Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Applied Soft Computing 111 (2021) 107831

Contents lists available at ScienceDirect

Applied Soft Computing

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

SANE: A sequence combined attentive network embedding model for COVID-19 drug repositioning

Xiaorui Su [a,b,c], Zhuhong You [d,*], Lei Wang [e,**], Lun Hu [a,b,c,***], Leon Wong [e], Boya Ji [f], Bowei Zhao [a,b,c]

[a] Xinjiang Technical Institutes of Physics and Chemistry, Chinese Academy of Sciences, Urumqi 830011, China
[b] University of Chinese Academy of Sciences, Beijing 100049, China
[c] Xinjiang Laboratory of Minority Speech and Language Information Processing, Urumqi 830011, China
[d] School of Computer Science, Northwestern Polytechnical University, Xi'an 710129, China
[e] Big Data and Intelligent Computing Research Center, Guangxi Academy of Science, Nanning, 530007, China
[f] College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, China

ARTICLE INFO

Article history:
Received 3 June 2021
Received in revised form 25 July 2021
Accepted 14 August 2021
Available online 23 August 2021

Dataset link: https://github.com/Blair1213/SANE

Keywords:
Drug–virus interactions
Depth-first-search
Drug repositioning
Attention
Network embedding
COVID-19

ABSTRACT

The COVID-19 has now spread all over the world and causes a huge burden for public health and world economy. Drug repositioning has become a promising treatment strategy in COVID-19 crisis because it can shorten drug development process, reduce pharmaceutical costs and reposition approval drugs. Existing computational methods only focus on single information, such as drug and virus similarity or drug–virus network feature, which is not sufficient to predict potential drugs. In this paper, a sequence combined attentive network embedding model SANE is proposed for identifying drugs based on sequence features and network features. On the one hand, drug SMILES and virus sequence features are extracted by encoder–decoder in SANE as node initial embedding in drug–virus network. On the other hand, SANE obtains fields for each node by attention-based Depth-First-Search (DFS) to reduce noises and improve efficiency in representation learning and adopts a bottom-up aggregation strategy to learn node network representation from selected fields. Finally, a forward neural network is used for classifying. Experiment results show that SANE has achieved the performance with 81.98% accuracy and 0.8961 AUC value and outperformed state-of-the-art baselines. Further case study on COVID-19 indicates that SANE has a strong predictive ability since 25 of the top 40 (62.5%) drugs are verified by valuable dataset and literatures. Therefore, SANE is powerful to reposition drugs for COVID-19 and provides a new perspective for drug repositioning.

1. Introduction

Coronaviruses are systematically classified into the genus Coronavirus of the family Coronaviridae of the order Nidovirales [1]. They are a group of enveloped, positive-strand RNA viruses [2] with non-segmented genomes of about 30,000 nucleotides and diameter of about 80-120nm [3]. Coronaviruses are widespread in nature [4] and prone to mutate, but it only infects vertebrates. Coronavirus has been widely concerned since 2000 owing to it has caused three serious outbreaks in the world to date, including the Severe Acute Respiratory Syndrome (SARS-CoV) in 2003 [5] and Middle East Respiratory Syndrome (MERS-CoV) in 2012 [6] and the epidemic we are experiencing COVID-19 (SARS-CoV-2) in 2019 [7]. To date, the COVID-19 outbreak has now spread to 6 continents including more than 60 countries [8], causing more than 112 millions of people infected, 2.5 millions of deaths [9,10] and the economic losses of over 300 billion dollars [11,12]. As a result, developing or discovering effective drugs is so important for preventing public health from being threaten by COVID-19 and reigniting global economy. However, traditional drug development experiments are often time-consuming, costly and risky, which decides that it is unrealistic to develop new effective drugs to treat COVID-19 in a short time. Therefore, it is urgent to adopt a new method to accelerate the drug discovery process and find effective drugs for COVID-19 treatment.

Drug repositioning [13] as an effective method to find new uses of existing drugs has received much attention and a large

* Corresponding author.
** Correspondence to: Guangxi Academy of Science, Big Data and Intelligent Computing Research Center, No. 98 Daling Road, Nanning, Guangxi, 530007, China.
*** Corresponding author at: Xinjiang Technical Institutes of Physics and Chemistry, Chinese Academy of Sciences, Urumqi 830011, China.
E-mail addresses: suxiaorui19@mails.ucas.ac.cn (X. Su), zhuhongyou@nwpu.edu.cn (Z. You), leiwang@gxas.cn (L. Wang), hulun@ms.xjb.ac.cn (L. Hu), lghuang@gxas.cn (L. Wong), byj@hnu.edu.cn (B. Ji), zhaobowei19@mails.ucas.ac.cn (B. Zhao).

https://doi.org/10.1016/j.asoc.2021.107831
1568-4946/

number of studies have been conducted and grow exponentially [14] in recent years. According to the records [15], there are in total 94 cases in which a repositioned drug made it to the market, such as Aspirin [16], Thalidomide [17] and Sildenafil [18]. More importantly, in addition to the advantages in time and cost, drug repositioning still has a number of advantages in drug adverse effects avoiding and economy [19]. Therefore, drug repositioning is a promising strategy for accelerating drug identifying of COVID-19 and minimizing the translational gap in drug development.

At the beginning of the development of drug repositioning, biological methods such as molecular docking, account for the majority. Molecular docking can directly determine drug targets, but it usually requires experiments on all drugs, which is lack of efficiency. In recent years, benefiting from the rapid development of machine learning and artificial intelligence, computational methods are gradually proposed and applied widely in the field of bioinformatics. According to the methods used in computational models, these methods can be grouped into two classifies: similarity-based and network-based. Similarity-based methods focus on drug similarity and protein similarity, and recommend drugs according to similarities. Though similarity-based methods utilized the basic information of drug and other molecules comprehensively, they only get the information or similarities in 2D space and have the limitation in high dimension network or graph structure data.

With the considerable advancements in social network and knowledge graph, network representation learning becomes a significant research tool in many fields. In drug repositioning task, a host of network-based models have been proposed in the last several years. The network-based methods, including Random Walk, graph neural network and knowledge graph embedding model, are employed to learn the network topology feature and repurpose potential drugs for diseases. Compared with similarity-based methods, network-based methods have the stronger learning ability and can capture complex information in high dimension space. However, network-based methods still have limitations in computational efficiency and lacking of node basic information. To address the above limitations, in this paper, we proposed a sequence combined attentive network embedding model SANE to repurpose drugs for COVID-19 by integrating drug SMILES and virus sequence information into an attention-based pre-search network embedding. Firstly, we collected drug SMILES and virus sequence information as their basic information. Simultaneously, we also collected a valuable drug–virus interaction dataset HDVD. Secondly, encoder–decoder is adopted to extract sequence feature. Then, in order to improve the efficiency of SANE, an attention-based Depth-First-Search (DFS) is applied to decrease network scale and reduce noises. After that, an attentive network embedding is used to learn representation for each node by aggregating basic information. By this way, the final representation contains both basic information and network topology information. In summary, our main contributions are as follows:

(1) we address the drug repositioning task from network perspective and design an efficient attention-based DFS network embedding to identify potential drugs against COVID-19;

(2) we integrate the drug and virus basic sequence information into network embedding to enhance information granularity, which also contributes to accuracy improvement;

(3) we use attention-based DFS to reduce redundancies and noises before representation learning;

(4) we test proposed model SANE and the results show that SANE achieves the new state-of-the-art results with significant improvements over baselines.

The article is organized as follows. In the next section, we introduce the related works and the motivation of our work. In Section 3, we introduce the datasets used in our work and elaborate the three sub-models contained in SANE. The experiment settings and detailed experimental results are provided in Section 4. We discuss the results obtained by SANE in Section 5 and Section 6 concludes this article.

2. Related work

Our research work is inspired by two lines of research: the inherent limitations of traditional similarity-based methods and the remarkable success of network representation learning algorithm in the interaction prediction of molecular network.

2.1. Similarity-based methods for drug repositioning

According to the resources used in similarity-based models, they can be divided into two groups: single-source similarity and multi-source similarity. The models using single-source similarity are simple and understandable. For example, the research work conducted by Li et al. [20] calculated the drug pair similarity based on drug structural information and repurposed drugs according to drug similarity. Afterwards, with the development of bioinformatics, more interactions data related to drugs are available, such as drug–target interactions, drug–disease interactions, protein sequence and disease structural information. As a result, the similarity in that stage was calculated by multi-source. Zhang et al. [21] integrated drug chemical structures, drug target proteins and drug–disease associations to extract similarity matrices of drugs and diseases, respectively and predicted potential drugs based on collaborative filtering. Moreover, Azad et al. [22] addressed drug repositioning by compiling heterogenous information for an exhaustive set of small-molecule drugs and integrated multiple sources to calculate drug similarity. Currently, similarity-based methods are also used to repurpose drugs for COVID-19 and there are a large number of works [23, 24] have been done on COVID-19. Meng et al. [25] proposed a similarity constrained matrix factorization model to identify new drug–virus interactions by calculating drug SMILES (Simplified Molecular Input Line Entry System) similarity and virus sequence similarity. Though similarity-based methods have been widely used and achieve good performances, they are still limited in 2D space as the similarity is calculated between two molecules, which lead to the lacking of global perspective.

2.2. Network-based methods for drug repositioning

Different from similarity-based methods, network-based methods address drug repositioning from a global perspective. This kind of methods concentrate on network topology feature learning based on network representation learning algorithms. For example, Luo et al. [26] proposed an efficient approach to capture global information of drug based on Random Walk for drug repositioning to prioritize candidate drugs for disease and the proposed approach outperformed similarity-based methods. Similarly, multi-sources are also used in network-based methods. In 2017, Luo et al. [27] integrated drugs, proteins, diseases and side-effects information to construct a heterogeneous network for drug–target interactions and repurposing existing drugs and validated experiment proved that the repositioned drug they found was able to prevent inflammatory disease. Network-based methods are also used to repurpose drugs for COVID-19, Zeng et al. [28] integrated multi-source to build a comprehensive knowledge graph to discover drugs for COVID-19 by using knowledge embedding method RotatE and achieved a promising result with 41 repositioned drugs identified. However, network-based methods often focus on the network structure, but ignore the

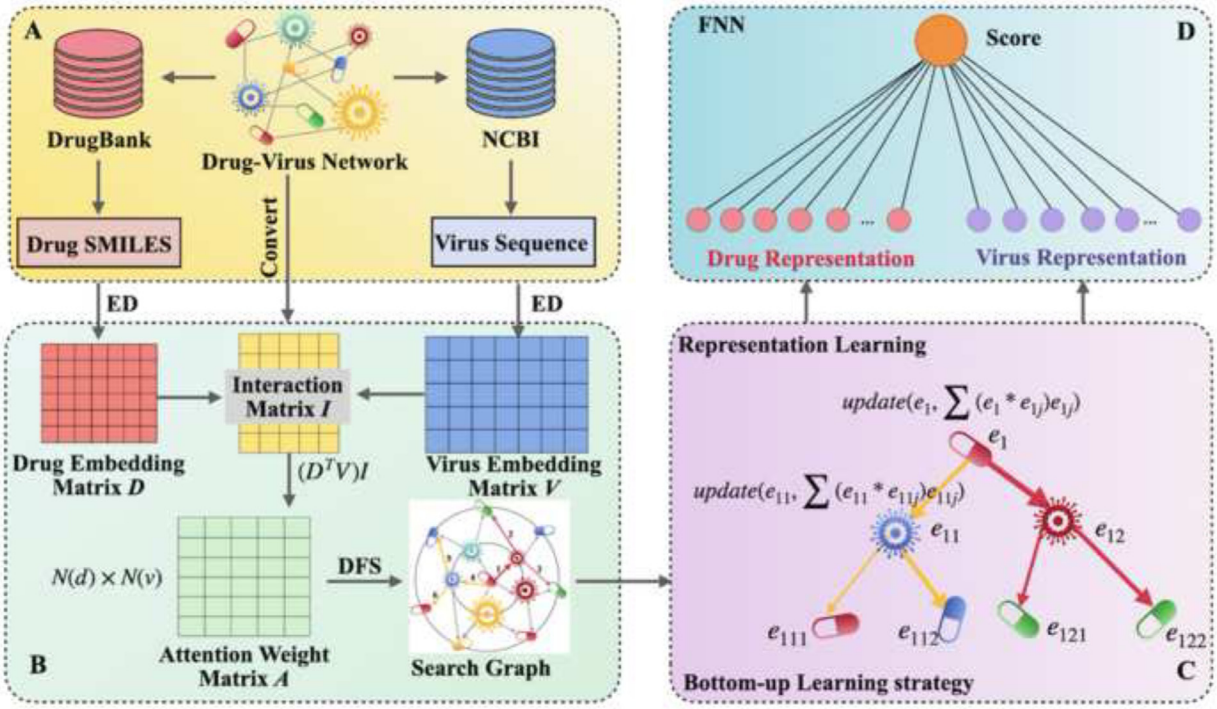

Fig. 1. The overview of SANE. (A) shows the data collecting, (B) shows feature extracting and attention-based DFS, (C) denotes representation learning and (D) is forward neural network for classifying.

node attribute, such as drug SMILES and virus sequence. Additionally, network-based methods may suffer from high computational complexity and low efficiency, which limits the application in large scale network.

3. Materials and methods

Proposed model involves five steps: (i) data collecting, (ii) extract sequence feature by encoder–decoder, (iii) reduce network noises and redundancies by attention-based DFS, (iv) extract network global feature by network representation learning, and (v) use the forward neural network to help suggest potential therapeutic drug scores for COVID-19. The whole model can be view in Fig. 1. Before introducing proposed model, some notations used should be illustrated. Lowercase bold letters (e.g. $\boldsymbol{v} \in \mathbb{R}^d$) denotes vectors and uppercase boldface letters (e.g. $\boldsymbol{M} \in \mathbb{R}^{m \times n}$) denotes matrices.

3.1. Data collecting

In the study, we collected the recently constructed human drug–virus interactions network (HDVD) [25] as the training dataset to measure the model performance. HDVD assembled a significant number of experimentally validated drug–virus interaction entries from literature by text mining technology. The statistics of HDVD is shown in Table 1 and all the interactions in HDVD are supported by experiments.

In addition, we also collected basic information of drugs and viruses contained in HDVD. Drug SMILES is one of the most popular molecular structure 1D representations. Thus, we downloaded the drug SMILES from DrugBank (V5.1.7) [29]. Generally, virus is represented by its RNA or DNA sequence. As a result, the genome nucleotide sequences of viruses are downloaded from the National Center for Biotechnology Information (NCBI).

3.2. Sequence Feature Extraction by Encoder–Decoder

In data collecting stage, we have collected drug SMILES and virus genome nucleotide sequences. Both drug SMILES and virus genome nucleotide sequences are the most representative molecular representations and have been widely used to extract features respectively. Different from previous work, we adopted encoder–decoder to extract sequence information. In fact, encoder–decoder can be viewed as a variant of Recurrent Neural Network (RNN). Long Short-Term Memory (LSTM) as one of the widely used basic units in RNN is also adopted in this work. We used Bi-LSTM in encoder layer and LSTM in decoder layer. The advantages of encoder–decoder structure in this work is that it can accept unequal length input and process the long sequence reasonably.

We denote the input as *seq*, which represents the drug SMILES or virus sequence. Firstly, an embedding layer is needed to transfer the unequal sequence to a machine understandable vector $\boldsymbol{s} \in \mathbb{R}^d$. Secondly, the embedding vector is sent into encoder layer to obtain encoder output, hidden state and cell state. Thirdly, encoder output is sent into decoder layer as its input and initialize decoder layer using hidden state and cell state. Then, the final output is obtained after decoder layer processing. Specifically, both encoder layer and decoder layer are stacked by LSTM as shown in Fig. 2 and single LSTM is shown in Fig. 3. Therefore, the encoder–decoder process can be formulated by:

$$\boldsymbol{s} = embedding\ (seq) \tag{1}$$

$$\boldsymbol{encoder}_{\boldsymbol{output}}, \boldsymbol{hiddenstate}, \boldsymbol{cellstate} = Bi - LSTM\ (\boldsymbol{s}) \tag{2}$$

$$\boldsymbol{decoder}_{\boldsymbol{output}} = softmax\left(LSTM\left(\boldsymbol{encoder}_{\boldsymbol{output}}, \boldsymbol{initial}_{\boldsymbol{state}}\right)\right) \tag{3}$$

Each LSTM unit mainly completes the following task: receive current data, transmit previous information, send information to next unit and update current cell state. Unlike the traditional RNN, LSTM controls the inflow and outflow of information by setting up three gates, which are input gate, forget gate and output date. In particular, input gate is responsible for processing the input of the current sequence position. Suppose the current

