Future Generation Computer Systems 125 (2021) 24–31

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

A computational drug repositioning model based on hybrid similarity side information powered graph neural network

Sumin Li *, Xiuqin Pan

School of Information Engineering, Minzu University of China, Beijing 100081, China

ARTICLE INFO

Article history:
Received 4 March 2021
Received in revised form 29 May 2021
Accepted 8 June 2021
Available online 18 June 2021

Keywords:
Computational drug repositioning
Drug–disease association prediction
Graph neural networks
Side information
Dimensionality reduction algorithm

ABSTRACT

Computational drug repositioning technology aims to rediscover the potential use of drugs already on the market and can significantly accelerate the traditional drug development process, reducing significant drug development costs and drug development instability

In this work, in order to capture valid and robust hidden feature representations of drugs and diseases, we introduce a new computational drug relocation model, HSSIGNN, based on hybrid similarity side information powered graph neural network, by drawing on the application of graph neural networks and Side information in recommender systems. Its advantage is to utilize the learning capability of graph neural networks to capture the effective hidden feature representation of drugs and diseases, which is used to infer the probability of whether a drug can treat the disease of interest, as a way to improve the generalization capability of the model. In addition, dimensionality reduction algorithms and side information of drugs and diseases are used to overcome the cold start problem encountered by traditional computational drug relocation models. Finally, the experimental results of the proposed model on two real drug–disease association datasets are analyzed to verify its superiority and effectiveness. Comprehensive experimentations on several real-world datasets show the efficiency of HSSIGNN.

1. Introduction

Medication exploration is really a huge expense as well as huge threat procedure [1,2]. Among one of the most essential action in medication project is composed in presuming possible signs for unique particles as well as in the repositioning of authorized medications [3,4]. Particularly medication repurposing has the advantage of beginning with well-characterized particles, thus decreasing the dangers in medical stages as well as the expense of tests [5,6].

Computational drug repositioning technology aims to rediscover the potential use of drugs already in the market, and it can significantly accelerate the traditional drug development process, reducing significant drug development costs and drug development instability [7,8]. Computational drug repositioning techniques have attracted the attention of a large number of researchers and companies due to their intrinsic and significant economic value [9].

In this work, in order to be able to obtain effective and robust hidden feature representations of drugs and diseases, we introduced a new computational drug repositioning model, HSSIGNN, based on hybrid similarity side information powered graph neural network, drawing on the application of graph neural networks [10] and side information [11] in recommender systems. Firstly, in order to obtain the effective hidden features of drugs and diseases, the HSSIGNN model draws on the graph neural network operator in High-Order GNN , which is used to compute the effective hidden feature values of drugs and diseases. Secondly, to be able to obtain a robust hidden feature representation, the HSSIGNN model uses the drug–disease side information to extract another hidden feature representation by feding the drug-to-disease similarity matrix and the disease-to-disease similarity matrix into the PCA algorithm. Then, in order to be able to consider the contribution of both hidden feature representations to the final predicted values, the HSSIGNN model performs a splicing operation of the two hidden features of the drug and the disease, and subsequently inputs the spliced hidden features of the drug and the disease into a three-layer autoencoder to extract the respective final hidden feature representations. Finally, the final hidden feature representation of the drug and the disease is element-wise multiplied and fed into a single layer fully connected network to obtain the final predicted value, the magnitude of which represents the probability of the drug being able to treat the disease.

The main contributions made by this work are as follows.

* Corresponding author.
E-mail address: smli@muc.edu.cn (S. Li).

https://doi.org/10.1016/j.future.2021.06.018
0167-739X/

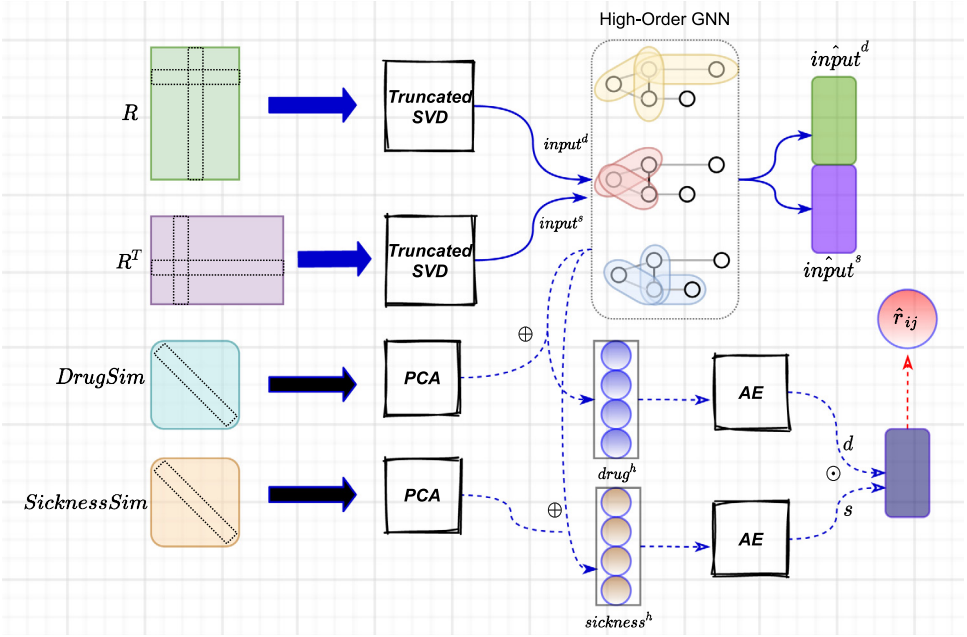


Fig. 1. The framework of our proposed model, HSSIGNN.

