

전력데이터 패턴 추출의 효율성 향상을 위한 변형된 K-means 기반의 분석 프로세스

정세훈[†], 신창선^{**}, 조용윤^{***}, 박장우^{****}, 박명혜^{*****},
김영현^{*****}, 이승배^{*****}, 심춘보^{*****}

Analysis Process based on Modify K-means for Efficiency Improvement of Electric Power Data Pattern Detection

Se Hoon Jung[†], Chang Sun Shin^{**}, Yong Yun Cho^{***}, Jang Woo Park^{****}, Myung Hye Park^{*****},
Young Hyun Kim^{*****}, Seung Bae Lee^{*****}, Chun Bo Sim^{*****}

ABSTRACT

There have been ongoing researches to identify and analyze the patterns of electric power IoT data inside sensor nodes to supplement the stable supply of power and the efficiency of energy consumption. This study set out to propose an analysis process for electric power IoT data with the K-means algorithm, which is an unsupervised learning technique rather than a supervised one. There are a couple of problems with the old K-means algorithm, and one of them is the selection of cluster number K in a heuristic or random method. That approach is proper for the age of standardized data. The investigator proposed an analysis process of selecting an automated cluster number K through principal component analysis and the space division of normal distribution and incorporated it into electric power IoT data. The performance evaluation results show that it recorded a higher level of performance than the old algorithm in the cluster classification and analysis of pitches and rolls included in the communication bodies of utility poles.

Key words: Electric Power IoT Data, K-means Algorithm, Principal Component Analysis, Pattern, Clustering

※ Corresponding Author : Chun Bo Sim, Address: (540-742) Maegok-Dong, Suncheon-si Jeollanam-do, Korea, TEL : +82-61-750-3834, FAX : +82-61-750-3830, E-mail : cbsim@sunchon.ac.kr

Receipt date : Jun. 13, 2017, Revision date : Oct. 11, 2017
Approval date : Nov. 3, 2017

[†] Dept. of Multimedia Eng., Sunchon National University (Gwangyang SW Convergence Institute)
(E-mail : iam1710@hanmail.net)

^{**} Dept. of Information Communication and Multimedia Eng., Sunchon National University
(E-mail : csshin@sunchon.ac.kr)

^{***} Dept. of Information Communication and Multimedia Eng., Sunchon National University
(E-mail : yycho@sunchon.ac.kr)

^{****} Dept. of Information Communication and Multimedia Eng., Sunchon National University
(E-mail : jwpark@sunchon.ac.kr)

^{*****} Korea Electric Power Research Institute
(E-mail : myunghye.park@kepco.co.kr)

^{*****} Korea Electric Power Research Institute
(E-mail : younghyun.kim@kepco.co.kr)

^{*****} Korea Electric Power Research Institute
(E-mail : sblee83@kepco.co.kr)

^{*****} Dept. of Information Communication and Multimedia Eng., Sunchon National University

※ This research was supported by Korea Electric Power Corporation through KEPSCO Research Institute.(Grant Number R15CA03) and The reserach was supported by 'Software Convergence Technology Development Program', through the Ministry of Science, ICT and Future Planning(No. S0177-16-1065) and This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(No.2017R1D1A3B03035379)

1. 서론

데이터 생성과 분석의 사이클이 좁아지고 있는 현재의 빅데이터 시대에는 여러 현상에 대한 분류와 분석에 대한 연구가 필요하다. 정보 통신 기술의 발전은 빅데이터 시대와 접목되어 사회의 다양한 분야에서 새로운 가치 창출을 발생시키고 있다[1]. 특히 에너지 관련 산업에서는 전력망과 정보 통신 기술을 결합한 스마트 그리드에 대한 연구가 진행되고 있다[2-4]. 그리고 전 세계적으로 다양한 산업분야에서는 전력에너지의 사용량이 증가되면서 안정적인 전력 공급과 전력시스템의 안정적인 분석이 절대적으로 필요하다. 안정적인 전력 분석 시스템은 공급적인 측면에서의 전력 분석과 수요 측면에서의 분석 관리가 절대적으로 필수한 사항이다.

전력 분석 시스템은 시간적인 순서에 따른 데이터 분석 방법인 시계열분석, 수집되는 데이터에 대한 내부요인을 분석하기 위한 상관분석, 외부요인과 같은 특징을 분석하기 위한 클러스터링 기법으로 구분할 수 있다[3]. 시계열분석은 자기회귀 모형, 이동 평균 모형, ARIMA 모형, 추세요인 분해, 계절요인 분해, 순환요인 분해, 불규칙요인 분해 등 시간적인 특징을 활용한 연구방식이 연구되고 있다. 그리고 수집되는 전력 데이터의 내부 요인 분석으로 두 데이터간의 관계 및 영향도를 분석하기 위하여 상관계수를 활용한다. 클러스터링은 수집되는 전력 데이터 내부의 특징을 분석하기 위한 기법으로 데이터간의 연결도를 통해 유사한 성격을 가진 비슷한 데이터끼리의 집단화를 형성한다. 최근 군집분석인 클러스터링 기법을 통해 전력 분석 시스템에 대한 연구가 진행되고 있다[5-9]. 특히 클러스터링 기법에서는 군집간의 민감도에 따라 발생하는 이상점(Outlier)의 추출과 이상점에 따라 전력 데이터 상의 문제점을 병행하여 분류 및 분석할 수 있다.

