

데이터 마이닝과 칼만필터링에 기반한 단기 물 수요예측 알고리즘

Short-term Water Demand Forecasting Algorithm Based on Kalman Filtering with Data Mining

최기선, 신강욱*, 임상희, 전명근
(Gee-Seon Choi, Gang-Wook Shin, Sang-Heui Lim, and Myung-Geun Chun)

Abstract: This paper proposes a short-term water demand forecasting algorithm based on kalman filtering with data mining for sustainable water supply and effective energy saving. The proposed algorithm utilizes a mining method of water supply data and a decision tree method with special days like Chuseok. And the parameters of MLAR (Multi Linear Auto Regression) model are estimated by Kalman filtering algorithm. Thus, we can achieve the practicality of the proposed forecasting algorithm through the good results applied to actual operation data.

Keywords: demand forecasting, Kalman filtering, data mining, water supply system

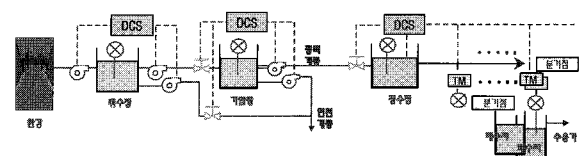
I. 서론

수요예측 분야는 경제성장, 시스템 보안, 그리고 경영 및 계획 등 많은 응용분야에 있어서 근간이 되어왔다. 특히 전력 계통의 안정적 운영을 위한 전력 수요예측에 대한 연구가 지속적으로 이루어지고 있다[1-4]. 뿐만 아니라 주택수요예측, 이동통신망 수요예측, 그리고 풍속에 대한 예측 등 다양한 분야에서 연구되고 있다[5-7]. 이와 같은 많은 응용 분야 중 물 수요예측에 있어서는 다른 분야와 달리 다음의 두 가지 중요 인자에 의한 의존성이 높다할 수 있다. 하나는 낮과 밤 동안의 인간 사회활동과 습관에 의한 것, 그리고 나머지 하나는 날씨 조건에 따른 영향을 나타낸다. 이러한 두 가지 중요 인자에 대한 조사를 통하여 다양한 물 수요예측알고리즘을 개발하려는 시도가 있어왔다. 물 수요예측을 기간별로 크게 구분하면, 장기예측과 단기예측으로 나눌 수 있다. 장래의 수도시설의 건설 혹은 확장계획 등의 연간 계획을 결정하는 일을 목적으로한 물 수요예측을 장기예측으로 한다. 한편, 수도시설의 합리적인 물 운용과 유지관리계획을 목적으로 하루 또는 시간계획의 물 수요예측을 단기 물 수요예측이라 한다. 특히 국토개발이나 도시계획에 의한 인구 증감에 따른 장기 물수요예측은 일반적으로 양호한 예측결과를 도출하였다[8]. 그러나 안정적인 물 공급과 에너지의 효율적 사용을 위한 일별 및 시간별 단기 물 수요예측에 대한 연구는 간헐적으로 진행되었지만, 실제 적용할 수 있는 수준의 결과도출이 미흡한 실정이다. 시계열 모델은 수요예측을 위해 사용된 가장 인기있는 예측모델로서 자기회귀(AR), 이동평균(MA), 자기회귀이동평균(ARMA), 그리고 자기회귀 적분 이동평균(ARIMA) 등 다양하다. 이러한 시계열 모델을 활용한 수요예측에 대한 연구

가 꾸준히 진행되어 왔으며, 특히 전력부하 예측분야에 있어서는 칼만, 퍼지, 뉴럴 모델등에 의한 양호한 연구결과가 발표되고 있다[9,10]. 물 수요예측에 있어서도 광역상수도 수운영시스템을 대상으로 시계열 예측모델, 칼만필터 모델, 그리고 뉴럴과 칼만 결합모델 등 다양한 시도가 있었고 국외에서는 강우량에 따른 주간 단위 수요예측과 도시 지역전체의 물 수요예측에 대한 연구가 진행되어 왔으나 현장 적용성이 높지 못하다[11-14]. 따라서, 본 연구에서는 단기 물 수요예측을 위한 시계열분석을 통하여 다중선형 자기회귀모델과 칼만필터링의 결합모델을 제안하였으며, 데이터의 이상치 감시 및 연관성 분석, 그리고 의사결정시스템을 구현하여 단기 물 수요예측의 최적 모델을 제시하고자 한다. 이러한 연구관점은 지금까지 단기 물 수요예측에 대한 적용성이 전무한 상황에서 특정 사업장에서의 적용성을 확인하고자 하였다. 본 논문의 II 장에서는 물 공급계통에 대한 일반사항과 데이터의 특성 분석을 실시하였으며, III 장에서는 본 연구에서 제안한 데이터 마이닝과 칼만필터의 결합모델을 제안하였고, IV 장에서는 제안된 알고리즘의 성능평가를 통하여 적용성을 검증하였다.

