Journal of the Korea Industrial Information Systems Research Vol.25 No.6, Dec. 2020 :1-13
http://dx.doi.org/10.9723/jksiis.2020.25.6.001 ISSN:1229-3741

원전 계측 신호 오류 식별 알고리즘 개발+

(Development of Nuclear Power Plant Instrumentation Signal Faults Identification Algorithm)

김 승 근1)*
(SeungGeun Kim)

요 약 본 논문에서는 원전 비상 상황 발생 시 다수의 신호 오류가 발생했을 때 어떤 신호에 오류가 발생했는지를 추정하는 신호 오류 식별 (Fault identification) 방법론을 개발하였다. 변분 오토인코더 (Variational autoencoder; VAE) 기반 모델은 기존의 이상 탐지 방법론과 같이 정상 신호 데이터만을 이용하여 훈련이 진행되며, 이후 각 신호에 대한 복원 오차 (Reconstruction error)와 복원 오차를 입력의 특정 부분으로 미분한 값을 이용하여 어떤 부분에 오류가 포함되어 있는지를 예측한다. 데이터 취득을 위하여 시뮬레이션을 수행하였으며, 일련의 실험으로부터 제시한 신호 오류 식별 방법이 적절한 오차 범위 내에서 오류가 발생한 신호를 특정할 수 있음을 확인하였다.

핵심주제어: 원자력 발전소, 계측 신호, 신호 오류 식별, 변분 오토인코더, 설계기준사고

Abstract In this paper, the author proposed a nuclear power plant (NPP) instrumentation signal faults identification algorithm. A variational autoencoder (VAE)-based model is trained by using only normal dataset as same as existing anomaly detection method, and trained model predicts which signal within the entire signal set is anomalous. Classification of anomalous signals is performed based on the reconstruction error for each kind of signal and partial derivatives of reconstruction error with respect to the specific part of an input. Simulation was conducted to acquire the data for the experiments. Through the experiments, it was identified that the proposed signal fault identification method can specify the anomalous signals within acceptable range of error.

Keywords: Nuclear power plant, Instrumentation signals, Signal fault identification, Variational autoencoder, Design basis accident

1. 서 론

원전과 같이 규모가 크고 복잡한 시스템을 효율적이고 안전하게 운영하기 위해서는 다양한 계측 신호에 기반하여 시스템의 상태를 파악하고 그에 따른 적절한 결정을 내려야 한다. 특히, 극도의 안전성이 요구되는 안전 필수 시스템

* Corresponding Author: sgkim92@kaeri.re.kr
+ 본 연구는 산업통상자원부(MOTIE)와 한국에너지기술평가원(KETEP)의 지원을 받아 수행한 연구 과제임(No. 20171510102040)
Manuscript received September 28, 2020 / revised December 04, 2020 / accepted December 04, 2020
1) 한국원자력연구원 미래전략본부 지능형컴퓨팅연구실, 제1저자 및 교신저자

(Safety-critical system)이 극한 상황에 처한 경우에는 적절하지 못한 의사결정이 막대한 인명 및 경제적 피해를 초래할 수 있으므로, 의사 결정의 근거가 되는 계측 신호의 건전성이 더욱 중요하다. 후쿠시마 사고의 경우 지진으로 인하여 소외 전원이 상실됨과 동시에 쓰나미로 인해 비상 디젤발전기 (Emergency diesel generator; EDG) 마저 가용하지 않게 되면서 소내정전 (Station blackout; SBO) 상황이 발생하였으며, 그에 따라 대다수의 계측 신호가 소실되어 적절하지 못한 의사결정이 이루어진 바 있다 (Yang, 2014).

이에 따라, 원자력 분야에서는 계측 오류의 탐지와 보정 (Detection and calibration) (Hines et al., 1998; Fantoni, 2000; Nair and Coble, 2017), 미계측 변수에 대한 예측 (Prediction) (Na et al., 2008; Lim et al., 2010; Kim et al., 2015; No and Seong, 2016; No et al., 2018). 소실된 계측신호의 복원 (Reconstruction) (Shaheryar et al., 2016; Shaheryar et al., 2018; Kim et al., 2020) 등 계측 신호의 건전성을 확보하기 위한 다양한 연구가 진행되어왔다. 이러한 연구들 중에서도, 신호의 오류 포함 여부를 판별하는 신호 오류 탐지 (Signal fault detection) 기술과 해당 오류에 대한 구체적인 정보를 제공하는 신호 오류 식별 (Signal fault identification) 기술은 신호의 오류와 관련된 다양한 연구의 기반이 되므로, 가장 기초적이면서도 중요한 기술이라고 할 수 있다.

대부분의 기존 신호 오류 탐지 및 식별 방법론은 신호 사이의 상관관계에 강하게 의존하므로, 이러한 상관관계가 크게 변하지 않는 상황에 대해서만 성공적으로 작동할 수 있다. 이는 다양한 상황을 고려하기 위해서는 각 상황에 대한 모델을 따로 개발해야 하며, 적용을 위해서 상황에 대한 사전정보가 필요하다는 한계점을 유발한다. 신호 오류는 극한 상황에서 발생할 가능성이 높고, 원전 비상 상황은 원전에서 사용하는 절차서나 확률론적 안전성 분석 (Probabilistic safety analysis; PSA)에서 확인할 수 있듯 사고의 유형뿐만 아니라 사고의 위치, 심각도, 가용한 계통 및 기기 등에 따라 매우 광범위하게 시나리오가 분화될 수 있어, 기존의 방법론을 그대로 적용하기 어렵다. 또한 다수의 신호에 오류가 발생한 경우에는 상황에 대한 올바른 진단을 내리기가 어려워지는데, 이 때는 오류 탐지 및 식별을 위한 모델을 특정하기도 힘들어지기 때문에 위와 같은 한계점이 더욱 치명적으로 작용하게 된다.

