

[Oceanography] Article

선박 분포 예측을 위한 학습용 데이터셋 구축: 제주도 북부 해역을 대상으로

박용길 · 김태훈 · 한현경 · 이철용*

한국해양과학기술원 해양빅데이터센터, 부산 49111, 대한민국

Construction of a Training Dataset for Vessel Distribution Prediction: The Northern Seas of Jeju Island

Yonggil Park · Taehoon Kim · Hyeon-Gyeong Han · Cholyoung Lee*

Marine Bigdata Center, Korea Institute of Ocean Science and Technology, Busan 49111, Republic of Korea

Received: 16 June 2022, revised: 8 July 2022, accepted: 11 July 2022

요약문 최근 해상 선박 상호간 충돌 예방, 불법 선박 탐지, 선박 경로 예측 등 해상사고 및 안전 관련 연구의 관심이 높아지고 있다. 이 때 선박위치정보 기반의 선박 분포도는 해양 안전 관리를 위한 의사결정을 지원할 수 있으며, 선박 분포를 예측할 수 있다면 조업 안전 관리 및 불법 조업 방지 등 해상 보안을 위한 선제적 대응이 가능하다. 이를 위해 본 연구에서는 선박패스(V-Pass)자료, 기상특보 및 해양환경 자료를 수집하여 선박 분포 예측을 위한 학습용 데이터셋을 구축하였다. 선박위치정보의 전송주기를 리샘플링한 결과를 격자 자료에 매핑하여 선박 밀집 정도를 평가하였으며, 연구대상지역에 대해 총 1,314,000개의 학습용 데이터를 구축하였다. 향후 선박 분포 예측 모델링을 수행하여 정확도를 평가하는 연구가 수행되어야 할 것이다.

주요어: 선박 분포, 선박패스, 학습용 데이터셋, 선박위치정보

Abstract Recently, interest in maritime accidents and safety-related research, such as preventing collisions between marine vessels, detecting illegal vessels, and predicting vessel routes, is increasing. Vessel location data-based vessel distribution map can support decision-making for maritime safety management, and if the vessel distribution can be predicted, it is possible to take a preemptive response for maritime security such as fishing safety management and illegal fishing prevention. In this study, a training dataset for vessel distribution prediction was constructed by collecting V-Pass data, weather warnings, and marine environment data. The result of resampling of reporting interval of vessel location data was mapped to grid data to evaluate the vessel density, and a total of 1,314,000 of training data were constructed for the study area. In the future, research to evaluate the accuracy by performing vessel distribution prediction modeling should be conducted.

Keywords: Vessel distribution, V-Pass, Training dataset, Vessel location data

*Corresponding author: chlee82@kiost.ac.kr

1. 서론

최근 국제교역량의 증가에 따른 해상 교통량이 증가하면서 선박 상호간 충돌 예방, 불법 선박 탐지, 선박 경로 예측 등 해상사고 및 안전 관련 연구가 활발히 수행되고 있다. 이러한 해상 안전을 위한 연구에는 선박위치정보의 활용이 필수적이며 이에 따라 선박위치정보를 활용한 연구의 관심이 높아지고 있다(Hwang et al., 2016). 국내에서는 AIS (Automatic Identification System, 선박자동식별시스템)와 V-Pass(선박패스) 장치의 설치를 법으로 규정하고 있으며 이를 이용하여 선박위치정보를 수집 및 활용하고 있다(IMO, 1998).

과거 선박위치정보는 해양사고 분석 등 선박의 운항 및 안전관리를 위한 항적 조회 및 관리를 목적으로 활용되고 있었으나 최근에는 해상 교통량 분석, 선박 분포도, 조업 지역 분석 등 다양한 분야에 활용되고 있다(Pokrajac et al., 2005). 특히 선박 분포도는 조업 안전을 위한 해경 및 어업관리단의 경험적 근거에 의한 순찰 지역 선정에 과학적 분석에 의한 선박 밀집 조업 지역을 특정할 수 있다. 또한 조업 시준에 따른 시기별 집중 안전 관리지역 선정이나 선박안전관리를 위한 시설 설치 등 해상사고 및 조업 안전을 위한 과학적 근거를 제시할 수 있다.

한편 중국의 불법 조업은 해양생태계 파괴 및 국내 어업자원의 고갈을 일으키는 주요 원인으로 지목되고 있으며, 이를 방지하고 위해 해양 경찰에서는 조업 안전을 확보하고 불법어업 선박을 단속하기 위해 위한 순찰을 수행하고 있다(Lim, 2014). 중국의 불법조업 선박의 활동 감시 및 국내 선박 조업안전관리 등을 위한 해양경찰의 순찰이 강화되고 있으나 해경선 배치 수에 한계가 있기 때문에 모든 해역을 동시에 관할할 수 없으며, 적절한 해경선의 배치가 순찰의 중요한 의사결정 사항으로 떠오르고 있다. 이 때 선박 분포도를 활용한다면 선박 밀집 지역은 조업 안전을 위한 해경선 배치를 수행할 수 있으며, 선박이 없는 지역은 불법 조업을 방지하기 위한 순찰 지역으로 해경선 배치를 수행할 수 있다. 이외에도 국가해양공간계획의 어업 구역을 획정하기 위한 용도 등 다양한 목적을 달성하기 위해 선박 분포도를 활용하고 있다.

이 때 선박 분포도를 제작 및 활용에 그치지 않고 미래의 선박 분포를 예측할 수 있다면 조업 안전 관리 및 불법 조업 방지 등에 대해 선제적 대응이 가능하다. 그러나 선박위치정보는 VTS (Vessel Traffic Service, 해상교통통제)를 위한 연구분야에서 주로 활용되고 있으며, 특히 어선에 대한 선박위치정보 활용 사례는 찾아보기 어려운 실정이다. 또한 선박위치정보의 획득 절차가 어렵기 때문에 선박 분포도 제작 및 활용이 제한적이며, 연평균 수집되는 80억 건의 AIS 자료와 35억 건의 V-Pass 자료를 처리하기 위한 기술 또한 요구되고 있는 실정이다(Ryu et al., 2014).

