INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 3 ISSUE 2 (2017) PAGES 1786 - 1790
RECEIVED: 09-03-17 PUBLISHED: 24-03-17.

March 24, 2017

BIG DATA ANALYTIC APPROACH IN IDENTIFYING HIGH UTILITY PRODUCT WITH USER INTEREST BEHAVIORAL ANALYSIS

[1]P.Keerthana, [2]S.Rithika, [3]S.Sowmya,

[1,2]UG Scholar, Dept of computer science and Engineering, Sri Balaji Chockalingam Engineering College, Irumbedu, Arni

[3]Assistant Professor, Dept of computer science and Engineering, Sri Balaji Chockalingam Engineering College, Irumbedu, Arni

Abstract

In the Tweets their raw form, while being informative, can also be overwhelming. It is a nightmare to plow through millions of tweets which contain enormous amount of noise and redundancy , we are analyzing the overall transaction of all the users and we are extracting maximum profit yielding purchase of frequency item set is analyzed. The process will exhibit the maximum profit based analysis. we are including the User profile based Purchase system in the twitter application. Twitter like application is designed where Users Likes in this page & likes in the Purchase website are monitored parallel. Purchase Portal will have two options like General Purchase & Profile based Purchase. In Profile based purchase, Items are displayed based on the User's Interest. Related Items and Items which are purchased more often are also displayed to the user based on the User Interest.

Index Terms: Tweet Stream Clustering, FIM, HUI, TWU.

1. INTRODUCTION:

Frequent item set mining (FIM) is a fundamental research topic in data mining. The traditional FIM may discover a large amount of frequent but low-value item sets and lose the information on valuable item sets having low selling frequencies. It cannot satisfy the requirement of users who desire to discover item sets with high utilities such as high profits. To address these issues, utility mining emerges as an important topic in data mining and has received extensive attention in recent years. In utility mining, each item is associated with a utility (e.g. unit profit) and an occurrence count in each transaction (e.g. quantity). The utility of an item set represents its importance, which can be measured in terms of weight, value, quantity or other information depending on the user specification. An item set is called high utility item set (HUI) if its utility is no less than a user-specified minimum utility threshold min_util.HUI mining is essential to many applications such as streaming analysis market analysis mobile computing and biomedicine. A classical TWU model-based algorithm consists of two phases. Initially we called the complete set of HTWUIs are found. Another is called , all HUIs are obtained by calculating the exact utilities of HTWUIs with one database scan. Depending on the threshold, the output size can be very small or very large.

The 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)**.**

2. RELATED WORKS:

Mining top-k Frequent patterns of memory constraint of a practicably interesting mining task to retrieve *top-k* (*closed*) item sets in the presence of the memory constraint. An improving the mining efficiency or on reducing the memory size by best effort, we attempt to specify the available upper memory size that can be utilized by mining frequent item sets. Mining high utility item sets from a transactional database refers to the discovery of item sets with high utility like profits. The problem of producing a large number of candidate item sets for high utility item sets.

The without candidate generation of Mining frequent patterns in transaction databases, time-series databases, and many other kinds of databases has been popularly in data mining research. Most of the adopt an Apriori-like candidate set generation-and-test approach.A new pruning approach for efficiently mining high utility item sets. The proposed approach removes unpromising items early from the database, thus causing better upper bounds of utility values and pruning more item sets.

3. ECHNICAL RIDE:

3.1.USER REGISTRATION:

In client side user can enter all details. Then user can login using particular username and password. All the inserted also updated items are added into the product list. Then select user wanted items then add 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.

3.2. TWITTER LIKE APPLICATION:

We create twitter like application, user can register in twitter application 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 the display of several products is to be done. Based on user interest he go for likes to the products. So this likes is going to monitor by server and stored in data base. This information giving input to hadoop server.

3.3. PURCHASE PORTAL

Consumer buying behaviour is the sum total of a consumer's attitudes, preferences, intentions and decision regarding the consumer's behaviour in the market place when purchasing a product or service. The consumer behaviour draws upon social science disciplines of sociology, and economics. At this stage consumer will make a purchasing decision. The ultimate decision may be based on factors such as price or availability.

INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 3 ISSUE 2 (2017) PAGES 1786 - 1790
RECEIVED: 09-03-17 PUBLISHED: 24-03-17.
March 24, 2017

3.4. 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 in their 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.

3.5. FEEDBACK:

The 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 behaviour so that the entity may adjust its current and future behaviour to achieve the desired result. Feedback occurs when an environment reacts to an action or behaviour.

3.6. 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. 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.

