An Enhanced Quality Approach for Ecommerce Website using DEA and High Utility Item Set Mining

Varsha Kulkarni[1], Vilas Jadhav[2].

[1] PG Scholar, Dept. of Computer Engineering, MGMCET, Kamothe, Navi Mumbai, India

[2] Professor, Dept. of Computer Engineering, MGMCET, Kamothe, Navi Mumbai, India

ABSTRACT: Today, E-commerce is booming everywhere. Customer's satisfaction is an important issue in every Business-to-Consumer website. Sometimes it is difficult to evaluate performance of website and to find utility for the product. It depends on customer to visit, see or purchase any product. Entrepreneur is unaware of its nature. Here, we have adopted DEA (Data Envelopment Analysis) method which is working on dataset. DEA is used to evaluate performance of DMU's (Decision Making Units). Here, we have taken up Utility Pattern Growth algorithm for finding High utility item sets. Our scheme E-commerce website Quality and Utility Analysis System (EQUAS) will analyse the website and will results in popularity of website to an entrepreneur. Our scheme will contribute to the improvement of overall quality of website and promote the healthy development in e-commerce. The goal of utility mining is to discover all the high utility item sets whose utility values are beyond a user specified threshold in a transaction database. Utility Mining identifies the item sets with highest utilities, by considering profit, quantity, cost or other user preferences. DEA analyzes quality by calculating user popularity rates and utility item set mining investigates profit earning items.

KEYWORDS: data envelopment analysis (DEA), decision making units (DMU's), E-commerce website Quality and Utility Analysis System (EQUAS).

I. INTRODUCTION

Quality of e-commerce website affects on its performance directly. With the rapid development and wide application of e-commerce, academia and business are desperately in need of an effective website evaluation technique to compare the relative merits of website which will be helpful for website construction [1]. DEA is one of the techniques to evaluate the quality of e-commerce website. DEA is relatively new "data oriented" approach for evaluating the performance of a set of entities called DMU's which convert multiple inputs into multiple outputs. Proposed model will calculate inefficiencies in units and results will be used for improvement of website. In addition to DEA, we have included utility mining technique for finding high utility item sets. Data mining is the extraction of hidden, previously unknown and potentially useful information from databases. Data mining is sometimes called as knowledge discovery. Based on user queries it analyzes the relationships in stored data. Data mining is mostly used in financial data analysis, retail industry, intrusion detection and other scientific applications. Data mining models can be applied to specific scenarios such as forecasting, risk and probability. To discover the useful patterns from database, frequent pattern mining has been applied to different databases. Searching of frequent patterns from large databases is very important in many applications. The goal of frequent item set mining is to identify all frequent item sets and it collects the set of items that occur frequently together. An item set is called high utility item set if its utility is not less than user specified minimum utility threshold [2]. The information of high utility item sets is maintained in a tree based structure called utility pattern tree. High utility item sets can be generated from UP-Tree efficiently. We have designed E-commerce website Quality and Utility Analysis System (EQUAS) to calculate inefficiency of a website as well as perform utility mining of high utility item sets. DEA will work on website log data and utility mining will work on transactional databases to form a tree structure.

