CNRIA

Construction et exploration de résumés de grands ensembles de règles d'association

Ndiaye Marie[*,**] – Diop Cheikh Talibouya[**] – Giacometti Arnaud[*] – Marcel Patrick[*] – Soulet Arnaud[*]

[*]Laboratoire d'Informatique
Université François Rabelais Tours – Antenne Universitaire de Blois
3 place Jean Jaurès, 41000 Blois (FRANCE)
{marie.ndiaye – arnaud.giacometti – patick.marcel – arnaud.soulet}@univ-tours.fr

[**]Laboratoire d'Analyse Numérique et d'Informatique
Université Gaston Berger de Saint-Louis
BP 234 Saint-Louis (SENEGAL)
talibouya1@yahoo.fr

RÉSUMÉ. Les algorithmes d'extraction génèrent souvent de grands ensembles de règles d'association. La représentation de ces ensembles par des résumés est un moyen généralement utilisé pour faciliter leur analyse. La plupart des méthodes de résumé proposées dans la littérature produisent des ensembles de règles difficiles à analyser car ils ne sont présentables que sous forme de listes. Dans ce travail, nous proposons des résumés basés sur un schéma qui permet de les présenter sous forme de tableau croisé. Nous proposons ensuite des opérations permettant de naviguer entre eux. Enfin, nous définissons une mesure appelée homogénéité, qui évalue l'information contenue dans les règles et qui est perdue dans les résumés.

ABSTRACT. Mining algorithms often generate huge sets of association rules. Summaries are generally used to represent these sets in order to facilitate their analysis. Most of the summary methods proposed in the literature produce rule sets which are difficult to analyze because they are presentable only in the form of lists. In this work, we propose summaries based on a schema that enables to present them as cross-tables. Then, we propose operations to navigate between them. Finally, we define a measure called homogeneity, which evaluates the information in the rules that is lost in the summaries.

MOTS-CLÉS : règles d'association, résumé, mesure de qualité de résumé

KEYWORDS : association rules, summary, quality measure of summary

1. Introduction

La grande quantité de règles d'association [2] produites par les algorithmes d'extraction constitue un problème majeur pour leur exploration. Plusieurs travaux ont été menés dans le but de réduire le nombre de règles extraites en supprimant les règles redondantes [12] ou peu intéressantes pour l'utilisateur [3], ou en introduisant des contraintes lors de l'extraction [10]. Ces techniques peuvent diminuer considérablement le nombre de règles. Toutefois, la quantité reste souvent très importante. Pour interpréter plus facilement ces règles qui restent encore nombreuses, des méthodes de résumé ont été proposées dans la littérature. Certains de ces travaux concernent le résumé d'ensembles de motifs [11, 4, 1, 6, 5] de manière générale. D'autres s'intéressent en particulier au résumé d'ensembles de règles d'association [8]. Cependant, les résumés générés par ces méthodes sont généralement difficiles à analyser, car leur représentation n'est pas prise en compte lors de leur construction.

La première principale contribution de ce papier est la proposition de résumés de grands ensembles de règles d'association basés sur un schéma. Un ensemble de règles peut être résumé selon plusieurs schémas. Le schéma permet de les présenter sous forme de tableaux croisés et de définir des opérateurs pour naviguer entre eux. La deuxième principale contribution est la proposition d'une mesure de qualité appelée homogénéité qui évalue la perte d'information des résumés. Elle correspond à la quantité totale d'information des règles d'origine qui est perdue dans les résumés. Nous présentons dans ce papier les idées générales de nos contributions qui sont développées dans [7] avec de plus amples détails.

Le reste de l'article est organisé comme suit. Quelques définitions et notations sont exposées dans la section 2. La section 3 est consacrée à la description des résumés basés sur un schéma. Dans la section 4, nous présentons ces résumés avec les tableaux croisés et nous proposons des opérations élémentaires qui permettent de naviguer entre eux. Ensuite, nous proposons dans la section 5 une mesure pour évaluer la qualité des résumés. Enfin, une conclusion et quelques perspectives pour nos travaux futurs sont présentées dans la section 6.

