

시공간 데이터베이스에서 선택도 추정 연구[†]

* 신병철, * 김차돌, * 이종연^o

[†] 충북대학교 컴퓨터교육과, * 충북대학교 컴퓨터공학과

suemirr@nate.com, suph92@lycos.co.kr, jongyun@chungbuk.ac.kr^o

A Study on the Selectivity Estimation in Spatio-Temporal Databases

[†]Byung-Chul Shin, *Cha-Dol Kim, and *Jong-Yun Lee^o

[†]Department of Computer Education, Chungbuk National University

*Department of Computer Engineering, Chungbuk National University

요약

최근 시공간 데이터에 대한 관심이 늘어나면서 시간에 따라 이동하는 객체에 대한 처리와 이력 공간 정보를 처리하는 영역에 대한 연구가 활발하게 일어나고 있다. 이러한 시공간 데이터베이스에서의 효과적인 질의 처리를 위해 선택도 추정 기술의 요구가 생겨났다. 본 논문에서는 이력 공간 정보에 대한 선택도 추정을 위하여 T-Minskew 히스토그램을 구축하고 이를 이용하여 선택도 추정 방법을 제시한다. 또한 임계치 기법을 이용한 효과적인 히스토그램 유지 기법을 제시한다.

1. 서론

최근 시공간 데이터에 대한 관심이 늘어나면서 이를 사용하는 여러 애플리케이션이 개발되었고 시공간 DBMS의 필요성이 대두 되었다. 이러한 시공간 DBMS는 시간에 따라 변화하는 공간객체들의 변화를 효과적으로 관리할 수 있어야 한다. 시공간 DBMS의 영역에는 크게 이동 객체에 대한 부분과 이력 공간정보에 대한 부분이 있다. 이동 객체에 대한 질의는 현재의 객체 정보에 기반하여 미래의 어느 시점에서 이 객체의 공간 정보가 어떻게 변화하는 것인가를 예측하는 것이다. 예를 들면, "Find all objects that overlap with query area A during 30 minutes."과 같은 질의를 들 수 있다. 이력 공간 질의는 과거의 어떤 시점에서 질의의 공간 영역과 겹치는 객체들을 찾아내는 것이다. 예를 들면, "Find all objects that overlap with query area A at time t"과 같은 질의를 들 수 있다.

시공간 선택도 추정은 질의 최적화에 매우 중요하다. 시공간 선택도 추정을 하기 위해서는 실제 존재하거나 존재했던 공간 데이터를 기반으로 요약 정보를 만들고 이를 선택도 추정에 이용한다. 공간정보를 이용하여 선택도 추정을 하는 기술에는 [1], [2], [3], [4] 등 이미 많은 기술들이 나와 있고 시공간 선택도 추정 기술에는 [5], [6], [7], [8]이 나와 있다.

[5], [6], [7], [8]과 같은 기존의 시공간 데이터베이스 선택도 추정은 이동 객체에 대한 연구가 주로 이루어 졌으나 본 논문에서는 히스토그램을 이용하여 이력 공간 정보를 유지하도록 한다. 본 논문에서 제안하는 히스토그램의 특징은 다음과 같다. (i) Minskew 히스토그램을 timestamp 별로 구축하고 유지한다. (ii) 너무 많은 히스토그램 재구축을 방지하기 위해 히스토그램 임계치를 두어 재구축 횟수를 줄이면서 만족스러운 선택도 추정률을 유지하도록 한다. 본 논문에서 제안하는 히스토그램은 기본적으로 2차원 객체의 특정 시점에 대한 질의를 다루고 있으며 향후 연구에서 범위 질의를 다룰 수 있는 히스토그램으로 확장할 것이다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구를

기술하고 3장에서는 본 논문에서 제안하는 T-Minskew 히스토그램 구축과 이를 이용한 선택도 추정 기법을 기술한다. 4장에서는 제안한 기술을 위한 실험 평가 및 결론을 내린다.

2. 관련 연구

2.1 기존의 시공간 선택도 추정

[5]에서는 이동하는 객체가 어떤 시간동안 고정된 질의 영역에 겹칠 수 있는지 여부에 초점을 맞추고 있다. 선택도 추정을 위하여 우선 전체 공간을 Minskew 알고리즘을 사용하여 버킷으로 분할하고 각 버킷에 대한 선택도 추정을 한 후 이를 모두 합하여 전체 공간에 대한 선택도를 추정하는 기술을 제시하고 있다. 2차원 공간 선택도 추정은 각 차원 별로 질의와 이동 객체를 사상시켜 1차원 환경에서 선택도를 추정한 뒤 각 차원별로 구해진 선택도를 곱하여 2차원 공간에서의 선택도 추정을 한다. 따라서 [5]에서 제시하는 선택도 추정 기술은 2차원 공간에서는 겹치지 않는 객체가 1차원 공간으로 사상하여 겹칠 수 있는 가능성을 가지기 때문에 다차원 공간에서의 선택도 추정에 좋지 못한 성능을 보인다. 만약 객체의 속도가 [0, V]에 존재하고, 공간적으로 [0, U]내에 균일하게 분포되어 있다면 질의는 T 시간동안 질의의 공간 영역과 겹치는 객체들을 찾아내는 것이다. 이 때 실제적인 선택도 Sel은 [6]에서 다음과 같이 제시하고 있다.

$$Sel = \frac{VT}{U^2} [(q_{R.xmax} - q_{R.xmin}) + (q_{R.ymax} - q_{R.ymin})] + \frac{(q_{R.xmax} - q_{R.xmin})(q_{R.ymax} - q_{R.ymin})}{U^2} \quad (1)$$

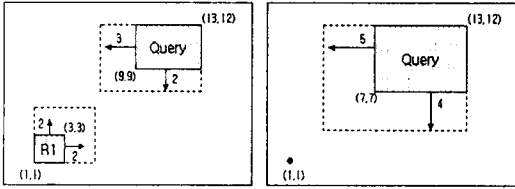
이 때 [5]에서 제시한 방법으로 선택도 추정을 한다면 각 차원 별로 수식(2)와 같은 선택도 추정이 가능하고 이를 곱하면 수식(3)과 같아지므로 선택도가 원래보다 높게 추정이 됨을 알 수 있다.

$$Sel_i = \frac{q_{m+} - q_{m-}}{U} + \frac{VT}{2U} \quad (2)$$

[†] 이 논문은 2004년도 충북대학교 학술연구지원사업의 연구비 지원에 의해 연구되었음.

$$Sel = \frac{VT}{U^2} [(q_{R,xmax} - q_{R,xmin}) + (q_{R,ymax} - q_{R,ymin})] + \frac{(q_{R,xmax} - q_{R,xmin})(q_{R,ymax} - q_{R,ymin})}{U^2} + \frac{V^2 T^2}{4U^2} \quad (3)$$

[6]에서는 [5]에서 제시한 이동 객체 표현을 2차원 공간으로 확장시켰다. 그리고 [5]에서의 초과 선택도 추정에 대하여 2차원 공간을 1차원 공간으로 사상시키는 것을 회피함에 따라 선택도가 높아지는 현상을 해결하였다. 어떤 객체의 현재 시간 0에서의 위치를 (x, y)라 하고 각 축에 따른 속도를 (v_x, v_y)라 할 때 [5]는 Minskew 기술을 사용하여 4차원 점 (x, y, v_x, v_y)을 표현할 수 있는 4차원 히스토그램을 생성하였다. 이 때 히스토그램의 각 버킷은 영역 MBR과 속도 MBR인 VMBR을 가지며 버킷안의 객체들은 영역 MBR과 VMBR안에서 균일하게 분포되도록 히스토그램을 구축한다. 이동하는 사각형 객체와 질의에 대해서는 그림 2.1와 같이 객체들을 간소화하는 기술을 사용하여 풀어내고 있다. 히스토그램의 재구축은 전체 데이터집합에서의 갱신률이 정해진 한계 값을 초과할 경우에만 일어나므로 재구축 빈도가 [5]에 비하여 줄어들었다.

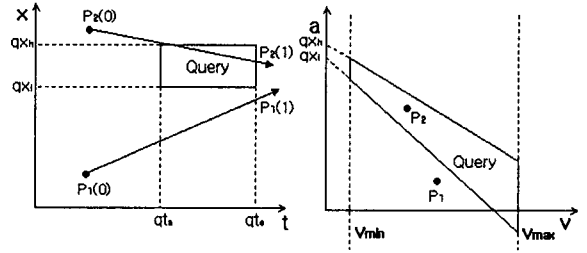


(a) 이동 객체와 질의 (b) 이동 객체의 간소화
그림 2.1 이동 객체의 간소화 과정

[7]에서는 공간-시간 그래프를 Hough 변환을 통한 속도-절편 그래프로 변형한 뒤 MinSkew 기술을 적용하여 이동 객체에 대한 선택도를 추정한다. 공간-시간 그래프에서 점의 이동에 따른 궤적은 속도-절편 그래프에서 하나의 점으로 표현될 수 있으며 질의 사각형은 어떤 공간 영역으로 표현된다. 만약 속도-절편 그래프에서 점이 질의 영역에 포함된다면 공간-시간 그래프에서 그 점에 해당하는 점의 궤적인 선은 질의 사각형을 지나간다고 할 수 있다. 그림 2.2는 공간-시간 그래프와 속도-절편 그래프의 예를 보인다. [7]에서의 선택도 추정은 점 객체들이 존재하는 전체 공간을 절편-속도 그래프에서 MinSkew 알고리즘을 사용하여 버킷들로 분해한 뒤 질의 영역과 겹치는 버킷들의 영역을 계산함으로써 구할 수 있다.

2.2 Minskew

[1]에서는 공간 선택도 추정을 위한 히스토그램인 Minskew를 제안하고 있다. Minskew 히스토그램은 편중된 객체들의 버킷 분할을 통한 분배를 통하여 균일하게 만드는데 있다. 버킷 분할의 기준은 객체들의 편중도 Skew에 기반하고 있으며 분할 가능한 경우의 수중에 분할되는 두 버킷의 편중도의 가중치가 가장 낮은 분할을 선택하여 이진 공간 분할(BSP)을 한다. 각 축별로 분할 가능한 모든 경우를 검사하기에는 많은 복잡도를 가질 수 있으므로 분할 적합 축을 우선 선택한 후에 그 축을 기준으로 분할을 하는 것으로 분할 알고리즘의 효과를 높이고 있다. B_i.num은 i번째 버킷에 포함되는 점 객체의 수 또는 사각형 객체의 중심점이 버킷의 영역에 포함되는 수를 저장한다. C_i.den은 셀을 가리키며 C_i.den은 i셀에 겹치는 객체수를 저장한다. Avg(den)은 셀의 평균 den수를 말하고 |C|는 셀의 수를 가리킨다. 이 때 버킷의 편중도 B_i.skew는 수식(4)과 같이



(a) 공간-시간 그래프 (b) 속도-절편 그래프

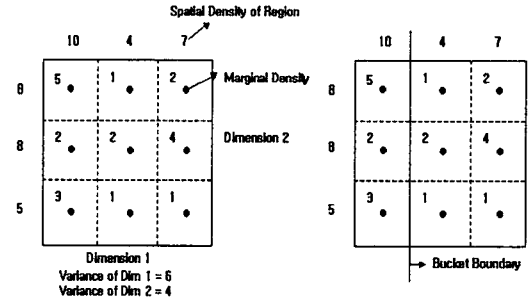
그림 2.2 Hough 변환을 사용한 그래프 변환

표현하며 전체 편중도의 가중치는 수식(5)로 표현한다. 최종 분할 경우의 선택은 가중치가 가장 작은 것으로 한다.

