시계열자료의 계층분리기법을 이용한 하천유역의 홍수위 예측

Flood Stage Forecasting using Class Segregation Method of Time Series Data

김성원*swkim68@phenix.dyu.ac.kr

Abstract

In this study, the new methodology which combines Kohonen self-organizing map(KSOM) neural networks model and the conventional neural networks models such as feedforward neural networks model and generalized neural networks model is introduced to forecast flood stage in Nakdong river, Republic of Korea. It is possible to train without output data in KSOM neural networks model. KSOM neural networks model is used to classify the input data before it combines with the conventional neural networks model. Four types of models such as SOM-FFNNM-BP, SOM-GRNNM-GA, FFNNM-BP, and GRNNM-GA are used to train and test performances respectively. From the statistical analysis for training and testing performances, SOM-GRNNM-GA shows the best results compared with the other models such as SOM-FFNNM-BP, FFNNM-BP, and GRNNM-GA and FFNNM-BP shows vice-versa. From this study, we can suggest the new methodology to forecast flood stage and construct flood warning system in river basin.

Key words : Flood stage forecasting, Unsupervised algorithm, Kohonen self-organizing map, Neural networks model

1. 서 론

하천홍수위의 예측은 수자원의 계획과 운영에 있어서 주요한 입력정보를 형성하며, 만약 주어진 수문년 이내에 연속적인 수위측정을 기준으로 이루어진다면 하천홍수위의 예측은 다음 두 가지의 측면에서 도움이 될 것이다. 첫 번째는 홍수시의 임박한 홍수의 예·경보에 대한 실시간 정보를 제공할 수 있으며, 두 번째는 홍수시에 저수지의 방류량을 제어하는데 도움이 될 것이다. 하천 홍수위의 자료는 일반적으로 자동계측기에 의하여 계측에 편리한 하천수위표지점에서 산정할 수 있으며, 이러한 자료 및 이 자료에 기초를 둔 장래홍수사상의 예측은 강우, 침투와 같은 토양특성과 같은 매개변수들의 변동이 비선형으로 하천홍수위에 영향을 많이 끼치기 때문에 매우 어려운 작업이라고 할 수 있다.

본 연구에서는 낙동강유역의 주요 수위표지점인 진동지점에서 하천홍수위의 예측을 위하여 코호넨 자기조직화함수(Kohonen self-organizing map, KSOM) 신경망모형을 보편적인 신경망모형과 연계한 운영기법을 제시하였으며, 입력층 노드로는 고령교 및 적포교 지점의 하천홍수위 자료를 이용하였다. 또한 제시한 모형의 운영기법에 대한 훈련 및 검증과정을 통하여 최적 매개변수의 선정 및 모형의 적합성을 검토하였으며, 실시간 홍수 예·경보 시스템 구축에 이용할 수 있을 것으로 판단된다.

2. 코호넨 자기조직화 함수(Kohonen self-organizing map, KSOM)

2.1 코호넨 자기조직화 함수의 기본개념

코호넨 자기조직화함수 신경망모형(Kohonen self-organizing map neural networks model, KSOM-NNM)은 훈련에 있어서 입력자료를 유사한 몇 개의 그룹(Class)으로 분류한다. 이것은 코호넨층의 연결강도를 적합시키기 위하여 수행되어 지며, 유사한 입력자료는 동일한 코호넨층의 노드를 활성화시킨다. 입력자료의 선

* 정회원, 공학박사, 수자원개발기술사, 동양대학교 철도토목학과 조교수 ; E-mail : swkim68@phenix.dyu.ac.kr

행과정은 매우 유용하며, 아것은 네트워크에 입력자료를 적용하기 전에 정규화된 모든 입력자료를 포함하고 있다. 코호넨층을 설정할 경우에는 첫 번째로 연결강도 벡터는 적절하게 초기화 되어져야 하며, 이것은 연결강도 벡터가 무작위 방향인 것을 나타내고 있다. 두 번째로 연결강도 벡터와 입력자료는 통일하여 정규화되어야 한다. 그러한 정규화는 임의 상황에서는 정보손실의 원인이 되고, 만약 발생할 경우에는 처리할 수 있는 다수의 방법이 존재한다. 연결강도 벡터가 무작위로 분포되어 있고, 그 이후에 각 노드의 연결강도 벡터가 입력자료에 얼마나 가깝게 있는지를 결정한다고 가정하자. 그때 노드는 훈련의 권리에 대하여 경쟁하며, 입력자료와 연결강도 벡터의 가장 큰 내적(Inner product)을 가진 노드는 승자(Winner)라고 지칭한다. 이 노드는 출력값을 발생시키도록 허락되어 질 수 있는 단 하나의 노드이며, 모든 다른 노드의 출력치는 0으로 고정된다(Kohonen, 1982, 1990; Tsoukalas and Uhrig, 1997; 김대수, 1996).

