

신경회로망 기반 우리나라 산업안전시스템의 모델링

최기흥^{1†}

Neural Network-based Modeling of Industrial Safety System in Korea

Gi Heung Choi^{1†}

†Corresponding Author

Gi Heung Choi

Tel : +82-2-760-4322

E-mail : gihchoi@hansung.ac.kr

Received : May 17, 2022

Revised : October 19, 2022

Accepted : February 3, 2023

Abstract : It is extremely important to design safety-guaranteed industrial processes because such process determine the ultimate outcomes of industrial activities, including worker safety. Application of artificial intelligence (AI) in industrial safety involves modeling industrial safety systems by using vast amounts of safety-related data, accident prediction, and accident prevention based on predictions. As a preliminary step toward realizing AI-based industrial safety in Korea, this study discusses neural network-based modeling of industrial safety systems. The input variables that are the most discriminatory relative to the output variables of industrial safety processes are selected using two information-theoretic measures, namely entropy and cross entropy. Normalized frequency and severity of industrial accidents are selected as the output variables. Our simulation results confirm the effectiveness of the proposed neural network model and, therefore, the feasibility of extending the model to include more input and output variables.

Copyright©2023 by The Korean Society of Safety All right reserved.

Key Words : industrial safety, artificial intelligence, neural network, entropy, cross entropy (discrimination information)

1. 서론

산업현장에서의 안전 즉 산업안전은 사고(재해)로부터 자유로운 정도로 이해할 수 있다. 산업안전은 산업재해의 원인을 제거하기 위한 접근방법으로 실제 적용면에서 인명과 재산의 손실을 효과적으로 제어하기 위한 다양한 방법을 사용하게 된다. 그러나 이러한 접근방법의 이론적 근거가 부족한 경우도 많으며 미래 지향적인 인공 지능 기술의 활용측면에서 더욱 그러하다. 인공지능(Artificial Intelligence, AI)은 인간의 학습능력, 추론능력, 지각능력을 기계 등에 인공적으로 시연(구현)한 것이다¹⁾. 즉, 인공 지능(AI)은 인간 지능을 모방하여 작업을 수행하고 수집한 정보를 기반으로 반복적으로 개선할 수 있는 시스템 또는 기계를 의미한다.

산업안전 분야에서 인공 지능 기술의 활용은 산업재해 관련 방대한 데이터에 기초한 산업안전시스템 모델(Model of Industrial Safety System)과 이를 활용한 사고

의 예측, 그리고 예측에 기초한 예방으로 이루어진다. 이를 통해 산업재해 예방예산의 효율성과 산업안전 정책의 효율성을 확보하여 우리나라 산업안전제도의 완성도를 크게 높일 수 있다.

우리나라 산업안전시스템 모델링에 필요한 산업재해 관련 데이터는 업무상 사고 및 질병으로 승인받은 사망 또는 3일 이상 휴업이 필요한 부상재해를 대상으로 산업재해 조사표에 근거하여 작성된다. 산업재해 조사표는 재해와 연관된 사업장과 근로자의 기본정보, 재해의 발생형태, 기인물, 원인발생형태, 요양기간과 근로손실일수를 포함한 재해강도, 그리고 재해개요 등 다양한 정보를 포함하고 있다²⁾.

본 연구에서는 우리나라 산업안전 분야에 인공지능 기술을 적용하기 위한 첫 단계로 산업재해 데이터에 근거하여 산업안전시스템을 모델링한다. 구체적으로는, 산업재해 조사표에 나타난 다양한 입력변수 중 재해의 빈도와 강도 등 출력변수에 가장 민감한 변수를 선정하는 방법을 제안하고 이들 입력변수와 출력변수

¹한성대학교 기계시스템공학과 교수(Department of Mechanical Systems Engineering, Hansung University)

간 신경회로망 모델을 정립한다. 본 연구에서 제안하는 체계적인 입력변수 선정체계와 입, 출력변수를 적용하고 신경회로망에 기초한 산업안전시스템 모델은 바람직한 산업안전 전략의 수립에 활용될 수 있다.

2. 모델링을 위한 입력 및 출력변수의 선정

2.1 위험성평가를 고려한 출력변수의 선정

일반적으로 산업안전 관련 위험성(Risk)은 Eq.(1)과 같이 재해빈도와 강도에 근거하여 평가할 수 있다.

$$\text{위험성(Risk)} = \text{재해빈도(Frequency)} \times \text{재해강도(Severity)} \quad (1)$$

재해빈도는 위험이 사고로 발전될 확률로 폭로빈도를 의미하며 재해강도는 부상 및 건강장애의 정도 또는 재산손실의 크기를 의미한다. 본 연구에서는 우리나라 산업안전시스템 모델의 출력변수로 재해빈도를 나타내는 재해건수와 재해강도를 나타내는 사망재해건수를 적용한다.

