Nearly Periodic Facts in Temporal Relational Databases

Paolo Terenziani

Abstract—Despite the huge amount of work devoted to the treatment of time within the relational context, few relevant temporal phenomena still remain to be addressed. One of them is the treatment of "*nearly periodic events*", i.e., events/facts that occur in intervals of time which repeat periodically (e.g., a meeting occurring *twice each Monday*, possibly *not at regular times*). Nearly periodic events are quite frequent in everyday life, and thus in many applicative contexts. Their treatment within the relational model is quite challenging, since it involves the *integrated* treatment of three aspects: (i) the *number* of repetitions, (ii) their *periodicity*, and (iii) *temporal indeterminacy*. Coping with this problem requires an *in-depth extension* of current temporal relational database techniques. In this paper, we introduce a new data model, and new definitions of relational algebraic operators coping with the above issues. We ascertain the properties of the new model and algebra, with emphasis on the *expressiveness* of our representation model, on the *reducibility* property, and on the *correctness* of the algebraic operators.

Index Terms—Temporal databases, database design, modeling and management

1 Introduction

Time is pervasive of reality. Many database approaches cope with it. Since the 80's, the scientific community has highlighted that time has a special status with respect to the other data, and dedicated techniques have to be devised to deal with it within the *relational database* context [1]. Since the 80', the scientific community has proposed hundreds of approaches to cope with time in *temporal relational databases* (TDB in the following; see, e.g., [2], [3]). Many TDB approaches focus on *individual occurrences* of facts, whose time of occurrence (*valid time* [4]) is exactly known. However, in many real world applications, one has to cope with facts/events that *repeat* several times. Tuzhilin and Clifford [5] distinguished among (1) "strongly periodic" events, occurring at equally distant intervals of time (e.g., Mondays, weeks); (2) "nearly periodic" events, occurring at regular intervals of time, but not necessarily at equally distant intervals (e.g., a person going to the cinema once each week –and thus at regular intervals–, but not necessarily in the same day –thus, not at equally distant intervals); (3) "intermittent" events, which occur repeatedly in time, but without any regularity (e.g., a man visiting a pub "periodically", meaning that the visits can be quite irregular). While several approaches considered the first class above, and the recent approach in [6] focused on the third class, the second class has been neglected by the TDB literature yet. Nearly periodic facts occur in many different domains. In such a sense, they can be thought as a domain-independent phenomenon that takes place whenever the periodicity does not constitute the *valid time* of the fact, but just characterizes periodic intervals of time *containing* the valid times of the instances of the facts. Additionally, we also consider the possibility that the facts occur more than once in each periodic interval (as in Example 1, considering the melphalan treatment), and that the number of repetitions may be approximated by a minimum and a maximum value (as in Example 2).

• *The author is with DISIT, Università del Piemonte Orientale "Amedeo Avogadro", Alessandria 15121, Italy. E-mail: paolo.terenziani@unipmn.it.*

Manuscript received 10 Jan. 2016; revised 23 May 2016; accepted 19 June 2016. Date of publication 0 . 0000; date of current version 0 . 0000.
Recommended for acceptance by K. Yi.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TKDE.2016.2585483

Example 1. The therapy for multiple mieloma is made by six cycles of 5-day treatment, each one followed by a delay of 23 days (for a total time of 24 weeks). Within each cycle of 5 days, 2 inner cycles can be distinguished: the melphalan treatment, to be provided twice a day, for each of the 5 days, and the prednisone treatment, to be provided once a day, for each of the 5 days.

Example 2. Each working day from 1/2/2015 to 30/6/2015, Ann attended between 2 and 4 classes of the Computer Science Program.

In many cases, as in the above examples, the *exact* time of facts is not known, and can only be *approximated*, so that *temporal indeterminacy* [14] has to be faced. Additionally, many facts and human activities are *periodically repeated* in time, and repetitions are expressed in a temporally indeterminate way. For instance, cases like Examples 1 and 2 above arise in many tasks (e.g., scheduling, planning, office automation) and domains, ranging from the description of medical treatments (see Example 1) to the recording of human activities (see Example 2), from manufacturing (e.g., *100 machines are produced each working hour*) to auditing (e.g., *between 10 and 20 phone calls have been registered each hour, from 1/1/2015 to 8/1/2015*) and monitoring (e.g., *John had between 70 and 80 heart-beats per minute from 30/6/2015 at 8:00 to 30/6/2015 at 9:30*). In all such domains, it is quite unrealistic to pretend that the exact time of each episode (each one of the repetitions) is exactly known. For instance, in the therapy for multiple myeloma, only the periodicity and number of repetitions of melphalan and prednisone administration can reasonably be specified, and not the exact time of each administration. Similar considerations hold for all the other domains elicited above. However, despite their diffusion, dealing with *nearly periodic* facts in the relational context is a new goal in TDBs. In TDBs, only *strongly periodic* events (see, e.g., the survey by Terenziani [7]) have been considered, plus a very recent approach coping with *intermittent* events [6].

Coping with nearly periodic events requires the joint treatment of: (i) the *number* (called *cardinality*) of repetitions (in each instance of the periodicity), (ii) the *periodicity*, and (iii) *temporal indeterminacy*, since the instances of the periodicity are not the exact valid times of the repetitions, but just time intervals *containing* them[1]. In the current TDB literature, the above phenomena have only been faced independently of each others (see Section 5). In Section 2 we show that coping with issues (i), (ii), and (iii) together is a challenging problem, requiring an in-depth extension of current TDB techniques (as shown in Section 4, where Property 5 clarifies the differences between current algebrae and our extended one). In this paper, we extend the current TDB literature to cope with such phenomena. Proofs, as well the definitions of the auxiliary functions ($Ext^t, Gran^*$, and *fragments*), are reported in the digital library as supplementary material, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TKDE.2016.2585483.

