# Predicting Adverse Drug-Drug Interactions via Semi-supervised Variational Autoencoders

Meihao Hou[1,2], Fan Yang[3,4], Lizhen Cui[1,2(✉)], and Wei Guo[1,2]

[1] School of Software, Shandong University, Jinan, China
hmhhhg@163.com
[2] Joint SDU-NTU Centre for Artificial Intelligence Research(C-FAIR), Shandong University, Jinan, China
{clz,guowei}@sdu.edu.cn
[3] School of Public Health, Cheeloo College of Medicine, Shandong University, Jinan, Shandong, China
fanyang983@gmail.com
[4] Institute for Medical Dataology, Shandong University, Jinan, Shandong, China

**Abstract.** Adverse Drug-Drug Interactions (DDIs) are a very important risk factor in the medical process, which may lead to readmission or death. Although a part of DDIs can be obtained through *in vitro* or *in vivo* experiments in the drug development stage, a large number of new DDIs still appear after the market, more and more researchers begin to pay attention to the research related to drug molecules, such as drug discovery, drug target prediction, DDIs prediction, etc. In recent years, many computational methods for predicting DDIs have been proposed. However, most of them only used labeled data and neglect a lot of information hidden in unlabeled data. Moreover, they always focus on binary prediction instead of multiclass prediction, although the exact DDI type is very helpful for our reasonable choice of medication. In this paper, a Semi-Surpervised Variational Autoencoders (SPRAT) method for predicting DDIs is proposed, which is composed of a neural network classifier and a Variational autoencoders (VAE). Classifier is the core components, VAE plays a role of calibration. In the end, the predicted label is a multi-hot vector which indicates specific DDI types between drug pairs. Finally, the experiments on real world dataset demonstrate the effectiveness of the proposed method in this paper.

**Keywords:** Drug-drug interaction · Semi-surpervised learning · Variational autoencoders · Prediction

## 1 Introduction

Drug-Drug interactions usually occur when patients take the combination drugs, because a drug can affect the activity of another drug in the body [9], which is likely to cause serious incidence rate and mortality [11]. Although some adverse


X. Wang et al. (Eds.): APWeb-WAIM 2020, LNCS 12318, pp. 132–140, 2020.
https://doi.org/10.1007/978-3-030-60290-1_10

DDIs have been screened out by *in vivo* and *in vitro* experiments during the drug development stage, many new DDIs have been found accidentally after the drug was put on the market [8]. With the abundance of drug-related molecular data and adverse event reporting data, such as Drugbank and FDA Adverse Event Reporting System (FAERS), a large number of researchers have focused their attention on the study of adverse DDIs.

Although there are many computational methods to predict potential DDIs, such as similarity based [5,7,14], classification based [3,10] and network-based [15,16], there are still some problems that are not highlighted.The first problem is that many works regard DDI prediction as binary prediction rather than multiclass prediction.However, it is very important to predict the specific types of DDI for patient's medication [13]. Another problem is that most of the previous methods are supervised learning, which depends on adequate labeled data. Many existing databases contain a lot of known DDI, but compared with the whole search space, it is still insufficient. Therefore, the supervised method will be affected by over fitting, which will reduce the prediction accuracy. Similarly, when unsupervised learning is used, the model will be a two-step rather than an end-to-end learning model, which will also greatly affect the prediction performance.

In order to overcome the above two problems, this paper focuses on the prediction of specific adverse drug reactions, and makes better use of data resources to improve the prediction performance. Therefore, we develope a Semi-Surpervised Variational Autoencoders (SPRAT) method, which actually can be seen as an ensemble model composed of a deep network classifier and a Variational Autoencoders (VAE). Discrimination classifier is the core module, which is calibrated by VAE. When the input is unlabeled data, the optimization of VAE reconstruction error is helpful to get a better classifier, so as to get more accurate ADRs prediction results. Finally, the experimental results using real datasets demonstrate our SPRAT can get a better ADRs predictive performance than other baseline methods.

## 2 Preliminaries

In order to describe the proposed method, in this section, some notations will be briefly defined, and introduce a main construction of the model: variational autoencodesr.

**Definition 1.** *Set of Labeled Drug Pairs and Set of Unlabeled Drug pairs*

In order to make better use of the information in the data set, we use the known DDI as the tag data, and the other DDI as the unmarked data. Therefore, set $D_L$ is the set of labeled drug pairs and $D_U$ is the set of unlabeled drug pairs.

