

# Usage Des Mesures Pour La Génération Des Règles d'Associations Cycliques

Eya Ben Ahmed\*, Ahlem Nabli\*\*  
Faïez Gargouri\*\*\*

\* Université de Tunis, Institut Supérieur de Gestion de Tunis  
eya.benahmed@gmail.com

\*\*Université de Sfax, Faculté des Sciences de Sfax  
ahlem.nabli@fsegs.rnu.tn

\*\*\*Université de Sfax, Institut Supérieur d'Informatique et de Multimédia de Sfax  
faiez.gargouri@isimsf.rnu.tn

**Résumé.** L'analyse en ligne (OLAP) fournit aux utilisateurs une navigation interactive des données multidimensionnelles. Cependant, aucun moyen pour expliquer les corrélations existantes entre les données n'est offert. Ainsi, le couplage de l'OLAP et de la fouille de données, plus particulièrement les règles d'association a efficacement apporté une solution satisfaisante à ce problème.

Dans ce cadre, on s'intéresse à une classe particulière de règles d'association qui est les règles d'association cycliques. Ces règles visent la découverte de modèles qui se reproduisent périodiquement à des intervalles réguliers définis par l'utilisateur. Généralement, les motifs générés ne prennent pas en considération les spécificités du contexte multidimensionnel à savoir, l'existence des mesures et leurs agrégations. Dans cet article, nous proposons une méthode d'extraction de règles d'association cycliques à partir de mesures et nous redéfinissons les métriques d'évaluation de la qualité de ces motifs en s'inspirant de l'additivité temporelle des mesures à travers l'intégration des fonctions d'agrégation appropriées. Pour prouver l'utilité de notre approche, nous menons une étude empirique sur un entrepôt de données réel.

## 1 Introduction

Dans le modèle multidimensionnel, OLAP offre des capacités de modélisation analytique. En effet, il peut générer des résumés, des agrégations, et des hiérarchies à chaque niveau de granularité de dimension. Cependant, ces fonctionnalités sont limitées à une tâche interactive exploratoire. Bien que l'utilisateur ne peut pas découvrir des modèles cachés ou expliquer de telles régularités à partir de grandes quantités de données, le couplage des techniques de fouille de données et de la technologie OLAP peut efficacement assurer ces tâches. Ainsi, l'utilisateur pourra souvent naviguer dans l'entrepôt de données, sélectionner des portions de données pertinentes, les analyser à des granularités différentes, et de fournir des connaissances sous plusieurs formes (Kamber et *al.*, 1997). Parmi les algorithmes de fouilles de données employés

dans ce cadre, on s'intéresse aux règles d'association et plus particulièrement aux règles d'associations cycliques (RAC). En fait, les règles d'association cycliques sont définies comme des règles qui surviennent à des intervalles réguliers définis par l'utilisateur dans la base de données. Par exemple, considérons le contexte donné par la table 1 où sont stockées les transactions trimestrielles d'une société tunisienne spécialisée dans la production pharmaceutique. Un exemple de règle d'association cyclique extraite est " *Trimestriellement, la gamme sous licence est vendue en Jordanie*". Comparé aux travaux consacrés à l'aspect algorithmique pour l'extraction efficace des règles d'association cycliques [(Ozden et al., 1998), (Han et al., 1999), (Thuan, 2004, 2008), (Ben Ahmed et Gouider, 2010)], seule l'étude de (Ben Ahmed et Gargouri, 2010) a abordé la génération des règles d'association cycliques à partir de plusieurs dimensions mais ce travail n'a pas tiré profit des spécificités structurelles du cube de données, à savoir l'existence des mesures et lors prise en compte pour générer des règles d'association cycliques à partir de mesures dans un contexte multi-dimensionnel.

Par contre, les règles d'association basées sur les mesures ont été initialement introduites en utilisant la fonction COUNT (Han et al., 1999) où les valeurs globales exigées pour découvrir des règles d'association sont pré-calculées et sauvegardées dans une cellule COUNT dans le cube de données. Le calcul du support et de la confiance est inspiré de ces valeurs dans ces cellules. Une autre contribution propose une redéfinition de ces métriques en se basant sur la fonction d'agrégation SUM (Ben Messaoud et al., 2006). Néanmoins, dans le processus d'analyse, les utilisateurs se concentrent généralement sur des données multidimensionnelles et leurs associations selon des mesures plus élaborées que des fréquences ou des simples additions. Ainsi, la restriction des fonctions globales considérées au SUM et COUNT limite les mesures considérées qui peuvent être sommables par d'autres fonctions et néglige l'évaluation significative des motifs générés. En fait, certaines mesures ne peuvent pas être agrégées en utilisant l'addition à travers aucune dimension. De telles mesures non-additives temporellement sont fréquentes et renferment de la connaissance cachée que les experts peuvent extraire sous forme de règles à partir d'un tel contexte (Plantevit et al., 2008). Pour mieux illustrer cette problématique, nous proposons un exemple motivant notre apport.

## 2 Exemple et motivations

Nous supposons que la table  $\mathcal{T}$  illustrée par la table 1, est définie par une dimension temporelle, à savoir : le *Temps T* des transactions, et les trois mesures, à savoir : La *Quantité vendue Q*, le *Taux de Change TC*, le *Chiffre d'Affaires Externe CAE*. Le premier tuple dans la table  $\mathcal{T}$  se rapporte au premier tuple quand la *quantité vendue* atteint 77520 unités et elle est exportée avec un *chiffre d'affaires externe* égal à 750 selon un *taux de change* qui atteint 1.948%.

Temps <i>T</i>	Quantité vendue <i>Q</i>	Taux de change <i>TC</i>	Chiffre d'Affaires Externe <i>CAE</i>
Premier trimestre 2007	77520	1,948%	750
Second trimestre 2007	75000	1,947%	550
Troisième trimestre 2007	77520	1,948%	750
Quatrième trimestre 2007	80000	1,946%	900
Premier trimestre 2008	71100	1,382%	450
Second trimestre 2008	80000	1,382%	800

TAB. 1 – Table  $\mathcal{T}$

Cependant, la *quantité vendue* ne peut pas être sommée selon la dimension de *temps*, pareillement, le *taux de change* n'est pas sommable sur aucune dimension, et le *chiffre d'affaires externe* ne peut pas être expressivement agrégé par l'addition sur la dimension temporelle. Nous visons l'extraction des règles d'association cycliques extraites à partir de ce contexte multidimensionnel en se basant sur le concept d'additivité par rapport au temps caractérisant les mesures du cube de données lors de l'évaluation des patterns générés. Par exemple, nous visons la génération de règles d'association combinant plusieurs mesures selon le temps telles que " *Chaque trimestre impair, la quantité vendue atteint 77250 unités est vendue avec un chiffre d'affaires externe arrivant à 750 selon un taux de change égal à 1.948%* ". Les métriques d'évaluation d'une telle règle sont basées sur les fonctions d'agrégation applicables sur les mesures par rapport au temps.

