

## 데이터 마이닝을 이용한 단기부하예측 시스템 연구

### A Study on Short-Term Load Forecasting System Using Data Mining

김 도완\*, 박 진배\*, 김 정찬\*\*, 주 영훈\*\*

(Do Wan Kim, Jin Bae Park, Juhg Chan Kim, and Young Hoon Joo)

\* 연세대학교 전기전자공학부(전화:(02)2123-2773, E-mail : {dwkim,jbark}@control.yonsei.ac.kr)

\*\* 군산대학교 전자정보공학부(전화:(063)469-47067, E-mail : yhjoo@kunsan.ac.kr)

**Abstract:** This paper presents a new short-term load forecasting system using data mining. Since the electric load has very different pattern according to the day, it definitely gives rise to the forecasting error if only one forecasting model is used. Thus, to resolve this problem, the fuzzy model-based classifier and predictor are proposed for the forecasting of the hourly electric load. The proposed classifier is the multi-input and multi-output fuzzy system of which the consequent part is composed of the Bayesian classifier. The proposed classifier attempts to categorize the input electric load into Monday, Tuesday~Friday, Saturday, and Sunday electric load. Then, we construct the Takagi-Sugeno (T-S) fuzzy model-based predictor for each class. The parameter identification problem is converted into the generalized eigenvalue problem (GEVP) by formulating the linear matrix inequalities (LMIs). Finally, to show the feasibility of the proposed method, this paper provides the short-term load forecasting example.

**Keywords :** Short-term load forecasting, data mining, fuzzy-Bayesian classifier, Takagi-Sugeno (T-S) Fuzz model-based predictor, linear matrix inequality (LMI), generalized eigenvalue problem (GEVP).

#### I. 서 론

복잡하고 대규모화된 전력 계통의 최적 운용 계획을 위해서 연간 최대 전력 부하 및 시간별 전력 부하에 대한 장기 및 단기간의 전력 부하 예측이 필요하다. 특히 단기 부하 예측은 발전 비용 및 신뢰도와 밀접한 관계를 가지며 그 결과는 예측에 가까운 날의 수요 패턴과 기상 변화 등에 많은 영향을 받으며 전력 계통의 제어 및 단기 계획, 전력 조류 계산 등의 입력 자료로 사용된다. 단기 부하 예측은 시계열 분석 기법[3], 회귀 기반 기법 [4]과 인공 지능(AI)형 기법[5,6], 등과 같이 다양한 기법으로 연구되고 있다. 하지만 전력 부하는 요일 별로 매우 상이한 형태의 유형을 가지므로 한 가지 예측 모델로 단기 부하 예측을 시행할 시 많은 오차를 유발하는 문제점이 있다.

본 논문은 시간별 전력 부하 예측을 위한 퍼지 모델 기반 분류기와 예측기를 다룬다. 제안된 분류기는 다중 입력출력 퍼지 모델을 이며, 강한 성능을 위해서 모델의 후반부는 베이저안 분류기가 적용되었다. 이러한 분류기를 통해서 얻어진 각 클래스별 전력 부하 데이터를 기반 하여 우리는 Takagi-sugeno (T-S) 퍼지 모델 기반 예측기를 설계한다. 파라미터 동정 문제는 선형 행렬 부동식 (LMIs) 형태의 generalized eigenvalu problem (GEVP) 문제로 대체함으로써, 주어진 조건에서 항상 유일한 최적 해가 존재한다. 마지막으로, 제안된 분류기와 예측기의 실효성 검증을 위한 실제 전력 부하 예측이 수행된다.

#### II. 전력 부하 예측을 위한 퍼지 모델 기반 분류기 및 예측기 설계

##### 1. 퍼지- 베이저안 분류기

우리는 전력 부하 데이터를 월요일, 화요일~금요일, 토요일, 일요일로 분류하였다. 적용된 기법은 다음과 같은 퍼지-베이저안 모델이다.

$$R_i: \text{IF } x_1 \text{ is } \mathcal{E}_{i1} \text{ and } \dots \text{ and } x_{48} \text{ is } \mathcal{E}_{i48}, \text{ THEN } y_i \quad (1)$$

여기서,  $[x_1, \dots, x_{24}]$  는 전날의 시간별 전력 부하이므로,  $[x_{25}, \dots, x_{48}]$  는 전주의 시간별 전력 부하이므로,

$\mathcal{E}_{ij}$  는  $i=[1,2,\dots,4]$  번째 규칙의  $j$  번째 입력의 소속 함수이다. 소속 함수는 입력의 평균과 표준 편차로 쉽게 구성할 수 있도록 다음과 같이 정한다.

