

지능형 에이전트의 위치기반 서비스를 위한

사용자의 위치이동패턴 학습

한상준^o, 강현지, 조성배

연세대학교 컴퓨터과학과

{sjhan, spechere, sbcho}@sclab.yonsei.ac.kr

Learning User's Moving Patterns for Location-based Services with Intelligent Agent

Sang-Jun Han, Hyun Jee Kang and Sung-Bae Cho

Dept. of Computer Science, Yonsei University

요 약

사용자의 위치정보는 에이전트가 상황에 적합한 서비스를 제공하는데 중요하게 사용될 수 있으며 정확한 위치 추적 및 활용 방안에 대한 활발한 연구가 진행되고 있다. 그중에서 사용자의 다음 위치를 예측하는 것은 사용자에게 필요한 서비스를 명시적인 요청없이 미리 제공하는데 유용하게 쓰일 수 있다. 본 논문에서는 GPS신호를 이용하여 사용자의 위치 이동경로를 학습하고 사용자의 이동에 기반한 서비스 제공 방법을 제안한다. GPS에 의해 관측된 위치 이동경로는 시간 순서의 데이터에 적합하도록 SOM을 변형한 RSOM과 마르코프 모델을 이용하여 학습되며, 새로 관측된 사용자 위치 데이터에 대해 다음 이동 패턴을 예측하는 기능을 가진다. 실제 캠퍼스에서 수집된 데이터를 이용하여 제안한 방법의 가능성을 평가한다.

1. 서론

최근 유비쿼터스 컴퓨팅의 등장으로 저수준의 센서를 이용하여 사용자 상태를 예측하기 위한 사용자 모델링 방법에 관한 연구가 활발하다. 그 중에서도 위치 정보는 사용자 상태를 예측하여 이에 맞는 서비스를 제공하기 위한 중요하기 때문에 이를 활용한 연구가 활발히 진행되고 있다.

위치 정보의 활용에 관한 연구로는 크게 정확한 위치 추적을 위한 방법에 대한 연구와 위치 정보를 이용한 다른 정보를 추론하는 연구가 있다. 위치 추적에 사용되는 방법으로는 GPS(global positioning system) 위성신호, 무선 네트워크 신호, 초음파 신호, 카메라, 압력 센서 등의 다양한 방법이 사용되어 왔으며 이러한 여러 가지 방법을 결합하여 더 정확한 추적을 가능하게 하려는 연구가 진행 중이다[1]. 이렇게 추적된 위치 정보는 사용자의 다음 위치, 현재 사용하고 있는 교통수단 등 다양한 다른 정보의 추론에 사용될 수 있다. 그중에서도 사용자의 다음 위치를 예측하는 것은 사용자에게 필요한 서비스를 명시적인 요청없이 미리 제공하는데 유용하다. 본 논문에서는 사용자의 위치이동 패턴을 학습하여 다음에 움직일 위치를 추론하는 기법과 이를 실제 서비스 제공에 활용하는 방법을 제안한다.

2. 관련연구

사용자 위치이동을 이용한 선행연구로는 다음과 같은 것들이 있다. D. Patterson 등은 이를 위해 GPS(Global Positioning System) 신호로 사용자 위치를 추적하고 베이직인 네트워크로 GPS신호를 모델링하여 현재 사용자가 이용하고 있는 이동수단을 예측하는 방법을 제안하였다[2]. D. Ashbrook 등은 GPS 데이터를 k-means 알고리즘을 사용해 클러스터링하여 사용자가 오래 머무르는 유효한 위치를 찾아내고 이 위치간의 이동을 마르코프 모델을 사용하여 모델링 하여 사용자 위치를 예측하도록 하였다[3]. 하지만 사용자가 머무르는 위치만을 사용하기 때문에 이동 중 경로에 따른 예측은 불가능 하다. 또한 마르코프 모델

은 매우 결정적이기 때문에 예측하는데 있어 유연성이 떨어지는 단점이 있다.

