ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

A Survey on Using Product Reviews for Mining Insights to Predict Sales

Salama H. Shaikh, Prof L.M.R.J Lobo

M.E Student, Dept. of Computer Science and Engineering, Walchand Institute of Technology, Solapur, Maharashtra, India

Associate Professor, Dept. of Computer Science and Engineering, Walchand Institute of Technology, Solapur, Maharashtra, India

ABSTRACT: People are now a day's shopping online, doing banking transactions online, purchasing products or services online and posting reviews for used product or service to help the other. There are mainly two types of review readers, one who wants to purchase particular commodity & the other who is selling commodity i.e. a customer and a vendor. The review information is useful to both. This particular system is predicting the sales influence i.e. economic impact and expected helpfulness to the customer. This system is one of the part of business Intelligence. It is helpful to vendor to take business decisions to increase sales of a particular commodity or service as well as it is helpful to customers to choose appropriate product for their requirements.

KEYWORDS: Business Intelligence, Data Mining, Opinion Mining, Prediction, Reviews Mining, Sentiment Analysis.

I. INTRODUCTION

Giving your reviews online has become very popular way for people to express opinions and sentiments about the commodity bought or services received. Analysing the large volume of online reviews available would produce useful actionable information that could be of economic & social values to vendors, customers and other interested parties. However, the high volume of reviews information that are typically published for a single product makes harder for customers as well as manufacturers to locate the best reviews and understand the true underlying quality of a product or service.

Data mining is the computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems [10].

A business Intelligence (BI) system plays a vital role in effective decision making in order to improve the business performance and opportunities by understanding the organization's environments through the systematic process of information.[9] Business intelligence is a business management term used to describe technologies which are used to gather, provide access to and analyze data and information about the organization, to help make better business decisions [6]. The purpose of business intelligence is to upgrade the quality of product.

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[14][15]. Generally speaking, 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 affective state, or the intended emotional communication [13](that is to say, the emotional effect the author wishes to have on the reader). A basic task in sentiment analysis is classifying the polarity of a given text at the document, sentence, or feature/aspect level whether the expressed opinion in a document, a sentence or an entity feature/aspect is positive, negative, or neutral. Advanced, "beyond polarity" sentiment classification looks, for instance, at emotional states such as "angry," "sad," and "happy"[2][11][12].Sentiment is nothing but determining an opinion about a product whether it is positive or negative or neutral. Sentiment classification is a special case of text categorization problem, where the classification is done on basis of attitude expressed by the consumers in discussion forums or the blogs etc. Sentiment analysis requires a deep understanding of the document under analysis because the concern here is how the sentiment is being communicated [2].Sentiment analysis is used to determine the attitude of a speaker or a writer with respect to some

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

topic or product. The attitude may be his or her judgment or the intended emotional communication. According to Kevin Swingler Prediction is Predicting the identity of one thing based purely on the description of another, related thing. Not necessarily future events, just unknowns. Based on the relationship between a thing that you can know and a thing you need to predict. It has become a very common practice for e-commerce websites to provide the facilities for people to publish their reviews, with a prominent example being Flipkart (www.flipkart.com), Twitter. Reviews are also very common and frequent thing in blog posts, social networking websites like Facebook, Twitter as well as other dedicated review websites. So those online reviews present a wealth of information on the services and products, and if properly utilized, can provide the vendors highly valuable network intelligence and social intelligence to facilitate the improvement of their business. They will also help to support the strategic market decisions like what changes we need to do in our product or should we continue with this product or not. As a result, review mining has recently received a great deal of attention. A growing number of recent studies have focused on the economic values of the reviews, exploring relationship between the sales performance of products and their reviews. Since what the general public thinks of a product can no doubt influence how well a particular service or product sells, then understanding the opinions and the sentiments expressed in the relevant reviews is of a high importance, because finally such reviews reflect what the general public think and thus can be very good indicator of the product's future sales performance.[2]

The basic idea of this paper is to create a system that would be helpful to take business decisions for a manufacturer to improve sales performance & for customers to improve social outcomes such as their perceived usefulness. The key idea is to use social networking sites reviews which are very popular now a day for business intelligence. The project has excellent demand in predicting sales performance of a particular product or service.

This research integrates issues from Social media, Business Intelligence, Data Mining, Sentiment Analysis, Review Mining, Opinion Mining and Prediction.

