일일 대표 부하패턴의 분별력을 높이기 위한 반복적인 소규모 군집화를 이용한 고객 군집화 방법

논 문
58-11-30

Customer Clustering Method Using Repeated Small-sized Clustering to improve the Classifying Ability of Typical Daily Load Profile

김 영 일† · 송 재 주* · 오 도 은** · 정 남 준*** · 양 일 권§
(Young-Il Kim · Jae-Ju Song · Do-Eun Oh · Nam-Joon Jung · Il-Kwon Yang)

Abstract - Customer clustering method is used to make a TDLP (typical daily load profile) to estimate the quater hourly load profile of non-AMR (Automatic Meter Reading) customer. In this paper, repeated small-sized clustering method is supposed to improve the classifying ability of TDLP. K-means algorithm is well-known clustering technology of data mining. To reduce the local maxima of k-means algorithm, proposed method clusters average load profiles to small-sized clusters and selects the highest error rated cluster and clusters this to small-sized clusters repeatedly to minimize the local maxima.

Key Words : Customer clustering, K-means, Typical Load Profile , AMR (Automatic Meter Reading)

1. 서 론

배전선로의 15분 단위의 부하분석은 배전선로의 안정적인 운영을 위해 필요한 기술 중의 하나이다. 배전선로의 15분 단위의 부하분석을 위해서는 선로를 구성하는 회선이나 구간에 대한 15분 단위의 부하곡선이 필요하다. 기존에는 변전소에서 공급하는 회선에 대한 15분 단위 부하를 측정하여 SOMAS(Substation Operating results MAnagement System)에 저장하고 회선에 대한 부하분석에 활용하였다. 최근에는 과부하의 위험이 있는 일부 설비에 자동검침 장비를 이용하여 특정 구간에 대한 부하를 분석하는 연구도 진행되고 있다. 그러나 모든 구간에 대한 부하분석을 하고자 하는 경우에는 모든 설비에 자동검침 장비를 설치해야 되므로 비용적인 측면에서 어려움을 갖게 된다. 따라서 자동검침 장비가 설치된 고객의 15분 단위의 부하곡선을 이용하여 고객들을 군집화하고 각 군집별로 대표 부하패턴을 생성하여 자동검침 장비가 설치되지 않은 고객들을 분류(classification) 기법을 이용하여 부하형태가 가장 유사할 것으로 예상되는 군집의 대표 부하패턴을 선택하여 월 전기사용량을 대입하여 가상 부하곡선을 만드는 방법이 연구되고 있다[1][2]. 자동검침 고객들의 실제 부하곡선과 미 검침 고객들의 가상 부하곡선을 이용하면 모든 구간에 대한 부하분석이 가능하게 된다[3]. 자동검침 고객의 실제 부하곡선과 미검침 고객의 가상 부하곡선을 이용할 경우 배전망의 관리 및 자동화[4], 부하 예측, 노후설비의 수명예측[5] 등의 분야에서의 활용을 통해 보다 안정적인 배전선로의 운영이 가능하게 된다

† 교신저자, 정회원 : 한국전력공사 전력연구원
E-mail : yikim@kepco.co.kr
* 정 회 원 : 한국전력공사 전력연구원
** 정 회 원 : 한국전력공사 전력연구원
*** 정 회 원 : 한국전력공사 전력연구원
§ 정 회 원 : 한국전력공사 전력연구원
접수일자 : 2009년 8월 12일
최종완료 : 2009년 9월 30일

2. 고객 부하곡선을 이용한 군집화 방법

자동검침 고객의 부하곡선을 군집화하는 방법으로 종래에는 k-means 알고리즘을 이용한 군집화를 주로 사용하였다. k-means 알고리즘은 공간상에 존재하는 데이터를 서로 가까운 데이터들 끼리 묶어서 몇 개의 군집으로 나누는 군집화 방법 중 하나로 데이터가 벡터 공간을 이룬다고 가정하고, 각각의 묶음에 속하는 점들이 다른 묶음의 중심에서의 유클리디안 거리보다는 자신이 속한 묶음까지의 유클리디안 거리가 최소화 되도록 군집화하여 묶음의 분산도를 최소화하는 방식이다[6].

