

A Review of Various Sentiment Analysis Techniques in Map-Reduce Environment

Behjat U Nisa

M.Tech Scholar, Department of CSE, Swami Devi Dyal Institute of Engineering and Technology, Haryana, India.

Abstract: Content development within the Internet lately has made a considerable volume of statistics on hand. This data is exhibited in various codecs, as an instance, posts, information articles, remarks, and surveys. Particularly in the car, hardware and movie segments, customers have composed audits approximately gadgets or their features. By collecting and analysing those surveys, new clients find out others' opinion approximately numerous highlights of the object. They can evaluation the items with every different to locate the satisfactory one that addresses their issues. In this way, makers will tackle the mentioned troubles and utilize the commercial enterprise knowledge in the back of the exam for destiny speculations. Thus, analysing the sentiments of the users become very important in optimizing the product or service. By extracting user estimation for decision making plus market analysis sentiment analysis is one of fast increasing in addition to reliable tool nowadays. In this paper, review of various sentiment analysis techniques has been presented.

Keywords: Big data, Sentiment analysis, Map reduce, Data mining

1. INTRODUCTION

Enormous amounts of data are being produced in structured, unstructured, and semi-structured type. The conventional techniques for data dealing out is not effective nor enough for processing this

massive amount of data which requiring more powerful computational analysis strategies. Big Data analytics along with storage propose scalable solutions moreover new methods to grip that massive composite amounts of data. The term “Big Data” has newly been functional to datasets that produce so large that they turn into uncomfortable to work with by means of usual database management systems. They are data sets whose size is away from the capability of normally used software tools in addition to storage systems to detain, store, manage, as well as process the data within a tolerable elapsed time. Big data is data whose scale, allocation, diversity, and/or timeliness entail the utilization of original technical architectures, analytics, in addition to tools in order to allow insights that undo new sources of business value. Three main type distinguish big data: volume, variety, along with velocity, or the three V's. [1]

Big data are realistic valuable in a variety of areas. Location tracking is useful in logistic companies to ease risks in transport, speed as well as dependability of delivery Data analysed from Facebook, Twitter and Google monitor behaviour and transactions that advertisers use to run embattled campaigns. Efficiency of (expensive) medication can be considered along with endemic outbreaks can be forecasted in an premature stage [2]. Big data analytics is where superior analytic

techniques are functional on big data sets. Analytics based on large data samples states in addition to leverages business modify. Conversely, the outsized the set of data, the additional hard it becomes to supervise [3].

Individuals are often communicating, discussing plus sharing in sequence through internet. Due to these reasons internet is one of the necessary part of human life. The information in it occupies a wide range of areas such as academic information, feedback or opinion about goods comments about social issues. Sentimental analysis is the process of discovering sentiment like positive or negative from a text data.[4] Sentimental analysis is the analytical knowledge of people's view, their view along with emotions. Sentimental Analysis of tweets from the twitter is a bit dissimilar from the ordinary text processing [5]. Sentiment analysis, or opinion mining, aims at user's attitude along with opinions by investigating, analyzing in addition to extracting discriminatory texts linking users' opinions, preferences along with sentiment. Natural human language hold two varieties of information: objective information about reality along with critical information with human biased sentiment. The rapid development of forum, BBS, Blog, and review websites contributes to the exploding amount of such critical information, which reflects users' attitudes, viewpoints. Sentiment analysis has grown into a hot investigate field in natural language processing. Sentiment analysis first created a sensation at Text Retrieval Conference (TREC), and since 2006 related evaluation errands have appeared every year. The annual multi-lingual opinion analysis task (MOAT) also started in 2006, counting standard libraries in

three languages, Chinese, English and Japanese [6].

Along with lexicon based and linguistic method, machine learning will be careful as one of the largely used approach in sentiment classification.

Classification Techniques: A text classification matter required to be determined is sentiment examination. Machine learning method plus Lexicon based method are the two broader categorizations of these classification method. **Machine Learning Approaches** The text is determined into classes by the machine learning based method with the help of classification techniques. The two broader categorizations of these machine learning techniques are:

- **Unsupervised learning:** There is no group concerned in addition to the targets are not given by them at all. Thus, clustering is measured to be a significant thing here.
- **Supervised learning:** The labelled dataset is required for making this method. When the classification method is to be planned, the labels are provided to the representation.

