TEXT MINING APPROACH TO EXTRACT KNOWLEDGE FROM SOCIAL MEDIA DATA TO ENHANCE BUSINESS INTELLIGENCE

R. Jayanthi[1], Dr. Ananthi Sheshasaayee[2]

[1]*Research Scholar and Assistant Professor,* [2]*Research Supervisor and Associate Professor,*

PG and Research Department of Computer Science, Quaid-E-Millath Government College for Women

(Autonomous), Chennai 600 002, TamilNadu (India)

ABSTRACT

In recent years, organizations are facing numerous challenges in dealing with the tremendous volumes of unstructured texts available in social media. Internet has become world-wide famous and important for content sharing, until now, the content that is generated from these websites remains largely unused. Social media data largely contains unstructured text which is not precise and complete. If the incomplete text data is considered for business decision making, then uncertainty is propagated to the results. Text mining refers to the process of extracting interesting and non-trivial information and knowledge from unstructured text. Business Intelligence is one of the applications where we can apply text mining effectively. Business Intelligence (BI) increases the competitive advantage of a business by intelligent use of available information collected from the users to make wise decisions. This paper presents a text mining approach to extract valuable information from social media content to predict real-world decisions that enhance business intelligence.

Keywords-Business Intelligence, Information Extraction, Social Media, Text Mining,Unstructured text.

I INTRODUCTION

Social networking sites like Facebook, Twitter, LinkedIn, Google Plus or Pinterest are increasingly used to provide value to have online conversations and mainly good at promoting discussion and events, connecting with like-minded organizations and the media and an interesting way to network with potential new supporters[1]. There are multiple sources of data for businesses in the form of web pages, emails, video and image files, news and reports which are called semi structured or unstructured data. They are in different formats and searching, analyzing such data is a time consuming activity and may lead to poor decisions.

Business Intelligence (BI) refers to applications of techniques for collecting, storing and analyzing business data that finally helps the enterprise to make better decisions [2]. Mostly, BI techniques are computer based, they help in analyzing past, current and future trends of the enterprise. Analyzing the data plays a major role in decision support system, query and reporting, data mining, complex event processing, online analytical processing, process mining, business performance management, text mining, statistical and predictive analysis.

 In addition to traditional data found in databases, other sources of information found in non-traditional sources can be equally, if not more, valuable. For instance, organizations can pool customer sentiment information from

CRM applications and gather relevant information from external sources.   An important aspect of BI is knowledge management that helps companies in making good strategies through proper insight and experiences. BI helps companies to analyze their tremendous volumes of data for decision making but unfortunately not all data is structured and simple to understand as some data exists in unstructured or semi-structured form which results in time consuming search and interpretation.

The dominant tool for business intelligence is text mining. The concept of text mining may seem to be complicated, but understanding the process is easy if the task is broken down step by step. Using a simplified approach that breaks down the overall process into steps offers insight into how text mining works. Also, organizations can identify which areas best suit their requirements and how to best apply text mining within their current processes. Text mining is generally defined as the process of deriving high quality information from text.

II TEXT MINING - AN EFFICIENT TOOL FOR BUSINESS INTELLIGENCE

The aim of text mining is to collect huge amounts of unstructured data available in any form from enterprises or from the internet, and use it to gain insights to solve real-world problems[3]. The available sources of data for businesses can be gathered from web pages, emails, video and image files, news and reports. These data are either semi structured or unstructured. Since they are in different formats, searching and analyzing such data is a time consuming activity and this would lead to poor decisions. To solve this problem, application of techniques of text mining for information extraction, automatic categorization or summarization of documents has become mandatory [4]. The Text Mining Process involves the following steps:

2.1 Text Pre-processing

This is a step-by-step method involving following mechanisms.

   a. Noise Removal- this process involves removal of stopwords and punctuation marks from text data so that the resultant text is free from redundancy and unwanted symbols.
   b. Tokenization- it is parsing of unstructured text in which the text is split into sets of tokens on the basis of words, sentences or paragraphs.
   c. Parts of Speech Tagging- in this process, grammar rules are applied on the text with their corresponding parts of speech.
   d. Word-sense Disambiguation- it allows finding different meanings or sense of a word implied in different situations.
   e. Text Transformation/ Attribute Generation- Attribute generation generates labels/ attributes from the text based on their features.

