

단기 하천수질 예측을 위한 신경망모형

Artificial Neural Networks for Forecasting of Short-term River Water Quality

김만식* · 한재석**

Kim, Man-Sik · Han, Jae-Seok

Abstract

The purpose of this study is the prediction of pollutant loads into Seonjin river watershed using neural networks model. The pollutant loads into river watershed depend upon the water quantity of inflow from the upstream as well as the water quality of the inflow into the river. For the estimation of pollutants into river, a neural networks model which has the features of multi-layered structure and parallel multi-connections is used. The used water quality parameters are BOD, COD and SS into Seonjin river. The results of calibration are satisfactory, and proved the availability of a proposed neural networks model to estimate short-term water quality pollutants into river system.

Keywords : Forecasting of short-term water quality, Neural networks model, BOD, COD, SS.

요 지

본 논문의 목적은 섬진다목적댐 유역의 하천을 대상으로 강우시에 단기 수질상태를 예측하기 위하여 병렬다중결선의 계층 구조를 갖는 신경망이론을 이용하였다. 본 연구에 적용한 신경망이론의 학습알고리즘으로는 역전파알고리즘을 사용하였으며, 최적모형의 개발을 위해 모멘트법-적응학습율기법을 이용하였다. 하천 수질오염 부하량에 영향을 미치는 요소로서 상류로부터 유입되는 유입량과 수질인자인 BOD, COD, SS를 고려하였다. 섬진다목적댐 유역에 대해 단기 수질을 예측할 수 있는 다층신경망모형을 개발하기 위해 은닉층 노드수와 학습회수에 변화를 주어 각 수질인자 별로 4가지씩 총 12개의 모형을 구성하여 학습을 실시하였다. 제안된 신경망모형의 검증을 위해 학습시키지 않은 수 질자료를 예측한 결과 양호한 것으로 분석되었고, 하천수계의 단기 수질오염 예측에 활용할 수 있을 것으로 사료되었다.

주요어 : 단기 수질예측, 신경망모형, BOD, COD, SS

* 정회원 · 신성대학 토목정보과 교수 · 공학박사

** 충청대학 생명공학부 교수 · 공학박사

1. 서론

하천은 인류문명이 시작되면서 생활터전으로서 우리의 생활환경에 중요한 부분을 차지하여 왔고 자연적인 것으로 조건없이 사용되어 왔다. 그러나, 우리나라는 1970년대 이후 경제발전에 힘입어 국민 생활수준의 급격한 향상으로 물 사용량이 급격히 증가되었고, 이에 따라서 수질오염도 심각한 수준으로 되어 80년대 이후에는 수질오염의 방지 및 저감, 오·폐수 처리시설의 확대와 첨단화를 도모하여 환경 및 생태계 변화에 적극 대응하고 있는 실정이다.

일반적으로 환경분야에서 사용하는 수질예측방법은 물수지(mass-balance)방정식에 기초한 오염부하량의 입·출력간의 상관관계를 근본으로 수치모델을 사용하고 있다.

그 예로 WQRRS, WASP4, QUAL2E 모형 등이 사용되고 있으나, 하천수계의 경우에 있어서는 계산을 수행할 때 하천연장이 비교적 긴 경우 올바른 오염부하량의 산정이 곤란하고, 또한 필요한 수문자료의 확보 또한 어려운 실정이다. 더욱이 많은 매개변수들에 대한 기본 자료가 정리되어 있지 못한 현실로 그 실제 적용을 위해서는 앞으로 많은 연구 및 자료의 확보가 필요한 실정이다. 특히, 우리나라 하천과 같이 갈수 및 홍수의 특성이 두드러지는 경우 하천의 수질은 더욱 악화되어 급수 대상지역에 대한 제한급수 등 일련의 조치가 취해지기도 한다. 특히 다목적댐 저수지 수계는 도시의 생활용수 및 공업용수 공급에 절대적으로 중요한 역할을 하므로 수자원의 양적 관리보다는 질적인 관리가 우선시 되어 다루어져야 할 것이다. 따라서 상기 모형을 활용시 많은 입력자료가 요구되고 최적의 매개변수를 산정하기가 매우 어렵기 때문에 모형 개발에 있어 특별한 구조나 매개변수 산정, 자료의 변환 등이 필요치 않고 자료의 축척에 따라 학습을 통하여 모형의 예측 능력을 향상시킬 수 있는 신경망이론을 사용하였다. 신경망이론은 인간의 두뇌와 같이 생각하고 판단하며, 인식할 수 있는 능력을 구현하기 위해 인간뇌의 정보처리 메커니즘을 수학적 으로 표현한 이론으로서 다수의 원소가 동시에 작동하는 병렬 처리 기계처럼 하드웨어 및 소프트웨어적인 병렬 분산 처리를 하는 시스템이며, 다른 환경에 맞도록 변화시킬 수 있는 학습 능력을 가지고 있다.

