다항회귀함수를 이용한 이동객체의 불확실한 위치 추정

양은주°, 안윤애, 오인배, 류근호
충북대학교 데이터베이스 연구실
{ejyang, yeahn, iboh, khryu}@dblab.chungbuk.ac.kr

Estimating Moving Object's Uncertain Position using Polynomial Regression Function

Eun Joo Yang°, Yoon Ae Ahn, In Bae Oh, Keun Ho Ryu
Database Laboratory, Chungbuk National University

요 약

샘플링되지 않은 불확실한 이동객체의 위치값을 추정하기 위한 기존의 연구방법 중 가장 보편적으로 사용하고 있는 방법은 선형 보간법이다. 선형 보간법을 사용할 경우 샘플링 구간을 좁게하여 오차를 줄일 수 있고 계산 시간을 단축할 수 있지만, 연속적인 이동객체의 경로는 직선이라기 보다는 곡선으로 나타내어지므로 샘플링되지 않은 이동객체의 위치값에 대해 불확실한 위치정보를 사용자에게 반환하게 된다. 따라서 이 논문에서는 샘플링된 이동객체의 위치값에 오차가 없다는 가정하에서 모든 위치점을 지나는 보간 다항식을 구해서 처리하는 선형 보간법 대신 이동객체의 위치값 자체의 오차범위까지 고려하는 다항회귀모형(polynomial regression model)을 이용한 이동객체의 불확실한 이동위치 추정 방법을 제시한다. 다항회귀모형을 이용할 경우 선형 보간법 보다 추정된 위치값에 대한 오차를 줄일 수 있으며, 이동객체의 과거 및 미래 위치값을 사용자에게 반환해 줄 수 있는 장점을 가진다.

1. 서론

실세계의 공간 즉, 점들의 집합 내에서 시간에 따라 위치나 형태가 변하는 객체를 시공간 이동객체(Spatiotemporal Moving Objects)라 하며, 이동점(Moving Point)과 이동영역(Moving Region)으로 나눌 수 있다. 이들을 데이터베이스화할 때는 점 혹은 영역형태로 표현하고 저장하게 된다. 이동점은 시간에 따라 위치가 변하며 특정 시점에 대한 질의값은 그 시간에 이동객체가 존재하는 위치를 나타내는 점으로 반환되며, 이동영역은 시간에 따라 위치와 형태가 변하며 특정 시점에 대한 질의는 그 시점에 이동영역이 존재하는 위치와 모양을 표현하는 영역을 반환한다. 이러한 이동객체의 샘플링 되지 않은 과거 혹은 미래 시점의 위치에 대한 질의 시 시스템은 이에 대한 위치값을 추정해서 질의자에게 반환해줘야 할 필요가 있다.

이를 위해 기존의 연구방법은 이동객체의 샘플링 되지 않은 과거 위치를 선형보간법으로 처리하며[Pfos99], 샘플링 되지 않은 미래 위치는 이동객체의 현 속력이나 방향 같은 동적 속성값을 가지고 이동객체의 향후 이동위치를 미리 계산해서 미래 테이블에 저장한다[Sist97]. 후자의 연구에서는 동적 속성값이 매번 변할 때마다 미래 테이블을 연속적으로 갱신해야 하므로 갱신비용이 많이 들며, 과거의 이력정보는 저장하지 않는다. 즉, 과거 위치에 대한 질의시 응답을 처리하지 못한다는 단점을 지니고 있다. 전자의 연구에서와 같이 측정장비로 측정시 오차가 없다는 가정하에 선형보간법을 이용할 경우, 계산 시간을 단축할 수 있지만 이는 아주 큰 오차를 포함하는 이동객체 위치 추정치를 반환한다[Bart97]. 따라서 이동객체의 샘플링된 시간이 GPS와 같은 센서 시스템에 의해 실제 측정된 시간이 아닌 데이터베이스에 기록되는 트랜잭션시간으로 저장이 되기 때문에 위치값 자체에 오차를 포함하고 있으므로, 오차 추정치를 최소화하는 통계적인 접근방법을 이용해 유효시간구간(valid time interval)내의 이동객체의 움직임을 선형회귀함수로 표현한다[More99]. 그러나, 연속적인 이동객체의 이동경로는 직선이라기 보다는 곡선에 가깝다. 따라서, 이 논문에서는 샘플링된 이동객체의 위치값 자체의 오차 범위를 고려하는 통계적인 기법인 다항회귀함수를 통해 샘플링되지 않은 위치에 대한 추정 방법을 제시한다.

