

METHODOLOGY

Open Access



NEDD: a network embedding based method for predicting drug-disease associations

Renyi Zhou¹, Zhangli Lu¹, Huimin Luo^{1,2}, Ju Xiang^{1,3}, Min Zeng¹ and Min Li^{1*}

From The 18th Asia Pacific Bioinformatics Conference
Seoul, Korea. 18-20 August 2020

* Correspondence: limin@csu.edu.cn

¹Hunan Provincial Key Lab on Bioinformatics, School of Computer Science and Engineering, Central South University, Changsha, China
Full list of author information is available at the end of the article

Abstract

Background: Drug discovery is known for the large amount of money and time it consumes and the high risk it takes. Drug repositioning has, therefore, become a popular approach to save time and cost by finding novel indications for approved drugs. In order to distinguish these novel indications accurately in a great many of latent associations between drugs and diseases, it is necessary to exploit abundant heterogeneous information about drugs and diseases.

Results: In this article, we propose a meta-path-based computational method called NEDD to predict novel associations between drugs and diseases using heterogeneous information. First, we construct a heterogeneous network as an undirected graph by integrating drug-drug similarity, disease-disease similarity, and known drug-disease associations. NEDD uses meta paths of different lengths to explicitly capture the indirect relationships, or high order proximity, within drugs and diseases, by which the low dimensional representation vectors of drugs and diseases are obtained. NEDD then uses a random forest classifier to predict novel associations between drugs and diseases.

Conclusions: The experiments on a gold standard dataset which contains 1933 validated drug-disease associations show that NEDD produces superior prediction results compared with the state-of-the-art approaches.

Keywords: Drug repositioning, Heterogeneous network, Network embedding, Meta path

Background

Drug discovery is known for a large amount of money and time it consumes and the high risk it takes [1]. The investments grow continuously in recent years, but the total number of approved drugs remains constant [2]. Therefore, drug repositioning has become a popular approach to save cost by finding novel indications for approved drugs. Since these commercialized drugs have passed various clinical tests, it would save



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, 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 changes were made. 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/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

tremendous effort if we could reuse them directly. As reported, relaunching a repositioned drug can save about 80% of the cost compared with launching a reformulation of an existing drug [3].

The goal of drug repositioning is to find potential new target diseases for an existing drug and apply the newly identified drug to the treatment of diseases other than the drug's originally intended ones [4]. Historically, the discovery of finding new indications for existing drugs is mostly the result of a better understanding of a drug [5] or serendipity [6]. Then as the omics data accumulate, new bioinformatics methods emerge and play an increasingly important role. The newly proposed methods generally can be categorized as 'drug based' or 'disease based' [7]. For instance, IDMap [8] is a drug-based method and it mainly focuses on exploring the chemical structure information of drugs. Later, with the growth of drug-related data and the initiative of open data, recent studies pay more attention to integrating heterogeneous information. For example, Gottlieb et al. proposed PREDICT [9], a method that integrates various drug-drug similarity and disease-disease similarity from different sources.

These computational repositioning approaches can be roughly divided into three types: machine learning methods, text mining methods, and network-based methods [5, 10].

The aforementioned method PREDICT is an example of machine learning methods. The authors used similarities as features and applied a logistic regression classifier to predict novel indications for drugs. Moreover, Napolitano et al. proposed an approach which used a combination of drug-related data to train a multi-class SVM (Support Vector Machine) classifier to identify latent drug-disease associations [11]. Besides, there are also researches comparing traditional machine learning methods and deep learning methods [12, 13].

Electronic health records (EHR) of the patients and other literature contains a vast amount of information about drugs and diseases that can be explored using the text mining technique. For instance, Zhu et al. explored pharmacogenomics studies and modelled FDA-approved breast cancer drugs by using Semantic Web notions which support automated semantic inference [14]. Chen et al. integrated and annotated data from public datasets and developed a statistical model called Semantic Link Association Prediction (SLAP) to assess drug–target associations based on semantic links [15].

Network-based methods have been widely used for computational drug repositioning. Martínez V et al. proposed DrugNet [16], which is based on a heterogeneous network prioritization approach that can utilize heterogeneous information. Luo, Y et al. developed a pipeline called DTINet [17]. It originally aims to find interactions between drugs and targets, but can also be applied in drug repositioning. DTINet uses a matrix completion method to calculate the low dimensional feature vectors which capture the topological properties of nodes in the network and uses these features to predict novel associations. Luo, H et al. proposed MBiRW [18], which used the Bi-Random Walk algorithm to predict potential novel indications of a drug. However, current network-based methods often show a certain preference for drugs that have more known drug-disease associations. Therefore, they are not good at finding novel indications for drugs that are less explored or new drugs.

In this work, we propose NEDD, a network embedding based method for predicting novel interactions between drugs and diseases using heterogeneous information. NEDD tries to solve the above problems by adopting the concept of inductive learning and

meta path. The results of experiments show that NEDD outperforms other methods in drug repositioning.

Methods

In this section, we will introduce our method NEDD. Generally, the whole procedure of NEDD consists of three steps. First, based on the heterogeneous information related to drugs and diseases, an undirected graph with weighted edges is constructed. Second, we train a meta-path-based representation learning model to learn the embedding vectors of each entity. Last, using the vectors learned, we train a classifier to identify potential associations between drugs and diseases.

