6th International Conference on Multidisciplinary Research (ICMR-2019)
Osmania University Campus, Hyderabad (India)
30th-31st May 2019 www.conferenceworld.in
ISBN : 978-93-87793-89-7

Feature Based Product Rating System Using Opinion Mining

Anil V. Deorankar[1], Shubham V. Pandey[2]

1 (Department of Computer Science and Engineering,
2 Government College of Engineering, Amravati, Maharashtra, India)
3 (Department of Computer Science and Engineering,
4 Government College of Engineering, Amravati, Maharashtra, India)

ABSTRACT

E-commerce and online shopping sites provide their users a platform to share opinions views and experience about the product and/or service used by the customers. Online customer review is an important resource for other customers in the process of product selection and decision making. These online reviews are available in large volume and still increasing day by day. Due to large number of reviews, it is difficult for customers to read analysis review data manually. And thus there is a need of an system which automatically analyze this reviews and present summarized result for the users. In aspect-based sentiment analysis (ABSA) the task is to extract aspect (or feature) and sentiment word in the review sentence. In the proposed system, dependency relation between the words of review sentences are used to identify the aspect terms and their associated opinion word. Negation handling is also performed using the dependency relation and context-based sentiment word dictionary.

Keywords- Opinion mining, aspect category detection, dependency parsing, natural language processing, sentiment analysis.

1. INTRODUCTION

Opinions and feedback are very useful for people specially while purchasing some goods or services. Opinion has major importance in decision making. These opinions are available in large number for internet users in the form of online reviews. The online reviews are useful for both the customers and the merchants . But the problem is that these online reviews are increasing day by day . It is difficult to read and analyze each and every review manually and make decisions. Also the five-star product rating method does not reflect the actual opinions of customers about the different features(aspect) of product like battery , performance, price of laptop product. Hence there is a need of an automated system which can analyze the reviews of product at the feature level and produce feature wise product summary.

Aspect level sentiment analysis system is needed to detect features of product and associated sentiment with each feature mentioned in the review text. For example aspects (features) of laptop products can be battery, cpu

speed, screen resolution, price etc. The proposed system deals with laptop reviews, and sentiment analysis is performed at the aspect level. One of the task in aspect level sentiment analysis process is to detect the aspect terms (or feature of the product/service) mentioned in review. Once the aspect terms are detected, next task is to detect the sentiment specified for the aspect terms in given review. The sentiment can be positive negative or neutral. In the proposed system we have considered only positive and negative sentiments for sentiment classification.

In the proposed system we have used the stanford corenlp depedency parser to detect the aspect terms of laptop product. Different grammatical relations of stanford typed dependencies are used to extract features and sentiment words from laptop review. To detect implicit opinions we have used knn algorithm. Implicit feature category is determined by the algorithm using the sentiment words used in review text.

2. RELATED WORKS

Opinion mining or sentiment analysis (SA) can be performed at three different levels: document level, sentence level and aspect level[2]. Sentiment analysis at sentence level and document level does not capture precisely people's opinion about the entity mentioned in review document. Thus aspect level SA is used to detect the people's sentiments about fine grained features of product.

Anh-Dung Vo et al. in [1] addressed the problem of feature-based summaries of product reviews. To do this a set of natural language processing (NLP) tools were used in this research. Dependency relations are used to extract opinion-aspect relationship knowledge from the review. Bo Pang at el. in [3] used three different learning algorithms for sentiment classification of movie review. . The machine learning algorithm used in this work are : Naive Bayes, maximum entropy classification, and SVM . among these algorithms SVM performance is best.

The work in [4] used the noun and noun phrase frequency to detect the aspect terms. A POS tagger is used to detect the noun phrase and frequently used nouns are used for analysis. Work in [6] proposed the approach to extract aspects and opinions from reviews. Their proposed approach first detects frequently used aspect terms and associated opinions. They used association rule mining to find all frequent itemsets. The opinion words are used to extract the infrequent aspect terms. Work in [5] proposed a sequential pattern-based approach to detect objective aspects and performance of aspect extraction was improved by pruning such objective aspects. The sequential patterns was defined on the basis of word occurrence in the product review. Q. Liu et al. in [7] worked to improved the accuracy of aspect extraction by using semantic similarity among aspects.

