Mining Association Rules in OLAP Cubes

Riadh Ben Messaoud[♭], Omar Boussaid[‡], and Sabine Loudcher Rabaséda[‡]
Laboratory ERIC – University of Lyon 2
5 avenue Pierre Mendès-France, 69676, Bron Cedex, France
[♭]rbenmessaoud@eric.univ-lyon2.fr, [‡]{ boussaid | sabine.loudcher }@univ-lyon2.fr

Abstract

On-line analytical processing (OLAP) provides tools to explore data cubes in order to extract interesting information. Nevertheless, OLAP is not capable of explaining relationships that could exist within data. Association rules are one kind of data mining techniques which finds associations among data. In this paper, we propose a framework for mining association rules from data cubes according to a sum-based aggregate measure *which is more general than frequencies provided by the* COUNT *measure. Our mining process is guided by a meta-rule context driven by analysis objectives and exploits aggregate measures to revisit the definition of support and confidence. We also evaluate the interestingness of mined association rules according to Lift and Loevinger criteria and propose an algorithm for mining inter-dimensional association rules directly from a multidimensional structure of data.*

1 Introduction

Data warehousing and OLAP technology has known important progress since the 90s. In addition, with efficient techniques developed for computing data cubes, OLAP users have become widely able to explore multidimensional data, navigate through hierarchical levels of dimensions, and therefore extract interesting information according to multiple levels of granularity. Nevertheless, the OLAP technology is quite limited to an exploratory task and does not provide automatic tools to explain relationships and associations within data. Users are usually supposed to explore the data cube according to multiple dimensions in order to manually find an explanation for a given phenomenon.

In the recent years, many studies addressed the issue of performing data mining tasks on data warehouses. Some of them were specifically interested to mining patterns and association rules in data cubes. For instance, Imieliński *et al.* state that *OLAP is closely intertwined with association rules and shares the goal with association rules for finding patterns in the data* [4]. Data mining techniques such as association rule mining can be used together with OLAP to discover knowledge from data cubes. The aggregate values needed for discovering association rules are already precomputed and stored in the data cube. The COUNT cells of a cube store the number of occurrences of the corresponding multidimensional data values. With such summary cells, it is straightforward to calculate the values of the support and the confidence of association rules.

The COUNT measure corresponds to the frequency of facts. Nevertheless, in an analysis process, users are usually interested in observing multidimensional data and their associations according to measures more relevant than simple frequencies. In this paper, we establish a general framework for mining *inter-dimensional* association rules from multidimensional data. We use the concept of *inter-dimensional meta-rule* which allows users to guide the mining process and focus on a specific context from which rules can be extracted. Our framework also allows a redefinition of the support and confidence measures based on the SUM aggregate functions over cube indicators (measures). Therefore, the computation of support and confidence according to the COUNT measure becomes a particular case in our proposal. In addition to support and confidence, we use two other descriptive criteria (*Lift* and *Loevinger*) in order to evaluate interestingness of mined associations. These criteria reflect interestingness of associations in a more relevant way than what is offered by support and confidence. We developed an efficient *bottom-up* algorithm which adapts the traditional Apriori algorithm in order to handle multidimensional data.

This paper is organized as follows. In Section 2, we expose a state of the art about association rule mining from multidimensional data. In Section 3, we define the concept of *inter-dimensional meta-rule*, present the general computation of support and confidence based on measures, and provide advanced evaluation of mined association rules. Section 4 describes our algorithm of mining inter-dimensional association rules and shows its efficiency with experiments. Finally, in Section 5, we conclude and propose some future research directions.

2 Related work

To the best of our knowledge, Kamber *et al.* [5] were the first who addressed the issue of mining *association rules from multidimensional data*. They introduced the *metarule-guided mining* which uses rule templates defined by users in order to guide the mining process of inter-dimensional association rules. Zhu divides the problem of mining association rules from data cubes to *inter-dimensional*, *intra-dimensional*, and *hybrid* association mining [10]. In [3], Chen *et al.* mine intra-dimensional association rules by adding *features* from other dimensions at multiple levels. *Extended association rules* [8] consist of repetitive predicates by involving attributes from user defined non-item dimensions. Tjioe and Taniar [9] extract associations from multiple dimensions by focusing on summarized data. They prepare multidimensional data for the mining process by pruning rows in the fact table which have less than the average quantity.