Naive Bayes (NB), Maximum Entropy (ME), as well as Support Vector machines (SVM) are little between the extensively used machines learning method for sentiment classification.

- **NAIVE BAYES CLASSIFIER:** The significant information of kind are used in feature vector through Naive Bayes classifier. As these sort are independent equally, analyzing them wholly is essential.
- **SUPPORT VECTOR MACHINE CLASSIFIER:** Huge margin is utilized for classification through SVM classifier. A

hyper plane is utilized to distinguish the tweets

- **MAXIMUM ENTROPY CLASSIFIER:** With respect to the relationship amongst features, no assumptions are careful in maximum entropy classifier. The conditional allocation of class label is approximate by maximizing the entropy of arrangement through this classifier.
- **ENSEMBLE CLASSIFIER:** Dissimilar type of ensemble classifiers are urbanized For making the best classification, all the sort of all the best classifiers are used in this classifier. Naive Bayes, Maximum entropy as well as SVM are the three approaches utilized by the base classifiers. The voting law is required for creating an ensemble classifier [7].

II. LITERATURE REVIEW

In this [8] authors proposed a fresh sentimental analysis method based on shared information to advance the competence of feature selection, which is dissimilar from the conventional sentimental analysis method, and a original weighting system is also utilized in the feature weighting procedure; (2) since ELM is a quick learning representation and has been productively functional in numerous investigation fields, Authors suggested a forecast representation which shared mutual information-based sentimental analysis with kernel-based ELM named as MISA-K-ELM. This representation has the payback of equally statistical sentimental analysis furthermore ELM, which can well equilibrium the needs of both prediction correctness plus prediction speed. The experiments were taken on HKEx 2001 stock market datasets to authenticate the presentation of

the presented MISA-K-ELM. The market historical price plus the market news are implemented in the MISA-K-ELM. For examination of the efficiency of MISA, the authors first evaluate the prediction correctness of ELM model with the help of MISA with ELM model using traditional sentimental analysis. Then, after that the authors presented MISA-K-ELM with existing state-of-the-art learning algorithms, like Back-Propagation Neural Network (BP-NN), and Support Vector Machine (SVM). Our experimental results show that (1) MISA model can help get higher prediction accuracy than traditional sentimental analysis models; (2) MISA-K-ELM and MISA-SVM have a superior prediction accuracy than MISA-BP-NN as well as MISA-B-ELM; (3) both MISA-K-ELM plus MISA-B-ELM can get faster prediction speed than MISA-SVM as well as MISA-BP-NN in most cases; (4) in mainly cases, MISA-KELM has higher forecast accuracy than the other three methods.

Having a high quantity of first-class research, the major issues of sentiment analysis have been solved except sentiment detection from the noun. Sometimes noun has sentiment base on the application, in this research [9], authors have solved this with a probable explanation for dual sentimental words. With alive technology, authors used some rules as well as made a model to do this.

The work here in [10] show a sentimental image leading graph subject process that can notice the topic from the heterogenous data along with mine the sentiment of each matter In details, authors design a topic model to relocate both the low-level visual modality along with the high-level text modality into a semantic various, in addition to get

better the discriminative power of CNN feature by jointly optimizing the output of both convolutional layer and fully-coupled layer. In addition, since the sentimental shock is extremely important for sympathetic the intrinsic meaning of subject, authors set up a semantic score of slanted sentences to compute the response on the support of the background sentence structure. The experiments on the public cross-modality benchmark show the promising performance of our model. So this method by means of AI knowledge will assist the intellectualization of 5G.

A hybrid sentimental thing recognition model (HSERM) has been calculated in [11]. Using 100 million together communication from Twitter, the hashtag is supposed as the label for sentimental classification. In this moment, features as emoji furthermore N-grams have been found in addition to confidential the composed topic communication into four dissimilar sentiment parts based on the circumplex sentimental representation. At last, machine learning methodology are utilized to categorize the sentimental data set, in addition to an 89 % exact result has been gained. Additional, entities that are behind emotions could be achieved with the assist of SENNA deep learning representation.

