International Journal of Computational Intelligence and Applications
AUTHENTIC USER BEHAVIORAL AND OPINION PATTERN MINING ON PRODUCT PERFORMANCE ANALYSIS
--Manuscript Draft--

Abstract:	With the rapid expansion of the Internet, consumers' ability to create and distribute material has resulted in the establishment of active electronic communities that provide a plethora of product knowledge. However, because of the large number of evaluations that are typically available for a single product, it is more difficult for individuals as well as manufacturers to choose the top reviews and discover a product's genuine underlying quality. We re-examine the impact of reviews on economic outcomes such as product sales, as well as how different elements influence social outcomes such as perceived usefulness. Various strategies for estimating the usefulness and economic impact of product reviews by mining text and reviewer characteristics have been discussed. To discover major text-based elements, we look at a variety of components of review text, such as subjectivity levels, various readability events, and the number of spelling errors. In this paper, we look at the economic impact of product reviews with a maximum threshold, as well as related studies. Comparisons are also made between various plans to demonstrate the benefits and drawbacks. The performance study of the Estimating the Helpfulness and Economic Impact of Product Reviews on the basis of decision threshold, classification rate, false pattern rate, and product index is presented in this work.
Suggested Reviewers:	

AUTHENTIC USER BEHAVIORAL AND OPINION PATTERN MINING ON PRODUCT PERFORMANCE ANALYSIS

S.Saravanankumar[1*], T.Saravanan[2], M. Narayanan [3], T. Poongothai[4], Dr. P. Santosh Kumar Patra[5] , S. Bavankumar[6]

[1*]Department of CSE, St.Martin's Engineering College, Secunderabad
sar112113118@gmail.com

[2]Department of CSE, St.Martin's Engineering College, Secunderabad
tsaravcse@gmail.com

[3]Department of CSE, St.Martin's Engineering College, Secunderabad
narayanancse2020@gmail.com

[4]Department of CSE, St.Martin's Engineering College, Secunderabad
drpoongothaicse@smec.ac.in

[5]Department of CSE, St.Martin's Engineering College, Secunderabad
principal@smec.ac.in

[6]Department of CSE, St.Martin's Engineering College, Secunderabad
sbavankumarcse@gmail.com

ABSTRACT

With the rapid expansion of the Internet, consumers' ability to create and distribute material has resulted in the establishment of active electronic communities that provide a plethora of product knowledge. However, because of the large number of evaluations that are typically available for a single product, it is more difficult for individuals as well as manufacturers to choose the top reviews and discover a product's genuine underlying quality. We re-examine the impact of reviews on economic outcomes such as product sales, as well as how different elements influence social outcomes such as perceived usefulness.

Various strategies for estimating the usefulness and economic impact of product reviews by mining text and reviewer characteristics have been discussed. To discover major text-based elements, we look at a variety of components of review text, such as subjectivity levels, various readability events, and the number of spelling errors. In this paper, we look at the economic impact of product reviews with a maximum threshold, as well as related studies. Comparisons are also made between various plans to demonstrate the benefits and drawbacks. The performance study of the Estimating the Helpfulness and

Economic Impact of Product Reviews on the basis of decision threshold, classification rate, false pattern rate, and product index is presented in this work.

1 INTRODUCTION

The significant resource in modern marketing strategies is Customers. Therefore, it is necessary to enterprises and organization to productively obtain new customers and keep hold of high value customers. To attain these aims, many enterprises plan to expand their own customers' data with a lot of of database tools which can be analyzed to accomplish the customer behavioral and applied to expand new business strategies.

With the fast growth of the Internet, product connected word of-mouth conversations have migrated to online markets, creating active electronic communities that offer a wealth of information. Reviewers donate time and energy to produce reviews, enabling a social structure that provides benefits both for the users and the firms that crowd electronic markets. In such a context, "who" says "what" and "how" they say it, matters.