3. 제안하는 방법

사용자의 위치는 GPS 장비를 이용하여 추적하였다. GPS는 실내에서의 위치 추적이 불가능하고 10m 정도의 오차가 생길 수 있는 단점이 있지만 별다른 추가적인 알고리즘 없이 손쉽게 센서만으로 위치를 알아낼 수 있다는 장점이 있다. 사용자의 위치 이동경로는 SOM(self-organizing map)과 마르코프 모델을 이용하여 학습한다. SOM은 사용자의 위치이동 경로를 클러스터링하여 사용자의 주 이동 패턴을 찾아내며 GPS 데이터 학습의 복잡도를 줄이는 역할을 한다. 마르코프 모델은 각 경로 패턴을 학습하고 현재 사용자가 어떤 패턴에 가장 가깝게 이동하는지를 판단한다. 제안하는 방법의 개요는 그림 1과 같다.

3.1 이동패턴 클러스터링

하나의 분류기를 이용해 모든 사용자의 이동패턴을 학습하기는 어렵기 때문에, 본 논문에서는 먼저 유사한 이동패턴을 모아 지역 데이터 셋을 만들고 각 데이터 셋을 학습하여 지역 모델을 만든 후 새로운 데이터를 지역모델로 평가해 사용자의 이동패턴을 예측하도록 하였다. 본 논문에서는 이를 위해 RSOM(recurrent SOM)을 사용하였다.

SOM은 대표적인 비교사학습 신경망으로 클러스터링이나 벡터 양자화 문제에 많이 사용되어 왔다. RSOM은 시퀀스 데이터의 처리에 적합하도록 변형된 것으로 일반적인 SOM이 가지고 있는 특징들을 모두 계승한다[4]. 기본적으로 SOM은 비교사학습을 하기 때문에 사전지식이 없이도 유사한 이동패턴들의 집합을 발견할 수 있어 사용자의 간섭없이 일상 이동 패턴을 클러스터링할 수 있는 장점이 있다. 또한 시퀀스 입력에 따른 최정확 노드의 변화 과정은 이동 패턴의 상태 변화로 볼 수 있으므로 GPS 데이터 대신 이를 학습할 경우 문제의 복잡도를 줄일 수 있다.

