Fuzzy based Sentiment Analysis of Online Product Reviews using Machine Learning Techniques

Haseena Rahmath P
M.Tech Scholar, Dept.of Computer Science and Engineering, Al-Falah School of Engineering
Dauj, Haryana, India

Tanvir Ahmad, PhD
H.O.D. Dept. Of Computer Engineering
Jamia Millia Islamia,New Delhi, India

ABSTRACT

The exponential growth of internet opened a new platform where people can freely express and exchange their suggestions, ideas and feedback about any product or services. People prefer e-commerce websites to buy or sell products or services and they like to review and analyze the opinions of others while purchasing any product or services. The social-medias, e-commerce websites, review websites, forums, blogs etc. encourage users to share their views, opinions, suggestions and feedback about different aspects that touch their day to day life. This trend lead to a huge accumulation of user generated content on internet. The processing and analyzing this huge unstructured content, which are written in natural language is a challenging task. These factors motivated the development of an opinion mining and sentiment analysis system that can automatically extract, classify and summarize users' reviews.

The present work proposes a multi-step opinion mining system that involves pre-processing to clean the document, a rule-based system to extract features and a scoring mechanism to tag their polarity. The proposed system can be used for binary as well as fine-grained sentiment classification of user reviews. The proposed technique utilizes fuzzy functions to emulate the effect of various linguistic hedges such as dilators, concentrator and negation on opinionated phrases that make the system more accurate in sentiment classification and summarization of users' reviews. Experimental evaluation indicates the system can perform the sentiment analysis with an accuracy of 93.85 %.

General Terms

Opinion Mining, Sentiment Analysis, Sentiment Classification

Keywords

Opinion Mining; Sentiment Analysis; Machine Learning Techniques; Rule Based System; Online Product Reviews; Review Classification; Review Summarization

1. INTRODUCTION

The exponential growth of internet changed the entire human society. Now internet has become integral part of our life: whether it is for finding valuable information about something or for purchase any product/service or for enrolling in any course or for booking any ticket etc. for everything now we depend up on internet. People always prefer e-commerce websites to buy or sell products or services. The social-medias, e-commerce websites, review websites, forums, blogs etc. encourage users to express or exchange their views, opinions, suggestions and feedbacks about different aspects that touch their day to day life. This trend lead to a huge accumulation of user generated content on internet.

Customers, manufactures, policymakers, lawmakers' etc., everyone can take advantage of these accumulated data by extracting and analyzing it properly. Customers can consider others opinion and experience while purchasing any product or services and they can also compare different competing brands. Manufactures can improve the quality of their product by taking valuable feedback, suggestion and opinions from user reviews of their product and they can also look into strength and weakness of their competitors. Policy makers or law makers can take more citizen friendly policy by considering citizens' real feeling or desire towards certain government policy or law. But most of these data are written in natural language and are in unstructured nature. These factors make the manual extraction and analysis of this huge content an impossible task. This situation in turn opened a new research area called Opinion Mining and Sentiment Analysis, which automatically extract and classify the sentiments expressed in these contents related to some aspects. The increased popularity of social-medias and micro blogging, and the high commercial importance of summarization of user reviews fuelled much research interest in Opinion Mining and Sentiment Analysis among natural language processing communities [1-8].

Opinion Mining and Sentiment Analysis, an extension of data mining, is a natural language processing and text analytic technique [9] that determines people's emotion or feeling or attitude towards some topic by processing huge unstructured internet content. Opinion Mining extracts the sentiments or opinions bearing words present in the free text while Sentiment analysis determines sentiment polarity, whether positive or negative or neutral by analyzing each opinionated word or phrase. Sentiment Analysis summarizes the opinion of a writer or speaker about a particular topic and it can be done at word or aspect level, sentence level and document level. Sentiment can be classified by various ways: 1) supervised classification techniques, 2) unsupervised classification techniques and 3) hybrid classification techniques or by combining the above two approaches.

This paper proposes a fine-grained sentiment classification of online product reviews by using unsupervised machine learning technique. It incorporates fuzzy linguistic hedges-modifiers on opinion descriptors. The effect of various fuzzy linguistic hedges on opinion descriptors are emulated by fuzzy functions. The Proposed opinion mining system is multi-step process such as 1) pre-processing step, (2) feature-set generation step and 3) sentiment classification step. The system classifies the reviews into three classes: positive, negative and neutral. The proposed fuzzy function to emulate the effect of fuzzy linguistic hedges on opinionated words gives better accuracy than other contemporary approaches [10 - 12]. Relatively very few studies are conducted which considers the effect of linguistic hedges on opinion descriptors [10 - 13].

The rest of the paper is organized as follows. Related works conducted in this area are surveyed in Section 2 while the proposed fuzzy based opinion mining framework is explained in Section 3. Section 4 discusses the experimental evaluation and the result obtained from the proposed strategy. Finally, Section 5 concludes the paper and gives suggestions for future work in this field.

