JPNT 14(1), 1-10 (2025)
https://doi.org/10.11003/JPNT.2025.14.1.1

JPNT Journal of Positioning, Navigation, and Timing

익명의 사용자 이동 패턴 학습과 지리적 이상감지 연구

김지형, 유재현†

A Study for Learning Anonymous Multi-Trajectory Patterns and Geographical Anomaly Detection

Jihyoung Kim, Jaehyun Yoo†
School of AI Convergence, Sungshin Women's University, Seoul 02844, Korea

ABSTRACT

Personal safety and crime prevention have become pressing societal concerns. While wearable devices such as smartwatches offer features including Global Positioning System (GPS) tracking and emergency alerts, their ability to proactively recognize deviations from a user's usual path is limited. This study proposes a Long Short-Term Memory (LSTM) based trajectory learning algorithm that leverages anonymized user data without additional identifiers. It enables detection of changes from a user DB of usual trajectories and thus allows recognition of anomalies in real-time. Experimental results demonstrate that the model achieves relatively consistent performance in predicting distance errors for paths, although the time prediction performance may vary depending on path characteristics. In anomaly detection analyses, normal paths maintained stable values without exceeding the set threshold, while anomalous paths exhibited increasing error values over time, eventually exceeding the threshold.

Keywords: trajectory learning, personal protection, LSTM, anomaly detection
주요어: 경로 학습, 신변 보호, LSTM, 이상 감지

1. 서론

Global Positioning System (GPS) 데이터를 기반으로 한 이동 패턴 분석은 보안, 물류, 그리고 위치 기반 서비스 등 다양한 분야에서 중요한 역할을 차지하고 있다. 특히 스마트 워치와 같은 웨어러블 장치가 신변 보호 용도로 상용화되고 있으며, 이는 사용자의 위치를 신속하게 파악할 수 있는 장점이 있다. 하지만 여러 명의 사용자 정보가 포함된 GPS 데이터를 중앙시스템에서 관리할 경우 개인정보 침해의 우려가 있다. 따라서 신변 보호 시스템의 일환으로 GPS 데이터를 사용할 때 사용자의 식별 정보를 요구하지 않으며 익명성을 보호하는 효과적인 이상 감지 알고리즘이 필요하다. 또한 위치 정보만을 이용한 지리적 이상 감지는 Geofencing (Nam et al. 2023, Shevchenko & Reips 2024)과 같은 기존 기술로도 효과적인 감지가 가능하지만, 본 논문에서는 시간적 경로 이탈 및 복합적 시공간 이상 감지를 위해 Long Short-Term Memory (LSTM) (Hwang & Shin 2020, Ji et al. 2020, Kim et al. 2021, Yoon et al. 2022, Shin et al. 2024) 기반 모델을 활용한다.

GPS 데이터는 Time-series data 즉 시간순으로 관측된 데이터로, 이전 정보와 현재 정보가 다음 정보에 영향을 미친다. 시계열 데이터를 처리하는 모델은 목적에 따라 Recurrent Neural Networks (RNN), Convolutional Neural Networks (CNN), Transformer, Clustering 및 Classification Model, 혹은 Survival Analysis Model 등 다양하게 존재한다. 그 중에서 순차적인 데이터를 처리하고 예측하는데 효과적인 RNN 유형의 LSTM 모델을 통해 연구를 진행한다. LSTM은 기존의 순환 신경망이 가진 Long-Term Dependency (장기 의존성 문제, 현재 정보로부터 멀리 떨어진 과거 정보를 오랫동안 기억할 수 없다는 한계)를 보완

Received Nov 12, 2024 **Revised** Dec 11, 2024 **Accepted** Dec 23, 2024

†**Corresponding Author** E-mail: jhyoo@sungshin.ac.kr

한 모델로, 과거 정보를 비교적 오랫동안 기억할 수 있다. 따라서 본 연구에서는 과거 정보와 현재 정보를 바탕으로 미래 정보를 예측하는 LSTM이 적합하다.

기존의 GPS 기반 이동 패턴 연구에는 대표적으로 Geofencing, Trajectory Generation Using Generative Adversarial (TrajGAN) (Rao et al. 2020), DeepMove (Zhou & Huang 2018), Spatio-Temporal Residual Network (ST-ResNet) (Zhang et al. 2017), 이 외에 앙상블 (Lee et al. 2023) 및 클러스터링 기법 (Lan & Yoon 2023) 등이 있다. Geofencing은 가상의 경계를 미리 설정하여 그 경계 주변에서 발생하는 단순 위치 이탈을 감지한다. TrajGAN은 Generative Adversarial Networks (GAN)을 통해 GPS 데이터 기반 가상 이동 패턴을 생성해 데이터 부족 문제를 해결하고 다양한 패턴을 학습한다. DeepMove는 LSTM과 어텐션 매커니즘을 결합하여 사용자의 장단기 이동 패턴을 학습 및 예측하는 모델로, 주로 사용자의 GPS 데이터와 사용자 활동 로그, 그리고 기타 보조 데이터를 활용한다. ST-ResNet은 다양한 데이터를 바탕으로 CNN과 잔차 네트워크를 통해 시공간 관계를 분석하며 교통량 예측에 특화되어 있다. 이 외에도 GPS를 클러스터링 하여 이동 패턴을 분석하여 사람의 이동에서 이상을 감지하거나 여러 LSTM 모델을 조합하여 이상 감지 정확도를 높이는 앙상블 기법이 활용된다.

그러나 기존의 연구는 시공간 정보를 통합하여 실시간으로 이상을 감지하거나, 사용자의 이동 패턴 이상을 감지하는 데 한계가 있다. Geofencing은 설정 경계를 기준으로 단순한 위치 이탈을 감지할 수 있지만 복잡한 패턴이나 시간적 정보를 고려하여 위치 이탈을 감지하지 않는다. DeepMove의 경우 장-단기 이동 예측에는 높은 성능을 보이지만 시간과 공간을 모두 고려한 복합적 데이터에 대한 이상 감지에는 제약이 있다. ST-ResNet은 교통 예측에 특화된 모델로 개인 사용자의 이동 패턴을 분석하고 이상을 효과적으로 감지하기 어렵다. 클러스터링과 앙상블 기법을 활용한 연구는 시간에 따른 예측에서 제한적이며, 긴 학습 시간과 높은 계산 비용이 가장 큰 문제점이라 볼 수 있다. 즉, 기존 연구에서는 사용자의 이동 패턴을 시공간 변화에 따라 분석하고 이상을 실시간으로 감지하는 접근이 충분하지 않다.

