SubM: Predicting Drug-Drug Interaction Events using Long Short Term Memory Networks

Monalika Padma Reddy
Department of Computer Science
Old Dominion University

Shannon M. Stephens
Department of Computer Science
Old Dominion University

Abstract

This paper addresses drug-drug interaction (DDI) event prediction by introducing a long short term memory (LSTM) classification network. Multi-class classification was performed using a modified dataset derived from the DrugBank. LSTM models have been gaining prominence within the natural language processing field, but have yet to be introduced within DDI prediction tasks. The introduced model utilizes SMILES, drug substructure, and Mordred molecular descriptor data. Model prediction results show model underperformance within the real-world data.

1 Introduction

Drug interactions occur when two (or more) phamecuticals interact, as well as when a drug interacts with a person's diet or taken supplements. A drug interaction can also result from an underlying medical issue. Drug interactions can have negative consequences or alter how well a drug(s) works. Drug–drug interactions (DDIs) is the pharmacological action between drug components, which can occur during a pharmaceutical co-administration. Finding possible DDI allows the investigation of the mechanism underlying drug interactions or adverse responses in order to prevent the negative effects. DDI increased a range of possible hazards for patients, such as the simultaneous use of multiple medications in cancer patients. After clinical use was approved, numerous DDIs that were unknown at the time of the clinical trial were reported (van Leeuwen, 2015). Because it can significantly lower the potential risk brought on by combination medications, it is crucial to monitor DDI and investigate suspected DDIs prior to taking a medication.

Even though numerous artificial intelligence techniques are used to mine and forecast probable DDIs, these techniques often don't fully utilize structural data from within their symbolic representation and do not completely account for the role that molecular substructure plays in DDIs. Machine Learning methods based on statistics are proposed. The model representations based on machine learning usually depends mainly on the effective data feature extraction. Initially the FBK IRST (Chowdhury, 2013) was proposed, a model in which two independent steps – DDI detection and a binary support vector machine classifier were used to train using context and shallow language features. However, deep learning methods have been recently prefered due to not relying on detailed pre-extraction of features. Early deep learning models used to predict DDIs include manifold regularization (Zhang, 2018). (Deng, 2020) learnt the representation of the DDIs in the representation modules using the chemical structure to learn the and then predicted events in comparing module and (Gao,2021) used the deep neural networks and different drug features to predict DDIs.

The Simplifed Molecular Input Line Entry System (SMILES) sequence is frequently used for the input of drug representations, including the previously mentioned work and this current work. Alternativily, graph-based representation learning has recently been applied to Drug-Drug interaction prediction models. Molecular graphs are used to represent the drugs, where the atom is represented as a node and the chemical bond is represented as the edge. In (Feng, 2020) DDI prediction was performed using a Graph Convolutional Network (GCN) and Deep Neural Network (DNN). (Asada, 2018)

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

used the molecular structure as the input to a GCN model. This was subsequently concatenated with a Convolutional Neural Network (CNN) model. Later, they used SCIBERT input text and drug description concatenated GNNs of the input drug smiles. This makes use of the structural or descriptive characteristics of drugs. (Schutt, 2018) developed Schnet to use distance embeddings as the location information into the aggregation information of the model. In (Li, 2020), attention weight of the molecular graph model, the lowest unoccupied molecular orbital and highest occupied molecular orbital was obtained. (Nyamabo, 2021) predicted DDIs by interactions between drug substructures. A heat map was plotted to demonstrate the substructure contribution in the molecules to predict the drug relationship. (Nyamabo, 2022) used gating devices to learn the chemical substructures of drugs. However, these proposed methods consider only the sequence, structure or the interaction between the drugs without considering the synergistic effects between the drugs that would be captured using a descriptor calculator. In the DDI networks, the diversity of the nodes will be lost because of the multi-layer GNN stacking. The GNNs can only extract the local features of the atoms in the molecular graphs but, the propagation of information is difficult to remotely capture the global features of the molecular graphs.

This work aims to combine molecular substructures and calculated molecular descriptors to classify DDI events by utilizing long short term memory networks (LSTM). These Recurrent Neural Networks (RNNs) originate in natural language processing, allowing the model to better predict later occuring text, using previous text. LSTM models offer reduced dimensionality compared with more traditional neural networks by working in loops when learning parameter weights.

2 Methods

2.1 Dataset

Our model's dataset was created using the Drugbank database (Wishart, 2006) and based off of the dataset for the 3DGT-DDI model (He, 2022), META-DDIE methodology (Deng, 2022), and the Mordred molecular descriptor calculator (Moriwaki, 2018). The Drugbank is an exhaustive database that contains drug names, targets, stuctures (via SMILES and InChi codes), etc., compiled into a unique drug ID system. DDI's between different pharmacuticals are among the data included. The 3DGT-DDI dataset created unique entries for each DDI event and classified them into five potential classes (Table 1).

