Towards Interpretable Adverse Drug Reaction Prediction Using Deep Graph Fusion

Pengwei Hu^a, Tiantian He^b, Zhaomeng Niu^c, Jing Mei^a, Bibo Hao^a, Shaochun Li^a

^aIBM Research, Beijing, China ^bNanyang Technological University, Singapore ^cRutgers Cancer Institute of New Jersey, New Brunswick, USA

Abstract

Adverse drug reaction (ADR) is the unexpected harmful reaction of patients after medication, which can increase the burden of the hospital, restrict the benefit and cause death. To better predict ADRs, we propose a machine learning method based on multi-network fusion called DF-ADRs. First, we used a correlation degree to determine the multitype drug features of each drug and the co-occurrence of each ADR and represent each ADR with these important features. On this basis, multiple similarity networks are constructed by introducing multitype features of drugs as new representations of these chemical substructures. Then, the network fusion algorithm is used to construct an attribute graph for each ADR. Finally, we used the deep network to obtain a set of interpretable subgraphs and used the classifier to predict the ADR of drugs. Extensive experiments using real-world data show that DF-ADRs can be potentially used to enable the clinical team to optimize the stimulation parameters.

Introduction

Adverse drug reactions (ADRs) is one of the main issues that doctors consider when using drugs, and it can cause patients' discomfort and even death (Edwards and Aronson 2000; Murphy 2011). ADR, unlike dose-induced toxicity, is usually caused by the properties of the structure of the drug and the misses of the compound. Hospitals often bear the financial cost of adverse events, and pharmaceutical companies must invest resources in screening a large number of candidate compounds to avoid potential side effects of new drugs. Statistical results show that over 40% of ADRs can be avoided in advance. This has inspired researchers to introduce machine learning techniques to improve drug safety, especially the detection of adverse drug reactions, which has become a common practice. Computational methods have been used to capture the relationship between adverse reactions and drug molecules hidden under a large number of clues, but they have not been able to pinpoint the exact small molecules that cause the adverse reactions (Page et al. 2012; Liu et al. 2012). For the industry, practical algorithms should be able to detect important molecular patterns and explain which are the small molecules that

cause ADRs. In response to this demand, a series of interpretable machine learning algorithms have emerged, all of which are innovative in identifying potential links between drug substructures and ADRs(Campillos et al. 2008; Xiao et al. 2017; Pauwels, Stoven, and Yamanishi 2011; Hu et al. 2017; Atias and Sharan 2011; Zhang et al. 2013; Wang et al. 2014b).Interpretation and accuracy are not always compatible, so these methods cannot achieve high precision. The problem is that they ignore one of the properties of drugs, which is that the basic unit of action is the compound, not the individual molecule. Therefore, the extraction of a set of interpretable models from primitive molecules to explain the predicted results is urgently needed by pharmaceutical enterprises(Bichindaritz 2008). In this paper, a new method, the deep network fusion for interpretable adverse drug reaction prediction (DF-ADRs), is proposed, which means that each ADR is required to learn the important subgraphs of its associated molecular graph, so as to enable these subgraphs to be used for ADR prediction. To achieve this, DF-ADRs first obtains richer and more meaningful drug substructure expression forms by calculating the drug features of various types and the correlation between each ADR. Then, a molecular graph was established for each new drug substructure expression, and an attempt was made to fuse the above multiple networks by introducing a network fusion model. The fused graph is sent to the deep learning model to cluster the sub-graphs. We believe that these subgraphs contain many molecular substructures. Once these potential subgraphs are identified, we use the classifier to predict ADRs in candidate drug molecules. The subplots are small molecular structures associated with ADRs that are strongly interpretable. In figure 1 below, we summarize the framework of DF-ADRs. To the best of our knowledge, there have been no studies that have constructed subgraphs of drug molecules from multiple types of data sources and used them to predict ADRs.

Methods

Extraction of Multi-type ADR Related Features

Each drug has more than 1,000 potential ADRs, so we trained the prediction model separately for each ADR. For each drug, there are multiple types of feature expression, and we tried to define the correlation with ADR for each type of

Copyright © 2019, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.