All the proposed approaches are restricted to the COUNT measure in the mining process of associations. In this paper, we use the notion of metarule-guided mining proposed by Kamber *et al.* [5] to guide a general process of mining inter-dimensional associations with non-repetitive predicates. The main contribution of our proposal consists in integrating the measures of a data cube in the computation of the support and the confidence of association rules. We also use advanced criteria in order to evaluate interestingness of mined associations. An Apriori-based algorithm is also adapted in order to handle multidimensional data.

3 The proposed framework

3.1 Notations and terminologies

Let $\mathcal{C}$ be a data cube with a non empty set of d dimensions $\mathcal{D} = \{D_i\}_{(1 \leq i \leq d)}$ and a non empty set of measures $\mathcal{M}$. Each dimension $D_i \in \mathcal{D}$ encloses a non empty set of hierarchical levels. We assume that H^i_j is the j^{th} $(j \geq 0)$ hierarchical level in D_i. The coarse level of D_i, denoted H^i_0, corresponds to its total aggregation level All. Let $\mathcal{H}_i$ be the set of hierarchical levels of dimension D_i where each hierarchical level $H^i_j \in \mathcal{H}_i$ consists of a non empty set of members denoted $\mathcal{A}_{ij}$.

Let $\mathcal{D}' \subseteq \mathcal{D}$ be a non empty set of p dimensions $\{D_1, \ldots, D_p\}$ from the data cube $\mathcal{C}$ $(p \leq d)$. The p-tuple $(\Theta_1, \ldots, \Theta_p)$ defines a sub-cube on $\mathcal{C}$ according to $\mathcal{D}'$ iff $\forall i \in \{1, \ldots, p\}$, $\Theta_i \neq \emptyset$ and there exists a unique j such that $\Theta_i \subseteq \mathcal{A}_{ij}$. A sub-cube according to a set of dimensions $\mathcal{D}'$ corresponds to a portion from the initial data cube $\mathcal{C}$. It consists in setting for each dimension from $\mathcal{D}'$ a non empty subset of member values from a single hierarchical level of that dimension. Each cell from the data cube $\mathcal{C}$ represents an OLAP fact which is evaluated in $\mathbb{R}$ according to one measure M from $\mathcal{M}$. We evaluate a sub-cube according to its *sum-based aggregate measure*. The sum-based aggregate measure of a sub-cube $(\Theta_1, \ldots, \Theta_p)$ on $\mathcal{C}$ according to $M \in \mathcal{M}$, noted $M(\Theta_1, \ldots, \Theta_p)$, is the SUM of measure M of all facts in the sub-cube.

We define a dimension predicate α_i in a D_i as the predicate $\langle a \in \mathcal{A}_{ij} \rangle$ which takes a dimension member as a value. Let $\mathcal{D}' \subseteq \mathcal{D}$ be a non empty set of p dimensions $\{D_1, \ldots, D_p\}$ from the data cube $\mathcal{C}$ $(2 \leq p \leq d)$. We also define $(\alpha_1 \wedge \cdots \wedge \alpha_p)$ as an inter-dimensional predicate in $\mathcal{D}'$ iff $\forall i \in \{1, \ldots, p\}$, α_i is a dimension predicate in D_i.