2 Main Problems and Solutions

We aim at identifying a *data model* and *relational algebraic operators* to cope with nearly periodic facts (like the ones in Examples 1 and 2 above) in relational TDBs. This is a challenging and innovative goal, since an integrated treatment of phenomena (i), (ii), and (iii) above must be devised.

1. Notably, this aspect sharply distinguishes strongly periodic facts from nearly periodic ones, since in the former the instances of the periodicity are *exactly* the valid time of each repetition (so that *no temporal indeterminacy* occurs), while in the latter they are just intervals *containing* them (and thus *temporal indeterminacy* occurs). The main difference with respect to *intermittent* facts is that such facts are repeated but not at periodic time, so that *periodicity* has *not* to be faced.

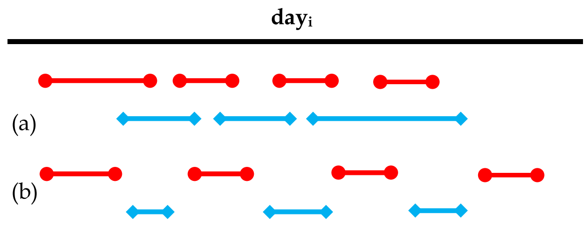

Fig. 1. Different scenarios for Q1.

As regard the cardinality of repetitions, duplicates could be used: one tuple for each occurrence of the fact could be inserted in the same relation. However, duplicates are space-consuming, and cannot cope with the case in which the *exact* number of repetitions is unknown (as in Example 2). We thus propose a compact "intensional" solution, where numeric attributes are used to represent the *number of repetitions* of a fact in each instance of a periodicity.

The second phenomena is *periodicity*. There is a trivial way to cope with periodic data, namely by explicitly storing all of them. Such an approach, usually called "extensional" (or "explicit"), reduces periodic data to standard non-periodic ones. E.g., a standard "extensional" approach represents Ex.1 above using, for each patient, 90 tuples, modeling 60 melphalan applications, and 30 prednisone applications. Despite its simplicity, the extensional approach has a main disadvantage: it is very expensive in terms of physical disk I/O's, due to the high storage size. Also, the extensional approach is not possible in case the exact number of repetitions is unknown (as in Example 2). For such reasons, as, e.g., in [8], [9], [10], [11], [12], [13], we propose an "intensional" representation of periodic facts. Our algebraic operators directly operate on such a representation, providing a compact intensional representation as a result. Our result is directly computed only on the basis of the input intensional representation, without resorting to its extension. This procedure is efficient since it only requires a manipulation on a compact representation, but demands a proof of correctness (see Property 4 below): we have to prove that the intensional manipulation provides the same results that would be obtained by operating on the extensions (i.e., on the explicit repetitions).

The third phenomenon is *temporal indeterminacy* [14]. Indeed, in Example 2, each working day is *not* the *exact* time when a class takes place: it is a span of time *containing* the classes. Such a form of temporal indeterminacy makes the definition of *relational algebraic operators* quite challenging, as shown below. Consider, for instance, Example 3 below, and suppose we want to know

Q1) when did both Ann and John had a drug administration?

Example 3. Each day from 1/1/2016 to 10/1/2016, Ann had 3 or 4 administrations of drug X, and John had 3 administrations of drug Y.

In each day, different scenarios are possible. Two of them are shown in Fig. 1 (where Ann's administrations are red, and we consider 4 administrations, and John's ones are in blue). Scenarios (a) and (b) show two extreme situations: in (a) the cardinality of the intersection is the maximum possible between four and three intervals (i.e., six). In (b) the intersection is empty (its cardinality is zero).

Abstracting from the specific example, even in case two periodic facts occur over the same (or intersecting) instances of a periodicity (e.g., on the same days in Ex.3), since we don't know the exact temporal location of the facts (but just the instance of the periodicity containing them), we cannot know (i) the exact location of the intersections (but just the instance of periodicity containing them), and (ii) the exact number of the intersections, but just a minimum (zero in the example) and maximum (six in the example) bound for them. Indeed, the example in Fig. 1 demonstrates that, even in case the exact input cardinalities are known, the cardinalities obtained after the application of relational operators may only be bounded by a minimum and a maximum value. However, we stress that such a behaviour is not due to our choice of the data model and algebraic operators, but is an intrinsic feature of the phenomena we cope with. In the rest of the paper, the above ideas are detailed.

3 Data Model

Tuples are associated with *valid time* (for the sake of brevity, *transaction time* is not considered in this paper). The timeline is partitioned into granules of a chosen *basic granularity*. As is BCDM [4; Chap.X] (which is the semantic model underlying many TDB approaches, including the "consensus" TSQL2 [4]), the time domain is totally ordered and is isomorphic to the subsets of the domain of natural numbers. The domain of valid times D_{VT} is given as a set $D_{VT} = \{t_1, t_2, \ldots, t_k\}$ of granules.

