

Applications of Data Science to Molecules and Materials 2022

Topical Conference at the 2022 AIChE Annual Meeting

Phoenix, Arizona, USA
13-18 November 2022

ISBN: 978-1-7138-7918-3

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571



Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2022) by AIChE
All rights reserved.

Printed with permission by Curran Associates, Inc. (2023)

For permission requests, please contact AIChE
at the address below.

AIChE
120 Wall Street, FL 23
New York, NY 10005-4020

Phone: (800) 242-4363
Fax: (203) 775-5177

www.aiche.org

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2633
Email: curran@proceedings.com
Web: www.proceedings.com

TABLE OF CONTENTS

APPLICATIONS OF DATA SCIENCE IN CATALYSIS AND REACTION ENGINEERING

283a Predicting the Performance of Mesophase Formation and Properties of Mesophase Pitch Based on Experimental Investigation and Machine Learning	1
<i>Wenjia Wang, Madison Cooley, Karissa Jolley, Robert M. Kirby, Eric Eddings</i>	
283c Bayesian Reaction Optimization of Rac-Lactide Polymerization Catalyzed By Aluminum Complexes	2
<i>Yang Huang, Maverick Lin, Rong Tong, Hongliang Xin</i>	
283d Inferring Structure-Entropy Correlations of Zeolite-Adsorbate Interactions Using Monte-Carlo Simulations and Machine Learning	3
<i>Christopher Rzepa, Srinivas Rangarajan</i>	

APPLICATIONS OF DATA SCIENCE IN MOLECULAR SCIENCES I

152a Combining functional group and Graph Neural Networks Towards Interpretable Molecular Property Models	4
<i>Adem Rosenkvist Nielsen Aouichaoui, Fan Fan, Seyed Soheil Mansouri, Jens Abildskov, Gurkan Sin</i>	
152b Evaluation of Sampling Algorithms to Explore the Latent Space Created By Deep Generative Models	6
<i>Fangxi Wang, Parisa Farzeen, Charles Plate, Lakshmikumar Kunche, Sanket Deshmukh</i>	
152c Automatic Creation of Molecular Substructure Descriptors for Estimation of Pure Compound Properties	7
<i>Qiong Pan, Jie Li, Xiaolei Fan</i>	
152d Data-Driven Chemical Property Models for Energetic Materials Using Transfer Learning	10
<i>Joshua Lansford, Brian C. Barnes, Betsy Rice, Klavs Jensen</i>	
152e Machine-Learning a Solution for Reactive Simulations of Complex Chemical Systems	11
<i>Rebecca Lindsey, Laurence E. Fried, Nir Goldman, C. Huy Pham, Sorin Bastea</i>	
152f Capturing Molecular Interactions in Graph Neural Networks: A Case Study in Multi-Component Phase Equilibrium	12
<i>Shiyi Qin, Shengli Jiang, Prasanna Balaprakash, Reid Van Lehn, Victor Zavala</i>	
152g A Statistical Analysis of Computational Hydrogen Bond Features in Hydrophobic Non-Ion Deep Eutectic Solvents and Non-Deep Eutectic Solvents	14
<i>Usman Abbas, Joseph Tapia, Yuxuan Zhang, Mohammad Selim, Jian Shi, Jin Chen, Qing Shao</i>	
152h Predicting Temperature-Dependent Activity Coefficients of Ionic Liquid-Solute Systems through Graph-Based Machine Learning	15
<i>Jan G. Rittig, Karim Ben Hicham, Artur M. Schweidtmann, Manuel Dahmen, Alexander Mitsos</i>	

152i Property Prediction of Amine-Functionalized Ionic Liquids for Multi-Scale Carbon Capture Design and Optimization	17
<i>Austin Keller, Pratik Kelkar, Michael Baldea, Mark Stadtherr, Joan Brennecke</i>	
152j Machine-Learning Enabled Screening of MOFs for Ion Selective Membranes.....	19
<i>Shuwen Yue, Aditya Nandy, Heather Kulik</i>	