II. RELATED WORK

In Here they have handled two types of mechanism consumer oriented & vendor oriented. For consumers how reviews will be helpful means estimating the social impact & for vendors how reviews will be going to influence on sales. They have conducted two level study first they have performed explanatory econometric analysis for usefulness of reviews & second they have build a predictive model using Random Forests for sales influence. They have used real data set of 411 products monitored over 15 months of period on Amazon.com.[1] Facebook reviews are collected about a product and they are classified as positive, negative and neutral. For doing it they have used a word dictionary. It is vendor oriented approach [2].

They have done sentiment analysis in movie domain for sales prediction. The sentiment factor, they have proposed Sentiment PLSA i.e. S-PLSA, in which a review is considered as a document record generated by a number of hidden sentiment factors, in order to capture the complicated nature of sentiments. Training an S-PLSA model enables them to obtain a concise summary of the sentiment data embedded in the reviews. Based on S-PLSFA, they have proposed ARSA, an Autoregressive Sentiment-Aware model for sales prediction. They were also predicting the quality of a review in the absence of user-supplied indicators, and present an Autoregressive Sentiment and Quality Aware model i.e. ARSQA, to utilize sentiments and quality for predicting product sales performance [3].

The visual analytics toolkit extracts data from social media, they have been collecting streaming movie data. This toolkit is useful to extract, visualize and clean information from social media [4].For a specific product, the system extracts a list of the discussed features and their associated sentiment scores. Online products reviews and review characteristics are extracted from www.Amazon.com. A two level filtering approach is adapted to choose a set of reviews that are perceived to be useful by customers. The filtering process is based on the concept that the reviewer generated textual content and other characteristics of the review, influence peer customers in making purchasing choices. The filtered reviews are then processed to obtain a relative sentiment score associated with each feature of the product that has been discussed in these reviews. Based on these scores, the customer's impression of each feature of the product can be judged and used for the manufacturer benefit [5].

They have shown a Business Intelligence conceptual architecture in that they collecting data in warehouses & data marts. Then it goes to business views (logical model layer) and lastly the BI front end applications. They described a service-oriented architecture for business intelligence that makes possible a seamless integration of technologies into a coherent business intelligence environment, thus enabling simplified data delivery and low-latency analytics [6].

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

This complex system was developed with Self Organized Multi Agent technology that would reduce the building cost without affecting the scalability and reliability of the system. The paper presents a novel framework based on Self Organized Multi Agent technology for building the low cost BI systems [9].

III. PROPOSED METHODOLOGY

3.1 System Architecture:

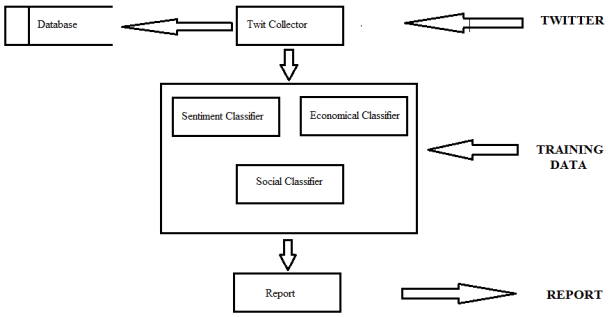

Fig. 1: System Architecture

In this system shown in Figure 1 we are fetching reviews of particular product and collecting tweets in the database. By providing training data we are building a social, economic & sentiment classifier. Then printing a report based on economic classifier for a particular vendor.

3.2 Flow of the System

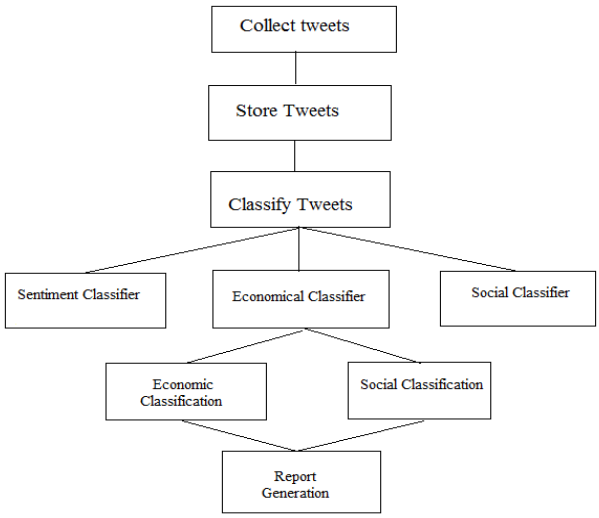

Fig 2 Flow of system

The proposed system aims to work as follows :

1. Collecting tweets from tweeter about specific product.
2. Storing tweets in the database.
3. Classifying tweets as Sentiment, Economical & Social.
4. Building social, economic and sentiment classifier.
5. In particular product it gives helpfulness for customer.
6. It will give expected influence on sales to vendor by generating a report.