Table 1: Dataset Classification Targets

	Training Set	Test Set
Total Drug Pairs	10215	3748
Negative Pairs	8862	3163
Advice	241	133
Effect	417	2
Interaction	64	46
Mechanism	631	229

Each drug entry contains a SMILES formula that encodes the structural data of the drug molecule. These SMILES codes were broken up by substructures using a molecular fingerprint like system (Huang, 2019). Each substructure was further encoded into a numerical value in which each number corresponds to a unique dictionary entry. Model comparisons were performed between the default dictionary encoding and with the dictionary was arranged by molecular weight. These comparisons are shown in Section 3.[1] The Mordred molecular descriptor calculator was used calculate more then 1600 2D and 3D descriptors of the drugs. Considering that each row within the dataset contained two drugs, with greater then 1600 hundred descriptors each, the entries were required to be reduced to simplify the model input. The number of Mordred descriptors was reduced from 3226 total entries to 567 by removing the descriptors that did not change between different drug rows in the dataset,

[1] One-hot encoding of each substructure was attempted There was considerable difficulty in employing this technique due to the formatting of the dataset.

removing descriptors that were correlated to each other by greater than 95%, and by performing a random forest algorithum to remove features that were unimportant to the classification task. The model also used the drug's SMILES strings as input.

2.2 Model Development

Long Short Term Memory networks (LSTMs) are explicitly designed to avoid the long-term dependency problem. Instead of having a single neural network layer, there are four "chained together" that interact with each other. The LSTM's cell state works as the model's long term memory, acting kind of like a conveyor belt. It runs straight down the entire chain, with only some minor linear interactions. It's very easy for information to just flow along it unchanged. The LSTM does have the ability to remove or add information to the cell state through gates. Gates are composed out of a sigmoid neural net layer and a pointwise multiplication operation. The sigmoid layer outputs numbers between zero and one, describing how much of each component should be let through. A value of zero means "let nothing through," while a value of one means "let everything through!" An LSTM has three of these gates, to protect and control the cell state. The first step in our LSTM is to decide what information is going to throw away from the cell state. This decision is made by a sigmoid layer called the, forget gate layer. It outputs a number between 0 and 1 for each number in the cell state. The next step is to decide what new information is going to be stored in the cell state. First, a sigmoid layer called the "input gate layer" decides which values are going to be updated. Next, an additional layer creates a vector of new candidate values that could be added to the state. These two are subsequently combined to create an update to the state. This output will be based on the cell state but will be a filtered version using an addictional sigmoid layer.

