Proliferate Business Competitors By Mining Unstructured Datasets

R.Aiswarya[1], R.Pradeepa[2], J.Subhashri[3], V.Vijipriya[4],

G.Nanthakumar[5] - Associate Professor.

[#]*students, Department of Computer Science and Engineering,*

Anjalai Ammal Mahalingam Engineering College, Kovilvenni.

Thiruvarur-614403, Tamilnadu , India

ABSTRACT

Competition is always a good thing .The ability to learn faster than your competitors may be a sustainable competitive advantage. In the real world particulary in Competitive Business, success relies on the potential to label an item or product better fascinating to the customers. This depends on the factors: how to certify and librate the competitiveness between two items? Who are the competitors? What are the attributes of the product that influence the competitiveness? In this paper, we represent a conventional definition of competitiveness between two items based on the market segments that the items can cover. Our estimation employs customer exploration or reviews, a bountiful source of data that is accessible in a broad field of domain .We present a effective pattern to assess the competitiveness in massive datasets and then label the natural problem of discovering the top-K competitors for an item. The information of various items and products is collected by comparing content related information and excluding unrelated contents using C-Miner Algorithm. At last the quality of our results and the scalability were analyzed by using datasets from huge domains.

Keywords: *Data Mining, C-Miner Algorithm, Web Mining, Reviews, Skyline*

I.INTRODUCTION:

Over the past decades, many studies have been done on improving the quality of strategic importance of recognizing and examining a firm's competitors. For Businesses, Data Mining is used to discover patterns and relationships in the data in order to help make better business decisions. The primary goal of analyzing the Competitors is to find out the strength and weaknesses of their competition. To achieve and maintain a competitive advantage, a complete knowledge about the competition, a positive identification of the competitor's strength and

weaknesses, and a mechanism must be developed to enhance a effective strategy.

Definition 1:

The competitiveness between two items depends on whether they compete for the attention and business of the same group of customers. For Example, two restaurants residing in different countries are not competitive as there is no convergence between their target groups. We present a approach for measuring all the segments in a given market based on mining huge review datasets. This approach allows us to operationlize our definition of competitiveness and addresses the problem of finding the top-K competitors of an item through a highly scalable framework involving an efficient algorithm ad an appropriate index. Our Methodology trounces the dependence on previous works on sparse comparative evidence mined from text. An impulsive examination is that the size of the segments has a direct impact on the competitiveness. The potentiality to assess the competitiveness between two items delegates the recommendation system to strategically choose the order in which the items should be approved.

Definition 2:

The pair-wise coverage of a feature f by two items i , j is the percentage of all possible values of f that can be covered by both i , j.

Binary and Categorical Features:

Categorical Features take one or more limited values from finite space and can be encoded via a set of binary features.

Numeric Features:

Numeric Features take value from a pre-defined range. The pair-wise coverage of a numeric feature f by two items i and j can be easily calculated as the least value achieved for f by any one item.

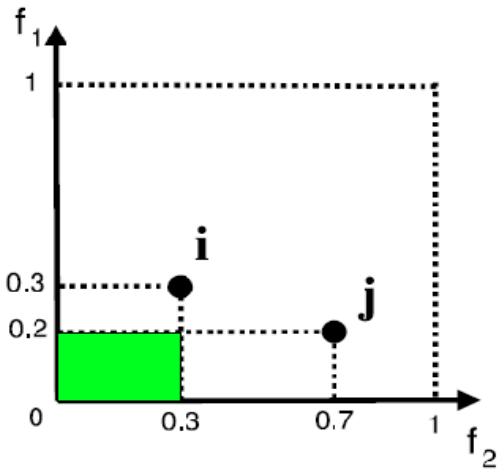
Ordinal Features:

Ordinal Features take values from a a ordered list. The pair-wise coverage of an ordinal feature f by two items i , j can be computed as follows:

$$\mathbf{V^f_{i,j} = min(f[i] ,f[j]) / |V^f|}$$

Pair-wise Coverage of a query:

Let us consider a query which includes two numeric features f_1 and f_2 and two competitive items i and j.



The pair-wise coverage of a given query q by two items i; j can be determined as the area of the hyper-rectangle indicated by the pair-wise coverage given by the two items for each feature f ∈ q .

Definition 3:

Let a market consists of a set of items I and a set of features F. Then it is clear than an item $i \in I$ dominates another item$j \in I$, if f[i]>=f[j] for every future$f \in F$.

