

Building Recommendation System for Hotel

Mohammad Aamir
PG Student
AKG Engineering College
Adhyatmik Nagar, GZB (UP) India

Mamta Bhusry
Professor
AKG Engineering College
Adhyatmik Nagar, GZB, (UP) India

ABSTRACT

As we move into the third decade of the World Wide Web (WWW), there has been a vast change in the availability of online information. Discovering information has never been more mechanized as of now, just a mouse click away. The objective of Opinion Mining can be achieved by executing a cluster of search results based on the features and quality for a given item. For rating the product and providing opinions, examination of customer evaluation is most significant-which is a challenging problem. Thus in the above context this paper attempts to discuss about the techniques and tools used by the opinion mining.

Keywords

Opinion mining, Opinion Retrieval, Opinion Classification, Opinion Summarization

1. INTRODUCTION

Information in the form of text can be roughly described into two categories, facts and opinions. Objective statements are called as Facts that describes about events and entities, whereas subjective statements are called as Opinions [1] which reflects an individual's perception regarding the events and entities. Processing the text information [2] for the retrieval of factual information for this a vast amount of research has been carried out for e.g. web search, text mining and various natural language processing tasks. For processing of opinions minimal amount of work has been carried out till now, whereas opinions play a vital role; whenever one needs to make a decision one wants to hear opinion of others [3]. This fact is true for both individuals and also for organizations. Before the evolution of web if one wanted to make a selection, one wanted to know the opinions from families and friends. Surveys and focused groups were conducted by the organization regarding their products and services whenever they need to find the opinion of the general public. Nowadays the world has been changing with the increasing growth of the user generated content on the web [4]. User generated content is a mixture of internet forums, discussion groups and blogs- where one can post reviews and opinions on almost anything. Now it is no longer necessary to enquire about any product regarding its reviews and opinions from friends and families as the availability of web allows us to take the reviews and opinions of other users of that particular product [5]. Similarly for organizations it is no longer necessary to organize focused groups and conducting surveys in order to become aware about the consumer opinions or reviews related to its product [6]. It is very arduous for a user to quest for a reliable source on order to extract proper sentences, and then read them, summarize them and categorize them into usable forms this whole process requires an automated opinion mining and summarization system [7].

In the area of opinion mining research has been started by identifying opinion words for example amazing, great, bad wonderful and many more. Opinion words have their own semantics (i.e., positive or negative) [8,9] the research has been carried out for identifying the semantic orientation of the opinion words. From a colossal collection of data the researchers have pointed out various linguistic rules for identifying opinion words from the data set. With the help of Word Net-a lexical database, which can group English words into sets of synonyms called synsets [10] one can find the synonyms and antonyms of a given opinion word. Word Net is a mixture of thesaurus and dictionary. At Document Level Sentiment classification [11] is another crucial development in terms of product reviews. The aim is to examine that weather a document is indicating a positive or negative opinion about a product [12]. A new area of research is sentence level sentiment classification which classifies every single expression as positive or negative opinion. A new prototype of feature-based opinion mining and summarization has been put forward which provides a more unabridged formulation of the opinion mining problem [13]. It shows that how from unstructured texts structured opinion summary can be produced by identifying and mining the key pieces of information [14].

Many of these techniques require sentiment analysis or feature scoring methods. However, existing sentiment analysis methods are limited when determining the sentiment polarity of context-sensitive words, and existing feature scoring methods are limited when only the overall user score is used to evaluate individual product features. Proposed summarization approach, uses context-sensitive information which is used to determine sentiment polarity while opinioned-feature frequency is used to determine feature scores. Based on experiments with actual review data, this method improves the accuracy of the calculated feature scores and improves the existing methods.

2. OPINION MINING OVERVIEW AND CONCEPTS

Sentiment classification is a major problem in this area where the evaluations of a target object for example a book, can be labeled as either positive or negative. Then there are two ways in which evaluation of opinion can be done [15].

