

최대 빈발 2-시퀀스를 이용한 최적 이동 패턴 추출

이연식, 고현, 김광종
군산대학교

yslee@kunsan.ac.kr, khyun001@kunsan.ac.kr, kkjkim@kunsan.ac.kr

Extraction of Optimal Moving Pattern using Maximum Frequent 2-Sequence

Yon-sik Lee, Hyun Ko, Kwangjong Kim
Kunsan University

요 약

최근 사용자들의 특성에 맞게 개인화되고 세분화된 위치 기반 서비스를 개발하기 위한 목적으로 이동 객체의 다양한 패턴들 중 의미있는 지식인 유용한 이동 패턴을 탐사하는 문제가 주요 이슈로 부각되고 있다. 이에 본 논문에서는 방대한 이동 객체의 이력 데이터 집합으로부터 특정 지점들 간의 최적 이동 경로나 정해진 시간내의 스케줄링 경로 탐색과 같이 복합적인 시간 및 공간 제약을 갖는 최적 이동 패턴을 탐사하는 문제에 대해 정의하고, 다양한 이동 패턴들 중 가장 빈발하게 발생하는 패턴이 최적의 비용을 소요할 것이라는 가정을 기반으로 최대 빈발 2-시퀀스를 추출하는 방법을 제안한다. 후보 시퀀스 집합으로부터 지지도 계산을 통해 추출되는 빈발 2-시퀀스들의 순차적인 조합은 패턴 탐사를 수행하는 각 패스 진행 시 후보 시퀀스 항목의 차수가 점차 감소하여 최적 이동 패턴 탐사 방법에 효과적으로 적용된다.

1. 서 론

최근 다양한 응용 분야에서 활용할 수 있는 새로운 위치 기반 서비스를 개발하고자 이동 객체의 동적인 위치 변화에 따른 특성 분석을 통해 유용한 정보를 추출하기 위한 이동 패턴 탐사에 대한 연구들이 지속적으로 수행되고 있다. 순차 패턴(sequential pattern)이나 주기 패턴(periodic pattern)과 같은 빈발(frequent) 이동 패턴을 탐사하는 기존의 시공간 이동 패턴 탐사 기법들은 패턴 탐사 시 후보 이동 시퀀스(candidate moving sequence)의 생성 및 검증, 후보 시퀀스들의 지지도 계산을 위해 빈번하게 데이터베이스를 스캔(scan)하는 문제와 많은 양의 데이터를 탐사 대상으로 하거나 낮은 최소지지도(minimum support)를 가지는 패턴들을 탐사하는 문제로 인해 탐사 수행 시간(mining execution time)이 증가하는 단점이 있다[1, 2, 3, 4, 7]. 또한, 후보 시퀀스 집합이 크거나 데이터의 양이 많을 경우 탐사 시 소요되는 메모리의 양이 급격히 증가하는 문제가 있다 이에 따라 일부 기법들은 탐사 대상 데이터양에 따른 탐사 수행 시간 및 메모리 소요량을 감소시키기 위해서 데이터 일반화(generalization) 방법을 적용하였고[1, 3, 4, 9, 11, 12], 일부는 잦은 데이터베이스의 스캔에 따른 탐사 수행 시간을 감소시키기 위해서 빈발1-패턴의 조합 또는 일반화된 이력 데이터로부터 패턴 트리(pattern tree)를 생성하여 지지도 계산 및 후보 이동 패턴(candidate moving sequence)에 대한 검증을 수행하였다[1, 7].

본 연구는 한국과학재단 특정기초연구(R01-2007-20989-0) 지원으로 수행되었음

지만 이러한 기법들은 일반화 방법에 있어 실세계의 시간 및 공간에 대한 고려 없이 사용자 정의에 따라 물리적인 형태로 일반화를 수행하기 때문에 일반화된 데이터로부터 실세계의 공간 및 시간에 대한 의미를 얻는 것이 불가능하고, 이동 데이터의 양에 따라 패턴 트리를 생성하고 유지하는 비용이 커지는 문제가 있다 따라서 데이터베이스 접근을 최소화하여 탐사 수행 시간을 단축하거나 이동 객체의 데이터양 최소지지도, 이동 패턴 트리 등과 같은 여러 요인으로 인한 메모리 소요량을 최소화하기 위한 새로운 기법이 필요하다.

또한, 기존의 탐사 기법들은 특정 지점들 간의 빈발한 이동 경로나 정해진 시간 내에서의 효율적인 스케줄링 경로 탐색과 같이 사용자에게 의해 지정된 시공간 범위 내에서 시간 및 공간 제약 조건을 갖는 빈발한 패턴을 탐사하는 문제에는 적용하기 어렵다 따라서 이러한 문제들에 대한 근본적인 해결을 통해 보다 효율적인 탐사 수행 성능을 가진 새로운 시공간 이동 패턴 탐사 기법이 필요하다.

