

인공신경망기법에 상관계수를 고려한 서울 강우관측 지점 간의 강우보완 및 예측

Rainfall Adjust and Forecasting in Seoul Using a Artificial Neural Network Technique Including a Correlation Coefficient

안정환* · 정희선** · 박인찬*** · 조원철****

Jung, Heesun · Ahn, Jeongwhan · Park, Inchan · Cho, Woncheol

Abstract

In this study, rainfall adjust and forecasting using artificial neural network(ANN) which includes a correlation coefficient is application in Seoul region. It analyzed one-hour rainfall data which has been reported in 25 region in seoul during from 2000 to 2006 at rainfall observatory by AWS. The ANN learning algorithm apply for input data that each region using cross-correlation will use the highest correlation coefficient region. In addition, rainfall adjust analyzed the minimum error based on correlation coefficient and determination coefficient related to the input region.

ANN model used back-propagation algorithm for learning algorithm. In case of the back-propagation algorithm, many attempts and efforts are required to find the optimum neural network structure as applied model.

This is calculated similar to the observed rainfall that the correlation coefficient was 0.98 in missing rainfall adjust at 10 region. As a result, ANN model has been for suitable for rainfall adjust. It is considered that the result will be more accurate when it includes climate data affecting rainfall.

Key Words : Artificial neural network(ANN), Rainfall forecasting, correlation-oefficient, determination coefficient, back-propagation algorithm

1. 서 론

최근 들어 지구온난화가 계속되면서 매년 한반도 지역에 장마 전·후로 국지성 집중호우로 인한 홍수피해가 발생하고 있다. 그러나 우리나라의 경우 도시지역에 대한 홍수 예·경보 업무는 거의 이루어지지 않고 있는 실정이다. 호우로 인한 피해를 예방하기 위해서는 홍수량의 정확한 예측이 필요하며, 그 정확도를 높이기 위해서는 강우 결측치에 대한 보완과 강우의 정확한 예측이 필요하다.

본 연구의 목적은 1시간강우 자료를 이용하여 지점간의 상관분석을 실시하고, 결측 지점의 강우 보완 및 예측을 하는데 있다. 강우의 일반적 특성을 고려해 볼 때 일정 유역면적을 가진 유역 전체에 균일한 강우가 발생될 경우는 드물다. 강우 지속시간 동안의 강우량은 호우 중심지역으로부터 면적이 증가할수록 점차 감소하게 된다.

본 연구에서는 서울지역 25개 AWS 관측소에서 관측된 시강우 자료로부터 서울지역 강우의 공간분포 특성을 분석하였다. 입력강우자료와 대상지역의 예측 강우량간의 비선형 상관관계를 잘 고려 할 수 있는 인공신경망 기법을 적용하여 그 결과를 분석함으로써 강우 결측치 보완 및 강우예측에 대한 인공신경망기법의

* 연세대학교 대학원 토목공학과 박사과정 · 공학석사 · E-mail:daliza@yonsei.ac.kr
** 한국시설안전관리공단 · 공학석사 · E-mail:jhs@yonsei.ac.kr - 발표자
*** 연세대학교 대학원 토목공학과 박사과정 · 공학석사 · E-mail:icpark@yonsei.ac.kr
**** 정희원 · 연세대학교 사회환경시스템공학부 교수 · 공학박사 · E-mail:woncheol@yonsei.ac.kr

실제 적용가능성을 검토하고자 한다.

2. 연구방법 및 범위

본 연구에서는 서울지역 25개 AWS 관측소에서 관측된 2000년 5월~2006년 10월 시자료를 기상청을 통해 확보하였고, 2006년 7월 27~29일 강우사상을 보완 및 예측 하였으며 다음과 같은 절차로 행해진다.

1. 서울지역의 25개 AWS 관측소로부터 시우량 자료 수집
-일 지점 일강우량 20mm이상인 동시 강우 사상을 연구대상으로 선정
2. 시우량 자료로부터 기본통계분석
3. 교차상관을 이용하여 지점별 강우량 자료의 상관성 분석
4. 강우의 공간분포 특성을 분석
5. 각 지점의 T시간 자료간의 상관성 분석
6. 교차상관계수가 0.8이상인 자료를 선정
7. 선정된 자료를 이용하여 인공신경망 알고리즘을 학습
8. 학습된 인공신경망알고리즘을 이용하여 T시간대 결측치 보완
9. 인공신경망 기법을 이용하여 이후의 강우를 예측한 후 실측값과의 비교를 통해 그 적합성 여부를 판단

3. 모형의 적용

3.1 대상지역

본 연구의 대상지역으로 서울지역을 선정하였다. 현재 서울지역의 경우 각 구마다 기상청에서 관리하는 28개 AWS 관측소를 바탕으로 분단위의 강수량자료를 수집하고 있으며, 영등포와 한강지점에는 각각 두 개의 강우기록계를 설치하여 운영하고 있다. 따라서 영등포와 한강지점에 대해서는 지점의 관측소중 각각 한 개만 사용하였고 28개 관측소 중에서 과천지점을 제외한 총 25개 지점을 선정하였다.

3.2 자료 수집 및 분석

모형의 학습 및 예측을 위해서 서울시 25개 AWS 관측소에 설치되어 있는 자동기상관측장비로 측정된 2000년 5월~2006년 7월까지의 일시우량 자료를 사용하였다. 서울지역의 25개 지점 기상관측소에서 관측된 2000~2006년 시강우의 자료에서 무 강우를 제외한 실질적 강우사상을 채택하여, 강우지속기간과 강우량에 대한 대표적인 통계값인 평균(mean), 분산(variance), 표준편차(standard deviation), 최대(max), 최소(min)와 같은 기본 통계치를 산정하고 그림 1과 같이 전체적인 분포도를 파악하였다.

교차 상관분석에 의한 이웃 관측소를 설정하는 방법은 시계열자료에 대한 지역화분석으로, 다른 용어로 군집화 분석의 첫 단계에 해당되는 매우 중요한 작업이다. 이웃관측소를 군집하기 위하여 0.6~0.8 사이의 적절한 상관계수를 정하였으며, 이 거리는 상호상관거리(correlation distance)로 불리며 하나의 관측소를 중심으로 이웃관측소를 군집화할 때 동일군집 영역의 반지름 결정에 기준이 된다. 그림 2에 대상 관측소와 그 외 관측소간의 교차상관계수를 도식화하여 나타내었다.

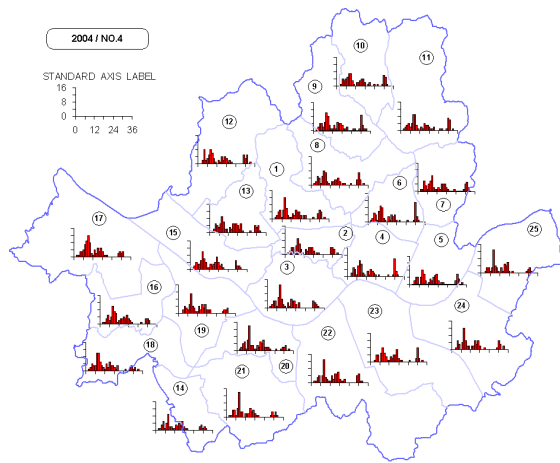


그림 1. 강우 공간분포 도식화

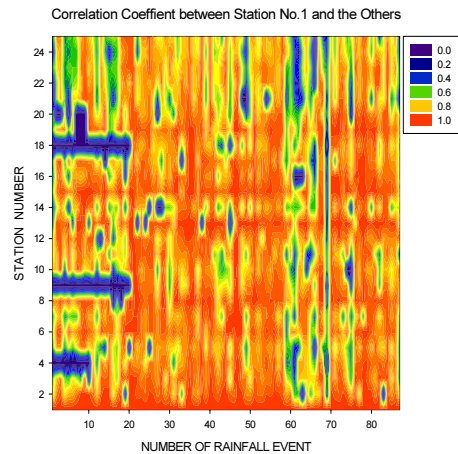


그림 2. No.1 관측소와 그 외 관측소간의 주요 강우사상별 교차상관계수

3.2 자료의 선정

각 지점에 대한 기본적인 통계분석과 상관분석을 실시하여 선정된 학습강우 사상은 표 1과 같다. 여기서 입력층 처리소자나 출력층 처리소자의 개수는 입출력패턴에 따라 정해지나 은닉층 처리소자의 개수를 결정하기 위한 명확한 기준이 없으므로 은닉층 처리소자의 개수에 따른 학습효과를 분석하기 위해 동일한 반복 횟수(epoch) 5000번에 대해 각 지점별로 은닉층 처리소자의 수를 10개로 두고 모의하였다. 지점1(종로)부터 지점25(강동)까지 대체적으로 해당 지점의 왼쪽지점으로부터 상관계수가 높게 나타났다.

표 1. 학습 강우사상

지 점	입력지점(No.)	지 점	입력지점(No.)
1	2, 3, 6, 8, 13, 15, 19	14	3, 15, 19, 20, 21
2	1, 3, 4, 6, 8, 15, 19	15	1, 2, 12, 13, 16, 17, 18, 19
3	1, 2, 4, 6, 19, 22, 23	16	1, 13, 15, 17, 18, 19
⋮	⋮	⋮	⋮
11	1, 2, 4, 5, 6, 7, 8	24	4, 5, 7, 22, 23, 25
12	1, 8, 13, 15, 16, 17	25	4, 5, 7, 23, 24
13	1, 2, 8, 12, 15, 16, 19		

4. 모형 적용결과

각 층의 연결강도 초기치는 어떠한 값을 사용해도 매개변수가 최적해로 수렴하게 되므로 본 연구에서는 -0.5~0.5사이의 난수를 발생시켜 초기화하였다. 또 적응학습율을 momentum방법과 함께 사용하여 momentum상수 0.7을 다층인공신경망의 학습에 사용하였다. 그리고 분산처리를 통한 효율적인 학습을 위하여 입력층 및 은닉층에 두는 bias를 1로 가정하였다.

각 지점의 부근에 위치한 관측지점들이 가지는 동질성이 양호하면 할수록 정확한 자료를 획득 할 수 있으며, 10개 이상의 부근 관측점을 선택하는 것이 통상이다. 각 모형에 따른 정확성 및 적합도를 알아보기 위해 예측한 결과에 대한 관측결과의 통계적 검증을 실시하였으며 표 2에 제시하였다. 그림 3과 4에 No.1 지점에 대하여 강우 관측치와 모형을 통한 예측치를 비교하여 도시하였다.

표 2. 지점별 결측치 보완결과 정확도의 통계적 검증

구 분	상관계수	결정계수	RMSE	MAPE	TSE
지점1	0.96	0.87	28.60	21.42	-21.42
지점2	0.91	0.81	30.51	21.03	-21.03
지점6	0.98	0.90	22.96	16.51	-16.51
지점10	0.98	0.93	18.94	16.34	-16.34
지점19	0.97	0.88	22.67	19.76	-19.76
지점22	0.95	0.89	23.12	24.13	-24.13

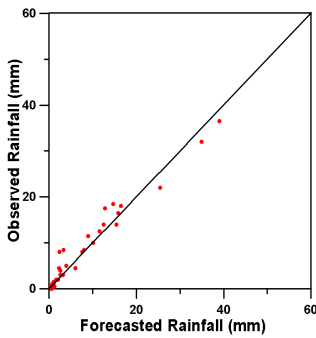


그림 3. N0.1 지점의 강우 예측결과 Scatter diagram

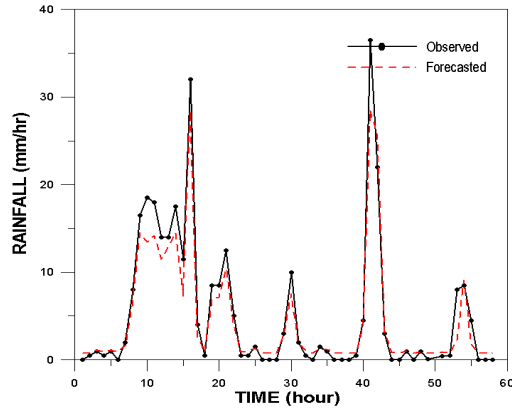


그림 4. No.1 결측치 보완결과

5. 결 론

본 연구에서는 서울지역 AWS 강우관측 지점별 상관계수(Pearson's correlation coefficient)를 분석하고, 인공신경망기법을 이용하여 임의 지점 강우자료의 보완 및 예측의 적합성을 분석하였다. 강우예측을 위하여 다층인공신경망을 구성하고 그 학습을 위하여 오차역전파 알고리즘을 이용하였다. 입·출력패턴에 따라서 오차가 최소화되도록 최적의 인공신경망 구조를 결정하였다. 이와 같은 절차를 통하여 임의 지점 강우자료의 결측치 보완과 예측 가능성을 검토하여 다음과 같은 결론을 도출하였다.

시강우 강우자료의 보완을 위해서 대상지점과의 상관계수가 0.8이상인 지점을 선택하여 인공신경망기법에 적용하였으며, 무작위로 선정된 6개 지점(종로, 중구, 동대문, 도봉, 영등포, 서초)에 대하여 결측치 보완을 모의한 결과 결정계수가 0.90이상으로 그 적합성이 매우 뛰어난 것으로 나타났다.

참고문헌

1. 김광섭 (2006). "상층기상자료와 신경망기법을 이용한 면적강우 예측." 한국수자원학회 논문집, 한국수자원학회, 39(8), 717~726.
2. 김주환 (1992). "신경회로망을 이용한 하천유출량의 수문학적 예측에 관한 연구." 박사학위논문, 인하대학교.
3. French, M.N., Krajewski, W.F., and Cuykendall, R.R. (1992). "Rainfall forecasting in space and time using a neural network." Journal of Hydrology, 137, 1~31.
4. Hsu, K.L., Gupta, H.V., and Soroosh, S. (1995). "Artificial neural network modeling of the rainfall-runoff process." Water Resources Research, AGU, 31(10) 2517~2530.
5. Luk, K.C., Ball, J.E., and Sharma, A. (2000). "A study of optimal model lag and spatial inputs to artificial neural network for rainfall forecasting." Journal of Hydrology, 227, 56~65.