機械学習によるブロック共重合体の物性推算

坂 下 竜 $-^{*1}$ 小 峯 拓 0^{*1}

Property Prediction of Block Copolymers by Machine Learning

Ryuichi SAKASHITA Takuya KOMINE

The machine learning which estimates elastic modulus of block copolymers has been trained using calculated values of dissipative particle dynamics (DPD) and finite element method (FEM) as training data set. The resultant estimator predicts elastic modulus of block copolymers from their compositions and interaction parameters with high accuracy in short time.

1. はじめに

機械学習を始めとしたインフォマティクスは近年著 しい発展を見せ、画像認識・言語解析・ビッグデータ 解析など多くの分野で盛んに応用されており、その重 要性は年々増してきている。化学分野においても薬 理活性探索などを中心にインフォマティクスが盛んに 活用されるようになって久しい。しかしながら、イン フォマティクスを高分子等のソフトマテリアルの物性 推算に適用した例はまだ少ない1)-3)。これは、ソフト マテリアルが階層構造を持つために、低分子化合物に 比べて特徴変数が多いことが原因であろう。高分子の 場合、このような変数として、一次構造では共重合組 成・分子量・立体規則性・分岐鎖・架橋状態、高次構 造では結晶状態やパッキング、さらに配合処方として 可塑剤等の添加剤などが挙げられる。機械学習を行う ためには良質な(前提条件が揃い、偏りのない)多数 のデータが必要であるが、これら膨大な変数を十分に 網羅した実験データを集めるには多大な労力を必要と される。ポリマーごとにガラス転移温度などの物性を まとめたデータベースは存在するが、同一の組成でも 分子量や結晶状態、成形履歴等で物性は変化するので、

限定的な情報しか得られない。

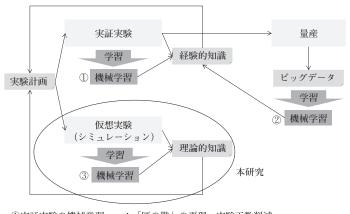
しかしながら、必ずしもインフォマティクスの対象 を実測値に限定する必要はない。高分子研究の現場に おいては、分子動力学・自己無撞着場(SCF)法・有 限要素法などによるシミュレーションが盛んに用いら れている。このような各種の高分子シミュレーショ ンは、均質で多数のデータを用意するうえでも有用で あり、インフォマティクスとのシナジーが期待される (Fig. 1)。「シミュレーション結果を予測する」こと は一見して無意味に思えるが、高分子シミュレーショ ンの弱点である大きな計算負荷と、それに伴う大規模 スクリーニングの困難を解決することが期待される。 今回我々は、高分子シミュレーションによる物性スク リーニングの高速化の為に機械学習を適用することを 試みた。シミュレーション結果を教師データとして機 械学習を行うことで、類似した系について短時間・低 負荷で物性推算が可能になると期待される。

2. ブロック共重合体の弾性率推算

[1] 背 景

ブロック共重合体のミクロ相分離構造はマクロな諸 物性に大きな影響を与えるため実用上重要であり、相 分離構造・及びその物性のシミュレーションは広範に

^{*1} ファンクショナルポリマー研究所 光学材料グループ



①実証実験の機械学習 : 「匠の勘」の再現・実験工数削減 ②ビッグデータの機械学習:プロセス技術改良・顧客の傾向分析 ③仮想実験の機械学習 : 計算負荷削減・大規模計算スクリーニング

Fig. 1 研究開発における機械学習の活用ポイント

研究されてきた⁴⁾。SCF法・全原子モデルや粗視化モデルに基づく分子動力学・散逸粒子動力学(DPD)⁵⁾ などのシミュレーション手法は、ミクロ相分離構造について重要な洞察を与える。しかし、計算負荷の観点から大規模な系を網羅的にスクリーニングするには困難があった。

