Data Mining Techniques for E-Business Intelligence

Mohiuddin Ali Khan, Dr. Sateesh K Pradhan

Abstract— Online product reviews are a major source of business intelligence that helps managers and marketers understand customers concerns and interests. The large volume of review data makes it difficult to manually analyze customers concerns. Automated tools have emerged to facilitate
this analysis, however most lack the capability of extracting the relationships between the reviews' rich expressions and the customer ratings. Managers and marketers often resort to manually read through voluminous reviews to find the relationships. To address these challenges, we propose the development of a new class of BI systems based on rough set theory, inductive rule learning, and information retrieval methods.

Index Terms— E-Business, Data Mining, Knowledge, business intelligence, shopping, behavior analysis, Cross-marketing

———————————— ◆ ————————————

1 INTRODUCTION

With data mining, business patterns can be discovered, relationships between obscure variables can be examined, and long term trends can be detected. In short data mining fulfills the expectations of data warehousing in many regards. An interesting question that immediately arises is, can data mining be done without building a data warehouse? The answer is yes. But, can data mining be done effectively in the face of no data warehouse? The answer is no. Data mining, which is defined as the process of extracting previously unknown knowledge, and detecting interesting patterns from a massive set of data [1], has been a very active research. As results, several commercial products and research prototypes are even available nowadays. However, most of these have focused on corporate data typically in Alphanumeric database. Even though relatively less research has been performed, very interesting and important studies have been published, and systems have been developed in the areas of data mining for business intelligence.

Using reviews of different products from Amazon.com, we conducted both qualitative and quantitative experiments to evaluate the performance of a BI system developed based on the framework. The results indicate that the system achieved high accuracy and coverage related to rule quality, and produced interesting and informative rules with high support and confidence values. The findings have important implications for market sentiment analysis and e-commerce reputation management [5]

Amazon launched in 1995 but could not reach the customers satisfaction and had to struggle a lot initially, After 2005 the Amazon boosted its business quickly to high profits because of their new strategies of catalog design and attractive online offers to the Customers, and proper advertising about the products to the customer the most

important factor they have considered was customer satisfaction by giving the service on time and the catalogue design and their daily, weekly offers, Deal of the day etc. Today Amazon.com is one of the largest online shopping website on the globe and it has reached This point by applying the new business techniques.

2.0 THE CONCEPT

In large transactional datasets, mining of frequent itemsets develops into new associations and correlations. With huge amounts of data continuously being collected and stored, today's market is deeply interested in mining such patterns from the databases. The discovery of interesting correlation relationships among huge amounts of business transaction records can help in many business decision-making processes such as catalog design, cross-marketing, and customer shopping behavior analysis [1]. A typical example of frequent itemset mining is market basket analysis. The process analyzes customer buying habits by finding associations between the different items that customers place in their "shopping baskets". The discovery of these associations can help retailers develop marketing strategies by gaining insight into which items are frequently purchased together by customers. For instance, if customers are buying milk, how likely are they to also buy bread (and what kind of bread) on the same trip to the supermarket? This information can lead to increased sales by helping retailers do selective marketing

Following figure describes extraction of knowledge from large data and Databases in order to apply Data mining techniques for business intelligence [3]

Applying Data Mining Techniques for E-Business from Utkal University,Bhubaneshwar,Orissa,India.E-mail: moinkku@gmail.com
2. Dr.Sateesh Kumar Pradhan *Head*,Post Graduate Department of Computer Science,Utkal University, Vani Vihar,Bhubaneswar, India 751 004.. E-mail sateeshind@yahoo.com

[1] Mohiuddin Ali khan is currently pursuing PhD program in
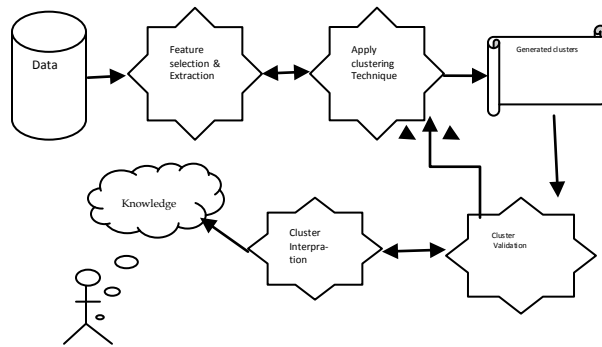
The aim of association rule mining is to discover all association rules that have support and confidence higher than the user-specified thresholds. When mining association rules from online reviews, the task is to find association rules that indicate what textual features contribute significantly to certain numerical ratings. Suppose that Fi = {f1, f2, …, f_m} denote all the textual features (single words) extracted from online reviews, represented by Di, where each review contains some of the textual features found in F_i. Also let

