

시간단위 전력수요자료의 함수적 군집분석: 사례연구[†]

윤상후¹ · 최영진²

^{1,2}한국의국어대학교 차세대도시농림융합기상사업단

접수 2015년 5월 6일, 수정 2015년 6월 30일, 게재확정 2015년 7월 1일

요약

전력시스템의 안정적이고 효과적인 운영을 위해선 전력수요예측이 필요하다. 본 연구에서는 일별 전력수요패턴의 시간에 따른 커브를 군집분석 하려고 한다. 2009년 1월 1일부터 2011년 12월 31일 까지의 일별 시간단위 전력수요 자료는 추세성분 제거와 로그변환을 통해 계절성분과 오차성분으로 구성된 시계열자료로 변환되었다. 변환된 자료는 Ma 등 (2006)이 제안한 함수적 군집모형을 사용하여 분석되었고, 모수는 EM알고리즘과 일반화교차검정을 통해 추정되었다. 군집의 수는 휴일과 평일을 잘 분류하는 10개로 결정하였다. 분석결과 월요일, 평일 (화요일~금요일), 토요일, 일요일 또는 공휴일과 계절요인으로 전력수요 평균곡선이 설명된다. 함수적 군집분석을 통한 전력수요패턴의 과학적인 분류는 향후 단기전력수요예측에 활용된다.

주요용어: 군집의 수, 시계열자료, 전력수요, 함수적 군집분석.

1. 서론

전력수요 예측은 전력시스템의 안정적이고 경제적인 운영을 위해 필수적이다. 최근 에너지 효율향상에 대한 관심은 전력의 공급 측면보다 수요 측면에서의 관리를 강조하고 있다. 전력부하관리시스템인 에너지저장장치 (energy storage system; ESS)와 에너지관리시스템 (energy management system; EMS) 등을 예 들 수 있다 (Moon, 2013; Lee와 Lee, 2013; Kwak 등, 2013; Lee 등, 2013). 즉, 전력수요 예측모형의 정확도는 에너지 효율성과 연관이 있으므로 이를 향상시키기 위해선 통계적, 수리적 모형이 필요하다.

전력수요예측 기법은 매우 다양하게 발전되어 왔다. 수요 예측은 크게 과거 시계열적인 특성을 반영한 연구와 기상과 같은 외부요인과의 관계를 규명을 위한 연구로 나눌 수 있다 (Amjadi, 2001; Weron, 2006). Hong 등 (2011), Song 등 (2005), Lee 등 (2014)은 과거 시계열적 특성을 활용한 이중계절 지수평활법, 삼중계절 방식, ARIMA 모형 등을 통해, Jeong과 Kim (2014), Song 등 (2014), Taylor (2010)은 온도와 같은 기상요소를 고려한 모형을 통해 전력수요를 예측하였다. 최근에는 군집분석을 기반 전력수요예측이 제안되고 있다 (Cho 등, 2013).

군집분석을 이용한 전력수요예측은 시간단위 전력수요패턴이 일별로 다르다는 가정에서 시작한다. Song 등 (2013)은 크게 주중, 주말, 특수일, 기타로 그룹으로 세분화하여 전력수요를 예측하였고,

[†] 본 연구는 기상청 차세대도시농림융합스마트기상서비스개발 (WISE) 사업의 지원으로 수행되었음 (KMIPA-2012-0001).

¹ 교신저자: (449-791) 경기도 용인시 처인구 모현면 외대로 81, 한국의국어대학교 차세대도시농림융합기상사업단, 선임연구원. E-mail: statstar@hufs.ac.kr

² (449-791) 경기도 용인시 처인구 모현면 외대로 81, 한국의국어대학교 차세대도시농림융합기상사업단, 단장.

Hwang 등 (2015)은 계층적 군집분석을 통해 건물의 전력수요패턴을 8개의 군집으로 나누어 접근하였다. Lim (2014)은 K -평균 군집분석을 통해 특수일 최대전력수요예측을 9개의 군집으로 나누어 분석하였다. 이 외에 시계열 자료의 군집분석에 관한 연구로 Kim 등 (2015)과 Yeo (2011) 등의 연구가 있다.

시계열자료의 함수적 자료접근은 자료의 정보 손실을 최소화하면서 통계적인 분석이 가능해 보다 정확한 연구를 할 수 있다 (Lee와 Lee, 2012). 시간단위 전력수요량 자료는 곡선이나 함수의 형태로 나타나므로 함수자료 군집분석을 실시하고자 한다. 함수자료란 시간이나 공간 위에 정의된 함수의 형태로 관측된 자료를 의미하며, 본 연구에서는 시간단위로 관측된 일별 전력수요량이 이에 해당한다 (Ramsey와 Silverman, 2005). 함수자료의 분석은 자료를 적절하게 연속된 함수의 형태로 변환하여 접근하고, 이를 통해 중요한 특징이 잘 반영되도록 한다 (Lee 등, 2014). 함수자료의 군집화에서 중요한 사항은 적절한 군집 수 결정과 정확한 군집 결과이다.

