KIET IJCE KIET International Journal of Communications &Electronics
Volume. No. 2, Issue No. 2, July – Oct 2014, ISSN: 2320 - 8996

An Intelligent Opinion Mining for Customer Reviews

Minal M.Thawakar[1], Prof.Sheikh Phiroj[2]
Department of Computer Technology, Priyadarshini College of Engineering, Nagpur, India

***Abstract*- Today opinion mining is fast growing topic as more and more people use internet in their day to day life and share information globally. Due to increase in social networking and micro-blogging people share their views, opinion and emotions with each other. This information is necessary for the organization that are selling or manufacturing products in order make changes in design and other configuration of the product. As number of customers give their reviews on same product so that it is difficult for the new user to take decision whether to buy a product or not. The proposed system gives an enhanced summarized result for newly users in order to take fast decision. It extends the level of feature based opinion classification. A soft computing technique is used to classify opinion into positive and negative opinion. The result is shown in textual as well as in graphical form. It not only gives the summary of individual model of a car but also compare two models of a car. By using proposed system, it is easy for user to learn the characteristics of a model as well as to take decision.**

Keywords: **Opinion mining, feature extraction, soft computing, ruled based fuzzy logic, opinion classification, summarization.**

I. INTRODUCTION

World Wide Web contain huge amount of information which consist of online opinion such as political affair comments, news comments, product comments etc. Internet users express their personal views on review websites, blogs, and discussion forum and so on. This information is publically available to users through internet. However large amount of opinion on the Web makes it difficult to get important information. Reading all reviews is a time consuming and confusing task. The research has been going on mining customer reviews which is called as opinion mining.

Opinion mining is the field of study people's emotions, views, experiences from written language. It is the field of natural language processing, data mining, machine learning, and artificial intelligence and so on. The interest in opinion mining has been increased due to increase in social media services such as review sites, micro-blogs, online social network etc. Today online opinion is very beneficial for businessman for collection of feedback from customers so that necessary changes can be made with product. Opinion mining is not only useful for consumer but also for producer.

In this research, we propose a feature based opinion summarization on customer reviews of automobiles. The task is performed in following steps:

1. Identify the features of automobiles that customer have commented.
2. Show all reviews of automobile according to selected features.
3. Now according to selected feature divide reviews into positive and negative reviews.
4. Show overall summary of reviews using graph which consist of histogram of positive and negative review according to features.
5. According to features show individual graphical summary of positive and negative reviews.
6. Finally comparison is made between two models and show graphical summary

We give a simple example to demonstrate the above steps. Consider the review of automobile having company and model as Maruti Suzuki and Alto K10 respectively.

Company: Maruti Suzuki
Model: Alto K10
Review all:
Look and style:
<individual comments>
< individual comments>
< individual comments>

Comfort:
<individual comments>
< individual comments>
< individual comments>

Positive reviews:
Look and style:
<individual comments>
< individual comments>

< individual comments>

Comfort:
<individual comments>
< individual comments>
< individual comments>

Negative reviews:
Look and style:
<individual comments>
< individual comments>
< individual comments>

Comfort:
<individual comments>
< individual comments>
< individual comments>

For the above example review all consist of all reviews that is the combination of positive and negative reviews and positive and negative reviews contains positive opinion reviews and negative opinion reviews from review all respectively.

Our task is different from others in number of ways. We textually as well as graphically represent the summary of reviews. We mainly focused on features of automobiles that customers have opinion and also whether the opinion is positive or negative according to features.

A soft computing technique is used to perform classification of opinion into positive and negative opinion. Here soft computing technique called rule-based fuzzy logic is used where fuzzy function is used to calculate the score of the sentence.

Section 2 describes related work on feature mining. Section 3 contains detail working of proposed system. Section 4 describes the experimental results. Finally we give conclude and give direction for future work in opinion mining field.

II. RELATED WORK

The following section explains the survey of various papers based on feature based opinion mining. Several methods have been proposed to extract feature from customer reviews.

Hu and Liu [3] worked on customer reviews. They extract features of product from customer reviews. They also show whether the opinion is positive or negative and finally summarize the result. The main problem with the method employed by this system is it does not give proper relation between feature and its opinion. In our work we not only decide whether the opinion is positive or negative but also calculate the score of the word which is used for the plotting graph.