본 연구에서는 신경망이론 중에서 역전파알고리즘과 최적 매개변수 추정에 사용되는 모멘트-적용학습방법을 이용하였으며, 학습회수 및 은닉층 그리고 노드수에 변화를

주어 여러 다층신경망 모형을 구축하여 섬진강유역의 쌍치와 관촌지점에 적용하였다. 그리고 학습된 결과중에 효율이 좋은 모형을 선정하여 학습되지 않은 자료의 수질인자의 예측을 실시하였다. 또한, 예측결과를 분석·고찰하였으며, 하천의 단기간 수질예측에 신경망모형의 실용가능성을 검증하였다.

2. 신경망이론

2.1 다층신경망모형

다층신경망모형은 입력자료를 받아들이는 입력층(input layer), 결과를 나타내는 출력층(output layer), 그리고 두 층 사이에 은닉층(hidden layer)으로 구성되어 있으며, 이 층들을 결합하는 점을 시냅스(synapse)라 하고 이 시냅스의 연결강도(결합강도) 변화에 따라 입력에 따른 출력이 결정되어지며 이를 그림 1에 도식화하여 나타내었다. 다층신경망모형에서 시냅스의 연결강도는 모형에 따라 미리 계산에 의해 주어지는 경우와 학습에 의해 결정되는 경우가 있으며, 본 연구에서의 연결강도는 임의의 초기값에서 계속적인 개선이 이루어지며 적절한 연결강도를 구하는 학습(learning)과정을 사용하였다. 학습방식에 따라 지도학습(supervised learning)모형과 자율학습(unsupervised learning)모형으로 나눌 수 있으며, 입력자료와 출력자료가 2진수(binary)의 형태를 갖거나 실수형태의 연속적인 값을 가지는 두 가지 형태로 나누어진다.

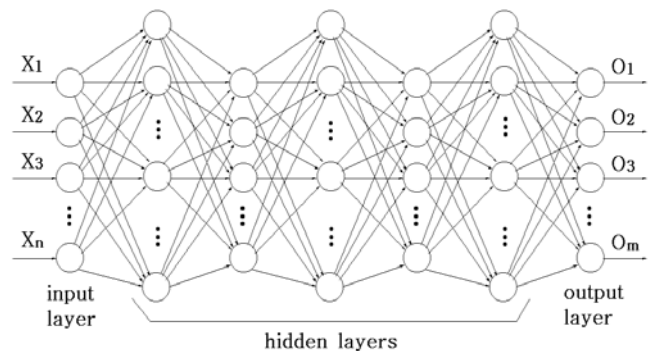


그림 1. 다층신경망의 구조

본 연구에 적용된 다층신경망(multilayered neural network)모형은 1개 이상의 은닉층이 존재하는 모형은

로, 입력층에서 은닉층으로 또는 은닉층에서 출력층으로 연결될 때 연결강도의 벡터들과 곱해서서 특정한 출력함수를 거쳐 출력을 산정한다. 이때 사용하는 출력함수(또는 전이함수)는 대표적인 비선형 함수이며 로그-시그모이드(log-sigmoid)함수를 사용하였으며, 그림 2에 신경망모형을 구성하고 있는 각 요소들의 흐름도를 나타내었다.

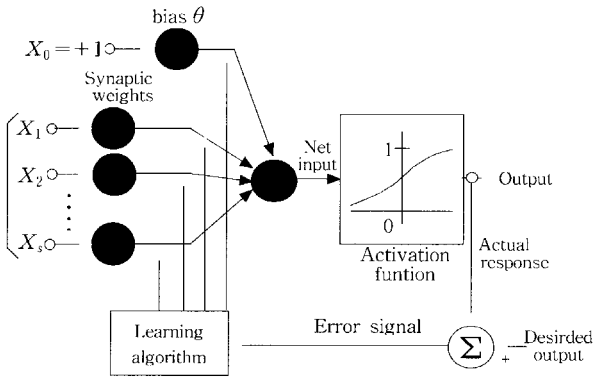


그림 2. 신경망을 구성하는 각 요소의 흐름도

2.2 역전파알고리즘

다층구조의 신경망에서 각 뉴런과 각 층(layer)의 모든 뉴런에 임계값(threshold values) 사이에 연결 가중치를 주기 위해서 학습 알고리즘(learning algorithm)이 필요하다.

본 연구에서는 학습 알고리즘으로 러멜하트(rumelhart, 1986) 등에 의해 제안된 역전파 알고리즘(backpropagation algorithm)을 사용하였다. 역전파 알고리즘은 입력층의 각 유니트에 입력패턴이 주어지면 이 신호는 각 유니트에서 가중값에 의해 변환되어 은닉층에 연결되고 마지막에 출력층으로 전파해 나가는데, 이때 네트워크에서 출력된 신호 패턴이 목표 패턴과 일치하지 않을 경우 역방향인 출력 패턴에서 목표 패턴으로 오차의 값들을 감소시킬 수 있도록 네트워크의 연결강도를 조절하여 학습하는 방법이다.

2.3 모멘트(moment)법과 적응학습율(adaptive learning rate)

다층신경망모형에서 역전파 알고리즘에 의한 학습과정에서 발생할 수 있는 국부최소값(local minima value) 문제와 오차를 적게 하고, 수렴속도를 빨리하여 학습속도

를 향상시킬 수 있는 모멘트법과 적응학습율을 사용하였다. 모멘트법은 간단하면서도 효율적이기 때문에 자주 이용되는 방법으로 모멘트항을 학습에 포함시킨다. 적응학습율은 모멘트 방법과 병행하여 학습율 μ 를 조절하며 학습속도를 향상시킨다. 학습 후에 오차함수의 감소여부를 확인하여 크게 감소하였을 때에는 학습율을 일정 비율만큼 증가시키고, 반대로 오차함수가 정해진 비율 이상으로 증가하였을 때에는 학습율을 정해진 비율만큼 감소시키게 된다.

3. 하천의 단기 수질 예측모형

신경망모형은 수집된 또는 측정된 입력자료와 출력자료의 학습패턴을 정리해야 하는데, 자료의 특성을 변화시키지 않는 범위에서 입·출력 학습패턴을 정규화하여 입력값과 출력값의 차원을 일치시킨다. 다음 단계로 입·출력패턴으로 학습을 수행하여 최적 매개변수를 산정하기 위해 신경망모형의 구조를 입·출력자료의 수와 각 층의 수, 각 층에 대한 노드의 수, 각 층의 출력함수 등에 따라 결정하여 모형을 구축하고, 학습을 완료한 후 학습되지 않은 실측 자료를 이용하여 모형의 적합성을 검증하게 된다.

3.1 모형화

다목적담 유입량의 수질은 발생시간에 따라 일치하지 않는 비정규성 및 비선형성을 나타낸다. 신경망은 수질인자와 유량과의 상관관계에 의하여 입력되는 변수와의 연결강도를 조정하여 각 처리소자에 입력값을 만드는 활성화값(activation value)을 가감시켜 매개변수를 추정하므로 학습을 통해 시스템의 최적 매개변수를 추정하는 동정과정을 자동화할 수 있다. 수질 예측의 모형화를 위해 시간흐름에 따른 비선형 수질인자 및 유량을 이용한 수질모형은 식 (1)과 같이 정의할 수 있다.

$$QUAL(t) = f [QUAL(t - k_i), Q(t - k_j)]$$

$$\text{for } k_i = 0, \dots, n, \quad k_j = 0, \dots, m \quad (1)$$

여기서, $QUAL(t)$: 수질예측인자, $QUAL(t - k_i)$: k_i 전의 수질예측 인자, $Q(t - k_j)$: k_j 전의 유량, k_i, k_j : n, m 간격의 시계열(시간, 일, 달, 년)을 의미한다.

