

# AUGMENTED COMPETITOR MINING WITH C-MINER ALGORITHM BASED ON PRODUCT REVIEWS

Gokkul V.S<sup>#1</sup> and Angel Pemala G<sup>\*2</sup>

<sup>#</sup> Assistant Professor, Department of Computer Science & Engineering, J.K.K.N. College of Engineering and Technology, Komarapalayam, Tamil Nadu, India.

<sup>\*</sup> Student, M.E. - Department of Computer Science & Engineering, J.K.K.N. College of Engineering and Technology, Komarapalayam, Tamil Nadu, India.

**Abstract**— In the current competitive business scenario, there is a need to analyze the competitive features and factors of an item that most affect its competitiveness. The evaluation of competitiveness always uses the customer opinions in terms of reviews, ratings and abundant source of information's from the web and other sources. This paper develops an augmented competitor mining using product reviews. The itemsets are analyzed for selecting the relevant features. Using the c-miner, the frequent items are discovered and then represented by skyline operators. However, if all customer data is inserted into a database, the resulting records will provide a detailed profile of these customers and their interactions with one another, and will be an important resource for businesses that wish to probe customer data, customer needs, and customer satisfaction levels. Experimental analysis has shown the efficiency of the proposed algorithm.

**Index Terms**— Competitive, relevant features, C-miner, skyline and frequent items

## 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 [1].

In the earlier works in the literatures, many authors analyzed such big customer data intelligently and efficiently [2]. 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 [3]. 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.

The data collected from the web are sometimes semi-structured or unstructured. The semi-structured data's are in the format of XML, JSON etc., the unstructured data sources are in a different format, which is not fall under any predefined category [4]. When managing thousands of customers, business will have difficulty sustaining the rising costs created by interactions among people. However, if all customer data is inserted into a database, the resulting records will provide a detailed profile of these customers and their interactions with one another, and will be an important resource for businesses that wish to probe customer data, customer needs, and customer satisfaction levels.

The rest of the paper is organized as follows: Section II describes related work; Section III presents the proposed work; Section IV presents the experimental analysis and results and Section V presents the conclusion.

## II. RELATED WORK

This section presents the prior work suggested in competitor mining. Authors in [5] developed an automatic system that discovers competing companies from public information sources. In this system data is crawled from text and it uses transformation oriented learning to obtain appropriate data normalization, combines structured and unstructured information sources, uses probabilistic modeling to represent models of linked data, and succeeds in autonomously discovering competitors. Bayesian network for competitor identification technique is used. The authors also introduced the iterative graph reconstruction process for inference in relational data [6], and shown that it leads to improvements in performance. To find the competitors, the authors used machine learning algorithms and probabilistic approaches. They also validate system results and deploy it on the web as a powerful analytic tool for individual and institutional investors. However, the technique has many

problems like finding alliances and market demands using the machine learning approach.

In the paper [7], authors presented a formal definition of the competitiveness between two items. Authors used many domains and handled many shortcomings of previous works. In this paper, the author considered the position of the items in the multi-dimensional feature space, and the preferences and opinions of the users. However, the technique addressed many problems like finding the top-k competitors of a given item and handling structured data. Authors in [8] proposed a new online metrics for competitor relationship predicting. This is based on the content, firm links and website log to measure the presence of online isomorphism, here the Competitive isomorphism, which is a phenomenon of competing firms becoming similar as they mimic each other under common market services [9]. Through different analysis they find that predictive models for competitor identification based on online metrics are largely superior to those using offline data. The technique is combined the online and offline metrics to boost the predictive performance. The system also performed the ranking process with the considerations of likelihood.

Several works in the same strategy in literature have discussed the need for accurate identification of competitors and provided theoretical frameworks for that. Given the expected isomorphism between competing firms, the process of competitor identification through pair-wise analysis [10] of similarities between focal and target firms is well founded. The unit of analysis is a pair of firms since competitor relationship is seen as a unique interaction between the pair. Authors in [11] have suggested frameworks for manual identification of competitors. The manual nature of these frameworks makes them very costly for competitor identification over a large number of focal and target firms, and over time [12].

