Big Data Analytic Approach in Identifying High Utility Product with User Interest Behavioral Analysis

A.Mary
Assistant Professor, Department of Computer Science And Engineering
Alpha College Of Engineering, Thirumazhisai, Tamil Nadu, India

Elakya B
UG Scholar, Department of Computer Science And Engineering
Alpha College Of Engineering, Thirumazhisai, Tamil Nadu, India

Lavanya R
UG Scholar, Department of Computer Science And Engineering
Alpha College Of Engineering, Thirumazhisai, Tamil Nadu, India

Ezhil Arasi S
UG Scholar, Department of Computer Science And Engineering
Alpha College Of Engineering, Thirumazhisai, Tamil Nadu, India

Abstract- In the existing system, the overall transaction of all the users is analyzed and maximum profit yielding purchase of item sets is extracted. This process will exhibit the maximum profit based analysis. In this paper as a modification we are adding up User profile based Purchase system. The user behavioral interest is gathered through processing of tweets by using tweet stream clustering algorithm. Items are displayed based on the User's Interest. Related Items that are brought by users with similar interest are recommended to other users by grouping users with similar interest together.

Keywords – Utility mining, high utility item set mining, top-k pattern mining, top-k high utility item set mining

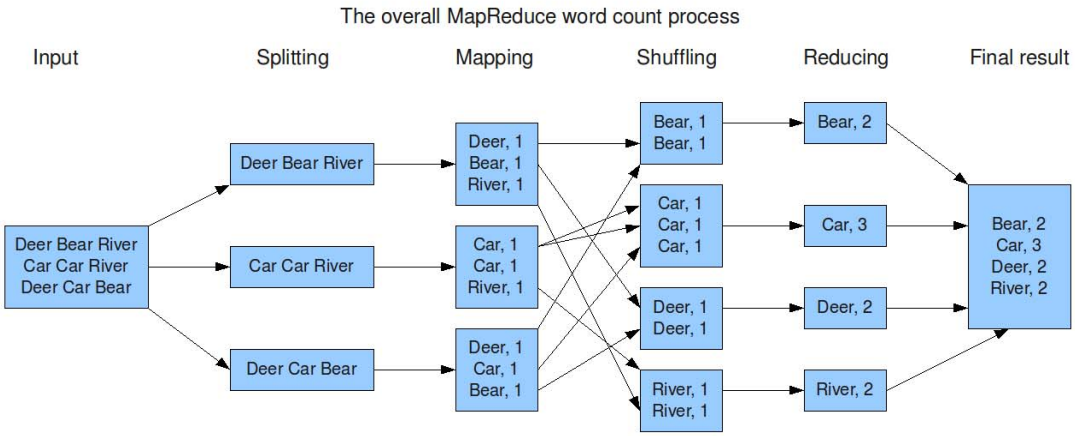
I. INTRODUCTION

The use of "micro-blogging" services, such as Twitter, has exploded exponentially in recent years. For example, currently, millions of Twitter users post millions of 140-character messages, called "Tweets," about topics ranging from daily activities, to opinions, to links to funny pictures. Beyond the large collection of user generated text, Twitter also has a social network aspect, allowing users to publicly message one another directly, and set up a social network of people who follow one another's Tweets. This rich relational and textual setting has spurred research in a number of areas (beyond traditional network analysis) .In this paper, we focus on discovering interest of a particular Twitter user. we use map and reduce framework to identify informative tweets .Using the obtained interest we identify top k high utility item sets. In traditional HUI mining, the search space can be efficiently pruned by the algorithms by using a given min_util threshold. However, in the scenario of top-k HUI mining, no min_util threshold is provided in advance over until being satisfied with the results. This process is both inconvenient and time-consuming. To precisely control the output size and discover the item sets with the highest utilities without setting the thresholds, a promising solution is to redefine the task of mining HUIs as mining top-k high utility item sets (top-k HUIs). The idea is to let the users specify k, i.e., the number of desired item sets, instead of specifying the minimum utility threshold. Setting k is more intuitive than setting the threshold because k represents the number of item sets that the users want to find whereas choosing the threshold depends primarily on database characteristics, which are often unknown to users.

