

Yashoda Multidisciplinary Research Journal

IDENTIFICATION OF TOP-K COMPETITORS USING MINING

Ms. Yojana Gaikwad

Department of Computer Engineering

JSPM's BSIOTR, Wagholi

yojanagaikwad101@gmail.com

Abstract- In many of the Competitive business, success is based on the ability to satisfies the customers requirement than competition. A lots of questions arise in this context like how do we formalize and quantify the competitiveness between items? Who are main competitors of a given item? What are the characteristics of an item that most affect its competitiveness? Despite the impact and relevance of this problem to many domains, only a limited amount of work has been devoted toward an effective solution. In this paperWe give effective methods which can search product in large review datasets and can help to solve the problem of finding the top-k competitors of a given item and when we get products related to searched query we are implementing algorithm called Cminer which is going to help to get best productand its competitors which are related to query searched andbased on reviews of products user can buy good product fromlarge unstructured dataset and likewise we are going to help userto get best products based on its search.

Index Terms—Index TermsData mining, Electronic commerce, Information Search and Retrieval, Web mining

I. INTRODUCTION

This paper is based on suggesting good product among all to customer according to his requirement. Whenever the customer tries to buy a new product and he is aware of his requirement then he has to enter the specific product keyword in search box and using these keywords the proposed system finds out best and suggests the products which are approximately close to searched product. But as there is large unstructured dataset present it can be little bit difficult to find required product. So, we have use the Cminer algorithm which requires input as features of different products which are in competition with other

products and which is required by different users and according to different users requirement we take query from user about what they want which has features of product they are interested then we take this query as input to system and we find keywords of products using this. We search in database match keywords of products and get a list of products with matched one then by using Cminer algorithm we find the best top K competitors among all choices.

A. Motivation of Project

In any business where there is lot of competition so seller

Yashoda Multidisciplinary Research Journal

focus on how to sell product to customer and so we focus on how we help seller how he can sell product to customer in business competition and which will suggest buyer which can be good product for him/her.

B. Existing Methodology of project

What was use to occur that when customer tries to buy new product he /she is not aware of products which can be good for him so they used to buy them on basis of reviews and not on what is there need and they can buy product which can be costly and features can be not what he wants.so to solve this problem we are using new method to suggest relevant product.

C. Proposed Methodology

In proposed system we are using cminer called algorithm which helps to suggest user most relevant product according to his needs. The input to cminer includes the set of items, the set of features, the item of interest, the set of queries and their probabilities and the skyline pyramid DI. We will try to run algorithm parallely for fast retrieval of top-k items and then first retrieves the items that dominate interested item and then will merge all to get result in less time. Then it sets lower bound so that it will help to find competitors related to interested item. Then use collected top k items to find the relevant items and top competitors that can be used in place of interested items.

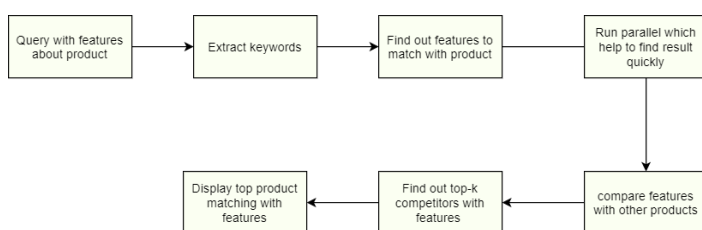
Fig. Framework of the Proposed Work

II. LITERATURE SURVEY

1. In Web Footprints of Firms: Using Online Isomorphism for Competitor Identification by Gautam Pant, Olivia R. L. Sheng work they found that predictive models which work for identification of competitor using online numbers which are more effective than those which are used as offline data. And also, combining online and offline metrics can increase the prediction process. And they also found that such models are useful in identifying nuances of competitor relationships. But this work will help domain experts, which will also like to check how good our predictive models can identify competitors and try to search specific types of pages in firms website.

2. In paper A probabilistic rating inference framework for mining user preferences from reviews by Edison Marrese-Taylor, Juan D.Velasqueza, Felipe Bravo-Marquez, Yutaka this can be a major problem since a lot of opinions could can have positive or negative sentiment and then based on the product the opinions are given.

