RESEARCH



A novel weighted pseudo-labeling framework based on matrix factorization for adverse drug reaction prediction



Junheng Chen¹, Fangfang Han^{1,2,3*}, Mingxiu He¹, Yiyang Shi¹ and Yongming Cai^{1,2,3*}

*Correspondence: hanff@gdpu.edu.cn; cym@gdpu. edu.cn

 ¹ School of Medical Information and Engineering, Guangdong Pharmaceutical University, Guangzhou 510006, China
 ² NMPA Key Laboratory for Technology Research and Evaluation of Pharmacovigilance, Guangzhou 510006, China
 ³ Guangdong Provincial Traditional Chinese Medicine Precision Medicine Big Data Engineering Technology Research Center, Guangzhou 510006, China

Abstract

Adverse drug reactions (ADRs) are among the global public health events that seriously endanger human life and cause high economic burdens. Therefore, predicting the possibility of their occurrence and taking early and effective response measures is of great significance. Constructing a correlation matrix between drugs and their adverse reactions, followed by effective correlation data mining, is one of the current strategies to predict ADRs using accessible public data. Since the number of known ADRs in real-world data is far less than the number of their unknown counterparts, the drug-ADR association matrix is very sparse, which greatly affects the classification performance of machine learning methods. To effectively address the problem of sparsity, we proposed a novel weighted pseudo-labeling framework that mines potential unknown drug-ADR pairs by integrating multiple weighted matrix factorization (MF) models and treating them as pseudo-labeled drug-ADR pairs. Pseudo-labeled data is added to the training set, and the MF model is fine-tuned to improve the classification performance. To prevent overfitting to easily found pseudo-labels and improve the quality of pseudo-labels, a novel weighting approach for pseudo-labels was adopted. This paper reproduces the baselines under the same experimental conditions to evaluate the performance of the proposed method on sparse data from the Side Effect Resource (SIDER) database. Experimental results showed that our method outperformed other baselines in the Area Under Precision-Recall and F1-scores and still maintained the best performance in sparser scenarios. Furthermore, we conducted a case study, and the results showed that our proposed framework efficiently predicted ADRs in the real world.

Keywords: Adverse drug reaction prediction, Weighted pseudo-labeling, Matrix factorization, Semi-supervised learning

Background

Adverse drug reactions (ADRs) are one of the major factors causing high morbidity and mortality during treatment and are also a source of economic burden for the medical system [1]. According to the recently released U.S. Food and Drug Administration (FDA) data, the total number of ADR reports in 2022 reached 2,347,431, with



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

the highest proportion of cases reporting serious ADRs reaching 53.86%, and the death rate reaching 7.46%. In vitro and in vivo experiments in the preclinical stage, human safety testing in the clinical stage, and monitoring of ADRs after marketing should be evaluated for drug safety to reduce the negative impact of ADRs [2]. However, it is difficult to fully detect ADRs before the drug is approved for marketing due to the insufficient number of volunteers in the clinical stage. According to statisticians, monitoring approximately 10,000–20,000 patients during clinical trials to rule out an incidence rate of 1/3000–1/6000 ADRs is undoubtedly difficult [3]. While it is impossible to verify all ADRs through experiments, using machine learning methods to predict unknown ADRs that may occur presents a feasible solution.

Through our investigation, we discovered that Matrix factorization (MF) and Neural Networks (NNs) are the two most commonly used methods for predicting ADRs in recent studies. Firstly, Matrix factorization is a straightforward and effective technique utilized in ADR prediction tasks. Galeano et al. [4] decomposed the drug-ADR association matrix into a drug latent matrix and an adverse reaction latent matrix by MF, and achieved an AUPR of 0.342 for the reconstructed drug-ADR association matrix. In 2015, Li et al. [5] integrated drug-target interactions as features and combined them with matrix decomposition to predict ADRs. In 2018, Zhang et al. [6] proposed a featuredriven graph regularized matrix factorization (FGRMF) model, which predicts ADRs by integrating multiple types of similarities, including drug structure similarity and associated protein similarity. Fukuto et al. [7] proposed the LogitMF model, which employs Kernel principal component analysis (KPCA) to reduce the dimensionality of the 2048-dimensional extended-connectivity fingerprints (ECFP) features of the drug to 100 dimensions, subsequently inputting them into the MF model. In addition to MF-based methods, Neural Network-based methods are also widely used for ADR prediction. Convolutional Neural Networks (CNNs), which have outperformed in image classification tasks, were employed by Uner [8] in DeepSide model to extract similarity information between drugs, achieving satisfactory prediction results. Molecular graphs are key data for representing drug structures. And Graph Neural Networks (GNNs) are used to encode the molecular graphs because they can capture complex relationships and topologies within graphs through massage passing and readout mechanisms [9]. A recent model, named idse-HE, encodes molecular graphs and drug-ADR bipartite graphs using GNNs and predicts ADRs by reconstructing the drug-ADR association matrix [10]. Xu et al. [11] developed a Graph Attention Network to predict the frequencies of ADRs and achieved an AUC of 0.922. Recurrent Neural Networks (RNNs) and Transformers have been widely used for extracting drug-ADR association information from text, thereby providing more information for ADR prediction [12–14].

Although many improved methods for ADR prediction have been proposed, there are still challenges associated with the above two basic methods. For NNs, there is the issue of unavoidable negative sample selection errors, and for MF, there is the problem of matrix sparsity. Most models require converting the ADR prediction into a binary classification task and treat unknown drug-ADR pairs as negative samples. Unknown drug-ADR pairs include (1) cases where the drug does not cause the ADR, (2) cases where the drug can cause an ADR, and this fact is known but missing from the source database, and (3) cases where the drug can cause the ADR but is not yet known [15]. Therefore,

treating unknown drug-ADR pairs as negative samples would undoubtedly introduce bias in model training. For NNs, data balancing methods and treating indications of drug as negative samples are commonly used in ADR prediction and have achieved great performance on balanced test sets [8, 15]. However, the performance of NNs-based methods in ADR prediction under imbalanced conditions needs to be verified. Additionally, the limited number of drug-indication pairs may not comprehensively capture the negative samples. Recent studies show that MF-based methods perform well in imbalanced situations, constructing all known and unknown ADR pairs into a drug-ADR association matrix [10-13]. Despite this, they are still influenced by noisy labels and matrix sparsity. Recent studies have demonstrated that employing semi-supervised learning is an effective approach to addressing these challenges. Ding et al. [16] proposed the Maximize the Cosine Similarity-based Multiple Kernel Learning method (MCS-MKL), which improves the performance of MF in predicting ADRs by using soft-labels with Weighted K Nearest Known Neighbors (WKNKN) to preprocess unknown drug-ADR pairs. Based on the common assumption that similar drugs tend to exhibit similar interaction and non-interaction patterns with ADRs, MCS-MKL assigns probabilities to unknown drug-ADR pairs by calculating their similarity with the K most similar drugs. Although more prior knowledge is given to unknown drug-ADR pairs through soft labels, they are still treated as negatives. An intuitive solution to these challenges is to treat unknown drug-ADR pairs as unknown rather than negative. Unlike soft-labeling, pseudo-labeling uses an existing model to make predictions on unlabeled data and assigns those with high confidence as labels for this data. However, it is often used in conjunction with NNs [17, 18].