이에 본 논문에서는 방대한 이동 객체의 이력 데이터 집합으로부터 복합적인 시간 및 공간 제약을 갖는 빈발 이동 패턴을 탐사하는 문제를 정의하고 최대 빈발 2-시퀀스를 추출하는 방법을 이용하여 응용에 효과적으로 적용할 수 있는 최적 이동 패턴을 생성하는 방법을 제안한다. 제안된 방법은 특정한 지점들 사이를 이동한 객체의 이력 데이터들 중 객체가 가장 빈번하게 이동한 경로를 탐색하여 최적 경로로 결정하는 패턴 빈발도를 이용한 방법으로, 객체의 위치값과 공간영역 간의 위상 관계를 고려하여 이동 객체의 위치 속성에 대한 최하위 수준에서의 공간 일반화를 통해 보다 효율적으로 패턴 탐사를

수행할 수 있다.

본 논문의 구성은 2장에서 최적 이동 패턴 탐사의 특성에 따른 연구 목적과 범위를 정의한다 3장에서는 최대 빈발 이동 패턴을 추출하기 위한 이동 시퀀스 생성 후보 시퀀스 집합 및 최대 빈발 2-시퀀스 추출 방법과 알고리즘을 제안한다. 마지막으로 4장에서 결론을 제시한다.

2. 연구 목적 및 범위

최적 이동 패턴을 탐사하는 문제는 기존의 패턴 탐사 문제와는 다르게 탐색하고자 하는 대상 패턴을 추출하기 위해서 시간적인 제약이나 공간적인 제약 혹은 두 제약 모두를 수반하게 된다. 이러한 제약은 기존의 패턴 탐사 문제에서 단순한 탐사 범위의 시간적인 혹은 공간적인 한정과는 다르다. 기존의 탐사 기법들은 시간 및 공간 한정자에 의해 제한된 범위의 시간 및 공간 영역에서 반복적이거나 주기적 혹은 공통적으로 발생하는 빈발 패턴 탐사를 주목적으로 함으로써 특정한 제약을 가진 패턴을 탐사하는 문제에 쉽게 대처할 수 없다[1, 2, 3, 6, 13]. 최적 이동 패턴 탐사는 특정한 지점들 간을 이동한 객체들의 패턴들 중 가장 빈번하게 발생하는 이동 패턴을 탐사하기 위한 방법이다 즉, 특정한 지점들 사이의 최소지 지도를 만족하는 모든 빈발 이동 시퀀스들을 탐색하여 최대의 지지도를 가지는 패턴들을 하나의 최적 패턴으로 결정하는 기법이다 제안된 방법의 기본 개념은 특정한 지점 간을 이동하는 다양한 패턴들 중 가장 빈번하게 발생하는 패턴이 가장 최적의 비용을 소요할 것이라는 가정을 기반으로 한다

최적 이동 패턴 탐사는 빈발 패턴을 탐사하는 방식에 있어 기존의 Apriori 계열의 빈발 패턴 탐사 방법과는 차이가 있다. 기존의 Apriori 계열의 알고리즘들은 대부분 빈발 패턴을 탐사하는데 있어 k-시퀀스 집합이 (k+1)-시퀀스 집합을 발견하기 위해서 시퀀스를 구성하는 항목의 차수가 점차 확장되는 패스 단위의 반복 접근법을 사용한다. 제안하는 방법에서는 패스 단위의 반복 접근법을 사용하지만 각 패스마다 서로 다른 길이를 가진 후보 시퀀스 집합으로부터 길이가 2인 빈발 시퀀스를 탐사하는 과정에 있어 후보 시퀀스들의 항목 차수가 점차 줄어드는 패스 단위의 반복 접근법을 사용한다 즉, 기존의 Apriori 계열의 알고리즘들은 빈발 k-시퀀스 집합으로부터 빈발 (k+1)-시퀀스 집합을 추출하는데 있어 빈발 k-시퀀스 집합 자신을 셀프조인(self join)하여 후보 시퀀스 집합을 생성하기 때문에 후보 시퀀스 항목의 차수가 하나씩 증가하지만 제안된 방법은 i번째 패스에서 추출된 빈발 2-시퀀스의 첫 번째 항목이 i+1번째 패스에서 생성되는 후보 이동 시퀀스들의 항목에 포함되지 않아야 하므로 후보 시퀀스 항목의 차수가 점차 감소하도록 한다. 이러한 특성은 패턴 탐사를 수행하는 각각의 패스에서 후보 시퀀스 집합으로부터 지지도 계산을 통해 추출되는 빈발 2-시퀀스들의 순차적인 조합으로부터 유도되어 최적 이동 패턴 탐사 방법에 적용된다