The proposed work in [8] have focused on subjective aspects. They proposed a rule based hybrid approach to use sequential patterns and normalized Google distance (NGD) to extract explicit and implicit aspects. Dependency analysis-based unsupervised approach is used [9] to extract Appraisal Expression Pattern (AEP) from reviews. they worked on reviews from hotel, restaurants, MP3 players, and camera domain. Syntactic based approach was used [10] in which syntactic dependency along with aggregate score of sentiment words, SentiWordNet and aspect table together are used for opinion mining. Stanford dependency parser was used for opinion word extraction. aspect level polarity score is calculated using SentiWordNet.

3. METHODOLOGY

The proposed system for opinion mining process is given in the Fig.1. The system processes laptop reviews and provides opinion summary for user of the system. Review dataset of laptop product from the SemEval-2016 task 5 is used in the system. The system consists of two modules that is user module and admin module. In user module, users can view product details and write review about the laptop product. Opinion mining task on user review is performed by system admin, in admin module of proposed system.

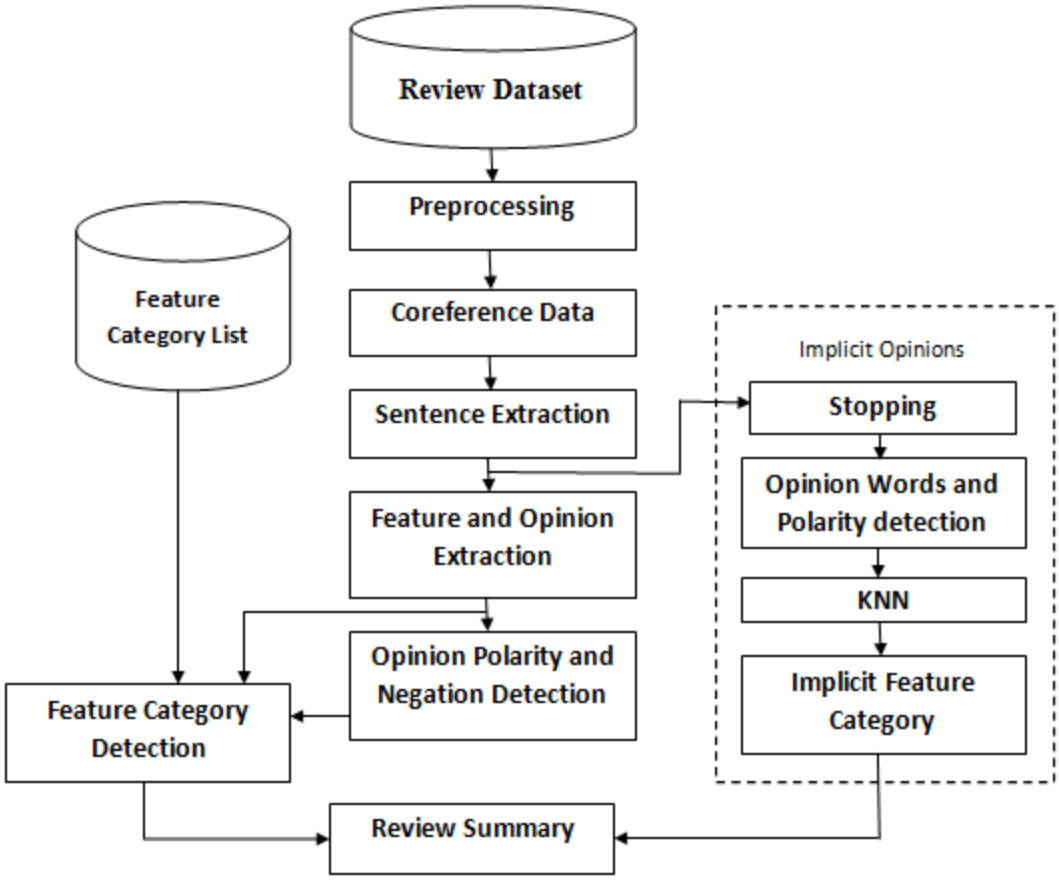

Fig. 1 Proposed System

The following section describes the proposed opinion mining system with each steps in detail.

