Extracting top competitors from unorganized data-Review

Kavinya[1], Nanthini.K[2], Indumathi.B[3]

[1]*PG Scholar, Biometrics and cyber security, Department of Information technology, PSG College of technology, Coimbatore, Tamilnadu, India.*

[2,3] *UG Scholar, Department of Information technology, Sri Ramakrishna engineering college, Coimbatore, Tamilnadu, India.*

[1]*Kavinyaapalanisamy@gmail.com,*[2]*knanthini43@gmail.com,*[3]*Indumathia30@gmail.com.*

Abstract

The ability to make a product more desirable to consumers than competition is central to the success of every competitive business. The web application allows the user to see products and their functionalities together with the potential to comment on the product and can also show other customers ' comments. A potential customer finds it difficult to read and determine from the broad comments. The competition of two items based upon market segments that both can cover is determined by this approach. A "CMiner" algorithm is provided to find the top competitors for a particular item to predict competition using the customer reviews. This system returns the competitors of products correctly and reliably, as compared to previous models based on subjective and comparative Web expressions. Business organizations are not only able to identify competitiveness, but they are also able to benefit from meeting user needs.

Keywords: Customer reviews, Competitor mining, Data mining, Firm analysis, Information Search and Retrieval, Item Dominance

1. Introduction

Data extraction is a method used to derive useful data from a wider range of raw data. It is the common research field that supports the business development process, such as mining user preferences, mining data for product or services and mining competitions. The web application of E-Commerce allows the user to access items, their features and the analysis option for the product. It enables the customer to see the product check and reaches all the users. E-commerce is growing fast.

1.1 Competitor Mining

Organization must operate in a competitive environment for the industry. To ensure the survival and growth of businesses, it is essential to understand the strengths and weaknesses of competition. An analysis of competition is a method for obtaining and predicting competitive behaviors while taking better business decisions with information about major competitors. An

analysis of competitors is a key part of a marketing plan. Using competitive analyzes, a business may assess how its product or service is distinctive and which characteristics it uses to draw its target market. Competition between businesses has to be carefully evaluated annually

1.2 Review Analysis

Online reviews have become an important information source that allow consumers to search for detailed and reliable information by sharing past consumption experiences. Product review is simply a mirror of the consumer's thoughts relating that product. Reviews written by other customers describe the usage experience and perspective of customer with similar needs. Online customer reviews serve to highlight various aspects of their business, including products, services, purchase interactions or customer support engagements. The analysis of product review can actually help you in understanding consumer interests and would provide you with a piece of marketing intelligence about the type of products which the consumers are more willing to purchase.
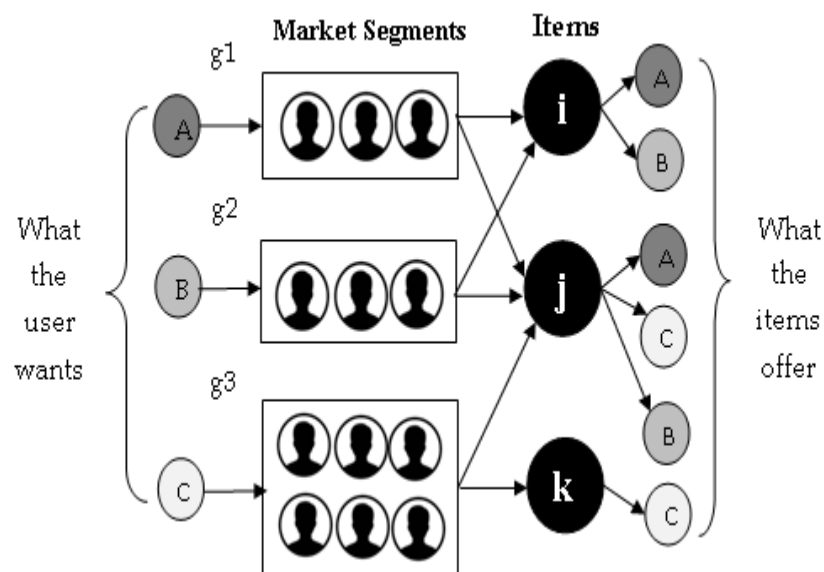

Fig. 1 Competitor Paradigm

2. Existing System

Mark Bergen and Margaret explain a broadly based management approach that compares companies to meet market needs to assess competitive threats on the basis of their capabilities. This model offers a two stage like classification of competitors and assessment of competitors [1].

Rui Li and Shenghua Bao explain the "CoMiner" algorithm that extracts a set of comparative candidates of the input entity, ranks them according to the comparability and then extracts the competitive fields. It calculates the product's overall ranking score from the directed weighted graph derived for the reviews. [2-4] recent trends have indicated that large numbers of customers are switching to online shopping. This methodology presents a feature based ranking technique that mines thousands of customer reviews. The product features are identified to analyze the frequencies and relative usage.

Bushra Anjum and Chaman Lal Sabharwal explain about the "Product Ranking Algorithm" which calculates the entropy measure of product reviews. Reviews are subjective opinions and judgment about a product or the service. [5-7].It explores a hybrid approach such as Entropy, Bilinear and statistical measures that analyze and rank products heterogeneous customer data. The

ranking of the product is based on text reviews, QA data and star rating of products shown in table.1

Table.1. The ranking of the product is based on text reviews

Title	Merits	Demerits
Competitor Identification and Competitor Analysis: A Broad-Based Managerial Approach	This framework compare firms based on their capabilities to meet market needs for evaluating competitive threats	It is a manual process so it takes more time to identify competitors
Web Scale Competitor Discovery Using Mutual Information	An algorithm "CoDis" discovers the competitive rank of firms from social media	Social media have privacy related issues while mining competitors
CoMiner: An Effective Algorithm for Mining Competitors from the Web	This framework used the "CoMiner" algorithm to extracts a set of comparative candidates of the input entity and then ranks them according to the comparability, and then extracts the competitive fields	This algorithm extracts the candidates only using small number of predefined patterns
Voice of the Customer: Mining Online	This framework mines the	This method uses manually
Customer Reviews for Product-Feature Based Ranking	feature-specific product graph constructed from subjective and comparative sentences in reviews	defined product features
Mining Millions of Reviews : A Technique to Rank Products Based on Importance of Reviews	This framework calculates the product's overall ranking score from the directed weighted graph derived for the reviews	Features with the same weight produce ambiguous in calculating ranking score
Identifying Customer Preferences about Tourism Product using an Aspect-based opinion mining approach	The framework called Aspect-based opinion mining analyzes the opinion about product features using predefined rules and patterns	This method needs more tests for all possible features
Mining Business Competitiveness from User Visitation Data	This framework ranks the stores based on check-in data extracted from location based social network	This method does not incorporate the customer reviews and it is not applicable for the online shopping stores

3.Proposed Work

3.1 Problem Statement

Along the line of research has demonstrated the strategic importance of identifying and monitoring competitors of a firm. Mining competitors from online reviews are important task for competitive analysis. Prompt to this problem, marketing and management community have focused on mining comparative expressions from the web or other textual sources. Even though these expressions indeed the indicators of competitiveness, they does not produce more accurate results and it only supports for limited domains.

3.2 Overview of the Project

This work collects data from a customer's online reviews of a particular product using the number of customers who are interested in each feature of a product. To determine competitiveness, the competition score of each item is calculated. The coverage and probability of functions is provided in pairs. The coverage is used in pairings to define the features that each product meets and the likelihood of features to determine the number of potential consumers from a large customer market segment. Thus artifacts are arranged to find superiority of items with the help of the competitive score. This helps reduce the time required for the identification of competitors. Skyline performance and list of items and their features are given to the CMiner algorithm. The CMiner algorithm will then be used to identify top-K contestants of a given item for the k value specified by the user.

3.3 Flow Diagram

The Block diagram of extracting top competitors from unorganized data Shown in Fig.2.

3.4 Module Description

This project deals with the user reviews to find competitive score of the items which helps in identifying top k competitors of a user interested item. The modules of the work are as follows:

1. Calculating competitive score
　　1.1 Pairwise coverage
　　1.2 Feature probability
2. Finding item dominance using competitive score
3. Identifying top K competitors using CMiner algorithm

3.5 Calculating Competitive Score

In order to find the competitor of an item, there is a need to calculate the score of an individual item. It includes two paradigms i.e., user required features and number of customers recommending the same feature. Competitive score is calculated for individual item and also for competitiveness between the two items. Feature Probability $P(f)$ is the percentage of users represented by particular feature f that belongs to set of features F and $V^f_{i,i}$ be the pairwise coverage of all possible values of f that can be covered by both items i,j or by particular item.

3.6 Pairwise coverage

Pairwise coverage $v^f_{i,j}$ of a feature is defined as possible values of f that can be covered by both items i and j or by the individual item i. Binary features indicating the coverage of feature's possible values, in this case the features can be fully covered(1) or not covered(0).

