More and more uncertain data, for example:
- Sensor networks
- Information extraction
- Data integration

Positive relational algebra queries well studied on uncertain databases, but little attention on:
- Updates
- Set Difference

Why?

We study the problem of updating uncertain and probabilistic databases in the form of U-relations.

Four research questions:
- How can we model updates on U-relations?
- How can we model set difference on U-relations?
- How can we optimize U-relations?
- How can we perform in practice?

What?

Updates

We need set difference

Positive relational algebra is not sufficient to model arbitrary updates. We reduce the problem of updating a U-relation to computing set difference.

\[ A \setminus B = A \cap \overline{B} \]

To compute set difference we use the inverse \( B \).

Inverse

We present two algorithms to compute the inverse:
- Decompression
- Negation

Optimization

\( \Sigma_2 \)-complete

We prove that there is no efficient algorithm that shrinks a U-relation to the minimally necessary size.

Heuristics

We present two tractable optimization heuristics:
- Subset elimination
- Merging

P-Time

Set difference is hard in general, but with “skipping” we get polynomial time on specific instances.

Better

We define translations of relational algebra queries for both. Both keep the advantages of U-relations, while giving an exponential benefit for set difference, and thus for updates. We prove that U-relations always need less space than U-relations.

New Representations

Before

U-relations use formulas in DNF over equalities between variables and constants. We present two extensions.

\( U \)-relations

Additionally allow inequalities.

\( U^{\text{int}} \)-relations

Additionally allow intervals.

Experimental Results

We have extended the uncertain database system MayBMS. The experimental evaluation shows that:
- Our approach is feasible in practice and scales well with the database size.
- \( U \)-relations and \( U^{\text{int}} \)-relations are a clear improvement.

New Representations of Uncertain Databases for Efficient Updates

Andreas Hubmer

Technische Universität Wien
Institut für Informationssysteme
Arbeitsbereich: Databases and Artificial Intelligence
Betreuer: Univ.-Prof. Dr. Reinhard Pichler
Mitwirkung: Projass. Vadim Savenkov, MSc
Projass. Dipl.-Ing. Sebastian Skritek

Kontakt: andreas.hubmer@alumni.tuwien.ac.at