본 연구에서는 이러한 한계점을 타개하기 위해, 생성 모델 (Generative model)의 일종인 변분 오토인코더 (Variational autoencoder; VAE) (Kingma and Welling, 2014)에 기반하여 원전 비상 상황에서 다수의 계측 신호에 오류가 발생한 경우에도 이를 탐지하고 식별해내는 방법론을 개발하고자 하였다. VAE는 이미 다양한 분야에서 이상 탐지 (Anomaly detection)를 위해 적용된 바 있으나 이상 식별 (Anomaly identification)에 적용된 사례는 거의 없다. VAE를 오류가 발생한 신호를 식별하는데 적용하기 위하여 기존의 이상 탐지 방법에 추가적으로 각 신호별로 복원 오차 (Reconstruction error)를 비교하는 방법과 변화 분석 (Gradient analysis)을 도입하였다. 제시한 방법론은 상황에 대한 사전 정보 없이도 오류가 발생한 다수의 계측 신호를 탐지하고 식별할 수 있으며, 하나의 모델로 다양한 원전 비상 상황을 고려할 수 있다.

본 연구는 2장에서 VAE의 개념과 그에 기반한 신호 오류 탐지 방법론에 대해 서술하며, 3장에서는 제시한 VAE 기반 신호 오류 및 식별 방법론에 대해 서술하였다. 4장에서는 제시한 방법론의 성능 검증을 위한 일련의 실험 과정을 기술하였으며, 마지막으로 5장에서는 본 연구의 결론을 정리하였다.

2. 배 경

2.1 오토인코더 기반 이상 탐지

오토인코더 (Autoencoder)는 인공신경망 (Artificial neural network) 구조의 일종으로 비지도학습 (Unsupervised learning)에 널리 사용되며, 다른 모델의 일부분으로도 자주 도입된다.

오토인코더는 서로 연결되어있는 전단의 인코

더 (Encoder)와 후단의 디코더 (Decoder)로 구성된다. 최초 입력을 받는 인코더는 이전 층(Layer)에 비해 이후의 층이 더 적은 수의 노드(Node)를 갖는, 전체적으로 점점 좁아지는 형태로 이루어진다. 반대로 최종 출력을 내는 디코더는 이전 층에 비해 이후의 층이 더 많은 수의 노드를 갖는, 전체적으로 점점 넓어지는 형태로 되어있다. Fig. 1은 오토인코더의 일반적인 구조를 도식화한 것이다.

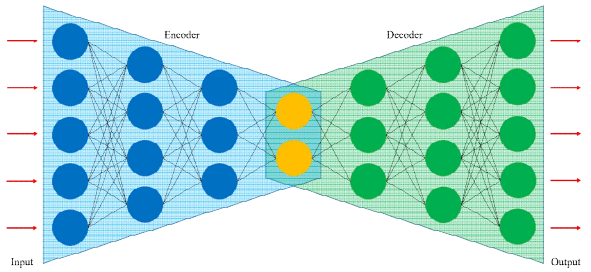

Fig. 1 Schematic of an Autoencoder

오토인코더는 일반적으로 입력과 출력이 최대한 같아지도록 훈련이 진행된다. 이 과정에서 인코더는 입력 데이터를 압축하여 더 적은 수의 매개변수 (Parameter)로 표현하는 차원 축소(Dimensional reduction)를 수행하게 되며, 디코더는 이렇게 압축된 데이터를 다시 기존의 입력 데이터로 복원하는 역할을 수행하게 된다. 인코더와 디코더 사이에 있는 층이 전체 모델에서 가장 적은 수의 노드를 가지므로, 해당 부분에서 입력 데이터의 가장 압축된 표현을 추출할 수 있다. 이 때 입력과 출력이 최대한 같아지도록 훈련이 진행되었으므로 차원 축소 과정에서 최소한의 정보 손실만이 발생하도록 훈련이 진행된다.

오토인코더는 이러한 특성으로 인하여 데이터의 특성을 추출 (Feature extraction)하거나 데이터의 잡음을 제거 (Denoising)하는데 널리 적용되었으며, 이상 탐지에도 적용되었다 (Kim and Hong, 2017; Kim et al., 2019; Minar et al., 2020).

오토인코더 기반 이상 탐지 과정에서는 입력과 출력 사이의 오차인 복원 오차를 이상 탐지의 기준으로써 이용한다. 이 때 오토인코더 모델은 이상이 없는 데이터로만 훈련이 진행되며, 따라서 이상이 없는 데이터에 대해서 복원 오차가 최소화되는 방향으로 훈련이 진행된다. 이렇게 훈련된 모델에 이상이 있는 데이터가 입력으로 주어질 경우 이상이 없는 데이터에 비하여 복원 오차가 크게 나타날 것이라는 가정이 오토인코더 기반 이상 탐지의 핵심적인 부분이며, 적절한 문턱값 (Threshold)을 설정함으로써 정상 데이터와 이상 데이터를 분류하게 된다. Algorithm 1은 오토인코더 기반 이상 탐지 알고리즘을 나타낸 것이다.

Algorithm 1. Autoencoder-based Anomaly Detection Algorithm

Input: Normal dataset X_n, i-th anomalous data $x_{a,i} \in X_a (i = 1, \ldots, I)$, reconstruction error threshold α

Output: Anomaly detection (classification) results for all anomalous data

$f_{enc}, f_{dec} \leftarrow$ Train an autoencoder using X_n

for $i = 1$ to I, **do**

  $E_i = [f_{dec}(f_{enc}(x_{a,i})) - x_{a,i}]$

  if $E_i \geq \alpha$

    $x_{a,i}$ is an anomaly

  else

    $x_{a,i}$ is not an anomaly

  end if

end for

f_{enc}: encoder function of autoencoder

f_{dec}: decoder function of autoencoder

E_i: reconstruction error for i-th anomalous data

2.2 변분 오토인코더 기반 이상 탐지

오토인코더는 단순한 구조에 비해 좋은 성능을 보여주기는 하지만 과적합 (Overfitting) 문제에 다소 취약하다. 또한 같은 입력에 대하여 항상 같은 출력을 도출하는 결정론적(Deterministic) 모델이라는 점에서, 오토인코더 기반 이상 탐지 방법은 입력이 얼마나 정상 데이터 또는 이상 데이터에 가까운지를 확인하기

가 힘들다는 단점을 가지고 있다.

