

# 1 A Survey on Data Mining Algorithm for Market Basket Analysis

2 Dr. M. Dhanabhakym<sup>1</sup>

3 <sup>1</sup> Bharathiar University

4 *Received: 13 April 2011 Accepted: 12 May 2011 Published: 26 May 2011*

---

## 6 Abstract

7 Association rule mining identifies the remarkable association or relationship between a large  
8 set of data items. With huge quantity of data constantly being obtained and stored in  
9 databases, several industries are becoming concerned in mining association rules from their  
10 databases. For example, the detection of interesting association relationships between large  
11 quantities of business transaction data can assist in catalog design, cross-marketing, lossleader  
12 analysis, and various business decision making processes. A typical example of association rule  
13 mining is market basket analysis. This method examines customer buying patterns by  
14 identifying associations among various items that customers place in their shopping baskets.  
15 The identification of such associations can assist retailers expand marketing strategies by  
16 gaining insight into which items are frequently purchased jointly by customers. It is helpful to  
17 examine the customer purchasing behavior and assists in increasing the sales and conserve  
18 inventory by focusing on the point of sale transaction data. This work acts as a broad area for  
19 the researchers to develop a better data mining algorithm. This paper presents a survey about  
20 the existing data mining algorithm for market basket analysis.

---

22 *Index terms*— Association Rule Mining, Apriori Algorithm, Market Basket Analysis

## 23 1 Introduction

24 he majority of the recognized organizations have accumulated masses of information from their customers for  
25 decades. With the e-commerce applications growing quickly, the organizations will have a vast quantity of data  
26 in months not in years. Data Mining, also called as Knowledge Discovery in Databases (KDD), is to determine  
27 trends, patterns, correlations, anomalies in these databases that can assist to create precise future decisions.

28 Mining Association Rules is one of the most important application fields of Data Mining. Provided a set of  
29 customer transactions on items, the main intention is to determine correlations among the sales of items. Mining  
30 association rules, also known as market basket analysis, is one of the application fields of Data Mining. Think a  
31 market with a gathering of large amount of customer transactions. An association rule is  $X \rightarrow Y$ , where X is referred  
32 as the antecedent and Y is referred as the consequent. X and Y are sets of items and the rule represents that  
33 customers who purchase X are probable to purchase Y with probability %c where c is known as the confidence.  
34 Such a rule may be: "Eighty percent of people who purchase cigarettes also purchase matches". Such rules assists  
35 to respond questions of the variety "What is Coca Cola sold with?" or if the users are intended in checking the  
36 dependency among two items A and B it is required to determine rules that have A in the consequent and B in  
37 the antecedent.

38 Figure1 shows a typical Market basket analysis. This is a perfect example for illustrating association rule  
39 mining. It is a fact that all the managers in any kind of shop or departmental stores would like to gain knowledge  
40 about the buying behavior of every customers. This market basket analysis system will help the managers to  
41 understand about the sets of items are customers likely to purchase. This analysis may be carried out on all  
42 the retail stores data of customer transactions. These results will guide them to plan marketing or advertising  
43 approach. For example, market basket analysis will also help managers to propose new way of arrangement in

44 store layouts. Based on this analysis, items that are regularly purchased together can be placed in close proximity  
45 with the purpose of further promote the sale of such items together. If consumers who purchase computers also  
46 likely to purchase anti-virus software at the same time, then placing the hardware display close to the software  
47 display will help to enhance the sales of both of these items.

48 Classification rule mining intends to determine a small set of regulations in the database that forms a perfect  
49 classifier. Association rule mining discovers all the rules offered in the database that assures some minimum  
50 support and minimum confidence constraint. In the case of association rule mining, the goal of discovery is not  
51 pre-determined, while for classification rule mining there is only one predetermined goal. This paper provides  
52 various existing data mining algorithms for market basket analysis.

