

신경망 모델과 확률 모델의 풍수해 예측성능 비교

최 선 화[†]

요 약

최근 급증하는 기상이변 및 기후온난화 현상은 풍수로 인한 피해를 더욱 가속시키고 있어 풍수해 발생가능성을 미리 예측하여 선제적으로 대응할 방안 마련이 필요하다. 재난·재해의 위험성 분석은 주로 확률·통계기법에 기반한 수식모델 연구가 주류를 이루고 있고 소방방재청 국립방재연구소에서 구축한 태풍위원회 재해정보시스템(TCDIS: Typhoon Committee Disaster Information System) 또한 지역별 풍수해 위험성 분석에 확률모델을 활용하고 있다. 본 논문에서는 경험적 패턴인식에 탁월한 성능을 가진 신경망 알고리즘을 활용하여 개발한 풍수해 예측모델을 소개하고 이 모델과 TCDIS의 KDF 확률밀도함수를 이용한 풍수해 예측모델의 성능 비교 결과를 제시하여 기존 TCDIS의 위험성 분석기능에 신경망 모델을 적용함으로써 시스템의 강건성과 예측 정확도 향상이 가능함을 보이고자 한다.

키워드 : 풍수해 예측, 신경망, 패턴인식, 모델 최적화, 확률밀도함수

Performance Comparison between Neural Network Model and Statistical Model for Prediction of Damage Cost from Storm and Flood

Seonhwa Choi[†]

ABSTRACT

Storm and flood such as torrential rains and major typhoons has often caused damages on a large scale in Korea and damages from storm and flood have been increasing by climate change and warming. Therefore, it is an essential work to maneuver preemptively against risks and damages from storm and flood by predicting the possibility and scale of the disaster. Generally the research on numerical model based on statistical methods, the KDF model of TCDIS developed by NIDP, for analyzing and predicting disaster risks and damages has been mainstreamed. In this paper, we introduced the model for prediction of damage cost from storm and flood by the neural network algorithm which outstandingly implements the pattern recognition. Also, we compared the performance of the neural network model with that of KDF model of TCDIS. We come to the conclusion that the robustness and accuracy of prediction of damage cost on TCDIS will increase by adapting the neural network model rather than the KDF model.

Keywords : Damage from Storm and Flood, Prediction of Damage, Neural Network, Pattern Recognition, Model Optimization, Kernel Density Function

1. 서 론

미래의 발생 가능한 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 현상에 존재하는 법칙을 알아내는 것이 어려울 뿐만 아니라, 그 경험적 규칙성은 우리가 알 수 있을 만큼 명확하지도 않다[1].

전통적으로 재해로 인한 위험성 분석 및 피해예측에는 과

거의 자료를 기반으로 하는 확률·통계적 접근방법이 주로 사용되었다. 확률·통계론을 활용한 가장 대표적인 재난위험도 평가 모델로는 리스크 곡선을 활용한 재난 발생 특성 분석기법이다. 이 모델은 인적재난 중 화재, 폭발, 붕괴, 해난 사고, 산불을 대상으로 하여 인명피해 및 피해규모, 발생장소 등을 분석하는데 활용하였다[2]. 또한, 임의 현상자료의 확률밀도함수를 나타내는 KDF(Kernel Density Function)를 활용하여 기상현상별 시설물별 피해내역으로 재해기상에 적합한 경향함수를 구축하고 지역별 재해 위험성 분석을 수행한 연구도 있었다[3].

하지만, 재해발생과 위험도와 같은 임의성이 높은 자료를 활용하여 수학적 모델로 문제를 해결하려는 확률·통계적 접근은 예측의 정확도가 낮고 소요시간이 길다는 한계를 가지

[†] 정 회 원 : 소방방재청 국립방재연구소 책임연구원
논문접수: 2011년 3월 29일
수정일: 1차 2011년 7월 4일
심사완료: 2011년 7월 4일

고 있다. 이처럼 수치적으로 해결이 곤란한 문제를 패턴 인식에 뛰어난 신경망(Neural Network) 알고리즘을 활용하여 해결하려고 하는 연구에 관심이 급증하고 있다. 신경망은 인간의 뇌의 뉴런들이 상호작용하고 경험을 통해 배우는 생물학적 활동을 모델화한 것이다. 이 모델은 포착된 현상의 데이터를 이용해 그 데이터를 지배하는 경험적 규칙성을 학습하고 획득하는데 다른 기법보다 탁월한 성능을 지니고 있다고 알려져 있다. 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[4].

신경망과 유전자 알고리즘을 이용하여 풍수해를 예측하려는 연구인 [5]는 신경망 매개변수 중 모멘텀과 학습률을 사용자가 입력하여 생성한 모델과 유전자 알고리즘을 활용하여 생성한 모델을 비교하였다. 하지만 두 개의 사용자 매개변수와 은닉층의 개수를 달리한 단순한 모델구조로는 피해 예측의 최적모델을 찾는 데는 한계가 있었다.