4차산업의 발전과 함께 빅데이터 처리 기술이 발전함에 따라 선박 분포도의 제작과 선박의 분포를 예측할 수 있는 기반이 조성되고 있으나 인공지능을 활용하여 선박 분포를 예측하기 위한 학습용 데이터셋을 제작하기 위한 기준은 정의되어 있지 않다. 따라서 본 연구에서는 V-Pass 자료를 이용하여 연속적으로 이동하는 선박의 분포의 시공간적 변화를 예측하기 위해 격자 형태의 학습용 데이터셋을 구축하였다. 이를 위해 선박위치정보의 전처리와 함께 선박 분포를 예측하기 위한 보조 자료의 수집을 수행하였으며, 시공간적 학습을 위한 일정한 간격의 학습용 데이터셋을 구축하였다.

2. 본론

2.1 연구 방법

선박의 분포를 예측하기 위해서는 과거의 선박의 행적을 통한 밀집 구역을 탐지하고 이동하는 선박의 움직임

을 포함하여 밀집 지역의 변화를 예측할 수 있어야 한다. 이는 시간의 흐름에 따라 공간적 변화가 일어나는 것이며, 시공간적 변화를 예측할 수 있는 방법을 제안할 필요가 있다. 또한 어선의 출항 및 어업에 영향을 줄 수 있는 다양한 원인요소들을 분석하여 가장 적합한 예측 모델을 만들 필요가 있다.

선박밀집도를 예측하기 위해서는 선박의 위치를 파악할 필요가 있다. 이러한 선박위치정보는 AIS를 통해 수집된다. AIS는 선박을 식별할 수 있는 정보와 함께 선박의 위치정보를 선박의 항해 상태와 속도에 따라 2 초에서 3 분 간격으로 무선 전송하는 시스템이다(Kim, 2014; IMO, 1998). 국내의 경우 150 톤 이상의 여객선은 AIS 장치를 설치하도록 법으로 규정되어 있다. 어선의 경우 대부분의 선박이 150 톤 미만이기 때문에 AIS 설치가 의무는 아니지만 어선의 출항 및 입항 신고 자동화를 위한 V-Pass를 장착하고 있다. V-Pass 장치도 선박의 항해 상태 및 속도에 따라 선박 식별정보와 위치정보를 무선으로 전송한다.

해양 공간은 예측을 하기 위한 대상으로서 빈 공간이 너무 넓기 때문에 선박 각각의 움직임을 예측하여 밀집 지역을 판단하기 어렵다. 선박이 많이 밀집되어 어업을 수행하는 지역일지라도 같은 선단이 아니라면 선박 사이의 간격이 1 km 이상씩 떨어져 있는 모습이 많이 나타난다. 시공간적 예측을 수행하기 위해 적절한 데이터의 형태는 격자형태이다. 본 연구에서 활용된 격자는 국가해양공간계획의 기본단위로 활용되고 있는 격자이며 6단계의 격자 중 3단계 격자를 활용하였다. 3단계 격자는 3분 간격의 격자로 약 5.5 km x 5 km의 크기를 가지는 직사각형 격자이며, 18 kn의 속도로 이동하는 선박이 10분간 이동하였을 때 격자간 이동을 나타낼 수 있는 크기이다. 선박 위치정보를 격자화하기 위해 위경도 데이터를 이용하여 3단계 격자에 매핑하였다. 각 격자별로 격자 ID가 부여되어 있으며, 선박위치정보의 위경도를 이용하여 격자와 매핑하기 위한 코드를 개발하였다. 이를 위해 Fig. 1과 같이 3단계 격자 25개에 포함되는 선박의 위치정보를 연구대상으로 선정하였으며, 2018년 1월부터 2018년 12월까지 1 년간의 V-Pass 선박위치정보를 수집하여 활용하였다.

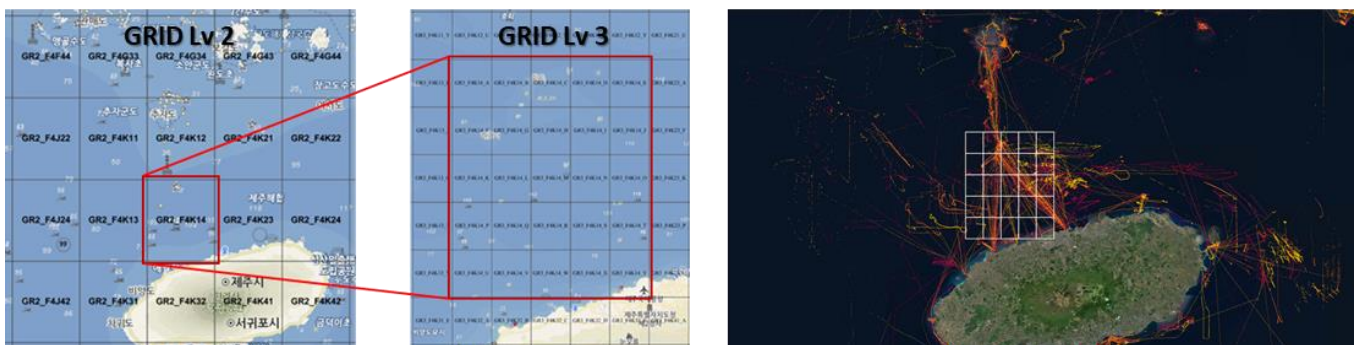


Figure 1. Study area: northern seas of Jeju Island.