- We use the learning capability of graph neural networks to capture the valid hidden feature representations of drugs and diseases, which are used to infer the probability of whether a drug can treat the disease of interest, as a way to improve the generalization capability of the model.
- We use dimensionality reduction algorithms and drug or disease side information to overcome the cold-start problem experienced by traditional computational drug repositioning models.
- We verified the superiority and validity of the model proposed in this work by analyzing its experimental results on two real drug–disease association datasets.

The subsequent sections of this work are structured as follows. In Section 2, "Related Work ", we present the results of the current mainstream computational drug repositioning models. Then, in Section 3, "Method", we will analyze the implementation details of the HSSIGNN model. In Section 4, "Experiment and Discussion", we will discuss the experimental results of the proposed HSSIGNN model on several real-world drug–disease association datasets and compare the results with other classical classification models. Finally, in Chapter Section 5, "Conclusion", we will conclude our work.

2. Related work

Over the last few years, research workers have actually proposed a range of computational medication repurposing methods [11–19], which includes graph-based approaches, matrix factorization based techniques, Collective filtering system and so on.

Based upon the presumption that resembling medications are generally connected with resembling illness as well as the other way around, Luo et al. [12] introduced an unique computational approach called MBiRW, which makes use of several detailed resemblance procedures as well as Bi-Random stroll (BiRW) formula to determine prospective new indicators for a provided medication.

Luo et al. [13] introduced a professional recommendation method to deal with the issue of medication repurposing. An unique computational approach for medication repurposing, called DRRS (Drug Repositioning Recommendation System), is created to determine new illness indicators for provided medications. In DRRS, a multiple medication–illness network is built by incorporating medication–medication network, illness-condition graph as well as medication–condition relationship graph.

Wang et al. [14] introduced an unique approach, called DrPOCS, to determine prospect indicators of marked medications based upon forecast onto convex collections (POCS). By using the combination of medication framework as well as illness phenotype info, DrPOCS forecasts possible relationships among medications as well as illness with matrix factorization.

Yang et al. [11] introduced an unique approach for computational medication repurposing, called ANMF. The ANMF network utilizes drug–drug resemblances as well as disease–disease resemblances to boost the depiction info of medications as well as illness so as to conquer the issue of information sparsity.

Xuan et al. [15] introduced a system based on non-negative matrix factorization called DisDrugPred, the novelty of this method is making use of a brand-new type of drug similarity is calculated based upon their associated diseases.

Zhang et al. [16] put forward a computational approach "SCMFDD" to forecast unnoticed drug–disease associations. SCMFDD integrate drug feature-based comparabilities and also disease semantic similarity right into the matrix factorization structure. The uniqueness of this technique is that it integrates medicine attributes and also disease semantic info right into the matrix factorization framework.

However, the shortcomings of previous models [9] in learning ability prevent them from obtaining implicit feature representations that can effectively infer the final predicted values. Also the sparsity of the drug–disease dataset leads to the inability of previous correlation models to learn robust feature representations of drugs and diseases, making them more susceptible to cold-start problems. The above two problems are the main reasons for the poor generalization performance of the correlation model.

3. Method

In this section, we will analyze the implementation details of each part of the HSSIGNN model and the related formulas. Fig. 1

Table 1
Statistics for two real data sets.

Datasets	Drugs	Diseases	Proven associations
Fdataset	~593	~313	~1933
Cdataset	~663	~409	~2532

Algorithm 1 Our proposed Model.

Input-1: Drug-Disease association matrix, R
Input-2: Drug-Drug similarity matrix, $DrugSim$
Input-3: Sickness-Sickness similarity matrix, $SicknessSim$
Output: Probability of treatment of disease j by drug i, $\hat{r}_{ij}$

for $(i, j) \in$ Unknown drug-sickness associations **do**
  Mining hidden features of drugs and diseases based on Higher-order Graph Neural Networks and Truncated SVD.
  $input^d \leftarrow PCA(R)$
  $input^s \leftarrow PCA(R^T)$
  $[drug^g, sickness^g] \leftarrow HO - GNN([input^d, input^s], R)$
  $[\hat{input}^d, \hat{input}^s] \leftarrow f(W_g[drug^g, sickness^g] + b_g)$
  Mining hidden features of drugs and diseases based on side information.
  $drug^s \leftarrow PCA(DrugSim)$
  $sickness^s \leftarrow PCA(SicknessSim)$
  Output predicted value
  $drug^h = [drug^g, drug^s]$
  $sickness^h = [sickness^g, sickness^s]$
  $d \leftarrow f(W_1^T drug^h + b_1)$
  $\hat{drug}^h \leftarrow g(V_1^T d + b_d)$
  $s \leftarrow f(W_2^T sickness^h + b_2)$
  $\hat{sickness}^h \leftarrow g(V_2^T d + b_s)$
  $\hat{r}_{ij} \leftarrow f(W^T(d_i \odot s_j) + b)$
  Return $\hat{r}_{ij}$
end for

Fig. 2. The Pseudocode of our proposed model.

shows the algorithm flowchart of the HSSIGNN model, which contains four input matrices, namely, the drug–disease association matrix R and its transpose matrix R^T, the drug-drug similarity matrix *DrugSim*, and the disease-disease similarity matrix *SicknessSim*.