본 연구에서는 풍수해 예측을 위한 신경망 모델 생성을 위해 지역별 풍수해 자료와 기상개황 자료뿐만 아니라 지역의 지형적 특성을 신경망 모델의 입력으로 추가하였고, 신경망 모델의 구조와 매개변수를 [5]보다 다양화하여 모델 생성·검증·평가 과정을 수행하였다. 즉, 신경망의 초기 가중치, 은닉층의 노드수, 모멘텀, 학습률을 달리하여 모델들을 생성하고 정확도(accuracy)의 평균과 TPR(True Positive Rate)과 FPR(False Positive Rate)를 척도로 성능을 비교하여 최적모델을 결정하였다. 결정된 풍수해 예측 최적의 신경망 모델과 소방방재청 국립방재연구소에서 구축한 태풍위험회재해정보시스템(TCDIS: Typhoon Committee Disaster Information System, 이후 TCDIS로 표기)의 KDF 확률밀도 함수를 기반으로 한 지역별 풍수해 위험도 분석 모델을 평가 데이터에 반영하여 성능을 비교하였다. 그 결과 TCDIS의 지역별 위험성 분석 기능의 강건성(Robustness)과 정확도 향상에 신경망 모델이 KDF 분석 모델보다 탁월함을 확인하였다.

2. 풍수해 예측 신경망 모델

2.1 모델 생성

소방방재청의 재해연보로부터 1991년부터 2005년까지 풍수로 인한 피해자료와 기상청에서 획득한 재해기간의 기상개황자료, 지역별 하천의 개수와 과거 풍수해 발생횟수를 활용하여 우리나라 232개 행정구역의 누적강우량과 최대풍

속, 재해사상 발생 5일 이내의 선행강우량, 총 피해액[5], 그리고 지역의 풍수해 발생 영향요인을 지역적 특징으로 정의하고 풍수해 예측모델 학습을 위한 데이터를 구성하였다. 최대 성과를 내기위해 신경망의 입력변수와 출력변수를 [0,1,0.9] 사이의 값으로 정규화하고, 호우, 태풍, 폭풍 등의 풍수해 212개 사건에 대해 약 6,000여건의 전체 데이터 집합에서 랜덤 추출하여 학습 데이터는 전체 데이터의 60%(3,999건), 검증 데이터는 30%(2,000건), 평가 데이터는 10%(666건)로 분할하여 한 세트(holdout)를 구성하였다. 이와 같은 방법으로 학습·검증·평가 데이터를 총 5세트(holdout)로 구성하고 각 모델마다 학습, 검증, 그리고 평가 작업을 5세트를 이용하여 반복 수행하였다.

풍수해 예측을 위한 신경망 모델 생성은 교사학습(supervised learning)의 다층 퍼셉트론 구조(multi-layer perceptron)와 오류 역전파 알고리즘(Back propagation)을 사용하였다. 신경망의 구조 및 학습 매개변수 값을 <표 1>과 같이 달리하여 8,379가지의 모델 구조 및 학습 매개변수 조합을 만들어 내고 이를 이용해 모델을 초기화하고 학습하였다. 신경망의 층은 입력, 은닉, 그리고 출력층의 3개 층으로 정의하였고, 누적강우량, 최대풍속, 선행강우량, 지역적 특징 값을 입력변수로 총피해액을 출력변수로 정의하였다. (그림 1)은 풍수해 예측모델 학습의 개념도이다.

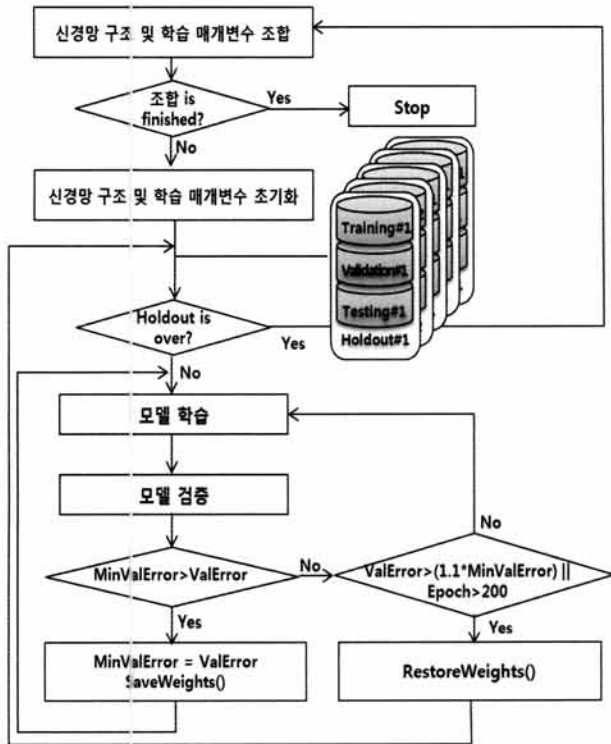
<표 1> 신경망의 구조 및 학습 매개변수

구분		값의 범위	간격	경우의 수
신경망의 구조	초기 가중치	±0.5, ±1.0, ±2.0	-	3
	은닉층 개수	1	-	-
	은닉층의 노드수	6 ~ 12	1	7
학습 매개변수	모멘텀	0.0 ~ 2.0	0.1	21
	학습률	0.1 ~ 1.0	0.05	19
입력변수	누적강우량	정규화값		
	최대풍속			
	선행강우량			
	지역적 특징값			
출력변수	총피해액			

신경망 구조 및 학습 매개변수 조합에 따라 신경망의 은닉층 노드수와 초기 가중치, 모멘텀, 그리고 학습률을 정의하고 모델을 초기화 한다. 미리 정의한 학습 데이터 5세트 중 첫 번째 세트의 학습 데이터를 이용하여 오류 역전파 알고리즘을 수행한다. 학습된 모델에 대해 첫 번째 세트의 검증 데이터를 이용하여 모델의 오류(E)를 계산한다. 검증 데이터에 대한 모델의 오류는 $E = \frac{1}{2} \sum_{i=1}^N (t_i - p_i)^2$ 으로 평가되며, N은 데이터의 개수, t_i 는 i번째 데이터의 목표값, p_i 는 i번째 데이터에 대한 모델의 예측값을 의미한다[6][7].