3.1 Data pre-processing

The first step in proposed system is pre-processing of review dataset. In preprocessing, extract white-spaces before and after review text is removed first. For implicit category detection we have used knn algorithm which extracts opinion words from review sentence. To do this, laptop review is first split into sentences and then stop words are removed from each review sentence. Punctuations symbols, useless words which are very common in user review(like 'a', 'an', 'the', 'in' etc.) are identified as stop word using pre-defined dictionary of words and then this stop words are removed.

3.2 Coreference Data

Next step is to resolve co-references mentioned in review. While writing opinion about products, customers generally describe product details once in review, then they refer the product by using pronouns like 'it' 'this', 'this product', 'it's' etc. For example the review "This laptop is very fast. It's cheap." contains the coreference " It's " which refers to the entity " This laptop ". In coreference resolution step we need to find all words and expressions that refer to the same entity in review text. Fig. 2 shows the output of Stanford corenlp coreference annotator.

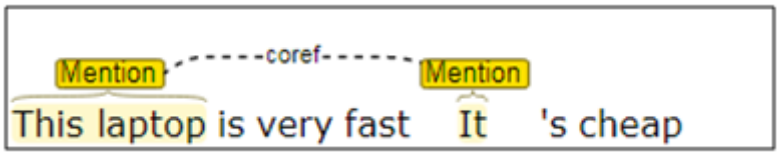

Fig 2 : Coreference resolution using Stanford corenlp

This step is performed before review text is split into individual text sentences. The entire review text is used as input for corefrence resolution. Word index of each reference is stored into database along with reference and it's mention entity. For example the database record for review shown in Fig 2 would be " It - 5 - This laptop ". The first value represents reference for mentioned entity, second value is word position of reference from the beginning of review, and the last entry in the record is the entity itself for which reference is used.

3.3 Sentence Extraction and Stop Word Removal

After coreference resolution step is performed, reviews are split into sentences. Full stop and exclamation sign (!) are used as separators of sentences. For explicit features (the product features which are mentioned in words) the original sentence is used for further processing. For implicit features of product (this are the feature which are not mentioned directly but implied by the meaning of review sentence), stop words like "the", "a" etc. are removed and then resulting sentence without stop words is used for further processing.

3.4 Feature and Opinion Extraction

In this step explicit features and their opinions are extracted from the review sentence. Stanford dependency parser [11] is used to detect the product feature and sentiment word present in review text. Stanford dependency relations[12] are used to detect laptop feature and sentiment word pair. A list of dependency relation used for feature and opinion extraction is given in Table 1.

Table 1 : Dependency relation used for feature and opinion extraction

Relation	Feature	Opinion word
nsubj(JJ, NN)	dep(NN)	gov(JJ)
Nsubj (VBZ/VB, --) + child of governor having POS Tag (JJ, RB)	dep(PRP)	child of gov(JJ/RB)
Nsubj (VBP , --) + child of governor having POS Tag(NN, NNS)	child of gov(NN, NNS)	gov(VBJ)
nsubj(JJ , NN/ NNP) + cop (JJ , --)	dep(NN , NNP)	gov (JJ)
nsubj(JJ/RB , PRP) + child of governor having POS Tag(VBZ)	dep (PRP)	gov (JJ/RB)
nsubj(VB/VBN , PRP) + child of governor having POS Tag(NN)	child of gov(NN)	gov (VB/VBN)
nsubj(VB/VBN , PRP) + child of governor having POS Tag(VBG/VBP)	gov (VB/VBN)	child of gov(VBG/VBP)
case(NN , IN)	gov (NN)	dep (IN)
dobj (VB, PRP)	dep (PRP)	gov (VB)
amod (NN, JJ) + child of governor having relation compound	gov (NN)	dep (JJ)
advmod (NN, RB)	gov (NN)	dep (RB)
advmod (JJ, RB)	dep (RB)	gov (JJ)
nmod(VBD, JJ)	gov(VBD)	dep(JJ)

The dependency relation is a triple representing relation between two words. The first column of Table 1 shows the relation used for opinion-feature pair extraction. Relation name is given outside parenthesis in first line of first column. Upper case words in parenthesis indicates parts-of speech tag (POS tag) of a word present in sentence.

