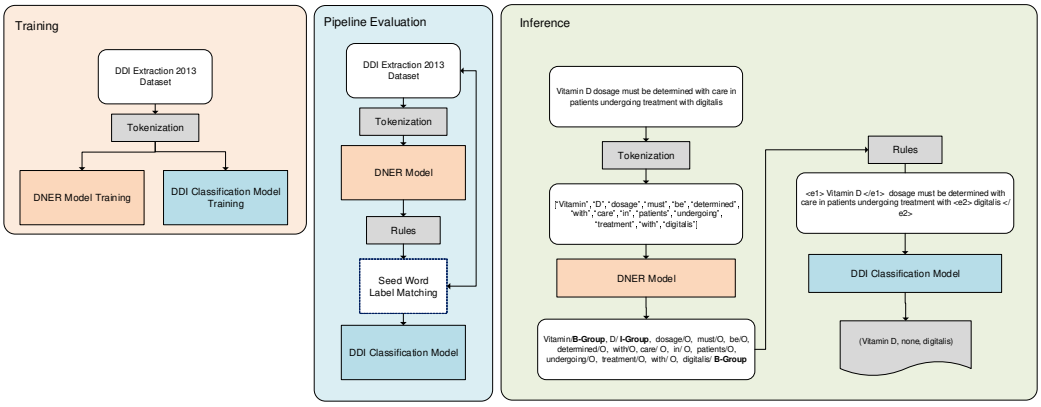
Graphical Abstract

TP-DDI: Transformer-based Pipeline for the extraction of Drug-Drug Interactions

Dimitrios Zaikis, Ioannis Vlahavas



Highlights

TP-DDI: Transformer-based Pipeline for the extraction of Drug-Drug Interactions

Dimitrios Zaikis, Ioannis Vlahavas

- This research study proposed an end-to-end Drug-Drug Interaction pipeline
- The proposed approach out-performs current state-of-the-art end-to-end systems
- A set of rules is used to mitigate error propagation in the pipeline
- The proposed approach is based on the Transformer architecture

TP-DDI: Transformer-based Pipeline for the extraction of Drug-Drug Interactions

Dimitrios Zaikis[a,*], Ioannis Vlahavas[a]

[a]*School of Informatics, Aristotle University of Thessaloniki, Greece*

Abstract

Drug-Drug Interaction (DDI) extraction is the task of identifying drug entities and the potential interactions between drug pairs from biomedical literature. Computer-aided extraction of DDIs is vital for drug discovery, as this process remains extremely expensive and time consuming. Therefore, Machine Learning-based approaches can reduce the laborious task during the drug development cycle. Numerous traditional and Neural Network-based approaches for Drug Named Entity Recognition (DNER) and the classification of DDIs have been proposed over the years. However, despite the development of many effective methods, achieving good prediction accuracy is an area where significant improvement can be made. In this article, we present a novel end-to-end approach that tackles the overall DDI extraction task as a pipelined method via the Transformer model architecture and biomedical domain pre-trained weights. In our approach, the tasks of DNER and DDI classification are executed successively to extract the drug entities and to classify their relationship respectively. The proposed approach, TP-DDI, integrates prior knowledge by using pre-trained weights from BioBERT and improves in both the Drug Named Entity Recognition and the overall DDI extraction task over the current state-of-the-art approaches on the DDI Extraction 2013 corpus.

Keywords: Drug-Drug Interaction, Relationship Extraction, Drug Named Entity Recognition, Relation Classification, Pipeline

*Corresponding author at: School of Informatics, Aristotle University of Thessaloniki, Thessaloniki, 54124, Greece.

Email addresses: dimitriz@csd.auth.gr (Dimitrios Zaikis), vlahavas@csd.auth.gr (Ioannis Vlahavas)

1. Introduction

Studies ranging from clinical trials and meta-analyses to systematic reviews are published at an exponential rate. This information is crucial to medical practitioners and researchers who rely on the latest published findings for their clinical decision-making process in order to provide better patient care. However, due to the exponential growth of biomedical literature, following the latest developments requires spending an exorbitant amount of valuable time. Even though Machine Learning (ML), with the use of Natural Language Processing (NLP) has gained ground towards advancing the field of Biomedical Informatics [1], Drug-Drug Interactions (DDI) identification from biomedical literature, constitutes an area with a lot of room for improvement.

During the drug discovery and development process, the preclinical research phase aims to determine the safety of a new potential drug candidate where new drugs get tested for efficacy, toxicity and Pharmacokinetic (PK) information. These trials are conducted by scientists and aim to determine the side effects, adverse events and the possible interactions with other drugs. Changes in the PK and Pharmacodynamic (PD) properties of a drug are the main cause of DDIs, which may result in Adverse Drug Reactions (ADR). DDI identification refers to the task of identifying the effect produced by a combination of two or more drugs.

Numerous drug databases, such as DrugBank, PharmGKB, Stockley, DailyMed, WebMD, National Drug File and Kyoto Encyclopedia of Genes and Genomes exist, which provide medical professionals the ability to retrieve DDI information. However, due to the aforementioned rapid growth of biomedical literature, a large quantity of valuable DDI information remains hidden in articles and publications, making the task of maintaining an up-to-date drug knowledge base a challenging endeavor. Studies suggest further facilitation of access to this type of sources due to the high number of interacting drug combinations and the limited ability of prescribers to identify them [2]. Therefore, the effective automatic extraction of drug entities and their interactions can contribute significantly to pharmacovigilance, also known as drug safety, and provide up-to-date information to drug databases.

DDI information retrieval requires an extensive workload involving topic identification, evidence search, evidence synthesis and recommendations generation. The process of locating evidence is the most critical step, due to the absence of a single archive containing all available information on DDI. The

broad range of sources in combination with the emergence of new evidence make the extraction of DDI information extremely difficult. Current methods of extraction employed by medical practitioners and researchers rely on a comprehensive search strategy for manually locating relevant information from clinical trials, case reports and systematic reviews [3]. Therefore, the automatic extraction of drug entities and their interactions from biomedical literature aims to significantly speed up this process by identifying the drug names and their relationships, retrieving the most relevant DDI information.

The objective of the DDI task is to discover mentions of drug named entities in text and extract drug interaction relations between drug entity pairs. The entity and interaction types studied in this work are from the gold standard dataset that was introduced with the SemEval DDI Extraction Challenge 2013 [4, 5].

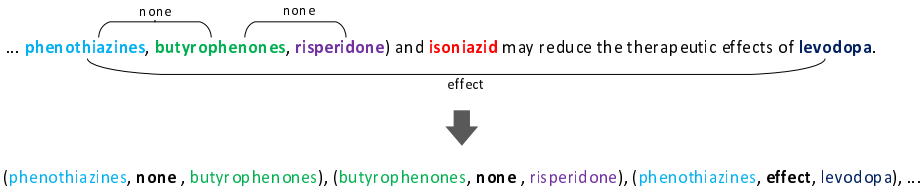


Figure 1: Example of a DDI extraction task sentence and the resulting drug entities and interaction triples.

In recent years, various methods for the extraction of DDIs have been proposed, based either on a single task, Drug Name Entity Recognition (DNER) or DDI classification, or DDI extraction in an end-to-end approach, which can be divided into joint and pipelined approaches.

