

셀룰러 공간에 존재하는 이동객체 궤적의 유사성 측정

강 혜 영* · 김 준 석* · 황 정 래** · 이 기 준***

Similarity measures for trajectories of moving objects in cellular space

Hye-Young Kang* · Joon-Seok Kim* · Jung-Rae Hwang** · Ki-Joune Li***

요 약

대부분의 지리정보 시스템들은 유클리디안 공간을 기반으로 하고 있지만, 셀룰러 공간 역시 많은 지리정보 시스템 응용을 위한 기반 공간으로 사용될 수 있다. 셀룰러 공간에 존재하는 이동객체의 이동 형태를 분석하기 위해서는, 유클리디안 공간과 분명히 다른 특징을 갖는 셀룰러 공간상에 존재하는 이동객체들의 유사도에 대한 새로운 정의가 필요하다. 본 논문에서는 셀룰러 공간에 존재하는 이동객체의 특성에 대해 살펴보았다. 이에 따라, 셀룰러 공간에 존재하는 궤적들 간의 유사도를 정의할 수 있는 2가지의 비교지수를 제시하고, 실험을 통해 제시한 비교 지수들 간의 차이를 분석하였다.

주요어 : 이동객체 궤적, 유사도, 셀룰러 공간

ABSTRACT : While most GIS are based on Euclidean space, cellular space can be used as an alternative type of space for a large number of GIS applications. In order to analyze the pattern of moving objects in cellular space, we need new definitions of similarity between their trajectories since the trajectory in cellular space significantly differs from those in Euclidean space. In this paper, we study the properties of moving objects in cellular space. Based on these observations, we propose several similarity measures between trajectories in cellular space. We analyze the difference of the proposed measures by experiments.

Keywords : moving objects trajectory, similarity, cellular space

*부산대학교 컴퓨터공학과 박사과정

**부산대학교 컴퓨터 및 정보통신연구소 기금교수

***부산대학교 정보컴퓨터공학부 교수

1. 서 론

최근 몇 년 사이, 이동 컴퓨팅과 센서, GPS (Global Positioning System) 기술의 발달에 따라 실외뿐만 아니라 실내에서도 대용량의 시공간 데이터 수집이 가능하게 되었다(정지안 외, 2006). 이러한 기술들에 의해 수집된 대용량의 시공간 데이터들은 다양한 응용분야에서 사용되기 위해서 분석할 필요성이 증가되었다.

데이터 마이닝은 이러한 데이터들의 주요 응용 분야중 하나로써, 데이터 마이닝 분야에서 유사성 검색은 다른 객체들을 유사한 그룹으로 클러스터링하거나 혹은 기준 값에 따라 객체를 분류하기 위한 중요한 연산중의 하나이다(이병길, 2003). 예를 들어, 대형 쇼핑몰에서 유사한 경로를 따라 움직이거나, 특정한 행동 패턴을 따르는 고객들의 공통되는 행동 양식을 발견하는 것은 쇼핑센터의 고객의 성향을 분석하는데 유용하며, 마케팅 계획을 세우는데도 도움이 된다. 이와 같이 이동객체의 궤적의 특정한 패턴을 발견하기 위해서는 유사도 기준이 정의되어야 하고, 유사성 검색이 필요하다.

유사한 이동객체를 찾는 대부분의 연구들이 유클리디안 공간을 기반으로 하고 있으나, 실세계에서 사람이나 차량과 같은 대부분의 이동 객체들은 도로 상이나 실내 공간에 존재한다. 이러한 이동 객체들의 움직임은 궤적으로 표현되고, 그것들 간의 유사성을 검색하는 연구가 많이 이루어지고 있다. 특히, 실내 공간과 같은 셀룰러 공간에 존재하는 이동 객체들의 움직임을 분석하는 연구가 최근에 소개되고 있다. 하지만, 기존의

대부분의 유사성을 검색하는 방법은 유클리디안 공간을 기반으로 하거나 도로 네트워크 공간을 기반으로 이루어졌다. 기존의 대부분의 방법에서 유클리디안 공간이나 도로 네트워크 공간의 이동 객체들의 궤적은 (x, y, t) 의 집합으로 표현되었다. 하지만, 실세계에서의 위치는 (x, y, t) 와 같은 좌표보다는 셀에 의해 참조된다. 예를 들어, “422호에는 누가 있는가?”라는 공간 질의를 처리하기 위해서는 방번호로 참조되는 위치가 (x, y, t) 와 좌표보다 더 유용하다.

셀룰러 공간에서의 공간 표현방법은 유클리디안 공간에서의 공간 표현방법과 확연히 다르며, 셀룰러 공간에 존재하는 이동객체의 궤적 역시 다르게 표현되어야 한다. 따라서 유클리디안 공간이나 도로 네트워크 공간에 존재하는 이동 객체들의 궤적들 간의 유사성을 측정하는 방법은 셀룰러 공간의 이동 객체들의 궤적에 적합하지 않다.

