9
Summary

Despite technological progresses and improved understanding of biological systems, discovery of drugs is an inefficient, arduous and expensive process. Research and development cost for discovery of new molecular entities is expensive, largely attributed to high attrition rate of candidate drugs due to adverse drug reactions. Computational models have been shown to be effective in modeling and prediction of phenotypic side effects (adverse drug reactions). Rational drug design works through modulation of a 'target', the key protein anticipated to be critical for molecular mechanisms underlying disease phenotype. Interaction of drugs with 'off-targets' cause undesired phenotypic side effects. Modeling of mechanisms of side effects has been attempted through various methods such as machine learning and graph theory.

The action of drug needs to be seen from the systems perspective knowing that cellular mechanisms form a web of interactions. Availability of data regarding the molecular interaction of drugs (DrugBank), their phenotypic side effects (SIDER) as well as that of molecular interactomes (HPRD) have facilitated systems modeling of drug-target-side effects, aimed at accurate prediction of side effects. In this thesis, we aimed to incorporate drug-profiles based on their molecular interactions, reported side effects and their chemical descriptors to build quantitative models. With the help of graph theoretical modeling and canonical correlation analysis, we worked towards finding effective set of features and computational complexity.

We implemented generalized canonical correlation models to integrate 2D and 3D chemical properties of drugs as well as their systems-level target profiles for prediction of adverse drug reactions [Kanji et al., 2015]. We observed that the model incorporating chemical features outperformed that incorporating target profiles. Jurs and Electrostatic chemical descriptors, that carry electronic information, were identified to yield best results, thereby implying their relevance in specifying adverse drug reactions. Further, from the analysis of 'genomic space' that embeds information of target genes and their similarity, we propose 'degree' as one the most useful topological features on the basis of its prediction performance and time complexity ($o(n^2)$). We have also developed a partial canonical correlation model to investigate interdependence among drug features so as to facilitate enumeration of contribution from individual drug features towards prediction of side effects. It also assists in identification of best set of features that could be used effectively in an integrative model. Beyond predicting side effects without any knowledge of known adverse reactions of drugs, we have also implemented canonical correlation model in conjunction with rank aggregation method with different proportion of known side effects to predict unknown side effects. Our integrative models that incorporate relevant aspects of drugs, such as chemical features and system-level properties of genomic space, offer important insights into mechanisms of adverse drug reactions and provide data-driven methods for their prediction.

…