[2] 手 法

散逸粒子動力学(DPD)によりブロック共重合体の相分離構造を生成し、それらの弾性率を有限要素法(FEM)により求めた。高分子シミュレーションには無償公開されている OCTA システム ⁶⁾ を用いた。OCTA システムの入出力は Python インターフェースで制御されるため、Python の豊富な機械学習ライブラリと連携を取りやすい。検討は以下のようなスキームで行った(Fig. 2)。

(1) 一次構造の生成

多様な一次構造ライブラリを用意する為、仮想モノマー種 A, B, C からなる鎖長 N=100 の 3 元ブロック共重合体を、ブロック数 $1\sim5$ の範囲でランダムに生成させた。モノマー間相互作用パラメータ a_{ii} は 25

~30の範囲でランダムに設定した。同種のモノマー同士の a_{ii} は 25 に固定した。 a_{ij} は Flory-Huggins 理論の χ パラメータに (a) 式で対応することが知られている。

$$a_{ii} = 25 + 3.28 \chi_{ii}$$
 (a)

(2) 散逸粒子動力学(DPD)

全境界に周期境界条件を設定した立方体中に密度=3となるように分子鎖を配置し、平衡状態を得るまでDPD計算を実施した。計算ソルバーにはOCTAシステムに含まれるCOGNAC ver. 9.0.3 を用いた。得られた構造には、相溶・海島・シリンダ・ラメラ・ジャイロイド等の多様な相構造がランダムに含まれていることを確認した。

(3) 有限要素法(FEM)

(2) の計算結果から各モノマーの三次元濃度分布を取り出して入力とし、FEM 計算を実施した。変位面を壁面境界・それ以外の面を周期境界と設定し、各方向について伸長・ずりを与えて応力を求めた。計算ソルバーには OCTA システムに含まれる MUFFIN

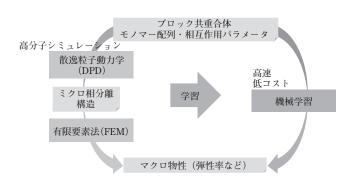


Fig. 2 高分子シミュレーションと機械学習のスキーム

version 5.0 E beta を用いた。

計算された応力より、応力弾性率テンソル (C) を求めた (Fig. 3)。弾性率テンソルより、式 (b) (c) をそれぞれ用いて体積弾性率 (K)・ずり弾性率 (G) を算出した。

$$\begin{split} &K = \frac{1}{9} \{ (C_{11} + C_{22} + C_{33}) + 2(C_{23} + C_{31} + C_{12}) \\ &G = \frac{1}{24} \{ \{C_{11} + C_{22} + C_{33}\} + 3(C_{44} + C_{55} + C_{66}) \end{split}$$
 (b)

 $-(C_{23}+C_{31}+C_{12})$ (c)

モノマー種 A, B, C の弾性率として、全てがガラス状態となっているものを想定した T_1 条件と、A はガラスであるが B, C が溶融したと仮定した T_0 条件

		XX	уу	ZZ	yz	ZX	xy
	XX	C ₁₁	C ₁₂	C ₁₃	C ₁₄	C ₁₅	C ₁₆
	уу	対称	C22	C23	C24	C25	C26
Cijkl=	ZZ	対称	対称	C33	C ₃₄	C ₃₅	C ₃₆
	yz	対称	対称	対称	C44	C ₄₅	C46
	ZX	対称	対称	対称	対称	C55	C56
	zy	対称	対称	対称	対称	対称	C66

Fig. 3 弾性率テンソル

を用意し、それぞれについて計算した(Fig. 4)。

(4) 機械学習(サポートベクタ回帰)

計 280 個の(1)で生成したポリマー構造及び a_{ij} と、(3)で求めた $K \cdot G$ との相関を、サポートベクタ回帰 (SVR) を用いて学習させた $^{7)}$ 。ポリマー構造と a_{ij} は 標準偏差で正規化した上で入力し、n=10 の交差検 証法(n-fold cross-validation estimate)にて求めた 決定係数 R^2 値を学習精度として評価した。[3](4)以降の学習については SVR のハイパーパラメータ (C, γ) を Grid Search で最適化した。計算には scikit learn Ver.0.18.1 80 を用いた。

