Macromolecules

pubs.acs.org/Macromolecules Article

A Recommender System for Inverse Design of Polycarbonates and Polyesters

Nathaniel H. Park,* Dmitry Yu. Zubarev,* James L. Hedrick, Vivien Kiyek, Christiaan Corbet, and Simon Lottier

Cite This: *Macromolecules* 2020, 53, 10847−10854 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The convergence of artificial intelligence and machine learning with material science holds significant promise in rapidly accelerating the development timelines of new high-performance polymeric materials. Within this context, we report an inverse design strategy for polycarbonate and polyester discovery based on a recommendation system that proposes polymerization experiments that are likely to produce materials with targeted properties or characteristics. Following the recommendations of the system driven by the historical ring-opening polymerization results, we carried out experiments targeting specific ranges of monomer conversion and dispersity for polyesters or polycarbonates. The results of the experiments were in close agreement with the recommendation targets with few false negatives or positives obtained for each class.

INTRODUCTION

The accelerated discovery of new materials has the potential to broaden the economic impacts arising from both the decreased cost of development timelines and improved material performance.[1−5] To meet this challenge, traditional, labor-intensive research workflows are being redesigned to utilize a combination of automated or autonomous experimentation, high-throughput characterization, high-performance computing, and artificial intelligence to improve experimental quality and rate.[6−12] In this context, new automated synthesis platforms for polymeric materials are critical and offer significant benefits by enabling programmatic control over experimental outcomes and polymer features such as monomer conversion (MC), the degree of polymerization (DP), dispersity (Đ), and architecture.[13−17] Coupling of automated systems with advances in predictive models for experimental outcomes, polymer structural features, and polymer properties will undoubtedly afford progress toward the rapid development and commercialization of new materials with improved performance characteristics.[10,11,18]

A common problem, however, with automated systems is the additional labor required to utilize the system. This can be in the form of equipment setup or background experimentation needed to use the system effectively. For the ring-opening polymerization (ROP) of lactones and cyclic carbonates, background investigation on the reaction conditions are critical, as poor choice of conditions can lead to reactor fouling, undesired broadening of dispersity, or potentially no reaction at all in automated systems using continuous-flow.[19] The selection of appropriate conditions for controlled ROP of lactones and cyclic carbonates depends heavily on the interplay between monomer reactivity[20−23] and catalyst activity.[13,24−27] While the thermodynamics and kinetics of ROP have been well-studied,[20−23] these insights do not provide a robust theoretical framework to enable a priori catalyst selection. Especially, since even relatively small changes in the monomer structure can have an outsized influence on reactivity.[20,28] Instead, experimentalists must rely on these principles in combination with the available literature and their own experience to guide catalyst and reaction condition selection. Design of experiments—a statistical method to help determine the influence of different experimental factors—can be a powerful tool to help arrive at an optimized set of reaction conditions, but still necessitates a significant number of experiments to be run.[29,30] Thus, given the breadth of catalysts

Received: September 17, 2020
Revised: November 13, 2020
Published: November 27, 2020

https://dx.doi.org/10.1021/acs.macromol.0c02127
Macromolecules 2020, 53, 10847−10854

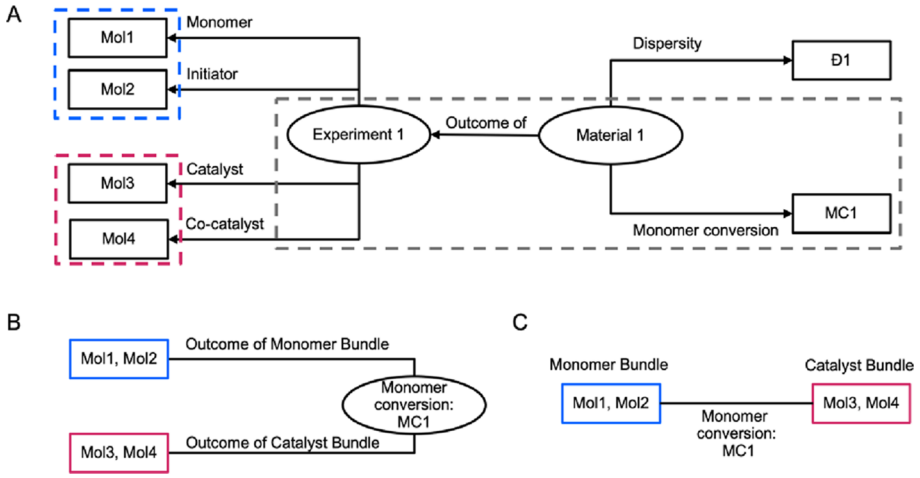

Figure 1. Transformation of the "experiment knowledge graph" (eKG) into a network amenable to link prediction via representation learning. Panel A: an RDF graph constructed from the relational database of the experimental parameters and outcomes (not shown). Continuous experimental parameters and measured properties are converted into categorical values via binning. Dashed outlines show the patterns of node collapse. Panel B: simplified RDF graph where the nodes of the experimental parameters (objects) are collapsed forming bundles and the categorical value of the measured property is treated as a subject. The same procedure applies to MC (shown) and Đ (not shown). Panel C: RDF triples are transformed into a bipartite graph, where nodes representing bundled experimental parameters (RDF objects) are directly connected to each other via a link; the value of the measured property (RDF subject) is assigned to the link as an attribute.

available for ROP,[31−33] a predictive framework capable of using historical data to match catalysts to lactone and carbonate monomers for controlled polymerization would both dramatically reduce experiment overhead and accelerate material development timeframes using automated synthetic platforms.

