ISSN: 2277-5528
Impact Factor- 4.015

SENTIMENT ANALYSIS OF TECHNICAL WEB DATA MINING

Sanjay Singh Bhadoria[#1], Dr. Dhanraj Verma[*2]

[#1]Research Scholar, [*2] Professor, Dept. Of Computer Application, College of Engineering,
Dr. A.P.J. Abdul Kalam University, Indore (M.P.), India
[1]sanjay.bhadoria@gmail.com, [2]dhanrajmtech@gmail.com

Abstract

Educational Datamining (EDM) has inspired the development of innovative approaches and improvements in instructional settings. The Vast Array of practice and research in this area has enforced significant possibilities and software out of personalization and adaptation Design and pedagogy decisions centered on students' needs. Learning Analytics (LA) and EDM play with an important Role in improving learning procedures by offering advanced software of analytics Techniques. This also Results in the understanding discovery regarding the learning procedures, and advancement and integration of personalized, flexible, and interactive informative surroundings. Technology enhanced learning (TEL) surrounds to boost the information and abilities of students. Inquiry based learning (IBL) targets contexts where students are intended to detect knowledge as opposed to passively memorizing the theories it eases learning and improve learning accomplishments of the students.

Index Terms: Educational Datamining (EDM), Learning Analytics (LA), Technology, Enhanced learning (TEL), Inquiry based learning (IBL)

Introduction

A huge quantity of data has been generated daily through different trades in businesses, social media, communicating systems, etc.. Substantial data is really a word which reflects vast amounts of top speed, complex and changeable data which want complex procedures and technologies to allow the capture, storage, control, and investigation of their data. Significant data investigation could be the potential for representing of use information from these types of large data sets. As a result of attributes such as volume, ethics, and speed, big statistics investigation is now among the very challenging research issues. Mining purposeful patterns from enormous input information for conclusion, forecast, etc. are at the heart of significant data investigation. While big-data is a considerable issue, the actual problem is making feeling of data that is big and finding info information inside it, which helps users create smarter conclusions. Semantic could be considered as a magical term to bridge the gap in the diver- sity of data. Semantics can be used in meaningful data integration, decision system which makes it possible to detect inconsistency of data, discovers new (hidden) knowledge, etc. Semantic analysis presents the data in a more efficient manner and makes it useful as a source for knowledge discovery and comprehension. Sentiment analysis is method to better understand the implied or practical meaning of the input dataset.

Semantic Analysis

Given such large sizes of text data sets, mining programs, that arrange the writing data sets into structured knowledge, will enhance efficient document access. This eases semantic search and at precisely the same period, provides an efficient platform for classification as data becomes extremely huge. Semantic analysis can be a process to better understand the implied or practical significance of the input data set. It is largely applied with ontology to analyze content chiefly from web resources. Ontology fitting is a remedy to the semantic heterogeneity problem. It finds correspondences between semantically related entities of ontology. All these correspondences can be employed for various tasks, such as for example ontology blending, query answering, or

ISSN: 2277-5528
Impact Factor- 4.015

statistics translation. More conventional machine learning algorithms aren't efficient enough to extract the semantic information (hidden) generally exhibited in big data. By extracting such attributes, semantic analysis empowers using classification, forecast, which is essential when developing models to manage the scale of data that is big. The telltale knowledge is used to aid sentiment investigation of queries like emotion mining, popularity analysis, recommendation systems, user profiling, etc.. Separate sources together with advice could be linked through semantic annotations.

SENTIMENT ANALYSIS

These days, clients rely on the Internet for Those opinions on various products and services. As a result of rapid growth of data accessibility in the internet, it's very tricky to manage these to a program. Again, these statistics are in heterogeneous formats as well as rapidly changing in nature. Hence, there's a demand for a highly effective method to categorize and analyze web reviews being a significant statistics investigation. The task of refining and assessing such collective net information together is called opinion mining, also it is also called sentiment analysis. Sentiment analysis refers to the computational techniques for extracting, perfecting, comprehension, and analyzing the opinions expressed in news, commentaries, along with other relevant content. An important component of our information-gathering is to find out what other people think. Before the web has come to be widely distributed, people asked friends for remarks on different subjects in order to make a better and wiser decision. It offers exciting opportunities for understanding the semantics of the general public and customers seeing social events, political moves, company strategies, promotion campaigns, and product preferences. Sentiment analysis is a very ambitious and promising field which uses both intersections of information retrieval and computational linguistic methods to deal with the reviews expressed from the source material.