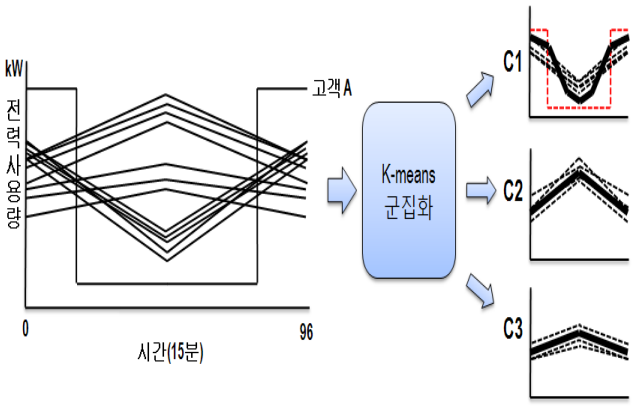

그림 1 K-means 방법을 이용한 고객 군집화 예
Fig. 1 Example of customer clustering using k-means algorithm

고객 부하곡선을 k-means 알고리즘을 이용하여 군집화하는 방법을 예를 들어 설명하면 그림 1과 같다. 군집화를 위한 자동검침 데이터로 13명의 자동검침 고객의 실제 부하곡선이 있다고 가정할 경우, 이를 데이터마이닝에서 주로 사용되는 k-means 알고리즘을 이용하여 3개의 군집으로 군집화할 경우 C1, C2, C3 군집이 생성되게 된다. 이때 각 군집에 속한 고객의 실제 부하곡선에 대한 평균값이나 대표값을 계산하여 이를 대표 부하패턴으로 정의하게 된다. 그러나 고객 군집화를 위해 k-means 알고리즘을 이용할 경우, 데이터가 많이 편중된 지역의 오차를 줄이기 위해 중심에서 먼 지역의 소수의 데이터들의 오차는 무시되는 지역편중 현상이 발생하게 되어 그림 1의 고객 A의 경우에는 하나의 군집이 되지 못하고 군집 C1에 속하게 되어 고객 A의 패턴이 무시되는 현상이 발생하게 된다. 고객 군집화의 주된 목적은 군집화를 통해 얻어진 대표 부하패턴을 이용하여 자동검침 장비가 설치되지 않은 고객을 분류 기법을 이용하여 군집에 할당하여 가상 부하패턴을 생성하는 것이다. 이때 C2와 C3의 경우에는 대표 부하패턴이 유사하여 분류 성능이 떨어지는 현상이 발생하게 된다. 따라서 유사한 형태의 대표 부하패턴을 가지면서 가상 부하패턴의 오차를 줄이는 것 보다는 약간의 오차의 증가를 감수하면서 다양한 형태의 대표 부하패턴을 갖도록 하는 것이 효과적이다.

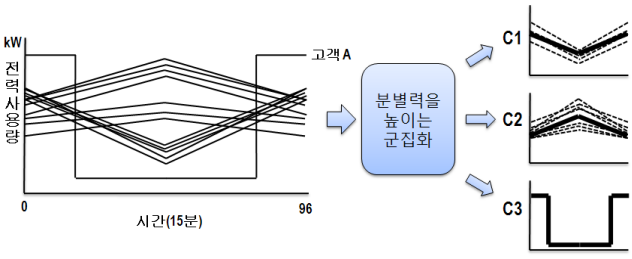

그림 2 분별력을 높이는 고객 군집화 예

Fig. 2 Example of customer clustering to improve the classification ability

그림 2를 살펴보면 가상 부하패턴의 오차는 증가하겠지만, 고객 A가 하나의 군집으로 분류되어 세 가지 군집의 대표 부하패턴이 그림 1에 비해 분별력이 높다는 것을 알 수 있다.

본 논문에서는 자동검침 고객의 부하곡선을 군집화하는 방법으로, k-means 알고리즘을 이용한 군집화 방법의 지역편중 현상을 해결하고자 하나의 군집을 소규모의 군집들로 나누고, 각각의 군집들 중에서 평균 오차율이 가장 높은 군집을 다시 소규모의 군집으로 나누는 과정을 반복적으로 수행하는 군집화 방법을 제안하였다.

