Journal of Computer Science, 9 (4): 433-438, 2013
ISSN 1549-3636

doi:10.3844/jcssp.2013.433.438 Published Online 9 (4) 2013 (http://www.thescipub.com/jcs.toc)

Intelligent Apriori Algorithm for Complex Activity Mining in Supermarket Applications

[1]V. Ganesh Kumar and [2]K. Muneeswaran

[1]Sree Sowdambika College of Engineering, Aruppukottai, Tamil Nadu, India
[2]Mepco Schlenk Engineering College, Sivakasi, Tamil Nadu, India

Received 2012-07-03, Revised 2013-03-11; Accepted 2013-05-09

ABSTRACT

As shopping becomes a shared experience and joint process with friends or family members nowadays, the most important problems arise with variety of products and the product information available in the supermarkets. This study proposes a system that uses Intelligent Apriori algorithm to support consumers in getting the required items from various supermarkets. Also this work intelligently suggests the best movement and reducing unwanted movement of the customer and quickly finds out the next operation which includes the next supermarket which is visited by the customer for the next item he/she purchases. This approach can further be extended to the world of mobile communication where the next movement of the mobile user can be predicted and used intelligently to arrange necessary requirements at the destination before he actually reaches. The feasibility of this approach is tested under simple conditions and the results are presented in this study.

Keywords: Data Mining, Apriori Algorithm, Activity Mining

1. INTRODUCTION

In the current world, consumers are often overwhelmed by the huge amount of in-store promotions, special offers and products on display. It is not an easier activity to decide, where to buy the required items with smaller movement? Varieties of choices allow better satisfaction of the individual needs. On the other hand variety can also be confusing and deterrent that can turn shopping into an endless decision process (Iyengar, 2010). In this study, we use the Apriori algorithm to quickly find out the next operation which includes the next supermarket the customer visits and next item which he/she purchases. Additionally, this study suggests, the supermarkets with minimum distance for the consumer to purchase the next interested item instead of sticking on the Apriori algorithm in turn reducing the distance to be travelled. This approach can further be extended to the world of mobile communications where the next movement of the mobile user can be predicted and used intelligently to arrange necessary requirements at the destination. Consider a set of super markets (S_1, S_2...S_p), a set of items (I_1, I_2...I_q) and a set of users (U_1, U_2...U_r) distributed across the different locations as shown in the **Fig. 1**. In this study the movement and purchase patterns of each user is studied and appropriate decision making is enabled.

1.1. Activity Mining Techniques

Data mining is an analytic process designed to explore data (usually large amount of data-typically business or market related) in search of consistent patterns and/or systematic relationships between variables and then to validate the findings by applying the detected patterns to new subsets of data. The process of data mining consists of three stages: (i) The initial exploration of data (ii) Model building or pattern identification with validation/verification and (iii) Deployment of the application of model to new data to generate predictions (Neel, 2011).

Corresponding Author: V. Ganesh Kumar, Sree Sowdambika College of Engineering, Aruppukottai, Tamil Nadu, India
Tel: 91-9443664026

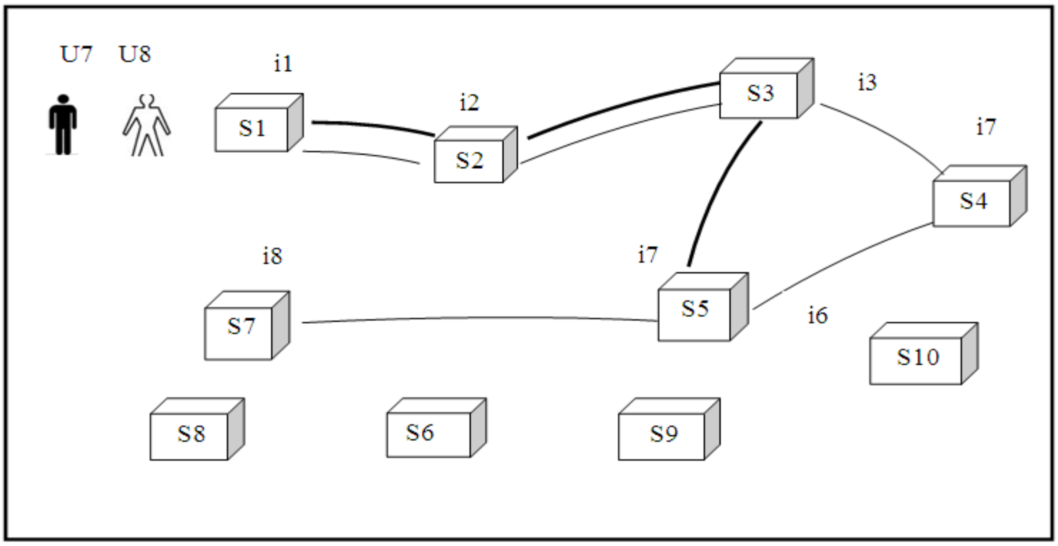

Fig. 1. User behaviour patterns in supermarkets

In the following are some of the data mining techniques and their uses: In data mining, a Decision Tree is a predictive model which can be used to represent both classifiers and regression models. Decision tree can also be used to estimate the value of continuous variable, although there are many techniques suitable to that task (Linoff and Berry, 2011). Neural Network methods are commonly used for data mining tasks, because they often produce comprehensible models. A neural network is a computational technique that benefits from techniques similar to ones employed in the human brain. The promise of neural networks lies in their ability to learn patterns in a complex signal (Sasaki *et al*., 2010). Clustering is a tool for data analysis, which solves classification problems. Its objective is to distribute cases (people, objects, events) into groups, so that the degree of association can be strong between members of the same cluster and weak between members of different clusters. Clustering is often done as a prelude to some other form of data mining or modelling (Linoff and Berry, 2011). K-means clustering is an example for clustering method. In data mining, Association Rule (Padmaja and Poongodai, 2011) is a popular and well researched method for discovering interesting relations between variables in large databases. Piatetsky-Shapiro describes analyzing and presenting strong rules discovered in databases using different measures of interests. Based on the concept of strong rules, Kamsu-Foguem *et al*. (2012) introduced association rules for discovering regularities between products in large scale transaction data recorded by Point-of-Sale (POS) systems in supermarkets. Factor Analysis is an essential step in effective clustering and classification procedures. There are several developed approaches for factor analysis. The recent development, Genetic Algorithms (GAs) have been very useful in finding optimal solutions because a GA can search a large space with comparatively less computation time (Sasaki *et al*., 2010). In statistics, signal processing and many other fields, a time series is a sequence of data points, measured typically at successive times, spaced at (often uniform) time intervals.

1.2. Proposed Approach

In our work, we analyze the user behaviour patterns in a supermarket purchase and formally characterize the idea of complex activities. We then argue that identifying user activities holds the key toward effective data management in any environment. Normally users travel from one place to another to purchase different items ShopSavvy Blog. Based on the activities analysed, we can characterize the basic user behaviour patterns into three categories:

- Moving to supermarket only and no items purchase patterns (S-type): Sequences of supermarkets that are repeatedly visited by users
- Item Purchase in the current supermarket patterns (I-type): Sequences of products that are repeatedly purchased by users

- Moving to another supermarket and purchasing item patterns (SI-type): Sequences of supermarkets-products pairs that are repeatedly visited and purchased by users

1.3. Intelligent Apriori Algorithm

Algorithm 1 : Intelligent Apriori Algorithm
Procedure IntelligentApriori (Activitydatabase Db)
Initialize $A_1 = Ø$; $SI_1 = Ø$; // A-Activity, SI-Supermarket name and Item Purchased Set
for each row t in Db do
 for each activity a in t do
  if a.supermarket ^ a.item ≠ null then
 add activity (a.supermarket, a.item), to SI_1;
  if a:supermarket ≠ null then
 add activity (a.supermarket) to A_1;
  if a.item ≠ null then
  add activity (a.service) to A_1;
 Increment the count of element *a* in SI_1 or A_1;
 end
end
Remove elements of count < support in SI_1;
Reduce the count for duplicate entries in A_1;
Remove elements of count < support in A_1;
$A_1 = A_1$ U SI_1;
for (k = 2; Ak_1 ≠ 0; k++) do
  Initialize Ck = Ø;
  Combine all the entries with $item_{1 \text{ to } k-2}$ are equal and $item_{k-1}$ is different in to C_k;
  for every subset s of c in C_k
  Delete c from C_k if s is in A_{k-1};
  Find the rows in Db that contains the activity c of C_k and Increment count;
  Remove element of count < support in c;
  Ak = A_k U C_k;
end
$A_k = U_k$ Ak;
n = | a.supermarket |; // Count no. of unique in D
for (i = 1; i < n; i++) do
  for (j = 1; j < n; j++) do
   d(i,j) = min_dist(a_i,a_j);
  end
end
for (i = 2; i < n; i++) do
  a = A_k(i-1); b = A_k(i);
  p = find_supermarket(b.item);// find out the supermarket with product b.item
  if(min_dist(a.supermarket, b. supermarket)> min_dist(a.supermarket, p.supermarket)
   then
    Choose alternative supermarket (p) instead of supermarket (b) as the low distance.
end