On the flip side, a huge number of reviews for a solitary product may also make it harder for individuals to track the gist of users' discussions and assess the true fundamental quality of a product. Recent work has exposed that the sharing of an overwhelming majority of reviews posted in online markets is bimodal. Reviews are moreover allotted a tremendously high rating or an enormously low rating. In such situations, the average numerical star rating assigned to a product may not express a lot of information to a potential buyer or to the manufacturer who tries to understand what aspects of its product are important.

As an alternative, the reader has to convert the actual reviews to inspect which of the positive and which of the negative attributes of a product are of interest. So far, the most excellent attempt for ranking reviews for consumers comes in the form of "peer reviewing" in review forums, where customers give "helpful" votes to additional reviews in order to signal their informative ness. Regrettably, the helpful votes are not a useful feature for ranking recent reviews: the helpful votes are accumulated over a long period of time, and hence cannot be used for review placement in a short- or medium term time frame.

Similarly, merchants need to know what aspects of reviews are the most informative from consumers' perspective. Such reviews are probable to be the most helpful for merchants, as they hold valuable information about what aspects of the product are driving the sales up or down.

Economic theory has recognized that a business derives 80% of its income from 20% of its customers. Conversely, instead of targeting all prospects evenly or providing the similar offers to all customers, enterprises choose only those individuals that gather particular productivity levels based on preceding behavior or entity needs. As a result, assuming there is identical pattern between customer behaviors. Many method have been introduced to achieve better knowing of customer behaviors, the "behavioral scoring models" is one of the majority successful technique that help decision makers to understand their customer behaviors. Behavioral scoring models assist to examine purchasing behavior of customers. These models are extremely applying data mining approaches.

2 LITERATURE REVIEW

Analysis of customer's behavior that isn't limited to customer's benefit in their general uses of banking services and also is measured as channel's charges. K-means for

customer segmentation according to important characteristic of customers and their behaviors is used a channel [1]. The result of segmentation of customer's profile is according to their behavior which will assist the bank for preservation strategies of existing customers and attracting new customers. Author [2] inspect the relative importance of the three broad feature categories: 'reviewer-related' features, 'review subjectivity' features, and 'review readability' features, and discover that by means of any of the three feature sets results in a statistically equivalent presentation as in the case of using all available features. It is a first study that integrates econometric, text mining, and predictive modeling techniques toward a more complete analysis of the information captured by user-generated online reviews in order to estimation their helpfulness and economic impact.

A large analysis of illegal activity reveals that all criminal performance shares a common set of universal principles [3]. A micro simulation model can be drawn out by interlinking the universal principles with the attributes of the individuals for profiling the criminal behavior. The process of mining technology works on common event logs that have no workflow cases reference, and name the new technology as behavior pattern mining [4]. It gives out a concise survey on issues, challenges, approaches and related tools in behavior pattern mining area, and compares behavior pattern mining with workflow mining technology, which is the other sub field of process mining.

A well-known statistical method which relies on the empirical probability distribution is used to expose trends in the power signal data [5]. These trends are altered if a) different data sampling rates are assumed, and b) a privacy algorithm is applied to defend the power data of different home appliances. Our results propose that the isolation of individual behavior types is uncovered even if comparatively uncommon measurements are obtained.

User Navigation Behavior Mining (UNBM) mostly studies the problems of extracting the attractive user access patterns from user access sequences (UAS), which are frequently used for user access prediction and web page suggestion [8]. Through analyzing

the real world web data, we discover the majority of user access sequences carrying hybrid features of dissimilar patterns, rather than a solitary one. Therefore, the methods that classify one access succession into a single pattern can barely get hold of good superiority results.

We expand and implement a domain-specific interactive QA system oriented to Artificial Intelligence [10]. The course ontology, predefined to describe the skeleton of AI course, is used to produce the structure of our interactive QA system. Students can pretense and look through questions and answers on their favorite boards.