변분 오토인코더 (Variational autoencoder; VAE) (Kingma and Welling, 2014)는 오토인코더의 일종으로, 오토인코더와 같이 인코더와 디코더로 구성된다. 하지만 VAE의 인코더는 디코더에 대한 입력을 직접 출력하는 대신 그 입력에 대한 확률 분포의 매개변수를 출력하고, 이때 일반적으로 정규분포 (Normal distribution, Gaussian distribution)를 확률 모델로 사용하여 평균과 분산에 대한 값을 출력한다. 디코더의 입력은 해당 확률분포로부터 샘플링을 통하여 도출되며, 이 때문에 같은 입력에 대하여 항상 같은 출력을 도출하는 결정론적인 특성을 갖는 일반적인 오토인코더와는 달리 VAE는 확률론적 (Probabilistic)인 특성을 갖게 된다. 이는 VAE가 기존의 오토인코더처럼 차원 축소를 위한 모델로써 활용될 수 있을 뿐만 아니라, 입력 데이터의 분포에 속하는 새로운 데이터를 생성할 수 있는 생성 모델 (Generative model)로도 활용될 수 있도록 한다. 실제로 VAE는 생성적 적대 신경망 (Generative adversarial network; GAN) (Goodfellow et al., 2014)과 함께 대표적인 생성 모델로써 다양한 분야에 적용되고 있다. 이 외에도 VAE를 기존의 오토인코더와 같은 목적으로 적용할 경우 오토인코더보다 훈련 데이터에 대한 일반화 (Generalization)을 더 잘 수행하며, 과적합 문제에 보다 강건 (Robust)하다는 장점이 밝혀진 바 있다.

VAE는 이러한 특성으로 인하여 오토인코더의 적용 분야에 더하여, 새로운 데이터를 생성하여 부족한 양의 데이터를 보충하기 위한 데이터 증강 (Data augmentation) 등의 여러 생성 모델 응용 분야에 적용되고 있다. Fig. 2는 VAE의 일반적인 구조를 도식화한 것이다.

VAE 기반 이상 탐지 방법 (An and Cho, 2015) 역시 오토인코더 기반 이상 탐지 방법과 유사하다. 가장 핵심적인 차이점은 기존의 오토인코더 기반 이상 탐지 방법은 정상 데이터와 이상 데이터의 분류를 결정론적인 관점으로 접근하지만, VAE 기반 이상 탐지 방법은 이를 확률론적으로 접근한다는 것이다. 이는 정상 데이터와 이상 데이터를 구분하는 기준을 더 정교하게 세울 수 있도록 할 뿐만 아니라, 어떤 데이터가 얼마나 정상 데이터 또는 이상 데이터에 가까운지를 확인할 수 있도록 한다는 장점을 부여한다 (여전히 오토인코더 기반 이상 탐지 방법에서처럼 적절한 문턱값을 설정함으로써 정상 데이터와 이상 데이터를 분류하게 된다). Algorithm 2는 VAE 기반 이상 탐지 방법의 알고리즘을 나타낸 것이다.

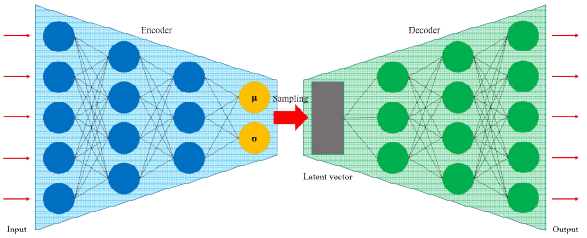

Fig. 2 Schematic of a VAE

Algorithm 2. VAE-based Anomaly Detection Algorithm

Input: Normal dataset X_n, i-th anomalous data $x_{a,i} \in X_a (i = 1, \dots, I)$, reconstruction probability threshold α, number of sampling K

Output: Anomaly detection (classification) results for all anomalous data

$f_{enc}, f_{dec} \leftarrow$ Train a VAE using X_n

for $i = 1$ to I, **do**

$\mu_{z,i}, \sigma_{z,i} = f_{enc}(z|x_{a,i})$

for $k = 1$ to K, **do**

draw sample of $z_{ik} \sim N(\mu_{z,i}, \sigma_{z,i})$

$\mu_{x,ik}, \sigma_{x,ik} = f_{dec}(x|z_{ik})$

end for

$$P_i^{rec} = \frac{1}{K}\sum_{k=1}^{K} p(x_{a,i}|\mu_{x,ik}, \sigma_{x,ik})$$

if $P_i^{rec} \leq \alpha$

$x_{a,i}$ is an anomaly

else

$x_{a,i}$ is not an anomaly

end if

end for

f_{enc}: encoder function of VAE
f_{dec}: decoder function of VAE
$\mu_{z,i}$: mean vector for i-th anomalous data(output of encoder)
$\sigma_{z,i}$: standard deviation vector for i-th anomalous data(output of encoder)
z_{ik}: latent vector for k-th sample of i-th anomalous data
$N(\mu, \sigma)$: Normal distribution with mean value μ and standard deviation value σ
$\mu_{x,ik}$: mean vector for k-th sample of i-th anomalous data(output of decoder)
$\sigma_{x,ik}$: standard deviation vector for k-th sample of i-th anomalous data(output of decoder)
P_i^{rec}: reconstruction probability for i-th anomalous data

3. 변분 오토인코더 기반 신호 오류 식별 방법론

본 연구에서 제시한 VAE 기반 신호 오류 식별 방법론은 기존의 VAE 기반 신호 오류 탐지 방법론을 확장한 것으로, 신호 오류를 식별하기 전 탐지하는 과정까지는 동일하다. 신호 오류를 식별하는 과정은 모델 사전훈련 (Pre-training), 신호 오류 탐지, 신호 오류 식별의 세 단계로 이루어진다.

3.1 모델 사전훈련

모델 사전훈련 단계는 VAE 모델을 다양한 상황 하에서 관찰될 수 있는 계측 신호에 대하여 적은 오차로 복원을 수행할 수 있도록 훈련시키는 단계이다. 이 과정을 통해 VAE 모델은 신호 오류가 포함되지 않은 데이터에 대하여 모사할 수 있는 능력을 갖추게 되며, 사전 훈련이 완료되면 모델은 고정되어 이후 단계에 적용된다.

사전 훈련된 모델이 얼마나 정확하게 입력을 복원할 수 있는지에 따라 오류 식별 과정에서의 정밀도가 달라지게 되며, 모델의 복원 정확도가 높아질수록 더욱 경미한 오류에 대해서도 오류로 판별할 수 있게 된다. 따라서 요구되는 오류 식별의 정밀도에 따라 모델의 성능을 조절하여야 한다. 이 때 모델의 층 또는 노드 수를 증가시키거나 보다 장기간 훈련을 진행하여 모델을 과적합시키는 방법 등을 고려할 수 있다.