Figure 1: The flowchart of the DF-ADRs pipeline.

feature. Here, let's use the chemical substructures as an example. Suppose we have N drug samples, R substructures for each drug, S ADRs and M measurements (for example, substructures, pathways, treatments). In introducing our approach, we will use the drug substructure as an example to facilitate generalization to multiple expression types. In the case of drugs, to understand the relationship between drug substructure and side effects, the goal was to determine the connection strength between (df_i) and S ADRs (ar_j) . To do this, we first obtain the observed frequency (o (df_i, ar_i)) for each ADR caused by all drugs and the expected frequency $(e(df_i, ar_i))$, and then calculate the difference. In this case, the measure of whether df_i causes ar_i is difference scale. From this, we can infer that drugs containing df_i may be related to ADR ar_i . Here we introduce a statistical method that has been effectively applied to calculate the correlation between ADR and drug substructure (Ching, Wong, and Chan 1995; Hu et al. 2017), which is defined as follows:

 $e(df_i, ar_j) = \frac{df_{i+} * ar_{+j}}{T} \tag{1}$

where

$$df_{i+} = \sum_{j=1}^{5} df_{ij} \tag{2}$$

$$ar_{+j} = \sum_{i=1}^{R} ar_{ij} \tag{3}$$

$$T = \sum_{i, j} o(df_i, ar_j) \tag{4}$$

Using the above results, we calculated the relationship between drug substructure and ADR using the following measurement standard R:

$$R(df_i, ar_j) = \frac{Z_{ij}}{\sqrt{\left(1 - \frac{df_{i+}}{T}\right)\left(1 - \frac{ar_{+j}}{T}\right)}}$$
(5)

where $\left(1 - \frac{df_{i+}}{T}\right)\left(1 - \frac{ar_{+j}}{T}\right)$ is used to adjust likelihood of $Z(df_i, ar_j)$.

$$Z(df_i, ar_j) = \frac{o(df_i, ar_j) - e(df_i, ar_j)}{\sqrt{e(df_i, ar_j)}}$$
(6)

As a hyperparameter, R can be predefined as a value, and when greater than this value, the correlation between df_i and ar_j can be determined. The measure of R approximately follows the standard normal distribution. This step can be used to identify a drug substructure that have a significant relationship with ADR. For the other feature types, we will do the same to get the desired feature. In other words, through the above steps, we can respectively obtain the features related to ADR in enzyme, pathway, ADR in target and treatments respectively.

Molecular Graph Representation

The basic structure of drugs is the chemical substructure, and the interaction between drugs and targets or pathways is caused by small chemical molecules. Therefore, explainable ADRs predictions need to find out what kind of molecular structure composition will lead to ADR. Here, the more abundant the expression of the substructure of the drug, the better the interpretability and performance of the prediction. Thanks to the availability of multiple drug features, we can form multiple drug substructure expression such as substructures-treatments. For this purpose, each kind of features can calculate a network expression for the drug substructure. These ADR related features can be used to enrich the expression of drug molecular structure in the next step.

Deep Fusion Model

For molecular graph, diverse profiles may have different contributions in network learning. We cross-referenced several different databases and obtained several drug expression information. We use this information to generate the substructure similarity network of drug substructures. The similarity measurement method also has an important influence on network learning, so we introduce Jaccard similarity to calculate the similarity network based on the above different expression profile. Biochemical mechanisms, including Chemical Substructures, Target Proteins, Transporters, Enzymes and Pathways expression exists between different types of features, such as inner intersections. Our goal is to make full use of the above features to calculate the most complete edges, so as to improve the accuracy of the prediction model. We believe that multiple sources of information can promote the integrity of the expression of drug substructures, so we propose the concept of network integration. Unlike common feature fusion, network fusion seeks to produce the most similar representation of all networks in the existing domain. After multiple expressions form multidrug molecular structure similar networks, we introduce the iterative network fusion method which is suitable for synthesizing multi-biochemical networks (Wang et al. 2014a). In this method, the first step of network fusion is to express the full kernel on the vertex set V as a normalized weight matrix $P = D^{-1}W$. W is a symmetric matrix to describe the similarity between two drugs. D is a diagonal matrix that entries $D(i, i) = \sum_{j} W(i, j)$, and $\sum_{j} P(i, j) = 1$. Next, we assume that the diagonal entries of W are free, that is, a better normalization has been achieved by discarding selfsimilarity, and keep $\sum_{i} P(i, j) = 1$:

