicly Available Semantic Resource for Opinion Mining," In AAAI fall symposiums: commonsense knowledge, vol. 10, p. 02, March 2010.
[13] Soni, V., Patel M. R., "Unsupervised Opinion Mining from Text Reviews Using SentiWordNet," International Journal of Computer Trends and Technology, volume 11 number 5, May 2014.
[14] Li, C. and R. Li, "Lexicon Construction: A Topic Model Approach," In International Conference on Systems and Informatics (ICSAI), pp. 2299-2303. IEEE, 2012.
[15] Karamibekr, M. and Ghorbani, A., "Sentiment Analysis of Social Issues," International Conference on Social Informatics (Social Informatics), IEEE 2012, December 2012.
[16] Mukwazvure, A., K.P. Supreethi, "A Hybrid Approach to Sentiment Analysis of News Comments. Reliability," In 4th International Conference on Infocom Technologies and Optimization (ICRITO)- Trends and Future Directions, IEEE, 2015.
[17] Jha, V.,Manjunath, N., Shenoy, V.,Venugopal, K. R.andPatnaik, L., "HOMS: Hindi Opinion Mining System," IEEE 2nd International Conference on Recent Trends in Information Systems (ReTIS), M. 2015.
[18] Abbasi, Ahmed, Hsinchun Chen, and Arab Salem. "Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums." ACM Transactions on Information Systems (TOIS) Volume 26 issue 3, pp.1- 12, 2008.
[19] Mudinas, Andrius, Dell Zhang, and Mark Levene. "Combining lexicon and learning based approaches for concept-level sentiment analysis." Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining. ACM, pp330-333, 2012.
[20] Neviarouskaya, A., Prendinger, H. and Ishizuka, M. "SentiFul: A Lexicon for Sentiment Analysis," IEEE transactions on affective computing, vol. 2, no. 1, January-March 2011.
[21] Feldman, Ronen. "Techniques and applications for sentiment analysis." Communications of the ACM 5,Volume 56 Issue4, pp.82-89,2013.
[22] Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. "Thumbs up?: sentiment classification using machine learning techniques." In Proceedings of the ACL-02 conference on Empirical methods in natural language processing- Association for Computational Linguistics, Volume 10, pp. 79-86,2002.
[23] Li, Shoushan, Zhongqing Wang, Sophia Yat Mei Lee, and Chu-Ren Huang. "Sentiment Classification with Polarity Shifting Detection" In Asian Language Processing (IALP), 2013 International Conference on, pp. 129-132. IEEE, 2013.

★★★