3.2 신호 오류 탐지

신호 오류 식별 과정은 데이터에 오류가 존재함을 확인한 후에만 의미가 있으므로 신호 오류 탐지 과정이 선행되어야 한다. 기존의 VAE 기반 신호 오류 탐지 방법론과 같이 정상 데이터에 기반하여 훈련된 모델은 이후 오류가 포함된 데이터를 입력으로 받은 경우 더 적은 복원 확률을 보이게 되며, 적절한 복원 확률값을 기준으로 하여 정상 데이터와 오류가 포함된 데이터를 구분할 수 있다.

3.3 신호 오류 식별

VAE 기반 신호 오류 탐지 방법론은 입력과 출력 사이의 복원 오차를 거시적으로 확인하여 오류가 존재하는지의 여부를 판별한다. 본 연구에서 제시한 신호 오류 식별 방법론은 입력과 출력 사이의 복원 오차를 미시적으로도 확인하여 데이터 내 오류의 존재 여부뿐만 아니라 입력의 어느 부분이 큰 복원 오차를 유발했는가를 밝혀낸다 (신호 탐지 방법론과는 달리 복원 확률을 이용하지 않고 복원 오차를 이용한다). 이는 입력의 특정 부분이 큰 복원 오차를 유발했다면, 해당 부분에 오류가 존재할 가능성이 높다는 가설에 기반한다. 이러한 개념은 단순하고 직관적이지만, 다음 두 가지 문제에 대하여 고려하여야 한다.

첫 번째는 가설 자체의 타당성과 관련된 문제로써, 모델의 입력 노드 간에는 상호작용이 없으나 각각의 노드는 그 이전 층의 모든 노드로부터 영향을 받으므로, 입력의 일부분에만 오류가 포함되어 있더라도 그 영향이 출력의 여러 노드에 골고루 영향을 미칠 수 있는 가능성이

존재한다는 것이다. 이러한 경우 복원 오차를 입력의 특정 부분으로 미분하더라도 그 값이 큰 차이를 보이지 않아 위와 같은 가설에 기반하여 신호 오류 식별을 수행할 수 없기 때문에, 4장에서 소개할 일련의 실험과정을 통하여 위와 같은 가설이 유효한가를 검증하였다.

두 번째는 특성이 상이한 여러 신호를 데이터에 포함시킬 경우, 모델이 복원을 수행하는 과정에서 각 신호에 대한 복원 성능이 다르게 나타난다는 점이다. 신호 오류를 탐지하는 과정에서는 거시적인 관점에서 복원 확률을 계산하므로 이러한 신호간의 차이가 고려되지 않아도 되지만, 신호 오류를 식별하는 과정에서는 이러한 모델의 성능 차이가 결과를 왜곡시킬 수 있다. 이에 입각하여, 본 연구에서는 VAE 모델을 사전 훈련하는 과정에서 기존의 모든 신호와 시간에 대한 평균 복원오차 뿐만 아니라 각 신호에 대한 평균 복원오차를 따로 계산하여, 이를 신호 오류 식별을 위한 판정 기준에 반영하였다.

상기한 두 가지 문제를 고려하여, 최종적으로 각 신호의 오류 포함 여부를 판별하는 기준은 각 신호에 대한 평균 복원 오차와 각 신호에 대한 복원 오차를 입력의 특정 부분으로 편미분한 값을 모두 포함한 형태로 제시하였다. 각 신호의 오류 포함 여부를 판별하는 기준을 수식으로 나타내면 식(1)과 같다.

$$C_j = W_1\overline{E_{n,j}} + W_2\nabla\overline{E_{n,j}} \tag{1}$$

식 (1) 에서 j는 j번째 신호를 의미하며, C_j는 j번째 신호에 대한 판별 기준, $\overline{E_{n,j}}$는 정상 신호의 j번째 신호에 대한 평균 복원 오차, $\nabla\overline{E_{n,j}}$는 정상 신호에 대한 복원 오차를 입력의 j번째 신호로 편미분한 값의 평균이다. W_1과 W_2는 가중치로써, 적절한 가중치의 값은 모델 훈련 과정에서의 초매개변수 (Hyperparameter)와 같이 실험적으로 탐색되어야 한다. Algorithm 3은 VAE 기반 신호 오류 식별 알고리즘을 나타낸 것이다.

Algorithm 3. VAE-based Signal Faults Identification Algorithm

Input: Normal dataset X_n, i-th anomalous data $x_{a,i} \in X_a\,(i = 1, \dots, I)$, j-th signal of i-th anomalous data $x_{a,i}^j\,(j = 1, \dots, J)$, number of sampling K, weighting factors W_1, W_2,

Output: Anomaly identification results for all anomalous data and signals

$f_{enc}, f_{dec} \leftarrow$ Train a VAE using X_n

$\overline{E_{n,j}} \leftarrow$ mean reconstruction error of j-th signal on normal dataset

$\nabla\overline{E_{n,j}} \leftarrow$ mean gradient of reconstruction error w.r.t. j-th signal on normal dataset

$C_j = W_1\overline{E_{n,j}} + W_2\nabla\overline{E_{n,j}}$

for $i = 1$ to I, **do**

$\mu_{z,i}, \sigma_{z,i} = f_{enc}(z|x_{a,i})$

for $k = 1$ to K, **do**

draw sample of $z_{ik} \sim N(\mu_{z,i}, \sigma_{z,i})$

$E_{a,i,k} = f_{dec}(z_{ik}) - x_{a,i}$

end for

$E_{a,i} = \frac{1}{K}\sum_{k=1}^{K} E_{a,i,k}$

for $j = 1$ to J, **do**

retrieve $E_{a,i}^j \in E_{a,i}$

$A_{i,j} = E_{a,i}^j + \frac{\partial E_{a,i}}{\partial x_{a,i}^j}$

if $A_{i,j} \geq C_j$

$x_{a,i}^j$ is faulty

else

$x_{a,i}^j$ is not faulty

end if

end for

end for

f_{enc}: encoder function of VAE
f_{dec}: decoder function of VAE
C_j: anomaly criteria for j-th signal
$\mu_{z,i}$: mean vector for i-th anomalous data
$\sigma_{z,i}$: standard deviation vector for i-th anomalous data
z_{ik}: latent vector for k-th sample of i-th anomalous data
$N(\mu, \sigma)$: normal distribution with mean value μ and standard deviation value σ
$E_{a,i,k}$: reconstruction error for k-th sample of i-th anomalous data
$E_{a,i}$: mean reconstruction error for i-th anomalous data
$E_{a,i}^{j}$: mean reconstruction error for j-th signal of i-th anomalous data
$A_{i,j}$: anomaly score for j-th signal of i-th anomalous data