전력 IoT 내부 데이터는 다양한 항목으로 구성되어 있다. 본 연구에서 데이터를 분석하고 진행할 전신주 에너지 내부 데이터는 다음과 같다. 전주, 주상기 기류, 전선, 애자류, 지중구조물, 지상변압기, 지상개폐기, 지중케이블, 대기환경, 온도, 습도, 가속도, 소리감지, 초음파, 빛 감지, 지자기, 전류, 압력, 오염도가 저장된다. 그리고 다량의 에너지 데이터 항목간의 관계를 분석하고 영향도를 확인하여 내부 데이터

간의 상관관계 분석을 처리한다. 또한 검증 가설을 설정하여 전주의 기자재 및 데이터간의 관계 설정 시 이상점과 특이 구간을 추출하여, 향후 발생할 수 있는 전력 IoT 데이터에 대한 대응책을 마련할 수 있는 학습 데이터로 활용될 수 있다.

기존 연구에서 활용되는 지도학습 및 비지도학습은 다양한 문제점이 포함하고 있다. 분석의 효율성을 높이기 위하여 지도학습 기법인 상관분석은 전형적이지 않은 내부 데이터를 분석하기 위해서는 반드시 산포도를 기반으로 관계를 미리 예측한 후 관계식을 추정하여 검증받는 방법이 포함되어야 한다. 전력 IoT 데이터를 분석하기 위한 상관분석 연구의 또 다른 문제점은 특정 데이터 항목이 두개의 데이터 항목에 복합적으로 영향을 주는 경우를 포함하기 어렵다는 문제점이 존재한다. 또한 비지도학습은 전력 IoT 데이터를 분석하기 위해서는 검증 가설 및 클러스터링 기준의 민감도를 조정하여 이상점의 데이터 분류가 용이하도록 해야 하는 문제점이 있다.

이에 본 논문에서는 지도학습 기법이 아닌 비지도 학습 기법중의 하나인 K-means의 단점을 수정한 변형된 K-means를 이용하여 전력 IoT 데이터 분석의 효율성을 향상시키는 분석 시스템을 제안한다. 기존 K-means 알고리즘은 데이터 민감도를 미리 지정하여 이상점을 군집의 경계선 외부 분포 데이터로 구분하고 임의적으로 데이터를 확인하였지만, 본 연구에서 제안하는 변형된 K-means 알고리즘은 이상점의 발생 확률을 기준으로 이상점을 미리 확인하고 주성분 분석을 통하여 데이터 클러스터링의 크기를 자동적으로 결정하는 알고리즘을 소개한다. 이를 통해 다양하게 발생할 수 있는 전력 IoT 데이터를 다양한 관점에서 분석할 수 있으며, 클러스터링의 경계선을 벗어난 분포 데이터의 가설 검증과 분석 평가를 간편하도록 프로세스를 제안한다.

2. 이 론

2.1 전력 데이터 분석 모델

[2]의 연구에서는 시간 단위의 시계열 데이터를 기준으로 전력사용량 패턴 분석을 위한 군집 및 분류분석 성능평가를 제안하였다. 본 연구에서는 전력수요 유형을 나누기 위해 군집 분석을 활용하였고, 최적의 군집수의 결정을 위해 분류분석을 활용하였다. 분류

분석을 위해 기온, 강수량, 풍속, 습도, 일조량, 휴일 여부와 같은 외부요인을 고려하였으며, 의사결정나무, 랜덤포레스트, SVM, 나이브베이지 등 4가지 분류방법으로 전력수요 패턴을 분류한 결과를 제시하였으며, 랜덤포레스트 기법을 통한 분류방법의 우수한 성능을 제시하였다. 또한 기상청 자료를 바탕으로 하루 동안의 평균 전력수요패턴을 랜덤포레스트 방법을 통해 예측할 수 있는 장점을 보였다. 그러나 분류분석의 정확도를 바탕으로 휴리스틱한 방법을 이용한 최적의 군집 수 선택은 자동화된 최적의 군집 선택이 아닌 반복적인 학습을 통한 군집 수 선택 방법으로 분석 비용이 증가하는 문제점이 존재한다.

[3]의 연구에서는 다양한 클러스터링 방법을 이용하여 전력 사용량 데이터를 일간 부하 패턴과 연간 부하 패턴 두 가지 측면에서 클러스터링을 활용하였으며, KSIC(한국표준산업분류, Korean Standard Industrial Classification) 레이블과의 연관성 분석하였다. 클러스터링을 위한 군집 분석 처리 기법으로 와드 연결법의 계층적 클러스터링을 활용하였다. 전력 부하 패턴을 일간, 연간 두 측면에서 분석을 처리하였으며, 분석 시 패턴 데이터의 분포가 선형적으로 연속된 분포로 이루어져 있어 밀도 기반 방식인 DBSCAN 방식을 제외하였다. 그러나 해당 연구에서는 선형적으로 연속된 분포 시 발생할 수 있는 이상점(Outlier)의 추출은 계층적 클러스터링 기법보다는 DBSCAN 알고리즘의 이상점 추출이 더 적합한 부분이 반영되지 않았다.

[4]의 연구에서는 대규모의 전력설비에 계측장치를 설치하지 않고도 모든 전력 고객으로부터 15분 또는 월 단위로 취득되는 검침데이터를 이용하여 순간순간 변화하는 회선이하 설비의 부하를 15분 단위로 분석할 수 있는 분석 모델을 제안하였다. 부하계산 및 검중에 필요한 여러 운영 시스템을 연계하여 원격검침, 월별 검침, 변압기 부하, 회선 부하등을 추출하고, 추출된 데이터를 이용하여 패턴 분석 모델 데이터 웨어하우스를 구축하였다. 부하 패턴 예측을 위해 클러스터링과 분류, 그리고 시계열 데이터 마이닝 기법을 활용하였다.