To provide an implicit ("intensional") representation of periodicities (elsewhere called *periodic granularities*), we start from the below definition and property, taken from [15], and widely used within the TDB community:

Definition 1 (Periodic Granularity). *A periodic granularity H is a granularity periodic with respect to the bottom granularity G.*

Property 1 (Finite Representation). *A periodic granularity H can be finitely described in terms of granules of G providing the following information:*

(1) the finite sets $S_0, \ldots, S_{i-1}$ of indexes of G each one describing the composition of one of the n repeating non-empty granules of H;
(2) the number of granules in which the pattern (1) repeats
(3) the indexes of the first and last non-empty granules in H, if their value is not infinite.

Example 4. Consider, for instance, the definition of school-days (e.g., from hour 9 to hour 13, from Monday to Friday) in the frame time from day 1 to day 100. Let us indicate with G(1) the first hour of day 1 and let us suppose that day 1 is a Monday.

(1) {{9, 10, 11, 12, 13}, {33, 34, 35, 36, 37}, {57, 58, 59, 60, 61}, {81, 82, 83, 84, 85}, {105, 106, 107, 108, 109}}
(2) 168
(3) [9, 2389]

Our relational representation of each periodicity is based on Property 1. The periodicity is represented through three temporal attributes: **Ppat** (*periodic pattern*) indicates the first component above, as a set of time intervals (e.g., {[9,13], [33,37], [57,61], [81,85], [105,109]} considering Ex.4), **PrepT** (*periodic repetition time*) the second component (e.g., 168, i.e., the number of hours constituting a week, in Ex.4), and **FT** (frame time) the frame of time of interest (e.g., [1, 2400] in Ex.4; notice that [9,2389] can be easily deducted from [1, 2400] and the other two components of the definition). Two additional attributes, **N** and **M**, are introduced to model the minimum and maximum number of occurrences of the nearly periodic fact in each instance of the periodicity.

Definition 2 (NP Relation and Tuple). *The schema of a "NP" (nearly periodic) temporal relation* $R = (A_1, \ldots, A_n | N, M, PrepT, Ppat, FT)$ *consists of an arbitrary number of non-temporal attributes* $A_1, \ldots, A_n$*, encoding some fact, of a minimal cardinality attribute N (domain* $\mathbb{N}$*), of a maximal cardinality attribute M (domain* $\mathbb{N}^+$*), an attribute PrepT (domain* $\mathbb{N}^+$*) representing the number of granules in which the periodic pattern repeats, an attribute Ppat representing the pattern (for the sake of simplicity, and with no loss of generality, we assume that*

TABLE 1
Relation ADMIN^{NP}

Patient	Drug	N	M	PrepT	Ppat	FT
Sue	Mel	2	2	28	{[2,2],[3,3],[4,4],[5,5],[6,6]}	[2,169]
Sue	Pre	1	1	28	{[2,2],[3,3],[4,4],[5,5],[6,6]}	[2,169]
Ann	X	3	4	1	{[1,1]}	[1,10]
John	Y	3	3	1	{[1,1]}	[1,10]

the first periodic pattern–with respect to the granule G(0)- is stored) through a temporal element *[4] on the domain* $2^{\mathrm{D_{VT}}\times\mathrm{D_{VT}}}$*, and of an attribute FT representing a frame time (i.e., a non-empty time interval, on the domain* $\mathrm{D_{VT}}\times\mathrm{D_{VT}}$*). Thus, a "NP tuple"* $\mathrm{x}=(a_1,\ldots,a_n|n_1,m_1,d,p,f)$ *in an NP relation r(*R*) on the schema R consists of a* n*-tuple of values for the non-temporal attributes associated with a minimum cardinality* n_1*, a maximum cardinality* m_1*, a natural number* d *(the value of PrepT) a notempty set* p $(p=\{[s_1,e_1],\ldots,[s_j,e_j]\})$ *of ordered and disjoint time intervals, and a time interval* $f=[\mathrm{s},\mathrm{e}]$ *such that: (i)* $\mathrm{n}_1\le\mathrm{m}_1,\mathrm{n}_1\ge0,\ \mathrm{m}_1>0$*; (ii)* $\mathrm{d}\ge(\mathrm{e_j}-\ \mathrm{s}_1)$*, (iii)* $s_1\le e_1<\cdots<s_j\le e_j$*, and (iv)* $\mathrm{s}\le\mathrm{e}$*. x represents the fact that* there are between n_1 and m_1 occurrences of the fact $a_1,\ldots,a_n$ in each interval $[k*d+s_i,k*d+e_i]$ such that $[s_i,e_i]\in p,\mathrm{k}\in\mathrm{Z}$, and $[k*d+s_i,k*d+e_i]\subseteq f$.

As an example, consider the relation ADMIN^{NP} modelling drug administrations at the granularity of days. The schema of ADMIN^{NP} is <Patient, Drug | N, M, PrepT, Ppat, FT>. The first two tuples represent the melphalan ("Mel") and prednisone ("Pre") administrations to Sue, starting her treatment on 2/1/2016, the other two tuples model Example 3. Day "1" represents 1/1/2016.

Notation 1. *Given a tuple x defined on the schema* $\mathrm{R}=(\mathrm{A}_1,\ldots,\mathrm{A_n}|\mathrm{N},\mathrm{M},\mathrm{PrepT},\mathrm{Ppat},\mathrm{FT})$*, we denote by A the set of attributes* $\mathrm{A}_1,\ldots,\mathrm{A_n}$*.* $\mathrm{x}[\mathrm{A}]$ *denotes the values of the A attributes in x, while* $\mathrm{x}[\mathrm{Y}](\mathrm{Y}\in\{\mathrm{N},\mathrm{M},\mathrm{PrepT},\mathrm{Ppat},\mathrm{FT}\})$ *denotes the value of attribute Y in the tuple x.*

3.1 Expressiveness and Consistent Extension Properties

Property 2 grants that our representation is expressive enough: it can represent all periodic granularities (i.e., periodicities), as defined in the TDB literature. We can thus represent nearly periodic facts occurring at any periodicity (i.e., at any level of granularity).