II. RELATED WORK

Zijuan Liu [1] chose input oriented BCC model which is ideal for modelling of purchase situations. It is done by comparing with hypothetical DMU's to calculate constant rates. It highlights difference between CCR and BCC model. It defined input oriented BCC model. Vincent S. Tseng et Al [2] explains about discovery of high utility item sets like profit to eliminate unpromising items. It is mainly used for transactional databases. Here, they used UP-Growth and UP-Growth+ algorithms. They can effectively used for long transactions. Jyothi Pillai, O.P.Vyas [3] explained efficiency of utility mining in the field of data mining. Utility mining discovers item sets that frequently occur. Chuen Tse Kuah et al formalized significance of DEA and its different models. The popularity of DEA is due to its ability to measure relative efficiencies of multiple-input and multiple-output DMUs without prior weights on the inputs and outputs. Corrado lo Storto [4] evaluated website quality from user viewpoint using DEA. The framework is implemented to compare websites that sell products online. Jyothi Pillai, O.P.Vyas [5] defined emerging technologies in mining. It illustrated utility mining in terms of profit, cost and other expressions of user experiences. Vincent S. Tseng [6] defined UP-Growth(Utility Pattern Growth) algorithm for discovering high utility item sets and removing unpromising items by providing threshold like profit, quantity etc. A tree structure called UP-Tree is created to to generate candidate item sets to avoid repeated scanning of database. Smita R. Londhe et al [7] tried to avoid space consumption and execution time for long transactions by defining two algorithms namely, UP-Growth and UP-Growth+ algorithms. Adinarayanareddy B et al [8] mainly focus on Up- Growth algorithm and their production of large number candidate item sets which requires more space and affects mining performance.

III. PROPOSED ALGORITHM

A. *Design Considerations:*
   - All the ports which are required should be free.

B. *Description of the Proposed Algorithm:*

We have designed E-commerce website Quality and Utility Analysis System (EQUAS). We have implemented this system using software Java 1.6 and database SQL server 2005. Our aim is to build a quality evaluation system which will diagnose datasets on the basis of their efficiency and investigates profit earning items. Our reference paper has built a system which uses BCC model. A BCC model is input oriented and used for evaluation purpose. A model is established based on input data and output data of derived from CCR [2]. DEA is used to evaluate efficiency of E-commerce website. In order to find unpromising items and to overcome the drawback of excess scanning of database, *utility mining* is also performed along with DEA. Proposed model will calculate inefficiencies in units and results will be used for improvement of website. In this the unit of analysis represents a production unit. We conceptualize consumer-website interaction during online shopping as a production process in which the customer conducts a purchase and inputs are converted into outputs.

Finding Utility

A novel algorithms as well as a compact data structure for efficiently discovering high utility item sets from transactional databases. Utility pattern growth (UP-Growth) [4] and a compact tree structure, called utility pattern tree (UP-Tree), for discovering high utility item sets and maintaining important information related to utility patterns within databases are proposed. High-utility item sets can be generated from UP-Tree efficiently with only two scans of original databases. To improve mining performance and avoid repeated scanning of database, we used here utility pattern tree [2].

To analyse performance of website we propose an enhanced method called, E-commerce website Quality and Utility Analysis System (EQUAS) which calculates inefficiency of website functionality and utility mining to obtain high quality item sets.

In this section we discussed our EQUAS system's design and algorithm and in next section we focus implementation and results.
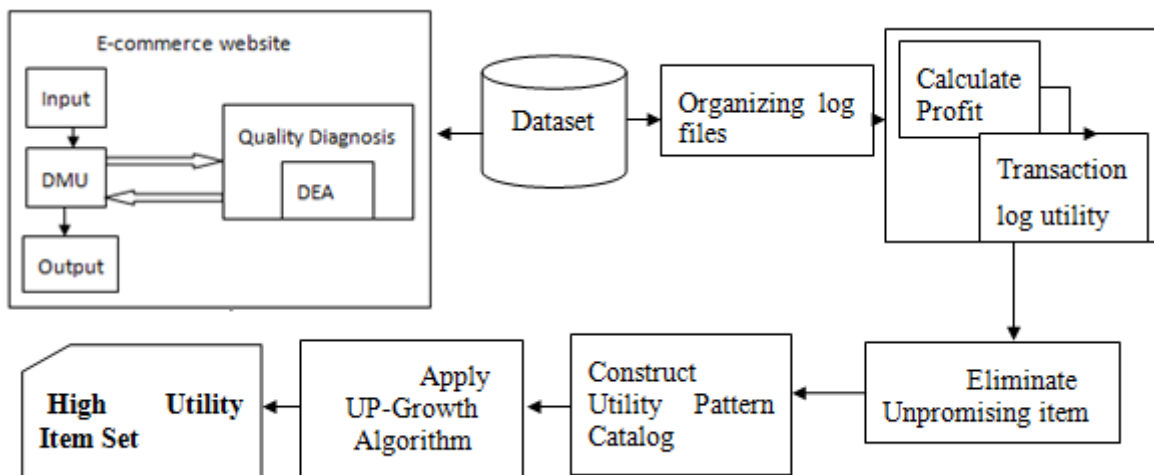
Figure 1: Proposed System