본 논문에서, 우리는 셀룰러 공간에서 이동하는 객체들의 궤적간의 유사성에 관심을 둔다. 우리는 셀룰러 공간이라는 새로운 공간을 정의하고, 그 안에서 움직이는 이동객체의 궤적을 표현하는 방법에 대해서 논의한다. 또한, 셀룰러 공간에서 이동객체의 궤적이 가지는 특징을 반영하여, 이동객체의 궤적들 간의 유사도를 측정하는 방법을 제안한다.

본 논문은 다음과 같이 구성된다. 2장에서는 관련연구와 연구 동기에 대하여 소개한다. 3장에서는 셀룰러 공간에 존재하는 이동객체의 궤적 표현방법과 유사도를 측정하는 방법을 제안한다. 실험결과는 4장에서 보이며, 마지막으로 5장에서 결론을 맺는다.

2. 관련 연구

이동 객체의 궤적을 다루는 연구는 크게 다음과 같이 세 가지 이슈를 가진다.

- 이동 객체의 위치 정보를 획득하는 방법
- 이동 객체의 궤적을 표현하는 방법
- 이동 객체 궤적들에 대한 유사성 측정 방법

이동 객체의 위치 정보를 획득하기 위해서 이동 객체 추적에 관한 많은 연구가 이루어지고 있다. 끊임없이 바뀌는 이동 객체의 위치정보를 매순간마다 유지하는 것은 통신 부하를 가중시키고, 위치 정보 획득을 위한 자원의 사용이 증가되어 비효율적이다(정재승 외, 2006). 이에, 이동 객체의 위치를 추적하기 위한 다양한 방법들이 연구되고 있다. 본 논문에서는 이동 객체의 위치 정보가 이미 존재한다는 것을 가정하고 유사성 측정 방법을 제시하므로, 이동 객체의 추적에 관한 것은 다루지 않는다. 따라서, 본 장에서는 기존의 관련 연구 중에서 이동 객체의 궤적을 표현하는 방법과 유사성 측정 방법에 대하여 살펴보고자 하겠다.

먼저, 이동 객체의 궤적을 표현하는 방법과 관련한 많은 연구가 이루어져 왔다. 기존의 이동 객체의 궤적을 표현하는 방법은 대부분 유클리디언 공간을 기반으로 하고 있다(Ralf 외, 2000; Ludger 외, 2004). 유클리디언 공간에서 이동 객체의 위치는 (x,y) 혹은 $(x,y,time)$ 으로 표현되며, 그 궤적의 정보는 $TR = \{(x_1, y_1, t_1), (x_2, y_2, t_2), \dots, (x_n, y_n, t_n)\}$ 과 같은 형태로 표현된다.

그러나 실세계에 존재하는 대부분의 이동 객체들은 단순히 유클리디언 공간보다는 도

로 네트워크 공간이나 셀룰러 공간 등에서 움직이고 있다. 도로 네트워크 상의 이동 객체들의 궤적에 대한 다양한 표현 방법과 질의 처리 방법들은 많이 연구되어온 반면(Michalis, Ouri, 2001; Nirvana, Rolf, 2003), 실내 공간과 같은 셀룰러 공간에 존재하는 이동 객체의 궤적에 대한 연구는 아직 미비하다. 셀룰러 공간에 존재하는 이동 객체의 위치를 표현하기 위하여 몇몇 방법들이 제안되었다(Christian Frank, 2005). 이 연구에서는 실내 공간에서의 위치 정보를 표현하기 위한 심볼릭 위치 모델들을 제시하고 있다. 그러나 제시한 3가지 위치 모델은 모두 개념적인 것으로, 실제의 기술적인 측면에 대한 고려가 없다.

둘째로, 이동 객체들의 궤적에 대한 유사성은 최근 많이 연구되고 있다. 이동 객체 궤적들 간의 유사성을 정의하는데 있어서, 가장 중요한 것은 거리함수이다. 기존의 많은 연구들이 이러한 민콥스키 거리의 L_2 거리인 유클리디언 거리 함수를 사용하고 있다(Rakesh 외 1993; Kin, Ada, 1999). 이러한 방법들은 각각의 값들을 벡터로 맵핑한 후, 각 점간의 유클리디언 거리를 이용하여 유사성을 측정하는 것이다. 유클리디언 거리 함수를 사용하여 계산이 용이한 장점이 있지만, 서로 다른 길이의 궤적에 대해서는 적용이 불가능하다는 단점이 있다.

