An Extendable Framework for Managing Uncertain Spatio-Temporal Data

Tobias Emrich, Maximilian Franzke, Hans-Peter Kriegel, Johannes Niedermayer, Matthias Renz, Andreas Züfle
Institute for Informatics, Ludwig-Maximilians-Universität München
{emrich,franzke,kriegel,niedermayer,renz,zuefle}@dbs.ifi.lmu.de

ABSTRACT

This demonstration presents our Uncertain-Spatio-Temporal (UST) framework that we have developed in recent years. The framework allows not only to visualize and explore spatio-temporal data consisting of (location, time, object)-triples but also provides an extensive codebase easily extensible and customizable by developers and researchers. The main research focus of this UST-framework is the explicit consideration of uncertainty, an aspect that is inherent in spatio-temporal data, due to infrequent position updates, due to physical limitations and due to power constraints. The UST-framework can be used to obtain a deeper intuition of the quality of spatio-temporal data models. Such models aim at estimating the position of a spatio-temporal object at a time where the object's position is not explicitly known, for example by using both historic (traffic-) pattern information, and by using explicit observations of objects. The UST-framework illustrates the resulting distributions by allowing a user to move forward and backward in time. Additionally the framework allows users to specify simple spatio-temporal queries, such as spatio-temporal window queries and spatio-temporal nearest neighbor (NN) queries. Based on recently published theoretic concepts, the UST-framework allows to visually explore the impact of different models and parameters on spatio-temporal data. The main result showcased by the UST-framework is a minimization of uncertainty by employing stochastic processes, leading to small expected distances between ground truth trajectories and modelled positions.

Categories and Subject Descriptors

H.3.3 [**INFORMATION STORAGE AND RETRIEVAL**]: Information Search and Retrieval—*Query Processing*

Keywords

Uncertain Trajectories, Uncertain Spatio-Temporal, Framework

1. INTRODUCTION

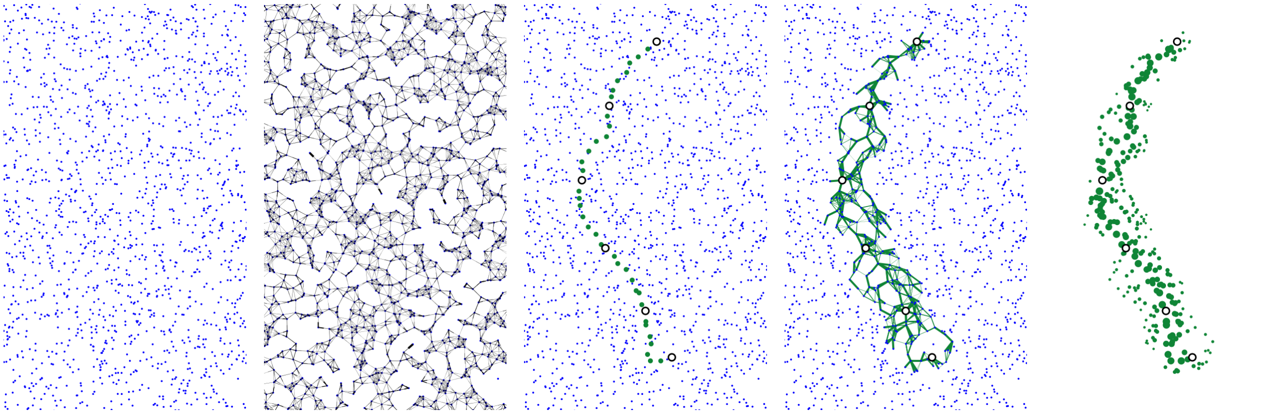
Both the current trends in technology such as smartphones, general mobile devices, stationary sensors and satellites as well as a new user mentality of utilizing this technology to voluntarily share information produce a huge flood of geo-spatio-temporal data. The main goal of this demonstration is to provide an intuitive yet powerful framework to evaluate and explore techniques for managing and querying historical spatio-temporal data in explicit consideration of the inherent uncertainty. By *historical spatio-temporal data* ([5]), we refer to a collection of data triples having the form (location, time, object). In a plethora of research fields and industrial applications, these techniques can substantially improve decision making, minimize risk and unearth valuable insights that would otherwise remain hidden. In practice, such triples correspond to observations (e.g. GPS positions, RFID signals, or visual observations) of *moving objects* in the past. Thus, each triple corresponds to an observation of the *trajectory* of a moving object.

In many applications, the position of an object is observed at discrete times only, leading to an inherent uncertainty between these discrete times and creating the notion of *uncertain spatio-temporal data* – an aspect raising an imminent need for scalable and flexible data management. In practice, moving objects for which the position at any time cannot be determined deterministically, are denoted as *uncertain moving objects*. The indeterministic trajectory of an uncertain moving object is an *uncertain trajectory*. Existing works to model, manage and query uncertain spatio-temporal data, including interpolation models ([8, 9, 10, 12, 11]) and geometric approximation models ([14, 15, 16, 6, 4, 13]), identify trajectories that may possibly satisfy a user-specified query predicate. These approaches cannot return to the user an indication of the likelihood of these trajectories to satisfy the query predicate, i.e., these approaches cannot give the user any information about the quality and significance of the returned result. First approaches to alleviate this problem have been proposed in [6, 18, 1, 17]. These approaches return uncertain moving objects associated with a probability estimating the likelihood of the corresponding object to satisfy the query predicate. They treat realizations of uncertain moving objects at different points of time as mutually independent random variables. However, the location of objects at subsequent discrete points of time are highly correlated, as commanded for example by laws of physics and speed limits. Ignoring this positive correlation leads to systematic and significant errors as shown in [7].

The shortcomings of these models have been addressed in previous work ([3, 2, 7]), by applying models from statistics, namely stochastic processes and Bayesian inference, to treat uncertain moving objects in a probabilistic way. In the following Section 2, the theoretic base of these works is sketched. Section 3 describes the system architecture, functionality and the main purpose of the UST-framework. Finally, Section 4 describes the demonstrator, by describing the impact on the field of managing spatio-temporal data, and explaining how the framework's visualization tools will be used to convince the audience of this impact.

SIGMOD'14, June 22–27, 2014, Snowbird, UT, USA.

http://dx.doi.org/10.1145/2588555.2594523.



(a) State space (b) A-priori transitions (c) Single trajectory (d) Transition probabil. (e) State probabilities

Figure 1: Data Visualization

2. PRELIMINARIES

A main challenge of uncertain spatio-temporal data is to infer the position of an uncertain object o at time t when the exact position has not been observed. The quality of this inference depends on our ability to effectively use the wealth of information stored in a spatio-temporal database. Firstly, we need to consider information about movement patterns, such as turn probabilities at a road intersection, learned empirically from historical data. Secondly, observations of o that may temporally precede or succeed t need to be considered to predict the position of o at time t.

The key idea of the demonstrated approach to model uncertain spatio-temporal data is to model possible object trajectories by a stochastic process, more precisely a Markov chain. Employing the Markov chain model for representing spatio-temporal data has three major advantages over previous work:

1. It allows answering queries such that results are associated with their corresponding probabilities.
2. Dependencies between object locations at consecutive points in time are taken into account.
3. It is often possible to express queries using simple matrix operations only, thus allowing to utilize highly efficient existing solutions for interactive query processing.

The Markov model is estimated empirically. The resulting model is called *a-priori model* and contains general transition probabilities; i.e., probabilities of any object moving from one node to an adjacent node in the network. However, using the a-priori model, i.e., the usage of a Markov random-walk model yields unacceptable inference quality in applications such as traffic monitoring, where objects aim at moving on a more-or-less shortest path, rather than randomly selecting a new direction at each intersection of a road network.

To solve this problem, the main challenge that we approached in our most recent publication ([7]) is to enrich these general movement patterns given by the Markov model by knowledge of individual observations of an uncertain moving object. We pre-compute the probabilities that, at a given time, a given object performs a transition between two nodes of the network, given all observations of an object, at past, present and future times. These adapted transition probabilities define an *a-posteriori* model, which combines all sources of information given by a spatio-temporal database. A main experimental result of this work shows that the adapted a-posteriori model, i.e., the fusion between empirical trajectory data and observation data of objects, allows to accurately estimate the position of an uncertain moving object at times between observations. In particular, the expected error in terms of distance between the real trajectory and the modelled position (which is a random variable) was shown to be vastly reduced by this data integration. For a detailt evaluation of the runtime of queries under the Markov model we refer to the corresponding papers [3, 2, 7].

Figure 1(a) visualizes the set of spatial states used for the Markov model. During a preprocessing phase, these states have to be extracted from a spatio-temporal data set. Furthermore it is necessary to build an a-priori Markov model as described in our initial work ([3]). The resulting data is then imported into the database. The model is visualized in Figure 1(b), where line segments indicate possible transitions, and the thickness of a line indicates the probability of making a transition between the pair of connected nodes.[1] The ground-truth object trajectories are visualized in Figure 1(c). This figure shows the full trajectory of an object, independent of time. In the above picture, observations are highlighted by larger black circles. Based on such observations, the a-posteriori model can be derived using our techniques of [7]. Figure 1(d) illustrates the a-posteriori model, analogous to Figure 1(b). Given the adapted model, the probability to visit each node of the network is visualized in Figure 1(e).

