



사용자 유사도 기반 경로 예측 기법

남수민*, 이석훈**

User Similarity-based Path Prediction Method

Sumin Nam*, Sukhoon Lee**

이 논문은 2019년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임.
(NRF-2017R1C1B2012065).

요 약

라이프로그를 이용한 경로 예측 기법은 정확한 경로 예측을 위하여 많은 양의 학습 데이터를 요구하며, 학습 데이터가 부족할 경우 경로 예측 성능이 저하된다. 학습 데이터 부족은 사용자의 이동 패턴이 유사한 다른 사용자의 데이터를 이용하여 해결이 가능하다. 따라서 이 논문은 사용자 유사도 기반 경로 예측 알고리즘을 제안한다. 이를 위하여 제안 알고리즘은 경로를 3단 그리드 패턴으로 학습하고 코사인 유사도 기법을 이용하여 사용자 간 유사도를 측정한다. 이후, 측정된 유사도를 학습된 모델에 적용하여 경로를 예측한다. 평가를 위하여 기존 경로 예측 기법들과 제안 기법의 경로 예측 정확도를 측정 및 비교한다. 그 결과, 제안 기법의 정확도는 66.6%로 다른 기법들에 비해 평균 1.8% 더 높은 정확도를 가진 것으로 평가된다.

Abstract

A path prediction method using lifelog requires a large amount of training data for accurate path prediction, and the path prediction performance is degraded when the training data is insufficient. The lack of training data can be solved using data of other users having similar user movement patterns. Therefore, this paper proposes a path prediction algorithm based on user similarity. The proposed algorithm learns the path in a triple grid pattern and measures the similarity between users using the cosine similarity technique. Then, it predicts the path with applying measured similarity to the learned model. For the evaluation, we measure and compare the path prediction accuracy of proposed method with the existing algorithms. As a result, the proposed method has 66.6% accuracy, and it is evaluated that its accuracy is 1.8% higher than other methods.

Keywords

path prediction, user similarity, lifelog, triple grid pattern

* 군산대학교 소프트웨어융합공학과 학부생
- ORCID: <https://orcid.org/0000-0001-7929-4977>
** 군산대학교 소프트웨어융합공학과 교수(교신저자)
- ORCID: <http://orcid.org/0000-0002-3390-5602>

· Received: Nov. 21, 2019, Revised: Dec. 09, 2019, Accepted: Dec. 12, 2019
· Corresponding Author: Sukhoon Lee
Dept. of Software Convergence Engineering, Kunsan National University,
Korea,
Tel.: +82-63-469-8914, Email: leha82@kunsan.ac.kr

I. 서 론

최근 웨어러블 디바이스는 사물인터넷 기술의 발전과 함께 성장해 왔다. 특히 Fitbit, 애플 워치, 갤럭시 워치 등과 같은 스마트워치는 그 시장 규모가 폭발적으로 증가하고 있다[1]. 스마트워치는 GPS, 3축 가속도, 심박계, 자이로 센서 등을 지니고 있으므로 사용자는 일상생활을 하는 동안 착용하며 사용자 스스로 모니터링하기 용이하다[2]. 따라서 스마트워치는 실시간으로 사용자의 위치 정보, 생체 정보, 활동량 등을 얻을 수 있으므로 라이프로그 수집에 적절하다.

한편, 라이프로그는 사용자가 생활하면서 발생하는 모든 형태의 기록을 의미한다. 라이프로그는 특정 개인의 영상, 위치 정보, 메시지, 행동 로그 등의 기록을 통하여 사용자의 경로 예측, 행동 분석, 질병 예측 등과 같은 다양한 영역에서 활용된다[3]. 특히 경로 예측은 사용자의 위치 정보를 이용하여 행동 및 생활 패턴을 분석할 수 있으므로 라이프로그가 적절히 활용될 수 있는 영역이다.

대부분의 경로 예측 알고리즘들은 GPS를 이용한 사용자의 위치 정보들을 학습 데이터로 요구한다[4]. 또한, 경로 예측 기법은 정확한 경로 예측을 위하여 많은 양의 학습 데이터를 요구하는데, 사용자의 위치 데이터가 부족할 경우 경로 예측의 성능이 저하된다는 문제를 지닌다. 예를 들어 사용자가 어떤 지역을 처음 방문하거나 새로운 경로로 진입하였을 때, 기존의 경로 예측 알고리즘들은 학습된 데이터가 없으므로 경로 예측 결과를 제시하지 못한다.

한편, 협력 필터링(Collaborative filtering)은 자신과 유사한 성향을 지닌 사용자를 이용하여 아이템을 추천해 주는 기법이다[5]. 협력 필터링은 사용자에게 맞춤 서비스를 제공하기 위하여 Amazon, 알라딘, 왓챠, 다음 쇼핑하우 등에서 추천 시스템으로 활용되고 있다[6]. 이를 적용하면 앞에서 제시한 학습 데이터의 부재에 대한 문제를 해결할 수 있다. 많은 사람들은 친구 또는 직장 동료들과 함께 집단적으로 이동하는 경우가 많다. 이러한 경우 서로 같은 이동 패턴을 가지게 되므로 서로 간의 유사도가 높아지며 이를 경로 예측 알고리즘에 적용할 경우 사용자의 학습 데이터가 없는 지역에서도 유사도가

높은 다른 사용자의 데이터를 활용할 수 있다.

