스마트부표 센서 데이터의 바이오파울링 예측을 위한 AI 개발: 용존산소 시계열 데이터 기반의 순환신경망 예측 모델

Development of AI for Biofouling Prediction Using Smart Buoy Sensor Data: Recurrent Neural Network Prediction Model Based on Dissolved Oxygen Time Series Data

Abstract

기후 변화에 따라 해양 환경이 급변하면서, 해양 환경 관측 기기의 중요성이 높아지고 있다. 특히 부표에 설치된 센서와 같은 기기는 바이오파울링(biofouling)으로 인해 데이터 정확도가 저하될 수 있는데, 이는 해양 생물이 센서에 부착해 기기의 기능을 방해하는 현상이다. 현재는 주기적인 점검으로 관리하고 있으나, 환경 변화로 인해 발생 주기를 예측하기 어려워 센서 데이터의 이상 징후를 통한 사전 예측이 필요한 상황이다.

기존의 해양 환경 관측 데이터 품질 관리는 전문가의 수작업과 통계적 임계값 기반 접근에 의존해왔으나, 이는 비선형적 패턴에서의 이상 탐지에 한계가 있었으며, 특히 바이오파울링과 같은 점진적 성능 저하 현상을 조기에 감지하는 데 제약이 있었다. 본 연구는 완도 지역 3 개 부표의 용존산소 데이터를 수집하여 바이오파울링 발생 시점(DO ≤ 3.0mg/L)을 기준으로 데이터를 레이블링하고, 720 개의 연속 데이터를 하나의 시퀀스로 정의하여 학습 데이터셋을 구성하였다.

RNN, GRU, LSTM 모델을 비교 분석한 결과, GRU 모델이 F1 스코어 0.99 로 가장 우수한 성능을 보였으며, 특히 작은 hidden size(32)로도 높은 예측 정확도를 달성하였다.

본 연구는 해양 센서의 바이오파울링 예측에 있어 실용적이고 즉시 활용 가능한 AI 기반 접근법을 제시함으로써 해양 관측 시스템의 효율적 운영과 관리를 지원할 수 있음을 보여준다.

1. Introduction

최근 이상 기후 현상에 따라 해양 환경이 급격하게 변화하면서, 해양 환경 관측 기기의 중요성이 더욱 부각되고 있다. 실시간으로 수집되는 해양 데이터는 생태계 보호와 해양 자원의 효율적 관리에 가장 기본적이고 핵심적인 요소이다. 그러나 부표에 설치한 센서와 같이 해상에 직접 배치된 관측 기기는 해양에서 발생하는 외부 요인에 의해 오작동하기도 한다. 특히 지속적, 반복적으로 발생하는 바이오파울링(Biofouling)에 의해 센서의 성능이 저하되고 수집 중인 데이터의 정확도가 크게 떨어질 수 있다.

바이오파울링이란 따개비나 해조류 등 해양 생물이 선박이나 부표 등 구조물에 부착하여 본래의 기능 수행을 어렵게 만드는 현상을 말한다. 이러한 바이오파울링은 센서의 측정 정확도를 떨어뜨려 해양 데이터의 신뢰성을 저해하며, 이는 생태계 보호와 해양 자원 관리에 부정적인 영향을 미칠 수 있다. 현재는 주기적인 현장 점검 및 부착된 유기물 제거를 통해 바이오파울링 발생에 대처하고 있다. 그러나 최근 해양 환경의 급격한 변화로 인해 바이오파울링 발생 주기의 예측이 점점 어려워지고 있어, 주기적인 점검만으로는 효율적인 대응에 한계가 있다.

이러한 상황에서 센서 데이터의 이상 징후를 통해 바이오파울링을 조기에 예측하고 대응할 수 있는 방법이 필요하다. 실시간으로 센서 데이터에서 나타나는 이상 패턴을 감지하면 바이오파울링으로 인한 센서 성능 저하와 데이터 신뢰성 감소 문제를 효과적으로 보완할 수 있다. 이는 센서의 유지보수 효율을 높이고 해양 데이터의 정확성과 신뢰성을 확보하는 데 필수적이다.