$$K(i,j) = \frac{W(i, j)}{2\sum_{k \neq i} W(i, k)}$$
(7)

subject to the constraints: $i \neq j$, otherwise $K(i, j) = \frac{1}{2}$. Let N_i represent a set of v_i 's neighbors including v_i in G. We assume that local affinity Q can be obtained by using K nearest neighbors (KNN) through the following methods:

$$Q(i,j) = \frac{W(i,j)}{\sum_{k \in N_i} W(i,k)}$$
(8)

subject to the constraints: $j \in N_i$, otherwise Q(i, j) = 0 For multiple similarity networks, fusion is the iterative updating of these matrices:

$$K^{(v)} = Q^{(v)} \times \left(\frac{\sum_{k \neq v} K^{(k)}}{m-1}\right) \times (Q^{(v)})^T, \ v = 1, 2, 3, \dots, m$$
(9)

where $K_{t=0}^{(v)}$ said initial v state matrix, and $Q_{t=0}^{(v)}$ is the kernel matrix. The main program of network fusion is to update the state matrix and generate v parallel exchange diffusion processes. The final status needs to be updated t times to obtain:

$$K^{(v)} = \frac{\sum_{v=1}^{m} K^{v}}{m}$$
(10)

Following the above method execution, the generated fusion graph can be defined as $G_f = (V_f, E_f)$. Vertex set $V_f = v_{f1}, \ldots, v_{fn}$ included all chemical substructures. For each pair of vertices, there is a fused similarity s_{ij} between two chemical substructures in the similarity matrix S. It is our important goal to decompose ADR related subgraphs from the new fused graphs. Stacked auto encoder(Bengio, Courville, and Vincent 2013; Tian et al. 2014) and spectral clustering are similar, it can eliminate edges between two nodes while learning nonlinear embedding. Each node in the fused graph will be normalized and fed to the Stacked Auto-Encoder. Then the training model is optimized to minimize the error between output and input. And then we apply C-Means(Gu et al. 2017) to extract C molecular subgraphs. By the above steps we can recreate the feature expression of each drug, which is expressed in the form of one-hot more than C molecular subgraphs, 0 for does not exist a subgraph in drug molecular structure, and 1 means there is. Next, we can introduce any interpretable classifier to complete the final ADR prediction. Specifically, suppose we have N drug samples and C molecular subgraphs (integration from substructures, enzyme, pathways, target and treatments), we will try to model new integration (if_i) features and label ADR (ar_i) . Inspired by the previous work(Hu et al. 2017), we used Naive Bayes as the classifier to ensure the interpretability of the results and to contrast the effect of the new molecular subgraphs.

Results

Evaluation Data Set and Metrics

In this study, the Liu's dataset and Mizutani's dataset which we used to predict ADRs are obtained from(Liu et al. 2012; Mizutani et al. 2012). They have been widely used to test the effectiveness of ADRs prediction. Liu's dataset builds a dataset containing 832 drugs and 1385 kinds of ADR and include six kinds of drug feature: Chemical Substructures, Target Proteins, Transporters, Enzymes, Pathways, Treatment indications. Mizutani's dataset builds a dataset containing 658 drugs and 1339 kinds of ADR and include two kinds of drug features: Chemical Substructures and Target Proteins. Liu's dataset has 881-dimensional binary chemical substructures to represent drugs which collected from PubChem Compound Database (Chen, Wild, and Guha 2009), Drug-Bank (Law et al. 2013). The data of drug ADRs come from SIDER(Kuhn et al. 2010) which collects ADRs data from FDA Adverse Event Reporting System (FAERS). In addition, this dataset also includes 786 target proteins, 72 transporters, 111 enzymes, 173 pathways and 869 treatment indications which collected from KEGG DRUG(Kanehisa et al. 2009), DrugBank, KEGG, SIDER respectively. Mizutani's dataset also has 881-dimensional binary chemical substructures which collected from PubChem Compound Database and DrugBank. This dataset also includes 1368 target proteins collected from KEGG. Even the most common ADRs are extremely rare because every drug on the market has gone through rigorous clinical trials. Regarding imbalance prediction performance evaluation, we introduce AUC (area under the ROC curve) to evaluate our model performance as the prediction performance can be safely unbiased. We infer interactions and compare against the held-out interactions, measuring performance using the AUC for our evaluations. To further evaluate the performance of the proposed method, we also use the several more measure like following criteria: the overall prediction accuracy, recall, precision, and F-measure were calculated.