Our algorithm for activity mining is based on the popular Apriori algorithm to identify all primitive and complex activities from a database of user behaviour logs. An action is a (Supermarket, Item) pair to denote that a user purchases an item in a particular supermarket. When the purchase of a product is made in the same supermarket, a list of items associated with that supermarket is maintained. When the item is null, it is a simple visit of the supermarket without purchasing any product. A pair of null values is not considered as a meaningful action. A behaviour transaction is a sequence of actions taken by a user. A behaviour database is a set of transactions recorded for the set of users in the area of interest.

Apriori algorithm shown in (Wu and Fan, 2010) is extended with special emphasis on activity mining. The function works by enumerating all SI-, S- and I-type activities first. Then, these SI-type activities with not enough support are removed. Since each remaining SI-type activity also constitutes an S-type and an I-type activity, we need to deduce the count from the corresponding S-, I-type activities. Finally, all S- and I-type activities with enough support are joined with SI-type activities to form the activity set. The rest of the activity mining is essentially a direct adaptation of the Apriori algorithm to the mining of complex activities. Our proposed Intelligent Apriori Algorithm shown in Algorithm1 is the modified version of Apriori algorithm used by Wu and Fan (2010) that can suggest the best movement (supermarket) for the consumer to purchase the next interested item instead of sticking on the Apriori algorithm in turns reducing the distance to be travelled.

1.4. Experimental Results

For performance evaluation, we have different types of structures with different numbers of supermarkets and products. Databases with different number of transactions consisting of various combinations of supermarkets, items are created. Then our intelligent Apriori algorithm is applied on the database and the following operations are performed:

- Identifying the possible combinations of movement of user at next level and the count of occurrences of those patterns is identified. This leads to the generation of the structure A_i
- As the consideration of patterns with lower number of occurrences leads to confusion and complexity of the work, all those patterns with count less than support are removed. Only the remaining activities are added to the A_i structure

Table 1. Input to all scenarios of our work

No of Supermarkets	Name of the Supermarket	No of Products available in the Supermarket	Products ID
10	S_1	15	i_1 i_2 i_3 i_4 i_5 i_6 i_7 i_8 i_9 i_{10} i_{11} i_{12} i_{13} i_{14} i_{15}
	S_2	10	i_1 i_2 i_3 i_4 i_5 i_6 i_7 i_8 i_9 i_{10}
	S_3	6	i_1 i_2 i_3 i_4 i_5 i_6
	S_4	10	i_1 i_2 i_3 i_4 i_5 i_6 i_7 i_8 i_9 i_{10}
	S_5	8	i_1 i_2 i_3 i_4 i_5 i_6 i_7 i_8
	S_6	8	i_1 i_2 i_3 i_4 i_5 i_6 i_7 i_8
	S_7	13	i_1 i_2 i_3 i_4 i_5 i_6 i_7 i_8 i_9 i_{10} i_{11} i_{12} i_{13}
	S_8	4	i_1 i_2 i_3 i_4
	S_9	6	i_1 i_2 i_3 i_4 i_5 i_6
	S_{10}	10	i_1 i_2 i_3 i_4 i_5 i_6 i_7 i_8 i_9 i_{10}

Table 2. Distance between Supermarkets

DistMat	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
S_1	0	4	6	3	7	3	5	8	3	5
S_2	4	0	5	4	7	3	8	5	7	2
S_3	6	5	0	3	6	9	2	4	5	2
S_4	3	4	3	0	12	7	3	2	5	1
S_5	7	7	6	12	0	6	3	4	2	2
S_6	3	3	9	7	6	0	5	4	2	3
S_7	5	8	2	3	3	5	0	8	3	4
S_8	8	5	4	2	4	4	8	0	2	6
S_9	3	7	5	5	2	2	3	2	0	5
S_{10}	5	2	2	1	2	3	4	6	5	0

- The standard Apriori algorithm is then applied to extract the possible movements of the user
- To add the intelligence to the standard Apriori algorithm, the following steps are used:
  - The minimum distances between different supermarkets are found out
  - Based on the each step movement of user for purchasing the items, identify next item to be purchased and find out the distance between them
  - Find out the alternative location of the supermarket for purchasing the next product to be purchased by the user and calculate the distance for the alternate supermarket
  - Compare the distance calculated in step b and c
  - If c is less than b, then suggest the user to get the product in the alternate supermarket as it can be reachable earlier

