Mining Closed Gradual Patterns

Sarra Ayouni[1], Anne Laurent[2] Sadok Ben Yahia[1], and Pascal Poncelet[2]

[1] Faculty of Sciences of Tunis, 1060, Campus Universitaire, Tunis, Tunisie,
ayouni@lirmm.fr, sadok.benyahia@fst.rnu.tn
[2] LIRMM − CNRS, 161 rue Ada, Montpellier, France
laurent@lirmm.fr, poncelet@lirmm.fr

1 Introduction

With the steady development of the computing tools, we attended last three decades a considerable increase of the quantity of data stored in databases. So, extracting knowledge from this data is of paramount importance. Data mining is becoming an inescapable tool to reach this goal. Association rule extraction is one of the important tasks in data mining. This powerful technique has a wide range of applications in many areas of business practice and also research. A scrutiny of the related work shows that another type of rules called *gradual rules* also paid attention within the data mining community. Gradual rules of the form − *"the more A, the more B"*− mainly grasped the interest within recommendation and command system fields [2]. Several approaches and semantics dealing with this kind of rules have been proposed in literature. However, the relevance and the usefulness of the mined knowledge seems no to be the main concern in these approaches. In fact, it is expected that an overwhelming quantity of gradual rules will be drawn even from low sized contexts. The main thrust of this paper is to address to lossless reduction of the mined knowledge. To reach this goal, a possible solution consists in using results of Formal Concept Analysis that has been shown to provide useful seeds to tackle such knowledge extraction problem. However, no work has addressed the use of the FCA framework for gradual patterns. Hence, we introduce a novel Galois connection that is a *sine qua non* issue for extracting closed gradual patterns. These latter patterns will act as a lossless reduced-size nucleus of patterns.

The remainder of the paper is organized as follows. Section 2 reviews the related work focused on mining gradual rules and some basic notions of the FCA framework. Section 3 introduces our novel Galois connection definition and shows its validity and soundness. Section 4 validates the importance of our approach at reducing the hudge number of the extracted gradual closed patterns through experiments carried out over synthetic datasets. Section 5 sketches our future perspectives and presents concluding remarks.

2 Related work

In this section, we present an overview of approaches dealing with gradual rules and some basic notions related to Formal Concept Analysis.

2.1   Gradual rules

Gradual rules are applied on data sets with m attributes $(X_1, \ldots, X_m)$ defined on numeric domains $\text{dom}(X_i)$. A data set $\mathcal{D}$ is a set of rows (m-uplets) of $\text{dom}(X_1)\times, \ldots, \times \text{dom}(X_m)$. In this scope, a gradual item is defined as a pair of an attribute and a variation $* \in \{\leq, \geq\}$. Let A be an attribute then the gradual item $A^{\geq}$ means that the attribute A is increasing. It can be interpreted by "*the more A*". A gradual itemset, or gradual tendency, is then defined as a non-empty set list of several gradual items. For instance, the gradual itemset $M = A^{\geq}\ B^{\leq}$ is interpreted as "*The more A and the less B*".

Example 1. $(Salary)^{\geq}$ is a gradual item meaning that the *"Salary"* is increasing. $((Age)^{\geq}, (Salary)^{\geq})$ is a gradual itemset.

A gradual rule, denoted by $M \Rightarrow M'$, is defined as a pair of gradual itemsets on which a causality relationship is imposed. Different measures and semantics have been proposed to extract and assess this kind of rules. In what follows, we review the related works that focused on mining gradual rules.

Gradual dependencies were introduced in [4], where they are called tendency rules and denoted by $A \rightharpoonup_t B$. Hüllermeier proposed to perform a linear regression analysis on the contingency diagram depicted from the data set. The validity of the rule is assessed on the basis of the regression coefficients α, β of the line that approximates the points in the contingency diagram. The quality of the regression is given by the R^2 coefficient. A tendency rule contains one or more attributes in the condition part and only one in the conclusion part. When it contains several attributes in the condition part, the author proposes to use a logical conjunction modeled by means of a so-called t-norm[3] This method leads to a high computational cost for the linear regression and the identification of the relevant item combinations.

Another definition has been proposed in [1]. The semantic of a gradual dependence is quite different, since the authors only consider the variation fulfillment. The authors define a gradual dependence as being similar to a functional dependence by considering the degrees variation between two objects. According to [1], the gradual dependence $A \Rightarrow B$ holds in a database $\mathcal{D}$ if $\forall\ o=(x, y)$ and $o'=(x', y') \in \mathcal{D}$, $A(x) < A(x')$ implies $B(y) < B(y')$.

