ADRNet: A Generalized Collaborative Filtering Framework Combining Clinical and Non-Clinical Data for Adverse Drug Reaction Prediction

Haoxuan Li
Peking University
Beijing, China
hxli@stu.pku.edu.cn

Taojun Hu
Peking University
Beijing, China
hutaojun@stu.pku.edu.cn

Zetong Xiong
Yale University
New Haven, USA
zx232@yale.edu

Chunyuan Zheng
University of California, San Diego
San Diego, USA
czheng@ucsd.edu

Fuli Feng
University of Science and Technology of China
Hefei, China
fulifeng93@gmail.com

Xiangnan He
University of Science and Technology of China
Hefei, China
xiangnanhe@gmail.com

Xiao-Hua Zhou*
Peking University
Beijing, China
azhou@math.pku.edu.cn

ABSTRACT

Adverse drug reaction (ADR) prediction plays a crucial role in both health care and drug discovery for reducing patient mortality and enhancing drug safety. Recently, many studies have been devoted to effectively predict the drug-ADRs incidence rates. However, these methods either did not effectively utilize non-clinical data, i.e., physical, chemical, and biological information about the drug, or did little to establish a link between content-based and pure collaborative filtering during the training phase. In this paper, we first formulate the prediction of multi-label ADRs as a drug-ADR collaborative filtering problem, and to the best of our knowledge, this is the first work to provide extensive benchmark results of previous collaborative filtering methods on two large publicly available clinical datasets. Then, by exploiting the easy accessible drug characteristics from non-clinical data, we propose ADRNet, a generalized collaborative filtering framework combining clinical and non-clinical data for drug-ADR prediction. Specifically, ADRNet has a shallow collaborative filtering module and a deep drug representation module, which can exploit the high-dimensional drug descriptors to further guide the learning of low-dimensional ADR latent embeddings, which incorporates both the benefits of collaborative filtering and representation learning. Extensive experiments are conducted on two publicly available real-world drug-ADR clinical datasets and two non-clinical datasets to demonstrate the accuracy and efficiency of the proposed ADRNet. The code is available at https://github.com/haoxuanli-pku/ADRnet.

CCS CONCEPTS

• **Applied computing** → **Bioinformatics**; • **Information systems** → **Collaborative filtering**.

KEYWORDS

adverse drug reaction; sided effect; drug-ADR prediction

ACM Reference Format:
Haoxuan Li, Taojun Hu, Zetong Xiong, Chunyuan Zheng, Fuli Feng, Xiangnan He, and Xiao-Hua Zhou. 2023. ADRNet: A Generalized Collaborative Filtering Framework Combining Clinical and Non-Clinical Data for Adverse Drug Reaction Prediction. In *Seventeenth ACM Conference on Recommender Systems (RecSys '23), September 18–22, 2023, Singapore, Singapore.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3604915.3608813

1 INTRODUCTION

Adverse drug reactions (ADRs, also known as side effects) refer to harmful effects produced by drugs that are detrimental to the patient's treatment in normal therapy [11, 19]. Until now, thousands types of ADRs have been reported, many of which have led to serious unwanted and harmful consequences [39]. Timely and effective warning of ADRs can help regulate as well as guide the production of drugs with limited side effects. Therefore, the accurate prediction of ADRs is meaningful in both health care [17] and drug discovery [2] for improving drug safety and reducing patient mortality.

Many previous approaches are based on association rule mining and statistical significance tests [1] to identify the significance of drug-ADR associations [6, 12, 15, 26, 34, 36], but with limited performance on prediction tasks. An alternative class of deep learning-based methods uses pharmacovigilance networks [5, 9], ensemble approaches [18], and deep neural models [31, 37], and has achieved

*Xiao-Hua Zhou is the corresponding author.

RecSys '23, September 18–22, 2023, Singapore, Singapore

ACM ISBN 979-8-4007-0241-9/23/09…$15.00
https://doi.org/10.1145/3604915.3608813

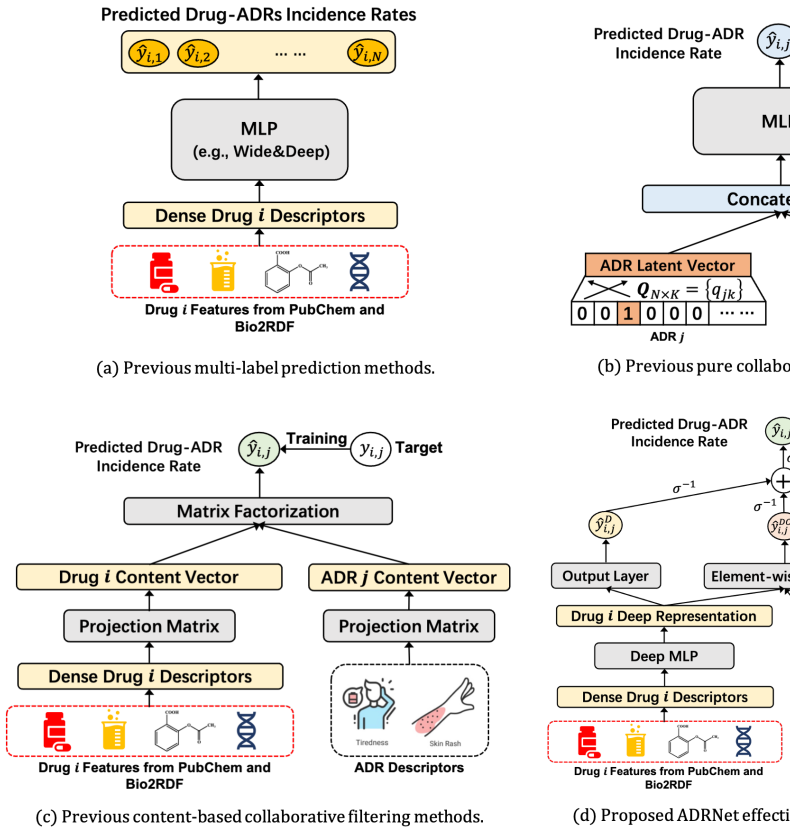
promising predictive performance for specific adverse reactions, as shown in Figure 1(a). However, as drugs may cause multiple ADRs simultaneously, the direct repeat use of the single-label prediction methods may lead to time-consuming and suboptimal prediction performance.

To tackle this issue, collaborative filtering methods can efficiently and accurately predict the relationship between drugs and multiple ADRs. The underlying motivation is that drugs with similar interactions in clinical data tend to have similar ADRs. For example, matrix factorization (MF) [21, 29] and neural matrix factorization (NMF) [16] first learn a latent embedding for each drug and ADR, and then perform drug-ADR prediction. However, as shown in Figure 1(b), the pure collaborative filtering methods ignore the non-clinical data (e.g., drug descriptors [20]), which are readily available and can significantly improve the drug-ADR prediction performance. Recently, content-based collaborative filtering methods have achieved impressive performance in recommendation, as shown in Figure 1(c), by incorporating user-side and item-side features, the prediction performance can be further improved. *However, unlike traditional recommender systems where users and items have symmetric information and status, data collected for drugs and ADRs usually contain non-clinical drug characteristics but lack of ADR characteristics (e.g., cough and fever), due to the inconvenience of the ADR characteristics to be obtained and processed, which poses a great challenge for developing effective drug-ADR predictions using both clinical and non-clinical data.*

In this paper, we first formulate the prediction of multi-label ADRs as a drug-ADR collaborative filtering problem, and to the best of our knowledge, this is the first study to provide extensive benchmark results of previous collaborative filtering methods on two large publicly available clinical datasets. Next, by noting the easy accessibility of drug characteristics and the asymmetric information of drugs and ADRs that differs from previous collaborative filtering, we propose ADRNet, a generalized collaborative filtering framework combining clinical and non-clinical data for drug-ADR prediction. Specifically,ADRNet has a shallow collaborative filtering module and a deep drug representation module, which can exploit the high-dimensional drug descriptors to further guide the learning of low-dimensional ADR latent embeddings, which incorporates both the benefits of collaborative filtering and representation learning.

Notably, the proposed ADRNet effectively exploits both the advantages of deep content-based and latent collaborative filtering-based approaches. On one hand, compared with the deep drug feature-based networks, our approach has fewer parameters and is convenient. Previous studies mapped drugs to ADR space [10, 23, 40] or to metabolic reaction space [32] and applied flux variability analysis to represent drug-protein/gene interactions [24], whereas our approach further utilize the learned ADR latent embedding from collaborative filtering to guide the deep representations learning of drugs. On the other hand, compared with the latent collaborative filtering-based approaches, the proposed approach has the clear advantage of exploiting the chemical, physical and biological characteristics of drugs contained in non-clinical data [4] to help reveal the mechanisms of drug-ADR and improve the prediction performance.

The contributions of this paper are summarized as follows.



(a) Previous multi-label prediction methods.

(b) Previous pure collabo

(c) Previous content-based collaborative filtering methods.

(d) Proposed ADRNet effecti

Figure 1: Comparisons of (a) previous multi-label prediction methods; (b) previous pure collaborative filtering methods; (c) previous content-based collaborative filtering methods; (d) proposed ADRNet as a generalized collaborative filtering framework combining clinical and non-clinical data for drug-ADR prediction, contains (i) a deep drug representation module; (ii) a shallow latent collaborative filtering module; and (iii) a drug collaborative filtering module.

- We formulate the prediction of multi-label ADRs as a drug-ADR collaborative filtering problem, and to the best of our knowledge, this is the first study to provide extensive benchmark results of previous collaborative filtering methods on two large publicly available clinical datasets.
- By exploiting the easy accessible drug characteristics from non-clinical data, we propose ADRNet, a generalized collaborative filtering framework combining clinical and non-clinical data for drug-ADR prediction.
- We further propose a sharing mechanism for ADR latent vectors that is easily applicable to previous pure collaborative filtering methods, and a learning approach to jointly train a deep drug representation and a shallow collaborative filtering network to better trade-off the predictions of each sub-network.
- Extensive experiments are conducted on two publicly available real-world drug-ADR clinical datasets and two non-clinical datasets to demonstrate the accuracy and efficiency of the proposed ADRNet.

2 PRELIMINARIES AND PROBLEM SETUP

Let M and N be the number of drugs and ADRs, respectively. Based on the clinical data containing drug-ADR records from electronic health or adverse reporting systems [3, 22], we define the drug-ADR interaction matrix $\mathbf{Y} \in \mathbb{R}^{M \times N}$ as

$$y_{i,j} = \begin{cases} 1, & \text{if interaction (drug } i, \text{ ADR } j \text{) is observed;} \\ 0, & \text{otherwise.} \end{cases}$$

The multi-label ADR prediction problem is formalized as a collaborative filtering problem to accurately predict drug-ADR incidence rate. Specifically, the model-based approach is abstracted as learning $\hat{y}_{i,j} = f(i, j \mid \Theta)$, where $\hat{y}_{i,j}$ is the predicted incidence rate of drug i to ADR j, and f is a parametric prediction model, Θ is the model parameters.

