University of Zurich[UZH]

Department of Informatics

University of Zürich
Department of Informatics
Binzmühlestr. 14
CH-8050 Zürich
Phone. +41 44 635 43 11
Fax +41 44 635 68 09
www.ifi.uzh.ch/dbtg

UZH, Dept. of Informatics, Binzmühlestr. 14, CH-8050 Zürich

Andrin Betschart

Prof. Dr. Michael Böhlen
Professor
Phone +41 44 635 43 33
Fax +41 44 635 68 09
boehlen@ifi.uzh.ch

Zürich, November 4, 2013

MSc Thesis

Topic: Exact and Approximate Confidence Computation in Temporal Probabilistic Databases

Each tuple in a temporal probabilistic database corresponds to an atomic event and is associated with a random variable with value in the domain {true, false}. Lineage is a propositional formula constructed from atomic events combined with binary operations $\wedge$ *(logical and)*, $\vee$ *(logical or)*. It is used to indicate the way the result tuples have been produced from the argument tuples, during algebra operations.

The goal of this project is to efficiently implement and evaluate methods for computation of the probability value corresponding to a propositional formula Φ over discrete random variables. Exact computation methods will be studied as well as approximate computations within an allowed error ϵ. [1, 2, 3, 4, 5, 6]

Tasks

1. Efficient representation of the lineage formula of each tuple
2. Adaptation of the algebra operations according to the lineage representation
3. Computation of confidence values using lineage through one exact and one approximate method
4. Experimental evaluation and comparison of both methods
5. Written thesis (approximately 50 pages)
6. 25-minute Presentation of the results in a group meeting

References

[1] Maximilian Dylla, Iris Miliaraki, and Martin Theobald. A temporal-probabilistic database model for information extraction. In *VLDB*, 2014.

[2] Robert Fink, Dan Olteanu, and Swaroop Rath. Providing support for full relational algebra in probabilistic databases. In *ICDE*, 2011.

[3] Christoph Koch and Dan Olteanu. Conditioning probabilistic databases. 2008.

[4] Dan Olteanu and Jiewen Huang. Secondary-storage confidence computation for conjunctive queries with inequalities. SIGMOD, 2009.

[5] Dan Olteanu, Jiewen Huang, and Christoph Koch. Sprout: Lazy vs. eager query plans for tuple-independent probabilistic databases. In *ICDE*, 2009.

[6] Dan Olteanu, Jiewen Huang, and Christoph Koch. Approximate confidence computation in probabilistic databases. In *ICDE*, 2010.

Supervisor: Prof. Dr. Michael Böhlen
Start date: 04-11-2013
End date: 03-05-2014

University of Zürich
Department of Informatics

Prof. Dr. Michael Böhlen