이동 객체 궤적의 최소경계사각형 영역을 효율적으로 분할하는 알고리즘에 관한 연구

박주현* · 조우현**

A Study on Efficient Split Algorithm for Minimum Bounding Box of Moving
Object Trajectory

Ju-Hyun Park* · Woo-Hyun Cho**

요 약

최근, 무선 네트워크의 발달로 인해 이동 객체에 대한 위치 정보를 수집하여 실생활에 활용하는 다양한 위치 기반 서비스의 증가하고 있다. 그에 따라서, 이동 객체의 연속적인 위치를 효율적으로 검색하는 새로운 색인 구조가 필요하게 되었다.

본 논문에서는 이동 객체의 좌표 사이의 거리가 긴 경우 탐색 공간을 줄이기 위해 효율적으로 분할하는 방법을 제안한다. 궤적의 적절한 분할 위치를 찾아서 평균적인 질의의 크기를 고려하여 형성되는 확장된 최소 경계 사각형(EMBR)의 영역을 이용한다. 추정 분할 방법은 최소 경계 사각형을 최소화하게끔 고안되었고 이를 실험하였다. 실험 결과 제안하는 추정 분할 방법이 기존의 방법에 비해서 EMBR의 면적을 더 효율적으로 줄여 줄 수 있었다.

ABSTRACT

With the recent development of wireless network technologies, there have been increasing usage of various position base services. Position based services basically collect position information of moving object for the utilization of them in real life. Accordingly, new index structures are required to efficiently retrieve the consecutive positions of moving objects.

In the paper, we consider volume of Extended Minimum Bounding Rectangles(EMBR) to be determined by average size of range queries. We proposed the method that split efficiently moving object with long distance between location, and split moving object for decrease searching space an Estimated-Split algorithm that minimizes the volume of MBRs is designed and simulated. Our experimental evaluation confirms the effectiveness and efficiency of our proposed splitting policy

키워드

이동 객체, 궤적 분할, 시공간 색인

Key word

Moving Object, Trajectory Splitting, Spatiotemporal Indexing

** 정회원 : 부경대학교(주저자, holy4001@empal.com)

** 정회원 : 부경대학교(교신저자)

접수일자 : 2012. 11. 03

심사완료일자 : 2012. 11. 28

I. 서 론

이동통신, 무선 네트워크, GPS(Global Position System)의 발전으로 인해 이동 객체들의 위치 정보를 수집하고 저장하여 다양한 어플리케이션과 여러 가지 위치 정보 서비스 분야에 활용되고 있다.

이동 객체란 시간의 변화에 따라 공간적인 위치가 연속적으로 변화는 시공간 데이터이고 이러한 위치 정보 데이터를 저장하고 효율적으로 검색하기 위해서 여러 가지 색인 기법들이 연구되었다.

기존의 색인 방법을 보면 크게 3가지로 나눌 수 있는데 첫 번째로 질의의 종류에 따라 이동 객체의 이력 정보를 검색하기 위한 방법[2, 4]과 두 번째로 이동 객체의 궤적을 검색하기 위한 방법[1, 3, 5], 그리고 세 번째로 이동 객체 궤적을 분석하여 미래를 예측하기 위한 방법[6, 7]으로 나눌 수 있다.

본 논문에서는 두 번째의 이동 객체의 궤적을 검색하기 위한 방법에서 궤적 기반 데이터를 검색하기 위해 궤적 기반 질의에 초점을 맞추고 있다.

궤적 기반 질의는 궤적들이 서로 교차하는 위상적인 관계나 속도와 이동 방향 등의 정보를 이용하여 최소 경계사각형(MBR : Minimum Bounding Rectangle)을 구성하여 공간 질의와 MBR이 교차하는지의 유무에 따라 질의에 포함된다. 따라서 이 MBR을 어떻게 구성하는지가 공간 질의를 처리하는데 있어 아주 중요한 요소이다. 하지만 이동 객체상의 모든 궤적의 포인트를 MBR로 구성하고 이를 검색하는 것은 시간이 많이 소요된다. 그리고 궤적의 포인트와 포인트의 거리가 너무 길 경우에는 거짓버림(False Drop)이 너무 많이 발생을 한다. 따라서 본 논문에서는 검색 시간 단축과 거짓버림을 줄이기 위한 궤적의 효율적인 분할 방법을 제안한다.

제안하는 방법은 제시된 공식을 이용하여 모든 궤적의 포인트와 포인트 사이의 거리가 길 경우 새롭게 궤적을 분할한다. 여기에 질의범위의 크기를 고려한 확장된 최소경계사각형(EMBR : Extended MBR)과 우선순위 큐(Priority Queue)를 이용하여 전체 EMBR의 면적이 최소가 되게 궤적을 분할하는 방법이다.