In relation column, the first line represent the relation triple. First phrase within parenthesis is called governor (gov) of relation and second phrase after comma (,) is called dependent (dep) word or node of dependency tree.

POS tag of gov and dep are used to indicate the relation triple. The relation having dash (-) as gov or dep indicates this entry of dependency relation is not considered for feature and opinion extraction. Second column of table indicates the feature word and third column indicates sentiment/opinion word for feature in second column. The Penn Treebank POS tag are used to indicates POS tags of dep , gov and child node of gov/dep.

3.5 Opinion Polarity Detection and Negation Handling

To detect the initial polarity of opinion word we used dataset of positive and negative words. If the opinion word is found in positive word dataset, we assign polarity as positive. Similarly negative opinion are detected. We used the "neg" dependency relation to handle the negative word present in sentence which inverts polarity of opinion words.

3.6 Feature Category Detection

In the proposed system we have used five different category of product feature namely : overall, cpu, quality, price, usability, battery and graphics. A predefined dataset is used to map the extracted feature into its feature category.

3.7 Implicit Opinion Detection

We have used the knn algorithm to detect implicit opinions mentioned in review sentence. The model is trained for seven feature category mentioned above. In this step sentiment words are used to detect the feature category of implicit opinion.

3.8 Review Summary

Review summary is generated for each feature category of given product for which user provide reviews. The summary is represented in percentage form. We calculate the positive and negative opinions for each feature category of product and then show it in percentage form as shown in fig 3.

ISBN : 978-93-87793-89-7

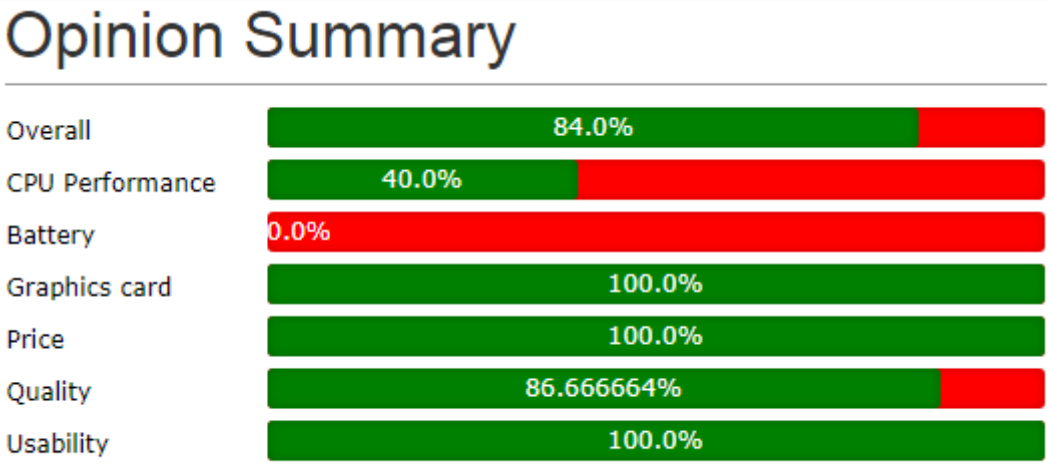

Fig. 3. Review Summary for Product Feature Category

4. RESULTS

The results of the proposed system are measured by the opinion mining parameters i.e. Recall, Precision, F-Measure and Accuracy. As seen in fig. 4 the proposed system has 80.51% precision. We have used 118 review for feature extraction. This review contains 511 individual sentences. Fig 5 shows the frequency graph of dependency relation present in review dataset.

