BIG DATA OPINION ANALYSIS SYSTEM FOR SOCIAL SET

Aher Bhagyodaya S. [1], Pawar Pratibha D.[2], Nikam Kishori R.[3], Naik Bhavini N.[4], Prof. D.S.Thosar[5]

[1,2,3,4] BE Student, Computer Department, SVIT Chincholi, Nashik
[5] Professor, Computer Department, SVIT Chincholi, Nashik

Abstract

Current analytical approaches in Computational Social Science can be characterized by four dominant paradigms: text analysis (information extraction and classification), social network analysis (graph theory), social complexity analysis (complex systems science),social simulations (cellular automata and agent-based modeling). However, when it comes to organizational and societal units of analysis, there exists no approach to conceptualize, model, analyze, explain and predict social media interactions as individuals' associations with ideas, values, identities, etc. To address this limitation, based on the sociology of associations and the mathematics of set theory, this paper presents a new approach to big data analytics called Social Set Analysis. Social Set Analysis consists of a generative framework for philosophies of computational social science, theory of social data, conceptual and formal models of social data, and an analytical framework for combining big social datasets with organizational and societal datasets. Three empirical studies of big social data are presented to illustrate and demonstrate Social Set Analysis in terms of fuzzy set-theoretical sentiment analysis, crisp set-theoretical interaction analysis and event-studies oriented set-theoretical visualizations. Implications for big data analytics, current limitations of the set-theoretical approach, and future directions are outlined.

Keyword : - Big social data, Formal Models, Social Set Analysis, Big data visual Analytics, New Computational Models for Big Social Data.

1. Introduction

A large multinational food corporation is assessing consumer preferences for its fast food product line comprising of instant noodles other rice and wheat based products. The questionnaire is focused on how to increase consumption frequency based on pack size, positioning it as snack between meals, ready-to-eat capability, variety of available flavours promotional tie-ins with other products. The Product Manager has commissioned a market research agency to do a consumer study to rank peoples preferences for increased consumption with respect to all these variables. The agency has come back with data from 1000 participants. Verifying Text: The basic question asked in Sentiment Analysis is whether a given piece of text contains any subjective content (opinions, emotions, etc.) or not. This task aims to tackle this problem of differentiating between subjective and objective content. Verifying discrete polarities: Once the subjective part is determined, the next step is to determine if the content is positive or negative. This problem can be looked upon as a classification problem. Identifying an ordinary value: Some applications require not just the type of polarity but the intensity as well. For example, movies are typically rated on a 5 point scale. Thus, this step aims at identifying such an ordinal value. Identifying subjective portions of text: The same word can be treated as subjective in one context, while it might be objective in other things. This makes it difficult to identify the subjective (sentiment-bearing) part of text. Example: - The language of the author was so crude. - Crude oil is extracted from the sea . The same word crude" is used as an opinion in first sentence, while it is completely objective in the second sentence. Associating sentiment with specific keywords: Many sentences indicate an extremely strong opinion, but it is difficult to pinpoint the source of these sentiments. Hence an association to a keyword or phrase is mostly difficult. For example: - Each time I read 'Pride and Prejudice' I Need to dig her up and beat her over the skull with her own shin-bone. In this example, her" refers to the character in the book Pride and Prejudice", which is not externally mentioned. In these cases the negative sentiment must be associated with the character in book.