Knowledge about computer users is very helpful for assisting them, predicting their prospect actions or detecting masqueraders. An approach for creating and recognizing mechanically the behavior profile of a computer user is presented [6].The EM algorithm that iteratively maximizes this criterion [7] is based on the maximization of a widespread probability criterion, which can be interpreted as a quantity of agreement between the statistical model and the unsure observations.

The importance of spatial and spatio-temporal data mining is rising with the growing occurrence and consequence of large geo-spatial datasets such as maps, repositories of remote-sensing images, trajectories of moving objects generated by mobile devices, etc [9]. It is used to examine spatial and spatio-temporal data to take out interesting, useful, and non-trivial patterns.

Web usage mining focuses on techniques that might forecast user behavior while the user interacts with the Web. It tries to build sense of the data generated by the Web surfer's sessions or behaviors. There is a try to provide an impression of the state of the art in the research of web usage mining [15], while discussing the most pertinent tools obtainable in the sphere as well as the niche requirements that the present variety of tools lack.

Web mining solution to business intelligence is to find out hidden patterns and business strategies from their customer and web data [11]. Web mining attempts to decide useful knowledge from secondary data obtained from the interactions of the users with the web. Web Usage Mining is the procedure of applying data mining techniques to the discovery of usage patterns from Web data and is under attack towards applications. It mines the secondary data derived from the connections of the users throughout certain period of Web sessions [12]. Web usage mining consists of three phases, specifically preprocessing, pattern discovery, and pattern analysis.

Cluster similar Web user [14], by consider two factors that the page-click number and Web browsing time, which is stored in the Web log, and the dissimilar quantity of influence of the two factors. An approach for discovering and tracking developing user profiles describes how the exposed user profiles can be enriched with unambiguous information need that is contingent from search queries extracted from Web log data [13].

The above listed research gap, motivated us to develop an authentic user behavior and opinion pattern mining suited to applications.

3 METHODOLOGIES

The different work involved in "Performance analysis of authentic user behavioral and opinion pattern mining on product reviews" is:

Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics (MTRC)

Beyond the product-specific data, we also composed all reviews of a product since the product was unconfined into the market. For each review, we recover the definite textual content of the review and the review rating of the product agreed by the reviewer. The rating that a reviewer allocates to the reviewed product is denoted by a number of stars

on a scale of 1 to 5. From the textual content, we generated a set of variables at the lexical, grammatical, and at the stylistic level.

Review Helpfulness: Amazon has a voting system whereby community members offer accommodating votes to rate the reviews of additional community members. Preceding peer ratings emerge straight away above the posted review, in the form, "[number of helpful votes] out of [number of members who voted] found the subsequent review helpful." These obliging and total votes allow us to calculate the portion of votes that evaluated the evaluation as helpful. To have as much correct representation of the proportion of customers that establish the review helpful, we collected the votes, ensuring that there is a important time period after the time the assessment was posted and that there is a important numeral of peer rating votes accumulated for the review.

Reviewer Disclosure: While review valence is probable to authority consumers, there is cause to consider that social information about reviewers themselves (rather than the product or vendor) is probable to be a significant forecaster of customers' buying decisions. On many sites, social information about the reviewer is at least as famous as product information. For example, on sites such as Amazon, information about product reviewers is graphically depicted, highly salient, and sometimes more complete and capacious than information on the products they review.

Reviewer History: Since one of our goals is to forecast the future helpfulness of a review, we wanted to examine whether the past history of a reviewer can be used to predict the usefulness of the future reviews written by the same reviewer. For this, we composed the precedent reviews for each reviewer, and composed the helpful and total votes for every of the past reviews. Using this information, we constructed for each reviewer and for each point in time the past performance of a reviewer.

Particularly, we created two variables, by micro averaging and macro-averaging the past votes on the reviews. The variable reviewer history macro is the relation of all past helpful votes separated by the total number of votes. Likewise, we also shaped the changeable reviewer history micro, in which we first computed the average helpfulness for each of the past reviews and then computed the standard across all past reviews. The difference with the macro and micro versions is that the micro version gives equivalent burden to the neighborliness of all past reviews, while the macro version weights more greatly the significance of reviews that established a large number of votes.

