

# 클러스터링 기법을 이용한 전력 고객의 대표 부하패턴 생성에 대한 연구

김영일\*, 최훈\*\*

\*한국전력공사 전력연구원 정보통신그룹

\*\*충남대학교 컴퓨터공학과

e-mail : yikim@kepcoco.kr

## A Study for Load Profile Generation of Electric Power Customer using Clustering Algorithm

Young-Il Kim \*, Hoon Choi \*\*

\*Information Communications Group, Korea Electric Power Research Institute

\*\*Dept. of Computer Engineering, Chungnam National University

### 요 약

한전에서는 연간 전력 사용량이 높은 고압 고객에 대하여 전자식 전력량계를 설치하여 15 분 단위로 전력 사용량을 수집하는 자동검침시스템을 운영하고 있다. 본 연구에서는 자동검침시스템을 통해 수집된 데이터를 이용하여 배전선로에 대한 부하를 분석하기 위해 자동검침 고객의 부하 데이터를 이용하여 클러스터링 기법을 통해 대표 부하패턴을 생성하는 방식을 제안하였다. 기존에는 계약종별 코드가 동일한 고객들의 부하패턴을 이용하여 15 분 단위의 평균 사용량을 계산하여 대표 부하패턴을 생성하는 방식을 사용하였으나, 같은 계약종별 코드를 갖는 고객이라 할지라도 부하패턴이 다른 경우가 많아서 부하분석의 정확도를 떨어뜨렸다. 본 연구에서는 동일한 계약종별 코드를 갖는 고객에 대하여 15 분 단위 자동검침 데이터를 이용하여 k-means 기법을 통해 고객을 분류하고 각 그룹마다 대표 부하패턴을 생성하는 방식을 제안하였다.

### 1. 서론

한전에서는 배전선로를 효과적으로 운영하기 위해 자동검침시스템을 운영하고 있다. 초기의 자동검침시스템은 검침 비용 절감을 위해 도입되었으나 최근에는 배전선로의 회선 및 구간에 대한 부하정보 분석에 많이 활용되고 있다. 고객의 부하정보를 15 분 단위로 수집하여 제공하면 고객은 에너지 사용 효율을 최적화 할 수 있는 전력 운용 계획을 세울 수 있으며, 전력사에서는 시간대별 가격정책과 부하관리, 수요자 관리, 배급망 최적화, 배전 운용 통신 등의 분야에 활용할 수 있게 된다. 특히 부하 및 배전망 최적화 부분에 있어서는 상당한 효과를 얻고 있으며, 최근에는 실시간 배전 최적화를 위한 응용 프로그램 개발에 대한 연구가 진행되고 있다.

기존의 배전선로의 부하분석을 위한 방식은 변전소를 통해 각 회선으로 제공되는 15 분 단위의 전력량을 SOMAS (Substation Operating results MAnagement System)를 통해 얻을 수 있을 뿐 이었다. 자동검침시스템은 회선에 연결된 고객이나 설비의 15 분 단위 전력량을 제공할 수 있어 보다 정확하고 빠른 정보분석 능력을 갖출 수 있게 한다. 그러나 모든 설비에 자동검침 장비를 설치하여 데이터를 수집하는 방식은 비용적인 측면에서 또 다른 문제점을 발생시킨다.