The authors in [12] states that when contrasted with the substitute between sentimentally expensive gifts plus gifts with shallow attributes that contest the preferences of the recipient, givers give the latter much more often than recipients would prefer to receive such gifts. This irregularity show to be determined by givers sentiment comparatively certain that favourite-matching gifts will be well-liked by recipients, but relatively

uncertain that the same is true for sentimentally valuable gifts. Three studies expose this gift-giving mismatch and confirm the planned mechanism across a range of gift-giving occasions along with giver-receiver relationship things. The part of these result to the gift-giving depiction, as well as commands for outlook investigate, are shared.

Authors in [13] offered a Pattern based subject discovery in addition to Analysis System (PTDAS) on Weibo, a Twitter-like stage in China. As one of the description mechanism of the absolute organization, an FP growth-like algorithm is working to extract cosine motivating patterns from a rest of tweets, along with then review them as topics. Especially in charge to realize topics in real-time, authors parallelize the algorithm on Spark for efficient mining. Along with pattern-based topic detection, authors also represented some analytic methods, counting both topic developing analysis in addition to sentimental analysis. Widespread experiments on the real-world data set display the effectiveness as well as competence of PTDAS.

In this work [14], authors begin a gated consistent neural network with sentimental kindred (GRNN-SR) to incarcerate the sentimental relations' information from sentiment modifier environment as well as mock-up their belongings in texts. At each time measure, GRNN-SR separately encodes the information of sentiment polarity furthermore sentiment modifier context. The new sentiment inputs are personalized multiplicatively by the prior resolute sentiment modifier background before they are updated into current states of sentiment polarity, which is more effective than the approach of traditional GRNNs. The new grades

demonstrate that our representation not simply can detain sentimental relations but as well is a growth over state-of-the-art gated standard neural network baselines.

The work in this paper [15] presents a materialist perceptive of the contributions spectacles that situates them within an accepted habits of sentimentality full in making as well as imagining forms of global community through social practices of exchange. It draws on the feminist scholarship on sentimental cultures and their imbrication with social reform movements in addition to service capitalism to demonstrate how Band Aid can be silent as part of a popular society of sentimental exchange, in which scarcity relief images, stories, tears, money as well as supplies were approved along in sentimental connections that also involved sentimental stories and personalized commodities and capital such as wedding rings, household furniture with allowances. The movement of feeling, concretized in the switch of goods as well as money, complete the shared hope of global community, imagined through the provisions of intimate love and familial gift exchange. When joint with local, national as well as international product markets that authorized information, goods plus images to travel among strangers, global gift giving appeared to restore geopolitical alliances in addition to financial security with an open, barrier-free, affective economy of love and cooperation.

Researchers in [16] outlined a penalty-based purpose in generators to power each of them to construct diversified instincts of a specific sentiment label. In addition, the employment of multiple generators in addition to one multi-class discriminator can make each generator focus on

creating its own texts of a specific sentiment label precisely. Experimental results on a diversity of datasets display that the SentiGAN representation constantly plays several state-of-the-art text production scheme in the sentiment accuracy along with superiority of generated texts. In addition, experiments on conditional text generation tasks confirm that our C-SentiGAN model has high-quality forecast for specific text generation errands.

In this project [17] Naive Bayes classifier method is being utilized for the reason of classification. The dataset is utilized here is the amazon product re-examine dataset made from the UCI repository. The dataset contained about 600 records and every record is being studied using the Naive Bayes approach, this is a probabilistic approach and finally the result is made in a form of matrix. The correctness of the projected approach is ultimately considered with the facilitation of the uncertain matrix.

The sentimental analysis is studied on the data extracted from Twitter as well as Stock Twits in [18]. The data is studied to calculate the humour of user's comment. These comments are considered into four category which are happy, up, down plus rejected. The division index along with market data is given to an artificial neural network to forecast the results.

III. CONCLUSION

Data is made out of more than one resources and it's far ever-increasing. This has given upward thrust to the concept of Big Data. Big Data has the principle characteristics of quantity, speed and variety. Social network have obtained an upward surge and facts generated from them is attaining higher values day by day. Twitter is one of the

most used social community. There are tweets concerning many things inclusive of enterprise establishments, films, political events, educational institutions, clinical initiatives and so forth. These tweets mirror sentiment of the human beings regarding diverse topics. Companies will benefit immensely with the aid of getting this sentiment information. So we need to categorise the information based on sentiment. We get records from a social community and analyse it into Hadoop. Then we implement a type algorithm using MapReduce framework on the way to classify the statistics we've got gathered based on what sentiments are hidden in it.

