

The University of Texas at Dallas HLTRI at TAC 2019

Ramon Maldonado, Maxwell Weinzierl, and Sanda M. Harabagiu

Human Language Technology Research Institute, The University of Texas at Dallas

{ramon, max, sanda}@hlt.utdallas.edu

Abstract. The Text Analysis Conference (TAC) Drug-Drug Interaction Extraction from Drug Labels track is an evaluation of Natural Language Processing (NLP) techniques for identifying drug-drug interactions in Food & Drug Administration (FDA) Structured Product Labeling (SPL) documents. We participated in all four tasks corresponding to: (1) entity identification; (2) sentence-level relation identification; (3) concept normalization; and (4) normalized relation identification. We developed a multi-task neural architecture based on BERT to jointly perform tasks (1), (2), and part of task (4). BERT is used to generate a contextualized representation for each word-piece token in a given sentence, which is shared between three prediction modules, the Boundary CRF (BCRF), the Relation classifier (RC), and the Pharmacokinetic Effect classifier (PKEC). Task (3) was performed using string matching on a MED-RT/UNII/SNOMED lexicon. The normalized interactions produced for Task (4) were entailed by the results of tasks (2) and (3) and the Pharmacokinetic effect codes identified by the PKEC. The results indicate that for tasks (1) and (2), a pipeline of special purpose systems out-perform joint learning. However, for task (4) – the focus of the track – the joint learning approach outperforms the pipeline.

1. Introduction

Structured Product Labeling (SPL) documents produced by the U.S. Food and Drug Administration (FDA) convey in-depth information characterizing prescription drugs. SPLs are comprised of natural language descriptions of the essential scientific information needed for the safe and effective use of a drug¹, referred to as the Labeled Drug. As such, Natural Language Processing (NLP) methods are necessary to extract information from the SPLs in order to use them in automated systems. The Text Analysis Conference (TAC) 2019 Drug-Drug Interaction (DDI) Extraction from Drug Labels track is designed by the FDA and National Library of Medicine (NLM) to address one such information extraction task: drug-drug interactions. Known drug-drug interactions, like those reported in FDA SPLs, are a preventable cause of adverse events, the eighth leading cause of death in the United States [1]. The TAC-DDI track provides a set of SPLs with manually annotated drug-drug interactions meant to facilitate the development of NLP systems that can automatically recognize such interactions.

The 2019 TAC-DDI track provides four subtasks²:

¹ <https://open.fda.gov/data/spl/>

² <https://bionlp.nlm.nih.gov/tac2019druginteractions/>

1. Entity Recognition. Extract mentions of interacting drugs/substances and their effects when combined with the Labeled Drug. This task would be more properly referred to as ‘interaction argument identification’ since only mentions of those substances and effects which participate in DDIs are to be identified.
2. Relation Identification (sentence-level). Identify interactions at the sentence level. This task is to connect mentions identified in task 1 and predict the type of interaction.
3. Normalization. Normalize each interacting substance and effect into a particular ontology of interest.
4. Normalized Relation Identification. Generate a global list of unique interactions indicated in a SPL. This task is entailed by the other three tasks by correctly normalizing each mention identified in task 1, correctly connecting each related mention in task 2, then filtering out duplicates.

There are three types of mentions indicating interaction arguments to be identified in task 1: Precipitants, Triggers, and Specific Interactions (effects). A precipitant is a substance that interacts with the Labeled Drug, as it *precipitates* the interaction. A trigger is a word or phrase that indicates an interaction event. A Specific Interaction (SI) is the result of an interaction or its effect.

There are three types of drug-drug interactions of interest in this track: Pharmacodynamic interactions, Pharmacokinetic interactions, and Unspecified Interactions. Pharmacodynamic interactions (PDI) are interactions between the Labeled Drug and a precipitant indicated by a trigger that results in a specific interaction. Consider this example from the Label for Accupril:

EX1: “Patients taking concomitant [*mTOR inhibitor*]_P (e.g., [*temsirolimus*]_P) therapy may be at [*increased risk*]_T for [*angioedema*]_{SI}.”