A new definition of gradual dependence was proposed in [8] using fuzzy association rules. The authors take into account the variation strength in the degree of fulfilment of an imprecise property by different objects. Hence, a gradual dependence holds in a database $\mathcal{D}$ if $\forall\ o=(x, y)$ and $o'=(x', y') \in \mathcal{D}$, $v_{*1}(A(x), A(x'))$ implies $v_{*2}(B(y), B(y'))$, where v_* is a variation degree of an attribute

---

[3] A triangular norm t-norm is a function $\top : [0, 1] \times [0, 1] \rightarrow [0, 1]$ verifying, $\forall$ x and y $\in [0, 1]$, these following properties:

$\top$ is commutative $\top$ (x, y) = $\top$ (y, x),

$\top$ is associative $\top$ (x, $\top$ (y, z)) = $\top$ ($\top$ (x, y), z),

$\top$ is increasing $\top$ (x, y) $\leq$ $\top$ (z, t) if x $\leq$ z and y $\leq$ t,

$\top$ (x, 1) = x.

between two different objects. In both propositions [1] and [8], the authors propose to build a modified data set $\mathcal{D}'$ that contains as many rows as there are pairs of distinct objects in the initial data set $\mathcal{D}$.

Another definition of support and confidence of a gradual itemset, as defined above, was proposed in [5]. In fact, the support of a gradual itemset $A_1^{*_1}, \ldots, A_p^{*_p}$, is defined as the maximal number of rows $\{r_1, \ldots, r_l\}$ for which there exists a permutation π such that $\forall\, j \in [1,\, l-1]$, $\forall\, k \in [1,\, p]$, it holds $A_k(r_{\pi_j}) *_k A_k(r_{\pi_{j+1}})$. More formally, denoting $\mathcal{L}$ the set of all such sets of rows the support of a gradual itemset is defined as follows.

Definition 1. *Let* $s = A_1^{*_1}, \ldots, A_p^{*_p}$ *be a gradual itemset, we have:*

$$supp(s) = \frac{\max_{L_i \in \mathcal{L}} |L_i|}{|\mathcal{D}|}.$$

The authors propose a heuristic to compute this support for gradual itemsets, in a level-wise process that considers itemsets of increasing lengths. Recently Di Jorio et al. [6], considered the same definition considered the same definition that was proposed within the conflict set based approach, and proposed an efficient method based on the precedence graph. In this method, named GRITE that stands for GRadual ITemset Extraction, the data is represented through a graph whose nodes are defined as the objects in the data, and the vertices express the precedence relationships derived from the considered attributes.

In [7], the authors propose to calculate the support in a different way by using the *Kendall tau ranking correlation coefficient*. This coefficient calculates the number of n-uplets pairs that can be ordered in the database according to the considered gradual pattern.

Unfortunately, in all surveyed approaches, reducing the quantity of mined patterns was not the main concern. In the following subsection, we recall some key settings from the FCA framework presenting some pioneering results towards defining a concise representation of gradual patterns.

2.2 Formal concept analysis

Formal Concept Analysis (FCA) is based on the mathematical theory of complete lattices (for further information please refer to [3]). This theory provides a powerful mathematical framework that has been used in many fields of computer science.Indeed, this theory has been employed in data mining to extract a representative set of itemsets that can be used to extract association rule in an efficient manner. To do so, a formel context must be defined as a triplet ($\mathcal{R}$, $\mathcal{O}$, $\mathcal{I}$) where $\mathcal{O}$ is a set of transactions or objects, $\mathcal{I}$ is a finite set of items and $\mathcal{R}$ is a binary relation $\mathcal{R} \subseteq \mathcal{O} \times \mathcal{I}$. The first step to construct the lattice is to define a Galois connection between two derivation operators: one mapping a set of objects into a set of items and the other one mapping a set of items into a set of objects.

For a set $O \in \mathcal{O}$ and $I \in \mathcal{I}$ the two mapping operators denoted by f and g are defined, respectively, as follows:

- $f : \mathcal{P}(\mathcal{O}) \to \mathcal{P}(\mathcal{I})$, $f(O) = \{i \in \mathcal{I} \text{---} (o, i) \in \mathcal{R}, \forall\, o \in O\ \}$
- $g : \mathcal{P}(\mathcal{I}) \to \mathcal{P}(\mathcal{O})$, $g(I) = \{o \in \mathcal{O} \text{---} (o, i) \in \mathcal{R}, \forall\, i \in I\ \}$

These two mapping operators f and g induce a Galois connection between the powerset of objects and the powerset of items. This Galois connection means that f and g are dually adjoint, *i.e.*, $O \subseteq g(I) \Rightarrow I \subseteq f(O)$ for a set of objects O and a set of items I. The two composite operators $f \circ g$ and $g \circ f$ are *closure operators* (*i.e.*, they keep the properties of monotonicity, extensivity and idempotency).
A formal concept is a pair (O, I) of a set of objects $O \in \mathcal{O}$ and a set of items $I \in \mathcal{I}$, where $f(O) = I$ and $g(I) = O$. The set of all concepts that can be extracted from a gradual formal context form a complete *lattice* provided with a partial order relation $\leq$, such that $\forall\, c_1 = (O_1, I_1)$ and $c_2 = (O_2, I_2)$ two concepts, if $c_1 \leq c_2 \Leftrightarrow I_2 \subset I_1\ (O_1 \subset O_2)$.