2.2 Text Representation

It allows representation of text by features and their occurrences using approaches of "Bag of words" and "Vector space" where each word is represented as an individual variable having numeric weight.

2.3 Feature Generation and Feature Selection

- Feature Generation This process selects features of a document so as to improve its representation which may be earlier misleading or redundant. It uses approaches of selection before use or selection based on use.

- Feature Selection- It involves reduction in text dimensionality and irrelevant attributes to deal with issues of scarcity of resources and feasibility. It is used to improve text representation by selecting a subset of features either before using them in a classifier or how well they perform in classifier. Though the approach of selection before use is independent of many classifiers and has low costs in computation but it is less effective in comparison to the approach of selection based on use as it evaluates on performance.

2.4 Interpretation/ Evaluation

It is the final step in text mining process which could lead to either termination or iteration. Text mining process could be finally terminated if well-suited results are achieved for business intelligence or the process could be finally iterated if the results are not up to the mark or are used as a part of further inputs.
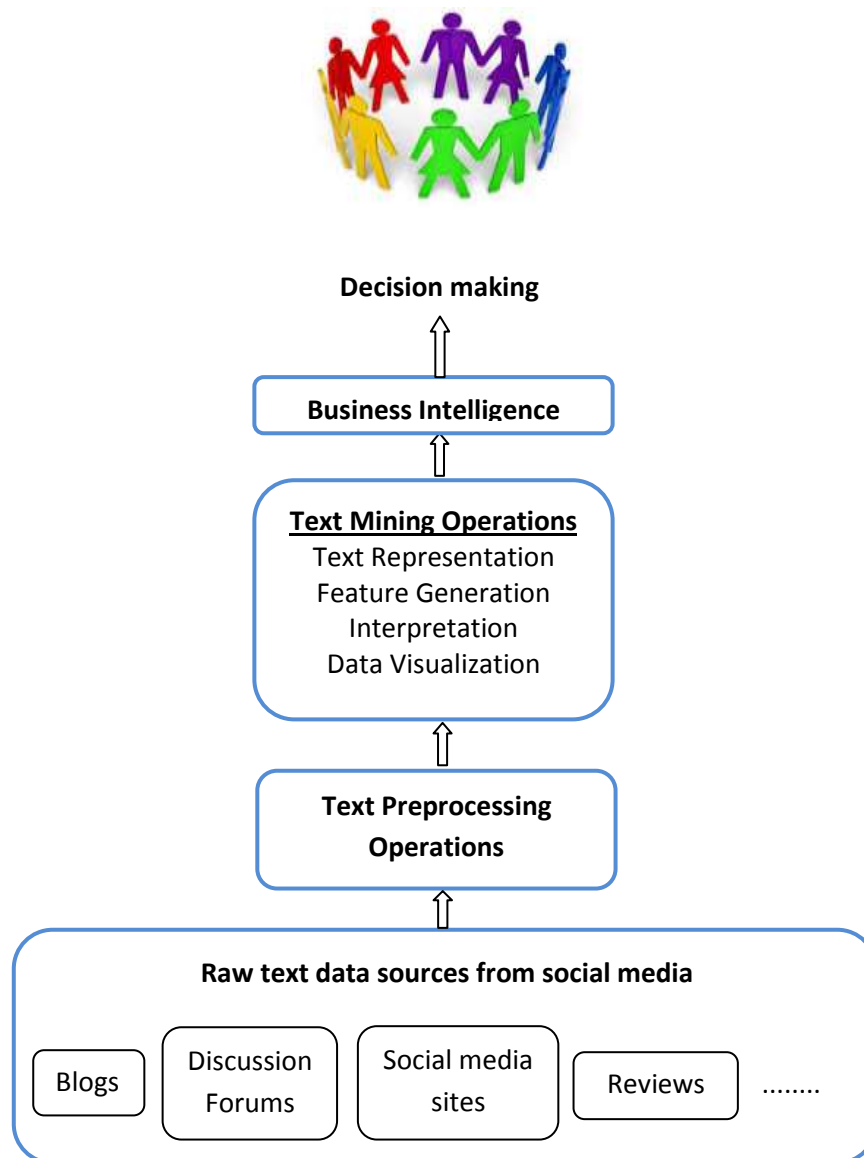