Lawrence and David [4] developed an automatic method which distinguishes positive and negative reviews. Aciar [5] proposed a feature extraction method based on ontology. But the main problem with this is, whenever new feature is added there is a need of construction and updating of ontology each time. Hana Jeong, Dongwook Shin, and Joongmin Choi [11] developed an enhanced feature extraction technique called FEROM which effectively extract correct feature from the review data but the main problem with this method it does not calculate the strength of the opinion and does not give proper summarize form. But in our approach we not only calculate the strength of the opinion but also give a detailed summary of each model according to selected features

Various machine learning methods and training sets are used to perform automatic text classification. Machine learning methods such [13] as Support Vector machine (SVM), Artificial intelligence [14], Naïve Bayesian [12] or hybrid approaches [15, 16] are used to improve the efficiency of classification. But all these methods are not focused on generating extractive summaries.

III. PROPOSED TECHNIQUE

The following section describes the design of our proposed feature based opinion mining system based on rule-based fuzzy logic. Our proposed opinion systems automatically extract the opinion from unstructured user reviews and classify the opinion into positive and negative opinion according to the assigned polarity. Polarity is a sometime considered as intensity. The proposed systems consist of following steps: 1) Data Preprocessing 2) Feature and opinion generation 3) Opinion Classification 4) Summary.

We perform opinion mining on online reviews of various car models. Reviews are collected from different automobiles websites having engine capacity between 800-1000 cc. The websites that we are using for collection of reviews are www.carwale.com and www.cardekho.com. We collect approximately 450 reviews for 9 car models where each model having approximately 50 reviews.

Figure 1 shows the system architecture. The different steps of the system architecture are explained below.

A) Data Preprocessing: User's opinions are generally expressed in natural language which contains errors in spelling, grammar, mistakes in punctuations and so on. Before mining user-

generated reviews need preprocessing in order to remove noise. An openNLP tool is used to perform this task. openNLP tool is a machine learning natural language processing tool used to perform following task such as sentence detection, tokenization, POS tagging, chunking, parsing. The version of openNLP tool use is 1.5.3. openNLP tool consist Standford parser which is a well known linguistic parser. It automatically corrects the unstructured text and produces a clean text.

Consider an example; a user-generated review is in the form of

"the look and style of a car is so nice!!!"

After preprocessing the sentence would read as,

"The look and style of a car is so nice!"

In the above example, the first sentence is capitalized and the repetitive exclamation mark occurs only once.

Thus the preprocessing step generates a clean text which is given as input to the next step of the system.

B) Feature and Opinion generation: In this step we generate a feature set for opinion mining from cleaned reviews generated in first step by using linguistic parser. Here we have created a table name as Feature and Opinion generation table (FOGT). FOGT table consist of car model features as well as positive, negative and inverter words associated with that feature. We manually assign weight to each word in the table. We use a standard WordNet directory for making FOGT table. Table 1. Shows the entries in FOGT. Here POS (part-of-speech) tagging is used to performed feature extraction. Frequently occur Noun (N) and Noun Phrase (NP) is considered as feature and Adjective or Adverb is considered as opinion word for the sentence.

For example, consider a review of sentence for model of a car creates after preprocessing is:

"The look of car is nice."

When POS tagging is done by using openNLP the output is generated as:

"The look [.n] of [.p] car [.n] is [.v] nice [.a]"

In the above example, [.n] denotes noun, [.p] denotes preposition, [.a] denotes adjective and [.v] denotes verb. Here "look" is considered as feature which is described by adjective "good". Adjective "good" shows the opinion about feature "look". Whenever POS tagging is done noun, noun phrases, adjective or adverb of the sentence is checked with the entries in the FOGT.

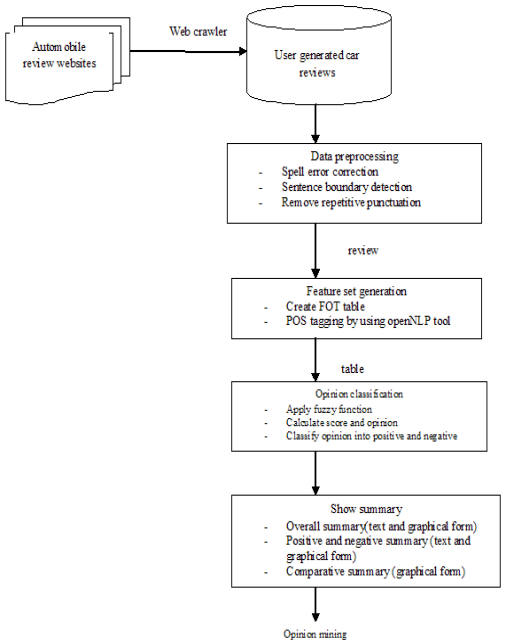

Figure 1. System architecture

Feature	Positive polarity words (p=1)	Negative polarity words (p=-1)	Inverter
Look and style	Excellent, good, best, decent, attractive, great, beautiful, elegant, amazing, stylish	Worst, bad, poor	Not, Never
Comfort	Comfortable, extraordinary	congested	Not
Mileage	Decent, satisfactory, reliable		Not
Pickup	Average, sufficient, high, fine	Slow, less	
Overall experience	Happy	Mistake, down, cheat, waste	
Fuel economy	Superb, fantastic	Low	Not

