



DTW 군집화 기반 롤링 IQR 규칙 강화학습을 통한 이상치 탐지 모델에 관한 연구

권우현¹, 정민성¹, 이정섭¹, 이충호², 허태욱², *이상금¹

*국립한밭대학교¹, 한국전자통신연구원²

EcoAI Lab

Abstract

- (연구 배경) 시스템 오류나 결함으로 인해 발생하는 이상치로 인한 분석 결과 저하 및 치명적인 위험 초래를 방지하기 위한 이상치 탐지 모델 설계의 필요성 확대
- (데이터 전처리 및 환경 구성) 연속적인 시계열 데이터의 결측치를 K-NN IMPUTER로 보간 후, 윈도우를 구성하여 모델의 입력으로 사용
- (DTW기반 클러스터링) 데이터의 전력 소모 분포에 따른 값을 동적으로 분류하기 위해, 두 시계열 간의 유사도를 측정하여 군집화하는 DTW 기반의 클러스터링 알고리즘 사용
- (롤링 IQR) 롤링 IQR를 사용하여 이상치 임계값을 파라미터 K로 설정하고, 롤링 윈도우를 사용하여 각 시간대의 특성을 모델의 입력으로 구성해 학습 파라미터의 다양성을 늘리는 것에 기여
- (강화학습) 파라미터 K를 강화학습 에이전트의 정책으로 설정하고, 모델의 이상치 탐지 능력에 따라 보상을 극대화 하는 방식으로 최적의 K값 설정