This sentence indicates two Pharmacodynamic interactions precipitated by the mentions “*mTOR inhibitor*” and “*temsirolimus*”, indicated by the trigger “*increased risk*” and having the effect “*angioedema*”. Pharmacokinetic interactions (PKI) occur between the Labeled Drug and a precipitant indicated by a trigger, e.g. “*The rate and extent of quinapril [absorption]_T are [diminished]_T moderately when ACCUPRIL tablets are administered during a [high-fat meal]_P.” Notice that the trigger “*absorption... diminished*” is indicated by a discontinuous span. The effects of Pharmacokinetic interactions are not explicitly annotated as their effects fall into a closed set of twenty categories defined by the National Cancer Institute Thesaurus. The effect of the PKI in the example sentence is C54356: Decreased Drug Level. Since the text indicating this effect is not explicitly annotated, PKI effects are only evaluated as part of task 4. Unspecified interactions (UI) are between the Labeled Drug and a precipitant indicated by a trigger indicating a general warning of*

risk against combining the Labeled Drug with the precipitant, e.g. “*Lithium generally [should not be given]_T with [diuretics]_P.*”

In the normalization task, each precipitant and SI mention is to be mapped into a particular ontology as follows:

- Precipitants are mapped to MED-RT NUIs if they are drug classes and UNII otherwise.
- Specific Interactions are mapped to SNOMED-CT
- PKI effects are mapped to the NCI Thesaurus

Finally, in the normalized relation identification task, the unique set of DDIs between normalized precipitants and effects involving the Labeled Drug are to be identified.

We present a pipeline performing all four tasks based on multi-task learning using massively pre-trained transformer models. The multi-task neural network jointly performs mention identification (task 1) and relation extraction (task 2) as well as pharmacokinetic interaction effect classification (task 4). On task 4, the multi-task network outperforms dedicated networks trained on tasks 1 and 2 separately, even though those networks produce superior intermediary results.

2. The Approach

In this work we present the FDA Label Drug-drug Interaction Identification Pipeline (LDIIP) for performing all four TAC-DDI tasks. LDIIP uses the Multi-task Transformer network for identifying Drug-Drug Interactions (MTTDDI) to perform tasks 1 and 2, and part of task 4. MTTDDI is a multi-task network, based on BERT [2] that uses the pretrained transformer to develop a shared multi-task representation that is fed to a series of four prediction modules: (1) the sentence classifier; (2) the mention boundary detector; (3) the relation extractor; and (4) the Pharmacokinetic Effect (PKE) classifier.

LIIP is depicted in Figure 1. LDIIP first ingests the three available training datasets via a preprocessing module. This module performs annotation propagation as in [3] and two-stage tokenization. Next, the MTTDDI model is used to extract raw mentions and relations as well as PK effects. The mentions and relations are post-processed by the Postprocessing module resulting in predicted mention spans (for task 1) and predicted relations (for task 2). The mentions are normalized into ontology codes by the Normalization module for task 3. Finally, the normalized mentions and predicted relations are unified and filtered by uniqueness to derive the label interactions for task 4.

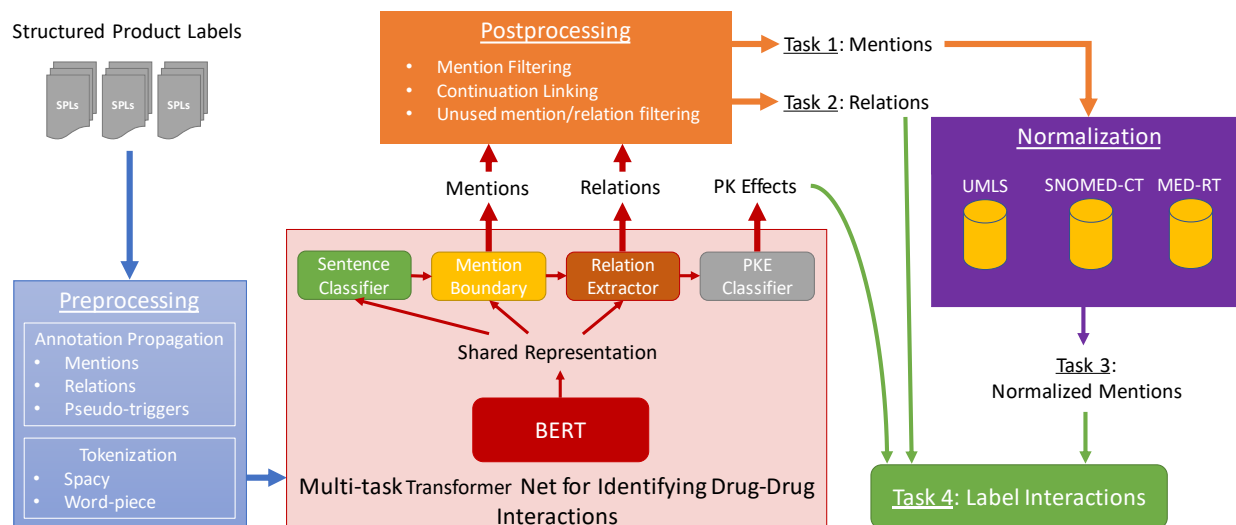


Figure 1. FDA Label Drug-drug Interaction Identification Pipeline (LDIIP)

The remainder of this section is organized as follows: in Section 2.1 we describe the preprocessing module, in Section 2.2 we describe the MTTDDI model, in Section 2.3 we describe the Postprocessing module, and finally in Section 2.4 the normalization module is described.

2.1 Preprocessing

Preprocessing began with reading mentions and sentence-level interactions from SPL files, where interactions were transformed into binary relations between triggers and either specific interactions or precipitants based on the interaction type. Trigger mentions were created for specific interactions which functioned as their own triggers in a pharmacodynamic interaction. Pharmacokinetic interaction effects treated as an attribute of the corresponding PK relation since they had no mention associated with them. Relation and mention propagation were also performed, where other matching string instances of any mentions participating in a relation within the sentence were also added as mentions with corresponding relations as in [3]. This is meant to ease the difficulty of the learning problem.

Tokenization was performed in two steps: at the word level, and at the word-piece level. First spaCy³ is used to extract word tokens from each sentence in a SPL. Word piece tokenization was performed on each sentence utilizing the token vocabulary of BERT, and word pieces were mapped back to their overlapping mentions. We adopted the C-IOBES tagging scheme for mention prediction due to the prevalence of disjoint spans in this corpus. In C-IOBES tagging, each word-piece token is assigned a tag in {O,I,B,E,S,C-I,C-B,C-E-C-S} depending on if it outside of a mention, inside of a mention, the beginning of a mention or

³ <https://spacy.io>

a single token representing a mention. The C- tags denote that a token is a part of a continuation span. C-IOBES boundary tagging was performed on each sentence, where trigger, precipitant, and specific interaction C-IOBES tags were assigned to each word piece token based on their overlapping mentions.

2.2 The Multi-Task Transformer network for identifying Drug-Drug Interactions

The Multi-Task Transformer network for identifying Drug-Drug Interactions (MTTDDI), depicted in Figure 2, is a multi-task neural network built on the pretrained BERT transformer model to perform end-to-end drug-drug interaction identification from FDA drug labels. MTTDDI consists of five modules: (1) the BERT sentence encoder; (2) the sentence classifier; (3) the mention boundary labeler; (4) the relation extractor; and (5) the PKE classifier. MTTDDI operates on the sentence level, determining if the sentence contains a drug-drug interaction using the sentence classifier. If so, MTTDDI applies the mention boundary labeler to identify the arguments of the DDI and the relation extractor to classify the type of the relation. If the relation is found to be a pharmacokinetic interaction, the PKE classifier is applied to classify the effect of the interaction from one of the pre-defined classes in the NCI Thesaurus. Each module is trained jointly (including fine-tuning the BERT model) using a linear interpolation of their loss functions.

