HAL open science

AOC-posets: a scalable alternative to Concept Lattices for Relational Concept Analysis

Xavier Dolques, Florence Le Ber, Marianne Huchard

To cite this version:

Xavier Dolques, Florence Le Ber, Marianne Huchard. AOC-posets: a scalable alternative to Concept Lattices for Relational Concept Analysis. CLA: Concept Lattices and their Applications, Oct 2013, La Rochelle, France. pp.129-140. hal-00916850

HAL Id: hal-00916850

https://hal.science/hal-00916850v1

Submitted on 10 Dec 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

AOC-posets: a scalable alternative to Concept Lattices for Relational Concept Analysis

Xavier Dolques[1], Florence Le Ber[1], and Marianne Huchard[2]

(1) ICube, Université de Strasbourg/ENGEES, CNRS, Strasbourg,
{xavier.dolques , florence.leber}@engees.unistra.fr
(2) LIRMM, Université de Montpellier 2 et CNRS
huchard@lirmm.fr

Abstract. Relational Concept Analysis (RCA) is a useful tool for classification and rule discovery on sets of objects with relations. Based on FCA, it produces more results than the latter but also an increase in complexity. Besides, in numerous applications of FCA, AOC-posets are used rather than lattices in order to reduce combinatorial problems. An AOC-poset is a subset of the concept lattice considering only concepts introducing an object or an attribute. AOC-posets are much smaller and easier to compute than concept lattices and still contain the information needed to rebuild the initial data. This paper introduces a modification of the RCA process based on AOC-posets rather than concept lattices. This work is motivated by a big set of relational data on river streams to be analysed. We show that using AOC-poset on these data provides a reasonable concept number.

1 Introduction

Relational Concept Analysis (RCA) [1] is based on iterative use of the classical Formal Concept Analysis algorithm to handle relational data: formal objects are described with formal attributes as in FCA, and with their relationships with other formal objects. Because RCA groups formal objects using relationships to formal objects at any distance, it often comes with a combinatorial explosion, and patterns of interest are difficult to extract from the huge set of built concepts. Various strategies can be used to cope with this complexity, including separating the initial formal object sets into smallest ones after a first analysis, or introducing queries [2]. Here we propose to adapt the RCA process in order to use only posets of concepts introducing objects or attributes (AOC-poset) rather than to build full concept lattices at each step of the process. Indeed AOC-posets are smaller and easier to compute than concept lattices [3] and their sets of concepts have interesting properties for extracting implication rules.

The context of this research is the FRESQUEAU project[1] which aims at developing new methods for studying, comparing and exploiting all the parameters available concerning streams and water areas. The data we deal with have been

[1] http://engees-fresqueau.unistra.fr/

© Manuel Ojeda-Aciego, Jan Outrata (Eds.): CLA 2013, pp. 129–140,
ISBN 978–2–7466–6566–8, Laboratory L3i, University of La Rochelle, 2013.

collected during 3 years over 40 sites in the Alsace plain. Two other data sets are available on the Rhin-Meuse district (North-east of France, 3400 sites, 10 years) and on the Rhône-Méditerranée district (South-east of France, 18000 sites, 40 years). In a previous work, FCA has been used for classifying characteristics of aquatic species [4]. RCA is now used for discovering relationships between those species characteristics and the characteristics of their habitat.

To perform this analysis, we propose to adapt RCA relying on AOC-poset. We show that this approach provides a reasonable concept number in our case.

The paper is organized as follows. Section 2 gives some useful definitions for our presentation. Section 3 details the RCA process based on AOC-poset. Section 5 presents the data set and results that can be obtained with AOC-posets. Related work is presented in section 6 and section 7 concludes the paper opening some perspectives of this work.

2 FCA Basics

Formal Concept Analysis (FCA) [5] aims at extracting an ordered set of concepts from a dataset, called a Formal Context, composed of objects described by attributes. We will denote by $\mathcal{K} = (\mathcal{G}, \mathcal{M}, \mathcal{I})$ a formal context, where $\mathcal{I} \subseteq \mathcal{G} \times \mathcal{M}$. Table 1 is a Formal Context $\mathcal{K}_{Animals} = (\mathcal{G}_{Animals}, \mathcal{M}_{Animals}, \mathcal{I}_{Animals})$ which

