ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

Finding Top-k competitors using Differential Evaluation

Rani R. Raut [1] , Prof. R. V. Mante [2] , Dr. K. P. Wagh [3]

P.G. Student, Computer Science and Engineering, Government College of Engineering, Amravati, India[1]

Assistant Professor, Computer Science and Engineering, Government College of Engineering, Amravati, India [2]

Assistant Professor, Information Technology, Government College of Engineering, Amravati, India [3]

ABSTRACT: Data mining is very popular area for research which helps us in business improvement. Using Data mining techniques helps for mining customer preferences and helps in mining information of customer reviews. This paper is concerned with the problem of mining competitors from the Web. Nowadays, the fierce competition in the market necessitates every company to know not only which companies are its primary competitors but also in which domains the company's rivals compete with itself and what is the competitors strength in a specific competitive domain. For finding competitors need to use customer opinion and how many customer purchasing product according to their features. 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. Our approach is evaluated against strong baselines via user study and experiments on multiple datasets from various domains.

KEYWORDS**:** Data mining , Differential evaluation, Unstructured dataset.

I. INTRODUCTION

Data mining is defined as extracting information from huge sets of data. We can say that data mining is a process of mining knowledge from different sources. In data mining we can use data analysis techniques for identifying patterns and establishing relationship in datasets. of mining knowledge from data. Along line of research has demonstrated the strategic importance of identifying and monitoring a firm's competitors [1]. Motivated by this problem, the marketing and management community have focused on empirical methods for competitor identification as well as on methods for analysing known competitors[1].Suppose we are comparing names of the products like "sony , onida" and "Seagate, Hp", Then we can get comparative companies by writing query to web. Generally competition is occur between two companies i.e. suppose we are comparing two companies and this companies are in different countries so we cant compare them no competition exist between them. Competitiveness is based on some paradigm. There are different type of competitors, some competitors compete for attention and business of the same group. consider an example shown below.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

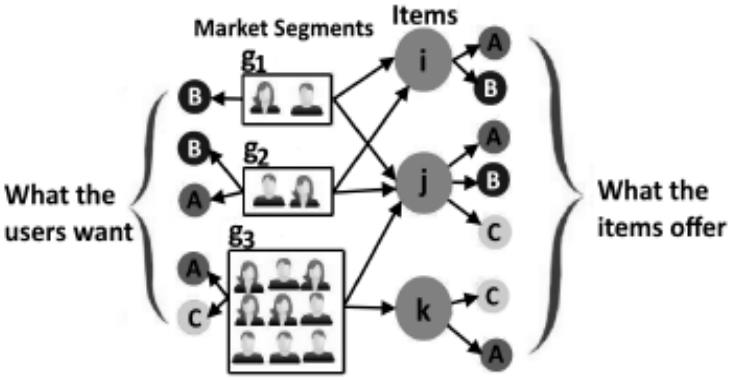

Fig. 1: Example of our competitiveness paradigm

This figure shows illustration of competitiveness between three items namely i,j and k. Each item is mapped to the features set so we can offer it to the customer.
Consider three features A,B and C. This example consider binary features namely available and not available. On the left side we have three group of customer i.e. g1 ,g2 and g3.Here users are grouped according to their preferences with respect to features. For example, the customers in *g*2 are only interested in features *A* and *B*. We observe that items *i* and *k* are not competitive, since they simply do not appeal to the same groups of customers. On the other hand, *j* competes with both *i* and k. initially, an interesting observation is that *j* competes for 4 users with *i* and for 9 users with *k*. In other words, *k* is a stronger competitor for *j*, since it claims a much larger portion of its market share than *i*.

Problem Definition:
In e-commerce application it is very difficult to identify the competition among the product. In market the execution and comparative analysis of the product is going to achieve on the basis of manual comparative analysis of the reviews comments and on the basis of that evaluation of product will done. So we need to provide the atomized way to make all process efficient and give the results on single click.