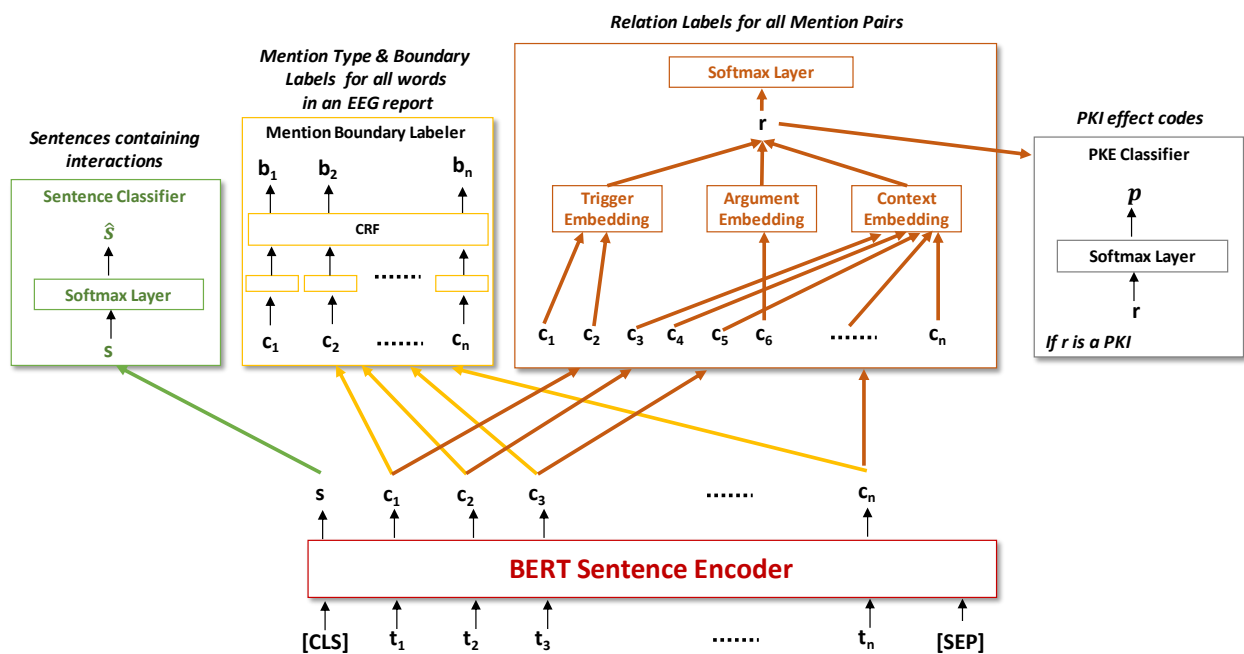


Figure 2. Multi-Task Transformer network for identifying Drug-Drug Interactions (MTTDDI).

The BERT sentence encoder applies the pre-trained transformer described in [2] to generate (1) a sentence embedding and (2) contextualized token embeddings for each token in the input sentence. The sentence embedding is passed to the sentence classifier while the contextualized token embeddings are passed to the mention boundary labeler and the relation extractor.

The input to the BERT sentence encoder is a sequence of word-piece tokens with a special leading '[CLS]' token and a trailing '[SEP]' token, as in [2]. Again as in [2], the embedding of the '[CLS]' token is used as a sentence embedding for sentence-level classification by the sentence classifier module. Formally, given a sequence of n word-piece tokens, t_1, t_2, \dots, t_n the BERT sentence encoder produces a sentence embedding, s and a sequence of contextualized token embeddings, c_1, c_2, \dots, c_n , as depicted in Figure 2.

