

인공 신경망 이론을 이용한 영산강의 유출량 예측

(The Prediction of Runoff using Artificial Neural Network in the Young-San River)

이관수, 박성천, 이한민, ○진영훈

1. 서 론

비선형성을 내포하고 있는 자연 현상인 강우-유출 현상을 모형화하기 위해서 일반적으로 개념적인 모형(conceptual modeling)과 이론적인 모형(system theoretical modeling)이 사용되어지고 있다. 이는 대개 물리적인 현상의 반영정도에 따라 나누어지게 되고, 개념적인 모형은 물리적인 과정을 단순화하여 표현하게 되며, 이론적인 모형은 흔히 Black-Box 모형이라고도 한다.

또한 실제적인 상황에서 모형을 개발하는 주요한 목적은 특정 유역에 대한 유출량의 예측 등을 정확하게 하는 데 있다. 따라서 이러한 상황은 개념적인 모형을 이용하기 위해 많은 시간과 노력을 투자하기보다는 좀 더 간단한 이론적인 모형을 사용하도록 하는 이유가 된다. 그러나 이론적인 모형의 적용은 매개변수를 결정하기 위해 사용된 자료에 국한 될 수 있으며, 결정된 매개변수는 특정 유역에만 적용되는 특성치가 된다.

본 연구에서는 이론적인 모형을 이용하여 영산강을 대상으로 하여 모형을 개발하였으며, 현재 까지도 널리 사용되어지고 있는 ARIMA와 같은 선형 시계열 모형이 자연 현상 자체에 포함되어져 있는 비선형성을 나타내기에는 부적합하므로 최근 많은 분야에서 적용되어지고 있는 인공 신경망(artificial neural network) 이론을 적용하여 일유출량 및 시유출량의 예측을 실시하였다.

2. 인공 신경망

인간의 두뇌는 매우 복잡하며, 비선형적이고 병렬처리를 행하는 컴퓨터와 같다. 또한 이러한 행위 및 계산을 가능하게 하는 뉴런(neuron)이라 불리는 자체의 구성체를 조직하는 능력을 가지고 있다. 이러한 두뇌는 선천적으로 경험(experience)을 통해 자기 자신의 규칙을 만들 수 있는 구조와 능력을 지니고 있으며, 경험은 시간이 지남에 따라 축적되어진다. 이와 같이 인간의 두뇌가 특별한 기능이나 임무를 수행하도록 하는 방법을 모형화한 것이 인공 신경망이며, Haykin은 “신경망이란 자연적으로 경험적인 지식을 저장하고 유용하게 만드는 경향을 지닌 간단한 처리 단위로 구성된, 거대하며 병렬적으로 분포된 연산 처리 장치이다. 이는 다음의 두 가지 면에서 인간의 두뇌와 유사하다. 첫 번째, 지식은 학습 과정을 통해 환경으로부터 신경망에 의해 얻어진다. 두 번째, 뉴런간의 연결강도는 습득된 지식을 저장하기 위해 사용되어진다.”라고 정의하고 있다. 인공 신경망은 인간 두뇌의 뉴런과 같은 기능을 수행하며, 수학적인 연산능력을 가진 노드들이 상호 연결된 것으로 적절한 학습규칙에 의하여 작동한다. 즉, 각 노드들은 결합 함수와 전달 함수(활성화

이관수 전남대학교 토목공학과 교수,
이한민 전남대학교 토목공학과 박사과정.

박성천 동신대학교 환경공학과 조교수,
○진영훈 전남대학교 토목공학과 공학석사.

함수)에 의해 수학적 연산을 수행한다. 노드로 입력되는 신호는 식(2.1)의 s_j 와 같이 가중치가 부여된 값들의 합이 되며, 이 값은 다시 전달 함수라 불리는 비선형 함수를 통과하여 식(2.2)의 y_j 가 된다.

$$s_j = \sum_{i=0}^m w_{ji}x_i \cdots \cdots \cdots (2.1) \quad y_j = f(s_j) = \frac{1}{1 + \exp(-s_j)} \cdots \cdots \cdots (2.2)$$

여기서 s_j 의 범위는 $\pm\infty$ 이며, y_j 의 범위는 0에서 1까지이다.

본 연구에서는 위의 식과 같이 발생된 신호들을 처리하기 위해 다층(multi-layer) 신경망을 사용하였으며, 특히 입력층과 은닉층, 출력층으로 구성되어지는 3계층 신경망 구조를 채택하였다. 또한 구성되어진 인공 신경망을 훈련시키기 위한 규칙적인 학습 알고리즘은 현재 널리 사용되어지고 있는 역전파 학습 알고리즘(back-propagation learning algorithm)을 사용하였다. 이와 더불어 보정을 위한 훈련 과정의 효율성의 증진을 위해 Early Stopping Method of Training을 적용하였으며, 모멘텀 상수와 적응식 학습율을 포함하여 훈련 과정을 구성하였다. 이에 포함된 활성화 함수로는 은닉층에서의 단극성 시그모이드(sigmoid) 함수와 출력층에서의 선형(linear)함수를 각각 적용하였다.

3. 대상 지점 및 자료

3.1 대상 유역 및 지점

본 연구의 대상 지점인 나주 지점의 수위 관측소는 전라남도 나주시 삼도동의 나주교(동경 126° 44' 00", 북위 35° 02' 00") 직하에 위치하며, 영산강 본류를 대표한다. 나주 지점에 대한 유역 면적은 2,058.72 km²로 전 유역 면적의 약 60%에 해당되며 유로 연장이 66.2 km인 영산강 본류의 중류부를 대표하는 지점이다. 또한 선암 지점의 수위 관측소는 광주광역시 광산구 도산동의 장록교 직하(동경 126° 47' 08", 북위 35° 07' 49")에 위치하며, 황룡강 유역을 대표한다. 유역 면적은 약 526.56 km²로 영산강 전 유역의 15.6%에 해당되며 유로 연장은 약 58 km이며, 하류 합류 점으로부터의 거리는 4 km이다.

3.2 적용 자료 및 기간

본 연구에서는 모든 자료에 대해 전처리 과정을 위한 방법으로 평균이 0, 표준편차가 1이 되도록 하였으며, 나주 지점에 대해서는 일유출량만을 예측하기 위해 1995년 자료를 제외한 1993년~1998년까지의 일유출량 및 일강우량 자료를 시차를 두어 구성된 모형에 적용하였다. 그리고 적용된 일강우량에 대한 자료는 Thiessen 가중법을 이용한 유역 평균 강우량을 사용하였으며 각 해의 6월 1일~10월 31일까지만을 사용하였다. 나주 지점에 적용된 자료 중 보정을 위한 훈련 과정에는 1996년~1997년의 일자료를 적용하였으며, 훈련의 효율의 증진을 위해 사용된 Early Stopping Method of Training에 적용되는 검증 자료로는 1998년의 자료를 사용하였다. 그리고 일련의 훈련 및 검증의 과정에서 얻어진 매개 변수인 가중치(weights)를 1993년과 1994년의 자료에 적용하여 모형의 재검증을 실시하였다.

선암 지점에 대해서는 일유출량 및 시유출량의 예측을 실시하였으며, 각각 다른 시차를 두어 모형을 구성하였다. 일유출량을 예측하기 위해 선암 지점에 적용된 일자료로는, 훈련 과정에 1997년 자료를 사용하였으며, 1998년 자료를 검증 자료로 사용하였다. 이는 1996년 11월에 수위표의 이설로 인해 1997년 이후의 자료만을 적용하였다. 일강우량의 경우 나주 지점에서와 마찬가지로

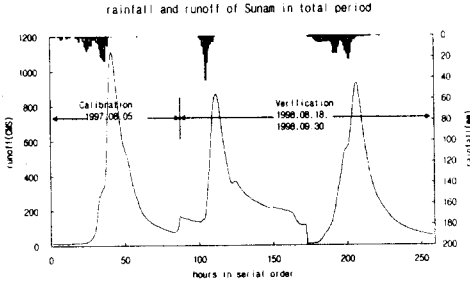


그림 3.1 선암 지점의 홍수 사상에 대한 유출량 및 강우량

유역 평균 강우량을 사용하였다. 또한 시유출량 예측을 위해 1997년~1998년 사이에 발생한 홍수 사상을 조사하여 유출량 및 강우량에 시차를 두어 사용하였다. 훈련 자료로는 1997년 8월 3일 10시~8월 6일 23시까지의 시자료를 사용하였다. 그리고 Early Stopping Method of Training을 위한 검증 자료로는 1998년 8월 17일 10시~8월 20일 23시까지의 자료를 사용하였으며, 이의 재검증을 위해 1998년 9월 29일 10시~10월 2일 23시에 발생한 홍수 사상을 적용하였다. 선암 지점의 시유출량 및 시강우량에 대한 도시는 그림 (3.1)과 같다.

4. 모형의 구성

4.1 일유출량 예측을 위한 모형

나주 지점에 대한 신경망 모형을 구성하기 위해 다음 식(4.1)~식(4.4)로 표현되는 4가지 모형을 적용하였으며, 선암 지점에 대해서는 다음의 식(4.5)와 식(4.6)을 적용하여 모형을 구성하였다. 각 모형들은 결정계수(R^2) 및 $RMSE$ 값을 비교하여 적절한 모형을 선별하였다.

$$q_k = \Phi(q_{k-1}, q_{k-2}, q_{k-3}, r_{k-1}, r_{k-2}) \dots \dots \dots (4.1)$$

$$q_k = \Phi(q_{k-1}, q_{k-2}, q_{k-3}, q_{k-4}, q_{k-5}, r_{k-1}, r_{k-2}, r_{k-3}, r_{k-4}, r_{k-5}) \dots \dots \dots (4.2)$$

$$q_k = \Phi(q_{k-1}, q_{k-2}, q_{k-3}, q_{k-4}, r_k, r_{k-1}, r_{k-2}, r_{k-4}) \dots \dots \dots (4.3)$$

$$q_k = \Phi(q_{k-1}, q_{k-2}, q_{k-3}, q_{k-4}, q_{k-5}, r_k, r_{k-1}, r_{k-2}) \dots \dots \dots (4.4)$$

여기서 $q_{k-1}, q_{k-2}, q_{k-3}, q_{k-4}, q_{k-5}, r_k, r_{k-1}, r_{k-2}, r_{k-3}, r_{k-4}, r_{k-5}$ 는 나주 지점의 각 선행 유출량과 유역 평균 강우량이다.

$$q_k = \Phi(q_{k(sa)-1}, q_{k(sa)-2}, r_{k(dg)-1}, r_{k(dg)-2}, r_{k(jsd)-1}, r_{k(jsd)-2}, r_{k(ss)-1}, r_{k(ss)-2}) \dots \dots \dots (4.5)$$

$$q_k = \Phi(q_{k(sa)-1}, q_{k(sa)-2}, r_{k(sa)}, r_{k(sa)-1}, r_{k(sa)-2}, r_{k(js)-1}, r_{k(js)-2}, r_{k(jsd)-1}, r_{k(jsd)-2}) \dots \dots \dots (4.6)$$

여기서 $q_{k(sa)-1}, q_{k(sa)-2}, r_{k(sa)}, r_{k(sa)-1}, r_{k(sa)-2}$ 는 선암 지점의 선행 유출량 및 소유역 평균 강우량이며, $r_{k(dg)-1}, r_{k(dg)-2}, r_{k(js)-1}, r_{k(js)-2}, r_{k(jsd)-1}, r_{k(jsd)-2}, r_{k(ss)-1}, r_{k(ss)-2}$ 는 각각 동곡 지점, 장성 지점, 장성댐 지점, 삼서 지점의 소유역 평균 강우량이다.

4.2 시유출량 예측을 위한 모형

그림 3.1에서 보는 바와 같이 평균 강우와 침투 유출간의 시차는 4~7까지로 판단되며, 입력 변수가 과대해지는 것을 막기 위해 두 부분으로 나누어 시차를 적용하였다. 그리고 선암 지점의 시유출량을 예측하기 위해 사용되어지는 강우량 자료는 Thiessen 가중법을 적용하지 않은 강우 관측소의 원자료를 그대로 적용하였으며, 그 모형들은 다음 식(4.7)과 식(4.8)과 같이 나타내었다.

$$q_k = \Phi \left(\begin{matrix} q_{k(sa)-1}, & q_{k(sa)-2}, & q_{k(sa)-3}, & r_{k(bg)-4}, & r_{k(bg)-5}, \\ r_{k(bg)-6}, & r_{k(dg)-4}, & r_{k(dg)-5}, & r_{k(dg)-6}, & r_{k(jsd)-4}, \\ r_{k(jsd)-5}, & r_{k(jsd)-6}, & r_{k(ss)-4}, & r_{k(ss)-5}, & r_{k(ss)-6} \end{matrix} \right) \dots \dots \dots (4.7)$$

$$Q_k = \Phi \begin{pmatrix} q_{k(sa)-1}, & q_{k(sa)-2}, & q_{k(sa)-3}, & r_{k(bg)-5}, & r_{k(bg)-6}, \\ r_{k(bg)-7}, & r_{k(dg)-5}, & r_{k(dg)-6}, & r_{k(dg)-7}, & r_{k(jsd)-5}, \\ r_{k(jsd)-6}, & r_{k(jsd)-7}, & r_{k(ss)-5}, & r_{k(ss)-6}, & r_{k(ss)-7} \end{pmatrix} \dots \dots \dots (4.8)$$

여기서 $q_{k(sa)-*}$ 는 시차별 선행 유출량이며, $r_{k(bg)-*}$, $r_{k(dg)-*}$, $r_{k(jsd)-*}$, $r_{k(ss)-*}$ 는 각각 북이, 동곡, 장성댐, 삼서 강우 관측소의 시차별 강우량이다.

5. 모의 결과

5.1 일유출량 예측 결과

각 모형들의 결정계수 (R^2) 및 $RMSE$ 값을 비교하여 적절한 모형을 선별하였으며, 나주 지점에 대해 채택된 모형들에 대한 값들은 표 5.1과 같다. 나주 지점에 대한 모형들은 모두 식(4.2)에 의한 것이며, 입력층의 노드의 수가 10개이고 은닉층의 노드의 수가 각각 23, 26, 28개이며, 출력층의 노드의 수는 1개인 것이 선택되었다. 채택되어진 모형들에 대한 보정(1996년~1997년)과 검증(1998년) 자료에 대한 관측값과 모의값의 도시는 다음 그림 5.1~5.6과 같으며, 관측값은 선으로 나타내었고 모의값은 점으로 표시하였다. 나주 지점에 대해 선별된 모형들의 보정 과정의 R^2 의 값은 모두 0.9 이상으로 나타났으며, 그에 대한 검증 결과는 1998년 자료가 가장 잘 맞는 것으로 나타났다. 또한 1993년과 1994년의 경우도 R^2 의 값이 각각 0.8 이상으로 양호한 결과를 보여주고 있다.

표 5.1 나주 지점의 일유출량 예측 결과

Model	R ² & RMSE		Calibration		Verification	R ²	RMSE
			1996~1997		1993	0.802	54.452
ANN(10,23,1)			R ²	RMSE	1994	0.881	14.936
			0.911	56.242	1998	0.909	92.671
			1996~1997		1993	0.821	51.584
ANN(10,26,1)			R ²	RMSE	1994	0.844	16.968
			0.909	56.714	1998	0.915	91.703
			1996~1997		1993	0.841	48.827
ANN(10,28,1)			R ²	RMSE	1994	0.858	16.197
			0.915	55.08	1998	0.916	90.920