3. 소규모 군집화 수행 절차

그림 3은 반복적인 군집화를 수행하는 절차를 나타낸다. 자동검침 데이터는 자동검침 장비가 설치된 고객들로부터 수집된 매 15분 단위의 사용전력(l_i)을 이용하여 하루 단위로 표현된 각각의 고객에 대한 일일 부하곡선을 말한다. 고객 c에 대한 일일 부하곡선(LP: Load Profile)은 아래의 식과 같이 나타난다.

$$LP_{day}^{c} = [l_1^c, l_2^c, ..., l_i^c, .., l_{96}^c]$$

전처리 과정에서는 하나의 고객에 대해서 매월 평일에 대하여 동일한 시각에 대한 매 15분 단위의 사용전력의 평균값(μ_i^c)을 계산하여 고객의 일일 평균 부하곡선(ALP: Average Load Profile)을 만들게 된다.

$$\mu_i^c = \frac{\sum_{day=1}^{20} l_i^c}{20} \quad (\text{주 } 5\text{일} \times 4\text{주} = 20\text{일})$$

$$ALP_{day}^{c} = [\mu_1^c, \mu_2^c, ..., \mu_i^c, ...\mu_{96}^c]$$

자동검침 고객의 일일 평균 부하곡선의 생성이 완료되면 고객의 일일 평균 부하곡선의 최대값이 1이 되도록 하는 정규화 과정을 거쳐 정규화된 일일 평균 부하곡선을 만들게 된다. 모든 그래프의 y축의 값을 최대 1로 만드는 정규화 과정은 전력 사용량을 나타내는 그래프의 패턴 유사도를 손쉽게 찾을 수 있도록 도와준다.

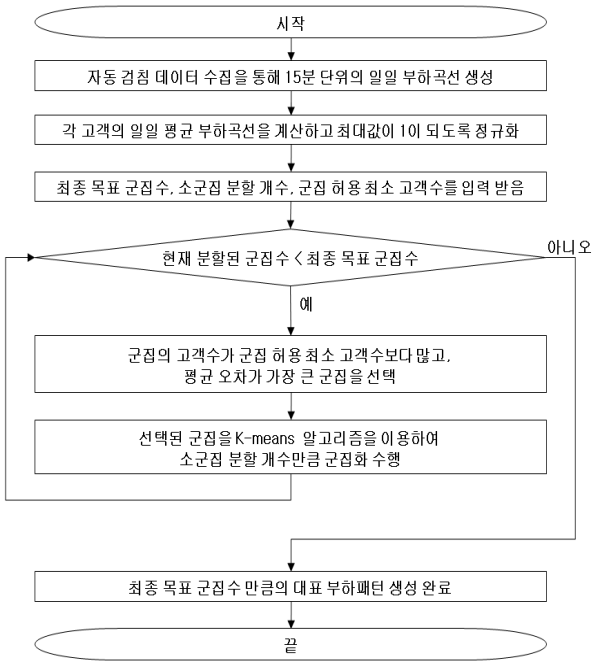

그림 3 반복적인 소규모 군집화 수행 절차

Fig. 3 Flow chart of repeated small-sized clustering process

군집화를 수행하기 위해서는 최종 목표 군집수, 소군집 분할 개수, 군집 허용 최소 고객수를 입력받게 된다. 최종 목표 군집수는 입력받은 군집을 최종적으로 몇 개의 군집으로 만들 것인가를 나타내며, 소군집 분할 개수는 하나의 군집을 k-means 처리기로 군집화할 때 몇 개의 소규모 군집으로 분할 할 것인가를 나타낸다. 군집의 분할 허용 최소 고객수는 하나의 군집을 소규모 군집으로 분할하고자 하는 경우 해당 군집에 최소한으로 존재하는 고객의 수를 나타내는 것으로 최소 고객 수보다 적은 경우에는 군집의 분할을 허용하지 않게 된다.