$$B_i.skew = \frac{1}{|C|} \sum_{i=1}^{number\ of\ cell\ in\ Bi} (C_i.den - Avg(den))^2 \quad (4)$$

$$Minskew = \sum_{i=1}^n (B_i.num \times B_i.skew) \quad (5)$$

그림 2.3는 Minskew 히스토그램에서 버킷 분할의 예를 보이고 있다. 그림 2.3a를 보면 Dim 1의 분산은 6이고 Dim 2의 분산은 4임을 알 수 있다. 이것은 분산이 높은 축이 객체의 편중도가 높다는 것이고, 따라서 이러한 편중도를 낮추기 위하여 편중도가 높은 축을 기준으로 공간을 분할하는 것이 좋다는 것을 미리 알 수 있다. 그림 2.3b는 선택된 분할 축을 기준으로 가능한 분할 경우의 수에서 수식(4)을 이용하여 분할된 두 버킷의 skew를 구하고 수식(5)를 이용하여 최종 가중치를 구한 뒤 가중치가 가장 작은 분할 선을 기준으로 공간을 두개의 버킷으로 분할한 것이다. 수식(5)를 이용하여 각 분할 가능한 경우의 가중치를 구하면 첫 번째 경우의 가중치는 28.14가 나오며 두 번째 경우는 37.38이 나온다. 따라서 첫 번째 분할이 편중도를 낮출 수 있는 방법임을 알 수 있고 그림 2.3b와 같다.



(a) (b)
그림 2.3 Minskew 히스토그램의 공간 분할

3. T-Minskew 히스토그램

이 장에서는 T-Minskew 히스토그램을 이용한 이력 공간 질의의 선택도 추정 방법을 제안한다. T-Minskew 히스토그램은 2차원 공간 구조에 시간을 고려한 히스토그램이다. 히스토그램에서 시간 정보를 가지고 있는 것은 timestamp에 따른 각각의 히스토그램 밖에 없고 히스토그램내의 버킷은 명시적으로 공간 정보를 가지면서 히스토그램의 timestamp내에 존재한다는 것을 보장한다. 그림 3.1

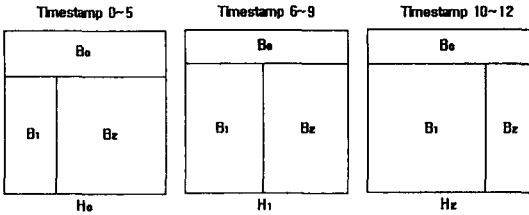


그림 3.1 T-Minskew 히스토그램

은 간단한 T-Minskew 히스토그램의 timestamp 0부터 12까지 형태의 예를 보이고 있다. timestamp 0에 히스토그램 H_0 가 생성되어 5까지 지속되다가 timestamp 6에 새로운 히스토그램 H_1 이 생성되고 timestamp 10에 히스토그램 재구축에 의해 다시 새로운 히스토그램 H_2 가 생성되었음을 보이고 있다.

본 논문에서는 기본적으로 사각형 질의 Qr 에 대한 선택도 추정을 하고 있다. 또 timestamp별로 히스토그램을 생성하므로 여러 개의 히스토그램이 존재하고 각 히스토그램은 유효 시간 $[t_{start}, t_{end}]$ 를 가지고 있다. 각각의 히스토그램 내에는 여러 버킷이 존재하고 공간영역 정보와 버킷이 포함하는 객체수를 저장한다. 전체 선택도 Sel 은 각각의 버킷에 대한 선택도 추정 Sel_i 을 합하여 구한다. 다음 표 3.1은 본 논문에서 사용하고 있는 기호들에 대한 정리를 하고 있다.