검증 오류가 최소로 수렴하면 학습을 종료하지만 그렇지 않을 경우 최대 200번 반복학습을 수행한 후 학습을 종료한다. 즉, (그림 4)에 제시한 바와 같이 이전 학습한 모델의 검증 오류 중 최소값이 현재 학습한 모델의 검증 오류보다 크면 최소 검증 오류를 현재 학습모델의 검증오류로 다시 설정한 후 학습을 반복한다. 만약 현재 학습 모델의 검증 오류가 최소 검증 오류보다 크다고 하더라도 최소 검증 오류*1.1 한 값보다 크지 않다면 최대 200번 반복학습을 수행하지만, 그렇지 않은 경우는 학습을 종료하고 학습과정 중 최소 검증 오류를 낸 모델을 최종 학습 모델로 선정한 후 학습을 종료한다.

미리 정의한 5세트의 학습 데이터에 대해 위와 같은 모델 학습 수행이 완료되면 새로운 신경망 구조 및 학습 매개변수 조합에 따라 모델을 초기화하여 새 모델에 대해 학습과정을 반복 수행한다. 학습된 각 모델의 정확도는 5세트 검증 데이터 각각의 정확도의 평균으로 구한다.



(그림 1) 모델 학습 개념도

2.2 모델 검증 및 평가

실제 피해액을 정확하게 예측하는 모델개발은 현실적으로 어려운 문제이므로 풍수해 예측의 문제를 분류문제로 전환하여 최적의 모델을 찾았다.

모델의 정확도 계산을 위해 Confusion Matrix¹⁾에 TP(True Positive), FN(False Negative), FP(False

Positive), TN(True Negative)를 구하고

$$\frac{TP + TN}{TP + TN + FP + FN}$$

식을 이용하여 정확도를 계산하였다. 이 정확도는 5세트 검증 데이터 각각에 대해 계산하였으며, 모델의 최종 정확도는 그들의 평균으로 정의하였다.

데이터 집합의 클래스가 균등하게 분포하지 않는 경우, 즉 데이터 집합이 특정 클래스로 편중되어 있는 경우 정확도는 모델의 성능을 평가하는데 한계가 있는 척도이다. 예를 들어 어떤 데이터 집합이 클래스 A는 990개고 클래스 B는 10개라면 이 데이터 집합은 클래스 A에 데이터가 편향되어 있다고 할 수 있다. 이 경우 모델이 모든 데이터를 클래스 A로 분류하였다면 모델의 정확도는 99%(990/1000)로 높게 나타나며 클래스 A의 인식률은 100%지만, 클래스 B의 인식률은 0%로 나타나 모델은 클래스 B를 전혀 인식하지 못함에도 불구하고 높은 정확도를 갖게 된다. 따라서 데이터 집합의 데이터의 분포가 균등하지 않은 경우 정확도가 모델의 성능을 정확히 반영하지 못하게 된다.

본 연구에서 사용하는 데이터 집합의 클래스 분포를 살펴보면 총피해액이 0과 100만원 사이의 클래스 0이 약 76%를 차지하고 있어 0번 클래스에 편향되어 분포하고 있다. 따라서 본 연구에서는 정확도가 모델 간 성능의 변별력이 떨어지므로 정확도 이외에 모델 성능 평가지수로 ROC 공간상의 TPR과 FPR를 도식하여 성능을 평가하였다. TPR²⁾은 모델의 민감도(Sensitivity)를 나타내는 것으로 특정 클래스에 해당하는 모든 데이터를 해당 클래스로 분류하는 능력이다. 이와 반대로 FPR³⁾은 특정 클래스에 해당하지 않은 모든 데이터 중에서 특정 클래스로 잘못 분류되는 비율을 나타낸다. 이 값은 1-특이도(Specificity)라고도 표현하기도 한다. ROC 공간의 x축과 y축은 FPR과 TPR로 각각 정의되며 이익(True Positive)과 비용(False Positive)의 균형을 표현할 수 있다.

풍수해 예측을 위해 학습된 모델들의 성능을 ROC 공간상에 도식한 그래프는 (그림 2)와 같다. ROC 공간에 도식된 (TPR,FPR)이 (1,0)에 가까울수록 모델의 성능이 우수한 것이며 대각선으로 표시된 Random 예측선을 기준으로 위쪽으로 갈수록 모델 성능이 좋아지고 예측선 기준으로 아래로 갈수록 모델 성능은 떨어진다. 본 연구에서는 TPR이 0.5 이상이며 FPR이 0.5미만인 모델 중에서 Random 예측선에서 가능한 멀리 떨어진 모델 17개⁴⁾를 선택하여 최적모델 결정을 위한 후보 모델로 정의하였다.

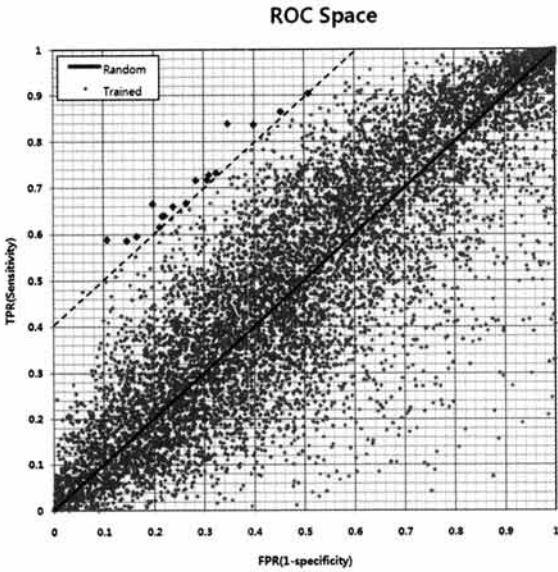
17개 후보 모델을 대상으로 최적모델을 결정하기 위해 5세트 평가 데이터를 적용하여 정확도의 평균을 각 모델의 정확도로 정의하고 TP, FP, TN, FN 각각의 평균을 각 모델의 해당 값으로 이용하여 (그림 3)의 ROC 그래프를 구하였다.