** 이 연구는 1999년도 KOSEF 특정기초(#1999-2-303-006-3) 연구비 지원으로 수행되었음.

2. 관련연구

시공간 시스템은 시간에 따라 변하는 공간 객체만을 취급한다. 이들 시공간 응용은 지구 과학, 경제학과 사회경제학, 도시계획, 교통 제어, 토지정보시스템, 환경 보호나 의료영상과 같은 여러 종류의 도메인에서 시간과 공간 컴포넌트들을 통합함으로써 유용하게 쓰일 수 있다. 그러나 두 컴포넌트는 독립적으로 연구되어 왔다. 기존의 시간 데이터베이스는 시간에 따라 연속적으로 변하는 객체에 대한 정보 표현은 지원하지 않고 있다. 이 시간 데이터베이스는 객체의 이동이나 회전 혹은 변형된 정보를 조정해야하는 GIS 응용이나 온도 혹은 기압과 같은 연속적으로 변하는 속성들을 취급하는 시스템에 유용하다. 시간 연속모델에서 시간은 실수로 취급되는데, 이런 무한한 집합들을 컴퓨터에 의해 직접 저장하거나 조작하기란 불가능하다[More99]. 따라서 이산적으로 샘플링된 이동객체의 위치를 기반으로 샘플링되지 않은 위치값을 추정시 이에 대한 불확실성이 발생하게 된다. 시공간 도메인에서의 이동 객체의 불확실성에 대한 연구는 그리 많지 않다. DOMINO 프로젝트[Sist97]에서는 시간에 따라 연속적으로 변하는 속성을 취급하는 시스템을 위해 MOST(Moving Objects Spatio-Temporal) 데이터모델과 FTL(Future Temporal Logic)언어를 소개하고 있다. 이 모델에서는 이동객체의 방향이나 속도가 이동객체의 이동 바운드를 초과할 때마다 자동적으로 선형함수가 변하게 된다. FTL 질의어는 이동객체의 미래 위치값에 대한 질의를 처리해준다. [Yeh94]에서는 이동객체의 위치가 어떻게 전개되는지를 표현해주는 행위함수(behavioural function)로서 punctual function과 step function, linear function, interpolation function를 소개하

고 있다. 이동객체 위치 추정을 위하여 행위함수 중 [Pfos99]에서는 보간함수를 사용하고 있으며, [Sist97]에서는 선형회귀함수를 사용하고 있다. [Pfos99]는 이동객체 데이터를 위한 시공간 불확실성을 소개하였으며 이동객체의 이동경로 에러 정보를 계산하고 이용하는 방법에 대한 접근방법을 소개하고 있다.

3. 이동점 객체 모델

3.1 이동객체

이동점 객체 타입은 크게 자유로운 이동경로를 갖는 객체와 제한된 이동경로를 갖는 객체로 나눌 수 있다[More99]. 자유로운 이동경로는 n차원의 공간상 이동점객체의 움직임에 있어 아무런 제한점이 없음을 의미한다. 예를 들어, 바다를 항해하는 배의 이동경로를 예로 들수 있다. 그와는 반대로 제한된 이동경로는 이동점객체의 움직임이 공간상 많은 제한을 받음을 의미하며, 철도 선로나 도로를 따라 운행하는 열차 혹은 자동차의 이동경로를 예로 들 수 있다. 이 논문에서는 제한된 이동경로를 따라 움직이는 이동점객체의 이력을 표현해주는 데이터모델을 기반으로 하고 있으며, 모든 이동점객체에 대해 샘플링된 위치값들은 서로 독립적이며 이들은 정규적인 시간주기에 따라 샘플링된다고 가정하고 있다.

