

A SURVEY ON CONCEPTUAL ADVANCEMENTS IN SENTIMENT ANALYSIS

Devesh Rathi B¹, Shinde Shrutika S², Dhoka Shubham R³, Jagruti Mahajan R⁴

¹UG Student, Computer Engineering, GHRCOEM, Maharashtra, India

²UG Student, Computer Engineering, GHRCOEM, Maharashtra, India

³UG Student, Computer Engineering, GHRCOEM, Maharashtra, India

⁴Associate Professor, Computer Engineering, GHRCOEM, Maharashtra, India

Abstract

Sentiment analysis is defined as the process of computationally identifying and categorizing the opinions expressed in the written text to analyze the writer's attitude and finding out whether the written text is positive, negative or neutral in nature. Over the past few years, it attracted a great deal of attentions from both academia and industry due to many challenging research problems and a wide range of applications. These opinions are generally obtained in text format through various databases such as twitter, customer forum, and complaint websites. In these survey paper we have studied and shown the advancements of the sentiment analysis over the past few years and have inhabited the growth of various concepts under it. This paper gives a review of various facts about the different perceptions related to sentiment analysis and opinion mining. We have also suggested a system as a future scope which works on the analysis of various opinions and creates a module that can handle multiple entities and queries related to it.

Key Words: Sentiment analysis, Opinion mining, Hadoop Framework, Node architecture.

-----***-----

1. INTRODUCTION

The current era states sentiment analysis as opinion mining, which has been widely applied to various Sectors of data processing, such as movies or product reviews, webpages and blogs. Opinion is defined as an individual's view towards a specific entity. These entities can be an individual, company, product, customer's reviews, media, sports, politics and real time events such as violent crimes and natural calamities. Also the customizations can be done to get the views of different entities according to our needs, which include various specific information or opinion for the entity such as its popularity of the entity or its views or followers or the authenticity of the opinion.

One of major examples that we have studies is of Canada Expedia. The company launched a music video which was lashed out onto in various social networking sites due to its continuous telecast, based on the opinions of the public they build a model which helped them to understand the ills of their video, after considering this opinions they launched a spoof video for the same previous video which was laughed at and appreciated by everyone. The Indian government has also adapted this concept and made a website that work on this analysis and framework (data.gov.in). We can even collect the data from various sectors and build a statistic records so as to help the purpose of improvement for various sectors.

We suggest the use of Hadoop framework instead of the traditional frameworks as they are not able to handle the large datasets, where HADOOP has potential to work with large datasets and process them efficiently. The paper also suggests the use of multi-node architecture over single node

architecture. The advantages of multi-node architectures are described below:

1. Horizontal Scalability which helps in increased transfer flow capacity.
2. High availability enables patch management on active cluster and the node automatically restarts after fail over.
3. With Traffic management, it balances load between nodes.

There are two possibilities where a node manager detects a node failure such as:

1. If it is a local node fail over, the node manager automatically attempts to restart the node on the local server.
2. If it is a remote node fail over, the node manager wait for the remote node manager to restart the node. If the remote node manager does not restart node before timeout, the local node manager will restart the node on local server.

After the node is restarted, whether local or remote, it will complete all transfer requests that were active when the failure occurred.

2. RELATED WORK

In 2010, Sentiment analysis approach was discovered [2]. Traditionally sentiment analysis was applied for minor organizations, small scale industries. In the past years it attracted a great deal of attentions from both academia and industry due to many challenging research areas and applications [3]. Then afterwards it was referenced to a broad range of fields of NLP (Natural Language

Processing), Computational linguistic and text mining & classification of reviews, remarks has emerged as the most useful application in the area of sentiment analysis. Then this concept is further expanded for social media with age. Social media was referred for the purpose of sentiment analysis as it offers sentiments rich data and that data is in the form of the tweets, blogs, status updates or posts. Devi. K and Bhaskaran V. M. enhanced the concept of sentiment analysis by automatically classifying the emotions and mood [6].

3. LITERATURE SURVEY

Hu and Bing Liu in 2004 suggested that the opinion summarization of different reviews of customers for various products placed on e-commerce websites is quite different than the normal text categorization because we mine only the features of the products. In their paper they created categories for the purpose of classification and based on the opinions the sorted them accordingly [1].

Bing Liu in 2010 stated opinions are major part of many research problems and many applications because whenever we need to make a decision we want to hear others opinions. This is a major context for not only an individual but also for the multiple organizations. He also gave some computational studies for the purpose of opinion mining which were not present or yet discovered because of the lack of opinionated text [2].

Khan A. and Baharudin S in 2011, proposed a method which classified subjective and objective sentences from reviews and blog comments. With the help of SentiWordNet **sematic** scores of these subjective sentences were calculated in order to get their polarity as positive, negative or neutral based [3].

Natural Language Processing, computational linguistics and text mining are some of the major fields that refer sentiment analysis. Bag of words and feature based sentiment analysis were most popular approaches earlier but nowadays researchers are concentrating more on products. Karamibekr M. and Ghorbani A. A. in 2012 implemented all these concepts to focus integration of sentiment analysis for social issues [4].

Neethu M. S. and Rajasree R. in 2013 gave a better way to classify the opinions and tweets for opinion mining from twitter using a feature called vector since in general the data is tough to classify due to presence of slang languages and misspellings [5].