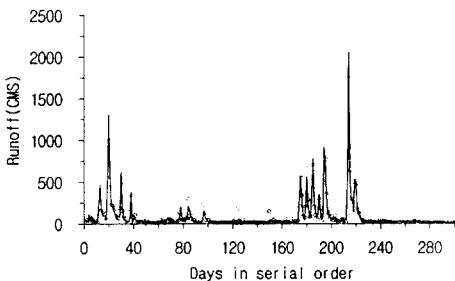


그림 5.1 ANN(10,23,1)모형의 보정 자료

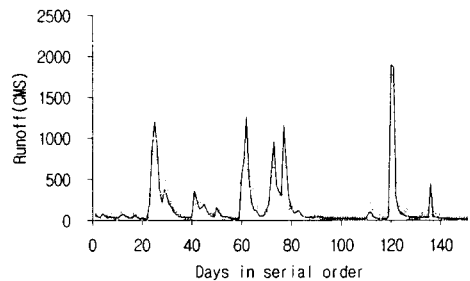


그림 5.2 ANN(10,23,1)모형의 검증 자료

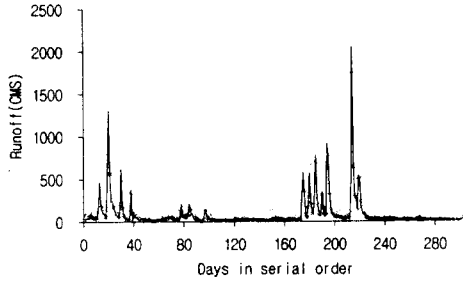


그림 5.3 ANN(10,26,1)모형의 보정 자료

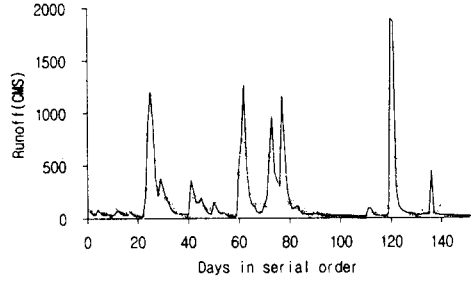


그림 5.4 ANN(10,26,1)모형의 검증 자료

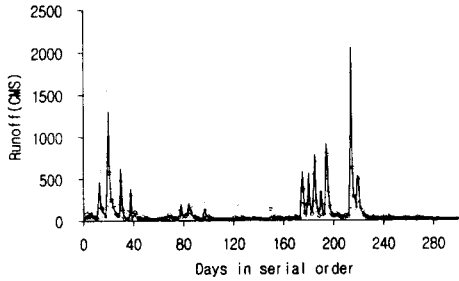


그림 5.5 ANN(10,28,1)모형의 보정 자료

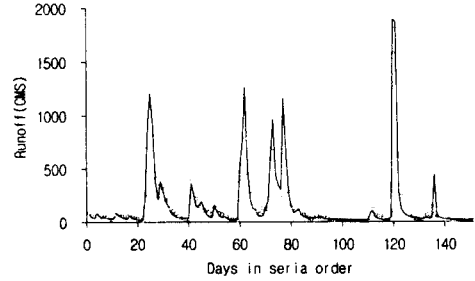


그림 5.6 ANN(10,28,1)모형의 검증 자료

선암 지점의 일유출량 예측에 대한 결과는 표 5.2에 나타내었다. 이는 식 4.6에 의한 것으로서 입력층의 노드의 수가 9개, 은닉층의 노드의 수는 각각 11개와 19개이며, 출력층의 노드의 수는 1개로 나타났다. 보정 과정의 R^2 값은 모두 0.9 이상의 값을 보여주었으며, 검증 과정에서도 역시 각각 0.907과 0.876의 값을 보여주어 대체적으로 양호한 결과를 보여 주었다.

표 5.2 선암 지점의 일유출량 예측 결과

	Calibration		Verification	
	R^2	RMSE	R^2	RMSE
ANN_SA(9,11,1)	0.993	5.752	0.907	29.497
ANN_SA(9,19,1)	0.989	7.015	0.876	35.060

5.2 시유출량 예측 결과

선암 지점에 대한 시유출량 예측 결과는 다음 표 5.3에 나타내었으며, 나주 및 선암 지점에 대한 일유출량 예측 결과와 비교하여 월등히 뛰어난 결과를 보여주고 있다. 각 모형의 구조는 입력층의 노드의 수가 15개이며, 은닉층의 노드의 수가 각각 19와 28개, 출력층의 노드의 수가 1개로 구성되었다. ANN_SAH1(15,19,1) 모형은 시차가 4~6까지 적용되어 있으며, ANN_SAH2(15,28,1) 모형은 시차가 5~7까지 적용되어 있다. 이와 같은 시차의 분할 적용은 입력층의 노드의 수가 과대해지는 것을 방지하기 위한 것이며, 그 결과 보정 및 검증 과정의 R^2 값은 0.9528에서 0.9998까지 나타났다. 그림 5.11~5.13까지는 ANN_SAH1(15,19,1) 모형에 대한 관측값과 모의값을 도시한 것이며, 관측값은 선으로 표현하였고, 모의값은 점으로 나타내었다. 침투 유량 부근에서 약간의 차이를 보이고 있으나, 대체로 뛰어난 예측력을 보여주고 있다.