DNER approaches aim to recognize drug entity mentions in biomedical texts and classify them into predefined categories. However, while DNER is related to the conventional Named Entity Recognition (NER) task, domain-specific challenges exists due to variations in the naming of a drug, frequent occurrences of abbreviations and acronyms [6] and complex naming schemes with numbers and symbols.

Comparably, DDI classification approaches focus solely on the task of classifying the relation of drug pairs in biomedical texts. The drug entities are from datasets where each entity pair is labeled with the predefined relation types. The types *"advice"*, *"mechanism"*, *"effect"* and *"int"* denote the types of interactions between two drugs and correspond to the positive class. The

types *"false"* or *"none"* that are used interchangeably, denote the absence of an interaction between two drugs and correspond to the negative class.

For the end-to-end DDI extraction task, joint and pipelined methods focus on the overall task of DDI extraction by implementing both DNER and DDI classification in a single system. Joint modeling methods approach the tasks of recognizing drug entities and classifying their relation as a single biomedical entity and relation extraction task. However, due to the many overlapping relations in biomedical texts, the current proposed methods convert the task into a tagging problem.

Pipelined methods separate the biomedical relation extraction and classification into two distinct tasks and address them in a sequential manner. Initially, the drug entities are extracted from the given literature using DNER techniques and all possible drug entity pairs in a given text are formed. Subsequently, the pairs are classified into predefined task-specific categories, forming the entities-relation triple, as shown in Figure 1.

Early approaches used pattern-based methods that rely on hand-crafted patterns to classify drug interactions, which are time-consuming and rely on domain expert knowledge. With the emergence of annotated corpora, ML approaches have achieved great success and recent research has shown great promise in using Deep Neural Networks (DNN) for all DDI extraction related tasks. However, traditional ML and DNN approaches rely heavily on laborious feature engineering and feature selection.

In our previous work [7], we proposed an attention-based "Bi-LSTM-CNN" model for the single task of DDI classification. This paper expands on our previous methodology by proposing an end-to-end Neural Network-based learning approach to the pipelined extraction of biomedical entities and the classification of the interactions between them. We aim to provide a simplified approach to the recognition of named drug entities and the classification of their interactions by taking advantage of the Transformer architecture [8] and the BERT Language Model (LM) that have been shown to improve NLP tasks [9].

First, with the use of in-domain pre-trained weights, we expand on BioBERT to recognize the drug named entities and classify them into four categories. Then, we apply a set of rules to correct possible misclassifications from the previous step, and create all combination of drug pairs and filter sentences where no relation exists. Finally, with the use of BioBERT, the relation of the drug pairs is classified into one of the five aforementioned categories. The main contribution of our work can be summarized as follows:

1. We approach this task with no preprocessing and feature engineering, eliminating the complexity of data preparation.
2. We develop a complete system by using a pipelined approach to extract drug entities and classify their relations.
3. We apply a set of non-complex rules to prepare the data for the relation classification step.
4. We explore the effectiveness of different pre-trained weights from different domains. The experimental results show that the pre-trained weights from the biomedical domain are the most effective and can further improve the performance.

We conducted the experiments on the DDI Extractions 2013 dataset and our results show that our pipelined method outperforms the existing approaches to the DDI extraction task and achieves state-of-the-art performance in both the Drug Named Entity Recognition task and the overall DDI extraction task.

2. Related Work

The Drug-Drug Interaction extraction task is a Relationship Extraction (RE) task that extracts semantic relationships between different entities from text. The subtasks consists of the recognition of named entities and the classification of their relationships, extracting triples using NLP techniques [10]. Drug names are extracted using Drug Named Entity Recognition (DNER) techniques and the interactions between drugs are classified using Relationship Classification (RC) techniques. The appearance of the SemEval-2013 DDI Task [11] extraction challenge enabled researchers to evaluate the effectiveness of NLP-based DDI extraction methods on the same gold standard corpus. As a result, various end-to-end DDI extraction models focusing on pipelined and joint methods have been proposed.

The pipelined methods, as mentioned previously, treat the extraction of DDIs as two separated tasks, DNER and Relation Classification. DNER is a traditional Named Entity Recognition task, specific to the biomedical domain. Typical NER methods are based on Deep Learning (DL) techniques while DNER methods utilize manually generated semantic and syntactic features. These methods are evaluated on the CoNLL 2003 dataset [12], which is considered as the benchmark corpus. State-of-the-art NER systems take advantage of Transformers, a novel architecture that handles

long-range dependencies in sequence-to-sequence tasks. Transformers, in the form of stacked encoders also serve as bases for BERT [13], while the best performing NER system similarly uses shallow bidirectional Transformers. However, the Global Context enhanced Deep Transition (GCDT) architecture described in Liu Y. et al. [14], which has no statistically significant differences in performance, make use of combinations of contextualized text representations and deep Recurrent Neural Networks (RNNs), along with an encoder for sequence classification to achieve similar performance.

Transitioning to the biomedical domain, the existence of varying in scope corpora that can be considered as benchmark datasets, render the identification of a clear state-of-the-art DNER system difficult. However, BioBERT [15], which is a fine-tuned BERT model trained on biomedical literature from PubMed, appears to outperform most DNER systems in almost all datasets. As an exception, CollaboNet[16] outperforms BioBERT on the JNLPBA dataset [17] for cell-line identification. Their proposed method uses a combination of three pre-trained Bi-LSTM-CRF architecture DNERs on chemicals, diseases and genes, to be used as extra-linguistic information in tandem with a weighted-pooling mechanism.

Similarly, due to the different proposed approaches focusing on either the complete DDI extraction task or only the RC sub-task, a clear state-of-the-art DDI extraction system is difficult to identify in the literature. Recent publications range from focusing on a single task in the extraction process, either DNER or RE, to joint and pipelined end-to-end systems [7]. Most RE and RC models treat this task as a supervised multiclass classification problem, with the exception of a few clustering methods [18]. The supervised approaches can be roughly divided into two categories: feature-based and DNN-based.

Current feature-based approaches rely heavily on manually generated features such as Part-of-Speech (POS) tags, syntactic and dependency parsing, obtained with laborious feature engineering and feature selection [19]. Likewise, kernel-based approaches that use syntactic information also proved effective for this work [20, 21]. DNN-based approaches which are able to learn the latent semantic features and better representations through the training process and consequently minimize the dependency on feature engineering and preprocessing techniques, prove to be very effective in the RC task [22, 23]. Similarly, graph-based models, based on Graph Convolutional Networks, have been applied to this task and achieved good results with the use of the Entity Pair Graph concept in combination with a Graph Neural

Network model that is able to incorporate semantic features from a sentence and topological features for relation classification [24].

State-of-the-art systems employ joint entity and relation modelling methods instead of pipelined methods, converting the DNER and RC tasks to a single task. The approach of Luo et al. [25], called "Att-BiLSTM-CRF", uses a combination of three embeddings, pre-trained word embeddings from a word2vec model, pretrained ELMo embeddings and character embeddings that are learned in the process by a very simplistic NER system. The main model consists of a BiLSTM network that creates latent representations from the three concatenated inputs, an attention mechanism over the hidden states of the Bi-LSTM to assign scores to the latent features produced, and a CRF layer used for predictions. The character embeddings are following the approach of Ma et al. [26] to extract features based on the characters. Additionally, in order to overcome the vast amount of overlapping relations that are present in the biomedical literature, a tagging scheme and extraction rules in combination with ELMo embeddings was employed to improve the performance of the "Att-BiLSTM-CRF" system.

