1130 A Data-driven Scheme to Search for Alternative Composite Materials

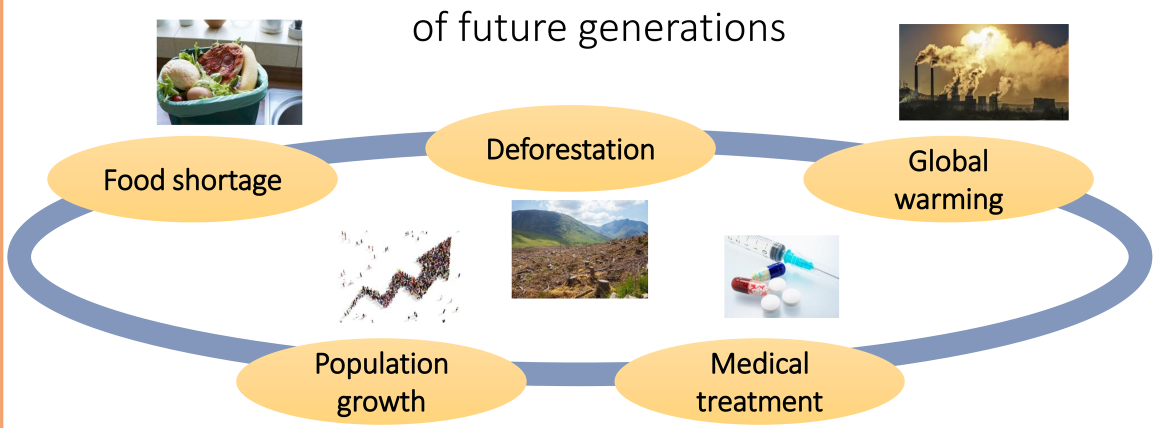
Michihiro Okuyama[a)], Yukihito Nakazawa[a)] and Kimito Funatsu[b)]

a) Konica Minolta, Inc, b) Nara Institute of Science and Technology

1. Introduction

Realization of a sustainable society based on data and in-house technology

Increasing problems that negatively affect the lives of future generations



Solving the above problems through new technologies that combines data and in-house technology

Initiatives to Achieve the Goal: Reduction of industrial waste (especially composite materials)

Solution

Development of composites substituted with recycled materials and biomass.

Existing composites (ex: products)	Constituent 1	Constituent 2	Constituent 3
Substituted composites	Constituent 1	Constituent 2	Recycled materials or Biomass

- Numerous candidate combinations of constituents and addition ratios
- Extremely difficult to experiment across all candidates

Data-driven development is essential

Obstacles to data-driven development of composite materials

- Interactions between constituent materials
- Physicochemical properties of constituent materials are unknown
- Time-consuming to measure performance

