Machine Learning to Predict Drug Bioactivities from Tandem Mass Spectra

May 21, 2019

Vilma Jägerroos, a student of Life Science Technologies at Aalto University, developed a machine learning pipeline to predict drug bioactivities from tandem mass spectra (MS/MS) in her Master’s thesis titled “Predicting Drug Bioactivities from Tandem Mass Spectra”. Predicting bioactivities based on MS/MS spectra can be used to prioritize the most promising samples for further experimental testing and therefore accelerate the drug discovery process from natural products.

Natural products have been the single most productive source of lead compounds for the modern drug development. In traditional drug discovery from natural products, concentrated extracts prepared from, e.g., plant samples were screened to determine their biological activity. These extracts are complicated mixtures of multiple compounds. Thus, determination of bioactivity may be confounded, e.g., by synergistic effects of several compounds. However, isolating each compound from the extract prior to screening would be inefficient when large number of samples are screened. Structures of compounds in a natural product sample are unknown in advance. Analytical methods, such as tandem mass spectrometry, are used to identify constituents of the samples in almost every stage of drug discovery process from natural products.

In the thesis, a machine learning pipeline was developed to predict bioactivities from MS/MS spectra. Relation between bioactivities and MS/MS spectra is learned by combining information from multiple databases. Furthermore, it is shown that predicting model can be build even when there is no drugs for which both bioactivities and MS/MS spectra is measured. The method achieves comparable performance to models that use known structure of drug compound to predict bioactivities.

More information:
Vilma Jägerroos
vilma.jagerroos@aalto.fi