

공간 개념 계층에 따른 STOMP(FW) 알고리즘의 정확도 분석

이연식^{1*}, 김영자², 박성숙³

¹군산대학교 컴퓨터정보공학과, ²한국폴리텍II대학 컴퓨터정보과

³한국폴리텍V대학 유비쿼터스시스템학과

Precision Analysis of the STOMP(FW) Algorithm According to the Spatial Conceptual Hierarchy

Yonsik Lee^{1*}, Youngja Kim² and Sungsook Park²

¹Dept. of Computer Information Engineering, Kunsan National University

²Dept. of Computer Information, Korea Polytechnics II

³Dept. of Ubiquitous System, Korea Polytechnics V

요약 기존의 패턴 탐사 기법들은 대부분 객체가 갖는 공간 정보의 연속적인 변화에 대한 패턴 탐사는 가능하나, 추출하고자 하는 패턴에 반드시 포함되어야 하는 공간 정보에 대한 제약이 없어 특정 지점들 사이의 최적 경로 탐색 문제나 단위기간 동안 이동 객체가 순회해야 하는 지점들에 대한 경로 예측 문제 등에 적용하기 어렵다. 본 논문에서는 Road Network 상의 특정한 지점들 사이를 이동한 객체의 위치 데이터들 중 최다 빈발 패턴과 시간 및 비용 등의 가중치를 복합적으로 고려하여 최적의 이동 경로를 탐색하는 방법(Spatial-Temporal Optimal Moving Pattern(with Frequency & Weight) algorithm)[13]을 이용하여, 공간 개념 계층에 따른 경로 탐색의 정확도를 분석한다. 분석의 결과는 패턴 탐사 과정에 있어 공간 제약을 적용하여 검색 데이터 범위를 축소함으로써 데이터베이스 검색 시간을 최소화함을 보이고, 또한 공간 추상 계층의 각 계층별 영역 내 포함여부를 고려함으로써 효율적으로 최적 이동 패턴을 탐색하여 제공하도록 한다.

Abstract Most of the existing pattern mining techniques are capable of searching patterns according to the continuous change of the spatial information of an object but there is no constraint on the spatial information that must be included in the extracted pattern. Thus, the existing techniques are not applicable to the optimal path search between specific nodes or path prediction considering the nodes that a moving object is required to round during a unit time. In this paper, the precision of the path search according to the spatial hierarchy is analyzed using the Spatial-Temporal Optimal Moving Pattern(with Frequency & Weight) (STOPM(FW)) algorithm which searches for the optimal moving path by considering the most frequent pattern and other weighted factors such as time and cost. The result of analysis shows that the database retrieval time is minimized through the reduction of retrieval range applying with the spatial constraints. Also, the optimal moving pattern is efficiently obtained by considering whether the moving pattern is included in each hierarchical spatial scope of the spatial hierarchy or not.

Key Words : Spatio-Temporal Pattern Mining, Optimal Path Search, Spatial-Temporal Optimal Moving Pattern(with Frequency&Weight) Algorithm, Spatial Conceptual Hierarchy

1. 서론

효율적 위치 기반 서비스를 위하여 이동 객체의 위치

이동 데이터로부터 보다 의미있는 지식인 유용한 패턴을 탐색하기 위한 시간 패턴 탐사가 필요하며[1,2], 현재 시간 패턴의 유형별로 순차 패턴, 주기 패턴, 시간 관계에

본 논문은 2009년도 정부의 재원으로 한국학술진흥재단(No. 2009-0074891)의 지원과 2010년도 교육과학기술부 및 한국산업기술진흥원의 지역혁신인력양성사업으로 수행된 연구결과임.

*교신저자 : 이연식(yslee@kunsan.ac.kr)

접수일 10년 09월 24일

수정일 (1차 10년 11월 08일, 2차 10년 12월 13일)

게재확정일 10년 12월 17일

관한 많은 기법들이 제시되었다[2-6]. 하지만, 기존의 시간 패턴 탐사 기법들은 시간과 공간 속성을 동시에 고려해야 하는 이동 객체 데이터의 패턴 탐사에 있어 공간 속성들의 변화를 고려하지 않아 이동 패턴 탐사 문제에 적합하지 않다. 이에 대한 해결 방법으로 이동 객체에 대한 시간 패턴 탐사를 위해서 Apriori를 확장한 이동 패턴 마이닝 알고리즘들[4,7]이 제안되었다. 이러한 방법들은 이동 객체의 공간 속성을 고려하여 공간 정보가 연속적으로 변화하는 이동 패턴 탐사는 가능하나, 제약을 가진 공간 정보의 연속적인 변화에 대한 패턴 탐사에는 적용하기는 어렵다. 즉, 추출하고자 하는 패턴에 공간 정보에 대한 제약이 없어 특정 지점들 사이의 최적 이동 경로 탐색 문제나 단위 기간 동안 이동 객체가 순회해야 지점들에 대한 스케줄링 경로 예측 문제 등에 적합하지 않다[8,9].

이에 본 논문에서는 이동 객체의 위치 이동 데이터들에 대한 시간적/공간적인 속성들을 고려하여 다양한 이동 패턴들 중 최적 이동 경로나 스케줄링 경로 예측과 같은 위치 기반 서비스에 적용 가능한 새로운 이동 패턴 탐사 기법(STOMP(FW) 알고리즘(Spatial-Temporal Optimal Moving Pattern(Frequency&Weight) Algorithm))[13]을 이용하여, 공간 개념 계층에 따른 경로 탐색의 정확도를 분석한다. 분석 결과는 패턴 탐사 과정에 있어 공간 제약을 통한 검색 데이터 범위의 축소를 통해 데이터베이스 검색 시간을 최소화하고, 공간 추상 계층의 각 계층별 영역 내 포함여부를 고려하여 각 공간 개념 계층에서의 적용 알고리즘들의 정확도를 분석하여 최적의 이동 패턴을 탐색 및 제공할 수 있도록 한다.

본 논문의 구성은 다음과 같다. 2장에서는 이동 객체의 시공간 패턴 탐사를 위한 정의와 이동 객체의 위치 일반화와 영역 대표유효시간 추출 방법 등 이론적 배경을 제시하고, 3장에서는 최적 이동 패턴 탐사와 적용 알고리즘을 설명한다. 4장에서는 빈발 패턴과 시간 및 비용 등의 가중치를 복합적으로 고려하여 가장 효율이 좋은 이동 경로를 탐색하는 방법(STOMP(FW) 알고리즘)을 이용하여, 공간 개념 계층 레벨에 따라 거리 중심, 패턴 빈발도 중심, 패턴 빈발도와 가중치 중심으로 이동 객체가 이동한 경로 탐색의 정확도를 실험을 통하여 비교 분석하고 그에 따른 결론을 제시한다.

2. 이동 객체의 시공간 패턴 탐사를 위한 정의 및 요구사항

최적 경로 탐색을 위한 이동 객체의 시간 패턴 탐사를 방대한 이동 객체의 위치 이동 데이터들로부터 특정한

지점들 간을 이동하는 객체의 다양한 패턴들 중 가장 최적의 비용을 소요하는 이동 패턴을 탐색하기 위함이다 [2,4,7]. 본 논문에서는 공간 개념 계층에 따른 경로 탐색의 정확도를 분석하기 위하여, 전체 이동 객체 데이터베이스로부터 빈발 패턴 집합을 추출하는데 있어 이동 객체의 연속적 위치변화를 보다 효과적으로 패턴화하기 위해서 공간 영역 일반화 접근법을 사용하는 MP 알고리즘 [4,7]을 변형하여 사용하고, 최적 경로를 탐색하기 위한 방법으로 시간/공간 제약 빈발 패턴 및 가중치를 이용하는 방법을 사용한다. 이 장에서는 이동 패턴 마이닝에 있어서의 이동 객체의 표현 방법에 대해 정의하고, 이동 객체의 시공간 패턴 탐사 문제 정의 및 이동 객체의 위치 일반화와 영역 대표유효시간 추출방법을 제시한다.

2.1 이동 객체 및 최적 이동 패턴 탐사 정의

이동 객체의 위치 변화는 이산적 또는 연속적으로 발생할 수 있으며, 시점(time points) 또는 시간 구간(time periods)으로 기록될 수 있다.

본 논문에서는 연속적으로 이동하는 객체의 이동 위치가 이산적인 시점에서 추출됨을 가정하며, 이동 점 객체의 공간 속성은 이동 객체의 평면상의 x, y 좌표로써 표현된다.

정의 1 - 이동 점 객체

$$Mpoint = oid, \{(VT_1, L_1), (VT_2, L_2), \dots, (VT_n, L_n)\}$$

oid : 유일한 특성을 갖는 객체의 식별자

VT_i : 유효시간

L_i : VT_i 에 샘플링된 객체의 위치 (x_i, y_i)

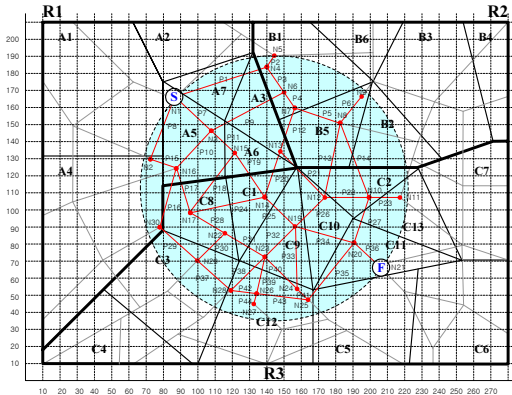
빈발도와 가중치를 이용한 최적 이동 패턴 탐사 방법은 객체의 다양한 이동 패턴들에 대해 빈발도(지지도)를 추출한 후 지지도 임계치를 적용하여 최소지지도 이상이고 가중치(거리, 시간, 비용 등)가 가장 적게 소요되는 이동 패턴을 탐사하는 방법이다. 본 논문에서 적용하는 방법의 기본 개념은 특정 지점 간을 이동한 객체의 패턴들 중 빈번하게 발생한 패턴이 대체로 최적의 비용을 소요하는 패턴일 확률이 높다는 가정을 기반으로 하지만 가장 빈번한 패턴이 반드시 최적의 비용을 소요한다고는 단정할 수 없다. 또한, 지지도에 대한 고려뿐만 아니라 사용자가 의도하는 최적 패턴에 대한 평가 요소(거리, 시간, 비용 등)가 최소일 때 최적 경로로 선택될 확률이 높다. 따라서 이러한 가정을 만족하는 패턴이 가장 최적 경로일 확률이 높다[13]. 이러한 조건에 따라 빈발도와 가중

치를 기반으로 한 최적 이동 패턴 탐사 문제를 정의하면 다음과 같다.