이 추정 분할(Eliminated Split) 알고리즘은 기존에 제안한 개선된 분할 알고리즘[9]에 궤적의 위치 사이가 먼 경우를 강제 분할하는 추정 공식을 사용하여 만들어졌다. 이 알고리즘을 사용하면 색인 구성 후 공간 질의에

대한 거짓버림이 줄어들고 불필요한 탐색공간을 감소시킬 수 있다. 성능 평가 결과로 합병 분할 방법(Merge Split)[8]과 비교해 본 결과 EMBR의 총합에 있어서 더 최적의 분할이 된다는 것을 알 수 있었다.

본 논문의 구성은 다음과 같다. 2절에서는 알고리즘에 필요한 관련 연구를 설명하고, 3절에서는 연속적인 두 개의 포인트 사이를 분할하는 이론적인 공식과 효율적인 분할을 만드는 추정 분할 알고리즘에 대해 제안한다. 4절에서는 제안하는 알고리즘과 기존의 알고리즘과의 성능 평가를 통해 성능을 비교 분석한다. 5절에서는 결론 및 향후 연구 과제에 대해 기술한다.

II. 관련 연구

공간 질의가 주어졌을 때 이동 객체에 포함되어 있는 궤적의 데이터를 검색하기 위해 MBR을 사용한다.

2.1. MBR

MBR은 아래의 그림 1과 같이 2차원의 궤적이 주어졌을 때 속도나 방향을 고려하여 현재의 위치 P_i 와 다음의 위치 P_{i+1} 사이를 사각형으로 나타낸다. 이 사각형 영역과 공간 질의가 겹치는지에 따라서 질의 결과가 반환된다.

그림 1. 최소경계사각형
Fig. 1 Minimum Bounding Rectangle

2.2. EMBR

EMBR은 범위 질의가 주어졌을 때 질의가 MBR과 중첩하는지 여부를 찾아내기 위하여 MBR의 영역을

공간 질의의 크기의 반만큼 확장을 해 놓은 영역이다. 이 영역을 비교 검색함으로써 공간 질의에 대한 결과가 반환된다.

본 연구에서는 3차원인 이동 객체 궤적의 위치 정보를 간단하게 나타내기 위해서 궤적을 위치 정보 x 와 시간 t 의 2차원 공간으로 나타낸다.

EMBR면적의 계산은 그림 2와 같이 주어지는 범위 질의 q_x, q_t 를 MBR의 범위에서 $qx/2, qt/2$ 만큼 MBR에서 확장된 사각형 면적을 계산한다.

그림 2. 질의의 크기를 고려한 확장된 최소 경계 사각형
Fig. 2 Extended MBR considered Size of range query

III. 추정분할 알고리즘

3.1. 추정분할 공식

그림 3. 하나의 MBR을 분할한 형태
Fig. 3 Form of splitting one MBR

위의 그림 3과 같이 m_x 와 m_t 로 이루어진 MBR이 있다면 임의의 분할 개수가 S 일 때 q_x 와 q_t 로 이루어진 공간 질의를 적용한 실선의 형태로 표시된 EMBR들의

총합을 $g(S)$ 라 둔다. 그러면 아래의 공식이 성립된다. $g(S)$ 는 m_x 를 분할 개수 S 로 나눈 크기에 q_x 를 더한 값과 m_t 를 분할 개수 S 로 나눈 크기에 q_t 를 더한 값을 곱하면 분할된 하나의 EMBR의 면적이 연산되고 이를 분할 개수만큼 곱하면 분할된 전체의 면적과 같다.

$$\begin{aligned} g(S) &= S \left(\frac{m_x}{S} + q_x \right) \left(\frac{m_t}{S} + q_t \right) \\ &= S \left(\frac{m_x m_t}{S^2} + \frac{m_x q_t}{S} + \frac{m_t q_x}{S} + q_x q_t \right) \\ &= q_x q_t S + m_x q_t + m_t q_x + \frac{m_x m_t}{S} \end{aligned}$$

$g(S)$ 가 최소가 되는 S 를 구하기 위해 미분한다.

$$g'(S) = q_x q_t - m_x m_t \frac{1}{S^2} = 0$$

※ $g''(S) = 2m_x m_t \frac{1}{S^3} > 0$ 이므로 $g'(S) = 0$ 인 S 에

대해 $g(S)$ 는 최소값을 갖는다. 따라서

$$S^2 = \frac{m_x m_t}{q_x q_t}$$

∴ $S = \sqrt{\frac{m_x m_t}{q_x q_t}}$ 일 때 $g(S)$ 는 최소값이 된다.

