

An Artificial Neural Network-based Model for Estimating Damage of Natural Disasters

Keun-Chae Jeong⁺

School of Civil Engineering, Chungbuk National University, Chungdae-ro 1, Seowon-Gu, Cheongju, Chungbuk 28644, Korea

Abstract

In the previous research, we developed a regression model for estimating damage of natural disasters based on the public database. Although this model considers nonlinearities among variables by using log transformation for the dependent variable, it reveals limitations in improving estimation accuracy because of its inherent characteristics of linearity assumption between independent and dependent variables. In this study, we proposed an artificial neural network (ANN) based model to predict the amount of damages due to natural disasters more accurately, which does not require the linearity assumption among the variables. For verification of the proposed model, we compared the model estimates with those from the regression model, including the Natural Disaster Risk Index (NDRI), Regional Safety Grades (RSG), and actual damage amounts. According to the results of analysis, we can confirm that the estimates from the ANN-based model reveal a higher correlation with the actual damage amounts than those from the regression model or the assessment results of NDRI and RSG.

Key words: natural disaster, damage amount, artificial neural network, forecast, non-linearity

1. 서론

1. 연구 배경 및 목적

최근 10년 동안 우리나라에 발생했던 자연재해 피해액을 분석해보면, <Figure 1>에 나타나 있는 바와 같이 연 평균 피해액이 3,441억 원에 달하며, 발생한 피해를 복구하기 위해 피해액의 두 배가 넘는 연 평균 7,709억 원이 소요된 것을 알 수 있다. 발생한 피해액을 자연재해 원인별로 살펴보면, 전체 피해액에서 차지하는 비율이 각각 호우 44.0%, 태풍 43.7%, 대설

7.0%, 지진 2.9%, 강풍 및 풍랑 2.5% 순으로 호우와 태풍으로 인한 피해액이 전체 피해액의 87.7%로 피해액의 대부분을 차지하고 있다(Ministry of Interior and Safety, 2002-2019).

이와 같이 심각한 자연재해로 인한 피해액을 줄이기 위해서는, 효율적이고 효과적인 자연재해 저감 정책의 수립이 필요하다. 이를 위해서는 방재예산을 투입하기 위한 사전 작업으로 어느 지역이 자연재해에 더욱 취약한가를 분석할 필요가 있다. 이는 자연재해에 취약한 지역에 보다 많은 재해 예방 예산을 투입함

⁺ Corresponding author: Keun-Chae Jeong, Tel. +82-43-261-2401, Fax. +82-43-275-2377, e-mail. kcjeong@cbnu.ac.kr

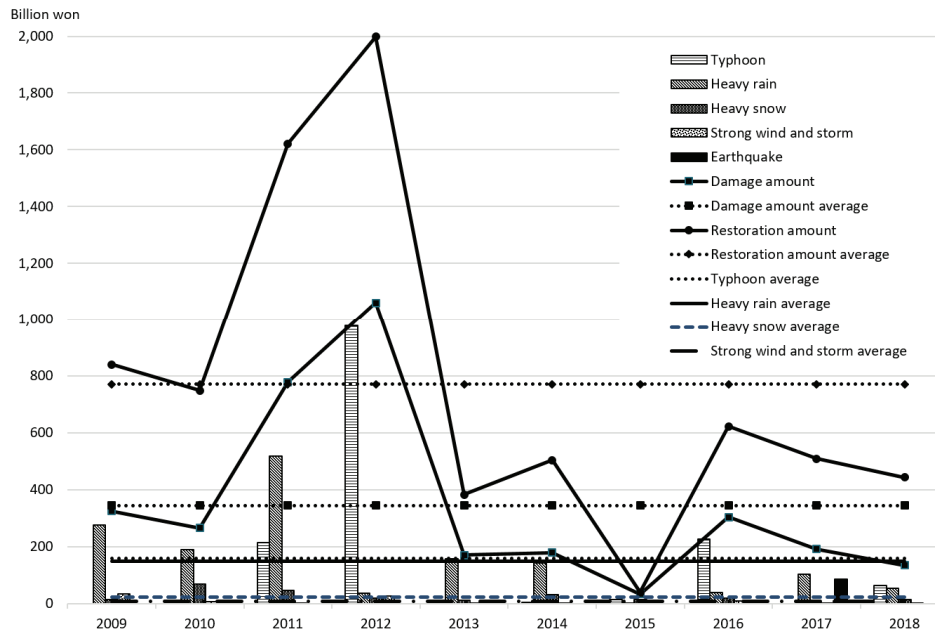


Figure 1. Damage and restoration costs from natural disasters in the recent 10 years

으로써, 자연재해로 인한 피해액 감소 효과를 극대화하기 위함이다.

이러한 노력의 일환으로 2011년부터 소방방재청에서는 지역안전도 진단 평가 지표를 개발하여 이용하고 있다. 이 지표는 잠재적 자연 재해 발생 가능성 및 환경적 위험을 진단하기 위한 위험환경, 재해를 저감시키기 위한 행정적인 노력을 진단하기 위한 위험관리능력, 지역의 구조적인 재해방어능력을 진단하기 위한 방재성능의 세 가지 측면을 종합평가하고 있다(National Emergency Management Administration, 2011). 그러나 지역안전도 지표는 재해 전문가가 직접 현장을 방문하여 평가를 수행하는 정성적 방식을 부분적으로 이용함으로써, 정량적인 자료만을 이용하는 국제적인 재해 평가 지표들에 비해 투입인력과 비용부담이 커지고 평가기간이 길어지는 단점을 갖는다(Yeon, et. al., 2016).

이러한 단점을 극복하기 위해, 국내에서도 공공분야의 정량자료만을 활용하여 재해위험도를 평가하는 자연재해 위험지표 모형이 개발되었다(Yeon, et. al., 2016). 이 모형은 위험지표 평가에 필요한 각종 자료를 모두 공공 부문 데이터베이스에서 수집하며, 자연재

해로 인한 위험성, 재해에 대한 사회경제적 취약성, 자연재해 전후의 대응능력, 자연재해 피해 대상 시설의 분포 등 네 가지 측면에서의 평가 결과를 바탕으로 종합적 성격의 자연재해 위험지표 값을 산출한다.

이들 기존 연구에서는, 지역안전도가 낮거나 자연재해 위험지표가 높은 지역이 실제 자연재해로 인해 발생한 피해액이 높을 것이라는 합리적 추론을 바탕으로, 지역안전도 및 자연재해 위험지표 평가 결과와 해당 지역의 실제 피해액과의 상관관계를 분석하였다. 분석 결과, 지역안전도와 실제 피해액 간 상관관계수가 0.470, 자연재해 위험지표와 실제 피해액 간 상관관계수가 0.462로 어느 정도의 상관관계는 갖지만 그 정도가 매우 높지는 않은 상황이었다(Yeon, et. al., 2016).

이에 자연재해로 인한 피해액을 직접적인 방법으로 추정하기 위한 연구가 진행되었다. 이를 위해 먼저 자연재해 피해액과 자연재해 발생 빈도 및 강도 간의 상관관계를 분석하는 연구가 선행되었다. 이 연구는 회귀분석 방법론을 기반으로 개별 자연재해의 피해액을 종속변수로 발생 빈도와 강도를 독립변수로 설정한 후, 독립변수에 대한 표준화 회귀계수를 추정하여 이를 바탕으로 종속변수와 상관관계를 분석하였다

(Yeon & Jeong, 2018). 또한, 기상자료를 바탕으로 회귀분석과 인공신경망에 이용하여 자연재해 피해를 예측하고자하는 연구가 진행되었다(Kim & Yoon, 2017). 그러나 이모형들은 호우, 태풍, 대설, 강풍 및 풍랑 등 개별적인 자연재해로 인한 피해액은 예측할 수 있었으나 전체 피해액을 예측하지는 못하며, 발생 빈도와 강도, 기상자료 등 자연재해 자체의 영향요인 이외의 다른 사회, 경제, 정책적 영향요인들을 독립변수로 포함시키지 못했다는 한계점을 가졌다.

