Application of Sentiment Analysis in Business Intelligence

MOHAMMED ISSA M. AL-KHARUSI, ABUBAKAR IDRIS USMAN &
JAMILU AWWALU
Baze University, Abuja, Nigeria

ABSTRACT Data mining tasks such as classification or prediction are applied in a variety of domains including manufacturing and business. Sentiment Analysis has significantly grown over recent years in our day-to-day engagement information and technology. As a field of study, sentiment analysis research has correlation with aspects of artificial intelligence research including Business Intelligence. This study reviews Sentiment Analysis as applied to Business Intelligence in terms of evolution and application. Following our review of articles and journals relating to Business Intelligence, we found that despite the advancements in Sentiment Analysis and Business Intelligence in terms of evolution, challenges such as technological, ethical and legislative persist.

Keywords: Sentiment Analysis, Business Intelligence, Natural Language Processing, Big Data.

Introduction

The purpose of Sentiment analysis as described by (Aalderks & Falls, 2012) is to determine the inclination or attitude of a communicator through the contextual polarity of their writing or speaking. Sentiment Analysis referred to opinion mining. It is also sometimes characterize as a sub-category of text mining (TM), which falls under Natural Language Processing and Data Mining. Text Mining as described by (Al-azmi, 2013) deals with textual data rather than records; and it differs from data mining in terms of methodology and techniques used, such as Text Mining using complex Natural Language Processing (NLP) techniques. A breakthrough in Natural Language Processing (NLP) was achieved in 1980's when computer processing power in conjunction with machine learning capabilities enabled the exponential growth in the abilities of machines to 'appear' intelligent (Badenhorst and Fitzgerald, 2012). Business Intelligence according to (Costa and Souza, 2012) comprises of two steps. First, extracting, transforming, and data loading from unstructured (e.g. social networks websites, emails) or structured (e.g. ERP, CRM) sources resulting in data warehouse which includes repository of data that is integrated, topic-oriented, time variant, and non-volatile. Second; using analytical tools for the dissemination and analysis of knowledge.

Businesses need to have a complete understanding of their customer's opinions and needs on their products or services they offer, but they face the challenge of dealing of unstructured text from source of customer's opinions and need. Consumer products and services sentiments as stated by (Aalderks and Falls, 2012) is now not just a source of customer reviews and references, but a source for customer services, business intelligence, and product brand reputation management. As stated by (Grimes, 2008), some of challenges and needs that make organizations want to answer fundamental problems in the voice of the customer are:

- Are the customers satisfied with services, products, and support?
- What customers think of products and services offered by competitors
- What influences the market and how opinions propagate
- What do the customers like
- What problems do customers have
- And, what additional features would the customer like to have and are willing to pay for.

These challenges include handling noise and linking with structured data (Venkata et al. 2009). Ad-hoc nature and terse documentation of techniques adopted by researchers, which has proven limited in a sense that they are laboriously tailored to the intricacies of the underlying data and intended use cases (Alberto et. al. 2012).

Business Intelligence as described by (Venkata *et al.* 2009) involves the use of technologies and methodologies for the collection, integration, and analysis of relevant information in a business for the purpose of better decision making in that business. The criteria for true Business Intelligence as identified by Bogza (2008) are:

- Breadth: Functions and technologies from accros an organization,
- Depth: Disseminate across all functional and organizational levels,
- Completeness: End to end platform,
- Advanced Analytics: Deliver predictive insights not just hindsight,
- Data Quality: Ensuring the right data is used.

Industry studies as stated by (Chen and Storey, 2012) show that Business Intelligence is increasingly becoming an important aspect in both business communities and academic areas over the past decades. IBM Tech Trends Reports (2011) identified business analytics as one of four major technology trends from over 4,000 information technology professional from 93 countries in the 2010s. Companies, especially those offering consumer services and products might want to know the opinions and sentiments of their consumers and opinions on their products and services so that they can makes strategic decisions in order to attract more customers, maintain existing customers, or competitive edge over rivals.

The information used for analyzing customers sentiments and opinions are mostly obtained from sources such as customers or users review on services and products on Amazon (Aalderks and Falls, 2012).

