TEST Engineering & Management

January-February 2020
ISSN: 0193-4120Page No. 6778 - 6781

Predicting Missing Items in Shopping Carts using Machine Learning

[1]I.Vinay Kumar Reddy, [2]R. Senthil Kumar
[1]Student, [2]Assistant Professor
Department of Computer Science and Engineering, Saveetha School of Engineering,
Saveetha Institute of Medical and Technical Sciences, Chennai
[1]vinay.ivinu1999@gmail.com, [2]rsenthilmecse@gmail.com[2]

Article Info
Volume 82
Page Number: 6778 - 6781
Publication Issue:
January-February 2020

Article History
Article Received*: 18 May 2019*
Revised*: 14 July 2019*
Accepted*: 22 December 2019*
Publication*: 01 February 2020*

Abstract
With the fast growth of e-commerce, large number of products is sold online, and a lot more people are purchasing products online. Users who are purchasing products online depict a similar kind of purchasing pattern. And these patterns are being repeated for very next transaction by those users. So we are analyzing the pattern of these peoples on the basis of their previous transactional data and we are trying to predict the product for respective customer in their next transaction. Analyzing the patterns will help grocery stores to maintain their stock as well as to notify the customer about the offers of the product which he tends to buy for his next purchase. And we will also analyze the rush hours and happening days of the customers so this will helps the respective grocery store websites to maintain their server traffic. In this project we are collecting purchased transactional data of users from INSTACART, which is a grocery ordering and delivery application.

Keywords: *E-commerce, patterns, INSTACART, Grocery.*

1. Introduction

Recommendation system is one of the major components of e-commerce, allowing the e-commerce company to provide individual users with personalized service, maximize order size by suggesting accessories at checkout, and improve customer loyalty and engagement. Recommendation system has become a key component of modern e-commerce. Latest research on suggestion systems has focused primarily on enhancing the relevance or profitability of the specific items suggested. But users are typically exposed to a collection of items in reality, and they can buy multiple items in a single order. Thus, the importance or profitability of one item may depend on the other items in the set. The collection of suggestions, in many other words, is a package of items iterating with one another. In today's world every transaction made by user during online shopping is being stored and processed to predict the nature of user and to predict what are the products in which the user is interested. These transaction data is used as a tool to predict the products which a user is more likely to purchase, or what is the peak time for the order, these details are being saved to maintain the server traffic.

2. Literature Survey

Title: Predicting Missing Items in Shopping Carts using Fast Algorithm
Author: Srivatsan. M, Sunil Kumar. M, Vijayshankar. V, Leela Rani P.
Year: 2011.
Description:
Prediction in shopping cart uses partial data about the contents of a shopping cart for the forecasting what else the customer would likely buy.

To reduce the cost of rule mining, a quick algorithm is proposed that generates frequent item sets without creating candidate item sets. The algorithm utilizes Boolean vector with relational AND operation to find

January-February 2020
ISSN: 0193-4120Page No. 6778 - 6781

frequent item sets and create the rule of association. Association rules are utilized to identify relations between a database set of items. Boolean Matrix initially is generated by converting the database into Boolean values. The frequent item sets are created from the matrix in Boolean. Association rules are then to be created from the frequent item sets previously created. The rules created by the association form the foundation for forecasting. The incoming item set i.e. the incoming shopping cart content will also be depicted by a Boolean vector and then each transaction vector will perform AND operation to create the association rules. Finally, the rules for getting the predictions are mixed. Dempster‟s rule of combination (DRC) is utilized to integrate the proofs. Lastly we suggest to the user the anticipated items.

Title: Predicting Missing Items in Shopping Cart Using Associative Classification Mining
Author: K. Jagadish Kumar, Sahini Sairam
Year: 2013.
Description:

The primary task of association rule mining is to identify regularly happen groupings of items in transactional databases. The aim is to use this knowledge for reasons of anticipation. Several studies has primarily concentrated on how to expedite the search for regularly co-occurring clusters of items in "shopping cart" and much less concentration has been paid to the techniques used for predictive purposes to use these "frequent item sets."This paper relates to the latter task by introducing a methodology that utilizes the incomplete data about the contents of a shopping cart to predict what else the client is likely to purchased, for instance, if bread, butter, and milk always occur in the similar item, then the existence of butter and milk in a shopping cart indicates that the client can also purchase bread. More commonly knowing which items a shopping cart holds, we need to also anticipate items that are likely to be added by the customer before continuing to checkouts. So this paper introduces a method called the "Combo Matrix" whose main diagonal elements reflect the relationship between items and looking at the main diagonal elements, the consumer can choose what else the different items can be bought with the shopping cart's current content and also lowers the rule mining cost. The association between items is display through Graph. The frequent item sets are created from the Combo Matrix. The association rules are then to be created from the frequent item sets previously created. The association rules created form the foundation for prediction. The incoming item sets i.e. the shopping cart contents will be depicted by a set of specific indexed numbers and the combination between items will be created via the Combo Matrix. The anticipated items are finally recommended to the Client.

