

## A study on electricity demand forecasting based on time series clustering in smart grid

Hueng-Goo Sohn<sup>a</sup> · Sang-Wook Jung<sup>a</sup> · Sahn Kim<sup>a,1</sup>

<sup>a</sup>Department of Applied Statistics, Chung-Ang University

(Received December 17, 2015; Revised December 30, 2015; Accepted December 30, 2015)

---

### Abstract

This paper forecasts electricity demand as a critical element of a demand management system in Smart Grid environment. We present a prediction method of using a combination of predictive values by time series clustering. Periodogram-based normalized clustering, predictive analysis clustering and dynamic time warping (DTW) clustering are proposed for time series clustering methods. Double Seasonal Holt-Winters (DSHW), Trigonometric, Box-Cox transform, ARMA errors, Trend and Seasonal components (TBATS), Fractional ARIMA (FARIMA) are used for demand forecasting based on clustering. Results show that the time series clustering method provides a better performances than the method using total amount of electricity demand in terms of the Mean Absolute Percentage Error (MAPE).

Keywords: time series clustering, TBATS model, double seasonal Holt-Winters model, electricity demand forecasting, dynamic time warping

---

### 1. 서론

스마트 그리드 환경하에서의 전력 체계는, 스마트 미터기를 통한 전력공급자와 소비자 간의 정보 교환을 통하여 전력의 사용 및 생산의 효율성을 높인다. 국내 전력시장에서 전력 생산방식은 최대 수요량을 기반으로 15%의 예비율을 두는 방식으로 생산되고 있으나, 2014년도 “제2차 에너지기본계획” 정책 이후 전력 관리 방침은 전력 공급 중심에서 전력 수요관리 중심으로 전환되어 가는 중이다. 전력 수요관리 중심으로의 변환은 효율성 문제에서 만이 아닌 환경적으로도 중요한 문제이다.

전력 공급 중심에서는 사용이 없을지 소모되는 전력으로 인하여 효율성이 떨어졌을 뿐만 아니라 석탄, 석유, 가스 등을 태우는 과정에서 온실가스 배출 등의 문제가 발생하였다. 온실가스 배출의 문제는 이전까지 환경에 위협을 가하는 정도로서만 문제가 제기되었으나, 2008년 7월 개최된 G7 확대 정상회의에서 한국은 온실가스 감축목표에 대하여 언급하였고, 2009년에 2020년 온실가스배출전망(Business As Usual; BAU) 대비 30% 감축의 국가 온실가스 감축목표가 확정하면서, 온실가스 배출은 환경 위협의 문제 만이 아닌 국가적 요구사항이 되었다. 이러한 온실가스 감축에서 핵심 영역이 발전 영역으로서, 스마트 그리드 환경을 도입을 통하여 전력 생산 감소를 통한 온실가스 감축을 목표로 하고 있는 것이다.

This research was supported by the Chung-Ang University Research Scholarship Grants in 2013.

<sup>1</sup>Corresponding author: Department of Applied Statistics, Chung-Ang University, 221 Heukseok-Dong, Dongjack-Gu, Seoul 06974, Korea. E-mail: [sahn@cau.ac.kr](mailto:sahn@cau.ac.kr)

즉, 스마트그리드와 스마트미터를 이용한 고효율 전력망 도입을 통하여, 불필요한 전력 생산을 막음으로써 온실가스 배출을 줄이는 것이다.

전력 수요예측의 연구 동향을 살펴보면, 국내 연구로서 Park 등 (2009)은 건물기반 시스템에서의 단기 전력 시스템의 수요와 공급 문제를 위하여 시계열 분석을 적용한 패턴인식을 이용한 전력수요를 예측을 위하여 SOM 알고리즘과 k-means 기법을 적용하여 요일별, 시간별 자료를 군집화하여, 지수평활기법과 ARIMA 모형을 비교 분석하였다. Kim과 Jeon (2010)는 장기 전력수요 예측절차 방법을 제안하여, 단기수요예측은 지수평활법을 이용한 시계열예측방법을 사용하는 반면, 장기 전력수요예측은 장기 전력수급계획의 중요한 기본 자료로서 사용되지만 이는 스마트 그리드 방식에서는 적합하지 않다고 밝혔다. Lee 등 (2010)은 전력피크에 대하여 회귀분석을 사용하였으며, 고온과 저온 및 GDP을 고려한 관계를 연구하여 지수평활법(exponential smoothing)을 포함한 회귀분석(regression)과 신경망 이론을 사용하였고, Lee와 Shin (2011)은 SVR(Support Vector Regression)에 기반한 전력 수요 예측을 하여 자기회귀 모형과 비교를 하였다. Hwang (2015)는 계층적 군집분석을 이용하여 건물의 전력 수요예측에 대한 연구가 이루어졌다.

