Knowledge Discovery from Qualitative Spatial and Temporal Data

Abderrahmane Boukontar, Jean-François Condotta, Yakoub Salhi

Knowledge Discovery from Qualitative Spatial and Temporal Data

Abderrahmane Boukontar
CRIL-CNRS UMR 8188
Université d'Artois
Lens, France
boukontar@cril.fr

Jean-François Condotta
CRIL-CNRS UMR 8188
Université d'Artois
Lens, France
condotta@cril.fr

Yakoub Salhi
CRIL-CNRS UMR 8188
Université d'Artois
Lens, France
salhi@cril.fr

***Abstract*—Qualitative reasoning formalisms facilitate the representation and interpretation of information involving complex entities. We use in this paper qualitative spatial and temporal reasoning to introduce novel data mining tasks, which consist in extracting knowledge from quantitative databases that are transformed into collections of qualitative relation networks (QRNs). After describing our qualitative data mining framework, we first propose an Apriori-like algorithm that exploits monotonicity and QRN consistency for pruning the search space: the validity of a pattern candidate depends on the supports of the larger patterns that include it and on its consistency. We then introduce an encoding of our data mining tasks into the well-known problem of frequent itemset mining. We finally show the feasibility of our approach by providing preliminary experimental results using real-world datasets about the movements of football players during matches.**

***Index Terms*—qualitative reasoning, knowledge discovery, data mining**

I. Introduction

By abstracting and simplifying aspects of common-sense and domain-specific knowledge, qualitative reasoning can be the basis for a variety of systems that allow producing reasoning and explanations understandable by humans. The literature witnesses a broad spectrum of qualitative formalisms for reasoning about time and space. In particular, Point Algebra (PA) [17] and Interval Algebra (IA) [3] are the most prevalent formalisms in artificial intelligence for reasoning about time. Regarding qualitative formalisms for reasoning about space, we can mention Cardinal Direction Algebra (CDA) [5] and the variants of Region Connection Calculus (RCC) [10]. All the previous formalisms use binary relations for encoding relations between temporal and spatial entities. Data mining aims at discovering patterns that represent pieces of knowledge from, generally large, datasets (e.g. see [1]). Dealing directly with data from the physical world with continuous aspects, such as time and space, can lead to complex patterns, making difficult the task of interpretation. To remedy this problem, we use in this work qualitative temporal and spatial formalisms for simplifying data as well as patterns.

It is worth noting that a similar approach was proposed in [12]. This approach uses qualitative reasoning for abstracting data having different types in the same dataset: every dataset is transformed into a set of qualitative relation networks (QRNs) where each one encodes the relations between the data rows w.r.t. a data attribute, i.e., every QRN concerns a single entity. However, in our approach, every QRN is used to encode a single data row, which means that a dataset can be seen as a set of configurations about the same set of entities. Thus, our framework introduces new data mining tasks different from those proposed in [12].

In this work, we first introduce the notion of qualitative database (q-database) that can be seen as a collection of QRNs. Then, a qualitative pattern (q-pattern) is defined as a QRN, and its quality in our data mining tasks is determined through the notion of support, which corresponds to the number of QRNs in a q-database that are included in the q-pattern. The inclusion here refers to the inclusion between the qualitative relations occurring in the QRNs. In this context, a q-pattern is said to be frequent if its support is greater than or equal to a given threshold. After defining the condensed representation of frequent q-patterns and describing some interesting properties, we show that algorithms used for frequent itemsets mining can be adapted to our mining tasks. We focus in this work on Apriori algorithm which is the first and most influential algorithm for generating frequent itemsets [2]. Our Apriori-like algorithm mainly results from the validity of the monotonicity property: if a QRN is not frequent then all its sub-QRNs are not frequent. Moreover, we introduce an encoding of our data mining tasks into the problem of frequent itemset mining. We finally provide a preliminary experimental evaluation that shows the feasibility of our approach. We consider in our experiments real-world datasets about the movements of football players during matches.

II. Preliminaries

A. Frequent Itemset Mining

Let Ω be a finite non empty set of symbols, called *items*. From now on, we assume that this set is fixed. An *itemset* I over Ω is defined as a non empty subset of Ω. A *transaction* is an ordered pair (i, I) where I is an itemset and i a positive integer, called *transaction identifier*. A *transaction database* $\mathcal{T}$ is defined as a finite non empty set of transactions where each transaction identifier refers to a unique itemset. The *cover* of

an itemset I in a transaction database $\mathcal{T}$, denoted $\mathsf{Cov}(I, \mathcal{T})$, is defined as follows:

$$\mathsf{Cov}(I, \mathcal{T}) = \{i \in \mathbb{N} \mid (i, J) \in \mathcal{T} \text{ and } I \subseteq J\}.$$

The *support* of I in $\mathcal{T}$, denoted $\mathsf{Supp}(I, \mathcal{T})$, corresponds to the cardinality of the cover of I in $\mathcal{T}$, that is, $\mathsf{Supp}(I, \mathcal{T}) = |\mathsf{Cov}(I, \mathcal{T})|$.

Tid	Itemset
1	S_1, S_3, D_1
2	S_2, D_2
3	S_1, D_1
4	S_1, S_2, D_2
5	S_1, D_1

TABLE I
A TRANSACTION DATABASE $\mathcal{T}$.

Given a transaction database $\mathcal{T}$ and a minimum support threshold σ, the *frequent itemset mining problem* (FIM problem), considered as the root of pattern mining field in the literature [1], consists in finding the itemsets occurring in $\mathcal{T}$ with supports greater than or equal to σ. Formally, it consists in computing the following set: $\{I \subseteq \Omega \mid \mathsf{Supp}(I, \mathcal{T}) \geq \sigma\}$.

In this work, we are also interested in the two following well-known condensed representations of frequent itemsets. An itemset I is said to be *closed* in a database $\mathcal{T}$ if $\mathsf{Supp}(I \cup \{a\}, \mathcal{T}) < \mathsf{Supp}(I, \mathcal{T})$ for every a in $\Omega \setminus I$. It is said to be *maximal* w.r.t. a minimum support threshold σ if $\mathsf{Supp}(I, \mathcal{T}) \geq \sigma$ and $\mathsf{Supp}(I \cup \{a\}, \mathcal{T}) < \sigma$ for every a in $\Omega \setminus I$.

For instance, consider the transaction database $\mathcal{T}$ described in Table I, where the items $S_{i=1,2,3}$ designate symptoms, and $D_{i=1,2}$ designate diseases. We have $\mathsf{Cov}(\{S_1, D_1\}, \mathcal{T}) = \{1, 3, 5\}$ and $\mathsf{Supp}(\{S_1, D_1\}, \mathcal{T}) = 3$. Note that $\{S_1, D_1\}$ is closed and maximal w.r.t. the threshold $\sigma = 3$.

