Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/42/20230683

Adverse Drug Reactions prediction by combining wide & deep learning and POLY2

Daoming Liu[1, †], Zexin Fu[2,3, †]

[1]School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi, 214122, China
[2]School of Computer Science and Technology, Shandong University, Qingdao, 266237, China

[3]jasonfu@mail.sdu.edu.cn
[†]These authors contributed equally to this work and should be considered co-first authors

Abstract. Accurate prediction of Adverse Drug Reactions (ADRs) holds immense importance in the field of clinical medicine and drug development. The requirement of accurate prediction spans various stages, ranging from drug design and clinical trials to marketing monitoring. The traditional ADR forecasting method has the disadvantage that it requires a lot of computing resources and is not suitable for large-scale forecasting. To address this issue, this study introduces the Wide & Deep model. This model combines the abilities of memorization and generalization to enhance the accuracy of ADR predictions. Additionally, we identify a shortcoming in the wide component of the traditional Wide & Deep model – the lack of nonlinear transformation. Therefore, we propose the inclusion of POLY2 in the Wide & Deep model to rectify this shortcoming. By incorporating POLY2, our aim is to retain the model's memorization and generalization abilities, leverage the nonlinear relationship between features, and capture the interaction effect between drug chemical substructures for better model performance. To validate our proposed method, we conduct experiments on two datasets: the FDA Adverse Event Reporting System (FAERS) and PubChem. The evaluation metric utilized is the Area Under the Curve (AUC) score, which demonstrates that our method outperforms the original model. The results indicate that by combining POLY2 feature crosses with the Wide & Deep model, we have achieved significant improvements in the prediction of ADRs.

Keywords: Adverse drug reactions prediction, Wide & Deep model, POLY2, FDA Adverse Event Reporting System, PubChem.

1. Introduction

1.1. Research Background

The wide application of drugs in the medical field has made great achievements and made great contributions to human health. However, the use of drugs is also accompanied by the inevitable risk of Adverse Drug Reactions (ADRs). ADRs refer to the harmful physical reactions of patients when receiving drug treatment. These reactions are caused by various factors such as drug properties, chemical

DOI: 10.54254/2755-2721/42/20230683

structure, and dosage. Any drug may cause ADRs in the human body, and these ADRs are becoming the main cause of morbidity and death in the medical process. At the same time, ADRs will not only affect the health and quality of life of patients, but also increase the burden of medical resources [1, 2]. Therefore, reducing the occurrence of ADRs is one of the important challenges in the medical field. At present, clinical trials are the main means to detect ADRs, but due to the need for sufficient experimental samples and time, this method cannot completely avoid undetected ADRs after the drug is marketed. Therefore, accurate prediction of potential ADRs has become one of the important tasks in clinical medicine and drug development.

1.2. Traditional Methods and Feature Engineering

Early ADRs prediction methods are based on biochemical and cellular determination of compounds for safety analysis [3]. The cost and efficiency of these traditional methods have gone beyond the acceptable range of drug research and development and clinical treatment. With the continuous development of machine learning and deep learning technology, scholars have begun to explore how to use related technologies to improve the prediction accuracy of ADRs prediction models. In this context, Wide & Deep model, as a method that combines wide component and deep component, has the abilities of memorization and generalization at the same time, and has achieved success in the fields of recommendation system and advertising click-through rate [4]. However, due to the complexity and nonlinear relationship of drug features, the traditional Wide & Deep model cannot calculate the nonlinear relationship between drug features, which leads to the fact that the Wide & Deep model cannot be directly used in ADRs prediction [5, 6]. To address this issue, our research suggests employing the POLY2-based approach for the wide component of the Wide & Deep model. This method incorporates POLY2 into the wide component to assess the non-linear association between chemical substructures of drugs. However, the inclusion of POLY2 brings about new challenges, namely the curse of dimensionality and an increase in the ratio of irrelevant information [7- 9]. To mitigate the complications arising from the integration of POLY2, this study undertakes dimensionality reduction on the input data prior to feature crosses. This approach aims to tackle the issue of the curse of dimensionality and enhance the proportion of valuable data.