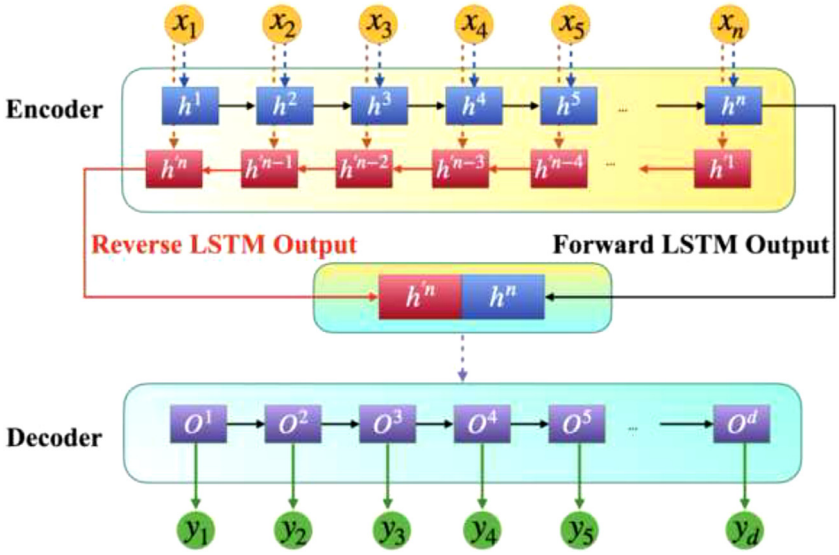

Fig. 2. The structure of encoder–decoder. The yellow circle nodes represent input at different time. In encoder layer, the blue block represents forward LSTM unit and the red block represents reverse LSTM unit. The word in them denotes the output of a single unit. The middle layer in diagram is encoder output and then it is sent into decoder layer. The purple block represents LSTM unit in decoder layer and green nodes represent the output of decoder.

Table 1
The statistics of HDVD dataset.

Dataset	#Drugs	#Virus	#Interactions	#Data Type
HDVD	219	34	455	Confirmed human drug–virus interactions

Table 2
Results of five-fold cross-validation achieved by SANE on HDVD dataset.

Method	Fold	ACC. (%)	F1 Score (%)	AUC	AUPR
Five-fold CV	1	84.62	85.42	0.9322	0.9175
	2	75.82	78.00	0.8493	0.8428
	3	87.91	88.89	0.9327	0.9354
	4	72.53	75.25	0.8681	0.8701
	5	89.01	89.58	0.8984	0.7879
	Average	**81.98±6.62**	**83.43 ± 5.79**	**0.8961 ± 0.0335**	**0.8707 ± 0.0529**

time is t and we denote the last time output as $\boldsymbol{h}^{t-1}$ and current time input as $\boldsymbol{x}^{t}$. The input gate as the pink lines shown in Fig. 3. can be formulated by:

$$i^t = \sigma\left(\boldsymbol{W}_i\boldsymbol{h}^{t-1} + \boldsymbol{U}_i\boldsymbol{x}^t + \boldsymbol{b}_i\right) \tag{4}$$

$$\alpha^t = \tanh\left(\boldsymbol{W}_a\boldsymbol{h}^{t-1} + \boldsymbol{U}_a\boldsymbol{x}^t + \boldsymbol{b}_a\right) \tag{5}$$

Here, i^t can be viewed as a probability that how much information will be passed in and α^t represents the information that the current unit receiving. The second gate, forget gate as the cyan lines shown in Fig. 3. is responsible for processing the hidden state from last time. The forget gate can be formulated by following and f^t represents that how much previous hidden information can be transmitted into current unit.

$$f^t = \sigma\left(\boldsymbol{W}_f\boldsymbol{h}^{t-1} + \boldsymbol{U}_f\boldsymbol{x}^t + \boldsymbol{b}_f\right) \tag{6}$$

The third gate, output gate, controls the output of current unit, which can be denoted as $\boldsymbol{h}^t$ and as the green lines shown in Fig. 3. During this process, there is another parameter is added to the calculation, called cell state $\boldsymbol{C}$. Cell state is updated by input gate and forget gate. Thus, the output of current unit can be represented by:

$$C^t = C^{t-1} \odot f^t + i^t \odot \alpha^t \tag{7}$$

$$o^t = \sigma\left(\boldsymbol{W}_o\boldsymbol{h}^{t-1} + \boldsymbol{U}_o\boldsymbol{x}^t + \boldsymbol{b}_o\right) \tag{8}$$

$$\boldsymbol{h}^t = o^t \tanh \odot \left(C^t\right) \tag{9}$$

In our work, encoder–decoder aims to learn sequence feature. As a result, the encoder output is the concatenation of forward LSTM representation and reverse LSTM representation. We denote the final output of forward LSTM as $\boldsymbol{h}^n$ and the output of reverse LSTM as $\boldsymbol{h}'^n$. Then, the encoder output, hidden state and cell state can be rewritten by:

$$\boldsymbol{encoder}_{\boldsymbol{output}} = \boldsymbol{hiddenstate} = \left[\boldsymbol{h}^n; \boldsymbol{h}'^n\right] \tag{10}$$

$$\boldsymbol{cellstate} = \left[\boldsymbol{C}^n; \boldsymbol{C}'^n\right] \tag{11}$$

3.3. Depth-first-search based on attention

Previous network representation learning models can be grouped into three groups (MF-based, DW-based and NN-based) according to learning strategy. Unlike these existing models, a preprocess is used in proposed model for accurate learning, called attention-based DFS, which is shown in Fig. 4. It has the advantage of reducing noises and redundancies and promoting accuracy of representation learning.

In detail, there are two hyperparameters in this process, which are search depth and sampling number. Search depth represents the maximum layer and sampling number controls the selected neighbor number. At each depth, we select top N targeting neighbor nodes by its attention weight, where N is equal to sampling number. The attention weight is calculated by neighbor node sequence $\boldsymbol{e}^{\boldsymbol{neigh}}$ and head node sequence $\boldsymbol{e}^{\boldsymbol{head}}$ obtained from

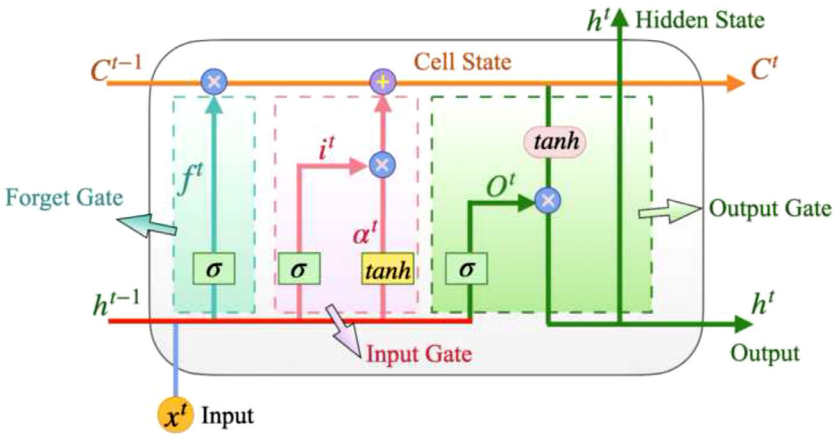

Fig. 3. The diagram of a single LSTM unit. The red line represents common part, which is the output from last unit. The blue line represents input of current time. Forget gate denoted by cyan line, the pink line represents input gate and the green line represents the output gate. Cell state is denoted by orange line and the output of current unit is denoted by green line.

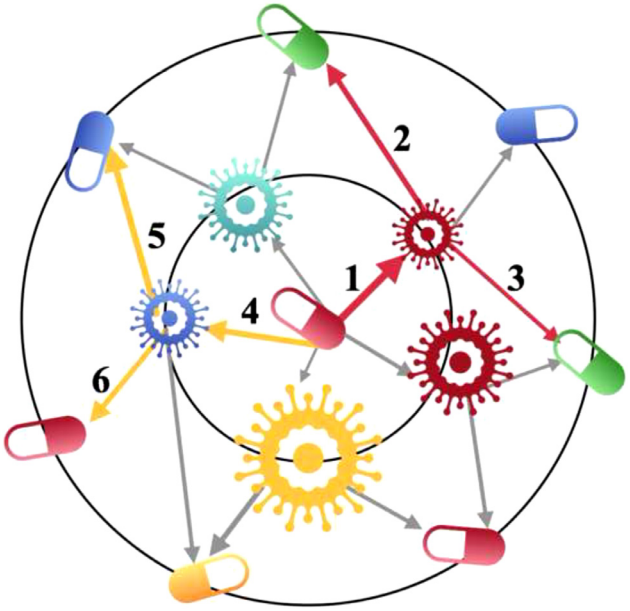

Fig. 4. The diagram of DFS-based on attention with search depth is 2 and sampling number is 2. The black circle represents the search depth and the lines between drugs and viruses denote real interactions. The linewidth represents attention weight. The red and yellow lines denote selected nodes and the number in it represents search path.