In the paper[13], authors attempts to accomplish a novel task of mining competitive information with respect to an entity , the entity such as a company, product or person from the web. The authors proposed an algorithm called “CoMiner”, which first extracts a set of comparative candidates of the input entity and then ranks them according to the comparability, and finally extracts the competitive fields. But the CoMiner [14] specifically developed to support for specific domain. However the effort for the further domains is still challenging. Authors in [15] have proposed ranking methods to give the competitor in a ranked way. They have used data from location based social media. Authors proposed the use of Page-Rank model and it’s variant to obtain the Competitive Rank of firms. However mining competitors from the social media developed many privacy related issues.

### III. PROPOSED WORK

This section presents the working of our proposed model. The proposed model composes of four phases and it’s explained as follows:

#### A. Administrator phase:

Administrator phase is the first step who takes the responsibility of the administering the user’s activities and upload the items like hotels, recipes and cinemas etc. They also check the profile details, customer queries and their

interests using C-miner. C-miner falls under the class of frequent sequence mining. It operates on discovering correlations in data blocks. Each data item is mapped with the blocks. In similar way, the search operation is processed in sequence to sequence mapping. By doing so, the frequently searched item are extracted and sorted to top k itemsets.

#### B. Customer phase:

Customer phase is the second step which operates on requirements of customers. Based on the queries given by the customers, the results will be displayed. For each item, the database is created. Relied upon the customer requirements, the requested data item is displayed.

#### C. C-miner algorithm:

This step depicts how the top k itemsets are retrieved. C-miner algorithm is used to discover the top k competitors for given item. It iterates like skyline pyramid which reduces the dimensionality. Each item is mapped with its corresponding blocks. The items are selected on basis of correlation score. Therefore, during competitor mining process, unstructured data is not taken into account and much valuable service information is lost. Structured systems are those where the data and the computing activity is predetermined and well-defined. Unstructured systems are those that have no predetermined form or structure and are usually full of textual data.

#### D. Skyline operator:

The skyline operator is performed for the coordinate points of data items. It helps to determine the subset of points which dominates the other set of points. Generally, the skyline function is given as sky (I). Given the skyline Sky (I) of a set of items I and an item  $i \in I$ , let Y contain the k items from Sky (I) that are most competitive with i. Then, an item  $j \in I$  can only be in the top-k competitors of i, if  $j \in Y$  or if j is dominated by one of the items in Y.

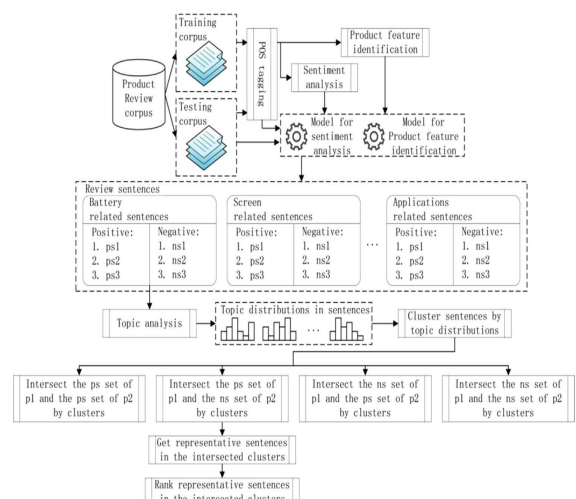


Fig.1 Proposed architecture

### IV. EXPERIMENTAL RESULTS AND ANALYSIS

This section presents the experimental analysis of our proposed work.



Fig.2 Administrator's page

The fig.2 explains the role of administrator to derive relevant information from unstructured data.