3. In Mining comparative opinions from customer reviews for Competitive Intelligence by Kaiquan Xu , Stephen Shaoyi Liao a, Jiexun Li b, Yuxia Song proposed a new graphical model which helps to extract and plot comparative relations between different products based on the customer reviews with the dependancies internally among the relations , which will help enterprises find out capable risks and further design for new products and marketing strategies but there is need of empirical evaluation process of the proposed model for larger scale with other product types. They also planned to extend the



Yashoda Multidisciplinary Research Journal

model for recognition which work jointly to get the comparative relations and entities so as to decrease the errors collected in the pipeline process.

4. In this work stated as CoMiner: An Effective Algorithm for Mining Competitors from the Web by Rui Li, Shenghua Bao, Jin Wang, Yong Yu, Yunbo Cao works on algorithm called CoMiner which first find out sets of comparative entities of the input entity and then they are ranked by comparability, and finally extraction of the competitive fields but this can be continued CoMiner figure outs in more domains and then tries to improve the CoMiner for mining in more competitive context from the web.

5. In this work called Identifying the Most Influential Data Objects with Reverse Top-k Queries by Akrivi Vlachou, Chris-tos Doulkeridis, Kjetil Norvag and Yannis Kotidis helps to overcome difficult problem of processing queries that sorts the top-m most influential products for customers, where influence is the grouping of the reverse top-k result set they studied number useful in practice and they can be supported by our algorithms.

III. CONCLUSION

From this work, we are having well known definition for competitiveness. Also overcoming the shortcomings of previous paper. The framework in proposed system is efficient and applicable to domains with very large populations of items. Unlike in an existing system we addressed the challenging

REFERENCES

[1] C. W.-K. Leung, S. C.-F. Chan, F.-L. Chung, and G. Ngai,

“A probabilistic rating inference framework for mining user preferences from reviews,” *World Wide Web*, vol. 14, no. 2, pp. 187–215, 2011.

- [2] K. Lerman, S. Blair-Goldensohn, and R. McDonald, “Sentiment summarization: evaluating and learning user preferences,” in *ACL*, 2009, pp. 514–522.
- [3] E. Marrese-Taylor, J. D. Velasquez, F. Bravo-Marquez, and Y. Mat-suo, “Identifying customer preferences about tourism products using an aspect-based opinion mining approach,” *Procedia Computer Science*, vol. 22, pp. 182–191, 2013.
- [4] C.-T. Ho, R. Agrawal, N. Megiddo, and R. Srikant, “Range queries in olap data cubes,” in *SIGMOD*, 1997, pp. 73–88.
- [5] Y.-L. Wu, D. Agrawal, and A. El Abbadi, “Using wavelet decomposition to support progressive and approximate range-sum queries over data cubes,” in *CIKM*, ser. CIKM '00, 2000, pp. 414–421.
- [6] D. Gunopulos, G. Kollios, V. J. Tsotras, and C. Domeniconi, “Approximating multi-dimensional aggregate range queries over real attributes,” in *SIGMOD*, 2000, pp. 463–474.
- [7] M. Muralikrishna and D. J. DeWitt, “Equi-depth histograms for estimating selectivity factors for multi-dimensional queries,” in *SIGMOD*, 1988, pp. 28–36.
- [8] N. Thaper, S. Guha, P. Indyk, and N. Koudas, “Dynamic multidimensional histograms,” in *SIGMOD*, 2002, pp. 428–439.
- [9] K.-H. Lee, Y.-J. Lee, H. Choi, Y. D. Chung, and B. Moon, “Parallel data processing with mapreduce: a survey,” *ACM SIGMOD Record*, vol. 40, no. 4, pp. 11–20, 2012.
- [10] S. Borzsonyi, D. Kossmann, and K. Stocker, “The skyline operator,” in *ICDE*, 2001.
- [11] D. Papadias, Y. Tao, G. Fu, and B. Seeger, “An optimal and progressive algorithm for skyline queries,” ser. SIGMOD '03.
- [12] G. Valkanas, A. N. Papadopoulos, and D. Gunopulos, “Skyline ranking a la IR,” in *ExploreDB*, 2014, pp. 182–18