

# Learning Method for Regression Model by Analysis of Relationship Between Input and Output Data with Periodicity

Hye-Jin Kim<sup>†</sup> · Ye-Seul Park<sup>††</sup> · Jung-Won Lee<sup>†††</sup>

## ABSTRACT

In recent, sensors embedded in robots, equipment, and circuits have become common, and research for diagnosing device failures by learning measured sensor data is being actively conducted. This failure diagnosis study is divided into a classification model for predicting failure situations or types and a regression model for numerically predicting failure conditions. In the case of a classification model, it simply checks the presence or absence of a failure or defect (Class), whereas a regression model has a higher learning difficulty because it has to predict one value among countless numbers. So, the reason that regression modeling is more difficult is that there are many irregular situations in which it is difficult to determine one output from a similar input when predicting by matching input and output. Therefore, in this paper, we focus on input and output data with periodicity, analyze the input/output relationship, and secure regularity between input and output data by performing sliding window-based input data patterning. In order to apply the proposed method, in this study, current and temperature data with periodicity were collected from MMC(Modular Multilevel Converter) circuit system and learning was carried out using ANN. As a result of the experiment, it was confirmed that when a window of 2% or more of one cycle was applied, performance of 97% or more of fit could be secured.

Keywords : Neural Network, Fault Diagnosis, Sensor Data, Regression Model, Thermal Estimation

## 주기성을 갖는 입출력 데이터의 연관성 분석을 통한 회귀 모델 학습 방법

김 혜 진<sup>†</sup> · 박 예 슬<sup>††</sup> · 이 정 원<sup>†††</sup>

## 요 약

최근 로봇이나 설비, 회로 등에 센서 내장이 보편화 되고, 측정된 센서 데이터를 학습하여 기기의 고장을 진단하기 위한 연구가 활발하게 수행되고 있다. 이러한 고장 진단 연구는 고장 상황이나 종류를 예측하기 위한 분류(Classification) 모델 개발과 정량적으로 고장 상황을 예측하기 위한 회귀(Regression) 모델 개발로 구분된다. 분류 모델의 경우, 단순히 고장이나 결함의 유무(Class)를 확인하는 반면, 회귀 모델은 무수히 많은 수치 중에 하나의 값(Value)을 예측해야 하므로 학습 난이도가 더 높다. 즉, 입력과 출력을 대응시켜 고장을 예측을 할 때, 유사한 입력값이 동일한 출력을 낸다고 결정하기 어려운 불규칙한 상황이 다수 존재하기 때문이다. 따라서 본 논문에서는 주기성을 지닌 입출력 데이터에 초점을 맞추어, 입출력 관계를 분석하고, 슬라이딩 윈도우 기반으로 입력 데이터를 패턴화 하여 입출력 데이터 간의 규칙성을 확보하도록 한다. 제안하는 방법을 적용하기 위해, 본 연구에서는 MMC(Modular Multilevel Converter) 회로 시스템으로부터 주기성을 지닌 전류, 온도 데이터를 수집하여 ANN을 이용하여 학습을 진행하였다. 실험 결과, 한 주기의 2% 이상의 윈도우를 적용하였을 때, 적합도 97% 이상의 성능이 확보될 수 있음을 확인하였다.

키워드 : 신경망, 고장진단, 센서 데이터, 회귀 모델, 온도예측

## 1. 서 론

센서 데이터는 전류, 전압, 온도 등과 같이 로봇이나 회로

와 같은 산업 설비에 내장된 센서로부터 측정된다[1-2]. 특히, 센서 데이터는 고장이나 결함을 예측하기 위해 활용되는 데, 예측해야 하는 데이터에 따라 분류와 회귀 방식의 학습으로 분류된다(예: 범주형 데이터-분류, 연속형 데이터-회귀). 예를 들어, 진동 데이터를 이용하여 베어링의 결함을 분류하는 분류 모델[3]과 연속적인 온도 데이터의 값을 예측하여 냉각 시스템의 결함을 감지하는 회귀 모델[4] 등이 존재한다. 이 때, 분류 모델은 고장 혹은 결함의 종류(예: 마모, 기어 손상 등)를 확률 범위로 나뉜 클래스로 예측하지만, 회귀 모델은 무수히 많은 수치 중에 하나의 값을 예측해야 하므로 상대적

※ 이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2020R1A2C1007400).

† 비 회 원 : 아주대학교 전자공학과 학사

†† 준 회 원 : 아주대학교 AI융합네트워크학과 박사

††† 종신회원 : 아주대학교 전자공학과/AI융합네트워크학과 교수

Manuscript Received : November 23, 2021

First Revision : February 9, 2022

Accepted : March 8, 2022

\* Corresponding Author : Jung-Won Lee(jungwony@ajou.ac.kr)

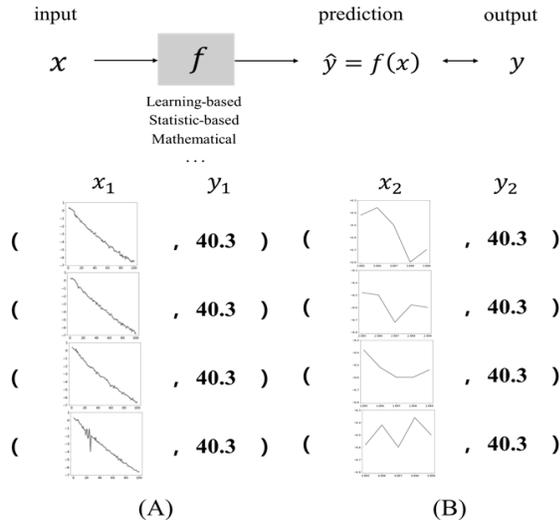


Fig. 1. Definition of Regularity between Input and Output Data

으로 학습 난이도가 더 높다.

