

# 에폭시 고분자 유리전이온도 예측: 데이터 기반 접근

정예진<sup>01</sup>, 최완선<sup>02</sup>, 이동현<sup>1</sup>

<sup>1</sup> 고려대학교 수학과, <sup>2</sup> 고려대학교 보험계리금융공학협동과정  
yejin\_mds@korea.ac.kr, choi430@korea.ac.kr, holy@korea.ac.kr

## A Data-driven Approach for Predicting Glass Transition Temperature of Epoxy Polymers

Yejin Jeong<sup>01</sup>, Wanshan Cui<sup>02</sup>, Donghun Lee<sup>1</sup>

<sup>1</sup>Department of Mathematics, Korea University

<sup>2</sup>Program in Actuarial Science and Financial Engineering, Korea University

### 요약

에폭시 접착제는 다양한 산업 분야에서 중요한 역할을 하며, 그 특성은 조성 및 경화 조건에 따라 크게 영향을 받는다. 본 연구에서는 에폭시 접착제의 유리전이온도를 데이터과학적 기법으로 예측하기 위해 RANSAC과 머신러닝을 결합한 방법을 제안한다. RANSAC을 활용하여 이상치를 식별하고 제거한 후, 실험 데이터를 기반으로 각종 머신러닝 모델을 학습시켰다. 그 결과, Gradient Boosting 모델이 학습 및 테스트 데이터셋에서 각각 0.840, 0.839의 높은 결정계수를 달성하며 우수한 성능을 보였다. 이러한 연구 결과는 계속 노이즈가 큰 고분자의 물성 예측과 같은 산업적 응용에 대한 중요한 정보를 제공할 것으로 기대된다.

### 1. 서론

에폭시 접착제는 다양한 소재에 대한 강력한 접착력과 뛰어난 기계적 성질로 인해 자동차, 항공우주, 토목 등 여러 산업에서 구조용 접착제로 광범위하게 활용되고 있다. 에폭시 접착제는 주요하게 에폭시 수지와 경화제로 구성되며, 필요에 따라 촉매와 충전제와 같은 첨가제를 포함할 수 있다. 이와 같은 조성은 믹서를 통해 배합되며 상온, 고온 또는 자외선 조사 등 조건에서 경화 공정을 거쳐 접착제로 제조된다.

에폭시 접착제의 특성은 조성 원료 및 함량, 경화 시간 및 온도 등 다양한 요인에 의해 결정된다. 이는 경화 과정에서 일어나는 화학 반응에 기인한 것으로, 에폭사이드(epoxide)는 경화제와 가교 반응을 일으켜 새로운 결합을 만들어내고 최종적으로 3차원 네트워크 구조의 고분자를 형성한다[1]. 이러한 반응을 통해 기계적, 열적, 물리화학적 특성이 개선된 재료를 제조할 수 있다. 때문에 용도에 적합한 접착제의 설계를 위해 다양한 에폭시 조성 및 공정에 대한 연구는 중요한 의의가 있다.

동적 기계적 분석(Dynamic Mechanical Analysis; DMA)은 고분자의 점탄성 거동을 분석하는 데 필수적인 도구로써 가교 네트워크를 연구하는 데 중요한 역할을 한다. DMA는 시편에 진동력을 가해 온도, 시간 및 주파수의 변화에 따라 재료의 반응을 측정하는 기술이다. DMA로 측정되는 유리전이온도는

고분자의 열적 특성이 급격히 변하는 지점을 나타내는 물성으로, 에폭시 접착제의 특성화에 관건적인 정보를 제공한다[2].

데이터 기반 머신러닝은 편향(bias)에 민감하기 때문에 데이터 품질은 모델의 학습과 정확도에 상당한 영향을 미친다. 데이터셋 내의 이상치를 식별하고 처리하는 것은 강건한 모델 구축과 신뢰도 높은 예측을 위해 매우 중요하다[3]. 화학 실험에서는 변수를 통제하기 위한 노력에도 불구하고, 인적오류나 실험의 변동성 등 요인이 여전히 발생할 수 있다. 이러한 요인들은 실험의 품질 및 생성물의 특성에 직접적인 영향을 미칠 수 있다.

본 연구에서는 임계값에 근거한 이상치 탐지 기술을 사용하여 데이터의 변동성을 완화하고 머신러닝을 활용하여 강건한 모델을 구축하여 다양한 에폭시 조성과 공정 조건에서의 유리전이온도를 예측하고자 한다. 이를 통해 이상치의 영향을 최소화하고 예측 신뢰성을 향상시키는 것을 목표로 한다.

