



트 샘플)하였으며 사용자 ID가 없는 이상 값을 제거하는 전처리 작업을 추가 진행하였다.

## II-B. 사용자별 충전 행동 특징 추출 및 클래스 분류

시스템 모델의 EV 충전 전력량 예측 성능을 개선하기 위해, 2019년도의 사용자별 과거 충전 행동 패턴을 분석하고 유사한 클래스끼리 분류하여 제안한 시스템의 입력 특성으로 추출하는 작업을 진행하였다. 먼저 사용자 ID를 기준으로 2019년도 EV 충전 이벤트 개별 샘플  $x$ 들을 사용자별 충전 행동 패턴을 대표하는 tuple의 원소들에 기반하여 3차원 벡터로 나타내었다. 표 1은 tuple을 구성하는 3가지 원소에 관한 설명이다.

$P_{avg}$ : 개인 사용자의 EV 충전 평균 통계량
$P_{std}$ : 개인 사용자의 EV 충전 표준편차 통계량
$t_{connect\_std}$ : 개인 사용자의 EV 충전시작시간 표준편차 통계량

표 1. 사용자별 충전 행동 패턴을 대표하는 tuple의 원소

$$\underset{S}{\operatorname{argmin}} \sum_{i=1}^k \sum_{x \in S_i} |x - u_i|^2 \quad (S = \{S_1, S_2, \dots, S_k\}) \quad (1)$$

( $x$ : 사용자별 데이터 샘플,  $u_i$ : 클러스터 중심점,  $S$ 는 해당 클러스터의 집합)

또한, 충전 행동 패턴 데이터를 적절한 클래스의 수로 클러스터링하기 위해, Elbow 알고리즘 논문을 참고하여 클러스터  $u_i$  (클러스터의 개수 =  $K$ )내 오차 제곱 합의 변화율을 최소화하는 식 (1)을 고려하였다[6]. 그림 2와 같이 오차 제곱합과 연산 시간을 고려한 결과,  $K=5$ 를 바탕으로 비지도학습의 하나인 K-means 알고리즘을 이용해 정규화된 충전 행동 패턴 tuple (표 1)을 5개의 클래스로 분류하였다.

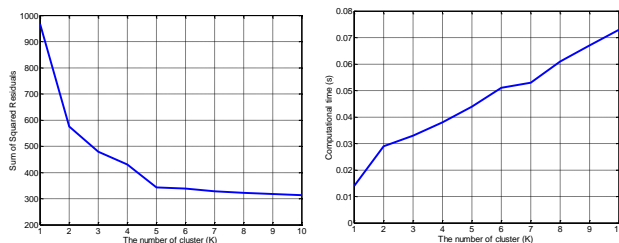


그림 2. 클러스터 개수( $K = 1 \sim 10$ )에 따른 오차 제곱합 (왼쪽)과 연산 시간(오른쪽)

분류한 결과를 바탕으로 클래스 번호와 해당 사용자 ID 리스트의 정보가 저장된 매핑 테이블을 설계하였다. 신규 사용자가 충전을 할 경우 k-NN(k-Nearest Neighbor) 알고리즘을 적용하여 가장 유사한 충전 행동 패턴을 보이는 클래스로 분류하도록 하였다. 또한 신규 사용자의 ID 정보와 클래스 번호는 매핑 테이블에 업데이트 추가하였다. 이 작업을 통해 신규 사용자가 들어올 때마다 전체 데이터 셋을 다시 클러스터링하는 것을 피함으로써 복잡도를 개선하였다. 나아가, 매핑 테이블은 원-핫 인코딩하여 EV 충전 전력량 예측 시스템의 새로운 입력 특성으로 추출하였다.