표 3.1 기호 테이블

기호	의미
Qr	윈도우 질의
$Qr.t$	윈도우 질의의 timestamp
$B_i.MBR = [B_i.x_{min}, B_i.y_{min}, B_i.x_{max}, B_i.y_{max}]$	i번째 버킷의 2차원 공간 좌표
$B_i.num$	i번째 버킷이 포함하는 객체 수
H_i	i번째 히스토그램
$H_i.t = [H_i.t_{start}, H_i.t_{end}]$	i번째 히스토그램의 유효시간
$H_i.var$	i번째 히스토그램의 객체 변화량
$timestamp_{now}$	현재시간
Sel_i	i번째 버킷에 대한 선택도
Sel	전체 선택도
$OverlapArea(B_i.MBR)$	i번째 버킷과과 겹치는 질의 영역

3.1 T-Minskew 히스토그램 구축

[가정 1] 히스토그램 H_i 의 시간 간격은 H_i 가 생성이 된 시점부터 $i+1$ 번째 재구축이 일어난 $timestamp_{now} - 1$ 까지로 한다.

히스토그램 H_0 가 실제 객체를 기반으로 timestamp 0에 구축되었을 때 유효 시간은 $[0, NOW]$ 을 가진다. 시간이 지나 timestamp가 1이 되었을 때 재구축을 해야 되는 지에 대한 판단을 하게 되는데 객체의 변화율과 임계치를 비교함으로써 히스토그램 재구축 여부를 결정한다. 만약 timestamp 1에서의 객체 변화율이 임계치를 넘어서지 않았을 경우는 timestamp 0에 구축된 히스토그램 H_0 는 timestamp 1에도 유지가 되며 유효시간은 여전히 $[0, NOW]$ 로써 변화가 없게 된다. timestamp 4에서 객체 변화율이 임계치를 넘어섰을 경우 히스토그램의 유효시간의 끝을 timestamp 3으로 기록하여 유효시간을 $[0, 3]$ 으로 갱

신하고 timestamp 4에는 이 시점에서 살아 있는 객체들을 기반으로 새로운 히스토그램 H_1 를 생성한 뒤 유효시간을 $[4, NOW]$ 로 할당한다.

그림 3.2는 이러한 히스토그램 재구축 알고리즘이다. 단계 1에서 이전 timestamp까지의 객체 변화량에 현재 timestamp의 변화량을 더한다. 단계 2에서는 변화량에 전체 객체수를 나눈 값이 임계치를 넘어섰는지를 검사한다. 만약 임계치를 넘어 선다면 단계 3에서 i번째 히스토그램의 유효 시간의 끝인 $H_i.t_{end}$ 를 현재 timestamp의 바로 이전으로 기록하고 단계 4에서 새로운 히스토그램 H_{i+1} 을 생성한다. 단계 5와 6에서 각각 새로운 히스토그램 H_{i+1} 의 유효시간을 기록하고 알고리즘을 종료한다.

Algorithm rebuildHistogram

```

1   $H_i.var +=$  calculate variation of objects
    in universal space
    during  $[timestamp_{now} - 1, timestamp_{now}]$ 
2  if  $(H_i.var / N)$  over threshold
3   $H_i.t_{end} = timestamp_{now} - 1$ 
4  Create new histogram  $H_{i+1}$ 
5   $H_{i+1}.t_{start} = timestamp_{now}$ 
6   $H_{i+1}.t_{end} = *$ 
    end if
End rebuildHistogram
    
```

그림 3.2 히스토그램 재구축 알고리즘

히스토그램 재구축을 결정짓는 객체의 변화율은 히스토그램이 생성된 시점부터 실제 객체의 생성, 삭제, 갱신 횟수가 전체 객체 수에 비해 얼마나 일어났는지에 따른다. 예를 들어 전체 객체 수 N 이 100이고 히스토그램 유지 기간 동안의 객체 변화 횟수가 20 이라면 20%의 객체 변화율을 가지게 되며 만약 임계치가 15%라면 재구축이 일어난게 된다.