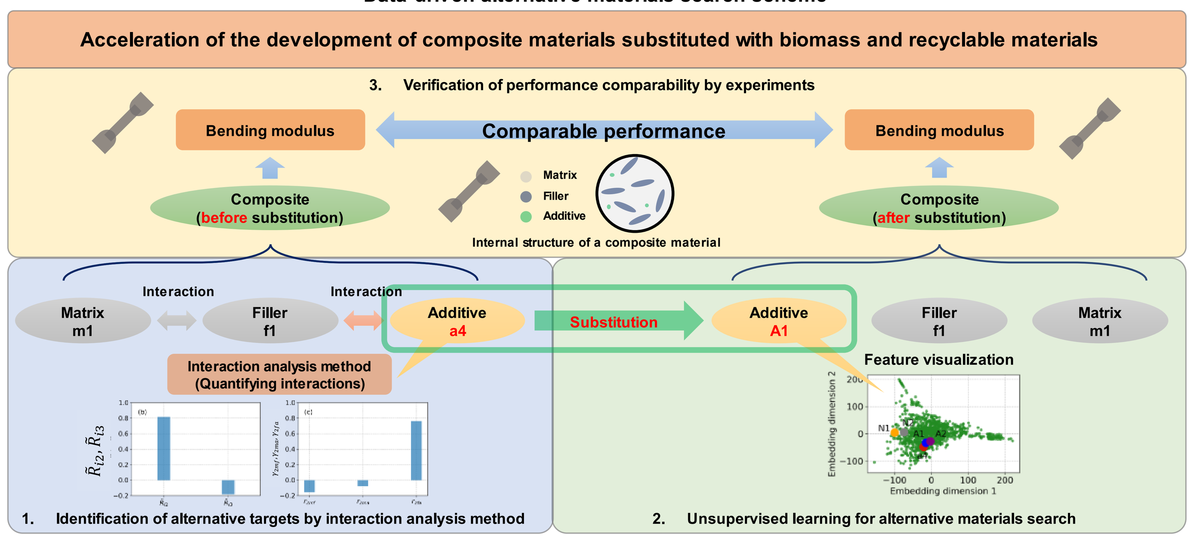
Issues to be solved

A) Quantification of interactions
B) Determination of alternative targets based on evaluation of interactions
C) Material search based on unsupervised learning

Important to establish a data-driven materials exploration strategy that resolves A) - C)

2. Strategy and method

Our Strategy with machine learning*[1]



Newly developed methods to implement our strategy

Method	Detail
Interaction analysis method	Quantifying interactions
Unsupervised learning method for alternative material search	Material Search without Regression

Interaction analysis method

A) Quantification of interactions

Interaction $x_i^{n_i}x_j^{n_j}$ $x_i^{n_i}x_j^{n_j}x_k^{n_k}$ Left: 2-factor interaction Right: 3-factor interaction

As an example, here, consider composite materials consisting of three constituents

Predictive model $\ln y = c_0 + \sum_i c_i x_i$ y : Performance (ex: Bending modulus) x_i: Addition ratio of constituents

Brief definition of interaction indicators (I - IV)

1. Solve the predictive model for y $y_\alpha = \mathcal{A}e^{c_{m_\alpha}x_{m_\alpha}}e^{c_{f_\alpha}x_{f_\alpha}}e^{c_{a_\alpha}x_{a_\alpha}}$ $\mathcal{A} = e^{c_0}$
2. McLaurin expansion $y_\alpha = \mathcal{A}\sum_{n_m,n_f,n_a=0}^{\infty}\left(\frac{c_{m_\alpha}^{n_{m_\alpha}}c_{f_\alpha}^{n_{f_\alpha}}c_{a_\alpha}^{n_{a_\alpha}}}{n_{m_\alpha}!n_{f_\alpha}!n_{a_\alpha}!}\right)x_{m_\alpha}^{n_{m_\alpha}}x_{f_\alpha}^{n_{f_\alpha}}x_{a_\alpha}^{n_{a_\alpha}}$ Interaction

The following equation is used $\sum_{n=1}^{\infty}\frac{1}{n!}c^n x^n = e^{cx}-1$

3. Separate each interaction and transform the equation into a computable form

$$
\begin{aligned}
y_\alpha = \mathcal{A}\{1 &+ \underbrace{c_{m_\alpha}x_{m_\alpha} + c_{f_\alpha}x_{f_\alpha} + c_{a_\alpha}x_{a_\alpha}}_{\tilde{y}_l} \\
&+ \underbrace{(e^{c_{m_\alpha}x_{m_\alpha}}-1) + (e^{c_{f_\alpha}x_{f_\alpha}}-1) + (e^{c_{a_\alpha}x_{a_\alpha}}-1) - c_{m_\alpha}x_{m_\alpha} - c_{f_\alpha}x_{f_\alpha} - c_{a_\alpha}x_{a_\alpha}}_{\tilde{y}_{n_1}} \\
&+ \underbrace{(e^{c_{m_\alpha}x_{m_\alpha}}-1)(e^{c_{f_\alpha}x_{f_\alpha}}-1) + (e^{c_{m_\alpha}x_{m_\alpha}}-1)(e^{c_{a_\alpha}x_{a_\alpha}}-1) + (e^{c_{f_\alpha}x_{f_\alpha}}-1)(e^{c_{a_\alpha}x_{a_\alpha}}-1)}_{\tilde{y}_{n_2}} \\
&+ \underbrace{(e^{c_{m_\alpha}x_{m_\alpha}}-1)(e^{c_{f_\alpha}x_{f_\alpha}}-1)(e^{c_{a_\alpha}x_{a_\alpha}}-1)}_{\tilde{y}_{n_3}}\}
\end{aligned}
$$