이러한 정의와 연구 범위를 기반으로 3장에서는 빈발도

가 가장 높은 최적 이동 패턴을 생성하기 위한 최대 빈발 2-시퀀스 추출 방법을 제안한다

3. 최대 빈발 2-시퀀스 추출

최대 빈발 2-시퀀스 추출 방법은 이동 시퀀스 집합 추출 단계인 전처리 과정을 거쳐 후보 시퀀스 집합을 생성하고, 이를 이용하여 최대 빈발 2-시퀀스를 추출한다. 각 단계의 설계과정은 최적 이동 패턴 탐사가 공간 상의 특정 지점들 간을 이동 객체가 이동하려 할 때 가장 최적의 경로에 해당하는 이동 패턴을 탐사하는 것을 목적으로 하기 때문에 상세한 수준에서의 이동 경로를 제공해야 함을 기반으로 한다.

3.1 전처리 단계(이동 시퀀스 생성 단계)

이동 시퀀스는 이동 객체의 이력 데이터를 각 이동 객체 별로 시간 순차에 따라 나열한 한 쌍의 공간 및 시간 속성값으로 구성된 순차 리스트이다 이러한 이동 시퀀스의 생성은 표 1과 같이 이동 객체 데이터베이스를 각 이동 객체별로 유효시간을 기준으로 정렬하여 추출하고 시퀀스를 구성하는 위치들 간에 최대 시간 간격에 대한 제약 조건을 두어 이를 만족해야만 의미있는 이동 시퀀스로 생성될 수 있도록 한다.

표 1. 이동 객체의 이력 데이터 예

OID	X	Y	VT	OID	X	Y	VT
1	84	166	2005/11/20/12/10/08	6	84	166	2005/11/20/12/10/24
	108	146	2005/11/20/12/21/19		108	146	2005/11/20/12/27/15
	121	133	2005/11/20/12/29/34		88	125	2005/11/20/12/36/34
	95	99	2005/11/20/12/43/36		95	99	2005/11/20/12/50/45
	139	108	2005/11/20/12/56/16		139	108	2005/11/20/13/02/12
	157	90	2005/11/20/13/04/25		157	90	2005/11/20/13/10/54
	191	81	2005/11/20/13/11/16		191	81	2005/11/20/13/22/24
	226	66	2005/11/20/13/15/16		226	66	2005/11/20/13/31/14
2	88	125	2005/11/20/12/02/55	7	78	90	2005/11/20/12/37/43
	95	99	2005/11/20/12/09/41		100	70	2005/11/20/12/45/39
	116	87	2005/11/20/12/15/15		116	87	2005/11/20/12/53/26
	139	72	2005/11/20/12/21/13		139	72	2005/11/20/12/59/34
	157	90	2005/11/20/12/26/15		157	90	2005/11/20/13/16/14
	191	81	2005/11/20/12/33/34		191	81	2005/11/20/13/27/12
	226	66	2005/11/20/12/41/45		226	66	2005/11/20/13/36/32
3	84	166	2005/11/20/12/33/35	8	116	87	2005/11/20/12/14/35
	108	146	2005/11/20/12/54/14		100	70	2005/11/20/12/23/54
	121	133	2005/11/20/13/15/11		119	52	2005/11/20/12/33/27
	139	108	2005/11/20/13/29/12		134	51	2005/11/20/12/39/19
	157	90	2005/11/20/13/43/36		163	48	2005/11/20/12/48/35
	191	81	2005/11/20/13/52/29		191	81	2005/11/20/13/00/26
	226	66	2005/11/20/13/01/19		226	66	2005/11/20/13/07/07
4	95	99	2005/11/20/12/10/05	9	150	169	2005/11/20/12/31/53
	139	108	2005/11/20/12/20/54		156	160	2005/11/20/12/34/24
	157	90	2005/11/20/12/28/29		148	134	2005/11/20/12/42/34
	157	53	2005/11/20/12/39/38		173	108	2005/11/20/12/49/25
	163	48	2005/11/20/13/00/27		199	108	2005/11/20/12/56/43
	191	81	2005/11/20/13/12/06		191	81	2005/11/20/13/05/16
	226	66	2005/11/20/13/20/46		226	66	2005/11/20/13/12/15
5	121	133	2005/11/20/12/15/45	10	195	166	2005/11/20/12/15/04
	95	99	2005/11/20/12/22/54		182	151	2005/11/20/12/20/43
	139	108	2005/11/20/12/34/07		173	108	2005/11/20/12/29/48
	157	90	2005/11/20/12/45/46		199	108	2005/11/20/12/34/23
	191	81	2005/11/20/12/55/15		191	81	2005/11/20/13/20/26
	226	66	2005/11/20/13/04/33		226	66	2005/11/20/13/26/00

