

## 군집 분석과 편차를 이용한 확률론적 가정용 부하 패턴 모델링

장문석, 박혜리, 배성우<sup>†</sup>  
한양대학교

### A Probabilistic residential load pattern modeling approach using clustering algorithm and statistical errors

Munseok Chang, Herie Park, Sungwoo Bae<sup>†</sup>  
Hanyang University

**Abstract** - 본 논문은 가정용 부하 패턴을 대상으로 군집 분석을 실시하여 대표적인 패턴과 군집의 편차를 추정하는 부하 모델링 방법을 제안하였다. 1년 동안의 가정용 부하 패턴은 먼저, 주성분 분석을 통하여 24개 차원에서 1개의 차원으로 축소되어 후속 군집 분석의 시간 및 용량 부담을 줄이도록 하였다. 군집 분석을 위하여, k-평균 알고리즘이 사용되었으며 365개의 부하 패턴은 비슷한 속성으로 묶인 6개의 대표 부하 패턴으로 도출되었다. 이후, 대표적인 부하 패턴과 이에 상응하는 군집 편차는 적합한 확률 밀도 함수를 사용하여 그 분포를 추정되었다. 최종적인 가정용 부하 패턴은 대표 부하 곡선과 그 편차를 추가함으로써 도출될 수 있다.

군집은 군집 중심 (부하 대표 패턴)과 이에 상응하는 군집 데이터 (부하 패턴)와의 제곱 유클리디안 거리의 합을 최소화하도록 설계된 군집 분석 방법이다.

#### 2.2.2 군집 평가

군집 결과의 타당성을 검증하기 위하여 여러 군집 평가 지표가 연구되고 있다. 최적의 군집 수는 군집 내부의 조밀도와 군집 사이간 분리도를 측정하는 평가 지표를 기반으로 결정될 수 있다. 본 논문에서는 군집 수가 증가할수록 평가 지표 값이 점진적으로 감소하는 4개의 측정 방법을 선정하였다.

평균 제곱 오차 함수 ( $J$ )는 동일한 가중치로 군집 데이터와 해당 데이터가 속한 군집 중심과의 거리를 측정한다 [1].

$$J = \frac{1}{N} \sum_{k=1}^K \sum_{x_i \in \Omega_k} d^2(x_i, c_k) \quad (1)$$

평균 거리 적합성 (Mean Index Adequacy ; MIA)은 군집에 할당된 데이터와 해당 군집 중심과의 평균 거리를 나타낸다 [1].

$$MIA = \sqrt{\frac{1}{K} \sum_{k=1}^K \sum_{x_i \in \Omega_k} d^2(x_i, c_k) / N_k} \quad (2)$$

군집 분산 지수 (Clustering Dispersion Indicator ; CDI)는 동일 군집에 속한 데이터 간 평균 거리와 서로 다른 군집 중심 사이의 거리 비율을 의미한다 [1].

$$CDI = \sqrt{\frac{1}{K} \sum_{k=1}^K \frac{1}{2N_k} \sum_{x_i \in \Omega_k} d^2(x_i, \Omega_k)} / \sqrt{\frac{1}{2K} \sum_{k=1}^K d^2(c_k, C_k)} \quad (3)$$

군집 내 제곱 거리와 군집 변동 비율 (Ratio of Within Cluster Sum of Squares to Between Cluster Variation ; WCBCR)은 군집 데이터와 해당 군집 중심간 제곱 거리 합과 군집 중심간 거리를 측정한다 [1].

$$WCBCR = \sum_{k=1}^K \sum_{x_i \in \Omega_k} d^2(x_i, c_k) / \sum_{1 \leq q < p} d^2(c_p, c_q) \quad (4)$$

최적의 군집 수는 위에서 언급한 4개 평가 지표의 변곡점 (knee point)에서 결정될 수 있다. 변곡점은 수식 (5)으로 표현되는 연속 차이 값(Successive difference value) 중 가장 높은 값을 갖는 지점을 의미한다 [3].

$$F_\varphi(k) = \varphi(k-1) + \varphi(k+1) - 2 \cdot \varphi(k) \quad (5)$$

여기서, 함수  $\varphi$ 는 평가지표,  $k$ 는 군집 수이다. 그림 1은 수식 (5)의 결과를 나타내며, 군집 수가 6, 7, 8개일 때 높은 연속 차이 값을 보였다. 군집 수 6, 7, 8개에 대하여 최적의 군집

## 1. 서 론

전력계통의 안정성 분석은 부하 수요의 모델링을 기본으로 한다. 하지만 계통 해석에 필요한 여러 요인 (다양한 발전원, 시나리오 등)을 고려하면, 장기간의 모든 부하 수요를 계통 해석에 여과없이 적용하기에는 부담이 가중될 수 있다. 따라서 전체 부하 수요의 대표성을 반영할 수 있는 부하 모델링 분석이 매우 중요하다.