Creating Evolving User Behavior Profiles Automatically (UBPA)

The proposed approach for automatic clustering, classifier design and classification of the behavior profiles of users. The narrative developing user actions classifier is based on Evolving Fuzzy Systems and it takes into account the information that the performance of any user is not fixed, but is quite altering. The proposed approach can be applied to any performance represented by a sequence of events. In order to categorize an observed behavior, our approach, as many other agent modeling methods, creates a collection which contains the different expected behaviors.

However, in our case, this records is not a prefixed one, but is mounting, learning from the explanation of the users real behaviors and, furthermore, it starts to be crowded in "from scratch" by transmission provisionally to the library the main observed user as a prototype. The library, called Evolving-Profile-Library (EPLib), is continually altering, mounting predisposed by the varying user behaviors observed in the environment.

Thus, the proposed approach includes at each step the subsequent two main actions:

1. Creating and evolving the classifier. This action involves in itself two sub actions:

Creating the user behavior profiles: This sub action analyzes the sequences of commands typed by diverse UNIX users online (data stream), and creates the corresponding profiles.

Evolving classifier: This sub action includes online learning and updating of the classifier. It including the potential of each performance to be a prototype, stored in the EPLib.

2. User classification. The user profiles shaped in the previous action are connected with one of the prototypes from the EPLib, and they are confidential into one of the lessons formed by the prototypes.

Maximum likelihood estimation from Uncertain Data in the Belief Function Framework (BFF)

Let X is a variable enchanting value in a restricted domain called the frame of discernment. Uncertain information about X may be represented by a mass function m. Function m is said to be normalized of any subset A of such that m (A) > 0 is called a focal element of m. Two special cases are of interest:

1) If m has a single focal element A, it is believed to be categorical and denoted as mA. Such a mass function encodes a piece of evidence. There is a one-to one correspondence between subsets A of and categorical mass functions mA.

2) If all focal elements of m are singletons, then m is said to be Bayesian. There is a one-to-one correspondence flanked by probability distributions. Bayesian mass functions are thus correspondent to probability distributions. To each normalized mass function m, we might connect belief and plausibility functions

Each quantity Bel(A) may be interpreted as the degree to which the proof supports A, while Pl(A) can be interpreted as an upper bound on the degree of support that could be assigned to A if more specific information became available. If m is Bayesian, then

functions Bel is equal to Pl and it is a probability measure. Another special case of interest is that where m is consonant, i.e., its focal elements are nested. The contour function pl is then the connected possibility distribution. As a result, the theory of belief functions can be measured as having superior expressive power than possibility theory.

Mining the Banking Customer Behavior Using Clustering (CBC)

Banks looking for newer and improved behavior to distinguish themselves from their competitors, customer clustering one of significant method to rich this consequence; Customer clustering is the make use of past transaction data to separate customer to the analogous groups. The results fashioned are based on the assumptions that the customer behavior follows patterns comparable to past pattern and repeats in the future.

Therefore, there might not be an improved time than at the present to examine the significance of an effectual new marketing policy using the customer behavior examine. The decisions to be made comprise which target groups of customers will be confident to use supplementary, what terminal type to assign, how predictable likelihood of receiving new products, whether to encourage new products to target groups of customers, and, how to administer groups of customers to affluent the customer approval and direct marketing.

Conversely, attempts to create high-quality customer behavior analysis might be restricted by the poor quality of data, poor relevant of data, or the volume of data needing to be processed. Database marketing (DM) is a methodical approach to the gathering, consolidation, and processing of customer data to assist the marketers' improved target their markets efforts to obtainable customers.

Further DM analyzes customer data to appear for patterns to use these patterns for an additional targeted variety of the consumers. Over the decades, many database marketing tools were urbanized and used in a variety of marketing. Some of the majority

accepted tools include: the RFM (recency, frequency, monetary), Formula, the behavior segment of the obtainable customers and the lifetime value of a customer.