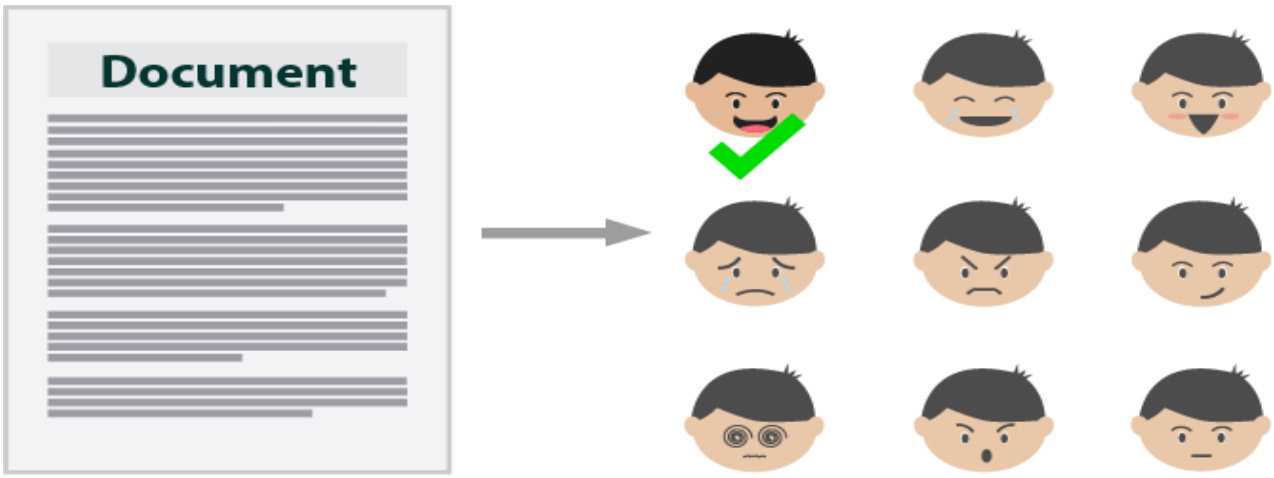

Figure 3.1 : Sentiment Analysis Concept

ISSN: 2277-5528
Impact Factor- 4.015

MOTIVATION

The main limitation of traditional data models is that these are incapable of handling unstructured, heterogeneous, complex data,

- Massive volumes of data need a suitable semantic representation to use them correctly and efficiently.
- Semantic could be considered as a magical term to bridge the gap in the diversity of data. Semantics can be used in meaningful data integration, effective decision system, discovers new (hidden) knowledge, etc. Big data analysis such as knowledge representation, information tagging, fast information retrieval etc. can be better addressed with the aid of semantic analysis. More traditional machine learning algorithms are not efficient enough to extract the semantic information (hidden) generally presented in big data. By extracting such features, semantic analysis enables the use of classification, prediction, which is important when developing models to deal with the scale of big data.
- In today's world, users can post their comments on any internet forums like review sites, blogs, discussion groups etc. These are commonly known as user content generator which contains the important information.
- Currently, there is not any recognized effective and efficient data model to manage enormous data with semantic understanding. It is an excellent challenge to effectively describe prescriptive features of big data as well as to build semantic organization model to bridge the semantic gap of varied data sources. Customers remain fighting with hunting for actual information in addition to filtering large volumes of articles to get relevant information for their own interests.

PROPOSED APPROACH

Since the number of Internet users of networking platforms and services grow fast, more and more data from these platforms can be used for data mining studies as well as sentiment analysis. Web can be considered as a huge source of short texts (e.g. comments) containing user opinions. Useful conclusions can only be extracted when huge amounts of text or comment are analyzed. However, standard solutions cannot handle large text data in a reasonable time. In this way, making sentiment analysis on comments is challenging on performance when huge amounts of comments are to be processed. To address this challenge we propose to take advantage of parallel architectures using big data technologies.

We presents a new semi –supervised approach for sentiment analysis of large web-based data based on big data technologies. The review of relevant literature highlights the most commonly encountered challenges which are context sensitivity, negation, sarcasm, domain sensitivity etc.

ISSN: 2277-5528
Impact Factor- 4.015

We have considered large web reviews data and negation as well as intensity of sentiment challenges in sentiment analysis. Here, we convert the input text data into sentiment term matrix with numerical values. MapReduce programming model can work efficiently with these large numerical datasets.