2. Cadre général

Dans cette section, nous introduisons quelques définitions et notations nécessaires à la description de notre approche.

2.1. Règles d'association

Considérons l'ensemble des attributs du domaine étudié noté $\mathcal{A}$. Soit A un attribut de $\mathcal{A}$, on note $dom(A)$ le domaine de l'attribut A et $dom(\mathcal{A})$ le produit cartésien des domaines des attributs de $\mathcal{A}$, i.e. $dom(\mathcal{A}) = \times_{A \in \mathcal{A}} dom(A)$.

Un *item* x défini sur $\mathcal{A}$ est un couple attribut-valeur noté (A, a) avec $A \in \mathcal{A}$ et $a \in dom(A)$.

Un *itemset* X défini sur $\mathcal{A}$ est un ensemble d'items définis sur $\mathcal{A}$. Soit un itemset $X = \{(A_1, a_{i_1}), ..., (A_k, a_{i_k})\}$ défini sur $\mathcal{A}$, on appelle *schéma* de X noté $sch(X)$ l'ensemble d'attributs $\{A_1, ..., A_k\} \subseteq \mathcal{A}$.

Une *règle d'association* est une relation $X \Rightarrow Y$ où X et Y sont des itemsets et $X \cap Y = \emptyset$. X est appelé le corps et Y la tête de la règle. On appelle schéma de $X \Rightarrow Y$

noté $sch(X \Rightarrow Y)$ le couple $\langle sch(X), sch(Y) \rangle$ formé par le schéma de X et celui de Y. Le tableau 1 présente un ensemble de règles. Chaque ligne correspond à une règle. La première colonne présente les identifiants des règles. La deuxième et la troisième colonne contiennent respectivement leur corps et leur tête.

Règle	**Corps**	**Tête**
r_1	{(CONTROL,auto)}	{(STABILITY,stab)}
r_2	{(CONTROL,auto)}	{(STABILITY,stab),(VISIBILITY,yes)}
r_3	{(CONTROL,auto)}	{(VISIBILITY,yes)}
r_4	{(STABILITY,stab)}	{(VISIBILITY,yes)}
r_5	{(STABILITY,stab)}	{(CONTROL,auto)}
r_6	{(STABILITY,stab)}	{(CONTROL,auto),(VISIBILITY,yes)}
r_7	{(VISIBILITY,yes)}	{(STABILITY,stab)}
r_8	{(VISIBILITY,yes)}	{(CONTROL,auto),(STABILITY,stab)}
r_9	{(VISIBILITY,yes)}	{(CONTROL,auto)}

Tableau 1. *Règles d'association*

2.2. Résumés d'ensembles de règles d'association

Une relation généralement utilisée pour construire des résumés de motifs est la relation de couverture. Dans ce papier, nous utilisons comme relation de couverture la relation suivante.

Définition 1 (Couverture) *Soient $r : X \Rightarrow Y$ et $r' : X' \Rightarrow Y'$ deux règles d'association. r est couverte par r' si r' est plus générale que r, i.e. $X' \subseteq X$ et $Y' \subseteq Y$.*

Cette définition de couverture est une adaptation aux règles de celle qui a été proposée dans [4] pour les itemsets. Elle est plus générale que la couverture utilisée dans [8]. En effet, les auteurs considèrent que r' couvre r seulement si $X' \subseteq X$ et $Y = Y'$ sachant que Y est constitué d'un seul item. Considérons un ensemble de règles R et une règle r, l'ensemble composé des règles de R couvertes par r est noté $cover(r, R)$.
La définition 2 est une adaptation aux règles d'association de la définition de résumé d'itemsets proposée dans [4] qui repose sur la relation de couverture.

Définition 2 (Résumé) *Soit R un ensemble de règles. Un résumé de R est un ensemble de règles S tel que : (i) pour toute règle r de R, il existe une règle s de S telle que s couvre r ; (ii) chaque règle de S couvre un sous-ensemble non vide de R.*

Notons que d'après cette définition, un résumé n'est pas forcément inclus dans l'ensemble qu'on résume. Par ailleurs, un résumé trivial d'un ensemble de règles est l'ensemble lui même. Cependant, il n'est pas intéressant car un des buts qu'on veut atteindre en résumant un ensemble de règles est d'obtenir un ensemble plus petit. Pour trouver des résumés intéressants, nous nous intéressons en particulier aux résumés minimaux.