### 2. 관련 연구

최근 연구에서는 고분자의 다양한 특성을 예측하기 위해 머신러닝 모델을 점점 더 많이 활용하고 있다. 특히, DMA 데이터를 활용하여 유리전이온도를 추정하는 것이 그 중요한 목표 중 하나로 떠오르고 있다[4]. DMA 데이터를 활용하여 이러한 특성을 예측하면 고분자의 열적 및 기계적 특성에 대한 이해와 예측 능력이 향상되어, 제품의 설계 및 개발에 귀중한 도구로 활용될 수 있다.

고분자 과학 분야에서는 실제 실험 데이터셋을 구축하는 것이 비용과 시간이 매우 많이 소모되는 작업이다. 따라서, 선행

\* 본 연구는 산업통상자원부와 한국산업기술진흥원의 소재부품산업기술개발 기반구축사업(과제번호 P0022334)으로 수행되었습니다.

<sup>0</sup>: These authors contributed equally.

연구에서는 소규모의 데이터셋을 사용하거나 실험 데이터를 보완하기 위해 컴퓨터 시뮬레이션 또는 웹 스크래핑 기술을 활용해왔다[5]. 그러나 본 연구에서는 실제 실험 데이터셋만을 기반으로 머신러닝 모델을 활용하여 유리전이온도를 예측하고자 한다.

다만, 다양한 외부 요인으로 인해 데이터에서 오차가 발생할 수 있는 상황이므로, 데이터 전처리 과정이 매우 중요하다. 따라서 본 논문에서는 RANSAC 을 사용하여 이상치를 식별하고 제거함으로써 주어진 데이터의 안정성을 전처리로 높인 뒤에 모델을 추정하고자 한다.

### 3. 실험 방법

#### 3.1 데이터 전처리 모델

RANSAC(RANdom SAmples Consensus)은 이상치로부터 영향을 최소화하고 정확한 모델을 추정하는 데 사용되는 반복적인 알고리즘이다[6]. 먼저, 무작위로 샘플을 선택하여 모델을 학습하고, 이를 사용하여 잔차를 계산한다. 그런 다음, 잔차가 특정 임계값보다 작은 데이터 포인트들을 사용하여 새로운 모델을 학습한다. 이 과정을 반복하여 최적의 모델을 찾는다. 본 연구에서는 이렇게 학습된 모델을 기반으로 특정 임계값이 넘어가는 데이터를 제거하는 전처리 과정을 진행하였다. 이는 고분자와 같이 데이터의 오차가 발생하기 쉬운 경우에 특히 유용하며, 데이터의 신뢰성을 높이고 정확한 분석을 위해 사용하였다.

#### 3.2 학습 모델

선형모델, 비선형모델, 앙상블 모델, 그리고 AutoML 모델은 각각의 장단점을 가지고 있으며, 데이터의 특성과 예측 목표에 맞게 적합한 모델을 선택하는 것이 중요하다.

선형모델인 Linear Regression 은 특징 간의 선형 관계를 가정하여 예측을 수행한다. 이 모델은 간단하고 해석하기 쉬우며 대규모 데이터셋에 대해 효율적으로 작동한다. 그러나 데이터의 복잡한 비선형 관계를 잘 표현하지 못하는 경우가 있다.

비선형모델인 KNN(K-Nearest Neighbors)은 데이터 포인트 간의 유사성을 기반으로 예측을 수행한다. 이 모델은 주변 데이터 포인트의 정보를 활용하여 예측을 하기 때문에 데이터의 로컬 패턴을 효과적으로 포착할 수 있다. 따라서 선형 모델보다는 데이터가 비선형인 경우 예측 정확도가 향상될 수 있다.

앙상블 모델인 Gradient Boosting 은 여러 개의 결정 트리를 순차적으로 학습하여 예측 성능을 향상시키는 방식이다. 이 모델은 이전 트리의 오차를 보완하고 점진적으로 더 나은 예측을 위해 학습한다. 이 모델은 복잡한 데이터 구조와 상호작용을 잘 모델링할 수 있으며, 일반적으로 다른 모델보다 더 높은 예측 정확도를 제공한다.

AutoML 의 대표적인 도구인 TPOT(Tree-based Pipeline Optimization Tool)[7]는 자동화된 방식으로 여러 가지 머신러닝 파이프라인을 탐색하고 최적화한다. 이 모델은 다양한 모델과 전처리 단계를 조합하여 최상의 예측 성능을 내는 방식이다. 이를 통해 모델 최적화 과정을 간소화하고, 더 나은 예측 결과를 쉽게 얻을 수 있다. TPOT 는 전통적인 머신러닝

방법에 비해 시간 및 예측 측면에서 효율성을 높이는 획기적인 방법론으로 평가된다.