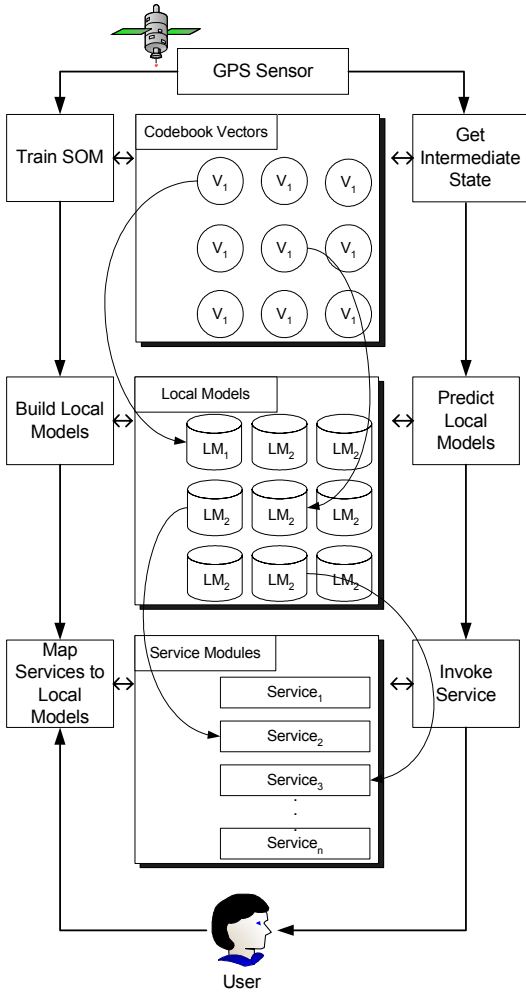


그림 1. 제안하는 방법의 개요

사용자의 이동 중에 수집된 GPS 데이터의 시퀀스들이 RSOM의 입력으로 사용되며 한 개의 시퀀스는 한 번의 이동을 나타낸다. RSOM의 학습 알고리즘은 다음과 같다. $x(n)$ 를 시퀀스의 n 번째 데이터, α 를 현재 데이터에의 가중치, $w_i(n)$ 을 i 번째 노드의 가중치라고 할 때, 시퀀스의 n 번째 데이터에서 출력층의 I 번째 노드의 출력값 $y_i(n)$ 은 다음과 같이 계산된다.

$$y_i(n) = (1 - \alpha)y_i(n-1) + \alpha(x(n) - w_i(n)) \quad (1)$$

이때 바로 전 데이터 값의 출력 노드의 값 $y_i(n-1)$ 도 반영함으로써 데이터사이의 상태정보를 유지하게 한다.

시퀀스의 n 번째 단계에서 입력 시퀀스와 가장 가까운 최정합 노드(best matching node) $b(n)$ 은 다음과 같이 계산된다.

$$b(n) = \arg \min_{i \in V_M} y_i(n) \quad (2)$$

$b(n)$ 이 계산된 후 다음 단계에서의 가중치 $w_i(n+1)$ 은 $\gamma(n)$ 을 학습률, $h(b(n), i)$ 를 이웃반경 함수라고 할 때 다음과 같이 결정된다.

$$w_i(n+1) = w_i(n) + \gamma(n)h(b(n), i)y_i(n) \quad (3)$$

이러한 과정을 거쳐 한 개의 입력 시퀀스의 처리가 끝나면 모든 출력 노드의 값을 0으로 설정한 후 새로운 데이터로 위의 과정을 반복한다. GPS 데이터 시퀀스는 가장 마지막 최정합 노드에 따라 클러스터링 되고 시퀀스 데이터 입력에 따른 최정합 노

드 변화 시퀀스 $B = b(1), b(2), \dots, b(N)$ 의 집합은 지역 데이터 셋으로 사용된다.

3.2 지역 모델 학습

RSOM을 이용해 지역 데이터 셋으로 클러스터링된 사용자 이동 패턴은 마르코프 모델에 의하여 학습된다. 출력 노드수 만큼의 지역 모델이 마르코프 모델을 이용하여 만들어 진다. 마르코프 모델은 초기 상태 분포 Q 와 상태 전이 확률 분포 행렬 P 로 이루어진다. 시간 t 의 상태 i 에서 시간 $t+1$ 의 상태 j 로 전이될 확률을 p_{ij} 라고 할 때 상태 전이 확률 행렬 P 는 다음과 같이 정의 되는데 한 상태에서 다른 상태로 전이될 확률을 모두 합하면 1이 된다.

$$P = \begin{bmatrix} p_{11} & p_{12} & \Lambda & p_{1n} \\ p_{21} & p_{22} & \Lambda & p_{2n} \\ \mathbf{M} & \mathbf{M} & \mathbf{M} & \mathbf{M} \\ p_{n1} & p_{n2} & \Lambda & p_{nn} \end{bmatrix} \quad (4)$$

$$\sum_{j=1}^{j=n} p_{ij} = 1 \quad (5)$$

그리고 시간 0에서 상태 i 가 될 확률을 q_i 라고 할 때 초기 확률 분포 Q 는 다음과 같다.

$$Q = [q_1 \quad q_2 \quad \Lambda \quad q_n] \quad (6)$$

마르코프 체인 모델의 학습은 데이터로부터 상태 전이 확률 분포 행렬과 초기 확률 분포를 얻어내는 과정이다. 시간 0부터 $N-1$ 까지의 상태변화 시퀀스 $X_0, X_1, X_2, \dots, X_{N-1}$ 가 학습 데이터로 주어졌다고 할 때 다음과 같은 식에 의해 P 와 Q 를 얻어 낼 수 있다.