3 Defining new Galois mapping operators for gradual patterns

To the best of our knowledge, no previous study in the literature has paid attention to apply the Galois connection in gradual rules extraction problem. Given that classical FCA was developed for binary relationships, adapting the former results to the gradual case turns out to be an interesting formalization problem.

Within this work we aim at using FCA theory in order to formalize a new closure system characterizing gradual data.

A gradual rule is defined as a special kind of association rule reflecting a variation in the degree of membership of itemsets in a sequence of objects. A gradual rule is formulated as *"The more/less X, the more/less Y"*, where X and Y are gradual patterns. An example of a gradual rule is *"The higher the Age, the higher the Salary"*. In order to satisfy the graduation property, we must consider a set of object sequences. Indeed, in the classical FCA case the domain of an itemset is a set of objects. In our case the domain of a gradual itemset is a set of sequences satisfying this itemset. The set of sequences will be ordered by the properly defined relation $\preceq$. In what follows, we propose to define the notion of sequence and the related mathematical operations that can be applied over the set of these sequences.

3.1 Handling object sequences

Let $\mathcal{O} = o_1, \ldots, o_n$ be a set of objects. We consider a sequence to be an ordered list of objects described over attributes (items). This sequence can be represented as $\langle o_1, \ldots, o_m \rangle$. This means that objects are sorted and each object o_i has an order in the sequence.

Definition 2. *A sequence $S = \langle o_1, \ldots, o_p \rangle$ is **included** in another sequence $S' = \langle o'_1, \ldots, o'_m \rangle$, denoted by $S \subseteq S'$, if there exist integers $1 < i_1 < i_2, \ldots, < i_p < m$ such that $o_1 = o'_{i_1}, \ldots, o_p = o'_{i_p}$.*

Example 2. Let $S_1 = \langle o_1, o_4, o_6 \rangle$, $S_2 = \langle o_2, o_1, o_3, o_4, o_5, o_6 \rangle$, and $S_3 = \langle o_2, o_1, o_6, o_3, o_4 \rangle$ be three sequences, we have $S_1 \subseteq S_2$ but $S_1 \nsubseteq S_3$.

Definition 3. *Let $\mathcal{S}$ be a collection (i.e., a set) of sequences, $S \in \mathcal{S}$ is said to be maximal if $\nexists\, S' \in \mathcal{S},\ S' \neq S$ such that $S \subset S'$.*

Definition 4. *The **intersection** of two sequences S_1 and S_2 is the set of all maximal subsequences of both S_1 and S_2:*
$S_1 \cap S_2 = \{s_i | s_i \subseteq S_1,\ s_i \subseteq S_2 \text{ and } \nexists\, s_i \subset s_i' \text{ such that } s_i' \subseteq S_1 \text{ and } s_i' \subseteq S_2\}.$

Example 3. Let $S_1 = \{\langle o_1, o_2, o_4, o_7 \rangle\}$, $S_2 = \{\langle o_2, o_5, o_4, o_6, o_8, o_7 \rangle\}$ be two sequences. $S_1 \cap S_2 = \{\langle o_2, o_4, o_7 \rangle, \langle o_1, o_4, o_7 \rangle\}$.

Definition 5. *A set of sequences $\mathcal{S}$ is **included** in another set $\mathcal{S}'$, denoted by $\mathcal{S} \preceq \mathcal{S}'$, if $\forall\, S$ in $\mathcal{S}$, $\exists\, S' \in \mathcal{S}'$ s.t, $S \subseteq S'$.*

Example 4. The set of sequences $\mathcal{S}_1 = \{\langle o_5, o_6, o_7 \rangle, \langle o_2, o_4, o_7 \rangle\}$ is included in the set of sequences $\mathcal{S}_2 = \{\langle o_5, o_6, o_8, o_7 \rangle, \langle o_1, o_2, 0_4, o_7 \rangle\}$.

