Modeling and Querying Evidential Databases: First Steps towards a Strong Representation System

Fatma Ezzahra Bousnina
Mouna Chebbah    Mohamed Anis Bach Tobji    Allel HadjAli
Boutheina Ben Yaghlane

May 18th, 2017
Outline

1. Context and Background Materials
   - Typology of Imperfection
   - Theory of Evidence and Evidential Databases

2. Modeling Evidential Databases
   - Evidential Databases as Possible Worlds
   - Implementation of the Evidential Database Model

3. Querying Evidential Databases
   - Extended Relational Operators
   - Evidential Top-k Queries

4. Conclusion and Future works
Outline

1. Context and Background Materials
   - Typology of Imperfection
   - Theory of Evidence and Evidential Databases

2. Modeling Evidential Databases
   - Evidential Databases as Possible Worlds
   - Implementation of the Evidential Database Model

3. Querying Evidential Databases
   - Extended Relational Operators
   - Evidential Top-k Queries

4. Conclusion and Future works
Typology of Imperfection

Imperfect Information

- Imperfection
  - Uncertainty
  - Imprecision
Imperfect Information

Imperfection

- Uncertainty
- Imprecision

Example

- John has at least two children and I am not sure about it
  ⇒ Imprecise and Uncertain
Typology of Imperfection

**Imperfect Information**

- **Imperfection**
  - Uncertainty
  - Imprecision

- **Example**
  - John has at least two children and I am not sure about it
    - Imprecise and Uncertain
  - John has at least two children and I am sure about it
    - Imprecise and Certain
### Imperfect Information

#### Imperfection
- Uncertainty
- Imprecision

#### Example
- John has at least two children and I am not sure about it
  \[\Rightarrow\text{Imprecise and Uncertain}\]
- John has at least two children and I am sure about it
  \[\Rightarrow\text{Imprecise and Certain}\]
- John has three children and I am not sure about it
  \[\Rightarrow\text{Precise and Uncertain}\]
Imperfect Information

Imperfection

- Uncertainty
- Imprecision

Example

- John has at least two children and I am not sure about it
  ⇒ Imprecise and Uncertain

- John has at least two children and I am sure about it
  ⇒ Imprecise and Certain

- John has three children and I am not sure about it
  ⇒ Precise and Uncertain

- John has three children and I am sure about it
  ⇒ Precise and Certain
Theories and database models of Imperfection

- Probability Theory
  - Laplace, 1812
- Possibility Theory
  - Zadeh, 1978
- Fuzzy sets Theory
  - Zadeh, 1965
- Evidence Theory
  - Dempster, 1967; Shafer, 1976
Theories and database models of Imperfection

- Probability Theory
  [Laplace, 1812]
Theories and database models of Imperfection

- Probability Theory
  [Laplace, 1812]
- Possibility Theory
  [Zadeh, 1978]
- Fuzzy sets Theory
  [Zadeh, 1965]
- Evidence Theory
  [Dempster, 1967; Shafer, 1976]
### Typology of Imperfection

(R.Cavallo, M.Pittarelli, 1987)

<table>
<thead>
<tr>
<th>ID</th>
<th>Weather</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>rainy</td>
<td>0.6</td>
</tr>
<tr>
<td>2</td>
<td>sunny</td>
<td>0.4</td>
</tr>
</tbody>
</table>

Probabilistic Databases
Possibilistic Databases

<table>
<thead>
<tr>
<th>ID</th>
<th>Weather</th>
<th>Possibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>rainy</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>sunny</td>
<td>0.4</td>
</tr>
</tbody>
</table>

[H.Prade, C.testemale, 1984]
Typology of Imperfection

[S.K. Lee, 1992]
Theory of Evidence

- Theory of evidence is also called theory of belief functions or Dempster-Shafer theory.
Theory of Evidence

- **Theory of evidence** is also called **theory of belief functions** or **Dempster-Shafer theory**.
Theory of Evidence

- **Theory of evidence** is also called theory of belief functions or Dempster-Shafer theory.
- Popularized by Smets in 1988 with the Transferable Belief Model (TBM).
Theory of Evidence

- Theory of evidence is also called theory of belief functions or Dempster-Shafer theory.
- Popularized by Smets in 1988 with the Transferable Belief Model (TBM).
- Theory of evidence provides an explicit representation of uncertainty, imprecision and inconsistency.
Evidential Databases

An *Evidential Database* (EDB) on its *compact form* has $N$ tuples and $D$ attributes. The value of an attribute $c$ for an object $l$ is called *evidential value*, $V_{lc}$.