Property 2 (Expressiveness). *Our data model can represent periodic granularities, as defined in [15].*

Nearly periodic facts are inherently temporally indeterminate facts, since the exact time of occurrence of each repetition is not known. Recently, Anselma et al. [17] have proposed a family of data models and algebrae to cope with different forms of temporal indeterminacy (but not with cardinality and periodicity). In particular, in the ITE approach [17], a (possibly non-convex) set of granules $\mathrm{G_S}$ is used to represent the valid time of a tuple, meaning that the tuple may hold at any possible subset of the granules in $\mathrm{G_S}$. Our data model is a consistent extension of the "ITE" model.

Property 3. *"ITE" relations can be modelled by NP relations in our approach.*

E.g., the NP tuple $<\mathrm{Mary},\mathrm{X}|1,1,7,\{[1,7]\},[1,7]\}>$ in our approach represents the fact that Mary had a drug administration of the drug X during the first week of 2016. To model a single fact, the minimum and maximum cardinalities are set to "1", FT correspond to the valid time, Ppat coincides with FT and PrepT with its duration.

Notice also that, since we do not model the exact valid time of nearly periodic facts (which is unknown), but just the instances of periodicity containing them, and the number of occurrences in such instances, our model applies to both the case in which facts are *durative* and *punctual* (i.e., to both "states" and "events" [4]).

4 Temporal Relational Algebra

Codd defined as complete any query language that is as expressive as his set of five relational algebraic operators: relational union ($\cup$), relational difference (–), selection (σ_P), projection (π_A), and Cartesian Product ($\times$) [18]. We propose a temporal extension of Codd's operators to query the data model in Section 3. Several temporal extensions to Codd's operators have been provided in the TDB literature [16]. In most cases, such extensions behave like standard non-temporal operators on the non-temporal attributes, and involve the application of set operators on the temporal attributes. For instance, in TSQL2 "consensus" approach, (i) Cartesian Product involves pairwise concatenation of the values of non-temporal attributes and pairwise intersection of their temporal values, (ii) difference r-s operates in the standard way on non-temporal attributes, and make the difference of valid times (by subtracting from each tuple f∈r the valid times of all the tuples $\mathrm{f}'\in\mathrm{s}$ *value equivalent* [4] to it), and (iii) relational union, non-temporal selection, and projection operate in the standard way on the non-temporal part, and do not operate on the temporal part.

4.1 Relational Algebra for Irregular Repetitions

We ground our approach on such a "consensus" background, extending the algebraic operators to cope with the new attributes. In the rest of the paper, we use the superscript "NP" (Nearly Periodic) for our operators, while $\cup,\cap,-$, denote standard set operators. We preliminarily define the set $VE\_INT(v,t_1,s)$, and the function $\mathbf{Ext^t}$. In the following, "$\{\mathrm{x}\backslash\ldots\}$" stands for "all x's such that".

Definition 3 (VE_INT: Set of Value Equivalent Tuples). *Given two relations r and s defined over the schema* $\mathrm{R}=(\mathrm{A}_1,\ldots,\mathrm{A_n}|\mathrm{N},\mathrm{M},\mathrm{PrepT},\mathrm{Ppat},\mathrm{FT})$*, and a tuple* $\mathrm{f}=<\mathrm{v}|\mathrm{n}_1,\mathrm{m}_1,\mathrm{d}_1,\mathrm{p}_1,\mathrm{t}_1>\ \in\mathrm{r}$*, we define* $VE\_INT(f[\mathrm{A}_1,\ldots,\mathrm{A_n}],f[FT],s)=\{\mathrm{f}'\backslash\mathrm{f}'\in\mathrm{s}\wedge\mathrm{f}'[\mathrm{A}_1,\ldots,\mathrm{A_n}]=\mathrm{f}[\mathrm{A}_1,\ldots,\mathrm{A_n}]\wedge\mathrm{f}[\mathrm{FT}]\cap\mathrm{f}'[\mathrm{FT}]\neq\emptyset\}$ *as the set of all and only the tuples* f' *in s that are value equivalent to f and whose frame time temporally intersects the frame time f[FT].*

Definition 4 (Temporal Extension $\mathbf{Ext^t(p,d,I)}$). *Given a periodic pattern* p*, a periodic repetition time* d*, and a time interval I,* $\mathrm{Ext^t(p,d,I)}$ *is the set of all time intervals obtained by making explicit all the repetitions of p in I.*

As an example, $\mathrm{Ext^t}(\{[2,2],[3,3],[4,4],[5,5],[6,6]\},28,[2,169])$ denotes the set $\{[2,2],[3,3],[4,4],[5,5],\ [6,6],[30,30],[31,31],[32,32],[33,33],[34,34],\ \ [58,58],\ldots,[146,146]\}$ of all the time intervals (denoting a single day) in which prednisone has been administered to Sue.

In the definition of difference, we also use the function fragments. ***fragments***$(t,\{t_1,\ldots,t_n\})$ provides in output an ordered list of intervals, obtained by partitioning the time interval *t* into non-overlapping covering parts, having as endpoints the starting and ending points of t and of $\mathrm{t}_1,\ldots,\mathrm{t_n}$. As a simple example, fragments ([50, 100], {[30, 70], [60, 80], [50,120], [20,150]}) = {[50, 60), [60, 70), [70, 80), [80, 100]}. Since, as in TSQL2, our union, projection and non-temporal selection do not modify the temporal attributes, we do not report them in Definition 5.