전처리가 완료되면 하나의 군집에 속한 일일 평균 부하곡

선들을 k-means 알고리즘을 이용하여 소규모의 군집으로 나누고, 각 군집(Gj)에 속한 일일 평균 부하곡선들의 동일 시각에 대한 평균값(μ_i^{Gj})을 계산하여 일일 대표 부하패턴(TLP: Typical Load Profile)을 계산하게 된다.

$$\mu_i^{Gj} = \frac{\sum_{Ck=1}^{m} l_i^{Ck}}{m} \quad \text{(m: Gj 군집에 속한 고객의 수)}$$

$$TLP_{day}^{Gj} = [\mu_1^{Gj}, \mu_2^{Gj}, ..., \mu_i^{Gj}, ...\mu_{96}^{Gj}]$$

소규모 군집화가 완료되면 군집에 속한 각 고객의 일일 평균 부하곡선과 일일 대표 부하패턴과의 15분 단위의 오차를 계산하고 이를 이용하여 대표 부하패턴의 평균 오차를 계산하게 된다. Gj 군집의 중심값(μ_i^{Gj})과 Gj 군집에 속한 Ck 고객 간의 15분 단위 오차(d_i^{Gjk}) 및 Gj 군집의 평균 오차(D_{avg}^{Gj})는 아래의 수식과 같이 계산된다.

$$d_i^{Gjk} = \left| \mu_i^{Gj} - l_i^{Ck} \right|$$

$$D_{avg}^{Gj} = \frac{\sum_{i=1}^{96}\sum_{k=1}^{m} d_i^{Gjk}}{96 \times m}$$

오차에 대한 계산이 완료되면 사용자로부터 입력받은 최종 목표 군집수, 소군집 분할 개수, 군집 허용 최소 고객수를 이용하여 소규모 군집화의 반복수행 여부를 판단하게 된다. 반복적으로 군집화를 수행하면서 현재 분할되어 있는 군집의 수가 최종 목표 군집의 수보다 적은 경우, 군집의 고객수가 군집의 분할 허용 최소 고객수보다 많고, 평균 오차가 가장 큰 군집을 선택하여 k-means 알고리즘을 이용하여 소규모 군집화를 수행하게 된다. 소규모 군집화를 반복하여 최종적으로 생성된 군집의 수가 목표 군집수와 동일하게 되면 군집화를 종료하고 목표한 만큼의 대표 부하곡선의 생성을 완료하게 된다.

4. 실험 결과 분석

4.1 K-means 군집화 실험 결과

그림 4는 2887명의 고객의 자동검침 데이터를 이용하여 기존의 k-means 알고리즘만을 이용하여 10개의 군집으로 군집화한 일일 대표 부하패턴을 나타낸 그래프이다. 하나의 그래프는 하나의 군집을 나타내며, 회색실선으로 표시된 그래프는 군집에 속한 각 고객의 일일 평균 부하곡선을 나타내며, 굵은 실선으로 표시된 그래프는 군집의 일일 대표 부하패턴을 나타낸다.

표 1은 k-means 알고리즘을 이용하여 군집화한 결과 각 군집에 속한 고객의 수와 고객의 실제 부하곡선과 고객이 속한 군집의 대표 부하패턴과의 오차거리에 대한 합과 평균, 표준편차, 분산을 나타낸다. 각 군집의 일일 대표 부하패턴을 살펴보면 TLP 2, TLP 5, TLP 6, TLP 9의 경우에는 부하패턴이 유사하고 고객의 수도 높은 비중을 차지하고 있다.

4.2 반복적인 군집화 실험 결과

그림 5는 본 논문에서 제시한 반복적인 소규모 군집화를 통한 군집화 절차에 따라 소군집 분할 개수는 3, 군집 허용 최소 고객수는 100, 최종 목표 군집수는 10으로 군집화하는 과정을 나타낸 그림이다.