Result Comparison

To evaluate the performance of DF-ADRs, we also made use of current interpretable approaches which include the GraphSE(Hu et al. 2017), Naive Bayes, OCCA(Pauwels, Stoven, and Yamanishi 2011), SCCA(Pauwels, Stoven, and Yamanishi 2011) and LDA(Xiao et al. 2017) to predict ADRs. To further evaluate the performance of DF-ADRs, we adopted 5-fold cross validation. Table 1 reports the AUC scores, F-measures, ACC, Precision and Recall of different algorithms on the same datasets. This result show that DF-ADRs extracted more meaningful subgraphs to drug substructures from fused network and this approach is shown here to have the potential to improve prediction performance.

Interpretable Results

To find out the relationship between the drug ADR and the multi-type features include enzyme, pathways, target, and treatments, we take a significance analysis first. In this step, we will obtain various types of features that have a statistical correlation with ADR. These features can be considered as single factors affecting the generation of ADR. We will use the features received from this step as new features of each information type. Then, we can use them to calculate multiple graph representations in terms of drug chemical

Methods	AUC		ACC		E-measures		Precision		Recall	
memous	Liu	Mizutani	Liu	Mizutani	Liu	Mizutani	Liu	Mizutani	Liu	Mizutani
DF-ADRs	0.906	0.902	0.917	0.921	0.467	0.469	0.413	0.398	0.622	0.626
GraphSE	0.883	0.881	0.892	0.898	0.445	0.439	0.389	0.382	0.597	0.595
NB	0.879	0.85	0.881	0.886	0.409	0.405	0.352	0.367	0.581	0.580
OCCA	0.841	0.852	0.856	0.851	0.386	0.383	0.339	0.335	0.557	0.561
SCCA	0.860	0.870	0.877	0.873	0.396	0.393	0.361	0.359	0.578	0.572
LDA	0.844	0.850	0.854	0.855	0.389	0.397	0.357	0.363	0.562	0.553

Table 1: Metric scores of different methods evaluated by 5-fold cross-validation



Figure 2: The information used for this visualization is based on SIDER database and represents some of the ADRs that commonly caused by same drug targets. Different color represents share different targets and the different drug be assigned to same color according to same targets.

structure. An example of the relative relationship between the drug targets and ADRs is given in Fig.2. This visual figure shows the human target that drug use is the same and so there are laws connect with same ADRs but the correlations is sparse and small range gathered. Meanwhile, the output of deep fusion model can also be interpreted, and the correlation between single factor and ADR can be extended to the correlation between groups composed of multiple chemical molecules and ADRs. Obviously, the above example shows the explainable ability for the ADRs prediction of DF-ADRs.

Conclusion

In this paper, an approach, DF-ADRs that can be used for network fusion for interpretable prediction of ADRs is proposed. DF-ADRs addresses several key challenges in interpretable ADRs associated patterns and multiple molecular network fusion. First, DF-ADRs takes the smallest computable substructure of the drug as the node to construct the graph. Secondly, it makes full use of heterogeneous information embedded in the expression of drug substructure. Additionally, the interpretable classifier is friendly enough to use the newly generated subgraphs as input features to construct a classification model. Experimental results on the real-world dataset demonstrate that DF-ADRs is able to achieves a good performance. DF-ADRs is an interpretable model with potential applications. Its interpretability can improve the confidence of doctors in the use of drugs and provide researchers with a reliable basis in drug development.

References

Atias, N., and Sharan, R. 2011. An algorithmic framework for predicting side effects of drugs. *Journal of Computational Biology* 18(3):207–218.

Bengio, Y.; Courville, A.; and Vincent, P. 2013. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence* 35(8):1798–1828.

Bichindaritz, I. 2008. Case-based reasoning in the health sciences: Why it matters for the health sciences and for cbr. In *European Conference on Case-Based Reasoning*, 1–17. Springer.

Campillos, M.; Kuhn, M.; Gavin, A.-C.; Jensen, L. J.; and Bork, P. 2008. Drug target identification using side-effect similarity. *Science* 321(5886):263–266.