Business Intelligence as a sub type of sentiment analysis has evolved with advent of technologies on the web, from web 1.0, 2.0, to 3.0. The evolution is shown in the figure 1 as sourced from (Chen and Storey, 2012)
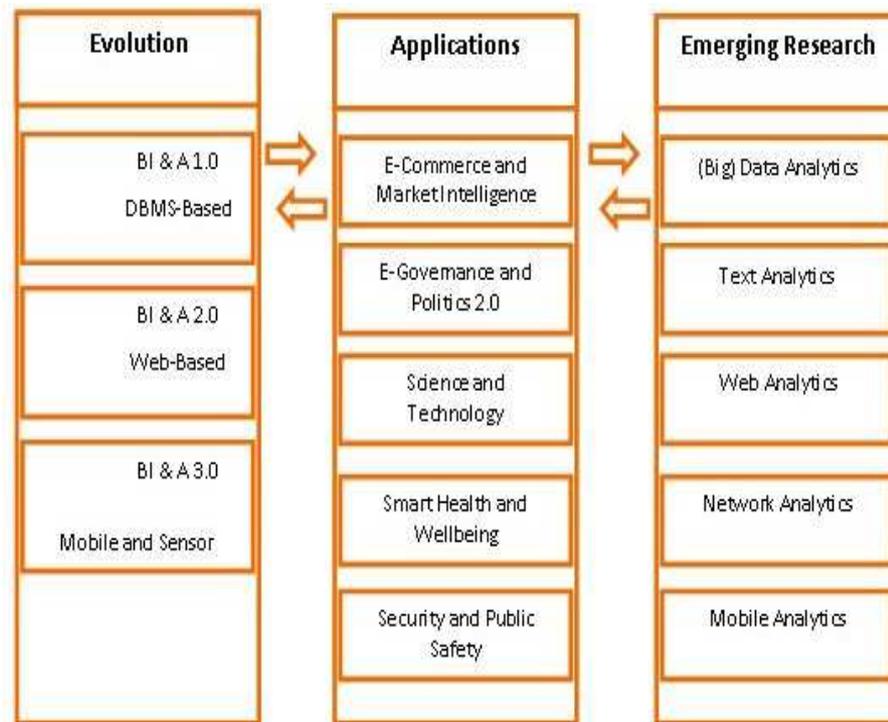
Figure 1: Overview of evolution, applications, and research on Business Intelligence

Related Works and Findings

There has been a number of research implement on mining opinions and sentiments for business intelligence. A research by (Funk and Bontcheva, 2008) using Natural Language Processing (NLP) techniques as input to machine learning instrument was used to conduct opinion analysis for business intelligence applications. The machine learning used was a support vector machine (SVM) of 10 folds cross validation was implemented in GATE including the application of NLP components for English language such as the Sentence Splitter, Morphological Analyzer, Sentence Splitter, and Hepple Parts of Speech (POS) Tagger in order to classify marked user comments as a thumb up or down rating. The dataset used was obtained from a website[1]. The website was crawled to collect corpus by fetching comments which numbered 92 documents, each containing an instance of a review by an individual, and a rating of thumbs up numbered 37% or thumbs down which was 67% of the data collected entered in the forum by a user. The findings from the study were that increasing the number of features does not improve performance in terms of accuracy; rather it makes the training data more sparse. Also, increasing the n in N-gram tends to decrease performance accuracy. Also, the unigram results i.e. 74% obtained was more favourable in terms of relating to the benchmark movie review dataset.

Visual Analytics has been highlighted as a method of analyzing unstructured data (Lu, *et al.*, 2014). The authors have demonstrated that Visual Analytics can play a role in enabling effective analysis of uunstructured data collected from social networks websites, blogs or other related sources.