이러한 한계점을 극복하기 위해 최근에는 자연재해 피해액을 추정할 수 있는 보다 개선된 형태의 회귀분석 모형이 개발되었다(Jeong, 2019). 이모형은 자연재해에 대한 발생 빈도와 강도 이외에도 추가적으로 사회, 경제, 정책적 영향요인을 포함시켜 독립변수를 구성하였으며 종속변수로는 자연재해로 인한 총 피해액을 이용하였다. 모형에 대한 예측력 분석 결과, 이모형을 이용한 예측 피해액과 실제 피해액 간의 상관관계수가 0.720으로 지역안전도와 자연재해 위험지표 모형에 비해 실제 피해액과의 상관관계가 50% 이상 향상되는 결과가 나타났다.

그러나 제시된 회귀분석 모형이 종속변수의 로그 변환을 통해 독립변수와 종속변수의 비선형성을 반영하였음에도 불구하고, 회귀분석 모형 자체의 독립변수와 종속변수 간 선형성에 대한 가정은 모형의 예측력을 높이는 데 한계점으로 작용하였다. 한편, 해외에서는 인공신경망을 이용하여 홍수자료, 위성영상 등을 독립변수로 삼아 자연재해 발생을 조기에 감지할 수 있는 방법론에 대한 연구(Wan, 2013; Santoso, 2015)가 진행되었다. 또한, 자연재해에 대한 조기 경보를 넘어서, 인공신경망을 통해 다양한 사회, 환경, 경제적 자료를 기반으로 자연재해로 인한 직간접적 손실을 예측하고자 하는 시도가 있었다(Rim, 2010). 이에 본 연구에서는 독립변수와 종속변수 간의 상호관계에 대한 선형성 가정이 필요하지 않은 인공신경망 모형을 이용함으로써, 자연재해로 인한 실제 피해액에 대한 예측력을 더욱 높일 수 있는 모형을 제안하고자 한다.

2. 연구 범위 및 방법

재난 및 안전관리 기본법 제 3조에 의하면 자연재난은 “태풍, 홍수, 호우, 강풍, 풍랑, 해일, 대설, 한파, 낙뢰, 가뭄, 폭염, 지진, 황사, 조류 대발생, 조수, 화산활동, 소행성·유성체 등 자연우주물체의 추락·충돌, 그 밖에 이에 준하는 자연현상으로 인하여 발생하는 재해”로 정의되어 있다(Disaster and Safety Management Basic Law, 2004). 본 연구에서는 이들 자연재난 중 재해연보에 피해액이 보고된 태풍, 호우, 대설, 강풍 및 풍랑 등 4가지 자연재해로 연구범위를 한정한다(Ministry of Interior and Safety, 2002-2015). 본 연구에서 종속변수로 사용된 자연재해 피해액과 독립변수로 사용된 자연재해 발생 빈도 및 강도와 사회, 경제, 정책적 요인 등 모든 정량적 자료는 전국 225개 시군구 단위의 기초지방자치단체별로 집계되었으며, 기존 연구들과의 비교·분석을 위해 시간적 범위는 기존 연구들과 동일하게 2001년부터 2014년까지로 한정하였다.

본 연구의 방법론은 <Figure 2>에 묘사되어 있는 바와 같이, 인공신경망 모형 구축 단계와 검증 단계로 양분되어 있다. 먼저 구축 단계에서는, 선행 연구인 회귀분석 기반 자연재해 피해액 예측 모형에서 사용되었던 14개 독립변수를 사용하여 자연재해 피해액을 추정하기 위한 인공신경망 모형을 수립한다. 이후, 모형에서 은닉층의 개수, 은닉층의 활성화 함수 종류, 출력층의 활성화 함수 종류를 다양하게 설정하여 인공신경망 모형의 성능을 평가한 후, 평가 결과를 바탕으로 최적 인공신경망 예측 모형을 선정한다.

다음으로 검증 단계에서는, 기존 다른 모형들과의 자연재해 피해액 예측력 비교·분석을 위해 본 연구에서 제안하는 인공신경망 기반 모형을 이용하여 자연재해 피해액 예측을 수행한다. 이후, 예측 결과를 바탕으로 기존의 지역안전도 등급, 자연재해 위험지표, 회귀분석 기반 모형 예측 피해액, 그리고 실제 자연재해 피해액 등과 상관분석을 수행함으로써, 제안된 인공신경망 기반 자연재해 예측 모형의 유용성을 검증한다.

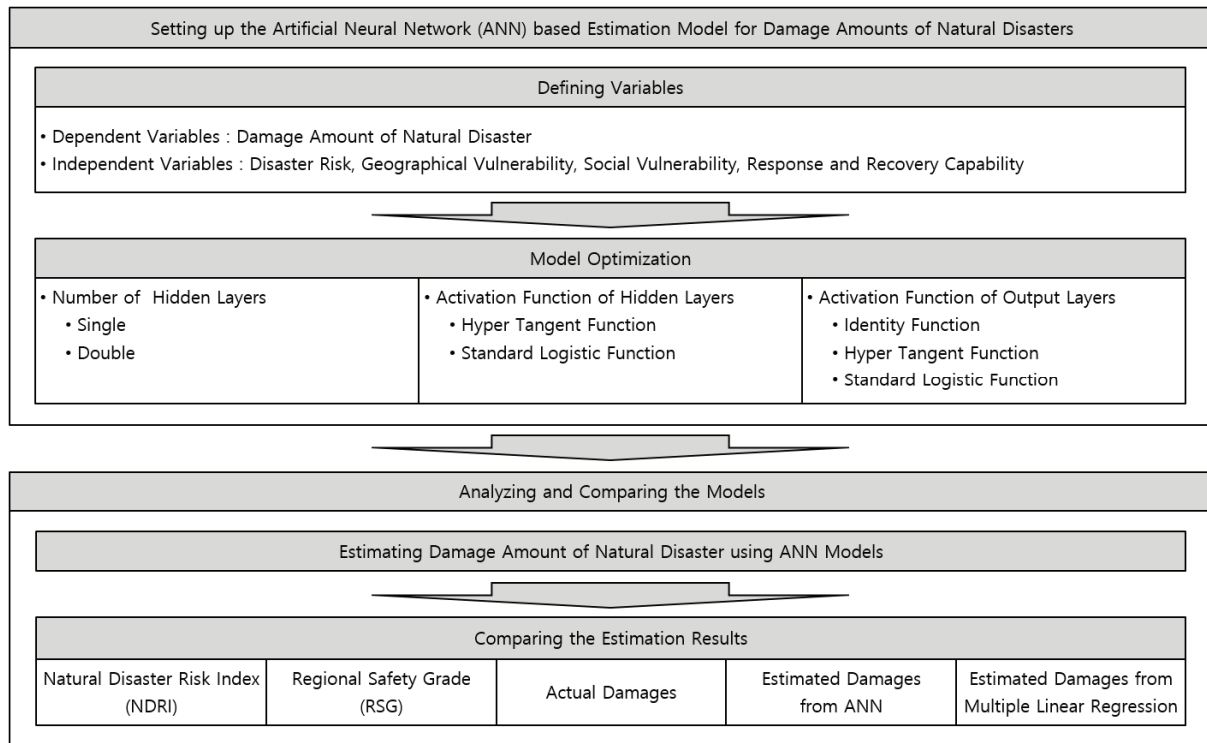


Figure 2. Research methodology

II. 회귀분석 모형

1. 독립변수와 종속변수 정의

선행 연구에서는 공공 데이터베이스를 기반으로 자연재해 피해액을 추정할 수 있는 회귀분석모형을

제안한 바 있다(Jeong, 2019). 이모형은 종속변수인 자연재해 피해액을 예측하기 위해 <Table 1>에 나타나 있는 바와 같이, 위험도, 지리적·사회적 취약성, 대응 능력 등을 대변하는 14개 독립변수를 사용한다. 이 변수들에 대한 시계열 자료는 재해연보, 국가통계포털,

Table 1. Independent variables of the regression model for estimating damage amounts of natural disasters

Number	Independent Variable	Definition	Direction	Dimension
1	Heavy Rain Frequency	Number of Annual Heavy Rain Occurrences	+	number/year
2	Heavy Rain Strength	The Maximum Annual Rainfall	+	mm
3	Typhoon Frequency	Number of Typhoon Rain Occurrences	+	number/year
4	Typhoon Strength	The Maximum Annual Typhoon Wind Speed	+	m/second
5	Heavy Snow Strength	Number of Annual Heavy Snow Occurrences	+	number/year
6	Heavy Snow Frequency	The Maximum Annual Snowfall	+	cm
7	Heavy Wind and Storm (HW&S) Strength	The Maximum Annual HW&S Wind Speed	+	m/second
8	River Density	River Area / Total Area	+	m ² /m ²
9	Ship Density	Ship Tons / Total Area	+	ton/m ²
10	Population Density	Population / Total Area	-	man/m ²
11	Higher Education Ratio	Number of University Graduates / Population	-	man/man
12	Foreigner Ratio	Registered Foreigners / Population	-	man/man
13	Disaster Prevention Budget Ratio	Disaster Prevention Budget / Total Budget	-	won/won
14	Population per Fireman	Population / Number of Firemen	+	man/man

※ Source: A Quantitative Data based Natural Disaster Risk Index Model in South Korea(2016: 173).