The standard computational selection strategy for evaluating monomer−catalyst interaction would rely on the direct evaluation of thermodynamic and kinetic parameters via state-of-the-art quantum chemistry. However, the computational task explodes when accounting for the experimental conditions needed for polymer design. Instead, complementary, data-centric approaches rooted in statistical learning and exhibiting a different balance of advantages and disadvantages has been gaining relevance lately.[34−43] In the chemical domain, the bulk of machine learning (ML) studies, including deep learning (DL) as a subset of ML, rely on the construction of structure−property and structure−activity relations informed by the structural features of chemical compounds. As polymers are stochastic macromolecules, establishing the exact make-up of the ensemble of polymer chains—particularly those with cross-links or network interpenetration—is often impractical both from computational and experimental perspectives. Therefore, identifying alternatives to the molecular structure of the polymer chains as the source of features in statistical learning tasks is critical. Assuming that materials are produced in a deterministic manner, consistency in the experimental conditions guarantees a degree of reproducibility in the observed material characteristics and properties that is sufficient for data-centric modeling. In addition to circumventing these issues with the stochastic nature of polymers, the advantage of using experiment−property relationships is avoiding the common bottleneck in models focused on the structure−property relationships: determining a route to synthesize the predicted material candidate. Instead, using experiment−property relationships enables the model to cast predictions in terms of experimental conditions that are directly actionable by researchers.

Having identified experiment−property modeling as the overarching theme of our efforts, we should distinguish the cases of the forward and inverse material design. The forward design of polymers encompasses models that enable one-to-one mapping of polymer structure to the property. The structure can be captured explicitly, such as the list of atomic positions in three dimensions, or implicitly, such as parameters of the experiments producing the polymer. The inverse material design includes the models that map polymer characteristics or properties on the structure in a one-to-many manner. In other words, given a value or range of polymer properties or characteristics, the inverse design model is expected to produce several distinguishable hypotheses about material structure—or experimental conditions to provide such a structure—that would exhibit the desired property or characteristic.[18,44−46] Inverse design strategies offer higher computational efficiency of the design and discovery process in terms of the computational cost and time to solution. In practice, the forward and inverse strategies should be seen as complementary in order to ensure the flexibility and robustness of the computational design and discovery workflow. Thus, in order to improve the experimental outcomes for the ROP of cyclic carbonates and lactones, we report an inverse design strategy for hypothesis generation about experimental conditions that are likely to produce polyesters or polycarbonates with predefined characteristics—in this case, the experimental outcome of monomer conversion to polymer and the dispersity of the obtained polymer. Specifically, given the target range of the measurable property or characteristic of a polymer, we generate recommendations for the values of the experimental degrees of freedom, both categorical and quantitative, that are immediately actionable.

RESULTS

The recommendation engine for experimental conditions performs link completion on the network of the historical experiments—in other words, recommender "links" catalysts with monomers thereby recommending experimental con-

https://dx.doi.org/10.1021/acs.macromol.0c02127
Macromolecules 2020, 53, 10847−10854