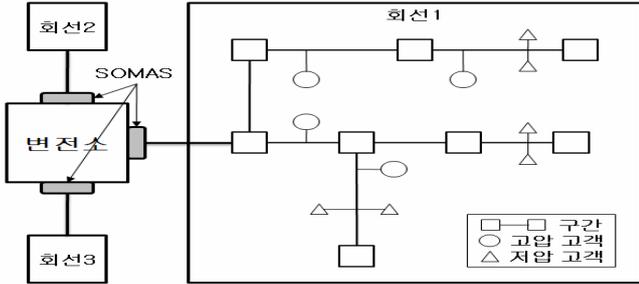
본 연구에서는 최소한의 설비에 검침장비를 설치하

고 수집된 데이터를 바탕으로 데이터 분석을 통해 각 설비에 대한 대표 부하패턴을 생성하고 검침장비가 설치되지 않은 설비의 월 사용량을 대표 부하패턴에 적용하여 가상의 15 분 단위 부하패턴을 생성하였다. 현재 한전에서는 신규 고객을 등록할 때 계약종별, 전기사용용도, 고객의 산업종류 및 계약전력량 등과 같은 정보를 입력하여 고객을 분류하고 있으며, 계약종별에 따라 고객의 대표 부하패턴을 생성하여 배전선로의 부하분석 자료로 사용하고 있다. 그러나 이러한 방식은 고객정보 입력 과정에서 고객에 대한 정확한 분류가 어렵고, 계약종별 코드만으로 생성한 대표 부하패턴의 오차가 커서 분석에 어려움을 갖게 된다. 본 연구에서는 동일한 계약종별 코드를 갖는 자동검침 고객의 부하패턴을 이용하여 k-means 기법을 적용하여 고객을 분류하였다. 고객 분류를 위한 클러스터의 개수(k)는 대표 부하패턴의 오차가 수렴하는 지역에 대한 값으로 선정하였다.

### 2. 대표 부하패턴을 이용한 부하분석

#### 2.1 자동검침 개요

배전선로는 그림 1 과 같이 변전소에서 여러 개의 회선으로 나누어지고, 회선에서는 여러 개의 구간으로 나누어져 구성된다.



(그림 1) 배전선로의 구성요소

고전적인 배전선로의 운영은 각 회선의 사용전력을 SOMAS 를 통해 15 분단위로 검침한 데이터를 바탕으로 이루어졌다. 그러나 SOMAS 데이터 만으로는 회선 전체의 개략적인 운영상태는 파악할 수 있지만, 회선 내에 각 구간별, 각 변압기별 운영상태는 파악이 어렵게 된다. 이러한 정보를 파악하기 위해서는 구간이나 변압기에 운영정보를 취득할 수 있는 계측장비를 설치하고 이를 통해 수집된 데이터를 분석해야만 가능하다. 자동검침은 배전설비에 15 분 단위로 전력량을 계량할 수 있는 장비를 부착하고 이를 PLC 와 같은 유선망이나 CDMA 와 같은 무선망을 이용하여 중앙 서버로 전송하는 방식으로 운영된다. 15 분 단위의 데이터를 활용할 경우 시간대 별로 회선 및 구간에 대한 사용량 및 피크 로드 등의 분석이 가능하게 되며 배전선로의 재구성 및 추후 배전망의 확충을 위한 분석자료로 활용이 가능하게 된다.

2.2 대표 부하패턴을 이용한 부하분석

최근에는 자동검침 데이터를 활용하여 부하를 분석하는 방식이 연구되고 있다. [1]은 약 10 만호의 고압고객을 대상으로 15 분 단위의 전력 사용량을 수집하여 각 계약종별 일일 대표 부하패턴을 생성하였다. 이를 활용하여 자동검침을 하지 않는 고압고객의 월 검침량을 각 계약종별 대표 부하패턴에 적용하여 15 분 단위의 가상 부하 패턴을 생성하였다. 이렇게 생성된 부하패턴을 Forward Sweeping[2] 기법을 적용하여 조류 계산을 통해 회선 및 구간에 대한 15 분 단위 부하 패턴을 생성할 수 있었다. 이러한 분석 방식을 통해 각 구간에 대하여 15 분/시간/일/주/월 단위의 최대부하 및 부하패턴 등을 분석할 수 있으며, 이를 통해 부하절체나 신규 수용, 과부하선로 해소, 고장선로 부하분석 등의 현장업무에 효과적으로 적용할 수 있게 되었다.