In contrast to the above systems, our approach removes the dependency on feature-engineering and preprocessing, incorporating a rule-set in a pipeline, producing an end-to-end system that can be used to extract entities and the relations between them.

3. Materials and Methods

This section describes the dataset and our method in detail and provides an overview of the pipelined approach to recognizing drug named entities and classifying their interactions in pairs. The data preparation, training, tuning and evaluation phase, are also described as shown in Figure 3. The pipeline consists of two distinct classification models for Drug Named Entity Recognition and Relation Classification. We apply a small set of hand-crafted rules to filter the output of the DNER model before passing it as input to the RC model and finally classifying drug entity pairs in a sentence. In both phases, the only preprocessing step consists of tokenization and no additional features are generated. Our models are trained, evaluated and tested with the published task-specific dataset and hyperparameter tuning was performed on the evaluation dataset. The process is described in detail in the following sections.

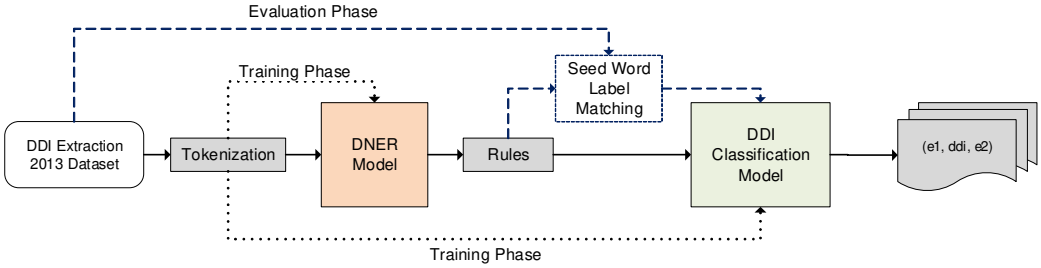


Figure 2: The TP-DDI Pipeline.

3.1. Dataset

The DDI Extraction 2013 corpus [4] is a semantically annotated corpus of documents that consists of sentences describing drug entities and drug-drug interactions from the DrugBank database and MedLine abstracts. DrugBank consists of manually curated texts that combine detailed drug data with comprehensive drug target information, while MedLine is a bibliographic database that contains biomedical publications. It has been manually annotated with pharmacological substances (drug named entities) and the interaction for all possible drug pair combinations. It serves as the gold standard dataset for the DDI extraction task and it has been annotated and reviewed by two expert annotators.

The DDI corpus is comprised of 784 documents describing drug interactions from the DrugBank database and 233 MedLine abstracts selected from the query 'drug-drug interactions'. A summary of the main features and statistics of the corpus is presented in Table 1. The target of the task of drug-drug interactions recognition in biomedical literature is to determine whether there is a relationship between two candidate drug entities in a given sentence, as defined in SemEval-2013 Shared Task [5].

The entities for the recognition of the drug named entities in the corpus are annotated with the types *drug, group, brand* and *drug_n*. Generic drug names are defined as *drug*, while branded drug names and drug group names as *brand* and *group* respectively. Finally, active substances that are not approved for human use are defined as *drug_n*.

The drug-drug interactions for the classification task in the corpus are annotated with the following types:

- advice: Advice is the type that is designated to the drug-drug interactions within which a recommendation or advice relating to the con-

comitant use of a pair of drugs involved in them is described.

- effect: Effect is assigned when the impact of the drug-drug interaction is described. It can be a pharmacological effect, signs or symptoms, a clinical finding, an unspecified alternation of the effect or action of a single drug, a rise of the toxicity or protective effect, or therapeutic failure. Furthermore, this type is assigned when a pharmacodynamic mechanism or effect of interaction is described in the sentence.

- mechanism: The type mechanism is assigned when a pharmacokinetic mechanism is described, including changes in levels or concentration of the entities. However, this type can also be assigned when the mechanism of interaction is pharmacodynamic.

- int: This type is assigned when an interaction in the sentence occurs but does not provide any information about the type of interaction, so none of the other types can be assigned.

- false: This type is assigned when the a drug pair that occurs in the sentence has no interaction.

The dataset provides a single training set for both named entity recognition and relation classification tasks and test sets specific to each of the two tasks. The sets are comprised of instances separated in documents that contain paragraphs separated by sentences. For each sentence in the dataset, the drug entities and their types, the drug entity off-sets and the drug pairs along with their interaction are annotated. The drug entity off-sets contain the starting and ending position of each entity in the corresponding sentence. Sentences that contain more than one drug pair have all possible drug pairs annotated, leading to multiple instances with the corresponding interaction from a single sentence.

Furthermore, the dataset is extremely unbalanced in relation to both named entities and interaction types. For the drug entities, 64% of the instances belong to the type *drug* and 23% to the type *group*, while the types *brand* and *drug_n* constitute only 10% and 3% of the instances respectively. For the interaction types, which include *"advice"*, *"mechanism"*, *"effect"*, *"int"* and *"none"* or *"false"*, 85% of the instances are negative and the remaining 15% positive. Moreover, the distribution of each type in the positive samples is unbalanced, where the number of instances for the type *"int"* is remarkably less than the other types.

		Training Set		DNER Test Set		RC Test Set	
		DB	ML	DB	ML	DB	ML
	Documents	572	142	54	58	158	33
	Sentences	5675	1301	145	520	973	326
Entity Types	drug	8197	1228	180	171	1518	346
	group	3206	193	65	90	626	41
	brand	1423	14	53	6	347	22
	drug_n	103	401	5	115	21	119
DDIs	mechanism	1260	62	-	-	279	24
	effect	1548	152	-	-	301	62
	advice	819	8	-	-	215	7
	int	178	10	-	-	94	2
	false/none	22216	1555			4381	356

Table 1: DDI corpus statistics

3.2. BERT Language Model

BERT [27] (Bidirectional Encoder Representations from Transformers), is a context-sensitive word representation model that utilizes bidirectionality in Transformers [28] trained on large-scale unsupervised corpora to obtain contextualized representations of each word in a sentence. Since BERT aims to generate a language representation, only the encoder part of the Transformers is used and comes in two architectures, BERT-base and BERT-large with 12 layers and 24 layers in the encoder stack respectively. Previous Language Models (LM) incorporated unidirectional LMs (left-to-right or right-to-left), while BERT uses a Masked Language Model (MLM) that predicts randomly masked words in a sequence, making it able to learn bidirectional representations. According to the authors of BERT, incorporating information from bidirectional representations, rather than single directional representations, is crucial for representing words in natural language.

As a general purpose LM, BERT is pre-trained on large-scale corpora from the English Wikipedia and BooksCorpus. However, biomedical domain-specific literature contains a substantial number of domain-specific proper nouns and terms that are not present in general purpose corpora. As a result, NLP models designed for general purpose language understanding often obtain poor performance in biomedical information retrieval tasks.

To alleviate this, in this work, we use BioBERT [15], which is a biomedical domain-specific Language Representation Model (LRM) based on the

BERT LM and pre-trained on large-scale biomedical corpora from PubMed abstracts (PubMed) and PubMed Central full-text articles (PMC). According to the authors of BioBERT, bidirectional representations are also critical in biomedical text mining since complex relationships between biomedical terms often exist in a biomedical corpus [29].

