Qualitative Reasoning and Data Mining

Yakoub Salhi
CRIL - CNRS & Université d'Artois, Lens, France
salhi@cril.fr

Abstract

In this paper, we introduce a new data mining framework that is based on qualitative reasoning. We consider databases where the item domains are of different types, such as numerical values, time intervals and spatial regions. Then, for the considered tasks, we associate to each item a constraint network in a qualitative formalism representing the relations between all the pairs of objects of the database w.r.t. this item. In this context, the introduced data mining problems consist in discovering qualitative covariations between items. In a sense, our framework can be seen as a generalization of gradual itemset mining. In order to solve the introduced problem, we use a declarative approach based on the satisfiability problem in classical propositional logic (SAT). Indeed, we define SAT encodings where the models represent the desired patterns.

2012 ACM Subject Classification Information systems → Data mining; Information systems → Association rules; Theory of computation → Constraint and logic programming; Computing methodologies → Knowledge representation and reasoning

Keywords and phrases Qualitative Database, Qualitative Pattern Mining, Declarative Approach, SAT Modeling

Digital Object Identifier 10.4230/LIPIcs.TIME.2019.9

1 Introduction

Data mining techniques are applied on different data types, such as transactions, sequences, graphs, texts, etc. In order to consider complex aspects of the real world, it is interesting to extend these techniques for knowledge discovery to new complex data, such as spatio-temporal pieces of information. However, it is important in this context to take into account the simplicity of the pattern structure. Thus, the challenge in this work is to propose a framework that allows us to deal with different complex data types and discovering patterns having a simple structure.

Qualitative reasoning is concerned with facilitating reasoning about complex entities and pieces of information through symbolic representation formalisms. In particular, this kind of reasoning is strongly related to human one and, for instance, it can be used for dealing with pieces of information that come from natural language. In the literature, the qualitative formalisms are widely used for reasoning about two physical entities of the world that are time and space (e.g. see [21]). Indeed, qualitative spatial and temporal reasoning is an important research field in Artificial Intelligence in general, and knowledge representation in particular. The spatial and temporal representation formalisms allow reasoning about configurations by abstracting numerical quantities of space and time thanks to qualitative relations, such as *inside*, *before*, *after*, etc. One of the best known qualitative representation formalisms is the Point Algebra [31], which allows representing and reasoning about the possible relative positions between two points on the timeline. The Interval Algebra [2, 3], for its part, is used for reasoning about the possible positions between two intervals. Furthermore, regarding qualitative spatial reasoning, the Region Connection Calculus RCC8 [25] is one of the most studied formalisms in qualitative reasoning, which concerns topological relations between two spatial regions.

26th International Symposium on Temporal Representation and Reasoning (TIME 2019).
Editors: Johann Gamper, Sophie Pinchinat, and Guido Sciavicco; Article No. 9; pp. 9:1–9:15
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

In this work, we propose a framework for data mining using qualitative reasoning, which allows considering different data types, such as numerical values, time intervals and spatial regions. To this end, we first introduce the notion of qualitative database, which is defined by associating to each item a constraint network in a qualitative formalism representing the relations between the pairs of objects of the database w.r.t. this item. Then, we describe data mining tasks for discovering qualitative covariations, called qualitative itemsets, in the previous kind of databases. For instance, the desired patterns can capture pieces of information of the form "a variation of an item a w.r.t. the qualitative relation r_1 is associated with a variation of b w.r.t. the qualitative relation r_2". In a sense, the proposed tasks can be seen as a generalization of those related to gradual itemsets where the qualitative relations that are considered in the extracted patterns are only $\leq$ and $\geq$ on numerical values [7, 10, 11, 20]. We express the interestingness predicate on the qualitative itemsets in a database through two different definitions of support. The first definition takes into consideration a local view by reasoning about the pairs of objects that satisfy the partial order induced by the itemset, while the second is obtained by reasoning about the sequences respecting the previous partial order. These two definitions allow extracting interesting recurrent pieces of information. Finally, we use a declarative and flexible solution for solving the introduced data mining tasks based on the use of the satisfiability problem in classical propositional logic (SAT). Indeed, we define for each task a SAT encoding whose models allow us to obtain all the desired patterns. Thus, we follow in our solution the constraint programming based approach for data mining initiated in [24, 13], which offers a declarative and flexible representation model.

The rest of the paper is organized as follows. After describing related works in Section 2, we introduce in Section 3 the notion of qualitative database. In Section 4, we present the data mining tasks proposed in this work. In Section 5, we describe our SAT-based encodings for solving these tasks, while Section 6 concludes the paper.

2 Related Works

The most related data mining tasks to our framework are those concerned with extracting gradual itemsets [7, 10, 11, 20]. A gradual itemset is a pattern expressing covariations of items having as domains sets of numerical values. For instance, the gradual itemset containg three gradual items $\{sport^{\geq}, weight^{\geq}, diseases^{\leq}\}$ can be used to express the fact "the higher the time of physical activity, the higher the weight loss, and the fewer the number of diseases". The gradual itemset structure allows analyzing numerical data in a simple and intuitive way, since it avoids the quantitative aspects of the considered data.

The data mining framework introduced in this work can be seen as a generalization of that of mining gradual itemsets in the case of numerical data. Indeed, instead of using only the inequality relations $\leq$ and $\geq$, many binary qualitative relations on different data types can be used in our framework, in particular qualitative relations on time intervals and spatial regions.

It is worth noting that we use in our framework measures for determining the quality of a qualitative itemset similar to those proposed in the case of gradual itemsets. In fact, in the same way as in gradual itemset mining, we consider two distinct definitions of support: the first definition considers the pairs of objects that respect the itemset, while the second definition is obtained by reasoning about the length of the sequences that respect the pattern. More precisely, the first definition of support corresponds to the numbers of pairs of objects that satisfy the partial order associated to the pattern, and the second definition corresponds the length of the longest sequences of objects that are ordered using the partial order induced by the pattern.

The use of a declarative approach for data mining was originally proposed in [24] for performing different tasks. Specifically, the authors have demonstrated that constraint programming is an appropriate tool in many respects in itemset mining. One of the main motivations lies in the fact that this framework offers a flexible and generic representation model. Indeed, new constraints often require new implementations for specialized data mining approaches, which can often be integrated in a fairly simple way into declarative frameworks, since it is not needed to change the solving tools. In addition, the continual evolution in the efficiency of tools dedicated to problems that can be used for data mining modeling, like ASP (*Answer Set Programming*), CSP (*Constraint Satisfaction Problem*) and SAT, is a strong argument in favor of using approaches based on these problems. Thus, from this first work, a new line of research has emerged within the data mining community. Indeed, in recent years, many works using CSP and SAT for different data mining tasks have been proposed in the literature (e.g. [13, 16, 12, 30, 19, 9]). In particular, in [8], the authors show that their SAT-based approach achieves better performance than state-of-the-art specialized techniques. In this work, we use a SAT-based approach for solving all the considered data mining tasks. Let us note that a SAT-based approach was recently used for extracting gradual patterns in [22].