II. 정수공급시스템

본 연구에서 검토된 대상 사업장은 일일 공급 가능한 급수량이 약 100 만톤이며 급수 인구는 약 200 만명인 정수



(DCS: Distributed control system)

그림 1. 계통도.

Fig. 1. Configuration of Water Treatment Process or Water Supply.

* 책임저자(Corresponding Author)

논문접수: 2009. 6. 24., 수정: 2009. 7. 20., 채택확정: 2009. 8. 7.

최기선, 신강욱, 임상희: 한국수자원공사

(kschoi@kwater.or.kr/gwshin@kwater.or.kr/sang@kwater.or.kr)

전명근: 충북대학교(mgchun@chungbuk.ac.kr)

장이다. 대상 정수장은 그림 1에서와 같이 취수장으로부터 유입된 유량을 정수장에서 정수처리 후 배수지 혹은 수용가로 공급하도록 구성되어 있다. 각 배수지 및 수용가에서의 물 사용량의 합에 대한 신뢰성 있는 데이터 확보의 어려움으로 인하여 정수장 유출 단계에서의 통합 유량을 통한 단기 물 수요예측알고리즘을 도출하고자 한다.

1. 물 공급량 분석

A 정수장에서의 연간 물공급량을 분석하기 위하여 2006년부터 2008년까지 3년 동안의 물 공급데이터를 조사하였다. 그림 2는 일일 공급량에 대한 시계열도를 나타낸 것이며, 시계열도에서 알 수 있는 바와 같이 여름철에서의 물 공급량이 대체로 많은 것을 알 수 있다. 또한, 년도별 일평균 공급유량을 살펴보면 2006년도 44만톤, 2007년도 47만7천톤, 그리고 2008년도 약 50만톤의 물이 공급되었다.

2. 요일별 유량 특성분석

3년간의 물공급 데이터에 대하여 요일별 일원분산분석과 특이일에 대한 일원분산분석을 실시하였다. 표 1에서와 같이 특이일은 설 연휴와 추석 연휴를 말하는 것으로, 평균 물 공급량이 평일의 공급량보다 평균적으로 약 15% 적게 공급되는 것으로 분석되었다. 또한, 특이일에서의 휴일 전후에 대한 물 사용량에 있어서는 평일의 물 사용량과 별 차이가 없으나, 편차가 크게 발생함을 알 수 있다.

3. 시간대별 유량 특성분석

A 정수장의 3년간의 시간대별 공급유량에 대하여 분석하였으며, 시간대별 그래프는 그림 3과 같다. 그림에서 알

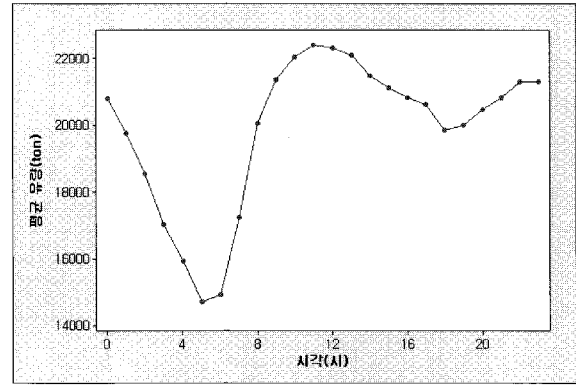


그림 3. 시간대별 일원분산 분석.
Fig. 3. Timely one-way ANOVA.

표 2. 기상데이터 상관성분석.

Table 2. Correlation analysis of weather data.

기상 변수	기온			평균 풍속	운량	
	최고	평균	최저		유량	
상관 계수	0.304	0.316	0.320	-0.087	0.019	
기상 변수	상대습도	일조시간	강수량	적설량	유량	
					1일전	2일전
상관 계수	0.124	-0.021	0.013	-0.009	0.742	0.617

수 있는바와 같이 새벽 0시부터 5시까지 물 공급량이 감소하는 것을 알 수 있다. 또한 낮 12시경 소비량이 가장 많은 것으로 물 공급량이 최대치를 나타냈다.

4. 상관성 분석

다중자기회귀 분석을 위하여 2006년과 2007년도 기상데이터와 공급 유량 데이터를 이용하여 상관관계를 분석하였다. 수집된 기상데이터는 최고기온, 평균기온, 최저기온, 평균풍속, 상대습도, 일조시간, 운량, 강수량, 적설량 등이다. 공급유량과의 상관계수를 구한 결과 표 2와 같은 결과를 얻었으며 본 연구에서는 상관관계가 높은 기온과 전일유량을 매개변수로 선정하였다.

