International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 – 5161

Available at http://inpressco.com/category/ijcet

Research Article

A Survey on Sentiment Analysis and Opinion mining

Mahima Gaur and Jyotika Pruthi*

The NorthCap University, Gurgaon, India

Accepted 17 March 2017, Available online 20 March 2017, **Vol.7, No.2 (April 2017)**

Abstract

For a business to run effectively it is helpful to know the opinion viewpoints or reviews of the consumers and make changes in the strategies and services accordingly. Similarly for the costumers it is very useful to know about the quality of the products and services beforehand. Opinion mining is a method of distilling the information and studying the emotions associated with a particular review and therefore finding out the polarity of the review. This paper provides a comprehensive overview of sentiment analysis and techniques and methods to achieve it.

Keywords: *Sentiment Analysis, Opinion mining etc.*

Introduction

Humans are subjective creatures and opinions are important therefore sentiment analysis aims at building a system which analyses the mood of an individual about a particular product, topic or event expressed in a text span made in a review, blog post, comment or tweet.

Mining opinions in a user generated content is very challenging but practically very useful. Sentiment analysis can be done on three levels: document level, sentence level or attribute level. Document level sentiment analysis works by reducing the whole document to a single opinion. But most of the time a document does not represent a single opinion. The same document might contain different contradicting opinions about the same entity. Sentiment analysis on sentence level classifies the sentiment expressed in each sentence. The first task is to classify whether the sentence is subjective or objective.

Objective: I went to the new DSW store to shop yesterday.

Subjective: they have a beautiful collection of shoes.
The opinion holder can give different reviews on different aspects of an entity.

The food is good but very expensive. Helpings are small. The seating is good though.

Aspect level sentiment analysis aims at classifying opinions for different aspects of the same entity. The first task here is to identify the different aspects of the entity and then the opinion on those aspects.

*Corresponding author: **Jyotika Pruthi**

The main data sources for opinion mining are review sites, micro-blogging and social network. Opinion of people on different debatable issues like a political party or spots or an opinion about an event or incident also makes a source of data.

Applications

Business and organisations

For the establishment and smooth functioning of a business it is very important to find out what the users think about their services for their product and service benchmarking. User's opinion is the basic criterion for the improvement of the quality of products and enhancement of the services. Companies therefore spend a huge amount of money to find out costumer sentiment and opinions

Individuals

For any costumer making a purchase knowing the opinion of other users can play an influential role in his decision making. Internet provides a large number of user generated reviews of numerous products. E-commerce websites such as amazon.com (product reviews), zomato.com (restaurant reviews), reviewcentre.com have millions of reviews by costumers.

Detection of Hatred

It is possible to scan the social media and blog posts to detect the arrogant words to find out the disagreement or anger over a particular issue. Monitoring the emails or tweets can mine the hatred language used which can prove to be harmful.

Detection of Spam

With internet being as easily available as it can it allows anyone to put anything on the web. This therefore increases the possibility of the spam content on the internet. With opinion mining and sentiment analysis it is possible to classify the content into spam or not-spam.

Challenges

Irrelevant Objects

First task in opinion mining is to identify the objects on which the opinion is expressed. This becomes a challenge since a sentence might contain objects which are irrelevant to us. Therefore it is a difficult task to figure out which objects are relevant and which are not.

Identifying Features

It's a challenge to identify the features on which the opinion is made. Usually the noun in the sentence is considered to be the feature and it is comparatively easier to identify but some sentence also have the verb as the feature which makes the task to identify the feature difficult.

Different Words for Same Feature

A lot of times the same feature of a product is addressed by using different words. Like both voice and sound quality refer to the same feature of a TV. In such cases it is not enough to simply pick out the feature from the sentence. Grouping synonyms to make the analysis more accurate is necessary.

Context Dependency

The same sentence can be considered positive in one situation and negative in the other. The remark made my humans are mostly context dependent which are very easy to be understood by humans but very hard to be understood by the machine in order to assign a polarity to it.