따라서 이 논문은 사용자 간의 유사도를 측정하고 이를 기반으로 경로 예측에 적용하는 알고리즘을 제안한다. 제안 기법은 경로 학습을 위하여 시간, 3축 가속도, 심박 수, GPS의 위치 정보를 이용한다. 또한, 제안 기법은 사용자의 위치 정보를 일정 크기의 그리드로 간소화하고, 사용자의 이동 궤적을 3단 그리드 패턴(TGP, Triple Grid Pattern)으로 변환하여 학습한다. 사용자 유사도의 계산을 위하여 코사인 유사도 공식을 이용한다. 이후, 이 논문은 경로 예측을 위하여 사용자 간 유사도와 TGP의 가중치를 이용한 알고리즘을 제안한다. 평가를 위하여 이 논문은 제안 알고리즘과 기존 경로 예측 알고리즘의 예측 정확도를 측정한다.

이 논문은 다음과 같이 구성된다. 2장에서는 관련 연구들을 기술하며, 3장에서는 제안 방법을 소개한다. 4장에서는 실험을 통해 제안 알고리즘의 정확도를 비교 평가한다. 5장에서는 결론과 향후 연구를 기술한다.

II. 관련 연구

이 장은 경로 예측 알고리즘에 관련한 연구들을 기술한다. 많은 경로 예측 알고리즘들은 개인의 데이터를 활용하여 경로 학습 및 예측을 수행한다[7][8]. GPS 기반의 개인 경로 모델 학습 및 이동 경로 예측에 관련된 연구는 너비 우선 탐색 알고리즘을 이용하여 라인 골격을 그래프로 변환하고 이를 경로 패턴으로 학습한다[7]. 이후 경로 예측을 위하여 최소 경계를 이용한 후보 경로 결정 단계와 최하 경계를 이용한 유효성 검증 단계를 거친다.

동적 베이저안 망(Dynamic bayesian network) 모델을 이용한 경로 예측 기법은 확률에 기반한 개인화된 경로 예측 기법 중 하나로 RBPF(Rao-Blackwellized Particle Filtering)를 이용하여 실시간으로 경로를 예측한다[8].

하지만 이런 연구들은 사용자의 이전 경로를 학습하여 개인화 예측을 하므로 해당 사용자의 데이터가 부족할 경우 경로 학습이 어렵다는 문제가 있다. 즉, 사용자가 여행, 출장 등으로 인해 학습되지 않은 지역을 이동하게 된다면 개인화 경로 예측 기

법은 사용자의 이전 학습 데이터가 없으므로 예측이 불가능하다.

이러한 문제 해결을 위해, 여러 사용자들로부터 패턴을 학습하여 개인에게 적용하는 연구들이 진행되었다[9][10]. 집단행동 패턴(CBP, Collective Behavior Pattern) 기반 경로 예측 기법은 사용자의 위치를 스마트폰의 기지국을 통하여 식별하고 이를 Naive Bayes 모델로 학습한다. 이후, 다른 사용자의 이동 패턴과 혼합하여 경로를 예측한다[9]. 집단행동 패턴이란 집단의 행동이 개인의 행동에 영향을 미칠 수 있다는 개념으로 개인화 경로 예측의 문제점 해결을 위해 적용될 수 있다.

빠른 투영 후보 알고리즘은 센서 레지스트리 시스템의 정보를 사용자에게 제공하기 위해서 근거리의 경로를 예측하고 해당 경로 주변의 센서 정보를 사용자에게 제공해 준다[10]. 빠른 투영 후보 알고리즘은 단순하지만 빠른 속도의 정보 제공을 위하여 모든 사용자의 경로 패턴을 통합하여 학습한다.

하지만, 이러한 CBP 기반의 경로 예측 기법들은 학습 데이터 부족의 문제 해결을 위해 고안되었기 때문에, 학습 데이터가 커질수록 개인화된 경로 예측 기법들보다는 성능이 떨어질 수밖에 없다. 따라서 개인화된 경로 예측 기법과 CBP 기반의 경로 예측 기법들은 상호 보완적인 관계를 지닌다.

한편, 경로를 표현함에 있어서, 3간선 패턴(TEP, Three-Edge Pattern) 기반 경로 예측 기법은 경로 학습 시 사용자의 이동 경로를 연속된 3개의 도로 간선으로 패턴화한다[11]. 하지만 TEP는 목표 영역을 선정하고 해당 영역 안의 도로 정보를 구축해야 하므로 높은 비용이 요구된다. 또한, 경로를 도로로 치환하여 표현하기 때문에 도로가 아닌 광장과 같은 영역의 경로를 정확하게 표현하지 못한다. 이는 사용자가 해당 영역을 지나가는 경우 학습이 불가능하므로 경로 예측의 정확도가 저하된다는 문제점을 지닌다[12].

이러한 문제는 일정 범위를 균등한 그리드(Grid)로 나누어 이동 경로를 패턴화시키는 방법으로 해결할 수 있다. 그리드 기반 경로 예측 기법은 사용자의 위치 및 이동 경로를 그리드로 식별하여 패턴화한다[13]. 그리드 기반 기법은 도로 정보 기반 기

법과 비교하였을 때 유지 보수 비용이 적고 적은 전처리 비용과 오류를 줄일 수 있다는 장점을 지닌다.

III. 유사도 기반 경로 예측 기법

이 장에서는 제안 기법인 사용자 유사도 기반 경로 예측 알고리즘을 기술한다. 이를 위하여 경로 학습 및 예측 프로세스를 정의하고, TGP를 이용한 경로 학습 기법을 기술한다. 이후 코사인 유사도 공식을 이용한 사용자 유사도 측정 기법을 설명하고, 사용자 유사도에 기반한 경로 예측 알고리즘을 기술한다.

3.1 제안 프로세스

그림 1은 이 논문에서 제안하는 경로 학습 및 예측 프로세스를 보인다. 첫 번째 단계에서는 라이프로그를 수집한다. 이후, 수집된 사용자의 이동 위치를 그리드로 간소화하고, 경로를 TGP로 학습한다. 다음 단계로는 학습된 TGP의 가중치를 벡터화하여 사용자들 간의 유사도를 측정한다. 사용자 유사도는 코사인 유사도 기법을 이용한다. 이후, 학습된 TGP 가중치와 측정된 사용자 유사도를 이용하여 경로 예측 모델을 생성한다. 마지막으로 제안 모델과 기존 경로 예측 모델들의 정확도를 측정하는 것으로 비교 평가를 진행한다.

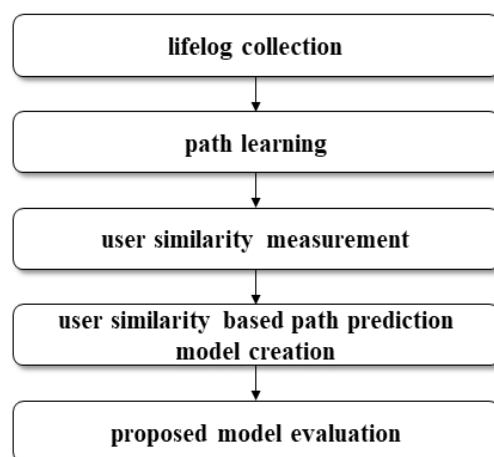


그림 1. 경로 학습 및 예측 프로세스
Fig. 1. Processes of path learning and prediction

3.2 경로 학습

경로의 패턴을 생성하기 위해서는 사용자의 위치 정보들을 간소화하여 표현할 필요가 있다. 사용자가 동일한 경로를 이동하더라도 수집되는 GPS 위치 정보는 동일하지 않기 때문이다.

따라서 이 논문은 그리드를 이용하여 사용자의 위치 정보를 간소화한다. 그리드는 지도의 특정 범위를 일정하게 나눈 직사각형의 면적을 의미한다. 만약 사용자의 위치 정보를 그리드로 매핑하여 표현한다면, 사용자의 이동 경로는 그리드의 순서로 패턴화되어 학습하기 용이해진다.

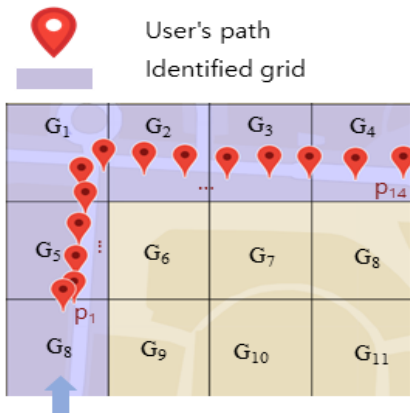


그림 2. 이동 궤적의 그리드 패턴 표현
Fig. 2. Grid pattern expression of a trajectory

표 1. TGP 기반 경로 학습의 예
Table 1. Example of path learning based on TGP

Pattern (tgp)	User	Previous grid	Current grid	Next grid	Frequency (f)	Weight (w)
tgp ₁	User ₁	G ₈	G ₅	G ₁	45	0.069
tgp ₂	User ₁	G ₅	G ₁	G ₂	80	0.123
tgp ₃	User ₁	G ₁	G ₂	G ₃	50	0.077
tgp ₄	User ₁	G ₂	G ₃	G ₄	70	0.108
tgp ₅	User ₁	G ₇	G ₈	G ₂	30	0.046
tgp ₆	User ₁	G ₇	G ₈	G ₃	35	0.054
tgp ₇	User ₁	G ₇	G ₈	G ₄	25	0.038
tgp ₈	User ₁	G ₇	G ₈	G ₉	90	0.138
tgp ₉	User ₁	G ₇	G ₈	G ₁₄	65	0.100
tgp ₁₀	User ₁	G ₇	G ₈	G ₁₃	55	0.085
tgp ₁₁	User ₁	G ₇	G ₈	G ₁₂	60	0.092
...
tgp _n	User _m	G _{prev}	G _{curr}	G _{next}	X	W _{Userm,tgpn}

그림 2는 사용자 이동 궤적을 그리드로 간소화하여 패턴으로 표현한 것이다. 그리드는 G_i 같이 나타내며, 사용자의 점은 p_j로 나타낸다. 사용자가 빨간 점 p₁부터 p₁₄까지 순차적으로 이동할 때, 이동 패턴은 G₈ → G₅ → G₁ → G₂ → G₃ → G₄로 표현할 수 있다.

이동 궤적이 그리드 패턴으로 매핑되었다면, 이 패턴을 TGP로 학습한다. TGP는 3개의 연속적인 그리드로 묶은 패턴으로 각 그리드는 이전 그리드, 현재 그리드, 다음 그리드로 {G_{previous}, G_{current}, G_{next}}와 같이 구성된다. 이는 경로 예측 시 사용자가 위치한 현재 그리드(G_{current})를 기준으로 이전에 어디서 왔는지에 대한 이전 그리드(G_{previous}), 다음으로 이동할 위치인 다음 그리드(G_{next})를 묶은 것이다.

따라서 그림 2의 그리드 기반 이동 패턴을 TGP로 표현한다면 {G₈, G₅, G₁}, {G₅, G₁, G₂}, {G₁, G₂, G₃}, {G₂, G₃, G₄}과 같이 4개의 TGP로 추출할 수 있다. 이와 같이 모든 사용자의 이동 궤적에 대하여 TGP를 학습한다.

표 1은 사용자의 이동 궤적들을 학습하였을 때의 TGP의 가중치를 보인다. 각 이동 궤적 별로 동일한 TGP 패턴이 추출될 수 있으므로 frequency는 패턴이 추출되는 횟수를 의미한다. 또한, 사용자 간 유사도 측정 시 벡터화를 위하여 정규화가 요구되기 때문에 weight는 전체 사용자에 대한 해당 패턴의 비율을 의미한다.

3.3 유사도 비교 방법

이 논문은 CBP의 개념을 경로 예측에 적용하기 위해 사용자 유사도 측정 기법을 TGP에 적용한다. 사용자들의 이동 궤적을 TGP 기반 모델로 학습한 후 이 모델의 비교를 통하여 사용자의 유사도를 측정한다. 그림 3은 사용자별 TGP의 가중치를 벡터화하여 표현한 것이다.

사용자 유사도 측정을 위해 각 사용자 간 벡터의 거리를 구하는 방법은 유클리드 거리 공식을 사용한 K-NN 기법[14]과 벡터 사이의 각도를 구하는 코사인 유사도 측정 기법이 있다[15]. 이 논문은 TGP 벡터를 이용하였을 때 더 정확도가 높은 코사인 유사도 기법을 활용한다[16].

$$\begin{aligned}
 User_1 &= \{w_{u_1, tgp_1}, w_{u_1, tgp_2}, w_{u_1, tgp_3}, \dots, w_{u_1, tgp_n}\} \\
 User_2 &= \{w_{u_2, tgp_1}, w_{u_2, tgp_2}, w_{u_2, tgp_3}, \dots, w_{u_2, tgp_n}\} \\
 User_3 &= \{w_{u_3, tgp_1}, w_{u_3, tgp_2}, w_{u_3, tgp_3}, \dots, w_{u_3, tgp_n}\} \\
 &\vdots
 \end{aligned}$$

그림 3. 사용자별 패턴 가중치 벡터화
Fig. 3. Vectorizing pattern weight per user

이 논문에서 사용자들 사이의 유사도를 구하기 위하여 사용하는 코사인 유사도 기법은 식 (1)과 같다.

$$sim(u_i, u_j) = \frac{\sum_{k=1}^n (w_{u_i, tgp_k} \cdot w_{u_j, tgp_k})}{\sqrt{\sum_{k=1}^n w_{u_i, tgp_k}^2} \sqrt{\sum_{k=1}^n w_{u_j, tgp_k}^2}} \quad (1)$$

3.4 제안 알고리즘

이 절은 이 논문에서 제안하는 TGP 기반 경로 예측의 방법을 기술한다. 제안하는 경로 예측 기법은 사용자가 현재 위치한 그리드와 이전의 그리드를 이용하여 다음 그리드를 예측하는 방식이다. 그림 4는 TGP를 이용한 경로 예측 방식을 표현한다. G₇과 G₈이 각각 이전 그리드와 현재 그리드라고 할 때, 다음으로 이동할 그리드는 G₂, G₃, G₄, G₉, G₁₂, G₁₃, G₁₄ 중 하나로 예측할 수 있다. 이때, 다음으로 이동할 그리드는 패턴별 가중치를 비교하여 가장 높은 가중치를 지니는 패턴을 선정한다. 표 1에서 이전과 현재 그리드가 G₇, G₈일 때, 다음 그리드의 후보들 중 확률이 가장 높은 tgp₈로 선정할 수 있으며 다음 그리드는 G₉가 된다.

한편, 경로 예측에 사용자 유사도 기법을 적용하기 위해서는 식 (2), (3)을 사용한다.

$$tgp_k = \{G_{prev}, G_{curr}, G_{next}\}$$

$$P_{u_i, tgp_k} = \frac{f_{u_i, tgp_k}}{\sum_{j=1}^{|USER|} f_{u_j, tgp_k}} \quad (2)$$

$$Sim W(u_i, tgp_k) = \alpha \cdot \sum_{j=1}^{|USER|} (sim(u_i, u_j) \cdot P_{u_j, tgp_k}) + (1 - \alpha) \cdot P_{u_i, tgp_k} \quad (3)$$

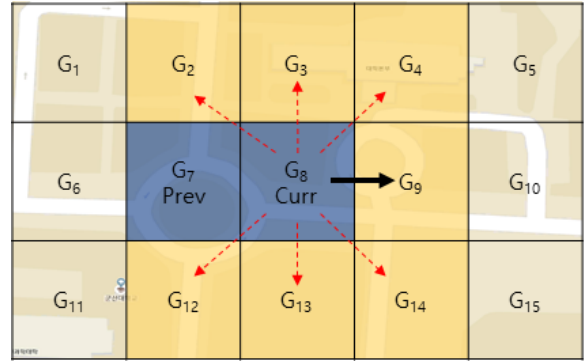


그림 4. TGP 기반 경로 예측 방법
Fig. 4. TGP based path prediction method

USER는 u₁, u₂, ..., u_n을 지니고, TGP는 tgp₁, tgp₂, ..., tgp_m을 지닌다. 그리드는 G₁, G₂, ..., G_l로 구성된다. u_i의 tgp_k에 대한 가중치 값을 구하기 위해서는 u_i와 자신을 제외한 다른 사용자와 코사인 유사도를 구한 뒤 다른 사용자의 tgp_k의 확률을 곱한다. 그리고 u_i의 tgp_k에 대한 확률을 더 해준다. 이때, α는 tgp_k의 이동 빈도(frequency)의 크기에 따라 조정하여 최적의 값을 찾아 선정한다.

IV. 실험 및 평가

4.1 실험 방법

이 논문은 실험을 위하여 사용자의 GPS 위치 정보를 포함한 라이프로그 데이터 셋인 DSEM-Traj2018을 사용한다[17]. 이 데이터 셋은 국내 한 대학교와 대학로 부근에서 스마트워치로부터 수집된 라이프로그를 대상으로 한다.

DSEM-Traj2018은 총 170일 동안 39명이 자유롭게 이동하며 수집된 데이터로 이루어져 있으며, 사용자로부터 787,966 레코드로 이루어진 총 9,330개의 궤적을 수집하였다. 데이터 수집에 참여한 실험자 중 남자 27명, 여자 12명으로 이루어져 있으며, 교수 5명과 학생 34명으로 이루어져 있다.

이 논문은 비교 평가를 위하여 제안 기법과 3가지 기존 경로 예측 기법을 비교 대상으로 선정한다.

첫 번째 비교 대상은 CBP 기반 경로 예측 알고리즘이다. 즉, 이 기법은 사용자의 구분 없이 경로 패턴을 학습하기 때문에 모든 사용자가 동일한 경로 예측 모델을 사용함을 의미한다.

두 번째 비교 대상은 개인화 경로 학습을 의미하는 개별 행동 패턴(IBP, Individual Behavior Pattern) 기반 경로 예측 알고리즘이다. 이 기법은 사용자가 자신의 데이터만을 이용하여 학습한 모델이므로 사용자별로 경로 예측 모델을 지님을 의미한다.

세 번째 비교 대상으로는 IBP와 CBP를 혼합한 하이브리드(Hybrid) 기법이다. 이 기법은 개인화 패턴의 빈도수에 따라 IBP와 CBP의 모델을 차등하게 적용하는 방식이다.

한편, 그림 5는 사용자의 이동 경로를 그리드로 식별하기 위한 크기별 지도를 보인다. 이 논문은 지도의 대상 구간을 (5×5, 10×10, 15×15, 20×20)와 같은 동일한 크기의 그리드로 나눈다. 그리드의 크기는 단위 면적에 따라 사용자의 경로를 세분화시킬 수 있다. 이는 곧 학습되는 경로 패턴의 양과 경로 예측의 정확도에 영향을 주기 때문에, 실험을 통하여 적절한 크기의 그리드를 선정한다.

각 그리드의 면적은 5×5(48,348m²), 10×10(12,087m²), 15×15(5,372m²), 20×20(3,022m²)이며, DSEM-Traj2018의 이동 궤적 데이터들을 모두 그리드로 식별하여 간소화한다.

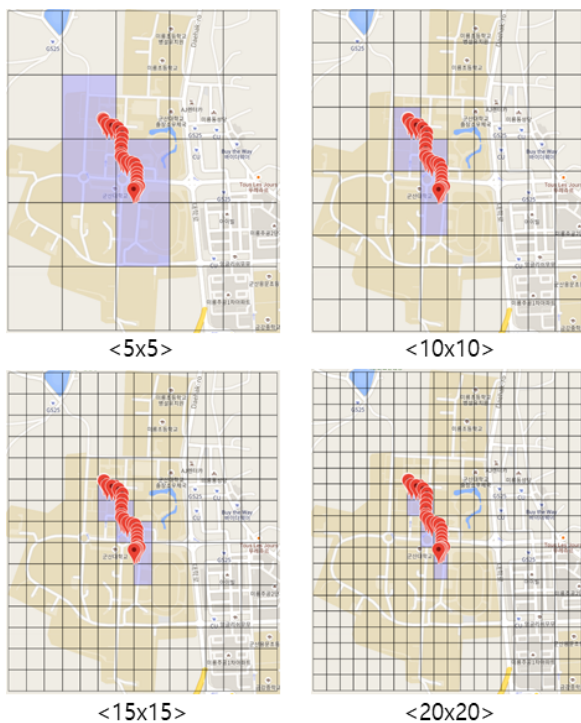


그림 5. 그리드 크기별 이동 궤적의 식별
Fig. 5. Identification of a trajectory by grid size

첫 번째 실험으로는 최적의 그리드 크기를 선정하기 위하여 실험 데이터를 각각의 그리드에 식별한 후 경로 예측 모델을 학습하고 정확도를 측정한다. DSEM-Traj2018의 전체 이동 궤적 중 임의의 10%를 테스트 데이터로 추출하고 나머지는 학습 데이터로 이용한다.

두 번째 실험으로는 사용자별로 진행되며 사용자들의 학습 데이터양에 따른 정확도를 비교한다. 이를 위하여 수집된 이동 궤적의 수가 100개 이상인 사용자들에 한하여 학습 데이터의 수를 해당 사용자의 10%~90%까지 조절하며 학습한다. 테스트 데이터는 사용자의 임의의 10%로 고정하여 학습 데이터의 양에 따른 정확도의 추세를 측정한다. 이 실험에서 사용자 한 명의 데이터를 학습할 때, CBP 모델 학습 및 사용자 유사도 측정을 위하여 해당 사용자를 제외한 다른 모든 사용자의 데이터를 학습한다. 이후 해당 사용자의 학습 데이터 수를 조절해 가며 추가로 학습한다. 이러한 실험은 사용자가 기존의 학습 데이터의 양에 따라 얼마나 경로 예측의 결과에 영향을 미치는지를 확인할 수 있다.

4.2 실험 결과

표 2는 코사인 유사도 기법을 기반으로 사용자들 간의 유사도를 측정한 값이다. 사용자들 간의 유사도는 최대 0.982, 최소 0.109 값을 가진다.