2. RELATED WORKS

Opinion Mining and Sentiment Analysis involves extraction of sentiment words from user reviews and automatic classification and summarization of sentiments. The sentiment words present in the free text can be identified by considering the following: adjectives or adverbs [14], uni-grams [15] or n-grams [16] with their frequency of occurrence, the POS (parts of speech) tagging of words [17] the negation of words [18]. The automatic text classification can be done through various machine learning techniques. The machine learning technique may be supervised learning technique such as Naive Bayesian [18], support vector machines [19], Artificial Neural Networks [21] or unsupervised learning technique [16, 19, 20] or hybrid approaches [22, 24]. The hybrid methods combine the supervised and unsupervised techniques to yield maximum accuracy in sentiment classification.

However, the classification and summarization of sentiments expressed by the user in the free text or documents is more comprehensive process and it is quite different from simple text mining approaches. It is not an extractive summary or classification of entire documents by simply considering topic-indicative words or phrases. Instead, Sentiment classification involves tasks like generation of semantic feature-set, sentiment words or opinionated word identification corresponding to the features, determination of the semantic polarity orientation of the feature-opinion pairs, and find out overall sentiment by aggregating the mined results [1, 2, 23].

People prefer to write reviews in natural language and in free format. The spelling and grammatical mistakes are quite common. Hence, the documents should be cleaned before processing [23, 25]. Features can be extracted efficiently by using parts-of-speech (POS) tagging [23, 26], it is language-dependent semantic process, and it tags each word by considering its position in the grammatical context. Usually linguistic parsers such as link grammar parser [25, 27] and Stanford parser [23, 36] are used for POS tagging of sentences present in user-generated contents. Various approaches like frequent term identification based on Apriori algorithm [6, 28, 29], seed-set expansion from the initial seed set of features [23, 26], or multi words-based feature extraction [25, 30–32] and rule based feature extraction [17, 33, 34] are adopted for extracting relevant features present in the documents.

Adjectives or adverbs are main indicators of polarity orientation of opinionated phrase. In [35] the authors utilized conjunctions between adjectives as a method for determining orientation of adjectives. In [20, 36] authors proposed word association measures such as PMI (point-wise mutual information) and LSA (latent semantic association) to determine semantic polarity orientation. Another approach is by considering synonyms and antonyms of opinionated word in which first start the polarity determination with an initial known seed list of sentiment words and then gradually expand the seed list by considering synonyms and antonyms from lexical resource like WordNet [37]. Since the semantic orientations of sentiment words are depend closely with its usage, it is not accurate to assign fixed polarity orientation to

each sentiment descriptor. Hence, most of the recent researches are utilized SentiWordNet API [38] to determine semantic polarity orientation of opinionated words [12, 25, 33, 34]. The SentiWordNet assigns a score for each sentiment descriptor in the form of a triplet which represents its positivity/objectivity/ negativity scores.

The presence of linguistic hedges or modifiers has high impact on determination of semantic orientation of opinionated word [10, 39, 40]. For example, semantic orientation of opinion descriptor "good" is different from "very good" and it is entirely different from "not good". In second and third cases, there exist modifies: "very" and "not". In fact, it is changing the semantic orientation of "good" completely. Some authors consider linguistic hedges as modifiers and they showed that it can affect polarity of an opinioned phrase [10, 11]. In [12] a hybrid scoring technique is used which combined SentiWordNet with PMI. Initial sentiment value of opinion descriptor is calculated manually and if any linguistic hedges present then it is adjusted using fuzzy functions. This paper proposed novel fuzzy functions for emulating the effect of fuzzy linguistic hedges on opinionated phrases and this approach yielded better accuracy compared to other contemporary approaches.

3. PROPOSED OPINION MINING FRAMEWORK

This section discusses the design of proposed fuzzy based opinion mining and sentiment analysis system, which automatically extracts features, opinions and linguistic hedges (modifiers) from unstructured user-generates reviews and based on their sentiment orientation, it classifies reviews as "positive reviews", "negative reviews" and "neutral reviews".

During sentiment score calculation of opinionated word, it considers the affect of linguistic edges or modifiers on those opinionated words. For instance, "x is good" describes no hedge; "x is very good" describes an intensifying or concentrating hedge and in "x is not good" describes an inverting or modifying hedge.

The cleaned and parsed documents obtained from the pre-processing stage are given to the feature extraction stage. The features extraction stage extract the features by using various rules and the irrelevant features are filtered out by considering frequency of occurrence features. If the max frequency measure of any feature is below the threshold limit then it will be eliminated.

