Drug2Vec: Knowledge-aware Feature-driven Method for Drug Representation Learning

Ying Shen
ICNLAB, School of Electronics and Computer Engineering
Peking University (Shenzhen)
Shenzhen, China
shenying@pkusz.edu.cn

Kaiqi Yuan
ICNLAB, School of Electronics and Computer Engineering
Peking University (Shenzhen)
Shenzhen, China
kqyuan@pku.edu.cn

Yaliang Li
Medical Artificial Intelligence Lab
Tencent America
Palo Alto, USA
yaliangli@tencent.com

Buzhou Tang
School of Computer Science and Technology
Harbin Institute of Technology
Shenzhen, China
tangbuzhou@hit.edu.cn

Min Yang
SIAT
University of the Chinese Academy of Sciences
Shenzhen, China
min.yang@siat.ac.cn

Nan Du
Medical Artificial Intelligence Lab
Tencent America
Palo Alto, USA
ndu@tencent.com

Kai Lei*
ICNLAB, School of Electronics and Computer Engineering
Peking University (Shenzhen)
Shenzhen, China
leik@pkusz.edu.cn
(Corresponding Author)

***Abstract*— Proper representations of drugs have broad applications in healthcare analytics, such as drug-drug interaction (DDI) prediction and drug-drug similarity (DDS) computation. However, drug application involves accurate drug representation and rich annotated data, requiring tremendous expert time and effort. Thereby, drug feature sparseness creates a substantial barrier for drug representation learning, making it difficult to accurately identify new drug properties prior to public release. To alleviate these deficiencies, we propose a knowledge-aware feature-driven method (Drug2Vec) for exploring the interaction between two drugs. The method of Drug2Vec captures the medical information, taxonomy information and semantic information of drugs. The results of experiments demonstrate that compared with existing methods, Drug2Vec can effectively learn the drug representation and discover accurate drug-drug interaction.**

Keywords—drug representation learning, feature processing, drug-drug interaction.

I. INTRODUCTION

Efficient representations of high dimensional concepts has gained increasing attention in numerous natural language processing (NLP) tasks [1] across question answering, semantic textual similarity, and knowledge generative discovery. Many studies have demonstrated that it is possible to learn efficient representations of medical concept by improving the performance of medical predictive or classification models [2]. Despite this progress, learning efficient representations of drug concepts is still a relatively new territory and challenging task for three reasons: (i) Impressive drug representation learning was achieved in domains where a complete dictionary or a knowledge base is available [3]. However, the lack of clinical data and application data is almost inevitable for new drugs. (ii) Drug description information plays a crucial role in learning drug representation. Despite its usefulness, the application of drug description information in drug presentation is still under-explored [4]. (iii) The interactions between different drug features derived from various text and knowledge bases have received little attention in existing drug representation learning methods [5]. The assumption of feature independence may affect its representation learning.

To alleviate these challenges, we propose a knowledge-aware feature-driven method named Drug2Vec for drug representation learning and drug-drug interaction prediction. Specifically, we first employ knowledge embedding methods to pre-train the drug embeddings with respect to pharmacological features, drug class features and drug textual description features by taking advantage of various textual corpora and knowledge bases. Then, we design a joint learning method to learn the interaction information between different features as well as reduce the interference of noise information. Finally, we conduct experiments on a real-world dataset on the drug-drug interaction prediction.

The main contributions of this paper are summarized as follows: (1) We learn the drug knowledge by leveraging the pharmacological features, drug class features and drug textual description features within neural network architecture, addressing the incompleteness of drug attributes from single data source; (2) We propose a knowledge-aware interactive learning to exploit the interrelations among features and discover more informative information based on the relevancy of various drug features; (3) The experimental results show that Drug2Vec consistently outperforms the state-of-the-art methods.

978-1-5386-5488-0/18/$31.00 ©2018 IEEE

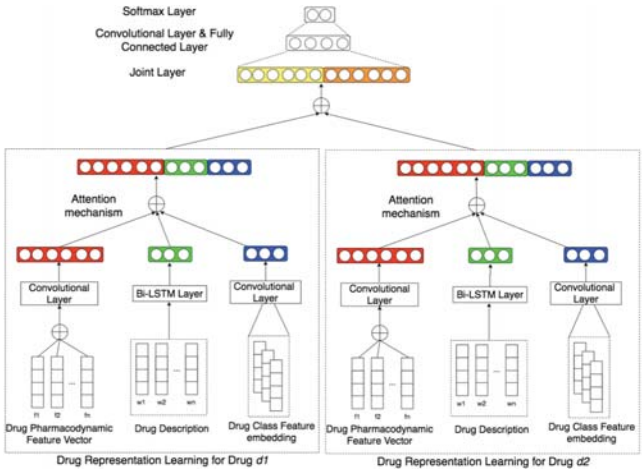

Fig. 1. Drug2Vec framework for drug-drug interaction prediction.

