KATHOLIEKE UNIVERSITEIT LEUVEN

DEPT COMPUTERWETENSCHAPPEN
DEPT OF COMPUTER SCIENCE

Towards prediction of structured values

Pieter-Jan Drouillon
Dept. of Computer Science

1 Introduction

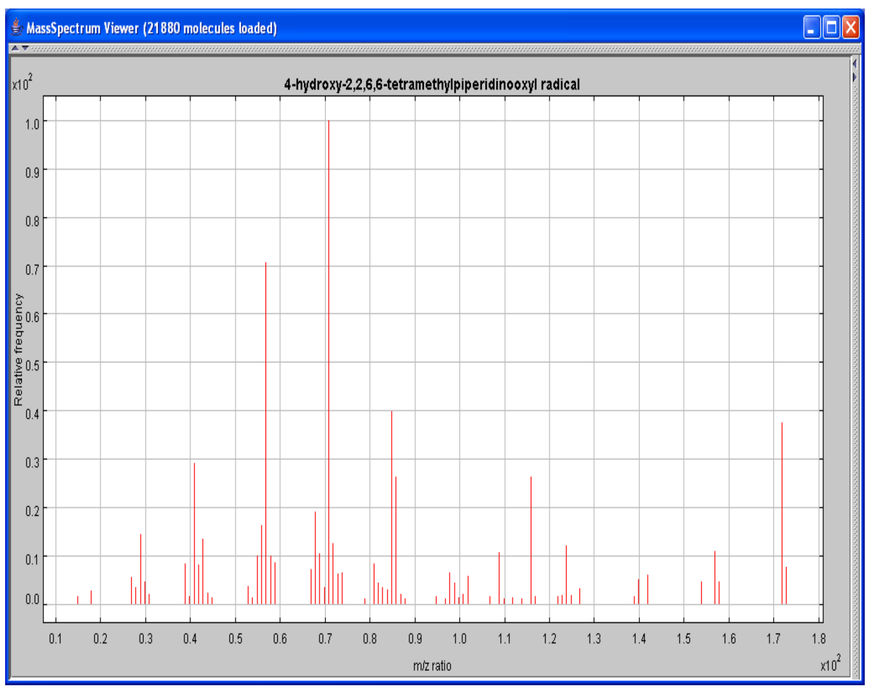
Learning predictive models is a common task. Based on input values, a predictive model delivers a set of output values. These input values are a mixture of different types such as numerical, categorical or structured values. They are entered in the model which computes the outcome. The output is typically a single value or a vector of values.

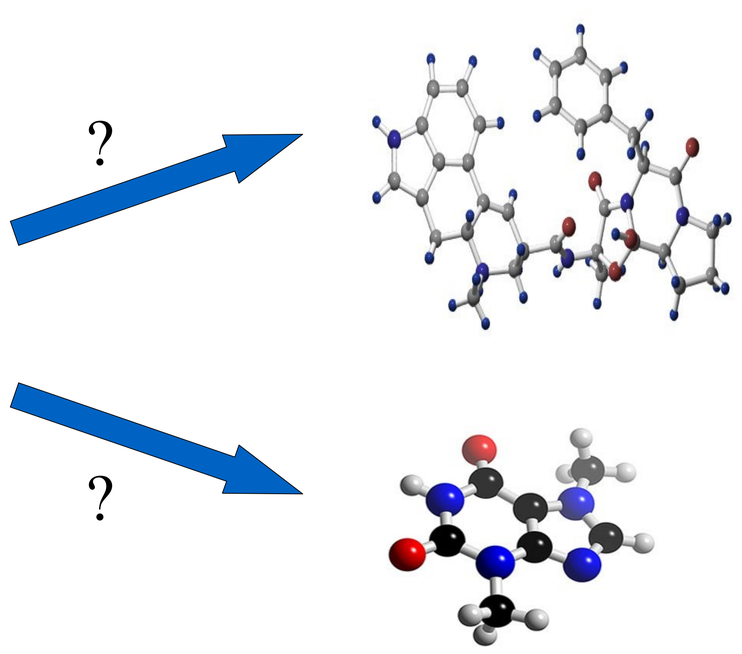
$$f : R^n \rightarrow R^m : x_1, \ldots, x_n \rightarrow f(x_1, \ldots, x_n) = y_1, \ldots, y_m$$

2 Problem description

- Most common output values: numerical/categorical, what about structured values?
- How to learn a model, predicting a structured value?
  - What are the needs?
  - What are possible restrictions?
- An example application: deriving the structure of a molecule based on its mass spectrogram.

$$f : \textit{mass spectrogram} \rightarrow \textit{molecule structure}$$





3 Approach

- Use of methods that can handle structured values, such as ILP
- Predict numerical features such as number of atoms
- Predict categorical features such as class of molecule (alcohol, alkane,...)
- Combine all this with physical and chemical constraints

Further information
Pieter-Jan Drouillon
K.U.Leuven – Dept. Computerwetenschappen
Declarative Languages and Artificial Intelligence research group
Celestijnenlaan 200A
3001 Leuven (Heverlee), Belgium

Supervisor: Hendrik Blockeel

Pieter-Jan.Drouillon@cs.kuleuven.be
http://www.cs.kuleuven.be/~pieterjd