Electronic Supplementary Material (ESI) for ChemComm.

Supplementary Information for Modern Machine Learning for Tackling Inverse Problems in Chemistry: Molecular Design to Realization

Bhuvanesh Sridharan[†], Manan Goel[†] and U Deva Priyakumar[†*]

[†] *Center for Computational Natural Sciences and Bioinformatics, International Institute of Information Technology, Hyderabad 500032, India*

∗ Email: deva@iiit.ac.in

Table 1: Representative list of publications that have proposed methodologies related to molecular generation

Title	Methods	Authors
Automatic chemical design using a data-driven continuous representation of molecules	VAE with RNN Encoder and Decoder	Gómez-Bombarelli *et al.*[1]
Junction tree variational autoencoder for molecular graph generation	VAE with Graph Encoder and Decoder	Jin *et al.*[2]
GraphVAE: Towards Generation of Small Graphs Using Variational Autoencoders	VAE with Graph Encoder and Decoder	Simonovsky and Komodakis[3]
Constrained graph variational autoencoders for molecule design	VAE with Gated Graph NN Encoder and Decoder	Liu *et al.*[4]
Molecular generative model based on conditional variational autoencoder for de novo molecular design	VAE with RNN Encoder and Decoder	Lim *et al.*[5]
Grammar variational autoencoder	VAE with RNN Encoder and Decoder	Kusner *et al.*[6]
Learning continuous and data-driven molecular descriptors by translating equivalent chemical representations	VAE with RNN/CNN Encoder and RNN Decoder	Winter *et al.*[7]
Constrained Bayesian optimization for automatic chemical design using variational autoencoders	VAE with RNN Encoder and Decoder with Bayesian Optimization	Griffiths and Hernández-Lobato[8]
Efficient multi-objective molecular optimization in a continuous latent space	VAE with RNN Encoder and Decoder with Particle Swarm Optimization	Winter *et al.*[9]
druGAN: an advanced generative adversarial autoencoder model for de novo generation of new molecules with desired molecular properties in silico	Autoencoder with adverserial training	Kadurin *et al.*[10]
Application of generative autoencoder in de novo molecular design	Autoencoder with adverserial training followed by Bayesian Optimization	Blaschke *et al.*[11]
Deep reinforcement learning for de novo drug design	RNN with Reinforcement Learning	Popova *et al.*[12]
MoleGuLAR: Molecule Generation using Reinforcement Learning with Alternating Reward	RNN with Reinforcement Learning	Goel *et al.*[13]
Objective-reinforced generative adversarial networks (ORGAN) for sequence generation models	GAN with Reinforcement Learning	Guimaraes *et al.*[14]
Optimizing distributions over molecular space. An objective-reinforced generative adversarial network for inverse-design chemistry (ORGANIC)	GAN with Reinforcement Learning	Sanchez-Lengeling *et al.*[15]
Reinforced adversarial neural computer for de novo molecular design	GAN with Reinforcement Learning	Putin *et al.*[16]
Adversarial threshold neural computer for molecular de novo design	GAN with Reinforcement Learning	Putin *et al.*[17]
MolGAN: An implicit generative model for small molecular graphs	WGAN with Reinforcement Learning	De Cao and Kipf[18]
Mol-CycleGAN: a generative model for molecular optimization	GAN with Reinforcement Learning	Maziarka *et al.*[19]
A de novo molecular generation method using latent vector based generative adversarial network	Combination of VAE and GAN	Prykhodko *et al.*[20]
Graph convolutional policy network for goal-directed molecular graph generation	Graph NN with Reinforcement Learning	You *et al.*[21]
DeepGraphMolGen, a multi-objective, computational strategy for generating molecules with desirable properties: a graph convolution and reinforcement learning approach	Graph NN with Reinforcement Learning	Khemchandani *et al.*[22]
Optimization of molecules via deep reinforcement learning	Graph NN with Reinforcement Learning	Zhou *et al.*[23]

Table 2: Representative list of publications related to application of modern ML methods for mapping retrosynthesis pathways