Table 1. Formal Context of animals described by their life traits $\mathcal{K}_{Animals}$

	size (1-2cm) (**att0**)	resistance to dessication (**att1**)	Micro-habitat in sand (**att2**)	reproduction clutches terrestrial (**att3**)	transversal distribution in lakes (**att4**)	transversal distribution on banks (**att5**)
Anc.	×					×
Ani.	×	×				×
Ano.			×			
Ant.	×		×	×		
Aph.	×		×		×	
Ase.	×				×	×
Ath.	×		×	×	×	

describes animals by characteristics they may own. The considered animals here are macro-invertebrates that can be found in rivers. We took the examples of the following kinds of animals : *Ancylus (Anc.), Anisus (Ani.), Anodonta (Ano.), Anthomyiidae (Ant.), Aphelocheirus (Aph.), Asellus (Ase.) and Athericidae (Ath.).*

Given a $\mathcal{K} = (\mathcal{G}, \mathcal{M}, \mathcal{I})$ formal context, a formal concept is a $C = (Extent(C), Intent(C))$ pair where:

$Extent(C) = \{g \in \mathcal{G} | \forall m \in Intent(C), (g,m) \in \mathcal{I}\}$ is the extent of the concept, $Intent(C) = \{m \in \mathcal{M} | \forall g \in Extent(C), (g,m) \in \mathcal{I}\}$ is the intent of the concept. Given two formal concepts $C_1 = (E_1, I_1)$ and $C_2 = (E_2, I_2)$ of $\mathcal{K}$, the concept specialization order $\leq_s$ is defined by $C_1 = (E_1, I_1) \leq_s C_2 = (E_2, I_2)$ if and only if $E_1 \subseteq E_2$ (and equivalently $I_2 \subseteq I_1$).

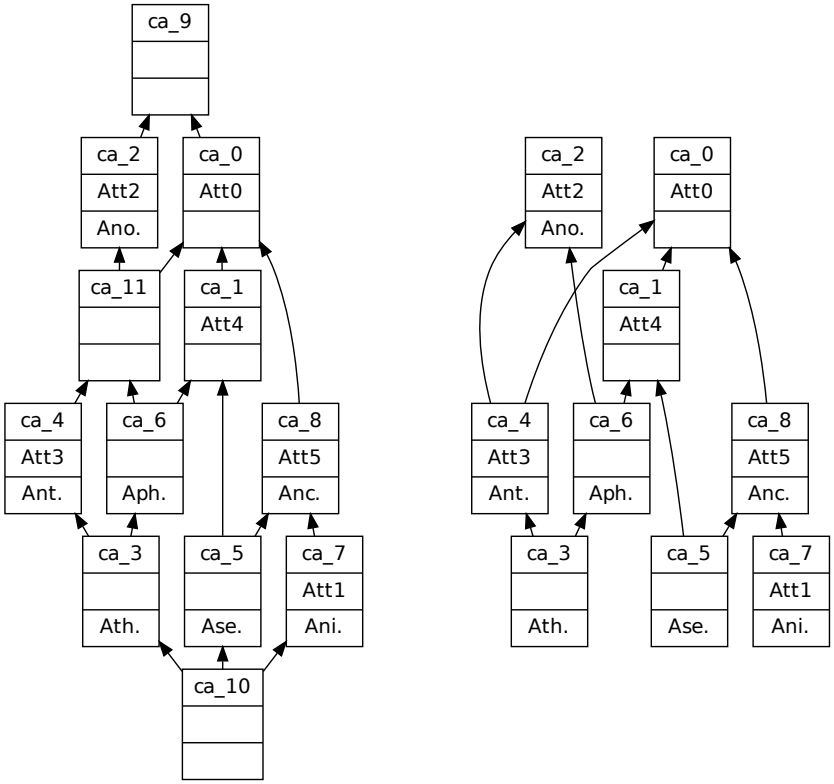

Fig. 1. Concept Lattice of animals (left), AOC poset $AOC_{animals}$ (right)

For simplicity's sake, most lattice representations show attributes (*resp.* objects) solely where they are introduced. They are said to show the simplified intents and simplified extents which will be denoted respectively $Intent_S(C)$ and $Extent_S(C)$ for a given concept C.

Let $\mathcal{C}_\mathcal{K}$ be the set of all concepts of a $\mathcal{K}$ formal context. This set of concepts provided with the ($\mathcal{C}_\mathcal{K}$, $\leq_s$) specialization order has a lattice structure, and is called the concept lattice associated with $\mathcal{K}$.

Left-hand side of Fig. 1 shows the concept lattice associated with the formal context of Table 1. In several FCA applications (*e.g.* in [3,6]; only the *object concepts* (which introduce at least one object) and the *attribute concepts* (which introduce at least one attribute) are used. In Fig. 1 (left-hand side), ca_5 and

ca_8 are examples of object concepts; ca_1 and ca_2 are examples of attribute concepts; ca_9, ca_10 and ca_11 do not introduce any object or any attribute.

The so-called AOC-poset (for Attribute-Object-Concept poset) is the sub-order of $(\mathcal{C}_{\mathcal{K}}, \leq_s)$ restricted to object-concepts and attribute-concepts. It is also called Galois Sub-Hierarchy, a term we consider less explicit. Right-hand side of Fig. 1 shows the AOC-poset for the context of Table 1. One may have a large difference of complexity between the two structures, because the concept lattice may have $2^{min(|\mathcal{G}|,|\mathcal{M}|)}$ concepts, while the number of concepts in the AOC-poset is bounded by $|\mathcal{G}| + |\mathcal{M}|$.

3 A variant of Relational Concept Analysis with AOC-poset

Relational Concept Analysis aims at extending Formal Concept Analysis to take into account a dataset where objects of several categories are described by attributes and by relations to objects [1,7]. The dataset is called a Relational Context Family.

Definition 1 (Relational Context Family (RCF)). *A Relational Context Family (denoted RCF) is a* $(\mathbf{K}, \mathbf{R})$ *pair where:*

- $\mathbf{K} = \{\mathcal{K}_i\}_{i=1,\ldots,n}$ *is a set of* $\mathcal{K}_i = (G_i, M_i, I_i)$ *contexts*
- $\mathbf{R} = \{r_j\}_{j=1,\ldots,m}$ *is a set of* r_j *relations where* $r_j \subseteq G_{i_1} \times G_{i_2}$ *for some* $i_1, i_2 \in \{1, \ldots, n\}$.

An example of a RCF is composed of an Animal context, denoted $\mathcal{K}_{Animals}$ (Table 1), a Site context, denoted $\mathcal{K}_{Sites}$ (Table at the left-hand side of Fig. 2), and a *contains* relation, denoted $r_{contains}$ (Table 2). Sites are determined by locations on a river where samples are done; two kinds of samples are here considered: physico-chemical samples that measure physical and chemical properties of the water (*e.g.* temperature or presence of a chemical compound) and biological samples that measure the level of life.

Table 2. Relation $r_{contains}$

contains ↗	***Anc.***	***Ani.***	***Ano.***	***Ant.***	***Aph.***	***Ase.***	***Ath.***
Site0				×	×		×
Site1	×			×	×		×
Site2		×		×	×		×
Site3	×	×				×	

When objects of a category (*e.g.* Sites) are connected to objects of another category (*e.g.* Animals) *via* a relation (*e.g.* contains), concepts formed on top of objects of the latter category can be used to form concepts on top of objects

	NH4	SO4
Site0	×	×
Site1		×
Site2	×	×
Site3		

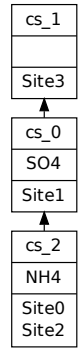

Fig. 2. Formal Context of Sites $\mathcal{K}_{Sites}$ (left), AOC-poset of Sites (right)

of the former category. Let us for example consider the AOC-poset of the right-hand side of Fig. 1 and the `ca_8` concept which groups animals with a transversal distribution on banks (Att5), namely *Ancylus (Anc.)*, *Anisus (Ani.)* and *Asellus (Ase.)*. Now Site1, Site2 and Site3 can be grouped because they contain at least one animal with Att5, that is at least one animal in $Extent$(`ca_8`), *e.g. Ancylus (Anc.)* for Site 1 (Table 2). In RCA, this relationship between objects of a category and concepts formed on objects of another category is implemented thanks to *scaling operators*. This results in the creation of special attributes called relational attributes. The most used *scaling operators* are:

- the *existential* scaling operator which encodes the fact that an object o is in relation by $\exists r$ with a concept C if $r(o)$ has a non-empty intersection with $Extent(C)$
- and the *strict universal* scaling operator which encodes the fact that an object o is in relation by $\forall r$ with a concept C if $r(o)$ is non-empty and included in the extent of C.

For a given r relation in a given analysis, its associated scaling operator is denoted by $\rho(r)$. Now, since we rely on AOC-posets, the existential scaling operator is defined slightly differently from [7]. Nevertheless the reference [7] can be read to have more details on the RCA process which is simplified below for space reasons.

Definition 2 (Existential scaling operator). *Let $\mathcal{K} = (G, M, I)$ be a context, and r a relation, where G is the domain of r, G_{i_r} is the range of r, and $\mathcal{K}_{i_r} = (G_{i_r}, M_{i_r}, I_{i_r})$ is another context. Let also $\mathcal{C}_{i_r}$ be a set of concepts*[2] *built on $\mathcal{K}_{i_r}$. The $\mathbb{S}_\exists$ existential scaling operator is applied to the context $\mathcal{K}$, denoted by $\mathbb{S}_\exists(\mathcal{K}, r, \mathcal{C}_{i_r}) = \mathcal{K}^+ = (G^+, M^+, I^+)$, with:*

- $G^+ = G$
- $M^+ = \{\exists r(C) \mid C \in \mathcal{C}_{i_r}\}$, *where each $\exists r(C)$ is a relational attribute*

[2] rather than a concept lattice built on $\mathcal{K}_{i_r}$ in [7].

– $I^+ = \{(o, \exists r(C)) \mid o \in G, C \in \mathcal{C}_{i_r}, r(o) \cap Extent(C) \neq \emptyset\}$

Table 3 shows $\mathbb{S}_\exists(K_{Sites}, contains, \mathcal{C}_{Animals})$ where $\mathcal{C}_{Animals}$ is the set of concepts of the AOC-poset of right-hand side of Fig. 1. In this case I^+ contains $(Site3, \exists\, contains(ca_5))$ and $(Site3, \exists\, contains(ca_7))$ because Site3 contains one animal from the extent of these two concepts.

Table 3. Existential Scaling of Formal Context of sites $\mathbb{S}_\exists(K_{Sites}, contains, \mathcal{C}_{Animals})$. **ca_i** is a short expression for $\exists\, contains(ca_i)$.

$\exists contains$	**ca_0**	**ca_1**	**ca_2**	**ca_3**	**ca_4**	**ca_5**	**ca_6**	**ca_7**	**ca_8**
Site0	×	×	×	×	×		×		
Site1	×	×	×	×	×		×		×
Site2	×	×	×	×	×		×	×	×
Site3	×	×				×		×	×

Then, for each $\mathcal{K}$ context of $\mathbf{K}$, the *apposition* of $\mathcal{K}$ (denoted by symbol '|') with the respective results of the scaling upon each r_j of $\mathbf{R}$ with G as domain $(1 \leq j \leq k)$ and the chosen $\rho(r_j)$, is used to build a new set of concepts (notations are taken from Def. 2). This apposition is the relational extension of the $\mathcal{K}$ context considering ρ and a set of concepts $\mathbf{C}$ which is a union of concept sets including $\mathcal{C}_{i_{r_j}}$, $1 \leq j \leq k$:

$$\mathbb{E}_{\rho,\mathbf{C}}(\mathcal{K}) = \mathcal{K} \mid \mathbb{S}_{\rho(r_1)}(\mathcal{K}, r_1, \mathcal{C}_{i_{r_1}}) \mid \ \ldots \mid \mathbb{S}_{\rho(r_k)}(\mathcal{K}, r_k, \mathcal{C}_{i_{r_k}})$$

By extension, $\mathbb{E}^*_{\rho,\mathbf{C}}(\mathbf{K})$ denotes the relational extension of $\mathbf{K}$, which is composed of all the relational extensions of all $\mathcal{K}_i$ in $\mathbf{K}$ (and $\mathbf{C}$ is a union of concept sets associated with all ranges of all relations).

$$\mathbb{E}^*_{\rho,\mathbf{C}}(\mathbf{K}) = \{\mathbb{E}_{\rho,\mathbf{C}}(\mathcal{K}_1), \ldots, \mathbb{E}_{\rho,\mathbf{C}}(\mathcal{K}_n)\}.$$

For example a relational extension of our $\mathbf{K}$ example is composed of Table 1 (no outgoing relation), and the left table of Fig. 2 apposed to Table 3. The AOC-poset built from this extended context is shown on figure 3.

Now a whole construction process consists in building a (possibly infinite) sequence of contexts and AOC-posets associated with $(\mathbf{K}, \mathbf{R})$ and ρ. The first set of contexts (step 0) is $\mathbf{K}^0 = \mathbf{K}$. The contexts of step p are used to build the associated AOC-posets. The $\mathbf{C}_p$ set composed of the sets of the concepts of AOC-posets at step p is used to calculate the relational extension. The set of contexts at step $p+1$ is defined using the relational extension: $\mathbf{K}^{p+1} = \mathbb{E}^*_{\rho,\mathbf{C}_p}(\mathbf{K})$.

Considering independently the contexts: for some $i \in \{1, \ldots, |\mathbf{K}|\}$ the i-th context at the step $p+1$ is the relational extension of the context of the same rank with the concepts from step p: $\mathcal{K}_i^0 = \mathcal{K}_i$, $\mathcal{K}_i^{p+1} = \mathbb{E}_{\rho,\mathbf{C}_p}(\mathcal{K}_i)$. The difference here with classical RCA is that we rely on $\mathcal{K}_i$ and concepts of the AOC-posets of step p to build $\mathcal{K}_i^{p+1}$, rather than on $\mathcal{K}_i^p$ and lattices of step p.

In our example, $\mathbf{K}^0 = (K_{Animals}, K_{Sites})$ and $\mathbf{R} = \{contains\}$ (see Table 2). If $\rho(contains) = \exists$, then

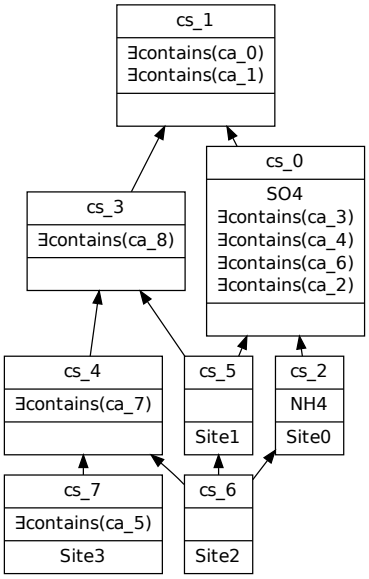

Fig. 3. AOC-posets for the K_{Sites} formal context and the $\mathcal{C}_{Animals}$ set of concepts, existential scaling

- At step 0, we obtain the AOC-posets of Fig. 1 (right) and Fig. 2 (right); $\mathbf{K}^1 = \{K_{Animals}, K_{Sites} \mid \mathbb{S}_\exists(K_{Sites}, contains, \mathcal{C}_{Animals})\}$.
- At step 1, we obtain the AOC-posets of Fig. 1 (right) and Fig. 3 (left).

In this simple example, the following steps do not produce any new concept (a fix-point is reached). This process will be denoted by RCA-AOC in the following.

4 Extracting Implication Rules

Implication rules are implications that are verified by the whole data set considered. Some implication rules can be extracted from the AOC-poset concepts by considering their simplified intent.

An attribute a from the simplified intent of a concept C is an attribute that is not contained in the intent of any concept more general than C *i.e.* the set of all the objects sharing a is the extent of C. The objects from the extent also all share the attributes of the full intent of C, but they may not be the only ones. By consequence, for every object o in the data set the presence of a for a given object o implies the presence of all the attributes from the intent of C.

For instance in Fig. 1 we can consider the concept `ca_7`. $Intent(\texttt{ca_7}) = \{Att1, Att0, Att5\}$ and $Intent_S(\texttt{ca_7}) = \{Att1\}$ which means that in all the dataset the rule $Att1 \rightarrow Att0 \wedge Att5$ is verified.

All the concepts with no empty simplified intent can be found in the AOC-poset meaning that all the rules having one element at the left hand side can be found with an AOC-poset.

Extracting rules from an existing AOC-poset is straightforward as it consists in reading the simplified intent and the full intent of each concept. The concept order permits to extract rules ordered by support, the rules extracted from the most general concepts having a larger support than more specific rules. The use of RCA permits to use relational attributes in implication rules.

5 A case study

We rely on a large database collecting data on Alsatian streams and water areas (North-east of France) [8], but more data are available through the current FRESQUEAU project, concerning larger areas and periods. The data are either issued from samples (*e.g.* physical, chemical and biological data collected on stream sites), synthetic data (*e.g.* biological indices, land cover) or general information issued from the literature (*e.g.* information about the aquatic species living in the streams). More precisely in this paper we work with three tables. The first one gives values of 27 physico-chemical parameters (*e.g.* temperature, pH, SO4, NH4, organic matters) collected on 49 stream sites. The second table gives the level of population for 197 macro-invertebrates (*e.g. Ancylus*, *Anisus*, *Anodonta*) collected on the same 49 sites. The third one describes the macro-invertebrates with 22 different life traits, *i.e.* their characteristics and functioning, *e.g* life cycle, reproduction mode, etc. each life trait being represented by several modalities (*e.g.* for the life trait *life cycle* there are two possible modalities : *less than a year* or *more than a year*). The sum of the modalities for all life traits is 116. We look for rules combining life traits and physico-chemical parameters, *e.g. "the M modality of the T life trait is associated with a high value of the C physico-chemical parameter"*.

We modeled our data within 4 formal contexts: *stream sites*, *physico-chemical parameters*, *life traits* and *macro-invertebrates* and we considered the three relations between them that are described by our tables: level of physico-chemical parameter, population of macro-invertebrates and life trait of macro-invertebrates. The relations being originally numerical ones we applied a preprocessing to separate values into a few classes. So the *level of physico-chemical parameter* relation has been split into 5 binary relations describing 5 different levels, the *population of macro-invertebrates* relation has also been split into 5 different binary relations and the *life trait of macro-invertebrates* relation has been split into 6 binary relations. This binarisation process has been accomplished under the guidance of a domain expert.

Applying RCA on this data led to a combinatorial explosion that forced us to try other methods of classification as our implementation could not sustain the number of concepts created reaching the limits of time and size. On this same data RCA-AOC gives a result in a few seconds, and can then be used on much bigger data sets. The implementation of RCA-AOC relies on the same algorithm

as RCA except for the use of the Hermes algorithm [9] for building AOC-posets instead of an FCA algorithm.

From the obtained AOC-posets we extracted two sets of rules: the rules with a physico-chemical character on the left hand side and the rules with animal traits on the left hand side.

We obtained 49179 rules for the first set and 56 for the second one. The difference is coming from the granularity of data which is different for the physico-chemical characters and the biological characters. It comes directly from the choices of the binarization guided by domain experts in order to support the extraction of rules in one direction. Here the goal is to infer biological characters from physico-chemical characters, the former being more expensive to measure than the latter.

The rule presented below is one rule extracted from the experimental data with a support of 67% of the considered sites.

$$\begin{aligned} \mathit{Site} \sqcap \exists \mathit{lvl1}.\mathit{Lig} \sqsubseteq \exists \mathit{lvl2}.(\mathit{Animal} \sqcap \exists \mathit{aff0}.(\mathit{Trait} \sqcap \mathit{FR4}) \\ \sqcap \exists \mathit{aff3}.(\mathit{Trait} \sqcap \mathit{SA1})) \end{aligned}$$

For any *site* of the data set the presence of the component *Lig* at a concentration of level 1 implies the presence of *animals* at a concentration of level 2 that have the *trait* identified by the code *FR4* (modality *diapause or dormancy* of trait *resistance kind*) with an affinity of 0 and the *trait* identified by the code SA1 (modality *fresh water* of the trait *salinity*) with an affinity of 3.

The levels that are referred here are defined specifically to each physico-chemical character, each animal species, and each biological character. For the physico-chemical characters they are usually attached to a level of concentration. The same goes for the animals presence. For the biological characters it refers to an affinity of the species to the modality of a particular biological trait (*e.g.* for its habitat, an animal may have a stronger affinity with *sand* than *gravel*, but may be found in both of them).

Although the modeling can be considered as naive and several approaches could be used to limit the number of concepts, we saw here that RCA has a scale issue as we are only working on a small part of the data. Using AOC-poset allows to handle big data without exploding the number of concepts. Besides, reducing the number of concepts means that the extracted information is also reduced. For the stream site concepts, from which we extract the rules, the AOC-poset keeps the interesting data as we still have the concepts where each attribute is introduced and their order. For the macro-invertebrates however, several concepts are lost that represent combinations of shared life traits. We still have to measure the impact of the missing concepts on the results but it would seem appropriate to combine the RCA-AOC approach with a more traditional lattice building approach that would compute the relevant concepts for the macro-invertebrates while keeping the number of stream site concepts low with AOC-poset.

6 Related Work

Other approaches to integrate relations in FCA have been proposed, including *power context family* [10], and *Logical Concept Analysis* [11]. The originality of RCA is to compute in iterative manner (with a possible stop at each step) several concept lattices from data represented in relational format. The concept lattices are connected by links that abstract the relations between objects. Several operators, borrowed to Description Logics, build the links between concepts.

The original RCA framework, using whole concept lattices, has been used to the analysis and modernization of UML elements [12,13], namely in class diagrams and in use case diagrams. In [14], concept lattices exploit relations between methods and between methods and attributes to detect and fix design defects. Model transformations are learned from transformation examples thanks to several kinds of relations between model elements (*e.g.* between elements inside a model, transformation links between source elements and target elements) [15]. In [16], relations between abstract tasks in an abstract orchestration are used to classify relevant Web services to instantiate the tasks. Other applications can be found in ontology engineering [17,18,19]. In these applications, the datasets are medium-size guarantying the feasibility of the approach. In the FRESQUEAU project, we have larger sets of data where AOC-posets are more suitable than concept lattices. Furthermore, for some issues we want to deal with in the project, like extracting part of the implication rules, AOC-posets are relevant because they contain all the concepts that introduce attributes.

To our best knowledge, the AOC-posets have been introduced by Godin et al [3] in the domain of software engineering (object-oriented programming). The AOC-poset has also been used in applications of FCA to non-monotonic reasoning and domain theory [20] and to produce classifications from linguistic data [21,22]. Specific parts of the AOC-poset (mainly the attribute-concepts part) are used in several works. They include approaches for rebuilding class hierarchy [23] and a recent study for extracting feature tree (in the domain of software product lines) from a set of products [24].

Many approaches exist to extract logical rules from data either in a supervised context for building decision tree [25] or classification rules [26] or in an unsupervised context for association rules learning. Our approach is unsupervised, but AOC-posets only allows the extraction of association rules of confidence 1.

However, the scaling operation permits to consider different kinds of granularity for the relations between the concepts. The relational aspect of the approach permits the extraction of rules from complex relational data which would require to be transformed by propositionalization approaches [27] for many learning approaches. It also could be compared to Inductive Logic Programming [28] in some ways as it will result in first order logic formulas. ILP being a supervised approach, its goal differs as it will try to find the right premise for a given conclusion. The expressivity of the results of our approach are far more restricted than with ILP, even with all the scaling operators that can be imagined, leading to better performances but with a restricted output language.

7 Conclusion

This paper has introduced a RCA process based on AOC-poset in order to deal with computational complexity over big datasets. AOC-posets reduce the number of concepts, with no information lost as the context can still be retrieved from an AOC-poset. In the future we will work on specifying the convergence conditions of AOC-poset based RCA. Indeed, more complex datasets may include cycles between objects. Convergence is ensured with the RCA specification from [7] where the set of concepts used at each step is the set of concepts of the whole lattice. With AOC-poset the convergence is not guaranteed when there are cycles between objects. We also will develop a tool for vizualising the results and assisting their exploration. Finally the approach will be tested on the various datasets of the FRESQUEAU project, and compared with other approaches for relational data mining such as statistical approaches [29] or propositionalization [27].

Acknowledgement. This work was funded by ANR11_MONU14. We warmly acknowledge Corinne Grac (LHYGES) for advices about the dataset, Clémentine Nebut (LIRMM) for the example and Alain Gutierrez (LIRMM) for the implementation of the AOC-poset algorithm (`http://www.lirmm.fr/~gutierre/gsh/`).

References

1. Huchard, M., Rouane-Hacène, M., Roume, C., Valtchev, P.: Relational concept discovery in structured datasets. Ann. Math. Artif. Intell. **49**(1-4) (2007) 39–76
2. Azmeh, Z., Huchard, M., Napoli, A., Rouane-Hacène, M., Valtchev, P.: Querying relational concept lattices. In: CLA'11. (2011) 377–392
3. Godin, R., Mili, H.: Building and Maintaining Analysis-Level Class Hierarchies using Galois Lattices. In: OOPSLA '93. Volume 28. (1993) 394–410
4. Bertaux, A., Le Ber, F., Braud, A., Trémolières, M.: Identifying ecological traits: a concrete FCA-based approach. In: ICFCA 2009. LNAI 5548, Springer-Verlag (2009) 224–236
5. Ganter, B., Wille, R.: Formal Concept Analysis, Mathematical Foundations. Springer-Verlag (1999)
6. Loesch, F., Ploedereder, E.: Restructuring variability in software product lines using concept analysis of product configurations. In: CSMR 2007. (2007) 159–170
7. Hacene, M.R., Huchard, M., Napoli, A., Valtchev, P.: Relational concept analysis: mining concept lattices from multi-relational data. Ann. Math. Artif. Intell. **67**(1) (2013) 81–108
8. Grac, C., Le Ber, F., Braud, A., Trémolières, M., Bertaux, A., Herrmann, A., Manné, S., Lafont, M.: Programme de recherche-développement *Indices* – rapport scienfique final. Contrat pluriannuel 1463 de l'Agence de l'Eau Rhin-Meuse, LHYGES – LSIIT – ONEMA – CEMAGREF (2011)
9. Berry, A., Huchard, M., Napoli, A., Sigayret, A.: Hermes: an efficient algorithm for building Galois Sub-hierarchies. In: CLA 2012. (2012) 21–32
10. Prediger, S., Wille, R.: The Lattice of Concept Graphs of a Relationally Scaled Context. In: ICCS'99, Springer (1999) 401–414

