

## 자동검침 데이터를 이용한 고객 분류 기법에 대한 연구

김영일, 신진호, 송재주, 이봉재  
한국전력공사 전력연구원 정보통신그룹

### A Study for Customer Clustering Mechanism using Automatic Meter Reading Data

Young-Il Kim, Jin-Ho Shin, Jae-Ju Song, Bong-Jae Yi

Information Communications Group, Korea Electric Power Research Institute (KEPRI)

**Abstract** - 배전선로의 효과적인 운영을 위해 최근 들어 자동검침 데이터를 활용한 부하분석에 대한 연구가 진행되고 있다. 일반적인 부하분석 방식은 자동검침 고객의 데이터를 이용하여 대표 부하패턴을 생성하고 이를 이용하여 미 검침 고객의 부하패턴을 생성하여, 전체 배전선로의 회선 및 구간에 대한 15분/시간/일/주/월 단위의 최대부하 및 부하패턴 등을 분석하는 방법이다. 기존에는 고객을 분류하기 위해 계약종별 코드만을 사용하였으나, 같은 계약종별 코드를 갖는 고객이라 하더라도 부하패턴이 다른 경우가 많아서 부하분석의 정확도를 떨어뜨렸다. 본 연구에서는 고객의 계약종별 코드뿐 아니라 다양한 고객속성 정보와 15분 단위 자동검침 데이터를 이용하여 k-means 기법을 통해 고객을 분류하는 방식을 제안하였다.

#### 1. 서 론

초기의 자동검침(AMR: Automatic Meter Reading) 시스템은 단순한 검침비용 절감의 측면에서 도입되었으나, 최근들어 고객 서비스 수준을 향상시키고 경쟁력 있는 에너지 가격 정책을 유지하기 위해 다양한 분야에서 활용되고 있다. 이 시스템을 통해 고객의 에너지 사용 효율을 높일 수 있도록 돕기 위해 시간대 별로 구체적인 고객의 전력 사용량을 제공할 수 있으며, 시간대별 가격정책과 부하관리, 수요자 관리, 배급량 최적화, 배전 운용 통신 등의 분야에 활용할 수 있게 된다. 특히 부하 및 배전망 최적화 부문에 있어서는 상당한 이익을 창출하고 있으며, 최근에는 실시간 배전 최적화를 위한 응용 프로그램 개발에 대한 연구가 진행되고 있다. 배전선로의 효과적인 운영을 위해 전력사에서는 매년 막대한 비용을 투입하고 있으나, 인구밀도나 지역적인 특성에 따라 설비의 운용률과 효율성이 많은 차이를 갖게 되어 운영에 많은 어려움을 갖게 된다. 이러한 설비운영 방식은 기존의 배전계통에 대한 효율성을 분석하기 위한 상세한 데이터를 얻기가 어려워 경험적인 방식에 의한 설비운영에 의존하고 있기 때문이다. 기존의 방식은 변전소에 운영 정보를 SOMAS (Substation Operating results MAnagement System)를 통해 15분 단위 데이터를 취득하고 이를 이용하여 변전소에 연결된 회선 정보를 얻을 수 있을 뿐이었다. 자동검침시스템은 15분 단위의 설비의 운영 정보를 제공할 수 있어 보다 정확하고 빠른 정보 분석 능력을 갖출 수 있게 한다. 그러나 모든 설비에 자동검침 장비를 설치하여 데이터를 수집하는 방식은 비용적인 측면에서 또 다른 문제점을 발생시킨다. 따라서 최소한의 설비에 검침장비를 설치하고 수집된 데이터를 바탕으로 데이터 분석을 통해 각 설비에 대한 대표 부하패턴을 생성하고 이를 검침장비가 설치되지 않은 설비에 적용함으로써, 전체 배전선로에 대한 운영 정보를 얻고자 하는 연구가 진행되고 있다.

본 논문에서는 일부 고압고객에 설치된 검침장비를 통해 수집된 데이터를 이용하여 대표 부하패턴을 만드는 방식을 제시하고자 한다. 기존의 대표 부하패턴을 생성하는 방식은 고객의 계약종별 코드를 이용하여 단순 분류하고 해당 그룹에 속한 고객들의 사용량의 평균을 계산하여 대표 부하패턴으로 사용하는 방식이다. 이러한 방식은 계약종별 코드만으로는 고객의 대표 부하패턴을 생성하기 어려워 실제 사용방식과 맞지 않게 입력될 가능성이 있다. 현재 한전에서는 신규 고객을 등록할 때 계약종별, 전기사용용도, 고객의 산업종류 및 계약전력량 등과 같은 정보를 입력하여 고객의 정보를 관리하고 있다. 본 논문에서 고객의 분류를 위해 다양한 고객 정보뿐 아니라 고객의 자동검침 데이터를 이용하여 클러스터링 기법을 통해 부하 패턴이 유사한 고객을 분류하여 대표 패턴을 생성하는 방법을 연구하였다.