$$\mathcal{E}_{ij} = \exp\left(\frac{-(x_j - m_{ij})^2}{2\sigma_{ij}^2}\right) \quad (2)$$

퍼지-베이저안 분류기 (1)의 후반부는 다음과 같이 베이저안 분류기가 사용된다.

$$y_i = \begin{bmatrix} d_1(x) - d_i(x) \\ \vdots \\ d_i(x) \\ \vdots \\ d_4(x) - d_i(x) \end{bmatrix} \quad (3)$$

여기서,

$$d_i(x) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-m_i)^T \Sigma^{-1}(x-m_i)\right) P(C_i) \quad (4)$$

따라서 가중치 평균에 의한 추론을 이용하여 (1)의 전체 결과는 다음 식에 의해서 결정된다.

$$\hat{y} = \frac{\sum_{i=1}^4 w_i y_i}{\sum_{i=1}^4 w_i} \quad (5)$$

$$w_i = \prod_{j=1}^{48} \mathcal{E}_{ij}(x_j), \quad \hat{y} = \begin{bmatrix} \hat{y}_1 \\ \vdots \\ \hat{y}_i \\ \vdots \\ \hat{y}_4 \end{bmatrix}$$

여기서,

결국, 주어진 특성 벡터에 대해서 제안된 모델은 다음과 같은 분류를 실행한다.

$$\hat{y}_i > \hat{y}_r, \quad \forall i \neq r, \quad i \in \{1, 2, \dots, 4\} \quad (6)$$

퍼지-베이지안 분류기의 모델의 복잡도를 낮추기 위해 우리는 특성 변수 제거 알고리즘을 제안한다. 우선 다음 식과 같은 퍼지 집합 정확도 측정 기법을 정의하자.

$$C(\mathcal{E}_{ij}) = \frac{\sum_{x_i \in \mathcal{K}_y} \mathcal{E}_{ij}}{\sum_x \mathcal{E}_{ij}} \quad (7)$$

식 (7)을 기반으로 제안된 알고리즘은 먼저 퍼지 규칙을 선택 후, 그 규칙의 제거될 특성 변수를 선택한다. 선택 기준은 앞서 설명한 정확도 측정 기법에 기반한다. 우선 규칙 선정은 그것의 입력으로 사용된 특성 변수의 정확도 평균이 가장 높은 순서로 선택한다. 가장 높은 정확도 평균을 가지는 규칙이 표현하는 클래스가 다른 클래스와 분리되어 있을 가능성이 높기 때문에 상대적으로 적은 특성 변수 개수로도 그 클래스를 잘 표현할 수 있다. 규칙을 선택 후, 정확도 정도가 가장 낮은 특성 변수를 후보로써 선정하고 특성 인식이 떨어지지 않는다면 그 특성 변수가 제거된다.

## 2. T-S 퍼지 모델 기반 예측기

기존 많은 연구에서 전력 부하는 전날의 전력부하와 전주의 전력부하와 매우 흡사한 패턴을 형태임이 검증되었다. 따라서 우리는 전날과 전주의 전력 부하를 입력을 가지는 선형 전력 부하 예측 모델을 설정하였다.

$$\hat{x}(k) = A_1 x(k-1) + B_1 \quad (8)$$

$$\hat{x}(k) = A_2 x(k-7) + B_2 \quad (9)$$

여기서,  $\hat{x}(k) = [x_1(k), x_2(k), \dots, x_{24}(k)]^T$  는  $k$  번째 날의 24시간 전력 부하 예측 값이며,  $x(k-1)$  와  $x(k-7)$  는 각각 전날과 일주일전의 24시간 실제 전력 부하 크기이고  $A_1, A_2, B_1, B_2$  는 동정될 행렬이

다.

모델 동정 문제는 모든 주어진  $k$  개의 데이터에 대해서 가능한 다음과 같은 등식을 만족하도록 하는 행렬  $A_1, A_2, B_1, B_2$  을 찾는 것이다.

$$x(k) = \hat{x}(k) \quad (10)$$

그러나 모든  $k$  개의 데이터에 대해서 식(10)을 만족하는 행렬  $A_1, A_2, B_1, B_2$  를 해석적으로 구하는 것은 사실 불가능하다. 이러한 문제를 해결하기 위해, 우리는 모델 동정 문제를 수치 최적화 기법을 이용하여 행렬  $A_1, A_2, B_1, B_2$  을 찾는다. 우리의 중심 아이디어는  $x(k)$  와  $\hat{x}(k)$  사이의 놈(norm) 거리를 가능한 최소화하는 행렬  $A_1, A_2, B_1, B_2$  을 구하는 것이다. 구체적으로 모델 동정 문제는 다음과 같이 설정된다.