표 2. 사용자 유사도 측정 값
Table 2. User similarity measure

	GAH0 001	GAH0 002	GAH0 003	...	GAH0 039
GAH0 001	0	0.6526	0.7035	...	0.1952
GAH0 002	0.6526	0	0.8092	...	0.1721
GAH0 003	0.7035	0.8092	0	...	0.2256
...
GAH0 039	0.1952	0.1721	0.2256	...	0

그림 6은 첫 번째 실험의 결과인 그리드 크기별 경로 예측 정확도를 보인다. 각 기법별로 CBP

62.0%, IBP 63.9%, Hybrid 64.3%, 제안 기법 64.7%와 같은 평균 정확도를 가진다. 즉, 제안 기법이 가장 높은 정확도를 지니는 것을 확인할 수 있다. 또한, 5×5, 10×10, 15×15, 20×20 크기 별 평균은 각각 61.2%, 65.3%, 63.1%, 65.5%로 20×20이 가장 높은 평균 정확도를 가진다. 이렇듯 그리드는 그 크기에 따라 각각 생성되는 패턴의 수 및 가중치 크기가 다르게 학습되며 이는 정확도에 영향을 미칠 수 있다.

첫 번째 실험의 또 다른 목적은 실제 경로 예측 기법의 적용을 위하여 최적의 그리드 크기를 선정하는 것이다. 앞서 언급했듯 가장 높은 평균 정확도

를 가진 그리드 크기는 20×20이지만, 10×10의 경우 제안 기법이 전체 조건 중에서 최고 정확도를 기록하였다. 또한, 향후 가장 높은 정확도를 지니는 알고리즘만을 선택하여 사용한다는 것을 고려하면 10×10이 최적의 그리드 크기로 분석된다.

두 번째 실험은 사용자별로 학습 및 테스트가 진행되었다. 경로 학습 시 목표 사용자를 제외한 모든 사용자들의 경로를 학습한 후 목표 사용자의 데이터는 차등을 두어 가며 학습하고 정확도를 측정하였다.

그림 7은 각 사용자들이 수집한 데이터양과 경로 예측 기법들의 정확도를 보인다.