APPLICATIONS OF DATA SCIENCE TO HIGH THROUGHPUT EXPERIMENTATION

88a A Bayesian Experimental Design Framework for Optimizing Microbial Communities.....	20
<i>Jaron Thompson, Ophelia Venturelli, Victor Zavala</i>	
88b Solution Coating of Polymer Blend Libraries for High-Throughput Experimentation Via Passive Mixing	22
<i>Aaron L. Liu, Ezgi Melis Dogan-Guner, Michael McBride, Rahul Venkatesh, Miguel A. Gonzalez, Jun Amano, Elsa Reichmanis, Martha Grover, J Carson Meredith</i>	
88d Design of an Automatic Platform for Machine-Learning Model Based Molecular Property Optimization.....	23
<i>Matthew McDonald, Brent Koscher, Richard Canty, Seung Kyun Ha, Camille Bilodeau, Kevin P. Greenman, Charles J. McGill, Rafael Gomez-Bombarelli, William Green, Klavs Jensen</i>	
88e Simulating Closed-Loop Catalyst Discovery Processes Using an Experimental Band Gap Surrogate Model	24
<i>Kirby Broderick, Zachary Ulissi</i>	
88f Improved Characterization of Membrane Transport Properties through Advanced Data Analytics.....	25
<i>Xinhong Liu, Jonathan Ouimet, Laurianne Lair, William Phillip, Alexander Dowling</i>	
88g Autonomous Optimization of Robotic Process Parameters for Accurate Experiments	27
<i>Martin Seifrid</i>	
88h High-Throughput Screening of Binary and Ternary Organic Redox Active Materials-Based Deep Eutectic Solvents.....	28
<i>Maria Politi, Jaime Rodriguez Jr., Stuart Adler, David Beck, Lilo Pozzo</i>	

INNOVATIONS IN METHODS OF DATA SCIENCE

416a Global Systems Analysis Using Deep Learning on Industrially Relevant Large Datasets	29
<i>Gurkan Sin, Adem Rosenkvist Nielsen Aouichaoui</i>	
416b Uncertainty Quantification for Molecular Property Predictions Using Automatic Graph Neural Architecture Search.....	31
<i>Shengli Jiang, Shiyi Qin, Prasanna Balaprakash, Reid Van Lehn, Victor Zavala</i>	
416c Projecting the Effectiveness of Deep Ensembles.....	34
<i>Charles McGill, William Green</i>	
416d Automatic Strain Sensor Design Via Active Learning and Data Argumentation for Soft Machines	35
<i>Po-Yen Chen, Tianle Chen</i>	
416e A Framework for the Statistical Value of Extreme Points for Function Approximation	37
<i>Arijit Chakraborty, Krishnan Kumaran, Venkat Venkatasubramanian</i>	

416f Graph Hysteria – Comparing the Generative Performance of Graph and String-Based Translation Vaes for Molecular Design	38
<i>Orion Dollar, Nisarg Joshi, David Beck, Jim Pfaendtner</i>	
416g Extending Bigsmiles to Non-Covalent Bonds in Supramolecular Polymer Assemblies	39
<i>Weizhong Zou, Alexis Martell-Monterroza, Yunxin Yao, S. Cem Milik, Morgan Cencer, Nathan Rebello, Haley Beech, Melody Morris, Tzyy-Shyang Lin, Cleotilde Castano, Julia Kalow, Stephen Craig, Alshakim Nelson, Jeffrey Moore, Bradley Olsen</i>	
416h Formal Proofs in the Chemical Sciences Using the Lean Theorem Prover	40
<i>Parivash Feyzishendi, Maxwell P. Bobbin, Catherine M. Wraback, Samiha Sharlin, Tyler R. Josephson</i>	
416i Machine Learning with Differential Equation Priors	41
<i>Aditi Krishnapriyan</i>	

APPLICATIONS OF DATA SCIENCE IN MOLECULAR SCIENCES II

218a Discovery of O ₂ -Selective Metal-Organic Frameworks Via Bayesian Optimization	42
<i>Eric Taw, Jeffrey B. Neaton</i>	
218c Cript: A Scalable Polymer Material Data Structure	43
<i>Dylan Walsh, Bradley Olsen, Klavs Jensen</i>	
218d Data-Driven Design of Polymer-Protein Hybrid Materials	44
<i>Michael Webb</i>	
218e Applications of Machine-Learned Electron Densities of Nucleic Acids.....	45
<i>Alex Lee, Joshua A. Rackers, William P. Bricker</i>	
218f Using Machine Learning Approaches to Estimate Enzyme Kinetic Parameters	46
<i>Veda Sheersh Boorla, Costas D. Maranas</i>	
218g Classifying the Toxicity of Pesticides to Honey Bees Via Support Vector Machines with Random Walk Graph Kernels.....	47
<i>Ping Yang, Adrian Henle, Xiaoli Fern, Cory Simon</i>	
218h Hallucinating Inexpensive, Diverse and Native-like Antibody Binders with Deep Learning	48
<i>Sai Pooja Mahajan, Jeffrey Ruffolo, Rahel Frick, Jeffrey J. Gray</i>	
218i Identifying Novel Fentanyl Analogs from Mass Spectral Measurements	50
<i>Arun S. Moorthy, Anthony J. Kearsley, William E. Wallace III</i>	
218j Developing a Defeatured Atom-Additive Model to Predict Single Component Partition Coefficients with FT-ICR MS Data	51
<i>David Kenney, Heather LeClerc, Randy Paffenroth, Michael T. Timko, Andrew R Teixeira</i>	

Author Index