본 연구에서는 추가적인 식별자를 데이터셋에 포함하지 않고도 여러 사용자의 이동 데이터를 하나로 통합하여 단일 LSTM 모델을 학습시킨다. 이는 사용자의 데이터를 익명성을 유지한 상태에서 모델이 학습 및 추론할 수 있도록 설계되었다는 점에서 기존 연구와 차별화된다. 기존 연구에서는 개별 식별자를 기반으로 데이터를 분류하여 학습하는 경우가 많았다. 예를 들어, Geofencing은 특정 위치로부터 가상의 경계를 설계할 때 사용자별로 특정 위치 정보를 요구하며, TrajGAN과 DeepMove에서는 학습 데이터를 특정 식별자(ID)를 통해 구분하여 모델을 학습한다. 반면, 본 연구는 다수 사용자의 데이터를 하나의 LSTM 모델로 통합하여 학습하고, 이를 통해 일반화된 이상 감지 알고리즘을 제안한다. 이는 사용자가 개별 식별 정보를 노출하지 않고도 단일 모델을 활용하여 이상 감지 서비스를 제공할 수 있는 기반을 마련한다.

또한 이상 감지 알고리즘에서 시간 오차 지표(m_T)를 활용한 점에서 기존 연구와 차별점을 가진다. 기존의 이상 감지 연구는 주로 위치 오차 지표를 기반으로 이상 경로를 감지하거나 단순한 지리적 이탈에 중점을 두었다. 반면, 본 연구에서는 시간과 공간 데이터를 결합하여 '특정 시간에 특정 좌표에 위치해야 할 사용자가 예상 시간과 공간에서 벗어나는 경우'를 이상 경로로 정의한다. 이를 통해 모델은 시간과 공간을 상호 연관적으로 학습하며, 실시간으로 이상 여부를 감지할 수 있는 기능을 갖추게 된다.

아울러 가상 경로와 실제 경로 데이터를 결합하여 다양한 상황에서 발생할 수 있는 경로를 포괄적으로 학습한다. 이러한 접근은 예상된 시간과 위치에 기반한 비정상적인 이동 패턴을 효과적으로 탐지할 수 있으며, 특히 신변보호 대상자와 같은 민감한 상황에 처한 사용자를 위한 실시간 이상 감지 서비스의 가능성을 제시한다.

본 논문은 2장에서 익명의 사용자 이동 패턴 학습과 가상 경로 데이터를 설명하고, 3장에서는 모델 구조 및 학습 데이터 구조를 서술한 뒤, 모델 추론 결과를 기반으로 이상 감지 알고리즘에 대해 정의하였다. 4장은 실험 환경과 결과를 기술하고, 5장에서 결론을 맺는다.

2. 연구 배경

2.1 익명 사용자의 다중 이동 패턴 학습

본 논문에서는 서로 다른 사용자들의 이동 패턴을 효과적으로 학습하기 위해 LSTM 모델이 사용자 식별 정보를 사용하지 않고 경로 데이터를 학습함으로써, 모델이 익명성을 유지한 채로 다양한 경로 패턴을 이해하고 서버에서 실시간으로 다음 위치를 예측할 수 있다는 이점을 보여준다. 익명성은 모델이 개별 사용자의 식별 정보를 사용하지 않고, 사용자별 구분 없이 경로 데이터를 학습하고 추론함으로써 보장된다. 즉 예측할 때 하나의 모델이 모든 경로를 통합한 데이터셋을 학습하며, 사용자별 개별 모델이 존재하지 않기 때문에 모델이 특정 사용자의 경로를 구별할 수 없다. 다만, 하나의 이동 패턴을 구분하여 데이터를 처리해야 하기 때문에 실제 경로 데이터 수집 시에는 익명으로 수집되지 않지만, 학습 및 예측 단계에서는 익명성을 유지한다.

이상 감지라는 최종 목표를 위해, 본 연구는 실제 사용자의 경로 데이터와 가상 경로 데이터를 통합하여 하나의 LSTM 모델이 학습하는 방식을 사용한다. 이러한 방식은 모델이 다양한 경로 패턴을 학습하고, 이상을 감지할 수 있도록 한다. 다양한 사용자 패턴을 하나의 모델로 통합 학습하여 실환경에서 발생할 수 있는 경로 이탈 상황에 대비할 수 있다. 따라서 본 연구에서는 실제 및 가상 경로를 학습한 LSTM 모델의 추론 오차를 분석하여 이상 감지 알고리즘을 설계하고 그 효과성을 평가한다.

2.2 가상 경로 데이터

가상 GPS 데이터는 네이버 클라우드 플랫폼의 Map application programming interface (API)를 통해 수집되었다. 출발지와 목적지, 출발 시간과 요일을 입력하면 운전 경로를 따라 GPS 위치와 시간이 출력된다. 각 GPS 위치 데이터는 4~6초 간격으로 수

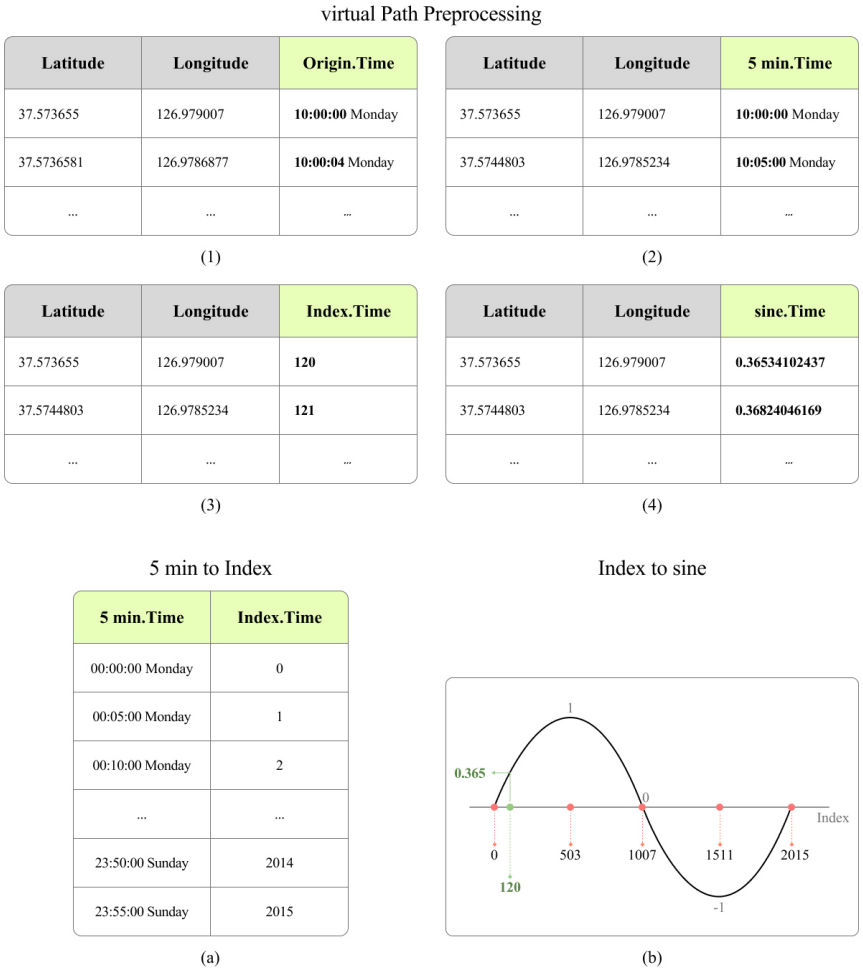


Fig. 1. Input data preprocessing. (1) represents a sample trajectory dataset collected using the Map API. The time column in this dataset is recorded at approximately 4-second intervals based on vehicle movement patterns. (2) transforms 15 rows of data collected at 4-second intervals into a single (latitude, longitude, time) data point, with the time redefined in 5-minute intervals. For example, the second row in (1), with a time value of '10:00:04 Monday,' is transformed into '10:05:00 Monday' in (2). (3) assigns indices to the data based on a weekly schedule. As shown in (a), '00:00:00 Monday' is indexed as 0, '00:05:00 Monday' as 1, and '23:55:00 Sunday' as 2015. (4) applies a sine function to the indices defined in (3). This transformation process is visually represented in (b)'s sine graph. The X-axis of the graph represents the indices. For instance, indices 0, 1007, and 2015 have a sine value of 0, index 503 has a sine value of 1, and index 1511 has a sine value of -1. Additionally, as an example from (3) and (4), applying the sine function to Index 120 results in a sine value of approximately 0.365341.