2.3 Implementation Details

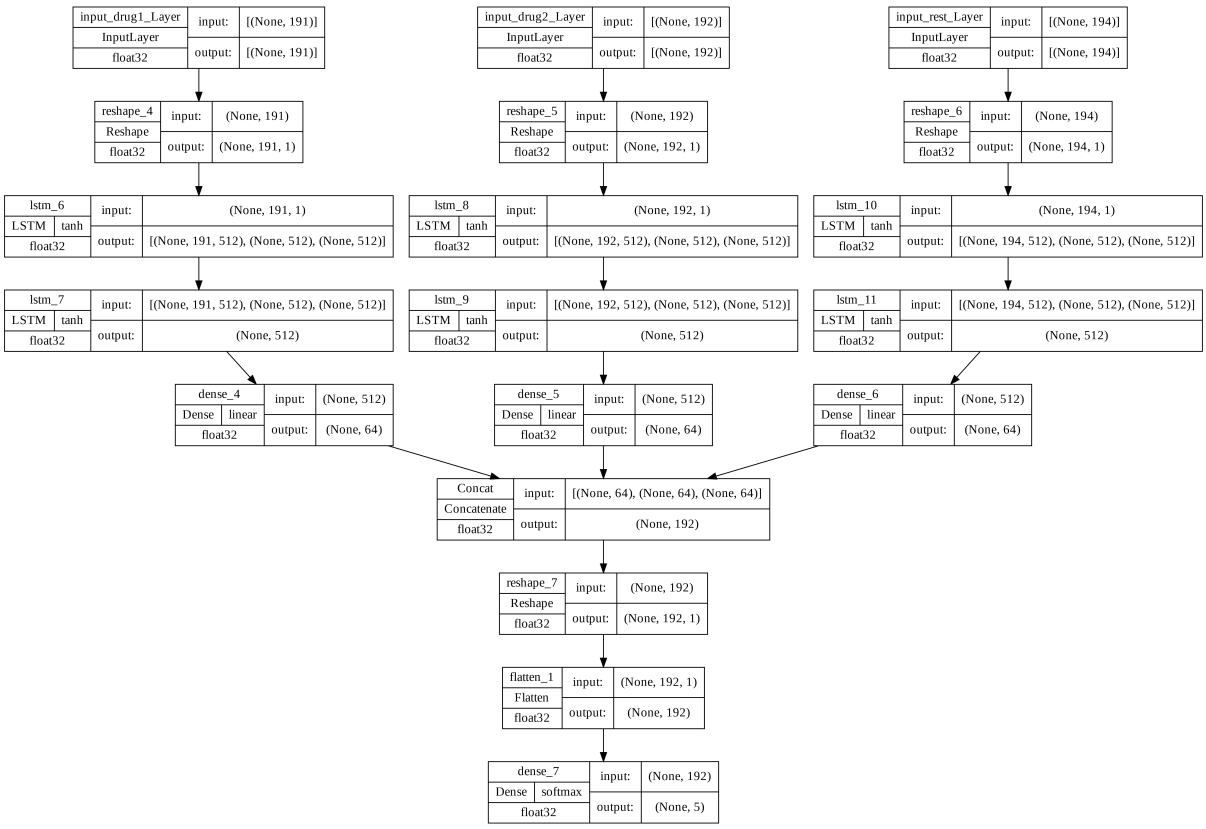

Figure 1: LSTM model architecture and design

Table 2: Implimentation Details

Layer (type)	Output Shape	Param #	Connected to
Drug 1 input layer	[(None, 191)]	0	[ ]
Drug 2 input layer	[(None, 191)]	0	[ ]
input_rest_layer	[(None, 192)]	0	[ ]
reshape_4 (reshape)	(None, 191,1)	0	['input_drug1_Layer[0][0]']
reshape_5 (reshape)	(None, 191,1)	0	['input_drug2_Layer[0][0]']
reshape_6 (reshape)	(None, 196,1)	0	['input_rest_Layer[0][0]']
lstm_6 (LSTM)	[(None, 191,512)	1052672	['reshape_4[0][0]']
(None, 512)			
(None, 512)]			
lstm_8 (LSTM)	[(None, 191,512)	1052672	['reshape_5[0][0]']
(None, 512)			
(None, 512)]			
lstm_10 (LSTM)	[(None, 192,512)	1052672	['reshape_6[0][0]']
(None, 512)			
(None, 512)]			
lstm_7 (LSTM)	(None, 512)	2099200	['lstm_6[0][0]']
'lstm_6[0][1]'			
'lstm_6[0][2]'			
lstm_9 (LSTM)	(None, 512)	2099200	['lstm_8[0][0]']
'lstm_8[0][1]'			
'lstm_8[0][2]'			
lstm_10 (LSTM)	(None, 512)	2099200	['lstm_10[0][0]']
'lstm_8[0][1]'			
'lstm_8[0][2]'			
dense_4 (Dense)	(None,64)	32832	['lstm_7[0][0]']
dense_5 (Dense)	(None,64)	32832	['lstm_9[0][0]']
dense_6 (Dense)	(None,64)	32832	['lstm_11[0][0]']
Concat (Concatenate)	(None, 192)	0	['dense_4[0][0]', 'dense_5[0][0]', 'dense_6[0][0]']
reshape_7 (Reshape)	(None, 192, 1)	0	['Concat[0][0]']
flatten_1 (Flatten)	(None, 192)	0	['reshape_7[0][0]']
dense_7 (Dense)	(None,5)	965	['flatten_1[0][0]']

Total parameters: 9,555,077
Trainable parameters: 9,555,0777
Non-trainable parameters: 0

Table 2 and Figure 1 includes the implimentation details of the model design. The final model architecture include three LSTM modules, trainined on Drug 1, Drug 2, and the Mordred calculations. The modules outputs were concatenated and used for DDI classification. A softmax activation function was utilized with a sparse categorical cross-entropy loss function employed.

3 Results

The model results are included in Table 3 and Table 4. For each of the model configurations, the training and test datasets were combined and split according to 80% training and 20% test. This mixture of the two datasets was performed due to drugs occuring in the test set that did not occur in the training set, to allow the model to predict the best possible classification performance. Random undersampling was performed to test model performance with a more balanced dataset.