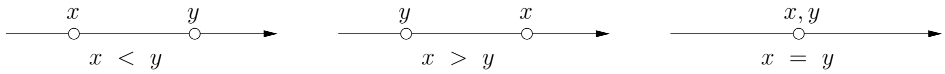
3 Qualitative Database

In this section, we introduce the notion of qualitative database. The main idea consists in associating to each item a constraint network in a qualitative formalism representing the relations between the pairs of objects of the database w.r.t. this item. To illustrate our proposal, we consider three distinct qualitative formalisms for reasoning about time and space, namely Point Algebra [31], Interval Algebra [2, 3] and Region Connection Calculus RCC8 [25].

Given a finite set S, we use $\mathcal{P}(S)$ and $|S|$ to denote respectively the powerset and the cardinality of S. Given a finite set of items $\mathcal{I}$, V_a is used to denote the domain of the item $a \in \mathcal{I}$. The domain of an item can be a numerical value, a temporal interval, a spatial region, etc. Further, we associate to each item a a finite set of qualitative base relations B_a, which consists of *jointly exhaustive* and *pairwise disjoint* relations, i.e., for each $(v, v') \in V_a \times V_a$, there exists exactly one $b \in B_a$ such that $(v, v') \in b$. Further, we only consider the set of qualitative base relations B_a that contains the identity relation $id = \{(v, v') \in V_a \times V_a \mid v = v'\}$, and is closed under the inverse operation $(\cdot)^{-1}$, namely whenever b is in B_a, the inverse $(b)^{-1}$ is also in B_a. A qualitative relation is said to be *universal* if it contains all the base relations.

The *weak composition* of two base relations b and b' in B_a, denoted $b \diamond b'$, is defined as the set of base relations $\{b'' \in B_a \mid \exists (v, v') \in b \;\&\; (v', v'') \in b' \;\&\; (v, v'') \in b''\}$. The weak composition operation is extended to the relations in $\mathcal{P}(B_a)$ as follows: $r \diamond r' = \bigcup_{b \in r, b' \in r'} b \diamond b'$. In this context, it is worth mentioning that the *composition* $\circ$ of two relations is defined as follows: $r \circ r' = \{(v, v') \mid \exists v'', (v, v'') \in r \;\&\; (v'', v') \in r'\}$. In other words, $r \diamond r'$ is the largest set of base relations where each one shares at least one value with $r \circ r'$.

For example, consider the point algebra (PA) qualitative formalism described in Figure 1, which has been mainly used for temporal reasoning. Indeed, PA can be used to encode temporal relations between two points in the timeline. We also describe in Figure 2 the base relations of two other qualitative formalisms: interval algebra (IA) and region connection calculus RCC8. The formalism IA allows encoding relative relations between intervals, while RCC8 allows encoding topological relations between two regions. For instance, the expression $DC(Region1, Region2)$ represents the fact that the two spatial regions $Region1$ and $Region2$ are disconnected.



(a) The base relations of Point Algebra.

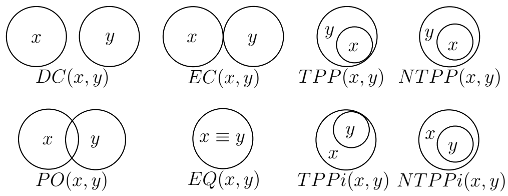
b	$(b)^{-1}$
$<$	$>$
$>$	$<$
$=$	$=$

(b) The inverse table of Point Algebra.

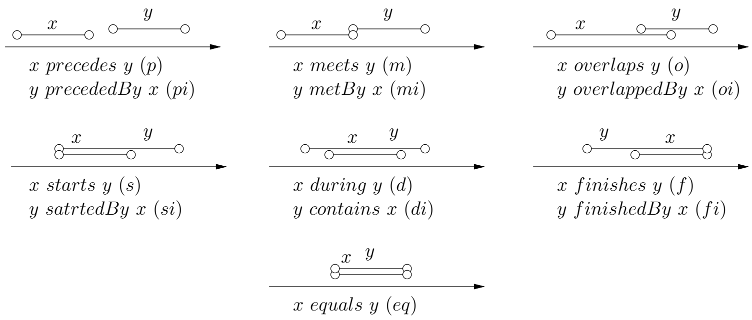
$\diamond$	$<$	$>$	$=$
$<$	$\{<\}$	$\{<,>,=\}$	$\{<\}$
$>$	$\{<,>,=\}$	$\{>\}$	$\{>\}$
$=$	$\{<\}$	$\{>\}$	$\{=\}$

(c) The composition table of Point Algebra.

Figure 1 Point Algebra.



(a) The base relations of RCC8.



(b) The base relations of Interval Algebra.

Figure 2 The qualitative formalisms RCC8 and Interval Algebra.

▶ **Definition 1** (Qualitative Column). *A* q-column *is a structure of the form* $c = (a, \mathcal{O}, R)$*, where* a *is an item, denoted* $item(c)$*,* $\mathcal{O}$ *is a finite non empty set of objects, denoted* $obj(c)$*, and* R *is a mapping form* $\mathcal{O} \times \mathcal{O}$ *to* B_a*, denoted* $rel(c)$*.*

Let us now introduce the notion of qualitative database, which is defined by associating to each item a constraint network in a qualitative formalism representing the relations between the pairs of objects of the database w.r.t. this item.

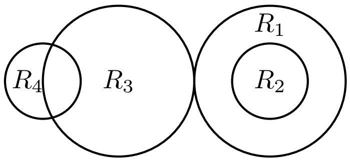
▶ **Definition 2** (Qualitative Database). *A* qualitative database *is a structure of the form* $(\mathcal{O}, \mathcal{I}, \mathcal{C})$*, where* $\mathcal{O}$ *is a finite non empty set of objects,* $\mathcal{I}$ *is a finite non empty set of items and* $\mathcal{C}$ *is a set of q-columns s.t. (i)* $|\mathcal{C}| = |\mathcal{I}|$*, (ii)* $\forall c \in \mathcal{C}$*,* $obj(c) = \mathcal{O}$*, and (iii)* $\forall a \in \mathcal{I}$*, there exists exactly one* $c \in \mathcal{C}$ *s.t.* $item(c) = a$*.*

In the sequel, we sometimes use R_a to denote $rel(c)$ where c is the qualitative column associated to the item a.

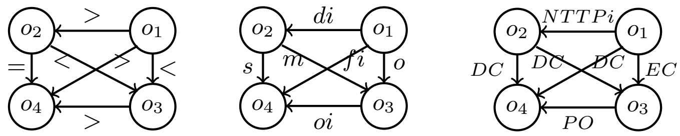
For example, we describe in Figure 3 a qualitative database: we provide in (a) a database using values in item domains, in (b) the concrete situation of the considered spatial regions, and in (c) the qualitative database. For instance, the edge between o_1 and o_2 in the left-hand graph represents the qualitative base relation in PA $>(o_1, o_2)$, usually denoted $o_1 > o_2$.

objects	a	b	c
o_1	2	[1,4]	R_1
o_2	1	[2,3]	R_2
o_3	4	[3,6]	R_3
o_4	1	[2,4]	R_4

(a) A database using values in item domains.



(b) A representation of the real situation the regions R_1, R_2, R_3 and R_4.



(c) The qualitative database corresponding to the database in (a).

Figure 3 A Qualitative Database.