There are three databases commonly used for ADRs prediction: FDA Adverse Event Reporting System (FAERS) database is an official channel provided by the US Food and Drug Administration; Side Effect Resource (SIDER) database contains a large amount of information on ADRs. Healthcare Cost and Utilization Project (HCUP) database contains a large amount of data on patient healthcare costs and utilization. In this paper, the PubChem dataset is introduced to conduct experiments on the basis of using the FAERS dataset. PubChem is a chemical module database maintained by the National Centre for Biotechnology Information, which contains a large amount of information on chemical structures, biochemical experimental data, chemical structure information of compounds, and raw data. These data can provide important information for ADRs prediction.

1.3. Innovation Point of the Research

The innovation of this paper is to introduce the POLY2 method and combine it with the wide component of the Wide & Deep model to improve the ADRs prediction model. This innovation aims to address the complex nonlinear relationships between drug features that are difficult to capture by traditional ADRs prediction methods and simple collaborative filtering models. Feature crosses allow us to obtain higher-order interactions between features, thus more accurately predicting the ADRs. In the Wide & Deep model, the wide part is mainly responsible for capturing the association of known linear features, and the POLY2 method is introduced to enhance the ability of capturing the nonlinear feature association of the wide part, so as to better deal with the nonlinear relationship between features. With this innovation, we attempt to address a long-standing challenge in the field of ADRs prediction, which is how to provide the model with memorization and generalization abilities to improve prediction performance when given a new drug. In this study, by performing experimental validation on two datasets, namely FAERS and PubChem, and employing AUC as the metric for evaluation, we illustrate the efficacy of the suggested

DOI: 10.54254/2755-2721/42/20230683

approach. Our study aims to further improve the accuracy of ADRs prediction, thus providing more reliable support for clinical practice and drug research and development.

The remaining sections of this paper are as follows: Section II discusses related work on ADRs prediction and briefly describes the Wide & Deep model and POLY2. Section III discusses the research problem setting of this paper. Section IV describes the Wide & Deep model and POLY2 model in detail. Section V discusses the experimental setup and results. Finally, section VI concludes the paper.

2. Related Work

Numerous studies have been dedicated to enhancing the efficacy of prediction models in the field of ADRs prediction. These studies aim to provide superior guidance in drug discovery and clinical application. This chapter delves into the research pertaining to ADRs prediction, with specific emphasis placed on the utilization of collaborative filtering, machine learning, and deep learning techniques.

2.1. ADR Prediction Method Based on Drug Characteristics

Existing research on ADRs prediction mainly utilizes three sources of information: biological information [10], drug molecular structure information and market monitoring, drug-ADRs interaction data. Many data types belong to the third drug-ADRs interaction data and market monitoring data, which require a lot of time and economic costs, so it is more practical to choose biological information or drug molecular structure information [11]. In this paper, FAERS dataset and PubChem dataset are selected to perform ADRs prediction based on drug structural features. Similarity-based and machine learning-based methods are the two existing structure-based methods [12].

ADRs prediction based on drug structural similarity, as described by previous studies [13-15], is conducted by identifying molecules that exhibit structural resemblance to currently available drugs. Although this method is easier to implement than the other method, when the drug structure is diversified and the data structure is multi-dimensional, the prediction accuracy of this method will be relatively reduced. At the same time, this method deals with all structural features with the same weight, which cannot improve the prediction result for a specific ADR. Moreover, it is more difficult to find the chemical structures associated with ADRs with this method. ADRs prediction based on machine learning establish ADRs prediction models by utilizing molecular fingerprints [4], circular fingerprints [16-18], and various types of models (e.g., Bayesian models and decision trees [19, 20]). Most existing machine learning-based prediction methods have pre-defined fingerprints, resulting in the possibility that it is unable to compute, search for all secondary chemical structures. At the same time, this method is not suitable for large-scale data prediction. Therefore, how to make full use of all drug structures and expand the range of data volume has become the focus of research.

