

정수장 운영효율 향상을 위한 ELM 기반 단기 물 수요 예측

(ELM based Short-term Water Demand Prediction for Effective Operation of Water Treatment Plant)

최기선* · 이동훈 · 김성환 · 이경우 · 전명근

(Gee-Seon Choi · Dong-Hoon Lee · Sung-Hwan Kim · Kyung-Woo Lee · Myung-Geun Chun)

요 약

본 논문에서는 단기 물 수요 예측에 대한 모델구현을 위해 MLP의 과도학습 문제를 해결할 수 있고 빠른 학습이 가능한 ELM 기반 단기 물 수요 예측 알고리즘을 제안한다. 제시된 알고리즘의 검증을 위해 2007년도와 2008년도 충남지역 광역상수도인 A정수장에서 취득된 데이터를 분석하여 알고리즘 구현의 정확도 분석에 사용하였다. 실험 결과 MLP모델은 MAPE가 5.82[%]인 반면, 제안된 방법인 ELM 기반 모델은 5.61[%]로 성능이 향상된 것으로 나타났다. 또한, MLP모델은 학습에 소요된 시간이 7.57초인 반면, ELM 기반 모델은 0.09초로 빠른 학습이 가능함을 알 수 있었다. 따라서 제안된 ELM 기반 알고리즘은 정수장의 효율적 운영을 위한 단기 물 수요 예측에 활용할 수 있음을 보였다.

Abstract

In this paper, we develop an ELM(Extreme Learning Machine) based short-term water demand prediction algorithm which solves overfitting problem of MLP(Multi Layer Perceptron) and has quick training time. To show effectiveness of proposed method, we analyzed time series data collected in A water treatment plant at Chung-Nam province during 2007~2008 years and used the selected data for the verification of developed algorithm. According to the experimental results, MLP model showed 5.82[%], but the proposed ELM based model showed 5.61[%] with respect to MAPE, respectively. Also, MLP model needed 7.57s training time, but ELM based model was 0.09s. Therefore, the proposed ELM based short-term water demand prediction model can be used to operate the water treatment plant effectively.

Key Words : ELM, Water Demand Prediction, Water Treatment Plant

1. 서 론

산업사회의 발달과 더불어 물 수요량이 급격하게 증가함에 따라 수돗물의 안정적인 공급과 효율적인 이용이 보다 절실하게 요구되고 있다. 이와 같은 요구에 부응하기 위해서는 체계적이고 과학적인 연구

* 주저자 : 한국수자원공사 교육원 교수
Tel : 042-870-7266, Fax : 042-870-7549
E-mail : kschoi@kwater.or.kr
접수일자 : 2009년 7월 10일
1차심사 : 2009년 7월 13일
심사완료 : 2009년 7월 21일

로 수도물 수요를 정확하게 예측하여야 한다[1].

정수장 및 급배수 시설의 급수량 예측 방법은 시점에 따라 단기예측과 장기예측으로 분류할 수 있다. 단기예측은 예측기간을 시간단위나 일단위로 구분하여 급수량을 추정하며, 예측된 급수량은 급·배수량을 실시간으로 제어하는 데 사용한다. 반면에, 장기예측은 예측구간을 월별, 계절별 또는 연별로 구분하여 급수량을 예측하며, 이는 상수도시설을 계획하거나 설계시의 참고 자료로 활용하고 있다[2].

국내 물 수요 예측에 관련된 연구는 지속적인 물 수요량의 증가에 대응하기 위하여 단기 물 수요량 예측보다는 장기 물 수요량 예측에 대한 연구가 주로 수행되어 왔다[3]. 장기 수요예측에 비하여 단기 수요예측(Short-term Prediction)은 해당 정수장의 공급지역 형태, 급수인구 및 시간대별 특성 등의 개별적인 상황과 계절적인 다양한 패턴이 혼합되어 있으며, 이러한 복잡한 외부적 요인으로 인해 단순한 수요예측 모델로는 물 수요량을 정확히 예측하는 데는 무리가 있다[4-5].

실제 광역상수도의 대규모 정수장과 급배수 시설을 효율적으로 운영하기 위해서 일단위 수요예측만으로는 실시간으로 변화하는 기상 등의 외부 요인에 대응하지 못하는 문제가 있다. 따라서 본 논문에서는 시간단위 기상정보와 공급유량 자료를 바탕으로 정수장 및 급배수 시설의 효율적인 실시간 운영에

활용할 수 있도록 ELM 기반의 단기 물 수요 예측 알고리즘을 제안하여 실제 운영중인 정수장에 활용하고자 한다.

본 연구에서는 ELM 기반의 단기 물 수요 예측 모델을 제안하고, 이 모델을 통해 MLP 알고리즘의 과도학습 문제를 해결하고 빠른 학습이 가능함을 실험을 통해 나타냄으로서, 정수장의 운영효율 향상을 위한 단기 물 수요 예측에 활용할 수 있음을 보였다.

2. 수처리 시스템 개요

정수처리의 목적은 정수시설의 단위공정별로 유량과 수질의 양론적 물질수지를 기초로 하여 원수 중에 포함된 현탁물질(제거대상물질)을 제거하여 양질의 수도물을 충족하게 생산·공급하는데 있다. 수처리 시스템은 그림 1에서 보는 바와 같이 취수펌프장, 착수정, 약품주입시설, 혼화시설, 응집시설, 침전 시설, 여과시설, 소독(살균)시설, 정수지, 가압(송수) 시설 등으로 구성되어 있다.