R = {1, 2, …, n} be the set of numerical ratings that reviewers provide.Each review can be modeled as a tagged collection of textual features. Therefore, the task of discovering decision rules from the reviews is to identify the association rules in the form: "$X \Rightarrow Y$" where $X \subseteq F$ is the set of extracted textual features, $Y \subset R$ is a numerical rating, and $X \cap Y = \phi$. To filter the potentially large number of association rules identified from the review, the threshold support and confidence are used. The support of the aforementioned rule is the proportion of the reviews containing both feature set X and rating Y. The confidence is the proportion of the reviews that are rated as Y if these reviews already contain feature set X. To extract association rules from online reviews, we selected the Apriori Algorithm [4]. in our framework due to the wide applicability and high efficiency of the algorithm. In using the algorithm, we consider the presence or absence (instead of term frequency) of textual features in the reviews because the algorithm handles only binary values of each feature. Two major steps are involved in the Apriori Algorithm: (1) to find all sets of features that have support exceeding the minimum support threshold (the resulting feature sets are called frequent feature sets) and (2) to generate from the frequent feature sets association rules that have confidence exceeding the minimum confidence threshold. Because the number of association rules can be very large, filtering is needed in three aspects. First, rules with feature

items in the consequents are removed because the consequents should contain only the numerical ratings. Second, rules that contain numerical ratings in the antecedents are removed because only textual features should contribute to the decision of numerical ratings. Third, rules with zero feature values (i.e., absence of a feature) are removed because they are much less meaningful than rules showing presence of a feature.
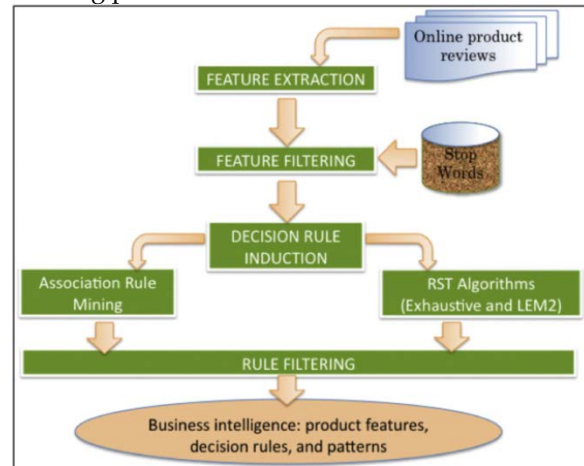
A rule-induction framework for discovering BI from online product reviews

Methodology:

In this paper, the case study of Amazon.com is used to validate and compliment findings from existing research. In information systems (IS) research, case studies are more suitable for theory building when existing theory is limited [7]. there are virtually no studies which cover both success and failure occurring within a company in a short period of time. The key advantage of studying success and failure within a company is the ability to identify the impacts of a variable in the three following scenarios: (1) if it was applied, success or failure followed, (2) when it was removed, success or failure followed, and (3) after the company failed, removal or restoration of a variable that was previously applied or removed, particularly led to regained success.

This case study in this paper can be viewed as a single case or multiple cases. From a single company perspective, Amazon was selected because it was a successful major catalogue retailer after 2005 but before it was struggling for its survival and were facing lots of issues from unable to satisfy the customers because of various strategic errors.

After 2005 the amazon boosted its business quickly to high profits because of their new strategies of catalog design and attractive online offers to the Customers, and proper advertising about the products to the customer the most Important factor

they have considered was customer satisfaction by giving the services On time and the catalogue design and their daily, weekly offers, Deal of the day etc.Today, Amazon.com is one of the largest online shopping we site on the globe and it has reached This point by applying the new business techniques. becomes successful giant in online shopping business.