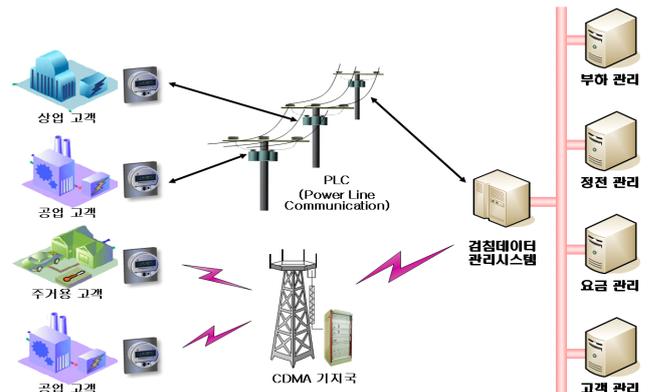
#### 2. 본 론

##### 2.1 자동검침데이터를 이용한 부하분석

고전적인 배전선로의 운영은 SOMAS를 통해 검침되는 데이터를 바탕으로 이루어졌다. 그러나 변전소 인출단(CB)에서 측정된 SOMAS 데

이터 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 15분 단위로 전력량을 계량할 수 있는 장비를 부착하고 이를 PLC와 같은 유선망이나 CDMA와 같은 무선망을 이용하여 중앙 서버로 전송하는 방식으로 그림 1과 같다. 각 배전설비를 통해 전송된 15분 단위의 데이터를 취합할 경우 시간대 별 회선 및 구간에 대한 사용량 및 피크 로드 등의 분석을 통해 배전망의 부하 분포를 이해할 수 있으며, 부하율이 높은 지역과 낮은 지역을 판단하여 배전선로의 계구성 및 추후 배전망의 확충을 위한 분석자료로 활용이 가능하게 된다.

최근에는 자동검침 데이터를 활용하여 부하를 분석하는 방식이 연구되고 있다. [1]은 약 10만호의 고압고객을 대상으로 15분 단위의 전력 사용량을 수집하여 각 계약종별 일일 대표 부하패턴을 생성하였다. 이를 활용하여 자동검침을 하지 않는 고압고객의 월 검침량을 각 계약종별 대표 부하패턴에 적용하여 15분 단위 부하 패턴을 생성하였다. 설비의 부하는 설비자체에서 발생하는 손실을 고려하지 않는다면 각 설비에서 전력을 공급하고 있는 고객들의 부하의 합과 동일하다. 따라서 설비의 15분 단위 부하 패턴은 설비에 연결된 모든 고객들의 15분 단위 검침 데이터를 이용하여 다양한 알고리즘을 통해 계산이 가능하게 된다. 본 연구에서는 Forward Sweeping 기법을 적용하여 조류 계산을 통해 회선 및 구간에 대한 15분 단위 부하 패턴을 생성할 수 있었다. 이러한 분석 방식을 통해 각 구간에 대하여 15분/시간/일/주/월 단위의 최대부하 및 부하패턴 등을 분석할 수 있으며, 이를 통해 부하절감이나 신규 수용, 과부하선로 해소, 고장선로 부하분석 등의 현장업무에 효과적으로 적용할 수 있게 되었다.