Sentiment analysis as stated by (Liu, 2010) starts with subjectivity classification which determines whether a document or sentence is objective or subjective. A study by (Modha, 2013) addressed this common problem in sentiment analysis which is determining sentiments from subjective statements and ignoring objective sentences or statements. In the study, an approach was proposed using SVM, Naïve Bayes, Bag of Words (BOW), Parts of Speech (POS) Tagging, sentiment lexicons (i.e. SentiWord-Net, grammar rules, and N-gram) which is composed of four steps shown in figure 3 below. Step one: classify sentences as containing opinions or not. Step two: classifying sentences with opinions from step one into objective or subjective sentences. Step three: classifying subjective sentences into neutral, positive, or negative categories—i.e. attaching semantic orientation or context in a case where the sentence is a complex one. Step four: classifying objective sentences into neutral, positive, or negative categories—attaching semantic orientation or context in a case where the sentence is a complex one.
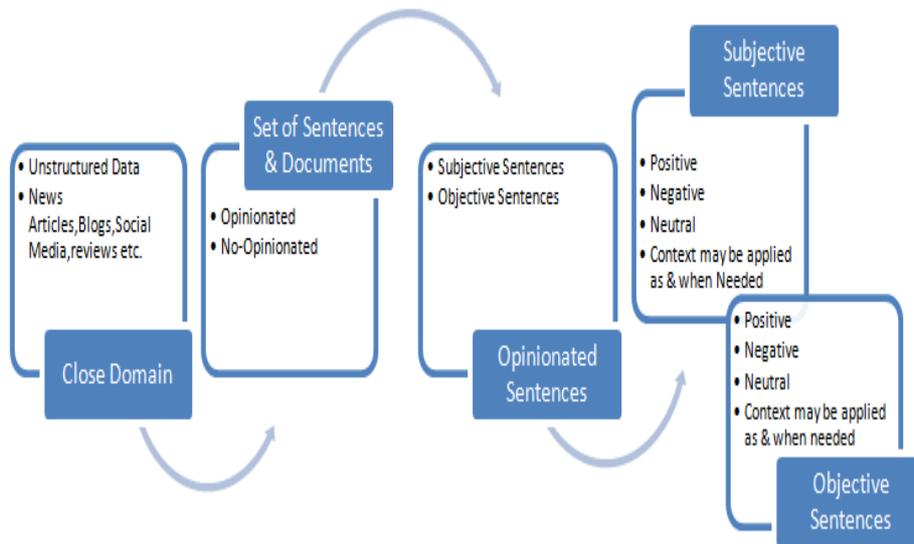

Figure 2 Proposed Classification of Sentiment Analysis. *Source* (Modha, 2013)

Depending on the task to be carried out and viewpoint of the person conducting the sentiment analysis, the approach to be analysis can be relationship-driven, discourse-driven, keyword-driven, or language-model driven. The proposed approach in figure 2 is one of the common approaches in sentiment classification. Domain independence, the biggest challenge in sentiment analysis as stated by Haseena (2014) can affect the

proposed approach in figure 2. Domain dependence occurs when one feature set gives very good performance in a specific domain, at the same time performing very poorly in some other domains.

Challenges and problems in mining technologies such as data mining (DM), web mining (WM), and text mining (TM) which sentiment analysis is a subcategory can be grouped into three, according to (Al-azmi, 2013)—such as legislative, ethical, and technological challenges—each of which is discussed as follows:

Legislative: New levels of transparency have resulted in BI due to the large amount of data made available across the net willingly or unwillingly, example is Wikileaks.

Ethical: The concern raised from public about data collected in the Internet. Customers' profiles details contain private data. Although most data is anonym-zed before use, still, concerns are exist on how enterprises use individuals' data

Technical: Limitation in terms of software accessibility and the elaborate and huge requirement for infrastructure. The need for dedicated high end hardware and distributed computational grids can be a challenge to meet in terms of cost and need for expert IT staff. Software limitation in terms of scalability is also a technical challenge because not all solutions are scalable or adaptable to every business environment.

Methods, Techniques and Approaches

Several techniques are available in conducting sentiment analysis as shown in table 1; some of these techniques that can be implemented for Business Intelligence or general purpose sentiment analysis. Some common methods of determining sentiment analysis as identified by (Aalderks and Falls, 2012) are: Scaling Systems, Subjectivity or Objectivity Identification, and Bales Interaction process Analysis (IPA).