본 논문에서는 수처리 시스템 중에서 수용가의 요구에 맞는 수요량을 효율적으로 공급하기 위한 물 수요 예측에 중점을 두고 있다. 배수지로 물을 공급하기 위한 펌프 제어방식은 지령을 운전원이 수동으로 지령하는 방식(수동제어)과 계측기 등으로부터 신호에 의하여 자동적으로 지령이 주어지는 방식(자

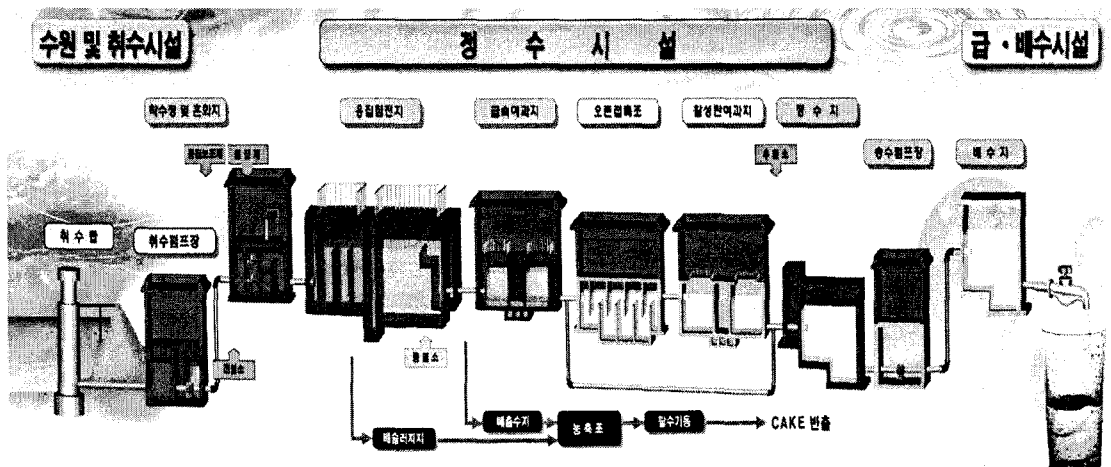


그림 1. 수처리 시스템
Fig. 1. Water treatment system

정수장 운영효율 향상을 위한 ELM 기반 단기 물 수요 예측

동제어) 등이 있다. 이들 중에서 어떤 방식으로 할 것인가는 그 시설의 규모, 신뢰성, 경제성 및 운전관리방법 등을 종합적으로 검토하여 결정한다.

유량을 제어하는 방법으로는 크게 펌프의 운전대수를 제어하는 방법과 회전속도를 제어하는 방법, 그리고 밸브의 개도를 제어하는 방법 등이 있다. 또한 회전속도 제어와 밸브 개도를 제어하는 방법에 의하여 펌프의 토출압력을 제어할 수 있으며 설치하는 기기와 운전효율에 개별적인 특성이 있으므로 제어방식은 이들의 특징을 살려서 결정한다.

위의 방법 중에서 가장 보편적으로 사용되는 유량 제어 방법은 그림 2에 나타낸 바와 같이 운전대수를 변경함으로써 유량을 제어하는 방식이다. 제어방법이 간단하고 대수 분할로 위험을 분산할 수 있지만, 제어량이 단계적으로 된다는 단점이 있다. 흡입 또는 토출수면의 수위와 연동하여 운전대수를 제어하는 자동운전인 경우에는 수위의 변동이 심하면 기동·정지가 몇 번씩 반복되는 헌팅(hunting) 현상이 발생하는 경우도 있고, 이러한 경우에는 전동기가 과열되거나 밸브와 기동기의 소모가 심하여 기기의 수명을 단축시킬 우려가 있으므로 면적과 펌프 1대당 용량과의 비가 지나치게 작지 않도록 주의하여야 한다. 운전대수의 자동제어는 각 펌프의 운전시간을 평균화하기 위하여 예비기를 포함한 계통의 모든 펌프를 대상으로 순차적으로 운전할 수 있도록 한다. 정전 후 전원이 복구되었을 때에는 1대씩 순차적으로 기동시키기 위하여 동시기동 방지회로를 설치해야 하며, 운전중인 기기의 고장으로 인하여 펌프가 정지되었을 경우에는 그 다음 순서의 펌프를 자동으로 기동시키는 회로를 설치해야 한다. 또한 운전지령계가 고장일 경우에 예비기를 포함한 모든 펌프가 운전됨으로써 계약전력을 초과하는 일이 생기지 않도록 인터록(Interlock)회로를 설치해야 한다[6].

본 논문에서 제안한 물 수요 예측 모델을 통해 정수장 공급 물량을 추정하여 취수장, 정수장, 배수지의 연계 운영을 통해 원활한 물 공급이 가능하게 할 수 있으며, 최적화된 운영기법으로 펌프제어를 체계적으로 자동화함으로써 효율적인 정수장 운영이 가능하게 할 수 있다.