B. Qualitative Spatial and Temporal Formalisms

Each binary qualitative formalism [8] considers a set of elements $\mathcal{D}$ allowing to represent spatial or temporal entities. It is based on a finite set of non-empty binary relations $\mathcal{B}$ defined on $\mathcal{D}$ called *base relations*. Each base relation of $\mathcal{B}$ represents a particular temporal or spatial qualitative configuration between two entities. The set $\mathcal{B}$ forms a partition of $\mathcal{D} \times \mathcal{D}$. Moreover, it is assumed that the identity relation on $\mathcal{D}$, denoted by $\mathsf{id}_\mathcal{B}$ and defined by $\mathsf{id}_\mathcal{B} = \{(x, x) \mid x \in \mathcal{D}\}$, belongs to $\mathcal{B}$. Additionally, the converse of each base relation $b \in \mathcal{B}$, denoted by b^{-1}, is also one element of $\mathcal{B}$, i.e., there exists $b' \in \mathcal{B}$ such that $b' = b^{-1} = \{(y, x) \mid (x, y) \in b\}$ for all $b \in \mathcal{B}$. The set $\mathcal{B}$ is also equipped with the *weak composition* operation also called *algebraic composition*, denoted by $\diamond$ and defined by $b \diamond b' = \{b'' \in \mathcal{B} \mid \exists x, y, z \in \mathcal{D}$ with $(x, y) \in b, (y, z) \in b'$ and $(x, z) \in b''\}$. Hence, the weak composition of two base relations $b \diamond b'$ is the set of base relations corresponding to the possible qualitative configurations between two entities x and z while there is a third entity y such that the qualitative configuration between the entities x and y (respectively the entities y and z) corresponds to the base relation b (respectively b').

As illustration, consider the temporal qualitative formalisms called Point Algebra (PA) [17] and Interval Algebra (IA) [3]. PA represents the temporal entities by the points of the line whereas IA uses intervals of the line. Hence, the domain $\mathcal{D}$ can be defined by $\mathcal{D} = \mathbb{Q}$ for PA and by $\mathcal{D} = \{(x, y) \in \mathbb{Q} \times \mathbb{Q} \mid x < y\}$ for IA. The base relations of PA (resp. IA) are depicted in Figure 2 (resp. Figure 1). In Figure 3 are given the tables of converse and weak composition of PA.

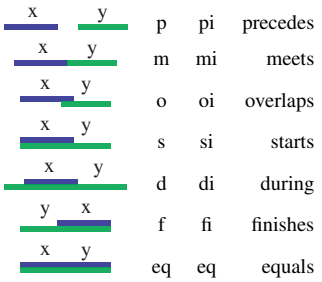


Fig. 1. The thirteen base relations of IA.

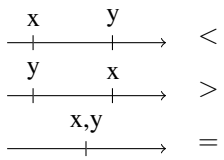


Fig. 2. The three base relations of PA.

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

b \ b'	$<$	$=$	$>$
$<$	$\{<\}$	$\{<\}$	$\{<, =, >\}$
$=$	$\{<\}$	$\{=\}$	$\{>\}$
$>$	$\{<, =, >\}$	$\{>\}$	$\{>\}$

Fig. 3. The table of converse and the table of weak composition of PA.

Concerning spatial qualitative reasoning, the Cardinal Direction Algebra (CDA) [7] considers points to represent spatial entities and is based on a set of nine base relations $\mathcal{B} = \{\mathsf{N}, \mathsf{NW}, \mathsf{W}, \mathsf{SW}, \mathsf{S}, \mathsf{SE}, \mathsf{E}, \mathsf{NE}, \mathsf{EQ}\}$ corresponding to compass relations between two entities in the plane (see Figure 5). The family of the Region Connection Calculi (RCC) formalisms are also widely used for spatial qualitative reasoning. These formalisms are based on binary topological relations between the regions of a topological space. As illustration, in Figure 4 are depicted the eight base relations of the RCC8 [10] calculus.

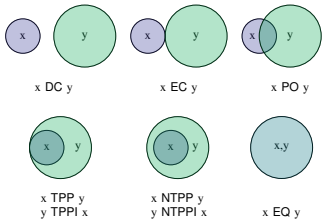


Fig. 4. The eight base relations of RCC8.

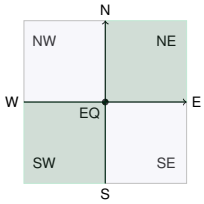


Fig. 5. Qualitative spatial directions in CDA.

Definite knowledge between any two entities can be represented by one base relation of $\mathcal{B}$ whereas indefinite knowledge can be defined by the union of the possible base relations

between the two entities. Such a union of base relations is called *qualitative relation* or just *relation* and is represented by the set of its included base relations. Hence, the set $2^{\mathcal{B}}$ represents the relations of the considered qualitative formalism based on $\mathcal{B}$. Among the relations of $2^{\mathcal{B}}$, the particular relation $\mathcal{B}$ is called the *universal relation* whereas the particular relation $\{\}$ is called the empty relation. The universal relation allows to represent the absence of any information between two entities. This relation is always satisfied between two entities whereas the empty relation is never satisfied. The set of relations $2^{\mathcal{B}}$ is equipped with the usual set-theoretic operations (intersection and union). Moreover, the weak composition is extended to $2^{\mathcal{B}}$ in the following way: for all $r, r' \in 2^{\mathcal{B}}$, $r \diamond r' = \bigcup_{b \in r, b' \in r'} \{b \diamond b'\}$ [11]. Note that $r \diamond r'$ corresponds to the strongest relation of $2^{\mathcal{B}}$ containing the elements of the usual relational composition $r \circ r' = \{(x, z) \in \mathcal{D} \times \mathcal{D} \mid \exists y \in \mathcal{D}$ with $(x, y) \in r$ and $(y, z) \in r'\}$.

In what follows, we generically consider a qualitative formalism defined from a set of base relations $\mathcal{B}$.

III. Qualitative Databases and Mining Tasks

We first describe the notion of qualitative relation network which underlies the notions of qualitative database and pattern.

Definition 1 (QRN[a])**.** *A* Qualitative Relation Network (QRN) *is an ordered pair $\mathcal{N} = (E, R)$ where E is a set of (spatial or temporal) entities and R a function that associates a relation in $2^{\mathcal{B}}$ to each element of $E \times E$ s.t. $R(e, e) = \{id_{\mathcal{B}}\}$ and $R(e, e') = (R(e', e))^{-1}$.*

We use $\mathsf{QRN}(E)$ to denote the set of all QRNs over E and $\mathsf{E}(\mathcal{N})$ to denote the set of entities occurring in $\mathcal{N}$. The QRN defined on the set of entities E such as $R'(e, e') = \mathcal{B}$ for every $e, e' \in E$ is denoted by $\top^E$. $\mathcal{N}_{[e,e']/r}$ with r in $2^{\mathcal{B}}$ is the QRN $\mathcal{N}' = (E, R')$ defined by $R'(e, e') = r$ and $R'|_{(E\times E)\setminus\{(e,e')\}} = R|_{(E\times E)\setminus\{(e,e')\}}$. A QRN $\mathcal{N}_1 = (E, R_1)$ is a sub-QRN of $\mathcal{N}_2 = (E, R_2)$ if for every $e, e' \in E$ we have $R_1(e, e') \subseteq R_2(e, e')$ and we note $\mathcal{N}_1 \subseteq \mathcal{N}_2$.