Table 1. Feature and Opinion Generation Table

Consider the first entry in the FOGT, it indicate the feature "look and style" for a car model. This feature can take fuzzy values like "excellent", "good", "best", "decent", "attractive", "great", "beautiful", "elegant", "amazing" and "stylish" which has positive polarity. On the other hand it can take fuzzy values like "worst", "bad" and "poor" which have negative polarity. The polarity of words is reversed by inverter which has words like "not" and "never."

The FOGT table is used in next step to calculate the score of the user review sentence and classify it into positive and negative review.

C) Opinion Classification: In this step, the reviews are classified into positive and negative reviews. We classify new user review by calculating fuzzy score. The fuzzy score is calculated by using following steps. 1) Extract feature and words from FOGT look up table. 2) Identify the polarity and initial weight of the word. 3) Calculate overall score using fuzzy function. The first two steps of fuzzy score calculation are explained in above second step of system architecture. The fuzzy score of the word can be calculated as:

$$f(x) = 1-(1-w(x)) \quad (1)$$

Where $f(x)$ represents fuzzy score of the word and $w(x)$ represents the initial weight of the word which s assigned in FOGT table. Consider $\forall x, (x) \in [0, 1]$ which indicate the output value of $f(x)$ should be in the range of [0, 1]. For example, the initial assigned weight of the word "good" is 0.5, hence fuzzy score for the particular word is calculated as $f(x) = 0.5$. The FOGT table contains polarity for identification of positive and negative opinion. Hence the modified fuzzy score can be calculated as:

$$f(x)_{new} = p_i f(x) \quad (3)$$

In the above equation, the first term indicate the polarity of the feature. If polarity is positive then the value of $p_i = 1$ otherwise the value of $pi = -1$. If inverter is present (ex. "not") then simply reverse the value of polarity which is indicated by '*pi'*. The inverter only changes the polarity of the feature but the magnitude remains unchanged. Equation (3) is used to plot overall graph. The value of $f(x)_{new}$ calculated using (3) remains in the range of [-1,1] so we need to normalized the value of $f(x)_{new}$ .The normalized value should be in the range of [0,1],
It is calculated in the following manner:

$$f_n(x) = \frac{f(x)new + 1}{2} \quad (4)$$

The normalized value is generally used to plot positive/negative graph. The accuracy of this classification is verified by comparing them. Comparison is made on the basis of features of two different models of car.

D) Summary: This is the last step of system architecture. Here we have generated a detailed summary of reviews according to selected features in textual as well as in graphical format. It consists of three types of graphs. 1) Overall graph. 2) Positive/Negative graph and 3) Comparison Graph. Figure 1, figure 2 and figure 3 show the graphical summary of reviews for car model Alto-K10.