2.2 K-means 및 Principal Component Analysis

K-means[10] 알고리즘은 군집의 수 K를 미리 정하여 각 데이터가 특정 클러스터에 배치되는지를 분

석하는 기법이며, 상호 배반적 클러스터링 기법이다. K-means 알고리즘의 절차는 다음과 같다.

- ① 사용자에 의해 미리 클러스터 수(K)를 결정한다.
- ② 선정된 클러스터 수 K개의 각 클러스터에 전체 데이터 중 한 개를 각 클러스터에 포함시킨다.
- ③ 클러스터링을 위한 모든 데이터는 거리 계산법을 통해 가장 가까운 클러스터의 중심으로 배속된다.
- ④ ③ 단계 과정을 실시한 후 배속된 데이터들의 중심을 해당 클러스터의 새로운 중심점으로 정하고 클러스터를 재 할당한다.
- ⑤ 데이터의 이동이 없을 때까지 ③단계와 ④단계를 반복적으로 실시한다.

임의적인 p 차원 상의 관찰값을 분류하기 위하여 임의의 군집 중심점 값인 $X=(x_1, x_2, x_3, \dots, x_p)$ 와 임의의 관찰값인 $Y=(y_1, y_2, y_3, \dots, y_p)$ 간의 유클리드 거리는 $d(X, Y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_p - y_p)^2} = \sqrt{(X - Y)^T (X - Y)}$ 로 정의된다. 유클리드 거리를 이용하고 군집의 중심을 그 군집에 속한 관찰치들의 평균으로 정의하므로 모든 데이터가 연속형 데이터여야 한다. K-means 알고리즘은 적절한 의 결정에 어려움이 있고 이상치 데이터가 있을 경우 초기값에 민감하다는 문제점이 존재한다.

주성분분석(PCA, Principal Component Analysis)은 분석하여 데이터의 다차원 입력 벡터를 가능한 정보 손실을 낮춘 후 낮은 차원의 벡터로 환원시키는 비지도학습 기법이다. 몇 개의 주성분 값으로 나타내는 다변량 데이터 처리 기법 중의 하나이다. n 차원의 벡터가 있을 경우, 식 (1)과 식 (2)를 적용해 나온 평균 벡터와 분산 공분산 행렬을 통해 고유벡터를 구한 뒤에 대응되는 고유값 크기에 따라 고유벡터를 정렬하여 새로운 행렬을 추가한다.

$$m_x = \frac{1}{M} \sum_{k=1}^M x_k \tag{1}$$

$$C_x = \frac{1}{M} \sum_{k=1}^M x_k x_k^T - m_x m_x^T \tag{2}$$

이 새로운 행렬은 변환 행렬로 적용해 식 (3)과 같이 벡터 x 를 벡터 y 로 변환하면, y 열에 있는 새 변수들은 비상관성을 가지며 단조 감소 분산 순서로 배열되어 분산 값이 큰 주성분들로 차원을 줄일 수

있다.

$$y = nMatrix(x - m_x) \quad (3)$$

3. 제안하는 분석 시스템 설계

3.1 변형된 K-means기반의 데이터 분석 시스템 구성

본 연구에서는 전력 IoT 데이터 분석을 위하여 비지도학습인 변형된 K-means 알고리즘[10]을 활용한다. 그러나 K-means 알고리즘은 빅데이터 분석에 적용하기에는 데이터 분류 구조가 최적화 되지 않은 단점이 존재한다. 빅데이터의 형성 조건인 Variety, Volume, Velocity에서 기존 K-means 알고리즘은 Volume과 Velocity의 관점에서 데이터를 분류하기에는 한계점이 존재한다. 또한 본 연구에서 추진하는 전력 IoT 데이터 역시 Volume과 Velocity가 빠르게 생성되고 변화는 특징이 존재하므로 기존 K-means 알고리즘을 적용하기에는 여러 가지 문제점이 존재한다. 이에 본 연구에서는 변형된 K-means 알고리즘을 적용한 분석 시스템을 설계 및 구현한다. Fig. 1은 본 연구에서 제안하는 분석 시스템 구성도이다. 전력 IoT 데이터는 빅데이터 형태를 포함하고 있기 때문에 기존 K-means 알고리즘을 통한 클러스터 분류는 정확률적인 측면에서 낮게 나타날 수 있는 문제점이 있다. 이에 본 논문에서는 다량의 전력 IoT 데이터를 군집화 할 수 있는 자동화된 K값 선택 알고리즘을 적용하여 전주의 센서노드 데이터를 분류 및

분석한다. 자동화된 K-means의 K값 추출은 입력되는 다차원 전력 IoT 데이터에서 주성분분석의 차원 축소를 통해 결정한다. 차원 축소는 전체 데이터의 80%를 설명할 수 있는 범위까지 지정하며, 80%까지 속한 주성분의 범위를 K-means의 클러스터링 수 K로 선택한다.