집되었으며, 파일당 약 10분에서 3시간(row가 약 200~3600개 사이)의 정보를 포함한다. 이렇게 수집된 raw 데이터는 위도 • 경도 • 시간을 컬럼으로 둔 2차원 행렬이다.

동일 궤적에 대한 데이터라 하더라도 실제 이동 시에는 위치와 시간에 미세한 차이가 발생한다. 이러한 차이에 대해 학습 모델은 구조(layer, unit size에 따른 복잡도)에 따라 민감하게 반응할 수 있다. 따라서 작은 오차로 인한 영향을 최소화하기 위해 학습 데이터의 가상 경로에 오차(noise)를 추가한다. 위도와 경도에 해당하는 각 행마다 (−0.000001, 0.000001) 범위의 임의 값이 더해지고, 시간에는 (−5, 5)분 범위의 값이 무작위로 더해진다. 위치 데이터에 더해지는 (−0.000001, 0.000001) 값은 위도 기준으로 약 ±0.111 m, 경도 기준으로 약 ±0.088 m 범위의 오차를 줄 수 있다.

4~6초마다 기록된 raw 경로 데이터를 5분 단위로 변환하여 저장한다. 시간 데이터는 일주일을 5분 간격으로 나눈 2,016개의 인덱스로 표현되며, 이 인덱스를 sine 값으로 변환해 사용한다. 예를 들어 '00시 00분 월요일'에 해당하는 시간 데이터는 인덱스 '0'으로, sine 값 '0.0'이 저장되고, '00시 05분 화요일'에 해당하는 시간 데이터는 인덱스 '288'으로, sine 값 '0.7818'이 저장된다. Fig. 1에서 위 과정을 보여 준다.

시간 데이터를 인덱싱한 후 sine 함수 값으로 변환한 것은 모델의 추론 오류를 줄이기 위함이다. 인덱스 값을 그대로 사용하면 처음(index 0)과 끝(index 2015) 사이의 차이가 커져 오류가 발생할 수 있지만 sine 함수를 적용하면 이러한 값들이 주기적으로 연결되어 시계열 데이터의 연속성을 효과적으로 반영할 수 있다. 이는 transformer 모델이 모든 입력 데이터를 동시에 병렬로 처리하면서도 시퀀스의 순서를 인식할 수 있도록 하는 positional encoding의 sine 함수 활용 방식에서 착안하였다.

실제 데이터의 경우 스마트 워치에서 5분 단위로 GPS와 관측 시간을 수집하도록 설정했기 때문에 가상 데이터의 인덱스 과정부터 동일한 전처리가 적용된다.

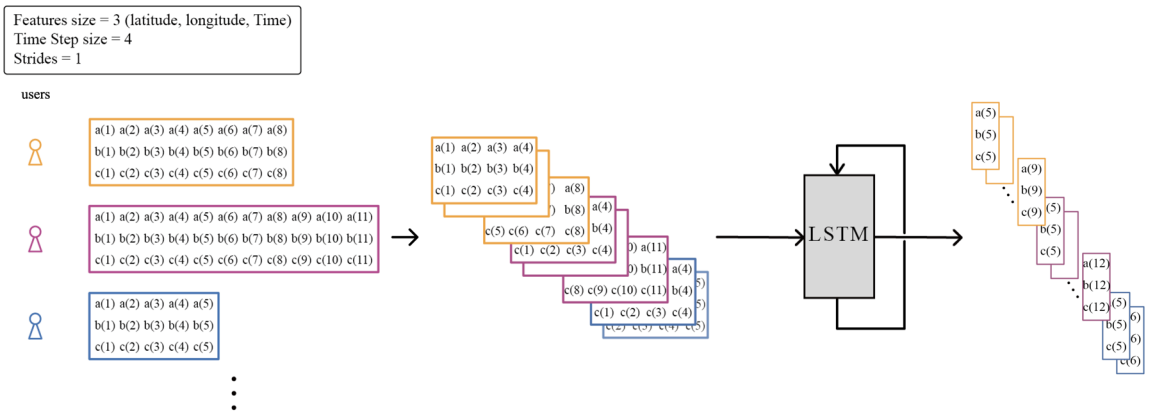


Fig. 2. The overall training process of the LSTM model. The movement trajectory data of each individual is represented by three features: latitude (a), longitude (b), and time (c). The feature size of each data point is set to 3, as indicated in the top-left corner of the figure. All data are divided into overlapping time windows, with a time step size of 4. The stride is set to 1, meaning the sliding window moves forward one step at a time, generating overlapping subsequences. For example, a user's data is divided into overlapping sequences such as [a(1), a(2), a(3), a(4)] and [a(2), a(3), a(4), a(5)] in the Latitude value. This windowing process is applied to the data of all individuals. Each time window, consisting of a sequence of 4 data points, serves as input to the LSTM model. Sequences generated from different individuals are combined into batches for training. The LSTM model processes each input sequence to predict the features (latitude, longitude, and time) of the next time step. For example, given the input sequence [a(1), a(2), a(3), a(4)] in the Latitude value, the model predicts [a(5)]. In summary, the model is trained to predict the next data point (e.g., [a(5)]) based on the input sequence (e.g., [a(1), a(2), a(3), a(4)]). The training process is performed iteratively over all sequences in the dataset, with the model's weights adjusted to minimize prediction errors.

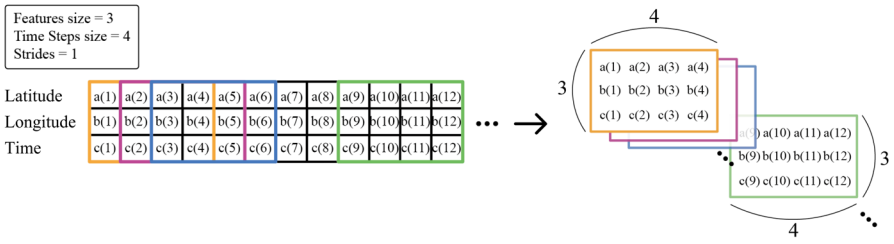


Fig. 3. Input data structure. When the first data chunk containing the position and time information for (1) to (4) is input into the training model, the position and time information corresponding to (5) is output. The second data chunk contains information from (2) to (5), and the trained model predicts the information for (6). In other words, when the information from (n - 4) to (n) is input into the training model, it outputs the information for (n + 1) where $n \in N$, $n>4$.