1) Confusion Matrix의 각 행에는 예측된 클래스를 표시하고 열에는 실제 클래스를 표시하는 Matrix로 분류문제의 정확도를 쉽게 판단할 수 있는 장점이 있다.

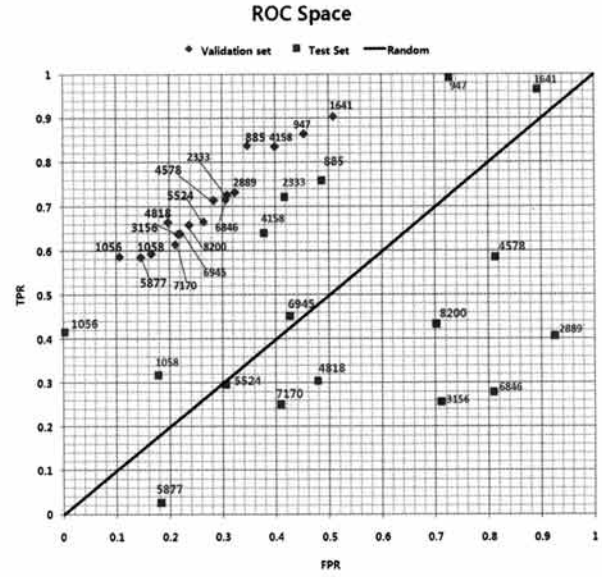
2) TPR = TP/(TP+FN)

3) FPR = FP/(FP+TN)

4) (그림 2)에서 점선과 (TPR,FPR)=(1,0) 사이의 점



(그림 2) 학습된 모델의 ROC 그래프



(그림 3) 후보모델의 ROC 그래프

<표 2> 후보모델 평가결과

No	Model	Validation Set				Test Set				Difference	
		Accuracy	TPR	FPR	Dist. from Random	Accuracy	TPR	FPR	Dist. from Random	TPR	FPR
1	885	0.6484	0.8380	0.3470	2.5	0.6482	0.7585	0.4865	1.5	-0.0795	0.1395
2	4818	0.6444	0.6646	0.1978	2.5	0.6485	0.3047	0.4787			
3	1056	0.6407	0.5870	0.1062	2.5	0.6464	0.4153	0.0027			
4	5877	0.6336	0.5856	0.1464	1.5	0.6393	0.0270	0.1837			
5	4158	0.6442	0.8354	0.3991	1	0.6524	0.6401	0.3777	1.5	-0.1953	-0.0214
6	4578	0.6482	0.7151	0.2841	1	0.6521	0.5856	0.8131			
7	1058	0.6477	0.5942	0.1661	1	0.6476	0.3183	0.1784			
8	1641	0.6619	0.9036	0.5084	0.5	0.6643	0.9657	0.8929			
9	947	0.6526	0.8642	0.4534	0.5	0.6584	0.9921	0.7271			
10	2889	0.6473	0.7315	0.3240	0.5	0.6476	0.4065	0.9248			
11	2333	0.6490	0.7265	0.3093	0.5	0.6521	0.7216	0.4170	2	-0.0049	0.1077
12	6846	0.6403	0.7159	0.3064	0.5	0.6377	0.2788	0.8092			
13	8200	0.6504	0.6590	0.2380	0.5	0.6514	0.4328	0.7019			
14	6945	0.6356	0.6387	0.2211	0.5	0.6467	0.4521	0.4263			
15	3156	0.6420	0.6381	0.2168	0.5	0.6444	0.2567	0.7111			
16	5524	0.6513	0.6655	0.2644	0	0.6473	0.2965	0.3058			
17	7170	0.6570	0.6143	0.2110	0	0.6467	0.2505	0.4090			

17개 후보모델들은 검증 데이터에 대해 ROC 그래프에서 Random 예측선 위에 존재하는 것으로 우수모델이었으나 평가 데이터에 적용한 결과 9개 모델이 ROC 그래프에서 Random 예측선 아래에 나타났다. 또한, TPR이 0.5이상이고 FPR이 0.5미만인 모델은 885, 4158, 2333의 3개 모델로 확인되었다. 최적의 모델을 결정하는데 정확도의 변별력이 떨어져서 ROC 그래프를 이용하여 우수모델로 판단되는 3개

모델, 885, 4158, 2333 중 최적 모델을 결정하였다.

<표 2>에 제시된 10번째 열값은 모델을 평가 데이터에 적용하여 ROC 그래프에서 모델과 Random 예측선의 거리를 나타낸다. 이 거리가 클수록 좋은 모델을 나타내는 지표이므로 이 값을 이용한다. 또한 모델을 검증 데이터에 적용

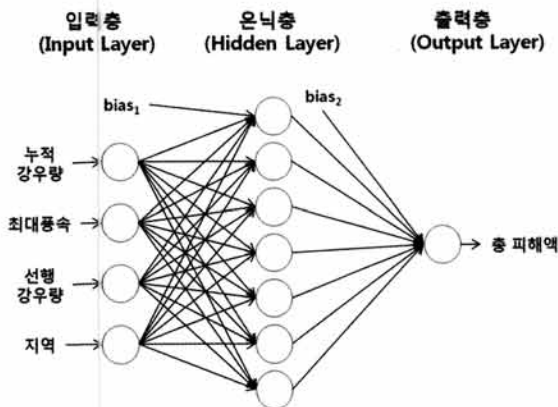
5) Random 예측선의 거리는 (그림 3)에 Random 예측선과 모델의 (TPR,FPR)좌표 사이에 존재하는 격자의 개수임.

한 경우보다 평가 데이터에 적용한 경우의 TPR이 증가할수록 FPR은 감소할수록 좋은 성능을 갖는 모델이므로 이 두 값들의 차이를 선정 지표로 사용하였다. 그 결과 2333 모델이 Random 예측선과의 거리가 가장 크고, TPR이 가장 작게 감소됨과 동시에 FPR의 증가치도 가장 작아 검증 데이터 적용 결과와 평가 데이터 적용 결과에 큰 차이가 없는 안정된 모델임을 알 수 있다. 따라서 2333모델을 풍수해 예측을 위한 최적모델로 선정하였으며, <표 3>은 이 모델의 구조 및 평가 결과이다.

은닉층 개수는 본 연구의 조정 대상이 아니었으나, 향후 연구의 가능성을 검토하기 위해 위와 같이 정의한 풍수해 예측 최적 신경망 모델에 대해 은닉층 개수와 노드수⁷⁾를 조

<표 3> 풍수해 예측 최적 모델

구분	값	비고	
모델 구조	층 수	3	
	입력층의 노드수	5	bias 포함
	은닉층 개수	1	
	은닉층의 노드수	7	
	출력층의 노드수	1	
	가중치	±1.0 범위값	모델학습결과 얻어진 가중치
사용자 매개변수	모멘텀	1.9	
	학습률	0.95	
오류 ⁶⁾ 의 평균	1.64041E-04	학습	
	1.85600E-04	검증	
	1.27628E-04	평가	
정확도의 평균	0.6490	검증	
	0.6521	평가	
TFR의 평균	0.7265	검증	
	0.7216	평가	
FFR의 평균	0.3093	검증	
	0.4170	평가	

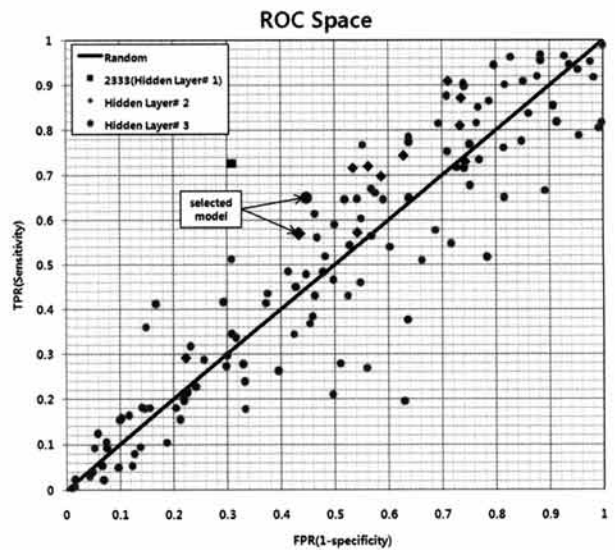


(그림 4) 풍수해 예측 최적 모델구조

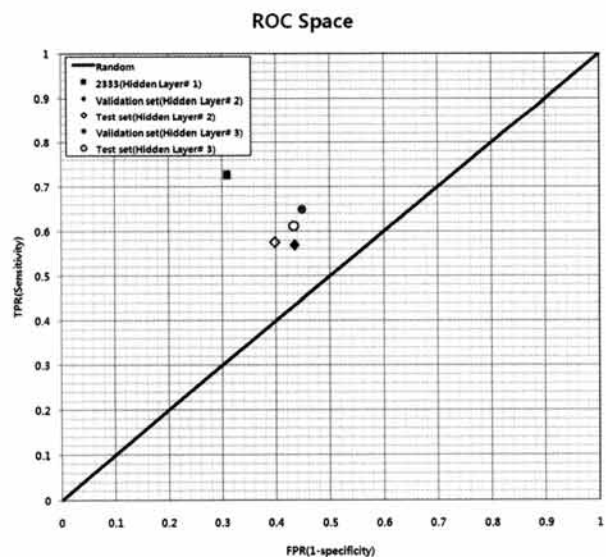
정하여 모델의 변화를 실험하였다. 그 결과, (그림 5)에 제시된 바와 같이 1개 은닉층을 갖는 2333 모델의 성능이 은닉층의 개수가 더 많은 구조의 모델보다 우수하였다.

은닉층 개수가 2인 모델과 3인 모델 중 성능이 우수한 모델을 선택하여 비교해보면 은닉층 개수가 3인 모델이 2인 모델보다 우수한 것처럼 보이지만 평가 데이터에 적용한 결과 성능의 큰 차이를 보이지 않았다.

위 실험은 초기 가중치, 첫 번째 은닉층 노드의 수, 그리고 사용자 매개변수인 모멘텀과 학습률을 <표 3>의 값으로 고정한 후 은닉층 개수 증가에 따른 성능을 비교한 것이고 그 결과 은닉층 개수 1인 모델에서 최적의 성능을 보였다. 향후 모델구조 및 매개변수 값을 결정하는 초기단계에서부터 은닉층 개수를 고려하는 연구를 통해 성능을 비교할 필요가 있다.



(그림 5) 은닉층 개수별 모델의 성능



(그림 6) 은닉층 개수별 모델의 평가 데이터 적용결과

6) 오류 = $\frac{1}{N} \left(\frac{1}{2} \sum_{i=1}^N (t_i - p_i)^2 \right)$, N = 데이터 수, t_i = 목표값, p_i = 예측값

7) 은닉층의 개수 : 2개 ~ 3개, 은닉층의 노드수 : 2개 ~ 12개

8) (그림 5)에 "Selected mode"로 표기된 모델

3. KDF를 활용한 풍수해 예측 모델

과거의 재해 기상에 해당하는 강수량 및 최대풍속과 유형별 피해내역은 매우 산포된 자료이므로 통계적 접근보다는 KDF를 이용해 부정형 구간에 대한 부드러운 분포함수를 추정함으로써 풍수해를 예측할 수 있다[3].

3.1 KDF 개요

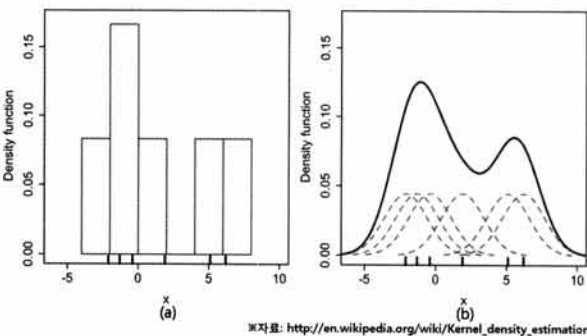
KDF(Kernel Density Function)는 임의현상자료의 확률밀도함수를 나타내는 한 방법으로, 예를 들어 KDF를 이용하면 인구에 관한 표본을 이용하여 전체 인구에 대한 경향 파악이 가능하다. 도수 분포(Histogram)는 등 간격의 표본 검색을 이용한 KDF의 일종으로 생각할 수 있다.

확률적으로 독립적인 임의 현상의 표본을 x_1, x_2, \dots, x_N 라고 하면 이에 대한 확률밀도함수를 KDF

는 $\hat{f}_h(x) = \frac{1}{Nh} \sum_{i=1}^N K\left(\frac{x-x_i}{h}\right)$ 으로 나타낼 수 있는데 이때 K는 커널(kernel), h는 빈도폭(bandwidth 또는 smoothing parameter)이다. 가장 흔히 사용되는 K는

$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$ 같이 0과 1을 각각 평균과 분산으로 갖는 표준 가우스 함수이다.

비록 도수분포와 같이 부드럽지 않은 밀도함수도 점근적으로는 일관성을 갖도록 할 수 있지만 때때로 불연속적이거나 수렴속도가 매우 느린 밀도 분포가 존재한다. 따라서 표본을 일정한 빈도폭으로 분류하기보다 각각의 표본을 각각의 기준으로 삼아 커널 함수를 결정함으로써 최종적인 밀도함수는 이러한 각각의 커널의 총체로서 나타낼 수 있다. 이 경우 일반적인 도수분포와 달리 매우 부드러운 경향을 나타낼 수 있다. 예를 들어 임의현상자료 x_i 가 -2.1, -1.3, -0.4, 1.9, 5.1, 6.2로 주어질 때, 같은 자료를 일반적인 도수분포에 의해 표현한 경우9)와 KDF로 주어진 표본의 확률밀도함수10)로 표현한 경우를 비교해보면 KDF를 활용하여 같은 자료를 부드러운 분포로 나타낼 수 있음을 알 수 있다[8].

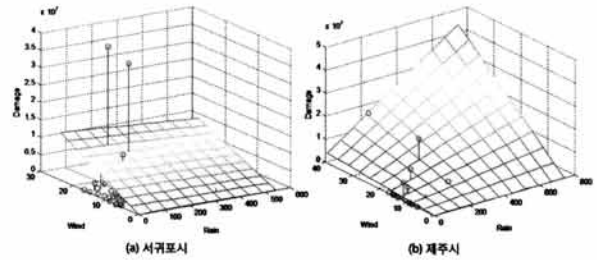


(그림 7) 도수분포와 KDF 비교

9) (그림 7)의 (a) : 위의 자료는 빈도폭이 2인 6개 bin으로 표현되며 임의자료가 특정 bin에 포함되는 경우 그 bin의 막대높이가 1/12씩 증가하도록 작성
10) (그림 7)의 (b) : 점선은 각 자료의 커널(K)이며 실선은 KDF로 표본의 확률밀도함수를 표현한 것

3.2 KDF 분석을 통한 경향함수 도출

지역별 재해연보에 나타난 재해 기상에 따른 누적강수량과 해당기간의 최대풍속을 독립인자로 KDF 분석을 수행하였고, 이 때 종속인자는 유형별 피해내역이다. (그림 8)은 서귀포시와 제주시의 피해 경향함수를 나타내고 있다. 서귀포시는 주로 풍속(Wind)에 의해 피해 규모가 결정되지만 제주시는 풍속과 강수량(Rain)에 의해 결정되는 것으로 나타났다.



(그림 8) 총 피해액 경향함수

4. 신경망 모델과 KDF 분석 모델 비교

지역별 풍수해 예측을 위한 신경망 모델과 KDF 분석 모델의 성능을 비교하기 위하여 평가 데이터 5세트에 각각의 모델을 적용하여 정확도의 평균과 TPR, FPR의 분포를 비교하였다.