**문제 1.** 모든  $k$  개 데이터에 대해서 가능한 다음 부등식 제약 조건을 만족하는 행렬  $A_1, A_2, B_1, B_2$  을 찾으시오.

*Minimize*  $A_1, B_1, \alpha$

$$(i) \quad \|x(k) - A_1 x(k-1) - B_1\| < \alpha^2 I$$

*Minimize*  $A_2, B_2, \beta$

$$(ii) \quad \|x(k) - A_2 x(k-7) - B_2\| < \beta^2 I$$

주목할 점은 문제 1은 볼록(convex) 최적화 문제라고 주어졌으므로 선형 행렬 부등식 (LMIs)으로 구성함으로써 수치적으로 해결할 수 있다는 것이다. 전력 부하 예측 모델의 주요 결과는 다음과 같이 요약된다.

**정리 1.** 모든  $k$  에 대해서 다음과 같은 GEVPs의 해가 존재하도록 하는 행렬  $A_1, A_2, B_1, B_2$  와 실수  $\alpha$  와  $\beta$  가 존재한다면,

GEVP 1

*Minimize*  $A_1, B_1, \alpha$

$$\begin{bmatrix} \alpha I \\ x(k) - A_1 x(k-1) - B_1 \\ \alpha I \end{bmatrix}^* < 0 \quad (11)$$

GEVP 2

*Minimize*  $A_2, B_2, \beta$

$$\begin{bmatrix} \beta I \\ x(k) - A_2 x(k-7) - B_2 \\ \beta I \end{bmatrix}^* < 0 \quad (12)$$

전력 부하 예측 모델 (8,9)은 실제 전력 부하와 거의 같은 결과를 예측할 수 있다. 여기서, '\*' 는 대응하는 대각(symmetric) 위치에서 전치 행렬이다.

**증명:** 문제 1로부터 다음과 같은 부등식이 주어진다.

$$\|x(k) - A_1 x(k-1) - B_1\| < \alpha^2 I \quad (13)$$

$$\|x(k) - A_2 x(k-7) - B_2\| < \beta^2 I \quad (14)$$

2놈(induced-2 norm)정의로부터 다음 부등식은 유효하다.

$$(x(k) - A_1 x(k-1) - B_1)^T (x(k) - A_1 x(k-1) - B_1) < \alpha^2 I \quad (15)$$

$$(x(k) - A_2 x(k-7) - B_2)^T (x(k) - A_2 x(k-7) - B_2) < \beta^2 I \quad (16)$$

Schur complement 이용해서, 식(15) 와 식(16)은 각각

식(11)과 식(12) 같이 선형 행렬 부등식 (LMIs)으로 표현될 수 있다.■

참고 1. 정리 1에서 행렬  $A_1, A_2$  를 일반 행렬, 대각 (symmetric) 행렬, 또는  $a_1I, a_2I$  행렬 형태로 설정함으로써 모델의 복잡성을 조절할 수 있다. 따라서 과잉적합 (overfitting) 문제를 해결할 수 있는 가능성이 있다.

우리는 이론 1을 통해서 잘 동정된 전력 부하 예측 모델 (8)과 (9)를 얻을 수 있다. 그런데 전력 부하  $x(k)$  을 예측하기 위해서 예측 모델 (8)과 (9) 선택 문제가 있다. 이러한 문제에 대한 해결책으로 우리는 다음과 같은 후건부가 예측 모델 (8)과 (9)로 구성된 T-S 퍼지 모델 기반 시간별 전력 부하 예측기를 제안한다.

$$\begin{aligned}
 R_1: & \text{ IF } e(x(k-1)) \text{ is } \Gamma_{11} \text{ and } e(x(k-7)) \text{ is } \Gamma_{12}, \\
 & \text{ THEN } \hat{x}_1(k) = A_1 x(k-1) + B_1 \\
 R_2: & \text{ IF } e(x(k-1)) \text{ is } \Gamma_{21} \text{ and } e(x(k-7)) \text{ is } \Gamma_{22}, \\
 & \text{ THEN } \hat{x}_2(k) = A_2 x(k-7) + B_2
 \end{aligned}
 \tag{17}$$

