

Original Article

LSTM-Autoencoder를 이용한 부산 해역 해파리 대량 발생 예측 및 조기 경보 방안 연구

류제완¹ · 김태윤^{2,†}

¹한국환경연구원 환경평가본부 초빙연구원

²한국환경연구원 환경평가본부 선임연구위원

Prediction and Early Warning of Jellyfish Bloom in Busan Coastal Region based on LSTM-Autoencoder

Jewan Ryu¹ and Taeyun Kim^{2,†}

¹Invited Research Fellow, Environmental Assessment Group, Korea Environment Institute, Sejong 30147, Korea

²Chief Research Fellow, Environmental Assessment Group, Korea Environment Institute, Sejong 30147, Korea

요 약

최근 기후변화와 환경오염 등이 원인이 되어 해파리 대량 발생이 지속적으로 증가하고 있다. 특히 노무라입깃해파리는 국내에 대량 발생으로 피해를 입히는 주요 종으로 지적되며 적절한 대응과 조기 경보의 필요성이 대두되었다. 본 연구는 부산 해역의 해파리 대량 발생에 영향을 주는 수온, 유속 등 다양한 시계열 데이터를 해양수질자동측정망과 해수관측부이 데이터로부터 수집하고 LSTM-AE 모델을 활용하여 노무라입깃해파리의 대량 발생을 사전에 감지하고자 하였다. 학습을 마친 LSTM-AE 모델은 조기 경보를 위한 복원 오차를 0.7869로 설정하였을 때 0.6162의 정밀도 및 재현율, 그리고 0.7469의 정확도를 보였다. 제시된 LSTM-AE 모델을 활용하여 해파리 대량 발생을 2주 이전에 감지가능한 것이 확인되어 추후 노무라입깃해파리의 대량 발생으로 인한 피해 저감에 기여할 수 있을 것으로 기대된다.

Abstract – Recently, the climate change and environmental pollution in ocean have been resulted in jellyfish blooms. In particular, the *Nemopilema nomurai* is indicated as one of major jellyfish species to have damage by its bloom in South Korea so that it requires the appropriate response and early warning. This study collected various time series data in Busan coastal region such as seawater temperature and current speed which affect jellyfish blooms. Our LSTM-AE model learned collected time series data after appropriate preprocessing and succeeded to pre-detected blooms of *Nemopilema nomurai* two weeks before. The reconstruction error for early warning was suggested as 0.7869 and our LSTM-AE model showed 0.6162 of both precision and recall, and 0.7469 of accuracy with given reconstruction error. The LSTM-AE model succeeded to detect jellyfish blooms and it is expected to contribute damage reduction in the future.

Keywords: Jellyfish bloom(해파리 대량발생), Artificial intelligence(인공지능), LSTM-Autoencoder(LSTM-오토인코더), Anomaly detection(비정상 탐지), Marine Environmental Monitoring System(해양환경측정망)

1. 서 론

최근 기후변화와 환경오염, 수산자원 고갈 등의 이유로 국내 해역에서 해파리가 대량 발생하는 사례가 지속적으로 증가하고 있다. 1960년대 이후 점차 해양이 해파리 대량 발생에 적합한 환경으로 조성되면서 대량 발생이 증가하기 시작하였고(Parsons and Lalli

[2002]), 국내에서는 1990년 이후부터 보름달물해파리가, 2000년대 후반부터 노무라입깃해파리가 대량으로 발생하기 시작하였다(Kim *et al.*[2014]). 해파리는 생식능력이 탁월하여 먹이생물이 풍부한 곳에서 급속히 증가하며 대량 발생이 일어날 시 어족자원을 포식하여 해양생태계에 부정적인 영향을 끼친다. 또한, 해양에 위치한 산업시설(발전소 등)의 취수구를 막히게 하여 산업적 피해를 유발하기도 한다. 독성을 가진 해파리는 여름철 해수욕장 이용객을 쏠아 인명 피해가 발생하기도 하는 등 해파리의 대량 발생은 다양

[†]Corresponding author: kimty@kei.re.kr

한 환경적·경제적·사회적 문제를 발생시킨다. 우리나라에서 출현하고 있는 해파리는 모두 124종으로 보고되고 있으며, 약 100여 종이 독성 해파리로 알려져 있다(Wrobe and Mills[1998]; Park *et al.*[2015]; Yu[2016]; Lee *et al.*[2019]).

특히, 우리나라 연근해에서는 여름철 중국으로부터 발생한 것으로 추정되는 노무라입깃해파리가 각별한 주의가 필요한 종으로 인식되고 있다. 타 해파리와 비교하여 더 큰 체장을 가져 그물에 쉽게 걸리는 특징이 단점으로 작용하여 무거운 하중으로 그물을 끊어내야 하는 등 어업에 심각한 피해를 유발한다. Yoon *et al.*[2012]의 연구에 따르면 2012년 노무라입깃해파리의 대량 발생은 어획량감소, 상품가치하락, 어구파손 등 어업 전 분야에 상당한 피해를 끼쳤으며 보름달물해파리와 더불어 대부분의 어업 피해를 유발한 원인으로 조사되었다. 또한, 강력한 독성을 지니고 있어 해수욕장 이용객을 쏘아 사망에 이르게 한 사례도 있다.

국내에서는 해파리 대량 발생으로 인한 피해를 최소화하기 위해 다양한 노력을 기울이고 있다. 「해양생태계 관리 및 보전에 관한 법률」에서는 해파리를 유해해양생물로 지정하여 해파리 부착유생 제거 및 모니터링을 실시하도록 한다. 국립수산물과학원에서 운영하는 해파리정보시스템을 통해 해파리 모니터링이 시행되고 있다. 해파리 모니터링 자료에는 다양한 해파리들이 조사되고 있으며, 지역별 출현율은 노무라입깃해파리, 보름달물해파리, 기타해파리로 분류하여 산정하고 있다(Park *et al.*[2015]; Oh *et al.*[2021]; National Fisheries Research and Development Institute[2022]; Marine Environment Information Portal[2022]). 또한, 수층별 분포 현황(Kim *et al.*[2015]), 해파리 피해 저감 장치의 효과성 분석(Park *et al.*[2015]), 기후요소를 고려한 출현 현황 분석(Song *et al.*[2015])과 같은 다양한 연구가 진행되었다.

