데이터 군집화를 이용한 시계열 데이터 기반 예측 모델

최영렬[1], 안나현[1,2], 임종구[3], 전상준[3],
한인수[3], 문일[2], 조형태[1], 김정환[1,†]
[1]한국생산기술연구원 친환경재료공정연구그룹; [2]연세대학교 화공생명공학과; [3]GS칼텍스(주)
(kjh31@kitech.re.kr[†])

연속적으로 운전되는 화학 공정의 시계열 데이터는 시간에 따라 다른 특성을 가지는 동적 데이터이다. 데이터 기반 예측 모델은 학습하는 데이터들의 특성에 따라 성능이 달라진다. 따라서 시계열 데이터를 이용할 경우 정확한 예측 모델을 개발하기 위해 시간에 따른 데이터들의 특성을 일정한 패턴으로 분류하여 사용해야 한다. 본 연구에서는 시계열 데이터인 2,3-BDO 증류 공정의 데이터를 군집화하고 군집화된 데이터들이 예측 모델의 학습 데이터로 사용되는 비율에 사례연구를 진행하였다. 군집 분석 방법으로 K-평균 알고리즘을 사용하였으며 Elbow 기법을 사용하여 데이터의 군집 개수를 설정하였다. 본 연구 결과, 군집화된 데이터를 학습 데이터로 사용하면 군집화를 실시하지 않은 데이터를 사용할 때보다 적은 수의 데이터를 사용하면서 더 높은 예측 성능을 보임을 보였다.