Association Rules Inducer

The study's aim is to attempt to discover out most significant and Practical patterns in bank databases so as to it might better appreciate diverse behaviors about dissimilar customers and expand new strategies to offer better service and rewarding their requirements better than the opposition. K-means clustering model to categorize customers into clusters with communal characteristics. The employment of mining association rules was used to generate customer profile in each cluster. The purpose of association rule extraction is to find out significant relationships among items or features that happen regularly in a transaction database.

4 PERFORMANCE RESULT

In this section, it demonstrates the performance analysis of various product review schemes through experiments by examining the user behavioral and opinion pattern mining. It is measured in terms of

   i)   Decision threshold

   ii)   Classification rate

   iii)   False pattern rate

   iv)   Product Index

Decision Threshold

No. of users	Decision Threshold			
	MTRC Method	UBPA Method	BFF Scheme	CBC Model
10	83	70	51	62
20	88	72	49	59
30	92	69	46	61
40	96	75	53	65
50	97	78	59	67
60	99	82	63	67
70	99	84	65	66

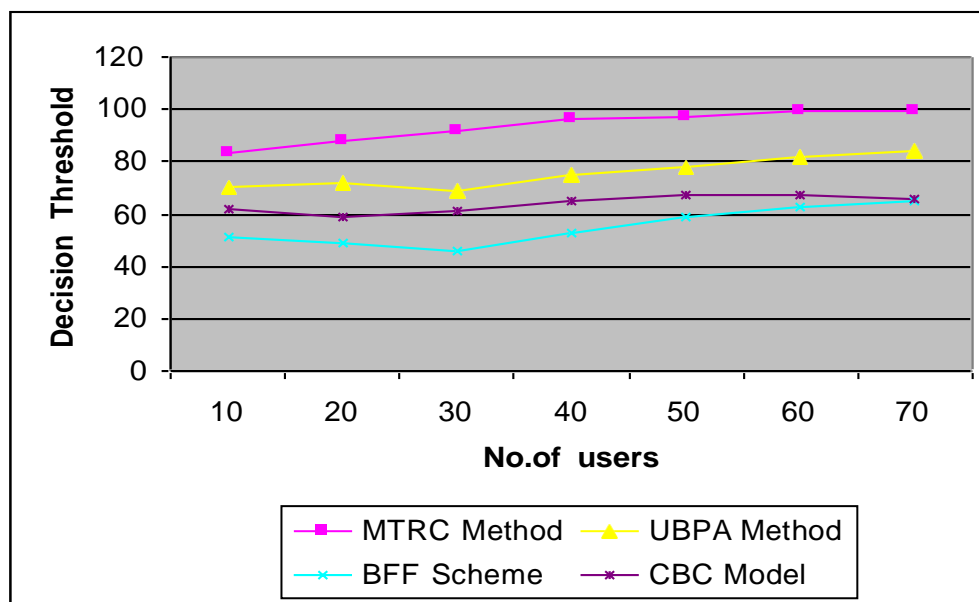**Table 4.1 No. of users Vs Decision Threshold**

Fig 4.1 No. of users Vs Decision Threshold

Fig. 4.1 plots the no. of users with the decision threshold using different product review schemes. This result shows that as the number of user's increases and the decision threshold also increases dramatically. Decision Threshold is the boundary beyond which a radically different state of affairs exists. Decision Threshold is in higher ratio in Mining Text and Reviewer Characteristics (MTRC) Scheme compared with UBPA Scheme, BFF Model, and Customer behavior using Clustering. In this experiment, Mining Text and Reviewer Characteristics (MTRC) model produces better result than other schemes.

Classification rate

No. of users in training data	Classification rate (%)			
	MTRC Method	UBPA Method	CBC Model	BFF Scheme

50	90	80	70	48
100	92	82	68	52
150	93	83	71	55
200	95	85	75	58
250	95	86	76	59
300	96	86	79	57
350	99	87	77	59

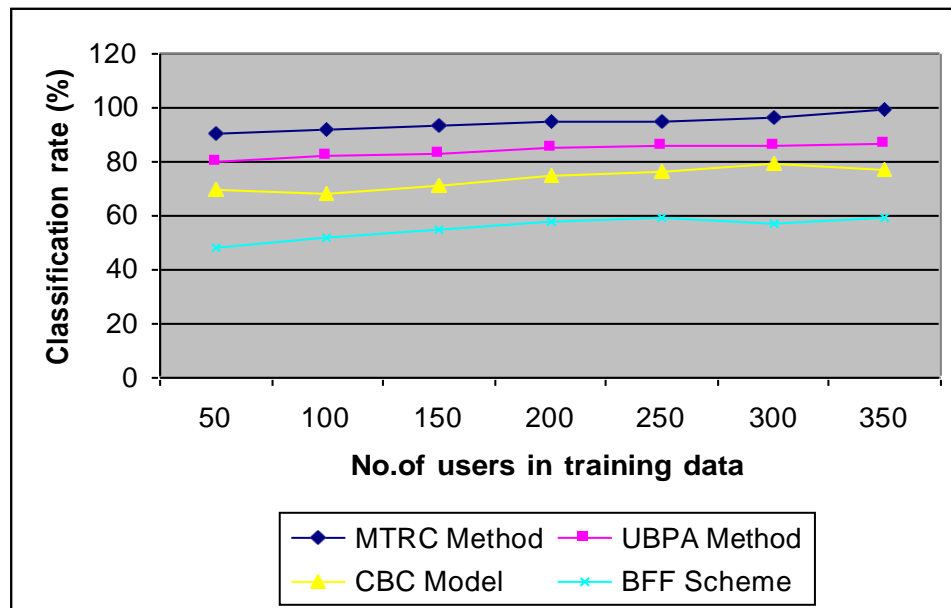**Table 4.2 No. of users in training data Vs Classification rate**

Fig 4.2 No. of users in training data Vs Classification rate

Fig 4.2 shows classification rate of various user training data. Particularly, our analysis relies greatly on MTRC scheme. To check the performance of classification rate from the training data, setup a test which considers the classification of the MTRC on data ranging from 50 to 350 users. From the Figure 4.2 it can be seen that the MTRC provides a prolong classification when compared to the other existing system. It can also be applied to other type of users such as users of e-services, digital communications, etc.

False Pattern rate

Users Product Trend ratio	False Pattern Rate			
	BFF Scheme	MTRC Method	UBPA Method	CBC Model
15	0.2	1.1	1.8	2.6
30	0.4	1.3	2.1	2.8
45	0.5	1.4	2.3	2.9
60	0.6	1.5	2.3	2.9
75	0.7	1.6	2.5	3.1
90	0.7	1.6	2.6	3.2
100	0.9	2.0	2.7	3.2