본 연구에서는 유해해양생물로 지정되어 심각한 피해를 입히는 노무라입깃해파리의 대량 발생을 예측하는 인공지능망 모형을 구축하고자 한다. 인공지능망 모형은 부산 해역의 노무라입깃해파리 대량 발생 징후를 사전에 감지하는 것이 목표이며, 해양수질자동측정망과 해양관측부이 데이터를 학습하여 해파리 모니터링 주간보고 자료에서 추출한 노무라입깃해파리 대량 발생 여부를 예측하도록 한다.

2. 이론적 배경

2.1 노무라입깃해파리의 거동 및 대량 발생 특성

노무라입깃해파리의 대량 발생 거동은 동중국해에서 발생하여 5월경 최초 발견되며, 발생 이후 서해, 남해, 동해로 이동하는 것으로 알려져 있다(Uye[2013]). Yoon *et al.*[2012]의 연구에 따르면 노무라입깃해파리 기원 추정지의 동계~춘계 수온을 분석한 결과 노무라입깃해파리 대량 출현 시기의 수온이 비출현 시기보다 높은 것으로 확인되어 수온이 노무라입깃해파리의 대량 발생을 일으키는 요인 중 하나로 추정된다. 또한, 노무라입깃해파리의 생활사는 부유생활시기와 고착생활시기로 구분되는데 고착생활시기에 무성생

식을 통해 개체수를 늘릴 수 있으며, 이때 수온, 염분, 먹이, 광조건 등이 주요 인자로 작용할 가능성이 있다. 특히, 노무라입깃해파리의 주요 무성생식 방법인 podocyst 생성은 20°C이상의 비교적 높은 수온에서 주로 이루어진다. 따라서 해수 유동, 기상 조건, 해양수질 등 노무라입깃해파리의 유영 및 발생과 생장에 영향을 끼치는 자료를 인공신경망의 입력자료로 활용가능할 것으로 기대된다.

2.2 LSTM-AE 모델을 이용한 비정상 데이터 검출

최근 인공지능 기술의 발달로 다양한 분야에 인공지능 기술이 적용되기 시작하면서 해파리 출현 예측, 해파리 무리 인식 등의 연구에도 활용되었다. 한 선행연구는 심층신경망(Deep Neural Network, DNN)을 사용하여 기온, 기압, 수온, 유속, 풍속, 염분, 조위, 파고를 특징(feature)으로 활용하여 해파리의 출현을 예측하였다(Hwang and Han[2019]). 출현율의 학습 및 검증은 국립수산물과학원의 해파리 모니터링 주간보고 자료를 이용하였으며, 해운대 해수욕장을 연구해역으로 예측을 진행하였다. 특히, 원본 데이터가 247개로 학습에 충분치 않은 점을 보완하기 위해 부트스트래핑을 통해 부족한 데이터를 생성하고 예측 모형의 정확도를 향상시켰다. Jeon *et al.*[2019]는 합성곱신경망(Convolutional Neural Network, CNN)을 이용하여 해파리 군집 인식 연구를 수행했다. 해양감시선에서 녹화한 영상을 분석하여 해양쓰레기와 해양유해생물을 구별하였다. 경상남도 마산 진동 바다에서 촬영한 영상을 학습 데이터로 사용하여 해파리 군집 인식 CNN 모형을 구축하였다.

Malhotra[2016]에 의해 제안된 LSTM-Autoencoder(이하 LSTM-AE)는 오토인코더와 장단기메모리(Long short-term memory, LSTM) 모델이 결합된 것으로 두 모델의 장점을 모두 가져 시계열의 비정상 데이터 검출에 효과적이다. 오토인코더는 비지도학습의 한 가지로 인코더와 디코더로 구성되며 예측(prediction), 분류(classification), 비정상 데이터 검출(anomaly detection) 등 다양한 분야에 널리 사용되는 모델 중 하나이다. Fig. 1은 LSTM-

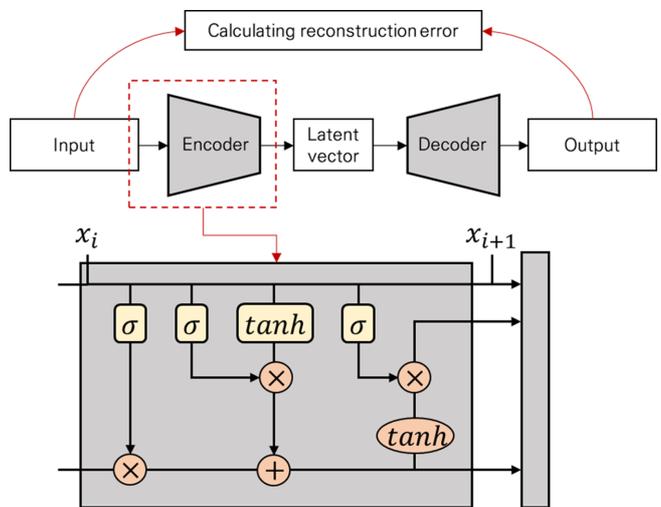


Fig. 1. Architecture of LSTM-AE model.

Autoencoder 모델을 도식화한 것으로 기존의 오토인코더 모델과 같이 인코더와 디코더로 이루어져 있다. 인코더를 통과한 입력 데이터가 특징벡터(latent vector)로 압축되고, 다시 디코더를 통해 복원하는 방법으로 학습이 이루어진다. 따라서, 정상 데이터에 대해서만 학습이 이루어진 후, 비정상 데이터가 입력 데이터로 주어질 시 복원된 출력 데이터와 입력 데이터 사이의 복원 오차(reconstruction error)가 크게 나타난다. 인코더와 디코더에 기존의 장단기메모리(Long Short-term memory) 모델을 활용하여 특히 신호, 시계열과 같은 시퀀스 형태의 데이터를 효율적으로 학습할 수 있다는 특징을 갖는다. 즉, 시계열 데이터에 해당하는 해수 거동, 기상, 해양 수질 등 해파리 대량 발생의 영향 인자들을 효과적으로 학습 가능하기 때문에, 시계열 데이터를 통해 해파리 대량 발생을 조기 감지하려는 본 연구에 적합한 모델로 판단된다.

