# Extracting Hierarchies of Closed Partially-Ordered Patterns Using Relational Concept Analysis

Cristina Nica, Agnès Braud, Xavier Dolques, Marianne Huchard, Florence Le Ber

# Extracting Hierarchies of Closed Partially-Ordered Patterns using Relational Concept Analysis

Cristina Nica[1], Agnès Braud[1], Xavier Dolques[1]
Marianne Huchard[2], Florence Le Ber[1]

[1] ICube, University of Strasbourg, CNRS, ENGEES
`cristina.nica@engees.unistra.fr`, `agnes.braud@unistra.fr`,
`xavier.dolques@engees.unistra.fr`, `florence.leber@engees.unistra.fr`
`http://icube-sdc.unistra.fr`
[2] LIRMM, University of Montpellier, CNRS
`huchard@lirmm.fr`
`https://www.lirmm.fr`

**Abstract.** This paper presents a theoretical framework for exploring temporal data, using Relational Concept Analysis (RCA), in order to extract frequent sequential patterns that can be interpreted by domain experts. Our proposal is to transpose sequences within relational contexts, on which RCA can be applied. To help result analysis, we build closed partially-ordered patterns (cpo-patterns), that are synthetic and easy to read for experts. Each cpo-pattern is associated to a concept extent which is a set of temporal objects. Moreover, RCA allows to build hierarchies of cpo-patterns with two generalisation levels, regarding the structure of cpo-patterns and the items. The benefits of our approach are discussed with respect to pattern structures.

## 1 Introduction

Different approaches have been designed to explore datasets containing relational data [10]. Relational Concept Analysis (RCA, [19]) classifies sets of objects described by attributes and relations, allowing the discovery of knowledge patterns and implication rules in relational datasets. RCA has been applied to various data, e.g. for software model analysis and re-engineering [2,9]. The RCA result is a family of interconnected concept lattices, where each lattice can have a huge number of concepts. Consequently, in order to facilitate the analysis step of the RCA output some special procedures for selecting relevant concepts or facilitating the navigation are compulsory.

This paper focuses on exploring qualitative temporal data using RCA, relying on its capability of classifying relational data and its hierarchical results which facilitate the analysis step. Although there exist well-known methods for mining temporal data, our aim is to explore the benefits of RCA, that structures the temporal patterns in a lattice allowing the navigation amongst them. We present

a theoretical framework for this purpose, that is based on a *temporal data model* emphasizing a *main lattice* (i.e. containing the objects of interest) from the RCA results. Any concept of the main lattice corresponds to a set of sequential patterns that is synthesized into a closed partially-ordered pattern (cpo-pattern, [6]) to ease the analysis step. Indeed, such patterns are compact, contain the same information as the sets of sequential patterns they synthesize, and are easy to interpret. On this basis, we proceed as follows (Fig. 1). Firstly, we apply RCA on a *relational context family* containing the temporal data. Secondly, we extract the sequential patterns starting from the main lattice concepts, and build the cpo-patterns. Furthermore, thanks to the hierarchical structure of the RCA output, more or less general patterns can appear, allowing the exploration of the space of cpo-patterns from common to particular trends or vice versa without extra processing being required.
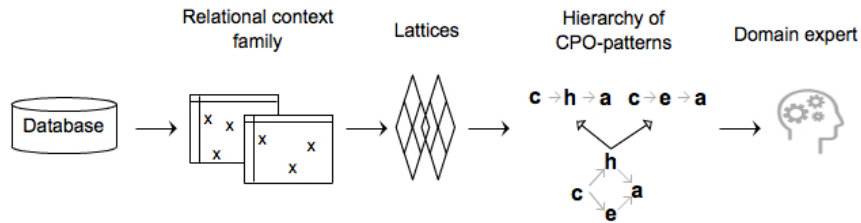


Fig. 1: Analysis process schema

The result obtained can be compared to pattern structures [14], that allow to build lattices on various data and especially graphs. As far as we know, this is still an open question to connect graph pattern structures and the set of graphs that can be built from the set of interconnected lattices of specific instantiations of RCA (like we do). To answer the question in the general case, we must consider the variety of scaling operators. Here we focus on the existential scaling operator.

The paper is structured as follows. Section 2 gives some theoretical background of our work. Section 3 introduces RCA relying on a simple temporal example. Section 4 details the RCA properties allowing to extract and to organise cpo-patterns into a hierarchy, and discusses our approach with respect to pattern structures. In addition, the whole method is described based on the same simple example. Section 5 presents related work. Section 6 concludes and gives a few perspectives of this work.

## 2   Background

Our approach relies both on sequential patterns and formal concept analysis domains.

### 2.1   Sequences, Sequential Patterns and PO-patterns

Let $\mathcal{I} = \{I_1, I_2, ..., I_m\}$ be a set of *items*. An *itemset IS* is a non empty, un-ordered, set of items, $IS = (I_{j_1}...I_{j_k})$ where $I_{j_i} \in \mathcal{I}$. Let $\mathcal{IS}$ be the set of all itemsets built from $\mathcal{I}$. A *sequence S* is a non empty ordered list of itemsets, $S = \langle IS_1 IS_2...IS_p \rangle$ where $IS_j \in \mathcal{IS}$. The sequence $S$ is a *subsequence* of another sequence $S' = \langle IS'_1 IS'_2...IS'_q \rangle$, denoted as $S \preceq_s S'$, if $p \leq q$ and if there are integers $j_1 < j_2 < ... < j_k < ... < j_p$ such that $IS_1 \subseteq IS'_{j_1}, IS_2 \subseteq IS'_{j_2}, ..., IS_p \subseteq IS'_{j_p}$.