II. PROPOSED ALGORITHM

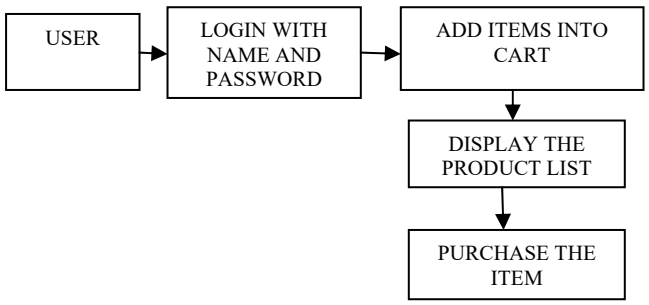
A.TWEET PROCESSING

Twitter like webpage, user can register in twitter site and go for login by giving valid user name and password. If the user name and password is valid the user can login into home page. Once we login in home page based on user interest they go for likes and tweets about the products. So this likes is going to monitor by server and stored in data base. All the tweets of the user is also retrieved and stored in database. This information giving input to hadoop server. Map and reduce frame work is used to identify the informative tweets.



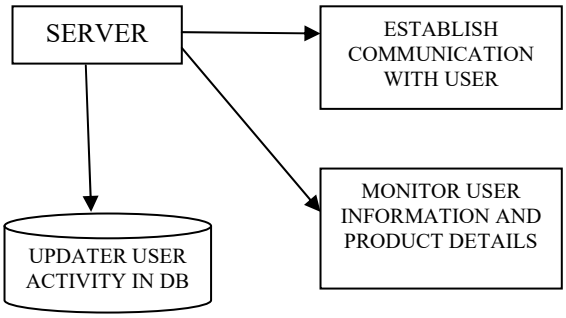
B.PURCHASE PORTAL

Then user can login using particular username and password. All the inserted also updated items are added into the product list. Then user selects wanted items then adds all items into cart products with count of the each item. A warning message will display in dialogue box when the customer type the quantity above the constraint value mentioned in the database. All selected items are displayed in the cart product list, then purchase the required items.



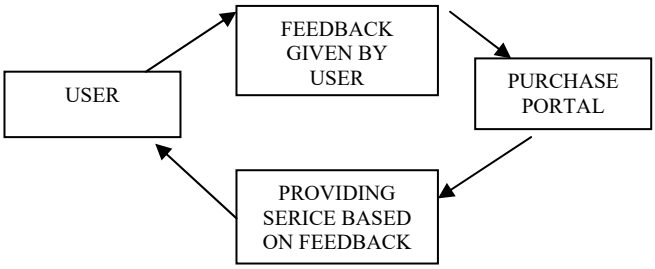
C.BIG DATA SERVER

The Server will monitor the entire User's information in their database and verify them if required. Also the Server will store the entire User's information and product information in the database. Also the Server has to establish the connection to communicate with the Users. The Server will update the each User's activities in its database. The Server will authenticate each user before they access the Application. So that the Server will prevent the Unauthorized User from accessing the Application.



D.FEEDBACK

Process in which the effect or output of an action is 'returned' (feed-back) to modify the next action. Feedback is essential to the working and survival of all regulatory mechanisms found throughout living and non-living nature, and in man-made systems such as education system, online shopping system and economy. As a two-way flow, feedback is inherent to all interactions, whether human-to-human, human-to-machine, or machine-to-machine. In an organizational context, feedback is the information sent to an entity (individual or a group) about its prior behavior so that the entity may adjust its current and future behavior to achieve the desired result. Feedback occurs when an environment reacts to an action or behavior. For example, 'customer feedback' is the buyers' reaction to a firm's products and policies, and 'operational feedback' is the internally generated information on a firm's performance. Response to a stimuli (such as criticism or praise) is considered a feedback only if it brings about a change in the recipient's behavior.



E.PRODUCT RANKING

Based on the feedback value we rate the promising items. Then find out the promising items. Candidate item sets can be generated efficiently with only two scans of database. Efficient top-k HUI mining refers to the discovery of item sets with high utility like profit. So the users can the feedback base product to purchase. This will be useful for the new user to buy the product.

Based on the feedback value we rate the promising items. Then find out the promising items. Candidate item sets can be generated efficiently with only two scans of database. Mining high utility item sets from database refers to the discovery of item sets with high utility like profit. So the users can the feedback base product to purchase. This will be useful for the new user to buy the product.