정의 2 - 최적 이동 패턴 탐사

이동 객체 데이터베이스 MD , 사용자가 지정한 특정지점들 S 와 F (공간제약), 시간기간 T_1 과 T_2 (시간제약), 최소지지도 min_sup 이 주어졌을 때, 공간 및 시간제약을 만족하는 이동 시퀀스 집합으로부터 최소지지도를 만족하는 S 에서 F 까지의 모든 빈발한 부분 시퀀스들을 추출하여 가중치가 가장 적은 시퀀스들의 집합을 탐색하는 것이다. 단, 부분 시퀀스는 S 에서 F 까지를 도달하는 하나의 시퀀스에 포함되어야 한다.

최적 이동 패턴의 탐사 과정을 보이기 위해, 표 1은 그림 1의 도로 네트워크 상에서 각 경로 구간을 이동할 때 소요되는 비용, 즉 가중치 값을 나타낸 것이고, 표 2는 이동 객체들의 궤적을 각 노드에서 샘플링 한 이력 데이터이다. 가중치는 이동 객체의 각 이동 정보들 간의 효율성을 상대적으로 비교하여 백분율로 표현한 것으로, 가중치의 기준요소는 이동시간 또는 이동거리, 이동 소요 비용 등이 될 수 있다.



[그림 1] 공간지역과 도로 네트워크 예

[표 1] 이동 경로별 가중치 예

Path	weight
P1	0.76
P2	0.12
P3	0.24
P4	0.18
.	.
P42	0.55
P43	0.79
P44	0.47

[표 2] 이동 객체의 이력 데이터 예

OID	X	Y	VT
1	121	133	2009/11/20/12/29/34
	95	99	2009/11/20/12/43/36
	139	108	2009/11/20/12/56/16
2	191	81	2009/11/20/13/52/29
	226	66	2009/11/20/13/01/19
.	.	.	.
.	.	.	.
n	121	133	2009/11/20/12/15/45
	95	99	2009/11/20/12/22/54
	139	108	2009/11/20/12/34/07
	157	90	2009/11/20/12/45/46

2.2 이동 객체의 위치 일반화와 영역 대표유 효시간 추출

실세계에서 이동 객체의 의미있는 패턴을 획득하기 위해서 이동 객체의 상세한 위치 정보를 공간상의 상위 개념으로 추상화할 필요가 있다. 이동 객체의 위치 속성은 공간상의 특정 지역 수준의 의미로 일반화되기 위하여 공간 지역에 대한 공간 개념 계층이 요구된다.

정의 3 - 공간 개념 계층 레벨의 집합 정의

공간 개념 계층 레벨 집합 L_s 의 정의는 다음과 같다.

여기서, $L_s \subseteq L_{s_{k-1}}$, $0 \leq k \leq j$ 이다.

$$L_s = \{L_{s_j} | L_{s_j} \in L, L_{s_k} \subseteq L_{s_{k-1}}, \text{ for } \forall j\}$$

[정의 3]에서 공간 개념 계층의 집합 L_s 는 개념적인 전체 레벨 L 의 원소가 될 수 있다. 또한 L_s 의 원소인 L_{s_k} 와 $L_{s_{k-1}}$ 의 관계에 있어 L_{s_k} 가 $L_{s_{k-1}}$ 보다 레벨 수준을 의미하는 k 가 한 수준 낮기 때문에 $L_{s_k} \subseteq L_{s_{k-1}}$ 이다.

정의 4 - 공간 개념 계층 레벨에 대한 지역 인스턴스 객체 정의

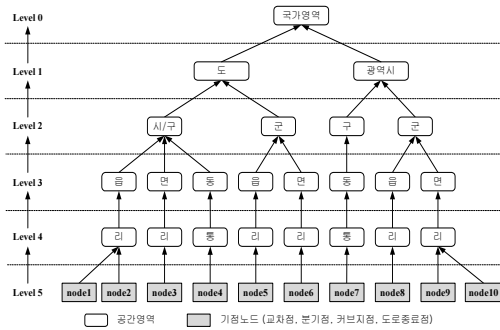
공간 개념 계층 레벨 L_s 에 대한 R 의 인스턴스 객체 $L_{s_{R_i}}$ 는 다음과 같다.

$$L_{s_{R_i}} = \{L_{s_{k_R}} | L_{s_{k_R}} \in R, L_{s_{k_R}} \subseteq L_{s_{k-1_R}}, 1 \leq k \leq j \text{ and for } \forall i\}$$

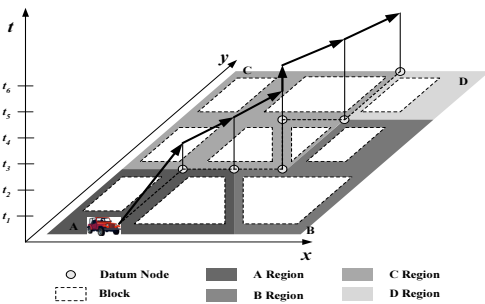
정의 4에서 공간 레벨 수준이 k 인 공간지역 인스턴스 객체 R_i 는 전체 공간지역의 집합 R 의 원소이고, 정의 3

에 따라 L_{S_k} 레벨에 포함되는 R_i 는 $L_{S_{k-1}}$ 레벨에 포함된 R_i 의 부분집합이 될 수 있다.

