A uniform framework for temporal functional dependencies with multiple granularities

Carlo Combi[1], Angelo Montanari[2], Pietro Sala[3]

[1,3]Department of Computer Science, University of Verona, Italy

[2]Department of Mathematics and Computer Science, University of Udine, Italy

[1]carlo.combi@univr.it

[2]angelo.montanari@uniud.it

[3]pietro.sala@univr.it

SSTD 2011

Outline

Introduction

Motivation: several computer applications require the ability of representing and managing data changing over time.

Temporal databases allow one to describe the temporal evolution of information by associating one or more temporal dimensions with the stored data.

Temporal Functional Dependencies: allow one to express Functional Dependencies over time.

Various representation formalisms for TFDs have been proposed in the literature (Jensen et al., Bettini et al., Wijsen, Vianu) which differ a lot in their structure as well as in the underlying data model.

Purpose: we propose a general logical framework for specifying TFDs which subsumes existing formalisms for TFDs and allows one to encode TFDs which are not captured by existing systems.

Standard Functional Dependencies

Let R be a relational schema over the set of attributes U and $X \subseteq U$, $Y \subseteq U$ be two subsets of U. An instance r of R satisfies a functional dependency $R : X \rightarrow Y$ if and only if

$$\{t \mid \exists t'(r(t) \wedge r(t') \wedge t[X] = t'[X] \wedge t[Y] \neq t'[Y])\} = \emptyset$$

Informally an instance r over the relational schema R satisfies a functional dependency $R : X \rightarrow Y$ if and only if every pair of tuples of r, which agree on the values for the attributes in X, agrees on the values for the attributes in Y.

What is this talk about?

In this work, we propose a general framework that makes it possible to formally specify TFDs, possibly involving multiple time granularities, and to check whether or not a given database instance satisfies them.

We will prove that the proposed framework subsumes all existing formalisms for TFDs, and it allows one to express TFDs which are not captured by them.

Moreover, we will show the effectiveness of the approach by applying it to a real-world medical domain, related to the administration of chemotherapies.

TFDs on a Real World Scenario

To illustrate the relevance of properly expressing and checking temporal constraints on data, we consider a real-world example taken from the domain of chemotherapies for oncology patients.

Example

Recommended FAC and CEF regimes for treatment of breast cancer:

"The recommended FAC regimen consists of 5-fluorouracil on days 1 and 8, and doxorubicin and cyclophosphamide on day 1.
This is repeated every 21 days for 6 cycles (that is, 6 cycles of 21 days each)"

"The recommended CEF regimen consists of 14 days of oral cyclophosphamide, and intravenous injection of epirubicin and 5-fluorouracil on days 1 and 8.
This is repeated every 28 days for 6 cycles."

TFDs on a Real World Scenario

Consider a relation schema Patient to be used for storing information about patients who underwent several chemotherapies.

tuple#	*Chemo*	*P-Id*	*BG*	*Phys*	*Drug*	*Qty*	*VT*
1	FAC	1	0+	Smith	Flu	500	1
2	FAC	1	0+	Hubbard	Dox	50	1
3	FAC	1	0+	Verdi	Cyc	500	1
4	FAC	1	0+	Smith	Flu	500	8
5	CEF	2	AB-	Verdi	Cyc	600	1
6	CEF	2	AB-	Hubbard	Flu	600	1
7	CEF	2	AB-	Hubbard	Epi	60	1
8	CEF	2	AB-	Smith	Cyc	600	2
9	CEF	2	AB-	Verdi	Cyc	500	3
	..	..	..	..	..	..	..
20	CEF	2	AB-	Verdi	Cyc	600	8
21	CEF	2	AB-	Hubbard	Flu	600	8
22	CEF	2	AB-	Hubbard	Epi	60	8
	..	..	..	..	..	..	..
33	CEF	3	AB-	Verdi	Cyc	550	1

TFDs on a Real World Scenario

We want to impose the following constraints over our `Patient` relation;

1 "A patient may take any given drug at most one time per day"

2 "For patients undergoing a specific chemotherapy regimen, with the same drug being prescribed on two consecutive days, the quantity of the drug administered on the latter day depends solely on the drug quantity administered on the former"

3 "For any pair of successive administrations of the same drug to the same patient within the same month, the quantity of the second administration uniquely depends on (the drug and) the quantity of the first administration."

