

Study of Polymor Properties Prediction Using Nonlinear SEM Based on Gaussian Process Regression

Moon Kyung-Yeol[†] · Park Kun-Wook^{††}

ABSTRACT

In the development and mass production of polymers, there are many uncontrollable variables. Even small changes in chemical composition, structure, and processing conditions can lead to large variations in properties. Therefore, Traditional linear modeling techniques that assume a general environment often produce significant errors when applied to field data. In this study, we propose a new modeling method (GPR-SEM) that combines Structural Equation Modeling (SEM) and Gaussian Process Regression (GPR) to study the Friction-Coefficient and Flexural-Strength properties of Polyacetal resin, an engineering plastic, in order to meet the recent trend of using plastics in industrial drive components. And we also consider the possibility of using it for materials modeling with nonlinearity.

Keywords : Gaussian Process Regression, Structural Equation Modeling, Multivariate Analysis, Polymer Development, Polyacetal(PA)

가우시안 프로세서 회귀 기반의 비선형 구조방정식을 활용한 고분자 물성거동 예측 연구

문 경 렬[†] · 박 건 옥^{††}

요 약

고분자 분야의 개발 및 양산과정에는 제어가 안되는 많은 변수가 있으며, 화학적 조성, 구조, 가공 조건 등 작은 변화에도 물성편차가 크게 발생하기에 보편적인 환경을 가정한 기존의 선형적 모델링 기법으로는 현장 데이터 적용시 많은 오차가 발생한다. 이에 본 연구에서는 최근 산업용 구동부품의 플라스틱 채용경향에 맞추어 엔지니어링 플라스틱인 Polyacetal 수지의 내마모성 및 내굴곡성 강화 연구에 다변량 분석기법인 구조방정식과 가우시안 프로세스 회귀를 결합한 모델링 방식(GPR-SEM)을 제안하고, 비선형성을 가지는 물질 모델링에 활용 가능성을 고찰하고자 한다.

키워드 : 가우시안 프로세스 회귀, 구조방정식 모델링, 다변량분석, 고분자 개발, 폴리아세탈

1. 서 론

물질을 만들어내고 활용하는 것은 인간이 문명을 이루는 가장 중요한 요소 중 하나이며, 오래전부터 물질에 대한 조성과 예측방법론이 연구되어 왔다. 최근 머신러닝 기법이 물질 모델링 분야에 적용되면서 금속, 단결정 구조 물질, 저분자 물질이나 단백질합성 같은 바이오 분야는 상당한 수준으로 모델링이 가능한 수준에 이르렀다. 더불어 주요 선진국 및 글로벌 기업들은 분야별 물질정보 데이터베이스(이하 물질정보DB)를 구축해 왔으며 국내도 출원연구기관들 중심으로 물질정보DB 구축 연구사업을 진행하여 방대한 자료들이 수집되어 있다.

위의 물질정보DB를 활용하여[1] 이차전지 양극재 스크리닝[2]이나 태양광 광흡수성 향상[3] 등을 위해 물질을 조합하고 기능을 예측하는 연구가 활발히 진행되고 있다.

그럼에도 불구하고 엔지니어링 플라스틱, 탄성체 같은 고분자(폴리머)에 대한 모델링 방법론에 대한 연구는 상대적으로 덜 진행되어 왔다. 유한요소해석[4]과 딥러닝을 활용한 부분적인 상관관계 연구((1,5))는 꾸준히 진행되어 왔으나, 상용 폴리머들의 조성비 및 물성관계에 대한 범용적인 모델링 기법은 거의 없다고 볼 수 있다. 주된 이유는 폴리머 개발 및 양산과정에 제어가 안되는 많은 변수가 있으며 투입되는 재료와 물성간 비선형성 관계로 인해 정량화가 어렵기 때문이다. 또한 폴리머 물성은 화학적 조성, 첨가제 분자구조, 가공 조건 등 작은변화에도 크게 달라지므로 보편적인 환경을 가정한 모델의 결과값과 현장의 실제 데이터와는 오차가 발생할 수밖에 없다[6].

[†] 정 회 원 : 한국소재융합연구원 생산기술연구단 수석연구원

^{††} 비 회 원 : 한국소재융합연구원 혁신소재연구단 수석연구원

Manuscript Received : October 17, 2023

Accepted : November 30, 2023

*Corresponding Author : Moon Kyung-Yeol(moon@kimco.re.kr)

본 연구에서는 공정의 정량화는 제외하고 폴리머 조성비와 물성간의 관계에 초점을 맞추어 보다 범용적인 모델링 방식을 제안하고 활용 가능성을 살펴보고자 한다.

수식적으로 모델링하기 어려웠던 폴리머 조성비와 물성관계를 레시피를 바탕으로, 기존 구조방정식에서 사용되는 상관계수 또는 표준화계수를 비선형성 회귀함수로 대체하여 구조방정식 형태로 모델링하였다. 비선형적 패턴을 보이는 물성에 대해 방대한 실험데이터 기반이 아니어도 정밀한 예측이 가능하며 자가학습 특징을 갖고 있는 가우시안 프로세스 회귀(이하 GPR) 기법을 적용하여 각각의 첨가제(독립변수)와 물성(종속변수)의 회귀방정식을 구해 매트릭스화 하였고, 이를 구조방정식 형태로 모델링하고 검증용 실험치를 입력하여 예측 정확도를 측정하였다.

2. 가우시안 프로세스와 구조방정식

2.1 가우시안 프로세스(Gaussian Process)

GP는 함수 $f(x)$ 의 값을 확률적으로 추정하는 모델이라고 할 수 있다. 랜덤변수들의 임의의 조합이 모두 가우시안 분포를 갖는 랜덤 프로세스로 정의된다.