encoder–decoder layer, which is formulated by:

$$score\left(\boldsymbol{e}^{\boldsymbol{head}}, \boldsymbol{e}^{\boldsymbol{neigh}}\right) = \boldsymbol{e}^{\boldsymbol{head}^T}\boldsymbol{e}^{\boldsymbol{neigh}} \tag{12}$$

In real training process, the attention-based DFS is implemented for each node. In order to decrease the computational complexity, the attention weight can be calculated as following before DFS. Suppose the interaction matrix is $\boldsymbol{I} \in \mathbb{R}^{N(d)\times N(v)}$, drug embedding matrix is $\boldsymbol{D} \in \mathbb{R}^{N(d)\times d}$ and virus embedding matrix is $\boldsymbol{V} \in \mathbb{R}^{N(v)\times d}$, where d represents embedding dimension, the attention weight can be calculated by:

$$\boldsymbol{Score}_{\boldsymbol{attention}} = (\boldsymbol{D}^{\boldsymbol{T}}\boldsymbol{V}) \times \boldsymbol{I} \tag{13}$$

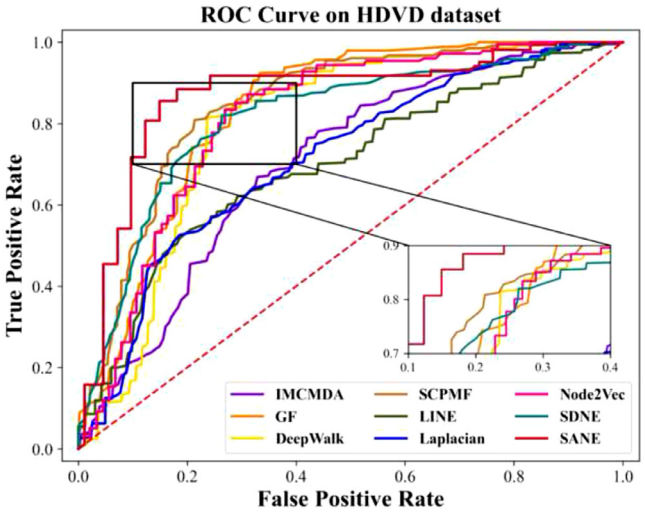

Fig. 5. Receiver operating characteristic (ROC) curves of various methods on HDVD dataset.

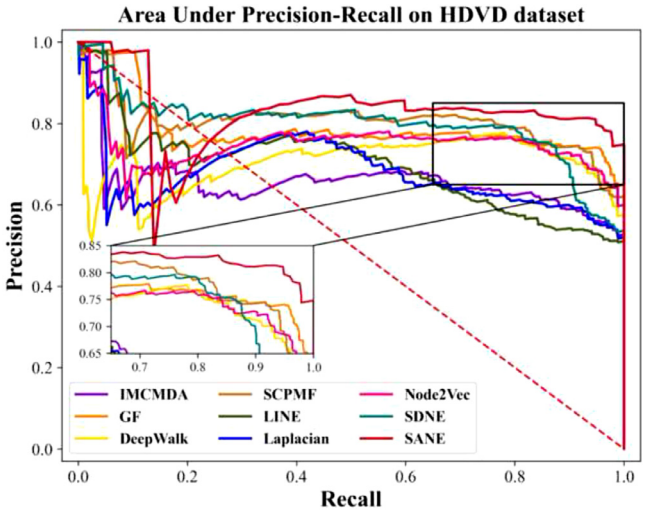

Fig. 6. Precision–recall (PR) curves of various methods on HDVD dataset.

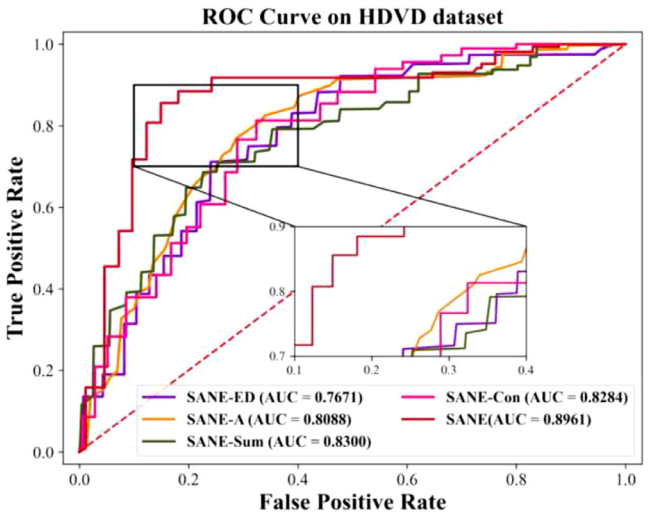

Fig. 7. Receiver operating characteristic (ROC) curves of different variants on HDVD dataset.

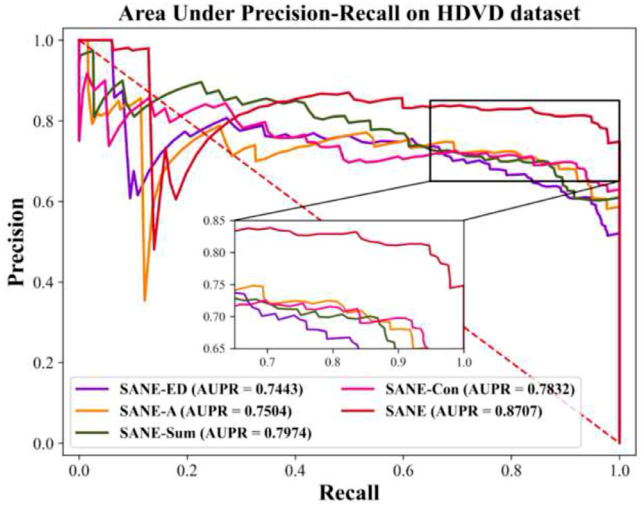

Fig. 8. Precision–recall (PR) curves of different variants on HDVD dataset.

3.4. Global feature extraction

After applying attention-based DFS to decrease graph scale and reduce noises, network embedding is used to learn network topology feature. Conceptually inspired by spatial-based GNN method [30]. Our work here can be regarded as that the targeted node is represented by nodes in its DFS network. The DFS network can be understood as the view of targeted node in different layers.

For first layer, neighbor representation $N_1(t)$ of targeted node t is calculated by attention weights and neighbor initial representation obtained from encoder–decoder layer, which can be formulated by:

$$\boldsymbol{N_1}(t) = \sum_{j \in V_1(t)} score\left(\boldsymbol{e_t}, \boldsymbol{e_j}\right) \boldsymbol{e_j} \tag{14}$$

$V_1(t)$ represents receptive field of targeted node t in first layer. Then, using first receptive field representation to update targeted node:

$$\boldsymbol{R_1}(t) = update\left(\boldsymbol{e_t}, \boldsymbol{N_1}(t)\right) \tag{15}$$

With the increasement of layer, updating each node in DFS network using its receptive field in next layer. Therefore, the whole process is a recursive process and can be represented by:

$$\boldsymbol{R_d}(t) = \begin{cases} \boldsymbol{e_i} & , d = 0 \\ \boldsymbol{update}\left(\boldsymbol{R_{d-1}(t)}, \sum_{j \in V_d(t)} score\left(\boldsymbol{e_t}, \boldsymbol{e_j}\right) \boldsymbol{e_j}\right) & , 0 < d < depth \end{cases} \tag{16}$$

Here, inspired by previous work [31], we also designed three kinds of update or aggregation function, including sum, concatenation and neigh. After obtaining the node representation, the concatenation of drug representation and virus representation is sent into a forward neural network for predicting the probability of interaction and sigmoid function is used in forward neural network.

3.5. Performance evaluation

In the experiment, Adam algorithm is adopted to optimize all trainable parameters and five-fold cross-validation is used to evaluate the performance of proposed method. The approved drug–virus associations as positive samples are randomly divided into training, validation and testing sets in the 8:1:1 manner, and we randomly sample the complement set of positive samples as negative samples, with an equal number of positive and negative samples in all phase. Moreover, two evaluation indicators are adopted to verify proposed model performance, including accuracy (ACC.), F1 Score, which can be defined as following formulations:

$$ACC. = \frac{TP + TN}{TP + TN + FP + FN} \tag{17}$$

$$F1\,Score = \frac{2TP}{2TP + FP + FN} \tag{18}$$

TP, TN, FP and FN in the formulation represent the number of correctly predicted positive samples, correctly predicted negative samples, incorrectly predicted samples and incorrectly predicted negative samples, respectively. Additionally, the area under the receiver operating characteristic (ROC) curve (AUC) and the area under precision–recall curve (AUPR) are calculated to reflect the model performance more comprehensively.