본 논문의 구조는 2절에서 Ma 등 (2006)이 제안한 함수적 군집모형과 본 연구에서 군집 수를 결정하는 방법을 살펴보고 3절에서 시간단위 전력수요 자료에 적용시켜보고 각 군집의 특징 및 결속력을 확인한다. 마지막으로 4절에서 연구 결과 및 연구 한계점을 정리한다.

2. 함수자료 군집분석 방법

이번 절에서는 Ma 등 (2006)이 제안한 시간 함수적 자료를 위한 군집모형에 대해 설명하고자 한다.

$g(t_i)$ 를 t 시점에서의 i 번째 개체 (곡선)의 참값이라 하자. t_i 에서의 전력 수요량 Y_i 는 시간 t_i 에 따라 변하므로 평활함수를 이용하여 i 번째 개체 (곡선) 값을 식으로 표현하면

$$Y_i = f(t_i) + \epsilon_i, \quad i = 1, \dots, n \quad (2.1)$$

여기서 n 은 개체의 수를 나타내고 ϵ_i 는 평균이 0이고 분산이 σ^2 인 가우시안분포(Gaussian distribution)를 따른다. 일반적인 평활함수는 잔차제곱합 (residual sum of squares; RSS)을 최소화하는 함수이다.

$$RSS = \sum_{i=1}^T [y_i - f(t_i)]^2 \quad (2.2)$$

모든 관측지점을 지나는 많은 함수 중 $f(t_i)$ 의 모수 과적합을 피하기 위해 미리 지정한 기본 함수 (basis function)를 이용한다. 조금 더 유연한 방법으로 다음과 같은 평활 조건을 채용할 수 있고, 여기서 η 는 관측지점을 부드럽게 설명하는 기본 함수의 평활정도를 의미한다.

$$\sum_{i=1}^T [f''(t_i)]^2 dt < \eta. \quad (2.3)$$

라그랑지 방법 (Lagrange multiplier method)에 의해, 식 (2.3)를 만족하면서 식 (2.2)을 최소로 값은 식 (2.4)을 최소로 하는 값이다.

$$\sum_{i=1}^T [y_i - f(t_i)]^2 + \lambda T \int [f''(t_i)]^2 dt_i \quad (2.4)$$

식 (2.4)의 곡선은 cubic smoothing spline으로 알려져 있다.

본 연구에서는 시간에 따른 전력수요량 자료에 대한 모형을 함수적 자료 기반으로 군집화하고자 한다. 이를 위해서 평활곡선으로 표현되는 각 군집의 평균 곡선을 모형화 해야 한다. 특정일 j , 시간대 i , 군집

k 에 대해 t_{ij} 의 관측된 시간단위 전력수요량은 다음과 같다.

$$y_{ij} = \mu_k(t_{ij}) + b_j + \epsilon_{ij}, \quad (2.5)$$

여기서 μ_k 는 평균 곡선이고 b_j 는 특정일의 분포를 의미하고 평균이 0이고 분산이 σ_{bk}^2 인 가우시안분포를 따르고, ϵ_{ij} 는 평균이 0이고 분산이 σ^2 인 가우시안분포를 따른다 ($y_j \sim N(\mu_k, \sum_k)$). 시간대별 전력수요량 군집은 일반적으로 알려져 있지 않으므로 시간대별 y_i 에 대한 혼합가우시안분포는

$$y_i \sim p_1 N(\mu_1, \sum_1) + p_2 N(\mu_2, \sum_2) + \dots + p_k N(\mu_k, \sum_k)$$

로 적어진다. 여기서 k 는 총 군집의 수를 의미하고 p_1, p_2, \dots, p_k 는 군집의 상대적 크기를 의미한다. 만약 k 군집이 n 개 일의 시간대별 자료로 구성된다면, 혼합효과모형의 penalized 로그-우도함수는 다음과 같다:

$$\sum_{i=1}^n \log p_{J_i} - \sum_{i=1}^n \sum_{j=1}^T \left[\frac{(y_{ij} - \mu_k(t_{ij}) - b_i)^2}{2\sigma^2} \right] - \sum_{i=1}^n \frac{b_i^2}{2\sigma_{bk}^2} - \lambda_k T \int [\mu_k''(t)]^2 dt + C. \quad (2.6)$$

Penalized 로그 우도함수를 최대화 하는 모수를 수학적으로 계산할 수 없으므로, EM 알고리즘과 Generalized Cross-Validation (GCV)를 이용하여 모수를 추정할 수 있다. EM 알고리즘의 E 스텝은 특정일이 모든 모수들로 구성된 모형에 속할 확률로 다음과 같다:

$$P(\text{day}_i \in k) = \frac{p_k N(\mu_k, \sum_k)}{p_1 N(\mu_1, \sum_1) + \dots + p_K N(\mu_K, \sum_K)}. \quad (2.7)$$

EM 알고리즘의 M 스텝은 각각 군집의 penalized 로그-우도함수의 가중치 (2.8)를 최대화 하면 된다:

$$- \sum_{k=1}^K \left\{ \sum_{i=1}^n P(\text{day}_i \in k) \left(\sum_{j=1}^T \frac{(y_{ij} - \mu_k(t_{ij}) - b_i)^2}{2\sigma^2} + \frac{b_i^2}{2\sigma_{bk}^2} \right) - \lambda_k T \int [\mu_k''(t)]^2 dt + C \right\}. \quad (2.8)$$

특정 군집의 추정된 전력수요량 (\hat{y}_{ij})은 관측된 전력수요량들의 선형결합으로 표현되고, a_{ijlm} 을 적절히 배치함으로써 얻어지는 행렬을 평활행렬 (smoothing matrix)라 한다.

$$\hat{y}_{ij} = \sum_{l=1}^n \sum_{m=1}^T a_{ijlm} y_{lm}.$$

σ_{bk}^2 과 λ_k 는 leave-one-out 교차검증을 기반으로한 GCV를 통해 계산된다. Gu와 Ma (2005)에 따르면 실제값과 추정치 차이가 점근적으로 최소화된다고 알려져 있기 때문이다. σ^2 는 자료의 RSS로 추정될 수 있다. 군집의 가중치를 의미하는 모수 p_k 는 다음과 같이 업데이트 된다:

$$p_k = \frac{\left[\sum_{i=1}^n P(\text{day}_i \in k) + a_k \right]}{\left(n + \sum_{k=1}^K a_k \right)}. \quad (2.9)$$

이 단계는 수렴할 때까지 계속 반복 계산된다.

EM 알고리즘 중 M 스텝은 모든 자료의 가중치로 구성된 군집들의 합을 계산해야하므로 그 계산시간이 매우 비싸고 안정적이지 못하다. 이 문제를 해결하기 위해 Liu 등 (1998)의 Rejection-controlled EM (RC-EM) 알고리즘을 이용하였다. 먼저 day 와 군집의 관계 확률에 c (예, $c = 0.05$)라는 문턱치를

설정한다. 만약 일자료의 군집별 확률이 c 보다 낮다면 0으로 배정한다. 그러면 M 스텝에서 많은 값들이 0으로 배정되어 합을 계산하는 시간을 줄여주게 된다. 만약 $c = 0$ 이면, RCEM 알고리즘은 EM 알고리즘과 같게 되고, $c = 1$ 이면 Wei와 Tanner (1990)가 제안한 Monte carlo EM algorithm과 같게 된다. 따라서 제안된 알고리즘은 M 스텝의 계산시간을 줄이면서 E 스텝이 수렴하도록 만들 수 있다. 하지만, 국소 최적해 (local optima)로 빠지는 문제가 발생할 수 있으므로 여러 개의 체인을 이용한다. 이 단계를 간단히 정리하면 다음과 같다.

E step

1. day_i 가 각 군집에 속할 사후확률을 계산한다.
2. 사후확률을 기반으로 day_i 를 군집 k 로 배정한다.

M step

1. day_i 의 군집의 평균 곡선과 다른 모수들을 업데이트하여 우도값을 최대로 한다.
2. GCV를 통해 최적 평활모수 λ 와 σ_{bk}^2 를 추정한다.

E와 M step을 수렴할때까지 반복 수행한다.

3. 자료 분석

Ma 등 (2006)이 개발한 R의 'MFDA' 패키지를 이용하여 2009년 1월 1일부터 2011년 12월 31일까지 한국 전력거래소에서 수집된 1,095일의 시간단위 전력수요 자료를 분석하였다. 시간에 따른 전력수요 원자료 시계열은 Figure 3.1과 같다. 전력수요자료는 추세성, 계절성을 보이고 있고 설 연휴와 추석 연휴에는 정규적인 계절패턴과 다르게 갑자기 전력수요가 떨어지는 패턴을 보인다. 또한 추세와 함께 시간이 흐름에 따라 자료의 변동 폭이 커짐을 알 수 있다. 시간의 흐름에 따라 자료의 변동 폭이 증가하므로 로그변환을 실시하였고, 추세성을 제거시키면 Figure 3.2와 같다. 계절, 요일 외에 추세성분도 군집분석에 중요한 요인으로 분석되어 본 연구는 로그변환과 추세성을 제거시킨 자료를 토대로 함수적 군집분석이 실시하여 시간에 무관한 계절성만 반영된 결과를 살펴보고자 한다.

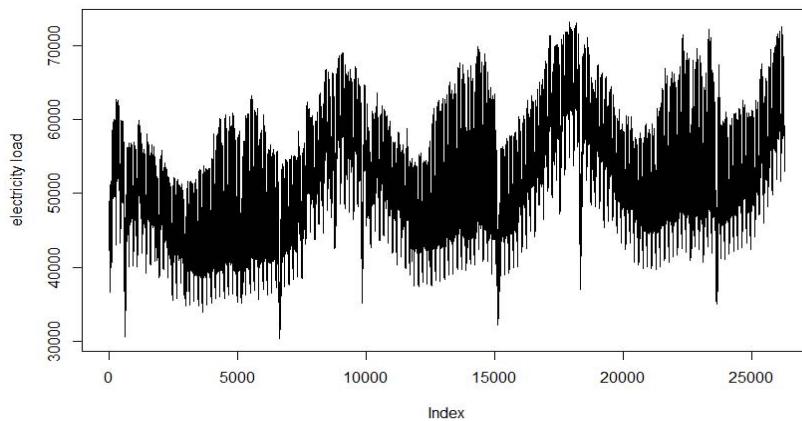


Figure 3.1 Original time series data

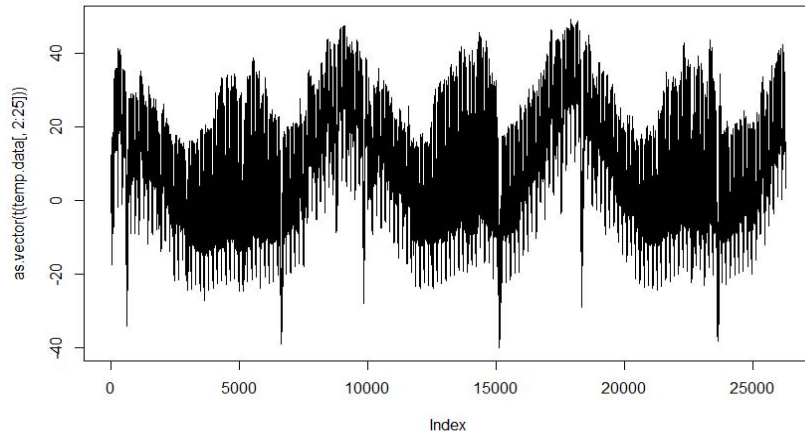


Figure 3.2 Time plot without trend after square root transformation

군집분석의 어려움 중 하나는 군집의 수의 결정이다. 아직까지 최적 군집수를 선정하는 방법은 알려져 있지 않다. Hwang 등 (2015)은 24시간 자료에 대해 계층적 군집분석을 실시하여 총 8개의 수요패턴을 구분하였고, 이를 토대로 수요예측패턴을 예측하였다. Ma 등 (2006)은 likelihood와 자료의 수를 기반으로 한 BIC로 군집의 수를 제안하였다. BIC는 적합성과 모형의 복잡성을 모두 고려하는 장점을 지니고 있으나, 본 자료의 경우 군집의 수가 증가할수록 BIC가 낮아진다 (Table 3.1, Figure 3.3). 즉, 시간별 전력수요 패턴을 적은 수의 군집으로 세분화하기엔 BIC가 적절하지 않다. 본 연구에서는 7개부터 25사이의 군집 결과와 공휴일, 주말 등의 관계를 통해 군집 수를 결정하였다.

Song 등 (2005)에 따르면 전력수요는 사회적 이벤트에 의하여 다양한 특징을 나타내고, 일별 24시간 전력 수요 패턴은 크게 평일, 주말, 특수일, 기타 총 4가지 패턴으로 구분하였다. 일반적으로 월요일은 다른 평일인 화요일 금요일 시간단위 패턴에 비해 밤에서 낮까지의 전력수요가 낮은 특성을 보인다. 이는 주말 전력수요 영향이 월요일 오전까지 미치기 때문이다. 주 5일제의 영향으로 토요일은 휴일이지만, 주 5일제가 실시되지 않은 기업이나 학교의 영향으로 낮 시간 전력수요가 발생한다. 일요일은 토요일에 비하여 더욱 많은 전력수요가 감소하게 된다. 여름에는 낮 시간의 높은 기온으로 인한 냉방장비의 전력수요가 증가하지만 겨울에는 밤 시간의 낮은 기온으로 인한 난방장비의 전력수요가 증가한다. 특수일은 설날, 추석, 삼일절과 같은 공휴일과, 공휴일과 주말 사이의 평일 (징검다리 날)로 구성되어 있다. 공휴일은 일요일과 비슷한 패턴을 보이지만 추석, 설날과 같은 3일 이상의 공휴일은 상대적으로 낮 시간 전력수요가 낮다. 반면 징검다리 날은 일요일과 평일의 패턴이 복합적으로 나타난다.