Decision making

Business Intelligence

Text Mining Operations
Text Representation
Feature Generation
Interpretation
Data Visualization

Text Preprocessing Operations

Raw text data sources from social media

Blogs   Discussion Forums   Social media sites   Reviews   ....

Fig.1: Functional Architecture of Text Mining for Business Intelligence System

2.5 Data visualization

For text-based information in a way that can be analyzed by organizations is becoming more important as organizations increasingly mine this data. Business intelligence tools offer organizations a wide array of visualization options to present mined text results in a user-friendly way that will identify trends and report on findings to benefit planning. Beyond the actual presentation of this data, business intelligence allows organizations to store the data in a data warehouse. Also, data can be combined with other data to draw an overall picture of both structured data and unstructured data combined [5].

Fig.1 highlights the Text Mining approach to extract information from social media sites. Text mining is an important aspect of Business Intelligence that helps users and enterprises in analyzing stored text in an improved way to enable better decisions, resulting in customer satisfaction and gaining competitive advantage[6], [7]. Text Mining is better than data mining as it provides deeper insight into the expanding business domain and extracts more fruitful data for business intelligence from unstructured information sources. Its main objective is to derive new information from multi sources of raw text information, an attempt for which was rarely thought of before. There are many business intelligence applications of text mining system like Scientific data analysis, marketing, identifying potential customers and market,   segmentation, biomedical Sciences, document warehouse for SAP, fraud detection and various other relevant fields[8].

Text mining uses a number of software and technologies to help decision support system of an enterprise and keeps on generating alerts on market changeovers, mergers, poor performance and competition that in turn help the business to take corrective, measurable and preventive steps and be the leader.

III CONCLUSION

This paper presents a text mining approach to extract information from voluminous data sources to enhance business decisions. As businesses understand how text mining can be employed and its associated benefits, the transition to adoption becomes quite realistic. Forward thinking organizations are now at the threshold of identifying how text mining fits within their current business intelligence framework.

REFERENCES

[1] "Comparing Social Networking Websites", Published *on Idealware, August 2010*

[2] Li Gao et al, "Powerful Tool to expand Business Intelligence – Text Mining", *World Academy of Science, Engineering & Technology 2007*

[3] Shantanu Godbole et all, " Building Re-usable Dictionary Repositories for Real-world Text Mining", Annual text mining industry conference, CIKM'10, October 26–30, 2010, ACM 978-1-4503-0099-5/10/1

[4]   Lyndsay Wise, "Understanding the role of text mining within BI The 4th in a series ", *Dashboard Insight Article July 27, 2007*

[5] Palak Gupta et al, "Role of Text Mining in Business Intelligence", *Gian-Jyoti E-journal, volume 1 issue 2,2012*

[6] Adam Funk et al, "Opinion Analysis for Business Intelligence Applications", ACM OBI '08 Proceedings of the first international workshop on Ontology-supported business intelligence, ISBN: 978-1-60558-219-1 doi>10.1145/1452567.1452570

[7] Anil Kumar Pantangi et al, "Customer Intelligence, Classification of Customers Textual Responses via application of Topic Mining", *SAS Global Forum 2012*

[8] Vishal Gupta, Gurpreet S. Lehal, "A Survey of Text Mining Techniques and Applications", *Journal of Emerging Technologies in Web Intelligence, Vol. 1, No. 1, August 2009, pages 60-76*