### 3.3 데이터셋

본 실험에 사용된 에폭시 접착제 물성데이터를 구성하기 위하여 서울대학교 재료공학과 유기재료 설계 및 합성 연구실에서 재료합성, 서울대학교 산림과학부 접착과학 및 바이오복합재료 연구실에서 물성계측을 담당하였다. 전체 1,087 개의 실험 데이터를 사용했으며 수집한 데이터를 바탕으로 DMA 에서 측정된 유리전이온도 값 ( $158.7^{\circ}\text{C} \pm 7.8^{\circ}\text{C}$ )을 종속변수로, 시편 제작에서의 배합 및 경화 공정 변수를 기반으로 총 18 개의 설명변수를 구축하였다.

조성은 에폭시 수지, 경화제, 촉매로 구성되었으며 이들은 다양한 비율로 실험에 사용된다. 각 원료의 투입량(g)과 함량(wt%)으로 총 12 개의 설명변수를 산정하였다. 에폭시 수지는 총 네 가지로 이들은 단일 원료 또는 두 가지로 혼합되어 실험에 사용되었다. 그 중 DGEBA(YD-128, 국도화학), CTBN(KR-450, 국도화학), MBS(KDAD-7101, 국도화학), dimer 변성 에폭시 수지(YD-172, 국도화학)는 각각 전체 데이터에서 98%, 20%, 28%, 10%의 비중으로 사용되었다. 경화제와 촉매로는 각각 Dicyandiamide (DYHARD® 100S, AlzChem)와 1,1'-(4-methyl-m-phenylene) bis(3,3'-dimethylurea) (DYHARD® UR500, AlzChem)를 사용하였다. 배합을 마친 조성은 다양한 온도( $90^{\circ}\text{C}$ - $160^{\circ}\text{C}$ )와 시간(0.5h-3h/회)으로 최소 1 회에서 최대 3 회의 경화 과정을 거치게 되는데, 이때의 조건 값을 6 개의 설명변수로 산정하였다.

### 4. 실험 결과

#### 4.1 데이터 전처리 결과

RANSAC 를 활용하여 전체 1,087 개의 데이터 중 일부를 선별하였다. 이 과정에서는 기본적으로 Linear Regression 모델을 사용했으며, 최소한의 샘플 수는 전체 데이터셋의 36%를 사용하도록 설정했다. 또한, 이상치를 식별하고 제거하기 위해 잔차의 임계값을 4.0 으로 설정하였다. 이 임계값은 유리전이온도의 특성을 고려하여 결정되었다. RANSAC 모델에서 inlier 로 선택된 데이터를 최종 데이터셋으로 사용했다. 이 과정을 통해 최종적으로 총 649 개의 데이터가 선택되어 유리전이온도 예측 모델 학습과 테스트에 사용되었다.

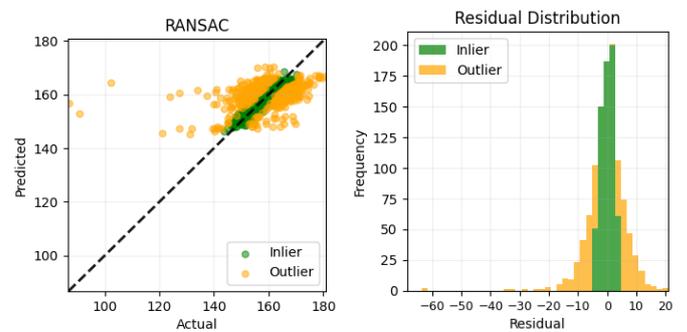


그림 1. RANSAC 에서 임계값을 기준으로 분리된 Inlier 와 Outlier 의 (좌) 유리전이온도 예측값과 (우) 잔차의 분포

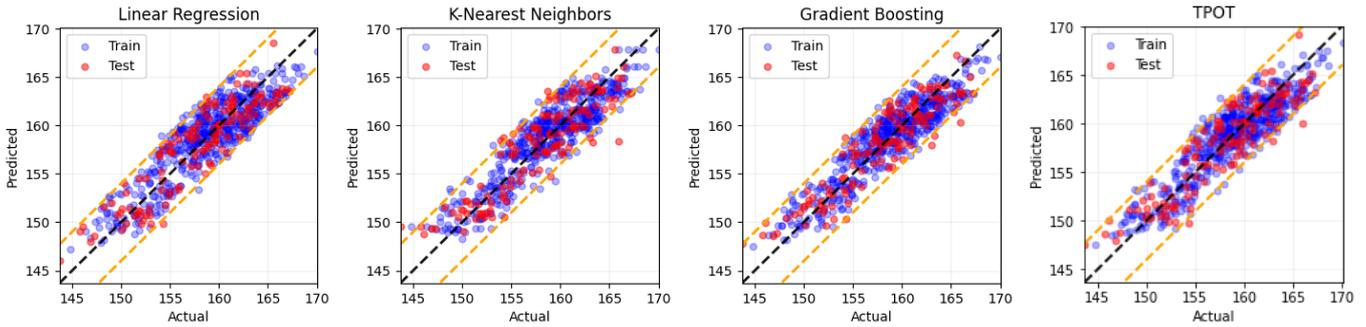


그림 2. 모델별 회귀 산점도 (좌측부터 Linear Regression, K-Nearest Neighbors, Gradient Boosting, TPOT)