선박 분포도에서 밀집도를 평가하기 위한 방법으로는 밀집도를 측정하려는 시간의 모든 선박 위치를 격자에 매핑한 후 각 격자에 포함된 선박의 숫자를 세어 그 개수를 선박의 밀집도로 정의하였다. 이 때 선박위치정보의 전송 주기는 일정하지 않기 때문에 전송 시간을 일치시키기 위한 처리가 필요하다. 이를 위해 본 연구에서는 선박위치정보의 전송주기를 리샘플링하여 그 결과를 활용하였다.

선박밀집도 예측의 정확도를 확인하기 위해서는 선박의 위치이동이 많이 이루어지는 지역을 연구대상으로 선정해야 한다. 제주도 북부 해역은 Fig. 1과 같이 선박의 통행이 많은 곳이며 특히 낚시어선이 많은 비율을 차지하고 있는 지역이다. 어선의 밀집도가 시간, 일별로 수시로 변하는 것을 확인할 수 있었으며, 본 연구의 목적에 부합하여 연구 대상지역으로 선정하였다.

2.2 선박위치정보 보간 및 매핑

선박 분포도를 제작하기 위해서는 밀집도를 측정하려는 시간의 모든 선박 위치를 알 필요가 있으나 선박위치 정보 전송 주기가 일정하지 않기 때문에 밀집도를 판단하기 어렵다. 수집된 선박위치정보를 그대로 활용하여 일정 시간 사이에 격자 내에 존재하는 선박위치정보 개수를 합할 경우 선박위치정보의 전송 주기가 짧은 선박이 더 많은 선박위치정보를 전송하게 되어 해당 선박이 머물렀던 격자가 더 높은 밀집도를 나타내는 오류가 발생하게 된다. 이 문제를 해결하기 위해서는 모든 선박의 선박위치정보의 전송 주기 및 전송 시간이 같아야 하지만 실제로는 불가능하기 때문에 별도의 자료처리 방식이 요구된다. 또한 Fig. 2와 같이 선박위치정보의 자료를 처리과정 없이 전송 시간 순으로 나열했을 경우 불규칙한 전송주기로 인해 알 수 없는 데이터의 나열로 나타난다. 따라서 본 연구에서는 선박위치정보의 전송주기를 일정한 간격으로 리샘플링하고 리샘플링 결과로 생성된 선박위치정보를 보간하였다.

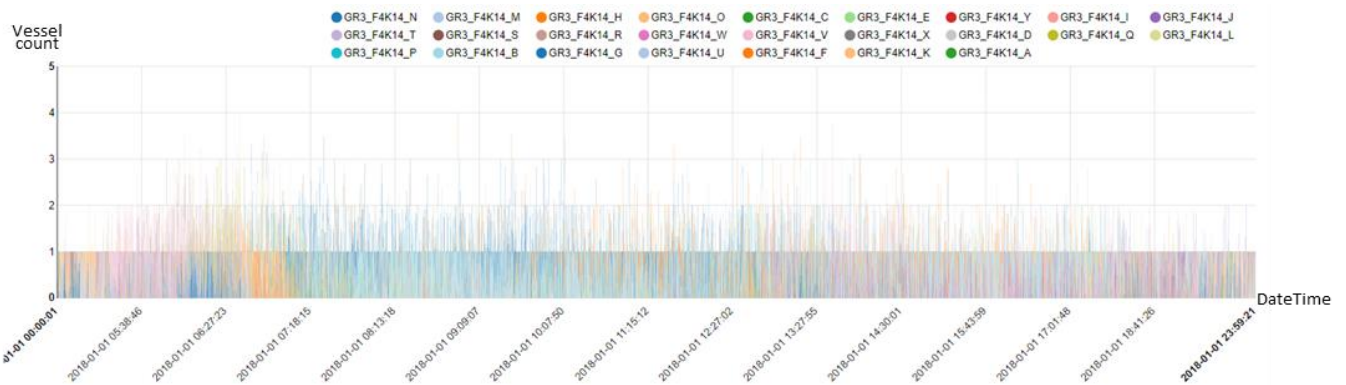


Figure 2. Example of reporting interval distribution of vessel location data (1-Jan-2018).

선박위치정보의 전송주기를 일정한 간격으로 조정하게 된다면 선박위치정보 또한 현재 전송된 시간의 위치에서 변경된 시간의 선박위치로 이동시켜야 한다. 여객선이나 화물선 등 정해진 항로를 따라 이동하는 선박은 대부분 직선으로 이동하기 때문에 수집된 선박위치 간 이동을 직선으로 이동했다고 가정한 후 위치를 보간할 수 있다. 어선의 경우 필요에 따라 경로를 수정하기 때문에 칼만필터 등 유연한 움직임을 구현하기 위한 알고리즘을 적용하여 보간할 필요가 있다. 그러나 본 연구에서는 선박의 위치정보가 5.5 km 격자 내에 포함되도록 수정이 이루어지며, 10 분 간격으로 보간하였을 때 그 위치정보가 분포도 작성에 미치는 영향이 적음을 고려하여 어선 또한 직선으로 이동했다고 가정하고 선박위치의 보간을 수행하였다. 이 때 연속된 선박의 이동을 보간하는 것이기 때문에 동일한 선박ID를 가지고 있는 위치정보의 전송 시간이 30 분 이상 차이가 날 경우 신호의 누락이나 항구에 정박한 것으로 판단하고 보간하지 않도록 하였다. Fig. 4는 선박위치정보의 전송주기를 1 분 간격으로 리샘플링한 결과이다. 전송주기를 일치시킴으로써 매 1 분마다 모든 선박의 위치를 동시에 파악할 수 있으며 이를 이용하여 선박 분포도를 제작할 수 있다. 다만 선박 분포도 제작 시 선박의 종류, 톤수, 선폭 및 선박길이 등 선박의 물리적인 정보와 연근해 어업, 통발어업 등 어업 면허를 이용하여 선종 분류 및 분류별 선박 분포도를 제작할 필요가 있었으나 이러한 정보는 개인정보 보안에 해당되는 정보로 활용하기 어려운 문제가 있어 본 연구에서는 적용하지 못하였다.