III. 단기 물 수요예측

1. 다중선형 자기회귀 모델

일반적인 선형 회귀모델을 위한 종속변수와 독립변수 사이의 관계식은 식 (1)과 같이 나타낼 수 있다. 여기서 종속변수의 변화를 설명하기 위하여 두개 이상의 독립변수가 사용되어 종속변수와의 관계를 선형으로 가정하는 회귀모델을 다중선형 회귀모델이라고 한다.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \epsilon \quad (1)$$

여기서, $\beta_0, \beta_1, \dots, \beta_k$ 는 회귀모수로서의 상수값이며, X_1, X_2, \dots, X_k 는 각각 독립변수이며, ϵ 은 오차항이다. 위 식의 회귀모수들을 추정하기 위하여 일반적으로 최소제곱법을 이용한다. 여기서 최소제곱법에 의한 회귀모수 추정값을 얻기 위해서는 식 (2)를 최소화 시킨 값으로 구할 수 있다.

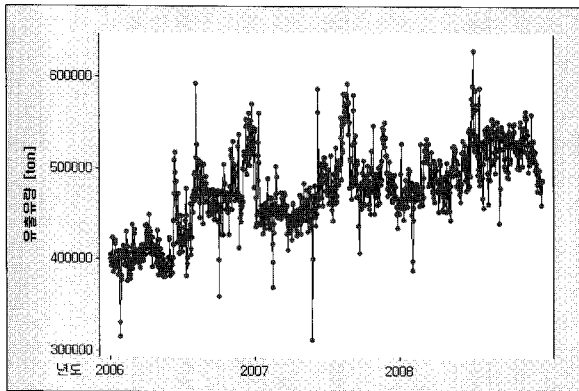


그림 2. 일 유량 시계열도.
Fig. 2. Time series of daily water supply.

표 1. 요일별 일원분산 분석.

Table 1. Daily one-sample T examine.

수준	데이터 수	평균	표준 편차
월	153	479293	43765
회수목	458	476389	47967
금	151	476699	46080
토	151	470235	42019
일	152	467451	40845
휴일전	6	474982	43351
설추석	19	416817	53819
휴일후	6	446332	48498

$$\sum_t (y - \hat{\beta}_0 - \hat{\beta}_1 X_1 - \dots - \hat{\beta}_k X_k)^2 \quad (2)$$

모수 β_0 와 β_1 의 추정값 $\hat{\beta}_0$ 와 $\hat{\beta}_1$ 은 최소제곱 기준에 의해서 구해지며, 하나의 예측변수만 있을 때 $\hat{\beta}_0$ 와 $\hat{\beta}_1$ 에 대한 추정 공식은 다음과 같이 구해진다.

$$\hat{\beta}_1 = \frac{SS_{yy}}{SS_{xx}}, \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \quad (3)$$

여기서,

$$SS_{xx} = \sum X^2 - \frac{(\sum X)^2}{n} \quad \bar{X} = \frac{\sum X}{n}$$

$$SS_{yy} = \sum XY - \frac{(\sum X) \cdot (\sum Y)}{n} \quad \bar{Y} = \frac{\sum Y}{n}$$

본 연구에서는 일일 물 공급량을 예측하기 위하여 표 2 에서와 같이 상관성 분석을 실시한 결과 평균기온은 0.316, 1일전 유량은 0.742, 2일전 유량은 0.617로서, 상대습도 0.124, 강수량 0.13 보다 상대적으로 상관성이 높은 것으로 나타났다. 따라서, 본 연구에서는 이를 주요 인자로 선정하였으며, 다음과 같이 다중선형 자기회귀모델의 일반식을 표현할 수 있다.

$$y(k) = a_0 + a_1 y(k-1) + a_2 y(k-2) + b_0 t(k-1) \quad (4)$$

여기서, $y(k)$ 는 물 공급 예측량, $y(k-1)$ 는 1일전 물공급량, $y(k-2)$ 는 2일전 물공급량, $t(k-1)$ 는 1일전 평균온도를 각각 나타낸다.

2. 칼만필터 모델

칼만필터는 불규칙한 외란을 포함한 공정의 최적 상태를 효율적으로 추정하기 위한 알고리즘으로서, 매 샘플링 시간마다 최소오차 분산의 반복적 알고리즘을 통하여 상태값을 추정한다[15].