The K-Means clustering algorithm present in the Weka data mining software is used to classify the datasets using the generated matrix from the feature extraction stage which had already been converted into ARFF format as input. The System consists of three major steps as described in Figure 1: pre-processing step, feature selection step, and classification and summarization step. Figure 2 depicts the detailed diagram of fuzzy based opinion mining and sentiment analysis framework. The system used a non-supervised sentiment classification approach for sentiment classification and it is evaluated using dataset of online customer reviews of mobile phones.
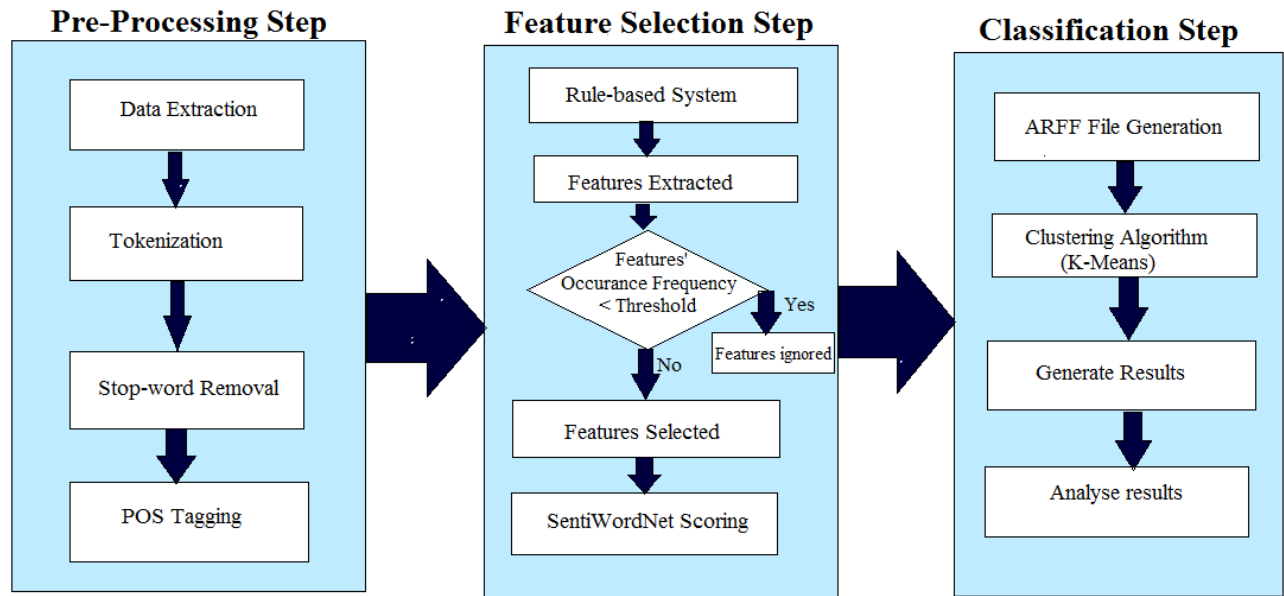
Figure 1: Block diagram of the system

3.1 Preprocessing Phase

In order to achieve maximum accuracy in feature extraction process, the noises present in the user generated content should be eliminated [8, 12]. The noises are usually in the form of spelling mistakes, grammar mistakes, mistakes in punctuation, incorrect capitalization, and usage non-dictionary words such as abbreviations or acronyms of common terms and so on. The main reason for this is these reviews are mostly written by non-experts and in short informal texts. After downloading the datasets from internet, the proposed system cleaned the documents by removing the html tags present in the document and it correct spelling errors. The texts are tokenized into tokens and the stop-words are detected and removed. Since words like preposition, digits, articles and proper nouns like name of cell phone etc. are considered as valueless in the sentiment analysis, hence these words are included in the stop word list. The sentences generated in this pre-processing can be parsed automatically by any linguistic parser. The proposed system used Stanford Linguistic parser [41] for POS tagging of each word present in the sentences. POS tagger parses each sentences and tags each term with its part of speech.

3.2 Feature Extraction

 The features, modifies/linguistic hedges and opinion words are extracted by applying various rules. The System used six feature selection rules. They are given below:

Rule-1: In a dependency relation R, if there exist relationships nsubj(w1,w2) and nn(w2,w3) such that POS(w1) = JJ* and POS(w2) = POS(w3) = NN* and w1, w2 and w3 are not stop-words then (w3 w2) is considered as the feature and w1 as an opinion. Thereafter, the relationships advmod(w1,w4) and neg(w1,w5)are searched. If both exist together then (w5 w4) will be the modifier or if advmod(w1,w4) only exist then the modifier will be w4 or if only neg(w1,w5)exist then modifier will be w5.

Rule-2: In a dependency relation R, if there exist relationships nsubj(w1,w2) and appos(w2,w3) such that POS(w1) = JJ* and POS(w2) = POS(w3) = NN* and w1, w2 and w3 are not stop-words then (w3 w2) is considered as the feature and w1 as an opinion. Thereafter, the relationships advmod(w1,w4) and neg(w1,w5)are searched. If both exist together then (w5 w4) will be the modifier or if advmod(w1,w4) only exist then the modifier will be w4 or if only neg(w1,w5)exist then modifier will be w5.