2.2 입력변수의 선정방법

Fig. 1은 산업안전시스템 모델링을 위한 지표(일반적으로 모델의 출력변수 값을 의미)를 평면상에 도시한 것이다³⁾. 개별 입력변수 내 Class 간 출력변수 값이 가장 크게 차별화되고(Class 1과 Class 2의 중심 간 거리가 최대) 동시에 각 Class 안에서는 가장 적게 차별화되는(Class 1과 Class 2의 반경이 최소) 입력변수를 찾아 모델링에 활용하는 것이 본 연구의 목표에 부합한다. 이와는 반대로 Class 간 차별화가 어렵지만 Class 안에서는 크게 차별화되는 입력변수 즉, Class 중심 간

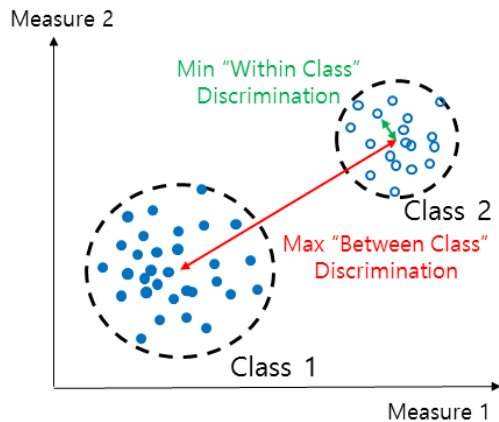


Fig. 1. Selection of input variables where classes are clearly discriminated based on the output variables.

거리보다 Class 내 반경이 큰 입력변수는 본 연구의 목표에 부합하지 않는다.

본 연구에서는 위에서 언급한 입력변수 선정방법을 적용할 대상으로 엔트로피(Entropy) 개념을 고려하였다. 엔트로피 개념은 물리학자 볼츠만에 의해 도입되었으며 어떤 상태에서 프로세스의 불확실성 또는 이와 동등한 의미로 평균 정보량을 표현한다⁴⁾.

x 가 가능한 상태 Set D 를 갖는 어떤 시스템의 특정 상태 즉 $x \in D$ 이고 D 에서 정의된 확률 q 의 Set이 Q 인 경우

$$q(x) \geq 0$$

$$\int_D dx q(x) = 1 \quad (2)$$

을 만족한다고 가정한다. 이제 확률 q 로 표시되는 프로세스의 엔트로피는 다음과 같이 주어진다:

$$E[q] = - \int_D q(x) \log(q(x)) dx \quad (3)$$

Set D 가 이산분포 $D = \{x_1, \dots, x_n\}$ 일 경우, 엔트로피(또는 평균 정보량)는 개별 상태와 모든 상태가 포함된 프로세스 전체에 대해 각각 다음과 같이 정의된다:

$$E_i = - \log q_i \quad (4)$$

$$E = \sum_i E_i = - \sum_i \log q_i \quad (5)$$

엔트로피 값은 모든 상태의 확률이 동일할 때 즉, $q_i = q(const)$ 일 때 최대가 되며 가장 불확실한 상태를 의미한다. 반대로 어떤 특정 상태의 확률이 1이고 나머지 사건의 확률이 0이면 즉, $q_i = 1, q_j = 0, j \neq i$ 일 때 엔트로피 값은 최소 (0)가 되며 불확실성이 없는 상태이다.

크로스 엔트로피(또는 차별화 정보량, Discrimination Information)는 엔트로피 개념을 사전(Prior) 확률분포가 존재할 경우로 확장하여 일반화한 것이다^{5,6,7,8)}. 즉, x 가 가능한 상태 Set D 를 갖는 어떤 시스템의 특정 상태 즉 $x \in D$ 이고 D 에서 정의된 확률 q 의 Set이 Q 이며 Eq.(2)를 만족한다고 가정한다. 이제 사전(Prior) 확률분포 $p \in Q$ 는 $p(x \in D) > 0$ 를 만족하는 순양(Strictly Positive) 값일 때 사후(Posterior) 확률분포 $q \in Q$ 를 통해 새로 얻는 정보의 양(차별화 정보량)은 다음과 같이 크로스 엔트로피에 의해 주어진다:

$$H[q, p] = - \int_D dx q(x) \log \left(\frac{q(x)}{p(x)} \right) \quad (6)$$

이산분포의 경우 Eq(6)은 다음과 같이 쓸 수 있다. 즉, 독립 상태 i 에 대하여 크로스 엔트로피는

$$H_i[q, p] = - \log \frac{q_i}{p_i} \quad (7)$$

와 같이 주어지며 모든 상태에 대하여 다음과 같이 표현된다:

$$H[q, p] = \sum_i H_i[q, p] = - \sum_i \log \frac{q_i}{p_i} \quad (8)$$

2.4 입력변수별 Class 정의 및 정규화

우리나라 산업안전시스템의 신경회로망 모델링은 주어진 입력변수에 대하여 출력변수인 재해의 빈도(재해건수)와 재해의 강도(사망재해 건수)를 예측하는 것을 목표로 한다. 산업재해 조사표에 나타난 다양한 입력변수 중 사업장 관련 정보인 업종, 기업규모(근로자수) 그리고 근로자 관련 정보인 성별, 직종, 근속기간 그리고 근로자의 나이 등 6가지를 본 연구에서 제안하는 신경회로망 모델의 입력변수로 고려할 수 있다. 다만, 근로자의 성별(남 또는 여)은 단순한 형태의 입력력으로 차별화되는 정보를 제공하지 못하므로 제외하였다.

이제 다섯 개 각 입력변수 내 Class는 산업안전보건공단에서 작성하는 분류기준에 따라 Table 1~Table 5과 같이 정의할 수 있다. 이 때, 입력변수는 모델링 중

Table 1. Classes and normalized values of input variable 1 (Business Type(Large)) used in neural network-based modeling of industrial safety system in Korea

Input variable	Class	Description	Value
(1) Business type	1	Construction	0
	2	Mining	0.111
	3	Finance and insurance	0.222
	4	Others	0.333
	5	Agriculture	0.444
	6	Fishery	0.556
	7	Transportation, warehouse and telecommunication	0.667
	8	Forestry	0.778
	9	Electricity, gas, vapor and water supply	0.889
	10	Manufacturing	1

Table 2. Classes and normalized values of input variable 1 (Company Size (Number of Workers)) used in neural network-based modeling of industrial safety system in Korea

Input variable	Class	Description	Value
(2) Company size (Number of workers)	1	~5:	0
	2	5~9	0.091
	3	10~15	0.182
	4	16~29	0.273
	5	30~49	0.364
	6	50~99	0.455
	7	100~199	0.545
	8	200~399	0.636
	9	400~499	0.727
	10	500~999	0.818
	11	1000~1999	0.909
	12	2000~	1

Table 3. Classes and normalized values of input variable 5 (Age of Workers) used in neural network-based modeling of industrial safety system in Korea

Input variable	Class	Description	Value
(3) Age of worker	1	Not classified	0
	2	~18	0.071
	3	18~24	0.143
	4	25~29	0.214
	5	30~34	0.286
	6	35~39	0.357
	7	40~44	0.429
	8	45~49	0.500
	9	50~54	0.571
	10	55~59	0.643
	11	60~64	0.714
	12	65~69	0.786
	13	70~74	0.857
	14	75~79	0.929
	15	80~	1

Table 4. Classes and normalized values of input variable 1 (Work Duration) used in neural network-based modeling of industrial safety system in Korea

Input variable	Class	Description	Value
(4) Work duration	1	Not classified	0
	2	~1M	0.071
	3	1~2M	0.143
	4	2~3M	0.214
	5	3~4M	0.286
	6	4~5M	0.357
	7	5~6M	0.429
	8	6~12M	0.500
	9	1~2Y	0.571
	10	2~3Y	0.643
	11	3~4Y	0.714
	12	4~5Y	0.786
	13	5~10Y	0.857
	14	10~20Y	0.929
	15	20Y~	1

Table 5. Classes and normalized values of input variable 1 (Work Type) used in neural network-based modeling of industrial safety system in Korea.

Input variable	Class	Description	Value
(5) Work type	1	Manager	0
	2	Expert and related worker	0.125
	3	Office worker	0.250
	4	Service worker	0.375
	5	Sales worker	0.500
	6	Farmer and fisherman	0.625
	7	Technician	0.750
	8	Equipment/Machine operator and assembler	0.875
	9	Labor worker	1

입력변수의 영향을 고르게 반영하기 위해 [0, 1] 사이에서 정규화된 값(Normalized Value, 각 Table 중 "Value"로 표기)을 각각의 Class에 적용하였다.

