

제한적 선박 정보와 무작위의 숲 분류기를 이용한 선종 예측

† 전호군 · 한재림*

† 한국해양과학기술원 해양빅데이터센터 UST학생연구원, † 과학기술연합대학원대학교 응용해양과학전공 박사과정생
*한국해양과학기술원 해양빅데이터센터 연구원

Ship Type Prediction using Random Forest with Limited Ship Information

† Ho-Kun Jeon · Jae Rim Han*

† UST Student Researcher, Korea Institute of Ocean Science and Technology, Busan, 49111, Korea
† Ph.D. Student, University of Science and Technology(UST), Daejeon 34113, Korea
*Researcher, Korea Institute of Ocean Science and Technology, Busan, 49111, Korea

요 약 : 주변선박에 대한 선종 식별은 선박의 향후 이동 경로와 조종 특성을 유추할 수 있으므로 항해사와 VTS 관제사에게 모두 중요한 정보이다. 그러나 AIS메시지 전송 중 통신 문제, 항해사의 AIS 이용 미숙지로 인해 선박의 선종 확인이 불가능한 경우가 빈번하다. 따라서 이 연구는 AIS데이터를 학습 및 테스트 데이터셋으로 분할, 무작위의 숲 분류기(Random Forest Classifier)에 AIS데이터의 선박 특성(feature)과 선종을 훈련 및 예측하는 방법을 제시한다. 연구를 위해 2018년 한해 울산 앞바다 AIS데이터를 이용하였다. 이 방법을 사용하면 다수의 항해사 또는 VTS 관제사가 토론을 통해 선종예측 경험을 공유하는 것과 같은 기능을 할 수 있다.

핵심용어 : 선종, 예측, 분류, 자동식별장치, 무작위의 숲, 기계학습

Abstract : The ship type identification of the surrounding ship is important information for navigators and VTS officers since they can estimate the maneuverability and near-future route of the ships. However, it is more than frequent that the information is not provided due to transmission trouble and seafarers' unfamiliarity with AIS. Thus, this study suggests predicting ship types through the Random Forest classifier after preparing a training and test dataset that contains ship features and types. The AIS data for Ulsan coast in 2018 was used for this study. The method may provide the effect that many navigators and VTS officers discuss and share the experience of predicting ship types.

Key words : Ship type, Prediction, Classification, AIS, Random Forest, Machine learning

1. 서 론

항해사와 해상교통관제사(이하 관제사)가 당직 중 행하는 선종 확인은 육안식별과 선박자동식별장치(Automatic Identification System, 이하 AIS)의 정적메시지 확인으로 구분된다. 야간 또는 해무로 인해 육안 식별이 어려운 경우 AIS를 확인한다. 그러나 AIS를 통해 움직임을 파악할 수 있는 선박 중 25~60%가 선종 정보를 포함하고 있지 않다[1]. 이러한 문제는 기기결합, 메시지 전송 불량, 항해자의 AIS 이용 미숙지에 기인한다[2].

선박 종류(이하 선종)에 따라 주로 사용하는 항로가 결정되므로[3], 주변 선박의 선종을 알 수 있다면 해당 선박의 조종 성능뿐 아니라 가까운 미래의 진행경로, 더 나아가 선박 간의 조우 관계를 경험적으로 예측할 수 있다. 따라서 선종 확인은 안전운항과 항만안전을 도모하는 항해사 및 관제사에게 중요한 요소이다.

따라서 이 연구에서는 선박의 전장, 폭, 시각, 위치, 속도, 침로를 특성(feature)으로 하여 무작위의 숲 분류기(Random

Forest classifier, 이하 RFC)를 통해 예측하는 방법을 제안한다.

2. 연구영역 및 데이터 전처리

연구영역은 울산 앞바다(129.3-129.6E, 35.3-35.6N)이다. 2018년 8월 1일 AIS에 기록된 395척의 선박 중 140척(35.4%)이 탱커선으로 가장 높은 비중을 차지한다. 항만 인근에서 입출항, 묘박, 정박 중인 선박으로부터 영향을 최소화하기 위해 해안선으로부터 2km 외측 경계선을 설정하고 경계선 밖의 데이터만 사용하였다.

AIS는 선종을 코드로서 구분하고 있다. 선종코드는 0~99번 정수로 정의된다. 이 중에서 여객선(passenger), 화물선(cargo), 탱커선(tanker), 수중익선(WIG), 고속정(HSC), 기타(other), 예비(reserved), 여유(spare) 등 명칭은 동일하나 코드가 상이한 선종을 단일화하고, 그 외 선종명은 상이하나 성격 면에서 유사한 선종들도 일정한 기준을 통해 그룹을 지어 간소화하였다. 시간 관점에서 AIS데이터는 일정한 시간 간격을

가지지 않는다. 공간 관점에서 선종에 따른 주요항로를 고려 하되 동일한 범위를 설정할 필요가 있다. 따라서 시간적인 단일화를 위해 시간은 5분 등간격으로 내삽보간(interpolation)하고 공간은 약 0.05°(약 5.5km) 등간격으로 해역(zone)을 구분하여 각 영역에 좌표번호 X, Y를 할당했다.