II. METHODOLOGY

Given a drug, we first learn the initial drug representation by processing the features with respect to pharmacology, drug catalog, and drug description information. Then we propose a joint learning method within deep neural network to learn interrelations among features. After learning the drug embeddings, we perform binary classification (positive correlation and negative correlation) between a pairwise drug. Fig. 1 illustrates the overall architecture of Drug2Vec.

A. Representation Learning of Pharmacological Feature

We consider the following Pharmacological features simultaneously:

Side effect: Given a drug d, its side effect embedding $Sider(d)$ can be obtained by learning the side effect resource SIDER using the IDF weighting method. The value of element s of $Sider(d)$, denoted $Sider(d)[s]$, is IDF(s, Drugs) if it is one of the side effects of drug d, otherwise it is 0. $IDF(s, Drugs)$ can be calculated as:

$$IDF(s, Drugs) = log\ ((|Drugs| + 1)/ (DF(s, Drugs) + 1)\ ), \quad (1)$$

where $Drugs$ is the set of drugs, s stands for a side effect, $DF(s, Drugs)$ is the number of drugs with side effect s.

Pharmaceutical composition, Physiological effects, Drug action and Drug targets: *Pharmaceutical composition* is a mixture that has been designed as a useful product. Many products are complex mixtures in which each chemical has a particular purpose. *Physiological effects* of drug vary by the type of drug, resulting in some imbalance to the overall human system, or some specific part of it. *The action of drugs* is that the drugs stimulate certain receptors or ion channels, act on enzymes or transporter proteins, causing the human body to react in a specific way. A *drug target* is a molecule in the body, usually a protein, that is intrinsically associated with a particular disease process and that could be addressed by a drug to produce a desired therapeutic effect. We learn the vectors of the aforementioned drug features by the same IDF-weighted mechanism as mentioned in the previous paragraph.

Drug chemical structure: The chemical structure of a drug determines its physicochemical properties, and ultimately affects its pharmacological activity. We adopt MACCS Key 166 [1] that can generate molecular fingerprints to learn the embeddings of drug chemical structure. For example, we has MACCS fingerprint "0101" for a testing compound (only for illustration), the likelihood of MACCS key can be calculated as:

$$\Pr(Fp = 0101|Cl) = \Pr(Fp_1 = 0|Cl)\times \Pr(Fp_2 = 1|Cl)\times\Pr(Fp_3 = 0|Cl)\times\Pr(Fp_4 = 1|Cl). \quad (2)$$

The presence or absence of predefined MACCS features are thus considered in the drug encoding. To reduce the vector dimension, we first concatenate the feature embeddings of drug d_i as:

$$d_i = [f_i^s : f_i^a : f_i^{pf} : f_i^{ps} : f_i^t : f_i^{cs}], \quad (3)$$

where [:] is the concatenation operation, f_i^s is its side effect embedding, f_i^a is the drug action embedding, f_i^{pf} is the pharmaceutical composition embedding, f_i^{pe} is the relative root feature embedding, f_i^t is the drug target embedding, and f_i^{cs} is the chemical structure embedding.

Afterwards, we input the feature embeddings of drug d_i to a Convolutional Neural Network (CNN). The fully connected layer of CNN model reduces the dimension of feature vectors from over 6000 dimensions to 500 dimensions, so as to improve the calculation and visualization of embeddings.

B. Representation Learning of Drug Class Feature

A drug class is a set of medications that have similar chemical structures, the same mechanism of action, a related mode of action, and/or are used to treat the same disease. Given the drug class taxonomy referred from dictionary ChemOnt [6], a CNN is designed to learn the drug class representation from drug taxonomy.

Input representation: DeepWalk, node2vec, and LINE are used to learn drug chemical taxonomy separately. Vectors learned by these network embedding methods are input to different channels in the convolution layer of the CNN, so as to make full use of all learnt taxonomy information.

Convolution layer: In the convolutional layer, we first perform convolution operation over a sliding window then max-pooling to learn the stack vector of drug class embeddings $\mathbf{D_n}$, where $\mathbf{D_n} \in \mathbb{R}^n$.