3.2 Inter-dimensional meta-rules

We consider two distinct subsets of dimensions in the data cube $\mathcal{C}$: (1) $\mathcal{D}_\mathcal{C} \subset \mathcal{D}$ is a subset of p *context dimensions*; and (2) $\mathcal{D}_\mathcal{A}$ is a subset of $(s+r)$ *analysis dimensions*. An inter-dimensional meta-rule is of the following form:

$$\left| \begin{array}{l} \text{In the context } (\Theta_1, \ldots, \Theta_p) \\ (\alpha_1 \wedge \cdots \wedge \alpha_s) \Rightarrow (\beta_1 \wedge \cdots \wedge \beta_r) \end{array} \right. \quad (1)$$

$(\Theta_1, \ldots, \Theta_p)$ is a sub-cube on $\mathcal{C}$ according to $\mathcal{D}_\mathcal{C}$. It defines the portion of cube $\mathcal{C}$ to be mined. Unlike [5], our meta-rule allows the user to target a mining context by identifying a particular sub-cube to be explored. We note that $\forall k \in \{1, \ldots, s\}$ (respectively $\forall k \in \{1, \ldots, r\}$), α_k (respectively β_k) is a dimension predicate in a distinct dimension from $\mathcal{D}_\mathcal{A}$. Therefore, $(\alpha_1 \wedge \cdots \wedge \alpha_s) \wedge (\beta_1 \wedge \cdots \wedge \beta_r)$ is an inter-dimensional predicates in $\mathcal{D}_\mathcal{A}$.

3.3 Measure-based support and confidence

Traditionally, the support (SUPP) of an association rule $X \Rightarrow Y$, in a database of transactions $\mathcal{T}$, is the probability that the population of transactions contains both X and Y [1]. The confidence (CONF) of $X \Rightarrow Y$ is the conditional probability that a transaction contains Y given that it already contains X. In the case of a data cube $\mathcal{C}$, the aggregate values needed for discovering association rules are already computed and stored in $\mathcal{C}$. In fact, a data cube stores the particular COUNT measure which represents precomputed frequencies of OLAP facts. Nevertheless, with the COUNT measure, only number of occurrences of facts are taken into account to decide whether a rule is *large* (respectively *strong*) or not. However, in the OLAP context, users are usually interested to observe facts according to summarized values of measures more expressive than their simple number of occurrences. It is naturally significant to observe association rules according to the sum of these measures. Let us consider a data cube of *Sales* by taking once the COUNT measure (Table 1(a)) and then the total profit measure (Table 1(b)). In this example, with a *minsupp* $= 0.2$, the itemsets $(\langle$ America $\rangle$, $\langle$ MP3 $\rangle$, $\langle$ 2004 $\rangle)$ and $(\langle$ America $\rangle$, $\langle$ MP3 $\rangle$, $\langle$ 2005 $\rangle)$ are *large* according to the COUNT measure (grayed cells in Table 1(a)). Whereas these itemsets are not *large* in Table 1(b). The *large* itemsets according to the profit measure are rather (⟨ Europe ⟩, ⟨ Laptop ⟩, ⟨ 2004 ⟩) and (⟨ Europe ⟩, ⟨ Laptop ⟩, ⟨ 2005 ⟩).

Table 1. Sales cube according to the (a) COUNT and the (b) profit measure

	2004		2005	
	America	Europe	America	Europe
Desktop	1,200	800	950	500
Laptop	2,500	2,700	2,800	3,200
MP3	10,600	5,900	11,400	9,100

(a)

	2004		2005	
	America	Europe	America	Europe
Desktop	$ 60,000	$ 33,000	$ 28,000	$ 10,000
Laptop	$ 500,000	$ 567,000	$ 420,000	$ 544,000
MP3	$ 116,000	$ 118,000	$ 57,000	$ 91,000

(b)

In the OLAP context, the rule mining process needs to handle any measure from the data cube in order to evaluate its interestingness. Therefore, a rule is not merely evaluated according to probabilities based on frequencies of facts, but needs to be evaluated according to quantity measures of its corresponding facts. The choice of the measure closely depends on the analysis context according to which a user needs to discover associations within data. For instance, if a firm manager needs to see strong associations of sales covered by achieved profits, it is more suitable to compute the support and the confidence of these associations based on units of profits rather than on unit of sales themselves. Therefore, we define a general computation of support and confidence of inter-dimensional association rules according to a user defined measure $M \in \mathcal{M}$ from the mined data cube. Consider a general rule R which complies with the defined inter-dimensional meta-rule (1):

$$R: \left| \begin{array}{l} \text{In the context } (\Theta_1, \ldots, \Theta_p) \\ (x_1 \wedge \cdots \wedge x_s) \Rightarrow (y_1 \wedge \cdots \wedge y_r) \end{array} \right.$$