시간 제약 조건은 시퀀스 내에서의 연속적인 이동으로

특정 위치에서 인접한 다른 위치로 이동한 경우인접한 위치로 이동이 발생한 시간 $t_j - t_{j-1}$ 을 의미한다. 최대 시간 간격은 max_gap 으로 $t_j - t_{j-1} \leq max_gap$ 이고, $2 \leq j \leq k$ 이다. 즉, 이동 객체의 공간 속성에 대한 샘플링 시간을 검사하여 특정 위치에 머문 시간이 최대 시간 간격 max_gap 을 초과하면 초과 이전까지의 이동 객체의 순차 리스트에 대한 이동 시퀀스와 초과 이후의 이동 시퀀스로 분리한다. 가령, 표 1에서 객체 4의 이동 경로는 각 위치점 간을 이동한 시간 간격을 고려하지 않았을 때 $(95,99) \rightarrow (139,108) \rightarrow (157,90) \rightarrow (157,53) \rightarrow (163,48) \rightarrow (191,81) \rightarrow (226,66)$ 로 표현될 수 있다. 만약 위치점 간의 이동 시간 간격 max_gap 이 최대 1시간이라면 위치점 $(157,53)$ 에서 멈추었다가 다시 출발하는데 걸린 시간이 1시간 이상이기 때문에 이동 경로 $(95,99) \rightarrow (139,108) \rightarrow (157,90) \rightarrow (157,53) \rightarrow (163,48) \rightarrow (191,81) \rightarrow (226,66)$ 는 $\langle (95,99) (139,108) (157,90) (157,53) \rangle$ 와 $\langle (163,48) (191,81) (226,66) \rangle$, 두 개의 시퀀스로 분리되어야 한다. 다음 표 2는 표 1의 이력 데이터를 시간 간격을 고려하여 이동 시퀀스로 생성한 것이다

표 2. 단위 이동 시퀀스 추출 예

OID	X	Y	VT	OID	USID	X	Y	VT
1	84	166	2005/11/20/12/10/08	1	S1	84	166	2005/11/20/12/10/08
	108	146	2005/11/20/12/21/19		S1	1081	146	2005/11/20/12/21/19
	121	133	2005/11/20/12/29/34		S1	21	133	2005/11/20/12/29/34
	95	99	2005/11/20/12/43/36		S1	95	99	2005/11/20/12/43/36
	139	108	2005/11/20/12/56/16		S1	139	108	2005/11/20/12/56/16
	157	90	2005/11/20/13/04/25		S1	157	90	2005/11/20/13/04/25
	191	81	2005/11/20/13/11/16		S1	191	81	2005/11/20/13/11/16
226	66	2005/11/20/13/15/16	S1	226	66	2005/11/20/13/15/16		
...
4	95	99	2005/11/20/17/10/05	4	S4	95	99	2005/11/20/17/10/05
	139	108	2005/11/20/17/20/54		S4	139	108	2005/11/20/17/20/54
	157	90	2005/11/20/17/28/29		S4	157	90	2005/11/20/17/28/29
	157	53	2005/11/20/17/39/38		S4	157	53	2005/11/20/17/39/38
	163	48	2005/11/20/18/51/27		S5	163	48	2005/11/20/18/51/27
191	81	2005/11/20/19/12/06	S5	191	81	2005/11/20/19/12/06		
226	66	2005/11/20/19/20/46	S5	226	66	2005/11/20/19/20/46		
...
7	78	90	2005/11/20/18/37/43	7	S8	78	90	2005/11/20/18/37/43
	100	70	2005/11/20/18/45/39		S8	100	70	2005/11/20/18/45/39
	116	87	2005/11/20/18/53/26		S8	116	87	2005/11/20/18/53/26
	139	72	2005/11/20/18/59/34		S8	139	72	2005/11/20/18/59/34
	157	90	2005/11/20/21/16/14		S9	157	90	2005/11/20/21/16/14
	191	81	2005/11/20/21/27/12		S9	191	81	2005/11/20/21/27/12
226	66	2005/11/20/21/36/32	S9	226	66	2005/11/20/21/36/32		
...

다음 표 3은 표 2의 이동 시퀀스 데이터의 공간상의 특정 지점들 간의 상세 수준의 경로 제공을 위하여 공간 속성을 노드 식별자로 대치하여 생성한 이동 시퀀스 집합이다.