Definition 5 (Temporal Algebraic Operators). *Let* $\mathrm{r^{NP}}$ *and* $\mathrm{s^{NP}}$ *denote NP relations in our model having the proper schema (A stands for the set of non-temporal attributes).*

$\mathbf{r^{NP}\times^{NP}s^{NP}}=\{<\mathrm{v}_1\cdot\mathrm{v}_2|\mathrm{n},\mathrm{m},\mathrm{d},\mathrm{p},\mathrm{t}>\backslash\exists\mathrm{r}_1\in\mathrm{r^{NP}}\wedge\exists\mathrm{s}_1\in\mathrm{s^{NP}}\wedge\mathrm{v}_1=\mathrm{r}_1[\mathrm{A}]\wedge\mathrm{v}_2=\mathrm{s}_1[\mathrm{A}]\wedge\mathrm{n}=0\wedge\mathrm{m}=\mathrm{r}_1[\mathrm{M}]+\mathrm{s}_1[\mathrm{M}]-1\wedge$

$\mathrm{d} = \mathrm{lcm}(\mathrm{r}_1[\mathrm{PrepT}], \mathrm{s}_1[\mathrm{PrepT}]) \wedge \mathrm{t} = \mathrm{r}_1[\mathrm{FT}] \cap \mathrm{s}_1[\mathrm{FT}] \wedge \mathrm{t} \neq \emptyset \wedge \mathrm{p} = (\mathrm{Ext}^{\mathrm{t}}(\mathrm{r}_1[\mathrm{Pper}], \mathrm{r}_1[\mathrm{PrepT}], [\mathrm{start}(\mathrm{t}), \mathrm{start}(\mathrm{t}) + \mathrm{d}]) \cap \mathrm{Ext}^{\mathrm{t}}(s_1[Pper], s_1[PrepT], [start(t), start(t) + d]))\}$

$$\mathbf{r}^{\mathbf{NP}} -^{\mathbf{NP}} \mathbf{s}^{\mathbf{NP}} = \{ < \mathbf{v} | \mathbf{n}, \mathbf{m}, \mathbf{d}, \mathbf{p}, \mathbf{t} > \backslash$$

$\exists \mathrm{r}_1 \in \mathrm{r}^{\mathrm{NP}} \wedge \mathrm{n} = \mathrm{r}_1[\mathrm{N}] \wedge \mathrm{m} = \mathrm{r}_1[\mathrm{M}] \wedge \mathrm{d} = \mathrm{r}_1[\mathrm{PrepT}] \wedge \mathrm{p} = \mathrm{r}_1[\mathrm{Pper}] \wedge \mathrm{t} = \mathrm{r}_1[\mathrm{FT}] \wedge \mathit{VE\_INT}(\mathrm{r}_1[\mathrm{A}], \mathrm{r}_1[\mathrm{FT}], \acute{s}^{NP}) = \emptyset) \vee \exists \mathrm{r}_1 \in \mathrm{r}^{\mathrm{NP}} \wedge \mathit{VE\_INT}(\mathrm{r}_1[\mathrm{A}], \mathrm{r}_1[\mathrm{FT}], \mathrm{s}^{\mathrm{NP}}) = \mathrm{s}_1, \ldots, \mathrm{s}_\mathrm{n}\} \wedge \exists\, \mathrm{f} \in \mathrm{fragments}(\mathrm{r}_1[\mathrm{FT}], \{s_1[FT], \ldots, s_n[FT]\}) \wedge \mathit{VE\_INT}(r_1[\mathrm{A}], \mathrm{f}, \mathrm{s}^{\mathrm{NP}}) = \{\mathrm{s}'_1, \ldots, \mathrm{s}'_\mathrm{k}\} \wedge n = 0 \wedge m = r_1[M] + \sum_{s_i \in \{s'_1, \ldots, s'_k\}} s_i[M] \wedge \mathrm{p} = (\mathrm{Ext}^{\mathrm{t}}(\mathrm{r}_1[\mathrm{Pper}], \mathrm{r}_1[\mathrm{PrepT}], [start(f), start(f) + d]) - (\mathrm{Ext}^{\mathrm{t}}(\mathrm{s}'_1[\mathrm{Pper}], \mathrm{s}'_1[\mathrm{PrepT}], [\mathrm{start}(\mathrm{f}), start(f) + d]) \cup \ldots \cup Ext^t(\mathrm{s}'_\mathrm{k}[\mathrm{Pper}], \mathrm{s}'_\mathrm{k}[\mathrm{PrepT}], [start(f), start(f) + d])))\}$

As motivated above, all our algebraic relational operators operate in the standard way on the non-temporal attributes. Considering Cartesian Product ($\times^{\mathrm{NP}}$), for each pair of tuples (one from r^{NP} and one from s^{NP}) the output is a tuple which has as non-temporal part the concatenation of the two non-temporal parts, and as temporal part the intersection of the temporal parts. In our approach, the intersection is obtained by (i) intersecting the frame times and (ii) intersecting the periodic patterns over a period of time which starts at the intersection of the frame times, and whose periodic repetition time is the repetition time of the output periodicity, i.e., the *least common multiple* (*lcm*) of the duration of the two input patterns. As discussed in Section 2, the minimum cardinality is 0 (see, e.g., Fig. 1b), and the maximum cardinality is the sum of the two input maximum cardinalities, minus 1 (see, e.g., Fig. 1a).