Direct opinion

are easier to work with and directly gives the opinion either positive or negative. For example, "This car has poor mileage"

Comparison

are more insightful. For example "The Toyota Corolla is not as good as Honda Civic" shows a comparison

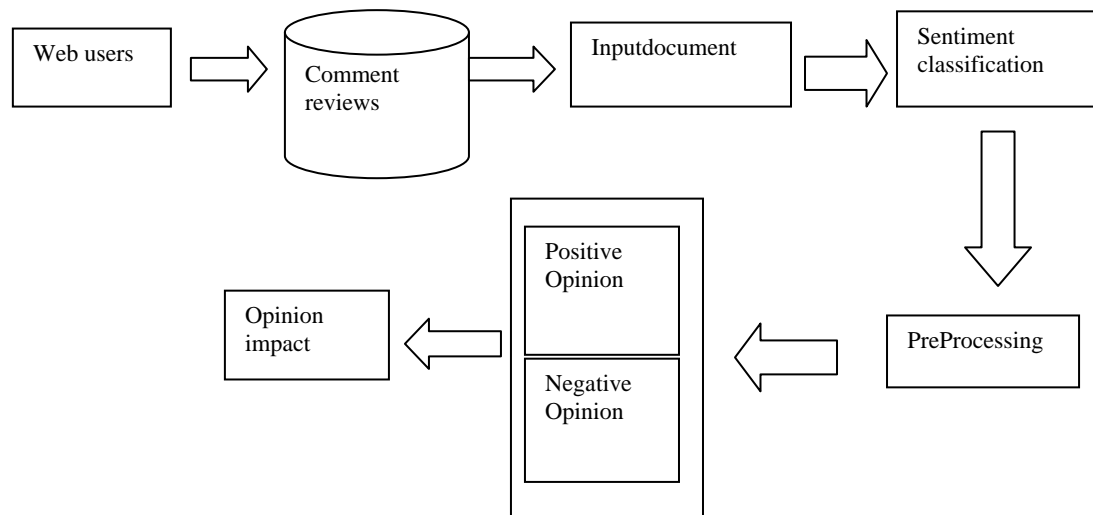


Figure1 Work Flow of Opinion Mining

Figure shows the workflow of opinion mining that how the opinions are being extracted from the reviews made by the users and posted in the form of comments.

2.1 Architecture Of Opinion Mining

The process of finding opinion of people towards a product or a topic is called opinion mining. Opinion mining is a conclusion of what people think about a product, topic or event let it be positive, negative, or neutral. There are three main steps involved in the process of opinion mining and summarization [16]. These are as follows

(1) Opinion Retrieval

Collection of review text from different review websites is the called as opinion retrieval. Various websites maintains a database of reviews for various products [17]. Web crawler an Information retrieval techniques were used to gather the review data from different sources and accumulate them in database.

(2) Opinion Classification

Classification of review text is the first step in sentiment analysis. The process is to categorize every single d_i in D , where D is a review document $D = \{d_1 \dots d_n\}$ with a label expressed from C , where C is a predefined set $C = \{\text{positive, negative}\}$. Classifying review text into positive and negative forms is the main aspect of this approach. Two types of popular approaches are there for this classification (1) Machine Learning and (2) Lexicon Based Approach [18, 19].

(3) Opinion Summarization

In opinion mining summarization of opinion plays a major part in the mining process. A large amount of work has been done till now on summarization of product reviews. Features or Subtopics that are described in reviews and summary of reviews should be provided based on those features and subtopics. There are mainly two types of approaches for opinion summarization process [20].

- (1) Feature based Summarization: this involves searching out for frequent features that are coming into sight in many reviews. Then a summary is produced by opting out sentences that carries distinctive feature information [21].
- (2) Latent Semantic Analysis Method: this method is used for identifying the feature present in review text. *Term Frequency* is defined as a tally of term

occurrences in a particular document. Higher frequency term means that term is to a greater extent important for summary presentation[22].

