

유향/유속 데이터 중심의 이상 검출 신경망 모델

현성길¹, 유동영²

¹홍익대학교 스마트시티형도시재생융합전공 ¹석사과정

²홍익대학교 소프트웨어융합학과 교수

aha1492@outlook.kr, ydy@hongik.ac.kr

A Study on Anomaly Detection Neural Network Model Based On Flow Direction/Velocity Data

Seong-Kil Hyun¹, Dong-Young Yoo²

¹Smart City & Urban Regeneration Converge, Hongik University

²Department of Software Convergence, Hongik University

요 약

해양의 영향을 많이 받는 우리나라의 지리적 특성상 해양 상황은 산업 및 생업과 밀접한 관계가 있다. OPEN API 를 이용하여 유향/유속, 조위 등 해양 환경 관련 실시간 시계열 데이터를 수집한 후 2 차원 공간에 표시하여 순환 신경망 모델을 이용하여 학습한다. 학습된 모델을 실제 데이터에 적용하여 파랑을 예측한다. 시계열의 성격이 있고 공간상에 표시할 수 있는 데이터라면 본 논문에서 제시한 체계를 통해 예측할 수 있을 것이라 예상된다.

1. 서론

삼면이 바다로 둘러 쌓인 우리나라는 지리적인 특성상 바다의 영향을 많이 받을 수밖에 없다. 해양에서의 기상 변화는 수산자원, 어촌어항, 양식산업 및 인명사고 등에 큰 영향을 미친다. 국립해양조사원의 해양조사 용어사전에서는 파랑은 바람이 해면 위를 지속적으로 불 때 생기는 풍랑과 다른 해역에서 발생한 풍랑이 전파되거나 바람이 멈출 때 생기는 너울을 총칭한다고 되어있다. 해상교통, 어업 등의 산업에 영향을 미치며 심한 경우 인명과 재산에 심각한 피해를 줄 수 있다. 파랑과 같은 심한 파도는 계속해서 선박을 위협에 빠뜨리고 해양 구조물을 손상시키며 선원에게 심각한 부상을 입힌다.[1] 장기적으로 정확하고 신뢰할 수 있는 해수 높이 예측을 위해 게이트 순환장치(GRU) 네트워크를 통해 단기 및 장기 해파 예측 방안을 제시한 논문[2]도 있으며 콘볼루션 인공 신경망을 이용하여 파력을 예측하기 위한 논문도 있다.[3] 연안 환경 변화를 정확하게 예측하기 위해 조위 및 조류 예측의 정확도를 개선하기 위한 방법이 연구되기도 하고[4] 정확한 데이터 수집 및 보정을 위해 결측되는 조위자료의 양상을 연구하기도 하였다.[5] 또한, 한반도에서의 파랑에 의한 해난사고 분석을 위해 파랑관측자료 분석 및 파랑후측도의 실험을 한 연구도 있다.[6] 본 연구는 파랑의 시간적 공간적 특성을

이용하여 2 차원 공간 평면상에 파랑에 영향을 미치는 조위, 유향/유속 데이터를 가시화한 후 인공지능으로 학습하여 파랑 예측을 할 수 있는 방법을 제시한다.

2. 관련 연구

2-1. 해양 환경 데이터

해양수산부는 대한민국의 중앙행정기관 중 하나로, 해양과 수산 분야의 정책을 수립하고 집행하는 부처이다. 해양수산부 및 관련 산하기관에서는 해양수산 분야의 다양한 데이터를 수집, 가공하여 다양한 데이터 포맷으로 제공하고 있다. 유향/유속, 조위, 수온 등 관측소를 통해 실시간으로 수집된 데이터, 선박현황, 어가 및 어가인구 증감 현황 등 정기적으로 생산되는 통계성 데이터, 수온, 파고 등 해양 예측 데이터 등이 서비스되고 있다. 직접 데이터 파일을 다운로드 받을 수 있고, 2 차원 지도를 통해 데이터를 확인할 수도 있다. 또한, OPEN API(Application Programming Interface)를 이용해 자동으로 데이터를 수집할 수도 있다.