따라서 위의 공식을 P_i 과 P_{i+1} 으로 만들어진 MBR에 적용하면 가장 최적의 분할의 개수를 구할 수 있다.

3.2. 추정분할 알고리즘

추정 분할 알고리즘은 그림 7과 같다. 이 방법은 첫 번째 과정 1)에서 그림 1과 같이 포인트 $P_0 \sim P_n$ 까지 P_i 와 P_{i+1} 의 높이와 넓이를 계산한 다음 추정 알고리즘의 공식에 대입하여 분할 개수가 2 이상이 되면 P_i 과 P_{i+1} 사이의 간격을 분할한다.

과정 2)에서 그림 4와 같이 새롭게 분할된 포인트 $P_0 \sim P_n$ 에서 연속적으로 인접한 두 개의 포인트를 이용하여 MBR을 생성한다.

과정 3)에서 그림 5와 같이 이웃한 두 EMBR의 값과 이 EMBR의 전체 값을 계산해서 $KEY_{(i,j)}$ 를 만든 후

차례대로 우선순위 큐에 삽입한다.

다음 네 번째 과정 4)에서 정렬된 우선순위 큐에서 첫 번째로 반환되는 키 값 $KEY_{(i,j)}$ 을 가지고 MBR 집합을 갱신한 뒤 다시 우선순위 큐를 재구성한다. 더 이상 음의 수를 가지는 키 값이 더 이상 존재하지 않을 때까지 4)의 과정을 반복 수행한 후 그림 6과 같이 최종적인 EMBR 집합을 만들어 낸다.

그림 4. 추정 분할 공식을 적용하여 분할된 궤적
Fig. 4 Splitted trajectory by applying The estimated split formula

그림 5. 궤적에 확장된 최소경계사각형의 적용
Fig. 5 Application of Extended MBR

그림 6. 최종 EMBR의 결과
Fig. 6 Result of Final EMBR

[정리] Algorithm-1의 시간 복잡도는 $O(n \log n)$ 이다.
증명.

1)에서 P_i 과 P_{i+1} 사이의 높이와 넓이를 계산하고 공식을 적용하기 위한 연산 횟수는 n 이고, 2)에서 인접한 두 개의 포인트를 이용하여 MBR을 생성시키는데 필요한 연산의 횟수는 n 이다. 3)에서 하나의 키 값을 계산하여 우선순위 큐에 삽입하는 연산 횟수는 n 이다. 4)에서 EMBR 집합의 합병과 새로운 우선순위 큐의 재구성을 위한 연산 횟수는 $n \log n$ 이다. 그래서 알고리즘의 총 연산 횟수는 $n + n + n + n \log n$ 이다. 그러므로 이 알고리즘의 시간 복잡도는 $O(n \log n)$ 이다.

Algorithm : Estimated Split for a Trajectory

Input: A set of Trajectory's Points($P_0, P_1, \dots, P_n$), Query size.

Output: A set of MBRs that cover the Trajectory.

- 1) **For each** i ($0 \leq i \leq n-1$)
Calculate width, height between P_i to P_{i+1}
Apply the Estimated split formula
- 2) Construct full split MBRs $M_i[o, n]$
- 3) **For each** i ($0 \leq i \leq n-1$)
Compute EMBRs volume for merging $M[i, i+1]$ with $M[i+1, i+2]$
store $KEY_{(i,i+2)}$ in a priority queue such that following holds:
 $KEY_{(i,i+2)} = E[i, i+2] - (E[i, i+1] + E[i+1, i+2])$
- 4) While ($KEY_{(i,j)}$ in the root node of the priority queue is negative)
Use the priority queue to merge the pair of consecutive $M_i[o, n]$ that give the biggest decrease in EMBR volume;
Update the priority queue with new merged MBRs;

그림 7. 추정 분할 알고리즘
Fig. 7 Estimated Split Algorithm

IV. 실험 및 성능 비교

본 논문에서 제안하는 이동 객체 궤적에 대한 추정 분할 방법의 성능실험은 인텔 i5 2.8GHz 프로세서와 메모리 1Gbyte, Windows 7 운영체제를 사용하는 시스템상에서 수행했으며, 알고리즘의 구현을 위해서 C++ 언어를

사용하였다. 그리고 성능 비교 평가를 위하여 기존의 합병분할 방법과 추정 분할 알고리즘을 구현하여 실험을 하였다. 본 실험에서는 이동 객체의 궤적 포인트의 개수를 2000개로 제한하였고 공간 질의의 크기를 10×10 , 20×20 , 30×30 으로 주어서 여러 크기의 질의에 따라 EMBR의 총면적이 어떻게 줄어드는지 실험을 했다. 실험에 사용된 위치 축인 x 축은 상하의 위치 변화를 50, 100, 150으로 주었고 t 축은 시간변화가 규칙적일 때 200으로 하고 불규칙적일 때 100~200으로 두어서 포인트의 생성을 임의의 생성하여 실험에 적용했다.

