KlickLabs at the TAC 2018 Drug-drug Interaction Extraction from Drug Labels Track

Gaurav Baruah and Maheedhar Kolla

Klick Inc., 175 Bloor Street East, Toronto, Ontario M4W 3R8, Canada
gbaruah@klick.com, mkolla@klick.com

Abstract. Detection and extraction of potential drug-drug interactions from collections of biomedical text is an important problem in automating pharmacovigilance. The 2018 TAC drug-drug interaction extraction track aims to compare systems that extract DDIs from Structured Product Labelings for drugs. We submitted a baseline system for this years iteration of the track, to gain a better understanding of the problem, and to build a base for future work.

Keywords: sentence classification · noun phrases

1 Introduction

Drug-drug interactions (DDI) can affect absorption of medications and cause adverse reactions, potentially endangering patients. The risk is exacerbated for patients taking multiple medications. Physicians and pharmacists need to keep track of potential DDIs in order to prescribe effective therapies while minimizing the risk of adverse effects for patients.

There are vast amounts of text available that describe drugs and their effects in the form of scientific articles, product labels, social media, electronic health records, amongst others. Mining these collections of text for potential drug-drug interactions would not only help researchers and clinicians keep up-to-date on latest findings, but also help them determine the best treatments for their patients. DDI detection/extraction could also help patients buying prescribed medication to be self-aware of possible conflicts.

The "Drug-Drug Interaction Extraction from Drug Labels" track at TAC 2018 works on a very specific task—to detect and extract potential drug-drug interactions from text, specifically from structured product labelings (SPLs).

The 2018 TAC DDI extraction problem can be broken down into the following steps:

1. Detecting entities (or mentions) of interest within a sentence: drugs, chemicals, families, classes, foods, symptoms, triggers, causes and effects
2. Extracting relationships (interactions) between entities and estimating the specific role of an entity in a detected interaction, e.g., is the entity a trigger, a symptom, or a precipitant

3. Estimating the specific type of an interaction: Pharmacokinetic, Pharmacodynamic, or Unspecified
4. Normalizing the detected entities and relationships into a standard medical vocabulary/ontology

We submitted the output of one baseline system for evaluation to the 2018 TAC DDI track. Our system consists of a sentence classifier that predicts whether a sentence contains an interaction, and a dependency parsing module that returns noun phrases as detected entities.

2 Data, Tasks, and Evaluation Metrics

2.1 Datasets

The track participants were provided with 22 SPLs for training[1] (`training22`). 2 sets of test SPLs were provided: `Test1` and `Test2`, containing 57 and 66 SPLs respectively.

The track organizers also suggested an additional dataset: the "NLM-DDI CD corpus DailyMed Cardiovascular Product Labels Annotated with Drug-drug Interactions"[2] (`NLM180`), prepared and manually annotated by the National Library of Medicine. The NLM180 dataset contains 180 SPLs.

2.2 Task Description

The main aim of the 2018 TAC drug-drug interaction (DDI) extraction task was to compare the performance of various NLP methods for the extraction of DDIs. The track's participants were mainly tasked to find: (i) the entities that play a role in an interaction, (ii) the interactions, as well as their specific types. Furthermore, the task explored normalizing the found entities and interactions to standard medical terminologies and coding standards like UMLS [1], NDF-RT NUI [3] and SNOMED CT [4].

Entities (also called *mentions*) can play one of the following roles in a DDI:

- Precipitants: a drug, a drug class/family, food substance, etc.
- Trigger: a word or phrase characterizes the interaction (e.g. avoid, increase the risk, etc).
- SpecificInteraction: a symptom or effect of the interaction.

Each entity/mention, once detected, was required to be classified into the particular role that it plays in an interaction.

Interactions are the relationships between the identified entities. As such, this task could potentially be tackled as an Entity-Relationship detection/extraction task. The interactions were required to be classified into one of the following types:

[1] https://bionlp.nlm.nih.gov/tac2018druginteractions/
[2] https://lhce-brat.nlm.nih.gov/NLMDDICorpus.htm
[3] https://www.nlm.nih.gov/research/umls/sourcereleasedocs/current/NDFRT/
[4] https://www.nlm.nih.gov/healthit/snomedct/

– Pharmacokinetic: phrases that indicate changes in physiological functions
– Pharmacodynamic: phrases that describe the effects of the drugs/precipitants
– Unspecified: phrases indicating caution

We did not attempt normalization sub-tasks for this year's iteration of this task, and we mainly focus on finding entities/mentions.

