New Algorithm for Cyclic Mining in Temporal Databases

Eya Ben Ahmed[1], Ahlem Nabli[2], Faiez Gargouri[3]
[1]University of Tunis
High Institute of Management of Tunis
[2]University of Sfax
Faculty of Sciences of Sfax
[3]University of Sfax
Higher Institute of Computer Science and Multimedia
eya.benahmed@gmail.com, ahlem.nabli@fsegs.rnu.tn, faiez.gargouri@isimsf.rnu.tn

ABSTRACT: *The mining of temporal databases using cyclic approach is gaining momemtum at present. In the last few years many works addressed the system of generating rules for producting better rules which have varying effectiveness. Through this work, we introduced a newer method of cyclic association rules which used many features and perceptions. Tis paper Besides, this paper has generated a new kind of algorithm and which we tested in the real life system.*

Keywords: Mining Algorithm, Temporal Databases, Cyclic Mining, Cyclic Association Rules

Received: 18 September 2011, Revised 3 December 2011, Accepted 12 December 2011

1. Introduction

Data mining with cyclic association rules in temporal databases in the form of rules from data is characterized by regular variation over time. Normally when we generate the Cyclic Association Rules from monthly sales data, we find the seasonal variation as certain rules are true, approximately, the same month each year. For example, such a cyclic rule that can be generated is *"A pharmaceutical company sells both products Astradol and Clarid cyclically every month."*

Unlike sequential association rules, the CAR express intensively the time factor. Indeed, the sequential rules which highlight correlations between events according to their chronological appearance. In fact, the sequential rules which express correlations between the events according to their chronological appearance bring only a chronological order and do not point out any other temporality aspect. Nevertheless, the cyclic patterns occur simultaneously and are repeated in each regular cycle. Intensive studies have been made about mining cyclic association rules drawn from datasets [1], [3], [5], [6], [8], [12], [13]. However, these researches are limited to consider only a single dimension to generate cyclic rules. Most of work considered it as the dimension *product*. More advanced proposals deal with the combination of several dimensions. For example, we extract a cyclic rule such as *"A company manufacturing a product such as Astradol, records a sales turnover bracket ranging from 50000 to 90000. Monthly, this fact reoccurs"*. Not only this rule combines two dimensions *turnover bracket* and *Product*, but also it combines them over time (the product *Astradol* once manufactured, will be sold with already predicted turnover bracket).

As far as we know, no algorithm for mining cyclic association rules from several dimensions is proposed. Our main claim, in this paper, is to generalize the multidimensionality so that the derived cyclic patterns combine several dimensions over time. In addition, we provide a comprehensive framework for the cyclic multidimensional pattern extraction. In this paper, we present a critical survey of dedicated literature. In light of the identified shortcomings, we propose the basic concepts of our contribution and introduce the dedicated algorithm RACYM.

The remainder of the paper is organized as follows. The section 2 introduces a motivating example illustrating our contribution. In section 3, we present a survey of some related work. We briefly sketch the foundations of our method in section 4. We describe our algorithm RACYM to mine multidimensional cyclic association rules in section 5. Results of experiments carried out on a real data warehouse are reported in section 6. The last section concludes the paper and points out future directions to follow.

2. Motivating Example

Throughout this paper, we use an example from the pharmaceutical area to illustrate our proposal. We consider a table *T* where are stored the sales of a Tunisian company specialized in pharmaceutical production. As shown in the table I, we assume that *T* is defined over six dimensions, namely *D the date of the sales*, *C the city* where the sales took Place (Tunis, Sfax, Nabeul), *P the product* (let us consider both products *Astradol* and *Clarid*) as well as the discretized measures considered also as dimensions, *i.e., the sold quantity bracket Q* of the product *P*, *the internal turnover bracket IT* and *the external turnover bracket ET*.

D	*C*	*P*	*Q*	*IT*	*ET*
1^{st} quarter 2007	Tunis	*Astradol*	{2500-3000}	{75000-80000}	{500-1000}
2^{nd} quarter 2007	Tunis	*Astradol*	{2500-3000}	{75000-80000}	{500-1000}
3^{rd} quarter 2007	Tunis	*Astradol*	{2500-3000}	{75000-80000}	{1500-1000}
4^{th} quarter 2007	Tunis	*Astradol*	{2500-3000}	{75000-80000}	{500-1000}
1^{st} quarter 2008	Sfax	*Clarid*	{3000-3500}	{80000-85000}	{500-1000}
2^{nd} quarter 2008	Sfax	*Clarid*	{3000-3500}	{80000-85000}	{1500-1000}

Table 1. TABLE T

For example, the first tuple in the table *T* refers to the first quarter when the sold quantity bracket of *Astradol* ranging between 2500 and 3000 units is sold internally or exported abroad. Therefore, we can mine multidimensional association rules drawn from the quarterly internal turnover of *Astradol* manufactured in Tunis in known quantity bracket. An example of multidimensional cyclic itemset with a length of cycle equal to a quarter is (*Astradol*, Tunis, {75000-80000}). Indeed, this occurrence appears in the first, the second, the third and the fourth tuple of the table relating respectively to the first, the second, the third and the fourth quarter of 2007.