In the definition of temporal difference $\mathrm{r}^{\mathrm{NP}} -^{\mathrm{NP}} \mathrm{s}^{\mathrm{NP}}$ we have to manage (as in TSQL2) the fact that an unpredictable number of value-equivalent tuples may be present in the input relations. Intuitively speaking, each tuple $x \in \mathrm{r}^{\mathrm{NP}}$ (*x* has the form $< v | n_1, m_1, d_1, p_1, t_1 >$) which has no value-equivalent tuple in s^{NP} intersecting its frame time is simply reported unchanged in output. Otherwise, the set of all the tuples $x_1, \ldots, x_n$ in s^{NP} value equivalent to *x* that intersect x's frame time (i.e., t_1) must be retrieved (*VE_INT* function). The frame time t_1 of *x* is split into "fragments" (i.e., subintervals) by the frame times of $x_1, \ldots, x_n$ (function *fragments*). There is an output tuple for each one of such fragments. For each one of these fragments *f*, only the value-equivalent tuples $y_1, \ldots, y_k$ in s^{NP} that cover it must be considered. The frame time of the output tuple is *f*, the periodic repetition time *d* is the *least common multiple* (*lcm*) of the repetition time of *x* and the repetition times of $y_1, \ldots, y_k$, and the periodic pattern is the pattern obtained by making the difference between the extension (Ext^t function) of the pattern p_1 of *x* (over the time interval starting at the start of *f*, and with duration *d*), and the union of the extensions of the patterns of $y_1, \ldots, y_k$ (over the time interval starting at the start of *f*, and with duration *d*). The minimum cardinality is 0, since the valid times of all repetitions of *x* can be "covered" by the valid times of tuples $y_1, \ldots, y_k$. The maximum cardinaliy is the sum of all the maximum cardinalities (i.e., the maximum cardinality m_1 of *x* plus the maximum cardinalities of $y_1, \ldots, y_k$. This is due to the fact that the difference between a pair of time intervals may generate a maximum of two time intervals).

As an example, (Q1) in Section 2 can be asked as follows:

$$(\sigma^{\mathbf{NP}}{}_{\mathbf{patient=Ann}}(\mathbf{ADMIN^{NP}})) \times^{\mathbf{NP}} (\sigma^{\mathbf{NP}}{}_{\mathbf{patient=John}}(\mathbf{ADMIN^{NP}}))$$

4.2 Properties of the Algebra

As motivated in Section 2, since our algebraic operators perform an intensional manipulation of the representation, a proof of correctness is required. Since we both manipulate (i) time intervals (the instances of our intensional representation of periodicity) and (ii) cardinalities (the number of occurrences of the fact within each time interval), we need to consider both in the proof. To deal with the correctness of the issue (i), we introduce the "extension" operator **Ext**. For each tuple, it adopts the function $\mathrm{Ext}^{\mathrm{t}}$ to make explicit the set S of all the repetitions of a periodic pattern in a frame time, and gives in output a new (value-equivalent) tuple for each one of the intervals in S.[2]

Definition 6 (Extension Ext($\mathbf{r}^{\mathbf{NP}}$)). *Given a NP relation* r^{NP} *defined on the schema* $\mathrm{R} = (\mathrm{A}_1, \ldots, \mathrm{A}_\mathrm{n} | \mathrm{N}, \mathrm{M}, \mathrm{PrepT}, \mathrm{Ppat}, \mathrm{FT})$, *and given* $\mathrm{R}' = (\mathrm{A}_1, \ldots, \mathrm{A}_\mathrm{n} | \mathrm{T})$ *the corresponding schema in the extensional (e.g., TSQL2) model,*

$$\mathbf{Ext}(\mathbf{r}^{\mathbf{NP}}) = \{\mathrm{z} \backslash \exists \mathrm{x} \in \mathrm{r}^{\mathrm{NP}}, \exists \mathrm{I} \in \mathrm{Ext}^{\mathrm{t}}(\mathrm{x}[\mathrm{Ppat}], \mathrm{x}[\mathrm{PrepT}], \mathrm{x}[\mathrm{FT}]) \wedge \mathrm{z}[\mathrm{A}] = \mathrm{x}[\mathrm{A}] \wedge \mathrm{z}[\mathrm{T}] = \mathrm{I}\}$$

Property 4(a) (Correctness of the Manipulation of Time Intervals). Our extended algebraic operators, operating on the intensional temporal model, operate correctly on time intervals: for each algebraic operator $\mathrm{Op}^{\mathrm{NP}}$ in our approach, $\mathrm{Ext}(\mathrm{r}^{\mathrm{NP}} \mathrm{Op}^{\mathrm{NP}} \mathrm{s}^{\mathrm{NP}}) = \mathrm{Ext}(\mathrm{r}^{\mathrm{NP}})\ \mathrm{Op}\ \mathrm{Ext}(\mathrm{s}^{\mathrm{NP}})$, where Op is the operator in an extensional approach (e..g., TSQL2) corresponding to $\mathrm{Op}^{\mathrm{NP}}$.