국외에서도 다양한 분야에 대한 연구가 이루어졌다. Sari와 Soytaş (2004)는 분해 모형(decomposition models)을 이용하여 오차분해 기법을 통하여 자국의 총수입의 분산의 크기를 이용하여 에너지소비와 경제성장 간의 관계를 규명하였으며, Lise와 Montfort (2007)는 실질국민소득과 에너지수요에 대하여 VEC(Vector Error Correction) 모형 추정 및 그레인저 인과관계 모형(Granger causality model)를 실시하였다. ARIMA 모형과 관련된 수요 예측으로 전력 부문에서 Erdogdu (2010)는 터키자료를 이용하여 ARIMA 모형을 통하여 천연가스 수요에서 단기, 장기요금 및 수입탄력성 예측하였으며, Bakhat와 Rossello (2011)는 ARMAX, GARCH를 이용하여 일일전력수요와 관광효과에 대한 연구를 실시하였다. 신경망 모형으로는 Sozen과 Arcaklioglu (2007)는 한국에서의 에너지 수요예측을 신경망 모형을 이용하여 실시하였으며, Pao (2009)는 Winters의 지수평활법, EGARCH(Exponential GARCH), SEGARCH(Seasonal Exponential GARCH), WARCH(Weighted ARCH), ANN(Artificial Neural Network) 모형 중 하이브리드 비선형모형(hybrid non linear model)인 SEGARCH-ANN, WARCH-ANN으로 전기와 석유 소비예측을 실시하였다. 지수평활법 모형으로는 Taylor 등 (2006)에서 일별 미래 예측을 이중 계절성에 대해 지수평활법 방법 및 주성분 분석(PCA)을 기반으로 하는 새로운 방법을 모형을 다른 모형과 비교했으며, Taylor (2010)는 삼중 계절성 모형을 적용시키며 초단기 삼중 계절성 모형을 개발했다. Hong (2009)은 SVR를 이용하여 중국자료에 대하여 분석결과 적합자료가 적을 경우 SVR의 우수성을 보였으며, Wang 등 (2009)은 RSME를 기준으로 각각의 변수에 가장 적합한  $\epsilon$ -SVR 모형을 탐색하며, MA와  $\epsilon$ -SVR로 단기전력수요 예측하였다. 이와 같이 전력 수요 예측에 대한 연구는 다양한 방법이 적요되고 있음을 확인 할 수 있다.

시계열 예측 모형에 대한 연구로서 De Livera 등 (2011)는 기존의 시계열 모형들에서의 주기성 및 비선형성 등을 해결하기 위하여 BATS(Box-Cox transform, ARMA errors, Trend and Seasonal components) 모형과 TBATS(Trigonometric, Box-Cox transform, ARMA errors, Trend and Seasonal components) 모형을 제안하여, 수요 예측에서의 주기성의 정수로 지정되는 문제를 해결하였으며, Liu 등 (1999)는 이러한 주기성 문제 해결을 위하여 장기 특성을 고려한 FARIMA 모형을 제안하였다.

시계열 군집분석 방법으로는 Montero와 Vilar (2014)가 시계열 군집분석 방법을 원 변수를 기반한 시계열 군집분석 방법과 상관관계를 고려한 군집분석 방법, 자기상관성 및 부분 자기상관성을 고려한 시계열 군집분석 방법, 주기성 및 웨이블릿 방법들을 고려한 군집분석들을 소개하며, 이를 기반으로 TSclust R패키지를 개발하였다. Hong 등 (2009)는 국내에서 시계열 자료의 군집분석을 위해 시간영역과 진동수 영역에서의 군집분석을 소개하며, 실제 KOSPI200을 대상으로 분석을 실시하여, 웨이블릿의 우수성

을 보였으며, 경험적 모드분석(EMD)가 차분을 이용하여 추세를 제거하는 것보다 스펙트럴 밀도함수를 이용한 군집분석에서 더 효율적이었음을 밝혔다. Jang 등 (2010)은 동적 타임 워핑(DTW)를 통한 군집화로 시계열 패턴인식을 실시하였다.

기존 연구에서 전체 전력을 통한 분석을 진행되었는데 Park과 Kim (2012)에서 FARIMA 모형과 AR-GARCH를 사용하여 전력 수요예측을 실시하였으며, Lee 등 (2013)은 온도를 고려한 모형에서의 우수성을 Reg-ARIMA를 통하여 예측을 실시하였으며, Jung과 Kim (2014)은 일별 최대 전력 수요 예측에 있어서 Reg-AR-GARCH 모형을 이용하여, 공휴일, 명절 및 온도등을 고려한 수요예측을 실시하였다.

본 논문에서는 스마트 그리드 환경하에서 ICT기반 시장에서의 수요관리시스템에서의 핵심 요소인 전력 수요 예측을 위하여, 전체 사용량을 기반으로 예측 하는 방식이 아닌, 패턴에 따른 군집분석을 통한 군집별 사용량의 예측량의 결합으로서 예측을 실시하였다.

이에 따라 2장에서는 시계열 모형을 소개하며, 3장에서는 시계열 군집분석 방법을 소개한다. 4장은 국내 기업들의 전력 자료를 이용한 실증 분석을 실시하며, 5장에서는 결론을 나타냈다.