    I.   Scaling Systems: Scaling Systems are one of popular methods of directly getting consumer sentiments online. It assigns a system for rating products or services, usually one to four or five stars as shown in figure 2 in order to determine customer appreciation of the a product or service rendered to him/her.

    II.   Subjectivity or Objectivity Identification: This is mostly in the form of documentation where any sentence or word of interest in a given document is determined to be either subjective or objective. However, this type of sentiment analysis is pretty difficult to implement because the meaning of a sentence or word can change when it is put into or taken out of contexts.

    III.   Bales Interaction Process Analysis (IPA): Scoring of interactions based on a specific unit of communication or interaction which is applied to a specific set of categories, then analysis is carried out based on the scores of respective categories.

In a categorization of methods and approaches in sentiment analysis, (Modha, 2013) highlighted five approaches as: Machine learning; using supervised and unsupervised machine learning. Natural language processing: Using Bag of Words, POS tagging, and N-gram algorithms. Text Mining. Techniques of Information Theory and Coding. Semantic Approach. And, Hybrid Approaches: A combination of any two or more sentiment analysis techniques.

55

S / No.	Title of Research Paper	Authors	Approaches, Techniques and Method
1	Automatic Sentiment Analysis of Twitter Messages	Ana C.E.S Lima and Leandro N.de Castro	Text mining approach, Naïve Bayes algorithm, word based approach, emotional based approach
2	Harnessing Twitter 'Big Data' for Automatic Emotion Identification	Wenbo Wang, Lu Chen, Krishnaprasad Thirunarayan, and Amit P.Shet	POS, SentiWordNet, N-gram, Naïve Bayes algorithm
3	Opinion mining and sentiment analysis	Bo Pang and Lillian Lee	Machine Learning Techniques, NLP Techniques like BOW, SentiWordNet
4	A sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts	Bo Pang and Lillian Lee	Graph Theory, Cut based classification
5	Sentiment Text classification of customers reviews on the web Based on SVM	Huising Xia, Min Tao and Yi Wang	Support Vector Machine Algorithm
6	Domain Independent Sentiment Classification with Many Lexicons	Bruno Ohana, Brenden Tierney and Sarah-Jane Delany	Large Sentiment Lexicon acquisition
7	Using Objective word in SentiWordNet to improve word-of-mouth sentiment classification	Chihil Hung and Hao-Kai Lin	Add Sentiment threshold values to the objectives words
8	Verb Orientation Sentiment Classification	Mostafa Karamibekr and Ali A. Ghorbani	Verb oriented and opinion dictionary
9	Sentiment classification based on Random Process	Jintao Mao and Jian Zhu	Random Process
10	Sentiment Analysis of Social Issues	Mostafa Karamibekr and Ali A. Ghorbani	Verb oriented and opinion dictionary

Table 1: Research Technique on Sentiment Analysis

Business Aspect	Business Intelligence Advantage	Benefits
Competitive Advantage	• Market Research • Risk Management • Manufacturing Optimization	• Finding Elements of Market Dominance • Bankruptcy Prediction, Better Investments • Better material usage, shipments, scheduling
Customer Relationship Management	• Customer's targeting • Pricing Discrimination • Market Baskets • Customers Satisfaction	• Target specific customers with the right products • Dynamic pricing • Better Marketing and Advertisement • Find the reasons and the costs of switching, chum, and satisfactory levels
Logistics and Supply chain Management	• Production Managements • Scheduling Supply Chain • Dynamic Reactions • Forecasting	• Prevent overproduction and underproduction • Help dynamically manage the supplies during their move through the chain • React immediately to changes to help sustain supply • Forecast the demand for production
Anomalies and Fraud Detection	• Fraud Detection • Anomaly Detection	• Help find fraudulence transactions, fraudsters, hackers, and possible counterfeiting • Find what data to leave out, why such anomalies happened, and avoid considering them.

Table 2: Summary of Advantages of BI. *Source: (Al-azmi, 2013)*

Advantages

Business Intelligence as a sub type of sentiment analysis offers arrange of benefits and advantages that is just focused to competitive advantage, but a more varied set of advan-

tages that are worthy of consideration to any company. Some of the advantages as stated by (Al-azmi, 2013) is shown in table 2.