Experiments have been conducted with different scenarios and the results obtained with some of them are given below. **Table 1 and 2** depict the input data to the activity mining process and distance between supermarkets. Letter S stands for supermarket and i represents item. For example, S_2 i_2 denotes the purchase of item i_2 in supermarket S_2. From the example, we can observe that the algorithm can successfully identify all primitive and complex activities. **Table 3** shows the distance between each super market. We have generated the database (inp8) with 6 transactions as follows:

Transaction 1: S_1 i_1, S_2 i_2, S_3 i_3, S_4 i_7, S_5 i_6, S_7 i_8
Transaction 2: S_1 i_1, S_2 i_2, S_3 i_3, S_4 i_4
Transaction 3: S_1 i_1, S_2 i_2, S_3 i_3, S_4 i_4
Transaction 4: S_1 i_1, S_4 i_4, S_6 i_5, S_7 i_3, S_8 i_2
Transaction 5: S_2 i_2, S_4 i_4, S_5 i_5
Transaction 6: S_1 i_1, S_2 i_2

Apriori algorithm is applied on this database and the user's next movements are predicted as: S_1i_1, S_2i_2, S_3i_3 and S_4i_4. From the result one can understand that, the user (U_8) started from the supermarket S1 and get the product i1 and Visits S2 for getting product i2. Then he visits S_3 and S_4 for getting i_3 and i_4 products. Here, our Intelligent Apriori algorithm suggests the user to get the products i_2, i_3 and i_4 from the supermarkets S_1, S_2 and S_3 respectively to reduce the travel distance of S_4 for getting the product i_4.

Table 3. Output of the Apriori and our proposed intelligent Apriori algorithms

User	Trans actions	Products purchased Name of the Supermarket and Product	A1		A2		A3		Apriori output	Intelligent Apriori output
			Data	Count	Data	Count	Data	Count		
U_1 (inp2)	T_1	$S_1 i_1, S_7 i_2, S_5 i_3, S_6 i_5$	S_7i_0	2	i_3, S_6i_0	2	S_1i_1, i_3, S_6i_0	2	$S_1i_1, i_3, S_6i_0, S_1i_1, S_2i_2, i_3$	$S_1i_1, i_3, S_6i_0, S_1i_1, S_1i_2, i_3$
	T_2	$S_1 i_1, S_2 i_2, S_4 i_3, S_6 i_4$	S_5i_0	2	S_1i_1, S_7i_0	2	S_1i_1, S_2i_2, i_3	2		
	T_3	$S_1 i_1, S_2 i_2, S_7i_3$	i_3	3	S_1i_1, S_5i_0	2				
	T_4	$S_1 i_1, S_5 i_5$	S_6i_0	2	S_1i_1, i_3	3				
			i_5	2	$S_1i_1, S_6 i_0$	2				
			S_1i_1	4	S_1i_1, i_5	2				
			S_2i_2	2	S_1i_1, S_2i_2	2				
			S_2i_2, i_3	2						
U_5 (inp5)	T_1	$S_1 i_1 S_2 i_2 S_3 i_3 S_5 i_4$	i_3	3	i_3, S_5i_0	2	S_1i_1,i_3, S_5i_0	2	$S_1i_1,S_2i_2, i_3, S_5i_0$	$S_1i_1, S_1i_2, i_3, S_5i_0$
	T_2	$S_1 i S_2 i_2 S_5 i_3 S_5 i_5$	S_5i_0	2	S_1i_1, i_3	3	S_1i_1, S_2i_2,i_3	3		
	T_3	$S_1 i_1 S_2 i_2 S_7 i_3$	S_1i_1	4	S_1i_1, S_5i_0	3	S_1i_1, S_2i_2, S_5i_0	2		
	T_4	$S_1 i_1 S_5 i_4$	S_2i_2	3	S_1i_1, S_2i_2	3	S_2i_2, i_3, S_5i_0	2		
			S_5i_4	2	S_1i_1, S_5i_4	2				
					S_2i_2, i_3	3				
					S_2i_2, S_5i_0	2				
U_7 (inp7)	T_1	$S_1 i_1, S_2 i_2, S_3 i_3, S_5 i_7$	i_3	3	i_3, S_5i_0	2	S_1i_1,i_3, S_5i_0	2	$S_1i_1, S_2i_2, i_3,S_5i_0$	$S_1i_1, S_1i_2, i_3, S_5i_0$
	T_2	$S_1 i_1, S_2i_2, S_5 i_3, S_5 i_5$	S_5i_0	4	S_1i_1, i_3	3	S_1i_1, S_2i_2, i_3	3		
	T_3	$S_1 i_1, S_2 i_2, S_7 i_3$	S_1i_1	4	S_1i_1, S_5i_0	3	S_1i_1, S_2i_2, S_5i_0	2		
	T_4	$S_1 i_1, S_5 i_8$	S_2i_2	3	S_1i_1, S_2i_2	3	S_2i_2, i_3, S_5i_0	2		
					S_2i_2, i_3	3				
					S_2i_2, S_5i_0	2				
U_8 (inp8)	T_1	$S_1 i_1, S_2 i_2, S_3 i_3, S_4 i_7, S_5 i_6, S_7i_8$	i_3	2	i_3, S_4i_4	2	S_1i_1, i_3, S_4i_4	2	$S_1i_1, S_2i_2, i_3, S_4i_4, S_1i_1, S_2i_2, S_3i_3, S_4i_4$	$S_1i_1, S_1i_2, i_3, S_2i_4, S_1i_1, S_1i_2, S_2i_3, S_3i_4$
	T_2	$S_1 i_1, S_2 i_2, S_3 i_3, S_4 i_4$	S_5i_0	2	S_1i_1, i_3	4	S_1i_1, S_2i_2, i_3	3		
	T_3	$S_1 i_1, S_2 i_2, S_3 i_3, S_4 i_4$	S_7i_0	2	S_1i_1, S_7i_0	2	S_1i_1, S_2i_2, S_3i_3	3		
	T_4	$S_1 i_1, S_4 i_4, S_6 i_5, S_7 i_3, S_8 i_2$	i_5	2	S_1i_1, S_2i_2	4	S_1i_1, S_2i_2, S_4i_4	2		
	T_5	$S_2 i_2, S_4i_4, S_5i_5$	S_1i_1	5	S_1i_1, S_3i_3	3	S_1i_1, S_3i_3, S_4i_4	2		
	T_6	$S_1 i_1, S_2i_2$	S_2i_2	5	S_1i_1, S_4i_4	3	$S_2i_2, i_3 S_4i_4$	2		
			S_3i_3	3	S_2i_2, i_3	3	S_2i_2, S_3i_3, S_4i_4	2		
			S_4i_4	4	S_2i_2, S_5i_0	2				
			S_2i_2, S_3i_3	3						
			S_2i_2, S_4i_4	3						
			S_3i_3, S_4i_4	2						
			S_4i_4, i_5	2						

