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

D. Calvanese\textsuperscript{1}, E. Güzel Kalaycı\textsuperscript{1}, V. Ryzhikov\textsuperscript{1}, G. Xiao\textsuperscript{1} and M. Zakharyaschev\textsuperscript{2}

\textsuperscript{1} Free University of Bozen-Bolzano, Italy
\textsuperscript{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

<table>
<thead>
<tr>
<th>time</th>
<th>field1</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

⇒

A(a)@t

P(a, b)@t

⇒

q(x₁, ..., xₙ, δ₁, ..., δₘ)

ontology

End-User Queries
Mappings: Example

Raw Log Data:

<table>
<thead>
<tr>
<th>station id</th>
<th>time</th>
<th>wind speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>KYHS</td>
<td>13:25</td>
<td>105</td>
</tr>
<tr>
<td>KYHS</td>
<td>13:45</td>
<td>120</td>
</tr>
<tr>
<td>KYHS</td>
<td>14:30</td>
<td>125</td>
</tr>
<tr>
<td>KYHS</td>
<td>14:40</td>
<td>95</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

mappings:

```
SELECT
    STATION_ID AS x,
    lag(TIME) over (partition
    by STATION_ID order by TIME) AS t1,
    TIME AS t2,
    "(" AS ⟨,
    "]" AS ⟩
FROM Weather
WHERE WIND_SPEED > 118
```

Conceptualized Temporal Data:

HurricaneForceWind(KYHS)@([13:25, 13:45]]

HurricaneForceWind(KYHS)@([13:45, 14:30]]

Concept HurricaneForceWind(x) is temporal

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

\[
\text{Hurricane}(x) \leftarrow \square_{\geq 0}^{1h} \text{HurricaneForceWind}(x)
\]

- \(\square_{\geq 0}^{1h}\) is a metric temporal operator ‘*during the previous hour*’

\[
\text{HurricaneForceWind}(KYHS)@(13:25, 13:45) \Rightarrow \text{Hurricane}(KYHS)@(14:25, 14:30)
\]

\[
\text{HurricaneForceWind}(KYHS)@(13:45, 14:30)
\]

We present temporal knowledge base language **datalogMTL**
datalogMTL: Syntax

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

$$\lambda \leftarrow \lambda_1 \land \cdots \land \lambda_k, \quad \bot \leftarrow \lambda_1 \land \cdots \land \lambda_k$$

- Literals: $\lambda ::= P(x) \mid O_{\triangledown e}^{\triangleleft d} \lambda,$ also $(x \neq x')$ and $(x = x')$ not in the head

- Temporal operators $O_{\triangledown e}^{\triangleleft d}$:
  - $\square_{\triangledown e}^{\triangleleft d}$ (always between $e$ and $d$ in the future),
  - $\square_{\triangledown e}^{\triangleleft d}$ (always between $e$ and $d$ in the past),
  - $\diamond_{\triangledown e}^{\triangleleft d}$ (sometime between $e$ and $d$ in the future),
  - $\diamond_{\triangledown e}^{\triangleleft d}$ (sometime between $e$ and $d$ in the past),

- $\triangleleft$ is either $<$ or $\leq$, $\triangledown$ is either $>$ or $\geq$. 
**datalogMTL: Semantics**

- Interpretation, $\mathcal{M}$, is based on the domain $\Delta$ of individuals in $\mathcal{D}$ and the temporal domain $\mathbb{R}$.
- For any $m$-ary predicate $P$, $m$-tuple $c$ from $\Delta$ and $t \in \mathbb{R}$, $\mathcal{M}$ specifies whether $P$ is *true on $c$ at $t$*, we write $\mathcal{M}, t \models P(c)$.
- $\lambda$ of shape $P(x)$ are interpreted as usual: $\mathcal{M}, t \models^\nu \lambda$ using assignment $\nu$.
- $\mathcal{M}, t \models^\nu \bigotimes e_d \lambda$ iff $\mathcal{M}, t' \models^\nu \lambda$ for all $t'$ such that $t' - t \triangleright e$ and $t' - t < d$.
- $\mathcal{M}, t \models^\nu \bigodot e_d \lambda$ iff $\mathcal{M}, t' \models^\nu \lambda$ for some $t'$ such that $t' - t \triangleright e$ and $t' - t < d$.
- Axioms $\lambda \leftarrow \lambda_1 \wedge \cdots \wedge \lambda_k$, $\bot \leftarrow \lambda_1 \wedge \cdots \wedge \lambda_k$ are interpreted globally (hold at all times).
Queries

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

Rain(x) ← PositiveTemp(x) ∧ Precipitation(x),
□[≤1 h] Hurricane(x) ← □[≤1 h] HurricaneForceWind(x),
□[≤24 h] ExcessiveHeat(x) ← □[≤24 h] TempAbove24(x) ∧
◇[≥0] TempAbove41(x),

HurricaneAffectedCounty(x) ← LocationOf(x, y) ∧ Hurricane(y),
SpreadRainCounty(x) ← LocationOf(x, y) ∧ LocationOf(x, z) ∧
(y ≠ z) ∧ Rain(y) ∧ Rain(z).

Queries: Hurricane(x)@δ, ExcessiveHeat(x)@δ
Consider HornMTL, a propositional fragment of datalogMTL:

\[ \lambda \leftarrow \lambda_1 \land \cdots \land \lambda_k, \quad \bot \leftarrow \lambda_1 \land \cdots \land \lambda_k, \]

\[ \lambda ::= P \mid \Box_e^d \lambda \mid \Diamond_e^d \lambda \mid \Box_e^d \lambda \mid \Diamond_e^d \lambda \]

Extending to Boolean clauses:

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

- A fragment $\text{datalogMTL}_{nr}$ of $\text{datalogMTL}$ that is decidable
- $\bigcirc \bigcirc \bigcirc^d \lambda$ and $\bigcirc \bigcirc \bigcirc^d \lambda$ do not appear in the head of rules
- Non-recursiveness: $P(x)$ cannot be defined in terms of $P(x)$
- 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 $\text{datalogMTL}$ and its fragments other than $\text{datalogMTL}_{nr}$
- Establish connection with works on Metric Temporal Logic and Modal Logic of Metric Spaces
- Consider other practical use-cases: monitoring of engines and devices

\[
\text{SmoothShutDown} \leftarrow \text{IdleRPM} \land \Box^{15\text{min}} > 0 \land \\
\Diamond^{25\text{min}} \geq 15\text{min} \land \\
\Diamond^{10\text{sec}} > 0 \land \\
\text{ConsHighVibration} \leftarrow \Box^{50\text{sec}} \land \\
\Diamond^{10\text{sec}} > 0
\]
Algorithm Idea

- Store data \( P(c_1, \ldots, c_n) @ (t_1, t_2) \) in tables \( P^* \) with attributes \( c_1, \ldots, c_n, t_1, t_2, \langle, \rangle \)

- For, e.g., \( P(x) \leftarrow Q(x) \land R(x) \) 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., \( \forall e \leq d \exists P(x) \leftarrow Q(x) \), 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., \( P(x) \leftarrow \forall e \geq d \exists Q(x) \), 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 \( Q^* \)