Journal of Engineering and Applied Sciences 12 (Special Issue 2): 6157-6159, 2017
ISSN: 1816-949X

Lexicon Based Sentiment Analysis on Facebook Page

[1]Mohammad Arif and [2]Gourav Bathla
[1]Department of M.E Bigdata and Analytics, Chandigarh University, Ajitgarh, Mohali, India
[2]Department of CSE, Chandigarh University, Ajitgarh, Mohali, India

Abstract: Sentiment analysis, also known as opinion mining is among the operations of NLP. Due to gradual increase in data like comments, reviews, blogs the direction of research is going or focusing over sentiment analysis. Sentiment analysis has grown interest of researchers of today's age, sinceits application are growing tremendously. Sentiment analysis is the calculation of conception, sentiments and subjectivity of text. Internet is one of the biggest source from where we can collect our data for analysis. The social networking sites like facebook and twitter are generating huge amount of data filled with sentiments, opinions, comments of people about their friends, situation, products. The sentiment behind every comment, review, blog can be analyzed by opinion mining. I will apply NLP to extract subjective information from text. I have chosen a facebook page of Kashmir named as Wazwan the great Kashmiri cuisine. Wazwan is the traditional cooking of the Kashmir.

Key words: Lexicon based method, sentiment analysis, social media, traditional cooking, facebook, Kashmiri

INTRODUCTION

Sentiment analysis is the approach where we judge that how a person behaves on commodity or some act or fact. It is very important to know that what people can conceive about something. In modern era, almost everything is available on social media and people keep posting their thoughts, ideas, reviews. Lot of data can be gathered from social media like facebook, twitter for analysis. The important fact is that we can retrieve data about people that we do not know personally. Social media makes data gathering very easy, i.e., almost at our footsteps. The analysis of such data can help us to clinch recent trends about an event or happening basedinterest of people. "Wazwan the great Kashmiri cuisine" is the Kashmiri facebook page which I have used to gather data.

The methods that are used for sentiment analysis are like SVM (support vector machine), Naive Bayes machine learning. In this study, I have used lexicon based approach. As my page is based on Kashmiri language which is totally different from English. So, I have converted the Kashmiri words into their appropriate English words and filtered out those words which contain sentiments regarding the Wazwan (Kashmiri food). By doing this I got categories of sentiment words like positive, negative and neutral. So, I can determine that lexicon based approach works good here.

By doing this analysis we can determine the current trend of Kashmiri foods anddetermine where should the investors pay their attention.

Sentiment analysis: In modern era, sentiment analysis has become hot sensational research topic due to reason that many common real life problems are associated with this topic. Sentiment analysis research emerged from years 2000 and has become most effective research topic until now. Sentiment analysis is widely experimented in information retrieval system, data mining and web mining. Sentiment analysis is mainly based on three methods.

Document level (Pang *et al.*, 2002): In document level approach, we analyze the overall sentiment and conclude with the central idea that weather it is expressing a negative, positive or neutral opinion.

Sentence level: In sentence level, we analyze the sentences and determine the expression given by each sentence, i.e., negative, positive or neutral.

Aspect level (Wilson *et al.*, 2004): Also, called as feature level. This approach works in a different manner as compared to document level and sentence level. Aspect level cannot work on sentences, clauses, paragraphs, it straightly finds the target and the opinion.

The opinion can be a negative, positive or neutral towards the target. For example, "I hate kabab" (kabab is a Kashmiri dish made of meat). In this example, the person has made a negative opinion and the target is kabab. Some researchers, have classified two more categories as the challenge. They are regular opinion and comparative opinion.

Corresponding Author: Mohammad Arif, Department of M.E Bigdata and Analytics, Chandigarh University, Ajitgarh, Mohali, India

MATERIALS AND METHODS

Lexicon based approach: Sentiments can be determined by the words that are used in a review, post, comment by the user. In English language there are words which can be used to show positive, negative and neutral opinion. For example, admire, beautiful, care, honest are positive words and give a positive opinion about the target. While as dirty, jealous, poor are words which give negative sentiments. The combination of some words with positive and negative words can generate a neutral sentiment for example "the food was preferably good". We cannot only use these three categories to analyze the sentiment. The lexicon based method has some issues like.

A negative or positive word can have different or opposite sentiments for example, terrible is the negative word and generates a negative sentiment but when it is used like "the food was terribly delicious" give a super positive sentiment.

Sometimes a sentence does not contain any negative or positive words but still the sentence gives a sentiment or opinion over the target for example, "we sat for hours, then the food was served". In this example, it clearly gives a negative sentiment about the place that their timing is not good, although it does not contain any sentimental words. Sometimes some sentences, posts do not contain any opinions for example "me and my friend were at the restaurant".

Sentiment analysis on social media page: Social media is the hub of data where we can find data related to trends, relationships between entities, products. It is possible to understand the real-world movement by analyzing data of social media groups (Jindal and Liu, 2006). The data can be analyzed to predict future trends. The data is very much beneficial to ecommerce organizations because the insights that are generated from social media data helps them to produce trends that people discuss or like or show interest.