Based on the binary inclusion relation of sequences set defined in definition 5, the following proposition holds:

Proposition 1. *Let $\mathcal{S}$ be a set of maximal sequences. $\mathcal{P}(\mathcal{S})$ provided with the binary relation $\preceq$ is a partially ordered set (or poset).*

Proof. The binary relation $\preceq$ over the set $\mathcal{P}(\mathcal{S})$ is *reflexive*, *antisymmetric*, and *transitive*, i.e., for all S_1, S_2, and S_3 in $\mathcal{P}(\mathcal{S})$, we have:

- $S_1 \preceq S_1$. (Reflexivity)

- Let us consider S_1 and S_2 such that $\begin{cases} S_1 \preceq S_2 \ (1) \\ S_2 \preceq S_1 \ (2) \end{cases}$ According to, respectively, *(1)* and *(2)* we have $\begin{cases} \forall s_1 \in S_1, \exists s_2 \in S_2 \, s.t. s_1 \subseteq s_2 \\ \forall s_2 \in S_2, \exists s_1 \in S_1 \, s.t. s_2 \subseteq s_1 \end{cases}$
  As the maximality prevents from having two sequences S_1 and S_2 in the same set such that $S_1 \subset S_2$ or $S_2 \subset S_1$, hence the antisymmetry property holds.

- Let us consider S_1 and S_2 such that $\begin{cases} S_1 \preceq S_2 \ (3) \\ S_2 \preceq S_3 \ (4) \end{cases}$ According to *(3)* and *(4)* we have: $\begin{cases} \forall s_1 \in S_1, \exists s_2 \in S_2 \, s.t. s_1 \subseteq s_2 \ (5) \\ \forall s_2 \in S_2, \exists s_3 \in S_3 \, s.t. s_2 \subseteq s_3 \ (6) \end{cases}$
  According to *(5)* and *(6)* we have $\forall\, s_1 \in S_1$, $\exists\, s_3 \in S_3$ s.t. $s_1 \subseteq s_3$. Hence, we have $S_1 \preceq S_3$ and the transitivity property is satisfied.

3.2 Gradual Galois Connection

In this paper, we propose a new definition of Galois connection taking into account the graduality aspect. Hence, we first define the notion of a gradual formal context.

Definition 6. *Gradual formal context A gradual formal context is defined as the quadruplet $\mathcal{K} = (\mathcal{O}, \mathcal{I}, \mathcal{Q}, R)$ describing a set of objects $\mathcal{O}$, a finite set $\mathcal{I}$ of attributes (or items), a finite set of quantities or values $\mathcal{Q}$ and a binary relation R (i.e., $R \subseteq \mathcal{O} \times \mathcal{I}$). Each pair $(o, i^q) \in R$, means that the value of the attribute (item) i belonging to $\mathcal{I}$ in the object o belonging to $\mathcal{O}$ is q.*

Example 5. An example of a gradual formal context $\mathcal{K} = (\mathcal{O}, \mathcal{I}, \mathcal{Q}, R)$ is sketched in Table 1. We have $(o_1, Age^{22}, Salary^{1200}) \in \mathcal{R}$.

	Age	Salary	Loan
o_1	22	1200	4
o_2	24	1850	2
o_3	30	2200	3
o_4	28	3400	1

Table 1. Formal gradual context

Let $\mathcal{K} = (\mathcal{O}, \mathcal{I}, \mathcal{Q}, R)$ be a gradual formal context, we define bellow the two closure operators f and g:

$$f : \mathcal{P}(\mathcal{S}) \rightarrow \mathcal{P}(\mathcal{I})$$
$$f(S) = \{i^* | \ \forall \ s \in S, \ \forall o_l, o_k \in s \ \text{s.t.} \ (o_l, i^{q_1}), (o_k, i^{q_2}) \in \mathcal{R} \ \text{and} \ k < l \ \text{we have} \ q_1 * q_2\}$$

The mapping function f returns all gradual items with their respective variation respecting all sequences in S.

$$g : \mathcal{P}(\mathcal{I}) \rightarrow \mathcal{P}(\mathcal{S})$$
$$g(I) = \{s \in S | \ s \ \text{is maximal in} \ S \ \text{and} \ \forall o_l, o_k \in s \ \text{s.t.} \ k < l \ and \ (o_l, i^{q_1}), (o_k, i^{q_2}) \in \mathcal{R}, \ \forall i^* \in I \ \text{we have} \ q_1 * q_2 \}$$

The mapping function g returns the set of maximal sequences respecting the variations of all items in I.

The two mapping functions g and f are respectively defined over the power set of $\mathcal{I}$ and the power set of sequences of $\mathcal{S}$. Given that the intersection of a set of object sequences may result more than one sequence, we consider the power set of sequences. The function f is applied on a set of sequences whereas g is applied on a set of gradual attributes. On the one hand, $f(S)$ returns the gradual itemset I, such that every gradual item $i \in I$ is provided with the corresponding

variation $*$ through the sequences of S. On the other hand, $g(I)$ looks for all sequences verifying the variation of each item in I.

