

# 딥러닝 모델 기반 AIS 데이터를 활용한 선박 경로 예측 연구

이원희, 윤승원, 장다현, 이규철(교신저자)

충남대학교 컴퓨터공학과

k334lwh@gmail.com, yoonenoch11@gmail.com, dahyunjang@gmail.com, kcleec@cnu.ac.kr

## Vessel Trajectory Prediction Study Using AIS Data Based On Deep Learning Model

Won-Hee Lee, Seung-Won Yoon, Da-Hyun Jang, Kyu-Chul Lee(corresponding author)

Chungnam Univ.

### 요약

해상 운송의 대부분을 차지하고 있는 선박의 경로를 예측하는 연구는 해상의 위험을 사전에 탐지하여 사고를 예방할 수 있다. 도로와 달리 해상에는 신호체계가 따로 존재하지 않고, 교통 관리가 어렵기에 해상 안정성을 위해 선박 경로 예측은 필수적이다. 선박의 경로를 예측하기 위하여 본 연구는 선박 경로 예측 딥러닝 모델을 생성하였다. 본 연구의 모델은 선박의 실시간 경로 정보를 담고 있는 AIS 데이터를 통해 선박의 이동패턴을 파악하여 이후에 위치할 선박의 gps 좌표를 예측하는 모델이다. 선박 경로 데이터는 시계열 데이터이므로 시계열 데이터 예측에 적합한 LSTM 딥러닝 모델을 개발하였다. 본 논문은 선박 경로 예측에 적합한 딥러닝 모델과 최적의 파라미터를 제시한다.

### I. 서론

선박을 이용한 해상 운송은 세계 무역의 최대 90%를 차지하고 있는 중요한 운송수단이다. 많은 화물 운송이 선박을 통해 이루어지고, 선박 관련 기술이 점차 발전하면서 해상에 선박의 밀도가 증가하고 해상 교통이 복잡해졌다[1]. 이로 인해 해상에서의 교통사고가 빈번해지고, 항해의 위험이 증가하였다. 따라서 해상에서 발생할 수 있는 사고를 미리 탐지하고 예방하기 위해 선박 경로 예측의 중요성이 커지고 있으며, 경로 예측과 관련된 연구들이 활발히 진행되고 있다.

선박 경로 예측을 위해 사용되는 AIS(Automatic Identification System)는 일정한 시간 간격으로 선박의 위치 및 운항 정보를 전송(broadcast)하여 주변의 다른 선박이나 육상 기지국에서 선박의 운항 상황을 자동으로 탐지할 수 있는 시스템이다. 송수신되는 정보들은 크게 선박과 관련된 정적정보와 항해함에 따라 변화하는 동적정보로 구분할 수 있다. 정적정보에는 선박의 ID 값인 MMSI 번호, 선박의 종류, 길이, 폭, 너비와 같은 정보를 담고 있고, 동적 정보에는 실시간 선박의 위치, 방향, 속도 등과 같은 정보를 담고 있다. 이처럼, AIS 데이터는 다량의 선박 궤적 데이터를 얻을 수 있고, 선박의 이동 특성을 반영할 수 있는 추가적인 정보들을 얻을 수 있어 선박 경로 예측에 유용한 데이터이다. AIS 데이터는 실시간으로 항해 정보를 송수신하기에 시간의 흐름을 갖는 시계열 데이터이다. 시계열 데이터는 시간의 순서가 중요하기에 여러 딥러닝 모델 중, 시간의 흐름을 저장하고 예측에 반영할 수 있는 LSTM 모델을 사용하는 것이 적합하다. LSTM은 RNN 모델의 한계인 장기 의존성 문제를 해결하여 시계열 데이터 처리에 특화된 모델이다. 또한, 딥러닝 모델 학습을 위한 본 연구만의 학습 데이터 단위 look\_back 및 이후 예측 지점을 나타내는 forward\_length를 설정하여 모델을 개발하였다.

본 논문은 다음과 같이 구성되어있다. 2장에서는 AIS 데이터를 활용한 선박 경로 예측 관련 연구와 개발한 선박 경로 예측 딥러닝 모델에 대해

설명하고 모델의 성능을 제시한다. 3장에서는 결론과 향후 연구 방향에 대해 제시한다.

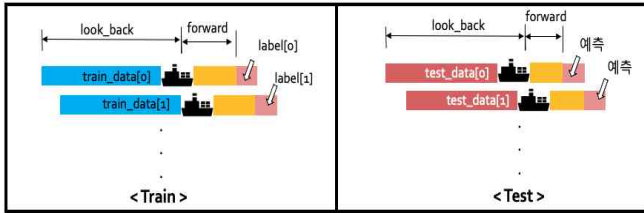
### II. 본론

경로 예측은 보행자 경로 예측, 차량 경로 예측과 같은 다양한 분야에서 연구되고 있다. 통계를 기반으로 하는 전통적인 경로 예측에서 기계 학습 기반 예측, 현재는 딥러닝 기반 예측으로 방법이 발전해왔다.

