Nature Reviews Chemistry | https://doi.org/10.1038/s41570-020-0183-2 | Published online 6 April 2020

MACHINE LEARNING

Fingerprints of molecular reactivity

> “We believe that our approach can help to lower the barrier to the use of machine learning techniques in organic synthesis”

Machine learning is becoming one of the ‘essentials’ in the toolbox of the modern chemist. Copious algorithms have been developed that can guide and enhance the discovery of new molecules and materials. The performances of these algorithms are undergoing constant optimization. An essential component of these optimizations is the simplification and generalization of models and input parameters that can be used to explore unknown chemical space. A group of chemists and computer scientists led by Frank Glorius at Westfälische Wilhelms-Universität Münster has now developed a machine learning approach that uses the molecular structure of reactants alone as input parameter to predict the compound reactivity.

Despite the many developments of machine learning algorithms, they have not yet been widely adopted among the organic chemistry community in the search for new synthetic methods. In some cases, these models are problem specific and cannot be generally applicable or are based on non-chemically meaningful parameters. “In our opinion, a major reason for the resistance of synthetic chemists to the use of these models is the high complexity of these approaches particularly when aiming for quantitative reaction outcome prediction, such as the prediction of yields and selectivities,” says Glorius. “These predictive approaches are mostly based on a prior understanding of the underlying processes and cannot easily be transferred to other problems. Nevertheless, we believe that computational models based on machine learning have the potential to change the way we approach organic syntheses.”

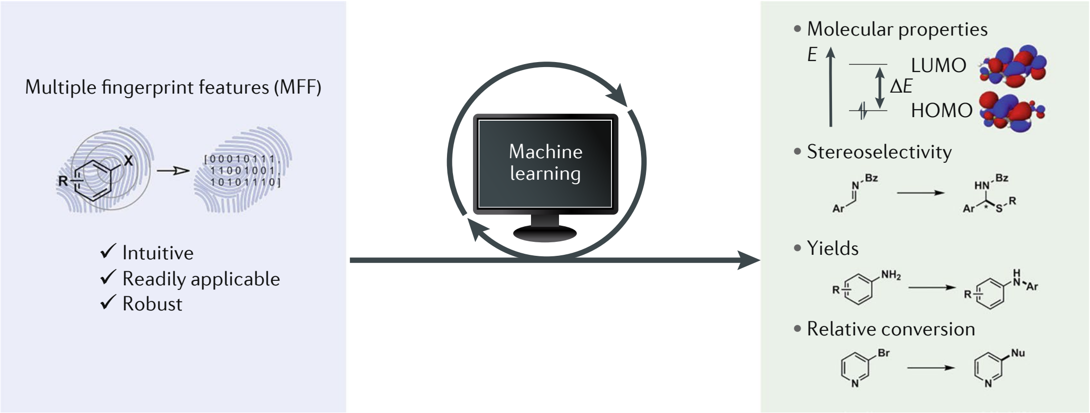
The approach proposed by Glorius and co-workers predicts the outcome of new chemical reactions, including quantitative predictions of enantioselectivities, yields and relative conversion by taking as input only the molecular structure of the reactants in the form of multiple fingerprint features (MFF). “Organic compounds can be represented as graphs. On these graphs, simple structural queries (typically yes/no queries, such as ‘is a carbon atom connected to two oxygens?’ or ‘does the molecule contain an amide?’) can be carried out. So-called molecular fingerprints are number sequences based on the combination of many such successive queries. We use a large number of different fingerprints (MFF) to represent the molecular structure of each compound as accurately as possible,” explains Felix Strieth-Kalthoff, co-author of the article.

The group tested this structure-based machine learning approach to predict common molecular properties such as HOMO–LUMO gaps and then to predict enantioselectivities and yields. These tests revealed that the structure-based approach can make predictions as accurate as those obtained from problem-oriented approaches but much more quickly. Finally, this model was tested to predict reaction performances on a new data set, never used with machine learning approaches before.

“Most importantly, our approach is robust and can be easily applied to diverse and very different problems. Moreover, the training data sets used within this study were relatively small and can easily be generated using contemporary high-throughput experimentation,” remarks Frederik Sandfort, first author of the article.

“When it comes to evaluating large amounts of complex data, computers are fundamentally superior to us. However, our goal is not to replace synthetic chemists with machines, but to support them as effectively as possible,” clarifies Glorius. “Models based on artificial intelligence can significantly change the way we approach chemical syntheses, but we are still at the very beginning. We believe that our approach can help to lower the barrier to the use of machine learning techniques in organic synthesis,” he concludes.

Gabriella Graziano



Credit: Adapted with permission from Sandfort, F. et al. *Chem* https://doi.org/10.1016/j.chempr.2020.02.017 (2020), Elsevier.

ORIGINAL ARTICLE Sandfort, F. et al. A structure-based platform for predicting chemical reactivity. *Chem* https://doi.org/10.1016/j.chempr.2020.02.017 (2020)