BERT and BioBERT obtain state-of-the-art performance on most general purpose and biomedical domain-specific NLP tasks respectively, while requiring minimal task-specific architectural modification. The basic structure of BERT and by extension BioBERT is comprised of self-attention encoders (SA-encoder) that obtain the corresponding context-specific representations using the sequence and the mask matrix. The downstream task layer enables the model to fine-tune to the task specific output.

In this work, we focus on the Drug Named Entity Recognition and Relation Classification tasks for extracting DDIs, which are further detailed in subsections 3.4 and 3.6, respectively.

3.3. TP-DDI pipeline

We present a DDI extraction pipeline consisting of two separate models, the DNER sequence-to-sequence model and the DDI classification model as a three-step process. Initially, the sentence sequences pass through the DNER model, classifying each token based on predefined labels. In the second step, the output passes through a set of rules to filter out the instances where no interaction between drug pairs is present and preemptively correct mislabeled tokens. Finally, the previous output is fed to the DDI classifier, categorizing the interaction between the drug pairs in each sentence. The output of the pipeline is an entity-interaction-entity triple.

For both classification tasks (DNER and DDI), the BERT architecture was used and additional layers were introduced to fine-tune each model for each corresponding task. Furthermore, to reduce the complexity in feature engineering and preprocessing, our approach consists of tokenization only. Therefore, we utilize WordPiece (WP) tokenization [30] which represents out-of-vocabulary words by frequent subwords (e.g. tetrahydropyridine - te ##tra ##hy ##dr ##opy ##rid ##ine). It was observed that using cased vocabulary results in slightly better performance in downstream tasks as presented in this work. In both models, the original vocabulary of BERT-base was used. This approach allows the interchangeable use of existing models based on both BERT and BioBERT.

Due to the compatibility of BioBERT with BERT, pretrained weights on general domain corpora can be re-used. Furthermore, existing models based on BERT and BioBERT can be interchangeably used and any out-of-vocabulary words can still be represented and fine-tuned for the biomedical domain using the original WordPiece vocabulary of BERT.

3.4. Drug Named Entity Recognition Model

The Drug Named Entity Recognition, which is a sequence-to-sequence task, is tackled with the first model in our pipeline. The model follows the BERT architecture and a dense layer is added to tag each token based on the tagging scheme. The tagging scheme follows Inside-Outside-Before (IOB) format for each entity type (i.e. I-DRUG_N, B-GROUP) when the type of drug is important. This format introduces less complexity to the DNER model and consequently to the overall pipeline compared to BIOES/BILOU formats which offer marginal improvement in performance. For split word tokens we assign an X annotation to all subsequent subwords and mirror the same action to the corresponding labels, maintaining the alignment between the tokens and their respective labels. Alternatively, we omit the type information from the IOB format from all drug named entities, treating them as a single type.

Initially, the input sequence passes through WP tokenization. Afterwards, a "[CLS]" and "[SEP]" token is added to the input of the word tokens at the beginning and the end of each sentence, respectively. Segment embeddings, to allow the encoder to distinguish between sentences, are added and finally position embeddings are also added to each token to indicate the position in the sentence. Fundamentally, the encoder stack maps sequences to sequences resulting in the output consisting of a sequence of the same size as the input vector. Finally, the output is a dense layer with $softmax$ classification where each token is a assigned a probability for each class. A detailed overview of the model architecture is shown in Figure 3.

3.5. Rules

A set of rules that is applied to all instances, filters the output of the DNER model before passing it as input to the DDI classification model. The output of the DNER model consists of line-separated word tokens represented by frequent subwords and a separate label set with the entity types for each word token. Depending on the phase (i.e. training/testing phase,

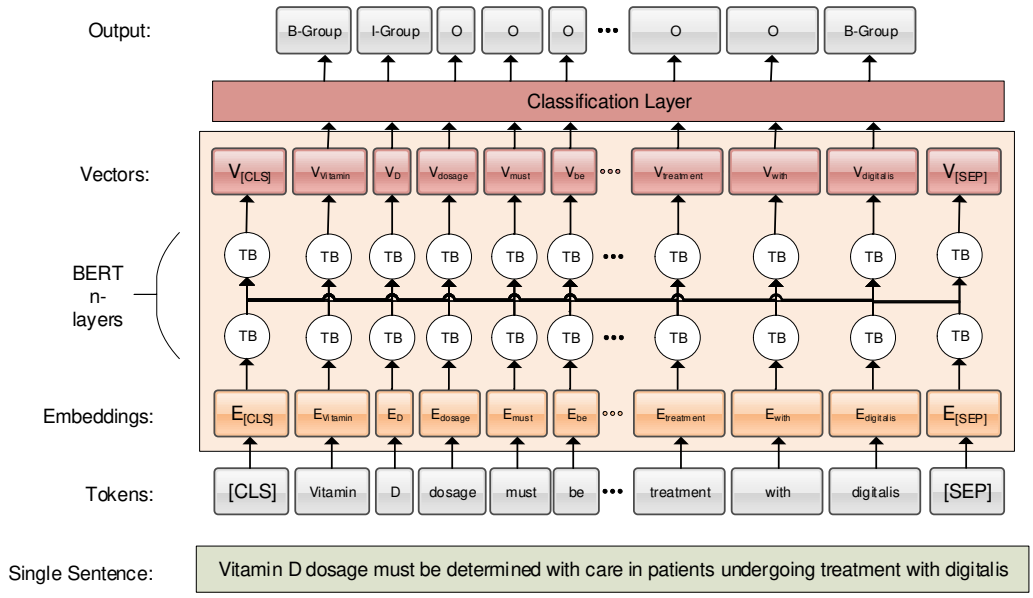


Figure 3: Drug Named Recognition model architecture. TB denotes the BERT Transformer Blocks.

inference) the sentences either pass through all transformations and filters or only through a subset of them.

Initially, the DNER model outputs get inverse transformed to build the sentences. To achieve this, we applied regular expressions to recreate each sentence based on on the "[CLS]" and "[SEP]" tokens. In the second step, the frequent subwords get merged and each word in each sentence is joined to form a whitespace-separated string. Additionally, tokens labeled with the X annotation get replaced with the appropriate label. Special cases where the WP tokenizer created tokens from special character ("!", "?", "/", etc.), are combined with the word token they belong to. Every transformation applied to the word tokens is mirrored to the label list to keep the correct indices for each word and label. Afterwards, by matching the joint sentences with the inferred labels from the DNER model, the entity type information is mapped to each word in each sentence.

Despite the satisfying performance of the DNER model, a perfect prediction is not possible. To preemptively correct possible mislabeled tags, we heuristically look for instances where the model misclassified a word token in the sequence. This validation process is executed efficiently as it takes place

within one iteration of the data structure containing the complete sequence list, resulting in a time complexity of $O(n)$. For example, in instances where a word token is labeled as Before (B) ahead of a Inside (I) of different entity types (i.e. B-DRUG, I-GROUP), the type information of the tag I is replaced with the prior tag of B. Afterwards, the sentences and the entity types get parsed and multi-word entities are treated as a single whitespace-separated entity. Instances where only one drug entity is present in the sentence, get discarded. The most commonly used rules are the ones responsible for decoding the subword tokens, replacing the X annotation with the appropriate label, while the label-correcting rules are applied to approximately 3% of all instances, reflecting the DNER models' performance. The rules are presented in Table 2.