A QRN $\mathcal{N} = (E, R)$ is said to be *consistent* if there exists a solution s, i.e. a function from E to the domain $\mathcal{D}$ such that, for every $(e, e') \in E \times E$, $(s(e), s(e')) \in b$ for some $b \in R(e, e')$.

In Figure 6, a consistent QRN $\mathcal{N}_1$ of PA and one of its solutions are depicted. Note that for every QRN, an entity is represented by a node, and a relation by an arc labeled with it. For the sake of simplicity, $id_{\mathcal{B}}$ loops ($R(e, e)$), converse relations and universal relations are omitted.

Definition 2 (Qualitative Database)**.** *A* qualitative database (q-database) *is a structure of the form $\langle E, C\rangle$ where E is a finite set of entities and C a finite set of ordered pairs $(i, \mathcal{S})$, called q-row, s.t. i is a positive integer, called* q-row identifier*, $\mathcal{S}$ a consistent QRN s.t. $\mathsf{E}(\mathcal{S}) = E$. Each identifier refers to a unique q-row, that is, for every two distinct q-rows $(i, \mathcal{S}^i)$ and $(j, \mathcal{S}^j)$, $i \neq j$.*

[a]A similar definition in the literature is that of qualitative constraint networks (QCNs) [14], [15]. However, we do not consider constraints on variables but relations between entities.

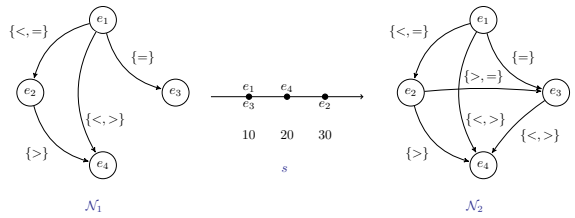


Fig. 6. A consistent QRN $\mathcal{N}_1$, a solution s of $\mathcal{N}_1$, and a minimally labelled QRN $\mathcal{N}_2$.

Alternatively stated, a q-database can be seen as a finite multiset of consistent QRNs. In the sequel, we fix $QB = \langle E, \{(0, \mathcal{S}^0), \ldots, (m-1, \mathcal{S}^{m-1})\}\rangle$ with $E = \{e_0, \ldots, e_{n-1}\}$.

Consider the set E of the temporal entities $\{e_1, e_2, e_3, e_4\}$ and the PA formalism. Figure 7 is an illustration of a qualitative database QB as a multi-set of 4 QRNs. We emphasize that the multiple presence of the same QRN is possible according to the definition ($\mathcal{S}^1 = \mathcal{S}^3$) and that, for more simplicity, a node e is omitted iff for every $e' \in E$ such that $R(e, e') = \mathcal{B}$.

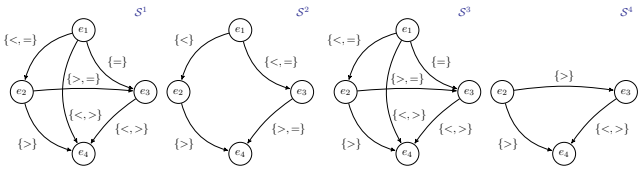


Fig. 7. Example of a qualitative database.

A *qualitative pattern (q-pattern)* is simply defined as a QRN. We say that a q-row $(i, \mathcal{S}^i)$ supports a q-pattern $\mathcal{N}$ if $\mathcal{S}^i \subseteq \mathcal{N}$. The *cover* of a q-pattern $\mathcal{N}$ in a q-database QB, denoted $\mathsf{Cov}(\mathcal{N}, QB)$, is the set of the identifiers of the q-rows in QB that support $\mathcal{N}$. That is, $\mathsf{Cov}(\mathcal{N}, QB) = \{i \in \mathbb{N} \mid \mathcal{S}^i \subseteq \mathcal{N}\}$. In our framework, the quality of a q-pattern in a q-database is determined through the notion of support defined as the size of its cover, i.e., $\mathsf{Supp}(\mathcal{N}, QB) = |\mathsf{Cov}(\mathcal{N}, QB)|$.

Definition 3 (FQP)**.** *Given a minimum support threshold $\sigma > 0$, the problem of generating* Frequent Q-Patterns (FQP) *consists in computing the set of q-patterns that have supports greater than or equal to σ, i.e., $\mathsf{FQP}(QB) = \{\mathcal{N} \in \mathsf{QRN}(E) \mid \mathsf{Supp}(\mathcal{N}, QB) \geq \sigma\}$.*

Proposition 4 (Monotonicity)**.** *If $\mathcal{N} = (E, R)$ is a frequent q-pattern in QB, then, for every $(e, e') \in E \times E$ with $e \neq e'$ and every base relation $b \in \mathcal{B}$, $\mathcal{N}_{[e,e']/R(e,e')\cup\{b\}}$ is also a frequent q-pattern in QB.*

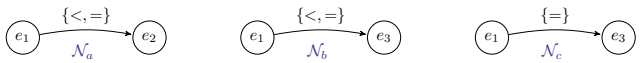
Proof. Let $\alpha = (i, \mathcal{S}^i)$ be a q-row in QB that supports $\mathcal{N}$, $(e, e') \in E \times E$ s.t. $e \neq e'$ and $b \in \mathcal{B}$. Then $\mathcal{S}^i \subseteq \mathcal{N}$ holds. Owing to $\mathcal{N} \subseteq \mathcal{N}_{[e,e']/R(e,e')\cup\{b\}}$, we obtain $\mathcal{S}^i \subseteq \mathcal{N}_{[e,e']/R(e,e')\cup\{b\}}$. Therefore, we have $\mathsf{Supp}(\mathcal{N}_{[e,e']/R(e,e')\cup\{b\}}, QB) \geq \mathsf{Supp}(\mathcal{N}, QB)$. □

The monotonicity property leads to the following condensed representations.