Title: Predicting Missing Items in Shopping Carts
Author: Kasun Wickramaratna, Miroslav Kubat, Kamal Premaratne
Year: 2009.
Description:
Current study in association mining concentrate primarily on how to improve the search for regularly co-occurring clusters of items in the "shopping cart" kind of transactions; lesser attention was paid to techniques that use these "frequent item sets" for predictive objectives. This paper relates to the latest goal by introducing a methodology that utilizes incomplete data about a shopping cart's contents to predict what else the client is likely to buy. Utilizing the newly suggested software framework of item set trees (IT-trees), we achieve from the partial shopping cart, in an effective mathematical way, all rules whose antecedents includes at least one item. Instead, we integrate these principles by analyzing techniques of ambiguity, including the classical Bayesian decision theory and a recent algorithm focused on the mixture of proof Dempster-Shafer (DS).

Title: Market Basket Analysis with Data Mining Methods
Author: Andrej Trnka
Year: 2010
Description:

This paper described the way Six Sigma method is implemented for Market Basket Analysis. Data mining techniques offer a lot of market prospects. One of them is analysis of the basket market. Methods Six Sigma uses various mathematical methods. With Market Basket Analysis (as part of Data Mining) being applied in Six Sigma (to one of its phases), we will enhance the results and modify the process's Sigma performance level. We have used the GRI (General Rule Induction) algorithm in our study to generate rules of association between products in the market basket. Such associations imply various among the products. To demonstrate the dependency we used a Web plot between the products. For analysis the final algorithm was CS.O. This algorithm was utilized for creating profiles depend on rules.

January-February 2020
ISSN: 0193-4120Page No. 6778 - 6781

3. Existing System

The existing system utilizes flagged item set trees for purposes of rule creation. An item set tree, T, consists of a root and a (possibly empty) set, {T1; . . . ;Tk}, each element of which is an item set tree. The root is a pair [s, f(s)], where s is an item set and f(s) is a frequency. If si denotes the item set associated with the root of the ith subtree, then s is a subset of si; s not equal to si, must be satisfied for all i. Twice the number of transactions in the initial database is upper-bounded by the number of nodes in the IT-tree. Remember that some of the item sets in the IT-tree [4] are similar to at least one of the transactions found in the initial database, while others were generated during the tree construction process where they came into being as common ancestors of lowest level transactions. They altered the initial algorithm for constructing tree by flagging every node that is similar with at least one transaction. These are shown by black dots. This is known as IT-tree flagged [4].

The drawbacks of current methods are

- Compared to Boolean matrix form, the time taken to build IT-Tree [4] is much more.
- This method demands more handling memory.

4. Proposed System

Proposed work should be possible by future work of existing framework. Machine learning and Data Mining are creating at a quick pace with a few new strategies being created and old systems being adjusted to upgrade execution, remembering this our work can be extended to join new techniques for order for result expectation and more highlights could be included alongside the ones as of now considered. Data mining techniques which are used in this paper are naïve bayes and Gradient Boosting. This paper adds to the last assignment by suggesting a method that utilizations incomplete data on the substance of a shopping cart for the forecast of what else the client is probably going to purchase for example foreseeing the missing items in shopping cart.

5. System Architecture

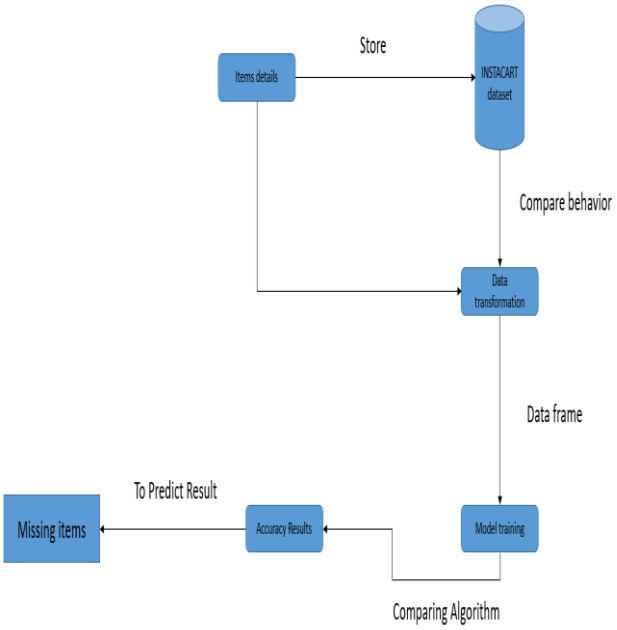

Figure 1: System Architecture

System configuration is the sensible model that portrays the structure, direct, and more points of view on a system. [13] A designing delineation is a customary depiction and depiction of a system, dealt with to such an extent that supports pondering the structures and practices of the structure. A system configuration shown in fig 1 can contain structure fragments and the sub-systems developed, that will coordinate to execute the general system. There have been attempts to formalize lingos to delineate structure plan; all things considered these are called designing depiction vernaculars (ADL).