표 3. 이동 시퀀스 집합 예 (공간 속성 일반화)

OID	Sequence
MO1	$\langle N1 N3 N15 N17 N14 N19 N20 N21 \rangle$
MO2	$\langle N16 N17 N22 N23 N19 N20 N21 \rangle$
MO3	$\langle N1 N3 N15 N14 N19 N20 N21 \rangle$
MO4	$\langle N17 N14 N19 N24 \rangle \langle N25 N20 N21 \rangle$
MO5	$\langle N15 N17 N14 N19 N20 N21 \rangle$
MO6	$\langle N1 N3 N16 N17 N14 N19 N20 N21 \rangle$
MO7	$\langle N30 N29 N22 N23 \rangle \langle N19 N20 N21 \rangle$
MO8	$\langle N22 N29 N28 N26 N25 N20 N21 \rangle$
MO9	$\langle N6 N7 N13 N12 N10 N20 N21 \rangle$
MO10	$\langle N3 N15 N14 N19 \rangle \langle N20 N21 \rangle$

3.2 후보 시퀀스 집합 및 최대 빈발 2-시퀀스 추출

후보 시퀀스 집합 생성 단계에서는 탐사 대상 이동 시퀀스 집합으로부터 각 시퀀스를 구성하는 항목들 중 첫 번째 항목이 S이고 마지막 항목이 F인 모든 후보 시퀀스 집합을 추출한 후 S 항목과 이에 직접 연결된 항목으로 구성되는 2-시퀀스 집합을 추출한다 최초의 시작점은 S이고 다음 패스에서의 시작점 S'은 다음 단계인 최대 빈발 2-시퀀스 추출 단계에서 추출하는 빈발 2-시퀀스의 두 번째 항목이 된다. 최대 빈발 2-시퀀스 추출 단계에서는 전체 이동 시퀀스 집합을 읽어 들여 앞선 단계에서 추출한 2-시퀀스 집합에 대해 지지도를 계산한 후 최대 지지도를 갖는 빈발 2-시퀀스를 추출한다. 추출된 최대 빈발 2-시퀀스를 구성하는 두 항목 중 두 번째 항목이 종료점 F와 같지 않으면 후보 시퀀스 집합 생성 단계를 반복하고 같으면 최대 빈발 2-시퀀스들을 순차적으로 연결하여 최적 이동 패턴을 추출하는 마지막 탐사 단계를 수행한다 빈발도 기반의 최적 이동 패턴 탐사는 이동 패턴 탐사를 위해 전처리 과정을 통해 생성된 이동 시퀀스 집합으로부터 최대지지도를 가지는 가장 빈발한 이동 시퀀스를 탐사하는 방법이다 그림 1은 표 3의 이동 시퀀스 집합으로부터 최대지지도를 가진 빈발 이동 시퀀스를 탐사하는 빈발도 기반의 최적 이동 패턴 탐사 예이다