3. Related Work

In this section, we focus on the various research work closely related to the domain of cyclic pattern extraction and multidimensional association rules mining.

3.1 Cyclic patterns

The extraction of the CAR is a major problem in data mining. It was introduced by Odzen *et al.* [8]. It involves the discovery of association rules characterized by regular cyclic variation over time. Similarly, these association rules can highlight the daily regular variation, weekly, quarterly, or annual that is cyclical in nature. Discovering such regularities in the behavior of association rules allow marketers, for example, to better identify sales trends and ensure a better prediction of future requests. The transactional data for analysis are time-stamped and the time intervals are specified by user to split the data into disjoint segments. Generally, users opt for a "natural" segmentation of data based months, weeks, days,...etc thanks to their comprehension of data, the users are the more privileged to make such decision.

We present briefly the basic concepts related to cyclic patterns. The databases, from which are drawn the cyclic patterns, contain three data closely related to the market basket analysis: the first is the client identifier customerID, the second is the list of products and the third represents the date when this customer bought this product. The database is composed of itemsets identified by date and customerID. A *cycle* is a period of time characterized by its length [8](*a quarter for example*). The database is therefore considered as a set of cycles in respect of a user-specified length of cycle. An item is *cyclic* if the assigned attribute to product is repeated cyclically according to the length of cycle [8], (*Astradol* or *Clarid* are cyclic items because they respectively occurred every quarter of 2007 and 2008). Cyclic itemset is a set of cyclic items. For example (*Astradol*, *Clarid*) is a cyclic itemset if it occurs at the first and the second quarter of 2007. A cyclic itemset is frequent if it is bought a number of times greater than a threshold in respect of the length of cycle considered [8].

The crucial challenge of cyclic association rules mining algorithms is the best extraction of the frequent cyclic patterns. Several algorithms were proposed such as INTERLEAVED and SEQUENTIAL introduced by Odzen *et al.* [8] or MTP presented by Thuan [12], [13] or the Chiang's method to combine cyclic and sequential patterns [3] or PCAR, proposed by *Ben Ahmed et al.* [1]. These propositions rely on *generate and prune paradigm* where candidates are generated then infrequent ones are pruned.

3.2 Multidimensional association rules

Combining several dimensions of analysis to extract knowledge allows a fitting description of data. Since its introduction by Kamber *et al.*, we can witness the presence of several methods to mine association rules from data warehouses [7]. In this

Method	Temporality			Multidimensionality			Hierarchy		Constraint	
	Non - temporal	Sequencial	Cyclic	Intra - dimensional	Inter- dimensional	Hybrid	Single - level	Multi - level	Constraint- based	Without constraints
Kamber *et al*. [7]	x				x		x			x
Zhu [14]	x				x x	x	x			x
Odzen *et al.* [8]			x		x		x			x
Dong *et al.* [4]	x				x		x			x
Thuan [12], [13]			x		x		x		x	
Tjioe and Taniar [11]	x				x			x	x	
Plantevit *et al.* [9]		x			x		x			x
Ben Messaoud *et al.* [2]	x				x			x		x
Chiang *et al.* [3]	x	x			x		x		x	
Plantevit *et al.* [10]		x			x			x	x	
Ben Ahmed and Gouider [1]			x		x		x			x
(Our approach)			**x**		**x**		**x**		**x**	

Figure 1. Approaches of cyclic patterns and multidimensional association rules trends

respect, depending on the number of involved dimensions, three categories of rules can be distinguished [14]: (i) *intradimensional rules* generated from a single dimension, (ii) *multidimensional rules* extracted from two or more dimensions, and (iii) *hybrid rules*, drawn from multiple dimensions with repetitive predicates. In this paper, we particularly focus on multidimensional rules. To this end, we shed light on the temporality aspect while conducting a survey of the dedicated approaches. Based on this criterion, we can distinguish three types of rules: non-temporal rules, sequential rules and cyclic rules that will be detailed in what follows.

