

International Journal of Computer Science and Mobile Computing



A Monthly Journal of Computer Science and Information Technology

ISSN 2320-088X

IMPACT FACTOR: 6.199

IJCSMC, Vol. 8, Issue. 8, August 2019, pg.77 – 81

Review of Various Sentiment Analysis Techniques of Twitter Data

Bipin Gupta

M.Tech Scholar

gbpin002@gmail.com

RIMT University

Gobindgarh

Mr. Ankur Gupta

Assistant Professor

ankurgupta@rimt.ac.in

RIMT University

Gobindgarh

Abstract: A sentiment is known as an attitude or judgments towards any particular event or person. Opinion mining is known as the extraction of these sentiments or opinions from given data. This technique also helps in analyzing the kinds of sentiments that people have towards specific objects or services. The best source available to collect the sentiments is internet. Twitter is known as a social networking platform that is accessed by users for posting their views online. Due to the unique properties of Tweets, there is an increase in new challenges. Higher analysis studies are needed in case of sentiments as compared to other domain applications due to their complexities. Analyzing the sentiments of product reviews for Amazon is the focus of this research. Classification as well as feature extraction is applied such that sentiment analysis can be performed.

KEYWORDS: *Sentiment Analysis, SVM, KNN*

Introduction

With the increase in amount of data available in different applications, many large sized databases have been designed to store them. People are using new methods to perform operations and the inputs and outputs are stored with time on these large sized databases. Data mining is a KDD based technology that examines and reveals the important information from the raw data collected in large databases [1]. Cluster analysis is known as an important method for data mining. Either the dissemination of data index can be understood or the calculation values can be performed over the available data through pre-processing using this extracted data. The data mining approach is applied for differentiating the new, legitimate, reasonable and potentially valuable designs. This technology aims to extract the prescient data that is available in huge databases for examining the data available in the information distribution centers. An approach also known as opinion mining in which the opinions of people related to particular services are categorized is known as sentiment analysis [2]. Based on the emotions and attitudes of certain event or object, the opinions and perspectives of humans are analyzed through sentiment analysis. In the applications like social media analysis or commercial product reviews, opinion mining is performed. For creating the recommender systems, semantic analysis is considered as a valuable technique. On the e-commerce and social networking

websites, several online reviews and comments are mentioned by users. These sources help in understanding the opinions of users in an effective manner [3]. For checking if the reviews of users about the products are positive, negative or neutral, the sentiment analysis is performed. These reviews help in defining the important or popularity of products in the competitive market. For a specific event, the opinions, feelings, thoughts and emotions are different for every human being. Each of the sentiment denotes a different category since every sentiment analysis can be considered as a separate task of classification process. Since it deals with the human and computer language interaction, the AI and computer science play an important role in NLP [4]. Due to the huge changes arising in market level of competition, more research needs to be done in sentiment analysis such that effective outcomes can be achieved. The opinions of speaker are determined by the sentiment analysis otherwise commonly known as opinion mining. Here, an appropriate review related to a product or service is provided through this method. It analyzes all the opinions and information provided only related to the product under review [5]. For analyzing the opinions of an individual user, the data collected from various users is improved. For posting their opinions or using blog posts, several social networking platforms are provided to the users online. The platforms like Instagram, Google, Twitter, and Facebook are some of the commonly known sites. The views and opinions of various users are available here. To improve the quality of services, the most important component to be considered is the user's opinion. The deliverable products are improved here. The reception level of products and services can be understood in a good way through the micro blogs, review sites, blogs and online posts. Different kinds of classifiers are applied to perform twitter sentiment classification through text classification. Some of the techniques are explained further [6].

a. **Lexicon Based Approach:** There are lexicographical resources included here. The collection of seeds of sentiment words and their orientation is the initial step. The sets can be expanded by finding their antonyms and synonyms. The iteration process will stop after no words are left to be analyzed.

- **Dictionary Based Approach:** There are lexicographical resources included here. The collection of seeds of sentiment words and their orientation is the initial step. The sets can be expanded by finding their antonyms and synonyms. The iteration process will stop after no words are left to be analyzed. Extracting the opinions is not possible here due to the specification of domain orientations, which is the major disadvantage of this designed approach [7].
- **Corpus-Based Approach:** When a single corpus is used for particular domain within the discovery process, the particular opinion of the domain and its orientation is identified. The similar word in the similar domain is identified here that does not show the domain based words but has various contexts.

b. **Machine Learning Approach:** These approaches are used to solve the problems associated with the classification of sentences. These approaches are broadly categorized among supervised and unsupervised learning approaches [8]. Huge amount of labeled training document is used to define the supervised learning. The two approaches are explained below:

- **Unsupervised learning:** Due to the absence of any category or correct information related to the appropriate targets, the clustering techniques completely rely on this kind of learning.
- **Supervised learning:** From these datasets, significant outputs are achieved when encountering them through decision making process.

c. **Hybrid Approaches:** These are the approaches that combine the elements from lexicon-based techniques and machine learning techniques. For determining the semantics available in highly appropriate way, these techniques are used as semantics networks and ontologies [9].

