Biomedical Event Extraction
as Multi-turn Question Answering

Expose for student research project

Xing Wang

December 18, 2019

1 Introduction

People mine text collections for knowledge in many different tasks and in a multitude of applications. Common applications include question answering, sentiment analysis, named entity recognition and relation extraction. Tasks like entity recognition and relation extraction help in structuring knowledge and in gaining a better understanding of phenomena presented in text [Feldman and Sanger, 2007]. Early relation extraction tasks try to determine whether a simple binary relationship holds between two given entities whereas newer event extraction tasks also try to distinguish different event types with a varying amount of arguments which themselves can be events to create nested structures. In order to learn relation and event structures, the algorithms often rely on tokenized and annotated text with information regarding whether an entity participates in a given event or not. Considering the increasing amount of data automating the extraction process is mandatory as manual annotation is labour-intensive and not feasible [Gonzalez et al., 2015].

In the context of biomedical event extraction [Björne, 2014], one is interested in extracting events with biomedical entities such as genes and proteins and their biochemical reactions. Found events can be combined with other entities or events to create nested event structures, e.g., to model parts of biological pathways and networks. These biological networks may help in understanding biological processes and their underlying molecular mechanisms, e.g., understanding of diseases affecting humans [Gonzalez et al., 2015].

2 Background and Related Work

Early approaches of biomedical event extraction often divide the task into several multiclass classification problems solving them with Support Vector Machines (SVM) [Miwa et al., 2010]. Miwa et al. used manually created features from text, e.g., token features like *has a capital letter or a number*, then trained an SVM with the previously chosen features to find event triggers and repeated the process for identifying features of simple and complex events by training another SVM classifier to predict them. Valenzuela-Escárcega et al. [2015] developed a decision system based on fixed rules to extract events. State-of-the-art deep learning systems try to automatize the identification of features for their respective tasks. Chen et al. [2015] proposed a pipelined approach via dynamic multi-pooling convolutional neural networks to detect events. They first identified and classified events triggers. If a sentence contained triggers, they would try to find event arguments, assign these arguments to their corresponding event triggers and group both together to an event. They automatically captured lexical-level features with a word embedding model and sentence-level features using a Convolutional Neural Network (CNN). Chen et al. evaluated their model on the common Automatic Content Extraction (ACE05) [Walker et al., 2006] dataset for relation and event extraction. The dataset contains broadcast transcripts, newswire and newspaper data split in eight event types and 33 subtypes. However, such pipelined approaches share a common problem: By training independent classifiers for triggers and arguments, they suffer from error propagation, i.e. events with not recognized event triggers are not found in subsequent stages.

Joint framework approaches try to alleviate error propagation by extracting triggers and arguments together. Li et al. [2013] describe a joint event extraction model via structured prediction and global features, but still rely on a large set of manually designed features. Feng et al. [2016] use a hybrid neural network containing a Long short-term memory (LSTM) and a CNN to model both sequence and chunk information from specific contexts. Nguyen et al. [2016] use Recurrent Neural Networks (RNN) for their joint event extraction model and automatic feature extraction. Miwa and Bansal [2016] perform end-to- end relation extraction using LSTMs to capture structures like word sequences and dependency trees. The joint representation of entities and relations have shared parameters in a single model. Feng et al., Nguyen et al. and Miwa and Bansal all evaluate their results on the ACE datasets.

RNNs are well suited to incorporate short-term dependencies in sequences. However, they often fail to capture meaningful long-term dependencies, i.e. a relation between two entities far away from each other in a sentence. To solve this problem, RNNs and CNNs have been combined with the attention mechanism to allow a model to focus on different, distinct parts of a whole sequence. In 2017, Vaswani et al. proposed the Transformer model which drops convolution and recurrence to rely solely on attention. Devlin et al. introduced BERT in 2018, a pretrained Transformer model which facilitates transfer learning. They pre-train general language representations from unlabeled text on two tasks, masked language model and next sentence prediction, to learn both token-level and sentence-level features. Their training corpus consists of English Wikipedia and the BooksCorpus [Zhu et al., 2015]. After pretraining, BERT can be adapted to a new task by swapping inputs and outputs and fine-tuning all parameters end-to-end.

In our work, we follow a multi-turn question answering framework by Li et al. [2019] based on BERT to extract biomedical events. In this approach, Li et al. conduct multiple turns of question answering to retrieve specific arguments of a relation structure. Although answers of subsequent turns rely on the results of earlier turns, the model is a joint framework as the same BERT model and its parameters are shared across all question answering turns. A general question answering procedure consists of a question followed by a text passage as an input whereas a span from the text passage is marked as the output. For single-answer questions, it is sufficient to mark the starting and ending indices of the answer in the input passage. Instead of limiting their model to single-answer questions like in common question answering tasks, Li et al. allow multiple answers by annotating every token in the input sequence as either belonging or not belonging to the answer. The question answering model by Li et al. builds on top of BERT and fine-tunes its pretrained parameters for entity-relation extraction. These kind of adaptations of relation extraction and other natural language processing tasks into question answering formulations has gained traction in the last few years [McCann et al., 2018].