The support and the confidence of this rule are therefore computed according to the following general expressions:

$$\text{SUPP}(R) = \frac{M(x_1,\ldots,x_s,y_1,\ldots,y_r,\Theta_1,\ldots,\Theta_p,\text{All},\ldots,\text{All})}{M(\text{All},\ldots,\text{All},\Theta_1,\ldots,\Theta_p,\text{All},\ldots,\text{All})} \quad (2)$$

$$\text{CONF}(R) = \frac{M(x_1,\ldots,x_s,y_1,\ldots,y_r,\Theta_1,\ldots,\Theta_p,\text{All},\ldots,\text{All})}{M(x_1,\ldots,x_s,\text{All},\ldots,\text{All},\Theta_1,\ldots,\Theta_p,\text{All},\ldots,\text{All})} \quad (3)$$

where $M(x_1, \ldots, x_s, y_1, \ldots, y_r, \Theta_1, \ldots, \Theta_p, \text{All}, \ldots, \text{All})$ is the sum-based aggregate measure of a sub-cube. Traditional support and confidence are particular cases of the above expressions which can be obtained by the COUNT aggregation. Nevertheless, in order to simplify notations, we keep on refering to our generalized support and confidence with the usual terms.

3.4 Advanced evaluation of association rules

Support and confidence are the most known criteria for the evaluation of association rule interestingness. These criteria are the fundamental principles of all Apriori-like algorithms [1]. However, they usually produce a large number of rules which may not be interesting.

Let consider again the association rule $R : X \Rightarrow Y$ which complies with the inter-dimensional meta-rule (1), where $X = (x_1 \wedge \cdots \wedge x_s)$ and $Y = (y_1 \wedge \cdots \wedge y_r)$ are conjunctions of dimension predicates. We also consider a user-defined measure $M \in \mathcal{M}$ from data cube $\mathcal{C}$. We denote by P_X (respectively P_Y, P_{XY}) the relative measure M of facts matching X (respectively Y, X and Y) in the sub-cube defined by the instance $(\Theta_1, \ldots, \Theta_p)$ in the context dimensions $\mathcal{D_C}$. We also denote by $P_{\overline{X}} = 1 - P_X$ (respectively $P_{\overline{Y}} = 1 - P_Y$) the relative measure M of facts not matching X (respectively Y), i.e., the probability of not having X (respectively Y). The support of R is equal to P_{XY} and its confidence is defined by the ratio $\frac{P_{XY}}{P_X}$ which is a conditional probability, denoted $P_{Y/X}$, of matching Y given that X is already matched.

$$P_X = \frac{M(x_1,\ldots,x_s,\text{All},\ldots,\text{All},\Theta_1,\ldots,\Theta_p,\text{All},\ldots,\text{All})}{M(\text{All},\ldots,\text{All},\Theta_1,\ldots,\Theta_p,\text{All},\ldots,\text{All})}$$

$$P_Y = \frac{M(\text{All},\ldots,\text{All},y_1,\ldots,y_r,\Theta_1,\ldots,\Theta_p,\text{All},\ldots,\text{All})}{M(\text{All},\ldots,\text{All},\Theta_1,\ldots,\Theta_p,\text{All},\ldots,\text{All})}$$

$$P_{XY} = \text{SUPP}(R) = \frac{M(x_1,\ldots,x_s,y_1,\ldots,y_r,\Theta_1,\ldots,\Theta_p,\text{All},\ldots,\text{All})}{M(\text{All},\ldots,\text{All},\Theta_1,\ldots,\Theta_p,\text{All},\ldots,\text{All})}$$

$$P_{Y/X} = \text{CONF}(R) = \frac{M(x_1,\ldots,x_s,y_1,\ldots,y_r,\Theta_1,\ldots,\Theta_p,\text{All},\ldots,\text{All})}{M(x_1,\ldots,x_s,\text{All},\ldots,\text{All},\Theta_1,\ldots,\Theta_p,\text{All},\ldots,\text{All})}$$