2.3 Evaluation metrics

Precision, Recall and F1 were the metrics utilized to measure performance for all subtasks of the track. For detected entities, the metrics were computed for exact as well as partial matches; the primary metric was the micro-averaged F1 for exact matches. For detected interactions, the primary metric was the micro-averaged F1.

3 Background and Related Work

Automatic drug-drug interaction (DDI) detection could help in enhancing pharmacovigilance. The key components DDI are drug identification and interaction detection. These key tasks roughly correspond to named-entity recognition and relationship extraction in conventional Natural Language Processing (NLP) techniques, although, in biomedical texts.

The 2013 SemEval task of extraction of drug-drug interactions from biomedical texts [8] has been previously attempted to compare systems that extract DDIs from the Drugbank [9] database and Medline abstracts. As in the 2018 TAC DDI task, the two subtasks in the 2013 SemEval tasks were to (i) recognize and classify pharmacological substances, and (ii) extract interactions between drugs. In contrast to the target classes specified for the 2018 TAC DDI task, the pharmacological substances were required to be classified as one of {generic drug name, brand name, drug group name, and active substances not approved for human use}. The interactions were required to be classified as {advice, effect, mechanism, int}.

The key findings from the 2013 SemEval task were that SVMs with non-linear kernels performed well for the detection of DDIs. For recognition of drugs or pharmacological substances, conditional random fields and dictionary-based methods worked best.

Recent advances in deep neural networks have also been used to tackle the DDI problem. Hierarchical bidirectional long short-term memory [7, 2] (Bi-LSTM) networks have successfully been developed to advance the state-of-the-art over the 2013 SemEval DDI dataset [10]. Another recent work tries to learn a joint model for entity and relation extraction [5], wherein Bi-LSTMs are used for drug recognition as well as extraction of DDIs.

Both methods take as input word embeddings over biomedical texts. Li et al. [5] argue that since a large number of new drugs and chemicals are invented each year, conventional dictionary-based methods may fail to scale. Con-

sequently, they developed character-based embeddings and develop a convolutional neural network model to extract and utilize morphological and lexical constructs as inputs for drug/entity recognition.

Interestingly, both these methods advocate that the shortest dependency path in the dependency parse-tree for a sentence, helps to identify semantic connections between interacting entities (drugs). They both also utilized parts-of-speech (POS) embeddings as inputs to their model.

4 System development

Recent developments in DDI using neural networks [10, 5] lend support to the use of dependency parsing and shortest dependency paths for DDI extraction.

For our baseline system, we performed dependency parsing on each sentence in an SPL, and we returned the noun phrases that were children of root verbs as entity mentions. We further tried to shortlist sentences on which to perform dependency parsing, in order to improve performance. We tried unsuccessfully to reproduce the system described in Li et al. [5] for generating runs; we leave this work for the future.

The `KLncLSTMsentClf` system from KlickLabs utilized a two-step process to find potential interactions and the mentions of interacting entities. Given a structured product label (SPL), for each sentence in the SPL,

1. A sentence classifier predicts whether a given sentence describes an interaction (Section 4.2).
2. If the sentence potentially contains an interaction, the sentence was parsed and noun phrases were returned as the interacting entities (Section 4.3).

4.1 Training data preparation

We expanded the provided `training-22` dataset with the NLM180 dataset. The training-22 and NLM180 datasets contain 309 and 2597 sentences respectively that describe interactions. Additionally, the training-22 and NLM180 datasets contain 294 and 3236 sentences that do not contain interactions. In total we have an *expanded* dataset containing 6436 sentences for training a sentence classifier to predict whether a sentence potentially contains an interaction. We utilized the mapping provided in the track guidelines [5] to translate labels from the NLM180 dataset to the TAC training-22 set. We did observe that the translated labels were noisy in nature.

4.2 Sentence Classifier for potential interactions

We initially trained two classifiers with 5-fold cross validation for sentence classification to predict potential interactions:

[5] https://bionlp.nlm.nih.gov/tac2018druginteractions/

- A random forest model; classification accuracy of 84.9% over the expanded training set
- A bidirectional LSTM neural network model; classification accuracy of 96.2% over the expanded training set

Thus, for step 1 in `KLncLSTMsentClf`, we utilized a neural network model with 2 layers of bidirectional LSTMs on the expanded training set to predict whether a sentence contains an interaction. The model takes as input word embedding representations of tokens in a sentence; these embeddings were sourced from http://bio.nlplab.org/#word-vectors and they were trained on Pubmed, PMC and Wikipedia articles [6].