이렇게 정리된 AIS데이터는 시각(Time), 대지선속(SOG), 대지침로(COG), 전장(length), 폭(breadth), 영역좌표 X, Y로 구성되어 있다.

3. 무작위의 숲 분류기

RFC는 여러 의사결정나무 분류기(Decision Tree classifier, DTC)들의 판단 결과를 종합하여 최종적인 판정 결과를 제시하는 모델이다(Ho, 1995). 여러 분류모델 및 예측모델 중에서 우수한 성능을 나타내왔다[4].

RFC는 Python의 Library {Scikit-learn}의 Random Forest Classifier 모듈을 이용해 구현했다[5]. RFC 최적화를 위해 {Scikit-learn}의 GridSearchCV 기능을 이용하여 길이 정보를 가진 경우 224개 경우의 수, 길이 정보 부재한 160개 경우 수에 대해 선종예측 성능을 평가하고 최적 예측기 수, 최대특성 수, 최대 깊이를 결정하였다.

4. 결과 및 논의

4.1 최적 초매개변수

기상의 영향이 적은 2018년 8월 1일에 대한 7일간의 학습데이터와 GridSearchCV를 통해 최적 초매개변수(hyperparameter)를 결정했다. 길이 정보가 있는 경우 Bootstrap은 사용하지 않으며, 예측기(estimator) 수가 크기 정보 부재 시 보다 4배나 낮았다. 또한 최대 특성 수와 최대 깊이가 부재 시보다 높았다. 즉, 적은 수더라도 깊게 생각하는 예측기가 필요했다. 반면 부재 시는 얇게 생각하는 예측기가 많이 필요했다.

4.2 일일 예측성능

크기 정보가 있는 모델은 Q1와 Q3가 0.8보다 높았으며 IQR 크기가 0.1 이하로 안정적이었다. 그 외 f1-score, recall, precision의 경우 Q1와 Q3가 0.6과 0.9사이에 위치하였고 IQR 크기가 0.2 이하로 accuracy보다 안정적이지 않지만 중간값이 0.7을 모두 넘어 높은 예측성능을 가진다고 볼 수 있다.

크기 정보 부재에 대한 모델은 accuracy를 제외한 나머지 성능지표들은 점수가 낮았다. Accuracy의 경우 3월, 7월, 10월 중 성능이 급감하는 날들이 있었으나 중간값은 0.7이상이고 IQR이 0.1보다 작아 안정적이었다. 반면 f1-score, recall, precision은 Q1, Q3, 중간값이 0.4보다 낮았으며 IQR은 accuracy와 큰 차이를 보이지 않았다.

5. 결 론

RFC를 활용하여 선종을 예측할 때, 선박의 길이 정보를 가진 경우 정확도 0.8 이상으로 신뢰할 수 있는 모델이 생성되었다. RFC는 다수의 DTC를 기반으로 하였기 때문에 다수의 항해사 또는 VTS관제사가 모여 다수결로 의사결정을 한 것과 같은 효과를 가져온다고 본다.

그러나 길이 정보를 가지지 않은 모델이 현저히 낮은 성능을 보여주어 이에 대한 개선점이 필요하다. 향후 연구에서는 각 영역의 좌표번호 X, Y 대신 선종별 주요 항로를 특성정보로 하여 선종 예측 성능을 향상시키고자 한다.

사 사

이 연구는 해양경찰청 “위성연계 접경수역 선박 모니터링 및 분포 예측 체계 개발(PM62980)” 과제 지원을 받아 수행되었습니다.

참 고 문 헌

- [1] Hong D. B.(2014), “Classification of Passing Vessels Around the Jeodo Ocean Research Station Using Automatic Identification System (AIS): November 21-30, 2013”, J. Kor. Soc. Mar. Env. Energy., 17, 4, pp.297-305.
- [2] Emmans, T., Amrit, C., Abdi, A. and Ghosh, M. (2021), “The promises and perils of Automatic Identification System data, The Promises and perils of Automatic Identification System data”, Expert Sys. with Applications, 178, 114975.
- [3] Lee, J.S. and Cho, I.S. (2022), “Extracting the Maritime Traffic Route in Korea Based on Probabilistic Approach Using Automatic Identification System Big Data”, applied sciences, 12, 635.
- [4] Lee, Y.J, Han, D.H., Ahn, M.H, Im, J.H., Lee, S.J. (2019), “Retrieval of total precipitable water from Himawari-8 AHI data: A comparison of random forest, extreme gradient boosting, and deep neural network”, Remote Sens., 11, 15, 1741.
- [5] Pedregosa. F. Varoquaux, G., Gramfort, A., Michel, V. Thirion, B. Grisel, O. Blondel, M. Prettenhofer, P., Weiss, R. 7Dubourg, V., Vanderplas, J. Passos, A. Cournapeau, D. Brucher, M. Perrot, M., Duchesnay, E. (2011), “Scikit-learn: Machine Learning in Python”, JMLR, 12, pp.2825-2830.