3. 본론

3.1 학습 데이터 입출력 구조

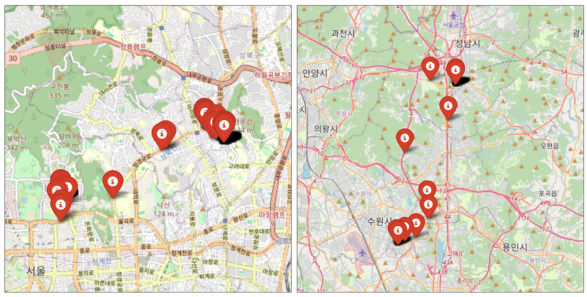
기본적으로 LSTM은 3차원 형식의 입력 데이터를 요구한다. 입력 데이터의 shape는 [Samples size, Time steps size, Features size]와 같이 3차원 행렬로 reshape 되어야 한다. 이에 따라 원본 2차원 데이터 셋에서 3차원으로 reshape한 GPS 경로 데이터의 samples size는 데이터의 전체 길이(1분 간격으로 관측한 모든 경로의 길이)이며, time steps size(4)는 LSTM layer에 들어가는 데이터의 개수이다. 여기서 Features size(3)는 데이터 셋이 위도, 경도, 시간 3개의 컬럼으로 이루어져 있기 때문에 3으로 설정된다. LSTM 모델의 전체 학습 과정은 Fig. 2를 통해 확인할 수 있다.

Time steps은 모델에 한 번 데이터 덩어리가 입력될 때, 몇 개의 데이터 덩어리를 포함할지를 나타낸다. 다시 말해, Fig. 3에서 (3 x 1) 크기의 데이터(위도, 경도, 시간)가 입력되는 것이 아니라, (3 x 4) 크기의 데이터가 하나의 세트로 묶여 모델에 입력된다. 이는 모델에 입력되는 하나의 데이터 덩어리가 20분 분량(5분 x 4)의 정보를 포함하고 있음을 의미한다. Fig. 3에서 (1) ~ (4)의 위치 및 시간 정보가 포함된 첫 번째 데이터 덩어리가 학습 모델에 입력되면 (5)에 해당하는 위치 및 시간 정보가 출력된다. 두 번째 데이터 덩어리는 (2) ~ (5)의 정보를 포함하며, 훈련 모델은 (6)에 해당하는 정보를 추론한다. 즉 (n – 4) ~ (n)의 정보가 훈련 모델에 입력되면, (n+1)의 정보를 출력한다 ($n \in N$, $n>4$).

3.2 실제 및 가상 경로 데이터의 구조

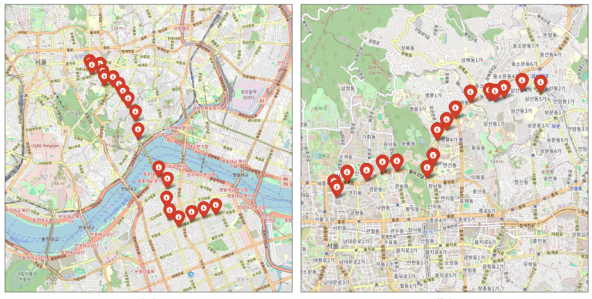
Fig. 4는 실제 사용자들의 GPS 데이터를 시각화한 지도 일부이다. Fig. 4 이외에도, 실제 경로 데이터는 서울시를 포함한 경기도 내에서 수집된 실제 사용자 이동 정보를 기반으로 한다. 본 데이터는 5분 간격으로 기록된 위치와 시간 정보를 포함하며, 학생 또는 직장인 등 특정 목적지를 반복적으로 오가는 사용자의 일상적인 이동 패턴을 반영한다. 실제 raw 경로 데이터에는 위도 및 경도 정보와 시간 정보(날짜 및 시간)가 저장되며, 이러한 실제 사용자의 전체 궤적을 포함한 학습 데이터셋은 LSTM 모델의 학습에 필요한 중요 정보들을 제공한다.

https://doi.org/10.11003/JPNT.2025.14.1.1



(a) (b)

Fig. 4. Real path maps. (a) and (b) represent the movement information of actual users within the Gyeonggi-do area displayed on a map. The sections in (a) and (b) where coordinates are missing correspond to instances where GPS data was not collected due to signal limitations, such as when the user was traveling through a subway or tunnel.



(a) (b)

Fig. 5. Virtual path maps. (a) represents the movement information of virtual users within the Seoul area, and (b) represents the movement information within the Gyeonggi-do area, both displayed on a map and illustrating the routes collected using the Naver Cloud Platform Map API.

가상 경로는 네이버 클라우드 플랫폼의 Map API를 사용하여 서울시와 경기도 범위 내 궤적을 수집하였다. 다만 해당 API에서는 오로지 차량의 이동 데이터를 기반으로 GPS 데이터를 출력하기에 가상 데이터와 실제 데이터는 경로의 형태와 이동 속도에서 서로 차이가 나타날 수 있다. 2.2절에서 언급한 바와 같이 가상 GPS 데이터는 4~6초 간격으로 수집된 차량 이동 데이터를 가공하여 학습에 활용한다. 차량과 보행자의 속도 차이를 고려하여 차량 기준 1분 동안 이동한 거리는 사용자가 약 5분 동안 이동하는 거리로 가정한다. 이를 기반으로 약 4초 간격으로 수집된 15개의 GPS 데이터를 하나의 GPS 데이터로 통합하여 학습과 테스트에 이용하였다. 가상 raw 경로 데이터 역시 위치와 시간 정보를 저장하고 있으며, 실제 경로 데이터에서 부족한 학습 데이터 셋을 보완한다. Fig. 5는 서울시와 경기도권 내의 가상 경로 데이터를 시각화한 지도이다.

위 과정을 통해 수집된 실제 경로와 가상 경로는 LSTM 모델의 학습 데이터 셋으로 사용된다. 가상 경로는 더 다양한 상황에서 LSTM 모델을 학습할 수 있도록 돕는다. 즉 가상 경로를 통해 실제 상황에서 발생할 수 없는 시나리오를 추가함으로써 모델이 예외적인 상황에도 잘 대응할 수 있도록 학습 데이터를 제공한다.

이러한 실제 경로 데이터와 가상 경로 데이터의 통합을 바탕으로 모델의 추론 가능한 상황을 확장했으며, 이를 통해 보다 신뢰성 있는 예측 결과를 도출할 수 있다.

Table 1. LSTM architecture.