다음 그림 2-(a)는 행정구획에 대한 6 레벨의 공간 개념 계층의 예로, 단말 수준의 노드는 도로 상의 교차점, 분기점, 커브지점, 도로종료점 등을 표현하고 각 레벨별 공간 범위는 포함관계를 통해 상위 레벨 영역으로 일반화된다. 그림 2-(b)는 시간 t축 상에서 불규칙적인 단위시간동안 단말노드인 기점노드 간을 이동하는 객체의 이력에 대한 표현 예이다.



(a) 행정구획에 대한 개념 계층



(b) 공간영역과 기점노드

[그림 2] 공간 개념 계층

이동 객체의 위치 정보에 대한 공간 일반화는 x, y 좌표 값으로 표현된 이동 객체의 공간 정보를 좌표가 포함되는 영역으로 나타나는 방법이다. 특정 시간 기간 VT_n 에서 VT_n 동안의 이동 객체의 위치 정보는 이동 객체의 공간 위치 속성을 나타내는 좌표의 유한집합 $L=\{l_1, l_2, \dots, l_m\}$ 로 표현될 수 있다. $l_i=(x_i, y_i)$ 이며 각 x_i, y_i 는 평면상에서의 이동 객체의 좌표이다. $A=\{a_1, a_2, \dots, a_n\}$ 은 이동 객체의 공간 위치 속성 값으로 표현되는 영역의 집합으로, 각 $1 \leq j \leq n$ 에 대해 $a_j=(l_1, l_2, \dots, l_k)$ 이고 $l_k=(x_k, y_k)$ 이다. 이동 객체의 위치 일반화는 영역이 가지는 대표 좌표 값들을 사용하여 좌표 값으로 표현된 이동 객체의 공간 위

치 속성을 공간연산을 통해 영역으로 변환한다[10].

이동 객체의 위치 속성은 일반화된 공간 영역으로 표현되며, 이동 객체의 이동 이력은 영역의 순차 리스트로 표현될 수 있다. 영역의 순차 리스트는 한 이동 객체에 대해 시간의 증가에 따라서 영역을 순서적으로 정렬한 이동 시퀀스(moving sequence)로 표현한다[11]. k 개의 길이를 가진 시퀀스, 즉 k 개의 영역으로 이루어진 시퀀스는 k -시퀀스로 표기되며, 하나의 영역은 한 시퀀스에서 여러 번 나타날 수 있다. a_k, b_k 등과 같이 이동 영역을 나타내는 식별자들로 구성된 두 이동 시퀀스에 대하여 시퀀스 $\langle a_1 a_2 \dots a_n \rangle$ 은 만약 $a_1=b_{i_1}, a_2=b_{i_2}, \dots, a_n=b_{i_n}$ 인 정수 $i_1 < i_2 < \dots < i_n$ 이 존재한다면 다른 시퀀스 $\langle b_1 b_2 \dots b_m \rangle$ 의 부분 시퀀스이다. 하나의 시퀀스 $MS=\{s_1, s_2, \dots, s_k\}$ 로 표현할 수 있다. 이 때, $s_j=(a_j, t_j)$ 이며, t_j 는 이동 객체가 a_j 영역 내에 있을 때의 대표유효시간이고 $a_i \in A$ 이다. 특정 영역(a_i) 내 대표유효시간 AT_i 는 특정 영역(A_i)에 들어갔을 때 처음 샘플링된 유효시간(A_iTim)에서부터 나오기 전 샘플링된 유효시간(A_iTout)까지의 평균유효시간이다. 한편, 이동 객체의 이동 이력에 대한 영역 순차 리스트인 시퀀스를 형성하기 위해서는 트랜잭션 시간 간의 시간 간격인 시간제약조건을 적용한다. 시간제약조건은 시퀀스 내에 연속적인 이동으로 영역이 포함되기 위해 인접한 이동이 발생한 시간 $t_j - t_{j-1}$ 을 의미한다. 최대 시간간격은 max_gap 으로 $t_j - t_{j-1} \leq max_gap$ 이고, $2 \leq j \leq k$ 이다. 이동 시퀀스의 집합 $S=\{ms_1, ms_2, \dots, ms_m\}$ 이라 할 때, 각 ms_i 는 이동 시퀀스를 나타내며, 이 때 $1 \leq i \leq m$ 이다. 한 시퀀스 ms 가 다른 시퀀스 ms' 의 부분 시퀀스이면 ms' 는 ms 를 포함한다고 한다. 시퀀스 ms 의 지지도는 ms 를 포함하는 전체 이동 시퀀스의 비율 즉, $sup(ms)=\{ms_i | ms \subseteq ms_i\} / m$ 으로 정의된다. 최소지지도(minimum support threshold) min_sup 는 시퀀스 ms 가 빈발한 것으로 평가되기 위해 만족해야 하는 지지도의 하한값으로 사용자가 명시한다. 하나의 시퀀스 ms 는 $sup(ms) \geq min_sup$ 이면 빈발 시퀀스이다. 시퀀스에서 하나의 지역이 여러 번 나타날 수 있지만, 하나의 시퀀스 안에서는 최대 한번만 세어진다.