Table 4.3 Users Product Trend ratio Vs False Pattern Rate

The above table (Table 4.3) described the false pattern rate of MTRC Scheme with the various existing system.
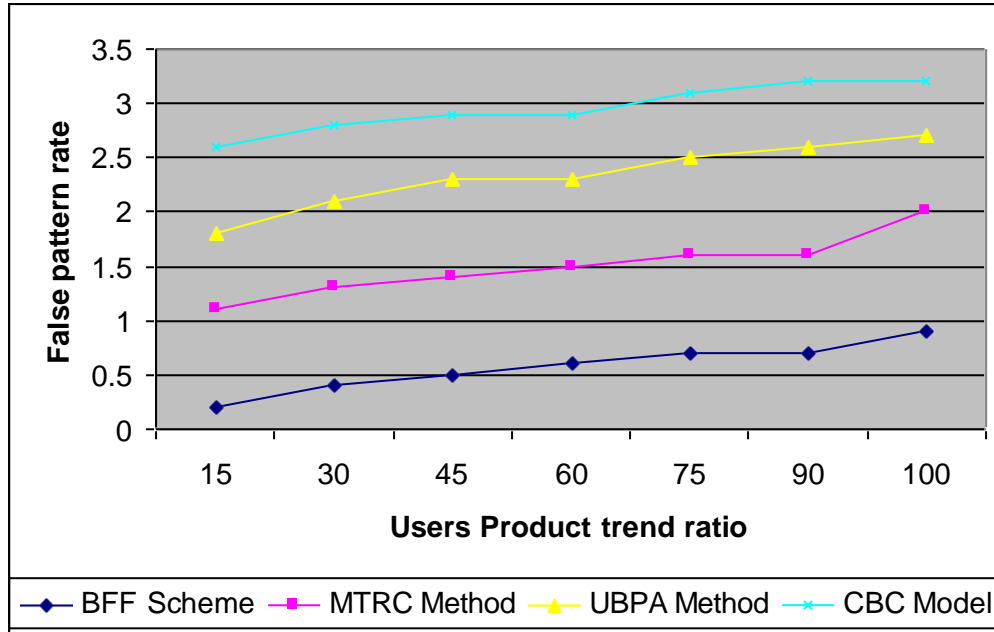

Fig 4.3 Users Product Trend ratio Vs False Pattern Rate

Fig 4.3, described the false pattern rate for the various product trend ratio. In the MTRC scheme the variance in the false pattern rate would be 10-15% high when compared to BFF method, UBPA Scheme and CBC Model.

Product Index

No. of users	Product Index (%)			
	MTRC Method	CBC Model	BFF Scheme	UBPA Method
5	99	89	85	70

10	100	91	87	72
15	96	92	88	75
20	97	92	89	76
25	98	94	89	76
30	98	94	90	78
35	99	94	91	79

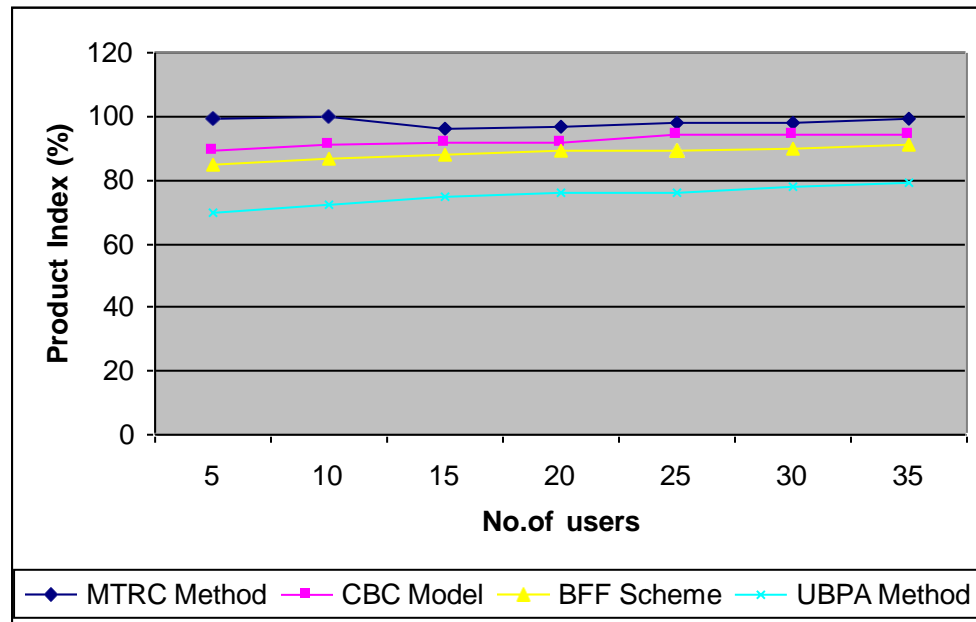**Table 4.4 No. of users Vs Product Index (%)**

Fig 4.4 No. of users Vs Product Index (%)

Fig 4.4, described the product index of four schemes. Comparison result of the MTRC Scheme with an other method is measured.When the user increases the product index also increases drastically. It is measured in terms of percentage. The MTRC and the CBA method is approximately 5% high in product index when compared to the BFF Scheme and User Behavior Profiles Automatically (UBPA).