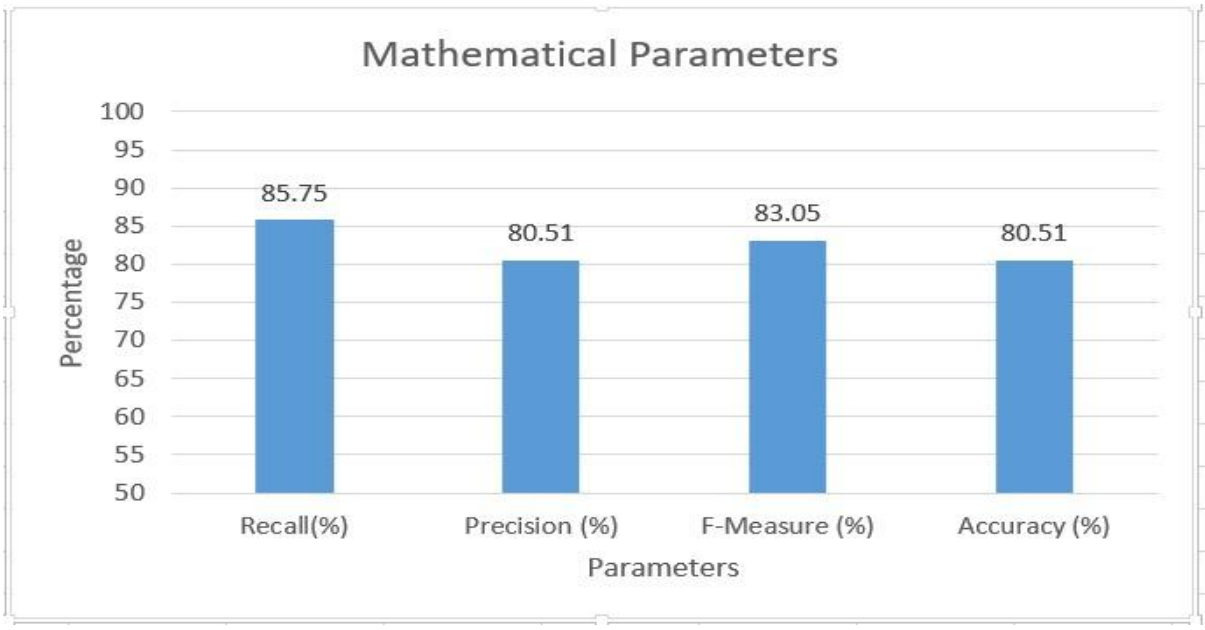

Fig. 4. Mathematical Parameters of Proposed System

ISBN : 978-93-87793-89-7

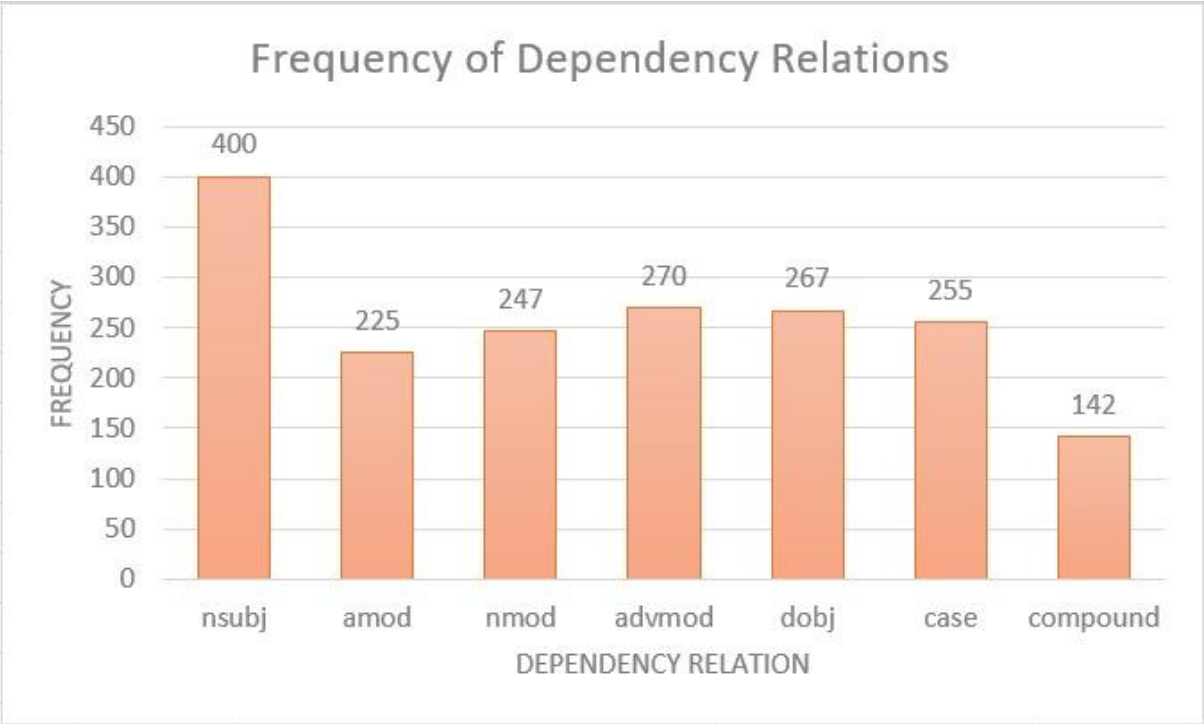

Fig. 5. Frequency of dependency relations found in review dataset.

5. CONCLUSION

The proposed system detects and summarize user opinions at aspect level and present the review summary for different feature categories namely: overall, cpu, battery, graphics, price, usability of laptop, and quality of laptop. Using the stanford dependency parser explicit opinions are detected, and for implicit opinions knn alorithm is used which detects the feature category using sentiment word used in review sentence. Negation handling is also performed in the proposed system.

REFERENCES

[1] Anh-Dung Vo, Quang-Phuoc Nguyen, and Cheol-Young Ock "Opinion aspect relations in cognizing customer feelings via reviews" in Proc. IEEE Access 2018 volume 6, pp 5415-5426

[2] B. Liu, ``Sentiment analysis and opinion mining," Synth. Lectures Human Lang. Technol., vol. 5, no. 1, pp. 1_167, 2012.

[3] B. Pang, L. Lee, and S. Vaithyanathan, ``Thumbs up?: Sentiment classification using machine learning techniques," in Proc. EMNLP, Philadelphia, PA, USA, 2002, pp. 79-86.

[4] M. Hu and B. Liu, "Mining and summarizing customer reviews," Proc. 2004 ACM SIGKDD Int. Conf. Knowl. Discov. data Min. KDD 04, vol. 4, p. 168, 2004

[5] T. A. Rana and T. A. Rana, "Exploiting Sequential Patterns to Detect Objective Aspects from Online Reviews," no. August, 2016.