4. Experiments

4.1. Experimental settings

There are four hyperparameters in proposed model, including search depth, sampling number at each layer, learning rate and embedding dimension. We conducted grid search to obtain the optimal values of parameters. We set the parameters search depth and sampling number in the range from 1 to 8 with step of 1 and found that the model achieved the best performance when the search depth was set as 3 and sampling number was set as 4. We set the embedding dimension in the range from 8, 16, 32, 64, 128, 256 and the model performed best when it was set as 32. Then, we set the learning rate in the range from 0.0 to 0.05 with step of 0.0002 and found that the model obtained better performance when learning rate was 0.005. The detailed results will be discussed in later section Parameter Sensitivity Analysis. As a result, we select the search depth as 3, sampling number as 4 and embedding dimensionality of 32 for each node and we set the learning rate as 0.005 during the training process.

In addition, in order to prove the predictive ability of proposed model, we also select four types of computational models as baseline models, including similarity-based models and representation learning-based models (MF-based models, RW-based models and NN based models). We select two representative computational models for each type. For example, SCPMF [25] and IMCMDA [32] of similarity-based models, Laplacian [33] and

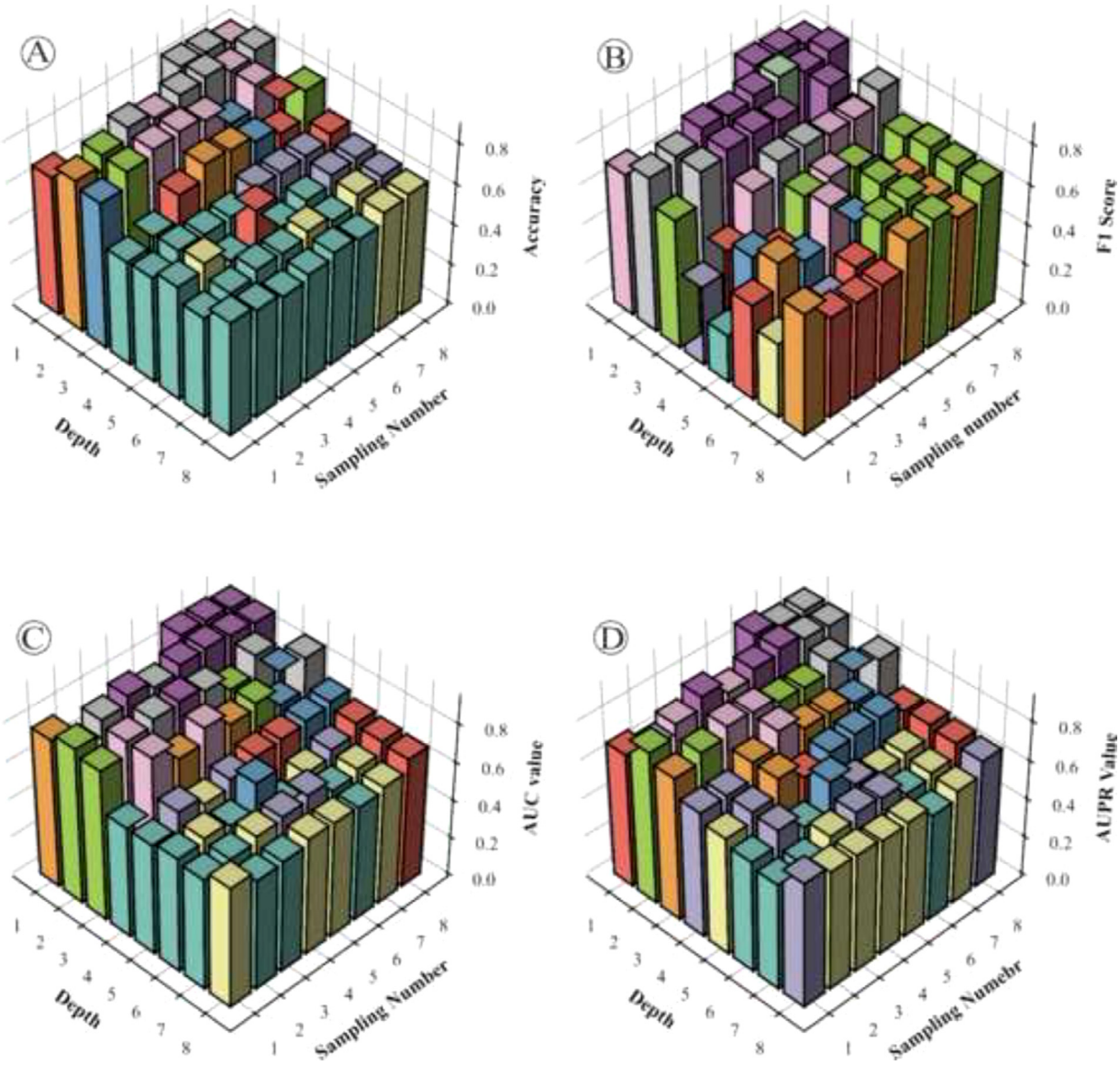

Fig. 9. Effect of parameters search depth and sampling number on SANE performance. (A) Effect of parameters search depth and sampling number on proposed model accuracy. (B) Effect of parameters search depth and sampling number on proposed model F1 Score. (C) Effect of parameters search depth and sampling number on proposed model AUC value. (D) Effect of parameters search depth and sampling number on proposed model AUPR value.

Graph Factorization [34] of MF-based models, Deepwalk [35] and Node2Vec [36] of RW-based models and LINE [37] and SDNE [38] of NN-based models. The parameters used in these baseline models are the same as their initial work.

4.2. Results and analysis

In the experiment, five-fold cross-validation is used to test proposed model on HDVD dataset. The proposed model performance is listed in Table 2. According to the results, it can be observed that proposed model achieved the prediction accuracy of 81.98% and its standard deviation was 0.662. As for the other evaluation indicators, proposed model obtained 83.43%, 0.8961 and 0.8707 on F1 score, AUC and AUPR, respectively.

To prove the effectiveness of proposed model, we compared it with above-mentioned baseline models using five-fold cross-validation for predicting drug–virus interactions. Table 3 showed the detailed results among four evaluation indicators and we also plotted ROC and PR curves to express the results intuitively (see Figs. 5 and 6). On the basis of experiment results,

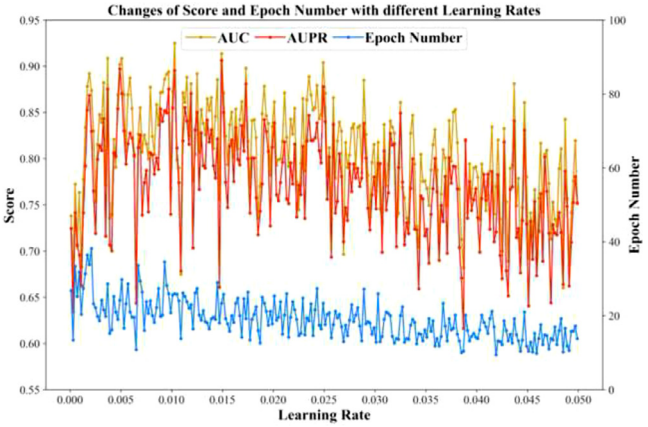

Fig. 10. The performance of SANE with the changes of learning rate.

Table 3
Results achieved by baseline models and SANE on HDVD dataset.

Model type	Method	ACC. (%)	F1 Score (%)	AUC	AUPR
Similarity-based	SCPMF	66.93 ± 6.31	69.35 ± 4.89	0.8631 ± 0.0456	0.5090 ± 0.0382
	IMCMDA	59.82 ± 6.89	60.41 ± 6.78	0.6423 ± 0.0683	0.1649 ± 0.0463
RL-MF-based	Laplacian	67.36 ± 2.74	65.38 ± 3.13	0.7217 ± 0.0323	0.6743 ± 0.0163
	GF	80.11 ± 2.83	81.10 ± 2.26	0.8541 ± 0.0247	0.7954 ± 0.0201
RL-RW-based	Deepwalk	77.14 ± 3.29	78.23 ± 2.88	0.8066 ± 0.0232	0.7085 ± 0.0166
	Node2Vec	78.57 ± 3.15	79.64 ± 2.69	0.8347 ± 0.0258	0.7600 ± 0.0239
RL-NN-based	LINE	66.26 ± 1.76	65.24 ± 1.56	0.6913 ± 0.0305	0.6950 ± 0.0279
	SDNE	80.55 ± 3.78	80.86 ± 3.42	0.8599 ± 0.0270	0.8091 ± 0.0170
Proposed Model	SANE	**81.98±6.62**	**83.43 ± 5.79**	**0.8961 ± 0.0335**	**0.8707 ± 0.0529**

Table 4
Results achieved by variants of SANE on HDVD dataset.

