

위치데이터 분석을 통한 일일 이동패턴 모델의 도출

김현욱, 송하윤, 최동연, 김동엽
홍익대학교 컴퓨터공학과

email : gizingun@gmail.com, hayoon@wow.hongik.ac.kr, dy3459@naver.com,
hatddum@gmail.com

Human Daily Mobility Patterns Model via Location Analysis of Positioning Information

Hyun-Uk Kim, Ha-Yoon Song, Dong-Yeon Choi, Dong-yup Kim
Dept of Computer Engineering, Hong-Ik University

요 약

본 논문에서는 인간의 일일 이동패턴을 모델링하기 위해 사람의 이동정보인 위치데이터를 바탕으로 위치분석(Location Analysis)을 통해 사람의 이동패턴이 날마다 어떤 형태로 나타나고 반복되는지 보이려고 한다. 이에 사람의 이동패턴은 자주 방문하거나 특정시간이상 머문 공간간의 이동이라고 정의하고, 해당 공간을 하나의 군집으로 하는 군집간의 이동 모습으로 인간의 이동 모습을 나타내고자 한다. 위치데이터를 일일 기반으로 위치분석을 하게 되면 일일 이동모습을 나타낼 수 있고, 이러한 일일 이동모습을 통합하여 분석하게 되면 사람의 전체 이동모습을 나타낼 수 있다. 이렇게 분석된 일일 이동모습과 전체 이동모습을 시간대별로 다시 분석하게 되면 전체 이동모습에 대해 일일 이동모습이 어떠한 형태로 중첩되는지 그 패턴을 찾아볼 수가 있다. 이와 같은 방식으로 우리는 위치데이터에서 일일 이동모습, 전체 이동모습, 그리고 시간대별 이동패턴을 찾아낼 수 있었다.

1. 서론

과학·산업이 발전함과 같이 인간의 이동패턴 모델에 관한 연구가 활발히 이루어지고 있다. 이러한 흐름과 같이 휴대용 위치정보 수집 장비나, 스마트폰 같은 장비가 대중화됨에 따라 사람들의 이동정보인 위치데이터 수집이 용이해지고 있다. 이러한 위치정보 수집의 용이성은 사람의 이동패턴에 관해 좀 더 높은 예측을 가능하게 만들었다.

위치정보 수집은 휴대용 위치수집 장비나 스마트폰과 같은 장비내의 GPS 수신기나 모바일 장비의 기지국 위치, 또는 Wifi 위치확인 등의 방식으로 이루어 질 수 있다. 수집된 위치정보는 <위도, 경도, 시간>의 형태로 나타낼 수 있고, 본 논문에서는 이렇게 수집된 위치데이터를 바탕으로 위치분석(Location Analysis)을 통해 일일 이동모습, 전체 이동모습, 그리고 시간대별 이동패턴을 찾아 나타내 보도록 하겠다.

2. 관련 연구

실제적인 인간의 이동모습은 심리학, 물리학, 도시공학, 컴퓨터 공학 등의 다양한 분야에서 연구되고 있다. 심리학에서는 인간의 이동모습을 심리적인 관점으로 연구를 진행하였다. 즉, 사람이 이동하는 이유와 이동 수단의 선택이 사람의 심리적 요소에 의해 결정된다고 얘기하고 있다 [1].

통계·물리학 측면에서는 인간의 이동을 실제 세계의

현상으로 바라보고 그 현상을 설명하려 하였다. 예를 들면, 지폐(Bank Note)의 이동을 가지고 인간의 이동모습을 설명하려 하였고 [2], 휴대폰의 기지국 정보를 바탕으로 인간의 이동이 중심부로부터 Power law 법칙을 따름을 보였고 이러한 방식으로 인간의 이동을 93%까지 예측 가능하다고 말하고 있다 [3].

컴퓨터 과학 측면에서는 무선 이동네트워크, 인공지능, 로봇, 무인자동차 등의 측면에서 실제적인 인간의 이동 모습을 요구하고 있다. 구체적으로 네트워크 분야에서 MANET(Mobile Ad Hoc Network)의 성능에 영향을 미치는 인간의 이동성이 가지는 통계학적 특성과 MANET 에서의 Protocol 에 대한 성능평가가 활발히 진행되었는데 그 중에서도 노드(Node) 간 통신이 연결되어 있는 시간인 CT(Contact Time), 노드 사이의 링크가 유지되어 있는 시간인 Link Duration, 그리고 Routing Path 의 Lifetime 에 대해 서로 다른 이동성 모델이 어떤 영향을 미치는지에 대한 연구가 진행되었다 [4], [5].