2.2. Integrated Application of Collaborative Filtering and Deep Learning in ADRs Prediction

Collaborative filtering is a widely used technique in the field of recommender systems, which predicts items that a user may like by mining the similarity between the user and the item based on the user-item interaction matrix [5, 21, 22]. In the field of ADRs prediction, collaborative filtering method is introduced to analyse the association between drugs and structures to predict ADRs. However, traditional collaborative filtering methods have limitations when facing the complex task of ADRs prediction. ADRs are usually affected by a variety of factors, including drug molecular structure, patient genotype, and environmental factors. Traditional collaborative filtering methods are difficult to effectively consider the diversity and complexity of data.

To overcome these limitations, we introduce the Wide & Deep model [4], a method that combines wide component and deep component. The wide part mainly utilizes the generalized linear model to learn the known correlations between features, while the deep part learns higher order nonlinear relationships through deep neural networks. The combination enables the model to grasp linear as well as nonlinear relationships of features to a certain extent. Through experimental proof, we expect that this method can provide more accurate ADRs prediction for drug development and clinical practice, and provide more reliable support for medical decision-making.

DOI: 10.54254/2755-2721/42/20230683

At the same time, in ADRs prediction, deep learning also shows its powerful ability. In particular, Multilayer Perceptron (MLP) [23-25], as a common deep learning architecture, performs well in ADRs prediction. MLP has multiple layers of neurons, which can effectively represent and learn high-dimensional features. Its nonlinear activation function and hierarchical structure allow it to extract more abstract features from the data that better capture the hidden associations between drug chemical substructures and ADRs. By optimizing weights and biases during training, MLP can be adapted to different data distributions and problems. In addition, the deep learning model applying MLP can also deal with large-scale data and has strong generalization ability. This is particularly important for ADRs prediction, as feature-ADR association data are often high-dimensional, sparse, and complex. Deep learning models applying MLP can also help to identify patterns and relationships hidden behind such data, thus improving prediction performance.

3. Problem Setup

To formally define the problem in our research, we adopt the formal description of collaborative filtering as a basis:

Given a matrix R containing drug-ADR associations, where $R_{i,j}\ \ i \in N, j \in M$ denotes whether drug i causes ADR j or not. Also, given a drug-structure feature association variable X, where $X_{i,k}\ \ i \in N, k \in D$ denotes whether drug i has chemical substructure j or not, where N is the set of drugs, M is the set of ADRs, and D is the set of drug chemical substructures. Our goal is to predict entries in the drug-ADR association matrix R that have not yet been observed, that is, to predict potential ADRs.

In this problem context, our study aims to explore how the Wide & Deep model as well as POLY2 can be utilized to improve the prediction performance of ADRs. We will verify the effectiveness and superiority of the proposed method by comparing the AUC evaluation metrics by conducting experiments on FAERS and PubChem datasets [26, 27]. We will also explore the limitations of the proposed method and the potential value in practical clinical applications. Through these experiments and analyses, we hope to provide valuable insights and contributions to research and practical applications in the field of ADRs prediction.

4. Method

4.1. The Wide & Deep Model

The Wide & Deep learning framework is a hybrid model composed of a single-layer wide component and a multi-layer deep component [4].

The generalized linear model, known as the wide component, can be expressed as $y = w^T x + b$, where y is the prediction, x is the feature vector, w are the model parameters, and b is the bias. The wide component can provide the Wide & Deep model with memorization ability.

The deep component consists of a neural network that operates in a feed-forward fashion. Initially, the original inputs undergo a conversion process to obtain an embedding vector. These embedding vectors are randomly initialized and subsequently trained to minimize the overall loss function throughout the model training phase. Subsequently, the embedding vectors are utilized as input to the hidden layers in the neural network during the forward pass. To be specific, each hidden layer is computed according to the following formula:

$$a^{(l+1)} = f(W^{(l)}a^{(l)} + b^{(l)}) \tag{1}$$

where l is the layer number and f is the activation function. $a^{(l)}$, $b^{(l)}$ and $W^{(l)}$ are the activations, bias, and model weights at l-th layer. In our research, we use Multilayer Perceptron (MLP) in the deep component.