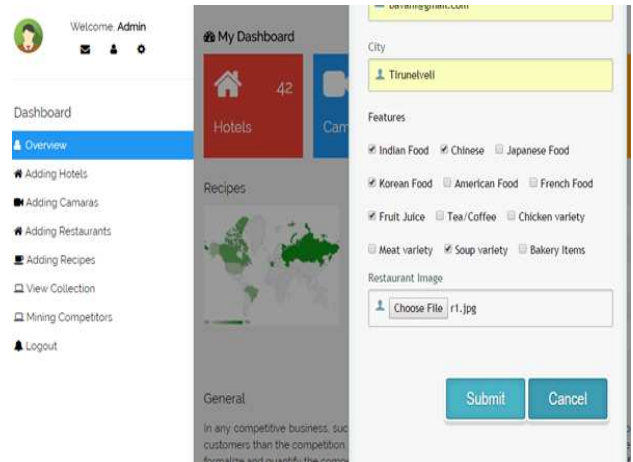


Fig.5. Adding items to the hotel section

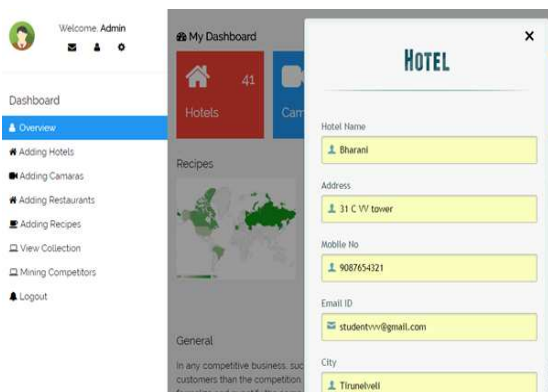


Fig.3 Dataset creation

The fig.3 presents the creation of dataset for the sample instance, hotels.