The HSSIGNN model first inputs R and R^T into the Truncated SVD model, and the output values are used as the initial feature values of the drugs and diseases in the GNN model. Subsequently, the graph neural network operator operation is used to obtain the effective hidden feature representations of drugs and diseases. Then the HSSIGNN model captures the second hidden feature representation of drugs and diseases using the dimensionality reduction algorithm PCA and the side information of drugs and diseases, *DrugSim* and *SicknessSim*. Finally, in the output prediction value module, in order to be able to consider the contribution of both hidden feature representations to the final prediction value, the HSSIGNN model performs a splicing operation on the two hidden features of drug and disease, and then inputs the spliced hidden features of drug and disease into the three-layer autoencoder respectively to extract the final hidden feature representation. Finally, the final hidden feature representations of the drug and the disease are element-wise multiplied and input into a single-layer fully connected network to obtain the final prediction value, which represents the probability of the drug being able to treat the disease.

How to obtain valid hidden feature representations of drugs and diseases using the graph neural network operator will be presented in Section 3.1. Subsequently, how to capture the second hidden feature representation of drugs and diseases will be described in Section 3.2. Finally in Section 3.3 the output predictive value module will be introduced.

3.1. Mining hidden features of drugs and diseases based on higher-order graph neural networks and Truncated SVD

Currently graph neural networks have achieved great success in many fields due to their unique learning ability. GNN models have strong learning ability, which has been verified in many papers on GNN. However, the shortcomings of previous drug repositioning models in learning ability prevent them from obtaining the hidden feature representation that can effectively infer the final predicted value. Therefore, the HSSIGNN model uses the graph neural network (GNN) operator in "Higher-Order Graph Neural Networks" to extract the hidden feature representations of drugs and diseases.

The drug–disease associations are too sparse, resulting in the presence of a large amount of worthless information. Therefore, in order to obtain the initialized drug–disease hidden feature representation, we use the Truncated SVD algorithm to downscale the drug–disease association matrix and its transpose matrix to obtain the initial hidden feature values. Truncated SVD is a variation of SVD that only calculates the maximum K singular values specified. Equations (1)–(2) are used to obtain the initialized drug–disease hidden feature representations.

$$input^d = TSVD(R) \tag{1}$$

$$input^s = TSVD(R^T) \tag{2}$$

where $input^d$ and $input^s$ are the initial hidden feature values of the drug and disease, respectively, and TSVD denotes the Truncated SVD algorithm.

The HSSIGNN model then borrows the graph neural network operator from "Higher-Order Graph Neural Networks" and uses it to extract the hidden features of drugs and diseases. The extraction operation is shown in Eqs. (3)–(4).

$$[drug^g, sickness^g] = HO - GNN([input^d, input^s], R) \tag{3}$$

$$[\hat{input}^d, \hat{input}^s] = f(W_g[drug^g, sickness^g] + b_g) \tag{4}$$

where $drug^g$ and $sickness^g$ are the hidden features after GNN computation. $hatinput^d$ and $hatinput^s$ are the original inputs after decoding, respectively. By minimizing the error between $[\hat{input}^d, \hat{input}^s]$ and $[input^d, input^s]$, the effective hidden features of drugs and diseases can be trained.

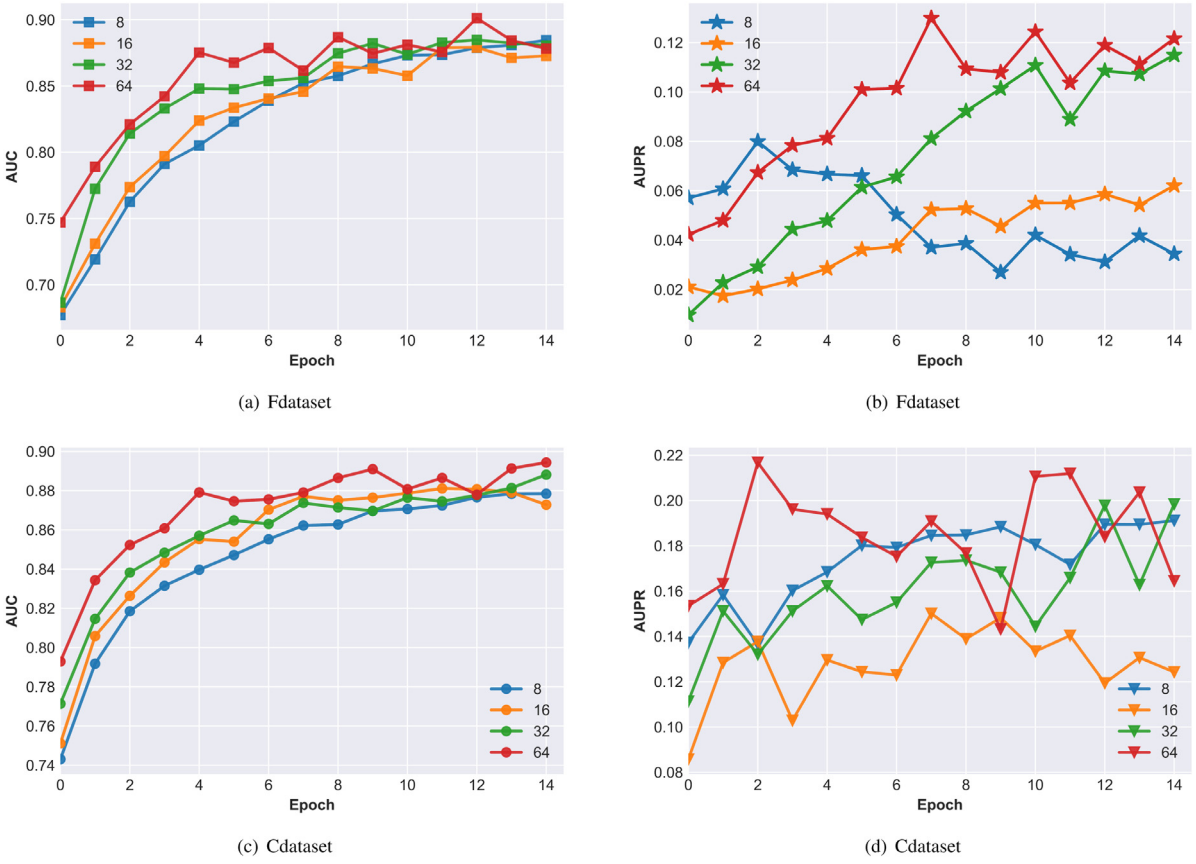
3.2. Mining hidden features of drugs and diseases based on side information