$$p_{ij} = \frac{N_{ij}}{N_i} \quad (7)$$

N_{ij} : 시퀀스중 상태 i 에서 j 로의 전이 회수
 N_i : 시퀀스중 상태 i 가 나타난 회수

$$q_i = \frac{N_i}{N} \quad (8)$$

N : 모든 관찰 시퀀스의 수
 N_j : 시퀀스중 상태 j 가 나타난 회수

주어진 모델로부터 관찰된 상태변화 시퀀스 $X_0, X_1, X_2, \dots, X_T$ 가 나올 확률 $P(X_0, \dots, X_T)$ 는 다음과 같이 얻어진다.

$$P(X_1, \dots, X_T) = q_{x_1} \prod_{t=2}^T P_{x_{t-1}x_t} \quad (9)$$

사용자의 이동에 따른 최정합 노드의 변화과정이 입력으로 사용되므로 마르코프 모델은 RSOM의 출력 노드 수만큼의 상태를 가진다.

3.3 예측 및 서비스 선택

이렇게 구축된 지역모델을 사용하여 사용자 위치 이동을 예측한다. 새로 들어온 데이터는 각 지역모델에 의해 평가되며 그중 특히 높은 확률을 보이는 지역모델이 있을 경우 즉시 선택된다. 주어진 시퀀스 X_0, \dots, X_i 가 지역모델 LM_b 에 의해 생성될 확률을 $P(X_0, \dots, X_i | LM_b)$ 라고 할 때 다음과 같이 자신을 제외한 다른 모델의 평가값의 평균과의 차이가 일정 값 이상이 되는 모델이 있을 경우 해당되는 모델이 선택된다.

$$P(X_0, \dots, X_i | LM_p) = \sum_{st, k \neq p}^N \frac{P(X_0, \dots, X_i | LM_s)}{N-1} > threshold \quad (10)$$

이와 같은 방법으로 사용자가 이동 중에 유사하다고 판단되는 패턴이 선택되는 즉시 적절한 서비스를 제공해 줄 수 있다. 제공될 서비스는 사용자에게 의해 각 지역 모델에 수동으로 지정되며 이는 각 지역모델이 선택되었을 때 실행된다.

4. 실험 및 결과

제안한 방법을 시험하기 위해 실제 GPS 데이터를 모으고 이를 이용해 성능을 평가하여 보았다. 데이터는 실제 대학생의 생활을 모델로 작성된 시나리오를 기반으로 연세대학교 캠퍼스내에서 수집되었다. 수집된 데이터는 총 10가지의 위치이동 경로를 가지며 각 이동 경로별로 2번씩 수집되어 총 20개의 위치이동 데이터로 이루어져 있다. 그림 2는 수집된 데이터를 지도위에 나타낸 모습이다.



그림 2. 지도위에 나타낸 수집된 데이터

사용된 RSOM의 출력층 크기는 7×7 , 초기 학습반경은 4, 학습률은 0.03, 학습 반복회수는 1000번이며 현재 데이터에의 가중치 α 는 0.3이다.

먼저 RSOM의 클러스터링 성능을 평가해 보았다. 그 결과 1, 10번 패턴과 5, 7번 패턴이 같은 집단으로 구분되었고 다른 패턴은 모두 다른 집단으로 구분되었다. 같은 집단으로 구분된 패턴을 분석해본 결과 사용자의 도착지가 같거나 비슷한 곳에 위치에 있었다. 또한 두 개의 노드가 사용된 패턴은 근처의 노드들이 사용됨을 볼 수 있었다.