## 2. 시계열 모형 소개

### 2.1. Fractional ARIMA(FARIMA) 모형

FARIMA(Fractional ARIMA) 모형은 Box 등 (1976)이 제안한 ARIMA 모형에 Hosking (1981)의 장기기억 특성을 결합하여 Liu 등 (1999)이 제안한 모형으로 장기기억 특성을 고려한 모형이다. 이에 따라 ARIMA( $p, d, q$ ) 모형에서  $d$ 가 실수를 가지는 모형으로  $\Delta^d$ 는 분수차분연산자로 식 (2.1)과 같이 정의되며, FARIMA( $p, d, q$ )에 대해 식 (2.2)와 같이 정의된다. 여기서  $\epsilon_t$ 는 백색잡음으로 모형의 정상성을 위해  $d \in (-0.5, 0.5)$ 이다.

$$\Delta^d = (1 - B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k, \quad (2.1)$$

$$\phi_p(B)\Delta^d Z_t = \theta_q(B)\epsilon_t. \quad (2.2)$$

### 2.2. Double Seasonal Holt-Winters(DSHW) 모형

Taylor (2003)는 가법 Holt-Winters에서 두개의 계절 주기를 가지는 DSHW(Double Seasonal Holt-Winters) 모형을 제안하였다. 이에 대하여 오차의 AR(1)을 고려하였다. 이에 따라 1일 주기인 24와 1주일 주기인 168을 모형에 적용한다. DSHW 모형은 식 (2.3)과 같이 정의된다.

$$\begin{aligned} y_t &= l_{t-1} + b_{t-1} + s_t^{(1)} + s_t^{(2)} + d_t, \\ l_t &= l_{t-1} + b_{t-1} + \alpha d_t, \\ b_t &= \phi b_{t-1} + \beta d_t, \\ s_t^{(1)} &= s_{t-m_1}^{(1)} + \gamma_1 d_t, \\ s_t^{(2)} &= s_{t-m_2}^{(2)} + \gamma_2 d_t, \end{aligned} \quad (2.3)$$

여기서  $l$ 은 평활 수준,  $b$ 는 평활가법추세를 의미하며,  $s_t^{(1)}$ 는 첫 번째 주기에 대한 계절지표를 나타내며,  $s_t^{(2)}$ 는 두 번째 주기에 대한 계절지표이며  $\alpha, \beta$ 는 수준과 추세에 대한 모수이고  $\gamma_1, \gamma_2$ 는 계절평활 모수이며,  $\phi$ 는 오차의 AR(1)을 고려한 모수이다. 모수  $\alpha, \beta, \gamma, \delta, \phi$ 의 추정은 훈련자료에서의 1차 시점 후 적합에서의 SSE(Sum of Squared Error)를 최소화하는 과정을 통해 계산된다.

### 2.3. Trigonometric, Box-Cox transform, ARMA errors, Trend and Seasonal components(TBATS) 모형

De Livera 등 (2011)는 기존의 지수평활법들이 변수들간의 비선형성을 고려하기 어려우며, 주기성이 정수인 한계 및 패턴의 반복주기가 배수가 되어야 한다는 한계를 감안하여, 복잡한 주기성을 고려한 일변화된 모형으로서 BATS(Box-Cox transform, ARMA errors, Trend and Seasonal components) 모형과 이에 대하여 주기성을 삼각함수의 합으로 나타낸 TBATS(Trigonometric, Box-Cox transform, ARMA errors, Trend and Seasonal components) 모형을 제안하였다.

BATS 모형과 TBATS 모형은 명칭 그대로 식 (2.4)와 같이 Box-Cox 변환을 통하여 종속변수의 비선형성 관련 문제를 해결하며, 식 (2.5)와 같이 지수평활법을 통한 추세성분(trend components)과 계절성분(seasonal components)을 고려하였다. 식 (2.6)은 주기성을 삼각함수의 합으로 고려하는 것을 나타내며, 식 (2.7)에서 ARMA 모형을 통하여 오차항의 자기상관성을 고려한다. 이에 따라서 모형은 식 (2.8)과 같이 나타난다.

$$y_t^{(\omega)} = \begin{cases} \frac{y_t^\omega - 1}{\omega}, & \omega \neq 0, \\ \log y_t, & \omega = 0, \end{cases} \quad (2.4)$$

$$\begin{aligned} l_t &= l_{t-1} + \phi b_{t-1} + \alpha d_t, \\ b_t &= (1 - \phi)\bar{b} + \phi b_{t-1} + \beta d_t, \end{aligned} \quad (2.5)$$

$$\begin{aligned} s_t^{(i)} &= \sum_{j=1}^{k_i} s_{j,t}^{(i)} + \gamma_i d_t, \\ s_{j,t}^{(i)} &= s_{j,t-1}^{(i)} \cos \lambda_j^{(i)} + s_{j,t-1}^{*(i)} \sin \lambda_j^{(i)} + \gamma_1^{(i)} d_t, \end{aligned} \quad (2.6)$$

$$\begin{aligned} s_{j,t}^{*(i)} &= -s_{j,t-1}^{(i)} \sin \lambda_j^{(i)} + s_{j,t-1}^{*(i)} \cos \lambda_j^{(i)} + \gamma_2^{(i)} d_t, \\ d_t &= \sum_{i=1}^p \psi_i d_{t-i} + \sum_{i=1}^q \theta_i \epsilon_{t-i} + \epsilon_t, \end{aligned} \quad (2.7)$$

$$y_t^{(\omega)} = l_{t-1} + \phi b_{t-1} + \sum_{i=1}^T s_{t-m_i}^{(i)} + d_i. \quad (2.8)$$