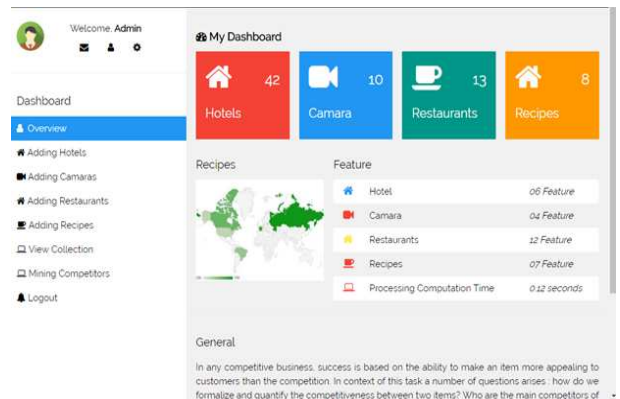


Fig.6. Viewing the list of items in database

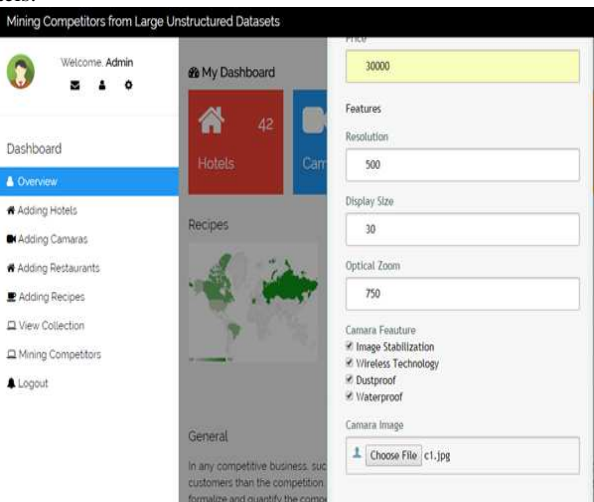


Fig.4. Creating database for the camera section

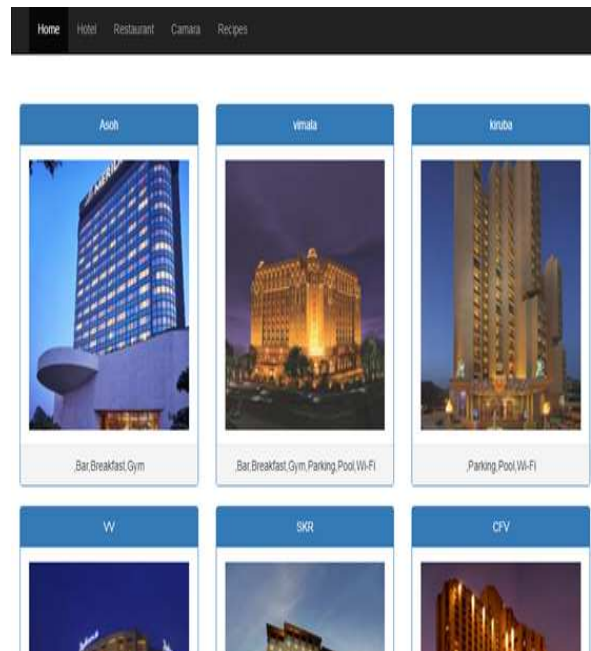


Fig.7. Viewing the list of hotels



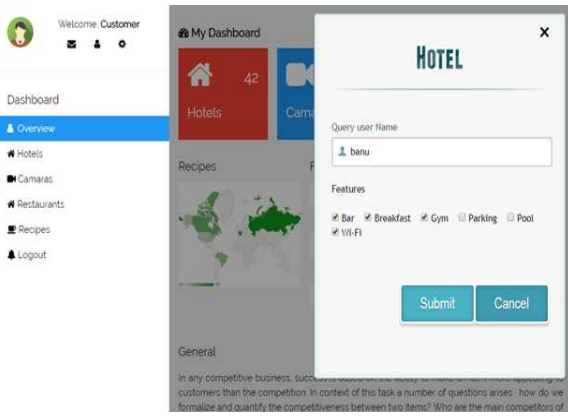


Fig. 8. User's entering the queries

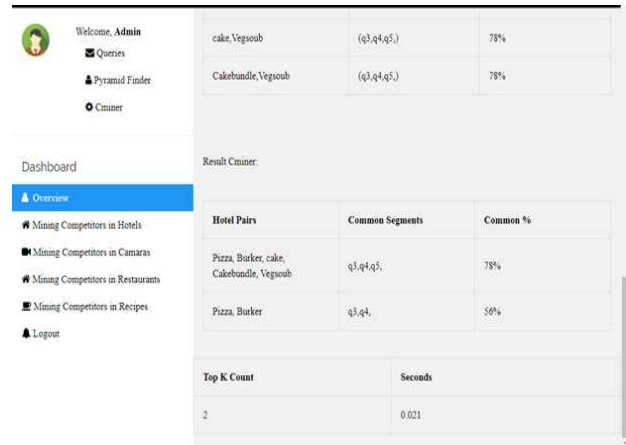


Fig.12 Finding top k competitors

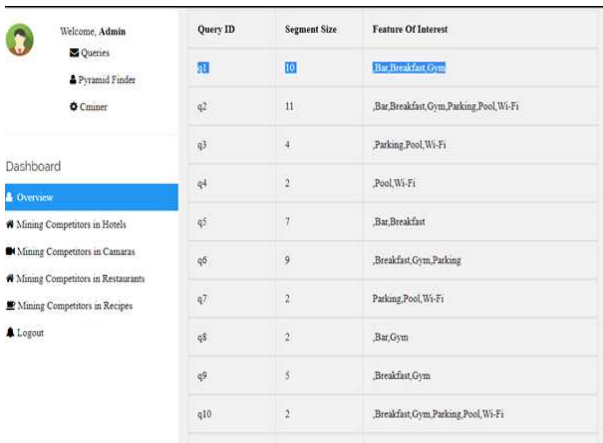


Fig.9. Creating the data into blocks using c-miner

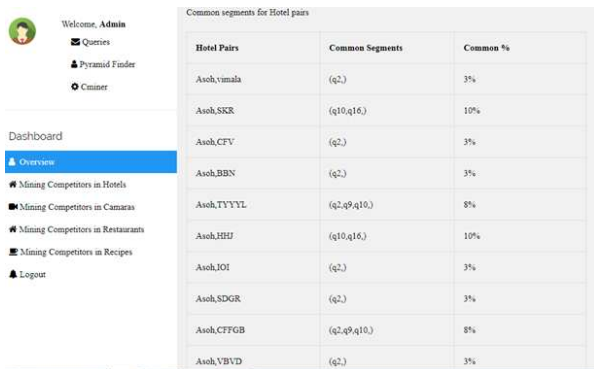


Fig.10. Working of pyramid miner

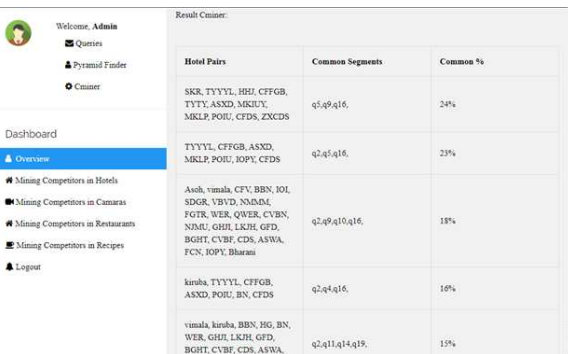


Fig.11 Arrangements of unstructured data into structured datasets

## V. 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 paper, we propose an enhanced c-miner algorithm which arranges the unstructured data into structured data. Based on the customer reviews, the itemset are selected. Since the aim is to analyze the competitors, the relevant features are picked from the reviews. With the help of c-miner and skyline approach, the information is represented and then top k item is extracted from the available resources. Experimental analysis has shown the efficiency of the proposed algorithm.

## REFERENCES

- [1] Ding, X., Liu, B., Yu, P.S., 2008. A holistic lexicon-based approach to opinion mining. In: Proceedings of the WSDM'08.
- [2] Abbasi, A., Chen, H., Salem, A., 2008. Sentiment analysis in multiple languages: feature selection for opinion classification in web forums. ACM Trans. Inf. Syst. 26 (3), 12:1–12:34
- [3] Chen, L., Qi, L., Wang, F., 2012. Comparison of feature-level learning methods for mining online consumer reviews. Expert Syst. Appl. 39 (10), 9588–9601.
- [4] Zhan, J., Loh, H.T., Liu, Y., 2009. Gather customer concerns from online product reviews – a text summarization approach. Expert Syst. Appl. 36 (2 Part 1), 2107–2115
- [5] Jin, Jian, Ping Ji, and Rui Gu. "Identifying comparative customer requirements from product online reviews for competitor analysis." Engineering Applications of Artificial Intelligence 49 (2016): 61-73.
