Metric Temporal Logic for Ontology-Based Data Access over Log Data

D. Calvanese[1], E. Güzel Kalaycı[1], V. Ryzhikov[1], G. Xiao[1] and M. Zakharyaschev[2]

[1] Free University of Bozen-Bolzano, Italy
[2] Birkbeck, University of London, UK

ONTOLP Workshop 2016, New York, USA

July 11, 2016

Temporal OBDA Setting

Raw Log Data	mappings	Conceptualized Temporal Data	ontology	End-User Queries

time	field1	...
...	...	...

$\Rightarrow$

$A(a)@\iota$

$P(a, b)@\iota$

$\Rightarrow$

$q(x_1, \ldots, x_n,$
$\delta_1, \ldots, \delta_m)$

Mappings: Example

Raw Log Data:

station id	time	wind speed
...	...	...
KYHS	13:25	105
KYHS	13:45	120
KYHS	14:30	125
KYHS	14:40	95
...	...	...

mappings:

$\mathrm{HurricaneForceWind}(x)@\langle t_1, t_2\rangle$

SELECT
STATION_ID AS x,
lag(TIME) over (partition by STATION_ID order by TIME) AS t_1,
TIME AS t_2,
"(" AS $\langle$,
"]" AS $\rangle$
FROM Weather
WHERE WIND_SPEED > 118

"wind with the speed greater than 118 km/h is hurricane force"

Conceptualized Temporal Data:

$\mathrm{HurricaneForceWind}(KYHS)@(13{:}25, 13{:}45]$

$\mathrm{HurricaneForceWind}(KYHS)@(13{:}45, 14{:}30]$

Concept $\mathrm{HurricaneForceWind}(x)$ is temporal

Temporal Ontologies

- Declarative language to define temporal concepts in terms of other temporal concepts
- Capture temporal events
- "hurricane is a hurricane force wind lasting one hour or longer"

$$\mathrm{Hurricane}(x) \leftarrow \boxminus_{\geqslant 0}^{\leqslant 1h} \mathrm{HurricaneForceWind}(x)$$

- $\boxminus_{\geqslant 0}^{\leqslant 1h}$ is a metric temporal operator *'during the previous hour'*

$$\begin{array}{l} \mathrm{HurricaneForceWind}(KYHS)@(13{:}25, 13{:}45] \\ \mathrm{HurricaneForceWind}(KYHS)@(13{:}45, 14{:}30] \end{array} \quad \Rightarrow \quad \mathrm{Hurricane}(\mathrm{KYHS})@(14{:}25, 14{:}30]$$

We present temporal knowledge base language datalogMTL

datalogMTL: Syntax

- Predicate symbols $P_0, P_1, \dots$, each of some arity $m \geq 0$
- Data instances $\mathcal{D}$ with $P(\mathbf{c})@\iota$, where $\mathbf{c}$ is an m-tuple of individual constants, and ι an interval of $\mathbb{R}$: $P(\mathbf{c})$ is true at ι
- An ontology $\mathcal{O}$ is a finite set of axioms:

$$\lambda \leftarrow \lambda_1 \wedge \cdots \wedge \lambda_k, \qquad \bot \leftarrow \lambda_1 \wedge \cdots \wedge \lambda_k$$

- Literals: $\lambda \ ::= P(\mathbf{x}) \ \mid \ \mathbf{O}^{\lhd d}_{\rhd e}\lambda$,

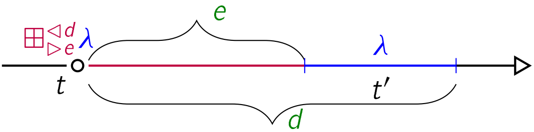
also $(x \neq x')$ and $(x = x')$ not in the head