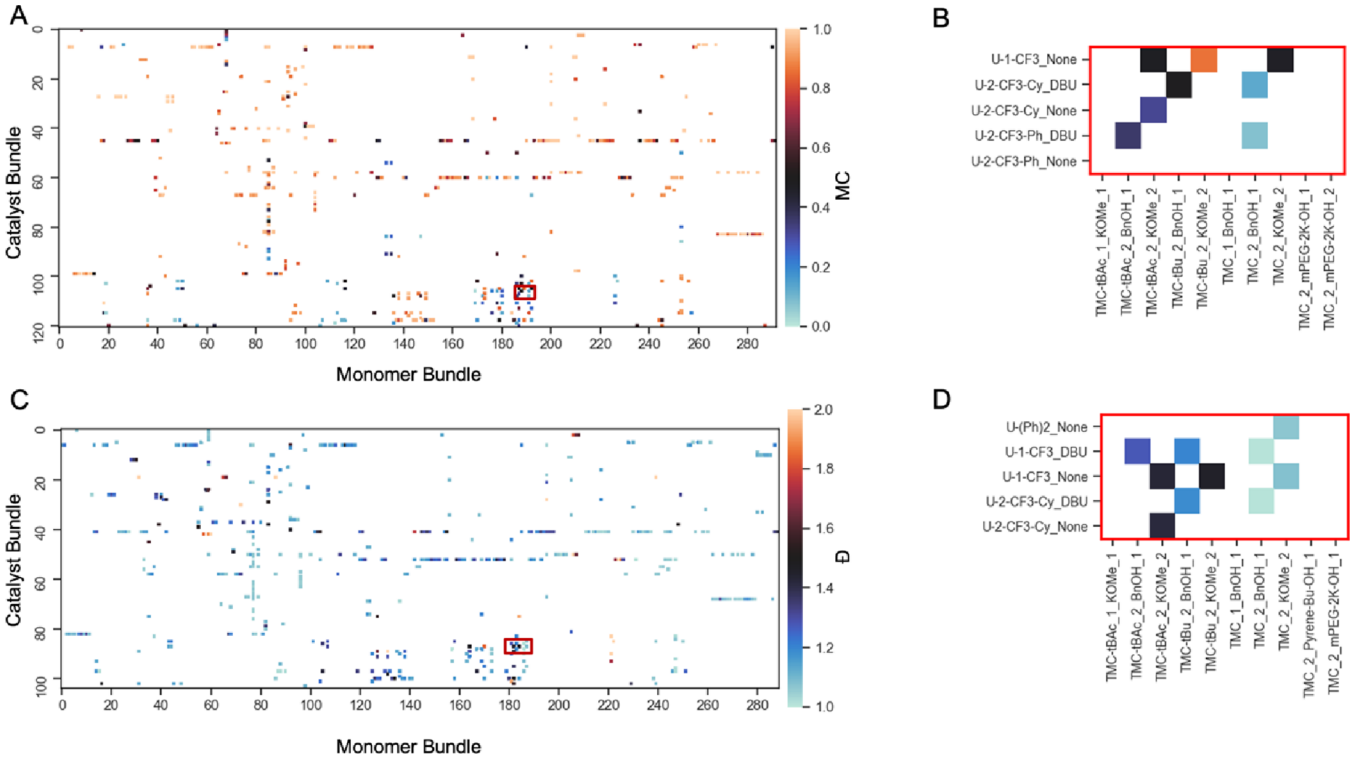

Figure 2. Historical data selected to train the recommender; the target properties are MC and Đ. Panel A: heatmap of MC values. The vertical axis represents the bundled degree of freedom and includes catalyst/cocatalyst combinations designated as "Catalyst Bundle"; horizontal axis represents the bundled degree of freedom that includes monomer/monomer concentration/initiator/monomer−initiator molar ratio combinations designated as "Monomer Bundle". Panel B: a subset of MC data bounded by the red box in panel A. Panel C: heatmap of Đ data. Panel D: a subset of Đ data bounded by the red box in panel C. Abbreviations: U-1-CF_3 = 1-phenyl-3-(3-(trifluoromethyl)phenyl)urea; U-2-CF_3-Cy = 1-(3,5-bis(trifluoromethyl)phenyl)-3-cyclohexylurea; U-2-CF_3-Ph = 1-(3,5-bis(trifluoromethyl)phenyl)-3-phenylurea; U-$(Ph)_2$ = 1,3-diphenylurea; KOMe = potassium methoxide; BnOH = benzyl alcohol; Pyrene-Bu-OH = 1-pyrenebutanol; mPEG-2 K-OH = mono-methoxy-terminated polyethylene glycol, M_n = 2000; TMC = trimethylene carbonate; TMC-tBAc = 2-(*tert*-butoxy)-2-oxoethyl 5-methyl-2-oxo-1,3-dioxane-5-carboxylate; and TMC-tBu = *tert*-butyl 5-methyl-2-oxo-1,3-dioxane-5-carboxylate.

ditions expected to produce polymers with target characteristics. Preparation of the network amenable to link completion can be approached in different ways. We describe an operational approach motivated by the potential utilization of the knowledge graphs (KG) in the polymer materials domain.[47] The first step is to convert a relational database of historical experiments into a resource description framework (RDF) graph (Figure 1A). We refer to this construction as "experiment knowledge graph" (eKG). The prediction of new experiments on eKG is an instance of the network completion problem,[48] where one infers a new "experiment" node connected to a new "material" node. On the eKG, the "experiment" node is linked to the nodes representing particular values of the control parameters and the "material" node is linked to the nodes representing particular values of measurable structural characteristics. The second step is to collapse the nodes on the historical eKG forming a network with multiple partitions (Figure 1A,B). In the case of ROP, we consider three partitions. The nodes in the first partition are bundles of control parameter values related to the catalyst utilization, including the catalyst identity and cocatalyst identity (red outlines, Figure 1A,B). The nodes in the second partition are bundles of control parameter values related to the monomer utilization, including the monomer identity, the initiator identity, the initial monomer concentration ($[M]_0$), and the initial monomer−initiator molar ratio ($[M]_0$:$[I]_0$) (blue outlines, Figure 1A,B). The nodes in the third partition are the values of polymer structural characteristics (gray outlines, Figure 1A,B). Finally, the multipartite graph is simplified by projecting out the partition of property values (Figure 1C). In the resulting bipartite network two nodes are connected via a link if the experiment with the respective parameters was carried out; the outcome of the experiment, i.e., the measured characteristics, are assigned to the link as attributes (Figure 1C). Both the original eKG and the generated bipartite network are suitable inputs for the network completion task. Structuring the data as a bipartite network (Figure 1C) enables network completion via both simple decomposition methods, such as non-negative matrix factorization (NMF)[49] and advanced representation learning approaches, such as *node2vec*[50,51] The presented study reports the results obtained with the continuous feature representation of the nodes in the bipartite network (Figure 1C) obtained using the *node2vec* framework.

Historical data prepared by the subject matter expert (SME) included 810 experiments involving a total of 83 catalysts, 24 cocatalysts, 80 monomers, and 61 initiators. In addition, experiment specification requires the selection of values with at least two continuous degrees of freedom: $[M]_0$ and $[M]_0$:$[I]_0$. To make the recommendation of the values of continuous variables tractable, we discretized the historical ranges of the respective degrees of freedom, producing 21 subranges, or bins, for each. Approaching the combinatorial design of the future experiments by strictly recycling the historical sets of

https://dx.doi.org/10.1021/acs.macromol.0c02127
Macromolecules 2020, 53, 10847−10854

options for the aforementioned degrees of freedom, one would have to navigate a set of 4 billion possible experiments.

We constrained the space of recommendations by considering only those historical experiments that resulted in the successful polymerization and evaluation of the MC (741 experiments). Considering MC to polymer as the target outcome characteristic, the successful historical experiments involved 77 catalysts, 24 cocatalysts, 68 monomers, and 50 initiators. Only 12 subranges of $[M]_0$ and 11 ranges of $[M]_0$:$[I]_0$ were utilized in these experiments. Specifically, 121 unique catalyst/cocatalyst combinations and 292 unique monomer/$[M]_0$/initiator/$[M]_0$:$[I]_0$ combinations were encountered, shaping up a combinatorial space of 35,000 possible experiments (Figure 2A). Successful historical experiments that measured Đ comprised 69 catalysts, 20 cocatalysts, 67 monomers, and 53 initiators. There were 104 unique catalyst/cocatalyst combinations and 289 unique monomer/$[M]_0$/initiator/$[M]_0$:$[I]_0$ combinations with the potential for ~30,000 experiments (Figure 2C).

Recommendations regarding experiments that are expected to produce polymers with target characteristics—Đ and MC—were generated in the following manner. The target range of MC was set at 90−100%; the target range of Đ was set at 1.05−1.15. These ranges were dictated by the practical constraints of the ongoing polymer synthesis efforts and consideration of the balancing positive and negative classes in the recommendation task (cf. Figure 3). Separate bipartite networks were constructed for each characteristic (cf. Figure 1C). The "monomer" and "catalyst" nodes in these networks were connected via links of weight 1.0, if the resulting experiments produced polymers with characteristics in the target ranges. Nodes comprising "unknown" experiments were connected via links of weight 0.5 per imputation strategies outlined in the studies of recommenders (see the "Recommender Methodology" section in the Supporting Information for the discussion). Finally, the nodes were disconnected if the resulting experiments produced polymers with characteristics outside the target ranges. The recommendation workflow comprised the node embedding stage utilizing the *node2vec* representation learning algorithm, the stage computing edge embedding as the Hadamard product of the node embeddings (see ref 50. for the details), and the edge evaluation stage using a random forest classifier (RFC) producing probabilities of the "unknown" edges to have weight 1.0 (see the Supporting Information for technical details and references). A cross-validation study was performed to select the parameter values for the stages of the recommendation workflow and to establish the expectations about its performance (see the Supporting Information). Overall, we observed an average precision (AP) score of 0.65 for MC recommendations and 0.67 for Đ recommendation.