3. 최적 이동 패턴 탐사

3.1 최적 이동 패턴 탐사

최적 이동 패턴의 탐사(정의 2)는 이동 객체의 이력 데이터를 변환하는 전처리 단계와 최적 이동 패턴을 탐사하는 단계로 구분된다. 전처리 단계는 이동 객체의 이력 데이터를 각 객체별로 시간 순차에 따라 나열한 한 쌍

의 공간 및 시간 속성 값으로 구성된 순차 리스트인 이동 시퀀스를 생성하는 단계로서, 이동 객체 데이터베이스를 각 객체별로 유효시간을 기준으로 정렬한 후 시퀀스를 구성하는 위치들 간에 최대 시간 간격 max_gap 에 대한 제약 조건을 두어 이를 만족할 때 이동 시퀀스를 생성한다. 최적 이동 패턴 탐사 단계는 전처리 과정을 통해 생성된 이동 시퀀스 집합으로부터 최소지지도 min_sup 이상의 지지도를 가진 빈발 2-시퀀스들 중 가장 최소의 가중치를 가지는 이동 시퀀스들을 탐사하는 단계이다.

패턴 탐사의 기본 원리는 먼저, 트랜잭션 데이터베이스로부터 시간 및 공간 제약을 만족하는 이동 시퀀스 집합을 생성한다. 그 다음 공간상의 특정한 두 지점(시작점, 종료점)을 항목으로 가지는 후보 시퀀스 집합을 생성하고, 각 후보 시퀀스들로부터 첫 번째와 두 번째 항목만을 추출하여 부분 2-시퀀스를 생성한다. 생성된 부분 2-시퀀스들에 대해 이동 시퀀스 집합의 각 시퀀스들을 탐색하여 지지도를 계산한 후 지지도가 min_sup 이상인 빈발 2-시퀀스들을 추출하고 이 중 가장 적은 가중치를 갖는 최소 가중 2-시퀀스를 생성한다. 이러한 과정은 최소 가중 2-시퀀스의 항목으로 종료점 F가 포함된 2-시퀀스를 추출할 때까지 반복적으로 수행되며, 모든 최소 가중 2-시퀀스들을 탐색했을 경우 각 시퀀스를 구성하는 항목들의 순서에 따라 조합하여 최적 이동 패턴을 추출한다. 이와 같은 탐사 과정은 마지막 최대 빈발 2-시퀀스를 결정하는 단계에서 빈발도와 가중치 기반은 최소가중치를 기준으로 결정한다.

3.2 최적 이동 패턴 탐사 알고리즘

빈발도와 가중치 기반의 최적 이동 패턴 탐사 알고리즘 $STOMP(FW)$ 는 최적 패턴을 탐사하는데 있어 이동 패턴의 빈발도에 대한 임계치인 최소지지도 min_sup 과 각 단위 경로에 대한 가중치 WT 를 사용한다[13]. 그림 3의 $STOMP(FW)$ 는 $STOMP(F)$ [12]와 마찬가지로 $SeqExtractor$ 함수를 통해 이동 시퀀스를 생성한 후 $OptPathExtractor-WT$ 함수를 통해 최대 빈발 시퀀스를 추출하여 최적 경로를 결정한다. $limitedDataSet$ 함수는 이동 객체 데이터베이스로부터 공간 및 시간 한정자 C_s 와 C_t , 공간제약조건 S와 F, 두 제약 조건 모두를 만족하는 객체의 이력 데이터 집합을 추출하는 함수이다.

```

Input : D(Database), Cs(Spatial Constraint), Ct(Temporal Constraint),
        S(Start Node), F(Final Node), max_gap(Time Interval Constraint),
        Ls(Spatial Level), min_sup(Threshold of Minimum Support),
        WT(Threshold of Weight)
Output : Opt.Path

Procedure STOMPFW(D, Cs, Ct, S, F, max_gap, Ls, min_sup, WT)
Begin
    PrevA = null;

    //Set of History Data satisfied Constraint of Spatial and Temporal
    D1 = limitedDataSet(D, Cs, Ct, S, F);

    // Extracting Set of Moving Sequence from D1.
    D2K = SeqExtractor(D1, max_gap);

    // Generalizing Region to Spatial Property of Moving Sequence in D2K
    D3(K)} = Contains(D2K, PrevA, Ls);

    // Moving Pattern Mining
    Opt.Path = Opt.PathExtractor - WT(D3(K)}, min_sup, WT, S, F);

    Return Opt.Path;
End
    
```

[그림 3] STOMP(FW) 알고리즘

$OptPathExtractor-WT$ 함수는 최적 경로 추출 알고리즘으로서, 반복적으로 $Freq_Link-WT$ 함수를 호출하여 출발점 S로부터 시작해 도착점 F에 도달할 때까지의 단위 최적 패턴을 탐사한다. 반복적인 단위 최적 패턴 탐사가 종료되면, $OptPath$ 의 최종 결과 값이 최적 이동 패턴으로 결정된다. 또한, 단위 최적 패턴 추출 알고리즘인 $Freq_Link-WT$ 함수는 $OptPathExtractor-WT$ 함수에서 전달받은 전체 이동 시퀀스 집합으로부터 출발점 S와 도착점 F를 포함하는 부분 이동 시퀀스 집합을 추출한다. 그 다음 최초 S로부터 시작하는 최소지지도 min_sup 이상의 1-빈발 패턴들을 탐색하고 그 중 가중치가 최소인 패턴을 추출한다. 추출된 패턴에서 S가 아닌 반대편 노드 S'(nextN)을 기준으로 하여 다시 F를 포함하는 모든 부분 이동 시퀀스 집합을 추출한 후 최소지지도 및 가중치 조건을 만족하는 새로운 단위 최적 패턴을 추출한다. 이러한 과정은 F에 도달하는 단위 최적 패턴을 탐사할 때까지 반복 수행한다.

