ACS OMEGA

Cite This: ACS Omega 2020, 5, 83−88

Mini-Review

http://pubs.acs.org/journal/acsodf

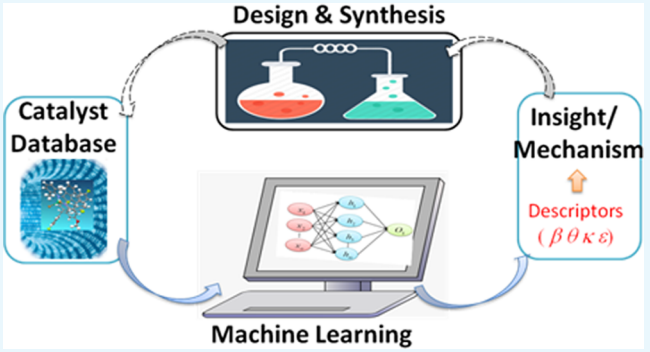
Machine Learning in Catalysis, From Proposal to Practicing

Wenhong Yang,[†,‡] Timothy Tizhe Fidelis,[†,‡] and Wen-Hua Sun*[,†,‡]

[†]Key Laboratory of Engineering Plastics and Beijing National Laboratory for Molecular Science, Institute of Chemistry, Chinese Academy of Sciences, Beijing 100190, China

[‡]CAS Research/Education Center for Excellence in Molecular Sciences and International School, University of Chinese Academy of Sciences, Beijing 100049, China

ABSTRACT: Recently, machine learning (ML) methods have gained popularity and have performed as powerfully predictive tools in various areas of academic and industrious activities. In comparison, their application in catalysis has been underdeveloped. Relying on the rapid development of different algorithms and their implementation, it is the right timing to harvest the potential of ML in catalysis across academy and industry spectra. Herein, we discuss the current applications in the field of homogeneous and heterogeneous catalysis by using various ML approaches. To the best of our knowledge, modern statistical learning techniques will be a strong tool for computational optimization and discovery. This in turn will accurately extract the underlying mechanism in the model that converts readily available data and precatalysts into their promising and useful ones.



INTRODUCTION

The past decade has experienced a rapid growth in the capability of network and computing devices to gather, store, and transport a huge amount of data, which is usually referred to as "Big Data". Regarding data mining and data analysis, artificial intelligence (AI) is penetrating and tends to be remarkable in various fields of science and technologies for developing a practical algorithm and software for computer vision, language processing, image recognition, and others. One of the most powerful strategies within AI is machine learning (ML), which uses algorithms to learn from data, detect patterns, and make fast and accurate prediction.[1] Actually, ML is not a new field of research in chemistry, and several seminal studies mark the initiation of ML which can be tracked back as early as the 1940s, but the peak in ML tools as a practical technology has only been attained in recent years.[2] The increasing attention for the application of ML signifies its vast acceptability and potential in resolving future scientific challenges. Despite the growing trends of ML in various domains, its application in the field of catalysis is still at the early stage of development. Typically, catalysts are designed and synthesized by trial and error with chemical intuition, which conventionally is a time-consuming and high-cost resource. It is found that the automated machine-learning process has been assisting in building better models, understanding the catalytic mechanism, and providing an insight into novel catalytic design.[3,4] This has been facilitated by the development of the latest algorithms and theory, the easy availability of experimental data, as well as low computational cost. Herein, we discuss current applications by using ML in the field of catalysis.

GENERAL METHODS OF MACHINE LEARNING

ML is an interdiscipline field of study combining computer science, statistics, and various subjects in data science. It is concerned with an automated learning process which progresses over time through decision making even under uncertainty. ML approaches can be classified in a variety of ways based on the particular task employed for solving practical problems. One of the wide classifications distinguishes supervised and unsupervised learning. The different definitions between the two methods are dependent on the type of output layer even when used for the training model. The supervised learning method is the most widely used ML method, and much of the practical success in deep learning has evolved from supervised learning methods for discovering a predictive model. According to the algorithm, the ML can be categorized into regression and classification types, including linear regression, artificial neural networks (ANNs), support vector machines (SVMs), random forest, and various forms of decision tree methods.[5] We chose three common methods to demonstrate the different methodologies of ML as shown in Figure 1.