- Temporal operators $\mathbf{O}^{\lhd d}_{\rhd e}$:
  - $\boxplus^{\lhd d}_{\rhd e}$ (always between e and d in the future),
  - $\boxminus^{\lhd d}_{\rhd e}$ (always between e and d in the past),
  - $\diamondplus^{\lhd d}_{\rhd e}$ (sometime between e and d in the future),
  - $\diamondminus^{\lhd d}_{\rhd e}$ (sometime between e and d in the past),
- $\lhd$ is either $<$ or $\leq$, $\rhd$ is either $>$ or $\geq$.

datalogMTL: Semantics

- Interpretation, $\mathfrak{M}$, is based on the domain Δ of individuals in $\mathcal{D}$ and the temporal domain $\mathbb{R}$
- For any m-ary predicate P, m-tuple $\mathbf{c}$ from Δ and $t \in \mathbb{R}$, $\mathfrak{M}$ specifies whether P is *true on* $\mathbf{c}$ *at* t, we write $\mathfrak{M}, t \models P(\mathbf{c})$
- λ of shape $P(\mathbf{x})$ are interpreted as usual: $\mathfrak{M}, t \models^{\nu} \lambda$ using assignment ν
- $\mathfrak{M}, t \models^{\nu} \boxplus_{\rhd e}^{\lhd d}\lambda$ iff $\mathfrak{M}, t' \models^{\nu} \lambda$

  for all t' such that $t' - t \rhd e$ and $t' - t \lhd d$



- $\mathfrak{M}, t \models^{\nu} \diamondplus_{\rhd e}^{\lhd d}\lambda$ iff $\mathfrak{M}, t' \models^{\nu} \lambda$

  for some t' such that $t' - t \rhd e$ and $t' - t \lhd d$
- Axioms $\lambda \leftarrow \lambda_1 \land \cdots \land \lambda_k$, $\bot \leftarrow \lambda_1 \land \cdots \land \lambda_k$ are interpreted globally (hold at all times)

Queries

- We consider atomic queries $P(\mathbf{x})@\delta$
  - P is a predicate symbol of arity m
  - δ is an interval variable
- Ontology-mediated query $\mathbf{q}(\mathbf{x}, \delta) = (\mathcal{O}, P(\mathbf{x})@\delta)$
- Certain answer to $\mathbf{q}(\mathbf{x}, \delta)$ over $\mathcal{D}$ is $(\mathbf{c}, \iota)$ such that
  - $\mathbf{c} = \nu(\mathbf{x})$ for some ν,
  - $\mathfrak{M} \models (\mathcal{O}, \mathcal{D})$ implies $\mathfrak{M}, t \models P(\mathbf{c})$ for all $t \in \iota$

Example: Weather Ontology

$$\begin{aligned}
\mathrm{Rain}(x) &\leftarrow \mathrm{PositiveTemp}(x) \land \mathrm{Precipitation}(x),\\
\boxminus_{\geqslant 0}^{\leqslant 1h}\mathrm{Hurricane}(x) &\leftarrow \boxminus_{\geqslant 0}^{\leqslant 1h}\mathrm{HurricaneForceWind}(x),\\
\boxminus_{\geqslant 0}^{\leqslant 24h}\mathrm{ExcessiveHeat}(x) &\leftarrow \boxminus_{\geqslant 0}^{\leqslant 24h}\mathrm{TempAbove24}(x) \land\\
&\quad \diamondminus_{\geqslant 0}^{\leqslant 24h}\mathrm{TempAbove41}(x),
\end{aligned}$$

$$\begin{aligned}
\mathrm{HurricaneAffectedCounty}(x) &\leftarrow \mathrm{LocationOf}(x, y) \land \mathrm{Hurricane}(y),\\
\mathrm{SpreadRainCounty}(x) &\leftarrow \mathrm{LocationOf}(x, y) \land \mathrm{LocationOf}(x, z) \land\\
&\quad (y \neq z) \land \mathrm{Rain}(y) \land \mathrm{Rain}(z).
\end{aligned}$$