Rule-3: In a dependency relation R, if there exist relationships nsubj(w1,w2) and amod(w2,w3) such that POS(w1) = POS(w3) = JJ* and POS(w2) = NN* and w1, w2 and w3 are not stop-words then (w3 w2) is considered as the feature and w1 as an opinion. Thereafter, the relationships advmod(w1,w4) and neg(w1,w5)are searched. If both existtogether then (w5 w4) will be the modifier or if advmod(w1,w4) only exist then the modifier will be w4 or if only neg(w1,w5)exist then modifier will be w5.

Rule-4: In a dependency relation R, if there exist a such that POS(w1) = JJ* and POS(w2) = NN* and w1and w2 are not stop-words then w2 is considered as the feature and w1 as an opinion. Thereafter, the relationships advmod(w1,w3) and neg(w1,w4)are searched. If both exist together then (w4 w3) will be the modifier or if advmod(w1,w3) only exist then the modifier will be w3 or if only neg(w1,w4)exist then modifier will be w4.

Rule-5: In a dependency relation R, if there exist relationships nsubj(w1,w2),   dobj(w1,w3) and nn(w3,w4) such that POS(w1) = VB* and POS(w2) = POS(w3) = POS(w4) = NN* and w1, w2, w3and w4 are not stop-words then (w4 w3) is considered as the feature and w1 as an opinion. Thereafter, the relationships advmod(w1,w5) and neg(w1,w6)are searched.
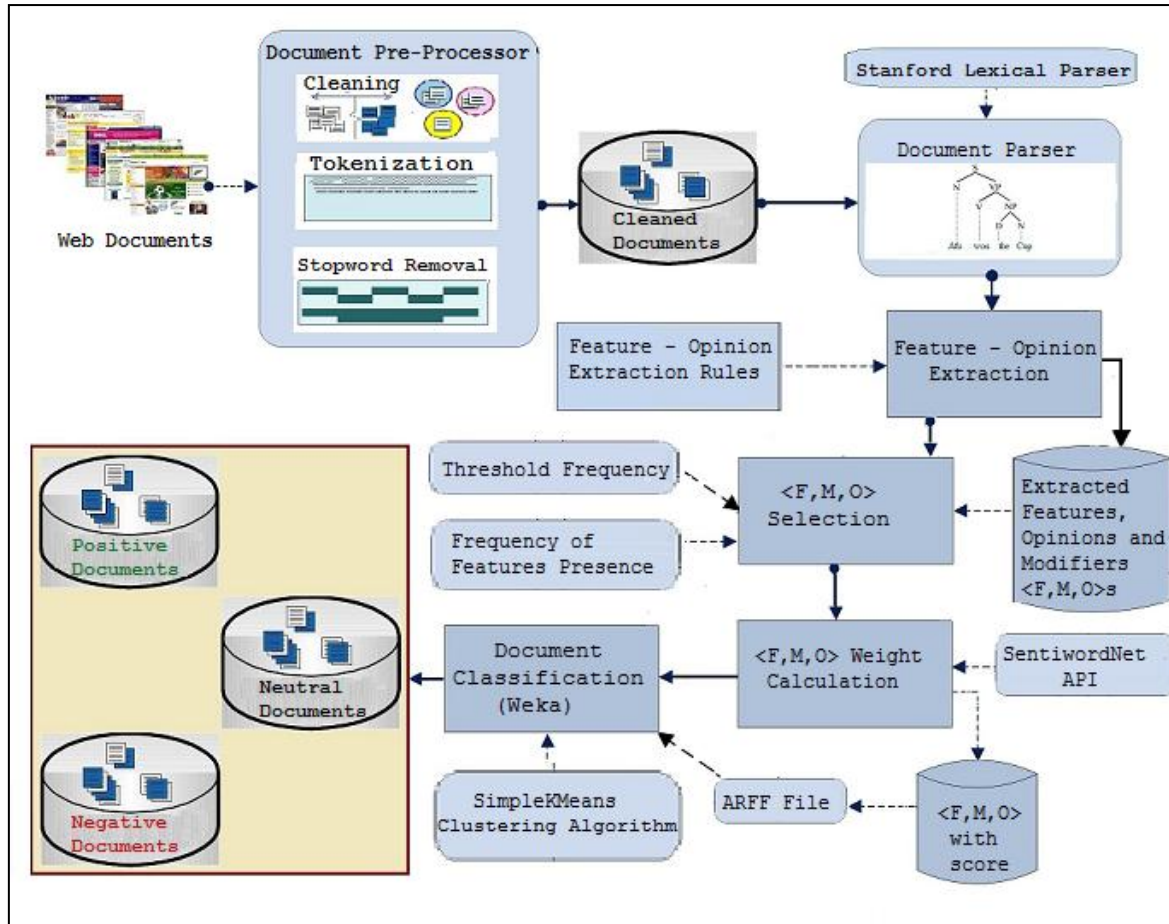
Figure 2: Fuzzy based opinion mining and sentiment analysis framework

If both exist together then (w6 w5) will be the modifier or if advmod(w1,w5) only exist then the modifier will be w5 or if only neg(w1,w6)exist then modifier will be w6.

