

ANN과 회귀모형의 오차 수정을 통한 댐 유입량 예측 향상

Improving Forecasts of Dam Inflow Using Rescaling Errors From ANN and Regression Model

장선우*, 유지영**, 김태웅***

Sun Woo Jang, Ji young Yoo, Tae-Woong Kim

요 지

수자원이 우리 생활의 전반적으로 중요한 역할을 차지하면서 댐의 효율적인 운영과 안정적인 용수공급에 대한 연구는 지속적으로 수행되어지고 있다. 1990년대 이후 비선형적인 특성을 잘 모의하는 장점을 가진 인공신경망(ANN)을 이용하여 유입량 예측에 대한 많은 연구가 수행되었다. 하지만 ANN 모형을 포함한 회귀모형은 월 강우 및 유입량의 예측에 대해 간편하게 사용을 할 수 있지만, 예측의 정확성에 한계를 가지고 있다. 본 연구에서는 ANN 모형과 회귀모형의 예측오차를 후처리 과정을 통하여 오차를 줄임으로써 예측모형의 성과를 향상시키는 방법을 제안하였다. 연구지역은 금강수계의 대청댐 유역으로, 1982년 9월부터 2005년 12월에 해당하는 유역 내 11개 지점의 강우관측소에서 관측한 월 강우와 댐 유입량을 수집하여 모형을 구축하였다. 강우량과 유입량 자료에 대해 자기상관함수와 교차상관함수를 이용하여 입력변수를 결정하였고, 정규화를 통한 전처리 과정을 거쳐 ANN 모형과 회귀모형을 이용한 예측모형을 구축하였으며, 예측성과의 향상을 위하여 군집 분석을 이용하여 오차를 재조정하였다. 이러한 오차 후처리 과정을 포함한 모형은 RMSE와 상관계수를 이용하여 비교 평가한 결과, 예측성과를 약 40% 정도 향상시켰다.

핵심용어 : 인공신경망모형, 회귀모형, 군집분석, 오차후처리

1. 서 론

우리나라 수자원 총량 중 약 14%는 댐 등 저수지에서 공급되고 있다(건설교통부, 2006). 따라서, 댐 유입량은 우리나라의 수자원에 있어서 중요한 공급처 중의 하나이고, 최근 기후변화로 인한 강우량의 불확실성으로 인해 댐 유입량 예측에 대한 관심은 높아지고 있다. 댐 유입량 예측의 방법은 여러 가지 인과 관계를 고려한 물리적인 방법과 기존의 유출량 자료를 이용한 통계학적인 접근방법이 있다. 물리적인 방법은 강우, 토양수분, 증발산, 토지이용 등의 자료들을 수집하여 유입량과의 관계를 분석하여 예측하는 방법이고, 확률론적 방법은 이전의 자료의 패턴을 이용하여 예측을 실시하는 방법이다. 유입량 예측은 재해에 대한 피해에 대비하기 위해 단기적 예측과 수자원의 활용과 댐 운영 및 관리를 위한 중·장기적 예측으로 구분하여 사용목적에 맞는 연구를 수행하고 있다. McCulloch and Pitts(1943)에 의해 소개된 인공신경망 모형(Artificial Neural Networks, ANN)은 비선형 관계를 잘 모의한다는 장점을 가지고 있어 우리나라는 1990년대 이후 다양한 분야에 적용되고 있다(박성천 등, 2009). 하지만 유출량 예측 시 저·갈수기에는 과대평가되

* 비회원·한양대학교 대학원 건설환경공학과 석사과정·E-mail : jjang6878@naver.com

** 정회원·한양대학교 대학원 건설환경공학과 박사과정·E-mail : 7924pooh@hanyang.ac.kr

*** 정회원·교신저자·한양대학교 건설환경공학과 조교수·E-mail : twkim72@hanyang.ac.kr

는 경향이 나타나고, 홍수기에는 과소평가되는 경향으로 인해 정확한 예측이 비교적 잘 이루어지지 않는다(박성천 등, 2009). 이에 본 연구는 기존의 ANN 모형과 회귀모형의 예측력을 향상시키기 위하여 k-means 방법을 이용한 군집분석을 도입하여 Error Rescaling의 후처리 과정을 통해 오차를 재조정하였다.

2. 대상지역 및 자료

본 연구는 금강수계의 대청댐 유역을 대상으로 월 유입량 예측모형을 구축하였다. 1982년 9월부터 2005년 12월까지의 대청댐 월 유입량 자료와 대청댐 유역내의 11개 강우 관측소 (안내, 옥천, 이원, 묘금, 청산, 능월, 송죽, 관기, 삼산, 동정, 삼가)의 관측 강우자료를 이용한 유역평균강우량으로 산정하여 자료를 구축하였다. 유입량 자료의 자기상관함수(Auto-Correlation Function, ACF)와 유입량과 강우량 자료의 교차상관함수(Cross Correlation Function, XCF)는 그림 1과 같다. ACF와 XCF를 분석하여 표 1과 같이 예측모형의 입력변수를 결정하였다.

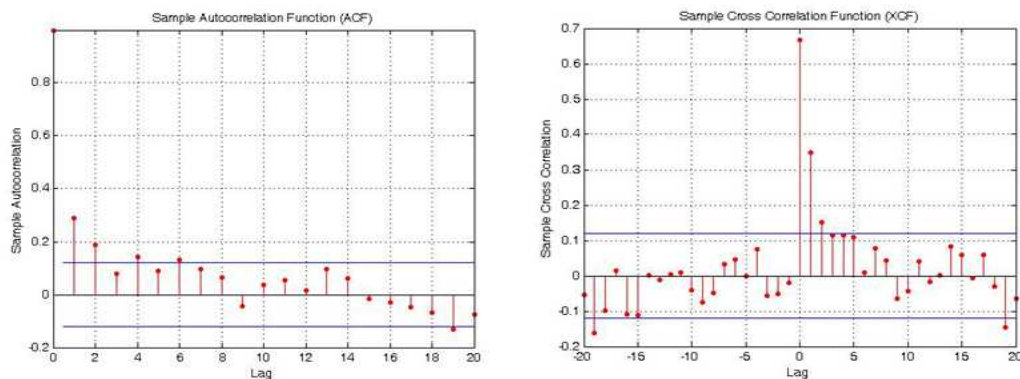


그림 1. 유입량의 자기상관함수(ACF)와 유입량과 강우량의 교차상관함수(XCF)

3. 이론적 배경

3.1 인공신경망 모형

ANN 모형은 1990년대 이후 공학 분야와 주식시장 예측 등 여러 분야에 적용되어지고 있고 (Kim, 2006), 수자원 분야에서도 강우, 유출뿐 아니라 가뭄 예측까지 광범위하게 적용되고 있다. 이와 같이 ANN 모형이 많이 사용되는 것은 Black-box 형태의 모형이기 때문에 물리현상에 대한 사전지식 없이 최적 해를 찾아내고, 비선형 문제 해결에 좋은 결과를 얻을 수 있다는 장점을 가지고 있어 다양한 분야에 적용하고 있다(Kim, 2006). 인공신경망 모형은 입력층, 은닉층, 출력층으로 구성된 구조로서 각 층의 뉴런들의 연결강도를 학습(Training)을 통해 최적의 모형을 구축하도록 되어 있다. 본 연구에서는 역전파알고리즘(Backpropagation, BPN)을 이용하여 뉴런의 연결강도를 최적화시켰으며, 활성화 함수는 은닉층에서의 Tangent Sigmoid Function을, 출력층에서는 Linear Function을 이용하였다. 역전파알고리즘(BPN)은 다층인식자(Multi-layer Perceptron, MLP) 모델 중 하나로 학습과정을 통해서 출력값(Output value)이 나오면 관측값(Target value)과 비교하여 오차가 최소가 되도록 연결강도의 가중치를 갱신하는 과정을 통해 출력값을 생성한다(이창환 등, 2009). 각 뉴런들은 활성화 함수에 의해 수학적 연산을 수행하여 가중치가 부여된 값들의 합으

로 이루어지는 것을 식 (1)과 같이 나타낼 수 있다.

$$y_k = \tilde{f} \left[\sum_{j=1}^M w_{kj} \cdot f \left(\sum_{i=1}^d w_{ji} x_i + w_{j0} \right) + w_{k0} \right] \quad (1)$$

여기서 w 는 뉴런간의 연결강도, x_i 는 입력 자료, f 는 활성화 함수이다.

3.2 회귀모형

회귀모형은 독립변수와 종속변수의 집단을 함수적인 상호관계식을 통해 모형화한 것이다. 본 연구에서는 식 (2)와 같은 다중회귀분석모형을 적용하였다.

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \quad i = 1, 2, 3, \dots, n \quad (2)$$

여기서 y_i 는 i 번째 측정된 종속변수의 값, β_0, β_1 는 회귀계수, x_i 는 i 번째 독립변수의 값, ϵ_i 는 i 번째 Y_i 의 오차 항으로 확률분포는 $N(0, \sigma^2)$ 이다.

4. 결과

본 연구에서는 1982년 9월부터 1995년 12월까지(학습기간)의 자료를 이용하여 모형을 구축하였으며, 1996년 1월부터 2005년 12월까지(검증기간)의 자료를 이용하여 모형의 예측성과를 비교하였다. 모형의 예측성과는 최소자승평균오차(Root Mean Square Error, RMSE)와 상관계수(Correlation Coefficient, CC)를 산정하여 평가되었다. 표 1은 검증기간동안의 모형의 예측성과를 정리한 것이다.

표 1. 검증기간(1996년 1월-2005년 12월) 동안의 예측성과

구분	ANN 모형		회귀모형	
	CC	RMSE	CC	RMSE
$Q_t = f(Q_{t-1}, Q_{t-2})$	0.49	3791	0.38	4376
$Q_t = f(Q_{t-1}, Q_{t-2}, P_{t-1}, P_{t-2})$	0.58	3653	0.57	3587
$Q_t = f(Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}, Q_{t-5}, Q_{t-6})$	0.50	3785	0.43	4534
$Q_t = f(Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}, Q_{t-5}, P_{t-1}, P_{t-2}, P_{t-3}, P_{t-4}, P_{t-5})$	0.60	3498	0.57	3564

본 연구에서는 이전 달의 관측 유입량을 예측변수로 사용하는 Persistence 모형($Q_t = Q_{t-1}$)을 기준모형으로 예측성과를 비교 분석하였다. Persistence 모형의 RMSE는 4793, CC는 0.38이다. Persistence 모형과 비교하여, ANN 모형은 약 28%, 회귀모형은 약 26% 향상된 모의 결과를 보여 주고 있다. 본 연구에서는 ANN 모형과 회귀모형의 예측력을 향상시키기 위하여 학습기간의 모의 결과를 이용하여 군집분석을 실시하였다. 학습기간의 모의결과를 2, 3, 4개의 군집으로 구분한 후, 군집분석 결과를 바탕으로 (1) 각 군집들의 중심점에 대한 오차를 산정하여 이를 검증기간의 군집분석 결과를 학습기간의 군집에 대한 오차만큼 이동 평균하여 보정하였고, (2) 각 군집들의 회귀 직선을 이용하여 이를 검증기간에 적용하여 오차를 보정하였다. ANN 모형과 회귀모형의 예측결과에 대하여 재 산정한 결과를 표 2에 정리하였다. 모의결과는 ANN 모형에서의 입력변수 10개의 뉴런 수 20개일 때와 회귀모형에서의 입력변수 10개일 때 가장 좋은 예측성과를 얻을 수 있었고, 이를 k-mean 군집분석을 이용하여 2-4개의 군집에 대해 두 가지의 방법을 이용하여 오차 스케일

을 줄인 결과, ANN모형이 군집 수 4일 때 평행이동을 이용한 방법과 회귀모형의 군집 수 2일 때 회귀모형을 이용한 방법일 때 가장 좋은 결과를 얻을 수 있었다. 그림 3은 ANN 모형을 이용하여 예측한 유입량을 관측된 유입량과 비교한 것이고, 그림 4는 군집분석을 이용하여 예측력을 향상시킨 유입량을 산포도를 이용하여 비교한 것이다.

표 2. ANN모형과 회귀모형의 Error Rescaling 적용 결과

구분	군집수	ANN 모형							
		오차조정 전				오차조정 후			
		Training		Validation		Training		Validation	
		RMSE	CC	RMSE	CC	RMSE	CC	RMSE	CC
평행이동 (1)	2	3041	0.53	3498	0.6	1778	0.87	2297	0.85
	3					1590	0.91	2190	0.87
	4					1465	0.92	2086	0.88
회귀모형 (2)	2					1499	0.91	2126	0.87
	3					1486	0.91	2114	0.87
	4					1379	0.92	2140	0.87

구분	군집수	회귀모형							
		오차조정 전				오차조정 후			
		Training		Validation		Training		Validation	
		RMSE	CC	RMSE	CC	RMSE	CC	RMSE	CC
평행이동 (1)	2	2970	0.55	3564	0.57	1807	0.87	2484	0.84
	3					1565	0.90	2405	0.85
	4					1494	0.91	2212	0.87
회귀모형 (2)	2					1454	0.91	2260	0.86
	3					1426	0.92	2256	0.86
	4					1392	0.92	2160	0.87

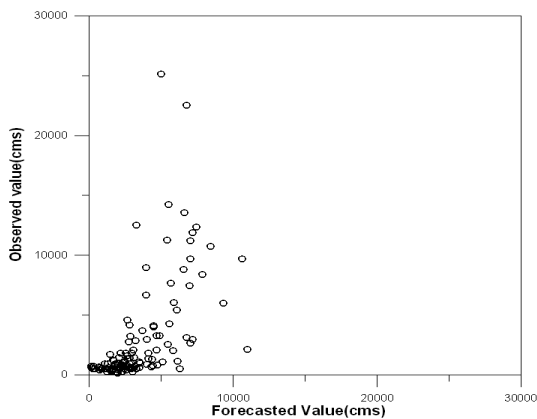


그림 3. 유입량 예측결과
(검증기간, 오차조정 전)

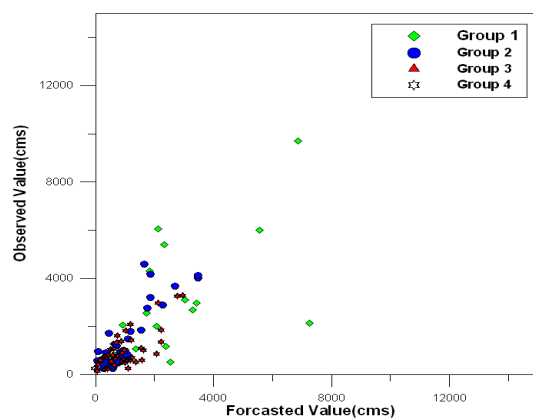


그림 4. 유입량 예측결과
(검증기간, 오차조정 후)

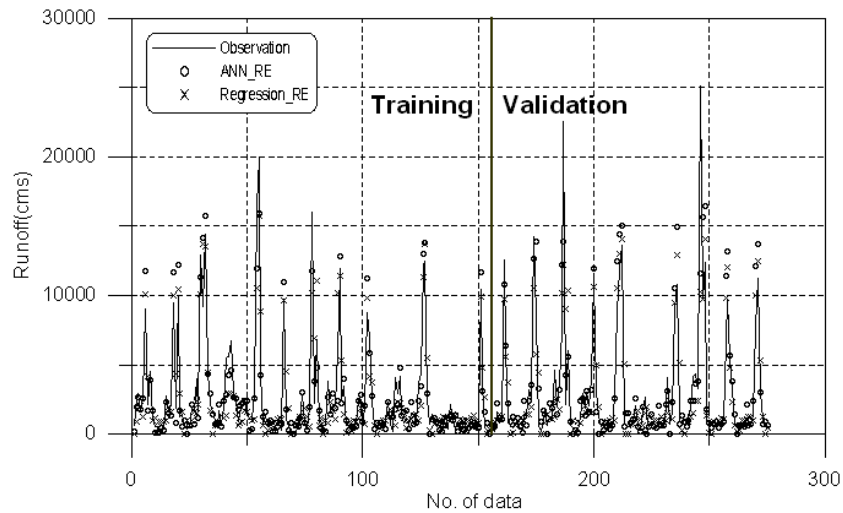


그림 5. 학습기간과 검증기간의 유입량 시계열

5. 결 론

ANN 모형과 회귀모형 모두 간편하게 예측모형으로 사용할 수 있다는 장점이 있는 반면, 저·갈수기의 과대추정 및 홍수기의 과소추정 되는 현상의 단점을 개선위해 본 연구에서는 군집분석을 통해 평행이동과, 회귀모형을 이용하여 Error Rescaling의 후처리 과정을 제안하였다. ANN모형이 군집 수 4일 때 RMSE 2086과 CC 0.88로 가장 좋은 결과를 얻었고, 약 40% 가까이 예측력을 향상 시켰다. 대체로 학습기간에서는 회귀모형을 이용한 후처리과정이 RMSE가 적게 산정된 반면, 검증기간에서는 높은 RMSE를 보였지만, 향후 회귀모형을 이용한 후처리과정에 적합한 군집분석이 이루어진다면 더 좋은 예측성과를 얻을 수 있을 것으로 기대된다.

감 사 의 글

이 논문은 2009년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구(No. 2009-0072476)입니다.

참 고 문 헌

1. 건설교통부(2006). 수자원장기종합계획(2006-2020).
2. 박성천, 김용구, 정천리, 진영훈(2009). 홍수 예·경보시스템 개선을 위한 인공지능망 이론의 적용, 한국수자원학회 학술대회논문집, pp. 1265-1271
3. 이창환, 안재현, 이주현, 김태웅(2009). 인공지능망모형과 군집분석을 이용한 교각 세굴심 예측, 대한토목학회논문집, 제29권, 제2B호, pp. 111-120.
4. Kim, T.W.(2006). Monthly precipitation forecasting using rescaling errors, Journal of Civil Engineering, KSCE, Vol. 10, No. 2, pp. 137-143.
5. McCulloch, W.S., and Pitts, W.(1943). A logical calculus of the ideas immanent in nervous activity, Bulletin of Mathematical Biophysics, Vol. 5, pp. 115-133.