Case Study on Amazon.com

Amazon.com is an American e-commerce company based in Seattle, Washington. Founded in 1994 by Jeff Bezos, and launched in 1995, Amazon.com began as an online bookstore before diversifying its product lines by adding VHSs, DVDs, music CDs, MP3s, computer software, video games, electronics, apparel, furniture, food, toys, and more. Amazon has since established separate websites in Canada, the United Kingdom, Germany, Austria, France, China, and Japan. In 1995, july amazon.com sells first book Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought, by Douglas Hofstadter.it has to struggle a lot initially as the people were not having any idea of online shopping. Since its founding, the website Amazon.com has attracted criticism and controversy from multiple sources, where the ethics of certain business practices and policies have been drawn into question. Amazon has faced numerous allegations of anti-competitive or monopolistic behaviour. This includes documented instances of price differentiation, enforcement of controversial patents against competitors, attempts to prevent discounted direct selling by publishers, and a declared intention to cease working with third-party print on demand services in favour of its own.Questions have been raised concerning the company's legal compliance. In 2002, Amazon faced a challenge to the legitimacy of their Canadian operations, although that case was subsequently dropped

In 2005, Amazon launches New Gifting Experience. Amazon.com announces record-breaking Watch Sales. Amazon announces Partnership with Celebrities and Announces Customers Ordered over 108 Million Items Worldwide During 11th Holiday Season . Designers of online shops are concerned with the effects of information load. Information load is a product of the spatial and temporal arrangements of stimuli in the webstore. Compared with conventional retail shopping, the information environment of virtual shopping is enhanced by providing additional product information such as comparative products and services, as well as various alternatives and attributes of each alternative. Retail success is no longer all

about physical stores, this is evident because of the increase in retailers now offering online store interfaces for consumers. With the growth of online shopping, comes a wealth of new market footprint coverage opportunities for stores that can appropriately cater to offshore market demands and service requirements.

Amazon and more than 2 million third-party sellers offer millions of unique, new, refurbished and used items in categories such as Books; Movies; Music; Video Games; Electronics & Computers; Home & Garden; Tools; Toys; Kids & Baby; Grocery; Health & Beauty; Clothing; Shoes & Jewelry; Health & Beauty; Sports & Outdoors; and Automotive & Industrial. Around 88,400 full-time and part-time employees worldwide [6]
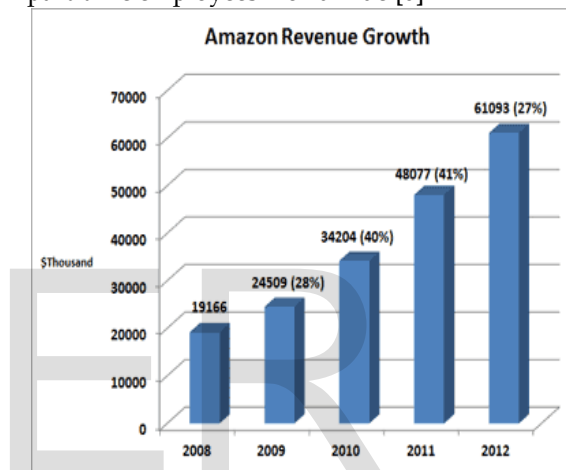
Amazon's revenue growth has been staggering. As shown in the chart below, over the last four years, its top line has grown 28%, 40%, 41%, and 27% respectively, taking it to $61 billion in annual sales and in the current year 2013 amazon.com is expected to reach $80 billion in annual sales and $100 billion mark by year 2015.

Customer Relationship Management

Customer Relationship Management is important today because it is often much more expensive to acquire new customers than to keep them. However, as relationships develop in stages, IT provides good tools to automate, maintain, and exploit them from the beginning over the lifespan of the relationships. focuses on established customers. Happily established customers are the richest source of data for mining A CRM system is a repository of customer information which contains all customer profiles. In addition to the traditional database roles, it has the capability of personalizing needs of individual customers by differentiating products or services for each unique customer. Popular strategies recommended to improve CRM include

the use of BI for price discrimination, lock-in/high switching costs, and BI tools.