**Definition 2.** *DDI Data*

The DDI data includes $n$ drugs, $w$ pairs of drug interaction and $v$ types of adverse drug reactions. We denote $D = \{d_1, d_2, ..., d_n\}$ as the set of drugs, $I = \{i_1, i_2, .., i_w\}$ as known DDI drug pairs and $R = \{r_1, r_2, ..., r_v\}$ as all types

of adverse drug reaction between drug pairs. Therefore, we can denote labeled data $D_L$ described in the last section as $D_L = \{(d_p, r, d_q)|d_p, d_q \in D, 0 < p, q < n, r \subset R\}$.

**Definition 3.** *Drug Side Effects*

Each drug has its own side effects, except for adverse drug-drug interactions. Define all side effects exist in our DDI data as $S$, $S = \{s_1, s_2, ..., s_k\}$. Then, we denote $s(s \subset S)$ as owned side effects of one drug.

**Variational Autoencoders**

VAE is a series of models in which the input data is transformed into a coding vector, and each dimension represents some learning attributes about the data and decoder network, and then obtains these values and attempts to reconstruct the original input. It is worth noting that original autoencoders output a single value for each encoding dimension in latent representation, whereas a variational autoencoder provides a probability distribution for each latent attribute. So that it can serve as a generative model as Generative Adversarial Networks (GAN) do.

## 3 Method

### 3.1 Overview of Our Method

The proposed method consists of three parts, as shown in Fig. 1: classifier, encoder and decoder. Moreover, the encoder and the decoder together form a VAE. In brief, classifier can be seen core component in our model with VAE as a calibrator to improve performance of the classifier. Specifically, the VAE encode its input into latent representation $z$, if we join a label $y$ into $z$, then decoder strive to reconstruct the input based on $z$ and $y$. As in conditional VAE, a more accurate $y$ can be more informative and helpful for the reconstruction process. Similarly the predicted label $y'$ of our classifier will do so as to. In this way, the classifier will be benefited when the input is unlabeled data, while we optimizing the reconstruction loss of VAE, which is exactly what we want.

**Adverse Drug-Drug Interaction Prediction Task** The task is to predict unknown new adverse reactions between drug pairs. The following part will specifically introduce the model architecture centered on our prediction task. First, we gain labeled data $(x, y)$ from $D_L$ and unlabeled data $x$ from $D_U$. Then, we input the concatenated feature vector representation $[x_i, x_j]$ of the drug pairs $(d_i, d_j)$ to classifier and encoder, whether it is labeled or unlabeled. In the next step, the training of the model will be divided into two situations. The first is that the data is labeled, we directly incorporate the label $y$ to latent variable $z$. In that case, both classifier loss and VAE loss need to be optimized. The other is the data is unlabeled, the predicted label $y'$ from classifier is added to latent representation $z$ and we just require to reduce the loss of VAE. At last, the overall loss of the model is the sum of the two parts, a multiple hot-label $y'$ representing adverse drug reactions is obtained.