## II-C. 기계학습 기반 시스템 설계와 모의 실험 성능 비교

2020년 Caltech 대학 주차장의 15분 단위 EV 전력 수요량을 예측하기 위해 논문[4]에서 사용한 입력 특성들을 모두 사용하였다. 추가로 사용자 충전 행동 특성을 클러스터링한 매핑 테이블을 원-핫 인코딩으로 변환한 입력 특성을 추출하였다. SVM, Decision Tree 기법과 여러 트리를 독립적으로 결합한 Random Forest와 순차적으로 트리를 만들며 이전 트리의 오차를 보완한 Gradient

Boosting 알고리즘을 기반한 4가지 회귀 예측 학습 모델을 실험하였다. 모든 기계학습 알고리즘은 Grid Search CV를 통해 10번의 교차 검증(cross-validation)을 수행하여 모델 각각의 최적 하이퍼 파라미터를 설정하여 오버피팅의 가능성을 줄였다. 표 2은 RMSE의 기준에 따라 전력 그리드 망의 EV 충전 전력량 예측 시스템에서 사용한 기계학습 모델들 중, 성능이 좋은 2가지 경우를 보여준다. 모의실험 결과, 사용자별 EV 충전 행동 패턴을 분류하고 입력 특성으로 추출하는 제안 시스템이 평균 16.5%(RMSE 1.35Wh 감소)의 성능 향상을 보였다.

	기존 논문 시스템 [4]		제안한 시스템	
Algorithm	Train	Test	Train	Test
Random Forest	<b>6.91</b>	8.00	<b>4.19</b>	6.75
Gradient Boosting	6.98	<b>7.99</b>	4.70	<b>6.65</b>

표 2. 기계학습 모델 성능 비교(기준값: RMSE (kWh))

## III. 결 론

본 논문에서는 전력 그리드 망의 15분 단위 EV 충전 전력량을 정확하게 예측하기 위해, 사용자별 충전 행동 패턴을 클러스터링하여 입력 특성으로 추출한 예측 시스템을 제안했다. 사용자별 EV 충전 행동 패턴을 분류하기에 앞서, 오차 제곱 합과 연산 속도를 고려하여 클래스의 개수를 설정하였다. K-means와 k-NN 기법을 사용하여 사용자들의 리스트와 클래스에 관한 매핑 테이블을 설계하였다. 나아가, 교차 검증을 통해 하이퍼 파라미터를 설정한 기계학습 알고리즘들을 적용한 결과 Random Forest 기반 전력량 예측이 가장 우수함을 확인하였다.

이 연구는 사용자별 충전 행동 특성을 클러스터링하여 실제 전력 그리드 망의 EV 전력 예측 성능을 향상시킨 새로운 방법을 제시하고 적용함에 의의가 있다.

## ACKNOWLEDGMENT

This work was supported in part by GIST Research Institute (GRI).

## 참 고 문 헌

- [1] A. Mehrabi, H.S.V.S Kumar Nunna, A. Dadlani, S. Moon, and K. Kim, "Decentralized Greedy-Based Algorithm for Smart Energy Management in Plug-in Electric Vehicle Energy Distribution Systems", *IEEE Access*, vol. 8, 2020.
- [2] M. B. Arias and S. Bae, "Prediction of Electric Vehicle Charging Demand Forecasting Model Based on Big Data Technologies," *Applied Energy*, vol. 183, 2016.
- [3] E. S. Xydias et al., "Forecasting Electric Vehicle Charging Demand using Support Vector Machines," in *Proc. Int. Universities Power Eng. Conf.*, 2013.
- [4] Y. Chung et al., "Ensemble Machine Learning-based Algorithm for Electric Vehicle User Behavior Pattern," *Applied Energy*, vol. 254, 2019.
- [5] Z. J. Lee et al., "ACN-Data: Analysis and Applications of an Open EV Charging Dataset," in *Proc. the Tenth ACM International Conference on Future Energy Systems*, vol.19, 2019.
- [6] D. Marutho et al., "The Determination of Cluster Number at K-means Using Elbow Method and Purity Evaluation on Headline News," in *Proc. Int. Seminar on Application for Technology of Information and Communication*, 2018.