도시 공학분야 에서 비슷한 연구로는 운전자들을 통해 수집된 위치데이터들의 궤적(Trajectory data)을 데이터 마이닝(Data Mining)을 통해 운전자들이 주로 어느 경로로 어떤 시간대에 밀집되어 이동하는지, 주로 어느 시간대에 이동의 빈도가 높아지는 지 등에 대한 분석을 한 연구도 있다 [6].

3. 위치분석(Location Analysis)

우리는 사람의 이동정보인 위치정보를 분석하기 위해 확률 기반 군집화 기법인 기대치 최대화 기반 군집화(Expectation Maximization Clustering)를 응용하여 분석해보았다.

EM 군집화를 통해 위치데이터를 분석하기 위해서는 위치정보의 특성을 잘 고려해야 한다. 일반적인 EM 군집화는 특정 데이터들이 가우시안 분포(Gaussian Distribution)를 따르는 초기모형을 바탕으로 반복-정제 과정을 통해 최적의 모델을 찾아나가게 된다. 그러나 위치데이터 분석에 가우시안 분포를 사용하였을 때, 원하던 결과를 얻을 수 없었다. 이에 본 연구진은 인간의 이동은 특정 중심지역으로부터 1~2Km 내에서 발생하고 그 이상거리로의 이동은 Power Law 를 따른다는 [3]. 연구사실에 근거해 위치정보 분석에 알맞은 새로운 확률밀도 함수를 제안하였다.

$$P(x)=e^{-\lambda x} \quad (1)$$

Λ : 해당 군집의 최대 거리와 평균 거리의 합의 역
 x : 해당 군집의 중심과 위치데이터 사이의 거리

그 다음 고려해야 될 점은 바로 위치정보의 상태(state)다. 위치데이터 분석을 통해 인간의 이동을 하나 이상의 군집간 이동모습으로 나타내기 위해서는 위치데이터의 상태가 고려되어야 한다. 위치데이터의 상태는 크게 머물러있는(stay) 상태와 이동중인(moving) 상태로 나눌 수 있다. 이때, 머물러있는(stay) 상태의 위치데이터가 많이 밀집되어 있는 곳이 특정시간 이상 머물러 있던 장소이거나 한번 이상 방문이 이루어진 장소라고 판단할 수 있다. 위와 같은 위치데이터의 상태를 판단하기 위해서는 시간대별 위치정보의 위치, 즉, 위치데이터의 속도를 기준으로 그 상태를 판단할 수 있다. 기준 속도 이하의 위치데이터는 머물러 있는 상태라고 판단하고, 기준 속도 이상의 위치데이터는 이동중인 위치데이터라고 판단하였다. 위와 같은 기준은 확률기반 군집화를 사용할 때, 머물러 있는 상태에 대한 위치데이터에는 이동중인 상태의 위치데이터보다 좀 더 높은 가중치를 주어 분석함으로써 결과의 신뢰성을 높일 수 있었다.

4. 이동모델 형성 과정

4.1 일일 데이터 분석

인간의 이동패턴의 분석을 위한 기대치 최대화 기반 군집화는 각 일일 위치데이터에 대해 다음의 4 단계 과정을 통해 실시된다.

1) 초기화(Initialization)

위치데이터를 초기화하는 과정이다. 해당 일의 위치데이터가 가지는 속도와 점 사상의 밀도를 고려하여 초기 군집의 개수 및 초기 군집의 위치 및 크기를 결정하게 된다.

2) 기대화 과정(Expectation, E step)

초기화된 군집들의 속성 값(군집 내 위치데이터 주 군집의 중심으로부터 떨어진 거리의 평균 등)을

바탕으로 각각의 위치데이터가 특정 군집에 속할 확률 및 가중치(weight or probability)를 계산한다. 이때 사용되는 확률 밀도 함수는 수식 (1)의 확률밀도 함수를 사용한다.

