

Analyse Relationnelle de Concepts pour l'exploration de données relationnelles

Xavier Dolques*, Florence Le Ber*, Marianne Huchard**, Clémentine Nebut**

*ICUBE, Université de Strasbourg/ENGEES, CNRS, Strasbourg,
prénom.nom@engees.unistra.fr

**LIRMM, Université de Montpellier 2 et CNRS, Montpellier,
prénom.nom@lirmm.fr

Résumé. L'Analyse Relationnelle de Concepts (ARC) est une extension de l'Analyse Formelle de Concepts (AFC), une méthode de classification non supervisée d'objets sous forme de treillis de concepts. L'ARC supporte en plus la gestion de relations entre objets de différents contextes ce qui permet d'établir des liens entre les concepts des différents treillis. Cette particularité lui permet d'être plus intuitive à utiliser pour extraire des connaissances à partir de données relationnelles et de donner des résultats plus riches. Malheureusement lorsque les jeux de données présentent de nombreuses relations, les résultats obtenus sont difficilement exploitables et des problèmes de passages à l'échelle se posent. Nous proposons dans cet article une adaptation possible de l'ARC pour explorer les relations de manière supervisée pour augmenter la pertinence des résultats obtenus et réduire le temps de calcul. Nous prenons pour exemple des données hydrobiologiques ayant trait à la qualité des milieux aquatiques.

1 Introduction

L'Analyse Formelle de Concepts (Ganter et Wille, 1999), notée de manière abrégée AFC, est une méthode de classification automatique d'objets décrits par des attributs au travers d'une relation binaire. Le résultat d'une telle classification est un treillis de concepts (appelé aussi treillis de Galois (Barbut et Monjardet, 1970)) où chaque concept regroupe tous les objets ayant en commun un ensemble d'attributs. On peut naviguer dans le treillis de manière simple et intuitive, des concepts les plus spécifiques (les concepts regroupant beaucoup de caractéristiques partagées par peu d'objets) aux moins spécifiques (les concepts regroupant beaucoup d'objets partageant peu de caractéristiques).

L'AFC est exploitée dans différents domaines en tant que méthode d'extraction de connaissances et les différentes publications sur le sujet, notamment Carpineto et Romano (2004); Valtchev et al. (2004), ont permis d'en identifier les forces et les limites. Certaines de ces limites ont pu être contournées en utilisant différentes approches.

L'Analyse Relationnelle de Concepts (ARC) (Huchard et al., 2007) est une extension de l'AFC qui permet de prendre en compte non seulement les caractéristiques des objets, mais aussi les relations que les objets entretiennent entre eux. L'ARC consiste à appliquer itérativement un algorithme de l'AFC pour gérer les données relationnelles : les objets sont décrits

par des attributs et par leurs relations vers d'autres objets. Les concepts découverts à une itération donnée sont propagés le long des relations, pour permettre la découverte de nouveaux concepts à l'itération suivante. L'ARC apparaît plus intuitive à utiliser sur des données relationnelles telles que des bases de données ou des langages de modélisation orientés objet comme UML. Nous proposons dans cet article une adaptation de l'ARC en vue de l'utiliser en tant que méthode d'extraction de connaissances sur des données de mesure de qualité de l'eau des cours d'eau d'Alsace.

Ce travail s'inscrit dans le projet ANR FRESQUEAU¹ dont le but est le développement de nouvelles méthodes d'étude, de comparaison et d'exploitation de tous les paramètres disponibles sur les cours d'eau et les plans d'eau. Il approfondit une étude précédente menée avec l'AFC (Bertaux et al., 2009a) et des approches statistiques (Bertaux et al., 2009b).

Propager le long des relations les concepts découverts d'une itération à une autre permet certes la découverte de concepts intéressants, mais engendre souvent une explosion combinatoire, et les motifs intéressants sont difficiles à extraire à partir du grand ensemble de concepts construits. Plusieurs stratégies peuvent être utilisées pour pallier cette complexité, incluant la séparation des objets initiaux en plusieurs sous-ensembles après une analyse préliminaire ou l'introduction de requêtes (Azmeh et al., 2011). Nous nous intéressons dans cet article à l'utilisation de l'ARC pour explorer interactivement les données en laissant l'utilisateur/trice choisir avant chaque itération de l'AFC quels contextes (formels et relationnels) il ou elle veut utiliser.