To perform text classification that can be applied for twitter sentiment classification, various classifiers have been designed among which few popular ones are described below:

- i. **Maximum Entropy Algorithm:** For providing extra semantic, syntactic features that are used with huge flexibility, this algorithm is applied.
- ii. **SVM Classifier:** To perform classification, the massive edge is provided by SVM classifier.

iii. Ensemble classifier: Today, several ensemble classifiers are available. By ensuring that all the embedded features of the base classifiers are used to the fullest, the best classification is provided here.

Literature Review

Dan Cao, et.al (2016) stated that Automatic Text Summarization approach intended to make a condensed adaptation of documents. This version should be able to cover all important contents and common information. In this study, all features that utilized metrics and thought of complicated network for scoring sentences were reviewed [10]. In this study, tested outcomes on individual module and mixture of various presented were analyzed. DUE 2002 data sets were utilized to evaluate quantitative and qualitative features. Shortened ways were identified as amazing for text summarization. With respect to the quality of produced summary, these ways attained maximum grades. An additional significance was the detecting those that featured mixtures with same assets of network and specified unbelievable effect on chosen sentences. It was identified that Sentence correlation among sentences became a necessary element in the retrieval of fine abstracts.

RasimAlguliyev, et.al (2016) presented a study in which main attention was given to extractive text summarization [11]. In this approach, a summary was created with the help of scoring and selection of sentences in the source text. Initially, the score of each sentence was evaluated and further most representative sentences were selected from the text in view of the fact that semantic resemblance among selected sentences would be low. In order to score sentences, one more formula was established. The proposed approach depicted accomplishments for finding equilibrium between coverage and repetition in an abstract. In this study, a human learning optimization algorithm was utilized to handle optimization problem.

NarendraAndhale, et.al (2016) stated that the procedure used for the generation of compressed structure of text document was known as text summarization. An important technique using which relevant information could be recognized from enormous documents was called automatic text summarization approach. The wide-ranging analysis of both approaches was presented within the text summarization in this study [12]. In this study, numerous extractive and abstractive sorts of summarization techniques were analyzed. An effective summary was to be generated by summarization approach in minimum time slot. This summary had less redundancy and included well-formed sentences. High-quality results were attained with the help of extractive and abstractive methodologies. These results could be utilized further by the users. The testing for hybridization was analyzed in this study for generating helpful, fine condensed and understandable abstracts.

Rupal Bhargava, et.al (2017) proposed a novel technique for analyzing various languages in order to discover sentiments in these languages and performed sentiment analysis [13]. The proposed technique implemented different machine learning methodologies for content inspection. In order to deal with various languages, Machine translation was utilized in this mechanism. In order to find sentiments in content, content was processed following the machine translation. The introduction of blogs, forums and online surveys resulted in the ample amount of online text. This sentiment regarding a specific topic or an object could be analyzed using this text. Therefore, it was advantageous to retrieve important text occurring within this text for reducing further processing. Thus, the presented structure utilized text summarization procedure for extracting significant elements of text. These parts were then utilized to examine sentiments regarding some specific matter and its features.

Archana N. Gulati, et.al (2017) proposed a new method for multiple documents and extractive text summarization was presented by considering this issue [14]. A condenser or summarizer for Hindi language was fabricated by considering it the most common language of India. Input to the system was applied in form of News editorials regarding sports and political affairs from Hindi newspapers which were available online. Eleven important aspects of the text had been utilized for retrieval procedure using Fuzzy inference system. A standard accuracy of about 73% over numerous Hindi documents was achieved by the proposed approach. The system produced summary was similar to the human produced summary up to some extent. The system generated summary showed good Precision, Recall and F-score values.

Manisha Gupta, et.al (2016) proposed a new scheme for summarizing Hindi text document was proposed in this study [15]. This approach was based on several linguistic rules. For generating smaller amount of words from the real document, dead wood words and phrases were eliminated from the actual text as well. The performance of resented approach was tested on

numerous Hindi inputs. The accuracy of proposed approach was obtained as amount of lines retrieved from actual document holding significant information of the actual text. The proposed approach reduced the text size of information up to 60% - 70 %. The extractive summary provided by user was generated by proposed approach.

Akshi Kumar, et.al (2017) stated that various algorithms had been utilized in text summarization process [16]. In this study, author had analyzed and compared the related performances of three different algorithms. Initially, the different text summarization methods were explained. Secondly, Extraction based techniques were utilized for the extraction of significant information or keywords involved in the summary. Abstraction based methods generated their own sentences for text summary. Three algorithms were elucidated and applied in python programming language. A parameter named ROUGE-1 was utilized for evaluating the effectiveness of retrieved keywords. The outcomes of all algorithms were compared with the handwritten summaries for evaluating the performance of these algorithms. It was analyzed that TextRank algorithm showed superior performance among all other algorithms.