신경망 모델과 달리 KDF 분석 모델은 예측 범위 밖의 풍속과 강수량이 입력되는 경우 예측을 실패하므로 각 모델의 예측 정확도를 비교하기 이전에 예측 가능성을 비교해 보았다. 신경망 모델은 666개 평가 데이터 입력에 대해 666개 모두 예측결과를 내어주는 반면에 KDF 모델은 666개중에서 평균 100개 데이터만을 예측할 수 있어 예측 가능성이 15%에 해당한다. 이는 대부분의 입력에 대해 예측결과를 제시하지 못하여 재해위험에 대한 상황판단을 적절히 지원하지 어려운 것으로 판단된다.

하지만 두 모델의 예측 정확도 비교를 위해 KDF 분석 모델이 예측에 성공한 경우만 고려하여 평가 데이터 5세트 정확도의 평균 및 TPR, FPR을 계산하고 신경망 모델 중 최적모델로 결정된 모델(2333)의 성능과 비교하였다. 그 결과는 <표 4>와 (그림 9)에 제시한 바와 같다.

KDF 모델은 예측가능율이 신경망 모델보다 현저히 낮아 실용화하기 어려우며 정확도 역시 신경망 모델이 두 배 이상 높고 ROC 공간의 TPR, FPR분포 또한 KDF 분석 모델은 Random 예측선 아래에 존재한 것으로 나타나 신경망 모델이 월등히 뛰어난 것을 확인할 수 있다.

이와 같이 신경망 모델이 KDF 모델보다 우수한 것은 첫째 모델의 입력요소의 차이이다. KDF는 풍수해에 직접적 요인인 기상상황(누적강수량과 최대풍속)만 고려하였지만 풍수해는 지역의 지형적 특징에 따라 피해가 다를 수 있어 피해의 원인을 조사할 때 피해지역 주변의 지형적 특징을 살피고 피해원인 규명 및 대책방안을 마련한다.

〈표 4〉 KDF 분석 모델 예측 가능성

평가세트	데이터 개수	예측실패	예측가능	예측가능율 (%)	정확도 (Accuracy)	민감도 (TPR)	1-특이도 (FPR)	비고
1	666	570	96	14.4%	0.3646	0.4216	0.5714	
2		572	94	14.1%	0.3298	0.3827	0.3846	
3		560	106	15.9%	0.3774	0.4255	0.75	
4		562	104	15.6%	0.3558	0.4111	0.6666	
5		567	99	14.9%	0.2626	0.2888	0.4	
KDF 모델 평균				15.0%	0.3380	0.3859	0.5545	
신경망 모델				100%	0.6521	0.7216	0.4170	2333모델

자연재해로 인한 피해는 풍수지리적으로 우수한 곳일수록 적어지는 음의 상관관계가 있다는 연구[10] 보고는 지형적 특징이 풍수해와 무관할 수 없음을 암시한다. 또한 과거 풍수해가 빈번하게 발생된 지역은 앞으로 풍수해 발생 가능성이 높으므로 과거 풍수해 발생횟수는 지역의 풍수해 취약정도를 나타낼 수 있는 지표라 할 수 있다.

따라서 본 논문에서는 풍수해와 관련있는 지역의 지형적 특징으로는 지역의 하천(국가, 지방1급, 지방2급, 소하천)개수를 풍수해 취약 정도로는 과거 풍수해 발생횟수를 활용하였다. 하천개수 및 풍수해 발생횟수와 피해액과의 상관성을 분석한 결과 모두(국가하천 개수 제외) 약한 양의 상관관계를 보였으며 상관성 또한 유의하였으므로 각 값의 가중평균을 지역의 풍수해 발생 영향요인인 지역적 특징으로 정의하고 이를 신경망 모델의 입력요소로 추가하였다[9]. 향후 풍수해와 상관성 있는 입력요소를 추가로 발굴하여 모델 정확도를 개선하는 연구가 필요하다.

신경망 모델이 KDF 모델보다 우수한 또 하나의 이유는 풍수해 예측과 같은 복잡한 자연현상을 수학적 모델로 해결

하기 보다는 데이터를 지배하는 경험적 규칙성을 학습하고 획득하는데 다른 기법보다 탁월한 성능을 지니고 있다고 알려진 신경망 알고리즘을 활용하였기 때문이다. 하지만 본 연구에서 제시한 최적모델의 정확도를 개선하기 위해서는 학습 데이터 편향성 제거와 신경망 모델 구조 및 매개변수의 지속적인 최적화 연구가 수행되어야 할 것이다.

5. 결론 및 향후 연구계획

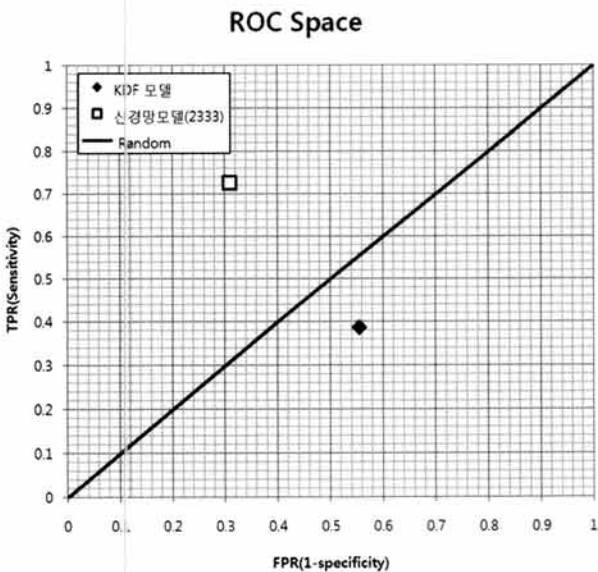
본 연구는 신경망을 활용한 풍수해 예측 모델을 제안하였다. 신경망 모델의 입력 변수로 누적강우량, 최대풍속, 선행강우량, 풍수해 발생에 영향을 줄 수 있는 지역적 특징 값을 사용하였고 출력변수인 총 피해액의 실측치 예측의 어려움이 크므로 분류 문제로 전환하여 모델을 평가함으로써 이전 모델[5]보다 정교한 모델을 생성할 수 있었다.

