

진동수주 파력발전장치를 위한 머신러닝 기반 압력 예측모델 설계 및 분석

서동우¹, 허태상^{1*}, 김명일¹, 오재원², 조수길²
¹한국과학기술정보연구원 가상설계센터, ²선박해양플랜트연구소

A Design and Analysis of Pressure Predictive Model for Oscillating Water Column Wave Energy Converters Based on Machine Learning

Dong-Woo Seo¹, Taesang Huh^{1*}, Myungil Kim¹, Jae-Won Oh², Su-Gil Cho²

¹Division of Modeling and Simulation, KISTI

²Technology Center for Offshore Plant Industries, KRISO

요약 최근 다양한 산업/제조 현장에서 운영 효율화를 위한 디지털 트윈(digital twin) 기술 연구가 활발하게 수행 중이고, 화석 연료의 점진적 고갈과 환경오염 문제는 파력발전소와 같은 신재생/친환경 발전방식을 요구한다. 하지만, 파도의 에너지에 의해서 전기를 생산하는 파력발전에서 변동성이 높은 파도에너지에 의해서 발전량과 고장 등의 운영 효율화 요소가 밀접하게 관련되어 있어 이들 사이의 관계를 이해하고 예측하는 것이 매우 중요하다. 따라서 첫 번째로 파고 데이터, 진동수주(OWC: Oscillating Water Column, 이하 OWC) 챔버의 센서 데이터 등과 같은 변동성이 높은 데이터 간에 의미 있는 상관관계 도출이 필요하다. 두 번째로 도출된 상관관계를 기반으로 추출된 데이터로 예측 상황을 학습함으로써 원하는 정보를 예측할 수 있는 방법론 연구가 이루어져야 한다. 본 연구에서는 파력발전 시스템의 디지털 트윈으로 스마트 운용 및 유지보수가 가능하도록 실제 파력발전소의 IoT 센서 데이터를 이용하여 OWC의 압력 예측을 위해 머신러닝 프레임워크를 활용한 워크플로우 기반의 학습모델을 설계하고, 검증 및 평가 데이터셋을 통한 압력 예측 분석의 유효성을 확인한다.

Abstract The Korea Nowadays, which is research on digital twin technology for efficient operation in various industrial/manufacturing sites, is being actively conducted, and gradual depletion of fossil fuels and environmental pollution issues require new renewable/eco-friendly power generation methods, such as wave power plants. In wave power generation, however, which generates electricity from the energy of waves, it is very important to understand and predict the amount of power generation and operational efficiency factors, such as breakdown, because these are closely related by wave energy with high variability. Therefore, it is necessary to derive a meaningful correlation between highly volatile data, such as wave height data and sensor data in an oscillating water column (OWC) chamber. Secondly, the methodological study, which can predict the desired information, should be conducted by learning the prediction situation with the extracted data based on the derived correlation. This study designed a workflow-based training model using a machine learning framework to predict the pressure of the OWC. In addition, the validity of the pressure prediction analysis was verified through a verification and evaluation dataset using an IoT sensor data to enable smart operation and maintenance with the digital twin of the wave generation system.

Keywords : Machine Learning, OWC, Pressure Prediction Model, Wave Energy Converter, Big Data, HPC Cloud

본 논문은 한국과학기술정보연구원(Grant No. K-20-L02-C06-S01, K-20-L02-C12-S01)과 선박해양플랜트연구소(Grant No. PES3590) 연구과제로 수행되었습니다.

*Corresponding Author : Taesang Huh(KISTI)

email: tshuh@kisti.re.kr

Received September 7, 2020

Revised October 5, 2020

Accepted November 6, 2020

Published November 30, 2020

1. 서론

최근 석유나 가스 등 화석 연료 의존에 따른 에너지 수급의 불균형과 고갈의 문제와 함께 심각해지는 환경오염에 따른 지구온난화 문제가 대두됨에 따라 영구적인 에너지원인 태양광, 풍력뿐 아니라 해양에너지와 같은 청정에너지에 관심이 쏠리고 있다. 대체 청정에너지로 주목 받는 해양에너지로는 해양의 조력, 파력, 해류, 온도 차 등을 이용하여 전기를 생산하는 발전 기술이 있다. 특히 제주도 서해는 국내에서 파랑에너지가 높은 지역으로 파력발전에 용이성이 있다[1]. 겨울철에는 대륙성 기후의 영향으로 파랑에너지가 높으며 12월에 평균적으로 파랑에너지 밀도가 가장 높게 관측되었고, 태풍 시즌인 8월과 9월에는 변동성과 최대 파랑에너지의 비가 가장 높은 특징을 보인다[2]. 파력발전을 위한 구조물은 운동에너지의 변환방식에 따라 월파형(overtopping devices), 가동물체형(wave-activated bodies)과 진동수주형(OWC: oscillating water column, 이하 OWC)으로 분류되기도 하며 OWC가 가장 유용성이 큰 것으로 알려져 있다 [3-5].

국내에서는 제주시 한경면 용수리 앞 1.2km 해상에 건설된 용수 파력발전소가 고정식 OWC 파력발전장치로 2014년도에 시공되어 현재 시범 운영을 준비하고 있다. 일반적으로 OWC의 파랑에너지에서 추출되는 전기동력은 60~500kW의 범위에 있으며, 용수 파력발전소는 500kW를 목표로 시험 중에 있다[6]. OWC는 공기실 내부의 수위 진동에 의해 발생한 공기 흐름으로 변환하여 공기터빈을 구동시켜 전기에너지를 회수하는 방식으로 유입되는 파랑의 주파수와 공기실의 고유주파수와 일치할 때 공진현상으로 에너지효율을 극대화할 수 있다. 챔버의 압력 예측은 이러한 에너지 효율성을 높이는데 가장 직접적인 변수일 뿐만 아니라 현재 용수 파력발전소의 발전기에서 크게 이슈가 되는 큰 압력에서의 고장문제를 사전에 제어하기 위한 필요 기술이다.

