

## PP 복합수지 물성 예측 모델의 성능 향상을 위한 categorization

주종호<sup>1,2</sup>, 박현도<sup>1,3</sup>, 신은철<sup>4</sup>, 서하규<sup>4</sup>, 조형태<sup>1</sup>, 김정환<sup>1,†</sup>

<sup>1</sup>한국생산기술연구원 친환경재료공정연구그룹; <sup>2</sup>건국대학교 화학공학부; <sup>3</sup>연세대학교 화공생명공학과; <sup>4</sup>GS칼텍스(주)  
(kjh31@kitech.re.kr<sup>†</sup>)

본 연구에서는 데이터 전처리 방법으로 categorization을 이용하여 폴리프로필렌(PP) 복합수지 물성 예측 모델의 성능을 향상시켰다. PP 복합수지 recipe 데이터는 recipe마다 사용된 재료의 종류와 개수가 달라 데이터 불균형 문제를 야기하여 물성 예측 성능을 감소시킨다. 본 연구에서는 이러한 문제들을 효과적으로 해결하기 위해 전처리 방법으로 categorization을 이용하였다. Categorization이란 사용된 재료가 같은 recipe들끼리 그룹화 하는 작업으로, 분류된 category로 데이터 세트를 만들어 데이터 불균형 문제를 해결할 수 있다. PP 복합수지의 물성 specific gravity(SG), flexural strength(FS), heat deflection temperature(HDT)을 포함한 recipe는 각각 798, 803, 803개이며, categorization을 진행한 결과, 각각 496, 494, 495개의 category가 생성되었다. 예측 성능 향상을 확인하기 위해 대표적인 선형 회귀 기법인 multi-linear regression(MLR)을 이용하여 물성 예측 모델을 개발하였으며, categorization 전과 후의 데이터로 SG, FS 그리고 HDT를 예측하고 실제값과 예측값의  $R^2$ 을 계산하여 비교하였다. 비교 결과, categorization 후 데이터로 물성을 예측한 경우, SG, FS 그리고 HDT에 대해  $R^2$ 이 각각 0.0134, 0.1044, 0.0989 증가하였다.