이동 객체들의 궤적에 대한 유사성을 측정함에 있어서 또 다른 방법으로, 시간정합(time wraping)을 이용하는 연구들이 진행되었다. Eamonn과 Michael(2000)은 서로 다른 길이를 갖는 시계열 데이터들 간의 유사성을 정의하는 방법을 제시한다. 또한 Yasushi 외(2005)는 시간 정합 거리의 근사값을 새로운 하한거리로 사용하는, 동적 시간 정합(Dynamic

Time Wrapping)을 위한 빠른 검색 방법을 제시하였다. 그러나, 이러한 방법들은 대부분 두 궤적간의 유사성을 측정하는데 있어서, 샘플링된 위치의 순서만을 중요하게 생각하였다. 이것은 시간 차원은 궤적간의 유사성을 판별하는데 있어서 무시되고 있다는 것을 의미한다. 이에, Nikos 외(2007)는 궤적 데이터 베이스에서의 궤적간의 유사성 검색을 크게 시공간 유사성과 공간 유사성으로 분류하고, 각각에 대한 유사성 검색 방법을 제시하였다.

실제로 셀룰러 공간에 존재하는 이동 객체의 궤적들 간의 유사성을 비교한 연구는 없으나, 궤적의 모양에 따른 유사성 비교 연구 가운데, Bin과 Jianwen(2008)이 제시한 그리드 표현방법을 이용한 유사성 비교 방법은 셀룰러 공간에서의 그것과 비슷한 형태를 보인다. Bin과 Jianwen은 기존의 선형적 표현 방법을 이용한 궤적들 간의 유사성 비교의 계산 비용이 너무 비싸다는 점에 착안하여, 궤적을 그리드로 표현하였다.

3. 셀룰러 공간의 이동 객체 궤적과 유사성

본 장에서는 셀룰러 공간의 기본 개념과 궤적의 표현방법을 소개하고, 이를 기반으로 궤적의 시공간적 속성을 반영할 수 있도록 최장 공통 부분열(Longest Common SubSequence)방법을 확장하여 궤적간의 유사성을 측정할 수 있는 지수를 제안한다.

3.1 셀룰러 공간의 이동객체 궤적

셀룰러 공간과 유클리디안 공간의 기본적

인 차이점은 셀룰러 공간은 공간적 속성의 기하적 표현을 기반으로 하지 않는다는 점이다. 본 논문에서는 단순성을 위하여 다음과 같은 속성들을 가정한다.

- 이동 객체가 존재하는 전체 공간은 셀들의 집합으로 이루어져 있다.
- 셀들 간의 겹침은 존재하지 않는다.
- 모든 이동 객체들은 반드시 하나의 셀 안에 존재한다.

이러한 가정에 따라, 이동 객체의 위치는 이동 객체가 위치한 셀 아이디어에 의해 참조된다. 셀룰러 공간의 셀과 유사한 개념이 액티브 맵지 위치 시스템에 의하여 소개되기도 하였다(Roy 외, 1992).

이동 객체의 위치를 셀 아이디어로 표현한다면, 이동 객체의 궤적은 이동 객체가 방문한 셀의 순서로 표현할 수 있다. 따라서, 셀룰러 공간에 존재하는 이동 객체 m 의 궤적 TR 은 다음과 같이 표현된다.

$$TR(m) = \{(c, I) \mid c \in C, I = [t_{in}, t_{out}]\}$$

여기서, t_{in} 과 t_{out} 은 객체 m 이 셀 c 에 진입한 시간과 탈출한 시간을 의미하며, c 는 이동 객체 m 이 시간 구간 I 동안 존재한 셀을 의미한다. 그림 1에서는 본 논문에서 제안하는 셀룰러 공간에 존재하는 이동 객체의 궤적의 예를 보인다. 서로 다른 이동 객체 a, b, c 가 각각 주어진 그림과 같이 움직였을 때, 이동 객체들의 위치는 그림 1과 같이 단순히 이동 객체가 존재했던 셀의 아이디어와 해당 셀에 머무른 시간 구간으로 표현된다. 그림 2는 그림 1에서 표현된 이동 객체의 궤적을

시간의 흐름에 따라 표현한 그림이다.

따라서 두 궤적간의 유사성은 이러한 궤적의 표현을 기반으로 정의되어야 한다. 비록, 이러한 방법에 따라 셀룰러 공간에서의 궤적들 간의 유사성을 비교하는 연구가 없지만, 이와 유사한 접근 방법으로 문자열간의 유사성을 비교하는 연구가 있다. 예를 들어, 최장 공통 부분열(LCSS)을 통하여 두 문자열간의 유사성을 정의하는 방법이 제안되었고(Daniel, 1977), 문자열은 셀룰러 공간의 궤적을 표현하는 셀의 순서열과 동일한 형태로 표현된다(Lasse 외, 2000). 이것은 최장 공통 부분열 방법이 셀룰러 공간의 이동 객체의 궤적들 간의 유사성을 정의하는데 확

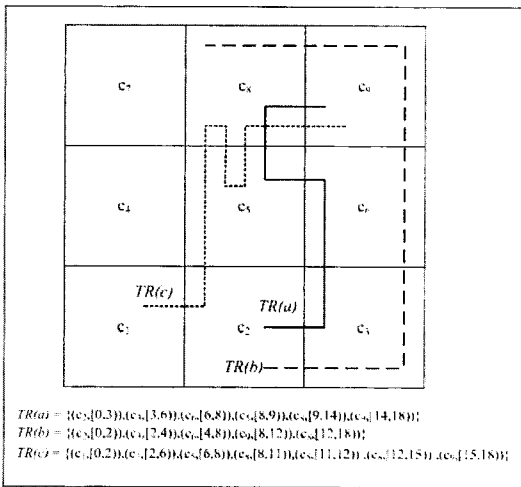
장되어 사용될 수 있음을 의미한다. 하지만, 최장 공통 부분열 방법을 셀룰러 공간에 존재하는 궤적들 간의 유사성 측정에 사용하기에는 2가지 문제점이 존재한다.