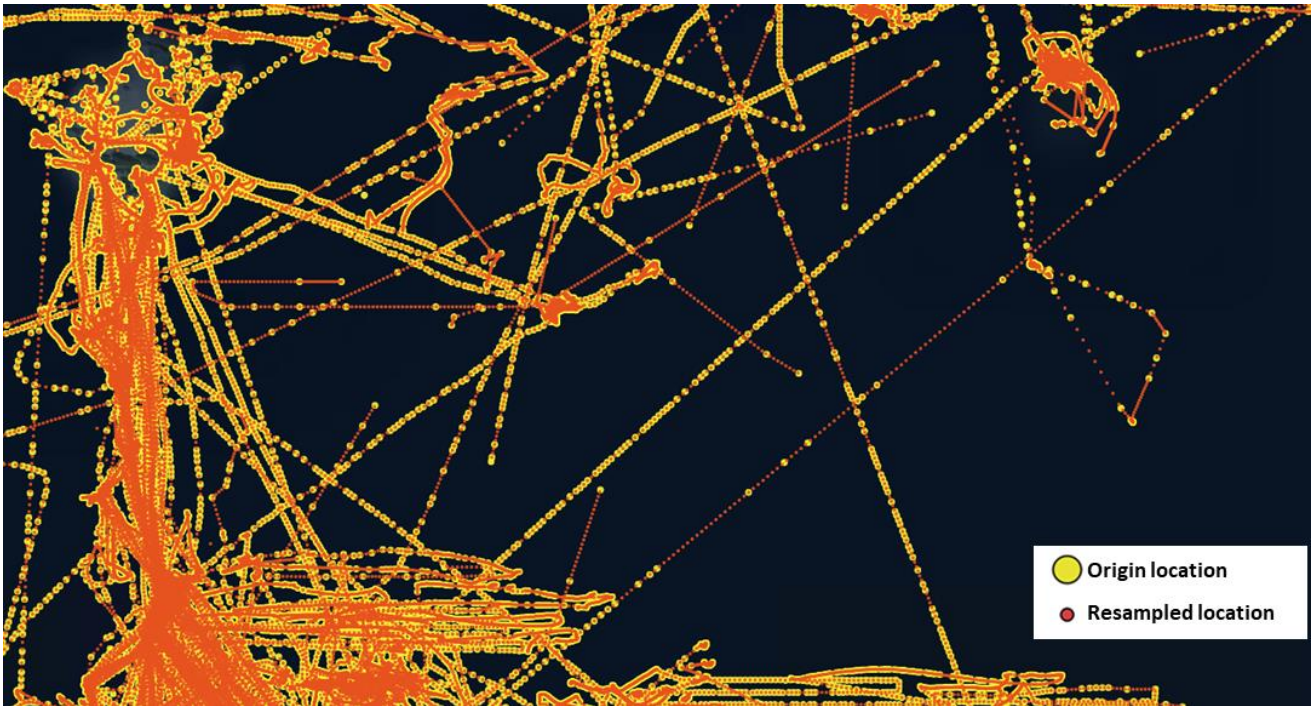


Figure 3. Example of resampled vessel location data.

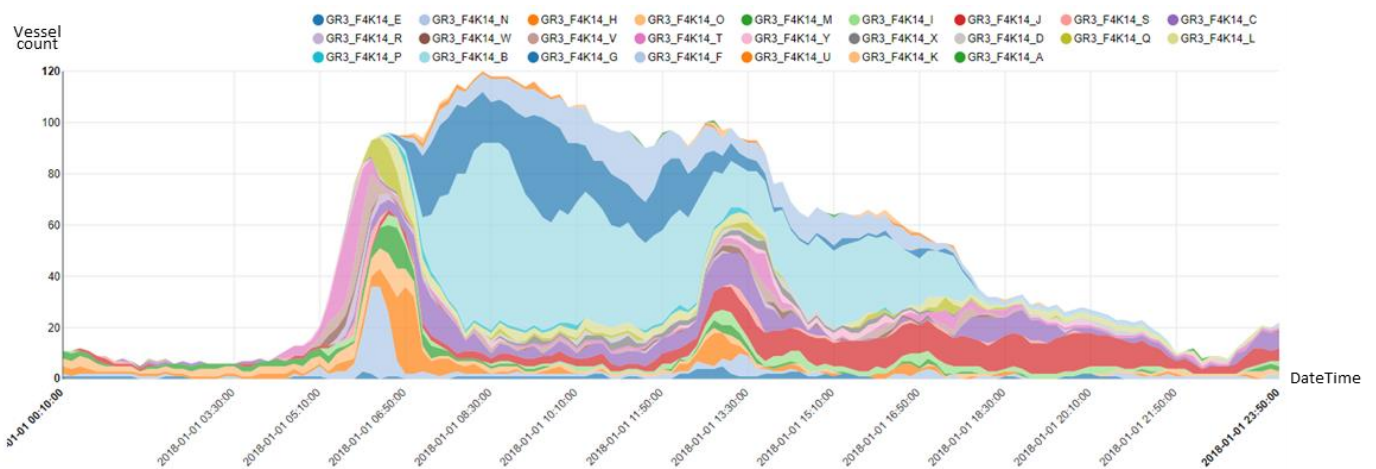


Figure 4. Example of reporting interval distribution of resampled vessel location data (1 min interval).

