Garrett Wankel [Aug 20]

Mechanism of Action Prediction

12th October 2020

Context

An initial step in a drug discovery program for a pharmaceutical company is defining a candidate's mechanism of action (MoA). This process involves testing within an in vitro setting and observing patterns of interactions between the newly formulated compound and human cells and genes. By identifying trends and measuring levels of the compounds inhibition or stimulating effects a MoA can be determined. Other variables like dosing regimen, which includes the recommended frequency of medicating and the strength of medication are used during the clinical trial stage of development, but before it reaches the clinical stage, determining its MoA becomes an important safety step by identifying a safe blood toxicity level. In short, it helps determine a safe level of medication to be present in the bloodstream in order to achieve the compound's desired effect. As the pharmaceutical industry matures, specialized medicine is becoming more popular with some orphan or rare genetic diseases finally becoming treatable, and discovering a reliably measurable MoA is a vital step in the process. Further, by validating a compound's MoA, a more predictive outcome will be achievable, thus leaving less variability to the clinical trial process and the creation of a comprehensive and robust argument for the candidate-compound to treat certain conditions. The value created by understanding a compound's MoA comes from its viability to treat a certain condition and once understood, a company can accurately design and power its clinical trials with an intentionally narrowed outcome in mind. So with the total cost of developing a marketable medication safely in the millions of dollars, mitigating risk and uncertainty of the drug's safety, activity and effectiveness is paramount to generating interest from investors as well as potentially shifting the paradigm with how medical conditions are treated.

Criteria For Success

For this dataset:

1. Create an accurate multilabel model, and achieve a low average log-loss on the MoA annotation for each drug.

Scope of Solution

This dataset and analysis is specific to the Research and Development branch of a pharmaceutical company. Creating an accurate algorithm will help the clinical team determine viable candidates for certain conditions, and the algorithm's reliability will also help ensure management that the chosen candidates have a high probability of clinical success relative to other candidates. Assisting both the clinical and managerial teams with this algorithm will create cost-savings and value-creation measures that could amount of millions or billions of dollars.

Constraints

1. Dosing strength:
    a. Total strength in milligrams or micrograms may be important to determine effectiveness when designing pivotal trials and choosing the optimal dose strength.

Key Stakeholders

1. Laboratory for Innovation Science at Harvard (LISH):
    a. They are providing the data for the Kaggle competition.
2. The Connectivity Map:
    a. The Connectivity Map at the Broad Institute provides the genome-scale library of cellular signatures that catalogs transcriptional responses to chemical, genetic and disease perturbation.
3. Kaggle:
    a. They are hosting this competition in conjunction with LISH and The Connectivity Map teams.

Key Data Sources

1. https://www.kaggle.com/c/lish-moa/data - link to the competition page and corresponding dataset.