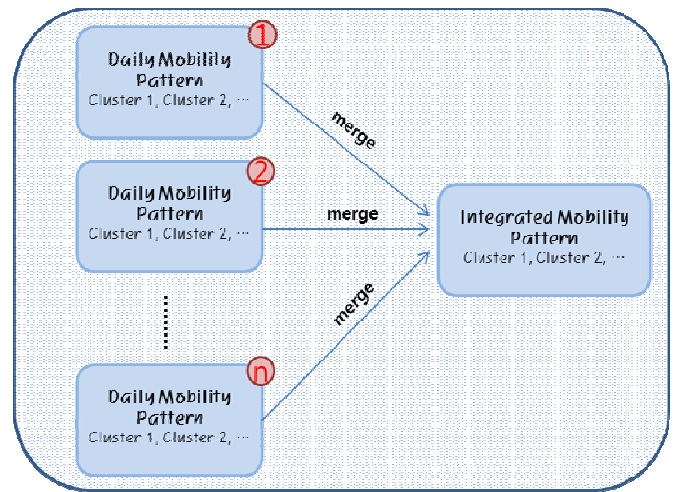
3) 최대화 과정(Maximization, M step)

기대화 과정(E step) 이후 계산된 각각의 위치데이터가 특정 군집에 속할 확률을 가지고 군집별로 최대 우도(likelihood) 값을 가지는 매개변수들을 재 계산한다. (평균, 분산 등)

4) 종료(Termination)

각각의 군집이 가지는 매개변수의 값 차이가 특정 임계 값 이하로 변하지 않을 때까지 2, 3 단계를 반복한다. 즉, 최대 우도(likelihood)값이 수렴할 때까지 위의 과정을 반복한다.

위의 4 단계를 마치게 되면 사람의 일일 이동패턴이라 할 수 있는 군집들이 공간상에 자리잡게 되고 각 군집간의 전이횟수, 각 군집의 속성값(중심점, 평균, 머문 시간 등)을 구할 수 있게 된다. 이렇게 계산된 속성 값 중 머문 시간과 각각의 군집 간의 전이(Transition) 횟수를 군집간의 이동을 상태간의 이동이라고 생각하여 연속시간 마르코프 연쇄(Continuous Time Markov Chain)의 형태로 나타낼 수 있다.



<그림 1> 일일 모델의 통합과정

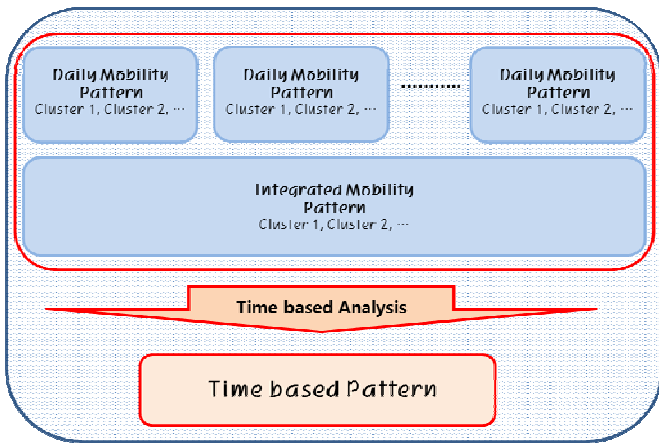
4.2 일일 이동모델의 통합

3.1의 과정을 통해 위치데이터는 일일 이동모델로 분석되었다. 이렇게 분석된 각 일일 이동모델은 통합되어 전체 모델로 나타낼 수 있다. <그림 1>은 각 일일 이동모델의 통합과정의 전체 모습을 보여주는 그림이다. <그림 1>과 같이 일일 이동모델이 통합되기 위해서는 일일 모델의 병합(Merge)과정이 필요하다. 일일 이동모델의 병합은 다음의 순서로 이루어진다.

- i. 통합 이동모델의 틀을 생성한다.(아직 그 어떤 일일 모델도 통합되지는 않았다)
- ii. 가장 처음 시간대의 일일 모델을 통합 모델에 병합한다. 이 경우 통합 이동모델은 틀이 없기 때문에 처음의 일일 모델이 통합모델

- iii. 그 자체가 된다)
 그 다음 모델의 이동모델을 병합한다. 이때, 앞서 저장된 이동모델의 군집들과 비교하여 중첩이 없는 경우에는 통합 이동모델에 새로운 군집을 생성시켜 주고, 중첩이 있는 경우 해당 군집의 속성값을 재계산하여 반영한다.