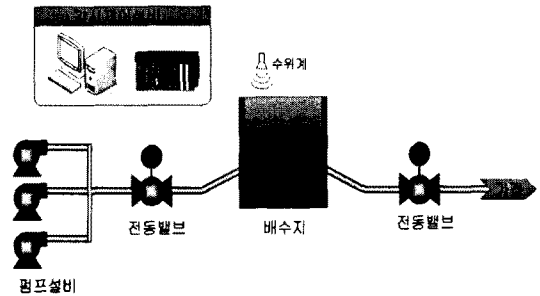


그림 2. 펌프운전 제어
Fig. 2. Pump operation control

3. ELM 기반 물 수요 예측 알고리즘

본 논문에서는 ELM 기반의 학습알고리즘인 ELM, I-ELM, EI-ELM을 통하여 물 수요예측을 구현하는 방법을 제안한다. 이러한 알고리즘에 대하여 설명하면 다음과 같다.

3.1 ELM 알고리즘

ELM(Extreme Learning Machine)은 그림 3에서 보는 바와 같이 단일 은닉층을 가지는 전방향 신경망으로 최근 Huang[7-8]등에 의하여 제안되었다. 최근 경사법(gradient descent method)에 기반한 신경 회로망의 학습은 일반적으로 좋은 성능을 얻기 위해 많은 반복 학습을 필요로 하며, 적절하지 못한 학습에 의하여 낮은 적응성을 보이거나 종종 국부 최소해(local minimum)에 수렴하는 단점을 가지고 있다. ELM에서 출력 가중치들은 반복적인 학습 구조 대신에 모어-페로스의 일반화된 역행렬(Moore-Penrose generalized

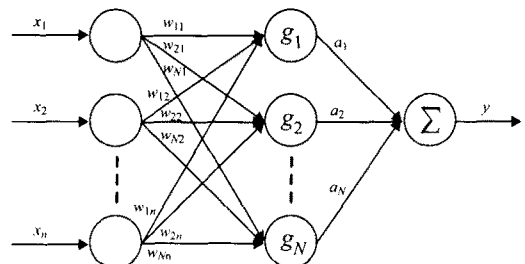


그림 3. ELM의 구조
Fig. 3. Structure of ELM

inverse)에 의해 분석적으로 계산되어진다.

ELM의 주된 특징으로 반복학습을 하지 않기 때문에 전형적인 반복 학습 방법들에 비하여 매우 빠른 학습이 가능하다는 점이다. ELM 알고리즘의 학습과정을 설명하면 다음과 같다.

m 개의 데이터 $\{(x_i, y_i), i=1, 2, \dots, m\}$ 와 한개의 출력노드를 가지고 N 개의 은닉노드들을 가지는 전방향 신경 회로망의 학습을 고려해보자. 여기서, 입력 $x_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T \in R^n$ 이고, $y_i \in R$ 이며, N 개의 은닉노드들의 활성화 함수 $g(x)$ 에 의한 전방향 신경 회로망의 출력은 다음과 같이 나타낼 수 있다.

$$g(w_j \cdot x_i) \cdot a = y_i, \quad i=1, 2, \dots, m \quad (1)$$

여기서, $w_j = [w_{j1}, w_{j2}, \dots, w_{jn}]^T, j=1, 2, \dots, N$ 은 j 번째 은닉노드와 입력노드들을 연결하는 가중치 벡터로 임의의 값으로 초기화 되며, $a = [a_1, a_2, \dots, a_N]^T$ 는 은닉노드들과 출력노드를 연결하는 가중치 벡터이다. $w_j \cdot x_i$ 는 w_j 와 x_i 의 내적을 의미한다.

식 (1)은 다음과 같이 간결하게 표현될 수 있다.

$$Ga = y \quad (2)$$

여기서

$$G(w_1, w_2, \dots, w_N, x_1, x_2, \dots, x_m) = \begin{bmatrix} g(w_1 \cdot x_1) \cdots g(w_N \cdot x_1) \\ \vdots \\ g(w_1 \cdot x_m) \cdots g(w_N \cdot x_m) \end{bmatrix} \quad (3)$$

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}_{m \times 1} \quad (4)$$

G 는 전방향 신경 회로망에서 은닉층의 출력으로 G 의 i 번째 열은 입력 x_1, x_2, \dots, x_m 과 관련된 i 번째 은닉노드들의 출력 벡터이다. 최종적으로, 다음과 같이 최소자승법에 의하여 출력 가중치 \hat{a} 을 구한다.

$$\hat{a} = G^+y \quad (5)$$

여기서 G^+ 는 은닉층의 출력 G 의 모어-페로스의 일반화된 역행렬이다. 만일 $G^T G$ 가 비정칙인 경우,

G 의 모어-페로스의 일반화된 역행렬은 다음과 같이 얻어진다.

$$G^+ = (G^T G)^{-1} G^T \quad (6)$$

최종 예측은 다음과 같이 얻어진다.

$$\hat{y} = G\hat{a} \quad (7)$$

3.2 I-ELM(Incremental-ELM)의 구조와 학습

위와 같은 ELM에 기반을 둔 방법들은 신경망의 구축시간이 빠르고, 일반적인 성능이 우수한 장점을 가지고 있지만, 사전에 사용자에게 의하여 신경망 구축에 필요한 은닉노드 개수를 선정하는 문제와 복잡한 비선형 회귀문제들에서 과잉 학습의 영향을 해결해야 하는 문제점들을 여전히 가지고 있다[9].

이러한 문제점을 개선하기 위하여 Huang등은 단일 은닉층을 가지는 전방향 신경망의 은닉노드 개수를 단계적으로 1개씩 추가하는 I-ELM(Incremental-ELM)을 제안하였다[10]. I-ELM은 신경망을 구축하기위해 사전에 은닉노드 개수를 결정해야 하는 어려움을 극복하고, 구축된 신경망이 범용 근사화(universal approximation)가 가능함을 이론과 실험을 통하여 증명되었다[10].

n 개의 은닉노드들에 의한 단일 은닉층을 가지는 신경망의 출력 함수는 다음과 같이 표현되어질 수 있다.

$$f_n(x) = \sum_{i=1}^n \beta_i g_i(x) = \sum_{i=1}^n \beta_i G(x, a_i, b_i) \quad (8)$$

$$a_i \in C^d, x \in C^d, b_i \in C, \beta_i \in C$$

여기서 g_i 또는 $G(x, a_i, b_i)$ 는 i 번째 은닉노드의 출력함수이고, β_i 는 i 번째 은닉노드와 출력 노드사이의 출력 가중치이다.

신경망의 출력 함수 f_n 과 목적 함수 f 의 유사도는 다음과 같이 구해 질 수 있다.

$$\|f_n - f\| = \left[\int_X (f_n(x) - f(x))(f_n(x) - f(x)) dx \right]^{1/2} \quad (9)$$

평수장 운영효율 향상을 위한 ELM 기반 단기 물 수요 예측

n 개의 은닉노드들에 의한 신경망의 잔류오차(residual error)는 $e_n = f - f_n$ 이 된다. 단계적으로 랜덤하게 추가되어지는 n 번째 은닉노드에서의 출력가중치 β_n 는 다음과 같이 계산되어진다.

$$\beta_n = \frac{\langle e_{n-1}, g_n \rangle}{\|g_n\|^2} \quad (10)$$

따라서 잔류 오차 e_n 은 $e_n = f - f_n = e_{n-1} - \beta_n g_n$ 이 되며, e_n 과 g_n 의 내적은 다음과 같다.

$$\begin{aligned} \langle e_n, g_n \rangle &= \langle e_{n-1} - \beta_n g_n, g_n \rangle \\ &= \langle e_{n-1}, g_n \rangle - \beta_n \|g_n\|^2 \end{aligned} \quad (11)$$

위의 식 (10)과 (11)로부터

$$\begin{aligned} \langle e_n, g_n \rangle &= \langle e_{n-1}, g_n \rangle - \frac{\langle e_{n-1}, g_n \rangle}{\|g_n\|^2} \cdot \|g_n\|^2 \\ &= \langle e_{n-1}, g_n \rangle - \langle e_{n-1}, g_n \rangle = 0 \end{aligned} \quad (12)$$

따라서 $\langle e_n, e_n - e_{n-1} \rangle = \langle e_n, -\beta_n g_n \rangle = 0$ 이 되며 이것은 $e_n \perp (e_n - e_{n-1})$ 를 의미한다.

I-ELM은 단순히 은닉노드들을 랜덤하게 선택하고, 단지 추가되어지는 은닉노드들의 가중치를 식 (12)와 같이 계산하는 것에 의하여 예측 잔류오차 e_n 은 0으로 수렴하게 된다. Huang등은 이러한 증분 알고리즘에 의하여 구축된 I-ELM이 범용 근사화가 가능함을 이론과 실험을 통하여 증명하였다.

3.3 EI-ELM(Enhanced I-ELM)의 구조와 학습

신경망에서 시스템에 적합한 은닉층의 개수와 은닉노드 개수를 결정하는 것은 신경망의 성능에 있어서 중요한 이슈로 자리 잡고 있다. 신경망에서 너무 적은 개수의 은닉노드들은 예측하고자 모델을 정확하게 표현하기가 어려운 반면, 너무 많은 은닉노드 개수의 확장은 학습 데이터에 대한 과잉학습의 우려와 연산시간의 효율 면에서 좋지 않은 결과를 얻을 수 있다.

즉, I-ELM은 원하는 수준의 성능을 얻기까지 많

은 개수의 은닉노드들과 오랜 구축 시간이 필요하다. 단점을 가지고 있다. 특히 적절하지 못한 은닉노드의 추가는 신경망의 출력에서 매우 적은 영향을 미치게 되고 결과적으로 신경망의 복잡성을 증가시킬 수 있다. 이러한 문제점을 개선하기 위하여 다수의 후보 노드들을 생성한 후 최적의 노드를 선택하는 EI-ELM(Enhanced I-ELM)이 최근 Huang와 Chen에 의하여 제안되었다[11].

EI-ELM을 각 단계별로 설명하면 다음과 같다.

훈련데이터 $\mathbf{x} = \{(x_i, t_i) | x_i \in \mathbb{C}^d, t_i \in \mathbb{C}, i = 1, \dots, N\}$ 일 때, 최대 은닉노드 개수 L_{\max} 와 후보 노드의 개수 k , 허용 오차 ϵ 를 설정한다.

[단계 1] 초기화: 은닉노드의 개수 $L=0$, 잔류오차 $E=t$ 로 초기화 한다. 여기서 $t = [t_1, \dots, t_N]^T$ 이다.

[단계 2] 학습 단계:

while $L < L_{\max}$ & $\|E\| > \epsilon$

(a) 은닉노드의 개수 L 를 1개 증가 시킨다

: $L = L + 1$.

(b) for $i = 1 : k$

(i) i 번째 후보 노드에 대하여 임의의 랜덤 파라미터 $(a_{(i)}, b_{(i)})$ 를 할당한다.

(ii) 출력 가중치 $\beta_{(i)}$ 를 다음과 같이 계산한다.

$$\beta_{(i)} = \frac{E \cdot H_{(i)}^T}{H_{(i)} \cdot H_{(i)}^T} \quad (13)$$

(iii) 새로운 후보 노드 i 를 추가한 후의 잔류오차를 계산한다.

$$E_{(i)} = E - \beta_{(i)} \cdot H_{(i)} \quad (14)$$

endfor

(c) k 개의 후보 노드들에 대하여 $i^* = \{i | \min_{1 \leq i \leq k} \|E_{(i)}\|\}$ 를 만족하는 i 를 선택하여 최종적으로 은닉노드에 추가한다. $E = E_{(i^*)}$, $a_L = a_{(i^*)}$, $b_L = b_{(i^*)}$,

$$\beta_L = \beta_{(i^*)}.$$

endwhile

4. 실험 및 결과 분석

4.1 물 수요 데이터의 특성분석

본 논문에서는 그림 4에서 보는 바와 같이 충청남도의 광역상수도 중에서 A정수장의 B시 공급유량에 대한 단기 물 수요 예측 모델 구현을 위해 물 수요 데이터 특성을 분석하였다. 연구 대상지역인 B시는 충청남도 북부지역에 위치하고 있으며, 편리한 교통여건과 최근 수도권 확대와 반도체 산업 등의 신규 공장 건설 등으로 급격한 성장 발전을 이루면서 물 수요량이 꾸준히 증가되고 있는 지역이며 급수인구는 약 56만명이다.

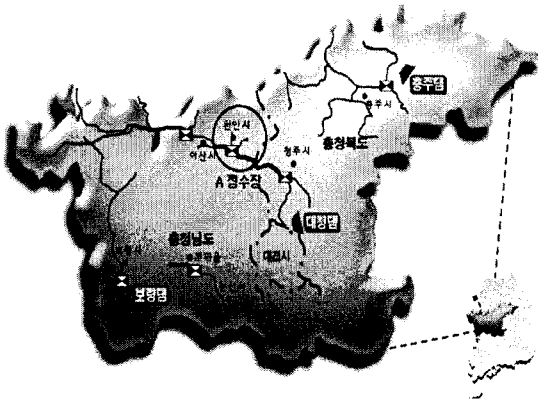


그림 4. 단기 물 수요예측 대상지역
Fig. 4. Site for short-term water demand prediction

본 논문에서 사용된 데이터는 A정수장의 2007~2008년 까지의 2년간 시간단위 공급유량과 기상데이터를 활용하였다. 대상사업장인 A정수장은 일 41.4천톤의 생산능력을 갖춘 대규모 정수장이다.

B시 물 수요 예측 모델을 정수장 운영에 실제 적용하기 위해 시간 및 기상데이터와 유량데이터의 상관성 분석을 통해 입력변수를 선정하였으며 상관 정도 분석결과를 표 1에 나타내었다.

상관성 분석은 2007년도, 2008년도 및 2007~2008년도 데이터로 각각 구분하여 분석을 하였다. 상관성 분석결과를 보면, 우선 2007년도 데이터는 시간특성에서는 시, 기상요인에서는 기온, 습도 및 풍속,

과거데이터에서는 3시간 전까지 데이터가 물 수요량에 밀접한 관계를 갖는 것으로 나타났다. 2008년도 데이터는 시간특성에서는 시, 기상요인에서는 기온, 습도 및 기압, 과거데이터에서는 3시간 전까지 데이터가 물 수요량에 밀접한 관계를 갖는 것으로 나타났다. 2007~2008년도 데이터를 모두 적용하여 분석한 결과 시간특성에서는 시, 기상요인에서는 기온, 습도, 과거데이터에서는 3시간 전까지 데이터가 물 수요량과 상관성이 상대적으로 높은 것으로 나타났다. 이를 종합하면, 일반적으로 물 수요량이 시간특성에서는 시간과의 상관성이 높은 것으로 나타났으며, 기상 특성에서는 기온 및 습도인자가 다른 외부인자에 비하여 비교적 높은 상관성을 가진 것으로 나타났다. 그러나, 풍향, 풍속, 기압, 강수량, 일타입(평일/휴일 구분), 계절타입 등은 예상과는 달리 낮은 상관성을 보여주고 있음을 알 수 있다. 또한 유량 데이터 중에서 3시간 전까지의 과거값이 현재 유량과의 높은 상관성을 보여주었다. 따라서 본 논문에서는 시간특성에서는 시, 기상요인에서는 기온과 습도, 과거데이터에서는 3시간 전까지의 데이터를 물 수요 예측에 적용하였다.

표 1. 연도별 상관계수 분석
Table 1. Correlation analysis

구분	2007	2008	2007~2008
월	-0.027	0.157	0.064
일	0.015	-0.052	-0.018
시	0.372	0.344	0.356
일타입	-0.038	-0.066	-0.052
계절타입	0.054	0.240	0.146
기온	0.250	0.491	0.369
습도	-0.416	-0.248	-0.328
풍향	0.102	0.004	0.050
풍속	0.238	0.189	0.219
기압	-0.113	-0.296	-0.200
강수량	0.000	0.046	0.010
1시간전	0.870	0.868	0.870
2시간전	0.691	0.688	0.692
3시간전	0.523	0.526	0.529
4시간전	0.367	0.383	0.381
5시간전	0.226	0.260	0.250
6시간전	0.109	0.157	0.141

4.2 물 수요 예측 결과

본 논문에서는 충남지역의 A정수장에서 취득된 2007~2008년의 물 공급유량 자료를 이용하여 실험을 하였다. 그림 5에서는 2007년도와 2008년도에 대한 시간단위 물 수요량을 나타내었다. 그림 5에서 보이는 것과 같이 2007년과 2008년의 물 공급량의 상관계수는 0.358(Pearson 상관 계수 기준)로, 연도별 물 공급량만을 고려할 경우 물 수요 예측이 매우 어려움을 짐작할 수 있고, 시간별 공급유량은 일일을 기준으로 유사한 패턴을 나타냄을 알 수 있다.

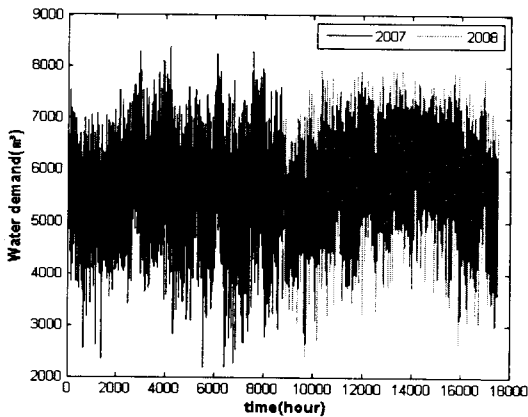


그림 5. A정수장에서 2007년도와 2008년도 물 수요
Fig. 5. Water demand for 2007 and 2008 in A water treatment plant

제안된 방법의 성능평가를 위해 학습데이터로 2007년도 데이터를 이용하였으며, 2008년도 데이터를 검증 데이터로 사용하였다. 성능지표로는 아래식의 절대평균퍼센트오차인 MAPE(Mean Absolute Percentage Error)와 절대평균오차인 MAE(Mean Absolute Error)를 고려하였다.

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{Actual_i - Forecast_i}{Actual_i} \right| \times 100 \quad (15)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |Actual_i - Forecast_i| \quad (16)$$

본 논문에서는 시간특성에서는 시, 기상요인에서

는 기온과 습도, 과거데이터에서는 3시간 전까지의 데이터를 물 수요 예측에 적용하였으며, 표 2에서는 적용방법에 따른 성능평가를 나타냈다.

MLP에서는 은닉노드의 개수를 6개에서 40개까지 2개씩 증가시키면서 실험을 하였으며, 각 실험마다 500번의 학습과 학습률(Learning rate)은 0.001, 모멘텀 항(Momentum term)은 0.2로 설정하였다. ELM에서는 은닉노드의 개수를 6개에서 40개까지 2개씩 증가시키면서 실험을 하였으며, I-ELM에서는 은닉노드의 개수를 1개씩 추가하면서 총 2000개까지 추가를 하였다. EI-ELM에서는 k가 10인 경우와 20인 경우로 나누어서 실험을 하였고, 일반적인 성능을 알기위하여 각각 5번씩 실험을 수행하여 평균과 표준편차를 비교 분석하였다. 모든 실험은 MATLAB으로 프로그램 하였고, 인텔 2.4[GHz] CPU, 2[G] RAM의 환경에서 실험하였다.

표 2. 실험 결과

Table 2. Experimental result

Algorithm	MAPE [%]		MAE	
	Training	Testing	Training	testing
MLP (10)	6.055 ±0.103	5.820 ±0.144	319.731 ±4.895	323.766 ±13.591
ELM (34)	5.713 ±0.022	5.615 ±0.030	300.116 ±1.326	309.700 ±2.213
I-ELM	6.586 ±0.110	6.603 ±0.141	348.398 ±6.302	367.017 ±8.817
EI-ELM (k=10)	6.047 ±0.175	6.011 ±0.241	320.608 ±10.708	334.537 ±15.073
EI-ELM (k=20)	6.020 ±0.209	5.870 ±0.208	317.421 ±10.658	324.746 ±11.364

표 2에서는 각 알고리즘별로 학습데이터에 대하여 성능이 가장 우수한 경우를 상호 비교하였다. MLP와 ELM에서 괄호부분은 은닉노드의 개수를 나타낸다. 알고리즘별로 실험결과를 정리한 표 2에서 알 수 있는 바와 같이 학습데이터에 대한 성능평가를 분석하면 MLP인 경우 은닉노드의 개수가 10일때 학습데이터는 6.05[%]이며, 검증데이터는 5.82[%]로 나타났다. ELM의 경우에는 은닉노드의 개수가 34일 때 학

습데이터에 대해서는 5.71[%], 검증데이터에 대해서는 5.61[%]로 나타났다. EI-ELM(k=20)일 때 학습데이터와 검증데이터 모두 MLP와 ELM에 비하여 인식률은 낮았으나 I-ELM보다는 우수하게 나타났다.