Inspired by the success of MF on ADR prediction and the success of pseudo-labels in addressing the label scarcity problem [19], this study explores the possibility of combining pseudo-labeling with MF model, which could play a significant role in ADR prediction tasks. Unlike NN-based methods, the output of a MF-based method consists of a limited number of elements. Drugs or ADRs that appear more frequently in the database are more easily predicted by MF, leading to a concentration of pseudo-labels on highfrequency drug-ADR pairs and overfitting of the model to these pairs (Supplement 2). To better combine the MF and pseudo-labeling, this study proposes a novel weighted pseudo-labeling framework with MF (WPLMF), aiming to predict potential unknown drug-ADR pairs by leveraging multiple MF models. In the WPLMF framework, node-2vec was initially used to embed drugs from the knowledge graph to capture their biological information. WPLMF treats positive samples predicted by MF from unknown drug-ADR pairs as pseudo-labels. Subsequently, these pseudo-labels are assigned different weights based on their prediction scores to prevent the model from overfitting to pseudo-labels that are easily discoverable. Pseudo-labels were then incorporated into the training set to fine-tune the MF model and optimize the classification hyperplane, thereby providing additional learnable data for the prediction model and addressing the matrix sparsity issue of the drug-ADR association matrix. The main contributions of this study are as follows. First, unlike traditional pseudo-labels combined with NNs, this study explored the possibility of combining MF with pseudo-labels. Second, we proposed novel pseudo-label weighting methods to address the issue of pseudo-labels concentrating on high-frequency adverse drug reactions when MF is combined with pseudo-labels. Finally, we proposed a method that combines pseudo-labeling with MF to solve the matrix sparsity problem of the drug-ADR association matrix, thereby improving the performance of ADR prediction.

Materials and methods

Datasets

The data were collected from the DrugBank (https://go.drugbank.com) and Side Effect Resource (SIDER, http://sideeffects.embl.de) databases. The source data was downloaded from SIDER and Drugbank, which contained 1430 and 14,315 drugs, respectively, with SIDER also including 5868 ADRs. Additionally, Drugbank provided extra drugprotein interaction information, with proteins consisting of targets, enzymes, transporters, and carriers. During the data preprocessing phase, drugs that could not be matched between the SIDER and DrugBank databases were removed, such as 'x' and 'v', among others. The ADRs were mapped to the preferred term (PT) using the Medical Dictionary for Regulatory Activities (MedDRA). Then we retained the proteins from the Drug-Bank that relate to the drugs that in our dataset. The final dataset included 1177 drugs and 4247 ADRs. Then all drugs and ADRs were constructed into an 1177×4247 drug-ADR association matrix M. If drug *i* can cause ADRs *j*, then the elements in the matrix $M_{ii} = 1$, otherwise $M_{ii} = 0$. After the statistics of the dataset, the ratio of the number of known drug-ADR pairs to the number of unknown drug-ADR pairs was approximately 0.027, which shows that the drug-ADR association matrix was sparse. In addition, our dataset contains 1749 proteins, including targets, enzymes, transporters, and carriers, as well as 10,715 drug-protein pairs. There were 49 types of drug-protein interactions in our dataset, such as inhibitors and substrates, among others.

To obtain more robust model evaluation results and prove the applicability of the proposed framework to different datasets, the SIDER4 dataset [20] was used as additional data for evaluation. The details on the benchmark datasets are shown in Table 1. The SIDER4 dataset comprises 5579 Lowest Level Term (LLT) ADRs as defined by Med-DRA, which differ from our preferred term (PT) ADRs. Hence, the number of ADRs in our dataset was lower than that in the SIDER4 dataset. In addition, the SIDER4 dataset is sparser than our compiled dataset, with the ratio of the number of known drug-ADR pairs to the number of unknown drug-ADR pairs being approximately 0.020.

Pseudo-labeling framework

Overview of pseudo-labeling framework

To solve the problem of matrix sparseness, we propose a pseudo-labeling framework to achieve data augmentation so that the MF model can obtain more reliable data to train and receive better performance. The overview of our proposed pseudo-labeling framework is shown in Fig. 1. The framework includes four crucial processes: (a) representation of drugs and ADRs; (b) pseudo-labeling; (c) label weighting; and (d) acquisition of the best prediction results.

Process a. Since the drug-ADR association matrix is sparse, it is difficult to effectively mine pseudo-labels using only the MF model, such as facing the cold start problem. Therefore, drug-protein association can be used as a feature to improve the efficiency of pseudo-label mining. We constructed a biological knowledge graph

 Table 1
 Details about benchmark datasets

Dataset	Drug	ADR	Substructure	Target	Transporter	Enzyme	Pathway	Indication	Carrier	Relation
Ours	1177	4247	881	1299	A.N	328	A.N	1546	122	51
SIDER4	1080	5579	881	1046	96	160	268	2537	N.A	9
N A means data	is not available									



Fig. 1 Overview of the pseudo-labeling framework

(KG) containing proteins and drugs, then used the Node2vec [21] algorithm for node embedding to obtain potential features of drugs in preparation for improving the accuracy of initial pseudo-label mining.

Process b. A generalized matrix factorization model (GMF) is used to predict ADRs. If the model's predicted probability for an element originally labeled 0 in the drug-ADR association matrix is higher than the preset threshold, it is regarded as a noisy label and relabeled as 1. It is noteworthy that multiple GMF models were used to predict ADRs simultaneously to obtain more robust prediction results.

Process c. To reduce the impact of erroneous pseudo-labels, we propose a novel weighting scheme to adjust the contribution of different samples to the loss by controlling the weight of the labels to achieve better model fine-tuning.

Process d. The mean ensemble was applied to combine all GMF model prediction results.

Processes b to *d* are reiterated until the output from *process d* is optimal, and the output is regarded as the final prediction result.

Knowledge graph representation

In recent years, KGs have been widely used in ADR prediction. Joshi et al. and Zhang et al. [7, 8] showed that using word2vec-based methods to embed biological maps of drug-associated proteins can effectively improve the performance of ADR prediction. From a biological perspective, incorporating drug-protein associations into the model is efficacious. Off-target effects are one of the causes of ADRs. Besides acting on targets related to disease treatment, drugs might also exert effects on other non-therapeutic targets, which could give rise to adverse effects. For instance, tyrosine kinase inhibitors (TKIs) are often not specific to the intended tyrosine kinase, and thus unintended off-target effects can occur. Some of these off-target effects can lead to cardiotoxicity, particularly arrhythmias [22]. To enhance the efficiency of initial pseudo-labels mining, we first obtained effective representations of drug and ADR features by embedding a biological KG before implementing MF. Specifically, we constructed this biological KG using data from the DrugBank and SIDER databases, which included entities such as drugs, indications (ADRs), targets, carriers, and enzymes. We then employed Node2vec, a variant of the Word2vec algorithm, to embed the nodes within the KG.

