ISSN: 2321-7782 (Online)
Volume 2, Issue 1, January 2014
International Journal of Advance Research in Computer Science and Management Studies
Research Paper
Available online at: www.ijarcsms.com

Recommender System using Assosiation Rule Mining

Amol Sharadchandra Chaudhari[1]
Information Technology Department
R.K.D.F. Institute of Science and Technology
Bhopal – India

Shrikant Lade[2]
Information Technology Department
R.K.D.F. Institute of Science and Technology
Bhopal – India

Balvirsing Ramsing Chauhan[3]
Computer Science and Engineering Department
R.K.D.F. Institute of Science and Technology
Bhopal – India

Harshal Chandrashekhar Sonar[4]
Electronics and Telecommunication Department
Government Polytechnic Nandurbar
Nandurbar – India

Abstract: ***A typical recommender system includes people recommendation as input, which then system then combines and direct to appropriate recipients. These recommender systems are widely presented in e-commerce websites, search engines, library catalogue and many other customer oriented systems. This paper attempt to propose a system that will work as Recommender system for the Market basket database. The proposed system retrieves the sensitive items presented in the application database using association rule mining. This paper also suggested privacy preserving technique for the customer choices and recommendation. Initially, the system will find out the frequent item sets and generate the association rules. The proposed system calculates the confidence levels between clicked products, between the products placed in the basket and between purchased products respectively. Simultaneously, the preference level is calculated through the linear combination of above three confidence level.***

Keywords: Association Rule Mining, Apriori, Recommender system.

I. INTRODUCTION

With the vast growth of data mining, several research and applications has been proposed in the areas of marketing, business, medical analysis, production, design, bio-informatics and space exploration etc. It has also been used in the computer science research such as association rules, decision tree, graphs and vertices etc. With considering an example of a Retail shop, associations are used for different items which the customer placed in the shopping cart. In market basket database, the association rule must be not visible for competency. Although, there have been lot of analysis is done for hiding sensitive association rules. The related algorithm accepts the some set of sensitive items manually. The market basket database contains large variety of products and items. The selection process from these databases generally takes more amount of time. This paper also presents the privacy preserving technique to preserve sensitive data or information. According to the study, two techniques have proposed namely output privacy and input privacy. The output privacy relates to any malicious inference attacks whereas the input privacy sanitized the raw data itself before performing mining. The proposed recommender system will work for customer's navigational and behavior patterns in market basket. The recommender system advises not only products which are potentially auctioned to each customer but also assists the customers in making purchases.

1.1 ASSOCIATIVE RULE MINING

We have proposed algorithm for discovering large item sets to make multiple passes over the data. Initially, the algorithm will count the support of items and measures the larger one of that i.e. have minimum support. The seed set pass generating new potentially large item-set found to be large in the previous pass. The set was used for generating new potentially large item-sets, called item-set and count the actual support for these candidate item sets are actually large and become the seed for the next

pass. This process continues until no new item sets are found. Many algorithms for generating association rules were presented over time. The known algorithms are Apriori, Eclat and FP-growth. On comparing to these Apriori is the best algorithm to mine association rules. It uses a breadth first search mechanism to count the support of item sets and uses a candidate generation function that attains the downward closure property of support. Examples of areas in which association rules have been used are Credit card transactions, Supermarket purchase, telecommunication product purchase, Banking services, Medical services and insurance claims etc.

1.2 MARKET BASKET ANALYSIS

Association rule of the form if p the q. For ex. 60 % of those who buy Laptop also buy Tablet; 80 % of those who buys Books online also buy DVD online .50% of those who have diabetes also are obese and have high cholesterol. These rules are actionable in that they can be used to target customers for marketing, or for product placing, or more generally to inform decision making. Let z (z_1…..,z_I) be the set of products. Let d be the database that contains set of transactions. Each transaction t→D is item set such that t is a proper subset of z. A transaction t supports X, a set of items in I, if X is proper subset of t. Assume that the items in a transactions or an item set are sorted in linguistics.

This paper is organized as follows: section 2 gives the overview of related studies and research done previously. Section3 provide the brief knowledge of the proposed system. Section 4 describes the conclusion and future work.