본 연구에서는 챔버의 압력을 예측하여 발전기 제어로 전력생산을 극대화하고 사전 고장방지를 목적으로 한다. 수집되는 센서 데이터를 기반으로 데이터 상관분석을 포함한 데이터 전처리와 머신러닝을 적용한 예측모델을 설계하고 학습모델의 평가로 챔버의 압력 예측분석의 유효성을 검증한다. 본 논문의 구성을 살펴보면, 2장에서는 관련 연구로서 OWC 파력발전장치, 데이터 분석 플랫폼 및 머신러닝에 대해 소개하고 3장에서 OWC 압력 예측 학습모델 설계 부분으로 데이터 수집, 입력 데이터의 상

관분석, 데이터셋 설계 및 구축을 논한다. 4장에서는 챔버의 압력 예측 학습모델 구축 및 평가 결과를 설명하고 5장에서는 결론으로 기대효과 및 향후 연구계획으로 마무리한다.

2. 관련 연구

2.1 OWC 파력발전장치

OWC 파력발전장치는 파도에너지를 기계적 에너지로 변환하고, 최종적으로 기계적 에너지를 전기 에너지로 변환한다.

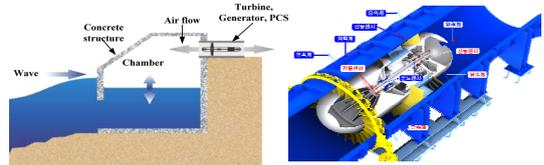


Fig. 1. OWC wave energy converter and air turbine

Fig. 1의 OWC 파력발전장치는 파랑에 의한 챔버 내의 공기가 상하로 진동하고 이에 따라 공기가 외부와 내부를 왕복운동 한다. 왕복운동을 하는 공기가 블레이드를 통과함으로써 터빈을 회전시키게 되고 이는 기계적 에너지로 변환한다. 터빈으로 전달된 기계 에너지는 발전기를 통해 전기 에너지로 변환된다. 하지만 신재생에너지 특성상 일정하지 않은 전압이 발생하기 때문에 이렇게 전달된 에너지는 전력변환장치(PCS: Power Conversion System, 이하 PCS)를 거쳐 일정한 전압과 주파수를 갖는 전기적 에너지로 변환된다. 즉, 기계적 에너지를 통한 발전기의 출력은 변화하는 주파수와 전압의 크기를 갖는 전기적 에너지이기 때문에 계통에 연계하여 파력발전장치를 운행하기 위해서는 PCS가 필수적으로 필요하다[7, 8].

2.2 데이터 분석 플랫폼 KNIME

KNIME 분석 플랫폼(analytics platform)은 오픈소스 기반의 예측분석(predictive analytics) 소프트웨어 플랫폼으로 유연성, 개방성 및 다른 소프트웨어 도구들과 통합이 용이하며, 데이터 수집, 데이터 변환, 데이터 통합, 데이터 분석에 이르기까지 예측분석을 지원하는 개방형 소프트웨어 플랫폼이다[9, 10].

Table 1. KNIME feature

특징	세부내용
개방성	Hadoop, Spark, R, Python, JAVA, SPSS, SAS 등의 다양한 소프트웨어와 통합이 용이
빅데이터 처리	범용 DBMS의 In-database 프로세싱 처리가 가능하며, In-Hadoop 프로세싱으로 데이터 핸들링 및 머신러닝의 수행이 가능
프로그램이 확장	R, Python 등의 언어와 그래픽을 지원하여 사용자 맞춤 대화상자 생성 가능
다양한 데이터 접근 및 핸들링	MySQL, Oracle, Teradata 등의 범용 DBMS뿐 아니라, Excel, XML, Twitter, Google, Facebook 등의 다양한 유형의 데이터 소스 접근 및 핸들링 기능 제공
다양한 예측분석	통계분석, 데이터마이닝 및 머신러닝 알고리즘을 지원하며, R과 WEKA의 머신러닝 알고리즘을 Visual 프로그래밍 방식으로 수행 가능

GUI를 통해 데이터 전처리, 모델링, 데이터 분석 및 시각화를 위한 노드의 구성이 가능하고, 다양한 프로그램과 연계하여 확장하여 사용이 가능 데이터 테이블이 커질 경우는 하드 드라이브에서 캐싱을 제공하며, 다양한 데이터 유형을 지원하고 Table 1과 같은 특징을 가진다 [11-13].

2.3 머신러닝

머신러닝은 인공지능의 한 분야로서, 데이터를 기반으로 알고리즘을 이용하여 데이터를 분석, 학습하고 새로운 입력 데이터의 출력을 판단 및 예측하도록 하는 알고리즘이다. 머신러닝은 학습형태에 따라 크게 지도 학습, 비지도 학습으로 분류할 수 있다. 지도 학습은 입력 데이터와 그에 대한 출력 데이터를 포함하는 레이블이 있는 데이터(labeled data)를 사용하여 학습을 진행하는 방법으로, 학습 데이터의 입력 데이터와 출력 데이터로 해당 시스템의 함수를 추론하여 새로운 입력이 들어왔을 때 해당 함수를 이용하여 대응하는 출력을 예측한다. 지도 학습은 연속적인 출력 데이터에 대한 예측인 회귀분석(regression)과 불연속적인 출력 데이터에 대한 예측인 분류(classification)로 나뉜다. 비지도 학습은 입력 데이터만을 가지는 레이블이 없는 데이터(unlabeled data)를 이용하여 학습법으로 입력 데이터의 특성을 이용하여 비슷한 특성을 가진 데이터들끼리 무리 짓는 군집화(clustering)가 대표적이다. 본 연구의 압력 예측은 연속적인 입력 데이터에 대해 연속적인 출력 데이터를 추론하는 회귀분석 알고리즘에 적합하다 할 수 있다. 대표적인 회귀분석 알고리즘으로는 LR(Linear Regression), SVR(Support Vector Regression), Ensemble methods,

RT(Regression Tree), NN(Neural Networks), DL(Deep Learning) 등이 있고 딥러닝로 많이 활용되는 알고리즘으로는 CNN(Convolutional Neural Networks), RNN(Recurrent Neural Networks), LSTM(Long Short-Term Memory) 등이 있다[14-16].