Definition 5 (Closed Q-Pattern)**.** *A q-pattern $\mathcal{N} = (E, R)$ is said to be* closed *in QB if, for every $(e, e') \in E \times E$ with $e \neq e'$ and every $b \in R(e, e')$, $\mathsf{Supp}(\mathcal{N}_{[e,e']/R(e,e')\setminus\{b\}}, QB) < \mathsf{Supp}(\mathcal{N}, QB)$.*

Definition 6 (Maximal Q-Pattern)**.** *A q-pattern $\mathcal{N} = (E, R)$ is said to be* maximal *in QB w.r.t. a given minimum support threshold σ if $\mathsf{Supp}(\mathcal{N}, QB) \geq \sigma$ (i.e., it is frequent) and, for every $(e, e') \in E \times E$ with $e \neq e'$ and every $b \in R(e, e')$, $\mathsf{Supp}(\mathcal{N}_{[e,e']/R(e,e')\setminus\{b\}}, QB) < \sigma$.*

One can easily see that every maximal q-pattern is closed. To illustrate the two definitions above, consider the database of Figure 7 and a minimum support threshold equals to 2. The QRN $\mathcal{N}_a$ below illustrate a maximal q-pattern while the QRN $\mathcal{N}_b$ illustrate a closed but not maximal q-pattern since $\mathcal{N}_{b[e_1,e_3]/\{<,=\}\setminus\{<\}} = \mathcal{N}_c$ is frequent with a $\mathsf{Supp}(\mathcal{N}_c, QB) = 2$.



Definition 7 (Minimally Labelled Q-Pattern)**.** *A QRN $\mathcal{N} = (E, R)$ is said to be* minimally labelled *if, for every $(e, e') \in E \times E$ with $e \neq e'$ and every $b \in R(e, e')$, $\mathcal{N}_{[e,e']/\{b\}}$ is consistent.*

As illustration, a minimally labelled QRN $\mathcal{N}_2$ of PA, and a consistent but not minimally labelled QRN $\mathcal{N}_1$ are depicted in Figure 6. For example, one can see that there is no solution meeting the choice of $R(e_2, e_3) = \{<\}$.

Proposition 8. *Let QB be q-database. If all the QRNs occurring in QB are minimally labelled, then the following properties are satisfied:*

1) *every closed q-pattern in QB is minimally labelled, and*
2) *every maximal q-pattern in QB is minimally labelled.*

Proof. We only need to show Property 1, the other being a direct consequence of Property 1 and the fact that every maximal q-pattern is closed.

Let $\mathcal{N} = (E, R)$ be a closed q-pattern in QB. For the sake of contradiction, suppose that $\mathcal{N}$ is not minimally labelled. Then, there exist $(e, e') \in E \times E$ with $e \neq e'$ and $b \in R(e, e')$ s.t. $\mathcal{N}_{[e,e']/\{b\}}$ is inconsistent. Let $\alpha = (i, \mathcal{S}^i)$ be a q-row in QB that supports $\mathcal{N}$. Then we have $\mathcal{S}^i \subseteq \mathcal{N}$. Thus, due to the fact that $\mathcal{S}^i$ is minimally labelled, we have $b \notin \mathcal{S}^i[e, e']$. Thus, $\mathsf{Supp}(\mathcal{N}_{[e,e']/R(e,e')\setminus\{b\}}, QB) = \mathsf{Supp}(\mathcal{N}, QB)$ holds and, since $\mathcal{N}$ is closed, we obtain a contradiction. □

The *minimal labelling problem* (MLP) consists in determining all the feasible base relations in a QRN [9]. In other words, this consists in determining the greatest minimally labelled QRN included in a given QRN. It might be interesting to replace every QRN occurring in a q-database with their corresponding minimally labelled QRN before generating frequent q-patterns. However, MLP is NP-hard in the case of several qualitative reasoning formalisms, such as IA and RCC8 [9].

One of the best trade-offs can be the use of the tractable method called *path-consistency* method or still $\diamond$-*consistency* method, consisting in the calculation of the closure by weak composition of the considered QRN and that by removing some unfeasible base relations through the operation of weak composition to obtain an equivalent $\diamond$-consistent sub-QRN [13]. A QRN $N = (E, R)$ is said $\diamond$-consistent or closed under weak composition if for every $e, e', e'' \in E$ we have $R(e, e') \subseteq (R(e, e'') \diamond R(e'', e'))$. The path-consistency method iterates the triangulation operation $R(e, e') \leftarrow (R(e, e'') \diamond R(e'', e'))$ for every $e, e', e'' \in E$ until a fix point is reached. This method can be realized in $\mathcal{O}(|E|^3)$. It is sound but not complete for the consistency problem of a QRN. Indeed, in general, it removes some but not all unfeasible base relations of the QRN. In the case where the empty relation is obtained, we can affirm that the initial QRN is inconsistent.

Similarly to Proposition 8, we have the following proposition.

Proposition 9. *Let QB be q-database. If all the QRNs occurring in QB are $\diamond$-consistent, then the following properties are satisfied:*

1) *every closed q-pattern in QB is $\diamond$-consistent and*
2) *every maximal q-pattern in QB is $\diamond$-consistent.*

Thus, whenever the database contains only $\diamond$-consistent QRNs, the $\diamond$-consistent frequent q-patterns can be seen as a condensed representation, which is weaker than closed frequent q-patterns.

IV. Apriori-like Algorithm

This section aims to show that algorithms used for frequent itemsets mining can be adapted to our mining tasks. We focus here on Apriori algorithm which is the first and most influential algorithm for generating frequent itemsets [2]. It proceeds by a level-wise search: it first computes the frequent itemsets of size one; then, assuming the frequent itemsets of size n known, it computes a set of candidates of size $n + 1$ and compute their supports, so that I is a valid candidate if and only if its subsets of size n are frequent (a consequence of monotonicity property); this procedure is iterated until no more candidate is found.

To define our Apriori-like algorithm, we represent a q-pattern by the set of base relation occurrences removed from the QRN whose relations are all universal, called *qr-set*; this allows us to have a compact representation of q-patterns and emphasize the informative part. The removed base relation occurrences that correspond to a q-pattern are written as a set of expressions of the form (e, b, e'), where b is a base relation, and e, e' are occurring entities belonging to E.

The expression $\top^E \setminus C$ with C a qr-set of the form $\{(e_0, b_0, e'_0), \ldots, (e_k, b_k, e'_k)\}$ with $e_0, \ldots, e_k, e'_0, \ldots, e'_k \in E$ and $b_0, \ldots, b_k \in B$, represents the QRN $\mathcal{N} = (E, R)$ defined by $R(e, e') = \mathcal{B} \setminus \{b \mid (e, b, e') \in C \text{ or } (e', b^{-1}, e) \in C\}$

for every $e, e' \in E$ and $e \neq e'$. The pairs of entities $(e_0, e'_0), \ldots, (e_k, e'_k)$ are not necessarily distinct.

For instance, consider the QRN $\mathcal{N}_1$ from Figure 6. This QRN is represented with the qr-set $\{(e_1, >, e_2), (e_1, <, e_3), (e_1, >, e_3), (e_1, =, e_4), (e_2, <, e_4), (e_2, =, e_4)\}$. It is worth mentioning that to check if a q-row supports a q-pattern, we only need to show that none of the relations occurring in the QRN associated with the q-row is empty after removing the elements in the set associated with the q-pattern.

