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

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

Zürich, February 27, 2014

MSc Project

Topic: Extending PostgreSQL to evaluate queries in Temporal Probabilistic Databases

Each tuple in a temporal probabilistic database corresponds to an atomic event. It is associated with a random variable, with value in the domain $\{$true, false$\}$, and with a confidence value in the domain (0,1], indicating how possible it is for the event to be true. The following tables correspond to the vacation plans of different people in March and the expected weather conditions at the time of their visit. For example, according to tuple p_1, there is 0.3 probability that Jim will be in Zurich at a specific time point between the 3rd and the 8th of March.

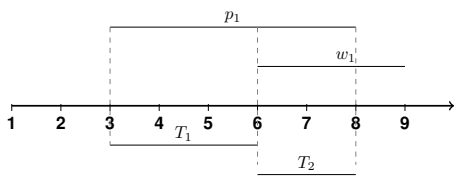
P (People)

	$Name$	$Dest$	T	p
p_1	Jim	ZH	[3, 8)	0.30
p_2	Ann	BE	[5, 10)	0.20

W (Weather)

	Loc	$Weather$	T	p
w_1	ZH	Snow	[6, 9)	0.80

Lineage is a propositional formula constructed from atomic events combined with binary operations $\wedge$ *(logical and)*, $\vee$ *(logical or)*. First of all, it is used to adapt the intervals of the tuples during temporal adjustment since we group time points which are expected to have the same lineage. The temporal aligner and the temporal splitter are the two temporal primitives defined to provide efficient modification of the intervals of a tuple based on the tuples it's expected to be combined with during an operation. In the following figure, we see the necessary temporal adjustment of relation **P** in order to be later joined with relation **W** based on the condition *Dest = Loc*.



Lineage is also used to indicate the way the result tuples have been produced from the argu-

ment tuples, during algebra operations, and thus as a means to compute the probability of a tuple. e.g. The lineage of tuple r_1 indicates that it will be true only if tuples p_1 and w_1 are both true. Its probability is computed considering that P($p_1 \wedge w_1$) = $P(p_1) \cdot P(w_1) = 0.3 \cdot 0.2 = 0.06$

Result

	$Name$	$Dest$	Loc	$Weather$	T	p	
r_1	Jim	ZH	ZH	Snow	[6, 9)	0.06	$p_1 \wedge w_1$

The goal of this master project is the understanding of the query evaluation and confidence computation process in temporal probabilistic databases and its efficient implementation in the kernel of PostgreSQL so that it offers native support.

Tasks

1. Adaptation of the temporal primitives [1, 2] to capture the notion of lineage.
2. Extension of the conventional operators so as to correctly compute the lineage of the result tuples.
3. Query execution without the need to be aware of the rules needed to reduce a temporal-probabilistic operation into a conventional one
4. Confidence computation based on lineage by using one exact and one approximate computation method [3, 4, 5, 6]
5. Generation of an appropriate dataset including information on time and probability
6. Experimental evaluation of both the lineage and confidence computation
7. Technical Report (20 pages)

References

[1] Anton Dignös, Michael H. Böhlen, and Johann Gamper. Temporal alignment. In *SIGMOD Conference*, pages 433–444, 2012.

[2] Anton Dignös, Michael H. Böhlen, and Johann Gamper. Query time scaling of attribute values in interval timestamped databases. In *ICDE*, 2013.

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

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

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

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

Supervisor: Prof. Dr. Michael Böhlen
Start date:
End date: