

신경망과 유전자 알고리즘을 이용한 자연재해 피해예측 모델 연구

최선화[○]

소방방재청 방재연구소

shchoi333@nema.go.kr

Natural Disaster Damage Cost Prediction Model based on Neural Network and Genetic Algorithm

Seonhwa Choi[○]

National Institute for Disaster Prevention

요 약

기후온난화, 국지성 호우 및 대규모 태풍으로 인한 피해가 증대되면서 사회·경제적 손실 또한 날로 증가하고 있어 재해로 인한 피해 발생가능성을 효율적으로 예측하는 모델을 통한 선제적 대응이 필요하다. 재난·재해의 위험성 분석 방법은 주로 확률·통계기법을 기반으로 하는 연구가 주류를 이루었으나, 본 논문에서는 포착된 현상의 데이터를 이용해 그 데이터를 지배하는 경험적 규칙성을 학습하고 획득하는데 다른 기법보다 탁월한 성능을 가진 신경망 모델을 적용하여 자연재해 피해예측 모델을 연구하였다. 1991년부터 2005년 사이에 우리나라에서 발생한 자연재해의 피해자료와 기상개황 자료를 이용하여 지역별 자연재해로 인한 피해를 예측하는 신경망 모델은 우리나라 232개 행정구역에 대하여 누적강우량과 최대풍속, 그리고 재해사상 발생 5일 이내의 선행강우량을 입력변수로 하고 총 피해액을 출력변수로 한다. 또한 학습을 통한 최적의 해를 찾기 위해 신경망의 매개변수 학습률, 모멘텀, 편의값을 유전자 알고리즘으로 결정하여 학습을 수행 하였다.

1 서 론

미래의 재난·재해로 인한 위험도를 평가하기 위해서는 그 시점에서의 위험에 대한 정보 즉, 재해유형에 따른 피해규모별 재난발생빈도와 발생확률, 이로 인한 사회적·경제적 파급 등을 고려한 사회적 위험지수와 같은 재난발생 양상을 정략적으로 나타낼 수 있는 재난위험 지표가 필요하다. 따라서 재난위험 지표 개발 및 위험지표 별 지수 거동을 신뢰성 있게 분석하는 기술개발이 요구된다. 재난선진국에서는 여러 가지 방법을 통하여 자연재해에 대한 위험평가를 실시하여 그 효용성을 입증하고 있다. 국내에서도 HAZUS 등을 활용한 지진 취약도 함수개발[1]이나 지역안전도 평가 시스템 구축[2] 등을 통해 재난관리에 활용되고 있다.

전통적으로 재해로 인한 위험성 분석 및 피해예측에는 결정론적 접근 방법이 아닌 과거의 자료를 기반으로 하는 확률·통계적 접근방법이 주로 이용되었다. 확률·통계론을 활용한 가장 대표적인 재난위험도 평가 모델로 리스크 곡선을 활용한 재난 발생 특성 분석기법은 인적재난 중 화재, 폭발, 붕괴,

해난사고, 산불을 대상으로 하여 인명피해 및 피해규모, 발생장소 등을 분석하는데 활용하였다[3].

[4]에서는 임의 현상자료(Random Variables)의 확률밀도함수인 KDF(Kernel Density Function)활용하여 지역별 재해 위험성 분석을 수행하였다. KDF는 부정형 구간에 대한 부드러운 분포함수를 추정할 수 있어서 지역별로 재해연보에 나타난 재해 기상에 따른 누적강우량과 해당기간의 최대풍속을 독립인자로 KDF해석을 수행함으로써 지역별 위험성 분석을 수행하고 종속인자로 유형별 피해를 사용하였다.

재해로 인한 각종 피해들은 자연계에 존재하는 하나의 현상으로 미래의 재난·재해 위험을 예측하는 것은 재해 위험을 지배하는 법칙에 대한 지식을 확보함과 동시에 경험적 규칙성에 대한 최소한의 지식을 알아내는 것이 목표다. 하지만 이러한 지식을 확보하려 할 때 문제가 되는 것은 존재하는 법칙을 알아내는 것이 매우 어려울 뿐만 아니라, 그 경험적 규칙성은 우리가 알 수 있을 만큼 명확하지도 않고, noise에 더럽혀져 있다.