Introduction & Research Background

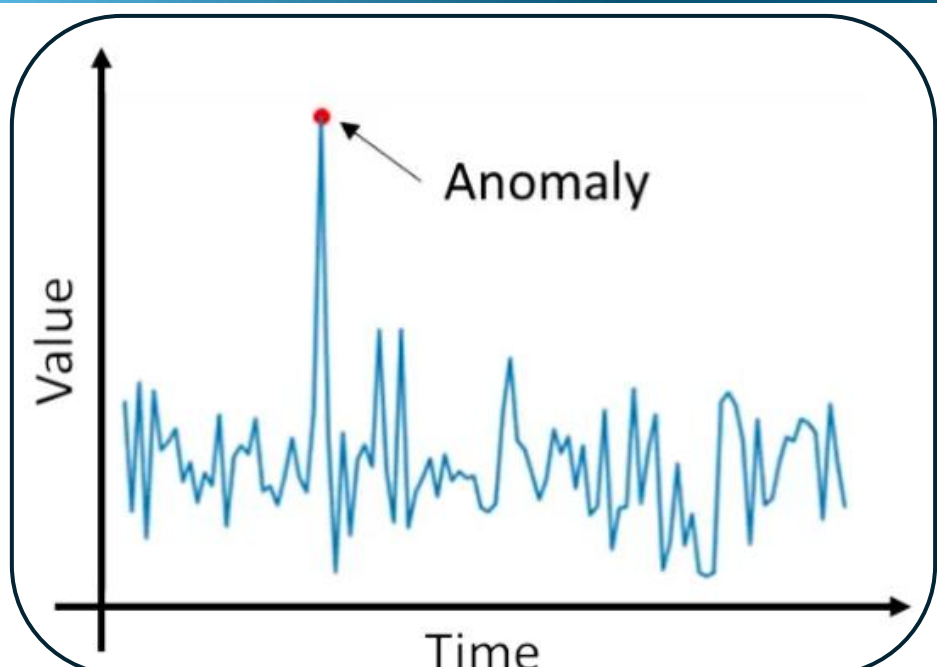


그림1. 시계열 데이터에서 이상치

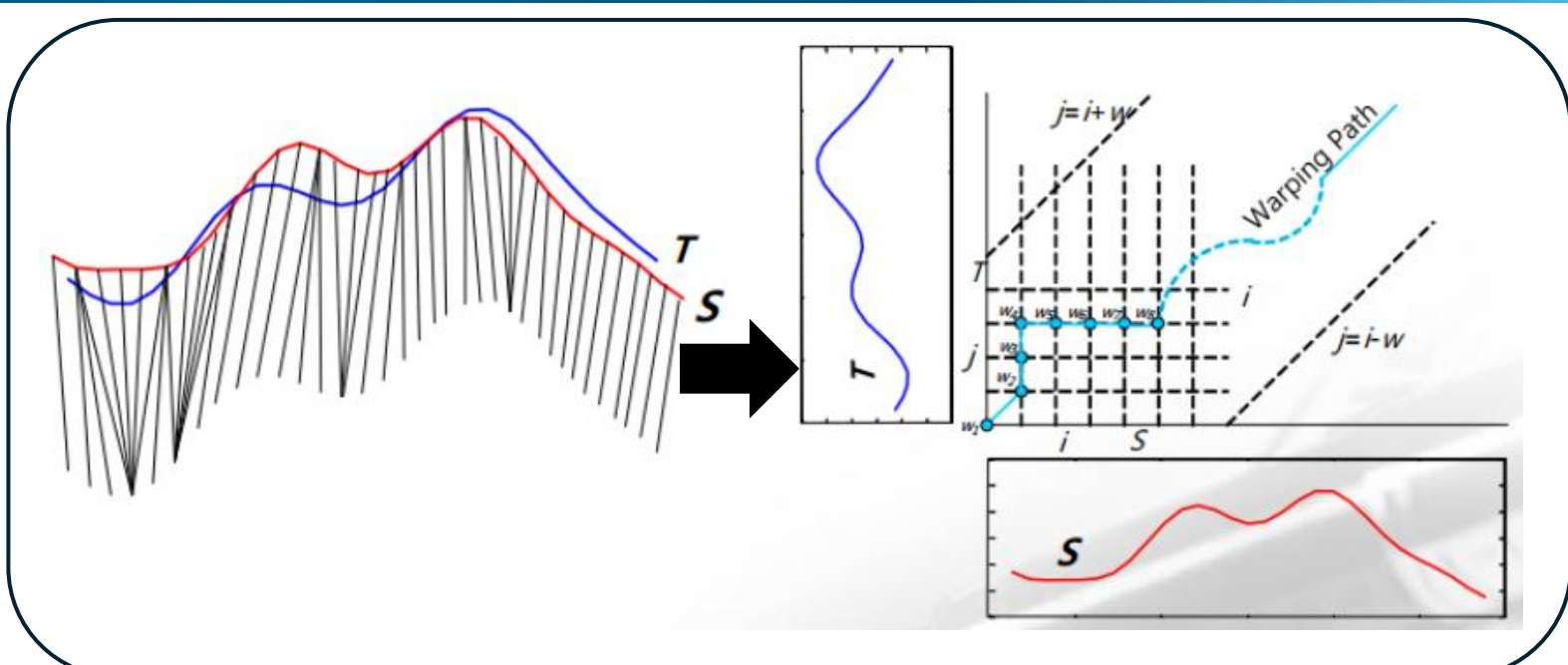


그림2. DTW 알고리즘 예시

- (그림 1) 시계열 데이터에서 이상치는 기존의 데이터 시퀀스에서 크게 벗어나는 값을 보임
- (그림 2) 이상치에 대한 탐지 정확도를 높이기 위해, DTW 알고리즘 기반의 클러스터링을 통해 데이터를 군집화하고, 롤링 IQR + 강화학습을 사용하여 파라미터 조정을 통해 모델 학습

