

An Overview of Sentiment Analysis in Bigdata Environment

#¹Kishor Kumar Gajula, #²Dr.R.Kamalakar

¹Ph.D Scholar, Department of CSE, Shri JTT University, Rajasthan, India

²Department of CSE, Shri JTT University, Rajasthan, India

ABSTRACT

Social network gives clients a stage to discuss successfully with companions, family, and partners, and furthermore gives them a stage to discuss their top pick (and minimum most loved brands). This "unstructured" discussion can give organizations significant knowledge into how shoppers see their image, and enable them to effectively settle on business choices to keep up their picture. Quick increment in the volume of assumption rich online networking on the web has brought about an expanded enthusiasm among scientists with respect to Sentimental Analysis and Opinion Mining. Be that as it may, with so much online networking accessible on the web, Sentiment Analysis is currently considered as a Big Data undertaking. The primary focal point of the exploration was to discover such a system, to the point that can effectively perform Sentiment Analysis on Big Data sets. In this paper Sentiment Analysis was performed on an expansive informational index of tweets utilizing Hadoop and the execution of the method was estimated in type of speed and precision. The trial result demonstrates that the strategy displays great effectiveness in taking care of enormous assessment informational indexes.

Keywords : Sentiment Analysis Approaches, Big Data Analytics, Hadoop, Lexicon, Machine learning.

I. INTRODUCTION

Big Data is slanting examination region in Computer Science and Sentiment Analysis is a standout amongst the most critical piece of this exploration territory. Enormous Data is considered as extensive measure of information which can be discovered effectively on web, Social media, remote detecting information and restorative records and so forth in type of organized, semi-organized or unstructured information and we can utilize these information for Sentiment Analysis.

Nostalgic Analysis is going to get the genuine voice of individuals towards particular item, administrations, association, motion pictures, news, occasions, issues and their attributes[1]. Feeling Analysis incorporates branches of Computer Science like Natural Language Processing, Machine Learning, Text Mining and

Information Theory and Coding. By utilizing approaches, strategies, procedures and models of characterized branches, we can arrange our unstructured information which might be as news articles, web journals, tweets, film surveys, item audits and so on into positive, negative or unbiased estimation as indicated by the notion communicated in them.

Sentiment Analysis is done on three levels [1].

1. Document level
2. Sentence level
3. Aspect or Entity level

1) Document Level Sentiment Analysis is performed for the entire record and afterward choose whether the report express positive or negative slant [1].

2) Entity or Aspect Level Sentiment Analysis performs fine-grained examination. The objective of element or angle level Sentiment Analysis is to discover assumption on substances or potentially part of those elements. For instance consider an announcement "My HTC Wildfire S telephone has great picture quality yet it has low telephone memory stockpiling." so opinion on HTC's camera and show quality is certain however the slant on its telephone memory stockpiling is negative.

3) Sentence level Sentiment Analysis is identified with discover estimation frame sentences whether each sentence communicated a positive, negative or unbiased slant Sentence level Sentiment Analysis is firmly identified with subjectivity arrangement. A large number of the announcements about substances are truthful in nature but then regardless they convey assessment. Current Sentiment Analysis approaches express the feeling of subjective proclamations and disregard such target articulations that convey opinion [1].

The decrease in the cost of both stockpiling and processing power is one of the primary factors that prompted the blasting of huge information. Preceding this time, organizations settled on choices in view of value-based information put away in social databases, while other possibly vital assets in non-conventional and less organized information are overlooked. The technique to use enormous information ranges from developing current venture information design to consolidating huge information and conveying business esteem. Huge information empowers organizations to make focused on, realtime choices that expansion piece of the overall industry. Enormous information is described by the volume, speed, veracity, assortment, esteem and unpredictability of information. All things considered, the fitting instruments are expected to gain, arrange and get an incentive from huge information to

underwrite one concealed connections and to recognize new bits of knowledge.

The refining and examination of enormous information can encourage a more intensive and quick comprehension of ventures, which can prompt improved efficiency, more grounded focused position and more noteworthy advancement. As per the potential that huge information offers, an expanding number of studies have concentrated on procedures for examining new and different computerized information streams to uncover new wellsprings of monetary esteem, give new bits of knowledge into client conduct and distinguish advertise inclines ahead of time (Bernabé-Moreno et al., 2015; Harrigan et al., 2014; Malthouse et al., 2013).