OPEN API 는 개발자가 소프트웨어나 웹 서비스의 기능을 이용할 수 있도록 공개한 메소드(함수)를 말한다. 보통 OPEN API 를 이용하려면 서비스를 제공하는 곳으로부터 서비스키와 같은 사용권한을 얻어야 하며, 기능에 맞는 URL(또는 함수)을 호출해 원하는 기능을 동작시킨다. SOAP 이나 REST 방식으로 제공될 수

있으며, 응답은 XML 이나 JSON 형태로 되어 있다. 최근에는 가볍고 쉽게 사용할 수 있는 RESTful 방식의 JSON 이 많이 사용되고 있다. 해양수산부에서도 다양하고 많은 데이터에 대해 OPEN API 를 제공하고 있으며, 사용자는 수집 프로그램을 제작하여 자동/실시간으로 데이터를 수집/적재할 수 있다.

2-2. 공간정보(지리정보)

공간정보는 공간상에 존재하는 모든 객체에 대한 정보를 말하며 지리정보는 공간정보의 하위 개념으로 볼 수 있다. OGC(Open Geospatial Consortium, 개방형 공간정보 컨소시엄)는 국제 표준을 이끌고 있으며 다양한 공간정보관련 기술을 표준화하고 QGIS, GeoServer 등과 같은 유용한 소프트웨어를 무료로 공개하고 있다. 2 차원 평면에 점, 선, 면으로 표시되는 평면 정보와 3 차원으로 표시되는 입체 정보로 구분할 수 있다. 3 차원 정보를 표시하기 위해서는 많은 컴퓨팅 자원이 필요하며, 특히 좌표연산을 위해 그래픽카드(GPU)는 필수라 할 수 있다. 과거에는 pc 에 설치되는 standalone 형태의 어플리케이션에서 공간정보를 다루었으나 최근 컴퓨터의 사양이 향상되고 웹브라우저의 기능이 좋아지면 웹에서도 3 차원 정보를 표시하는 사례가 늘어나고 있다. 또한, 성능 향상을 위한 다양한 기법들이 발표되고 있다. ESRI 사의 shape file 과 GeoServer 의 Web Map 성능을 비교한 논문도 있다.[7] 3 차원 데이터를 다룰 수 없는 GeoServer 에 3 차원 공간 데이터를 처리할 수 있도록 하기 위한 방안이 제시되기도 하였다.[8] A Research of publishing map technique based on geoserver 에서는 응답 시간을 최소화 하고 사용자 경험을 향상시키기 위해 타일-피라미드 모델을 분석하여 개선된 쿼드-트리 공간 인덱스 구조를 제시한다.[9] 타원모양의 실제 세계를 평면에 표시하기 위해서는 투영(Projection)이라는 작업이 필요하며 평면도법, 원통도법, 원추도법등 다양한 종류의 투영법이 존재한다.

2-3. 이상 탐지(Anomaly Detection)

데이터 속의 이상한 값이나 패턴을 찾는 기법으로 시계열 분석이나 인공 신경망과 같은 분야에서 활용된다. 일반적으로 데이터셋에서 작은 비율의, 다른 데이터의 범주와 확연히 구분되는 데이터를 찾고 싶을 때 그것을 outliers 또는 anomalies 라고하며 이상 탐지 문제로 정의한다. 오래전부터 연구되어 온 주제이며, 시간이 지나며 통계적, 경험적, 머신러닝등 다양한 탐지 기법들이 개발되었다.

통계학 측면에서 이상은 관측치들이 주로 모여 있는 곳에서 멀리 떨어져 있는 관측치로 정의된다 이상

값은 확률 분포에서 낮은 영역에 나타난다고 가정하며 모수적 기법과 비모수적 기법이 있다.