일반적으로 회귀 모델은 데이터의 특성에 적합한 네트워크를 선택하여 학습한다. 그러나 적합한 네트워크 선택만으로는 높은 학습 성능을 확보하기 어렵다. 특히 학습에 사용되는 입력 데이터가 출력을 예측하기 위한 특징이 모호하다면, 아무리 데이터의 특성을 반영한 네트워크(예: 시계열 데이터-RNN, 이미지·영상-CNN 등)일지라도 학습 자체가 이루어지지 않기 때문이다. 구체적으로 학습 데이터는 입력과 출력이 짝을 이루어 존재하고, 유사한 입력으로부터 하나의 출력을 결정할 수 있을 때 규칙(Regularity)이 존재한다고 할 수 있다. 즉, Fig. 1(A)와 같이, 동일한 출력값( $y_1$ )을 예측할 때, 입력( $x_1$ )이 유사한 경우 출력 값( $y_1$ )을 예측하기 위한 규칙이 명확해진다. 하지만 Fig 1(B)과 같이, 동일한 출력( $y_2$ )을 예측해야 함에도 불구하고, 입력( $x_2$ )이 서로 다른 패턴을 지닌다면 출력 값( $y_2$ )을 예측하기 위한 규칙을 정의할 수 없어 학습 성능이 떨어진다. 따라서 본 논문에서는 회귀 모델 입출력 데이터의 특성 분석을 통해, 예측 성능을 높일 수 있는 방법을 제안하고자 한다.

데이터 특성 분석은 데이터가 선형인지 비선형인지, 시간에 따라 변화하는 데이터인지, 현재의 데이터가 이전 데이터에 영향을 받는지 등 여러 관점의 특성 분석을 의미한다. 이 과정에서 중요한 부분은 모델에 입력과 대응되는 출력이 높은 성능으로 예측할 수 있도록 입출력 데이터에 대한 연관성을 분석하는 것이다. 입력과 출력 사이에 명확한 규칙을 생성하기 어려운 경우, 모델의 학습 또한 어렵기 때문이다. 특히, 주기성이 있는 센서 데이터에서는 주기에 따라 증감이 달라지기 때문에 데이터 셋을 구성하는데 더욱 문제가 될 수 있다. 여기에서 주기성이 있는 데이터는 일정한 주기를 가지고 반복되는 데이터를 말한다.

Fig. 2를 살펴보면  $x_1, x_2$ 는 동일한 출력에서 서로 다른 입력 값을 가지고,  $x_3, x_4$ 는 동일한 입력에서 서로 다른 출력

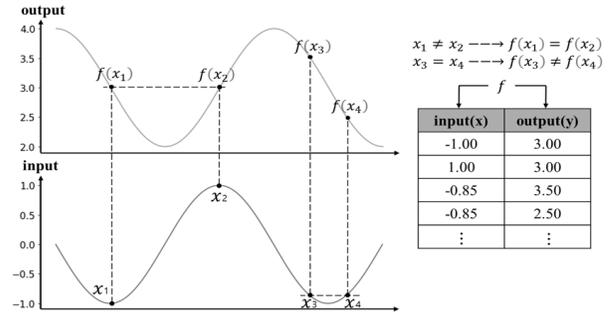


Fig. 2. Regularity Analysis of Periodical Input and Output Data

값을 가지는 것을 볼 수 있다. 따라서 주기성이 있는 입출력 데이터를 단순히 일대일로 대응시켜 회귀 모델의 학습을 수행한다면 동일한 입력에 서로 다른 출력을 예측해야 하는 모순적인 학습 과정이 발생하게 된다. 따라서 높은 성능을 확보하기 위해서는 동일한 출력을 예측할 때 유사한 입력 데이터가 구성되도록 데이터 셋을 구성하는 것이 필요하다.

따라서 본 논문에서는 주기성을 지닌 입출력 데이터의 연관성을 분석하여 데이터 셋을 구성하고, 회귀 모델을 학습하는 방법을 제안한다. 구체적으로 슬라이딩 윈도우 기법을 사용하여 윈도우 크기를 변경함에 따라 입출력 데이터 간 연관성이 높아지는 지점을 분석하고, 학습 성능이 우수한 윈도우 크기를 발견한다. 본 연구에서는 MMC(Modular Multilevel Converter) 회로 시스템으로부터 주기성이 있는 센서 데이터를 수집하여 회귀 모델을 학습하였다. 실험 결과, 한 주기의 10% 이상의 윈도우 크기인 경우,  $R^2$  성능이 0.9763로 가장 우수한 성능을 보이고, 2% 이상의 윈도우를 적용하였을 때, 97% 이상의 성능이 확보될 수 있음을 확인하였다.

## 2. 관련 연구

최근 시스템의 효율적인 운영을 보장하기 위해 센서 데이터를 활용한 고장, 결함 진단 연구가 활발히 수행되고 있다. [5]에서는 RNN(Recurrent Neural Network)을 적용하여 난방, 환기 시스템의 결함 감지 모델을 개발하였고, [6]에서는 센서 데이터를 이용하여 이상 소비 전력을 감지하는 ANN(Artificial Neural Network) 기반 회귀 모델을 개발하였다. 이때, 회귀 모델의 성능은 데이터의 품질, 학습 모델, 데이터 셋의 입력-출력의 구성에 영향을 받게 된다.

예를 들어, 센서 데이터는 실제 시스템에서 가져온 데이터로 노이즈, 결측 값과 같은 이상(Abnormal) 값을 나타낼 수 있다. 이상 값은 회귀 모델의 성능을 저하시키기 때문에, [7]의 연구에서는 노이즈, 결측 값을 처리하는 전처리를 통해 회귀 모델의 성능을 개선하는 방법을 제안하고 있다. [8-9]에서는 하이퍼 파라미터(예: 뉴런 수, 배치 크기 등) 튜닝을 통해 학습 성능을 개선한다. 그러나 하이퍼 파라미터 튜닝은 일정 수준 이상의 성능이 보장된 모델을 견고하게 고도화할 수 있지

만, 데이터 부족, 불균형, 불규칙성 등의 이유로 학습 자체가 이루어지지 않은 낮은 성능의 모델에서는 적용이 어렵다.