4 Mining Qualitative Itemsets

In this section, we introduce data mining tasks for discovering qualitative covariations in qualitative databases. For instance, the patterns in this context can be used to capture pieces of information of the form "a variation of a w.r.t. the qualitative relation r_1 is associated with a variation of b w.r.t. the qualitative relation r_2".

▶ **Definition 3** (Qualitative Itemset). *A qualitative itemset is a finite non empty set of qualitative items I, where a* qualitative item *is a structure of the form a^r where a is an item and $r \subseteq B_a$.*

Let us now describe the partial order on the objects of a database that is induced by a qualitative itemset, and also the notion of ordered sequence that is used for defining the support of a qualitative itemset.

▶ **Definition 4** (Induced Partial Order). *Let $\mathcal{D} = (\mathcal{O}, \mathcal{I}, \mathcal{C})$ be a qualitative database, $o, o' \in \mathcal{O}$ and $I = \{a_1^{r_1}, \dots, a_k^{r_k}\}$ a qualitative itemset. Then, we say that o precedes o' w.r.t. I, written $o \preceq_I o'$, if for all $i \in 1..k$, $R_a(o, o') \in r_i$ holds.*

▶ **Definition 5** (Ordered sequence of objects). *Let $\mathcal{D}$ be a qualitative database, $L = \langle o_1, \dots, o_k \rangle$ a sequence of distinct objects in $\mathcal{D}$ and I a qualitative itemset. We say that L respects I if it is ordered with respect to $\preceq_I$, i.e., $o_i \preceq_I o_{i+1}$ for every $i \in 1..k-1$.*

We here use $\mathcal{L}(\mathcal{D}, I)$ to denote all the sequences of objects occurring in $\mathcal{D}$ that respect the qualitative itemset I.

In the same way as in gradual itemset mining, we express the quality of an itemset in a database through two different definitions of *support*. The first definition captures a local view by taking into consideration the number of pairs that satisfy the partial order induced by the qualitative itemset ($\mathcal{D} = (\mathcal{O}, \mathcal{I}, \mathcal{C})$):

$$supp_1(I, \mathcal{D}) = \frac{|\{\{o, o'\} \subseteq \mathcal{O} \mid o \neq o', o \preceq_I o'\}|}{|\mathcal{O}| \cdot (|\mathcal{O}| - 1)/2}.$$

The second definition is obtained by reasoning about the sequences that respect the qualitative itemset. Indeed, it corresponds to the length of the longest sequences that respect the considered itemset:

$$supp_2(I, \mathcal{D}) = \frac{max\{|L| \mid L \in \mathcal{L}(\mathcal{D}, I)\}}{|\mathcal{O}|}.$$

Furthermore, we consider that it is more appropriate to allow the user to select the relations that can be associated to every item in a pattern. For example, it is not interesting to consider the universal or empty relations because they do not describe any variation.

Thus, we define two problems of enumerating qualitative itemsets as follows: given a function f that maps each item a to a subset of relations $f(a) \subseteq \mathcal{P}(B_a)$ which is closed under the inverse operation and the inclusion, and a minimum support threshold v, the problems QIE1 and QIE2 consist in computing respectively the sets of qualitative itemsets $\mathcal{QIE}1(\mathcal{D}, f, v) = \{I \mid supp_1(I, \mathcal{D}) \geq v \ \& \ \forall a^r \in I, r \in f(a)\}$ and $\mathcal{QIE}2(\mathcal{D}, f, v) = \{I \mid supp_2(I, \mathcal{D}) \geq v \ \& \ \forall a^r \in I, r \in f(a)\}$.

Let us consider now two condensed representations, which are similar to those that are widely considered in itemset mining. Before that, we need the following partial order relation. Given two qualitative itemsets I and J, we have $I \sqsubseteq J$ if, $\forall a^r \in I$, $\exists a^{r'} \in J$ s.t. $r' \subseteq r$. Moreover, we have $I \sqsubset J$ if $I \sqsubseteq J$ and $I \neq J$.

▶ **Definition 6** (Closedness). *Let $\mathcal{D}$ be a database and I a qualitative itemset. Then, I is said to be a* closed qualitative itemset *in $\mathcal{D}$ w.r.t. $supp_1$ (resp. $supp_2$) if, for all qualitative itemset J with $I \sqsubset J$, $supp_1(I, \mathcal{D}) > supp_1(J, \mathcal{D})$ (resp. $supp_2(I, \mathcal{D}) > supp_2(J, \mathcal{D})$) holds.*

In other words, a qualitative itemset is closed if there is no more informative qualitative itemset that has the same support.

▶ **Definition 7** (Maximality). *Let $\mathcal{D}$ be a database, v a minimum support threshold and I a qualitative itemset. Then, I is said to be a* maximal qualitative itemset *w.r.t. $supp_1$ (resp. $supp_2$) and the threshold v if, for all qualitative itemset J with $I \sqsubset J$, $supp_1(J, \mathcal{D}) < v$ (resp. $supp_2(J, \mathcal{D}) < v$) holds.*

A qualitative itemset is maximal if there is no more informative qualitative itemset that has a support greater than or equal to the minimum support threshold.

In the context of the condensed representations, one can easily see that we have the following property.

▶ **Proposition 8** (Anti-Monotonicity). *Let $\mathcal{D}$ be a qualitative database and I and J two qualitative itemsets in $\mathcal{D}$. If $I \sqsubseteq J$ then $supp_1(I, \mathcal{D}) \geq supp_1(I, \mathcal{D})$ and $supp_2(I, \mathcal{D}) \geq supp_2(I, \mathcal{D})$.*

Therefore, using the anti-monotonicity property, computing either the closed itemsets or the maximal itemsets in $\mathcal{QIE}1(\mathcal{D}, f, v)$ and $\mathcal{QIE}2(\mathcal{D}, f, v)$ allows getting all the elements of these two sets. Furthermore, the anti-monotonicity property can be used for defining Apriori-like algorithms for solving the problems QIE1 and QIE2 in a fairly simple way. Let us recall that Apriori algorithm was originally proposed in [1] for mining frequent itemsets.

It is worth mentioning that the qualitative relations are not necessarily transitive. For example, we have $1\{<,>\}2\{<,>\}1$ in PA ($x\{<,>\}y$ means that x is different from y) without having $1\{<,>\}1$. This has as a consequence the fact that a sequence respects a qualitative itemset does not implies that its sub-sequences (by avoiding intermediate objects) respect also this pattern. Thus, in order to have transitivity, a solution can consist in restricting our mining task to the relations that satisfy $\diamond$-*idempotence*: a qualitative relation r is said to be $\diamond$-*idempotent* if $r \diamond r = r$. For example, in PA the $\diamond$-idempotent relations are $\{=\}$, $\{<\}$, $\{<,=\}$, $\{>\}$, $\{>,=\}$ and $\{<,=,>\}$, i.e., all the relations except $\{\}$ and $\{<,>\}$. That being said, we provide in this work general methods for solving QIE1 and QIE2 without considering transitivity.