Table of Comparison

Authors Names	Year	Description	Outcomes
Dan Cao, et.al	2016	In this study, all features that utilized metrics and thought of complicated network for scoring sentences were reviewed.	It was identified that Sentence correlation among sentences became a necessary element in the retrieval of fine abstracts.
Rasim Alguliyev, et.al	2016	In this approach, a summary was created with the help of scoring and selection of sentences in the source text.	The proposed approach depicted accomplishments for finding equilibrium between coverage and repetition in an abstract.
Narendra Andhale, et.al	2016	In this study, numerous extractive and abstractive sorts of summarization techniques were analyzed.	The testing for hybridization was analyzed in this study for generating helpful, fine condensed and understandable abstracts.
Rupal Bhargava, et.al	2017	The proposed technique implemented different machine learning methodologies for content inspection.	These parts were then utilized to examine sentiments regarding some specific matter and its features.
Archana N. Gulati, et.al	2017	A condenser or summarizer for Hindi language was fabricated by considering it the most common language of India.	The system produced summary was similar to the human produced summary up to some extent.
Manisha Gupta, et.al	2016	A new scheme was proposed for summarizing Hindi text document was proposed in this study. This approach was based on several linguistic rules.	The proposed approach reduced the text size of information up to 60% - 70 %. The extractive summary provided by user was generated by proposed approach.
Akshi Kumar, et.al	2017	In this study, author had analyzed and compared the related performances of three different algorithms.	It was analyzed that TextRank algorithm showed superior performance among all other algorithms.

Conclusion

In this research study, the behavior of user is analyzed on the basis of sentiment analysis of twitter data. N-gram method is implemented in this study for sentiment analysis for analyzing certain features of input data. Moreover, a classification technique is applied to analyze the behavior of client. The N-gram method divides whole input dataset into several sections. All segments are analyzed individually in order to analyze sentiments. In this research study, a classification model called logistic regression is used for analysis. During the classification of data, various classes are produced.

References

- [1] Tharindu Weerasooriya, Nandula Perera, S.R. Liyanage. A method to extract essential keywords from tweet using NLP. 2016 16th International Conference on Advances in ICT for Emerging Regions(ICTer).
- [2] Ibrahim A. Hameed. Using Natural language processing for designing socially intelligent robots. 2016 Joint International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob).
- [3] L. Suanmali, M. S. Binwahlan, and N. Salim. Sentence features fusion for text summarization using fuzzy logic in Hybrid Intelligent Systems. 2009, HIS'09, Ninth International Conference on, vol. 1, IEEE, 2009, pp. 142-146.
- [4] L. Suanmali, N. Salim, and M. S. Binwahlan. Fuzzy logic based method for improving text summarization. arXiv pre print arXiv:0906.4690, 2009.
- [5] X. W. Meng Wang and C. Xu. An approach to concept oriented text summarization, Proceedings of ISCITS05, IEEE international conference, China, 1290-1293" 2005.
- [6] M. G. Ozsoy, F. N. Alpaslan, and I. Cicekli. Text summarization using latent semantic analysis. Journal of Information Science, vol. 37, no. 4, pp. 405-417, 2011.
- [7] AdyanMarendraRamadhani, Hong Soon Goo. Twitter Sentiment Analysis using Deep Learning Methods. 7th International Annual Engineering Seminar (InAES), Yogyakarta, Indonesia, 2017.
- [8] K. Kaviya, C. Roshini, V. Vaidhehi, J. DhaliaSweetlin. Sentiment for Restaurant Rating. 2017 IEEE International Conference on Smart Technologies and Management for Computing, Controls, Energy and Material (ICSTM).
- [9] Devika M D, Sunitha C, Amal Ganesh "Sentiment Analysis:A Comparative Study On Different Approaches", Procedia Computer Science, vol.87 , pp. 44-49, 2016
- [10] Dan Cao, LiutongXu. Analysis of Complex Network Methods for Extractive Automatic Text Summarization.2016 2nd IEEE International Conference on Computer and Communications, vol. 9, iss. 8, pp- 97-110, 2016.
- [11] RasimAlguliyev, RamizAliguliyev, NijatIsazade. A Sentence Selection Model and HLO Algorithm for Extractive Text Summarization, IEEE, vol. 9, iss. 8, pp- 97-110, 2016.
- [12] NarendraAndhale, L.A. Bewoor. An Overview of Text Summarization Techniques. IEEE, vol. 9, iss. 8, pp- 97-110, 2016.
- [13] RupalBhargavaandYashvardhan Sharma. MSATS: Multilingual Sentiment Analysis via Text Summarization, IEEE, vol. 9, iss. 8, pp- 97-110, 2017
- [14] ArchanaN.Gulati, Dr.S.D.Sawarkar. A novel technique for multi-document Hindi text summarization. 2017 International Conference on Nascent Technologies in the Engineering Field (ICNTE-2017), vol. 8, pp. 1-4, 2017.
- [15] Manisha Gupta, Dr.Naresh Kumar Garg. Text Summarization of Hindi Documents using Rule Based Approach, International Conference on Micro-Electronics and Telecommunication Engineering, vol. 8, pp. 1-4, 2016.
- [16] Akshi Kumar, Aditi Sharma, SidhantSharma,ShashwatKashyap. Performance Analysis of Keyword Extraction Algorithms Assessing Extractive Text Summarization. International Conference on Computer, Communication, and Electronics (Comptelix), 2017.