이로 인해, 근본적인 입력 데이터 변형을 통해, 학습 효율을 높이기 위한 시도가 있다. [10-12]에서는 슬라이딩 윈도우 기법을 적용하여 입력 데이터 셋 구성하고 회귀 모델의 성능을 개선한다. [10]의 연구에서는 주가를 예측하기 위한 회귀 모델 학습을 위해 중첩된 슬라이딩 윈도우 기법을 사용하여 데이터 셋을 구성한다. 그러나 선형 모델과 비선형 모델의 성능 비교에 초점을 맞추고 있어 입출력 데이터 분석을 통한 데이터 셋 구성을 확인하기 어렵다. [11]의 연구에서는 WMA (Weighted Moving Average)로 계산된 이전 날짜의 과거 데이터를 사용하여 데이터 셋을 구성하고 회귀 모델을 학습한다. 이처럼, 슬라이딩 윈도우 기반의 데이터 셋 구성은 이전 입력 값을 포함하여 단순히 값에 의한 예측이 아닌 추세에 의한 예측을 수행할 수 있게 한다. 그러나 [11]에서는 가장 높은 성능을 낼 수 있는 최적화된 입력 데이터 크기, 즉 윈도우의 크기를 측정하기 어렵다. [12]의 연구에서는 슬라이딩 윈도우 기법을 사용하여 시계열 센서 데이터의 이상 탐지 기반으로 전기 부하를 예측한다. 최적의 윈도우 크기를 결정하기 위해 1-10일 까지 윈도우의 크기를 증가시키며 학습 성능이 좋은 윈도우의 크기를 제시하고 있다. 그러나 [12]에서는 슬라이딩 윈도우 기반의 데이터 셋 구성을 수행하는 것이 입출력 데이터 간의 어떤 연관성을 제공하는 것인지 분석하기 어렵다. 따라서 본 논문에서는 입출력 데이터의 연관성 분석을 통해, 학습 효율을 올릴 수 있는 최적의 윈도우 크기를 확인하고 회귀 모델을 학습할 수 있는 방법을 제안한다.

### 3. 주기성을 가지는 입출력 데이터 간 연관성 분석 방법

회귀 모델링에서 입력 데이터와 출력 데이터는 크게 선형 관계 혹은 비선형 관계를 가질 수 있다. 3장에서는 선형 관계와 비선형 관계에 있는 데이터의 관계를 시각적으로 가시화하여 분석한다.

#### 3.1 선형/비선형 관계를 갖는 입출력 데이터 분석

선형 관계는 입력(X축) 데이터와 출력(Y축) 데이터가 직선 형태로 표현될 수 있는 관계를 의미한다. Fig. 3은 선형 관계를 갖는 입출력 데이터를 가시화한 것이다. 이에 대한 세부적인 분석은 다음과 같이 이어진다.

- (A) 양(Positive)의 선형 관계: 입력 데이터가 증가할 때 출력 데이터도 일정한 비율로 증가하는 것으로, 양의 기울기를 갖는 직선으로 표현된다.
- (B) 음(Negative)의 선형 관계: 입력 데이터가 증가할 때 출력 데이터는 일정한 비율로 감소하는 것으로, 음의 기울기를 갖는 직선으로 표현된다.

Fig. 3과 같이 선형 관계에 있는 입출력 데이터는 빨간색 직선으로 정의되며, 회귀 모델을 개발할 때 학습이 쉽게 이루어진다.

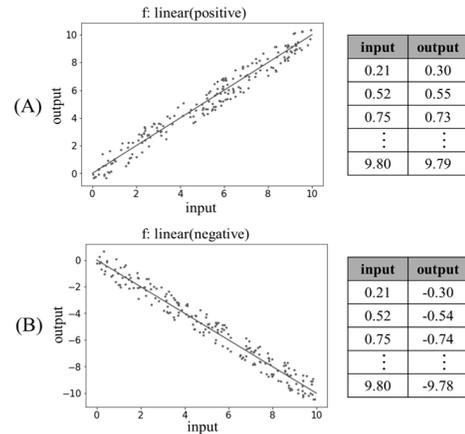


Fig. 3. Linear Relationship between Input and Output Data

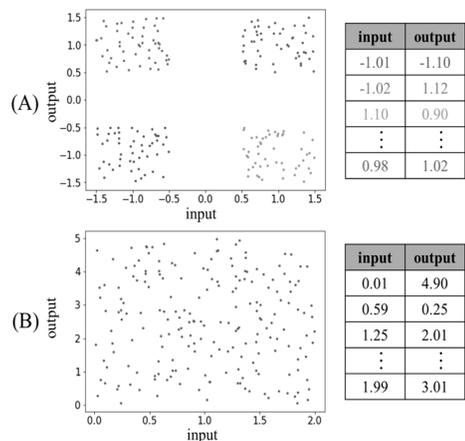


Fig. 4. Nonlinear Relationship between Input and Output Data

비선형 관계는 입출력 데이터 사이의 관계가 선형식(다항식)으로 표현될 수 없는 경우를 의미한다. Fig 4는 두 가지 비선형 관계를 나타낸다.

- (A) 군집화 가능 관계: (-1,-1), (-1,1), (1,-1), (1,1)을 중심으로 퍼져 있는 데이터를 나타낸 것으로 4개의 군집(빨강, 초록, 노랑, 파랑)으로 나누어진다.
- (B) 불규칙 관계: 무작위(Random)의 입력과 출력을 가지는 데이터로 입출력 사이에 아무런 관계를 표현할 수 없다. 일반적으로 데이터를 수집하게 되면 노이즈가 섞여 대부분 (B)와 같은 형태의 불규칙한 데이터가 얻어진다.

