Efficient Processing of Probabilistic Spatio-Temporal Range Queries over Moving Objects

Bruce S.E. Chung
Department of Computer Science
National Tsing Hua University
Hsinchu, Taiwan, R.O.C.
se.chung@yahoo.com.tw

Wang-Chien Lee
Dept of Comp. Science & Engineering
Pennsylvania State University
State College, PA 16801, U.S.A.
wlee@cse.psu.edu

Arbee L.P. Chen
Department of Computer Science
National Chengchi University
Taipei, Taiwan, R.O.C.
alpchen@cs.nccu.edu.tw

ABSTRACT
Range queries for querying the current and future positions of the moving objects have received growing interests in the research community. Existing methods, however, assume that an object only moves along an anticipated path. In this paper, we study the problem of answering *probabilistic range queries* on moving objects based on an uncertainty model, which captures the possible movements of objects with probabilities. We conduct a performance study, which shows our proposal significantly reduces the number of object examinations and the overall cost of the query evaluation.

Categories and Subject Descriptors
H.2.4 [**Database Management**]: Query Processing.

General Terms
Algorithms, Design, Performance.

1. PROBLEM DESCRIPTION AND IDEAS
There are growing demands for moving objects monitoring functions in numerous mobile applications. For instance, by continuously receiving location updates from buses on roads, a bus control system can perform a better bus scheduling. To avoid traffic congestion, a range query, which "retrieves all the buses that will arrive within 1 mile of the station in the next 10 minutes", may be issued to obtain estimated answers. Research on spatio-temporal databases that manage objects' moving information has produced fruitful. A number of efficient methods for managing moving objects have been proposed. Typically, they assume that an object moves on an anticipated path. As a result, an object's most possible path based on a linear function or recent information is maintained. In fact, the real location of an object is known *for certain* to the server only when an update is received. The uncertainty of the object's location increases as time grows until the next update is received. Thus, the uncertainty has a great impact on the accuracy of these proposed methods.

Since most objects move without deviating from their recent moving behaviors drastically, the uncertainty can be bounded, i.e., an object's future positions could be captured using an *uncertainty probability density function* (pdf) [1]. A pdf of an object O_i moving in a one-dimensional (1-D) space can be represented as $f_i(x,t)$ where f_i is the probability of O_i at the location x at the time instant t. Therefore, an object's location at any time can be estimated with an uncertainty probability defined by its pdf. Typically, the function f_i can be derived from O_i's past moving behavior or the moving velocity distribution. Thus, each object can be stored in the server database with a pdf instead of a location from the last update. A moving object's velocity distribution can be approximated as a normal distribution. We employ the *Brownian motion with drift process* as an uncertainty model to produce a pdf from a normal velocity distribution for all objects. The Brownian motion model can be derived with two parameters, i.e. mean velocity and variance. Thus, by adopting this model, computation cost at the server and communication cost between the server and the objects can be reduced.

CIKM'08, October 26–30, 2008, Napa Valley, California, USA.
ACM 978-1-59593-991-3/08/10.

To query moving objects using an uncertainty model, [1] has proposed *probabilistic range queries (PRQ),* e.g., "retrieve the identifications of all the buses that will come within 1 mile of the station in the next 10 minutes with a probability more than 0.4." Unlike a conventional range query, a probability is specified as part of query conditions. Additionally the answer is to be obtained over a *time period* instead of just a *time instant*. An object is returned as an answer if its probability inside the query range (1 mile) is greater than 0.4 at any moment during the next 10 minutes. This query, by taking into account both the location range over the time period, returns objects with probabilities satisfying the specified probability threshold.

Probabilistic condition in the query can only be evaluated on objects with expensive integral computation, incurring significant overhead. Several methods have been proposed to index objects with uncertain movement. A common idea is to pre-determine the uncertainty intervals of objects with the same probability bounds. Several different bounds are pre-defined with various probability values and grouped by the proposed indexes. A query is processed by first verifying these bounds to reduce the number of objects under consideration. However, these indexes are not applicable to *time-varying uncertain data*. In other words, the assumed uncertainty model is time-independent, i.e., an object's uncertain range and probability values remain unchanged no matter how much time passing by from the last update.

In this work, we propose efficient techniques for querying uncertain time-varying data. We observe that the expensive query cost is due to: (1) the waste of time in evaluating objects far away from the query range, and (2) the expensive integral computation required for evaluating the probability. In order to solve problem (1), we transform the uncertain movements of objects into points in a dual space using the Hough Transform [2]; therefore, these points can be indexed by a point access method, such as R-tree. After transforming the movements, however, only the most representative path with a timestamp is maintained in the index. As a result, some of the answers may get lost when querying on the index. We expand the query to avoid losing answers and transform the expanded query into a search range. Querying on the index using the search range allows pruning the unqualified points outside the search range. Thus, the cost of object examinations is reduced. After the elimination process, the remaining objects are examined to evaluate their probabilities. For problem (2), we develop an approximate approach to reduce the cost of integral operations in evaluating the probabilistic query condition. While the proposed approximate approach results in false positives, none of the objects that should be included in the answers is lost. We also develop an error function to compute the probability bound for the answers.

