Passion Project Proposal

Using Neural Networks to Predict Drug Toxicity

Sean Davern

Seattle, Fall Cohort, weeks 9-12

In 2014 the NIH National Center for Advancing Translational Sciences initiated what they called the Tox21 data challenge. [1] Referencing toxicology in the 21st century, the initiative was to compare computational methods for predicting the biological toxicity of chemicals and drugs based on their chemical structure data. The challenge uses over 800 structural attributes of over 12,000 chemical compounds and their toxicity as measured by 12 different assays to predict the toxicity of 650 other compounds. My goal would be to build and train my own neural network and see if it can outperform the machine learning approaches reported in the article: SVM, Random Forests, Elastic net. [2,3] Additionally, a listing of the final leaderboard for the 2014 challenge is available. [4]

Data

The data is provided by the Tox21 article authors. [2] The more than 800 chemical structure attributes are mimally described here by the Computational Biology & Drug Design Group. The toxicity assays are described in more detail here by NIH. However, the data analysis may not require a thorough analysis of the attribute nor toxicity assays.

The original data has been compiled into a format more easily consumed and processed today by the Institute of Bioinformatics. [2]

1. https://tripod.nih.gov/tox21/challenge/index.jsp↩

2. Mayr, Andreas; Klambauer, Guenter; Unterthiner, Thomas; Hochreiter, Sepp (2016). "DeepTox: Toxicity Prediction Using Deep Learning". *Frontiers in Environmental Science*. **3**: 80. Data↩↩↩

3. [Huang2016] Huang, R., Xia, M., Nguyen, D. T., Zhao, T., Sakamuru, S., Zhao, J., Shahane, S., Rossoshek, A., & Simeonov, A. (2016). Tox21Challenge to build predictive models of nuclear receptor and stress response pathways as mediated by exposure to environmental chemicals and drugs. *Frontiers in Environmental Science*, **3**:85.↩

4. https://tripod.nih.gov/tox21/challenge/leaderboard.jsp↩