CONCLUSION

This paper discussed the various methods of economic impact of product reviews. Comparisons are made to explain the advantages and limitations of different product reviews. Performance analyses of these schemes are evaluated through the experiments. Experimental results demonstrate that some of the schemes support opinion decision threshold and some of the techniques support user behavioral profile factor. Various schemes are examined and their performance is evaluated on four criteria: Decision Threshold, product index, false pattern rate, classification rate. From the experimental results, Estimating the Helpfulness and Economic Impact of Product Reviews by Mining Text and Reviewer Characteristics (MTRC) scheme performs well in three criteria Decision threshold, Classification rate and product index compared with User Behavior Profiles Automatically (UBPA), Belief Function Framework (BFF) and Customer Behavior Using Clustering (CBC).

REFERENCES

[1] Reza Baradaran Kazem Zadeh., Ahmad Faraahi., Amir Mastali., "Profiling bank customers behaviour using cluster analysis for profitability," International Conference on Industrial Engineering and Operations Management, 2011

[2] Anindya Ghose., Panagiotis G. Ipeirotis., "Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics" IEEE Transaction on Knowledge and Data Engineering, 2010

[3] Hussain, K.Z., Durairaj, M., Farzana, G.R.J., " Criminal behavior analysis by using data mining techniques" International Conference on Advances in Engineering, Science and Management (ICAESM), 2012

[4] Jinliang Song., Tiejian Luo., Su Chen., "Behavior Pattern Mining: Apply Process Mining Technology to Common Event Logs of Information Systems," IEEE International Conference on Networking, Sensing and Control, 2008.

[5] Kalogridis, G., Denic, S.Z., " Data Mining and Privacy of Personal Behaviour Types in Smart Grid.," International Conference on Data Mining Workshops (ICDMW), 2011

[6] Jose Antonio Iglesias., Plamen Angelov., Agapito Ledezma., Araceli Sanchis., "Creating Evolving User Behavior Profiles Automatically.," IEEE, 2012

[7] Thierry Denoeux., "Maximum likelihood estimation from Uncertain Data in the Belief Function Framework.," IEEE Transaction on knowledge and Data Engineering, 2010

[8] Li Xue., Ming Chen., Yun Xiong., Yangyong Zhu., "User Navigation Behavior Mining Using Multiple Data Domain Description," International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010

[9] Bogorny, V., Shekhar, S., "Spatial and Spatio-temporal Data Mining.," IEEE International Conference on Data Mining (ICDM), 2010

[10] Zhongying Zhao., Shengzhong Feng., Yongquan Liang., Qingtian Zeng ., Jianping Fan., "Mining User's Interest from Interactive Behaviors in QA System.," International Workshop on Education Technology and Computer Science, 2009.

[11] Tiwari, S., Razdan, D., Richariya, P., Tomar, S., "A web usage mining framework for business intelligence.," International Conference on Communication Software and Networks (ICCSN), 2011

[12] Etminani, K., Delui, A.R., Yanehsari, N.R., Rouhani, M., "Web usage mining: Discovery of the users' navigational patterns using SOM.," International Conference on Networked Digital Technologies, 2009.

[13] Nasraoui, O., Soliman, M., Saka, E., Badia, A., Germain, R.," A Web Usage Mining Framework for Mining Evolving User Profiles in Dynamic Web Sites," IEEE Transactions on Knowledge and Data Engineering, Volume: 20 , Issue: 2, 2008

[14] Yaxiu Yu., Xin-wei Wang., "Web Usage Mining Based on Fuzzy Clustering.," International Forum on Information Technology and Applications, 2009.

[15] Chhavi Rana., "A Study of Web Usage Mining Research Tools.," International journal on Advanced Networking and Applications Volume: 03, Issue: 06, 2012