위의 병합과정 중 서로 다른 군집간의 중첩이 이루어지는 경우는 다음의 경우로 정의했다. 먼저 통합된 군집의 중심점과 영역(군집내의 최대거리 + 군집 내의 평균거리)안에 새로 병합하고자 하는 군집의 중심점이 포함되는 경우 중첩이 이루어졌다고 판단한다. 이런 과정을 통해 전체 일일 이동모델을 하나의 통합 모델로 병합하여 분석할 수 있다.



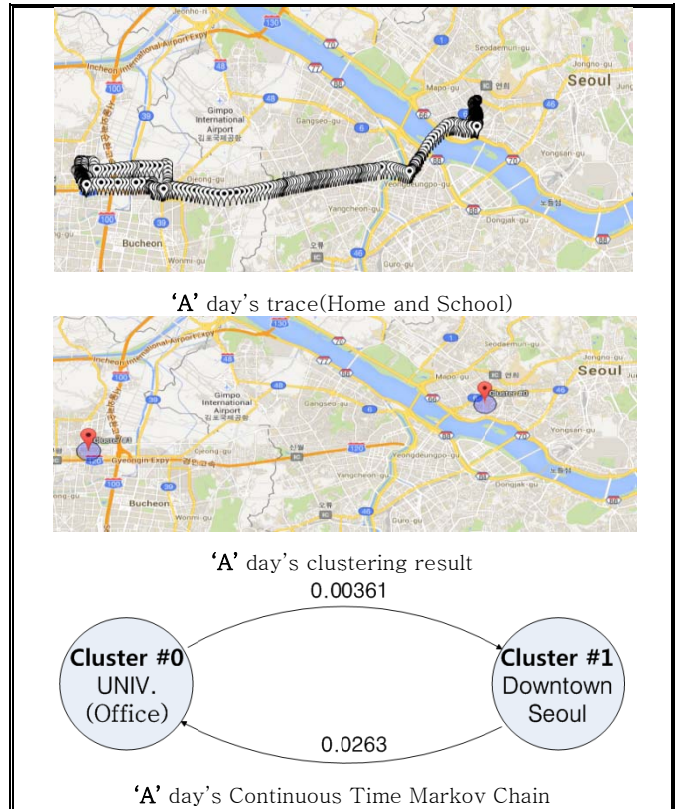
<그림 2> 시간대별 이동패턴

4.3 일일 이동패턴의 도출

3.1 과 3.2 의 과정을 통해 일일 이동모델과 전체 통합모델을 도출해낼 수 있었다. 위의 과정을 통해 통합된 이동모델에서 우리는 TimeRatio 값을 비교하여 주요 군집을 포착할 수 있다. TimeRatio 값은 전체 위치데이터가 수집된 시간 분에 각 군집에서 머물렀던 시간의 비율이다. TimeRatio 의 값이 높은 군집이 가지는 의미는 전체 위치데이터가 수집되는 기간 중 가장 오래 머물렀다는 의미가 되고 이는 보통 집 지역의 군집이거나 직장 지역 근처의 군집임을 알 수 있었다. 예를 들어, 보통 집과 직장만을 반복적으로 이동하고 그 밖의 장소로의 이동은 드문 사람이 있다고 하자. 이런 사람의 위치데이터가 분석되면 주요 군집은 집과 직장 지역이 될 것이고, 그 밖의 장소에 생성된 군집은 TimeRatio 가 낮은 값을 갖게 될 것이다. 이처럼 통합 이동모델에서 구분한 주요 군집과 비 주요군집은 다시 일일 이동모델과 시간대별 분석을 통해 일일 이동패턴으로 나타낼 수 있다. 통합 이동모델에서의 주요군집이 각 일일 이동모델에서 어느 시간대에 분포하는지를 모아 분석하고, 비 주요 군집으로의 이동이 어느 시간대에 분포해 있는지 종합하여 분석하게 되면, 하루마다 사람의 이동이 시간대별로 어떠한 패턴을 가지는지 알 수 있다. <그림 2>는 일일 이동패턴의 도출 방법에 대한 전반적인 모습을 그림으로 나타낸 것이다.

