

Enhancing Drug-Drug Interaction Extraction from Texts by Molecular Structure Information

Masaki Asada, Makoto Miwa and Yutaka Sasaki

Computational Intelligence Laboratory

Toyota Technological Institute

2-12-1 Hisakata, Tempaku-ku, Nagoya, Aichi, 468-8511, Japan

{sd17402, makoto-miwa, yutaka.sasaki}@toyota-ti.ac.jp

Abstract

We propose a novel neural method to extract drug-drug interactions (DDIs) from texts using external drug molecular structure information. We encode textual drug pairs with convolutional neural networks and their molecular pairs with graph convolutional networks (GCNs), and then we concatenate the outputs of these two networks. In the experiments, we show that GCNs can predict DDIs from the molecular structures of drugs in high accuracy and the molecular information can enhance text-based DDI extraction by 2.39 percent points in the F-score on the DDIExtraction 2013 shared task data set.

1 Introduction

When drugs are concomitantly administered to a patient, the effects of the drugs may be enhanced or weakened, which may also cause side effects. These kinds of interactions are called Drug-Drug Interactions (DDIs). Several drug databases have been maintained to summarize drug and DDI information such as DrugBank (Law et al., 2014), Therapeutic Target database (Yang et al., 2016), and PharmGKB (Thorn et al., 2013). Automatic DDI extraction from texts is expected to support the maintenance of databases with high coverage and quick update to help medical experts. Deep neural network-based methods have recently drawn a considerable attention (Liu et al., 2016; Sahu and Anand, 2017; Zheng et al., 2017; Lim et al., 2018) since they show state-of-the-art performance without manual feature engineering.

In parallel to the progress in DDI extraction from texts, Graph Convolutional Networks (GCNs) have been proposed and applied to estimate physical and chemical properties of molec-

ular graphs such as solubility and toxicity (Duvenaud et al., 2015; Li et al., 2016; Gilmer et al., 2017).

In this study, we propose a novel method to utilize both textual and molecular information for DDI extraction from texts. We illustrate the overview of the proposed model in Figure 1. We obtain the representations of drug pairs in molecular graph structures using GCNs and concatenate the representations with the representations of the textual mention pairs obtained by convolutional neural networks (CNNs). We trained the molecule-based model using interacting pairs mentioned in the DrugBank database and then trained the entire model using the labeled pairs in the text data set of the DDIExtraction 2013 shared task (SemEval-2013 Task 9) (Segura Bedmar et al., 2013). In the experiment, we show GCNs can predict DDIs from molecular graphs in a high accuracy. We also show molecular information can enhance the performance of DDI extraction from texts in 2.39 percent points in F-score.

The contribution of this paper is three-fold:

- We propose a novel neural method to extract DDIs from texts with the related molecular structure information.
- We apply GCNs to pairwise drug molecules for the first time and show GCNs can predict DDIs between drug molecular structures in a high accuracy.
- We show the molecular information is useful in extracting DDIs from texts.

2 Methods

2.1 Text-based DDI Extraction

Our model for extracting DDIs from texts is based on the CNN model by Zeng et al. (2014). When an input sentence $S = (w_1, w_2, \dots, w_N)$ is given, We prepare word embedding w_i^w of w_i and word