Les données avec lesquelles nous travaillons ne sont pas initialement sous forme de relation binaire mais de nombreux travaux traitent de l'échelonnement des données en vue d'obtenir une relation binaire (Ganter et Wille, 1999) ou une représentation sous forme de structures de patron (Ganter et Kuznetsov, 2001). Ces approches ont été appliquées précédemment sur des données similaires à celles de notre projet (Bertaux et al., 2009a) et nous considérons par la suite uniquement des données sous forme de relations binaires.

Dans cet article nous allons tout d'abord présenter l'AFC puis le principe général du processus d'ARC pour mettre en évidence les différents points de variations qui permettraient d'en améliorer l'utilisation dans un contexte de fouilles de données. Nous présenterons un exemple du type de données que nous avons dans le projet FRESQUEAU et les conséquences des variations sur ces données. Nous concluons ensuite par une courte discussion.

2 Analyse Formelle de Concepts

L'AFC telle que présentée par Ganter et Wille (1999) a pour but de classifier un ensemble d'objets décrits par des attributs, et présentés sous la forme d'un contexte formel. Un contexte formel \mathcal{K} est un triplet (O, A, I) où O est un ensemble d'objets, A un ensemble d'attributs et $I \subseteq O \times A$ est la relation d'incidence entre O et A telle que $(o, a) \in I$ si et seulement si a est un attribut de o . Le tableau 1 représente un contexte formel. L'ensemble des objets est ici un ensemble d'identifiants de stations de prélèvements dans différents cours d'eau. Chaque station est représentée par une ligne. Les attributs correspondent à des caractéristiques de description des différents cours d'eau. L'association entre une station et une caractéristique du cours d'eau dans lequel elle se trouve est représentée par une croix. Ainsi la station identi-

1. <http://engees-fresqueau.unistra.fr/>

	petit cours	gros cours	eau fraîche et calme	eau fraîche et vive	cours phréatique
BREI0001	x			x	
BRUMB001	x		x		
BRUN001					x
BRUN002					x
DOLL001	x			x	
FECH001		x		x	

TAB. 1 – Exemple de contexte formel. Les objets sont représentés par les lignes et les attributs par les colonnes.

fiée par BREI0001 est située dans un petit cours d'eau fraîche et vive. Les stations BRUN001 et BRUN002 sont situées sur la même rivière mais à différents endroits.

L'application de l'AFC sur un contexte $\mathcal{K} = (O, A, I)$ engendre la création de concepts. Un concept est un couple (X, Y) où $X \subseteq O$ et $Y \subseteq A$ tel que $X = \{o \in O \mid \forall a \in Y, (o, a) \in I\}$ et $Y = \{a \in A \mid \forall o \in X, (o, a) \in I\}$. On appelle X l'*extension* du concept et Y l'*intension* du concept. De manière informelle l'extension d'un concept représente l'ensemble maximal des objets possédant l'ensemble des attributs de l'intension et l'intension d'un concept représente l'ensemble maximal des attributs possédés par l'ensemble des objets de l'extension.

L'AFC permet la création de tous les concepts d'un contexte donné, concepts formant un treillis de concepts aussi appelé treillis de Galois. Un concept c_1 est plus général (resp. plus spécifique) qu'un concept c_2 si l'extension de c_1 contient (resp. est contenue par) l'extension de c_2 . De manière duale l'intension d'un concept est contenue par l'intension d'un concept plus spécifique. Deux concepts donnés ont une unique borne supérieure et une unique borne inférieure.