II.RELATED WORK:

A) User Preferences:

Users have a strong preference for summarizers that model sentiment over non-sentiment baselines, but have no broad overall preference between any of the sentiment-based models [1]. The study of User Preferences is implemented on aspect based opinion mining that offers a larger scope method to understand aggregated user preferences [2]. The preferences can be used to determine future sales in both online and offline environments. Online Reviews have become a major information source for consumer prior to making a purchase [3].

B) *Competitor Mining:*

The task of competitor mining includes mining all the information such as competitors, competing fields and competitor's strength [4]. An explanatory Research is performed to derive competition from visits performed by large number of users and developed models to rank the stores based on this data that converts user visits to competitive probabilities [5]. A Novel Graphical Method is proposed to extract and visualize the comparative relations between the products from customer reviews with interdependencies among the relations taken into consideration to help enterprises discover potential risks and further design new products and marketing strategies [6]. Competitor identification is referred to as a classification process through which competitors of a focal firm are identified based on "relevant similarities" [9].

C) Skyline:

The Skyline is defined as those points which are not dominated by any other points. A point dominates another point if it is as good or better in all dimensions or at least in one dimension. The Skyline Queries can be extended into an existing database system with a new logical operator known as Skyline Operator [7].Branch and Bound Skyline Algorithm[BBS] is efficient fro both progressive and complete skyline computation independently of that data characteristics and it can easily handle user preferences and process numerous alternative skyline queries[8].

III) SYSTEM MODULE

A new formalization of the competitiveness between two items is suggested, based on the market segments that the items can both cover. A method for estimating all the segments in a given market based on mining large review datasets is described. Our Proposed System is the first to address the evaluation of competitiveness via the analysis of large unstructured , without the need for direct comparative evidence. Our Approach overcomes the dependency on past work.A Dignified approach for the recognition of various types of customers in a market and also for the evaluation of percentage of customers that belong to each type. This paper builds on and obviously broaden our preclusive work on the analysis of competitiveness.,

COMPETITIVENESS

Give U a chance to be the number of inhabitants in every single conceivable client in a given market. We look at that as a thing I covers a client u ∈

U in the event that it can cover the greater part of the client's necessities. At that point, the competitiveness between two things I, j is relative to the quantity of clients that they can both cover

PYRAMID FINDER

Given the horizon Sky (I) of an arrangement of things I and a thing i ∈ I, let Y contain the k things from Sky (I) that are most aggressive with I. At that point, a thing j ∈ I must be in the best k contenders of i, if j ∈ Y or if j is overwhelmed by one of the things in Y. we don't have to consider the whole arrangement of applicants with a specific end goal to generate the top-k contenders. This propels us to build the horizon pyramid. A structure that unimaginably diminishes the amount of things that ought to be considered. We allude to the calculation used to build the horizon pyramid as Pyramid Finder

1) Administrator Module:

This is the first step where the admin takes the responsibility of administering the customers and upload data regarding items or features in the database. Later, the admin can check all the uploaded details and user queries along with their interests. At last top-k competitors will be evaluated using C-Miner Algorithm and the result is available to the admin.

2) Customer Module:

It is the second step which operates on customer requirements where the customer based features are developed and the customer can give query for any item , the query will be processed which will be useful for finding the competitors . Relied upon the customer requirements, the customer requested data item is displayed.

3) C-Miner Algorithm Module:

An Exact Algorithm for discovering the top-k competitor for a given item. Our Algorithm makes use of skyline pyramid in order to minimize the number of items need to be considered. This algorithm only cares about the top-k competitors that calculate the score of each candidate and terminate when it is given assurance that top-k has been emerged. It iterates the skyline pyramid which reduces the dimensionality. C-Miner falls under the category of frequent sequence mining.

C-Miner Algorithm:

Step 1: To find top-k← $master(i)$

Step 2: get the sets of competitive items.

A=set (), B=set ().

Step 3: if k≤ $|top - k|$ then

Step 4: return top-k.

Step 5: else

Step 6: X← $update(top - k)$

Step 7: top-k← $merge(top - k, X)$

Step 8: end if

Step 9: return top-k.

4) Skyline Operator Module:

The Skyline operator is performed for the coordinate points of data items. It helps to determine the subset of points that are not dominated by any other points in the same population.