Suppose now that there is a partial order on the items, $(\mathcal{I}, \leq)$. Then the order on itemsets is defined as follows: $IS \subseteq_H IS'$ if $\forall I_j \in IS, \exists I_{j'} \in IS', I_{j'} \leq I_j$ and $j \neq k \rightarrow j' \neq k'$. The order on sequences is defined accordingly.

Sequential patterns have been defined by [1] as frequent subsequences found in a sequence database. A sequential pattern is associated to a support, i.e. the number of sequences containing the pattern. Formally, the support of a sequential pattern $M$ extracted from a sequence database $\mathcal{D_S}$ is defined as:

$$Support(M) = |\{S \in \mathcal{D_S} | M \preceq_s S\}| \tag{1}$$

Partially ordered patterns, *po-patterns*, have been introduced by [6], to synthesise sets of sequential patterns. Formally, a *po-pattern* is a directed acyclic graph $G = (\mathcal{V}, \mathcal{E}, l)$. $\mathcal{V}$ is the set of vertices, $\mathcal{E}$ is a set of directed edges such that $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$, and $l$ is a labelling function mapping each vertex to an itemset. With such a structure, we can determine a strict partial order on vertices $u$ and $v$ such that $u \neq v$, i.e. $u < v$ if there is a directed path from $u$ to $v$. However, if there is no directed path from $u$ to $v$, these elements are not comparable. Each path of the graph is a sequential pattern as defined before. The set of paths in $G$ is denoted by $\mathcal{P}_G$. A po-pattern is associated to the set of sequences $\mathcal{S}_G$ that contain all paths of $\mathcal{P}_G$. Following Eq. 1, the support of a po-pattern is defined as:

$$Support(G) = |\mathcal{S}_G| = |\{S \in \mathcal{D_S} | \forall M \in \mathcal{P}_G, M \preceq_s S\}| \tag{2}$$

Furthermore, let $G$ and $G'$ be two po-patterns with $\mathcal{P}_G$ and $\mathcal{P}_{G'}$ their sets of paths. $G'$ is a sub po-pattern of $G$, denoted by $G' \preceq_g G$, if $\forall M' \in \mathcal{P}_{G'}, \exists M \in \mathcal{P}_G$ such that $M' \preceq_s M$. A po-pattern $G$ is closed, denoted cpo-pattern, if there exists no po-pattern $G'$ such that $G \prec_g G'$ with $\mathcal{S}_G = \mathcal{S}_{G'}$.

### 2.2   FCA and Pattern Structures

Formal Concept Analysis (FCA, [15]) considers a formal context which is a set of objects described by attributes, and builds from it a concept lattice used to analyse the objects. Concisely, a formal context $K$ is a 3-tuple $(G, M, I)$, where $G$ is a set of objects, $M$ a set of attributes, and $I$ the incidence relation, $I \subseteq G \times M$. $C = (X, Y)$ where $X = \{g \in G | \forall m \in Y, (g, m) \in I\}$ and $Y = \{m \in M | \forall g \in X, (g, m) \in I\}$ is a formal concept built from $K$. $X$ and $Y$ are respectively the extent and the intent of the concept. Let $\mathcal{C}_K$ be the set of all formal concepts that can be built on $K$. Let $C_1 = (X_1, Y_1)$ and $C_2 = (X_2, Y_2)$ be two concepts from $\mathcal{C}_K$, the concept generalisation order $\preceq_{\mathcal{C}_K}$ is here defined by $C_1 \preceq_{\mathcal{C}_K} C_2$ if

and only if $X_1 \subseteq X_2$ (which is equivalent to $Y_2 \subseteq Y_1$). $\mathcal{L}_K = (\mathcal{C}_K, \preceq_{\mathcal{C}_K})$ is the concept lattice built from $K$. We denote by $\top(\mathcal{L}_K)$ the concept from $\mathcal{L}_K$ whose extent has all the objects, and by $\bot(\mathcal{L}_K)$ the concept from $\mathcal{L}_K$ whose intent has all the attributes.

FCA is designed to deal with binary contexts, whereas attributes can be of various forms, e.g. intervals, multi-valued, etc. To generalise the FCA approach, [14] proposed to use pattern structures. A pattern structure $(G, (D, \sqcap), \delta)$ gives a description of a set of objects $G$ by a set of descriptions (patterns) in $D$, which is provided with a similarity operation $\sqcap$, such that $(D, \sqcap)$ is a meet-semilattice. $\delta : G \to D$ maps objects to their description and should verify that $\{\delta(g) | g \in G\}$ is a complete subsemilattice of $(D, \sqcap)$. Patterns can be of different types, such as vectors of intervals [16], sequences [5] or labelled graphs [14].

## 3    Relational Analysis of Temporal Data

We propose a general temporal modelling of sequential datasets that allows the assessment of relationships between *qualitative temporal objects*. Here, we use a toy example from the medical domain, which is illustrated in Fig. 2, to explain our general approach.