Layer (type)	Shape	
	Input	Output
Input (Input layer)	(None, 4, 3)	(None, 4, 3)
LSTM_0 (LSTM)	(None, 4, 3)	(None, 4, 128)
LSTM_1 (LSTM)	(None, 4, 128)	(None, 4, 64)
LSTM_2 (LSTM)	(None, 4, 64)	(None, 4, 32)
Dense_0 (Dense)	(None, 4, 32)	(None, 4, 16)
Dense_1 (Dense)	(None, 16)	(None, 3)

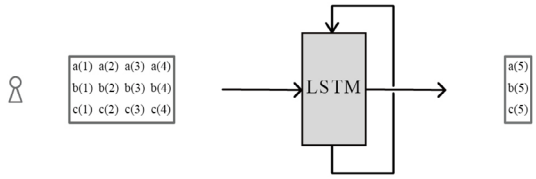


Fig. 6. Testing process of the trained LSTM model. When 20 minutes of data is received from an anonymous user, the trained model predicts the next location and time. During the testing process, an input sequence consisting of four consecutive data points, including latitude (a), longitude (b), and time (c), is provided to the model. This sequence represents 20 minutes of GPS data, with each data point assumed to be collected at 5-minute intervals. The input sequence is passed to the trained LSTM model, which processes it to predict the data for the next time step. Based on the four previous data points, the LSTM model infers the user's latitude, longitude, and time at the next predicted location.

3.3 모델 구조 및 테스트 데이터 학습

Table 1의 LSTM 모델의 구조는 학습 데이터셋에 맞춰 다운 사이징한 모델로, 1개의 input layer와 3개의 LSTM Layer, 2개의 dense layer로, 총 6개의 layer로 구성된다. 설정한 LSTM layer의 unit의 수는 Table 1의 output의 마지막 차원 수와 같다. Dense_0에서는 RELU 활성화 함수를 사용하였다.

학습 모델에 들어갈 테스트 데이터는 익명 사용자의 경로 데이터(위도, 경로, 시간)로, 5분 간격으로 수집한 사용자의 위치 및 시간 데이터 4개를 모델에 입력하여 그 다음 5분 후의 경로 데이터를 추론한다. Fig. 6에서 (1) ~ (4)의 정보가 학습 모델에 입력되면 (5)의 추론 결과가 출력된다.

3.4 이상 감지 알고리즘

이상 감지 알고리즘을 설계하기 위해, 오차 지표와 관련된 변수들을 활용하여 감지 기준을 정의할 수 있다. 이 과정에서 각 지표의 누적 값(m)을 공식화하고, 임계치를 설정해 이상 여부를 판별할 수 있도록 한다.

오차 지표는 시간 오차를 통해 다음과 같은 시간 변수를 정의할 수 있다. T_{avg}는 각 지점에서의 평균적인 시간 오차를 나타내는 평균 오차 시간이며, T_{max}는 최대 오차 시간이다. 시간 누적 오차 m_T는 평균 오차 시간(T_{avg})과 최대 오차 시간(T_{max})의 합으로 표현되며, 식 (1)과 같다.

$$m_T = T_{avg} + T_{max} \tag{1}$$

시간 오차는 추론 시간과 실제 시간 간의 차이를 계산하여 산

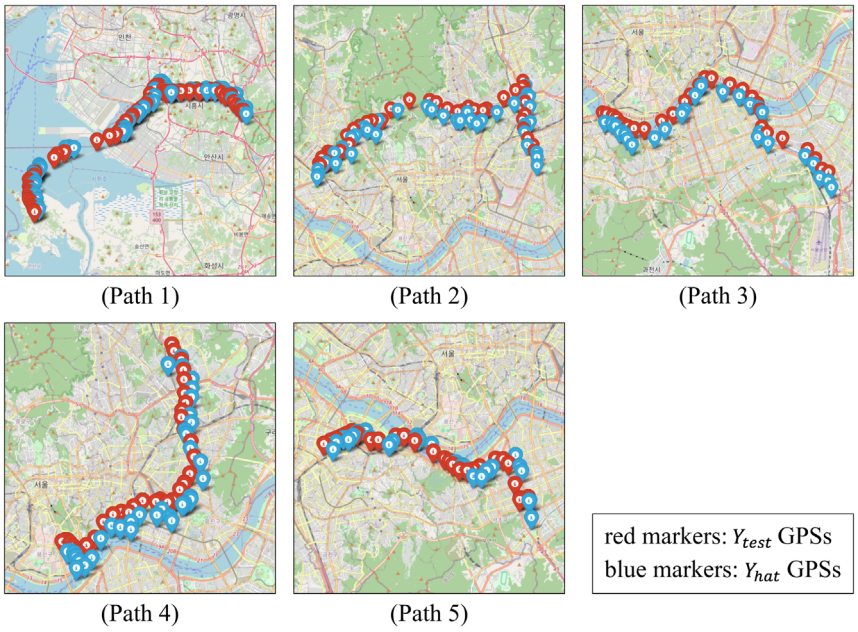


Fig. 7. Maps of Path 1 to 5 in Table 2. The red map markers represent the GPS coordinates of the actual observations (y), while the blue map markers indicate the GPS coordinates of the predicted values ($\hat{y}$).

출되며, 이를 통해 평균 값과 최대 값을 구해 경로의 정확도를 평가한다. 식 (1)의 누적 오차를 활용해 경로의 이상 여부를 감지할 수 있다. 이후, 누적 오차 값이 설정된 임계치를 초과하는지 여부를 기준으로 이상 경로를 판단한다. 즉 m_T 값이 임계치를 넘으면 해당 경로를 이상이라 판단한다. 이는 서론에서 정의한 이상 경로에 해당하며, 사용자가 특정 시간대에 원래 이동 경로에 있지 않음을 나타낸다.

이와 같은 방식으로 이상 감지 알고리즘은 LSTM 모델의 추론 결과인 $\hat{y}$과 실제 측정치인 y 사이의 시간 오차 지표 m_T를 계산하여 익명 사용자의 경로에서 이상을 감지한다. 여기서 $\hat{y}$은 5분 간격으로 수집된 (위도, 경도, 시간) 데이터 4개를 입력 받은 LSTM이 추론한 그 다음 (위도, 경도, 시간) 데이터 1개를 의미하며, y는 실제 관측된 그 다음 (위도, 경도, 시간) 데이터 1개를 의미한다.

$\hat{y}$과 y는 원래 위도, 경도, 시간 정보를 포함하는 데이터지만, 이상 감지 알고리즘에서는 이들 데이터의 시간 정보만을 사용하여 m_T를 계산하고 임계치와 비교한다. 따라서 m_T는 식 (1)을 통해 $\hat{y}$의 시간 정보와 y의 시간 정보 사이의 오차를 나타내는 지표로 정의된다. 이후, 추론 경로와 실제 경로 간의 시간 차이를 통해 산출된 m_T 값이 임계치를 초과할 경우 이를 이상으로 판단한다.

4. 실험 결과

4.1 가상 경로 추론 결과

Tables 2와 3은 제안한 추론 모델을 사용하여 가상 경로 Path 1~5의 위치 및 시간 추론 오차 비교 결과를 보여준다. Path 1~5의 GPS 지도는 Fig. 7과 같다. 다양한 metric을 통한 각 경로에 대한 위치 및 시간 오차를 바탕으로 LSTM 모델의 성능을 종합적으로 분석할 수 있다.