이제 군집분석에서 세분화된 군집이 Song 등 (2005)의 분류기준 (휴일, 요일, 월)으로 어떻게 표현되는지 살펴보자. 의사결정나무를 통해 분석한 결과 중요한 요인은 휴일여부, 요일, 월 순으로 그 결과는 Figure 3.5와 같다. 일 전력패턴은 사회적 이벤트에 영향으로 받고 이를 사회적 이벤트를 구성하는 요인으로 분류했을 때 의사결정나무로 예측된 군집과 실제 군집이 상이한 오분류된 경우는 Table 3.1과 같다. 군집의 수를 7부터 25까지 증가시키며 오분류 일 수를 계산하면 군집의 수가 10일 때 35일로 오분류 일 수가 가장 낮았다. 이에 본 연구의 군집의 수는 Song 등 (2005)의 분류기준으로 분류하였을 때, 오분류된 일 수가 가장 낮은 10개로 결정하였다.

Table 3.1 The selection of the number of cluster

k	7	8	9	10	11	12	...	25
BIC (unit:1,000)	129.6	127.3	123.7	121.0	119.7	118.6	...	105.72
Misclassification	90	39	38	35	38	47	...	40

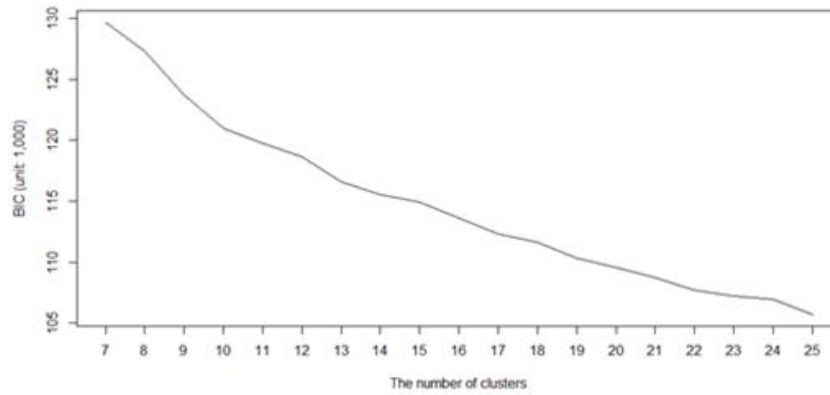


Figure 3.3 The BIC values based on the number of clusters

군집의 수를 10으로 했을 경우, 함수적 군집분석 결과는 Table 3.2와 같다. 군집의 결집력은 각 군집의 2.5%와 97.5%에 해당되는 분위수 사이의 거리와 R^2 을 통해 살펴보았다. R^2 가 높은 군집일수록 결집력이 높다고 볼 수 있다. 10개의 군집으로 할당된 각 군집 비율 범위는 4.8%~18.9%이고 R^2 의 범위는 0.696에서 0.993이다. 0.8을 기준으로 군집의 결속력을 살펴보면 군집 4의 결속력이 낮다고 볼 수 있다. 또한 각 군집의 2.5%와 97.5% 분위수 사이의 거리를 통해 각 군집을 구성하는 케이스의 결속력을 시각적보면 Figure 3.4와 같다. R^2 결과와 비슷하게 R^2 가 높을수록 시간대별 95% 범위가 짧고, R^2 가 낮을수록 시간대별 95% 범위가 길었다.

Table 3.2 The result of functional clustering

Cluster	1	2	3	4	5	6	7	8	9	10
n	53	81	111	82	106	118	113	144	80	207
%	4.8%	7.4%	10.1%	7.5%	9.7%	10.8%	10.3%	13.2%	7.3%	18.9%
R^2	0.875	0.888	0.908	0.696	0.966	0.996	0.974	0.977	0.993	0.966

각 군집의 케이스를 토대로 군집 유형을 휴일별, 요일별, 계절별로 정리하면 Table 3.3과 Figure 3.5이다. 유형 구분은 선행연구 결과와 비슷하지만 계절에 따라 군집분석 결과가 세분화되었다. Figure 3.5의 각 군집의 특징을 살펴보면, 토요일에 해당되는 군집은 군집 3, 군집 4, 군집 10이다. 여름 (군집 3)에는 냉방기기의 사용으로 인해 낮 시간대에 높은 전력수요량이 발생되고, 봄과 가을 (군집 10)에는 여름에 비해 상대적으로 낮 시간대의 전력수요량이 낮다. 반면 겨울 (군집 4)에는 밤 시간대의 전력수요량이 높게 나타나고 있다. 일요일 또는 공휴일에 해당되는 군집은 군집 1, 군집 2, 군집 3으로 토요일에 비해 업무시간 (6시 20시)사이의 전력수요량이 상대적으로 낮다. 일요일의 여름과 봄, 가을 패턴은 비슷하지만 겨울은 전혀 다른 패턴을 보인다. 월요일에 해당되는 군집은 군집 9와 군집 10으로 자정부터 새벽시간 (0시 6시) 사이의 전력수요량 패턴은 일요일 또는 공휴일과 비슷하지만 이후에 전력수요가 가파르게 올라가고 있다. 일요일과 비슷하게 봄, 여름, 가을 패턴은 비슷하지만 겨울의 패턴은 상이하 다. 물론 여름철 기온이 높은 시간대의 전력수요량이 상대적으로 높다. 평일 (화요일 금요일)에 해당되는 군집은 군집 5, 군집 6, 군집 7, 군집 8이다. 겨울은 난방기의 영향으로 밤낮 시간 관계없이 전력수요가 높고 여름은 냉방기의 영향으로 낮 시간대에 전력수요가 높다. 봄과 가을을 구분하는 군집 5와 군집 8은 패턴이 비슷하지만 군집 5는 밤 시간대의 낮은 기온으로 인해 군집 8에 비해 상대적으로 22시 12시 까지의 전력수요가 높았다.

