Sentiment Analysis using Machine Learning for Business Intelligence

Saumya Chaturvedi
Ph.D. scholar
AKTU Lucknow, INDIA
saumyanmishra5@gmail.com

Vimal Mishra
Director
IERT Allahabad, INDIA
vimal.mishra.upte@gmail.com

Nitin Mishra
drnitinmishra10@gmail.com

Abstract— **this paper suggests use of sentiment analysis classification as an effective method for examining textual data coming from variety of resources on internet. Sentiment analysis is a method of data mining that evaluates textual data consuming machine learning techniques. Due to tremendous expanse of opinions of users, their reviews, feedbacks and suggestions available over the web resources, it is so much indispensable to discover, analyze and consolidate their views for enhanced decision making. Sentiment analysis presents an effective and efficient opinion of consumers in real time which can greatly affect the decision making process for business domain. We have seen an increment in level of activity during last ten year period and emphases on exploratory research approaches. We noticed that several procedures are inattentive from the pond of Business Intelligence research. We also recognized potential zones that requisite additional exploration.**

Keywords— ***sentiment analysis, Business Intelligence, Machine learning, Big Data***

I. INTRODUCTION

People now a days practice dissimilar styles of online forums for societal engagement as well as social media alternatives such as Twitter and Facebook. This collective media consumer engagement occurs in real time. These communications provides wide range of opportunities to business intelligence [14]. People of different countries, gender, class and race shares their experiences and opinions by using internet. The upsurge of social media, usage of mobile devices, and Internet of things embroils huge real-time data generation. More than 1200 Exabyte's of new data is transported each year from a variety of sources [1]. Eighty percentage of this data is unstructured and is hard to be stored, processed and analyzed with commonly used tools [1, 2]

Analytics allows making sense of big data by renovating into intelligent information [3]. The complete course of extraction from big data involves of two step procedure. First, data management and Second, analytics. Data management encompasses "process and supporting technologies to acquire and store data and to prepare and retrieve it for analysis" whereas analytics mentions "technologies used to analyze and acquire intelligence from big data" [4]. Researchers have been working since long back to improve information system that affords business intelligence. Business Intelligence includes a product and a process. The product is information that permits organizations to forecast the conduct of their competitor's, consumers, providers, technologies, acquisitions, marketplaces, goods and services, and general business atmosphere by an amount of conviction. The process includes methods that firms practice to ripen valuable information, or intellect, which in turn benefit officialdoms endure and flourish in the global economy. [5]

II. SENTIMENT ANALYSIS

A. What is Sentiment Analysis?

Sentiments are feelings that reflect attitude, emotions and opinion. Sentiment Analysis can be classified as computation study of opinions, evaluations, attitudes, subjectivity, and views expressed in text. An emotive valuation of a condition is a general assessment of that state that can be either positive or negative depending on bodily or mental responses. The importance of sentiments is not new to business. To a business process customer responses are indirect motivators of purchase behavior. Sentiment Analysis addresses the issues by systematically collecting and analyzing online sentiments from variety of sources and from a very huge sample of customer in real time. The phases of sentiment analysis process are shown in Fig-1.