IV. SYSTEM FLOW

1. An e-commerce website will be developed on which DEA will work to analyze quality of the website and utility mining will be performed to calculate the profit earned by the website.

2. Collecting web log data from datasets to find out value of input and output variables through DMU.

3. Calculating unit inefficiencies to analyze website functionality.

4. Calculating conversion rates from one page to another and obtaining average value of each conversion rate

5. Organizing web log data to calculate profit by using item set and its unique identifier.

6. Remove unpromising data by extracting only utility item sets whose threshold is greater than or equal to minimum threshold.

7. Applying UP-Growth algorithm to generate UP-Tree which is representation of information of high utility item sets.

V. RESULTS

We can divide the result into two parts. One is the quality analysis system developed by using DEA and another one is utility mining based on transactional database. First is to enhance background performance of the system and another is to increase the profit. We can see that rate of people who tends to purchase product is lower than product list. So, here we have scope to enhance our website quality. Some ways are arrange products properly, transaction should not be complex. Also here in second graph Transaction unit (TU) and Transaction weighted unit (TWU) are shown.

TU is addition of each item in single transaction. Each transaction is number of items purchased by single user.

TWU is addition of TU's which contain same element.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

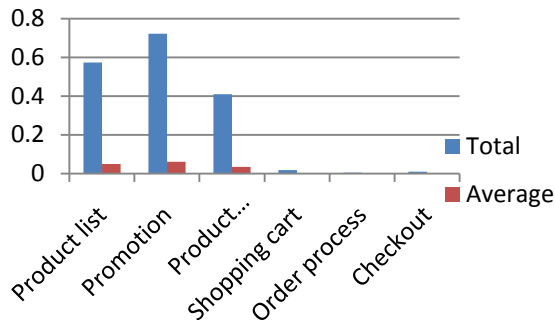**Vol. 3, Issue 11, November 2015**
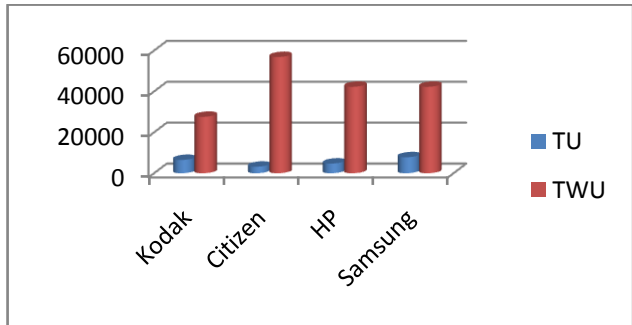