또한, 풍수해 예측을 위한 최적의 신경망 모델을 생성하기 위하여 신경망의 아키텍처 중에서 초기 가중치, 은닉층 노드의 수, 그리고 사용자 매개변수인 모멘텀과 학습률을 8,000여 개의 다양한 조합을 이용하여 모델을 생성하였고 최적 모델을 선정하기 위해 정확도와 ROC 그래프를 이용하였다.

이렇게 선정된 최적의 신경망 모델과 TCDIS에 적용된 KDF 분석모델의 성능을 비교하였다. 그 결과 신경망 모델의 강건함이 KDF 분석모델보다 우수하였으며 정확도 역시 KDF 모델보다 높았다. 이는 재해위험 상황판단을 지원하는 시스템에 신경망 모델을 적용함으로써 강건하고 정확한 시스템 구축을 가능하게 할 것으로 판단된다.

향후 본 연구에서 제안하는 풍수해 예측 신경망 모델을 기준으로 풍수해 예측 모델의 정확도 향상을 위해 풍수해와 관련성 있을 특징들을 좀 더 검토하고 발견하여 입력 변수 선정 연구가 더 필요하다. 또한 최적의 풍수해 예측 신경망 모델을 정의하기 위해서 본 연구에서 고려하지 않았던 은닉층의 개수, 활성화 함수 등을 고려한 후속 연구가 필요할 것이다.

신경망은 예측 및 분류 등의 문제를 학습하여 해결하는데 그 우수성이 증명되었으나, 그 동안 재난 및 재해의 위험성



(그림 9) 신경망 모델과 KDF 모델의 평가 데이터 적용 결과

을 예측하는 연구에는 활용되지 못했다. 본 연구는 신경망을 이용한 풍수해 예측 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.

참 고 문 헌

[1] Hegazy, T., and Moselhi, O. "Analogy-based solution to markup estimation problem", *ASCE Journal of Computing Civil Engineering*, Vol.8, No.1, pp.72-87, 1994.

[2] "리스크 곡선을 활용한 재난발생 특성분석 방안 연구", 국립방재연구소, 2008.

[3] "웹GIS를 활용한 국내 재해정보DB 및 태풍위원회 재해정보시스템 구축", 국립방재연구소, 2008.

[4] 김광희, 안성훈, 조형근, "신경망과 유전자알고리즘을 이용한 공사비예측 모델의 예측정확도 비교에 관한 연구-공동주택 공사비를 중심으로-", 대한건축학회논문집, Vol.20, No.2, pp.81-89, 2004.

[5] 최선화, "신경망과 유전자 알고리즘을 이용한 자연재해 피해예측 모델 연구", 2010 한국컴퓨터종합학술발표 논문집, Vol.37, No.1, pp.380~384, 2010.

[6] 양원직, 이원호, "초고층 건축물의 부등축소량 예측을 위한 새로운 알고리즘", 대한건축학회 논문집, Vol.23, No.4, pp.35-42, 2007.

[7] 윤여창, "신경망의 개선된 학습 과정을 이용한 구조인자들의 효과 비교", 2008 한국정보과학회 학술발표 논문집, Vol.35, No.2(C), pp.272-276, 2008.

[8] Scott, D., "On optimal and data-based histograms", *Biometrika* Vol.66, pp.605-610, 1979.

[9] 최선화, "풍수해 예측을 위한 신경망 모델", 정보과학회논문지, Vol.38, No.3, pp.115-123, 2011.

[10] 신진동, 이우종, 이창수, "입항성격 및 시기에 따른 전통마을 입지특성 연구", 대한국토·도시계획학회논문지, Vol.43, No.1, pp.7-25, 2008.

[11] Andres Krogh and Jesper Vedelsby, "Neural Network Ensembles, Cross Validation, and Active Learning", *Advances in Neural Information Processing Systems 7*, MIT Press, pp.231-238, 1995.

[12] Liu, D., Chang, T.S. and Zhang, Y., "A New Learning Algorithm for Feedforward Neural Networks", *Proceedings of IEEE International Symposium on Intelligent Control*, pp.39-44, 2001.

[13] Gunaseeli, M. and Karthikeyan, N., "A Constructive Approach of Modified Standard Backpropagation Algorithm with Optimum Initialization for Feedforward Neural Networks", *Proceedings of International Conference on Computational Intelligence and Multimedia Applications*, pp.325-331, 2007.

[14] Taylor, T.J., *Neural Networks and Their Applications*, John Wiley and Sons, Chichester, 1996.

[15] 장병탁, "차세대 기계학습 기술", 정보과학회지 Vol.25, No.3, pp.96-107, 2007.



최 선 화

e-mail : shchoi333@gmail.com

1995년 광주대학교 전자계산학과(학사)
 2002년 전남대학교 전산학과(이학석사)
 2006년 전남대학교 전산학과(이학박사)
 2006년~2008년 서울시립대학교 기계정보
 공학과 연구교수

2008년~현 재 소방방재청 국립방재연구소 책임연구원
 관심분야: 언어처리, 시맨틱 검색, 재난정보관리 등