The set of gradual itemsets can be ordered by the standard inclusion binary relation $\subseteq$. However, the set of sequences is ordered by the binary relation $\preceq$.

Example 6. Let us consider the context illustrated in Figure 1. Thus, we have for example $f(\langle o_1, o_2, o_4\rangle, \langle o_1, o_2, o_3\rangle)= \{Age^{\geq}\ Salary^{\geq}\}$ and $g(\{Age^{\geq}\ Loan^{\leq}\})=\{\langle o_1, o_2, o_3\rangle, \langle o_1, o_3\rangle\}$

Based on the definitions and propositions introduced above, we can now demonstrate that we have construct a Galois Connection framework for the gradual case. As stated above, this will allow us to mine for concise representations, thus reducing the size of the results presented to end-users. It should be noted that this reduction is of great importance as users are often drawn in resulting patterns.

Proposition 2. *For sets of sequences S and $S' \in \mathcal{S}$, and sets of gradual itemsets I and I' the following properties hold:*

1) $S \preceq S' \Rightarrow f(S') \subseteq f(S)$ **1')** $I \subseteq I' \Rightarrow g(I') \subseteq g(I)$

2) $S \preceq g(f(S))$ **2')** $I \subseteq f(g(I))$

Proof. Each property is respectively proved as follows:

1) $S \preceq S'$ means that $S'= S \cup \{s'_1, \ldots, s'_p\}$. Then every gradual item i^* belonging to $f(S')$ holds on $S \cup \{s'_1, \ldots, s'_p\}$ and thus holds on S. Hence, $f(S)$ includes all gradual items i^* included in $f(S')$, and may include some other ones. Therefore $f(S) \subseteq f(S')$.

2) We have:
- $f(S) =\{i^*| \forall s \in S, \forall o_l, o_k \in s$ s.t. $(o_l, i^{q_1}), (o_k, i^{q_2}) \in \mathcal{R}$ and $k < l$ we have $q_1 * q_2\}$
- $g(f(S))= \{s \in S|\ s$ is maximal in S and $\forall o_l, o_k \in s$ s.t. $k < l$ *and* $(o_l, i^{q_1}), (o_k, i^{q_2}) \in \mathcal{R}, \forall i^* \in f(S)$ we have $q_1 * q_2\ \}$

Obviously, if $s \in S$ then $s \in g(f(S))$. Hence, $S \preceq g(f(S))$.

1') For every $s' \in g(I')$ we have $\forall i'^* \in I'$, i'^* has an uniform variation $*$ in s'. In particular, for every $i^* \in I$ as $I \subseteq I'$. Thus, i^* has also the same uniform variation in S'. Therefore, $s' \in g(I)$ which means that $g(I') \preceq g(I)$.

2') We have:
- $g(I) =\{s \in S|\ s$ is maximal in S and $\forall o_l, o_k \in s$ s.t. $k < l$ *and* $(o_l, i^{q_1}), (o_k, i^{q_2}) \in \mathcal{R}, \forall i^* \in I$ we have $q_1 * q_2\ \}$
- $f(g(I)) =\{i^*| \forall s \in g(I), \forall o_l, o_k \in s$ s.t. $(o_l, i^{q_1}), (o_k, i^{q_2}) \in \mathcal{R}$ and $k < l$ we have $q_1 * q_2\}$

Obviously, if $i^* \in I$ then $i^* \in f(g(I))$. Hence, $I \subseteq f(g(I))$.

Proposition 3. *The composite operators $f \circ g$ and $g \circ f$ form two closure operators, respectively defined on the sets of sequences and the set of itemsets.*

Proof. Considering both functions f and g as previously defined, the following properties hold:

- Monotonicity : $I \subseteq I' \Rightarrow f{\circ}g(I) \preceq f{\circ}g(I')$
  We have $I \subseteq I'$, property 1') yields to $g(I) \preceq g(I')$, and by property 1) we get $f{\circ}g(I) \subseteq f{\circ}g(I')$.
- Extensivity : $I \subseteq f{\circ}g(I)$
  Proved in property 2').
- Idempotency : $f{\circ}g(I)= f{\circ}g(f{\circ}g(I))$
  According to property 2') we have $f{\circ}g(I) \subseteq f{\circ}g(f{\circ}g(I))$ With $S = g(I)$ and according to property 2) we have $g(I) \preceq g{\circ}f{\circ}g(I)$, and property 1) yields to $f{\circ}g(I) \supseteq f{\circ}g{\circ}f{\circ}g(I)$.
  Hence we can conclude that $f{\circ}g(I)=f{\circ}g(f{\circ}g(I))$.

These properties are valid in a dual manner for closure operators $f{\circ}g$ and $g{\circ}f$. Indeed:

- Monotonicity : $S \preceq S' \Rightarrow g{\circ}f(S) \preceq g{\circ}f(S')$
  We have $S \preceq S'$, property 1) yields to $f(S') \subseteq f(SS)$, and by property 1') we get $g{\circ}f(S) \preceq g{\circ}f(S')$.
- Extensivity : $S \preceq g{\circ}f(S)$
  Proved in 2).
- Idempotency : $g{\circ}f(S)= g{\circ}f(g{\circ}f(S))$
  According to property 2) we have $S \preceq g{\circ}f(S)$. With 1), it yields that $f(S) \supseteq f \circ g{\circ}f(S)$ and according to property 1'), $g{\circ}f(S) \supseteq g{\circ}f{\circ}g{\circ}f(S)$. With $I = f(S)$ and according to property 2') we have $f(S) \subseteq f{\circ}g{\circ}f(S)$, and property 1') yields to $g{\circ}f(S) \succeq g{\circ}f{\circ}g{\circ}f(S)$. Equality follows from the fact that the closure operator, by definition of g, always gives a set where no sequence is a subsequence of another.

The result of all these propositions define a new framework for gradual closed patterns which are a concise representation of gradual freaquent patterns. In fact, let us consider the definitions below:

Definition 7. *Gradual formal concept* *The pair (S, I), such that $S \in \mathcal{S}$ and $I \in \mathcal{I}$, is called a gradual concept if $f(S) = I$ and $g(I) = S$. O is called the extension and I the intension of the gradual concept.*

Definition 8. *Gradual closed itemset* *Let us consider the formal context $\mathcal{K}$ = (O, I, Q, R), a gradual subset $I \subseteq \mathcal{I}$, I is called gradual closed itemset if and only if it is equal to its closure, i.e., $f{\circ}g(I)= I$.*

Definition 9. *Minimal gradual generator* *A gradual itemset $h \subseteq \mathcal{I}$ is called minimal gradual generator of another gradual closed itemset I if $f{\circ}g(h)= I$ and does not exist $h' \subseteq \mathcal{I}$ such that $h' \subset h$. The set $\mathcal{GGM}$ of all gradual minimal generators of a gradual closed itemset I is defined as bellow:*
$\mathcal{GGM}= \{h{\subseteq} \mathcal{I}|f{\circ}g(h)=I \wedge \nexists\ h' \subset h\ tel\ que\ f{\circ}g(h')= I\}$

The closure operator as defined above infers an equivalence relation over the power set of $\mathcal{I}$ (i.e., $\mathcal{P}(\mathcal{I})$). Hence, $\mathcal{P}(\mathcal{I})$ is partitionned into several disjoined subsets called equivalence classes. In one classe all elements have the same support and the minimal gradual genarators are the smallest uncomparable elements

in the class. However, gradual closed itemsets are the largest elements in that class ([9]).

Proposition 4. *A set of gradual formal concepts $\mathcal{GC}_\mathcal{K}$ extracted from a formal context $\mathcal{K}$, ordered using the set inclusion relation, form a complete lattice $\mathcal{L}_\mathcal{K}$ = $(\mathcal{GC}_\mathcal{K}, \subseteq)$, called gradual Galois lattice.*

Example 7. Let us consider the gradual formal context given by Table 1, a part of its related gradual Galois lattice is depicted in Figure 1.
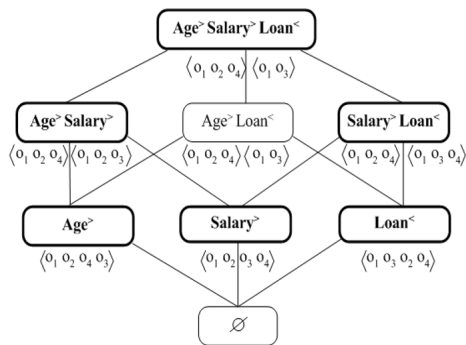

Fig. 1. A part of the gradual Galois lattice related to the gradual formal context depicted in Table 1.

Remark 1. When only considering gradual closed itemsets with the inclusion relation $\subseteq$, the resulting structure only preserves the join operator which is a sub-order of the gradual itemset lattice, thus often is much smaller. This structure is represented in bold in Figure 1. The gradual closed itemset lattice can be used as a formal framework for discovering gradual itemsets given the basic properties that the support of a gradual itemset is equal to the support of its closure. On the other hand, using the gradual itemset lattice, we can directly generate a reduced set of gradual rules without loss of information.

