# Proposal for an Association Rules Algorithm Based on Logical AND Operation

**Prof. Dr. Hilal Hadi Saleh*** **Dr. Ahmed Tariq Sadiq***
**Emad Kadhum Jabbar***

Received on: 29 / 4 / 2004
Accepted on: 1 /11/2004

**Abstract**

Data mining is the task of discovering interesting knowledge from large amounts of data where the data can be stored in database, data warehouse, or other information repositories. A knowledge discovery process includes data cleaning, data integration, data selection, data transformation, data mining, pattern evolution, and knowledge presentation. Data mining is a good way for extracting or mining knowledge from amount of data for classification, predication, estimation, clustering or association rules or any activities, which need decision. Association rules identify relationships between attributes in a database. Association rule mining consists of first finding frequent itemsets which satisfy a minimum support threshold, and then computes confidence percentage for each k-itemsets to construct strong association rules. The proposed algorithm aims to produce association rules depending on logical AND operation by converting the database transaction into binary representation and neglecting any sum (column) less than threshold to find the identical column in (k-1)-itemset table with the column in k-itemset table which represent the association rules.

**الخلاصة**

إن تنقيب البيانات هو مهمة لاستكشاف المعلومات من الكم الهائل للبيانات المخزونة في قواعد البيانات، أو في مستودعات البيانات، أو أي مستودعات أخرى للبيانات. وتتضمن عملية استكشاف المعرفة التنظيف، والتكامل، والاختيار، والتحويل، والتنقيب، وتطوير النماذج، وتمثيل المعرفة. يعتبر تنقيب البيانات طريقة جيدة لاستخلاص المعرفة من البيانات لغرض التصنيف، والتنبؤ، والتخمين، والتجميع، أو إيجاد العلاقات الترابطية أو أية أنشطة أخرى تحتاج إلى قرار. وتعرف العلاقات الترابطية العلاقات ما بين عناصر قاعدة البيانات. ويتضمن التنقيب عن العلاقات الترابطية إيجاد تكرارات العناصر التي تحقق الإسناد الأدنى على وفق حد عتبة معين، ومن ثم حساب نسبة الوثوق لكل عنصر لتكوين علاقات ترابطية قوية. وتهدف الخوارزمية المقترحة إلى إنتاج علاقات ترابطية، إستناداً إلى دالة الربط المنطقي AND، وذلك من خلال تحويل الحركات في قواعد البيانات إلى تمثيل ثنائي وإهمال الأعمدة التي يقل مجموعها عن حد عتبة معين لإيجاد الأعمدة المتطابقة في كل جدول والجدول السابق له، والتي هي بمثابة علاقات ترابطية.

*Department Of Computer Science, University Of Technology, Baghdad, Iraq.

data mining are [2]:

1- **Data cleaning:** it is the step at which noise and inconsistent data are removed.

2- **Data integration:** where multiple data sources combined together such as relational database, flat file, and on-line transaction records, data integration techniques are applied to ensure consistency in naming conventions, encoding structure, attribute measures, and so on.

3- **Data selection:** To choose data that is appropriate for data mining system, it's important to have multiple views of data; like data type, data means, data structure, format and other characteristics to analyze task and retrieve from the database.

4- **Data transformation:** Data is transformed or consolidated into forms appropriate for mining by performing summary or aggregation operations, to produce the analytical data model of the data selected. This is a crucial phase to the accuracy and validity of the final result. In general, data transformation, such as normalization, improves the accuracy and efficiency of mining algorithms, which involve distance measurement.

5- **Data mining:** It is an essential process where intelligent methods are applied in order to extract data patterns, and it is the main step in the knowledge discovery. The data-mining step may interact with the user or acknowledge base. The interesting patterns are presented to the user, and may be stored as new knowledge in the knowledge base. Note that according to this view, data mining is only one step in the entire process, albeit an essential one since it uncovers hidden patterns for evaluation [5].

6- **Pattern evaluation:** To identify the truly interesting patterns representing knowledge based on some interestingness measure.

7- **Knowledge presentation:** Where visualization and knowledge representation techniques are used to present the mined knowledge to the user [4].

## 4- The Proposed Algorithm

The proposed algorithm aims to produce an algorithm to mine association rules depending on logical AND operator.