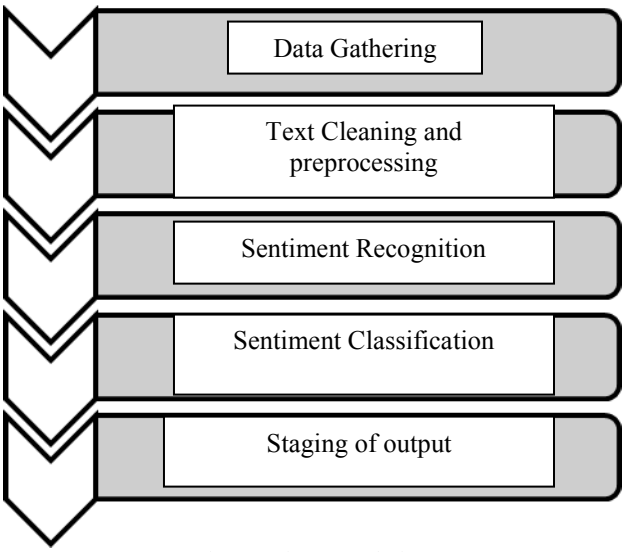

Fig-1 Sentiment Analysis Process

978-1-5386-0814-2/17/$31.00 ©2017 IEEE

Sentiment analysis is a methodical examination of online expressions. Precisely, sentiment analysis emphases on weighing attitudes and opinions on a topic of interest using machine learning techniques. The description of sentiment analysis in data mining can be pronounced from two standpoints: functional and operational. The functional aspects focus on practical uses of the technique. For illustration, sentiment analysis is described as a process that categorizes a body of textual information to determine feelings, attitude and emotions towards a particular issue or object [6]. The definition points to the way sentiment analysis works and describes the outcome of polar classification.

Another facet of sentiment analysis in data mining focuses on the operations of the technique as a sub-field of computational linguistics. Sentiment analysis is described as automated subjectivity analysis comparable to opinion mining and assessment withdrawal which focuses on extracting and classifying texts with machine language and computer programming [7]. Despite the differences in both perspectives, the general narrative is the same. In other words, sentiment analysis is a data mining technique that practices natural language processing, computational linguistic and text analytics to recognize and excerpt content of interest from a body of textual data.

B. Business Process Decision

Business process can be defines as a collection of interrelated, organized activities or tasks that yield specific service or product that serves certain goal for a particular customer or customers [18].

In today's era Business is rapidly changing to fulfill the customer's needs and to meet the requirement of changing environment. So these changes need to be implemented in business process (BP) and information systems (IS). So for maintaining these changes and take decisions accordingly need a new approach of dynamic business process modeling and predictive intelligence [19, 20].

Researchers have been working since long back to improve information system that affords business intelligence. Business Intelligence includes a process and a product. The process includes methods that organizations practice to ripen useful information, or intellect, that can benefit organizations survive and flourish in the global economy. The product is information that permits organizations to forecast the conduct of their competitor's, suppliers, customers, technologies, acquisitions, markets, products and services, and general business environment by an amount of conviction [5].

III. MACHINE LEARNING APPROACHES

The goal of machine learning is to cultivate an algorithm that optimizes performance of a model using historical experience as training data. In traditional approaches, we provide data (input) and program to the computer (machine) which processes it and gives us result (output). But, in case of machine learning approach we provide data (input) and result (output) and computer (machine) itself deduce a program based on input output relation [15]. Machine learning paradigm is shown in Fig-2.

Based on this devices program (algorithm) model is build and applied to unseen data for producing or predicting results.

Machine learning algorithms can be classified into three broad categories known as Supervised, Unsupervised and semi-supervised. Supervised learning is inductive in nature mean training data incudes the required output. Supervised learning is used in classification and regression. In unsupervised learning training data does not include desired output. Unsupervised learning is used in clustering and pattern recognition [17].

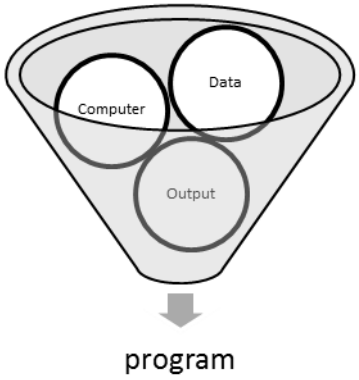

Fig-2 Machine Learning Paradigm

A computational learning model should be clear about the following aspects as shown in Fig-3.

Classification task itself is combination of different tasks such as

- Data preprocessing which includes cleaning of data.
- Feature selection and/or feature reduction which attempts to reduce the dimensionality that is, reduction in number of features.
- Representation includes the way in which objects are going to be learn.
- Classification is the actual categorization after all these steps.
- Post processing includes the presentation of output.

