International Journal on Applications of Information and Communication Engineering
Volume 1: Issue 2: February 2015 , Pages:59-62 www.aetsjournal.com ISSN (Online) : 2394-6237

Smart Sentiment Analysis Technique Using Big Data

Dr.Jayaraj, Dr Deepa, H.Shaheen

Abstract— Many of the companies employ big data in the form of analytics to provide the organizations with measurable data.. Sentiment Analysis is the method of detecting the contextual polarity of text. Here data analytics can help to analyze such big data. It will score the entire document as positive or negative, and it will also gain the sentiment of individual words or phrases in the document This has given rise a thirst for carrying out the study on sentiment analytics, big data and use of some smart algorithm to discover correct sentiments or opinions from unstructured big data. The approach uses natural language processing techniques of Artificial Neural Network to extract features of interest from textual data retrieved from a micro blogging platform in real-time and, hence, generate appropriate executable code for the Decision Science and get predetermined means of social communication. So by enriching semantic knowledge bases using Fuzzy Logic (for fitness approximation) for Opinion Mining in Big Data Applications with predetermined means, suggested user action decisions can be improved.

Keywords— Artificial Neural Network, Big Data, Decision Science, Fuzzy Logic, Opinion Mining

I. INTRODUCTION

Sentiment analysis(also known as opinion mining) refers to the use of natural language processing, text analysis and computational linguistics to identify and extract subjective information in source materials. Sentiment Analysis aims to determine the attitude of a speaker or a writer with respect to some topic or the overall contextual polarity of a document. The attitude may be his or her judgment or evaluation (see appraisal theory), affective state (that is to say, the emotional state of the author when writing), or the intended emotional communication (that is to say, the emotional effect the author wishes to have on the reader).

Existing approaches to sentiment analysis can be grouped into four main categories: keyword spotting, lexical affinity, statistical methods, and concept-level techniques. Keyword spotting classifies text by affect categories based on the presence of unambiguous affect words such as happy, sad, afraid, and bored. Lexical affinity not only detects obvious affect words, it also assigns arbitrary words a probable ―affinity‖ to particular emotions. Statistical methods leverage on elements from machine learning such as latent semantic analysis, support vector machines, "bag of words" and Semantic Orientation — Point wise Mutual Information. More sophisticated methods try to detect the holder of a sentiment (i.e. the person who maintains that affective state) and the target (i.e. the entity about which the affect is felt). To mine the opinion in context and get the feature which has been opinionated, the grammatical relationships of words are used.

Grammatical dependency relations are obtained by deep parsing of the text. Unlike purely syntactical techniques, concept-level approaches leverage on elements from knowledge representation such as ontologies and semantic networks and, hence, are also able to detect semantics that are expressed in a subtle manner, e.g., through the analysis of concepts that do not explicitly convey relevant information, but which are implicitly linked to other concepts that do so. A human analysis component is required in sentiment analysis, as automated systems are not able to analyze historical tendencies of the individual commenter, or the platform and are often classified incorrectly in their expressed sentiment. Automation impacts approximately 23% of comments that are correctly classified by humans. Sometimes, the structure of sentiments and topics is fairly complex. Also, the problem of sentiment analysis is nonmonotonic in respect to sentence extension and stop-word substitution (compare THEY would not let my dog stay in this hotel vs I would not let my dog stay in this hotel). To address this issue a number of rule-based and reasoning-based approaches have been applied to sentiment analysis, including Defeasible Logic Programming. Also, there is a number of tree traversal rules applied to syntactic parse tree to extract the topicality of sentiment in open domain setting.

The best way to track the sentiment of nearly all demographics is to monitor social media. Currently there are a variety of narrow AI-based subscription services that provide the ability to track consumer comments in real time. However, they are expensive and most offer limited flexibility.

A. Natural language processing

Natural language processing gives machines the ability to read and understand the languages that humans speak. A sufficiently powerful natural language processing system would enable natural language user interfaces and the acquisition of knowledge directly from human-written sources, such as newswire texts. Some straightforward applications of natural language processing include information retrieval (or text mining) and machine translation.

Dr.Jayaraj , Professor, Department of Electronics and Communication Engineering, Nehru Institute of Engineering and Technology, Coimbatore, Tamilnadu, India.

Dr.Deepa , Professor, Department of Electronics and Communication Engineering, Nehru Institute of Engineering and Technology, Coimbatore, Tamilnadu, India.

