International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 07 | July – 2025 DOI: 10.55041/ISJEM04892
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Investigating the Various Methods for Predicting Drug–Drug Interactions Based on Machine Learning Model

M.Arunkumar
Research Scholar,
PG & Research Department of Computer Science,
A. Veeriya Vandayar Memorial Sri Pushpam College (Autonomous), Poondi - 613503, Thanjavur,
E-Mail: arunk145@gmail.com

Dr. T.S. Baskaran
Associate Professor& Research Supervisor,
PG & Research Department of Computer Science,
A Veeriya Vandayar Memorial Sri Pushpam College (Autonomous), Poondi - 613503, Thanjavur,
E-Mail: t_s_baskaran@yahoo.com

"Affiliated to Bharathidasan University, Tiruchirappalli-620024", TamilNadu, India.

Abstract

Drug–drug interactions performance a vigorous role in drug research. However, they may also cause adverse reactions in patients, with serious consequences. Manual detection of drug–drug interactions is time-consuming and expensive, so it is urgent to use computer methods to solve the problem. There are two ways for computers to identify drug interactions: one is to identify known drug interactions, and the other is to predict unknown drug interactions. In this paper, we review the research progress of machine learning in predicting unknown drug interactions. Among these methods, the literature-based method is special because it combines the extraction method of DDI and the prediction method of DDI. We first present the common databases, then briefly describe each method, and summarize the advantages and disadvantages of some prediction models. Finally, we discuss the challenges and prospects of machine learning methods in predicting drug interactions.

Keywords: machine learning, drug-drug interactions, comparation, prediction

INTRODUCTION

Drug–drug interactions (DDI) can occur when two or more drugs are used in combination (Baxter and Preston, 2010). Such interactions may enhance or weaken the efficacy of drugs, cause adverse drug reactions (ADRs) that can even be life-threatening in severe cases (Classen et al., 1997; Agarwal et al., 2020), and cause a drug to be withdrawn from the market (Lazarou et al., 1998). According to the U.S. Centers for Disease Control and Prevention, more than 10% of people take five or more drugs at the same time. Even worse, 20% of older adults take at least 10 drugs (Hohl et al., 2001), which greatly increases the risk of ADR. With an increasing number of approved drugs, the possibility for interactions between drugs increases accordingly (Khori et al., 2011). Therefore, predicting DDI in advance is both urgent and increasingly difficult in clinical practice.

In vivo and *in vitro* experiments can facilitate the identification of DDI, but cannot be performed in some cases

due to laboratory limitations and/or high cost (Safdari et al., 2016). Thus, it is mainly important to develop computational approaches to solve problems of identifying DDI. Current computational approaches to identify DDI can be divided into two categories: 1) extraction of DDI from literature, electronic medical records, and spontaneous reports; 2) use of known DDI to predict unknown DDI.

Extraction of DDI

A large number of DDI are contained in formless articles, but with the explosion of biomedical literature, it has become a huge challenge to identify useful information from the vast literature and coordinate it within drug databases (Rodríguez-Terol et al., 2009; Pathak et al., 2013). Extraction of DDI is achieved by one of two approaches: pattern-based approaches and characteristics- based machine learning. The current pattern-based approach is being phased out because it relies on domain knowledge to manually classify DDI. With the emergence of annotated corpus (Segura et al., 2013), the method of extracting DDI based on machine learning becomes more and more popular. Moreover, extracting DDI from unstructured text data does not provide an early warning or identify unknown DDI, while machine learning can effectively predict it in advance (Kanehisa et al., 2010; Chen et al., 2019; Song et al., 2021).

Prediction of DDI

Only known DDI can be extracted from amorphous articles. However, if the relevant DDI can be predicted in advance before a drug is put onto the market, drugs that cannot be used in combination can be identified. These identified DDI's can prevent many medical errors. We first divide machine learning into old-style and non-outdated categories. In traditional machine learning methods, it is divided into similarity—based method and classification—based method. There are four broad categories of non-traditional machine learning. 1) Network propagation-based approach. The network propagation-based approach can be divided into link prediction and graph embedding according to different methods of network (graph) processing. The link prediction method takes biomedical entities as nodes and their complex interactions as edges to predict unknown relationship interactions and identify false or missing interactions. The method of graph embedding is to transform the known network (graph) into a low-dimensional space through the embedding layer and retain the information of the network (graph). 2) Matrix factorization. The matrix factorization method is to decompose the known drug interaction matrix into N low-dimensional space matrices using different decomposition methods, and then recombine them to obtain the matrix predicting drug interaction. 3) Ensemble-based approaches. The ensemble-based approaches which combine various methods for predicting drug interactions with the goal of achieving better results. 4) Literature-based methods. This approach first uses NLP to extract drug interactions from unstructured data as data sets.