6. Conclusion

Recommendation analysis deals with the classification of products based on their previous order transaction.
From analysis, we conclude that

1) Fresh fruits and Fresh vegetables are the most reordered product.
2) Dairy departments have most re ordered factor.
3) Sales of products is higher on Saturday.
4) Reorder ratio of Tuesday is lowest.
5) Produce departments have largest number of products.
6) Most of the items purchased between 9am to 6 pm.
7) Products that are added to the cart initially are more likely to be reordered again compared to the ones added later.

January-February 2020
ISSN: 0193-4120Page No. 6778 - 6781

7. Result

Here proposed Naïve Bayes by calculating the posterior probability and Gradient Boost, which takes the following outcomes as info. Since input is about user transaction data that means labels are assigned to each order id we use supervised learning.

References

[1] Kasun Wickramaratna, Miroslav Kubat and Kamal Premaratne, "Predicting Missing Items in Shopping Carts", IEEE Trans. Knowledge and Data Eng., vol. 21, no. 7, July 2009.

[2] M. Anandhavalli, Sandip Jain, Abhirup Chakraborti, Nayanjyoti Roy and M.K. Ghose "Mining Association Rules Using Fast Algorithm", Advance Computing Conference (IACC), 2010 IEEE 2nd International.

[3] H.H. Aly, A.A. Amr, and Y. Taha, "Fast Mining of Association Rules in Large-Scale Problems," Proc. IEEE Symp. Computers and Comm. (ISCC "01), pp. 107-113, 2001.

[4] R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules," Proc. Int"l Conf. Very Large Databases (VLDB "94), pp.487-499, 1994.

[5] K.K.R.G.K. Hewawasam, K. Premaratne, and M.-L. Shyu, "Rule Mining and Classification in a Situation Assessment Application: A Belief Theoretic Approach for Handling Data Imperfections," IEEE Trans. Systems, Man, Cybernetics, B, vol. 37, no. 6 pp. 1446-1459, Dec. 2007.

[6] Apriori Algorithm Reference URL: http://www2.cs.uregina.ca/~dbd/cs831/notes/item sets/items et_prog1.html

[7] P. Bollmann-Sdorra, A. Hafez, and V.V. Raghavan, "A Theoretical Framework for Association Mining Based on the Boolean Retrieval Model," Data Warehousing and Knowledge Discovery: Proc. Third Int"l Conf. (DaWaK "01), pp. 21-30, Sept. 2001.

[8] W. Li, J. Han, and J. Pei, "CMAR: Accurate and Efficient Classification Based on Multiple Class-Association Rules," Proc. IEEE Int"l Conf. Data Mining (ICDM "01), pp. 369376, Nov./Dec. 2001.

[9] M. Kubat, A. Hafez, V.V. Raghavan, J.R. Lekkala, and W.K. Chen, "Item set Trees for Targeted Association Querying," IEEE Trans. Knowledge and Data Eng., vol. 15, no. 6, pp. 1522-1534, Nov./Dec.2003.

[10] Agrawal. R, Imielinski. T and Swami. A, "Mining Association Rules between Sets of Items in Large Databases," Proc. ACM Special Interest Group on Management of Data (ACM SIGMOD), pp. 207-216, 1993.

[11] C. C Aggarwal, C. Procopius, and P. S. Yu, "Finding Localized Associations in Market Basket Data," IEEE Transaction on Knowledge and Data": Eng. , vol. 14, no. 1, pp. 51-62, Jan. /Feb. 2002.

[12] R. Bayardo and R. Agrawal, "Mining the Most Interesting Rules," Proc. ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, pp. 145-154, 1999.

[13] Liu. B, Hsu, Y. M. WAND Ma,: "Integrating Classification and Association Rule Mining," Proc. ACM SIGKDD Int'l Conf. Know. Disc. Data. Mining (KDD '98), pp. 80-86, Aug. 1998.

[14] Li. W, Han. J, and Pei. J, "CMAR: Accurate and Efficient Classification Based on Multiple Class Association Rules," Proc. IEEE Int'l Conf. Data Mining (ICDM '01), pp. 369-376, Nov. /Dec. 2001.

[15] Yen S. J. and Chen A "An efficient approach to discovering knowledge from large database". In proc. Of the IEEE/ACM International Conference on parallel distributed Information system, Pages 8-18, 1996.

[16] Lee K. L, G Lee and Chen A. L. P. "Efficient Graph based Algorithm for discovering and maintaining association rules in large database".

.