여기서,

$$\begin{aligned}
 e(\cdot) &= \frac{|x(k) - (\cdot)|}{x(k)} * 100 \\
 \Gamma_{11} &= -\frac{1}{\max_{k} \sqrt{k} e(x(k-1))} e(\hat{x}_1(k)) + 1 \\
 \Gamma_{22} &= -\frac{1}{\max_{k} \sqrt{k} e(x(k-7))} e(\hat{x}_2(k)) + 1 \\
 \Gamma_{21} &= 1 - \Gamma_{11} \\
 \Gamma_{12} &= 1 - \Gamma_{22}
 \end{aligned}$$

따라서 가중치 평균 추론을 이용 (17)의 전체 결과는 다음과 같다.

$$\hat{x}(k) = \frac{\Gamma_{11}\Gamma_{12}\hat{x}_1(k) + \Gamma_{21}\Gamma_{22}\hat{x}_2(k)}{\Gamma_{11}\Gamma_{12} + \Gamma_{21}\Gamma_{22}}
 \tag{18}$$

### III. 모의실험 결과 및 분석

본 논문은 제안된 기법의 검증용 위해 1999년 12월 17일부터 12월 23일 실제 국내 전력 부하량을 예측하였다. 우선, 우리는 1999년 10월 4일부터 12월 19일 기간 동안의 국내 시간별 전력 부하 데이터 (그림 1)를 이용해서 제안된 분류기와 예측기를 동정하였다. 동정된 분류기는 주어진 입력을 표1과 같이 정해진 그룹으로 완벽히 분류하였다. 한편, 주어진 10/1일부터 12/16일 전력부하 데이터 크기는 월요일, 토요일, 일요일 전력 부하 예측기를 동정하기에는 상대적으로 적으므로, 과잉적합 (overfitting) 문제가 발생가능성이 크다. 이러한 문제점을 해결하기 위해서 우리는 월요일, 토요일, 일요일 전력 부하 예측기의 행렬  $A_1$ 과  $A_2$ 를 행렬  $a_1I$ 와  $a_2I$ 로 치환을 통해 모델의 복잡성을 상당히 낮추었다. 그림2부터 그림 5는 예측기의 동정 결과를 보여준다. 마지막으로, 우리는 동정된 분류기와 예측기의 성능을 1999년 12월 20일부터 12월 26일의 시간별 전력 부하 데이터 대해서 검증하였다. 표1과 그림 6을 통해서 제안된 분류기의 우수성이 입증되었다.

표 1. 전력 부하 데이터의 분류 결과(%)

10/1~12/16 전력 부하			12/17~12/23 전력 부하				
월	화~금	토	일	월	화~금	토	일
100	100	100	100	100	100	100	100

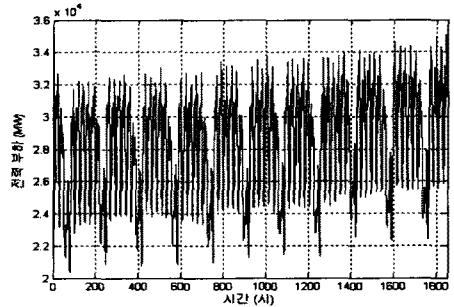


그림 1. 1999년 10월 4일부터 12월 19일까지 국내 시간별 전력 부하 (MW)

Fig.1 Domestic hourly electric load for 10/4~12/19

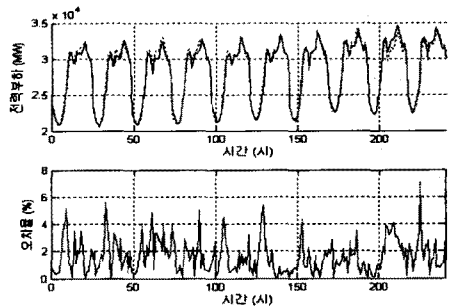


그림 2. 그림 2. 10/1~12/16 기간 중 월요일 전력 부하에 대한 예측 결과 (실선-실제 전력 부하, 점선-예측 전력 부하).

Fig. 2. Prediction results for Monday electric load data between 10/1~12/16.

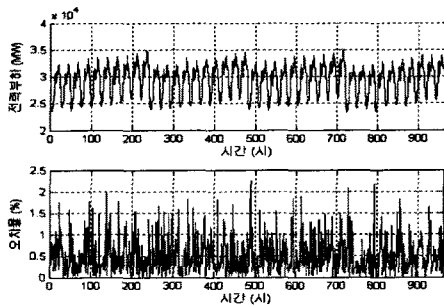


그림 3. 10/1~12/16 기간 중 화요일부터 금요일 전력 부하에 대한 예측 결과 (실선: 실제 전력 부하, 점선: 예측 전력 부하).

Fig. 3. Prediction results for Tuesday~Wednesday Electric load data between 10/1~12/16.