Model	ACC. (%)	F1 Score (%)	AUC	AUPR
SANE-ED	70.77 ± 3.97	73.78 ± 2.90	0.7671 ± 0.0257	0.7443 ± 0.0242
SANE-A	74.28 ± 7.93	75.28 ± 5.24	0.8088 ± 0.0452	0.7504 ± 0.0419
SANE-Sum	79.74 ± 6.88	74.48 ± 5.98	0.8300 ± 0.0395	0.7974 ± 0.0537
SANE-Con	73.63 ± 6.74	74.77 ± 6.10	0.8284 ± 0.0401	0.7832 ± 0.0533
SANE	**81.98±6.62**	**83.43 ± 5.79**	**0.8961 ± 0.0335**	**0.8707 ± 0.0529**

Table 5
Top 40 potential drugs against COVID-19 predicted by SANE.

Rank	Accession number	Drug name	Evidence
1	DB00558	Zanamivir	PMID: 15200845
2	DB04786	Suramin	–
3	DB00811	Ribavirin	PMID: 22555152
4	DB06412	Oxymetholone	PMID:12815555; PMID:32194980
5	DB00932	Tipranavir	–
6	DB13729	Camostat	PMID: 22496216
7	DB02187	Equilin	PMID: 27169275; PMID: 32194980
8	DB01065	Melatonin	PMID: 25262626; PMID: 20070490
9	DB15622	Triazavirin	–
10	DB13393	Emetine	PMID: 32147496
11	DB01072	Atazanavir	–
12	DB11758	Cenicriviroc	–
13	DB01029	Irbesartan	PMID: 32129518
14	DB12129	Tideglusib	–
15	DB07715	Emodin	PMID: 21050882; PMID: 16940925
16	DB00715	Paroxetine	PMID: 29272110; PMID: 32194980
17	DB01004	Ganciclovir	PMID: 32166607
18	DB00970	Dactinomycin	PMID: 1335030; PMID: 32194980
19	DB01394	Colchicine	PMID: 28795759; PMID: 32194980
20	DB00959	Methylprednisolone	–
21	DB14126	Tenofovir	PMID: 32222463
22	DB13609	Umifenovir	PMID: 18756809
23	DB13879	Glecaprevir	–
24	DB00864	Tacrolimus	–
25	DB01103	Quinacrine	PMID: 23301007; PMID: 31307979
26	DB00290	Bleomycin	–
27	DB13068	Nim811	–
28	DB00539	Toremifene	PMID: 31474372
29	DB01211	Clarithromycin	–
30	DB15661	EIDD-2801	PMID: 32253226
31	DB00477	Chlorpromazine	PMID: 8811199; PMID: 23529728
32	DB00244	Mesalazine	PMID: 17555580
33	DB00608	Chloroquine	PMID: 32074550
34	DB00478	Rimantadine	PMID: 31133031; PMID: 15288617
35	DB01264	Darunavir	PMID: 32671131
36	DB01024	Mycophenolic Acid	PMID: 5799033
37	DB00594	Amiloride	–
38	DB00441	Gemcitabine	PMID: 24841273
39	DB06290	Simeprevir	–
40	DB01118	Amiodarone	–

it can be found that proposed model achieved superior performance and outperformed those previous representative or state-of-the-art methods: SCPMF (AUC = 0.8631, AUPR = 0.5090), IMCMDA (AUC = 0.6423, AUPR = 0.1649), Laplacian (AUC = 0.7217, AUPR = 0.6743), Graph Factorization (AUC = 0.8541, AUPR = 0.7954), Deepwalk (AUC = 0.8066, AUPR = 0.7085), Node2Vec (AUC = 0.8347, AUPR = 0.7600), LINE (AUC = 0.6913, AUPR = 0.6950) and SDNE (AUC = 0.8599, AUPR = 0.8091). This is mainly because that proposed model SANE can capture high-dimension feature compared with similarity-based methods and SANE is able to learn both node basic information and drug–virus

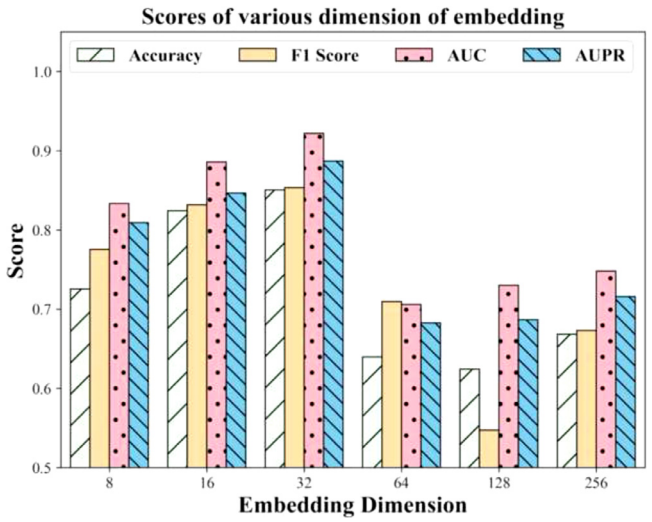

Fig. 11. The performance of SANE with the changes of embedding dimension.

network topology information when compared with representation learning-based methods.

4.3. Ablation study

Proposed model is constructed by three basic units, including encoder–decoder for extracting sequence feature, attention-based DFS to reduce redundancies and noises, aggregation strategy to calculate final representation. However, it is difficult to reflect the effectiveness of each basic unit in general. As a result, we conducted an ablation study to discuss the contribution of basic units and illustrate the reason why proposed model has a better performance than baseline models.

As a result, we redesigned four variants of proposed model, which were SANE-ED, SANE-A, SANE-Sum and SANE-Con. The differences between them and proposed model are that there are no sequence information and reduction in noises and redundancies contained in SANE-ED and SANE-A, respectively. The aggregation strategy is different in SANE-Sum and SANE-Con when compared with proposed model. One is sum aggregation and the other is concatenation. The other settings of the four variants are the same as proposed model. Five-fold cross-validation is used to test these models and the detailed results are listed in Table 4 and ROC curves and PR curves are shown in Figs. 7 and 8.

Firstly, integrating drug SMILE sequence and virus sequence can comprehensively improve the expressiveness of proposed model since proposed model obtained about 11%, 10%, 0.13 and 0.13 higher performance than SANE-ED on accuracy, F1 score, AUC and AUPR, respectively. However, even if SANE-ED has a normal performance, it still outperforms Laplacian and LINE on AUC and accuracy. Secondly, resulting from the experiment results between SANE-A and proposed model SANE, adopting attentive DFS has a positive effect on model performance. Actually, other search methods can also be applied in this study, such as Breadth-First-Search (BFS). But finally, we selected the DFS because it is a more efficient way to select nodes, takes less memory and facilitate the calculation of subsequent nodes. On the other hand, the performance of SANE-A is better than half of baselines on accuracy, AUC and AUPR, which further indicates the positive influence of attentive DFS. Furthermore, we also tested the variants with different aggregation strategies. It can be found that aggregation functions influence model results lightly. In fact, the model with both encoder–decoder and attentive DFS has overpassed more than a half of baseline models, but proposed model with neigh aggregation achieves the best performance over all evaluation indicators.

4.4. Hyperparameter sensitivity analysis

Hyperparameter sensitivity analysis is significant for the performance of a model in different scenarios [25]. There are four hyperparameters used in proposed model, including search depth, sampling number, learning rate and embedding dimension.

Firstly, we mainly focused on the search depth and sampling number and conducted five-fold cross-validation on the HDVD dataset to select parameters. Specifically, we investigated the influence of search depth and sampling number by varying from 1 to 8. Fig. 9 shows the proposed model performances with different parameter combinations. The *x-axis* represents search depth, *y-axis* represents sampling number and *z-axis* represents the value of evaluating indicator. As seen in Fig. 9(A)(B)(C)(D), proposed model has a stronger predictive ability at lower depth with same sampling number because more noises and irrelevant information are introduced into computational model with depth increasing. From this perspective and considering the model performance, we finally select 3 as default search depth. In addition, it also can be observed that the model has better performance with sampling number increasing. Taking all the factors together, including running time etc., we select sampling number as 4 owing to its the highest accuracy and AUC.