2.5 확률분포에 기초한 엔트로피 계산

차별화되는 출력변수 값을 나타내는 입력변수를 선정하기 위해서는 출력변수 값에 기초하여 개별 입력변수에 해당되는 엔트로피 또는 크로스 엔트로피 값을 구해야 한다. 이들 값은 확률분포에 의해 계산되므로 개별 입력변수 내 각 Class에 해당되는 출력변수의 확률분포를 알면 구할 수 있다. 즉, 개별 입력변수 내 각 Class에 해당되는 재해건수와 사망재해 건수를 알면 각 Class에 해당되는 확률을 다음의 Eq.(9)에 의해 구할 수 있으며 개별 입력변수 내 Class별 확률분포도 구할 수 있다.

$$(q_a)_{i,j} = \frac{(N_a)_{i,j}}{\sum_j (N_a)_{i,j}}, (q_d)_{i,j} = \frac{(N_d)_{i,j}}{\sum_j (N_d)_{i,j}} \quad (9)$$

이 때, N_a , q_a 는 각각 재해빈도와 관련 확률, 그리고 N_d , q_d 는 각각 재해강도와 관련 확률을 나타내며 첨자 i, j 는 각각 입력변수와 Class를 의미한다.

2.6 신경회로망 모델 및 출력변수의 정규화

인공지능을 구현하는 한 방법인 신경회로망(Neural Network)은 생물학적 뇌의 작동원리를 그대로 모방하여 새로운 형태의 계산도구를 제공함으로써 지식획득 과정 및 획득된 지식의 표현을 보다 자유롭게 하는 특징이 있다. 신경회로망에 관한 기본적인 정보는 참고문헌 [9]에 나타나 있다. 본 연구에서는 Fig. 2와 같이 입력층, 은닉층, 출력층을 갖는 신경회로망 모델을 적용하였다.

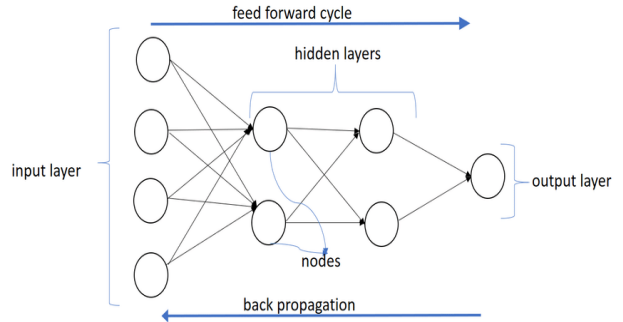


Fig. 2. Structure of neural network used in modeling of industrial safety system in Korea.

본 연구에서는 세 개 또는 네 개의 입력변수와 2개의 출력변수가 적용된 신경회로망 모델을 사용하며 모델링 방법은 다음과 같다:

- 노드간 전달함수: 시그모이드(Sigmoid) 함수

$$\frac{1}{1 + e^{-x}} \quad (10)$$

- 학습방법: 머신러닝(Machine Learning) 기법 중 지도학습(Supervised Learning)을 적용하되 하강기울기(Descent Gradient)에 기초한 역전파 알고리즘(Back Propagation Algorithm)을 적용
- 사용 소프트웨어: TensorFlow¹⁰⁾

지도학습의 경우 오버피팅(Over fitting) 문제를 해소하기 위해 필요한 은닉층(Hidden Layer)의 노드의 수는 다음의 Eq.(11)을 적용하여 구할 수 있다¹¹⁾.

$$N_h = \frac{N_s}{(\alpha \cdot (N_i + N_o))} \quad (11)$$

N_h = 은닉층 노드 수 (100)

N_i = 입력층 노드 수 (3 또는 4)

N_o = 출력층 노드 수 (2)

N_s = 학습샘플 수 (1,080 또는 1,800)

α = 임의의 수로 보통 2~10 사이에서 결정된다. 은닉층 노드 수는 Layer 수와 각 Layer당 노드수에 의해 결정되며 본 연구에서는 다섯 개의 Layer와 Layer 당 20개의 노드 등 총 100개의 은닉층 노드를 적용하였다. 따라서 α 는 3개의 입력노드 기준 샘플의 수에 따라 2.16 또는 3.6이다.