Rule-6: In a dependency relation R, if there exist relationships amod(w1,w2) and amod(w1,w3) such that POS(w1) = JJ* and POS(w2) = POS(w3) = NN* and w1,w2 and w3 are not stop-words then (w3 w1) is considered as the feature and w2 as an opinion. Thereafter, the relationships advmod(w2,w4) and neg(w2,w5)are searched. If both exist together then (w5 w4) will be the modifier or if advmod(w2,w4) only exist then the modifier will be w4 or if only neg(w2,w5)exist then modifier will be w5.

3.2.1 Feature Filtering
The features, modifiers and opinions <F,M,O> collected using the above rules again filtered out by applying a threshold frequency limit. If the feature occurrences are below the specified threshold limit then it will be ignored. System used a threshold frequency of 3 for eliminating irrelevant feature.

3.2.2 SentiWordNet Score Calculation
The score of each opinionated word is calculated using SentiWordNet database. In order to find the corresponding score opinionated term and its part-of-speech is required. A score of zero is assigned to those terms that are not present in the SentiWordNet. For instance, happy: a: 0.2881 where happy is opinionated word and a represents adjective.

3.2.3 Fuzzy Functions to Adjust SentiWordNet Score
In order to emulate the effect of linguistic hedges on opinionated word, various fuzzy functions are used. If any modifier or linguistic hedges associated with any opinionated word, then its SentiWordNet score is adjusted with novel fuzzy functions. Let score S be the score of the opinionated word then,

1. If there is no modifier or linguistic hedge exist then

$$adjScore = S$$

2. If the modifier or linguistic hedge is negator such as not, never etc. then
$$adjScore = 1 - S$$

3. If the linguistic hedge is a dilator such as somewhat then
$$adjScore = 1 - (\sqrt{1 - S})$$

4. If linguistic hedge is concentrator such as very, extremely then
$$adjScore = 1 - (1 - S)^2$$

5. If the modifier or linguistic hedge is a combination of 2 and 3 then
$$adjScore = \sqrt{S} * (1 - S)$$

6. If the modifier or linguistic hedge is a combination of 2 and 4 then
$$adjScore = S^2 * (1 - S)$$

3.2.4 *Feature-wise Total Score Calculation*

The SentiWordNet score calculated in the previous step is utilized for feature wise SentiWordNet score calculation. Feature wise SentiWordNet score is calculated by using the following equations:

$$P_S = \sum_{i=1}^{n} S_i^+ \qquad (1)$$

$$N_s = \sum_{j=0}^{m} S_j^- \qquad (2)$$

Where n is total number of features having positive SentiWordNet score, m is total number of features having negative SentiWordNet score, S^+ is positive SentiWordNet score of the opinion, S^- is negative SentiWordNet score of the opinion, P_s is total positive SentiWordNet score of the feature S and N_s is total negative SentiWordNet score of the feature S. Total SentiWordNet score is calculated using the equation 3.

$$T_s = P_s - N_s \qquad (3)$$

Based on the total score calculated in the previous step, the features are segregated as positive features and negative features.

Using the negative and positive attributes obtained from the previous step, a feature matrix is created for the testing dataset that is an n by m matrix where n represents the total number of documents and m the total number of terms or attributes in the entire dataset. This matrix has a unique id for each term stored as the attribute heading and an id for each document stored as rows. At the intersection of each column and row is the score of the term i.e., if the particular positive term contains in the document, its score will be 1. Similarly if any negative terms exist in the document then its value will be -1.

3.3 Sentiment Classification Step

For classifying the documents, System applied a SimpleKMeans Clustering Algorithm, which is a part of Weka Data Mining software. The ARFF file generated using the previous step output is given as input to the clustering algorithm.

4. EXPERIMENTAL EVALUATION AND RESULT

This section presents the results of experimental evaluation of the system. The system is evaluated using a dataset of over 2500 user-generated reviews of smart phones (downloaded from www.cs.uic.edu/~liub/FBS/CustomerReviewData.zip and www.cs.uic.edu/~liub/FBS/Reviews-9-products.rar). System used 35% of the review database as training set and 65% as the test set. The working of the system can be summarized as follows:

- Firstly, as explained in Section 3.1, the documents are cleaned by removing the html tags present in the document and by correcting spelling errors. The texts are then tokenized into tokens and the stop-words are detected and removed. Stanford Linguistic parser is used for POS tagging of each term.