칼만필터의 이산상태 방정식은 다음 식과 같이 표현할 수 있다.

$$x(k+1) = A(k)x(k) + w(k) \quad (5)$$

$$z(k) = C(k)x(k) + v(k) \quad (6)$$

여기서, $x(k)$ 는 $n \times 1$ 시스템 상태벡터, $A(k)$ 는 $n \times n$ 시변상태전이행렬, $z(k)$ 는 $m \times 1$ 측정 벡터, $C(k)$ 는 $m \times n$ 시변 출력 행렬, $w(k)$ 는 $n \times 1$ 시스템 오차, 그리고 $v(k)$ 는 $m \times 1$ 측정 오차를 나타낸다.

칼만이득방정식 :

$$K(k) = [A(k)P(k)C^T(k)][C(k)P(k)C^T(k) + Q_1]^{-1} \quad (7)$$

상태예측 방정식 :

$$\hat{x}(k+1) = A(k)\hat{x}(k) + K(k)[z(k) - C(k)\hat{x}(k)] \quad (8)$$

오차 공분산 방정식 :

$$P(k+1) = [A(k) - K(k)C(k)]P(k)[A(k) - K(k)C(k)]^T + K(k)Q_2K^T(k) \quad (9)$$

여기서, $P(k)$ 는 오차 공분산 행렬, $\hat{x}(k)$ 는 k 시점의 상

태 예측값, Q_1 은 추정오차의 공분산 행렬, 그리고 Q_2 는 시스템오차의 공분산 행렬을 각각 나타낸다.

3. 제안된 데이터 마이닝 결합모델

데이터 마이닝은 대규모 데이터 저장소에서 유용한 정보를 자동적으로 탐색하는 과정이며 새롭고 유용한 패턴을 탐색하기 위해 적용된다[16].

데이터 마이닝의 핵심 작업은 다음 4가지로 나타낼 수 있다. 첫째, 예측 모델링은 목표 변수를 설명 변수의 함수 모델로 생성하는 작업으로서, 이산형 목표 변수에 사용하는 분류와 연속형 목표 변수에 사용하는 회귀의 두 가지 유형으로 나눌 수 있다.

둘째, 연관 분석은 데이터에 강하게 연관된 특징을 설명하는 패턴을 발견하는 데 사용한다.

셋째, 군집 분석은 동일한 군집에 속하는 관측들은 다른 군집에 속하는 관측보다 더 유사하도록 긴밀하게 관련된 관측의 그룹을 탐색하는 것이다.

넷째, 이상치 탐지는 특징이 다른 나머지 데이터들과 현저히 다른 관측들을 식별하는 작업이다.

본 연구에서 제안한 데이터 마이닝 결합모델에 대한 근사 모델을 그림 4와 같이 나타낼 수 있다. 그림에서와 같이 기상정보를 포함한 일일 물 사용량에 대한 데이터 마이닝을 실시하고 다중선형 회귀모델의 회귀모수를 칼만필터를 사용하여 추정함으로써 동적시스템으로의 구현이 가능하게 된다. 또한, 명절 연휴의 특징을 나타내는 특이일에 대해서는 과거 데이터를 이용하여 구할 수 있는 유사일 기법을 적용한다.

본 연구에서는 데이터 마이닝 기법 적용을 통한 모델 도출을 위하여 물 공급량에 대한 이상치 탐지를 적용하였다. 이상치 탐지의 기준은 전일 공급량 대비 10 % 범위를 벗어난 경우 물 공급시스템의 이상이나 공급계통의 변경에 따른 변화로 인식하여 이상치로 분류하였으며, 이상치의 마이닝은 과거 실적데이터를 적용하였다. 과거 실적데이터는 전일 2일간 물 공급량의 평균값으로 설정하여 적용하였다. 이러한 이상치는 명절 등의 특이일이거나 또는 평일에 대한 추정오차에 의한 경우로 나눌 수 있다. 우선 특이일인 경우에는 그림 5에서 보는 바와 같이 과거의 유사일 선택에 의해 물 수요량을 예측하였다. 특이일이 아닌 경우에는 다중선형 자기회귀모델과 칼만필터에 의해 물 수요량을 추정하였다. 본 연구에서 적용한 유사일 기법은 명절 연휴만

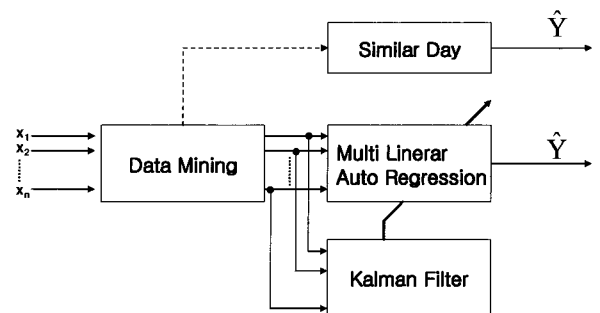


그림 4. 제안된 수요예측모델 블록도.

Fig. 4. Block diagram of proposed demand forecasting model.

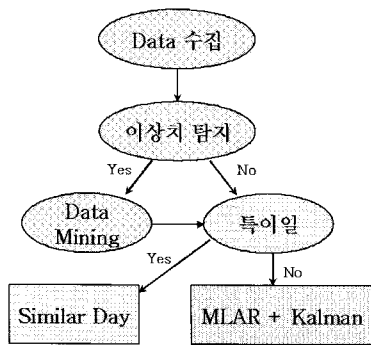


그림 5. 의사결정 트리기법 순서도.
Fig. 5. Flow of decision tree model.

해당되기 때문에 1년중 6일로서, 전년도 명절 연휴의 해당 일에 대한 공급량 실적을 적용하였다.

본 연구에서 적용된 단기 물 수요예측 모델식은 식 (4)로 표현된 1일전 공급량과, 2일전 공급량, 그리고 1일전 평균기온의 함수인 다중선형 자기회귀모델이다. 식 (4)의 회귀 모수를 칼만 필터의 상태변수로 선정하여 추정값이 최적화되도록 구성한다. 따라서, 회귀 모수값을 구하기 위한 칼만필터의 관측 방정식 $z(k) = C(k)x(k)$ 를 다음 식과 같이 정의하였다.

$$y(k) = [1 \ y(k-1) \ y(k-2) \ t(k-1)] [a_0 \ a_1 \ a_2 \ b_0]^T \quad (10)$$

여기서, 상태벡터 $x = [a_0 \ a_1 \ a_2 \ b_0]^T$ 는 칼만필터에 의해 계산하고, 계산된 상태벡터를 이용하여 자기회귀모델을 구한다. 따라서, 1일전 공급량과 2일전 공급량, 그리고 1일전 기온값과 구해진 상태변수값을 이용하여 물 수요량을 나타내는 y 값을 예측한다. 그리고, 칼만필터 모델의 적용을 위하여 식(5)에서의 A는 단위행렬을 적용하고, 식(6)에서의 $C(k)$ 값은 $[1 \ y(k-1) \ y(k-2) \ t(k-1)]$ 이다. 또한, 칼만필터에 적용된 시스템오차와 측정오차에 대한 공분산 행렬은 단위행렬을 적용하였다. 제안된 결합모델에 대한 의사결정 트리를 통한 순서도는 위 그림 5와 같이 나타낼 수 있다.

IV. 시뮬레이션 및 결과분석

본 연구에서는 A정수장의 2006년부터 2008년까지의 3년간의 물 공급량과 기온 데이터를 바탕으로 다중선형 자기회귀 모델, 칼만필터 모델, 그리고 제안된 모델에 대하여 각각 비교분석을 실시하였다. 세가지 모델에 대한 적용결과와 성능평가를 위하여 적용된 방법은 절대평균 백분위 오차(MAPE: Mean Absolute Percentage Prediction Error)를 사용하였으며, 다음 식과 같이 표현할 수 있다.

$$MAPE(\%) = \frac{1}{N} \sum_{k=1}^N \frac{|y_k - \hat{y}_k|}{y_k} \times 100 \quad (11)$$

여기서, y_k 는 실제 물 공급량, \hat{y}_k 는 예측 물 공급량이다.

1. 다중선형 자기회귀 모델

다중선형 자기회귀 모델의 적용성을 분석하기 위하여 추

정오차는 2006년과 2007년 2년에 대하여 적용하였으며, 예측오차는 2008년 운영데이터에 의한 결과와의 비교를 통하여 MAPE를 구하였다. 모델의 차수가 9차 이상일 경우 절대평균 백분위 오차가 최대한 줄어들지만, 본 연구에서는 오차 변화폭이 적고 현장 적용성을 높일 수 있는 2차 모델식을 적용하였다. 따라서, 표 3에서는 다중선형 자기회귀 모델의 식 (4)와 같이 1일전 공급량, 2일전 공급량과 1일전 평균기온에 의한 모델식을 도출한 결과를 나타내었다. 또한, 물 공급량의 계절적 특성 및 주기 특성을 도출하기 위하여 평일, 주말, 특이일, 여름 등으로 나누어 각각의 모델식을 구하여 예측값을 구하였다. 표 3의 결과에서와 같이 평일의 측정오차 및 예측오차는 각각 3.73 %와 3.30 %로 나타났으며, 주말에 대한 결과는 각각 3.78 %와 3.13 %로 나타났다. 그러나 특이일에서의 경우는 8.75 %와 12.2 %로 측정오차 및 예측오차가 매우 높게 나타났다. 이와 같이 평일, 주말, 그리고 특이일로 나누어 분석한 결과 년 평균 추정오차가 3.81 %이며, 예측오차는 3.35 %로 나타났다. 계절적 특성을 고려하여 여름철과 비 여름철에 대한 다른 추정 오차와 예측오차의 경우 여름철 모델에 대한 오차가 크게 나타남을 알 수 있었다. 위와 같이 요일별 특성과 계절별 특성을 분류한 결과와 비교하여 전체 데이터에 대한 다중선형 회귀모델의 추정 결과 요일별 특성과 계절별 특성에 비하여 양호함을 알 수 있다.

2. 칼만필터 모델

칼만필터모델에 의한 물 수요량 예측 결과는 그림 6과 표 4와 같으며, 물 사용량의 급격한 변화가 발생하는 경우에는 예측 정확도가 낮아지는 것을 알 수 있다. 여름철과 명절 등의 특이일에 특히 많은 오차를 내포하고 있음을 알 수 있다. 이러한 경향은 2006년부터 2008년까지 전반적으로 유사하게 나타나고 있다. MAPE값 분석 결과를 살펴보면 2006년도에서는 4.20 %, 2007년도는 3.91 %, 2008년도에서는 3.57 %로 각각 나타났다. 각 그림의 상단부는 실적치와 예측치에 대한 그래프를 나타낸 것이며, 하단부는 실적치와 예측치의 오차에 대한 그래프를 나타내었다.

표 3. 다중선형 자기회귀모델 수요예측결과.

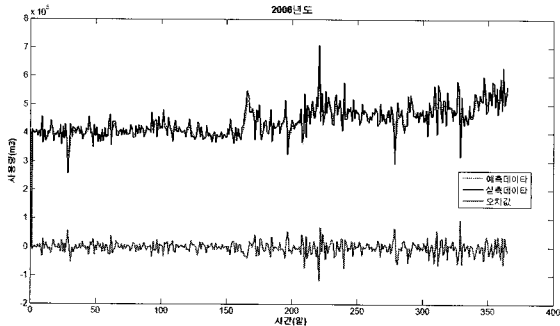
Table 3. Forecasting result of multi linear auto regression.

구분	모델식	추정 오차 (%)	예측 오차 (%)
평일	$y(k) = 59192 + 0.673y(k-1) + 0.194y(k-2) + 168t(k-1)$	3.73	3.30
토일	$y(k) = 75994 + 0.621y(k-1) + 0.211y(k-2) + 59t(k-1)$	3.78	3.13
특이일	$y(k) = 419960 - 0.064y(k-1) - 0.022y(k-2) + 1450t(k-1)$	8.75	12.2
년 평균		3.81	3.35
여름	$y(k) = 58987 + 0.66y(k-1) + 0.156y(k-2) + 1211t(k-1)$	4.06	4.41
여름외	$y(k) = 52518 + 0.65y(k-1) + 0.234y(k-2) + 18t(k-1)$	3.51	3.00
년 평균		3.65	3.35
모든일	$y(k) = 56416 + 0.674y(k-1) + 0.2y(k-2) + 137t(k-1)$	3.64	3.31

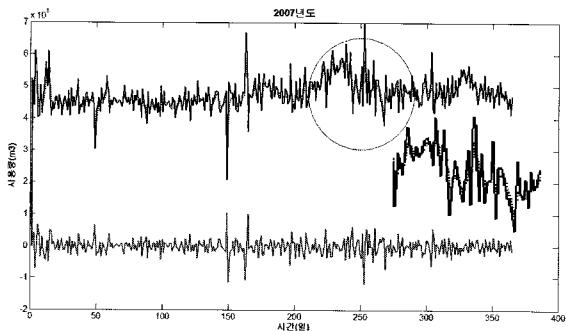
표 4. 칼만필터 수요예측결과.

Table 4. Demand forecasting result of Kalman filtering.

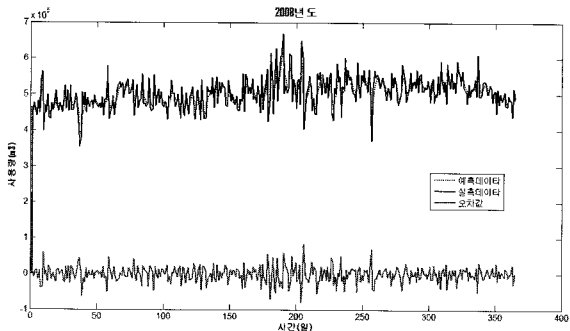
구 분	2006년 MAPE(%)	2007년 MAPE(%)	2008년 MAPE(%)
년 평균	4.20	3.91	3.57



(a) 2006년도



(b) 2007년도



(c) 2008년도

그림 6. 칼만필터 수요예측.

Fig. 6. Demand forecasting using Kalman filtering.