3.2 주성분분석을 통한 변형된 K-means 알고리즘

입력되는 전력 IoT 데이터들은 다차원 형태로 주어지고 주성분분석을 통한 차원 축소를 진행한다. 주성분분석의 공분산(Covariance)은 주어진 전체 데이터에서 선택된 클러스터간의 연관성을 확인할 수 있다. 데이터간 공분산(Covariance)을 가진다면 각각의 데이터들은 서로 독립적인 역할을 수행하며, 서로간의 연관성(Correlation)이 최소가 된다는 특징을 포함하고 있다. 주성분분석을 위해 각 차원별 입력 데이터 행렬은 $Y = (y_1, y_2, \dots, y_i)$ 이며, 초기 입력 데이터의 임의 중심은 $y_i = X_i - \bar{X}$ 로 처리된다. 공분산행렬은 $YY^T = \sum_i (X_i - \bar{X})(X_i - \bar{X})^T$ 로 정의하며, u_k, v_k 는 각각 주성분방향, 주성분요소를 정의한다.

$$YY^T u_k = \lambda_k u_k \quad (4)$$

$$Y^T Y v_k = \lambda_k v_k \quad (5)$$

$$v_k = Y^T u_k / \lambda_k^{\frac{1}{2}} \quad (6)$$

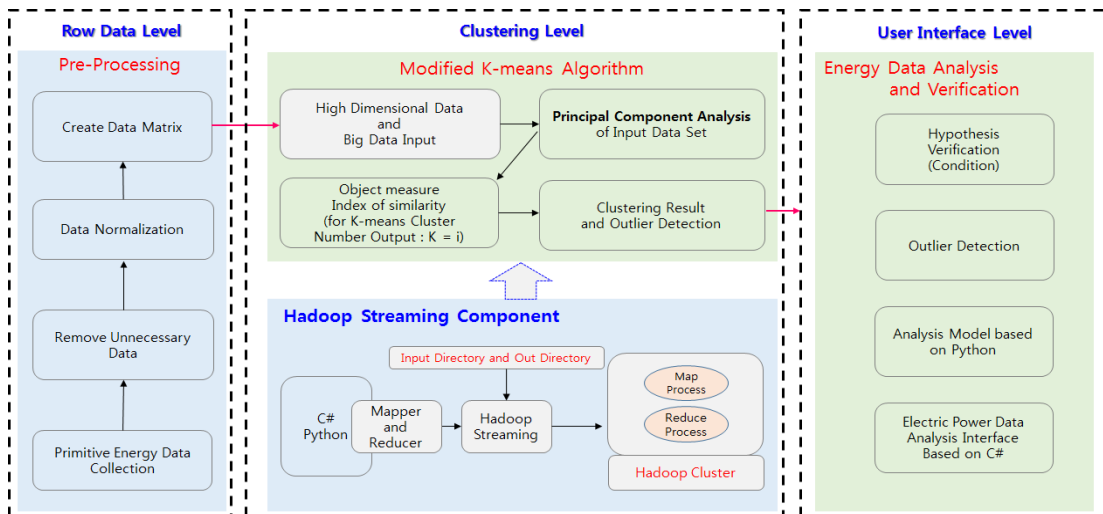


Fig. 1. Proposed Analysis System Structure.

주성분분석을 통해 클러스터간의 주성분을 확인할 수 있다. 클러스터 지표 벡터의 주성분은 v_i 로 정의되며, 식 (7)과 같이 정의한다. 초기 중심 값의 최적화된 조건에 만족하는 범위는 식 (8)과 같이 정의된다.

$$k_1 = \{i | v_1(i) \leq 0\}, k_2 = \{i | v_1(i) > 0\} \quad (7)$$

$$ny^2 - \lambda_1 < k_{k=2} < ny^2 \quad (8)$$

임의 클러스터 내의 데이터 데이터와 임의로 선택된 중심점과의 평균 거리 측정은 유클리드 거리 제곱의 합으로 정의한다. 식 (9)은 임의로 선택된 중심점과 입력된 모든 데이터 데이터와의 측정된 평균 거리이다. 식 (11)의 $S(k)$ 는 식 (9)의 임의 클러스터 내의 데이터에 대한 평균 거리인 A_k (응집도)와 식 (10)의 임의 클러스터 외의 데이터에 대한 평균 거리의 최소값인 B_k (분리도)의 차에 의한 결합으로 최대값을 주는 클러스터 최적 비유사도 $S(k)$ 가 클러스터 수 K 가 되며, 식 (12)와 같다. 단, $S(k)$ 는 -1과 1사이의 값으로 1에 가까울수록 최적화된 클러스터 수로 선택한다.

$$A_k = \sum_{i=1}^k \sum_{i \in C_k} (X_i - m_k)^2 \quad (9)$$

$$B_k = \min\{(X_i - m_k)^2\} \quad (10)$$

$$S(k) = \frac{1}{N} \sum_{i=1}^k \frac{B_k - A_k}{\max\{A_k, B_k\}} \quad (11)$$

$$S(k) = \begin{cases} 1 - A_k/B_k, & \text{if } A_k < B_k \\ A_k/B_k - 1, & \text{if } A_k > B_k \\ -1 \leq S(k) \leq 1 \end{cases} \quad (12)$$

3.3 공간분할을 이용한 클러스터의 초기 중심점 선정 알고리즘

전력 IoT 입력 데이터는 반복되는 클러스터의 초기 중심값 설정을 위하여 기존 알고리즘 방식인 임의 점 선택이 아닌 클러스터링 정확도 및 효율성 향상을 위하여 순차적인 공간분할 알고리즘을 적용한다. 초기 중심점을 선택하기 위하여 데이터간의 유사도 및 밀집도의 분포와 표준 정규 분포의 위치를 활용한다. 입력되는 전체 벡터 데이터 중 첫 번째 클러스터 C_1 의 중심점 m_k 는 식 (13)과 같이 표준 정규 분포 $\phi_{\mu, \sigma^2}(x_k y_k)$ 를 기준으로 $P(\bar{X} \geq x_k y_k) = \pm 0.9$ 이상 범위에 분포할 경우 초기 중심점에 대한 관측값 항목으로 분류한다.