3.2.1 Non-temporal rules

Ignoring the time criteria, two types of association rules can be derived : the multi-levels association rules and the constraint-based association rules. The constraint-based mining efficiently overcomes the drawbacks of irrelevance and uselessness of extracted rules especially from multidimensional databases. Initially, Kamber *et al.* advanced this proposal [7]. Then, Dong *et al.* introduced the cubegrades which are a generalization of association rules [4]. Indeed, the latter describe how a set of measures is influenced by changing a cube through specialization (rolling down or drilling down), generalization (rolling up or drilling up) and mutation (altering one of the cube's dimensions). Cubegrades are drastically more expressive than association rules for determination of trends and patterns in data because they apply subjective aggregate measures (*i.e.*, SUM, MIN, MAX, AVG, COUNT). Unlike the other approaches, Ben Messaoud *et al.* propose a metarule based framework for mining association rules from data cubes according to the COUNT count-based aggregate [2]. As for the multidimensional multi-level rules, *Tjioe et al.* present a method for extracting association rules from multiple dimensions and several levels of abstraction by focusing on summaries of data obtained through the COUNT measure [11]. In fact, the authors propose an efficient data initialization through one of the four methods:VAVG, HAVG, WMAVG and MODUSFILTER. The output will be used to generate multilevel rules.

3.2.2 Sequential Rules

Sequential rules are combined with constraints or extended with the multi-level ones. In the context of sequential rules, Plantevit *et al.* propose a complete method to mine such multidimensional sequential patterns [9]. Furthermore, they generalize multidimensional sequential patterns where they consider patterns in which some of the dimension values may not be instantiated called *jokerized* patterns. Plantevit *et al.* extend this approach by taking into account hierarchies [10]. A multidimensional definition of sequential patterns is advanced using taxonomies for each dimension of analysis. The critical survey of the dedicated literature points out that only sequential mining methods from multidimensional context are characterized by a temporal order. To the best of our knowledge, as illustrated in figure 1, no method has been proposed to mine cyclic relationships from several dimensions. Nevertheless, the actual data are mostly defined on different dimensions and existing methods do not allow the extraction of cyclic hidden knowledge from multidimensional context. To overcome this insufficiency, we introduce a new algorithm for extracting multidimensional cyclic association rules. The main contribution of our method is to combine several dimensions of analysis when generating cyclic association rules.

4. Mining Multidimensional Cyclic Rules

In this section, we introduce the basic concepts that will be of use in the remainder.

4.1 Dimensions Partition

We consider that all is set in a multidimensional context.The three necessary data for cyclic mining drawn from classic context (Customer, Product, Date) become sets in a multidimensional context.

Thus, we consider that the table *T*, related to the data sales issued by customers, defined on a set *D* of n dimensions is partitioned into two sets: *Context dimension* D_C which concerns the investigated dimensions, and *out of context dimensions* $D_{\overline{C}}$ related to the rest of uninvestigated dimensions or the complementary dimensions.

The context dimensions can be divided into three subcategories: (i) *temporal dimension* D_T : introducing a relation of temporal order (date in classical context), (ii) *reference dimensions* D_R : the table is segmented according to the reference dimensions values (customer in classical context), and (iii) *analysis dimensions*: $D_A = \{D_1, \ldots, D_m\}$ with $D_i \subset \mathrm{Dom}(D_i)$ corresponding to products in the classic context and relative to dimensions from which the cyclic correlations will be extracted.

All reference dimensions D_R can be a conjunction of several dimensions where each dimension can have a single attribute value or a set of occurrences. For example, if one considers that the reference dimension is the city dimension *C*, this attribute can have a fixed value as Tunis or set of values such as Tunis and Sfax.

Definition 1. (Sub-cube)

Let $D' \subset D$ *be a nonempty set of p dimensions* $\{D_1, \ldots, D_p\}$ *extracted from the data cube C* ($p \leq d$). *The p-tuple* $(\delta_1, \ldots, \delta_p)$ *is called a sub-cube of data C according to D' iff* $\forall i \in \{1, \ldots, p\}$, $\delta_i \neq \phi$; *and* $\delta_i \in Dom(D_i)$.

As defined above, a sub-cube is defined by a set of dimensions *D'* extracted from the initial data cube after the determination of the context dimensions and the assignment of values to attributes included in the reference dimensions.

For example, referring to the table 1, if we define our context by ignoring the *external Turnover bracket* dimension, we fix the temporal dimension D_T to *the date* dimension in the table *T*, the reference dimension to *the city* dimension limited to Tunis and Sfax values, analysis dimensions to *the product* dimension and *internal turnover bracket* dimension. In the sub-cube, each p-tuple c = (d_1 , . . . , d_p) can be written in the form of a triple c = (r, a , t) where r, a and t are the restrictions on c respectively D_R , D_A and D_T . Thus, the sub-cube has as dimensions *the product, date, internal turnover bracket* and filters *the city* dimension according the two affected values (*i.e.*, Tunis and Sfax).