A notre connaissance, peu de méthodes permettent l'extraction de motifs cycliques à partir du contexte multidimensionnel contenant des mesures. L'originalité de notre approche est la considération de telles mesures lors de l'extraction des motifs et la prise en compte du fait qu'elles ne peuvent pas être agrégées en utilisant la fonction COUNT ou SUM par rapport au temps. Par conséquent, nous redéfinissons les concepts de base de cette fouille articulée sur les mesures ainsi que les métriques d'évaluation des règles par le biais des fonctions d'agrégations appropriées.

Cet article est organisé comme suit. Dans la section 3, nous présentons un aperçu sur la discrétisation des attributs quantitatifs, particulièrement dans le contexte multidimensionnel. Puis, nous présentons les motifs cycliques. Ensuite, nous passons en revue les travaux antérieurs dédiés à l'extraction de règles multidimensionnelles. Nous clôturons cette section par la présentation du concept d'additivité de mesure par rapport au temps. Dans la section 4, nous introduisons nos propres concepts de base. La section 5 décrit notre contribution. La section 6 détaille le fonctionnement de notre algorithme. Dans la section 7, nous menons une étude empirique qui prouve l'utilité de notre algorithme. La conclusion et les perspectives font l'objet de la section 8.

### 3 Panorama des travaux antérieurs

Dans cette section, nous présentons un survol des techniques permettant la discrétisation des attributs quantitatifs. Puis, nous nous intéressons au processus d'extraction en présentant un aperçu de la littérature dédiée aux règles d'association multidimensionnelles. Enfin, nous introduisons le concept d'additivité temporelle des mesures.

#### 3.1 Discrétisation des attributs quantitatifs

De nombreuses techniques de discrétisation sont utilisées pour transformer les attributs quantitatifs en des attributs qualitatifs (catégoriques) Liu et al. (2002), à savoir, (i) *Largeur égale* qui divise les valeurs possibles en  $N$  intervalles ; (ii) *Fréquence égale* qui divise les valeurs possibles en intervalles de même taille  $N$  ; (iii) *Discrétisation basée sur l'entropie* qui utilise la distribution de classe dans la détermination des points de partitionnement ; (iv) *Discrétisation basée sur le clustering* pour partitionner les valeurs en groupes ; (v) *Binning* employant la technique 1R DISCRETIZER de Holte pour calculer l'erreur résultante lors du partitionnement des intervalles ; (vi) *Dépendance* à travers plusieurs techniques telles que  $\chi^2$

ou STATISTIC Liu et Setiono (1995). Ces techniques déterminent la similitude des intervalles adjacents à travers la mesure du niveau de signification.

Dans le contexte multidimensionnel, Palaniappan et Hong (2008) ont pu démontrer que les méthodes simples de discrétisation (i.e. largeur égale et fréquence égale) aboutissent à des résultats respectables et consomment significativement moins de temps.

### 3.1.1 Motifs cycliques

Plusieurs travaux de recherches sont menés sur l'extraction des RAC. Initialement, ce problème est abordé par Ozden et al. (1998). Il s'agit de l'extraction de règles d'association à partir d'articles caractérisés par leur variation régulière au fil du temps. En effet, ces règles d'association peuvent mettre en exergue la variation régulière naturellement cyclique, qui peut être quotidienne, hebdomadaire, trimestrielle ou annuelle. La découverte de telles régularités dans le comportement des règles d'association permettent d'améliorer, par exemple, les services du marketing à travers une meilleure identification des tendances de ventes et une prédiction pertinente de futures demandes. Les données dans l'analyse transactionnelle sont horodatés et les intervalles de temps sont spécifiés par l'utilisateur divisant les données en segments disjoints, les utilisateurs optent pour une segmentation des données "naturelle" basée sur les mois, semaines, jours, etc. En effet, les utilisateurs sont les plus capables de déterminer telle valeur vue leur compréhension des données.

Nous présentons brièvement les concepts de base liés aux motifs cycliques. Un *cycle* est une période dans le temps caractérisée par sa longueur (un trimestre dans notre cas). La base de données est donc considérée comme un ensemble de cycles de longueur fixe désignée par l'utilisateur. Un *item cyclique* est une valeur affectée à l'attribut produit et qui se répète cycliquement selon la longueur du cycle (*Gamme sous licence* est vendue chaque trimestre de 2007). Un *itemset cyclique* est un ensemble d'items cycliques. Par exemple (*Gamme sous licence, modèle hôpital*) est un itemset cyclique s'il apparaît durant le premier et le second trimestre de 2007. Un *item cyclique* est *fréquent*, dans notre cas d'étude, si cet item a été acheté un nombre de fois supérieur à un seuil prédéfini pour la période de cycle considérée. Le défi des algorithmes de fouille de données cycliques est l'extraction la plus efficace des motifs cycliques fréquents. Plusieurs algorithmes ont été alors proposés tels que INTERLEAVED et SÉQUENTIEL introduits par Ozden et al. (1998) ou MTP présenté par Thuan (2004, 2008) ou la méthode de Chiang qui combine les motifs cycliques et séquentiels Chiang et al. (2009), ou encore PCAR proposé par Ben Ahmed et Gouider (2010). Ces propositions reposent sur le paradigme *Générer/Elaguer* où les candidats sont générés puis les non fréquents sont élagués.

## 3.2 Dérivation des règles d'association multidimensionnelles

Dans cette sous-section, on présente un survol de l'état de l'art. Plusieurs travaux ont généré des règles sans prise en compte de mesure lors de l'extraction des motifs multidimensionnels. Tout d'abord, (Kamber et al., 1997) ont introduit deux algorithmes pour l'extraction des règles d'association à partir d'un cube de données sous contraintes<sup>1</sup>.

---

1. Elles s'articulent sur la formalisation des attentes des experts à travers la spécification du type de la connaissance, ou des contraintes sur les données ou des contraintes d'intérêts ou des contraintes sur les règles.

Méthode	Temporalité			Hiérarchie		Contraintes		Evaluation de la règle		
	Non temporelles	Séquentielles	Cycliques	uni-niveau	Multi-niveaux	Sous contraintes	Sans contraintes	COUNT	SUM	Toutes les fonctions
(Kamber et al.,1997)	x			x			x	x		
(Zhu,1998)	x			x			x	x		
(Odzen et al.,1998)		x		x		x		x		
(Nestorov et Jukic,2003)	x				x	x		x		
(Thuan,2004,2008)		x		x		x		x		
(Tjioe et Taniar,2005)	x				x	x		x		
(Plantevit et al.,2005)		x		x		x		x		
(Ben Messaoud et al.,2006)	x			x			x		x	
(Chiang et al.,2009)		x	x	x		x		x		
(Plantevit et al.,2009)		x		x		x		x		
(Plantevit et al.,2010)		x			x	x		x		
(Ben Ahmed et Gouider,2010)			x	x			x	x		
(Ben Ahmed et Gargouri,2010)				x		x		x		
<b>(Notre approche, 2011)</b>		x			x		x			x

FIG. 1 – Comparaison des approches dédiées aux RAC et aux règles d'association multidimensionnelles.