랜덤변수 집합에서 정의된 함수 $m: X \rightarrow R$, $k: X \times X \rightarrow R$ 이 주어질 때, GP는 입력 $x \in X$ 에 대해 확률변수 $GP(x)$ 를 출력으로 가지며 m 을 평균함수, k 를 공분산으로 갖는다.

$$f(x) \sim GP(m, k) = GP(\mu(x), k(x, x'))$$

가우시안 프로세스 회귀(이하 GPR)는 함수의 측정값을 바탕으로 모집단 함수를 식별, 추정하는 방법으로 알려져 있다. 이러한 GPR은 잡음이 없는 정확한 측정값이 충분히 제공되면 추정하려는 사후평균이 실제 값과 매우 가까운 값을 출력하며, 더 많은 측정값이 주어질수록 추정값은 더 정밀해진다. 단점으로는 데이터가 많아질수록 $O(N^3)$ 복잡도를 갖게 되어 계산량이 많아진다[7,8].

전통 회귀 모델링과 GPR방식의 회귀 모델링 차이는 다음과 같다. 전통적 회귀 모델링은 선형인 경우 연구자 가설에 의존하며, 추세선이 직선 또는 n차곡선임을 가정한다. 비선형성 인 경우 특정수식으로 표현이 어려우며, 과적합 가능성이 많아 별도의 튜닝이 필요하다. 예측결과로 스칼라 값만을 제공하며, 모델이 개발이 완료되면 수정이 불가능하고 모델수정 필요시 모델개발 과정을 다시 수행해야 한다(학습 후 검증용 데이터셋으로 모델을 검증해야 함). 즉 비선형/비정형 도메인에 적용하기 어렵고, 측정데이터가 많아야 효과적이며, 비선형성을 모델하는 경우 파라미터 튜닝이 필요하다.

반면 가우시안 프로세스 회귀(GPR, 비선형회귀 모델)의 경우는 연구자 가설에 영향이 적으며, 측정데이터에 기반하여 자가학습 구조를 가진다. 연구자 역할은 커널함수를 생성 및 조합하고 매개변수를 조절한다. 이때 도메인 전문가의 경험치

를 정량화하여 커널함수에 적용해야 정밀한 모델이 도출될 수 있다. GPR 방식은 선형이든 비선형이든 상관없이 거의 모든 패턴의 관측치나 분포를 수식(범함수)으로 표현할 수 있다. 예측결과와 불확실성 정도를 확률정보로 제공하므로 예측범위가 향상되며 특히 모델개발을 완료한 이후에도 새로운 관측치를 통해 모델을 계속 업데이트하여 모델이 점점 정밀해질 수 있다. 정리하자면 GPR은 비선형/비정형 도메인에 적합하고, 관측치가 적은 경우에도 효과적이며, 반면 관측치가 많은 경우 계산량이 많아지는 단점이 있으며 도메인에 적합한 커널 구성 및 파라미터 튜닝이 매우 중요하다 할 수 있다[7,8].

2.2 구조방정식

구조방정식모델(SEM)은 확인적 요인분석과 경로분석이 결합된 분석기법으로서 사회과학분야에서 많이 쓰이는 다변량 모델링 기법이며, 잠재변수의 사용 및 측정오차의 추정, 동시 추정, 총효과, 직접효과, 간접효과 등의 측정 등 다양한 통계기법과 표현기법이 사용된 하이브리드형 통계분석 기법이라고 할 수 있다[9].

구조방정식은 잠재변수와 측정변수를 사용하여 모델을 구성하며, 크게 측정모델과 구조모델로 구성된다. 측정모델은 잠재변수를 측정하는 측정변수 간의 관계를 설명하는 모델이며, 측정변수들은 잠재변수의 측정오차를 포함하고 있으므로, 측정모델은 잠재변수와 측정변수 간의 선형 관계를 설명하는 선형 회귀모형으로 표현된다. 구조모델은 잠재변수 간의 인과관계를 설명하는 모델이다. 잠재변수들은 관측이 불가능한 추상적인 개념이기 때문에, 구조모델은 잠재변수 간의 인과관계를 설명하는 경로분석 모형으로 표현된다.

모형평가는 적합도지수를 활용하여 수행하는데 구체적인 검정방법 및 범위는 아래와 같다[10].

많은 통계기법들의 장점들을 종합한 구조방정식 모델이지만 고분자 물성 모델링에는 거의 사용되지 않는 분석방법이다. 그 이유는 앞서 언급한대로 고분자 개발분야는 대표적인 비선형계 도메인 이기에 모델 검증 및 해석이 어렵기 때문이다. 또한 기능을 구현을 위해 다수의 첨가제들이 배합에 투입되고 첨가제와 물성들간의 관계가 다대다의 관계가 많으며 상관성 또한 비선형인 경우가 대부분이다. 그리고 제어되지 않는 공정변수(롤링/믹싱 수작업, 재료 물성편차 등)들이 존재하고, 재료간에도 서로 상호작용 일어나고, 가공과정에서 중간 반응과정이 생기기도 한다. 이러한 요인들을 기존 구조방정식 모델링 절차에 따라 경로계수로 표현하면 새로운 실험데이터 적용시 큰 오차가 발생하여 실제상황에서 모델 정확도가 떨어지기 때문이다.

그래서 본 논문에서는 적은 관측치로도 비선형계의 변화를 정밀하게 표현할 수 있고 모델변화가 동적인 GPR 비선형함수를 표준화계수로 치환하고 변수관계를 매트릭스를 만들어 구조방정식 형태로 고분자 첨가제와 물성간의 관계를 모델링해 보았다.

Table 1. Criteria for Evaluating Fitness Index

Fitness index	Criteria for optimal model
χ^2 (p-value)	(.05)
RMSEA	< .08
GFI	.90 - 1
AGFI	.85 - 1
CFI	.90 - 1
NFI	.90 - 1
NNFI(TLI)	.90 - 1
RMR	< .05

그런데 관계계수에 비선형함수가 대입되게 되면 검증방식을 새롭게 고찰해 볼 필요가 있다. 기존 구조방정식 모델은 연구가설이 선형성이라는 가정을 하였기에 모델링 결과로 피팅된 관계계수(표준화계수 등)는 상수값이 되고 새로운 값이 들어와도 관계계수(표준화계수 등)는 변화하지 않는다. 모델이 잘되었거나 잘못 만들어진 것이다.

반면 GPR 방식은 “Bayesian optimization”에 기반한 함수 추정이기 때문에 관측 값이 입력될 때 마다 모델이 변한다. 다시 말해 실험을 할 때마다 모델은 변화하고 측정값이 기존모델과 일치하면 다행이고, 예측값이 분산범위 밖에 있어도 데이터에 맞게 GPR함수가 변화한다. 모델이 잘못된 만들어진 것이 아니고 입력데이터에 따라 변화한다는 관점이다. 연구자가 개입할 수 있는 부분은 커널함수 구성과 하이퍼파라메타 설정이다. 이와 같은 이유로 GPR 기반 모델을 검증하기 위해서는 커널작성과 하이퍼파라메타가 중요한 기준이 되어야 할 것이다[7].

본 연구에서는 모델검증 및 평가방법은 포함하지 않으며, GPR기반 구조방정식(이하 GPR-SEM)을 이용한 다변수/다변량 모델링 기법이 고분자 물성추정에 대해 활용 가능한지에 대해서만 살펴볼 것이다.

3. 기계적 물성&거동

3.1 기계적 물성

고분자 물질 중 엔지니어링 플라스틱은 높은 굴곡강성, 내구성 등으로 다양한 산업분야에 사용되는 소재이다. 최근 구동부품의 고기능화를 위하여 다양한 소재들이 사용되고 있으며, 이로 인한 굴곡강도 향상 및 마찰계수 저감 요구가 높아지고 있다. 이에 따라 본 연구에서는 Polyacetal(Polyoxymethylene) 수지를 기반으로 굴곡강도 향상을 위한 보강제 및 내마모성, 마찰계수 저감을 위한 슬립(Slip) 첨가제를 적용하고, 이에 따른 내굴곡성 및 마찰계수 영향성을 확인하고자 하였다.

3.2 실험

아래 표와 같이 Polyacetal(POM)수지를 기반으로 보강제인 chopped Glass Fiber (Hankuk Fiber, 1.5mm, GF)을 0, 5, 10, 15 phr, 슬립첨가제인 Graphite를 0, 1, 3, 9, 12 phr 각각 혼합하고, 플라스틱 사출기로 굴곡강도 및 마찰계수 측정

시편을 제작하였다. 제작된 시편은 아래의 Fig. 1의 Friction Coefficient Tester(Davenport사)를 활용하여 건조(Dry) 상태에서 마찰계수값을 측정하였으며, Fig. 2의 Texture Analyzer (BrookField사)를 활용하여 굴곡강도를 측정하였다. 측정 결과는 Table 2에 나타내었다.

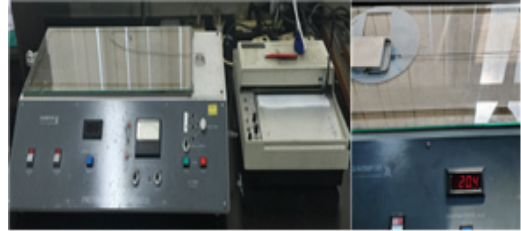


Fig. 1. Friction Coefficient Tester

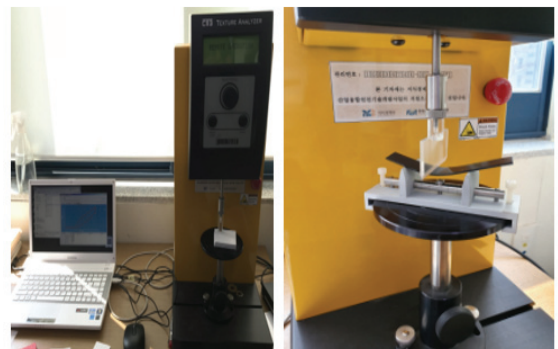


Fig. 2. Texture Analyzer

Table 2. Test Result

No.	Recipes			Properties	
	POM	GF	Graphite	Flexural Strength	Friction Coefficient
	phr	phr	phr	kgf/cm2	-
Training Dataset					
S1	100	0	0	950	0.250
S2	100	0	1	971	0.242
S3	100	0	3	987	0.231
S4	100	0	9	923	0.210
S5	100	0	12	879	0.198
S6	100	5	0	1,280	0.275
S7	100	5	1	1,298	0.267
S8	100	5	3	1,313	0.257
S9	100	5	9	1,209	0.246
S10	100	5	12	1,104	0.238
S11	100	10	0	1,590	0.296
S12	100	10	1	1,640	0.289
S13	100	10	3	1,723	0.284
S14	100	10	9	1,584	0.276
S15	100	10	12	1,432	0.268
Validation Dataset					
S16	100	15	0	1,730	0.302
S18	100	15	3	1,746	0.296
S19	100	15	9	1,670	0.284
S20	100	15	12	1,530	0.274