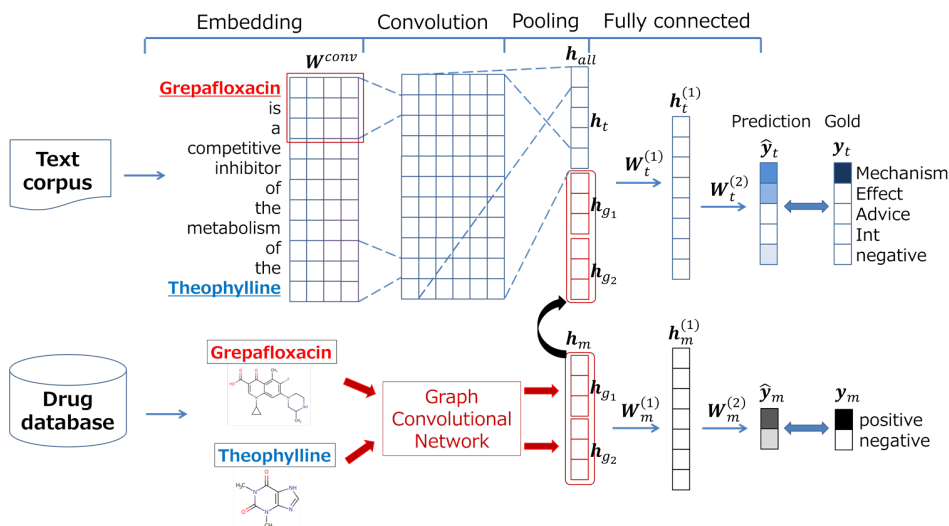


Figure 1: Overview of the proposed model

position embeddings $w_{i,1}^p$ and $w_{i,2}^p$ that correspond to the relative positions from the first and second target entities, respectively. We concatenate these embeddings as in Equation (1), and we use the resulting vector as the input to the subsequent convolution layer:

$$w_i = [w_i^w; w_{i,1}^p; w_{i,2}^p], \quad (1)$$

where $[\cdot]$ denotes the concatenation. We calculate the expression for each filter j with the window size k_l .

$$z_{i,l} = [w_{i-(k_l-1)/2}, \dots, w_{i-(k_l+1)/2}], \quad (2)$$

$$m_{i,j,l} = \text{relu}(\mathbf{W}_j^{conv} \odot z_{i,l} + b^{conv}), \quad (3)$$

$$m_{j,l} = \max_i m_{i,j,l}, \quad (4)$$

where L is the number of windows, \mathbf{W}_j^{conv} and b^{conv} are the weight and bias of CNN, and \max indicates max pooling (Boureau et al., 2010).

We convert the output of the convolution layer into a fixed-size vector that represents a textual pair as follows:

$$m_l = [m_{1,l}, \dots, m_{J,l}], \quad (5)$$

$$h_t = [m_1; \dots; m_L], \quad (6)$$

where J is the number of filters.

We get a prediction \hat{y}_t by the following fully connected neural networks:

$$h_t^{(1)} = \text{relu}(\mathbf{W}_t^{(1)} h_t + b_t^{(1)}), \quad (7)$$

$$\hat{y}_t = \text{softmax}(\mathbf{W}_t^{(2)} h_t^{(1)} + b_t^{(2)}), \quad (8)$$

where $\mathbf{W}_t^{(1)}$ and $\mathbf{W}_t^{(2)}$ are weights and $b_t^{(1)}$ and $b_t^{(2)}$ are bias terms.

2.2 Molecular Structure-based DDI Classification

We represent drug pairs in molecular graph structures using two GCN methods: CNNs for fingerprints (NFP) (Duvenaud et al., 2015) and Gated Graph Neural Networks (GGNN) (Li et al., 2016). They both convert a drug molecule graph G into a fixed size vector h_g by aggregating the representation h_v^T of an atom node v in G . We represent atoms as nodes and bonds as edges in the graph.

NFP first obtains the representation h_v^t by the following equations (Duvenaud et al., 2015).

$$m_v^{t+1} = h_v^t + \sum_{w \in N(v)} h_w^t, \quad (9)$$

$$h_v^{t+1} = \sigma(\mathbf{H}_t^{deg(v)} m_v^{t+1}), \quad (10)$$

where h_v^t is the representation of v in the t -th step, $N(v)$ is the neighbors of v , and $\mathbf{H}_t^{deg(v)}$ is a weight parameter. h_v^0 is initialized by the *atom features* of v . $deg(v)$ is the degree of a node v and σ is a sigmoid function. NFP then acquires the representation of the graph structure

$$h_g = \sum_{v,t} \text{softmax}(\mathbf{W}^t h_v^t), \quad (11)$$

where \mathbf{W}^t is a weight matrix.

GGNN first obtains the representation h_v^t by using Gated Recurrent Unit (GRU)-based recurrent neural networks (Li et al., 2016) as follows:

$$m_v^{t+1} = \sum_{w \in N(v)} \mathbf{A}_{e_{vw}} h_w^t \quad (12)$$

$$h_v^{t+1} = \text{GRU}([h_v^t; m_v^{t+1}]), \quad (13)$$

where $A_{e_{vw}}$ is a weight for the *bond type* of each edge e_{vw} . GGNN then acquires the representation of the graph structure.

$$\mathbf{h}_g = \sum_v \sigma(i([\mathbf{h}_v^T; \mathbf{h}_v^0])) \odot (j(\mathbf{h}_v^T)), \quad (14)$$

where i and j are linear layers and \odot is the element-wise product.

We obtain the representation of a molecular pair by concatenating the molecular graph representations of drugs g_1 and g_2 , i.e., $\mathbf{h}_m = [\mathbf{h}_{g_1}; \mathbf{h}_{g_2}]$.

We get a prediction $\hat{\mathbf{y}}_m$ as follows:

$$\mathbf{h}_m^{(1)} = \text{relu}(\mathbf{W}_m^{(1)}\mathbf{h}_m + \mathbf{b}_m^{(1)}), \quad (15)$$

$$\hat{\mathbf{y}}_m = \text{softmax}(\mathbf{W}_m^{(2)}\mathbf{h}_m^{(1)} + \mathbf{b}_m^{(2)}), \quad (16)$$

where $\mathbf{W}_m^{(1)}$ and $\mathbf{W}_m^{(2)}$ are weights and $\mathbf{b}_m^{(1)}$ and $\mathbf{b}_m^{(2)}$ are bias terms.

2.3 DDI Extraction from Texts Using Molecular Structures

We realize the simultaneous use of textual and molecular information by concatenating a text-based and molecule-based vectors: $\mathbf{h}_{all} = [\mathbf{h}_t; \mathbf{h}_m]$. We normalize molecule-based vectors. We then use \mathbf{h}_{all} instead of \mathbf{h}_t in Equation 7.

In training, we first train the molecular-based DDI classification model. The molecular-based classification is performed by minimizing the loss function $L_m = -\sum \mathbf{y}_m \log \hat{\mathbf{y}}_m$. We then fix the parameters for GCNs and train text-based DDI extraction model by minimizing the loss function $L_t = -\sum \mathbf{y}_t \log \hat{\mathbf{y}}_t$.

3 Experimental Settings

In this section, we explain the textual and molecular data and task settings and training settings.

3.1 Text Corpus and Task Setting

We followed the task setting of Task 9.2 in the DDIExtraction 2013 shared task (Segura Bedmar et al., 2013; Herrero-Zazo et al., 2013) for the evaluation. This data set is composed of documents annotated with drug mentions and their four types of interactions: *Mechanism*, *Effect*, *Advice* and *Int*. For the data statistics, please refer to the supplementary materials.

The task is a multi-class classification task, i.e., to classify a given pair of drugs into the four interaction types or no interaction. We evaluated the performance with micro-averaged precision (P),

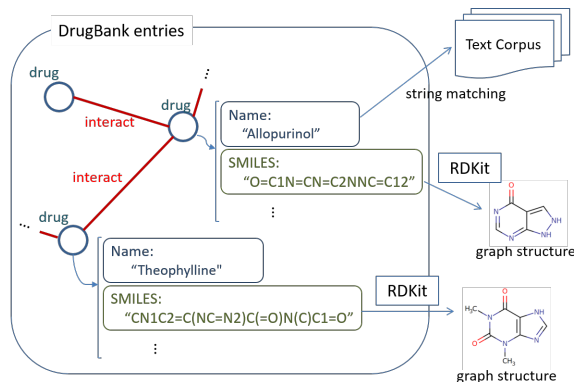


Figure 2: Associating DrugBank entries with texts and molecular structures

recall (R), and F-score (F) on all the interaction types. We used the official evaluation script provided by the task organizers.

As preprocessing, we split sentences into words using the GENIA tagger (Tsuruoka et al., 2005). We replaced the drug mentions of the target pair with *DRUG1* and *DRUG2* according to their order of appearance. We also replaced other drug mentions with *DRUGOTHER*. We did not employ negative instance filtering unlike other existing methods, e.g., Liu et al. (2016), since our focus is to evaluate the effect of the molecular information on texts.