입력노드 값은 Table 1~Table 5에 제시된 정규화된 값(Value)을 적용하였으며 출력변수 값은 재해빈도(출력변수 1)와 재해강도(출력변수 2) 모두 학습샘플 중 가장 큰 값을 1로 놓고 이 값을 기준으로 [0, 1] 구간에서 정규화하여 출력노드 값으로 사용하였다.

3. 적용 예

3.1 입력변수의 선정

본 연구에서는 2016년도 산업재해 데이터 81,209건을 신경회로망 모델링에 활용하였다. 2017년도 산업재해 데이터 79,972건은 입력변수가 출력변수에 미치는 영향을 비교할 목적으로 활용하였다. 처리해야 하는 데이터의 양이 많고 단순한 입, 출력구조를 갖는 모델이 학습하기 쉬우며 활용편의성 면에서 유리함을 고려할 때 개별 입력변수 내 Class 간 재해빈도 및 재해강도가 가장 차별화되는 입력변수를 모델링에 우선적으로 활용하는 것이 필요하다. Fig. 3과 Fig. 4는 2016년도 산업재해 데이터를 사용하여 계산된 개별 입력변수 내 각 Class의 확률분포를 나타낸 것이며 Table 6은 이러한 확률분포를 사용하여 계산된 엔트로피 값과 재해빈도와 재해강도 간 크로스 엔트로피값을 나타낸 것이다.

Fig. 3과 Fig. 4에 따르면 입력변수 3은 Class간 확률분포의 변화가 나머지 입력변수에 비해 적다. 실제로 Table 6에 따르면 각 입력변수 내 Class간 차별화 정도를 나타내는 엔트로피값은 재해빈도와 재해강도 모두에 대해 입력변수 1-5-2-4-3 (업종 > 직종 > 사업장 규모 > 근로자 근속기간 > 근로자 나이)의 순으로 작다 (차별화 정도가 크다). 또한, 두 출력변수 사이에 차별화되는 정도를 나타내는 크로스 엔트로피값은 클수록 재해빈도와 재해강도 사이에 큰 차별화가 가능하므로

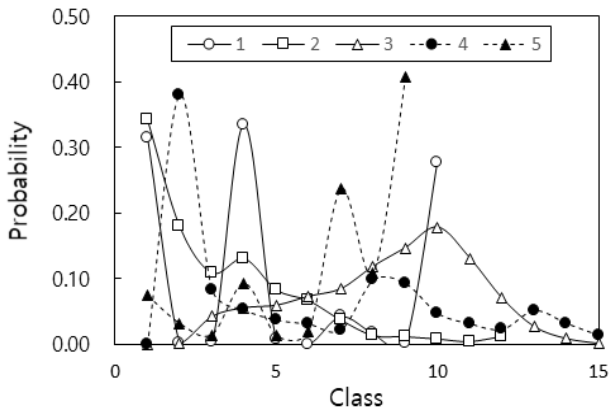


Fig. 3. Probability distribution of output variable 1 (frequency of accident) for classes of 5 input variables used in neural network-based modeling of industrial safety system in Korea.

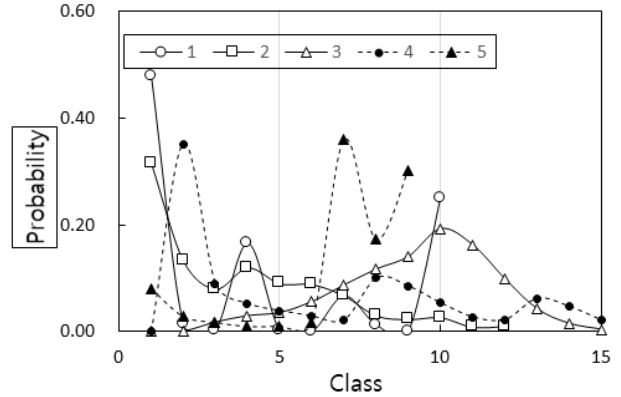


Fig. 4. Probability distribution of output variable 2 (severity of accident) for classes of 5 input variables used in neural network-based modeling of industrial safety system in Korea.

Table 6. Probability distribution of output variable 1 (frequency of accident), output variable 2 (severity of accident) and cross entropy between output variable 1 and 2 for classes of 5 input variables used in neural network-based modeling of industrial safety system in Korea

Input variable	Entropy based on output variable 1 (Norm. frequency)	Entropy based on output variable 2 (Norm. severity)	Cross entropy
1	1.371	1.354	0.116
2	1.927	2.093	0.052
3	2.316	2.264	0.041
4	2.162	2.226	0.008
5	1.668	1.573	0.119

입력변수 5-1-2-3-4 (직종 > 업종 > 사업장 규모 > 근로자 나이 > 근로자 근속기간)의 순으로 차별화 정도가 크다. 따라서 본 연구에서는 Fig. 3, Fig. 4 그리고 Table 6의 결과에 기초하여 입력변수 1, 2, 4 및 5(사업장의 업종과 규모, 근로자의 직종과 근속기간)을 신경회로망 모델링의 입력변수로 사용하였다.