$T$ 개의 주기성을 가정한 TBATS 모형은  $TBATS(\omega, \phi, p, q, \{m_1, k_1\}, \dots, \{m_T, k_T\})$ 로 표시하는데 여기서  $\omega$ 는 Box-Cox 변환 모수이고,  $\phi$ 는 Damping 모수,  $p$ 와  $q$ 는 ARMA 오차의 차수이다.  $\psi_i$ 와  $\theta_i$ 는  $ARMA(p, q)$ 의 계수이며, 단기적인 추세  $b_t$ 의 예측치는 장기적으로  $\bar{b}$ 로 수렴하여  $\bar{b}$ 는 장기 추세 값이다.  $\alpha, \beta$ 는 수준과 추세에 대한 모수이다.  $\gamma_1^{(i)}$ 와  $\gamma_2^{(i)}$ 는 평활(smoothing) 모수이며,  $\lambda_j^{(i)} = 2\pi/m_i$ ,  $k_i$ 는  $i$ 번째 주기성( $S_t^{(i)}$ )을 이루는 삼각함수의 수이며,  $S_{j,t}^{(i)}$ 는  $S_t^{(i)}$ 의 확률적 수준을 말하며,  $S_{j,t}^{*(i)}$ 는  $S_t^{(i)}$ 의 확률적 성장부문을 나타낸다.

### 3. 시계열 군집분석

시계열 군집분석은  $M$ 이 관측간의 모든 가능한 매칭형태라고 할 경우,  $r = ((X_{a_1}, Y_{b_1}), \dots, (X_{a_m}, Y_{b_m}))$ 가 매칭을 위한 정의로서 두 시계열 간의 거리를 최소화 시키며(이때,  $a_i, b_j \in \{1, \dots, T\}$ ) 집단간의 유사성에 따라 군집화 하는 것을 말한다.

### 3.1. 동적 시간 왜곡 거리(Dynamic Time Warping(DTW) distance)

모형에 제약 없이 접근하는 대표적 방법으로, Berndt와 Clifford (1994)에 의하여 제안된 방법으로 동적 시간왜곡(Dynamic Time Warping; DTW)은 두 시계열 간의 거리를 최소화하는 하는 방향으로 움직이면서 매칭시켜, 누적 거리를 계산하여 최소가 되는 집단들을 군집화한다.

$X_T = (X_1, \dots, X_T)^t$ 과  $Y_T = (Y_1, \dots, Y_T)^t$ 간의 근접성을 측정하는 간단한 방법으로 그 값의 근접도에 기초하여 사용하는 방법으로 DTW 거리는 식 (3.1)의 정의에 따라 나타난다.

$$d_{DTW}(X_T, Y_T) = \min_{r \in M} \left( \sum_{i=1, \dots, m}^T |X_{a_i} - Y_{b_i}| \right). \quad (3.1)$$

### 3.2. Periodogram 거리(Periodogram distance)

Caiado 등 (2006)는 Periodogram 기반의 방법으로  $I_{X_T}(\lambda_k) = T^{-1} |\sum_{t=1}^T X_t e^{-i\lambda_k t}|^2$ 과  $I_{Y_T}(\lambda_k) = T^{-1} |\sum_{t=1}^T Y_t e^{-i\lambda_k t}|^2$ 는  $X_T$ 와  $Y_T$ 의 Periodogram로서, 이때 진동수  $\lambda_k = 2\pi k/T$ ,  $k = 1, \dots, n$ 이다( $n = \lfloor (T-1)/2 \rfloor$ ). 이때, 값의 크기 보다는 상관 구조에 더 관심이 있을 경우에 정규화 Periodogram의 좌표 사이의 유클리드 거리를 사용하여 더 나은 결과를 얻을 수 있다.  $X_T$ 와  $Y_T$ 의 표본분산  $\hat{\gamma}_{0, X_T}$ ,  $\hat{\gamma}_{0, Y_T}$ 에 의하여  $NI_{X_T}(\lambda_k) = I_{X_T}(\lambda_k)/\hat{\gamma}_{0, X_T}$ ,  $NI_{Y_T}(\lambda_k) = I_{Y_T}(\lambda_k)/\hat{\gamma}_{0, Y_T}$ 는 정의된다. 이에 따라 표준화 거리의 정의는 식 (3.2)와 같다.

$$d_{NP}(X_T, Y_T) = \frac{1}{n} \sqrt{\sum_{K=1}^N (NI_{X_T}(\lambda_k) - NI_{Y_T}(\lambda_k))^2}. \quad (3.2)$$