Définition 3 (Résumé minimal) *Soit R un ensemble de règles et S un résumé de R. S est dit minimal s'il n'existe pas un ensemble de règles $S' \subset S$ tel que S' est un résumé de R.*

Un résumé est minimal s'il ne couvre plus toutes les règles de l'ensemble de départ lorsqu'on lui enlève ne serait-ce qu'une règle. Les résumés minimaux de R sont ses plus petits résumés possibles.

Dans la section 3, nous présentons la première contribution de ce papier qui est la proposition de résumés minimaux basés sur un schéma.

3. Résumé minimal basé sur un schéma

Intuitivement, un résumé S d'un ensemble de règles R est basé sur le schéma $\langle \mathcal{B}, \mathcal{H} \rangle$ si $\langle \mathcal{B}, \mathcal{H} \rangle$ est le schéma de toutes ses règles. Chaque règle de S résume celles de R qu'elle couvre. Par la suite, on dira que S est un résumé de schéma $\langle \mathcal{B}, \mathcal{H} \rangle$. Dans cette section, nous proposons une définition formelle de résumés basés sur un schéma et nous décrivons la méthode de construction de tels résumés.

3.1. Définition

En pratique, un sous-ensemble de R dont les règles ont le même schéma $\langle \mathcal{B}, \mathcal{H} \rangle$ couvre rarement la totalité de R. En effet, on peut trouver des attributs de $\mathcal{B}$ et $\mathcal{H}$ qui n'apparaissent pas respectivement dans la tête et dans le corps de certaines règles de R. Par conséquent, ces règles ne sont pas couvertes. Afin de permettre la construction de résumés qui couvrent la totalité des ensembles de règles, nous introduisons une extension des itemsets qui constitueront la tête et le corps des règles des résumés.

Soit $\mathcal{A}$ un ensemble d'attributs. Soit $null$ une valeur n'appartenant à aucun domaine $dom(A)$, $A \in \mathcal{A}$. $dom(A)^+$ représente le domaine de A auquel la valeur $null$ est ajoutée, i.e. $dom(A)^+ = dom(A) \cup \{null\}$. Ainsi, $dom(\mathcal{A})^+$ correspond au produit cartésien des ensembles $dom(A)^+$, i.e. $dom(\mathcal{A})^+ = \times_{A \in \mathcal{A}} dom(A)^+$. L'itemset X^+, appelé extension de l'itemset X, est défini par $X^+ = X \cup \{(A, null) \mid A \in \mathcal{A} \setminus sch(X)\}$. Intuitivement, pour tout attribut A de $\mathcal{A}$ n'apparaissant pas dans $sch(X)$, l'item $(A, null)$ est ajouté à X. Ainsi, Tous les itemsets peuvent être représentés avec le même schéma $\mathcal{A}$. Par la suite, on note $\mathcal{R}^*$ l'ensemble de toutes les règles telles que la valeur de chaque item (A, a) de sa tête et de sont corps appartient à $dom(A)^+$ ($A \in \mathcal{A}$). La définition suivante est une extension de la définition de couverture (voir définition 1) sur $\mathcal{R}^*$.

Définition 4 (Couverture étendue) *Soit $X_1 \Rightarrow Y_1$ une règle de $\mathcal{R}$ et $X_2 \Rightarrow Y_2$ une règle de $\mathcal{R}^*$. $X_1 \Rightarrow Y_1$ est couvert par $X_2 \Rightarrow Y_2$ si $X_2 \Rightarrow Y_2$ est plus général que $X_1^+ \Rightarrow Y_1^+$, i.e. $X_2 \subseteq X_1^+$ et $Y_2 \subseteq Y_1^+$.*

La couverture étendue permet de prendre en compte dans les résumés l'absence de valeur de certains attributs au niveau des règles. Ainsi, la définition suivante de résumé basé sur un schéma repose sur cette couverture.