분류는 머신러닝 방법 중 지도 학습의 대표적인 방법으로 SVM(Support Vector Machine), NN(Nearest Neighbor)등이 있다. 학습 데이터의 정상 데이터와 이상 데이터를 라벨링할 수 있다면 지도학습으로 학습할 수 있다. 학습 데이터에 정답 레이블이 있어 분류기를 학습한 뒤, 학습된 모델로 새로운 데이터가 각각의 클래스에 속할 확률을 예측하는 방법이다.

비지도 학습은 정답이 없는 데이터에서 비슷한 특징끼리 군집화하여 새로운 데이터에 대한 결과를 예측하는 방법으로 군집화(Clustering), 차원축소(Dimensionality Reduction)등이 있다.

3. 본론

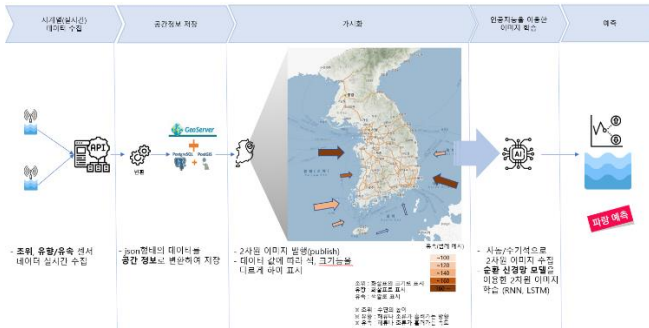
3-1 모델

Ki-Su Kim 은 디노이즈 자동엔코더와 콘볼루션 LSTM 을 이용하여 해양 날씨를 예측하는 방안을 제시하였고[10] Christoph Jorges 는 LSTM 으로 측정된 수심 데이터를 학습하여 해수 높이를 예측하려고 하였다.[11] 유항/유속, 조위등의 값은 시간에 따라 변동되는 시계열성 데이터로 머신러닝 학습을 위해 순환 신경망 모델이 적절하다. 하지만, Wang, Yi(2019)처럼 WNN(wavelet neural network)을 이용하여 해양 관측 데이터의 이상 탐지 방안을 제시한 경우도 있다.[12] 순환 신경망 모델은 시간의 흐름에 따라 변화하는 데이터를 학습하기 위한 인공 신경망의 한 종류로, 순방향 신경망과 달리 과거의 출력 데이터를 재귀적으로 참조하는 특징이 있다. 순환 신경망 모델에는 여러가지 변형이 있는데, 그 중에서 LSTM(Long Short-Term Memory, 장단기메모리) 모델은 오랜 시간 동안 정보를 기억하고 전달할 수 있는 구조로, 복잡한 문제를 해결하는데 효과적이다. 시계열 데이터 학습에 효과적이며 이전에 관측된 값을 기반으로 미래의 값을 예측할 수 있는 모형이다. 시계열의 예로는 시간 경과에 따른 온도, 습도, 주가등이 있으며 입력은 시간에 따라 연속적으로 나타나는 신호이다.

3-2 체계

해양수산부에서 OPEN API 형태로 제공하는 유항/유속, 조위 데이터와 비전인식, 이상 검출을 위한 순환 신경망 모델을 활용하여 파랑 예측 모델을 만든다. 48 개소 조위관측소에서 측정된 조위 데이터와, 38 개소 해양관측부이로부터 측정된 유항/유속 데이터를 OPEN API 를 이용하여 (근)실시간으로 수집한다. 수집 프로그램은 java 로 구현하여 리눅스 서버에 설치(복사)한다. 리눅스의 cron 을 이용하여 프로그램을 주기적으

로 실행시키며, 실행된 프로그램은 OPEN API 를 이용하여 데이터를 수집한다. 수집된 데이터는 2 차원 공간상에 표시하기 위해 실시간으로 공간정보와 속성정보로 변환되어 데이터베이스에 저장된다. 공간정보 데이터를 공유하고 조회할 수 있도록 하기 위해 OGC(Open Geospatial Consortium, 개방형 공간정보 컨소시엄)에서 무료로 제공하는 오픈 소스 GIS 소프트웨어 GeoServer 를 사용하였다. GeoServer 의 WMS(Web Map Service)를 이용하여 공간정보를 웹 브라우저상에 2 차원 이미지로 표시한다. 파도의 높이, 해수의 방향/속도 데이터에 따라 각기 다른 아이콘이 2 차원 지도상에 표시된다. 학습용 프로그램은 자동/주기적으로 GeoServer 에 2 차원 이미지를 요청하며 수집된 이미지는 순환 신경망 모델로 학습된다. 학습된 모델을 이용하여 파랑 발생을 예측한다.