이와 같은 비선형 관계 데이터는 입력과 출력 사이의 관계를 표현하기 어려워 이를 예측하기 위한 함수를 쉽게 결정할 수 없다. 이로 인해, 회귀 모델을 학습하는 경우 낮은 성능을 보일 수 있다.

#### 3.2 주기성을 갖는 입출력 데이터의 특성 분석

비선형 관계 중에서, 입출력 데이터가 주기성을 갖는 경우, 생성되는 관계를 명확히 결정할 수는 없지만 특수한 패턴을 갖는 것을 확인할 수 있다. 주기성을 갖는 입출력 데이터는 비선형 관계이며, 세부적인 관계는 Fig. 5, Fig. 6과 같이 예시된다.

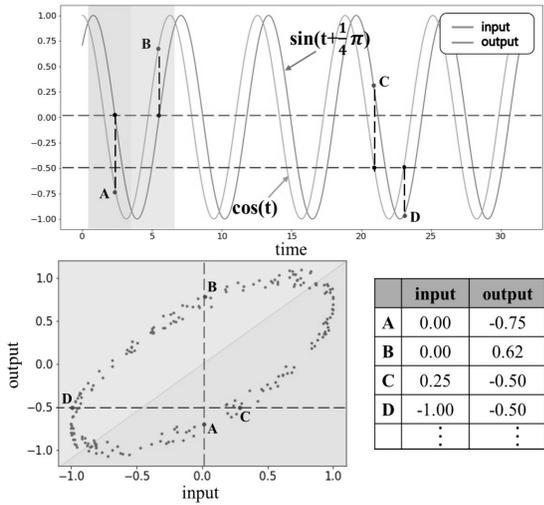


Fig. 5. Analysis of Periodic Input and Output Data (Input:  $\sin(t+0.25\pi)$ , Output:  $\cos(t)$ )

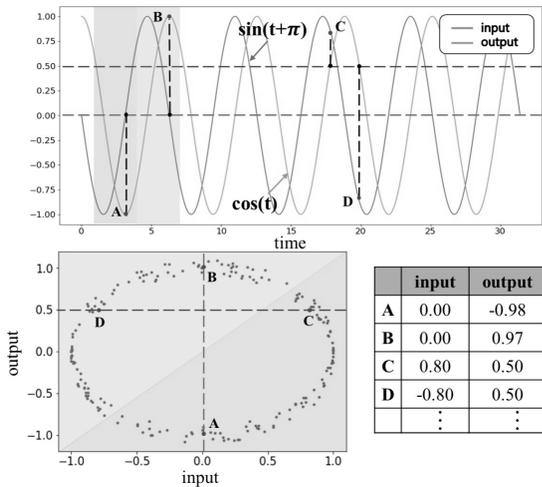


Fig. 6. Analysis of Periodic Input and Output Data (Input:  $\sin(t+\pi)$ , Output:  $\cos(t)$ )

(A) 입력:  $\sin(t+0.25\pi)$ , 출력:  $\cos(t)$

Fig. 5의 하단에서 입력이 증가할 때 출력이 증가하는 형태를 볼 수 있다. 구체적으로, Fig. 5 상단에서 파란색 음영 부분에서 입력이 감소할 때 출력이 더 작은 값을 가지고 감소하는 것에 대응되고, 초록색 음영 부분에서 입력이 증가할 때 출력이 더 큰 값을 가지고 증가하는 것에 대응된다. 또한, 그림에서 표시한 A-D는 앞서 Fig. 2에서 설명된 상황에 대응되어 설명될 수 있다. A와 B는 동일한 입력(0)에서 서로 다른 출력(-0.75, 0.62)을 가지고, C와 D는 동일한 출력(-0.5)에서 서로 다른 입력(0.25, -1)을 가지는 상황을 나타낸다.

(B) 입력:  $\sin(t+\pi)$ , 출력:  $\cos(t)$

Fig. 6의 상단에서 파란색 음영 부분은 입력과 출력의 교차 지점 이후 입력이 1이 될 때까지 입력이 증가할 때 출력이 감소하는 것을 나타내고 있다. 이후, 입력이 1에서 0까지 감소할

때 출력은 더 작은 값을 가지고 감소하는 것에 대응되고, 입력이 0보다 작아지면 입력은 감소하고 출력은 증가하는 것에 대응된다. 초록색 음영 부분도 이와 같이 대응되고, Fig. 6의 하단에서 원의 형태를 갖게 된다. 또한, A와 B는 동일한 입력(0)에서 서로 다른 출력(-0.98, 0.97)을 갖고, C와 D는 동일한 출력(0.5)에서 서로 다른 입력(0.75, -0.75)을 갖는다.

그러나 이와 같은 비선형 관계는 슬라이딩 윈도우 기법을 통해, 입력 데이터를 패턴화하여 입출력 간의 명확한 관계를 형성할 수 있다. 이에 대한 방법론은 4장에서 자세히 알아보도록 한다.

#### 4. 입력 데이터 패턴화를 통한 회귀 모델 학습

앞장에서 분석된 주기성을 지닌 입출력 데이터를 기준으로, 입출력 데이터 간의 대응 규칙성을 높이기 위한 방법을 제안한다. 먼저, 슬라이딩 윈도우를 활용하여 동일 출력을 예측할 때, 입력 데이터 간 패턴이 유사하도록 구성한다. 그리고 이를 출력과 대응시켜 입출력 데이터 셋을 구성하고, 인공 신경망 기반의 회귀 모델을 학습한다.