Fig. 4는 리샘플링한 연구대상지역의 선박위치정보를 시간순으로 나열한 그래프이며 Fig. 2에서 의미를 파악할 수 없었던 정보와 달리 시간의 흐름에 따라 출항한 선박의 총 개수와 선박이 출항하여 이동하는 것을 격자의 이동 패턴으로 확인할 수 있으며, 특정 지역에서 여러 대의 선박이 어업활동을 한 후 귀항하는 선박의 이동을 확인할 수 있었다. 리샘플링된 선박위치정보는 시간의 흐름에 따라 연속적인 격자의 변화가 나타나고 있으며 Fig. 4와 같이 의미를 분석할 수 없는 정보가 선박의 이동 패턴을 분석할 수 있는 정보로 변환된 것을 확인할 수 있었다. 이것을 이용하여 선박의 이동 패턴 및 어업 지역을 학습한다면 선박의 분포를 예측할 수 있는 모델링이 가능하며, 따라서 리샘플링된 선박위치정보를 이용하여 학습용 데이터셋의 구축이 가능함을 나타내고 있다.

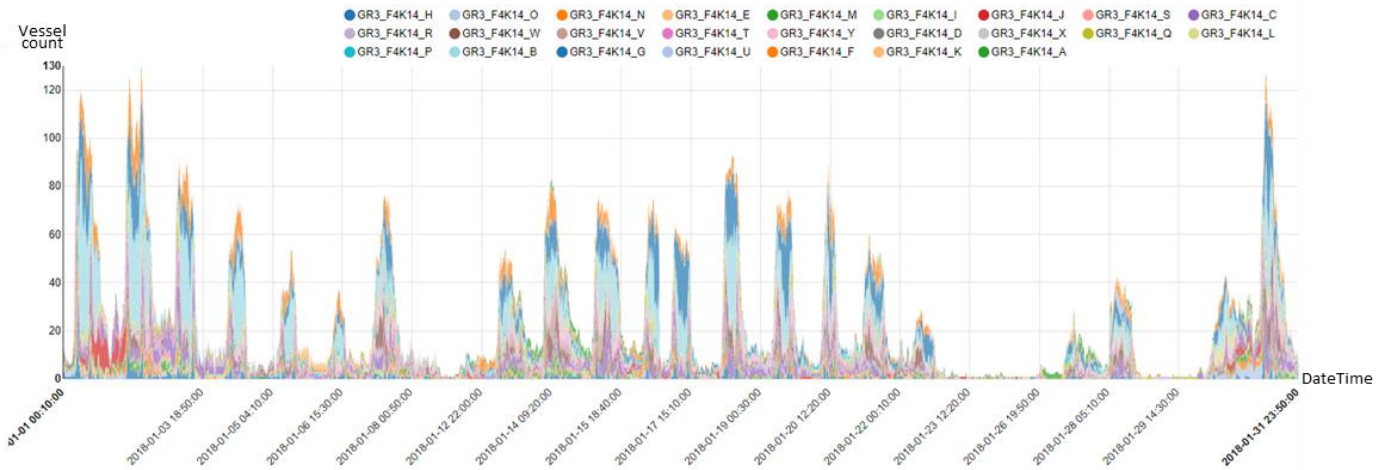


Figure 5. Example of reporting interval distribution of resampled vessel location data (10 min interval).

Fig. 5는 2018년 1월의 V-Pass 선박위치정보를 10 분 간격으로 리샘플링한 결과이다. 선박의 출항 대수 및 패턴을 확인할 수 있으며, 특정 기간에 출항을 하지 않는 것도 확인할 수 있었다. 이는 1월 22일부터 1월 24일까지 풍랑주의보가 발표되어 출항하지 못한 것으로 확인되었으며, 1월 28일부터 1월 30일까지도 풍랑주의보가 발표된 기간이다. 이를 통해 기상이 선박의 출항에 큰 영향을 미치고 있는 것을 확인할 수 있었다.

2.3 학습용 데이터셋 제작

본 연구에서는 V-Pass 선박위치정보 데이터를 10분간격으로 리샘플링한 결과를 이용하여 2018년 1월 1일 00시부터 2018년 1월 31일 23:50:00까지의 학습용 데이터셋을 구축하였다. 이 때 선박의 분포 예측을 위해서는 선박위치정보 이외에도 선박의 분포에 영향을 미치는 자료를 Table. 1과 같이 수집하여 함께 모델링할 필요가 있다.

Table 1. List of supplement data

| Category | Detail | Source |
|-------------------|--|---|
| Weather warning | Regional weather warning | Korea Meteorological Administration |
| Tide table | Daily times and levels of high and low tides | Korea Hydrographic and Oceanographic Agency |
| Water Temperature | 3' grid modeling data (1 day) | Ministry of Oceans and Fisheries |
| Salinity | 3' grid modeling data (1 day) | Ministry of Oceans and Fisheries |
| Sea wind | 3' grid modeling data (1 hour) | Ministry of Oceans and Fisheries |
| Wave height | 3' grid modeling data (3 hour) | Ministry of Oceans and Fisheries |

앞선 예시와 같이 풍랑주의보, 강풍주의보 등 기상특보가 선박의 출항에 많은 영향을 끼치는 것을 확인하였으며, 풍랑주의보, 풍랑경보, 태풍주의보 및 태풍경보 등의 기상특보를 기상청의 기상자료개방포털 (<https://data.kma.go.kr/data/weatherReport/wsrList.do>)에서 수집하여 학습 보조자료로 활용하였다. 또한 물때에 따라 어업 방식 및 어업 지역이 변경되는 것을 고려하여 국립해양조사원의 물때 정보를 수집하여 학습 보조자료로 활용하였다. 이와 함께 수온, 해상풍, 염분 등 해양환경 정보는 어획량에 큰 영향을 미치는 요소이며, 어업 시기 및 어업 지역을 결정짓는 요소로 작용할 수 있다. 따라서 해양수산부 해양수산 빅데이터플랫폼

(<https://www.vadahub.go.kr/user/aiModel/selectAiModelCont.do>)의 수온, 염분, 해상풍 및 파랑 모델링 자료를 수집하여 함께 활용하였다. 각 모델 결과는 3단계 격자를 기준으로 모델링 되어 있으며 각 모델링 값의 시간 간격은 Table. 1과 같다.