최근 연구에서 바이오파울링 감지를 위해 비전(카메라) 기반 시스템을 활용하려는 시도가 있었으나[1], 카메라 설치와 유지 보수에 많은 비용이 소요되고, 실시간 데이터 전송 및 처리를 위한 서버 운영의 복잡성으로 인해 비효율적이라는 한계가 있다. 특히, 카메라 렌즈에 바이오파울링이 직접 발생하면서 장기간 관측 시 신뢰성을 유지하기 어려운 문제도 있다. 그에 비해,

부표와 같은 해양 관측 장비의 센서 데이터를 활용하여 바이오파울링 발생 시점을 예측하는 접근은 아직까지 제한적이다.

본 연구에서는 부표에 설치된 해양 센서가 측정한 데이터를 활용하여 시계열 패턴을 학습하고 바이오파울링 발생을 예측하는 AI 모델을 개발하였다. 데이터의 흐름을 추적하면서 바이오파울링을 실시간으로 예측할 수 있다면, 보다 합리적이고 체계적인 센서 바이오파울링 대응이 가능해질 것이다. 또한 수집된 해양 데이터의 신뢰성을 보장하고, 데이터 관리의 자동화를 지원하여 효율적인 시스템 관리 및 운영에 기여할 수 있을 것으로 기대한다.

2. Related Work

2.1 기존 해양 환경 이상 탐지 연구

과거 연구에서는, 해양 측정 기기의 데이터 품질을 주로 전문가가 주로 수작업으로 처리하고, 이상 탐지는 전문가의 경험에 의존하여 수행되었다.

최근에는 Argro Profile Float 라는 해양 관측 장치를 통해 해양의 여러 정보를 수집하고 있다. Argro Profile Float 는 일정 기간 동안 수심 1000m 에서 2000m 사이의 해양 깊이를 수직으로 이동하면서, 일반적으로 10 일 동안 해양의 온도, 염도, 깊이에 대한 정보를 총체적으로 수집한다. 하지만 이 관측 장치도 갑작스러운 수괴층의 변화나 해양 오염에 노출될 경우 데이터가 왜곡될 가능성이 있다. 이 문제를 해결하기 위해 Trajectory Clustering 방법을 활용한 이상 탐지 연구가 진행되었으며, 기존의 Local Outlier Factor 및 Density Based Spatial Clustering of Applications with Noise 방식보다 우수한 성능을 보여, 해양 데이터의 신뢰성을 높이는 데 중요한 기여를 하고 있다고 보고되었다[2].

2.2 AI 기반 센서 이상 탐지

과거 연구에서 센서 데이터의 이상 탐지는 전통적으로 통계적 방법이나 임곗값 기반 접근법을 통해 이루어졌다. 평균, 분산, 상관계수 등의 통계 지표를 활용하여 데이터의 이상 여부를 판단하고, 사전에 정의된 임곗값을 초과하는 경우 이상으로 간주하는 방식이다. 이러한 방법들은 직관적이며 모델의 설명력이 우수하나, 비선형적 패턴에서의 이상 탐지에는 한계가 있다[3].

최근에는 인공지능의 발전으로 비지도 학습과 딥러닝을 활용한 이상 탐지 기법이 등장하였다. LSTM(Long Short-Term Memory)과 같은 시계열 모델을 활용하여 시간에 따른 데이터 변화를 효과적으로 모델링함으로써 이상을 탐지할 수 있었다.

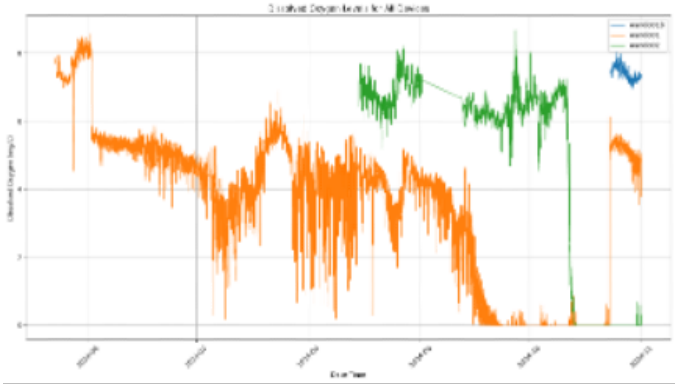
이러한 연구들을 통해 센서 데이터 이상 탐지 방법이 전통적인 통계적 접근법과 전문가의 경험에 의존하는 방식에서 AI 기반의 지능형 시스템으로 발전했음을 확인할 수 있다. 본 연구에서는 이러한 최신 기법들을 적용하여 해양 환경에서 바이오파울링 발생 여부를 보다 정확하게 예측하고자 한다.