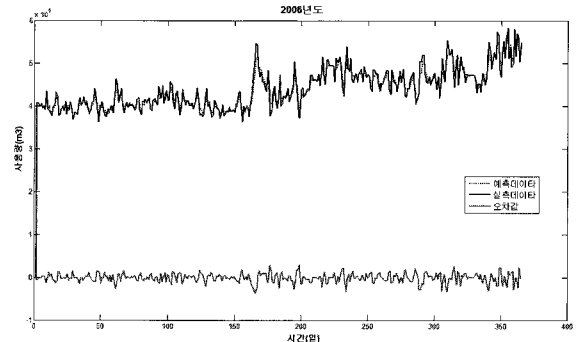
3. 제안된 데이터 마이닝 결합모델

본 연구에서 제안된 알고리즘의 특성을 분석하기 위하여 요일별 특성과 계절별 특성, 그리고 전체일에 대한 적용성을 실험하였다. 실험 결과 값은 표 5와 같으며 평일에서의 경우 주말과 특이일에 비해 양호한 결과를 얻었으며 이는 표 3에서의 다중선형 자기회귀모델과 비교하면 아주 양호함을 알 수 있다. 뿐만 아니라 계절별 특성과 특이일을 포함한 년간 평균값에서도 아주 양호한 결과를 얻었다. 특히 특이일에 대하여 유사일기법을 도입한 결과 전체적인 MAPE값이 2.23 %와 2.28 %로 매우 양호한 결과를 얻었다.

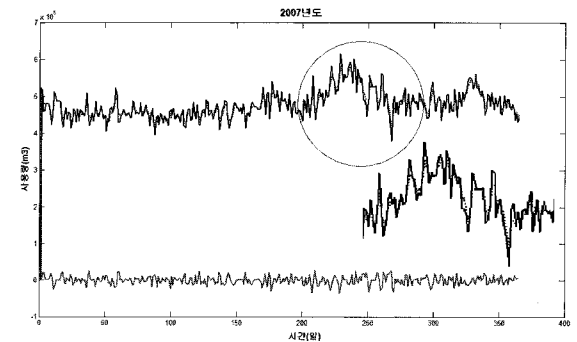
표 5. 제안된 알고리즘의 수요예측결과.

Table 5. Demand forecasting result of proposed algorithm.

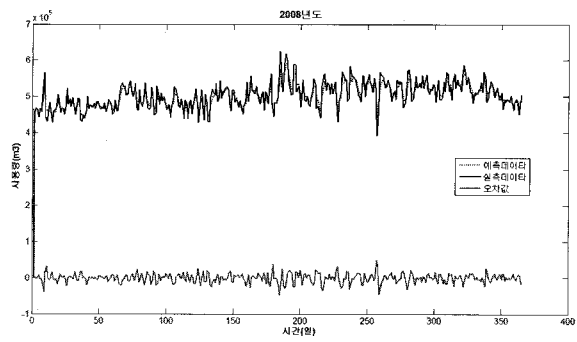
구 분	2006년 MAPE(%)	2007년 MAPE(%)	2008년 MAPE(%)
월회수목금	2.33	2.49	2.27
토일	3.14	2.85	3.20
특이일	19.01	18.27	19.23
년 평균	2.83	2.85	2.81
여름(6월,7월,8월)	3.19	2.87	3.40
여름의 기간	2.18	2.35	2.16
년 평균	2.47	2.52	2.51
특이일	-	6.15	6.23
특이일의 기간	2.12	2.17	2.22
년 평균	2.12	2.23	2.28



(a) 2006년도



(b) 2007년도



(c) 2008년도

그림 7. 제안된 수요예측알고리즘.

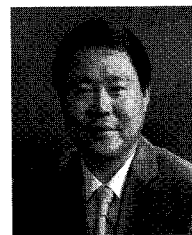
Fig. 7. Proposed demand forecasting.

V. 결론

본 연구에서는 물 공급량의 일일 수요 예측알고리즘을 개발하기 위하여 경기도에 위치한 A정수장에서의 3년간의 실제 운영데이터를 수집하였으며, 이를 이용하여 요일별 특성을 포함한 계절적 특성 등에 대한 상관성 분석을 실시하였다. 상관성 분석후 다중선형 자기회귀모델과 칼만필터 모델을 도출하였으며 이에 따른 실험을 통하여 MAPE 기준 약 3.35% ~ 4.2%의 오차를 보였다. 본 연구에서는 다중선형 자기회귀모델의 매개변수에 대하여 칼만필터링을 통한 적응형 결합모델을 제안하였다. 또한, 원시데이터의 실시간 데이터 마이닝기법과 특이일에 대한 유사일기법 적용을 통한 의사결정트리기법을 제안하였다. 본 연구에서 제안된 모델에 대한 실험결과 2006년, 2007년, 2008년에 대하여 연간 MAPE 약 2.12% ~ 2.28%의 양호한 결과를 얻었다. 따라서, 본 연구에서 제안된 데이터 마이닝기법 적용모델의 유용성을 확인하였으며, 이를 통하여 현장 적용이 가능한 것으로 판단된다. 향후에는 시간별 특성 분석을 통하여 본 연구에서 제안된 데이터 마이닝 적용모델의 일일 예측량을 기준으로 시간별 예측으로의 확대 적용을 위한 추가적인 연구를 수행하고자 한다.

참고문헌

- [1] 송경빈, 하성관, “단기 수요예측 알고리즘,” 대한전기학회, vol. 53A, no. 10, pp. 529-535, 2004.
- [2] S. Chentur Pandian, K. Duraiswamy, C. Chistober Asir Rajan, and N. Kanagaraj, “Fuzzy approach for short term load forecasting,” *Electric power systems research*, vol. 76, pp. 541-548, 2006.
- [3] P. Mandal, T. Senjyu, N. Urasaki, and T. Funabashi, “A neural network based several-hour-ahead electric load forecasting using similar days approach,” *Electrical power & energy system*, vol. 28, pp. 367-373, 2006.
- [4] D. Srinivasan, “Energy demand prediction using GMDH networks,” *Neurocomputing*, vol. 72, pp. 625-629, 2008.
- [5] 양승원, 박근준, “주택수요 예측을 위한 주택량과 상수도보급률의 상관성 분석,” 한국건설관리학회, 제6권 제2호, pp. 61-68, 2005.
- [6] H. S. Hwang, “Fuzzy GMDH-type neural network model and its application to forecasting of mobile communication,” *Computer & Industrial Engineering*, vol. 50, pp. 450-457, 2006.
- [7] J. L. Torres, A. Garcia, M. De Blas, and A. De Francisco, “Forecast of hourly average wind speed with ARMA models in Navarre(Spain),” *Solar Energy*, vol. 79, pp. 65-77, 2005.
- [8] 김신결, 변신숙, 김영상, 구자용, “시스템 다이내믹스법을 이용한 서울특별시의 장기 물수요예측,” 환경공학회, 제20권 제2호, pp. 187-198, 2006.
- [9] H. M. Al-Hama and S. A. Soliman, “Short-term electric load forecasting based on Kalman filtering algorithm with moving window weather and load model,” *Electric power systems research*, no. 68, pp. 47-59, 2004.
- [10] 이동철, 홍연찬 “GMDH를 이용한 전력 수요 예측 알고리즘 개발”, 퍼지 및 지능시스템학회, 제13권 제3호, pp. 360-365, 2003.
- [11] 한태환, 남의석 “칼만필터의 적응형모델 기법을 이용한 광역상수도 시스템의 수요예측 모델 개발,” 한국조명전기설비학회, 제15권 제2호, pp. 38-48, 2001.
- [12] 남의석, 우천희, 김학배, “광역상수도의 최적운영 및 제어를 위한 수운영시스템 개발,” 제어·자동화·시스템공학회, 제9권 제7권, pp. 489-497, 2003.
- [13] J. Bougadis, K. Adamowski, and R. Diduch, “Short-term municipal water demand forecasting,” *Hydrological Processes*, vol. 19, pp. 137-148, 2005.
- [14] S. L. Zhou, T. A. McMahon, A. Walton, and J. Lewis, “Forecasting daily urban water demand: a case study of melbourne,” *Journal of Hydrology*, vol. 236, pp. 153-164, 2000.
- [15] R. E. Kalman, “A new approach to linear filtering and prediction problems,” *Transaction of the ASME-Journal of Basic Engineering*, pp. 35-45, 1960.
- [16] P.-N. Tan, “Introduction to data mining,” *Addison-Wesley*, 2006.

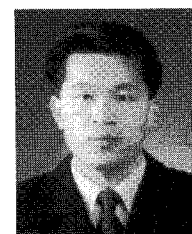


최기선

1992년 충남대학교 재료공학과 졸업.
2002년 한밭대학교 제어계측공학과 공학석사.
2009년 충북대학교 대학원 박사과정 수료.
1992년~현재 한국수자원공사 k-water 교육원 차장. 관심분야는 수운영시스템, 수요예측, RFID/USN.

신강욱

제어·로봇·시스템학회 논문지 제15권 제7호 참조.



임상혜

1992년 조선대학교 전자공학과 졸업.
1992년~현재 한국수자원공사 수도권 운영처 차장. 관심분야는 수도시설의 통합운영체계, 수운영시스템 구축 및 운영, 수처리공정 자동제어.

전명근

제어·로봇·시스템학회 논문지 제15권 제5호 참조.