3. OWC 압력 예측 학습모델 설계

실(real) 데이터 기반의 학습모델을 개발하기 위해 실제 시험·운용 중인 OWC 파력발전 시스템의 외부파랑과 OWC 파고계 등에서 생산되는 센서 데이터 확보하였다. 필터링 등의 데이터 전처리 후에 학습셋(training set), 테스트셋(test set)으로 분류하여 학습모델 개발에 활용하였다.

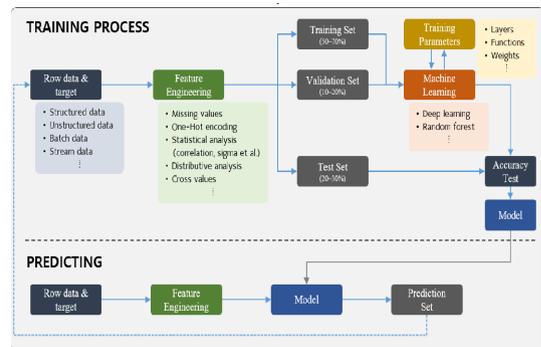


Fig. 2. Development process of pressure predictive model based on machine learning in OWC

데이터 분석 모델을 구축하기 위해 오픈소스 기반의 KNIME 플랫폼과 Tensorflow 활용하여 데이터 통합, 변환, 분석, 가시화 등의 머신러닝에 필요한 일련의 프로세스를 처리하고, 머신러닝 알고리즘을 검증하기 위해 머신러닝 프레임워크인 Tensorflow를 이용해서 DT(Decision Tree), RF(Random Forest), SVM, RNN 등의 알고리즘을 적용·시험하여 OWC 예측에 가장 효과적인 기법을 제시하고, 향후 센서의 종류/위치/데이터가 변경되는 경우에도 핵심 기법의 수정 없이 활용 가능한 학습모델 개발하였다. Fig. 2는 데이터 수집, Feature Engineering, 데이터셋 구성, 학습을 통한 OWC 압력 예측모델 개발의 전주기 프로세스를 보여준다.

3.1 데이터 수집

데이터 수집은 Table 2와 같이 외부파랑 정보 수집을 위한 OWC 파고계 데이터와 OWC와 파력발전소 데이터를 선정하였다.

Table 2. Data selection

구분	설명
파랑(wave) 데이터	HS, TP, 파향(wave direction) 등
OWC 및 파력발전 데이터	공압, 유속, 발전량 등

실제 시험·운용 중인 OWC 파력발전 시스템의 외부파랑 확보를 위하여 Fig. 3 같이 파력발전소에서 1km 거리에 있는 OWC 파고계에서 생산되는 Table 3 센서 데이터를 확보하여 활용하였다.



Fig. 3. Location of sensors and wave power plant

3.1 파고 데이터의 상관성 분석

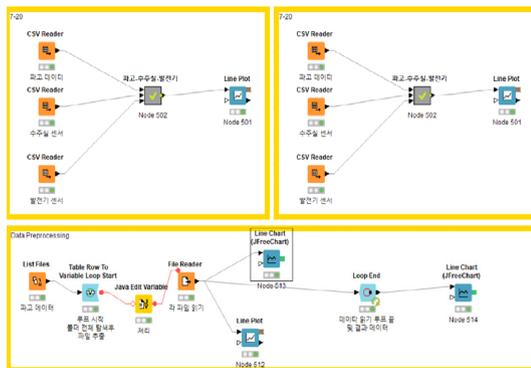


Fig. 4. Workflow for correlation analysis

Fig. 4 상관관계 분석 워크플로우는 파고계 데이터의 영향을 확인하기 위해 48시간 파고계로부터 측정된 파고 데이터의 Target feature(압력)와 전력간의 상관관계 분석을 목적으로 설계되었고 파고계 데이터의 경우 파랑 에너지(파고2x파주기)로 변환하여 활용하였다.

Table 3. Raw data collected from wave-height meter

필드명	설명
Date/Time	날짜와 시간
T	파도의 평균 주기
T(Hmax)	파도의 최대 주기
H	평균 파고
Hmax	최대 파고
Direction	파향

수집주기(30분), 수집시간 (48시간)

Table 4. Raw data related to chamber, power generation (such as internal pressure, RPM, torque, etc). Data selection

필드명	설명
Turbine_V	터빈에너지
SG	유속
Press_R	수주실 압력
TORQUE	모터의 회전력
RPM	모터의 1분동안의 회전수
POWER	전력량

수집주기(1초), 수집시간(48시간)

입력 데이터는 Table 3 파고, Table 4 수주실 정보를 포함한 발전기 데이터를 선정하고, 파력발전소에서 1km에서 떨어진 파고계의 데이터와 발전소의 영향에 대하여 분석하는 데 활용하였다.

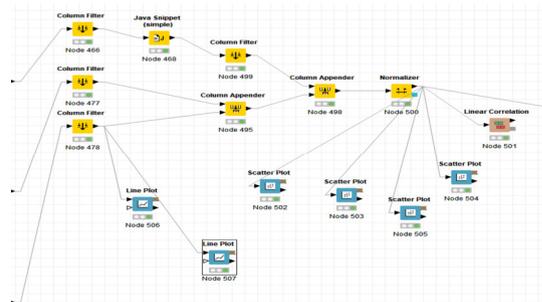


Fig. 5. Metanode workflow for detail correlation analysis

Fig. 5는 상관관계 분석을 위한 [파고수주형 발전기] 메타노드의 상세설계로 데이터 전처리와 상관관계 분석을 위한 워크플로우이며 상관관계 분석을 위한 전처리 과정에서는 정규화 등의 전처리 등을 수행하였다. 분석 기법은 SWAN 모델과 선형 상관(Linear Correlation)을 사용하여 분석하였다. 파랑 에너지와 압력 데이터는 시간 축을 맞추기 위하여 계측된 파랑 에너지의 최소 간격인 30분을 기준으로 하였으며, Fig. 6은 48시간 동안 30분 단위의 분석을 수행하였다. 지리적으로 떨어져 있는 파고 센서와 파력발전소의 영향을 분석하기 위하여 SWAN모델을 활용하였고, 분석결과 30분이라는 다소 긴 시간에도 불구하고 1km라는 거리로 인한 시간 지연과 편차로 인하여 이를 활용하기는 어려운 것으로 판단되었다.