Objective

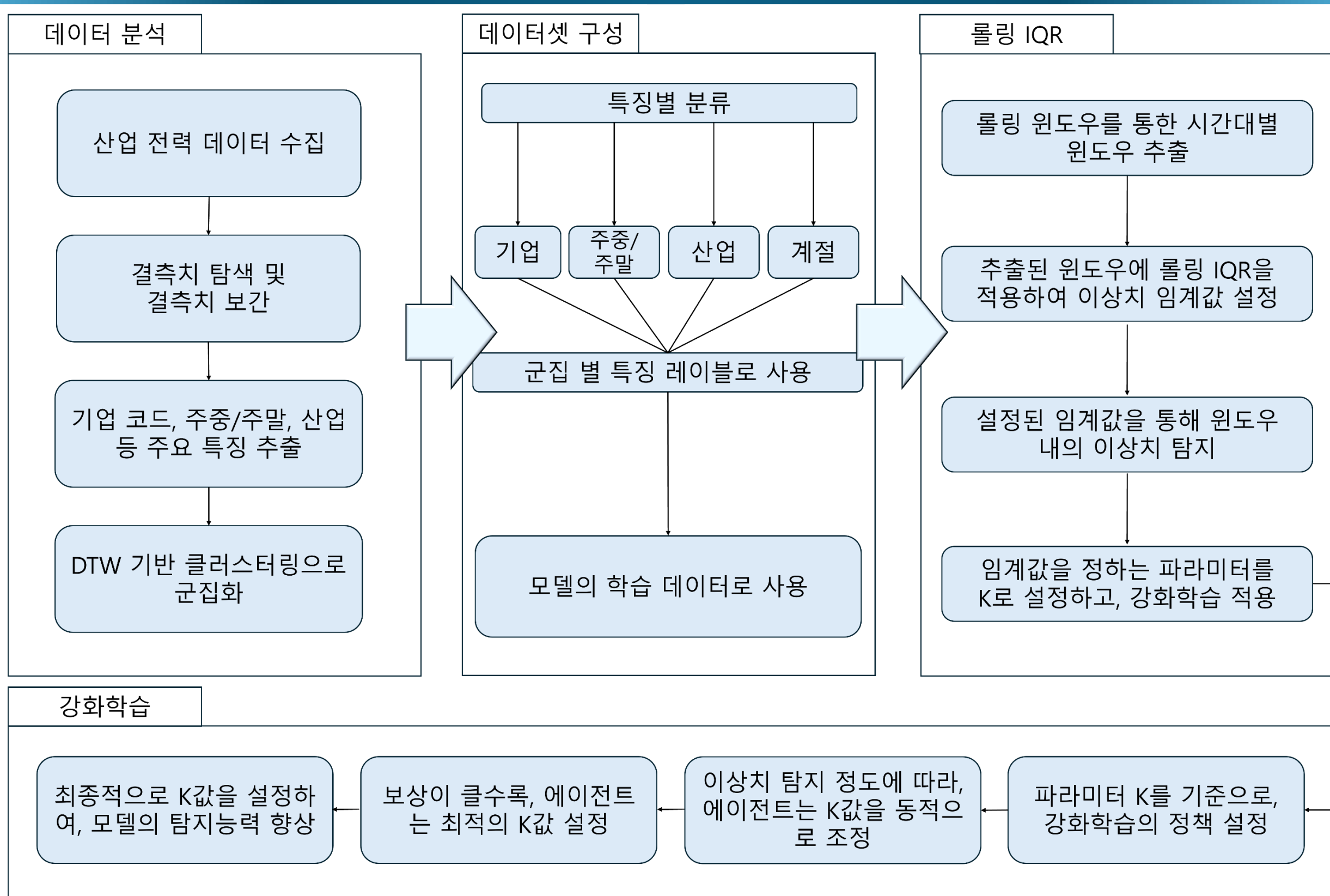


그림3. 모델 프레임워크 구조도

Materials & Method

데이터 전처리

- (2020~2022) 3년 간 수집된 산업 전력 데이터
- 기업, 주중/주말, 산업 등 사용할 특징을 기반으로 데이터셋 재구성
- 결측치 탐색 및 K-NN IMPUTER를 사용한 결측치 보간