2.3 계약종별 코드를 이용한 대표 부하패턴 생성

기존의 대표 부하패턴을 생성하는 방식은 동일한 계약종별 코드를 갖는 고객들의 15 분 단위 데이터를 합하여 평균을 내는 방식으로 산출되어 왔다.

특정 고객 c 에 대한 하루 부하패턴을  $L_c$ 라고 할 경우,

$$L_c = [L_c^1, L_c^2, \dots, L_c^i, \dots, L_c^k]$$

- $L_c$ : 고객 c 의 하루 부하패턴
- : 고객 c 의 i 번째 15 분 단위 부하량
- i: 15 분 단위 부하량에 대한 일련번호

T: 15 분 단위로 하루 측정 시 최대값  $T=4 \times 24 = 96$  이 된다. 이를 이용하여 계약종별 코드가 산업용인 고객에 대해 각 15 분 단위 부하량의 평균을 내는 방식으로 일일 대표 부하패턴을 나타내면 다음과 같다.

$$SLP_{day}^{industry} = \left[ \frac{\sum_{c=1}^{cn} L_c^1}{cn}, \frac{\sum_{c=1}^{cn} L_c^2}{cn}, \dots, \frac{\sum_{c=1}^{cn} L_c^i}{cn}, \dots, \frac{\sum_{c=1}^{cn} L_c^k}{cn} \right]$$

(cn: 계약종별 코드가 산업용인 고객의 수)

2.4 클러스터링 기법을 이용한 대표 부하패턴 생성

본 연구에서는 고객들의 자동검침 데이터를 이용하여 부하패턴이 유사한 고객들을 분할하기 위해 k-means[3] 알고리즘을 선택하였다. k-means 알고리즘은 주어진 데이터를 특정 성질에 기초해서 k 묶음으로 나누는 방법 중 하나로 데이터가 벡터 공간을 이룬다고 가정하고, 각 묶음의 분산도를 최소화하는 방식이다. 이것을 식으로 나타내면 다음과 같다.

$$V = \sum_{i=1}^k \sum_{j \in C_i} |x_j - \mu_i|^2$$

- : k 개의 묶음,  $i = \{1, 2, \dots, k\}$
- : 각 묶음의 중심(대표)값

K-means 알고리즘은 맨 처음 각 점들을 k 개 집합으로 나눈 후, 각 집합의 중심 값을 구한다. 이 후, 각각의 점들을 방금 구한 중심 값 중에서 제일 가까운 것에 새롭게 연결하면서 새로운 집합으로 나누게 된다. 이 작업을 반복하면 점들이 소속된 집합이 바뀌지 않거나, 중심 값이 바뀌지 않는 상태로 수렴하게 되어 최종 중심 값을 얻을 수 있게 된다.

K-means 알고리즘을 이용하여 계약종별 코드가 산업용인 고객에 대하여 k 개로 분할한 일일 대표 부하패턴을 나타내면 다음과 같다.

$$SLP_{day}^{industry} = \{LP_{day}^1, LP_{day}^2, \dots, LP_{day}^i, \dots, LP_{day}^k\}$$

$$LP_{day}^i = [L_c^1, L_c^2, \dots, L_c^i, \dots, L_c^k]$$

$LP$ : 클러스터  $C_i$  의 대표 부하패턴

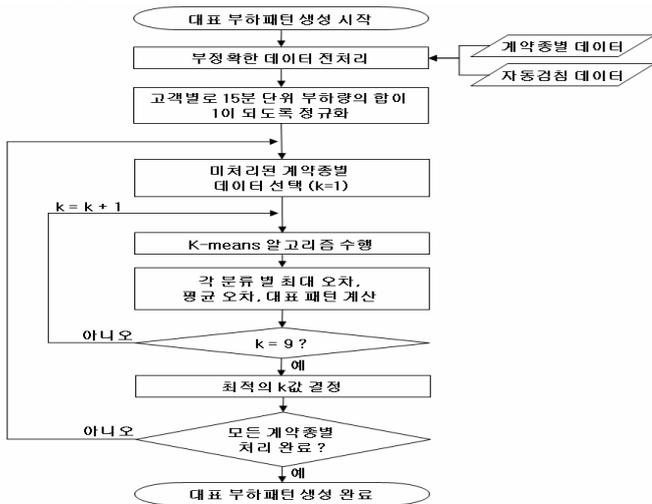
$i$ : 클러스터  $C_i$  의 j 번째 15 분 단위 대표 부하량