이동객체의 내부에 장착되어 있는 GPS와 같은 센서시스템이 (Moid, t, x, y)순서쌍으로 이동객체의 정보를 정규적인 시간주기에 따라 순차적으로 캡처하면, 이들 정보는 데이터베이스내에 저장된다. 이동객체의 시간과 위치 이외에도 이동객체의 움직임을 표현하는 속성으로는 속력과 방향 등 많은 요소들을 고려해 볼 수 있지만, 너무나 많은 설명변수를 회귀모형에 포함할 경우, 이동객체의 위치정보를 획득하므로써 변수들의 변화를 탐지, 관리하는데 많은 비용이 요구되므로[Chap94] 이동객체의 위치와 시간에 대해서만 고려한다. 이 논문에서는 추정된 이동객체의 위치정보들을 연속적으로 시간에 따라 행위함수를 적용해 미리 데이터베이스에 저장하는 것이 아닌, 이동객체의 불확실한 위치정보에 대한 질의 발생시마다 기존에 샘플링된 이동객체의 위치정보를 기반으로 이에 대한 위치값 혹은 시간을 추정해주는 접근방법에 중점을 두고 있다. 이동객체의 위치값에 대한 시간에 따른 움직임과 추정하고자하는 위치값을 표현하면 다음과 같다.

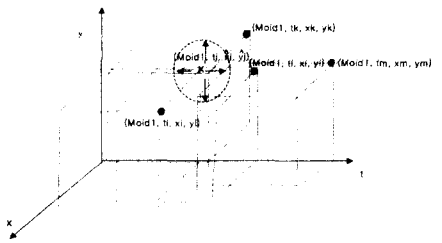

그림 1. 시간에 대한 이동객체의 x, y 위치값

위 그림에서의 (Moid1, ti, xi, yi), (Moid1, tk, xk, yk), (Moid1, tl, xl, yl) 위치값은 id가 1인 이동객체의 시간 인스턴트 i, k, l 때에 데이터베이스에 샘플링된 위치값이며, 이들은 위치값 자체에 오차를 포함하고 있으므로 그림 1에서 x로 표시된 추정하려는 위치값 (Moid1, tj, $\hat{x}j$, $\hat{y}j$) 또한 오차를 포함하고 있으므로 이 논문에서는 통계적인 접근방법을 사용한다.

3.2 다항회귀 함수를 이용한 위치추정

회귀모형은 독립변수 X가 변화함에 따라 종속변수 Y가 어떠한 함수형태를 가지고 변화하는지를 추정 혹은 예측하는 접근방법이다. 통계적 함수관계 $Y=f(X)+\varepsilon$는 독립변수의 값에 대하여 종속변수의 값이 유일하게 결정되지 않는 관계를 의미한다. 여기서 ε는 확률 오차항이다. 오차항 ε은 이동객체의 이동경로를 중심으로 무작위 하게 흩어져 이는 변동을 나타내며 확률적 오차는 종속변수에 중요한 영향을 미치지만 회귀모형에 생략된 다른 독립변수들의 영향을 포함한다고 볼 수 있다. 예를 들면 시공간 이동객체의 매 샘플링 시간과 이동위치의 관계를 분석할 때 샘플링 시간의 기준이 일정하더라도 이동위치의 차이는 있을 수 있는데, 그 이유는 이동객체의 이동위치는 샘플링 시간 이외에 운전상태에 영향을 줄 수 있는 기후, 도로상태, 운전자의 건강 상태, 자동차의 상태 등 다른 변수들의 영향을 받을 수 있기 때문이다. 여기서 다른 변수들의 영향은 오차항에 포함되어 있다고 볼 수 있다. 시공간데이터베이스의 이동객체에 대한 회귀모형 도출 시 선형회귀모형을 이용할 경우, 독립변수들과 종속변수의 관계가 선형적이지 않으면 추정된 회귀선과 실제 샘플링된 이동객체의 경로간의 오차가 발생할 수 있으므로 이 논문에서는 종속변수의 변화가 독립변수에 대해 곡선 관계일 때 사용하는 방법인 다항회귀모형을 이용한다. 시간 독립변수에 따라 영향을 받는 설명변수는 x, y좌표값이므로, 이들 각각에 대한 다항회귀모형은 다음 식과 같다.