(i) The pairwise coverage of a binary feature f for an individual item is calculated as,

$$v^f_{i,i} = f(i) \qquad (1)$$

(ii) The pairwise coverage of a binary feature f for items i and j can be computed as,

$$v^f_{i,j} = f(i) \times f(j) \qquad (2)$$

3.7 Feature Probability

Feature probability estimation process requires the abundant resource that is the customer reviews. Each review includes a customer opinion on a particular feature of reviewed item. Feature probability is computed by dividing the frequency of each feature f in review corpus dataset R and the sum of frequency of all features. $P(f)$ is computed as follows,

$$P(f) = \frac{frequency(f,R)}{\sum_{f \in F} frequency(f,R)} \qquad (3)$$
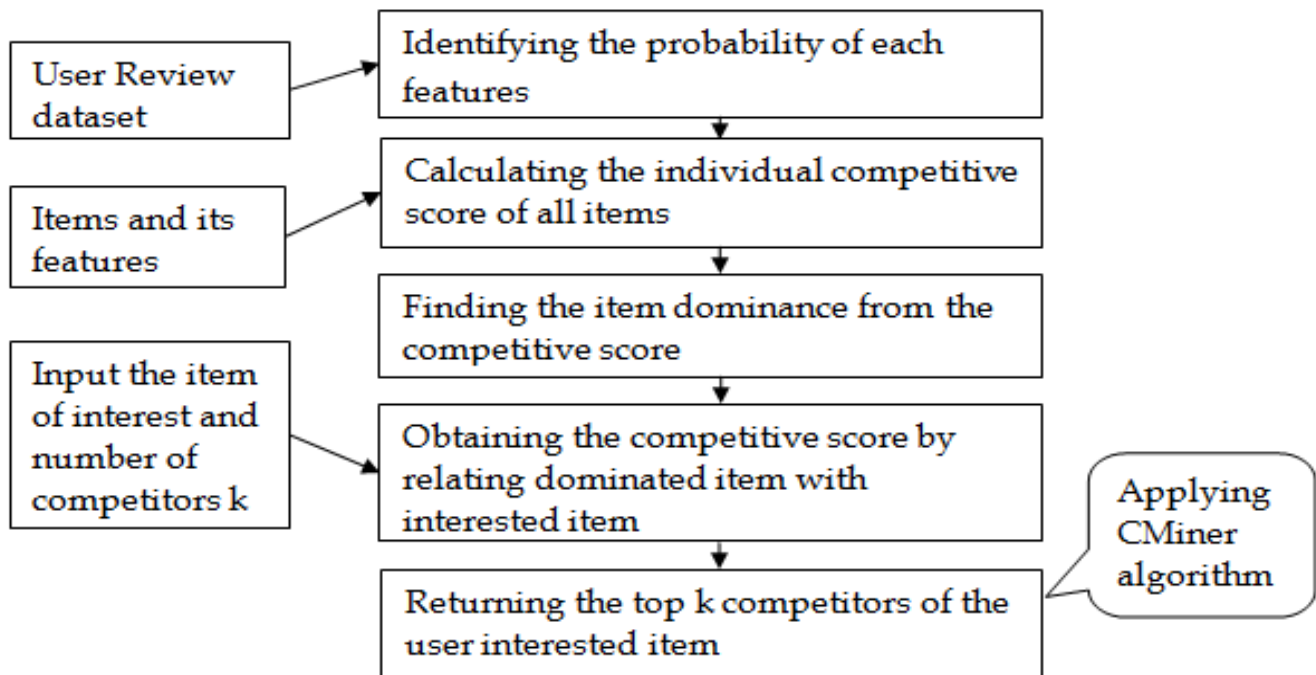
Fig. 2. Block diagram of extracting top competitors from unorganized data

3.8 Finding item dominance using competitive score

Item dominance is a structure consists of all the items with its individual competitive score. Competitive score of individual items are arranged in an order which represent the item dominance. While calculating competitive score, all the features and their probability are considered as the main item dominance factors. The item with the highest competitive score dominates all other items. This approach is also called as skyline which refers the item not dominated by any other item. Competitor identification approach becomes easier by using item dominance. It greatly reduces the number of items need to be considered to identify the competitors.

The main purpose of item dominance is to reduce the time to find top k competitor, because itself provides a result when the user interested item is dominated by k items. When required number of competitors is not achieved, then the item dominance is given as input to the CMiner algorithm.

3.9 Identifying top Competitor using CMiner algorithm

Initially, the dominating items identified from item dominance are stored in the database as topk and k value is decreased accordingly. The items dominated by i are stored separately to find remaining competitors. And for that each item, the competitive score is calculated by relating it with i for all features. If the competitive score is less than lower bound, the item gets eliminated. Otherwise the item is added in topk database.