Les treillis sont généralement représentés par leur diagramme de Hasse. Le treillis du tableau 1 est représenté par la figure 1. Les flèches représentent la relation de généralisation, c'est-à-dire que le concept pointé est plus général que le concept d'origine. Compte-tenu de l'inclusion de l'intension d'un concept dans les concepts plus spécifiques et de l'inclusion de l'extension d'un concept dans les concepts plus généraux, chaque objet (resp. attribut) n'est indiqué qu'une fois dans le concept le plus spécifique (resp. le plus général) le contenant. Par exemple le concept 6 regroupe les stations BREI0001 et DOLL001 qui possèdent les attributs petit cours et eau fraîche et vive que l'on retrouve par la relation de généralisation vers les concepts 2 et 7.

3 Extension de l'ARC pour l'analyse exploratoire

L'Analyse Relationnelle de Concepts (ARC) (Huchard et al., 2007) est une extension de l'AFC permettant de prendre en compte, en plus des caractéristiques des objets, les relations existant entre ces objets.

L'algorithme 1 présente les principales étapes de l'ARC. Le paramètre d'entrée de l'ARC est une famille relationnelle de contextes $RCF = (K, R)$ composée de n contextes objet-

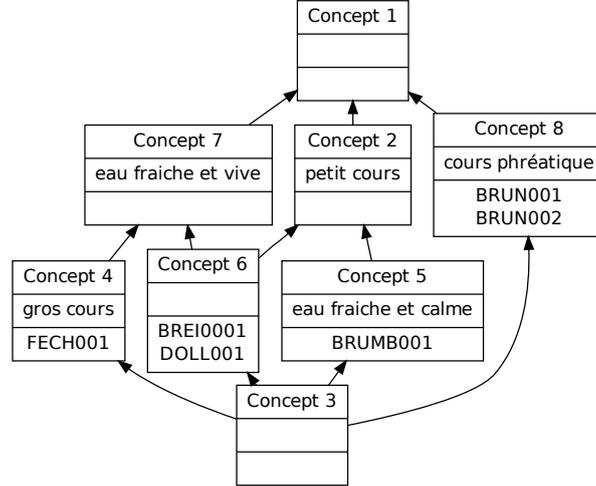


FIG. 1 – Diagramme de Hasse du treillis de concepts issu du contexte décrit par le tableau 1.

attribut $\mathcal{K}_i = (O_i, A_i, I_i)$, $i \in [1..n]$, et m contextes objet-objet $\mathcal{R}_j = (O_k, O_l, I_j)$, $j \in [1..m]$ où O_k et O_l sont les ensembles d'objets de \mathcal{K}_k et \mathcal{K}_l . On peut voir dans le tableau 2 un exemple de famille relationnelle de contextes. On peut voir sur la partie gauche deux contextes objet-attribut taxons et stations et sur la droite le contexte objet-objet `presenceDeTaxon` qui relie les objets du contexte stations aux objets de taxons².

```

1: proc MULTI-FCA( Entrée: (K, R) une RCF,
2: Out: L tableau [1..n] de treillis)
3:  $p \leftarrow 0$ ; arrêt  $\leftarrow$  faux
4: pour  $i$  de 1 à  $n$  faire
5:    $\mathbf{L}^0[i] \leftarrow$  CONTRUIRE-TREILLIS ( $\mathcal{K}_i^0$ )
6: tant que non arrêt faire
7:    $p++$ 
8:   pour  $i$  de 1 à  $n$  faire
9:      $\mathcal{K}_i^p \leftarrow$  ETENDRE-REL ( $\mathcal{K}_i^{p-1}$ ,  $\mathbf{L}^{p-1}$ )
10:     $\mathbf{L}^p[i] \leftarrow$  MAJ-TREILLIS ( $\mathcal{K}_i^p$ ,  $\mathbf{L}^{p-1}[i]$ )
11:  arrêt  $\leftarrow \bigwedge_{i=1..n}$  ISOMORPHE( $\mathbf{L}^p[i]$ ,  $\mathbf{L}^{p-1}[i]$ )

```

Algorithme 1: Processus de l'Analyse Relationnelle de Concepts.