ISBN : 978-93-87793-89-7

[6]M. Hu and B. Liu, "Mining opinion features in customer reviews," in Proceedings of the 19th national conference on Artifical intelligence, 2004, vol. 4, no. 4, pp. 755–760.

[7] Q. Liu, B. Liu, Y. Zhang, D. S. Kim, and Z. Gao, "Improving opinion aspect extraction using semantic similarity and aspect associations," in Thirtieth AAAI Conference on Artificial Intelligence (AAAI), 2016.

[8] T. A. Rana and Y.-N. Cheah, "Hybrid rule-based approach for aspect extraction and categorization from customer reviews," in Proceedings of the 9th International Conference on IT in Asia (CITA), 2015, pp. 1–5.

[9] X. Zheng, Z. Lin, X. Wang, K.-J. Lin, and M. Song, "Incorporating appraisal expression patterns into topic modeling for aspect and sentiment word identification," Knowledge-Based Syst., vol. 61, pp. 29–47, May 2014.

[10] Heeyoung Lee, Yves Peirsman, Angel Chang, Nathanael Chambers, Mihai Surdeanu, Dan Jurafsky. 2011. Stanford's Multi-Pass Sieve Coreference Resolution System at the CoNLL-2011 Shared Task. In Proceedings of the CoNLL-2011 Shared Task.

[11] C. D. Manning et al., "The Stanford CoreNLP natural language processing toolkit," in Proc. 52nd Annu. Meeting Assoc. Comput. Linguist. Syst. Demonstrations, 2014, pp. 55–60. [Online]. Available at : http://www.aclweb.org/anthology/P/P14/P14-5010.

[12] de Marneffe, M.-C. & Manning, C. D. (2008), "Stanford typed dependencies manual". [Online]. Available : http://nlp.stanford.edu/software/dependencies_manual.pdf.