In order to illustrate the mining tasks described previously, we provide now a simple example. Consider the database described in Table 1. It represents pieces of information related to a set of workers about time at work, productivity and satisfaction degree. For the corresponding qualitative database, we consider interval algebra for time at work, and point algebra for both productivity and satisfaction degree. Moreover, we only consider QIE2 with a support threshold equal to 3 without any restriction on the considered qualitative relations in the patterns on `time`, but we only consider $\{<,\leq,>,\geq\}$ on both `productivity` and `satisfaction`. A first interesting qualitative pattern is $I = \{\texttt{time}^{\{p,o,m\}}, \texttt{productivity}^{\leq}\}$, which has a support equal to 4 since it is satisfied by the sequence $\langle w_1, w_2, w_3, w_4\rangle$. In a sense, it expresses that starting work earlier increase productivity. The pattern I is not closed since it has the same supports as $J = \{\texttt{time}^{\{p,o,m\}}, \texttt{productivity}^{<}\}$. Moreover, J is closed since $J \cup \{\texttt{satisfaction}^{\leq}\}$ and $J \cup \{\texttt{satisfaction}^{\geq}\}$ are respectively 2 and 3. Moreover, $J \cup \{\texttt{satisfaction}^{>}\}$ is a maximal patterns since its support is equal to the fixed threshold and it is not included in any other pattern.

Table 1 A description of a database.

`worker`	`time`	`productivity`	`satisfaction`
w_1	5am to 9am	100	1
w_2	8am to 12am	80	4
w_3	12am to 4pm	60	5
w_4	5pm to 9pm	50	3

5 SAT-based Approach for Enumerating Qualitative Itemsets

In this section, we introduce a SAT-based approach for solving the problems QIE1 and QIE2. We first describe the satisfiability problem in classical propositional logic. We then introduce our SAT encodings for QIE1 and QIE2: the computation of the models of each encoding corresponds to the computation of the desired qualitative itemsets. We here follow the constraint programming based approach for data mining initiated in [24, 13].

5.1 Classical Propositional Logic

We here describe the syntax and the semantics of classical propositional logic. We use Prop to denote the set of propositional variables. The propositional formulas of classical propositional logic (CPL) are built using Prop, the constants $\top$, denoting *true*, and $\bot$, denoting *false*, the unary logical connective $\neg$ and the usual binary connectives $\wedge$, $\vee$, $\rightarrow$ and $\leftrightarrow$. The grammar is defined as follows:

$$\phi ::= p \mid \top \mid \bot \mid \phi \wedge \phi \mid \phi \vee \phi \mid \phi \rightarrow \phi \mid \phi \leftrightarrow \phi \mid \neg\phi$$

with $p \in \mathsf{Prop}$. The set of propositional formulas is denoted Form. We use the letters p, q, r, s to denote the propositional variables, and the Greek letters ϕ, ψ and χ to denote the propositional formulas. Moreover, given a syntactic object o, we use $Var(o)$ to denote the set of propositional variables occurring in o.

A *Boolean interpretation* $\mathcal{B}$ of a formula ϕ is defined as a function from the set of variables $Var(\phi)$ to $\{0,1\}$ (0 stands for $false$ and 1 for $true$). It is inductively extended to propositional formulas as usual:

$$\begin{array}{ll} \mathcal{B}(\top) = 1 & \mathcal{B}(\bot) = 0 \\ \mathcal{B}(\neg\phi) = 1 - \mathcal{B}(\phi) & \mathcal{B}(\phi \to \psi) = max(1 - \mathcal{B}(\phi), \mathcal{B}(\psi)) \\ \mathcal{B}(\phi \wedge \psi) = min(\mathcal{B}(\phi), \mathcal{B}(\psi)) & \mathcal{B}(\phi \wedge \psi) = min(\mathcal{B}(\phi), \mathcal{B}(\psi)) \end{array}$$
$$\mathcal{B}(\phi \leftrightarrow \psi) = 0 \text{ if } \mathcal{B}(\phi) \neq \mathcal{B}(\psi),\ \mathcal{B}(\phi \leftrightarrow \psi) = 1 \text{ otherwise}$$

A formula ϕ is satisfiable if there exists a Boolean interpretation $\mathcal{B}$ of ϕ such that $\mathcal{B}(\phi) = 1$, and $\mathcal{B}$ is called a *model* of ϕ in this case. We use $Mod(\phi)$ to denote the set of all the models of ϕ.

Consider for instance the formula $(p \wedge q) \leftrightarrow p$, which has exactly three models: $\mathcal{B}_1$ with $\mathcal{B}_1(p) = \mathcal{B}_1(q) = 0$; $\mathcal{B}_2$ with $\mathcal{B}_1(p) = \mathcal{B}_1(q) = 1$; and $\mathcal{B}_3$ with $\mathcal{B}_3(p) = 0$ and $\mathcal{B}_1(q) = 1$.

A propositional formula in *Conjunctive Normal Form* (CNF) is a conjunction of clauses, where a *clause* is a disjunction of literals. It is well-known that every propositional formula can be translated to CNF w.r.t. the satisfiability problem using Tseitin's linear encoding [29]. The problem of determining whether there exists a model that satisfies a given CNF formula, abbreviated as SAT, is one of the most studied NP-complete problems.

A *cardinality constraint* is an inequality of the form $\sum_{i=1}^{n} p_i \geq m$. Several polynomial encodings of this kind of constraints into propositional formulas have been proposed in the literature (e.g. [4, 26, 5]). An *AtMostOne constraint* is a particular case of the form $\sum_{i=1}^{n} p_i \leq 1$, which can be linearly encoded in SAT. For instance, the encoding using sequential counter [26, 23] is defined as follows:

$$(\neg p_1 \vee q_1) \wedge (\neg p_n \vee q_{n-1})$$
$$\bigwedge_{1<i<n} ((\neg p_i \vee q_i) \wedge (\neg q_{i-1} \vee q_i) \wedge (\neg p_i \vee \neg q_{i-1}))$$

where q_i is a fresh propositional variable for $i = 1, \ldots, n-1$.

5.2 A SAT Encoding for QIE1

In this section, we propose a SAT encoding for the problem of enumerating qualitative itemsets QIE1. More precisely, we associate to every instance of QIE1 a propositional formula so that its models allow us to obtain all the corresponding qualitative itemsets.

Let $\mathcal{D} = (\mathcal{O}, \mathcal{I}, \mathcal{C})$ be a qualitative database, f a function that maps each $a \in \mathcal{I}$ to a subset of $\mathcal{P}(B_a)$ closed under the inverse operation and the inclusion, and v a minimum support threshold. We here use the integer α defined as the value $v \cdot (|\mathcal{D}| \cdot (|\mathcal{D}| - 1)/2)$.

In order to define our encoding, we associate to each pair of an item a and a relation $r \in f(a)$ a distinct propositional variable denoted p_{a^r}. The variable p_{a^r} is used to express the qualitative itemset in the sense that it is true if and only if a^r belongs to the current qualitative itemset. Furthermore, we associate to each ordered pair of different objects (o, o') in $\mathcal{D}$ a distinct propositional variable denoted $q_{(o,o')}$. In the proposed encoding, a variable $q_{(o,o')}$ is true if and only if o precedes o' with respect to the current qualitative itemset. In order not to take into account both symmetric couples of objects in support computation, we also associate a variable denote $s_{\{o,o'\}}$ to each pair of distinct objects $\{o, o'\}$ in $\mathcal{D}$.