The sparsity of drug–disease datasets leads to the inability of previous correlation models to learn robust feature representations of drugs and diseases, making them more susceptible to the cold-start problem, which leads to the degradation of the generalization performance of the models. Meanwhile, side information is often used in the field of recommender systems to alleviate the cold-start problem. Therefore, HSSIGNN uses the similarity information between drugs and similarity information between diseases to obtain the respective second hidden features. The process of extracting the hidden feature representation is shown in Eqs. (5)–(6).

$$drug^s = PCA(DrugSim) \tag{5}$$

$$sickness^s = PCA(SicknessSim) \tag{6}$$

where $drug^s$ and $sickness^s$ are the second hidden features of drug and disease, respectively, PCA is the principal component analysis algorithm, which plays a role in extracting effective hidden features. DrugSim is the similarity matrix between drugs, and SicknessSim is the similarity matrix between diseases.



(a) Fdataset (b) Fdataset

(c) Cdataset (d) Cdataset

Fig. 3. The experimental results of our proposed model with different hidden feature dimensions.

3.3. Output predicted value

Through the above steps we obtain two types of hidden features for drugs and diseases respectively. Therefore, in order to be able to consider the contribution of both implicit feature representations to the final predicted values, the HSSIGNN model first splices the two implicit features of drugs and diseases as in Eqs. (7)–(8).

$$drug^h = [drug^g, drug^s] \tag{7}$$

$$sickness^h = [sickness^g, sickness^s] \tag{8}$$

The splicing vectors of drug and disease, $drug^h$ and $sickness^h$, respectively, are then fed as raw inputs to AE model to extract robust and efficient final hidden feature representations as in Eqs. (9)–(12).

$$d = f(W_1^T drug^h + b_1) \tag{9}$$

$$\hat{drug}^h = g(V_1^T d + b_d) \tag{10}$$

$$s = f(W_2^T sickness^h + b_2) \tag{11}$$

$$\hat{sickness}^h = g(V_2^T d + b_s) \tag{12}$$

where equations (9) and (11) are encoding operations and (10) and (12) are decoding operations. The d and s are the final hidden feature representations of the drug and the disease. Minimizing the error between the input and output thus allows to obtain d and s.

Finally, the final hidden feature representations of drugs and diseases, d and s, are subjected to the element pair multiplication operation [20] as in Eq. (13) and input to a single-layer fully connected network to obtain the final predicted values.

$$\hat{r}_{ij} = f(W^T(d_i \odot s_j) + b) \tag{13}$$

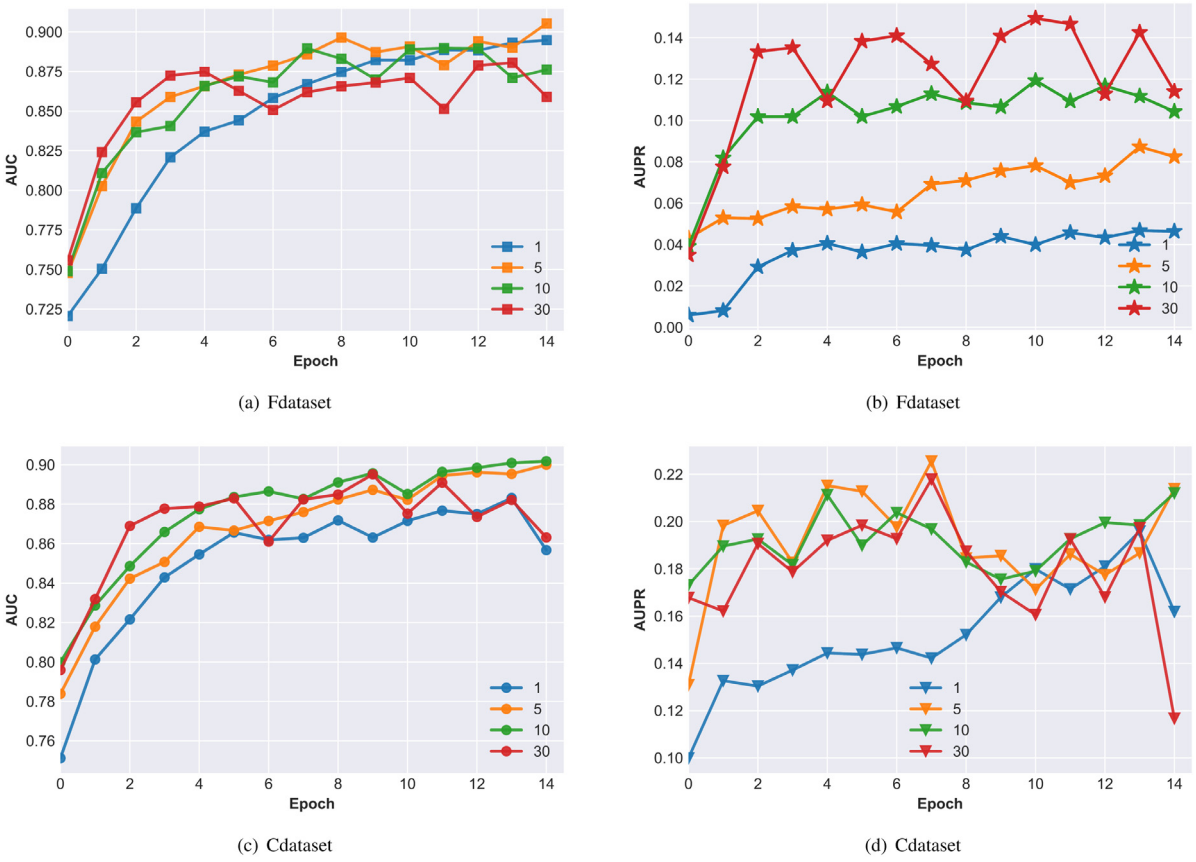
where $\hat{r}_{ij}$ is the final predictive value, the magnitude of which represents the probability that the drug i can treat the disease j. The pseudo-code of the HSSIGNN model is shown in Fig. 2.