Rule	**Outcome**
if token.startwith('X') AND prev_tag='B'	Replace 'X' with 'I' if it is a subword token and previous tag is 'B'
if token.startwith('X') AND prev_tag='I'	Replace 'X' with 'I' until the tag changes
if prev_tag='B' AND cur_tag='I' AND prev_label NOT cur_label	Correct the label of 'I' with the label of the previous tag
if prev_tag='I' AND cur_tag='I' AND prev_tag NOT cur_tag	Correct the label of 'I' with the label of the previous tag

Table 2: Summary of rules

To prepare the data for the DDI classification model, all possible combinations of the recognized drug entities are created. Sentences that contain more than one drug pair (i.e. more than two drug entities), have all possible drug pairs combinations generated, leading to multiple instances with the corresponding interaction from a single sentence. Finally, the special tokens "$< e1 >$", "$< /e1 >$" and "$< e2 >$", "$< /e2 >$" are inserted in front and at the end of each entity of the target entity pair for each sentence.

3.6. Drug-Drug Interaction Classification model

The DDI classification, which is a Relation Classification task, is tackled in the second and final model in our pipeline. The model is based on S. Wu et al. [31], where entity information is used to enrich the pre-trained BERT model. As shown in Figure 4, for each sentence that contains two target entities $e1$ and $e2$, the special tokens "$< e1 >$", "$< /e1 >$" and

"$< e2 >$", "$< /e2 >$" at both the beginning and end of each entity are inserted, in order to make BERT capture the location information of the two entities. Similarly to the DNER models step, "[CLS]" and "[SEP]" tokens are added to the beginning and end of each sentence accordingly. For example, after inserting the special entity separation tokens, for a sentence with target entities *"Fenfluramine"* and *"guanethidine"* the text will be converted to: *"[CLS] "$< e1 >$" Fenfluramine "$< /e1 >$" may increase slightly the effect of antihypertensive drugs, e.g., "$< e2 >$" guanethidine "$< /e2 >$", methyldopa, reserpine. [SEP]"*

Given a sentence with a target entity pair, the final hidden layer output of the model before classification is the concatenated output of three hidden layers. The first hidden layer consists of a fully connected layer with a *tanh* activation function for the first token (i.e. "[CLS]") in the sentence. The vectors for each entity are calculated by averaging the hidden state vectors representing each entity $e1$ and $e2$ accordingly. Afterwards, the resulting entity vectors are passed through a fully connected layer with *tanh* activation. The final hidden state output of the the DDI classification model consists of the outputs of the vectors for the first token and both entities $e1$ and $e2$. A $softmax$ activation function is applied to the final output, generating a probability for the sentence to belong to each class.

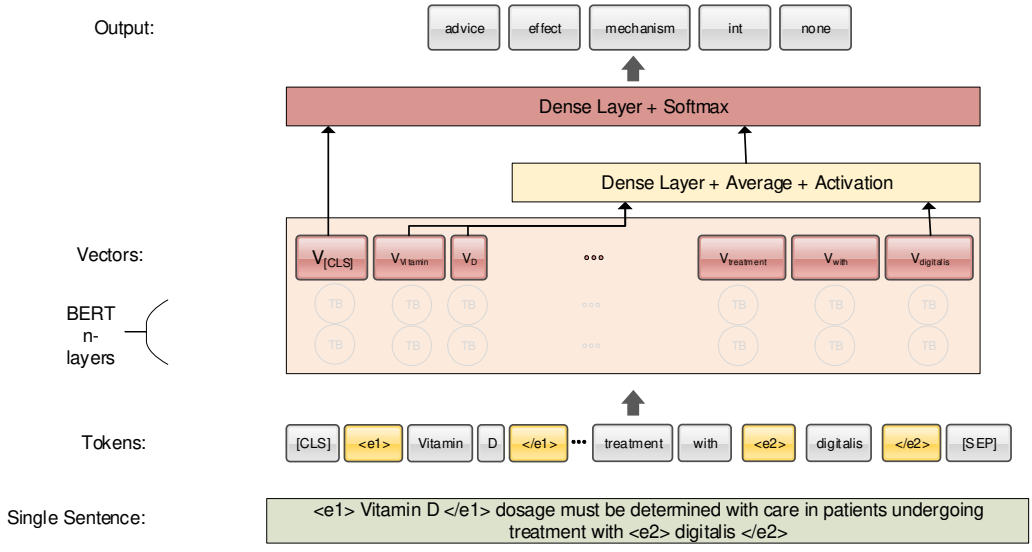


Figure 4: DDI classification model architecture.

4. Results and discussion

4.1. Experimental setup

In order to evaluate the performance of our pipelined approach, we used the F1-score on the DDI Extraction 2013 dataset as described in Section 3.1. The F1-score metric is the harmonic mean of the Precision and Recall metrics, where Precision is the ratio of correctly predicted positive observations to the total predicted positive observations and Recall the ratio of correctly predicted positive observations to all observations in the actual class. The contribution of all classes are aggregated to calculate the average score as it was used in the SemEval DDI Extraction challenge and in related studies.

We trained both Drug Named Entity Recognition and Drug-Drug Interaction classification models on the 6976 sentences contained in 714 abstract documents from both DrubBank and MedLine using predefined train and test splits. We used a subset of the train set to create the evaluation sets and separate test sets for each model in the pipeline. We evaluated the DNER model with a test set consisting of 665 sentences contained in 112 abstract documents and the DDI classification model on 1299 sentences in 181 abstract documents.

We experimented with pre-trained weights from BERT and BioBERT for both BERT-base (base) and BERT-large (large) architectures. The pre-trained weights used in BERT are trained on English Wikipedia and BooksCorpus and in BioBERT from PubMed and PubMed Central. In the base architecture, 12 encoder layers are stacked with a hidden layer size of 768, while in the large architecture, the encoder stack is comprised of 24 encoder layers with a hidden layer size of 1024. Sentences were padded to the maximum sentence length in the training set. AdamW [32] was used as an adaptive optimizer that decouples the weight decay from the optimization step, allowing for separate optimization. A learning rate of 0.001 and decay of 0.01 per epoch were applied and the models were trained for 4 epochs.

Both DNER and DDI models were individually trained using the gold standard dataset. For the evaluation of the pipeline, we seeded the labels from the dataset and assigned the labels to each sentence from the DNER output, according to the sentence text and target drug entity pair.

The experiments were conducted on a computer with a single Volta V100 16GB graphics card and a 40-core Intel CPU and we implemented our TP-DDI model with the Tensorflow library and the Python programming language.

4.2. Experimental results

To investigate the contributing factors in improving the performance of our proposed approach, we extended the experiments to the tagging scheme used for training the DNER model. Initially, we used the IOB-scheme for each drug entity type (drug, drug_n, group, brand). The primary approach was to retain the drug entity type information during the DDI classification process. Since the classification of the interactions between drug entities is of importance, the type of drug is not relevant to this task. Therefore, we reduced the complexity of the drug entity names by discarding any additional drug entity types (EntA) and labeling all types as drugs for the DNER task, since interactions can occur in all cases. Consequently, by omitting the drug entity types from the IOB tags, the total amount of labels and the models complexity is reduced. Similarly, for the DDI classification task, we experimented with replacing all drug names in each target drug entity pair with 'drug_a' and 'drug_b' and every other drug mention present in the sentence with 'drug_n' (TDEM). For each drug pair we encounter in a sentence, the sentence is reproduced, changing which entities are annotated with 'drug_a', 'drug_b' and 'drug_n' to consider all possible combinations in the classification process. This approach leads to a drug name-agnostic model and is applied to the best performing approach.

	English Wiki and BookCorpus		PubMed v1.1	
TP-DDI Pipeline	NER F1	DDI F1	NER F1	DDI F1
Base	0.743	0.723	0.959	0.816
+ EntA	0.792	0.719	**0.961**	**0.817**
+ EntA + TDEM	0.792	0.719	0.961	0.817
Large	0.897	0.764	0.969	0.821
+ EntA	0.916	0.778	**0.971**	**0.824**
+ EntA + TDEM	0.916	0.778	0.971	0.824

Table 3: Comparison of proposed pipeline with different architectures and pre-trained weights. "EntA" denotes drug entity types consolidation to a single class for the DNER task, removing the drug entity type from the IOB tags. "TDEM" denotes the masking of target drug names in the DDI classification task by replacing the drug mentions with 'drug_a', 'drug_b' and 'drug_n'.

As shown in Table 3, the pipeline with pre-trained weights from the biomedical domain in combination with target drug name masking achieves the best performance in the overall DDI extraction task. Additionally, the

larger architecture that uses twice the amount of encoder stacks and larger hidden layer sizes, achieves marginally better results than the base architecture. However, the training and inference time are greatly increased. Training times for each DNER and DDI model vary between 6 to 7 hours for each base model and from 32 to 36 hours for each large model. An analogous increase is observed for inference as well, varying from 1 to 2 minutes and 8 to 12 minutes, respectively. Furthermore, the added complexity of the large BERT architecture introduces difficulties in GPU memory management.

With the implementation of the special entity separation tokens in the DDI classification task, the models are able to focus on the target drug pair in order to create better hidden representations. Therefore and most notably, replacing the the drug names for the DDI classification task yielded the same results in all cases while adding complexity to the preprocessing step of the pipeline. The experimental results show that both base and large architectures achieve the best results by reducing the drug entity types to a single class.

4.3. Performance comparison

State-of-the-art systems use varying techniques for this task, ranging from negative instance filtering to converting the task to a tagging problem. Furthermore, extensive feature engineering is required to achieve their reported performances. Therefore, in order to evaluate the effectiveness of our DDI extraction pipeline, we compare the performance to the baseline approaches presented in [25, 33, 34] and to similar pipelined and joint approaches, where drug names are recognized and subsequently their interactions classified.

We compare our pipeline to "BiLSTM", "SCNN" [33] and "HRNN" [34], which are the baseline relation extraction methods on golden standard entities where only DDI relationships get classified. The "BiLSTM" model, as implemented by Luo et al. [25], acts as baseline and uses word and entity position embeddings as inputs only. Additionally, we compare to "BiLSTM-CRF + HRNN", which is the previous state-of-the-art pipelined method and finally to "Att-BilSTM-CRF + Elmo" which is the previous state-of-the-art joint method.

The experimental results in Table 4 show that our end-to-end method, labeled TP-DDI achieved the best overall performance and the best individual task performance for DNER and DDI respectively. Our set of rules in combination with the fine-tuned BioBERT architectures and pre-trained

weights are the main contributing factors. Our rules can capture the multi-token entities and filter instances that do not contain relationships between drug entities while the Transformer-based mechanisms are able to capture long distance dependencies and prioritize important words for the pipelines' predictions. The use of the large BioBERT architecture increases the performance by a slight margin (approximately 1%) in both DNER and DDI tasks. However, the increased complexity introduced to the overall system, by doubling the encoder layers from 12 to 24, increases the training and inference time considerably.

Additionally, compared to the existing approaches, both pipelined and joint, our method removes the dependence on any kind of feature engineering and complex preprocessing. Furthermore, error propagation is mitigated due to the excellent performance of the DNER model and our rules that capture possible misclassified drug entities.

	DNER			DDI (RE)		
Method	P	R	F1	P	R	F1
BiLSTM	-	-	-	0.684	0.665	0.674
SCNN	-	-		0.722	0.651	0.686
HRNN	-	-	-	0.741	0.718	0.729
BiLSTM-CRF + HRNN	0.932	0.861	0.895	0.692	0.707	0.692
Att-BilSTM-CRF + Elmo	0.905	0.939	0.922	0.750	0.752	0.751
TP-DDI (base)	0.954	0.967	**0.961**	0.859	0.779	**0.817**
TP-DDI (large)	0.974	0.968	**0.971**	0.864	0.788	**0.824**

Table 4: Performance comparison

4.4. Case study

To analyze the advantages and disadvantages of our method, we compare a few prediction results of the TP-DDI (base) method with the gold standard labels from the DDI-Extraction 2013 dataset as shown in A.5.

For sentence 1, the sentence describes a recommendation towards the co-administration of two drugs. Our method identifies the drug entities and each respective label correctly. Similarly, all possible drug pair combinations are generated and labeled correctly, consequently resulting in a correct extraction process.

For sentence 2, the sentence indicates that the simultaneous use of drugs belonging to two drug groups may cause specific symptoms. This example

contains a combination of multi-token entities, initialisms, i.e. abbreviations of a word to its initials, as well as a drug entity that is specific to an aforementioned brand. Although our method is able to identify all drug entities and the labels for each entity correctly, the possible drug pair combinations generated for classification exceed the pairs in the gold dataset. The main reason is that our method is unable to identify which entities are initialisms or acronyms of previous mentions in the sentence and creates pairs that include these initialsms with their full form. However, our method extracts all the true drug pairs and classifies them with the correct labels.

For sentence 3, the sentence describes an antagonistic effect between two drugs and includes a recommendation to counter this effect. Our method manages to correctly identify the drug entities and their labels and generates the correct pair combinations found in the sentence. However, the classification model mistakenly classifies the interaction type as advice instead of effect.

Conclusively, our method is capable of extracting drug entity mentions and their relations, but needs further improvement in determining the type of interaction between the drug mentions. Furthermore, both DNER and DDI models are limited by the pre-trained (PubMed) and the DDI Extraction 2013 datasets. This may not generalize well outside the biomedical and more specific the drug domain. Finally, our method, similarly to related works, is unable to extract relations between entities across sentences as this limitation is introduced by the DDI extraction task and the provided dataset.

5. Conclusion

In this paper, we propose an end-to-end Entity Recognition and Relationship Classification pipeline for the extraction of Drug-Drug Interactions from biomedical literature, that achieves state-of-the-art performance. Specifically, we presented a Drug Named Entity Recognition (DNER) model to extract drug named entities, rules that are applied to the DNER models output and a Drug-Drug Interaction (DDI) Classification model to classify the interactions between the target drug pairs. Initially, drug named entities are extracted from biomedical texts in a sequence-to-sequence classification task. Then a set of rules and filters are applied before finally classifying the interactions between drug pairs.

Mitigation of error propagation is achieved by preemptively correcting possible mislabeled tags while filtering instances that do not contain drug

pairs. Furthermore, by avoiding the use of feature engineering and tedious preprocessing, we reduce the overall complexity of the pipelined approach to extracting DDIs. The partitioning of the overall task into two separate subtasks allows for easier modification of each part in the pipeline. Consequently, both DNER and DDI classification models could be fine-tuned to identify adverse event mentions and classify them respectively, leveraging the underlying BERT architecture. With this contribution, we aim to aid in the drug development process as well as in the identification of possible adverse drug event due to simultaneous use of more than one drug.