- 문자열 비교에서는 시간적 측면이 고려되지 않는다.
- 문자열의 출현 순서와 달리, 셀들의 출현 순서에는 셀들 간의 위상과 같은 공간적 특성이 반영되어 있다.

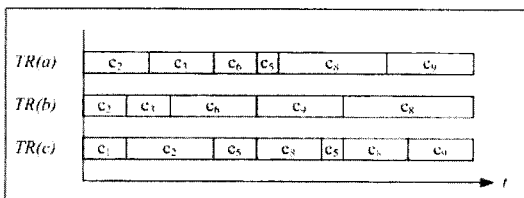
이에, 본 논문에서는 셀룰러 공간에서의 궤적들 간의 유사성 측정에 LCSS를 확장하여 사용하고자 한다. 또한, LCSS의 문제점을 해결하기 위하여, 본 논문에서는 궤적들 간의 공간적 유사성뿐만 아니라, 시간적 측면을 반영하고자 한다.

3.2 최장 공통 부분열(Longest Common Subsequence)

본 논문에서 제안하는 이동 객체 궤적간의 유사성을 측정하는 방법은 Daniel(1977)에 의해 제안된 LCSS 방법을 이용하는 것이다. 이동 객체 a, b 가 각각 n, m 개의 셀을 차례대로 방문하였을 경우, 두 이동 객체의 궤적들 간의 유사성을 LCSS의 정의를 이용하여 재정의하면 다음과 같다.



[그림 1] Example of trajectories in cellular space



[그림 2] Visiting time intervals at each cell

정의 1. LCSS

$$LCSS_{a,b}(i, j) = \begin{cases} 0, & \text{만약 } i = 0 \text{ 또는 } j = 0 \\ LCSS_{a,b}(i-1, j-1) + 1, & \text{만약 } c_i > 0, j > 0, a_i = b_j \\ \max\{LCSS_{a,b}(i, j-1), LCSS_{a,b}(i-1, j)\} + 1, & \text{그 외} \end{cases}$$

정의 1에서, a_i 와 b_j 는 객체 a 와 b 가 각각 i 번째와 j 번째 방문한 셀을 의미한다. $LCSS_{a,b}(i, j)$ 는 이동 객체 a 의 i 번째 셀과 이동 객체

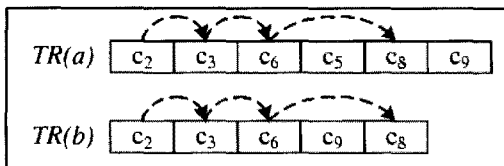
b 의 j 번째 셀까지의 최장 공통 부분열의 길이를 의미한다. 위와 같은 재귀적 정의에 의해, 그림 1에서 보았던 이동 객체 궤적 $TR(a)$ 와 $TR(b)$ 간의 LCSS값을 표 1과 같이 구할 수 있다. 표 1에서 i 는 이동 객체 a 의 방문 셀의 색인이고, j 는 이동 객체 b 의 방문 셀의 색인이며 표의 각 값은 a 객체의 i 번째 셀과 b 객체의 j 번째 셀까지의 최장 공통 부분열의 길이를 나타낸다.

예를 들어, 이동 객체 궤적 $TR(a)$ 와 $TR(b)$ 의 최장 공통 부분열의 길이는 4이고, 최장 공통 부분열은 $\{(c_2,(1,1)), (c_3,(2,2)), (c_6,(3,3)), (c_8,(5,5))\}$, $\{(c_2,(1,1)), (c_3,(2,2)), (c_6,(3,3)), (c_9,(6,4))\}$ 로 2개이다. 그림 3은 $TR(a)$ 와 $TR(b)$ 의 최장 공통 부분열의 예를 나타낸다. 최장 공통 부분열의 표현에서 원소 $(c_i,(i,j))$ 는 이동 객체 궤적 a 의 i 번째 셀과 b 의 j 번째 셀이 c 로 동일함을 의미한다.