2.2 코호넨 자기조직화 함수의 훈련과정

경쟁학습에서는 네트워크의 노드는 서로다른 입력자료에 대하여 점차적으로 민감하게 되어가는 것을 채택한다. KSOM 신경망모형에서는 일반적으로 2개의 층 즉 입력층과 코호넨층으로 구성되어 있다. 입력층의 각 노드는 보통 2차원으로 형성된 코호넨층에 완전히 연결되어 있으며, 코호넨층의 진행인자(Processiong elements, PE) 혹은 출력노드중 어떠한 것도 서로서로 연결되어 있는 것은 없다. 코호넨층에서의 PE는 입력자료에 대하여 그들의 연결강도의 거리를 측정한다.

3. 보편적인 신경망모형의 적용

3.1 일반화된 회귀신경망모형(Generalized regression neural networks model, GRNNM)

본 연구에 적용된 일반화된 회귀신경망모형(GRNNM)은 반경기초함수 신경망모형(Radial basis function neural networks model, RBFNNM)의 변형된 형태의 신경망모형이다. GRNNM은 입력층, 은닉층, 합산층 및 출력층의 4개의 층으로 구성되었으며, 비선형 회귀이론에 기초를 둔 신경망모형이다. 입력층, 은닉층 및 합산층노드는 완전히 연결된 구조로 형성되어 있으나, 출력층의 노드는 합산층의 노드 중 몇 개로만 연결이 되어 있다. 입력층과 은닉층 사이에서의 훈련과정은 반경기초함수 신경망모형의 경우처럼 자율훈련(Unsupervised training)으로 구성되어서 K-means나 OLS 알고리즘과 같은 특별한 군집알고리즘을 필요로 하며, 군집알고리즘의 반경은 훈련이 시작하기 전에 결정되어 져야 한다. 또한 은닉층과 합산층 사이에서의 훈련과정은 은닉층에서의 출력값에 대하여 평균제곱오차의 최소화과정에 기초를 둔 지도훈련(Supervised training)으로 구성된다(김성원 등, 2001; Kim and Jee, 2006; Specht, 1991; Tsoukalas and Uhrig, 1997; Wasserman, 1993).

3.2 전방향 신경망모형(Feedforward neural networks model, FFNNM)

전방향 신경망모형은 입력층, 출력층 및 입력층과 출력층 사이에 한개 혹은 그 이상의 은닉층을 가지고 있다. 각 층에서의 노드는 다음 층의 모든 노드에 완전히 연결되어 있으며, 한 층에서의 노드는 인접한 다음 층의 노드에만 연결되어 있다. 하나의 노드에서 다음 노드로의 전달하는 신호의 힘은 상호연결 사이에의 연결강도에 달려있다. 전방향 신경망모형은 일반적으로 역전파 훈련 알고리즘을 이용하여 훈련된다. 훈련수행은 연결강도와 편차를 적합시키는 과정이며, 그 결과는 원하는 가장 최적의 값에 일치시키게 된다(Simpson, 1990; Haykin, 1994).

4. 신경망모형의 구성 및 적용

4.1 분석대상지점 및 수문자료의 선정

본 연구에서의 신경망모형을 적용하기 위한 주요 수위표지점은 낙동강유역의 본류에 위치한 진동 수위표지점으로 선정하였으며, 입력자료로는 고령교 및 적포교 수위표지점의 홍수위자료를 이용하였다. 따라서 각 수위표지점의 홍수위사상은 기왕의 자료중에서 자료의 신뢰성이 높은 4개의 시간 홍수위자료를 선택하였다. 또한 홍수위사상 1-3까지를 신경망모형의 훈련에 사용하였으며, 홍수위사상 4를 신경망모형의 검증에 사용하

였다. 본 연구에 사용된 자료는 국가 수자원관리종합정보 시스템(Water management information system, WAMIS)의 홈페이지(www.wamis.go.kr)의 무료 공개자료를 이용하였다. 다음 표 1은 본 연구에 사용된 홍수위자료를 나타내고 있다.

