ISSN:0975-9646

A Peer Review of Feature Based Opinion Mining and Summarization

Padmapani P. Tribhuvan [#1], S.G. Bhirud[* 2] , Amrapali P. Tribhuvan[$3]

[#]*Deogiri Institute of Engineering and Management Studies, Auarangabad, India.*
[*]*Veermata Jijabai Technological Institute, Mumbai, India.*
[$]*Marathwada Institute of Technology, CIDCO, Aurangabad, India*

Abstract-Any organization needs to conduct surveys and collect reviews, in order to improve their product quality. There are number of websites which deals with product reviews. All these reviews are nothings but the opinions of people all over the world about different products. These reviews are very huge and difficult to analyze. Opinion mining is a discipline which deals with analysis of such reviews. Feature based or feature level opinion mining is one of the opinion mining tasks. This paper focuses on feature based opinion mining.

Keyword- Opinion Mining, Information Retrieval ,Sentiment Classification, SentiWordNet, Feature based opinion mining and summarization.

I. INTRODUCTION

Opinion mining is the computational study of people's opinions, appraisals and emotions toward entities, events and their attributes. It involves techniques from different disciplines like information retrieval, natural language processing and data mining [1]. It is very challenging to mine opinions from reviews which are in natural language.

Opinions are so important that whenever one needs to make a decision, one wants to hear others' opinions. This is true for both individuals and organizations. If an individual wants to purchase a product, it is useful to see a summary of opinions of existing users so that he/she can make decision. This is better than reading a large number of reviews. He/she can also compare the summaries of opinions of different products, instead of reading a large number of reviews.

Using opinion mining a review can be evaluated at three different levels- at document level, sentence level and feature level. When review is evaluated at document level, whole review is classified into either positive or negative depending upon the opinion expressed in that review. When review is evaluated at sentence level, then each sentence in a review is classified into either positive or negative. Whereas feature level or feature based opinion mining gives summary which feature of product is liked or disliked by reviewer.

Document level opinion mining consider as a simple text classification problem. It is also called as Sentiment Classification. Feature based opinion mining and summarization can't be considered as text classification problem. How feature based opinion mining and summarization can be done is discusses in this paper.

This paper is organized as follows: Section 2 discusses feature level opinion mining and summarization and some basic steps for sit, Section 3 focus on different approaches proposed by researchers, Section 4 discusses performance measures.

II. FEATURE BASED OPINION MINING AND SUMMARIZATION

As stated above, opinion mining can be done at document level, sentence level and feature level. This paper focus on feature level/ feature based opinion mining. The major tasks of feature based opinion mining are - (1) to identify the products features in review, (2) to determine opinion expressed by the reviewer (positive, negative or neutral), (3) summarize discovered information.

Some researchers have proposed ontology based feature based opinion mining [2], domain-specific opinion extraction [3], and some other automated techniques for feature based opinion mining [1]. Almost all of these follow some basic steps to achieve their goal. These steps are shown in following Fig.1:

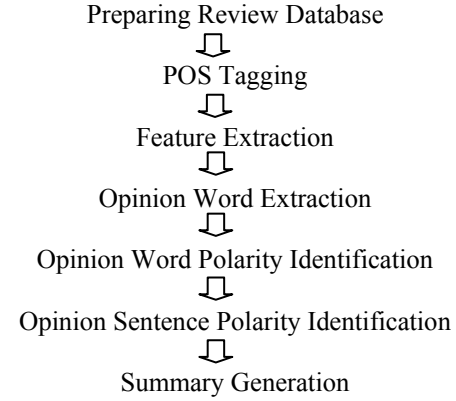

Fig.1 Basic Steps for Feature Based Opinion Mining and Summarization

A. Preparing Review Database

Reviews are extracted from different websites and then we can store those reviews into storing into review database. Each website has its own structure. Web crawlers can be used to download reviews. Web crawlers, also known as spiders or robots, are programs that automatically download Web pages. A crawler can visit many sites to collect information that can be analyzed and mined in a central

location, either online (as it is downloaded) or off-line (after it is stored).