3. 계약종별 고객분류 과정 및 방법

계약종별로 클러스터링을 통해 고객을 분류하기 위해 본 연구에서는 강남지역의 고객에 대한 15 분 단위 자동검침 데이터를 이용하여 k-means 알고리즘을 적용하고 대표 부하패턴을 생성하였다. 대표 부하패턴을 생성하기 위한 과정은 그림 2 와 같다.

대표 부하패턴 생성을 위해 고객의 계약종별 데이터와 자동검침 데이터를 추출하고, 전처리 과정을 통해 통신오류 등으로 15 분 단위 사용량이 완벽하게 검침되지 않은 부정확한 데이터들을 제거한다. 전처리된 데이터는 각 고객별로 15 분 단위 부하량의 합이 1 이 되도록 정규화 한다. 이 정규화 방식을 사용하면 총 사용량은 많은 차이가 있다고 하더라도 15 분 단위의 사용량의 증감 방식이 유사한 고객들을 같은 그룹으로 분류할 수 있을 뿐 아니라, 정규화된 대표 부하패턴을 생성한 후 각 고객의 총 사용량을 곱하여 손쉽게 15 분 단위 부하 값을 계산해 낼 수 있

기 때문이다.



(그림 2) 계약종별 부하패턴 계산 절차

정규화를 완료하면, 하나의 계약종별 데이터를 선택하여 k 값을 1부터 9까지 증가시키며 k-means 알고리즘을 수행하고 각 분류 별로 최대 오차와 평균 오차, 대표 부하패턴을 계산한다. 이때 k가 1인 경우는 클러스터링 기법을 적용하지 않는 기존의 방식과 동일한 결과를 갖는다. 최대 k 값까지 적용하였으면, 평균 오차를 이용하여 최적의 k 값을 결정하고 이를 해당 계약종별의 분류 개수로 결정한다.

#### 4. 고객분류 결과 분석

##### 4.1 실험 데이터 및 기존의 대표 부하패턴 분석

본 연구를 위해 한전 강남지점의 2007년 1월 8일의 자동검침 고객 데이터를 이용하였다. 데이터의 전처리 과정을 통해 부정확한 데이터를 제거하고 7가지 계약종별에 대하여 총 3093명의 고객의 자동검침 데이터를 이용하여 계약종별 분류를 수행하였으며 수행결과는 표 1과 같다. k 값에 따른 계약종별 최대 오차( $SLP_{max}^k$ )는 각 k 값에 따라 분류된 클러스터들의 최대 오차( $LP_{max}^k$ ) 중에서의 최대값이 된다. 각 클러스터의 최대 오차는 클러스터의 대표 부하패턴( $I_j^i$ )과 해당 클러스터의 고객의 15분 단위 부하량( $I_j$ )과의 거리 중 최대값이 된다.

$$SLP_{max}^k = \text{Max}(LP_{max}^{c_1}, LP_{max}^{c_2}, \dots, LP_{max}^{c_i}, \dots, LP_{max}^{c_k})$$

$$LP_{max}^{c_i} = \text{Max}(d_{max_1}^{c_i}, d_{max_2}^{c_i}, \dots, d_{max_m}^{c_i}, \dots, d_{max_n}^{c_i})$$

$$d_{max_j}^{c_i} = \text{Max}(|I_j^{c_i} - I_j^{i1}|, \dots, |I_j^{c_i} - I_j^{im}|, \dots, |I_j^{c_i} - I_j^{in}|)$$