4. 실 험

4.1 데이터 취득 및 전처리

원전에 비상 상황이 발생한 사례는 극히 드물기 때문에, 실험은 시뮬레이션으로부터 취득된 데이터를 이용하여 수행되었다. 시뮬레이터는 한국원자력연구원에서 개발한 CNS (Compact nuclear simulator)를 이용하였다 (Korea Atomic Energy Research Institute, 1990). CNS가 모사한 원전은 웨스팅하우스 (Westinghouse) 사의 3-loop 900MW 가압경수로 (Pressurized water reactor; PWR)로, 국내에 있는 원전 중 고리 3, 4호기와 영광 1, 2호기가 이에 해당한다.

원전 비상 상황으로는 설계 기준 사고 (Design basis accident; DBA) 중 저온관 및 고온관 파단 냉각재 상실 사고 (Cold-leg/hot-leg loss-of-coolant accident, cold-leg/hot-leg LOCA), 증기발생기 전열관 파단 (Steam generator tube rupture; SGTR), 주증기관 파열 (Main steam line break; MSLB)의 네 가지를 고려하였다.

각 상황에 대하여 10㎠부터 100㎠까지 1㎠ 간격으로 파단 크기를 변화시키며 시뮬레이션을 수행하였다. MSLB의 경우 동일한 파단 크기에서 상대적으로 긴 시간이 소요된 후에 원자로 정지가 발생하여, 시뮬레이션의 효율성을 높이기 위해 10배의 파단크기를 적용하였다. 취득한 계측 신호의 종류로는 원전 비상 상황 발생 시 의사결정을 내리는 데 중요한 역할을 하는 31가지의 계측 신호를 선정하였다 (Table 1. 참고). 시뮬레이션은 원자로 긴급 정지로부터 5분간 수행하였다.

Table 1 Collected Signals and Their Units

Signals	Units
CTMT* sump level	m
CTMT radiation	mrem/h
CTMT relative humidity	%
CTMT temperature	℃
CTMT pressure	kg/㎠
Core outlet temperature	℃
Hot-leg temperature (loop 1/2/3)	℃
Cold-leg temperature (loop 1/2/3)	℃
Delta temperature (loop 1/2/3)	℃
PRT* temperature	℃
PRT pressure	kg/㎠
Hydrogen concentration	%
Reactor vessel water level	%
Pressurizer temperature	℃
Pressurizer level	%
Pressurizer pressure (wide range)	kg/㎠
S/G* pressure (loop 1/2/3)	kg/㎠
S/G level (narrow range, loop 1/2/3)	%
Feedwater flow rate (loop 1/2/3)	ton/h

* CTMT: containment
* PRT: pressurizer relief tank
* S/G: steam generator

시뮬레이션 후에는 각 신호에 대하여 최소최대정규화 (Min-max normalization)를 적용하여 모든 신호가 0과 1 사이의 값을 가지도록 하였

다. 이 때, 이후 오류 주입 과정에서 0과 1 사이의 값을 벗어날 수 있으므로 최댓값과 최솟값의 차이의 5%에 해당하는 여유 구간을 두고 정규화를 적용하였다. 또한 선형 내삽 (Linear interpolation)을 통해 데이터의 시간 간격을 1초로 균일하도록 하였으며, 5분간의 데이터를 30초 길이를 갖는 단위데이터로 가공하여 총 98,280개의 단위 데이터를 생성하였다.

4.2 모델 사전훈련

모델의 기본 구조로서, 가장 일반적인 형태의 VAE를 이용하였다. 확률 모델로는 정규분포를 이용하였으며, 인코더와 디코더는 각각 입력 층과 출력 층을 포함하여 6개의 Fully-connected 층으로 구성되어 있다. Fig. 3은 적용된 VAE 모델을 나타낸 것이다.

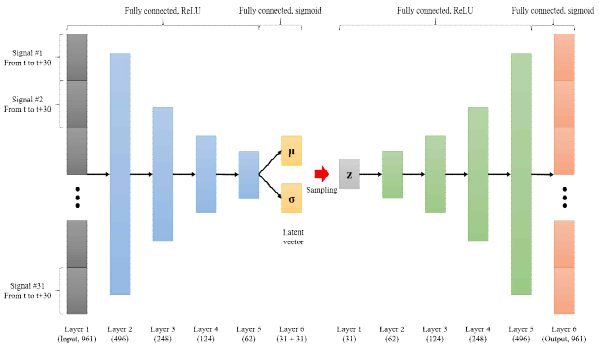

Fig. 3 Schematic of an Applied VAE Model with the Activation Functions and the Number of Nodes

훈련 데이터로는 전체 데이터를 모두 이용하였으며 (이후 실험단계에서 인위적으로 오류를 주입한 데이터를 이용하므로 전체 데이터를 이용하였다), 모델 사전 훈련 단계는 노드 개수와 배치 크기 (Batch size)등의 초매개변수를 변화시켜가면서 반복적으로 수행하였다. 훈련 과정에서 이용되는 손실 함수의 두 요소인 복원 오차와 KL (Kullback-Leibler)-divergence 사이의 가중치 β는 두 요소 사이의 값이 큰 차이를 보이지 않음에 따라 1로 고정하였다. 샘플링 과정에서는 1,000개의 샘플을 추출하도록 하였으며, 최적화 방법으로는 Adam (Kingma and Ba, 2014) 을 사용하였다. Algorithm 4는 VAE 모델 훈련 과정을 나타낸 것이다.