The Node2vec calculates the transition probabilities for all pairs of nodes that exist in the KG using a fixed length *l* of the node sequences. Assuming $c_0 = u$ is the starting node of a sequence, c_i represents the *i*th node, and its probability coming from c_{i-1} is:

$$P(c_i = x | c_{i-1} = v) = \begin{cases} \frac{\pi_{vx}}{Z} & \text{if } (v, x) \in E\\ 0 & \text{otherwise} \end{cases}$$
(1)

where π_{vx} is the non-normalized transfer probability from node v to x, and Z is the normalized term. Nevertheless, the transition probabilities above cannot fully consider the homogeneity and structure of the graph. A factor α_{pq} is utilized to control the random walk strategy. Specifically, $\alpha_{pq}(t, x)\pi_{vx}$ takes place of π_{vx} in *formula* 1.

$$\alpha_{pq}(t,x) = \begin{cases} 1/p \ if \ d_{tx} = 0\\ 1 \ if \ d_{tx} = 1\\ 1/q \ if \ d_{tx} = 2 \end{cases}$$
(2)

where *t* is the previous node c_{i-1} and *x* is the next node c_{i+1} . If node *t* and node *x* are the same, $d_{tx} = 0$; if node *t* and node *x* are connected directly, $d_{tx} = 1$; otherwise, $d_{tx} = 2$. *p* and *q* are two hyperparameters. When the value of *p* is large (> max(q, 1)), the current node c_i is more likely to wander to the previous node t to capture the structure of the graph. If *p* is small (< min(q, 1)), it is more likely to return to the previous node t capture homogeneity. For the parameter *q*, if *q* > 1, it tends to wander near the previous node t, and if *q* < 1, it tends to wander to a farther node. Finally, the skip-gram is used to train all sequences obtained through partial random walks to obtain the latent feature representation of all nodes.

Pseudo-labeling

MF is one of the effective methods for recommendation systems and ADR prediction [4, 23, 24]. Therefore, we use MF to achieve efficient pseudo-label prediction. This model is classically modeled as:

$$Y \approx Y' = PQ^T \tag{3}$$

where *P* is the $m \times k$ drug feature matrix, which contains the k-dimensional features of all *m* drugs. Q is an $n \times k$ ADR feature matrix, which contains the k-dimensional features of all *n* ADRs, $k \ll \min\{m, n\}$.

We used the GMF model with better performance to replace the traditional MF model. The definition of GMF is shown in Eq. (5), and binary cross-entropy loss is used as the objective function.

$$y_{ij} \approx y_{ij'} = \sigma\left(\left(p_i \odot q_j\right) w^T\right) \tag{4}$$

$$loss = -\sum_{i} \sum_{j} y_{ij} \log (y_{ij'}) + (1 - y_{ij}) \log (1 - y_{ij'})$$
(5)

where *w* is a shared $k \times 1$ dimensional parameter matrix, p_i and q_j are the representation of the *i* th drug and the representation of the *j* th ADR respectively. $\sigma(\cdot)$ is the sigmoid activation function, which converts the predicted value into a probability value between

0 and 1. Drug-ADR pairs originally labeled as 0 in the matrix but whose output value is higher than the preset threshold is regarded as pseudo-labels, and their labels are relabeled as 1 to form a new drug-ADR association matrix for fine-tuning the GMF model.

Threshold adjustment is an effective method to achieve a trade-off between pseudolabel quality and quantity [25]. However, this method may not be suitable for ADR prediction because numerous unknown drug-ADR pairs exist in the matrix. A fixed threshold was utilized to screen suitable pseudo-labels. With the addition of pseudolabels, the classification hyperplane is optimized to a certain extent, and the overall prediction score by the model increases, eventually leading to more pseudo-labels exceeding the threshold, as shown in Supplement1(a). Therefore, using a fixed threshold to filter pseudo-labels can still make pseudo-labels play a similar role as curriculum learning [26]. Specifically, simple and high-confidence samples in the early stages of training are labeled as pseudo-labels to fine-tune the model first, thereby guiding the model to achieve a better optimal solution.

As shown in Supplement1 (b), when GMF was trained with 50 epochs, the hitting rate of the pseudo-labels was the highest, and pseudo-labels with a higher hitting rate can better improve the pseudo-label method. Therefore, we trained each GMF model with 50 epochs.

Label weighting

In Sect. Pseudo labeling, we mentioned that the number of pseudo-labels will gradually increase with the fine-tuning training of GMF. This leads to the problem of excessive labeling, which reduces the quality of pseudo-labels and ultimately results in poor model performance. GMF is more likely to predict high-frequency ADRs (i.e., drugs or ADRs that occur frequently in the dataset), causing pseudo-labels to concentrate on high-frequency drug-ADR pairs and leading to overfitting of GMF to such pseudo-labels. To address this issue, we propose a novel weighting method inspired by Focal loss and Error L2-Norm (EL2N) [27, 28]. This method is designed to mitigate the influence of errone-ous pseudo-labels and overfitting to easily detectable pseudo-labels by assigning appropriate weights to them. Specifically, the weight of each label is defined as follows:

$$w_{ij} = \begin{cases} -\left(\alpha^{\gamma} \ln\left(y_{ij'} + \varepsilon\right) + (1 - \alpha)^{\gamma} \ln\left(1 - y_{ij'} + \varepsilon\right)\right) & x \text{ is pseudo} \\ 1/t & otherwise \end{cases}$$
(6)

where α is the penalty factor, which makes the model concentrate on the quality of pseudo-labels rather than the quantity; γ is the adjustment factor; ε is the minimum value; and $y_{ij'}$ and $1 - y_{ij'}$ are the forms of EL2N scores in the binary classification problem. Since the label before the pseudo-labeling is marked as 0, $y_{ij'}$ is used to measure the distance of the model output to the original label, while $1 - y_{ij'}$ is the distance to the output of 1. The smallest weight of the pseudo-label with a model output is $y_{min'} \approx \frac{\alpha^{\gamma}}{\alpha^{\gamma} - (1-\alpha)^{\gamma}}$. We treat $y_{min'}$ as a dividing point; when the prediction score of the pseudo-label is higher than $y_{min'}$, the weight of the pseudo-label becomes larger, and vice versa. Thus, at large values of alpha, pseudo-labels with relatively high scores are assigned low weights, while those with relatively low scores are assigned high weights. *t* represents the current number of times of pseudo-label labeling. For non-pseudo-label

samples, a weight of 1/t is given to prevent overfitting. After combining the weights, the objective function is redefined as:

$$loss = -\sum_{i} \sum_{j} w_{ij}(y_{ij} \log (y_{ij'}) + (1 - y_{ij}) \log (1 - y_{ij'}))$$
(7)

Model evaluation

The MF model often treats ADR prediction as a binary classification problem, that is, pairs of existing drug-ADR pairs are regarded as positive samples, while pairs of unknown ADRs are regarded as negative samples. As mentioned in Sect. 2.1, there are far more unknown drug-ADR pairs than known ones; thus, our drug-ADR association matrix for training is extremely sparse. The following indicators were used to evaluate the performance of the model.

AUPR: Area Under Precision-Recall (AUPR) curve, which is widely used in imbalance problems.

F1: The F1 score can be interpreted as a harmonic mean of the precision and recall.

MRR: Mean Reciprocal Rank, which focuses on measuring the accuracy of highly predicted samples. For better presentation, we use the accumulation of the reciprocal ranking instead of taking the average.

Precision: Precision indicates the proportion of positive identifications that are actually correct.

Recall: Recall is the proportion of actual positive cases that are correctly identified by the model.

Precision@15: Top15 precision, macro mean precision of the top 15 ADRs of all drugs. *Recall@15*: Top 15 recall, macro mean recall of the top 15 ADRs of all drugs. The metrics above are formulated as follows:

$$F1 = 2 \times \frac{precision \times recall}{precision + recall}$$
(8)

$$MRR = \frac{1}{N} \sum_{i=1}^{|N|} \frac{1}{rank_i}$$
(9)

We believe that to predict ADRs, more attention should be paid to ADRs with higher model prediction scores because they can better provide warnings for safe drug use. Therefore, we used Precision@15 and Recall@15 to replace the traditional Precision and Recall.

Results and discussion

Experiment setup

The present study used five-fold cross-validation to evaluate the classification prediction performance of the model. In each fold, we constructed a biological KG using all drug-protein pairs in the training set. Then Node2vec with five different walking strategies or parameters was used to embed the nodes of the biological KG, and five GMF models were initialized with the obtained five different sets of initial representations of drugs and ADRs. Five different parameter settings of Node2vec were as follows: p = [0.5, 0.7, 1, 1.5, 2] and q = [2, 1.5, 1, 0.7, 0.5]. The length of the Node2vec walking sequence was set to 5, and each node obtained 100 sequences through partial random walk; the embedding dimension of the node was 12. The hyper-parameters p and q were aimed at capturing homophily or structural equivalence from the KG, thereby providing different representations of drugs and improving the overall generalization ability of the model.

For the pseudo-labeling framework, five GMF models were tuned using an Adam optimizer and 0.02 learning rate to minimize the cross-entropy loss with 50 epochs at each step. Then the pseudo-labels were marked and added to the training set. The threshold for labeling pseudo-labels was set to 0.8. Alpha and gamma were set to 0.9 and 2 respectively. The entire framework was implemented on PyTorch.

The detailed settings for the baseline methods are shown in Table 2:

Performance analysis and comparison

In this section, we compared the performance of WPLMF with other baselines. In FGRMF, a graph was constructed based on a drug features, and graph regularization, which preserves the structure of the drug, is incorporated into FGRMF to fully take advantage of the drug's characteristics [12]. idse-HE captures the drug-ADR association using a GNN, and multi-view structure features of drugs, transformed by MPNN+Set-2Set and molecular fingerprint algorithms, are used to provide abundant features [15]. Galeano's method projects the drug-ADR relationship into a low-dimensional space, uncovering the latent features of drugs and side effects [10]. LogitMF employed Kernel Principal Component Analysis (KPCA) to reduce the dimensionality of the drug's 2048-dimensional extended-connectivity fingerprints (ECFP) features to 100 dimensions, thereby removing redundant structural features to enhance performance [13]. MSC-MKL improved the performance of MF in predicting ADRs by using soft-labeling with Weighted K Nearest Known Neighbors to preprocess unknown drug-ADR pairs and employing multi-kernel learning to capture more information through different similarity matrices of drugs and ADRs [21].

Existing MF algorithms typically treat ADR prediction as a binary classification problem. However, unlabeled data far exceeds labeled data, which limits their effectiveness. Therefore, we propose a new weighted pseudo-labeling MF framework that can address

	5	
Method	Features	Setting
MCS-MKL	CS, DAA	$\mu = 2e - 3, v = 2e - 5, k = 17$
FGRMF	CS, DAA, DPI	$\mu = 8, \lambda = 4$
idse-HE	CS, DAA	$lr = 1e-3$, dropout = 0.1, $a = 0.02$, $\lambda = 5e-3$
Galeano's	DAA	$lr = 1e - 2, \lambda = 15$
Logit MF	CS, DAA	$lr = 1e-2, \lambda = 1e-4, \alpha = 1, \beta = 0.2$
WPLMF	DAA, DPI	lr = 1e-2, $p = [0.5, 0.7, 1, 1.5, 2]$, $q = [2, 1.5, 1, 0.7, 0.5]$, $a = 0.9$, $\gamma = 2$, threshold = 0.8

 Table 2
 The detailed settings for the baseline methods

CS Chemical Structure, DAA Drug-ADR association, DPI Drug-Protein Interaction

Method	AUPR	F1	MMR	Precision@15	Recall@15	Precision	Recall
MCS-MKL [16]	0.6428 (1.3e-3)	0.6082 (1.2e-3)	10.1247 (1.4e–2)	0.6567 (7.6e–3)	0.7813 (1.9e-2)	0.6145 (1.2e-2)	0.6082 (1.1e-2)
FGRMF [6]	0.5117 (4.5e—3)	0.4993 (2.6e-3)	9.7799 (7.6e — 2)	0.4932 (5.9e-3)	0.7842 (1.3e — 2)	0.4953 (1.4e-2)	0.5043 (1.5e — 2)
idse-HE [10]	0.5888 (1.1e–2)	0.5564 (1.1e — 2)	8.128 1(7.4e—1)	0.5676 (8.5e-3)	0.9136 (1.3e–2)	0.4738 (3.5e-2)	0.6804 (3.9e-2)
Galeano's [4]	0.6294 (2.8e-3)	0.5924 (1.9e–3)	10.0803 (3.4e-2)	0.6081 (5.2e-3)	0.7748 (1.1e — 2)	0.6049 (1.9e-2)	0.5804 (1.2e-2)
Logit MF [7]	0.6124 (3.3e-3)	0.5766 (3.5e-3)	10.0619 (3.0e-2)	0.6370 (6.6e-3)	0.7263 (5.0e-3)	0.5819(1.3e-2)	0.5714 (1.6–3)
WPLMF (ours)	0.6553 (3.0e–3)	0.6095 (1.8e–3)	10.0918 (1.9e—1)	0.6440 (11.8e–2)	0.77981 (5.5e — 3)	0.6172 (1.4e-2)	0.6024 (1.1e-2)