4. Results and discussion

In this section, we will explore the experimental results of the HSSIGNN model proposed in this work on several real-world drug–disease association datasets, including the analysis of the results on some important parameters and the comparison results with other classical classification models. Firstly, the relevant datasets used for the experiments will be presented in Section 4.1. Secondly, in Section 4.2, the evaluation metrics used in the experiments are presented. Then in Section 4.3, the impact of some important parameters on the performance of the HSSIGNN model will be analyzed. Finally, Section 4.4 will compare the experimental results of HSSIGNN model with some mainstream classification models.

4.1. Dataset

This section both uses 2 real datasets, Fdataset and Cdataset, where Fdataset contains 593 drugs, 313 diseases and 1933 validated drug–disease associations, and Cdataset contains 663 drugs, 409 diseases and 2532 validated drug–disease associations. The types of drugs and diseases in the Cdataset are actually an expansion of the Fdataset, and it can also be said that the Fdataset is a subset of the Cdataset. The above datasets and the associated drug or disease similarity information are available in [21–25] and the dataset used for the experiments in this section can be downloaded at "https://github.com/bioinfomaticsCSU/MBiRW/tree/master/Datasets". Table 1 shows the relevant data statistics for the two datasets.



(a) Fdataset (b) Fdataset

(c) Cdataset (d) Cdataset

Fig. 4. The experimental results of our proposed model with different number of negative samples.

4.2. Evaluation criteria

We sliced the known drug–disease associations in a 9:1 ratio. Ninety percent of the associations are used as the training set to train the model. The remaining 10 percent of the associations are used as the test set to evaluate the generalization performance of the model. To be able to numerically evaluate the performance of the model, we use two mainstream evaluation metrics, AUC and AUPR, for evaluating the performance of the experimental results of the HSSIGNN model. The algorithm proposed in this paper is done in a CPU environment.

4.3. Related important parameters experiment

In this section, we analyze the performance of the HSSIGNN model with different hidden feature dimensions, number of negative samples and learning rate. The variation interval of the hidden feature dimension is [8, 16, 32, 64], the variation interval of the number of negative samples is [1, 5, 10, 30], and the variation interval of the learning rate is [0.0001, 0.001, 0.005, 0.01]. We are using the grid method to find the best combination of parameters.

4.3.1. Hidden feature dimensions

The size of the hidden feature dimension determines the ability of the HSSIGNN model to capture high-level drug–disease associations. The appropriate dimensionality can enhance the generalization ability of HSSIGNN model to a great extent. Fig. 3 shows the experimental results of the HSSIGNN model with different hidden feature dimensions for the experimental datasets Fdataset and Cdataset. the evaluation metrics are AUC and AUPR.

The horizontal coordinates of Fig. 3 are the number of model training iterations, and the vertical coordinates are the corresponding AUC or AUPR values. On the Fdataset, the magnitudes of AUC and AUPR values are basically proportional to the magnitudes of the hidden feature values. Meanwhile, compared with the AUC metric, the value of AUPR value changes more jitterily, while the AUC metric changes more smoothly. the experimental results of Cdataset are basically similar to those of Fdataset.