두 궤적간의 LCSS값이 크다는 것은, 두

<표 1> An example of LCSS table

i \ j	1	2	3	4	5
1	1	1	1	1	1
2	1	2	2	2	2
3	1	2	3	3	3
4	1	2	3	3	3
5	1	2	3	3	4
6	1	2	3	4	4



[그림 3] An example of longest common subsequence

궤적이 공통된 셀을 많이 지났음을 의미한다. 이것은 두 궤적의 이동 패턴이 유사함을 나타낸다. 그러나, LCSS방법에서는 시간에 대한 고려가 전혀 없다. 예를 들어, 그림 1에서 이동객체 a 와 c 는 공통적으로 c_9 를 지나지만, 이동 객체 a 와 c 가 c_9 를 방문하는 시간 구간은 서로 다르다.

3.2 공통 방문시간 (Common Visit Time Interval)

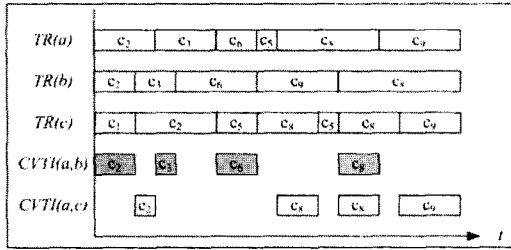
정의 1에서 제시한 LCSS방법은 시간에 대한 고려가 전혀 없다. 이에, 본 논문에서는 두 이동객체가 동일한 셀에 머문 시간의 크기를 유사도의 정의에 반영하고자 한다. 만약, 두 객체가 서로 동일한 시간구간 동안 동일한 셀에 머물렀다면, 두 객체의 궤적은 시간구간의 겹침이 없는 경우보다 더 유사하다고 간주할 수 있다. 이러한 관점에서, 본 논문에서는 이동객체 a 와 b 의 궤적간의 공통 방문시간(CVTI)을 통한 유사성을 정의한다. n 과 m 이 각각 이동 객체 a 와 b 가 방문한 셀의 개수를 나타낸다고 할 때, 공통 방문 시간은 다음과 같이 정의할 수 있다.

정의 2. 공통 방문 시간(CVTI)

$$CVTI(a,b) = \sum_{i=1, j=1}^{n,m} |a_i \cdot I \cap b_j \cdot I|, \text{ where } a_i \cdot c = b_j \cdot c$$

정의 2에 따라, CVTI는 두 이동 객체 a 와 b 의 전체 궤적에서 동일한 셀에 머문 시간구간의 합을 의미한다. 그림 1에서 나타난 이동 객체 궤적 $TR(a)$ 와 $TR(b)$, $TR(c)$ 간의 공통 방문 시간은 그림 4와 같이 계산된다.

그러나, CVTI의 정의에는 단지, 동일한 시



[그림 4] Common Visit Time Interval

간에 동일한 공간에 있었던 시간만을 고려할 뿐, 이동 객체의 공간적 이동 형태에 대한 고려가 없다. 이에, 본 논문에서는 CVTI를 구할 때, LCSS의 최장 공통 부분열에 나타나는 셀만을 고려하여 공통 방문 시간을 계산하는 방법을 제안한다. 최장 공통 부분열을 S_{LCSS} 라고 하고, 모든 최장 공통 부분열의 집합을 \mathcal{S} 라고 하자. S_{LCSS} 은 앞서 설명한 것과 같이 (c_i, j) 의 순서로 나타나며, 이를 이용하여 공통 방문 시간구간을 다음과 같이 재 정의할 수 있다.

정의 3. $CVTI_{LCSS}$

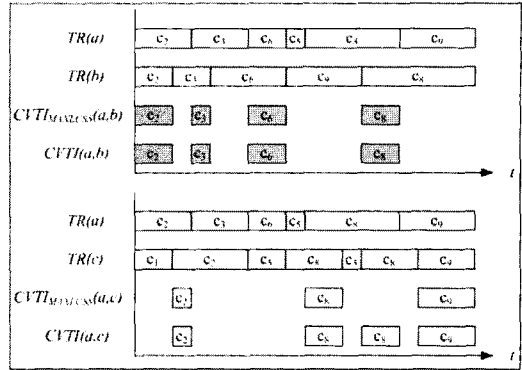
$$CVTI_{LCSS}(S_{LCSS}(a,b)) = \sum_{(c_i, j) \in S_{LCSS}(a,b)} (a_i, I \cap b_j, I)$$

두 궤적간의 최장 공통 부분열이 1개 이상 존재하므로, $CVTI_{LCSS}$ 값 역시 1개 이상 나타나게 된다. 따라서 본 논문에서는 모든 최장 공통 부분열의 공통 방문시간들 중의 최대값을 선택하여 두 이동 객체 궤적 사이의 유사도를 평가한다.

정의 4. $CVTI_{MAXLCSS}$

$$CVTI_{MAXLCSS}(a,b) = \max_{S_{LCSS} \in \mathcal{S}} CVTI_{LCSS}(S_{LCSS})$$

정의 4에 따라서, 정의 3에 의해 구해진 $CVTI_{LCSS}$ 값들 중에서 가장 큰 값을 $CVTI_{MAXLCSS}$



[그림 5] Comparison between CVTI and CVTI_{MAXLCSS}

라고하면, 그림 1에서 나타난 이동 객체 a와 이동 객체 b의 궤적 사이의 $CVTI_{MAXLCSS}(a,b)$ 값은 7이다.