Une autre proposition<sup>2</sup> présentée par (Zhu, 1998) a validé le même choix. (Nestorov et Jukic, 2003) ont proposé la dérivation des règles d'association des entrepôts de données en utilisant les requêtes. (Tjioe et Taniar, 2005) génèrent les associations à partir de plusieurs dimensions en se focalisant sur les données résumées. (Plantevit et al., 2005) proposent une nouvelle définition des motifs séquentiels multidimensionnels. (Ben Messaoud et al., 2006) proposent d'utiliser la mesure afin de calculer le support et la confiance des règles d'association multidimensionnelles. (Plantevit et al., 2010) dérivent des motifs séquentiels multi-niveaux<sup>3</sup>. Dans la même optique, (Ben Ahmed et Gargouri, 2010) ont introduit la génération de motifs cycliques multidimensionnels à partir de plusieurs dimensions.

Toutes ces approches génèrent des règles multidimensionnelles exclusivement à partir des dimensions sans aucune considération des mesures lors de l'extraction des motifs. Seule la contribution de (Plantevit et al., 2008) initie au problème d'extraction de règles contenant des mesures à partir du cube. Comme le montre la figure 1, les méthodes proposées sont limitées à l'utilisation des mesures dans le cadre de l'évaluation des règles d'associations en employant les fonctions COUNT ou SUM. Néanmoins, dans des situations réelles, les mesures peuvent être corrélées et le cumul de certains attributs peut être sémantiquement non significatif par rapport à la dimension temporelle. Pour cette raison, l'utilisation exclusive et limitée aux deux fonctions COUNT et SUM est la plupart du temps inadéquate vu qu'elle induit à la génération de règles incorrectement évaluées.

2. Classifiant les règles en intra-dimensionnelles (générées à partir d'une seule dimension.), inter-dimensionnelles (extraites à partir de deux dimensions ou plus) et hybrides (extraites à partir de plusieurs dimensions avec des prédicats répétitifs).

3. Vue la difficulté de trouver des associations robustes à partir de données élémentaires, les règles multi-niveaux permettent de regrouper les données élémentaires en classes plus générales.

### 3.3 Additivité temporelle des mesures

Lenz et Shoshani (1997) distinguent trois types de mesures : mesure de type flot, mesure de type stock et mesure de type unité. Cette catégorisation détermine les opérateurs d'agrégation utilisables pour chaque mesure par rapport au temps.

(i) *Mesure additive* (de type flot) : Ses valeurs peuvent être combinées de façon significative pour n'importe quelle dimension. Par exemple, il est judicieux de sommer le total des ventes ou le chiffre d'affaires externe par rapport au temps, cela ne provoquera pas de chevauchement entre les phénomènes du monde réel que peuvent engendrer les valeurs individuelles.

(ii) *Mesure semi-additive* (de type stock) : Ses valeurs ne peuvent être sommées sur une ou plusieurs des dimensions, le plus souvent la dimension temporelle. Généralement elles sont alliées à des faits instantanés. Par exemple, la sommation de l'inventaire d'une gamme spécifique telle que la gamme sous licence n'a pas de signification car les valeurs seront comptées plusieurs fois, mais il est utile de résumer les niveaux d'inventaire des différentes gammes.

(iii) *Mesure non additive* (de type unité) : Ses valeurs ne peuvent pas être sommées sur aucune dimension. Des exemples typiques sont les prix des articles ou des gammes et le coût par unité d'une gamme telle que le modèle hôpital.

Dans cet article, nous utilisons la notion d'additivité des mesures par rapport aux hiérarchies temporelles proposée par Lenz et Shoshani (1997) pour introduire un processus de fouille de règles d'associations cycliques doté d'une évaluation spécifique de chaque mesure dépendamment des fonctions d'agrégation qui lui sont appropriées. Pour ce faire, un algorithme inspiré de l'Apriori Agrawal et al. est introduit tout en intégrant les mesures dans les règles cycliques extraites des cubes en un premier volet. L'intégration de l'additivité temporelle des mesures lors du calcul des métriques d'évaluation des règles, à savoir le support et la confiance des règles d'association est la contribution majeure dans un second volet.

## 4 Formalisme et notions de base

Dans cette section, nous proposons nos concepts de base inspirés des travaux de Plantevit et al. (2005).

### 4.1 Partition du Contexte

Nous considérons que tout est ensemble dans un contexte multidimensionnel. Les trois données nécessaires pour l'extraction de motifs cycliques dans un contexte classique (Client, Produit, Date) deviennent dans un contexte multidimensionnel des ensembles.

Considérons une table relationnelle  $\mathcal{T}$ , où les transactions effectuées par les clients  $y$  sont stockées, définie sur une dimension temporelle et un ensemble  $M$  de  $m$  mesures. La table est partitionnée en deux ensembles :

1. *Contexte endogène* composé d'une dimension temporelle  $\mathcal{D}_T$  et des mesures de contexte  $\mathcal{M}_C$  incluses dans le processus d'analyse ;
2. *Contexte Exogène* lié à toutes les dimensions et les mesures exclues de l'analyse.

Nous nous concentrons dans ce qui suit sur le contexte endogène.

#### 4.1.1 Mesures de contexte

Les mesures de contexte  $\mathcal{M}_C$  peuvent être partitionnées en deux sous-catégories :

- (i) *Mesure de Référence*  $\mathcal{M}_R$  (table est partitionnée en fonction des valeurs affectées aux mesures de référence) ;
- (ii) *Mesure d'analyse*  $\mathcal{M}_A$  (tuples contenant les valeurs de mesure qui seront exploitées pour extraire des tendances cycliques).

Les mesures d'analyse peuvent être divisées en quatre sous-catégories :

- *Ensemble de mesures additives*  $\mathcal{M}_{SUM}$  lorsque la fonction SUM peut être utilisée pour l'agrégation des données par rapport au temps ;
- *Ensemble de mesures maximales*  $\mathcal{M}_{MAX}$  lorsque la fonction MAX peut être appliquée pour agréger les données par rapport au temps ;
- *Ensemble de mesures minimales*  $\mathcal{M}_{MIN}$  lorsque la fonction d'agrégation MIN peut être utilisée en se référant au temps ;
- *Ensemble de mesures moyennes*  $\mathcal{M}_{AVG}$  lorsque la fonction d'agrégation AVG peut être appliquée par rapport au temps.

Dans notre exemple, nous considérons la table 1 comme étant notre contexte endogène composé de :

1. *Dimension temporelle*  $\mathcal{D}_T = \{T\}$  ;
2. *Mesures de contexte*  $\mathcal{M}_C = \{Q, TC, CAE\}$  avec les mesures d'analyse égales à toutes les mesures existantes et aucune mesure de référence n'est limitée à une valeur fixe de mesure avec  $\mathcal{M}_{MAX} = \{Q\}$  et  $\mathcal{M}_{MIN} = \{CAE\}$  et  $\mathcal{M}_{AVG} = \{TC\}$ .

#### Définition 1. (Sous-cube Endogène)

Soit  $\mathcal{D}' \subset \mathcal{D}$  un ensemble non vide contenant une dimension temporelle  $\{\mathcal{D}_1\}$  extraite du cube de données  $C$  ( $1 \leq d$ ) avec  $d$  la cardinalité de toutes les dimensions et  $\mathcal{M}' \subset \mathcal{M}$  un ensemble non vide de  $q$  mesures  $\{\mathcal{M}_1, \dots, \mathcal{M}_q\}$  extrait du cube de données  $C$  ( $q \leq m$ ) avec  $m$  la cardinalité de toutes les mesures.