Algorithm 4. VAE Model Training

Input: Normal data $x_{n,i} \in X_n \, (i = 1, \ldots, I)$, number of sampling K, weighting factor β

Output: Encoder function f_{enc}, decoder function f_{dec}

Initialize parameters of f_{enc} and f_{dec}

repeat

for $i = 1$ to I, **do**

$\mu_i, \sigma_i \leftarrow f_{enc}(x_{n,i})$

for $k = 1$ to K, **do**

draw sample of $\epsilon_{i,k} \sim N(0,1)$

$z_{i,k} = \mu_i + \sigma_i \epsilon_{i,k}$

end for

$$L_{recon} = \frac{1}{IK} \sum_{i=1}^{I} \sum_{k=1}^{K} (f_{dec}(z_{i,k}) - x_{n,i})^2$$

$$L_{KL} = \frac{1}{2} \sum_{i=1}^{I} (\sigma_i^2 + \mu_i^2 - \ln(\sigma_i^2) - 1)$$

$$L = L_{recon} + \beta L_{KL}$$

end for

Update parameters of f_{enc} and f_{dec} using gradients of L

until convergence of parameters of f_{enc} and f_{dec}

μ_i: mean vector for i-th data

σ_i: standard deviation vector for i-th data

$\epsilon_{i,k}$: sampled number

$N(\mu, \sigma)$: normal distribution with mean value μ and standard deviation value σ

$z_{i,k}$: latent vector for i-th data

L_{recon}: reconstruction error loss

L_{KL}: Kullback-Leibler divergence loss

L: total loss

이후 신호 오류 탐지 및 식별 단계에서는 다양한 모델 중 가장 적은 복원 오차를 보이는 모델을 최종적으로 선정하여 적용하였다. 최종 선정된 모델은 약 0.87%의 평균 복원 오차로 정상 데이터를 복원할 수 있었다.

4.3 신호 오류 탐지 및 식별

모델 사전훈련 단계에서 선정된 모델은 성능 확인을 위해 신호 오류 탐지 및 식별에 적용되었다. 원전 비상 상황을 모사할 수 있는 대부분의 시뮬레이터가 신호의 오류를 모사할 수 없으므로, 오류가 포함된 신호를 생성하기 위해 취득된 데이터의 인위적으로 오류를 주입하는 방식으로 실험을 진행하였다.

전체 단위 데이터에서 임의로 2,000개를 선정하였고, 그 중 1,000개의 단위 데이터에 인위적인 오류를 주입하였다. 오류가 주입된 신호의 개수는 3개부터 10개까지 증가시키면서 실험을 진행하였으며, 이 때 31가지의 계측 신호 중 임의의 계측 신호를 선정하여 오류를 주입하였다.

오류의 종류로는 Gaussian white noise, step error, ramp error를 고려하였으며, 주입한 오류의 정도는 모델의 정상 데이터에 대한 복원 오차가 대부분 1.5% 이내였다는 점에 착안, Gaussian white noise의 경우 평균 0, 표준편차 0.02, step error의 경우 ±0.02, ramp error의 경우 ±0.001/sec의 신호 오류를 주입하도록 하였다. 데이터 전처리 과정에서 모든 신호가 0과 1 사이의 값을 가지도록 정규화를 수행하였으므로 step error는 2%의 오류가 갑자기 발생한 경우를, ramp error는 오류의 정도가 초당 0.1%씩 점진적으로 증가하는 경우를 고려한 것이다 (단위 데이터의 길이가 30초이므로, 단위 데이터의 끝 부분에서 3%의 오류가 주입되었다).

신호 오류 탐지는 연구의 본 목적이 아니므로, 모델의 유효성 검증을 위해 비교적 간단하게 진행되었다. 정상 단위 데이터와 여러 종류의 오류가 주입된 단위 데이터를 선정하여 신호 오류 탐지 과정에 적용하였으며, 복원 확률 문턱값을 90% 또는 95%로 설정하였다. 그 결과, 복원 확률 문턱값을 95%로 설정했을 때, 약 98.1%의 정확도로 오류가 포함된 단위 데이터를 판별하여, VAE 모델이 신호 오류 탐지를 적절히 수행할 수 있음을 확인하였다. Fig. 4는 정상 신호와 오류가 주입된 신호의 예시를, Fig. 5는 정상 신호와 VAE 모델이 해당 데이터를 복원한 예시를 나타내고 있다.

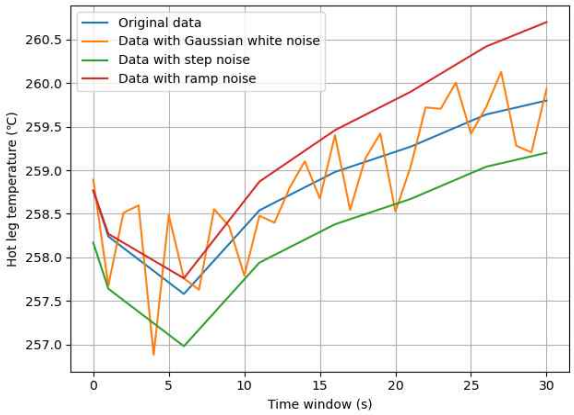

Fig. 4 Example of Original Data and Noisy Data (Hot Leg Temperature)

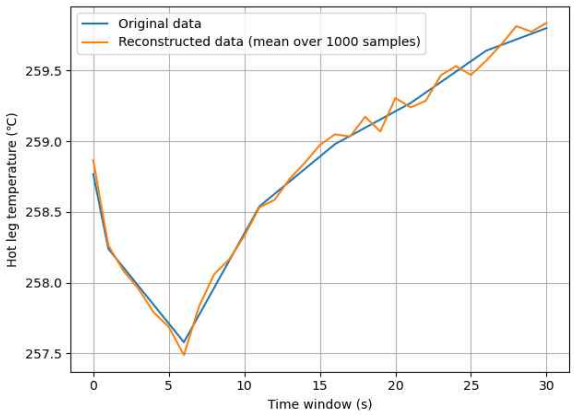

Fig. 5 Example of Original Data and Reconstructed Data (Hot Leg Temperature)

신호 오류를 식별하는 과정에서는 판별 기준식에 포함된 가중치 W_1와 W_2의 값을 격자 탐색 (Grid search) 방법에 따라 변경하며 실험을 반복하였다. W_1의 값을 0으로 설정할 경우, 다양한 W_2 값을 적용하더라도 적절한 오류 신호

식별이 불가능했으며, W_2의 값을 0으로 설정한 경우에는 일부 W_1값에 대하여 오류 신호 식별이 가능하였으나 최적의 성능을 보이지는 않았다. 최종적으로 가중치 W_1와 W_2의 값을 각각 1.5와 1,000 으로 설정했을 때 최적의 성능을 보였으며 (미분값이 복원 오차의 값보다 매우 작아 가중치가 큰 차이를 보인다), 이로써 최초 가설대로 미분값 만으로는 오류 식별을 수행할 수 없음을 확인하였으나 미분값이 오류 식별에 있어 보조적인 역할을 수행할 수 있음을 확인하였다. Table 2 와 Fig. 6, Fig. 7는 신호 오류 식별의 결과를 나타내고 있다.