Conflict of interest statement

The authors have no conflicts of interest to declare.

References

[1] Y. Wang, L. Wang, M. Rastegar-Mojarad, S. Moon, F. Shen, N. Afzal, S. Liu, Y. Zeng, S. Mehrabi, S. Sohn, H. Liu, Clinical information extraction applications: A literature review, Journal of Biomedical Informatics 77 (June 2017) (2018) 34–49. doi:10.1016/j.jbi.2017.11.011.
URL https://doi.org/10.1016/j.jbi.2017.11.011

[2] A. L. Hincapie, T. L. Warholak, L. E. Hines, A. M. Taylor, D. C. Malone, Impact of a drug-drug interaction intervention on pharmacy and medical students' knowledge and attitudes: A 1-year follow-up, Research in Social and Administrative Pharmacy 8 (5) (2012) 472–477. doi:10.1016/j.sapharm.2011.11.003.
URL https://linkinghub.elsevier.com/retrieve/pii/S1551741111001331

[3] A. J. Grizzle, J. Horn, C. Collins, J. Schneider, D. C. Malone, B. Stottlemyer, R. D. Boyce, Identifying Common Methods Used by Drug Interaction Experts for Finding Evidence About Potential Drug-Drug Interactions: Web-Based Survey., Journal of medical Internet research 21 (1) (2019) e11182. doi:10.2196/11182.
URL http://www.ncbi.nlm.nih.gov/pubmed/30609981http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC6682289

[4] M. Herrero-Zazo, I. Segura-Bedmar, P. Martínez, T. Declerck, The DDI corpus: An annotated corpus with pharmacological substances and drug-drug interactions, Journal of Biomedical Informatics 46 (5) (2013) 914–920. doi:10.1016/j.jbi.2013.07.011.

[5] I. Segura-Bedmar, P. Martínez, M. Herrero-Zazo, SemEval-2013 task 9 : Extraction of drug-drug interactions from biomedical texts (DDIExtraction 2013), in: Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013), Association for Computational Linguistics, Atlanta, Georgia, USA, 2013, pp. 341–350.
URL https://www.aclweb.org/anthology/S13-2056

[6] S. Liu, B. Tang, Q. Chen, X. Wang, Drug name recognition: Approaches and resources, Information (Switzerland) 6 (4) (2015) 790–810. doi:10.3390/info6040790.

[7] D. Zaikis, I. Vlahavas, Drug-Drug Interaction Classification Using Attention Based Neural Networks, in: 11th Hellenic Conference on Artificial Intelligence, ACM, New York, NY, USA, 2020, pp. 34–40. doi:10.1145/3411408.3411461.
URL https://dl.acm.org/doi/10.1145/3411408.3411461

[8] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, J. Davison, S. Shleifer, P. von Platen, C. Ma, Y. Jernite, J. Plu, C. Xu, T. Le Scao, S. Gugger, M. Drame, Q. Lhoest, A. Rush, Transformers: State-of-the-art natural language processing, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, Association for Computational Linguistics, Online, 2020, pp. 38–45. doi:10.18653/v1/2020.emnlp-demos.6.
URL https://www.aclweb.org/anthology/2020.emnlp-demos.6

[9] N. Stylianou, I. Vlahavas, Transformed: End-to-end transformers for evidence-based medicine and argument mining in medical literature, Journal of Biomedical Informatics 117 (2021) 103767. doi:https://doi.org/10.1016/j.jbi.2021.103767.
URL https://www.sciencedirect.com/science/article/pii/S1532046421000964

[10] S. K. Sahu, A. Anand, Drug-drug interaction extraction from biomedical texts using long short-term memory network, Journal of Biomedical Informatics 86 (2018) 15–24. doi:10.1016/j.jbi.2018.08.005.
URL https://linkinghub.elsevier.com/retrieve/pii/S1532046418301606

[11] I. S. Bedmar, P. Martinez, M. H. Zazo, Extraction of Drug-Drug Interactions from BioMedical Texts, https://www.cs.york.ac.uk/semeval-2013/task9/ (2013).

[12] E. F. Tjong Kim Sang, F. De Meulder, Introduction to the CoNLL-2003 shared task, in: Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003 -, Vol. 4, Association for Computational Linguistics, Morristown, NJ, USA, 2003, pp. 142–147. doi:10.3115/1119176.1119195.
URL http://portal.acm.org/citation.cfm?doid=1119176.1119195

[13] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding (oct 2018). arXiv:1810.04805.
URL http://arxiv.org/abs/1810.04805

[14] Y. Liu, F. Meng, J. Zhang, J. Xu, Y. Chen, J. Zhou, GCDT: A Global Context Enhanced Deep Transition Architecture for Sequence Labeling (jun 2019). arXiv:1906.02437.
URL http://arxiv.org/abs/1906.02437

[15] J. Lee, W. Yoon, S. Kim, D. Kim, S. Kim, C. H. So, J. Kang, BioBERT: a pre-trained biomedical language representation model for biomedical text mining, Bioinformatics (sep 2019). doi:10.1093/bioinformatics/btz682.
URL https://academic.oup.com/bioinformatics/advance-article/doi/10.1093/bioinformatics/btz682/5566506

[16] W. Yoon, C. H. So, J. Lee, J. Kang, CollaboNet: collaboration of deep neural networks for biomedical named entity recognition, BMC Bioinformatics 20 (S10) (2019) 249. doi:10.1186/s12859-019-2813-6.
URL https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-019-2813-6

[17] S. Pyysalo, F. Ginter, J. Heimonen, J. Björne, J. Boberg, J. Järvinen, T. Salakoski, BioInfer: a corpus for information extraction in the biomedical domain, BMC Bioinformatics 8 (1) (2007) 50. doi:10.1186/1471-2105-8-50.
URL https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-8-50

[18] C. Quan, M. Wang, F. Ren, An Unsupervised Text Mining Method for Relation Extraction from Biomedical Literature, PLoS ONE 9 (7) (2014) e102039. doi:10.1371/journal.pone.0102039.
URL https://dx.plos.org/10.1371/journal.pone.0102039

[19] V. Suárez-Paniagua, R. M. Rivera Zavala, I. Segura-Bedmar, P. Martínez, A two-stage deep learning approach for extracting entities and relationships from medical texts, Journal of Biomedical Informatics 99 (2019) 103285. doi:10.1016/j.jbi.2019.103285.
URL https://linkinghub.elsevier.com/retrieve/pii/S1532046419302047

[20] S. Kim, H. Liu, L. Yeganova, W. J. Wilbur, Extracting drug–drug interactions from literature using a rich feature-based linear kernel approach, Journal of Biomedical Informatics 55 (2015) 23–30. doi:10.1016/j.jbi.2015.03.002.
URL https://linkinghub.elsevier.com/retrieve/pii/S1532046415000441

[21] A. Ben Abacha, M. F. M. Chowdhury, A. Karanasiou, Y. Mrabet, A. Lavelli, P. Zweigenbaum, Text mining for pharmacovigilance: Using machine learning for drug name recognition and drug–drug interaction extraction and classification, Journal of Biomedical Informatics 58 (2015) 122–132. doi:10.1016/j.jbi.2015.09.015.
URL https://linkinghub.elsevier.com/retrieve/pii/S1532046415002099