### 3.3. 예측 거리(prediction distance)

예측기반의 방법은 예측 시점에서의 유사성 개념을 기반으로 군집화 하는 방법으로서,  $X_T$ 와  $Y_T$ 가  $X_t = \phi(X_{t-1}) + \epsilon_t$  형태로 나타날 경우 Vilar 등 (2010)은 식 (3.3)과 같은 거리를 제시하였다. 이때  $h$ 는 예측 시점으로서 본 논문은 전력 대비 시간으로서 6시점 후로 사용하였다.

$$d_{PRED, h}(X_T, Y_T) = \int \left| \hat{f}_{X_{T+h}}(u) - \hat{f}_{Y_{T+h}}(u) \right|. \quad (3.3)$$

## 4. 실제 자료 분석

본 연구에서 사용된 자료는 기업의 전력 사용량으로서 전력측정에 있어 결측이 없는 스마트그리드 환경하의 86개 기업을 대상으로 한다. 2014년 2월 12일 부터 2014년 4월 28일까지의 11주간의 시간별 자료를 사용하여 분석을 실시하였으며, 2014년 2월 12일 부터 2014년 4월 21일까지의 자료로 모형을 적합하였다. 적합한 모형을 사용하여 4월 22일부터 28일까지의 1주일을 대상으로 예측 정확도를 평가하였다.

모형 적합시 대하여 DSHW 모형과 TBATS 모형, FARIMA 모형을 적합하였으며, 각 군집별 MAPE가 가장 작게 나온 방법을 통하여 예측을 실시하여, 86개 기업 사용량 합계에 대한 예측 한 결과와 군집별 예측의 합계와의 비교를 실시하였다.

MAPE(Mean Absolute Percentage Error) 통계량은 식 (4.1)과 같이 나타난다. 이때,  $n$ 은 전체 기간,

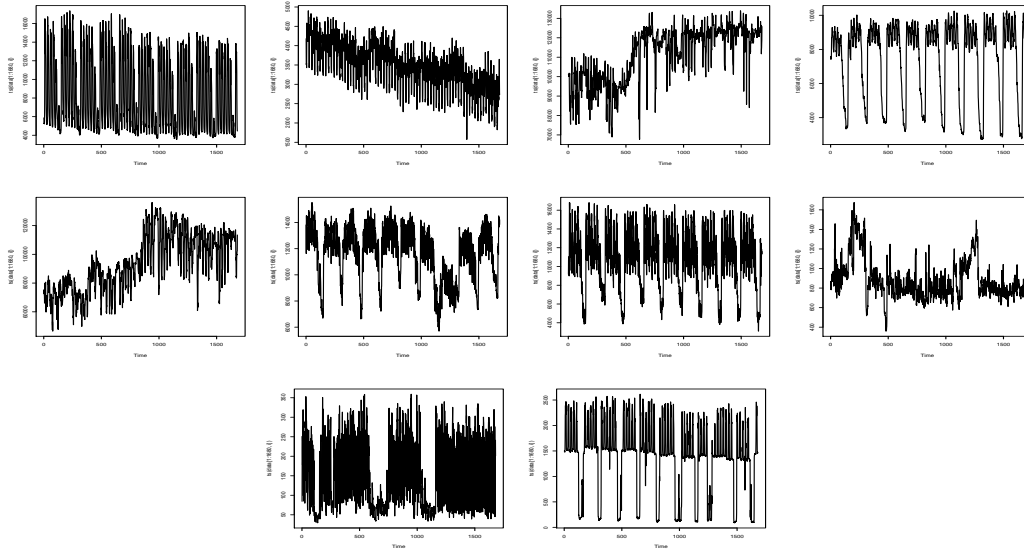


Figure 4.1. Time plots for Cluster1~Cluster10.

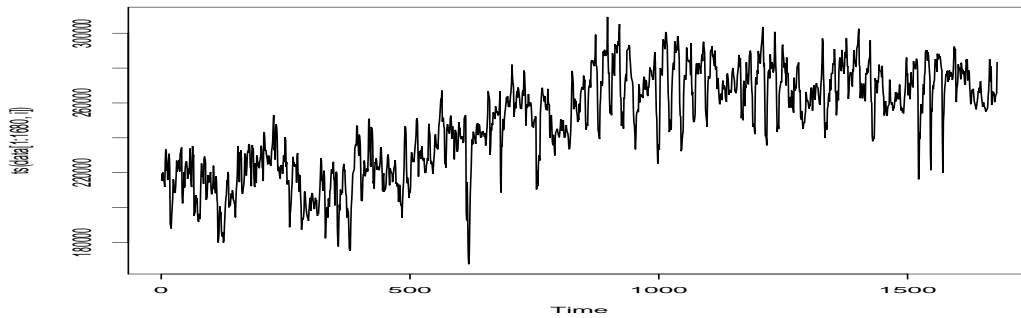


Figure 4.2. Time plot for total electricity demand.

$m$ 은 자료 적합 기간을 나타내며,  $Y_t$ 는 실제값,  $F_t$ 는 예측값을 나타낸다.

$$\text{MAPE} = \frac{1}{n-m} \sum_{t=m+1}^n \left| \frac{Y_t - F_t}{Y_t} \right|. \quad (4.1)$$

Figure 4.1은 정규화를 통한 Periodogram 기반 군집 분석 방법을 통한 10개 군집별 시계열도표이며, Figure 4.2는 전체 전력사용량의 시계열 도표이다. 이에 따라 각 군집들이 상이한 패턴을 보임을 확인할 수 있다. 또한 각 군집들이 주별 주기를 가지고 있는 것을 확인할 수 있다.