DOI: 10.54254/2755-2721/42/20230683

4.2. POLY2

POLY2 is the simplest method of feature crosses and can be used to process high-dimensional sparse data in machine learning. Its main purpose is to combine all the features in pairs to generate a linear model. The mathematical expression of POLY2 is as follows:

$$y = w_0 + \sum_i^n w_i x_i + \sum_{j_1}^n \sum_{j_2=j_1+1}^n w_{j_1 j_2} x_{j_1 j_2} \quad (2)$$

where x is the feature vector, w are the weights of the features and w_0 is the intercept.

5. Experiment

5.1. Experimental Setup

5.1.1. Datasets. FAERS database contains information on adverse events and medication errors reported to FDA for all marketed drug and therapeutic biologic products. We obtain 1,358 drugs, and 2,707 ADRs, using binary features to indicate whether each drug can cause a certain ADR.

PubChem database provides drug chemical substructures. We also obtain 1,358 drugs and 881 drug chemical substructures. There is a one-to-one correspondence between the IDs of drugs in PubChem database and FAERS database.

5.1.2. Evaluation Protocol. To assess our performance, we employ the evaluation protocol of Area Under the Curve (AUC) [26, 27]. The AUC represents the area enclosed by the ROC curve with the axis. The ROC curve plots the False Positive Rate (FPR) on the horizontal axis and the True Positive Rate (TPR) on the vertical axis. FPR and TPR are defined as follows:

$$FPR = \frac{False\ Positive}{False\ Positive + True\ Negative} \quad (3)$$

$$TPR = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (4)$$

5.2. Result

We use different parameters to train the model and take out the best AUC results in multiple training results, and compare the AUC of the Wide & Deep model without using POLY2 for feature crossover with that of the Wide & Deep model using POLY2 for feature crossover. Seven of the typical results are presented in Table 1. In Table 1, parameter embed_dims represent the dimension of the embedding vector in the deep component, parameter mlp_dims represents the number of neurons in each hidden layer of the MLP in the deep component, and parameter dropout represents the proportion of neurons discarded during training. From Table 1, we can see that the AUC of the Wide & Deep model using POLY2 for feature crossover is better than that of the Wide & Deep model without POLY2 for feature crossover in the first six cases. And in some special cases, such as the last case in Table 1, it fails to outperform the model without POLY2.

Table 1. Comparison of AUC of the two methods with different parameters.

Parameters	Wide & Deep	Wide & Deep + POLY2
embed_dim=16, mlp_dims=(128,64,32), dropout=0.2	0.581	0.662
embed_dim=16, mlp_dims=(512,128,64), dropout=0.2	0.602	0.623
embed_dim=16, mlp_dims=(512,128,64), dropout=0.3	0.604	0.614
embed_dim=16, mlp_dims=(512,128,64), dropout=0.4	0.612	0.621
embed_dim=2, mlp_dims=(512,256,128), dropout=0.5	0.585	0.599
embed_dim=8, mlp_dims=(512,256,128), dropout=0.5	0.627	0.643
embed_dim=128, mlp_dims=(512,256,128), dropout=0.5	0.614	0.605

DOI: 10.54254/2755-2721/42/20230683

5.3. Discussion

We use a combination of POLY2 and the wide component of Wide & Deep model as our method and we apply the memorization and generalization abilities of Wide & Deep model to predict ADRs based on drug chemical substructures. Considering the interaction effect of drug chemical substructures on ADR, we use POLY2 for feature crossover in the wide component to better consider the nonlinear relationship between features. In general, the performance of the model using POLY2 for feature crossover is superior, and it can predict ADR more effectively.