$$f_1(t)=\beta_{10}+\beta_{11}t+\beta_{12}t^2+\varepsilon_1 \quad 식(1)$$
$$f_2(t)=\beta_{20}+\beta_{21}t+\beta_{22}t^2+\varepsilon_2 \quad 식(2)$$

β_{i0}, β_{i1}, β_{i2}는 알지 못하는 값으로서 이를 회귀계수 또는 모수라 한다. 회귀모형에서는 샘플링한 이동객체의 위치값에 대한 오차항 ε의 값이 최소가 되는 회귀계수를 추정하는데 목적이 있다. 따라서 OLS(Ordinary Least Squares)추정을 이용해서, $f_1(t)$과 $f_2(t)$의 회귀계수들을 추정한다. 이를 추정하기 위하여 식(3)과 식(4)를 이용한다.

$$\hat{\beta}_1 = (T'T)^{-1}T'Y \quad 식(3)$$
$$\hat{\beta}_2 = (T'T)^{-1}T'X \quad 식(4)$$

oid가 1인 이동객체의 위치값 $\{(t_1,x_1),(t_2,x_2),\ldots,(t_k,x_k)$와 $\{(t_1,y_1),(t_2,y_2),\ldots,(t_k,y_k)$ 데이터베이스에 샘플링되어 있을 경우, 이때 샘플링되지 않은 t_c 시간때의 이동객체의 위치값을 추정하고자 할 경우 알고리즘은 다음과 같다.

```
algorithm PositionEstimateFunction (Moid, tc)
input: an object's identifier and time instant tc for estimating
output: the x and y coordinate value at instant tc
begin
    search the tuples that have the input oid and
          a time interval containing the time point tc from a relation
          that includes object's position information
    if (the tuple exist) then
          search before two and next two tuples of the result tuple,
          extract attribute values from five tuples and
          assign them to t[], X[], Y[]
          x_result ← PolyRegressFunction( t[], dataX[])
          y_result ← PolyRegressFunction( t[], dataY[])
    return x_result and y_result
end PositionEstimateFunction

algorithm PolyRegressFunction( t[], dataX[], dataY[])
input: a array of time points and a array of x or y value correspond to t
output: estimated regression coefficients about x and y coordinate values
begin
  for(int i=0; i<(order+1) * (order+1); i++)
      dataX[i] = X[matrix[i]];
  for(int i=0; i<order+1; i++)
      dataY[i] = Y[matrix[i]];
  GaussElimination(dataX, dataY, Coeff, order+1, order+1);

  GaussElimination(double dataX[], double dataY[], double coeff[], int num1, int num2){
      for(int i = 0; i < num1; i++) {
          double d = dataX[FindPosInArray(num2, i, i)];
          for(int j = i + 1; j < num1; j++) {
          double d1 = dataX[FindPosInArray(num2, j, i)] / d;
          for(int k = i; k < num2; k++) {
          dataX[FindPosInArray(num2, j, k)] = dataX[FindPosInArray(num2, j, k)] - d1 *
dataX[FindPosInArray(num2, i, k)];
                dataY[j] = dataY[j] - d1 * dataY[i];
          }
      }

      for(int i = num1 - 1; i >= 0; i--) {
          double d2 = 0.0;
          for(int j = num2 - 1; j > i; j--) {
              d2 += coeff[j] * dataX[FindPosInArray(num2, i, j)];
          }
          coeff[i] = (dataY[i] - d2) / dataX[FindPosInArray(num2, i, i)];
      }
  }
end PolyRegressFunction
```