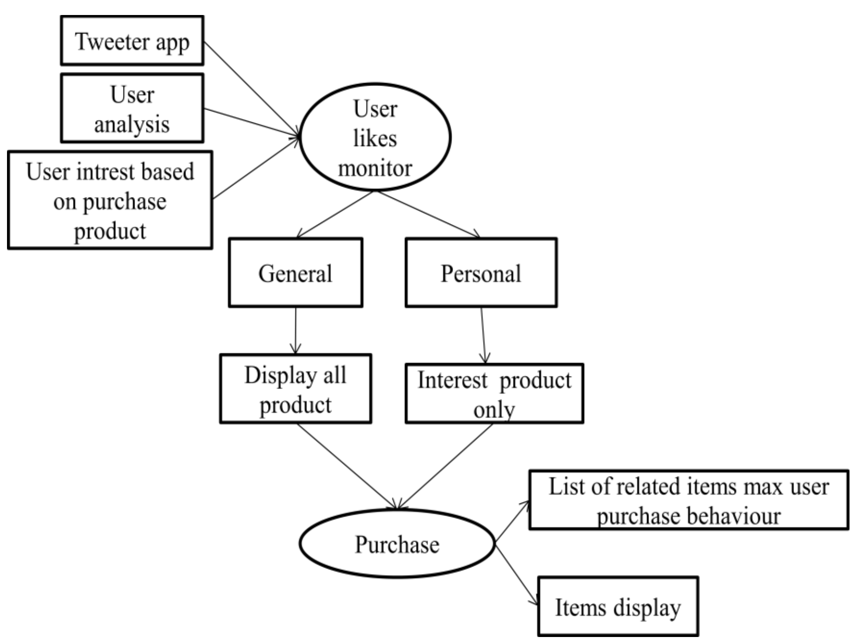

Fig.1.Tweet Stream Clustering Model

4. FUNCTIONAL RIDE:

Social media and news reporting platforms represent a valuable source of information to monitor events occurring around the world in real-time.

The size and speed at which this information is generated and the fact that this information is often produced in a free-text it can be an overwhelming challenge to fully benefit from such data sources.

The most useful feature for humans to group tweets together.

To calculate distances between tweet texts, we chose to process as follows:

1) Text cleaning: we remove punctuation marks, common words in the language used (called stop words) and URLs; we also transform all letters to lower case.

2) Text to words: we split the texts into words.

3) Unit-unit distance: we calculate the Jaccard distance between pairs of texts, which are now sets of words.

Common user similarity- $Su(Pi,Pj)=\frac{\sum u \in Ui \cap Uj\ Piu\ Pju}{\sqrt{\sum u \epsilon Ui\ Piu^2}\sqrt{\sum u \epsilon Uj\ Pju^2}}$ to calculate a tweet ID $St(Pi, Pj) = \frac{|Pi \cap Pj|}{\sqrt{|Pi|}\sqrt{|Pj|}}$.

CONCLUSION:

We have 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.

FUTURE ENHANCEMENT:

Although we have proposed a new framework for top-k HUI mining, it has not yet been incorporated with other utility mining tasks to discover different types of top-k high utility patterns such as top-k high utility episodes, top-k closed high utility item sets, top-k high utility web access patterns and top-k mobile high utility sequential patterns. These leave wide rooms for exploration as future work.

REFERENCES:

[1] N. D. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, and A.T. Campbell, “A survey of mobile phone sensing,” Commun. Mag.,vol. 48, no. 9, pp. 140–150, Sep. 2010.

INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 3 ISSUE 2 (2017) PAGES 1786 - 1790
RECEIVED: 09-03-17 PUBLISHED: 24-03-17.
March 24, 2017

[2] K. K. Rachuri and M. E. Musolesi, "Emotionsense: A mobile phones based adaptive platform for experimental social psychology research," in Proc. 12th ACM Int. Conf. Ubiquitous Comput.,2010, pp. 281–290.
[3] E. C. E. Larson, "Accurate and privacy preserving cough sensing using a low-cost microphone," in Proc. 13th Int. Conf. Ubiquitous Comput., 2011, pp. 375–384.
[4] Y. Jiang and K. E. Li, "Maqs: A mobile sensing system for indoor air quality," in Proc. Int. Conf. Ubiquitous Comput., 2011,pp. 493–494.
[5] J. Lester, T. Choudhury, and G. Borriello , "A practical approach to recognizing physical activities," in Pervasive Computing. Berlin ,Germany: Springer, 2006, vol. 3968, pp. 1–16.
[6] K. Lewin and D. Cartwright, Field Theory in Social Science: Selected Theoretical Papers. London: U.K.: Tavistock, 1952.
[7] C. M. Bishop and N. M. Nasrabadi, Pattern Recognition and Machine Learning. New York, NY, USA: Springer, 2006, vol.1.
[8] R. Bennett, "Queues, customer characteristics and policies for managing waiting-lines in supermarkets," Int. J. Retail Distrib. Manage., vol. 26, no. 2, pp. 78–87, 1998.
[9] I. Constandache and X. Bao, "Did you see bob?: Human localization using mobile phones," in Proc. 16th Annu. Int. Conf. Mobile Comput.Netw., 2010, pp. 149–160.