Multiple linear regression analysis (MLRA) is the simplest ML method based on the property (Y) to be modeled by a linear combination of the descriptors (x_i), as shown in Figure 1a. The advantage of linear regression is the ease of explanation and interpretation of the modeled relationship between independent and dependent variables. The artificial neural

Received: October 31, 2019
Accepted: December 10, 2019
Published: December 24, 2019

DOI: 10.1021/acsomega.9b03673
ACS Omega 2020, 5, 83−88

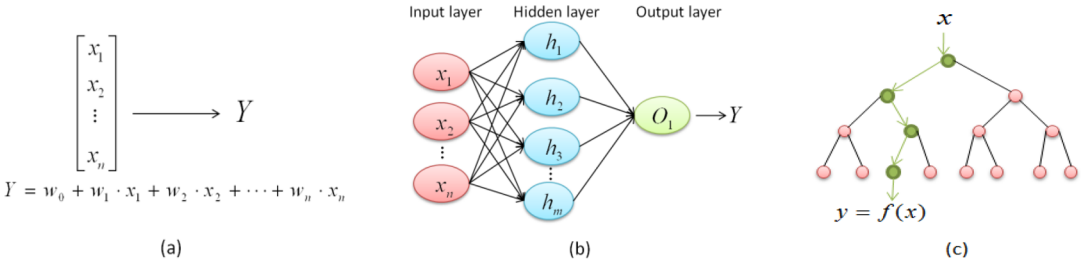
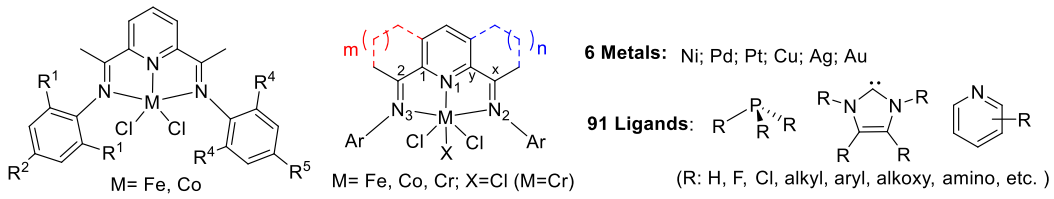


Figure 1. Schematic diagram for common machine learning methods, including (a) linear regression, (b) artificial neural network, and (c) random forest.

Scheme 1. Structures for the Catalysts of (A) Bis(imino)pyridine Fe/Co Complexes, (B) Cycloalkyl-Fused Bis(arylimino)pyridine Metal Complexes, and (C) Suzuki−Miyaura C−C Cross-Coupling in References 11−13, Respectively



network (ANN) as the name implies is a technological design to mimic the working principle of the biological neurons in the human brain. The method is the most popular algorithm of ML and is widely used in both regression and classification analysis. Generally, it includes input, hidden, and output layers as depicted in Figure 1b. Each layer consists of a certain number of neurons, receiving a signal from the input layer, and transmits a response in a forward direction across the different layers, with the treatment of nonlinear activation function such as LOGSIG and TANSIG, and finally to produce a reliable output.

Regarding random forest (RF), the principle is on the basis of constructing a training set and building a predictor tree that maps out observation of a given variable to obtain accurate and stable prediction of the target. It classifies a data point, starting from the top of the tree to different branches down through the several nodes and reaching the terminal leaf, as the decision flow in green color is depicted in Figure 1c. RF is easier to interpret relative to ANN due to its simplicity and clarity. In addition, they give more insights into the relative importance of descriptors on how they influence the target.

In addition, different ML methods have their corresponding advantages and disadvantages, and the selection of an individual method depends on the size and feature of the database. Generally speaking, for a small data set, the linear regression and classification methods can handle such tasks, including Logistic and Naive Bayes methods. On the other hand, for the big data set, the nonlinear method is preferable, such as ANN and K nearest-neighbor (KNN) methods. To begin with the ANN method, huge data usually yield better model performance due to its easy learning approach. Furthermore, there are also some methods that can be used for both linear and nonlinear analysis, such as SVM, which is easier to interpret than ANN, but its disadvantange is sensitivity to the selection of parameters and functions. If the database lost the attribute value and there is no correlation for characteristics, the decision tree may be a better consideration, while it tends toward overfitting compared with that of RF. Regarding unsupervised learning, clustering is the most common technique, and the K-means method is a good choice, which is simple and fast to deal with the big data set to find a hidden pattern in unlabeled data.