MACHINE LEARNING-BASED APPROACH

Similarity-Based Approach

The basic concept of traditional similarity-based approaches for prediction of DDI is as follows: if drug A and drug B interact with each other to produce a specific effect, then drugs like drug A (or drug B) are likely to produce the same effect with drug B (or drug A). With regard to drug similarity, interactions between new drugs are predicted through the fusion of similar characteristics of multiple drugs (Su et al., 2019a; Zeng et al., 2019; Zeng et al., 2020a; Fu et al., 2020; Mo et al., 2020; Zhuang et al., 2020; Shaker et al., 2021).

multiple types of drugs caused by inhibition of metabolic enzymes, transporters, and even pharmacological targets. To obtain molecular similarity, the authors first collected and processed drug molecules, then represented the resulting molecular structure as a bit vector that encoded the presence or absence of molecular features, where each feature was assigned a specific location. Finally, the calculation and data representation of similarity measurement are presented. Tanimoto coefficient (TC) was used to measure molecular fingerprints. 0 indicates the greatest dissimilarity, and 1 indicates the greatest similarity.

Ferdousi et al. (2017) calculated the similarity of drug pairs using the Rus-Rao approach based on similarity measurements of 12 binary vectors. The greater the similarity, the greater the likelihood of drugs interactions. Pharmacokinetic DDI (Zhang et al., 2009) describes the process by a drug affects the absorption, distribution, metabolism, or excretion of another drug; whereas, pharmacodynamic DDI (Imming et al., 2006) involves the process by which two or more drugs affect the same receptor to cause synergistic or harmful effects. Gottlieb et al. (2012) used a logistic classifier to infer interactions between pharmacodynamics and pharmacokinetics, as well as their severity, by integrating the similarity measurements of seven different drugs and building classification characteristics.

Classification-Based Approach

The traditional classification-based approach involves simulating the DDI prediction task as a binary classification problem. DDI pairs and non-DDI pairs are used to build classification models. For binary classification, known interactions are used as inputs, and other drug pairs may have undetected or unobserved interactions that need to be predicted. In machine learning, similar problems are generally converted to semi-supervised learning tasks (Zhao et al., 2020; Hu et al., 2021a). In the classification task, a model is often built using classifiers such as logistic regression, Bayesian, k-nearest neighbor, random forest, and support vector machines (SVM) to predict DDI.

Li et al. (2015) designed a probability ensemble approach employing a Bayesian network model and similarity algorithm to predict drug pairs from molecular and pharmacological characteristics. Jian-Yu et al. (2016) proposed a new semi- supervised fusion algorithm based on a local classification model and Dempster–Shafer evidence theory. With this approach, new DDI may be predicted based on structural and side-effect similarity (Zhao et al., 2019). Kastrin et al. (2018) treated the process of predicting DDI as a binary

classification task by predicting unknown interactions of randomly selected drugs in five large DDI databases using a link-prediction technique, and enhanced the network topology characteristics using four semantic characteristics.

Similarity based on the traditional method and based on the traditional classification method are obtained to predict the unknown drug interactions between very good results, but in these methods, the characteristics of drugs and drug interactions cannot get a good integration between known, there would be no way to use the known information to fully predict drug interactions. So, we need to develop more efficient computational methods to predict unknown drug interactions.