성능 지표로는 평균 정확도 (Mean accuracy), 평균 정밀도 (Mean precision), 평균 재현율 (Mean recall)을 이용하였다. True positive (TP)는 오류가 포함된 신호를 오류가 포함된 신호로 올바르게 식별한 경우, True negative (TN)는 오류가 포함되지 않은 신호를 오류가 포함되지 않은 신호로 올바르게 식별한 경우, False positive (FP)는 오류가 포함되지 않은 신호를 오류가 포함된 신호로 틀리게 식별한 경우, False negative (FN)는 오류가 포함된 신호를 오류가 포함되지 않은 신호로 틀리게 식별한 경우를 나타낸다. 평균 정확도, 평균 정밀도, 평균 재현율은 식 (2)-(4)와 같은 식으로 계산된다.

$$mean\,accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{2}$$

$$mean\,precision = \frac{TP}{TP+FP} \tag{3}$$

$$mean\,recall = \frac{TP}{TP+FN} \tag{4}$$

오류가 포함된 신호가 오류가 포함되지 않은 신호에 비하여 적으므로, 평균 정확도와 평균 정밀도로는 성능을 온전히 평가하기 힘들다. 평균 정확도의 경우, 오류가 포함된 신호가 매우 적을 때 모델이 모든 신호를 오류가 포함되지 않았다고 식별을 하더라도 높은 값을 보일 수 있다는 문제가 있다. 평균 정밀도의 경우, 오류가 포함된 신호보다 오류가 포함되지 않은 신호의 수가 많아 True positive 비율에 비해 False negative 비율이 지나치게 높아져 실제 모델의 성능을 왜곡시킬 수 있다. 이에 반해 평균 재현율은 실제로 오류가 포함된 신호의 수와 오류가 포함되었다고 식별된 신호의 수 사이의 비율을 나타내므로, 오류가 포함된 신호가 그렇지 않은 신호에 비하여 적은 것이 문제되지 않아 성능 평가에 적합하다. 다만, 평균 재현율 단독으로는 오류가 포함되지 않은 신호를 어떻게 식별했는가에 대해서 알 수 없으므로 평균 정확도와 동시에 확인하는 것이 바람직하다.

실험 결과, 제시한 모델은 전체 31가지의 계측 신호 중 3가지 신호에 오류가 포함된 경우

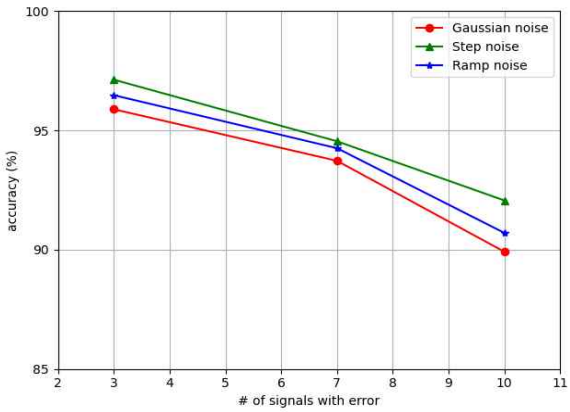

Fig. 6 Signal Faults Identification Accuracies according to Number of Noisy Signals

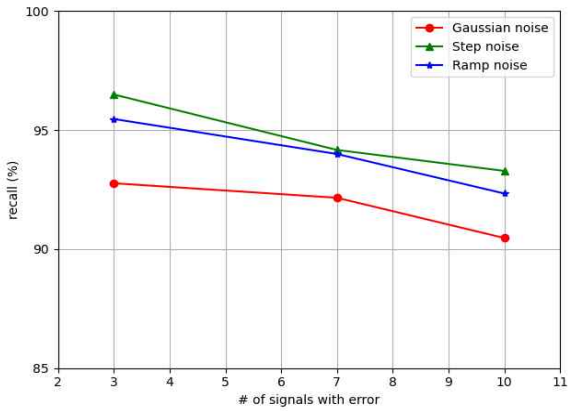

Fig. 7 Signal Faults Identification Recalls according to Number of Noisy Signals

Table 2 Experiment Results
($W_1 = 1.5, W_2 = 1000$)

Noise type	# of noisy signals	Mean accuracy	Mean precision	Mean recall
Gaussian (μ=0, σ=0.02)	3	0.9589	0.7247	0.9277
	7	0.9373	0.8222	0.9216
	10	0.8992	0.8064	0.9047
Step (±0.02)	3	0.9713	0.7858	0.9650
	7	0.9455	0.8073	0.9417
	10	0.9207	0.8392	0.9329
Ramp (±0.001/s)	3	0.9648	0.7501	0.9547
	7	0.9426	0.8288	0.9400
	10	0.9071	0.8137	0.9234

Gaussian white noise의 경우 95.9%, Step noise의 경우 97.1%, Ramp noise의 경우 96.5%의 정확도로 오류 신호를 식별할 수 있었다. 평균 재현율 또한 각각 92.8%, 96.5%, 95.5%로 높은 성능을 보였다. 7가지와 10가지 신호에 오류가 포함된 경우에는 식별 정확도와 재현율이 하락했으나, 10가지 신호에 오류가 포함된 경우에도 90% 전후의 식별 정확도와 재현율을 보였다. 이는 평균적으로 오류가 포함된 10가지 신호 중 한 가지를 제외한 나머지 신호들을 식별해낼 수 있음을 의미한다.