Table 2. Distance error results for the virtual path.

Metric	Path 1	Path 2	Path 3	Path 4	Path 5
Total distance (km)	36.06	20.32	18.7	16.38	12.59
Distance MSE (km^2)	0.72	0.66	0.80	0.61	0.36
Average error distance (km)	0.70	0.76	0.83	0.74	0.55
Max error distance (km)	3.05	1.41	1.85	1.19	1.32

Table 3. Time error results for the virtual path.

Metric	Path 1	Path 2	Path 3	Path 4	Path 5
Time MSE (25 minute)	1.47	26.89	8.54	2.97	21.29
Average error time (hours)	0.09	0.43	0.24	0.13	0.38
Max error time (hours)	0.17	0.58	0.33	0.25	0.50

Table 2의 거리 오차 결과를 보면, Path 1는 Total Distance가 36.06 km로 길며, 다른 경로에 비해 위치 오차도 높은 편이다. 반면, 짧은 경로인 Path 5(12.59 km)는 Average Error Distance가 상대적으로 가장 작다. 전반적으로 Total Distance가 길수록 위치 오차가 증가하는 경향을 보여준다. 특히, Distance MSE는 경로에 따라 0.36 km^{2}에서 0.72 km^2 사이로 분포한다. Distance MSE는 모델이 추론한 GPS 좌표와 실제 GPS 좌표 간의 거리 오차를 제곱한 값의 평균으로, 두 지점 간의 거리는 Haversine formula를 이용해 계산된다. Average Error Distance는 0.55 km에서 0.76 km로 비교적 안정적으로 유지되며, Max Error Distance는 Path 1에서 3.05 km로 가장 높고, Path 5에서 1.32 km로 가장 낮다.

Table 3의 시간 오차를 살펴보면, Path 2과 Path 5은 서로 다른 이동거리를 가졌음에도 불구하고 Time MSE와 Average Error Time이 높게 나타났다. Path 2의 Time MSE는 26.89, 평균 시간 오차는 0.43시간으로, 이는 가상 경로에서 시간 예측이 이동거리와 관련이 없고 복잡성이나 경로의 특성에 더 큰 영향을 받을 수

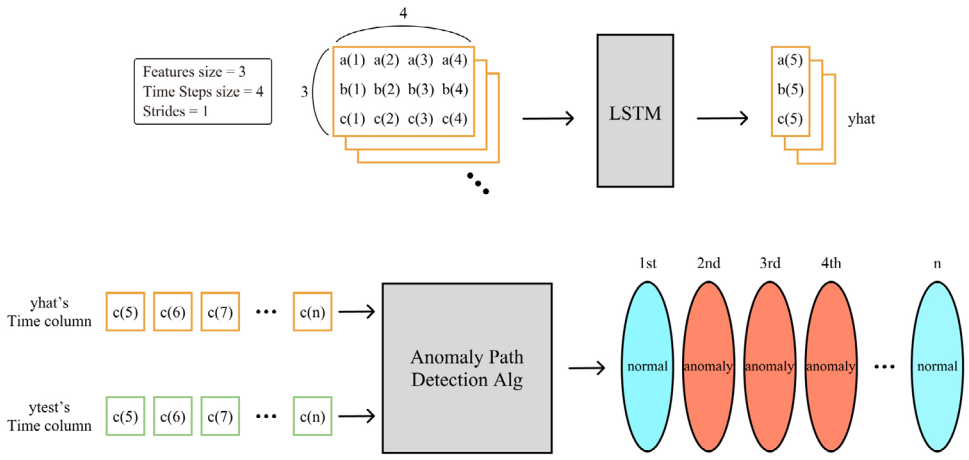


Fig. 8. The overall anomaly detection algorithm process. When Y_{hat} is produced by the trained model, m_T is calculated using the actual results Y_{test}, and m_T is computed for each interval. If the value exceeds the threshold, it is considered an anomaly. The anomaly detection algorithm calculates the time error metric (m_T) by comparing the predicted results ($\hat{y}$) from the LSTM model with the actual observed data (y). This comparison is performed using the time column (c) from both datasets, and the cumulative time error is computed for each interval. The time error metric (m_T) is calculated for predefined intervals (e.g., 10 minutes, 30 minutes, or 1 hour) and is evaluated against a predetermined threshold to determine anomalies. If the m_T value for a specific interval exceeds the threshold, that interval is classified as an 'anomaly'. As depicted in the figure, the anomaly detection algorithm analyzes the overall trajectory data by evaluating the m_T value for each interval, providing a basis for identifying abnormal segments within the trajectory.

Table 4. Path scenario.

Scenario	Distance	Time	Description	Path
S1	Normal	Normal	Normal paths	s1_01 s1_02 s1_03 s1_04
S2	Normal	Anomaly	Anomaly paths	s2_01 s2_02 s2_03 s2_04

있음을 시사한다. 여기서 Time MSE는 예측된 시간과 실제 시간의 차이를 제곱한 값의 평균으로, 시간 오차의 전반적인 분포와 크기를 수치화 한다. Average Error Time은 Path에 따라 0.09시간에서 0.43시간 사이, Max Error Time은 Path 1에서 0.17시간, Path 5에서 0.50시간으로 나타났다.

이를 통해, 가상 경로에서 LSTM 모델의 거리 추론에서 비교적 일관성을 보이지만, 시간 추론에서는 경로 특성에 따라 성능이 변동할 수 있다. 추가적인 데이터 분석과 모델 최적화를 통해 이러한 예측 성능을 더욱 향상시키기 위한 연구가 필요하다.

4.2 이상 감지 알고리즘 시나리오

학습 모델의 추론 결과에서 위치 및 시간 오차의 양상을 분석하기 위해 테스트 데이터를 정상과 이상 시나리오로 분류하고, 각 시나리오에 이상 감지 알고리즘을 적용하여 그 결과를 비교하고자 한다. Table 4에서 정상 시나리오는 기존 학습 데이터셋의 경로에 기반하며, 2.2절과 같이 위치 및 시간 정보에 noise를 추가해 학습 데이터와 완전히 동일하지 않은 새로운 정상 경로(S1)를 생성한다. 이상 시나리오 S2는 시간적으로 이상이 있는 데이터로, 정상 시나리오와 동일한 방식으로 noise를 추가하여 이상 경로를 수집한다.

이상 감지 알고리즘은 3.4절에서 정리한 바와 같다. 시나리오별로 학습 모델의 추정치 $\hat{y}$을 이상 감지 알고리즘에 입력해 N분마다 m_T를 계산하고, 각 지표가 임계치를 초과하는지에 따라 경로 이상 여부를 판단한다. 예를 들어 1시간 마다 이상을 감지한다고 설정한다면, 12개의 추정치 $\hat{y}$과 12개의 실제 측정치 y를 모아 각각 1시간 데이터로 구성한 뒤, 알고리즘을 적용한다. Fig. 8은 전반적인 알고리즘 적용 과정을 나타낸다.