4. 실험

이동객체의 샘플링되지 않은 불확실한 위치값을 추정하기 위한 방법 중 가장 보편적으로 사용되고 있는 방법은 선형 보간법이다. 선형 보간법을 사용할 경우 샘플링 구간을 좁게하여 오차를 줄일 수 있고 계산 시간을 단축할 수 있지만, 연속적인 이동객체의 경로는 직선이라기 보다는 곡선으로 나타내어지므로 샘플링되지 않은 이동객체의 위치값에 대해 불확실한 위치 정보를 사용자에게 반환하게 된다[Wolf98]. 이 논문에서는 곡선 적합 프로그램인 CurveExpert[Curv01]를 이용하여 이동객체의 위치값을 선형보간법과 다항회귀곡선에 적합하여 추정된 위치 값들을 비교한다. 이때, 데이터베이스에 샘플링된 이동객체의 위치값은 GSTD데이터셋[Pfos00b]을 통해 얻은 데이터이다.

x좌표	y좌표
0.02225	0.01216
0.02568	0.03056
0.02756	0.29800
0.03565	0.54223
0.04193	0.63960
0.04199	0.79344
0.04719	0.79421
0.05130	0.79552
0.05726	1.0000

표 1. 샘플링된 위치값

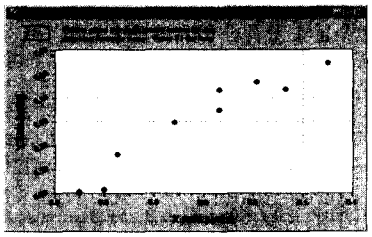

그림 2 샘플링된 위치데이터 그래프

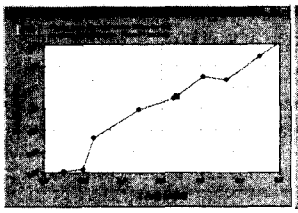

그림 3. 선형보간을 적용한 예

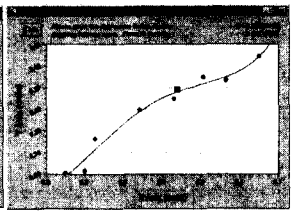

그림 4. 다항회귀를 적용한 예

그림2는 원래의 샘플링된 이동객체의 위치값과 그에 대한 이동경로이며, 그림 3과 4는 6번째 샘플링된 위치값(0.04199, 0.79344)을 제거한 후 이에 대해 선형보간법과 다항회귀기법을 이용해 이동경로 및 6번째 위치값을 추정한 예이다.

x좌표	y좌표
0.02225	0.01216
0.02568	0.02056
0.02756	0.01998
0.05130	0.37552
0.05726	0.33056
0.06554	0.21198
0.09990	0.25501
0.10789	0.21542
0.11329	0.13331
0.12278	0.19933
0.13375	0.32321
0.15335	0.45611
0.16314	0.55521
0.16327	0.68923
0.17587	0.65231
0.18551	0.54321
0.29738	0.58781
0.19878	0.42111
0.24235	0.68112
0.24445	0.99911