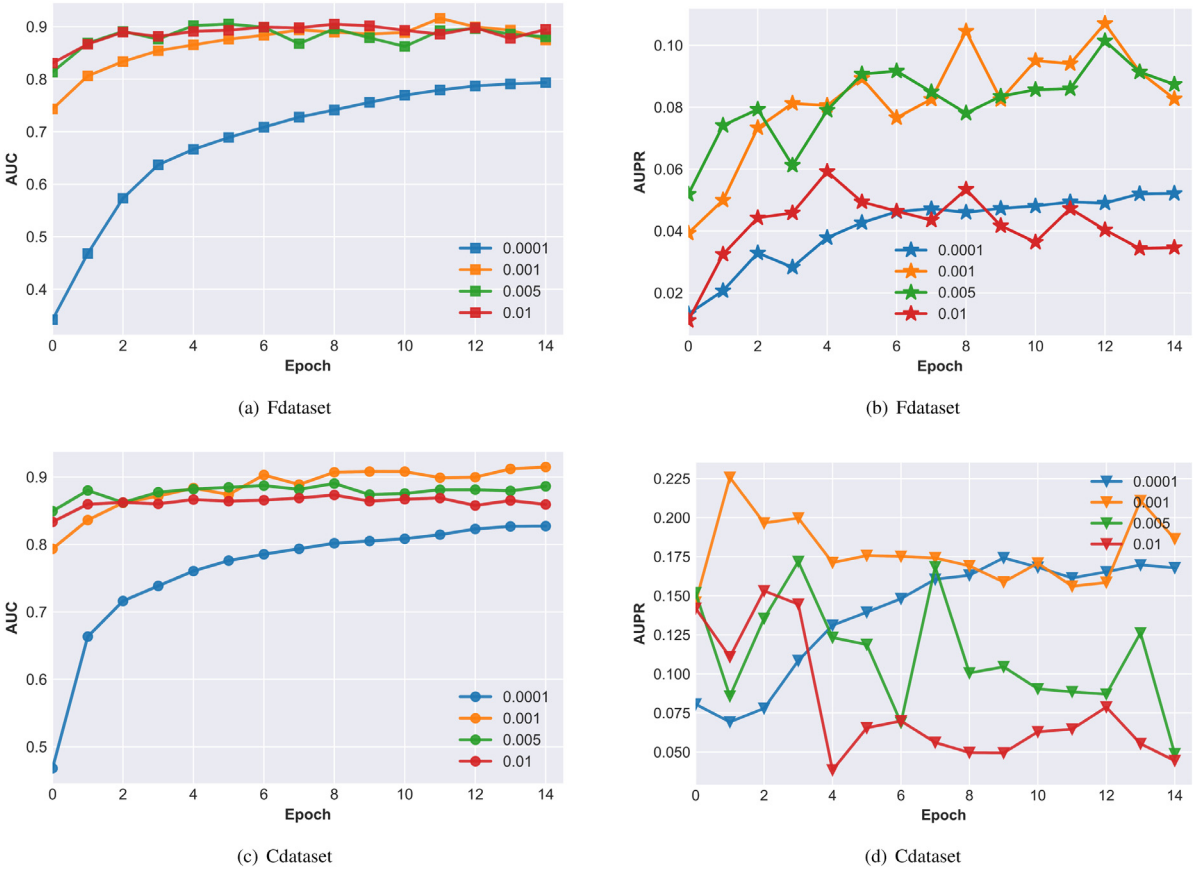
4.3.2. Number of negative samples

The size of the number of negative samples can adjust the ratio of positive and negative samples in the training set. The appropriate ratio of positive and negative samples can improve the performance of HSSIGNN model. Fig. 4 shows the experimental results of HSSIGNN model with different number of negative samples, the experimental datasets are Fdataset and Cdataset. the evaluation metrics are AUC and AUPR.

The horizontal coordinates of Fig. 4 are the number of model training iterations and the vertical coordinates are the corresponding AUC or AUPR values. On the Fdataset, a larger number of negative samples instead achieves a smaller AUC value, and the model achieves the largest AUC value when the number of negative samples is 5. However, with the AUPR metric, a larger number of negative samples achieves the best AUPR value. Also unlike Fdataset, on Cdataset, the model achieved the largest AUC and AUPR values when the number of negative samples was 10 and 5, respectively.

4.3.3. The size of the learning rate

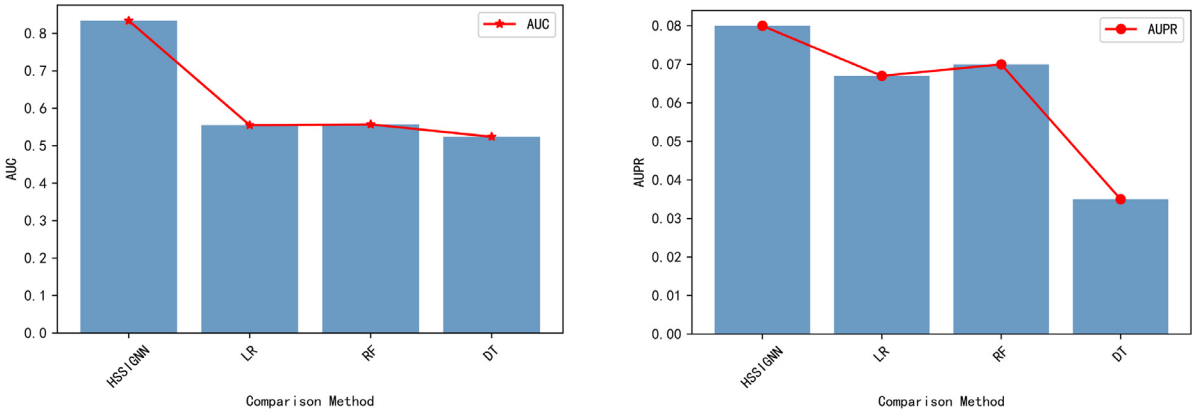
The size of the learning rate determines the learning speed of HSSIGNN model parameters. A suitable learning rate enables the model to find the right combination of parameters and enhance



(a) Fdataset (b) Fdataset

(c) Cdataset (d) Cdataset

Fig. 5. The experimental results of our proposed model with different learning rate.



(a) Fdataset (b) Fdataset

Fig. 6. The experimental results of our proposed model and three classical classification algorithms on Fdataset.

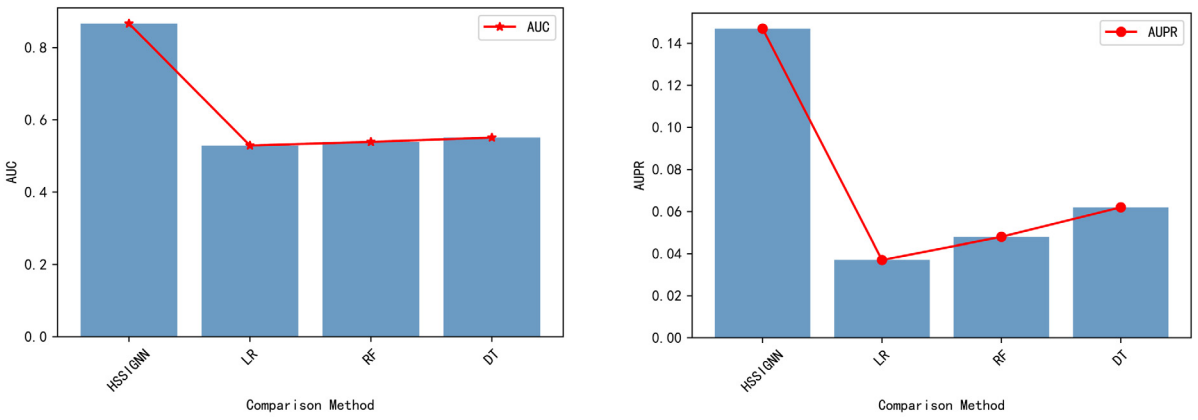
the experimental effect of the model. Fig. 5 shows the experimental results of the HSSIGNN model with different learning rates for the experimental datasets Fdataset and Cdataset. the evaluation metrics are AUC and AUPR.

The horizontal coordinate of Fig. 5 is the number of model training iterations, and the vertical coordinate is the corresponding AUC or AUPR value. On Fdataset, too large or too small learning rate cannot achieve good AUC and AUPR values, and when the learning rate is 0.001, the model achieves the maximum AUC value and the best AUPR value. the experimental results of Cdataset are basically similar to Fdataset. Too large a learning rate leads to a non-converging model, while too small a learning rate leads to a particularly slow convergence or failure to learn. Therefore, too large or too small learning rates cannot achieve good AUC and AUPR values.

4.4. Benchmark comparison

In order to be able to objectively evaluate the performance of the HSSIGNN model, we will compare its experimental results with the following three classical machine learning based classification algorithms.

(1) Logistic Regression (LR): LR is an artificial intelligence approach utilized to address a binary issue to predict the possibility of one particular thing.



(a) Cdataset (b) Cdataset

Fig. 7. The experimental results of our proposed model and three classical classification algorithms on Cdataset.

(2) Random Forest (RF): The RF method is a classifier which contains numerous decision trees as well as whose outcome classes are identified by the majority of the classes outcome by the each trees.
(3) Decision Tree (DT): The DT approach is a fundamental classification method. Its high interpretability makes it one of the mainstream algorithms today.

Figs. 6–7 show the experimental results of the HSSIGNN model, LR model, RF model and DT model on Fdataset and Cdataset, respectively, with the evaluation metrics of AUC and AUPR. observing Fig. 6, the HSSIGNN model achieves the best results on Fdataset with both metrics, substantially ahead of the other three comparison algorithms. Similarly, the experimental results in Fig. 7 show that the HSSIGNN model achieves the best AUC and AUPR values on Cdataset, which is also better than the other 3 comparison algorithms. Because these comparison benchmarks are machine learning methods, the method proposed in this manuscript is based on graph neural networks, which are deep learning methods. The graph neural network has better learning ability and generalization ability, so the experimental results of the comparison benchmarks are not as good as the results of the method proposed in this manuscript. Based on the information derived from Figs. 6–7, it can be inferred that the HSSIGNN model has certain excellence and effectiveness.

5. Conclusion

Computational drug repositioning technology aims to rediscover the potential use of drugs already on the market and can significantly accelerate the traditional drug development process, reducing significant drug development costs and drug development instability

In this work, in order to capture valid and robust hidden feature representations of drugs and diseases, we introduce a new computational drug relocation model, HSSIGNN, based on hybrid similarity side information powered graph neural network, by drawing on the application of graph neural networks and Side information in recommender systems. its advantage is to utilize the learning capability of graph neural networks to capture the effective hidden feature representation of drugs and diseases, which is used to infer the probability of whether a drug can treat the disease of interest, as a way to improve the generalization capability of the model. In addition, dimensionality reduction algorithms and side information of drugs and diseases are used to overcome the cold start problem encountered by traditional computational drug relocation models. Finally, the experimental results of the proposed model on two real drug–disease association datasets are analyzed to verify its superiority and effectiveness.