3. 연구 방법

3.1 분석 장소 및 데이터 선정

Uye[2013]와 Yoon *et al.*[2012]의 선행연구에서 밝혀진 바와 같이 노무라입깃해파리는 동중국해에서 발생하여 우리나라의 서해, 남해, 동해로 순차적으로 이동하기 때문에 시기의 차이는 있지만 국내 대부분 해역에서 노무라입깃해파리의 대량 발생을 관측할 수 있다. 해파리 모니터링 주간보고 자료가 모니터링 요원의 활동을 기반으로 작성된다는 점에서 미루어 볼 때, 해파리 모니터링 요원의 수가 많은 지역일수록 노무라입깃해파리의 출현율이 더 정확할 것으로 기대할 수 있다. 또한, 해역 특성을 반영하기 위해서는 해양관측부이와 해양수질자동측정소의 위치가 동일한 해역에 속해야

하며 가까울수록 해역 특성이 더 정확하게 반영될 것으로 판단되었으며, 해양수질자동측정소의 경우 측정소에 따라 측정 항목이 달라지기 때문에 측정 항목의 수가 분석 해역 선정의 중요한 기준이 되었다. 해파리 모니터링 요원의 수, 데이터 결측, 측정 항목의 수, 해양관측부이와 해양수질자동측정소의 공간적 근접성 등을 고려하여 부산 해역을 해파리 대량 발생 예측 대상 지점으로 선정하였다. 부산 해역의 해양수질자동측정 자료는 낙동명지 측정소의 데이터를 활용하였으며, 해양관측부이 자료는 해운대 해수욕장의 측정소 데이터를 활용하였다.

3.2 데이터 수집 및 전처리

상기 서술한 바와 같이 LSTM-AE 모델의 학습을 위해 국립수산과학원의 해파리 모니터링 주간보고 자료, 해양환경정보포털의 해양수질자동측정망 데이터, 국립해양조사원의 바다누리 해양정보서비스에서 제공되는 해양관측부이 측정 데이터를 활용하였다. 해파리 모니터링 주간보고 자료와 해양관측부이 측정 데이터는 OpenAPI가 제공되어 이를 활용하였으며, 해양수질자동측정망 데이터는 홈페이지에서 제공되는 원본 데이터를 다운로드하여 활용하였다.

국립수산과학원은 해파리 출현 및 이동경로를 모니터링 하기 위해서 2004년부터 어민과 지자체 및 관련 공무원으로 구성된 해파리 모니터링 네트워크를 운영하고 있다. 해파리 모니터링 주간보고 자료는 이러한 모니터링 활동을 통해서 확인된 한국 연근해 해파리 출현 중, 출현 위치, 출현율 등의 자료를 포함하고 있다. 본 연구에서는 해파리 모니터링 주간보고 자료에서 제공되는 부산 지역 노무라입깃해파리 출현율을 수집 및 활용하였다. 해파리 모니터링 주간보고는 대략 5월부터 12월까지 이루어지며, 주간보고가 이루어

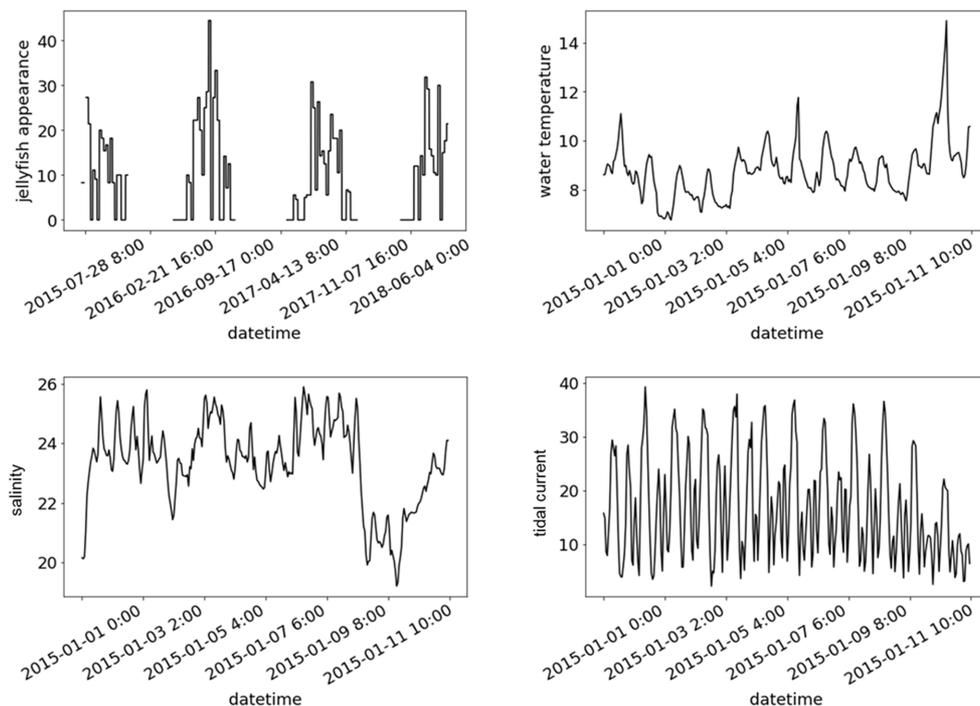


Fig. 2. Graph representation of collected data.

어지지 않는 시기의 노무라입깃해파리 출현율은 직관적으로 0으로 가정할 수 있다.