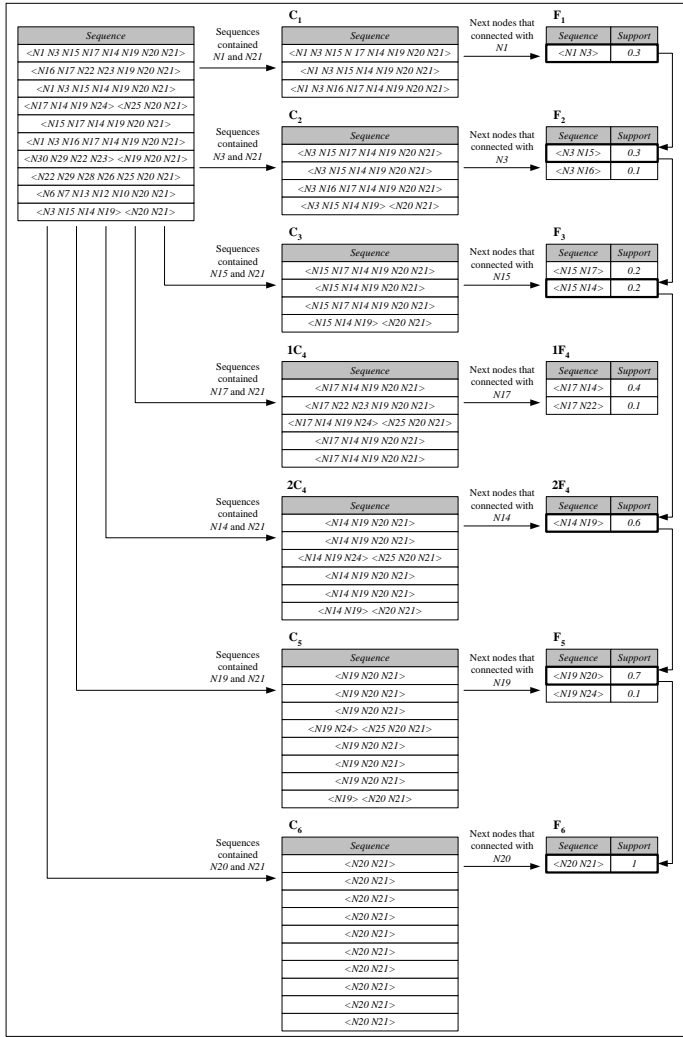


그림 1. 빈발도 기반의 최적 이동 패턴 탐사 예

그림 1에서 출발지점 S가 N1이고 도착지점 F가 N21이라고 할 때, 첫 번째 패스에서는 탐사 대상 이동 시퀀스 집합으로부터 N1과 N21을 시퀀스 항목으로 가지는 모든 이동 시퀀스들을 검색한다. 검색된 이동 시퀀스들은 첫 번째 후보 시퀀스 집합인 C1을 구성한다. C1은 총 3개의 시퀀스로 구성되어지며 각 시퀀스의 첫 번째 항목과 두 번째 항목을 추출하여 부분 2-시퀀스를 생성하고 지지도를 계산하여 빈발 시퀀스 집합인 F1을 구성한다. F1의 시퀀스는 <N1 N3>으로 단 한 개만 생성되기 때문에 이를 최대 빈발 2-시퀀스로 결정한다.

두 번째 패스에서는 첫 번째 패스에서 찾아진 빈발 2-시퀀스인 <N1 N3>에서 두 번째 항목 N3을 새로운 출발지점 S'으로 지정하고 N3과 N21을 시퀀스 항목으로 가지는 모든 이동 시퀀스들을 검색하여 두 번째 후보 시퀀스 집합인 C2를 구성한다. 총 4개의 시퀀스로 구성된 C2에서 첫 번째 패스의 방법과 같이 각 시퀀스의 첫 번째 항목과 두 번째 항목을 추출하여 부분 2-시퀀스를 생성하고 각 시퀀스의 지지도를 계산하여 F2를 구성한다. F2의 시퀀스들로는 <N3 N15>와 <N3 N16>, 총 두 개의 시퀀스들이 생성되어지며 지지도가 각각 0.3과 0.1이기 때문에 최대지지도를 가진 <N3 N15>를 최대

빈발 2-시퀀스로 결정한다.

한편, 세 번째 패스에서는 앞선 과정과 마찬가지로 먼저 N15와 N21을 시퀀스 항목으로 가지는 후보 시퀀스 집합 C3을 구성하고 C3의 각 시퀀스들로부터 첫 번째와 두 번째 항목을 추출하여 부분 2-시퀀스를 생성한다. 부분 2-시퀀스는 <N15 N17>과 <N15 N14>로, 총 두 개의 시퀀스가 생성되며 지지도를 계산하여 F3을 구성한다. 최대 빈발 2-시퀀스의 추출은 빈발 시퀀스 집합의 각 시퀀스에 대한 지지도를 비교하여 추출하게 되는데 F3의 두 부분 2-시퀀스는 모두 동일한 지지도를 가지므로 하나의 최대 빈발 2-시퀀스를 추출할 수 없다. 이러한 경우 두 시퀀스 모두를 대상으로 다음의 탐사 패스를 수행하여 추출되는 모든 부분 2-시퀀스들 중 최대지지도를 가지는 빈발 2-시퀀스를 탐사한다. 탐사된 최대 빈발 2-시퀀스의 항목 중 첫 번째 항목이 이전 패스의 어느 부분 2-시퀀스에 포함되는지 검사하여 해당 부분 2-시퀀스를 이전 패스의 최대 빈발 2-시퀀스로 결정한다.

네 번째 패스에서는 세 번째 패스의 두 부분 2-시퀀스 <N15 N17>과 <N15 N14>에 대해 탐사 과정을 적용한다. 먼저 시퀀스 <N15 N17>은 N15와 N17을 항목으로 가지는 모든 이동 시퀀스들을 검색하여 후보 시퀀스 집합 1C4를 구성하고, 시퀀스 <N15 N14>도 N15와 N14를 항목으로 가지는 모든 이동 시퀀스들을 검색하여 2C4를 구성한다. 그리고 앞선 과정과 마찬가지로 두 후보 시퀀스 집합에서 각 시퀀스의 첫 번째 항목과 두 번째 항목을 추출하여 부분 2-시퀀스들을 생성하고 각각의 지지도를 계산하여 1F4와 2F4를 구성한다. 집합 1F4와 2F4의 모든 2-시퀀스들을 합하여 지지도를 계산하였을 때 가장 높은 지지도를 가지는 <N14 N19> 시퀀스가 네 번째 패스의 최대 빈발 2-시퀀스로 결정된다. 세 번째 패스 단계의 최대 빈발 2-시퀀스로는 네 번째 패스의 최대 빈발 2-시퀀스가 N14를 포함하므로 <N15 N14> 시퀀스가 최대 빈발 2-시퀀스로 결정된다. 이러한 탐사 패스 과정을 반복적으로 수행하여 도착지점인 N21을 포함하는 최대 빈발 2-시퀀스가 결정되면 탐사 과정을 마친다. 최종적으로 각 패스 단계에서 결정된 최대 빈발 2-시퀀스를 조합하여 하나의 최적 이동 패턴을 생성한다. 그림 1에서는 <N1 N3 N15 N14 N19 N20 N21> 시퀀스가 최적의 이동 패턴으로 결정된다.