The first propositional formula of our encoding for QIE1 allows avoiding the empty itemset by requiring at least one item:

$$\bigvee_{a \in \mathcal{I}} \bigvee_{r \in f(a)} p_{a^r}. \tag{1}$$

Indeed, this formula corresponds to a single clause that expresses that there is at least one variable of the form p_{a^r} assigned to true.

The following conjunction of AtMostOne constraints allows avoiding the association of multiple variations to an item in the same pattern:

$$\bigwedge_{a \in \mathcal{I}} \sum_{r \in f(a)} p_{a^r} \leq 1. \tag{2}$$

More precisely, each AtMostOne constraint is associated to a distinct item and means that there is at most one qualitative relation associated to this item in the pattern.

The following formula allows establishing that each variable $q_{(o,o')}$ is true if and only o precedes o' w.r.t. the qualitative itemset:

$$\bigwedge_{o,o' \in \mathcal{O}, o \neq o'} \neg q_{(o,o')} \leftrightarrow \bigvee(\{p_{a^r} \mid a \in \mathcal{I}, r \in (f(a) \setminus \{r' \in f(a) \mid R_a(o,o') \in r'\})\}. \tag{3}$$

We exactly express in the previous formula that $q_{(o,o')}$ is false if and only if there is a qualitative item a^r such that $r(o,o')$ does not hold.

We now introduce the formula that is used for symmetry breaking by considering in the support computation at most one of the couples (o,o') and (o',o):

$$\bigwedge_{o,o' \in \mathcal{O}, o \neq o'} s_{\{o,o'\}} \leftrightarrow (q_{(o,o')} \vee q_{(o',o)}). \tag{4}$$

Finally, the following cardinality constraint expresses that support of every qualitative itemset in $\mathcal{D}$ has to be greater than or equal to v:

$$\sum_{o,o' \in \mathcal{O}, o \neq o'} s_{\{o,o'\}} \geq \alpha. \tag{5}$$

Let us note that the use of α in the previous constraint is clearly equivalent to the use of v as a minimum support threshold.

We use $\mathcal{ENC}(\mathcal{D}, f, v)$ to denote the conjunction of the previous formulas: $(1) \wedge (2) \wedge (3) \wedge (4) \wedge (5)$.

There are three important properties related to our encoding $\mathcal{ENC}(\mathcal{D}, f, v)$. First, the soundness property means that every model encodes a frequent qualitative itemset. Second, the completeness property expresses that every frequent qualitative itemset is encoded in a model of the encoding. Third, the non-redundancy property is used to capture the fact that there is a bijective mapping between the set of the models and the set of the frequent qualitative itemsets.

▶ **Proposition 9** (Soundness). *Given an instance $(\mathcal{D}, f, v)$ of QIE1, if $\mathcal{B}$ is a model of $\mathcal{ENC}(\mathcal{D}, f, v)$ then $I_{\mathcal{B}} = \{a^r \mid \mathcal{B}(p_{a^r}) = 1\} \in \mathcal{QIE}1(\mathcal{D}, f, v)$.*

Proof. First, using the formula (1), we clearly have $|I_{\mathcal{B}}| \geq 1$. Then, using (2), we know that an item occurs at most once in every pattern. Moreover, using $(3) \wedge (4)$, we obtain $\{s_{\{o,o'\}} \mid \mathcal{B}(s_{\{o,o'\}}) = 1\} = \{\{o,o'\} \subseteq \mathcal{O} \mid o \neq o', o \preceq_{I_{\mathcal{B}}} o'\}$. Thus, using the cardinality constraint (5), we obtain $|\{s_{\{o,o'\}} \mid \mathcal{B}(s_{\{o,o'\}}) = 1\}| \geq \alpha$ and we have thereby $supp_1(I_{\mathcal{B}}, \mathcal{D}) \geq v$. Therefore, $I_{\mathcal{B}}$ belongs to $\mathcal{QIE}1(\mathcal{D}, f, v)$. ◀

▶ **Proposition 10** (Completeness). *Given an instance $(\mathcal{D}, f, v)$ of QIE1, if $I \in \mathcal{QIE}1(\mathcal{D}, f, v)$ then there exists a Boolean interpretation $\mathcal{B}_I$ that satisfies the encoding $\mathcal{ENC}(\mathcal{D}, f, v)$, where $I = \{a^r \mid \mathcal{B}_I(p_{a^r}) = 1\}$.*

Proof. Let us define $\mathcal{B}_I$ as follows:

1. for every pair of an item a and a relation $r \in f(a)$, $\mathcal{B}_I(p_{a^r}) = 1$ iff $a^r \in I$;
2. for every ordered pair of distinct objects (o, o'), $\mathcal{B}_I(q_{(o,o')}) = 1$ iff $o \preceq_I o'$;
3. for every pair of distinct objects $\{o, o'\}$, $\mathcal{B}_I(s_{\{o,o'\}}) = 1$ iff $o \preceq_I o'$ or $o' \preceq_I o$.

Using the fact that $|I| \geq 1$, $\mathcal{B}_I$ satisfies (1). Then, using the fact that an item cannot occur more than once in I, $\mathcal{B}_I$ satisfies (2). Further, using the properties 1 and 2 in the definition of $\mathcal{B}_I$, we obtain that $\mathcal{B}_I$ satisfies (3). Using the fact that $\mathcal{B}_I$ satisfies (3) and the property 3 in the definition of $\mathcal{B}_I$, we also obtain that (4) is also satisfied by $\mathcal{B}_I$. Moreover, the formula (5) is satisfied since $supp_1(I, \mathcal{D}) \geq v$. ◀

▶ **Proposition 11** (Non-Redundancy). *Given an instance $(\mathcal{D}, f, v)$ of QIE1, for all two distinct models $\mathcal{B}$ and $\mathcal{B}'$ of $\mathcal{ENC}(\mathcal{D}, f, v)$, $\{a^r \mid \mathcal{B}(p_{a^r}) = 1\} \neq \{a^r \mid \mathcal{B}'(p_{a^r}) = 1\}$ holds.*

Proof. This property is a direct consequence of the fact that we use the equivalence logical connective in the formulas (3) and (4). Indeed, the support is encoded using the variables of the form $q_{(o,o')}$ and $s_{\{o,o'\}}$, and a qualitative itemset cannot have two distinct values for the support. ◀

It is worth noting that having a bijective mapping between the set of the models and the set of the frequent qualitative itemsets allows us to adapt in a fairly simple way our encoding for many variants of QIE1, such as counting the number of patterns.

Let us now introduce the notion of complementary qualitative itemset, which is mainly used for reducing the search space.

▶ **Definition 12** (Complementary Qualitative Itemset). *Let $I = \{a_1^{r_1}, \ldots, a_k^{r_k}\}$ be a qualitative itemset. The complementary of I, denoted I^c, is the qualitative itemset $\{a_1^{(r_1)^{-1}}, \ldots, a_k^{(r_k)^{-1}}\}$.*

We clearly have the following proposition.

▶ **Proposition 13.** *The following two properties are satisfied, for all qualitative database $\mathcal{D}$ and for all qualitative itemset I:*

- $supp_1(I, \mathcal{D}) = supp_1(I^c, \mathcal{D})$
- $supp_2(I, \mathcal{D}) = supp_2(I^c, \mathcal{D})$.