3. Methods

3.1 Data

본 연구에서는 스마트 부표 디바이스를 실제 운영 중인 기업의 스마트 부표를 통해 수집한 해양 데이터를 바탕으로 바이오파울링 예측 모델을 수립하였다. 본 연구는 부표에 탑재된 센서 중 바이오파울링에 민감한 광학센서인 용존산소 센서 데이터를 주로 활용하였다. 실제 데이터는 동일한 지역에 설치된 세 개의 부표(Wando01, Wando01b, Wando02)에서 수집된 자료를 사용하였으며, Wando01 은 5 월 23 일부터 11 월 1 일까지, Wando01b 는 10 월 24 일부터 11 월 1 일까지, Wando02 는 8 월 15 일부터 11 월 1 일까지의 데이터를 수집하여 분석하였다.
원시 데이터에서 Fig. 1 과 같이 시계열 예측 및 기기 구분을 위해 기기 ID, 측정 시간, 용존산소 수치만을 추출하여 전처리하였다.

Fig. 1. 기기별 용존산소 시계열 그래프



해양 생물 다양성과 저산소 상태에 관한 선행 연구[4]에 따르면, <Table 1>과 같이 용존산소 농도가 3.0 mg/L

이하일 때 저산소 상태로 정의된다. 본 연구에서는 무산소 상태인 0 mg/L 를 제외한 평균 용존산소 농도가 5.272 mg/L 로 나타났다. 해양 환경에서 용존산소 농도가 단기간에 5.0 mg/L 에서 3.0 mg/L 로 급격히 감소하는 현상은 비정상적인 상태를 시사하며, 바이오파울링의 시작 기준으로 3.0 mg/L 를 설정하였다. 비정상 데이터는 3.0 mg/L 로 감소하는 시점의 719 개 이전 데이터를 포함하며, 그 외의 데이터를 정상 데이터로 정의하였다.

<Table 1> 용존 산소에 따른 해양 상태

용존 산소(mg/L)	상태
DO ≥ 5.0 mg/L	정상
3.0 mg/L ≤ DO < 5.0 mg/L	경계저산소
2.0 mg/L ≤ DO < 3.0 mg/L	저산소
DO < 2.0 mg/L	심각한 저산소
DO ≈ 0 mg/L	무산소

원시 데이터는 1 분 이내의 불규칙한 시간 간격으로 수집되었기 때문에, 시계열 데이터 예측에 적합하도록 <Table 1>과 같은 칼럼을 추출한 뒤 2 분 간격으로 샘플링하고, 결측치는 보간법으로 처리하여 최종 데이터셋을 구성하였다. 이 과정에서 비정상 데이터는 72,682 개, 정상 데이터는 116,570 개로 총 189,152 개의 데이터를 확보하였다.

<Table 2> 원시 데이터 개수

상태	개수	샘플링 후 개수
전체 데이터	229,287	-
정상 데이터	161,395	116,570
비정상 데이터	67,892	72,582

본 연구는 해양데이터가 일반적으로 하루를 주기로 특정 패턴이 반복되는 것을 고려하여 720 개의 연속적인 데이터를 하나의 시퀀스로 정의하였으며, 슬라이딩 윈도우 기법을 사용하여 정상 데이터 시퀀스 159,238 개와 비정상 데이터 시퀀스 66,454 개를 구축하였다. 전체 데이터는 <Table 3>과 같이 학습(train), 검증 (validation), 테스트(test) 용도로 70:15:15 의 비율로 무작위 분할하여 구성하였다.