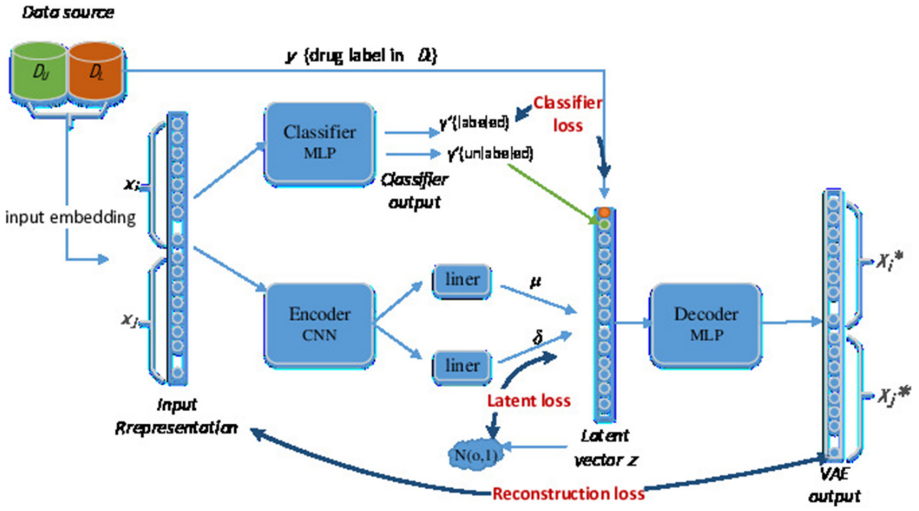


**Fig. 1.** The overview of our SPRAT model for DDIs prediction

### 3.2 Conditional VAE Model

Conditional variational autoencoders (CAVE) refer to adding label $y$ to the model to aid in the generative module. CVAE involves a series of models instead of one, depending on how the label $y$ is added. For this model we have two cases to consider, in the first case $y$ is given directly, in the other y can be calculated by a discriminative model. Accurately, our method is based on the second situation exactly for this knowledge can help us form a better classifier and predict the labels of unseen data.

In the first case, the variational lower bound is similar to Eq. (1). In the second case, assume $q(y|x)$ represents the predicted layel via discriminative network. Therefore, the variational lower bound is shown in the following Eq. (2).

$$\begin{aligned} \log(p(x,y)) &\geq E_{q(z|x,y)} \log(p(x|y,z)) - \mathrm{KL}(q(z|x,y)||p(z)) + \log p(y) \\ &= -\mathcal{L}(x) \end{aligned} \tag{1}$$

$$\begin{aligned} \log p(x) &\geqslant E_{q(y,z|x)}[\log p(x|y,z) + \log p(y) + \log p(z) - \log(y,z|x)] \\ &= \textstyle\sum_y q(y|x)(-\mathcal{L}(x,y)) + \mathcal{H}(q(y|x)) \\ &= -\mathcal{U}(x) \end{aligned} \tag{2}$$

**Loss Function**

The training process for the model is described in detail below.

**Goal** Since the model is composed of a classifier network and a VAE, and the data can be divided into labeled and unlabeled, our training object is to optimize the following three losses:

$$\mathcal{J} = \mathcal{L}(x)_{(x,y)\subset D_L} + \mathcal{U}(x)_{(x)\subset D_U} + \mathrm{Classifier\_loss}_{(x,y)\subset D_L} \tag{3}$$

The input mentioned above for our classifier and encoder is the feature vector of ($d_i$,$d_j$). Next, there are two situations. The first is when input comes from $D_L$, $y$ is an exact vector given by classifier. Thus, the optimization of VAE is just concatenating $y$ to the latent variable $z$ on the original basis. The specific calculation formula of loss during training is shown in Eq. (4).

$$\begin{aligned} &\mathcal{L}(x)_{(x,y)\subset D_L} \\ &= \min\{\mathcal{D}(x*,x)^2 + \mathrm{KL}(q(z|x,y)||p(z))\} \end{aligned} \tag{4}$$

$\mathcal{D}(x*,x)^2$ represents reconstruction loss which can be obtained directly with machine learning tools. Moreover, $q(z|x,y)$ generally selects normal gaussian distribution and $p(z)$ is unit gaussian distribution. Finally, formula to be optimized is shown in Eq. (5)

$$\begin{aligned} &\mathrm{KL}(q(z|x,y)||p(z)) \\ &= \mathrm{KL}(N(\mu,\delta^2)||N(0,1)) \\ &= \frac{1}{2}(\mu^2 + \delta^2 - \log\delta^2 - 1) \end{aligned} \tag{5}$$

Another situation is, $y$ is output by classifier when input comes form $D_U$. We assume that the predicted label $y'$ is added to only affect the mean of the distribution without affecting the variance. As a result, the loss function of this situation can be expressed as Eq. (6)

$$\begin{aligned} &\mathcal{U}(x)_{(x)\subset D_U} \\ &= \mathcal{D}(x*,x)^2 + \frac{1}{2}(\left(\mu - \mu^Y\right)^2 + \delta^2 - \log\delta^2 - 1) \end{aligned} \tag{6}$$

## 4 Experiments

This section briefly describes the datasets and experimental settings used. At the end, the experimental results on these datasets are explained.

### 4.1 Datasets

**DDI Data**
The FDA Adverse Event Reporting System (FAERS) is a database designed to support FDA post-marketing surveillance programs for pharmaceuticals and therapeutic biologics, including all adverse event information and medication error information collected by FDA. As a sub-database extracted from the FAERS, TWOSIDES [14] contains all types of adverse reactions between reported and confirmed drug pairs.

