

멀티모달 데이터 기반 위험 발생 유사성 비교 방법

권은정¹ · 신원재¹ · 이용태¹ · 이규철²

¹한국전자통신연구원 · ²충남대학교

A Method of Comparing Risk Similarities Based on Multimodal Data

Eun-Jung Kwon¹ · WonJae Shin¹ · Yong-Tae Lee¹ · Kyu-Chul Lee^{2*}

¹Electronic and Telecommunications Research Institute (ETRI) · ²Chungnam National University

E-mail : ejkwon@etri.re.kr, thunder9001@etri.re.kr, ytleee@etri.re.kr / kcleee@cnu.ac.kr

요 약

최근 공공안전 분야에서 위험상황을 감지하거나 선제적인 예측을 통해 안전을 보장하기 위한 요구사항이 대두되고 있다. 모바일 단말의 보급화로 인해 다양한 센서 데이터를 융합하여 분석할 경우 데이터의 잠재적 가치를 안전보장 측면에서 활용할 수 있다는 점에서 주목받고 있다. 스마트폰, 웨어러블 기기 등에서 발생하는 센서 데이터를 결합하여 사용자의 이동 경로 패턴, 행동 패턴을 분석하고, 사용자의 현재 위치에서 제공하는 위치 기반 범죄 위험 데이터를 융합하여 공공안전을 보장할 수 있는 효과적인 모델링 기법이 필요하다. 본 논문은 사용자의 과거 이동 궤적의 패턴과 행동 패턴을 분석하고, 사용자 별 이동 궤적들이 동일 공간 내에 다른 사용자와의 유사도를 분석하여 안전한 경로를 추천하게 된다. 본 논문은 위치기반 멀티모달 센서 데이터를 결합하여 사용자의 안전을 보장하기 위한 위험 예측 방법을 제안한다.

ABSTRACT

Recently, there have been growing requirements in the public safety sector to ensure safety through detection of hazardous situations or preemptive predictions. It is noteworthy that various sensor data can be analyzed and utilized as a result of mobile device's dissemination, and many advantages can be used in terms of safety and security. An effective modeling technique is needed to combine sensor data generated by smart-phones and wearable devices to analyze users' moving patterns and behavioral patterns, and to ensure public safety by fusing location-based crime risk data provided.

키워드

Public Safety, Prediction Model, Risk Prediction, Multi-Modal Data

I. 서 론

최근 국민의 일상을 위협하는 사회적 문제들이 지속적으로 발생하면서 이를 해결하여 삶의 질에 대한 수준을 높이기 위해 ICT의 사회적 역할이 강조 되고 있다. 특히, AI, 빅데이터 등 첨단 ICT 기술을 이용하여 공공정보의 체계적 수집·분석 등을 통해 다양한 사회문제의 원인을 파악하고 해결 방안을 제공하는데 있어 사회문제해결에 큰 역할을

기대하고 있다[1]. 이러한 추세에 따라 모바일 단말을 통해 생성되는 사용자의 다양한 데이터와 공공안전과 관련된 위치별 범죄 발생 정보를 융합하여 위험도에 따른 사용자 간 유사성을 판단할 수 있는 기법을 제시 한다. 이러한 방법은 공공안전의 응용 서비스로 활용될 수 있을 것으로 기대한다.

II. 관련연구

대용량의 데이터로부터 유용한 지식을 찾아내는

* corresponding author

데이터마이닝 기법은 대용량데이터로부터 데이터를 선별하는 과정(data selection), 선별된 데이터를 가공 및 정제하는 과정(cleaning), 분석의 용이함을 위해 데이터를 변환(transformation)하는 과정, 변환된 데이터에 적절한 알고리즘을 적용하는 과정(data mining), 알고리즘의 결과에 따라 실제 데이터를 적용하여 통합(integration)하는 과정을 거치게 된다 [2].

일정한 공간 안에서 사람들의 이동 경로를 분석하여 위치 기반 서비스를 제공하기 위한 기존연구는 빈발 패턴 마이닝이나 순차적인 패턴 마이닝 기법을 주로 이용하였다. 빈발 패턴 마이닝의 경우, 트랜잭션 내에서 빈발하게 발생하는 동일한 항목집합의 패턴을 추출하기 때문에 트랜잭션 내에서 항목집합의 존재여부에 관심을 갖는다. 따라서 사용자가 이동하면서 시간적인 순서에 따라 발생하는 다양한 패턴을 고려할 수 없다. 반면, 연관 규칙에 시간(time)의 개념을 첨가한 순차 패턴 마이닝은 시퀀스인 빈도가 많은 최대시퀀스(maximal sequence)를 효과적으로 처리하는 방안의 중점을 두고, 잠재적인 후보 시퀀스를 찾은 후 시퀀스 간의 유사도 계산을 통해 다양한 서비스에 활용하고 있다[2]-[4].

