|| Volume 3 || Issue 9 || September 2018 || ISSN (Online) 2456-0774
INTERNATIONAL JOURNAL OF ADVANCE SCIENTIFIC RESEARCH
AND ENGINEERING TRENDS

Survey on Techniques for Mining Industries and Competitors

Jyoti Pandurang Mohite[1], Prof. Dr. B. K. Sarkar[2]
Department of Computer Engineering,Padmbhushan Vasantdada Patil Institute of Technology, Bavdhan Pune, Maharashtra[1 2]

Abstract— **In the present world Completive business, the achievement is totally in light of the capacity to make a thing more engaging clients than the opposition. Huge information is a trendy expression that is utilized for expansive size information which incorporates organized information, semi-organized information and unstructured information. In the current competitive business scenario, there is a need to analyse the competitive features and factors of an item that most affect its competitiveness. We study on a formal definition of competitiveness between two items, which we validated both quantitatively and qualitatively. Finally the paper provides the challenges and importance in the competitor mining tasks with optimal improvements.**

Keywords: *Competitiveness, Data Mining, Quantitatively, Qualitatively and Web Mining.*

I INTRODUCTION

The strategic importance of detecting and observing business competitors is an inevitable research, which motivated by several business challenges. Monitoring and identifying firm's competitors have studied in the earlier work. Data mining is the optimal way of handling such huge information's for mining competitors. Item reviews form online offer rich information about customers' opinions and interest to get a general idea regarding competitors. However, it is generally difficult to understand all reviews in different websites for competitive products and obtain insightful suggestions manually. In the earlier works in the literatures, many authors analyzed such big customer data intelligently and efficiently [1] [2] [3].For example, a lot of studies about online reviews were stated to gather item opinion analysis from online reviews in different levels. However, most researchers in this field ignore how to make their findings be seamlessly utilized to the competitor mining process.

Recently, a limited number of researches were noted to utilize the latest development in artificial intelligence (AI) and data mining in the e-commerce applications [4]. These studies help designers to understand a large amount of customer requirements in online reviews for product improvements. But, these discussions are far from sufficient and some potential problems. These have not been fully investigated such as, with product online reviews, how to conduct a thorough competitor analysis. Actually, in a typical scenario of a customer-driven new product design (NPD), the strengths and weakness are often analyzed exhaustively for probable opportunities to succeed in the fierce market competition.

Mostly, competitiveness paradigm is based on the following observation: the competitiveness between two items is based on whether they compete for the attention and business of the same groups of customers (i.e. the same market segments). For example, two restaurants that exist in different countries are obviously not competitive, since there is no overlap between their target groups. Consider the example shown in Figure 1.

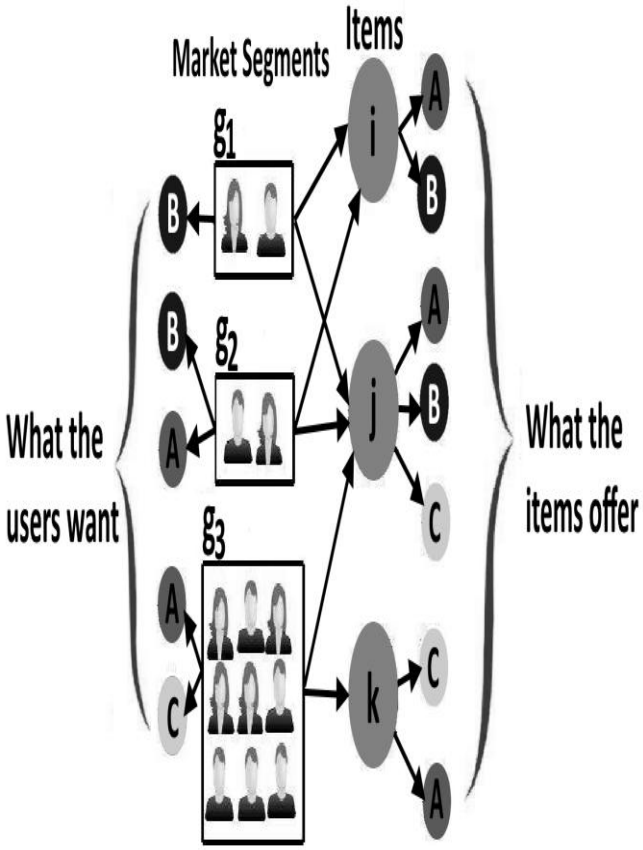

Figure 1 Example of our Competitiveness Paradigm

II LITERATURE REVIEW

This literature review provides the various methodologies implemented to mine competitors with reference to customer lifetime value, relationship, opinion and behaviour using data mining techniques:

In this paper [1] propose and evaluate an approach that exploits company citations in online news to create an intercompany network whose structural attributes are used to infer competitor relationships between companies. As noted earlier the company citations in news may not necessarily represent competitor relationships. They find that such a

citation-based network carries latent information and the structural properties can be used to infer competitor relationships. In this they present evaluation prompt three broad observations. First, the intercompany network captures signals about competitor relationships. Second, the structural attributes, when combined in various types of classification models, infer competitor relationships. For imbalanced portions of the data, for this require more advanced modeling techniques (e.g., data segmentation, DTA) to achieve reasonable performance. Third, quantify the degree to which two commercial data sources are incomplete in their coverage of competitors and estimate the extent to which approach extends them while still maintaining adequate performance.

In this paper [2] they present proposed ranking methods using data from location-based social media by turning check-ins into competitions between restaurants and their neighbours. The first assumption is that the stores to be ranked are of the same type. The second assumption is that there are competitions between stores that are near each other. They have evaluated performance on the real dataset from Foursquare and found probability options P_{NAR} and P_{NUR} behave similarly. They have also qualitatively analysed the results through cases studies and verify the correctness of this model via the "ground truth".

This paper [3] proposes an improved method about comparative sentence recognition. They construct the mixed rule base combined with entity name dictionary. Compared with previous study, the proposed method can achieve better recognition precision with less artificial work and supervision. As shown in the experimental result, this method outperforms many methods. By classifying sentences directly into 'Equitize', 'Non-Equal' and 'Non-Comparison', the recognition result is good, then the mining work can be simplified. Finally, based on above method, a visualized restaurant competitiveness analysis is made which testifies the practical value of this paper. To realize real business intelligence, the technology of comparative sentence recognition can be combined with product feature mining, sentiment analysis, and comparative network construction etc. As a result, the potential useful business value can be extracted and assist the decision makers to make correct decision or judgment.

In this paper [4] they proposed metrics of online isomorphism based on content and linkage structure of firm's web sites. Then they utilize the presence of online isomorphism for the competitor identification problem. Competitor identification has been highlighted as a critical and challenging step in competitive analysis and strategy but there is limited literature on automatic identification of competitors. They use online metrics as inputs in predictive models that classify pairs of firms as competitors or non-competitors. They find the resulting predictive models provide high accuracy, F measure, and AUC. The models also indicate that using a variety of web metrics as suggested by us provides a clear benefit as compared to just using the individual control metrics that are derived from previous literature. The benefit is observed for data sets with different proportions of competitor and non-competitor pairs of firms. They benchmark the predictive models that use online metrics against those that use online metrics.

Managers routinely seek to understand firm performance relative to the competitors. Recently, competitive intelligence (CI) has emerged as an important area within business intelligence (BI) where the emphasis is on understanding and measuring a firm's external competitive environment. A requirement of such systems is the availability of the rich data about a firm's competitors, which is typically hard to acquire. This paper [5] proposes a method to incorporate competitive intelligence in BI systems by using less granular and aggregate data, which is usually easier to acquire. They motivate, develop, and validate an approach to infer key competitive measures about customer activities without requiring detailed cross-firm data. Instead, in this method derives these competitive measures for online firms from simple "site-centric" data that are commonly available, augmented with aggregate data summaries that may be obtained from syndicated data providers.

In this [6] they proposed models to define and extract opinions from web documents present a simple, yet relatively effective manner of transforming the unstructured data about opinions available on the web. However, the algorithm for aspect expressions extraction, based on frequent nouns and NPs appearing in reviews, achieved a poor performance in the tourism domain. Results show that, in fact, multiple expressions are used to denote the same attribute or component of a tourism product in reviews. Therefore, not only the most frequent words need to be considered when extracting aspect expressions in order to achieve a better recall for this task.