We linked mentions in texts to DrugBank entries by string matching. We lowercased the mentions and the names in the entries and chose the entries with the most overlaps. As a result, 92.15% and 93.09% of drug mentions in train and test data set matched the DrugBank entries.

3.2 Data and Task for Molecular Structures

We extracted 255,229 interacting (positive) pairs from DrugBank. We note that, unlike text-based interactions, DrugBank only contains the information of interacting pairs; there are no detailed labels and no information for non-interacting (negative) pairs. We thus generated the same number of pseudo negative pairs by randomly pairing drugs and removing those in positive pairs. To avoid overestimation of the performance, we also deleted drug pairs mentioned in the test set of the text corpus. We split positive and negative pairs into 4:1 for training and test data, and we evaluated the classification accuracy using only the molecular information.

To obtain the graph of a drug molecule, we took

Methods	P	R	F (%)
Liu et al. (2016)	75.29	60.37	67.01
Zheng et al. (2017)	75.9	68.7	71.5
Lim et al. (2018)	74.4	69.3	71.7
Text-only	71.97	68.44	70.16
+ NFP	72.62	71.81	72.21
+ GGNN	73.31	71.81	72.55

Table 1: Evaluation on DDI extraction from texts

as input the SMILES (Weininger, 1988) string encoding of the molecule from DrugBank and then converted it into the graph using RDKit (Landrum, 2016) as illustrated in Figure 2. For the *atom features*, we used randomly embedded vectors for each atoms (i.e., C, O, N, ...). We also used 4 *bond types*: single, double, triple, or aromatic.

3.3 Training Settings

We employed mini-batch training using the Adam optimizer (Kingma and Ba, 2015). We used L2 regularization to avoid over-fitting. We tuned the bias term $b_t^{(2)}$ for negative examples in the final softmax layer. For the hyper-parameters, please refer to the supplementary materials.

We employed pre-trained word embeddings trained by using the word2vec tool (Mikolov et al., 2013) on the 2014 MEDLINE/PubMed baseline distribution. The vocabulary size was 215,840. The embedding of the drugs, i.e., *DRUG1* and *DRUG2* were initialized with the pre-trained embedding of the word *drug*. The embeddings of training words that did not appear in the pre-trained embeddings were initialized with the average of all pre-trained word embeddings. Words that appeared only once in the training data were replaced with an *UNK* word during training, and the embedding of words in the test data set that did not appear in both training and pre-trained embeddings were set to the embedding of the *UNK* word. Word position embeddings are initialized with random values drawn from a uniform distribution.

We set the molecule-based vectors of unmatched entities to zero vectors.

4 Results

Table 1 shows the performance of DDI extraction models. We show the performance without negative instance filtering or ensemble for the fair comparison. We observe the increase of recall and F-score by using molecular information,

DDI Type	Mech.	Effect	Adv.	Int (%)
Text-only	69.52	69.27	79.81	48.18
+ NFP	72.70	72.44	79.56	46.98
+ GGNN	73.83	71.03	81.62	45.83

Table 2: Performance on individual DDI types in F-scores

Methods	Accuracy (%)
NFP	94.19
GGNN	98.00

Table 3: Accuracy of binary classification on DrugBank pairs

Methods	P	R	F	Acc. (%)
NFP	15.56	48.93	23.61	45.78
GGNN	15.11	57.10	23.90	37.72

Table 4: Classification of DDIs in texts by molecular structure-based DDI classification model

which results in the state-of-the-art performance with GGNN.

Both GCNs improvements were statistically significant ($p < 0.05$ for NFP and $p < 0.005$ for GGNN) with randomized shuffled test.

Table 2 shows F-scores on individual DDI types. The molecular information improves F-scores especially on type *Mechanism* and *Effect*.

We also evaluated the accuracy of binary classification on DrugBank pairs by using only the molecular information in Table 3. The performance is high, although the accuracy is evaluated on automatically generated negative instances.