SARCASM

Humans make sarcastic comments and remarks in their reviews and it sometimes becomes very difficult for other humans beings to understand sarcasm. It's even harder to build a machine system which understands can differentiate the sarcastic comments from the non-sarcastic ones.

Machine Learning Approaches

Machine learning is a subfield of artificial intelligence that aims at making the computer able to work without being programed explicitly. Machine learning look through data to search for patterns and uses them to improve the programs understanding so that the program can grow and change according to the data it is exposed to various machine learning techniques tried to classify reviews:

- Naïve Bayes classifier
- Maximum entropy classifier
- Support vector machine

Lexicon based Approaches

Lexicon based approach assumes the semantic orientation of the entire text is equal to the sum of individual semantic orientation of words and phrases. Lexicon approach uses a dictionary of words along with their respective semantic orientation. The dictionary can be created manually or automatically using seed words. Usually the adjective words are used to identify he semantic orientation of the text. First all the adjective words are extracted from the text and their respective semantic orientation values are extracted. These values are then used to calculate the overall polarity of the text.

Statistical Approaches

Statistical analysis basically involves collecting and inspecting data samples to observe trends or patterns. Statistical model considers the data to be a combination of aspects and ratings and the motive is to extract the aspects from the data and to identify and classify the sentiments into ratings. A word occurring more frequently in positive texts is considered to have a positive polarity whereas a word whose occurrence is more frequent in negative texts is considered negative. A word with similar frequency of occurrence is considered neutral.

Supervised and Unsupervised Learning

Supervised Learning

In supervised learning the machine is trained using output datasets so as to receive the desired result. In supervised learning the categories the data is assigned to is known before computation. The model basically defines the effect one set of observations have on another set of observations.

Unsupervised Learning

In unsupervised learning the data is clustered in classes and algorithm is made to differentiate. The model is not provided with the correct results during the training. Latent variables are the cause of all observations.

Since supervised learning tries to find a correlation between two sets of data it is more convenient to employ unsupervised learning for models with deep

hierarchies. With unsupervised learning time increases linearly in number of levels of model hierarchy since learning can proceed from one level to another and only one step is required at a hierarchy level.

Conclusion

Opinion mining and sentiment analysis have wide range of applications. Also there are many challenges to researches focusing in this field. This therefore has been a very active research area in recent years. It is not possible to consider a classification model better than the other. No model consistently outperforms the other. Different classification methods can be used together to overcome each other's drawbacks. Combining different algorithms in an efficient way enhances sentiment classification performance.

A lot of challenges however continue to exist in this field of research and we have not yet developed a fully automated system for sentiment analysis. Although the techniques are advancing very fast, further more work is required to be done in future to overcome the drawbacks of existing models.

References

G.Vinodhini, RM.Chandrasekaran (June 2012) Sentiment Analysis and Opinion Mining: A Survey, Volume 2 Issue 6, www.ijarcsse.com.

Neesha Jebaseeli, E.Keerubakaran (2012), A Survey on Sentiment Analysis of (Product) Reviews, Volume 47, International Journal of Computer Applications (0975-888)

Khan, Aamera ZH, Mohammad Atique, and V. M. Thakare (2015), Combining lexicon-based and learning-based methods for Twitter sentiment analysis. International Journal of Electronics, Communication and Soft Computing Science & Engineering (IJECSCSE): 89.

Modha, Jalaj S., Gayatri S. Pandi, and Sandip J. Modha (2013), Automatic sentiment analysis for unstructured data. International Journal of Advanced Research in Computer Science and Software Engineering 3.12: 91-97.

Revathy, K., and B. Sathiyabham (2013), A hybrid approach for supervised twitter sentiment classification. International Journal of Computer Science and Business Informatics 7

Go, Alec, Richa Bhayani, and Lei Huang (2009),Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford 1: 12.

H Saif, M Fernández, Y He, H Alani (2014),On stopwords, filtering and data sparsity for sentiment analysis of Twitter.: 810-817.

Tripathy, Abinash, Ankit Agrawal, and Santanu Kumar Rath (2016), Classification of sentiment reviews using n-gram machine learning approach. Expert Systems with Applications 57: 117-126.