To accurately estimate parameters Θ in pure collaborative filtering, we associate each drug i and ADR j with a real-valued vector of hidden features, denoted as $\mathbf{p}_i$ and $\mathbf{q}_j$, respectively. Then many standard collaborative filtering methods can be adopted, e.g., matrix factorization (MF) [21, 29] using the inner product of $\mathbf{p}_i$ and $\mathbf{q}_j$ for drug-ADR interaction prediction as $\hat{y}_{i,j} = f\left(i, j \mid \mathbf{p}_i, \mathbf{q}_j\right) = \mathbf{p}_i^\top \mathbf{q}_j = \sum_{k=1}^{K} p_{ik} q_{jk}$, where K denotes the dimension of the latent space. More generally, the backbone can be substituted by neural networks $\hat{y}_{i,j} = f\left(\mathbf{P}, \mathbf{Q} \mid \Theta\right)$, where $\mathbf{P} \in \mathbb{R}^{M \times K}$ and $\mathbf{Q} \in \mathbb{R}^{N \times K}$, denoting the latent factor matrix for drugs and ADRs, respectively. However, these methods do not exploit drug features in easily available non-clinical data, leading to suboptimal performance in practice.

Content-based collaborative filtering further utilizes non-clinical data that contain chemical, physical and biological information about the drugs, which can be beneficial for the prediction of drug-ADR relations [8, 13, 35, 42]. Specifically, drug descriptors in the non-clinical data can be categorized into two types: physical, chemical descriptors (PC-descriptors), and biological descriptors (BIO-descriptors) [28]. To enhance the prediction accuracy of drug-ADR, we use PubChem [20] as the PC-descriptors and Bio2RDF [4] as the BIO-descriptors for drugs, and denote them with $\mathbf{x}_i^{PC}$ and $\mathbf{x}_i^{BIO}$, respectively.

Unfortunately, unlike drug descriptors, ADRs in non-clinical data always appear in a descriptive manner (e.g., cough and fever), and these features are difficult to be processed and utilized directly to improve the prediction performance. Such drug-ADR information asymmetry poses a serious challenge to the previous collaborative filtering approaches. In order to simultaneously take the advantage of the latent collaborative filtering and exploit the informative drug descriptors, we propose a general asymmetric collaborative filtering framework for accurate drug-ADR prediction next.

3 PROPOSED ADRNET: A GENERALIZED COLLABORATIVE FILTERING FRAMEWORK

In this section, we propose ADRNet, a generalized collaborative filtering framework combining clinical and non-clinical data for drug-ADR prediction. By learning the latent embeddings from clinical data and leveraging high-dimensional drug descriptors from non-clinical data, the proposed method combines the benefits of collaborative filtering and deep representation learning. We introduce the structure of ADRNet as shown in Figure 1(d) in the following subsections, including (i) a deep drug representation module in Section 3.1; (ii) a shallow latent collaborative filtering module in Section 3.2; and (iii) a drug collaborative filtering module in Section 3.3. We discuss the training of ADRNet and summarize its advantages in Section 3.4.

3.1 Deep Drug Representation Module

The deep drug representation module first concatenates and maps sparse and high-dimensional drug PC-descriptors and BIO-descriptors to a low-dimensional and dense real-valued vector, also known as the drug embedding vector. These embedding vectors are randomly initialized, then fed into a feed-forward neural network, and obtain a drug representation *with the same dimensions* as the drug/ADR latent space in latent collaborative filtering. Given a learned drug representation, as in Figure 1(d), the representation is fed into a fully connected (FC) layer to obtain drug-level predictions $\hat{y}_{i,j}^D = \hat{y}_i^D$. In addition, this representation is shared to make predictions together with the ADR latent vector obtained from the collaborative filtering part. Specifically, the deep drug representation layers under our ADRNet framework are defined as

$$\mathbf{z}_1^D = \phi_1(\mathbf{x}_i^{PC}, \mathbf{x}_i^{BIO}) = \begin{bmatrix} \mathbf{x}_i^{PC} \\ \mathbf{x}_i^{BIO} \end{bmatrix}, \quad \mathbf{z}_2^D = \phi_2^D(\mathbf{z}_1^D) = a_2\left((\mathbf{W}_2^D)^\top \mathbf{z}_1^D + \mathbf{b}_2^D\right), \quad \ldots$$

$$\mathbf{z}_L^D = \phi_L^D(\mathbf{z}_{L-1}^D) = a_L\left((\mathbf{W}_L^D)^\top \mathbf{z}_{L-1}^D + \mathbf{b}_L^D\right), \quad \hat{y}_{i,j}^D = \hat{y}_i^D = \sigma\left((\mathbf{h}^D)^\top \mathbf{z}_L^D\right),$$

where $\mathbf{W}_l, \mathbf{b}_l$, and $a_l(\cdot)$ denote the weight matrix, bias vector, and activation function of the l-th layer's perceptron, and $\sigma(\cdot)$ denote the sigmoid function.