An association rule means that a relationship exists between all items generating it and this associated relationship must have attendant items. So an AND operator is very useful and is perfect to express such kind of relation. For instance, the relation between A and B means that if A exists, then B also exists.

To simplify the task of AND operator, the database must be converted into binary representation. This helps to construct binary tables of different relations (1-1), (2-1), (3-1), (2-2),... between all the items of database transactions.

The following steps represent the proposed association rule algorithm:

***Step 1***: Convert items of database transaction into binary representation as in table (1-itemset):

Table (1-itemset)

| A | B | C | D | E |
|---|---|---|---|---|
| 0 | 1 | 1 | 0 | 1 |
| 0 | 1 | 1 | 1 | 1 |
| . | . | . | . | . |
| . | . | . | . | . |

***Step 2***: From table (1-itemset) construct table (2,3,4,5,k-itemset) by using the AND operator between each item and

other items.

***Step 3:*** Generate association rules from each construct Table (k-item sets) by:

3-1 Neglect any column in each table (1,2,3…k-item set), which have Sum (column) < threshold, when initial k=1.

3-2 Determine column of table ((k-1)-item set), which is a subset of table (k-item set) when initial k=2.

Such that:

Column (table ((k-1)-item set)) $\subseteq$ column (table (k-item set))

Table (1-itemset) is subset of AB in table (2-itemset)

3-3 Find which column of table ((k-1)-item set) is identical to Column of table (k-itemset) when initial k=2.

Such that:

IF column ((k-1)-itemset) $\equiv$ column (k-item set) then

Item (column ((k-1)-itemset)) imply to association rule with

Item (column (k-itemset))

Like A $\rightarrow$ AB this means that A has an association rule with B

Example:

Consider the following transaction DB with set items {A, B, C, D, E}. Each item represents an attribute name with transaction identifiers TID 1,2,3,4,5,6 as follows:

Table-DB Transaction

| TID | Transaction |
|---|---|
| 1 | B, C, E |
| 2 | B, C, D, E |
| 3 | A, B, C, D, E |
| 4 | B, C, D |
| 5 | A, B |
| 6 | A, B, C, E |

**Solution:**

Now due to proposed algorithm

1- From table-DB Transaction construct binary table 1-itemset

Table 1-itemset

| A | B | C | D | E |
|---|---|---|---|---|
| 0 | 1 | 1 | 0 | 1 |
| 0 | 1 | 1 | 1 | 1 |
| 1 | 1 | 1 | 1 | 1 |
| 0 | 1 | 1 | 1 | 0 |
| 1 | 1 | 0 | 0 | 0 |
| 1 | 1 | 1 | 0 | 1 |

2- From table 1-itemset construct table 2-itemset by using AND operator With each pairs items

Table 2-itemset

| AB | AC | AD | AE | BC | BD | BE | CD | CE | DE |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 |
| 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 0 |

3- From table 1-itemset construct table 3-itemset by using AND operator With each pairs items

Table 3-itemset

| ABC | ABD | ABE | ACD | ACE | ADE | BCD | BCE | BDE | CDE |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |

4- From table 1-itemset construct table 4-itemset by using AND operator With each four items

Table 4-itemset

| ABCD | ABCE | ABDE | ACDE | BCDE |
|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 1 |
| 1 | 1 | 1 | 0 | 1 |
| 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 |
| 0 | 1 | 0 | 0 | 0 |

5- From table 1-itemset construct table 5-itemset by using AND operator With each five items

Table 5-itemset

| ABCDE |
|---|
| 0 |
| 0 |
| 1 |
| 0 |
| 0 |
| 0 |

6- Neglect all columns in all tables k-item which have sum(item) < threshold when threshold =2 (these columns are pointed by gray color on its tables)

Neglected columns are:

| | | |
|---|---|---|
| Column | AD | In table 2-itemsts |
| Column | ACD | In table 3-itemsts |
| Column | ABD | In table 3-itemsts |
| Column | ADE | In table 3-itemsts |
| Column | ABCD | In table 4-itemsts |
| Column | ABDE | In table 4-itemsts |
| Column | ACDE | In table 4-itemsts |
| Column | ABCDE | In table 5-itemsts |