Table 3 Results of the proposed WPLMF framework and baselines in Ours Dataset (The best performance is highlighted in bold)

Table 4 Results of the proposed WPLMF framework and baselines in SIDER4 (The best performance is highlighted in bold)

Method	AUPR	F1	MMR	Precision@15	Recall@15	Precision	Recall
MCS-MKL	0.5747 (3.6e-3)	0.5624 (2.1e-3)	9.8289 (5.8e-2)	0.6061 (1.1e-2)	0.7741 (2.0e-2)	:0.5693 (1.2e-2)	0.5590 (1.7e–2)
FGRMF	0.5434 (4.4e-3)	0.5347 (3.4e-3)	9.7910 (1.5e–2)	0.7304 (1.0e-2)	0.5416 (6.7e-3)	0.5347 (9.2e — 3)	0.5351 (1.1e — 2)
idse-HE*	0.5303 (7.7e — 3)	0.5069 (2.6e–2)	8.1523 (7.0e — 2)	0.5228 (1.9e-2)	0.9026 (1.3e–2)	0.4267 (5.8e–2)	0.6402 (4.5e–2)
Galeano's	0.5698 (3.6e—3)	0.5457 (7.1e — 3)	9.7946 (5.8e — 2)	0.7500 (1.3e–2)	0.5567 (2.3e — 2)	0.5617 (9.4e-3)	0.5306 (1.3e–2)
Logit MF*	0.5479 (4.9e-3)	0.5402 (4.1e - 3)	9.6442 (3.9e-3)	0.7324 (1.3e-2)	0.6135 (6.7e—3)	0.5379 (1.8e–2)	0.5425 (2.1e — 2)
WPLMF (ours)	0.6031 (3.8e–3)	0.5744 (2.9e–3)	9.9028 (1.3e–2)	0.61004 (7.5e-3)	0.7816 (8.2e-3)	0.5858 (9.4e–3)	0.5635 (4.3e—3)

The asterisk (*) indicates that SMILES data is not included in the SIDER4 dataset; instead, an 881-dimensional Pubchem fingerprint is used

these limitations. Firstly, it reduces the obstacles caused by the matrix sparsity problem during training through pseudo-labeling. Secondly, pseudo-labeling is equivalent to entropy regularization, which reduces the overlap between different categories of data [29]. Thus, pseudo-labeling has certain theoretical advantages.

Tables 3 and 4 present the results of WPLMF and baselines in our dataset and SIDER4 dataset, with the average values and standard deviations obtained from a fivefold validation. WPLMF exhibits superior performance compared to all baselines, particularly in terms of AUPR and MRR. The AUPR reached 0.6553, and the F1-score reached 0.6095. These results indicate that our method excels at distinguishing existing drug-ADR pairs, attributing greater credibility to its predictions compared to other baselines. However, MCS-MKL performs better in MRR, precision@15, and recall@15. This suggests that MCS-MKL is more accurate in predicting ADRs with higher scores. In the sparser SIDER4 dataset, the proposed WPLMF still maintained the best performance, exceeding MCS-MKL in MRR, precision@15, and recall@15. Hence, our model may maintain better performance in sparse scenarios. MRR, precision@15, and recall@15 are metrics used to evaluate the performance of the model for samples with high prediction scores.

The performance of WPLMF in predicting samples with high prediction scores is poorer than that of MCS-MKL. This may stem from the fact that WPLMF shows minimal improvement in predicting the adverse effects of high-frequency drugs, as detailed in Fig. 5.

To further investigate the effectiveness of the pseudo-labeling framework in sparse scenarios, we varied the ratio of our training set from 100 to 60% and reported the F1-scores of WPLMF and all baselines. Figure 2 shows that when the data is sparser, the pseudo-labeling framework maintains the best performance, while the performance of MCS-MKL is unsatisfactory. When 40% of the known drug-ADR pairs were removed, the F1-score of MCS-MKL dropped to 0.5607, close to the performance of Galeano's method, while our model still performed better than all other baseline models. The reason for the poor performance of MCS-MKL in sparse scenarios may be that MCS-MKL relies on known drug-ADR pairs to calculate drug-drug similarity and ADR-ADR similarity. This approach reduces the available data when the ADR data is sparser, leading to poorer similarities, and eventually, the errors are transferred to the prediction results, resulting in reduced model performance. Additionally, we examined our framework without pseudo-labeling (WPLMF_NOPL) in sparse scenarios. In contrast to WPLMF, WPLMF_NOPL exhibits inferior classification performance in various sparse scenarios.



Fig. 2 Results of WPLMF and other baselines in sparse scenarios. The blue line denoted WPLMF, with the best F1-score in sparse scenarios

Nevertheless, as the matrix becomes increasingly sparser, the disparity in F1-score between them gradually diminishes. The results above show that the addition of pseudo-labels can effectively improve the performance of MF in sparse scenarios, but the effect of pseudo-labels may be limited in extremely sparse scenarios.

Trade-off between quality and quantity

The trade-off between the quality and quantity of pseudo-labels is crucial. It is difficult to improve model performance if the number of pseudo-labels is too high with poor quality, or if the pseudo-labels are of grate quality but fewer in number. The proposed WPLMF can achieve a better balance between the quality and quantity of pseudo-labels in two ways. The hit rate of pseudo-labels, which means the accuracy of the pseudo labels, was used to measure their quality. The first way is by adjusting the filtering threshold for pseudo-labels. To explore the effect of the threshold on the trade-off between quality and quantity of pseudo-labels, we examined the hit rate and number of pseudo-labels along with F1-score of WPLMF. As shown in Fig. 3, a high threshold is more likely to yield high-quality pseudo-labels. The F1-score increases until the threshold reaches 0.6, which indicates that WPLMF performs better with high-quality pseudo-labels.

Another way is to adjust the penalty coefficient α value in Formula 6. As shown in Fig. 4, we fixed $\gamma = 2$. As α increases, numerous pseudo-labels are abandoned and replaced by higher-quality ones. When α is small, pseudo-labels have a higher quantity. This is attributed to the fact that an increase in α can restrain the learning of high-confidence pseudo-labels in the model, thereby enabling the model to selectively disregard them. A high value of α can constrain the number of pseudo-labels, thus avoiding the filling of the objective drug-ADR association matrix with noisy pseudo-labels. In our experiments, when α was set to 0.8, the number of pseudo-labels was twice that when it was set to 0.1, but the hit rate was reduced by approximately 0.2. The weighting method



Fig. 3 Impact of threshold on Trade-off between quality and quantity of pseudo labels



Fig. 4 Impact of different values of α on the quality and quantity of pseudo-labels. When the hit rate is high, the color of the grid tends toward green; otherwise, it tends toward blue

TUDIC 9 Hestills of the Ablation study	(The best performance is highlighted	
Variants	AUPR	F1
Without ensemble	0.6403	0.5981
Without pseudo labeling	0.6522	0.6076
Without node2vec	0.6548	0.6082

 Table 5 Results of the Ablation Study (The best performance is highlighted in bold)

we purposed improved the F1 score of the model from 0.6082 to 0.6095 under fivefold cross-validation.