3.2 Shallow Latent Collaborative Filtering Module

The shallow latent collaborative filtering module concatenates the drug and ADR latent vectors and performs pure collaborative filtering using a feed-forward neural network, which is able to capture the non-linearity of interactions between $\mathbf{p}_i$ and $\mathbf{q}_j$. Compared with the drug representation module, the collaborative filtering module does not directly use the drug features, and therefore has shallower layers. However, in contrast to neural collaborative filtering (NCF) [16], the learned drug representations can further guide the training of ADR latent vectors in Figure 1(d), which provides a stronger supervised signal for latent vectors compared with collaborative filtering using only clinical interactions, and thus leads to more accurate predictions. Another advantage is induced from the joint learning of the deep network in Section 3.1 and the shallow collaborative filtering. As suggested by the previous studies, jointly learning a shallow and a deep network can further improve the model's "memory" and "generalization" abilities [7, 38]. Specifically, the "memory" ability means that the model directly learns and utilizes the co-occurrence frequency of drugs and ADRs from historical data, and the "generalization" ability refers to letting the model mine the drug-ADR associations, expecting the model to be able to accurately predict sparse or even unobserved drug or ADR features. Formally, the shallow latent collaborative filtering module is formulated as

$$\mathbf{z}_1^{CF} = \phi_1(\mathbf{p}_i, \mathbf{q}_j) = \begin{bmatrix} \mathbf{p}_i \\ \mathbf{q}_j \end{bmatrix}, \quad \mathbf{z}_2^{CF} = \phi_2^{CF}(\mathbf{z}_1^{CF}) = a_2\left((\mathbf{W}_2^{CF})^\top \mathbf{z}_1^{CF} + \mathbf{b}_2^{CF}\right), \quad \ldots$$

$$\mathbf{z}_S^{CF} = \phi_S^{CF}(\mathbf{z}_{S-1}^{CF}) = a_S\left((\mathbf{W}_S^{CF})^\top \mathbf{z}_{S-1}^{CF} + \mathbf{b}_S^{CF}\right), \quad \hat{y}_{i,j}^{CF} = \sigma\left((\mathbf{h}^{CF})^\top \mathbf{z}_S^{CF}\right),$$

where $\mathbf{W}_s, \mathbf{b}_s$, and a_s denote the weight matrix, bias vector, and activation function for the s-th layer's perceptron.

3.3 Drug Collaborative Filtering Module

The drug collaborative filtering module performs element-wise product of learned drug representations from the deep network and ADR latent vectors from the shallow network, and then feeds the element-wise product into an FC layer to obtain $\hat{y}_{i,j}^{DCF}$ as in Figure 1(d) that

$$\hat{y}_{i,j}^{DCF} = \sigma\left((\mathbf{h}^{DCF})^\top\left(\phi_L^D(\mathbf{z}_{L-1}^D) \odot \mathbf{q}_j\right)\right),$$

which has the same form as generalized matrix factorization (GMF) [16], but should be considered as an *asymmetric* collaborative filtering. The key observation is that the learned drug representations are derived from PC-descriptors and BIO-descriptors, which inherently contain extensive structural, chemical, physical and biological features about the drug. Also, the drug representations are fed to an FC layer for drug-ADR prediction in Section 3.1, so that these representations can retain a strong association to ADR. Notably, the ADR latent vectors are *shared* between deep representations and shallow collaborative filtering, which is further interpreted as drug representations can guide the training of ADR latent vectors. We also empirically validate this in our experiments.

3.4 Joint Training of ADRNet

We jointly train the deep drug representation network and the shallow latent collaborative filtering network in ADRNet. Specifically, the drug representation ϕ^D, its element-wise product with the ADR latent vector ϕ^{DCF}, and the collaborative filtering vector ϕ^{CF} are concatenated and fed into an FC layer for drug-ADR prediction that

$$\phi_{i,j}^D = \phi_L^D(\ldots(\phi_2^D(\phi_1(\mathbf{x}_i^{PC}, \mathbf{x}_i^{BIO}))\ldots), \quad \phi_{i,j}^{CF} = \phi_S^{CF}(\ldots(\phi_2^{CF}(\phi_1(\mathbf{p}_i, \mathbf{q}_j))\ldots), \quad \phi_{i,j}^{DCF} = \phi_{i,j}^D \odot \mathbf{q}_j,$$

$$\hat{y}_{i,j} = \sigma\left(\mathbf{h}^\top\begin{bmatrix}\phi_{i,j}^D\\ \phi_{i,j}^{CF}\\ \phi_{i,j}^{DCF}\end{bmatrix}\right) = \sigma\left(\sigma^{-1}(\hat{y}_{i,j}^D) + \sigma^{-1}(\hat{y}_{i,j}^{CF}) + \sigma^{-1}(\hat{y}_{i,j}^{DCF})\right),$$

where $\mathbf{h} = [(\mathbf{h}^D)^\top, (\mathbf{h}^{CF})^\top, (\mathbf{h}^{DCF})^\top]^\top$ denote the edge weights of the output layer. We use mini-batch stochastic optimization to back-propagate the gradients in both the deep and shallow networks, and train the prediction model by minimizing the training loss

$$L(\theta) = -\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} y_{i,j}\log\hat{y}_{i,j} + (1-y_{i,j})\log(1-\hat{y}_{i,j}).$$

We summarize the advantages of ADRNet over previous works as follows: (a) provides a convenient and flexible framework to combine drug descriptor-based models with latent-based collaborative filtering; (b) has better "memory" and "generalization" ability due to joint learning of shallow and deep models; (c) learns more informative drug and ADR latent vectors compared with pure collaborative filtering.

4 EXPERIMENTS

In this section, we aim to answer the following research questions on real-world clinical and non-clinical datasets:

- How does the proposed method perform compared with the existing models in terms of drug-ADR prediction? Does non-clinical data help enhance the previous collaborative filtering methods?
- What factors (shallow collaborative filtering, deep drug representation, latent embedding, drug descriptor, sharing mechanism) influence the validity of our method?
- Does our method stably outperform the previous models at varying latent embedding sizes?
- Are deeper layers in collaborative filtering and drug representation helpful for drug–ADR prediction?