4 "For any given chemotherapy, the quantities of drugs taken by patients with successive administrations that take place one 7 days after the other cannot change"

Temporal data model: our proposal

r is a relation on a schema R

$$R = \{\ \ \mathcal{U}\ \ \cup\ \ \{VT\}\}$$

Temporal data model: our proposal

r is a relation on a schema R

$$R = \{ \; U \; \cup \; \{VT\}\} \quad \text{Set of atemporal attributes}$$

Temporal data model: our proposal

r is a relation on a schema R

$$R = \{ \ U \ \cup \ \{VT\}\}$$

valid time

Adding suitable views: Next View

$next$ view

Let r be a temporal relation with schema $U \cup \{VT\}$

$\chi_Z^k(r)$ with $Z \subseteq U$ and $k \geqslant 1$

joins t,t' (tuples of r)

if and only if

$t[Z] = t'[Z]$ and $t'[VT] = t[VT] + k$

Adding suitable views: Next View

next view

Let r be a temporal relation with schema $U \cup \{VT\}$

$\chi^k_Z(r)$ with $Z \subseteq U$ and $k \geqslant 1$

joins t, t' (tuples of r)

if and only if

$t[Z] = t'[Z]$ and $t'[VT] = t[VT] + k$

Chemo	*P-Id*	*BG*	*Phys*	*Drug*	*Qty*	*VT*
..	..	..	..	..	..	..
CEF	2	AB-	Verdi	Cyc	600	1
CEF	2	AB-	Hubbard	Flu	600	1
CEF	2	AB-	Hubbard	Epi	60	1
CEF	2	AB-	Smith	Cyc	600	2
CEF	2	AB-	Verdi	Cyc	500	3
..	..	..	..	..	..	..
CEF	2	AB-	Verdi	Cyc	600	8
CEF	2	AB-	Hubbard	Flu	600	8
CEF	2	AB-	Hubbard	Epi	60	8
..	..	..	..	..	..	..

Adding suitable views: Next View

$\chi^{7}_{\{\mathrm{P-Id},\mathrm{Drug}\}}(r)$

Chemo	*P-Id*	*BG*	*Phys*	*Drug*	*Qty*	*VT*	$\overline{\text{Chemo}}$	$\overline{\text{P-Id}}$	$\overline{\text{BG}}$	$\overline{\text{Phys}}$	$\overline{\text{Drug}}$	$\overline{\text{Qty}}$	$\overline{\text{VT}}$
..	..	..	..	..	..	..	..	..	..	..	..	..	..
CEF	2	AB-	Verdi	Cyc	600	1	CEF	2	AB-	Verdi	Cyc	600	8
CEF	2	AB-	Hubbard	Flu	600	1	CEF	2	AB-	Hubbard	Flu	600	8
CEF	2	AB-	Hubbard	Epi	60	1	CEF	2	AB-	Hubbard	Epi	60	8
..	..	..	..	..	..	..	..	..	..	..	..	..	..

Chemo	*P-Id*	*BG*	*Phys*	*Drug*	*Qty*	*VT*
..	..	..	..	..	..	..
CEF	2	AB-	Verdi	Cyc	600	1
CEF	2	AB-	Hubbard	Flu	600	1
CEF	2	AB-	Hubbard	Epi	60	1
CEF	2	AB-	Smith	Cyc	600	2
CEF	2	AB-	Verdi	Cyc	500	3
..	..	..	..	..	..	..
CEF	2	AB-	Verdi	Cyc	600	8
CEF	2	AB-	Hubbard	Flu	600	8
CEF	2	AB-	Hubbard	Epi	60	8
..	..	..	..	..	..	..