11. Ferré, S., Ridoux, O., Sigonneau, B.: Arbitrary Relations in Formal Concept Analysis and Logical Information Systems. In: ICCS'05. Volume 3596 of LNCS., Springer (2005) 166–180
12. Arévalo, G., Falleri, J.R., Huchard, M., Nebut, C.: Building Abstractions in Class Models: Formal Concept Analysis in a Model-Driven Approach. In: MoDELS 2006. (2006) 513–527
13. Dolques, X., Huchard, M., Nebut, C., Reitz, P.: Fixing Generalization Defects in UML Use Case Diagrams. Fundam. Inform. **115**(4) (2012) 327–356
14. Moha, N., Hacene, A.R., Valtchev, P., Guéhéneuc, Y.G.: Refactorings of Design Defects Using Relational Concept Analysis. In: ICFCA 2008. (2008) 289–304
15. Saada, H., Dolques, X., Huchard, M., Nebut, C., Sahraoui, H.A.: Generation of Operational Transformation Rules from Examples of Model Transformations. In: MoDELS 2012. (2012) 546–561
16. Azmeh, Z., Driss, M., Hamoui, F., Huchard, M., Moha, N., Tibermacine, C.: Selection of Composable Web Services Driven by User Requirements. In: ICWS 2011. (2011) 395–402
17. Bendaoud, R., Napoli, A., Toussaint, Y.: Formal Concept Analysis: A unified framework for building and refining ontologies. In: EKAW 2008. LNCS 5268 (2008) 156–171
18. Rouane-Hacene, M., Valtchev, P., Nkambou, R.: Supporting Ontology Design through Large-Scale FCA-Based Ontology Restructuring. In: ICCS 2011. (2011) 257–269
19. Shi, L., Toussaint, Y., Napoli, A., Blansché, A.: Mining for Reengineering: an Application to Semantic Wikis using Formal and Relational Concept Analysis. In: ESWC'11. Volume 6644 of LNCS., Springer (2011) 421–435
20. Hitzler, P.: Default Reasoning over Domains and Concept Hierarchies. In: Proc. of KI 2004. Volume 3238 of LNCS., Springer Verlag (2004) 351–365
21. Osswald, R., Pedersen, W.: Induction of Classifications from Linguistic Data. In: Proc. of ECAI'02 Workshop. (July 2002)
22. Petersen, W.: A Set-Theoretical Approach for the Induction of Inheritance Hierarchies. In: ENTCS. Volume 51., Elsevier (July 2001)
23. Huchard, M., Dicky, H., Leblanc, H.: Galois Lattice as a Framework to specify Algorithms Building Class Hierarchies. Theoretical Informatics and Applications **34** (2000) 521–548
24. Ryssel, U., Ploennigs, J., Kabitzsch, K.: Extraction of feature models from formal contexts. In: SPLC 2011 Workshops. (2011) 4
25. Quinlan, J.: Induction of decision trees. Machine Learning (1986) 81–106
26. Clark, P., Boswell, R.: Rule induction with cn2: Some recent improvements. In: EWSL. (1991) 151–163
27. Lachiche, N.: Propositionalization. In Sammut, C., Webb, G.I., eds.: Encyclopedia of Machine Learning. Springer (2010) 812–817
28. Muggleton, S., de Raedt, L.: Inductive logic programming: Theory and methods. The Journal of Logic Programming **19**(20) (1994) 629–679
29. Doledec, S., Chessel, D., ter Braak, C., Champely, S.: Matching species traits to environmental variables: a new three-table ordination method. Environmental and Ecological Statistics **3** (1996) 143–166