4.3 이상 감지 알고리즘 실험 결과

Table 4의 S1은 위치와 시간 모두 정상인 가상 경로이며, S2는 시간에 이상이 있는 가상 경로이다. 본 실험에서는 이상이 없는 학습 데이터셋과 테스트 데이터셋에 대한 추론 결과에 오차 지표 m_T를 계산하고, 이를 바탕으로 이상과 정상을 감지할 수 있는 임계치를 0.0091로 설정하였다. 임계치 0.0091은 정상 데이터에서 산출한 m_T와 비정상 데이터의 m_T를 실험적으로 비교하여 도출하였다.

Fig. 9에서 y축은 시간 오차 지표인 m_T를 나타내며, x축은 10분 간격으로 설정된 interval을 의미한다. 각 점은 10분 동안의 데이터를 기반으로 계산된 누적 값(m)이며, 이는 시간 오차 지표 m_T와 의미가 같다. 그래프에서 노란색 점선은 임계치를 나타내며, 초록색 꺾은선은 각 interval에서 계산된 경로 데이터의 m_T 값을 순차적으로 연결한 것이다.

Figs. 9a–d는 정상 경로 시나리오(S1)에 해당하며, 이들의 y축 범위는 약 0.001~0.009이다. 이 중에서 Fig. 9a를 살펴보면, interval 0에서 m_T 값이 약 0.0063으로 나타나며 최소 m_T는 약 0.0047, 최대 m_T는 0.007이다. 정상 경로에 대한 이상 감지 알고리즘 결과 그래프는 임계치보다 낮은 m_T 값을 유지한다. 반면, Figs. 9e–h는 서론에서 정의한 이상 경로 시나리오(S2)에 해당한다. 이 그래프의 y축 범위는 0.007에서 0.011로 설정하였다. x

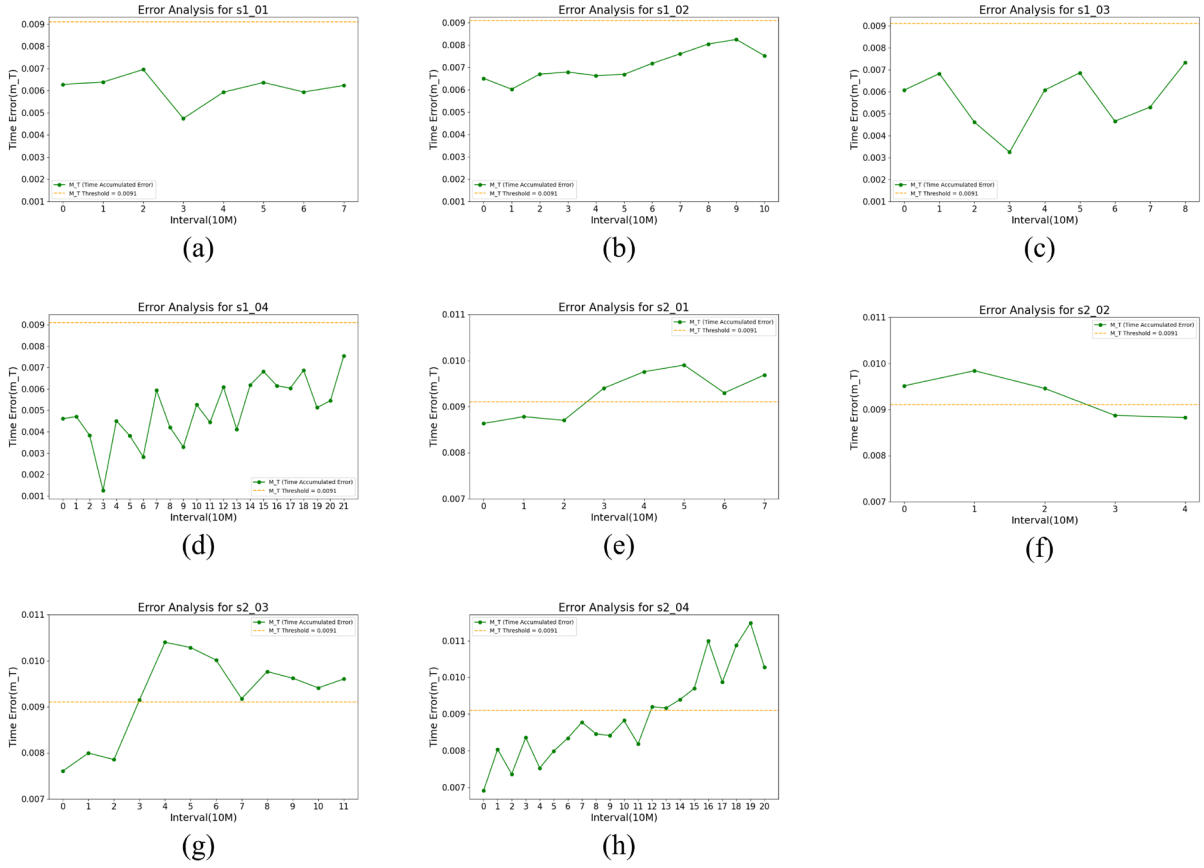


Fig. 9. Anomaly detection algorithm results. (a) to (d) graphs showing the Interval-based m_T values calculated by applying the anomaly detection algorithm to the predicted results Y_{hat} and actual results Y_{test} of the normal path, while (e) to (h) represent graphs based on the predicted and actual results of the anomalous path. It shows graphs where the y-axis represents the time error metric (m_T) and the x-axis represents intervals set at 10-minute intervals. This reflects the cumulative value (m) calculated based on 10 minutes of data, which corresponds to m_T The graphs depict the threshold (0.0091) as a yellow dashed line and the m_T values calculated for each interval as a green polyline. (a) to (d) represent normal path data, with the y-axis range set between 0.001 and 0.009. For instance, in graph (a), the m_T value at interval 0 is approximately 0.0063, with a minimum m_T value of around 0.0047 and a maximum value of approximately 0.007. (e) to (h) correspond to anomalous path data, where the y-axis range is set between 0.007 and 0.011. In graph (e), the m_T value does not exceed the threshold until interval 3, after which it surpasses it. Here, the minimum m_T value is approximately 0.0086, and the maximum value is about 0.0099. Additionally, graphs like (f) show cases where m_T exceeds the threshold from interval 0.

축의 interval 간격은 정상 경로 그래프와 동일하다. Fig. 9e를 보면 interval 2까지 m_T 값이 임계치보다 낮은 값을 유지하다가 interval 3부터 초과한다. Fig. 9e의 최소 m_T는 0.0086이며, 최대 m_T는 0.0099이다. 일반적으로 이상 경로는 Fig. 9e와 같이 점진적으로 m_T가 증가하는 양상을 보이지만 Fig. 9f처럼 초기 interval부터 임계치를 초과하는 경우도 존재한다.