2.2.1 Sentence Classifier

The sentence classifier module is used to determine if a given sentence contains a drug-drug interaction. While not one of the four tasks provided by TAC DDI, we found that filtering out sentences identified by the sentence classifier as not containing a DDI beneficial, experimentally. The sentence classifier module consists of a single softmax layer operating on the [CLS] embedding produced by the BERT sentence encoder. The sentence classifier is trained using sigmoid cross-entropy.

2.2.2 Mention Boundary Detector

The Mention Boundary Detector (MBD) uses a Conditional Random Field (CRF) [5] to generate the most likely C-IOBES boundary tag sequences for a sentence in order to identify trigger, precipitant and effect mentions in the sentence. In order to accomplish this, the MBD passes each contextualized token embedding c_i through a fully connected layer to produce a vector of potentials for each possible tag, $\tilde{b}_i \in \mathbb{R}^9$. A CRF is trained to extract the highest likelihood tag sequence given a sequence of potential vectors, $\tilde{b} = \tilde{b}_1, \dots, \tilde{b}_n$. Three separate fully connected layers are trained to produce potentials for triggers, precipitants, and effects, respectively, which are then fed to the same CRF. The MBD is trained to maximize the log likelihood of the tag sequences it predicts.

2.2.3 Relation Extractor

The relation extractor is used to extract relations between a pair of mentions and classify the type of relation between them, if any. Again, using the shared contextualized token embeddings created by the BERT sentence encoder, the relation extractor distills the token embeddings into: (a) a trigger embedding; (b) an argument embedding; and (c) a context embedding. The trigger embedding represents the relation trigger, while the argument embedding represents either a precipitant or a specific interaction (effect), depending on what was identified by the mention boundary detector. The context embedding represents the *context* in which the two relation arguments occur – i.e. it is derived from the rest of the sentence. Formally, the trigger

embedding is calculated using max-pooling, $\tau = \phi(c_i^t, \dots, c_j^t)$ where ϕ is the max-pooling operation and c_i^t, \dots, c_j^t are the tokens corresponding to the trigger mention. Similarly, the argument embedding is calculated as $\alpha = \phi(c_i^a, \dots, c_j^a)$ where c_i^a, \dots, c_j^a are the tokens corresponding to the argument mention and the context embedding is calculated as $\delta = \phi(c_i^s, \dots, c_j^s)$ where c_i^s, \dots, c_j^s are the tokens from the sentence that appear in neither argument.

The three embeddings are concatenated to derive the relation embedding, $r = [\tau, \alpha, \delta]$ which is passed to a fully connected softmax layer. The relation extractor is trained using softmax cross-entropy.

2.2.4 Pharmacokinetic Effect Classifier

For pharmacokinetic interactions (PKIs), the pharmacokinetic effect (PKE) classifier is used to predict the effect code of the interaction. The PKE classifier consists of a single softmax layer that operates on the relation embedding of the candidate relation. The PKE classifier is trained using softmax cross-entropy jointly along with the other four modules.

2.3 Postprocessing

Postprocessing began with reading predictions from our model's output and cleaning up predicted C-IOBES tags by removing malformed predicted spans. Predicted spans were transformed into predicted mentions, where continuation spans were linked to the closest same-type mention. First occurrences of mention text in the sentence were the only mentions kept, and interactions were reconstructed from predicted binary relations with the same head trigger. Predicted mentions which participated in no interactions were also removed. Unique interactions were maintained for each drug document, and at the end of the sentence-level predictions all the unique document-level interactions were collected and saved to XML.