그림 6에서는 은닉노드의 개수에 따른 MLP와 ELM의 성능을 상호 비교하여 나타냈다. 학습데이터에 대하여 MLP에서는 10개의 은닉노드에서 ELM은 34개의 은닉노드의 개수에서 가장 좋은 성능을 나타냈다. 그림 6에서 보는 바와 같이 ELM를 이용하는 경우 은닉 노드 개수의 변화에 일관적인 성능을 얻을 수 있는 반면에 MLP를 이용한 경우 은닉노드의 개수 변화에 따라 성능의 변화가 큰 단점을 가짐을 알 수 있다. 이를 확인하기 위해 그림 7에서는 은닉 노드의 개수에 따른 표준편차의 변화를 나타냈다. 그림 7에서 보는 바와 같이 MLP의 경우에는 은닉 노드의 변화에 따라 편차가 급격히 변화는 반면에 ELM의 경우는 은닉노드의 증가에 따라 편차가 크게 나타나지를 않았다. 이는 제안된 방법을 적용할 경우 MLP의 문제로 대두되는 과도학습 문제를 효과적으로 해결할 수 있는 것으로 판단된다.

그림 8에서는 은닉노드의 개수에 따른 소요된 학습시간을 상호 비교하여 나타내었다. 그림 8에서 보는 바와 같이 은닉 노드가 증가될수록 MLP의 학습시간은 기하급수적으로 증가한 반면에 ELM의 학습시간은 은닉 노드의 수에 상관없이 일정한 시간이 소요된다. 표 3에서는 각 알고리즘별로 학습에 소요된 시간을 상호 비교하였다. 표 3에서 보는 바와 같이 ELM을 물 수요예측에 적용한 결과 물 수요예측 성능뿐만 아니라 학습 시간도 MLP에 비하여 매우 우수함을 확인할 수 있다.

표 3. 학습시간
Table 3. Training time

단위 : (second)

Algorithm	Time(s)
MLP(10)	7.5753
ELM(34)	0.0938
I-ELM	1.8750
EI-ELM(k=10)	19.5128
EI-ELM(k=20)	39.8331

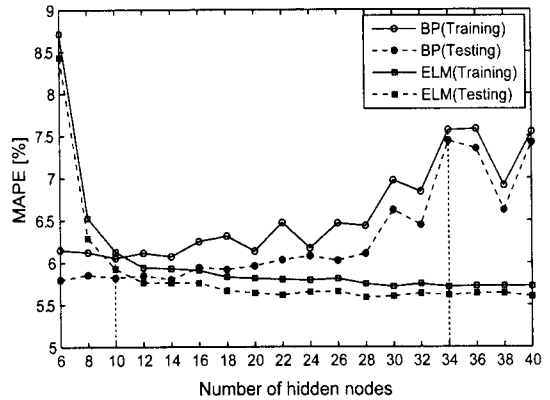


그림 6. 은닉노드의 개수에 따른 MLP와 ELM의 성능
Fig. 6. Performance of MLP and ELM according to number of hidden node

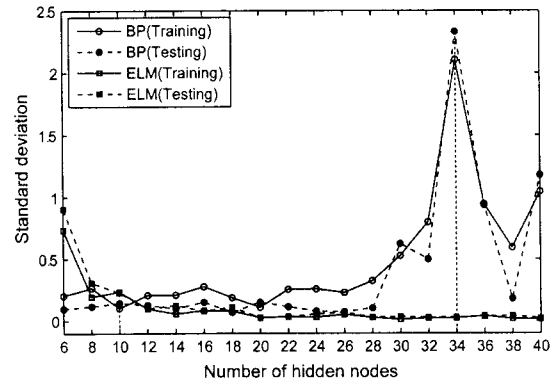


그림 7. 은닉노드의 개수에 따른 MLP와 ELM의 표준편차
Fig. 7. Standard deviation of MLP and ELM according to number of hidden node

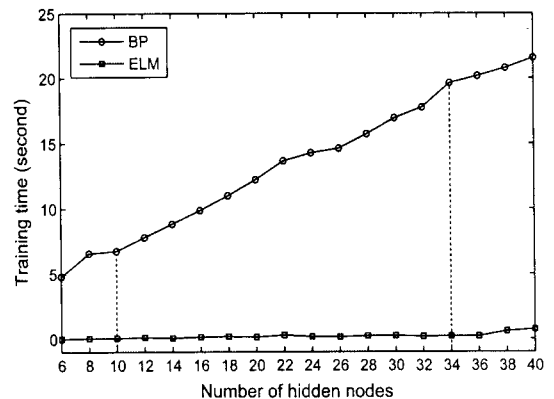


그림 8. 은닉노드의 개수에 따른 MLP와 ELM의 학습시간
Fig. 8. Training time of MLP and ELM according to number of hidden node