Property 4(b) (Correctness of the Manipulation of Cardinalities). *Our extended algebraic operators provide as output the correct cardinalities.*

Reducibility is fundamental for all TDB approaches, to grant that the new operators, which extend simpler operators to cope with new phenomena, reduces to simpler operators when the new phenomena are disregarded [4], [16]. Since we cope with repetitions for which the *exact* valid time is not known, it is appropriate to reduce our approach to a TDB approach coping with temporal indeterminacy, specifically to the "ITE" approach [17] (see also Section 3.1). First, we define a reduction operator $\mathrm{R}^{\mathrm{ITE}}$. In the following, we use the auxiliary function $Gran^*(\{I_1, \ldots, I_n\})$ that takes in input a set of (convex) time intervals, and returns the granules it contains (e.g., $\mathrm{Gran}^*(\{[3,5],[7,8]\}) = \{3,4,5,7,8\}$).

The reduction operator $\mathrm{R_I}^{\mathrm{ITE}}$ is defined as follows.

Definition 7. $\mathbf{R_I}^{\mathbf{ITE}}$. *Let* r^{NP} *a NP relation, defined on the schema* $\mathrm{R} = (\mathrm{A}_1, \ldots, \mathrm{A}_\mathrm{n} | \mathrm{N}, \mathrm{M}, \mathrm{PrepT}, \mathrm{Ppat}, \mathrm{FT})$, *let* $\mathrm{R}' = (\mathrm{A}_1, \ldots, \mathrm{A}_\mathrm{n} | \mathrm{T})$ *BE the corresponding schema in the ITE model, where* $\mathrm{T} \in 2^{\mathrm{D_{VT}}}$, *and I a time interval.*

$$\mathrm{R_I}^{\mathrm{ITE}}(\mathrm{r}^{\mathrm{NP}}) = \{\mathrm{z} \backslash \exists \mathrm{x} \in \mathrm{r}^{\mathrm{NP}} \mathrm{z}[\mathrm{A}] = \mathrm{x}[\mathrm{A}] \wedge \mathrm{x}[\mathrm{M}] > 0 \wedge \mathrm{z}[\mathrm{T}] = \mathrm{Gran}^*(\mathrm{Ext}^{\mathrm{t}}(\mathrm{x}[\mathrm{Ppat}], \mathrm{x}[\mathrm{PrepT}], \mathrm{x}[\mathrm{FT}])) \cap \mathrm{Gran}^*(\mathrm{I})\}$$

Given the above definition, Property 5 holds.

Property 5 (Reducibility to ITE). *Our algebra is reducible to ITE's algebra, i.e., for each time interval* $\mathrm{I}, \mathbf{R_I}^{\mathbf{ITE}}(\mathbf{r}^{\mathbf{NP}} \mathbf{Op}^{\mathbf{NP}} \mathbf{s}^{\mathbf{NP}}) = (\mathbf{R_I}^{\mathbf{ITE}}(\mathbf{r}^{\mathbf{NP}})\ \mathbf{Op}^{\mathbf{ITE}}\ \mathbf{R_I}^{\mathbf{ITE}}(\mathbf{s}^{\mathbf{NP}}))$, *where* $\mathrm{Op}^{\mathrm{NP}}$ *and* $\mathrm{Op}^{\mathrm{ITE}}$ *represent corresponding relational operators in our algebra and in the ITE algebra respectively.*

Since they manipulate the implicit representation, it is worth reporting the complexity of our operators (see the supplementary material, available online for a detailed analysis).

Property 6 (Complexity of Algebraic Operators). *Union, projection and non-temporal selection behave like traditional operators. Cartesian product operates in a time proportional to the product of the number of tuples of the input relations. For each pair of tuples, temporal intersection is performed in a time linear in the number of intervals in the periodic pattern. Difference, as regards the nontemporal part, behaves like standard difference. As regards the temporal component,*

2. For the sake of clarity, here we simplify our approach, considering the case in which the minimum and maximum cardinalities are both equal to one. In the Supplementary Material, available online, we generalize such a definition.

difference can involve a computation which is exponential in the number of value-equivalent tuples.

Such an exponential factor is not due to our approach: in *any* approach in which the (intervals in the) periodic patterns are explicitly represented, the union of n periodic patterns involves the evaluation of a number of intervals which may grow in an exponential way with respect to n.

5 Related Works

Despite their relevance, *nearly periodic events* have not been studied by the TDB literature yet, except by Clifford et al. [5], where they are defined, and an abstract (*not relational*) representation has been provided for them. Differently from [5], we provide a *relational* representation for them (considering also *cardinalities*, not coped with in [5]), and an *algebra* operating on it, and studying its properties. To cope with nearly periodic events, we have proposed the *first integrated* TDB approach considering *together* (i) *cardinality* of repetitions, (ii) *periodicity*, and (iii) *temporal indeterminacy*. Such three issues have been coped with only *separately* in the previous TDB literature.

In TDB, *periodicity* has been studied to deal with "*strongly periodic*" events (see, e.g., [5], [7]). Different approaches have been proposed to cope in an *intensional* way with periodicity. They can be divided into three mainstreams (the terminology is derived from [8], [9]): (i) *Deductive rule-based* approaches, using deductive rules. For instance, Chomicki and Imielinsky [10] dealt with periodicity via the introduction of the successor function in Datalog; (ii) *Constraint-based* approaches, using mathematical formulae and constraints (e.g., [11]); *Symbolic* approaches, like, e.g., the ones by Leban et al., [12] and Terenziani [13], providing symbolic languages to cope with temporal periodicity in a compositional way.

The *cardinality* of repetitions (but not their periodicity) has been recently addressed by Terenziani [6], who proposed the only TDB approach coping with "*intermittent events*" [5]. [6] provides both a relational representation model and an algebra for them, and studies the reducibility of the algebra.

