Product Recommendation System Based on Competitor Mining From Huge Data

1Prof. Sudam Pawar, 2Akshay Handgar, 3Arati Shegade, 4Kiran Jonwal

Abstract: In competitive business environment success is based on the ability to make an item more appealing to customers. Number of questions comes in the context of this task like 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? Despite the impact and relevance of this problem too many domains, only a small amount of work has been given toward an effective solution. 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.

Keywords: Camera products, Laptop products, Cminer algorithm, Naïve Bayes Classifier, etc.

Introduction:
In this paper, we present a formal definition of the competitiveness between two items, based on the market segments that they can both cover. Nowadays in such environment where every buyer tries to search or find product of his choice in large unstructured datasets is quite difficult job as it can bring outputs where user can get confused what to buy and what not to so, this system is proposed where it will suggest best product and its competitors to buyers where they can buy what they want according to their requirements so when user searches product they have to enter product details what they want. These details work as features to system where we use these keywords to find product they are searching. We 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 product and its competitors which are related to query searched and based on reviews of products user can buy good product from large unstructured dataset and likewise we are going to help user to get best products based on its search. 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.

Problem Definition:
To develop the recommendation System to suggest the best product to customer.

Related work
1. Mining competitor relationships from online news: A network
   Description: Addressing the challenging problem of processing queries that identify the top-m most influential products to customers
   Limitation: To examine our approach with news stories written in another language.

2. Identifying the Most Influential Data Objects with Reverse Top-k Queries.
   Description: Identify the top-m most influential products to customers.
   Limitation: Influence is defined as the cardinality of the reverse top-k result set

3. Parallel Data Processing with Map-Reduce: A Survey
   Description: Map-Reduce is simple but provides good scalability and fault-tolerance for massive data processing.
   Limitation: Map-Reduce is unlikely to substitute DBMS even for data warehousing.

4. A probabilistic rating inference framework for mining user preferences from reviews.
   Description: The sentimental orientations (SO) and strength of the opinion words using our proposed Relative-frequency-based method.
   Limitation: Failed to find model text-based CF as an information retrieval (IR) problem.
5. Identifying Customer Preferences about Tourism Products using an Aspect-Based Opinion Mining Approach

**Description:** To discover consumer preferences about tourism products, using opinions available on the Web as reviews.

**Limitation:** A major problem is that lot of opinions could imply a positive or negative sentiment depending on the product the opinion is given on.

**Motivation:**
In any business where there is a lot of competition so the seller focuses on how to sell the product to a customer. So we focus on how we help seller how he can sell a product to the customer in business competition. And which will suggest the buyer which can be a good product for him/her?

**Tools Used:**

**Software Requirement:**
Operating System: Windows 8 and above.
Application Server: Tomcat5.0/6.0
Language: Java
Front End: JSP, HTML
Database: MySQL

**Hardware Requirement:**
The hardware design of the system includes designing the hardware units and the interface between those units.

Processor - Intel core i3
RAM - 4 GB (min)
Hard Disk - 40 GB

**System Architecture:**
Figure shows the detailed flow of mining products. As shown in figure, 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, and the number of top competitors to retrieve. The algorithm first retrieves the items that dominate interested item. 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.

![System Architecture Diagram](image)

**Algorithms**

1. **CMiner:** In this paper, we used CMiner Algorithm for top-k competitors of a given item. Our algorithm makes use of the skyline pyramid in order to reduce the number of items that need to be considered. Given that we only care about the top-k competitors, we can incrementally compute the score of each candidate and stop when it is guaranteed that the top-k has emerged.
2. Naïve Bayes:
It is a classification technique based on Bayes with an assumption of independence among predictors. Naïve Bayes model is easy to build and particularly useful for very large data sets. Along with simplicity, Naïve Bayes is known to outperform even highly sophisticated classification methods. It is a classification technique based on Bayes’ Theorem with an assumption of independence among predictors. Naïve Bayes model is easy to build and particularly useful for very large data sets. Along with simplicity, Naïve Bayes is known to outperform even highly sophisticated classification methods.

DESIGN OF THE STUDY

Propose Algorithm:-

Step 1: User registration
Step 2: User login
Step 3: All Product Showing
Step 4: Search Product
Step 5: Naïve Bayes classifier using for related product shows.
Step 6: Cminer algorithm using for Top-K product find out.

Conclusion:

• We provided a formal definition of competitiveness between two items.
• Based on the competitiveness definition, we also solve the computationally challenging problem of finding the top-k competitors of a given item.
• The proposed framework is very efficient and can be applied to domains with very large populations of items.

References: