BERTChem-DDI : Improved Drug-Drug Interaction Prediction from text using Chemical Structure Information

Ishani Mondal
Microsoft Research Lab
Lavelle Road, Bengaluru, India
t-imonda@microsoft.com/ishani340@gmail.com

Abstract

Traditional biomedical version of embeddings obtained from pre-trained language models have recently shown state-of-the-art results for relation extraction (RE) tasks in the medical domain. In this paper, we explore how to incorporate domain knowledge, available in the form of molecular structure of drugs, for predicting Drug-Drug Interaction from textual corpus. We propose a method, ***BERTChem-DDI***, to efficiently combine drug embeddings obtained from the rich chemical structure of drugs (encoded in SMILES) along with off-the-shelf domain-specific BioBERT embedding-based RE architecture. Experiments conducted on the DDIExtraction 2013 corpus clearly indicate that this strategy improves other strong baselines architectures by 3.4% macro F1-score.

1 Introduction

Concurrent administration of two or more drugs to a patient to cure an ailment might lead to positive or negative reaction (side-effect). These kinds of interactions are termed as Drug-Drug Interactions (DDIs). Predicting drug-drug interactions (DDI) is a complex task as it requires to understand the mechanism of action of two interacting drugs. A large number of efforts by the researchers have been witnessed in terms of automatic extraction of DDIs from the textual corpus (Sahu and Anand, 2018), (Liu et al., 2016), (Sun et al., 2019), (Li and Ji, 2019) and predicting unknown DDI from the Knowledge Graph (Purkayastha et al., 2019), (Karim et al., 2019). Automatic extraction of DDI from texts aids in maintaining the databases with high coverage and help the medical experts in their diagnosis and novel experiments.

In parallel to the progress of DDI extraction from the textual corpus, some efforts have been observed recently where the researchers came up with various strategies of augmenting chemical structure information of the drugs (Asada et al., 2017) and textual description of the drugs (Zhu et al., 2020a) to improve Drug-Drug Interaction prediction performance from corpus and Knowledge Graphs.

The DDI Prediction from the textual corpus has been framed by the earlier researchers as relation classification problem. Earlier methods (Sahu and Anand, 2018), (Liu et al., 2016), (Sun et al., 2019), (Li and Ji, 2019) for relation classification are based on CNN or RNN based Neural Networks.

Recently, with the massive success of the pre-trained language models (Devlin et al., 2019), (Yang et al., 2019) in many NLP classification / sequence labeling tasks, we formulate the problem of DDI classification as a relation classification task by leveraging both the entities and sentence-level information. We propose a model that leverages both domain-specific contextual embeddings (BioBERT) (Lee et al., 2019) from the target entities and also external Chemical Structure information of the target entities (drugs). In the recent years, representation learning has played a pivotal role in solving various machine learning tasks. In addition to information of drug entities from the text, we make use of the rich hidden representation obtained from the molecule generation using Variational Auto-Encoder (Gómez-Bombarelli et al., 2018) representation of the drugs to learn the chemical structure representation. During unsupervised learning of chemical structure information of the drugs using Variational AutoEncoder (Kingma and Welling, 2014), we make use of the canonical SMILES representation (**S**implified **M**olecular **I**nput **L**ine **E**ntry **S**ystem) obtained from the DrugBank (Wishart et al., 2008). We illustrate the overview of the proposed method in Figure 1. Experiments conducted on the DDIExtraction 2013 corpus (Herrero-Zazo et al., 2013) reveals that this method outperforms the existing baseline models and is in line with the new direction of research of fusing various infor-

Proceedings of the First Knowledgeable Natural Language Processing Workshop, pages 27–32
December 7, 2020.

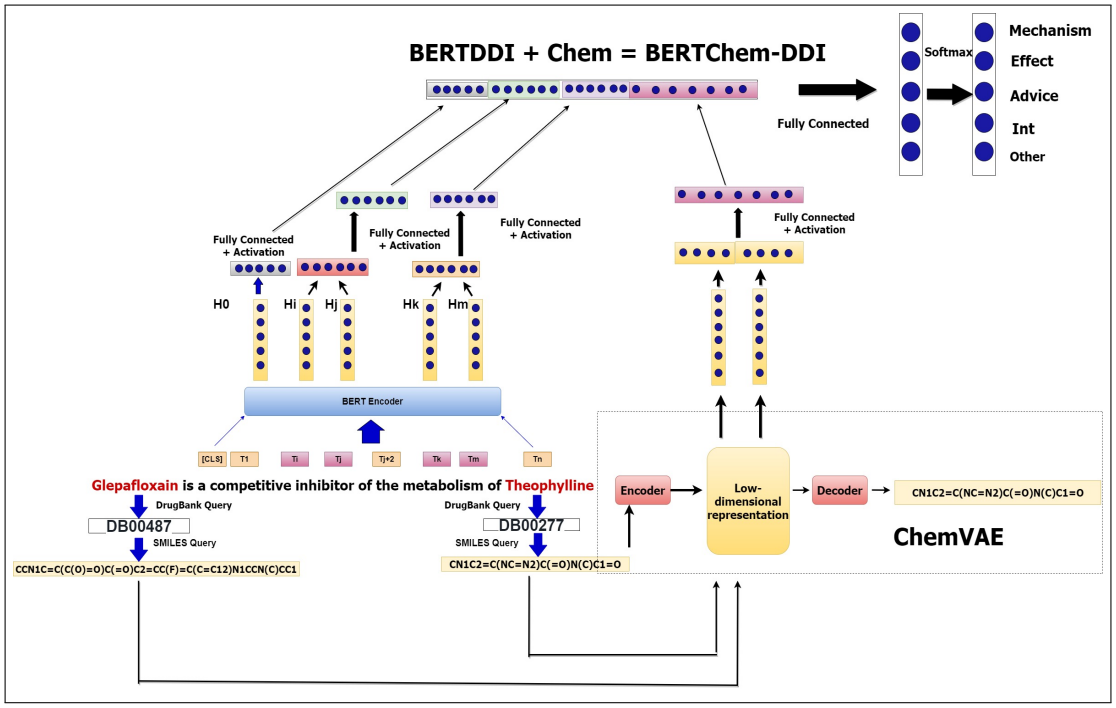

Figure 1: Schematic Representation of *BERTChem-DDI* with the input sentence *"Glepafloxain is a competitive inhibitor of the metabolism of Theophylline"* tagged with two drug entities **Glepafloxacin** and **Theophylline**.

mation to boost DDI classification performance.

In a nutshell, the major contributions of this work are summarized as follows:

- We propose a method that jointly leverages textual and external Knowledge information to classify relation type between the drug pairs mentioned in the text.
- We show the molecular information from the SMILES encoding using Variational AutoEncoder helps in extracting DDIs from texts.
- Our method achieves new state-of-the-art performance on DDI Extraction 2013 corpus.

2 Methodology

Given a sentence s with target drug entities d_1 and d_2, the task is to classify the type of relation (y) the drugs hold between them, $y \in (y_1, ..., y_N)$, where N denotes the number of relation types.

2.1 Text-based Relation Classification

Our model for extracting DDIs from texts is based on the pre-trained BERT-based relation classification model by (Wu and He, 2019). Given a sentence s with drugs d_1 and d_2, let the final hidden state output from BERT module is H. Let the vectors H_i to H_j are the final hidden state vectors from BERT for entity d_1, and H_k to H_m are the final hidden state vectors from BERT for entity d_2. An average operation is applied to obtain the vector representation for each of the drug entities. An activation operation *tanh* is applied followed by a fully connected layer to each of the two vectors, and the output for d_1 and d_2 are $H_1^{'}$ and $H_2^{'}$ respectively.

$$H_1^{'} = W_1[tanh(\frac{1}{(j-i+1)}\sum_{t=i}^{j} H_t] + b_1 \quad (1)$$

$$H_2^{'} = W_2[tanh(\frac{1}{(m-k+1)}\sum_{t=k}^{m} H_t] + b_2 \quad (2)$$

We make W_1 and W_2, b_1 and b_2 share the same parameters. In other words, we set $W_1 = W_2$ and keep $b_1 = b_2$. For the final hidden state vector of the first token ('[CLS]'), we also add an activation operation and a fully connected layer, which is formally expressed as:

$$H_0^{'} = W_0(tanh(H_0)) + b_0 \quad (3)$$

Matrices W_0, W_1, W_2 have the same dimensions, i.e. $W_0 \in R^{d*d}$,$W_1 \in R^{d*d}$, $W_2 \in R^{d*d}$, where d is the hidden state size from BERT.
We concatenate $H_0^{'}$, $H_1^{'}$ and $H_2^{'}$ and then add a fully connected layer and a softmax layer, which can be expressed as :

$$h^{''} = W_3[concat(H_0^{'}, H_1^{'}, H_2^{'})] + b_3 \quad (4)$$

$$y_t^{'} = softmax(h^{''}) \quad (5)$$

$W_3 \in R^{N*3d}$, and $y_t^{'}$ is the softmax probability output over N. In Equations (1), (2), (3), (4) the bias vectors are b_0, b_1, b_2, b_3. We use cross entropy as the loss function. We denote this text-based architecture as *BERT-DDI*.