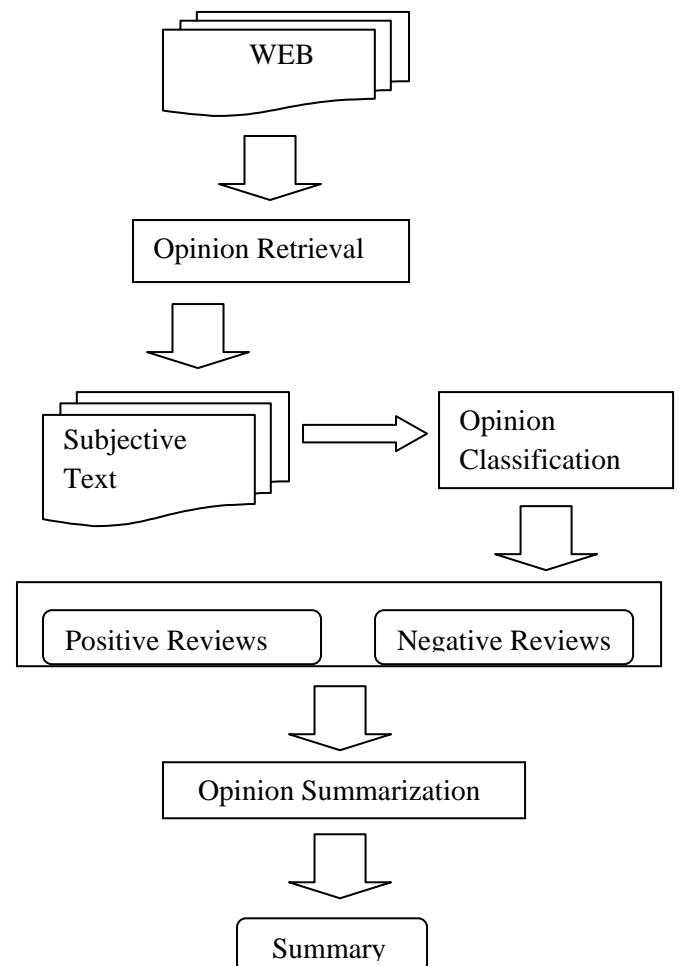


Figure2 Architecture of Opinion Mining

Figure shows the architecture of opinion mining [23] which shows how the input is being classified in different steps in order to summarize the reviews.

3. IMPLEMENTATION AND AREA OF APPLICATIONS

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.

3.1 Preparing Review Database

Reviews are extracted from different websites and then one can store those reviews into storing into review database. Each website has its own structure. After this preprocessing is done where unwanted text (other than product reviews) is removed and then reviews are stored into database.

3.2 Part-of-Speech Tagging (POS Tagging)

The aim of 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 hotel”

In above sentence, hotel (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.

3.3 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. Explicit Features are those which are mentioned directly in a sentence and implicit features are those which are not mentioned directly. For example,

“The hotel has a limited Wi-Fi connection”

In the above sentence Wi-Fi connection has been mentioned directly by the reviewer hence it is a explicit feature. Such features can be extracted easily.

Now consider following sentence,

“The hotel was put on generator many times in a day”

In the above sentence the electricity is mentioned by the reviewer but not directly in the sentence. Hence the implicit feature here is electricity.

Extracting such features from a sentence and understanding them is a difficult task.

3.4 Opinion Word Extraction

Opinion words are identified in opinion word extraction. A sentence is called as opinion sentence if it contains one or more opinion words and one or more product features. Generally opinion words are adjectives.

3.5 Opinion Word Polarity Identification

Semantic orientation is identified of each opinion word in opinion word polarity identification. Identifying whether a opinion word can be expressed as positive opinion, negative opinion or neutral opinion is called as semantic orientation.

3.6 Opinion Sentence Polarity Identification

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

“This is not a good hotel”

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 its 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.