The machine learning approach solves classification problem in steps as shown in Fig-4:

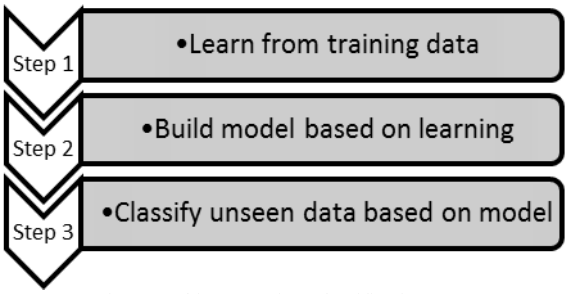

Fig-4 Machine Learning Classification steps

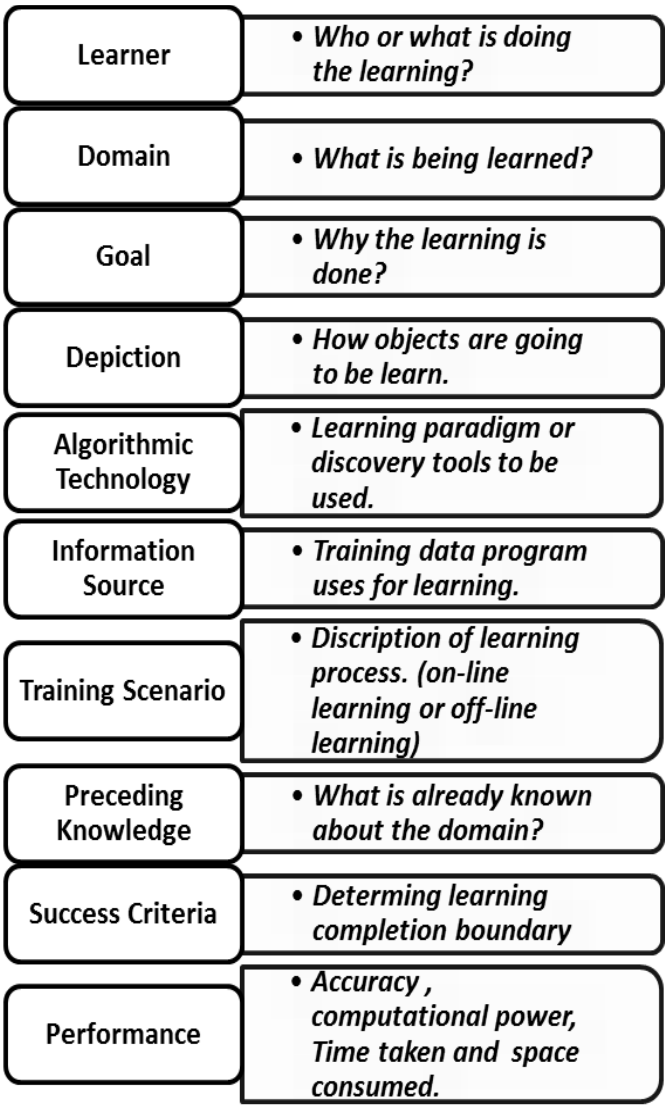

Fig-3 Computational Learning Model

IV. MACHINE LEARNING SOLUTIONS

Sentiment analysis of natural language texts is a huge and emergent field which can be exploit for every business firm in order to make decision points for their services and products [21]. Sentiment analysis or Opinion Mining is the computational handling of opinions, sentiments and subjectivity of text. Sentiment analysis is a Natural Language Processing and Information Extraction task that targets to obtain writer's feelings articulated in positive sense or negative sense, interrogations and appeals, by evaluating a large numbers of documents.

Transforming a portion of written data to a feature vector is the elementary stage in any data compelled methodology to Sentiment analysis. Another well-thought-out essential part in traditional Information Retrieval and Text Classification task is term frequency. But Pang-Lee [8] established that presence of term is more significant to Sentiment analysis rather than term frequency. That is, binary-valued feature vectors in which the entries merely indicate whether a term occurs (value 1) or not (value 0).It also reported that unigrams outperform bigrams when classifying movie reviews by sentiment polarity. As a result, the sentiment analysis research from the determination of the semantic orientation of the terms. Determining semantic orientation of words Hatzivassiloglou and McKeown [10] hypothesize that adjectives separated by "and" have the same polarity, while those separated by "but" have opposite polarity .Starting with small seed lists, this information is used to group adjectives into two clusters such that maximum constraints are satisfied.