<table>
<thead>
<tr>
<th>ID</th>
<th>Disease</th>
<th>Symptom</th>
<th>CL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Diabetes 1</td>
<td>Fatigue 0.4</td>
<td>[0.3 ; 0.9]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>{Fatigue, Nausea}</td>
<td>0.6</td>
</tr>
</tbody>
</table>
Evidential Databases

Mass Function

A mass function, $m$, is a mapping from $2^\Theta$ to $[0;1]$. The basic belief mass of an hypothesis $x$, $m(x)$ represents the degree of truth about that hypothesis $x$ such that:

$$\sum_{x \subseteq \Theta} m^\Theta(x) = 1$$

<table>
<thead>
<tr>
<th>ID</th>
<th>Disease</th>
<th>Symptom</th>
<th>CL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Diabetes</td>
<td>Fatigue 0.4, Nausea 0.6</td>
<td>[0.3 ; 0.9]</td>
</tr>
</tbody>
</table>
Confidence Level

A confidence level $CL$ is a measure that quantifies the degree of belief $bel$ and plausibility $pl$ about the existence of each tuple in the database, $CL = \{[bel, pl]| bel, pl \in [0, 1]; bel \leq pl\}$.

<table>
<thead>
<tr>
<th>ID</th>
<th>Disease</th>
<th>Symptom</th>
<th>CL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Diabetes 1</td>
<td>Fatigue 0.4</td>
<td>[0.3 ; 0.9]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>{Fatigue, Nausea}</td>
<td>0.6</td>
</tr>
</tbody>
</table>
Belief Function

The belief function, \( \text{bel} \), is the minimal degree of belief given to an hypothesis \( x \).

\[
\text{bel}(x) = \sum_{y, x \subseteq \Theta : y \subseteq x} m(y)
\]
Plausibility Function

The plausibility function, $pl$, is the maximal amount of belief on the hypothesis $x$.

$$pl(x) = \sum_{y, x \subseteq \Theta: x \cap y \neq \emptyset} m(y)$$
Outline

1. Context and Background Materials
   - Typology of Imperfection
   - Theory of Evidence and Evidential Databases

2. Modeling Evidential Databases
   - Evidential Databases as Possible Worlds
   - Implementation of the Evidential Database Model

3. Querying Evidential Databases
   - Extended Relational Operators
   - Evidential Top-k Queries

4. Conclusion and Future works
Modeling and Querying Evidential Databases: First Steps towards a Strong Representation System

- Modeling Evidential Databases
- Evidential Databases as Possible Worlds

![Diagram of Imperfect Database and Candidates](image-url)
Modeling and Querying Evidential Databases: First Steps towards a Strong Representation System

Modeling Evidential Databases

Evidential Databases as Possible Worlds

Probabilistic Database

Probabilistic Candidate 1

Probabilistic Candidate 2

Probabilistic Candidate 3

[Abiteboul, Kanellakis and Grahne, 1991]
Modeling Evidential Databases

Evidential Databases as Possible Worlds

[Bosc, Liétard and Pivert, 2002]
Evidential Databases as Possible Worlds
### EDB as Possible Worlds

<table>
<thead>
<tr>
<th>ID</th>
<th>Disease</th>
<th>BloodType</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Anemia 1</td>
<td>B 0.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>{B, O} 0.7</td>
</tr>
<tr>
<td>2</td>
<td>{Asthma, Flu}</td>
<td>A 1</td>
</tr>
</tbody>
</table>
## Generating Possible Worlds from an EDB: Example

<table>
<thead>
<tr>
<th>ID</th>
<th>Disease</th>
<th>BloodType</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Anemia</td>
<td>B</td>
</tr>
<tr>
<td>2</td>
<td>Asthma</td>
<td>A</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ID</th>
<th>Disease</th>
<th>BloodType</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Anemia</td>
<td>O</td>
</tr>
<tr>
<td>2</td>
<td>Asthma</td>
<td>A</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ID</th>
<th>Disease</th>
<th>BloodType</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Anemia</td>
<td>B</td>
</tr>
<tr>
<td>2</td>
<td>Flu</td>
<td>A</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ID</th>
<th>Disease</th>
<th>BloodType</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Anemia</td>
<td>O</td>
</tr>
<tr>
<td>2</td>
<td>Flu</td>
<td>A</td>
</tr>
</tbody>
</table>

\[ m(\{W_1; W_2\}) = 0.3 \quad m(\{W_1; W_2; W_3; W_4\}) = 0.7 \]
The non existence of an implementation of an Evidential Database in the literature
Evidential Database Design
Implementation