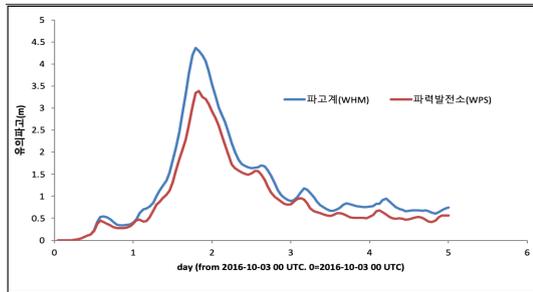


Fig. 6. Analysis result of wave with power generator

따라서 파고 데이터를 활용하는 것은 시간 지연과 편차의 문제로 활용하는데 제약이 있어 OWC와 발전기 데이터를 활용하기 위한 분석을 수행하였다. OWC와 발전기 데이터를 활용하는 이유는 Fig. 7과 같이 OWC는 파도 에너지의 영향을 받아 압력 에너지를 생성하고 그 압력은 유속 등의 에너지로 변환되어 터빈을 가동하는 구

조를 가지기 때문에 긴밀한 상관관계가 있을 것으로 판단되기 때문이다.

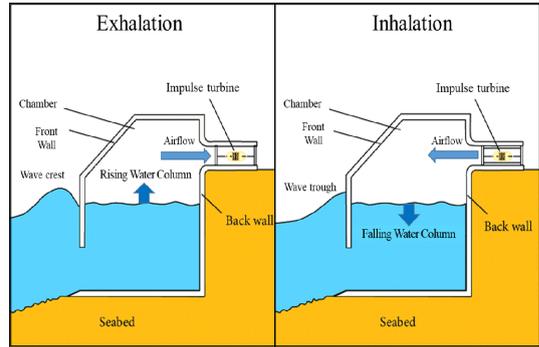


Fig. 7. Schematic layout of OWC

Fig. 7의 현상을 데이터로 분석해보기 위하여 OWC와 전력(IG/SG)은 계측된 파고 데이터의 최소 간격인 1초를 기준으로 수집하였으며, Table. 5는 48시간 동안 1초 단위의 OWC와 발전기의 상관관계 분석결과를 보여준다. 분석결과 OWC의 압력과 전력량 등의 발전기 데이터들과는 높은 양의 상관관계를 가짐을 볼 수 있다. 따라서 OWC 압력이 파력발전기의 운영을 위해서 매우 중요한 변수임을 추측할 수 있어 머신러닝을 활용한 OWC의 미래 데이터 예측이 파력발전기를 운영하는데 있어 매우 중요하다고 볼 수 있다.

3.2 데이터셋 설계 및 구축

3.2.1 데이터 전처리

Table 5. Result of Correlation analysis

구분	Press_R (OWC압력)	SG (유속)	TORQUE (모터의 회전력)	RPM (모터의 회전수)	POWER (전력량)
Press_R (OWC압력)	1	0.976678	0.919996	0.913444	0.905568
SG (유속)	0.314410	1	0.452299	0.452299	0.452299
TORQUE (모터의 회전력)	0.392105	0.681115	1	0.952299	0.952299
RPM (모터의 회전수)	0.324449	0.612237	0.892699	1	0.952299
POWER (전력량)	0.352299	0.652211	0.852299	0.852299	1

Table 6. Feature Engineering for data preprocessing

기법 항목	세부내용
누락된 값이 많은 열 제거	· 누락된 값이 많은 열은 입력 데이터로 활용 불가
누락된 값 처리	· 평균값이나 계산된 값으로 대체 · 일반적으로 누락 값을 1이나 0으로 변환
정수 인코딩	· 문자형 데이터가 존재하는 경우 · 단어별로 클래스화한 후, 인덱스로 변환
중복적이고 상호연관성이 높은 변수 제거	· 복제데이터, 중복데이터 또는 상호연관성이 높은 변수 최소화 · 산점도(scatter plot), PCA(Principal Component Analysis) 등 활용
편향된 값 수정	· 연속변수는 정규분포를 따라야 함 · 분포 경향 검토 후 편향된 값은 보정 · 검토: 통계 척도 및 히스토그램 활용 · 보정: 로그변환, 역수, 제곱근 변환, 제곱 등 활용
극단 값과 특이값 처리	· 극단 값은 되도록 활용 · 특이값(표준편차의 3배 이상 벗어난 값)은 중요하지 않을 경우 제거

데이터 전처리를 위해서는 도메인 지식으로부터 Feature를 만드는 과정이 필요하며, 머신러닝 모델을 위한 데이터 테이블의 컬럼(Feature)을 생성하거나 선택하는 작업을 의미하며 입력 데이터를 만들기 위해 주어진 원시 데이터로부터 특징을 가공하고 생성하는 일련의 과정을 거친다. Feature Engineering은 모델 성능에 미치는 영향이 매우 크기에 학습용 데이터 생성을 위해 Table 6과 같은 전처리 기법을 사용한다. 예측모델은 일반적으로 올바른 데이터를 가정으로 작동하기 때문에 파고계 데이터와 내부 압력 및 발전데이터의 분석을 통한 학습 데이터 생성을 위한 정제 방법 도출한다. 우선적으로 Feature들의 상관관계 분석을 수행하기 위해 계측 주기가 다른 센서 데이터들의 간격을 맞추기 위한 처리와 포맷을 맞추는 전처리를 수행하였다. Target Feature(압력)를 선정하고 Target Feature을 기준으로 센서 데이터의 Feature별 상관관계 분석을 수행하여 Input Feature를 선정하는 데이터셋 구축 방안을 수립하고 이를 수행하였다.