Ensemble-Based Approach

The ensemble-based approach combines multiple approaches to predict unknown DDI. Zhang P et al. (2015) proposed that the computational burden of multi-label cases may be reduced by selecting appropriate information dimensions based on the mutual characteristics and side effects of drugs. Combined use of genetic algorithms and the multi-label k-nearest neighbor algorithm can define the optimal characteristic size and enables development of prediction models. A novel multi-label K-nearest adjacency method based on function selection (FS- MLKNN) is proposed, which can simultaneously determine key feature sizes and construct high-precision multi-label prediction models. FS-MLKNN takes two steps to establish the relationship between characteristic vectors and side effects. Firstly, information dimensions are selected by mutual information between functional dimensions and side effects to reduce the computational burden of multi-label learning. Then, genetic algorithm (GA) and multi-label K-nearest neighbor point method (MLKNN) were combined to determine the optimal feature size and develop a prediction model.

Zhang et al. (2017) built a prediction model based on various characteristics of drugs and known data about DDI according to neighbor-recommendation, random walk, and matrix disturbance approaches, which use flexible and diverse frameworks to combine different models with different ensemble rules. Deepika and Geetha (2018) predicted DDI through positive-unlabeled (PU) learning (Elkan and Keith, 2008) and meta-learning (Lemke et al., 2015), and proposed a learning framework for semi-supervised classifiers based on SVM. The PU-based classifier was used to generate meta- knowledge from the network, and the meta-classifier was designed to predict the probability of DDI from the generated meta-knowledge.

CONCLUSION

The occurrence of DDI affects the treatment of patients and has become a serious problem for patient safety and drug management. The harm caused by DDI will be greatly reduced if machine learning can be used to efficiently predict DDI. To this end, it is urgent to develop better-performing machine learning approaches. This article describes existing machine learning- based approaches for predicting DDI.

REFERENCES

Agarwal, S., Agarwal, V., Agarwal, M., and Singh, M. (2020). Exosomes: Structure, Biogenesis, Types and Application in Diagnosis and Gene and Drug Delivery. *Curr. Gene Ther.* 20, 195–206. doi:10.2174/1566523220999200731011702

Bach, S. H., Broecheler, M., Huang, B., and Getoor, L. (2015). '*Hinge-loss Markov Random fields and Probabilistic Soft Logic*. arXiv preprint arXiv:1505.04406.

Baxter, K., and Preston, C. L. (2010). *Stockley's Drug Interactions*. Pharmaceutical Press London.

Cai, L., Lu, C., Xu, J., Meng, Y., Wang, P., Fu, X., et al. (2021). Drug Repositioning Based on the Heterogeneous Information Fusion Graph Convolutional Network. *Brief. Bioinformatics* 22 (6), bbab319. doi:10.1093/bib/bbab319

Cai, L., Wang, L., Fu, X., Xia, C., Zeng, X., and Zou, Q. (2020). ITP-pred: an Interpretable Method for Predicting, Therapeutic Peptides with Fused Features Low-Dimension Representation. *Brief. Bioinformatics* 22 (4), bbaa367. doi:10.1093/bib/bbaa367

Cami, A., Manzi, S., Arnold, A., and Reis, B. Y. (2013). Pharmacointeraction Network Models Predict Unknown Drug-Drug Interactions. *PLoS One* 8, e61468. doi:10.1371/journal.pone.0061468

Chen, X., Shi, W., and Deng, L. (2019). Prediction of Disease Comorbidity Using HeteSim Scores Based on Multiple Heterogeneous Networks. *Curr. Gene Ther.* 19, 232–241. doi:10.2174/1566523219666190917155959

Cheng, L., Hu, Y., Sun, J., Zhou, M., and Jiang, Q. (2018). DincRNA: a Comprehensive Web-Based Bioinformatics Toolkit for Exploring Disease Associations and ncRNA Function. *Bioinformatics* 34, 1953–1956. doi:10.1093/bioinformatics/bty002

Classen, D. C., PestotnikEvans, S. L. R. S., Evans, R. S., Lloyd, J. F., and Burke, J. P. (1997). Adverse Drug Events in Hospitalized Patients. Excess Length of Stay, Extra Costs, and Attributable Mortality. *Jama* 277, 301–306. doi:10.1001/ jama.1997.03540280039031

Deepika, S. S., and Geetha, T. V. (2018). A Meta-Learning Framework Using Representation Learning to Predict Drug-Drug Interaction. *J. Biomed. Inform.* 84, 136–147. doi:10.1016/j.jbi.2018.06.015