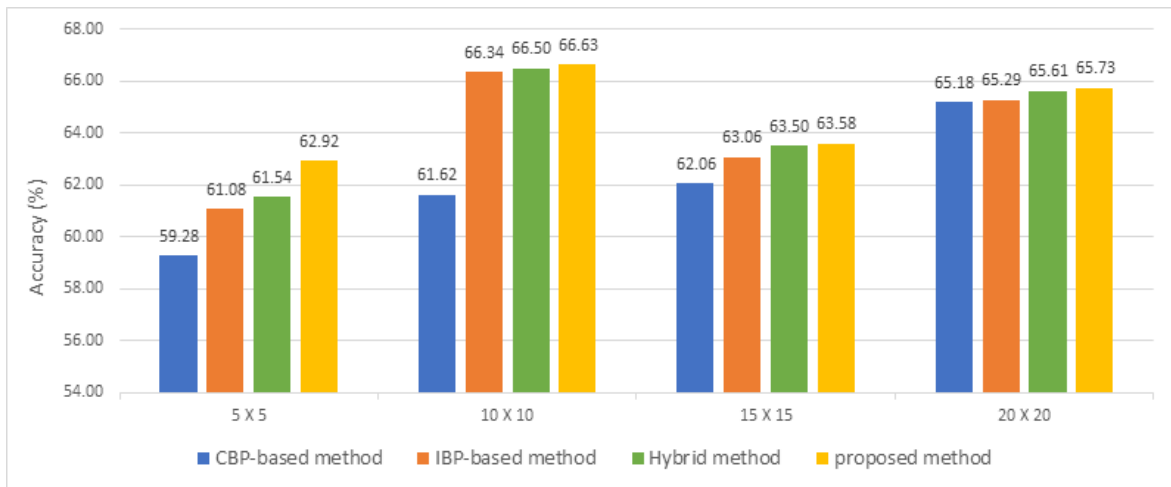


그림 6. 그리드 크기별 정확도
Fig. 6. Accuracy by grid size

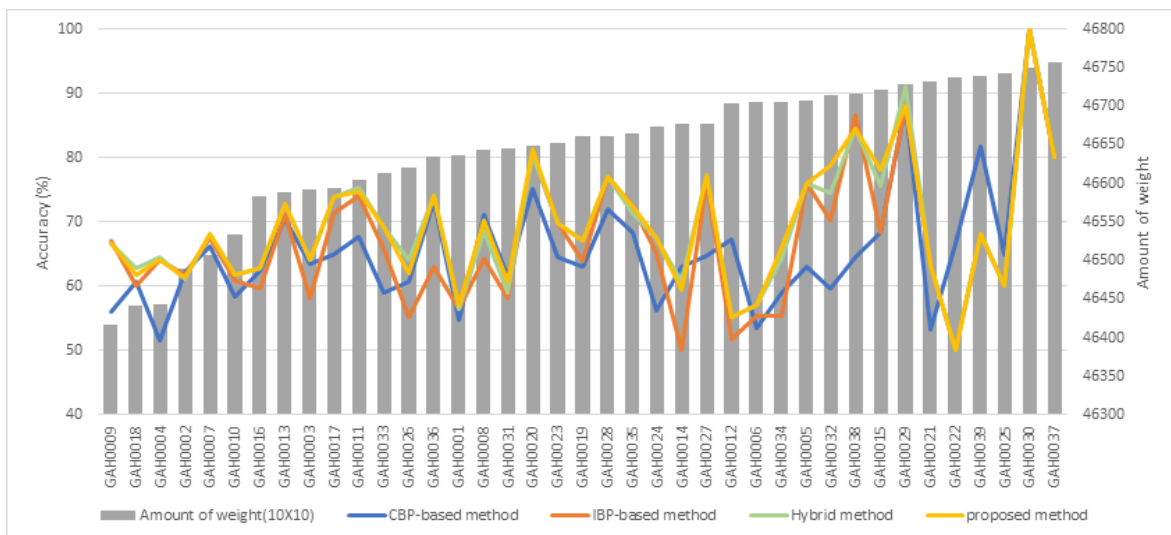


그림 7. 사용자별 정확도 및 가중치의 양
Fig. 7. Accuracy and amount of weight per user

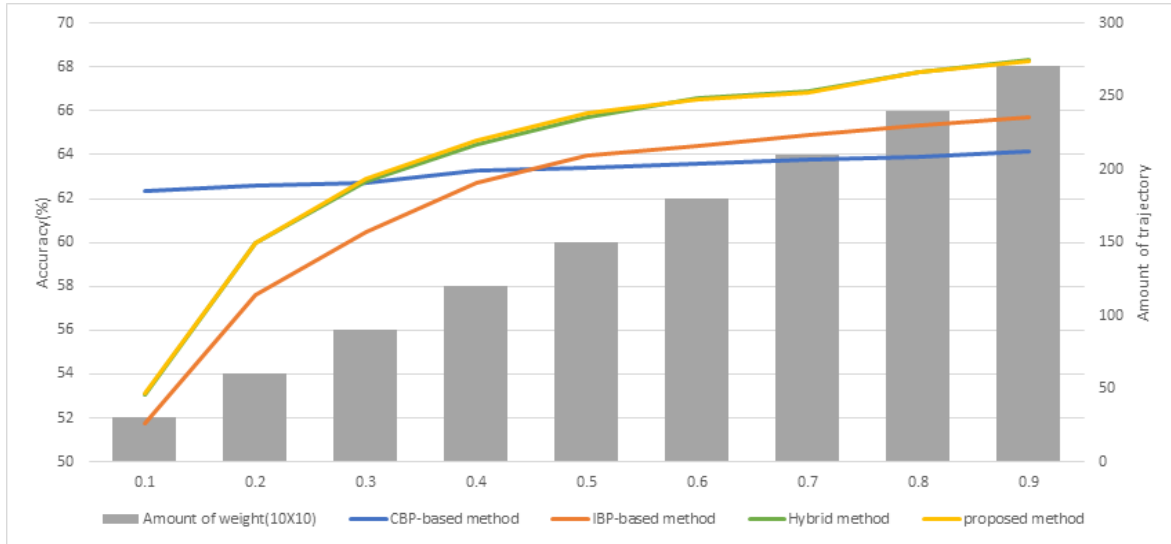


그림 8. 학습 데이터 양에 따른 평균 정확도와 평균 이동 궤적 수

Fig. 8. Average accuracy and average number of trajectory according to amount of training data

각 사용자들의 경로 학습 시 측정되는 패턴의 양으로 정렬을 해 본 결과, 사용자별 데이터 수집 양과 경로 예측의 정확도는 연관성이 없어 보이는 것으로 판단된다. 즉, 사용자에게 따라 데이터 수집 양과 무관하게 경로 예측의 정확도는 다를 수 있다는 것을 알 수 있다.

그림 8은 사용자별 경로 학습 시 사용자의 학습 데이터의 증가에 따른 정확도의 변화를 비교한 그래프이다. 그림 7에서 사용자별로 데이터의 양과 정확도는 무관한 것으로 평가되었지만, 그림 8에서는 평균적으로 사용자의 학습 데이터양이 점차 증가할수록 정확도도 높아지는 것을 확인하였다. 이는 사용자 개개인을 중심으로 보았을 때 해당 사용자의 학습 데이터가 많을수록 생활 패턴이나 이동 경로의 변화가 단조롭거나 복잡해진다고 볼 수 없지만, 전체적인 사용자들의 추세로 보았을 때는 데이터가 많아질수록 경로 예측의 정확도가 높아진다는 것을 알 수 있다.

또한, 그림 8의 기법별 정확도를 보면 학습률이 0.2 이하일 때는 CBP 기법의 정확도는 62.6%로 다른 기법들에 비해 가장 높다. 하지만, 학습률이 0.3 부터는 제안 기법이 62.8%로 가장 높은 정확도를 나타내며, 학습률이 높아짐에 따라 점차 상승하여 제안 기법은 0.9일 경우 68.3%로 가장 높은 정확도를 보인다.

V. 결 론

이 논문은 기존의 경로 예측 기법들의 성능 향상을 위하여 사용자 유사도 기반 경로 예측 알고리즘을 제안하였다. 이를 위하여 사용자의 GPS 위치 정보를 그리드로 식별한 후 TGP로 변환시킨 후 경로 예측 모델을 학습하였다. 이후 사용자별로 경로 예측 모델을 벡터화시키고 코사인 유사도 기법을 통하여 사용자 간 유사도를 측정하였다. 측정된 유사도는 제안 알고리즘에 의해 경로 예측에 활용된다.

제안 기법의 평가를 위하여 DSEM-Traj2018 데이터 셋을 이용하여 경로를 학습하고 예측을 진행한다. 제안 기법과의 비교 대상으로는 CBP, IBP, Hybrid 기법들을 각각 선정하였으며, 실험 결과 10x10 그리드 크기에서 제안 기법은 66.6%로 다른 기법들보다 더 우수한 성능을 지니는 것으로 평가되었다. 또한, 사용자별 학습률에 따른 정확도 비교에서는 학습률이 0.2 이하일 경우 CBP가 62.6%로 가장 높았지만, 학습률이 높아짐에 따라 제안 기법은 학습률이 0.3에서 0.9로 점차 증가함에 따라 62.8%에서 68.3%로 가장 높은 정확도의 상승을 보였다.