##### 4.1 슬라이딩 윈도우 기반 입력 데이터 패턴화

주기성을 갖는 데이터는 동일한 입력에서 서로 다른 출력을 갖기 때문에, 일대일 대응 방식으로 학습이 어렵다. 따라서 본 연구에서는 슬라이딩 윈도우 기법을 이용하여 입력 데이터를 패턴화하여 학습 데이터 셋을 구성하는 방법을 제안한다. 슬라이딩 윈도우 방식은 본래 데이터 프레임을 연속적으로 전송하기 위한 방법이다. 본 논문에서는 이 기법을 응용하여 학습의 대상이 되는 데이터를 현재의 입력 값뿐만 아니라, 해당 시점 이전의 입력 값을 포함하도록 하여 단순히 값에 의한 예측이 아닌 값의 추세에 의한 예측을 수행할 수 있게 한다. Fig. 7은 슬라이딩 윈도우를 이용한 데이터 셋 구성 방법을 나타낸다. 시간에 따른 입력과 출력 데이터를 나타내며, 입력의 다음 값을 출력 데이터 셋으로 구성한다. 윈도우의 크기를 3으로 가정하고 데이터 셋을 구성하면 첫 번째 데이터 셋은 (입력: 1-3, 출력: 4)로 구성되고, 다음 데이터 셋은 (입력: 2-4, 출력: 5)로 구성된다.

주기성이 있는 입출력 데이터에 슬라이딩 윈도우 데이터셋 구성 과정을 적용할 경우, 비선형 관계의 데이터를 선형 관계의 데이터로 표현할 수 있다. Fig. 8은 Fig. 6의 데이터를 슬라이딩 윈도우를 적용하여 가시화한 그림이다. Fig. 8 왼쪽의 그래프의 X축은 입력 값, 그리고 Y축은 입력 값에 대응되는 출력을 나타낸다. 그림에서 확인할 수 있듯이, 기존 데이터에서는 동일 입력에 여러 개의 출력값이 분산되어 대응되는 것을 확인할 수 있다. Fig. 8 오른쪽 그래프의 X축은 출력에 대응되는 입력을 윈도우로 구성하도록 하여, 해당 윈도우의 그래디언트(Gradient)를 특징으로 추출한 값이다. 그래디언트는 예측해야 하는 출력 부근의 입력 데이터 순간 변화율을 나타낸다. 또한, Y축은 해당 윈도우에 대응되는 출력 값을 나타

Time	1	2	3	4	5	...	n-2	n-1	n
Output									
Time	1	2	3	4	5	...	n-2	n-1	n
Input									

window size

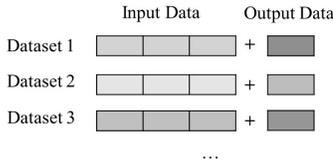


Fig. 7. Patterned Dataset using the Concept of Sliding Window

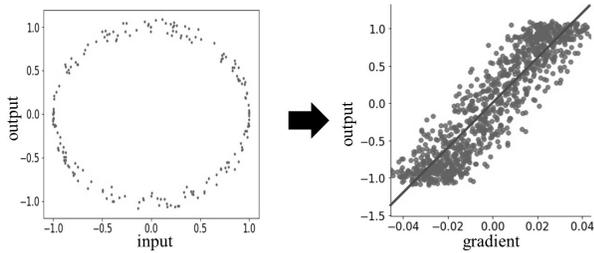


Fig. 8. Linear Relationship through a Sliding Window

낸다. 빨간색 선으로 표시한 부분을 살펴보면 양의 선형 관계로 표현되는 것을 확인할 수 있으며, 기존(왼쪽 그래프)에 비해 동일 입력에 대응되는 출력 값의 편차가 매우 줄어든 것을 확인할 수 있다.

이때, 입출력 데이터 셋 구성 과정에서 가장 중요한 것은 적절한 윈도우 크기를 설정하는 것이다. 동일 출력을 예측해야 하는 경우, 입력 데이터가 균일한 특징을 나타내고 있어야 학습이 수월하게 이루어지며, 이와 같은 특징이 포함되도록 윈도우 크기가 설정되어야 하기 때문이다. Fig. 9은 이와 같은 특징을 잘 보여주는 예시로, 동일한 출력을 예측할 때 윈도우 크기에 따른 데이터 3가지의 패턴을 나타낸다. 즉, 윈도우 크기는 하나의 출력을 예측하기 위한 시점 이전 입력 데이터 크기를 의미하며, (A), (B), (C)는 각각 윈도우 크기가 1, 5, 100인 경우이다.

- (A) window의 크기가 1인 경우: 슬라이딩 윈도우를 적용하지 않고 하나의 입력으로 하나의 출력을 예측하는 것이다. 그래프에서 볼 수 있듯이 동일한 출력을 예측해야 하지만, 입력이 모두 다른 값을 가지고 있어 규칙성이 없는 것을 확인할 수 있다.
- (B) window의 크기가 5인 경우: 윈도우의 크기가 작아 노이즈에 민감하게 반응하여 데이터의 패턴이 생성되지 않아 높은 규칙성을 확인할 수 없다.
- (C) window의 크기가 100인 경우: 노이즈에 민감하게 반응하지 않고, 동일한 출력을 예측함에 있어 균일한 패턴을 나타내는 것을 확인할 수 있다. 데이터의 추세를 포함하는 윈도우 크기일 경우 규칙성이 있는 데이터

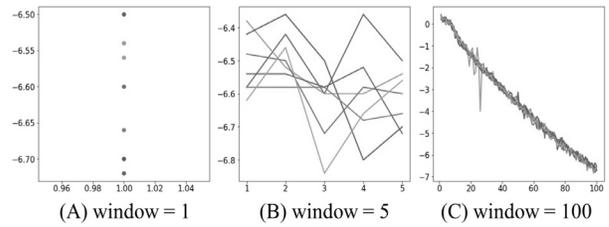


Fig. 9. Input Data Patterns According to Window Size

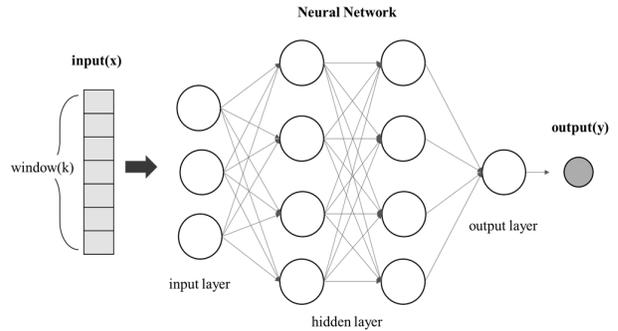


Fig. 10. Artificial Neural Network-based Regression Model Structure

셋을 구성할 수 있다. 다만, 윈도우의 크기가 커질 경우 계산 복잡도가 증가하기 때문에 무조건 크게 설정하지 않는다.