지방자치단체통계연보, 국가재난안전포털, 국가기후데이터센터, 국가태풍센터 등에서 제공하고 있다 (Yeon, et. al., 2016).

위의 독립변수는 각각 서로 다른 방향성과 단위를 갖기 때문에, 종속변수에 미치는 영향을 균형화하기 위해 14개 독립변수를 다음과 같이 표준화하여 사용한다. 기초지방자치단체 j 에 대한 연도 t 의 표준화된 독립변수 i 값, x_{ijt} 는 변수의 방향성에 따라 다음과 같이 정의된다.

양의 방향성(+):

$$x_{ijt} = (X_{ijt} - X_{i_{\min}}) / (X_{i_{\max}} - X_{i_{\min}}) \quad (\text{식 1})$$

음의 방향성(-):

$$x_{ijt} = (X_{i_{\max}} - X_{ijt}) / (X_{i_{\max}} - X_{i_{\min}}) \quad (\text{식 2})$$

- i : 독립변수를 나타내는 인덱스, $i = 1, 2, \dots, 14$
- j : 기초지방자치단체를 나타내는 인덱스, $j = 1, 2, \dots, 225$
- t : 연도를 나타내는 인덱스, $t = 2001, 2002, \dots, 2014$
- X_{ijt} : 연도 i 에 대한 기초지방자치단체 i 의 독립변수 i 값
- $X_{i_{\min}}$: 모든 j 와 t 에 대한 X_{ijt} 중 최솟값
- $X_{i_{\max}}$: 모든 j 와 t 에 대한 X_{ijt} 중 최댓값

선행 연구에서 제안한 회귀분석 모형은, 종속변수인 피해액과 독립변수 간의 비선형 관계를 선형으로 변환하기 위해 먼저 피해액에 자연로그를 취한 후, 종속변수가 독립변수와 마찬가지로 0부터 1사이의 값을 갖도록 표준화된 피해액으로 변환하여 사용한다. 연도 t 에 대한 기초지방자치단체 j 의 표준화된 피해액 d_{jt} 는 다음과 같이 정의된다.

$$d_{jt} = (\ln D_{jt} - \ln D_{\min}) / (\ln D_{\max} - \ln D_{\min}) \quad (\text{식 3})$$

D_{jt} : 연도 t 에 대한 기초지방자치단체 j 의 단위면적당 피해액(천원/ha)

$\ln D_{\min}$: 모든 j 와 t 에 대한 $\ln D_{jt}$ 중 최솟값

$\ln D_{\max}$: 모든 j 와 t 에 대한 $\ln D_{jt}$ 중 최댓값

2. 모형 수립 및 계수 추정

앞서 정의된 독립변수와 종속변수를 이용하여, 연도 t 에 대한 기초지방자치단체 j 의 자연재해 피해액 추정을 위한 회귀분석 모형을 수립하면 다음과 같다.

$$d_{jt} = \beta_0 + \sum_{i=1}^{14} \beta_i x_{ijt} + \epsilon_{jt} \quad (\text{식 4})$$

- β_i : 독립변수 x_{ijt} 에 대한 회귀계수, $i = 0, 1, 2, \dots, 14$, 단, β_0 는 절편
- ϵ_{jt} : 오차항

위의 모형을 바탕으로, d_{jt} 의 추정치 \hat{d}_{jt} 과 D_{jt} 의 추정치 \hat{D}_{jt} 은 회귀계수 β_i 의 추정치 b_i 를 이용하여 다음과 같이 계산될 수 있다. <Table 2>는 회귀계수에 대한 추정치 b_i 를 보여주고 있다.

$$\hat{d}_{jt} = b_0 + \sum_{i=1}^{32} b_i x_{ijt} \quad (\text{식 5})$$

$$\hat{D}_{jt} = e^{\ln D_{\min} + \hat{d}_{jt}(\ln D_{\max} - \ln D_{\min})} \quad (\text{식 6})$$

Table 2. Regression coefficients for the regression model

Number	Independent Variable	Regression Coefficient
0	Intercept	-0.5551
1	Heavy Rain Frequency	0.1528
2	Heavy Rain Strength	0.6130
3	Typhoon Frequency	0.1223
4	Typhoon Strength	0.5363
5	Heavy Snow Strength	0.2997
6	Heavy Snow Frequency	0.3205
7	Heavy Wind and Storm Strength	0.1857
8	River Density	0.1014
9	Ship Density	0.2882
10	Population Density	0.0863
11	Higher Education Ratio	0.1563
12	Foreigner Ratio	0.1834
13	Disaster Prevention Budget Ratio	0.3424
14	Population per Fireman	0.2534

III. 인공신경망 모형

1. 네트워크 구조

인공신경망은 <Figure 3>과 같이 크게 입력층, 은닉층, 출력층으로 구성된 계층적 구조를 갖는다. 입력층 노드의 값, 입력층과 은닉층 간의 연결 가중치, 은닉층 편향(Bias)을 이용하여 은닉층 노드의 값을 결정하고, 이후 은닉층 노드의 값, 은닉층과 출력층 간의 연결 가중치, 출력층 편향을 기반으로 출력층 노드의 값이 결정된다. 이때 입력층과 출력층 간 상관관계의 복잡도에 따라 <Figure 3>의 a)와 b)와 같이 은닉층을 한층 또는 두층으로 구성할 수 있다.

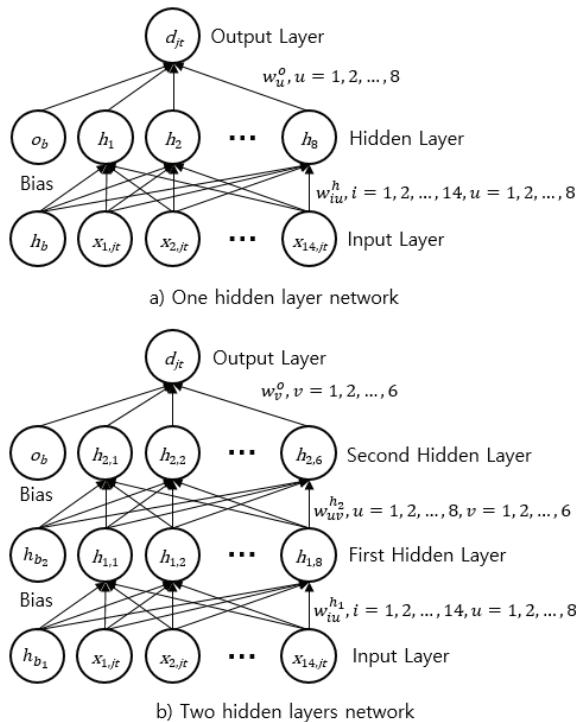


Figure 3. Artificial neural network structure

2. 활성화 함수

인공신경망의 은닉층과 출력층에 위치한 노드에서는 입력값을 바탕으로 출력값을 계산한다. 이때 사용하는 함수가 바로 활성화 함수이다. 활성화 함수로는, 하위 계층 노드의 값, 연결 가중치, 편향을 이용하여 계산된 가중합을 그대로 출력하는 항등함수(Identity Function)를 이용하거나, 계산된 가중합에 대한 함수

변환을 통해 출력값을 산출하는 S자 모양의 시그모이드 함수(Sigmoid Function)를 이용한다. 시그모이드 함수로는 표준 로지스틱 함수(Standard Logistic Function)와 쌍곡 탄젠트 함수(Hyperbolic Tangent Function)가 주로 이용된다. <Figure 4>에 나타나 있는 바와 같이, 표준 로지스틱 함수는 중앙점이 0이고, 최솟값과 최댓값이 각각 0과 1이며, 로지스틱 증가율이 1인 로지스틱 함수이며, 쌍곡 탄젠트 함수는 중앙점이 0이고, 최솟값과 최댓값이 각각 -1과 1인 쌍곡선 함수이다. 본 연구에서 사용하고 있는 활성화 함수의 정의는 다음과 같다.

