

상호정보량 기법과 인공신경망을 이용한 실시간 강우 자료 보정

Calibration of Real Time Rainfall Data Using Mutual Information and Artificial Neural Network

성경민*, 구여주**, 김태순***, 허준행****

Kyungmin Sung, Yeojoo Goo, Taesoon Kim, Jun-Haeng Heo

요 지

이러한 강우자료의 결측값이나 오자료를 보정하는 것은 그 지역의 정확한 수문학적 특성 파악 및 안전한 수공구조물의 설계에 영향을 미치게 되므로 매우 중요하다고 할 수 있다. 최근 이러한 강우자료를 비선형적 모델인 인공신경망(Artificial Neural Network)을 이용하여 보정하는 연구가 활발히 진행되고 있다(오재우 등, 2008). 그러나 이러한 인공신경망을 적용하는 경우, 선택한 신경망 구조의 형태와 학습(training)을 위해 사용되는 자료가 전체 자료의 특성을 반영하고 있는 정도에 따라 정확도에 차이를 보인다(한광희 등, 2010). 따라서 자료보정을 위한 입력 자료의 선택은 인공신경망을 이용한 결측치 보정의 중요한 과정이다.

본 연구에서는 이러한 입력 자료의 선택을 위한 여러 가지 기법 중 입력 변수간의 상호정보량 (Mutual Information)을 이용한 방법을 적용하여 대상 결측 지점을 보정할 강우지점을 선별한 후 선택된 지점만으로 인공신경망을 구성하여 강우자료를 보정하고 주변 자료를 모두 이용한 결과와 상관성분석으로 얻어진 결과와 비교하였다.

핵심용어 : 상호정보량 기법, 인공신경망, 강우자료 보정

1. 서론

전 지구적인 기후변화로 인하여 날이 갈수록 수문자료의 정확한 해석의 중요성이 강조되고 있다. 특히 실시간으로 관측되는 강우 자료는 수문 자료의 근간이 되는 자료로서 실시간 저수지 운영을 위한 유입량을 구하기 위한 자료로 사용하는 등 매우 중요하게 이용되고 있다. 하지만 주변 여건이나 기기의 오작동 등에 의하여 자료가 결측되거나 오자료가 생기는 사례가 빈번히 발생하고 있어 이러한 자료들을 수정 또는 보정할 필요성이 증대되고 있다. 최근에 이러한 오자료 검보정을 위하여 비선형적 방법인 인공신경망(Artificial Neural Network, ANN)기법을 이용하여 자료를 보정하는 연구가 활발히 진행되고 있다. 또한 이런 인공신경망 기법을 적용하는 경우, 주변 자료 모두를 이용하는 것보다 해당 결측 지점과 상관성이 비교적 큰 지점만을 입력 자료로 선택하는 것이 결측치 보정에 어떠한 영향을 미치는지 알아보려고 하였다. 이런 상관성을 분석하기 위하여 상호정보량 (Mutual Information)기법을 사용하였다.

2. 결측자료 보정기법

2.1 상호정보량 (Mutual Information)

입력 자료의 적절한 선택을 위하여 상호정보량 기법을 적용하였다. 상호정보량은 Shannon의 정보이론에서 제시되었다. 이산 확률변수 X의 각 사건에 대한 상호정보량을 식.(1)과 같이 가중 평균한 것을 평균 정

* 정회원 · 연세대학교 대학원 토목공학과 석사과정, E-mail : kmini33@yonsei.ac.kr

** 정회원 · 연세대학교 대학원 토목공학과 석사과정, E-mail : goo27@yonsei.ac.kr

*** 정회원 · 한국수력원자력, 사원, 공학박사, E-mail : tskim9505@gmail.com

**** 정회원 · 연세대학교 사회환경시스템공학부 토목환경공학과 교수, E-mail : jhheo@yonsei.ac.kr

보량 또는 엔트로피라고 부른다(Shannon, 1948). 엔트로피에 대한 상호정보량은 평균 상호정보량, 또는 기대 상호정보량이라고 부르며 식.(2)와 같이 나타낼 수 있다(Fano, 1961).

$$H(X) = - \sum_{x \in X} p(x) \log p(x) \quad (1)$$

$$MI(X; Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)} \quad (2)$$

두 변수의 연관성을 출현빈도에 근거해서 판단하는 상황에서 정보량과 상호정보량과의 관계를 도식화하면 다음과 같다(이재운, 2003).

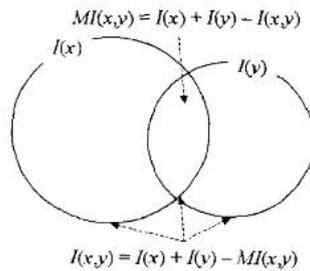


그림 1. 정보량과 상호정보량과의 관계

2.2 인공신경망의 구성

인공신경망은 다음 그림과 같이 입력 자료에 대응하는 출력자료를 하나의 자료세트로 이용하고 신경망을 구성하기 위한 은닉층(Hidden layer)에 존재하는 뉴런(Neuron)의 가중치를 여러 가지 학습기법을 이용하여 조절하여 출력값을 계산하는 형태로 구성된다(김태순 등.,2009). 식.(3)은 학습시간 t에 대한 오차함수를 표현한 것이다. 학습이란 임의의 초기 값에서 다음과 같이 표현되는 오차가 최소가 되도록 연결내의 가중치를 적절히 조정하는 것을 뜻한다.

$$E(t) = \frac{1}{2} \sum (y_i(t) - d_i(t))^2 \quad (3)$$

3. 대상 유역

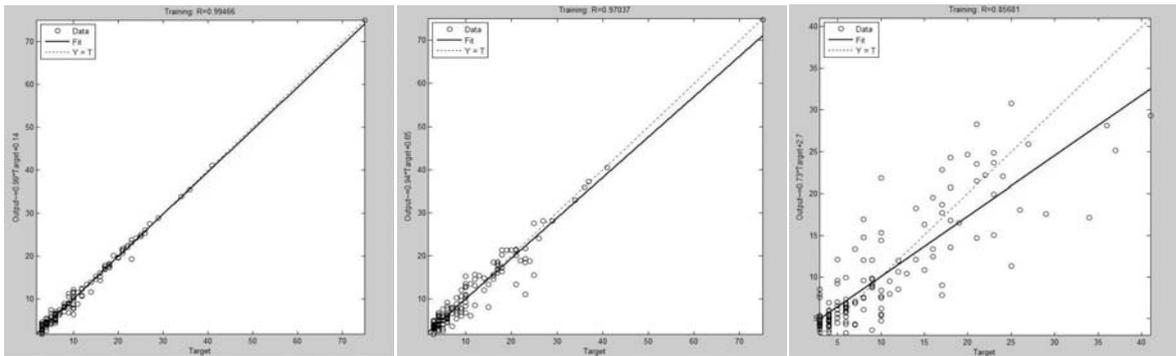
결측치 보정에 이용된 대상지점은 청평댐 유역 내에 있으며, 검토된 강우 관측소는 2005년 1월 1일부터 2009년 12월 31일까지의 시강우 자료를 사용하였다. 청평댐 주변의 여러 강우 관측소 중 가평(RMLT 10134020), 북방(RMLT 10144090), 노천(RMLT 10144080)을 분석 대상 지점으로 선정하고 그 주변의 지점들을 후보지점으로 선정하였다. 선정한 모든 지점의 강우량 자료 중 모두 결측이 아니면서 3mm 보다 크거나 같은 경우의 자료만을 분석 대상으로 추출하였다. 아래의 표 1은 3 개의 대상지점에 관한 후보지점들을 나타낸 표이다. 숫자는 대상지점과 후보지점간의 상관계수를 의미하며, 굵은 글씨에 이탤릭체로 된 지점은 상관성 분석을 통해 선정된 지점이고 밑줄과 빗금 있는 지점은 상호정보량 분석을 통하여 선정된 지점이다.

인공신경망은 모든 후보지점을 대상으로 구성된 구조 및 상관성 분석에 의해 선택된 지점으로 구성된 것, 그리고 상호정보량 분석으로 선택된 지점으로 구성된 세 가지의 구조를 구성하여 결측치 보정을 하였다.

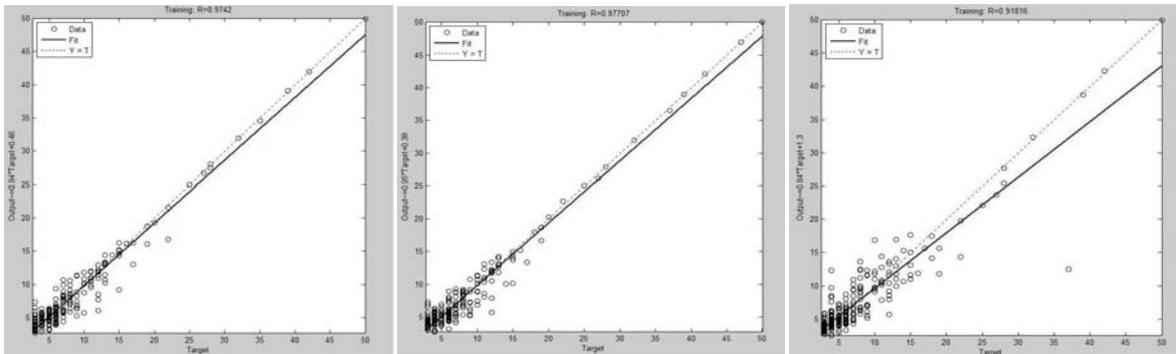
표 1. 선정된 지점의 상관성 분석 및 상호정보량 분석 결과

가평	현리 0.79	가평 0.86	대성리 0.45	남산 0.62	춘천 0.71	화악 0.68	하면 0.84	청평 0.59	노곡 0.58	
북방	용문산 0.59	남산 0.63	청운 0.6	춘천 2 0.57	홍천 2 0.34	화악 0.56	춘천 0.69	홍천 0.57	삼포 0.55	두촌 1 0.67
노천	유치 0.41	홍천 1 0.18	삼포 0.53	내촌 0.49	서석 0.73	춘당 0.47				

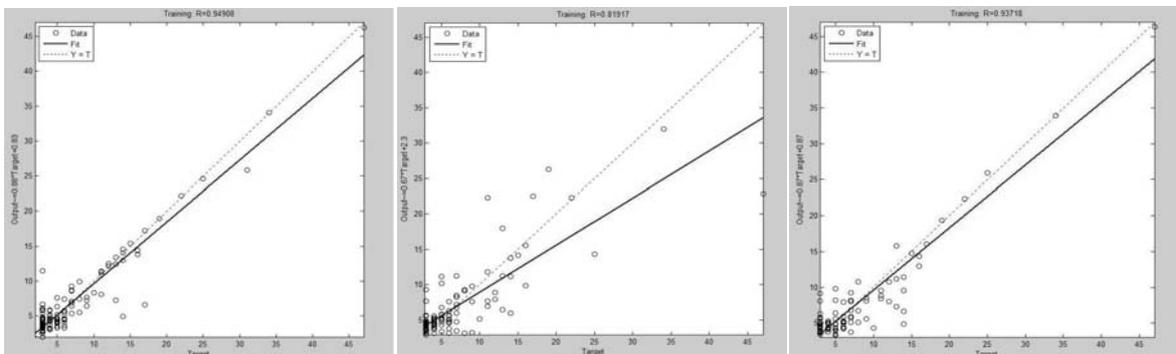
그림 2는 세 가지 구조에 대한 학습결과이며, 그림 3은 테스트 결과를 나타낸 것이다. 학습결과는 전반적으로 세 지점이 모든 구조에서 높은 결정계수를 나타내었다. 노천의 경우는 상호정보량 분석(MI)에 의한 학습결과가 더 높은 결정계수를 가졌으나, 가평과 북방의 경우는 그 외의 결과가 더 좋은 것으로 나타났다.



모든지점 (결정계수 0.99466) 상관성 분석 (결정계수 0.97037) MI (결정계수 0.85681)
(a) 가평 학습 결과

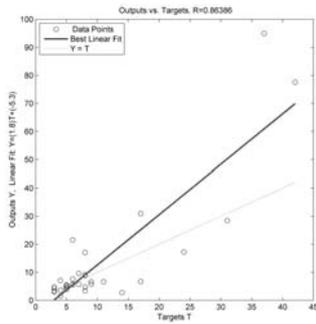


모든지점 (결정계수 0.9742) 상관성 분석 (결정계수 0.97707) MI (결정계수 0.91816)
(b) 북방 학습 결과

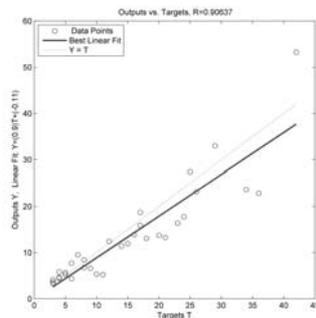


모든지점 (결정계수 0.94908) 상관성 분석 (결정계수 0.81917) MI (결정계수 0.93718)
(c) 노천 학습 결과

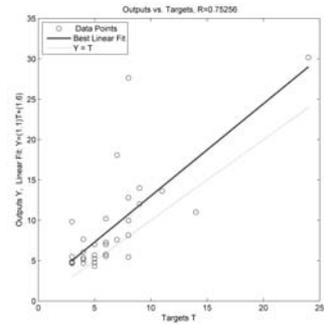
그림 2. 대상 지점의 학습 결과



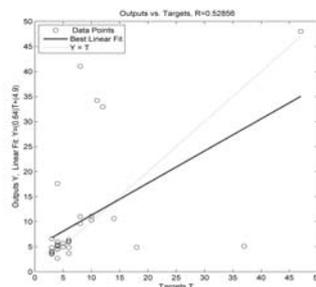
모든지점 (결정계수 0.86386)