## 53 2 Literature Survey

54 Zhixin et al., [1] recommended an improved classification technique based on predictive association rules.  
55 Classification dependent predictive association rules (CPAR) is a type of association classification methods which  
56 unites the benefits of associative classification and conventional rule-based classification. For generation of the  
57 rule, CPAR is highly effective when compared to the conventional rule-based classification because most of the  
58 repeated calculation is ignored and multiple literals can be chosen to produce multiple rules at the same time.  
59 Even though the benefit mentioned above avoids the repeated calculation in rule generation, the prediction  
60 processes have the disadvantage in class rule distribution inconsistency and interruption of inaccurate class rules.  
61 Further, it is ineffective to instances satisfying no rules. To ignore these difficulties, the author recommends Class  
62 Weighting Adjustment, Center Vector-based Preclassification and Post-processing with Support Vector Machine  
63 (SVM). Experimental observations on Chinese text categorization corpus TanCorp proves that this approach  
64 gains an average enhancement of 5.91% on F1 score compared with CPAR.

65 Qiang et al., [2] presented association classification based method on compactness of rules. Associative  
66 classification provides maximum classification accurateness and strong flexibility. On the other hand, this  
67 associative classification suffers from a difficulty of over fitting because the classification rules satisfied least  
68 support and lowest confidence are returned as strong association rules return to the classifier. In this paper,  
69 proposed an innovative association classification technique based on neatness of rules, it extends Apriori Algorithm  
70 which considers the interestingness, importance, overlapping relationships among rules. Experimental observation  
71 proves that the proposed approach has better classification accuracy in comparison with CBA and CMAR are  
72 highly intelligible.

73 Wang et al., [3] suggested a novel rule weighting approach in classification association rule mining.  
74 Classification association rule mining (CARM) is a latest classification rule mining technique that constructs  
75 an association rule mining based classifier by utilizing classification association rules (CARs). The specific  
76 CARM algorithm which is used is not considered, a similar set of CARs is constantly produced from data, and a  
77 classifier is generally presented as a structured CAR list, depending on a selected rule ordering approach. Many  
78 numbers of rules ordering approaches have been established in the recent past, which can be classified as rule  
79 weighting, support-confidence and hybrid. In this approach an alternative rule-weighting method, called CISRW  
80 (Class-Item Score based Rule Weighting) and build up a ruleweighting based rule ordering mechanism depending  
81 on CISRW. Later, two hybrid techniques are additional developed by merging (1) and CISRW. The simulation  
82 results indicates that the three proposed CISRW based/related rule ordering techniques do well by means of  
83 accuracy of classification.

84 Bartik [4] presented association based classification for relational data and its use in web mining. Classification  
85 according to the mining association rules is a technique with better accuracy and human understandable  
86 classification scheme. The intention of the author is to put forward a alteration of the fundamental association  
87 based classification technique that can be helpful in data gathering from Web pages. In this paper, the  
88 alteration of the technique and required discretization of numeric characteristics are provided. Sumithra et al., [5]  
89 proposed a distributed Apriori association rule and classical Apriori mining algorithms for grid based knowledge  
90 discovery. The intention of this paper is to obtain knowledge with the help of predictive apriori and distributed  
91 grid dependent apriori algorithms for association rule mining. The author provides the implementation of an  
92 association rules discovery data mining task with the help of Grid technologies. A consequence of implementation  
93 with a contrast of existing apriori and distributed apriori is also provided by the author. Distributed data mining  
94 systems offers an effective utilization of multiple processors and databases to accelerate the execution of data  
95 mining and facilitate data distribution. For evaluating the effectiveness of the described technique, performance  
96 investigation of apriori and predictive apriori techniques on a standard database have been provided using weka  
97 tool. The key intention of grid computing is to offer the organizations and application builders the capability  
98 to generate distributed computing environments that can make use of computing resources on demand. Hence,  
99 it can assist amplify the effectiveness and decrease the cost of computing networks by reducing the time for  
100 data processing and optimizing resources and distributing workloads, thus permitting users to attain much faster  
101 outcome on large operations and at lesser costs.

102 Trnka [6] uses Data Mining Methods for Market Basket Analysis. This paper provides the technique for  
103 Market Basket Analysis implementation to Six Sigma technique. Data mining techniques offers more prospects  
104 in the market sector. Basket Market Analysis is one among them. Six Sigma technique utilizes various statistical  
105 techniques. With execution of Market Basket Analysis to Six Sigma, the results can be enhanced and Sigma

---

106 performance level of the method can also be modified. The author used GRI (General Rule Induction) technique  
107 to construct association rules among products in the market basket. These associations provide a variety among  
108 the products. Web plot is utilized here for representing the dependence among the products.