However, introducing POLY2 can easily cause the problem of curse of dimensionality. In order to solve this problem, the method we adopt is to reduce the dimension of the features first, which can effectively reduce the dimension after feature crossover. This method simply and directly solves the problem of curse of dimensionality brought by POLY2, but this method also has certain limitations. In the process of dimensionality reduction of features, it is inevitable to lose some information in the features. In the process of feature crosses, the impact caused by the loss of information may also be amplified: the correspondence between some original features and ADRs is amplified, while the correspondence between other features and ADRs is reduced. This deficiency reduces the memorization ability of the wide component, and thus reduces the overall performance of the model, which is the part that needs to be improved.

During our experiments, in most cases the accuracy of the prediction results of the model combining Wide & Deep and POLY2 are able to exceed that of the original model. By reducing the dimensionality of the original features, we reduce the cost of model training, making it possible to use POLY2 to perform feature crosses on data with high dimensions. Overall, this approach has more advantages than disadvantages. The ability of the model to extract abstract features and the ability to generalize reduce overfitting and enable the model to achieve good results in the prediction of ADRs for new drugs, which is a major advantage of the model.

6. Conclusion

In this paper, we apply the Wide & Deep model to the field of ADRs prediction. POLY2 is used for feature crossover in the wide part, and the features after feature crossover are used to train the generalized linear model in the wide component. In the deep component, we embed the drug chemical substructures into the embedding layer and then input them into the MLP. Finally, the output of the linear model of the wide component is mixed with the MLP output of the deep component for forward propagation. We contrast the performance of ADRs prediction of the model using POLY2 for feature crossover with the model without feature crossover. The results show that the model using POLY2 for feature crossover can achieve more accurate results overall. The generalization ability of deep component of Wide & Deep model enables our model to have good prediction results for potential ADRs of new drugs, which is of great significance for clinical medicine and drug research and development.

Since we use the dimensionality reduction method in the wide component to avoid the curse of dimensionality caused by POLY2, we will lose some information in the features. This leads to a decrease in the memorization ability of the wide component, which affects the overall performance of the model. It is a limitation of our current research results. Our research results can be improved on the feature crossover method in the future to avoid the curse of dimensionality while avoiding the loss of feature information. And the overall framework we use can become a new idea in the field of ADRs prediction and be improved to better solve this problem.

Acknowledgment

Daoming Liu and Zexin Fu contributed equally to this work and should be considered co-first authors.

References

[1] Lin X, Li J, Pan J, et al. Effect of “Shenque” Qi distribution theory on malignant Bone cancer patients’ quality of life (QOL) during chemotherapy[C]//2012 IEEE International Conference on Bioinformatics and Biomedicine Workshops. IEEE, 2012: 352-354.