항등함수 : $f(z) = z$

표준 로지스틱 함수 : $f(z) = 1/(1 + e^{-z})$

쌍곡 탄젠트 함수 : $f(z) = (e^z - e^{-z}) / (e^z + e^{-z})$

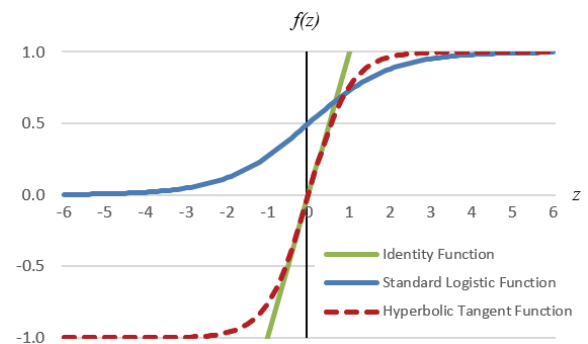


Figure 4. Shape of activation functions

3. 인공신경망 모형

앞서 정의한 네트워크 구조와 활성화 함수를 이용하면, 인공신경망 모형을 수학적으로 정의할 수 있다. 이때, 인공신경망의 수학적 정의란 입력층의 노드 값으로 기초지방자치단체 j 에 대한 연도 t 의 표준화된 독립변수 i 의 값, x_{ijt} 를 입력받아, 연도 t 에 대한 기초지방자치단체 j 의 표준화된 피해액 d_{jt} 를 출력층의 노드 값으로 계산하는 과정을 의미한다. 이 과정은 인공신경망을 구성하는 은닉층의 개수에 따라 다음과 같이 정의될 수 있다. 본 연구에서는 <Figure 3>과 같이 예비 실험을 통해, 1개 은닉층 네트워크에 대해서는

은닉층에 8개 노드를, 2개 은닉층 네트워크에 대해서는 첫 번째 은닉층에 8개 노드, 두 번째 은닉층에 6개 노드를 배치하였다.

One hidden layer network :

$$d_{jt} = f_o(o_b + \sum_{v=1}^8 w_v^o f_h(h_b + \sum_{i=1}^{14} w_{iv}^h x_{ijt}))$$

- w_{iv}^h : 입력층 노드와 은닉층 노드를 연결하는 가중치
- w_u^o : 은닉층 노드와 출력층 노드를 연결하는 가중치
- h_b : 은닉층 편향
- o_b : 출력층 편향
- f_h : 은닉층 노드에 대한 활성화 함수
- f_o : 출력층 노드에 대한 활성화 함수

Two hidden layers network :

$$d_{jt} = f_o(o_b + \sum_{v=1}^6 w_v^o f_{h_2}(h_{b_2} + \sum_{u=1}^8 w_u^{h_2} f_{h_1}(h_{b_1} + \sum_{i=1}^{14} w_{iu}^{h_1} x_{ijt})))$$

- $w_{iu}^{h_1}$: 입력층 노드와 첫 번째 은닉층 노드를 연결하는 가중치
- $w_{uv}^{h_2}$: 첫 번째 은닉층 노드와 두 번째 은닉층 노드를 연결하는 가중치
- w_v^o : 두 번째 은닉층 노드와 출력층 노드를 연결하는 가중치
- h_{b_1} : 첫 번째 은닉층 편향
- h_{b_2} : 두 번째 은닉층 편향
- f_{h_1} : 첫 번째 은닉층 노드에 대한 활성화 함수
- f_{h_2} : 두 번째 은닉층 노드에 대한 활성화 함수

4. 모형 비교 분석

인공신경망은 은닉층의 개수, 은닉층과 출력층에서 사용하는 활성화 함수의 종류에 따라 다양한 형태로 정의될 수 있다. 본 연구에서는 <Table 3>에 정의되어 있는 바와 같이, 은닉층의 개수를 1층과 2층으로, 은닉층에서 사용하는 활성화 함수를 표준 로지스틱 함수, 쌍곡 탄젠트 함수의 두 종류로, 출력층에서 사용하는 활성화 함수를 항등함수, 표준 로지스틱 함수, 쌍곡 탄젠트 함수의 세 종류로 하여, 총 12개의 실험 케이스를 정의하고 이를 비교·분석하여 자연재해 피해액 예측에 가장 적합한 모형을 선정하고자 한다.

본 연구에서는 인공신경망의 학습을 위해서 IBM SPSS Statistics 25에서 제공하는 다층 퍼셉트론 분석 기능을 이용하였다. 모든 훈련 데이터 레코드를 전달한 후에만 시냅스 가중치를 갱신하는 배치 방식의 훈련유형을 이용하였으며, 스케일링된 공역 구배법을 이용하여 최적화하였다. 학습을 위한 매개변수인 람다와 시그마의 초기값은 각각 0.0000005와 0.00005를, 구간 중심 및 구간 오프셋은 각각 0과 ±0.5를 이용하였다. 또한, 학습 정지 규칙을 정의하기 위해 오차 감소가 없는 최대 단계 수로는 1, 최대 훈련 시간으로는 15분, 훈련 오차의 최소 상대 변화량은 0.0001, 훈련 오차 비율의 최소 상대 변화량으로는 0.001을 이용하였다.

인공신경망의 특성 상, 학습 과정에 무작위성이 개입되므로 동일한 모형이라해도 학습의 결과, 즉, 가중

Table 3. Cases for alternative artificial neural networks

Case	Number of hidden layers	Activation function for hidden layer	Activation function for output layer
1	1	Hyperbolic tangent	Identity
2	1	Hyperbolic tangent	Hyperbolic tangent
3	1	Hyperbolic tangent	Standard logistic
4	1	Standard logistic	Identity
5	1	Standard logistic	Hyperbolic tangent
6	1	Standard logistic	Standard logistic
7	2	Hyperbolic tangent	Identity
8	2	Hyperbolic tangent	Hyperbolic tangent
9	2	Hyperbolic tangent	Standard logistic
10	2	Standard logistic	Identity
11	2	Standard logistic	Hyperbolic tangent
12	2	Standard logistic	Standard logistic

Table 4. Results of t-test for comparing correlation coefficients among 12 cases

Case	Average ρ	p-values between two cases										
		3	8	2	7	4	6	1	12	10	11	5
9	0.8301	0.799	0.144	0.050*	0.010**	0.014*	0.007**	0.002**	0.000**	0.000**	0.000**	0.004**
3	0.8298		0.181	0.059	0.010**	0.015*	0.007**	0.002**	0.000**	0.000**	0.000**	0.040*
8	0.8276			0.762	0.324	0.333	0.253	0.069	0.002**	0.001**	0.001**	0.019*
2	0.8271				0.402	0.416	0.303	0.070	0.001**	0.000**	0.001**	0.022*
7	0.8260					0.952	0.825	0.233	0.004**	0.001**	0.002**	0.040*
4	0.8260						0.889	0.291	0.012	0.003**	0.004**	0.044*
6	0.8258							0.317	0.008**	0.001**	0.003**	0.047*
1	0.8243								0.221	0.117	0.031	0.113
12	0.8226									0.652	0.100	0.227
10	0.8223										0.118	0.254
11	0.8198											0.711
5	0.8184											

ρ : Correlation coefficient between estimated and actual amounts of damages

* : Two means are different under significant level of 0.05

** : Two means are different under significant level of 0.01

치와 편향은 매번 다르게 나타난다. 따라서 본 연구에서는 이러한 무작위성이 결과에 영향을 미치는 것을 배제하기 위해 12개 실험 케이스 각각에 대해, 서로 다른 학습과정을 수행하여 가중치와 편향을 추정한 후, 이들을 이용하여 예측하는 과정을 10번씩 반복하여 추정치를 산출하고, 이를 바탕으로 실험 케이스 간 비교·평가를 수행하였다. 즉, 12개의 실험 케이스 각각에 대해, 인공신경망을 이용하여 자연재해 피해액을 10번씩 반복 추정한 후, 이 추정값들과 실제 피해액 간에서 산출된 10개의 상관계수를 이용하여 케이스들의 예측 성능을 비교·분석하였다. 상관계수의 차이를 검증하기 위해, 12개 실험 케이스를 2개씩 조합하여 수행한 66(12C2)번의 t-검정 결과가 <Table 4>에 정리되어있다. 검정 결과, 실험 케이스 3, 8, 9가 가장 높

은 상관계수 값을 제공함을 알 수 있었다.