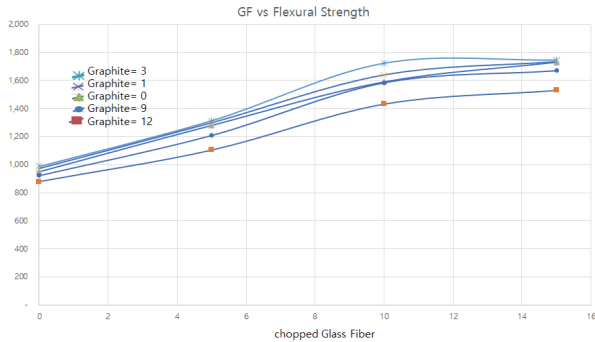


Fig. 3. GP vs Flexural-Strength

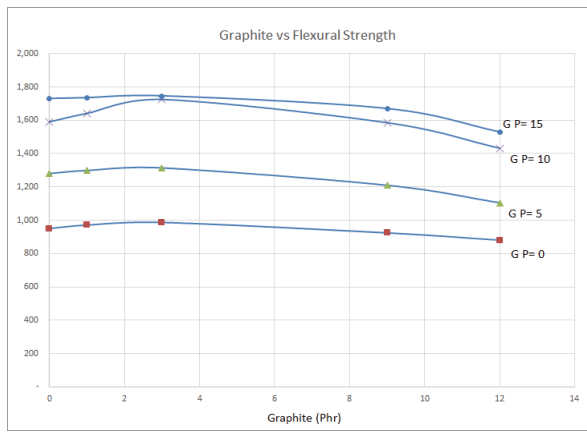


Fig. 4. Graphite vs Flexural-Strength

3.3 결과

chopped Glass Fiber 함량이 증가할수록 굴곡강도는 증가하였으며, 마찰계수도 증가하는 경향을 확인하였다. 반면 Graphite는 함량이 6phr까지는 증가할수록 굴곡강도가 증가되나, 6phr 이후부터는 감소하는 경향을 확인하였으며, 마찰계수는 감소되는 것을 확인할 수 있었다.

4. 고분자 레시피와 물성간의 관계 모델링

4.1 모델링 절차

일반적인 머신러닝 검증절차에 따라 훈련 데이터셋과 검증 데이터셋을 3:1로 분리하여 훈련 데이터셋으로 도출된 모델식에 검증 데이터를 적용하고 비교하여 물성치가 실험현장의 신뢰구간내에 포함되는지 확인할 것이다. 커널함수와 하이퍼파라메타에 따라 분산범위를 조정할 수 있는데 도메인 특성에 맞게 오버피팅이나 언더피팅이 되지 않도록 해당도메인 전문가의 의견을 수집하고 모델에 반영하는 것이 중요하다 하겠다.

본 모델링은 python의 scikit-learn 패키지중 gaussian processes 클래스를 활용하여 구현하였다. GPR 추정에서는 커널함수의 선정과 구성이 매우 중요하며 최소자료 입력(small-data)으로 다변수/다변량 구조방정식 모형을 구성하는 것이 본 연구의 목표이며 상세 절차는 다음과 같다.

① GPR 모델링

GPR을 활용하여 조성비별 물성추정 범함수 방정식을 생성한다(커널함수 구성 및 하이퍼파라메타 설정). Table 2의 S1~S5 실험세트를 대상으로 대표 커널함수인 RBF(방사기저함수/제곱지수함수)를 사용하여 GPR 모델링과 일반선형회귀(OLS)의 모델적합도를 단순비교해 보았다(Table 3).

Table 3. GPR & Linear Regression Result

GPR Results			
=====			
GPR_SCORE	:	0.999988	
gp.kernel	:::	31.6**2 * RBF(length_scale=10)	
OLS Regression Results			
=====			
Dep. Variable:	y	R-squared:	0.400
Model:	OLS	Adj. R-squared:	0.100
Method:	Least Squares	F-statistic:	1.334
		Prob (F-statistic):	0.367
Time:	11:30:28	Log-Likelihood:	-17.361
No. Observations:	4	AIC:	38.72
Df Residuals:	2	BIC:	37.50
Df Model:	1		
Covariance Type:	nonrobust		

Table 4. GPR Prediction Result

	Mixing ratio			Flexural Strength
	POM	GF	Graphite	
	phr	phr	phr	kgf/cm ²
train data	100	0	0	950
	100	0	1	971
	100	0	3	987
	100	0	9	923
validate data	100	0	12	real = 879 predict = 869

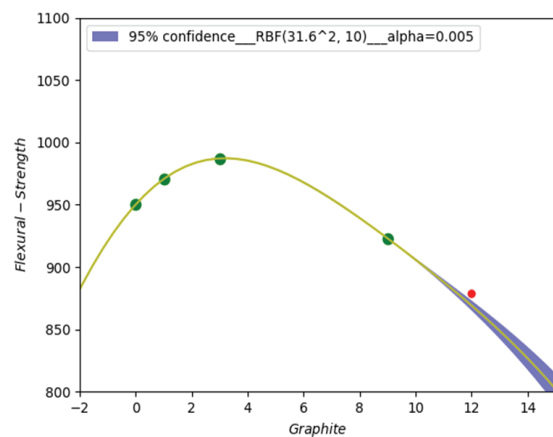


Fig. 5. Graphite vs Flexural-Strength

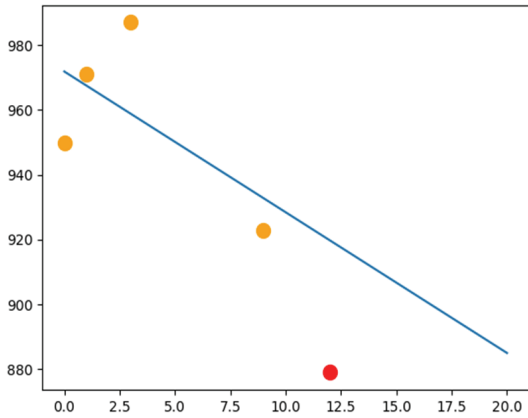


Fig. 6. OLS Regression

S1~S5 레시피 세트에 대해 모델을 평가를 하였는데 비선형 자료이기에 선형회귀로는 모델 설명력이 부족(R2=0.4)하며 (Fig. 6), GPR방식이 모델 정확도(R2=0.99)와 예측 정확도가 높게 나타남을 볼 수 있다(Fig. 5). GPR은 주어진 데이터에 가장 가까운 추세를 표시하기에 (Fig. 5) 그래프 자체로는 유의 미하다고 보기 어렵다. 많은 검증데이터셋을 통해 실제상황에서의 모델을 검증하지 않았기에 모델에 대한 엄밀한 검증은 부족하고 할 수 있다. 그러나 고분자 개발을 위해 다수의 첨가제를 활용하여 레시피를 설계하고 실험을 하는데 많은 비용과 시간이 소요되므로, 한그룹의 레시피로 많은 실험을 하기보다는 한그룹의 레시피에 작은수의 실험을 하고 이러한 실험세트를 다양한 조합으로(레시피의 분할) 실행하는 것이 실험모수를 줄이면서 목표로 하는 최적점을 찾는 일반적인 고분자 개발방식이다[11]. 따라서 최소 관측자료(small-data)를 바탕으로 부분예측 가능한 모델을 만들고 (Fig. 5) 관측치 추가를 통해 모델을 정밀하게 업데이트하는 방식이 현업에 적합하다 할 수 있겠다[12].