REFERENCES

1. Elgendy, N., & Elragal, A. (2014). *Big Data Analytics: A Literature Review Paper. Lecture Notes in Computer Science*, 214–227. doi:10.1007/978-3-319-08976-8_16
2. Willems, S. M., Abeln, S., Feenstra, K. A., de Bree, R., van der Poel, E. F., Baatenburg de Jong, R. J., ... van den Brekel, M. W. M. (2019). *The potential use of big data in oncology. Oral Oncology*, 98, 8–12. doi:10.1016/j.oraloncology.2019.09.003
3. Russom, P.: Big Data Analytics. In: TDWI Best Practices Report, pp. 1–40 (2011)
4. Shahana, P. H., & Omman, B. (2015). Evaluation of Features on Sentimental Analysis. *Procedia Computer Science*, 46, 1585–1592. doi:10.1016/j.procs.2015.02.088
5. Kisan, H. S., Kisan, H. A., & Suresh, A. P. (2016). Collective intelligence & sentimental analysis of twitter data by using StanfordNLP libraries with software as a service (SaaS). 2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC). doi:10.1109/iccic.2016.7919697
6. Nanli, Z., Ping, Z., Weiguo, L., & Meng, C. (2012). *Sentiment analysis: A literature review. 2012 International Symposium on Management of Technology (ISMOT)*. doi:10.1109/ismot.2012.6679538
7. Tyagi, P., & Tripathi, R. C. (2019). A Review towards the Sentiment Analysis Techniques for the Analysis of Twitter Data. *SSRN Electronic Journal*. doi:10.2139/ssrn.3349569
8. Wang, F., Zhang, Y., Rao, Q., Li, K., & Zhang, H. (2016). *Exploring mutual information-based sentimental analysis with kernel-based extreme learning machine for stock prediction. Soft Computing*, 21(12), 3193–3205. doi:10.1007/s00500-015-2003-z
9. Chaki, P. K., Hossain, I., Chanda, P. R., & Anirban, S. (2017). *An Aspect of Sentiment Analysis: Sentimental Noun with Dual Sentimental Words Analysis. 2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC)*. doi:10.1109/ctceec.2017.8455159
10. , Liang Li, Wenchao Li, Jiyong Zhang & Chenggang Yan , Enabling 5G: sentimental image dominant graph topic model for cross-modality topic detection, *Wireless Networks* (2019)
11. hibo Wang, Xiaohui Cui, Lu Gao, Qi Yin, Lei Ke & Shurong Zhang , A hybrid model of sentimental entity recognition on mobile social media, *EURASIP Journal on Wireless Communications and Networking* volume 2016,

12. JulianGivi,JeffGalak, Sentimental value and gift giving: Givers' fears of getting it wrong prevents them from getting it right, *Journal of Consumer Psychology*,Volume 27, Issue 4, October 2017, Pages 473-479

13. Zhang, L., Wu, Z., Bu, Z., Jiang, Y., & Cao, J. (2017). *A pattern-based topic detection and analysis system on Chinese tweets. Journal of Computational Science*. doi:10.1016/j.jocs.2017.08.016

14. ChaotaoChen,RunZhuo,JiangtaoRen, Gated recurrent neural network with sentimental relations for sentiment classification, *Information Sciences* Volume 502, October 2019, Pages 268-278

15. Cheryl Lousley, 'With Love from Band Aid': Sentimental exchange, affective economies, and popular globalism, *Emotion, Space and Society* Volume 10, February 2014, Pages 7-17

16. K.Wang ,X.Wan' Automatic generation of sentimental texts via mixture adversarial networks, *Artificial Intelligence*,Volume 275, October 2019, Pages 540-558

17. Prabha, B Subbulakshmi, Sentimental Analysis using Naive Bayes Classifier, Publisher: IEEE

18. Sunil Kumar Khatri ; Ayush Srivastava, Using sentimental analysis in prediction of stock market investment, Publisher: IEEE