Pour $\mathcal{R}_j \subseteq O_k \times O_l$, nous appelons O_k le domaine et O_l l'image. L'étape d'initialisation (lignes 4-5) consiste à construire, pour tout $i \in [1..n]$, le treillis $\mathbf{L}^0[i]$ associé au contexte \mathcal{K}_i . La figure 2 présente les deux treillis obtenus après cette étape d'initialisation sur notre exemple. On remarque que la relation `presenceDeTaxons` n'est pas encore prise en compte à cette étape du processus et les deux treillis sont indépendants.

À l'étape p :

2. Le terme taxon recouvre les différents qualificatifs utilisés pour la dénomination des êtres vivants qu'il s'agisse d'espèces, de genres ou de familles.

Contextes objet-ottribut				Contextes objet-objet																																															
<table border="1"> <thead> <tr> <th>taxons</th> <th>≤ 1 an</th> <th>> 1 an</th> <th></th> </tr> </thead> <tbody> <tr> <td>Athericidae</td> <td>x</td> <td></td> <td></td> </tr> <tr> <td>Bithynia</td> <td>x</td> <td>x</td> <td></td> </tr> <tr> <td>Boreobdella</td> <td></td> <td>x</td> <td></td> </tr> </tbody> </table>				taxons	≤ 1 an	> 1 an		Athericidae	x			Bithynia	x	x		Boreobdella		x		<table border="1"> <thead> <tr> <th>presenceDeTaxon</th> <th>Atheri-cidae</th> <th>Bithy-nia</th> <th>Boreob-della</th> <th></th> <th></th> <th></th> </tr> </thead> <tbody> <tr> <td>BREI0001</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td>x</td> </tr> <tr> <td>BRUN001</td> <td>x</td> <td>x</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>FECH001</td> <td>x</td> <td></td> <td></td> <td></td> <td></td> <td>x</td> </tr> </tbody> </table>				presenceDeTaxon	Atheri-cidae	Bithy-nia	Boreob-della				BREI0001						x	BRUN001	x	x					FECH001	x					x
taxons	≤ 1 an	> 1 an																																																	
Athericidae	x																																																		
Bithynia	x	x																																																	
Boreobdella		x																																																	
presenceDeTaxon	Atheri-cidae	Bithy-nia	Boreob-della																																																
BREI0001						x																																													
BRUN001	x	x																																																	
FECH001	x					x																																													
<table border="1"> <thead> <tr> <th>stations</th> <th>petit cours</th> <th>eau fraiche et vive</th> <th>cours phréatique</th> </tr> </thead> <tbody> <tr> <td>BREI0001</td> <td>x</td> <td>x</td> <td></td> </tr> <tr> <td>BRUN001</td> <td></td> <td></td> <td>x</td> </tr> <tr> <td>FECH001</td> <td></td> <td>x</td> <td></td> </tr> </tbody> </table>				stations	petit cours	eau fraiche et vive	cours phréatique	BREI0001	x	x		BRUN001			x	FECH001		x																																	
stations	petit cours	eau fraiche et vive	cours phréatique																																																
BREI0001	x	x																																																	
BRUN001			x																																																
FECH001		x																																																	

TAB. 2 – Exemple de famille de contextes relationnelle. Les objets sont représentés par les lignes et les attributs par les colonnes.

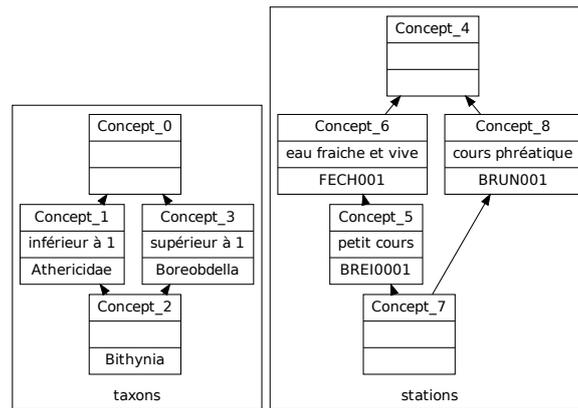


FIG. 2 – Famille de treillis issue de la famille relationnelle de contextes du tableau 2 après l'étape d'initialisation 0.

stations	petit cours	eau fraiche et vive	cours phréatique	\exists presenceDeTaxon			
				Concept_0	Concept_1	Concept_2	Concept_3
BREI0001	x	x		x			x
BRUN001			x	x	x	x	x
FECH001		x		x	x		x

TAB. 3 – Échelonnement de la relation presenceDeTaxons et extension du contexte stations à l'étape 1.