해양수질자동측정망의 측정 항목에서 수온, 일사량, 기온, 풍속, 염분, 화학적 산소 요구량을 입력값으로 활용하였고, 해양관측부이의 측정 항목에서 파고, 풍속, 유속을 입력값으로 활용하였다. 2015년 1월 1일부터 2021년 12월 31일까지의 데이터를 수집하였고, 분석에는 2015년 1월 1일부터 2018년 9월 31일까지의 데이터를 사용하였다. 2018년 10월부터 현 시점까지의 데이터는 측정 항목이 누락되거나 결측치가 지나치게 길어 사용할 수 없는 점 등의 이유로 분석에서 제외되었다. 해양수질자동측정망과 해양관측부이는 모두 5분 간격으로 측정되고 있다. 통신 장애, 유지보수 등의 원인으로 결측값이 측정 항목에 따라 다양한 길이로 존재하며, 연속적으로 긴 시간에 걸쳐 발생한 결측값의 경우 단순한 보간법을 사용할 수 없게 하는 제약이 된다. 본 연구에서는 결측치 및 노이즈의 영향을 일차적으로 감소시키기 위해서 일평균 데이터를 입력값으로 사용하였으며, 또한 KNN(k-nearest neighbor) 알고리즘을 사용하여 결측치를 보간하였다. 전처리가 완료된 데이터 세트는 1,369의 길이를 갖는 시계열 데이터이다. Fig. 2는 수집된 원 데이터 일부의 그래프로 부산해역 노무라입깃해파리 출현율, 해양수질자동측정망의 수온과 염분, 해양관측부이의 유속 데이터를 대표적으로 나타내었다. 노무라입깃해파리 출현율 외의 데이터는 가시성을 위해 가장 앞선 300개의 데이터만 표시되었다. 측정 항목 별로 측정 주기가 다르므로 측정 주기에 따라 그래프에 표현되는 일시가 다른 점에 유의해야 한다.

3.3 LSTM-AE 모델 구성 및 학습

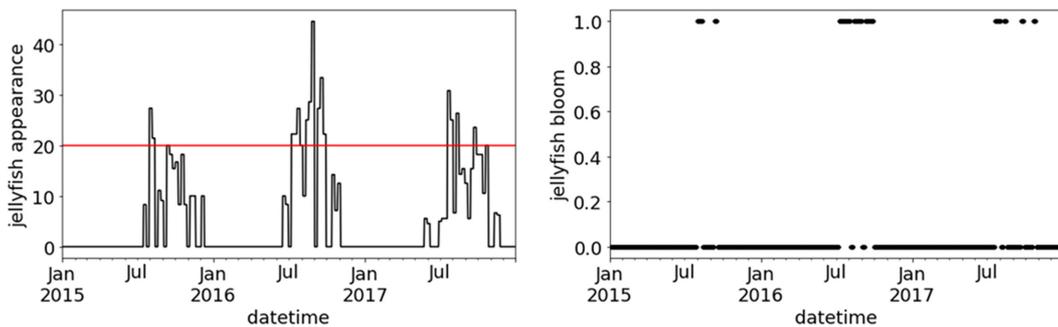


Fig. 3. Jellyfish appearance binarized by a threshold to indicate jellyfish bloom.

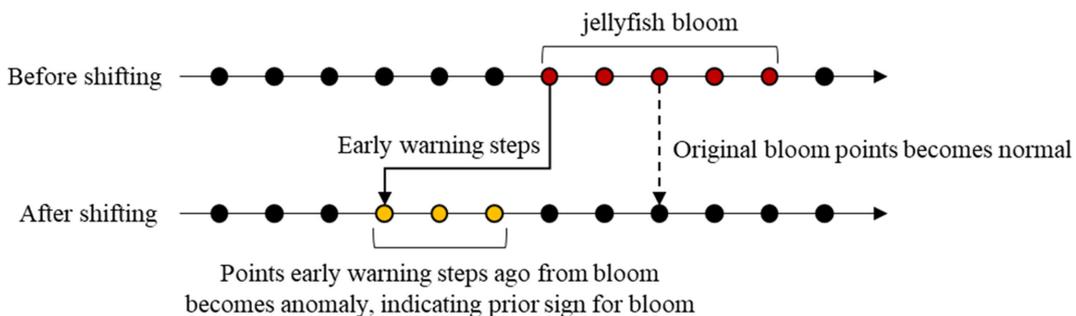


Fig. 4. Description of data shifting to learn the prior sign for jellyfish bloom.

수집 및 전처리가 완료된 전체 데이터 세트를 학습, 검증, 테스트 데이터로 나누기 위해 일반적으로 사용되는 비율이 있으나, 본 연구에서는 학습, 검증, 테스트 데이터에 모두 해파리 대량 발생 시기가 포함되게 하도록 2015년 1월 1일부터 2016년 12월 31일까지의 데이터를 학습 데이터로, 2017년 1월 1일부터 2017년 12월 31일까지를 검증 데이터로, 2018년 1월 1일부터 2018년 9월 30일까지의 데이터를 테스트 데이터로 사용하였다.

본 연구에서는 노무라입깃해파리의 대량 발생을 사전에 예측하는 것이 목표이므로, 조기 경보 기간을 설정하여 조기 경보 기간 이전부터 해파리 대량 발생 시점까지 대량 발생의 징후가 나타난다고 가정하고 이를 학습하도록 하였다. 국내에서 해파리 주의보는 단위 면적 당 해파리 개체수를 기준으로 발령되며, 해파리 출현율에 기반한 발령 기준은 전무하기 때문에, 부득이하게 본 연구에서는 해파리 출현율 20%를 해파리 대량 발생의 기준으로 설정하였다. Fig. 3의 좌측 그래프는 해파리 모니터링 요원의 활동으로 산정된 부산해역 노무라입깃해파리 출현율을 의미하고, 우측 그래프는 20% 이상의 출현율을 대량 발생으로 판단하여 이진화한 것이다. 즉, 좌측의 그래프에서 출현율이 20% 이상일 경우 대량 발생으로, 20% 미만일 경우 대량 발생하지 않음으로 판단하고 우측 그래프에서 각각 1.0과 0의 값을 갖도록 가공한 것을 의미한다. 수집된 시계열 데이터가 내포하고 있는 해파리 대량 발생의 사전 징후를 나타내기 위해서 Fig. 3과 같이 이진화된 해파리 출현율을 Fig. 4와 같이 가공하여 비정상 데이터로 판단한다. 해파리 대량 발생 기간 중의 데이터는 이미 해파리 대량 발생이 일어난 시점이므로 정상 데이터로 판단한다. 해파리 대량 발생이 연속적으로 일어날 시 가장 먼저 관측된 대량 발생 시점의 사전 징후가 연속된 대량 발생 시기를 대