첫 번째 실험의 결과는 위치변화를 50, 100, 150으로 시간 구간은 200으로 하고 포인트를 2000개 생성하여 공간 질의의 크기가 10×10 , 20×20 , 30×30 일 경우 추정 알고리즘을 적용했을 때와 분할 알고리즘을 적용했을 때의 EMBRs의 총면적을 계산하여 비교하였다.

그림 8-10은 이동 객체의 위치 변화를 다양하게 변화하게 하여 합병 분할 알고리즘과 추정 분할 알고리즘을 적용하여 최종 EMBR의 총면적의 값을 도식화한 것이다. 그림 8의 도표는 이동 객체의 좌표 (x, t) 에서 t 의 좌표는 200로 일정하게 변화시키고 이동 객체의 현 위치에서 다음 위치로의 x 좌표의 변화를 상하 50 이내로 제한하여 좌표를 생성시킨 후 추정 분할 알고리즘을 적용하여 EMBR의 총면적을 계산 후 도식화 시킨 그림이다. 그림 9은 x 좌표의 변화를 100으로, 그림 10은 x 좌표의 변화를 150로 주어서 이동 객체 궤적의 좌표를 생성시킨 후 EMBR의 총면적을 계산한 결과를 도식화한 것이다.

그림 8. 위치변화 정도 50이내,
측정시간 구간 200으로 규칙적일 경우
Fig. 8 location change within 50
The regular case of measurement time 200

그림 9. 위치변화 정도 100이내,
측정시간 구간 200으로 규칙적일 경우
Fig. 9 location change within 100
The regular case of measurement time 200

그림 10. 위치변화 정도 150이내, 측정시간 구간
200으로 규칙적일 경우
Fig. 10 location change within 150
The regular case of measurement time 200

도표에서 보는 것과 같이 EMBR의 총면적이 더 작음을 알 수 있다. 이는 추정 분할 방법이 인접한 두 좌표 사이 거리를 효율적으로 분할했음을 알 수 있다.

두 번째 실험은 시간 구간의 좌표를 불규칙적으로 생성했을 때 각 알고리즘에 의해 만들어지는 EMBR의 총합을 시험하였다. 이동 객체의 좌표 (x, t) 에서 시간 구간의 좌표 이동 객체의 좌표 t 를 100~200 사이에서 불규칙적으로 생성하고 x 의 좌표를 위치 변화를 각각 상하 50, 100, 150로 제한하여 이동 객체의 위치 정보를 생성하여 실험을 했다. 그 외의 다른 조건은 첫 번째 실험과 동일한 조건을 주었다.

그림 11. 위치변화 정도 50이내, 측정시간 구간 100~200으로 불규칙적일 경우

Fig. 11 location change within 50 The irregular case of measurement time 100~200

그림 12. 위치변화 정도 100이내, 측정시간 구간 100~200으로 불규칙적일 경우

Fig. 12 location change within 100 The irregular case of measurement time 100~200

그림 13. 위치변화 정도 150이내, 측정시간 구간 100~200으로 불규칙적일 경우

Fig. 13 location change within 150 The irregular case of measurement time 100~200

그림 11-13의 도표는 두 번째 실험을 통한 각 궤적 분할 방법의 성능 비교 결과이다. 첫 번째 실험과 마찬가지로 제안하는 추정 분할 방법을 통해 만들어지는 EMBR 총면적이 합병 분할 방법의 EMBR의 총면적보다 더 효율적으로 줄어들음을 알 수 있다.

위의 결과에서 보는 것 같이 합병 분할과 추정 분할 두 알고리즘의 성능 비교 결과를 보면 이동 객체에 적용되는 공간 질의의 크기가 작을수록 좌표의 위치 변화가 클수록 본 논문에서 제안하는 추정 분할 방법을 이용해서 인접한 두 좌표 사이를 분할하는 것이 EMBR 총면적에 있어 합병 분할 방법을 통해 만들어지는 EMBR 총면적보다 더 효율적으로 줄어든다는 것을 알 수 있다. EMBR의 총면적이 작다는 것은 필요 정보를 탐색한데 있어 탐색 공간이 더 줄어든다는 것이고 거짓버림 역시 줄어든다는 것이다. 이 결과는 본 논문에서 제안하는 추정 분할 알고리즘이 다른 방법에 비해 탐색공간을 더욱 효율적으로 줄여 주고, 평균적인 질의의 크기를 고려하여 거리가 긴 궤적을 분할하는 것이 더 이점이 있다는 것을 의미한다.