109 Mining Interesting Rules by Association and Classification Algorithm is put forth by Yanthy et al., [7].

110 The main intention in data mining is to disclose hidden knowledge from data and several techniques have been  
111 suggested so far. But the demerit is that characteristically not all rules are interesting -only little portions of  
112 the created rules would be of interest to several provided user. Therefore, numerous measures like confidence,  
113 support, lift, information gain, etc., have been suggested to find the best or highly interesting rules. On the  
114 other hand, some techniques are good at creating rules high in one interestingness measure but not good in other  
115 interestingness measures. The relationship among the techniques and interestingness measures of the created  
116 rules is not clear until now. The author studied the relationship among the techniques and interesting measures.  
117 The author used synthetic data so that the outcome result is not restricted to particular situations.

118 Market Basket Analysis Based on Text Segmentation and Association Rule Mining is suggested by Xie et  
119 al., [8]. Market basket analysis is very useful in offering scientific decision support for trade market by mining  
120 association rules between items people purchased collectively. The author provides an innovative market basket  
121 analysis technique by mining association rules on the items' internal features that are obtained with the help of  
122 automatic words segmentation technique. This technique has been used for dynamic dishes recommend system  
123 and results better in the experimental results.

124 Chiu et al., [9] proposed a market-basket analysis with principal component. Market-basket analysis is a  
125 well-known business crisis that can be solved computationally with the help of association rules, mined from  
126 transaction data to reduce the crossselling results. The author model the market-basket analysis as a finite  
127 mixture density of human consumption activities based on social and cultural activities. This results in the usage  
128 of principle component analysis and perhaps mixture density analysis of transaction data that was not obvious  
129 previously. The author contrast PCA and association rules mined from a set of benchmark transaction data, to  
130 discover common and differences among these two data exploration tools.

### 131 3 Market Basket Analysis of Library Circulation

132 Data is provided by Cunningham et al., [10]. Market basket analysis technique have lately seen extensive usage  
133 in evaluating consumer purchasing patterns particularly, in identifying products that are often purchased. The  
134 author utilized the a-priori market basket tool to the process of detecting subject classification grouping that co-  
135 occur in transaction records of books borrowed from a university library. This data can be utilized in directing  
136 users to extra portions of the gathered that may consists of documents that are related to their information  
137 requirement, and in finding a library's physical layout. These results can also offer insight into the amount of  
138 scatter that the classification method provokes in a particular gathering of documents.

139 Zongyao et al., [11] that have possible association patterns and processes LMAS for every object in the  
140 reference objects for some interested spatial relation. Hence, the result of the algorithm is a LMAS distribution  
141 map that replicates association potential variations over the examination area. Spatial interpolation for LMAS  
142 is recommended to generate a continuous LMAS distribution that can be utilized to investigate hot spots that  
143 reveal strong association patterns. This proposed technique was applied in an ecological system research.

144 Mining association rules based on apriori algorithm and application is given by Pei et al., [12]. In the data  
145 mining research, mining association rules is an significant subject. Intended at two difficulties of discovering  
146 frequent itemsets in a large database and mining association rules from frequent itemsets, the author carries  
147 some analysis on mining frequent itemsets algorithm with the help of apriori algorithm and mining association  
148 rules algorithm with the help of enhanced measure system. Mining association rules technuque with the help  
149 of support, confidence and interestingness is enhanced, aiming at generating interestingness ineffective rules and  
150 losing helpful rules. Useless rules are cancelled, creating more reasonable association rules including negative  
151 items. The suggested technique is utilized to mine association rules to the 2002 student score list of computer  
152 dedicated field in Inner Mongolia university of science and technology.

153 Yong et al., [13] proposed a mining association rules with new measure criteria. In recent days, association  
154 rules mining from bulk databases is an active research field of data mining motivated by many application areas.  
155 But, there are some difficulties in the strong association rules mining depending on supportconfidence framework.  
156 Initially, there are a huge number of redundant association rules are created, then it is complicated for user to  
157 discover the interesting ones. Then, the correlation among the features of specified application areas is avoided.  
158 Therefore innovative measure criteria called Chi-Squared test and cover should be introduced to association  
159 rules mining, and the more important aspect is the use of Chi-Squared test to reduce the amount of rules. The  
160 Chi-Squared test and cover of measures are utilized by author for association rules mining for the purpose of  
161 eliminating the itemsets that are statistic free, while frequent itemsets or rules are created. Therefore the number  
162 of patterns itemsets reduced and it is effortless for user to gather the highly noticeable association rules. The  
163 simulation results suggest that the Chi-Squared test is efficient on decreasing the quantity of patterns through  
164 merging support and cover constrain. Pattern choosing according to Chi-Squared test can remove few irrelevant  
165 attributes and the efficiency and veracity of mining association rules are enhanced.