[3] 結果

(1) シミュレーション結果

FEM で計算した $K \ge G$ は、モノマー数(組成)から加成則で計算した値より低下し、大きな相分離ドメインを持つものほど弾性率が低下する傾向にあった (Fig. 5)。これは大きな相分離ドメインを持つと柔らかいドメインが大きく歪むことにより応力が緩和され、マクロな弾性率が低くなるためである。また、溶融ドメインを含む T_2 条件の方が、相分離による軟化

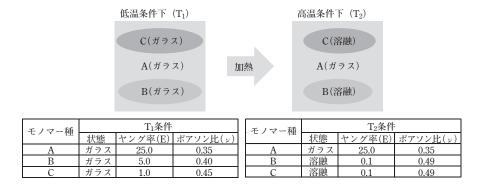


Fig. 4 仮想モノマーの弾性率仮定値

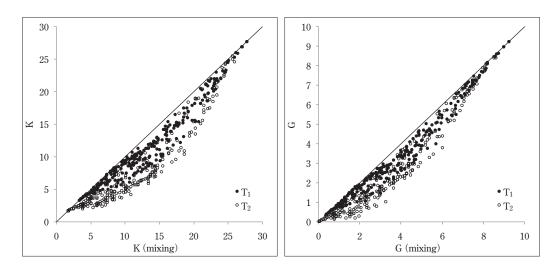


Fig. 5 FEM 計算値(縦軸)とモノマー数からの加成則値(横軸)

の寄与が大きい。

このように、ミクロ相分離構造を持つ材料のマクロ な弾性率は、相分離ドメインの大きさや形状、ドメイ ンの弾性率に強く依存するため、モノマーの体積分率 から単純な線形回帰で求めることができない。

(2) 説明変数の最適化

機械学習において、入力データとそのフォーマット、アルゴリズムの選定は結果に大きな影響を与える (Fig. 6)。今回、アルゴリズムとしてサポートベクトル回帰 (SVR) を選定した。SVR はカーネル法を用いたノンパラメトリックな非線形回帰アルゴリズムであり、関数やパラメータをあらかじめ指定することなく複雑な相関関係を学習できるため、相分離のように複

雑な非線形関係が予想される場合に適している。

入力データとしては、ブロック共重合体のモノマー配列と相互作用パラメータ(a_i 値)を入力した。モノマー配列については、どのような形式で表現するかが課題となる。そこで Table 1 の 3 方式を試験したところ、Fig. 7 の学習曲線を得た。学習曲線は教師データを増やした時の学習精度の変化をグラフ化したものである。序数としてのモノマー種とブロック長を組み合わせたシーケンス形式(Table 1 ①)や、単純に序数化したモノマー種を 100 個並べた序数形式(Table 1 ②)は人間の感覚的には一見して理解しやすいが、 R^2 < 0.6 であり、よい学習結果を与えなかった。対して、各モノマーの有無を二進数で表し、全 100 モノマーを

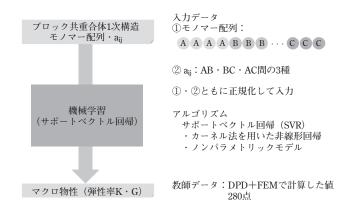


Fig. 6 機械学習スキーム

Table 1 モノマー配列のベクトル設計

① sequence format

 \cdot monomer type(A=1, B=2, C=3) \cdot number of monomer $\times 5$

: 10 dimensions

	block1		block2		block3		block4		block5	
	monomer	mumber of								
	type	monomer								
A50C50	1	50	3	50	0	0	0	0	0	0
B1A49C50	2	1	1	49	3	59	0	0	0	0

2 ordinary format

· monomer type (A=1, B=2, C=3) \times 100

	monomer 1	monomer 2	 monomer 49	monomer 50	•••	monomer 100
A50C50	1	1	 1	3		3
B1A49C50	2	1	 1	3		3