PROCEDURE: TopK-HUI-Search

Input: (1) $u(P)$: utility-list for a prefix P;
(2) $Class[P]$: a set of itemsets w.r.t. the prefix P;
(3) $ULS[P]$: a set of utility-lists w.r.t.the prefix P;
(4) δ: border minimum utility threshold min_util_{Border};
(5) *TopK-CI-List*: a list for storing candidate itemsets;

Results: (1) Use *TopK-CI-List* to capture all the top-k HUIs

```
01.  For each X={x1, x2,..., xL} ∈Class[P] do
02.  {  If (SUM(X.iutils) ≥ δ)
03.     {  //Raise min_util_Border by the strategy RUC;
04.        δ ← RUC(X, TopK-CI-List);
05.     }
06.     If (SUM(X.iutils) + SUM(X.rutils) ≥ δ)
07.     {  Class[X] ← Ø; ULS[X] ← Ø;
08.        For each Y = {y1, y2,..., yL} ∈Class[P] | yL ≻ xL do
09.        {  Z ← X ∪ Y;
10.           ul(Z) ← Construct(ul(P), X, Y, ULS[P]);
11.           Class[X] ← Class[X] ∪ Z;
12.           ULS[X] ← ULS[X] ∪ ul(Z);
13.        }
14.        TopK-HUI-Search(X, ULS[X], Class[X], δ, TopK-CI-List);
15.     }
16.  }
```

III. Experiment and Result

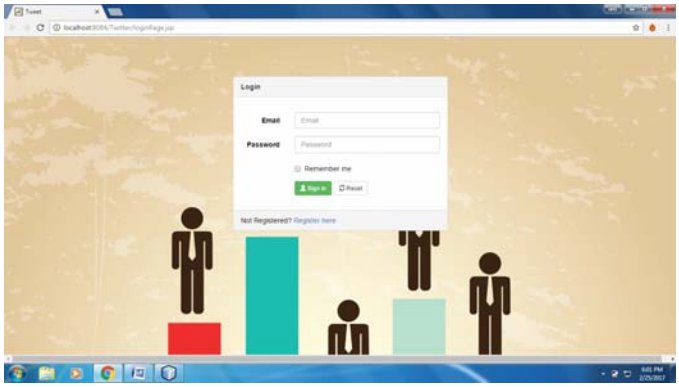


FIG:1

FIG:2

IV.CONCLUSION

In this paper, we have studied the problem of top-k high utility item sets mining, where k is the desired number of high utility item sets to be mined. Two efficient algorithms TKU (mining Top-K Utility item sets) and TKO (mining Top-K utility item sets in One phase) are proposed for mining such item sets without setting minimum utility thresholds. TKU is the first two-phase algorithm for mining top-k high utility item sets, which incorporates five strategies PE, NU, MD, MC and SE to effectively raise the border minimum utility thresholds and further prune the search space. On the other hand, TKO is the first one-phase algorithm developed for top-k HUI mining, which integrates the novel strategies RUC, RUZ and EPB to greatly improve its performance. Empirical evaluations on different types of real and synthetic datasets show that the proposed algorithms have good scalability on large datasets and the performance of the proposed algorithms is close to the optimal case of the state-of-the art two-phase and one-phase utility mining algorithms

REFERENCES

[1] K. Chuang, J. Huang, and M. Chen, “Mining top-k frequent patterns in the presence of the memory constraint,” VLDB J., vol. 17, pp. 1321–1344, 2008
[2] J. Han, J. Pei, and Y. Yin, “Mining frequent patterns without candidate generation,” in Proc. ACM SIGMOD Int. Conf. Manag. Data, 2000, pp. 1–12
[3] J. Han, J. Wang, Y. Lu, and P. Tzvetkov, “Mining top-k frequent closed patterns without minimum support,” in Proc. IEEE Int. Conf. Data Mining, 2002, pp. 211–218
[4] Y. Lin, C. Wu, and V. S. Tseng, “Mining high utility itemsets in big data,” in Proc. Int. Conf. Pacific-Asia Conf. Knowl. Discovery Data Mining, 2015, pp. 649–661
[5] V. S. Tseng, C. Wu, B. Shie, and P. S. Yu, “UP-Growth: An efficient algorithm for high utility itemset mining,” in Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2010, pp. 253–262
[6] J. Yin, Z. Zheng, L. Cao, Y. Song, and W. Wei, “Mining top-k high utility sequential patterns,” in Proc. IEEE Int. Conf. Data Mining, 2013, pp. 1259–1264
[7] S. Krishnamoorthy, “Pruning strategies for mining high utility itemsets,” Expert Syst. Appl., vol. 42, no. 5, pp. 2371–2381, 2015.
[8] S. Petrovic, M. Osborne, and V. Lavrenko, “Streaming First Story Detection with application to Twitter,” in Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, 2010, pp. 181–189
[9] S. Kumar, H. Liu, S. Mehta, and L. V. Subramaniam, “From tweets to events: Exploring a scalable solution for Twitter streams,” arXiv preprint arXiv:1405.1392, 2014.
[10] C. Aggarwal and K. Subbian, “Event Detection in Social Streams,” in Proceedings of SIAM International Conference on Data Mining (SDM), 2012, pp. 624–635.