학습용 데이터셋은 빅데이터 플랫폼을 이용하여 구축되었다. 35억 건의 V-Pass 선박위치정보를 10 분 간격으로 리샘플링하였으며, 각 격자번호로 매핑한 후 10 분 간격으로 격자에 포함된 선박의 수의 합계를 밀집도로 나타내었다. 수집한 보조자료는 선박위치정보와 동일하게 격자 매핑을 수행하였으며, 빅데이터 플랫폼 내에서 시간 및 격자 번호를 기준으로 결합(Join)하여 학습용 데이터셋을 구축하였다.

3. 연구결과

3.1 우리나라 전국 선박분포도 제작

우선 리샘플링한 선박위치정보를 이용하여 선박 분포도를 제작하였다. 2018년의 V-Pass 자료를 그대로 사용하면 선박위치정보를 더 많이 전송한 선박이 있는 곳의 밀도가 올라가는 오류가 발생하지만 리샘플링한 선박위치정보를 이용하여 선박 분포도를 작성한 결과는 선박의 분포를 정확하게 나타낼 수 있다.

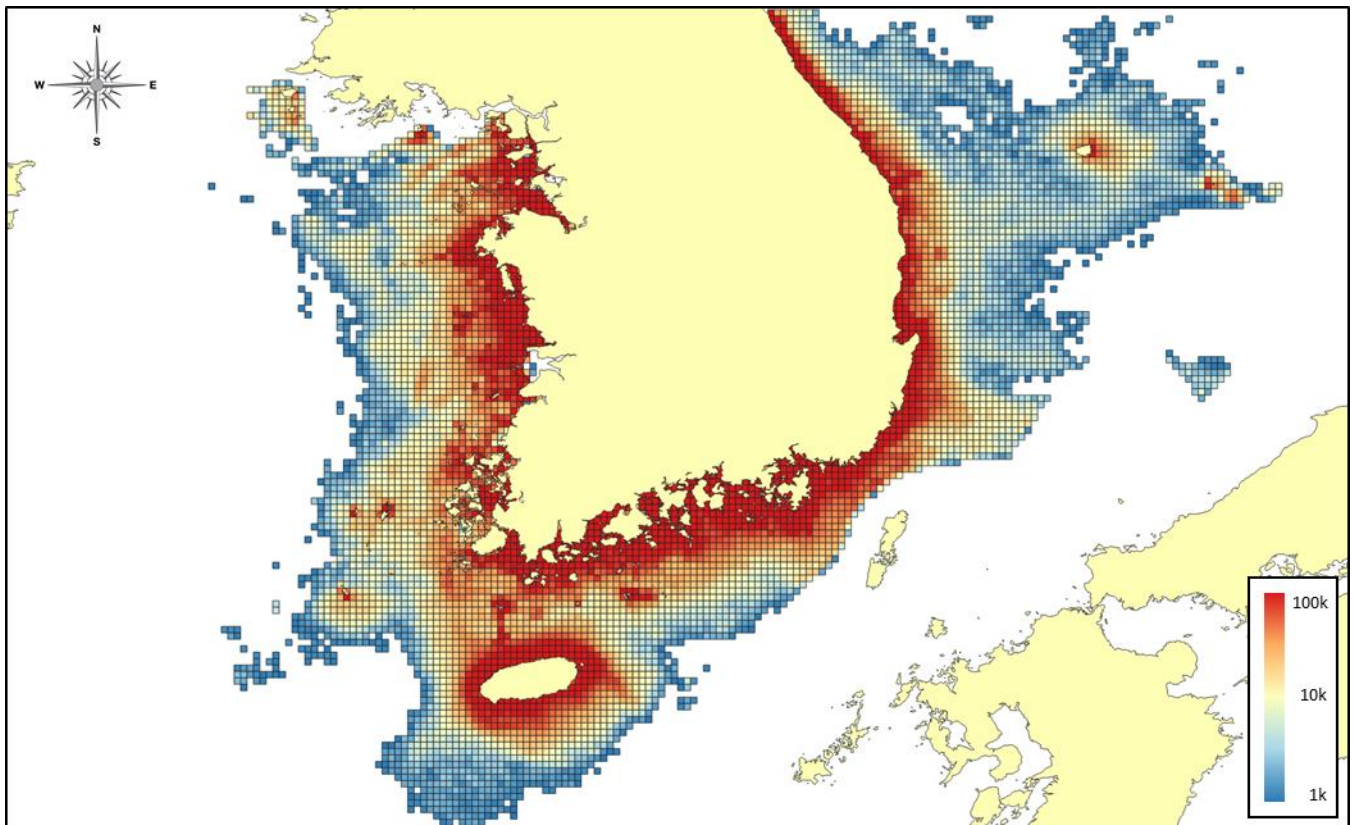


Figure 6. Vessel distribution map of 2018 V-Pass data.