Although the early studies by ML were reported more than half a century ago, important applications of ML techniques have been reached just in the past several years in different branches of chemistry science, such as organic and inorganic chemistry.[6−8] As far as the field of catalysis is concerned, ML is making an impact on homogeneous and heterogeneous catalysis research.

APPLICATIONS OF ML IN HOMOGENEOUS CATALYSIS

In terms of homogeneous catalysts, most of the reported studies apply linear regression analysis to establish the quantitative relationship between the structural descriptors and catalytic activity and/or other properties. For example, a multiple linear regression algorithm was applied to predict the catalytic activity of different analogues of bis(arylimino)-pyridine metal complexes.[9] Seven descriptors were extracted from the electronic and steric effects, including the Hammett constant, effective net charge, HOMO−LUMO energy gap, energy difference, open core angle, and bite angle. Different combinations of descriptors were tried to get better fitting results by using as few descriptors as possible. The results revealed that the combination of effective net charge and bite angle shows correlation coefficient values over 0.9 for both Fe and Co analogue systems, and the variation of catalytic activity is mainly determined by the effective net charge.

There have been a small number of reports of quantitative structure−activity/properties relationship (QSA/PR) studies of olefin polymerization using metallocene catalysts by Cruz and co-workers. They investigated the polymerization activity and molecular weight of the resulting polyethylenes for a series of biscyclopentadienyl Zr(IV), bisdimethylsilane-indenyl Zr-(IV), bisindenylHf(IV) metallocene, and bis(imino)pyridine iron complex catalyst systems, respectively.[10] The results illustrate that the catalytic activities are affected by the charge

DOI: 10.1021/acsomega.9b03673
ACS Omega 2020, 5, 83−88

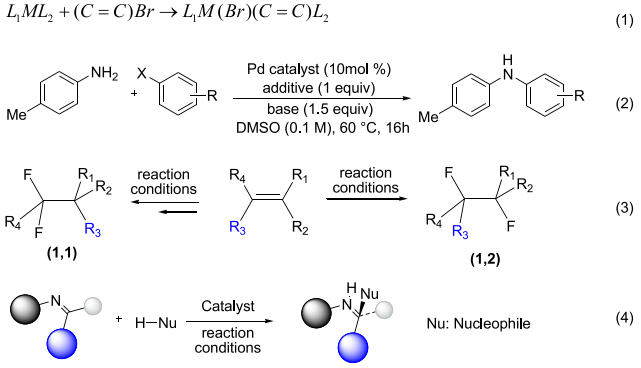
distribution around the aromatic ligands and the steric hindrance around the active site; meanwhile, the steric effect during the reaction of monomer insertion influences the molecular weight of the product. Recently, both 2D- and 3D-QSPR modeling were utilized to analyze various catalytic performances of a data set of 55 bis(imino)pyridine Fe/Co complexes as shown in Scheme 1A, including the catalytic activity, molecular weight, and melting temperature of the product.[11] The results revealed that the descriptors favorable to catalytic activity are unfavorable to the molecular weight of product. This may explain the reason why the high activity hardly exists with high molecular weight, at the same time regarding the catalytic performance of one catalyst.

Furthermore, the catalytic performance of cycloalkyl-fused bis(arylimino)pyridine metal precatalysts toward ethylene polymerization is investigated to explain the effect of different sizes of fused rings.[12] The size of fused member rings is changed from five to seven, as listed in Scheme 1B. It is found that the number of aromatic bonds in the complex accounts for over 50% of catalytic activity, indicating that in addition to the ring strain the degree of conjugation in the complex is the main responsibility for the catalytic activity. This work illustrates the mechanism of the higher catalytic activities by seven-membered fused-ring Fe derivatives compared with that of six-membered fused-ring analogues.