Construction of the drug-disease network

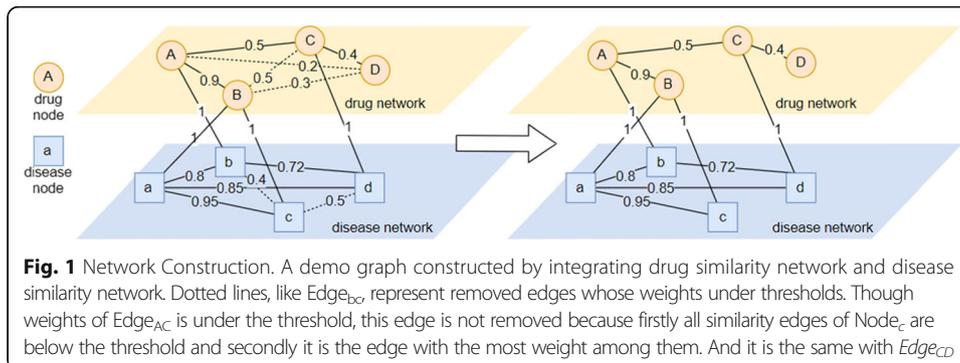
We construct the network as an undirected graph that consists of two node types (drug node n_{dr} and disease node n_{di}) and three edge types (drug-disease edge r_{dr-di} , drug-drug edge r_{dr-dr} , and disease-disease edge r_{di-di}). An edge of r_{dr-di} represents the association between a drug and a disease. An edge of r_{dr-dr} represents the connection between two drugs which have a high similarity and an edge of r_{di-di} denotes the connection between two diseases which have a high similarity.

After calculating all the similarities among drugs and that among diseases, we filter the edges with a certain threshold. In particular, the threshold of drug similarity is 0.8 and the threshold of disease similarity is 0.7. The thresholds are determined by experiments. Each similarity edge below its threshold is removed unless it has the greatest weight (similarity) among other homogeneous edges for a drug or a disease and removing this edge may cause one node to be isolated. The insight of filtering is that drug pairs with low similarity have an insignificant probability in indicating common diseases while drug pairs with high similarity have a strong probability to indicate common diseases [18], and the same is also true for disease pairs. This step is illustrated in Fig. 1.

Though here we only make use of three types of relationships within drugs and diseases, other types of information, like disease-target interactions and so on, can be used to expand the graph.

Network embedding

Most existing network representation learning methods can be summarized into two steps: proximity matrix construction and dimension reduction [19]. NEDD uses



random walk to perform the first step. One of the strengths of random walk is that it is efficient in both time and space. In the next step, NEDD adopts a neural-network-based method, HIN2vec [20], to learn network embedding vectors.

The key concept of the first step is the meta path. Given a drug-disease network $G(V, E)$, the node type set T_V and the edge type set T_E , a meta path can be defined as a triplet (n_h, n_t, m) , where $n_h, n_t \in T_V$ are the head node type and the end node type, and m is a sequence of edge types $r_1 \rightarrow r_2 \rightarrow \dots \rightarrow r_l$ ($r_1, r_2, \dots, r_l \in T_E$) which indicates a composite relation between the two node types. For example, a meta path $(n_{di}, n_{di}, r_{dr-di} \rightarrow r_{dr-di})$ indicates the relationship that two diseases share the same treatment, while a meta path $(n_{dr}, n_{dr}, r_{dr-di} \rightarrow r_{di-di} \rightarrow r_{dr-di})$ describes the relationship of two drugs that could treat two similar diseases.

NEDD first uses a random walk algorithm to generate long possible paths in the graph. In this process, node numbers and node types are recorded. Then, these paths are cut into shorter paths with lengths from 1 to W . These shorter paths represent the network proximity between two nodes of the 1st to the W^{th} order. In our experiments, W is set to 6. Besides, negative sampling [21] is used to generate negative data entries.

Then, NEDD adopts HIN2vec [20] to learn network embedding vectors. HIN2vec is a representation learning method. It assigns a low dimensional embedding vector for each entity in the graph, including drug nodes and disease nodes, and each meta path. It trains a neural network classifier with one hidden layer to identify if two nodes have a certain relationship and take weights of the hidden layer as embedding vectors like word2vec [21]. The prediction which node v_i and node v_j have an association that matches a particular meta path R is given as below:

$$P(R|v_i, v_j) = \text{sigmoid} \left(\sum e_{v_i} \odot e_{v_j} \odot f_{01}(e_R) \right)$$

where $e_{v_i}, e_{v_j}, e_R \in R^D$ is the embedding vector of the nodes v_i, v_j , and the meta path R , f_{01} is a regularization function that regularizes values in e_R within 0 and 1, and \odot is the element-wise product. HIN2vec uses cross-entropy loss to measure the prediction error and uses stochastic gradient descent to update embedding vectors.