4. Definition of an indicator of interaction based on the formula in 3.

I. Proportion of nonlinear terms among linear and nonlinear terms

$$R_n = \frac{\tilde{y}_n}{|\tilde{y}_n| + |\tilde{y}_l|} = \frac{Nonlinear\ term}{|Linear\ term| + |Nonlinear\ term|}$$

$$\tilde{y}_n = \tilde{y}_{n1} + \tilde{y}_{n2} + \tilde{y}_{n3} \qquad \tilde{y}_l = c_{m_\alpha}x_{m_\alpha} + c_{f_\alpha}x_{f_\alpha} + c_{a_\alpha}x_{a_\alpha}$$

a. $|R_n|$: Proportion of the nonlinear term among the linear and nonlinear terms
b. $R_n > 0$: Contributes to increasing performance
c. $R_n <= 0$: Contributes to lowering performance

We defined the remaining three indicators with properties a. – c.
(The indicators, I - IV are used to quantitatively evaluate the interaction between the constituents.)

II. Proportion of interactions in the total nonlinear term

$$\tilde{R}_i = \frac{\tilde{y}_i}{|\tilde{y}_{n1}| + |\tilde{y}_i|} = \frac{Interaction}{\underbrace{|1-factor\ nonlinear\ term|}_{x_i^{n_i}} + \underbrace{|Interaction|}_{\tilde{y}_{n_2}+\tilde{y}_{n_3}}}$$

III. Proportion of each interaction in the interaction

$$\tilde{R}_{ia} = \frac{\tilde{y}_{na}}{|\tilde{y}_{n2}| + |\tilde{y}_{n3}|} = \frac{2(or\ 3) - factor\ Interaction}{|Interaction|} \qquad a = 2, 3$$

IV. Proportion of two-factor interactions that are accounted for by interactions between constituents

$$r_{2\gamma\delta} = \frac{(e^{c_\gamma x_\gamma}-1)(e^{c_\delta x_\delta}-1)}{|(e^{c_{m_\alpha}x_{m_\alpha}}-1)(e^{c_{f_\alpha}x_{f_\alpha}}-1)| + |(e^{c_{m_\alpha}x_{m_\alpha}}-1)(e^{c_{a_\alpha}x_{a_\alpha}}-1)| + |(e^{c_{f_\alpha}x_{f_\alpha}}-1)(e^{c_{a_\alpha}x_{a_\alpha}}-1)|} \qquad \gamma, \delta = m_\alpha, f_\alpha, a_\alpha$$

$$= \frac{Interaction\ AB}{|Interaction\ AB| + |Interaction\ AC| + |Interaction\ BC|}$$ (Example for constituent A and B)

Unsupervised learning method for alternative material search

Information used for search

- Physicochemical properties of materials to be substituted
- Physicochemical properties of alternative candidate materials

How to search

Search among alternative candidates for materials with similar physicochemical properties to those of the material to be substituted*[2]

Method

Manifold learning : ISOMAP*[3]

In this presentation, we use the physicochemical properties as features transformed from 2D molecular descriptors computed from alvaDesc*[4] to principal components.

3. Case study

Application to Fiber Reinforced Plastics

Alternative material search for the constituents involved in the interactions that contribute the most to performance.