이동 경로 분석을 위해 GPS 데이터를 기본적으로 궤적 데이터로 활용하지만, 좀 더 정확한 사용자의 위치에 따른 상황을 이해하기 위해 물리적인 지역을 특정 셀(cell)로 나누고 사용자가 특정 셀에서 체류한 시간에 따라 의미를 부여하여 시퀀스를 생성하는 기법이 제시되기도 했다[3][4].

그 밖에 물리적인 지역을 특정 셀로 나누고, 두 패턴간 유사도를 계산하기 위해 최대시퀀스를 찾는 방법에서 시간이 많이 소요되는 단점을 해결하기 위해 동적으로 두 패턴간의 유사도를 계산하는 알고리즘을 제시하기도 하였다. 패턴간 유사도를 계산하고, 실제적으로 각 사용자 간의 유사도를 비교하기 위해서 지지도(support)를 적용하였다[3].

III. 이동경로기반 사용자 위험 유사성 비교 모델링

본 장에서는 보행자 위험 예측 모델링을 위해 지역별 범죄 통계 데이터, 사용자의 스마트 기기(스마트폰, 웨어러블 기기)로부터 생성되는 데이터(GPS, 가속도, 자이로, 심박 데이터)를 융합하여 보행자의 데이터 패턴을 생성하고, 이를 사용자 간 비교하여 유사한 사용자를 추출하고자 한다.

먼저 위치정보를 특정 셀로 구분하기 위해 사용자가 이동한 경로 데이터를 지역별 범죄 통계 데이터에서 기준으로 삼고 있는 지역 정보와 매칭하는 과정을 거치게 된다. 지역별 범죄 통계 데이터는 사용자가 위치한 곳과 근접한 치안 안전 블록으로 매칭하는 과정을 거친다. 이 과정에서 사용자

의 이동 경로에 따라 매칭되는 치안 안전 블록 정보가 시퀀스를 구성하는 요소가 된다. 또한, 사용자가 이동하면서 발생된 사용자 행동(뛰기, 걷기, 멈춤, 심박 이상)이 분석되어 위치 정보와 함께 시퀀스를 구성하는 요소로 생성된다.

그림1과 같이 사용자가 장소를 이동함에 따라 스마트폰 탑재 가속도 센서 데이터 및 GPS 데이터, 웨어러블 기기의 심박 데이터가 수집된다. 이후 1단계, GPS 데이터를 활용하여 사용자가 체류한 장소가 의미기반 경로 정보로 변환 및 추출되는 과정을 거치게 된다. 즉 수치적인 GPS의 X, Y 좌표 정보가 실제 사용자가 체류한 장소와 결합하여 사용자 별 의미있는(semantic) 있는 경로에 대한 패턴이 구축된다.

2단계는 사용자별로 구축된 이동 경로에 대한 패턴들은 사용자 간 유사성을 비교를 통해 위치기반 위험 상황을 선제적으로 회피할 수 있는 정보를 제공받을 수 있도록 한다.

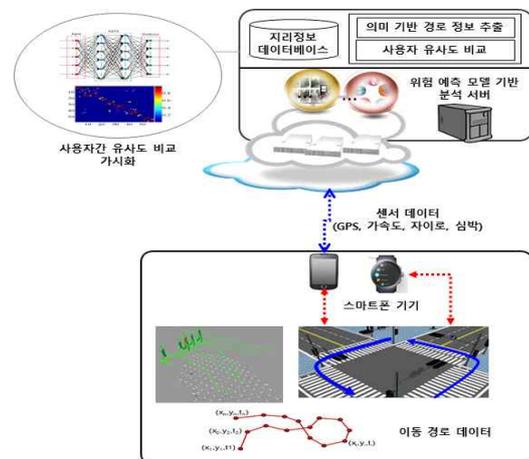


그림 1. 멀티모달 데이터 기반 위험 유사성 비교를 위한 프레임워크

1단계 및 2단계에 대한 세부적인 처리 절차는 그림2와 같다. 셀 단위의 의미기반 경로 정보를 시간의 흐름에 따라 장소 정보로 변환하는 과정은 셀 단위에 대한 설정, 사용자가 셀에서 체류하는 시간에 대한 설정이 전제되어야 한다. 본 논문에서 셀 단위는 과거 범죄사건 발생건과 관련하여 치안 블록으로 매핑 정보를 구성한다. 실제 치안 블록별로 과거 발생한 범죄사건 시간, 피해자 성별, 사건 발생 시간에 대한 정보 및 해당 블록의 위험도와 함께 매핑 정보로 지리정보로 관리된다.

이동 경로의 패턴에 대한 유사성 검출 방식은 이동경로 1(P)과 이동 경로 2(Q)에 대한 공통부분(셀에 포함된 특정 장소)이 많을수록 유사도가 높으므로, 식 1 LCS (Longest Common Sequence) 준용하여 유사도를 비교한다.