Adding suitable views: Next Tuple View

$nexttuple$ view

Let r be a temporal relation with schema $U \cup \{VT\}$

$\tau_Z(r)$ with $Z \subseteq U$ and $k \geqslant 1$

joins t,t' (tuples of r) **if and only if**

$t[Z] = t'[Z]$ and t' is the tuple immediately following t with respect to VT

Adding suitable views: Next Tuple View

$$\tau_{\{P-Id,Drug\}}(r)$$

Chemo	*P-Id*	*BG*	*Phys*	*Drug*	*Qty*	*VT*	$\overline{\text{Chemo}}$	$\overline{\text{P}-\text{Id}}$	$\overline{\text{BG}}$	$\overline{\text{Phys}}$	$\overline{\text{Drug}}$	$\overline{\text{Qty}}$	$\overline{\text{VT}}$
..	..	..	..	..	..	..	..	..	..	..	..	..	..
CEF	2	AB-	Verdi	Cyc	600	1	CEF	2	AB-	Smith	Cyc	600	2
CEF	2	AB-	Smith	Cyc	600	2	CEF	2	AB-	Verdi	Cyc	500	3
CEF	2	AB-	Verdi	Cyc	500	3	CEF	2	AB-	Verdi	Cyc	600	8
CEF	2	AB-	Hubbard	Flu	600	1	CEF	2	AB-	Hubbard	Flu	600	8
CEF	2	AB-	Hubbard	Epi	60	1	CEF	2	AB-	Hubbard	Epi	60	8
..	..	..	..	..	..	..	..	..	..	..	..	..	..

Chemo	***P-Id***	***BG***	***Phys***	***Drug***	***Qty***	***VT***
..	..	..	..	..	..	..
CEF	2	AB-	Verdi	Cyc	600	1
CEF	2	AB-	Hubbard	Flu	600	1
CEF	2	AB-	Hubbard	Epi	60	1
CEF	2	AB-	Smith	Cyc	600	2
CEF	2	AB-	Verdi	Cyc	500	3
..	..	..	..	..	..	..
CEF	2	AB-	Verdi	Cyc	600	8
CEF	2	AB-	Hubbard	Flu	600	8
CEF	2	AB-	Hubbard	Epi	60	8
..	..	..	..	..	..	..

Expressing TFDs

Given a relation schema R on attributes $U \cup \{VT\}$, we define the TFD:

$$[E - \mathrm{Exp}(R), \quad t - \mathrm{Group}(i)] \quad X \rightarrow Y$$

Expressing TFDs

Given a relation schema R on attributes $U \cup \{VT\}$, we define the TFD:

$$[E - Exp(R), \quad t - Group(i)] \quad X \rightarrow Y$$

Evolution Expression: a relational calculus expression on R (possibly) involving the views χ_Z^k and τ_Z.

Expressing TFDs

Given a relation schema R on attributes $U \cup \{VT\}$, we define the TFD:

$$[E - Exp(R), \quad t - Group(i)] \quad X \rightarrow Y$$

Temporal Grouping: is a mapping which groups tuples according to the values of their attribute VT.

Expressing TFDs

Given a relation schema R on attributes $U \cup \{VT\}$, we define the TFD:

$$[E - Exp(R), \quad t - Group(i)] \quad X \rightarrow Y$$

Temporal Grouping: is a mapping which groups tuples according to the values of their attribute VT.

Granularity $\longrightarrow$ $\boxed{t - Group(i) \stackrel{\text{def}}{=} G(i)}$

Expressing TFDs

Given a relation schema R on attributes $U \cup \{VT\}$, we define the TFD:

$$[E - Exp(R), \quad t - Group(i)] \quad X \rightarrow Y$$

Temporal Grouping: is a mapping which groups tuples according to the values of their attribute VT.