Chen, B.; Wild, D.; and Guha, R. 2009. Pubchem as a source of polypharmacology. *Journal of chemical information and modeling* 49(9):2044–2055.

Ching, J. Y.; Wong, A. K. C.; and Chan, K. C. C. 1995. Class-dependent discretization for inductive learning from continuous and mixed-mode data. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 17(7):641–651.

Edwards, I. R., and Aronson, J. K. 2000. Adverse drug reactions: definitions, diagnosis, and management. *The lancet* 356(9237):1255–1259.

Gu, J.; Jiao, L.; Yang, S.; and Zhao, J. 2017. Sparse learning based fuzzy c-means clustering. *Knowledge-Based Systems* 119:113–125.

Hu, P.; Chan, K. C.; Hu, L.; and Leung, H. 2017. Discovering second-order sub-structure associations in drug molecules for side-effect prediction. In 2017 IEEE In-

ternational Conference on Bioinformatics and Biomedicine (BIBM), 2250–2253. IEEE.

Kanehisa, M.; Goto, S.; Furumichi, M.; Tanabe, M.; and Hirakawa, M. 2009. Kegg for representation and analysis of molecular networks involving diseases and drugs. *Nucleic acids research* 38(suppl_1):D355–D360.

Kuhn, M.; Campillos, M.; Letunic, I.; Jensen, L. J.; and Bork, P. 2010. A side effect resource to capture phenotypic effects of drugs. *Molecular systems biology* 6(1).

Law, V.; Knox, C.; Djoumbou, Y.; Jewison, T.; Guo, A. C.; Liu, Y.; Maciejewski, A.; Arndt, D.; Wilson, M.; Neveu, V.; et al. 2013. Drugbank 4.0: shedding new light on drug metabolism. *Nucleic acids research* 42(D1):D1091–D1097.

Liu, M.; Wu, Y.; Chen, Y.; Sun, J.; Zhao, Z.; Chen, X.-w.; Matheny, M. E.; and Xu, H. 2012. Large-scale prediction of adverse drug reactions using chemical, biological, and phenotypic properties of drugs. *Journal of the American Medical Informatics Association* 19(e1):e28–e35.

Mizutani, S.; Pauwels, E.; Stoven, V.; Goto, S.; and Yamanishi, Y. 2012. Relating drug–protein interaction network with drug side effects. *Bioinformatics* 28(18):i522–i528.

Murphy, R. F. 2011. An active role for machine learning in drug development. *Nature chemical biology* 7(6):327.

Page, D.; Costa, V. S.; Natarajan, S.; Barnard, A.; Peissig, P.; and Caldwell, M. 2012. Identifying adverse drug events by relational learning. In *Twenty-Sixth AAAI Conference on Artificial Intelligence*.

Pauwels, E.; Stoven, V.; and Yamanishi, Y. 2011. Predicting drug side-effect profiles: a chemical fragment-based approach. *BMC bioinformatics* 12(1):169.

Tian, F.; Gao, B.; Cui, Q.; Chen, E.; and Liu, T.-Y. 2014. Learning deep representations for graph clustering. In *Twenty-Eighth AAAI Conference on Artificial Intelligence*.

Wang, B.; Mezlini, A. M.; Demir, F.; Fiume, M.; Tu, Z.; Brudno, M.; Haibe-Kains, B.; and Goldenberg, A. 2014a. Similarity network fusion for aggregating data types on a genomic scale. *Nature methods* 11(3):333.

Wang, F.; Zhang, P.; Cao, N.; Hu, J.; and Sorrentino, R. 2014b. Exploring the associations between drug side-effects and therapeutic indications. *Journal of biomedical informatics* 51:15–23.

Xiao, C.; Zhang, P.; Chaovalitwongse, W. A.; Hu, J.; and Wang, F. 2017. Adverse drug reaction prediction with symbolic latent dirichlet allocation. In *Thirty-First AAAI Conference on Artificial Intelligence*.

Zhang, P.; Wang, F.; Hu, J.; and Sorrentino, R. 2013. Exploring the relationship between drug side-effects and therapeutic indications. In *AMIA annual symposium proceedings*, volume 2013, 1568. American Medical Informatics Association.