166 Mining traditional association rules using frequent itemsets lattice is given by Vo et al., [14].

167 Numerous methods have been formulated for the enhancement of time in mining frequent itemsets. However,

### 3 MARKET BASKET ANALYSIS OF LIBRARY CIRCULATION

---

168 the methods which deal with the time of mining association rules were not put in deep research. In reality,  
169 in case of database which contains many frequent itemsets (from ten thousands up to millions), the time of  
170 mining association rules is much larger than that needed for mining frequent itemsets. In this paper, developed  
171 an application of lattice in mining conventional association rules which will significantly decrease the time for  
172 mining rules. This technique comprises of two stages: (a) construction of frequent itemsets lattice and (b) mining  
173 association rules from lattice. For the quick determination of association rules, the parent-child relationships in  
174 lattice is used. The experiments observation proves that the mining rule from lattice is more efficient than the  
175 straight mining from frequent itemsets by means of hash table.

176 Rastogi et al., [15] presents mining optimized association rules with categorical and numeric attributes. Mining  
177 association rules on bulky data sets has gained significant attention recently. Association rules are helpful for  
178 predicting correlations among the features of a relation and contain applications in marketing and many retail  
179 sectors. Moreover, optimized association rules are an efficient approach to focus on the most interesting features  
180 linking certain attributes. Optimized association rules are allowed to include uninstantiated attributes and the  
181 difficulty is to find out instantiations such a way that either the support or confidence of the rule is maximized.  
182 In this approach, the optimized association rules difficulty is simplified in three ways: (a) association rules  
183 are permitted to include disjunctions in excess of uninstantiated features, (b) association rules are allowed to  
184 contain a random number of uninstantiated features, and (c) uninstantiated features can be either categorical  
185 or numeric. This generalized association rules permits to mine more helpful information about seasonal and  
186 local patterns linking multiple features. This paper also suggests an efficient method for pruning the search space  
187 when calculating optimized association rules for both categorical and numeric features. Experimental result shows  
188 that pruning techniques are effective for a huge number of uninstantiated features, disjunctions, and values in  
189 the domain of the features. Wang et al., [16] performs an investigation on association rules mining based-on  
190 ontology in ecommerce. Commercial actions carried out with the use of Internet turn out to be more and more  
191 popular. And plenty of transaction logs are created, which helps to gather useful information by data mining. In  
192 this manner, association rule mining is very important in ecommerce. However there are various problems occur  
in the existing association rules mining systems. The <sup>1 2 3 4 5 6</sup>