0.6553

0.6095

Above all, it is beneficial to obtain higher-quality pseudo-labels by setting a higher filtering threshold and a higher penalty coefficient α .

Ablation study

WPLMF

An ablation study was implemented to explore the real effect of the components of WPLMF on ADR prediction. We removed one of the major components of WPLMF, such as the ensemble process, and recorded the change in the model's performance. As shown in Table 5, the performance of WPLMF worsens when major components are removed. The ensemble process shows significant improvement in WPLMF.

Ensemble methods can reduce generalization errors by integrating diverse models [30]. Ensemble methods in WPLMF can improve the accuracy of the pseudo-labels, thereby improving the overall performance of WPLMF. In addition, pseudo-labeling and node2vec play key roles in WPLMF. Pseudo-labeling provides more drug-ADR association information to mitigate the impact of matrix sparsity and bias. In addition to providing more information about drug-protein associations, node2vec also differentiates between different base models in the ensemble method by setting different hyperparameters to capture homophily or structural equivalence from the drug-protein network.

Furthermore, we conducted a comprehensive study of the prediction performance of WPLMF for drugs and ADRs ranked with different frequencies. We sorted and binned the drugs or ADRs in descending order based on their frequency rankings and compared the prediction errors on the test set under different conditions: when pseudo-labels were used, when weighting was not employed, and when pseudolabels were not used. Specifically, the frequency ranking of a drug or ADR refers to the number of times the drug or ADR appears in the database. The frequency of a drug was equal to the sum of the corresponding row in the drug-ADR association matrix (Fig. 5). Similarly, the ADR frequency was calculated in the same manner. Subsequent binning involved 200 ADRs per bin for ADRs and 50 drugs per bin for drugs. The results are shown in Fig. 5: pseudo-labels mainly reduced the prediction error in high-frequency (high-ranking) drugs and high-frequency (high-ranking) adverse drug reactions (ADRs), as well as in mid-frequency (mid-ranking) drugs. However, for extremely low-frequency (low-ranking) ADRs and low-frequency (low-ranking) drugs, the model could not provide better predictions. This suggests that prediction error can be reduced by using weighted pseudo-labels, which also helps prevent the model from overfitting to easily detected pseudo-labels.



Fig. 5 Prediction errors of drugs and ADRs with different frequency rankings. Drugs or ADRs are ordered by their frequency of concurrence in descending order. NOPL: WPLMF without pseudo-labeling, NOWT: WPLMF without weighting. a Predict relevant drugs given ADRs. b Predict relevant ADRs given drugs

Case study

Considering that not all possible drug-ADR pairs are recorded in the SIDER database, traditional model evaluation indicators such as F1-score may not reflect the true performance of the model. Therefore, case studies should be performed to further evaluate the model's performance by analyzing drug-ADR pairs that are not recorded in SIDER. To achieve this, we selected the top 30 drug-ADR pairs that are not recorded in SIDER and were predicted by WPLMF for a case study. In other words, their labels in the drug-ADR matrix are 0, and the supporting evidence was screened from drugs.com and other literature databases. The case study results are shown in Fig. 6.

Of note, we assume that a given drug can causes ADR A, then ADR B may be associated with the drug in following situations:

• ADR A may lead to ADR B. For example, Octreotide has been reported to cause stomach pain, and stomach pain can induce upper abdominal pain. Therefore, we believe that Octreotide may cause epigastric pain.



Fig. 6 Illustration of the links of the top 30 drug-ADR pairs predicted as positive using our method. However, they are considered negative samples in our dataset. The blue nodes represent drugs, and the orange nodes represent ADRs. The supported drug-ADRs pairs are linked by black lines. The drug-ADR pairs connected by the red line indicate that they have not been proven so far

- ADR A and ADR B exhibit a parent–child relationship. For example, desmopressin can cause corneal edema, and corneal edema and eye edema have a father-son relationship. Therefore, we postulate that desmopressin may cause ocular edema.
- ADR B is a synonym for ADR A.

Out of the top 30 unknown drug-ADR pairs predicted by model pseudo labeling framework and original labels of 0, 23 pairs were verified through additional evidence. Among the top 10 pairs of adverse drug reactions, 8 pairs were verified; among the top 20 pairs of adverse drug reactions, 15 pairs were verified. Hence, it can be asserted that the approach introduced in this study exhibits a robust capability to forecast unknown drug-ADR associations. However, case study for top 30 drug-ADR pairs cannot fully evaluate the true performance of our method owing to the long-tail problem in SIDER, especially for ADR [31]. For this reason, a case study comprising only the top 30 predictions cannot fully evaluate our method. According to the Pareto Principle, we divide drugs or ADRs according to their frequency. The drugs or ADRs with the top 20% of the frequencies are considered head drugs or ADRs, whereas drugs or ADRs with the bottom 80% of the frequencies are considered long-tail drugs or ADR. Taking long-tail drugs as an example, we delete all drug-ADR pairs associated with head drugs in the prediction results, and randomly select 10 drug-ADR pairs predicted by our method as positive but with actual labels of negative from the remaining prediction results, and then external database is utilized to verify the selected drug-ADR pairs. The same applies to long-tail ADRs. The case study results for long-tail situation are presented in Tables 6 and 7. The former refers to a case study outcome for long-tail drug-ADR pairs, while the latter relates to long-tail ADR-drug pairs. The results demonstrate that our method achieves a precision of approximately 0.5 for long-tail drug-ADR pairs, significantly surpassing random chance. Importantly, the performance of random for long tail pairs is lower than 0.5, as they account for a small proportion of the dataset. In addition, although some examples predicted by our method have not been verified, some relevant cases have been reported. For example, WPLMF predicts that tumor lysis syndrome (TLS) is an ADR of cisplatin. According to a recent study, TLS occurred in a 58-year-old patient after transcatheter arterial infusion of cisplatin and embolization therapy for liver metastases of melanoma [32]. Therefore, we cannot completely rule out

Low-frequency Drug	ADR	Verified	Resource
Flecainide	Dysarthria	F	_
Felodipine	Hypertension	F	_
Ranolazine	Disturbance in sexual arousal	Т	Drugs.com
Diethylpropion	Syncope	Т	Drugs.com
Bimatoprost	Endophthalmitis	Т	Drugs.com
Micafungin	Weight decreased	F	-
Nortriptyline	Dyskinesia	Т	Drugs.com
Vancomycin Angioedema		Т	[33]
Dantrolene	Oedema	Т	Drugs.com
Deferasirox	Rhabdomyolysis	F	-

 Table 6
 Results of the case study for low-frequency drugs

Drug	Low-frequency ADR	Verified	Resource
Decitabine	Non-cardiac chest pain	Т	Drugs.com
Cefoxitin	Clostridium difficile colitis	F	-
Paroxetine	Glucose tolerance Decreased	F	_
Sulfasalazine	Hepatic cytolysis	Т	Drugs.com
Aripiprazole	Electrocardiogram st segment depression	Т	Drugs.com
Medroxyprogesterone	Acetate hepatic cancer	F	_
Amiodarone	Haemolytic uraemic syndrome	F	_
Lenalidomide	Non-cardiac chest pain	Т	Drugs.com
Aripiprazole	High density lipoprotein decreased	F	_
Clomiphene	Ectropion of cervix	F	_

Table 7 Results of the case study for Low frequency ADI	Table 7 Results of the	Case Study for	Low-Frequenc	y ADRs
---	--------------------------------	----------------	--------------	--------

the association between cisplatin and tumor lysis syndrome. In summary, the developed model can accurately predict real-world ADRs.