Sentiment classification is a fresh sub discipline of text classification which is concerned not with the topic a document is about, but with the opinion it expresses. Functional to the extraction of opinions from text is the determination of the orientation of "subjective" terms contained in text, i.e. the determination of whether a term that carries opinionated content has a positive or a negative connotation [11]. Esuli and Sebastiani proposed new way for defining the alignment of subjective terms. The method is based on the quantitative analysis of the glosses of such terms that is the meanings that these terms are given in online dictionaries and on the use of the resulting term representations for semi-supervised term classification. Sentiment classification can be divided into several specific subtasks: determining subjectivity, determining orientation, determining the strength of orientation [11]. Esuli and Sebastiani [12] described SENTIWORDNET, which is a lexical resource in which each WordNet synset is associated with three numerical scores, i.e., Obj(s), Pos(s), and Neg(s), thus describing how objective, positive, and negative the terms contained in the synset.

Traditionally, sentiment classification can be regarded as a binary-classification task [8, 9].Dave, Lawrence, Pennock [9] use structured reviews for testing and training, identifying appropriate features and scoring methods from information retrieval for determining whether reviews are positive or negative.. These results perform as well as traditional machine learning method then use the classifier to identify and classify review sentences from the web, where classification is more difficult. Various supervised or data-driven techniques to Sentiment analysis like Naïve Byes, Maximum Entropy and SVM. Pang Lee [8] compared the performance of Naïve Bayes, Maximum Entropy and Support Vector Machines in Sentiment analysis on different features like considering only unigrams, bigrams, combination of both, incorporating parts of speech and position information, taking only adjectives etc. The result observance demonstrated in Fig-5.

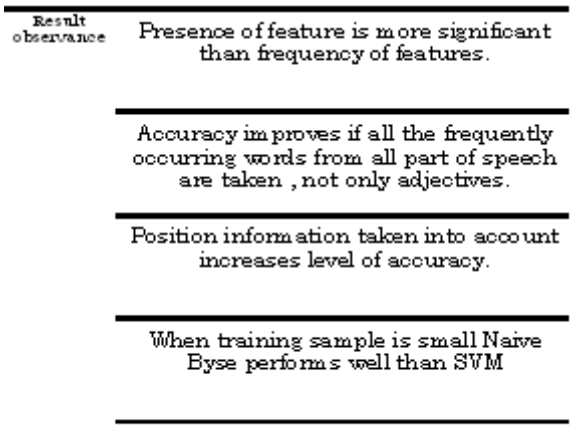

Fig-5 Comparison of performance of Naïve Bayes, Maximum Entropy and Support Vector Machines in Sentiment analysis on different features

According to Sonosy et al. [22] for understanding the behavior of business trends necessitate acquisition of hefty amount of data from sundry sources. Location grounded social Networks can offer such bulky amount of data that could be utilized for analysis to comprehend the business behaviors. Data collected is employed to predict business gatherings. Prediction is fulfilled consuming spatial regression model machine learning technique. Authors' scrutinized spatial regression models over and done with qualified study in order to find data set features relationships for business behavior prediction.

Singh, Kushwaha & Vyas [23] suggested that sentimental analysis plays very important role in business intelligence in order to know consumer feedback and behavior which turns out in better prediction and decision making. In order to do so they have combined two techniques NLP (Natural Language Processing) and Machine learning approach.

According to Chen J., Huang C., & Cheng C [24] machine learning approaches can be employed over monitoring system for business support system for estimating performance parameters and sometimes to know the causes of performance degradation. Though it is not always useful for some critical business support systems but it can sped up the trouble shooting and can be ready some time before for the emergency breakout. They have used Random Forest machine learning model for predicting healthy state of a system before the emergency with an hour with average 14 points error and with some spare time for trouble-shooting.