Proposition 13 can be used to avoid unnecessary computations. Indeed, at each found model, we can avoid in the next step both the corresponding qualitative itemset and its complementary itemset. It is worth noting that a similar property is used in the case of gradual patterns [7, 10, 11, 20].

Let us now consider the condensed representations corresponding to the closed and the maximal qualitative itemsets. In order to obtain the closed qualitative itemsets, we first need to conjunctively add to the encoding $\mathcal{ENC}(\mathcal{D}, f, v)$ the following formula:

$$\bigwedge_{a \in \mathcal{I}} \bigwedge_{r \in f(a)} \Big(\big(\bigwedge_{o, o' \in \mathcal{O}, o \neq o'} (q_{(o,o')} \rightarrow R_a(o, o') \in r) \big) \rightarrow \bigvee_{r' \subseteq r} p_{a^{r'}} \Big). \quad (6)$$

Indeed, this propositional formula means that, for all qualitative item a^r, if we have $supp_1(I, \mathcal{D}) = supp_1(I \cup \{a^r\}, \mathcal{D})$, then there exists $r' \subseteq r$ such that $a^{r'}$ belongs to I, where I is the qualitative itemset associated to the current model. In other words, it allows making the current qualitative itemset more informative without changing the support.

Then, we add the following formula to express that it is not possible to reduce the size of any relation in the pattern without changing the support:

$$\bigwedge_{a \in \mathcal{I}} \bigwedge_{r \in f(a), |r| > 1} (p_{a^r} \rightarrow \bigwedge_{r' \subset r} (\bigvee_{o, o' \in \mathcal{O}, o \neq o'} q_{(o,o')} \wedge R_a(o, o') \notin r')). \tag{7}$$

We use $\mathcal{ENC} - \mathcal{C}(\mathcal{D}, f, v)$ to denote the SAT encoding for the problem of enumerating the closed qualitative itemsets: $\mathcal{ENC}(\mathcal{D}, f, v) \wedge (6) \wedge (7)$.

Similarly, to compute the maximal qualitative itemsets, we only need to conjunctively add to $\mathcal{ENC}(\mathcal{D}, f, v)$ the following two formulas:

$$\bigwedge_{a \in \mathcal{I}} \bigwedge_{r \in f(a)} (\sum_{o, o' \in \mathcal{O}, o \neq o'} (q_{(o,o')} \wedge R_a(o, o') \in r) \geq \alpha \rightarrow \bigvee_{r' \subseteq r} p_{a^{r'}}) \tag{8}$$

$$\bigwedge_{a \in \mathcal{I}} \bigwedge_{r \in f(a), |r| > 1} (p_{a^r} \rightarrow \bigwedge_{r \subset r} \sum_{o, o' \in \mathcal{O}, o \neq o'} (q_{(o,o')} \wedge R_a(o, o') \notin r') < \alpha). \tag{9}$$

The formula (8) allows maximizing the size of the current qualitative itemset while keeping the support greater than or equal to v, (9) states that it is not possible to reduce the size of any relation without reducing the support to a value smaller than v. We use $\mathcal{ENC} - \mathcal{M}(\mathcal{D}, f, v)$ to denote the SAT encoding $\mathcal{ENC}(\mathcal{D}, f, v) \wedge (8) \wedge (9)$.

5.3 A SAT Encoding for QIE2

We here propose a SAT encoding for the problem QIE2, which combines formulas defined for QIE1 and new ones that are described in this section.

Let $\mathcal{D} = (\mathcal{O}, \mathcal{I}, \mathcal{C})$ be a database, f a function that maps each $a \in \mathcal{I}$ to a subset of $\mathcal{P}(B_a)$ closed under the inverse operation and the inclusion, and v a minimum support threshold. We here use the integer β defined as the value $v \cdot |\mathcal{D}|$. We now describe an encoding that allows one to obtain all the elements of $\mathcal{QIE}2(\mathcal{D}, v)$.

In the same way as the previous encoding, we also use in the same way the propositional variables of the forms p_{a^r} and $q_{(o,o')}$: the variables of the form p_{a^r} are used to encode the qualitative itemset, and those of the form $q_{(o,o')}$ to encode its support. Moreover, we associate to each integer $i \in 1..\beta$ and object o in $\mathcal{D}$ a fresh propositional variable t^i_o, which is used to express that the object o is used at the location i in a sequence in $\mathcal{L}(\mathcal{D}, I)$, where I is the current qualitative itemset.

The first formula in our encoding is the conjunction of $(1) \wedge (2) \wedge (3)$ of the previous encoding $\mathcal{ENC}(\mathcal{D}, f, v)$. Indeed, (1) is used to express that every qualitative itemset contains at least one qualitative item, (2) is used to avoid multiple occurrences of an item in the same itemset, and (3) says that $q_{(o,o')}$ is false if and only if there is a qualitative item a^r such that $R_a(o, o') \in r$ does not hold. As a consequence, every model of the previous conjunction encodes a qualitative itemset, where the variables of the form $q_{(o,o')}$ encode the pairs of objects that satisfy the partial order induced by this itemset.

Using the fact that the propositional variables of the form t^i_o are used to build an ordered sequence of objects, the following formula means that an object cannot be used more than once in a sequence:

$$\bigwedge_{o\in\mathcal{O}} \sum_{i=1}^{\beta} t^i_o \leq 1. \tag{10}$$

The following formula says that there is exactly one object at each location:

$$\bigwedge_{i=1}^{\beta} \sum_{o\in\mathcal{O}} t^i_o = 1. \tag{11}$$

Clearly, the previous formula allows us to only consider the qualitative itemsets that have supports greater than or equal to v w.r.t. $supp_2$.

In order to require the ordering induced by the qualitative itemset, the following formula is used to capture the fact that if two objects o and o' occur in successive locations, then the couple (o, o') respects the qualitative itemset, which is expressed by the truth of the variable $q_{(o,o')}$:

$$\bigwedge_{o,o'\in\mathcal{O},o\neq o'} \bigwedge_{i=1}^{\beta-1} ((t^i_o \wedge t^{i+1}_{o'}) \rightarrow q_{(o,o')}). \tag{12}$$

We use $\mathcal{ENC}2(\mathcal{D}, f, v)$ to denote the encoding that corresponds to the following conjunction: (1) ∧ (2) ∧ (3) ∧ (10) ∧ (11) ∧ (12).