최근 전 산업분야에서 널리 이용되고 있는 신경망(Neural Network)은 인간의 뇌의 뉴런들이

상호작용하고 경험을 통해 배우는 생물학적 활동을 모형화한 것이다. 이 모형은 포착된 현상의 데이터를 이용해 그 데이터를 지배하는 경험적 규칙성을 학습하고 획득하는데 다른 기법보다 탁월한 성능을 지니고 있다는 점이다. 즉 데이터를 제공하고 그 학습 알고리즘을 사용하여 학습시키면, 그 데이터에 내재하는 규칙성이 수학적으로는 표현될 수 없지만, 신경망 내부의 가중치에 저장되고, 이후 비슷한 입력이 존재하면 신경망은 학습된 결과에 따라 비슷한 출력을 낼 수 있는 기능을 가지고 있다. 신경망이 여러 분야에 응용되어 좋은 결과를 얻는 것은 통계적 방법에 비해 예측의 정확도나 예측하는 데 걸리는 시간 등의 관점에서 많은 이점을 가지고 있다. 하지만 많은 장점이 있음에도 불구하고 사용자가 결정하여야 할 요소가 많기 때문에 실제로 좋은 성능의 신경망을 설계하는 것은 대단히 어려운 일이다.

본 연구에서는 우리나라 자연재해 피해예측 모델을 개발하기 위해 자연재해로 인한 피해자료와 기상개황자료를 이용하여 지역별 피해를 예측하는 신경망 모델을 개발하였다. 신경망은 다층 퍼셉트론(multi-layer perceptron)을 이용한 교사학습(supervised learning)으로 학습알고리즘은 오류 역전파(back propagation algorithm)를 사용하였다. 또한, 최적의 신경망 학습을 위해 유전자 알고리즘(Genetic Algorithm)을 사용하여 신경망 매개변수를 결정함으로써 예측 정확성을 높였다.

2 과거 기상자료 및 피해자료

2.1 데이터 생성

2.1.1 재해기간·행정구역별 기상개황자료

지역별 기상개황과 피해내역을 활용하여 기상 별 피해를 예측하기 위하여 과거 재해기간별·행정구역별 기상개황자료 생성하였다.

기상청에서 획득할 수 있는 과거 기상정보를 재해기간을 중심으로 재해기간별 기상정보를 생성한다. 이 과정에서 우리나라는 행정구역별로 기상관측소가 존재하는 것이 아니기 때문에¹ 기상관측소가 없는 행정구역에 대해서는 기상개황자료를 시/군 행정구역에 적합하도록 변경하였다. 즉, 행정구역 내에 기상관측소가 없는 경우 주변의 기상관측소의 자료를 활용하여 행정구역의 기상개황을 확보한다. 예를 들어, 표 1과 같이 음성군 내에 기상관측소가 없으므로 주변의 청주, 이천, 충주의 기상관측소를 음성군 기상개황자료를 확보하는 데 활용한다. 각 기상관측소에 역 거리법을 이용해 생성한 적용계수를 반영하여 음성군의 기상개황을 확보한다. 표 2는 이와

같은 방법으로 얻은 음성군의 기상개황자료이다.

표 1 행정구역별 사용 기상대

| 시도 | 시군구 | 차수 | 기상대 | 기상대번호 | 적용계수 |
|------|-----|----|-----|-------|-----------|
| 충청북도 | 음성군 | 1 | 이천 | 203 | 0.2028309 |
| 충청북도 | 음성군 | 1 | 충주 | 127 | 0.5925837 |
| 충청북도 | 음성군 | 1 | 청주 | 131 | 0.2045854 |
| 충청북도 | 제천시 | 1 | 제천 | 221 | 1 |
| 충청북도 | 증평군 | 1 | 청주 | 131 | 0.7824335 |
| 충청북도 | 증평군 | 1 | 충주 | 127 | 0.2175665 |

표 2 재해기간의 지역 기상

| 시도 | 시군구 | 재해코드 | 최대일강수량 | 누적강수량 | 최대풍속 |
|------|-----|----------|----------|----------|----------|
| 충청북도 | 음성군 | 20030901 | 527.5558 | 734.2385 | 81.01755 |

2.1.2 피해자료 변환

과거 재해이력인 피해 규모 등은 과거시점의 행정구역과 현물가치를 기준으로 DB화 되어있어 지역별 위험성 분석을 통한 미래 피해예측을 위해서는 행정구역은 물론 경제적 가치의 현재 시점 자료로의 변환이 필요하다. 따라서 본 연구에서는 그림 1에서 보는 바와 같은 단계적 정보변환 과정을 이용하여 현재 시점의 행정구역 및 현물 가치에 상응하는 피해자료를 구축하였다.