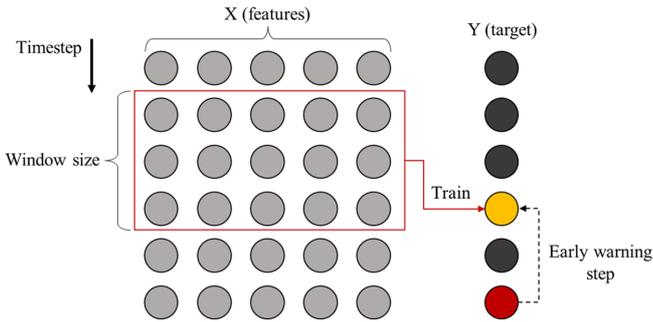


Fig. 5. Description of input data to LSTM-AE model.

표한다. Fig. 5는 LSTM-AE모델의 입력 데이터 구조를 나타낸다. 학습에 사용된 입력 변수는 총 8개가 사용되었고, 해파리 대량 발생 여부를 타겟으로 학습이 이루어졌다. 전체 시계열 데이터는 순차적으로 Window size의 길이를 갖는 시계열로 나누어져 LSTM-AE 모델의 가중치를 업데이트한다. 또한, 해파리 대량 발생을 사전에 예측하기 위해 Early warning step을 도입하여 일정 시간 이후의 해파리 대량 발생 여부를 예측하도록 하였다. 즉, 해파리 모니터링 주간 보고 자료의 노무라입깃해파리 부산 해역 출현율을 Fig. 3과 같이 적절한 기준을 통해 대량 발생 여부를 이진화하고, Fig. 4와 같이 조기 경보를 위한 사전 징후를 내포하는 비정상 데이터로 나타낸다. 이후 최종적으로 Fig. 5와 같이 LSTM-AE 모델이 입력 데이터를 학습하여 설정된 조기 경보 기간 이후의 해파리 대량 발생 여부를 예측하도록 하였다.

LSTM-AE 모델의 학습에 사용된 하이퍼파라미터가 Table 1에 정리되어 있다. 해파리 대량 발생으로 판단하는 출현율은 20% 이상으로 설정하였고, 조기 경보 기간은 해파리 출현율이 주간으로 보고되는 점을 감안하여 2주로 설정하였다. LSTM-AE 모델의 인코더와 디코더에 포함된 LSTM 셀의 은닉층은 4개의 뉴런으로 구성하였으며, 활성화함수는 Relu 함수를 사용하였다. 과적합을 방지하기 위해 20%의 확률로 뉴런이 학습에 사용되지 않도록 드롭아웃을 적용하였고, 손실 함수는 평균절대오차(Mean Absolute Error, MAE)를 사용하였다. 옵티마이저는 Adam을 사용하였고 배치 사이즈를 2로 설정하여 가중치 업데이트가 이루어지도록 하였다. 최대

Table 1. Experimental setup and hyperparameters for LSTM-AE model

Appearance threshold	equal or more than 20%
Early warning steps	14 days
Window size	7 days
Hidden neurons	4 for both encoder and decoder
Activation function	Relu
Dropout	0.2
Loss function	Mean absolute error
Optimizer	Adam
Batch size	2
Epoch	1000 (early stopping)
Learning rate	0.0001

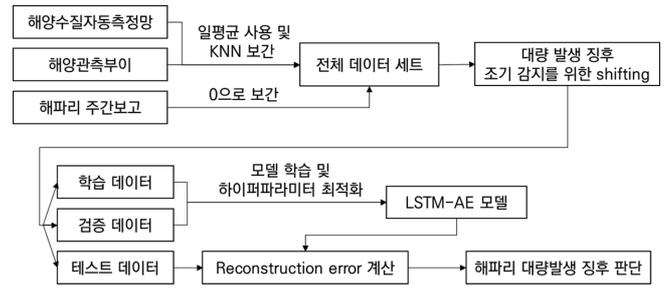


Fig. 6. Data preparation and training process.

에폭은 1000으로 설정하였으나, 조기 종료를 적용하여 검증 손실 값이 50에폭까지 개선되지 않으면 학습을 종료하도록 했다. 하이퍼파라미터는 trial-and-error를 통해 최적화되었으며, 학습 과정에서 학습 및 검증 손실의 수렴을 모니터링하여 과적합이 일어나지 않은 것을 확인하였다. 학습 과정에서 시계열 데이터의 연속성을 보존하기 위해 별도의 교차 검증은 시행하지 않았다. 데이터 전처리, LSTM-AE 모델 학습 및 비정상 데이터 검출 과정이 Fig. 6에 요약되어 있다.