5. 논의

본 연구에는 몇 가지 한계가 존재한다. 첫째, 본 연구에서 제안한 이상 감지 알고리즘은 시간 오차 지표를 활용하여 이상 여부를 판단하였으며, 이를 통해 사용자가 특정 시간대에 예상된 경로를 이탈하는 상황을 감지할 수 있다. 시간 정보를 제외한 위치 오차 지표는 정상 및 이상 경로를 판단하는데 한계가 있었는데, 이는 다중 경로를 학습한 LSTM의 일반화 능력이 학습데이터에 존재하지 않았던 새로운 경로 패턴도 추정을 하였기 때문이다. 둘째, 학습 데이터셋에 실제 사용자의 데이터를 충분히 포함하지 못해 모델의 실효성을 실환경에서 평가하는 데 한계가 있다. 이를 극복하기 위해, 모델의 학습 및 평가 단계에서 가상 데이터를 사용하여 다양한 경로 시나리오를 반영하려 노력하였다. 그러나 실제 환경의 복잡성을 온전히 반영하기 위해 추가적인 실제 데이터 수집과 검증이 필요하다.

본 연구는 기존의 개별 사용자 기반 모델링과 달리, 익명의 다중 경로 데이터를 하나의 LSTM 모델로 통합하여 학습한 점에서 의의가 있다. 이는 사용자 식별 정보를 요구하지 않고도 다양한 경로 패턴을 학습할 수 있도록 설계된 네트워크를 구축한 것으로, 익명성과 데이터 통합의 균형을 맞춘 접근법을 제시하였다. 이러한 방식은 신변보호와 같은 실시간 이상 감지 서비스에서 효율적인 데이터 처리와 확장 가능성을 제공할 수 있다.

향후 연구에서는 실제 사용자 데이터를 추가로 수집하여 학습 데이터셋의 다양성과 신뢰성을 높이는 동시에, 위치 및 시간 데이터를 통합적으로 활용할 수 있는 알고리즘 개선이 필요하다. 또한, 경로의 복잡성과 특성에 따라 변동하는 시간 예측 성능을

개선하기 위한 최적화된 모델 개발이 이루어진다면, 본 연구 결과의 실효성과 일반화를 강화할 수 있을 것이다.

6. 결론

기존 GPS 기반 이동 패턴 분석 및 이상 감지 연구는 시공간 정보를 통합하여 실시간 이상 감지나 개인 사용자의 이동 패턴 이상 감지에 있어 부족함이 있었다. 본 논문에서는 익명의 다중 경로 데이터를 통합하여 하나의 LSTM 모델로 전체 이동 패턴을 분석하고, 이를 기반으로 이상을 감지하는 알고리즘을 제안하였다. 이상 감지 알고리즘 적용 결과, 정상 경로의 경우 설정된 임계치 이하로 예측 오차를 유지하여 안정성을 보였으며, 이상 경로에서는 시간 경과에 따라 오차가 점차 증가하여 임계치를 초과하는 경향을 보였다. 향후 연구에서는 실제 사용자의 GPS 수집을 통해 제안한 모델의 실효성을 분석할 것이다.

ACKNOWLEDGMENTS

This work was supported by Protection Technology for Socially vulnerable individuals Program (www.kipot.or.kr) funded by Korean National Police Agency (KNPA, Korea) [Project Name: Development of an Integrated Control Platform for Location Tracking of Crime Victim based on Low-Power Hybrid Positioning and Proximity Search Technology/Project Number: RS-2023-00236101].

AUTHOR CONTRIBUTIONS

Conceptualization, J.Y.; methodology, J.K.; software, J.K.; validation, J.K.; formal analysis, J.Y.; investigation, J.Y.; experiment, J.K.; writing—original draft preparation, J.K; writing—review and editing, J.Y.; funding acquisition, J.Y.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Hwang, C.-H. & Shin, K.-W. 2020, CNN-LSTM Combination Method for Improving Particular Matter Contamination (PM2.5) Prediction Accuracy, JKIICE, 24, 57-64. https://doi.org/10.6109/jkiice.2020.24.1.57

Ji, Y., Wang, L., Wu, W., Shao, H., & Feng, Y. 2020, A Method for LSTM-based Trajectory Modeling and Abnormal Trajectory Detection, IEEE Access, 8, 104063-104073. https://doi.org/10.1109/ACCESS.2020.2997967

Kim, T. H., Song, M. J., Choi, E. J., Kim, B. S., & Moon, Y. H. 2021, Flight data prediction method using LSTM based-deep learning model, In Proceedings of the Fall Conference of KSAS, Seoul, 17-19 November 2021, pp.968-969.

Lan, D. T. & Yoon, S. 2023, Trajectory Clustering-Based Anomaly Detection in Indoor Human Movement, Sensors, 23, 3318. https://doi.org/10.3390/s23063318

Lee, G., Yoon, Y., & Lee, K. 2023, Anomaly Detection Using an Ensemble of Multi-Point LSTMs, Entropy, 25, 1480. https://doi.org/10.3390/e25111480

Nam, G., Kim, J., Min, D., & Lee, J. 2023, Along-Track Position Error Bound Estimation using Kalman Filter-Based RAIM for UAV Geofencing, JPNT, 12, 51-58. https://doi.org/10.11003/JPNT.2023.12.1.51

Rao, J., Gao, S., Kang, Y., & Huang, Q. 2020, LSTM-TrajGAN: A Deep Learning Approach to Trajectory Privacy Protection, GIScience, Poznań, Poland, 27-30 September 2021, 177, 12:1-12:17. https://doi.org/10.4230/LIPIcs.GIScience.2021.I.12

Shevchenko, Y. & Reips, U. D. 2024, Geofencing in location-based behavioral research: Methodology, challenges, and implementation, Behavior Research Methods, 56, 6411-6439. https://doi.org/10.3758/s13428-023-02213-2

Shin, Y., Lee, C., Jung, D., & Kim, E. 2024, Long Short-Term Memory Network for INS Positioning During GNSS Outages: A Preliminary Study on Simple Trajectories, JPNT, 13, 137-147. https://doi.org/10.11003/JPNT.2024.13.2.137

Yoon, S.-W., Lee, W.-H., & Lee, K.-C. 2022, Pedestrian GPS Trajectory Prediction Deep Learning Model and Method, JKSCI, 27, 61-68. https://doi.org/10.9708/jksci.2022.27.08.061

Zhang, J., Zheng, Y., & Qi, D. 2017, Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction, Proceedings of the AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4-9 February 2017, 31. https://doi.org/10.1609/aaai.v31i1.10735

Zhou, Y. & Huang, Y. 2018, DeepMove: Learning Place Representations through Large Scale Movement Data, IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 10-13 December 2018. https://doi.org/10.1109/BigData.2018.8622444

Jihyoung Kim is under B.S. degree in the School of AI Convergence, Sungshin Women's University. Her research interests are machine learning and anomaly detection.

Jaehyun Yoo received the Ph.D. degrees in the School of Mechanical and Aerospace Engineering, Seoul National University, Seoul, in 2016. He was a postdoctoral researcher at the School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Stockholm, Sweden. He is currently a Professor at the School of AI, Sungshin Women's University. His research interests include machine learning, indoor localization, automatic control, and robotic systems.

https://doi.org/10.11003/JPNT.2025.14.1.1