CALIFORNIA STATE SCIENCE FAIR
2016 PROJECT SUMMARY

Name(s)

Tanisha M. Joshi

Project Number

S1411

Project Title

A Novel Cognitive Knowledge Harvesting Approach to Train a Self Learning System for Drug Predictive Models

Abstract

Objectives/Goals

The goal is to develop an incremental self-learning model that helps in the prediction of drug behavior effectively increasing the net knowledge without human intervention

Methods/Materials

I built an iterative learning model that makes measurable progress with each iteration. I designed a BioEntity based feature-vector abstraction for capturing multidimensional harvested interactions. It taps into ontological specificity across knowledge networks to exploit existing knowledge, and explore new knowledge.

Results

My knowledge network generated unprecedented novel cross validations that include 1.16 million BioEntities, 2.008 million relationships, and 6.578m properties. The final knowledge sets were generated as a result of successive approximations of incremental knowledge gain/loss and the number of iterations have a finite upper bound that converges to an order of O(N*LogN) in terms of cost complexity. Through each iteration, new knowledge insights were generated that asymptotically and monotonically converge to the expected knowledge set with an 81% accuracy level. The P-value of the confidence aggregate score of the interactions consensus across 30 sample runs was 0.9236 at 0.05 significance level.

Conclusions/Discussion

A set of Bio-Entities is the initial input which represents a specific assay and constitutes the research problem of interest. This assay, when studied for harvesting knowledge generation created a meaningful knowledge set that more closely aligns with an associated confidence score. The generated knowledge was corroborated with a Lapatinib use case subgraph that included EGFR/HER2 receptor binding, MAPK signaling pathways reported in literature.The project successfully demonstrates an operational self learning system that can evolve, update knowledge continuously, and improve drug behavior predictions in randomized cell# based assays. Knowledge was corroborated with a Lapatinib use case sub graph that included EGFR/HER2 receptor binding, MAPK signaling pathways reported in literature.

Summary Statement

Developing iterative self-learning knowledge network topologies from curated life science datasets which converge to expected knowledge, and performing ontological feature based analysis using this knowledge to help predict drug behavior.

Help Received

Archana Gangakhedka, Drug Researcher & Dr. Lato provided verification of experimental methodology, validation of results, tips for presentation. My teacher Mrs. Taylor sponsored my project.