4. 결과 분석 및 논의

정밀도(Precision)와 재현율(Recall) 및 정확도(accuracy)를 성능 지표로 사용하여 LSTM-AE 모델의 비정상 데이터 검출(해파리 대량 발생 징후 감지) 성능을 확인하였다. 정밀도는 수식 (1)과 같이 모델이 참으로 예측한 데이터 중 실제 참인 데이터의 비율을 의미하며, 수식 (2)의 재현율은 실제로 참인 데이터 중 모델이 참으로 예측한 비율을 의미한다. 즉, 정밀도는 LSTM-AE 모델이 2주 뒤 대량 발생이 나타날 것으로 감지한 경우 중 실제로 대량 발생이 나타난 경우의 비율을 의미하며, 재현율은 실제로 2주 뒤 해파리 대량 발생이 나타난 경우 중 LSTM-AE 모델이 대량 발생 징후를 감지해낸 비율을 의미한다. 정밀도와 재현율이 1에 가까울수록 모델이 높은 성능을 가진다.

$$\text{Precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}} \quad (1)$$

$$\text{Recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}} \quad (2)$$

$$\text{Accuracy} = \frac{\text{true positives} + \text{true negatives}}{\text{true positives} + \text{false positives} + \text{true negatives} + \text{false negatives}} \quad (3)$$

비정상 판단 기준이 되는 복원 오차의 기준값(threshold)에 따른 모델 정밀도와 재현율을 Fig. 7과 같이 나타내었다. 가로축은 비정상 판단하기 위한 복원 오차의 기준값을 의미한다. Fig. 7에서 확인할 수 있듯이 비정상 판단하기 위한 복원 오차의 기준을

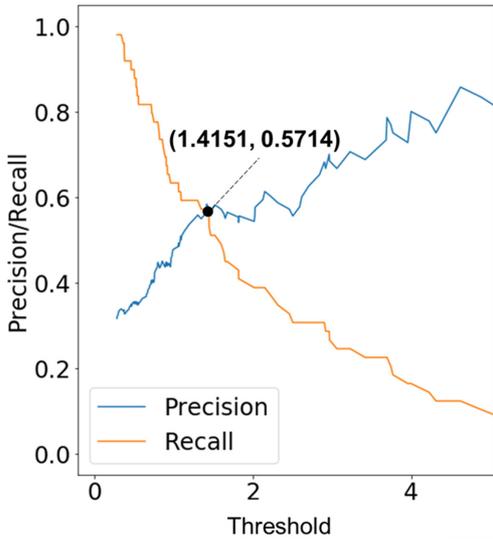


Fig. 7. Precision and recall calculated from the result of LSTM-AE model.

낮게 설정하면 모델의 정밀도는 낮고 재현율은 높다. 이는 비정상적으로 판단한 데이터의 수는 많지만 그중 실제 비정상 데이터는 낮음을 의미한다. 반면, 복원 오차의 기준을 높게 설정하면 정밀도는 높아지고 재현율은 낮아진다. 이는 모델이 찾은 데이터의 수는 작지만, 그중에서 실제 찾으려는 대상의 비율이 높은 것을 의미한다. 복원 오차의 기준값이 0일 때, 모델은 모든 데이터를 실제 해파리

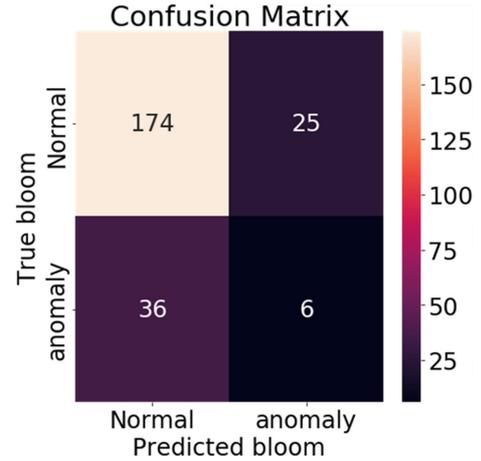


Fig. 8. Confusion matrix of LSTM-AE model.

대량 발생 여부에 상관없이 모두 대량 발생으로 판단한다. 이후 복원 오차의 기준값을 조금씩 증가시키면 대량 발생으로 판단하는 데이터가 줄어들고, 정상 상태로 판단하는 데이터가 늘어나면서 정밀도와 재현율이 교차하는 지점이 존재한다. 이 지점을 지나 계속 복원 오차의 기준값을 증가시키면 점차 더 많은 데이터를 정상 상태로 판단하게 된다. 이처럼 비정상 판단 기준이 되는 복원 오차를 지나치게 낮게 설정할 경우 모델은 대부분의 데이터를 비정상적으로 판단하게 되고, 반대로 지나치게 높게 설정할 경우 비정상 데이터까지

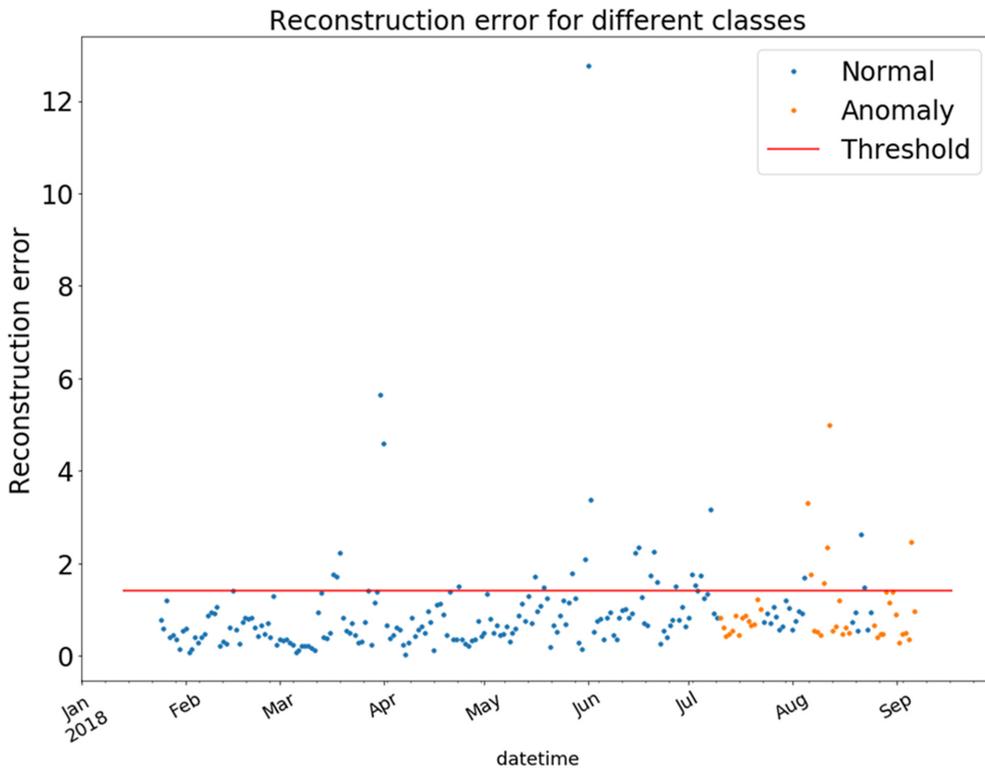


Fig. 9. Anomaly detection result of LSTM-AE model with given test set.