**SQL3 implementation:**
- Facilitate the complex structure of an evidential database
- Optimize the I/O cost
- Accelerate the information extraction thanks to the indexes
### Table: Contribution of caches to queries re-execution

<table>
<thead>
<tr>
<th>Database size</th>
<th>First execution time (s)</th>
<th>Next executions’ times (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>0.2</td>
<td>0.03</td>
</tr>
<tr>
<td>5000</td>
<td>0.8</td>
<td>0.04</td>
</tr>
<tr>
<td>10 000</td>
<td>4.1</td>
<td>0.06</td>
</tr>
<tr>
<td>50 000</td>
<td>5.8</td>
<td>0.12</td>
</tr>
<tr>
<td>70 000</td>
<td>6.4</td>
<td>0.16</td>
</tr>
<tr>
<td>100 000</td>
<td>17</td>
<td>1.2</td>
</tr>
</tbody>
</table>
Outline

1. Context and Background Materials
   - Typology of Imperfection
   - Theory of Evidence and Evidential Databases

2. Modeling Evidential Databases
   - Evidential Databases as Possible Worlds
   - Implementation of the Evidential Database Model

3. Querying Evidential Databases
   - Extended Relational Operators
   - Evidential Top-k Queries

4. Conclusion and Future works
The compact form is the only feasible model in practice but the non-compact form is fundamental to prove if the model is a **Strong Representation System**.

**Strong Representation System**

A model is a *strong representation system* (SRS), when the result of querying the *compact form* is *equivalent* to the result of querying the set of its *possible worlds* (the non-compact form).
Compact Form

Non Compact Form

$W_1$

$W_j$

EDB
Compact Form

Non Compact Form

$W_1$

$Q_{op}$

$R_1$

$W_j$

$Q_{op}$

$R_u$

$Q_{op}$
Definition

Let $Q$ be the query processed on each possible world $W_j$. Querying each possible world $W_j$, gives a possible answer $R = Q(W_j)$.
Querying Possible Worlds

**Definition**
Let $Q$ be the query processed on each possible world $W_j$. Querying each possible world $W_j$, gives a possible answer $R = Q(W_j)$

**Confidence Level**
For each tuple $t$ belonging to the set of responses $Q(W_u)$, a confidence level $CL$ is calculated. $t.CL = [\text{bel}(S), \text{pl}(S)]; S = \{W_u | t \in Q(W_u)\}$
Querying Possible Worlds

**Definition**

Let $Q$ be the query processed on each possible world $W_j$. Querying each possible world $W_j$, gives a possible answer $R = Q(W_j)$

**Confidence Level**

For each tuple $t$ belonging to the set of responses $Q(W_u)$, a confidence level $CL$ is calculated. $t.CL = [\text{bel}(S), \text{pl}(S)]; S = \{W_u | t \in Q(W_u)\}$
Modeling and Querying Evidential Databases: First Steps towards a Strong Representation System

Evidential Top-$k$ Queries

Evidential Database

Result 1
[0.2, 1]

Result 2
[0.1, 0.9]

Result 3
[0.15, 0.8]

Result n
[0.25, 0.85]
Which are the $k$ best results?

How to rank them?

![Diagram showing ranked results]

- Result 1: [0.2 1]
- Result 2: [0.1 0.9]
- Result 3: [0.15 0.8]
- Result n: [0.25 0.85]
Top-

k queries also known as Ranking queries.
- **Top-k queries** also known as **Ranking queries**.

- Introduced in the multimedia systems by Fagin in 1996.
- **Top-k queries** also known as **Ranking queries**.

- Introduced in the multimedia systems by Fagin in 1996.

- **Order** queries’ results in order to give only the most interesting answers.
- **Top-k** queries also known as **Ranking queries**.

- Introduced in the multimedia systems by Fagin in 1996.

- **Order** queries’ results in order to give only the most interesting answers.