Although the data set for organometallic catalysts (~50) in the latest studies is comparably increased, it is still relatively small. Therefore, all the reports mentioned above were conducted by the linear regression analysis method. The linear relationship between each descriptor and the properties of interest, which characterize the linear fitting model, makes it meritorious for a clear relationship and easy interpretation. However, if more accurate quantitative prediction was necessary, a larger data set for training would be desirable. Unfortunately, homogeneous catalysis studies usually involve small-scale batch experiments. Each individual reaction is an independent experiment and not directly related to others. Particularly for organometallic catalysis, experimental studies on catalytic reactions are time-consuming and require a large amount of resources; therefore, it is a great challenge to build the big data set from the kinetic studies. Furthermore, the catalytic properties are influenced by too many experimental reaction variables.

Cross-coupling reactions have also been the subject of several studies. Lilienfeld and Corminboeuf applied kernel ridge regression machine learning models to predict the energy of the oxidative addition process for organometallic catalysts for Suzuki−Miyaura C−C cross-coupling reactions indicated as reaction 1 in Scheme 2.[13] The huge library of catalysts was comprised of the different combinations of 6 metals and 90 ligands, as indicated in Scheme 1C. The energy values can be used as a descriptor to estimate the activity of catalysts via molecular volcano plots, which were obtained by density functional theory (DFT) calculation instead of experiments. The trained ML models were subsequently exploited to predict the energy-based descriptor of 18 062 potential out-of-sample catalysts with negligible computational cost. More importantly, Dreher and Doyle reported the ML work on 4600 palladium-catalyzed Buchwald−Hartwig cross-coupling reactions as shown in reaction 2, which were obtained via high-throughput experimentation, in order to predict the performance of the synthetic reaction in multidimensional chemical space.[14] The atomic, molecular, and vibrational descriptors were computed

Scheme 2. Catalytic Reactions for Cross-Coupling in References 13 and 14 and Regio-/Enantioselectivities in References 15 and 16, Respectively



and extracted as input variables. By contrast, a random forest algorithm and a single-layer neural network provide significantly improved predictive performance for predicting the reaction yield over linear regression analysis, support vector machines, and *k*-nearest neighbors. By evaluating the relative importance of descriptors, the proposed resource of deleterious side reactivity was verified by further experiments, suggesting its value in facilitating the adoption of synthetic methodology.

How to predict and control various forms of selectivities, such as enantio- and/or regioselectivities, has been a long-term goal in chemical catalysis. For the regioselective difluorination catalytic reaction of alkenes as reaction 3, Sunoj investigated the reaction outcome by using various ML tools.[15] The neural networks accurately predicted the different difluorinated product. The chemical insights were deciphered by a combination of decision tree and the random forest for more rational choices of the reactant alkene for the desired regioisomeric product. Sigman and co-workers presented a holistic, data-driven workflow for deriving statistical models by the linear regression algorithm of one set of enantioselectivity reactions in asymmetric catalysis.[16] The reaction of BINOL-based phosphoric-acid-catalyzed nucleophilic additions to imines was chosen as a general reaction as indicated in reaction 4. The obtained linear fitting models revealed the general interactions that impart asymmetric induction and allow the quantitative transfer of this information to new reaction components.

APPLICATIONS OF ML IN HETEROGENEOUS CATALYSIS

In contrast to homogeneous catalysis, heterogeneous catalysis involves the interaction of a molecule with a substrate, giving heterogeneous catalysis unique features. In heterogeneous catalysis, the data set is generally generated from continuous operation. This provides the chance for the variation of a limited number of parameters and makes it easier to directly generate a huge data set. Therefore, there are more reports in the field of heterogeneous catalysis compared with that of homogeneous catalysis.[17,18] The combination of ML with quantum mechanics (QM) calculations has aided researchers to accelerate the discovery of catalyst candidates in combinatorial big spaces, such as bimetallic alloys. Bimetallic catalysts are good candidates for the most complicated thermal and electrochemical reactions, but modeling the diverse active sites on polycrystalline samples is a topic of discussion.

DOI: 10.1021/acsomega.9b03673
ACS Omega 2020, 5, 83−88

Undoubtedly, high-level quantum-chemical calculations can be used to accurately obtain reactivity descriptors, but the expensive computational cost is its limitation. The ML method, on the contrary, provides an alternate pathway to model the reactivity of catalysts on the basis of the correlations between structural descriptors and reactivity properties.