Full specifications of the recommended experiments were reviewed by the SMEs to perform the basic sanity checks and provide the initial level of the assessment of the prediction quality. We attempted to predict the reaction time for each experiment using several standard regressors and treating the combination of binned continuous parameters and one-hot encoded categorical parameters as the features. In order to facilitate SME review and prioritization of the proposed experiments, we introduced the following novelty classes for the proposed experiments. The class of recommendations with the highest novelty comprises pairs of monomers and catalysts that have not been investigated before. Medium-novelty class comprises combinations of monomers, initiators, catalysts, and cocatalysts that have not been attempted. This class comprises performance-tuning hypotheses where new initiators and cocatalysts accompany known monomer−catalyst pairs. Finally, the class with the lowest novelty of recommendations comprises combinations of monomers, initiators, catalysts, and

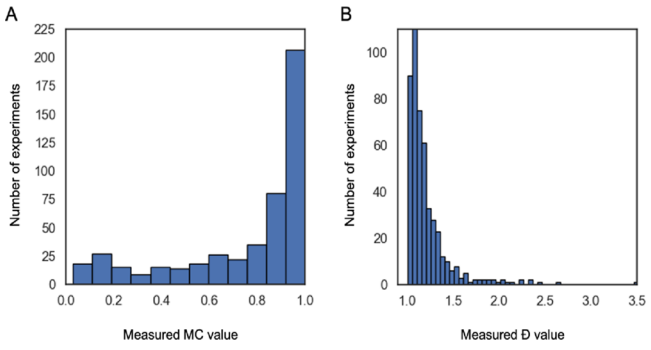

Figure 3. Histograms of the outcomes of historical experiments. Panel A: observed MC values. Panel B: observed Đ values.

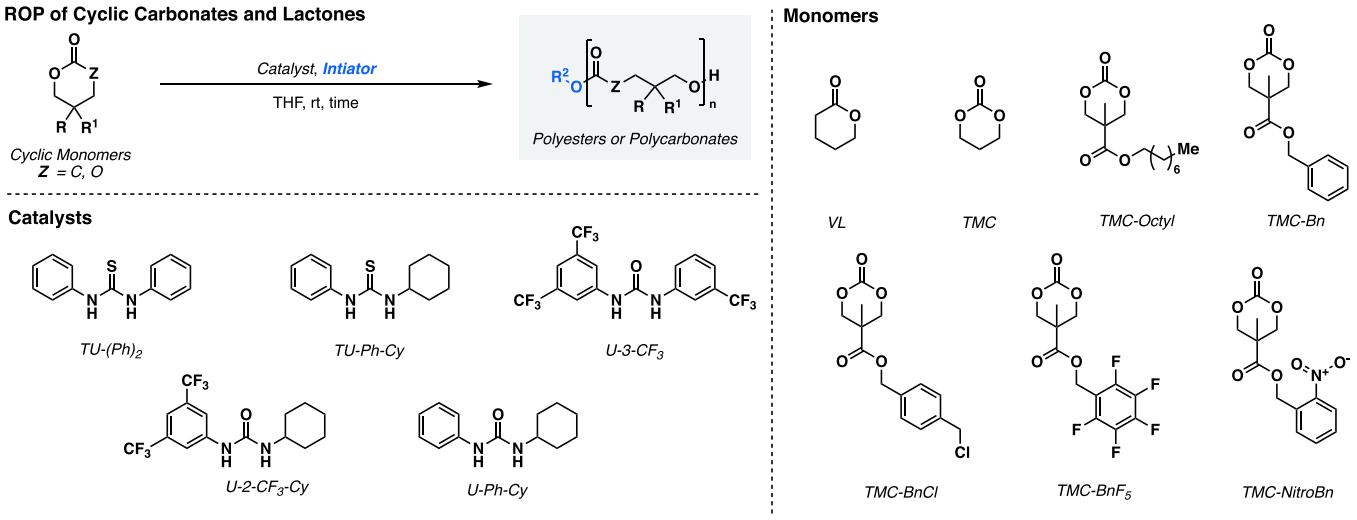

Figure 4. Monomers and catalysts used in validation experiments.

https://dx.doi.org/10.1021/acs.macromol.0c02127
Macromolecules 2020, 53, 10847−10854

Table 1. Experimental Evaluation of Recommendations

entry	novelty	class[a]	initiator	catalyst	cocatalyst	monomer	time (s)	conv. (%)[b]	M_n (GPC)[c]	Đ[c]	MC[d]	Đ[d]
1	high	C(+),Đ(+)	BnOH	U-3-CF_3	KH	TMC-BnF_5	180	96	11,388	1.39	TP	FP
2	***high***	***C(+),Đ(+)***	***BnOH***	***U-3-CF_3***	***KH***	***TMC-BnF_5***	***5***	***91***	***14,288***	***1.07***	***TP***	***TP***
3	medium	C(+),Đ(+)	Pyrene-Bu-OH	U-3-CF_3	KH	VL	186	97	14,629	1.07	TP	TP
4	high	C(+),Đ(+)	BnOH	TU-$(Ph)_2$	DBU	TMC	300	0	—	—	FP	FP
5	medium	C(+),Đ(−)	KOMe	U-Ph-Cy	—	TMC-BnCl	720	97	5340	1.7	TP	TN
6	***medium***	***C(+),Đ(−)***	***KOMe***	***U-Ph-Cy***	—	***TMC-BnCl***	***5***	***96***	***8729***	***2.24***	***TP***	***TN***
7	medium	C(+),Đ(−)	KOMe	U-2-CF_3-Cy	—	VL	15,007	98	14,672	1.77	TP	TN
8	***medium***	***C(+),Đ(−)***	***KOMe***	***U-2-CF_3-Cy***	—	***VL***	***1800***	***98***	***15,791***	***1.62***	***TP***	***TN***
9	medium	C(+),Đ(−)	KOMe	U-2-CF_3-Cy	—	TMC	27,901	99	7835	2.16	TP	TN
10	***medium***	***C(+),Đ(−)***	***KOMe***	***U-2-CF_3-Cy***	—	***TMC***	***600***	***99***	***8696***	***2.01***	***TP***	***TN***
11	high	C(+),Đ(−)	BnOH	TU-Ph-Cy	DBU	TMC-Octyl	225	38	—	—	FP	TN
12	high	C(−),Đ(+)	BnOH	U-3-CF_3	DBU	TMC-Bn	60	95	9276	1.15	FN	TP
13	high	C(−),Đ(−)	BnOH	U-Ph-Cy	DBU	TMC-Octyl	141	20	—	—	TN	TN
14	medium	C(−),Đ(−)	KOMe	U-2-CF_3-Cy	—	TMC-NitroBn	190	87	2311	1.83	FN	TN

[a]Recommender output classes: **C(+)**—monomer conversion >90% after allotted experiment time; **C(−)**—monomer conversion <90% after allotted experiment time; **Đ(+)**—dispersity of the resulting polymer within 1.05−1.15; **Đ(−)**—dispersity of the resulting polymer greater than 1.15. Bold entries are ones that have SME-adjusted reaction times. [b]Determined by ^{1}H NMR spectroscopy of the crude reaction mixture. [c]Determined by gel-permeation chromatography using THF as the eluent and calibration with polystyrene standards. THF = tetrahydrofuran; KOMe = potassium methoxide; KH = potassium hydride; DBU = 1,8-diazabicyclo[5.4.0]undec-7-ene; BnOH = benzyl alcohol; and Pyrene-Bu-OH = 1-pyrene butanol. [d]Evaluation of the predicted class labels: TP = true positive; FP = false positive, TN = true negative, and FN = false negative. See the Supporting Information for details on reaction conditions and characterization data.

cocatalysts that exist in historical data but new quantitative parameters, such as concentrations and ratios. Considerations of the novelty of the proposed experiments were further constrained by the availability and/or accessibility of the required monomers and catalysts. In our case, there were seven monomers and five catalysts available for the immediate recommender-driven action (Table 1).

The recommender produced approximately 16,000 hypotheses above a 0.51 threshold (see the "Recommender Methodology" section in the Supporting Information for the discussion of the threshold selection) comprising monomers, initiators, $[M]_0$ (discretized), $[M]_0$:$[I]_0$ (discretized), catalysts, and cocatalysts that are expected to afford MC within the target range; 199 hypotheses were based on the available/preferred monomers and catalysts only and were considered actionable. Further, 52 actionable hypotheses belonged to the highest novelty class and were based on unique combinations of the monomers and catalysts, and 48 actionable hypotheses in the medium-novelty class were based on unique combinations of the monomers and catalysts. In the case of Đ as the target characteristic, approximately 16,000 experiments were recommended with the threshold set at 0.6 (see the "Recommender Methodology" section in the Supporting Information for the discussion of the threshold selection), including 238 based on the available/preferred components and were considered actionable; 61 actionable hypotheses had the highest novelty and were based on unique combinations of the monomers and catalysts, and 36 actionable hypotheses had medium novelty and were based on unique combinations of the monomers and catalysts. Overall, 10 recommendations were selected by the SME for the experimental follow-up (Figure 4 and Table 1) driven primarily by the availability of the monomers and catalysts and novelty class of the experiment. These experiments included 10 unique combinations of the components; five of which were ranked as having high novelty and five as having medium novelty.

Table 1 summarizes the data obtained for each of the 10 recommendations provided by the recommender engine and four experiments where the SME experimenter adjusted the reaction times from those provided by the linear regression (entries 2, 6, 8, and 10, Table 1). Additional recommended experiment parameters and experimental results are provided in Table S1. The novelty column in Table 1 labels the recommended combination of monomers, catalysts, initiators, and cocatalysts as unique compared to the historical data set used for training. In this regard, only experiments with "medium" or "high" novelty were selected for validation of the model. The class column contains labels for the expected output class of experiment result. As discussed above, the inverse design strategy for materials will attempt to propose experimental conditions that will afford experimental outcomes and polymers within targeted characteristic range, which includes both positive and negative class examples. Since we targeted two commonly measured characteristics for ROP experiments and polymers—MC and Đ—each validation experiment contains a class label for MC (**C**) and Đ (**Đ**). In addition, each class label is appended with a positive (+) sign to indicate that it falls within the prespecified range for MC (>90%) and Đ (1.05−1.15) or a negative (−) sign to indicate that it falls outside of those ranges, giving a total of four possible predicted output classes. Finally, based on the experimental outcome, each entry is assigned a true positive (TP), false positive (FP), true negative (TN), or false negative (FN) label. The TP and FP labels indicate whether or not the experimental results matched the predicted positive class label for MC or Đ (e.g., **C(+)** and **Đ(+)**). TN or FN refers to cases where experimental results did or did not match the predicted label for negative classes (e.g. **C(−)** and **Đ(−)**). Overall, TP and TN labels indicate cases where the recommender engine was successful in assigning the correct class labels for the experimental conditions to afford polyesters and polycarbonates within the targeted positive or negative ranges for the characteristics of MC or Đ.

https://dx.doi.org/10.1021/acs.macromol.0c02127
Macromolecules 2020, 53, 10847−10854

DISCUSSION

The implemented recommender system is intrinsically sensitive to the connectivity of the networks subject to *node2vec* embedding. If only successful historical experiments are considered, the bipartite networks connecting "monomer bundle" nodes and "catalyst bundle" nodes include multiple connected components (Figure S1). This is a form of the "cold start" problem, i.e., lack of the data required to produce a recommendation: if only successful experiments are expressed as links, there are pairs of monomers and catalysts that do not have any paths between them. Our imputation strategy (expressing unknown experiments as links with weight 0.5) enables *node2vec* to carry out random walks between disconnected components and produce a unified embedding. Of course, the obtained embeddings for the nodes that belong to the different connected components lead to recommendations equivalent to random guesses which drag down the performance metrics (AP score of 0.65 for MC recommendations and 0.67 for Đ recommendation). The presence of the multiple connected components in the network of successful experiments is easy to interpret in the context of the historical patterns of data acquisition, such as low tolerance for failing experiments leading to "survivor's bias" and "frozen accidents" in the selection of the catalytic platforms. Exposure of this feature of the historical data topology in the recommender workflow helps to explicitly address the "cold start" problem via the acquisition of the critically missing experimental data.