2.2 Chemical Structure Representation

For the purpose of constructing an encoder from which a continuous latent representation is obtained, molecular representation of drugs has been used as both input and output.

Gómez-Bombarelli et al. (Gómez-Bombarelli et al., 2018) converted the discrete SMILES representations of the drug molecules into a continuous multi-dimensional representation using the unsupervised deep learning algorithm Variational Auto-Encoder(VAE) (Kingma and Welling, 2014). This representation has also been leveraged by (Purkayastha et al., 2019). The input x = (x_1 ,x_2 ,....,x_n) to VAE is represented by $x_i \in X$ where X = C, =, (,), O, F, 1, 2, $\cdots$ 9 in the SMILES representation. Each x_i is a X-dimensional one-hot vector. We denote this VAE architecture used in our experiments as *ChemVAE* and is explained as follows: As an encoder it uses three 1D convolutional layers, followed by a single fully-connected layer. The decoder uses three layers of GRU networks. The objective of this work is to maximize the probability distribution of generation of SMILES representation of drug molecules with the help of latent representation as presented in equation below:

$$P(X_{SMILES}) = \int P(X_{SMILES}|z)P(z)dz \quad (6)$$

In equation 6, X_{SMILES} denotes the drug molecules, z represents the latent SMILES representation, $P(X_{SMILES})$ denotes the probability distribution of drug molecules. The *ChemVAE* model takes SMILES representation of the drugs as input and encodes the drugs into continuous latent representation (z). The decoder then samples a string from the probability distribution over characters in the input SMILES representation. Finally, the hidden representation for each of the drug entities is treated as its chemical structure representation from *ChemVAE*.

	Train Set	Test Set
No. of unique drugs	2931	1055
No. of Normalized drugs	2670	997
No. of DDI Pairs	27779	5713

Table 1: Statistics of the DDI Extraction corpus 2013.

2.3 BERTChem-DDI

From the sentence s containing two target drug entities d_1 and d_2, we obtain the chemical structure representation of two drugs c_1 and c_2 respectively using *ChemVAE*. We concatenate these two embeddings c_1 and c_2 and pass those through a fully connected layer as represented as follows:

$$chm = W[concat(c1, c2)] + b \quad (7)$$

W and b are the parameters of the fully-connected layer of the chemical structure representation of d_1 and d_2. The final layer of *BERTChem-DDI model* contains the concatenation of all the previous text-based outputs (see Section 2.1) and chemical structure representation as expressed in the equations:

$$o^{'} = W_3[concat(H_0^{'}, H_1^{'}, H_2^{'}, chm)] + b_3 \quad (8)$$

$$y_t^{'} = softmax(o^{'}) \quad (9)$$

Finally the training optimization is achieved using the cross-entropy loss (L_t) :

$$L_t = \sum_t y_t \log y_t^{'} \quad (10)$$

3 Experimental Setup

In this section, we explain the dataset and experiments of using *ChemVAE* and *BERTChem-DDI*.

3.1 Dataset and pre-processing

We have followed the task setting of Task 9.2 in the DDIExtraction 2013 shared task (Herrero-Zazo et al., 2013) for the evaluation. This data set comprises of documents annotated with drug mentions and five types of interactions: *Mechanism*, *Effect*, *Advice*, *Int* and *Other*. The task is a multi-class classification to classify each of the drug pairs in the sentences into one of the types and we evaluate using Precision (P), Recall (R) and F1-score (F1) for each relation type.

During pre-processing, we obtain the *DRUG* mentions in the corpus and map those into unique *DrugBank* identifiers. This mention normalization

Embeddings on BERT-DDI	Test set Macro F1
bert-base-cased	0.806
scibert-scivocab-uncased	0.812
biobert v1.0 pubmed pmc	0.818
biobert v1.1 pubmed	0.822

Table 2: Ablation of the contextual embeddings.

