Accelerated Materials Discovery by Materials Informatics: Informatics-Enabled Strong Polymer Composite Discovery

Jun AMANO＊, Karsten BRUENING＊, Michael MCBRIDE＊＊, Nils PERSSON＊＊, Ezgi DOGAN-GUNER＊＊, Aaron LIU＊＊, Zihao QU＊＊, Guoyan ZHANG＊＊, Elsa REICHMANIS＊＊, Carson MEREDITH＊＊, and Martha GROVER＊＊

要旨

マテリアルズインフォマティクス（MI）は，多様な材料データの取得，管理，分析，普及により，新素材の開発，生産，導入にかかる時間，リスク，コストの削減を可能とする。たとえば，新しいポリマー複合材料発見の時間とコストをMIを使って削減することで，材料開発イノベーションインフラを大幅に改善でき，新たに発見されたポリマー複合材料を様々な用途に迅速に展開できる。材料開発イノベーションインフラの改善による経済的利益の推定は，米国だけでも1,000億ドル／年を超えている。MIを有効に活用するには，データマイニングによる文献データの収集に加えて，多量の有効なデータをいかにスピーディに低コストで生成させることも重要である。

今回，我々は産業的に重要なポリマー複合材料の開発において必要なMIのインフラを開発したので，報告する。インフラは材料種を文献からのデータマイニングによる選択，技術的に非常に困難なソリューションプロセスで組成比を変化させたサンプルを迅速に作成するハイスループット高温製膜装置，およびハイスループット物性測定装置から成り立つ。本インフラは，様々な複合材料の開発に適用でき，MIによる新高機能性複合材料の迅速な開発に貢献すると期待される。

Abstract

Materials Informatics (MI) enables the reduction of time, risk, and cost for development, production, and deployment of a new material by acquisition, management, analysis, and dissemination of diverse materials data. For example, MI will significantly improve materials innovation infrastructure by reducing the time and cost of new polymer composite discovery resulting in a rapid deployment of the newly discovered polymer composite to various applications. The estimated overall economic benefit of improved materials innovation infrastructure is over hundred billion $/year in USA alone. In order to make effective use of MI, It is important to generate a large amount of effective data quickly at a low cost in addition to the collection and analysis of published literature data by data mining.

We report development of the infrastructure consisted of data mining for selection of industrially important polymer composite materials and process, a high-throughput (HT) and high-temperature gradient composition film coater using technologically very challenging solution process for quickly creating the samples with various material species and composition ratios critical for composite materials discovery, and high-throughput characterization tool to quickly measure physical properties. This infrastructure can be applied to the development of various composite materials and is expected to contribute to the rapid development of new high-functional composite materials.

＊Konica Minolta Laboratory USA, Konica Minolta Business Solutions U.S.A., Inc.
＊＊Chemical & Biomolecular Engineering, Georgia Institute of Technology

1 Introduction

Traditionally, the discovery of new materials depends on the domain knowledge of experienced materials scientists and relies on a process with many trials and errors. It usually takes a decade to come up with new functional materials. In order to accelerate this highly expensive and time-consuming process of materials discovery, data-driven materials informatics approaches have become more popular among the materials science community. According to Wikipedia[1], "Materials informatics is a field of study that applies the principles of informatics to materials science and engineering to better understand the use, selection, development, and discovery of materials." The estimated overall economic benefit of improved materials innovation infrastructure is over hundred billion $/year in USA alone[2].

We have conducted a collaborative research project between Konica Minolta and Georgia Institute of Technology (GIT). The goal was to develop the infrastructure consisted of selection of industrially important polymer composite materials and process by data mining of the published papers, a high-throughput and high-temperature gradient composition film coater using technologically very challenging solution process for quickly creating the samples with various material species and composition ratios critical for composite materials discovery and high-throughput characterization tool to quickly measure physical properties.

In this article, the data-driven materials informatics process, e.g., data mining, high-throughput experimentation for data creation, and data analytics using machine learning (ML) is described. The main process workflow is shown in Fig. 1.

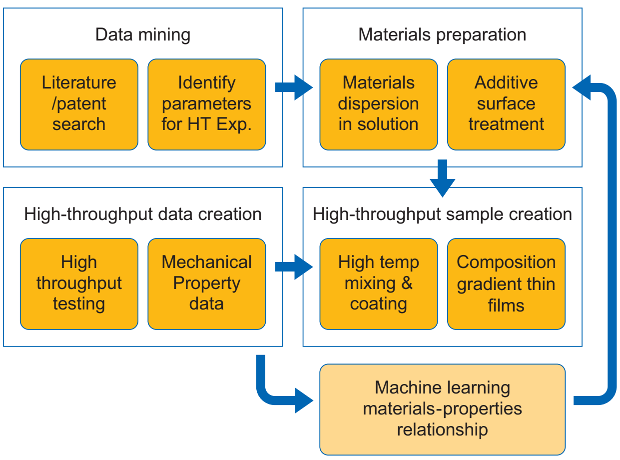

Fig. 1 Overall process workflow for discovery of high- strength polymer composite.