Définition 5 (Résumé basé sur un schéma) *Soit un ensemble de règles R et un schéma de règles $\langle \mathcal{B}, \mathcal{H} \rangle$. Le résumé de R base sur $\langle \mathcal{B}, \mathcal{H} \rangle$ est l'ensemble $\sigma_{\langle \mathcal{B}, \mathcal{H} \rangle}(R)$ de toutes les règles $X \Rightarrow Y$ telles que $X \in dom(\mathcal{B})^+$, $Y \in dom(\mathcal{H})^+$ et $cover(X \Rightarrow Y, R) \neq \emptyset$.*

3.2. Construction de résumé basé sur un schéma

Pour déterminer le résumé S de R basé sur le schéma $\langle \mathcal{B}, \mathcal{H} \rangle$, on construit d'abord les règles candidates en effectuant toutes les combinaisons possibles entre les itemsets de $dom(\mathcal{B})^+$ et ceux de $dom(\mathcal{H})^+$. Ensuite, on construit $cover(s, R)$ pour chaque candidat s. Si $cover(s, R)$ n'est pas vide, alors la règle s est sélectionnée. Les règles ainsi retenues forment le résumé S. Le tableau 2 présente les règles construites à partir du schéma $\langle$\{VISIBILITY\},\{CONTROL\}$\rangle$ sachant que $dom(VISIBILITY)^+ = \{null, yes, no\}$

et $dom(CONTROL)^+ = \{null, auto, noauto\}$. La dernière colonne du tableau montre les règles couvertes par chaque règle générée. Le résumé de schéma $\langle${VISIBILITY},{ CONTROL}$\rangle$ est constitué des règles s_1, s_2, s_3 et s_4. Les règless_5, s_6, s_7, s_8 et s_9 n'appartiennent pas au résumé car elles ne couvent aucune règle de $\mathcal{R}$.

Règle	**Corps**	**Tête**	**Règles couvertes**
s_1	$\{(VISIBILITY, null)\}$	$\{(CONTROL, null)\}$	$\{r_1, r_2, r_3, r_4\}$
s_2	$\{(VISIBILITY, null)\}$	$\{(CONTROL, auto)\}$	$\{r_5, r_6\}$
s_3	$\{(VISIBILITY, yes)\}$	$\{(CONTROL, null)\}$	$\{r_7\}$
s_4	$\{(VISIBILITY, yes)\}$	$\{(CONTROL, auto)\}$	$\{r_8, r_9\}$
s_5	$\{(VISIBILITY, null)\}$	$\{(CONTROL, noauto)\}$	
s_6	$\{(VISIBILITY, yes)\}$	$\{(CONTROL, noauto)\}$	
s_7	$\{(VISIBILITY, no)\}$	$\{(CONTROL, null)\}$	
s_8	$\{(VISIBILITY, no)\}$	$\{(CONTROL, auto)\}$	
s_9	$\{(VISIBILITY, no)\}$	$\{CONTROL, noauto)\}$	

Tableau 2. *Construction d'un résumé basé sur un schéma*

Chaque règle de l'ensemble qu'on résume est couverte par une seule et unique règle du résumé. Ainsi, les résumés présentés à l'utilisateur ne comportent pas de redondances. Pair ailleurs, les résumés obtenus sont minimaux. En effet, la suppression d'une règle quelconque du résumé entraine la non couverture d'au moins une règle de R. Par la suite, le schéma est utilisé pour présenter les résumés sous forme de tableaux croisés et pour construire des opérateurs de navigation.

4. Exploration de résumés

La présentation des règles découvertes ainsi que leur navigation constituent des facteurs déterminants dans l'étape d'exploration. Dans cette section, nous décrivons une manière de présenter les résumés et nous proposons des opérateurs qui permettent de naviguer entre eux.

4.1. Tableaux croisés

Les tableaux croisés constituent un moyen simple et intuitif d'exploration de données multidimensionnelles. Ils sont particulièrement utilisés pour afficher les données provenant de cubes OLAP. Leur structure est tout aussi adaptée pour présenter des résumés basés sur un schéma. La figure 1 présente sous forme de tableau croisé, le résumé de schéma $\langle${VISIBILITY,STABILITY},{CONTROL}$\rangle$.