Secondly, learning rate is essential in training process, which determines the convergence rate and performance of the model. In order to obtain a robustness model, we tested the model performance with different learning rate, increasing it from 0 to 0.05 with a step of 0.0002. The change of model performance is shown in Fig. 10, the blue line represents epoch number and the other two lines are evaluating indicators. On the basics of Fig. 10, we finally choose 0.005 since it not only ensures the accuracy but also the training speed of the proposed model.

Finally, embedding dimension is also one of the important hyperparameters. We conducted embedding dimension experiments by increasing from 8 to 256. According to Fig. 11, proposed model achieved the best performance on accuracy, F1 score, AUC and AUPR when embedding dimension is 32. As a result, the default parameter of embedding dimension is 32.

4.5. Case study

In this section, case study is conducted to estimate the ability of proposed model to identify potential drugs against COVID-19. Specifically, COVID-19 virus sequence and 219 drug SMILES are sent into proposed model and then rank them according to the prediction score. After that, search predicted drugs in published literatures and known COVID-19 drug repurposing database to test proposed model predictive performance. In this way, we listed all proposed model-predicted drugs that scored more than 0.98 for COVID-19 and showed the ranking, drug Accession Number, literature-reported evidence and scores in descending order.

According to the results Table 5, 7 of top 10 (70%), 13 of the top 20 (65%), 18 of the top 30 (60%) and 25 of the top 40 (62.5%) drugs were verified by valuable dataset Excelra and literatures. For example, Zanamivir, which achieved the highest score, can selectively bind and inhibits virus neuraminidase-mediated cleavage of sialic acid residues in host cell membrane-bound glycoprotein receptors, preventing the release of progeny viruses from host cell surfaces [39]. Studies in vitro revealed that Zanamivir inhibits SARS coronavirus infection. The third drug, Ribavirin [40] was initially recommended in clinical practice for the China 2019-nCoV pneumonia diagnosis and Treatment Plan Edition 5-Revised. As for the fifth drug Oxymetholone, it was found to be effective against wasting associated with HIV infection in a clinical trial [41]. How it might be helpful in treating COVID19 infection is debatable [42] but it has been recorded in Excela COVID-19

drug repurposing dataset. The sixth drug, Camostat, can block SARS-CoV-2 infection of lung cells and could be considered for off-label treatment of COVID-19 infections [43,44]. Owing to Equilin blocked the cellular entry of a pseudovirus formed by an HIV-core packed with the Zaire Ebola virus glycoprotein in an in-vitro experiment and Zaire Ebola and COVID-19 belong to coronavirus [45]. Therefore, Equilin is recognized as a potential drug to treat COVID-19 [42]. The eighth drug Melatonin has been shown to target pathological alterations associated with an Ebola infection [46] such as endothelial disruption, disseminated intravascular coagulation and multiple organ hemorrhage. Melatonin plays an inhibitory role on lung oxidative stress induced by respiratory syncytial virus infection in mice [47] and it is also considered as one of the potential drugs to treat COVID-19 [42]. Additionally, Ref. [48] indicated the combination of remdesivir and emetine therapy may provide better clinical benefits. These empirical results indicate that proposed model has a strong predictive ability and can narrow the scope of candidates for further biological experiments.

5. Discussion

Drug repositioning as an effective drug development method provides a far more rapid option on the clinic than traditional drug discovery. In this study, we proposed a network-based drug repositioning method SANE to identify potential drugs for COVID-19.

There are three reasons why SANE has superior performances and outperforms representative models. First, SANE mines the associations in a deeper level, compared with similarity-based methods. SANE preserves the local structural feature by adopting aggregating neighbor information and learns the network topology by increasing the targeted node receptive field. By this way, SANE is not limited in 2D space, it can capture high dimensional and complex potential features. Second, compared with previous representative network-based methods, SANE introduces drug SMILES and virus sequence into model as node initial feature, increasing information granularity. Though previous work has integrated above two kinds of information, their models learn two features respectively and simply concatenate them together as the final representation. Unlike them, SANE embeds the attribute information into network representation learning. By doing this, the representation obtained by SANE contains not only network structural information but node attribute information. Third, SANE adopts attention-based learning strategy to ensure the stability and accuracy. Previous models select the nodes or paths randomly, but in our study SANE selects the neighbor nodes by attention weight, which is a more reasonable strategy. Then, the model can concentrate on the crucial part of network and reduce the random noises.

Though SANE performs well in COVID-19 drug repositioning, it still faces some challenges. SANE will be limited by the incomplete sequences or information lacking due to the attribute feature is extracted from drug SMILES or virus sequences. Besides, though SANE is able to get reliable initial node embeddings, it is challenged to maintain the node attribute similarity in the low dimension space.

6. Conclusion

In this study, we proposed an attention-based computational model SANE to repurpose drugs for COVID-19 based on drug–virus associations network. SANE integrated the drug SMILES and virus sequence into attention-based network embedding to infer commercially available drugs that could be applied to experimental therapy options against COVID-19. The result show that SANE has a strong ability in associations prediction and outperforms several classical network embedding models and similarity-based models. Further case study on COVID-19 show that SANE is a promising model to repurpose drugs against COVID-19 since 7 of top 10, 13 of top 20, 18 of top 30 and 25 of top 40 drugs are verified by recent literatures and some of them has been used to COVID-19 treatment in clinical trials. We hope that predicted approved drugs of SANE may be helpful in the future prevention of the transmission of COVID-19.

However, there is still room for further improvement. In the future, we will adopt various optimization methods [49] to further improve the prediction performance. Moreover, we will enlarge the network by integrating associations related to drugs, such as drug–drug interactions and drug–target interactions, to further enhance the information dimension, and adopt probability-based negative set sampling strategy to enhance the stability of the model.

CRediT authorship contribution statement

Xiaorui Su: Conceptualization, Methodology, Software. **Zhuhong You:** Supervision. **Lei Wang:** Project administration. **Lun Hu:** Project administration. **Leon Wong:** Data curation. **Boya Ji:** Formal analysis. **Bowei Zhao:** Visualization.

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.

Availability

The data underlying this article are available in https://github.com/Blair1213/SANE.

Acknowledgments

The authors would like to thank all anonymous reviewers for their constructive advice.

Funding

This work was supported in part by the National Natural Science Foundation of China, under Grants 62172355, in part by Awardee of the NSFC Excellent Young Scholars Program, under Grant 61722212, in part by the Tianshan youth - Excellent Youth, under Grant 2019Q029, in part by the Qingtan scholar talent project of Zaozhuang University, in part by the Natural Science Foundation of Xinjiang Uygur Autonomous Region under grant 2021D01D05 and the Pioneer Hundred Talents Program of Chinese Academy of Sciences.

References

[1] S.G. Siddell, The coronaviridae, in: The Coronaviridae, Springer, 1995, pp. 1–10.
[2] D.A.J. Tyrrell, J.D. Almeida, D.M. Berry, C.H. Cunningham, D. Hamre, M.S. Hofstad, L. Malluci, Mcintosh, Virology: Coronaviruses, Nature 220 (1968) 650-650.
[3] P.C.Y. Woo, Y. Huang, S.K.P. Lau, K.-Y. Yuen, Coronavirus genomics and bioinformatics analysis, Viruses 2 (2010) 1804–1820.
[4] J.A. Robb, C.W. Bond, Coronaviridae, in: Comprehensive Virology, Springer, 1979, pp. 193–247.
[5] Y. Ding, L. He, Q. Zhang, Z. Huang, X. Che, J. Hou, H. Wang, H. Shen, L. Qiu, Z. Li, Organ distribution of severe acute respiratory syndrome (SARS) associated coronavirus (SARS-CoV) in SARS patients: implications for pathogenesis and virus transmission pathways, The J. Pathol.: A J. Pathol. Soc. Great Britain Ireland 203 (2004) 622–630.