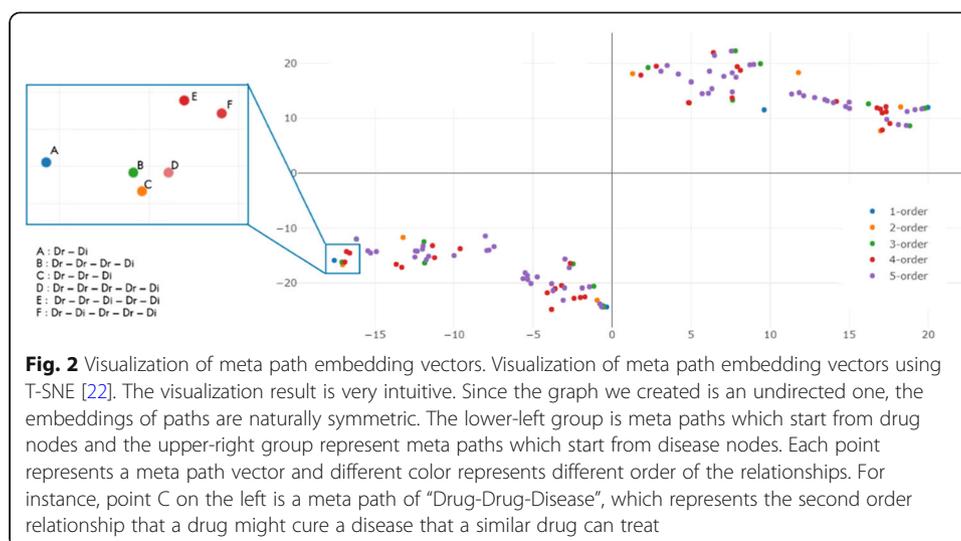
A visualization using T-SNE [22] of meta path embedding vectors is shown in Fig. 2.

Using embedding vectors to predict novel associations

In this step, we use a random forest classifier to make final predictions. We apply the element-wise product to aggregate a drug node embedding vector and a disease node embedding vector together as the input of the random forest classifier. The output of the random forest classifier is the predicted probability that a drug and a disease have associations:

$$P(\text{drug}_i \text{ treats disease}_j) = \text{random_forest_classifier}(e_{\text{drug}_i} \odot e_{\text{disease}_j})$$

We use scikit-learn [23] to implement the random forest classifier. While optimizing the parameters of the random forest, we evaluated the forest on the same training set and validation set. Because random forests are less likely to overfit, we started with a large classifier and gradually cut down the scale. In the end, we set the max depth of trees to 25, the number of estimators to 300, the minimum number of samples to split to 2, and use the Gini coefficient as the criterion.



Results

In this section, we evaluate the performance of NEDD on the gold standard dataset. First, we introduce the evaluation metrics. Then, by performing these measurements, we compare NEDD with five other methods: MBiRW [18], DTINet [17], HGBI [24], NBI [25], and JUST [26].

Dataset

In the dataset, drug-disease associations are collected from multiple data sources. This gold standard dataset which has been used in reference [9] includes 593 drugs from DrugBank [27], 313 diseases from the Online Mendelian Inheritance in Man (OMIM) [28] and 1933 validated drug-disease associations.

The drug similarity data is calculated by the Chemical Development Kit (CDK) [29] based on SMILES [30] chemical structures and the disease similarity data is obtained from MimMiner [31] which is based on disease phenotype similarity using text mining analysis of their medical descriptions information in the OMIM database.

Evaluation metrics

In order to evaluate the ability of NEDD in finding new possible target diseases of a specific drug, we conduct 10-fold cross-validation and perform the top-ranked candidate disease analysis.

In 10-fold cross-validation, all 1933 known drug-disease associations in gold standard datasets are randomly divided into 10 partitions with each roughly equal in size. Then, 1 of 10 partitions in turn serves as the test set, while the remaining as the training set. After the whole process, each possible association is given a score representing the confidence of the association. Then these associations are sorted in descending order according to their score. Next, for each ranking threshold, true positive rate (TPR), which measures the proportion of known associations that are correctly identified, and false positive rate (FPR), which measures the proportion of unverified associations that are predicted as real associations, are calculated based on the ranking results. By changing

settings of the threshold, we can get various pairs of TPR and FPR. Based on these pairs, the receiver operating characteristic curve (ROC curve) can be drawn with FPR as the x-axis and TPR as the y-axis. The area under ROC (AUC) is then calculated to measure the performance.

Besides, as top ranked predicting results may be important in practice, we also test our method in terms of top ranked results. We count the number of true drug-disease associations proved by other sources, e.g. other datasets or literatures to evaluate NEDD and other methods.