Experimental Result

In this section, we describe the experiments conducted to evaluate our methodology. The individual competitive score of each hotel was calculated to form Item Dominance. Our experiment gives the user friendly interface so the user can get to know about the competition of their interested item and number of competitors to be retrieved. The proposed algorithm "CMiner" takes the user inputs 7 market segments into consideration to find the top competitors of user specified item. As our approach includes Item dominance as a major factor which produces better

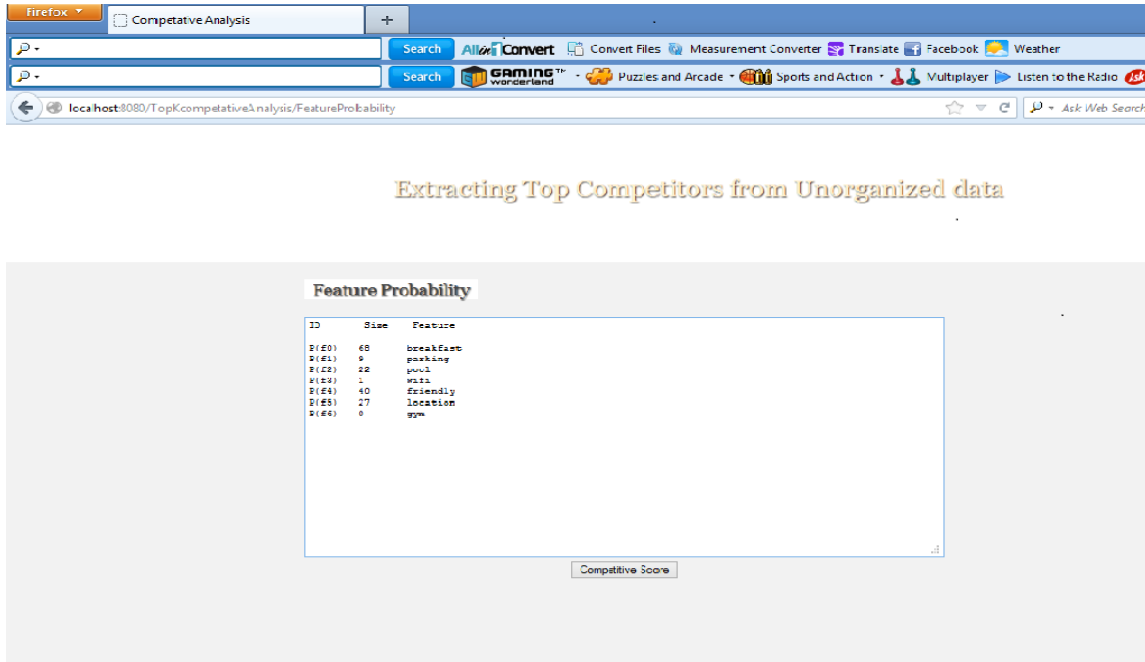results and reduces time complexity shown in Fig 3,4,5.
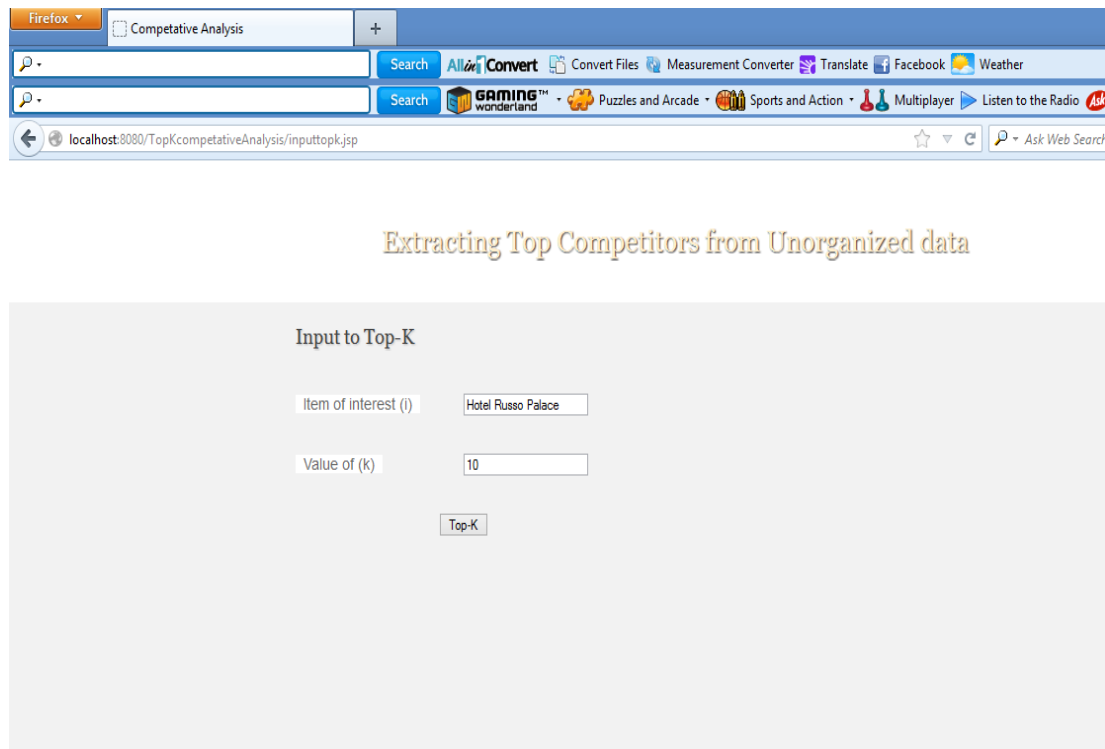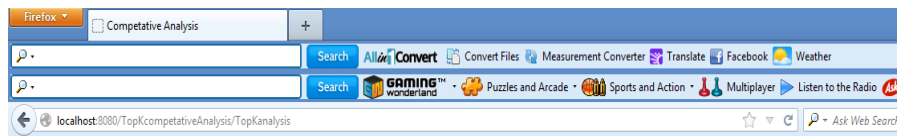