② 통합추세선 작성

첨가제 대비 물성영향이 큰 순서대로 레시피를 순차적으로 정렬하여 범함수 방정식을 통합하여 추세선을 만든다(Fig. 11~14).

③ 구조방정식 모델링

구조방정식에서 모수를 추정하는 방법은 크게 공분산기반 CB-SEM과 선형회귀를 기반으로 하는 PLS-SEM으로 구분 되는데 본 연구에서는 위의 방식이 아닌 비선형 회귀식인 GPR을 회귀계수 대신 활용한다(GPR-SEM). GPR 결과는 고정된 함수가 아닌 범함수식으로 표현되고, 추가 데이터가 입력될 때마다 경로계수(표준화계수)가 변화된다. 선형회귀에 기반한 구조방정식은 각변수들의 경로계수들을 집계하기 위해 가중합을 사용하지만 비선형 구조방정식(GPR-SEM)에서는 위의 방식을 사용하는데 어려움이 있다. 따라서 비선형 함수들의 가중합을 구하기 위해 측정 구간내 값들을 x1에 해당하는 y값을

0으로 하고, x1을 기준으로 나머지 y값들을 선형변환한 다음 적절한 값(θ)을 곱하여 각 지점별 가중치로 사용한다(Table 5). “①”을 바탕으로 첨가제와 물성간의 상관관계 매트릭스를 생성하고 모델을 도식화 한다(Fig. 15).

④ 모델검증 및 업데이트

검증데이터를 추가하여 추정값이 모델의 분산 범위내에 포함되는지 확인하고, 필요시 커널함수를 수정하거나 재작성한다. 추가 데이터(조성비)는 입력값은 새로운 관측치가 되고, 위의 단계를 반복하여 경로계수(물성방정식)와 구조방정식 모델을 업데이트 한다(Table 6).

4.2 모델링 결과

1) GPR 모델링

가장 기본적인 커널인 RBF와 Constant커널을 활용하여 가장 적합한 모형의 파라메타를 적절히 조정하여 아래와 같이 커널함수를 구성하였다.

$$\text{RBF kernel : } k(x_i, x_j) = \exp\left(-\frac{d(x_i, x_j)^2}{2l^2}\right)$$

$$\text{Constant kernel : } k(x_1, x_2) = \text{constant_value } \forall x_1, x_2$$

$$\text{Kernel} = \text{RBF} * \text{Constant}$$

Table 2의 S0~S5 레시피를 Graphite와 굴곡강도(Flexural Strength)의 기저관계(basis)로 설정하고 오차범위를 0.005로 피팅하여 Fig. 7에 그래프로 나타내었다. 오차범위와 분산범위 선정은 고분자개발 전문가의 경험치(휴먼팩터)를 반영하였다.

Table 2의 S0~S5 레시피를 Graphite와 마찰계수(Friction Coefficient)의 기저관계로 설정하고 오차범위를 0.0001로 피팅하여 Fig. 8에 그래프로 나타내었다. 오차범위와 분산범위 선정은 고분자개발 전문가의 경험치를 반영하였다. Fig. 7에 비해 분산범위가 넓은 이유는 마찰계수 실험측정값의 변동성이 크기 때문이다.

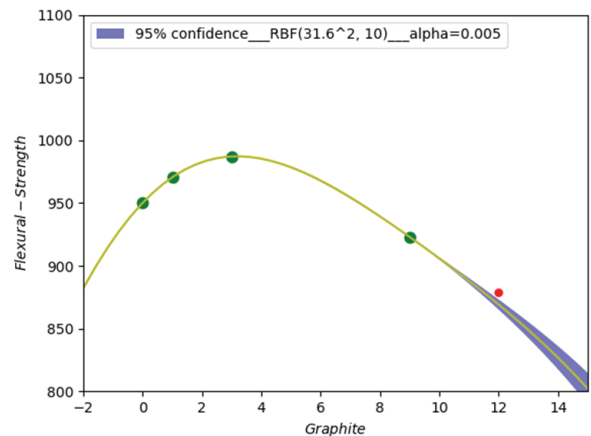


Fig. 7. Graphite vs Flexural-Strength = f1(x)

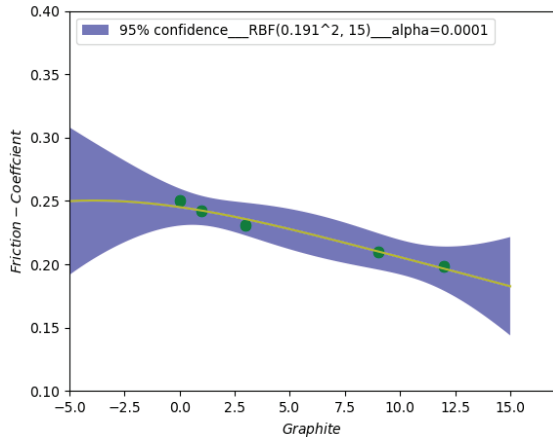


Fig. 8. Graphite vs Friction-Coeff = $g_1(x)$

Table 2의 S1, S6, S11 레시피를 Glass-Fiber와 굴곡강도간의 기저관계, Glass-Fiber와 마찰계수간의 기저관계로 설정하고 피팅하여 Fig. 9-10에 그래프로 나타내었다. 오차범위와 분산범위 선정은 고분자개발 전문가의 경험치를 반영하였다. Fig. 10에서 분산범위가 넓은 이유는 마찰계수 실험 측정값의 변동성이 크기 때문이다.