Models	Embeddings	Macro F1
BERT-DDI	*biobert v1.0 pubmed pmc*	0.818
BERTChem-DDI	*biobert v1.0 pubmed pmc*	0.829
BERT-DDI	*biobert v1.1 pubmed*	0.822
BERTChem-DDI	*biobert v1.1 pubmed*	0.838

Table 3: Probing deeper into the influence of chemical structure information into the BERT-based models for DDI Relation Classification.

has been performed based on the longest overlap of drug mentions in the DrugBank. This mention normalization has been done for obtaining the corresponding *SMILES* representation to encode molecular structure information. The dataset statistics of the total drugs and the normalized drugs are enumerated in table 1. We initialize the non-normalized drug representations using pre-trained word2vec trained on PubMED [1].

3.2 Training Details

We make use of the pre-trained contextual embeddings such as ***bert-base-cased***, ***scibert-scivocab-uncased*** (Beltagy et al., 2019) and domain-specific ***biobert v1.0 pubmed pmc*** and ***biobert v1.0 pubmed*** as the initialization of the transformer encoder in **BERTChem-DDI**. We uniformly keep the maximum sequence length as 300, batch size 16, initial learning rate for ADAM optimizer as 2e-5, drop out 0.1 for all the embedding ablations and trained for 5 epochs. During unsupervised training of *ChemVAE* with drugs from ZINC (Irwin and Shoichet, 2005), the input SMILES representation has been trimmed to 120. The hidden dimension of *ChemVAE* encoder is 200 and for the decoder it is 500. Finally, a 292-dimensional representation of the drugs has been ultimately used for initialization of the *BERTChem-DDI* model's chemical structure representations of the drugs.

4 Results and Discussion

In this section, we provide a detailed analysis of the various results and findings that we have observed during experimentation. We have demonstrated

[1] http://evexdb.org/pmresources/ngrams/PubMed/

Methods	Adv F1	Eff F1	Mch F1	Int F1	Tot F1
(Sahu and Anand, 2018)	79	67	76	43	71
(Asada et al., 2018)	81	71	73	45	72
(Zhang et al., 2017)	80	71	74	54	72
(Sun et al., 2019)	80	73	78	**58**	75
(Vivian et al., 2017)	85	76	77	57	77
(Zhu et al., 2020b)	86	80	84	56	80
Our method	**88**	**80**	**87**	58	**83**

Table 4: Comparison of F1 scores for all the relation types using existing baselines on test set. *Adv* indicates 'Advice', *Mch* denotes 'Mechanism', 'Eff' means 'Effect', 'Tot' means overall.

strong empirical results based on the proposed approach for both text and chemical structure. We further want to understand the specific contributions by the chemical structure component besides the pre-trained BERT and its other domain-specific variants. For this purpose, we refer to our experimental configurations in meaningful ways while enumerating the results.

Ablation of Embeddings on BERT-DDI: During ablation analysis, we observe that the incorporation of domain-specific information in *biobert v.1 pubmed* boosts up the predictive performance in terms of macro-F1 score (across all relation types) by 2.3% compared to *bert-base-cased*. Moreover, the *scibert-vocab-cased* embedddings due to the scientific details obtained during fine-tuning achieves reasonable boost in performance. *biobert v.1 pubmed based BERT-DDI* is thus the best-performing text-based relation classification model. The results are enumerated in Table 2.

Advantage of Chemical Structure embeddings on BERTChem-DDI: During empirical analysis of the *BERTChem-DDI* model, we observe how much performance gain can be achieved by augmenting the chemical structure information. From the results enumerated in terms of macro F1-score on all the relation types in table 3, we observe that the best-performing *BERT-DDI* model achieves a performance boost of 1.6% after adding chemical structure information in *BERTChem-DDI*. Probing deeper, we observe that the relation types *Mechanism* (3.2%) and *Advice* (2.11%) achieve significant performance improvement over *BERT-DDI*.

Comparison with the existing baselines: We compare our best-performing model with some of the best-performing existing baselines. Our method achieves the state-of-the-art performance based on

the results in Table 4.

5 Conclusion

In this paper, we develop an approach for DDI relation classification based on pre-trained language model and chemical structure representation of drugs. Experiments on the benchmark DDI dataset proves the efficacy of our method. Possible directions of further research might be to explore Knowledge Graph based drug representation combined with textual description and other relation specific embeddings obtained from various ontologies.

Acknowledgement

This work is an extension of the thesis work by the author during her course at the Indian Institute of Technology, Kharagpur. Besides, the author would also like to thank the anonymous reviewers for their insightful comments and feedback on the paper.

References

Masaki Asada, Makoto Miwa, and Yutaka Sasaki. 2017. Extracting drug-drug interactions with attention CNNs. In *BioNLP 2017*, pages 9–18, Vancouver, Canada,. Association for Computational Linguistics.