따라서 본 논문에서는 동일 출력을 예측할 수 있는 적합한 입력 데이터 크기를 정의하여, 인공신경망 기반의 회귀 모델을 설계한다.

#### 4.2 인공신경망 기반 회귀 모델 설계

Fig. 10는 인공신경망 기반의 회귀 모델 구조를 나타낸다. 슬라이딩 윈도우를 적용하여 구성된 데이터 셋이 회귀 모델에 입력된다. 입력층(Input Layer)에서는 윈도우(k)로 이루어진 초기 입력 데이터가 들어가는 부분으로 윈도우 크기의 입력 데이터가 개별 노드로 정의되어 은닉층으로 전달되고, 출력층에서는 최종 예측값을 반환한다. 입력층 뉴런의 계산 복잡도는 Equation (1)과 같고, 입력이 증가하면 계산 복잡도는 선형적으로 증가한다.

$$y = g(w_{11}x_1 + w_{12}x_2 + \dots + b_1) \quad (1)$$

Fig. 11은 회귀 모델의 예측 과정을 나타낸다. 슬라이딩 윈도우 크기만큼의 데이터(파란색 음영 박스 부분)가 모델에 입력되어, 윈도우의 바로 다음 값을 예측하게 된다. 학습을 수행하는 경우, 초록색 실선의 실측 값(Ground Truth)과 주황색 실선의 예측 값(Prediction) 차이를 계산하여, 모델의 손실(Loss)을 역전이(Back Propagation)한다. 그림의 예를 들어, 예측 값이 39.4이고 실측 값이 40.1인 경우 두 값의 차이인 0.5의 값이 손실로써 전이되어 가중치와 편향을 학습하도록 한다.

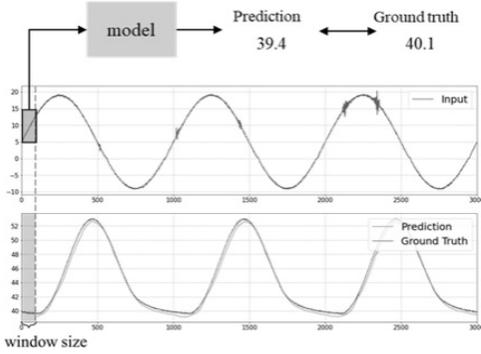


Fig. 11. Regression Model-based Real-time Prediction Process

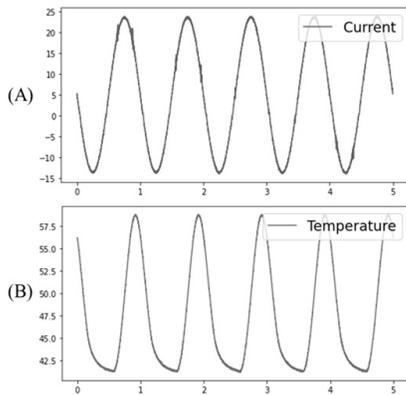


Fig. 12. Visualization of Current and Temperature Data in the Circuit System

### 5. 구현 및 평가

본 연구에서는 MMC 회로 시스템으로부터 주기성이 있는 센서 데이터(전류와 온도)를 수집하고, 제안하는 방법을 적용하여 회귀 모델 학습 실험을 수행하였다. 이 장에서는 슬라이딩 윈도우 기법을 적용하여, 윈도우 크기를 변경하며 학습 성능을 비교한 실험을 제시한다.

#### 5.1 실험 환경

본 연구에서는 회로 시스템의 전류 및 온도 데이터를 수집하여, 전류 기반의 온도 예측 회귀 모델을 개발하였다. Fig. 12는 실험에서 사용한 데이터를 나타낸다. (A)는 AC15, DC5인 전류 그래프이고, (B)는 진폭이 6.79인 온도 데이터를 나타낸다. 전류, 온도 데이터 모두 주기가 1초로 동일하고, 0.001초 단위로 측정되어 1,000개의 데이터가 한 주기가 된다.

전류-온도 데이터는 비선형 데이터로 Fig. 13과 같이 입력 대비 출력의 그래프에서 원형으로 표현된다. 이때, A, B는 동일한 전류에서 서로 다른 온도를 갖고, C, D는 동일한 온도에서 서로 다른 전류를 갖는다. 본 연구에서는 이와 같은 원시 데이터 셋으로부터 입력과 출력의 규칙적인 연관성을 확보할 수 있도록, 슬라이딩 윈도우를 적용하여 관계 분석을 수행하였다.

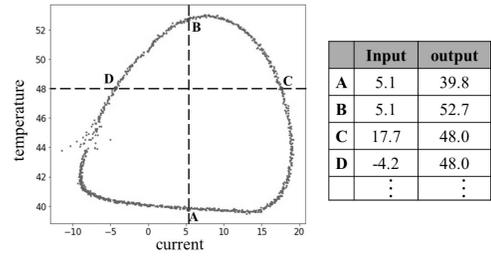


Fig. 13. Current-temperature Data Analysis

Table 1. Regression Model Design Conditions

hidden layer1	hidden layer2	hidden layer3
64	128	128

#### 5.2 학습 모델 설계

먼저, Table 1과 같이 세 개의 은닉층으로 구성된 ANN 기반의 회귀 모델을 설계하여 실험하였다. 슬라이딩 윈도우에 따라 전류-온도의 입출력 데이터 셋 관계를 파악하기 위해, 한 주기의 0.1%, 0.2%, 0.3%, 0.4%, 0.5%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%로 총 15가지의 입력 데이터 크기를 변경하여 데이터 셋을 구성하였다. 이와 같은 15개의 조건에서 구성된 데이터 셋에서 분석된 입출력 연관성은 Fig. 14와 같다. 그림은 전류-온도 데이터에 슬라이딩 윈도우를 적용하여 3가지 조건(window = 0.1%, 1%, 10%)의 실험 결과만 가시화한 그림이다. Fig. 14는 3차원 그래프의 X축은 윈도우의 그래디언트를 나타내고, Y축은 윈도우의 평균, 그리고 Z축은 온도를 나타낸다.