4.2 훈련과정

본 연구에서의 신경망모형은 크게 4가지의 종류로 구분하였다. 첫 번째는 훈련의 선행과정으로 KSOM 신경망모형에 의해 최적 분류된 자료를 사용하는 경우 및 실측자료를 모두 사용하는 경우이다. 두 번째는 훈련에 사용되는 보편적인 신경망모형이 일반화된 회귀신경망모형(GRNNM) 및 전방향 신경망모형(FFNNM)의 경우이다. 마지막으로 신경망모형의 훈련을 위하여 역전파 훈련알고리즘(Backpropagation, BP) 및 유전자 알고리즘(Genetic algorithm, GA)을 사용하는 경우이다. 따라서 본 연구에 적용되는 신경망모형은 SOM-FFNNM-BP, SOM-GRNNM-GA, FFNNM-BP 및 GRNNM-GA로 4종류로 구성된다. 본 연구에서 유전자 알고리즘은 각 입력층 노드에 대한 적절한 개별적인 평활인자(Smoothing factor)뿐만 아니라 전반적인 평활인자를 결정하는데 사용된다. 유전자 알고리즘을 이용한 GRNNM에 대한 훈련 수행과정은 크게 두 부분으로 진행되어 진다. 첫 번째는 훈련자료를 이용하여 GRNNM을 훈련시키는 것이며, 두 번째는 첫 번째 과정에서 형성된 GRNNM을 이용하여 테스트자료에 대하여 가장 최적운영을 위한 평활인자의 전 범위를 테스트하는 것이다. 유전자 알고리즘은 각 입력층 노드에 대한 평활인자 승수(Multiplier)를 찾고 있으며, 비록 훈련에 대하여 탐색기법으로 유전자 알고리즘을 사용하지 않은 경우보다 시간은 좀 더 걸리지만 테스트자료에 대하여 가장 양호하게 운영될 수 있는 GRNNM을 생산할 수 있다(Neuroshell 2, 1993). KSOM 신경망모형의 선행과정은 학습율을 0.5, 초기연결강도 0.3, 반복횟수 10,000회 및 이웃의 크기는 3으로 각각 설정하였으며, FFNNM-BP는 학습율과 모멘텀을 각각 0.1, 초기 연결강도를 0.3으로 설정하였으며, 훈련조건은 반복횟수 5,000회 및 평균제곱오차 0.001로 설정하여 이들 중 어느 하나에 도달하면 훈련은 정지한다. 또한 FFNNM-BP의 경우 훈련동안에 훈련자료가 아닌 검증자료에 최적값을 산정하도록 설정하였다. 이런 경우에는 검증자료의 통계분석결과가 훈련자료의 경우보다 양호한 결과치를 나타낼 수도 있다.

4.3 검증과정

본 연구의 검증과정은 SOM-FFNNM-BP, SOM-GRNNM-GA, FFNNM-BP 및 GRNNM-GA의 과적합문제(Overfitting problem)를 극복하기 위하여 교차검증법(Cross-validation method)을 적용하였다. 교차검증법은 최소 RMSE에 도달할 때까지 모든 훈련자료를 훈련시키게 하는 것이 아니고, 각 훈련단계의 마지막에 테스트자료로서 교차검증하는 방법이다. 만약 과접합문제가 발생되면 테스트자료의 평균제곱오차의 수렴과정은 훈련자료가 여전히 훈련되고 있는 경우에도 감소하지 않고 증가하는 경향을 나타낼 것이다(Bishop, 1994; Haykin, 1994; 김성원, 2005). 다음 그림 2의 (a)-(b)는 진동 수위표지점에 대한 홍수위예측의 훈련 및 검증결과에 대한 비교를 나타낸 것이다. 또한 다음 표 2는 SOM-FFNNM-BP, SOM-GRNNM-GA, FFNNM-BP 및 GRNNM-GA의 훈련 및 검증에 대한 통계분석 결과를 나타내고 있다. 그림 2(b)의 경우 가장 최적의 통계결과치를 나타내는 SOM-GRNNM-GA 조차도 첨두부의 홍수위를 예측하는데 있어서 상당한 오차를 나타내고 있는 것으로 나타났는데, 이는 훈련과정에 사용된 자료의 특성에 기인한다고 판단된다. 즉 검증자료의 경우 태풍 매미가 내습할 당시에 나타난 첨두홍수위가 훈련자료에서는 발생하지 않아서 본 연구에서 제시된 신경망모형들이 광범위한 자료를 이용하여 훈련하지 못한 결과로 판단된다.

