Self Driving Pharma: A Novel Cognitive Knowledge Harvesting Approach to Train a Self-Learning System For Drug Predictive Models through Multi-Dimensional Bio-Entity Feature-Vector Based Topological Data Analyses

Joshi, Tanisha

82m researchers are updating public curated databases with information about bio-molecular interactions and the role they play in drug behavior. This has led to an era of dynamic big data processing often involving brute force computational methods. The purpose of this project is to develop a novel iterative self-learning system to predict how drugs will perform and allow the pharmaceutical industry to be better prepared to optimize their multi-year multi-billion dollar investments in drug trials. Correlation analysis of a candidate drug with similar drug topologies helps recycle validated methods on new drug targets along with molecular differentiation. Proposed method models complex knowledge networks enriched with 15 different micro-structural, anatomical and functional ontologies by projecting a bio-ranked graph topology that uses an Euclidean distance metric to compute knowledge similarity. A data ingestion system, MedGopher was developed to access data (on genes, proteins, drugs, enzymes, ligands and other entities) from public databases. A bio-entity based feature-vector abstraction was created for capturing multi-dimensional interactions harvested using a framework created called MedCognition and incorporates a feedback loop that uses the assay results to train the next iteration cycle. Through each iteration, new knowledge insights were generated that asymptotically 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. 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.