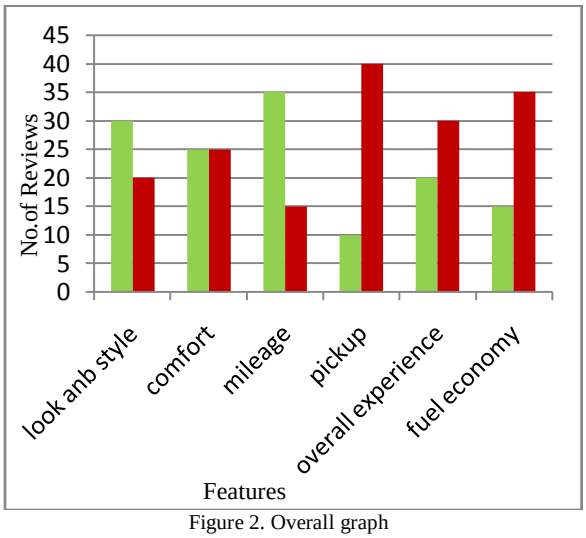

Figure 2. Overall graph

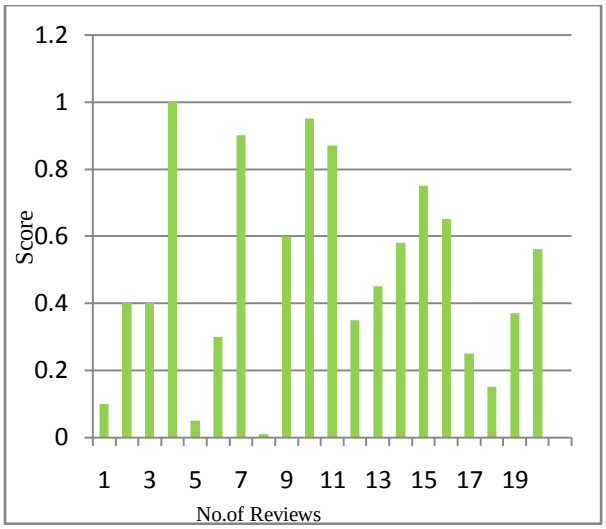

Figure 3. Positive graph

The overall graph contains total number of positive and negative reviews according to selected features. We have taken 50 reviews for each model. Consider for feature overall experience there are 20 positive reviews and 30 negative reviews so the overall graph contain 20 positive and 30 reviews for feature overall experience.

Positive and negative graph contain individual positive and negative reviews according to the score calculated by using equation (4). Overall graph and Positive/ Negative graph is used for verifying individual car model and comparison graph is used for verifying two different models of car.

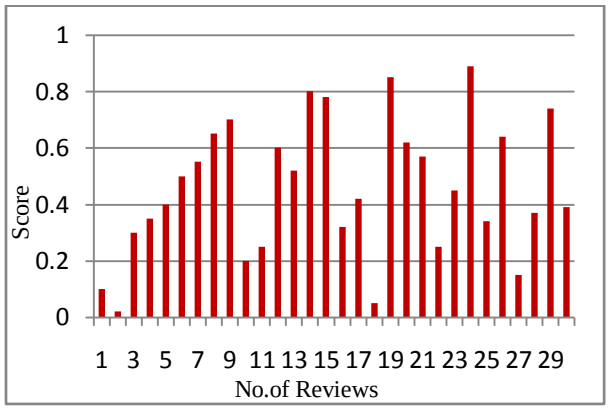

Figure 4. Negative graph

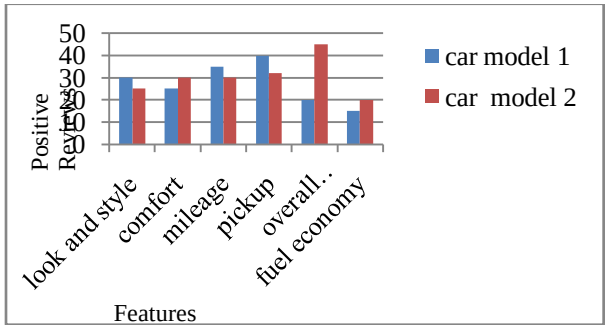

Figure 5.Positive Review Comparison graph

IV. EXPERIMENT RESULTS

This section describes the result of evaluation of our opinion mining strategy. We have conducted the experiment on customer reviews of four different car companies which consist of total nine models. The websites from where we collected the reviews are carwale.com cardekho.com. For each model we first loaded 50 reviews for each model. These reviews were cleaned by using linguistic parser. Feature extraction and opinion word were found by using FOGT table and fuzzy logic respectively. Figure 6. represents the evaluation result. Here customer can get detail summary about car model by selecting company, model and features

V. CONCLUSION

In this paper, we propose a framework for automobiles review extraction based on features. Our experimental results indicate that the proposed technique is very promising in performing the task of opinion mining. We not only predict the positivity and negativity of opinion word, but also calculate the score. The summary is not only useful to common customers, but also for product manufactures.

The primary area of future work is to improve the method of calculate the score/strength of words and use our summarization system end to end in practice

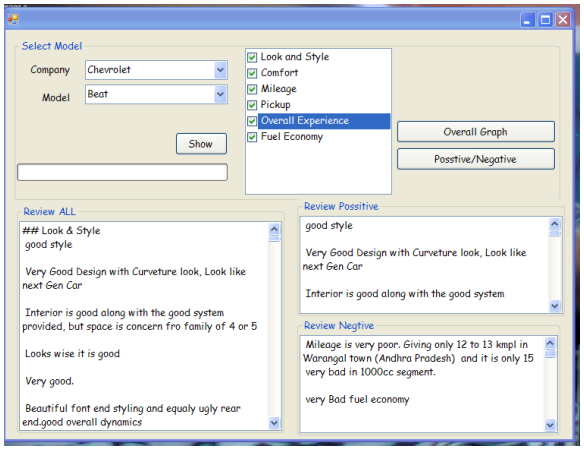

Figure 6. Experiment result

VI. REFERENCES

[1] LiZhen Liu, WenTao Wang, HangShi Wang. Summarizing Customer Reviews Based On Product Features. 2012 5th International Congress on Image and Signal Processing (CISP 2012)

[2] Bing Liu. Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers, 2012.