그림 5는 이동 객체 궤적 $TR(a)$ 와 $TR(b)$, $TR(a)$ 와 $TR(c)$ 간의 CVTI와 $CVTI_{MAXLCSS}$ 를 비교하여 나타내고 있다. 궤적 $TR(a)$ 와 $TR(b)$ 의 경우, 그림 5에서 보듯이 CVTI와 $CVTI_{MAXLCSS}$ 값은 동일하게 나타나지만, 궤적 $TR(a)$ 와 $TR(c)$ 은 CVTI값이 $CVTI_{MAXLCSS}$ 값보다 크게 나타난다. 이러한 차이는 이동 객체 궤적 $TR(a)$ 와 $TR(c)$ 의 S_{LCSS} 을 구성하는 각 원소들의 관계에 의해서 발생한다. 그러나, 방문순서를 고려하지 않는 CVTI값은 방문순서를 고려하는 $CVTI_{MAXLCSS}$ 값보다 항상 크거나 같으며, 이것은 CVTI값이 두 궤적사이의 $CVTI_{MAXLCSS}$ 값이 가질 수 있는 상한값이 된다.

4. 실험

본 논문에서는 셀룰러 공간에 존재하는 이동 객체 궤적들 간의 유사성을 측정할 수 있는 기준을 두 가지 제시하였다. 이에, 본 장에서는 본 논문에서 제시한 유사성 측정

지수들의 성능을 분석하기 위하여 몇 가지 실험을 수행하였다.

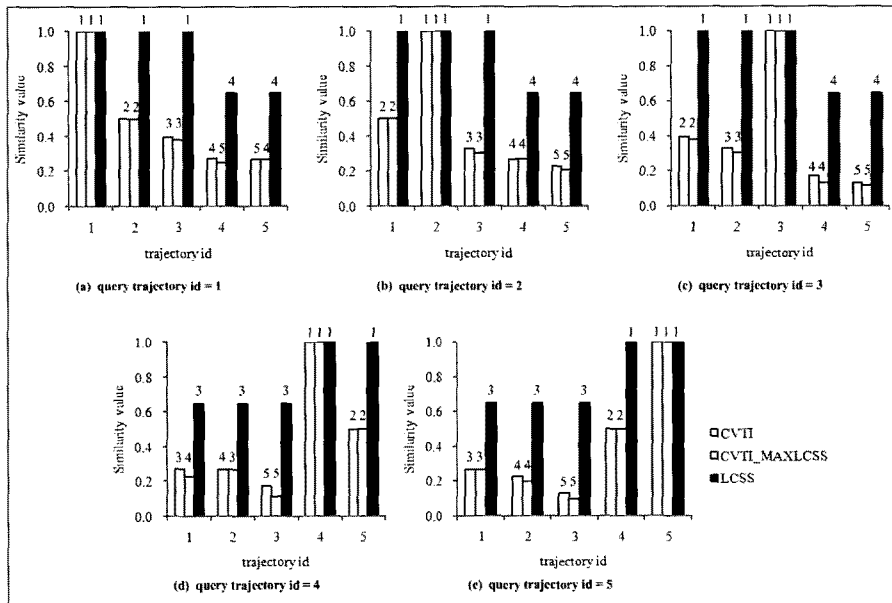
첫째로, 실험을 위하여 데이터 생성기(Yannis, Mario, 2000)로부터 생성한 이동 객체 궤적을 본 논문에서 제시한 셀룰러 공간상의 궤적으로 수정하였다. 또한, 원본 궤적으로부터 서로 다른 속성을 가지는 3세트의 데이터 궤적을 생성하였다. 각각의 데이터 세트들은 원본 궤적과 수정된 4개의 궤적으로 구성된다. 표 2는 생성된 데이터 궤적들의 특징을

원본 데이터와 비교해서 정리한 것이다.

그림 6은 이러한 데이터 셋트에 대하여 LCSS, CVTI, CVTIMAXLCSS 방법으로 각각의 유사도 값을 계산한 평균을 나타낸다. 각 그래프의 x축은 데이터 궤적의 아이디를 나타내고, y축은 유사도 값을 나타낸다. 각 막대위의 숫자는 각 제안된 각 방법에 따른 질의 궤적에 대한 각 데이터 궤적의 유사도 순위를 의미한다. 예를 들어, 그림 6-a)에서 LCSS 방법에 따르면 궤적 TR_1 , TR_2 와 TR_3 의 순위가 동

<표 2> Properties of derived data trajectories

데이터 궤적 아이디	절대적 시간영역	셀에서 머문 시간의 크기	방문 순서
TR_1	원본	원본	원본
TR_2	다름	원본과 동일	원본과 동일
TR_3	원본과 동일	다름	원본과 동일
TR_4	원본과 동일	원본과 동일	다름
TR_5	다름	원본과 동일	다름
TR_6	다름	다름	다름