3.7 Summary Generation

Summary 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.

4. EXPERIMENTAL RESULTS USING ACTUAL REVIEW DATA

The results indicate that the sentiment polarity and opinion strength weighting adjustments can effectively calculate a user’s feature evaluation.

In most previous work, the distribution of document-level user scores was used to calculate feature scores. Thus, in order to calculate feature scores more accurately, feature-level evaluations were necessary. To handle context-sensitive words, one considers the contexts of the opinion words. In addition, a method has been used that utilizes user scores, frequency of opinioned-features, and sentiment polarities of opinion words to increase the accuracy of the feature scores derived during the feature scoring step. The summary provides a set of calculated scores for each product feature

Score obtained by the hotels based on all the parameters

- (1) Staff
- (2) Room
- (3) Service
- (4) Food
- (5) Breakfast
- (6) Locale
- (7) Pool
- (8) Buffet
- (9) Price
- (10) Lounge

Table1: Review Summary of the Hotel.

S.NO	Hotel Name	Score Obtained
1	Hotel City Star	471.81
2	J W Marriott Hotel New Delhi Aero city	466.776

3	The Metropolitan Hotel and Spa New Delhi	421.314
4	Crowne Plaza New Delhi Rohini	416.448
5	Kempinski Ambience Delhi	404.04
6	The Suryaa New Delhi	403.09
7	Red Fox Hotel Delhi Airport	396.128
8	Le Meridien New Delhi	396.124
9	Welcome Hotel New Delhi	395.58
10	The Imperial Hotel	389.122
11	Dusit Devarana New Delhi	383.28
12	Hotel Delhi Aerocity	378.38
13	Holiday Inn New Delhi International Airport	362.84
14	The Leela Palace New Delhi	359.84
15	Shangri-La's Eros Hotel	337.394
16	Bloomrooms @ Link Rd	327.698
17	ITC Maurya New Delhi	321.556
18	Hyatt Regency Delhi	316.34
19	Radisson Blu Hotel New Delhi Paschim Vihar	301.14
20	Eros Hotel New Delhi Nehru Place	298.68
21	Hotel Ajanta	295.17
22	Hotel Hari Piorco	286.1
23	Radisson Blu Hotel New Delhi Dwarka	282.2
24	Hotel Toronto	270.758
25	The Lalit New Delhi	264.2
26	Piccaddily Hotel New Delhi	255.6
27	Lemon Tree Premier, Delhi Airport	241.36
28	Taj Palace Hotel	239.76
29	Country Inn By Carlson Saket New Delhi	228.9
30	Bloomrooms @ New Delhi Railway Station	109.3

Table 2: Accuracy of the Algorithm.

ACTUAL FEATURES	940
EXTRACTED FEATURES	877
MATCHED FEATURES	256
RECALL	93.2978723404%
ACCURACY	29.1904218928%

5. FUTURE SCOPE AND CONCLUSION

The results indicate that the sentiment polarity and opinion strength weighting adjustments can effectively calculate a user's feature evaluation. In this approach, features with positive opinions in a low-scored review and features with negative opinions in a high-scored review are scored according to the opinion for that feature, the errors that may occur when all feature scores are inherited from the overall user score are decreased and precision increases. Hence it Helps hotels to review their facilities and services in view of the customers' demand and expectations.