- (A) 0.1% of one cycle: 윈도우 크기가 1로 윈도우를 적용하지 않았다는 의미다. XY평면에서 동일한 값 x1에서 서로 다른 Z(온도) 값이 존재한다.
- (B) 1% of one cycle: 윈도우: 윈도우가 10으로 XY평면 위의 두 점 x1, x2가 유사한 값을 가지며, x1과 x2에서 서로 다른 Z(온도)와 대응된다.
- (C) 10% of one cycle: 윈도우가 100으로 XY평면 위의 서로 다른 두 점 x1, x2는 서로 다른 Z(온도)와 대응되고, x1, x2는 뚜렷하게 구분된다. 이때, XY평면의 동일한 점은 하나의 Z 값만을 갖게 된다.

이와 같이, 슬라이딩 윈도우로 데이터 셋을 구성할 경우, 입력에 대응되는 출력이 단 하나로 결정된다.

#### 5.3 모델 성능 평가

모델의 성능 평가를 위해, 학습 데이터와 검증 데이터는 400,000개의 데이터를 8:2의 비율로 나누어 실험하였다. 또한, 회귀 모델의 성능은  $R^2$ (R-square)[13]를 사용하여 평가하였다. Table 2는 윈도우 크기에 따른  $R^2$  성능과 입력층의 계산 복잡도를 수치적으로 나타낸 표이다. 계산 복잡도는 Equation (1)에서 나타내는 입력층의 연산 횟수를 의미하고, 첫 번째 은닉층의 유닛수와 윈도우의 크기가 증가할수록 계

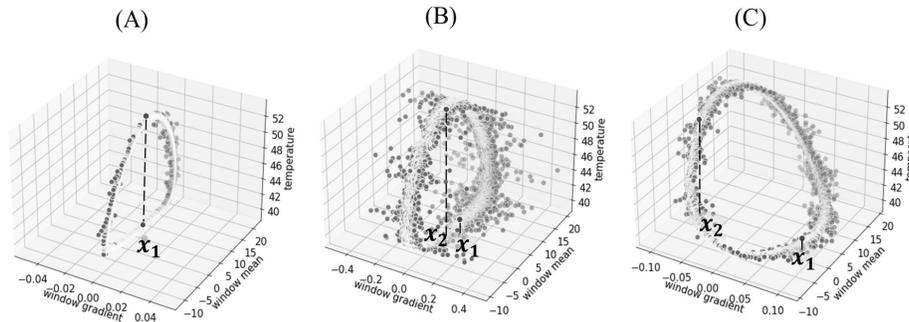


Fig. 14. Current-temperature Data Analysis using Sliding Window

Table 2. Performance Evaluation Based on Window Size

Window size	R <sup>2</sup> (R-square)	# of operaitons
0.1% of one cycle	-6.4050	192
0.2% of one cycle	-1.9432	320
0.3% of one cycle	-0.2264	448
0.4% of one cycle	0.3861	576
0.5% of one cycle	0.7871	704
1% of one cycle	0.9426	1344
<b>2% of one cycle</b>	<b>0.9706</b>	<b>2624</b>
3% of one cycle	0.9740	3904
4% of one cycle	0.9746	5184
5% of one cycle	0.9761	6464
6% of one cycle	0.9755	7744
7% of one cycle	0.9732	9024
8% of one cycle	0.9759	10304
9% of one cycle	0.9741	11584
10% of one cycle	0.9769	12864

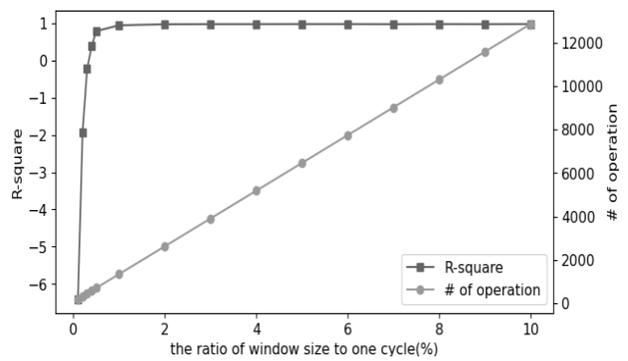


Fig. 15. Analysis of Regression Model Performance Change According to Window Size

산 복잡도는 선형적으로 증가한다. Fig. 15는 윈도우에 따른 R<sup>2</sup> 성능과 계산 복잡도 가시화하여 나타낸 그래프이다. Fig. 15의 파란색 선은 회귀 모델 학습을 통한 그래프를 나타낸다. 실험 결과, 윈도우의 크기가 주기의 10%로 R<sup>2</sup>가 0.9769로 가장 높은 성능을 보였다. 그러나, 계산 복잡도를 고려하여 윈도우를 설정한다면 주기의 2%에서 낮은 계산 복잡도와 0.97 이상의 성능을 갖는 것을 볼 수 있다.