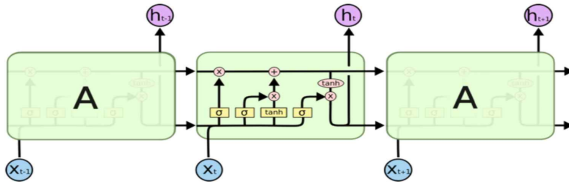
전통적인 방법의 시계열 데이터 예측에는 주로 Gaussian maximum likelihood 모델, Hidden Markov 모델 등의 확률적 예측 모델이 활용되었다[2][3].

본 연구는 AIS 데이터를 활용하여 연구를 진행하였다. AIS 데이터에는 MMSI(선박의 ID), timestamp, latitude, longitude, speed 등의 다양한 선박에 관련된 세부 정보들이 존재한다. 본 연구에서는 MMSI, timestamp, latitude, longitude의 4가지 정보만 사용한다. 원본 데이터를 MMSI 별로 데이터를 분할한 후, 시간 순서대로 정렬하는 과정을 통해 선박별로 데이터를 전처리하였다.

본 연구에서 사용하는 시계열 데이터는 시간의 경과와 함께 일정한 간격마다 관측 값이 기록된 데이터이다. 시계열 데이터는 과거의 데이터에 종속적이다. 그렇기에 시계열 데이터를 예측하기 위해서는 이전 시간 간격의 데이터를 많이 학습할 수 있어야 한다. 때문에, 이러한 시계열 데이터 예측에 최적화된 딥러닝 모델인 LSTM(Long Short-Term model) 모델을 개발하였다[4]. RNN 모델은 시퀀스를 가지고 있는 시계열 데이터를 처리할 때 효과적인 모델이지만 시퀀스의 길이가 길어질수록 중요한 정보를 가진 이전의 데이터를 고려하지 못하게 되어 장기 의존성 문제가 발생한다[5]. 이를 해결하기 위해 [그림2]에서 RNN의 hidden state에 cell-state를 추가하여 장기 의존성 문제를 개선한 모델인 LSTM 모델을 사용하였다.



[그림1] look\_back을 활용한 데이터 라벨링



[그림2] LSTM 모델 구조

본 연구에서 개발한 선박 경로 예측 LSTM 모델은 AIS 데이터의 gps 좌표 (위도, 경도)를 입력으로 받아 다음 시간의 위치(위도, 경도)를 예측하는 모델이다. 학습을 할 때, 한 번에 모든 학습데이터를 통해 학습하는 것이 아닌 일정한 시간 간격을 지정해 해당 간격만큼 데이터를 학습한 후 그 다음 위치를 예측하기 위해 look\_back을 사용한다. [그림1]에서 look\_back은 학습을 하기 위한 데이터의 개수를 의미한다. look\_back 길 이만큼의 데이터를 묶어 하나의 학습 데이터로 만든다. forward는 학습데이터의 라벨을 forward\_length만큼 떨어진 위치로 설정하겠다는 의미이다. 학습된 지점 바로 다음 위치를 예측하는 것이 아닌, 학습된 지점 이후 더 먼 위치를 예측하는 딥러닝 모델을 생성하기 위해 해당 개념을 적용하였다. 학습데이터 길이의 단위가 되는 look\_back의 길이와 예측 위치를 나타내는 forward\_length에 따라 모델의 성능이 변함을 확인하였다. 또한, MinMaxScaler를 사용하여 입력데이터를 0과 1 사이의 값으로 스케일링 하였다. 원본 데이터를 그대로 딥러닝 모델이 학습하도록 하면 학습이 느리거나 문제가 발생하게 된다. 그래서 딥러닝 모델에 데이터를 입력하기 전에 스케일링을 사용하여, 본래 데이터의 분포를 유지하면서 오버플로우나 언더플로우를 방지하고 최적화 과정에서의 안정성 및 수렴 속도를 향상시킨다. 또한 손실함수(loss function)를 최소화하고 학습속도를 빠르고 안정적이게 하기 위해 사용하는 최적화(optimization) 과정에서는 아담(Adam) 최적화 알고리즘을 사용하였다[6]. 아담 알고리즘은 stepsize가 gradient의 rescaling에 독립적이라는 특징이 있어 파라미터(weight)마다 학습률을 달리 조정이 가능하다는 장점이 있다. 손실함수에는 평균제곱오차(Mean Squared Error)를 사용하였다. 평균제곱오차는 예측값과 실제값의 차이, 즉 오차값에 제곱을 한 후 평균을 취한 것으로, 딥러닝 모델의 손실함수로 주로 사용된다. Drop\_out은 모델이 학습할 때, 신경망의 뉴런을 부분적으로 생략하여 모델의 과적합(overfitting)을 해결해 주기 위해 사용하는 방법이다. 본 모델에서는 Drop\_out을 0.4로 적용하였다. 최적의 성능을 내는 본 연구의 LSTM 모델의 초매개변수(hyper-parameter)는 다음과 같다. 레이어는 총 3개로 설정하였고, 전체 학습의 횟수를 나타내는 에폭(Epoch)은 100으로, 학습의 단위가 되는 배치(Batch)사이즈는 70으로 설정하였다. Train 데이터는 1652개의 AIS 데이터를 활용하였고, Test 데이터는 413개의 AIS 데이터를 활용하였다. 본 연구의 실험은 Intel(R) Core(TM) i7-7700 CPU @ 3.60GHz, 32GB RAM, GeForce GTX 1080 Ti GPU 환경에서 진행하였다.