Finally, we applied the molecular-based DDI classification model trained on DrugBank to the DDIExtraction 2013 task data set. Since the DrugBank has no detailed labels, we mapped all four types of interactions to positive interactions and evaluated the classification performance. The results in Table 4 show that GCNs produce higher recall than precision and the overall performance is low considering the high performance on DrugBank pairs. This might be because the interactions of drugs are not always mentioned in texts even if the drugs can interact with each other and because hedged DDI mentions are annotated as DDIs in the text data set. We also trained the DDI extraction model only with molecular information by replacing h_{all} with h_m , but the F-scores were quite low ($< 5\%$). These results show that we cannot predict textual relations only with molecular information.

5 Related Work

Various feature-based methods have been proposed during and after the DDIExtraction-2013 shared task (Segura Bedmar et al., 2013). Kim et al. (2015) proposed a two-phase SVM-based approach that employed a linear SVM with rich features that consist of word, word pair, dependency graph, parse tree, and noun phrase-based constrained coordination features. Zheng et al. (2016) proposed a context vector graph kernel to exploit various types of contexts. Raihani and Laachfoubi (2017) also employed a two-phase SVM-based approach using non-linear kernels and they proposed five groups of features: word, drug, pair of drug, main verb and negative sentence features. Our model does not use any features or kernels.

Various neural DDI extraction models have been recently proposed using CNNs and Recurrent Neural Networks (RNNs). Liu et al. (2016) built a CNN-based model based on word and position embeddings. Zheng et al. (2017) proposed a Bidirectional Long Short-Term Memory RNN (Bi-LSTM)-based model with an input attention mechanism, which obtained target drug-specific word representations before the Bi-LSTM. Lim et al. (2018) proposed Recursive neural network-based model with a subtree containment feature and an ensemble method. This model showed the state-of-the-art performance on the DDIExtraction 2013 shared task data set if systems do not use negative instance filtering. These approaches did not consider molecular information, and they can also be enhanced by the molecular information.

Vilar et al. (2017) focused on detecting DDIs from different sources such as pharmacovigilance sources, scientific biomedical literature and social media. They did not use deep neural networks and they did not consider molecular information.

Learning representations of graphs are widely studied in several tasks such as knowledge base completion, drug discovery, and material science (Wang et al., 2017; Gilmer et al., 2017). Several graph convolutional neural networks have been proposed such as NFP (Duvenaud et al., 2015), GGNN (Li et al., 2016), and Molecular Graph Convolutions (Kearnes et al., 2016), but they have not been applied to DDI extraction.

6 Conclusions

We proposed a novel neural method for DDI extraction using both textual and molecular informa-

tion. The results show that DDIs can be predicted with high accuracy from molecular structure information and that the molecular information can improve DDI extraction from texts by 2.39 percent points in F-score on the data set of the DDIExtraction 2013 shared task.

As future work, we would like to seek the way to model the textual and molecular representations jointly with alleviating the differences in labels. We will also investigate the use of other information in DrugBank.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number 17K12741.

References

- Y-Lan Boureau, Jean Ponce, and Yann LeCun. 2010. A theoretical analysis of feature pooling in visual recognition. In *Proceedings of ICML*. pages 111–118.
- David K Duvenaud, Dougal Maclaurin, Jorge Iparraguirre, Rafael Bombarell, Timothy Hirzel, Alan Aspuru-Guzik, and Ryan P Adams. 2015. Convolutional networks on graphs for learning molecular fingerprints. In *Proceedings of NIPS*, pages 2224–2232.
- Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. 2017. Neural message passing for quantum chemistry. In *Proceedings of ICML*. pages 1263–1272.
- María Herrero-Zazo, Isabel Segura-Bedmar, Paloma Martínez, and Thierry Declerck. 2013. The DDI corpus: An annotated corpus with pharmacological substances and drug–drug interactions. *Journal of biomedical informatics* 46(5):914–920.
- Steven Kearnes, Kevin McCloskey, Marc Berndl, Vijay Pande, and Patrick Riley. 2016. Molecular graph convolutions: moving beyond fingerprints. *Journal of computer-aided molecular design* 30(8):595–608.
- Sun Kim, Haibin Liu, Lana Yeganova, and W John Wilbur. 2015. Extracting drug–drug interactions from literature using a rich feature-based linear kernel approach. *Journal of biomedical informatics* 55:23–30.
- Diederik Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *Proceedings of ICLR*.
- Greg Landrum. 2016. Rdkit: Open-source cheminformatics.