II. RELATED WORK

Data mining is a vast technique and several theory and research works have been proposed and implemented. [1] presented the view of mining association rules in large databases. An efficient algorithm is presented that generates all significant association rules between items in databases. [2] includes a security and privacy implications in data mining in which a possible solution is given to this problem. [3] presented the algorithms for balancing privacy and knowledge information in association rule mining that algorithms require only two scans, regardless of the database size and the number of restrictive association rules that must be protected. [4] suggested the work of recommended system related to the ecommerce websites in order with international transactions and communication strategy. [5] proposed an Constant Time Collaborative Filtering Algorithm that uses universal queries to elicit real-valued user ratings on a common set of items and applies principal component analysis (PCA) to the resulting dense subset of the ratings matrix.[6] suggested a novel based recommender system based on collaborative filtering is developed for ecommerce site. The proposed approach analyzes the data captured from the navigational and behavioral patterns of customers, estimates the preference levels of a customer for the products which are clicked but not purchased, and CF is conducted using the preference levels for making recommendations.[7] presented a survey of analysis on recommendation for e-commerce. The challenge was to improve the quality of recommendations to the customers.[8] proposes the work related to the cosmetic business which shows the effective command for relationship between customer and product with help of content based, collaborative filtering and data mining techniques. Our proposed system will perform the similar task with different techniques that includes a appropriate combination of associative mining rules, Hiding sensitive association rule to determine the recommendations in market basket database. The proposed system will results the appropriate and suitable recommendations for the market basket database.

III. METHODOLOGY

The proposed system is divided into three modules:

A. Input database (Data collection and manipulation)

All the related data are related to the product purchase is gathered

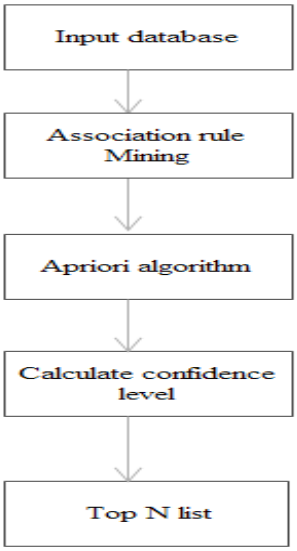

Fig1. General architecture

B. Market basket Analysis

Let D be the database that contains transactions. Each transactions t→D is an item set such that t is a proper subset of I. A transactions t supports Z, a set of items in I, if P is a proper subset of t. An association rule is an implication of the form P→Q, where P and Q are subsets of I and P ∩ Q = φ.

The support of rule P→Q can be calculated by the following equation:

$$\text{Support}(P \rightarrow Q) = \frac{|P \rightarrow Q|}{|D|}$$

where |P→ Q| denotes the transactions containing the item set PQ in the database, |D| denotes the number of transactions in the data.

$$\text{Confidence}(P \rightarrow Q) = \frac{|P \rightarrow Q|}{|P|}$$

where |P| is the number of transactions in database D that contains item set Z.

A rule P→Q is strong if support (P→ Q) ≥ min_support and confidence(P→Q) ≥ min_confidence,

where min_support and min_confidence are two given minimum threshold.

C. Apriori

After the calculation of min_Support and min_Confidence the apriori algorithm is used to generate frequent item sets. The algorithm counts the item in the transactions and discards the infrequent items in the first scan of the database. At this process frequent 1- item set are found. During the second database scan support of 2- candidates are counted. Let (i1,i2) and (i3,i4) where i1<i2 and i3. The two item sets can form a 3 candidate if i1=i3 and (i2,i4) is also frequent. The resulting 3-candidate is the following: (i1, i2, i4). The algorithm terminates if no candidates can be generated or no frequent item sets are found. In this process frequencies (F) of individual items and calculated and stored. The association rules are generated from database using minimum support and minimum confidence. Select all the rules containing min_support. Finally select all the rules and join them. Maximum occurrences of items are selected by user specified value in terms of percentage. The selected sensitive items are used to hide sensitive rules.

D. Confidence level calculation

Three confidence levels between clicked products, between the products placed in the cart, between purchased products are derived as follows:

- Click confidence level(C_{ij}) is defined for the rule P→Q where the former consists of another clicked product and a variable representing a customer's consists of another clicked product and a variable representing a customer's navigational and behavioral pattern, while the result consist of another clicked product.
- Basket placement confidence (B_{ij}) is performed similar to above stage and basket placement of $Item_i$ to $Item_j$ is determined.
- This is performed in similar to the above steps. For ex. "Purchase $Item_i$ and Length of reading time (High) → Purchase $Item_j$".
- The confidence levels calculated in Phase II are used to estimate the preference level for each pair of products. The preference level (r_{ij}) of product i to product j for each customer are estimated through a linear combination of the above three confidence levels. r_{ij} is defined as:

$$\mathbf{r_{ij} = \alpha C_{ij} + \beta B_{ij} + \gamma P_{ij}}$$

where $\mathbf{0 \le \alpha \le 1}$, $\mathbf{0 \le \beta \le 1}$, $\mathbf{0 \le \gamma \le 1}$ and $\mathbf{\alpha + \beta + \gamma = 1}$

E. Making Top N list

After making a Top N list based on the preference levels of a customer for the products not clicked, recommendations are made to each customer.