$d_{max_j}^{c_i}$ : 클러스터  $C_i$ 의  $j$  번째 대표 부하패턴 부하량과 해당 클러스터의 고객들의  $j$  번째 부하량과의 거리 중 최대값

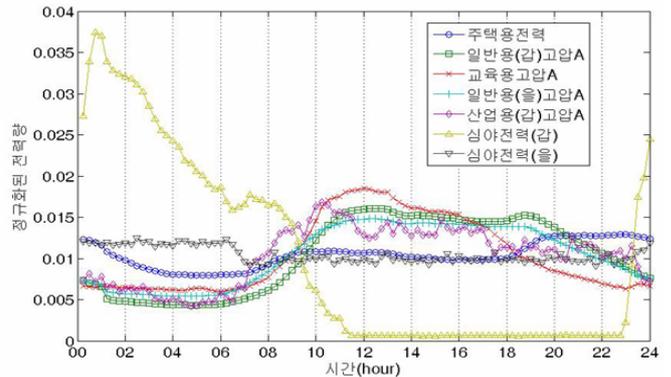
$I_j^i$ : 클러스터  $C_i$ 에 속한  $m$  번째 고객의  $j$  번째 15분 단위 부하량

$n$ : 클러스터  $C_i$ 의 총 고객수

k 값에 따른 계약종별 평균 오차는 각 k 값에 따라 분류된 클러스터의 대표 부하패턴과 해당 클러스터내의 모든 고객의 실제 부하량과의 거리의 평균값이 된다.

$$SLP_{avg}^k = \frac{\sum_{i=1}^k \sum_{j=1}^m \sum_{m=1}^n (|I_j^{c_i} - I_j^{im}|)}{\sum_{i=1}^k T \times n_i}$$

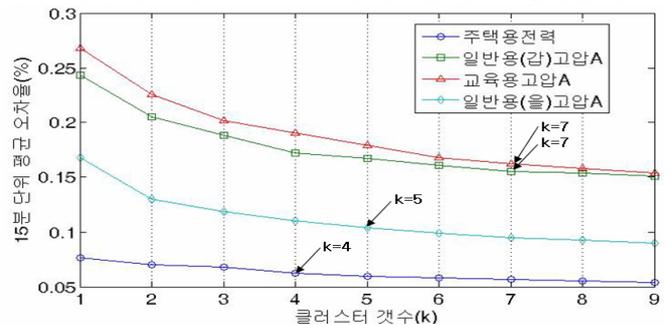
계약종별 분류 과정에서 k 값이 1인 경우는 계약종별로 고객을 분류하지 않은 경우로 기존의 계약종별 대표 부하패턴 계산 방식과 동일하며, 이때의 대표 부하패턴은 그림 3과 같다. 심야전력(갑)의 경우를 살펴보면 주간에는 부하량이 거의 없으며 심야에만 부하량이 급증하고 있어 심야전력(갑)의 대표 부하패턴을 잘 보여주고 있다.



(그림 3) 기존 방식의 계약종별 대표 부하패턴

##### 4.2 클러스터링 기법의 k 값 선정

계약종별 k 값을 결정하기 위해 k 값을 1에서 9까지 적용하여 계산된 평균오차( $SLP$ )를 이용하여 k 값 증가에 따른 평균오차의 차이값을 이용하였다. 그림 4는 k 값 증가에 따른 15분 단위 평균 오차율을 나타낸다. 이 그래프에서 k 값 증가에 따른 평균오차의 차이값은 기울기가 되며, 기울기 값이 0.005 이하인 경우 수렴하는 것으로 판단하여 k 값을 결정하였다.



(그림 4) 계약종별 k 값 증가에 따른 평균 오차율

##### 4.3 클러스터링을 이용한 계약종별 분류 결과 분석

본 연구를 통해 분류된 일반용(갑)고압 A와 일반용(을)고압 A의 분류 결과를 살펴보면 표 2와 같다. 그림 5는 k가 7일 때의 일반용(갑)고압 A에 대한 클러스터 별 대표 부하패턴을 나타낸다.