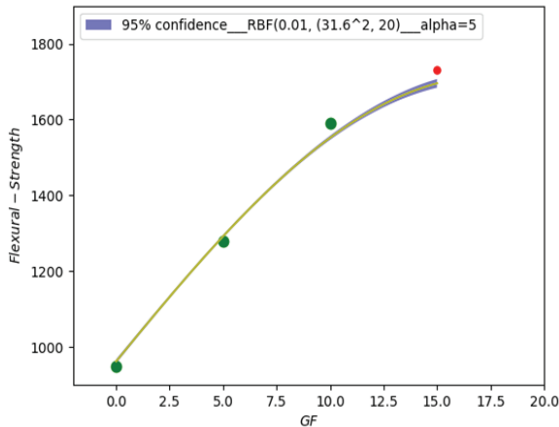


Fig. 9. GF vs Flexural-Strength = $f_2(x)$

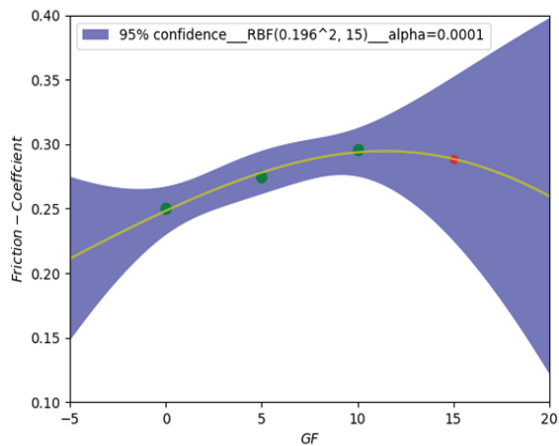


Fig. 10. GF vs Friction-Coeff = $g_2(x)$

Table 5. GPR-SEM Matrix (Material vs Properties)

		Properties	
		Flexural Strength	Friction Coefficient
Additive	Graphite	$f_1(x) \sim GP(m,k)$ $\sigma^2 = 31.62$ $L = 10$ $Weight_2 = \theta_1 * f_1(x)$	$g_1(x) \sim GP(m,k)$ $\sigma^2 = 0.191$ $L = 15$ $Weight_2 = \theta_2 * g_1(x)$
	GF	$f_2(x) \sim GP(m,k)$ $\sigma^2 = 31.62$ $L = 10$ $Weight_2 = \theta_3 * f_2(x)$	$g_2(x) \sim GP(m,x)$ $\sigma^2 = 0.196$ $L = 15$ $Weight_2 = \theta_4 * g_2(x)$

2) 첨가제와 물성간의 관계 매트릭스

1)에서 생성한 개별 기저관계 모델들을 통합하여 첨가제와 물성간의 관계 매트릭스를 Table 5로 나타내었다. 커널함수 구성에 있어서 전문가 경험치 확보는 모델정확도를 결정짓는 중요한 요소 중 하나이다. 다른 종류의 고분자를 모델링하는 경우에도 다수의 도메인 전문가가 모델구성에 참여하고, 경험치가 많이 수집될수록 더 정교한 모델이 만들어질 수 있다.

3) 기계적 물성 그래프

통합추세선 그래프는 아래와 같다(Fig. 11-14). RBF 커널을 활용하는 경우 통상 발산하는 성질을 보인다. 반면 고분자 배합의 경우 수렴하는 경우가 다수이므로, 발산과 수렴을 결정하는 지점 및 수준을 적절히 결정하는 것이 중요한 부분이다. Fig. 12, 14는 정규화를 거친 통합추세선을 보여주고 있다. 급격하게 발산하던 모델이 부드럽게 수렴 또는 발산됨을 볼 수 있다. 특히 Fig. 13은 우상향으로 발산하는데 반해 Fig. 14는 현실에 맞게 우하향으로 완만하게 발산하는 모델로 변경되었다. 추가실험이 입력되는 경우 정규화를 거친 모델이 예측정확도가 더 높게 나옴을 예상할 수 있겠다.

4) 비선형 구조방정식 모형 도식화

Table 5 매트릭스를 기반으로 구조방정식 모형으로 도식화한 최종모델 구조도를 Fig. 15에 나타내었다. 비선형 함수기반 이므로 구조도에서 관계계수를 표현하기는 어려우며 관계화살표와 대표적인 함수명만을 명시하였다.

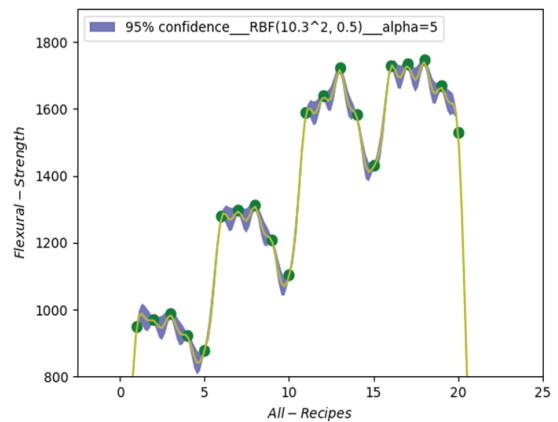


Fig. 11. All Recipes vs Flexural-Strength

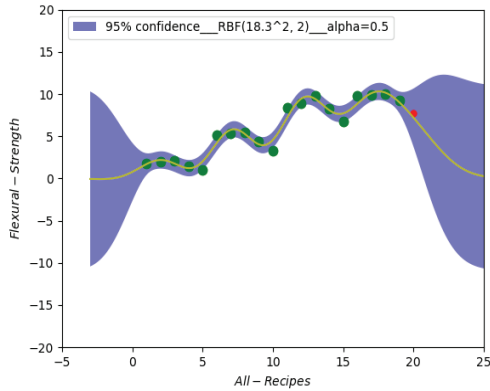


Fig. 12. All Recipes vs Flexural-Strength (normalized)