3.2 T-Minskew 히스토그램을 이용한 선택도 추정

시공간 이력 공간 질의에 대한 선택도 추정을 위해서는 우선 질의의 timestamp와 겹치는 히스토그램을 찾은 뒤 히스토그램 내에서 질의의 공간 영역과 겹치는 버킷들을 찾는다. 질의 공간 영역과 겹치는 각 버킷의 공간 영역을 해당되는 버킷의 전체 공간 영역으로 나누어 질의와 겹치는 공간 영역의 전체 공간 영역에 대한 비율을 구하고 이 비율에 버킷이 가지는 객체수를 곱함으로써 질의와 겹쳐지는 버킷내의 객체수를 추정할 수 있다. 마지막으로 각 버킷에서 추정된 객체수를 더함으로써 전체 선택도를 추정할 수 있게 된다. 수식 (6)과 (7)는 이러한 과정을 수식으로 나타낸 것이다.

$$Sel_i = B_j.num * \frac{OverlapArea(B_j.MBR)}{area(B_j.MBR)} \quad (6)$$

$$Sel = \sum_{j=0}^k Sel_j \quad (7)$$

그림 3.3은 T-Minskew를 이용한 선택도 추정 방법의 알고리즘을 나타내고 있다.

Algorithm Selectivity

```

1  Find a histogram  $H_i$  that satisfy
     $(H_i.t \supset Qp.t \text{ or } Qr.t)$ ,
    where  $(0 \leq i \leq \text{number of histogram})$ 
2  Find buckets that overlap
    with query area within  $H_i$ 
    
```

3 For $j=0$ to number of buckets overlapped with query
 4 $Sel_j = B_j\text{-num} * \text{OverlapArea}(B_j\text{-MBR})$
 / $\text{area}(B_j\text{-MBR})$
 end for
 5 $Sel = \sum_{j=0}^k Sel_j$, where k is
 number of buckets overlapped with query
 End Selectivity

그림 3.3 선택도 추정 알고리즘

단계 1에서 질의의 Timestamp를 포함하는 히스토그램을 찾고, 단계 2에서는 검색된 히스토그램 내에서 질의의 공간 영역과 겹치는 버킷들을 찾는다. 단계 3과 4에서는 질의와 겹치는 버킷의 선택도를 추정한다. 이 때 버킷 선택도를 추정 공식은 수식(6)과 같다. 단계 5에서 전체 선택도를 구하기 위해 각각 구해진 버킷의 선택도를 모두 합하는 것으로 알고리즘은 종료된다.

예를 들어 그림 3.4는 그림 3.1에서 질의의 timestamp와 겹치는 히스토그램 H_2 를 가져온 것이다. 질의 사각형의 공간 영역과 각 버킷의 정보는 다음과 같다.

$Qr.MBB = [5, 5, 10, 10]$
 $B_0.MBB = [0, 8, 10, 10], B_0.num = 3$
 $B_1.MBB = [0, 0, 7, 8], B_1.num = 9$
 $B_2.MBB = [7, 0, 10, 8], B_2.num = 2$

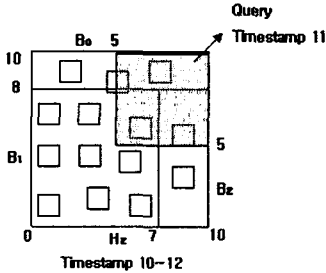


그림 3.4 T-Minskew를 이용한 선택도 추정

이 때 그림 3.4에서 각 버킷의 선택도를 추정과 전체 선택도 추정은 다음과 같이 계산된다.

$$Sel_1 = 1.5 = 3 * 10 / 20$$

$$Sel_2 = 0.96 = 9 * 6 / 56$$

$$Sel_3 = 0.75 = 2 * 9 / 24$$

$$Sel = \sum_{j=0}^3 Sel_j = 1.5 + 0.96 + 0.75 = 3.21$$

4. 실험 평가 및 결론

이번 절에서는 본 논문에서 제시한 T-Minskew 히스토그램을 위한 이론적인 실험 모델로써 오류율 계산법과 실제 실험 평가 기준을 제시한다.