**Drug Feature Information**
Like TWOSIDES, OFFSIDES [14] is also a sub-database extracted from FAERS, which includes the drug side effects information about drugs. According to [1], it can be used as a phenotypic feature information to improve our performance

of prediction. In addition, the drug structure fingerprint can be received from PubChem [6], where the drug is represented as an 881 dimensional binary vector, and each bit represents whether there is a chemical substructure.
As described in the previous datasets, we eventually have 635 drugs, 63,055 distinct drug pairs and each drug own a 5,394-dimensional binary feature vector (4,513 side effect types and 881-dimensional fingerprint). We construct a DDI matrix used to query whether a group of drug pairs is labeled data, where 1 indicating that the two drugs interact and 0 means that the two drugs do not interact as far as we know, by the way, the diagonal elements are all set to 0.

### 4.2 Evaluations

In order to optimize the prediction results, cross validation is designed as an effective method to estimate the generalization error. First, we randomly assigned 10% of the drugs and all DDIs associated with these drugs to be tested. The remaining 90% were then divided into K groups (generally equal). Each group is verified once, and the rest k-1 subset is used as training set. Therefore, under the condition of k-CV, the average value of classification accuracy of the final verification set of K model is used as the performance index of the classifier. To increase the randomness, repeat the K-CV process 50 times. In addition, because of the imbalance of DDIs data, that is, the known positive samples are much smaller than the unknown samples, so the area under the precise recall curve (AUPR) [2] should be used as a measure of model performance.

### 4.3 Experimental Setup

We construct a DDI matrix used to query whether a group of drug pairs is labeled data, where 1 indicating that the two drugs interact and 0 means that the two drugs do not interact as far as we know, by the way, the diagonal elements are all set to 0. From a macro point of view, our SPRAT method is composed of a classifier and a VAE. From a macro point of view, our SPRAT method is composed of a classifier and a VAE. But in more detail, it contains three components across which the classifier and the decoder of VAE is realized with general deep network, the decoder of VAE is implemented by CNN.

The proposed method is compared with other methods which are completely different but perform well in DDIs prediction: Concatenated drug features, Dyadic prediction [4] and LoNAGE graph embedding [12].

### 4.4 Experimental Results

**Experiment I**: *Binary prediction*
First of all, in order to verify the advantages of the proposed model in ADRs prediction, the comparative method is applied to binary prediction, that is, only to find out whether two drugs interact, rather than to predict specific drug types. Table 1 shows the results w.r.t AUROC score and AUPR score. From the results we can

draw a conclusion our method performs better than other alternative approaches. Simultaneously, we find that deep learning based approaches (LoNAGE and our model) perform much better that general machine learning methods because of its powerful non-linear learning ability. Furthermore, AUROC score improves while AUPR score reduces when the K value of Cross-validation is increased.

**Table 1.** Comparison with Three State-of-the-Art Approaches by 50 Runs of 5-CV and 10-CV

| Evaluation | Method | AUROC | AUPR |
|---|---|---|---|
| 5-CV | Concatenated drug features | 0.793 | 0.456 |
| | Dyadic prediction | 0.705 | 0.375 |
| | LoNAGE graph embedding | 0.857 | 0.482 |
| | SPRAT(our model) | 0.876 | 0.513 |
| 10-CV | Concatenated drug features | 0.798 | 0.429 |
| | Dyadic prediction | 0.739 | 0.331 |
| | LoNAGE graph embedding | 0.883 | 0.447 |
| | SPRAT(our model) | 0.893 | 0.486 |

**Experiment II**: *Multi-class prediction*
Most importantly, we need to prove the effectiveness of our method in multiple ADRs prediction. In the LoNAGE method, latent representations output by encoder are input into a SVM for prediction label y. As shown in Fig. 2, compared with binary prediction, the multi-class prediction performance of all methods is reduced, which reveals that comprehensive prediction is more difficult than binary prediction. Despite this, our approach is still better than others in terms of both AUROC and AUPR.