H.Shaheen, Assistant Professor, Department of Computer Science and Engineering, Nehru Institute of Engineering and Technology, Coimbatore, Tamilnadu, India. (Email: shaheen66@gmail.com)

The combination of the two technologies ie NLP and Big Data provides more documents processed per hour for fewer dollars spent than any other available solution for natural language processing on tens of millions of documents. Analysis teams can potentially save millions of dollars a year and better devote human resources to the more nuanced business of reasoning about risk exposure.

B. Big Data

Big Data, Mining, and Analytics ties together big data, data mining, and analytics to elucidate how readers can leverage them to extract valuable insights from their data. Illustrate basic approaches of business intelligence to the more complex methods of data and text mining, guides through the process of extracting valuable knowledge from the varieties of data currently being generated in the brick and mortar and internet environments. Data analytics is the science of examining raw data with the purpose of drawing conclusions about that information.

The science is generally divided into exploratory data analysis (EDA), where new features in the data are discovered, and confirmatory data analysis (CDA), where existing hypothesis are proven true or false. Qualitative data analysis (QDA) is used in the social sciences to draw conclusions from non-numerical data like words, photographs or video. Data analysis is used to determine whether the systems in place effectively protect data, operate efficiently and succeed in accomplishing an organization's overall goals.

C. Issues & Challenges with Big Data

According to Stephen Kaisleret. al. the data stored with machine plays very important role in decision making and knowledge discovery. A major challenge for IT researchers and practitioners is that growth rate is fast exceeding our ability to both: (1)Design appropriate systems to handle the data effectively (2) Analyze it to extract relevant meaning for decision making.

Big data is the realization of greater business intelligence by storing, processing, and analyzing data that was previously ignored due to the limitations of traditional data management technologies. The ultimate purpose of big data is to discover new insights that lead to useful applications .Big Data is also behind a range of new business models such as online dating and social media websites. Big Data went mainstream with the development of cloud computing Organizations of any size can now analyze their data using massive computing power without having to invest in expensive and high maintenance hardware. Big Data is also creating new jobs in the interpretation of the results. Hadoop is a distributed file system and data processing engine that is designed to handle extremely high volumes of data in any structure. The focus is on supporting redundancy, distributed architectures, and parallel processing. NoSQL focuses on a schema-less architecture.

II. LITERATURE REVIEW

Till today different sentiment analysis techniques are being implemented with different aspects of evaluation for big data opinion analysis.

S.No	Sentiment Analysis Technique	Takes into account
1	Document level sentiment analysis	Classifying the whole document as positive or negative
2	Supervised learning techniques Unsupervised learning techniques	"terms and their frequency", "parts of speech", "sentiment words and phrases", "sentiment shifters" Use of fixed syntactic patterns that occur in an opinion.
3	Sentence level Sentiment Analysis	Associated with a phrase or sentence
4	Aspect Based Sentiment Analysis	sentiment on entities and/or aspect of those entities
5	Oracle Advanced Analytics	Database into a comprehensive advanced analytics

The above table no. 1 shows comparative analysis of different sentient analysis techniques and what it takes into account for implementation base. Polarity, subjective detection and opinion identification all are very important things in this kind of sentiment analysis. Document level sentiment analysis classifies the whole document as positive and negative statement documents. Supervised learning verifies terms and frequency, 'parts of speech', _sentiment words and phrases', _sentiment shifters'. Unsupervised learning technique uses fixed syntactic patterns that occur in an opinion. Sentence level sentiment analysis is associated with phrase or sentence evaluation. Aspect based sentiment analysis evaluates sentiment and aspect of entities. And Oracle Advanced analytics provides database into comprehensive advanced analytics.

III. RESEARCH GAP

The creation of analytics and the consumption of analytics are two different things. The real challenge is transforming the people and the processes to analyze unstructured big data to extract relevant meaning. Thus researcher pointed a problem to design and develop sentiment analysis techniques for unstructured big data for translating analytics into good decisions.

IV. PROBLEM STATEMENT

As information technology system become less monolithic and more distributed, real-time big data analysis will become less exotic and more common place. At that point, the focus will shift from data science to next logical frontier: decision science.

Researches would like to focus big data for the study to design an efficient sentiment analytic technique for unstructured big data. Such Sentiment analysis is very useful to identify and predict current and future trends, product

reviews, people opinion for social issues, effect of some specific event on people.