II. RELATED WORK

We have to address the evaluation of competitiveness using large unstructured datasets.
In management literature survey, it is found that they focus on how manager can identify competitors. In some works they perform competitor identification as mental categorization and use to classify candidate firms[2]. Other manual categorization methods are based on market and resource-based similarities between a firm and candidate competitors[3]. Finally, managerial competitor identification has also been presented as a sense- making process in which competitors are identified based on their potential to threaten an organizations identity [4].
Some competitors identify key points like market share and share of e-wallet and show their values ,perform according to their information by mining details of customer transaction and aggregate data for competitors. But this approach is not good. Doan et al. explore user visitation data, such as the geo-coded data from location-based social networks, as a potential resource for competitor mining [5]. Pant and Sheng hypothesize and verify that competing firms are likely to have similar web footprints, a phenomenon that they refer to as online isomorphism [6]. Their study considers different types of isomorphism between two firms , such as the overlap between the in-links and out- links of their respective websites, as well as the number of times that they appear together online (e.g. in search results or new articles). Similar to our own methodology, their approach is geared toward pairwise competitiveness. However, the need for isomorphism features limits its applicability to firms and make it unsuitable for items and domains where such features

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

are either not available or extremely sparse, as is typically the case with co-occurrence data. In fact, the sparsity of co-occurrence data is a serious limitation of a significant body of work [7].
Recent work [8], [9] has explored competitiveness in the context of product design. The first step in these approaches is the definition of a dominance function that represents the value of a product. The goal is then to use this function to create items that are not dominated by other, or maximize items with the maximum possible dominance value. A similar line of work [10] represents items as points in a multidimensional space and looks for subspaces where the appeal of the item is maximized. While relevant, the above projects have a completely different focus from our own, and hence the proposed approaches are not applicable in our setting.

III. EXISTING SYSTEM

In existing system they have given definition of competitiveness and how customers of same group are competing. Existing system works on following contributions:

- A formal definition of the competitiveness between two items, based on their appeal to the various customer segments in their market. Our approach overcomes the reliance of previous work on scarce comparative evidence mined from text.
- A formal methodology for the identification of the different types of customers in a given market, as well as for the estimation of the percentage of customers that belong to each type.
- A highly scalable framework for finding the top-k competitors of a given item in very large datasets.

In existing system we have formal definition of competitiveness between the two items. Here we are evaluating competitiveness between the items on the basic of customer reviews which is available on web.
The main purpose of existing system is to calculate top –k competitors. For calculating top-k a competitors we have to extract data from the datasets .According to customer reviews means according to customers good and bad reviews we can calculate it. For calculating top-k competitors existing system used different algorithms. While using different algorithms ,it is notice that this algorithm have a different time complexity.

IV. PROPOSED SYSTEM

The working of propose system is same as the existing system but difference is that we are using another algorithms. In proposed work we used E-Commerce example. Here working is same as existing system like extracting unstructured data from dataset i.e extracting customers good as well as bad reviews and product features.
Existing system used different algorithm for calculating top-k competitors and got different result with different time complexity. Here our aim is to reduce the computing time for calculate top-k competitors.
For reducing computing time for calculating top-k competitors we used differential evaluation algorithms.
But in proposed work calculated top-k competitors sequence may change means competitors calculated by using C-Miner and DE may or may not same.

Differential Evaluation Algorithm:

The DE algorithm is a population based algorithm like genetic algorithms using the similar operators; crossover, mutation and selection. The main difference in constructing better solutions is that genetic algorithms rely on crossover while DE relies on mutation operation. This main operation is based on the differences of randomly sampled pairs of solutions in the population. The algorithm uses mutation operation as a search mechanism and selection operation to direct the search toward the prospective regions in the search space. The DE algorithm also uses a non-uniform crossover that can take child vector parameters from one parent more often than it does from others. By using the

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

components of the existing population members to construct trial vectors, the recombination (crossover) operator efficiently shuffles information about successful combinations, enabling the search for a better solution space[12].

Flowchart:

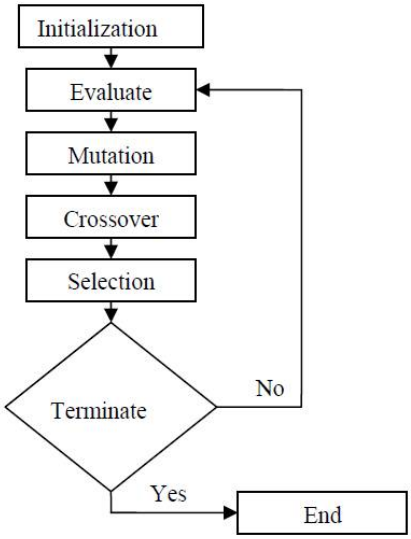

Fig .2 The procedure of Differential Evolution

Initialization:
The main assignment is to decide the above parameter, and to ensure the initial value of solution vectors control in restriction scope to make sure solution's rationality [12].