3.2 신경회로망 모델링

선정된 네 개의 입력변수 모두를 적용한 1-2-3-4 신경회로망 모델의 경우 Table 1~Table 5에 제시된 각 입력변수별 Class의 수를 고려하면 총 16,200 (10×12×9×15) 입력변수의 조합이 가능하다. 그러나 매년 발생하는 사망재해 건수는 700건 내외이므로 위 총 입력변수 조합 중 사망재해가 발생하지 않는 조합이 대부분이다. 즉, 각 입력변수 내 Class간 대부분의 확률이 0이므로 확률분포가 무의미해진다. 반면에 선정된 4개의 입력변수 중 1, 2, 3 입력변수를 적용한 1-2-3 신경회로망 모델의 경우 총 1,080 (10×12×9) 입력변수 조합과 이에 대응하는 출력변수가 사용되며 1, 2, 4 입력

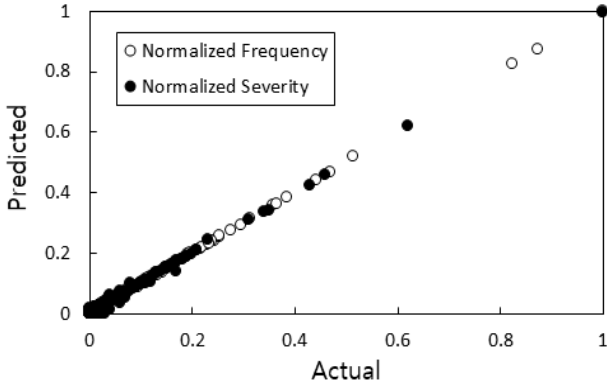


Fig. 5. Comparison of the actual and the predicted frequency and severity of 3-level 1-2-3 input neural network model (Business type-company size-work type model).

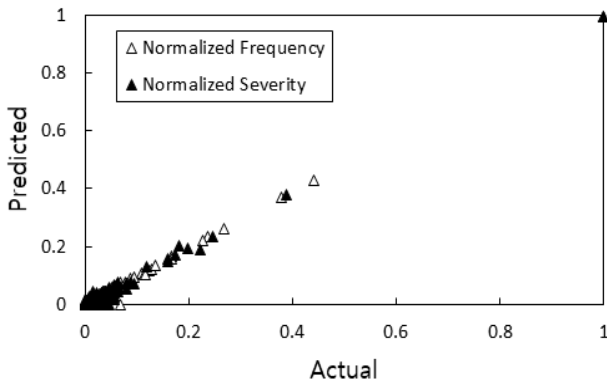


Fig. 6. Comparison of the actual and the predicted frequency and severity of 3-level 1-2-4 input neural network model (Business type-company size-work duration model).

변수를 적용한 1-2-4 신경회로망 모델의 경우 총 1,800 (10×12×15) 입력변수 및 이에 대응하는 출력변수 값이 적용되어 대부분의 입력변수 조합이 유의미한 출력값을 갖는다. 이러한 이유로 본 연구에서는 1-2-3 신경회로망 모델과 1-2-4 신경회로망 모델을 고려하였다.

1-2-3 신경회로망 모델의 경우, 2016년도 산업재해 데이터 81,209건을 사용하여 총 1,080개 입력조합에 대한 정규화된 출력변수 값을 구하고 이를 학습샘플로 활용하였다. Fig. 5에 도시된 바와 같이 모델링에 사용된 1,080개 학습샘플의 정규화된 재해빈도와 재해강도의 실제 값과 학습된 신경회로망 모델로부터 얻은 예측 값은 거의 일치한다. 실제 값과 예측 값의 MSE (Mean Squared Error)는 2.46×10^{-5} 이었다. Fig. 6에 도시된 바와 같이 동일한 과정을 거쳐 얻어진 1-2-4 신경회로망 모델의 경우 모델링에 사용된 1,800개 학습샘플의 실제 값과 신경회로망의 예측 값은 거의 일치하며 MSE는 3.44×10^{-5} 이었다.

