Machine Learning in Drug Discovery

Article *in* Journal of Chemical Information and Modeling · August 2018

DOI: 10.1021/acs.jcim.8b00478

CITATIONS
10

READS
999

3 authors, including:

Sepp Hochreiter
Johannes Kepler University Linz
193 PUBLICATIONS **40,546** CITATIONS
SEE PROFILE

Günter Klambauer
Johannes Kepler University Linz
76 PUBLICATIONS **1,884** CITATIONS
SEE PROFILE

JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING
Cite This: J. Chem. Inf. Model. 2018, 58, 1723−1724
Editorial
pubs.acs.org/jcim

Machine Learning in Drug Discovery

Currently, machine learning methods drive the success of artificial intelligence in academia and industry. Among machine learning methods, Deep Learning has emerged as a game-changer in many fields and has already impacted a wide range of scientific areas.[1] Deep Learning is founded on novel algorithms and architectures together with the recent availability of very fast computers and massive data sets. In its core, Deep Learning discovers multiple levels of distributed representations of the input, with higher levels representing more abstract concepts. These representations considerably improved data analysis in many research areas. In particular, deep neural networks (DNNs) substantially increased the performance in computer vision, speech recognition, among many other fields. Surprisingly, models developed in a number of fields have reached human performance or above. Nevertheless, the risk of overfitting remains and care has to be taken to not overestimate the generality models can achieve.

A long tradition connects machine learning to drug discovery. The first application to drug discovery was reported by Corwin Hansch in 1962,[2] and since this time, statistical methods, producing quantitative models for complex biological, chemical, or physical phenomenon, have found their place in the computational chemist's toolbox. Since biological systems are often too complex to capture all relevant features, machine learning methods are an attractive alternative to physics-based models. Consequently, machine learning methods have attracted significant attention with DNNs being among the top performers in the Merck Molecular Activity Challenge 2013[3] and the Tox21 Data Challenge 2015.[4] Since the initial successful applications, many novel machine learning methods in diverse application areas of drug discovery have been described.[5] Recently, complex models predicting synthesis routes for compounds were reported yielding synthetic strategies that an experienced chemist could not distinguish between a human or computer-generated approach.[6] Many of the major pharmaceutical companies initiated machine learning projects, and drug discovery start-ups with a focus on machine learning have sprouted up in recent years.

Despite all of the success stories, drug discovery has never been and certainly will never be an easy playground for machine learners. In contrast to other fields, it remains a challenge to find the right representations of complex objects like molecules or molecular complexes, which allow the development of a causal relationship with respect to experimental data. While standard, substructure-oriented fingerprints yield good results in some applications, they might fall short when it comes to the description of bioactivity. Handling molecular shape and pharmacophores, hydrogen bonds, and desolvation effects highlight the importance of three-dimensional representations, which in turn require a proper handling of conformational space and molecular flexibility.

Machine Learning methods strongly depend on experimental training data which opens a second, even more challenging consideration. For the most desirable predictions, the corresponding data likely has high noise levels and a high level of uncertainty and sometimes can even be inconsistent. Experiments are expensive and result in sparse and unbalanced data sets. If detailed experimental conditions influence the data, their comparability across different experiments or laboratories can be a significant consideration. New challenges come with new experiments, how to deal with microscopic images stemming from high-content screening, and how to deal with heterogeneous data sets. Although image recognition is a clear strength of ML it is yet unclear whether these data sets can be handled with convolutional neural networks (CNNs) and the currently available computational power.

Nevertheless, these challenges will be faced and the opportunities for sophisticated machine learning techniques in drug discovery are manifold. The full potential will only be exploited if chemists, cheminformaticians, and machine learners join forces to develop tailored solutions. Of equal importance will be the development of realistic method assessments allowing the objective evaluation of the predictive power of novel machine learning models. Finally, close collaboration with experimentalists to create curated and validated data sets is crucial for the field. We are convinced that this *Journal of Chemical Information and Modeling* special issue will contribute and advance this burgeoning field. We envision that this will spurn the creation of better machine learning methods in drug discovery and provide an objective view of their capabilities.

Sepp Hochreiter[*,†]
Guenter Klambauer[†]
Matthias Rarey[*,‡]

[†]LIT AI Lab & Institute of Bioinformatics, Johannes Kepler University, 4040 Linz, Austria

[‡]ZBH−Center for Bioinformatics, Universität Hamburg, 20146 Hamburg, Germany

AUTHOR INFORMATION

Corresponding Authors

*E-mail: hochreit@bioinf.jku.at.
*E-mail: rarey@zbh.uni-hamburg.de.

ORCID

Guenter Klambauer: 0000-0003-2861-5552
Matthias Rarey: 0000-0002-9553-6531

Notes

Views expressed in this editorial are those of the authors and not necessarily the views of the ACS.

REFERENCES

(1) LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521* (7553), 436.

(2) Hansch, C.; Maloney, P.; Fujita, T.; Muir, R. Correlation of Biological Activity of Phenoxyacetic Acids with Hammett Substituent Constants and Partition Coefficients. *Nature* **1962**, *194*, 178−180.

Published: August 15, 2018

DOI: 10.1021/acs.jcim.8b00478
J. Chem. Inf. Model. 2018, 58, 1723−1724

(3) Ma, J.; Sheridan, R. P.; Liaw, A.; Dahl, G. E.; Svetnik, V. Deep neural nets as a method for quantitative structure−activity relationships. *J. Chem. Inf. Model.* **2015**, *55* (2), 263−274.
(4) Mayr, A.; Klambauer, G.; Unterthiner, T.; Hochreiter, S. DeepTox: toxicity prediction using deep learning. *Front. Environ. Sci.* **2016**, *3*, 80.
(5) Gawehn, E.; Hiss, J. A.; Schneider, G. Deep learning in drug discovery. *Mol. Inf.* **2016**, *35* (1), 3−14.
(6) Segler, M. H.; Preuss, M.; Waller, M. P. Planning chemical syntheses with deep neural networks and symbolic AI. *Nature* **2018**, *555* (7698), 604.

DOI: 10.1021/acs.jcim.8b00478
J. Chem. Inf. Model. 2018, 58, 1723−1724