L'e-tuple  $(\delta_1, \dots, \delta_i)$  est appelé un sous-cube endogène de données  $C$  avec :

- $\mathcal{D}'$  with  $i = 1, \delta_i \neq \emptyset$  et  $\delta_i \in \text{Dom}(\mathcal{D}_1)$ .
- $\mathcal{M}'$  iff  $\forall j \in \{1, \dots, q\}, \gamma_j \neq \emptyset$  et  $\gamma_j \in \text{Dom}(\mathcal{M}_j)$ .

Dans notre exemple, dans le sous-cube endogène, chaque e-tuple  $e = (d_1, m_1, \dots, m_q)$  peut être écrit sous la forme de triple  $e = (r, a, t)$  avec  $r$  est la restriction sur  $\mathcal{M}_R$ ,  $a$  est la restriction sur  $\mathcal{M}_A$  et  $t$  est la restriction sur  $\mathcal{D}_T$ .

## 4.2 Item cyclique dimensionnel et Itemset cyclique multidimensionnel basés sur les mesures

#### Définition 2. (Item Cyclique Dimensionnel basé sur la mesure)

Soient la dimension temporelle  $\mathcal{D}_T$ , les mesures d'analyse  $\mathcal{M}_A = \{\mathcal{M}_1, \dots, \mathcal{M}_q\}$  et une longueur de cycle  $l$ . Un item cyclique dimensionnel basé sur la mesure  $\alpha$  est un item satisfaisant la condition suivante : il appartenant à l'une des mesures d'analyse, à savoir,  $\mathcal{M}_1$  et ayant une valeur de  $\alpha = m_1$  à la date  $t$  et obligatoirement à la date  $t + l$  tel que  $\forall l \in [1, q], m_l \in \text{Dom}(\mathcal{M}_1)$ .

Un exemple typique d'item cyclique dimensionnel basé sur la mesure, considéré dans le contexte multidimensionnel représenté par la table 3 et la délimitation de contexte considérée précédemment  $\alpha = (77500)$  car il appartient à la mesure *quantité vendue*  $Q$ , faisant partie de la mesure d'analyse et de sa valeur 77500 appartient au domaine de la *quantité vendue* et est répété chaque semestre impair de 2007.

**Définition 3. (Itemset Cyclique Multidimensionnel basé sur les mesures)**

Un itemset cyclique multidimensionnel basé sur les mesures  $I$  défini sur  $\mathcal{M}_A = \{\mathcal{M}_1, \dots, \mathcal{M}_j\}$  est un ensemble non vide d'items  $I = \{\alpha_1, \dots, \alpha_l\}$  avec  $\forall \pi \in [1, \theta]$ ,  $\alpha_\pi$  est un item cyclique dimensionnel basé sur la mesure défini sur  $\mathcal{M}_A$  et  $\forall \pi, e \in [1, \theta]$ ,  $\alpha_\pi \neq \alpha_e$ .

**Exemple 1.** Un exemple d'itemset cyclique multidimensionnel basé sur les mesures est  $I : [CAE = 600, Q = 77500]$  car il est composé de deux items cycliques dimensionnels basés sur les mesures i.e.,  $\alpha_1 = (CAE = 600)$  et  $\alpha_2 = (Q = 77500)$ . Il est répété chaque trimestre de 2007.

**4.3 Support et confiance Basés sur les fonctions d'agrégation**

**Définition 4. : (Support d'Item Cyclique Dimensionnel basé sur la mesure)**

- Le support d'un item cyclique dimensionnel basé sur la mesure  $\alpha$  est dénoté par  $Supp(\alpha)$ . Vu que  $\alpha$  appartient à la mesure d'analyse  $\mathcal{M}_1$ , le support est calculé comme suit :

1.  $\mathcal{M}_1 \in \mathcal{M}_{SUM}$ , le support de  $\alpha$  représente le quotient de la valeur de  $\alpha$  par la somme de la mesure car cette dernière est agrégée à travers une sommation par rapport au temps :

$$Supp(\alpha) = \frac{\alpha}{SUM(\mathcal{M}_1)};$$

2.  $\mathcal{M}_1 \in \mathcal{M}_{MAX}$ , le support de  $\alpha$  représente le rapport de la valeur de  $\alpha$  par le maximum de la mesure vue que cette mesure est agrégée temporellement en utilisant le maximum de la mesure :

$$Supp(\alpha) = \frac{\alpha}{MAX(\mathcal{M}_1)};$$

3.  $\mathcal{M}_1 \in \mathcal{M}_{MIN}$ , le support de  $\alpha$  représente le quotient de la différence entre la valeur de  $\alpha$  et la valeur du minimum de la mesure par la différence entre le maximum et le minimum de cette mesure courante :

$$Supp(\alpha) = \frac{\alpha - MIN(\mathcal{M}_1)}{MAX(\mathcal{M}_1) - MIN(\mathcal{M}_1)};$$

4.  $\mathcal{M}_1 \in \mathcal{M}_{AVG}$

- Si  $(\alpha \leq AVG(\mathcal{M}_1))$ , le support de  $\alpha$  représente le rapport de la valeur de  $\alpha$  par la moyenne de la mesure si la valeur de  $\alpha$  est inférieure à cette moyenne :

$$Supp(\alpha) = \frac{\alpha}{AVG(\mathcal{M}_1)};$$

- Sinon le support de  $\alpha$  représente le quotient de la différence entre la valeur de  $\alpha$  et la valeur du minimum par la différence entre le maximum et le minimum de la mesure courante :

$$Supp(\alpha) = \frac{\alpha - MIN(\mathcal{M}_1)}{MAX(\mathcal{M}_1) - MIN(\mathcal{M}_1)};$$

Dans notre exemple, se référant à la table 3, on considère les items suivants :

-  $(\alpha_1=77500) \in Q$  avec  $Q \in \mathcal{M}_{MAX}$ , son support est égal à

$$Supp(\alpha_1) = \frac{\alpha_1}{MAX(\mathcal{M}_1)} = \frac{\alpha_1}{MAX(Q)} = \frac{77500}{85000} = 0.911$$

–  $(\alpha_2=700) \in CAE$  avec  $CAE \in \mathcal{M}_{MIN}$ , son support est égal à

$$Supp(\alpha_2) = \frac{\alpha_2 - MIN(\mathcal{M}_1)}{MAX(\mathcal{M}_1) - MIN(\mathcal{M}_1)} = \frac{\alpha_2 - MIN(CAE)}{MAX(CAE) - MIN(CAE)} = \frac{700 - 600}{700 - 600} = 1;$$

–  $(\alpha_3=1,947\%) \in TC$  et  $TC \in \mathcal{M}_{AVG}$  avec  $AVG(TC)=\frac{(1,947*4)}{4} = 1,947$ .  
 $\alpha_3=1,947\% \leq AVG(TC) = 1,947$  alors  $\alpha_3$ , son support est égal à

$$Supp(\alpha_3) = \frac{\alpha_3}{AVG(TC)} = \frac{1,947}{1,947} = 1.$$

**Définition 5. : (Support d'un Itemset Cyclique Multidimensionnel basé sur les mesures)**

On considère un itemset cyclique multidimensionnel basé sur les mesures  $I = \{\alpha_1, \dots, \alpha_p\}$  composé de  $p$  mesures de  $\mathcal{M}_A$ .