### 6. 결론 및 향후 연구

본 논문에서는 주기성을 지닌 입출력 데이터에 초점을 맞추어, 입출력 관계를 분석하고, 슬라이딩 윈도우 기반의 입력 데이터 패턴화를 수행하여, 입출력 규칙성을 발견하는 방법을 제시하였다. 또한, 제안하는 방법을 검증하기 위해, 회로 시스템으로부터 동일한 주기를 지닌 전류, 온도 데이터를 수집하여 실험을 수행하였다. 제안하는 기법은 가장 높은 성능을 보장하는 윈도우 크기를 제시할 수 있고, 실험 결과 한 주기의 10%를 윈도우 크기로 적용하였을 때 성능이 0.9769로

가장 우수한 성능을 확보될 수 있음을 확인하였다. 그러나 계산 복잡도를 고려하여 한 주기의 2%로 윈도우의 크기를 설정하면 낮은 계산 복잡도에서 0.97 이상의 성능을 확보할 수 있다. 빠른 학습이 필요한 경우, 윈도우 크기를 2%로 설정할 수 있다. 이처럼, 제안하는 방법은 입출력 데이터를 단순히 일대일로 대응하는 경우, 입력과 출력 간의 규칙성이 결여되어 학습이 이루어지지 않는 문제를 해결하고, 개발자의 경험에 의존하여 설계되던 기존의 회귀 모델 학습 과정을 체계화하여 효과적인 학습 성능 확보를 지원한다.

### References

- [1] Y. Liu, X. Yan, C. Zhang, and W. Liu, "An ensemble convolutional neural networks for bearing fault diagnosis using multi-sensor data," *Sensors*, Vol.19, No.23, pp.5300, 2019.
- [2] Z. Li, R. Outbib, S. Giurgea, D. Hissel, A. Giraud, and P. Couderc, "Fault diagnosis for fuel cell systems: A data-driven approach using high-precise voltage sensors," *Renewable Energy*, Vol.135, pp.1435-1444, 2019.
- [3] S. S. Udmale, S. K. Singh, R. Singh, and A. K. Sangaiah, "Multi-fault bearing classification using sensors and ConvNet-based transfer learning approach," *IEEE Sensors Journal*, Vol.20, No.3, pp.1433-1444, 2019.

- [4] D. A. T. Tran, Y. Chen, and C. Jiang, "Comparative investigations on reference models for fault detection and diagnosis in centrifugal chiller systems," *Energy and Buildings*, Vol.133, pp.246-256, 2016.
- [5] M. S. Mirnaghi and F. Haghighat, "Fault detection and diagnosis of large-scale HVAC systems in buildings using data-driven methods: A comprehensive review," *Energy and Buildings*, Vol.229, pp.110492, 2020.
- [6] A. Capozzoli, F. Lauro, and I. Khan, "Fault detection analysis using data mining techniques for a cluster of smart office buildings," *Expert Systems with Applications*, Vol.42, No.9, pp.4324-4338, 2015.
- [7] J. A. Cortés-Ibáñez, S. González, J. J. Valle-Alonso, J. Luengo, S. García, and F. Herrera, "Preprocessing methodology for time series: an industrial world application case study," *Information Sciences*, Vol.514, pp.385-401, 2020.
- [8] H. Kaneko and K. Funatsu, "Fast optimization of hyper-parameters for support vector regression models with highly predictive ability," *Chemometrics and Intelligent Laboratory Systems*, Vol.142, pp.64-69, 2015.
- [9] A. Ahrens, C. B. Hansen, and M. E. Schaffer, "lassopack: Model selection and prediction with regularized regression in Stata," *The Stata Journal*, Vol.20, No.1, pp.176-235, 2020.
- [10] S. Selvin, R. Vinayakumar, E. A. Gopalakrishnan, V. K. Menon, & K. P. Soman, "Stock price prediction using LSTM, RNN and CNN-sliding window model," *2017 international conference on advances in computing, communications and informatics (icacci)*. IEEE, 2017.
- [11] H. S. Hota, R. Handa, and A. K. Shrivastava, "Time series data prediction using sliding window based RBF neural network," *International Journal of Computational Intelligence Research*, Vol.13, No.5, pp.1145-1156, 2017.
- [12] S. Park, S. Jung, S. Jung, S. Rho, and E. Hwang, "Sliding window-based LightGBM model for electric load forecasting using anomaly repair," *The Journal of Supercomputing*, Vol.77, No.11, pp.12857-12878, 2021.
- [13] O. Renaud and M. P. Victoria-Feser, "A robust coefficient of determination for regression," *Journal of Statistical Planning and Inference*, Vol.140, No.7, pp.1852-1862, 2010.



**김혜진**

<https://orcid.org/0000-0003-4113-2678>  
e-mail : khjin501@ajou.ac.kr  
2021년 아주대학교 전자공학과(학사)  
관심분야 : Bigdata Analysis, Sensor  
Data Anomaly Detection,  
Artificial Intelligence,  
Embedded Software



**박예슬**

<https://orcid.org/0000-0003-2584-7489>  
e-mail : yeseuly777@gmail.com  
2015년 아주대학교 전자공학과(학사)  
2017년 아주대학교 전자공학과(석사)  
2022년 아주대학교 AI융합네트워크학과  
(박사)

관심분야 : Bigdata Analysis, Machine Learning, Deep Learning, Collaborative Robot, Predictive Maintenance, Intelligent System



**이정원**

<https://orcid.org/0000-0001-8922-063X>  
e-mail : jungwony@ajou.ac.kr  
1993년 이화여자대학교 전자계산학과(학사)  
1995년 이화여자대학교 전자계산학과(석사)  
1995년 ~ 1997년 LG종합기술원  
주임연구원

2003년 이화여자대학교 컴퓨터학과(박사)  
2003년 ~ 2006년 이화여자대학교 컴퓨터학과 BK교수, 전임강사  
(대우)  
2006년 ~ 현재 아주대학교 전자공학과/AI융합네트워크학과  
교수

관심분야 : Context awareness, Big Data Analysis, Predictive Maintenance, Collaborative Robots, Intelligent Embedded Software