After this preprocessing is done where unwanted text (other than product reviews) is removed and then reviews are stored into database.

B. Part-of-Speech Tagging (POS Tagging)

The aim feature based opinion mining is to find out product features and opinion words (opinion words means words which express opinion). And then find polarity of opinion word. In general, opinion words are adjectives and product features are nouns. Consider following example

"This is good phone"

In above sentence, phone (product feature) is noun and good (opinion word) is adjective.

In part-of-speech (POS tagging), each word in review is tagged with its part- of- speech (such as noun, adjective, adverb, verb etc). After POS tagging now it is possible to retrieve nouns as product features and adjectives as opinion words.

There are different freely available POS taggers like Stanford POS Tagger. Fig1 shows how above sentence will be tagged using Stanford POS Tagger.

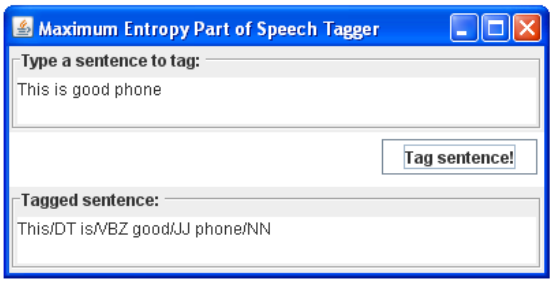

Fig 2: Output of Stanford POS Tagger

In tagged sentence, good is tagged with tag JJ which indicates 'good' is an adjective where a 'phone' is tagged as NN which indicates noun.

C. Feature Extraction

In feature extraction, product features are extracted from each sentence. Product features are generally nouns, so each noun is extracted from sentence.

In review, features may be mentioned explicitly or implicitly by the reviewer. Features which are mentioned in a sentence directly are called as explicit features and features which are not mentioned directly are called implicit features. For example,

"Battery Life of a phone is less"

In this sentence reviewer has mentioned battery life directly so it is explicit feature. It is easy to extract such features. Now consider following sentence,

"This phone needs to charge many times in a day"

In this sentence reviewer is talking about battery of phone but it is not mentioned directly in the sentence. So here battery is implicit feature. It is difficult to understand and extract such features from sentence.

D. Opinion Word Extraction

In opinion word extraction, opinion words are identified. If a sentence contains one or more product features and one or more opinion words, then the sentence is called an opinion sentence. As stated above opinion words are generally adjectives.

E. Opinion Word Polarity Identification

In opinion word polarity identification, semantic orientation of each opinion word is identified. Semantic orientation means identifying whether opinion word is expressing positive opinion, negative opinion or neutral opinion.

F. Opinion Sentence Polarity Identification

Opinion sentence polarity identification predicts the orientation of an opinion sentence. Consider following sentence-

"This is not good phone"

Above sentence contains opinion word 'good' which expresses positive opinion. But sentence expresses negative opinion because of negation word 'not'. Therefore after finding opinion word polarity identification it is necessary to find polarity of opinion sentence.

For opinion sentence polarity identification a list of negation words such as 'no', 'not', 'but' etc. can be prepared and negation rules can be formed. For example, if a sentence contains odd number of negation words then it's polarity will be opposite of polarity opinion word in that sentence. Otherwise sentence will have same polarity as that of polarity of opinion word in it.

G. Summary Generation

Summary generation is generated after opinion sentence orientation identification. This summary is based on features of product. With the help of information discovered in previous steps summary can be generated. Summary can be generated in the form of tables or graphs. A table or graph will give summary of all the reviews related to a product.

III. DIFFERENT APPROACHES FOR FEATURE- BASED OPINION MINING AND SUMMARIZATION

Following are some approaches which are used for opinion word polarity identification.