Table 3: LSTM model classification scores for full dataset

DrugBank Label	Original Arrangement	Substructures Rearranged
	Precision Score\ Recall\ F1-score	Precision Score\ Recall\ F1-score
Negative Pair	0.86 \ 1.00 \ 0.93	0.86 \ 1.00 \ 0.93
Advice	0.00 \ 0.00 \ 0.00	0.00 \ 0.00 \ 0.00
Effect	0.00 \ 0.00 \ 0.00	0.00 \ 0.00 \ 0.00
Interaction	0.00 \ 0.00 \ 0.00	0.00 \ 0.00 \ 0.00
Mechanism	0.00 \ 0.00 \ 0.00	0.00 \ 0.00 \ 0.00
Model Accuracy	0.86	0.86

Table 4: LSTM model classification scores for undersampled dataset

DrugBank Label	Original Arrangement	Substructures Rearranged
	Precision Score\ Recall\ F1-score	Precision Score\ Recall\ F1-score
Negative Pair	0.40 \ 0.89 \ 0.55	0.00 \ 0.00 \ 0.00
Advice	0.00 \ 0.00 \ 0.00	1.00 \ 0.03 \ 0.05
Effect	0.29 \ 0.26 \ 0.27	0.31 \ 0.79 \ 0.45
Interaction	0.00 \ 0.00 \ 0.00	0.00 \ 0.00 \ 0.00
Mechanism	0.00 \ 0.00 \ 0.00	0.24 \ 0.34 \ 0.28
Model Accuracy	0.37	0.29

The results tables both indicate very poor model perforance. The classification report indicates that the model was unable to learn from the SMILES string, substructure numbers, and Mordred calculation inputs. Model performance did not markedly improve using undersampling. Future work on DDI classification problems will include re-processing the dataset to a format that the model is better able to learn from. Specifically, addressing drug substructure data differently could lead to better model accuracy.

References

Chowdhury, M. F. M., & Lavelli, A. (2013, June). FBK-irst: A multi-phase kernel based approach for drug-drug interaction detection and classification that exploits linguistic information. In *Second Joint Conference on Lexical and Computational Semantics (* SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)* (pp. 351-355).

Deng, Y., Xu, X., Qiu, Y., Xia, J., Zhang, W., & Liu, S. (2020). A multimodal deep learning framework for predicting drug–drug interaction events. *Bioinformatics*, 36(15), 4316-4322.

Deng, Y., Qiu, Y., Xu, X., Liu, S., Zhang, Z., Zhu, S., & Zhang, W. (2022). META-DDIE: predicting drug–drug interaction events with few-shot learning. *Briefings in Bioinformatics*, 23(1), bbab514.

Feng, Y. H., Zhang, S. W., & Shi, J. Y. (2020). DPDDI: a deep predictor for drug-drug interactions. *BMC bioinformatics*, 21(1), 1-15.

He, H., Chen, G., & Yu-Chian Chen, C. (2022). 3DGT-DDI: 3D graph and text based neural network for drug–drug interaction prediction. *Briefings in Bioinformatics*, 23(3), bbac134.

Huang, K., Xiao, C., Glass, L., & Sun, J. (2019, December). Explainable substructure partition fingerprint for protein, drug, and more. In *NeurIPS Learning Meaningful Representation of Life Workshop*.

Lyu, T., Gao, J., Tian, L., Li, Z., Zhang, P., & Zhang, J. (2021). MDNN: A Multimodal Deep Neural Network for Predicting Drug-Drug Interaction Events. *IJCAI*, (pp. 3536-3542).

Moriwaki, H., Tian, Y. S., Kawashita, N., & Takagi, T. (2018). Mordred: a molecular descriptor calculator. *Journal of cheminformatics*, 10(1), 1-14.

Nyamabo, A. K., Yu, H., & Shi, J. Y. (2021). SSI–DDI: substructure–substructure interactions for drug–drug interaction prediction. *Briefings in Bioinformatics*, 22(6), bbab133.

Nyamabo, A. K., Yu, H., Liu, Z., & Shi, J. Y. (2022). Drug–drug interaction prediction with learnable size-adaptive molecular substructures. *Briefings in Bioinformatics*, 23(1), bbab441.

Schütt, K. T., Sauceda, H. E., Kindermans, P. J., Tkatchenko, A., & Müller, K. R. (2018). Schnet–a deep learning architecture for molecules and materials. *The Journal of Chemical Physics*, 148(24), 241722.

van Leeuwen, R. W., Jansman, F. G., van den Bemt, P. M., de Man, F., Piran, F., Vincenten, I., ... & van Gelder, T. (2015). Drug–drug interactions in patients treated for cancer: a prospective study on clinical interventions. *Annals of Oncology*, 26(5), 992-997.

Wishart, DS., Knox, C., Guo, AC., Shrivastava, S., Hassanali, M., Stothard, P., Chang, Z., & Woolsey, J. (2006) Drugbank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids* Res. Jan 1;34 (Database issue):D668-72. 16381955.

Zhang, W., Chen, Y., Li, D., & Yue, X. (2018). Manifold regularized matrix factorization for drug-drug interaction prediction. *Journal of biomedical informatics*, 88, 90-97.

Zhu, Y., Li, L., Lu, H., Zhou, A., & Qin, X. (2020). Extracting drug-drug interactions from texts with BioBERT and multiple entity-aware attentions. *Journal of biomedical informatics*, 106, 103451.