Target Composite Materials

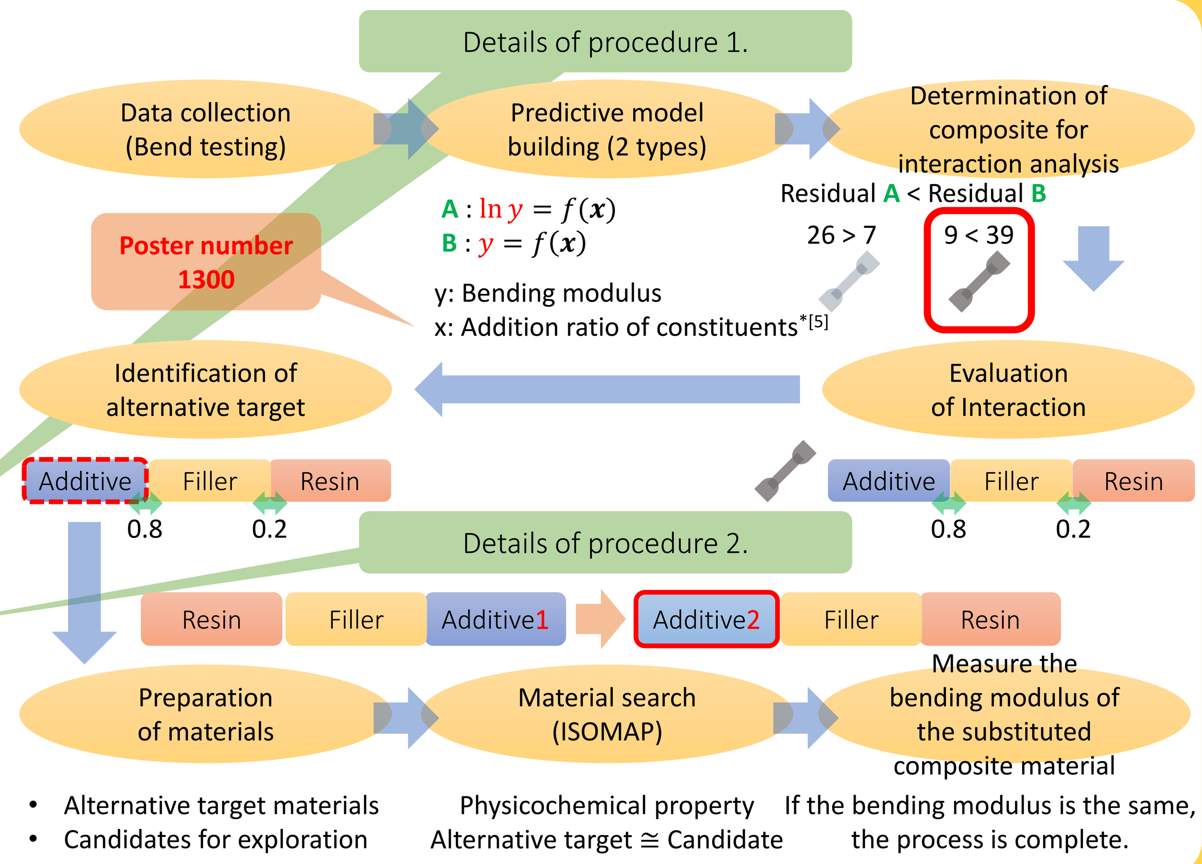
Composite consisting of three different constituent materials

Resin | Filler | Additive

Overview of search procedure

1. Determine the composite that contains the constituents involved in the interaction that contributes the most to the performance.
2. Design a new composite material with performance equivalent to that of the composite in 1. (the composite before substitution) by substituting that constituents with another one.

- Performance: Bending modulus
- Additives: Organic compounds with low molecular weight