4. 실험 및 평가

본 장에서는 Road Network 상의 특정한 지점들 사이를 이동한 객체의 위치 데이터들 중 객체가 가장 빈번하게 이동

한 경로를 탐색하여 최적 경로로 결정하는 빈발 패턴을 이용한 경로 탐색 방법과 시간 및 비용 등의 가중치를 복합적으로 고려하여 가장 효율성이 좋은 이동 경로를 탐색하는 방법(STOMP(FW) 알고리즘(Spatial-Temporal Optimal Moving Pattern (Frequency&Weight) Algorithm))[13]을 이용하여, 본 논문에서 정의한 공간 개념 계층 레벨에 따라 거리 중심, 패턴 빈발도 중심, 패턴 빈발도와 가중치 중심으로 이동 객체가 이동한 경로 탐색의 정확도를 실험을 통하여 비교 분석한다.

구현은 Windows XP 기반하에서 Eclipse 3.2, JDK 1.6.0과 JBuilder X 프로그램을 이용하였으며, 데이터베이스는 Oracle 10g를 사용하였다. 성능 평가를 위해 사용된 시스템은 Pentium PC 2.4GHz 2Gbyte 사양의 PC 2대를 사용하였다. 실험을 위한 Geometry 데이터로 서울시 행정 구획 데이터와 도로 네트워크 데이터를 사용하였고, 이동 객체 데이터로 서울시의 도로 네트워크 상에서 택시들의 운행 기록을 추위하여 이력 데이터를 생성하였다. 이력 데이터는 각 탐색 횟수(100~500번)에 대비하여 출발점(서울시 중구 내 임의노드)을 중심으로 5개의 동심원 형태로 반경을 점차 늘려가며 해당 반경에 포함되는 도착점까지의 거리에 대한 결과를 측정하였으며, 도로 네트워크 상에서의 교차점이나 분기점, 종료점 등을 노드로 설정하여 각 노드에서 이동 객체의 위치 정보를 샘플링 하였다. 다음 표3과 표 4는 성능 평가 실험에 이용되는 각각 이력 데이터의 유형과 실험 데이터에 대한 특성을 나타내며, 데이터 집합의 이름에서 N은 이동 객체 수, D는 이동기간, G는 샘플링 위치를 의미한다.

[표 3] 성능 평가를 위한 실험 데이터 유형

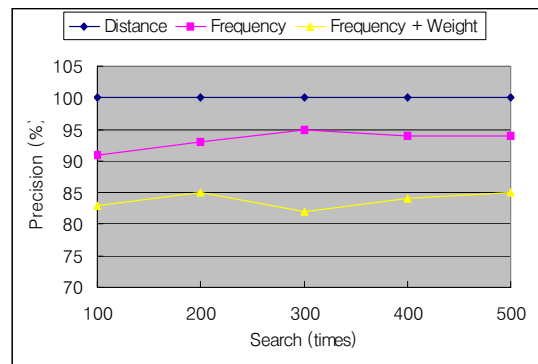
MOID	NODE		Valid Time
	x	y	
15440001	121.731123	133.203405	2009/11/20/12/29/34
15440001	121.731208	133.206347	2009/11/20/12/30/10
15440001	121.745882	133.212650	2009/11/20/12/30/58
15440001	121.740665	133.217732	2009/11/20/12/32/11
...

[표 4] 실험 데이터 특성

데이터 집합 이름	이동 객체 수(N)	이동기간 (D)	실험데이터 크기
N2000-D1120-G _{node}	2000	1일	31.5 MB

실험 방법은 데이터베이스 서버로부터 정해진 시간동안 특정 지점에서 목표 지점까지 운행한 이력 데이터를 검색하여 먼저, 거리 중심으로 이동한 경로를 탐색하고, 기존 연구에서 제안한 탐사 방법인 패턴 빈발도를 이용한 방법과 패턴 빈발도와 가중치를 이용한 방법으로 각 경로를 탐색하여 거리 중심의 경로와의 비교를 통해 어느 정도 경로들이 일치하는지에 대해서 평가하였다.