표 5.3 선암 지점의 시유출량 예측 결과

	ANN_SAH1(15,19,1)		ANN_SAH2(15,28,1)	
	R ²	RMSE	R ²	RMSE
Calibration (1997.08.05)	0.9996	5.9931	0.9998	3.7797
Verification (1998.08.18)	0.9671	36.2530	0.9528	42.2770
Verification (1998.09.30)	0.9865	34.6167	0.9870	41.8035

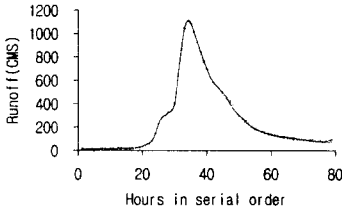


그림 5.11 ANN_SAH1(15,19,1) 모형의 보정 자료(1997.08.05)

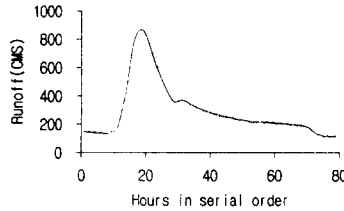


그림 5.12 ANN_SAH1(15,19,1) 모형의 검증 자료(1998.08.18)

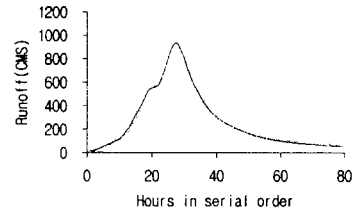


그림 5.13 ANN_SAH1(15,19,1) 모형의 검증 자료(1998.09.30)

6. 결 론

본 연구에서는 영산강 유역을 대상으로 하였으며, 특히 영산강 유역을 대표하는 나주 지점과 영산강의 지류인 황룡강 유역을 대표하는 선암 지점을 선택하여 일유출량 및 시유출량 예측을 실시하였다. 유출량 예측을 위한 적용 방법은 최근 많은 분야에서 널리 이용되고 있는 인공 신경망 이론을 이용하였다. 특히 본 연구에서 사용된 인공 신경망의 구조는 입력층, 은닉층, 출력층으로 구성되는 3계층 구조이며, 학습 알고리즘으로는 모멘텀 상수와 적응식 학습율을 적용한 역전파 학습 알고리즘을 사용하였다.

은닉층의 노드의 수에 따라 생성되는 다수의 모형들 중 R^2 값과 $RMSE$ 값을 비교하여 우수한 모형들을 선별하였다. 각 구조의 R^2 값은 보정 과정에서 0.909~0.915이며, 검증 과정에서 모두 0.8 이상을 나타내어 대체로 양호한 결과를 보여주었다. 특히 1998년의 검증 자료에 대해서는 0.909~0.916까지 나타났다. 그리고 선암 지점의 일유출량 예측 위해 선택된 모형들에 대한 R^2 값은 보정 과정에서 0.98 이상의 결과를 보여주었고, 검증 과정에서는 0.907과 0.876을 보여줌으로써 양호한 결과를 나타내었다. 또한 시유출량의 예측을 위한 경우에는 적용되는 시차의 차이를 두어 크게 두 가지로 신경망 구조를 구분하였다. 이에 의한 결과로는 보정 과정에서 R^2 값이 0.9996과 0.9998로 나타났으며, 검증 과정의 R^2 값은 0.9528~0.9870으로 나타나 일유출량 예측 결과와 비교하여 월등히 뛰어난 결과를 보여주었다.

참고문헌

- 이관수 외 3인, "역전파 학습 알고리즘과 적응식 학습율을 적용한 강우-유출 모형에 관한 연구", 전남대학교 공업기술연수지, 2000.
- 진영훈, "하천의 유출량 예측을 위한 인공 신경망 이론의 적용", 전남대학교 토목공학과 석사학위논문, 2000.
- 영산강 홍수 통제소, "영산강 홍수 예경보", 건설교통부, 1998.
- Kuo-lin Hsu, et al., "Artificial neural network modeling of the rainfall-runoff process", Water Resources Research, 1995.
- N.Sajikumar, et al., "Fundamentals of Neural Networks", Prentice Hall, 1994.
- Simon Haykin, "Neural Networks", Prentice Hall, 1999.