(seule une borne est connue) et peut être très grand selon les données.

Les treillis relationnels s'interprètent différemment des treillis de concepts classiques car ils doivent être considérés ensemble. Le treillis `stations` de la figure 3 doit être considéré avec le treillis `taxons` pour pouvoir être correctement interprété. On trouve dans les intensions de concepts des attributs faisant référence à d'autres concepts. Par exemple `Concept_8` possède l'attribut relationnel `presenceDeTaxons` : `Concept_2` ce qui signifie que tous les objets de `Concept_8` sont liés par la relation `presenceDeTaxons` à au moins (car l'opérateur d'échelonnement utilisé est l'opérateur \exists) un objet de l'extension de `Concept_2` appartenant au treillis `taxons`.

L'avantage d'un tel processus est que les concepts obtenus ont dans leur intension des relations à d'autres concepts en plus des attributs classiques. Ces relations permettent l'extraction de motifs construits à partir de plusieurs contextes interconnectés, comme cela a été fait dans Dolques et al. (2009) et Dolques et al. (2010), qui ne pourraient pas être facilement obtenus à partir du processus classique de l'AFC.

Cependant un problème majeur de ce type de processus est la difficulté potentielle à appréhender le résultat. Dans des travaux précédents en ingénierie dirigée par les modèles, les données extraites de modèles de taille moyenne peuvent être facilement appréhendées par l'ARC. Mais dans un contexte de fouille de données, la taille des données est beaucoup plus importante. Le temps de calcul est dépendant du nombre de concepts à générer et celui-ci est exponentiel par rapport au nombre minimum des attributs ou des objets dans le pire des cas. Ainsi, si les relations entre objets sont nombreuses et ont peu de similarité d'un objet à l'autre, le temps de calcul peut augmenter de manière exponentielle et le résultat peut apparaître difficile à comprendre par un utilisateur à cause du nombre de concepts à considérer simultanément. Ceci est particulièrement avéré quand seuls de petits motifs sont nécessaires alors que de nombreuses relations connectent les objets entre eux et que ces relations forment un circuit. Dans de tels cas, nous pensons qu'il peut être plus pertinent d'avoir une approche exploratoire.

Nous listons ci-dessous les différentes variations possibles sur l'algorithme pour mettre en pratique une approche exploratoire. Nous énumérons les points de variation possibles dans l'algorithme qui peuvent affecter le résultat en changeant les contextes pris en compte à chaque étape. Nous proposons pour chaque point de variation un scénario alternatif à partir du processus précédemment décrit qui implique l'utilisateur en lui demandant d'effectuer des choix. Toutes ces variations ou seulement un sous-ensemble peuvent être appliquées selon la granularité voulue.

- **étape d'initialisation, lignes 4 à 5** Construire les treillis pour des contextes objet-attribut choisis concaténés à des contextes objet-objet choisis.
- **ÉTENDRE-REL, ligne 9** Plutôt que d'utiliser toutes les relations et d'échelonner toutes les relations objet-objet à chaque étape, sélectionner un sous-ensemble de la famille relationnelle de contextes et différents opérateurs d'échelonnement pour chaque contexte objet-objet sélectionné. Note : les treillis pour les images des relations objet-objet sélectionnées auront dû être calculés lors d'une étape précédente (mais pas nécessairement $p - 1$). À cette étape, des contextes objet-attribut peuvent aussi être sélectionnés et le treillis correspondant peut être construit.
- **MAJ-TREILLIS, ligne 10** Mettre à jour seulement les treillis pour les relations sélectionnées.
- **arrêt, ligne 11** Si un point fixe n'est pas atteint, laisser la décision d'arrêter à l'expert.