3.3 최적 이동 패턴 추출 알고리즘

- 최적 경로 추출 알고리즘: OptPathExtractor

OptPathExtractor 함수는 최적 이동 패턴을 구성하는 각각의 단위 패턴을 찾기 위해 최초의 출발점 S를 OptPath 변수에 저장하고 S와 연결된 노드를 탐색하기 위해 Freq_Link 함수를 호출한다.

```

Input :  $D_{gen}$  (Generalized Moving Sequence),  $S$  (Start Node),  $F$  (Final Node)
Output :  $OptPath$ 

Procedure  $OptPathExtractor (D_{gen}, S, F)$ 

Begin
    PrevA = null;
    OptPath = null;
    nextN = S

    While (nextN  $\neq$  F) {
        OptPath add nextN;
        nextN = Freq_Link(nextN, F,  $D_{gen}$ );
    }

    OptPath add F;

    Return OptPath;

End

```

그림 2. OptPathExtractor 알고리즘

● 최대 빈발 2-시퀀스 추출 알고리즘 : Freq_Link

Freq_Link 함수는 최적 이동 패턴을 추출하는 함수인 OptPathExtractor에서 전달받은 전체 이동 시퀀스 집합으로부터 출발지점 S와 도착지점 F를 포함하는 부분 이동 시퀀스 집합을 추출하여 최대 빈발 이동 패턴을 탐사하게 된다. 이동 패턴을 탐사하는 과정은 최초 S로부터 시작하는 1-최대 빈발 패턴을 탐색하고 그 패턴과 연결된 S가 아닌 반대편 노드를 기준으로 하여 다시 F를 포함하는 모든 부분 이동 시퀀스 집합을 추출하여 다시 1-최대 빈발 패턴을 추출한다. 이러한 과정은 F에 도달하는 1-최대 빈발 패턴을 탐사할 때까지 반복 수행한다

```

Input : currN (Current Node), F (Final Node),  $D_{gen}$  (Generalized Moving Sequence)
Output : Node

Procedure  $Freq\_Link (currN, F, D_{gen})$ 

Begin
    subD = Search Sequence that contained Node currN and Node F in  $D_{gen}$ ;
    For each moving sequence  $S \in subD$ 
         $c = subset(currN, S)$ ; // Link from currN to nextN that connected with currN
        If  $c \notin C_s$  Then // Set of Links between the decided Node and optional Node
             $C_s$  add  $c$ ;
        End If
        count++;
    End For

     $c_{max\_sup} = find\ 2-sequence\ with\ maximum\ support\ from\ C_s$ ;
    nextN = another Node that connected with currN in  $c_{max\_sup}$ ;

    Return nextN;

End

```

그림 3. Freq_Link 알고리즘

최대 빈발 2-시퀀스 추출을 통한 최적 이동 패턴 탐사는 다음과 같은 특징을 갖는다. 첫째, 기존의 Apriori 계열의 빈발 패턴 탐사 기법들은

각 탐사 패스에서 후보 시퀀스 집합을 이전 패스의 빈발 시퀀스를 이용하여 조인연산이나 곱 연산 등을 통해 생성하지만, 제안된 기법에서는 탐사 대상 이동 시퀀스 집합으로부터 특정한 공간 제약예를 들면 시작점, 종료점)을 만족하는 시퀀스들만을 추출하여 후보 시퀀스 집합을 생성한다. 또한, 한정된 공간 영역에서의 불특정한 빈발 패턴을 탐사하는 것이 아닌 사용자에 의해 의도된 공간상에서의 빈발 패턴을 탐사하는 것을 목적으로 한다.