$$\begin{aligned} \phi_{\mu, \sigma^2}(x_k y_k) &= \left(\frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \right) * \left(\frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} \right) \\ &= \frac{1}{\sigma} \phi\left(\frac{x-\mu}{\sigma}\right) * \frac{1}{\sigma} \phi\left(\frac{y-\mu}{\sigma}\right) \end{aligned} \quad (13)$$

4. 전력 IoT 데이터 분석 시스템 결과

4.1 실험 데이터 구성

본 연구에서는 전신주의 전력 IoT 데이터를 분류 및 분석하기 위하여 한국전력에서 제공된 2016년 4월 특정지역 전신주의 센서노드를 기반으로 제안하는 시스템의 실험을 진행한다. 전주 센서노드는 변압기 본체, 부하 개폐기 완급, 전주, 통신용함체로 구분되어 있으며, 각각 가속도, 온도, 습도로 구분하여 데이터를 수집한다. 그리고 전주 기본정보인 위치, 코드, 시설, 일시, 날짜, 시간, Pole, Position 정보를 수집하여 분석 시스템에 적용한다. 실험데이터는 총 12,280개의 전력 데이터를 활용한다.

4.2 전력 IoT 데이터 분석 결과

본 연구에서는 전신주의 전력 IoT 데이터를 변형된 K-means 알고리즘을 통해 데이터를 분류 및 분석한다. Table 1은 데이터별 클러스터링을 위한 주성분분석 결과를 나타낸다. 다차원의 데이터를 축소하여 최소로 설명할 수 있는 데이터의 클러스터링 수를 결정하기 위한 기준으로 결과는 다음과 같다.

데이터간 클러스터링을 위하여 주성분분석을 통한 결과값은 Temp.-Pitch, Temp.-Roll, Roll-Pitch의 변량 비율은 각각 PC3, PC4, PC4에 해당할 경우 전체 변수를 설명할 수 있는 80% 기준을 확보하게 된다. 이를 기준으로 클러스터링의 클러스터 수 K 값을 결정하게 된다. Fig. 2는 Temperature-Pitch 클러스터링 결과를 K 값에 따라 클러스터링의 결과가 다르게 나타내고 있다. 확인결과 $K=3$ 일 경우 Temperature-Pitch간의 분포 설명이 최적화될 수 있으며, 클러스터링의 수가 증가할수록 이상점 및 데이터

Table 1. Proportion of Variance by Variable

Part	PC1	PC2	PC3	PC4
Temp.-Pitch	0.486	0.686	0.846	0.912
Temp.-Roll	0.294	0.548	0.689	0.945
Roll-Pitch	0.348	0.457	0.784	0.891

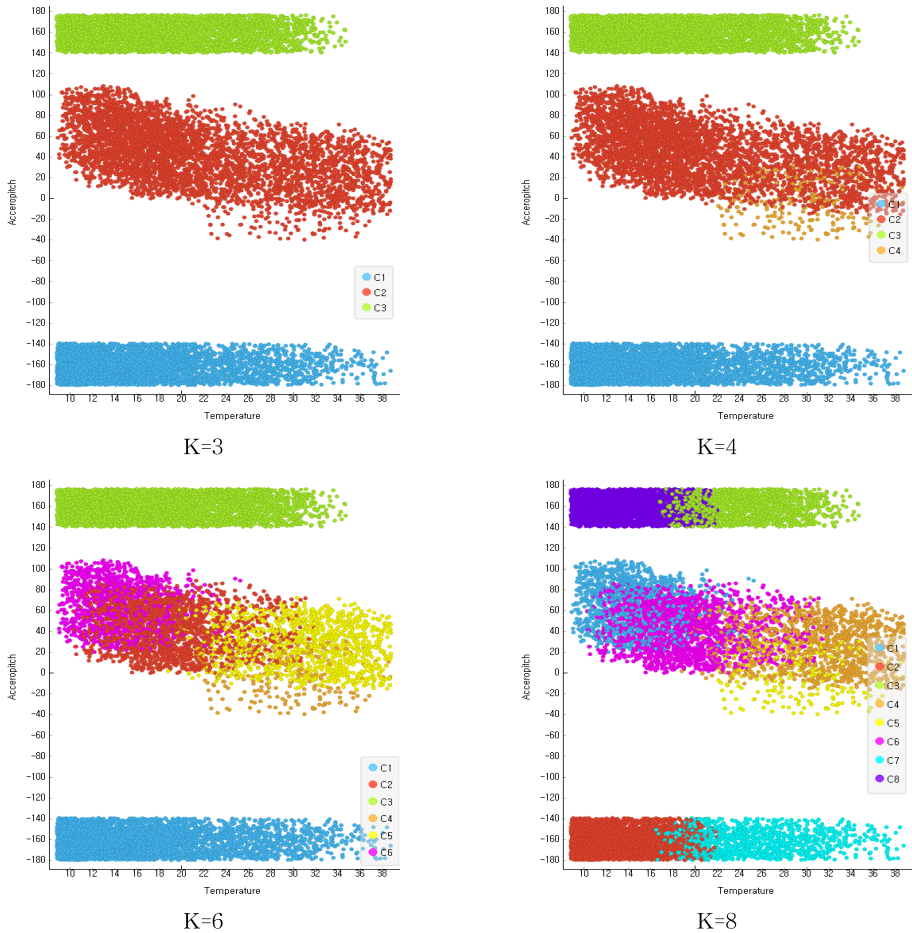


Fig. 2. Clustering Result by K Value (Temperature–Pitch).

구분 경계가 명확해지지 않은 문제점이 존재한다.