Table 3.3 The characteristic of cluster based on weekday and month

type \ month	1	2	3	4	5	6	7	8	9	10	11	12
Sunday or holiday	2				1	3			1		2	
Saturday	4								3		10	
Tuesday to Friday	7	5				8		6		8		7
Monday	10								9		10	

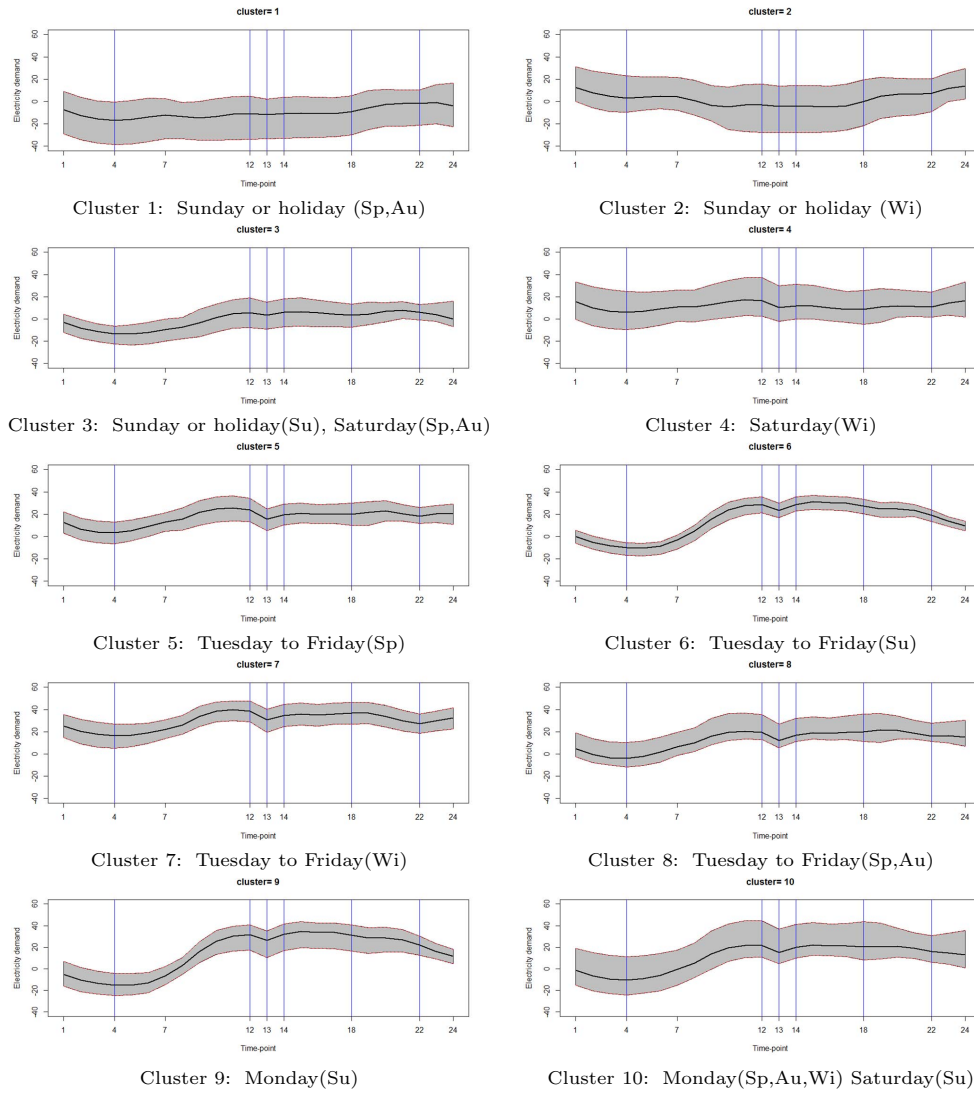


Figure 3.4 The mean curve and 95% confidence interval for each cluster)
(Sp: Spring, Su: Summer, Au: Autumn, Wi: Winter)