4.1 Experimental Setup

4.1.1 Dataset and Preprocessing. We follow the previous studies [28, 44] to use two widely used real-world clinical datasets Liu's [22] and AEOLUS [3] that contain drug-ADR interactions. We then select the drugs that appear in DrugBank, as well as require ADRs to occur on more than 50 drugs. Specifically, the Liu's dataset contains 58,810 drug-ADR interactions occurring on 828 drugs and 1,385 ADRs, and the AEOLUS dataset contains 605,121 interactions occurring on 1,358 drugs and 2,707 ADRs. In addition, to verify the effectiveness of the proposed generalized collaborative filtering framework, we also adopt two real-world non-clinical data PubChem [20] and Bio2RDF [4]. Specifically, for each drug, PubChem provides pre-defined 881 bits of structural PC-descriptors, and Bio2RDF provides 6,712 bits of chemical, physical, and biological BIO-descriptors. We concatenate the PC-descriptors and BIO-descriptors into 7,593 bits and feed them into the deep drug representation layers.

4.1.2 Baselines. To comprehensively evaluate the proposed method, we compare with the following multi-label prediction methods: **LNSM** [27, 44], **SVM** [18], **RF** [9, 22], **LR** [5, 22], **CCA** [5, 42]. We also compare with the following collaborative filtering methods: **DrugCF**[1], **MF** [21, 29], **NCF** [16], **NMF** [16], **PNN** [30], **FNN** [43], **Deep Crossing** [33], **Wide&Deep** [7], **Deep&Cross** [38], **DeepFM** [14], **NFM** [16], **AFM** [41], and **UltraGCN** [25].

4.1.3 Experimental Protocols and Details. Following previous studies [10, 29, 42, 44], we use 10-fold cross-validation for our experiments. Two common metrics are used to evaluate the prediction performance, i.e., area under the ROC curve (AUC) and area under the precision-recall curve (AUPR). All the experiments are implemented on PyTorch with Adam as the optimizer. We tune the learning rate in $\{0.001, 0.005, 0.01, 0.05\}$, weight decay in $\{1e{-}6, 1e{-}5, 1e{-}4, 1e{-}3\}$, and embedding size in $\{16, 32, 64, 128, 256, 512, 1024\}$ for both Liu's and AEOLUS.

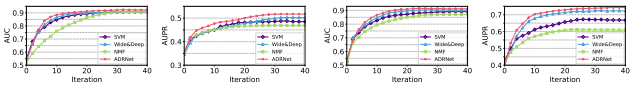
4.2 Performance Comparison

Overall Performance. We compare the proposed ADRNet with a wide range of baseline methods on two real-world datasets Liu's and AEOLUS, and the results are shown in Table 1. We find the proposed ADRNet outperforms the best baseline in terms of AUC and AUPR at 0.01 statistical significance. We also compare the running times of the various methods in Table 1. Compared with the content-based collaborative filtering, ADRNet has a lower running time due to the faster convergence resulting from the joint training of the deep network with latent collaborative filtering.

[1] DrugCF is implemented through drug similarity-based collaborative filtering.

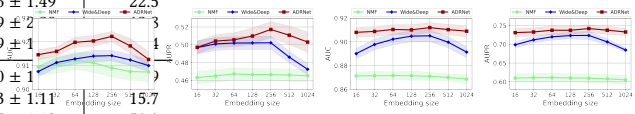
Table 1: Performance comparison of drug–ADR prediction models on Liu's. We bold the best model, and underline the best single-label prediction model and collaborative filtering model. * means statistically significant results (p-value $\leq$ 0.01) using the paired-t-test compared with the best baseline.

Dataset	Liu's			AEOLUS		
Method	AUC ($\times 10^{-2}$)	AUPR ($\times 10^{-2}$)	Time (s)	AUC ($\times 10^{-2}$)	AUPR ($\times 10^{-2}$)	Time (s)
LNSM	90.31 ± 0.25	46.64 ± 0.69	32.6	84.98 ± 0.64	57.40 ± 1.65	89.7
SVM	<u>90.98 ± 0.22</u>	<u>48.79 ± 0.71</u>	8863.3	<u>89.25 ± 0.46</u>	67.52 ± 1.41	20378.0
RF	87.77 ± 0.32	44.00 ± 0.59	7.7	88.56 ± 0.40	66.56 ± 1.49	22.5
LR	89.25 ± 0.23	46.54 ± 0.83	9.5	89.02 ± 1.02	66.89 ± 2.[illegible]	[illegible]
CCA	63.48 ± 0.83	19.50 ± 0.94	72.0	58.59 ± 1.51	26.69 ± 1.[illegible]	[illegible]
DrugCF	87.88 ± 0.35	45.53 ± 0.75	2.2	84.79 ± 0.58	58.40 ± 1.[illegible]	[illegible]
MF	90.89 ± 0.26	45.00 ± 0.65	10.9	87.15 ± 0.37	61.03 ± 1.11	15.7
NCF	89.84 ± 0.36	44.00 ± 0.67	17.2	84.68 ± 0.41	59.35 ± 1.13	50.0
NMF	91.21 ± 0.37	46.73 ± 0.67	19.3	87.16 ± 0.32	61.03 ± 1.11	43.5
PNN	91.10 ± 0.29	46.48 ± 1.06	10.1	90.22 ± 0.35	69.81 ± 0.[illegible]	[illegible]
FNN	91.05 ± 0.43	45.17 ± 0.86	17.6	89.64 ± 0.42	66.72 ± 0.82	19.4
Deep Crossing	90.79 ± 0.38	48.13 ± 1.09	17.2	88.44 ± 0.47	69.48 ± 0.[illegible]	[illegible]
Wide&Deep	<u>91.42 ± 0.25</u>	<u>50.23 ± 0.85</u>	4.6	<u>90.50 ± 0.45</u>	<u>72.38 ± 1.11</u>	9.2
Deep&Cross	90.03 ± 0.27	46.34 ± 1.12	11.7	87.95 ± 0.57	60.40 ± 2.91	43.2
DeepFM	91.35 ± 0.20	50.22 ± 0.75	18.9	90.46 ± 0.44	72.35 ± 1.04	34.2
NFM	91.41 ± 0.25	50.20 ± 0.99	20.3	90.40 ± 0.50	71.85 ± 0.[illegible]	[illegible]
AFM	88.76 ± 0.53	46.05 ± 1.16	3.3	86.71 ± 0.57	65.63 ± 1.41	5.0
UltraGCN	91.29 ± 0.28	47.01 ± 0.69	5.0	89.54 ± 0.40	71.23 ± 1.14	14.6
ADRNet (ours)	**92.23 ± 0.21***	**51.72 ± 0.89***	5.9	**91.20 ± 0.42***	**74.21 ± 1.04**	15.2