표 1. 이동패턴의 클러스터링 결과

패턴	최종 최정합 노드	패턴	최종 최정합 노드
1	(4, 0)	6	(0, 0), (0, 1)
2	(2, 6)	7	(6, 0)
3	(0, 4)	8	(5, 3), (4, 3)
4	(0, 6)	9	(5, 6), (3, 3)
5	(6, 0)	10	(4, 0)

그림 3은 위치이동에 따른 최정합 노드의 변화과정을 보여준다. 그림에서 볼 수 있듯이 패턴에 따라 최종 최정합 노드와 시간에 따른 최정합 노드의 변화과정이 다른 것을 알 수 있다. 따라서 SOM의 최정합 노드의 변화는 이동에 따른 상태 변화를 나타내는 것을 확인 할 수 있고 이를 학습하여 문제의 복잡도를

줄일 수 있다.

학습된 지역모델을 가지고 예측 성능을 테스트 해 보았다. 그 결과 사용자가 도착하기 전에 모두 올바른 지역 모델이 선택되어 사용자의 이동패턴을 잘 예측할 수 있음을 확인하였다. 사용자의 평균 이동시간은 5분 25초였으며 평균 예측에 걸리는 시간은 1분 31초였다.

5. 결론

본 논문에서는 RSOM 및 마르코프 모델을 사용하여 사용자 위치 이동 패턴을 학습하여 다음 이동할 위치를 예측하고 그에 맞는 서비스를 제공하는 기법을 제안하였다. 실제 데이터로 실험해본 결과 이동 경로에 따른 패턴이 RSOM에 의해서 구분될 수 있으며 마르코프 모델을 사용하여 구분된 사용자의 이동경로를 학습하고 사용자가 목적지에 도착하기 전에 예측 할 수 있음을 확인하였다.

향후 연구로는 체계적으로 좀 더 많은 데이터를 수집하여 예측 정확도, 예측에 걸리는 시간 등의 관점에서 제안한 방법의 유용성을 검증해볼 것이다. 또한 RSOM의 클러스터링 성능을 높이기 위한 GPS 데이터의 적절한 전처리에 대한 연구가 이루어져야 할 것이다.

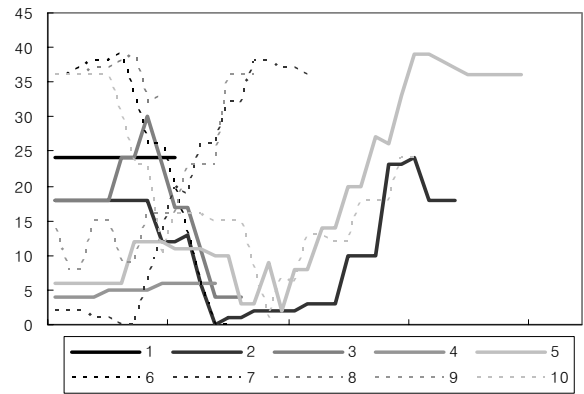


그림 3. 위치 이동에 따른 최정합 노드의 변화

감사의 글

본 연구는 프론티어 연구사업의 지원에 의한 것임

참고 문헌

- [1] J. Hightower and G. Borriello, "Location Systems for Ubiquitous Computing," *IEEE Computer*, vol. 34, no. 8, pp. 57-66, 2001.
- [2] D. Patterson, L. Liao, D. Fox, and H. Kautz, "Inferring High-Level Behavior from Low-Level Sensors," *Proc. of The Fifth International Conference on Ubiquitous Computing*, pp. 73-89, Seattle, WA, October, 2003.
- [3] D. Ashbrook and T. Starner, "Learning Significant Locations and Predicting User Movement with GPS," *Proc. of IEEE Sixth Int. Symp. on Wearable Computing*, Seattle, WA, October 2002.
- [4] T. Koskela, M. Varsta, J. Heikkonen, and K. Kaski, "Temporal Sequence Processing using Recurrent SOM," *Proc. of Second Int. Conf. on Knowledge-Based Intelligent Engineering Systems*, vol 1, pp. 290-297, Adelaide, Australia, April 1998.