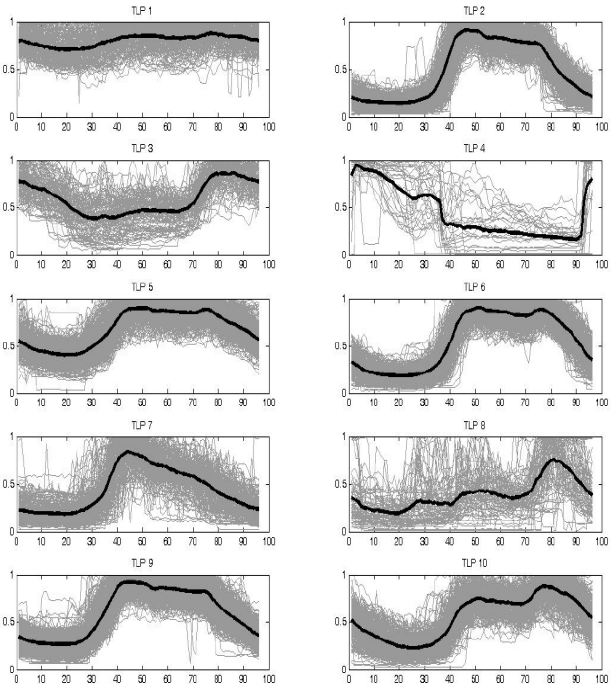

그림 4 k-means 알고리즘을 이용한 군집화 결과
Fig. 4 Clustering result using k-means algorithm

표 1 군집별 고객수 및 평균 오차거리(k-means 알고리즘)
Table 1 Number of customers and average error distance for clusters (k-means algorithm)

군집 번호	고객수	실제부하곡선과 대표부하패턴과의 오차거리			
		합	평균	표준편차	분산
1	177	249.3	1.41	0.35	0.12
2	492	468.1	0.95	0.27	0.07
3	108	258.1	2.39	0.37	0.14
4	29	134.8	4.65	0.53	0.29
5	462	494.0	1.07	0.30	0.09
6	426	421.4	0.99	0.27	0.07
7	208	425.3	2.04	0.46	0.21
8	75	297.5	3.97	0.54	0.29
9	629	571.1	0.91	0.27	0.07
10	281	451.0	1.61	0.32	0.10
전체	2887	3771.1	1.30	0.40	0.16

2887명의 고객에 대한 정규화된 일일 평균 부하곡선은 k-means 알고리즘을 이용하여 3개의 소군집으로 분할된다. 첫 번째 단계에서 3개로 분할된 군집은 각각 5.05, 1.95, 1.80의 평균 오차거리(AED)를 갖게 된다. 이 중에서 AED가 가장 큰 첫 번째 군집을 선택하여 다시 3개의 군집으로 소규모 군집화를 수행하여 총 5개의 소규모 군집이 생성된다. 5개의 소규모 군집의 AED를 살펴보면 4.28, 3.49, 2.41, 1.95, 1.80이 된다. 이 중에서 AED가 4.28로 가장 큰 군집은 고객의 수가 26명으로 군집 허용 최소 고객수보다 적으므로 배제하고, AED가 3.49로 두 번째로 큰 군집을 선택하여 다시

3개의 군집으로 소규모 군집화를 수행하게 된다. 이와 같은 방법으로 소규모 군집화를 총 5회 수행하여 최종적으로 10개의 군집에 대한 일일 대표 부하패턴을 생성하게 된다.

