Chapter 5

Machine Learning Methods in Computational Toxicology

Igor I. Baskin

Abstract

Various methods of machine learning, supervised and unsupervised, linear and nonlinear, classification and regression, in combination with various types of molecular descriptors, both "handcrafted" and "data-driven," are considered in the context of their use in computational toxicology. The use of multiple linear regression, variants of naïve Bayes classifier, *k*-nearest neighbors, support vector machine, decision trees, ensemble learning, random forest, several types of neural networks, and deep learning is the focus of attention of this review. The role of fragment descriptors, graph mining, and graph kernels is highlighted. The application of unsupervised methods, such as Kohonen's self-organizing maps and related approaches, which allow for combining predictions with data analysis and visualization, is also considered. The necessity of applying a wide range of machine learning methods in computational toxicology is underlined.

Key words Computational toxicology, Machine learning, Support vector machines, Random forest, Neural networks, Deep learning

1 Introduction

In view of the wide variety of different types of toxicity, endpoints, and mechanisms of action, computational toxicology intensively uses approaches of different types, including rule-based expert systems, molecular docking, pharmacophore (toxicophore) modeling, quantum chemistry studies, and building structure–activity models using machine learning methods. In recent years, due to the accumulation of a large amount of data on different toxicity endpoints in databases, the rapid grows of computer power, the development of sophisticated data analysis algorithms, the role of machine learning methods in computational toxicology has become crucial. This is reflected in several reviews [1–11] concerning computational toxicology. The use of machine learning in building structure–activity models, including quantitative structure–activity relationships (QSAR) models, is comprehensively analyzed in several reviews [12, 13].

In this chapter, the application of various methods of machine learning in computational toxicology is surveyed. Unlike other

Orazio Nicolotti (ed.), *Computational Toxicology: Methods and Protocols*, Methods in Molecular Biology, vol. 1800,
https://doi.org/10.1007/978-1-4939-7899-1_5,