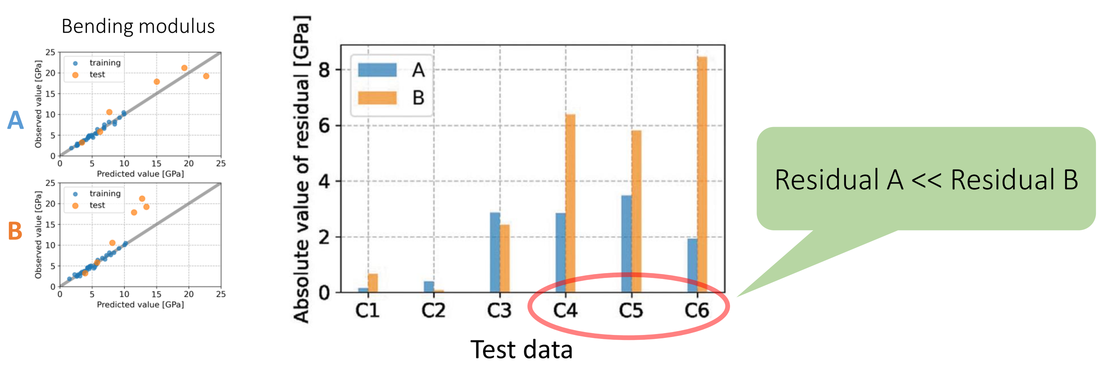


Predictive model building and determination of composite for interaction analysis

Training data	Test data
32	6

Method : Partial Least Squares Regression (PLS)

Select the target of interaction analysis from the composite of test data



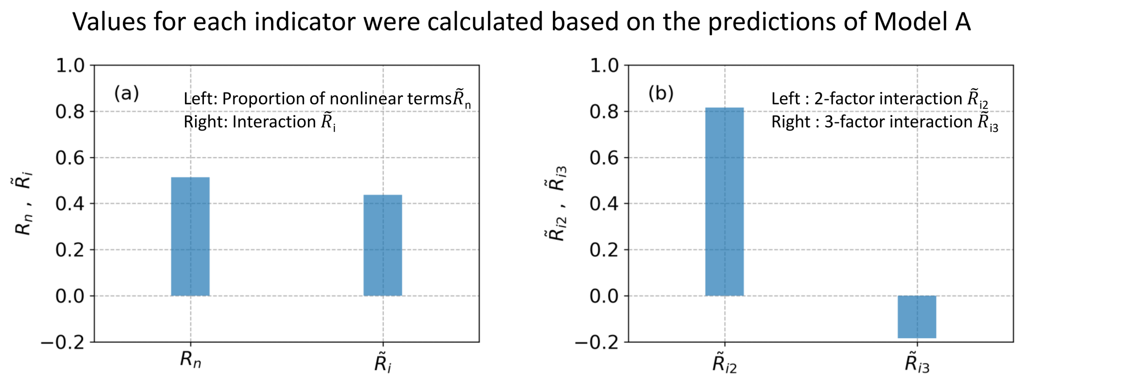
Composites with non-negligible interactions are C4 - C6

In this case, we chose C4 as the target of analysis for the interaction analysis method.

Interaction Analysis of Composite C4

Proportion of non-linear terms (R_n), and interactions ($\tilde{R}_i, \tilde{R}_{i2}, \tilde{R}_{i3}$)

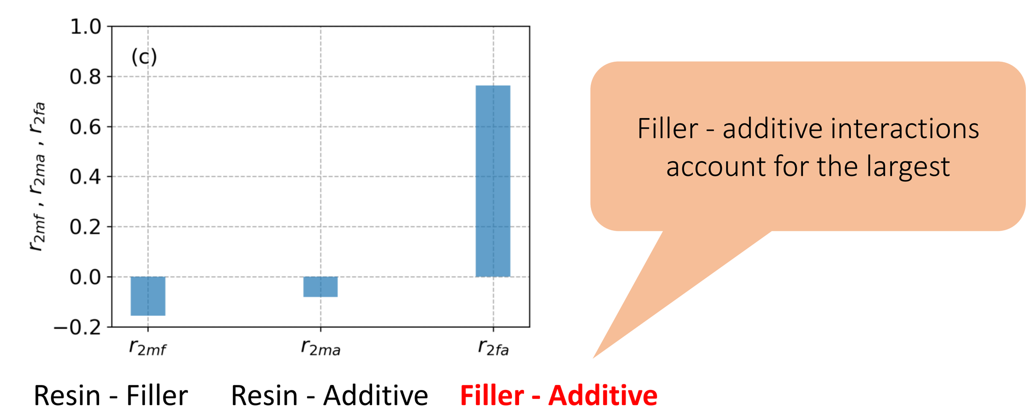
Values for each indicator were calculated based on the predictions of Model A



- The proportion of the nonlinear term and that of the linear term are the same. ($R_n \cong 0.5$)
- The proportion of the one-factor nonlinear term and that of the interaction are almost the same. ($\tilde{R}_i \cong 0.4$)
- The interaction consists almost entirely of two-factor interactions. ($|\tilde{R}_{i2}| \cong 0.8$)

The interaction that contributes the most to the bending modulus is the 2-factor interaction.

