

Computational methods for inference and prediction of novel drug-drug interactions

Min Qian^{1, a}, Liao Jun^{2, b}

¹School of Sciences, China Pharmaceutical University, Nanjing, 210009, China

²School of Sciences, China Pharmaceutical University, Nanjing, 210009, China

^aemail:minqian2006@126.com, ^bemail:liaojun@cpu.edu.cn

Keywords: in silico; DDI; algorithm

Abstract. It is critically important for drug development and clinical safety to identify and assess drug-drug interactions (DDIs) early and accurately. In addition to in vitro laboratory studies, *in silico* methods exist that may supplement the screening and detection of DDIs in large-scale databases. Here we review recent novel approaches that uncover underlying DDIs from the databases constructed with drug information, adverse event reports, electronic medical records, and other sources to give revelation to the researchers for further surveys.

Introduction

A 2007 report described that DDIs caused 4.8% of hospital admissions among the elderly, as they are often administrated multiple medications, increasing their risk almost 8.5-fold relative to the general population[1]. Thus detecting DDIs early in the drug discovery process can cut down on the number of hospitalizations owing to drug interactions and allow more beneficial drugs to reach the market quickly and at a lower cost. Besides adverse effects, potentially beneficial effects of DDIs are equally important. Some drugs show synergistic effects against their targets which increases the efficacy of treatments and improve the selectivity[2]. Antagonistic interactions can be used to detect the mechanisms of action of drugs, and even suppressing interactions have been found to be potentially related to the fight against antibiotic-resistant pathogens[3].

Both traditional in vivo and in vitro laboratory studies exploring likely DDIs, and modeling methods, such as pharmacokinetic models of drug levels within the body make contributions to the detection of DDIs. Unfortunately, the above approaches are slow and costly processes and may not uncover all the interactions. Thus, computational methods in predicting DDIs based on large-scale databases developed quickly in recent years due to their efficiency and accuracy. Here, we summarize the existing *in silico* methods, and concluded a prevalent main workflow of these papers in Figure 1. In the figure you can find that first, data were integrated from multiple sources, including drug essential data, drug interaction data and mechanism data. Next, typical features were extracted and calculated from the data sources. Finally different algorithms were selected to construct models to predict DDI, and evaluations were performed for the results based on modeling experiments.

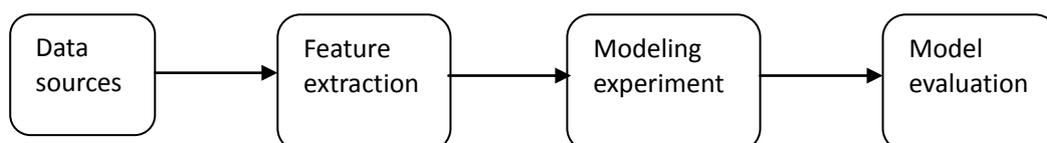


Fig1. Overview of the main workflow.

Different type of models

We find that most of the models have their special emphases on one of the first three steps that is to say we can divide the models into three typical groups according to their different emphases.

Type 1

In this type, models put their focus on the construction of data sources. The adverse drug events spontaneous reporting system is not only used to detect DDIs but also to predict novel ADRs and drug targets. However, there are some shortcomings in this data source such as low reporting numbers, under reporting, and uncharacterized confounding factors which may limit the use of signal detection methods in the analysis of such data. To address the problem Tatonetti *et al.* [4] used an adaptive data-driven method for correcting these factors-- the so-called unknown "covariates." With this information, they constructed a database of drug effects (O_{FFSIDES}) and a database of drug-drug interaction side effects (T_{WOSIDES}) and then used their new database to pinpoint drug targets and detect drug-class interactions.

Type 2

In this group, authors are likely to pay more attention to integrating features from data sources. Typically, it referred to similarity measure and descriptor-based method. In the similarity measure, comparing drugs by looking at their shared features can detect novel DDIs, and uncover information about potential mechanisms and effects additionally. Gottlieb *et al.* used seven similarity measures and combined them into 49 features to calculate the max similarity between query drug pair and all the known DDIs existed in the database[5]. Their method was not only able to reveal latent DDIs but also DDI recommendations and related P450s. INDI predicts the severity level of each DDI of the involved drugs including: contraindicate, generally avoid, adjust dosage and monitor[5]. The model can also made prediction of the CYP enzymes as the condition of patient-specific genetic polymorphisms related to certain CYP enzymes which provided important leads for physicians and pharmaceutical companies alike.

From 2012 to 2014 Vilar and colleges used molecular structure similarity, Interaction profile fingerprints, and 3D pharmacophoric similarity to predict novel DDIs respectively. Recently, they used the above three similarities added with drug-target similarity and adverse drug effects similarity to construct a new protocol[6] with the predicted effect correct in 36 out of 43 DDIs (84%). Not only added with two novel descriptors the new protocol also used principal component analysis (PCA) to show the results of combining the 5 different similarity measures with the STATISTICA software.