4.3 Mentions extraction from sentences

For step 2 in `KLncLSTMsentClf`, we utilized SpaCy [4] (https://spacy.io/) to tokenize and perform dependency parsing [3] on each sentence shortlisted in step 1. We observed that most of the noun-phrases in sentences contained mentions of the entities in the sentences that contained interactions. Given that there exist many multi-word symptoms, disease names, and chemical reactions, we returned the entire noun phrase as a mention.

5 Results and Discussion

For test set `Test1`, for detecting entities/mentions, `KLncLSTMsentClf` gets an F1 of 9.04 and 12.78, when evaluated with and without entity role types respectively. Our sentence classifier identifies sentences describing DDIs with an accuracy of 74.7%, with precision of 0.47 and a recall of 0.62, for `Test1`.

For test set `Test2`, for detecting entities/mentions, `KLncLSTMsentClf` gets an F1 of 8.01 and 11.10, when evaluated with and without entity role types respectively. Our sentence classifier identifies sentences describing DDIs with an accuracy of 76.6%, with precision of 0.49 and a recall of 0.67, for `Test2`.

Even though our DDI sentence classifier achieves reasonable accuracy, it has low precision. Additionally, for our true-positive set of sentences, we observed that many multi-word noun phrases (that we returned) did not correspond to mentions in the gold standard. Although our returned words seem to overlap well with the gold standard, the returned token strings were not in the correct sequence or did not have the correct length, as per the mentions in the gold standard. As such, our baseline system performs well below the median performance for the track.

6 Conclusion

We submitted one system to the Drug-drug interaction extraction from drug labels track at TAC 2018. Our system consisted of a sentence classifier—that

shortlisted sentences containing potential interactions, and a noun chunker—that attempted to return multi-word mentions of precipitants, triggers, and specific interactions. In the future, we plan to implement and research deep neural network architectures for this task.

Acknowledgments

We thank Michael Li for conducting a preliminary literature survey, and Peter Leimbigler for proofreading and feedback. We would like to thank the KlickLabs team at Klick Inc. for their feedback, and support, for this research.

References

1. Bodenreider, O.: The unified medical language system (umls): integrating biomedical terminology. Nucleic acids research **32**(suppl_1), D267–D270 (2004)
2. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (Nov 1997). https://doi.org/10.1162/neco.1997.9.8.1735, http://dx.doi.org/10.1162/neco.1997.9.8.1735
3. Honnibal, M., Johnson, M.: An improved non-monotonic transition system for dependency parsing. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. pp. 1373–1378. Association for Computational Linguistics, Lisbon, Portugal (September 2015), https://aclweb.org/anthology/D/D15/D15-1162
4. Honnibal, M., Montani, I.: spacy 2: Natural language understanding with bloom embeddings, convolutional neural networks and incremental parsing. To appear (2017)
5. Li, F., Zhang, M., Fu, G., Ji, D.: A neural joint model for entity and relation extraction from biomedical text. BMC bioinformatics **18**(1), 198 (2017)
6. Pyysalo, S., Ginter, F., Moen, H., Salakoski, T., Ananiadou, S.: Distributional semantics resources for biomedical text processing. In: Proceedings of LBM 2013. pp. 39–44 (2013), http://lbm2013.biopathway.org/lbm2013proceedings.pdf
7. Schuster, M., Paliwal, K.K., General, A.: Bidirectional recurrent neural networks. IEEE Transactions on Signal Processing (1997)
8. Segura-Bedmar, I., Martínez, P., Zazo, M.H.: 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). vol. 2, pp. 341–350 (2013)
9. Wishart, D.S., Knox, C., Guo, A.C., Cheng, D., Shrivastava, S., Tzur, D., Gautam, B., Hassanali, M.: Drugbank: a knowledgebase for drugs, drug actions and drug targets. Nucleic acids research **36**(suppl_1), D901–D906 (2007)
10. Zhang, Y., Zheng, W., Lin, H., Wang, J., Yang, Z., Dumontier, M.: Drug–drug interaction extraction via hierarchical rnns on sequence and shortest dependency paths. Bioinformatics **34**(5), 828–835 (2017)