4.2 Dimensional Cyclic Item and Multidimensional Cyclic Itemset

Definition 2. (Dimensional Cyclic Item)

Let the analysis dimensions $D_A = \{D_1,..., D_m\}$ *and a cycle length l. A dimensional cyclic item* α *is an item belonging to one of the analysis dimensions, namely* D_k *and having a value of* d_k *for the date t and the date t + l such that* $\forall\, k \in [1,m]$, $d_k \in Dom(D_k)$;

Example 1. *A typical example of a dimensional cyclic item, considered in the multidimensional context, represented by the table I and the delimitation of the context considered previously, is* α = (*Astradol*) *because it belongs to the Product dimension P, being a part of analysis dimension and its value Astradol belongs to the product domain and is repeated each quarter of 2007.*

Definition 3. (Multidimensional Cyclic Itemset)

A multidimensional cyclic itemset I defined on $D_A = \{D_{i1},...,D_{im}\}$ *is a nonempty set of items* $I = \{ \alpha_i,...,\alpha_p\}$ *where* $\forall\, \boldsymbol{j} \in [1, p]$, α_j *is a dimensional cyclic item defined on* D_A *at the date t and it is repeated at each date t + l with* $\forall$ j ,k $\in [1, p]$, $\alpha_j \neq \alpha_k$.

Example 2. *An example of multidimensional cyclic itemset is I =* [(*Astradol*,{*75000-80000*},*Tunis*)] *because it is composed of three dimensional cyclic items* i.e., α_1 = (*Astradol*), α_2=({*75000-80000*}) *and* α_3=(*Tunis*). *It is repeated quarterly.*

4.3 Support and frequency of Multidimensional Cyclic Itemset

Definition 4 : (Support and frequency of Multidimensional Cyclic Itemset)

- *The support or the absolute support of multidimensional cyclic itemset I, denoted Supp*(*I*) *is the number of tuples that contain the itemset.*

$$Supp\,(I) = COUNT(I).$$

Supp(*I*) *varies between 0 and the number of tuples in the sub-cube.*

- *The relative support or the frequency of the multidimensional cyclic itemset I, denoted Freq*(*I*), *is equal to the ratio of the number of tuples that contain the itemset to the total number of tuples in the sub-cube.*

$$Freq\,(I) = \frac{COUNT(I)}{COUNT(ALL)}$$

The frequency of I is then Freq $(I) \in [\,0, 1\,]$.

Example 3. *Consider the context shown by the table I, the temporal dimension* D_T={*T*}, *the reference dimension* D_R={*Q* ={2500-3000}}, *the analysis dimensions* D_A = {*P, C, IT*} *and the length of cycle equal to a quarter.*

The multidimensional cyclic itemset I =(*Astradol*,{75000- 80000},*Tunis*) *has an absolute support related to the sales of the product Astradol in Tunis and having a local turnover bracket ranging between 75000 and 80000 by ignoring the external turnover bracket;*

$$\boldsymbol{Supp}(Astradol;\ \{75000 - 80000\};\ Tunis) = \boldsymbol{COUNT}(\boldsymbol{P} = Astradol;\ \boldsymbol{C} = Tunis;\ \boldsymbol{IT} = \{75000 - 80000\}) = 4$$

σδφφ

and a relative support equal to the ratio of the already computed support to the total sales.

$$\textbf{\textit{Freq}}(Astradol, \{75000 - 80000\}, Tunis) = \frac{\textbf{\textit{COUNT}}(\textbf{\textit{P}} = Astradol, \textbf{\textit{C}} = Tunis, \textbf{\textit{IT}} = \{75000 - 80000\})}{\textbf{\textit{COUNT}}(\textbf{\textit{P}} = ALL, \textbf{\textit{C}} = ALL, \textbf{\textit{IT}} = ALL)} = \frac{4}{6} = 0.666.$$

4.4 Rule Support and Confidence

Definition 5. : (Rule Support)

Let X and Y be two multidimensional cyclic itemsets. The rule support $R : X \Rightarrow Y$ *, denoted Supp*(*R*)*, is equal to the ratio of the number of tuples that contain X and Y to the total number of tuples in the sub-cube.*

$$Supp(R) = \frac{COUNT(X \cup Y)}{COUNT(ALL, ALL)} ;$$

The support of de R, $Supp(R) \in [0, 1]$.