In this example, we study the symptoms (S), e.g. fever, cough and fatigue, that indicate the presence of viruses (V), e.g. influenza and hepatitis, in patients. The symptoms and viruses are detected by medical examinations (ME) and viral tests (VT), respectively. These physical examinations are identified by temporal objects *(Object, Date)* where: *Object* represents the patient, and *Date* designates the time when the physical examination was done. A patient can do several medical examinations and viral tests. Symptoms, viruses, medical examinations and viral tests are sets of objects. Viral tests are linked to viruses by some qualitative binary relations *has virus* differentiated by the type of diagnosed virus, e.g. *A*, *B* or *C*. Similarly, medical examinations are linked to symptoms by qualitative relations *has symptom* (mS or hS) differentiated by the type of identified symptoms, e.g. *moderate* or *high*. Viral tests/medical examinations and medical examinations are linked by a temporal binary relation *is preceded by* (ipb) that associates a viral test/medical examination to a medical examination if the viral test/medical examination is preceded in time by the medical examination. There is no temporal binary relation between viral tests since our aim is to study the symptoms that prognosticate distinct types of viruses. In the following, based on the relational character of the toy example, we apply RCA in order to mine these qualitative temporal data.

RCA extends the purpose of FCA to relational data. RCA applies iteratively FCA on a Relational Context Family (RCF). An RCF comprises a set $\mathcal{K}$ of object-attribute contexts and a set $\mathcal{R}$ of object-object contexts. $\mathcal{K}$ contains $n$ object-attribute formal contexts $K_i = (G_i, M_i, I_i), i \in \{1, ..., n\}$. $\mathcal{R}$ contains $m$ object-object relational contexts $R_j = (G_k, G_l, r_j), j \in \{1, ..., m\}$, where $G_k$ that we call the domain of the relation and $G_l$ that we call the range of the relation are respectively the sets of objects of $K_k$ and $K_l$, and $r_j \subseteq G_k \times G_l$,

with $k, l \in \{1, ..., n\}$. New attributes, called *relational attributes*, extend the formal contexts by using object-object relations and already created concepts. A relational attribute takes the syntactic form $qr_j(C)$, where $q$ is a quantifier, $r_j$ is a relation and $C$ is a concept whose extent contains objects from the range of $r_j$. This paper uses the *existential* quantifier which, for a relational context $R_j = (G_k, G_l, r_j)$, creates a relation $\exists r_j$ between an object $o \in G_k$ and a concept $C = (X, Y)$ of the lattice $\mathcal{L}_{K_l}$ if $r_j(o) \cap X \neq \emptyset$. RCA process consists in applying FCA first on each object-attribute context of an RCF, and then iteratively on each object-attribute context extended by the relational attributes created using the concepts from the previous step. The RCA result is obtained when the families of lattices of two consecutive steps are isomorphic and the contexts are unchanged.

Henceforth, we try to answer the following question by means of RCA: *Can outbreaks of Influenza A virus be recognised assessing the symptoms, e.g. FEVER and COUGH, felt by patients?*

Firstly, the RCA input (RCF) is built by following the temporal data model illustrated in Fig. 2. Table 1 depicts an example of RCF on medical sequential data collected during the last year. The three tables from the left hand side represent object-attribute contexts: KS (symptoms), KVT (viral tests) and KME (medical examinations). For example, KME has no column, i.e. a medical examination is described using qualitative binary relations, and the rows represent medical examinations identified by pairs such as P1_10/01, that is a medical examination done by patient P1 on 10th of January. There is no object-attribute context of viruses due to the set of viruses that contains only *Influenza A* virus. The four tables from the right hand side represent object-object contexts: RVT-ipb-ME (viral test *ipb* medical examination), RME-ipb-ME (medical examination *ipb* medical examination), RmS (medical examination detects a moderate symptom) and RhS (medical examination detects a high symptom). For instance, RVT-ipb-ME has viral tests as rows and medical examinations as columns. A cross indicates a link between objects, e.g. the cell identified by the viral test P1_20/01 and the medical examination P1_17/01 contains a cross since both are undergone by patient P1 and 20/01 is preceded in time by 17/01.
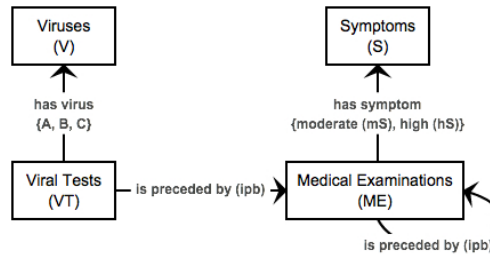


Fig. 2: The modelling of the medical sequential dataset

Table 1: RCF composed of object-attribute contexts: KS, KVT and KME; temporal object-object contexts: RVT-ipb-ME and RME-ipb-ME; qualitative object-object contexts: RmS and RhS.

**Object-attribute contexts**

KS

| KS | FEVER | COUGH |
|---|---|---|
| FEVER | × | |
| COUGH | | × |

KVT

| KVT | FEVER | COUGH |
|---|---|---|
| P1_20/01 | | |
| P1_28/12 | | |
| P2_30/02 | | |
| P2_15/07 | | |

KME

| KME |
|---|
| P1_17/01 |
| P1_10/01 |
| P1_25/12 |
| P2_28/02 |
| P2_20/02 |
| P2_11/07 |

**Object-object contexts**

RVT-ipb-ME

| RVT-ipb-ME | P1_17/01 | P1_10/01 | P1_25/12 | P2_28/02 | P2_20/02 | P2_11/07 |
|---|---|---|---|---|---|---|
| P1_20/01 | × | × | | | | |
| P1_28/12 | | | × | | | |
| P2_30/02 | | | | × | × | |
| P2_15/07 | | | | | | × |

RME-ipb-ME

| RME-ipb-ME | P1_17/01 | P1_10/01 | P1_25/12 | P2_28/02 | P2_20/02 | P2_11/07 |
|---|---|---|---|---|---|---|
| P1_17/01 | | × | | | | |
| P1_10/01 | | | | | | |
| P1_25/12 | | | | | | |
| P2_28/02 | | | | | × | |
| P2_20/02 | | | | | | |
| P2_11/07 | | | | | | |

RmS

| RmS | FEVER | COUGH |
|---|---|---|
| P1_17/01 | × | × |
| P1_10/01 | | |
| P1_25/12 | | × |
| P2_28/02 | × | × |
| P2_20/02 | | |
| P2_11/07 | × | |

RhS

| RhS | FEVER | COUGH |
|---|---|---|
| P1_17/01 | | |
| P1_10/01 | × | |
| P1_25/12 | × | |
| P2_28/02 | | |
| P2_20/02 | × | |
| P2_11/07 | | × |

Secondly, RCA is applied on the RCF shown in Table 1 and the result is given in Fig. 3. There is a lattice for each object-attribute context: $\mathcal{L}_{\text{KME}}$ (medical examinations), $\mathcal{L}_{\text{KS}}$ (symptoms) and $\mathcal{L}_{\text{KVT}}$ (viral tests). $\mathcal{L}_{\text{KME}}$ and $\mathcal{L}_{\text{KVT}}$ are modified during the iterative steps due to the temporal and qualitative object-object contexts that have the domain KME and KVT, respectively. Each concept is represented by a box structured from top to bottom as follows: concept name, simplified intent, simplified extent. The representation of each lattice is simplified as every attribute/object is top-down/bottom-up inherited. Thus, an attribute/object is shown only in the highest/lowest concept where it appears. For example, the intent of concept CKME_4 from $\mathcal{L}_{\text{KME}}$ contains the relational attributes ∃RmS(CKS_3) and ∃RhS(CKS_3) inherited from concepts CKME_10 and CKME_9, respectively; the extent contains the objects P1_25/12 and P2_11/07 inherited respectively from concepts CKME_3 and CKME_2. The arrows represent the generalisation order. The navigation amongst these lattices follows the concepts used to build relational attributes, e.g. the aforementioned ∃RhS(CKS_3) allows us to navigate from the concept CKME_4 from $\mathcal{L}_{\text{KME}}$ to concept CKS_3 out of $\mathcal{L}_{\text{KS}}$.

## 4     Extracting CPO-patterns from RCA Result

We focus here on lattices built on temporal objects. Furthermore, let us recall that we consider a main lattice, which contains the objects of interest (in the toy example, the viral tests allowing the diagnosis on a patient). In addition, concepts in these lattices contain two types of relational attributes: 1) qualitative relational attributes and 2) temporal relational attributes. The range of a qualitative attribute is a concept which represents *parameters* (e.g. virus, symptom). The range of a temporal attribute is a concept that represents *temporal observations* (e.g. medical examinations).

### 4.1     Proposition

In this section we give some useful properties of RCA results that help the extraction process of cpo-patterns. A temporal object in the extent of a concept
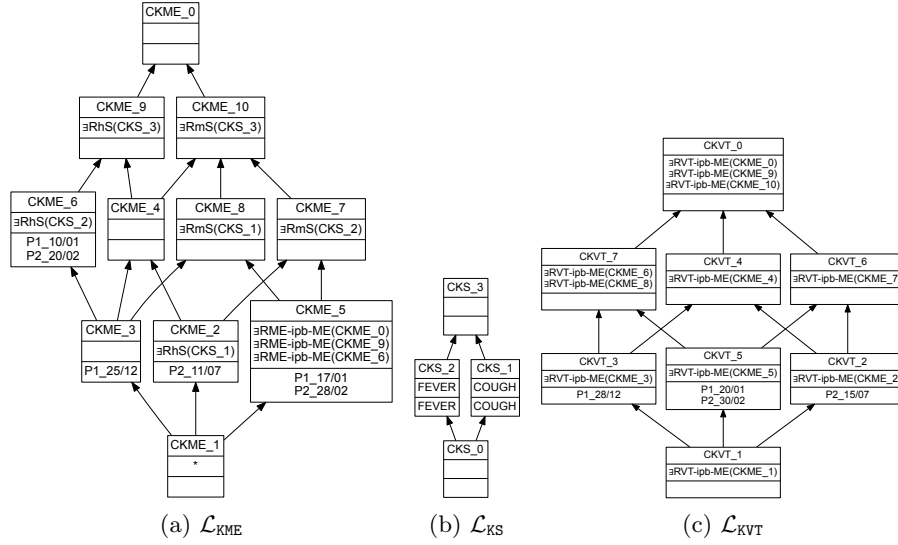
Fig. 3: The concept lattice family obtained by applying RCA on the RCF given in Table 1. The $*$ symbol represents all the relational attributes of KME

from the main lattice is associated to a temporal sequence (e.g. the sequence of medical examinations before the final viral test). A concept extent represents thus a set of sequences. Navigating the relational attributes allows to reveal sequential patterns contained in all these sequences, as explained below.

*Property 1.* Each temporal relational attribute in a concept intent of the main lattice allows to extract at least one sequential pattern. On the contrary, if there is no temporal relational attribute in a concept intent, this concept represents no sequential pattern.

Indeed, let $C$ be a concept of the main lattice and $\exists t(C_1)$ a temporal relational attribute of its intent. If $C_1$ intent contains a qualitative attribute $\exists q(C_2)$ then $C_1$ reveals an itemset of *qualitative values* (e.g. *Fever*$_{high}$); if $C_1$ intent contains a temporal attribute $\exists t(C_2)$ then $C_1$ leads to another itemset in the sequential pattern, depending on $C_2$ intent. The temporal relational attributes reveal the position of the itemset (in the sequential pattern). If $C_1$ intent contains no temporal attribute, the sequential pattern is finished.

A naive approach to extract sequential patterns out of a concept intent takes into account all its qualitative and temporal relational attributes. Nevertheless, some properties of the RCA results can be used to improve the extraction process. In the following, properties 2 and 3 are introduced to reduce redundancy by considering only the relational attributes pointing *the most specific concepts*, and to prune temporal relational attributes that can be deduced by transitivity.

*Property 2.* Let $C_1 = (X_1, Y_1) \in \mathcal{L}_K$ and $C_2 = (X_2, Y_2) \in \mathcal{L}_K$ be two concepts such that $C_1 \preceq_{\mathcal{C}_K} C_2$. Let $C = (X, Y)$ be a concept which intent has two relational attributes $\exists r(C_1)$ and $\exists r(C_2)$ (derived from the same relation $r$). Then $\exists r(C_1) \rightarrow \exists r(C_2)$.

*Proof.* $\exists r(C_1) \in intent(C) \leftrightarrow \forall o \in X, r(o) \cap X_1 \neq \emptyset$. Since $C_1 \preceq_{\mathcal{C}_K} C_2, X_1 \subseteq X_2$, and thus $r(o) \cap X_2 \neq \emptyset \leftrightarrow \exists r(C_2) \in intent(C)$.

Hence, the relational attributes are ordered and $\exists r(C_2)$ is redundant in the interpretation of $C$.

*Property 3.* Let $t$ be the temporal relation. Let $C$, $C_1$ and $C_2$ be three concepts such that $\{\exists t(C_1), \exists t(C_2)\} \subseteq intent(C)$, and $\exists t(C_2) \in intent(C_1)$. Then $\exists t(C_2) \in intent(C)$ can be deduced from $\exists t(C_1) \in intent(C)$.

*Proof.* The property 3 is directly obtained from the transitivity of the $t$ relation.

In order to facilitate the analysis step the sets of sequential patterns extracted from the RCA result are then converted into cpo-patterns. We use therefore the pruning and merging steps proposed by [12].

*Property 4.* Let $C$ be a concept of the main lattice whose intent contains at least one temporal attribute. Then $C$ can be associated to a cpo-pattern that summarises the set of sequential patterns deriving from $C$. Conversely, the cpo-pattern is associated to the extent of $C$ and its support is $|extent(C)|$.

*Proof.* A set of sequential patterns can be transformed into po-patterns [6]. A po-pattern associated to a concept is closed since the corresponding set of sequences is maximal or equivalently the concept extent is maximal.

*Property 5.* The set of cpo-patterns associated to the main lattice is ordered according to the inclusion on extents. This order corresponds to the subsumption on graphs $\preceq_g$ (cf. Section 2).

*Proof.* Let $G$ and $G'$ be two cpo-patterns with $\mathcal{P}_G$ and $\mathcal{P}_{G'}$ their sets of paths. Suppose $G$ (resp. $G'$) is associated to a concept $C = (X, Y)$ (resp. $C' = (X', Y')$) and $X \subseteq X'$. Then $Y' \subseteq Y \leftrightarrow \forall a \in Y', a \in Y$. Then $\forall M' \in \mathcal{P}_{G'}, \exists M \in \mathcal{P}_G, M' \preceq_s M \rightarrow G' \preceq_g G$.

## 4.2   Characterising CPO-patterns

To characterise the items in the extracted sequential patterns, the qualitative relational attributes are analysed. To this end, we define two types of relational attributes, depending on the generality or specificity of the concept they point to.

**Definition 1 (Vague/Defined Relational Attribute).** *The relational attribute $\exists r(C_1)$, where $C_1$ is a concept of lattice $\mathcal{L}_K$, is called Vague if $C_1 \equiv \top(\mathcal{L}_K)$, respectively it is Defined if $C_1 \prec_{\mathcal{C}_K} \top(\mathcal{L}_K)$.*

Relying on the taxonomy of items revealed by RCA, we define three types of items that disclose abstract and concrete information from the analysed data as follows:

- let $C$ be a concept whose intent has no qualitative relational attribute, e.g. CKME_0, see Fig. 3. Then the extracted item is an **abstract** item, denoted by "$?_?$". The abstract item describes a collection of objects which point out the occurrence of dissimilar parameters having dissimilar qualities;
- let $C$ be a concept whose intent contains a vague qualitative relational attribute $\exists quality(\top)$ (e.g. CKME_10). The extracted item is an **abstract qualitative** item, denoted by "$?_{quality}$". The abstract qualitative item describes a collection of objects which point out the occurrence of dissimilar parameters having the same quality;
- let $C$ be a concept whose intent contains a defined qualitative relational attribute $\exists quality(C_1)$, with $extent(C_1) = \{value\}$ (e.g. CKME_7). The extracted item is a **concrete qualitative** item, denoted by "$value_{quality}$". The concrete qualitative item describes a collection of objects which point out the occurrence of the same concrete parameter having the same quality.

These types of items allow us to define the partial order $(\mathcal{I}, \leq)$ on the extracted items. For every *value* from the parameter set and every *quality* from the qualitative relations the relation $\leq$ is defined as follows:

- $value_{quality} \leq ?_{quality}$
- $?_{quality} \leq ?_?$
- $value_{quality} \leq ?_?$
- $value_{quality} \leq value_{quality}$.

The various extracted items allow us to introduce three new types of cpo-patterns.

**Definition 2 (Abstract/Hybrid/Concrete cpo-pattern).** *A cpo-pattern is as follows:*

- *Abstract if it contains only abstract and/or abstract qualitative items;*
- *Hybrid if it contains abstract and/or abstract qualitative items and concrete qualitative items;*
- *Concrete if it contains only concrete qualitative items.*

Hybrid patterns can be characterised using a measure of precision referred to as *accuracy*.

**Definition 3 (Accuracy($v$)).** *Let $G$ be a cpo-pattern and $\mathcal{I}_G$ the multiset of items labelling the nodes of $G$ ($\forall I \in \mathcal{I}_G$, $I \in \mathcal{I}$). Let $\mathcal{I}_G^c$ be the subset of $\mathcal{I}_G$ containing the concrete qualitative items. The accuracy of $G$ is defined as the ratio of the number of items in $\mathcal{I}_G^c$ to the total number of items in $\mathcal{I}_G$.*

$$v(G) = \frac{|\mathcal{I}_G^c|}{|\mathcal{I}_G|} \in [0,1] \tag{3}$$

*If $G$ is abstract, $v(G) = 0$; if $G$ is concrete $v(G) = 1$.*

### 4.3   RCA based CPO-patterns vs. Pattern Structures

In a previous work [12], cpo-patterns were directly extracted from sequences, using a sequential pattern mining algorithm. In these cpo-patterns, there is no order on the set of items $\mathcal{I}$, i.e. there are only concrete items. The resulting set of cpo-patterns, $D$, with the intersection operation on graphs $\sqcap$ allows to build a pattern structure $(G, (D, \sqcap), \delta)$ where $G$ is the set of objects described by the cpo-patterns, through the $\delta$ relation.

The resulting lattice can be compared to the hierarchy of cpo-patterns built in our RCA-based approach. Firstly, in our approach, the hierarchy is built directly from the RCA result. Secondly our approach produces a hierarchy of items and thus more general cpo-patterns than in [12]. Finally, the RCA-approach allows both to navigate along the sequences and to synthesize them within cpo-patterns.

### 4.4   Implementation with the Toy Example

To illustrate our method, let us examine the concept CKVT_6 of $\mathcal{L}_{KVT}$ (Fig. 3(c)). Its intent contains four temporal relational attributes, the most specific being $\exists$RVT-ipb-ME(CKME_7) and $\exists$RVT-ipb-ME(CKME_9). On the contrary, the intents of concepts CKME_7 and CKME_9 (Fig. 3(a)) contain no temporal relational attribute. Accordingly, the chains of concepts starting from CKVT_6 are $\langle$(CKVT_6)(CKME_7)$\rangle$ and $\langle$(CKVT_6)(CKME_9)$\rangle$ as shown in Fig. 4. From these chains two sequential patterns can be extracted, denoted respectively $S1_{\text{CKVT}\_6}$ and $S2_{\text{CKVT}\_6}$.

Relying on the aforementioned types of items, the itemsets of $S1_{\text{CKVT}\_6}$ and $S2_{\text{CKVT}\_6}$ can be defined. Both extracted chains of concepts begin with CKVT_6 that by default represents the concrete qualitative item "$Influenza_A$". The intent of concept CKME_7 contains both types of qualitative relational attributes: the vague $\exists$RmS(CKS_3) and the defined $\exists$RmS(CKS_2). CKS_2 $\prec_{\mathcal{C}_{KS}}$ CKS_3 (Fig. 3(b)), therefore the interpretation of CKME_7 is based on $\exists$RmS(CKS_2) that represents the concrete qualitative item "$FEVER_{moderate}$". Accordingly, the extracted sequential pattern is $S1_{\text{CKVT}\_6} = \langle$(Influenza$_A$)(FEVER$_{moderate}$)$\rangle$. Following the same principle, the intent of CKME_9 has only the vague qualitative relational attribute $\exists$RhS(CKS_3) that represents the abstract qualitative item "$?_{high}$". Consequently,
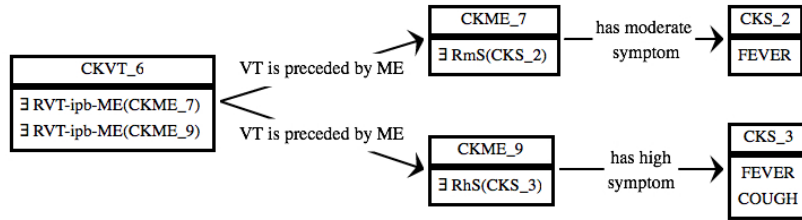


Fig. 4: The set of concept intents/extents used to interpret the concept CKVT_6

the extracted sequential pattern is $S2_{\text{CKVT\_6}} = \langle(\text{Influenza}_\text{A})(?_{\text{high}})\rangle$. The full set of concept intents and extents navigated by our approach is shown in Fig. 4. The cpo-pattern summarising the two sequential patterns is depicted in Fig. 5 (b).
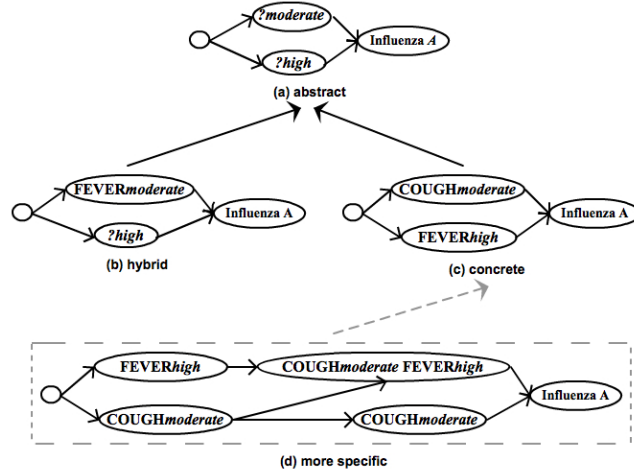


Fig. 5: An excerpt from the hierarchy of cpo-patterns obtained for the RCF out of Table 1. The (d) cpo-pattern is extracted by adding new medical examinations to the RCF

Following our method, i.e. extracting one cpo-pattern for each concept of the main lattice (Fig. 3(c)), a hierarchy of cpo-patterns is obtained. In Fig. 5, the abstract, hybrid and concrete cpo-patterns represent respectively the intent of concepts CKVT_0, CKVT_6 and CKVT_7. The abstract cpo-pattern (Fig. 5(a)) subsumes a group of cpo-patterns that share the less accurate common trend: *often before Influenza A virus patients felt in any order a high symptom and another moderate symptom*. The hybrid cpo-pattern (Fig. 5(b)) subsumes a subgroup of the aforementioned group, i.e. it is a specialisation of the abstract pattern (Fig. 5(a)). This subgroup encapsulates individual cpo-patterns that share the more or less accurate common trend: *less often before Influenza A virus patients felt in any order moderate FEVER and another high symptom*. The concrete cpo-pattern (Fig. 5(c)) is another specialisation of the abstract pattern (Fig. 5(a)). This pattern subsumes a subgroup of individual cpo-patterns that share the accurate common trend: *less often before Influenza A virus patients felt in any order moderate COUGH and high FEVER*. The concrete cpo-pattern depicted in Fig. 5(d) is extracted if new medical examinations are added to the RCF from Table 1. This pattern highlights the two generalization levels of the extracted hierarchy. Firstly, the structure of this pattern is more specific than the structure of the ancestors of the pattern, i.e. this pattern contains more vertices and more edges. Secondly, the generalisation of items is illustrated. For instance, the

pattern reveals the rule $\{COUGH_{moderate}, FEVER_{high}\} \Rightarrow \{Influenza_A\}$ that is a specialisation of the rule $\{FEVER_{high}\} \Rightarrow \{Influenza_A\}$ revealed by Fig. 5(c).

## 5   Related Work

To our knowledge, this is the first time that RCA is used to explore sequential datasets. There are, however, various related FCA approaches, e.g. [22] introduced Temporal Concept Analysis where objects are characterised with a date and a state (i.e. a set of attributes). Data are merged into a single context, and the resulting concept lattice is analysed thanks to the date element in the concepts, so that temporal relations between concepts are actually revealed by the analyst. This approach has been used to analyse sequential data about crime suspects [18]. In our RCA approach, the temporal relation between dates is considered as an object-object relation and it links concepts from several lattices. In [13], sequential datasets are processed without involving any partial order. In [6], closed subsequences are mined and then grouped in a lattice similar to an FCA lattice. In [4], sequential data are mapped onto pattern structures whose projections are used to build a pattern concept lattice. The authors combine the stability of concepts and the projections of pattern structures in order to select relevant patterns.

Besides, there exist various methods to explore qualitative sequential data. Indeed, sequential pattern mining is an active research area, in relation to the exponential growth of temporal and spatio-temporal databases. Sequential patterns have been introduced by [1] and used for different purposes, e.g. classification [7] or prediction [21]. Such an approach has been developed within the Fresqueau project and focused on cpo-patterns, which were selected through various measures [11]. Sequential pattern mining approaches are more efficient from a scalability point of view, but RCA enables to deepen the result analysis by navigating within the lattice family. Moreover it reveals a taxonomy from the data, that can be used to organise the cpo-patterns. Such results can be related to [20], where generalized sequential patterns are extracted based on a user-given taxonomy. RCA allows to discover this taxonomy.

Some authors proposed to combine RCA and pattern structures. In [8] RCA is adapted to integrate a description of $G_1$, a set of source objects with descriptors (coming from a pattern structure $(G_1, (D, \sqcap), \delta)$) and relational attributes to a set of concepts on a target formal context $(G_2, M_2, I_2)$. The relational attributes are built for a relation $r \subseteq G_1 \times G_2$ and the usual scaling operators, like $\exists$ or $\forall\exists$. This is formalized as a "heterogeneous pattern structure". An application to Information Research domain is described, where source objects are documents, descriptors are vectors of intervals of LV values, target objects are terms grouped into concepts when they have same meaning (represented by a synset), and the relation $r$ connects documents to their included terms. LV are Latent Variables that abstract hidden topics spread over the documents. Finally, there is no extraction of complex graph patterns as in our case, since there is only one relation and descriptors in the initial pattern structure are interval vectors.

## 6    Conclusion

In this paper, we have presented a theoretical framework for exploring temporal data using RCA. Our work proposes a comprehensive process for exploring sequential datasets which spans: 1) the relational analysis step that relies on a temporal data model which allows to emphasize the objects of interest in the study and 2) the extraction step of cpo-patterns from the RCA result.

The result is a hierarchy of cpo-patterns associated to sets of temporal objects, that can be compared to a lattice of pattern structures. With respect to pattern structures, the proposed approach, thanks to RCA, allows both to navigate along the sequences and to build a set of cpo-patterns including various levels of generalisation.

Our method was applied to sequential datasets, dealing with biological and physico-chemical parameters sampled in waterbodies [3]. Data were collected from french databases during the ANR 11 MONU 14 Fresqueau project. Results showing the effectiveness of our approach on large datasets are presented in [17] and are available at a website[3].

In the future, we will test our approach on other relational data such as spatial data, where various relations can be considered together (topology, distance, orientation).

## References

1. Agrawal, R., Srikant, R.: Mining sequential patterns. In: Int. Conference on Data Engineering. pp. 3–14 (1995)
2. 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. pp. 513–527 (2006)
3. Berrahou, L., Lalande, N., Serrano, E., Molla, G., Berti-Équille, L., Bimonte, S., Bringay, S., Cernesson, F., Grac, C., Ienco, D., Le Ber, F., Teisseire, M.: A quality-aware spatial data warehouse for querying hydroecological data. Computers & Geosciences 85, Part A, 126–135 (2015)
4. Buzmakov, A., Egho, E., Jay, N., Kuznetsov, S.O., Napoli, A., Raïssi, C.: On mining complex sequential data by means of FCA and pattern structures. International Journal of General Systems 45 (2016)
5. Buzmakov, A., Egho, E., Jay, N., Kuznetsov, S.O., Napoli, A., Raïssi, C.: FCA and pattern structures for mining care trajectories. In: Proceedings of the Int. Workshop FCA4AI at IJCAI 2013. CEUR Workshop Proceedings, vol. 1058, pp. 7–14. CEUR-WS.org (2013)
6. Casas-Garriga, G.: Summarizing sequential data with closed partial orders. In: 2005 SIAM Int. Conference on Data Mining. pp. 380–391 (2005)
7. Cheng, H., Yan, X., Han, J., Hsu, C.: Discriminative frequent pattern analysis for effective classification. In: Int. Conference on Data Engineering. pp. 716–725 (2007)
8. Codocedo-Henriquez, V.: Contributions to indexing and retrieval using Formal Concept Analysis. Doctoral thesis, Université de Lorraine (Sep 2015)

---

[3] `http://icube-sdc.unistra.fr/en/img_auth.php/c/c4/Mining_Hydroecological_Data_using_RCA.pdf`

9. Dolques, X., Huchard, M., Nebut, C., Reitz, P.: Fixing Generalization Defects in UML Use Case Diagrams. Fundam. Inform. 115(4), 327–356 (2012)
10. Džeroski, S.: Relational data mining. In: Maimon, O., Rokach, L. (eds.) Data Mining and Knowledge Discovery Handbook, pp. 869–898. Springer (2005)
11. Fabrègue, M., Braud, A., Bringay, S., Grac, C., Le Ber, F., Levet, D., Teisseire, M.: Discriminant temporal patterns for linking physico-chemistry and biology in hydro-ecosystem assessment. Ecological Informatics 24, 210–221 (2014)
12. Fabrègue, M., Braud, A., Bringay, S., Le Ber, F., Teisseire, M.: Mining closed partially ordered patterns, a new optimized algorithm. Knowledge-Based Systems 79, 68–79 (Jan 2015)
13. Ferré, S.: The efficient computation of complete and concise substring scales with suffix trees. In: Formal Concept Analysis, 5th Int. Conference ICFCA 2007, LNAI, vol. 4390, pp. 98–113. Springer (2007)
14. Ganter, B., Kuznetsov, S.O.: Pattern structures and their projections. In: Conceptual Structures: Broadening the Base, Proceedings of the 9th Int. Conference on Conceptual Structures, ICCS-01. LNCS, vol. 2120, pp. 129–142. Springer (2001)
15. Ganter, B., Wille, R.: Formal Concept Analysis: Mathematical Foundations. Springer (1999)
16. Kaytoue-Uberall, M., Assaghir, Z., Messai, N., Napoli, A.: Two complementary classification methods for designing a concept lattice from interval data. In: Foundations of Information and Knowledge Systems, 6th Int. Symposium, FoIKS 2010. LNCS, vol. 5956, pp. 345–362. Springer (2010)
17. Nica, C., Braud, A., Dolques, X., Huchard, M., Le Ber, F.: L'analyse relationnelle de concepts pour la fouille de données temporelles – Application à l'étude de données hydroécologiques. Revue des Nouvelles Technologies de l'Information Extraction et Gestion des Connaissances, EGC 2016, RNTI-E-30, 267–278 (2016)
18. Poelmans, J., Elzinga, P., Viaene, S., Dedene, G.: A Method based on Temporal Concept Analysis for Detecting and Profiling Human Trafficking Suspects. In: Artificial Intelligence and Applications, AIA 2010. pp. 1–9 (2010)
19. Rouane-Hacene, M., Huchard, M., Napoli, A., Valtchev, P.: Relational concept analysis: Mining concept lattices from multi-relational data. Annals of Mathematics and Artificial Intelligence 67(1), 81–108 (2013)
20. Srikant, R., Agrawal, R.: Mining Sequential Patterns: Generalizations and Performance Improvements. In: Advances in Database Technology, Proceedings of the 5th Int. Conference on Extending Database Technology. pp. 3–17. EDBT'96, Springer-Verlag (1996)
21. Wang, M., Shang, X., Li, Z.: Sequential pattern mining for protein function prediction. In: Advanced Data Mining and Applications, pp. 652–658. ADMA (2008)
22. Wolff, K.E.: Temporal Concept Analysis. In: ICCS-01 Workshop on Concept Lattice for KDD, 9th Int. Conference on Conceptual Structures. pp. 91–107 (2001)