Le support d'itemset cyclique multidimensionnel  $I$ , dénoté  $Supp(I)$  est calculé comme suit :

$$Supp(I) = \prod_{i=1}^p SUPP(\alpha_i);$$

**Exemple 2.** L'itemset cyclique multidimensionnel basé sur les mesures  $I=[CAE = 600, Q = 77500]$  a un support relatif aux ventes de 77500 unités avec un chiffre d'affaires externe 600

$$\begin{aligned} Supp(I) &= SUPP(CAE = 600) * SUPP(Q = 77500) \\ &= \frac{600 - MIN(CAE)}{MAX(CAE) - MIN(CAE)} * \frac{Q = 77500}{MAX(Q)} = \frac{700 - 600}{700 - 600} * \frac{77500}{85000} = 1 * 0.911 = 0.911. \end{aligned}$$

#### 4.4 Support et confiance des règles

**Définition 6. : (Support de Règle )**

- Le support de la règle  $\mathcal{R} : X \Rightarrow Y$ , dénoté  $Supp(\mathcal{R})$ , est égal au quotient du nombre de tuples contenant  $X$  et  $Y$  par le nombre total des tuples dans le sous-cube.

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

Le support de  $\mathcal{R}$ ,  $Supp(\mathcal{R}) \in [0, 1]$ .

**Définition 7. : (Confiance de Règle)**

- La confiance d'une règle  $\mathcal{R} : X \Rightarrow Y$ , dénotée  $Conf(\mathcal{R})$ , est égale au quotient du nombre des tuples contenant  $X$  et  $Y$  par le nombre de tuples contenant  $X$  dans le sous-cube.

$$Conf(\mathcal{R}) = \frac{Supp(\mathcal{R})}{Supp(X)};$$

La confiance de  $\mathcal{R}$ ,  $Conf(\mathcal{R}) \in [0, 1]$ .

Dans notre exemple, la règle  $\mathcal{R} : CAE = 600 \Rightarrow Q = 77500$  a :

$$Supp(\mathcal{R}) = Supp(CAE = 600 \cup Q = 77500) = SUPP(600) * SUPP(Q = 77500)$$

$$= \frac{700 - 600}{700 - 600} * \frac{77500}{MAX(Q)} = \frac{700 - 600}{700 - 600} * \frac{77500}{85000} = 1 * 0.911 = 0.911.$$

$$Conf(\mathcal{R}) = \frac{Supp(\mathcal{R})}{Supp(Y)} = \frac{Supp(\mathcal{R})}{Supp(CAE=600)} = \frac{0.911}{1} = 0.911.$$

## 5 Génération des règles d'association cycliques multidimensionnelles basées sur les mesures

Pour dériver des règles d'association cycliques multidimensionnelles basées sur les mesures, nous proposons le processus d'extraction articulé sur les deux phases suivantes :

- **Phase de prétraitement** contient deux étapes : (i) la dérivation du sous-cube endogène selon la spécification de l'expert en utilisant une requête SQL ; (ii) la discrétisation des mesures d'analyse en utilisant la technique de *fréquence égale* ;
- **Phase de traitement** consiste à l'extraction des règles d'association cycliques basées sur les mesures à partir du cube de données et l'évaluation de ces règles par des critères de qualité enrichis par les fonctions d'agrégation.

### 5.1 Phase de prétraitement

Cette phase comporte les deux étapes qu'on présentera dans cette sous-section.

#### 5.1.1 Dérivation du sous-cube endogène

En se basant sur la spécification de l'expert du contexte endogène, une requête SQL sera lancée pour obtenir le sous-cube endogène par la sélection des mesures d'analyse selon les restrictions des mesures de référence. Ceci s'accomplit grâce à l'utilisation d'une requête SQL qui permettra de délimiter le nouveau domaine d'analyse. Par exemple, considérons le cas décrit par la table 1 et le contexte délimité comme suit : la dimension temporelle est la dimension Temps, les mesures d'analyse sont la quantité vendue et le chiffre d'affaire externe avec la mesure de référence fixée à la mesure taux de change ayant des taux dépassant 1.5%. La requête lancée est illustrée par la figure 2.

```
SELECT Temps, QuantitéVendue, ChiffresDAffairesExterne  
FROM  $\mathcal{T}$   
WHERE TauxDeChange > 1.5% ;
```

FIG. 2 – Exemple de requête SQL lancée pour obtenir le sous-cube endogène.

#### 5.1.2 Discrétisation des mesures d'analyse en utilisant la technique de fréquence égale

Pour générer les règles d'association cycliques multidimensionnelles à partir du sous-cube endogène, une discrétisation des mesures de contexte est primordiale. Le choix de la technique de *fréquence égale* est motivé par le fait que les intervalles sont créés de telle sorte que la fréquence de chaque intervalle est constante et donc chaque intervalle contient à peu près le même nombre de valeurs. Dans notre cas, l'expert fixe la fréquence à quarante. En fait, les quarante valeurs les plus proches forment un intervalle indépendant. Par exemple, 77520, 75000, 77520 et trente-sept autres valeurs de la quantité vendue appartenant à l'intervalle de 75000 et 80000 composent l'intervalle [75000-80000]. Appliquant cette méthode, le sous-cube obtenu après l'étape de discrétisation est représenté par la table 2.

Temps $T$	Quantité vendue $Q$	Taux de Change $TC$	Chiffre d'affaires Externe $CAE$
Premier trimestre 2007	75000-80000	1,946%- 1,948%	450-750
Second trimestre 2007	75000-80000	1,946%- 1,948%	450-750
Troisième trimestre 2007	75000-80000	1,946%- 1,948%	450-750
Quatrième trimestre 2007	80000-90000	1,946%- 1,948%	800-1000

TAB. 2 – Table  $\mathcal{T}$  avec une discrétisation des mesures par la technique de **fréquence égale**

Une fois la discrétisation effectuée, nous utilisons la valeur médiane pour chaque intervalle de mesure afin qu'on puisse considérer l'additivité temporelle des mesures lors de l'évaluation des motifs générés. En fait, la valeur médiane peut être calculée en utilisant la formule suivante  $V_{mediane} = \frac{V_{max}+V_{min}}{2}$ .

Par exemple, dans notre cas (table 2), l'item considéré pour la mesure quantité vendue est  $\alpha = \frac{V_{max}+V_{min}}{2} = \frac{75000+80000}{2} = 77500$ . La sortie de la table est représentée par la table 3 qui sera une entrée pour notre algorithme.