2.4 Normalization

Each mention was normalized into one of three vocabularies: SNOMED-CT⁴, MED-RT⁵, and UNII⁶ using string matching against atoms from (a) the vocabularies themselves and (b) the Unified Medical Language System [4]. For each vocabulary, a dictionary was constructed using the 2019AA UMLS release augmented with primary names from the source vocabularies themselves. For MED-RT, only drug classes and their atoms were extracted. Specific Interactions were matched against SNOMED and precipitants were matched

⁴ <https://www.nlm.nih.gov/healthit/snomedct>

⁵ <https://www.nlm.nih.gov/research/umls/sourcereleasedocs/current/MED-RT>

⁶ <https://www.fda.gov/industry/fda-resources-data-standards/fdas-global-substance-registration-system>

against MED-RT first, then UNII if no match to MED-RT was found. In this way, the task of determining if a mention was a drug class was obviated.

3. Results

The results are presented in Table 1. LIIP is compared against the best submission for each subtask and the median submission for each subtask. In addition, we compare MTTDDI against our best submitted run (UTDHLTRI Run3) and a pipeline consisting of separately trained boundary, relation, and sentence classification models (Run3+Filtering). All evaluations are measured by F1 score.

| System | Task1 | Task2 | Task3 | Task4 |
|-------------------|--------------|--------------|--------------|--------------|
| Best Submission | 65.38 | 49.03 | 62.39 | 17.56* |
| Median Submission | 48.97 | 37.13 | 45.53 | 17.56 |
| UTDHLTRI Run3 | 35.04 | 27.48 | 28.66 | 17.56 |
| Run3 + Filtering | 56.03 | 42.29 | 45.73 | 24.07 |
| MTTDDI | 54.39 | 41.34 | 44.08 | 25.20 |

Table 1. Results on the four tasks of TAC DDI, measured in F1 score. Top performing systems are italicized and bolded. *Our submission was the lone attempt at task 4.

MTTDDI, UTDHLTRI Run3, and Run3+Filtering are all variations of the LIIP system. MTTDDI is the system described in this work while UTDHLTRI Run3 and Run3+Filtering replace MTTDDI with separately trained BERT models. UTDHLTRI Run3 consists of two models: BERT+the mention boundary detector and BERT+the relation extractor. UTDHLTRI Run3 was only trained on sentences containing DDIs, leading to very poor performance. However, when combined with sentence filtering the results are much more competitive. Run3+Filtering consists of the same boundary and relation models from Run3 along with a separately trained sentence classifier. It should be noted that MTTDDI is trained on the entire training corpus – not just the sentences with interactions as in Run3. Interestingly, the superior performance of Run3+Filtering to MTTDDI on tasks 1 and 2 indicates that training the boundary and relation detection modules only on sentences with interactions is useful for those tasks. However, while Run3+Filtering outperforms the jointly trained MTTDDI model in tasks 1-3, MTTDDI performs best in task 4 – which is the focus of the track.

4. Conclusion

In this work we present the FDA Label drug-drug Interaction Identification Pipeline (LDIIP) for identifying drug-drug interactions in FDA Labels. LDIIP uses a Multi-task neural model, MTTDDI, to perform end-to-end relation extraction. The results indicate that the multi-task paradigm is beneficial for the end-task of identifying unique drug-drug interactions in an FDA label.

References

- [1] Demner-Fushman, Dina, et al. "Overview of the TAC 2018 drug-drug interaction extraction from drug labels track." proceedings of the Text Analysis Conference (TAC 2018). 2018.
- [2] Devlin, Jacob, et al. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). 2019.
- [3] Dandala, Bharath, Diwakar Mahajan, and Ananya Poddar. "IBM Research system at TAC 2018: Deep learning architectures for drug-drug interaction extraction from structured product labels." Proceedings of the 2018 Text Analysis Conference (TAC 2018). 2018.
- [4] Lindberg, Donald AB, Betsy L. Humphreys, and Alexa T. McCray. "The unified medical language system." Yearbook of Medical Informatics 2.01 (1993): 41-51.
- [5] Sutton, Charles, and Andrew McCallum. "An introduction to conditional random fields." Foundations and Trends® in Machine Learning 4.4 (2012): 267-373.