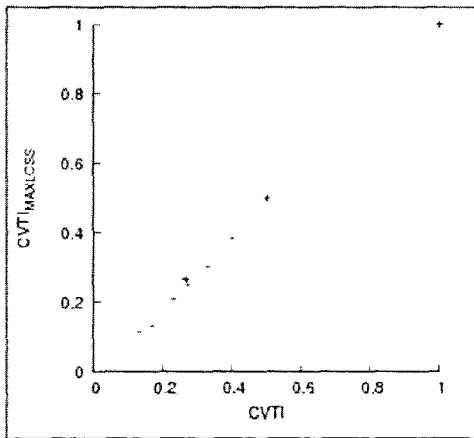


[그림 6] Average similarity value between data trajectories

일한 반면, CVTI와 $CVTI_{MAXLCSS}$ 방법에서는 각 궤적들이 서로 다른 유사도 순위를 가진다. 이것은 LCSS 방법이 셀룰러 공간의 궤적의 유사도를 측정하는데 있어서 부적합함을 의미한다. 대부분의 경우, CVTI와 $CVTI_{MAXLCSS}$ 방법에 대하여 각각의 데이터 궤적의 순위가 동일한 순위를 가지지만, 그림 6-d)와 같이 순위가 역전되는 현상이 나타나기도 한다. 이것은 CVTI 방법은 유사도를 측정할 때, 동일한 셀의 중복 방문을 반영하는데 반하여, $CVTI_{MAXLCSS}$ 방법은 중복 방문의 경우를 무시하기 때문에 발생한다.

이러한 실험 결과를 기반으로, 우리는 CVTI와 $CVTI_{MAXLCSS}$ 방법 간의 상관관계를 분석해 보았다. 그림 7은 CVTI와 $CVTI_{MAXLCSS}$ 방법에 의해 계산되는 유사도 값으로 산포도로 나타낸 것으로, 두 지수들이 거의 일치함을 보이고 있다. 또한, CVTI와 $CVTI_{MAXLCSS}$ 방법 간의 상관계수를 다음과 같이 계산해 보았다.

$$r = \frac{\sum (x_i - \bar{X})(y_i - \bar{Y})}{\sqrt{\sum (x_i - \bar{X})^2 \sum (y_i - \bar{Y})^2}} = 0.9996$$



[그림 7] Comparison CVTI and $CVTI_{MAXLCSS}$ between data trajectories

상관계수의 값이 거의 1에 근접함에 따라, 두 비교지수간의 상관성은 매우 강하다고 말할 수 있다. 따라서, CVTI와 $CVTI_{MAXLCSS}$ 방법이 밀접하게 관련되어 있으므로, 둘 중의 어느 방법이라도 셀룰러 공간상의 궤적들 간의 유사도를 분석하기에 충분하다고 할 수 있다. 그러나, 궤적의 특성을 좀 더 고려하고 있는 $CVTI_{MAXLCSS}$ 방법을 추천한다.

5. 결 론

대부분의 이동객체의 궤적들 간의 유사도 측정 지수들은 유클리디안 공간 혹은 도로 네트워크 공간을 기반으로 하고 있다. 그러나 셀룰러 공간에서의 위치는 좌표로 표현되는 유클리디안 공간과 달리 방 번호와 같은 셀 아이디로 표현된다. 따라서, 기존의 연구에서 제안된 유사도 측정 지수들을 셀룰러 공간에 적용하는 것은 부적합하다. 이에, 본 논문에서는 기본적인 셀룰러 공간에 대하여 정의하고, 셀룰러 공간에 존재하는 이동객체의 속성을 살펴보았다. 그리고 이를 기반으로 셀의 집합 형태로 셀룰러 공간을 나타내었으며, 이동 객체의 궤적은 방문한 셀과 시간 구간의 순서로 나타내었다.

셀룰러 공간의 이동 객체 궤적에 대한 유사도를 측정하기 위하여, 본 논문에서는 문자열간의 유사도를 비교하기 위하여 제안된 최장 공통 부분열 방법을 적용해 보았다. 제시한 두가지 방법들은 최장 공통 부분열 방법에서 고려되지 않았던 시간적 속성을 고려하여 유사도를 측정하는 방법으로, 이동객체 궤적의 공간적 유사성과 시간적 유사성을 모두 고려하는 방법이다. 이러한 지수

들은 셀룰러 공간을 위해 제안된 첫 유사도 측정 지수이다.