Ghiasi, Zimbra & Lee [25] discussed valuable feedbacks to firms about their brand and products from social media statement on twitter. In their work they focused on challenges accompanying with distinctive characteristics of the Twitter language and brand related tweet sentiment class distribution. They used SVM supervised learning model for classification of twitter sentiments related to two distinctive brands.

Social media can be recycled to expedite collaboration between people and can provide a substantial, unparalleled platform for extensive involvement of citizens in the governance measures. According to Kumar & Joshi [27] an ontology-based analytics on social media illustrates a brainy governance model in which sentiments can be excavated for dig up views of citizens towards government rules and regulations, policies, practices and monitoring performance.

V. ESTIMATION OF SENTIMENT ANALYSIS

The enactment of sentiment analysis for classification can be measured with four catalogs. They are named as Accuracy, Precision; Recall and F score [13]. The communal way of figuring these indexes is grounded on Confusion matrix. The challenge is that how to choose effective machine learning algorithm that fits to a particular domain with specific performance criteria and available resources.

VI. CONCLUSION

Some of the machine learning techniques has been discussed like Naïve Byes and Support Vector Machines. Recent development in sentiment analysis is also discussed. How this sentiment analysis can be exploit for business process decision. Challenges of big data are also discussed. This paper introduced and surveyed the field of sentiment analysis and their application in business environment. Sentiment analysis is a challenging task and can greatly affect decision points in any domain.

Though there are many of machine learning approaches that can be applied to some business domain, but still there is no generalized solution for which technique is best suitable. People in different organizations try one type of approach and then follow it again and again. There is no standard approach for business prediction to be followed. So there is need to dug up deep for some generalized standard machine learning approaches for different business domains.

References

[1].L. Issacs, "Rolling the Dice with Predictive Coding: Leveraging Analytics Technology for Information Governance", Information Management (47:1), 2013, pp. 22- 26.

[2].P. Malik, "Governing Big Data: Principles and Practices", IBM Journal of Research and Development (57 :¾), 2013, pp. 1-13.

[3].R.M. Hogarth, E. Soyer, "Using Simulated Experience to Make Sense of Big Data", MIT Sloan Management Review (spring), 2015, pp. 5-10.

[4].A. Gandomi, and M. Haider, "Beyond the Hype: Big Data Concepts, Methods, and Analytics", International Journal of Information Management (35), 2015, pp. 137-144.

[5].Vedder, R. G., Vanecek, M. T., Guynes, C. S., & Cappel, J. J. (1999). CEO and CIO Perspectives on Competitive Intelligence. Communications of the ACM, 42(8), 108–116.

[6].Liu, Bing. (2010). Sentiment analysis and subjectivity. Handbook of Natural Language Processing, 2nd ed. Chapman and Hall: Florida.

[7].Kumar, Akshi, and Sebastian, Teja, Mary. (2012). "Sentiment analysis. A perspective on its past present and future." International Journal of Intelligent Systems and Applications, 4 (10): 1-14.

[8].B. Pang, L. Lee, and S. Vaithyanathan,"Thumbs up?:Sentiment classification using machine learning techniques,"inProc.ACL-02Conf.Empirical Methods Natural Lang. Process., 2002, pp. 79–86.

[9].K. Dave, S. Lawrence, and D. M. Pennock, "Mining the peanut gallery: opinion extraction and semantic classification of product reviews," in Proc. 12th Int. Conf. World Wide Web, New York: ACM, 2003, pp. 519–528.

[10].V. Hatzivassiloglou and K. R. McKeown, "Predicting the semantic orientation of adjectives," in Proc. 8th Conf. Eur. Chap. Assoc. Comput.Linguist., Morristown, NJ: Assoc. Comput. Linguist, 1997, pp.174–181

[11]. A. Esuli and F. Sebastiani, "Determining the semantic orientation of terms through gloss classification," in Proc. 14th ACM Int. Conf. Inf. Knowl.Manage., 2005, pp. 617–624.