Fully connected layer: The vector obtained in the max pooling layer is fed to the fully connected softmax layer. In this study, the outputs of the fully connected layer is the embeddings of drug class feature.

C. Representation Learning of Textual Description Feature

In this study, we incorporate dependency information into deep neural networks to extract entities and the relations between entities from drug textual description for the representation learning. For two medical entities and a set of sentences containing both of them, the probability of the relation between them is measured.

[1] http://www.mayachemtools.org/docs/scripts/html/MACCSKeysFingerprints.html

1) Input representation.

Word embeddings: Given a sentence, its words are represented by real-valued vectors by looking up the pre-trained word embeddings.

Dependency embeddings: Dependency information can shorten the semantic distance between entities by organizing the whole sentence into a dependency tree [7]. We use the Stanford dependency parser [2] to extract the relative dependency features and dependency tags from the hierarchical structure of the dependency tree.

Dependency features also need to be transformed into vectors to jointly use with word embeddings. Then the word embeddings v_i^w and feature embeddings v_i^d are concatenated to represent each word:

$$word_i = [v_i^w, v_i^d]. \quad (4)$$

2) Bi-LSTM and Attention Mechanisms

In this paper, we employ Bi-LSTM model to capture the sequence information from both past and future contexts. We input the word vector sequence $W_l = [word_1, word_2, \dots word_j]$ and the inverse word vector sequence $W_{l_{reverse}} = [word_j, word_{j-1}, \dots word_1]$ into the forward layer and backward layer of Bi-LSTM respectively. The output hw_t at time step t, which combines the output of forward layer hf_t and backward layer hb_t, can be calculated as:

$$hw_t = hf_t + hb_t. \quad (5)$$

Sentence-level and word-level attention are adopted in our model to pay due attention to the useful information. Take sentence-level attention a_i as an example. It is given by:

$$a_i = \frac{\exp(e_i)}{\sum_i \exp(e_j)}, \quad (6)$$

where e_i scores the relativity between the sentence and the predicted relation. Given a drug and its description, the outputs of the Bi-LSTM is the embeddings of drug textual description.

D. Joint Learning Method

Using attention mechanism, the representation of drug class feature v_c are calculated as:

$$M_w = \tanh(W_{sw} H_w), \quad (7)$$

$$\alpha_w = \mathrm{softmax}(w_w^T M_w), \quad (8)$$

$$v_c = H_w \alpha_w^T, \quad (9)$$

where $M_w \in R^{dlxm}$ is a nonlinear mapping function, $W_{sw} \in R^{dlxdl}$ and $w_w \in R^{dl}$ are projection parameters, $\alpha_w \in R^m$ is the normalized attention. Other two types of features are processed by the same attention mechanism. Then these three type of feature embeddings are concatenated for the final knowledge-aware drug representations.

For the DDI prediction, there is a joint layer to join the final drug representations of drug 1 and drug 2. The outputs of the convolutional layer and fully connected layer then go through a softmax layer for binary classification:

$$y = softmax(W_o pr + b_o), \quad (10)$$

where each dimension of y denotes the normalized probability of a certain relation, i.e., positive correlation or negative correlation, in accordance with the fully connected layer, $W_o \in R^{2xdl}$ is the projection matrix, and $b_o \in R^2$ is the offset vector.

III. Experiments

A. Datasets

1) Datasets.

Knowledge bases. Drug side effect is collected from Side effect resource (SIDER[3]), drug action is learnt from and NDF-RT, and drug class features is extracted from ChemOnt[4]. The knowledge about the pharmaceutical formulation, physiological effects, drug targets and drug chemical structure are learnt from DrugBank[5].

Text corpus. Given a drug, its textual description can be obtained from the "title" and "abstract" section of Pubmed[6], the "drug label information" section of DailyMed[7], and the "description", "indication", "metabolism", "toxicity" and "pharmacodynamics" section of DrugBank[8].

2) Implementation Details

The kernel and the depth of the CNN model are set to 5 and 20 respectively. A Fully connected layer whose size is 500 is added after the CNN layer. In the Bi-LSTM, we employ dropout on the output layer to prevent overfitting. We use ReLU activation function, take cross-entropy as loss function, and adopt AdaGrad as optimizer. For both CNN and Bi-LSTM model, the learning rate and the dropout rate are set to 0.003 and 0.5 respectively. We train our models in batches with a size of 40. All other parameters are randomly initialized from [-0.1,0.1]. The maximum length of sentence is set to 100. For the base models, we follow exactly the same parameter settings as those in their original papers.