Granularity → $t - Group(i) \stackrel{\text{def}}{=} G(i)$

Sliding windows →

$$\begin{array}{ccc} t - Group(i) & \stackrel{\text{def}}{=} & \bigcup_{j=1}^{n} \{(i + \alpha_j)\} \\ & where & \alpha_1 = 0, \\ & & \forall j \in [1, n] \ \alpha_j \in \mathbb{N}, \\ & & \forall k \in [1, n-1] \ \alpha_k < \alpha_{k+1} \end{array}$$

Expressing TFDs

Given a relation schema R on attributes $U \cup \{VT\}$, we define the TFD:

$$[E - \mathrm{Exp}(R),\quad t - \mathrm{Group}(i)]\quad X \to Y$$

Temporal Grouping: is a mapping which groups tuples according to the values of their attribute VT.

Granularity $\rightarrow$

$$\cdots\ \underbrace{\overset{1}{\bullet}\ \cdots\ \overset{31}{\bullet}}_{\text{July 2011}}\ \underbrace{\overset{1}{\bullet}\ \cdots\ \overset{31}{\bullet}}_{\text{August 2011}}\ \cdots$$

Sliding windows $\rightarrow$

$$\begin{aligned} t - \mathrm{Group}(i) \quad &\overset{\text{def}}{=} \quad \bigcup_{j=1}^{n} \{(i + \alpha_j)\} \\ \textit{where} \quad & \alpha_1 = 0, \\ & \forall j \in [1, n]\ \alpha_j \in \mathbb{N}, \\ & \forall k \in [1, n-1]\ \alpha_k < \alpha_{k+1} \end{aligned}$$

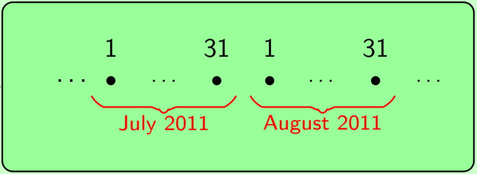
Expressing TFDs

Given a relation schema R on attributes $U \cup \{VT\}$, we define the TFD:

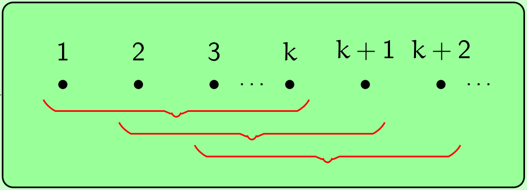
$$[E - Exp(R), \quad t - Group(i)] \quad X \to Y$$

Temporal Grouping: is a mapping which groups tuples according to the values of their attribute VT.

Granularity



Sliding windows



Expressing TFDs:

TFDs Classification

Pure temporally grouping TFDs. $E\text{-}Exp(R)$ returns the given temporal relation r. Tuples are grouped on the basis of $t\text{-}Group$.

Pure temporally evolving TFDs. where the evolution expression associates tuples representing some real world object evolution and there is no temporal grouping (i.e. all tuples are considered together).

Temporally mixed TFDs. First, $E\text{-}Exp(R)$ merges tuples modeling the evolution of a real-world object; then, temporal grouping is applied to the resulting set of tuples.

Temporally hybrid TFDs. First, $E\text{-}Exp(R)$ selects those tuples of the given temporal relation that contribute to the modeling of the evolution of a real-world object, that is, it removes isolated tuples; then, temporal grouping is applied to the resulting set of tuples.

Expressing TFDs: Pure temporally grouping TFDs

The evolution expression simply returns the given temporal relation and tuples are suitably grouped.

$E - Exp$ specifies the original temporal relation r over the schema $U \cup \{VT\}$

Example

1. A patient may take any given drug at most one time per day.

$[Patient, \ \{i\}] \ \ Pat - Id, Drug \rightarrow Chemo, BG, Phys, Qty$

at any time

Expressing TFDs: Pure temporally evolving TFDs

The evolution expression associates tuples representing some real world object evolution and there is no temporal grouping.

E-Exp specifies a relation over a schema $\subseteq \mathcal{U}\overline{\mathcal{U}} \cup \{VT, \overline{VT}\}$

Example

2. for patients undergoing a specific chemotherapy regimen, with the same drug being prescribed on two consecutive days, the quantity of the drug administered on the latter day depends solely on the drug quantity administered on the former.