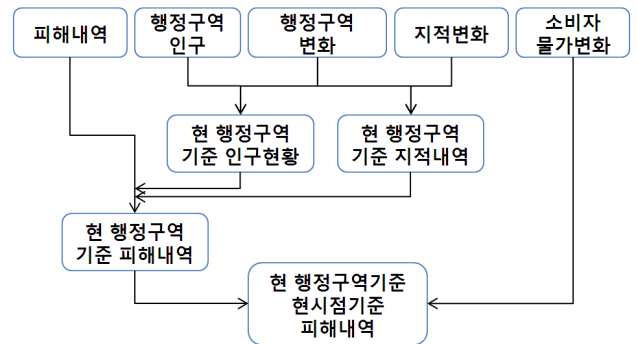


그림 1 피해자료 변환 개요도

먼저 법정 행정구역 변화, 행정구역 인구 변화 및 1993년 이후 5년 주기의 지적변화를 고려하여 과거 행정구역 기준의 인구분포 및 지적분포를 현재 시점 기준으로 변환하였으며, 이 결과를 이용하여 과거 시점의 피해 내역을 현재 법정 행정 구역 기준으로 재구성하였다. 이 결과에 소비자 물가지수의 연도별 변화율을 고려하여 과거 피해 이력을 2007년도 기준 경제가치로 변환하였다.

2.2 데이터 구성

우리나라 자연재해 피해예측 모델을 개발하기 위한 데이터는 소방방재청의 재해연보로부터 1991년부터 2005년까지의 자연재해로 인한 피해자료와 기상청에서

¹ 현재 우리나라에 위치한 기상관측지점은 76개이며 행정구역은 232개이다.

획득한 재해기간의 기상개황자료를 2.1.1과 2.1.2에서 설명한 방법으로 생성하였다.

호우, 태풍, 폭풍 등의 자연재해 212개 사건에 대해 약 6,000여건의 데이터를 생성하였고, 이 중에서 4,665개를 학습 데이터로, 1,500개를 테스트 데이터로, 500개를 평가 데이터로 사용하였다.

3 자연재해 피해예측 모델 개발

3.1 신경망 모델(Neural Network Model)

신경망은 인간의 학습능력을 모방한 것으로 인공 뉴런들의 망으로 구성돼 있고, 각 그룹으로 이루어진 뉴런들은 세 가지 형태의 층인 입력층, 은닉층, 출력층에 위치하게 된다. 입력층의 뉴런들은 외부로부터 정보를 받아들이고, 또한 정보를 은닉층 또는 출력층으로 내보내는 역할을 하며, 은닉층의 뉴런들은 모두가 입력층과 출력층 사이에 위치한다. 그리고 출력층의 뉴런들은 신경망의 결과값을 외부의 사용자에게 제공하는 역할을 한다. 뉴런은 다른 뉴런의 출력 값을 입력 값으로 받아서 뉴런의 출력값을 계산하게 되는데 이러한 뉴런들은 상호 연결돼 있고, 연결된 강도를 가중치(weight)라 한다.

신경망은 전이함수의 역할과 학습알고리즘에 따라서 각각 역방향(feedback)과 순방향(feed forward), 선형(linear)과 비선형(nonlinear), 그리고 교사학습(supervised learning)과 비교사학습(unsupervised learning)으로 구분된다. 가장 일반적으로 이용되는 신경망은 다층 퍼셉트론(multi-layer perceptron)이다. 다층 퍼셉트론은 교사학습이고, 학습알고리즘은 오류 역전파(Back propagation)를 사용[6]한다.

현재까지 많은 연구를 통해서 신경망이 예측, 분류 등의 분야를 학습하고 예측하는데 그 우수성이 증명되었으나, 재난 및 재해의 위험성을 분석하는 연구에는 활용되지 못했다.