There are two categories of frequently used evaluation criteria to capture the interestingness of association rules: *descriptive* criteria and *statistical* criteria. In general, one of the most important drawbacks of a statistical criterion is that it depends on the size of the mined population. In addition, it requires a *probabilistic* approach to model the mined population. This approach assumes advanced statistical knowledge of users, which is not particulary true for OLAP users. On the other hand, descriptive criteria are easy to use and express interestingness of association rules in a more natural manner. In addition to support and confidence, we add the Lift (LIFT) [2] and the Loevinger criterion (LOEV) [6]. These two criteria are descriptive, take the independence of itemsets X and Y as a reference, and are defined on rule R as follows:

$$\text{LIFT}(R) = \frac{P_{YX}}{P_X P_Y} = \frac{\text{SUPP}(R)}{P_X P_Y} \quad (4)$$

$$\text{LOEV}(R) = \frac{P_{Y/X} - P_Y}{P_{\overline{Y}}} = \frac{\text{CONF}(R) - P_Y}{P_{\overline{Y}}} \quad (5)$$

The Lift of a rule can be interpreted as the deviation of the support of the rule from the support expected under the independence hypothesis between the body X and the head Y [2]. For the rule R, the Lift captures the deviation from the independence of X and Y. This also means that the Lift criterion represents the probability scale coefficient of having Y when X occurs. By opposition to the confidence, which considers directional implication, the Lift directly captures correlation between body X and its head Y. In general, greater Lift values indicate stronger associations. The Loevinger criterion is one of the oldest used interestingness evaluation for association rules [6]. It consists in a linear transformation of the confidence achieved by centering it on P_Y and dividing it by the scale coefficient $P_{\overline{Y}}$. In other terms, the Loevinger criterion normalizes the centered confidence of a rule according to the probability of not satisfying its head.

4 Implementation, algorithm and experiments

We developed a Web application to mine association rules from data cubes according to our proposal. This application is a module evolving in a general Client/Server platform, called MiningCubes [7]. The platform enables connection to multidimensional data cubes stored in the *Analysis Services* of *MS SQL Server 2000*. A *Mining Association Rule Module* allows the definition of analysis dimensions $\mathcal{D}_\mathcal{A}$, context dimensions $\mathcal{D}_\mathcal{C}$, a meta-rule with its context sub-cube $(\Theta_1, \ldots, \Theta_p)$ and its inter-dimensional predicates scheme $(\alpha_1 \wedge \cdots \wedge \alpha_s) \Rightarrow (\beta_1 \wedge \cdots \wedge \beta_r)$, the measure M used to compute criteria of association rules, the minimum support threshold *minsupp*, and the minimum confidence threshold *minconf*.

Traditionally, frequent itemsets can be mined according to a *top-down* search or a *bottom-up* search. The *bottom-up* approach complies with the *antimonotony* property of the Apriori algorithm [1] which states that *for each non frequent itemset, all its super-itemsets are definitely not frequent*. This property enables the reduction of the search space, especially when it deals with large and sparse data sets, which is particulary the case of OLAP data cubes. As summarized in Algorithm 1, we proceed by an *bottom-up level wise* search for large i-itemsets, where level i is the number of items in the set. We denote by $C(i)$ the sets of *i-candidates*, i.e., i-itemsets that are potentially frequent, and $F(i)$ the sets of *i-frequents*, i.e., frequent i-itemsets.

At **the initialization step**, our algorithm captures the 1-candidates from user defined analysis dimensions $\mathcal{D}_\mathcal{A}$ over the data cube $\mathcal{C}$. These 1-candidates correspond to members of $\mathcal{D}_\mathcal{A}$, where each member complies with one dimension predicate α_k or β_k in the meta-rule R.