Fig. 8. Preprocessing workflow to build a dataset

Fig. 8은 상관관계 분석을 위한 전처리 워크플로우를 구성하여 파고 데이터들을 폴더에서 읽어서 파일 목록을 얻고 이를 통해 입력 데이터를 고려하였다. 시계열을 고려한 데이터 분석을 위하여 데이터의 시계열과 기간을 입력하여 처리할 수 있도록 노드를 구성하였고 해당 데이터를 구성하기 위한 노드를 추가적으로 디자인하여 데이터의 필터링, 병합, 컬럼 변경 등의 처리를 지원하도록 하였다. 데이터 내 각각의 클래스들이 차지하는 데이터의 비율이 균일하지 않고 한쪽으로 치우친 데이터 불균형(imbalanced data)을 해소하기 위해 분류 알고리즘의 성능을 향상시키거나, 학습 전에 리샘플링(re-sampling)을 통해 학습셋 내의 클래스 균형을 맞추는 방법을 고려하여 높은 비중을 차지하는 클래스의 값을 임의로 제거하는 오버샘플링(over-sampling)과 낮은 비중을 차지하는 클래스의 값을 임의로 늘리는 언더샘플링(under-sampling)을 시도하였다.

3.2.2 데이터셋 구축

머신러닝을 위한 데이터셋은 크게 학습셋과 테스트 셋으로 구분할 수 있고, 데이터가 충분히 확보된 경우에는 학습셋, 테스트셋과 검증 셋으로 구분하여 데이터를 추가로 검증이 가능하다. 원시 데이터로부터 데이터 패턴을 학습하기 위해 학습 데이터셋, 모델이 실제 시나리오에 따른 동작을 통한 모델 성능 확인을 테스트 데이터셋으로 구성하고, 튜닝 및 평가를 위한 단계로 검증 데이터셋으로 구성이 일반적이다. 하지만, 본 예측모델은 머신러닝 알고리즘 중 신경망 학습에 적합성이 높다고 판단되었고 원시 데이터셋이 크지 않아 학습셋(70%)과 테스트셋(30%)만으로 구성하여 Fig. 9와 같이 예측모델을 도출했다.

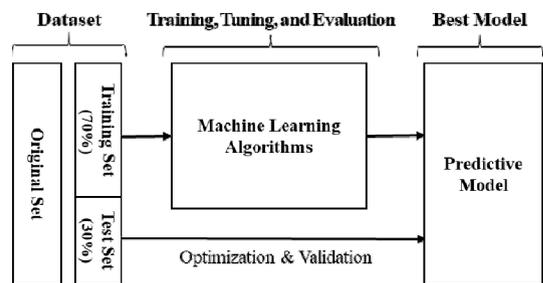


Fig. 9. Building Predictive model according to dataset

Fig. 10, Fig 11과 같은 기준을 토대로 OWC 압력 예측을 위한 학습 데이터셋을 구축했다. 학습 속도 및 정확도 측면에서 이점을 가지기 위해 특정 시간 동안에 피크 압력치가 발생할 경우 0, 그렇지 않은 경우를 1로 지정하여 분류 문제로 단순화하였다. 또한, 압력에 대해 일정한 시간(1분 40초)동안의 시계열 데이터를 입력 특성으로 지정하여 1일 100개의 입력과 1개의 대상 특성으로 구성된 985개의 레코드를 가지는 데이터셋을 구축하였다.

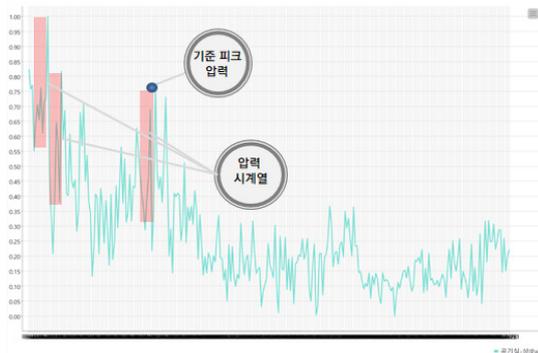


Fig. 10. Input/target feature criteria setting for building dataset

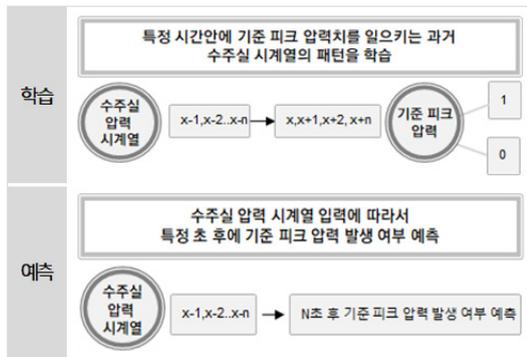


Fig. 11. Target feature and prediction method

4. 학습모델 구현 및 평가

4.1 머신러닝을 통한 학습모델

본 연구는 구축한 데이터셋은 대상 특성이 '0'과 '1'로 구성되어 있어 분류 방법을 활용할 수 있으나, 더 높은 정확도와 유연성을 확보하기 위해 회귀분석 출력층의 활성화 함수를 분류함수 형태로 수정하여 학습에 활용하였다.

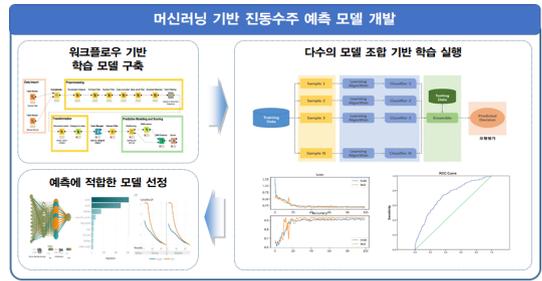


Fig. 12. Development process of pressure prediction model

Fig. 12에서와 같이 워크플로우 기반의 학습모델을 구축하여 모델별로 학습을 통해 적절한 모델을 선정한다. 선정된 학습모델로 테스트셋을 이용하여 예측 정확도와 정밀도 수준을 평가한다.

4.2 워크플로우 기반 학습모델 설계 및 구현

본 연구에서는 대규모 센서 데이터 수집과 시간 소요가 많은 인공지능 계산 지원을 위한 플랫폼을 개발하였다. 빅데이터 기반의 고성능 데이터 분석 및 처리 플랫폼 구조는 Fig. 13과 같으며 그에 대한 설명은 아래와 같다.