3.3 신경회로망 모델 기반 재해빈도 및 강도 추정

학습된 신경회로망 모델을 활용하면 임의의 입력변수 값 조합에 해당되는 정규화된 출력변수 값(재해빈도와 재해강도)을 추정할 수 있다.

3.3.1 (1-2-3) 신경회로망 모델

- 입력변수: 제조업, 근로자수 78명, 기능직

Class내 보간(Interpolation)에 의해 입력변수 값(1, 0.506, 0.750)을 1-2-3 모델에 입력하여 정규화 재해빈도 0.047과 정규화 재해강도 0.035를 얻을 수 있다. 이는 곧 발생가능한 최대 빈도대비 4.7%, 최대강도 대비 3.5% 정도의 재해를 예상할 수 있음을 의미한다.

3.3.2 (1-2-4) 신경회로망 모델

- 입력변수: 건설업, 근로자수 326명, 근속기간 12년

Class 내 보간에 의해 입력변수 값(0, 0.692, 0.943)을 1-2-4 모델에 입력하여 정규화 재해빈도 0.000105와 정규화 재해강도 0.003209를 얻을 수 있으며 발생가능한 최대 빈도대비 0.01%, 최대강도 대비 0.3% 정도의 재해를 예상할 수 있다.

3.3.3 신경회로망 모델

이제, 2017년도 산업재해 데이터 79,972건을 사용하고 동일한 과정을 거쳐 모델링을 수행한 후 출력변수인 정규화된 재해빈도와 재해강도의 예측 값을 2016년도 값과 비교하여 각각 Fig. 7과 Fig. 8에 도시하였다. 1-2-3 입력모델과 1-2-4 입력모델 모두에 대하여 정규화된 재해빈도의 변화는 거의 없으나 정규화된 재해강도의 변화는 다소 뚜렷하게 나타난다. 이는 사망재해의 경우 부상재해에 비해 빈도는 낮으나 불규칙하게 발생하는 특징을 나타내며 동시에 사업장 규모가 연동된 경우 근로자의 직종보다는 근속기간이 사망재해에 더 큰 영향을 미침을 나타낸다. 특히, 참고문헌 [12]에 따르면 근로자의 근속기간이 1년 미만이거나 20년 이상일 때 재해의 빈도와 강도가 높은 것으로 나타났다. 근로자 1,000명이상 사업장의 경우 근로자 채용 시 상대적으로 젊고 근무초기 안전교육이 잘 이루어지고 있는 반면에 근로자 100인 미만 사업장의 경우 고령 근로자의 비중이 상대적으로 높으므로 근속기간이 짧은 초기에 고령 근로자 중심의 안전관리와 안전교육이 재해예방에 중요함을 알 수 있다.

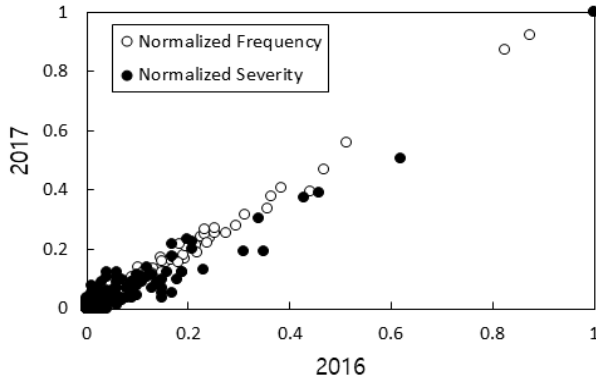


Fig. 7. Comparison of the normalized frequency and the severity predicted by the 3-level 1-2-3 input neural network model (Business type-company size-work type model).

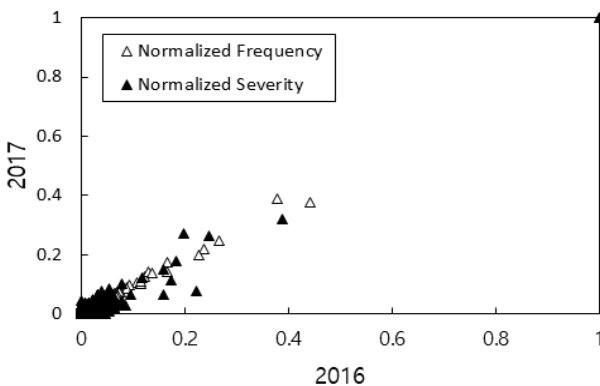


Fig. 8. Comparison of the normalized frequency and the severity predicted by the 3-level 1-2-4 input neural network model (Business type-company size-work duration model).