For each level i, if the set $C(i)$ is not empty and i is less than $(s + r)$, **the first step** of our algorithm derives frequent itemsets $F(i)$ from $C(i)$ according to: (i) an itemset $A \in C(i)$ should be an instance of an inter-dimensional predicates in $\mathcal{D}_\mathcal{A}$; and (ii) an itemset $A \in C(i)$ must have a support greater than *minsupp*.

```
input  : C, D_C, D_A, D_U, R, M, minsupp, minconf
output: X ⇒ Y, SUPP, CONF, LIFT, LOEV
C(1) ← ∅;
for i ← 1 to (s + r) do
|   C(1) ← C(1) ∪ A_ij;
end
i ← 1;
while C(i) ≠ ∅ and i ≤ (s + r) do
    F(i) ← ∅;
    foreach A ∈ C(i) do
        if A is an inter-dimensional predicates then
            SUPP ← COMPUTESUPPORT(A, M);
            if SUPP ≥ minsupp then F(i) ← F(i) ∪ {A};
        end
    end
    foreach A ∈ F(i) do
        foreach non empty B ∈ A do
            if A\B ⇒ B complies with R then
                CONF ← COMPUTECONFIDENCE(A\B, B, M);
                if CONF ≥ minconf then
                    X ← A\B;
                    Y ← B;
                    LIFT ← COMPUTELIFT(X, Y, M);
                    LOEV ← COMPUTELOEVINGER(X, Y, M);
                    return (X ⇒ Y, SUPP, CONF, LIFT, LOEV);
                end
            end
        end
    end
    C(i + 1) ← ∅;
    foreach A ∈ F(i) do
        foreach B ∈ F(i) that shares i − 1 items with A do
            if All Z ⊂ {A ∪ B} of i items are inter-dimensional
            predicates and frequent then
            |   C(i + 1) ← C(i + 1) ∪ {A ∪ B};
            end
        end
    end
    i ← i + 1;
end
```

Algorithm 1: **Algorithm for mining association rules in a data cube**

From each $A \in F(i)$, **the second step** extracts association rules with respect to: (i) an association rule $X \Rightarrow Y$ must comply with the user defined meta-rule R; and (ii) an association rule must have a confidence greater than *minconf*. The computation of confidence is also based on the user defined measure M according to formulas (2) and (3). When an association rule satisfies the two previous conditions, the algorithm computes its Lift and Loevinger criteria according to formulas (4) and (5). The computation of support, confidence, Lift and Loevinger criteria are performed respectively by functions: COMPUTESUPPORT, COMPUTECONFIDENCE, COMPUTELIFT and COMPUTELOEVINGER which directly pick up required precomputed aggregates from the data cube via MDX (Multi-Dimensional eXpression) queries.

Based on the Apriori property, **the third step** uses the set $F(i)$ of large i-itemsets to derive a new set $C(i + 1)$ of $(i+1)$-candidates. One $(i+1)$-candidate is the union of two i-itemsets A and B from $F(i)$ that respects three conditions: (i) A and B must have $i - 1$ commun items; (ii) all non empty sub-itemsets from $A \cup B$ must be instances of inter-dimensional predicates in $\mathcal{D}_\mathcal{A}$; and (iii) all non empty sub-itemsets from $A \cup B$ must be frequent itemsets.

According to the experiments[1] presented in Figure 1, for a support and a confidence thresholds equal to 5%, we notice that the efficiency of the algorithm closely depends on the number of extracted frequent itemsets and association rules. The generation of association rules from frequent itemsets is more time consuming than the extraction of frequent itemsets themselves. In fact, an Apriori-based algorithm is efficient for searching frequent itemsets and have a low complexity level especially in the case of sparse data. Nevertheless, the Apriori property does not reduce the running time of extracting association rules from a frequent itemset.