4. 결과

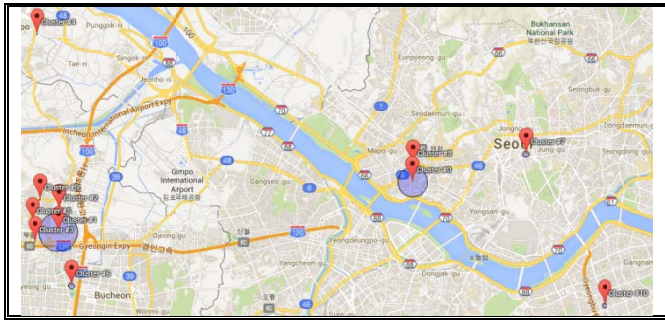
약 한 달 동안 수집한 KH 의 전체 위치데이터는 위치분석을 통해 일일 이동모델과 통합 이동모델로 나타낼 수 있었다. <그림 3>은 그 중 A 일의 위치데이터에 대해 A 일의 이동패턴, 위치분석 결과, 분석 후 생성된 이동모델을 연속시간 마르코프 연쇄의 형태로 나타낸 그림이다.



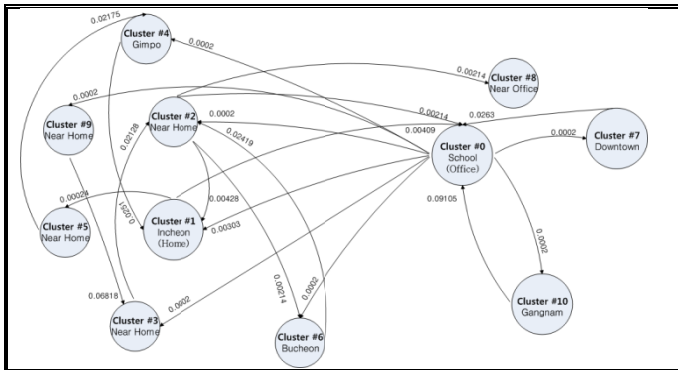
<그림 3> A 일의 일일 이동모델 분석 과정 및 결과

이러한 방법으로 전체 위치데이터에 대해 위치분석을 실시하면 날짜마다 특정시간 이상 머물렀거나 한번 이상 방문이 나타난 장소를 중심으로 위치데이터 군집이 형성된다. 이렇게 생성된 군집은 군집의 중심 위치, 중심 위치의 표준 편차, 군집 내 위치데이터가 가지는 최대 거리, 중심으로부터 각 위치데이터가 떨어진 거리의 평균, 머무른 시간, 머무른 시간의 비율 (TimeRatio), 해당 군집 내에서 머문 군집의 위치데이터의 수와 같은 속성 값들을 가지게 된다. 이렇게 생성된 일일 이동모델은 병합과정을 통해 통합 이동 모델로 나타낼 수 있다.

<그림 4>은 각 일일 이동모델을 병합 후 나타난 통합 모델을 지도상에 나타낸 그림이다. 위치분석 후 결과값 중 TimeRatio 를 비교했을 때, 주요 군집은 학교와 집 2 지역이 생성되었다. 통합 이동모델에서는 총 11 개의 군집이 나타났고 TimeRatio 값 비교와 통합 이동모델의 Transition Matrix 를 살폈을 때, 전체 이동모습은 집 지역과 학교 지역을 중심으로 이루어졌다. <그림 5>은 통합 이동모델을 연속시간 마르코프 연쇄의 형태로 나타낸 그림이다.