본 연구의 선박 경로 예측 모델의 최고성능은 look\_back이 10일 때, 0.0040의 MSE 오차를 갖는다. 또한 시계열 데이터 예측 성능에 영향을

look_back	학습시간(초)	MSE
10	183.8849	0.0038
30	450.5410	0.0039
50	632.5809	0.0043

[그림3]. look\_back 별 모델 성능표(forward\_length = 0)

Forward_length	학습시간(초)	MSE
5	185.6431	0.0051
10	173.1309	0.0089
15	180.4802	0.0106

[그림4]. forward\_length 별 모델 성능표(look\_back = 10)

끼치는 look\_back 길이에 따른 성능을 [그림3]으로 제시하였다. 선박 예측 모델의 최적 look\_back 길이는 10임을 확인하였다. look\_back의 길이가 10보다 길어질수록 예측성능은 조금씩 떨어지고, MSE는 증가함을 확인하였다. [그림4]에서는 forward\_length가 커질수록 예측하는 지점이 더 멀어지기에 성능이 저하되었으나, 낮은 MSE 값을 보이며 우수한 예측 성능 모델임을 확인할 수 있었다. 이를 통해 선박의 경로 예측을 위해서는 많은 양의 이전 경로 데이터를 학습하기보다는 바로 이전의 10개 데이터를 학습함으로 다음 위치를 예측하는 것이 가장 효율적임을 확인하였다.

### III. 결론

본 논문에서는 선박의 경로 예측을 위해 시계열 데이터인 AIS 데이터를 활용한 딥러닝 LSTM 모델을 개발하였고, 최적의 성능을 내는 파라미터를 제시하였다. 또한, 본 모델의 시계열 데이터 예측에 영향을 끼치는 look\_back과 forward를 사용하여 각 데이터에 따른 성능을 제시하였다. 향후 연구에는 학습 데이터의 시간 간격을 학습에 반영시켜 더욱 좋은 성능의 예측결과를 얻고자 한다. 또한, 선박뿐만 아니라 다양한 이동수단의 이동패턴을 파악하여 경로를 예측하는 경로 예측 모델 개발도 진행하고자 한다.

### ACKNOWLEDGMENT

본 과제(결과물)는 2023년도 교육부의 재원으로 한국연구재단의 지원을 받아 수행된 지자체-대학 협력기반 지역혁신 사업의 결과입니다. (2021RIS-004)

### 참 고 문 헌

- [1] Yu, Jun Ye, Moslem Ouled Sghaier, and Zofia Grabowiecka. "Deep learning approaches for AIS data association in the context of maritime domain awareness." 2020 IEEE 23rd International Conference on Information Fusion (FUSION). IEEE, 2020.
- [2] Varshneya, Daksh, and G. Srinivasaraghavan. "Human trajectory prediction using spatially aware deep attention models." arXiv preprint arXiv:1705.09436 (2017).
- [3] Endo, Yuki, et al. "Predicting destinations from partial trajectories using recurrent neural network." Pacific-Asia Conference on Knowledge Discovery and Data Mining. Springer, Cham, 2017.
- [4] Hochreiter, Sepp, and Jürgen Schmidhuber. "Long short-term memory." Neural computation 9.8 (1997): 1735-1780.
- [5] Zaremba, Wojciech, Ilya Sutskever, and Oriol Vinyals. "Recurrent neural network regularization." arXiv preprint arXiv:1409.2329 (2014).
- [6] Zhang, Zijun. "Improved adam optimizer for deep neural networks." 2018 IEEE/ACM 26th international symposium on quality of service (IWQoS). Ieee, 2018.