(a) AUC on Liu's (b) AUPR on Liu's (c) AUC on AEOLUS (d) AUPR on AEOLUS

Figure 2: Prediction performance of the ADRNet, compared with the best single-label prediction model (SVM), multi-label prediction model (Wide&Deep), latent collabrative filtering model (NMF), w.r.t. the number of iterations on Liu's and AEOLUS.



(a) AUC on Liu's (b) AUPR on Liu's (c) AUC on AEOLUS (d) AUPR on AEOLUS

Figure 3: Effects of the embedding size on Liu's and AEOLUS. the optimal performance when embedding sizes are around 128-256.

Analysis of Convergence. To further explore the convergence performance of ADRNet, we compare AUCs and AUPRs with respect to the number of iterations in Figure 2. We find that ADRNet using joint training has the fastest convergence speed on both datasets, while NMF converges the slowest. This indicates the guiding effect of drug descriptor-based representations on the latent vectors in collaborative filtering.

4.3 Ablation Studies

Shallow vs. Deep Layers. ADRNet has both shallow latent collaborative filtering and deep drug representation layers. In Table 2, we conduct ablation experiments to explore the effects of using the shallow network, using the deep network, and using both on drug-ADR prediction performance. We find that using both shallow and deep networks helps to improve the prediction performance in terms of AUC and AUPR, which is consistent with the discussion in Section 3.2.

Latent Vector vs. Drug Descriptor. We also explore the effect of latent vectors versus drug descriptors on prediction performance, and the results are presented in Table 2. We find that the high-dimensional drug descriptor-based models outperform the pure collaborative filtering models significantly. This is due to the structural, chemical, physical, and biological information about the drug contained in the drug descriptors.

Embedding Size. We next investigate how embedding size affects drug-ADR prediction performance. Specifically, we compare ADRNet with the best pure (NMF) and the best content-based (Wide&Deep) collaborative filtering that uses both shallow and deep networks, with the results shown in Figure 3. Notably, ADRNet stably outperforms NMF and Wide&Deep at all embedding sizes and achieves

5 CONCLUSION

We formulate the prediction of multi-label ADRs as a drug-ADR collaborative filtering problem, and to the best of our knowledge, this is the first study to provide extensive benchmark results of previous collaborative filtering methods on two large publicly available clinical datasets. Then, by noting the gap between previous studies on collaborative filtering with symmetric information and the asymmetry of available drug and ADR information for multi-label ADR predictions, we propose ADRNet as a generalized collaborative filtering framework combining clinical and non-clinical data for drug-ADR prediction. We describe and discuss in detail the three modules of ADRNet: a deep drug representation module, a shallow latent collaborative filtering module, and a drug collaborative filtering module. Notably, ADRNet inherits the "generalization" ability of the deep model and the "memory" ability of the latent collaborative filtering, and further shares the ADR latent vector to establish stronger connection. Finally, we jointly train the deep drug representation and the shallow collaborative filtering network to better trade-off the predictions of each sub-network. Extensive experiments are conducted on two publicly available real-world drug-ADR clinical datasets and two non-clinical datas, and the results show that the proposed ADRNet can effectively combine the advantages of drug descriptors and latent collaborative filtering to improve drug-ADR prediction performance.

REFERENCES

[1] Alan Agresti. 1992. A survey of exact inference for contingency tables. *Statistical science* 7, 1 (1992), 131–153.

[2] John Arrowsmith and Philip Miller. 2013. Trial watch: phase II and phase III attrition rates 2011-2012. *Nature reviews. Drug discovery* 12, 8 (2013), 569.

[3] Juan M Banda, Lee Evans, Rami S Vanguri, Nicholas P Tatonetti, Patrick B Ryan, and Nigam H Shah. 2016. A curated and standardized adverse drug event resource to accelerate drug safety research. *Scientific data* 3, 1 (2016), 1–11.

[4] François Belleau, Marc-Alexandre Nolin, Nicole Tourigny, Philippe Rigault, and Jean Morissette. 2008. Bio2RDF: towards a mashup to build bioinformatics knowledge systems. *Journal of biomedical informatics* 41, 5 (2008), 706–716.