본 연구는 신경망을 이용한 자연재해 피해예측 모델의 가능성을 검토하고 향후 모델의 정확성 향상을 위한 지속적 연구 및 영역확대를 통해 다양한 후속 연구가 가능하게 할 것이다.

본 연구에서는 자연재해 피해예측을 위한 신경망 모델 개발을 위해 교사학습의 다층 퍼셉트론 구조와 오류 역전파 알고리즘을 사용하였다.

신경망은 많은 장점이 있음에도 불구하고 사용자가 결정하여야 할 요소가 많기 때문에 실제로 좋은 성능의 신경망을 설계하는 것은 대단히 어려운 일이다.

3.2 유전자 알고리즘(Genetic Algorithm)

유전자 알고리즘은 확률적인 최적해 탐색 방법으로 다윈의 적자생존의 원리를 기초로 한 자연계의

진화과정의 메커니즘을 모델링한 것이다. 한 세대에서 각 개체는 일반적으로 자연계에서 염색체와 유전인자를 모방한 선형 스트링(String)으로 표현된다. 유전자 알고리즘은 문제에 대한 해를 최적의 개체로 하고 개체의 적합도를 최대화시키도록 한다.

유전자 알고리즘의 기본 알고리즘은 재생산(선택), 교배, 돌연변이 연산자이다. 재생산 연산자는 각 세대에서 최적개체의 유전적 특성이 다음 세대에 전달되는 자연선택과 비슷하며, 가정 널리 사용되는 재생산 연산자는 룰렛휠(Roulette Wheel) 방법이다. 교배 연산자는 적용된 교배확률에 따라 무작위로 두 개체를 선택하여 선택된 개체의 일정부분을 서로 교환함으로써 새로운 두 개의 개체를 생성하여 개체군에 새롭게 삽입(복사)하는 것으로 일점교배, 다점교배, 균등교배 등이 있고, 일반적으로 일점교배가 사용된다. 돌연변이 연산자는 주어진 돌연변이 확률만큼 단지 한 개의 개체에서 작용하는데 개체의 특정 인자를 바꾸는 것이다. 이러한 연산자에 의해서 생성된 개체를 자손이라고 하고 새로운 개체는 선택된 개체(부모)의 유전적 특성을 물려받게 된다.

유전자 알고리즘은 구조가 비교적 간단하여서 응용범위가 매우 넓으며, 특히 최적화 문제(Optimization problem)에 많이 적용되고 있다. 또한, 신경망, 퍼지이론(Fuzzy Logic)등과 조합은 기존의 알고리즘들이 가지고 있었던 한계를 유전자 알고리즘을 이용하여 극복할 수 있게 하였다.

본 연구에서는 신경망 모델을 학습하는데 정의되어야 할 매개변수를 유전자 알고리즘을 이용하여 결정함으로써 보다 좋은 성능의 모델을 설계하고자 한다. 신경망 구조를 결정하는 요소는 은닉층의 수, 은닉노드의 수, 출력노드의 수, 학습률(learning rate), 모멘텀(momentum), 편의값(bias) 등이 있다. 은닉층의 수는 시행착오적인 방법으로 실험하였고, 학습률과 모멘텀, 편의값은 유전자 알고리즘으로 결정하여 학습하였다.

3.3 자연재해 피해예측을 위한 신경망과 유전자 알고리즘 모델 구축

3.3.1 데이터 정규화

신경망은 입력변수와 출력변수가 [0,1] 사이에 있을 때 최대 성과를 낸다. 이러한 이유로 모든 변수는 신경망에 입력하기 전에 [0,1]로 변환되어야 한다. 본

연구에서는 다음과 같은 정규화·역정규화 방법을 사용하여 누적강우량, 최대풍속, 선행강우량, 총 피해액을 정규화하였다.

■ 정규화

$$newX = \frac{0.8}{(\max X - \min X)} \times X + (0.1 - 0.8 \times \frac{\min X}{(\max X - \min X)})$$

■ 역정규화

$$X = \frac{(newX - (0.1 - (0.8 \times \frac{\min X}{(\max X - \min X)})))}{0.8}$$

$$\frac{(\max X - \min X)}$$

행정구역을 나타내는 지역은 232개의 범주를 갖는 범주형 변수로 값의 의미가 없으므로 m-1개 더미변수로 변환하였다.