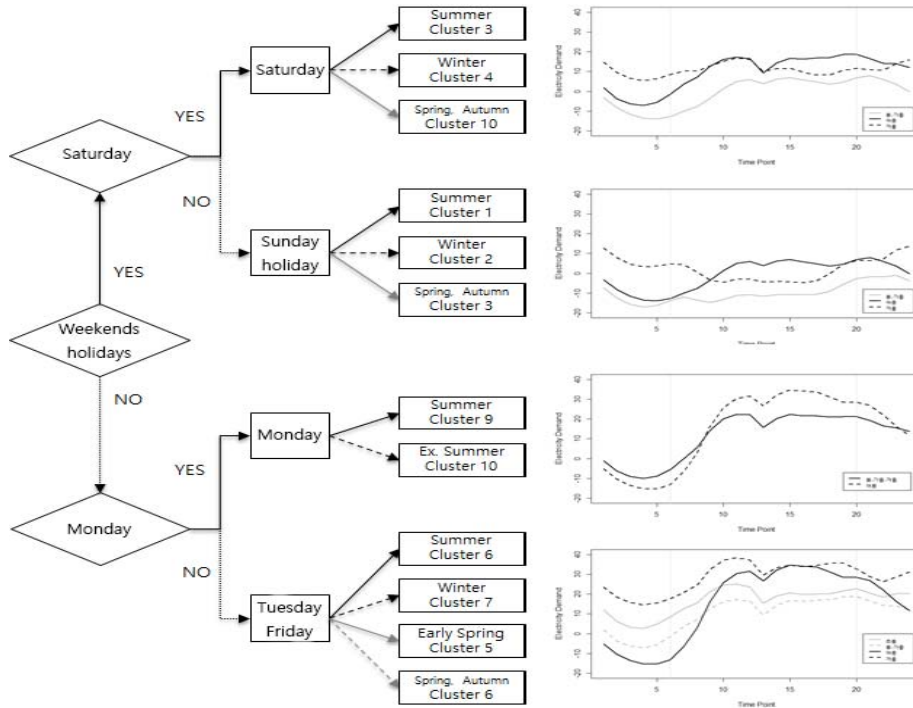


Figure 3.5 Decision tree for the result of functional clustering

4. 결론

계절과 요일에 따라 시간단위 전력수요패턴이 다르므로, 군집분석을 통한 시간단위 전력수요패턴 파악은 단기전력 수요예측에 필수적이다. Song 등 (2005)은 전력수요패턴을 평일, 주말, 특수일, 기타 이렇게 4가지로 구분하였지만, 본 연구에서 추계성분이 제거된 일별 시간단위 전력수요량을 함수적 군집분석을 진행하였다. 군집의 수는 휴일여부와 함수적 군집분석 결과를 토대로 10개로 결정하였으며, 각 군집은 요일성분과 계절성분을 토대로 각각 군집별 특징을 기술하였다. 분석결과 휴일여부와 계절은 시간단위 전력수요량에 중요 요소로 볼 수 있다.

본 연구의 한계점을 기술하면 다음과 같다. 첫째, 계절과 요일에 따른 군집 유형을 구분하였으나 (Table 3.3), 봄, 가을철은 기온, 날씨와 같은 외부요인 인한 냉방기기 및 난방기기 전력수요가 발생되어 월 단위로 구분하는데 어려움이 있다. 기온 또는 날씨와 같은 외부요인은 군집분석 결과의 직접적인 영향을 미치므로, 외부요인과 군집분석 결과를 자세히 살펴볼 필요성이 있다. 본 연구에서는 생략되었지만 평균기온, 최저기온, 최고기온에 대해 군집 간 평균차이를 분석하면 군집 9 < 군집 3 < 군집 8 < 군집 4 < 군집 10 < 군집 6 ≤ 군집 1 < 군집 2 < 군집 7 ≤ 군집 5 순이다. 둘째, 전력수요패턴은 전력수요예측을 위한 기초 연구이므로 함수적 군집분석과 시계열분석을 통해 예측성능에 대한 평가가 필요하다. 셋째, 함수적 군집분석의 계산시간이 계층적 군집분석 또는 K-평균 군집분석에 비해 상대적으로 비싸다. 향후 대용량 자료의 효율적인 군집분석을 위해선 알고리즘의 개선이 필요하다.