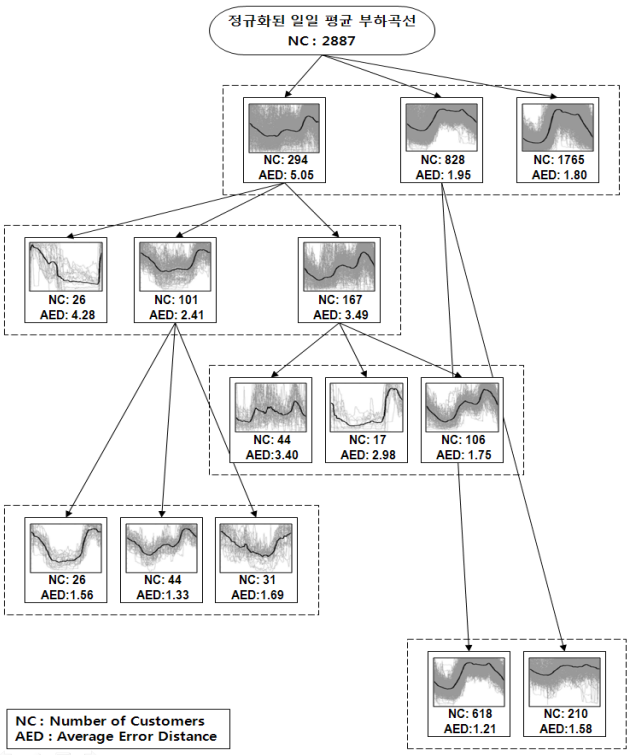

그림 5 반복적인 소규모 군집화의 진행과정 예

Fig. 5 Example of repeated small-sized clustering process

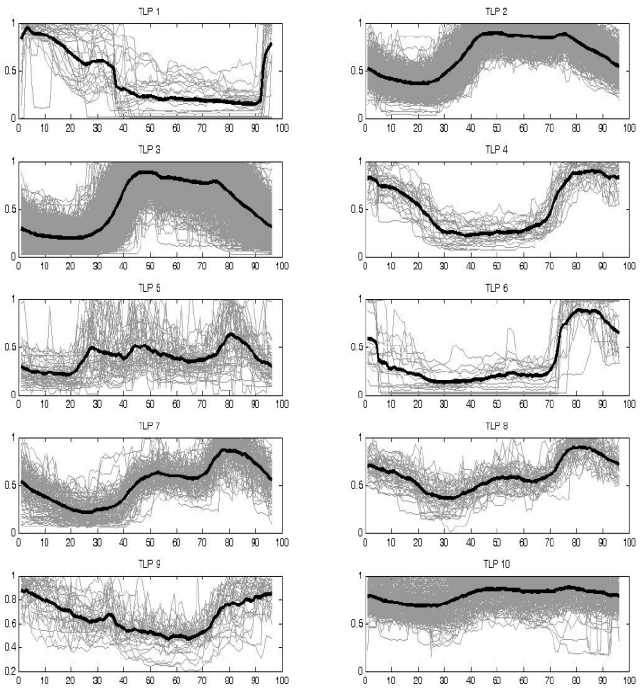

그림 6 반복적인 군집화 방법을 이용한 군집화 결과

Fig. 6 Clustering result using repeated clustering

표 2 군집별 고객수 및 평균 오차거리(제안된 방법)

Table 2 Number of customers and average error distance for clusters (proposed method)

군집 번호	고객수	실제부하곡선과 대표부하패턴과의 오차거리			
		합	평균	표준편차	분산
1	26	111.2	4.28	0.52	0.27
2	618	750.0	1.21	0.32	0.10
3	1765	3176.7	1.80	0.43	0.19
4	26	40.4	1.56	0.19	0.03
5	44	149.7	3.40	0.37	0.14
6	17	50.6	2.98	0.40	0.16
7	106	185.3	1.75	0.38	0.15
8	44	58.4	1.33	0.27	0.07
9	31	52.3	1.69	0.38	0.14
10	210	331.1	1.58	0.42	0.18
전체	2887	4906.2	1.69	0.43	0.18

4.3 코사인 유사도를 이용한 분별력 분석

표 1과 2의 전체 고객에 대한 분산을 살펴보면 제안된 방법이 k-means 방법에 비해 분산이 0.02정도 증가하는 것을 볼 수 있다. 그러나 본 논문에서는 지역편중 문제를 해결하여 대표 부하패턴의 분별력을 높이는 것을 목적으로 하고 있으므로 0.02정도의 분산의 증가는 감수하고 대표 부하패턴들 간의 코사인 유사도(Cosine Similarity)를 계산하여 두 방법의 분별력의 증가를 비교하였다. 코사인 유사도는 다차원 공간에 있는 임의의 두 점에 대해 원점에서 출발하는 방향 벡터를 만들고 두 벡터가 이루는 각도에 대한 코사인 값을 구한 것으로 -1에서 1사이의 값을 가지며 1에 근접할수록 두 벡터의 유사도는 증가하게 된다[7]. 하나의 대표 부하패턴 C1을 ($C1_1$, $C1_2$, …, $C1_{96}$) 벡터로 정의하면 두 대표 부하패턴 C1과 C2의 코사인 유사도는 아래와 같이 계산된다.