Temps $T$	Quantité vendue $Q$	Taux de Change $TC$	Chiffre d'affaires Externe $CAE$
Premier trimestre 2007	77500	1,947%	600
Second trimestre 2007	77500	1,947%	600
Troisième trimestre 2007	77500	1,947%	600
Quatrième trimestre 2007	85000	1,947%	700

TAB. 3 – Table  $\mathcal{T}$  finale obtenue suite à la phase de prétraitement

## 5.2 Phase de traitement pour la génération des règles cycliques multidimensionnelles basées sur les mesures

Suite à la dérivation du sous-cube endogène puis la discrétisation des mesures évaluées et enfin le calcul de médiane pour chaque intervalle trouvé, l'algorithme permettant l'extraction des règles d'association cycliques multidimensionnelles basées sur les mesures, appelé MICAR reçoit en entrée mis à part le sous-cube obtenu, le seuil minimum de support  $MinSupp$ , le seuil minimum de confiance  $MinConf$  et la longueur du cycle  $l$ .

Notation	Description
$SC$	Sous-cube endogène
$\alpha$	Item cyclique dimensionnel potentiel basé sur la mesure
$I$	Itemset cyclique multidimensionnel basé sur les mesures
$\mathcal{C}_i$ (resp. $\mathcal{F}_i$ )	Ensemble des candidates (resp. fréquents) d'i-itemsets cycliques multidimensionnels basés sur les mesures.
$MinSupp$ (resp. $MinConf$ )	Seuil minimum de support (resp. de confiance)
$\mathcal{D}_t$	Temps $t$
$l$	Longueur du cycle
$Supp(\mathcal{C})$	Support de l'itemset cyclique multidimensionnel $\mathcal{C}$ basé sur les mesures
$\mathcal{M}_{MAX}$ (resp. $\mathcal{M}_{MIN}, \mathcal{M}_{AVG}, \mathcal{M}_{SUM}$ )	Ensemble des mesures agrégées par la fonction MAX (resp. MIN, AVG, SUM)

TAB. 4 – Liste des notations utilisées pour l'algorithme MICAR.

**Algorithme 1 : MICAR : Règles d'Association Cycliques MultIdimensionnelles basées sur les Mesures**

**Data :**  $SC, \mathcal{M}_{insupp}, \mathcal{M}_{inconf}, l, \mathcal{M}_{MAX}, \mathcal{M}_{MIN}, \mathcal{M}_{SUM}, \mathcal{M}_{AVG}$   
**Result :**  $\mathcal{R}$  : Règles d'association cycliques multidimensionnelles basées sur les mesures dans  $SC$

**début**  
 $\mathcal{F}_1 =$  Trouver les 1-itemsets cycliques multidimensionnels basés sur les mesures ( $SC, l, \mathcal{D}_t, \mathcal{M}_{insupp}$ );  
**pour** ( $k=2; k \neq 0; k++$ ) **faire**  
      $\mathcal{C}_k =$  CandidatGeneration( $\mathcal{C}_{k-1}$ );  
     **si**  $\mathcal{C}_k$  est un itemset cyclique multidimensionnel basé sur les mesures **alors**  
         **Pour chaque** transaction  $\mathcal{T} \in SC$  à la date  $\mathcal{D}_t$  **faire**  
              $\mathcal{C}_i =$ subset( $\mathcal{C}_k, \mathcal{T}$ )  
             **Pour chaque** candidat  $\mathcal{C} \in \mathcal{C}_i$  **faire**  
                  $\mathcal{C}.count =$  CalculSupport( $SC, l, \mathcal{D}_t, \mathcal{C}$ );  
                  $\mathcal{F}_k = \{ \mathcal{C} \in \mathcal{C}_k, \mathcal{C}.count > \mathcal{M}_{insupp} \}$   
     **Retour**  $\mathcal{F}_k = \cup_k \mathcal{F}_k$ ;  
     // Génération de Règles  
     **pour** ( $i=2; i < k; i++$ ) **faire**  
         Générer tous les sous-ensembles de  $\mathcal{F}_i$ ;  
         **Pour chaque** sous-ensemble  $s$  non vide  $\mathcal{F}_i$  **faire**  
              $r = s \rightarrow (\mathcal{F}_i - s)$ ;  
             **si** ( $confiance(r) > \mathcal{M}_{inconf}$ ) **alors**  
                  $\mathcal{R} = \mathcal{R} \cup r$ ;  
     **Retour**  $\mathcal{R}$ ;  
**fin**

**Procédure** Trouver 1-itemsets cycliques multidimensionnels basés sur les mesures ( $SC, l, \mathcal{D}_t, \mathcal{M}_{insupp}$ )  
**Result :**  $\mathcal{F}_1$   
**début**  
     **tant que** (!Fin de tuples dans  $SC$ ) **faire**  
         **Pour chaque** transaction  $\mathcal{T} \in SC$  **faire**  
             **Pour chaque** item  $\alpha \in \mathcal{T}$  **faire**  
                 **Pour chaque** transaction  $\mathcal{T}' \in SC$  à la date  $\mathcal{D}_{t+i}$  **faire**  
                     //  $\alpha \in M_k$   
                     **si** ( $\alpha$  existe dans  $\mathcal{T}'$ ) et ( $M_k \in \mathcal{M}_A$ ) **alors**  
                         **si**  $M_k \in \mathcal{M}_{SUM}$  **alors**  
                              $Supp(\alpha) = \frac{\alpha}{SUM(M_k)}$ ; // avec  $m$  la mesure courante  
                         **si**  $M_k \in \mathcal{M}_{MAX}$  **alors**  
                              $Supp(\alpha) = \frac{\alpha}{MAX(M_k)}$ ;  
                         **si**  $M_k \in \mathcal{M}_{MIN}$  **alors**  
                              $Supp(\alpha) = \frac{\alpha - MIN(M_k)}{MAX(M_k) - MIN(M_k)}$ ;  
                         **si**  $M_k \in \mathcal{M}_{AVG}$  **alors**  
                             **si**  $\alpha \geq AVG(M_k)$  **alors**  
                                  $Supp(\alpha) = \frac{\alpha}{AVG(M_k)}$ ;  
                              $Supp(\alpha) = \frac{\alpha - MIN(M_k)}{MAX(M_k) - MIN(M_k)}$ ;  
                     **si** ( $Supp(\alpha) > \mathcal{M}_{insupp}$ ) **alors**  
                          $\mathcal{F}_1 = \mathcal{F}_1 \cup \alpha$ ;  
     **Retour**  $\mathcal{F}_1$ ;  
**fin**

**Procédure** CalculSupport ( $SC, l, \mathcal{D}_t, \mathcal{C}$ )  
**Result :**  $Supp(\mathcal{C})$   
**début**  
     NoMoreCyclic : Booléen;  
     NoMoreCyclic = faux;  
     **tant que** (!Fin de tuples dans  $SC$ ) et (!NoMoreCyclic) **faire**  
          $\mathcal{C}_k =$  CandidatGeneration( $\mathcal{C}_{k-1}$ );  
         **Pour chaque** transaction  $\mathcal{T} \in SC$  à la date  $\mathcal{D}_{t+i}$  **faire**  
             **si**  $\mathcal{C}$  existe dans  $\mathcal{T}$  **alors**  
                 **si** Itemset  $l$  de  $\mathcal{C}$  contient  $\mathcal{M}_A$  **alors**  
                      $Supp(\mathcal{C}) = \prod_{i=1}^n Supp(i)$ .  
             NoMoreCyclic = vrai;  
     **Retour**  $Supp(\mathcal{C})$ ;  
**fin**