2. System Architecture:

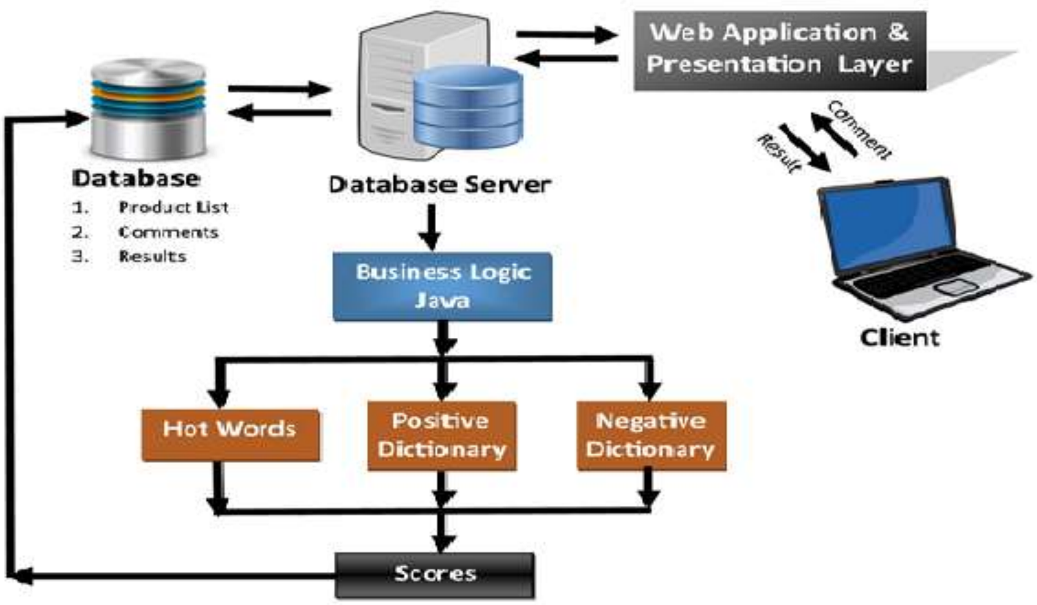

Fig: System Architecture

The system will accept the input in 2 formats: XML editor and existing database. The first type of input will be converted to XML format and then passed on to the XML parser. The XML parser will parse the XML file which is accepted as an input. The parser will extract information, and pass to matrix builder. The matrix builder will build the entity matrix and pass it to Normalizer. The matrix builder will also accept the input in the form of an existing database and pass it to the Normalizer. The Normalizer will normalize the data and store it.

Client give the comment through different social set. Comments are fetched by web application.Further this comments are downloaded into database server. As we are going to use Business Logic by implementing java, as java is platform independent, secure. Here, we separate hot words for analysis. The comparison is done in hot words by using prediction algorithm and social set theory. Now we will get Positive, Negative and Neutral analysis result using scores. Database will contain Product list, Comments, Result.

3. Construction Methodology:

We have proposed system to address this limitation, based on the sociology of associations and the mathematics of set theory, this paper presents a new approach to big data analytics called Social Set Analysis. Social Set Analysis consists of a generative framework for philosophies of computational social science, theory of social data, conceptual and formal models of social data, and an analytical framework for combining big social datasets with organizational and societal datasets. Three empirical studies of big social data are presented to illustrate and demonstrate Social Set Analysis in terms of fuzzy set-theoretical sentiment analysis, crisp set-theoretical interaction analysis and event-studies oriented set-theoretical visualizations. Implications for big data analytics, current limitations of the set-theoretical approach, and future directions are outlined..

Basic Method

In the first basic method, we can suggest to develop an web application. We are fetching comment from user. This comment will be fetched by web application. The comments are then proceed for analysis part.Here comments separated,then sentences are form.Splitting of the text if done. By using dictionary approach we display the polarity class of comments

Analysis of Comments

In this method, after separation of sentence, splitting of text is done using Gate processor. As we are going to design dictionary for hot words, dedicated list of words will be generated. In the first basic method, we can suggest to overcome drawback of existing system which are, Handlling AND and BUT clause,Negation handling, coreferencing.System will give N-Point scalling that is scoring. On the basis of score Positive,Negative and Neutral social set analysis is done.

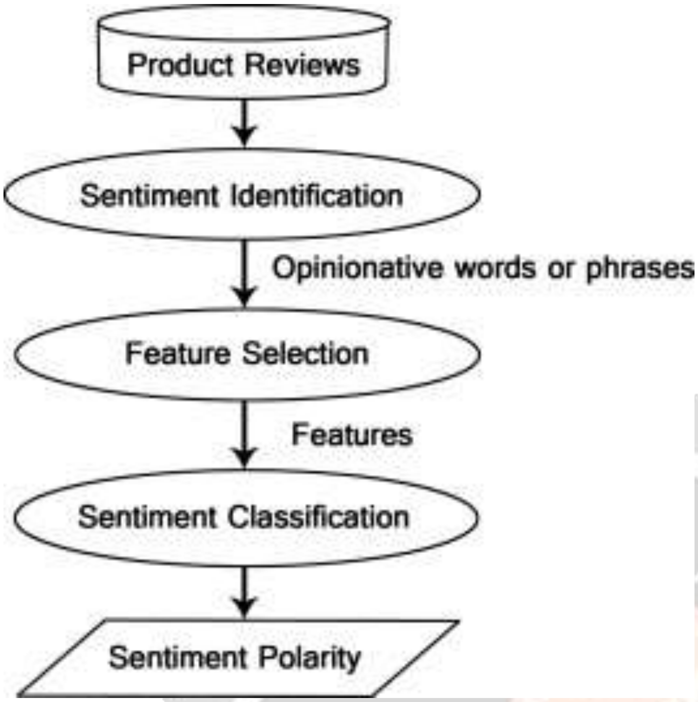

Figure 1. Sentiment analysis process on product reviews.

4. GATE

General Architecture for Text Engineering or GATE is a Java suite of tools originally developed at the University of Sheffield beginning in 1995 and now used worldwide by a wide community of scientists, companies, teachers and students for many natural language processing tasks, including information extraction in many language.

GATE includes an information extraction system called ANNIE (A Nearly-New Information Extraction System) which is a set of modules comprising a tokenizer, a gazetteer, a sentence splitter, a part of speech tagger, a named entities transducer and a coreference tagger. ANNIE can be used as-is to provide basic information extraction functionality, or provide a starting point for more specific tasks.