In our algorithm 1, we consider that there is an order on the set of entities E and a lexicographical order $\prec$ on the qr-set.

The algorithm 1 starts initially by generating the frequent q-patterns associated with the qr-sets of size 1, i.e. the set of $\{(e, b, e')\}$ for every $e, e' \in E$ with $e < e'$ and b a base relation of $\mathcal{B}$, and that by calculating the support of the qr-sets corresponding to them. Having thus the set $Freq_1$, we reiterate the procedure from line 2 to line 11 of the algorithm 1 until no more candidates can be found. We generate the set of candidates using GENERATE_CAND procedure, calculate their supports by a simple scan of the database, and finally keep those with minimum support of σ. The candidates' generation procedure remains the most important step of the algorithm 1 since it is at this stage that the search space is pruned using the two properties, monotonicity, and consistency. Having a set $Freq_i$ of frequent q-patterns of size i, GENERATE_CAND considers all possible ordered pairs of two qr-sets P, P' of $Freq_i$ which differ only by one base relation and reiterate the following. A candidate C is formed by joining P and P', however, the consistency of this one in addition to the monotonicity property is still to be verified. We emphasize here that the CONSISTENCY procedure noted in the algorithm 2, can correspond to a complete or incomplete approach; however, given the complexity of the consistency problem which is NP-complete for a large number of formalisms in general [8], it is more feasible to consider one of the incomplete approaches as the $\diamond$-consistency method. The monotonicity property is verified by assuring that all the sub-sets of size $i-1$ of the qr-set C belong to $Freq_{i-1}$. Finally, only candidates verifying the two conditions are retained and their supports are calculated. Candidates with a minimum support threshold σ only make the set $Freq_i$. Once the set of candidates is empty, we return the qr-sets associated with all frequent q-patterns and thus obtain all the frequent q-patterns.

Algorithm 1: Apriori-Like Procedure

Data: A q-database QB and a minimum support threshold σ
Result: The qr-sets associated with all frequent q-patterns

1 $Freq_1 \leftarrow \{\{(e, b, e')\} \mid e, e' \in E, e < e', b \in \mathcal{B}$ and $\mathsf{Supp}(\top^E \setminus \{(e, b, e')\}, QB) \geq \sigma\}$
2 **for** *($i \leftarrow 2$; $Freq_{i-1} \neq \emptyset$; $i++$)* **do**
3 $Cand_i \leftarrow$ GENERATE_CAND$(Freq_{i-1})$
4 **for** *($j \leftarrow 0$ to $m-1$)* **do**
5 **for** *($C \in Cand_i$)* **do**
6 **if** *$\mathcal{S}^j$ supports ($\top^E \setminus C$)* **then**
7 $C.supp++$
8 **end**
9 **end**
10 **end**
11 $Freq_i \leftarrow \{C \in Cand_i \mid C.supp \geq \sigma\}$
12 **end**
13 **return** $\bigcup_i Freq_i$

Algorithm 2: Procedure GENERATE_CAND

Data: A set $Freq_{i-1}$ of frequent q-patterns of size $i-1$
Result: A set of qr-sets of size i representing candidates

1 $Cand_i \leftarrow \emptyset$
2 **for** *($P, P' \in Freq_{i-1}$ with $|P \setminus P'| = |P' \setminus P| = 1$ and $P \prec P'$)* **do**
3 $C \leftarrow P \cup P'$
4 **if** *(*CONSISTENCY$(\top^E \setminus C)$ *and*
5 *$\forall S \subset C$ with $|S| = i-1$, $S \in Freq_{i-1}$)* **then**
6 $Cand_i \leftarrow Cand_i \cup \{C\}$
7 **end**
8 **end**
9 **return** $Cand_i$

V. A FIM-BASED ENCODING

In this section, we propose an encoding of the problem of generating frequent q-patterns into that of generating frequent itemsets. To this end, we use again the qr-sets associated with the QRNs.

We first associate a distinct item a^b_{ij} with every expression $e_i, e_j \in E$ with $i < j$ and $b \in \mathcal{B}$. Then, for every QRN $\mathcal{N}$, we associate the itemset $T(\mathcal{N}) := \{a^b_{ij} \mid 0 \leq i < j \leq n-1,$ and $b \in \mathcal{B} \setminus R(e_i, e_j)\}$. Finally, the transaction database $\mathcal{T}$ associated with QB, denoted $T(QB)$, corresponds to the set $\{(0, T(\mathcal{S}^0)), \ldots, (m-1, T(\mathcal{S}^{m-1}))\}$.

Given an itemset I, we use $QRN(I)$ to denote the QRN $\top^E \setminus \{(e_i, b, e_j) \mid a^b_{ij} \in I\}$.

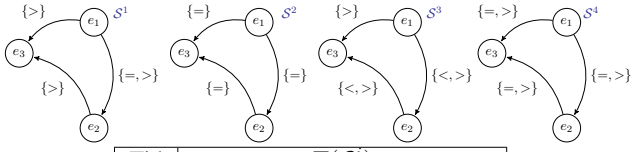
The following property is a direct consequence of the definitions of $T(\cdot)$ and $QRN(\cdot)$.

Proposition 10. *Let $\mathcal{N}$ be a QRN. Then, $I = T(\mathcal{N})$ iff $QRN(I) = \mathcal{N}$.*

Consider the q-database QB_2 in Figure 8 with four QRNs of three entities e_1, e_2 and e_3 on PA. The table illustrate the associated transaction database $T(QB_2)$ of QB_2.

Theorem 11. *The following properties are satisfied:*

1) *I is a frequent itemset in $T(QB)$ w.r.t. a given threshold iff $QRN(I)$ is a frequent q-pattern in QB w.r.t. the same threshold;*
2) *I is a closed itemset in $T(QB)$ iff $QRN(I)$ is a closed q-pattern in QB;*



Tid	$T(\mathcal{S}^i)$
1	$a^{<}_{12},\ a^{<}_{13},\ a^{=}_{13},\ a^{<}_{23},\ a^{=}_{23}$
2	$a^{<}_{12},\ a^{<}_{13},\ a^{<}_{23},\ a^{>}_{12},\ a^{>}_{13},\ a^{>}_{23}$
3	$a^{<}_{13},\ a^{=}_{13},\ a^{=}_{23},\ a^{=}_{12}$
4	$a^{<}_{12},\ a^{<}_{13},\ a^{<}_{23}$

Fig. 8. Example of a transaction database associated with the q-database QB_2.

