k-

:

...

, 2015

***k*-**

:

...

29 2015.

()	*()*	*()*
....	....	....
...	...	.

, 2015

()

....

...

. ...

. ...

, ...

Περίληψη

. , : (i) q, (ii) k , (iii) θ.
, , . , . , , . ‘ ’, q, k θ.
, . . , , . , .
, .

, , , , .

Abstract

The purpose of this diploma thesis is to develop and implement an algorithm for most likely nearest neighbors monitoring from specific focal points in a hypothetical service for smartphone users. Whenever a user submits a most likely nearest neighbors query, sets three criteria: (i) a focal point of interest q, (ii) the desired number k of nearest neigbors, and (iii) a probability threshold θ.

Because of privacy protection reasons, no user compromises their geographical position to the rest, but declares a wider *uncertainty region*. In this case, these regions are modelled according to the bivariate Gaussian distribution. Of course, uncertainty can acquire different parameters, expressing different scales of privacy. By using the term "*most likely nearest neighbors*", we mean that in a certain search region arount point q, k moving users with probabilistic coverage above a certain threshold θ have been found.

This thesis mainly focuses on developing indexing, filtering and pruning techniques which will enable us to reduce the cost and processing time of data. The suggested algorithm is deliberately chosen to be approximate in the calculation of probabilistic coverage of uncertain regions and provides a solution to the problem of answering probabilistic nearest neighbor queries for uncertain positions of moving objects. By utilizing the above techniques, a experimental study was conducted against synthetic datasets generated using the map of Athens. In addition, the expected performance on the execution times and accuracy of answers was confirmed. The overall conclusion of this thesis is that the algorithm is suitable for real time problems, where some accuracy may be sacrificed for the benefit of timely response.

Keywords

Uncertainty, Probabilistic nearest neighbor queries, bivariate Gaussian distribution, moving objects, data streams.

Περιεχόμενα

Κατάλογος Σχημάτων

Κατάλογος Πινάκων

Κεφάλαιο 1

. . , .

1.1

.) ( ),) . , . .

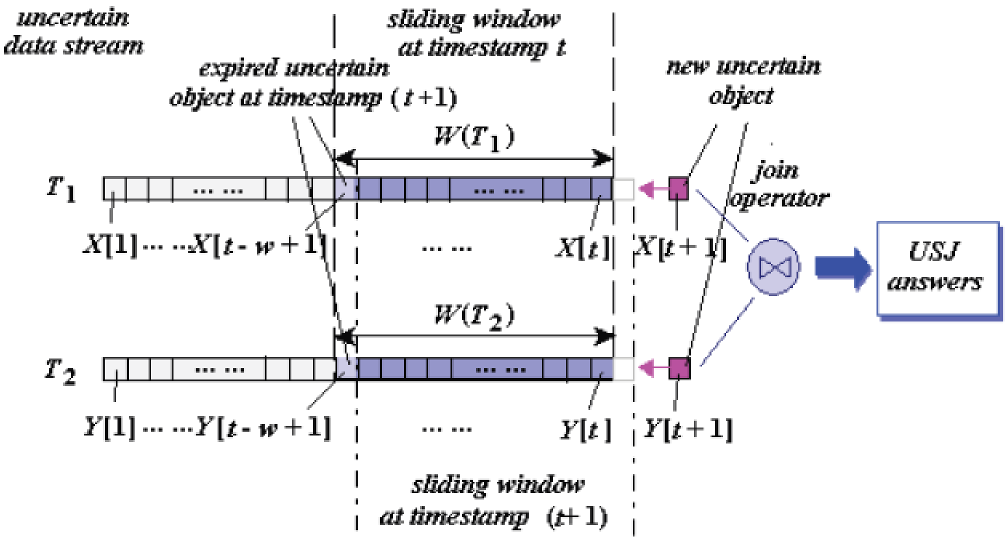
1.1.1

// . :
:

1. ...
2. ...
3. ...
4. ...
5. ...

1.2

: . ( ):
2. 3 . 4 ...
( ), . , .



Σχήμα 2.1: (: [1])

Κεφάλαιο 2

2.1

, . , .

2.2 ‘ 1’

// 1 . . , , . 2.1.

, , , , . . :

* : [8]
* : [1, 2, 3]
* : [4, 5, 7]
* : [6]
* : [9]

2.3 ‘ 2’

...

Κεφάλαιο 3

3.1

// . // . //, .

3.2 ‘ , , 1’

...

3.3 ‘ , , 2’

...

Κεφάλαιο 4

῾ ’, ..: $\boldsymbol{k}$-

4.1

2-3 .

4.2 ῾ , ..: ’

.

4.3 ῾ , ..: ’

, , , ...

