

Sentiment Categorization through Natural Language Processing : A Survey

Jyovita Christi*, Prof. Gayatri Jain

*1Information Technology Department, Gujarat Technological University, L.J. Institute of Engineering and Technology, Ahmedabad, Gujarat, India

ABSTRACT

Sentiment is an attitude, thought, or judgment prompted by feeling. Sentiment analysis, which is also known as opinion mining, studies people’s sentiments towards certain entities. Sentiment Analysis isn’t an unfamiliar term anymore. Today, smart phones, high speed Internet and various forums and social networks, have made it very common for people to give voice to their opinions. Therefore, a lot of textual data is available in various forms where people express their opinions. Analysing this data to know the underlying sentiment behind it has also become quite popular these days. Various techniques and applications have been created in the past and even today to perform sentiment analysis. This paper contributes towards understanding some of the modern techniques and in knowing which technique to use under what circumstances. It also studies feature extraction which is an important aspect of sentiment analysis. Feature extraction allows us to identify the features in the given text and analyse the sentiment for each feature.

Keywords : Sentiment Analysis, Categorization, Opinion Mining, Natural Language Processing, POS tagging, Aspects, Features

I. INTRODUCTION

Increased use of social media generates high-volume, high-velocity, high-variety, high-value, high-variability data termed as big social data. Several researchers have shown a keen interest in the exploitation of big social data in order to describe, determine and predict human behaviors in several domains. In fact, almost 80% of internet data is text [1], therefore, text analysis has become key element for public sentiment and opinion elicitation. Sentiment analysis, which is also called opinion mining, aims to determine people’s sentiment about a topic by analyzing their posts and different actions on social media.

Sentiment Categorization:

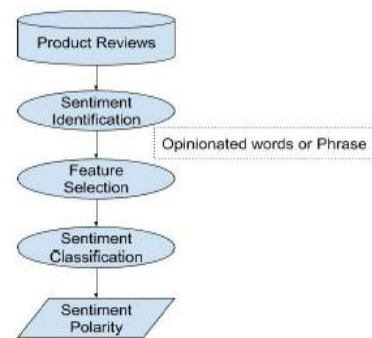


Figure 1. Sentiment analysis process on product reviews

Sentiments can be classified into several categories like positive, negative, neutral, happy, sad, angry, etc.

Natural Language Processing:

Natural language processing (NLP) is a subfield of linguistics, computer science, information engineering, and artificial intelligence concerned with the interactions between computers and human (natural) languages, in particular how to program computers to process and analyse large amounts of natural language data. [6]

Applications of Sentiment Analysis:

1) Online Commerce

All E-commerce websites allow users to write reviews about the products and their features. These reviews are publicly available and can be analysed to create a summary for a user to know the overall opinion about the product instead of having to go through all the reviews.

2) Voice of the Market (VOM)

Voice of the Market helps in determining the feelings of customers about products or services of competitors. This helps to manufacture products more suitable for the customers. Sentiment Analysis helps companies to know the real-time opinion of the people.

3) Voice of the Customer (VOC)

Voice of the Customer helps in determining what the individual customer or user is saying about products or services. It means examining the reviews, opinions and feedback of the customers. VOC is an important part of Customer Experience Management. To invent new products, VOC helps in finding such new opportunities. Extracting customer opinions also helps identify functional requirements of the products and some non-functional requirements like performance and cost.

4) Brand Reputation Management (BRM)

Brand Reputation Management is concerned about management of reputation in market. It focuses on product and company rather than customer.

Sentiment analysis helps in determining how company's brand, product or service is being perceived by community online.

5) Government

Sentiment analysis helps government in assessing their strength and weaknesses by analysing opinions from public. Sentiment analysis also helps in identifying strengths and weaknesses in a recruitment campaign in government job, assessing success of electronic submission of tax returns and many other areas.

II. LITERATURE REVIEW

In [1], the authors suggest an integrated framework which bridges the gap between lexicon-based and machine learning approaches to achieve better accuracy and scalability in Sentiment Analysis. A novel genetic algorithm (GA)-based feature reduction technique is proposed which can reduce the feature-set size by up to 42% without compromising the accuracy.

In [2], a general process for sentiment polarity categorization is proposed with detailed process descriptions for both sentence-level categorization and review-level categorization along with feature extraction.

In [3], online reviews from hotel booking websites are crawled along with their metadata and classified into predefined classes based on some common aspects. Then, Topic Modelling Technique called LDA is applied to identify hidden information and aspects which is followed by performing sentiment analysis on classified sentences and summarization.