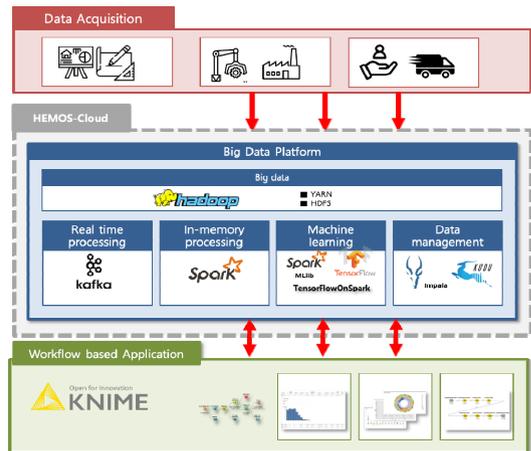


Fig. 13. Structure of Bigdata & AI platform

- (1) HPC 클라우드: 사용자는 인터넷 연결을 통해 원격으로 접근하여 플랫폼이 제공하는 분석 기능을 사용한다.
- (2) 빅데이터 모듈: 빅데이터를 수집, 처리, 저장, 분석을 지원한다.
- (3) 데이터 수집 모듈: 원본 데이터를 수집하여 빅데이터 플랫폼에 공급하는 역할을 한다.

- (4) 실시간 데이터 처리 모듈: IoT 센서 스트림 데이터를 실시간으로 처리 및 분석
- (5) 인공지능 기반 분석 라이브러리: 실시간 또는 배치 분석에 필요한 인공지능 기반 분석 기능 함수를 제공한다.
- (6) 응용 서비스: 실시간 데이터 모니터링, 데이터 브라우징, 분석 결과, 프로세스 워크플로우 편집, 빅데이터 클러스터 상태 모니터링 등의 기능을 제공한다.

빅데이터 플랫폼의 워크플로우 기능을 활용하여 OWC 압력 예측에 적합한 워크플로우 기반의 학습모델 구축하였다. 이는 데이터 정제/변환, 분석, 모델링, 시각화 및 출력까지 한 번에 동작하며, 모든 메뉴 및 과정은 GUI로 구현한 워크플로우에서 제공된다. 먼저 다양한 머신러닝 기법 적용 및 테스트를 통해 최적의 기법(조합) 선정하기 위해 머신러닝 방법을 메타노드로 구성하여 병렬 학습이 가능하도록 구성하고 선행회귀 분석을 위한 알고리즘을 적용하여 분석을 수행하여 예측에 적합한 모델을 선별하였다.

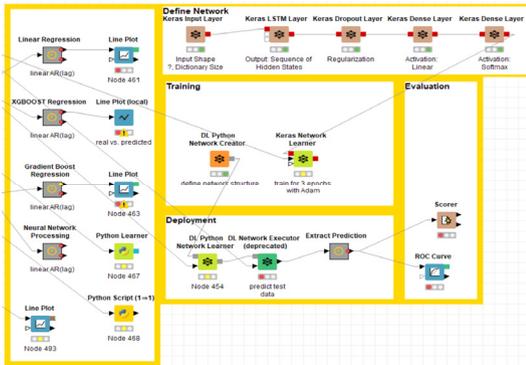


Fig. 14. Workflow for implementation of machine learning algorithms

Fig. 14는 OWC 파력발전장치의 압력에너지 예측을 위해 설계한 워크플로우 방식의 머신러닝 프로세스이며, 학습을 수행하기 위해 활용된 학습모델들을 나타낸다. OWC 파력발전장치의 압력에너지 예측을 위해 다양한 머신러닝 학습모델을 활용하였다. 각 학습모델을 하나의 전처리 프로세스와 연결하여 병렬로 학습을 수행하도록 구성하였고, 머신러닝 방법을 메타노드로 구성하여 병렬 학습이 가능하도록 하였다. 적용한 머신러닝 알고리즘으로는 선형회귀 알고리즘, 머신러닝 앙상블 기법 중 과적

합(overfitting) 방지가 가능하고, 분산/병렬 처리가 가능한 XGBoost 알고리즘, 딥러닝 기반 LSTM 알고리즘이며, 입력 레이어, 히든레이어, 출력 레이어 등을 추가적으로 구성하여 워크플로우를 설계하였다. 이때, 데이터셋을 나누고 학습 노드와 예측 노드를 이용하여 예측 결과값을 계산한다. 또한, 목적에 적합한 학습모델 구성을 위하여 파이썬(Python) 기반의 코드로 필요한 모델을 추가로 구현하였다.

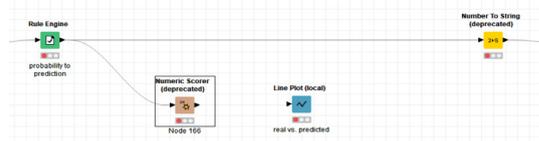


Fig. 15. Workflow for prediction evaluation

Fig. 15과 같이 머신러닝 모델에 맞게 예측값을 비교하기 위한 추가적인 메타노드를 구성하였고, 결과를 스코어와 그래프로 표현하도록 설계하였다.

4.3 학습모델 평가

개발된 학습모델에 대해 테스트셋을 이용하여 예측 정확도를 평가하였고, 가장 신뢰 있는 정확도를 가지는 모델을 개발하기 위해 베이지안 최적화 (Bayesian Optimization) 기법을 이용하였다.

최종적인 딥러닝 학습은 파라미터 최적화를 수행한 파라미터 결과(LSTM Layer : 2, optimizer : adam, batch_size : 30, epochs : 100)를 토대로 수행하였다.