3.3.2 유전자 알고리즘으로 최적화한 신경망 모델 학습 자연재해 피해예측 신경망 모델의 입력변수는 누적강우량, 최대풍속, 재해사상 발생 5일 이내의 선행강우량, 그리고 행정구역 구분이며 출력변수는 총 피해액이다.

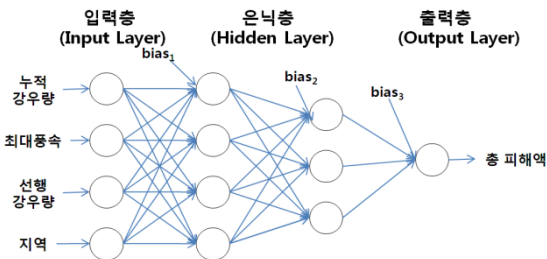


그림 2 자연재해 피해예측 신경망 모델 구조

신경망의 은닉층의 개수는 임의로 변경하여 모델의 정확성을 실험하였다. 또한, 유전자 알고리즘을 이용하여 신경망의 학습률, 모멘텀, 그리고 편의값을 결정하였다.

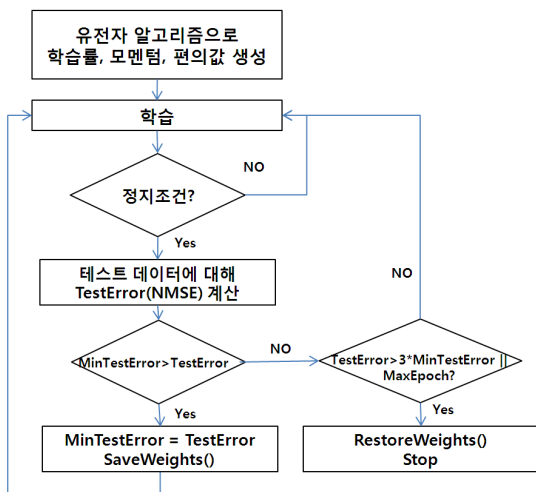


그림 3 신경망 모델 학습 절차

이렇게 결정된 신경망의 매개변수들을 가지고 정해진 학습 데이터를 이용해 반복적으로 학습을 수행하였고 반복 학습 수행 시 테스트 데이터를 대상으로 신경망의

오차인 NMSE(Normalized Mean Square Error),

$$NMSE = \frac{1}{N} \sum_{i=0}^{N-1} (x - \hat{x})^2 / \frac{1}{N} \sum_{i=0}^{N-1} (x^2)$$

NMSE와 비교하여 최적의 신경망 구조를 저장한다. 종료 임계값에 도달하거나 이전 학습의 NMSE와의 차이가 적을 경우 학습을 종료하도록 하였다. 그림 3은 신경망 모델의 학습절차를 도식화 한 것이다.

3.3.3 평가

학습이 완료된 신경망을 평가하기 위해 학습 데이터, 테스트 데이터, 평가 데이터를 적용하여 평가를 실시하였다. 또한, 신경망의 매개변수를 유전자 알고리즘으로 결정한 경우와 사용자 입력으로 결정한 경우를 비교하였으며, 신경망의 은닉층의 개수 변화에 따른 신경망의 정확도를 살펴보았다. 각 구분된 데이터에 대해 평가한 결과는 표 3과 같으며 오차(NMSE)를 백분율(%)로 표기하였다.

표 3 학습된 신경망 모델 평가 (%)

| A \ B | 1개 | | | 2개 | | |
|-------|-------|--------|-------|-------|--------|-------|
| | 학습 | 테스트 | 평가 | 학습 | 테스트 | 평가 |
| 사용자선택 | 0.481 | 10.180 | 0.574 | 0.486 | 10.160 | 0.572 |
| GA결정 | 0.408 | 10.038 | 0.415 | 0.327 | 10.930 | 0.405 |

A : 매개변수 결정방법, B : 은닉층의 개수

유전자 알고리즘으로 신경망의 매개변수를 결정하고 은닉층의 개수가 1개보다 2개일 경우 학습 데이터와 평가 데이터에서 오차율이 적게 나타났다. 이것은 향후 신경망의 은닉층의 개수 및 노드의 개수를 유전자 알고리즘으로 최적화할 필요가 있음을 나타내며, 또한 기타 신경망의 매개변수들을 유전자 알고리즘으로 결정하는 추가적인 연구가 필요하다고 하겠다.