Sentiment Analysis (SA) is one of the principle plan in enormous information that spotlights on different approaches to break down huge information to distinguish examples and connections, make educated expectations, convey noteworthy knowledge and pick up business understanding from this consistent deluge of data. SA is regularly used to dissect individuals' slants, assessments, examinations, dispositions, assessments and feelings towards such elements as associations, items, administrations, people, subjects, issues, occasions and their qualities, as exhibited online by means of content, video and different methods for correspondence. These interchanges can fall into three general classifications, in particular positive, unbiased and negative. These classifications include numerous names and marginally extraordinary assignments, for example, supposition mining, conclusion extraction, estimation mining, subjectivity investigation, client dissension, influence examination, feeling investigation, audit mining and survey examination.

II. RELATED WORK

Sentiment Analysis Issues in Big Data

In spite of the fact that SA is one of the fundamental motivation in enormous information, no known work has talked about whether SA approaches are reasonable for huge information framework. This area centers around this angle by beginning with a dialog of the general situation and difficulties of huge information examination, trailed by a work about the general SA structure.

Issues in Big Data Analysis

Enormous information is related with the 5V issues, to be specific volume, speed, veracity, assortment, esteem and instability of information. The huge sum and high volume of information are the primary attributes of enormous information and are, truth be told, the fundamental motivation behind why the term huge information was instituted. Having a nearby connection to volume is the speed factor, which is identified with the procedure by which continuous gushing information are being produced through sensors and along these lines should be examined. At the point when a gigantic volume of consistently produced information exists, the veracity issue emerges to address the vulnerability, legitimacy, chaos and dependability of the information. The quality and precision of the information are additionally viewed as, given that these variables are applicable to the assortment issue in light of the fact that different arrangements and styles of information are produced. Next is the issue on the estimation of the information, which ought to be misused instantly. This choice is related with the unpredictability or span in which the information are regarded legitimate and should in this way be put away.

The above certainties show that huge information brings new information composes and capacity components as well as new kinds of examination. Enormous information examination is a continuum

and isn't a confined arrangement of exercises that include making "sense" out of huge volumes of changed information that, in their crude shape, do not have an information model to characterize what every component implies with regards to the others. A few new issues ought to be considered while setting out on this new kind of examination; these issues incorporate revelation, emphasis, adaptable limit mining and forecast and choice administration (Asur and Huberman, 2010; Bravo-Marquez et al., 2014; Rao et al., 2014).

The disclosure issue is credited to the way that the estimation of the information is frequently concealed profound under the surface of the gathered dataset and must be resolved through an investigation procedure. Besides, the real connections inside the enormous measure of information are not generally known ahead of time. In this manner, revealing understanding is frequently an iterative procedure until the point when the appropriate responses are found. In any case, the nature of emphasis is identified with experimentation, to such an extent that it now and again leads down a way that ends up being a deadlock. An unavoidable issue identified with enormous information is the adaptable limit. In spite of the fact that distributed computing is misused for enormous information, the iterative idea of huge information examination requires the use of additional time and assets to take care of the current issues. This test is exacerbated by the way that huge information examination isn't a common highly contrasting choice. Recognizing, mining and anticipating how the different information components identify with each other are consistent issues. Choice administration is likewise considered regarding how the usage of every one of these activities can be computerized and improved.

Big Data Framework for Sentiment Analysis

SA for the most part centers around recognizing the feeling of the writer. The ways to deal with

accomplish this objective can be separated into two classifications, in particular substance particular and substance free. SA is firmly identified with conclusion mining, which is characterized as a quintuple feeling comprising of an objective question, highlight of the protest, an opinion estimation of the opinion, an opinion holder and the time when the opinion is expressed (Sharef and Haghanikhameneh, 2014).

In spite of the fact that reviews on SA have advanced over 10 years, yet without accentuation on enormous information, a few stages give SA administrations to huge information clients inferable from its vicinity to online networking investigation (Batinca and Treleaven, 2014; Conejero et al., 2013; Sharef, 2014) Table 1 indicates cases of huge information apparatuses. Given the extensive volume of movement in web-based social networking, the initial phase in examining online networking is to comprehend the extent of information that should be gathered for investigation. Regularly, information can be constrained to certain hash labels, records and catchphrases.