In [4], a dynamic dictionary of words' polarity based on a selected set of hashtags related to a given topic is constructed. This is then used in classifying the tweets under several classes by introducing new features that strongly fine-tune the polarity degree of

a post. The classified tweets are related to the 2016 US election.

In [5], brand-related tweets associated with five leading UK online retailers during important sales periods of the year like Black Friday, Christmas and New Year are analysed. A combination of data analytics approaches including time series analysis, sentiment analysis and topic modelling are used. Time series analysis is used to understand the reasons behind deviations in the sentiment due to a certain time period. Topic modelling is used to know what exactly drives these deviations.

III. METHODS AND MATERIAL

Approaches for Sentiment Analysis:

1. Lexicon analysis:

Uses a dictionary and categorizes the sentiments on the basis of semantic orientation of words in the text.

2. Machine Learning:

Builds models trained using a labeled dataset known as training data set which in turn can be used to classify the sentiments behind some other text.

3. Rule-based approach:

In rule based technique, if a rule has “if-then relation then it consists of an antecedent and its associated consequent. A consequent denotes a sentiment that can be either positive or negative.

4. Keyword-based approach:

In Keyword-based approach, main task is the construction of word lexicons. So that, text can be classified into affect category on the basis of presence of affect words like “happy”, “awesome”, “sad”, “bored”. [2]

IV. RESULTS AND DISCUSSION

Although Sentiment Categorization can be done in many ways, today’s data present a lot of challenges. Some of these challenges are spelling mistakes, short

forms, slang words, etc. Moreover, social networking data contains more hashtags and video links than actual text. This presents us with the question of effectively analysing such data. Another challenge is knowing how much data is enough because there is unlimited supply of data. So, in order to maintain accuracy, one has to work with Big Data tools and collect data over a period of time and then apply an appropriate method for analysing it. The dictionary used for analysis must also be a dynamic dictionary as that’s the only way in which one can cope up with rapidly changing linguistic patterns.

V. CONCLUSION

A novel approach to do sentiment categorization is to create a content specific dynamic and adaptable dictionary. This can be used to analyse sentiments from data obtained through web crawling or APIs. Hashtags are also an important part of the textual data these days and must be identified and categorized just like any other text. Effective pre-processing is quintessential to accurate sentiment categorization. This should include stop word removal, POS tagging, correcting spelling errors and more. No amount of data is enough for analysis these days. So, in future, analysis should be performed using Big Data tools and technology along with traditional methods.

VI. REFERENCES

- [1]. Farkhund Iqbal, Jahanzeb Maqbool Hashmi, Benjamin C. M. Fung, (Senior Member, Ieee), Rabia Batool, Asad Masood Khattak, Saiqa Aleem , And Patrick C. K. Hung, “A Hybrid Framework for Sentiment Analysis Using Genetic Algorithm Based Feature Reduction”. (08 February, 2019), DOI 10.1109/ACCESS.2019.2892852
- [2]. Xing Fang* and Justin Zhan “Sentiment analysis using product review data” Journal of Big Data (2015) 2:5, DOI 10.1186/s40537-015-0015-2

- [3]. Nadeem Akhtara*, Nashez Zubaira, Abhishek Kumara, Tameem Ahmada “Aspect Based Sentiment Oriented Summarization of Hotel Reviews” 7th International Conference on Advances in Computing & Communications, ICACC-2017, 22-24 August 2017, Cochin, India (Elsevier 2017)
<https://doi.org/10.1016/j.procs.2017.09.115>
- [4]. Imane El Alaoui* , Youssef Gahi, Rochdi Messoussi, Youness Chaabi, Alexis Todoskoff and Abdessamad Kobi “A novel adaptable approach for sentiment analysis on big social data” El Alaoui et al. J Big Data (2018) 5:12
<https://doi.org/10.1186/s40537-018-0120-0>
- [5]. Noor Farizah Ibrahima, Xiaojun Wang “Decoding the sentiment dynamics of online retailing customers: Time series analysis of social media” Computers in Human Behavior 96 (2019) 32-45
<https://doi.org/10.1016/j.chb.2019.02.004>
- [6]. https://en.wikipedia.org/wiki/Natural_language_processing

Cite this article as :

Jyovita Christi, Prof. Gayatri Jain, "Sentiment Categorization through Natural Language Processing : A Survey", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 6 Issue 6, pp. 104-107, November-December 2019. Available at doi : <https://doi.org/10.32628/IJSRSET196626>
Journal URL : <http://ijsrset.com/IJSRSET196626>