In our recent work we have used distance metrics, especially the expected distance between the ground truth data and the model based on observations, to numerically prove the vast improvement over other approaches. Yet, the intention of this demonstration is *not* to re-run existing experiments on modeling and querying uncertain spatio-temportal data. Rather, the main goal is to present to a broad audience the practical impact of this data integration of combining historical trajectory data and concrete observations of an object to reduce uncertainty in spatio-temporal data. The main modules of the UST-framework, which show this impact and which we wish to present to the SIGMOD audience, will be described in the following two sections.

[1]The image shows the maximum transition probabilities of both directions. On the other hand, it would also be possible to split line segment into two equi-distant parts, each having individual thickness.

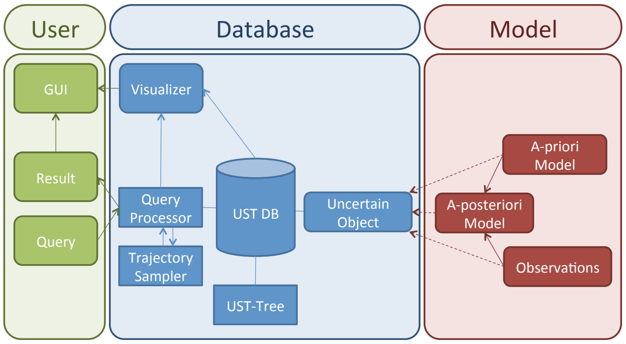


Figure 2: UST-framework: System Architecture

3. FRAMEWORK DESCRIPTION

For the purpose of exploring data models for uncertain spatio-temporal data, a first aspect of our demonstrator allows visualization of various aspect of such data. For this purpose, a data set consisting of (location, time, object)-triples can be imported into the framework. The data sets that we want to demonstrate include – in addition to a number of artificial spatio-temporal data sets – the T-Drive data set ([19]) pertaining to GPS-traces of taxis in the city of Beijing. In addition to simple data visualization, the framework can also be used to evaluate the model accuracy of uncertainty models, such as interpolation (which has been used previously in [8, 9, 10, 12, 11]), geometric models (that have been investigated in [14, 15, 16, 6, 4, 13]) and stochastic processes, especially under the Markov chain model ([3, 7]). Beside the Markov chain model we have implemented two exemplary approaches, an interpolation approach based on the shortest paths between observations, and a geometric approach representing the uncertainty area between observations by diamonds. While the accuracy of our approach has already been evaluated in [7], this numeric evaluation gives little insights on the impact this research can have on concrete systems. This demonstration shows how the fusion of both data sources, historic traffic data and current object observations, can very accurately describe the motion of an uncertain spatio-temporal object. The main features of the UST-framework are:

- Integration of the latest techniques for managing and querying uncertain spatio-temporal data.
- A powerful visualization toolbox used for spatio-temporal data sets and query results.
- An intuitive framework which can easily be extended by new query types, visualization techniques, and index structures.

The general structure of the framework is sketched in Figure 2: The database consists of a set of uncertain objects, each represented by a set of (location, time) pairs called observations. Uncertain objects are stored in a database (UST DB) that stores all uncertain objects and contains one or more index structures (for example a UST-Tree) to efficiently access them. The database stores an a-priori Markov model ([3]), i.e. a (global) transition matrix. This model is usually learned from the full historical data set stored in the database. Using observations of individual objects, the transition matrix can be adapted using the approach from [7]. The resulting a-posteriori Markov model accounts for these observartions by adapting transition probabilities such that all possible trajectories match all observations. While these models could be computed during query time, they are stored in the database to improve the runtime performance of the system.

The implemented interpolation and geometric models are stored in the system as well, together with their corresponding index structures. Queries are posed on a query processor that forwards the query to the database and underlying index structure. Depending on the hardness of the query, the query process either computes exact result probabilities, or employs a sampling approach to approximate result probabilities. The visualizer module is used to visualize both the database contents and query results.