Hadoop is helpful for pre-preparing information to distinguish full scale drifts or to discover chunks of data, for example, out-of-extend values. It empowers organizations to open potential incentive from new information utilizing cheap ware servers. Associations principally utilize Hadoop as an antecedent to cutting edge types of investigation. Hadoop is a famous decision for sifting, arranging, or pre-handling a lot of new information set up and refining such information to produce denser information that hypothetically contain more 'data'. Pre-preparing includes sifting new information sources to make them reasonable for extra examination in an information stockroom.

MapReduce empowers us to take unstructured information, change (delineate) information into

something significant and afterward total (lessen) the information for revealing. These means happen in parallel over all hubs in the Hadoop bunch. A basic case of MapReduce could delineate media presents on a rundown of words and tally their events. Such rundown is then decreased to a tally of the quantity of events of a word for every day (Nirmal and Amalarethnam, 2015b).

Once the significant information are put away in Hadoop, they can be stacked into a current venture Business Intelligence (BI) stage or broke down specifically utilizing intense self-benefit instruments, for example, PowerPivot and PowerView. Clients using SQL Server as their endeavor BI stage have an assortment of alternatives to get to their Hadoop information. These alternatives incorporate Sqoop, SQL Server Integration Services and Polybase.

Prophet has presented Oracle Advanced Analytics (OAA) to reveal concealed connections inside information by joining in-database calculations and open-source R calculations, which are available by means of SQL and R dialects. OAA joins elite information mining capacities with the open-source R dialect to empower prescient examination, information mining, content mining, measurable investigation, progressed numerical calculations and intelligent illustrations all inside the database.

Amazon Web Services (AWS) uses the AWS Cloud Formation stack, which gives a content to gathering online networking messages, for example, tweets. The tweets are put away in Amazon S3 and a guide for each record is modified for use with the Amazon EMR. An Amazon EMR bunch is then made. This bunch utilizes a SA program inside the Python NLTK program, which is actualized with a Hadoop gushing employment, to order the information. The yield records are then assessed to screen the totaled supposition of the tweets.

III. SENTIMENT CLASSIFICATION THROUGH MACHINE LEARNING

The Machine Learning (ML) approach applies the ML calculation and utilizations etymological highlights with the point of improving the execution of the framework utilizing illustration information. The huge information structure, for example, Mahout and Pentaho contain library and modules for the ML approach which can be executed to play out the assumption arrangement. With regards to huge information examination, a client ought to decide the kind of calculation that would be connected for the current information and such calculation is executed through enormous information investigation instruments for particular critical thinking purposes, for example, prescient investigation.

Ordinarily, two arrangements of archives are required in a ML-based order. These records are the preparation and testing sets. A preparation set is utilized by the classifier to take in the record attributes, though a testing set is utilized to approve classifier execution. The content order strategies utilizing the ML approach can be partitioned into regulated and unsupervised learning techniques. The administered techniques utilize countless preparing records. The unsupervised strategies are utilized when these marked preparing reports are hard to discover.

The regulated techniques accomplish sensible adequacy yet are normally area particular and dialect ward and they require marked information, which is frequently work escalated. In the mean time, the unsupervised strategies have appeal on the grounds that openly accessible information are frequently unlabelled and therefore require strong arrangements. Accordingly, semi-administered learning has been presented and has pulled in extensive consideration in slant order. In unsupervised learning, it utilizes a

lot of unlabelled information alongside named information to fabricate better learning models.

Various ML procedures have been embraced to play out the grouping errand in SA (da Silva et al., 2014; Go et al., 2009; Xia et al., 2011). The most prevalent ML systems that have made awesome progress in content categorisation are Support Vector Machine (SVM), Naive Bayes (NB) and Maximum Entropy (ME). The other surely understood ML techniques in characteristic dialect preparing are K-Nearest neighbor, ID3, C5, centroid classifier, winnow classifier and the N-gram show.

Support Vector Machine (SVM)

SVM is a measurable arrangement strategy that uses the auxiliary hazard administration guideline from computational learning hypothesis. SVM has been turned out to be exceedingly successful strategy for conventional content categorisation contrasted and other ML procedures, for example, NB and ME (Khairnar and Kinikar, 2013). SVM likewise shows the best execution for feeling grouping (Prabowo and Thelwall, 2009; Tan and Zhang, 2008; Xia et al., 2011; Zhang et al., 2011c). At the point when joined with another procedure, for example, the oblige subject model, SVM is fit for extricating the understood angle in inspected records (Wang et al., 2013a).

Naive Bayes (NB)

NB classifier is a basic probabilistic classifier in light of Bayes' hypothesis. NB is especially reasonable for utilize when the information sources have high dimensionality. NB is a basic however powerful calculation that has been generally utilized as a part of record characterization works (Ding et al., 2007; Melville et al., 2009; Tan and Zhang, 2008; Ye et al., 2009; Zhang et al., 2011a). NB beats SVM when the quantity of highlights is little (Pang et al., 2002). The calculation additionally can be enhanced when joined with different strategies, for example, senti-dictionary (Kang et al., 2012; Sharef and Shafazand, 2014; Zhou et al., 2013). A straightforward NB

classifier can be improved to empower a superior comprehension of more confused models through more proper component determination and undesirable element (commotion) expulsion (Narayanan et al., 2013).

Maximum Entropy (ME)

ME is another ML classifier that has been demonstrated powerful in various common dialect applications. Dissimilar to NB, ME makes no suspicions about the connection between highlights, to such an extent that it may perform better when contingent autonomy suppositions are not met. At times, for example, for the situation where words in the dictionary can't express the conclusion propensity, the ME entropy arrangement show outflanks lexiconbased strategies as far as recognizing notion words in a sentence (Fei et al., 2010).

Quality/Sentiment Scoring

Estimation quality is computed by controlling the recurrence of coordinated dictionaries as indicated by extremity. Broadened thinks about in this test incorporate earlier extremity (Ghazi et al., 2014; Kouloumpis et al., 2011; Loia and Senatore, 2013), reliance rules (Poria et al., 2014b), refutation ID (Wiebe et al., 2005) and summarisation (Kontopoulos et al., 2013; Zhan et al., 2009; Zhuang et al., 2006). These methodologies, notwithstanding, are still a long way from having the capacity to gather the intellectual and emotional data related with common dialect, given that they for the most part depend on learning bases that are still excessively constrained, making it impossible to process message effectively at the sentence level. Also, such content examination granularity may even now be inadequate, given that a solitary sentence may contain diverse feelings about various features of a similar item or administration. To this end, idea level SA (Kontopoulos et al., 2013; Poria et al., 2014a) plans to go past an insignificant word-level investigation of content to give novel

ways to deal with assessment mining and SA that empower more productive entry from unstructured literary data to organized machine-processable information in any space.

IV. APPLICATIONS OF SENTIMENT ANALYSIS

Late research demonstrates that the quantity of individuals and organizations utilizing web-based social networking applications as a client relationship administration device has drastically expanded (Bagheri et al., 2013; Fuchs et al., 2014; Kaplan and Haenlein, 2010). It is the standard to see an expansive number of audits, protestations and compliments posted and shared seconds after another item is discharged.

Dissecting this data encourages organizations to suit this developing pattern with a specific end goal to accomplish some business esteems like expanding the quantity of clients; upgrading client unwaveringness, consumer loyalty and friends notoriety; and accomplishing higher deals and aggregate income (Batrinca and Treleaven, 2014; Bravo-Marquez et al., 2014; He et al., 2015). Then again, this data can be utilized by the clients as tributes by separating the qualities and shortcomings of the recognizable highlights of every item, and in addition finding the fulfillment levels of different clients of those items. Other than the advantages in business, an investigation of political pages gives data to political gatherings with respect to individuals' perspective of their projects. Social associations may look for individuals' feeling on current civil arguments or on issues like the following presidential competitor. This data can be acquired by breaking down the assumption introduction of remarks, the quantity of preferences, offers or remarks on posted themes.

Utilizations of SA go from open voice examination, swarm observation, client care and social insight based SA to misuse the public's online substance age

for breaking down sources of info, for example, pandemic spreading, feeling and reactions towards nearby occasions. SA that spotlights on microblogging is extremely regular since this is the primary source that taps people in general's voice. SA on microblogging information is all the more difficult contrasted with traditional messages, for example, reports audit, because of the length, rehashed utilization of some informal and atypical words and the quick advance of dialect variety use.