<Table 3> 데이터셋 분할 개수

데이터 구분	개수
Train	157,984
Validation	33,853
Test	33,855

3.2 Modeling

본 연구에서는 해양 환경에서의 바이오파울링 발생 여부를 예측하기 위해 순환신경망 계열의 모델인 RNN, GRU, LSTM 을 활용하였다. 각 모델은 시계열 데이터 처리에 있어 고유한 특성을 지니고 있다. 기본적인 RNN 은 순차적 데이터를 처리하는 데 효과적이지만, 긴 시퀀스에서 기울기 소실 문제가 발생할 수 있다는 한계를 가진다. GRU 는 Update Gate 와 Reset Gate 를 통해 이러한 문제를 개선하였으며, 비교적 단순한 구조로 효율적인 학습이 가능하다. LSTM 은 Input Gate, Forget Gate, Output Gate 와 Memory Cell 를 활용하여 장기 의존성 문제를 효과적으로 해결할 수 있으며, 복잡한 시계열 데이터에서 정보의 저장과 전달을 선택적으로 제어함으로써 안정적인 학습이 가능하다.

바이오파울링 예측 문제는 시간에 따른 용존산소량의 변화 패턴을 분석해야 하는 특성을 가진다. 특히, 센서 데이터의 미세한 변동성과 장기적인 추세를 모두 고려해야 하므로, 단순한 통계적 방법이나 전통적인 시계열 분석 기법으로는 한계가 있다. 이에 순환신경망 계열의 모델들을 적용함으로써, 시간적 의존성을 고려한 효과적인 패턴 학습이 가능할 것으로 기대하였다. 또한, 과적합을 방지하고 안정적인 학습을 보장하기 위해 학습률 조정(learning rate scheduling), 조기 종료(early stopping) 기법을 적용하였다. 특히, 검증 손실이 개선되지 않을 경우 학습률을 자동으로 조정하는 `ReduceLROnPlateau` 스케줄러를 사용하여 학습의 안정성을 확보하였다.

모델의 학습 과정에서는 Cross-Entropy Loss 를 손실 함수로 사용하여 이진 분류 문제에 적합한 학습이 이루어지도록 하였다. Adam 옵티마이저를 통해 효율적인 파라미터 업데이트를 수행하였으며, 초기 배치 크기는 32 로 설정하였다. 이는 선행 연구[5]에서 배치 크기가 32 이하일 때 일반화 성능이 가장 우수하다는 실험 결과를 반영한 것이다.

3.2.1 Metrics

본 연구에서는 바이오파울링 예측 모델의 성능을 평가하기 위해 이상 탐지에서 일반적으로 사용되는 4 가지 평가 지표를 활용하였다. 정확도(Accuracy)는 전체 예측 중 정확히 분류된 비율을 나타내는 지표로, 전반적인 모델의 성능을 평가하는 데 사용된다. 그러나 클래스 불균형이 있는 경우 이 지표만으로는 모델의 성능을 정확히 평가하기 어렵다. 정밀도(Precision)는

모델이 비정상(바이오파울링 발생)이라고 예측한 케이스 중 실제로 비정상인 비율을 나타낸다. 이는 오탐지(false positive)를 최소화하는 것이 중요한 상황에서 유용한 지표이다. 재현율(Recall)은 실제 비정상 케이스 중 모델이 정확히 비정상이라고 예측한 비율을 의미한다. F1 Score 는 Precision 와 Recall 의 조화평균으로, 두 지표 간의 균형을 평가하는 데 사용된다. 이는 정밀도와 재현율 중 어느 한쪽으로 치우치지 않고 두 성능이 모두 우수한 모델을 선별하는 데 효과적이다.

3.2.2 Experimental Results

본 연구에서는 RNN, GRU, LSTM 모델의 성능을 비교하기 위해 각 모델별로 하이퍼파라미터 최적화를 수행하였다. 하이퍼파라미터 최적화는 Optuna 라이브러리를 활용하여 진행되었으며[6], 최적화 대상은 hidden size, learning rate, batch size 로 설정하였다. 각 모델에서 최적화된 하이퍼파라미터는 <Table 4>와 같다.

<Table 4> 모델별 최적 하이퍼파라미터

Model	hidden_size	learning_rate	batch_size
RNN	32	0.001	16
GRU	32	0.006	64
LSTM	64	0.001	32

최적화된 하이퍼파라미터를 적용하여 각 모델의 성능을 평가한 결과는 <Table 5>에 나타나 있다. 각 실험은 무작위 초기화와 데이터 셔플에 따른 변동성을 고려하여 10 회 반복 수행하였으며, 결과는 평균값으로 제시하였다.