B. Evaluation Tasks

1) Drug-Drug Interaction (DDI) Classification

We conduct a retrospective evaluation using the known DDIs and drug features in an earlier version of DrugBank (V4.5) to predict the DDI among newly developed drugs that presented in a more updated version of DrugBank (V5.0.7).

The experimental results are summarized in Table 1. Six state-of-the-art baselines are adopted for comparison: (1) **SVM** [9], support vector machines (SVM) with manually defined features; (2) **FBK-irst** [10], a multi-phase kernel based approach for DDI that exploits linguistic information; (3) **CNN** [11], a CNN model for DDI task consists of four layers: a look-up table layer, a convolutional layer, a max pooling layer, and a softmax layer; (4) **Att-BLSTM** [12], an attention-

[2] https://nlp.stanford.edu/software/lex-parser.shtml

[3] http://sideeffects.embl.de
[4] http://classyfire.wishartlab.com/
[5] https://www.drugbank.ca/
[6] https://www.ncbi.nlm.nih.gov/pubmed/
[7] https://dailymed.nlm.nih.gov/
[8] https://www.drugbank.ca/drugs/DB00316

TABLE I. DDI RETROSPECTIVE EVALUATION: TRAINING IN AN EARLIER VERSION OF DRUGBANK AND TESTING IN A MORE UPDATED VERSION OF DRUGBANK. DRUG2VEC CORRECTLY PREDICTS UP TO 92.19% OF THE DDIS FOUND AFTER 2016.

	Accuracy	Precision	Recall	F-score	AUROC	AUPR
FBK-irst (Chowdhury et al. 2013)	0.6533	0.6437	0.6867	0.6645	0.6807	0.7479
SVM (Björne et al. 2013)	0.7867	0.7622	0.8333	0.7962	0.8844	0.8694
CNN (Liu et al. 2016)	0.81	0.8039	0.82	0.8118	0.8892	0.8897
Att-BLSTM (Zheng et al. 2017)	0.7750	0.7749	0.7750	0.7750	0.8455	0.8486
Tiresias (Abdelaziz et al. 2017)	0.80	0.7885	0.82	0.8039	0.8869	0.8861
LP-AllSim (Zhang et al. 2015)	0.77	0.7547	0.8	0.7767	0.8544	0.8600
Drug2Vec (our model)	**0.9219**	**0.9191**	**0.9191**	**0.9191**	**0.9512**	**0.9568**
w/o pharmacology	0.8571	0.8570	0.8571	0.8571	0.8571	0.8571
w/o drug class	0.8854	0.8854	0.8855	0.8854	0.8854	0.9391
w/o textual description	0.9033	0.9032	0.9033	0.9033	0.9373	0.9432

based neural network model that uses RNN with LSTM units; (5) **Tiresias** [13], a similarity-based model that uses local and global features in a knowledge graph to predicts DDIs through link prediction; (6) **LP-AllSim** [14]: an label propagation model to predict DDIs based on clinical side effect.

To analyze the effectiveness of our model, we also report the ablation test in terms of discarding the pharmacological feature (w/o pharmacology), drug class feature (w/o drug class) and drug textual description feature (w/o textual description), respectively. There are multiple interesting observations from Table 1 as followings: (1) Compared with baseline systems, Drug2Vec boosts the DDIs prediction performance. It outperforms the current best system (CNN [11]) by 10% in F-score. (2) Top performing systems in Table 1 (e.g., SVM [9], Tiresias [13],) are performed based on various features such as features derived from medical resources or manually defined features. Compared with these top performing systems, the features used in the Drug2Vec are automatically learned during training. Moreover, avoiding errors caused by NLP toolkits. (3) We can observe from the ablation test that, all adopted features contribute, and it makes larger performance boosting to DDI prediction, demonstrating the necessity of simultaneous consideration of the adopted features.

IV. DISCUSSION AND CONCLUSION

This paper proposed a novel representation method termed as drug2vec of drugs from three different sources of information including pharmacological information, drug taxonomy information, and textual descriptions. The performance of drug2vec representation was tested on DDI prediction, and it demonstrated a superior performance compared with the existing methods.