이 연구는 기존의 경로 예측 기법들 중 해당 사용자들의 데이터의 양이 부족할 경우 다른 사용자의 데이터가 유용하게 활용될 수 있다는 것을 검증할 수 있다.

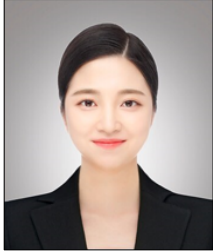
향후 연구에서는 DSEM-Traj2018에 수집된 요소 중 시간 및 행동에 대한 요소들을 활용하여 경로 예측 모델을 학습하였을 때 얼마나 성능이 향상되는지를 검증해 볼 예정이다.

References

- [1] B. Jeong, "Current Status and Prospects of the Wearable Devices Market", *Telecommunications Broadcasting Policy Periodicals*, Vol. 30, No 20, pp. 1-7, Nov. 2018.
- [2] S. Kwon and S. Lee, "Relational Database Model for Collecting Lifelog from Heterogeneous Smart Watches", *Journal of KIIT*, Vol. 16, No. 9, pp. 13-21, Sep. 2018.
- [3] C. Gurrin, A. F. Smeaton, and A. R. Doherty, "LifeLogging: Personal Big Data", *Foundations and Trends in Information Retrieval*, Vol. 8, No. 1, pp. 1-125, Jun. 2014.
- [4] J. M. Kim, H. Baek, and Y. T. Park, "Probabilistic graphical model based personal route prediction in mobile environment", *Applied Mathematics & Information Sciences*, Vol. 6, No. 6S, pp. 651S- 659S, Jan. 2012.
- [5] S. K. Lee, Y. H. Cho, and S. H. Kim, "Collaborative filtering with ordinal scale-based implicit ratings for mobile music recommendations", *Information Sciences*, Vol. 180, No. 11, pp. 2142-2155, Jun. 2010.
- [6] J. Son, S. B. Kim, H. Kim, and S. Cho, "Review and Analysis of Recommender Systems", *Journal of Korean Institute of Industrial Engineers*, Vol. 41, No. 2, pp. 185-208, Apr. 2015.
- [7] J. M. Kim, H. J. Baek, and Y. T. Park, "GPS based Smartphone Users' Personal Route Model Learning and Moving Route Prediction", *Journal of KIISE: Software and Applications*, Vol. 39, No. 1, pp. 56-65, Jan. 2012.
- [8] M. O. Heo, M. Kang, B. K. Lim, K. B. Hwang, Y. T. Park, and B. T. Zhang, "Real-time Route Inference and Learning for Smartphone Users using Probabilistic Graphical Models", *Journal of KIISE: Software and Applications*, Vol. 39, No. 6, pp. 425-435, Jun. 2012.
- [9] H. Xiong, D. Zhang, D. Zhang, and V. Gauthier, "Predicting Mobile Phone User Locations by Exploiting Collective Behavioral Patterns", *Proc. of the 9th International Conference on UIC/ATC*, Fukuoka, Japan, pp. 164-171, Sep. 2012.
- [10] D. Jeong and M. Doo, "A Path Prediction-Based Sensor Registry System for Stable Use of Sensor Information", *Journal of KIISE*, Vol. 42, No. 2, pp. 255-263, Feb. 2015.
- [11] S. Lee, D. Jeong, H. Jung, and D. K. Baik, "Three-Edge Pattern based Path Prediction Algorithm for Sensor Registry System", *Korea Information Processing Society 2015 Spring Conference*, pp. 798-801, Apr. 2015
- [12] H. Jung D. Jeong, and S. Lee, "Development of Sensor Registry System-Based Predictive Information Service Using a Grid", *Sensors*, Vol. 18, No. 11: 3620, pp. 1-17, Oct. 2018.
- [13] H. Jung, S. Lee, and D. Jeong, "Improving Path Prediction Using Grid for Sensor Registry System", *Journal of KIIT*, Vol. 16, No. 3, pp. 1-10, Mar. 2018.
- [14] Jiawei Han, Micheline Kamber, and Jian Pei, "Data Mining (3rd Ed.) - Concepts and Techniques", *Acompub*, pp. 542-545, 2015.
- [15] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl "Item-Based Collaborative Filtering Recommendation Algorithms", *Proc. of the 10th World Wide Web Conference*, Hong Kong, pp. 285-295, May 2001.
- [16] S. Nam, T. Ka, and S. Lee, "Comparison of User Similarity Methods based on User Movement Path Pattern", *Korea Information Technology Society Collective Papers*, pp. 310-311, Jun. 2019.
- [17] S. Nam, S. J. Kim, and S. Lee, "DSEM-Traj2018: Building Lifelog Dataset using Smartwatch for Path Prediction", *World IT Congress*, Feb. 2019.

저자소개

남 수 민 (Sumin Nam)



2016년 3월 ~ 현재 : 군산대학교
소프트웨어융합공학과 학부생
관심분야 : 경로 예측, 데이터 분석

이 석 훈 (Sukhoon Lee)



2009년 2월 : 고려대학교
전자및정보공학부(학사)
2011년 2월 : 고려대학교
컴퓨터·전파통신공학과(공학석사)
2016년 2월 : 고려대학교
컴퓨터·전파통신공학과(공학박사)
2016년 3월 ~ 2017년 3월 :

아주대학교 의료정보학과 연구강사

2017년 4월 ~ 현재 : 군산대학교 소프트웨어융합공학과
조교수

관심분야 : 사물인터넷, 메타데이터, 센서 레지스트리,
시맨틱 웹, 경로 예측