3 binary format

· binary monomer types $\times 100$

: 100 dimensions

:	300	dimer	isions
---	-----	-------	--------

	y															
	monomer1		m	onome	r2		mo	onome	49	mo	onome	r50	 mo	nomer	100	
	A	В	С	A	В	С		A	В	С	A	В	С	 A	В	С
A50C50	1	0	0	1	0	0		1	0	0	0	0	1	 0	0	1
B1A49C50	0	1	0	1	0	0		1	0	0	0	0	1	 0	0	1

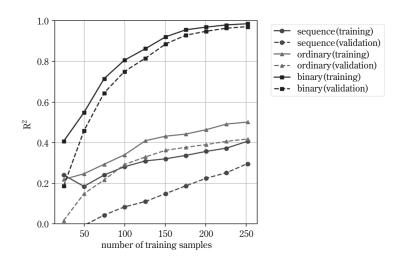


Fig. 7 入力ベクトルごとの学習曲線 (K, T₁条件)

300 要素のベクトルで表した場合(Table 1 ③)、 $\mathbb{R}^2 > 0.9$ と極めて効果的な学習がなされた。

これは、類似した配列が類似したベクトルで表現される場合、学習効率が向上することを示している。例えば、A50C50 と B1A49C50 のポリマーはモノマーが1つ交換したのみであり、ほぼ同様の物性を示すのが自然である。これらをシーケンス形式(Table 1①)で表現した場合、全く異なるベクトルになるのに対し、二進数形式(Table 1③)ならばモノマー1の行が変わるのみでほとんど同一に近いベクトルとなる。

(3) 系の対称性の利用

教師データの数には限りがあるが、対象とする系に 対称性が存在する場合、対称性を取り入れることで見 かけ上のデータ数を増やし、学習効率を向上させるこ とができる。 この系の場合、直鎖ポリマーに前後が存在しないため、逆シーケンス (例: A50B50 ⇔ B50A50) は同一のポリマーであり、したがって等しい物性を与える。逆シーケンスの組成を追加することにより、実質的にデータ数を 2 倍に増やすのと同様の効果を得ることができた (Fig. 8)。

(4) 各説明変数の感度の検証

マクロな弾性率は、各モノマーの数 (=体積分率) と相分離構造 (モノマー配列・χパラメータ) によって決定されると考えられる。それぞれがどの程度学習精度に寄与しているかを調べる為に、情報を意図的に欠損させて学習を行った (Table 2)。

 T_1 条件において、モノマー数のみを入力した場合 の結果は $R^2 = 0.943$ であり、これが体積分率のみか ら推測できる上限であると思われる。これに対し、モ

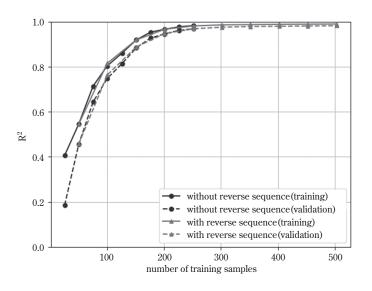


Fig. 8 逆シーケンスの追加による学習曲線の延長 (K, T₁条件)

Table 2 各説明変数の感度の検証(K, T₁, T₂条件)

T. 多件

11 / 11		
	モノマー配列	モノマー数のみ
a _{ij} あり	$R^2 = 0.989$	$R^2 = 0.951$
a _{ij} なし	$R^2 = 0.982$	$R^2 = 0.943$

T₂条件

	モノマー配列	モノマー数のみ
a _{ij} あり	$R^2 = 0.978$	$R^2 = 0.857$
a _{ij} なし	$R^2 = 0.964$	$R^2 = 0.839$

ノマー配列を入力すると $R^2 = 0.982$ に上昇し、さらに χ パラメータに対応する a_{ij} を加えると 0.989 まで精度が向上した。以上より、機械学習は一次構造に由来する相分離構造の差を学習していることがわかる。興味深いことに、モノマー配列を与えずとも、 a_{ij} を加えるのみでも精度の向上が見られた。これは、 a_{ij} が相分離構造と組成に相関するためと考えられる。

また、ガラス・溶融混合系である T_2 条件では、モノマー配列・ a_{ij} の精度向上効果はより大きく表れた。これは、両ドメインの弾性率が極端に違う系においては、弾性率に相分離構造の影響がより顕著に現れるからであると思われる。

以上より、ミクロ相分離構造の弾性率はモノマーの体積分率に最も影響されるが、相分離状態の影響も受けること、後者にはχパラメータより一次構造が強く影響していることがわかる。このように、マクロ物性に材料組成のどの変数が物性に対してどの程度の決定力を持っているのか検証することができた。

(5) ランダム生成組成の推算と実行速度

得られた SVR を用いて新たに 400 点の ランダム組成を予測し、DPD-FEM の検算結果と比較し、 $R^2=0.989$ (K, T_1 条件)と、新規データも学習時と同等の精度で予測できることを確認した。

また、実行速度を比較したところ、シミュレーショ

ン(DPD-FEM)に対して SVR による推算は極めて 高速であった(**Table 3**)。これにより、1000 組成の 網羅的推算も 1 秒以下で実行できた(**Fig. 9**)。1000 組成をスクリーニングすることにより、この系におい て設計可能な弾性率の領域がほぼ露わとなったのが見 てとれる。また、 $T_1 \cdot T_2$ それぞれの条件で任意の弾性 率を満たす組成を逆算することもできる。

3. まとめ

ブロック共重合体の弾性率を OCTA の散逸粒子動力学 (DPD) と有限要素法 (FEM) により計算し、それを教師データとして機械学習を検討した。得られた推定器はブロック共重合体組成から弾性率を高精度かつ短時間で推算することができた。

高分子シミュレーションの結果を教師データとした 機械学習は、類似した系を大量に対象とする必要があ る場合、短時間・低負荷でのスクリーニングを可能と

Table 3 各工程の実行時間

工程	計算時間
DPD	7.83 時間/1組成 (4コア)
FEM	140 秒 / 1 組成 (1コア)
SVR	0.36 秒 / 100 組成(1 コア)

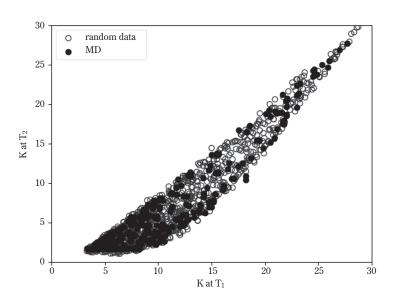


Fig. 9 機械学習による網羅的スクリーニング結果

する。本研究はシンプルで小規模な粗視化モデルであるが、今後の展開として、全原子シミュレーションへの適用や、弾性率以外の物性推算への応用等が想定される。高分子シミュレータと機械学習のシナジーによる計算科学の更なる発展が期待される。

4. 謝 辞

本研究は(公社)新化学技術推進協会の高分子シミュレーション技術セミナーにおいて行われました。以下のメンバーの皆様に感謝いたします。

滝本 淳一(山形大院・有機)・青柳 岳司(産総研)・ 諸星 圭(トヨタ自動車)・島 広志(ブリヂストン)・ 岩方 裕一(リンテック)・永井 良典(日立化成)・岩 間 立洋(旭化成)・槻尾 大輔(カネカ)・阿部 敬太 (3M Japan)・堀江 一司(日本触媒)・清水 陽平(ジェイテクト)(順不同・敬称略)。

参考文献

- [1] I-S. Han, C. Han, C-B. Chung, *J. Appl. Polym. Sci.* 95, 967 (2005)
- [2] A.V. Gubskaya et al., *Polymer*, 48, 5788 (2007)
- [3] T. Matsumoto, K. Tanabe, *JCPE journal*, 12, 49 (2000)
- [4] T. Aoyagi, T. Honda, and M. Doi, *J. Chem. Phys.* 118, 8153 (2002)
- [5] R. D. Groot, P. B. Warren, J. Chem. Phys. 107, 4423 (1997)
- [6] http://octa.jp/jp/
- [7] A. J. Smola, B. Schölkopf, Stat. and Compt. 14, 199 (1998)
- [8] http://scikit-learn.org/