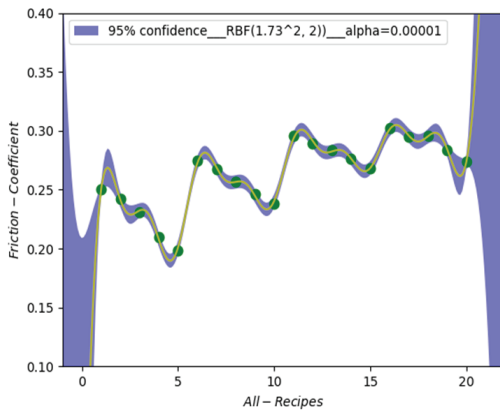


Fig. 13. All Recipes vs Friction-Coeff

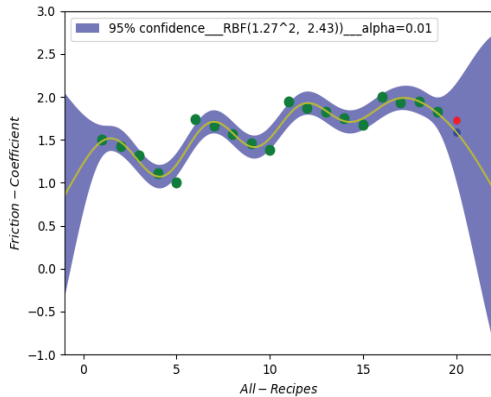


Fig. 14. All Recipes vs Friction-Coeff (normalized)

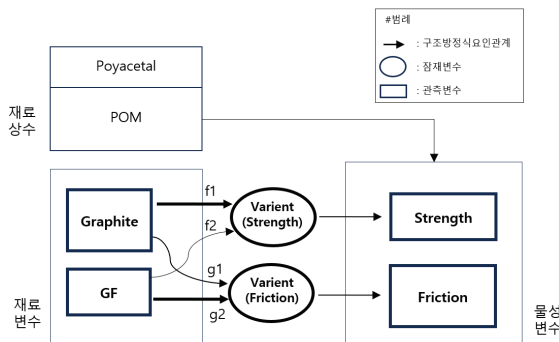


Fig. 15. GPR-SEM Model

5. 결 과

5.1 모델 성능평가

모델이 산출한 예측결과는 아래와 같다.

Table 6. Prediction Process

No.	Mixing ratio		Prediction					
			Flexural Strength			Friction Coefficient		
S16	GF	15	Weight ₁ f ₂ (15,0)=	1,696	1,696	Weight ₁ g ₁ (15,0)=	0.289	0.289
	graphite	0	Weight ₂ f ₁ (0,0)=	0		Weight ₂ g ₂ (0,0)=	0	
S18	GF	15	Weight ₁ f ₂ (15,0)=	1,696	1,744	Weight ₁ g ₁ (15,0)=	0.288	0.283
	graphite	3	Weight ₂ f ₁ (0,3)=	48.031		Weight ₂ g ₂ (0,3)=	-0.005	
S19	GF	15	Weight ₁ f ₂ (15,0)=	1,696	1,660	Weight ₁ g ₁ (15,0)=	0.288	0.280
	graphite	9	Weight ₂ f ₁ (0,9)=	-35.280		Weight ₂ g ₂ (0,9)=	-0.008	
S20	GF	15	Weight ₁ f ₂ (15,12)=	1,696	1,603	Weight ₁ g ₁ (15,0)=	0.288	0.279
	graphite	12	Weight ₂ f ₁ (0,12)=	-92.501		Weight ₂ g ₂ (0,12)=	-0.009	

Table 7. Prediction vs Measured-value

No.	Mixing ratio		Prediction		Measured value		Prediction error	
			Flexural Strength (kgf/cm ²)	Friction coefficient	Flexural Strength (kgf/cm ²)	Friction coefficient	Flexural Strength (kgf/cm ² ,%)	Friction coefficient (%)
S16	GF	15	1,696	0.289	1,730	0.302	-34	-0.013
	graphite	0					-2%	-4.5%
S18	GF	15	1,744	0.283	1,746	0.296	-2	-0.013
	graphite	3					-0.1%	-4.3%
S19	GF	15	1,660	0.280	1,670	0.284	-10	-0.004
	graphite	9					-0.6%	-1.5
S20	GF	15	1,603	0.279	1,530	0.274	73	0.005
	graphite	12					4.8%	1.8%

Table 8. Model Prediction Result (Flexural-Strength)

OLS Regression Results		
Dep. Variable:	y	R-squared: 0.895
Model:	OLS	Adj. R-squared: 0.843
Method:	Least Squares	F-statistic: 17.07
		Prob (F-statistic): 0.054
		Log-Likelihood: -18.941
No. Observations:	4	AIC: 41.88
Df Residuals:	2	BIC: 40.65
Df Model:	1	
Covariance Type:	nonrobust	

Table 9. Model Prediction Result (Friction-Coefficient)

Dep. Variable:	y	R-squared:	0.823
Model:	OLS	Adj. R-squared:	0.735
Method:	Least Squares	F-statistic:	9.323
		Prob (F-statistic):	0.0926
		Log-Likelihood:	19.982
No. Observations:	4	AIC:	-35.96
Df Residuals:	2	BIC:	-37.19
Df Model:	1		
Covariance Type:	nonrobust		

표본수가 작기에 통계분석을 통해 평가하기에 무리가 있으나, 예측성능은 R-squared = 0.895와 0.823 수준이며, 실험 값이 추가되어 진다면 예측정밀도가 더 높아질 것이라 예상된다. 고분자 개발에 있어서 오차관리는 중요한 항목이다. 고분자 개발에 있어서 측정물성 항목에 따라 허용오차가 달라지는데 굴곡강도와 마찰계수에 한해서 상기의 수치는 유의미한 결과값이라 할 수 있겠다[6].