DTW기반 클러스터링

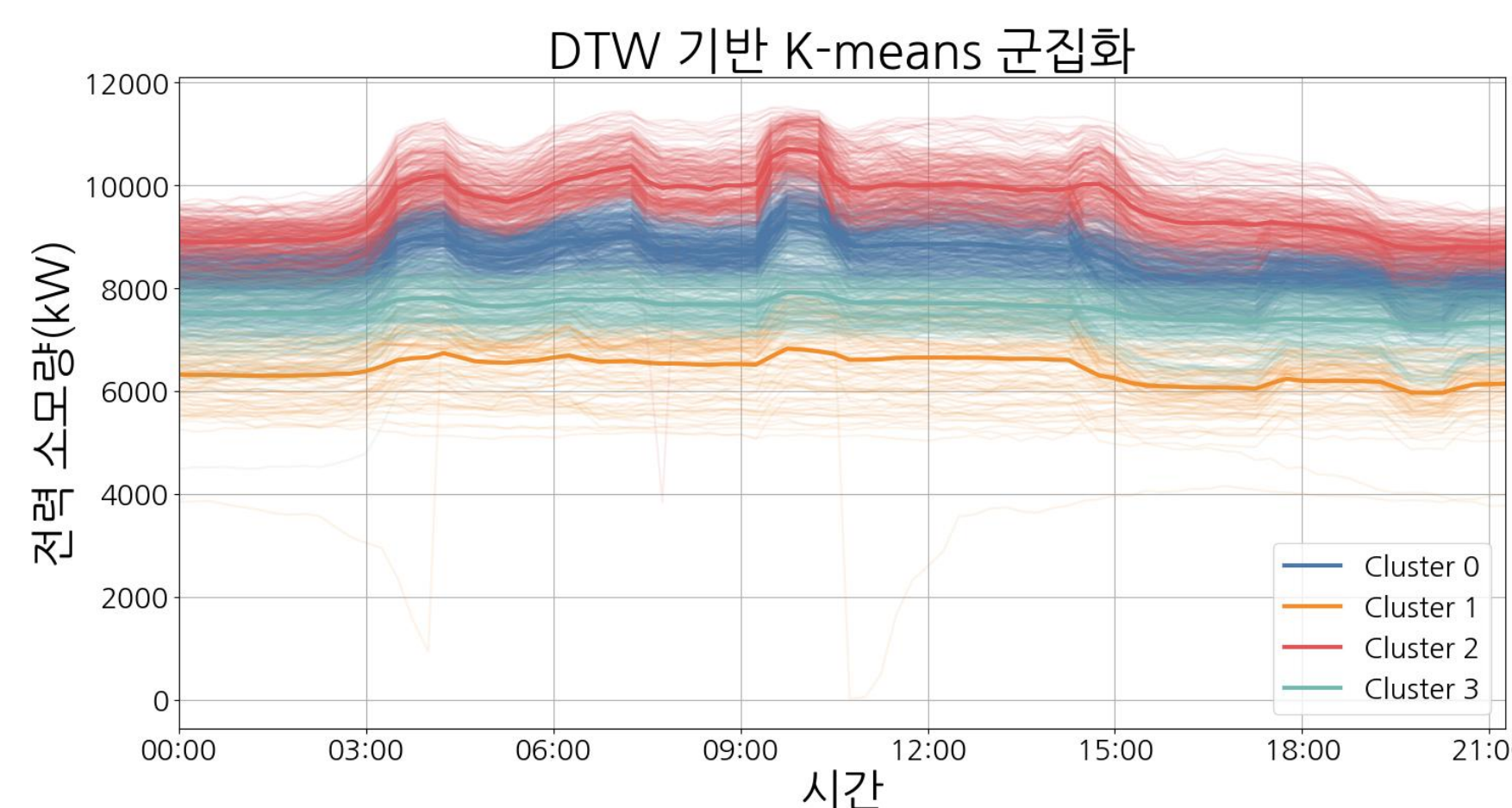


그림4. DTW 기반 K-means 군집화

- 각 색깔은 군집 별 클러스터
- 군집 별 데이터 분포
 - 6000~7000 (Cluster 1)
 - 7000~8000 (Cluster 3)
 - 8000~9000 (Cluster 0)
 - 10000~ (Cluster 2)