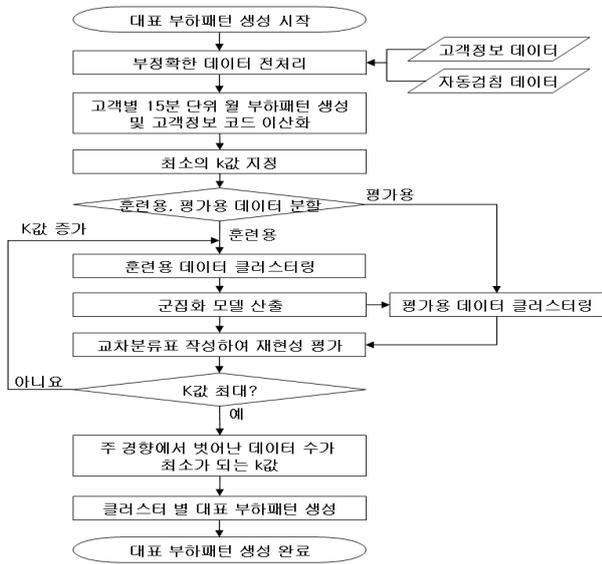


<그림 1> 자동검침시스템 구조도

##### 2.2 대표 부하패턴 생성을 위한 클러스터링 기법 적용

기존의 대표 부하패턴을 생성하는 방식은 동일한 계약종별 코드를 갖는 고객들의 15분 단위 데이터를 합하여 평균을 내는 방식으로 산출되어 왔다. 그러나 이러한 방식은 동일한 계약종별 코드를 갖는 고객이라 하더라도 대표 부하패턴과 상이한 부하패턴을 갖는 고객들이 많아 정확도가 떨어지는 문제점이 있다. 따라서 본 연구에서는 계약종별 코드뿐 아니라, 전기사용용도, 계약전력, 산업분류코드, 공급방식코드, 동력전 등구분코드 등 다양한 고객분류 정보를 추가하고, 이를 자동검침 데이터와 함께 클러스터링 기법을 사용하여 대표 부하패턴을 생성하는 연구를 수행하였다. 클러스터링 기법을 사용할 경우, 각 고객들을 객체로 보고 고객의 속성정보와 자동검침 데이터를 근간으로 각 객체의 유사성을 계산하여 그룹을 형성하게 된다.

대표 부하패턴을 생성하기 위한 일련의 과정은 그림 2와 같다. 고객 정보 데이터베이스와 자동검침 데이터베이스에서 각각의 정보를 추출하고, 전처리 과정을 통해 부정확한 데이터들을 제거한다. 각 고객의 15분 단위의 AMR 데이터는 월별로 부하 패턴을 생성하고 고객 정보 코드 값 중 연속적인 실수 값에 대해, 이산화 작업을 수행한다. AMR 데이터는 연산 속도를 높일 수 있도록 각 사용자의 월 사용량의 총합이 1이 되도록 정규화하였다. 총합이 1이 되도록 하는 정규화 방식을 선택한 이유는 대표 부하패턴을 생성할 경우 이 패턴에 해당하는 미 AMR 고객의 월 전력 사용량을 대입하여 손쉽게 15분 단위의 사용량을 계산해 낼 수 있기 때문이다.



〈그림 2〉 대표 부하패턴 계산 절차

전처리된 데이터를 클러스터링 결과에 대한 재현성 평가를 위해 훈련용 데이터와 재현성 평가용 데이터로 분할한다. 훈련용 데이터를 분류하기 위한 최소의 k 값을 지정하여 클러스터링하게 된다. 본 연구에서는 대표 부하패턴 생성을 위해 분할 방법 중의 하나인 k-means[2] 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. 이 알고리즘은 맨 처음, 각 점들을 k개 집합으로 나눈 후, 각 집합의 무게 중심을 구한다. 이후, 각각의 점들을 방금 구한 무게중심 가운데 제일 가까운 것에 연결 지음으로써 새로이 집합을 나눌 수 있다. 이 작업을 반복하면 점들이 소속된 집합을 바꾸지 않거나, 무게중심이 변하지 않는 상태로 수렴하게 된다. 최종 상태의 각 묶음을 하나의 그룹으로 선정하고 해당 그룹내의 무게중심을 대표 부하패턴으로 사용하게 된다. 최종 선정된 그룹을 이용하여 평가용 데이터를 처리하기 위한 군집화 모델을 산출한다. 산출된 군집화 모델을 이용하여 평가용 데이터를 클러스터링 하고, 두 방식의 클러스터링 결과를 토대로 교차분류표를 작성하여 재현성을 평가한다. 만약 적용된 클러스터링 모델이 최적화된 것이라면 이 표에서 행과 열은 강한 대응성을 보이게 된다. 위와 같은 방법을 k값을 증가시키면서 반복하여 주 경향에서 벗어난 데이터의 수가 최소가 되는 k 값을 선택하고, 해당 k값에서의 각 클러스터 별 대표 부하패턴을 생성하게 된다.