Il fournit en sortie de la liste des règles d'association cycliques multidimensionnelles basées sur les mesures. Les notations utilisées sont présentées par la table 4 et le pseudo-code de l'algorithme est illustré dans ce qui suit. En fait, MICAR, un processus itératif, fonctionne en trois étapes successives :

En premier lieu, nous procédons par un niveau croissant de recherche pour trouver les  $i$ -itemsets cycliques multidimensionnels larges basés sur les mesures, où le niveau de  $(i)$  représente le nombre d'items dans l'itemset. On note  $C(i)$  le candidat  $i$ -itemset cyclique multidimensionnel basé sur les mesures potentiellement fréquent, et  $F(i)$   $i$ -itemsets cycliques multidimensionnels basés sur les mesures fréquents. Pour chaque niveau de  $(i)$ , si l'ensemble  $C(i)$  n'est pas vide, la première étape de notre algorithme extrait les motifs cycliques multidimensionnels basés sur les mesures fréquents  $F(i)$  à partir de  $C(i)$  selon deux conditions :

- un itemset cyclique multidimensionnel basé sur les mesures  $A \in C(i)$  doit être une conjonction de membres de mesures d'analyse ;
- et un itemset cyclique basé sur les mesures doit avoir un support supérieur au seuil minimum de support  $Minsupp$  pour être inclus dans  $F(i)$ .

Par exemple,  $I : [CAE = 600, Q = 77500]$  est un *itemset cyclique multidimensionnel basé sur les mesures fréquent* si son support dépasse le seuil minimum de support.

La deuxième étape utilise les  $i$ -itemsets cycliques multidimensionnels basés sur les mesures larges  $F(i)$  pour obtenir un nouvel ensemble  $C(i+1)$  de  $(i+1)$ -candidats. Un  $(i+1)$ -candidat est formé par l'union de deux  $i$ -itemsets A et B à partir de  $F(i)$  en fonction de trois conditions :

- A et B doivent avoir  $(i-1)$  items cycliques dimensionnels basés sur les mesures communs ;
- Tous les sous-itemsets multidimensionnels cycliques  $A \cup B$  basés sur les mesures doivent être des instances de  $\mathcal{M}_A$ ,
- Tous les sous-itemsets cycliques de  $A \cup B$  basés sur les mesures doivent être des itemsets cycliques multidimensionnels basés sur les mesures fréquents.

Enfin, la troisième étape consiste à scanner  $F(i)$  niveau par niveau. De tous les  $A \in F(i)$ , on extrait les règles d'association cycliques multidimensionnelles basées sur les mesures ayant une confiance supérieure au seuil minimum de confiance  $MinConf$ . La règle  $\mathcal{R} : CAE=600 \Rightarrow Q=77500$  est un exemple de règle d'association cyclique multidimensionnelle basée sur les mesures dont sa confiance dépasse  $MinConf$ .

## 6 Évaluation expérimentale

Les expérimentations ont été menées sur un PC Pentium doté d'une fréquence d'horloge égale à 2 GHz et d'une mémoire centrale égale à 512 Mo. Nous présentons les résultats des tests réalisés sur un entrepôt de ventes réel<sup>4</sup>, contenant une table de faits de vente ayant trois dimensions, à savoir *le temps*, *le lieu* et *le client* et cinq mesures i.e., *la quantité vendue*, *le chiffre d'affaires interne*, *le taux de change*, *le chiffre d'affaires externe* et *le bénéfice réalisé*. Chaque dimension contient presque cinq attributs et a une hiérarchie de concept comportant trois niveaux. L'entrepôt de données se dote d'une architecture OLAP relationnel (ROLAP) et est modélisé par un schéma en étoile. La première figure 3.(a) montre le comportement de

4. L'entrepôt de données relatif aux ventes a été mis en place en se basant sur les données d'une entreprise cotée en bourse dont ses indicateurs d'activité sont disponibles à l'adresse <http://www.bvmt.com.tn/companies/?view=listed>.

## Usage Des Mesures Pour La Génération Des Règles d'Associations Cycliques

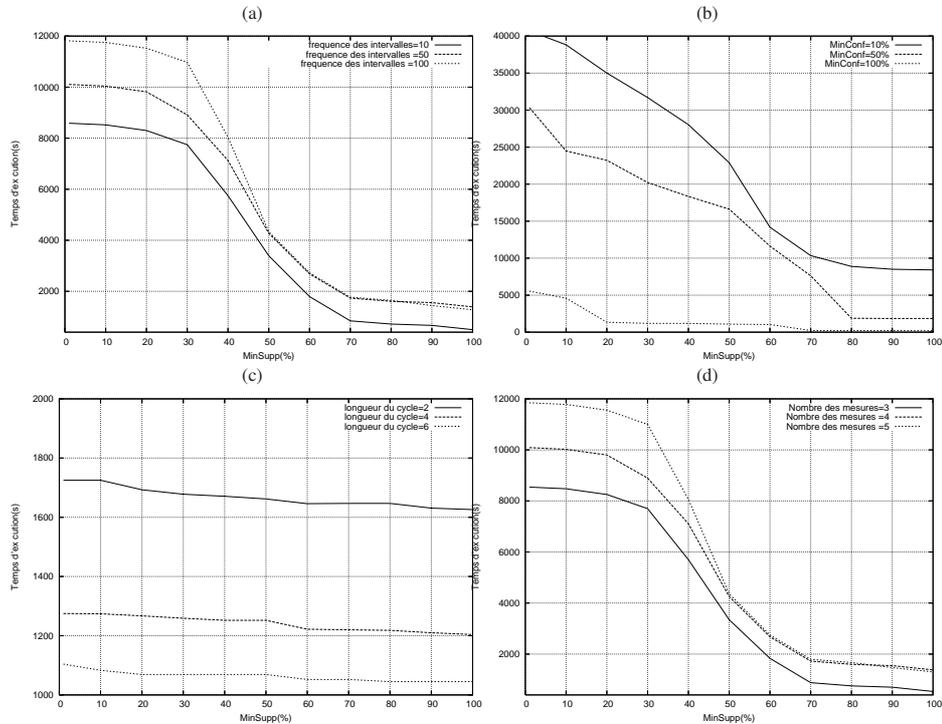


FIG. 3 – Le temps d'exécution de notre algorithme par rapport à la variation de (a) la fréquence d'intervalle, (b) le minimum de confiance, (c) la longueur du cycle, (d) le nombre de mesures.

notre approche en cas de changement de fréquence d'intervalle lors de la discrétisation. L'augmentation de la fréquence d'intervalles, diminue la performance de notre méthode car si on accroît la fréquence, le nombre de candidats qui pourront être éventuellement fréquents augmente aussi.