- (군집화) 데이터의 시계열 패턴을 안정적으로 반영하기 위해서, 시계열 시퀀스가 비슷한 데이터의 군집화를 목적으로 클러스터링을 진행
- (DTW 기반 클러스터링) 데이터의 거리를 기반으로 군집화 시 발생하는 데이터 소실 문제를 해결하기 위해 유사도를 측정하여 군집화하는 DTW 기반의 K-means 클러스터링 알고리즘 사용

롤링 IQR + 강화학습

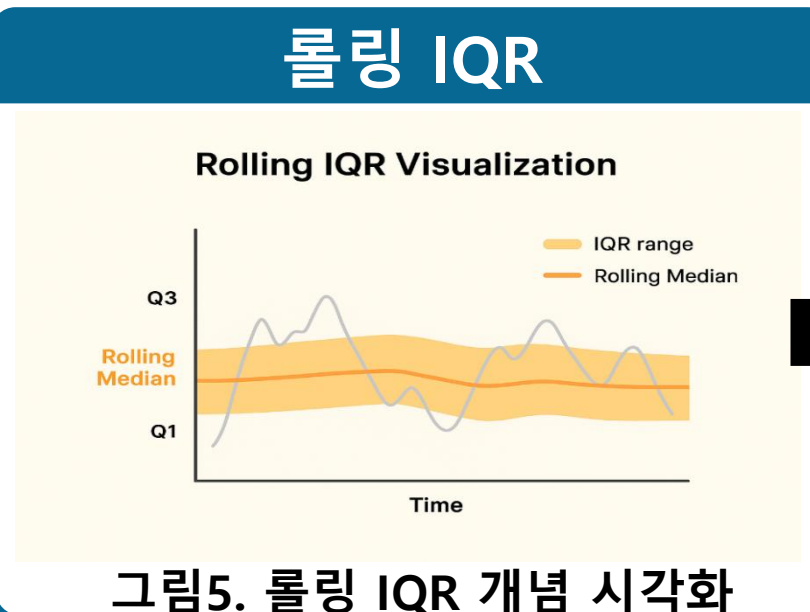


그림5. 롤링 IQR 개념 시각화

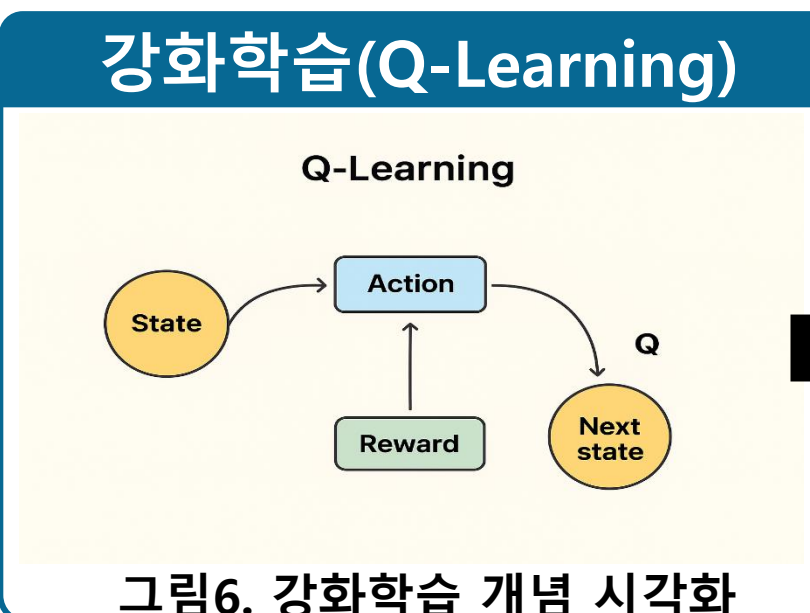


그림6. 강화학습 개념 시각화

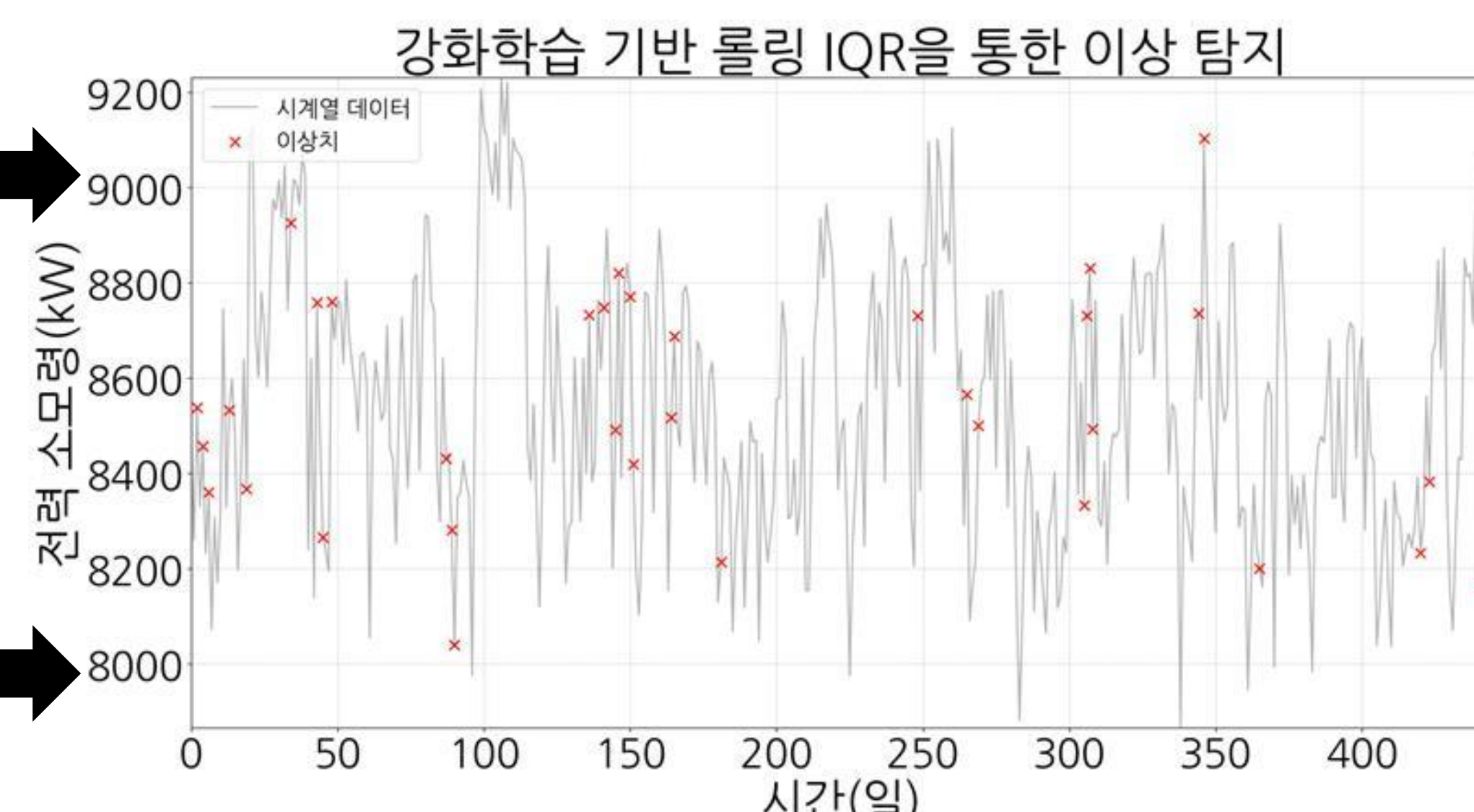


그림7. 강화학습 롤링 IQR 이상치 탐지 그래프

- (롤링 IQR) 롤링 IQR 알고리즘 내의 롤링 윈도우 개념을 적용하여 데이터셋의 시간대별 특성을 반영한 윈도우를 구성, 구성된 윈도우에 IQR을 적용하여 윈도우 별로 이상 구간 및 이상치 탐지
- (파라미터 K) 모델의 이상치 탐지 임계값 설정을 위한 파라미터 K 설정
- (강화학습) 모델의 이상치 탐지 능력을 향상시키기 위해서, 파라미터 K를 강화학습 에이전트의 정책으로 설정하고, 에이전트는 모델이 이상치를 안정적이고 정확하게 탐지할수록 보상을 극대화하면서, 최적의 K값을 설정
- (이상치 탐지) 최종적으로 이상치는 구성된 윈도우 내에서 극단적인 위치에 있는 곳이나 굴곡지고, 경사가 심한 곳에서 빈번하게 탐지됨