For each experiment selected for validation, the conditions were tested as given (see Table S1 for additional recommended parameters), meaning no additional fine-tuning of reaction parameters based on the knowledge of SME were performed. In select cases where the predicted reaction time determined by the SME was longer than necessary to provide a material which matched the predicted class labels, a second experiment was performed at a shorter reaction time based on the SME's estimation (entries 2, 5, 7, and 14; Table 1). The validation experimental results are summarized in Table 1 and are highly encouraging. Of the 10 recommended experiments and 40 potential outcomes (positive or negative within each class), only five experiments had outcomes containing either an FP or FN label. Notably, in the four cases where a second experiment was run with the SME-adjusted reaction time was needed (entries 1, 5, 7, and 9, Table 1), the only experiment with an FP label (entry 1, Table 1) was converted into a TP (entry 2, Table 1) via the adjusted reaction time. For all the others (entries 5, 7, and 9, Table 1), the TP and TN outcomes were maintained, indicating the robustness of the predicted classifications. In addition, the experimental results are consistent with the general reactivity trends for the ROP catalysts and monomers utilized, in particular for combinations classified as having high novelty.[24,25] Together, the experimental validation results demonstrate the overall accuracy of the recommendations in predicting experimental conditions to afford polycarbonates and polyesters with the targeted characteristics.

CONCLUSIONS

In this work, we demonstrated that a recommender system constructed from sparse historical experimental data can accurately provide recommendations to match catalysts with monomers for ROP. The process starts with curating a suitable data set as well as selecting the target property or characteristic class and proceeds to infer the parameters of the experiments that are expected to produce polymers with the target property or characteristic. The pattern of our recommendation workflow matches the pattern of the inverse design strategies actively pursued in the context of structure−property relationships (where property−structure is the inverse of structure−property). By exploring the property−experimental parameters paradigm of the discovery, we ensure that the computational predictions are "embarrassingly actionable," as in immediately transferrable into experimental validation. This approach is more streamlined compared to typical protocols that rely on an SME to de novo construct a polymerization experiment using experience and knowledge by matching an ROP catalyst, monomer, and reaction conditions—especially in cases of high novelty. Overall, this work highlights the potential for AI systems, informed by historical data, to reduce experimental workloads by enabling focus toward desired outcome classes. Elaboration of these systems and their combination with automated platforms will significantly enhance and accelerate materials development endeavors.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.macromol.0c02127.

Computational and experimental details, supplementary figures, and NMR spectra (PDF).

AUTHOR INFORMATION

Corresponding Authors

Nathaniel H. Park − *IBM Research–Almaden, San Jose, California 95120, United States;* orcid.org/0000-0002-6564-3387; Email: npark@us.ibm.com

Dmitry Yu. Zubarev − *IBM Research–Almaden, San Jose, California 95120, United States*; Email: dmitry.zubarev@ibm.com

Authors

James L. Hedrick − *IBM Research–Almaden, San Jose, California 95120, United States;* orcid.org/0000-0002-3621-9747

Vivien Kiyek − *IBM Research–Almaden, San Jose, California 95120, United States*

Christiaan Corbet − *IBM Research–Almaden, San Jose, California 95120, United States*

Simon Lottier − *IBM Research–Almaden, San Jose, California 95120, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.macromol.0c02127

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Dr. Tim Erdmann (IBM) and Dr. Pedro Arrechea (IBM) for helpful discussions and assistance in editing this manuscript.

REFERENCES

(1) Correa-Baena, J.-P.; Hippalgaonkar, K.; van Duren, J.; Jaffer, S.; Chandrasekhar, V. R.; Stevanovic, V.; Wadia, C.; Guha, S.; Buonassisi, T. Accelerating Materials Development via Automation, Machine

