Efficient Query Computing for Uncertain Possibilistic Databases with Provenance

Angelos Vasilakopoulos  Verena Kantere

National Technical University of Athens
Cyprus University of Technology

June 20, 2011
The problem we investigate:

How to compute answers to queries for uncertain data with attached “confidence values”?

**Problem 1**
Existing models for uncertain data (e.g., possibility theory) are not closed for SPJ queries.

**Problem 2**
Computing probabilities of SPJ answers in models that combine uncertainty and provenance is a problem with \#P complexity.
We propose:

A data model that combines uncertainty, provenance and possibilities.

**Benefits of the proposed model:**

- Closed for SPJ queries
- Computing possibilities of SPJ answers has PTIME complexity.
Motivating Example

Drives (name, car)

Hank, Mazda : 0.6
Hank, Toyota : 1

Amy, Mazda, Hank : 0.6
ID Q (U)
31

ID
Billy, Mazda, Hank : 0.4
Billy, Lexus : 1

ID
Amy, Mazda, Toyota : 0.8

ID
Drives (name, car)

Hank, Mazda : 0.6
Hank, Toyota : 1

Q_1 = Saw \times Drives

Suspects = \pi_{\text{person}}(Q_1(U))

\lambda(41,1) = ((11,1) \land (21,1)) \lor ((12,1) \land (21,1)) \lor ((11,2) \land (21,2))
**Possible Worlds:**

<table>
<thead>
<tr>
<th>ID</th>
<th>W1</th>
<th>ID</th>
<th>W2</th>
<th>ID</th>
<th>W3</th>
<th>ID</th>
<th>W4</th>
<th>ID</th>
<th>W5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Saw (witness, car)</td>
<td></td>
<td>Saw (witness, car)</td>
<td></td>
<td>Saw (witness, car)</td>
<td></td>
<td>Saw (witness, car)</td>
<td></td>
<td>Saw (witness, car)</td>
</tr>
<tr>
<td>11</td>
<td>Amy, Mazda : 1</td>
<td>11</td>
<td>Amy, Mazda : 1</td>
<td>11</td>
<td>Amy, Toyota : 0.8</td>
<td>11</td>
<td>Amy, Toyota : 0.8</td>
<td>11</td>
<td>Amy, Toyota : 0.8</td>
</tr>
<tr>
<td>12</td>
<td>Billy, Mazda : 0.4</td>
<td>12</td>
<td>Billy, Lexus : 1</td>
<td>12</td>
<td>Billy, Mazda : 0.4</td>
<td>12</td>
<td>Billy, Mazda : 0.4</td>
<td>12</td>
<td>Billy, Lexus : 1</td>
</tr>
<tr>
<td></td>
<td>Drive (name, car)</td>
<td></td>
<td>Drive (name, car)</td>
<td></td>
<td>Drive (name, car)</td>
<td></td>
<td>Drive (name, car)</td>
<td></td>
<td>Drive (name, car)</td>
</tr>
<tr>
<td>21</td>
<td>Hank, Mazda : 0.6</td>
<td>21</td>
<td>Hank, Mazda : 0.6</td>
<td>21</td>
<td>Hank, Mazda : 0.6</td>
<td>21</td>
<td>Hank, Toyota : 1</td>
<td>21</td>
<td>Hank, Toyota : 1</td>
</tr>
<tr>
<td></td>
<td>(\Pi(W1) = 0.4)</td>
<td></td>
<td>(\Pi(W2) = 0.6)</td>
<td></td>
<td>(\Pi(W3) = 0.4)</td>
<td></td>
<td>(\Pi(W4) = 0.4)</td>
<td></td>
<td>(\Pi(W5) = 0.8)</td>
</tr>
</tbody>
</table>

Efficient Query Computing for Uncertain Possibilistic Databases with Provenance
<table>
<thead>
<tr>
<th>ID</th>
<th>Q₁ = Saw ✶ Drives</th>
</tr>
</thead>
<tbody>
<tr>
<td>31</td>
<td>Amy, Mazda, Hank : 0.6</td>
</tr>
<tr>
<td>32</td>
<td>Billy, Mazda, Hank : 0.4</td>
</tr>
<tr>
<td>33</td>
<td>Amy, Toyota, Hank : 0.8</td>
</tr>
</tbody>
</table>

- Tuples 31 and 33 cannot coexist in any possible world.
- Adding provenance (lineage) makes possibilistic uncertain data model closed for SPJ queries.

Computing probabilities is intractable

$\text{Suspects} = \pi_{\text{person}}(Q_1(U))$

\[ \lambda(41, 1) = \{(11, 1) \land (21, 1)\} \lor \{(12, 1) \land (21, 1)\} \lor \{(11, 2) \land (21, 2)\} \]

ID | Suspects
---|---
41 | Hank

Axioms of Possibilistic theory:

- \( \Pi(X) = 1 \)
- \( \Pi(\emptyset) = 0 \)
- \( \Pi(E_1 \cup E_2) = \max(\Pi(E_1), \Pi(E_2)) \)
- \( \Pi(E_1 \cap E_2) \leq \min(\Pi(E_1), \Pi(E_2)) \)
- \( \Pi(E_1 \cap E_2) = \min(\Pi(E_1), \Pi(E_2)) \)
  (for not-interactive events)
- \( \max\{\Pi(E), \Pi(\bar{E})\} = 1 \)
- \( N(E) = 1 - \Pi(\bar{E}) \)
Running Example

\[ Q_1 = \text{saw} \land \text{drives} \]

\[ \text{Suspects} = \pi_{\text{person}}(Q_1(U)) \]

\[ \lambda(41,1) = \{(11,1) \land (21,1)) \lor ((12,1) \land (21,1)) \lor ((11,2) \land (21,2)) \} \]
Possible Worlds:

- Π(W1) = 0.4, N(W1) = 0
- Π(W2) = 0.6, N(W2) = 0
- Π(W3) = 0.4, N(W3) = 0
- Π(W4) = 0.4, N(W4) = 0
- Π(W5) = 0.8, N(W5) = 0

Efficient Query Computing for Uncertain Possibilistic Databases with Provenance
Ongoing Work:

- Extend the query language to extentions of conjunctive queries.
- Find for which class of query languages the problem remains in polynomial time.
- Find for which class of query languages the problem becomes intractable.
- Study complexity of new query languages that can query over uncertainty and provenance.
Thank you