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

IV EXPERIMENTAL RESULT

A. **Hardware Resources**
1. Operating System: Windows 2000/XP/2007/2008
2. RAM : Minimum 500 MB.
3. HDD: Minimum 500 GB

B. **Software Resources**
1. The Software will be developed using Java.
2. Database to be used is My SQL Server.

C. **Expected Results**
The System should generate a Report to predict a sale of a product in percentage form. Results should be in very well format that it should be helpful to vendor to take business decisions.
.

V. CONCLUSION

The system is helpful to all readers who want to take advantage of online reviews whether for shopping a commodity, going for movie or selling a product. This system gives sales prediction, so that the vendor can upgrade their product. The system is going to predict sales and generate a report to help vendor for taking business decisions.

REFERENCES

1. Anindya Ghose and Panagiotis G. Ipeirotis "Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics" Ieee Transactions On Knowlegde And Data Engineering, VOL. 23, NO. 10, October 2011
2. Mohsin Nadaf, Akshay Deshpande, Sneha Tirth "Using Business Intelligence for Mining Online Reviews for Predicting Sales Performance" IJECS Volume 4 Issue 5 May, 2015 {https://www.ijecs.in/issue/v4-i5/1%20ijecs.pdf}
3. Xiaohui Yu, Yang Liu, Jimmy Xiangji Huang, and AijunAn "Mining Online Reviews For Predicting Sales Performance: A case Study in the Movie Domain" IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 24, NO. 4, APRIL 2012{http://www.computer.org/csdl/trans/tk/2012/04/ttk2012040720-abs.html}
4. Yafeng Lu, Feng Wang, and Ross Maciejewski "Business Intelligence Analytics Business Intelligence from Social Media A Study from Vast Box Office Challenge" 0272-1716/14/$31.00 © 2014 IEEE
5. Soundarya.V, Siddareddy Sowmya Rupa, Sristi Khanna, G.Swathi, Dr.D.Manjula "Extracting Business Intelligence From Online Product Reviews" International Journal on Soft Computing (IJSC) Vol.4, No.3, August 2013
6. Liya Wu, Gilad Barash, Claudio Bartolini "A Service Oriented Architecture for Business Intelligence" IEEE International Conference on Service-Oriented Computing and Applications(SOCA'07) 0-7695-2861-9/07 © 2007
7. C.K.Gomathy, Dr.S.Rajalakshmi "A Business Intelligence Network Design For Service-Oriented Architecture" International Journal of Engineering Trends and Technology (IJETT) – Volume 9 Number 3 - Mar 2014
8. Venkatadri. M, Hanumat G. Sastry. M ,Manjunath. G "A Novel Business Intelligence System Framework" UniCSE, 112 -116, 2010
9. Zhu Zhang, Xin Li and Yubo Chen "Deciphering word-of-mouth in social media: Text-based metrics of consumer reviews"ACM(TMIS) Volume 3 Issue 1, April 2011
10. "Data mining - Wikipedia, the free encyclopedia " https://en.wikipedia.org/wiki/Data_mining
11. https://en.wikipedia.org/wiki/sentiment_analysis
12. http://www.ijecs.in/issue/v4-i5/1%20ijecs.pdf
13. http//www.definitions.net/definitions/Sentiment%20analysis
14. http://www.sciencedirect.com/science/article/pii/s1319157815001020
15. http://www.wikidata.org/wiki/Q2271421

BIOGRAPHY

Salama Hasan Shaikh is pursuing her Master Degree of Computer Science & Engineering from Walchand Institute of Technology, Solapur, Maharashtra, India. She has received Bachelor degree in Computer Engineering from Pune University. Her research interests include Data Mining, Big Data & Business Intelligence.

Prof. L.M.R.J. Lobo is an Associate Professor in Department of Computer Science and Engineering & HOD in Department of Information Technology, Walchand Institute of Technology, Solapur, Maharashtra, India. He is pursuing his Ph.D in Genetic Algorithm based approach to Data Mining . He has twenty-six years of teaching experience. His areas of interest include Data mining, Genetic Algorithm & Artificial Intelligence.