For small scale blogging SA, particularly Twitter, critical work (Cheong and Lee, 2010; Dodds and Harris, 2011; Khan et al., 2014; Kontopoulos et al., 2013; Sharef and Haghanikhameneh, 2014) has been done through boisterous marks, which are likewise called 'far off supervision'. Twitter is misused for the most part on the grounds that the idea of the information is printed, contrasted with the usage of Facebook (Eirinaki et al., 2012; Ortigosa et al., 2014) and YouTube (Cambria et al., 2011; Li and Wu, 2010). The interpersonal organization is additionally misused to distinguish the most compelling opinionators (Fukushima et al., 2008; Zhao et al., 2014) as a correspondence technique which is helpful amid races and fiascos. Full of feeling registering through SA encourages answers to inquiries, for example, 'What are the essential subjects that more than once include in client remarks?', 'What is the slant introduction of a particular sexual orientation about a particular post?' and 'What are the patterns of bliss and misery of the client after some time?' Emotions in content might be communicated unequivocally (for instance, emojis and dictionary) (Fukushima et al., 2008; Loia and Senatore, 2013; Ptaszynski et al., 2013) and additionally verifiably (Balahur et al., 2012; Lau et al., 2014; Wang et al., 2013b).

Full of feeling registering empowers organizations to think more about their clients (Bagheri et al., 2013) and is valuable for showcase expectation (Lassen et al.,

2014; Li and Li, 2013; Milea et al., 2012; Nassirtoussi et al., 2014; Zhang et al., 2009), helps with diagnosing patients' self-destructive levels (Desmet and Hoste, 2013; Pestian et al., 2010a; 2010b) and enables the related gatherings to check open discernment towards occasions (Loia and Senatore, 2013; Moreo et al., 2012). The progressions in emotional registering enable applications to detect and convey administrations custom-made to client needs, however issues, for example, protection should be watched.

SA has likewise been tried in multilingual points of view (Balahur et al., 2014; Denecke, 2008; Hogenboom et al., 2014; Lim and Kong, 2004; Yong et al., 2011) where the concentration was to determine the restrictions of dialect subordinate assumption dictionaries. A few methodologies exist in this investigation, for example, making an interpretation of content into a reference dialect in which an assessment dictionary is accessible before in this way breaking down the content and mapping assumption scores from a semantically empowered reference vocabulary to an objective vocabulary by crossing relations between dialect particular vocabularies. These standards have empowered numerous dialects, for example, Dutch (Hogenboom et al., 2014), Czech (Habernal et al., 2014), Malay (Saloot et al., 2014) and Arabic (Abdul-Mageed et al., 2014) to investigate the capability of SA.

V. CONCLUSION

Sentiment Analysis is being used for different applications and can be used for several others in future. It is evident that its applications will definitely expand to more areas and will continue to encourage more and more researches in the field. We have done an overview of some state-of-the-art solutions applied to sentiment classification and provided a new approach that scales to Big Data sets

efficiently. A scalable and practical lexicon based approach for extracting sentiments using emoticons and hash tags is introduced. Hadoop was used to classify Twitter data without need for any kind of training. Our approach performed extremely well in terms of both speed and accuracy while showing signs that it can be further scaled to much bigger data sets with similar, in fact better performance.

In this research, main focus was on performing Sentiment Analysis quickly so that Big Data sets can be handled efficiently. The work can be further expanded by introducing techniques that increase the accuracy by tackling problems like thwarted expressions and implicit sentiments which still needs to be resolved properly. Also as explained earlier, this work was implemented on a single node configuration and although it is expected that it will perform much better in a multimode enterprise level configuration, it is desirable to check its performance in such environment in future.

