Predicting drug-drug interactions based on integrated similarity and semi-supervised learning

ABSTRACT

A drug-drug interaction (DDI) is defined as an association between two drugs where the pharmacological effects of a drug are influenced by another drug. Positive DDIs can usually improve the therapeutic effects of patients, but negative DDIs cause the major cause of adverse drug reactions and even result in the drug withdrawal from the market and the patient death. Therefore, identifying DDIs has become a key component of the drug development and disease treatment. In this study, we propose a novel method to predict DDIs based on the integrated similarity and semi-supervised learning (DDI-IS-SL). DDI-IS-SL integrates the drug chemical, biological and phenotype data to calculate the feature similarity of drugs with the cosine similarity method. The Gaussian Interaction Profile kernel similarity of drugs is also calculated based on known DDIs. A semi-supervised learning method (the Regularized Least Squares classifier) is used to calculate the interaction possibility scores of drug-drug pairs. In terms of the 5-fold cross validation, 10-fold cross validation and de novo drug validation, DDI-IS-SL can achieve the better prediction performance than other comparative methods. In addition, the average computation time of DDI-IS-SL is shorter than that of other comparative methods. Finally, case studies further demonstrate the performance of DDI-IS-SL in practical applications.

**EXISTING SYSTEM**

Recently, based on machine learning models, many computational approaches have been developed to predict potential DDIs. Tatonetti *et al*. developed a signal discovery method to infer DDIs [18], main features of drugs used in this method are drug adverse event profiles. By combining drug chemical similarities, side effect similarities, proteinprotein interaction similarities and target sequence similarities, an INDI (INferring Drug Interactions) framework was developed to predict DDIs, which used two types of drug interactions( potential CYP (Cytochrome P450)-related

DDIs, and non-CYP-related DDIs (NCRDs)) [19].

By the combination of crizotinib with ketoconazole or rifampin, a PBPK (physiologically based pharmacokinetic) model was developed for predicting DDIs [20]. Based on properties of the drug metabolism, the text-mining and reasoning approaches were also used to discover novel DDIs [21]. Vilar *et al*. computed the molecular fingerprint similarity and the molecular structure similarity of drugs to predict DDIs [22].

With 2D and 3D molecular structures, interaction profiles, target and side-effect similarities, Vilar *et al*. further developed a protocol applicable on a large scale data to infer novel DDIs [23]. Based on drug phenotypic, therapeutic, chemical, and genomic properties and machine learning model, Cheng *et al*. proposed a computational method to predict DDIs [24]. Based on the drug molecular similarity

and phenotypic similarity, Li *et al*. developed a computational method to discover the combination efficacy of drugs with a Bayesian network model [25]. Based on a random forest model, Liu *et al*. proposed a computational method to predict DDIs by integrating chemical interactions, proteinprotein interactions between targets of drugs and target enrichment of KEGG pathways [26]. This method adopted a feature selection technique to obtain the important features of drugs.

Luo *et al*. developed a computational method to predict DDIs by implementing the chemical-protein interactome, which provided as a web server (called DDICPI) [27]. Based on the framework probabilistic soft logic, Sridhar *et al*. took a PSL (Probabilistic Soft Logic) method to predict novel DDIs by integrating networks of multiple drug similarities and known DDs [13]. With 2D structural similarities of drugs, Takako *et al*. developed a logistic regression model to infer potential DDIs [28]. Its prediction performance is further improved by combining targetrelated

and enzyme-related scores.

Based on inner productbased similarity measures (IPSMs), Ferdousi *et al*. provided an computational method to predict DDIs. This method also used the drug similarity constructed with key biological elements including carriers, transporters, enzymes and targets of drugs. In addition, based on the assumption that synergistic effects with drugs are often similar and vice versa, NLLSS (Network-based Laplacian regularized Least Square Synergistic drug combination prediction) was proposed to predict hidden synergistic drug combinations, but it can not predict DDIs for new drugs [29].

**Disadvantages**

* The system is not implemented Regularized least squares classifier.
* In conjunction with security threats, an emerging concern on ML-based solutions is not suitable for drugs test, namely the non ml classifies are very weak of information from the ML models to the adversaries.

Proposed System

In this study, by integrating the chemical, biological and phenotype information of drugs, we develop a computational method (called DDI-IS-SL) to predict DDIs. These drug information includes drug chemical structures, drugtarget interactions, drug enzymes, drug transports, drug pathways, drug indications, drug side effects, drug off side effects and known DDIs.

First, based on these pieces of drug information, we construct a high-dimensional binary vector to calculate the feature similarity of drugs via the cosine similarity method. Furthermore, we also compute the Gaussian Interaction Profile (GIP) kernel similarity [38] of drugs based on known DDIs.

The final drug similarity is constructed by their feature similarity and GIP similarity.

Then a Regularized Least Squares (RLS) classifier [39] is adapted to predict DDIs. For new drugs which do not have any interactions with other drugs, we also calculate their relational initial scores via performing the node-based drug network diffusion method. Therefore, our method can predict potential DDIs not only for known drugs but also for new drugs. The prediction performance of our method and other competing methods are systematically assessed by the 5-fold cross validation, the 10-fold cross validation and the de novo validation. The AUC (area under the ROC curve) is used as the metric to evaluate the performance of computational methods. In terms of AUC, our method is superior to other competing methods.

Specifically, in the 5-fold cross validation, the AUC value of our method is 0.9691, which is larger than the AUC of 0.9570 from the state-of-the-art L1E. Furthermore, in the 10-fold cross validation, the AUC value of our method raches 0.9745, which is also larger than the best result of L1E whose AUC value is 0.9599. Our method also obtain the best prediction performance in the de novo drug validation, its AUC value is 0.9292, which is also larger than the the best result of other methods (WAE (weighted average ensemble method) : 0.9073). In addition, the comparison of the average running time further improves that our method has the higher running

efficiency than other competing methods. Finally, the verification results of case studies also prove the prediction ability of our method in practical applications and show that DDI-IS-SL is an effective computational method to predict

new DDIs.

**Advantages**

* The proposed system implemented many ml classifies for testing and training on datasets.
* The proposed system developed a Regularized least squares classifier to find an accurate accuracy on the datasets.

**SYSTEM REQUIREMENTS**

➢ **H/W System Configuration:-**

➢ Processor - Pentium –IV

➢ RAM - 4 GB (min)

➢ Hard Disk - 20 GB

➢ Key Board - Standard Windows Keyboard

➢ Mouse - Two or Three Button Mouse

➢ Monitor - SVGA

**SOFTWARE REQUIREMENTS:**

* **Operating system :** Windows 7 Ultimate.
* **Coding Language :** Python.
* **Front-End :** Python.
* **Back-End :** Django-ORM
* **Designing :** Html, css, javascript.
* **Data Base :** MySQL (WAMP Server).