(그림 1) 데이터 수집 pipeline 및 학습 흐름도

4. 결론

본 연구는 시변적 동적 특성을 가진 유향/유속 데이터를 2 차원 공간상에 표시한 후 인공지능으로 학습하여 파랑을 예측하기 위한 인공지능 모델과 데이터 구축 pipeline 및 구현체계의 구성방법을 제시하였다. 속성 데이터를 공간 데이터로 변환하여 저장하기 위한 데이터 구축 pipeline 과 구현체계 제시, 이미지 학습을 통한 이상 탐지에 의의가 있다고 하겠다. 파랑은 어민들의 생명, 생업 및 산업에 큰 영향을 미친다. 해양수산부가 운영하는 해양수산 빅데이터 플랫폼에서 제공하는 파랑 예측에 필요한 시계열, 공간 데이터인 유향/유속, 조위데이터를 OPEN API 를 이용하여 실시간으로 수집하여 관계형 데이터베이스에 저장하였다. 수집된 데이터를 공간데이터로 변환하기 위해 GDAL 을 이용하였고, 아이콘화된 유향/유속 데이터를 GeoServer 를 이용하여 웹상에 2 차원 공간 이미지를 발행(publishing)하였다. 발행된 2 차원 이미지를 시변적 동적 특징을 모델링 할 수 있는 순환 신경망 모델을 이용하여 학습한 후 파랑을 예측하였다. 입력 데이터가 변경된다고 해도 시계열의 성격이 있고 2 차원 공간상에 표시할 수 있다면 본 연구의 제시방법을 이용하여 이상 탐지가 가능할 것으로 예상된다.

참고문헌

- [1] Thomas Breunung, Balakumar Balachandran, Data-Driven, high resolution ocean wave forecasting and extreme wave predictions, Ocean Engineering, 268, 2023
- [2] Multi-step forecasting of ocean wave height using gate recurrent unit networks with multivariate time series, Ocean Engineering, 248, 2022
- [3] Ocean wave power forecasting using convolutional neural networks, IET Renewable Power Generation, 15, 15, 2021, 3341-3353
- [4] 정태성, 조위 및 조류 예측 정확도의 개선 방법, 한국해양환경공학회지, 13, 4, 234-240
- [5] 조홍연외 2, 한반도 연안 조위자료의 결측 양상, 한국해양·해양공학회논문집, 23, 6, 496-501, 2011
- [6] 천후섭, 안경모, 조석 및 조류 효과를 고려한 황해역 광역 파랑 수치모의 실험, 한국해양·해양공학회 논문집, 30, 6, 286-297, 2018
- [7] JanRuzicka, Comparing speed of Web Map Service with GeoServer on ESRI Shapefile and POSTGIS, 15,1,3-9,2016
- [8] 김수진, 유형규, 이기준, 3 차원 공간데이터를 다루기 위한 GeoServer 확장, 대한공간정보학회, 10, 125-128, 2017
- [9] Sun,L., He,D.A, Zhao,P., Research of publishing map technique based on geoserver,Asian Journal of Applied Sciences8,3,185-195,2015
- [10] Ki-Su Kim 외 4, Prediction of Ocean Weather Based on Denoising AutoEncoder and Convolutional LSTM, Mar. Sci. Eng., 8, 10, 2020
- [11] Prediction and reconstruction of ocean wave heights based on bathymetric data using LSTM neural networks, Ocean Engineering, 232, 109046, 2021
- [12] Wang,Yi 외 4, Study on wavelet neural network based anomaly detection in ocean observing data series, Ocean Engineering, 186,106129,2019