[6] I.K. Oboho, S.M. Tomczyk, A.M. Al-Asmari, A.A. Banjar, H. Al-Mugti, M.S. Aloraini, K.Z. Alkhaldi, E.L. Almohammadi, B.M. Alraddadi, S.I. Gerber, MERS-CoV Outbreak in Jeddah—a link to health care facilities, N. Engl. J. Med. 372 (2015) (2014) 846–854.
[7] K.G. Andersen, A. Rambaut, W.I. Lipkin, E.C. Holmes, R.F. Garry, The proximal origin of SARS-CoV-2, Nat. Med. 26 (2020) 450–452.
[8] D. Wu, T. Wu, Q. Liu, Z. Yang, The SARS-CoV-2 outbreak: what we know, Int. J. Infect. Dis. 94 (2020) 44–48.
[9] W.H. Organization, Coronavirus disease (COVID-19), 2020.
[10] E. Dong, H. Du, L. Gardner, An interactive web-based dashboard to track COVID-19 in real time, The Lancet Inf. Dis. 20 (2020) 533–534.
[11] P.K. Ozili, T. Arun, Spillover of COVID-19: impact on the global economy, available at SSRN 3562570, 2020.
[12] W. McKibbin, R. Fernando, The global macroeconomic impacts of COVID-19: Seven scenarios, Asian Econ. Pap. (2020) 1–55.
[13] T.T. Ashburn, K.B. Thor, Drug repositioning: identifying and developing new uses for existing drugs, Nat. Rev. Drug Discov. 3 (2004) 673–683.
[14] J. Langedijk, A.K. Mantel-Teeuwisse, D.S. Slijkerman, M. Schutjens, Drug repositioning and repurposing: terminology and definitions in literature, Drug Discov. Today 3 (2015) 1027–1034.
[15] J. Jourdan, R. Bureau, C. Rochais, P.D. Allemagne, Drug repositioning: a brief overview, J. Pharm. Pharmacol. (2020).
[16] R.J. Vane, Inhibition of prostaglandin synthesis as a mechanism of action for aspirin-like drugs, Nature New Biol. 231 (1971) 232.
[17] N. Raje, K. Anderson, Thalidomide–a revival story, N. Engl. J. Med. 341 (1999) 1606–1609.
[18] J. Cummings, G. Lee, T. Mortsdorf, A. Ritter, K. Zhong, Alzheimer's disease drug development pipeline: 2017, Alzheimer's Dementia : Transl. Res. Clin. Interv. 3 (2017) 367–384.
[19] S. Naylor, J.M. Schonfeld, Therapeutic drug repurposing, repositioning and rescue: Part I-overview, Drug Discov (2015).
[20] J. Li, Z. Lu, A new method for computational drug repositioning using drug pairwise similarity, in: 2012 IEEE International Conference on Bioinformatics and Biomedicine, 2012, pp. 1-4.
[21] J. Zhang, C. Li, Y. Lin, Y. Shao, S. Li, Computational drug repositioning using collaborative filtering via multi-source fusion, Expert Syst. Appl. 84 (2017) 281–289.
[22] A. Azad, M. Dinarvand, A. Nematollahi, J. Swift, F. Vafaee, A comprehensive integrated drug similarity resource for in-silico drug repositioning and beyond, Brief. Bioinform. 3 (2021) bbaa126.
[23] J. Li, S. Zheng, B. Chen, A.J. Butte, S.S. Joshua, Z. Lu, A survey of current trends in computational drug repositioning, Brief. Bioinform. 17 (2016) 2–12.
[24] S. Dotolo, A. Marabotti, A. Facchiano, R. Tagliaferri, A review on drug repurposing applicable to COVID-19, Brief. Bioinform. 22 (2021) 726–741.
[25] Y. Meng, M. Jin, X. Tang, J. Xu, Drug repositioning based on similarity constrained probabilistic matrix factorization: COVID-19 as a case study, Appl. Soft Comput. 103 (2021) 107–135.
[26] H. Luo, J. Wang, M. Li, J. Luo, P. Ni, K. Zhao, F. Wu, Y. Pan, Computational drug repositioning with random walk on a heterogeneous network, IEEE/ACM Trans. on Computational Biology & Bioinformatics 16 (2018) 1890–1900.
[27] Y. Luo, X. Zhao, J. Zhou, J. Yang, Y. Zhang, W. Kuang, J. Peng, L. Chen, J. Zeng, A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information, Nature Commun. 8 (2017) 573.
[28] X. Zeng, X. Song, T. Ma, et al., Repurpose open data to discover therapeutics for COVID-19 using deep learning, J. Proteome Res., J. Proteome Res. 19 (2020) 4624–4636.
[29] D.S. Wishart, Y.D. Feunang, A.C. Guo, E.J. Lo, A. Marcu, J.R. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda, DrugBank 5.0: a major update to the DrugBank database for 2018, Nucleic Acids Res. 46 (2018) D1074–D1082.
[30] W.L. Hamilton, R. Ying, J. Leskovec, Inductive representation learning on large graphs, in: Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017, pp. 1025-1035.
[31] X. Lin, Z. Quan, Z.J. Wang, T. Ma, X. Zeng, KGNN: Knowledge Graph Neural Network for Drug-Drug Interaction Prediction, in: Twenty-Ninth International Joint Conference on Artificial Intelligence and Seventeenth Pacific Rim International Conference on Artificial Intelligence, 2020, pp. 2739-2745.
[32] X. Chen, L. Wang, Q. Jia, N.N. Guan, J.Q. Li, Predicting mirna-disease association based on inductive matrix completion, Bioinformatics (2018) 4256–4265.
[33] M. Belkin, P. Niyogi, P. Niyogi, Laplacian eigenmaps for dimensionality reduction and data, Neural Comput. 15 (2003) 1373–1396.
[34] P. Goyal, E. Ferrara, Graph embedding techniques, applications, and performance: A survey, Knowl.-Based Syst. 151 (2017) 78–94.
[35] B. Perozzi, R. Al-Rfou, S. Skiena, DeepWalk: online learning of social representations, in: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery, New York, USA, 2014, pp. 701–710.
[36] A. Grover, J. Leskovec, node2vec: Scalable Feature Learning for Networks, in Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining, San Francisco, USA, 2016, pp. 855-864.
[37] J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, Q. Mei, Line: Large-scale information network embedding, in: Proceedings of the 24th international conference on world wide web, International World Wide Web Conferences Steering Committee, 2015, pp. 1067-1077.
[38] D. Wang, P. Cui, W. Zhu, Structural deep network embedding, in: Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining, San Francisco, USA, 2016, pp. 1225-1234.
[39] E.L. Tan, E.E. Ooi, C.-Y. Lin, H.C. Tan, A.E. Ling, B. Lim, L.W. Stanton, Inhibition of SARS coronavirus infection in vitro with clinically approved antiviral drugs, Emerg. Infect. Diseases 10 (2004) 581.
[40] J.S. Khalili, H. Zhu, N.S.A. Mak, Y. Yan, Y. Zhu, Novel coronavirus treatment with ribavirin: Groundwork for an evaluation concerning COVID-19, J. Med. Virol. 92 (2020) 740–746.
[41] U.R. Hengge, K. Stocks, S. Faulkner, H. Wiehler, C. Lorenz, W. Jentzen, D. Hengge, G. Ringham, Oxymetholone for the treatment of HIV-wasting: a double-blind, randomized, placebo-controlled phase III trial in eugonadal men and women, HIV Clinical Trials 4 (2003) 150–163.
[42] Y. Zhou, Y. Hou, J. Shen, Y. Huang, W. Martin, F. Cheng, Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2, Cell Discovery 6 (2019) 1–18.
[43] M. Hoffmann, H. Kleine-Weber, S. Schroeder, N. Krüger, T. Herrler, S. Erichsen, T.S. Schiergens, G. Herrler, N.-H. Wu, A. Nitsche, SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor, Cell 181 (2020) 271–280, e278.
[44] Y. Zhou, P. Vedantham, K. Lu, J. Agudelo, R. Carrion Jr, J.W. Nunneley, D. Barnard, S. Pöhlmann, J.H. McKerrow, A.R. Renslo, Protease inhibitors targeting coronavirus and filovirus entry, Ant. Res. 116 (2015) 76–84.
[45] L. Wang, Q. Chen, L. Zhou, Y. Guo, Study of gonadal hormone drugs in blocking filovirus entry of cells in vitro, Yao Xue Xue Bao= Acta Pharmaceutica Sinica 50 (2015) 1545–1550.
[46] D.X. Tan, A. Korkmaz, R.J. Reiter, L.C. Manchester, Ebola virus disease: potential use of melatonin as a treatment, J. Pineal Res. 57 (2014) 381–384.
[47] S.H. Huang, X.J. Cao, W. Liu, X.Y. Shi, W. Wei, Inhibitory effect of melatonin on lung oxidative stress induced by respiratory syncytial virus infection in mice, J. Pineal Res. 48 (2010) 109–116.
[48] F. Touret, X.D. Lamballerie, Of chloroquine and COVID-19, Ant. Res. 177 (2020).
[49] M. Ali, M.G. Sarowar, M.L. Rahman, J. Chaki, J. Tavares, Adam deep learning with SOM for human sentiment classification, Int. J. Ambient Comput. Intell. 10 (2019) 92–116.