Fig. 3은 Temperature–Roll 클러스터링 결과를 K 값에 따라 클러스터링의 결과가 다르게 나타내고 있다. 확인결과 K=4일 경우 Temperature–Roll간의 이상점 분포가 최소화 되는 것으로 확인되었다. Fig. 4는 Roll–Pitch 클러스터링 결과를 K 값에 따라 클러스터링의 결과가 다르게 나타내고 있다. 확인결과 K=4일 경우 Roll–Pitch 변수를 통해 장치 외함의 움직임 여부를 조금 더 세부적으로 확인할 수 있다. K=3일 경우, Roll–Pitch 변수간의 움직임 중 Roll의 변화를 인지하지 못하는 결과를 도출하여 이상점으로 구분될 수 있는 문제점이 존재하고, K=5 또는 K=6일 경우 Pitch의 움직임을 다변화된 단계로 구분해야하는 문제점이 발생할 수 있다.

4.3 클러스터링 정확성 평가

본 연구에서는 클러스터링의 정확성 평가를 위하여 유효성 인덱스(Clustering Validity Index, CVI)를 활용하며, 평가를 위해 분류된 클러스터 내 데이터와 클러스터 중심점을 활용한다. 클러스터간 응집도의 전체 합을 클러스터들의 중심점 간 분리도로 나눈 Davies Bouldin(DB) 모형으로 CVI를 평가한다. 측정 결과 최소값을 주는 K에 대해 최적 군집 수로 결정할 수 있는 것으로 해석할 수 있다. 식 (14)는 DB 측정 지수를 구하는 정의식이다.

$$DB_k = \frac{1}{K} \sum_{i=1}^K \max_{i=1, \dots, K; j \neq i} \left\{ \sqrt{\frac{1}{n_i} \sum_{x \in C_i} d(x, z_i)^2} + \sqrt{\frac{1}{n_j} \sum_{x \in C_j} d(x, z_j)^2} / d(z_i, z_j) \right\} \tag{14}$$

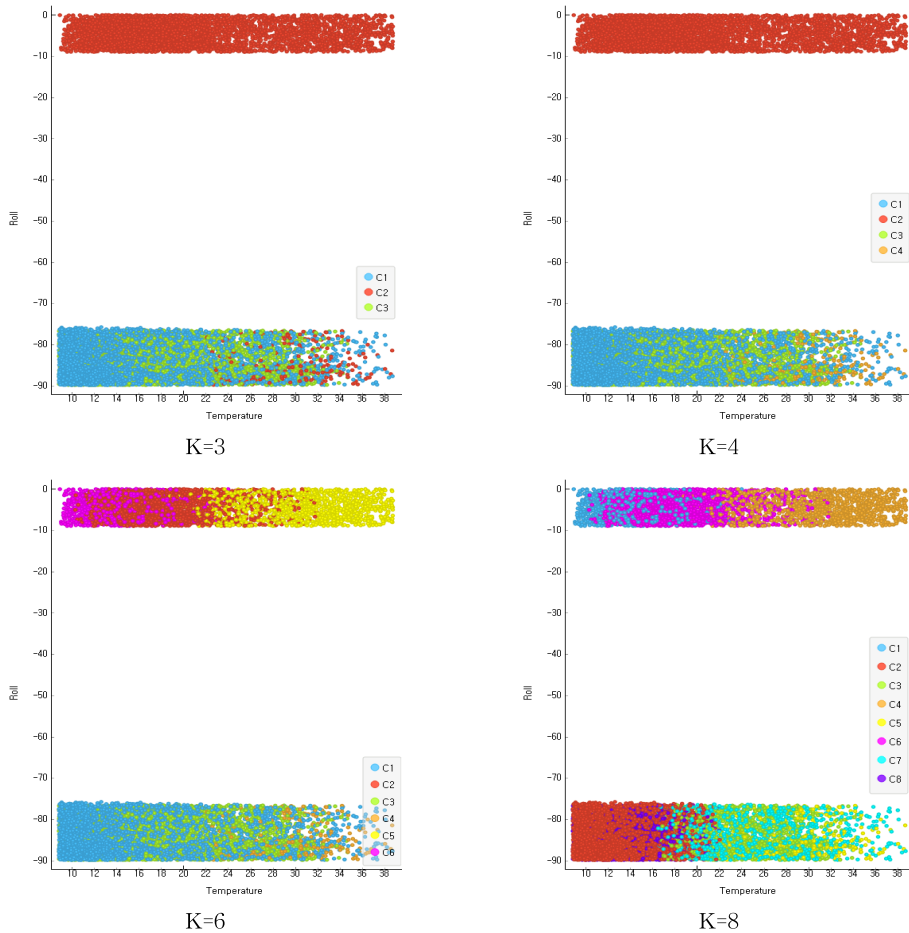


Fig. 3. Clustering Result by K Value (Temperature-Roll).

CVI를 통해 클러스터링의 유효성을 평가하기 위하여 비교분석 모델로는 클러스터링의 K값 평가 방법인 거리 측정 방식을 기반으로 하는 Silhouette 방법과 SSE(오차 제곱합)의 측정을 비교평가 하였다. Table 2는 전력 IoT 데이터 항목인 Temp.-Pitch간의 클러스터링 유효성 지수를 측정한 결과이다. DB 지수는 클러스터 수가 3일 경우(0.3314) 가장 유효성 있는 결과를 나타낼 수 있다. Silhouette와 SSE는 모두 클러스터의 수가 4일 경우로 본 연구에서 제안하는 최적의 클러스터 수와 차이가 발생하였다. Table 3은 전력 IoT 데이터 항목인 Temp.-Roll간의 클러스터링 유효성 지수를 측정한 결과이다. DB 지수, Silhouette, SSE 모두 클러스터 수가 4일 경우 (0.3314) 가장 유효성 있는 결과를 나타내고 있다. Table 4는 전력 IoT 데이터 항목인 Roll-Pitch간의

클러스터링 유효성 지수를 측정한 결과이다. DB 지수, Silhouette, SSE 각각 다른 결과를 도출하였다. 주성분분석을 통해 최적화된 클러스터 수는 K=4일 경우 이며, DB 지수를 통해 나타난 결과와 동일함을 알 수 있다.