Discussion

In addition to NNs-based methods, MF-based methods have achieved remarkable success in predicting ADRs by overcoming negative sample selection errors; however, their prediction performance is significantly influenced by label noise and matrix sparsity. In this study, we propose a novel weighted pseudo-labeling framework based on MF named WPLMF for ADR prediction tasks, aimed at solving these problems. The experimental results indicate that WPLMF outperforms other methods, achieving 0.6553 in AUPR and 0.6095 in F1-score. We then further investigate the effectiveness of the proposed WPLMF framework in sparse scenarios. As the sparsity of the drug-ADR association matrix increases, MF-based methods show a decline in prediction effectiveness to varying degrees. However, WPMLF, which combines MF with pseudo-labeling, maintains the best performance in F1-score, indicating that the weighted pseudo-labels can address the matrix sparsity problem faced by MF. Moreover, the results presented in Fig. 5 indicate that the incorporation of weighted pseudo-labels can improve the prediction efficacy of high-frequency and mid-frequency drug-ADR pairs. This simultaneously and indirectly implies that the proposed method of weighting pseudo labels can prevent overfitting to high or intermediate frequency ADRs. In the ablation experiments, after removing different modules in WPLMF respectively, the model's prediction declined to varying degrees. This suggests that these modules play an important role in WPLMF. For instance, the inability of MF to obtain sufficient prior knowledge of drug-protein interactions from the KG after the deletion of node2vec may lead to vulnerability to the cold start problem when discovering pseudo-labels. In our proposed novel weighting method, the processing can achieve a trade-off between the quality and quantity of pseudo-labels by adjusting the hyperparameters. It can also reduce the weight of those pseudo-labels that are easy to find, thus preventing the model from overfitting to this kind of pseudo-labels.

Data imbalance and matrix sparsity are crucial problems in machine learning. Data imbalance or matrix sparsity may cause the model to be biased toward the majority class, thus performing poorly on the minority class. The model may over-rely on the features of the majority class and neglect the features of the minority class, leading to inaccurate predictions for the minority class [34]. Taking gradient descent as an example, the majority classes have more training iterations, so they are easier to fit than the minority classes, which may eventually lead to overfitting of the majority class or underfitting of the minority class. The proposed WPLMF further balances the two classes by adding weighted pseudo-labels to avoid overfitting of the majority class or underfitting of the minority class as far as possible. However, our approach has some limitations. Firstly, the prediction of extremely low-frequency drug-ADR pairs using our pseudolabeling approach does not achieve significant improvements. To address this problem, we plan to incorporate more relevant features, such as the structure of the drug, into the model in future work. Multi-view features have been demonstrated to offer abundant features for drug representation [35]. Hence, transforming the structures of drugs via diverse methods to acquire more abundant features might be feasible for our future work. Secondly, WPLMF cannot predict ADRs for new drugs since node2vec is trained on a knowledge graph constructed from all known drug-protein pairs. In addition, the performance of GMF, as the basic method for mining pseudo-labels in the framework proposed in this paper, may not be optimal. In the future, we will explore the use of other classification models with better performance to replace the GMF in the pseudolabeling framework and improve the performance of the framework.

Conclusion

In this study, a novel weighted pseudo-labeling framework named WPLMF for ADRs prediction is proposed. The framework consists of three crucial steps, each with distinct roles. The first step involves the use of the Node2vec algorithm to extract features from biological networks, provide additional drug features for GMF models, and reduce the impact of data sparsity on pseudo-label mining for GMF for the first time. In the second step, pseudo-labeling is conducted using GMF and mean ensemble methods. Pseudolabels are incorporated into the training set based on their scores surpassing a predefined threshold. The combined outcomes are then considered as the final prediction results. In the third step, a new weighting method is employed to achieve a trade-off between the quality and quantity of pseudo-labels and prevent the model overfitting to high-frequency pseudo-labels. The second and third steps are repeated until the prediction result of the model ensemble in the second step reaches the best performance of ADRs prediction. This model can achieve great performance even when the matrix is sparse, demonstrating the feasibility of combining MF with pseudo-labeling. Therefore, the method proposed in this study can maintain excellent prediction of ADRs when data are hard to acquire and offers a more effective approach for the discovery of ADRs.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12859-025-06053-z.

Additional file1 Additional file2 Additional file3

Acknowledgements Not applicable.

Author contributions

Junheng Chen: Conceptualization, Methodology, Validation, Investigation, Formal analysis, Data curation, Visualization, Funding acquisition, Writing—original draft, Writing—review & editing; Mingxiu He: Investigation, Formal analysis, Writing—original draft; Viyang Shi: Data curation, Formal analysis; Fangfang Han: Supervision, Writing—review & editing; Yongming Cai: Project administration, Supervision, Writing—review & editing.

Funding

This paper was funded by the Guangzhou Science and Technology Bureau 2025 Municipal School (Institute) Enterprise Joint Funding Project "Research and Development of Precision Medication Platform Based on Pharmacogenomics Data" (No. 2025A03J3712).

Availability of data and materials

The dataset and code used in this project can be freely available at: https://github.com/WholsBalance/WPLMF.

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests

The authors declare no competing interests.