정상적으로 판단하게 되므로 적절한 기준을 설정할 필요가 있다. 본 연구에서는 정밀도와 재현율이 교차하는 지점의 복원 오차를 해파리 대량 발생 조기 경보의 기준값으로 제시한다. 해당 지점의 복원 오차는 약 1.4151이며, 이때 LSTM-AE 모델의 정밀도와 재현율은 약 0.5714로, 정확도는 0.7469로 계산된다.

제시된 복원 오차를 적용하여 LSTM-AE 모델의 해파리 대량 발생 예측 결과를 Fig. 8의 confusion matrix로 나타내었다. 테스트 데이터에 포함된 총 199개의 정상 데이터 중 약 87%에 해당하는 174개를 정상으로 판별하였고, 25개를 비정상적으로 판별하였다. 또한, 총 42개의 비정상 데이터 중 약 14%에 해당하는 6개를 비정상 데이터로 판별하고, 36개를 정상 데이터로 판별하였다. 비정상 데이터에 대한 예측력이 다소 부족한 것으로 판단되나, 조기 경보 기간으로 설정된 2주 이내에 비정상 데이터가 검출되었으므로 조기 경보의 활용 가능성은 충분한 것으로 보인다.

테스트 데이터로 사용된 241개 데이터를 Fig. 9에 나타내었다. 푸른 점은 정상 데이터를, 주황색 점은 비정상 데이터를 나타낸다. 비정상 판단 기준이 되는 복원 오차가 붉은 가로 실선으로 표현되었다. Fig. 9에서 확인할 수 있듯이 해파리 대량 발생 시기와 그 외 시기는 복원 오차의 차이가 명확하게 나타나지 않음에도, 어느 정도 해파리 대량 발생 시기를 예측 가능한 것이 확인된다. 해파리 대량 발생 시기와 그 외 시기를 더 명확하게 구분할 수 있는 데이터가 포함되면 모델의 성능이 향상될 것으로 기대된다.

해파리 대량 발생 기준이 되는 복원 오차를 보다 낮게 설정할 경우 더 많은 비정상 데이터를 감지할 수 있으나, 이때 정상 데이터를 비정상 데이터로 감지하는 경우가 함께 늘어나고 그로 인한 trade-off를 산정할 수 있는 근거가 부족하여 본 연구는 Fig. 7에서 설정한 것과 같이 정밀도와 재현율이 동일하게 나타나는 복원 오차를 비정상 탐지의 기준으로 제시하였다. 정상을 비정상적으로 탐지하는 경우와 비정상을 정상으로 탐지하는 경우의 손익을 비교할 수 있다면, 더 최적화된 조기 경보의 복원 오차 기준을 제시할 수 있다.

조기 경보 기간을 비교적 긴 시간으로 설정한 것도 예측력에 영향을 끼치는 요인이다. 조기 경보 기간을 길게 설정할수록 대량 발생 징후 감지에 성공할 시 유용하게 활용될 수 있으나, 모델의 전체적인 예측력은 낮아진다. 추후 연구를 통해서 조기 경보 기간을 조절하면서 모델의 예측력과 조기 경보 기간의 최적화가 이루어질 필요가 있다.

추가적인 학습 데이터가 주어지거나 데이터 전처리 방법 등에 따라 LSTM-AE 모델의 예측력이 향상될 수 있다. 특히, 본 연구에서는 해파리 대량 발생 지점과 해양 환경의 공간을 일치시키는 것에 초점을 맞추어 모든 데이터를 부산 해역에서 수집했지만, 노무라입깃해파리의 거시적인 발생 및 거동 행태를 반영할 수 있는 데이터가 학습 데이터로 주어진다면 모델의 예측력을 크게 향상시킬 수 있을 것으로 기대된다.

데이터 전처리 과정에서 해파리 주간보고가 이루어지지 않는 시기는 일반적으로 0으로 보간하였는데, 해파리 모니터링 주간보고가 이루어지는 시기에도 노무라입깃해파리 출현율이 0으로 측정된

경우가 확인된다. 선행연구에서 밝혀진 노무라입깃해파리의 발생 및 이동 행태, 과거 통계 자료, 모니터링 활동이 어민들에 의해 이루어지므로 어업 활동 여부의 영향을 받는다는 점을 고려하면 해파리 최초 발견 이후 모니터링 주간보고가 이루어지는 시기의 비출현 데이터(출현율 0)는 관측이 이루어지지 않은 결측값으로 판단하여 선형보간법을 적용하는 방법도 검토해볼 수 있다.

6. 결 론

최근 기후변화와 환경오염 등의 원인으로 해파리 대량 발생 사례가 점차 증가하고 있다. 국내 해역에서 다양한 해파리가 출현하고 있으며 특히 노무라입깃해파리는 강한 독성과 긴 체장으로 심각한 피해를 유발한다. 해파리로 인한 피해를 저감하기 위해 국내에서는 관련 법률을 제정하고 지속적인 모니터링을 수행하는 등 다양한 노력을 기울이고 있다.