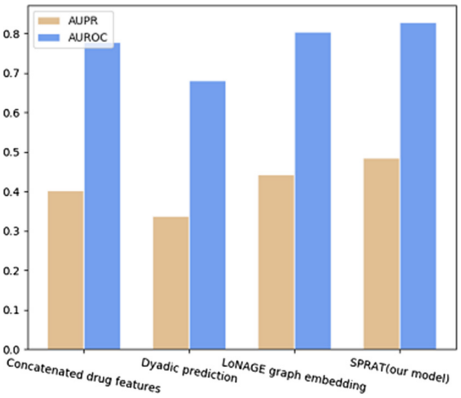


**Fig. 2.** The results of multi-class ADRs prediction under 5-CV.

## 5 Conclusion

This paper proposes a new method to predict the adverse reactions between drug combinations. In order to improve the prediction efficiency, this method uses the variational autoencoder to integrate the discriminant classifier. The greatest advantage of this model is that it is a semi-supervised learning method which can make full use of the effective information hiddened in unlabeled data. In order to verify the effectiveness of this method, experiments are carried out on a large real world dataset (Twosides) and compared with several representative methods. Finally, the experiments results demonstrate that SPRAT performs better than the other three state-of-the-art approaches. The direction of future work is to combine our method with the graph and add more drug-related data to improve the prediction accuracy of specific ADRs types.

**Acknowledgements.** This work is partially supported by the NSFC No. 91846205; the National Key R&D Program No. 2017YFB1400100; the Innovation Method Fund of China No. 2018IM020200; the Shandong Key R&D Program No. 2018YFJH0506, 2019JZZY011007.

## References

1. Campillos, M., Kuhn, M., Gavin, A.C., Jensen, L.J., Bork, P.: Drug target identification using side-effect similarity. Science **321**(5886), 263–266 (2008)
2. Davis, J., Goadrich, M.: The relationship between precision-recall and ROC curves, pp. 233–240 (2006)
3. Jamal, S., Goyal, S., Shanker, A., Grover, A.: Predicting neurological adverse drug reactions based on biological, chemical and phenotypic properties of drugs using machine learning models. Sci. Rep. **7**(1), 872–872 (2017)
4. Jin, B., Yang, H., Xiao, C., Zhang, P., Wei, X., Wang, F.: Multitask dyadic prediction and its application in prediction of adverse drug-drug interaction, pp. 1367–1373 (2017)
5. Kastrin, A., Ferk, P., Leskosek, B.: Predicting potential drug-drug interactions on topological and semantic similarity features using statistical learning. PLOS ONE **13**(5), e0196865 (2018)
6. Kim, S., et al.: PubChem substance and compound databases. Nucleic Acids Res. **44**, 1202–1213 (2016)
7. Li, P., et al.: Large-scale exploration and analysis of drug combinations. Bioinformatics **31**(12), 2007–2016 (2015)
8. Percha, B., Altman, R.B.: Informatics confronts drug-drug interactions. Trends Pharmacol. Sci. **34**(3), 178–184 (2013)
9. Qato, D.M., Wilder, J., Schumm, L.P., Gillet, V., Alexander, G.C.: Changes in prescription and over-the-counter medication and dietary supplement use among older adults in the united states, 2005 vs 2011. JAMA Intern. Med. **176**(4), 473–482 (2016)
10. Shi, J., Li, J., Gao, K., Lei, P., Yiu, S.: Predicting combinative drug pairs towards realistic screening via integrating heterogeneous features. BMC Bioinformatics **18**(12), 409 (2017)

11. Tatonetti, N.P., Ye, P.P., Daneshjou, R., Altman, R.B.: Data-driven prediction of drug effects and interactions. Sci. Transl. Med. **4**(125), 125ra31 (2012)
12. Tran, P.V.: Learning to make predictions on graphs with autoencoders, pp. 237–245 (2018)
13. Vilar, S., Friedman, C., Hripcsak, G.: Detection of drug-drug interactions through data mining studies using clinical sources, scientific literature and social media. Brief. Bioinform. **19**(5), 863–877 (2018)
14. Zhang, P., Wang, F., Hu, J., Sorrentino, R.: Label propagation prediction of drug-drug interactions based on clinical side effects. Sci. Rep. **5**(1), 12339–12339 (2015)
15. Zhang, W., Chen, Y., Liu, F., Luo, F., Tian, G., Li, X.: Predicting potential drug-drug interactions by integrating chemical, biological, phenotypic and network data. BMC Bioinformatics **18**(1), 18 (2017)
16. Zitnik, M., Agrawal, M., Leskovec, J.: Modeling polypharmacy side effects with graph convolutional networks. Bioinformatics **34**(13), 457–466 (2018)