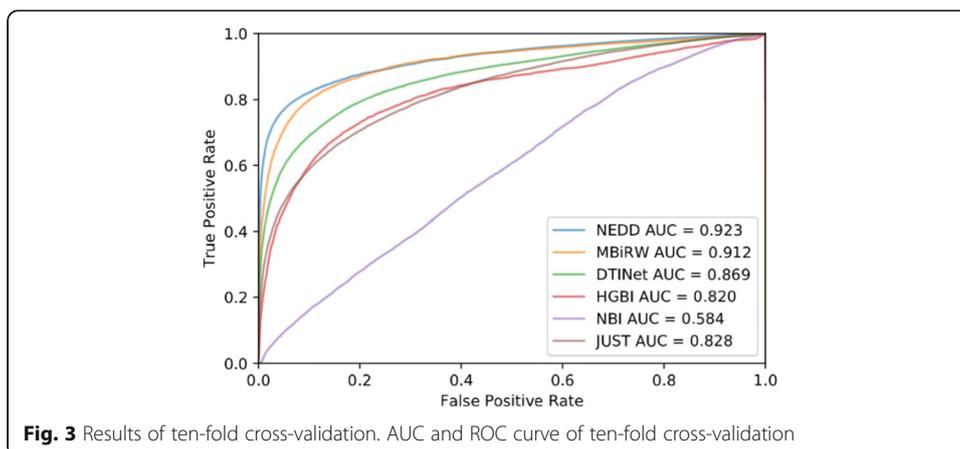
Comparison with other methods

NEDD is compared with five state-of-the-art methods: MBiRW [18], DTINet [17], HGBI [24], NBI [25], and JUST [26]. These five methods are all computational methods which are based on network and can utilize the heterogeneous network of drugs and diseases. NBI is a method based on a two-state diffusion model in a bipartite graph. HGBI is a method based on the guilt-by-association principle and an intuitive interpretation of information flow on a heterogeneous graph. MBiRW and DTINet have been introduced in the above. MBiRW is based on random walk and DTINet is based on matrix completion. Moreover, similar to NEDD, DTINet also adopts the concept of inductive learning. HGBI and NBI are also originally developed for drug-target association prediction but they have also been used in drug-disease association prediction [32]. JUST improves the random walk sampling method on the heterogeneous network which avoids using meta paths and use Skip-Gram model [21] to learn network embedding vectors. In the study, the parameter α is set to 0.3, and l, r to 2 for MBiRW according to the default parameter setting in [18]. The parameter α of HGBI is set to 0.4 as suggested in [24]. We change the lengths of vectors to 100 in DTINet, and we use the default settings for other parameters. Besides, we use additional information to train DTINet model because the method mainly focuses on integrating various information. The additional information includes drug-drug interactions and drug-protein interactions collected from DrugBank [27], associations between drugs and side-effects from SIDE R [33], and disease-gene associations from CTD [34]. For JUST, we set α to 0.4, m to 2 as suggested in the original paper [26]. The length of the embedding vectors is set to 128 and the window size of the Skip-Gram model is set to 10.

Through repeating the 10-fold cross-validation experiment specified in the above 1 hundred times with different random seeds, we calculate average AUC to estimate the performance of each method. In the 10-fold cross-validation test, the experiment results show that NEDD outperforms the other methods. NEDD achieves the AUC value of 0.923, while MBiRW, DTINet, HGBI, NBI, and JUST obtain inferior results of 0.912, 0.869, 0.820, 0.584 and 0.828, respectively. The results are illustrated in Fig. 3.

Case study

After confirming the ability of NEDD to predict potential drug-disease associations based on 10-fold cross-validation, we further conduct a case study to search evidences in other sources. In this step, all known associations in the gold standard dataset are used as the training set, and then the possible associations are ranked according to NEDD's prediction. Next, top-ranked predictions for each drug are verified based on CTD [34]. The top-N results of NEDD are summarized in Fig. 4. In this step, it shows



that 204 novel indications in the top 5 predictions are verified by CTD; 438 novel indications are verified in the top 10 and 849 novel indications are verified in the top 20.

Fifteen drug-disease associations with the highest prediction scores of all are listed in Table 1. None of these associations are verified by CTD. So, we conduct a further study to find supporting evidences by literature searching. The result shows that 7 out of 15 predictions are verified by literature. In predictions ranked 1 and 6, drugs are related to corresponding diseases, but may target another subtype of the diseases. Predictions ranked 2, 3, 5, 7, 11, and 12 have not been adequately investigated. Some details are provided below.

Levetiracetam can treat epilepsy but one of its side-effects is ataxia [35]. So, it may not be helpful in treating ataxia with myoclonic epilepsy and presenile dementia. Ifosfamide is predicted to treat reticulum cell sarcoma and it has been used in treating soft tissue sarcoma [43]. Enalapril is an orally-active antihypertensive agent that can suppress the renin-angiotensin-aldosterone system to lower blood pressure. NEDD predicts Enalapril can treat renal failure, which has been tested in [36]. Teniposide is used for the treatment of refractory acute lymphoblastic leukaemia. The prediction result suggests that it may be applied in the treatment of mismatch repair cancer syndrome (MMRCS) as well. Dihydrotychsterol is predicted to treat vitamin D-dependent rickets, type 2a, which has been studied in [37]. NEDD suggests using Ibandronate in the treatment of inclusion body myopathy with early-onset Paget disease with or without frontotemporal dementia 1 and the treatment of Paget

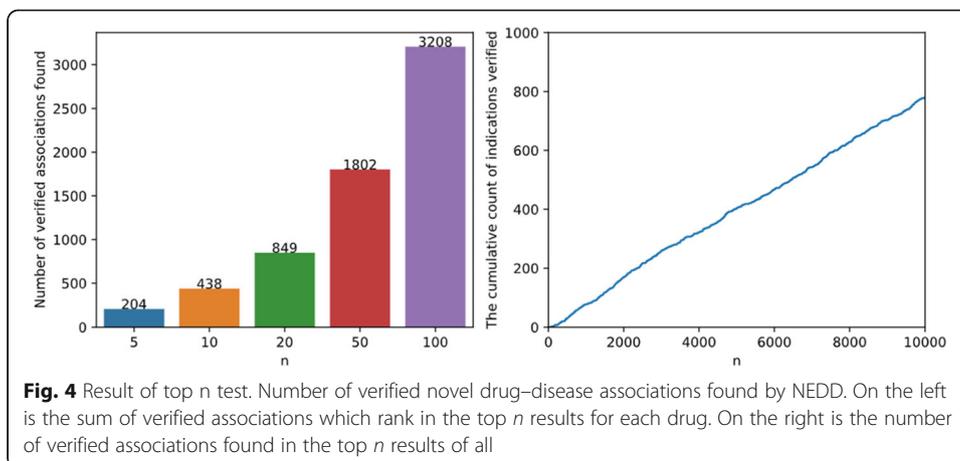


Table 1 Associations with the highest prediction scores

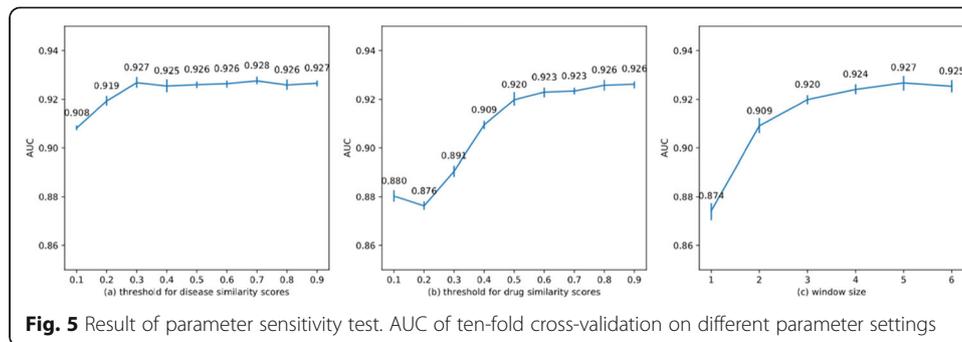
Rank	Drug		Disease		References
	ID (DrugBank)	Name	ID (OMIM)	Name	
1	DB01202	Levetiracetam	208,700	ATAXIA WITH MYOCLONIC EPILEPSY AND PRESENILE DEMENTIA	[35]
2	DB01181	Ifosfamide	267,730	RETICULUM CELL SARCOMA	–
3	DB00937	Diethylpropion	303,110	CHOROIDEREMIA, DEAFNESS, AND MENTAL RETARDATION	–
4	DB00584	Enalapril	161,900	RENAL FAILURE, PROGRESSIVE, WITH HYPERTENSION; RFH1	[36]
5	DB00444	Teniposide	276,300	MISMATCH REPAIR CANCER SYNDROME; MMRC5	–
6	DB01070	Dihydrotachysterol	277,440	VITAMIN D-DEPENDENT RICKETS, TYPE 2A; VDDR2A	[37]
7	DB00176	Fluvoxamine	131,300	CAMURATI-ENGELMANN DISEASE; CAEND	–
8	DB00710	Ibandronate	167,320	INCLUSION BODY MYOPATHY WITH EARLY-ONSET PAGET DISEASE WITH OR WITHOUT FRONTOTEMPORAL DEMENTIA 1; IBMPFD1	[38]
9	DB00710	Ibandronate	602,080	PAGET DISEASE OF BONE 2, EARLY-ONSET; PDB2	[38]
10	DB00282	Pamidronic acid	602,080	PAGET DISEASE OF BONE 2, EARLY-ONSET; PDB2	[39]
11	DB01551	Dihydrocodeine	147,530	INSENSITIVITY TO PAIN WITH HYPERPLASTIC MYELINOPATHY	–
12	DB00500	Tolmetin	147,530	INSENSITIVITY TO PAIN WITH HYPERPLASTIC MYELINOPATHY	–
13	DB00136	Calcitriol	241,519	HYPOPHOSPHATEMIA, RENAL, WITH INTRACEREBRAL CALCIFICATIONS	[40]
14	DB00214	Torsemide	256,370	NEPHROTIC SYNDROME, TYPE 4; NPHS4	[41]
15	DB01120	Gliclazide	600,496	MATURITY-ONSET DIABETES OF THE YOUNG, TYPE 3; MODY3	[42]

disease of bone 2, early-onset. And Ibandronate has long been used in treating Paget disease [38]. Pamidronic acid is used to prevent bone loss and to strengthen the bone in Paget disease [39], which verified the prediction of our method. Calcitriol, an active metabolite of vitamin D is predicted to treat renal hypophosphatemia with intracerebral calcifications. In real life, it is used in the treatment of hypophosphatemia [40]. Despite concerns that the use of calcitriol may contribute to vascular calcification, there is no clear evidence [44]. Torsemide is a high-ceiling loop diuretic [45]. It is predicted to treat nephrotic syndrome, type 4; this prediction is verified in [41]. For Gliclazide, the prediction to treat maturity-onset diabetes of the young of type 3 has been tested in clinical trials [42].