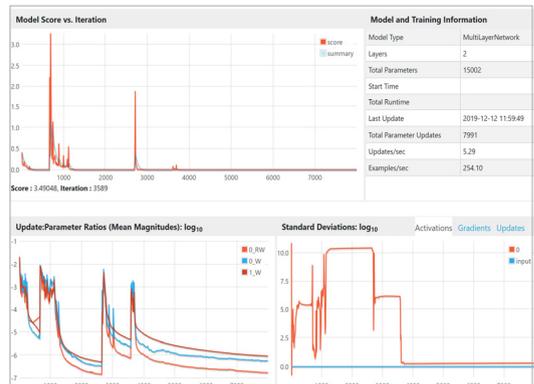


Fig. 16. Training and evaluation result

Fig. 16은 머신러닝 학습 및 평가의 결과를 실시간으로

로 제공하는 그림으로 사용자가 결과를 다양한 관점에서 검토할 수 있도록 해주었다. 학습결과 RNN을 확장/개선한 LSTM 모델의 정확도(Accuracy)는 88%로 가장 높았으며, 학습에 활용되지 않은 테스트 데이터셋을 압력에너지 예측에 활용했을 때의 정밀도(Precision)는 81% 수준으로 나타났다. 결과에서 정확도는 모델이 데이터를 얼마나 정확하게 분류했는지의 지표이다. 많은 머신러닝 시스템은 데이터를 활용하여 몇 가지 선택지 중에 가장 적합한 것을 고르는데, 이때 선택이 올바른지 아닌지를 측정하여 수치화시킨다. 정밀도는 양성으로 예측한 데이터 중에서 얼마나 정답을 맞췄는지를 평가한다. LSTM 모델은 히말라야 지역의 용설(snowmelt) 예측[17], 석유생산 예측[18] 교통통제, 경로선택 및 계획 제공을 위한 이동 시간 예측[19]과 같은 시계열에 따른 장기 의존성(Long-Term Dependency)이 높은 여러 연구에서 효과적인 기법으로 활용하고 있고, 직전 데이터뿐만 아니라 거시적으로 과거 데이터를 고려할 때 미래의 예측에 적합하다고 알려져 있다.

5. 결론

본 연구는 현재 운용 중인 용수 파력발전소의 IoT 센서 데이터 기반으로 디지털 트윈의 요소기술인 기계학습 기반의 예측 기술을 적용하여 그 유효성을 확인하였다. 파력발전소의 경우 파고 데이터, OWC의 센서 데이터 등과 같은 다수의 파라미터에 의해 의미 있는 데이터 정의와 함께 다수의 파라미터로부터 상관관계를 분석하고 이를 통해 원하는 정보를 예측할 수 있는 방법론 연구를 수행하였다. 원시 데이터셋을 기반으로 Feature Engineering을 통해 머신러닝 모델에 적합한 Feature를 도출하고 데이터 전처리와 거친 데이터를 기반으로 머신러닝의 여러 모델에 학습시켜 스코어가 높은 모델을 선정하였다. 그 결과, RNN의 확장형인 LSTM 알고리즘이 예측 정확도가 높은 모델임을 확인하였고, 정확도와 정밀도는 각각 88%, 85%로 나타났다. 본 연구를 통해, 파력발전기의 발전량과 고장 등에 밀접한 관계가 있는 OWC의 압력을 예측함으로써 운영의 효율화를 기대할 수 있었고, 파력발전 시스템의 디지털 트윈으로 스마트 운용에 필요한 머신러닝 기술의 유효성을 확인할 수 있었다.

본 연구에서는 IoT 센서로부터 충분한 데이터가 확보되지 않아 기술적용의 타당성 연구로 일부 제한되었다. 향후에는 OWC 파력발전기의 스마트 운용 및 유지보수

를 위한 정확성을 높인 예측분석 모델 설계와 실시간의 대규모 데이터 분석 연구가 추가적으로 요구된다. 특히 입력 벡터 수의 증가와 이에 따른 학습 및 분석 소요시간의 급격한 증가 문제가 예상되기 때문에 이에 따른 머신러닝 기반의 OWC 압력 예측모델 고도화, 딥러닝 모델의 하이퍼파라미터(hyperparameter) 튜닝으로 정확성 향상 연구와 실시간 대규모 센서 데이터 처리를 위한 분산/병렬 기반의 학습 환경에 대한 후속연구가 필요하다.

References

- [1] H. S. Lee, J. S. Oh, "Study of Power Output Characteristics of Wave Energy Conversion System According to Turbine Installation Method Combined with Breakwater", *Journal of Ocean Engineering and Technology*, Vol.29, No.4, pp.317-321, Aug. 2015. DOI: <http://dx.doi.org/10.5574/KSOE.2015.29.4.317>
- [2] J. S. Kim, J. Y. Park, J. S. Choi, S. H. Shin, "Analysis of Wave Energy Resource in the West Sea of Jeju based on Wave Observation Data", *Journal of the Korean Society for Marine Environment & Energy*, Vol.23, No.1, pp.37-46, Feb. 2020. DOI: <https://doi.org/10.7846/JKOSMEE.2020.23.1.37>
- [3] K. H. Lee, J. H. Lee, I. H. Jeong, D. S. Kim, "3-Dimensional numerical analysis of air flow inside OWC type WEC equipped with channel of seawater exchange and wave characteristics around its structure (in case of regular waves)", *Journal of Korean Society of Coastal and Ocean Engineers*, Vol.30, No.6, pp.242-252, Dec. 2018. DOI: <https://doi.org/10.9765/KSCOE.2018.30.6.253>
- [4] M. Kharati-Koopae, A. Fathi-Kelestani, "Assessment of oscillating water column performance: Influence of wave steepness at various chamber lengths and bottom slopes", *Renewable Energy*, Vol.147, pp.1595-1608, Mar. 2020. DOI: <https://doi.org/10.1016/j.renene.2019.09.110>
- [5] R. Cascajo, E. García, E. Quiles, A. Correcher, F. Morant, "Integration of marine wave energy converters into seaports: a case study in the Port of Valencia", *Energies*, Vol.12, No.5, pp.787. Feb. 2019. DOI: <https://doi.org/10.3390/en12050787>
- [6] Y. Cui, Z. Liu, X. Zhang, C. Xu, "Review of CFD studies on axial-flow self-rectifying turbines for OWC wave energy conversion", *Ocean Engineering*, Vol.175, pp.80-102, Feb. 2019. DOI: <http://dx.doi.org/10.1016/j.oceaneng.2019.01.040>
- [7] C. Roh, K. H. Kim, J. H. Kim, K. Y. Hong, "A Comparison of PWM methods for Power Converter of OWC Wave Power Generation System", *Journal of Korean Society for Marine Environment & Energy*, Vol.22, No.2, pp.84-94, May. 2019.