$$Cos(\theta) = \frac{C1 \cdot C2}{\|C1\|\|C2\|}$$

$$= \frac{(C1_1 \times C2_1 + C1_2 \times C2_2 + \cdots + C1_{96} \times C2_{96})}{\sqrt{C1_1^2 + C1_2^2 + \cdots + C1_{96}^2} \times \sqrt{C2_1^2 + C2_2^2 + \cdots + C2_{96}^2}}$$

표 3은 10개의 대표 부하패턴에 대해 서로간의 코사인 유사도를 계산하여 합과 평균, 표준편차, 분산을 구한 것이다. 유사도가 높아질수록 두 패턴은 비슷한 형태를 갖게 되어 분별력이 낮아지고, 유사도가 낮아질수록 두 패턴의 분별력은 높아지게 된다.

제안된 방법이 k-means 방법에 비해 대표 부하패턴의 분별력이 높다는 것을 통계적 추론(Statistical Inference) 방법의 하나인 가설검정(Hypothesis Testing) 방법으로 검정하였다[8].

- 귀무가설 H0 : 제안된 방법의 유사도(μ)는 k-means 방법의 유사도(μ_0)와 같다 ($\mu = \mu_0$)
- 대립가설 H1 : 제안된 방법의 유사도(μ)는 k-means 방법의 유사도(μ_0) 보다 작다. ($\mu < \mu_0$)

실험에서 얻은 두 방법의 대표 부하패턴간의 유사도는 정규분포를 따르지 않기 때문에 비모수(non-parametric) 방식인 Mann-Whitney 검정방법을 사용하였다[8]. 검정의 정확

도를 위해 각각 8, 10, 15, 20개의 군집으로 나누도록 실험하였으며, 10개의 군집으로 나누었을 경우를 살펴보면 유의수준 5%로 검정하면 귀무가설을 기각할 확률인 P값이 0.0415로 유의수준 0.05보다 작으므로 귀무가설을 기각하게 된다. 따라서 대립가설이 성립하여 제안된 방법이 k-means 방법에 비해 유사도가 작아지게 되므로 본 논문에서 제안된 방법이 기존의 k-means 알고리즘을 이용하는 방법에 비해서 분별력이 높은 일일 대표 부하패턴을 구할 수 있는 것을 알 수 있다.

표 3 두 방법의 코사인 유사도 및 가설검정을 위한 P값
Table 3 Cosine similarity and p-value for hypothesis testing of each clustering methods

군집수	군집방법	CosineSimilarity($Cos(\theta)$)				P값
		합	평균	표준편차	분산	
8	k-means	24.76	0.8844	0.1216	0.0148	0.0447
	proposed	23.89	0.8535	0.1201	0.0144	
10	k-means	39.75	0.8834	0.1240	0.0154	0.0415
	proposed	38.29	0.8732	0.1087	0.0118	
15	k-means	94.00	0.8952	0.1159	0.0134	0.0424
	proposed	92.45	0.8806	0.1133	0.0128	
20	k-means	171.10	0.9006	0.1189	0.0141	0.0037
	proposed	168.20	0.8853	0.1100	0.0121	

5. 결 론

본 논문에서는 고객의 일일 평균 부하곡선을 이용하여 고객에 대한 군집화를 수행하는 경우에 k-means 알고리즘만을 이용하여 군집화를 수행할 경우에 발생하는 지역편중 현상을 줄일 수 있도록 하나의 군집을 소규모의 군집으로 나누고, 평균 오차가 큰 군집을 선택하여 소규모 군집으로 나누는 작업을 반복적으로 수행하는 군집화 방법을 제안하고 있다. 본 논문에서는 실제 고객의 자동검침 데이터를 이용하여 기존의 k-means 알고리즘을 이용한 방법과 본 논문에서 제안한 방법을 이용하여 10개의 군집으로 군집화하고, 대표 부하패턴간의 코사인 유사도를 계산하여 Mann-Whitney 검정방법을 이용하여 제안된 방법이 대표 부하패턴의 분별력을 높일 수 있음을 검정하였다.