A. Lexicon-Based Approach

Xiaowen Ding, Bing Liu, Philip S. Yu proposed holistic lexicon-based approach to identifying the orientations of context dependent opinion words .Polarity of opinion word is find out using score function and negation rule. Following is score function-

$$score(f) = \sum_{wi:wi \in s \wedge wi \in V} \left(\frac{wi.SO}{dis(wi,f)} \right)$$

where *wi* is an opinion word, *V* is the set of all opinion words (including idioms) and *s* is the sentence that contains the feature *f*, and *dis*(*wi*, *f*) is the distance between feature *f* and opinion word *wi* in the sentence *s*. *wi.SO* is the semantic orientation of the word *wi*. The multiplicative inverse in the formula is used to give low weights to opinion words that are far away from the feature *f* [6].

Opinion word lexicons can be obtained through bootstrapping process using WordNet. In WordNet, adjectives are organized into bipolar clusters and share the same polarity of their synonyms and opposite polarity of their antonyms [7].

Aspect (or feature) Related Terms (ART) identification can also be done using bootstrapping. In [4], researchers' considers two types of ARTs for study-(1) nouns, adjectives, adverbs and verbs and (2) multiword terms. Researchers have defined a new bootstrapping algorithm, multi-aspect bootstrapping algorithm and a multi-aspect segmentation algorithm.

SenticNet is publicly available affective common sense resource for computing. As target lexicon and source of polarity information for polarity-based concept similarity measure, SenticNet can be used [8].

SenticWordNet 3.0 is also a publicly available lexical resource which is explicitly devised for supporting sentiment classification and opinion mining applications. In [9], proposed automatic annotation process for which consists of two steps-(1) a weak-supervision, semi-supervised learning step and (2) a random-walk step.

The support vector machine (SVM) has performed effectively for classification in the literature. SVM ac be use more effectively in combination with SentiWordNet for sentiment classification [10].

B. Statistical Approach

[11] proposed a supervised information extraction system which extracts features and associated opinions. They used frequency Bayesian classification technique to calculate probability distribution.

PMI (Pointwise Mutual Information) algorithm uses mutual information as measure of the strength of semantic association between two words. PMI-IR uses Pointwise Mutual Information (PMI) and Information Retrieval (IR) to measure the similarity of pairs of words or phrases. The semantic orientation of a given phrase is calculated by comparing its similarity to a positive reference word with its similarity to a negative reference word [12].

C. Intelligent Feature Selection Approach

The IFS approach uses a feature relation network (FRN). FRN utilizes two important syntactic n-gram relations-(1) subsumption and (2) parallel [13]. These two relations occur between two n-gram features categories. IFS can be also combined with larger feature sets for enhanced opinion-classification performance.

IV. PERFORMANCE MEASURES

For performance analysis of opinion word extraction and feature extraction, recall and precision these two measures are used. These are used generally in information retrieval. For opinion sentence polarity identification algorithm, performance is measured using accuracy. Other than these purity and entropy can be used [14].

A. Recall

In information retrieval, recall is the fraction of the documents that are relevant to the query that are successfully retrieved. In opinion word extraction or feature exaction algorithm, recall will be the fraction of the relevant opinion words or features that are relevant to that are successfully retrieved. So it can be calculated using following formula-

$$\text{Recall} = \frac{|\{\text{relevant opinion words/features}\} \cap \{\text{retrieved opinion words/features}\}|}{|\{\text{relevant opinion words/features}\}|}$$

B. Precision

In information retrieval, precision is the fraction of retrieved documents that are relevant to search. Similarly, in opinion word or feature exaction algorithm precision is the fraction of retrieved opinion words/ features that are relevant to search.

$$\text{Precision} = \frac{|\{\text{relevant opinion words features}\} \cap \{\text{retrieved opinion words/features}\}|}{|\{\text{retrieved opinion words/features}\}|}$$

C. Accuracy

For opinion sentence polarity identification algorithm, performance is measured by calculating accuracy.

V. CONCLUSION

Feature based opinion mining and summarization is challenging field for researchers. It is useful for individuals as well as for organization.

There are many approaches for it. But there is no automated technique that mines opinions that are hidden between the lines. In case of context-independent feature based opinion mining less work is done.

For feature based opinion mining and summarization different tools like RapidMiner, WordNet, SentiWordNet, POSTagger, Crawlers and Parsers can be used.

