

연구논문

뉴로 유전자 결합모형을 이용한 상수도 1일 급수량 예측

이경훈* · 강일환** · 문병석*** · 박진금****

전남대학교 토목공학과*, 경호엔지니어링(주) 상하수도부**,

서남대학교 토목공학과***, 대한주택공사 광주전남본부****

(2005년 4월 25일 접수, 2005년 7월 25일 승인)

Prediction of Daily Water Supply Using Neuro Genetic Hybrid Model

Kyoung-Hoon Rhee* · Il-Hwan Kang** · Byoung-Seok Moon*** · Jin-geum Park****

Department of Civil Eng., Chonnam National University*,

Department of Water and WasteWater, Kyong-ho Eng.**,

Department of Civil Eng., Seonam University***,

The Head Office of GwangJu Chonnam, Korea National Housing Corporation****

(Manuscript received 25 April 2005; accepted 25 July 2005)

Abstract

Existing models that predict of Daily water supply include statistical models and neural network model. The neural network model was more effective than the statistical models. Only neural network model, which predict of Daily water supply, is focused on estimation of the operational control. Neural network model takes long learning time and gets into local minimum.

This study proposes Neuro Genetic hybrid model which a combination of genetic algorithm and neural network. Hybrid model makes up for neural network's shortcomings. In this study, the amount of supply, the mean temperature and the population of the area supplied with water are use for neural network's learning patterns for prediction.

RMSE(Root Mean Square Error) is used for a MOE(Measure Of Effectiveness). The comparison of the two models showed that the predicting capability of Hybrid model is more effective than that of neural network model. The proposed hybrid model is able to predict of Daily water, thus it can apply real time estimation of operational control of water works and water drain pipes.

Proposed models include accidental cases such as a suspension of water supply. The maximum error rate between the estimation of the model and the actual measurement was 11.81% and the average error was lower than 1.76%. The model is expected to be a real-time estimation of the operational control of water works and water/drain pipes.

Key words : Local Minimum, Neural Network, Neuro Genetic, Hybrid Model

1. 서론

산업사회의 발달과 더불어 물 수요량이 급격하게 증가함에 따라서 수도물의 안정적인 공급과 효율적인 이용이 보다 절실하게 요구되고 있다. 이와 같은 요구에 부응하기 위해서는 체계적이고 과학적인 연구에 의해 수도물 수요가 정확하게 예측되어야 한다 (이경훈 등, 1994).

급수량 예측방법은 시점에 따라 단기예측과 장기예측으로 분류할 수 있다. 단기예측은 시간 및 일일예측으로 급·배수량의 실시간 제어에 이용되고, 장기예측은 순, 월, 계절, 연별 예측으로 상수도시설을 계획하거나 설계하는데 필요한 자료로 활용되고 있다.

급수량의 수요예측방법에는 크게 다변량 해석방법과 시계열 분석방법이 있다. 최근의 급수량 예측을 위한 연구(문병석, 1998)를 보면 급수량, 급수인구, 기후자료를 입력자료로 하여 단기 급수량을 예측하고자 하였고 또한 보다 양호한 모의결과를 얻기 위해 퍼지 이론과 신경망의 융합 기술 중 하나인 ANFIS (Adaptive Neuro - Fuzzy Inference System)를 적용해 1일 급수량을 예측하였다(강일환, 1998). 최근 인공지능 모형인 신경망과 퍼지 및 유전자 알고리즘은 여러 가지 공학분야 들(패턴인식, 음성인식, 시스템 동정화와 제어, 예측 등)에 성공적으로 이용되고 있다.

본 연구에서는 상수도시설을 효율적으로 운영하는 데 필요한 1일 급수량 수요를 예측하고자 신경망에 유전자 알고리즘을 결합한 뉴로 유전자 결합모형을 적용하고 기존의 신경망 모형과 비교·검토하였다. 결합모형은 가중치의 최적화를 위해 초기 연결 가중치는 유전자 알고리즘의 전역탐색능력을 이용하

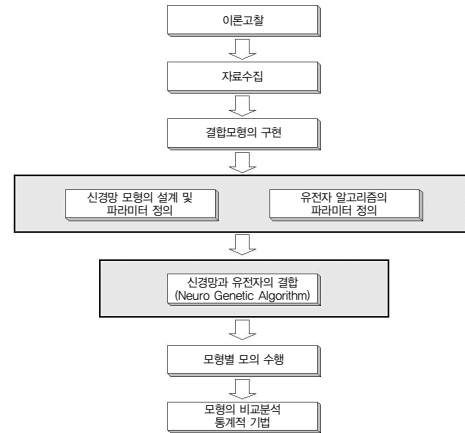


Figure 1. Procedure of Study

여 근사 가중치 값을 빠른 학습속도로 찾고, 최적의 연결가중치는 신경망의 역전파 학습알고리즘으로 찾는 방법이다.

본 연구의 목적은 기존 신경망 이론을 이용한 예측 방법에서 단점으로 지적된 과도한 학습시간과 진동 학습으로 인한 문제를 유전자 알고리즘의 결합형태를 통해 극복하고 결합모형이 실시간 예측모형으로서 적합함을 알리고자 한다.

또한 제안된 모형식은 사고 등의 인위적인 조작(단수 등)이 가해지는 시기를 포함하고도 실측지와 모형의 추정치와의 오차율이 적어 모형의 결과는 상수도 1일 급수량 추정에 필요한 시설에 적용 가능하다고 판단된다.

II. 1일 급수량 예측모형

1. 신경망모형

본 논문에서 사용된 인공신경망의 종류는 다층

(multi-layer)신경망이며, 입력층과 은닉층, 그리고 출력층으로 구성된다. 또한 사용된 학습알고리즘은 모멘텀 상수와 적응식 학습율이 적용된 최급강하법을 이용한 역전파 학습알고리즘이며, 이에 포함된 전달 함수는 식(1)의 선형(Linear)함수와 식(2)의 비선형(Nonlinear)인 단극성 시그모이드(Sigmoid)함수를 사용하였다.

$$\Phi(x) = x \tag{1}$$

$$\Phi(r) = (1 + e^{-r})^{-1} = \frac{1}{1 + e^r} \tag{2}$$

2. 뉴로 유전자 결합모형

위의 신경망에 관련된 연구에서는 공통적으로 역전파 신경망이 지역최소점(준최적해)에 빠질 수 있고, 학습패턴이 많은 경우 학습시간이 매우 길어지는 문제점을 지적하였다. 이러한 역전파 신경망이 가지는 단점을 극복하기 위해 유전자 알고리즘의 결합모

형을 구현한다. 다음은 신경망의 단점을 극복하기 위한 유전자 알고리즘의 결합이다.

일반적으로 유전자 알고리즘과 신경망의 결합에는 몇 가지 방법을 생각 할 수 있다.

첫째, 유전자 알고리즘을 이용하여 연결강도와 bias의 조정.

둘째, GA를 이용하여 신경망의 구조를 결정하는 것.

셋째, 또한 그것들의 결합 등이다.

본 연구에서는 먼저 신경망의 구조를 결정한 후 유전자 알고리즘을 이용하여 연결강도 등을 효율이 우수한 최적해 근방까지 학습하게 하였다. 이렇게 학습되어 국소해에 빠질 가능성을 감소시킨 연결강도 등을 다시 신경망에 사상시키는 방법을 사용하였다.