▶ **Proposition 14** (Soundness). *Given an instance $(\mathcal{D}, f, v)$ of QIE2, if $\mathcal{B}$ is a model of $\mathcal{ENC}2(\mathcal{D}, f, v)$ then $I_\mathcal{B} = \{a^r \mid \mathcal{B}(p_{a^r}) = 1\} \in \mathcal{QIE}2(\mathcal{D}, f, v)$.*

Proof. The soundness can be shown in the same way as in the case of QIE1. Using (1), we know that $I_\mathcal{B}$ contains at least one qualitative item. Then, using (2), each item occurs at most once in every qualitative itemset. Further, using (3), we obtain $a^r \in I_\mathcal{B}$ iff, for all $o, o' \in \mathcal{O}$, $\mathcal{B}(q_{(o,o')}) = 1$ iff $R_a(o, o') \in r$. Thus, using (10) ∧ (11) ∧ (12), we know that there exists a sequence $\langle o_1, \ldots, o_\beta\rangle$ which respects $I_\mathcal{B}$, where $\mathcal{B}(t^i_{o_i}) = 1$ for $i \in 1..\beta$. As a consequence, $supp_2(I_\mathcal{B}, \mathcal{D}) \geq v$ and $I_\mathcal{B}$ belongs to $\mathcal{QIE}2(\mathcal{D}, f, v)$ ◀

▶ **Proposition 15** (Completeness). *Given an instance $(\mathcal{D}, f, v)$ of QIE2, if $I \in \mathcal{QIE}2(\mathcal{D}, f, v)$ then there exists a Boolean interpretation $\mathcal{B}_I$ that satisfies the encoding $\mathcal{ENC}2(\mathcal{D}, f, v)$ where $I = \{a^r \mid \mathcal{B}_I(p_{a^r}) = 1\}$.*

Proof. First, given a sequence $s = \langle o_1, \ldots, o_\beta\rangle$ respecting I, we define $\mathcal{B}_I$ as follows:

- for every pair of an item a and a relation $r \in f(a)$, $\mathcal{B}_I(p_{a^r}) = 1$ iff $a^r \in I$;
- for every couple of distinct objects (o, o'), $\mathcal{B}_I(q_{(o,o')}) = 1$ iff $o \preceq_I o'$;
- for every object o and location $i \in 1..\beta$, $\mathcal{B}_I(t^i_o) = 1$ iff $o = o_i$.

For the same reasons described in the proof of Proposition 10, $\mathcal{B}_I$ satisfies (1) ∧ (2) ∧ (3). Then, using the fact that the length of s is β and the objects in this sequence are pairwise distinct, $\mathcal{B}_I$ satisfies also (10) ∧ (11). Finally, using the fact that s respects the partial order induced by I, $\mathcal{B}_I$ satisfies (12). ◀

Contrary to our previous encoding, $\mathcal{ENC}2(\mathcal{D}, f, v)$ does not satisfy the non-redundancy property, since the same qualitative itemset may be associated to distinct sequences. However, this is not a problem for enumerating the qualitative itemsets without redundancy, because we only need to conjunctively add the negation of the found qualitative itemset at each step instead of the negation of the found model. More precisely, if we found a model representing the qualitative itemset $I = \{a_1^{r_1}, \ldots, a_k^{r_k}\}$, then we conjunctively add the clause $\neg p_{a_1{}^{r_1}} \vee \cdots \vee \neg p_{a_k{}^{r_k}} \vee \bigvee_{a^r \notin I} p_a^r$ to avoid this itemset in the next steps.

In $\mathcal{ENC}2(\mathcal{D}, f, v)$, we use propositional variables that are associated to only β locations, since we aim at computing the qualitative itemsets having supports at least equal to v. However, for computing the closed qualitative itemsets, we need to have the exact value of the support, which means that we have to encode one of the longest sequences in each model of the SAT encoding. In order to avoid this problem, we propose an intermediate solution by restricting $\mathcal{ENC}2(\mathcal{D}, f, v)$ to the closed qualitative itemsets w.r.t. QIE1. In this context, we clearly have the following property.

▶ **Proposition 16.** *Let $\mathcal{D}$ be a qualitative database and I a qualitative itemset. If I is closed in $\mathcal{D}$ w.r.t. $supp_2$, then it is also closed in $\mathcal{D}$ w.r.t. $supp_1$.*

Proof. This property is a direct consequence of the fact that if $supp_2(I, \mathcal{D}) > supp_2(J, \mathcal{D})$, then $supp_1(I, \mathcal{D}) > supp_1(J, \mathcal{D})$ holds for every qualitative itemsets I and J with $I \sqsubset J$. ◀

Thus, the set of closed qualitative itemsets w.r.t. QIE2 is included in that of the qualitative itemsets obtained from the encoding $\mathcal{ENC}2(\mathcal{D}, f, v) \wedge (6) \wedge (7)$. As a consequence, the previous SAT encoding can be used for enumerating all the closed qualitative itemsets w.r.t. QIE2. Indeed, we only need in this context to select the largest patterns w.r.t. $\sqsubseteq$ for every value for the support.

Let us now consider the problem of enumerating the maximal qualitative itemsets. In this context, consider the following formulas:

$$\bigwedge_{a \in \mathcal{I}} \bigwedge_{r \in f(a)} \bigwedge_{o,o' \in \mathcal{O}, o \neq o'} \Big(\bigwedge_{i=1}^{\beta-1} ((t_o^i \wedge t_{o'}^{i+1} \wedge R_a(o,o') \in r) \rightarrow \bigvee_{r' \subseteq r} p_{a^{r'}})\Big), \tag{13}$$

$$\bigwedge_{a \in \mathcal{I}} \bigwedge_{r \in f(a), |r|>1} (p_{a^r} \rightarrow \bigwedge_{r \subset r} (\bigvee_{o,o' \in \mathcal{O}, o \neq o'} \bigwedge_{i=1}^{\beta-1} (t_o^i \wedge t_{o'}^{i+1} \wedge R_a(o,o') \notin r'))). \tag{14}$$

These two formulas express that, for a sequence of length equal to β, the associated qualitative itemset has to be the largest w.r.t. $\sqsubseteq$. Therefore, in the same way as our encoding for enumerating the closed qualitative itemsets, the encoding $\mathcal{ENC}2(\mathcal{D}, f, v) \wedge (13) \wedge (14)$ allows one to compute a set of patterns that contains all the maximal qualitative itemsets.

It is worth noting that the strategies proposed in [15, 18] for adapting Conflict-Driven Clause-Learning (CDCL) based SAT-solvers to the task of model enumeration can be directly used in the case of our encoding. Furthermore, it is also possible to directly use the decomposition method introduced in [17] for improving the SAT-based approach in solving data mining problems.

6 Conclusion and Perspectives

The first main contribution of this article is a definition of a framework for data mining using qualitative reasoning. This framework allows considering different data types, such as numerical values, time intervals and spatial regions. Moreover, the data mining tasks introduced in this work can be seen as a natural generalization of those related to gradual itemsets. The second main contribution is our declarative and flexible solution for solving the proposed data mining tasks based on the satisfiability problem in classical propositional logic (SAT): each task is modeled as a propositional formula whose models correspond to the desired patterns.

In our future work, we intend to further study qualitative reasoning in data mining following three main directions: (1) the use of disjunctions of base relations between objects, which allows, for instance, modeling vagueness; (2) considering qualitative formalisms that are not closed under the inverse operation, such as cardinal direction calculus [27, 28]; (3) considering some qualitative relations with arities greater than two in the case of some particular data types (e.g. [14, 6]). Furthermore, we plan to implement the proposed SAT-based methods to provide an experimental study on the use of our framework.

References

1 Rakesh Agrawal, Tomasz Imielinski, and Arun N. Swami. Mining Association Rules between Sets of Items in Large Databases. In *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, Washington, DC, USA*, pages 207–216. ACM Press, 1993.

