P-21

SOLUBILITY OF CHEMICAL COMPOUNDS

Mariam Pirashvili
University of Southampton

This is a report on joint work with the Joining the Dots group at the University of Southampton.

Given a set of chemical compounds, we wish to predict their solubility in water. This is a regression problem that has been tackled by traditional machine learning methods previously. Some of their results were available to us and we were able to compare them with the topological methods we empoyed.

The data is given in two forms. The first is as a matrix of descriptors, or features, of logical, categorical and continuous type. The second is as molecular graphs – that is, as undirected, connected, labelled graphs. We built different models based on the form of the data. The output variable, solubility, is continuous.

Besides the task of predicting solubility, we also tried to discover the shape of the space of these molecules, to gain further insight into the chemical and physical properties of molecules that predict their solubility.