5. 결론

본 연구에서는 전력 IoT 데이터의 패턴 추출의 효율성 향상을 위하여 변형된 K-means 알고리즘 적용한 분석 프로세스를 제안하였다. 전력 IoT 데이터는 2016년 4월 특정지역 전신주의 센서노드에서 수집된 정보를 활용하였으며, 전신주의 통신융합체에 대한 온도, 피치, 롤의 정보를 활용하여 클러스터링을 진행하였다. 주성분분석을 통해 80% 이상의 변량 비율을 설명할 수 있는 구간을 선택하여 최적의 클러스터

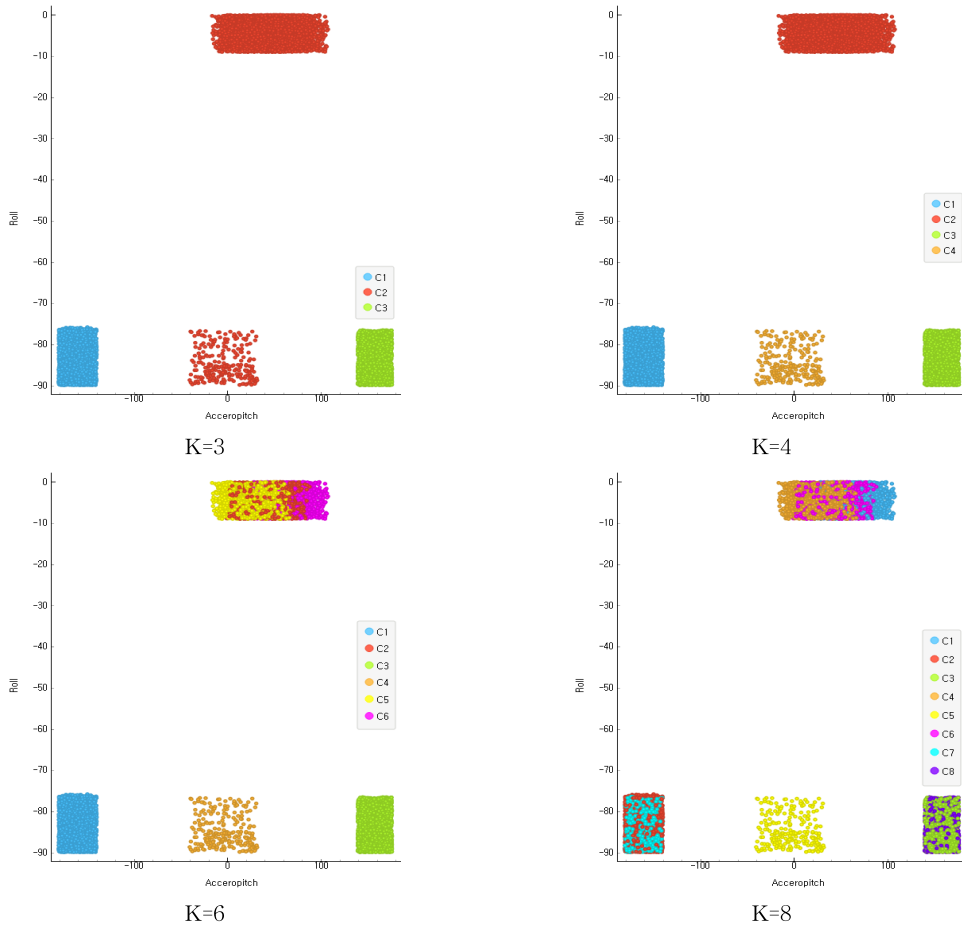


Fig. 4. Clustering Result by K Value (Roll-Pitch).

Table 2. Experiment of Clustering Availability (Temp.-Pitch)

Part	K=3	K=4	K=6	K=8
DB	0.3314	0.4855	0.5587	0.6117
Silhouette	0.7812	0.8445	0.6411	0.4844
SSE	220.1497	210.9422	168.4547	148.4548

Table 3. Experiment of Clustering Availability (Temp.-Roll)

Part	K=3	K=4	K=6	K=8
DB	0.8799	0.5315	0.6484	0.7181
Silhouette	0.8462	0.8948	0.6417	0.5478
SSE	150.4866	142.1811	128.4588	1200.1443

수 K값을 선택한다. 그리고 표준 정규 분포 기반의 공간분할 방식을 적용하여 초기 클러스터의 중심점을 선택하여 분류의 정확성을 높였다. 클러스터링 결과 클러스터의 효율성을 확인할 수 있는 DB 기반의

CVI 지수를 측정하였으며, 측정 결과 제안한 K-means 알고리즘을 통해 분류한 클러스터의 유효성 지수가 정확하다는 것을 확인하였다. 특히 Pitch와 Roll간의 클러스터 수를 기존 Silhouette 측정 결과에

Table 4. Experiment of Clustering Availability (Roll-Pitch)

Part	K=3	K=4	K=6	K=8
DB	0.4877	0.2311	0.5971	0.6132
Silhouette	0.7187	0.8482	0.6414	0.5284
SSE	198.5478	168.6233	154.5111	99.2648

따르면 K가 3일 경우를 가장 최적화로 나타냈지만 K=3일 경우 Roll의 움직임에 대한 설명변수의 이상 점이 발생하는 문제점이 나타난다. 그러나 본 연구에서 제안하는 클러스터 K값을 적용하면 Roll과 Pitch의 모든 상황을 설명할 수 있다. 향후 연구로는 전력 IoT 데이터를 비지도학습 기법으로 분석하기 위해서는 선행적인 전처리과정과 가설 설정을 통한 유의한 학습 모델을 구축하는 기술이 필요할 것으로 판단된다.