$[\chi_{Pat-Id,Chemo,Drug}(Patient) \quad Top(i)] \quad Drug, Qty \rightarrow \overline{Qty}$

next view

Expressing TFDs: Temporally mixed TFDs

Temporal grouping is applied together with the evolution expression over a schema $\subseteq \mathtt{U}\overline{\mathtt{U}} \cup \{\mathtt{VT}, \overline{\mathtt{VT}}\}$.

The evolution expression $\mathtt{E\text{-}Exp}$ specifies an evolution relation over a schema $\subseteq \mathtt{U}\overline{\mathtt{U}} \cup \{\mathtt{VT}, \overline{\mathtt{VT}}\}$ by means of join operations on some attribute subset of $\mathtt{U}$, while the temporal grouping groups together tuples according to their $\mathtt{VT}$ values.

Example

3. For any pair of successive administrations of the same drug to the same patient within the same month, the quantity of the second administration uniquely depends on (the drug and) the quantity of the first administration.

$$[\tau_{\mathtt{Pat-Id,Drug}}(\mathtt{Patient}),\quad \mathtt{Month(i)}]\ \ \mathtt{Drug}, \mathtt{Qty} \rightarrow \overline{\mathtt{Qty}}$$

nexttuple view

granularity grouping tuples in the same month

Expressing TFDs: Temporally hybrid TFDs

First, $\mathtt{E\text{-}Exp}(R)$ selects those tuples of the given temporal relation that contribute to the modeling of the evolution of a real-world object, that is, it removes isolated tuples; then, temporal grouping is applied to the resulting set of tuples.

The evolution expression $\mathtt{E\text{-}Exp}$ specifies an evolution relation over the schema $W = U \cup \{VT\}$ while the temporal grouping groups together tuples according to their VT values.

Example

4. For any given chemotherapy, the quantities of drugs taken by patients with successive administrations that take place one 7 days after the other cannot change.

$$[\mathtt{He}(\mathtt{Patient}),\ \ \mathtt{Top}(i)]\ \ \mathtt{Chemo}, \mathtt{Drug} \to \mathtt{Qty}$$

$$\left\{ t \mid \exists t' \left(\begin{array}{c} t' \in \tau_{\mathtt{PatId},\mathtt{Chemo},\mathtt{Drug}}(\mathtt{Patient}) \wedge t'[\overline{VT}] = t'[VT] + 7 \wedge \\ t[\mathtt{PatId}, \mathtt{Chemo}, \mathtt{Drug}] = t'[\mathtt{PatId}, \mathtt{Chemo}, \mathtt{Drug}] \wedge \\ t[VT] = t'[VT] \wedge \\ \left(\begin{array}{c} (t[\mathtt{BG}, \mathtt{Phys}, \mathtt{Qty}] = t'[\mathtt{BG}, \mathtt{Phys}, \mathtt{Qty}] \wedge t[VT] = t'[VT]) \\ \vee \\ (t[\mathtt{BG}, \mathtt{Phys}, \mathtt{Qty}] = t'[\overline{\mathtt{BG}}, \overline{\mathtt{Phys}}, \overline{\mathtt{Qty}}] \wedge t[VT] = t'[\overline{VT}]) \end{array} \right) \end{array} \right) \right.$$

Checking TFDs

The TFD:

$$[\text{E-Exp}(R), \text{t-Group}(i)]X \to Y$$

Can be verified on a relation s by checking the emptiness of the following query:

$$\sigma_{Cnd}(\text{E-Exp}(R) \bowtie_{X=\widehat{X}} \rho_{W \to \widehat{W}} \text{E-Exp}(R))$$

where W is the set of attributes of E-Exp(R) and Cnd stands for $\bigvee_{A \in Y}(A \neq \widehat{A}) \wedge \; SameTGroup$.

The predicate $SameTGroup$ verifies that all the given valid times belong to the same group, according to the expression t-Group.