5. 결 론

본 연구에서는, 원전 비상 상황에서 다수의 계측 신호에 오류가 발생한 경우에 대비하기 위하여 VAE 기반 신호 오류 식별 방법론을 개발하였다. 개발을 위해 기존의 VAE 기반 신호 오류 탐지 방법론을 확장하였으며, 오류 식별의 기준이 되는 문턱값의 결정 방법을 제시하였다.

일련의 실험 과정으로부터, 제시한 모델은 상황에 대한 사전 정보가 없이도 3가지 신호에 오류가 포함된 경우 95% 이상, 10가지 신호에 오류가 포함된 경우에도 90% 전후의 정확도로 이들을 식별할 수 있음을 확인하였다.

본 연구가 실용화되면 비상 상황 시 다수의 계측 신호에 오류가 발생한 상황에서도 운전원이 어떠한 계측 신호를 신뢰해야 하는지에 대한 충분한 정보를 제공할 수 있을 것으로 기대된다. 이는 오류가 발생한 계측 신호를 복원하는 기술과 연계되어, 최종적으로 운전원이 보다 적절한 의사결정을 내릴 수 있도록 도움을 줄 수 있다.

다만, 오류의 정도가 실험을 진행한 경우보다 완화될 경우 VAE 모델의 기본 복원 오차로 인하여 성능이 크게 하락할 것으로 예상된다. 따라서 더 정밀한 오류를 탐지하고 식별하기 위해서는 정상 데이터를 보다 정확하게 묘사할 수 있는 모델을 개발하여야 하며, 추후 발전된 모델과 더 많은 양의 데이터를 기반으로 후속 연구가 진행되어야 한다.

References

An, J. W., and Cho, S. Z. (2015). Variational autoencoder based anomaly detection using reconstruction probability, *Special Lecture on IE*, SNU Data Mining Center, 2, pp. 1-18.

Fantoni, P. F. (2000). A neuro-fuzzy model applied to full range signal validation of PWR nuclear power plant data, *International Journal of General Systems*, 29(2), 305-320.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets, *Advances in Neural Information Processing Systems 27 (NIPS 2014)*, Dec. 8-13, Montreal, Canada.

Hines, J. W., Uhrig, R. H., and Wrest, D. J. (1998). Use of autoassociative neural networks for signal validation, *Journal of Intelligent and Robotic Systems*, 21, 143-154.

Kim, S. G., Chae, Y. H., and Seong, P. H. (2020). Development of a generative -

adversarial - network - based signal reconstruction method for nuclear power plants, *Annals of Nuclear Energy*, 142, Article 107410.

Kim, S. G., No, Y. G., and Seong, P. H. (2015). Prediction of severe accident occurrence time using support vector machines, *Nuclear Engineering and Technology*, 47(1), 74-84.

Kim, S. S., Kim, J. I., and Jung, K. C. (2019). Portfolio system using deep learning, *Journal of the Korea Industrial Information Systems Research*, 24(1), 23-30.

Kim, S. S., and Hong, K. J. (2017). Development and performance analysis of predictive model for KOSPI 200 index using recurrent neural networks, *Journal of the Korea Industrial Information Systems Research*, 22(6), 23-29.

Kingma, D. P., and Welling, M. (2014). Auto-encoding variational Bayes, arXiv: 1312.6114 [stat] https://arxiv.org/abs/1312.6114 (Accessed on Dec. 07th, 2020).

Kingma, D. P., and Ba, J. L. (2014). Adam: a method for stochastic optimization, arXiv: 1412.6980v9 [cs.LG] https://arxiv.org/abs/1412.6980 (Accessed on Dec. 07th, 2020).

Korea Atomic Energy Research Institute (1990). *Advanced Compact Nuclear Simulator Textbook*, Nuclear Training Center in Korea Atomic Energy Research Institute.

Lim, D. H., Lee, S. H., and Na, M. G. (2010). Smart soft-sensing for the feedwater flowrate at PWRs using a GMDH algorithm, *IEEE Transactions on Nuclear Science*, 57(1), 340-347.

Minar, R. M., Tuan, T. T., and Ahn, H. J. (2020). An Improved VTON (Virtual-try-on) algorithm using a [air of cloth and human image, *Journal of the Korea Industrial Information Systems Research*, 25(2), 11-18.

Na, M. G., Park, W. S., and Lim, D. H. (2008). Detection and diagnostics of loss of coolant accident using support vector machines, *IEEE Transactions on Nuclear Science*, 55(1), 628-636.

Nair, A. M., and Coble, J. (2017). Bayesian inference for high confidence signal validation and sensor calibration assessment, *ANS 10th International Topical Meeting on Nuclear Plant Instrumentation, Control and Human-Machine Interface Technologies*, Jun. 11-15, San Francisco, CA, pp. 1688-1697.

No, Y. G., Lee, C. Y., and Seong, P. H. (2018). Development of a prediction method for SAMG entry time in NPPs using the extended group method of data handling (GMDH) model, *Annals of Nuclear Energy*, 121, 552-556.

No, Y. G., and Seong, P. H. (2016). Smart-sensing of the aux. feed-water pump performance in NPP severe accidents using advanced GMDH method, *Proceedings of the KNS 2016 Spring Meeting*, May 11-13. Jeju, Republic of Korea.

Shaheryar, A., Yin, X., Hao, H., Mahmood, Z., and Abuassba, A. (2018). Selection of optimal denoising-based regularization hyper-parameters for performance improvement in a sensor validation model, *Artificial Intelligence*, 50(3), pp. 341-382.

Shaheryar, A., Yin, X., Hao, H., Ali, H., and Iqbal, K. (2016). A Denoising based autoassociative model for robust sensor monitoring in nuclear power plants, *Science and Technology of Nuclear Installations*, https://doi.org/10.1155/2016/9746948.

Yang, J. E. (2014). Fukushima Dai-ichi accident: lessons learned and future actions

from the risk perspectives, *Nuclear Engineering and Technology*, 46(1), 27-38.

김 승 근 (SeungGeun Kim)

- 한국과학기술원 원자력 및 양자공학과 공학석사
- 한국과학기술원 원자력 및 양자공학과 공학박사
- (현재) 한국원자력연구원 지능형컴퓨팅연구실 선임연구원
- 관심분야: 원자력 계측제어, 인공지능, 딥러닝