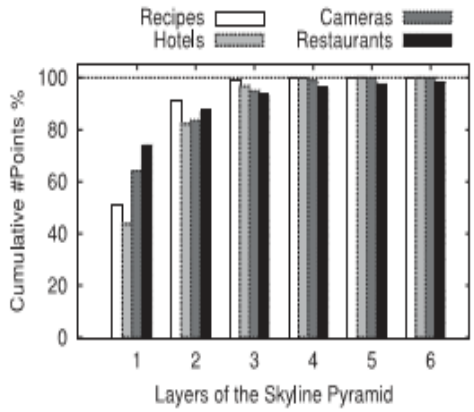


Figure 1: Cumulative Distribution of items.

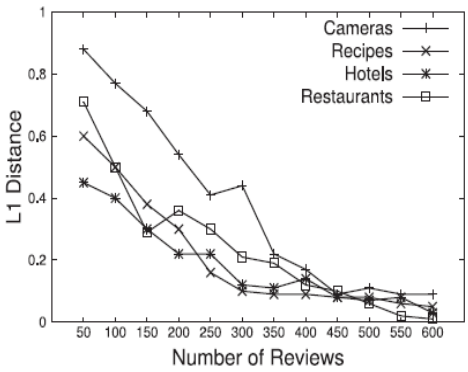


Figure 2: Convergence of Query Probabilities

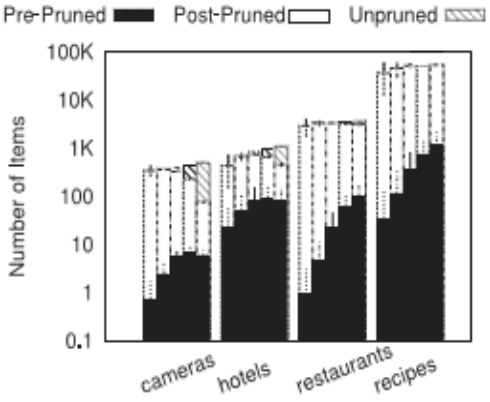


Fig 3: Pruning Effectiveness

IV) CONCLUSION

Data Mining is predominant in finding the patterns, forecasting, pioneering of knowledge and information etc. in various business domains. Machine Learning Algorithms are widely implemented in various applications. Every business related application makes use of data mining techniques. There is a need to support web mining to enhance such business or proffer felicitous competitors for the business to the user. The competitor mining is one such a way to analyze the competitors for the selected items. In this paper, we proposed an enhanced C-Miner algorithm that devises unstructured data into structured data. Based on the customer reviews, the item sets are selected. Since the aim is to analyze the competitors, the relevant features are picked from the reviews. With the help of C-Miner and skyline approach, the information is represented and then top-K item is extracted from the available resources. The proposed framework is very efficient and can be applied to large number of items.

REFERENCES

[1] K.Lerman, S.Blair-Goldensohn, and R. McDonald, **"Sentiment summarization: Evaluating and learning user preferences,"** in Proc. 12th Conf. Eur. Chapter Assoc. Comput. Linguistics, 2009, pp. 514–522.

[2] 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 Comput. Sci., vol. 22, pp. 182–191, 2013.

[3] R. Decker and M. Trusov, **"Estimating aggregate consumer preferences from online product reviews,"** Int. J. Res. Marketing, vol. 27, no. 4, pp. 293–307, 2010.

[4] S. Bao, R. Li, Y. Yu, and Y. Cao,**"Competitor mining with the web,"** IEEE Trans. Knowl. Data Eng., vol. 20, no. 10, pp. 1297–1310, Oct. 2008.

[5] T.-N. Doan, F. C. T. Chua, and E.-P. Lim, **"Mining business competitiveness from user visitation data,"** in Proc. Int. Conf. Social Comput. Behavioral-Cultural Model. Prediction, 2015, pp. 283–289.'

[6] K. Xu, S. S. Liao, J. Li, and Y. Song, **"Mining comparative opinions from customer reviews for competitive intelligence,"** Decision Support Syst., vol. 50, pp. 743–754, 2011.

[7] S. Borzsonyi, D. Kossmann, and K. Stocker, **"The Skyline Operator,"** in Proc. 17th Int. Conf. Data Eng., 2001, pp. 421–430.

[8] D. Papadias, Y. Tao, G. Fu, and B. Seeger, **"An optimal and progressive algorithm for skyline queries,"** in Proc. ACM SIGMOD Int. Conf. Manage. Data, 2003, pp. 467–478.

[9] T. Lappas, G.Valkanas, and D.Gunopulos, "Efficient and domain-invariant competitor mining," in Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2012, pp. 408–416.