표 1. 모델별 유리전이온도 예측 결과

Model	Train		Test	
	RMSE (°C)	$R^2$	RMSE (°C)	$R^2$
Linear Regression	2.1505	0.782	2.2088	0.815
K-Nearest Neighbors	1.9474	0.821	2.3338	0.794
Gradient Boosting	<b>1.8438</b>	<b>0.840</b>	<b>2.0651</b>	<b>0.839</b>
TPOT	1.9843	0.814	2.0658	0.838

#### 4.2 모델 학습 결과

다양한 머신러닝 모델을 사용하여 유리전이온도에 대해 회귀 예측한 결과(표 1), RMSE가 약 2°C로 산출되었다. 이는 앞서 RANSAC에서 설정한 4°C의 임계값보다 월등히 낮은 수준이다. 데이터 포인트별 예측 결과는 그림 2에서 확인할 수 있는 바, 대부분의 경우 유리전이온도 실제 측정치와 4°C 이내의 안정된 범위 내에서 일치하는 것으로 나타났다. 이는 화학 영역에서의 경험적 지식을 바탕으로 설정한 임계값으로, 결과를 통해 해당 연구에서 제시한 방법론의 타당성을 더욱 강조한다. 사용한 4종의 모델 중 Gradient Boosting 모델은 낮은 RMSE와 높은 결정계수로 학습 및 테스트 데이터셋에서 모두 우수한 성능을 보여주었다. 따라서 해당 데이터셋에서는 Gradient Boosting 모델이 가장 적합하다고 평가된다.

#### 5. 결론 및 고찰

본 연구에서는 다양한 에폭시 고분자 조성 및 공정으로 구성된 데이터를 기반으로, 에폭시 고분자의 물성 중 하나인 유리전이온도에 대한 회귀 분석을 수행하였다. 예측 결과, Gradient Boosting은 테스트 데이터셋에서 2.0651°C의 RMSE와 0.839의 결정계수를 보여주어 높은 모델 예측력과 설명력을 나타냈다.

전처리 단계에서 RANSAC을 활용하여 임계값 기반의 데이터 샘플링을 통해 이상치를 식별하고 유효한 데이터를 추출하였으나, 이 과정에서 상당량의 데이터가 손실되어 실제 모델 학습에 사용된 데이터가 제한적이다. 이에 대한 구체적인 원인을 분석하고 이를 감소하는 방법에 대해 추가적인 논의가 필요하다. 향후 연구에서는 고분자 데이터의 불안정성과 이상치에 대한 분석을 보다 심층적으로 수행하여 더욱 견고하고 정확한 예측 모델을 개발할 것이다. 또한, 다양한 실험적 접근과 알고리즘의 적용을 통해 데이터의 보존을 강화하고 예측 성능을 향상시키는 방안을 모색할 예정이다.

#### 참고문헌

- [1] Rudawska, A., & Czarnota, M. (2013). Selected aspects of epoxy adhesive compositions curing process. *Journal of Adhesion Science and Technology*, 27(17), 1933-1950.
- [2] Menard, K. P., & Menard, N. (2020). *Dynamic Mechanical Analysis* (3rd ed.). CRC Press.
- [3] Smiti, A. (2020). A critical overview of outlier detection methods. *Computer Science Review*, 38, 100306.
- [4] E. Alcobac a, S. M. Mastelini, T. Botari, B. A. Pimentel, D. R. Cassar, A. C. P. de Leon Ferreira de Carvalho, and E. D. Zanotto. (2020). Explainable machine learning algorithms for predicting glass transition temperatures. *Acta Materialia*, 188, 92-100.
- [5] Kang, H., Lee, J. H., Choe, Y., & Lee, S. G. (2021). Prediction of lap shear strength and impact peel strength of epoxy adhesive by machine learning approach. *Nanomaterials*, 11(4), 872.
- [6] Fischler, M. A., & Bolles, R. C. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6), 381-395.
- [7] Olson, R. S., & Moore, J. H. (2016). TPOT: A Tree-based Pipeline Optimization Tool for Automating Machine Learning. *Proceedings of the Workshop on Automatic Machine Learning*, in *Proceedings of Machine Learning Research* 64, 66-74.