다음으로는 은닉층의 개수, 은닉층 활성화 함수의 종류, 출력층 활성화 함수의 종류 중 어떤 요인이 자연재해 피해액 추정을 위한 인공신경망 모형의 예측력, 즉, 예측 피해액과 실제 피해액 간의 상관계수 값에 영향을 미치는가를 파악하기 위한 분산분석을 수행하였다. <Table 5>에 나타나 있는 바와 같이, 은닉층의 개수는 상관계수의 값에 영향을 미치지 못하는 반면, 은닉층과 출력층의 활성화 함수 종류는 인공신경망 모형의 예측력에 큰 영향을 미치는 것을 알 수 있다.

분산 분석의 결과를 바탕으로 은닉층과 출력층의 활성화 함수로서 어떤 함수를 사용하는 것이 가장 적합한 것인가를 밝히기 위한 분석이 수행되었다. 12개 실험 케이스에서 10번씩 반복하여 얻어진 예측 피해액과 실

Table 5. Results of analysis of variance (ANOVA)

Source	Sum of Square	Degree of freedom	Mean square	F	p-value
Intercept	81.670	1	81.670	3,891,000	0.000
Number of hidden layers	0.000	1	0.000	0.298	0.586
Activation function for hidden layer	0.001	1	0.001	35.639	0.000**
Activation function for output layer	0.000	2	0.000	7.287	0.001**
Error	0.002	115	0.000		
Total	81.674	120			

** : This factor affects the values of correlation coefficients under significant level of 0.01

Table 6. Results of t-test for comparing grouping correlation coefficients

Number of hidden layers	Average ρ	2	
1	0.8252	p-value : 0.645	
2	0.8247		
Activation function for hidden layer	Average ρ	Standard logistic	
Hyperbolic tangent	0.8275	p-value : 0.000**	
Standard logistic	0.8225		
Activation function for output layer	Average ρ	Identity	Hyperbolic tangent
Standard logistic	0.8271	p-value : 0.040*	p-value : 0.050*
Identity	0.8247		p-value : 0.255
Hyperbolic tangent	0.8232		

ρ : Correlation coefficient between estimated and actual amounts of damages
 * : Two means are different under significant level of 0.05
 ** : Two means are different under significant level of 0.01

제 피해액 간의 상관계수 총 120개를 은닉층과 출력층의 활성화 함수의 종류별로 집단화한 후 각 집단의 평균을 비교하기 위한 t-검정을 수행하였다. <Table 6>에 나타나 있는 검정 결과와 같이, 은닉층의 개수에 대해서는 집단 간 차이가 발생하지 않았으며, 은닉층과 출력층 활성화 함수의 종류에 대해서는 집단 간 차이가 발생하였다. 즉, 은닉층 함수는 쌍곡 탄젠트 함수일 때, 출력층 함수는 표준 로지스틱 함수일 때, 인공신경망 모형의 예측력이 최대가 되는 것을 알 수 있었다.

요인이 3개 이상인 출력층 함수에 대해서는 추가적으로 분산분석의 사후검정인 Scheffe 검정, LSD (Least

Significant Difference) 대응별 다중 비교 검정을 수행하였다. <Table 7>에 나타나 있는 바와 같이 이들 분석 결과 역시 분산분석과 t-검정의 결과와 동일하게 출력층 함수가 표준 로지스틱 함수일 때 가장 좋은 예측력을 보이는 것으로 나타났다. 또한, Duncan 다중 범위 검정 분석 결과, 출력층 함수는 항등 함수와 쌍곡 탄젠트 함수 집단과 표준 로지스틱 함수 집단으로 양분되는 것을 확인할 수 있었다. 이들 결과를 종합하면, 사후분석 역시 출력층 함수로서 표준 로지스틱 함수를 사용하는 것이 가장 적합함을 보여주고 있다는 것을 알 수 있다.

Table 7. Results of post-hoc tests for comparing grouping correlation coefficients

Test	Activation function for output layer (I)	Activation function for output layer (J)	Difference (I-J)	Standard error	p-value
Scheffe	Identity	Hyperbolic tangent	0.001	0.001	0.372
		Standard logistic	-0.002	0.001	0.065
	Hyperbolic tangent	Identity	-0.001	0.001	0.372
		Standard logistic	-0.004	0.001	0.001**
	Standard logistic	Identity	0.002	0.001	0.065
		Hyperbolic tangent	0.004	0.001	0.001**
LSD	Identity	Hyperbolic tangent	0.001	0.001	0.160
		Standard logistic	-0.002	0.001	0.020*
	Hyperbolic tangent	Identity	-0.001	0.001	0.160
		Standard logistic	-0.004	0.001	0.000**
	Standard logistic	Identity	0.002	0.001	0.020*
		Hyperbolic tangent	0.004	0.001	0.000**

* : Two means are different under significant level of 0.05
 ** : Two means are different under significant level of 0.01

5. 최적 인공신경망 모형

앞서 진행된 모형 비교 분석의 결과에 의하면, 1개의 은닉층과 2개의 은닉층을 갖는 케이스가 예측력에 있어서 유의미한 차이를 보이지 않으므로, 모형 간소화를 위해 단순한 형태의 단일 은닉층 네트워크 구조를 사용하는 것이 보다 바람직하다고 할 수 있다. 또한, 활성화 함수에 대해서는, 상대적으로 더 우수한 성능을 보인 함수들, 즉, 은닉층 활성화 함수로는 쌍곡 탄젠트 함수를 출력층 활성화 함수로는 표준 로지스틱 함수를 이용하는 것이 인공신경망 모형의 예측력을 보다 향상시킬 수 있다. 이들 실험 결과를 토대로, 본 연구에서 제안하는 인공신경망 모형은 다음과 같다.

$$d_{jt} = f_o(o_b + \sum_{v=1}^8 w_{uv}^o f_h(h_b + \sum_{i=1}^{14} w_{iu}^h x_{ijt})) \quad (\text{식 7})$$

$$f_h(z) = 1/(1 + e^{-z}), f_o(z) = (e^z - e^{-z})/(e^z + e^{-z})$$

최적 인공신경망 모형을 이용하면, d_{jt} 의 추정치 \hat{d}_{jt} 을 계산하기 위한 추정식을 다음과 같이 정의할 수 있다. 가중치와 편향에 대한 추정치는 기계 학습 과정에

서 사용한 난수에 따라 매번 다르게 결정될 수 있으며, 학습 과정을 통해 얻은 추정치는 <Table 8>에 예시되어 있는 바와 같은 형태이다.

$$\hat{d}_{jt} = f_o(\hat{o}_b + \sum_{v=1}^8 \hat{w}_{uv}^o f_h(\hat{h}_b + \sum_{i=1}^{14} \hat{w}_{iu}^h x_{ijt})) \quad (\text{식 8})$$

\hat{w}_{iu}^h : 입력층 노드와 은닉층 노드를 연결하는 가중치 w_{iu}^h 에 대한 추정치

\hat{w}_{uv}^o : 은닉층 노드와 출력층 노드를 연결하는 가중치 w_{uv}^o 에 대한 추정치

\hat{h}_b : 은닉층 편향 h_b 에 대한 추정치

\hat{o}_b : 출력층 편향 o_b 에 대한 추정치

IV. 모형 간 자연재해 피해액 예측 성능 비교 분석

1. 예측 성능 비교 분석

회귀분석과 인공신경망 모형의 예측 성능을 비교·분석하기 위해, 2001년부터 2014년까지 각 연도 t 및 지역 j 에 대한 자연재해 피해액 추정치 \hat{d}_{jt} 를 각각의 모형을 이용하여 예측하였다. 하나의 회귀계수 추정치를 갖는 회귀분석 모형과 달리 인공신경망 모형