본 연구는 부산 해역을 분석 지역으로 선정하고, 노무라입깃해파리를 대상으로 대량 발생에 영향을 끼칠 가능성이 있는 데이터들을 수집하여 인공지능 알고리즘에 학습시킨 후 대량 발생 징후를 사전에 감지하였다. 입력 데이터로는 해양수질자동측정망과 해양관측부의 측정 자료를 일평균 및 적절한 전처리를 마친 후 활용하였다. 타겟 데이터로는 해파리 모니터링 주간보고 자료의 부산 해역 노무라입깃해파리 출현율을 이진화 및 조기 경보 기간만큼 사전 징후를 보이는 것으로 가공하여 사용하였다.

학습을 마친 LSTM-AE 모델은 0.6162의 정밀도 및 재현율을, 0.7469의 정확도를 보이면서 정상 및 비정상 데이터를 탐지하는데 성공하였다. 이때의 해파리 대량 발생 판단을 위한 복원 오차는 0.7869로 제시되었다. 해당 모델은 정상 데이터 탐지 능력은 우수하나 비정상 데이터 탐지 능력은 다소 부족한 것으로 보인다. 그러나, 필요한 경우 해파리 대량 발생 판단을 위한 복원 오차를 조절함으로써 비정상 탐지 비율을 향상시키는 것이 가능하며, 또한 5장에서 논의된 바와 같이 학습 데이터 추가 확보와 후속 연구를 통해 예측력을 충분히 확보할 수 있을 것으로 기대된다.

후 기

본 논문은 한국환경연구원에서 환경부의 수탁과제로 수행된 「해상풍력발전 환경영향 조사·평가 고도화방안 연구 I (2022-040)」의 연구결과를 기초로 작성되었습니다.

References

- [1] Kim, B.-T., Eom, K.-H., Han, I.-S. and Park, H.-J., 2015, An Analysis of the Impact of Climatic Elements on the Jellyfish Blooms. *J. Fisheries and Marine Sciences Education*, 27(6), 1755-1763.
- [2] Hwang, C. and Han, M.-M., 2019, The Predictions of Appearance of Jellyfish through Deep Neural Network. *J. Internet Comput.*

- Serv., 20(5), 1-8.
- [3] Kim, D.-Y., Lee, J.-S. and Kim, D.-H., 2014, A Study on Direction of Industrial Utilization for Jellyfish in Korea. J. Fisheries and Marine Sciences Education., 26(3), 587-596.
- [4] Lee, D., Han, I., Chae, J., Yoon, W., Yang, Y., Ki, D. and Lee, K., 2019, Analysis of the Advantage and Disadvantage of Harmful Jellyfish's Damage Reduction Devices Strategy Types in the Beach. J. Fisheries and Marine Sciences Education, 31(4), 1230-1241
- [5] Marine Environment Information Portal 2022, <https://www.meis.go.kr/mes/marineLife/harmful/view2.do> (accessed 2022.10.23.)
- [6] Malhotra, P., Ramakrishnan, A., Anand, G., Vig, L., Agarwal, P., and Shroff, G., 2016, LSTM-based encoder-decoder for multi-sensor anomaly detection. arXiv preprint arXiv:1607.00148.
- [7] National Fisheries Research and Development Institute 2022, <https://www.nifs.go.kr/bbs?id=jellynews> (accessed 2022.10.23.)
- [8] Parsons, T.R. and Lalli, C.M., 2002, Jellyfish population explosions: revisiting a hypothesis of possible causes. La mer, 40(3), 111-121.
- [9] Song, S.H., Lee, S.-G. and Kim, H., 2015, A Study on the Management Performance of a Set Net Fishery According to the Blooming Frequency of Jelly Fish *Nemopilema nomurai* in Yeosu. J. Korean Soc Fish Ocean Technol, 51(1), 42-49.
- [10] Park, S., Lee, D.-G., Yang, Y.-S., Lee, K., Hahn, M. and Lee, T., 2015, Analysis on Underwater Stability of the Jellyfish Sting Protection Net Installed in the Heaundae Beach. J. Korean Soc Fish Ocean Technol, 51(1), 128-135.
- [11] Park, S., Lee, K., Yoon, W.-D., Lee, D.-G., Kim, S., Yang, Y.-S. and Lee, G.-H., 2015, A Study on the Damage Reduction Strategy Against a Harmful Aquatic Organism, Jellyfish's Bloom. J. Fisheries and Marine Sciences Education. 27(1), 49-62.
- [12] Oh, S., Kim, K.Y., Lim, W.A., Park, G., Oh, H., Oh, W. and Lee, K., 2021, Vertical Distribution of Giant Jellyfish (*Nemopilema nomurai*) in the Coastal Waters of Korea and its Correlation Analysis by Survey Method. J. Korean Soc Fish Ocean Technol, 57(4), 351-364.
- [13] Uye, S.-I., 2014, The Giant Jellyfish *Nemopilema nomurai* in East Asian Marginal Seas. Jellyfish blooms. 185-205.
- [14] Jeon, W.-S., Rhee, S.-Y. and Yoo, N.-H., 2019, Detection of Marine Garbage and Jellyfish Cluster using CNN and Post-processing. J. Korean Institute of Intelligent Systems, 29(1), 16-22.
- [15] Yoon, W.-D., Suh, Y.-S., Hwang, J.-D., Han, I.-S., Youn, S.-H., Lim, W.A., Lee, K.-H., Chang, S.-J., Yang, J.-Y. and Han, C.-H., 2012, Study on the Causes of and Countermeasure against Jellyfish Bloom. National Fisheries Research and Development Institute.
- [16] Wrobel, D. and Mills, C., 1998, Pacific Coast Pelagic Invertebrates: A Guide to the Common Gelatinous Animals. Monterey Bay Aquarium 108.
- [17] Yu, W.B., 2016, Study on the rDNA characteristics of harmful jellyfishes *Aurelia* sp.1 and *Nemopilema Nomurai* in Korea. Master Dissertation, Sangmyung University, Seoul, Korea, 1-54.

Received 28 November 2022

Revised 16 January 2023

Accepted 20 February 2023