Languages currently handled in GATE include English, Chinese, Arabic, Bulgarian, French, German, Hindi, Italian, Cebuano, Romanian, Russian, Danish.

Plugins are included for machine learning with Weka, RASP, MAXENT, SVM Light, as well as a LIBSVM integration and an in-house perceptron implementation, for managing ontologies like WordNet, for querying search engines like Google or Yahoo, for part of speech tagging with Brill or TreeTagger, and many more. Many external plugins are also available, for handling e.g. tweets.

GATE accepts input in various formats, such as TXT, HTML, XML, Doc, PDF documents, and Java Serial, PostgreSQL, Lucene, Oracle Databases with help of RDBMS storage over JDBC.

JAPE

JAPE transducers are used within GATE to manipulate annotations on text. Documentation is provided in the GATE User Guide.[10] A tutorial has also been written by Press Association Images

In computational linguistics, JAPE is the Java Annotation Patterns Engine, a component of the open-source General Architecture for Text Engineering (GATE) platform. JAPE is a finite state transducer that operates over annotations based on regular expressions. Thus it is useful for pattern-matching, semantic extraction, and many other operations over syntactic trees such as those produced by natural language parsers.
JAPE is a version of CPSL – Common Pattern Specification Language.
A JAPE grammar consists of a set of phases, each of which consists of a set of pattern/action rules. The phases run sequentially and constitute a cascade of finite state transducers over annotations. The left-hand-side (LHS) of the rules consist of an annotation pattern description. The right-hand-side (RHS) consists of annotation manipulation statements. Annotations matched on the LHS of a rule may be referred to on the RHS by means of labels that are attached to pattern elements.

5.Objective

The purpose of this document is to present a detailed description of the product rating Review Summarization. It will explain the purpose and features of the system, the interfaces of the system, what the system will do, the constraints under which it must operate and how the system will react to external stimuli. This task is commonly defined as classifying a given text (usually a sentence) into one of two classes: objective or subjective. This problem can sometimes be more difficult than polarity classification. The subjectivity of words and phrases may depend on their context and an objective document may contain subjective sentences (e.g., a news article quoting people's opinions). Moreover, results are largely dependent on the definition of subjectivity used when annotating texts. However, showed that removing objective sentences from a document before classifying its polarity helped improve performance.

we are going to add comparison of 2 models comment analysis which is depend on users choise .User when logins to our web app can click on" Do you want to compare ?"button, there he need to select 2nd model with which he wants to compare the choosen model from the option given in combo box and user can see the comparision of their respective comment analysis.we are trying for this in our proposed system.

6.Software,Hardware &Test Data Requirements:

6.1Hardware Requirement:

Hardware platform:
_ Hard disk drive: 40GB
_ RAM: Minimum 512MB

6.2Software Requirements:

Operating system:

_ Windows based operating system
_ Linux

Software :

_ Eclipse Galileo, Oracle 10g.

Language: Core Java, Servlets, JSP and HTML5.

7. Conclusion:

In this paper, we concentrate on subjective summarization for unsupervised sentiment classification. Initially we process on comments to categorize each word into different classes then summarized scores for given comment is generated using subjective summarization algorithm. The further scope for this system will be to obtain results for comparison between two or more products. Lot of work needs to be done on spam detection but we have implemented few through this paper. Some sentences make confusing situation for the human mind as well the system. Likewise, all things considered, applications, to give a totally automated arrangement are no place in sight. Be that as it may, it is conceivable to devise productive semi-automated arrangements.The key is to fully understand the whole range of issues and pitfalls, skillfully manage them, and determine what portions can be done automatically and what portions need human assistance. In the continuum between the fully manual solution and fully automated solution, we can push more and more toward automation.

8. References:

[1] R. Xia, F. Xu, C. Zong, Q. Li, Y. Qi and T. Li, "Dual sentiment

[2] Z. Hai, K. Chang, J. Kim, and C. C. Yang, "Identifying features in opinion mining via intrinsic and extrinsic domain relevance," IEEE Trans. Knowl. Data Eng., vol. 26, no. 3, pp. 447–462, Mar. 2014.

[3] R. Xia, T. Wang, X. Hu, S. Li, and C. Zong, "Dual training and dual prediction for polarity classification," in Proc. Annu. Meeting Assoc. Comput. Linguistics, 2013, pp. 521–525.

[4] C. Lin, Y. He, R. Everson, and S. Ruger, "Weakly supervised joint sentiment-topic detection from text," IEEE Trans. Knowl. Data Eng., vol. 24, no. 6, pp. 1134–1145, Jun. 2012.

[5] A. Abbasi, S. France, Z. Zhang, and H. Chen, "Selecting attributes for sentiment classification using feature relation networks," IEEE Trans. Knowl. Data Eng., vol. 23, no. 3, pp