III COMPARATIVE ANANLYSIS

Table 1: Survey Table

Sr. No	Paper Name	Author	Method Proposed	Limitations
1.	Mining competitor relationships from online news: A network-based approach.	ZhongmingMaa, Gautam Pant, Olivia R.L. Sheng	Approach that uses graph-theoretic measures and machine learning techniques to infer competitor relationships on the basis of structure of an intercompany network derived from company citations (cooccurrence) in online news articles.	In this research to address problems with a binary class label.

2.	Mining business competitiveness from user visitation data	DOAN T. N., F. C. T. Chua and E.-P. Lim	Ranking methods using data from location-based social media by turning check-ins into competitions between restaurants and their neighbors.	For this they do not considered features like the distance between user and store; comments and reviews from users; social relationships.
3.	Mining comparative opinions from customer reviews for competitive intelligence	Song Gao, H. Wang, Y. Song and Ting Lu.	Mining comparative sentences based on the achievements of linguistic study.	In experimental results detected that few cases of Non-Comparison sentences are classified as Comparison sentences.
4.	Web footprints of firms: Using online isomorphism for competitor identification	Gautam Pant and Olivia R. L. Sheng	Online metrics based on the content, in-links, and out-links of rms' web sites to measure the presence of online isomorphism as well as uncover its utility in predicting competitor relationships.	A strong need for alleviating the complexity of the problem through automated methods.
5.	From business intelligence to competitive intelligence: Inferring competitive measures using augmented site-centric data.	Z. Zheng, P. Fader, and B. Padmanabhan	Infer key competitive measures about customer activities without requiring detailed cross-firm data	Overall performance of this techniques is not good.
6.	Identifying customer preferences about tourism products using an aspect-based opinion mining approach	Edison Marrese-Taylora, Juan D. Velsqueza, Felipe Bravo-Marquezb, Yutaka Matsuoc	Aspect-based opinion mining technique to apply it to the tourism domain	The algorithms were only capable of extracting 35% of the explicit aspect expressions.

IV CONCLUSION

Data mining has importance regarding finding the patterns, forecasting, discovery of knowledge etc., in different business domains. Machine learning algorithms are widely used in various applications. Every business related application uses data mining techniques. To improve such business or providing appropriate competitors for the business to the user need the support of web mining techniques.The competitor mining is one such a way to analyze competitors for the selected items. In this we study on a formal definition of competitiveness between two items, which validated both quantitatively and qualitatively. Here considered a number of factors that have been largely overlooked in the past, such as the position of the items in the multi-dimensional feature space and the preferences and opinions of the users.

REFERENCES

1. ZhongmingMaa, Gautam Pant, Olivia R.L. Sheng,"Mining competitor relationships from online news: A network-based approach"-Electronic Commerce Research and Applications, Elsevier-2011.
2. DOAN T. N., F. C. T. Chua and E.-P. Lim "Mining business competitiveness from user visitation data",-Information SystemsResearch, vol. 26, no. 1, pp. 188–209, Springer-2015.
3. Song Gao, H. Wang, Y. Song and Ting Lu., "Mining comparative opinions from customer reviews for competitive intelligence", Decis. Support Syst., AIS Electronic Library 2011.
4. Gautam Pant and Olivia R. L. Sheng, "Web footprints of firms: Using online isomorphism for competitor identification", Information Systems Research, vol. 26, no. 1, pp. 188–209, 2015.
5. Z. Zheng, P. Fader, and B. Padmanabhan, "From business intelligence to competitive intelligence: Inferring competitive measures using augmented site-centric data," Information Systems Research, vol. 23, no. 3-part-1, pp. 698–720, 2012.
6. Edison Marrese-Taylora, Juan D. Velsqueza, Felipe Bravo-Marquezb, Yutaka Matsuoc, "Identifying customer preferences about tourism products using an aspect-based opinion mining approach", Procedia Computer Science, vol. 22, pp. 182–191, Elsevier-2013.