Parameter sensitivity

In this section, we investigate the parameter sensitivity. We change the thresholds for drug similarities and disease similarities and the window size to see how these parameters affect the result. We conduct 10-fold cross-validation five times for each parameter setting and evaluate the performance using AUC values. In each experiment, we only change one corresponding parameter and set the others as default—i.e. 0.7 for disease similarity threshold, 0.8 for drug similarity threshold, and 6 for window size.

The results are illustrated in Fig. 5. From Fig. 5, we can recognize a similar pattern. The performance rises initially when the values of the corresponding parameters rise.



However, after a certain point, NEDD becomes insensitive towards that parameter. For window size, this is because most of the useful information is already encoded in the embedding vectors. For thresholds of similarity scores, this is because most of the noise has already been ruled out at the beginning when we raise the thresholds.

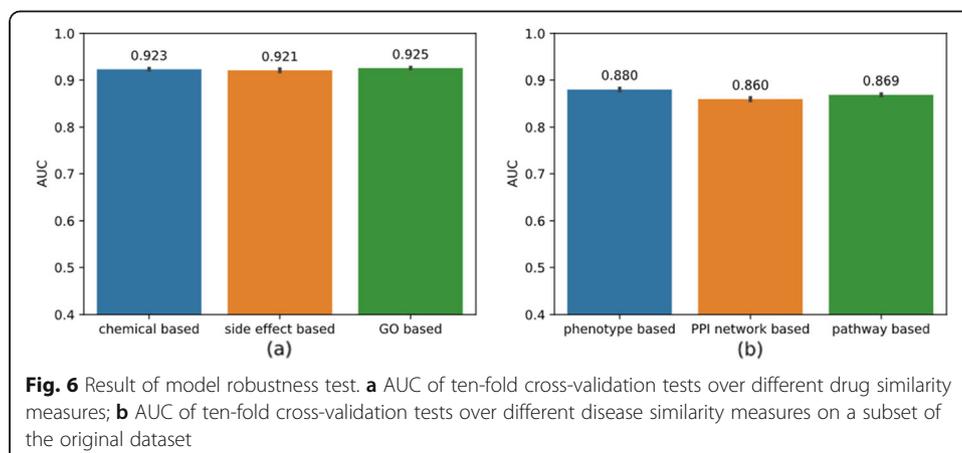
Model robustness

In this section, we investigate model robustness over different similarity measures.

We evaluate NEDD’s performance using three different disease similarity measures, i.e. MimMiner [31] which is in the golden test dataset and utilizes disease phenotype information, NetSim [46] which employs the protein interaction network, and RADAR [47] which we used to get similarity scores based on pathways. Since we use the disease similarity scores provided by the authors and they used other types of ID for disease and some IDs do not have any mapping information, the experiments are done on a subset of the original dataset, which consists of 196 diseases, 593 drugs, and 1052 drug-disease associations.

We also evaluate NEDD’s performance using three different drug similarity measures which utilize the information from the chemical structure, the corresponding side effects of each drug and the drug-related genes respectively. The three different types of drug similarity are calculated according to [9].

We repeat 10-fold cross-validation 10 times on each type of similarity scores. The results are illustrated in Fig. 6, in which NEDD produces similar AUC over different disease similarity measures.



Discussion

We think that the superior performance of NEDD in finding indications for new drugs stems from two aspects: inductive learning and meta path. Inductive learning methods can be applied to entities not seen at train time. Compared with MBiRW, HGBI and NBI, DTINet and NEDD, which used inductive methods, yields relatively higher AUC score in the test. And with the concept of meta path, NEDD is able to explicitly capture high order proximities. This is especially important in tasks like drug repositioning where many latent links between drugs and diseases are unknown. If the associations between two nodes are missing, their first-order proximity is zero, so it is essential to exploit high order information.

Besides, NEDD can be easily adopted in larger datasets with more types of biological entities such as target, gene, side-effect, etc.

However, the limitations of NEDD should also be acknowledged. First, in order to use various information like drug-target interactions, the maximum length of the meta path should be increased, which might significantly increase the computational cost. Second, because the trained embeddings are not specifically fine-tuned for association prediction between drugs and diseases, the difficulty in training the vectors is increased when adding more information to the network.

Conclusion

In this work, we present NEDD, a new computational approach for drug repositioning. NEDD uses a meta-path-based representation method to inductively learn node embedding vectors of drugs and diseases on a given graph. The graph is constructed by integrating heterogeneous biological information related to drugs and diseases. After learning the network embedding vectors, a random forest classifier is trained to predict the probabilities of drugs and diseases being associated.

NEDD shows competitive results in the 10-fold cross-validation test. The case study of NEDD gives fair results that remain to be further explored. In summary, the results prove that NEDD is practical in drug repositioning tasks toward existing drugs.

Abbreviations

SVM: Support vector machine; HER: Electronic health records; OMIM: Online Mendelian Inheritance in Man; CDK: Chemical Development Kit; TPR: True positive rate; FPR: False positive rate; AUC: Area under curve; ROC: Receiver operating characteristic

Acknowledgements

Not applicable.

