

OPINION MINING IN R

Ayushi Gupta

*Assistant Professor, Ajay Kumar Garg Engineering College, Ghaziabad, U.P. India
 guptaayushi@akgec.ac.in*

Abstract— In present era of computers and social networking we are available with a bulk of digital content related to what people believe, feel or think about certain products, social and economic issues, governmental policies etc. in form of comments, discussions, reviews, tweets on social sites, discussion forums, blogs on almost each and every topic. This plenty of information can be very well utilized in the process of decision making not only in organizations but also for humans as we humans are always concerned about what others think of a particular thing before taking decision. This is what opinion mining does using automated analysis techniques by extracting people’s opinion and classifying them in certain categories of sentiment such as joy, sadness, anger, disgust, fear etc. as manual analysis of unstructured data present in today’s Internet ecosystem is not feasible. This paper represents the basic concepts of opinion mining under following sections covered :- (I) Introduction, (II) Importance of opinion mining, (III) Opinion mining model, (IV) Opinion mining in real world, (V) Opinion mining in real world, (VI) Opinion mining and R.

Keywords— Opinion mining; Opinion; Sentiments; Reviews; Decision making;

I. INTRODUCTION

The field of study that analyses people’s opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes is known as Opinion mining (also known as Sentiment Analysis). It represents a large problem space. The terms opinion mining and sentiment analysis are synonymous. It encompasses all the terms like opinion extraction, affect analysis, sentiment mining, subjectivity analysis, emotional analysis and review mining etc.

Opinion mining is the hottest research area in present scenario. Before the advent of world-wide-web, companies used to conduct surveys, questionnaire, employ consultants for collecting reviews from people about their products, policies. Similarly people used to ask their peers for their suggestion before buying a particular product, which restaurant is good and all but opinion mining overtook all this cumbersome process with automated decision making. The plethora of

review based or opinion oriented posts available on web forms the building block of opinion mining.

II. IMPORTANCE OF OPINION MINING

Opinion mining has gained huge popularity in recent years because of its utmost importance in not only to customers decision making but of an organization as well. Opinion Following are the components of opinion mining model used: (Chinsha TC, 2015)

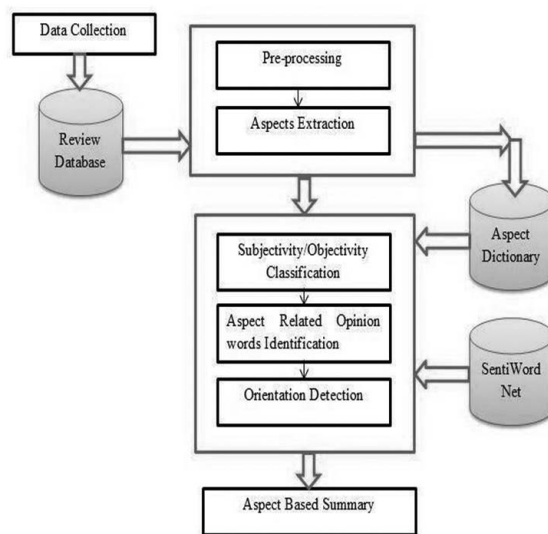


Fig 1: Proposed Opinion Mining Model (Chinsha TC, 2015)

A. Data Collection

The very first and important step is to collect data using which opinions are to be extracted for opinion mining. As plenty of unstructured data containing opinions is available online. It serves as an input to opinion mining system on being converted to structured data. We need to collect that data from which useful information is to be extracted. A review crawler parses the HTML page and extracts the users opinions from it. The data collected is stored in database and fed as input to aspect based opinion mining system for further processing.

B. Pre-processing

The pre-processing of reviews is a technique to remove unnecessary characters from the reviews. This technique improves the opinion mining process accuracy and avoids any unnecessary processing overhead.

C. *Extracting Aspects*

Aspects are the different features or characteristics of the product about which we are concerned. For example, display, picture quality and android version are different aspects of a smart phone. Aspects are important as they are features to which the reviewers give rating to. An aspect may be formed of a phrase or single word. Generally nouns and noun phrases are the aspects. Therefore nouns and noun phrases need to be searched in the reviews. For extracting the important aspects, first find the Part-Of-Speech tag of words in each sentence of a review then identify the nouns and noun phrases in it. For example, tags like NN (noun), NNS (noun plural), NNP (proper noun, singular) etc. of word need to be identified. Then find its frequency of occurrence, the words occurring more frequently are chosen and infrequently occurring ones are discarded. After this, the synonym words are grouped and an aspect dictionary of important aspects is created. Any available POS tagger can be used for tagging purpose.

D. *Subjectivity/Objectivity Classification*

As all the sentences in reviews do not express an opinion. A sentence of the review is helpful for analyses only when it contains an opinion. Such subjective sentences should be identified and other (objective) sentences should be discarded in order to avoid any further processing overhead. Subjective sentences are identified by looking presence of aspects and opinion words. SentiWordNet containing opinion words, and aspect dictionary, containing important aspects is used for this purpose.

E. *Identifying aspect related opinion words*

Aspect related opinion words are identified by the aspect related opinion words identifier. It takes reviews from the database to analyse each review sentence and look up for it in the aspect dictionary. Corresponding opinion words should be identified if any aspects are present in sentence. The POS information of a word like adjective, adverb, noun and verb are used for identifying the opinion words in a sentence. If any aspects are present in a sentence, then extract adjective, adverb, noun and verb by searching 5-gram forwards and backwards from the aspect position.

F. *Orientation Detection*

Orientation Detection identifies the polarity or orientation i.e. positive and negative of an opinion on a certain aspect.