[22] S. K. Sahu, A. Anand, What matters in a transferable neural network model for relation classification in the biomedical domain?, Artificial Intelligence in Medicine 87 (2018) 60–66. doi:10.1016/j.artmed.2018.03.006.
URL https://linkinghub.elsevier.com/retrieve/pii/S0933365717304050

[23] D. Zhou, L. Miao, Y. He, Position-aware deep multi-task learning for drug–drug interaction extraction, Artificial Intelligence in Medicine 87 (2018) 1–8. doi:10.1016/j.artmed.2018.03.001.
URL https://linkinghub.elsevier.com/retrieve/pii/S0933365717306310

[24] Y. Zhao, H. Wan, J. Gao, L. Youfang, Improving Relation Classification by Entity Pair Graph, Proceedings of The Eleventh Asian Conference on Machine Learning (2017) (2019) 1156–1171.
URL http://proceedings.mlr.press/v101/zhao19a.html

[25] L. Luo, Z. Yang, M. Cao, L. Wang, Y. Zhang, H. Lin, A neural network-based joint learning approach for biomedical entity and relation extraction from biomedical literature, Journal of Biomedical Informatics 103 (August 2019) (2020) 103384. doi:10.1016/j.jbi.2020.103384.
URL https://doi.org/10.1016/j.jbi.2020.103384

[26] X. Ma, E. Hovy, End-to-end Sequence Labeling via Bi-directional LSTM-CNNs-CRF (mar 2016). arXiv:1603.01354.
URL http://arxiv.org/abs/1603.01354

[27] J. Devlin, M. W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference 1 (Mlm) (2019) 4171–4186. arXiv:1810.04805.
URL https://arxiv.org/pdf/1810.04805.pdf

[28] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, in: Advances in neural information processing systems, 2017, pp. 5998–6008.

[29] M. Krallinger, O. Rabal, S. A. Akhondi, M. Pérez Pérez, J. Santamaría, G. Pérez Rodríguez, G. Tsatsaronis, A. Intxaurrondo, J. Antonio López, U. Nandal, E. Van Buel, A. Chandrasekhar, M. Rodenburg, A. Laegreid, M. Doornenbal, J. Oyarzabal, A. Lourenço, A. Valencia, Overview of the BioCreative VI chemical-protein interaction Track, Proceedings of BioCreative VI workshop 450 (9) (2017) 141–146.

[30] Y. Wu, M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey, J. Klingner, A. Shah, M. Johnson, X. Liu, Ł. Kaiser, S. Gouws, Y. Kato, T. Kudo, H. Kazawa, K. Stevens, G. Kurian, N. Patil, W. Wang, C. Young, J. Smith, J. Riesa, A. Rudnick, O. Vinyals, G. Corrado, M. Hughes, J. Dean, Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation (2016) 1–23arXiv:1609.08144.
URL http://arxiv.org/abs/1609.08144

[31] S. Wu, Y. He, Enriching Pre-trained Language Model with Entity Information for Relation Classification, in: Proceedings of the 28th ACM International Conference on Information and Knowledge Management, ACM, New York, NY, USA, 2019, pp. 2361–2364. doi:10.1145/3357384.3358119.
URL https://dl.acm.org/doi/10.1145/3357384.3358119

[32] I. Loshchilov, F. Hutter, Decoupled weight decay regularization, in: International Conference on Learning Representations, 2019.
URL https://openreview.net/forum?id=Bkg6RiCqY7

[33] Z. Zhao, Z. Yang, L. Luo, H. Lin, J. Wang, Drug drug interaction extraction from biomedical literature using syntax convolutional neural network, Bioinformatics (2016) btw486doi:10.1093/bioinformatics/btw486.
URL https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btw486

[34] Y. Zhang, W. Zheng, H. Lin, J. Wang, Z. Yang, M. Dumontier, Drug–drug interaction extraction via hierarchical RNNs on sequence and shortest dependency paths, Bioinformatics 34 (5) (2018) 828–835. doi:10.1093/bioinformatics/btx659.
URL https://academic.oup.com/bioinformatics/article/34/5/828/4565590

Appendix A. TP-DDI prediction examples

We manually selected three characteristic examples from the DDI-Extraction 2013 corpus and compare the gold labels with our methods predictions. These examples highlight both the cases where our proposed model performs well and and where it fails. Each gold entity is marked in bold and the type of drug is denoted in subscript for the entity contained in brackets. The underlined text denotes the wrong extraction results.

Sentence 1	[**Isocarboxazid**]$_{\text{drug}}$ should be administered with caution to patients receiving [**Antabuse**]$_{\text{brand}}$ ([**disulfiram**]$_{\text{drug}}$, Wyeth-Ayerst Laboratories)
Gold labels	Isocarboxazid, Antabuse → *advise* | Isocarboxazid, disulfiram → *advise* | Antabuse, disulfiram → *none*
TP-DDI	DNER: Isocarboxazid → *drug* | Antabuse → *brand* | disulfiram → *drug* DDI: Isocarboxazid, Antabuse → *advise* | Isocarboxazid, disulfiram → *advise* | Antabuse, disulfiram → *none*
Sentence 2	A rare, but serious, constellation of symptoms, termed serotonin syndrome, has been reported with the concomitant use of [**selective serotonin reuptake inhibitors**]$_{\text{group}}$ ([**SSRIs**]$_{\text{group}}$) and agents for migraine therapy, such as [**Imitrex**]$_{\text{brand}}$ ([**sumatriptan succinate**]$_{\text{drug}}$) and [**dihydroergotamine**]$_{\text{drug}}$.
Gold labels	selective serotonin reuptake inhibitors, Imitrex → *effect* | selective serotonin reuptake inhibitors, dihydroergotamine → *effect* | selective serotonin reuptake inhibitors, sumatriptan succinate → *effect* | SSRIs, Imitrex → *effect* | SSRIs, sumatriptan succinate → *effect* | SSRIs, dihydroergotamine → *effect* | Imitrex, dihydroergotamine → *none* | sumatriptan succinate, dihydroergotamine → *none*
TP-DDI	DNER: selective serotonin reuptake inhibitors → *group* | SSRIs: *group* | Imitrex: *brand* | sumatriptan succinate: *drug* | dihydroergotamine: *drug* DDI: <u>selective serotonin reuptake inhibitors, SSRIs → *none*</u> | selective serotonin reuptake inhibitors, Imitrex → *effect* | selective serotonin reuptake inhibitors, sumatriptan succinate → *effect* | selective serotonin reuptake inhibitors, dihydroergotamine → *effect* | SSRIs, Imitrex → *effect* | SSRIs, sumatriptan succinate → *effect* | SSRIs, dihydroergotamine → *effect* | Imitrex, sumatriptan succinate → *none* | Imitrex, dihydroergotamine → *none* | sumatriptan succinate, dihydroergotamine → *none*
Sentence 3	[**Epinephrine**]$_{\text{drug}}$ may antagonize the neuron blockade produced by [**guanethidine**]$_{\text{drug}}$ resulting in decreased antihypertensive effect and requiring increased dosage of the latter.
Gold labels	Epinephrine, guanethidine → *effect*
TP-DDI	DNER: Epinephrine → *drug* | guanethidine → *drug* DDI: Epinephrine, guanethidine → <u>*advice*</u>

Table A.5: Example predictions of the TP-DDI method.