5. 결 론

본 연구에서는 하천유역에서의 홍수위의 예측을 위하여 기존의 신경망모형과는 다른 적용기법을 이용하였다. 자율훈련 알고리즘을 가진 KSOM 신경망모형을 이용하여 자료의 분류를 통한 개별적인 훈련을 선행과정으로 도입하였다. 따라서 본 연구에서 구성한 SOM-FFNNM-BP, SOM-GRNNM-GA, FFNNM-BP 및 GRNNM-GA의 경우 KSOM 신경망모형을 선행조건으로 구성한 SOM-GRNNM-GA의 훈련 및 검증에 대한 통계분석결과가 가장 양호한 것으로 나타났으며, FFNNM-BP의 통계분석결과가 가장 양호하지 않은 것으로

나타났다. 그러나 본 연구에서는 홍수위예측에 대한 시간적인 분포에 해당하는 선행시간의 고려가 전혀 없었으며, 입력자료의 홍수위지점도 2개 지점만 선택하였고 홍수위 사상도 선택된 몇 개만 이용하였다. 따라서 언급한 사항에 대하여 모두 고려할 경우 본 연구에서 분석된 결과와 다른 결과를 유도할 수도 있을 것이다. 하지만 본 연구에서 제시된 KSOM 신경망모형의 선행과정은 홍수위예측에 있어서 양호한 결과를 유도할 것으로 판단되며, 홍수위예측의 새로운 방법론으로 제시할 수 있을 것으로 판단된다.

참고문헌

1. 김대수(1996). **신경망이론과 응용(Ⅰ)**, 하이테크정보, 서울.
2. 김성원(2005). "신경망모형에 의한 홍수위예측의 신뢰성분석 1. 모형의 개발 및 적용." **대한토목학회 논문집**, 대한토목학회, 제 25권, 제 6B호, pp. 473-482.
3. 김성원, 이순탁, 조정석(2001). "중소하천유역에서 Hybrid Neural Networks에 의한 수문학적 예측." **한국수자원학회 논문집**, 한국수자원학회, 제 34권, 제 4호, pp. 303-316.
4. Bishop, C.M.(1994). "Neural networks and their applications." *Rev. Scien. Instru.* Vol. 65, pp. 1803-1832.
5. Haykin, S.(1994). *Neural networks : A comprehensive foundation*, Macmillan College Pub. Comp., Inc., MA.
6. Kim, S., and Jee, H.(2006). "An expansion of the ungaged pan evaporation using neural networks model in rural regions, South Korea." *Proc. World Environmental & Water Resources Congress 2006*, ASCE/EWRI, Omaha, NE. [Printed in CD]
7. Kohonen, T.(1982). "Self-organized formation of topologically correct feature maps." *Biological Cybernetics*, Vol. 43, pp. 59-69.
8. Kohonen, T.(1990). "The self-organizing map." *Proc. IEEE 78* , Vol. 78, No. 9, pp. 1464-1480.
9. Neuroshell 2(1993). Ward systems group, Inc., MD.
10. Simpson, P.K.(1990). *Artificial neural systems : foundation, paradigms, applications and implementations,* Pergamom, New York.
11. Specht, D.F.(1991). "A general regression neural network." *IEEE Trans. on Neural Networks*, Vol. 2, No. 6, pp. 568-576.
12. Tsoukalas, L.H. and Uhrig, R.E.(1997). *Fuzzy and neural approaches in engineering,* John Wiley & Sons Incorporated, New York.
13. Wasserman, P.D.(1993). *Advanced methods in neural computing*, Van Nostrand Reinhold, New York.