4. DEMONSTRATION SCENARIOS

Per default, the spatio-temporal T-Drive [19] data set will be showcased, using the street network of Beijing, China. During the presentation, the presenter will browse the data set by moving the current focus to different parts of the city, moving forward and back in time and switching between data models. A snapshot of the framework is shown in Figure 3. In our demonstration we will present several application scenarios to the audience with the goal of providing visual insights to the practical impact resulting from our previous theoretical work. For this purpose we will address three different scenarios, which we will present using real-world as well as synthetic data sets:

4.1 Database Visualization

The framework visualizes a database of uncertain spatio-temporal objects. After opening the database connection, the user has access to a map showing the uncertain objects in the database at a given point in time, and a timeline visualizing the distribution of uncertain objects over the database time horizon. To illustrate the underlying road network (of Beijing in Figure 3), the demonstrator uses Open-Street-Map data. The map view provides bounding boxes of the uncertain objects approximating the position of each object at the given point of time. The visualized boxes correspond to spatial approximations of our spatio-temporal index structure ([2]) for the Markov model. The implemented geometric model is described by these bounding boxes as well. Certain and interpolated trajectories are visualized by their discrete positions. For a selected object (Markov model) the map shows non-zero probabilities of staying at the corresponding position as circles, while non-zero transition probabilities of moving from a given state to a neighboring state are visualized as lines between states.

4.2 Spatio-Temporal Query Processing

We further provide an easy-to-use interface for performing window queries to the user. The user interface for this feature is also depicted in Figure 3. To pose a window query on the database, the user has to first select a time-horizon on the timeline located in the bottom part of the user interface. Furthermore, the user is required to specify a spatial query window in the map view located at the upper-left part of the user interface. After defining the spatio-temporal window query, results are shown in a result view on the right-hand side. The type of result depends on the selected uncertainty model. In the case of interpolation models, a set of objects, corresponding to a raw guess (which might indeed not even be possible) is returned. In the case of geometric models, a set of possible results (which may indeed have a zero probability) are returned. If the Markov model ([3]) is selected, objects are returned together with their probabilities of being results. Furthermore, result objects can be highlighted in the map visualization. In the case of nearest neighbor queries, the user defines a query point on the map and a time interval on the timeline. After query processing, the user can highlight the objects with a nearest neighbor probability greater than zero for the given time interval on the map.

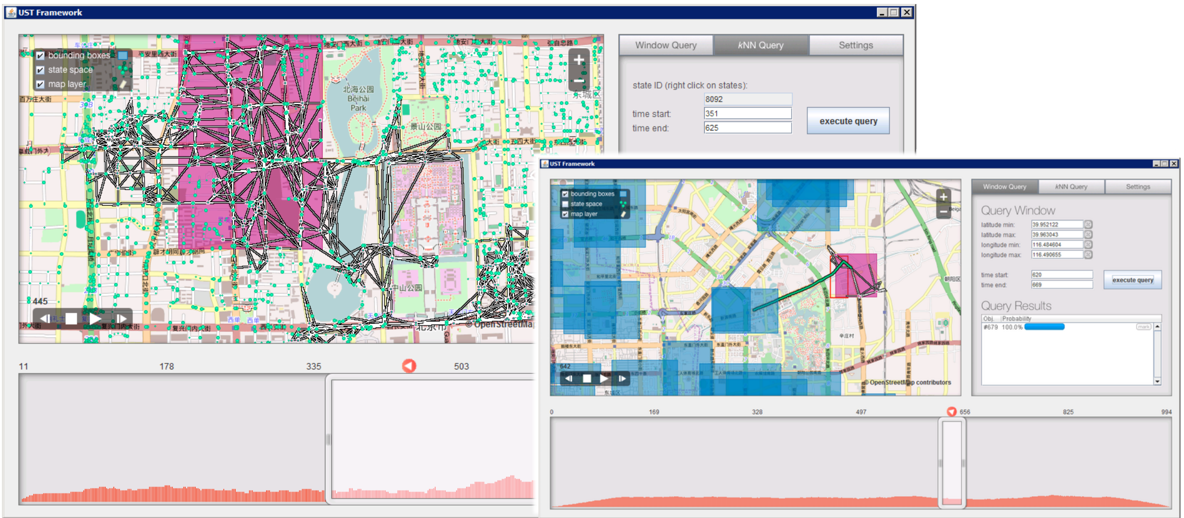


Figure 3: Graphical user interfaces for window queries (left) and NN queries (right).

4.3 Data Modification

Last but not least we will show how the user is able to insert new observations to uncertain trajectories. This is particularly of interest since the integration of new observations into the model yields new uncertainty regions of the corresponding object leading to a smaller degree of uncertainty. The uncertainty regions of all models are adapted to the new observation after insertion.