			TETE	
CORPS			(CONTROL, $null$)	(CONTROL, $auto$)
	(VISIBILITY, $null$)	(STABILITY, $null$)	3	0
		(STABILITY, stab)	1	2
	(VISIBILITY, yes)	(STABILITY, $null$)	1	2

Figure 1. *Tableau croisé du résumé de schéma $\langle${VISIBILITY,STABILITY},{CONTROL}$\rangle$*

Les items de $dom(\{VISIBILITY\})^+$ et de $dom(\{STABILITY\})^+$ sont imbriqués en ligne et ceux de $dom(\{CONTROL\})^+$ sont placés en colonne. La cellule correspondant à deux itemsets $X \in dom(\{VISIBILITY, STABILITY\})^+$ et $Y \in dom(\{CONTROL\})^+$ représente la règle $X \Rightarrow Y$. Elle contient des informations sur cette dernière. On peut afficher dans les cellules des mesures portant sur l'ensemble des règles couvertes. Chaque cellule de la figure 1 contient le nombre de règles couvertes par la règle du résumé correspondante. Cette représentation est facile à comprendre par sa simplicité et elle permet de présenter de manière intuitive de grands ensembles de règles d'association via leurs résumés.

4.2. Navigation entre résumés

Passer d'un résumé à un autre revient simplement à modifier le schéma du premier résumé en ajoutant ou en supprimant des attributs de $\mathcal{B}$ ou de $\mathcal{H}$. Dans ce papier, nous considérons que la navigation entre résumés correspond à l'application d'opérateurs au schéma des résumés. Un opérateur de navigation est une fonction qui modifie les composantes du schéma des règles d'un résumé. Nous distinguons quatre opérateurs élémentaires : $AddToBody$, $AddToHead$, $DeleteFromBody$ et $DeleteFromHead$. Ils permettent respectivement d'ajouter des attributs à $\mathcal{B}$ et $\mathcal{H}$ et de supprimer des attributs de ces ensembles.
Soit A un attribut de $\mathcal{A}$. Les opérateurs sont formellement définis comme suit.

$$\begin{array}{ll} AddToHead(\langle\mathcal{B},\mathcal{H}\rangle, A) = \langle\mathcal{B},\mathcal{H}\cup\{A\}\rangle & DeleteFromHead(\langle\mathcal{B},\mathcal{H}\rangle, A) = \langle\mathcal{B},\mathcal{H}\backslash\{A\}\rangle \\ AddToBody(\langle\mathcal{B},\mathcal{H}\rangle, A) = \langle\mathcal{B}\cup\{A\},\mathcal{H}\rangle & DeleteFromBody(\langle\mathcal{B},\mathcal{H}\rangle, A) = \langle\mathcal{B}\backslash\{A\},\mathcal{H}\rangle \end{array}$$

La figure 2 montre une utilisation concrète de l'opérateur $AddToBody$.

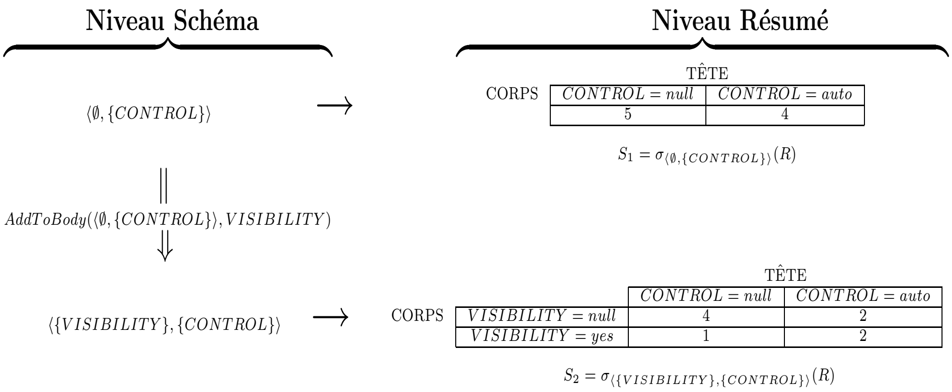

Figure 2. *Navigation entre deux résumés*

La navigation est un moyen très important pour analyser les règles d'association. Cependant, il n'y a pas d'indicateurs permettant de choisir les résumés à explorer. Pour guider l'utilisateur, nous introduisons dans la section 5 une nouvelle mesure de qualité qui évalue la représentativité d'un résumé basé sur un schéma.