[12]. A. Esuli and F. Sebastiani, "SENTIWORDNET: A publicly available lexical resource for opinion mining," in Proc. 5th Conf. Lang. Res. Eval., 2006, pp. 417–422

[13]. International Journal of Ad hoc, Sensor & Ubiquitous Computing (IJASUC) Vol.4, No.1, February 2013, "Opinion Mining and Sentiment Analysis –An Assessment of Peoples' Belief: A Survey"S Padmaja and Prof. S Sameen Fatima.

[14]. Hsinchun Chen, Roger H. L. Chiang(2012), BUSINESS INTELLIGENCE AND ANALYTICS: FROM BIG DATA TO BIG IMPACT, MIS Quarterly Vol. 36 No. 4, pp. 1165-1188/December 2012

[15].Ferguson, Rebecca (2012). Learning analytics: drivers, developments and challenges. International Journal of Technology Enhanced Learning, 4(5/6) pp. 304–317

[16].Vasant Dhar(2012), Data Science and Prediction, May 2012,http://hdl.handle.net/2451/31553

[17].Chid Apte(2010), The Role of Machine Learning in Business Optimization, in Proceedings of the 27th International Conference on Machine Learning, Haifa, Israel, 2010.

[18].Chiang, R. H. L., Goes, P., and Stohr, E. A. 2012. Business intelligence and analytics education, and program development: A unique opportunity for the information systems discipline. ACM Trans. Manage. Inf. Syst.3, 3, Article 12 (October 2012), 13 pages.

[19].Ranjit Bose (2008), Advanced Analytics, opportunities and challenges, industrial management & data systems ISSN:0263-5577

[20].Sharma, R., Mithas, S. and Kankanhalli, A. (2014). Transforming decision-making processes: a research agenda for understanding the impact of business analytics on organisations. European Journal of Information Systems, 23 (4), 433-441.

[21].Zack Jourdan, R. Kelly Rainer, and Thomas E. Marshall, Business Intelligence: An Analysis of the Literature, Information Systems Management, 25: 121–131Copyright b© Taylor & Francis Group, LLC ISSN: 1058-0530 print/1934-8703

[22].Sonosy, O. A., Rady, S., Badr, N. L., & Hashem, M. (2016). A study of spatial machine learning for business behavior prediction in location based social networks. 2016 11th International Conference on Computer Engineering & Systems (ICCES). doi:10.1109/icces.2016.7822012.

[23].Singh B., Kushwaha N., & Vyas O. P. (2016). An interpretation of sentiment analysis for enrichment of Business Intelligence. 2016 IEEE Region 10 Conference (TENCON). doi:10.1109/tencon.2016.7847950

[24].Chen J., Huang C., & Cheng C. (2016). The monitoring system of Business support system with emergency prediction based on machine learning approach. 2016 18th Asia-Pacific Network Operations and Management Symposium(APNOMS). doi:10.1109/apnoms.2016.7737239

[25]. Ghiassi, M., Zimbra, D., & Lee, S. (2016). Targeted Twitter Sentiment Analysis for Brands Using Supervised Feature Engineering and the Dynamic Architecture for Artificial Neural Networks. Journal of Management Information Systems, 33(4), 1034-1058. doi:10.1080/07421222.2016.1267526

[26].Cambria, E. (2016). Affective computing and sentiment Analysis. IEEE Intelligent Systems, 31(2), 102-107. doi:10.1109/mis.2016.31

[27].Kumar, A., & Joshi, A. (2017). Ontology driven Sentiment Analysis on Social Web for Government Intelligence. Proceedings of the Special Collection on eGovernment Innovations in India - ICEGOV 17. doi:10.1145/3055219.3055229

[28].Swain, A. K., & Cao, R. Q. (2017). Using sentiment analysis to improve supply chain intelligence. Information Systems Frontiers. doi:10.1007/s10796-017-9762-2

[29]V. Singh, R. Adhikari, and D. Mahata, "A clustering and opinion mining approach to socio-political analysis of the blogosphere," in Computational Intelligence and Computing Research (ICCIC), 2010 IEEE International Conference on, Dec 2010, pp. 1–4.