2. PRELIMINARY EVALUATION

We conduct experiments on uncertain object movements in one-dimensional space. We show the running time of querying with time instant and time period. All the algorithms are implemented in C++ and carried out on a 3.2GHz Intel Pentium IV PC with 1G main memory, running Windows XP SP2.

2.1 Experimental Setting

The experimental data is generated as follows. We simulate *N* objects moving on a line segment forthright [0, 200000], which has the length 200 kilometers. We vary *N* from 100*K* to 1*M*. We define the time unit is 1 second (1s). In the initial stage (i.e., at time $t = 0$), *N* objects are uniformly distributed on the forthright. The objects speeds are randomly generated from v_{min} = 10 *meter/sec* to v_{max}= *50 meter/sec* (10 *meter/sec* is equal to 36 *km/h* and 50 *meter/sec* is equal to 180 *km/h*.) and the direction is randomly positive or negative. The objects velocity variance is also randomly assigned from 4 $meter^2/sec$ to 16 $meter^2/sec$. Then the objects start moving. Each object re-generates its speed and variance until the distance between the location it moves to and the previous updated location is up to 500 *meters*. At each time instant we execute 100 random queries, where the length of the location is randomly chosen from 100 to 10000, the time length is from 1 to 600. Note that we have two kinds of queries with time instant and time period, respectively. We randomly generate a time instant from 1 to 3600 to assign the start time of a time period. We implement R-tree as our indexing structure. We keep all the information including the index and moving objects' data in the main memory instead of in the hard disk since performance of index is out of scope of this study.

2.2 Performance Study

Figure 1 presents the query execution time using our approach (labeled by HT since it's based on Hough Transform) in comparison with the non-indexing method for 500*K* and 1*M* objects. The HT-process includes two stages, i.e., 1) elimination, and 2) (approximate) examination. Non-index method retrieves all objects for evaluation by cumulative normal density function. The query execution time of HT-process decreases as the query probability increases till the probability threshold exceeds 0.5, because the generated queries with the same location interval are mapped to the same search range when the probabilities exceeds 0.5. Figure 2 shows the execution time for querying with time period. The execution time on 1*M* objects decreases more significantly than on 500*K* objects as the probability threshold increases because the elimination and examination are effective.

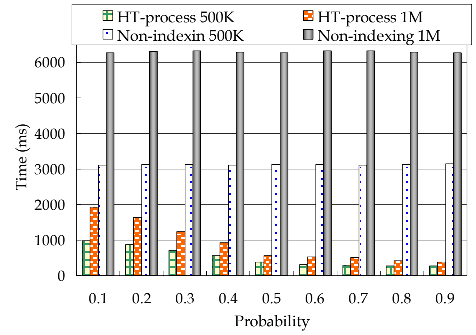

Figure 1. Execution time for the queries with time instant.

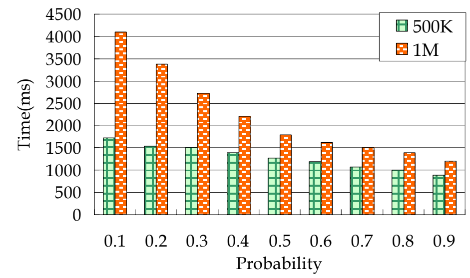

Figure 2. Execution time for the queries with time period.

3. CONCLUSION

In this paper, we investigate the problem of probabilistic query on objects with uncertain movements. We employ Brownian motion model for all the moving objects. In this model, every moving object's uncertain movements are represented as a probability density function. We extend conventional probabilistic range queries (PRQ) with a time period and a probability threshold. To process the query efficiently, we transform all objects uncertain movements into simple points and indexed these points for efficient querying. We developed approximate formulas and an algorithm with error bounds to evaluate probabilities of the moving objects and to ensure the correctness of the querying answers. Experimental results show the effectiveness and efficiency of our approach.

4. ACKNOWLEDGEMENT

Wang-Chien Lee was supported in part by NSF under Grant no. IIS-0328881, IIS-0534343 and CNS-0626709.

5. REFERENCES

[1] R. Cheng, D. V. Kalashnikov, and S. Prabhakar. Querying Imprecise Data in Moving Object Environments. In *IEEE Transactions on Knowledge and Data Engineering*, pp 1112-1127, 2004.

[2] P. V. C. Hough. *Method and Means for Recognizing Complex Patterns*, U. S. Patent No. 306964, 1962.