2 Data mining

One of the key processes in materials informatics is a careful data mining through published literature and patents. If a large volume of literature is available, it is possible to discover new materials solely by using an advanced text mining algorithm, such as Natural Language Processing (NLP), and careful analysis thereof, without experimentations[3]. For the case of limited data availability, one can use several ML algorithms, such as clustering and regression analysis to identify target materials with the desired properties[4]. For polymer composites, the available literature describing mechanical properties are often only on the order of hundreds. In those cases, it is better to use selected databases of high-performance materials for a new composite discovery[5].

After completion of the data mining process to define a promising polymer matrix and additives for the target mechanical properties, high-throughput experimentation needs to be conducted to validate the findings of the data mining and produce a sufficient database used for further machine learning-guided optimization.

3 High-throughput sample creation

In order to accelerate the fabrication of polymer composite samples and to probe multiple compositions in one sample, we developed a tool that produces polymer composite films with a gradient in composition using solution processing. Fig. 2 shows a schematic of such system; the two-channel composition gradient blade coater for polymer composite films with the mixing system and moving coating stage. In general, many technologically interesting polymers produce several technical challenges in solution processing due to low ambient solvent interactions and high viscosities. Accordingly, the automated flow, passive mixing, and blade coating systems must be capable of high temperature operation for solution-casting of polymer composite films. Fig. 3 shows a picture of a custom-designed passive high-temperature microfluidic chaotic mixer replacing a conventional drum and agitator mixer. The passive mixer consists of chevron-patterned mixing path suitable for flow of high viscosity solutions and filler particles. This system is capable of fabrication of gradient films with spatial variations in composition. Fig. 4 shows an example of the spatially gradient composite film. The left side of the film consisted of 100% polymer, and the right side of the film consisted of the highest additive concentration. The dye was added to the additive in order to show the gradient color change corresponding very smooth compositional gradient.

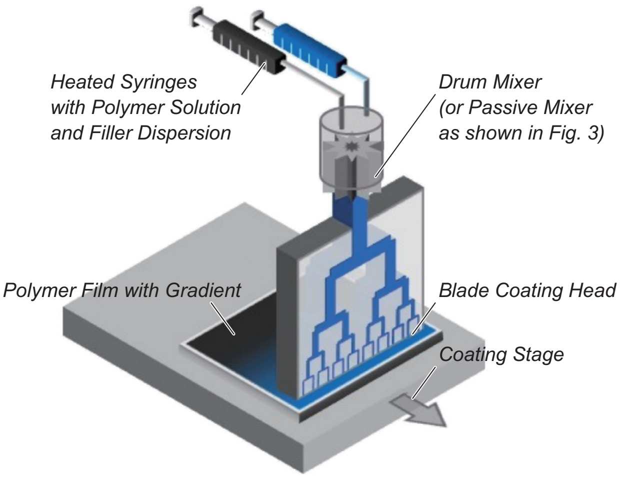

Fig. 2 A schematic diagram of two-channel composition gradient blade coater for polymer composite films with the mixing system and moving coating stage.

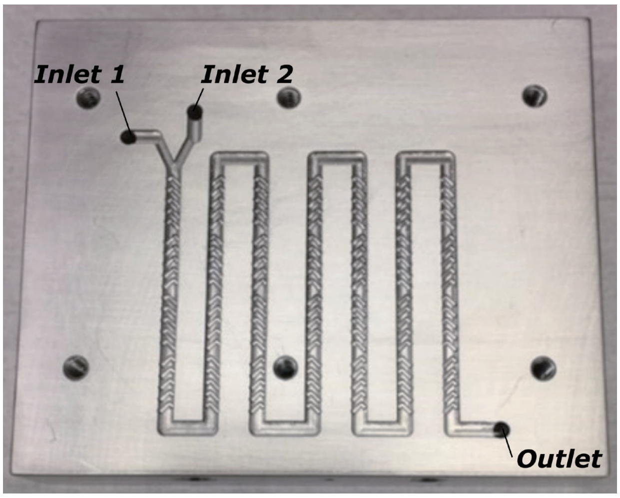

Fig. 3 Passive high-temperature micro fluidic chaotic mixer with a chevron patterned mixing path.

Fig. 4 Picture of a polymer composite film with very smooth compositional gradient.

4 High-throughput data creation

After fabricating gradient composition polymer composite films using the high temperature blade coater system, high-throughput mechanical property characterization is performed to rapidly map the mechanical properties in a large composition space. Fig. 5 illustrates the high-throughput mechanical characterization (HTMECH) tool to measure elastic (Young's) modulus and ultimate tensile strength of 100 points per sample with various compositions[6]. In the HTMECH tool, a pin attached to a force sensor pushes the composite films into small openings in the sample holder to measure elastic modulus and ultimate tensile strength, resulting in 100 data points per sample.

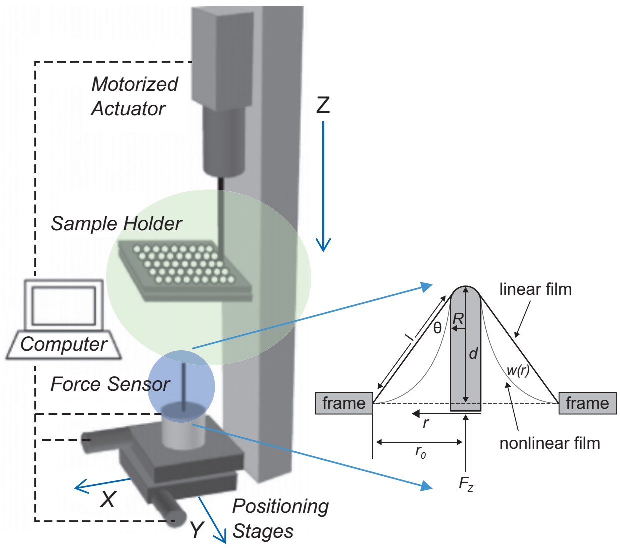

Fig. 5 High-throughput mechanical characterization (HTMECH) tool to measure elastic (Young's) modulus and ultimate tensile strength of 100 points/sample.

5 Machine learning

All measured data is now ready to be compiled and analyzed computationally to obtain materials-properties relationships and input this information back to materials preparation to start another experimentation to explore the optimum design space. At the initial stage of the Materials Informatics process, it is most helpful to use a traditional physical model-based machine learning process using manual feature extraction by an experienced materials scientist as indicated in Table 1. After a sufficiently larger labeled dataset has been obtained by completion of several experimental loops, deep learning with much fewer human inputs can be utilized. To obtain a reasonable accuracy in a new materials discovery process by deep learning, at least a few thousand labeled datasets may be required.

Some of the recent results of this project will be presented at the 2019 Fall Materials Research Society meeting in Boston, MA[7].

Table 1 Machine Learning vs. Deep Learning.

Parameters	Machine Learning	Deep Learning
Training dataset	Small	Large
Feature extraction	Manual	Automatic
Classifier	Many	Few
Training time	Short	Long

6 Conclusion

This article presents the workflow and tool sets to obtain materials-process-property relationships and to rapidly optimize polymer composites using a data-driven materials informatics approach with minimum time and resources requirements. This approach is applicable for many other materials systems.

Combining this approach with highly autonomous robotic experimental tools with deep learning/transfer learning processes, automated materials discovery and manufacturing system with a minimum human engagement may be possible in the near future.

Acknowledgements

The authors would like to thank Dr. Leiming Wang for initial phase of this research project.

References

1) https://en.wikipedia.org/wiki/Materials_informatics; retrieved on Sep 20, 2019
2) Scott T, Walsh A, Anderson B, O'Connor A, Tassey G, "Economic Analysis of National Needs for Technology Infrastructure to Support the Materials Genome Initiative", RTI Project Number 0215231, April 2018.
3) Tshitoyan V., Dagdelen J. Weston L., Dunn A., Rong Z., Kononova O., Persson K.A., Ceder G, Jain A. "Unsupervised word embeddings capture latent knowledge from materials science literature", Nature 571, 95–98 (2019).
4) Persson N, McBride M, Reichmanis E, and Grover M, "Machine learning approaches for extracting process-structure-property" Chapter 1. in Big Data Methods in Experimental Materials Discovery by World Scientific Publishing (November, 2019).
5) McBride M, Persson N, Reichmanis E, and Grover M, "Solving Materials' Small Data Problem with Dynamic Experimental Databases", Processes 6, (7), 79, (2018)
6) Sormana J-L, Chattopadhyaya S, and Meredith C, "High-throughput mechanical characterization of free-standing polymer films", Review of Scientific Instruments 76, 062214 (2005)
7) Dogan-Guner E, Liu A, McBride M, Grover M, Meredith C, Reichmanis E, Amano J, Bruening K, "Fabrication of Composition Gradient Polymer Films at Elevated Temperatures for High-Throughput Characterization", to be presented at 2019 Fall MRS meeting, Boston, MA.