Figure 1: T

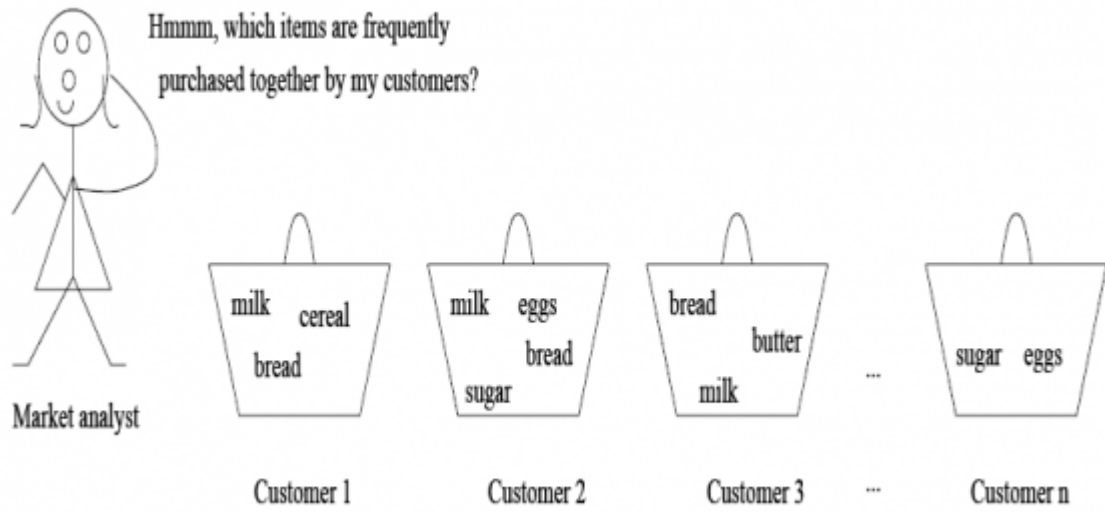


Figure 2:

<sup>1</sup>July

<sup>2</sup>© 2011 Global Journals Inc. (US) July © 2011 Global Journals Inc. (US)

<sup>3</sup>July

<sup>4</sup>© 2011 Global Journals Inc. (US) July © 2011 Global Journals Inc. (US)

<sup>5</sup>? ? ? July

<sup>6</sup>© 2011 Global Journals Inc. (US) July



193 existing conventional techniques can't solve these problems very well. With the intention of solving these  
 194 difficulties better, this paper proposes association rules mining depending on ontology. Generally researches  
 195 the specified three parts during data mining: (1) methods of ontology creation and principles of commodity  
 196 categorization; (2) simplifying R-interesting based on actual situations; (3) implementing association rules  
 197 mining depending on ontology by improved Apriori. Additionally, this paper tests the enhanced algorithm  
 198 using FoodMart2000, Java as the development language and Jena as the ontology engine, completes the entire  
 199 process of mining, and confirms the validity of the algorithm by the example of the database.

## 200 .1 III.

## 201 .2 Problems and Directions

202 The various existing data mining algorithm for market basket analysis are discussed in this paper. All the  
 203 techniques have its own advantages and disadvantages. This section provides some of the drawbacks of the  
 204 existing algorithms and the techniques to overcome those difficulties.

205 Among the methods discussed for data mining, apriori algorithm is found to be better for association rule  
 206 mining. Still there are various difficulties faced by apriori algorithm. The various difficulties faced by apriori  
 207 algorithm are It scans the database lot of times. Every time the additional choices will be created during the  
 208 scan process. This creates the additional work for the database to search. Therefore database must store huge  
 209 number of data services. This results in lack of memory to store those additional data. Also, the I/O load is not  
 210 sufficient and it takes very long time for processing. This results in very low efficiency.

211 Frequent item in the larger set length of the circumstances, leads to significant increase in computing time.

212 Algorithm to narrow face. At this situation, the algorithm will not result in better result. Therefore it is  
 213 required to improve, or even need to re-design of algorithms.

214 Those drawbacks can be overcome by modifying the apriori algorithm effectively. The time complexity for  
 215 the execution of apriori algorithm can be solved by using the fast apriori algorithm. This has the possibility of  
 216 leading to lack of accuracy in determining the association rule. To overcome this, the fuzzy logic can be combined  
 217 with the apriori algorithm. This will help in better selection of association rules for market basket analysis.

218 IV.

## 219 .3 Conclusion

220 Information is collected almost everywhere in our everyday lives. This leads to the huge increase in the amount of  
 221 data available. Physical analysis of these huge amount of information stored in modern databases is very difficult.  
 222 Data mining provides tools to reveal unknown information in large databases which are stored already. A well-  
 223 known data mining technique is association rule mining. Association rule mining and classification technique to  
 224 find the related information in large databases is becoming very important in the current scenario. Association  
 225 rules are very efficient in revealing all the interesting relationships in a relatively large database with huge amount  
 226 of data. The large quantity of information collected through the set of association rules can be used not only for  
 227 illustrating the relationships in the database, but also used for differentiating between different kinds of classes  
 228 in a database. This paper provides some of the existing data mining algorithms for market basket analysis. The  
 229 analysis of existing algorithms suggests that the usage of association rule mining algorithms for market basket  
 230 analysis will help in better classification of the huge amount of data. The apriori algorithm can be modified  
 231 effectively to reduce the time complexity and enhance the accuracy.