본 논문에서는 일별주기로 24와 주별 주기인 168을 고려하여 이중계절형 모델을 고려하였고, 이에 따라 본 연구에서는 전체 기업에 대하여 Table 4.1~Table 4.3은 TBATS 모형과 DSHW 모형, FARMIA 모형의 모수 추정 결과를 나타낸다.

**Table 4.1.** Parameter estimation of FARIMA

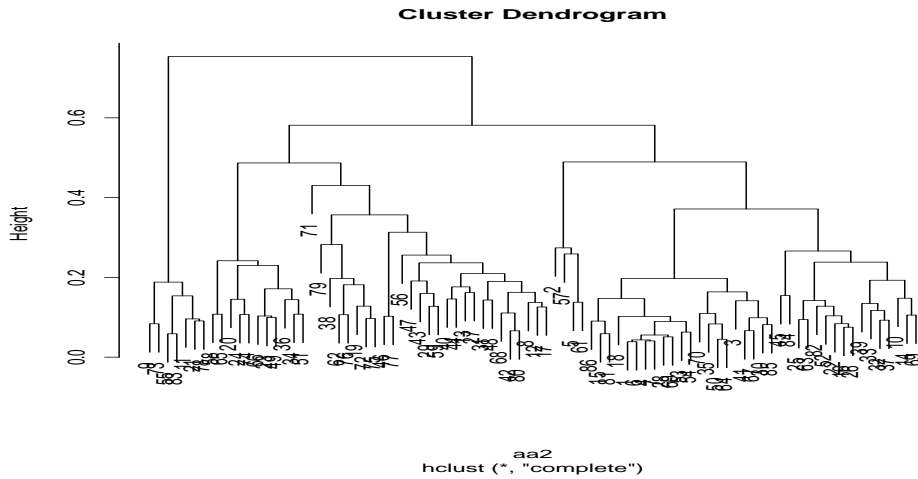
Parameter	Estimate	Parameter	Estimate
$\phi_1$	-0.2206	$\theta_1$	-1.1961
$\phi_2$	-0.2168	$\theta_2$	-1.2516
$\phi_3$	0.3206	$\theta_3$	-0.8860
$\phi_4$	0.5292	$\theta_4$	-0.1108
$\phi_5$	0.0900	추정된 차분치 $\hat{d}$	0.1258

**Table 4.2.** Parameter estimation of DSHW

Parameter	Estimate
Level( $\alpha$ )	0.8184
Trend( $\beta$ )	6.67E-08
SEASONAL1( $\gamma$ )	0.7804
SEASONAL2( $\delta$ )	0.9900
$\phi$	0.2728

**Table 4.3.** Parameter estimation of TBATS

Parameter	Estimate
Box-Cox( $\omega$ )	0.9617
Damping( $\phi$ )	0.8034
Level( $\alpha$ )	1.0615
Trend( $\beta$ )	-0.2336
$\gamma_1$	0.0153
$\gamma_2$	0.0010



**Figure 4.3.** Dendrogram for Periodogram-based clustering.

86개 기업에 대하여 군집분석 방법 중 정규화를 통한 Periodogram 기반 군집 분석 결과를 살펴보면, Figure 4.3에서 덴드로그램이 나타난다. Table 4.4에서는 10개 군집으로 구분 후 각 군집 별 3가지 모형을 훈련자료에 적용시의 MAPE를 산출결과이다. 이에 따라 FARIMA의 MAPE는 3번째 군집에

**Table 4.4.** MAPE of Training dataset by Periodogram-based Cluster

MAPE(%)	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	Total
FARIMA	11.19	5.27	2.26	2.42	3.51	5.10	10.51	5.77	31.34	61.81	2.08
DSHW	3.79	2.62	2.90	1.71	3.76	2.86	4.12	6.18	29.58	29.39	1.96
TBATS	4.86	2.98	2.26	2.17	3.49	3.20	6.36	5.43	39.19	34.70	1.99
obs.	21	4	17	7	16	6	11	2	1	1	86

MAPE = mean absolute percentage error, FARIMA = fractional ARIMA,

DSHW = double seasonal Holt-Winters,

TBATS = trigonometric, Box-Cox transform, ARMA errors, trend and seasonal components.

**Table 4.5.** Result of forecasting accuracy by MAPE

MAPE(%)	Total	$d_{DTW}$	$d_{Period}$	$d_{Pred}$
MAPE(%)	4.64	3.68	3.85	3.64

MAPE = mean absolute percentage error.