5. 결 론

본 연구는 광역상수도의 시간단위 수요량 예측을 위하여 충남지역 A정수장을 대상으로 물 공급량 데이터를 분석하여 단기 물 수요 예측모델을 구현하였다. 물 수요량의 상관성 분석결과 시간특성에서는 시, 기상요인에서는 기온과 습도, 과거데이터에서는 3시간 전까지의 데이터를 물 수요 예측에 적용하였다. 본 논문에서는 단기 물 수요 예측에 대한 모델구현을 위해 MLP의 과도학습문제를 해결할 수 있고 빠른 학습이 가능한 ELM 기반 단기 물 수요 예측 알고리즘을 제안하였으며, 제시된 알고리즘의 검증 을 위해 2007년도와 2008년도 충남지역 광역상수도 인 A정수장에서 취득된 데이터를 분석하여 알고리즘 구현의 정확도 분석에 사용하였다. 2007년 자료 로 모델링하고 2008년 자료로 검증하여 예측 모형의 성능을 평가한 결과 MLP모델은 MAPE가 5.82[%] 인 반면, 제안된 방법인 ELM 기반 모델은 5.61[%] 로 성능이 향상된 것으로 나타났다. 또한, MLP모델 은 학습에 소요된 시간이 7.57초인 반면 ELM 기반 모델은 0.09초로 빠른 학습이 가능함을 알 수 있었다. 따라서 제안된 ELM 기반 알고리즘은 외부 변동 요인에 실시간으로 대응할 수 있는 장점이 있어 정 수장의 효율적 운영을 위한 단기 물 수요 예측에 활 용할 수 있을 것으로 보인다.

References

- [1] Kyung-Hoon Rhee, IL-Hwan Kang, Byoung-Seok Moon, "Application of ANFS for Prediction of Daily Water Supply" Journal of the Korean Society of Water and Wastewater, Vol. 14, No. 3, pp. 281-290, 2000.
- [2] Kyung-Hoon Rhee, IL-Hwan Kang, Chang-Ju Oh, "Forecasting of Urban Daily Water Demand by using Backpropagation Algorithm Neural Network" Journal of the Korean Society of Water and Wastewater, Vol. 12, No. 4, pp. 43-52, 1998.
- [3] Shin-Geol Kim, Sin-Suk Pyon, Young-Sang Kim, Ja-Yong Koo, "Forecasting the Long-term Water Demand Using System Dynamics in Seoul", Journal of the Korean Society of Water and Wastewater, Vol. 20, No. 2, pp. 187-196, 2006.
- [4] Ishmael S. Msiza, Fulufhelo V. Nelwamondo, Tshilidzi Marwala, Water Demand Prediction using Artificial Neural Networks and Support Vector Regression," JOURNAL OF COMPUTERS. Vol. 3, No. 11, pp. 1-8, NOVEMBER 2008.

- [5] Mehmet Ali Yurdusev, Mahmut Firat, "Adaptive neuro fuzzy inference system approach for municipal water consumption modeling: An application to Izmir, Turkey," Journal of Hydrology. Vol. 365, pp. 225-234, 2009.
- [6] Ministry of Environment, Standard of waterworks facilities, KWWA, 2004
- [7] G. B. Huang, Q. Y. Zhu, and C. K. Siew, "Extreme learning machine: a new learning scheme of feedforward neural networks," in Proc. 2004 IEEE Int. Conf. Neural Networks, Vol. 2, pp. 985-990, 2004.
- [8] G. B. Huang, Q. Y. Zhu, and C. K. Siew, "Extreme learning machine: theory and applications," Neurocomputing, Vol. 70, No. 1-3, pp. 489-501, 2006.
- [9] Jae-Hoon Cho, Dae-Jong Lee, Myung-Geun Chun, "Parameter Optimization of Extreme Learning Machine Using Bacterial Foraging Algorithm," Journal of Korean Institute Intelligence System, Vol. 17, pp. 807-812, 2007.
- [10] G. B. Huang, L. Chen, and C. Siew, "Universal Approximation Using Incremental Constructive Feedforward Networks With Random Hidden Nodes," IEEE Transactions on Neural Networks, Vol. 17, pp. 879-892, 2006.
- [11] G. B. Huang and L. Chen, "Enhanced random search based incremental extreme learning machine," Neurocomputing, Vol. 71, pp. 3460-3468, 2008.

◆ 저자소개 ◆

최기선 (崔琪銑)

1965년 10월 8일생. 1992년 2월 충남대학교 졸업. 2002년 2월 한밭대학교 제어계측공학과 졸업(석사). 2006년~현재 충북대학교 제어계측공학과 박사과정. 1992년~현재 한국수자원공사 교육원 교수.

이동훈 (李東勳)

1976년 5월 5일생. 2001년 2월 경북대학교 전자전기공학과 졸업. 2003년 2월 포항공과대학교 전기전자공학과 졸업(석사). 2004년~현재 한국수자원공사 재직.

김성환 (金性煥)

1971년 7월 20일생. 1996년 2월 동신대학교 전자공학과 졸업. 2007년 8월 충북대학교 산업대학원 졸업(석사). 1996년~현재 한국수자원공사 재직.

이경우 (李慶雨)

1977년 4월 14일생. 2003년 2월 동아대 전자공학과 졸업. 2008년 3월~현재 충남대 전자전파정보통신공학과 석사과정. 2003년 2월~현재 한국수자원공사 재직.

전명근 (全命根)

1964년 7월 17일생. 1987년 부산대학교 전자공학과 졸업. 1989년 한국과학기술원 전기 및 전자공학과 졸업(석사). 1993년 한국과학기술원 전기 및 전자공학과 졸업(박사). 1993~1996년 삼성전자 자동화연구소 선임연구원. 2000~2001년 University of Alberta 방문교수. 1996년~현재 충북대학교 전자정보대학 제어로봇공학과 교수.