Conclusion

Companies today make use of technologies that can help them not only to reach out to their customers wherever they are, but to also study and predict customers' behaviour and responses to products or services offered to them. This made multinational companies and corporations adopt business intelligence systems built on Artificial Intelligence techniques and algorithms such as Natural Language Processing (NLP) to achieve benefits that Sentiment Analysis for Business Intelligence offers. Despite the advancements in Sentiment Analysis and Business Intelligence in terms of evolution, challenges such as technological, ethical and legislative persist.

Correspondence
Jamilu Awwalu
Baze University, Abuja, Nigeria
Email: jamilu.awwalu@gmail.com

References

Aalderks, D. and Falls, S. (2012). Sentiment Analysis, Analyzing the Polarity of Communication, available: https://web.njit.edu/~da225/media/Assignment%207&8%20Sentiment%20Analysis.pdf.

Alberto, B., Nicolas, B., and Latifa, G., (2012). Workshop on Mining Unstructured Data (MUD) . . . because "mining unstructured data is like fishing in muddy waters"! 2012 19th Working Conference on Reverse Engineering, IEEE Computer Society Conference Publishing Services (CPS). http://www.computer.org/cps.

Al-azmi, A. R. (2013). Data , Text , and Web Mining For Business Intelligence: A Survey. International Journal of Data Mining & Knowledge Management Process (IJDKP), 3(2), 1–21.

Badenhorst, B. K. and Fitzgerald, W. (2012). The application of text and sentiment analysis in market research, social sciences and business intelligence in South Africa. Southern African Marketing Research Association (SAMRA), available: http://www.samra.co.za/wp-content/uploads/2013/05/Badenhorst-Fitzgerald_Text-and-sentiment-analysis-Expert-QA-Materials.pdf.

Bogza, R.M., and Dorin, Z., (2008). "Business intelligence as a competitive differentiator", *AQTR*, 2008, International Conference on Automation, Quality and Testing, Robotics, International Conference on Automation, Quality and Testing, Robotics 2008, pp. 146-151, doi:10.1109/AQTR.2008.

Chen, H., and Storey, V. C. (2012). Business intelligence and analytics: from big data to big Impact. MIS Quarterly, 36(4), 1–24.

Costa, P. R. S., and Souza, F. F. (2012). Towards Integrating Online Social Networks And Business Intelligence. International Conferences Web Based Communities and Social Media (IADIS WBCSM 2012), available: http://homepages.dcc.ufmg.br/~fabricio/download/iadis-wbc2012.pdf.

Funk, A., and Bontcheva, K. (2008). Opinion Analysis for Business Intelligence Applications. Proceedings of f the first international workshop on Ontology-supported business intelligence 2008 Article No. 3, available: https://gate.ac.uk/sale/iswc08/obi-musing/musing-om-iswc08.pdf.

Grimes, B. S. (2008). Voice of the Customer. Text Analytics for the Responsive Enterprise. Business Intelligence Network. Powell Media and Alta Plana Corporation, available: http://altaplana.com/BIN-VOCTextAnalyticsReport.pdf.

Haseena, R.P., (2014). Opinion Mining and Sentiment Analysis - Challenges and Applications. International Journal of Application or Innovation in Engineering & Management (IJAIEM) Volume 3, Issue 5, May 2014.

Liu, B. (2010). Sentiment Analysis: A Multi-Faceted Problem. IEEE Intelligent Systems, 2010.

Lu, Y., Wang, F., & Maciejewski, R. (2014). Business Intelligence from Social Media: A Study from the VAST Box Office Challenge, 1–10.

Modha, J. S. (2013). Automatic Sentiment Analysis for Unstructured Data. International Journal of Advanced Research in Computer Science and Software Engineering, 3 (12), 91–97.

Venkata, L.S., Tanveer, A.F., Shajith, I., Shantanu, G., and Mukesh, K.M., (2009). Business Intelligence from Voice of Customer. In Proceedings of IEEE International Conference on Data Engineering. IEEE Press.