For example, Xin reported a holistic machine-learning framework for rapid screening of bimetallic catalysts toward methanol electro-oxidation.[19] The adsorption energies of *CO and *OH on {111}-terminated metal surfaces and fingerprint features of active sites were used as the output and input variables, which were calculated from density functional theory calculations. For around 1000 idealized alloy surfaces, the trained ML models show good predictive power in exploring the immense chemical space of bimetallic catalysts. Hundreds of possible active sites exist for every stable low-index facet of a bimetallic crystal. In order to systematically search for the active sites, Nørskov utilized a new approach using ML neural network potentials to model the nickel gallium bimetallic surface for the electrochemical reduction of CO_2, as illustrated in Figure 2.[20] The estimated results of activity by ML was

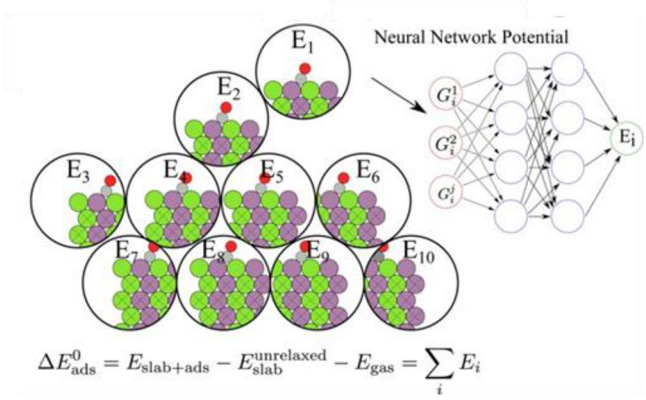


Figure 2. Cartoon of the neural network potential used to directly relax and predict adsorption energies for small molecules. The local region around each atom is used to generate a geometric fingerprint, which is fed through a neural network to provide an atomic contribution to the adsorption energy. The predicted adsorption energy is a summation over these atomic contributions. Reprinted with permission from reference 20. Copyright 2017 American Chemical Society.

present with an order of magnitude fewer explicit DFT calculations. The results show that the most promising active site motifs are isolated nickel atoms with surrounding gallium atoms, rationalized by recent experimental reports of nickel gallium activity.

Furthermore, Jinnouchi used a ML-based Bayesian linear regression method to predict the direct decomposition of NO on the Rh Au alloy nanoparticles (NPs) using a local similarity kernel, which allows interrogation of catalytic activities based on local atomic configurations, as shown in Figure 3. The proposed method can efficiently predict the binding energies of atoms and molecules with the NPs and formation energies of NPs using DFT data on single crystals.[21]

Later, Rappe investigated the catalytic activity of the hydrogen evolution reaction (HER) on $Ni_2P(0001)$ nanoclusters as a function of diverse adsorption site structures.[22] The influence of surface nonmetal doping, as shown in Figure 4A, on the surface structure, charge states, and HER activity,

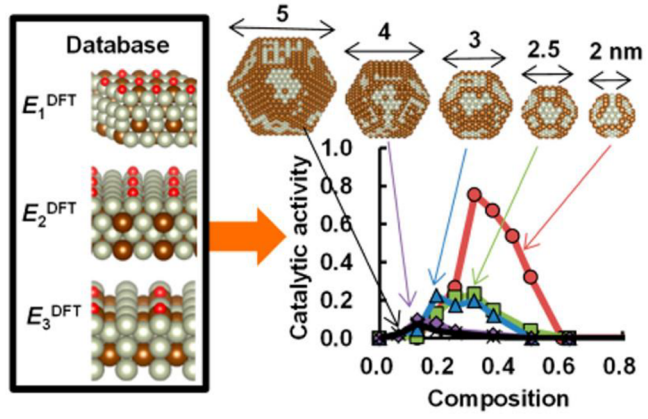


Figure 3. Predicted energetics of catalytic reactions for NO decomposition on nanoparticles using DFT data on single crystals. The kinetic analysis reveals detailed information on structures of active sites and size- and composition-dependent catalytic activities. Reprinted with permission from reference 21. Copyright 2017 American Chemical Society.