Fig. 6은 2018년 V-Pass 자료를 활용하여 제작한 2018년 선박 분포도를 나타내고 있다. 최대 값과 최소 값이 큰 차이를 보이고 있기 때문에 Log를 취한 값을 단계 구분 심볼로 표현하였다. V-Pass의 수신 범위가 연안으로부터 최대 60 km이기 때문에 먼 바다의 신호는 잡히지 않는 것을 확인할 수 있었다. 대부분의 어선들이 연안 및 섬 주변에서 이동 및 어업활동을 하고 있음을 확인할 수 있었으며, 일본과의 가상 해양경계 근처에 많은 선박이

분포하고 있는 것을 확인할 수 있었다. 대청도 꽃게, 서해 오징어 및 울릉도 오징어 어업 등 주요 조업 장소도 함께 확인할 수 있었다.

3.2 선박 분포 예측을 위한 제주도 북부해역의 학습용 데이터셋

선박 분포를 예측하기 위한 학습용 데이터셋은 제주도 북부 해역을 대상으로 25개 격자에 대하여 365일을 10분 단위로 리샘플링한 선박 밀집도와 함께 물때, 기상특보, 수온, 염분, 해상풍 및 파랑 정보를 포함하여 총 1,314,000개를 시범적으로 구축하였다. 기상특보 정보는 풍랑주의보와 풍랑경보를 1, 2 값으로 입력하고 태풍주의보와 태풍경보를 3, 4로 입력하여 바람의 강도에 따른 값을 적용할 수 있도록 하였다. 다만 구축된 해양환경정보의 경우 모델링 간격이 1 시간에서 1 일 간격으로 선박위치정보 리샘플링 간격인 10 분과 많은 차이가 있었다. 구축된 학습용 데이터셋은 Fig. 7과 같다.

| MessageTime | GR1 | Date_ | Time_ | Weekday | APM | Hour_ | Lunar | LunarDay | WaterTime | Gale | Temp | Sal | Wind | Wave | cnt |
|------------------|-------------|------------|----------|---------|-----|-------|------------|----------|-----------|------|------|------|------|------|-----|
| 2018-01-04 21:50 | GR3_F4K14_L | 2018-01-04 | 21:50:00 | Thu | PM | 21 | 2017-11-18 | 18 | 10 | 0 | 19.1 | 33.2 | 8.9 | 0.8 | 0 |
| 2018-01-04 21:50 | GR3_F4K14_M | 2018-01-04 | 21:50:00 | Thu | PM | 21 | 2017-11-18 | 18 | 10 | 0 | 19.1 | 33.2 | 8.2 | 0.9 | 0 |
| 2018-01-04 21:50 | GR3_F4K14_N | 2018-01-04 | 21:50:00 | Thu | PM | 21 | 2017-11-18 | 18 | 10 | 0 | 19.1 | 33.2 | 7.6 | 0.9 | 0 |
| 2018-01-04 21:50 | GR3_F4K14_O | 2018-01-04 | 21:50:00 | Thu | PM | 21 | 2017-11-18 | 18 | 10 | 0 | 19.1 | 33.2 | 8.6 | 0.9 | 0 |
| 2018-01-04 21:50 | GR3_F4K14_P | 2018-01-04 | 21:50:00 | Thu | PM | 21 | 2017-11-18 | 18 | 10 | 0 | 19.1 | 33.2 | 8.3 | 1 | 0 |
| 2018-01-04 21:50 | GR3_F4K14_Q | 2018-01-04 | 21:50:00 | Thu | PM | 21 | 2017-11-18 | 18 | 10 | 0 | 19.1 | 33.2 | 7.7 | 1 | 0 |
| 2018-01-04 21:50 | GR3_F4K14_R | 2018-01-04 | 21:50:00 | Thu | PM | 21 | 2017-11-18 | 18 | 10 | 0 | 19.1 | 33.2 | 6.9 | 1 | 0 |
| 2018-01-04 21:50 | GR3_F4K14_S | 2018-01-04 | 21:50:00 | Thu | PM | 21 | 2017-11-18 | 18 | 10 | 0 | 19 | 33.2 | 5.2 | -1 | 0 |
| 2018-01-04 21:50 | GR3_F4K14_T | 2018-01-04 | 21:50:00 | Thu | PM | 21 | 2017-11-18 | 18 | 10 | 0 | 19 | 33.2 | 8.6 | -1 | 0 |
| 2018-01-04 21:50 | GR3_F4K14_U | 2018-01-04 | 21:50:00 | Thu | PM | 21 | 2017-11-18 | 18 | 10 | 0 | 19.1 | 33.2 | 8.1 | 1.1 | 0 |
| 2018-01-04 21:50 | GR3_F4K14_V | 2018-01-04 | 21:50:00 | Thu | PM | 21 | 2017-11-18 | 18 | 10 | 0 | 19.1 | 33.2 | 7.5 | 1.1 | 2 |
| 2018-01-04 21:50 | GR3_F4K14_W | 2018-01-04 | 21:50:00 | Thu | PM | 21 | 2017-11-18 | 18 | 10 | 0 | 19.1 | 33.2 | 6.8 | 1.1 | 2 |
| 2018-01-04 21:50 | GR3_F4K14_X | 2018-01-04 | 21:50:00 | Thu | PM | 21 | 2017-11-18 | 18 | 10 | 0 | 19 | 33.2 | 5 | -1 | 0 |
| 2018-01-04 21:50 | GR3_F4K14_Y | 2018-01-04 | 21:50:00 | Thu | PM | 21 | 2017-11-18 | 18 | 10 | 0 | 19 | 33.2 | 4.3 | -1 | 0 |
| 2018-01-04 22:00 | GR3_F4K14_A | 2018-01-04 | 22:00:00 | Thu | PM | 22 | 2017-11-18 | 18 | 10 | 0 | 18.9 | 33.2 | 8.8 | 0.8 | 0 |
| 2018-01-04 22:00 | GR3_F4K14_B | 2018-01-04 | 22:00:00 | Thu | PM | 22 | 2017-11-18 | 18 | 10 | 0 | 18.9 | 33.2 | 9.1 | 0.7 | 0 |

Figure 7. Example of training dataset for vessel distribution prediction.