V. 결 론

본 논문에서는 검색 범위와 거짓버림을 줄이기 위해 궤적 정보를 효율적으로 줄이는 알고리즘을 제안하였다. 성능 평가의 결과를 보면 제안하는 궤적 분할 알고리즘은 길이가 긴 연속된 두 포인트 사이의 간격은 제시된 공식을 사용하여 분할함으로써 전체 EMBR의 면적을 줄일 수 있음을 알 수 있다. EMBR의 면적이 줄어들면 탐색을 위한 공간이 줄어들고 탐색 시간이 감소한다는 것을 의미한다. 또한 거짓버림이 줄어들게 되므로 검색의 오차를 줄일 수 있으므로 효율적인 데이터 처리를 할 수 있음을 알 수 있다. 그리고 이를 바탕으로 색인 구성을 더 효율적으로 할 수 있고 다양한 공간 질의에 대해서 효율적인 질의 처리를 할 수 있음을 알 수 있다.

향후 논문 과제는 색인을 만들기 위해 단일 궤적이 아닌 다양한 형태의 시공간 궤적들에 대하여 탐색 공간을 줄이고 검색의 효율을 높일 수 있는 분할 방법에 대한 실험을 수행할 예정이다.

참고문헌

- [1] D. Pfoser, C. S. Jensen and Y. Theodoridis, "Novel Approaches to the Indexing of Moving Object Trajectories", Proc of The 26th International Conference on Very Large Data Bases, pp. 395-406, 2000
- [2] X. Xu, J. Han and W. Lu, "RT-Tree: An Improved R-Tree Indexing Structure for Temporal Spatial Databases", Proc of The International Symposium on Spatial Data Handling, pp. 1040-1049, 1990
- [3] D. Pfoser, C. S. Jensen and Y. Theodoridis, "Novel Approaches to the Indexing of Moving Object Trajectories" Proc of The 26th International Conference on Very Large Data Bases. pp. 395-406. 2000
- [4] Y. Theodoridis, M. Vazirgiannis and T. Sellis, "Spatio-Temporal Indexing for Large Multimedia Application", Proc of The 3rd IEEE International Conference on Multimedia Computing and Systems, pp. 441-448, 1996
- [5] J. Nievergelt, H. Hinterberger and K. C. Sevcik, "The Grid File: An Adaptable, Symmetric Multikey File Structure", ACM Transactions on Database System, pp. 38-71, 1984.
- [6] T. Tzouramanis, M. Vassilakopoulos, "Overlapping linear quadtrees: a spatio-temporal access method", Proc of the 6th International Symposium on Advances in Geographic Information System, pp. 1-7, 1998.
- [7] V. P. Chakka, A. C. Everspaugh and J. M. Patel, "Indexing Large Trajectory Data Sets With SETI", Proc of The 1st Biennial Conference on Innovative Data Systems Research, 2003
- [8] S. Prabhakar, Y. Xia, D. V. Kalashnikov, W. G. Aref and S. E. Hanbrusch, "Query Indexing and Velocity Constrained Indexing: Scalable Techniques for Continuous on Computers", IEEE Transactions on Computers, pp. 1124-1140, 2002
- [9] 전현준, 박주현, 박희숙, 조우현, "이동 객체 궤적의 색인을 위한 개선된 분할 알고리즘", 정보처리학회 논문지D, v.16D, No2, pp

저자소개

박주현(Ju-Hyun Park)

2000년 경남대학교 컴퓨터공학과 (공학사)

2002년 경남대학교 대학원 컴퓨터공학과 (공학석사)

2013년 부경대학교 컴퓨터공학과(공학박사)

※관심분야: 공간 데이터베이스, 객체지향 데이터베이스

조우현(Woo-Hyun Cho)

1985년 경북대학교 전자공학과 전산공학전공(공학사)

1988년 경북대학교 대학원 전자공학과 (공학석사)

1998년 경북대학교 대학원 전자공학과 전산공학전공(공학박사)

1989년-현재 부경대학교 공과대학교 컴퓨터공학부 교수

※관심분야: 지능형 데이터베이스, 멀티미디어 인덱싱, 객체 데이터베이스 관리 기술