부하 모델링에는 군집 분석을 이용하여 특정한 부하 패턴을 추정하는 연구가 활발히 진행되고 있다. 연구 [1]에서는 다양한 군집 분석을 사용하여 사용자의 전력 소비 패턴을 분석하였다. 연구 [2]에서는 부하 패턴을 계절별, 낮과 밤으로 분류하여 편차 분포를 추정한 부하 모델링을 수행하였다. 이에 본 논문은 군집 분석으로 최적의 대표 곡선을 추정하고 편차를 조합한 확률론적 부하 모델링을 수행하였다.

## 2. 군집 분석과 편차를 이용한 가정용 부하 모델링

### 2.1 가정용 부하 패턴 데이터

가정용 부하 데이터는 범용적이고 접근하기 용이한 IEEE Reliability Test System (RTS)를 사용하였다. 해당 데이터에서 가정용 부하 값은 피크 부하를 기준으로 퍼센트로 표현되어 있다. 이를 행렬화하면, 1년 동안 1시간 간격인 부하 데이터는 365 일 × 24 시간/일로 표현될 수 있다.

### 2.2 가정용 부하의 대표 패턴 추정

앞서 소개한 IEEE RTS 데이터는 전처리 과정을 통해 차원이 축소된 후 군집 분석이 수행되었다. 최적의 군집 수는 여러 평가 지표를 가중치와 함께 조합하여 판단되었다.

#### 2.2.1 데이터 전처리 및 군집 분석

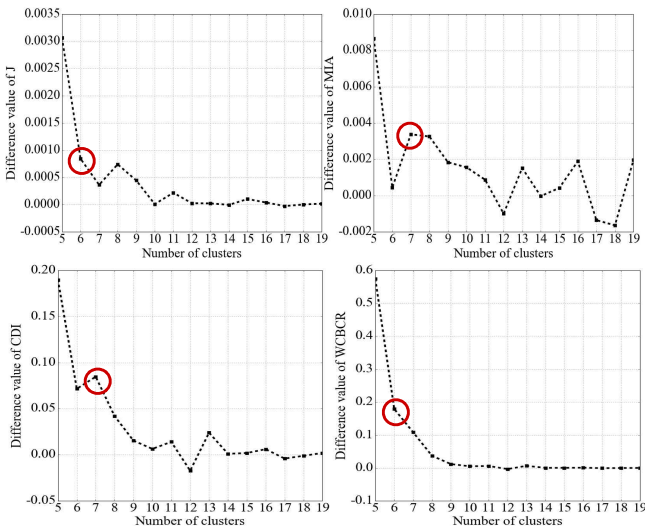
가용 데이터는 그 수와 차원 (특성)이 증가할수록 군집 분석의 동작에 시간 및 용량적 큰 부담을 주게 된다. 이를 완화하기 위한 대책으로 본 논문에서는 차원 축소 방법인 주성분 분석을 사용하였다. 24 차원인 IEEE RTS 데이터를 1 차원으로 축소할 시, 93.49 %의 분산이 보존됨을 확인하였다.

주성분 분석 후 변환된 가정용 부하 (365 일 × 1 시간/일)는 k-평균 군집 분석을 사용하여 대표 곡선이 도출되었다. k-평균

수를 선택하기 위하여, 4개 평가 지표의 연속 차이 값에 수식 (6)을 이용하여 가중치 함수를 적용하였다.

$$WEQ = w_1 \cdot F_J + w_2 \cdot F_{MIA} + w_3 \cdot F_{CDI} + w_4 \cdot F_{WCBCR} \quad (6)$$

가중치 함수  $w_1$ 과  $w_2$ 는  $(1-w)/4$ 로  $w_3$ 과  $w_4$ 는  $(1+w)/4$ 로 할당하였다. CDI와 WCBCR은 군집의 조밀도 뿐만 아니라 분리도를 포함한 평가 지표이므로 높은 가중치 함수를 적용하였다. 가중치 값  $w$ 는 0에서 1 사이 11 간격으로 나누었으며, 200회의 충분한 반복 수행을 실시하였다 (가중치 적용 수  $2200=200 \times 11$ ). 수식 (6)의 결과는 군집 수 6에서 1825.99 점, 7에서 1713.5 점, 8에서 1072.5 점으로 도출되었다. 가장 높은 점수는 군집 수 6에서 도출되어 최적 군집 수를 6으로 판단하였다.



〈그림 1〉 4개 평가 지표에 대한 연속 차이 값

### 2.3 군집별 편차 추정

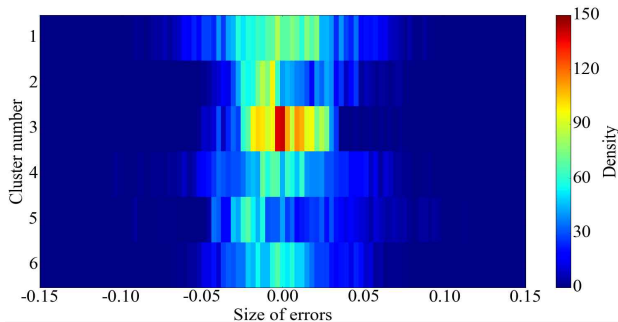
군집별 편차는 군집 중심에 대응되는 대표 부하 곡선에 군집 특성을 반영하기 위하여 계산되었다. 적합한 확률 밀도 함수를 편차 분포에 적용한 후, 편차는 대표 부하 곡선에 더해진다.