DOI: <https://doi.org/10.7846/jkosmee.2019.22.2.84>

- [8] S. Park, KH Kim, B. W. Nam, JS Kim, K. Hong. "A Study on the Performance Evaluation of the OWC WEC Applicable to Breakwaters using CFD", *J. of the Korean Society for Marine Environment & Energy*, Vol.21, No.4, pp.317-327, Nov. 2018.
DOI: <https://doi.org/10.7846/JKOSMEE.2018.21.4.317>
- [9] H. A. Abdelhafez, A. Amer. "The Challenges of Big Data Visual Analytics and Recent Platforms", *World of Computer Science & Information Technology Journal*, Vol.9, No.6, pp.28-33, 2019.
- [10] A. Fillbrunn, C. Dietz, J. Pfeuffer, R. Rahn, G. A. Landrum, M. R. Berthold, "KNIME for reproducible cross-domain analysis of life science data", *Journal of Biotechnology*, Vol.261, pp.149-156. Nov. 2017.
DOI: <http://dx.doi.org/10.1016/j.jbiotec.2017.07.028>
- [11] S. Bonthu, K. H. Bindu, "Review of Leading Data Analytics Tools", *International Journal of Engineering & Technology*, Vol.7, No.3.31, pp.10-15, 2018.
DOI: <http://dx.doi.org/10.14419/ijet.v7i3.31.18190>
- [12] Y. G. Kim, K. I. Moon, "Clustering of Smart Meter Big Data Based on KNIME Analytic Platform. The Journal of The Institute of Internet", *Broadcasting and Communication*, Vol.20, No.2, pp.13-20, Apr. 2020.
DOI: <https://doi.org/10.7236/JIIBC.2020.20.2.13>
- [13] Datasolution. Predictive analysis solution, KNIME Analytics Platform [Internet], c2016 [cited 2016]. Available From: <http://www.datasolution.kr> (accessed Aug. 10, 2020)
- [14] P. L. Bartlett, P. M. Long, G. Lugosi, A. Tsigler, "Benign overfitting in linear regression", *Proceedings of the National Academy of Sciences*, March 2020.
DOI: <https://doi.org/10.1073/pnas.1907378117>
- [15] M. Sevkli, A. Z. Sevkli, O. Cosgun, "Comparison of Data Mining and Mathematical Models for Estimating Fuel Consumption of Passenger Vehicles", *MWAIS 2019 Proceedings*, 34, 2019.
- [16] D. Qin, J. Yu, G. Zou, R. Yong, Q. Zhao, B. Zhang, "A Novel Combined Prediction Scheme Based on CNN and LSTM for Urban PM_{2.5} Concentration", *IEEE Access*, 7, 20050-20059, Feb. 2019.
DOI: <https://doi.org/10.1109/ACCESS.2019.2897028>
- [17] S. Thapa, Z. Zhao, B. Li, L. Lu, D. Fu, X. Shi, B. Tang, H. Qi, "Snowmelt-Driven Streamflow Prediction Using Machine Learning Techniques (LSTM, NARX, GPR, and SVR)", *Water*, Vol.12, No.6, pp.1734, Feb. 2020.
DOI: <http://dx.doi.org/10.3390/w12061734>
- [18] Y. Duan, L. V. Yisheng, F. Y. Wang, "Travel time prediction with LSTM neural network", *2016 ITSC*, IEEE, Rio de Janeiro, Brazil, pp.1053-1058, Nov. 2018.
DOI: <http://dx.doi.org/10.1109/ITSC.2016.7795686>
- [19] X. Song, Y. Liu, L. Xue, J. Wang, J. Zhang, J. Wang, Z. Cheng, "Time-series well performance prediction based on Long Short-Term Memory (LSTM) neural network model", *Journal of Petroleum Science and*

Engineering, 186, 106682. 2020

DOI: <http://dx.doi.org/10.1016/j.petrol.2019.106682>

서 동 우(Dong-Woo Seo)

[정회원]



- 2008년 2월 : 전남대학교 산업공학과 (공학석사)
- 2012년 2월 : 전남대학교 산업공학과 (공학박사)
- 2012년 10월 ~ 현재 : 한국과학기술정보연구원 선임연구원

<관심분야>

VR/AR based post-processing, Digital Product Design, Collaboration Interaction, Service Strategy

허 태 상(Taesang Huh)

[정회원]



- 2000년 2월 : 성균관대학교 전기전자컴퓨터공학부 (공학사)
- 2002년 8월 : 광주과학기술원 정보통신공학 (공학석사)
- 2017년 2월 : 현재 배재대학교 컴퓨터공학과 (공학박사)
- 2002년 ~ 현재 한국과학기술정보연구원 책임연구원

<관심분야>

데이터 플랫폼, 데이터 분석, e-Science, 분산 컴퓨팅, 메타 데이터 카탈로그, 디지털 트윈

김 명 일(Myoung-il Kim)

[정회원]



- 2000년 2월 : 중앙대학교 일반대학원 컴퓨터공학과 (공학석사)
- 2004년 2월 : 중앙대학교 일반대학원 컴퓨터공학과 (공학박사)
- 2004년 3월 ~ 현재 : 한국과학기술정보연구원 책임연구원

<관심분야>

Modeling & Simulation, 슈퍼컴퓨팅, 과학기술연구망

오 재 원(Jae-Won Oh)

[정회원]



- 2013년 2월 : 한양대학교 기계공학
학과 (공학석사)
- 2016년 8월 : 한양대학교 기계공학
학과 (공학박사)
- 2016년 9월 ~ 2018년 12월 : 한국
해양과학기술원 부설 선박해양
플랜트연구소 연수연구원
- 2018년 12월 ~ 현재 : 한국해양과학기술원 부설 선박해
양플랜트연구소 선임연구원

<관심분야>

다물체동역학, 기계설계, CAE, 시뮬레이션 기반 설계

조 수 길(Su-Gil Cho)

[정회원]



- 2008년 2월 : 한양대학교 기계공
학과 (공학사)
- 2015년 2월 : 한양대학교 미래자
동차공학과 (공학박사)
- 2015년 2월 ~ 2016년 7월 : 한국
해양과학기술원 부설 선박해양플
랜트연구소 연수연구원
- 2016년 7월 ~ 현재 : 한국해양과학기술원 부설 선박해양
플랜트연구소 선임기술원

<관심분야>

기계설계, 다분야통합최적화, 디지털트윈, RAMS