### 2.3 클러스터링 기법 적용결과 분석

본 연구를 위해 한전 강남지점의 2007년 1월에서 4월의 고압고객에 대한 자동검침 데이터를 이용하였다. 각 고객의 월별 데이터를 이용하여 그림 2와 같은 계산 절차와 재현성 평가를 통해 최적의 k값을 도출하였다. 각 월별로 재현성 선정을 위해 사용된 k값 중에서 주 경향에서 벗어난 데이터의 수가 적은 k값들은 표 1과 같다. 선전된 k값은 음영으로 표시하였다.

〈표 1〉 선정된 클러스터별 k값

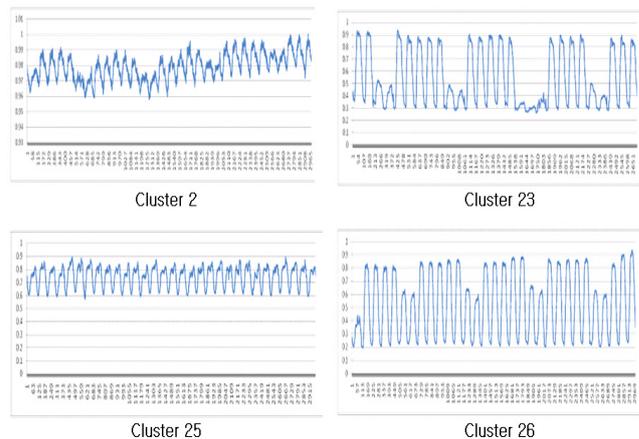
월	k	데이터수	오류 데이터수	오차율
1월	30	1553	41	2.64%
	31	1553	23	1.48%
	32	1553	23	1.48%
	33	1553	114	7.34%

2월	30	1559	102	6.54%
	31	1559	101	6.47%
	32	1559	62	3.98%
	33	1559	62	3.98%
	34	1559	148	9.49%
3월	27	1561	60	3.80%
	28	1561	46	2.94%
	29	1561	45	2.88%
	30	1561	133	8.50%
	31	1561	177	11.30%
4월	30	1146	101	8.81%
	31	1146	100	8.73%
	32	1146	23	2.00%
	33	1146	60	5.24%

본 연구에서는 기존에 고객의 계약종별 코드만을 이용하여 분류하던 방식의 정확성을 높이기 위해 클러스터링 기법을 통해 계약종별 코드만이 아닌 다양한 고객정보를 포함하여 고객을 분류하였다. 표 2를 살펴보면 같은 계약종별 코드를 갖는다 하더라도 다른 클러스터에도 할당된 것을 알 수 있다. 그림 4는 표 2에 나타난 클러스터들의 표준 부하패턴을 나타내고 있으며 이를 살펴보면 해당 클러스터에 할당된 고객들의 대표 부하패턴이 상당히 차이가 나는 것을 알 수 있다.

〈표 2〉 클러스터에 속한 고객들의 계약종별 코드 비교

클러스터번호	주택용	일반용(갑)	교육용	일반용(을)
C2	138	2141	81	16
C23	19	0	1	114
C25	2	0	0	5
C26	39	226	6	199



〈그림 3〉 클러스터의 대표 부하패턴

### 3. 결 론

자동검침 데이터를 이용하여 배전선로에 대한 부하분석을 수행하기 위해서는 부하패턴이 유사한 고객을 그룹으로 분류하여 대표 부하패턴을 생성하는 방법이 필요하다. 본 연구에서는 기존의 고객의 계약종별 코드만을 이용하여 분류하는 방식을 개선하여, 고객의 속성 정보와 15분 단위 자동검침 데이터를 추가로 이용하여 k-means 기법을 통해 고객을 분류하는 방식을 제안하였다. 클러스터링을 위한 최적의 k를 찾기 위해 신경망이나 나무 분류 및 회귀와 같은 지도학습 모델링에서 사용되는 데이터 분할 기법을 활용하였다.

향후에는 대표 부하패턴을 이용하여 자동검침을 실시하지 않는 고압고객에 부하패턴을 생성하고 이를 이용하여 배전선로의 분석 및 고장 방지, 부하절단 시뮬레이션 등의 기술을 연구할 예정이다.

### [참 고 문 헌]

- [1] 신진호, 김영일, 송재주, 이봉재, 이정일, "지리정보와 검침데이터를 이용한 배전계통 부하분석모델 개발", 대한전기학회학술대회, 7월, 2006, pp. 2124-2125.
- [2] Jain A. K. and Dubes R.C., 1988. "Algorithms for Clustering Data", Englewood Cliffs, NJ: Prentice-Hall.