4. 결론

본 연구에서는 우리나라 산업안전 분야에 인공지능 기술을 적용하기 위한 첫 단계로 산업재해 데이터에 근거하여 산업안전시스템을 모델링하였다. 특히, 산업재해원인 조사표에 나타난 다양한 입력변수별 Class를 정의 및 정규화하고 각 Class에 해당되는 출력변수(재해빈도 또는 재해강도)의 확률분포를 적용한 점은 일반적인 산업재해분석 기법과 차별화된다. 엔트로피 개념에 기초하여 출력변수인 재해빈도와 재해강도 모두에 대해 차별화되는 정도가 큰(즉, 두 출력변수 모두가 민감하게 반응하는) 입력변수는 업종 > 직종 > 사업장 규모 > 근로자 근속기간 > 근로자 나이 등의 순이다. 반면에, 크로스 엔트로피 개념을 적용하여 두 출력변수 재해빈도와 재해강도가 반응하는 정도가 차별화되는 입력변수는 직종 > 업종 > 사업장 규모 > 근로자 나이 > 근로자 근속기간 등의 순이다. 이러한 분석방법과 분석결과는 이전 연구에서도 시도한 적이 없는 것

으로 이들 입력변수와 정규화된 재해의 빈도와 강도 등 출력변수를 입, 출력 노드로 하고 은닉층을 갖는 신경회로망 모델을 머신러닝 기법을 활용하여 지도학습하였다. 이 때 사용된 학습샘플은 각 입력변수 내 Class의 조합에 대응하는 입력과 출력변수 값이다.

본 연구에서는 세 가지 입력변수 조합에 대한 신경회로망 모델링 예를 제시하였으며 다양한 입력변수 조합에 대한 신경회로망 모델을 정립하면 사고예측의 범위가 확장되어 활용성을 크게 높일 수 있다. 즉, 학습된 모델과 보간(Interpolation 또는 Extrapolation)을 이용하여 임의의 입력변수 값에 대응하는 출력변수 값의 추정이 가능해진다. 따라서 본 연구에서 제안된 체계적인 입력변수 선정체계와 입, 출력변수를 적용한 신경회로망에 기초한 우리나라 산업안전시스템 모델은 바람직한 규제의 형태와 타당성을 검증하고 산재예방 예산의 효율성과 보다 세밀한 산업안전 전략의 수립 및 효율성 제고에 활용될 수 있다.

Acknowledgement: This study was supported by Hansung University.

References

- 1) Artificial Intelligence, <https://ko.wikipedia.org/>
- 2) G. H. Choi, "Cause Analysis of Accidents Associated with Dangerous Machines and Devices Subject to Safety Certification", J. Korean Soc. Saf., Vol. 35, No. 4, pp. 1-8, 2020.
- 3) G. H. Choi, "Improvement of Reliability in Cause Analysis of Industrial Accidents", J. Korean Soc. Saf., Vol. 29, No. 6, pp. 1-8, 2014.
- 4) G. H. Choi, "Application of Information-theoretic Measure (Entropy) to Safety Assessment in Manufacturing Processes", Int. J. Safety, Vol. 4, No. 1, pp. 8-13, 2005.
- 5) G. H. Choi and B. H. Ryu, "Application of Discrimination Information (Cross Entropy) as Information-theoretic Measure to Safety Assessment in Manufacturing Processes", Int. J. Safety, Vol. 4, No. 2, pp. 1-5, 2005.
- 6) J. Shore, "Minimum Cross Entropy Spectral Analysis, Naval Research Laboratory Memorandum Report 3921, NRL, 1979.
- 7) John Shore and Rodney Johnson, "Properties of Cross Entropy Minimization", Transaction on Information Theory, Vol IT-27, No. 4, IEEE, pp. 472-482, 1981.
- 8) J. E. Shore and R. M. Gray, "Minimum Cross-Entropy

- Pattern Classification and Cluster Analysis”, IEEE Trans. On Pattern Analysis and Machine Intelligence, Vol. PAMI-4, No. 1, pp. 11-1103, 1982.
- 9) Neural Network, <http://www.aistudy.com>
 - 10) TensorFlow, <https://www.tensorflow.org/>
 - 11) <https://stats.stackexchange.com/questions/181/how-to-choose-the-number-of-hidden-layers-and-nodes-in-a-feedforward-neural-netw>
 - 12) G. H. Choi, “Cause Analysis of Accidents at Workplaces in Korea”, J. Korean Soc. Saf. (Submitted for publication)