A survey of TDB approaches to *temporal indeterminacy* has recently been provided in [14]. Dyreson and Snodgrass [19] and Dekhtyar et al. [20] have proposed *probabilistic* approaches. Recently, Anselma et al. [17] have introduced a *family* of algebraic approaches coping with different forms of temporal indeterminacy. Our model is a consistent extension of the ITE approach in [17] (see Property 3), and, if we disregard repetitions, our algebra can be reduced to ITE's one (Property 5).

6 Discussion and Conclusions

Despite their importance, our approach is the first one coping with *nearly periodic events* in the relational TDB area, providing a relational representation and an algebra for them. We have studied the expressiveness of our representation of periodicity, as well as the reducibility and the correctness of our relational algebra, thus providing the first comprehensive and theoretically-grounded approach to nearly periodic events in the relational context.

In our previous work, we have extended the relational model to cope with *strongly periodic* [13] and *intermittent* [6] facts. In our future work we plan to devise an homogeneous relational framework in which all types of repeated events (see [5]) are managed together. To achieve such a goal, we plan to propose (i) conversion operators, to switch from (relational models of the) different types of repetitions, and (ii) relational algebraic operators, applying to different types of relations (e.g., Cartesian Product between an NP relation and a strongly periodic relation).

Finally, temporal indeterminacy in nearly periodic facts may involve a degree of fuzziness about the distribution of fact instances into the time intervals containing them. To model such a phenomenon, in our future work, we envision the possibility of extending our approach with *fuzzy logic* or *probabilities*.

References

[1] R. T. Snodgrass, *Developing Time-Oriented Database Applications in SQL*. San Francisco, CA, USA: Morgan Kaufmann, Jul. 1999.

[2] Y. Wu, S. Jajodia, and X. S. Wang, "Temporal Database Bibliography Update," *Temporal Databases: Research and Practice*. Berlin, Germany: Springer, pp. 338–366, 1997.

[3] L. Liu and M. T. Özsu, *Encyclopedia of Database Systems*, Berlin, Germany: Springer, 2009, ISBN: 978-0-387-39940-9.

[4] R. T. Snodgrass, Ed., *The TSQL2 Temporal Query Language*. Norwell, MA, USA: Kluwer, 1995.

[5] A. Tuzhilin and J. Clifford, "On periodicity in temporal databases, information systems," *Inf. Syst.*, vol. 20, no. 8, pp. 619–639, 1995.

[6] P. Terenziani, "Irregular indeterminate repeated facts in temporal relational databases," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 4, pp. 1075–1079, Apr. 2016.

[7] P. Terenziani, "Temporal Periodicity," *Encyclopedia of Database Systems*. Berlin, Germany: Springer, 2009.

[8] M. Baudinet, J. Chomicki, and P. Wolper, "Temporal Deductive Databases," in *Temporal Databases*, A. Tansel, et al., Eds. San Francisco, CA, USA: Benjamin/ Cummings, 1993, pp. 294–320.

[9] M. Niezette and J.-M. Stevenne, "An efficient symbolic representation of periodic time," in *Proc. 1st Int. Conf. Inf. Knowl. Manage.*, Nov. 1992, pp. 161–168.

[10] J. Chomicki and T. Imielinsky, "Temporal deductive databases and infinite objects", in *Proc. 7th ACM Symp. Principles Database Syst.*, Mar. 1988, pp. 61–73.

[11] F. Kabanza, J.-M. Stevenne, and P. Wolper, "Handling infinite temporal data," in *Proc. ACM SIGACT-SIGMOD-SIGART Symp. Principles Database Syst.*, 1990, pp. 392–403.

[12] B. Leban, D. D. McDonald, and D. R. Forster, "A representation for collections of temporal intervals," in *Proc. 5th Nat. Conf. Artif. Intell.*, 1986, pp. 367–371.

[13] P. Terenziani, "Symbolic user-defined periodicity in temporal relational databases," *IEEE Trans. Knowl. Data Eng.*, vol. 15, no. 2, pp. 489–509, Mar./Apr. 2003.

[14] C. Dyreson, "Temporal Indeterminacy," *The TSQL2 Temporal Query Language*. Norwell, MA, USA: Kluwer, 1995.

[15] C Bettini and R De Sibi, "Symbolic representation of user-defined time granularities," *Ann. Math. Artif. Intell.*, vol. 30, no. 1–4, pp. 53–92, 2000.

[16] L. E. McKenzie, R. T. Snodgrass, "Evaluation of relational algebras incorporating the time dimension in databases," *ACM Comput. Surv.*, vol. 23, no. 4, pp. 501–543, 1991.

[17] L. Anselma, P. Terenziani, and R. T. Snodgrass, "Valid-time indeterminacy in temporal relational databases: Semantics and representations," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 12, pp. 2880–2894, Dec. 2013.

[18] E. F. Codd, "Relational completeness of data base sublanguages," in *Database Systems*, R. Rustin, Ed. San Jose, CA, USA: Prentice Hall, 1972, pp. 65–98, IBM Research Report RJ 987.

[19] C. E. Dyreson and R. T. Snodgrass, "Supporting valid-time indeterminacy," *ACM Trans. Database Syst.*, vol. 23, no. 1, pp. 1–57, 1998.

[20] A. Dekhtyar, R. Ross, and V. S. Subrahmanian, "Probabilistic temporal databases, I: Algebra," *ACM Trans. Database Syst.*, vol. 26, no. 1, pp. 41–95, 2001.