7- Search for identical columns in table 1-itemset and table 2-itemset to construct association rules (1→1)
When Column (table 1-itemset) ⊆ column (table 2-itemset)

| Table 1-itemset Columns | Identical | Table 2-itemset Columns | Imply to | Association Rules |
|---|---|---|---|---|
| A | ≡ | AB | → | A → B |
| C | ≡ | BC | → | C → B |
| D | ≡ | BD | → | D → B |
| E | ≡ | BE | → | E → B |
| D | ≡ | CD | → | D → C |
| E | ≡ | CE | → | E → C |

8- Search for identical columns in table 2-itemset and table 3-itemset to construct association rules (2→1) when Column (table 2-itemset) ⊆ column (table 3-itemset)

| Table 2-itemset Columns | Identical | Table 3-itemset Columns | Imply to | Association rules |
|---|---|---|---|---|
| AC | ≡ | ABC | → | AC → B |
| AE | ≡ | ABE | → | AE → B |
| AC | ≡ | ACE | → | AC → E |
| AE | ≡ | ACE | → | AE → C |
| BD | ≡ | BCD | → | BD → C |
| CD | ≡ | BCD | → | CD → B |
| BE | ≡ | BCE | → | BE → C |
| CE | ≡ | BCE | → | CE → B |
| DE | ≡ | BDE | → | DE → B |
| DE | ≡ | CDE | → | DE → C |

9- Search for identical columns in table 3-itemset and table 4-itemset to construct association rules (3→1) when Column (table 3-itemset) CC column (table 4-itemset)

| Table 3-itemset Columns | Identical | Table 4-itemset Columns | Imply to | Association rules |
|---|---|---|---|---|
| ABC | ≡ | ABCE | → | ABC → E |
| ABE | ≡ | ABCE | → | ABE → C |
| ACE | ≡ | ABCE | → | ACE → B |
| BDE | ≡ | BCDE | → | BDE → C |
| CDE | ≡ | BCDE | → | CDE → B |

10- search for identical columns in table 2-itemset and table 4-itemset to construct association rules (2→2) when column (table 2-itemset) CC column(table 4-itemset)

| Table 2-itemset Columns | Identical | Table 4-itemset Columns | Imply to | Association rules |
|---|---|---|---|---|
| AC | ≡ | ABCE | → | AC → BE |
| AE | ≡ | ABCE | → | AE → BC |
| DE | ≡ | BCDE | → | BE → BC |

11- Since no k-itemsets can generate other association rules so the Algorithm must be stopped.

Now the all association rules are:

| Association rules | |
|---|---|
| A → B | BE → C |
| C → B | CE → B |
| D → B | DE → B |
| E → B | DE → C |
| D → C | ABC → E |
| E → C | ABE → C |
| AC → B | ACE → B |
| AE → B | BDE → C |
| AC → E | CDE → B |
| AE → C | AC → BE |
| BD → C | AE → BC |
| CD → B | BE → BC |

## 5- Conclusions

Association rules identify relationships among items in database such as the presence or absence of one pattern implies the presence or absence of another pattern. Mining of such rules is one of the most popular pattern discovery methods in KDD, Apriori algorithm is used to generate association rules by finding large itemset $L_k$ by two-step process (join and prune) to reduce the search space. An algorithm that finds association rules needs too many scan and computation operations to get a lot of properly rules which need too much time. In the proposed algorithm, there is one scan operation to convert database transaction into binary representation and construct k-itemsets to identify identical column (k-itemset table) with column ((k-1)-itemset table) which represents association rules, so it does not need too much time and long computational operation to compute support and confidence percentage because this operation happens implicitly through identifying identical columns. All association rules discovered by this method have the same association rules discovered by Apriori method.

## References

[1] Peter Cabena, Pablo Hadjinian, Rolf Stadler, Jaap Verhees, Alessandro Zanasi, "**Discovering Data Mining From Concept to Implementation**", USA ,1997.

[2] Jiawei Han, Micheline Kamber, "**Data Mining Concept and Techniques**", USA, 2001.

[3] Michael J. A. Berry, Gordon S. Linoff, "**Mastering Data Mining the Art and Science of Customer Relationship Management** ", USA, 2000.

[4] Hipp J., Ulrich Guentzer, Gholamreza Nakhaeizader, "**Algorithms For Survey and Comparision** " USA, 2000.

[5] Alaa H. Al-Hamami, Abbas F. Kader, Hussein K. Al Khafagi, "**A New Approach to Mine Negative Association Rules**", Journal of Al-Rafidain Uni. Coll. No 10,2002.