6. REFERENCES

- [1] R. Ali, " Pro-Mining: Product recommendation using web based opinion mining," IJCET, vol. 4, Nos. 6, pp-299-313,2013
- [2] Bing Liu, Sentiment Analysis and Opinion Mining, San Francisco: Morgan & Claypool Publishers, 2012.
- [3] J. Siddiqui, "An Overview of opinion mining," IJARET, vol. 4, Nos. 7, pp-176-182,
- [4] S. S. Sohail, J. Siddiqui and R. Ali," Bookrecommendation system using web mining technique," Proceeding of International Conference on Advances in Computing, Communications and Informatics, pp-1609-1614
- [5] S. Kim and E. Hovy,"Determining the sentiment of opinions," COLING, pp- 1367-1374, 2004
- [6] V. Hatzivassiloglou and K. McKeown, "Predicting the semantic orientation of adjectives" EACL, pp- 174-181, 1997.
- [7] M. Eirinaki, S. Pisal, and J. Singh, "Feature-based opinion mining and ranking," journal of Computer and system Sciencesvol. 78, pp-1175-1184, 2012.
- [8] Bo Pang and Lillian Lee,"Opinion Mining and Sentiment Analysis", Foundations and TrendsR_in Information Retrieval Vol. 2, Nos. 1-2 (2008) 1–135
- [9] Lun-Wei Ku, Yu-Ting Liang and Hsin-Hsi Chen, "Opinion Extraction, Summarization and Tracking in News and Blog Corpora", 2006 American Association for Artificial Intelligence
- [10] Chaovalit,Lina Zhou, Movie Review Mining: a Comparison between Supervised and Unsupervised Classification Approaches, Proceedings of the 38th Hawaii International Conference on System Sciences – 2005
- [11] Christopher Scaffidi, Kevin Bierhoff, Eric Chang, Mikhael Felker, Herman Ng and Chun Jin, "Red Opal: product- feature scoring from reviews", Proceedings of 8th ACM Conference on Electronic Commerce, pp. 182-191, New York, 2007.
- [12] Kaiquan Xu , Stephen Shaoyi Liao ,Jiexun Li, Yuxia Song, "Mining comparative opinions from customer reviews for Competitive Intelligence", Decision Support Systems 50 (2011) 743–754.
- [13] Chen Mosha,"Combining Dependency Parsing with Shallow Semantic Analysis forChinese Opinion-Element Relation Identification", IEEE, 2010, pp.299-305.

- [14] Yuanbin Wu, Qi Zhang, Xuanjing Huang, Lide Wu, "Phrase Dependency Parsing for Opinion Mining", EMNLP '09 Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, 2009, Volume 3
- [15] Qi Zhang, Yuanbin Wu, Tao Li, Mitsunori Ogihara, Joseph Johnson, Xuanjing Huang, "Mining Product Reviews Based on Shallow Dependency Parsing", SIGIR '09, Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval, 2009
- [16] Himabindu Lakkaraju, Chiranjib Bhattacharyya, Indrajit Bhattacharya and Srujana Merugu, "Exploiting Coherence for the simultaneous discovery of latent facets and associated sentiments", SIAM International Conference on Data Mining (SDM), April 2011
- [17] Minqing Hu and Bing Liu, "Mining and summarizing customer reviews", KDD '04: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, 2004
- [18] T. L. Griffiths, M. Steyvers, D. M. Blei and J. B. Tenenbaum, "Integrating Topics and Syntax", Advances in Neural Information Processing Systems 17, 2005.
- [19] D. M. Blei, M. Jordan and A. Ng, "Latent Dirichlet Allocation", Journal of Machine Learning and Research, page 993-1022, 2003.
- [20] Ido Guy, Naama Zwerdling, David Carmel, Inbal Ronen, Erel Uziel, Sivan Yogev, Shila Ofek-Kolfman, (2009) Personalized Recommendation of Social Software Items Based on Social Relations, ACM.
- [21] Li S, Zhou G, Wang Z, Lee SYM, Wang R (2011) Imbalanced sentiment classification. In: Proceedings of CIKM-2011, pp 2469–2472.
- [22] Wan X (2011) Bilingual co-training for sentiment classification of Chinese product reviews. Computational Linguistics 37 (3):587-616
- [23] Ronak Talati, Sudeep Shrivastava, Sayali Khade, " Issue Analysis and Summarization using Opinion Mining" in International Journal of Advanced Research in Computer Science and Software Engineering, volume 4, issue 12, 2014.