Table 8. An Example set of estimators for the biases and weights

Hidden layer node		1	2	3	4	5	6	7	8	
Hidden layer	Bias	0.292	0.149	0.124	-0.418	-0.003	-0.920	0.579	0.017	
	Weight	Heavy Rain Frequency	-0.200	0.263	0.790	-0.788	0.137	-0.276	-0.792	-2.103
		Heavy Rain Strength	-0.118	1.029	0.185	-0.370	-0.210	0.756	-1.142	-1.839
		Typhoon Frequency	0.745	0.350	0.107	0.036	-0.507	-0.064	-0.923	-1.604
		Typhoon Strength	0.901	0.066	0.721	-0.626	0.217	0.624	-1.632	-2.084
		Heavy Snow Strength	0.007	0.715	0.114	-0.275	0.589	0.165	-2.579	-2.681
		Heavy Snow Frequency	-0.074	-0.108	0.178	-0.304	0.336	-0.137	-1.884	-2.143
		Heavy Wind and Storm Strength	0.521	0.623	-0.042	-0.883	-0.016	0.124	-1.356	-1.815
		River Density	0.127	-0.067	-0.561	0.142	-0.323	0.130	-0.094	0.174
		Ship Density	-0.346	0.463	-0.016	0.221	0.739	1.319	-0.221	0.347
		Population Density	0.314	0.141	0.337	-0.401	-1.370	-0.030	-0.765	-0.352
		Higher Education Ratio	0.322	0.117	0.642	0.464	-0.686	0.597	0.434	0.077
		Foreigner Ratio	0.518	0.303	0.685	0.412	-0.258	0.470	-0.309	0.223
		Disaster Prevention Budget Ratio	-0.045	-0.322	0.605	0.014	0.272	0.100	-0.455	-0.293
Population per Fireman	0.524	0.381	-0.042	-0.210	0.029	0.336	0.557	0.004		
Output layer	Bias	-1.722								
	Weight	-1.354	-0.223	-1.075	0.347	1.249	1.355	-2.598	-3.137	

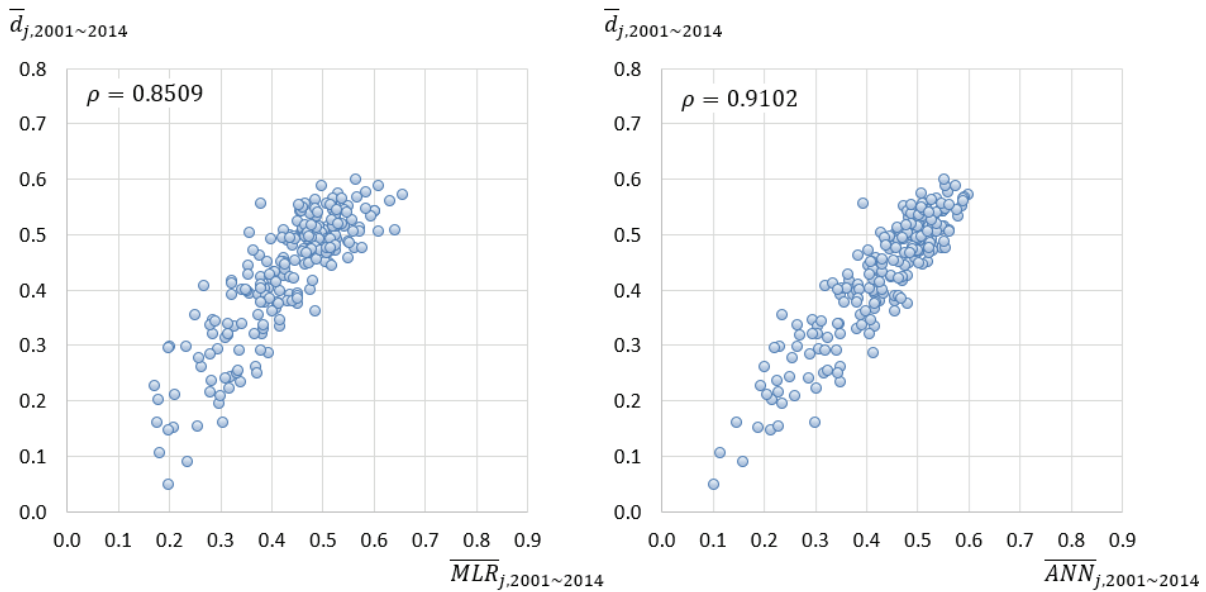


Figure 5. Scatter diagram for MLR and ANN models for years 2001 to 2014

은 어떤 난수 시드(Random Number Seed)를 사용했는가에 따라 가중치 및 편향에 대한 추정치가 다양하게 결정될 수 있다. 이러한 이유로 인공지능망 모형에 대해서는 30회의 학습과정을 반복 수행한 후, 서로 다른 가중치 및 편향을 갖는 30개 모형들로부터 얻은 추정치들의 평균을 산출하여 분석하였다.

회귀분석과 인공지능망 각각의 모형을 이용한 경우에 대해, 14년 동안의 피해액 추정치 평균($\overline{MLR}_{j,2001 \sim 2014}$, $\overline{ANN}_{j,2001 \sim 2014}$)과 실제 피해액의 평균($\overline{a}_{j,2001 \sim 2014}$)의 상관관계를 산점도를 이용하여 표출하면 <Figure 5>와 같다. 산점도에 나타나 있는 바와 같이, 인공지능망 모형이 회귀분석 모형에 비해 좁은 폭의 띠를 형성함으로써 자연재해 피해액을 보다 정확하게 예측하고 있음을 관찰할 수 있다. 구체적으로 살펴보면, 인공지능망 모형의 추정치와 실제 피해액과의 상관계수가 0.9102로 회귀분석 모형의 추정치에 대한 상관계수 0.8509에 비해 7% 이상 향상되었음을 알 수 있다.

또한, 2011년부터 2014년까지에 대해, 지역별 자연재해위험지표 (NDRI: Natural Disaster Risk Index) 평가치, 지역안전도 (RSG: Regional Safety Grades) 등급, 회귀분석 모형 (MLR: Multiple Linear Regression) 추정치, 인공지능망 모형 (ANN: Artificial Neural Network) 추

정치를 각각 산출하여 실제 자연재해 피해액과의 산점도를 <Figure 6>과 같이 작성하였다(Ministry of Public Safety and Security, 2015; Yeon, *et. al.*, 2016; Jeong, 2019). <Figure 6>은 3개년의 평균을 이용한 산점도로서, 14년 동안의 평균을 이용한 산점도인 <Figure 5>에 비해 평균화로 인한 분산 감소 현상이 둔화됨으로 인해, 띠의 폭이 다소 넓어진 것을 관찰할 수 있다. 이 경우에도 역시 인공지능망 모형이 다른 방법들에 비해 낮은 폭의 띠를 형성함으로써 자연재해 피해액을 보다 정확하게 예측하고 있음을 알 수 있다. 즉, 인공지능망 모형의 추정치와 실제 피해액과의 상관계수가 0.8310으로, 자연재해위험지표, 지역안전도, 회귀분석 모형에 대한 추정치와의 상관계수 0.4617, 0.4699, 0.7201에 비해, 적게는 15%부터 많게는 80% 이상 상관도가 향상되었다.

<Table 9>는 자연재해위험지표 평가치, 지역안전도 등급, 회귀분석 모형 추정치, 인공지능망 모형 추정치, 실제 피해액 간의 상관계수를 보여주고 있다. 회귀분석 모형과 인공지능망 모형은 실제 동일한 독립변수를 공유함으로써 인해 다른 방법론들에 비해 그 상관관계가 0.9059로 높게 나타나는 것을 볼 수 있다.