The demonstration will be side-kicked by a poster, to introduce the audience to the field of modeling uncertainty in spatio-temporal data. The main purpose of the poster is to present theoretic foundations of our recent publications ([3, 2, 7]) which are too complex (and thus time consuming) to be presented orally. Thus, the poster will provide the basics of our Markov model and the approach used to perform a data integration between historic movement patterns and individual object observations. Our UST-framework has been implemented in C++ and Java. For further information regarding the model and framework we refer to the UST project page.

Acknowledgements

This work is partially suppported by Deutsche Forschungsgemeinschaft (DFG), under grant RE 2667/5änd by German Accademic Exchange Service (DAAD), under grant "56048240".

5. REFERENCES

[1] J. Aßfalg, H.-P. Kriegel, P. Kröger, and M. Renz. Probabilistic similarity search for uncertain time series. In *Proc. SSDBM*, pages 435–443, 2009.

[2] T. Emrich, H.-P. Kriegel, N. Mamoulis, M. Renz, and A. Züfle. Indexing uncertain spatio-temporal data. In *CIKM*, pages 395–404, 2012.

[3] T. Emrich, H.-P. Kriegel, N. Mamoulis, M. Renz, and A. Züfle. Querying uncertain spatio-temporal data. In *Proc. ICDE*, pages 354–365, 2012.

[4] B. Kuijpers and W. Othman. Trajectory databases: Data models, uncertainty and complete query languages. In *JCSS*, pages 538–560, 2010.

[5] N. Mamoulis, H. Cao, G. Kollios, M. Hadjieleftheriou, Y. Tao, and D. W. Cheung. Mining, indexing, and querying historical spatiotemporal data. In *Proc. KDD*, pages 236–245, 2004.

[6] H. Mokhtar and J. Su. Universal trajectory queries for moving object databases. In *Mobile Data Management (MDM)*, pages 133–144, 2004.

[7] J. Niedermayer, A. Züfle, T. Emrich, M. Renz, N. Mamoulis, L. Chen, and H.-P. Kriegel. Probabilistic nearest neighbor queries on uncertain moving object trajectories. In *VLDB*, 2014.

[8] D. Pfoser, C. S. Jensen, and Y. Theodoridis. Novel approaches to the indexing of moving object trajectories. In *Proc. VLDB*, pages 396–406, 2000.

[9] S. Saltenis, C. S. Jensen, S. T. Leutenegger, and M. A. Lopez. Indexing the positions of continuously moving objects. In *Proc. SIGMOD*, pages 331–342, 2000.

[10] Y. Tao and D. Papadias. Time-parameterized queries in spatio-temporal databases. In *Proc. SIGMOD*, pages 334–345, 2002.

[11] Y. Tao, D. Papadias, and Q. Shen. Continuous nearest neighbor search. In *Proc. VLDB*, pages 287–298, 2002.

[12] Y. Tao, D. Papadias, and J. Sun. The tpr*tree: An optimized spatio-temporal access method for predictive queries. In *Proc. VLDB*, pages 790–801, 2003.

[13] G. Trajcevski, A. N. Choudhary, O. Wolfson, L. Ye, and G. Li. Uncertain range queries for necklaces. In *11th International Conference on Mobile Data Management (MDM 2010), Kansas City, Missouri*, pages 199–208, 2010.

[14] G. Trajcevski, R. Tamassia, H. Ding, P. Scheuermann, and I. F. Cruz. Continuous probabilistic nearest-neighbor queries for uncertain trajectories. In *Proc. EDBT*, pages 874–885, 2009.

[15] G. Trajcevski, O. Wolfson, K. Hinrichs, and S. Chamberlain. Managing uncertainty in moving objects databases. *ACM Trans. Database Syst.*, 29(3):463–507, 2004.

[16] G. Trajcevski, O. Wolfson, F. Zhang, and S. Chamberlain. The geometry of uncertainty in moving objects databases. In *Proc. EDBT*, pages 233–250, 2002.

[17] C. Xu, Y. Gu, L. Chen, J. Qiao, and G. Yu. Interval reverse nearest neighbor queries on uncertain data with markov correlations. In *Proc. ICDE*, 2013.

[18] M.-Y. Yeh, K.-L. Wu, P. S. Yu, and M. Chen. Proud: A probabilistic approach to processing similarity queries over uncertain data streams. In *Proc. EDBT*, pages 684–695, 2009.

[19] J. Yuan, Y. Zheng, C. Zhang, W. Xie, X. Xie, and Y. Huang. T-drive: Driving directions based on taxi trajectories. In *Proc. ACM GIS*, 2010.