3.3 Implementation

In this paper, we aim at presenting the importance of our approach in reducing the number of gradual closed patterns that can be extracted from a data set compared to the number of all gradual patterns. So, to reach this goal, we have

adopt a post-treatment method to extract equivalence classes (*i.e.*, the set of gradual closed itemsets and their respective minimal gradual generators) from the gradual frequent set provided by the approach of Di Jorio and al. [6]. To do so, we group together frequent gradual itemsets by their support. Hence, we build the set of equivalence class groups $\mathcal{GCE}$ in which every group $\mathcal{E}$ may contain one or several equivalent class(es). This can be explained by the double variation $\{\leq, \geq\}$ taken into account in our extraction process. Indeed, we can find in the same group two (or more) gradual closed itemsets having the same items but with different variations. The algorithm 1 illustrates the equivalence classes extraction process. In this algorithm, we call the function RECHGEN that determines the set of minimal gradual generators for every maximal gradual pattern, $\mathcal{C}$ in $\mathcal{E}$ (corresponding to a closed gradual pattern). For every gradual closed pattern $\mathcal{C}$ belonging to a group $\mathcal{E}$ taken as inputs, this function returns the gradual generators set of $\mathcal{C}$. The pseudo-code and the notations used in our approach are, respectively presented in algorithm 1 and table 2. The RECHGEN function is given by algorithm 2:

$\mathcal{IGF}$	: Gradual frequent patterns.
$\mathcal{GCE}$	: Set of equivalence classe groups.
$\mathcal{IGC}$	: Gradual closed patterns.
$\mathcal{GGM}$	: Minimal gradual generator of a graduel closed pattern.

Table 2. Notations used in the algorithm

Algorithm 1: Equivalence classes extraction algorithm

Input: Gradual frequent patterns ($\mathcal{IGF}$) and their respective support.

Output: Gradual closed patterns $\mathcal{IGC}$ and their respective minimal gradual generators.

begin

    Insert into $\mathcal{GCE}$ **Select** $*$ **From** $\mathcal{IGF}$ **Group By** *support* ;

    foreach *($\mathcal{E} \in \mathcal{GCE}$)* **do**

        $\mathcal{C} = \text{MAX}(\mathcal{E})$ **while** *(C)* $\neq \{\emptyset\}$ **do**

            $\mathcal{E} = \mathcal{E}$ - $\{\mathcal{C}\}$

            $\mathcal{C}.\mathcal{GGM} = \text{RECHGEN}(\mathcal{C}, \mathcal{E}$)

            $\mathcal{C} = \text{MAX}(\mathcal{E})$

        $\mathcal{IGC} = \bigcup \{\mathcal{C}\}$

end

---

4 Experiments

This section validates the interest of our approach. In fact, the number of gradual closed patterns that can be extracted from a dataset is much smaller than the

---

Algorithm 2: RECHGEN.

Input: A gradual closed pattern $(\mathcal{C})$ and a group of gradual frequent patterns $(\mathcal{E})$.

Output: Minimal gradual generators set $\mathcal{GGM}$ of $\mathcal{C}$.

begin
    foreach $(e \in \mathcal{E})$ **do**
        if $e \subset \mathcal{C}$ *et* $\nexists\, e' \subset e$ *such that* $(e' \in \mathcal{E})$ *and* $(e' \subset \mathcal{C})$ **then**
            $\mathcal{GGM} = \mathcal{GGM} \cup \{e\}$
    Return $\mathcal{GGM}$
end

---

total number of frequent gradual patterns. In the other hand, the set of gradual closed patterns and their respective gradual generators can be used to define an irreducible compact nuclei (*i.e.,* generic basis) of gradual rules (this point will be discussed with more details in a further paper). To rate the importance of our approach, we ran experiments on synthetic datasets. These datasets were generated by a modified version of IBM Synthetic Data Generation Code for Associations and Sequential Patterns[4]. Let us note that these datasets are very dense and, even for a high value of the minimal support, a huge number of gradual patterns can be extracted. In most of the cases, techniques allowing to obtain gradual knowledge are generally driven on bases containing a weak number of objects and attributes. In our experiments, we focus on the variation of gradual closed patterns with regard of frequent ones according to the minimal support (minsup) value, and the variation of the attributes and object number. The fig-
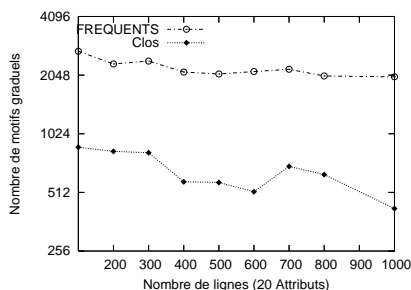

Fig. 2. Gradual closed vs. gradual frequent patterns with the variation of objects number.
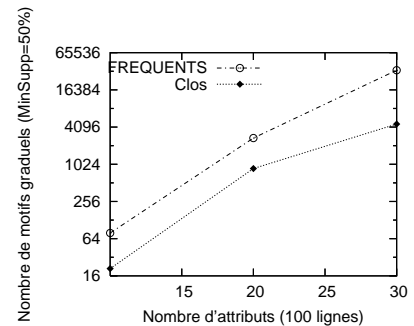