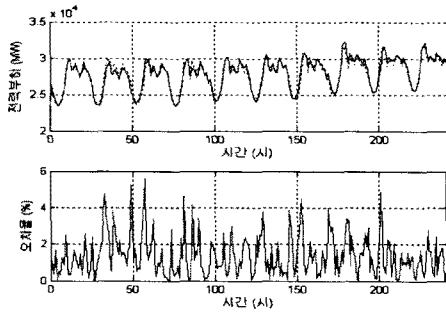


그림 4. 10/1~12/16 기간 중 토요일 전력 부하에 대한 예측 결과 (실선: 실제 전력 부하, 점선: 예측 전력 부하).

Figure 4. Prediction results for Saturday Electric load data between 10/1~12/16.

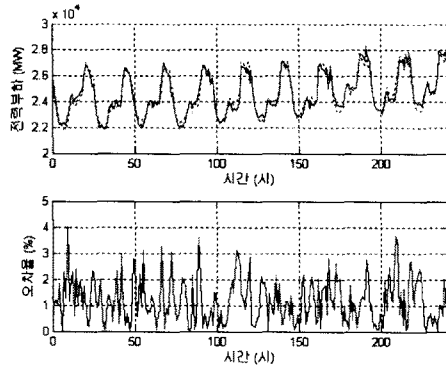


그림 5. 10/1~12/16 기간 중 일요일 전력 부하에 대한 예측 결과 (실선-실제 전력 부하, 점선-예측 전력 부하).

Fig. 5. Prediction results for Saturday Electric load data between 10/1~12/16.

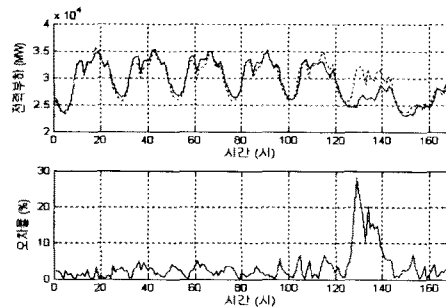


그림 6. 12/17~12/23의 전력 부하 예측 결과 (실선-실제 전력 부하, 점선- 예측 전력 부하).

Fig. 6. Prediction result for 12/17~12/23 Electric load data.

#### IV. 결론

본 논문은 시간별 전력 부하 예측을 위한 퍼지 모델 기반 분류기와 예측기를 제안하였다. 제안된 분류기는 다중 입출력 퍼지 모델을 이용하였으며, 강인한 성능을 위해서 모델의 후반부는 베이지안 분류기가 적용되었다. 이러한 분류기를 통해서 얻어진 각 클래스별 전력 부하 데이터를 기반 하여 우리는 퍼지 모델 기반 예측기를 설계하였다. 파라미터 동정 문제는 선형 행렬 부등식 (LMIs) 형태의 GEVP 문제로 대체함으로써, 우리는 주어진 조건에서 항상 유일한 최적 해를 얻었다. 마지막으로, 제안된 분류기와 예측기의 실효성은 실제 전력 부하 예측을 통해서 검증 되었다.

본 연구는 산업자원부의 지원에 의하여 기초전력공학 공동연구소(R-2003-B-078) 주관으로 수행된 과제임.

#### 참고 문헌

- [1] Y. H. Joo, H. S. Hwang, K. B. Kim, and K. B. Woo, "Linguistic model identification for fuzzy system," *Electron. Letter*, vol. 31, no. 4, pp. 330-331, 1995.
- [2] Y. H. Joo, H. S. Hwang, K. B. Kim, and K. B. Woo, "Fuzzy system modeling by fuzzy partition and GA hybrid schemes," *Fuzzy Set Syst.*, vol. 86, no. 3, pp. 279-288, 1997.
- [3] N. Amjady, "Short-term hourly load forecasting using time-series modeling with peak load estimation capability," *IEEE Trans. Power Syst.*, Vol. 16, No. 3, pp. 498-505, 2001.
- [4] S. J. Huang and K. R. Shih, "Short-term load forecasting via ARMA model identification including non-Gaussian process considerations," *IEEE Trans. Power Syst.*, Vol. 18, No. 2, pp. 673-679, 2003.
- [5] A. Khotanzad, E. Zhou, and H. Elragal, "A neuro-fuzzy approach to short-term load forecasting in a price-sensitive environment," *IEEE Trans. Power Syst.*, Vol. 17, No. 4, pp. 1273-1282, 2002.
- [6] H. S. Hippert, C. E. Pedreira, and R. C. Souza, "Neural networks for short-term load forecasting: a review and evaluation," *IEEE Trans. Power Syst.*, Vol. 16, No. 1, pp. 44-55, 2001.