REFERENCE

- [1] S.H. Jung, J.C. Kim, and C.B. Sim, "A Novel Data Prediction Model Using Data Weights and Neural Network Based on R for Meaning Analysis between Data," *Journal of the Korean Multimedia Society*, Vol. 18, No. 4, pp. 524-532, 2015.
- [2] D.I. Park and S.H. Yoon, "Clustering and Classification to Characterize Daily Electricity Demand," *Journal of the Korean Data and Information Science Society*, Vol. 28, No. 2, pp. 395-406, 2017.
- [3] S.H. Ryu, H.S. Kim, and D.E. Oh, and J.K. No, "Customer Load Pattern Analysis Using Clustering Techniques," *KEPCO Journal on Electric Power and Energy*, Vol. 2, No. 1, pp. 61-69, 2016.
- [4] J.H. Shin, *Load Pattern Analysis Model Using Data Mining Techniques for Electric Power Facilities*, Master's Thesis of Chungbuk National University of Electromagnetics, 2010.
- [5] J.H. Choi, D.O. Kang, H.J. Lee, and C.S. Bae, "Load-based Power Analysis Model and Low-Cost Power Prediction Method for Power Management in Data Centers," *The Journal of Korean Institute of Next Generation Computing*, Vol. 10, No. 2, pp. 45-55, 2014.
- [6] J.B. Lee, *A TFP Tree-based Incremental Emerging Patterns Mining for Analysing Electricity Date*, Master's Thesis of Chungbuk National University, 2011.
- [7] M.H. Park, Y.H. Kim, and S.B. Lee, "A Study on the Development of Energy IoT Platform," *The Korean Institute of Communication and Information Sciences*, Vol. 5, No. 4, pp. 311-318, 2016.
- [8] Y.I. Son, *Demand Power Forecasting with Data Mining Method in Smart Grid*, Master's Thesis of Kyungpook National University, 2016.
- [9] S.K. Eun, *Implementation of Big Data Analysis Based Power Management System*, Master's Thesis of Kyungpook National University, 2017.
- [10] S.H. Jung, *A Novel on Hybrid Machine Learning Method Based on Big Data Mining*, Master's Thesis of Sunchon National University of Multimedia Engineering, 2017.



정 세 훈

2010년 2월 순천대학교 멀티미디어공학과 공학사
 2012년 2월 순천대학교 멀티미디어공학과 공학석사
 2017년 2월 순천대학교 멀티미디어공학과 공학박사

2015년 1월~현재 광양만권 SW융합 연구소 팀장
 관심분야: 객체지향 모델링, 상황인식 시스템, 빅데이터 처리 및 확률 분석, 데이터마이닝



박 명 혜

1993년 2월 경북대학교 전자공학 공학사
 1995년 2월 경북대학교 전자공학 공학석사
 1995년~현재 한국전력 전력연구원 책임연구원

관심분야: 유·무선 통신망 설계, 사물인터넷



신 창 선

1996년 2월 우석대학교 전산학과 이학사
 1999년 2월 한양대학교 컴퓨터교육과 교육학석사
 2004년 2월 원광대학교 컴퓨터공학과 공학박사

2005년 3월~현재 순천대학교 정보통신·멀티미디어공학부 교수
 관심분야: 분산컴퓨팅, 실시간 객체모델, 기계학습, 시계열 분석



김 영 현

2002년 2월 한국항공대학교 통신정보공학 공학사
 2004년 2월 광주과학기술원 정보통신공학 공학석사
 2004년~현재 한국전력 전력연구원 선임연구원

관심분야: 유·무선 통신망 설계, 사물인터넷



조 용 운

1995년 2월 인천대학교 전산학과 이학사
 1998년 2월 숭실대학교 컴퓨터공학과 공학석사
 2006년 2월 숭실대학교 컴퓨터공학과 공학박사

2009년 3월~현재 순천대학교 정보통신·멀티미디어공학부 교수
 관심분야: 시스템 소프트웨어, 유비쿼터스 컴퓨팅, 기계학습



이 승 배

1993년 2월 청주대학교 행정학과 인문학사
 1996년 2월 충북대학교 행정학과 경영학석사
 1993년~현재 한국전력 전력연구원 책임연구원

관심분야: 전력자동화통신망, 사물인터넷



박 장 우

1989년 2월 한양대학교 전자공학과 공학사
 1991년 2월 한양대학교 전자공학과 공학석사
 1993년 2월 한양대학교 전자공학과 공학박사

1995년 3월~현재 순천대학교 정보통신·멀티미디어공학부 교수
 관심분야: SoC, USN, 기계학습, 시계열 분석



심 춘 보

1996년 2월 전북대학교 컴퓨터공학과 공학사
 1998년 2월 전북대학교 컴퓨터공학과 공학석사
 2003년 2월 전북대학교 컴퓨터공학과 공학박사

2005년 3월~현재 순천대학교 정보통신·멀티미디어공학부 교수
 관심분야: 멀티미디어 DB, 객체지향 모델링, 유비쿼터스 컴퓨팅