표 2. 샘플링된 위치값

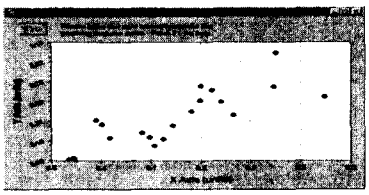

그림 5. 샘플링된 위치데이터 그래프

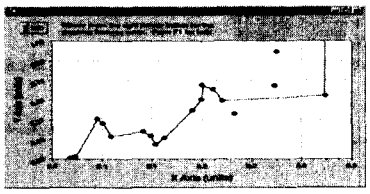

그림 6. linear interpolation를 적용한 예

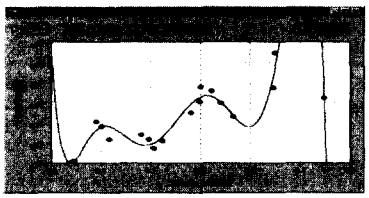

그림 7. polynomial regression를 적용한 예

그림 3과 4에서의 ■로 표시된 부분은 (0.04199, 0.79344)좌표와 근접한 좌표를 추정한 예이다. 두 그림에서 알 수 있듯이, 다항회귀기법을 이용한 그래프에서 (0.04199, 0.79344)에 훨씬 근접한 좌표가 반환된다. 그림 6과 그림7은 제한된 이동경로를 갖는 이동객체 외에도 자유로운 이동경로를 갖는 이동객체의 좌표값에 대한 실험결과이다. 11번째 샘플링된 위치값 (0.13375, 0.32321)을 제거한 후 이에 대해 11번째 위치값을 추정한 경우, 둘다 원래의 11번째 샘플링된 위치값과 거의 비슷한 결과를 보인다. 그러나, 선형보간의 경우, 시간에 대해 연속적으로 x, y좌표값이 커졌다가 작아지는 이동객체에 대해서는 그림 6에서 알 수 있듯이 보간을 해주지 못한다.

5. 결론 및 향후 연구

이동객체의 샘플링되지 않은 불확실한 위치값을 추정하기 위한 기법은 보간법, 퍼지집합론, 그리고 선형회귀함수를 이용한 통계학적인 접근 등이 있다. 이러한 방법 중 가장 보편적으로 사용하고 있는 방법은 선형 보간법이다. 그러나 이는 가장 단순한 접근방법이지만, 가장 큰 에러가 발생하게 된다. 따라서, 샘플링된 이동객체의 위치값 자체의 오차 범위를 고려하는 통계적인 기법인 다항회귀함수를 통해 샘플링되지 않은 위치에 대한 좀 더 정확한 위치 추정과 앞뒤 방향으로 이동하는 이동객체의 샘플링되지 않은 위치값 추정 또한 할 수 있음을 실험을 통해 알 수 있었다. 향후 과제로는 이 논문에서 제시한 알고리즘과 같이 추정하길 원하는 시간 인스턴트를 포함하는 시간구간에서의 샘플링된 위치값에 대해 다항회귀함수를 적용하여 샘플링되지 않은 과거 및 미래 위치값을 추정해서 리턴해주는 질의처리기를 구현하는 연구가 진행되어야 한다.

참고 문헌

[Bart97] R.Bartels, J.Beatty, and B.Barsky, "An Introduction to Splines for Use in Computer Graphics & Geometric Modeling", Morgan Kaufmann Publishers, Inc., 1987.

[Chap94] Steven C. Chapra, Raymond P. Canale, "Numerical Methods for Engineers", McGraw-Hill, 1994.

[Curv01] CurveExpert1.3 curve fitting and regression software, 2001, http://www.ebicom.net/~dhyams/cvxpt.htm.

[More99] Jose Moreira, Cristina Ribeiro and Jean-Marc Saglio, "Representation and Manipulation of Moving Points: An Extended Data Model for Location Estimation", Cartography and Geographic Information Systems (CaGIS), ACSM, Volume 26, No. 2, april 1999.

[Pfos99] Dieter Pfoser, Christian S. Jensen, "Capturing the Uncertainty of Moving-Object Representations", SSD, 1999.

[Pfos00a] Dieter Pfoser, Nectaria Tryfona, "Fuzziness and Uncertainty in Spatiotemporal Applications", CHOROCHRONOS Technical Report CH-00-04, Feb. 2000.

[Pfos00b] Dieter Pfoser, Yannis Theodoridis, Generating Semantics-Based Trajectories of Moving Objects", Technical Report, CHOROCHRONOS:TM Research Network Project, January, 2000.

[Sist97] A. P. Sistla, O.Wolfson, S.Chamberlain, S. Dao, "Modeling and querying moving objects", Proceedings of the 13 th International Conference on Data Engineering, pages 422-432, Birmingham, UK, April 1997.

[Wolf98] O.Wolfson, Bo Xu, S.Chamberlain, and L.Jiang, "Moving Objects Databases: Issues and Solutions", In Proc. Of the 10th Intl. Conference on Scientific and Statistical Database Management, Capri, Italy, 1998.

[Wolf99c] O. Wolfson, A. P. Sistla, S. Chamberlain, and Y. Yesha, "Updating and querying databases that track mobile units", Distributed and Parallel Databases, Vol. 7, No.3, 1999.