4 Exemple d'exploration

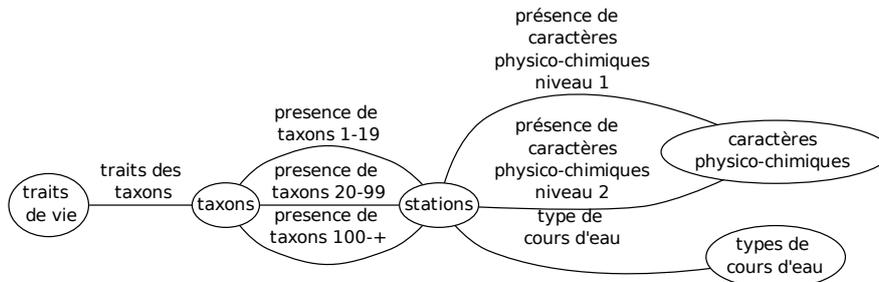


FIG. 4 – Schéma des données utilisées par notre exemple.

Dans cette section nous illustrons le potentiel d'une approche exploratoire avec l'exemple de cours d'eau alsaciens. Chaque cours d'eau est classifié en fonction de types généraux (par exemple un petit cours d'eau avec eau fraîche et vive). Pour évaluer la qualité des eaux d'un cours d'eau, les hydroécologues en sélectionnent des segments, appelés stations, sur lesquels ils effectuent des prélèvements et des mesures ainsi que l'échantillonnage des plantes et animaux présents (appelés taxons). Ces prélèvements et mesures respectent un ensemble de normes. Après analyse ou détermination en laboratoire, les stations d'un cours d'eau sont donc décrites par différents attributs quantitatifs : des données biologiques (par exemple le nombre d'individus appartenant aux différents taxons) d'une part et des données physico-chimiques (par exemple le pH, la température, le taux de matières organiques, d'oxygène dissous ...) d'autre part. Les taxons, quant à eux, sont caractérisés par des traits de vie, qui sont des données qualitatives, par exemple la durée de vie des invertébrés.

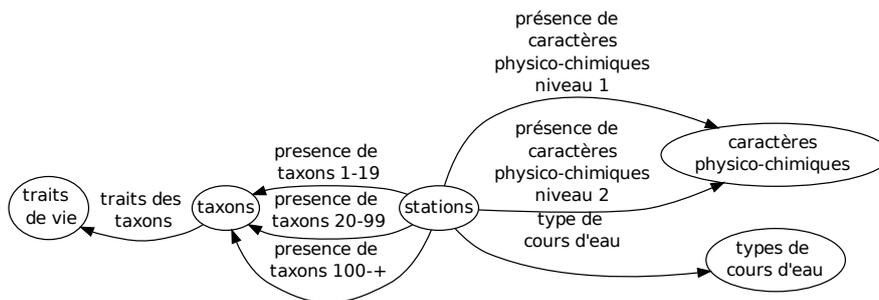


FIG. 5 – Schéma de données pouvant être utilisé par l'ARC.

La figure 4 représente les données par un schéma. Chaque nœud représente un ensemble d'objets qui est traduit en un contexte objet-attribut. Chaque arête étiquetée représente une relation entre les ensembles d'objets qui est traduite en un contexte objet-objet. Pour chaque arête nous considérons la relation dans les deux directions.

ARC pour l'exploration de données relationnelles

Concept_63. Concept_63 regroupe de son côté les taxons dont les individus ont une durée de vie supérieure à un an. Concept_41 regroupe les stations dont l'un des caractères physico-chimiques (de niveau 2, donc élevé) est DCO. On en tire l'implication déduite. À partir du treillis complet on peut obtenir l'ensemble des règles d'implication entre traits de vie et caractères physico-chimiques en considérant tous les cas où des caractères physico-chimiques sont introduits par un concept et des traits de vie sont introduits par un concept plus spécifique.

Mais des règles suivant le format « *la modalité M du trait de vie T peut apparaître quand le caractère physico-chimique C est présent* » peuvent aussi être pertinentes. Afin qu'elles puissent émerger, il serait nécessaire de changer (par rapport à la configuration précédente) la direction des relations entre traits de vie et taxons et entre taxons et stations. Il existe encore de nombreuses autres configurations dont les résultats sont potentiellement pertinents et faire varier les opérateurs d'échelonnement permet d'augmenter encore l'expressivité des règles que l'on peut obtenir grâce à l'ARC.

