Survey on Identification of Top-K Competitors Using Data Mining

Ms. Yojana Gaikwad[1], Dr. Gayatri Bhandari[2]

Student, Dept. Of Computer Engineering, JSPMs BSIOTR College, Maharashtra, India
Professor, Dept. Of Computer Engineering, JSPMs BSIOTR College, Maharashtra, India

---***---

Abstract - *In any kind of market, success is depends on the ability to make an item more appearing to customers than the competition. A variety of questions arise in terms of competition: How do we define the competitiveness between two items? Who are the main competitors of a given item? What are the most important benefits of an item that most affect its competitiveness? In this paper, we present a ceremonial definition of the competitiveness between two items, based on the market segments that they can both cover. Our appraisal of competitiveness utilizes customer reviews, which includes positive as well as negative responses from different customers. In addition we perform sentiment analysis by using Natural Language Processing (NLP), used to extract, identify or otherwise characterize the sentiment content of a text unit. Sentiment analysis is also called as 'Opinion Mining' used to determine customer reviews as positive, negative and neutral. According to customer reviews we can define the competitiveness between two items. We present an efficient method for evaluating competitiveness in large datasets and address the problem of identifying the Top-K competitors of a given item.*

Key Words: Data Mining, Information Search and Retrieval, Web Mining, Sentiment Analysis, Natural Language Processing.

1. INTRODUCTION

This paper presents a novel approach for identifying Top-K competitors from different domains using CMiner Technique.

Key concepts as follows:

- How do we formalize and quantify the competitiveness between two items?
- Who are the main competitors of a given item?
- What are the features of an item that most affect its competitiveness?

In this paper we give a formal definition of the competitiveness between two items depending on the market Groups (segments) that they can both cover. Our evaluation of competitiveness utilizes customer reviews, an abundant source of information that is available in a wide range of domains. We give effective methods so that we can evaluate competitiveness in large review datasets and solve the natural problem of finding the top-k competitors of a given item. And at last we calculate the quality of our results and the scalability of our approach using multiple datasets from different domains. This paper works on suggesting good product to customer according to his requirement. So, we have Cminer concept where it takes input of features about different products which are in competition with other products and which is required by different users and then we take query from user about what they want which features are they interested and using segment of features we calculate which product is more useful for customer. Our proposed evaluation of identifying competitors has been based on customer reviews. So in addition we perform Natural Language Processing (NLP) to divide customer reviews as positive, negative or neutral. According to reviews count as positive or negative the customer can easily identify the Top-K competitors of a given item. Our approach is based on following four datasets from different domains:

i) Cameras
ii) Mobiles
iii) TVs
iv) Laptops

Our proposed evaluation is a highly scalable framework for finding the Top-K competitors of a given item in a large datasets.

1.1 Motivation

We observe Competitors Identification Approaches in [8-9][15-27][23][32-35][44] that acts like brute-force method which compute competitiveness between item i and every possible candidate. We notice that it is impractical for larger datasets. We also observed the Bottleneck faced by A Naïve MapReduce Implementation motivates to implement Cminer Algorithm.

1.2 Existing Approach

The common user session on a typical review consists of following steps:

1. User can specify all required features in a query.
2. Submit the query to website`s search engine and retrieve the matching items.
3. Process the reviews of the returned items and then make the purchase decision.

In this type of process the items that covers most of the users requirement will be included in search engine`s response and will compete for users attention. One of the limitation to this process items that will not be considered by user will not have chance to compete.

2. PROPOSED APPROACH

In this paper, we propose a Cminer algorithm for identifying Top-K competitors of a given item which can be determined by using the customer reviews. Also use skyline pyramid to reduce the number of items that need to be considered into computation of finding Top-K Competitors. We propose algorithm to incrementally compute the score of each candidate and it stops working when top-K competitors have merged.

Mathematical model for proposed approach consisting of following:

- $S = \{s, e, X, Y,\}$

- S= Start of the Program
  1. Register/Login into the System
  2. Provide Query.

- e= End of the Program

- Identify the quality of the Product
  1. X= Input of the Program={Query}
     Query=Product with features/details
  2. Y= Output Of the Program
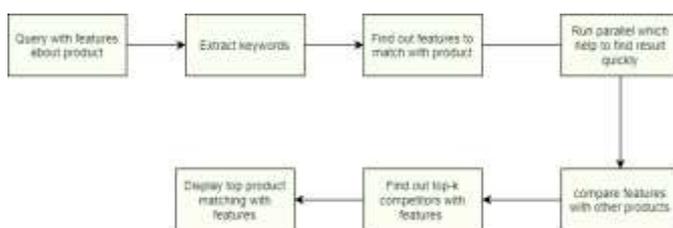
Fig -1: Proposed Approach for Identifying Top-K Competitors

In the proposed approach for identifying Top-K competitors the working of system begins with user query. The query submitted by user is in the form of products details/features. In next step, according to user specifications the keywords which are the features of products are extracted from query. After extracting keywords the system find out features that exactly match with the product, the system run parallel which help to find out result more quickly. In further steps, it compares the features of one product with other product. Next, the Cminer concept is used to find out Top-K competitors of a given items according to customer reviews and products features.

3. CONCLUSIONS

We present a formal definition of competitiveness between two items and identifying Top-K competitors of a given item according to user specifications. Our approach is applicable across different domains and overcoming the drawbacks of previous work. Our work introduces a methodology for mining such information from large datasets of customer reviews. Based on our work, we address the computationally challenging problem of identifying Top-K competitors. As we present the concept Sentiment Analysis, by using Natural Language Processing (NLP) we can easily determine the positive, negative and neutral customer responses.

ACKNOWLEDGEMENT

REFERENCES

[1] R. Li, S. Bao, J. Wang, Y. Yu, and Y. Cao, "Cominer: An effective algorithm for mining competitors from web," in ICDM, 2006.

[2] 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.

[3] 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.

[4] E. Marrese-Taylor, J. D. Velasquez, F. Bravo-Marquez, and Y. Matsuo, "Identifying customer preferences about tourism products using an aspect-based opinion mining approach", Procedia Computer Science, vol. 22, pp. 182-191, 2013.

[5] 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.

[6] 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.

[7] G. Pant and O. R. Sheng, "Web footprints of firms: Using online isomorphism for competitor identification," Information Systems Research, vol. 26, no. 1, pp. 188–209, 2015.

[8] K. Xu, S. S. Liao, J. Li, and Y. Song, "Mining comparative opinions from customer reviews for competitive intelligence," Decis. Support Syst., 2011.

[9] Q. Wan, R. C.-W. Wong, and Y. Peng, "Finding top-k profitable products," in ICDE, 2011.

[10] K. Hose and A. Vlachou, "A survey of skyline processing in highly distributed environments," The VLDB Journal, vol. 21, no. 3, pp. 359–384, 2012.