4 결론 및 향후 연구계획

신경망은 예측, 분류 등의 분야를 학습하고 예측하는데 그 우수성이 증명되었으나, 재난 및 재해의 위험성을 분석하는 연구에는 활용되지 못했다. 본 연구는 신경망을 이용한 자연재해 피해예측 모델에 대해 연구를 시도함으로써 가능성을 검토하였고 향후 피해예측 모델의 정확성 향상을 위한 지속적 연구 및 인적재난 위험성 분석 영역으로의 확대를 통해 다양한 후속 연구가 가능하게 될 것이다.

자연재해 피해예측을 위한 신경망 모델 개발을 위해 교사학습의 다층 퍼셉트론 구조와 오류 역전파 알고리즘을 사용하였고, 지역적 최고값(local maxima)의 문제에 빠지지 않고 최적의 신경망을 학습하기 위하여 신경망 매개변수인 학습률, 모멘텀, 편의값을 유전자 알고리즘으로 결정하였다.

재난·재해분야 위험도 예측을 신경망을 활용한 시범적 연구에 그치지 않고 모델을 재난관리 활용하기 위해서는 다음과 같은 연구가 지속적으로 수행되어야 한다.

첫째, 다양한 신경망의 아키텍처를 정의하고 실험을 통하여 모델 비교 분석 수행되어야 할 것이다. 신경망의 가중치, 활성화 함수, 은닉층의 수, 은닉노드의 수, 출력층의 수 등 신경망의 구조를 결정하는 매개변수들의 최적값을 결정하여 최적의 모델을 개발해야 한다.

둘째, 자연재해 피해예측 모델이 제시하는 예측 수준이 좀더 구체적이어야 한다. 먼저 출력변수인 총 피해액을 정량화하여 위험도로 정의하고 이것을 모델 학습에 이용하고 예측결과로 위험도를 제시해야 할 것이다. 또한 좀더 구체적인 피해항목별 위험도를 제시해야 한다. 예를 들면, 이재민, 인명, 주택, 선박, 농경지, 하천, 공공시설 피해 등 구체적인 위험도 제시하는 예측 모델의 활용성을 높일 수 있다.

마지막으로 지역에 대해 재해유형별 위험도를 예측해야 한다. 따라서 신경망의 입력변수에 대한 연구가 필요할 것이다.

자연재해 피해예측 신경망 모델의 정확성 향상과 재난관리 활용화를 위해 위에 제시한 문제를 해결하기 위한 연구를 지속적으로 수행할 계획이다.

참 고 문 헌

- [1] 국립방재연구소, “지진피해 취약도함수 국산화 연구”, 2008.
- [2] 국립방재연구소, “지역별 안전도 평가 시스템 구축 및 세부 평가수행방안에 대한 연구”, 2006.
- [3] 국립방재연구소, “리스크 곡선을 활용한 특성분석 방안 연구”, 2008.
- [4] 국립방재연구소, “웹GIS를 활용한 국내 재해정보DB 및 태풍위원회 재해정보시스템 구축”, 2008.
- [5] 김광희, 강경인, “유전자 알고리즘에 의한 신경망 구조의 최적화를 이용한 공동주택 초기 공사비 예측에 관한 연구”, 대한건축학회논문집, 제20권, 제2호, pp.81-89, 2004.
- [6] Hegazy, T., and Moselhi, O. “Analogy-based solution to markup estimation problem”, ASCE Journal of Computing Civil Engineering, 8(1), pp.72-87, 1994.
- [7] 국립방재연구소, “태풍위원회 통합 태풍재해정보분석 시스템 구축”, 2009.
- [8] 신대진, 정요원, 강환일, “유전 역전파알고리즘을 이용한 시계열 예측방법의 연구”, 한국통신학회 논문지, Vol.20, No.1, pp.445-448, 1999.