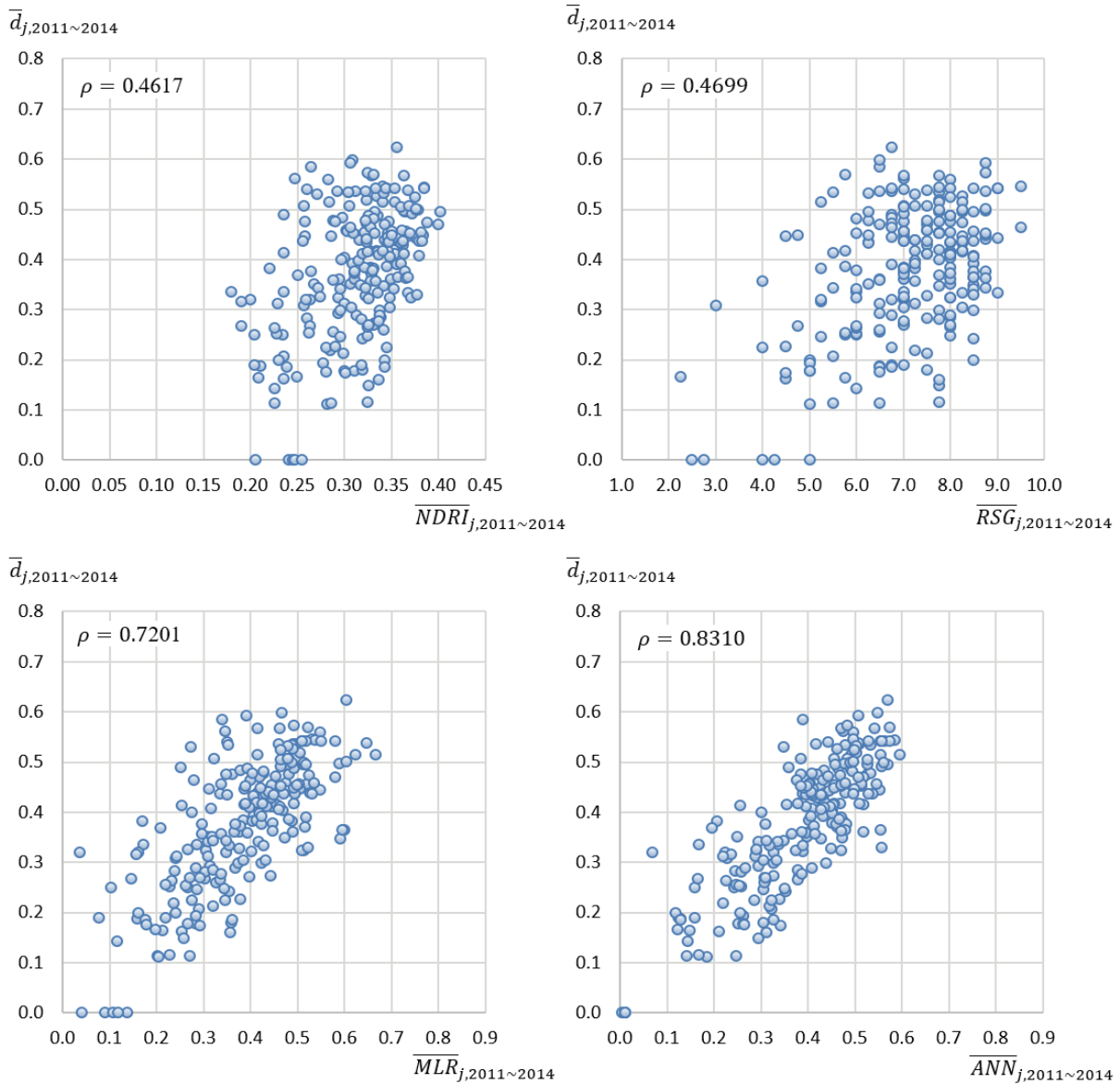


Figure 6. Scatter diagram for NDRI, RSG, MLR, and ANN model for years 2011 to 2014

Table 9. Correlation coefficients among the forecasting methodologies

	NDRI	RSG	MLR	ANN
$\bar{d}_{j,2011\sim 2014}$	0.4617*	0.4699*	0.7201*	0.8310*
NDRI	1	0.5760*	0.7867*	0.7241*
RSG		1	0.5375*	0.5395*
MLR			1	0.9059*

* p-value is 0

2. 독립변수 가중치 비교분석

본 절에서는 높은 상관관계를 보였던 회귀분석 모형과 인공신경망 모형의 각 독립변수들에 대한 중요도를 나타내는 가중치를 비교·분석하였다. 이 가중치들은 자연재해 위험지표 모형의 세부지표에 적용되

는 중요도 가중치와 유사한 의미로 해석될 수 있다. 이들 가중치가 <Table 10>에 나타나 있다. 회귀분석 모형에 대해서는 독립변수들의 표준화 회귀계수를 추정 후 계수의 합이 1이 되도록 조정된 수치를 표기하였고, 인공신경망 모형에 대해서는 민감도 분석을

Table 10. Weights for the independent variables

Variable	Weight	
	ANN	MLR
Heavy Rain Strength	0.2643	0.3995
Typhoon Strength	0.2634	0.3428
Heavy Snow Frequency	0.1951	0.1643
Ship Density	0.1638	0.0520
Heavy Snow Strength	0.1535	0.1003
Heavy Rain Frequency	0.1344	0.0854
Heavy Wind and Storm Strength	0.1331	0.1162
Typhoon Frequency	0.1246	0.0743
Disaster Prevention Budget Ratio	0.0808	0.0673
Foreigner Ratio	0.0686	0.0782
Population Density	0.0607	0.0646
Higher Education Ratio	0.0438	0.1073
Population per Fireman	0.0380	0.0656
River Density	0.0376	0.0439

수행한 후 회귀분석 모형과 마찬가지로 가중치의 합이 1이 되도록 조정한 수치를 표기하였다.

두 가중치의 차이를 좀 더 분명하게 나타내기 위해 <Figure 7>과 같은 형태의 그래프를 작성하였다. 그래프에서 확인할 수 있는 바와 같이, 회귀분석 모형과 인공신경망 모형에서 얻어진 가중치는 전반적으로 유사한 패턴을 보이면서도 개별적인 가중치의 값에서는 차이가 나타났다. 이 차이는 두 모형이 서로 다른 예측 결과를 만들어내게 된 표면적인 원인으로 해석될 수 있다. 그러나 보다 근본적으로는 독립변수와 종속변

수의 상호관계에 대한 선형성 가정이 필요 없는 인공신경망 모형만의 유연성, 즉, 비선형성에 대한 처리 능력이 결국 더 우수한 피해액 예측치를 제공한 동인으로 작용한 것이라 말할 수 있다.

V. 결론

본 연구에서는 공공부문 데이터베이스로부터 획득할 수 있는 14개 정량 자료를 기반으로 자연재해 피해액을 예측할 수 있는 인공신경망 모형을 제안하였다. 분석 결과, 자연재해 피해액 예측에 있어서 이전 연구에서 제안하였던 회귀분석 모형보다 더 우수한 예측력을 나타냄을 알 수 있었다. 즉, 인공신경망 모형의 자연재해 피해액에 대한 설명계수가 $0.8284(=0.9102^2)$ 로 회귀분석 모형의 설명계수 $0.7240(=0.8509^2)$ 보다 14.4% 이상 향상되었다. 이는 독립변수와 종속변수 간의 비선형적인 관계를 종속변수의 로그변환을 통해 제한적으로 처리했던 회귀분석 모형에 비해, 독립변수와 종속변수 사이에 은닉층 변수들을 포함시켜 비선형적 관계를 모사할 수 있는 인공신경망 모형만의 유연성이 자연재해 피해액 예측에 더 높은 정확도를 제공했기 때문이다.