Definition 6. : (Rule Confidence)

Let X and Y be two multidimensional cyclic itemsets. The rule confidence $R : X \Rightarrow Y$ *, denoted conf* (*R*)*, is equal to the ratio of the number of tuples that contain X and Y to the number of tuples that contain X in the sub-cube.*

$$conf\,(R) = \frac{Supp(R)}{Supp(X)} ;$$

The confidence of R, $conf\,(R) \in [\,0, 1\,]$.

Example 4. *In our running example, the rule R* : *Astradol*) {75000 - 80000} *has :*

- $Supp(R) = \textbf{\textit{COUNT}}(\textbf{\textit{P}} = \text{Astradol}, C = ALL, IT = \{75000 - 80000\}) = 4$
- $conf(R) = \dfrac{\textbf{\textit{COUNT}}(\textbf{\textit{P}} = \text{Astradol}, \textbf{\textit{C}} = ALL, \textbf{\textit{IT}} = \{75000 - 80000\})}{\textbf{\textit{COUNT}}(\textbf{\textit{P}} = \text{Astradol}, \textbf{\textit{C}} = ALL, \textbf{\textit{IT}} = ALL)} = \frac{4}{4} = 1$

5. Mining Multimensional Cyclic Associatoin Rules Method

Starting from a data cube, we proposethe following three steps to generate multidimensional cyclic association rules:

- Sub-cube derivation based on the user-delimitation of context and user-specification of temporal dimension, analysis and reference dimensions;
- Mining multidimensional cyclic association rules from data cube;
- Storage of generated multidimensional cyclic association rules on XML format.

In what follows, these steps are detailed.

5.1 Sub-cube derivation

Once the context dimensions is defined by the user, a determination of temporal, analysis and reference dimensions is fundamental to define the sub-cube. This operation is efficiently accomplished using the SQL query that will select the analysis dimensions and restrict the output according to the user-defined values related to the reference dimensions through the WHERE clause. Once the sub-cube was obtained, we can run our algorithm RACYM.

5.2 RACYM: Multidimensional Cyclic Association Rules

After deriving the sub-cube, RACYM algorithm takes as input the minimum threshold of support *Minsupp*, the minimum threshold of confidence *MinConf* and the length of cycle *l*.

It outputs the list of multidimensional cyclic association rules. The used notations are depicted by table II and its pseudo-code is illustrated by the algorithm in the following. In fact, RACYM, an iterative process, operates in three successive steps:

First, we proceed by an increasing level wise search for cyclic large i-itemsets, where the level (*i*) designs the number of items in the set. We denote by $C(i)$ the cyclic candidate i-itemsets potentially frequent, and $F(i)$ the cyclic frequent i-itemsets. For each

level (i), if the set $C(i)$ is nonempty, the first step of our algorithm derives the frequent cyclic patterns $F(i)$ from $C(i)$ with respect to two conditions:

- a cyclic itemset $A \in C(i)$ must be a conjunction of members from analysis dimensions;
- and a cyclic itemset must have a support above the minimum support threshold *Minsupp*.

For example, the multidimensional cyclic itemset [(*Astradol*,{75000-80000},Tunis)] is a *frequent* if its support exceeds the minimum support threshold. For an efficient extraction of frequent cyclic itemsets, we use the antimonotonicity property of the support in the multidimensional context. Indeed, any subset of a frequent cyclic set is frequent cyclic and any infrequent cyclic itemset, all its supersets will not be frequent so they will be pruned.

Notation	**Description**
SC	: Sub-cube
C_i (resp.F_i)	: Set of candidates (resp. frequent) multidimensional cyclic i-itemsets.
Minsupp	: Minimum Support Threshold
Mincon f	: Minimum Confidence Threshold
D_t	: Date t
l	: Length of cycle
Supp(C)	: Support of the multidimensional cyclic itemset C
r	: Generated cyclic rule
s	: nonempty subset s of F_i
R	: Set of the multidimensional cyclic rules

Table 2. List Of Used Notations In The Racym Algorithm

The second step uses the large cyclic i - itemsets $F(i)$ to derive a new set $C(i + 1)$ of $(i + 1)$ - candidates. $(i + 1)$ - candidate is formed by the union of two i-itemsets A and B from $F(i)$ according to three conditions:

- A and B must have (i - 1) common cyclic items;
- all cyclic sub-itemsets A U B must be instances of D_A,
- and all nonempty cyclic sub-itemsets of A U B must be frequent cyclic itemsets.

Finally, the third stage consists on scanning $F(i)$ level by level. From every $A \in F(i)$, we extract the multidimensional cyclic association rules with respect to condition, *i.e.* having a confidence above the minimum confidence threshold *MinConf*. The rule R : **P** = *Astradol* $\Rightarrow$ IT = {75000 - 80000} is a typical example of multidimensional cyclic rule if its confidence exceeds the minimum confidence threshold *MinConf*.