Mouthami K. et al. in 2013 implemented this concept for better understanding the various reviews of the different masses that place their opinions on different social media's for building a model for different sectors and companies to improve their performance [6].

Zhaoxia Wang et al. in 2014 stated that sometime the traditional method for collecting data from different social media fails; their proposed method overcomes this

disadvantage by not only enhancing the capability of algorithm but also performing analysis for non-English languages [7].

Shulong Tang et al. in 2014 described twitter as valuable platform for tracking and analyzing public sentiments, which provides critical information for decision making. In their work, they moved one step further to deduce sentiment variations. Based upon this observation, they proposed a Latent Dirichlet Allocation (LDA) based model, Foreground and Background LDA (FB-LDA), to purify forefront topics and filter out age-old circumstantial topics [8].

Ahmad S. R. et al. in 2015 showed different comparative studies for the purpose of selection of features for a specific product since that will help in deciding the criticism of people for that product; they also suggested the use of metaheuristic algorithm in sentiment analysis for better results [9].

Chinsha T. C. and Joshi S. in 2015 suggested aspect level opinion mining and proposed a syntactic based approach for opinion mining, which is based on syntactic dependency and aggregate score of opinion words. The experimental work was done on restaurant reviews. Their method achieved total of 78.04% for the dataset of restaurant reviews [10].

4. FUTURE SCOPE

Creation of a distinguished archetypal model would be a good way to get the overall overview of different emotions that a data or opinion would provide and could be easily classified into different aspects. Online public sentiment analysis is an increasingly popular topic in social network related research. There has been some research work focusing on assessing the relations between online public sentiment and real-life events (e.g., consumer confidence, stock market). They reported that events in real life indeed have a significant and immediate effect on the public sentiments in Twitter. Instead of using traditional frameworks for processing the data sets or the incoming data, the use of Hadoop framework is suggested. Since it has exceptional qualities to deal with large data volume and can process it with various functionalities. Availability of data can be through various sources such as social media, different databases, or various websites, but taking the data from twitter is quite easy due presence of variety in data sets, the concept of machine learning for training the system and making it more accurate for processing the data should also be considered.

5. APPLICATIONS

- Building a score system based upon the sentiment analysis of the opinions instead of using traditional methodologies.
- Sentiment analysis can help various politicians and political parties to understand the views of the mass people during elections for better campaigning.

- Small scale food industries can improve their product based upon the moods of people towards their product such as consistency, flavor and other factors.
- MNC's can create a hierarchical model for understanding the response of their employees regarding their payment structure or the schedule policies.
- Performing opinion mining on day-to-day products or brands to maximize the marketing of the product.
- For media and sports sectors, the information integrated into the market can be processed for sentiment analysis to make audience understand it in a better way, which helps them to visualize different aspects of an entity.

- [9] Ahmad S. R., Abu Bakar A., Yaakub M. R., "Metaheuristic algorithms for feature selection in sentiment analysis", Science and Information Conference (SAI), 2015.
- [10] Chinsha T. C., Joseph S., "A syntactic approach for aspect based opinion mining", Semantic Computing (ICSC), 2015 IEEE International Conference on, Anaheim, CA, 2015.

6. CONCLUSION

The overall advancements in the field of Sentiment Analysis are thoroughly described in this paper and also the actual work done and some innovations are to be seen here. The use of Hadoop Framework would ease the process of data managing and processing if used and also would increase the reliability of the system. Also we have observed that there is no concrete model for fulfilling the specific analysis for opinion mining and our proposed model might substitute for the requirement.

REFERENCES

- [1] Hu and Bing Liu. "Mining and Summarizing Customer Reviews", ACM SIGKDD International Conference on Knowledge Discovery and Data mining, Washington, USA, Aug 22-25, 2004.
- [2] Bing Liu. "Sentiment Analysis: Multifaceted Problem", IEEE Intelligent System, 2010.
- [3] Khan A., Baharudin S., "Sentiment classification using sentence-level semantic orientation of opinion terms from blogs", National Postgraduate Conference (NPC), 2011, July, 2011.
- [4] Karamibekr M., Ghorbani A. A., "Sentiment Analysis of Social Issues", Social Informatics (Social Informatics), 2012 International Conference on, 14-16 Dec. 2012.
- [5] Mouthami K., Devi K N., Bhaskaran V. M., "Sentiment analysis and classification based on textual reviews", Information Communication and Embedded systems (ICICES)", 2013 International conference on, Chennai, 21-22 Feb. 2013.
- [6] Neethu M. S., Rajasree R., "Sentiment analysis in twitter using machine learning techniques Computing", Communications and Networking Technologies (ICCCNT), 2013 Fourth International Conference on, Tirchengode, 4-6 July 2013.
- [7] Shulong Tan, Yang Li, Huan Sun, Ziyu Guan, Xifeng Yan, Member, Jiajun Bu, "Interpreting the Sentiment Variations on Twitter", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, MAY 2014.
- [8] Zhaoxia Wang, Jpp. V, Chuan Tong, Chan D, "Issues of Social Data Analytics with a New Method for Sentiment Analysis of Social Media Data", Cloud Computing Technology and Science (CloudCom), 2014 IEEE 6th International Conference on, Singapore.