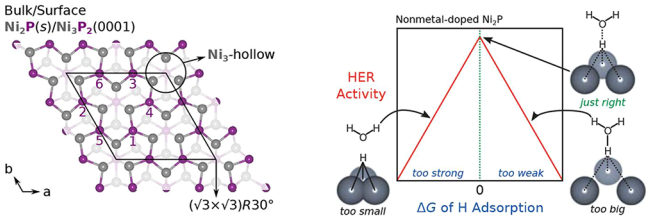


Figure 4. (A) Structure of the $Ni_3P_2(0001)$ surface of Ni_2P showing the $(\sqrt{3} \times \sqrt{3})$ $R30°$ supercell. The Ni_3-hollow sites, which bind H, are shown. The indices on P atoms indicate the preferred sequence of substitution with dopants. (B) The discovered Ni−Ni bond length is the most important descriptor of HER activity, which suggests that the nonmetal dopants induce a chemical pressure-like effect on the Ni_3-hollow site, changing its reactivity through compression and expansion. Reprinted with permission from reference 22. Copyright 2018 American Chemical Society.

was calculated by DFT. The regularized random forest machine learning algorithm was used to discover the relative importance of structure and charge descriptors, indicating the robust descriptor of the Ni−Ni bond length in determining the HER activity of $Ni_2P(0001)$ as in Figure 4B.

Activation energy is an important index for understanding how a chemical reaction would proceed and behave. ML is employed to find the underlying variation trends and regularity within multidimensional space in order to detect the key descriptors for the activation energy. Takahashi estimated the activation energies for a data set of 788 catalytic reactions constructed using first-principle calculations with the implementation of ML. Activation energy can be instantly predicted by chosen descriptors with over 90% accuracy using various methods, such as linear regression, random forest, and support vector machine.[23] Nøskov applied both the linear and nonlinear regression machine method to investigate the reaction barrier of 315 dissociation reactions of an assortment of molecules on a variety of surfaces.[24] The results indicated that the accuracy can be improved up to 2∼3 orders of magnitude by adding up to 7 additional descriptors beyond the traditionally used descriptor reaction energy.

DOI: 10.1021/acsomega.9b03673
ACS Omega 2020, 5, 83−88

In addition to the reaction properties, the reaction conditions play a crucial role in most reaction optimization studies, while rational solvent selection remains a significant challenge in process development. By using the Gaussian process surrogate model ML on a library of 459 solvents for $Rh(CO)_2(acac)$/Josiphos-catalyzed asymmetric hydrogenation, Lapkin demonstrated that it is possible to treat them algorithmically without resorting to expensive high-throughput experimentation.[25] The promising solvents had better reaction outcomes. Additionally, the composition of solvent mixtures and optimal reaction temperature were found using a black-box Bayesian optimization.

CONCLUSIONS

The application of machine learning (ML) has seen a rapid explosion in recent years. Despite its huge success in other fields, ML remains a young field particularly in catalysis with many underexplored research opportunities. This mini-review provides a platform for an integrated ML technique toward design and development in the area of homogeneous and heterogeneous catalysis, which is expected to shake the paradigm of catalysis by lowering the cost associated with the initial discovery. With the progressive increase in number of applications, we anticipate the method to accelerate dramatically in the near future and become the hallmark of a high-level computational tool in the catalytic community.

AUTHOR INFORMATION

Corresponding Author

*E-mail: whsun@iccas.ac.cn.

ORCID

Wenhong Yang: 0000-0003-2269-2987

Wen-Hua Sun: 0000-0002-6614-9284

Notes

The authors declare no competing financial interest.

Biographies

Wenhong Yang received her Doctoral degree in computational chemistry from the Institute of Chemistry, Chinese Academy of Sciences (ICCAS), in Beijing in 2009. After her postdoctoral research, she was appointed as assistant professor and promoted to associate professor at ICCAS in 2012. During the period from April 2017 to July 2018, she worked as a visiting professor at the Department of Chemistry, University of Tokyo. Her current research mainly focuses on the quantitative structure−property relationship (QSPR) modeling on the transition metal complex catalyst towards ethylene polymerization.