서 가장 작은 결과로 나타남에 따라 3번 군집은 예측에 있어서 FARIMA를 이용하여 예측을 실시하였으며, 1, 2, 4, 6, 7, 9, 10번째 군집에 경우는 DSHW 모형의 MAPE가 가장 낮은 값으로 나타나, DSHW 모형을 통하여 예측을 실시한다. 마지막으로 TBATS 모형을 통하여 5, 8번째 군집이 3가지 모형 중 가장 낮은 MAPE를 보여, TBATS를 통한 예측을 실시하였다. 전체 사용량에 대한 적용된 모형은 DSHW 모형이 가장 낮은 MAPE로서 선택되었다. 이에 따라 각 군집 및 전체 사용량에 대한 1주일의 예측 결과를 실시하여, Table 4.5에는 군집 분석법을 통하여 군집별 예측량들에 대하여 합산 예측값의 MAPE와 전체 사용량 자료를 통하여 예측결과의 MAPE가 제시된다. 전체 사용량의 예측에 대한 MAPE는 4.64%로 군집 별 예측을 통한 합산 예측 결과로서 DTW 군집분석을 통한 예측결과인 3.64%, Periodogram 기반의 군집분석을 통한 예측결과인 3.85%, 예측 기반 군집분석을 통한 결과인 3.85%가 더 높은 예측 정확도를 보였다. 이에 따라 군집분석을 통한 그룹별 예측이 전체 예측에 있어서 더 정확한 예측을 보여줌을 확인할 수 있었다.

## 5. 결론

본 연구는 스마트 그리드 하에서의 전력사용량 예측에 대하여 시계열 군집분석을 적용하여, 유사한 패턴을 보이는 자료간의 군집을 설정하였다. 군집분석 기법으로는 모형에 제약 없이 접근하는 대표적 기법으로 DTW 군집분석과 Periodogram 기반의 표준화 군집분석, 예측 기반의 군집분석을 사용하였으며, 각 군집들의 전력 사용량 예측에 있어서 시계열 모형으로는 DSHW 모형, TBATS 모형, FARIMA 모형을 사용하여 다중 계절성을 고려한 예측을 실시하였다.

예측평가 방법으로는 MAPE를 사용하였다. 각 군집별 전력 수요 예측량의 합산이 전체 전력 사용량을 통한 예측의 MAPE 보다 작음에 따라 시계열 군집분석을 통한 군집별 수요예측 방법이 더 높은 정확도를 보여주는 것으로 확인되었다.

이는 실제 스마트 그리드 방식에서 전력공급자와 소비자 간의 정보 교환을 통하여 전력 수요예측시 시계열 군집분석을 통하여 전력 생산에 있어 자원 낭비를 줄일 수 있는 방법으로 보여지며, 군집 별 특성에 따라 적합한 분석 모형의 선택이 가능할 것으로 보여진다.

향후 본 연구에 대한 확장 형태로 기업 정보의 확보를 통하여 군집 간 어떠한 특성이 있는지 확인을 하고자 하며, 군집 특성별 적합 모형의 대한 기준을 확보하고자 한다. 본 연구에 대하여 추가적으로 기업



전력 수요 예측 만이 아닌 전력 사용을 하는 여러 방법의 전력 수요 예측 방안을 고려되어야 할 것이며, 시계열 군집분석 방안에 대한 연구가 다양한 분야에서 연구 될 필요가 있다고 생각된다. 또한 실제 분야에 적용을 위한 자동화 방법들이 연구되어야 할 것으로 판단된다.

## References

- Bakhat, M. and Rossello, J. (2011). Estimation of tourism-induced electricity consumption: the case study of Balearics Islands, Spain, *Energy Economics*, **33**, 437–444.
- Berndt, D. J. and Clifford, J. (1994). Using dynamic time warping to find patterns in time series, In *KDD Workshop*, **10**, 359–370.
- Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (1976). Linear nonstationary models, *Time Series Analysis*, Fourth Edition, 93–136.
- Caiado, J., Crato, N., and Pena, D. (2006). A periodogram-based metric for time series classification, *Computational Statistics & Data Analysis*, **50**, 2668–2684.
- De Livera, A. M., Hyndman, R. J., and Snyder, R. D. (2011). Forecasting time series with complex seasonal patterns using exponential smoothing, *Journal of the American Statistical Association*, **106**, 1513–1527.
- Erdogdu, E. (2010). Natural gas demand in Turkey, *Applied Energy*, **87**, 211–219.
- Hong, H. W., Park, M. J., and Cho, S. S. (2009). Comparison study of time series clustering methods, *The Korean Journal of Applied Statistics*, **22**, 1203–1214.
- Hong, W. C. (2009). Electric load forecasting by support vector model, *Applied Mathematical Modelling*, **33**, 2444–2454.
- Hosking, J. R. (1981). Fractional differencing, *Biometrika*, **68**, 165–176.
- Hwang, H. M., Lee, S. H., Park, J. B., Park, Y. K., and Son, S. Y. (2015). Load forecasting using hierarchical clustering method for building, *The Transaction of the Korean Institute of Electrical Engineers*, **64**, 41–47.
- Jang, S. W., Park, Y. J., and Kim, G. Y. (2010). Time series pattern recognition based on branch and bound dynamic time warping, *Journal of KISS: Software and Applications*, **37**, 584–589.
- Jung, S. W. and Kim, S. (2014). Electricity demand forecasting for daily peak load with seasonality and temperature effects, *The Korean Journal of Applied Statistics*, **27**, 843–853.
- Kim, W. S. and Jeon, B. K. (2010). Overview of long-term electricity demand forecasting mechanism for national long-term electricity resource planning, *Transaction of Korean Institute of Electrical Engineers*, **59**, 1581–1586.
- Lee, H. N., Han, J. H., and Lee, M. H. (2010). Electricity peak equation: estimation and prediction, *Korean Energy Economic Review*, **9**, 83–99.
- Lee, H. R. and Shin, H. J. (2011). Electricity demand forecasting based on support vector regression, *IE Interfaces*, **24**, 351–361.
- Lee, J. S., Shon, H. G., and Kim, S. (2013). Daily peak load forecasting for electricity demand by time series models, *The Korean Journal of Applied Statistics*, **26**, 349–360.
- Lise, W. and Montfort, K. V. (2007). Energy consumption and GDP in Turkey: is there a cointegration relationship, *Energy Economics*, **29**, 1166–1178.
- Liu, J., Shu, Y., Zhang, L. and Xue, F. (1999). Traffic modeling based on FARIMA models, *IEEE Canadian Conference on Electrical and Computer Engineering*, 621–624.
- Montero, P. and Vilar, J. A. (2014). TSclust: An R package for time series clustering, *Journal of Statistical Software*, **62**, 1–43.
- Pao, H. T. (2009). Forecasting energy consumption in Taiwan using hybrid nonlinear models, *Energy*, **34**, 1438–1446.
- Park, S. S., Son, H. S., Lee, D. G., Ji, E. M., Kim, H.-S., and Ryu, K. H. (2009). Short-term power load forecasting using time pattern for u-city application, *Journal of Korea Spatial Information System Society*, **11**, 177–181.