Result

Model	F1-Score	Precision	Recall
Rolling IQR - RL	0.85	0.83	0.87
Rolling IQR	0.78	0.75	0.82
Isolation Forest	0.72	0.69	0.76
Autoencoder	0.76	0.78	0.74

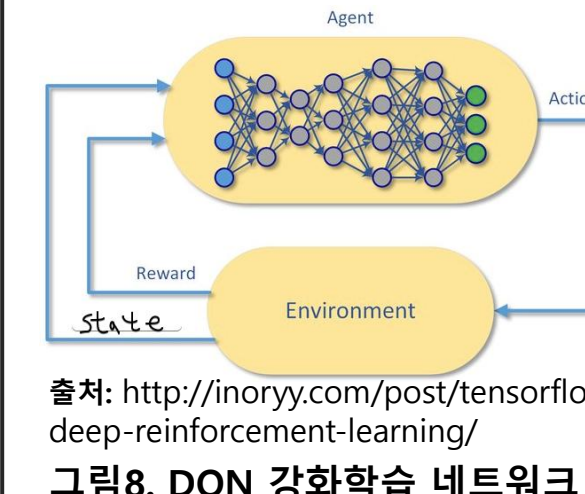
표1. 모델 성능 비교

- (표1) 표1은 기존 이상치 탐지 알고리즘인 Isolation Forest, Autoencoder와 단일 롤링 IQR 적용 시와 강화학습 기반의 롤링 IQR을 통해 F1-Score 및 Precision, Recall 점수를 산출한 표로, 롤링IQR + 강화학습 모델이 기존의 모델 대비 성능 향상을 보임
- (다양성) Isolation Forest와 Autoencoder는 각 시간대의 특성을 정밀하게 반영하지 못하는 반면, 롤링 IQR은 윈도우를 구성하여 보다 더 복잡한 파라미터를 모델의 입력으로 사용 가능
- (강화학습 적용) 이상치 판별 점수가 올랐다는 것은 강화학습 에이전트가 롤링 IQR의 파라미터 K값을 동적으로 최적화 하는 것에 성공적이었음을 시사

Future Works

- (강화학습을 통한 모델 확장) 본 연구에서 강화학습을 통하여 롤링 IQR이라는 단일 알고리즘의 매개변수를 조정하는 데 사용, 이를 확장한다면 다양한 매개변수를 정책으로 사용하여 보상과 상태를 학습시키면, 더 정밀한 모델의 확장으로 이어질 수 있음
- (통일성) 강화학습이라는 강력한 딥러닝 알고리즘은 주어진 데이터의 특징을 정책으로 설정할 수 있기 때문에, 다양한 산업군의 시계열 데이터를 통일하여 모델에 학습할 수 있다는 전망이 있음
- (스마트 팩토리) 제안된 프레임워크를 통하여, 공장에서 실시간으로 기록되는 전력 소모 시계열 데이터에서 효율적인 이상치 탐지를 기대함

(DQN 강화학습)



DQN 강화학습을 통한 모델 개선

- 더 복잡한 딥러닝 기반의 DQN 강화학습을 사용하여, 모델의 네트워크 깊이를 확장해서 더 복잡하고 다양한 특징을 학습
- 늘어난 특징은 더 강력한 모델 학습과 이상치 탐지 결과 제공을 기대

출처: <http://inory.com/post/tensorflow2-deep-reinforcement-learning/>
그림8. DQN 강화학습 네트워크 구조도