Timothy Tizhe Fidelis received his Bachelor degree in Chemistry from Modibbo Adama University of Technology in Nigeria in 2014. He is currently a graduate student in Prof. Wen-Hua Sun's group under close supervision of Assoc. Prof. Wenhong Yang in the Institute of Chemistry, Chinese Academy of Sciences (ICCAS). His current research interests focus on modeling of catalytic properties of transition metal complexes toward ethylene polymerization by machine learning.

Prof. Dr. Wen-Hua Sun heads a group of organometallic chemistry and catalysis at the Institute of Chemistry, Chinese Academy of Sciences, since October of 1999. Currently, he has also been a chair professor in the University of Chinese Academy of Sciences since 2014. He was selected as Fellow of the Royal Society of Chemistry (2011) and member of the European Academy of Sciences (2017). His contributions have had international impacts across the areas of polymerization catalysts for conversional α-olefins and polyolefins, biodegradable polymers, and biomassive processes as well as their intermediates and reaction mechanism. He received his B.Sc. in chemistry at Lanzhou University (1986) and his M.S./Ph.D. degrees in physical chemistry at Lanzhou Institute of Chemical Physics (LICP, 1989/1994). He worked in LICP as a Research Associate (1989) and Associate Professor (1993) and at Hokkaido University with fellowships from the Japan Society for the Promotion of Science (1995), Center of Excellence (1997), and Japan Science and Technology Corporation (1998). His group has been opened to domestic and international colleagues, and those collaborations have been visually recognized according to his publications indicating coauthor members from different affiliations. His major subjects include: polyolefin-oriented organometallics; developing polyolefin process; new donors for Ziegler−Natta catalysts; coupling reaction; fluorescent properties of metal complexes; pilot process of ethylene oligo/polymerization; as well as industrial processes

DOI: 10.1021/acsomega.9b03673
ACS Omega 2020, 5, 83−88

■ ACKNOWLEDGMENTS

This work was financially supported by the Innovated Cultivation Project of ICCAS (CXPY-19) and the National Natural Science Foundation of China (No. 21871275).