3) *I is a maximal itemset in $T(QB)$ w.r.t. a given threshold iff $QRN(I)$ is a maximal q-pattern in QB w.r.t. the same threshold.*

Proof. We only consider Property 1, the others being similar. *Part* $\Rightarrow$. Let I be a frequent itemset in $T(QB)$. Using Proposition 10, we obtain for every transaction $t_i = (i, T(\mathcal{S}^i))$ in $T(QB)$, if $I \subseteq T(\mathcal{S}^i)$ (t_i supports I) then $QRN(T(\mathcal{S}^i)) = \mathcal{S}^i \subseteq QRN(I)$ ($\mathcal{S}^i$ supports $QRN(I)$). Thus, the support of I in $T(QB)$ is equal to that of $QRN(I)$ in QB. Consequently, $QRN(I)$ is a frequent q-pattern in QB w.r.t. the same minimum support threshold.
Part $\Leftarrow$. Let I be an itemset such that $QRN(I)$ is a frequent q-pattern in QB. We have for every QRN $\mathcal{S}^i$ occurring in QB, if $\mathcal{S}^i \subseteq QRN(I)$ then $T(QRN(I)) \subseteq T(\mathcal{S}^i)$. Using again Proposition 10, $T(QRN(I)) = I$ holds. Due to the definition of $T(QB)$, the support of I in $T(QB)$ is equal to that of $QRN(I)$ in QB. Therefore, I is a frequent itemset in $T(QB)$ w.r.t. the same threshold. □

For instance, consider the q-database of Figure 8, a minimum support threshold equal to 3 and the q-pattern represented by the QRN $\mathcal{N}$ of Figure 9.

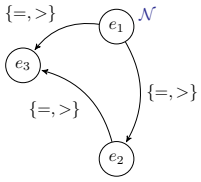


Fig. 9. Example of a maximal QRN with $\mathsf{Supp}(\mathcal{N}, QB_2) = 3$.

$\mathcal{N}$ is associated to the itemset $I = \{a^{<}_{12},\ a^{<}_{13},\ a^{<}_{23}\}$. I is maximal with a $\mathsf{Supp}(I, T(QB_2)) = 3$. One can easily check using Proposition 6 that $\mathcal{N}$ is maximal.

VI. Computational Experiments

In this section, we examine the practical feasibility of our approach by running experiments on real-world datasets of soccer games, which are available on Metrica-sports[b]. All experiments have been conducted on a node with Intel XEON E5-2643 Quad-Core 3,3 GHz, 64 Go of RAM, of a cluster of 34 nodes.

The datasets represent quantitative data from soccer games, describing the movement of the players and the ball by their x and y coordinates on the field over time. For the considered game the initial number of records is 141159 (25 records each second). After a necessary cleaning of these data (missing values, outliers) and conversion into qualitative data, we end up with 67891 q-rows. To get these q-rows, we have chosen to represent the relations between the players positions using three formalisms: PA formalism by choosing one axis of movement (the x axis), CDA formalism and finally an abstraction of CDA to an algebra of five base relations called CD5 and that by regrouping the relations N and NE into a single relation NEE and similarly E and SE into SEE, S and SW into SSW and finally W and NW into NWW. We used the FIM-based encoding described in Section V to translate the q-database into a transaction database.

In the first instance, we consider all binary relations between the 4 players (2,3,21, and 22) only. In a second instance, the ball (11) is taken into consideration, hence all the relations between the four players and the ball are considered as well. Only 1 record per 20 has been taken into account in the q-database ending up with 3394 q-rows. We made these choices, taking into account the size of the data on the one hand (the number of q-rows and the number of resulting relations which are of the order of $m \times n \times \frac{n-1}{2}$, n being the number of players and m the number of base relations of the formalism), and on the other hand considering that this application proves only the feasibility of the problem. Figure 10 represents a record of the game at a specific point in time and the 3 associated QRNs according to each formalism.

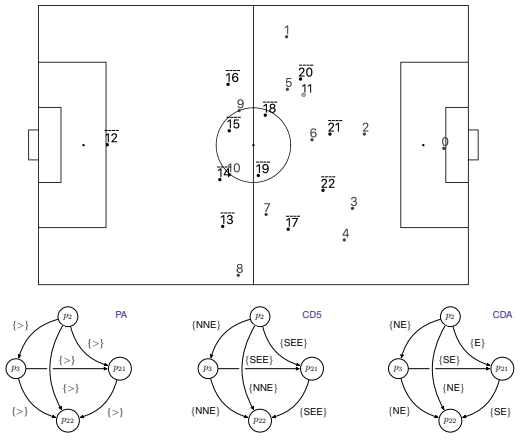


Fig. 10. The positioning of the players in a game at a given moment, where numbers from 0 to 10 are the players of team 1 and 12 to 22 those of team 2 and finally 11 correspond to the ball. On the left are the associated QRNs according to the three qualitative formalisms PA, CD5, CDA.

We resolved here the problems of mining frequent QRNs, closed and maximal frequent QRNs. A timeout of 4 hours was set and we used the Linear time Closed (*LCM*) itemset Miner

[b]https://github.com/metrica-sports/sample-data/tree/master/data/Sample_Game_2

[16] to resolve the 3 problems. LCM is the winner of the FIMI 2004 competition [4].

Table II (respectively III and IV) describes the number of frequent (#F), maximal (#M), and closed (#C) QRNs discovered in the q-database considering the four players (2,3,21 and 22) and according to the PA formalism (respectively CD5, and CDA) w.r.t a minimum support threshold σ and the time (in seconds) it took to obtain them.

$\sigma(\%)$	#F (time)	#C (time)	#M (time)
0.95	64 (0.0)	64 (0.0)	1 (0.0)
0.90	64 (0.0)	64 (0.0)	1 (0.0)
0.85	256 (0.0)	224 (0.0)	3 (0.0)
0.80	384 (0.0)	320 (0.0)	3 (0.0)
0.75	670 (0.0)	447 (0.0)	9 (0.0)
0.70	960 (0.0)	560 (0.0)	4 (0.0)
0.65	1024 (0.0)	576 (0.0)	1 (0.0)
0.60	1152 (0.0)	640 (0.0)	2 (0.0)
0.55	1408 (0.0)	736 (0.0)	3 (0.0)
0.50	1916 (0.0)	945 (0.0)	13 (0.0)
0.45	2368 (0.0)	1088 (0.0)	3 (0.0)
0.40	2816 (0.0)	1200 (0.0)	3 (0.0)
0.35	4032 (0.0)	1384 (0.0)	6 (0.0)
0.30	4864 (0.0)	1584 (0.0)	5 (0.0)
0.25	5768 (0.0)	1697 (0.0)	8 (0.0)
0.20	7168 (0.0)	1732 (0.0)	3 (0.0)
0.15	7832 (0.0)	1839 (0.0)	10 (0.0)
0.10	10360 (0.0)	2155 (0.0)	14 (0.0)
0.05	14848 (0.0)	2432 (0.0)	17 (0.0)

TABLE II
EXPERIMENTS WITH THE FOUR PLAYERS 2,3,21 AND 22 USING THE PA FORMALISM.