5. Mesure de qualité de résumé

Dans notre approche, la qualité d'un résumé dépend de sa perte d'information par rapport à la quantité d'information contenue dans l'ensemble de règles qu'il résume. Nous définissons dans cette section une mesure de qualité qui évalue la perte d'information de résumés basés sur un schéma. Elle repose sur une mesure usuelle de perte d'information appelée entropie conditionnelle de Shannon [9].

5.1. Entropie conditionnelle de Shannon

Considérons la variable aléatoire V_A correspondant à l'appartenance d'un item à la tête ou au corps d'une règle de R et V_S correspondant à la couverture d'une règle de R par une règle du résumé S. L'entropie conditionnelle de V_A sachant V_S, notée $I(V_A|V_S)$, est définie par :

$$I(V_A|V_S) = \sum_{\substack{a \in dom(A) \\ s \in S}} p(V_A = a, V_S = s) \ln(p(V_A = a|V_S = s))$$

$p(V_A = a, V_S = s)$ est la probabilité qu'une règle de R contienne l'item (A, a) dans sa tête ou dans son corps et qu'elle soit couverte par la règle s de S. $p(V_A = a|V_S = s)$ est la probabilité qu'une règle de R contienne l'item (A, a) sachant qu'elle est couverte par la règle s de S. L'entropie conditionnelle de Shannon évalue la similarité des règles couvertes par une même règle du résumé suivant un attribut. Deux règles sont similaires suivant un attribut lorsqu'elles ont dans la tête ou le corps la même valeur pour cet attribut. Plus les règles sont similaires, moins il y a de perte d'information. Maintenant, nous introduisons une définition formelle de la mesure de qualité, appelée homogénéité, qui évalue la perte d'information totale suivant tous les attributs de $\mathcal{A}$.

5.2. Homogénéité d'un résumé

Considérons une règle s du résumé $S = \sigma_{\langle \mathcal{B}, \mathcal{H} \rangle}(R)$. Par construction, toutes les règles de $cover(s, R)$ ont la même valeur pour les attributs de $\mathcal{H}$ et de $\mathcal{B}$. Cependant nous n'avons aucune information concernant les valeurs prises pour les autres attributs de $\mathcal{A}$. Considérons un attribut $A \in \mathcal{A} \backslash (\mathcal{B} \cup \mathcal{H})$. La valeur prise pour A dans les règles de $cover(s, R)$ peut être la même ou différente. S'il s'agit de la même valeur, alors on dit que les règles de $cover(s, R)$ sont homogènes par rapport à A. Par contre, s'il y a plusieurs valeurs prises, alors il y a un désordre engendré par la différence de ces valeurs. L'homogénéité d'un résumé est formellement définie comme suit :

Définition 6 (Qualité de résumé : homogénéité) *Soit R un ensemble de règles extraites à partir d'une relation de schéma $\mathcal{A}$ et S un résumé de R. L'homogénéité de S relativement à R est la moyenne des entropies conditionnelles des V_A, $A \in \mathcal{A}$ sachant V_S.*

$$\phi(R, S) = \frac{1}{|\mathcal{A}|} \sum_{A \in \mathcal{A}} I(V_A|V_S)$$

ϕ mesure l'homogénéité globale d'un résumé. Plus l'homogénéité augmente, plus la valeur de ϕ sera élevée. Cette valeur est négative ou nulle. Elle vaut 0 si R est parfaitement homogène, i.e. si dans chaque groupe $cover(s, R)$, les règles contiennent les mêmes valeurs pour tous les attributs de $\mathcal{A}$. Donc, chaque groupe contient une seule règle. Plus on ajoute des attributs dans $\mathcal{B}$ et $\mathcal{H}$, plus l'homogénéité obtenue augmente. Par exemple, le résumé de schéma $\langle \{VISIBILITY\}, \{CONTROL\} \rangle$ a une meilleure homogénéité que le résumé de schéma $\langle \emptyset, \{CONTROL\} \rangle$. Par conséquent, le résumé de schéma $\langle \emptyset, \emptyset \rangle$ a la plus mauvaise homogénéité. A l'opposé, le résumé de schéma $\langle \mathcal{B}, \mathcal{H} \rangle$ tel que $\mathcal{B} = \mathcal{H} = \mathcal{A}$ a une homogénéité maximale, i.e. $\phi(R, S) = 0$.