Κεφάλαιο 5

̔ ̓, ..: *k*-

5.1

, .

5.2 ̔ / 1, ..: ̓

, 1 . .. . /, .

Algorithm 1 Probabilistic $k\theta NN$ Monitoring

Procedure *VerifyCandidate* (focal query point q, threshold θ, object o, list of auxiliary objects P, distance $kMAXDIST$)
if $\Phi(o, kMAXDIST) \geq \theta$ **and** $L_2(q, o) \leq L_2(q, P.\text{top}())$ **then**
  P.pop(); *//Replace the most extreme element in P, since candidate o ...*
  P.push(o); *//... has enough probability and has its mean closer to focal q*
end if
End Procedure

5.3 ̔ / 2, ..: ̓

...

Κεφάλαιο 6

.

6.1

.

6.2

. , βλοςκ διαγραμ . , ‵ ′.

6.3

. ..) (,)) , ...
, .. 6.1:

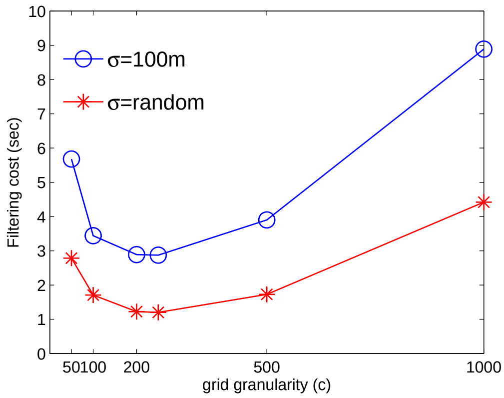
$c \times c$	50 × 50, 100 × 100, 200 × 200, **250 × 250**, 500 × 500, 1000 × 1000
σ	25m, 50m, 75m, **100m**, 150m, 200m
k	1, 2, **3**, 4, 5, 10, 20
θ	50%, 60%, 70%, **75**%, 80%, 90%, 99%

Πίνακας 6.1:

, :

6.3.1

...



Σχήμα 6.1:

6.4

, . 6.1. , . ( ) , .. .

6.5

. , .. 1. , 2. , ...

Κεφάλαιο 7

.

7.1

. , . , ...

7.1.1 ‘ 1’

...

7.1.2 ‘ 1’

...

7.2

, , , ηαρδωαρε, ... , . , .

Κεφάλαιο 8

.

8.1

, . .

8.2

.

Βιβλιογραφία

[1] D. J. Abadi, D. Carney, U. Çetintemel, M. Cherniack, C. Convey, S. Lee, M. Stonebraker, N. Tatbul, and S. Zdonik. Aurora: a new model and architecture for data stream management. *The VLDB Journal – The International Journal on Very Large Data Bases*, 12(2):120–139, 2003.

[2] N. Dalvi, C. Ré, and D. Suciu. Probabilistic databases: Diamonds in the dirt. *Communications of the ACM*, 52(7):86–94, July 2009.

[3] R. H. Güting, M. H. Böhlen, M. Erwig, C. S. Jensen, N. A. Lorentzos, M. Schneider, and M. Vazirgiannis. A foundation for representing and querying moving objects. *ACM Transactions on Database Systems (TODS)*, 25(1):1–42, 2000.

[4] N. Jain, S. Mishra, A. Srinivasan, J. Gehrke, J. Widom, H. Balakrishnan, U. Çetintemel, M. Cherniack, R. Tibbetts, and S. Zdonik. Towards a streaming SQL standard. In *Proceedings of the VLDB Endowment*, 1(2):1379–1390, 2008.

[5] K. Mouratidis, D. Papadias, and M. Hadjieleftheriou. Conceptual partitioning: an efficient method for continuous nearest neighbor monitoring. In *Proceedings of the 24th ACM SIGMOD International Conference on Management of Data*, pages 634–645, 2005.

[6] Oracle, Inc. Complex event processing CQL language reference. `http://docs.oracle.com/cd/E16764_01/doc.1111/e12048/intro.htm` 2009. Last accessed on 15/09/2013.

[7] K. Patroumpas and T. Sellis. Subsuming multiple sliding windows for shared stream computation. In *Advances in Databases and Information Systems*, volume 6909 of Springer *Lecture Notes in Computer Science*, pages 56–69, 2011.

[8] P. Rigaux, M. Scholl, and A. Voisard. *Spatial databases: with application to GIS*. Morgan Kaufmann, 2001.

[9] . . *DIPL-2015-02* , , 2015.

uncertainty
cumulative distribution function
query evaluation
sampling
indexing
continuous query
nearest-neighbor query
privacy
grid
moving object
window
multiplexing
data stream
focal point
aggregation
join
filtering
timestamp