둘째, 기존의 경로 탐색 기법들이 최적 경로를 탐색하기 위한 조건으로 이동시간이나 이동거리 소요비용 등을 기준으로 최소의 값을 갖는 경로들을 최적의 경로로 선택했던 방식과는 다르게 제안된 방법은 이동 객체의 실제 이동 경험치를 분석하여 빈발도가 가장 높은 경로를 최적 경로로 선택하는 방식이다 이는 각기 서로 다른 목적으로 특정한 위치를 목표로 하여 이동하는 객체들의 경우 스스로의 판단 기준에 따라 최적의 경로일 확률인 높은 경로를 선택하여 움직이는 성향을 가지고 있기 때문에 빈발도가 높은 이동 패턴이 최적의 이동 패턴이 될 수 있다.

4. 결 론

이동 객체의 위치 이력 데이터를 이용하여 다양한 응용 분야에서 활용할 수 있는 새로운 위치 기반 서비스를 개발하기 위한 목적으로 다양한 시공간 이동 패턴 탐사 기법에 대한 연구가 진행되어 왔으나 기존의 패턴 탐사 연구를 통해 개발되어 실질적으로 실생활에 적용되고 있는 위치 기반 서비스는 아직 미비한 실정이다 이에 본 논문에서는 시공간 패턴 탐사를 통한 실세계에 적용가능한 위치 기반 서비스의 응용으로 방대한 이동 객체의 이력 데이터 집합으로부터 복합적인 시간 및 공간 제약을 갖는 빈발 이동 패턴을 탐사하는 문제에 대해 정의하고, 최대 빈발 2-시퀀스 추출을 통한 최적 이동 패턴을 탐사하는 방법을 제안하였다 제안된 방법은 실 세계에서 특정 목적을 가진 이동 객체들이 가장 빈번하게 이동한 경로를 탐색하여 최적 경로로 결정하는 패턴 빈발도를 이용한 탐색 방법으로 패턴 탐사를 수행하는데 있어 SeqExtractor 연산을 통해 이동 객체의 이력 데이터로부터 이동 시퀀스를 추출하고 객체의 위치값과 공간영역 간의 위상 관계를 고려하여 이동 객체의 위치 속성에 대한 최하위 수준에서의 공간 일반화를 통해 일반화된 이동 시퀀스 집합을 생성하여 보다 효율적인 패턴 탐사를 수행할 수 있다. 또한 실제적으로 최대 빈발 2-시퀀스를 탐사하는 Freq_Link와 최적이동 패턴을 추출하는 OptPathExtractor 연산을 통해 패턴 탐사를 수행하도록 알고리즘을 설계하였다

제안 방법에서 빈발 패턴을 탐색하기 위한 각 패스 진행 시 후보 시퀀스 항목의 차수가 점차 감소하는 특성은 각각의 패스에서 후보 시퀀스 집합으로부터 지지도 계산을 통해 추출되는 빈발 2-시퀀스들의 순차적인 조합으로부터 유도되어 최적 이동 패턴 탐사 방법에 효과적으로 적용된다.

참고문헌

- [1] 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.
- [2] 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.
- [3] Y. Huang, L. Zhang and P. Zhang, "Finding Sequential Patterns from a Massive Number of Spatio-Temporal Events", SDM, SIAM, 2006.
- [4] 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.
- [5] S. Y. Hawng, Y. H. Liu, J. K. Chiu and E. P. Lim, "Mining Mobile Group Patterns : A Trajectory- based Approach", PAKDD, Lecture Notes in Computer Science, Springer, Vol.3518, pp.713-718, 2005.
- [6] Y. Wnag, E. P. Lim and S. Y. Hwang, "On Mining Group Patterns of Mobile Users", DEXA, Lecture Notes in Computer Science, Springer, Vol.2736, pp.287-296, 2003.
- [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] P. Laube and S. Imfeld, "Analyzing Relative Motion within Groups of Trackable Moving Point Object", in GIScience, Notes in Computer Science, Springer, Vol.2478, pp.132-144, 2002.
- [9] 백옥현, "위치 기반 서비스를 위한 이동 객체의 시간 패턴 탐사 기법", 충북대학교 대학원, 석사학위논문, 2002.
- [10] 이준욱, "위치 기반 서비스를 위한 이동 객체의 시간 패턴 탐사", 한국정보과학회 논문지, 제29권, 제5호, 2002.
- [11] 이준욱, "지식 탐사 프레임워크 기반의 시공간 이동 패턴 탐사 기법", 충북대학교 대학원, 박사학위논문, 2003.
- [12] 한선영, "시공간 이동 시퀀스 패턴 마이닝 기법, 이화여자대학교 대학원, 석사학위논문, 2006.
- [13] 박지웅, "시공간 이동 패턴 추출을 위한 효율적인 알고리즘", 건국대학교 대학원, 박사학위논문, 2006.