[5] Aurel Cami, Alana Arnold, Shannon Manzi, and Ben Reis. 2011. Predicting adverse drug events using pharmacological network models. *Science translational medicine* 3, 114 (2011), 114ra127–114ra127.

[6] YZ Chen and CY Ung. 2001. Prediction of potential toxicity and side effect protein targets of a small molecule by a ligand–protein inverse docking approach. *Journal of Molecular Graphics and Modelling* 20, 3 (2001), 199–218.

Table 2: Ablation studies on shallow *versus* deep layers, and latent collaborative filtering *versus* content-based methods. * means statistically significant results (p-value ≤ 0.01) using the paired-t-test compared with the best baseline.

	Model architecture				Liu's		AEOLUS	
Method	Shallow layers	Deep layers	Latent vector	Drug descriptor	AUC ($\times 10^{-2}$)	AUPR ($\times 10^{-2}$)	AUC ($\times 10^{-2}$)	AUPR ($\times 10^{-2}$)
MF	✓		✓		90.89 ± 0.26	45.00 ± 0.65	87.15 ± 0.37	61.03 ± 1.11
NCF		✓	✓		89.84 ± 0.36	44.00 ± 0.67	84.68 ± 0.41	59.35 ± 1.13
NMF	✓	✓	✓		91.21 ± 0.37	46.73 ± 0.67	87.16 ± 0.32	61.03 ± 1.11
LR	✓			✓	89.25 ± 0.23	46.54 ± 0.83	89.02 ± 1.02	66.89 ± 2.29
MLP		✓		✓	90.23 ± 0.29	45.64 ± 0.79	89.48 ± 0.08	68.65 ± 1.19
Wide&Deep	✓	✓		✓	91.42 ± 0.25	50.23 ± 0.85	90.50 ± 0.45	72.38 ± 1.11
ADRNet w/o share	✓	✓	✓	✓	92.16 ± 0.21*	51.71 ± 0.93*	91.14 ± 0.40*	73.91 ± 1.09*
ADRNet	✓	✓	✓	✓	**92.23 ± 0.21***	**51.72 ± 0.89***	**91.20 ± 0.42***	**74.21 ± 1.04***

[7] Heng-Tze Cheng, Levent Koc, Jeremiah Harmsen, Tal Shaked, Tushar Chandra, Hrishi Aradhye, Glen Anderson, Greg Corrado, Wei Chai, Mustafa Ispir, et al. 2016. Wide & deep learning for recommender systems. In *Proceedings of the 1st workshop on deep learning for recommender systems*. 7–10.

[8] R Todeschiniand V Consonni. 2009. Handbook of Molecular Descriptors, vol. 11.

[9] Behrooz Davazdahemami and Dursun Delen. 2018. A chronological pharmacovigilance network analytics approach for predicting adverse drug events. *Journal of the American Medical Informatics Association* 25, 10 (2018), 1311–1321.

[10] Giovanna Maria Dimitri and Pietro Lió. 2017. DrugClust: a machine learning approach for drugs side effects prediction. *Computational biology and chemistry* 68 (2017), 204–210.

[11] I Ralph Edwards and Jeffrey K Aronson. 2000. Adverse drug reactions: definitions, diagnosis, and management. *The lancet* 356, 9237 (2000), 1255–1259.

[12] Anton F Fliri, William T Loging, Peter F Thadeio, and Robert A Volkmann. 2005. Analysis of drug-induced effect patterns to link structure and side effects of medicines. *Nature chemical biology* 1, 7 (2005), 389–397.

[13] Francesca Grisoni, Davide Ballabio, Roberto Todeschini, and Viviana Consonni. 2018. Molecular descriptors for structure–activity applications: a hands-on approach. In *Computational Toxicology*. Springer, 3–53.

[14] Huifeng Guo, Ruiming Tang, Yunming Ye, Zhenguo Li, and Xiuqiang He. 2017. DeepFM: a factorization-machine based neural network for CTR prediction. *arXiv preprint arXiv:1703.04247* (2017).

[15] Felix Hammann, Heike Gutmann, Nadine Vogt, Christoph Helma, and Juergen Drewe. 2010. Prediction of adverse drug reactions using decision tree modeling. *Clinical Pharmacology & Therapeutics* 88, 1 (2010), 52–59.

[16] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural collaborative filtering. In *Proceedings of the 26th international conference on world wide web*. 173–182.

[17] Keith A Houck and Robert J Kavlock. 2008. Understanding mechanisms of toxicity: insights from drug discovery research. *Toxicology and applied pharmacology* 227, 2 (2008), 163–178.

[18] Md Jamiul Jahid and Jianhua Ruan. 2013. An ensemble approach for drug side effect prediction. In *2013 IEEE International Conference on Bioinformatics and Biomedicine*. IEEE, 440–445.

[19] Sarvnaz Karimi, Chen Wang, Alejandro Metke-Jimenez, Raj Gaire, and Cecile Paris. 2015. Text and data mining techniques in adverse drug reaction detection. *ACM Computing Surveys (CSUR)* 47, 4 (2015), 1–39.

[20] Sunghwan Kim, Paul A Thiessen, Evan E Bolton, Jie Chen, Gang Fu, Asta Gindulyte, Lianyi Han, Jane He, Siqian He, Benjamin A Shoemaker, et al. 2016. PubChem substance and compound databases. *Nucleic acids research* 44, D1 (2016), D1202–D1213.

[21] Yehuda Koren, Robert Bell, and Chris Volinsky. 2009. Matrix factorization techniques for recommender systems. *Computer* 42, 8 (2009), 30–37.