<그림 4> 통합 이동모델 결과

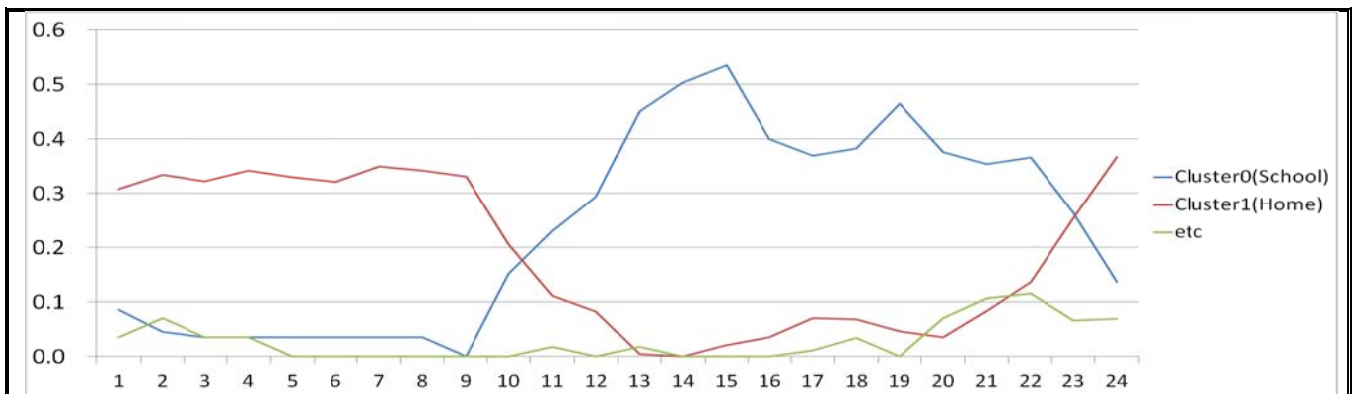


<그림 5> 통합 이동모델의 연속시간 마르코프 연쇄

KHU의 통합 이동모델에서 주요 두 군집(집, 학교)와 각 일일 이동모델에서 생성된 군집과 시간대별 분석을 통해 시간대별 이동패턴을 찾아 낼 수 있었다. KHU의 경우 새벽부터 오전 시간 동안 집 지역에서 주로 머무르고 있는 경향을 보였고 그 이후 시간을 보통 학교에서 머무르는 경향을 보였다. 종합 하자면 KHU의 경우 일일 이동패턴은 0~1 시에는 주로 집에서 머무르고 그 이후 시간은 주로 학교에서 머물러 있음을 도출해 낼 수 있었고 주요 군집이 아닌 다른 장소로의 이동은 주로 오후에 있음을 알 수 있었다. <그림 6>은 KHU의 일일 이동패턴 분석 결과를 그래프로 나타낸 그림이다. 빨간색 선은 0번 군집(학교 지역)의 시간대별 분포 확률을 나타내고 파란색 선은 1번 군집(집)의 시간대별 분포 확률을 나타낸다.

5. 결론

본 논문에서는 인간의 이동정보인 위치정보를 위치



<그림 6> 시간대별 일일 이동패턴

분석을 통하여 일일 이동모델과 통합 이동모델, 그리고 시간대별 이동패턴을 나타낼 수 있었다. 이렇게 생성된 이동모델은 시각적인 결과물을 제공할 수 있으며, 연속시간 마르코프 연쇄의 형태로도 나타낼 수 있어 다른 응용분야에서도 사용될 수 있다. 또한 통합 이동모델과 일일 이동모델의 시간대별 분석을 통해 시간대별 일일 이동패턴도 나타낼 수 있었다. 또한 이처럼 위치분석을 통해 얻은 일일 이동모델은 각각을 주간, 요일간, 월간, 연간 등의 주기적 병합을 통해 각 주기 단위로의 이동모델을 추출할 수 있는 토대를 만들었다는 점에 그 의미가 있다.

6. 사사

이 논문은 2013년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2012R1A2A2A03046473).

참고문헌

[1] R. W. Byrne, Memory for urban geography, Quarterly journal of experimental psychology 31, pp.147-154, 1979

[2] B. Verplanken, K. Aarts & A. V. Knippenberg, Habit, information acquisition, and the process of making travel mode choice, European journal of social psychology, vol.27, issue. 5, pp.539-560, 1997

[3] C. Song, Q. zehui, B. Nicholas, B. Albert-laszio: Limits of Predictability in Human Mobility, Science 19, Vol. 327 No. 5968 pp. 1018-1021, 2010

[4] F. Bai, N. Sadagopam, A. Helmy, Important: A framework to systematically analyze the Impact of Mobility on Performance of Routing Protocol for Ad hoc Networks, Twenty-second Annual Joint Conference Of The IEEE Computer And Communications Societies, vol.2, pp.825-835, 2003

[5] B. Zhou, K. Xu, M. Gerla, Group and swarm mobility models for ad hoc network scenarios using virtual tracks, IEEE Military communications Conference vol.1, pp.289-294, 2004

[6] F. Giannotti, M. Nanni, D. Pedreschi, F. Pinelli, C. Renso, S. Rinzivillo, R. Trasarti, Unveiling the complexity of human mobility by querying and mining massive trajectory data, The VLDB Journal, vol.20, issue.5, pp.695-719, 2011