232 [Wang et al. ()] 'A Novel Rule Weighting Approach in Classification Association Rule Mining'. Y J Wang , F Qin  
 233 Xin , Coenen . *Seventh IEEE International Conference on Data Mining Workshops*, 2007. 2007. p. . (ICDM  
 234 Workshops)

235 [Bartik ()] 'Association based Classification for Relational Data and its Use in Web Mining'. V Bartik . *CIDM*  
 236 '09, *IEEE Symposium on Computational Intelligence and Data Mining*, 2009. p. .

237 [Niu et al. ()] 'Association Classification Based on Compactness of Rules'. Qiang Niu , Lei Shi-Xiong Xia , Zhang  
 238 . *Second International Workshop on Knowledge Discovery and Data Mining*, 2009. 2009. p. .

239 [Hao et al. ()] 'Improved Classification based on Predictive Association Rules'. Zhixin Hao , Xuan Wang , Lin  
 240 Yao , Yaoyun Zhang . *SMC 2009, IEEE International Conference on Systems, Man and Cybernetics*, 2009.  
 241 p. .

242 [Wen-Xiu et al. ()] 'Market Basket Analysis Based on Text Segmentation and Association Rule Mining'. Xie  
 243 Wen-Xiu , Qi Heng-Nian , Huang Mei-Li . *First International Conference on Networking and Distributed*  
 244 *Computing (ICNDC)*, 2010. p. .

245 [Cunningham and Frank ()] 'Market Basket Analysis of Library Circulation Data'. S J Cunningham , E Frank .  
 246 *International Conference on Neural Information Processing*, 1999. 2 p. .

247 [Trnka ()] 'Market Basket Analysis with Data Mining Methods'. A Trnka . *International Conference on*  
 248 *Networking and Information Technology (ICNIT)*, 2010. p. .

### 3 MARKET BASKET ANALYSIS OF LIBRARY CIRCULATION

---

- 249 [Chiu et al. ()] ‘Market-basket Analysis with Principal Component Analysis: An Exploration’. K S Y Chiu , R  
250 W P Luk , K C C Chan , K F L Chung . *IEEE International Conference on Systems, Man and Cybernetics*,  
251 2002. 3.
- 252 [Wang et al. ()] ‘Mining Association Rules Based on Apriori Algorithm and Application’. Pei-Ji Wang , Lin Shi  
253 , Jin-Niu Bai , Yu-Lin Zhao . *International Forum on Computer Science-Technology and Applications* 2009.  
254 1 p. .
- 255 [Xu et al. ()] ‘Mining Association Rules with New Measure Criteria’. Yong Xu , Sen-Xin Zhou , Jin-Hua Gong .  
256 *International Conference on Machine Learning and Cybernetics*, 2005. 4 p. .
- 257 [Yanthy et al. ()] ‘Mining Interesting Rules by Association and Classification Algorithms’. W Yanthy , T Sekiya ,  
258 K Yamaguchi . *FCST '09. Fourth International Conference on Frontier of Computer Science and Technology*,  
259 2009. p. .
- 260 [Sha and Li ()] ‘Mining local association patterns from spatial dataset’. Zongyao Sha , Xiaolei Li . *Seventh*  
261 *International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, 2010. 3 p. .
- 262 [Rastogi and Shim ()] ‘Mining optimized association rules with categorical and numeric attributes’. R Rastogi ,  
263 Kyuseok Shim . *IEEE Transactions on Knowledge and Data Engineering* 2002. 14 (1) .
- 264 [Vo and Le ()] ‘Mining traditional association rules using frequent itemsets lattice’. B Vo , B Le . *International*  
265 *Conference on Computers & Industrial Engineering* 2009. p. .
- 266 [Xuping et al. ()] ‘Research on Association Rules Mining Based-On Ontology in E-Commerce’. Wang Xuping  
267 , Ni Zijian , Cao Haiyan . *International Conference on Wireless Communications, Networking and Mobile*  
268 *Computing*, 2007. p. .
- 269 [Sumithra and Paul ()] ‘Using Distributed Apriori Association Rule and Classical Apriori Mining Algorithms  
270 for Grid Based Knowledge Discovery’. R Sumithra , S Paul . *International Conference on Computing*  
271 *Communication and Networking Technologies (ICCCNT)*, 2010. p. .