- [6] Saxena, Prateek, David Molnar, and Benjamin Livshits. "SCRIPTGARD: automatic context-sensitive sanitization for largescale legacy web applications." Proceedings of the 18th ACM conference on Computer and communications security. ACM, 2011.
- [7] Ghamisi, Pedram, Jon Atli Benediktsson, and Johannes R. Sveinsson. "Automatic spectral-spatial classification framework based on attribute profiles and supervised feature extraction." IEEE Transactions on Geoscience and Remote Sensing 52.9 (2014): 5771-5782.
- [8] Petrucci, Giulio. "Information extraction for learning expressive ontologies." In European Semantic Web Conference, pp. 740-750. Springer, Cham, 2015.
- [9] Gentile, Anna Lisa, Ziqi Zhang, Isabelle Augenstein, and Fabio Ciravegna. "Unsupervised wrapper induction using linked data." In Proceedings of the seventh international conference on Knowledge capture, pp. 41-48. ACM, 2013.

- [10] K. Simon and G. Lausen, "ViPER: Augmenting Automatic Information Extraction with Visual Perceptions," Proc. 14th ACM Int'l Conf. Information and Knowledge Management, pp. 381-388, 2005
- [11] Zelenko, Dmitry, and Oleg Semin. "Automatic competitor identification from public information sources." *International Journal of Computational Intelligence and Applications* 2.03 (2002): 287-294.
- [12] Lappas, Theodoros, George Valkanas, and Dimitrios Gunopulos. "Efficient and domain-invariant competitor mining." Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2012.
- [13] Valkanas, George, Theodoros Lappas, and Dimitrios Gunopulos. "Mining Competitors from Large Unstructured Datasets." *IEEE Transactions on Knowledge and Data Engineering* (2017).
- [14] Pant, Gautam, and Olivia RL Sheng. "Web footprints of firms: Using online isomorphism for competitor identification." *Information Systems Research* 26.1 (2015): 188-209.
- [15] Bergen, Mark, and Margaret A. Peteraf. "Competitor identification and competitor analysis: a broad-based managerial approach." *Managerial and decision economics* 23.4-5 (2002): 157-169.