Fig. 3.Finding Feature Probability
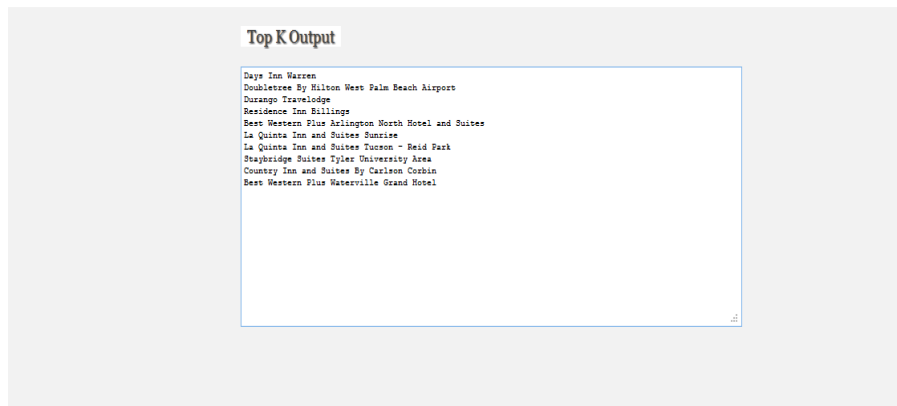

Fig.4. Input to top k

Fig. 5 Listing Top K Competitors

4. Conclusion

The proposed model is mainly focused on calculating competitive score of all items and obtaining top k competitors using CMiner algorithm. Compared to the Previous models based on subjective and comparative expression from web, this model returns the competitors of items accurately and also effectively. With the help of this identified competitors, business organizations not only found their competitors but also get benefited by satisfying user requirements. This method incorporates only the binary features of an item. But the features can also be ordinal (i.e.,) a value from finite ordered list that is user ratings on a product. In our future work, ratings of the product will also be included to extend the quality of result and to make it more efficient.

References

[1] George Valkanas, Theodoros Lappas, and Dimitrios Gunopulos," Mining Competitors from Large Unstructured Datasets", IEEE Transactions on Knowledge and Data Engineering, 1041-4347 (c) 2016

[2] M. Bergen and M. A. Peteraf, "Competitor identification and competitor analysis: a broad-based managerial approach," Managerial and Decision Economics, 2002.

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

[4] Kunpeng Zhang,Ramanathan Narayanan," Voice of the Customer: Mining Online Customer Reviews for Product-Feature Based Ranking"

[5] Kunpeng,Zhang,Yu,Cheng,Wei-Kang,Liao,Alok,Choudhary,"Mining Millions of Reviews : A Technique to Rank Products Based on Importance of Reviews"

[6] E. Marrese-Taylor, J. D. Vel´asquez, 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

[7] T.-N. Doan, F. C. T. Chua, and E.-P. Lim, "Mining business competitiveness from user visitation data," in International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction. Springer, 2015, pp. 283–289.