본 연구에서 제안한 지수들을 실험을 통하여 분석해 보았다. 실험 결과에 따르면, LCSS 방법은 시간적인 고려를 하지 않으므로 셀룰러 공간에 존재하는 이동 객체의 궤적간의 유사도 비교에 적합하지 않지만, 본 논문에서 제시한 2가지 지수(CVTI, $CVTI_{MAXLCSS}$)는 공간뿐만 아니라, 시간적이 측면도 고려하여 궤적간의 유사도를 비교하므로 셀룰러 공간에 존재하는 이동 객체의 궤적간의 유사도 비교에 사용할 수 있다. 또한, 본 논문에서 제시한 CVTI, $CVTI_{MAXLCSS}$ 방법은 각각 궤적의 다른 특성들을 고려하여 정의하였으나, 거의 유사한 결과를 보인다. 이러한 결과를 기반으로, 셀룰러 공간에 존재하는 이동 객체 궤적의 유사도를 측정하기 위한 방법으로 본 논문에서 제안한 $CVTI_{MAXLCSS}$ 방법을 추천한다.

감사의 글

본 연구는 2단계 두뇌한국21사업과 국토해양부 첨단도시기술개발사업 - 지능형국토정보기술혁신 사업과제의 연구비지원(07국토정보C04)에 의하여 일부 지원되었습니다.

참고문헌

이병길, 2003, “비즈니스 GIS에서 공간 데이터 마이닝(Spatial Data Mining)기법을 이용한 상권추출”, 한국 GIS학회지, 11권 2호, pp.171-184
정재승·박재민·김병국, 2006, “모바일매핑시스템에서의 이동객체 추적을 위한 연구”, 한

국 GIS학회지, 14권 2호, pp.235-244
정지안·장용구·강인준, 2006, “Indoor GPS수신기 정밀도에 관한 연구”, 한국GIS학회 GIS/RS 공동 춘계학술대회, pp.84-89
Bin Lin and Jianwen Su, 2008, One way distance : for shape based similarity search of moving object trajectories, Geoinformatica, 12(2), pp.117-142
Christian Becker and Frank Dürr, 2005, On location models for ubiquitous computing, Personal and Ubiquitous Computing, 9(1), pp.20-31
Daniel S. Hirschberg, 1977, Algorithms for the longest common subsequence problem. Journal of ACM, 24(4), pp.664-675
Eamonn J. Keogh and Michael J. Pazzani, 2000, Scaling up dynamic time warping for datamining applications, Proceedings of the 6th ACM SIGKDD international conference on Knowledge discovery and data mining, pp.285-289
Kin pong Chan and Ada Wai-Chee Fu, 1999, Efficient time series matching by wavelets, Proceedings of the 15th International Conference on Data Engineering, pp.126-133
Rakesh Agrawal, Christos Faloutsos, and Arun N. Swami, 1993, Efficient Similarity Search In Sequence Databases, Proceedings of the 18th International Conference on Foundations of Data Organization and Algorithms, pp.69-84
Lasse Bergroth, Harri Hakonen, and Timo Raita, T., 2000, A survey of longest common subsequence algorithms, Proceedings of the 7th International Symposium on String Processing Information Retrieval, pp.39-48
Ludger Becker, Henrik Blunck, Klaus Hinrichs, and Jan Vahrenhold, 2004, A framework for representing moving objects, Proceedings of the 15th International Conference on Database and Expert Systems Applications, pp.854-863
Michalis Vazirgiannis and Ouri Wolfson, 2001, A spatiotemporal model and language for moving

- objects on road networks. Proceedings of the 7th International Symposium on Spatial and Temporal Databases, pp.20-35
- Nikos Pelekis, Ioannis Kopanakis, Gerasimos Marketos, Irene Ntoutsis, Gennady L. Andrienko and Yannis Theodoridis, 2007, Similarity search in trajectory databases, Proceedings of the 14th International Symposium on Temporal Representation and Reasoning, pp.854-863
- Nirvana Meratnia and Rolf A. de By, 2003, Trajectory representation in location-based services: Problems & solution, Proceedings of the 3th International Workshop on Web and Wireless Geographical Information Systems, pp.18-24
- Ralf Hartmut Güting, Michael H. Bohlen, Martin Erwig, Christian S. Jensen, Nikos A. Lorentzos, Markus Schneider and Michalis Vazirgiannis, 2000, A foundation for representing and quering moving objects, ACM Transation of Database Systems, 25(1), pp.1-42
- Roy Want, Andy Hopper, Veronical Falcao and Honathan Gibbons, 1992, The active badge location system, ACM Transactions on Information Sysmtes, 10(1), pp.91-102
- Yannis Theodoridis, Mario A. Nascimento, 2000, Generating Spatiotemporal Datasets on the WWW, ACM SIGMOD Record, 29(3), pp.39-43
- Yasushi Sakurai, Masatoshi Yoshikawa and Christos Faloutsos, 2005, FTW: Fast similarity search under the time warping distance, Proceedings of the Twenty-fourth ACM SIGACT-SIGMOD -SIGART Symposium on Principles of Database Systems, pp.326-337
-
- 접수일 (2008년 6월 18일)
최종수정일 (2008년 11월 7일)
게재확정일 (2008년 11월 10일)