[3] Mingqing Hu and Bing Liu. Mining and Summarizing Customer Reviews. KDD'04, August 22-25,2004, Seattle, Washington, USA.

[4] David M.Blei, Andrew Y.NG and Michael I.Jordan. 2003. Latent Dirichlet allocation. Journal of Machine Learning Research,3(5):993-1022.

[5] S.Aciar et al., "Information Recommender:Basing Recommendations on Consumer Product Reviews," *IEEE Intell. Syst.* Vol. 53, no.9, 2001, pp.1375-1388.

[6]. Hana Jeong, Dongwook Shin, and Joongmin Choi. FEROM:Feature Extration and Refinement for Opinion Mining. ETRI Journal, Volume 33, Number 5, October 2011.

[7] Nikolay Archak, Anindya Ghose, and Panagiotis G. Ipeirotis. Show me the money!: deriving the pricing power of product features by mining consumer reviews. In Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '07, pages 56–65, New York, NY, USA, 2007. ACM.

[8] Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. Multi-facet rating of product reviews. In Proceedings of the 31th European Conference on IR Research on Advances in Information Retrieval, ECIR '09, pages 461–472, Berlin, Heidelberg, 2009. Springer-Verlag

[9] Wei Jin, Hung Hay Ho, and Rohini K. Srihari. Opinionminer: a novel machine learning system for web opinion mining and extraction. In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '09, pages 1195–1204, New York, NY, USA, 2009. ACM.

[10] Fangtao Li, Minlie Huang, and Xiaoyan Zhu. Sentiment analysis with global topics and local dependency. In Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence (AAAI), 2010.

[11] Hana Jeong, Dongwook Shin, and Joongmin Choi. FEROM:Feature Extration and Refinement for Opinion Mining. ETRI Journal, Volume 33, Number 5, October 2011.
[12] S. Kim, K. Han, H. Rim, and S. H. Myaeng, "Some effective techniques for na¨ıve bayes text classification," *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, no. 11, pp. 1457–1466, 2006.
[13] Z.-Q. Wang, X. Sun, D.-X. Zhang, and X. Li, "An optimal SVM-based text classification algorithm," in *Proceedings of the International Conference on Machine Learning and Cybernetics*, pp. 1378–1381, August 2006.
[14] 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.
[15] 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 Knowledge and Data Engineering*, vol. 20, no. 9, pp. 1264–1272, 2008.
[16] 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.
[17] 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 (ICCSNT '11)*, vol. 2, pp. 1190–1194, December 2011.
[18] LiZhen Liu, WenTao Wang, HangShi Wang, "Summarizing Customer Reviews Based On Product Features" 2012 5th International Congress on Image and Signal Processing (CISP 2012)
[19] Mohammad Khabbaz, Keivan Kianmehr, and Reda Alhajj "Employing Structural and Textual Feature Extraction for Semistructured Document Classification" IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS PART C: APPLICATIONS AND REVIEWS, VOL. 42, NO. 6, NOVEMBER 2012
[20] C. Caragea, A. Silvescu, S. Kataria, D. Caragea, and P. Mitra. "Classifying Scientific Publications Using Abstract Features." In: SARA, Parador de Cardona, Spain, 2011
[21] A. Esuli and F. Sebastiani. 2006. Sentiwordnet: A publicly available lexical resource for opinion mining. In Proc. of LREC 2006.
[22] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment Classification Using Machine Learning Techniques," Proc. EMNLP, Philadelphia, Jul. 2002, pp.79-86.
[23] Minqing H., Bing L. *Mining Opinion Features in Customer Reviews*. *In Proceedings of AAAI'04* Proceedings of the 19th national conference on Artificial intelligence Pages 755-760.
[24]Mita k. Dalal and Mukesh A. Zaveri, "Opinion Mining from Online User Reviews Using Fuzzy Linguistic Hedges", Applied Computational Intelligence and Soft Computing Volume 2014, Article ID 735942, 9 pages