본 연구는 자연재해로 인한 피해액을 예측할 수 있는 매우 효과적이고 효율적인 수단을 확보했다는 측

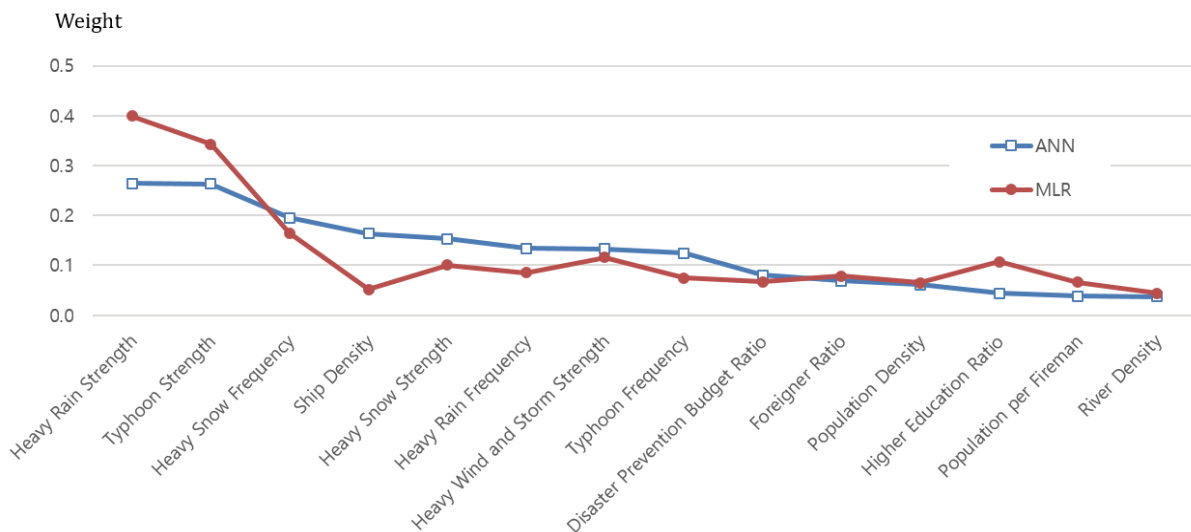


Figure 7. Weights for the independent variables

면에서 그 의미를 가진다. 향후, 기후 및 사회 변화에 따른 독립변수들의 미래 예측을 바탕으로 제안된 피해액 예측 모형을 활용한다면, 미래의 자연재해 피해가 어떤 형태로 나타날 것인가를 사전에 예측해 볼 수 있을 것이다.

감사의 글

이 논문은 행정안전부의 방재안전 분야 전문인력 양성사업의 지원을 받아 제작되었으며, 2018년도 정부(교육부)의 재원으로 한국 연구재단의 지원을 받아 수행된 기초연구사업(No. 2018R1D1A1B07043804)임.

References

- Disaster and Safety Management Basic Law. 2004. Amendment 2018.9.18. Law No. 13440.
- Jeong, Keun-Chae. 2019. Estimation Model for Damage Amounts of Natural Disasters Based on Regression Analysis Using Public Data. *Crisisonomy*. 15(3): 55-69.
- Kim, Keum Ji and Seong Hwan Yoon. 2017. Assessment of Natural Disaster Damage Using Weather Observation Data: Using Multiple Regression Analysis and Artificial Neural Network Analysis. *Journal of the Korean Society of Hazard Mitigation*. 17(5): 57-65.
- Ministry of Interior and Safety. 2002-2019. *2001~2018 Statistical Yearbook of Natural Disaster*.
- Ministry of Public Safety and Security. 2015. Regional Safety Diagnosis Results for Years 2011~2014.
- National Emergency Management Administration. 2011. *2011 Regional Safety Diagnosis Manual*.
- Rim, Jemli, Chtourou Nouri, and Feki Rochdi. 2010. Insurability Challenges under Uncertainty: An Attempt to use the Artificial Neural Network for the Prediction of Losses from Natural Disasters. *Panoeconomicus*. 57(1): 43-60.
- Santoso, Albertus Joko, Findra Kartika Sari Dewi, and Thomas Adi Purnomo Sidhi. 2015. Natural Disaster Detection Using Wavelet and Artificial Neural Network. *Science and Information Conference (SAI) 2015*. 761-764.
- Wan, Yu Qiong and Jun You Liang. 2013. A Nonlinear Heuristic-optimized Neural Network Modeling Approach for the Prediction of Natural Disasters. *Advanced Materials Research: AMR*. 694-697: 1310-1316.
- Yeon, Ho-Jae and Keun-Chae Jeong. 2018. Statistical Analysis for the Relationship among Frequency, Intensity, and Damage Amount of Natural Hazard. *Crisisonomy*. 14(3): 149-156.
- Yeon, Ho-Jae, Keun-Chae Jeong, and Tae-Young Heo. 2016. A Quantitative Data based Natural Disaster Risk Index Model in South Korea. *Crisisonomy*. 12(3): 169-186.
- [1] <http://sts.kma.go.kr>. National Climate Data Service System.
- [2] <http://typ.kma.go.kr>. National Typhoon Center.
- [3] <https://www.ibm.com/support/knowledgecenter>. IBM SPSS Manual
- [4] <http://www.kma.go.kr>. Korea Meteorological Administration.
- [5] <http://www.kosis.kr>. Korean Statistical Information Service.
- [6] <http://www.safekorea.go.kr>. National Disaster Safety Portal.
- Korean References Translated from the English*
- 국민안전처. 2015. 2011년~2014년 지역안전도 진단 현황.
- 김금지, 윤성환. 2017. 기상 관측자료를 이용한 자연재난 피해 영향평가: 다중회귀분석과 인공신경망분석을 이용하여. *한국방재학회논문집*. 17(5): 57-65.
- 소방방재청. 2011. 2011년 지역안전도 진단지침.
- 연호재, 정근채, 허태영. 2016. 정량자료 기반의 한국형 자연재해 위험지표 모형. *Crisisonomy*. 12(3): 169-186.
- 연호재, 정근채. 2018. 자연재해 빈도 및 강도와 피해액 상호관계에 대한 통계적 분석. *Crisisonomy*. 14(3): 149-156.
- 재난 및 안전관리 기본법. 2004. 개정 2018.9.18. 법률 제13440호.
- 정근채. 2019. 공공 데이터를 이용한 회귀분석 기반 자연재해 피해액 추정 모형. *Crisisonomy*. 15(3): 55-69.
- 행정안전부. 2002-2019. 2001년~2018년 재해연보. 행정안전부.

인공신경망 기반 자연재해 피해액 추정 모형

국문초록 기존에 공공 데이터베이스 정보를 이용하여 자연재해로 인한 피해액을 추정하기 위한 회귀분석 모형 개발 연구가 진행되었다. 그러나 이모형은 종속변수의 로그변환을 통해 독립변수와 종속변수 간의 비선형성을 반영했음에도 불구하고, 독립변수와 종속변수 간의 선형관계를 가정하는 회귀분석 모형의 고유한 특성으로 인해 예측 정확도 향상에 한계점을 나타냈다. 이에 본 연구에서는 독립변수와 종속변수 간의 선형성 가정이 요구되지 않는 인공신경망 모형을 기반으로 자연재해 피해액을 좀 더 정확하게 예측할 수 있는 모형을 제시한다. 제안된 모형의 유효성을 검증하기 위해, 인공신경망 모형을 이용한 피해액 예측치, 회귀분석 모형을 이용한 피해액 예측치, 자연재해 위험지표 평가결과, 지역안전도 등급, 자연재해로 인한 실제 피해액을 비교·분석하였다. 분석 결과, 인공신경망을 이용하여 도출된 추정치가 기존의 회귀분석을 이용하여 도출된 추정치뿐만 아니라 자연재해 위험지표와 지역안전도 평가결과에 비해서도 실제 자연재해 피해액과 더 높은 상관관계, 즉, 더 높은 예측력을 나타내는 것을 확인할 수 있었다.

주제어 : 자연재해, 피해액, 인공신경망, 예측, 비선형성

Profiles **Keun-Chae Jeong** : He received B.S. in Industrial Engineering from the Korea University in 1991. He received M.S. and Ph.D. in Industrial Engineering from the Korea Advanced Institute of Science and Technology (KAIST). He was with Department of Management Information Systems at the Jeju National of University as a full-time lecturer during 2000-2001 and with LG CNS as a senior consultant during 1997-2000. He is a professor of School of Civil Engineering at the Chungbuk National University, in which he has taught since 2001. His recent research area includes Disaster Economic Analysis, Decision Support Systems, Ubiquitous Construction, and Production Planning and Scheduling(kcjeong@cbnu.ac.kr).