Fig. 3. Gradual closed patterns vs gradual frequent patterns with the variation of attributes number

ure 2 shows, for 20 attributes and a minsupp equal to $0, 5$, the number of closed

---

[4] www.almaden.ibm.com/software/projects/hdb/resources.shtml

patterns and frequent patterns according to the number of lines. The reported values depicted in this figure show that both closed and frequents gradual patterns are varying with the number of lines in a linear manner. We have to note that we have a logarithmic scalability and the difference is very large. However this number varies in an exponential manner with the number of attributes and with the *minsup* values as, respectively, shown in Figures 3 and 4 for a dataset of 100 lignes and 40 columns.
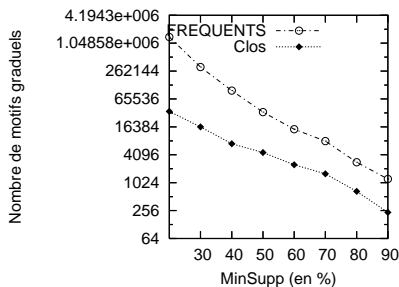

Fig. 4. Number of gradual closed patterns vs frequents with minsup variation
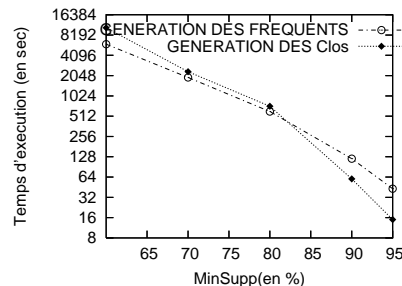

Fig. 5. Evolution of the computation time for discovering gradual closed and frequent patterns vs the variation of the minsup value.

In this paper, we aim at showing the importance reduction of the number of the extracted closed patterns with comparison with frequents ones. As mentioned above, our approach is a *post-traitement* of [6]. So, the run times are thus a little longer as shown in Figure 5

5 Conclusion and future issues

In this paper, we propose an approach for mining a concise representation of gradual patterns. Indeed, we proposed a novel closure system in order to extract gradual closed patterns. As it is expected, these gradual closed patterns will provide a high rate of compactness *vs* the total number of gradual patterns, allowing by the ease of their extraction as well as their manageability by the end-user. The dedicated algorithms and experiments show this fact. further works include the study of other optimizations in order to im- prove the efficiency of our algorithms. In the other hand, a compelling issue that we have to tackle is the definition of a cover of the gradual association rules. This task is of paramount issue since it allows to present to end-users only a manageable reduced quantity of gradual association rules. A theoretical connected task will be the proof of

the soundness and the completeness of a derivation mechanism from the cover of gradual association rules.

References

1. F. Berzal, J. Cubero, D. Snchez, M. Vila, and J. Serrano. An alternative approach to discover gradual dependencies. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 15(5):559– 570, 2007.
2. S. Galichet, D. Dubois, and H. Prade. Imprecise specification of ill-known functions using gradual rules. *International Journal of Approximate Reasoning*, 35:205–222, 2004.
3. B. Ganter and R. Wille. *Formal Concept Analysis*. Springer-Verlag, 1999.
4. E. Hüllermeier. Association rules for expressing gradual dependencies. In *Proceedings of the International Conference PKDD'2002, Helsinki, Finland*, pages 200–211, August 19-23, 2002.
5. L. D. Jorio, A. Laurent, and M. Teisseire. Fast extraction of gradual association rules: a heuristic based method. In *Proceedings of the 5th international conference on Soft computing as transdisciplinary science and technology (CSTST '08)*, pages 205–210, Cergy-Pontoise, France, 2008.
6. L. D. Jorio, A. Laurent, and M. Teisseire. Mining frequent gradual itemsets from large databases. In *Proceedings of the International Conference Intelligent Data Analysis (IDA'09)*, page to appear, Lyon-France, 2009.
7. A. Laurent, M.-J. Lesot, and M. Rifqi. Graank: Exploiting rank correlations for extracting gradual dependencies. In *Proc. of FQAS'09*, 2009.
8. C. Molina, J. Serrano, D. Snchez, and M. Vila. Measuring variation strength in gradual dependencies. In *Proceedings of the International Conference EUSFLAT'2007, Ostrava, Czech Republic*, pages 337–344, 2007.
9. S. B. Yahia and E. M. Nguifo. Approches déxtraction de règles dássociation basées sur la correspondance de galois. *Ingénierie des Systèmes d'Information*, 9(3-4):23–55, 2004.