■ REFERENCES

(1) Jordan, M. I.; Mitchell, T. M. Machine Learning: Trends, Perspectives, and Prospects. *Science* **2015**, *349*, 255−260.
(2) 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.
(3) Kitchin, J. R. Machine Learning in Catalysis. *Nat. Catal.* **2018**, *1*, 230−232.
(4) Li, Z.; Wang, S.; Xin, H. Toward Artificial Intelligence in Catalysis. *Nat. Catal.* **2018**, *1*, 641−642.
(5) Clarke, B.; Fokoue, E.; Zhang, H. H. *Principles and Theory for Data Mining and Machine Learning*; Springer Science&Business Media: 2009.
(6) Peiretti, F.; Brunel, J. M. Artificial Intelligence: The Future for Organic Chemistry? *ACS Omega* **2018**, *3*, 13263−13266.
(7) Janet, J. P.; Liu, F.; Nandy, A.; Duan, C.; Yang, T.; Lin, S.; Kulik, H. J. Designing in the Face of Uncertainty: Exploiting Electronic Structure and Machine Learning Models for Discovery in Inorganic Chemistry. *Inorg. Chem.* **2019**, *58*, 10592−10606.
(8) Mater, A. C.; Coote, M. L. Deep Learning in Chemistry. *J. Chem. Inf. Model.* **2019**, *59*, 2545−2559.
(9) Ahmed, S.; Yang, W.; Ma, Z.; Sun, W.-H. Catalytic Activities of Bis(pentamethylene)pyridyl (Fe/Co) Complex Analogues in Ethylene Polymerization by Modeling Method. *J. Phys. Chem. A* **2018**, *122*, 9637−9644.
(10) Cruz, V. L.; Martinez, S.; Ramos, J.; Martinez-Salazar, J. 3D-QSAR as a Tool for Understanding and Improving Single Site Polymerization Catalysts: A Review. *Organometallics* **2014**, *33*, 2944−2959.
(11) Yang, W.; Ma, Z.; Yi, J.; Ahmed, S.; Sun, W.-H. Catalytic Performance of Bis(imino)pyridylmetal Precatalysts in Ethylene Polymerization by 2D-/3D-QSPR Modeling. *J. Comput. Chem.* **2019**, *40*, 1374−1386.
(12) Yang, W.; Ahmed, S.; Fidelis, T. T.; Sun, W.-H. Effect of Cycloalkyl-fused Ring on the Catalytic Performance of Bis(imino)-pyridine Metal Complexes by QSPR Modeling. *Catal. Commun.* **2019**, *132*, 105820−105824.
(13) Meyer, B.; Sawatlon, B.; Heinen, S.; Von Lilienfeld, O. A.; Corminboeuf, C. Machine Learning Meets Volcano Plots: Computational Discovery of Cross-Coupling Catalysts. *Chem. Sci.* **2018**, *9*, 7069−7077.
(14) Ahneman, D. T.; Estrada, J. G.; Lin, S.; Dreher, S. D.; Doyle, A. G. Predicting Reaction Performance in C−N Cross-coupling using Machine Learning. *Science* **2018**, *360*, 186−190.
(15) Banerjee, S.; Sreenithya, A.; Sunoj, R. B. Machine Learning for Predicting Product Distributions in Catalytic Regioselective Reactions. *Phys. Chem. Chem. Phys.* **2018**, *20*, 18311−18318.
(16) Reid, J. P.; Sigman, M. S. Holistic Prediction of Enantioselectivity in Asymmetric Catalysis. *Nature* **2019**, *571*, 343−348.
(17) Grajciar, L.; Heard, C. J.; Bondarenko, A. A.; Polynski, M. V.; Meeprasert, J.; Pidko, E. A.; Nachtigall, P. Towards Operando Computational Modeling in Heterogeneous Catalysis. *Chem. Soc. Rev.* **2018**, *47*, 8307−8348.
(18) Goldsmith, B. R.; Esterhuizen, J.; Liu, J. X.; Bartel, C. J.; Sutton, C. Machine Learning for Heterogeneous Catalyst Design and Discovery. *AIChE J.* **2018**, *64*, 2311−2323.
(19) Li, Z.; Wang, S.; Chin, W. S.; Achenie, L. E.; Xin, H. High-Throughput Screening of Bimetallic Catalysts Enabled by Machine Learning. *J. Mater. Chem. A* **2017**, *5*, 24131−24138.
(20) Ulissi, Z. W.; Tang, M. T.; Xiao, J.; Liu, X.; Torelli, D. A.; Karamad, M.; Cummins, K.; Hahn, C.; Lewis, N. S.; Jaramillo, T. F.; Chan, K.; Nørskov, J. K. Machine-Learning Methods Enable Exhaustive Searches for Active Bimetallic Facets and Reveal Active Site Motifs for CO_2 Reduction. *ACS Catal.* **2017**, *7*, 6600−6608.
(21) Jinnouchi, R.; Asahi, R. Predicting Catalytic Activity of Nanoparticles by a DFT-Aided Machine-Learning Algorithm. *J. Phys. Chem. Lett.* **2017**, *8*, 4279−4283.
(22) Wexler, R. B.; Martirez, J. M. P.; Rappe, A. M. Chemical Pressure-Driven Enhancement of the Hydrogen Evolving Activity of Ni2P from Nonmetal Surface Doping Interpreted via Machine Learning. *J. Am. Chem. Soc.* **2018**, *140*, 4678−4683.
(23) Takahashi, K.; Miyazato, I. Rapid Estimation of Activation Energy in Heterogeneous Catalytic Reactions via Machine Learning. *J. Comput. Chem.* **2018**, *39*, 2405−2408.
(24) Singh, A. R.; Rohr, B. A.; Gauthier, J. A.; Nørskov, J. K. Predicting Chemical Reaction Barriers with a Machine Learning Model. *Catal. Lett.* **2019**, *149*, 2347−2354.
(25) Amar, Y.; Schweidtmann, A. M.; Deutsch, P.; Cao, L.; Lapkin, A. Machine Learning and Molecular Descriptors Enalbe Rational Sovent Selection in Asymmetric Catalysis. *Chem. Sci.* **2019**, *10*, 6697−6706.

DOI: 10.1021/acsomega.9b03673
ACS Omega 2020, 5, 83−88