Senti Word Net, which is a dictionary of aspect words, is used for this purpose of assigning polarity to opinion words. Each term in wordnet has been assigned a positive or negative Polarity. If there are words denoting negation like not, never, neither etc, the polarity is reversed.

For example, in the sentence “I did not like this restaurant at all” has positive opinion word like but the polarity is reversed

to negative due to presence of not. So such words need to be considered before calculating sentence score.

G. *Aspect based Summary*

Aspect based summary integrates the scores of each aspect collected from all reviews and generate an aspect based summary of it. Separate aggregation of positive and negative score of aspect is done. This serves as an input to create a sentiment profile of each entity. Visualization tools are an aid to this purpose for visualization of negative and positive polarity by a user.

IV. OPINION MINING IN REALWORLD

Opinion mining is not just used for checking the reviews of products but is implemented for a variety of applications ranging from simple like restaurant reviews to more complex ones like reforming the governmental policies based on people's opinion.

- **Online Review System**

The simplest and the most important application area of opinion mining which is also the most used purpose of opinion mining by people is writing about their opinion, review, feelings, beliefs, attitude towards an entity on web and reading the same by others for their decisionmaking.

- **Opinion mining in market analysis**

Organizations can check user opinions and use them to find current market trends and customer requirements without conducting surveys etc. For example, a laptop manufactured by a company is not able to bring profit to company as it is not successful in the market. A company would definitely like to check reasons behind its failure despite of its good specifications. It can be done by surfing the web for collecting users view regarding the laptop for analyzing what customer actually expected from that product and what can be done for its betterment.

- **Suggestion Systems**

Under suggestion systems, classification of what should be recommended and what not to the user can be done on the basis of positive and negative opinion.

- **Spam Identification**

The capability of opinion mining to differentiate between spam content put by some users with wrong intentions and valid content is an important application area of opinion mining.

- **Opinion mining in government policy making and voting**

People discuss online about the new policy formed by the government. This helps the government to keep track of what their countrymen think about the policies and reform them if

possible. Opinion mining can also predict the chances of winning of a particular candidate by analyzing the voting related tweets trending during voting time.

V. CHALLENGES IN OPINIONMINING

1. The authenticity of reviews cannot be judged on the internet because of irresponsible nature of individuals writing fake reviews to mislead, confuse others out of their grudges or else.
2. Plenty of effort and time needed to train machine by making them learn the training data so that they are able to evaluate future data to be analyzed.
3. The usage of short forms, jargons, slangs and certain linguistic also pose a problem during sentiment analysis.
4. The software to be used for opinion mining are very expensive to be affordable by only companies and large organizations.
5. It makes it difficult for opinion mining to analyze sarcasm or irony used in the reviews. For example, sarcastic comments are very much used in case of political discussions and movie reviews.
6. Sometimes the same sentiment word may show ambiguity in which context it is used. For example, the words *ckss* shows negativity in sentence “the new whatsapp update sucks” while it shows positivity in “this vacuum cleaner has really good sucking quality.”

VI. OPINION MINING AND R

R is a free and open source software for statistical computing and provides a wide variety of statistical models. It’s a programming language. R can be effectively used to perform opinion mining on the dataset collected from online reviews and generate the wordcloud which depicts the sentiment involved in during reviews giving by people.

Sentiment analysis in R can be done using package **sentiment** given by Timothy Jurka. It contains two convenient functions to serve the purpose: **classify_emotion** and **classify_polarity**. R also requires us to load certain packages without which opinion mining is not possible in R and then there is a piece of code to generate certain plots and wordcloud which aid to the visualization of useful opinions extracted giving people’s reviews on certain things.

VII. CONCLUSION

With the advent of technology and ease of access of social sites, the number of user generated comments, reviews, tweets is increasing day by day which is utilized by the opinion mining technology for automating decision making process by analyzing this plethora of content available in digital world. Opinion mining is not just limited to checking polarity of an entity for its positive or negative aspects but has much broader application areas as discussed above. However, it has certain limitations which need to be overcome.

REFERENCES

- [1] A G Dongre, SushmitDharurkar, SwanandNagarkar, Rahul Shukla, VivekPandita, “A Survey on Aspect Based Opinion Mining from Product Reviews,” International Journal of Innovative Research in Science, Engineering and Technology (A High Impact Factor, Monthly Peer Reviewed Journal), Vol. 5, Issue 2, pp. 1415-1418 February 2016.
- [2] EfthymiosKouloumpis, TheresaWilson, Johanna Moore, “Twitter Sentiment Analysis: The Good the Bad and the OMG!,” Proceedings of the Fifth International AAAI Conference on Weblogs and SocialMedia.
- [3] KiruthikaM;Sanjana;PriyankaGiri.(Apr.2016).”Sentiment Analysis of Twitter Data”.6[4], ISSN:2319-1058, IJJET,pp.264-273.
- [4] Kouloumpis;Efthymios;Wilson;Theresa;Moore;Johanna D. “Twitter Sentiment Analysis:The Good the Bad and the OMG!”. Icwsm 11,164, Proc. Of 5thInt. AAAI conference. pp(538-541),
- [5] Shibily Joseph; Chinsha T C .”Aspect based Opinion Mining from Restaurant Reviews”. International Journal of Computer Applications (0975 – 8887) ,Advanced Computing and Communication Techniques for High Performance Applications(ICACCTHPA-2014)
- [6] Bing Liu. “*Sentiment Analysis and Opinion Mining*”, Morgan & Claypool Publi

ABOUT THE AUTHOR



Ms Ayushi Gupta is working as an Assistant Professor in Ajay Kumar Garg Engineering College, Ghaziabad. Her area of Research is Data Mining, NLP and ML.