Sentiment analysis is the approach to generate or express positive, negative or neutral opinions towards an entity. There are mainly three components of sentiment analysis that are: "feeling" can be termed as opinion also. Opinion are used to bench, judge an entity. Opinions can be a decision, perception or a statement towards an entity. Opinions are personal feelings of every being that what they think or how they perceive things in mind. "Entity" anything that exists physically or has existence is called an entity. Entity may be a person, food, problem, organization. It is the entity on which the opinions are made. For example, "the sugar is sweet". "Subjectivity or emotions" emotions are feelings which depend upon mood and state of a person. Let us have an example from the facebook page Wazwan-the great Kashmiri cuisine. Posted by Piyush upadhyay 1 may 2014, Jammu city "goshtaba and roganjosh. I was disappointed at patnitop as far as Kashmiri food is concerned. It was impossible to find Wazwan there. Finally, my wish was fulfilled in Naz restaurant, Jammu. Not exactly like home cooked I was used to delicious nonetheless" (Goshtaba and Roganjosh are two famous Kashmiri dishes made of meat they are part of Wazwan) (Naz is a local restaurant in Jammu city) observations:

- This post contains a number of opinions
- Many entities are involved in this post
- The post contains both positive and negative sentiments

RESULTS AND DISCUSSION

My approach and experiment result

Data collection: As my analysis is on social media so I have chosen a facebook page Wazwan the great Kashmiri cuisine. The data can be downloaded from facebook API which is available for developers.

Formatting data: The data that was downloaded from facebook API is in JSON format. So, we have to do some formatting.

Formatting data for Naive Bayes classification: I tokenized the data into words then manually classified them into positive, negative, neutral, location time, etc., all the words are unique and redundant words were removed. There were around 7000 words.

Using Naves Bayes for this type of data had some issues. As we know that Naive Bayes is totally a probabilistic approach. Naive Bayes method works good for data like movies, novels. Here in our data we have words from Kashmiri language, Kashmiri written in English alphabet. Data is very noisy with a lot of spelling mistakes (Table 1).

As there is a lot of variation in the posts, so Naive Bayes approach could not be applied because it needs data with well-formed text corpuses like movies, novels. So, performance of Naive Bayes would have been decreased. So, I used lexicon based approach.

Table 1: Food status and problem

Status	Problem
The food was delicious.	No mention of which
But was very cold	food, veg or non-veg
Azz aies gamet aiss saalas,	Kashmiri written
tattie ous wariya maze	in English alphabet

Proposed solution: The counting of words was more efficient and then classifying them into positive, negative or neutral words. Since, people used different languages like Kashmiri, Urdu, English to post their reviews. It has been noted in the page that people have rated the food out of 10 sometimes. The score out of 10 can be also used to analyze the sentiment of the person. So, without reading the whole comment we can predict the sentiment by the score. The proposed solution:

- Retrieve data from facebook page
- Create a list of unique words that are used in all posts
- Remove words which do not provide any sentiment, location or time
- Rate the words which represent the sentiment and create a word list
- Check for the occurrences of words in word list
- Tag words to the posts that represent the sentiment
- When all the posts are tagged and sentiments are expressed create a macro level of the entity to generate the trend

Applications of sentiment analysis:

- Business intelligence
- Political scenarios
- Artificial intelligence
- Ecommerce organizations

Advantages of lexicon based sentiment analysis:

- Time saving and low cost approach of getting the sentiment
- Makes decision on knowing others interest
- Capability to act on user suggestions
- Helps to identify strength, weakness, threats and weakness of the organization
- Uses human perception

Sentiment expressed by emoticons: Facebook has evolved a lot, in modern era facebook uses six emoticons which are: like, love, haha, wow, sad and angry. Sentiments can be also expressed using these emoticons. People now a days prefer not to write comments, instead what they do is express their sentiment by using the emoticons. This can be rendered for future research that how can we exactly analyze the emoticons to get the proper sentiments.

CONCLUSION

In this study, I have shown how to analyze sentiments using facebook data which is very unstructured, cross language domain and noisy. Lexicon based works best on such type of data. The lexical approach is a method of teaching foreign languages described. At recent days researchers also working with lexicon based approach for sentiment analysis (Palanisamy *et al*., 2013; Taboada *et al*., 2011).

REFERENCES

Jindal, N. and B. Liu, 2006. Mining comparative sentences and relations. Proceedings of the 21st International Conference on Artificial Intelligence Vol. 2, July 16-20, 2006, ACM, Boston, Massachusetts, ISBN:978-1-57735-281-5, pp: 1331-1336.

Palanisamy, P., Y. Vineet and E. Harsha, 2013. Serendio: Simple and practical lexicon based approach to sentiment analysis. Proceedings of the 2nd Joint Conference on Lexical and Computational Semantics and Seventh International Workshop on Semantic Evaluation, June 14-15, 2013, Association for Computational Linguistics, Atlanta, Georgia, pp: 543-548.

Pang, B., L. Lee and S. Vaithyanathan, 2002. Thumbs up?: Sentiment classification using machine learning techniques. Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing Vol. 10, July 6-7, 2002, Association for Computational Linguistics, Stroudsburg, Pennsylvania, pp: 79-86.

Taboada, M., J. Brooke, M. Tofiloski, K. Voll and M. Stede, 2011. L exicon-based methods for sentiment analysis. Comput. Ling., 37: 267-307.

Wilson, T., J. Wiebe and R. Hwa, 2004. Just how mad are you? Finding strong and weak opinion clauses. Proceedings of the 19th International Conference on Artifical Intelligence Vol. 4, July 25-29, 2004, ACM, San Jose, California, ISBN:0-262-51183-5, pp: 761-767.