<Table 5> 모델별 성능 비교

Model	Accuracy	Precision	Recall	F1 Score
RNN	0.95	0.94	0.93	0.94
GRU	0.99	0.99	0.98	0.99
LSTM	0.98	0.98	0.96	0.97

실험 결과, GRU 모델이 F1 스코어 0.99 로 가장 우수한 성능을 보였다. GRU 모델은 작은 hidden size(32)로도 높은 예측 정확도를 달성하여 계산 효율성과 성능을 동시에 확보하였다. LSTM 모델은 F1 스코어 0.97 로 그 뒤를 이었으며, RNN 모델은 상대적으로 낮은 성능을 나타냈다.

GRU 모델의 우수한 성능은 바이오파울링으로 인한 용존 산소량의 시계열 패턴을 효과적으로 학습할 수 있는 게이트 구조 덕분으로 해석된다. 이는 GRU 모델이 바이오파울링 예측에 가장 적합한 모델임을 시사하며, 해양 센서 데이터의 이상 징후를 정확하게 포착하여 실시간 모니터링 및 대응에 활용될 수 있음을 의미한다.

4. Conclusion

본 연구에서는 완도 지역의 3 개 부표에서 수집한 실제 데이터를 활용하여 해양 센서의 바이오파울링을 예측하는 AI 모델을 개발하였다. RNN, GRU, LSTM 모델의 성능을 비교 분석한 결과, GRU 모델이 F1 스코어 0.99 를 달성하며 가장 우수한 성능을 보였다. 작은 hidden size(32)로도 높은 성능을 달성하여 계산 효율성과 예측 정확도를 동시에 확보하였다.

본 연구의 시사점은 해양 센서의 바이오파울링을 효과적으로 예측함으로써 센서의 효율적인 운영과 관리에 기여할 수 있다는 것이다. 높은 예측 정확도와 계산 효율성을 갖춘 GRU 모델을 활용하여 바이오파울링 발생을 실시간으로 모니터링할 수 있으며, 이를 기반으로 유지보수 일정의 최적화와 자동 알림 시스템 등의 통합 관리 플랫폼을 개발하고 있다. 이 플랫폼은 해양 데이터의 신뢰성을 보장하고 데이터 관리의 자동화를 지원하여 해양 자원의 효율적 관리와 생태계 보호에 이바지할 것으로 기대된다.

그러나 본 연구는 바이오파울링의 실제 발생 시점에 대한 정확한 레이블링 데이터가 부족하다는 한계를 지니고 있다. 향후 연구에서는 정기적인 현장 검사 결과 등을 활용하여 보다 정확한 바이오파울링 발생 데이터를 확보할 계획이다. 이를 통해 학습 데이터의 품질을 높이고, 모델의 예측 정확성을 개선할 수 있을 것이다. 또한, 다양한 지역의 데이터를 포함하고 수온, 염도, pH 등 복합적인 해양 환경 변수를 통합적으로 고려함으로써 모델의 일반화 가능성과 예측 능력을 더욱 강화할 예정이다.

Acknowledgement

본 연구는 과학기술정보통신부 및
정보통신기획평가원의 SW 중심대학사업의 연구결과로 수행되었음. (2021-0-01082)

References

[1] Signor, J., Schoefs, F., Quillien, N., & Damblans, G. (2023). Automatic classification of biofouling images from offshore renewable energy structures using deep learning. *Ocean Engineering*, 288, 115928.

[2] Cai, W.-Y., Liu, Z.-Q., & Zhang, M.-Y. (2020). *Trajectory clustering based oceanic anomaly detection using Argo profile floats*. In *ChinaCom 2019: International Conference on Communications in China* (pp. 498–508). Springer, Cham.

[3] Hodge, V. J., & Austin, J. (2004). *A survey of outlier detection methodologies.* Artificial Intelligence Review, 22(2), 85–126.

[4] Vaquer-Sunyer, R., & Duarte, C. M. (2008). "Thresholds of hypoxia for marine biodiversity." *Proceedings of the National Academy of Sciences*, 105(40), 15452–15457.

[5] Masters, D., & Luschi, C. (2018). "Revisiting Small Batch Training for Deep Neural Networks." arXiv preprint arXiv:1804.07612.

[6] Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). "Optuna: A Next-generation Hyperparameter Optimization Framework." Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2623–2631.