VI. REFERENCES

- [1]. Bing Liu, *Sentiment Analysis and Opinion Mining*, Morgan and Claypool Publishers, May 2012, p.18-19,27-28,44-45,47,90-101.
- [2]. Nitin Indurkha, Fred J. Damerau, *Handbook of Natural Language Processing*, Second Edition, CRC Press, 2010.
- [3]. Ronen Feldman, James Sanger, *The Text Mining Handbook-Advance Approaches in Analyzing Unstructured Data*, Cambridge University Press, 2007.
- [4]. Jiawei Han, Micheline Kamber, *Data Mining Concepts and Techniques*, Morgan Kaufmann Publications, 2006.
- [5]. Chihil Hung and Hao-kai Lin, "Using Objective Word in SentiWordNet to Improve Word-of-Mouth Sentiment Classification", IEEE Computer Society, P.47- 54, March-April 2013.
- [6]. Wikipedia article on supervised machine learning
http://en.m.wikipedia.org/wiki/Supervised_learning
- [7]. Jintao Mao and Jian Zhu, "Sentiment Classification based on Random Process", IEEE Computer Society, International Conference on Computer Science and Electronics Engineering, p.473-476, 2012.
- [8]. Sang-Hyun Cho and Hang-Bong Kang, "Text Sentiment Classification for SNS-based Marketing Using Domain Sentiment Dictionary", IEEE International Conference on Consumer Electronics (ICCE), p.717-718, 2012.
- [9]. Gautam Shroff, Lipika Dey and Puneet Agrawal, "Social Business Intelligence Using Big Data", CSI Communications, April 2013, p.11-16.
- [10]. Sentiment Analysis: Capturing Favorability Using Natural Language Processing Tetsuya Nasukawa IBM Research, Tokyo Research Laboratory Jeonghee Yi IBM Research, Almaden Research Center.
- [11]. ZHU Jian, XU Chen, WANG Han-shi, "Sentiment classification using the theory of ANNs", *The Journal of China Universities of Posts and Telecommunications*, July 2010, 17(Suppl.): 58-62. [16] Ziqiong Zhang, Qiang Ye, Zili Zhang, Yijun Li, "Sentiment classification of Internet restaurant reviews written in Cantonese", *Expert Systems with Applications* xxx (2011)
- [12]. Long-Sheng Chen, Cheng-Hsiang Liu, Hui -Ju Chiu, "A neural network based approach for sentiment classification in the blogosphere", *Journal of Informetrics* 5 (2011) 313-322.
- [13]. <http://twitter4j.org/en/index.html>
- [14]. *Building Machine Learning Algorithms on Hadoop for Bigdata* Asha T, Shravanthi U.M, Nagashree N, Monika M *International Journal*

- of Engineering and Technology Volume 3 No. 2, February, 2013
- [15]. Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll and Manfred Stede. 2011. Lexicon based methods for Sentiment Analysis. *Computational linguistics*, volume 37, number2, 267-307, MIT Press.
- [16]. Michael Wiegand, Alexandra Balahur, Benjamin Roth, Dietrich Klakow, Andres Montoyo. 2010. A survey on the role of negation in Sentiment Analysis. *Proceedings of the workshop on negation speculation in natural language processing 60-68*, Association for Computational Linguistics.
- [17]. Twitter Sentiment Classification using Distant Supervision Alec Go Stanford University Stanford, CA 94305 alecmgo@stanford.edu Richa Bhayani Stanford University Stanford, CA 94305 rbhayani@stanford.edu Lei Huang Stanford University.
- [18]. Kamps, Maarten Marx, Robert J. Mokken and Maarten De Rijke, "Using wordnet to measure semantic orientation of adjectives", *Proceedings of 4th International Conference on Language Resources and Evaluation*, pp. 1115-1118, Lisbon, Portugal, 2004.
- [19]. Andrea Esuli and Fabrizio Sebastiani, "Determining the semantic orientation of terms through gloss classification", *Proceedings of 14th ACM International Conference on Information and Knowledge Management*, pp. 617-624, Bremen, Germany, 2005.
- [20]. Ting-Chun Peng and Chia-Chun Shih, "An Unsupervised Snippet-based Sentiment Classification Method for Chinese Unknown Phrases without using Reference Word Pairs", 2010 IEEE/WIC/ACM International Conference on Web Intelligence and intelligent Agent Technology *JOURNAL OF COMPUTING*, VOLUME 2, ISSUE 8, AUGUST 2010, ISSN 2151-9617 .
- [21]. Prabu Palanisamy, Vineet Yadav, Harsha Elchuri, "Serendio: Simple and Practical lexicon based approach to Sentiment Analysis", Serendio Software Pvt Ltd, 2013.
- [22]. Peter Turney and Michael Littman. 2003. Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems* 21(4):315-346.
- [23]. Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll and Manfred Stede. 2011. Lexicon-based methods for sentiment analysis. *Computational linguistics*, volume 37, number2, 267-307, MIT Press.
- [24]. Albert Bifet and Eibe Frank. 2010. Sentiment knowledge discovery in twitter streaming data, *Discovery Science* 1-14, Springer.
- [25]. Michael Wiegand, Alexandra Balahur, Benjamin Roth, Dietrich Klakow, Andres Montoyo. 2010. A survey on the role of negation in sentiment analysis. *Proceedings of the workshop on negation and speculation in natural language processing 60-68*, Association for Computational Linguistics.
- [26]. Minqing Hu, Bing Liu. Mining and Summarizing Customer Reviews, Department of Computer Science, University of Illinois at Chicago, Research Track Paper.