#### 2.3.1 편차 분포 분석

군집별 편차는 군집 중심과 군집 데이터 사이의 유클리디안 거리 차이를 의미한다. 그림 2는 6개의 군집과 각 군집에 해당하는 대표 곡선과의 편차 분포를 나타내었다. 군집 편차 (확률변수)는 0을 중심으로 앞부분 혹은 뒤로 편향되어 와이블 확률 밀도 함수로 추정되어 질 수 있다.

#### 2.3.2 확률 밀도 함수의 파라미터 값 추정

와이블 분포 함수는 3 종류의 모수 (형상 모수, 척도 모수, 위치 모수)를 가진다. 6개 군집별 편차를 와이블 분포 함수로 적합화한 모수 값은 표 1에 제시하였다. 추정된 모수를 기반으로 군집별 편차는 재 생성되어 부하 패턴에 반영된다.



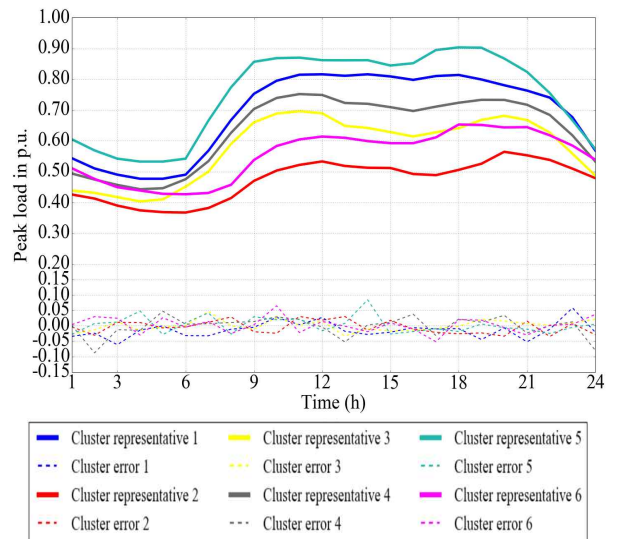
〈그림 2〉 6개의 군집 별 오차 분포

〈표 1〉 와이블 분포 함수 모수

군집 번호	형상 모수	척도모수	위치 모수
1	3.671	0.117	-0.105
2	2.219	0.055	-0.049
3	3.801	0.071	-0.064
4	4.263	0.128	-0.117
5	3.234	0.108	-0.097
6	2.900	0.078	-0.070

〈표 2〉 군집별 등장 확률

군집 번호	1	2	3	4	5	6
등장 확률	21.70	14.01	22.53	15.93	12.09	13.74



〈그림 3〉 대표 부하 패턴과 군집 오차

### 2.4 확률론적 가정용 부하 패턴 모델 결과

군집 알고리즘으로부터 도출된 6개의 가정용 부하 대표 패턴의 등장 확률은 표 2에 제시되었다. 또한, 대표 부하 패턴과 각 군집별 편차 분포는 그림 3에 나타내었다. 최종적으로 가정용 부하 패턴은 대표 곡선과 편차를 더하여 도출되어진다.

## 3. 결 론

본 논문은 군집분석을 통하여 전체 부하 패턴을 대표할 수 있는 최적의 부하 패턴 수를 추정하고, 이에 편차 분포를 더한 새로운 가정용 부하 패턴의 확률론적 모델링 방법을 제시하였다.

향후, k-평균 군집 분석 뿐 만아니라 다양한 군집 알고리즘의 적용을 통하여 부하 모델링에 가장 적합한 알고리즘을 찾는 연구 수행이 추가적으로 필요하다고 사료된다.

### 감사의 글

본 연구는 산업통상자원부(MOTIE)와 한국에너지기술연구원(KETEP)의 지원을 받아 수행한 연구 과제입니다. (No. 20161210200560)

### [참 고 문 헌]

- [1] J. George, et al, "Two-Stage Pattern Recognition of Load Curves for Classification of Electricity Customers," *IEEE Trans. Power Systems*, vol. 22, no.3, pp.1120 - 1128, Aug. 2007
- [2] 장문석, 배성우, "전기자동차 충전 수요에 기인한 제주 일부지역 시나리오별 배전계통 영향 평가," 대한전기학회 하계학술대회, July. 2018
- [3] Q. Zhao, et al, "Knee point detection on Bayesian information criterion," *IEEE ICTAI*, vol. 2, pp. 431 - 438, 2008