- Vivian Law, Craig Knox, Yannick Djoumbou, Tim Jewison, An Chi Guo, Yifeng Liu, Adam Maciejewski, David Arndt, Michael Wilson, Vanessa Neveu, et al. 2014. DrugBank 4.0: shedding new light on drug metabolism. *Nucleic acids research* 42(D1):D1091–D1097.
- Yujia Li, Daniel Tarlow, Marc Brockschmidt, and Richard Zemel. 2016. Gated graph sequence neural networks. In *Proceedings of ICLR*.
- Sangrak Lim, Kyubum Lee, and Jaewoo Kang. 2018. Drug drug interaction extraction from the literature using a recursive neural network. *PloS one* 13(1):e0190926.
- Shengyu Liu, Buzhou Tang, Qingcai Chen, and Xiaolong Wang. 2016. Drug-drug interaction extraction via convolutional neural networks. *Computational and mathematical methods in medicine*.
- Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Proceedings of NIPS*. pages 3111–3119.
- Anass Raihani and Nabil Laachfoubi. 2017. A rich feature-based kernel approach for drug-drug interaction extraction. *International Journal of Advanced Computer Science and Applications* 8(4):324–330.
- Sunil Kumar Sahu and Ashish Anand. 2017. Drug-drug interaction extraction from biomedical text using long short term memory network. *arXiv preprint arXiv:1701.08303*.
- Isabel Segura Bedmar, Paloma Martínez, and María Herrero Zazo. 2013. SemEval-2013 Task 9: Extraction of drug-drug interactions from biomedical texts (DDIExtraction 2013). In *Proceedings of the 7th International Workshop on Semantic Evaluations*. pages 341–350.
- Caroline F Thorn, Teri E Klein, and Russ B Altman. 2013. PharmGKB: the pharmacogenomics knowledge base. *Pharmacogenomics: Methods and Protocols* pages 311–320.
- Yoshimasa Tsuruoka, Yuka Tateishi, Jin-Dong Kim, Tomoko Ohta, John McNaught, Sophia Ananiadou, and Junichi Tsujii. 2005. Developing a robust part-of-speech tagger for biomedical text. In *Panhellenic Conference on Informatics*. pages 382–392.
- Santiago Vilar, Carol Friedman, and George Hripcsak. 2017. Detection of drug–drug interactions through data mining studies using clinical sources, scientific literature and social media. *Briefings in bioinformatics*.
- Quan Wang, Zhendong Mao, Bin Wang, and Li Guo. 2017. Knowledge graph embedding: A survey of approaches and applications. *IEEE Transactions on Knowledge and Data Engineering* 29(12):2724–2743.
- David Weininger. 1988. Smiles, a chemical language and information system. 1. introduction to methodology and encoding rules. *Journal of chemical information and computer sciences* 28(1):31–36.
- Hong Yang, Chu Qin, Ying Hong Li, Lin Tao, Jin Zhou, Chun Yan Yu, Feng Xu, Zhe Chen, Feng Zhu, and Yu Zong Chen. 2016. Therapeutic target database update 2016: enriched resource for bench to clinical drug target and targeted pathway information. *Nucleic acids research* 44(D1):D1069–D1074.
- Daojian Zeng, Kang Liu, Siwei Lai, Guangyou Zhou, Jun Zhao, et al. 2014. Relation classification via convolutional deep neural network. In *COLING*. pages 2335–2344.
- Wei Zheng, Hongfei Lin, Ling Luo, Zhehuan Zhao, Zhenguang Li, Yijia Zhang, Zhihao Yang, and Jian Wang. 2017. An attention-based effective neural model for drug-drug interactions extraction. *BMC Bioinformatics* 18(1):445.
- Wei Zheng, Hongfei Lin, Zhehuan Zhao, Bo Xu, Yijia Zhang, Zhihao Yang, and Jian Wang. 2016. A graph kernel based on context vectors for extracting drug–drug interactions. *Journal of biomedical informatics* 61:34–43.