R-squared뿐 아니라 RMSE 등 다양한 평가지표를 통해 GPR 모델을 평가할 수 있다[13]. 본 연구에서는 small-data 기반으로 단순검증만을 진행하였으며, GPR은 주어진 관측치에 꼭 맞는 추세선을 만드는 과적합 가능성이 높은 방식이므로 좀 더 많은 실험자료를 통해 성능평가가 진행되어야 할 것이다. 본 연구에서 제안한 비선형 구조방정식(GPR-SEM)을 활용하여 오차를 7% 내외로 다변수/다변량 결과를 계산할 수 있음을 보였다. 다만 앞서 언급한 대로 좀 더 엄밀한 평가방법에 대한 연구가 필요하다 하겠다.

6. 결 론

본 논문에서는 비선형 회귀 알고리즘을 구조방정식에 적용하여 고분자물성과 재료배합의 관계를 모델링해보았다. 다변수 및 다변량 구조를 반영할 수 있는 GPR기반 비선형 구조방정식(GPR-SEM) 모델을 통해 전통적 모델링 방식보다 정밀하고 범용적인 물성거동 예측모델링 방식을 제안하였다. 변수구조를 단순화 하였으며, 최소 레시피(small-data)를 기반으로 모델을 구성하였다. 그리고 검증데이터 결과값이 실험실 기준 오차범위 내에 있음을 확인하였고, 비선형계인 고분자 배합 및 물성거동을 모델링하는데 활용할 수 있음을 볼수 있었다.

향후 “비선형함수 기반 구조방정식”을 검증/평가를 위한 요인을 파악하고 정량적 기준을 마련하는 추가연구가 필요하다. 더 높은수준의 정확도를 달성하기 위한 필요한 실험데이터의 적정량과 범위를 도출해야 하며, 좀 더 복잡한 고분자 조성물에 대해 관측변수 및 잠재변수를 모델링하고, 다양한 고분자 물질군의 물성예측에 적합한 커널함수 개발 및 하이퍼파라미터 값을 찾는 연구가 요구된다.

더불어 가능성을 부여하기 위해 새로운 첨가제가 필요한 경우, 국내 또는 해외 물질정보 시스템에서 물질정보를 수집하고, 후보물질군을 스크리닝하는 머신러닝 기법[2]과 본 연구에서 제안한 “비선형함수 기반 구조방정식” 모델링 방법을 융합하면, 기능성 고분자를 개발하기 위해 소요되는 실험모수를 대폭 줄일 수 있으며, 나아가 디지털화 비율이 낮은 고분자 설계분야의 디지털 전환에 기여할 수 있을 것이다.

References

- [1] N. Lee, Y. Shin, and D. Shin, “Prediction of mechanical properties and behavior of polymer matrix composites based on machine learning,” *Journal of the Korean Institute of Gas*, Vol.25, No.2, pp.64-71, 2021.
- [2] M. Kim, S. Kang, and K. Min, “Searching for next-generation battery cathode materials using machine learning,” *Korean Society of Mechanical Engineers, 2021 Conference*, pp.1166-1168, 2021.
- [3] S. M. Kim, “Optical characterization and prediction with neural network modeling of various composition of perovskite using a hyper regression method,” *Polymer Science and Technology*, Vol.33, No.6, pp.548, 2022.
- [4] J.-H. Sim, “Characteristics analysis of highly elastic materials according to the graphite content and a simulation study of physical properties prediction using a nonlinear material model,” *Textile Coloration and Finishing*, Vol.34, No.4, pp.250, 2022.
- [5] K.-M. Lee, K.-Y. Kim, U. Oh, S.-K. Yoo, and B.-S. Song, “Prediction of multi-physical analysis using machine learning,” *Journal of IKEEE(Institute of Korean Electrical and Electronics Engineers)*, Vol.20, No.1, pp.99-100, 2016.
- [6] D. Lee I.-S. Hwang, “Analysis on the dynamic characteristics of a rubber mount considering temperature and material uncertainties,” *Korea Society of Computational Structural Engineering Journal*, Vol.24, No.4, pp.387-388, 2011.
- [7] C. E. Rasmussen and C. K. I. Williams, “Gaussian Processes for Machine Learning,” pp.19-46, pp.79, pp.114-122, pp.123-146, 2006.
- [8] C. M. Bishop, “Pattern recognition and machine learning,” pp.291-339, 2006.
- [9] J.-P. Yu, “The Criticisms and considerations of structural equation modeling,” *Journal of Product Research*, Vol.34, No.4, pp.83-93, 2016.
- [10] C.-S. Yoo, “A performance evaluation of fit indices for structural equation model,” pp.2-7, 2019.

- [11] T. Eckert, F. C. Klein, P. Frieler, O. Thunich, and V. Abetz, "A new experimental design method for targeted property optimization in polymer development," *Polymers*, 13-03147-v2, pp.6-10, 2021.
- [12] S. Pruksawan, G. Lambard, S. Samitsu, K. Sodeyama, and M. Naito, "Prediction and optimization of epoxy adhesive strength from a small dataset through active learning," *Science and Technology of Advanced Materials*, Vol.20, No.1, pp.1010-1011, 2019.
- [13] M. Hossin and M. N. Sulaiman, "A review on evaluation metrics for data classification evaluation," *International Journal of Data Mining & Knowledge Management Process (IJDKP)*, Vol.5, No.2, pp.5, 2015.



문 경 렬

<https://orcid.org/0009-0008-5140-305X>

e-mail : moon@kimco.re.kr

2002년 동의대학교 전산과학(석사)

2002년 ~ 현 재 한국소재융합연구원

생산기술연구단 수석연구원

관심분야 : Polymor Production

Modeling, Reinforce Learning



박 건 옥

<https://orcid.org/0009-0008-9432-6322>

e-mail : kupark@kimco.re.kr

2021년 부산대학교 화공생명·환경공학부

(박사수료)

2007년 ~ 현 재 한국소재융합연구원

혁신소재연구단 수석연구원

관심분야 : 고분자, 고분자가공, 적층가공, 복합재료