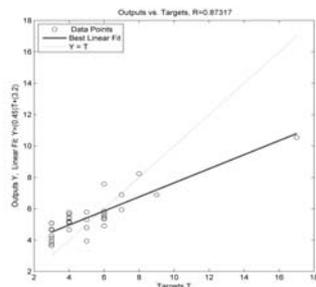
상관성 분석 (결정계수 0.90637)
(a) 가평 테스트 결과



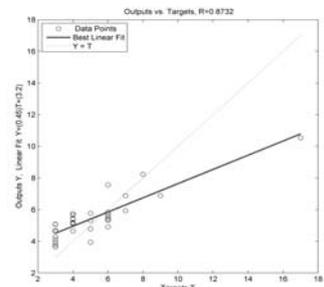
MI (결정계수 0.75256)



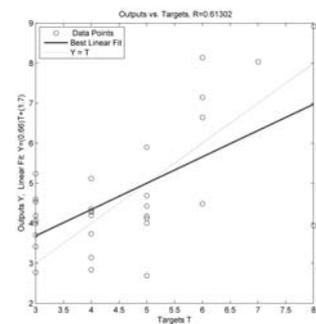
모든지점 (결정계수 0.52856)



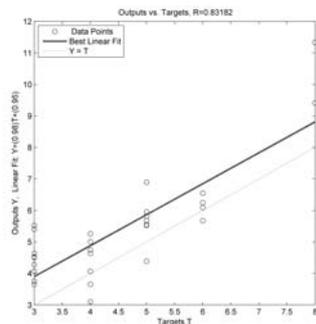
상관성 분석 (결정계수 0.87317)
(b) 북방 테스트 결과



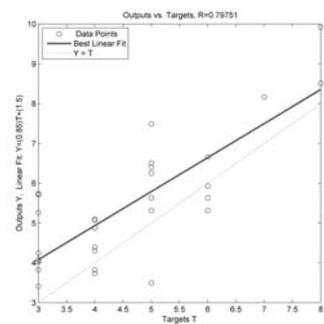
MI (결정계수 0.8732)



모든지점 (결정계수 0.61302)



상관성 분석 (결정계수 0.83182)
(c) 노천 테스트 결과



MI (결정계수 0.79751)

그림 3. 대상지점 테스트 결과

테스팅의 경우는 학습결과에 비해 낮은 결정계수를 가졌으나 전반적으로 모든 지점을 사용한 경우보다는 상관성 분석이나 상호정보량 분석을 통하여 지점을 선별한 결과가 더 좋은 결과를 보였다. 그러나 상관성 분석과 상호정보량 분석에 대한 결과는 특정 분석결과가 더 좋다고 판단할 만큼 일관성을 갖지 않았다. 이렇게 불규칙한 결과를 얻은 이유는 선별된 지점의 개수가 충분하지 못하였기 때문이며, 인공지능망의 결과가 자료의 특성에 크게 의존하기 때문으로 판단된다.

4. 결론

본 연구에서는 인공지능망을 구성하는 입력 자료를 선별하는 과정에서 상호정보량을 이용한 방법을 적용하여 기존의 방법과 비교하였다. 주변 지점을 모두 사용하는 방법에 비하여 테스트에서 우수한 결과를 보였으나, 상관계수를 통한 지점 선별 방법과 비교한 경우 일관성 있는 결과를 얻

지 못하였다. 따라서 인공신경망 구성을 위하여 더욱 향상된 결과를 얻기 위해서는 많은 지점에 대한 적용이 요구되며, 충분한 자료구성을 위한 다른 방법이 도입되어야 할 것으로 판단된다.

감사의 글

본 연구는 국토해양부가 출연하고 한국건설교통기술평가원에서 위탁시행한 건설기술혁신사업(08기술혁신 F01)에 의한 차세대홍수방어기술개발연구단의 연구비 지원에 의해 수행되었습니다.

참고 문헌

1. 김태순, 한광희, 허준행 (2009), 실시간 강우자료의 보정을 위한 인공신경망의 적용. 대한토목학회 정기학술발표논문집, 대한토목학회, pp. 2477-2480.
2. 오재우, 박진혁, 김영국 (2008). 신경망을 이용한 결측 수문자료 추정 및 실시간 자료 보정. 한국수자원학회 논문집, 제41권, 제10호, pp. 1059-1065.
3. 한광희, 류용준, 김태순, 허준행 (2010). 상호정보량 기법을 적용한 인공신경망 입력자료의 선정. 한국수자원학회 논문집, 제43권, 제1호, pp. 81-94.
4. 이재윤 (2003). 상호정보량의 정규화에 대한 연구. 한국문헌정보학회지, 제37권, 제4호, pp.177-198.
5. Fano, R. M. (1961). Transmission of information: A statistical theory of communications. New York: M.I.T press
6. Shannon, C. E. (1948). "A mathematical theory of communication". Bell system technical journal, Vol.27, pp.379-423.