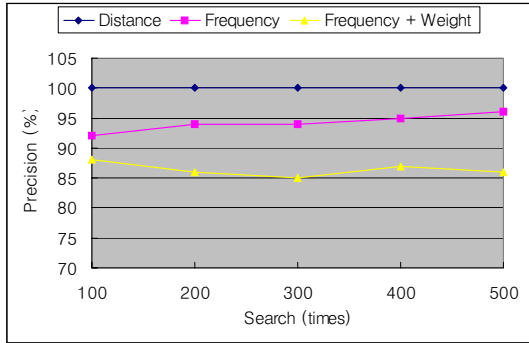
실험에서는 각 탐색 방법에 동일 질의를 수행한 결과, 탐색된 경로가 어느 정도 유사하게 이동했는지를 평가한다. 경로 탐색의 횟수는 각 탐색 방법 당 100 ~ 500번 수행하였다. 그림 4는 본 논문에서 정의한 공간 개념 계층의 최하위 레벨에 해당하는 Level 5 영역들 상에서 거리 중심, 패턴 빈발도 중심, 패턴 빈발도와 가중치 중심으로 이동 객체가 이동한 경로들의 탐색 결과의 비교를 보인다. 먼저 거리 중심의 경로를 기준으로 패턴 빈발도 중심의 경로는 탐색 횟수 범위 내에서 최대 9%에서 5%의 차이를 보였다. 또한 패턴 빈발도와 가중치를 중심으로 비교하였을 경우는 최대 18%에서 최소 15%의 차이를 보였다. 이는 패턴 빈발도의 경우, 운전자가 출발 지점에서 목표 지점까지의 최단 경로를 어느 정도 비슷하게 알고 있기 때문에 최단 경로와 근접하게 이동하였을 것으로 생각된다. 패턴 빈발도와 가중치의 경우, 최단 경로와 비슷한 경로뿐만 아니라 시간이나 비용과 같은 이동 경로 결정 요인을 어느 정도 인지하고 있기 때문에 패턴 빈발도를 이용하는 경우보다 좀 더 차이를 보인다.



[그림 4] 공간 개념 계층 Level 5에서의 경로 탐색 정확도 비교

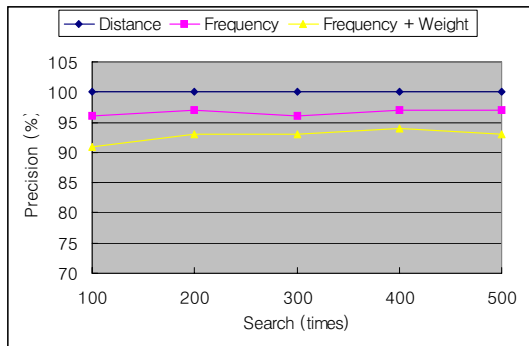
그림 5는 공간 개념 계층을 한 단계 높여 Level 4 영역들 상에서 동일 방법을 적용한 비교 결과로써, 거리 중심의 경로 탐색 방법과 비교하여 빈발도 중심은 8% ~ 4%의 차이를 보이며, 빈발도와 가중치 중심은 15% ~ 12%의 차이를 보인다. 이는 각 탐색 방법을 중심으로 탐색한

이동 경로의 유사율이 그림 4와 비교하여 더 상향되었음을 알 수 있다. 이는 작은 영역 간에서의 이동 경로보다 한 단계 위 레벨에서의 영역 간의 이동에 있어 경유하는 영역이 크게 변화하지 않음을 보여준다.



[그림 5] 공간 개념 계층 Level 4에서의 경로 탐색 정확도 비교

빈발도와 가중치를 사용하여 최적 경로를 탐색할 경우 공간 개념 계층 개념의 레벨이 상승함에 따라 최적 경로 탐색의 정확도가 높아질 수 있음을 보장하기 위하여, 보다 상위 단계인 Level 3 영역 상에서 경로 탐색을 수행한 결과는 다음 그림 6과 같다.



[그림 6] 공간 개념 계층 Level 3에서의 경로 탐색 정확도 비교

위의 실험들의 결과에서 알 수 있듯이 공간 개념 계층에 따라 빈발도 및 빈발도와 가중치 등을 이용하여 최적 경로들을 탐색할 때 탐색의 정확도에 차이가 있음을 알 수 있고, 공간 개념 계층을 구성하는 각 공간 영역들의 크기에 따라 최적 경로 탐색의 정확도에 차이가 있음을 알 수 있다. 이는 선행연구[12, 13]를 기반으로 개발 알고리즘들이 최적 경로 탐색을 위한 과정에서 필수적으로 요구되는 이동 객체들의 위치 일반화를 서로 다른 공간

개념 계층에 따라 실시할 때 탐색을 거리, 빈발도 및 빈발도와 가중치를 각각 사용하여 수행하였을 경우 각 알고리즘들의 정확도를 비교 평가한 것으로써, 이동 객체의 이동 환경이나 상황에 따라 위치 일반화 단위(크기)를 공간 개념 계층 레벨에 따라 가변적으로 설정함으로써 얻을 수 있는 알고리즘 수행시간 감소 및 다량의 복잡한 이력 데이터들의 수집, 저장, 및 관리를 효율적으로 할 수 있음을 보여준다.

STOMP(FW) 방법은 기존의 최적 경로 탐색 방법(STOMP(F), A*, Dijkstra)과 비교하여 패턴 탐사 시간을 최소화하고 최적 패턴의 정확도는 극대화하기 위한 방식이며[13], 이러한 방법들은 최적 물품 경로 서비스나 최적 경로 추천 서비스 등 다양한 응용에 제공될 수 있는데, 이들은 짧은 시간 내에 최적 경로를 제공하고 이동 시 소요되는 비용을 최소화할 필요가 있으므로 공간 개념 계층 레벨에 따른 알고리즘 수행 시간과 정확도가 주요 변수가 됨을 알 수 있다. 또한, 사용한 이동 객체의 이력 데이터는 실세계의 경험 이력 데이터를 기반으로 하기 때문에 특정 경로 상에서의 교통 체증이나 도로공사와 같은 특수한 상황에 대한 이력도 포함할 수 있으므로 빈발도와 가중치를 이용한 STOMP(FW) 방식을 공간 개념 계층 레벨별로 탐색의 정확도를 분석 적용하는 것은 보다 능동적 경로탐색에 효과적임을 보인다.

참고문헌

- [1] O. Wolfson, B. Xu, J. Zhou, S. Chamberlain, Y. Yesha and N. Rishe, "Tracking Moving Objects Using Database Technology in DOMINO", in proc. on The Fourth Workshop on Next Generation Information Technologies and Systems(NGITS), pp.112-119, July 1999.
- [2] D. O. Kim, H. K. Kang, D. S. Hong, J. K. Yun and K. J. Han, "STMPE : An Efficient Movement Pattern Extraction Algorithm for Spatio-temporal Data Mining", in proc. on International Conference on Computational Science and Its Applications(ICCSA), pp.259-269, 2006.
- [3] N. Mamoulis, H. Cao, G. Kollis, M. Hadjieleftheriou, Y. Tao and D. W. Cheung, "Mining, Indexing and Querying Historical Spatio-Temporal Data", in proc. on the International Conference on Knowledge Discovery and Data Mining, 2004.
- [4] H. Cao, N. Mamoulis and D. W. Cheung, "Mining Frequent Spatio-Temporal Sequential Patterns", in

proc. on the 5th IEEE International Conference on Data Mining(ICDM), pp.82-89, 2005.

- [5] Y. Huang, L. Zhang and P. Zhang, "Finding Sequential Patterns from a Massive Number of Spatio-Temporal Events", SDM, SIAM, 2006.
- [6] J. W. Lee, O. H. Paek and K. H. Ryu, "Temporal Moving Pattern Mining for Location-Based Service", The Journal of Systems and Software, Vol.73. 2004.
- [7] J. Gudmundsson, M. V. Kreveld and B. Speckmann, "Efficient Detection of Motion Patterns in Spatio-Temporal Data Sets", in proc. on the 12th annual ACM international workshop on Geographic Information Systems(GIS), pp.250- 257, 2004.
- [8] G. Yavas, D. Katsaros, O. Ulusoy and Y. Manolopoulos, "A Data Mining Approach for Location Prediction in Mobile Environments", Data & Knowledge Engineering, Vol.54, pp.121-146, 2005.
- [9] 이준욱, "위치 기반 서비스를 위한 이동 객체의 시간 패턴 탐사", 한국정보과학회 논문지, 제29권, 제5호, 2002.
- [10] 고현, 김광중, 이연식, "R* Tree와 Grid를 이용한 이동 객체의 위치 일반화 기법", 한국컴퓨터정보학회논문지, 제12권 2호, pp.231-242, 2007. 5
- [11] 한선영, "시공간 이동 시퀀스 패턴 마이닝 기법", 이화여자대학교 대학원, 석사학위논문, 2006.
- [12] 고 현, 이연식, "이동 시퀀스의 빈발도를 이용한 최적 이동 패턴 탐사 기법", 한국정보처리학회 논문지, 제16-D권, 제1호, 2009.
- [13] 이연식, 박성숙, "시퀀스 빈발도와 가중치를 이용한 최적 이동 패턴 탐사" 한국인터넷정보학회논문지, 제 10권 제5호, pp.79-93, 2009.10

이 연 식(Yonsik Lee)

[정회원]



- 1982년 2월 : 전남대학교 전자계산학과 졸업(학사)
- 1984년 2월 : 전남대학교 대학원 전자계산학과 졸업(이학석사)
- 1994년 2월 : 전북대학교 대학원 전산응용공학전공 졸업(공학박사)
- 1997년 ~ 1998년 : University of Missouri(Kansas City) 교환교수
- 2004년 ~ 2005년 : Ohio State University 교환교수
- 1986년 3월 ~ 현재 : 군산대학교 컴퓨터정보공학과 교수

<관심분야>

변역기 이론, 객체지향시스템, 능동시스템, 지능형 에이전트 미들웨어

김 영 자(Youngja Kim)

[정회원]



- 2000년 2월 : 순천대학교 교육대학원 컴퓨터학과(교육학 석사)
- 2008년 8월 : 군산대학교 대학원 컴퓨터학과(이학박사)
- 2001년 9월~ 현재 : 한국폴리텍Ⅱ 대학 컴퓨터정보학과 교수

<관심분야>

임베디스 시스템, 능동시스템, 지능형 에이전트

박 성 숙(Sungsook Park)

[정회원]



- 1992년 2월 : 원광대학교 컴퓨터공학과(학사)
- 1995년 2월 : 원광대학교 대학원 컴퓨터공학과(공학석사)
- 2009년 9월 : 군산대학교 대학원 컴퓨터정보공학과 박사과정수료
- 1996년 3월 ~ 현재 : 한국폴리텍V대학 유비쿼터스시스템학과 교수

<관심분야>

지능형 에이전트, 센서 네트워크 미들웨어