5.3 Storage of the multidimensional cyclic association rules

The generated multidimensional cyclic rules are stored in XML file according to the DTD shown by figure 2. The main idea behind such storage is to provide a great help for decision makers to select the best choices. Indeed, these rules conveying cyclic correlations can support the expert in many critical situations such as to efficiently decide about the quantity bracket of articles that will be periodically sold in Tunis.

6. Experimental Results

In this section, we report experiments performed on real data cube, simulating sales data of a pharmaceutical Tunisian company and containing 5000 tuples. Our sales data cube contains five dimensions, namely *product*, *city*, *shop*, *supplier*, *promotion* and four discretized dimensions, *i.e.*, *external turnover bracket*, *internal turnover bracket*, *investment bracket*, *sold quantity bracket* and *debt bracket*; obtained by discretizing measures, namely the external turnover, the internal turnover, the investment amount, the sold quantity and the amount of debt. This fact is due to inability to generate cyclic multidimensional association

rules from quantitative data. In addition, the data cube is composed of a temporal dimension, five analysis dimensions, namely *the product*, *the city*, *the local turnover bracket*, *the external turnover bracket*, *the investment bracket* and *the sold quantity bracket* selected as a reference dimension and *the debt bracket*, *the shop*, *the supplier*, *the promotion* as the ignored dimensions. All carried out experiments were conducted on a PC equipped with a 2GHz Pentium IV and 2GB of main memory running under Windows XP. The algorithm RACYM is implemented in Java. Through these experiments, we have a two-fold aim: first, we compare the runtime of RACYM according to mining based-criteria namely, the minimum support, the minimum confidence and

```
Data: SC, Minsupp, Minconf, l
Result: R: Multidimensional cyclic rules in SC
1  begin
2    F_1 = Find 1 - frequent cyclic itemsets in SC ;
3    // CandidateGeneration 3
4    for (k=2; k ≠ ϕ; k++ ) do
5      C_k = CandidatGeneration(C_k - 1);
6      if C_k is a multidimensional cyclic itemset then
7        foreach tuple T ∈ SC at date D_t do
8        C_t=subset(C_k , T );
9        foreach candidate C ∈ C_t do
10       C.count = SupportComputing(SC, l, D_t, C);
11       F_k = {C ∈ C_k, C.count > Minsupp ; }
12   F_k = ∪_k F_k
13   // RuleGeneration
14   for (i = 2; i < k; i++) do
15       Generate All nonempty subset of F_i;
16       foreach nonempty subset s of F_i do
17       r = s → (F_i-s);
18       if (confidence (r) > Minconf ) then
19         R = R ∪ r;
20   Return R;

21   Procedure SupportComputing (SC, l, D_t, C)
     Result: Supp(C)
22   begin
23       NoMoreCyclic: Boolean;
24       NoMoreCyclic = false;
25       while ((!End of tuples in SC) and (!NoMoreCyclic) do
26         C_k = CandidatGeneration(C_k - 1);
27         foreach tuple T ∈ SC at date D_t+l do
28         if C exists in T then
29           Supp(C)= Supp(C) + 1;
30         NoMoreCyclic = true;
31       Return Supp(C) ;
32   end
33 end
```

Algorithm 1: RACYM : Multidimensional Cyclic Association Rules Based on candidate generation

the length of the cycle. Second, we put the focus on the performance of our method in respect of warehousing criteria, *i.e.*, the number of dimensions (without discretization) and the number of intervals of dimensions obtained by discretizing measures.

```
<?xml version=”1.0" encoding= ”UTF-8"?>
<!ELEMENT ruleSet (rule+)>
<rule (support,confidence,cycle,premise,conclusion)
<!ELEMENT support (#PCDATA)>
<!ELEMENT confidence (#PCDATA)>
<!ELEMENT cycle (#PCDATA)>
<!ELEMENT premise (item+)>
<!ELEMENT conclusion (item+)>
<!ELEMENT item EMPTY>
<!ATTLIST item dimension CDATA #REQUIRED >
```

Figure 2. The DTD of output XML file

6.1 Mining based-criteria

The figure (3.a) plots the runtime of our algorithm RACYM when support changes according to several minimum support thresholds and several minimum confidence thresholds. Generally, the runtime of our algorithm decreases when the minimum support obviously increases. Indeed, the antimonotonicity property allows from the first iteration of the algorithm for high thresholds of minimum support, prunning significantly infrequent cyclic items. In addition, high levels of minimum confidence leads to an absolute decrease on the runtime of the algorithm. Similarly, the minimum confidence dramatically influences the performance of the algorithm. According to the figure (3.b), the shorter is the length of cycle, the more performant is our algorithm. Indeed, for a length of cycle equal to a half-year, RACYM requires the double of runtime compared with a length of cycle equal to 2 years with a minimum support = 20 %. This fact can be explained by the number of scans for a length of cycle equal to a half-year is 4 times larger than the number of scans for a length of cycle equal to 2 years.