Learning, and High-Performance Computing. *Joule* **2018**, *2*, 1410−1420.
(2) Coley, C. W.; Eyke, N. S.; Jensen, K. F. *Autonomous Discovery in the Chemical Sciences Part I: Progress*. Angew. Chem., Int. Ed. **2020**, anie.201909987 . DOI: 10.1002/anie.201909987.
(3) Coley, C. W.; Eyke, N. S.; Jensen, K. F. *Autonomous Discovery in the Chemical Sciences Part II: Outlook*. Angew. Chem., Int. Ed. **2020**, anie.201909989 . DOI: 10.1002/anie.201909989.
(4) MacLeod, B. P.; Parlane, F. G. L.; Morrissey, T. D.; Häse, F.; Roch, L. M.; Dettelbach, K. E.; Moreira, R.; Yunker, L. P. E.; Rooney, M. B.; Deeth, J. R.; Lai, V.; Ng, G. J.; Situ, H.; Zhang, R. H.; Elliott, M. S.; Haley, T. H.; Dvorak, D. J.; Aspuru-Guzik, A.; Hein, J. E.; Berlinguette, C. P. Self-Driving Laboratory for Accelerated Discovery of Thin-Film Materials. *Sci. Adv.* **2020**, *6*, eaaz8867.
(5) Häse, F.; Roch, L. M.; Aspuru-Guzik, A. Next-Generation Experimentation with Self-Driving Laboratories. *Trends Chem.* **2019**, *1*, 282−291.
(6) Caramelli, D.; Salley, D.; Henson, A.; Camarasa, G. A.; Sharabi, S.; Keenan, G.; Cronin, L. Networking Chemical Robots for Reaction Multitasking. *Nat. Commun.* **2018**, *9*, 3406.
(7) Dragone, V.; Sans, V.; Henson, A. B.; Granda, J. M.; Cronin, L. An Autonomous Organic Reaction Search Engine for Chemical Reactivity. *Nat. Commun.* **2017**, *8*, 15733.
(8) Bai, Y.; Wilbraham, L.; Slater, B. J.; Zwijnenburg, M. A.; Sprick, R. S.; Cooper, A. I. Accelerated Discovery of Organic Polymer Photocatalysts for Hydrogen Evolution from Water through the Integration of Experiment and Theory. *J. Am. Chem. Soc.* **2019**, *141*, 9063−9071.
(9) Kearsey, R. J.; Alston, B. M.; Briggs, M. E.; Greenaway, R. L.; Cooper, A. I. Accelerated Robotic Discovery of Type II Porous Liquids. *Chem. Sci.* **2019**, *10*, 9454−9465.
(10) Mannodi-Kanakkithodi, A.; Pilania, G.; Huan, T. D.; Lookman, T.; Ramprasad, R. Machine Learning Strategy for Accelerated Design of Polymer Dielectrics. *Sci. Rep.* **2016**, *6*, 20952.
(11) Chen, G.; Shen, Z.; Iyer, A.; Ghumman, U. F.; Tang, S.; Bi, J.; Chen, W.; Li, Y. Machine-Learning-Assisted De Novo Design of Organic Molecules and Polymers: Opportunities and Challenges. *Polymer* **2020**, *12*, 163.
(12) Granda, J. M.; Donina, L.; Dragone, V.; Long, D.-L.; Cronin, L. Controlling an Organic Synthesis Robot with Machine Learning to Search for New Reactivity. *Nature* **2018**, *559*, 377−381.
(13) Lin, B.; Hedrick, J. L.; Park, N. H.; Waymouth, R. M. Programmable High-Throughput Platform for the Rapid and Scalable Synthesis of Polyester and Polycarbonate Libraries. *J. Am. Chem. Soc.* **2019**, *141*, 8921−8927.
(14) Rubens, M.; Vrijsen, J. H.; Laun, J.; Junkers, T. Precise Polymer Synthesis by Autonomous Self-Optimizing Flow Reactors. *Angew. Chem., Int. Ed.* **2019**, *58*, 3183−3187.
(15) Rubens, M.; Junkers, T. Comprehensive Control over Molecular Weight Distributions through Automated Polymerizations. *Polym. Chem.* **2019**, *10*, 6315−6323.
(16) Zhou, Y.; Gu, Y.; Jiang, K.; Chen, M. Droplet-Flow Photopolymerization Aided by Computer: Overcoming the Challenges of Viscosity and Facilitating the Generation of Copolymer Libraries. *Macromolecules* **2019**, *52*, 5611−5617.
(17) Walsh, D. J.; Guironnet, D. Macromolecules with Programmable Shape, Size, and Chemistry. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 1538−1542.
(18) Kumar, J. N.; Li, Q.; Tang, K. Y. T.; Buonassisi, T.; Gonzalez-Oyarce, A. L.; Ye, J. Machine Learning Enables Polymer Cloud-Point Engineering via Inverse Design. *npj Comput. Mater.* **2019**, *5*, 1−6.
(19) Reis, M. H.; Varner, T. P.; Leibfarth, F. A. The Influence of Residence Time Distribution on Continuous-Flow Polymerization. *Macromolecules* **2019**, *52*, 3551−3557.
(20) Olsén, P.; Odelius, K.; Albertsson, A.-C. Thermodynamic Presynthetic Considerations for Ring-Opening Polymerization. *Biomacromolecules* **2016**, *17*, 699−709.
(21) Duda, A.; Kowalski, A. Thermodynamics and Kinetics of Ring-Opening Polymerization. In *Handbook of Ring-Opening Polymerization*; John Wiley & Sons, Ltd, 2009; 1−51. DOI: 10.1002/9783527628407.ch1.
(22) Duda, A.; Kowalski, A.; Libiszowski, J.; Penczek, S. Thermodynamic and Kinetic Polymerizability of Cyclic Esters. *Macromol. Symp.* **2005**, *224*, 71−84.
(23) Olsén, P.; Undin, J.; Odelius, K.; Keul, H.; Albertsson, A.-C. Switching from Controlled Ring-Opening Polymerization (CROP) to Controlled Ring-Closing Depolymerization (CRCDP) by Adjusting the Reaction Parameters That Determine the Ceiling Temperature. *Biomacromolecules* **2016**, *17*, 3995−4002.
(24) Lin, B.; Waymouth, R. M. Urea Anions: Simple, Fast, and Selective Catalysts for Ring-Opening Polymerizations. *J. Am. Chem. Soc.* **2017**, *139*, 1645−1652.
(25) Lin, B.; Waymouth, R. M. Organic Ring-Opening Polymerization Catalysts: Reactivity Control by Balancing Acidity. *Macromolecules* **2018**, *51*, 2932−2938.
(26) Pothupitiya, J. U.; Hewawasam, R. S.; Kiesewetter, M. K. Urea and Thiourea H-Bond Donating Catalysts for Ring-Opening Polymerization: Mechanistic Insights via (Non)Linear Free Energy Relationships. *Macromolecules* **2018**, *51*, 3203−3211.
(27) Spink, S. S.; Kazakov, O. I.; Kiesewetter, E. T.; Kiesewetter, M. K. Rate Accelerated Organocatalytic Ring-Opening Polymerization of l-Lactide via the Application of a Bis(Thiourea) H-Bond Donating Cocatalyst. *Macromolecules* **2015**, *48*, 6127−6131.
(28) Weilandt, K. D.; Keul, H.; Höcker, H. Synthesis and Ring-Opening Polymerization of 2-Acetoxymethyl-2-Alkyltrimethylene Carbonates and of 2-Methoxycarbonyl-2-Methyltrimethylene Carbonate; a Comparison with the Polymerization of 2,2-Dimethyltrimethylene Carbonate. *Macromol. Chem. Phys.* **1996**, *197*, 3851−3868.
(29) Bowden, G. D.; Pichler, B. J.; Maurer, A. A Design of Experiments (DoE) Approach Accelerates the Optimization of Copper-Mediated 18 F-Fluorination Reactions of Arylstannanes. *Sci. Rep.* **2019**, *9*, 11370.
(30) Shear, T. A.; Lin, F.; Zakharov, L. N.; Johnson, D. W. "Design of Experiments" as a Method to Optimize Dynamic Disulfide Assemblies: Cages and Functionalizable Macrocycles. *Angew. Chem., Int. Ed.* **2020**, *59*, 1496−1500.
(31) Zhang, X.; Fevre, M.; Jones, G. O.; Waymouth, R. M. Catalysis as an Enabling Science for Sustainable Polymers. *Chem. Rev.* **2018**, *118*, 839−885.
(32) Kamber, N. E.; Jeong, W.; Waymouth, R. M.; Pratt, R. C.; Lohmeijer, B. G. G.; Hedrick, J. L. Organocatalytic Ring-Opening Polymerization. *Chem. Rev.* **2007**, *107*, 5813−5840.
(33) Kiesewetter, M. K.; Shin, E. J.; Hedrick, J. L.; Waymouth, R. M. Organocatalysis: Opportunities and Challenges for Polymer Synthesis. *Macromolecules* **2010**, *43*, 2093−2107.
(34) Kuhn, C.; Beratan, D. N. Inverse Strategies for Molecular Design. *J. Phys. Chem.* **1996**, *100*, 10595−10599.
(35) Sanchez-Lengeling, B.; Aspuru-Guzik, A. Inverse Molecular Design Using Machine Learning: Generative Models for Matter Engineering. *Science* **2018**, *361*, 360−365.
(36) Elton, D. C.; Boukouvalas, Z.; Fuge, M. D.; Chung, P. W. Deep Learning for Molecular Design—a Review of the State of the Art. *Mol. Syst. Des. Eng.* **2019**, *4*, 828−849.
(37) Brocchini, S.; James, K.; Tangpasuthadol, V.; Kohn, J. A Combinatorial Approach for Polymer Design. *J. Am. Chem. Soc.* **1997**, *119*, 4553−4554.
(38) Kumar, J. N.; Li, Q.; Jun, Y. Challenges and Opportunities of Polymer Design with Machine Learning and High Throughput Experimentation. *MRS Commun.* **2019**, *9*, 537−544.
(39) Paradiso, S. P.; Delaney, K. T.; Fredrickson, G. H. Swarm Intelligence Platform for Multiblock Polymer Inverse Formulation Design. *ACS Macro Lett.* **2016**, *5*, 972−976.
(40) de Almeida, A. F.; Moreira, R.; Rodrigues, T. Synthetic Organic Chemistry Driven by Artificial Intelligence. *Nat. Rev. Chem.* **2019**, *3*, 589−604.
(41) Gómez-Bombarelli, R.; Aguilera-Iparraguirre, J.; Hirzel, T. D.; Duvenaud, D.; Maclaurin, D.; Blood-Forsythe, M. A.; Chae, H. S.; Einzinger, M.; Ha, D.-G.; Wu, T.; Markopoulos, G.; Jeon, S.; Kang,