ACKNOWLEDGEMENT

This work was supported by the National Natural Science Foundation of China (No.61602013), the Shenzhen Fundamental Research Projects (JCYJ20170818091546869), and Guangdong Science Fund (2018A030313017).

REFERENCES

[1] Y. Shen, Y. Deng, M. Yang, Y. Li, N. Du, W. Fan et al., "Knowledge-aware attentive neural network for ranking question answer pairs," in *Proceedings of the 41st International SIGIR Conference, Ann Arbor, MI, USA, July 08 - 12, 2018.* ACM: New York, 2018. pp.901-904.

[2] J. Minarro-Giménez, O, Marin-Alonso, M, Samwald. "Exploring the application of deep learning techniques on medical text corpora," Studies in health technology and informatics, 2014. pp.584-588.

[3] E. Choi, M. Bahadori, E. Searles, C. Coffey, and M. Thompson, "Multi-layer representation learning for medical concepts," in *Proceedings of the 22nd ACM SIGKDD*, San Francisco, 13-17 August, 2016. ACM: New York, 2016. pp. 1495-1504.

[4] S. Korkmaz, G. Zararsiz, and D. Goksuluk, "Drug/nondrug classification using support vector machines with various feature selection strategies", Computer methods and programs in biomedicine, vol.117, pp.51-60, 2014.

[5] E. Choi, M. Bahadori, L. Song, W. Stewart, and J. Sun, "GRAM: graph-based attention model for healthcare representation learning," in *Proceedings of the 23rd ACM SIGKDD, Halifax, NS, Canada, August 13 - 17, 2017.* ACM: New York, 2017. pp.787-795.

[6] Y. Djoumbou Feunang, R. Eisner, C. Knox, L. Chepelev, J. Hastings, G. Owen, et al., "ClassyFire: automated chemical classification with a comprehensive, computable taxonomy", Journal of Cheminformatics, vol. 8, no. 1, 2016.

[7] K. Tai, R. Socher, C. Manning. "Improved semantic representations from tree-structured long short-term memory networks," in *Proceedings of ACL, Beijing, China, July 26-31 2015.* 2015.

[8] Y. Pu, Z. Gan, R. Henao X. Yuan, C. Li, A. Stevens et al., "Variational autoencoder for deep learning of images, labels and captions," in *Proceedings of NIPS, Barcelona, Spain, Dec 4 - 9, 2016.* pp.2352-2360.

[9] J. Björne, S. Kaewphan, and T. Salakoski, "UTurku: drug named entity recognition and drug-drug interaction extraction using SVM classification and domain knowledge," in *Proceedings of SemEval 2013, Atlanta, Georgia, June 14-15, 2013.* pp.651-659.

[10] M. Chowdhury and A. Lavelli, "FBK-irst: a multi-phase kernel based approach for drug-drug interaction detection and classification that exploits linguistic information," in *Proceedings of SemEval 2013, Atlanta, Georgia, June 14-15, 2013.* pp.351-355.

[11] S. Liu, B. Tang, Q. Chen and X. Wang, "Drug-Drug Interaction Extraction via Convolutional Neural Networks", Computational and Mathematical Methods in Medicine, vol. 2016, pp. 1-8, 2016.

[12] W. Zheng, H. Lin, L. Luo, Z. Zhao, Z. Li, Y. Zhang, Z. Yang and J. Wang, "An attention-based effective neural model for drug-drug interactions extraction", BMC Bioinformatics, vol. 18, no. 1, 2017.

[13] I. Abdelaziz, A. Fokoue, O. Hassanzadeh, P. Zhang, M. Sadoghi, "Large-scale structural and textual similarity-based mining of knowledge graph to predict drug–drug interactions," in *Proceedings of World Wide Web, Perth, Australia, Apr 3 - 7, 2017*. pp.104-117.

[14] P. Zhang, F. Wang, J. Hu and R. Sorrentino, "Label Propagation Prediction of Drug-Drug Interactions Based on Clinical Side Effects", Scientific Reports, vol. 5, no. 1, 2015.

[15] D. Zeng, K. Liu, S. Lai, G. Zhou, and J. Zhao, "Relation classification via convolutional deep neural network," in *Proceedings of COLING 2014, Dublin, Ireland, 23rd August 2014.* pp.2335-2344.

[16] D. Zeng, K. Liu, Y. Chen, and J. Zhao, "Distant supervision for relation extraction via piecewise convolutional neural networks," In *Proceedings of the 2015 EMNLP, Lisbon, Portugal, September 17–21,* 2015. pp.1753-1762.