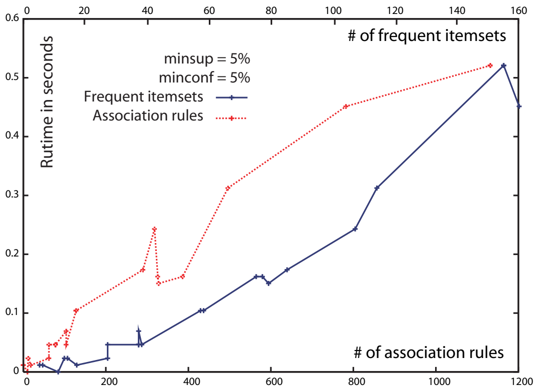

Figure 1. The running times of our algorithm according to # of frequent itemsets and # of association rules

5 Conclusion and perspectives

In this paper, we establish a general framework for mining inter-dimensional association rules from data cubes. We use inter-dimensional meta-rule which allows users to target the mining process in a particular portion in the mined data cube. We provide a general computation of support and confidence of association rules that can be based on any measure from the data cube. This issue is quite interesting since it expresses associations which consider wide analysis objectives and do not restrict users' analysis to associations only driven by the traditional COUNT measure. We also propose to evaluate interestingness of mined rules with two additional descriptive criteria in order to express the relevance of rules in a more precise way than what is offered by the support and the confidence. We developed an efficient *bottom-up* algorithm which adapts the traditional Apriori algorithm in order to handle the multidimensional structure of data.

Some future directions need to be addressed for this work. We plan to embed the measure in the expression of mined inter-dimensional association rules. In addition, we need to profit from the hierarchical aspect of cube dimensions to mine association rules at different level of granularities. Finally, we also have to cope with the visualization of mined association rules in the space representation of the data cube itself in order to make associations easier to interpret and to exploit.

[1]The experiments are conducted under Windows XP on a 1.60GHz PC with 480MB of main memory

References

[1] R. Agrawal, T. Imielinski, and A. Swami. Mining Association Rules between Sets of Items in Large Databases. In *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD 1993)*, pages 207–216, Washington, D.C., USA, May 1993. ACM Press.

[2] S. Brin, R. Motwani, and C. Silverstein. Beyond Market Baskets: Generalizing Association Rules to Correlations. In *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD 1997)*, pages 265–276, Tucson, Arizona, USA, May 1997. ACM Press.

[3] Q. Chen, U. Dayal, and M. Hsu. An OLAP-based Scalable Web Access Analysis Engine. In *Proceedings of the 2^{nd} International Conference on Data Warehousing and Knowledge Discovery (DaWaK 2000)*, Lecture Notes in Computer Science, pages 210–223, London, UK, September 2000. Springer-Verlag.

[4] T. Imieliǹski, L. Khachiyan, and A. Abdulghani. Cube-grades: Generalizing Association Rules. *Data Mining and Knowledge Discovery*, 6(3):219–258, 2002.

[5] M. Kamber, J. Han, and J. Chiang. Metarule-Guided Mining of Multi-Dimensional Association Rules Using Data Cubes. In *Proceedings of the 3^{rd} International Conference on Knowledge Discovery and Data Mining (KDD 1997)*, pages 207–210, Newport Beach, CA, USA, August 1997. The AAAI Press.

[6] J. Loevinger. A Systemic Approach to the Construction and Evaluation of Tests of Ability. *Psychological Monographs*, 61(4), 1947.

[7] R. B. Messaoud, O. Boussaid, and S. L. Rabaséda. A Data Mining-Based OLAP Aggregation of Complex Data: Application on XML Documents. *International Journal of Data Warehousing and Mining*, 2(4):1–26, 2006.

[8] S. Nestorov and N. Jukić. Ad-Hoc Association-Rule Mining within the Data Warehouse. In *Proceedings of the 36^{th} Hawaii International Conference on System Sciences (HICSS 2003)*, pages 232–242, Big Island, Hawaii, USA, January 2003. IEEE Computer Society.

[9] H. C. Tjioe and D. Taniar. Mining Association Rules in Data Warehouses. *International Journal of Data Warehousing and Mining*, 1(3):28–62, 2005.

[10] H. Zhu. On-Line Analytical Mining of Association Rules. Master's thesis, Simon Fraser University, Burnaby, British Columbia, Canada, December 1998.