REFERENCES

1. Dr.M S Vijaya, V Pream Sudha, "Research Directions in Social Network Mining with Empirical Study on Opinion Mining", CSI Communication Dec 2013 pp 23-26.
2. Larissa A. de Freitas, Renata Vieira, "Ontology-based Feature Level Opinion Mining for Portuguese Reviews", ACM- 978-1-4503-2038-2/13/05 WWW 2013.
3. Fermin L. Cruz, Jose A. Troyano, Fernando Enriquer, F.Javier Ortega, Carlos G. Vallejo,"A Knowledge-Rich Approach to Feature-Based Opinion Extraction from Product Reviews",ACM 978-1-4503-0386/10/10 SMUC'10,, 2010

4. Jingbo Zhu, Huizhen Wang, Muhua Zhu, Benjamin K. Tsou, Matthew Ma, "Aspect-Based Opinion Polling from Customer Reviews", IEEE Transactions on afferective Computing, VOL 2 No 1, Jan-March 2011
5. Minqing Hu and Bing Liu, "Mining and Summarizing Customer Reviews", KDD'04, August 22–25, 2004,
6. Xiaowen Ding, Bing Liu, Philip S. Yu, "A Holistic Lexicon-Based Approach to Opinion Mining", WSDM'08, 2008
7. M. Hu and B. Liu. Mining and summarizing customer reviews. KDD'04, 2004.
8. Soujanya Poria, Alexander Gelbukh, Amir Hussain, Dipankar Das, Sivaji Bandyopadhyay," Enhanced SenticNet with Affective Labels for Concept-based Opinion Mining",IEEE 2013
9. S. Baccianella, A. Esuli, and F. Sebastiani, "SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining," Proc. Int'l Conf. on Language Resources and Evaluation, 2010, pp. 2200-2204.
10. Chihli Hung, Hao-Kai Lin," Using Objective Words in SentiWordNet to Improve Sentiment Classification for Word of Mouth", IEEE 2013
11. N.Anwar, A. Rashid, S.Hassan,"Feature Based Opinion Mining of Online Free Format Customer Reviews Using Frequency Distribution and Bayesian Statistics", IEEE.
12. P. Turney," Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. "In Proc. of the Meeting of the Association for Computational Linguistics (ACL'02),2002.
13. Ahmed Abbasi, P"Intelligent Feature Selection for opinion classification", IEEE 2010.
14. Zhongwu Zhai, Bing Liu, Jingyuan Wang, Hua Xu, and Peifa Jia, "Product Feature Grouping for Opinion Mining", IEEE 2012.
15. Erik Cambria, Björn Schuller,Yunqing Xia, Catherine Havasi," New Avenues in Opinion Mining and Sentiment Analysis", IEEE 2013
16. Priti S. sajja, "Feature-Based Opinion Mining",IJDMET 2011.
17. Fermín L. Cruz, José A. Troyano, Fernando Enríquez, F. Javier Ortega, Carlos G. Vallejo," A Knowledge-Rich Approach to Feature-Based Opinion Extraction from Product Reviews", SMUC'10, 2010
18. Bing Liu, "Web Data Mining Exploring Hyperlinks, Contents, and Usage Data", Springer, 2007.
19. Dongjoo Lee, Ok-Ran Jeong, Sang-goo Lee ,"Opinion Mining of Customer Feedback Data on the Web", 2007
20. Bing Liu, Xiaowen Ding, "Systems and Methods for Opinion Mining, US Patent Application Publication" , *19 Feb.2009.*
21. Xiaowen Ding and Bing Liu, "The Utility of Linguistic Rules in Opinion Mining", *SIGIR'07, July 23–27, 2007 Amsterdam, Netherlands.*
22. Anindya Ghose, Panagiotis G. Ipeirotis, Arun Sundararajan, "Opinion Mining Using Econometrics: A Case Study on Reputation Systems", *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, 2007.*
23. K. Dave, S. Lawrence, and D. Pennock. Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews. *In Proc. of the 12th Intl. World Wide Web Conference (WWW'03)*