

“Mining Competitors for E-commerce Unstructured Datasets”

¹Shruti Kulkarni, ²Prajakta Jadhav, ³Sakshi Kankanwar, ⁴Gauri Bhagwat, ⁵Prof. Ila Savant

^{1,2,3,4} Student, Computer Engineering, MMCOE, Pune, India,

⁵Professor, Computer Engineering, MMCOE, Pune, India

Department of Computer Technology, Pune

Abstract: *Along line of research has demonstrated the strategic importance of identifying and monitoring a firm's competitors. Motivated by this problem, the marketing and management community have focused on empirical methods for competitor identification as well as on methods for analyzing known competitors. Extant research on the former has focused on mining comparative expressions (e.g. Item A is better than Item B) from the Web or other textual sources. Even though such expressions can indeed be indicators of competitiveness, they are absent in many domains.*

We present a formal definition of the competitiveness between two items, based on the market segments that they can both covers. Our evaluation of competitiveness utilizes customer reviews, an abundant source of information that is available in a wide range of domains. We present efficient methods for evaluating competitiveness in large review datasets and address the natural problem of finding the top-k competitors of a given item. Finally, we evaluate the quality of our results and the scalability of our approach using multiple datasets from different domains.

Keyword- *Data mining, Web mining, Information Search and Retrieval, Electronic commerce, CMiner, sentimental analysis.*

I. INTRODUCTION

Along line of research has demonstrated the strategic importance of identifying and monitoring a firm's competitors. Motivated by this problem, the marketing and management community have focused on empirical methods for competitor identification as well as on methods for analyzing known competitors. Extant research on the former has focused on mining comparative expressions (e.g. Item A is better than Item B) from the Web or other textual sources. Even though such expressions can indeed be indicators of competitiveness, they are absent in many domains. For instance, consider the domain of vacation packages (e.g. flight-hotel-car combinations). In this case, items have no assigned name by which they can be queried or compared with each other. Further, the frequency of textual comparative evidence can vary greatly across domains. For example, when comparing brand names at the firm level (e.g. “Google vs. Yahoo” or “Sony vs. Panasonic”), it is indeed likely that comparative patterns can be found by simply querying the web. However, it is easy to identify mainstream domains where such evidence is extremely scarce, such as shoes, jewelry, hotels, restaurants, and furniture. Motivated by these shortcomings, we propose a new formalization of the competitiveness between two items, based on the market segments that they can both covers.

II. PROBLEM STATEMENT

1. Problem Statement:

Many researcherrs conducted the experiment on item feature extracting data and competitor analysis. The problem of automatically extracting data records that are related to the user given may have two types of documents like structured and unstructured. Handling unstructured dataset in the web repository may always create many challenges. This method performs a novel data extraction by means of identifying the data regions and merging followed by segmentation and query result set identification of the records. The extracted data should be converted into structured one and nested structures are

identified. Even though the earlier work CMiner provides good result, it still limits in few cases like domain specification, data handling and dynamic data management issues.

2. Goals & Objectives:

- Satisfy the user and company via finding best competitor.
- Efficient way to find the competitor based on product and product feature.
- Develop system who can find the best competitor in unstructured data.

III. PROPOSED SYSTEM

In this paper, we present a formal definition of the competitiveness between different items, item features and item competitors based on the market segments that they can both covers. Our evaluation of competitiveness utilizes customer reviews, an abundant source of information that is available in a wide range of domains. We present efficient methods for evaluating competitiveness in large review datasets and address the natural problem of finding the top-k competitors of a given item. Finally, we evaluate the quality of our results and the scalability of our approach using multiple datasets from different domains.