- Step1: For each customer, clicked item are arranged on left hand side, and connected to items that exist in the consequent of association rule when clicked item are precedent.
- Step2: Among the products on the right hand side for each customer, if it has been clicked, then that Item is excluded from recommendation.
- Step3: If there are duplicate products which are associated with each customer, the higher preference level value is determined as the preference level of a customer for Item.

IV. CONCLUSION

The proposed system is a unique ARM (Association rule Mining)-based recommender system for ecommerce site. The effectiveness of the proposed method will be measured by implementing an experimental electronic products and accessories site. However, the approach is versatile and can be implemented for a variety of ecommerce sites of any scale.

Acknowledgement

I would like to thank to RKDF Institute of Science and Technology, Bhopal. For providing me such a graceful opportunity to become a part of its family. It has been a privilege for me to pursue M.Tech. inInformation Technology from this institute.

References

1. Agrawal, R., & Srikant, R. (1994). Fast algorithm for mining association rules. In Proceedings of the twentieth VLDB conference (pp. 487–499). Santiago.
2. Berson, A., Smith, K., & Thearing, K. (2000). Building data mining applications for CRM. New York: McGraw-Hill.
3. Billsus, D., & Pazzani, M.J. (1998). Learning collaborative information filters. In Proceedings on the fifteenth international conference on machine learning (pp. 46–54). Madison, WI.
4. Kim, D., & Yum, B.-J. (2005). Collaborative filtering based on iterative principal component analysis. Expert Systems with Applications, 28(4), 823–830.

5. Kim, Y. S., Yum, B.-J., Song, J., & Kim, S.-M. (2005). Development of a e-commender system based on navigational and behavioral patterns of customers in ecommerce sites.
6. Lawrence, R. D., Almasi, G. S., Korlyar, V., Viveros, M. S., & Duri, S. S. (2001).Personalization of supermarket product recommendations. Data Mining and Knowledge Discovery, 5(1), 11–32.
7. Changchien, S. W., Lee, C.-F., & Hsu, Y.-J. (2004). On-line personalized sales promotion in electronic commerce. Expert Systems with Applications, 27(1), 35–52
8. Cornell, J. A. (2001). Experiments with mixtures: Designs, models and the analysis of mixture data. A Wiley-Interscience Publication
9. Forsati, R., & Meybodi, M. R. (2010). Effective page recommendation algorithms based on distributed learning automata and weighted association rules. Expert Systems with Applications, 37(2), 1316–1330.
10. Goksedef, M., & Gunduz-Oguducu, S. (2010). Combination of web page recommender systems. Expert Systems with Applications, 37(4), 2911–2922.
11. Goldberg, K., Roeder, T., Gupta, D., & Perkins, C. (2001). Eigentaste: A constant time collaborative filtering algorithm. Information Retrieval Journal, 4(2), 133–151.
12. Hill, W., Stead, L., Rosenstein, M., & Furnas, G. (1995). Recommending and evaluating choices in a virtual community of use. In Proceedings of the 1995 ACM conference on factors in computing systems (pp. 194–201). New York.
13. Lawrence, R. D., Almasi, G. S., Korlyar, V., Viveros, M. S., & Duri, S. S. (2001). Personalization of supermarket product recommendations. Data Mining and Knowledge Discovery, 5(1), 11–32.
14. Leung, C. W.-k., Chan, S. C.-f., & Chung, F.-l. (2008). An empirical study of a crosslevel association rule mining approach to cold-start commendations.Knowledge-Based Systems, 21(7), 515–529
15. Atlanta, GA. Montgomery, D. C. (2000). Design and analysis of experiments. New York: Wiley. Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., & Riedle, J. (1994). Grouplens: An open architecture for collaborative filtering of netnews. In Proceedings of the ACM 1994 conference on computer supported cooperative work (pp. 175–186). Chapel Hill.
16. Russel, S., & Yoon, V. (2008). Applications of wavelet data reduction in a commender system. Expert Systems with Applications, 34(4), 2316–2325.
17. Sarwar, B., Karypis, G., Konstan J.A., & Riedl, J. (2000a). Analysis of ecommendation algorithms for e-commerce. In Proceedings of ACM E-commerce 2000 conference (pp. 158–167). Minneapolis, MN.
18. Sarwar, B.M., Karypis, G., Konstan, J.A., & Riedl, J. (2000b). Application of dimensionality reduction in recommender system – A case study. In Proceedings of ACM WebKDD workshop. Boston, MA.

AUTHOR(S) PROFILE

Amol Sharadchandra Chaudhari, B.E (Computer Engineering) and pursuing M.Tech (Information Technology). from R.K.D.F.Institute of Technology ,Bhopal