- **Top-k** queries order answers using a **score function** that returns the $k$ important answers with the highest scores.
Querying Evidential Databases

Evidential Top-k Queries

May 18\textsuperscript{th}, 2017
Modeling and Querying Evidential Databases: First Steps towards a Strong Representation System

Querying Evidential Databases

Evidential Top-\(k\) Queries
Modeling and Querying Evidential Databases: First Steps towards a Strong Representation System

Evidential Top-$k$ Queries

Querying Evidential Databases

Fatma Ezzahra Bousnina Mouna Chebbah

Modeling and Querying Evidential Database

May 18th, 2017
Modeling and Querying Evidential Databases: First Steps towards a Strong Representation System

Querying Evidential Databases

Evidential Top-\(k\) Queries

-May 18th, 2017 24 / 30
<table>
<thead>
<tr>
<th>ID</th>
<th>BookRate</th>
<th>CL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$b_1$ 0.3</td>
<td>[0.5 ; 1]</td>
</tr>
<tr>
<td></td>
<td>${b_2, b_3}$ 0.7</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>$b_2$ 0.5</td>
<td>[0.3 ; 0.8]</td>
</tr>
<tr>
<td></td>
<td>$b_4$ 0.5</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>${b_1, b_2, b_3}$ 1</td>
<td>[1 ; 1]</td>
</tr>
<tr>
<td>4</td>
<td>$b_3$ 1</td>
<td>[0.5 ; 0.9]</td>
</tr>
</tbody>
</table>

**Top-k Query**

The top-2 most appreciated books for readers?
Introducing Evidential Score

<table>
<thead>
<tr>
<th>item</th>
<th>Evidential Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b_1$</td>
<td>$R_1 = [0.0375; 0.325]$</td>
</tr>
<tr>
<td>$b_2$</td>
<td>$R_2 = [0.0375; 0.525]$</td>
</tr>
<tr>
<td>$b_3$</td>
<td>$R_3 = [0.125; 0.65]$</td>
</tr>
<tr>
<td>$b_4$</td>
<td>$R_4 = [0.0375; 0.1]$</td>
</tr>
</tbody>
</table>
- Adopted Preference Degree Equation

<table>
<thead>
<tr>
<th>item</th>
<th>EvidentialScore</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b_1$</td>
<td>$R_1 = [0.0375 ; 0.325]$</td>
</tr>
<tr>
<td>$b_2$</td>
<td>$R_2 = [0.0375 ; 0.525]$</td>
</tr>
<tr>
<td>$b_3$</td>
<td>$R_3 = [0.125 ; 0.65]$</td>
</tr>
<tr>
<td>$b_4$</td>
<td>$R_4 = [0.0375 ; 0.1]$</td>
</tr>
</tbody>
</table>

Final Ranking

$b_3 \succ b_2 \succ b_1 \succ b_4$.

The Top-2 appreciated books are:

- $b_3$ with a confidence level $[0.125 ; 0.65]$
- $b_2$ with a confidence level $[0.0375 ; 0.525]$
Defined A new Semantic for Imperfect Databases

Etop-2 Semantics

Books $b_3$ and $b_2$ are the 2 most appreciated credible answers from the set of results.
Outline

1. Context and Background Materials
   - Typology of Imperfection
   - Theory of Evidence and Evidential Databases

2. Modeling Evidential Databases
   - Evidential Databases as Possible Worlds
   - Implementation of the Evidential Database Model

3. Querying Evidential Databases
   - Extended Relational Operators
   - Evidential Top-k Queries

4. Conclusion and Future works
We introduced the possible worlds’ model of an Evidential Database.
Conclusion

- We introduced the possible worlds' model of an Evidential Database.
- We implemented the Evidential Database model on its compact form.
Conclusion

- We introduced the possible worlds’ model of an Evidential Database.
- We implemented the Evidential Database model on its compact form.
- We introduced two extended relational evidential queries: Selection and Projection.
Conclusion

- We introduced the possible worlds’ model of an Evidential Database.
- We implemented the Evidential Database model on its compact form.
- We introduced two extended relational evidential queries: Selection and Projection.
- We introduced a new imperfect top-k query, called Evidential Top-k Query.
Future Works

- Implementation of Etop-k queries.
Future Works

- Proof of the Strong Representation System for Evidential Databases.

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>1</td>
</tr>
<tr>
<td>$b_1$</td>
<td>0.8</td>
</tr>
<tr>
<td>${a_1, a_2}$</td>
<td>1</td>
</tr>
<tr>
<td>$b_2$</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>$b_1$</td>
</tr>
<tr>
<td>$a_1$</td>
<td>$b_2$</td>
</tr>
<tr>
<td>$a_2$</td>
<td>$b_2$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>$b_2$</td>
</tr>
<tr>
<td>$a_1$</td>
<td>$b_2$</td>
</tr>
<tr>
<td>$a_2$</td>
<td>$b_2$</td>
</tr>
</tbody>
</table>


Thank You