H.; Miyazaki, H.; Numata, M.; Kim, S.; Huang, W.; Hong, S. I.; Baldo, M.; Adams, R. P.; Aspuru-Guzik, A. Design of Efficient Molecular Organic Light-Emitting Diodes by a High-Throughput Virtual Screening and Experimental Approach. *Nat. Mater.* **2016**, *15*, 1120−1127.
(42) Pyzer-Knapp, E. O.; Suh, C.; Gómez-Bombarelli, R.; Aguilera-Iparraguirre, J.; Aspuru-Guzik, A. What Is High-Throughput Virtual Screening? A Perspective from Organic Materials Discovery. *Annu. Rev. Mater. Res.* **2015**, *45*, 195−216.
(43) Hautier, G.; Jain, A.; Ong, S. P. From the Computer to the Laboratory: Materials Discovery and Design Using First-Principles Calculations. *J. Mater. Sci.* **2012**, *47*, 7317−7340.
(44) Takeda, S.; Hama, T.; Hsu, H.-H.; Yamane, T.; Masuda, K.; Piunova, V. A.; Zubarev, D.; Pitera, J.; Sanders, D. P.; Nakano, D. *AI-Driven Inverse Design System for Organic Molecules.ArXiv200109038 Cs* 2020.
(45) Takeda, S.; Hama, T.; Hsu, H.-H.; Piunova, V. A.; Zubarev, D.; Sanders, D. P.; Pitera, J. W.; Kogoh, M.; Hongo, T.; Cheng, Y.; Bocanett, W.; Nakashika, H.; Fujita, A.; Tsuchiya, Y.; Hino, K.; Yano, K.; Hirose, S.; Toda, H.; Orii, Y.; Nakano, D. *Molecular Inverse-Design Platform for Material Industries*. *ArXiv200411521 Phys* . 2020, DOI: 10.1145/3394486.3403346.
(46) Freeze, J. G.; Kelly, H. R.; Batista, V. S. Search for Catalysts by Inverse Design: Artificial Intelligence, Mountain Climbers, and Alchemists. *Chem. Rev.* **2019**, *119*, 6595−6612.
(47) Zubarev, D. Y.; Pitera, J. W. Cognitive Materials Discovery and Onset of the 5th Discovery Paradigm. In *Machine Learning in Chemistry: Data-Driven Algorithms, Learning Systems, and Predictions*; ACS Symposium Series; American Chemical Society, 2019; *1326*, 103−120. DOI: 10.1021/bk-2019-1326.ch006.
(48) Kim, M.; Leskovec, J.. The Network Completion Problem: Inferring Missing Nodes and Edges in Networks. In *Proceedings of the 2011 SIAM International Conference on Data Mining; Proceedings; Society for Industrial and Applied Mathematics*, 2011; 47−58. DOI: 10.1137/1.9781611972818.5.
(49) Chen, B.; Li, F.; Chen, S.; Hu, R.; Chen, L. Link Prediction Based on Non-Negative Matrix Factorization. *PLoS One* **2017**, *12*, e0182968.
(50) Grover, A.; Leskovec, J. Node2vec: Scalable Feature Learning for Networks. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; KDD '16*; Association for Computing Machinery: San Francisco, California, USA, 2016; 855−864. DOI: 10.1145/2939672.2939754.
(51) Palumbo, E.; Rizzo, G.; Troncy, R.; Baralis, E.; Osella, M.; Ferro, E. Knowledge Graph Embeddings with Node2vec for Item Recommendation. In *The Semantic Web: ESWC 2018 Satellite Events*; Gangemi, A., Gentile, A. L., Nuzzolese, A. G., Rudolph, S., Maleshkova, M., Paulheim, H., Pan, J. Z., Alam, M., Eds.; Lecture Notes in Computer Science ; Springer International Publishing: Cham, 2018; 117−120. DOI: 10.1007/978-3-319-98192-5_22.

https://dx.doi.org/10.1021/acs.macromol.0c02127
Macromolecules 2020, 53, 10847−10854