- Park, W. G. and Kim, S. (2012). The performance of time series models to forecast short-term electricity demand, *Communications for Statistical Applications and Methods*, **19**, 869–876.
- Sari, R. and Soytas, U. (2004). Disaggregate energy consumption, employment and income in Turkey, *Energy Economics*, **26**, 335–344.
- Sozen, A. and Arcaklioglu, E. (2007). Prediction of net energy consumption based on economic indicators (GNP and GDP) in Turkey, *Energy Policy*, **35**, 4981–4992.
- Taylor, J. W. (2003). Short-term electricity demand forecasting using double seasonal exponential smoothing, *Journal of the Operational Research Society*, **54**, 799–805.
- Taylor, J. W. (2010). Triple seasonal methods for short-term load forecasting, *European Journal of Operational Research*, **204**, 139–152.
- Taylor, J. W., De Menezes, L. M., and McSharry, P. E. (2006). A comparison of univariate methods for forecasting electricity demand up to a day ahead, *International Journal of Forecasting*, **22**, 1–16.
- Vilar, J. A., Alonso, A. M., and Vilar, J. M. (2010). Non-linear time series clustering based on non-parametric forecast densities, *Computational Statistics & Data Analysis*, **54**, 2850–2865.
- Wang, J., Zhu, W., Zhang, W., and Sun, D. (2009). A trend fixed on firstly and seasonal adjustment model combined with the  $\epsilon$ -SVR for short-term forecasting of electricity demand, *Energy Policy*, **37**, 4901–4909.

# 스마트 그리드에서의 시계열 군집분석을 통한 전력수요 예측 연구

손흥구<sup>a</sup> · 정상욱<sup>a</sup> · 김삼용<sup>a,1</sup>

<sup>a</sup>응용통계학과, 중앙대학교

(2015년 12월 17일 접수, 2015년 12월 30일 수정, 2015년 12월 30일 채택)

---

## 요약

본 논문은 ICT기반 시장에서의 수요관리시스템에서의 핵심 요소인 전력 수요 예측을 위하여, 전체 사용량을 기반으로 예측 하는 방식이 아닌, 시계열 기반 군집분석을 통한 군집별 예측량의 결합을 실시하였다. 시계열 군집 분석 방법으로서 Periodogram 기반의 정규화 군집분석, 예측 기반의 군집분석, DTW(Dynamic Time Warping)를 이용하여 군집화를 시도하였으며, 군집 별 수요예측 모형으로서 DSHW(Double Seasonal Holt-Winters) 모형, TBATS(Trigonometric, Box-Cox transform, ARMA errors, Trend and Seasonal components) 모형, FARIMA(Fractional ARIMA) 모형을 사용하여 예측을 실시하였다. 전체 사용량을 기반으로 예측 하는 방식이 아닌, 군집분석을 통한 군집별 예측량의 결합이 더 낮은 MAPE로 나타남에 따라 우수한 예측 방법으로 판단되었다.

Keywords: 시계열군집분석, TBATS 모형, DSHW 모형, 전력수요예측, Dynamic Time Warping

---

이 논문은 2013년도 중앙대학교 연구장학기금 지원에 의한 것임.

<sup>1</sup>교신저자: (06974) 서울 동작구 흑석동 221, 중앙대학교 응용통계학과. E-mail: sahm@cau.ac.kr