3.2 모형 구성

예측되는 수질인자는 학습과정을 통하여 추정된 연결강도와 임계값을 이용하여 계산된다. 연결강도는 각 층의 상태를 매개변수로 저장시킨 값으로 각 처리소자간의 입력과 출력의 상관관계를 나타내며 연결되어 있는 처리소자간의 영향력에 정도를 나타낸다. 따라서 수질인자의 예측은 추정된 연결강도의 행렬을 이용하여 계산된다. 시간의 흐름에 따른 비선형 수질인자-유량의 관련성을 나타낸 식(1)을 이용하여 모형화 과정을 통해 신경망에 의한 예측모형을 정의하면 식(2)와 같이 나타낼 수 있다.

$$QUAL_{i,t} = g \left(\sum_{k=1}^K W_{jk} g \left(\sum_{j=1}^J W_{ij} (QQ(t-k) + \theta_j) + \theta_i \right) \right) \quad (2)$$

여기서, $QUAL_{i,t}$: 예측된 수질인자, $QQ(t-k)$: 학습된 시계열로서 과거의 수질인자와 유량, k 는 지체시간이다. 표 1에 수질예측을 위한 신경망모형들의 구성을 나타내었다.

표 1. 다층신경망 모형의 구성

water quality index	model	lag time (hour)	inflow lag time (hour)	input layer	hidden layer node	output layer	learning number
BOD	BOD-1, BOD-2 BOD-3, BOD-4	3	3	6	6, 12 18, 24	1	10,000 ~ 25,000
COD	COD-1, COD-2 COD-3, COD-4	3	3	6	6, 12 18, 24	1	
SS	SS-1, SS-2 SS-3, SS-4	3	3	6	6, 12 18, 24	1	

본 연구에서 예측하는 수질인자는 BOD, COD 및 SS농도로 하였으며 수질예측을 위한 최적 다층신경망모형을 구성하기 위해 표 1과 같이 입력층, 은닉층, 출력층의 노드수와 학습회수 그리고 지체시간을 정하였다.

4. 적용 및 결과

4.1 대상유역과 검증자료

본 연구를 위한 대상유역은 섬진강댐 저수지 상류유역으

로 섬진강은 전라북도 장수, 임실, 진안 등 3개군의 경계를 이루는 팔공산에서 발원하여 약 84.2km를 남서류하여 섬진강 다목적댐 저수지에 이르게 되며 그림 3에 나타내었다. 섬진강 다목적댐까지의 하천현황은 섬진강 본류와 39개의 제 1, 2 지류들로 구성되어 있으며, 유역면적이 약 763.82km²로 관촌유역이 301.776km², 댐저수지유역이 344.145km², 쌍치유역이 117.895km²이며, 하천 총연장은 356.5km이다. 하천 수질예측모형을 개발하기 위해 섬진강 다목적댐 수질관리기법(1, 2차년도, 한국수자원공사 수자원연구소)에 수록된 자료를 이용하였다. 자료는 섬진강 다목적댐 저수지의 상류유역인 관촌지점과 쌍치지점의 1998년(4회)과 1999년(6회)의 BOD, COD, SS 수질자료와 유출량자료를 사용하였다. 그리고 모형의 검증을 위해 학습되지 않은 자료를 이용하여 수질예측을 실시하였다.



그림 3. 섬진다목적댐 유역도

4.2 모형검증을 위한 통계지표

예측결과가 실측자료와 비교하여 얼마나 적합한 것인가 하는 문제는 단기 수질모형에 대한 신경망이론의 적용타당성과 관련된다. 따라서 본 연구에서는 예측문제에 적용되고 있는 통계적 검증방법으로 각 모형에서 계산된 오차와 비교 검토하여 예측에 사용될 최적 모형을 선정하기 위해 SSE(sum square error), CC(correlation coefficient), RMSE(root mean square error), R^2 를 이용하여 통계분석을 실시하였으며 각 통계지표의 방정식을 표 2에 나타내었다.

표 2. 신경망모형의 분석을 위한 통계량

index	relation equation	
sum square error(SSE)	$\frac{1}{N} \sum_{i=1}^N (X_i - Y_i)^2$	실측치: X_i 계산치: Y_i
correlation coefficient(CC)	$\frac{1}{N} \sum_{i=1}^N \left(\frac{X_i - X_{im}}{\sigma_X} \right) \left(\frac{Y_i - Y_{im}}{\sigma_Y} \right)$	예측치: Y_{for} 실측치 평균: X_{im} 계산치 평균: Y_{im}
root mean square error (RMSE)	$\sqrt{\frac{1}{N} \sum_{i=1}^N (X_i - Y_i)^2}$	실측치 표준편차: σ_X 계산치 표준편차: σ_Y
decision coefficient(R^2)	$1 - \frac{(\sum_{i=1}^N X_i - Y_{for})^2}{(\sum_{i=1}^N Y_{for} - X_{im})^2}$	자료개수: N

4.3 적용 및 분석

단기 수질예측모형을 구축하기 위해 1998년부터 1999까지 측정된 관촌과 쌍치지점의 수질자료중 9회의 자료에 대해 자기상관과 교차상관분석을 실시하여 유량은 2시간 전, 수질자료는 3시간 전 까지 영향을 받는 것으로 분석되었다. 그리고 각 수질인자별로 12가지의 신경망모형을 구축하여 학습을 실시하였으며 통계분석 결과는 표 3과 같다. 단기 수질예측을 위한 다층신경망 모형의 학습과정

에서 오차는 모든 모형에서 학습회수가 100회 이내에서 급격히 감소하였으며, 은닉층 노드수가 많을수록 좋은 결과를 나타내었다. 표 3에 나타난 결과를 볼 때, 거의 모든 모형들이 상관계수가 0.9이상이고, RMSE는 0.1이하, SSE는 0.001 이하로 나타나 매우 양호한 결과를 보여주고 있다. 이 모형들의 예측능력을 검토하기 위해서 학습에 사용하지 않은 1999년의 1회 수질자료에 대해 예측을 실시하였다.

표 3. 다층신경망모형의 통계분석 결과

지점	model	SSE	CC	RMSE	model	SSE	CC	RMSE	model	SSE	CC	RMSE
관촌	BOD-1	0.00249	0.979	0.0499	COD-1	0.00201	0.976	0.0448	SS-1	0.00089	0.988	0.0299
	BOD-2	0.00228	0.981	0.0478	COD-2	0.00214	0.979	0.0462	SS-2	0.00072	0.990	0.0268
	BOD-3	0.00226	0.979	0.0475	COD-3	0.00240	0.981	0.0490	SS-3	0.00090	0.989	0.0300
	BOD-4	0.00218	0.980	0.0467	COD-4	0.00307	0.983	0.0554	SS-4	0.00066	0.992	0.0257
쌍치	BOD-1	0.00311	0.965	0.0557	COD-1	0.00200	0.984	0.0447	SS-1	0.00172	0.974	0.0415
	BOD-2	0.00199	0.969	0.0446	COD-2	0.00060	0.985	0.0246	SS-2	0.00065	0.976	0.0278
	BOD-3	0.00134	0.970	0.0367	COD-3	0.00061	0.985	0.0247	SS-3	0.00077	0.978	0.0278
	BOD-4	0.00043	0.972	0.0207	COD-4	0.00069	0.987	0.0263	SS-4	0.00022	0.982	0.0148

4.4 예측 및 분석

선정된 모형을 학습에 이용되지 않은 1999년 1회 수질

자료를 예측한 결과에 대해 통계분석한 결과를 상기 표 4에 나타냈으며, 예측결과는 그림 4~5와 같다. 통계분석은 RMSE와 R^2 을 실시하였다.