Discussion on previous proposals

Wijsen TFDs

Tuples are grouped using a given a-priori-fixed binary ordered relation $\otimes \subseteq \{(i, j) \mid i, j \in \mathbb{N} \wedge i \leqslant j\}$

Discussion on previous proposals

Wijsen TFDs

Tuples are grouped using a given a-priori-fixed binary ordered relation $\otimes \subseteq \{(i,j) \mid i,j \in \mathbb{N} \wedge i \leqslant j\}$

$$Emp-Id, Role \rightarrow_T Salary$$

$Emp-Id$	$Role$	$Salary$	$Time$
Al	$Clerk$	1000$	1
...	...	...	...
Al	$Clerk$	1000$	15
...	...	...	...
Al	$Manager$	2000$	28
...	...	...	...

Months

$\otimes$	
i	j
1	2
1	3
...	
1	31
2	3
2	4
...	
2	31
...	

Discussion on previous proposals

Wijsen TFDs

Tuples are grouped using a given a-priori-fixed binary ordered relation $\otimes \subseteq \{(i,j) \mid i,j \in \mathbb{N} \wedge i \leqslant j\}$

$$Emp-Id, Role \rightarrow_T Salary$$

$Emp-Id$	$Role$	$Salary$	$Time$
Al	$Clerk$	1000$	1
...	...	...	...
Al	$Clerk$	1000$	15
...	...	...	...
Al	$Clerk$	2000$	28
...	...	...	...

Months

$\otimes$	
i	j
1	2
1	3
...	
1	31
2	3
2	4
...	
2	31
...	

Discussion on previous proposals

Vianu DFDs

Provide a mapping function which is a partial function from the tuples at time i to the tuples at time $i+1$, DFDs constrain tuples and their "updated" versions.

Discussion on previous proposals

Vianu DFDs

Provide a mapping function which is a partial function from the tuples at time i to the tuples at time $i+1$, DFDs constrain tuples and their “updated” versions.

$$\mathsf{Role}, \mathsf{Salary}, \overline{\mathsf{Role}}, \to_D \overline{\mathsf{Salary}}$$

Emp − Id	Role	Salary
Al	Clerk	1000$
...	...	...
Bob	Clerk	1000$
...	...	...
Sam	Clerk	2000$
...	...	...
Joe	Manager	1100$
...	...	...

Emp − Id	Role	Salary
Al	Manager	1600$
...	...	...
Bob	Manager	1600$
...	...	...
Sam	Manager	2500$
...	...	...
Adam	Clerk	1200$
...	...	...

Discussion on previous proposals

Vianu DFDs

Provide a mapping function which is a partial function from the tuples at time i to the tuples at time $i+1$, DFDs constrain tuples and their "updated" versions.

$$\mathrm{Role}, \mathrm{Salary}, \overline{\mathrm{Role}}, \to_{D} \overline{\mathrm{Salary}}$$

Emp − Id	Role	Salary
Al	Clerk	1000$
...	...	...
Bob	Clerk	1000$
...	...	...
Sam	Clerk	2000$
...	...	...
Joe	Manager	1100$
...	...	...

Emp − Id	Role	Salary
Al	Manager	1600$
...	...	...
Bob	Manager	1800$
...	...	...
Sam	Manager	2500$
...	...	...
Adam	Clerk	1200$
...	...	...

Discussion on previous proposals

		Wijsen's TFDs	Vianu's DFD	Our TFDs
1	"A patient may take any given drug at most one time per day"	✓	✗	✓
2	"For patients undergoing a specific chemotherapy regimen, with the same drug being prescribed on two consecutive days, the quantity of the drug administered on the latter day depends solely on the drug quantity administered on the former"	✗	✓	✓
3	"For any pair of successive administrations of the same drug to the same patient within the same month, the quantity of the second administration uniquely depends on (the drug and) the quantity of the first administration."	✗	✗	✓
4	"For any given chemotherapy, the quantities of drugs taken by patients with successive administrations that take place one 7 days after the other cannot change"	✗	✗	✓

Conclusions

In this paper we have proposed a new general notion of TFD.

We have proved that our proposal can express the previous TFDs proposed in literature, moreover we provide a significative set of constraints that can be captured by our formalism but not by the others.

Further Work:

Using TFDs for normalization;

Deriving new significative TFDs from the existing ones;

Deriving TFDs from given relations (Data Mining).