- By applying the six rules defined in Section 3.2, the features, opinions and modifiers are extracted. By applying a threshold frequency limit of 3, the irrelevant terms are filtered out. The polarities of extracted opinions are calculated using SentiWordNet API against each feature. If any linguistic hedge associated with the opinionated word the scores are re-calculated using the equations given in the Section 3.2.3.
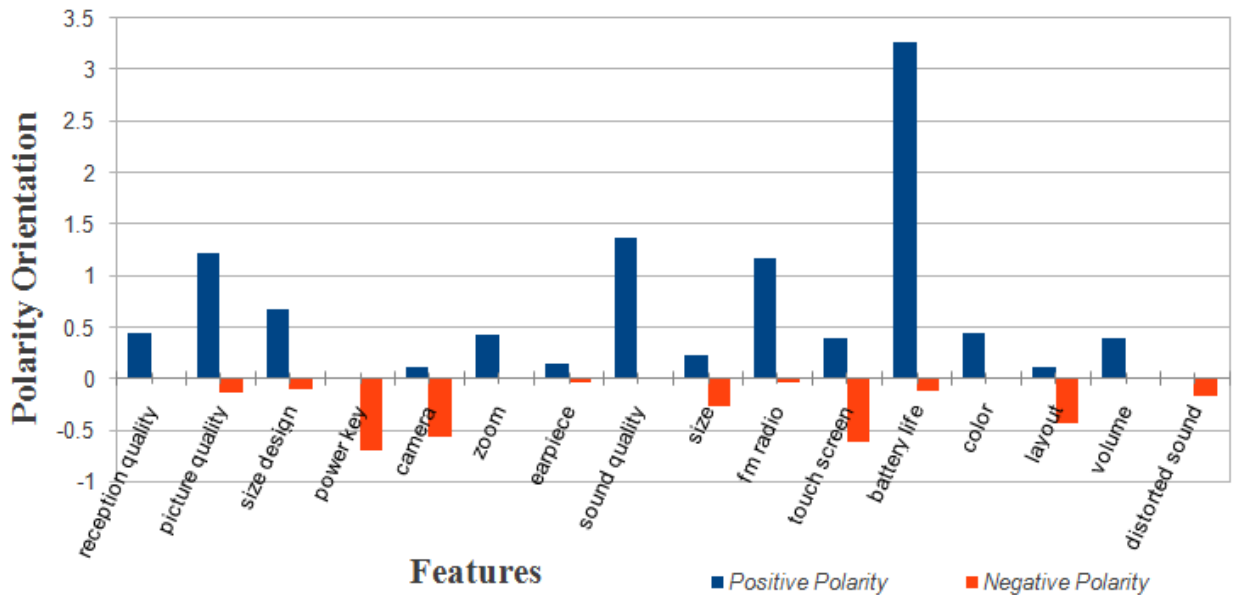

Figure 3: A feature-wise polarity orientation summary

- Against each feature, the adjusted scores of opinionated words are examined and the max functions (equations 1 and 2 in the Section 3.2.4) are applied to calculate the total positive and total negative score of the each feature. By using equation 3, features are grouped them into positive features set and negative features set.

- A feature-wise polarity orientation summary is generated from the negative and positive feature sets obtained in the previous step. A graphical representation of this summary is presented to the user as shown in Figure 3.

- Finally, the whole documents are classified using SimpleKmeans clustering algorithm, which is a part of WEKA data mining software, as explained in Section 3.3. This unsupervised clustering algorithm classified the given document into 3 categories: "positive", "negative", and "neutral" reviews as shown in Figure 4. Table 1 depicts the overall system performance while evaluating system with 2500 user-generated reviews of smart phones.
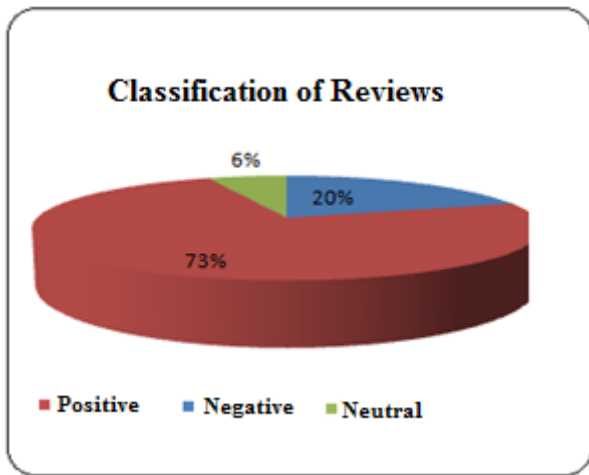

Figure 4: Classification of reviews of smart phone

Table 1: Performance of the system – dataset: customer reviews of smart Phone

Measures	Values
TP	43
FP	18
TN	445
FN	14
Precision (%)	70.49
Recall (%)	75.45
F1-measure (%)	72.88
Accuracy (%)	93.85

As shown in Table 1, the system achieved 70.49% precision and 75.45 % recall. The recall value is greater than precision indicates that the system recognized correct feature-opinion pairs. System yielded 93.85 % accuracy in the feature-opinion extraction process. System yielded 93.85 % accuracy in the feature-opinion extraction process. The system achieved an accuracy of 75.51% while classifying reviews using SimpleKMeans clustering algorithm.