표 4. 예측된 결과의 통계분석

지점	model	RMSE	R^2	model	RMSE	R^2	model	RMSE	R^2
관촌	BOD-1	0.131	0.642	COD-1	0.053	0.942	SS-1	0.105	0.846
	BOD-2	0.170	0.594	COD-2	0.027	0.947	SS-2	0.089	0.938
	BOD-3	0.163	0.224	COD-3	0.042	0.964	SS-3	0.107	0.964
	BOD-4	0.240	0.284	COD-4	0.059	0.955	SS-4	0.093	0.866
쌍촌	BOD-1	0.134	0.739	COD-1	0.222	0.390	SS-1	0.209	0.699
	BOD-2	0.122	0.666	COD-2	0.084	0.478	SS-2	0.263	0.525
	BOD-3	0.173	0.564	COD-3	0.134	0.479	SS-3	0.184	0.931
	BOD-4	0.147	0.641	COD-4	0.173	0.465	SS-4	0.267	0.168

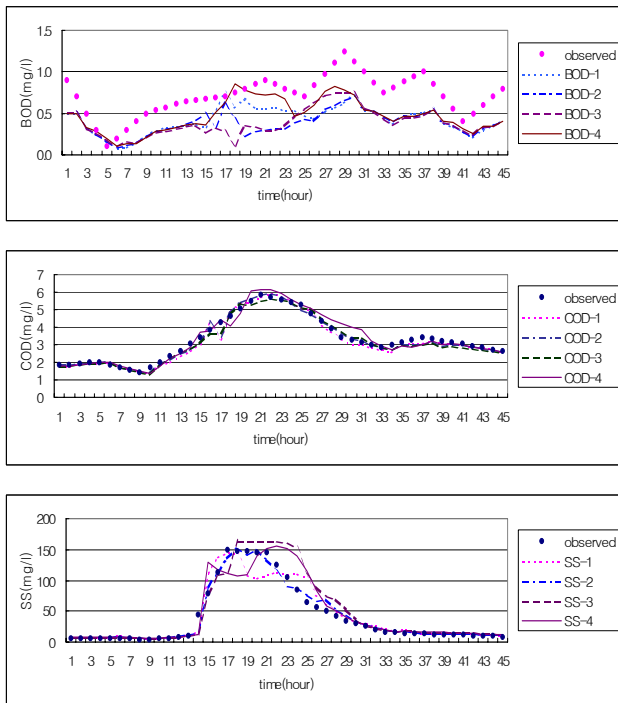


그림 4. 관촌지점의 수질 예측결과

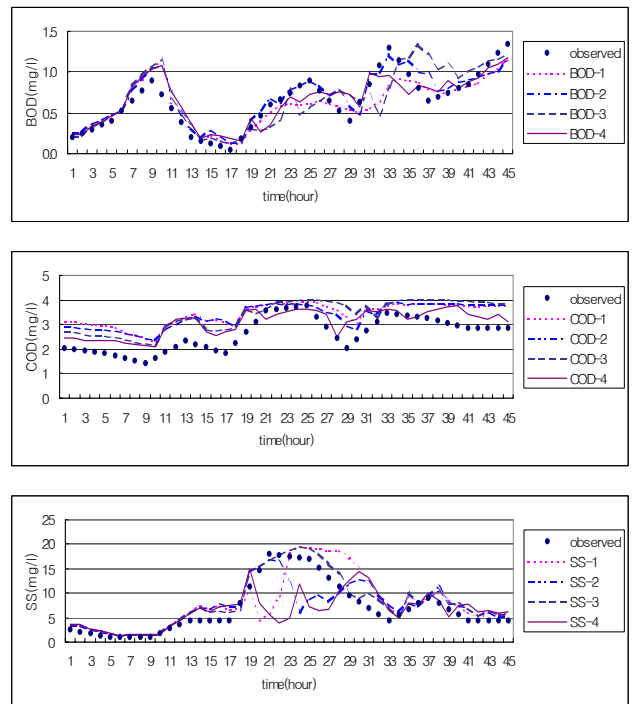


그림 5. 쌍치지점의 수질 예측결과

관촌지점의 예측된 결과를 보면, BOD모형에서는 다소 차이가 있으나 BOD-4모형의 예측결과가 관측치와 비슷한 경향을 나타내었고, COD모형에서는 모든 모형들이 실측값과 거의 비슷한 경향을 나타내었으며, SS모형은 SS-3모형이 오차가 가장 적은 결과를 나타내었다.