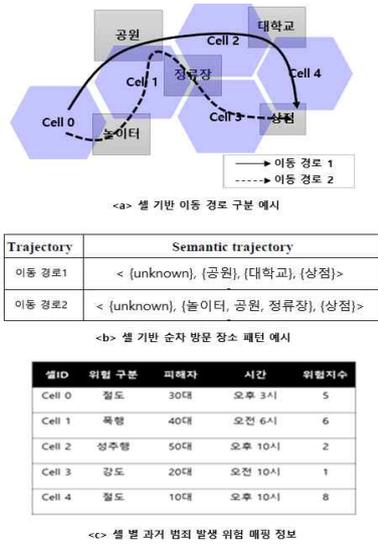


그림 2. 경로 기반 순차 패턴 데이터 세트 생성 예시

식 1을 기준으로, 각각 두 개의 경로에 대한 패턴의 유사성을 비교하기 위해 경로 1(T1)이 T1=<{공원}, {대학교}, {상점}>, 경로2(T2)가 T2=<{놀이터}, {공원}, {정류장}, {상점}>이라고 할 때, T1과 T2의 LCS는 LCS(T1,T2)=<{공원}, {상점}>으로 정의된다. 따라서, T2에 대한 T1과의 공통된 패턴의 비율은 식 1과 같다. 이를 기반으로 각 패턴이 공통된 패턴의 일정 부분을 포함한 T1, T2의 유사성도는 식 2를 준용하여 계산된다.

$$ratio(LCS(T1, T2), T1) = \frac{\sum_{i=1}^{|T1|} \sum_{j=1}^{|LCS(T1, T2)|} M(T1_i, LCS_j)}{|T1|}$$

$$where M(T1_i, LCS_j) = \begin{cases} \frac{|T1_i \cap LCS_j|}{|T1|} & (1) \\ 0 & \end{cases}$$

$$Patterns - Similarity(T1, T2) =$$

$$\frac{|T1|ratio(LCS(T1, T2), T1) + |T2|ratio(LCS(T1, T2), T2)}{|T1| + |T2|} \quad (2)$$

식 2를 이용하여, 위치 기반 과거 범죄사건 발생을 고려한 두 사용자간의 위험 유사도는 식3과 같다. 즉, 두 사용자간의 유사한 경로로 구성된 데이터 셋을 구성하고, 사용자가 방문한 지점의 셀(Cell)이 과거 범죄 발생 지역인 경우 그림2의 위험지수(R)를 이동 경로의 패턴이 차지하는 빈발성의 정도를 의미하는 지도도(S)의 가중치 값(W)으로 사

용한다. 식2의 결과 값이 클수록 사용자 유사도가 높기 때문에 관심 패턴으로 간주된다. 즉, 사용자(U1, U2)가 방문한 관심 패턴의 집합을 D1, D2이라고 할 때, D1={M1,M2,...,MN}, D2={N1,N2,...,NN}인 경우 아래 식4과 같이 위치기반 위험도를 고려한 사용자 간의 유사성을 계산할 수 있다.

$$Similarity(D1, D2) = \frac{\sum_{i=1}^m \sum_{j=1}^n W(S_i \times R_i) + (S_j \times R_j) + Patterns - Similarity(M_i, N_j)}{\sum_{i=1}^m \sum_{j=1}^n W(S_i \times R_i) + (S_j \times R_j)} \quad (3)$$

IV. 결론

본 논문은 사용자의 과거 이동 궤적에 따른 패턴을 분석하고, 사용자 간 관심 이동 궤적들 간 동일 다른 사용자와의 유사도를 분석할 수 있는 방법을 제안하였다. 제시된 방안은 위치 기반 범죄 데이터와 융합하여 위험지역에서의 사용자 별 관심 패턴의 비교 및 사용자의 위험 예측 모델링의 알고리즘으로 활용할 예정이다.

Acknowledgement

본 연구는 2017년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (2017-0-00336, 다중로그 기반 멀티모달 데이터융합 분석 및 상황 대응 플랫폼 기술 개발)

References

- [1] 조혜지, “사회문제 해결형 R&D 최신 동향 및 시사점,” 정보통신기술진흥센터, ICT Spot Issue, 04. 2018
- [2] Z. Chen, “Mining Individual Behavior Pattern Based on Significant Locations and Spatial Trajectories,” IBM PhD Forum on Pervasive Computing and Communications, 2012, Lugano, March 2012
- [3] J. J.-C. Ying, E. H.-C. Lu, W.-C. Lee, T.-C. Weng, V. S. Tseng, “Mining User Similarity from Semantic Trajectories,” In Proceedings of ACM SIGSPATIAL International Workshop on Location Based Social Networks (LBSN' 10), San Jose, California, USA, November 2010
- [4] J. J.-C. Ying, W.-C. Lee, T.-C. Weng, “Semantic Trajectory Mining for Location Prediction,” In Proceedings of ACM SIGSPATIAL GIS '11Chicago, IL, USA, November 2011