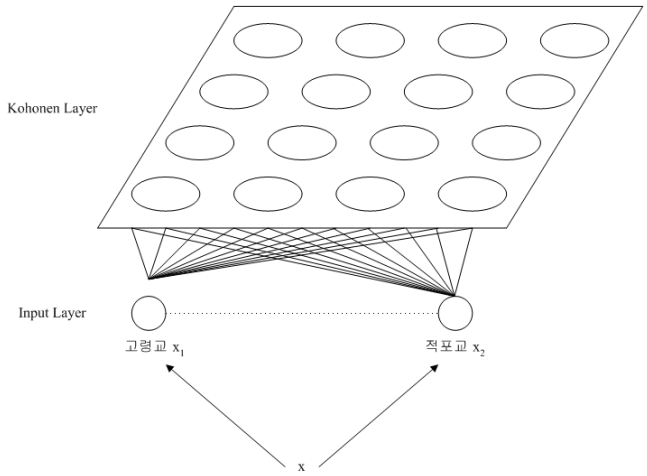

그림 1. 입력층에 2개의 입력변수와 코호넨층에 4-by-4의 PE로 이루어진 KSOM 신경망모형

그림 2(a) 홍수위의 비교(진동지점, 훈련자료)　　**그림 2(b) 홍수위의 비교(진동지점, 검증자료)**

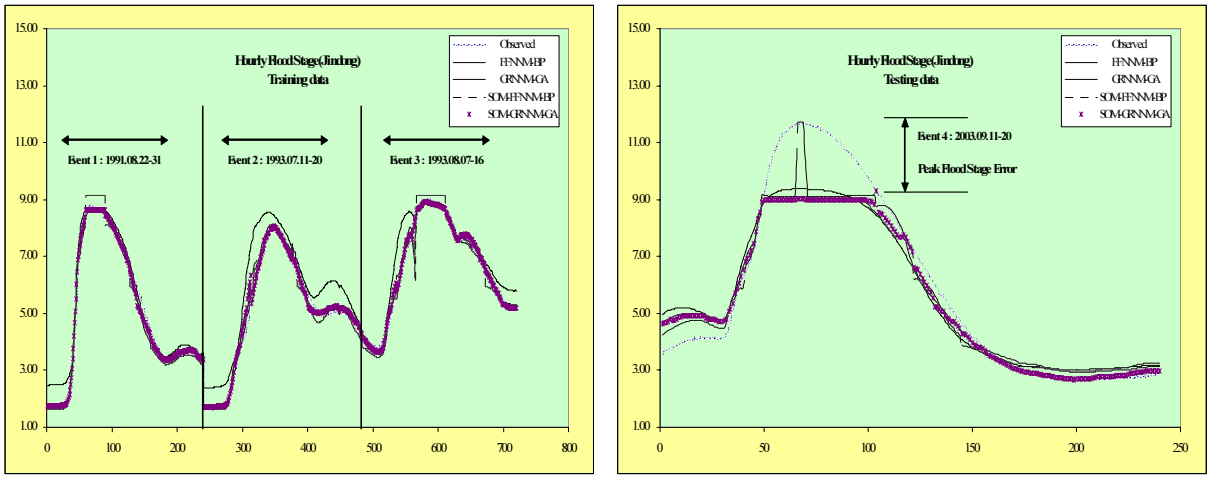

표 1. 본 연구에 사용된 홍수위자료

지점	일시	총 시간 (Hr)	시간최대수위 (m)	비고
고령교	1991.08.22 - 08.31	240	8.35	
	1993.07.11 - 07.20	240	8.21	
	1993.08.07 - 08.16	240	9.05	
	2003.09.11 - 09.20	240	12.07	태풍 매미
적포교	1991.08.22 - 08.31	240	8.34	
	1993.07.11 - 07.20	240	7.90	
	1993.08.07 - 08.16	240	8.97	
	2003.09.11 - 09.20	240	11.82	태풍 매미
진 동	1991.08.22 - 08.31	240	8.78	
	1993.07.11 - 07.20	240	8.01	
	1993.08.07 - 08.16	240	9.01	
	2003.09.11 - 09.20	240	11.70	태풍 매미

표 2. 각 신경망모형의 통계분석 결과 (진동지점)

지점	통계특성치	SOM-FFNNM-BP		SOM-GRNNM-GA		FFNNM-BP		GRNNM-GA	
		Training	Testing	Training	Testing	Training	Testing	Training	Testing
진동	CC	0.991	0.980	0.997	0.976	0.974	0.977	0.996	0.977
	RMSE(m)	0.312	0.969	0.172	1.034	0.625	1.007	0.190	1.005
	E	0.979	0.908	0.994	0.895	0.916	0.901	0.992	0.901
	MAE(m)	0.227	0.649	0.107	0.659	0.449	0.786	0.146	0.799