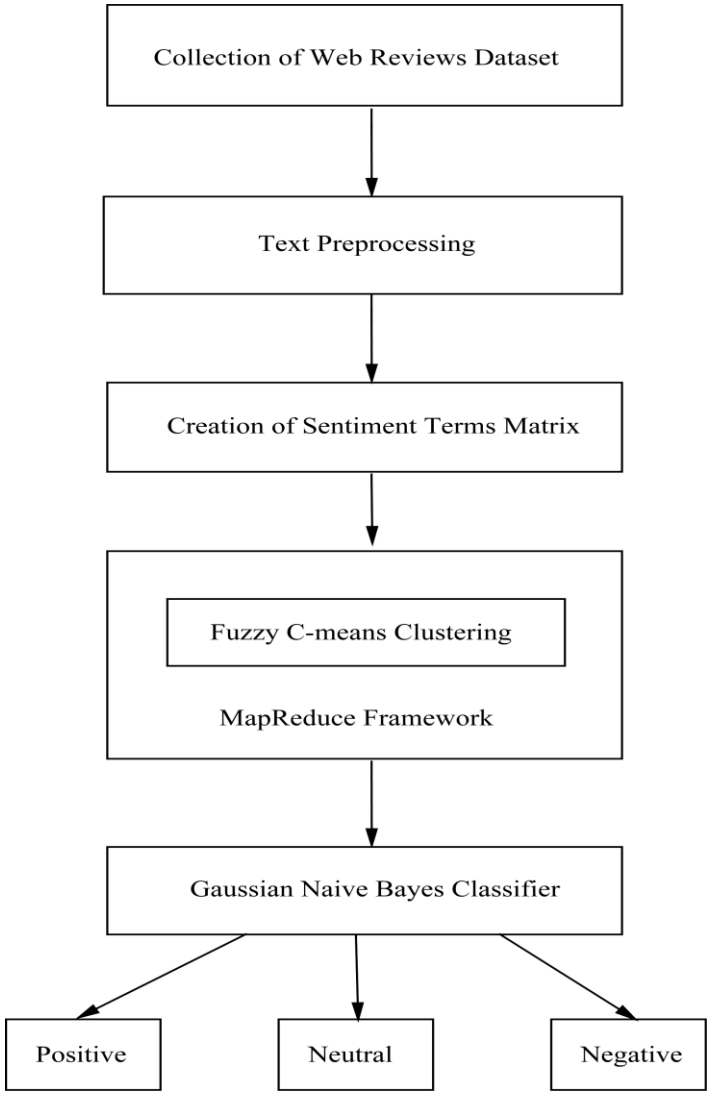

Figure 5.1: Block Diagram of the Proposed Model for Sentiment Analysis

ISSN: 2277-5528
Impact Factor- 4.015

We propose a novel sentiment analysis method using fuzzy c-means clustering in MapReduce platform and Gaussian naive Bayes classifier. Figure 5.1 shows various steps involved in the proposed method of sentiment analysis.

CONCLUSION

In this, we proposed a novel sentiment classification method with MapReduce programming for large web data. Traditional classification methods are not suitable for large datasets. In this method, large datasets are divided into several clusters. After that, we datasets are assigned classes by assigning class for each corresponding cluster centers. Here, we have applied MapReduce framework to divide large datasets into clusters efficiently. That is why, the proposed method becomes more effective and efficient in term of processing time.

REFERENCES

[1]J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. H. Byers, "Big data: The next frontier for innovation, competition, and productivity," 2011.

[2] S. Madden, "From databases to big data," *IEEE Internet Computing*, vol. 16, no. 3, pp. 4–6, 2012.

[3] B. Gerhardt, K. Griffin, and R. Klemann, "Unlocking value in the fragmented world of big data analytics," *Cisco Internet Business Solutions Group*, 2012.

[4] S. LaValle, E. Lesser, R. Shockley, M. S. Hopkins, and N. Kruschwitz, "Big data, analytics and the path from insights to value," *MIT sloan management review*, vol. 52, no. 2, p. 21, 2011.

[5] A. Weichselbraun, S. Gindl, and A. Scharl, "Enriching semantic knowledge bases for opinion mining in big data applications," *Knowledge-Based Systems*, vol. 69, pp. 78–85, 2014.

[6] B. Liu, "Opinion mining and sentiment analysis," in Web Data Ming, PP. 459-526, springer, 2011. *Web Data Mining*, pp. 459–526, Springer, 2011.

[7] D. Bhattacharyya, S. Biswas, and T.-h. Kim, "A review on natural language processing in opinion mining," *International Journal of Smart Home*, vol. 4, no. 2, pp. 31–38, 2010.

[8] F. Provost and T. Fawcett, "Data science and its relationship to big data and data-driven decision making," *Big Data*, vol. 1, no. 1, pp. 51–59, 2013.

[9] C. Curme, T. Preis, H. E. Stanley, and H. S. Moat, "Quantifying the se- mantics of search behavior before stock market moves," *Proceedings of the National Academy of Sciences*, vol. 111, no. 32, pp. 11600–11605, 2014.

ISSN: 2277-5528
Impact Factor- 4.015

[10] E. Cambria, B. Schuller, Y. Xia, and C. Havasi, "New avenues in opinion mining and sentiment analysis," *IEEE Intelligent Systems*, vol. 28, no. 2, pp. 15–21, 2013.

[11] E. Kontopoulos, C. Berberidis, T. Dergiades, and N. Bassiliades, "Ontology- based sentiment analysis of twitter posts," *Expert systems with applications*, vol. 40, no. 10, pp. 4065–4074, 2013.