$\sigma(\%)$	#F (time)	#C (time)	#M (time)
0.95	672 (0.00)	84 (0.00)	7 (0.00)
0.90	3480 (0.00)	435 (0.00)	21 (0.00)
0.85	10352 (0.00)	1294 (0.00)	40 (0.00)
0.80	26000 (0.00)	3250 (0.00)	77 (0.00)
0.75	49136 (0.00)	6142 (0.01)	108 (0.01)
0.70	79936 (0.01)	9984 (0.01)	110(0.01)
0.65	127584 (0.01)	15652 (0.02)	232 (0.02)
0.60	243824 (0.02)	28214 (0.03)	460 (0.03)
0.55	465744 (0.03)	49645 (0.04)	525 (0.04)
0.50	788592 (0.04)	76015 (0.06)	655 (0.07)
0.45	1261616 (0.05)	107758 (0.10)	944 (0.10)
0.40	1997240 (0.08)	152657 (0.14)	1230 (0.15)
0.35	3394888 (0.12)	218528 (0.2)	1518 (0.21)
0.30	5494672 (0.17)	297772 (0.28)	1696 (0.30)
0.25	8966832 (0.25)	397826 (0.38)	2270 (0.41)
0.20	14955208 (0.34)	508391 (0.50)	3348 (0.54)
0.15	26719936 (0.50)	673692 (0.69)	3556 (0.75)
0.10	49382984 (0.72)	848675 (0.90)	4268 (0.99)
0.05	99401072 (1.04)	1066489 (1.22)	5277 (1.32)

TABLE III
EXPERIMENTS WITH THE FOUR PLAYERS 2,3,21 AND 22 USING THE CD5 FORMALISM.

Table V, VI, and VII describe the same results discovered in a q-database with the four players (2,3,21 and 22) and where the ball (11) is taken into consideration as an other entity.

A first observation that is immediately obvious, is that the number of the q-patterns discovered depends on the granularity of the formalism. Indeed, the more the formalism is large in terms of base relations the more the number of discovered q-patterns will be important. This said, the choice of a formalism

$\sigma(\%)$	#F (time)	#C (time)	#M (time)
0.95	$\approx 1.0\times10^{10}$ (0.05)	38307 (0.05)	425 (0.05)
0.90	$\approx 5.5\times10^{10}$ (0.22)	208606 (0.25)	1896 (0.24)
0.85	$\approx 1.7\times10^{11}$ (0.57)	590563 (0.7)	2766 (0.66)
0.80	$\approx 4.2\times10^{11}$ (1.26)	1296021 (1.55)	4339 (1.48)
0.75	$\approx 8.0\times10^{11}$ (2.05)	2091131 (2.69)	6594 (2.54)
0.70	$\approx 1.3\times10^{12}$ (2.96)	3004016 (3.94)	5267 (3.76)
0.65	$\approx 2.1\times10^{12}$ (4.11)	4225057 (5.51)	10338 (5.25)
0.60	$\approx 4.0\times10^{12}$ (6.36)	6475899 (8.58)	17055 (8.24)
0.55	$\approx 7.7\times10^{12}$ (9.44)	9410051 (12.55)	17983 (12.28)
0.50	$\approx 1.3\times10^{13}$ (12.26)	11869287 (16.05)	17276 (15.36)
0.45	$\approx 2.1\times10^{13}$ (15.83)	15057971 (20.44)	24064 (19.78)
0.40	$\approx 3.3\times10^{13}$ (20.25)	18814955 (25.99)	26191 (29.25)
0.35	$\approx 5.7\times10^{13}$ (25.79)	23125139 (32.52)	26149 (31.14)
0.30	$\approx 9.2\times10^{13}$ (30.56)	26793663 (37.78)	21491 (36.13)
0.25	$\approx 1.5\times10^{14}$ (35.79)	30240298 (42.66)	23082 (47.13)
0.20	$\approx 2.5\times10^{14}$ (41.05)	33522666 (48.30)	25250 (45.59)
0.15	$\approx 4.5\times10^{14}$ (46.39)	36406732 (51.79)	17574 (50.37)
0.10	$\approx 8.3\times10^{14}$ (52.06)	38460444 (54.95)	13852 (52.73)
0.05	$\approx 1.7\times10^{15}$ (55.16)	39927173 (58.01)	11197 (55.10)

TABLE IV
EXPERIMENTS WITH THE FOUR PLAYERS 2,3,21 AND 22 USING THE CDA FORMALISM.

$\sigma(\%)$	#F (time)	#C (time)	#M (time)
0.95	1024 (0.0)	1024 (0.0)	1 (0.0)
0.90	1024 (0.0)	1024 (0.0)	1 (0.0)
0.85	4096 (0.0)	3584 (0.0)	3 (0.0)
0.80	9216 (0.0)	6400 (0.0)	4 (0.0)
0.75	18856 (0.01)	10651 (0.01)	55 (0.01)
0.70	43172 (0.01)	17448 (0.01)	78 (0.01)
0.65	70708 (0.01)	23193 (0.02)	68 (0.02)
0.60	125176 (0.01)	29767 (0.02)	24 (0.02)
0.55	200640 (0.02)	34174 (0.02)	27 (0.03)
0.50	285503 (0.02)	39549 (0.03)	126 (0.03)
0.45	333424 (0.02)	45890 (0.03)	39 (0.04)
0.40	442600 (0.03)	53948 (0.04)	66 (0.04)
0.35	613256 (0.03)	59063 (0.04)	64 (0.05)
0.30	820474 (0.04)	65074 (0.05)	48 (0.05)
0.25	1063264 (0.04)	71044 (0.06)	35 (0.06)
0.20	1408848 (0.04)	74951 (0.06)	43 (0.06)
0.15	1832384 (0.04)	78756 (0.06)	59 (0.07)
0.10	2374848 (0.05)	84007 (0.07)	95 (0.07)
0.05	4310656 (0.05)	92133 (0.08)	118 (0.08)

TABLE V
EXPERIMENTS WITH THE FOUR PLAYERS 2,3,21,22 AND THE BALL (11) USING THE PA FORMALISM.

with a few base relations can result in the lack of expressiveness of the information and therefore of the q-patterns. One solution would be to find a formalism, sufficiently expressive but also allowing for ease of computation. The CD5 here for instance, represents such a trade-off between the PA formalism that remains clearly unable to express correctly the spatial information of the movements of the players, since they move in two directions and CDA which shows, as we can see in table VII, a difficulty for the computation if the number of players exceeds 4.