Title	Methods	Authors
Efficient syntheses of diverse, medicinally relevant targets planned by computer and executed in the laboratory	Only Reaction Templates	Klucznik *et al.*[24]
Prediction and interpretable visualization of retrosynthetic reactions using graph convolutional networks	GCN with Reaction Templates	Ishida *et al.*[25]
Deep retrosynthetic reaction prediction using local reactivity and global attention	GLN with Reaction Templates	Chen and Jung[26]
Planning chemical syntheses with deep neural networks and symbolic AI	MCTS with Reaction Templates	Segler *et al.*[27]
Learning retrosynthetic planning through simulated experience	MCTS with Reaction Templates	Schreck *et al.*[28]
AiZynthFinder: a fast, robust and flexible open-source software for retrosynthetic planning	MCTS with Reaaction Templates	Genheden *et al.*[29]
Retro*: Learning Retrosynthetic Planning with Neural Guided A* Search	Best First Search with Reaction Templates	Chen *et al.*[30]
Retrosynthetic reaction prediction using neural sequence-to-sequence model	Template Free RNN	Liu *et al.*[31]
A transformer model for retrosynthesis	Template Free Transformer	Karpov *et al.*[32]
Predicting retrosynthetic reactions using self-corrected transformer neural networks	Template Free Transformer	Zheng *et al.*[33]
Valid, Plausible, and Diverse Retrosynthesis Using Tied Two-Way Transformers with Latent Variables	Template Free Transformer	Kim *et al.*[34]
Molecular graph enhanced transformer for retrosynthesis prediction	Template Free GNN and Transformer	Mao *et al.*[35]
GTA: Graph Truncated Attention for Retrosynthesis	Template Free GNN with Attention	Seo *et al.*[36]
Automatic retrosynthetic route planning using template-free models	Template Free Transformer with MCTS	Lin *et al.*[37]
Predicting retrosynthetic pathways using transformer-based models and a hyper-graph exploration strategy	Template Free Transformer with Beam Search	Schwaller *et al.*[38]
A graph to graphs framework for retrosynthesis prediction	Template Free Graph Convolutional Network	Shi *et al.*[39]
Learning graph models for template-free retrosynthesis	Template Free Message Passing Network	Somnath *et al.*[40]
Retroxpert: Decompose retrosynthesis prediction like a chemist	Semi-Template Based Graph Attention Network	Yan *et al.*[41]

Table 3: ML based studies that attempted to decipher spectra to molecule inverse problem

Title	Methods	Authors
Chiral Molecular Structure Determination for a Desired Compound Just from Its Molecular Formula and Vibrational Optical Activity Spectra	SVM	Wang *et al.*[42]
Spectral deep learning for prediction and prospective validation of functional groups	MLP	Fine *et al.*[43]
NMR-TS: de novo molecule identification from NMR spectra	Tree Search, DFT Calculations	Zhang *et al.*[44]
Deep imitation learning for molecular inverse problems	MLP, GCN	Jonas[45]
Spectra To Structure : Deep Reinforcement Learning for Molecular Inverse Problem	MCTS, GCN	Sridharan *et al.*[46]

References

[1] R. Gómez-Bombarelli, J. N. Wei, D. Duvenaud, J. M. Hernández-Lobato, B. Sánchez-Lengeling, D. Sheberla, J. Aguilera-Iparraguirre, T. D. Hirzel, R. P. Adams and A. Aspuru-Guzik, *ACS central science*, 2018, **4**, 268–276.

[2] W. Jin, R. Barzilay and T. Jaakkola, International conference on machine learning, 2018, pp. 2323–2332.

[3] M. Simonovsky and N. Komodakis, International conference on artificial neural networks, 2018, pp. 412–422.

[4] Q. Liu, M. Allamanis, M. Brockschmidt and A. L. Gaunt, *arXiv preprint arXiv:1805.09076*, 2018.

[5] J. Lim, S. Ryu, J. W. Kim and W. Y. Kim, *Journal of cheminformatics*, 2018, **10**, 1–9.

[6] M. J. Kusner, B. Paige and J. M. Hernández-Lobato, International Conference on Machine Learning, 2017, pp. 1945–1954.

[7] R. Winter, F. Montanari, F. Noé and D.-A. Clevert, *Chemical science*, 2019, **10**, 1692–1701.

[8] R.-R. Griffiths and J. M. Hernández-Lobato, *Chemical science*, 2020, **11**, 577–586.

[9] R. Winter, F. Montanari, A. Steffen, H. Briem, F. Noé and D.-A. Clevert, *Chemical science*, 2019, **10**, 8016–8024.

[10] A. Kadurin, S. Nikolenko, K. Khrabrov, A. Aliper and A. Zhavoronkov, *Molecular pharmaceutics*, 2017, **14**, 3098–3104.

[11] T. Blaschke, M. Olivecrona, O. Engkvist, J. Bajorath and H. Chen, *Molecular informatics*, 2018, **37**, 1700123.

[12] M. Popova, O. Isayev and A. Tropsha, *Science advances*, 2018, **4**, eaap7885.

[13] M. Goel, S. Raghunathan, S. Laghuvarapu and U. D. Priyakumar, 2021.

[14] G. L. Guimaraes, B. Sanchez-Lengeling, C. Outeiral, P. L. C. Farias and A. Aspuru-Guzik, *arXiv preprint arXiv:1705.10843*, 2017.

[15] B. Sanchez-Lengeling, C. Outeiral, G. L. Guimaraes and A. Aspuru-Guzik, 2017.

[16] E. Putin, A. Asadulaev, Y. Ivanenkov, V. Aladinskiy, B. Sanchez-Lengeling, A. Aspuru-Guzik and A. Zhavoronkov, *Journal of chemical information and modeling*, 2018, **58**, 1194–1204.

[17] E. Putin, A. Asadulaev, Q. Vanhaelen, Y. Ivanenkov, A. V. Aladinskaya, A. Aliper and A. Zhavoronkov, *Molecular pharmaceutics*, 2018, **15**, 4386–4397.