6. Conclusion et perspectives

Nous avons proposé des résumés basés sur un schéma pour représenter de grands ensembles de règles d'association. Puis, nous avons utilisé leur schéma pour les présenter sous forme de tableaux croisés et pour définir des opérateurs permettant de naviguer entre eux. Enfin, nous avons défini une mesure de qualité appelée homogénéité qui évalue la perte d'information des résumés.

Nous projetons de poursuivre ce travail en définissant de nouveaux opérateurs de navigation tels qu'un opérateur de zoom qui permet de se focaliser sur une partie de l'ensemble de règles, ou encore des opérateurs de changement de niveau dans des hiérarchies de domaines d'attributs. Enfin, nous envisageons de généraliser notre approche afin de pouvoir résumer d'autres types de motifs tels que les motifs séquentiels ou les graphes.

7. Bibliographie

[1] Foto Afrati, Aristides Gionis, and Heikki Mannila. Approximating a collection of frequent sets. In *KDD '04 : Proceedings of the tenth ACM SIGKDD*, pages 12–19. ACM, 2004.

[2] Rakesh Agrawal, Tomasz Imielinski, and Arun N. Swami. Mining association rules between sets of items in large databases. In Peter Buneman and Sushil Jajodia, editors, *Proceedings of the 1993 ACM SIGMOD*, pages 207–216, 1993.

[3] Liu Bing, Wynne Hsu, and Yiming Ma. Pruning and summarizing the discovered associations. In *KDD '99 : Proceedings of the fifth ACM SIGKDD.*

[4] Varun Chandola and Vipin Kumar. Summarization – compressing data into an informative representation. *Knowl. Inf. Syst.*, 12(3) :355–378, 2007.

[5] Ruoming Jin, Muad Abu-Ata, Yang Xiang, and Ning Ruan. Effective and efficient itemset pattern summarization : regression-based approaches. In *KDD '08 : Proceeding of the 14th ACM SIGKDD*, pages 399–407. ACM, 2008.

[6] Taneli Mielikäinen and Heikki Mannila. The pattern ordering problem. In *Proceedings of the 7th European Conference on Principles of Data Mining and Knowledge Discovery, Lecture Notes in Artificial Intelligence*, pages 327–338. Springer-Verlag, 2003.

[7] Marie Ndiaye, Cheikh T. Diop, Arnaud Giacometti, Patrick Marcel, and Arnaud Soulet. Construction et exploration de résumés de grands ensembles de règles d'association. In *25èmes journées Bases de Données Avancées*, 2009.

[8] Carlos Ordonez, Norberto Ezquerra, and Cesar A. Santana. Constraining and summarizing association rules in medical data. *Knowl. Inf. Syst.*, 9(3) :259–283, 2006.

[9] Claude E. Shannon. A mathematical theory of communication. *Bell system technical journal*, 27 :379–423, 1948.

[10] Ramakrishnan Srikant, Quoc Vu, and Rakesh Agrawal. Mining association rules with item constraints. In David Heckerman, Heikki Mannila, Daryl Pregibon, and Ramasamy Uthurusamy, editors, *KDD*, pages 67–73. AAAI Press, 14–17 1997.

[11] Xifeng Yan, Hong Cheng, Jiawei Han, and Dong Xin. Summarizing itemset patterns : a profile-based approach. In *KDD '05*, pages 314–323, New York, NY, USA, 2005. ACM.

[12] Mohammed J. Zaki. Generating non-redundant association rules. In *KDD '00 : Proceedings of the sixth ACM SIGKDD.*