5. CONCLUSION AND FUTURE WORK

Experimental results show that, the system performs very well in sentiment classification of user reviews with high accuracy. The implemented fuzzy functions to emulate the effect of various linguistic hedges such as dilators, concentrator and negation on opinionated phrases help the system to achieve more accuracy in sentiment classification and summarization of users' reviews.

As future work of this research, we can refine rule set to extract more dependency relations from datasets and that will help to improve the precision and recall values of the system. From the analysis of review documents, it was observed that most of the reviewers do not follow the grammatical rules while writing reviews due to which the parser fails to assign correct POS tag. It in turn fails the system from defining correct dependency relations between word pairs. If the system able to correct all the spelling and grammatical errors present in the review documents in the pre-processing step itself that will definitely improve the recall value of the system.

The system used SentiWordNet as a source of information for determining term orientation. It is a good idea to use SentiWordNet with other scoring measures to arrive at better scores for terms. It will surely generate a better score and will make up for the inaccurate scores generated sometimes from the SentiWordNet. From the analysis of review document, it was also realized that many review documents contain junk sentences, which opens a new area of research called review spam analysis. The other thing noted in the review documents was users generally prefer to express their emotions and feeling in short sentences and in abbreviations, hence the research should be carried out in that direction as well.

6. REFERENCES

[1] S. R. Das and M. Y. Chen. "Yahoo! for amazon: Sentiment extraction from small talk on the web", Management Science, 53(9):1375–1388, 2007.

[2] Sentiment Analysis - Wikipedia, the free encyclopedia. Retrieved on December 1, 2013, from the World Wide Web: http://en.wikipedia.org/wiki/Sentiment_analysis

[3] V. P. H. Binali and W. Chen. "A state of the art opinion mining and its application domains", In IEEE International Conference on Industrial Technology, pages 1–6, February 2009.

[4] Bing Liu. (2010), "Sentiment Analysis and Subjectivity", Handbook of Natural Language Processing, Second Edition by N. Indurkhya and F. J. Damerau.

[5] Kushal Dave, Steve Lawrence, and David M. Pennock. (2003), "Mining the peanut gallery: Opinion extraction and semantic classification of product reviews", In Proceedings of WWW. Pages: 519–528.

[6] M. Hu and Bing Liu. (2004), "Mining and summarizing customer reviews", Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data, Aug. 22-25, ACM Press, Washington, USA., pp: 168-177.

[7] Dunja Mladenic. (1999), "Text-Learning and Related Intelligent Agents: A Survey", In Intelligent Systems and their applications, IEEE, Vol. 14 Issue 4. Pages: 44-54.

[8] C. M. Kristina Toutanova, Dan Klein and Y. Singer. "Feature-rich partof-speech tagging with a cyclic dependency network". In HLT-NAACL, pages 252–259. ACM, 2003.

[9] Ms.K.Mouthami, Ms.K.Nirmala Devi, Dr.V.Murali Bhaskaran, "Sentiment Analysis and Classification Based On Textual Reviews", Information Communication and Embedded Systems (ICICES), 2013 International Conference on 21-22 Feb. 2013Page(s): 271 – 276.

[10] A. Kennedy and D. Inkpen, "Sentiment classification of movie reviews using contextual valence shifters", Computational Intelligence, vol. 22, no. 2, pp. 110–125, 2006.

[11] L. Polanyi and A. Zaenen, "Contextual valence shifters", in Computing Attitude and Affect in Text: Theory and Applications, vol. 20 of The Information Retrieval Series, pp. 1–10, 2006.

[12] A.D. Vo and C.Y. Ock, "Sentiment classification: a combination of PMI, SentiWordNet and fuzzy function", in Proceedings of the 4th International Conference on Computational Collective Intelligence Technologies and Applications (ICCCI '12), vol. 7654, part 2 of Lecture Notes in Computer Science, pp. 373–382, 2012.

[13] S. Nadali, M. A. A. Murad, and R. A. Kadir, "Sentiment classification of customer reviews based on fuzzy logic", in Proceedings of the International Symposium on Information Technology (ITSim' 10), pp. 1037–1044, mys, June 2010.

[14] M. Hu and B. Liu, "Mining and summarizing customer reviews", Proceedings of the tenth ACM international conference on Knowledge discovery and data mining, Seattle, 2004, pp. 168-177.

[15] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques", Proceedings of the ACL-02 conference on Empirical methods in natural language processing, vol.10, 2002, pp. 79-86.

[16] K. Dave, S. Lawrence, and D. M. Pennock, "Mining the peanut gallery: Opinion extraction and semantic classification of product reviews", Proceedings of WWW, 2003, pp. 519–528.

[17] R. Prabowo and M. Thelwall, "Sentiment analysis: A combined approach", Journal of Informetrics, vol. 3, pp.143-157, 2009.

[18] B. Pang and L. Lee, "Opinion mining and sentiment analysis", Foundations and Trends in Information Retrieval 2(1-2), 2008, pp. 1–135.