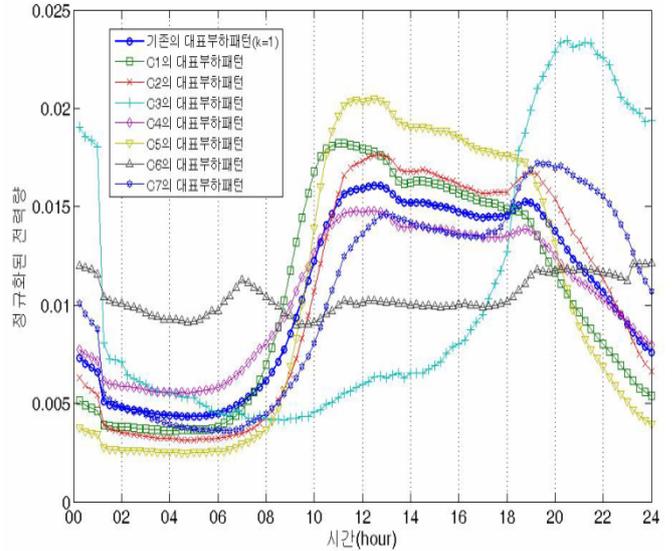
그림 5를 살펴보면 클러스터를 통해 분류된 고객들의 대표 부하패턴이 기존의 하나의 대표 부하패턴을 사용할 때와 많은 차이를 보이고 있음을 알 수 있다. 표 1에서 보이는 바와 같이 일반용(갑)고압 A에 대해 하나가 아닌 7개의 대표 부하패턴을 생성할 경우 매 15분 단위의 부하량에 대해 0.088%의 평균 오차를 더 줄일 수 있게 되며, 이러한 오차를 줄임으로

써 일일 또는 한 달의 부하를 예측하는 경우에 대한 오차를 크게 줄일 수 있게 된다.

도 등에 상관관계를 반영하여 보다 정확하게 가상 부하패턴을 생성하는 연구를 진행할 예정이다.

<표 2> 일반용(갑)고압 A 와 일반용(을)고압 A 에 대한 클러스터 별 고객수 및 최대, 평균 오차율

| 계약 종류      | 클러스터 번호 | 고객수 | 최대 오차율 | 평균 오차율 |
|------------|---------|-----|--------|--------|
| 일반용(갑)고압 A | C1      | 501 | 2.316  | 0.006  |
|            | C2      | 490 | 1.496  | 0.006  |
|            | C3      | 50  | 2.881  | 0.013  |
|            | C4      | 510 | 1.567  | 0.006  |
|            | C5      | 287 | 1.713  | 0.006  |
|            | C6      | 200 | 6.009  | 0.009  |
|            | C7      | 302 | 1.957  | 0.007  |
| 일반용(을)고압 A | C1      | 99  | 0.899  | 0.024  |
|            | C2      | 47  | 0.77   | 0.034  |
|            | C3      | 137 | 1.019  | 0.022  |
|            | C4      | 90  | 1.104  | 0.024  |
|            | C5      | 37  | 1.364  | 0.023  |



(그림 5) k=7 일 때의 일반용(갑)고압 A 에 대한 클러스터 별 대표 부하패턴

5. 결론 및 향후 연구

자동검침시스템이 도입되면서 자동검침 고객의 데이터를 이용하여 전체 배전선로의 회선 및 구간에 대한 15 분/시간/일/주/월 단위의 최대부하 및 부하패턴 등을 분석하는 방법이 많이 연구되고 있다. 기존의 대표 부하패턴 생성방식은 계약종별 코드가 같은 고객에 대하여 대표 부하패턴을 생성하였으나, 본 연구에서는 계약종별 코드가 같은 고객에 대하여 k-means 알고리즘을 이용하여 보다 세분화된 분류를 통해 대표 부하패턴을 생성하는 방법을 제안하였다.