[18] N. De Cao and T. Kipf, *arXiv preprint arXiv:1805.11973*, 2018.

[19] Ł. Maziarka, A. Pocha, J. Kaczmarczyk, K. Rataj, T. Danel and M. Warchoł, *Journal of Cheminformatics*, 2020, **12**, 1–18.

[20] O. Prykhodko, S. V. Johansson, P.-C. Kotsias, J. Arús-Pous, E. J. Bjerrum, O. Engkvist and H. Chen, *Journal of Cheminformatics*, 2019, **11**, 1–13.

[21] J. You, B. Liu, R. Ying, V. Pande and J. Leskovec, *arXiv preprint arXiv:1806.02473*, 2018.

[22] Y. Khemchandani, S. O'Hagan, S. Samanta, N. Swainston, T. J. Roberts, D. Bollegala and D. B. Kell, *Journal of cheminformatics*, 2020, **12**, 1–17.

[23] Z. Zhou, S. Kearnes, L. Li, R. N. Zare and P. Riley, *Scientific reports*, 2019, **9**, 1–10.

[24] T. Klucznik, B. Mikulak-Klucznik, M. P. McCormack, H. Lima, S. Szymkuć, M. Bhowmick, K. Molga, Y. Zhou, L. Rickershauser, E. P. Gajewska *et al.*, *Chem*, 2018, **4**, 522–532.

[25] S. Ishida, K. Terayama, R. Kojima, K. Takasu and Y. Okuno, *Journal of chemical information and modeling*, 2019, **59**, 5026–5033.

[26] S. Chen and Y. Jung, *JACS Au*, 2021, **1**, 1612–1620.

[27] M. H. Segler, M. Preuss and M. P. Waller, *Nature*, 2018, **555**, 604–610.

[28] J. S. Schreck, C. W. Coley and K. J. Bishop, *ACS central science*, 2019, **5**, 970–981.

[29] S. Genheden, A. Thakkar, V. Chadimová, J.-L. Reymond, O. Engkvist and E. Bjerrum, *Journal of cheminformatics*, 2020, **12**, 1–9.

[30] B. Chen, C. Li, H. Dai and L. Song, Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event, 2020, pp. 1608–1616.

[31] B. Liu, B. Ramsundar, P. Kawthekar, J. Shi, J. Gomes, Q. Luu Nguyen, S. Ho, J. Sloane, P. Wender and V. Pande, *ACS central science*, 2017, **3**, 1103–1113.

[32] P. Karpov, G. Godin and I. V. Tetko, International Conference on Artificial Neural Networks, 2019, pp. 817–830.

[33] S. Zheng, J. Rao, Z. Zhang, J. Xu and Y. Yang, *Journal of Chemical Information and Modeling*, 2019, **60**, 47–55.

[34] E. Kim, D. Lee, Y. Kwon, M. S. Park and Y.-S. Choi, *Journal of Chemical Information and Modeling*, 2021, **61**, 123–133.

[35] K. Mao, X. Xiao, T. Xu, Y. Rong, J. Huang and P. Zhao, *Neurocomputing*, 2021, **457**, 193–202.

[36] S.-W. Seo, Y. Y. Song, J. Y. Yang, S. Bae, H. Lee, J. Shin, S. J. Hwang and E. Yang, Proceedings of the AAAI Conference on Artificial Intelligence, 2021, pp. 531–539.

[37] K. Lin, Y. Xu, J. Pei and L. Lai, *Chemical Science*, 2020, **11**, 3355–3364.

[38] P. Schwaller, R. Petraglia, V. Zullo, V. H. Nair, R. A. Haeuselmann, R. Pisoni, C. Bekas, A. Iuliano and T. Laino, *Chemical science*, 2020, **11**, 3316–3325.

[39] C. Shi, M. Xu, H. Guo, M. Zhang and J. Tang, International Conference on Machine Learning, 2020, pp. 8818–8827.

[40] V. R. Somnath, C. Bunne, C. W. Coley, A. Krause and R. Barzilay, *arXiv preprint arXiv:2006.07038*, 2020.

[41] C. Yan, Q. Ding, P. Zhao, S. Zheng, J. Yang, Y. Yu and J. Huang, *arXiv preprint arXiv:2011.02893*, 2020.

[42] Z. Wang, X. Feng, J. Liu, M. Lu and M. Li, *Microchemical Journal*, 2020, **159**, 105395.

[43] J. A. Fine, A. A. Rajasekar, K. P. Jethava and G. Chopra, *Chem. Sci.*, 2020, **11**, 4618–4630.

[44] J. Zhang, K. Terayama, M. Sumita, K. Yoshizoe, K. Ito, J. Kikuchi and K. Tsuda, *Science and Technology of Advanced Materials*, 2020, **21**, 552–561.

[45] E. Jonas, Advances in Neural Information Processing Systems, 2019.

[46] B. Sridharan, S. Mehta, Y. Pathak and U. D. Priyakumar, *Unpublished Work*.