향후에는 반복적인 군집화를 위한 입력 값으로 요구되는 소군집 분할 개수나, 군집 허용 최소 고객수와 같은 값을 사용자가 입력하지 않고 최적의 군집화 결과를 얻을 수 있도록 자동으로 계산하여 사용할 수 있도록 알고리즘을 개선하고자 한다. 본 논문에서 제안한 방법을 활용할 경우 군집화를 통해 얻어지는 일일 대표 부하패턴의 분별력이 높아지게 되어 배전선로의 부하분석에 있어서 보다 정확한 분석 결과를 얻을 수 있을 것으로 판단된다.

감사의 글

본 연구는 지식경제부에서 시행한 전력산업연구개발사업인 "대수용가 서비스 모델 및 시장전략 개발, 시범사업 수행" 과제를 통해 수행되었습니다.

참 고 문 헌

[1] David Gerbec, Samo Gasperic, Ivan Smon, and Ferdinand Gubina, "Allocation of the Load Profiles to Consumers Using Probabilistic Neural Networks", IEEE Transactions on Power Systems, Vol. 20, No. 2, May 2005, pp. 548-555.

[2] Vera Figueiredo, Fatima Rodrigues, Zita Vale, and Joaquim Borges Gouveia, "An Electric Energy Consumer Characterization Framework Based on Data Mining Techniques", IEEE Transactions on Power Systems, Vol. 20, No 2, May 2005, pp. 596-602.

[3] 신진호, 김영일, 송재주, 이봉재, 이정일, "지리정보와 검침데이터를 이용한 배전계통 부하분석모델 개발", 대한전기학회 하계학술대회, 2006, 7월, pp. 2124-2125.

[4] 하복남, 한용희, 한병성, 이흥호, "배전자동화 투자비대 경제적 효과분석에 관한 연구", 대한전기학회논문지 전력기술부문A, 2003, 제52권 7호, pp. 407-413.

[5] 윤상윤, 김재철, "수용가 전력 소비 패턴을 고려한 배전용 변압기 과부하 판정기준", 대한전기학회논문지 전력기술부문A, 2004, 제53권 9호, pp. 513-520.

[6] Jain A. K. and Dubes R.C., 1988. "Algorithms for Clustering Data," Englewood Cliffs, NJ: Prentice-Hall.

[7] Van Rijsbergen, C. J., "Information Retrieval, 2nd edition", London: Buttersworth, 1979.

[8] Lehmann, E.L., and Joseph P. Romano, "Testing Statistical Hypotheses, 3rd edition", New York: Springer, 2005.

저 자 소 개

김 영 일 (金 榮 一)

1972년 11월 27일생
1998년 충남대 컴퓨터공학과 졸업
2000년 충남대 컴퓨터공학과 졸업(석사)
현재 한국전력공사 전력연구원 근무
E-mail : yikim@kepco.co.kr

송 재 주 (宋 在 周)

1967년 5월 25일생
1991년 충북대 전산통계학과 졸업
2004년 충북대 전자계산학과 졸업(석사)
현재 한국전력공사 전력연구원 근무
E-mail : jjsong@kepco.co.kr

오 도 은 (吳 度 垠)

1970년 4월 10일생
1993년 명지대 전산학과 졸업
2002년 충남대 컴퓨터과학과 졸업(석사)
현재 한국전력공사 전력연구원 근무
E-mail : baramsol@kepco.co.kr

정 남 준 (鄭 南 俊)

1966년 3월 8일생
1989년 조선대 컴퓨터공학과 졸업
2005년 충북대 전자계산학과(석사)
현재 한국전력공사 전력연구원 근무
E-mail : njjung@kepco.co.kr

양 일 권 (梁 日 權)

1954년 4월 24일생
1976년 조선대 전기공학과 졸업
1992년 미국 Indiana대 컴퓨터과학과 졸업
현재 한국전력공사 전력연구원 근무
E-mail : yangilk@kepco.co.kr