Mutation:
Selected several solution vector randomly, and acquire the difference between the vectors to multiply mutation faction further more added on target vector to assist target vector mutate. Here are two traditional common mutation types:

DE/rand：$V_{i,g+1} = X_{ri,g} + \boldsymbol{F}\,(X_{r2,g} - X_{r3,g})$

DE/rand：$V_{i,g+1} = X_{best} + \boldsymbol{F}\,(X_{r2,g} - X_{r3,g})$

Crossover:

Upon approaching the mutation operation, the donor vector will swap the information with the target vector (*xi*) randomly. After crossover, a trial vector u will be produced. The following formulation will be used to decide whether the component *i* is composed from target vector *xi* or donor vector *vi* in the generation *j*. Here, *rand* is a random number that obey uniform distribution between 0 and 1. *CR* records the crossover rate.

$$u_{j,i,G+1} = \begin{cases} V_{j,i,G+1} & \text{if rand} \le \text{CR} \\ V_{j,i,G+1} & \text{if rand} > \text{CR} \end{cases}$$

Here, *rand* is a random number that obey uniform distribution between 0 and 1. *CR* records the crossover rate.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

Selection:
After the mutation and crossover operations, the trial vector and target vector will approach to fitness functions to determine the one to be reserved for the next generation.

V. RESULTS

Below table shows the results of three algorithms that used to calculate top-k competitors. Here we used three algorithms namely Naive, C-Miner and DE performance of DE is good as compare to the Naive and C-Miner .The below result of algorithm shows its performance. Inserted product name is "Amazon Fire Tv" and features " Wifi,4G,HD Voice remote" Company name selected are HP, Sony, LG, Panasonic and Onida.

Algorithms	Competitors	Computing Time
Naive	LG Sony Onida HP PANASONIC	262271883 N S
C-Miner	LG Sony Onida HP PANASONIC	236531626 NS
Differential Evaluation	HP Sony LG Onida PANASONIC	232122939 NS

These are the time for calculating top-k competitors which is different for different algorithms and differential evaluation algorithm gives less computing time as compare to other but given top-k competitors are somewhat different in case of differential evaluation.

VI. CONCLUSIONS

We presented a formal definition of competitiveness between two items, which we validated both quantitatively and qualitatively. We addressed the computationally challenging problem of finding the top-k competitors of a given item. The proposed framework is efficient and applicable to domains with very large populations of items. Differential Evaluation algorithm provide the efficient way of identifying top-k competitors by considering various factors like customers reviews about products. We are able to manage large unstructured dataset and done with effective mining from large unstructured datasets.

REFERENCES

1. George Valkanas, Theodoros , Lappas and Dimitrios Gunopulos," Mining Competitors from Large Unstructured Datasets,"2017.

2. J. F. Porac and H. Thomas, "Taxonomic mental models in competitor definition," *Academy of Management Review*, vol. 15, no. 2, pp 224–240,

3. M. Bergen and M. A. Peteraf, "Competitor identification and competitors analysis: a broad-based managerial approach," *Managerial and Decision Economics*, 2002.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

4. W. T. Few, "Managerial competitor identification: Integrating the categorization, economic and organizational identity perspectives," *Doctoral Dissertaion*, 2007
5. 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.
6. 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.
7. K. Xu, S. S. Liao, J. Li, and Y. Song, "Mining comparative opinions from customer reviews for competitive intelligence," *Decis. Support Syst.*, 2011.
8. Q. Wan, R. C.-W. Wong, and Y. Peng, "Finding top-k profitable products," in *ICDE*, 2011.
9. Z. Zhang, L. V. S. Lakshmanan, and A. K. H. Tung, "On domination game analysis for microeconomic data mining," *ACM Trans. Knowl. Discov. Data*, 2009
10. T. Wu, D. Xin, Q. Mei, and J. Han, "Promotion analysis in multi dimensional space," *PVLDB*, 2009.
11. Ching-Wei Chien, Zhan-Rong Hsu, Wei-Ping Lee , **"**Improving the Performance of Differential Evolution Algorithm with Modified Mutation Factor",2009.
12. Dervis Karaboga, **"**A Simple and Global Optimization Algorithm for Engineering Problems: Differential Evolution Algorithm",2004.