향후에는 대표 부하패턴을 이용하여 자동검침을 실시하지 않는 고압고객에 대한 가상 부하패턴을 생성하고, 전력 사용량과 밀접한 관계가 있는 날씨 및 온

참고문헌

[1] 신진호, 김영일, 송재주, 이봉재, 이정일, “지리정보와 검침데이터를 이용한 배전계통 부하분석모델 개발”, 대한전기학회 하계학술대회, 2006, 7 월, pp. 2124-2125.  
 [2] G. W. Chang, S. Y. Chu, and H. L. Wang, “A Simplified Forward and Backward Sweep Approach for Distribution System Load Flow Analysis,” 2006 International Conference on Power System Technology, pp. 1-5.  
 [3] Jain A. K. and Dubes R.C., 1988. “Algorithms for Clustering Data,” Englewood Cliffs, NJ: Prentice-Hall.

<표 1> 계약종별 클러스터링 결과

| 계약종류       | 고객수  | 오차율 | k 값에 따른 오차율(%) |       |       |       |       |       |       |       |       |
|------------|------|-----|----------------|-------|-------|-------|-------|-------|-------|-------|-------|
|            |      |     | 1              | 2     | 3     | 4     | 5     | 6     | 7     | 8     | 9     |
| 주택용전력      | 213  | 최대  | 1.015          | 1.023 | 0.591 | 0.596 | 0.613 | 0.599 | 0.599 | 0.549 | 0.561 |
|            |      | 평균  | 0.076          | 0.07  | 0.068 | 0.063 | 0.06  | 0.058 | 0.056 | 0.056 | 0.054 |
| 일반용(갑)고압 A | 2340 | 최대  | 6.629          | 6.44  | 6.227 | 6.075 | 6.039 | 6.027 | 6.009 | 4.581 | 4.581 |
|            |      | 평균  | 0.243          | 0.205 | 0.188 | 0.172 | 0.167 | 0.161 | 0.155 | 0.154 | 0.151 |
| 교육용고압 A    | 90   | 최대  | 3.916          | 3.76  | 3.573 | 2.689 | 2.259 | 3.328 | 3.255 | 1.264 | 1.277 |
|            |      | 평균  | 0.268          | 0.226 | 0.202 | 0.19  | 0.179 | 0.168 | 0.162 | 0.158 | 0.154 |
| 일반용(을)고압 A | 410  | 최대  | 1.731          | 1.574 | 1.565 | 1.539 | 1.364 | 1.358 | 1.382 | 1.398 | 1.322 |
|            |      | 평균  | 0.168          | 0.13  | 0.119 | 0.11  | 0.104 | 0.099 | 0.095 | 0.092 | 0.09  |
| 산업용(갑)고압 A | 18   | 최대  | 3.435          | 3.305 | 2.528 | 2.473 | 2.429 | 2.429 | 2.369 | 2.161 | 2.161 |
|            |      | 평균  | 0.371          | 0.328 | 0.288 | 0.246 | 0.209 | 0.185 | 0.156 | 0.131 | 0.106 |
| 심야전력(갑)    | 10   | 최대  | 3.347          | 3.006 | 2.711 | 1.874 | 1.874 | 1.703 | 1.363 | 0.709 | 0.435 |
|            |      | 평균  | 0.34           | 0.258 | 0.182 | 0.153 | 0.123 | 0.098 | 0.06  | 0.037 | 0.015 |
| 심야전력(을)    | 12   | 최대  | 1.774          | 0.815 | 0.414 | 0.385 | 0.205 | 0.189 | 0.194 | 0.127 | 0.068 |
|            |      | 평균  | 0.187          | 0.071 | 0.054 | 0.04  | 0.029 | 0.022 | 0.016 | 0.01  | 0.006 |