V. OBJECTIVES

- To carry out comparative study of different big data analytics techniques for unstructured big data.
- Design and form a better sentiment analysis technique for unstructured big data.
- Generalization of sentiment analysis technique.
- While considering multiple parameters, with accuracy, expected to be fast, precise and improved. It will help to design the strategies and reduce the business loss.
- The techniques will be generalized, useful not only for Indian Educational System, but for any area wherever voluminous data is required to be accessed within shortest time.

VI. RESEARCH METHODOLOGY

The researcher has planned to follow Design and Creation research Strategy (figure no. 1). The strategy focuses on formation of new sentiment analysis technique for big data analytics

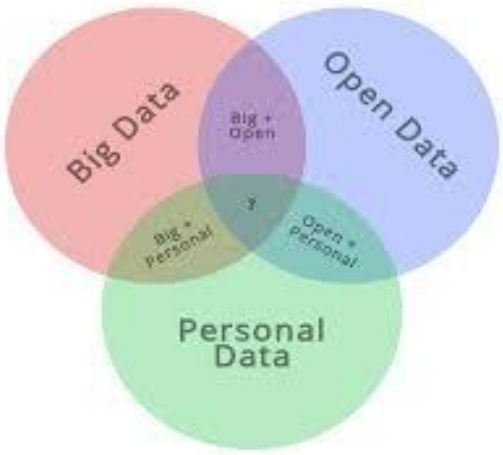

Figure 1. Collection and classification of Big Data

The above Figure No. 1 shows the manner of pre-processing of unstructured big data and classification so as to apply phase wise algorithmic training, data specification and pattern recognition or to check fitness of membership for specified pattern ranges. Unstructured data is data that does not follow a specified format for big data. Machine generated unstructured data includes Satellite images, scientific data such as seismic imagery, atmospheric data, and high energy physics, Photographs and video as security, surveillance, and traffic video, Radar or sonar data (vehicular, meteorological, and oceanographic seismic profiles) and human-generated unstructured data includes Text internal to your company all the text within documents, logs, survey results, and e-mails. Enterprise information actually represents a large percent of the text information in the world today, Social media data is generated from the social media platforms such as YouTube, Facebook, Twitter, LinkedIn, and Flickr, Mobile data (text messages and location information), website content: unstructured content, like YouTube, Flickr, or Instagram. Organisations stires such data and some organisations provides education, research and best practices to handle big data.However the technology didn‘t really support doing much with it except storing it or analysing it manually.