6.2 Warehousing based-criteria

The figures (3.c, 3.d) summarize the performance evaluation of our algorithm performed on data cubes with different volumes depending on the minimum support threshold. Each data cube is characterized by the number of dimensions that it contains. Figure (3.c) shows the experiments under variation of the number of dimensions which do not undergo any discretization while figure (3.d) shows the experiments under variation of the upper and lower bounds of intervals of dimensions obtained by discretizing measures.

The analysis of the figures show that the higher is the number of dimensions, the more needed runtime to extract multidimensional cyclic association rules. We notice that for large values of minimum support from 40 %, the number of dimensions deeply influences the performance of our algorithm. Similarly to the variation of number of intervals of dimensions obtained by discretizing measures, the introduction of a minimum threshold of support exceeding 40 % leads to an important reduction in the performance of our algorithm. However, for large values of minimum threshold of support, the number of dimensions respectively the number of intervals of dimensions (the discretized measures) has almost a slight in- fluence on the runtime of our algorithm. Indeed, the greater is the number of intervals of those dimensions, the more needed runtime to generate candidates from large multidimensional cyclic itemsets then to compute their support and finally to extract the related rules.

In respect of the collected results, it is noteworthy that: (i) The efficiency of the association rules mining is strictly dependent of the mining based-criteria, namely the minimum support threshold of support, the minimum confidence threshold and the length of cycle. (ii) The performance of our method is fundamentally related the size of the data cube. The latter is described through the dimension table (axes of analysis).

7. Conclusion

In this paper, we proposed a new method to extract cyclic association rules from multidimensional context such as *“A pharmaceutical company sells a product (*i.e. Astradol, *with a total turnover bracket ranging between 50000 and 90000 every month”*. Thus, a new definition of multidimensional cyclic patterns is provided and a new algorithm, called RACYM to

extract such patterns is introduced. The carried out experimental results showed that the performance of our algorithm is closely related to the data cube size, *i.e.*, dimensions as well as the mining based criteria, namely the minimum support threshold, the minimum confidence threshold and the cycle length.