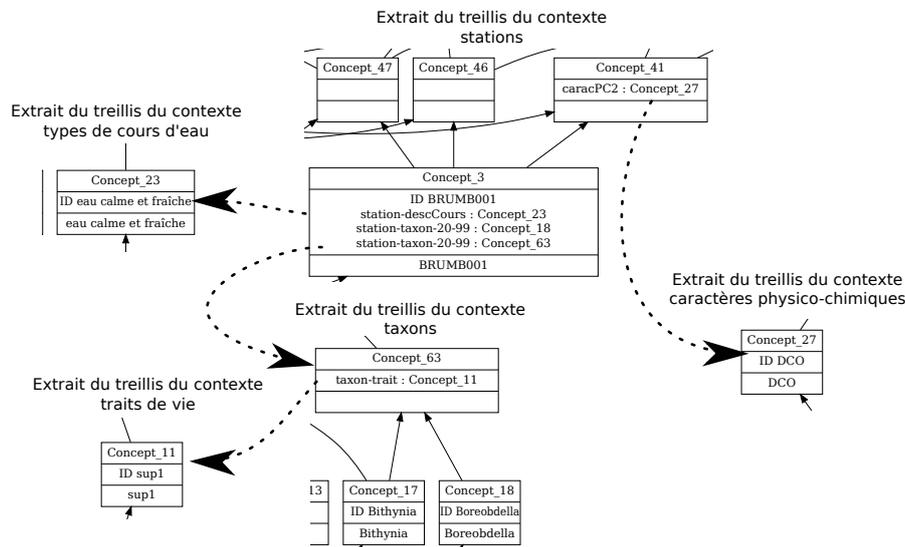


FIG. 6 – Extraits des treillis obtenus à partir de la famille relationnelle de contextes de la table 4.

Si nous considérons le schéma de la figure 4 comme un graphe, l'exploration consiste à analyser les différentes arêtes jusqu'à obtenir un chemin entre les caractères physico-chimiques et les traits de vie. En comparant les résultats obtenus lors d'une exploration avec ceux obtenus en utilisant l'ARC classique sur les mêmes relations, nous trouvons que les treillis résultants sont plus petits et plus faciles à lire dans le premier cas.

Combiner l'ensemble des configurations possibles n'est pas envisageable car la multiplication des relations risque d'entraîner une explosion combinatoire du nombre de concepts obtenus, augmentant ainsi le temps de calcul et la complexité des concepts obtenus. Dans le cas de notre exemple, considérer toutes les relations dans les deux directions échelonnées seulement par \exists entraîne la création de 120 concepts contre 66 et 63 pour les deux configurations évoquées précédemment. Nous envisageons donc une approche où l'utilisateur/trice explore

différentes configurations en effectuant différents choix à chaque étape du processus comme présenté dans la section 1.

5 Conclusion et discussion

Dans cet article, nous avons présenté une approche exploratoire pour assister l'utilisation de l'Analyse Relationnelle de Concepts de manière plus appropriée pour un processus d'extraction de connaissances. Nous avons plusieurs raisons de vouloir modifier le processus original de l'ARC : obtenir des résultats pertinents plus rapidement en calculant moins de treillis (de préférence seulement les treillis qui nous intéressent), diminuer la complexité de la fouille de données relationnelles, ou laisser l'expert guider le processus de découverte en se basant sur son intuition et les motifs d'apprentissage qui apparaissent au cours du processus.

Plusieurs questions se posent sur cette approche d'extraction de concepts à partir de données relationnelles. L'étape d'initialisation a un fort impact sur les structures qui peuvent être découvertes par la suite. Elle peut accélérer le processus, si les relations objet-objet contiennent les informations nécessaires à l'expert, ou à l'inverse, elle peut cacher à l'expert les informations pertinentes. Néanmoins, le problème le plus important vient du fait que les modifications à chaque étape rendent la construction de concepts non monotone et qu'il est possible de construire des exemples où le processus diverge (itérant sur plusieurs configurations récurrentes).

Dans le processus original de l'ARC, quand le point fixe est atteint, les treillis des deux dernières étapes sont isomorphes, ainsi quand un concept en référence un autre *via* un attribut relationnel, le concept référencé peut être trouvé dans un treillis de la même étape. Mais dans le processus exploratoire que nous proposons, quand un concept en référence un autre, le concept référencé est dans un treillis de l'étape précédente et peut référencer un concept lui-même dans une étape antérieure. Il est alors nécessaire de trouver des solutions pour présenter aux experts une information suffisamment simple à interpréter. Nous pensons malgré tout qu'une telle approche exploratoire est plus applicable qu'une approche systématique qui itère jusqu'à atteindre un point fixe et qui donne des résultats difficilement interprétables par un expert.

Remerciements

Nous souhaitons remercier les relecteurs anonymes pour leur retours constructifs, les hydrobiologistes et en particulier C. Grac (ENGEES-LHYGES) pour leur expertise sur les données fournies et le projet FRESQUEAU ANR11_MONU14 qui a financé en partie ces travaux.

Références

- Azmeh, Z., M. Huchard, A. Napoli, M. R. Hacene, et P. Valtchev (2011). Querying relational concept lattices. In *Proc. of the 8th Intl. Conf. on Concept Lattices and their Applications (CLA'11)*, pp. 377–392.
- Barbut, M. et B. Monjardet (1970). *Ordre et Classification : Algèbre et Combinatoire*, Volume 2. Hachette.

- Bertaux, A., F. Le Ber, A. Braud, et M. Trémolières (2009a). Identifying ecological traits : a concrete fca-based approach. In S. Ferré et S. Rudolph (Eds.), *7th International Conference on Formal Concept Analysis, ICFCA 2009, Darmstadt*, LNAI 5548, pp. 224–236. Springer-Verlag.
- Bertaux, A., F. Le Ber, P. Li, et M. Trémolières (2009b). Combiner treillis de Galois et analyse factorielle multiple pour l'analyse de traits biologiques. In G. d'Aubigny (Ed.), *Actes des XVIèmes Rencontres de la Société Francophone de Classification, Grenoble*, pp. 117–120.
- Carpineto, C. et G. Romano (2004). *Concept Data Analysis : Theory and Applications*. Wiley.
- Dolques, X., M. Huchard, et C. Nebut (2009). From transformation traces to transformation rules : Assisting model driven engineering approach with formal concept analysis. In *Supplementary Proceedings of ICCS'09*, pp. 15–29.
- Dolques, X., M. Huchard, C. Nebut, et P. Reitz (2010). Fixing generalization defects in UML use case diagrams. In *CLA'10 : 7th International Conference on Concept Lattices and Their Applications*, pp. 247–258.
- Ganter, B. et S. O. Kuznetsov (2001). Pattern structures and their projections. In *Proc. of the 9th Int. Conf. on Conceptual Structures (ICCS 2001)*, pp. 129–142.
- Ganter, B. et R. Wille (1999). *Formal Concept Analysis, Mathematical Foundations*. Springer.
- Huchard, M., M. R. Hacène, C. Roume, et P. Valtchev (2007). Relational concept discovery in structured datasets. *Ann. Math. Artif. Intell.* 49(1-4), 39–76.
- Valtchev, P., R. Missaoui, et R. Godin (2004). Formal concept analysis for knowledge and data discovery : New challenges. In *Proc. of the 2nd Intl. Conf. on Formal Concept Analysis (ICFCA'04)*, pp. 352–371.

Summary

Relational Concept Analysis (RCA) is an extension to the Formal Concept Analysis (FCA) which is an unsupervised classification method producing concept lattices. In addition RCA considers relations between objects from different contexts that allow for the creation of links between lattices. This feature makes it more intuitive to extract knowledge from relational data and gives richer results. However, data with many relations imply scalability problems and results that are difficult to exploit. We propose in this article a possible adaptation of RCA to explore relations in a supervised way in order to increase the performance and the pertinence of the results.