[22] Mei Liu, Yonghui Wu, Yukun Chen, Jingchun Sun, Zhongming Zhao, Xue-wen Chen, Michael Edwin Matheny, and Hua Xu. 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 (2012), e28–e35.

[23] Tal Lorberbaum, Kevin J Sampson, Raymond L Woosley, Robert S Kass, and Nicholas P Tatonetti. 2016. An integrative data science pipeline to identify novel drug interactions that prolong the QT interval. *Drug safety* 39, 5 (2016), 433–441.

[24] Radhakrishnan Mahadevan and Chrisophe H Schilling. 2003. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metabolic engineering* 5, 4 (2003), 264–276.

[25] Kelong Mao, Jieming Zhu, Xi Xiao, Biao Lu, Zhaowei Wang, and Xiuqiang He. 2021. UltraGCN: ultra simplification of graph convolutional networks for recommendation. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. 1253–1262.

[26] Jean-Louis Montastruc, Agnès Sommet, Haleh Bagheri, and Maryse Lapeyre-Mestre. 2011. Benefits and strengths of the disproportionality analysis for identification of adverse drug reactions in a pharmacovigilance database. *British journal of clinical pharmacology* 72, 6 (2011), 905.

[27] Emir Muñoz, Vít Nováček, and Pierre-Yves Vandenbussche. 2016. Using drug similarities for discovery of possible adverse reactions. In *AMIA Annual Symposium Proceedings*, Vol. 2016. American Medical Informatics Association, 924.

[28] Duc Anh Nguyen, Canh Hao Nguyen, and Hiroshi Mamitsuka. 2021. A survey on adverse drug reaction studies: data, tasks and machine learning methods. *Briefings in Bioinformatics* 22, 1 (2021), 164–177.

[29] Edouard Pauwels, Véronique Stoven, and Yoshihiro Yamanishi. 2011. Predicting drug side-effect profiles: a chemical fragment-based approach. *BMC bioinformatics* 12, 1 (2011), 1–13.

[30] Yanru Qu, Han Cai, Kan Ren, Weinan Zhang, Yong Yu, Ying Wen, and Jun Wang. 2016. Product-based neural networks for user response prediction. In *2016 IEEE 16th International Conference on Data Mining (ICDM)*. IEEE, 1149–1154.

[31] Suvash Sedhain, Aditya Krishna Menon, Scott Sanner, and Lexing Xie. 2015. Autorec: Autoencoders meet collaborative filtering. In *Proceedings of the 24th international conference on World Wide Web*. 111–112.

[32] Itay Shaked, Matthew A Oberhardt, Nir Atias, Roded Sharan, and Eytan Ruppin. 2016. Metabolic network prediction of drug side effects. *Cell systems* 2, 3 (2016), 209–213.

[33] Ying Shan, T Ryan Hoens, Jian Jiao, Haijing Wang, Dong Yu, and JC Mao. 2016. Deep crossing: Web-scale modeling without manually crafted combinatorial features. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 255–262.

[34] Nicholas P Tatonetti, Tianyun Liu, and Russ B Altman. 2009. Predicting drug side-effects by chemical systems biology. *Genome biology* 10, 9 (2009), 1–4.

[35] Bernard Testa and Lemont B Kier. 1991. The concept of molecular structure in structure–activity relationship studies and drug design. *Medicinal research reviews* 11, 1 (1991), 35–48.

[36] Angelo Vedani, Max Dobler, and Markus A Lill. 2005. In silico prediction of harmful effects triggered by drugs and chemicals. *Toxicology and Applied Pharmacology* 207, 2 (2005), 398–407.

[37] Chi-Shiang Wang, Pei-Ju Lin, Ching-Lan Cheng, Shu-Hua Tai, Yea-Huei Kao Yang, Jung-Hsien Chiang, et al. 2019. Detecting potential adverse drug reactions using a deep neural network model. *Journal of medical Internet research* 21, 2 (2019), e11016.

[38] Ruoxi Wang, Bin Fu, Gang Fu, and Mingliang Wang. 2017. Deep & cross network for ad click predictions. In *Proceedings of the ADKDD'17*. 1–7.

[39] Tai-Yin Wu, Min-Hua Jen, Alex Bottle, Mariam Molokhia, Paul Aylin, Derek Bell, and Azeem Majeed. 2010. Ten-year trends in hospital admissions for adverse drug reactions in England 1999–2009. *Journal of the Royal Society of Medicine* 103, 6 (2010), 239–250.

[40] Cao Xiao, Ping Zhang, W Chaovalitwongse, Jianying Hu, and Fei Wang. 2017. Adverse drug reaction prediction with symbolic latent dirichlet allocation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 31.

[41] Jun Xiao, Hao Ye, Xiangnan He, Hanwang Zhang, Fei Wu, and Tat-Seng Chua. 2017. Attentional factorization machines: Learning the weight of feature interactions via attention networks. *arXiv preprint arXiv:1708.04617* (2017).

[42] Yoshihiro Yamanishi, Edouard Pauwels, and Masaaki Kotera. 2012. Drug side-effect prediction based on the integration of chemical and biological spaces. *Journal of chemical information and modeling* 52, 12 (2012), 3284–3292.

[43] Weinan Zhang, Tianming Du, and Jun Wang. 2016. Deep learning over multi-field categorical data. In *European conference on information retrieval*. Springer, 45–57.

[44] Wen Zhang, Feng Liu, Longqiang Luo, and Jingxia Zhang. 2015. Predicting drug side effects by multi-label learning and ensemble learning. *BMC bioinformatics* 16, 1 (2015), 1–11.