2. CONCLUSION

In this work, a new Intelligent Apriori algorithm has been proposed and its performance has been tested in the supermarket environment. This study is an extension to work proposed by Wu and Fan (2010) where they have applied the Apriori algorithm which has been modified with special emphasis on activity mining. But in our study, we added more intelligent to the Apriori algorithm which can suggest the next movement based on the minimum distance, which can save a lot of movement in real-time environments.

This study can further be extended to the mobile communication environments, where the prediction of mobile user movement is an important task in helping the technology to make available the required resources at the destination. Moreover, the unnecessary movement of the user has also been avoided which in turn plays a major role in future communication technology.

3. REFERENCES

Iyengar, S., 2010. The Art of Choosing. 1st Edn., Grand Central Publishing, New York, ISBN-10: 0446558710, pp: 352.

Kamsu-Foguem, B., F. Rigal and F. Mauget, 2012. Mining association rules for the quality improvement of the production process. Expert Syst. Appli., 40: 1034-1045. DOI: 10.1016/j.eswa.2012.08.039

Linoff, G.S. and M.J. Berry, 2011. Data Mining Techniques for Marketing, Sales and Customer Relationship Management. 3rd Edn., John Wiley and Sons, Indianapolis, IN., ISBN-10: 1118087453, pp: 888.

Neel, B.M., 2011. Predictive data mining and discovering hidden values of data warehouse. ARPN J. Syst. Software.

Padmaja, V. and A. Poongodai, 2011. Mining weighted association rules. Int. J. Adv. Eng. Sci. Technol., 11: 153-153.

Sasaki, S., A.J. Comber, H. Suzuki and C. Brunsdon, 2010. Using genetic algorithms to optimise current and future health planning-the example of ambulance locations. Int. J. Health Geographics, 9: 1-10.

Wu, S.Y. and H.H. Fan, 2010. Activity-based proactive data management in mobile environments. IEEE Trans. Mobile Comput., 9: 390-404. DOI: 10.1109/TMC.2009.139