Queries: $\mathrm{Hurricane}(x)@\delta$, $\mathrm{ExcessiveHeat}(x)@\delta$

Related Work

Consider HornMTL, a propositional fragment of datalogMTL:

$$\lambda \leftarrow \lambda_1 \wedge \cdots \wedge \lambda_k, \qquad \bot \leftarrow \lambda_1 \wedge \cdots \wedge \lambda_k,$$

$$\lambda \ ::= P \ \mid \ \boxplus_{\triangleright e}^{\triangleleft d}\lambda \ \mid \ \boxminus_{\triangleright e}^{\triangleleft d}\lambda \ \mid \ \diamondplus_{\triangleright e}^{\triangleleft d}\lambda \ \mid \ \diamondminus_{\triangleright e}^{\triangleleft d}\lambda$$

Extending to Boolean clauses:

- Metric Temporal Logic
- Modal Logic of Metric Spaces
- Many known decidability+complexity/undecidability results

Our Contribution

- A fragment datalogMTL$^{\Box}_{nr}$ of datalogMTL that is decidable
- $\oplus^{\triangleleft d}_{\triangleright e}\lambda$ and $\diamondminus^{\triangleleft d}_{\triangleright e}\lambda$ do not appear in the head of rules
- Non-recursiveness: $P(\mathbf{x})$ cannot be defined in terms of $P(\mathbf{x})$
- Algorithm for query answering based on temporal joins and coalescing for intervals
- We constructed an SQL query (using WITH clause and RECURSIVE operator) for the weather ontology above
- We ran the query on real-world weather data: termination on reasonable size data in reasonable time

Future Work

- Study complexity of full datalogMTL and it fragments other than datalogMTL$^{\Box}_{nr}$
- Establish connection with works on Metric Temporal Logic and Modal Logic of Metric Spaces
- Consider other practical use-cases: monitoring of engines and devices

$$\begin{aligned}\mathrm{SmoothShutDown} \leftarrow\, & \mathrm{IdleRPM} \wedge \boxminus_{>0}^{\leqslant 15min} \mathrm{IntermRPM} \wedge \\ & \diamondminus_{\geqslant 15min}^{\leqslant 25min} \mathrm{RunningRPM},\\ \mathrm{ConsHighVibration} \leftarrow\, & \boxminus_{\geqslant 0}^{\leqslant 50sec} \diamondminus_{>0}^{\leqslant 10sec} \mathrm{HighVibration}\end{aligned}$$

Algorithm Idea

- Store data $P(c_1, \ldots, c_n)@\langle t_1, t_2 \rangle$ in tables P^* with attributes $c_1, \ldots, c_n, t_1, t_2, \langle, \rangle$
- For, e.g., $P(x) \leftarrow Q(x) \wedge R(x)$ add the tuple $c, t_1'', t_2'', [,]$ to P^* if there are $c, t_1, t_2, [,]$ in Q^* and $c, t_1', t_2', [,]$ in R^*, such that $[t_1, t_2] \cap [t_1', t_2'] \neq \emptyset$, where $[t_1'', t_2''] = [t_1, t_2] \cap [t_1', t_2']$
- For, e.g., $\boxplus_{\geqslant e}^{\leqslant d} P(x) \leftarrow Q(x)$, add the tuple $c, t_1 + e, t_2 + d, [,]$ to P^* if there is a tuple $c, t_1, t_2, [,]$ in Q^*
- For, e.g., $P(x) \leftarrow \boxplus_{\geqslant e}^{\leqslant d} Q(x)$, add the tuple $c, t_1 - e, t_2 - d, [,]$ to P^*, such that $c, t_1, t_2', [,]$ is in Q^*, $c, t_1', t_2, [,]$ is in Q^* and $[t_1, t_2]$ is covered by some set of intervals of shape $[t_1'', t_2'']$ for which $c, t_1'', t_2'', [,]$ is in Q^*