Poster M-5

Predicting Chemical Properties of Metabolites

Authors:
Roman Eisner (*University of Alberta*)
Craig Knox (*University of Alberta*)
Russ Greiner (*University of Alberta*)
David Wishart (*University of Alberta*)

Short Abstract: We present our efforts in learning to predict chemical properties of metabolites. These properties, such as solubility, are important for completing the annotations in the Human Metabolite Database (http://www.hmdb.ca), and also for facilitating further research on these compounds. This poster evaluates our predictors and discusses the results.

Long Abstract:

The Human Metabolome Project (HMP) is a large scale multi-year project launched in January 2005, funded by Genome Canada. The purpose of the project is to better understand the set of all small molecules that appear in the human body at concentrations greater than one micro-molar. In addition to identifying and cataloguing this set of metabolites, the project also aims to record many chemical and biological properties of each metabolite, as well as their NMR and MS spectra. We also connect each metabolite with associated diseases, which will help to improve disease identification, prognosis, and monitoring. This project will also provide insight into drug metabolism and toxicology and provide a linkage between the human metabolome and the human genome. This information, as well the software tools developed, will be freely accessible in an electronic format to all researchers through the Human Metabolome Database (http://www.hmdb.ca). In addition, all compounds will be publicly available through our Human Metabolome Library for further study.

Knowing the chemical properties of metabolites (such as solubility, logp, melting point, etc) is important to the HMP for two reasons. First, it is part of the project's mandate to catalogue various properties of all of the metabolites in the human body. Knowing these properties will help make the annotations in the database more complete. Second, knowing properties of chemicals aids in their further study through experimental methods. For example, the solubility of a compound is important for NMR sample preparation. Cataloguing these properties is important in its own sake, and the knowledge gained also aids in further research.

We present our efforts to predict several chemical properties of human metabolites. Our initial efforts using descriptors generated from the JOELib and Chemaxon libraries in conjunction with machine learning tools showed promise for predicting the water solubility of metabolites. Further experiments showed that the method worked for the prediction of other chemical properties as well. The methods presented use a variety of molecular descriptors which require knowing the structure of the metabolite a priori to generate. Our method is not

limited to human metabolites, generalizing to any small compound where the structure is known.

In this poster, we present our methods, and analyze the results of prediction. We present the quality of our results for cross-validation on experimentally annotated datasets, and also examine which descriptors are most important for each predictor.