Masaki Asada, Makoto Miwa, and Yutaka Sasaki. 2018. Enhancing drug-drug interaction extraction from texts by molecular structure information.

Iz Beltagy, Kyle Lo, and Arman Cohan. 2019. Scibert: A pretrained language model for scientific text. In *EMNLP/IJCNLP*.

J. Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *NAACL-HLT*.

Rafael Gómez-Bombarelli, Jennifer N. Wei, David Duvenaud, José Miguel Hernández-Lobato, Benjamín Sánchez-Lengeling, Dennis Sheberla, Jorge Aguilera-Iparraguirre, Timothy D. Hirzel, Ryan P. Adams, and Alán Aspuru-Guzik. 2018. Automatic chemical design using a data-driven continuous representation of molecules. *ACS Central Science*, 4(2):268–276. PMID: 29532027.

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.

John Irwin and Brian Shoichet. 2005. Zinc a free database of commercially available compounds for virtual screening. *Journal of chemical information and modeling*, 45:177–82.

Md. Rezaul Karim, Michael Cochez, Joao Bosco Jares, Mamtaz Uddin, Oya Beyan, and Stefan Decker. 2019. Drug-drug interaction prediction based on knowledge graph embeddings and convolutional-lstm network. In *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, BCB '19, page 113–123, New York, NY, USA. Association for Computing Machinery.

Diederik Kingma and Max Welling. 2014. Auto-encoding variational bayes.

Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. 2019. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4):1234–1240.

Diya Li and Heng Ji. 2019. Syntax-aware multi-task graph convolutional networks for biomedical relation extraction. In *Proceedings of the Tenth International Workshop on Health Text Mining and Information Analysis (LOUHI 2019)*, pages 28–33, Hong Kong. Association for Computational Linguistics.

Shengyu Liu, Buzhou Tang, Qingcai Chen, and Xiaolong Wang. 2016. Drug-drug interaction extraction via convolutional neural networks. *Computational and Mathematical Methods in Medicine*, 2016:1–8.

S. Purkayastha, I. Mondal, S. Sarkar, P. Goyal, and J. K. Pillai. 2019. Drug-drug interactions prediction based on drug embedding and graph auto-encoder. In *2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE)*, pages 547–552.

Sunil Kumar Sahu and Ashish Anand. 2018. Drug-drug interaction extraction from biomedical texts using long short-term memory network. *Journal of Biomedical Informatics*, 86:15 – 24.

Xia Sun, Ke Dong, Long Ma, Richard Sutcliffe, Feijuan He, Sushing Chen, and Jun Feng. 2019. Drug-drug interaction extraction via recurrent hybrid convolutional neural networks with an improved focal loss. *Entropy*, 21(1):37.

Vivian Vivian, Hongfei Lin, Ling Luo, Zhehuan Zhao, li Zhengguang, Zhang Yijia, Zhihao Yang, and Jian Wang. 2017. An attention-based effective neural model for drug-drug interactions extraction. *BMC Bioinformatics*, 18.

David Wishart, Craig Knox, An Chi Guo, Dean Cheng, Savita Shrivastava, Dan Tzur, Bijaya Gautam, and Murtaza Hassanali. 2008. Drugbank: a knowledgebase for drugs, drug actions and drug targets. nucleic acids res 36:d901-d906. *Nucleic acids research*, 36:D901–6.

Shanchan Wu and Yifan He. 2019. Enriching pre-trained language model with entity information for relation classification.

Z. Yang, Zihang Dai, Y. Yang, J. Carbonell, R. Salakhutdinov, and Quoc V. Le. 2019. Xlnet: Generalized autoregressive pretraining for language understanding. In *NeurIPS*.

Yijia Zhang, Wei Zheng, Hongfei Lin, Jian Wang, Zhihao Yang, and Michel Dumontier. 2017. Drug–drug interaction extraction via hierarchical RNNs on sequence and shortest dependency paths. *Bioinformatics*, 34(5):828–835.

Yu Zhu, Lishuang Li, Hongbin Lu, Anqiao Zhou, and Xueyang Qin. 2020a. Extracting drug-drug interactions from texts with biobert and multiple entity-aware attentions. *Journal of biomedical informatics*, 106:103451.

Yu Zhu, Lishuang Li, Hongbin Lu, Anqiao Zhou, and Xueyang Qin. 2020b. Extracting drug-drug interactions from texts with biobert and multiple entity-aware attentions. *Journal of Biomedical Informatics*, 106:103451.