References

- Amjady, N. (2001). Short-term hourly load forecasting using time-series modeling with peak load estimation capability. *IEEE Transactions on Power Systems*, **16**, 498-505.
- Cho, H., Goude, Y., Brossat, X. and Yao, Q. (2013). Modeling and forecasting daily electricity load curves: a hybrid approach. *Journal of the American Statistical Association*, **108**, 7-21.
- Gu, C. and Ma, P. (2005). Optimal smoothing in nonparametric mixed-effect models. *Annals of statistics*, **33**, 1357-1379.
- Hong, D. H., Lee, S. H. and Do, H. Y. (2011). Fuzzy linear regression analysis for fuzzy input-output data using shape preserving operations. *Fuzzy Sets and Systems*, **122**, 513-526.
- Hwang, H. M., Lee, S. H., Park, J. B., Park, Y. G. and Son, S. Y. (2015). Load forecasting using hierarchical clustering method for building. *The Transactions of the Korean Institute of Electrical Engineers*, **64**, 41-47.
- Jeong, S. W. and Kim, S. Y. (2014). Electricity demand forecasting for daily peak load with seasonality and temperature effects. *The Korean Journal of Applied Statistics*, **27**, 843-853.
- Kim, H. J., Kwak, H. R., Kim, Y. N. and Choi, Y. J. (2015). A spectrum based evaluation algorithm for micro scale weather analysis module with application to time series cluster analysis. *Journal of the Korean Data & Information Science Society*, **26**, 41-53.
- Kwak, Y. H., Jeon, S. H. and Heo, J. H. (2013). Application of EMS based simulation for potential of energy saving during the cooling season. *Journal of the architectural institute of Korea : Planning & design*, **29**, 255-262.
- Lee, J. I. and Lee, I. W. (2013). Energy storage system (ESS) technologies and standardization for smart grid. *Communications of the Korean Institute of Information Scientists and Engineers*, **31**, 33-42.
- Lee, M. A., Lim, Y. H., Park, C. G. and Lee, K.E. (2014). Functional clustering for clubfoot data: A case study. *Journal of the Korean Data & Information Science Society*, **25**, 1069-1077.
- Lee S. M. and Lee K. E. (2012). The study on unemployment rate variation using functional data analysis. *Proceedings of excellent graduate student paper (Statistics Korea)*, 57-80.
- Lee, Y. S., Heo, J. and Choi, Y. H. (2013). A study for space-based energy management system to minimizing power consumption in the big data environments. *The Journal of the Institute of Internet Broadcasting and Communication*, **13**, 229-235.
- Lim, H. W. (2014). Special Day Maximum Load Forecast Using K-means cluster, Master Thesis, Uiduk University, Kyeongju.
- Liu, J. S., Chen, R. and Wong, W. H. (1998). Rejection control and sequential importance sampling. *Journal of the American Statistical Association*, **93**, 1022-1031.
- Ma, P., Castillo-Davis, C. I., Zhong, W. and Liu, J. S. (2006). A data-driven clustering method for time course gene expression data. *Nucleic Acids Research*, **34**, 1261-1269.
- Moon, S. I. (2013). Utilization and economic analysis of energy storage system (ESS). *Journal of electrical world*, **442**, 44-48.
- Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data analysis*, Springer, New York.
- Song, K. B., Baek, Y. S., Hong, D. H. and Jang, G. S. (2005). Short-term load forecasting for the holidays using fuzzy linear regression method. *IEEE Transactions on Power Systems*, **20**, 96-101.
- Song, K. B., Kwon, O. S. and Park, J. D. (2013). Optimal coefficient selection of exponential smoothing model in short term load forecasting on weekdays. *The Transactions of the Korean Institute of Electrical Engineers*, **62**, 149-154.
- Sugar, C. A. and James, G. M. (2003). Finding the number of clusters in a dataset: An informationtheoretic approach. *Journal of the American Statistical Association*, **90**, 928-934.
- Taylor, J. W. (2010). Triple seasonal methods for short-term electricity demand forecasting. *European Journal of Operational Research*, **204**, 139-152.
- Wei, G. C. and Tanner, M. A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *Journal of the American statistical Association*, **85**, 699-704.
- Weron, R. (2006). *Modeling and forecasting electricity loads and prices: A statistical approach*, Wiley, Chichester.
- Yeo, I. K. (2011). Clustering analysis of Korea's meteorological data. *Journal of the Korean Data & Information Science Society*, **22**, 941-949.

Functional clustering for electricity demand data: A case study[†]

Sanghoo Yoon¹ · Youngjean Choi²

¹²WISE institute, Hankook University of Foreign Studies

Received 6 May 2015, revised 30 June 2015, accepted 1 July 2015

Abstract

It is necessary to forecast the electricity demand for reliable and effective operation of the power system. In this study, we try to categorize a functional data, the mean curve in accordance with the time of daily power demand pattern. The data were collected between January 1, 2009 and December 31, 2011. And it were converted to time series data consisting of seasonal components and error component through log transformation and removing trend. Functional clustering by Ma *et al.* (2006) are applied and parameters are estimated using EM algorithm and generalized cross validation. The number of clusters is determined by classifying holidays or weekdays. Monday, weekday (Tuesday to Friday), Saturday, Sunday or holiday and season are described the mean curve of daily power demand pattern.

Keywords: Electricity demand, functional clustering, number of cluster, time series data.

[†] This work was funded by the Korea Meteorological Administration Research and Development Program under Grant Weather Information Service Engine (WISE) project, KMIPA-2012-0001.

¹ Corresponding author: Senior researcher, WISE institute, Hankook University of Foreign Studies, Yongin 449-791, Korea. Email: statstar@hufs.ac.kr

² Director, WISE institute, Hankook University of Foreign Studies, Yongin 449-791, Korea.