[19] A. Harb, M. Planti, G. Dray, M. Roche, Fran, o. Trousset and P. Poncelet, "Web opinion mining: how to extract opinions from blogs?", presented at the Proceedings of the 5th international conference on Soft computing as transdisciplinary science and technology, Cergy-Pontoise, France, 2008.

[20] P. Turney, "Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews", Proceedings of the Association for Computational Linguistics (ACL), 2002, pp. 417–424.

[21] Z. Wang, Y. He, and M. Jiang, "A comparison among three neural networks for text classification", in Proceedings of the 8th International Conference on Signal Processing (ICSP '06), pp.1883–1886, November 2006.

[22] R. D. Goyal, "Knowledge based neural network for text classification", in Proceedings of the IEEE International Conference on Granular Computing (GrC '07), pp. 542–547, November 2007.

[23] L. Dey and Sk. M. Haque, "Opinion mining from noisy text data", International Journal on Document Analysis and Recognition, vol. 12, no. 3, pp. 205–226, 2009.

[24] D. Isa, L. H. Lee, V. P. Kallimani, and R. Rajkumar, "Text document preprocessing with the bayes formula for classification using the support vector machine", IEEE Transactions on Applied Computational Intelligence and Soft Computing 9 Knowledge and Data Engineering, vol. 20, no. 9, pp. 1264–1272, 2008.

[25] M. K. Dalal and M. A. Zaveri, "Semisupervised learning based opinion summarization and classification for online product reviews", Applied Computational Intelligence and Soft Computing,vol. 2013,Article ID 910706, 8 pages, 2013.

[26] L. Zhao and C. Li, "Ontology based opinion mining for movie reviews", in Proceedings of the 3rd International Conference on Knowledge Science, Engineering and Management, pp. 204–214, 2009.

[27] D. Sleator and D. Temperley, "Parsing english with a link grammar", in Proceedings of the 3rd International Workshop on Parsing Technologies, pp. 1–14, 1993.

[28] S. Shi and Y. Wang, "A product features mining method based on association rules and the degree of property co-occurrence", in Proceedings of the International Conference on Computer Science and Network Technology , vol. 2, pp. 1190–1194, December 2011.

[29] C.-P.Wei,Y.-M.Chen,C.-S.Yang, and C. C.Yang, "Understanding what concerns consumers: a semantic approach to product feature extraction fromconsumer reviews", Information Systems and e-Business Management, vol. 8, no. 2, pp. 149–167, 2010.

[30] W. Zhang, T. Yoshida, and X. Tang, "Text classification using multi-word features", in Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC '07), pp. 3519–3524, October 2007.

[31] M. K. Dalal and M. A. Zaveri, "Automatic text classification of sports blog data", in Proceedings of the Computing, Communications and Applications Conference, pp. 219–222, January 2012.

[32] W. Zhang, T. Yoshida, and X. Tang, "TFIDF, LSI andmulti-word in information retrieval and text categorization", in Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC '08), pp. 108–113, October 2008.

[33] Ahmad Kamal, Muhammed Abulaish and Tarique Anwar. "Mining Feature-Opinion Pairs and Their Reliability Scores from Web Opinion Sources", in ACM 978-1-4503-0915,2012.

[34] Abulaish Muhammad, Jahiruddin, M. N. Doja and T. Ahmad, "Feature and Opinion Mining for Customer Review Summarization", in Proceedings of the 3rd International Conference on Pattern Recognition and Machine Intelligence (PReMI 09), LNCS 5909, pp 219-224, 2009.

[35] V. Hatzivassiloglou and K. Mckeown, "Predicting the semantic orientation of adjectives", in Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics (ACL '98), pp. 174–181, 1998.

[36] P. D. Turney and M. L. Littman, "Measuring praise and criticism: inference of semantic orientation fromassociation", ACM Transactions on Information Systems, vol. 21, no. 4, pp. 315–346, 2003.

[37] G. A.Miller, "WordNet: a lexical database for English", Communications of the ACM, vol. 38, no. 11, pp. 39–41, 1995.

[38] S. Baccianella, A. Esuli, and F. Sebastiani, "SentiWordNet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining", in Proceedings of the 7th International Conference on Language Resources and Evaluation, pp. 2200–2204,2010.

[39] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning-II", Information Sciences, vol. 8, no. 4, part 3, pp. 301–357, 1975.

[40] T. Zamali, M. A. Lazim, andM. T. A. Osman, "Sensitivity analysis using fuzzy linguistic hedges", in Proceedings of the IEEE Symposium on Humanities, Science and Engineering Research, pp. 669–672, 2012.

[41] The Stanford Parser: A statistical parser, Available at: http://nlp.stanford.edu/software/lex-parser.shtml

[42] Penn Treebank Tagset, Retrieved on March 15, 2014, from the World Wide Web: http://www.computing.dcu.ie/~acahill/tagset.html