The comparison between the results in Tables II and V (III and VI), and (IV and VII) shows that the number of entities can also lead to an exponential increase in the computation run time. It should be noted that adding a single player to the database results in the addition of n new relations (all the binary relations between the player to be added and the

$\sigma(\%)$	#F (time)	#C (time)	#M (time)
0.95	11488 (0.0)	358 (0.0)	17 (0.0)
0.90	85248 (0.01)	2542 (0.01)	54 (0.01)
0.85	399072 (0.01)	110572 (0.02)	174 (0.01)
0.80	1741408 (0.03)	340349 (0.04)	810 (0.05)
0.75	6826624 (0.07)	127592 (0.13)	2262 (0.13)
0.70	22866367 (0.19)	336896 (0.35)	5224 (0.36)
0.65	65402816 (0.42)	742674 (0.76)	8558 (0.79)
0.60	$\approx 1.5\times10^{8}$ (0.79)	1365562 (1.47)	10962 (1.51)
0.55	$\approx 3.1\times10^{8}$ (1.35)	2284773 (2.40)	16406 (2.57)
0.50	$\approx 5.9\times10^{8}$ (2.43)	4085406 (4.46)	34907 (4.76)
0.45	$\approx 1.2\times10^{9}$ (4.66)	7703023 (8.54)	61940 (9.12)
0.40	$\approx 2.7\times10^{9}$ (8.89)	14116606 (16.04)	99241 (17.32)
0.35	$\approx 6.0\times10^{9}$ (16.68)	25552637 (29.88)	171580 (32.44)
0.30	$\approx 1.4\times10^{10}$ (32.04)	47264954 (56.21)	319418 (61.64)
0.25	$\approx 3.3\times10^{10}$ (63.62)	88064501 (107.05)	557722 (117.92)
0.20	$\approx 9.1\times10^{10}$ (126.09)	$\approx 1.6\times10^{8}$ (196.49)	884483 (221.48)
0.15	$\approx 2.7\times10^{11}$ (250.69)	$\approx 2.7\times10^{8}$ (351.35)	1417224 (389.8)
0.10	$\approx 9.0\times10^{11}$ (509.63)	$\approx 4.5\times10^{8}$ (616.66)	2197078 (688.64)
0.05	$\approx 4.1\times10^{12}$ (1154.05)	$\approx 7.3\times10^{8}$ (1130.35)	3159685 (1239.55)

TABLE VI
EXPERIMENTS WITH THE FOUR PLAYERS 2,3,21,22 AND THE BALL (11) USING THE CD5 FORMALISM.

$\sigma(\%)$	#F (time)	#C (time)	#M (time)
0.95	$\approx 9.8\times10^{15}$ (77.22)	$\approx 5.6\times10^{7}$ (90.94)	1124157 (85.18)
0.90	$\approx 8.1\times10^{16}$ (460.55)	$\approx 3.5\times10^{8}$ (564.37)	3703082 (524.55)
0.85	$\approx 3.9\times10^{17}$ (1715.52)	$\approx 1.3\times10^{9}$ (2101.44)	10058343 (1988.89)
0.80	$\approx 1.7\times10^{18}$ (4922.69)	$\approx 3.6\times10^{9}$ (6196.11)	18111949 (5859.21)
0.75	$\approx 6.9\times10^{18}$ (10699.64)	$\approx 7.7\times10^{9}$ (13458.67)	29467108 (12802.89)
≤ 0.70	-	-	-

TABLE VII
EXPERIMENTS WITH THE FOUR PLAYERS 2,3,21,22 AND THE BALL (11) USING THE CDA FORMALISM.

already existing n players in the database).

VII. CONCLUSION AND PERSPECTIVES

We have introduced a data mining framework that exploits qualitative reasoning formalisms for knowledge representation. In this context, we have proposed an adaptation of Apriori algorithm to our framework; this algorithm is the first and most influential algorithm for solving the problem of frequent itemset mining. We have also proposed an encoding of the considered data mining tasks into the problem of frequent itemset mining. Additionally, we have provided an experimental evaluation of our encoding by considering real-world datasets.

One of the main perspectives of this work is the generation of association rules from frequent qualitative patterns, similarly to the use of frequent itemsets in the computation of association rules [1]. Future work will also be concerned with the study of adaptations of other data mining algorithms to our framework, such as FP-Growth [6] and LCM [16].

REFERENCES

[1] R. Agrawal, T. Imieliński, and A. Swami, "Mining association rules between sets of items in large databases," in Proceedings of the ACM SIGMOD international conference on Management of data, 1993, pp. 207–216.

[2] R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules in Large Databases," in Proceedings of the 20th International Conference on Very Large Data Bases, 1994, pp. 487–499.

[3] J. F. Allen, "Maintaining Knowledge about Temporal Intervals," Communication of ACM, 1983, pp. 832-843.

[4] R. Bayardo, B. Goethals, and M. Javeed Zaki, "FIMI '04," in Proceedings of the IEEE ICDM Workshop on Frequent Itemset Mining Implementations, 2004.

[5] A. U. Frank, "Qualitative spatial reasoning with cardinal directions," Seventh Austrian Conference on Artificial Intelligence, 1991, pp. 157–167.

[6] J. Han, J. Pei, and Y. Yin, "Mining frequent patterns without candidate generation," ACM sigmod record, 2000, pp. 1–12.

[7] G. Ligozat, "Reasoning about cardinal directions," Journal of Visual Languages & Computing, 1998, pp. 23–44.

[8] G. Ligozat, Qualitative spatial and temporal reasoning, John Wiley & Sons, 2013.

[9] W. Liu and S. Li, "Solving minimal constraint networks in qualitative spatial and temporal reasoning," in Proceedings of the 18th International Conference on Principles and Practice of Constraint Programming, 2012, pp. 464–479.

[10] D. A. Randell, Z. Cui, and A. G. Cohn, "A spatial logic based on regions and connection," KR, 1992.

[11] J. Renz and G. Ligozat, "Weak composition for qualitative spatial and temporal reasoning," in 11th International Conference on Constraint Programming, 2005, pp. 534–548.

[12] Y. Salhi, "Qualitative reasoning and data mining," in 26th International Symposium on Temporal Representation and Reasoning, 2019.

[13] M. Sioutis and J. F. Condotta, "Efficiently enforcing path consistency on qualitative constraint networks by use of abstraction," in IJCAI, 2017, pp. 1262–1268.

[14] M. Sioutis, Y. Salhi, and J. F. Condotta, "On the use and effect of graph decomposition in qualitative spatial and temporal reasoning," in Proceedings of the 30th Annual ACM Symposium on Applied Computing, 2015, pp. 1874–1879.

[15] M. Sioutis, Y. Salhi, and J. F. Condotta, "A simple decomposition scheme for large real world qualitative constraint networks," in 28th International Flairs Conference, 2015.

[16] T. Uno, T. Asai, Y. Uchida, and H. Arimura, "LCM: An efficient algorithm for enumerating frequent closed item sets," in FIMI, 2003.

[17] M. Vilain, H. Kautz, and P. van Beek, "Constraint propagation algorithms for temporal reasoning: a revised report," in Readings in Qualitative Reasoning about Physical Systems, 1989, pp. 373–381.