Conclusions

As e-commerce continues to grow rapidly, managers and marketers should also find the results helpful in their analysis and understanding of market trend. For example, the keywords identified in the decision rules provide important clues to correlating between customers' concerns and product ratings. Managers are then able to use these keywords and rules strategically to enhance their products and to manage customer relationship. Marketers can promote their products on the Internet by strategically place the keywords in Web pages so as to increase their sites' ranking in search engines. Search engine optimization is thus supported by using the keywords and rules. Specifically, managers and marketers should find association rule mining (ARM) algorithm inducing more interesting rules, generating intuitive ideas, while RST algorithms producing more informative rules for deeper analysis.Future work can consider expanding the review datasets to cover different types of product, testing other methods for BI discovery, and studying reviews with varied distributions of ratings. Considering the rapid growth of e-commerce and the widespread use of online product reviews, companies that lack the capability of efficient and effective analysis of these reviews could lose significant competitive advantage.

REFERENCES

[1] Data Mining: Concepts and Techniques, Third Edition (The Morgan Kaufmann Series in Data Management Systems) by Jiawei Han, Micheline Kamber and Jian Pei (Jul 6, 2011)

[2] Data mining Techniques, Third Edition(Gordan S. Linoff, Michael J.A.Berry) Willy india 2012. ISBN: 978-0-470-65093-6

[3] An overview of the use of neural networks for data mining tasks, Frederic Stahl, Ivan Jordanov-2011 wiley online library

[4] Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules. In   J. B.Bocca, M. Jarke, & C. Zaniolo (Eds.), Proceedings of the 20th International Conference on Very Large Data Bases (pp. 487–499). Santiago, Chile: Morgan

Kaufmann.

[5] Aciar, S., Zhang, D., Simoff, S., & Debenham, J. (2007). Informed recommender: Basing recommendations on consumer product reviews. IEEE Intelligent Systems, 22, 39–47.

[6]http://phx.corporate-ir.net/phoenix.zhtml?c=176060&p=irol-factSheet

[7] I. Benbasat, D.K. Goldstein, M. Mead, The case research strategy in studies of information systems, MIS Quarterly 11, 1987, pp. 369–386.

[8] H. Zhang, A. Kantankanhalli, and S. Smoliar, "Automatic Partitioning of Full-Motion Video," *ACM Multimedia Systems*, vol. 1, no. 1, pp. 10-28

[9] An ontology-based for business intelligence application in a financial knowledge management system Original Research Article *Expert Systems with Applications*, *Volume 36, Issue 2, Part 2, March 2009*, *Pages 3614-3622* Hilary Cheng, Yi-Chuan Lu, Calvin Sheu

[10] Analysis on the Effect of Data-mining to Supermarket, The 5th International Conference on Computer Science & Education, Hefei, China. August 24–27, 2010, Yunkang Yue, Tiebao Zhang, Xingbing Xu

[11] Application of Data Mining in Supermarket, 978-1-61284-459-6/11 ©2011 IEEE, Zhang Aiguo, Jiang Lanling, Song Ping

[12] Towards Better Modeling of Supermarkets, *www.ismll.unihildesheim.de/pub/pdfs/Buza2010-Supermarket.pdf,* K. Buza, A. Buza and P.B. Kis

[13] A quantitative correlation coefficient mining method for business intelligence in small and medium enterprises of trading business, Expert Systems with Applications, Volume 39, Issue 7, 1 June 2012, C.F. Cheung, F.L. Li

[14] Data Mining: Multimedia, Soft Computing, and Bioinformatics (Hardcover) by Sushmita Mitra (Author), Tinku Acharya (Author)

[15] D**ata** Mining on Multimedia Data, Author Petra Perner**,** Publisher Springer-Verlag New York, Inc. Secaucus, NJ, USA.

[16] Multimedia Data Mining and Knowledge Discovery

Petrushin, Valery A.; Khan, Latifur (Eds.) 2007.

[17] H. Zhang, A. Kantankanhalli, and S. Smoliar, "Automatic Partitioning of Full-Motion Video," *ACM Multimedia Systems*, vol. 1, no. 1, pp. 10-28

[18] F. Kokkoras, H. Jiang, I. Vlahavas, A. Elmagarmid, E. Houstis, and W. Aref, "Smart VideoText: A Video Data Model Based on Conceptual

Graphs," *ACM/Springer Multimedia Systems,* vol. 8, no. 4, pp. 328-338, 2002.

[19] Y. Chen, W. Gao, Z. Wang, J. Miao, and D. Jiang. Mining audio/visual database for speech driven face animation. In Proc. of International Conference on Systems, Man and Cybernetics, pages 2638–2643,2001.

IJSER