In the descriptor-based method, QSAR models have been widely used to predict drug metabolism as in the case of ADMET parameters, drug metabolism is a key determinant of DDIs. In 2008, Li *et al.* reviewed the use of QSAR models to predict P450 metabolism, clearance and metabolic stability[7]. They also made rational recommendations for developing predictable and interpretable QSAR models. These recommendations include the importance of appropriate division of the training and test sets, choosing applicable cut points, and the construction of rationally designed molecular descriptors[7].

Type3

Papers in this type prefer to focus on the specific means or algorithm to build models. Networks are helpful in highlighting patterns in data sets. In 2013, Roger and colleges used Stochastic block models which mathematically formalize the idea of group-dependent interactions to predict novel DDIs without any pharmacological or biochemical information about the drugs, their targets or their mechanisms of action[8]. Their model was able to accurately detect interactions, not only in exhaustive pairwise interaction data between small sets of drugs, but also in large-scale databases. However, one limitation is also obvious in the paper that in the application of predicting new drug interactions, a seed interaction is necessary to find the remaining interactions while in reality not always attainable.

In 2011, a novel method of predicting DDIs was introduced by Sheikh-Bahaei and colleagues[9]. They used *in silico* hepatocyte cultures (ISHCs) to simulate drug metabolism by virtual transporters and P450 enzymes CYP1A2, CYP2C9, CYP2C19 CYP2D6, CYP2E1, and CYP3A4. The model was able to predict clearance rates at the accuracy of 79%.

Text mining is often used to address such models due to its advantages in driving information useful in detecting DDIs. In 2010, Rave *et al.* used a highly optimized and tailored Apriori

algorithm to detect novel post-marketed drug interaction adverse effects (DIAEs)[10]. They identified 1,868 latent drug interaction adverse effects of which 35% are known to interact, the remaining 65% are unknown interactions.

In the “evaluation step” models were usually evaluated with the k-fold cross validation to exam their ability of predicting novel DDIs or the accuracy of result. For example in a five-fold cross validation, the training dataset is divided into 5 subsets with the same size. The model is generated with 4 subsets except subset i for five times and then tested by the remaining subset i ($i=1-5$). Additionally, many models are also evaluated with by an external source other than the initial data source. A receiver operating characteristic curve (ROC) is often generated and the area under the curve (AUROC) represent more accurate interpretation of model performance.

Conclusions

Discovering and predicting DDIs is a tough mission for medicine and public health due to their diverse biological mechanisms and severity. Usually, DDIs rarely happen or are difficult to be detected through the entire drug approval process. Fortunately, the availability of drug-related data in scientific papers, electronic medical records, adverse drug reports from hospitals and patients and numerous databases covering drug interactions, mechanisms, side effects has increased tremendously which lead to the flourishing *in silico* predicting methods. In this paper, three typical groups of models or approaches are outlined, we hope that more papers will be associated with two or three groups to increase the accuracy and explicable mechanism of their models which can be helpful to accelerate the pace of probing the molecular mechanisms of drug response, understanding the basic pathophysiology of organism and disease and developing new drugs or discovering new uses for old drugs.

Acknowledgement

In this paper, the research was sponsored by the Fundamental Research Funds for the Central Universities (Project No. ZJ14040).

References

- [1]Becker, Kallewaard, Caspers, et al. Hospitalisations and emergency department visits due to drug-drug interactions: a literature review [J]. *Pharmacoepidemiology and drug safety*, 2007, 16(6): 641-51.
- [2]Lehar, Krueger, Avery, et al. Synergistic drug combinations tend to improve therapeutically relevant selectivity [J]. *Nature biotechnology*, 2009, 27(7): 659-66.
- [3]Chait, Craney, Kishony. Antibiotic interactions that select against resistance [J]. *Nature*, 2007, 446(7136): 668-71.
- [4]Tatonetti, Ye, Daneshjou, et al. Data-Driven Prediction of Drug Effects and Interactions [J]. *Sci Transl Med*, 2012, 4(125):
- [5]Gottlieb, Stein, Oron, et al. INDI: a computational framework for inferring drug interactions and their associated recommendations [J]. *Molecular systems biology*, 2012, 8(1):
- [6]Vilar, Uriarte, Santana, et al. Similarity-based modeling in large-scale prediction of drug-drug interactions [J]. *Nature protocols*, 2014, 9(9): 2147-63.
- [7]Li, Sun, Fan, et al. Considerations and recent advances in QSAR models for cytochrome P450-mediated drug metabolism prediction [J]. *Journal of computer-aided molecular design*, 2008, 22(11): 843-55.
- [8]Guimera, Sales-Pardo. A network inference method for large-scale unsupervised identification of novel drug-drug interactions [J]. *PLoS computational biology*, 2013, 9(12): e1003374.
- [9]Sheikh-Bahaei, Hunt. Enabling clearance predictions to emerge from *in silico* actions of quasi-autonomous hepatocyte components [J]. *Drug metabolism and disposition: the biological fate of chemicals*, 2011, 39(10): 1910-20.

[10]Harpaz, Haerian, Chase, et al. Statistical Mining of Potential Drug Interaction Adverse Effects in FDA's Spontaneous Reporting System [J]. AMIA Annual Symposium proceedings / AMIA Symposium AMIA Symposium, 2010, 2010(281-5).