La figure 3.(b) souligne la relation entre le temps d'exécution de notre algorithme et le support des règles obtenues en fonction de plusieurs seuils de confiance. Généralement, l'extraction de règles d'association nécessite moins de temps quand il s'agit de valeurs croissantes du support. Figure 3.(c) présente un test de notre algorithme avec la variation de la longueur du cycle. Pour les petites valeurs de support, le temps d'exécution augmente considérablement en diminuant la longueur du cycle. Toutefois, pour de grandes valeurs de support, la longueur du cycle a un moindre effet sur le temps de réponse de notre algorithme. La longueur du cycle est d'une année c'est à dire quatre trimestres vu que généralement les mêmes événements se répètent annuellement. Dans la figure 3.(d), nous signalons que l'efficacité de l'algorithme dépend étroitement du nombre de mesures. Lorsque le nombre de mesures est élevé, le nombre d'itemsets fréquents obtenus parmi plusieurs candidats est élevé. Cependant, lorsque ce nombre est réduit, le nombre des itemsets fréquents est également faible. On en conclut que la performance de notre algorithme s'améliore en diminuant le nombre de mesures.

En général, ces expériences illustrent un temps d'exécution acceptable. L'efficacité de notre algorithme est due à : (i) l'utilisation de la fouille de données sous contraintes réduit fortement l'espace de recherche des règles d'association cycliques et, par conséquent, diminue considé-

ablement le temps d'exécution du processus de fouille ; (ii) l'utilisation de la propriété Apriori qui est certainement adaptée à des cubes de données éparses et réduit considérablement la complexité de la recherche itemsets larges, (iii) l'utilisation des fonctions d'agrégation appropriées pour évaluer l'association multidimensionnelle basée sur les mesures pendant le processus de fouille.

## 7 Conclusion

Dans cet article, nous avons introduit une nouvelle méthode pour extraire des règles d'association cycliques multidimensionnelles basées sur les mesures en tenant compte de l'additivité de ces mesures par rapport au temps lors de l'évaluation des motifs extraits à partir du contexte OLAP. Ainsi, une nouvelle définition des motifs cycliques multidimensionnels à base de mesure est fournie. Une redéfinition des métriques d'évaluation inspirée de la caractéristique d'additivité temporelle des mesures est présentée en utilisant les fonctions d'agrégation adéquates qui caractérisent chaque mesure. En conséquence, une nouvelle méthode appelée MICAR pour extraire ces motifs est introduite. D'autres perspectives pour des futurs travaux concernent principalement les pistes suivantes : (i) l'utilisation d'autres techniques de discrétisation pour éviter la perte de l'information des mesures, (ii) la combinaison des techniques de discrétisation pour optimiser la répartition optimale des mesures en intervalles, (iii) l'extraction de motifs cycliques disjonctifs dans un contexte OLAP.

## Références

- Agrawal, R., T. Imielinski, et N. Swami. Mining association rules between sets of items in large databases. In *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, Washington, D.C., May 26-28, 1993*, pp. 207–216.
- Ben Ahmed, E. et F. Gargouri (2010). Règles d'association cycliques dans un contexte multidimensionnel. In *Atelier des Systèmes Décisionnels (ASD'10)*.
- Ben Ahmed, E. et M. Gouider (2010). Towards a new mechanism of extracting cyclic association rules based on partition aspect. In *IEEE Research Challenges In Computer Science (RCIS'10')*, pp. 69 – 78.
- Ben Messaoud, R., O. Boussaid, S. Rabaséda, et R. Missaoui (2006). Enhanced mining of association rules from data cubes. In *Proceedings of the 9 th ACM International Workshop on Data Warehousing and OLAP (DOLAP 2006)*, pp. 11–18.
- Chiang, D., C. Wang, S. Chen, et C. Chen (2009). The cyclic model analysis on sequential patterns. *IEEE Trans. on Knowl. and Data Eng.* 21, 1617–1628.
- Han, J., W. Gong, et Y. Yin (1999). Efficient mining of partial periodic patterns in time series database. In *ICDE*, pp. 106–115.
- Kamber, M., J. Han, et J. Chiang (1997). Metarule-guided mining of multi-dimensional association rules using data cubes. In *Proceedings of the 1997 International Conference on Knowledge Discovery and Data Mining (KDD'97), Newport Beach, CA.*, pp. 207–210.
- Lenz, H. et A. Shoshani (1997). Summarizability in olap and statistical data bases. pp. 132–143.

- Liu, H., F. Hussain, C. Tan, et M. Dash (2002). Discretization : An enabling technique. In *Data Mining and Knowledge Discovery*, Volume (6), pp. 393–423.
- Liu, H. et R. Setiono (1995). Chi2 : Feature selection and discretization of numeric attributes. In *IEEE Int. Conf. Tools with AI (ICTAI'95)*, pp. 388–391.
- Nestorov, S. et N. Jukic (2003). Ad-hoc association-rule mining within the data warehouse. In *Proceedings of the 36 th Hawaii International Conference on System Sciences (HICSS 2003)*, pp. 232–242.
- Ozden, B., S. Ramaswamy, et A. Silberschatz (1998). Cyclic association rules. In *Proceedings of the Fourteenth International Conference on Data Engineering, February 23-27, 1998, Orlando, Florida, USA*, pp. 412–421. IEEE Computer Society.
- Palaniappan, S. et T. Hong (2008). Discretization of continuous valued dimensions in olap data cubes. *International Journal of Computer Science and Network Security* 8, 116–126.
- Plantevit, M., A. Laurent, D. Laurent, M. Teisseire, et Y. Choong (2010). Mining multidimensional and multilevel sequential patterns. *ACM Transactions on Knowledge Discovery from Data* 4, 155–174.
- Plantevit, M., M. Laurent, et M. Teisseire (2008). Motifs séquentiels multidimensionnels et mesure. différentes techniques pour calculer le support. *Ingénierie des Systèmes d'Information* 13(6), 9–32.
- Plantevit, M., C. Y.W., A. Laurent, D. Laurent, et M. Teisseire (2005). M2sp : Mining sequential patterns among several dimensions. In *PKDD 2005, 9th European Conference on Principles and Practice of Knowledge Discovery in Databases*, Lecture Notes in Computer Science, pp. 205–216.
- Thuan, N. (2004). Mining cyclic association rules in temporal database. *The Journal Science and technology development, Vietnam National University* 7(8), 12–19.
- Thuan, N. (2008). Mining time pattern association rules in temporal database. In *SCSS (1)*, pp. 7–11.
- Tjioe, H. et D. Taniar (2005). Mining association rules in data warehouses. *IJDWM* 1(3), 28–62.
- Zhu, H. (December 1998). *On-line analytical mining of association rules*. Ph. D. thesis, Master's thesis, Simon Fraser University, Burnaby, British Columbia, Canada.

## Summary

The online analytical processing (OLAP) does not provide any explanation of correlations discovered between data. Thus, the coupling of OLAP and data mining, especially association rules, is considered as an efficient solution to this problem. In this context, we mainly focus on a particular class of association rules which is the cyclic association rules. These rules aimed to discover patterns that display regular variation over user-defined intervals. Generally, the generated patterns do not take an advantage from the specificities of the multidimensional context namely, the consideration of the measures and their aggregations. In this paper, we introduce a novel method for extracting cyclic association rules from measures, and we redefine the evaluation metrics of association rules quality inspired of the temporal summarizability of measures concept through the integration of appropriate aggregation functions. To prove the usefulness of our approach, we conduct an empirical study on a real data warehouse.