Figure 2: Result of DEA

Figure 3: TU and TWU calculation

After calculation of TU and TWU we have inserted user defined threshold. It is owner specified means website owner can put in some value. System categorizes items into promising and unpromising items. Suppose threshold is 40000 then result could be:
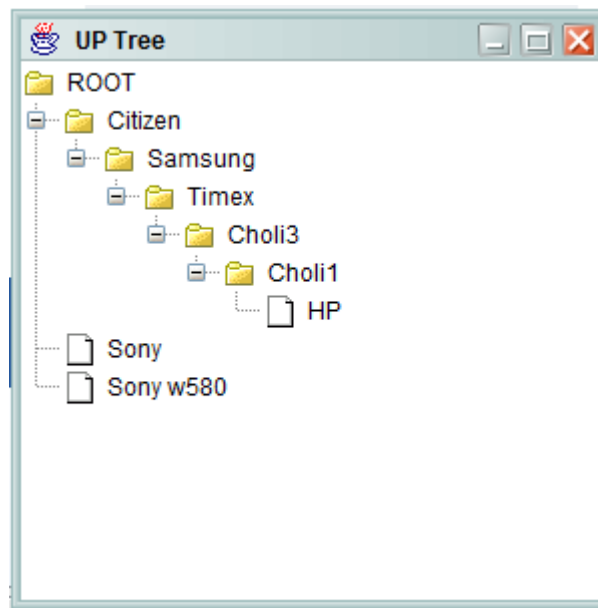

Figure 4: Output representation

VI. CONCLUSION AND FUTURE WORK

The overall performance of an e-commerce website can be analysed by this proposed technique. Using our approach EQUAS, we can verify website quality based on log data and utility mining helps to eliminate unpromising data by considering parameter profit as a threshold. Existing systems doesn't provide efficiency analysis as well as unpromising item removal mechanism. We used here both DEA and utility mining to check website popularity as well as removal of doubtful data to gain more profit. Implications of this study relate to a new model of a set of criteria that used to measure the efficiency of the design, usability and performance of websites. In case of DEA, it is only as good as the initial selection of input and output variables. With the help of utility mining, high utility item sets will be discovered whose values are beyond user specified threshold.

REFERENCES

1. Zijuan Liu, "Diagnosing Ecommerce website quality Based on DEA," 2nd International Conference on Computer Science and Network Technology, 13565627, 762 – 765, 2012.
2. Vincent S. Tseng, Bai-En Shie, Cheng-Wei Wu, and Philip S. Yu, "Efficient Algorithms for Mining High Utility Item sets from Transactional Databases," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 8, AUGUST 2013.
3. Chuen Tse Kuah, Kuan Yew Wong, Farzad Behrouzi, "A Review on Data Envelopment Analysis (DEA)," 2010 Fourth Asia International Conference on Mathematical/Analytical Modelling and Computer Simulation, 978-1-4244-7196-6, 168 – 173, 2010.
4. Corrado lo Storto, "Evaluating ecommerce websites cognitive efficiency: An integrative framework based on data envelopment analysis," Applied Ergonomics, 44, 1004-1014, 2013.
5. Jyothi Pillai, O.P.Vyas, "Overview of Item set Utility Mining and its Applications," International Journal of Computer Applications (0975 – 8887) Volume 5– No.11, August 2010.
6. Vincent S. Tseng , Cheng-Wei Wu , Bai-En Shie , and Philip S. Yu, "UP-Growth: An Efficient Algorithm for High Utility Item set Mining," KDD'10, July 25–28, 2010.
7. Smita R. Londhe, Rupali A. Mahajan, Bhagyashree J. Bhoyar, "Overview on Methods for Mining High Utility Item set from Transactional Database," International Journal of Scientific Engineering and Research (IJSER), Volume 1 Issue 4, December 2013.
8. Adinarayanareddy B, O Srinivasa Rao, MHM Krishna Prasad, "An Improved UP-Growth High Utility Item set Mining," IJCA, Volume 58– No.2, 25-28, November 2012.

BIOGRAPHY

Varsha Kulkarni is pursuing her Master degree of Engineering in Computer engineering from MGM's college of Engineering and Technology, Kamothe, Navi Mumbai, India. She also received Bachelor of Engineering in Computer engineering from MGM's college of Engineering and Technology, Kamothe, Navi Mumbai, India. Her research interests are data mining, web mining, web technology, website designing etc.

Vilas Jadhav is an Assistant Professor in Computer Engineering department, MGM's college of Engineering and Technology, Kamothe, Navi Mumbai, India. His research interests are Data mining, Software Engineering, Web Engineering etc.