A. SYSTEM ARCHITECTURE

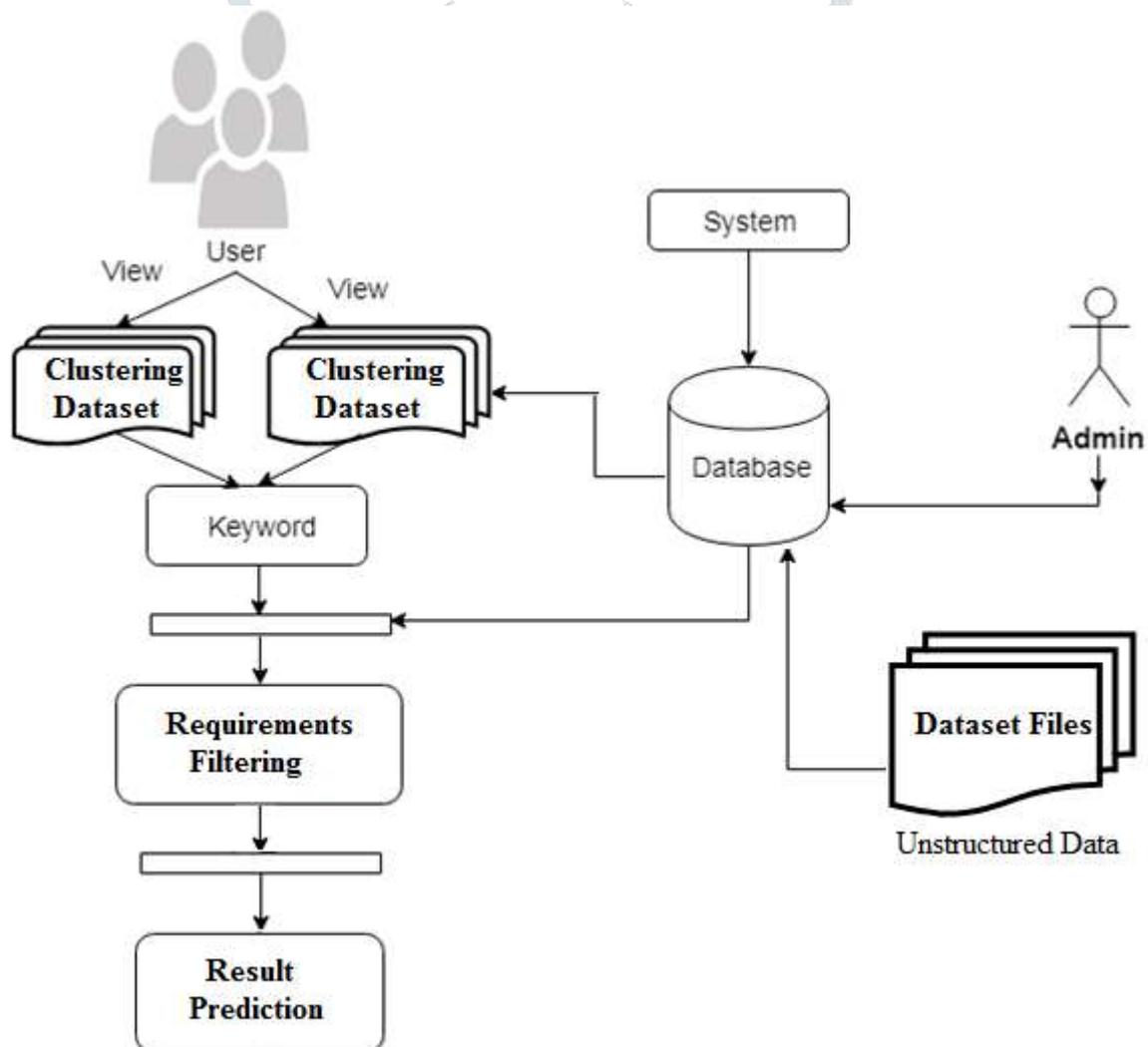


Fig.1: System Architecture

IV. Literature Survey

Sr. No.	Title	Authors	Description	Modules/Algo.
01	“A probabilistic rating inference framework for mining user preferences from reviews,”	C. W.-K. Leung, S. C.-F. Chan, F.-L. Chung, and G. Ngai,	In this paper formally define reverse top-k queries and introduce two versions of the query, monochromatic and biochromatic. First, we provide a geometric interpretation of the monochromatic reverse top-k query to acquire an intuition of the solution space. Then, we study in detail the case of biochromatic reverse top-k query, and we propose two techniques for query processing, namely an efficient threshold-based algorithm and an algorithm based on materialized reverse top-k views.	Top-k queries
02	A survey of skyline processing in highly distributed environments	Katja Hose · Akrivi Vlachou	data management and storage have become increasingly distributed. Advanced query operators, such as skyline queries, are necessary in order to help users to handle the huge amount of available data by identifying a set of interesting data objects.	Review
03	Identifying Customer Preferences about Tourism Products using an Aspect-Based Opinion Mining Approach	Edison Marrese-Taylor, Juan D. Velásquez, Felipe Bravo-Marquez, Yutaka	An experiment is also conducted, using hotel and restaurant reviews obtained from TripAdvisor, to evaluate our proposals. Results showed that tourism product reviews available on web sites contain valuable information about customer preferences that can be extracted using an aspect-based opinion mining approach	Customer preferences, natural language processing, web mining

Table.1: Literature Survey

B. REQUIREMENTS SOFTWARE AND HARDWARE:

Hardware Requirements Specification:

There should be required devices to interact with software.

- System : Pentium IV 2.4 GHz.
- Hard Disk : 40 GB.
- Ram : 256 Mb.

Software Requirements Specification:

- Operating system : Windows XP/7.
- Coding Language : JAVA
- IDE : Java eclipse

- Web server : Apache Tomcat 7.

V. CONCLUSION AND FUTURE WORK