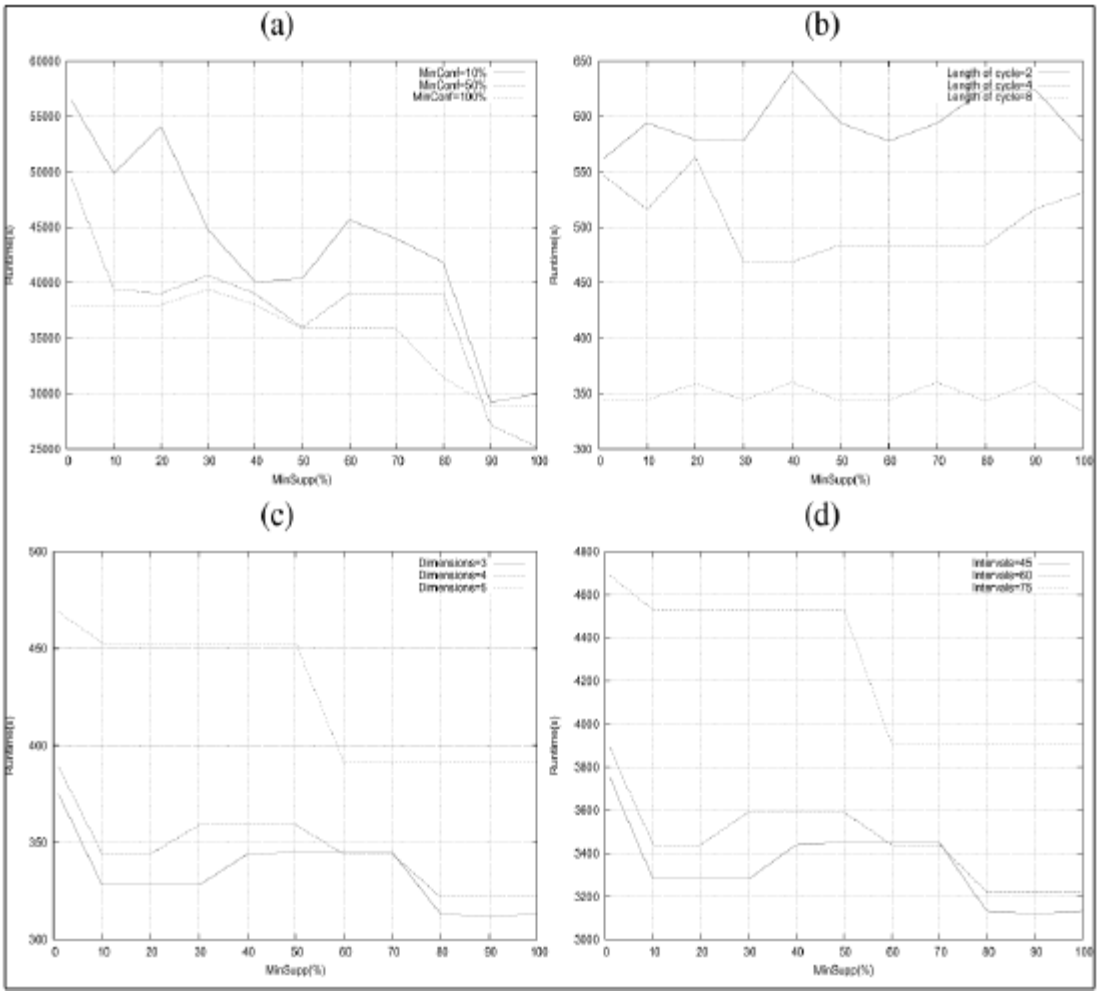

Figure 3. The running times of our algorithm according to the (a) minimum confidence, (b) length of cycle, (c) number of dimensions, (d) number of intervals of dimensions obtained by discretizing measures.

Other avenues for future work mainly address the following issues: (i) study of the relevance of the generated multidimensional cyclic association rules using condensed representations, (ii) extraction of multidimensional cyclic association rules under the convergence and divergence of cycles, (iii) consideration of the personalized dimension hierarchies during multidimensional cyclic association rules mining.

References

[1] Ben, E., Ahmed ., Gouider, M. S.(2010). *Towards a new mechanism of extracting cyclic association rules based on partition aspect*, RCIS, 69 - 78, IEEE.,

[2] Ben Messaoud, R., Boussaid, O., Rabasda, S. L., Missaoui, *R.* (2006). *Enhanced mining of association rules from data cubes, In*: Proceedings of the 9 th ACM International Workshop on Data Warehousing and OLAP (DOLAP'06), 11-18.

[3] Chiang, D., Wang, C., Chen, S., Chen, C. (2009). *The Cyclic Model Analysis on Sequential Patterns, IEEE Trans. on Knowl. and Data Eng*., 21 (11) November 1617-1628,USA.

[4] Dong, G., Han, J., Lam, J., Pei, J.,Wang, K., Zou, W. (2004). *Mining Constrained Gradients in Large Databases, IEEE Trans. on Knowl. and Data Eng.,* 16.

[5] Han, J., Gong, W., Yin, Y. (1998). Mining Segment-Wise Periodic Patterns in Time-Related Databases, KDD, 214-218.

[6] Han, J., Gong, W., Yin, Y. (1999). Efficient Mining of Partial Periodic Patterns in Time Series Database, ICDE, 106-115.

[7] Kamber, M., Han, M., Chiang, J, Y. (1997). *Metarule-guided mining of multidimensional association rules using data cubes*, KDD'97, Newport Beach, CA, 207-210.

[8] Ozden, B., Ramaswamy, S., Silberschatz, A. (1998). *Cyclic Association Rules*, ICDE, USA.

[9] Plantevit, M., Choong, Y. W., Laurent, A., Laurent, D., Teisseire, M. (2005). *M2SP: Mining Sequential Patterns Among Several Dimensions*, 2005, PKDD, LNCS.

[10] Plantevit, M., Laurent, A., Laurent, D., Teisseire, M, Choong, Y. (2004). *Mining multidimensional and multilevel sequential patterns*, ACM Transactions on Knowledge Discovery from Data, 4, p. 4-37.

[11] Tjioe, H. C., Taniar, D. (2005). Mining Association Rules in Data Warehouses, *IJDWM*, 1, 3, 28-62,

[12] Thuan, N. D. (2004). *Mining Cyclic Association Rules in Temporal Database , The Journal Science and Technology Developement*, Vietnam, 7, 8, Springer Netherlands, 12-19.

[13] Thuan, N. D. (2008). Mining Time Pattern Association Rules in Temporal Database, *SCSS* (1) 7-11,

[14] Zhu, H. (1998). On-line analytical mining of association rules, Master's thesis, Simon Fraser University, Burnaby, British Columbia, Canada.