쌍치지점은 BOD모형에서 다소 차이가 있으나 BOD-2모형이 예측결과가 가장 관측치와 비슷한 경향을 나타내었고, COD모형에서는 모든 모형들이 경향은 비슷하나 통계치의 정도가 많이 떨어지게 나타났으며, SS모형은 SS-3모형이 오차가 가장 적은 결과를 나타내었다.

이와 같이 예측결과가 BOD, SS, COD 모두가 다소 차이가 있으나 거의 유사한 경향을 나타냈고, 은닉층의 노드 개수는 입력층의 개수가 n 개 라면 $2n \sim 4n$ 개, 그리고 학습회수는 20,000~25,000회가 적당한 것으로 사료된다. 학습시간과 효율 등을 고려하여 수질모형에 대한 은닉층 노드수와 학습회수의 설정에 대한 세심한 검토가 요구되었으며, 좀더 많은 자료와 적절한 수질인자간의 상호 연관된 물리적 수질모형을 활용할 수 있다면 신경망이론을 응용한 수질추정 및 예측에 대해서도 충분히 활용할 수 있을 것으로 판단되었다.

5. 결론

본 연구에서는 신경망이론을 이용하여 섬진강다목적댐 상류유역의 관촌지점과 쌍치지점에 대한 단기 수질예측을 위한 활용가능성을 판단하기 위해 다층 신경망모형을 구성하였다. 1998년의 5회 수질자료, 1999년의 7회 수질자료, 그리고 동일한 기간의 유출량자료를 이용하여 BOD, COD, SS에 대해 총 12가지 모형을 학습하였고, 모형의 검증을 위해 학습에 이용되지 않은 1회 수질자료에 대해 예측하였다. 예측 결과는 비교적 양호한 결과를 나타냈으며, 좀 더 많은 연속된 수질자료가 있었다면 본 연구에서 개발한 신경망모형의 예측능력을 충분히 검증할 수 있었는데 수질자료가 없어 아쉬웠다. 향후 좀 더 향상된 결과를 얻기 위해서는 다양한 구조의 신경망모형의 구성, 양질의 많은 자료 그리고 수질자료의 특성을 명확히 구분할 수 있는 신경망모형에 대한 연구가 계속적으로 진행되어야 할 것으로 판단되었다.

감사의 글

본 논문은 충청대학 학술연구조성비의 자원에 의해 수정한 결과로서 이에 감사드립니다.

(접수일자 : 2002. 8. 27)

참 고 문 헌

1. 김만식(1998), 다목적 저수지의 홍수유입량 예측을 위한 신경망모형. 박사학위논문, 충북대학교. 300p.
2. 이요상, 강병수(2001), 댐저수지 유역의 오염부하 유출특성 - 섬진강댐 저수지를 중심으로 -, 한국수자원학회 논문집, Vol. 33, No. 6, pp. 757~764.
3. 김만식(1999), 하천수계의 장기 수질예측을 위한 신경망모형, 한국환경관리학회 논문집, Vol. 5, No. 1, pp. 145~152.
4. 김만식외 3인(1998), 신경망모형을 이용한 하천의 수질예측 연구, 2001년도 한국수자원학회 학술발표회 논문집(III), 한국수자원학회, pp. 925~930.
5. 한국수자원공사(1998), 섬진강 다목적댐 수질관리 기법 연구(1차년도), 연구보고서.
6. 한국수자원공사(1999), 섬진강 다목적댐 수질관리 기법 연구(2차년도), 연구보고서.
7. Demuth, H. and M. Beale(1994), Neural Network Toolbox : For Use with MATLAB User's Guide, The Math Works Inc., pp. 5.1~5.38.
8. Haykin, S.(1994), Neural Networks : A Comprehensive Foundation. Prentice Hall.
9. Peter A. K.(1994), and V. Novotny(1980), Water Quality Management. Academic Press.