실험을 위한 상대 오류율 계산법은 수식(8)와 같다. Sel은 T-Minskew를 이용하여 질의 결과를 추정한 값이고 Sel'는 실제 질의 결과값이다.

$$Err = (Sel - Sel') / Sel' \quad (8)$$

특정 질의에 대한 편중된 결과를 해결하기 위해 Q_n 개의 질의에 대한 선택도 추정 오류율을 측정하고 그것의 평균을 구함으로써 실험에 보다 높은 신뢰도를 가지도록 한다. 따라서 최종 오류율은 수식(9)과 같다.

$$Avg(Err) = (\sum_{i=1}^{Q_n} Err_i) / Q_n \quad (9)$$

만약 Minskew 히스토그램의 평균 오류율이 M_{err} 이라면

T-Minskew의 오류율은 M_{err} 보다 a 만큼 증가한다. 왜냐하면 임계치 기법에 의하여 히스토그램을 유지할 때 객체의 변화에 따른 히스토그램의 재구축을 하지 않고 이전 단계에서 사용했던 히스토그램을 그대로 쓰기 때문이다. 또한 추가된 오류율 a 는 높은 임계치일 수록 히스토그램의 재구축률이 감소하므로 증가하게 된다. 따라서 최종 T-Minskew의 오류율 TM_{err} 은 수식(10)과 같다.

$$TM_{err} = M_{err} + a \quad (10)$$

실제 실험은 다음의 변수에 따라 하여 T-Minskew 히스토그램에 대한 평가를 하도록 한다.

- (1) 고정된 버킷 수와 변화하는 임계에 대한 오류율
- (2) 고정된 임계치와 고정된 버킷에 대한 오류율
- (3) 고정된 버킷 수와 변화하는 임계치에 대한 히스토그램 재구축 횟수
- (4) 고정된 임계치와 고정된 버킷에 대한 히스토그램 재구축 횟수

T-Minskew 히스토그램은 공간 히스토그램인 Minskew 히스토그램을 특정 timestamp에 따라 구축하여 이력 공간 질의에 대한 선택도 추정을 가능하게 한다. 기본적으로 Minskew 히스토그램을 사용하므로 오류율은 임계치에 따라 다소 증가할 수 있으나 실험을 통하여 전체 데이터 개수에 대한 적절한 임계치의 사용함으로써 오류율이 증가하는 것을 어느 정도 막고 정확성을 유지할 수 있다.

앞으로의 연구과제는 특정 timestamp에서의 공간 선택도 추정뿐만 아니라 범위 timestamp 선택도 추정이 가능하게 확장하고 Minskew 이외의 효과적인 timestamp 히스토그램을 구축하여 공간 선택도 추정 오류율을 낮추는 것이다.

참고문헌

- [1] Acharya, S., Poosala, V., Ramaswamy, S., "Selectivity Estimation in Spatial Databases," ACM SIGMOD, USA, pages 13-24, 1999.
- [2] Abounaga, A., Naughton, J. "Accurate Estimation of the Cost of Spatial Selections," ICDE, pages 123-134, 2000.
- [3] Viswanath Poosala, Yannis E. Ioannidis, Peter J. Haas, and Eugene J. Shekita, "Improved Histograms for Selectivity Estimation of Range Predicates", ACM SIGMOD, NY, USA, pages 294-305, 1996.
- [4] Wang, M., Vitter, J. S., Lim, L. and Pdmanabhan, S., "Wavelet-Based Cost Estimation for Spatial Queries", The 7th International Symposium on Spatial and Temporal Databases(SSTD), CA, USA, pages 175-196, July 2001.
- [5] Choi, Y., Chung, C., "Selectivity Estimation for Spatio-Temporal Queries to Moving Objects," ACM SIGMOD, pages 440-451, 2002.
- [6] Tao, Y., Sun, J., Papadias, D., "Selectivity Estimation for Predictive Spatio-Temporal Queries," ICDE, pages 417-428, 2003.
- [7] Hadjieleftheriou, M., Kollios, H. and Tsotras, V. J., "Performance Evaluation of Spatio-temporal Selectivity Estimation Techniques", The 15th Int. conference on Science and Statistical Database Management (SSDBM), pages 202-211, 2003.
- [8] Zhan, Q. and Lin, X., "Clustering Moving Objects for Spatio-temporal Selectivity Estimation", ADC, pages 123-130, 2004.