CRediT authorship contribution statement

Sumin Li: Completed the algorithm framework, Writing - original draft. **Xiuqin Pan:** Proposed the idea, Checked the language and writing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] J.S. Shim, J.O. Liu, Recent advances in drug repositioning for the discovery of new anticancer drugs, Int. J. Biol. Sci. 10 (7) (2014) 654.
[2] M. Dickson, J.P. Gagnon, Key factors in the rising cost of new drug discovery and development, Nat. Rev. Drug Discov. 3 (5) (2004) 417–429.
[3] N.A. Tamimi, P. Ellis, Drug development: from concept to marketing!, Nephron Clin. Pract. 113 (3) (2009) c125–c131.
[4] S. Pushpakom, F. Iorio, P.A. Eyers, K.J. Escott, S. Hopper, A. Wells, A. Doig, T. Guilliams, J. Latimer, C. McNamee, et al., Drug repurposing: progress, challenges and recommendations, Nat. Rev. Drug Discov. 18 (1) (2019) 41–58.
[5] T.T. Ashburn, K.B. Thor, Drug repositioning: identifying and developing new uses for existing drugs, Nat. Rev. Drug Discov. 3 (8) (2004) 673–683.
[6] N. Nosengo, Can you teach old drugs new tricks?, Nature 534 (7607) (2016) 314–316.
[7] J.-L.E. Pritchard, T.A. O'Mara, D.M. Glubb, Enhancing the promise of drug repositioning through genetics, Front. Pharmacol. 8 (2017) 896.
[8] J.K. Yella, S. Yaddanapudi, Y. Wang, A.G. Jegga, Changing trends in computational drug repositioning, Pharmaceuticals 11 (2) (2018) 57.
[9] H. Luo, M. Li, M. Yang, F.-X. Wu, Y. Li, J. Wang, Biomedical data and computational models for drug repositioning: A comprehensive review, Brief. Bioinform. (2019).
[10] C. Morris, M. Ritzert, M. Fey, W.L. Hamilton, J.E. Lenssen, G. Rattan, M. Grohe, Weisfeiler and leman go neural: Higher-order graph neural networks, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 33, 01, 2019, pp. 4602–4609.
[11] X. Yang, Y. Liu, J. He, et al., Additional neural matrix factorization model for computational drug repositioning, BMC Bioinformatics 20 (1) (2019) 1–11.
[12] H. Luo, J. Wang, M. Li, J. Luo, X. Peng, F.-X. Wu, Y. Pan, Drug repositioning based on comprehensive similarity measures and Bi-random walk algorithm, Bioinformatics 32 (17) (2016) 2664–2671.
[13] H. Luo, M. Li, S. Wang, Q. Liu, Y. Li, J. Wang, Computational drug repositioning using low-rank matrix approximation and randomized algorithms, Bioinformatics 34 (11) (2018) 1904–1912.

[14] Y.-Y. Wang, C. Cui, L. Qi, H. Yan, X.-M. Zhao, DrPOCS: drug repositioning based on projection onto convex sets, IEEE/ACM Trans. Comput. Biol. Bioinform. 16 (1) (2018) 154–162.
[15] P. Xuan, Y. Cao, T. Zhang, X. Wang, S. Pan, T. Shen, Drug repositioning through integration of prior knowledge and projections of drugs and diseases, Bioinformatics 35 (20) (2019) 4108–4119.
[16] W. Zhang, X. Yue, W. Lin, W. Wu, R. Liu, F. Huang, F. Liu, Predicting drug-disease associations by using similarity constrained matrix factorization, BMC Bioinformatics 19 (1) (2018) 1–12.
[17] M. Gönen, S. Khan, S. Kaski, Kernelized Bayesian matrix factorization, in: International Conference on Machine Learning, 2013, pp. 864–872.
[18] W. Wang, S. Yang, J. Li, Drug target predictions based on heterogeneous graph inference, in: Biocomputing 2013, World Scientific, 2013, pp. 53–64.
[19] V. Martinez, C. Navarro, C. Cano, W. Fajardo, A. Blanco, Drugnet: network-based drug–disease prioritization by integrating heterogeneous data, Artif. Intell. Med. 63 (1) (2015) 41–49.
[20] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, T.-S. Chua, Neural collaborative filtering, in: Proceedings of the 26th International Conference on World Wide Web, 2017, pp. 173–182.
[21] D. Weininger, SMILES, A chemical language and information system. 1. Introduction to methodology and encoding rules, J. Chem. Inf. Comput. Sci. 28 (1) (1988) 31–36.
[22] C. Steinbeck, Y. Han, S. Kuhn, O. Horlacher, E. Luttmann, E. Willighagen, The chemistry development kit (CDK): An open-source java library for chemo-and bioinformatics, J. Chem. Inf. Comput. Sci. 43 (2) (2003) 493–500.
[23] C. Knox, V. Law, T. Jewison, P. Liu, S. Ly, A. Frolkis, A. Pon, K. Banco, C. Mak, V. Neveu, et al., Drugbank 3.0: A comprehensive resource for 'omics' research on drugs, Nucleic Acids Res. 39 (suppl_1) (2010) D1035–D1041.
[24] A. Hamosh, A.F. Scott, J.S. Amberger, C.A. Bocchini, V.A. McKusick, Online mendelian inheritance in man (OMIM), A knowledgebase of human genes and genetic disorders, Nucleic Acids Res. 33 (suppl_1) (2005) D514–D517.
[25] M.A. Van Driel, J. Bruggeman, G. Vriend, H.G. Brunner, J.A. Leunissen, A text-mining analysis of the human phenome, Eur. J. Hum. Genet. 14 (5) (2006) 535–542.

Sumin Li received the M.S. degree in 2003. She is currently a Lecturer at the School of Information Engineering, Minzu University of China, Beijing, China. His current research interests include big data, intelligent computing and intelligent information processing and system.

Xiuqin Pan received the B.E. degree in electric technology and the M.E. degree in power system and automation specialty from Zhengzhou University, Zhengzhou, China, in 1994 and 1999, respectively, and the Ph.D. degree in control theory and control engineering from the Beijing Institute of Technology, in 2002. She is currently a Professor with the School of Information Engineering, Minzu University of China. Her current research interests focus on parallel algorithm and intelligent systems.