

# 인공신경망과 유전자알고리즘을 이용한 수위예측에 관한 연구

## Study on Water Stage Prediction by Artificial Neural Network and Genetic Algorithm

여운기\*, 지홍기\*\*, 이순탁\*\*\*

Woon Ki Yeo, Hong Kee Jee, Soontak Lee

### 요 지

최근의 극심한 기상이변으로 인하여 발생하는 유출량의 예측에 관한 사항은 치수·이수는 물론 방재의 측면에서도 역시 매우 중요한 관심사로 부각되고 있다. 강우-유출 관계는 유역의 수많은 시·공간적 변수들에 의해 영향을 받기 때문에 매우 복잡하여 예측하기 힘든 요소이다. 과거에는 추계학적 예측모형이나 확정론적 예측모형 혹은 경험적 모형 등을 사용하여 유출량을 예측하였으나 최근에는 인공신경망과 퍼지모형 그리고 유전자 알고리즘과 같은 인공지능기반의 모형들이 많이 사용되고 있다. 하지만 유출량을 예측하고자 할 때 학습자료 및 검증자료로써 사용되는 유출량은 수위-유량 관계곡선식으로부터 구하는 경우가 대부분으로 이렇게 유도된 유출량의 경우 오차가 크기 때문에 그 신뢰성에 문제가 있을 것으로 판단된다. 따라서 본 논문에서는 선행우량 및 수위자료로부터 단시간 수위예측에 관해 연구하였다.

신경망은 과거자료의 입·출력 패턴에서 정보를 추출하여 지식으로 보유하고, 이를 근거로 새로운 상황에 대한 해답을 제시하도록 하는 인공지능분야의 학습기법으로 인간이 과거의 경험과 훈련으로 지식을 축적해듯이 시스템의 입·출력에 의하여 연결강도를 최적화함으로써 모형의 구조를 스스로 조직화하기 때문에 모형의 구조에 적합한 최적 매개변수를 추정할 수 있다. 따라서 정확한 예측이 어려운 하천수위를 과거의 자료로부터 학습된 신경망의 수학적 알고리즘을 통해 유출량의 예측에 적용할 수 있을 것이다.

유전자 알고리즘은 적자생존의 생물학 원리에 바탕을 둔 최적화 기법중의 하나로 자연계의 생명체 중 환경에 잘 적응한 개체가 좀 더 많은 자손을 남길 수 있다는 자연선택 과정과 유전자의 변화를 통해서 좋은 방향으로 발전해 나간다는 자연 진화의 과정인 자연계의 유전자 메커니즘에 바탕을 둔 탐색 알고리즘이다. 즉, 자연계의 유전과 진화 메커니즘을 공학적으로 모델화함으로써 잠재적인 해의 후보들을 모아 군집을 형성한 뒤 서로간의 교배 혹은 변이를 통해서 최적 해를 찾는 계산 모델이다.

따라서 본 연구에서는 인공신경망의 가중치를 유전자 알고리즘에 의해 최적화시킨후 오류역전파알고리즘에 의해 신경망의 학습을 진행하는 모형으로 감천유역의 선산수위표지점의 수위를 1시간~6시간까지 예측하였다.

**핵심용어 : 인공신경망, 유전자 알고리즘, 수위예측**

## 1. 서 론

최근의 극심한 기상이변으로 인하여 발생하는 유출량의 예측에 관한 사항은 치수·이수는 물론 방재의 측면에서도 역시 매우 중요한 관심사로 부각되고 있다. 강우-유출 관계는 유역의 수많은 시·공간적 변수들에 의해 영향을 받기 때문에 매우 복잡하여 예측하기 힘든 요소이다. 과거에는

\* 정회원·영남대학교 대학원 박사과정 E-mail: adonas@nafree.net  
\*\* 정회원·영남대학교 건설시스템공학부 교수 E-mail: hkjee@yu.ac.kr  
\*\*\* 정회원·영남대학교 석좌교수 E-mail: leest@yu.ac.kr

추계학적 예측모형이나 확정론적 예측모형 혹은 경험적 모형 등을 사용하여 유출량을 예측하였으나 최근에는 인공신경망과 퍼지모형 그리고 유전자 알고리즘과 같은 인공지능기반의 모형들이 많이 사용되고 있다. 하지만 유출량을 예측하고자 할 때 학습자료 및 검정자료로써 사용되는 유출량은 수위-유량 관계곡선식으로부터 구하는 경우가 대부분으로 이는 이렇게 유도된 유출량의 경우 오차가 크기 때문에 그 신뢰성에 문제가 있을 것으로 판단된다. 따라서 본 논문에서는 선행유량 및 수위자료로부터 단시간 수위예측에 관해 연구하였다.

인공신경망은 다양한 문제에 대해 적응력이 뛰어나며, 입력으로 들어오는 신호를 분석하여 미리 정의된 부류 중 어떤 것에 속하는지 알아내는 문자 인식, 음성 인식, 패턴 분류등에 주로 사용되고 있으나 과거의 정보를 기본으로 하여 새로운 상황에 대한 재현성이 뛰어나기 때문에 수문계의 모형화와 예측수단으로도 활용되고 있다. French 등(1992)은 신경망을 이용하여 1시간후의 사상에 대해 모의한 바 있으며, Kuligowski와 Barros(1998)은 강우관측소의 선행강우량을 이용하여 6시간까지의 누가강우량을 예측한 바 있다.

유전자 알고리즘은 독일의 Rechenberg(1965)에 의해 진화전략이 발표됨으로써 처음 도입되기 시작하였고 Fogel 등(1966)에 의해 진화 프로그램이 제안되어 교차 연산 없이 변이를 사용하였다. John Holland(1975)는 이진 스트링의 개체집단 위에서 모의진화를 일으켜 효율적으로 최적해를 탐색하는 알고리즘을 고안해 내어 최초의 유전자 알고리즘을 개발한 바 있다.

Goldberg (1989) 와 Michalewicz (1992)는 유전자 알고리즘을 이용하여 비선형 최적화문제를 해결하는 매커니즘과 강인성에 대해 논의 하였으며, Montana와 Davis (1988) 그리고 Maniezzo (1994)는 유전자 알고리즘을 이용하여 오류역전과 신경망의 훈련에 이용한 바 있으며, 우수한 결과를 보인 것으로 나타났다.

본 연구에서는 인공신경망의 가중치를 유전자 알고리즘에 의해 최적화시킨후 오류역전과 알고리즘에 의해 신경망의 학습을 진행하는 모형으로 감천유역의 선산수위표지점의 수위를 1시간~6시간까지 예측하였다.

## 2. 인공신경망과 유전자 알고리즘

### 2.1 인공신경망

인간의 뇌는 뉴런(Neuron)이라는 최소 단위의 신경 세포들이 시냅스(Synapse)라는 물질로 연결되어 있으며 이 연결이 복잡해져 거대한 그물망으로 구성되어 있는데 이러한 구조를 단순화하여 만든 연산모형을 인공신경회로망(Artificial Neural Network)라고 하며, 이 모델에서는 문제를 해결하기 위한 직접적인 지식이나 방법을 입력하지 않아도 뉴런들의 학습을 통한 최적화로 문제를 해결할 수 있다.

인공신경망의 계산은 입력층에 들어온 외부자극의 가중합을 활성화함수를 통해 적절한 반응을 출력하게 된다. 여기서 외부자극의 가중합은 식(1)과 같이 표현할 수 있다.

$$NET_j = \sum_{i=1}^n x_i w_{ij} \quad (1)$$

여기서,  $NET_j$ 는  $j$ 번째 출력뉴런에 들어오는 외부자극의 가중합을 의미하며,  $x_i$ 는 외부자극,  $w_{ij}$ 는 출력뉴런과  $x_i$ 사이의 연결강도를 의미한다.

이 때, 출력뉴런의 반응값은 활성화함수에 의해 결정되는데 아날로그 출력이 가능한 단극성 Sigmoid함수가 많이 사용되며, 이는 식(2)와 같다..

$$f(NE T_j) = \frac{1}{1 + \exp^{-\lambda NE T_j}} \quad (2)$$

여기서,  $\lambda$ 값에 따라 Sigmoid함수의 형태가 변하게 되는데 일반적으로는 오버플로우 혹은 언더플로우를 방지하기 위해  $\lambda = 1$ 을 사용한다.

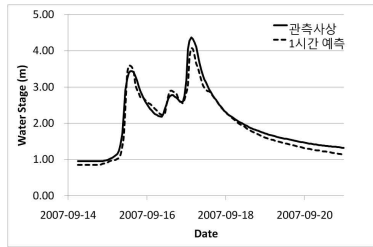
## 2.2 유전자 알고리즘

유전자 알고리즘은 적자생존의 생물학 원리에 바탕을 둔 최적화 기법중의 하나로 자연계의 생명체 중 환경에 잘 적응한 개체가 좀 더 많은 자손을 남길 수 있다는 자연선택 과정과 유전자의 변화를 통해서 좋은 방향으로 발전해 나간다는 자연 진화의 과정인 자연계의 유전자 메커니즘에 바탕을 둔 탐색 알고리즘이다. 즉, 자연계의 유전과 진화 메커니즘을 공학적으로 모델화함으로써 잠재적인 해의 후보들을 모아 군집을 형성한 뒤 서로간의 교배 혹은 변이를 통해서 최적 해를 찾는 계산 모델이다.

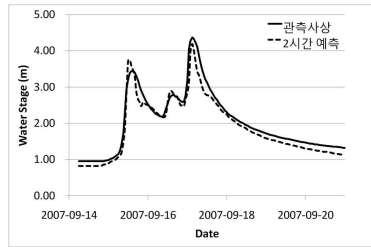
유전자 알고리즘은 목적함수에 따라 유전자와 군집의 크기를 결정하여 초기 유전자 집단을 생성한 후 적합도를 계산하고 이를 바탕으로 다음 세대를 결정한다. 재생산 과정은 선택(Selection), 교배(Crossover), 돌연변이(Mutation) 연산자를 이용하여 해 집단의 다양성을 높이고 적합도가 큰 개체의 발생을 기대한다. 적합도가 기준에 만족될 때까지 재생산과정을 반복한다. 일반적으로 최적화에 관련된 유전자알고리즘은 유전자집단을 0과 1의 2진 비트스트링(bit string)을 사용하는 경우가 많으나 본 연구에서는 유전자 알고리즘을 사용하여 신경망의 연결강도를 최적화하기 위해 실수코딩 유전자 알고리즘(Real-coded Genetic Algorithm)을 사용하였다.

## 3. 모형의 적용 및 결과

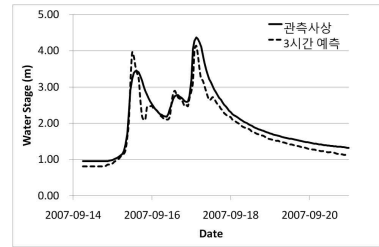
본 연구에서는 그림 3과 같이 낙동강의 중류지역에 위치한 감천유역의 선산수위표의 수위를 예측하기 위해 감천 유역내의 강우관측지점과 수위관측지점의 자료를 사용하여 모형을 구성하였다. 수위 예측을 위한 모형은 신경망과 유전자 알고리즘의 결합모형을 사용하여 유전자 알고리즘에 의해 신경망의 연결강도를 구한 뒤 오류 역전파 알고리즘에 의해 다시 학습시켰다. 입력자료는 감천유역의 시간평균강우량과 선산수위표의 상류수위표지점인 김천수위표의 수위자료를 사용하였다. 홍수시의 수위자료 예측을 위해 기존의 강우사상중에서 1991년, 1993년, 1997년, 2000년, 2001년의 자료를 사용하였으며, 모형의 검정을 위해 2007년의 사상을 사용하였다. 입력자료로 사용할 강우량 및 수위의 지체시간을 구하기 위해 선산수위표지점의 수위와 감천유역의 면적평균강우량 및 김천수위표지점의 수위에 대한 교차상관분석을 실시하였다. 그 결과 감천유역 면적평균강우량의 지체시간은 상관도가 가장 높게 나타난 12, 13, 14, 15시간으로 선정하였으며, 김천수위표지점의 지체시간은 3, 4, 5, 6시간을 선정하여 이들을 각각 입력변수로 구성하였다. 또한 은닉층은 1개 층을 사용하였으며, 은닉층의 노드수는 입력층의 노드수와 같은 8개를 사용하여 분석하였다. 이에 따라 1시간~6시간 후의 수위를 예측한 결과는 그림 2와 같다.



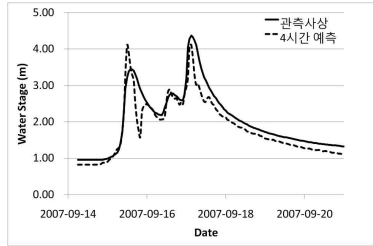
(a) 1시간 예측결과



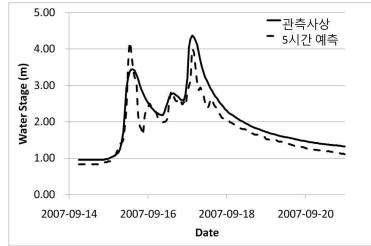
(b) 2시간 예측결과



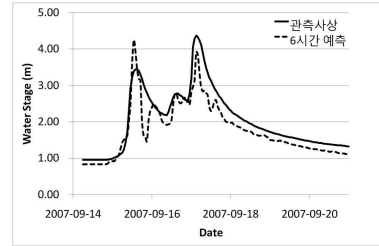
(c) 3시간 예측결과



(d) 4시간 예측결과



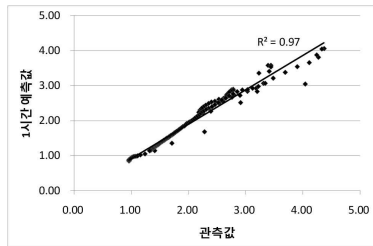
(e) 5시간 예측결과



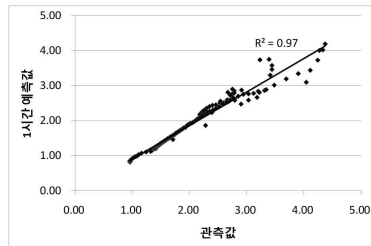
(f) 6시간 예측결과

그림 2. 수위예측 결과

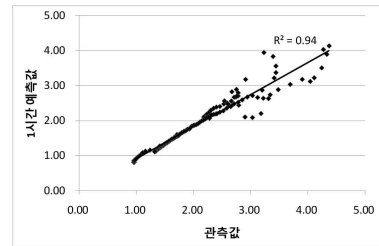
선산지점의 수위를 1시간~6시간 후의 수위를 예측한 결과 1시간 예측의 경우가 가장 실제의 값을 잘 예측하고 있으며 예측시간이 길어질수록 오차가 커지는 것을 알 수 있었다. 특히 검정에 이용된 사상이 두 개의 첨두를 포함하고 있는데 뒤의 첨두에 대한 예측은 예측시간이 길어져도 오차가 그렇게 커지지 않는 반면 앞의 첨두에 대해서는 오차가 갈수록 커지는 양상을 보였다. 각 예측에 대하여 상관분석을 해본결과 결정계수는 1시간 예측에 대해서 0.97로 상당히 높은 값을 보여 주었으며 예측시간이 길어질수록 결정계수의 값이 줄어들지만 6시간 예측에 대해서도 0.88로 상당히 높은 예측력을 보여주었다.



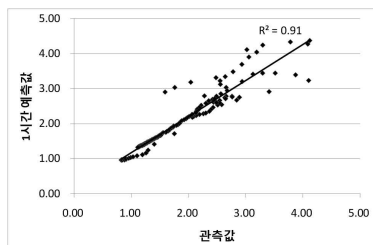
(a) 1시간 예측



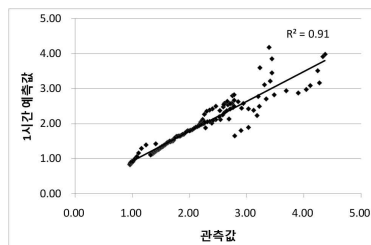
(b) 2시간 예측



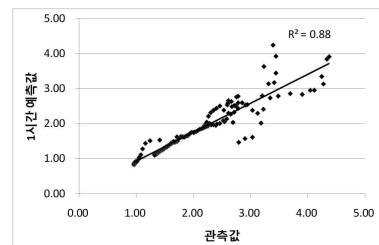
(c) 3시간 예측



(d) 4시간 예측



(e) 5시간 예측



(f) 6시간 예측

그림 3. 각 예측시간별 결정계수

#### 4. 결 론

본 연구에서는 인공신경망과 유전자 알고리즘을 결합한 모형을 이용하여 감천유역의 선산 수위표지점의 수위를 예측하기 위해 감천유역의 과거 강우사상과 감천수위표지점의 수위를 학습시키고 새로운 강우에 대해 1시간~6시간 후의 수위를 예측해 보았으며, 다음과 같은 결론을 얻을 수 있었다.

- 1) 강우 및 수위의 시계열자료는 비선형구조의 아날로그 데이터이며 이를 분석하기 위해 다층 신경망을 사용하였고 학습방법으로는 유전자 알고리즘과 오류역전파알고리즘을 사용하였으며, 활성화함수로는 단극성 시그모이드함수를 사용하였다.
- 2) 감천의 유역평균강우량과 감천지점의 수위를 학습시키고 새로운 강우에 대해 1시간~6시간 후의 수위를 예측한 결과 전체적으로 수위의 변화 추세를 잘 표현하고 있으며, 두 개의 첩두중에서 하나의 첩두는 예측의 정도가 매우 높았으나 다른 하나의 첩두에 대해서는 예측시간이 길어질수록 오차가 높아짐을 알 수 있었으며, 단일 첩두를 가진 사상에 대해서는 예측이 잘 될 것으로 기대된다.
- 3) 예측모형에 대한 결정계수는 각각 0.97~0.88로 예측시간이 길어질수록 값이 줄어들어 예측력이 조금씩 떨어짐을 알 수 있으나 가장 작은 결정계수도 0.88로 상당히 큰 값을 보여주고 있어 우수한 예측력을 가진 것을 알 수 있었다.

#### 참 고 문 헌

1. 오창석(2000), 뉴로 컴퓨터 개론, 내하출판사
2. 진강규(2000), 유전알고리즘과 그 응용, 교우사
3. 오성권(2002), 프로그래밍에 의한 컴퓨터지능(퍼지, 신경회로망, 진화알고리즘을 중심으로), 내하출판사
4. French, M.N., Krajewski W.F., Cuykendall R.R.(1992), Rainfall forecasting in space and time using a neural network, Journal of Hydrology, Vol.137, pp.1-31
5. Goldberg, D. E. (1989), Genetic algorithms in search, optimization, and machine learning. Reading, MA: Addison-Wesley
6. Maniezzo, V. (1994), Genetic evaluation of the topology and weight distribution of neural network. IEEE Transaction of Neural Network, Vol.5 No.1, pp.39 - 53
7. Michalewicz, Z. (1992), Genetic algorithms+ data structures = evolution programs. (3rd ed.) Springer-Verlag
8. Montana, D. J., & Davis, L. (1988), Training Feed-Forward Neural Networks using genetic algorithms. In: Kufmann, M. (Ed.), 11th International joint conference on artificial intelligence, Vol.1, pp.762 - .767, San Mateo: CA
9. Robert J. Kuligowski and Ana P. Barros (1998), Experiments in Short-Term Precipitation Forecasting Using Artificial Neural Networks, Monthly Weather Review, Vol.126, pp. 470-482
10. Whitley, D. (1988), Applying Genetic Algorithms to Neural Network Problems, International Neural Network Society p.230