This paper proposes 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. This paper evaluation prompts 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, it requires more advanced modeling techniques (e.g., data segmentation, DTA) to achieve reasonable performance. Third is quantifies the degree to which two commercial data sources are incomplete in their coverage of competitors and estimate the extent to which our approach extends them while still maintaining adequate performance.

ACKNOWLEDGEMENT

Thanks to our Gguide **Prof. Ila Sawant**, and MMCOE college management for providing resources and helping us in all possible ways. We also thank readers of this journal for showing interest in this topic and contributing towards enhancement of this topic as well.

REFERENCES

- [1] George Valkanas, TheodorosLappas, and DimitriosGunopulos:” Mining Competitors from Large Unstructured Datasets”, IEEE Transactions on Knowledge and Data Engineering, 2017.
- [2] M. E. Porter, Competitive Strategy: Techniques for Analyzing Industries and Competitors. Free Press, 1980.
- [3] R. Deshpand and H. Gatingon, “Competitive analysis,” Marketing Letters, 1994.
- [4] B. H. Clark and D. B. Montgomery, “Managerial Identification of Competitors,” Journal of Marketing, 1999.
- [5] W. T. Few, “Managerial competitor identification: Integrating the categorization, economic and organizational identity perspectives,” Doctoral Dissertation, 2007.
- [6] M. Bergen and M. A. Peteraf, “Competitor identification and competitor analysis: a broad-based managerial approach,” Managerialand Decision Economics, 2002.
- [7] J. F. Porac and H. Thomas, “Taxonomic mental models in competitor definition,” The Academy of Management Review, 2008.
- [8] M.-J. Chen, “Competitor analysis and interfirm rivalry: Toward a theoretical integration,” Academy of Management Review, 1996.

- [9] R. Li, S. Bao, J. Wang, Y. Yu, and Y. Cao, "Cominer: An effective algorithm for mining competitors from the web," in ICDM, 2006.
- [10] Z. Ma, G. Pant, and O. R. L. Sheng, "Mining competitor relationships from online news: A network-based approach," Electronic Commerce Research and Applications, 2011.
- [11] R. Li, S. Bao, J. Wang, Y. Liu, and Y. Yu, "Web scale competitor discovery using mutual information," in ADMA, 2006.
- [12] S. Bao, R. Li, Y. Yu, and Y. Cao, "Competitor mining with the web," IEEE Trans. Knowl. Data Eng., 2008.