About this supplement

This article has been published as part of *BMC Bioinformatics Volume 21 Supplement 13, 2020: Selected articles from the 18th Asia Pacific Bioinformatics Conference (APBC 2020): bioinformatics*. The full contents of the supplement are available online at <https://bmcbioinformatics.biomedcentral.com/articles/supplements/volume-21-supplement-13>.

Authors' contributions

Method design: RZ, ZL; overall study design: ML, HL; coding: RZ, ZL; literature and database search: JX, MZ; comparing with other methods: JX, MZ; analysis of results: ML, MZ, HL; paper writing: RZ, JX. All authors have read and approved the final manuscript.

Funding

Publication costs are funded by the National Natural Science Foundation of China under Grants (No. 61832019).

Availability of data and materials

The used dataset can be downloaded from bioinformatics.csu.edu.cn/resources/softs/DrugRepositioning/n_Web/NEDD.html.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Hunan Provincial Key Lab on Bioinformatics, School of Computer Science and Engineering, Central South University, Changsha, China. ²School of Computer and Information Engineering, Henan University, Kaifeng 475001, China.

³Neuroscience Research Center & School of Basic Medical Sciences, Changsha Medical University, Changsha, China.

Published: 17 September 2020

References

- Dudley JT, Deshpande T, Butte AJ. Exploiting drug–disease relationships for computational drug repositioning. *Brief Bioinform.* 2011;12(4):303–11.
- Ashburn TT, Thor KB. Drug repositioning: identifying and developing new uses for existing drugs. *Nat Rev Drug Discov.* 2004;3(8):673.
- Persidis A. The benefits of drug repositioning. *Drug Discov World.* 2011;12:9–12.
- Shim JS, Liu JO. Recent advances in drug repositioning for the discovery of new anticancer drugs. *Int J Biol Sci.* 2014;10(7):654.
- Yella J, Yaddanapudi S, Wang Y, Jegga A. Changing trends in computational drug repositioning. *Pharmaceuticals.* 2018;11(2):57.
- Bolgár B, Arany A, Temesi G, Balogh B, Antal P, Matyus P. Drug repositioning for treatment of movement disorders: from serendipity to rational discovery strategies. *Curr Top Med Chem.* 2013;13(18):2337–63.
- Shaughnessy AF. Old drugs, new tricks. *BMJ.* 2011;342:d741.
- Ha S, Seo YJ, Kwon MS, Chang BH, Han CK, Yoon JH. IDMap: facilitating the detection of potential leads with therapeutic targets. *Bioinformatics.* 2008;24(11):1413–5.
- Gottlieb A, Stein GY, Ruppin E, Sharan R. PREDICT: a method for inferring novel drug indications with application to personalized medicine. *Mol Syst Biol.* 2011;7(1):496.
- Li J, Zheng S, Chen B, Butte AJ, Swamidass SJ, Lu Z. A survey of current trends in computational drug repositioning. *Brief Bioinform.* 2015;17(1):2–12.
- Napolitano F, Zhao Y, Moreira VM, Tagliaferri R, Kere J, D'Amato M, Greco D. Drug repositioning: a machine-learning approach through data integration. *J Cheminform.* 2013;5(1):30.
- Aliper A, Plis S, Artemov A, Ulloa A, Mamoshina P, Zhavoronkov A. Deep learning applications for predicting pharmacological properties of drugs and drug repurposing using transcriptomic data. *Mol Pharm.* 2016;13(7):2524–30.
- Ramsundar B, Kearnes S, Riley P, Webster D, Konerding D, Pande V. Massively multitask networks for drug discovery. *arXiv preprint arXiv.* 2015:1502.02072.
- Zhu Q, Tao C, Shen F, Chute CG. Exploring the pharmacogenomics knowledge base (PharmGKB) for repositioning breast cancer drugs by leveraging web ontology language (OWL) and cheminformatics approaches. *Biocomput.* 2014;2014:172–82.
- Chen B, Ding Y, Wild DJ. Assessing drug target association using semantic linked data. *PLoS Comput Biol.* 2012;8(7):e1002574.
- Martínez V, Navarro C, Cano C, Fajardo W, Blanco A. DrugNet: network-based drug–disease prioritization by integrating heterogeneous data. *Artif Intell Med.* 2015;63(1):41–9.
- Luo Y, Zhao X, Zhou J, Yang J, Zhang Y, Kuang W, et al. A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. *Nat Commun.* 2017;8(1):573.
- Luo H, Wang J, Li M, Luo J, Peng X, Wu FX, Pan Y. Drug repositioning based on comprehensive similarity measures and bi-random walk algorithm. *Bioinformatics.* 2016;32(17):2664–71.
- Yang C, Sun M, Liu Z, Tu C. Fast network embedding enhancement via high order proximity approximation. In: *Proceedings of the 26th International Joint Conference on Artificial Intelligence.* United States: AAAI Press; 2017:3894–900.
- Fu TY, Lee WC, Lei Z. Hin2vec: explore meta-paths in heterogeneous information networks for representation learning. In: *Proceedings of the 2017 ACM conference on information and knowledge management.* United States: Association for Computing Machinery; 2017. 1797–806.
- Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J. Distributed representations of words and phrases and their compositionality. In: *Advances in neural information processing systems;* 2013. p. 3111–9.
- Maaten LVD, Hinton G. Visualizing data using t-SNE. *J Mach Learn Res.* 2008;9:2579–605.
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in python. *J Mach Learn Res.* 2011;12:2825–30.
- Wang W, Yang S, Li JING. Drug target predictions based on heterogeneous graph inference. *Biocomput.* 2013;2013:53–64.
- Cheng F, Liu C, Jiang J, Lu W, Li W, Liu G, et al. Prediction of drug-target interactions and drug repositioning via network-based inference. *PLoS Comput Biol.* 2012;8(5):e1002503.
- Hussein R, Yang D, Cudré-Mauroux P. Are Meta-paths necessary?: revisiting heterogeneous graph Embeddings. In: *Proceedings of the 27th ACM international conference on information and knowledge management.* United States: Association for Computing Machinery; 2018. p. 437–46.

27. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* 2006;34(suppl_1):D668–72.
28. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.* 2005;33(suppl_1):D514–7.
29. Steinbeck C, Han Y, Kuhn S, Horlacher O, Luttmann E, Willighagen E. The chemistry development kit (CDK): an open-source Java library for chemo- and bioinformatics. *J Chem Inf Comput Sci.* 2003;43(2):493–500.
30. Weininger D, Weininger A, Weininger JL. SMILES. 2. Algorithm for generation of unique SMILES notation. *J Chem Inf Comput Sci.* 1989;29(2):97–101.
31. Van Driel MA, Bruggeman J, Vriend G, Brunner HG, Leunissen JA. A text-mining analysis of the human phenome. *Eur J Hum Genet.* 2006;14(5):535.
32. Wang W, Yang S, Zhang X, Li J. Drug repositioning by integrating target information through a heterogeneous network model. *Bioinformatics.* 2014;30(20):2923–30.
33. Kuhn M, Letunic I, Jensen LJ, Bork P. The SIDER database of drugs and side effects. *Nucleic Acids Res.* 2015;44(D1):D1075–9.
34. Davis AP, Grondin CJ, Johnson RJ, Sciaky D, King BL, McMorran R, et al. The comparative toxicogenomics database: update 2017. *Nucleic Acids Res.* 2016;45(D1):D972–8.
35. Farooq MU, Bhatt A, Majid A, Gupta R, Khasnis A, Kassab MY. Levetiracetam for managing neurologic and psychiatric disorders. *Am J Health Syst Pharm.* 2009;66(6):541–61.
36. Abraham PA, Opsahl JA, Halstenson CE, Keane WF. Efficacy and renal effects of enalapril therapy for hypertensive patients with chronic renal insufficiency. *Arch Intern Med.* 1988;148(11):2358–62.
37. Rosen JF, Finberg L. Vitamin D-dependent rickets: actions of parathyroid hormone and 25-hydroxycholecalciferol. *Pediatr Res.* 1972;6(6):552.
38. Grauer A, Heichel S, Knaus J, Dösch E, Ziegler R. Ibandronate treatment in Paget's disease of bone. *Bone.* 1999;24(5 Suppl):87S–9S.
39. Seifi M, Amdjadi P, Tayebi L. Pharmacological agents for bone remodeling: an experimental approach. In: *Biomaterials for oral and dental tissue engineering.* United Kingdom: Woodhead Publishing; 2017. p. 503–23.
40. Imel EA, Econs MJ. Approach to the hypophosphatemic patient. *J Clin Endocrinol Metabol.* 2012;97(3):696–706.
41. Qavi AH, Kamal R, Schrier RW. Clinical use of diuretics in heart failure, cirrhosis, and nephrotic syndrome. *Int J Nephrol.* 2015;2015:975934.
42. Habeb AM, George ET, Mathew V, Hattersley AL. Response to oral gliclazide in a pre-pubertal child with hepatic nuclear factor-1 alpha maturity onset diabetes of the young. *Ann Saudi Med.* 2011;31(2):190–3.
43. Suppiah R, Wood L, Elson P, Budd GT. Phase I/II study of docetaxel, ifosfamide, and doxorubicin in advanced, recurrent, or metastatic soft tissue sarcoma (STS). *Investig New Drugs.* 2006;24(6):509–14.
44. Wolisi, G. O., & Moe, S. M. (2005). Vitamin D in health and disease: the role of vitamin D in vascular calcification in chronic kidney disease. In *Seminars in dialysis.* Oxford: Blackwell Science Inc.; 18(4):307–314.
45. Lopez B, González A, Hermida N, Laviades C, Díez J. Myocardial fibrosis in chronic kidney disease: potential benefits of torsemide: new strategies to prevent cardiovascular risk in chronic kidney disease. *Kidney Int.* 2008;74:S19–23.
46. Li P, Nie Y, Yu J. Fusing literature and full network data improves disease similarity computation. *BMC Bioinform.* 2016; 17(1):326.
47. Qin R, Duan L, Zheng H, Li-Ling J, Song K, Zhang Y. An ontology-independent representation learning for similar disease detection based on multi-layer similarity network. In: *IEEE/ACM transactions on computational biology and bioinformatics;* 2019.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