2 James F. Allen. An Interval-Based Representation of Temporal Knowledge. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence, IJCAI '81, Vancouver, BC, Canada*, pages 221–226. William Kaufmann, 1981.

3 James F. Allen. Maintaining Knowledge about Temporal Intervals. *Communications of the ACM*, 26(11):832–843, 1983.

4 Olivier Bailleux and Yacine Boufkhad. Efficient CNF Encoding of Boolean Cardinality Constraints. In *Principles and Practice of Constraint Programming - CP 2003, 9th International Conference, CP 2003, Kinsale, Ireland*, pages 108–122, 2003.

5 Olivier Bailleux, Yacine Boufkhad, and Olivier Roussel. A Translation of Pseudo Boolean Constraints to SAT. *JSAT*, 2(1-4):191–200, 2006.

6 Philippe Balbiani, Jean-François Condotta, and Gérard Ligozat. Reasoning about Cyclic Space: Axiomatic and Computational Aspects. In *Spatial Cognition III, Routes and Navigation, Human Memory and Learning, Spatial Representation and Spatial Learning*, pages 348–371. Springer, 2003.

7 Fernando Berzal, Juan C. Cubero, Daniel Sánchez, María Amparo Vila Miranda, and José-María Serrano. An Alternative Approach to Discover Gradual Dependencies. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 15(5):559–570, 2007.

8 Abdelhamid Boudane, Saïd Jabbour, Lakhdar Sais, and Yakoub Salhi. A SAT-Based Approach for Mining Association Rules. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI 2016, New York, NY, USA*, pages 2472–2478. IJCAI/AAAI Press, 2016.

9 Abdelhamid Boudane, Saïd Jabbour, Lakhdar Sais, and Yakoub Salhi. SAT-Based Data Mining. *International Journal on Artificial Intelligence Tools*, 27(1):1–24, 2018.

10 Lisa Di-Jorio, Anne Laurent, and Maguelonne Teisseire. Fast extraction of gradual association rules: a heuristic based method. In *CSTST 2008: Proceedings of the 5th International Conference on Soft Computing as Transdisciplinary Science and Technology, Cergy-Pontoise, France*, pages 205–210. ACM, 2008.

11 Lisa Di-Jorio, Anne Laurent, and Maguelonne Teisseire. Mining Frequent Gradual Itemsets from Large Databases. In *Advances in Intelligent Data Analysis VIII, 8th International Symposium on Intelligent Data Analysis, IDA 2009, Lyon, France*, pages 297–308. Springer, 2009.

12 Tias Guns, Anton Dries, Siegfried Nijssen, Guido Tack, and Luc De Raedt. MiningZinc: A declarative framework for constraint-based mining. *Artificial Intelligence*, 244:6–29, 2017.

13 Tias Guns, Siegfried Nijssen, and Luc De Raedt. Itemset mining: A constraint programming perspective. *Artificial Intelligence*, 175(12-13):1951–1983, 2011.

14 Amar Isli and Anthony G. Cohn. A new approach to cyclic ordering of 2D orientations using ternary relation algebras. *Artificial Intelligence*, 122(1-2):137–187, 2000.
15 Saïd Jabbour, Jerry Lonlac, Lakhdar Sais, and Yakoub Salhi. Extending modern SAT solvers for models enumeration. In *Proceedings of the 15th IEEE International Conference on Information Reuse and Integration, IRI 2014, Redwood City, CA, USA*, pages 803–810. IEEE Computer Society, 2014.
16 Saïd Jabbour, Lakhdar Sais, and Yakoub Salhi. Boolean satisfiability for sequence mining. In *22nd ACM International Conference on Information and Knowledge Management, CIKM'13, San Francisco, CA, USA*, pages 649–658. ACM, 2013.
17 Saïd Jabbour, Lakhdar Sais, and Yakoub Salhi. Decomposition Based SAT Encodings for Itemset Mining Problems. In *Advances in Knowledge Discovery and Data Mining - 19th Pacific-Asia Conference, PAKDD 2015, Ho Chi Minh City, Vietnam*, pages 662–674. Springer, 2015.
18 Saïd Jabbour, Lakhdar Sais, and Yakoub Salhi. On SAT Models Enumeration in Itemset Mining. *CoRR*, abs/1506.02561, 2015. `arXiv:1506.02561`.
19 Saïd Jabbour, Lakhdar Sais, and Yakoub Salhi. Mining Top-k motifs with a SAT-based framework. *Artificial Intelligence*, 244:30–47, 2017.
20 Anne Laurent, Marie-Jeanne Lesot, and Maria Rifqi. GRAANK: Exploiting Rank Correlations for Extracting Gradual Itemsets. In *Flexible Query Answering Systems, 8th International Conference, FQAS 2009, Roskilde, Denmark*, pages 382–393. Springer, 2009.
21 Gérard Ligozat. *Qualitative Spatial and Temporal Reasoning*. ISTE. Wiley, 2013.
22 Jerry Lonlac, Saïd Jabbour, Engelbert Mephu Nguifo, Lakhdar Saïs, and Badran Raddaoui. Extracting Frequent Gradual Patterns Using Constraints Modeling. *CoRR*, abs/1903.08452, 2019. `arXiv:1903.08452`.
23 João P. Marques-Silva and Inês Lynce. Towards Robust CNF Encodings of Cardinality Constraints. In *Principles and Practice of Constraint Programming - CP 2007, 13th International Conference, CP 2007, Providence, RI, USA*, pages 483–497, 2007.
24 Luc De Raedt, Tias Guns, and Siegfried Nijssen. Constraint Programming for Itemset Mining. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Las Vegas, Nevada, USA*, pages 204–212, 2008.
25 David A. Randell, Zhan Cui, and Anthony Cohn. A Spatial Logic Based on Regions and Connection. In *Proceedings of the 3rd International Conference on Principles of Knowledge Representation and Reasoning (KR'92). Cambridge, MA, USA*, 1992.
26 Carsten Sinz. Towards an Optimal CNF Encoding of Boolean Cardinality Constraints. In *Principles and Practice of Constraint Programming - CP 2005, 11th International Conference, CP 2005, Sitges, Spain*, pages 827–831, 2005.
27 Spiros Skiadopoulos and Manolis Koubarakis. Composing cardinal direction relations. *Artificial Intelligence*, 152(2):143–171, 2004.
28 Spiros Skiadopoulos and Manolis Koubarakis. On the consistency of cardinal direction constraints. *Artificial Intelligence*, 163(1):91–135, 2005.
29 Gregory S. Tseitin. On the complexity of derivations in the propositional calculus. In *Structures in Constructives Mathematics and Mathematical Logic, Part II*, pages 115–125, 1968.
30 Willy Ugarte, Patrice Boizumault, Bruno Crémilleux, Alban Lepailleur, Samir Loudni, Marc Plantevit, Chedy Raïssi, and Arnaud Soulet. Skypattern mining: From pattern condensed representations to dynamic constraint satisfaction problems. *Artificial Intelligence*, 244:48–69, 2017.
31 Marc B. Vilain and Henry A. Kautz. Constraint Propagation Algorithms for Temporal Reasoning. In *Proceedings of the 5th National Conference on Artificial Intelligence. Philadelphia, PA, USA. Volume 1: Science*, pages 377–382. Morgan Kaufmann, 1986.