Received: 2 January 2024 Accepted: 15 January 2025 Published online: 17 February 2025

References

- 1. Patton K, Borshoff DC. Adverse drug reactions. Anaesthesia. 2018;73(S1):76–84. https://doi.org/10.1111/anae.14143.
- Rao VS, Srinivas K. Modern drug discovery process: an in silico approach. J Bioinform Seq Anal. 2011;2(5):89–94.
 Nierenberg DW, Melmon KL, Morrelli H, Hoffman BB. Clinical pharmacology: basic principles in therapeutics. New
- York: McGraw-Hill Professional; 1992.
- Galeano D, Paccanaro A. A recommender system approach for predicting drug side effects. In: 2018 International Joint Conference on Neural Networks (IJCNN). IEEE; 2018:1–8. Accessed November 4, 2024. https://ieeexplore.ieee. org/abstract/document/8489025/
- Li R, Dong Y, Kuang Q, et al. Inductive matrix completion for predicting adverse drug reactions (ADRs) integrating drug-target interactions. Chemom Intell Lab Syst. 2015;144:71–9.
- Zhang W, Liu X, Chen Y, Wu W, Wang W, Li X. Feature-derived graph regularized matrix factorization for predicting drug side effects. Neurocomputing. 2018;287:154–62.
- Fukuto K, Takagi T, Tian YS. Predicting the side effects of drugs using matrix factorization on spontaneous reporting database. Sci Rep. 2021;11(1):23942.
- 8. Uner OC, Kuru HI, Cinbis RG, Tastan O, Cicek AE. DeepSide: a deep learning approach for drug side effect prediction. IEEE/ACM Trans Comput Biol Bioinf. 2022;20(1):330–9.
- Sun M, Zhao S, Gilvary C, Elemento O, Zhou J, Wang F. Graph convolutional networks for computational drug development and discovery. Brief Bioinform. 2020;21(3):919–35.
- Yu L, Cheng M, Qiu W, Xiao X, Lin W. idse-HE: Hybrid embedding graph neural network for drug side effects prediction. J Biomed Inform. 2022;131: 104098.
- 11. Xu X, Yue L, Li B, et al. DSGAT: predicting frequencies of drug side effects by graph attention networks. Brief Bioinform. 2022;23(2):bba586.
- Hiba C, Nfaoui EH, Loqman C. Fine-Tuning Transformer Models for Adverse Drug Event Identification and Extraction in Biomedical Corpora: A Comparative Study. In: Motahhir S, Bossoufi B, eds. Digital Technologies and Applications. Vol 668. Lecture Notes in Networks and Systems. Springer Nature Switzerland; 2023;957–966. https://doi.org/10. 1007/978-3-031-29857-8_95
- Keerthana PP, Roj R, Bhadra J, Dutta S. The Detection of Adverse Drug Reactions in Clinical Text Data Using Transformer Models. In: 2023 Fourth International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE). IEEE; 2023:1–6. Accessed November 5, 2024. https://ieeexplore.ieee.org/abstract/document/ 10584859/
- 14. Gupta S, Pawar S, Ramrakhiyani N, Palshikar GK, Varma V. Semi-supervised recurrent neural network for adverse drug reaction mention extraction. BMC Bioinform. 2018;19(S8):212. https://doi.org/10.1186/s12859-018-2192-4.
- Bean DM, Wu H, Iqbal E, et al. Knowledge graph prediction of unknown adverse drug reactions and validation in electronic health records. Sci Rep. 2017;7(1):16416. https://doi.org/10.1038/s41598-017-16674-x.
- 16. Ding Y, Tang J, Guo F. Identification of drug-side effect association via semi-supervised model and multiple kernel learning. IEEE J Biomed Health Inform. 2019;23(6):2619–32. https://doi.org/10.1109/JBHI.2018.2883834.
- Long P, Jian Z, Liu X. Uncertainty-Confidence Fused Pseudo-labeling for Graph Neural Networks. In: Liu Q, Wang H, Ma Z, et al., eds. Pattern Recognition and Computer Vision. Vol 14433. Lecture Notes in Computer Science. Springer Nature Singapore; 2024:331–342. https://doi.org/10.1007/978-981-99-8546-3_27
- Liu K, Ling S, Liu S. Semi-supervised medical image classification with pseudo labels using coalition similarity training. Mathematics. 2024;12(10):1537.

- Li Y, Yin J, Chen L. Informative pseudo-labeling for graph neural networks with few labels. Data Min Knowl Disc. 2023;37(1):228–54. https://doi.org/10.1007/s10618-022-00879-4.
- Zhang W, Chen Y, Tu S, Liu F, Qu Q. Drug side effect prediction through linear neighborhoods and multiple data source integration. *IEEE Int Conf Bioinformatics Biomed.* 2016: 427–434.
- 21. Grover A, Leskovec J. node2vec: scalable feature learning for networks. KDD. 2016;2016:855-64.
- Leiva O, Bohart I, Ahuja T, Park D. Off-target effects of cancer therapy on development of therapy-induced arrhythmia: a review. Cardiology. 2023;148(4):324. https://doi.org/10.1159/000529260.
- 23. Mnih A, Salakhutdinov R R. Probabilistic matrix factorization. NIPS. 2008: 1257–1264.
- 24. Koren Y, Bell R, Volinsky C. Matrix factorization techniques for recommender systems. Computer. 2009;8:30–7.
- 25. Zhang B, Wang Y, Hou W, et al. Flexmatch: boosting semi-supervised learning with curriculum pseudo labeling. Adv Neural Inf Process Syst. 2021;34:18408–19.
- Cascante-Bonilla P, Tan F, Qi Y, Ordonez V. Curriculum labeling: revisiting pseudo-labeling for semi-supervised learning. Proc AAAI Conf Artif Intell. 2021;35(8):6912–20.
- 27. Lin TY, Goyal P, Girshick R, He K, Dollar P. Focal loss for dense object detection. IEEE Trans Pattern Anal Mach Intell. 2020;42(2):318–27.
- Paul M, Ganguli S, Dziugaite GK. Deep learning on a data diet: Finding important examples early in training. Adv Neural Inf Process Syst. 2021;34:20596–607.
- 29. Lee DH. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In: Workshop on Challenges in Representation Learning, ICML. Vol 3. Atlanta; 2013. p. 896.
- Mahajan P, Uddin S, Hajati F, Moni MA. Ensemble learning for disease prediction: A review. In: Healthcare. Vol 11. MDPI; 2023. p. 1808. https://www.mdpi.com/2227-9032/11/12/1808. Accessed November 5, 2024.
- Pauwels E, Stoven V, Yamanishi Y. Predicting drug side-effect profiles: a chemical fragment-based approach. BMC Bioinformatics. 2011;12(1):169.
- 32. Nakamura Y, Nakamura Y, Hori E, et al. Tumor lysis syndrome after transcatheter arterial infusion of cisplatin and embolization therapy for liver metastases of melanoma. Int J Dermatol. 2009;48(7):763–7.
- 33. Levine DP. Vancomycin: A history. Clin Infect Dis. 2006;42(Supplement_1):S5-12.
- Kaur H, Pannu HS, Malhi AK. A systematic review on imbalanced data challenges in machine learning: applications and solutions. ACM Comput Surv. 2019;52(4):1–36. https://doi.org/10.1145/3343440.
- 35. Zhuang L, Wang H, Hua M, Li W, Zhang H. Predicting drug-drug adverse reactions via multi-view graph contrastive representation model. Appl Intell. 2023;53(14):17411–28. https://doi.org/10.1007/s10489-022-04372-9

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.