Proportion of interactions between constituents



Elements involved in interactions that contribute most to performance: filler and additive

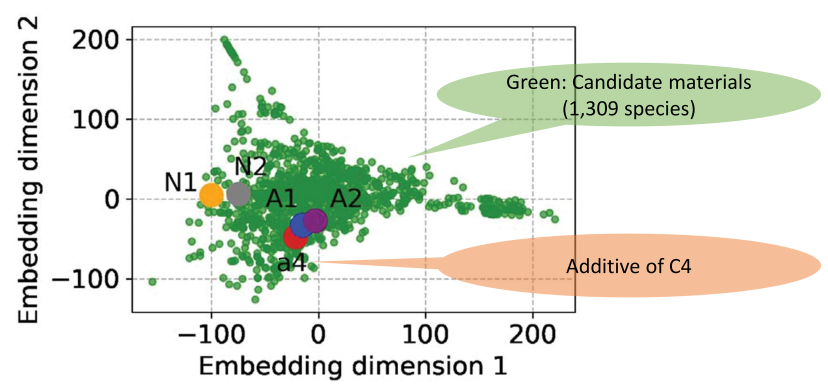
In this study, we will search for constituents that can substitute for additives of known chemical structure.

Unsupervised learning method for alternative material search

Search using ISOMAP

The following compounds were selected as experimental candidates:

- Compounds most similar to a4 features (A1, A2)
- Compounds not similar to a4 (N1, N2)



Experimental validation

- Bending modulus of A1 and A2 is equivalent to that of a4.
- Bending modulus of N1 and N2 are significantly different from a4

- Input for ISOMAP: Principal components converted from AlvaDesc 2D molecular descriptors (3,885)
- Candidate materials: Organic compounds with molecular weights of 100 - 900 in TCI*[6]

Results of bending test

Unit : GPa Δ: Absolute difference between bending modulus of 1 - 4 and C4

Composite	Additive	Similarity to a4	Bending modulus	Δ
C4	a4	-	17.89	0.0
1	A1	Similar	18.51	0.62
2	A2	Similar	18.80	0.91
3	N1	Not similar	20.10	2.21
4	N2	Not similar	12.86	5.03

(experimental value)

The bending modulus of C4 is value close to that of C4

By substituting the additive before substitution with an additive similar to the a4 feature, the bending modulus was obtained equivalent to that of the composite before the substitution.

As a result, this method is expected to be useful in the search for alternative materials for composite materials.

4. Summary and References

Summary

Develop a data-driven scheme to develop new composite materials with performance equivalent to that before substitution.

- Interaction Analysis Method
- Unsupervised learning alternative material search scheme including the above methods

The usefulness of this method was confirmed by applying it to a composite consisting of three materials: resin, filler, and additive.

- Search for alternative materials for the constituent (additive) involved in the interaction that contributes the most to the performance.
- Verify that materials with similar features to those of the additive before the substitution have the same performance as the composite before the substitution.

References

[1] M. Okuyama, et. al., Sci. Technol. Adv. Mater. Meth., 2, 118, (2022).
[2] Y. Zhang, et. al., Nat. Commun., 10, (1), 1. (2019).
[3] J. B. Tenenbaum, et. al., Science., 290, 2319, (2000).
[4] A. Mauri, https://www.alvascience.com/alvadesc/
[5] Y. Ikeda, et. al., J. Comput. Chem. Jpn. Int. Ed., 7, 2020, (2021).
[6] S. Kim, et. al., Nucleic. Acids. Res., 8, D1388, (2021).

Poster number 1300