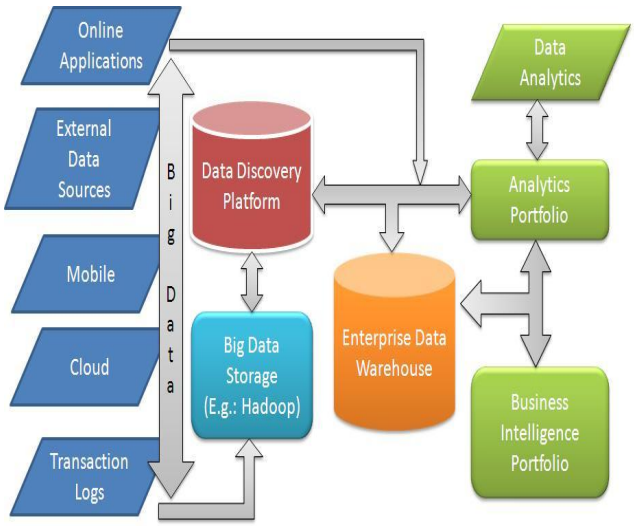

Figure No.2 Prescriptive Analysis

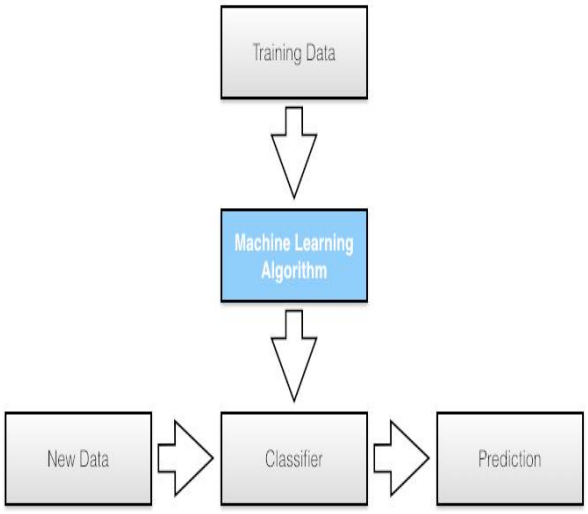

Figure 3 Classification of Big Data

As shown the figure no. 2, Prescriptive analytics automatically synthesizes big data, multiple disciplines of mathematical sciences and computational sciences, and business rules, to make predictions and then suggests decision options to take advantage of the predictions. So as shown in Figure no. 2, collection of structured and unstructured big data includes challenges as capture, analysis, search, sharing, storage, transfer, visualization, and privacy violations etc. Researcher would like to apply technique to classify data with range partitions and instead of manual analytic efforts, formation of algorithmic technique to get synthesized data to

be compared with fuzzy patterns. Figure No. 2 shows that, it needs to simplify the large and complex data sets into smaller but logically related container sets, so all documents firstly will be divided into opinionated and non-opinionated documents, so as to focus only opinionated documents to synthesize furthermore. The researcher would like to apply intelligent models to this synthesis, by applying data mining or fuzzy based rules to get prediction of categories. Then subjective and objective sentences of documents can be compared using automated rule based system to make predictions and to suggest decisions. Predicted results can be classified into three categories as positive, negative and neutral set of opinions.

By setting fuzzy weights to get identical features extraction and comparison for parameter passed as input. If sample data is collected, filtered and classified for datasets for analytics range partitions can help to form dataset in meaningful manner and then for every data set parameter membership for fuzzy specification can be evaluated to apply membership based rule to predict it's category of opinion (positive, negative, neutral) more accurately as shown in figure no.4. By applying algorithm to datasets which are made based on range partitions for its identical features, one can identify and derive related opinion category.

Algorithm to give Input and to get Output predictions

Step 1: Perform Extract, Load and Transfer data tasks on data warehouse

Step 2: Develop Patterns and analysis

Step 3: Train pattern using Artificial Neural Network

Step 4: Apply Rule based fuzzy logic to patterns.

Step 5: Input: Extract parameters from structured and unstructured data by using data cleaning and data pre-processing.

Step 6: Process: Execute fuzzy model to apply fuzzification for identifying membership of given input parameter (i.e. check fitness approximation).

Step 7: Process: Apply inference rule according to comparison of identified member category.

Step 8: Output: based on training decide analytics result for matching input parameter from trained network.

VII. Conclusion

In this paper, by using smart intelligent strategy by applying ANN and fuzzy logic algorithm is proposed for sentiment analysis techniques. The proposed algorithm simplifies the challenges of unstructured and structured data preprocessing and Artificial Neural Network addresses to pattern learning and fuzzy logic helps to determine fitness approximation comparing for proper decision base for different ranges. Proposed smart algorithm simplification and improvement in processing logic and speed apply to Big Data analytics and testing of algorithm for different business purposes may be future research domain. Significant smart sentiment analytics can help to predict habits, to take improved decisions, to recognise and design pattern for products and services, to design organisational business policies. Using narrow AI to track consumer sentiment, and separately extract specific relevant information from big data

References

[1] Edd Dumbill, Big Data 2012 Edition O'Reilly, Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472

[2] Jalaj S. Modha, Prof. and Head Gayatri S. Pandi, Sandip J. Modha, ―Automatic Sentiment Analysis for Unstructured Data‖, International Journal of Advanced Research in Computer Science and software Engineering, Volume 3, Issue 12, December 2013 pp no (91-97)

[3] Meena Rambocas, João Gama ―Marketing Research: The Role of Sentiment Analysis‖, FEP Working Papers, April 2013 ISSN: 0870-8541

[4] Felipe Bravo-Marqueza, Marcelo Mendoza, Barbara Poblete ―Meta-Level Sentiment Models for Big Social Data Analysis‖, Knowledge Based Systems May 2014

[5] Demystifying Big Data: A Practical Guide to Transforming the Business of Government, TechAmerica Foundation's Federal Big Data Commission, 2012

[6] George Gilbert, A guide to big data workload management challenges, May 2012, by Datastax.

[7] Michael Kozlowski, ―How big data helps the Education System, Jan. 2010.

[8] http://en.wikipedia.org/wiki/Prescriptive_analytics

[9] http://en.wikipedia.org/wiki/Big_data

[10] http://en.wikipedia.org/wiki/Artificial_intelligence