Fig. 7의 MessageTime 항목은 일시를 의미하며, GR1 항목은 격자 번호, Date_, Time_, Weekday 항목은 날짜, 시간과 요일, APM 항목은 오전 오후를 의미한다. Hour_는 시간만을 별도로 분류한 항목이며, WaterTime은 물때를 의미한다. Gale 항목은 풍랑주의보 및 태풍주의보를 나타내며, Temp, Sal, Wind, Wave 항목은 각각 수온, 염분, 풍속, 파랑을 의미한다. cnt 항목은 해당 격자 내에 포함된 선박 수를 나타낸다. Time_과 Hour_의 경우 Time_은 144 단계의 구분을 가진 시간 흐름이고, Hour_는 24단계의 구분을 가진 시간 흐름으로 학습 조건에 따라 활용할 수 있도록 항목을 분류하여 구축하였으며 학습 모델에 따라 항목을 변경하여 사용할 수 있도록 하였다. 날짜와 시간 항목은 DateTime형식의 자료로 구축하였고, 해양환경 자료는 실수형 자료로 구축하였으며 Gale 항목은 범주형 자료로 구축하였다.

학습용 데이터셋을 이용하여 인공지능 학습을 수행할 수 있는 기반을 제공할 수 있으며, 구축된 자료의 항목을 활용한 앙상블 머신러닝에 활용하거나, 격자밀도를 Fig. 6과 같이 이미지로 변환하여 시간의 흐름에 따른 공간적 변화를 연속적으로 모델링 할 수 있는 Convolutional LSTM (Long-Short Term Memory) 모델을 활용하여 선박의 분포를 예측할 수 있을 것으로 판단된다.

4. 결론 및 토의

본 연구에서는 V-Pass 자료를 활용하여 선박 분포 예측을 위한 인공지능 학습용 데이터셋을 구축하였다. 학습용 데이터셋 구축을 위해 자료의 시간 간격을 일정하게 변환하기 위한 방법으로 선박위치정보 전송주기를 리샘플링하는 방법을 적용하였으며, 그 결과 제주도 북부 해역에 대하여 25개 격자를 대상으로 1,314,000개의 학습용 데이터셋을 구축할 수 있었다. 학습 보조자료로 어업 활동 및 선박의 출항에 영향을 줄 수 있는 기상특보, 물때 및 해양환경 자료를 수집하여 학습용 데이터셋으로 구축하였다.

선박위치정보를 이용하여 구축한 선박 분포도를 활용하여 월별 일별 주기별 분포 지도를 제작함으로써 해양안전 및 안보 분야의 의사결정에 적극적으로 활용할 수 있음을 확인하였다. 향후 해양환경 정보의 모델링 결과의 간격을 줄이기 위한 방법으로 한국해양과학기술원의 KOOS 모델링 자료를 수집하여 학습 보조자료를 보완하고, 선박 분포 예측 모델링을 수행하여 그 정확도를 비교하는 연구가 수행되어야 할 것이다.

5. 사사

본 연구는 해양경찰청 R&D 과제인 "위성연계 접경수역 선박 모니터링 및 분포 예측 체계 개발" 사업의 일환으로 수행되었습니다.

6. 참고문헌

Hwang HG et al. (2016) A Development of Analysis System for Vessel Traffic Display and Statistics based on Maritime-BigData. JKIIICE 20(6): 1195-1202

IMO (1998) Resolution MSC.74(69), Annex 3, Recommendation on performance standards for an universal shipboard automatic identification system (AIS), pp 13-16

Kim BO (2014) A Study on the reporting intervals of shipborne AIS dynamic data, Proceedings of the Korean Institute of Information and Communication Sciences Conference, October 2014(10a): 305-308

Lim CH (2014) A Study on the Law Enforcement of Korea Coast Guard against the Illegal Chinese Fishing Vessels. Korean Soc. Mar. Environ. Saf. 20(1): 49-58

Pokrajac D, Filjar R, Desic S (2005) Provision of generalais functionalities in internet AIS. IEEE Intelligent Transportation Systems, Vienna, Austria, 16 Sep 2005

Ryu YH, Park KG, Kim HY (2014) A Study on the Advancement Structure Model of Maritime Safety Information System (GICOMS) using FSM. Journal of Korean Institute of Intelligent Systems 24(3): 337-342

7. 데이터셋에 대한 메타데이터

V-Pass 데이터는 개인정보를 포함하고 있어 공개가 어려운 실정이므로, 데이터 활용을 원할 경우 해양경찰청 및 해양수산부에 데이터의 활용목적과 범위를 명시한 자료요청 공문을 발송하여 허가를 득한 후 활용 가능하다.

| Sort | Field | Subcategory#1 | Subcategory#2 | |
|-------------|---------------------------|---------------------------|---|--|
| Essential | *Title | Vessel location | The dataset of V-Pass | |
| | *DOI name | | | |
| | *Category | Oceans | | |
| | Abstract | Vessel location | The dataset of V-Pass | |
| | *Temporal Coverage | 01/Jan/2018 ~ 31/Dec/2018 | | |
| | *Spatial Coverage (WGS84) | WGS84 | 126.25, 33.75 / 126.5, 33.5 | |
| | *Personnel | Name | | Yonggil Park |
| | | Affiliation | | Korea Institute of Ocean Science and Technology |
| | | E-mail | | ygpark32@kiost.ac.kr |
| *CC License | | | | |
| Optional | *Project | Korea Coast Guard | Development of satellite based system on monitoring and predicting ship distribution in the contiguous zone | |
| | *Instrument | | | |