본 연구에서 실시한 유전자 알고리즘에 의한 신경망 연결강도 등의 학습을 위한 구성도는 그림 2와 같으며, 신경망 연결강도 등을 유전자 알고리즘의 염색

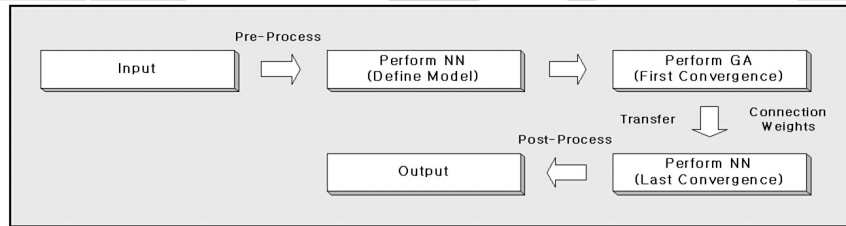


Figure 2. Combination of NN and GA

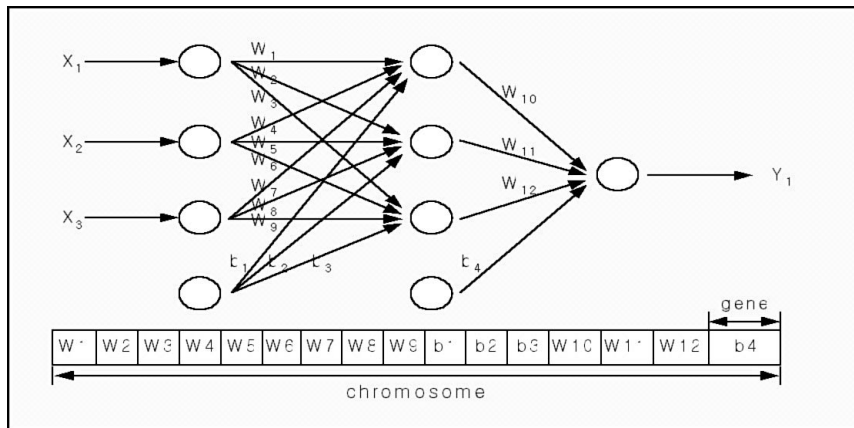


Figure 3. Chromosome of connection weights and biases

체 상에 사상시키는 방법은 직접 네트워크에 가중치 (weight)나 bias와 같은 연결강도 등을 표현하여 학습시키는 수법을 사용하였다.

본 연구에서 실시한 유전자알고리즘과 신경망의 결합 방법은 각 염색체에 실치표현(real value encoding)을 사용하였으며, 그 방법의 도식적인 표현은 그림 3과 같다.

3. 훈련 매개변수 결정

본 연구는 최적의 신경망모형을 선별하고자 하는 것이 아닌 신경망모형의 단점을 극복하는 결합모형을 구현하고자 하는 것이므로 입력층수(n)에 따른 은닉층 유닛의 개수를 n으로 고정하였다.

유전자 알고리즘을 실행하기 위해서 결정해야 할 제어 파라미터는 집단내 개체의 개수, 집단의 크기, 세대수, 교배율, 돌연변이율이 있다. 집단의 크기와 세대수의 경우 기준이 되는 값은 정해진 것이 아니라 교배율, 돌연변이율과 함께 실험을 통한 경험적인 값을 사용하게 되므로 대안별로 충분히 고려하여 모형의 최적 파라미터를 정하였다. 그리고 선택방식은 비선형 순위(non-linear ranking)에 의한 확률적 균등표본(stochastic sampling)방식을, 교배방식은 불연속 재조합법(Discrete recombination)을 사용하였다. 집단내 개체의 적합도 함수는 신경망의 목적함수인 평균 자승오차 함수(F)를 사용하여 오차를 최소화

로 하는 방향으로 탐색을 수행하게 한다.

$$F = \frac{(t_j - a_j)}{2}$$

t_j : j뉴런의목적값, a_j : j뉴런의출력값 (3)

한편, 본 연구에서 사용된 신경망과 유전자 알고리즘의 매개변수는 표 1과 같이 나타내었다.

III. 대상지역 및 연구자료분석

본 연구의 대상지역으로 광주광역시를 선정하였고, 1997년 1월 1일~2002년 12월 31일까지의 급수량, 기후, 급수인구 자료를 활용하였다. 이 자료 중 1997년~2001년의 급수량, 기후 및 급수인구 자료는 모형의 학습자료로 사용하였고, 2002년(1. 1~12. 31)의 자료를 검증자료로 활용하였다.

또한, 광주광역시 1일 급수량 예측을 위한 통계적 분석자료로 원시계열 자료를 사용하였다. 독립변수를 동정하기 위해 단계선택법을 이용하였으며, 자기상관 및 상호상관분석을 통해 동정된 변수에 대한 지체시간(lag time)을 결정하여 모형을 구성하였다. 이렇듯 본 연구에서 제안된 모형이 지체시간을 고려한 모형만으로 구성된 것은 우선 지체시간이 고려된 경우와 고려치 않은 경우의 신경망 모형을 구성하여 모의결과를 분석한 결과 지체를 고려한 경우의 모형이 지체를 고려치 않은 경우의 모형보다 우수한 결과를 보였기 때문이다. 이에 뉴로 유전자 결합모형의 경우에도 지체시간을 고려한 모형만을 본 연구에 제안하였다.

또한 기상청의 기후예보자료를 신뢰할 수 없는 경우에도 급수량을 예측할 수 있도록 급수량 자료만으로 모형을 구성해야 하는데, 모형을 구성하는 경우는 예측 5일전부터 예측전날(Q_{t-4} , Q_{t-3} , Q_{t-2} , Q_{t-1} , Q_t)까지의 급수량 자료만으로 모형을 구성하여야 할 것으로 판단된다.

Table 1. List of training parameters

Tools	Parameters	Variable	Value
Neural Network	Learning rate	lr	0.9
	Momentum factor	mc	0.7
	Iteration No.	epochs	10,000
	Terminate goal	goal	0.00001
Genetic Algorithm	Generation gap	ggap	0.8
	Insertion rate	insr	0.9
	Crossover rate	xovr	1
	Selective pressure	sp	2
	Mutaton rate	mutr	0.15
	Migration rate	migr	0.2
	Terminate value	termexact	0.00001

IV. 모형의 구성

연구자료분석에서 살펴보았듯이 광주광역시 1일 급수량(Q_{t+1}) 예측을 위한 모형은 표 2와 같다.

표 2는 독립변수와 알고리즘에 따른 모형의 구성을 보여준다. 독립변수는 동정된 급수량, 평균기온, 급수인구와 급수량만으로 되어 있고, 데이터의 범위는 학습자료로 97~2001년을 사용하였고 검증자료로 2002년(1. 1~12. 31)을 사용하였다. 학습알고리즘이 신경망인 경우의 모형은 NN, 뉴로 유전자 결합 모형인 경우는 N-G로 표기하였다. 또한 기후자료와 급수인구, 급수량 자료로 입력변수가 구성된 경우에는 각각의 모형에 1을 급수량 자료만으로 입력변수가 구성된 경우는 각각의 모형에 2를 표기하여 각각의 모형을 분류하였다.

Table 2. Construction of model according to input variable and algorithm

Range Model	Input variable	Algorithm
NN1	$Q_b, Q_{t-1}, Q_{t-2}, P_{t+1}, T_b, T_{t+1}$	Neural Network
NN2	$Q_b, Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}$	Neural Network
N-G1	$Q_b, Q_{t-1}, Q_{t-2}, P_{t+1}, T_b, T_{t+1}$	Neuro Genetic
N-G2	$Q_b, Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}$	Hybrid Algorithm

V. 분석방법 및 결과고찰 내용

1. 분석방법

1일 급수량 예측모형의 적용타당성을 평가하기 위하여 각 모형에서 계산된 결과를 오차 및 통계적인 검

증방법으로 검토하였다. 본 연구에서는 시계열 자료에 대한 적용타당성 문제를 다룰 때 가장 일반적으로 사용되고 있는 상관계수(Correlation Coefficient: CC), 결정계수(Determination Coefficient :R), 절대 평균오차율, 절대평균오차(Mean Absolute Error; MAE) 및 평균제곱근오차(Root Mean Square Error; RMSE)를 사용하였다.

2. 결과고찰

표 3은 제안된 모형의 오차 및 오차율 그리고 상관계수 및 결정계수를 나타내고 있다.

위의 분석을 통해 상수도 1일 급수량 예측에는 제안된 모형 중 N-G2(급수량만으로 구성된 뉴로 유전자 결합모형)이 절대평균오차 7,430m³, 평균제곱근오차 9,506m³, 최대오차율 11.81%, 평균오차율 1.76%로 가장 양호한 결과를 보이고 있다. 그리고 상관계수 및 결정계수도 0.94 및 0.88로 다른 모형에 비해 양호한 결과를 나타내고 있다. 또한 급수량만으로 구성된 모형이 급수량, 평균기온 및 급수인구로 구성된 모형보다 양호한 결과를 보였고, 학습알고리즘별로는 뉴로 유전자 결합모형, 신경망 모형순으로 우수한 예측결과를 보이고 있다.

그림 4는 신경망모형의 예측결과를 나타낸 그래프로 NN1은 급수량, 급수인구, 평균기온으로 구성된 모형이고 NN2는 급수량만으로 구성된 모형으로 NN2모형이 NN1모형에 비해 우수한 예측결과를 보임을 알 수 있다. 그림 5는 뉴로 유전자 결합모형의 예측결과를 나타낸 그래프로 N-G1은 급수량, 급수인구, 평균기온으로 구성된 모형이고 N-G2는 급수

Table 3. Errors & Coefficient of Correlation and Determination between the Estimation and the Actual Value

Error Model	MAE(m ³)	RMSE(m ³)	Max Error(%)	Mean Error(%)	Correlation Coefficient	Determination Coefficient
NN1	11,045	13,938	15.43	2.88	0.86	0.74
NN2	9,062	10,679	17.25	2.17	0.88	0.77
N-G1	8,356	11,082	16.64	2.01	0.94	0.88
N-G2	7,430	9,506	11.81	1.76	0.94	0.88

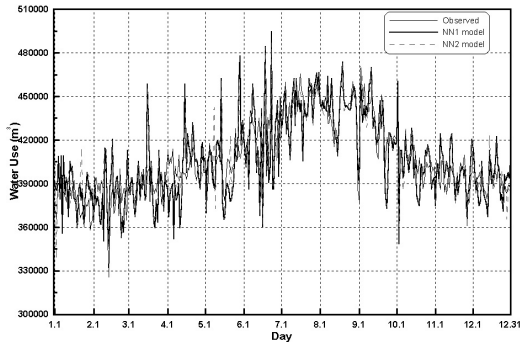


Figure 4. Estimation of daily urban water demand using NN1, NN2 model, 2002

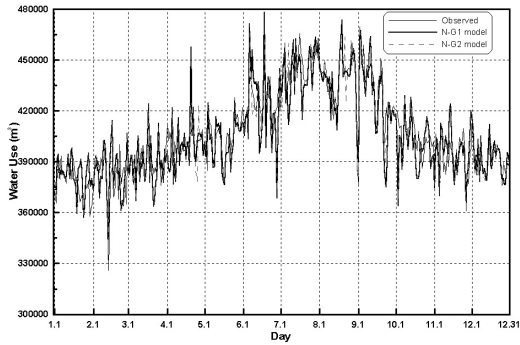


Figure 5. Estimation of daily urban water demand using N-G1, N-G2 model, 2002

량만으로 구성된 모형으로 N-G2모형이 N-G1모형에 비해 우수한 예측결과를 보임을 알 수 있다.

그림 6~9는 NN1과 NN2모형 및 N-G1과 N-G2 모형의 산포도를 도시한 것으로 실측치에 대한 예측치의 분포를 나타내고 있다. 위 그림을 통해 예측치가 실측치를 잘 반영하고 있음을 알 수 있다.

신경망 및 결합모형의 전체적인 통계값을 살펴본 결과 급수량, 급수인구, 평균기온으로 구성된 모형보다 급수량만으로 구성된 모형이 우수한 예측결과를 보임을 알 수 있었다. 또한 학습알고리즘별로는 뉴로 유전자, 신경망 모형순으로 우수한 예측값을 보이고 있다.

제한된 결합모형에서도 신경망모형과 마찬가지로 인위적인 조작(단수 등)이 가해지는 시기를 제외하면 높은 예측력을 보이고 있다.

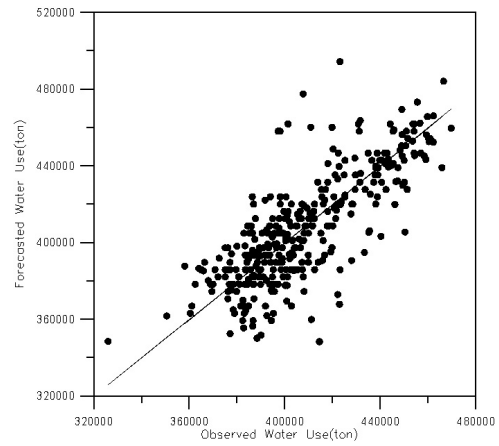


Figure 6. Scatter diagram of NN1

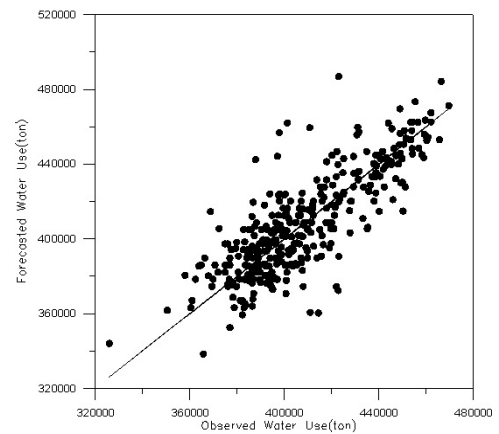


Figure 7. Scatter diagram of NN2

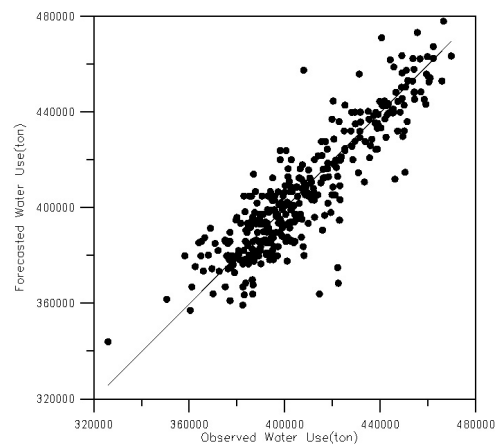


Figure 8. Scatter diagram of N-G1

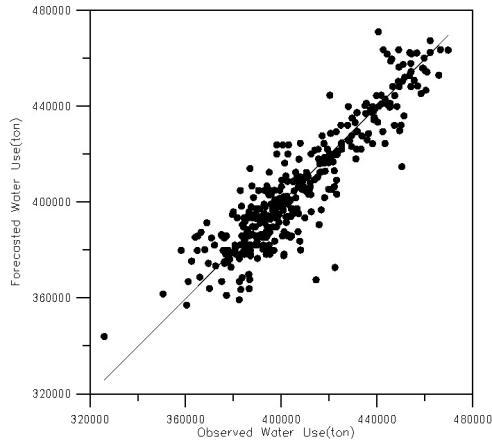


Figure 9. Scatter diagram of N-G2

VI. 결 론

본 연구에서는 상수도 1일 급수량 예측을 위하여 광주광역시를 대상으로 상수도 1일 급수량자료, 인구자료 및 기후자료를 이용하여 신경망 모형 및 뉴로 유전자 결합모형을 구성하여 예측정확도를 비교·검토하고 다음과 같은 결론을 얻었다.

1. 예측정도가 높은 순서는 N-G2>N-G1>NN2>NN1으로 신경망을 사용했을 때 보다 결합모형을 사용했을 때 더 좋은 예측력을 가진다는 것을 보았다. 또한 예측일의 급수량, 평균기온, 급수인구로 구성된 NN1, N-G1모형보다 예측 5일전부터 예측전일까지의 급수량만으로 구성된 NN2, N-G2모형으로 양호한 모의결과를 얻을 수 있었다.

2. RMSE(평균제곱근오차)를 비교적으로 각 모형을 비교하면 신경망 모형인 NN1, NN2는 각각 13,938m³, 10,679m³를, 결합모형인 N-G1, N-G2는 각각 11,082m³, 9,506m³으로 나타나 신경망만을 이용한 경우보다 결합모형을 이용한 경우가 양호한 예측력을 보였으며 이는 결합모형이 유전자 알고리즘을 이용하여 오차를 최소화하는 값으로 신경망의 초기 연결가중치를 조절하였기 때문이다.

3. 제안된 결합모형 중 N-G2 모형은 사고 등의 인위적인 조작, 단수 등이 가해지는 시기를 포함하고

도 실측치와 모형의 예측치와의 오차율이 최대 11.81%, 평균 1.76%, 절대평균오차 7,430m³, 평균 제공근오차 9,506m³, 상관계수와 결정계수는 각각 0.94와 0.88로 나타나, 모형의 결과는 상수도 시설의 운용 및 급·배수관망의 제어에 많은 도움을 주리라 생각된다.

본 연구가 실시간 수운용 관리에 적용되려면 과거 시간별 급수량 자료가 축적되는 것이 선행되어야 하고 자료가 방대해 지는 경우 분석시 학습에 소요되는 전체시간을 고려한 연구가 필요할 것으로 본다.

또한, 본 연구에서는 신경망의 단점인 지역 최소점에 빠지는 문제를 결합모형에서는 발견할 수 없었다. 하지만 결합모형 역시 지역 최소점에 빠지는 문제를 완전히 극복했다고 말할 수 없으며 단지 지역 최소점에 빠지는 문제를 확률적으로 줄였다고 말할 수 있다. 아울러 신경망의 매개변수(학습률과 모멘텀 상수, 은닉층 유니트수) 및 유전자 알고리즘의 다양한 교배방식과 재생산 방식을 적용한 결합모형의 연구가 필요하다.

참고문헌

- 노형진, 2000, Excel2000에 의한 통계적 조사방법, 형설출판사, 189~205/303~315.
- 문병석, 1998, 신경망 이론을 이용한 단기 급수량 예측, 전남대학교대학원 박사학위논문.
- 이경훈 · 문병석 · 강일환, 1998, ANFIS를 이용한 상수도 1일 급수량 예측에 관한 연구, 한국수자원학회지, 821-832.
- 이경훈 · 이삼노 · 문병석, 1994, 상수도 단기 급수량의 시간적 변화의 특성에 관한 연구, 한국수문학회지, 135-143.
- 원태연 · 정성원, 1998, 한글 SPSS 통계조사분석, SPSS 고려정보산업(주), 219-245/322-326.
- 손정아, 2001, 유전자 알고리즘을 이용한 신경망 모형선택, 대구카톨릭대학교대학원 석사학위

논문.

Li · Jun, 1997, Oil tanker markets modeling analysis and forecasting using Neural Networks & Fuzzy Logic and Genetic Algorithms, The University of Michigan. Pearson · David, 2003, Artificial Neural Nets and Genetic Algorithms, Springer-Verlay New York inc.

Chin-Teng Lin · C. S. George Lee, 1996, Neural Fuzzy Systems, Prentice hall, 186-189/707-708.

J.-S. R. jang · C.-T. Sun · E. Mizutani, 1997, Neuro-Fuzzy and soft computing, Prentice hall, 335-433.

최종원고채택 05. 07. 29