DOI: 10.54254/2755-2721/42/20230683

[2] Park Y H, Jung K H, Im S A, et al. Quality of life (QoL) in metastatic breast cancer patients with maintenance paclitaxel plus gemcitabine (PG) chemotherapy: results from phase III, multicenter, randomized trial of maintenance chemotherapy versus observation (KCSG-BR07-02)[J]. Breast cancer research and treatment, 2015, 152: 77-85.
[3] Whitebread S, Hamon J, Bojanic D, et al. Keynote review: in vitro safety pharmacology profiling: an essential tool for successful drug development[J]. Drug discovery today, 2005, 10(21): 1421-1433.
[4] Cheng H T, Koc L, Harmsen J, et al. Wide & deep learning for recommender systems[C]//Proceedings of the 1st workshop on deep learning for recommender systems. 2016: 7-10.
[5] Koren Y, Bell R, Volinsky C. Matrix factorization techniques for recommender systems[J]. Computer, 2009, 42(8): 30-37.
[6] Sun Y, Pan J, Zhang A, et al. FM2: Field-matrixed factorization machines for recommender systems[C]//Proceedings of the Web Conference 2021. 2021: 2828-2837.
[7] Luo Y, Wang M, Zhou H, et al. Autocross: Automatic feature crossing for tabular data in real-world applications[C]//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2019: 1936-1945.
[8] Chen L, Esfandiari H, Fu G, et al. Feature cross search via submodular optimization[J]. arXiv preprint arXiv:2107.02139, 2021.
[9] Li B, Vorobeychik Y. Feature cross-substitution in adversarial classification[J]. Advances in neural information processing systems, 2014, 27.
[10] Scheiber J, Chen B, Milik M, et al. Gaining insight into off-target mediated effects of drug candidates with a comprehensive systems chemical biology analysis[J]. Journal of chemical information and modeling, 2009, 49(2): 308-317.
[11] LaBute M X, Zhang X, Lenderman J, et al. Adverse drug reaction prediction using scores produced by large-scale drug-protein target docking on high-performance computing machines[J]. PloS one, 2014, 9(9): e106 8.
[12] Dey S, Luo H, Fokoue A, et al. Predicting adverse drug reactions through interpretable deep learning framework[J]. BMC bioinformatics, 2018, 19(21): 1-13.
[13] Liu R, Zhang P. Towards early detection of adverse drug reactions: combining pre-clinical drug structures and post-market safety reports[J]. BMC medical informatics and decision making, 2019, 19(1): 1-9.
[14] Muñoz E, Nováček V, Vandenbussche P Y. Using drug similarities for discovery of possible adverse reactions[C]//AMIA Annual Symposium Proceedings. American Medical Informatics Association, 2016, 2016: 924.
[15] Liu M, Wu Y, Chen Y, et al. Large-scale prediction of adverse drug reactions using chemical, biological, and phenotypic properties of drugs[J]. Journal of the American Medical Informatics Association, 2012, 19(e1): e28-e35.
[16] Wang Z, Liang L, Yin Z, et al. Improving chemical similarity ensemble approach in target prediction[J]. Journal of cheminformatics, 2016, 8: 1-10.
[17] Cereto-Massagué A, Ojeda M J, Valls C, et al. Molecular fingerprint similarity search in virtual screening[J]. Methods, 2015, 71: 58-63.
[18] Rogers D, Hahn M. Extended-connectivity fingerprints[J]. Journal of chemical information and modeling, 2010, 50(5): 742-754.
[19] Bender A, Scheiber J, Glick M, et al. Analysis of pharmacology data and the prediction of adverse drug reactions and off-target effects from chemical structure[J]. ChemMedChem: Chemistry Enabling Drug Discovery, 2007, 2(6): 861-873.
[20] Hammann F, Gutmann H, Vogt N, et al. Prediction of adverse drug reactions using decision tree modeling[J]. Clinical Pharmacology & Therapeutics, 2010, 88(1): 52-59.

DOI: 10.54254/2755-2721/42/20230683

[21] Resnick P, Iacovou N, Suchak M, et al. Grouplens: An open architecture for collaborative filtering of netnews[C]//Proceedings of the 1994 ACM conference on Computer supported cooperative work. 1994: 175-186.
[22] Sarwar B, Karypis G, Konstan J, et al. Item-based collaborative filtering recommendation algorithms[C]//Proceedings of the 10th international conference on World Wide Web. 2001: 285-295.
[23] Tolstikhin I O, Houlsby N, Kolesnikov A, et al. Mlp-mixer: An all-mlp architecture for vision[J]. Advances in neural information processing systems, 2021, 34: 24261-24272.
[24] Pinkus A. Approximation theory of the MLP model in neural networks[J]. Acta numerica, 1999, 8: 143-195.
[25] Taud H, Mas J F. Multilayer perceptron (MLP)[J]. Geomatic approaches for modeling land change scenarios, 2018: 451-455.
[26] Lobo J M, Jiménez-Valverde A, Real R. AUC: a misleading measure of the performance of predictive distribution models[J]. Global ecology and Biogeography, 2008, 17(2): 145-151.
[27] Huang J, Ling C X. Using AUC and accuracy in evaluating learning algorithms[J]. IEEE Transactions on knowledge and Data Engineering, 2005, 17(3): 299-310.