Li et al. apply their method to the relation extraction task of the ACE04, ACE05 and the CoNLL04 datasets. The relations described in the former three dataset are exclusively binary, therefore two turns of questions suffice to extract both arguments. To show the capability of inferring more complex relations, Li et al. introduce the RESUME dataset. The RESUME dataset is made up of 841 paragraphs from chapters containing people and their work history. It includes four entities and three relations, the entities *company, person, location, time* and the relations *Company(person, company)*, *Position(person, company, position)* and *Time(person, company, position, time)* which are ordered tuples

with different amounts of arguments. In each turn of question answering, one argument in the tuple is identified using its associated question template. In the first turn of question answering, the model queries for the first arguments of all relations, e.g., it is interested in all *person* entities. The question template for the *person* entity is *who is mentioned in the text?*. In the second turn, we use the answers from the first turn to fill the question template for the *Company* relation: "*Which companies did e1 work for?*", where *e1* denotes a slot to be filled with one of the extracted *person* entities. Not all extracted entities have to belong to a relation in the next turn. *None* answers are possible as well. In this case, there exists no relation involving the previously extracted entity. In the third turn of question answering, we use the previously extracted *person* and *company* to determine its corresponding *position*. The extraction of the *time* entity in the *Time* relation follows a similar schema. This iterative process allows the extraction of arbitrary hierarchical structures.

Extraction of biomedical events has been the subject of different BioNLP challenges, e.g., the Pathway Curation task from the BioNLP13 challenge [Ohta et al., 2013]. The objective in the Pathway Curation task is to identify and automatically extract a set of bio-molecular reactions in a given text. The reactions are represented in predefined event structures involving a variable number of physical entities, e.g., simple chemicals, proteins or genes, and/or other events. The input is given in form of raw text accompanied by the list of annotated physical entities. The expected output are event triggers, words indicating different events like *phosphorylates*, *transcription* or *binds*, and event arguments, e.g., themes or causes of an event. A cause is an entity or event that helps modifying the reactant entity or event (theme) into a new product. The Pathway Curation dataset consists of 525 biomedical abstracts and full texts relevant to pathway reactions with a total of 100,000 words, 16,000 entities and 12,000 events. The task uses the metrics recall, precision and F1-score for evaluation [Ohta et al., 2013].

3 Objective

This student research project aims to adapt the multi-turn question answering framework of Li et al. [2019] for extraction of biomedical events. To evaluate our implementation, we use data from the Pathway Curation task from the BioNLP13 challenge [Ohta et al., 2013] and compare the results with the baselines of Miwa and Ananiadou [2013] and TEES from Björne [2014].

4 Approach

We show how we plan to adapt Li et al.'s question answering approach to biomedical event extraction. In biomedical event extraction, we need to determine the event type by identifying the event triggers in an additional turn of question answering whereas the relation types in the ACE, CoNLL and RESUME datasets can be jointly extracted with the tail-entities of a relation. Challenges to overcome include dealing with overlapping and nested events. Overlapping arguments in multiple events can naturally be addressed with the allowance of multi-answer questions. Multi-answer questions allow us to distinguish events with the same head-entities but different tail-entities. We describe the two approaches we want to test illustrated with an example sentence in Figure 1 from the BioNLP13 challenge. The first approach contains only natural text in the input passage whereas the second approach incorporates special tagging tokens capturing already known tagging information in the input context.
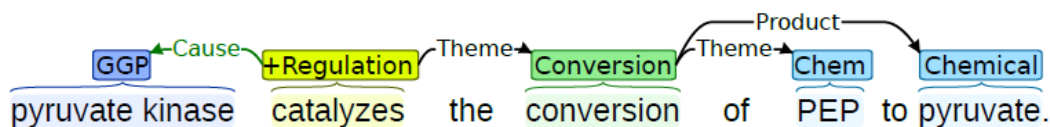

Figure 1: Event representation for a conversion reaction [Ohta et al., 2013]

In a first approach, we feed the question accompanied by the unmodified sentence or paragraph into BERT. For the first turn of question answering, we identify possible event triggers. In the subsequent turns, we continue by identifying the specific arguments for each previously detected event trigger. During these turns, we incorporate event arguments found in previous turn into our queries for new arguments. In case of *None* answers, all tokens in the text are labeled *O*. We do not distinguish between simple and nested events here, e.g., the question *What are the themes of the regulation event trigger "catalyzes"?* can return another event trigger like *conversion* as an answer for an event argument whereas in simple events only proteins are allowed as arguments. Nested structures are built in a subsequent step using all previously extracted arguments including event triggers. The question answering procedure for our example would start with a question for all triggers of conversion events: *Which conversion event triggers are mentioned in the text? Answer: conversion.* In the output, the *conversion* token would be labeled with a *B*-token (IOB2-style) whereas all other tokens would be labeled with an *O*-token. Given the conversion event triggers we would then pose the question *What are*

the themes of the conversion event trigger?. The answer would be the token *PEP* labeled with *B*. To discover the products of the conversion event we would continue to ask: *What are the products of the conversion event trigger with the theme PEP? Answer: Pyruvate.* Having found these arguments we would have extracted the whole conversion event with its corresponding event arguments, the theme and product.

In a second approach, we feed the question followed by a modified input passage into BERT. The modification includes special tagging tokens, e.g., *<T1>, <T2>* etc. denoting event triggers and proteins, who incorporate knowledge about the given physical entities and our already discovered event triggers with their respective arguments. During question answering, we only annotate the special tagging tokens with *B,I* or *O*-labels whereas all other tokens are annotated with an *O*. After extracting simple events we repeat the question answering procedure to include event tagging tokens, e.g., *<E1>, <E2>* etc., to identify complex events which have other events as arguments. The physical entities *pyruvate kinase, PEP* and *pyruvate* have already been recognized and are provided in the input data. We assign the tagging tokens *<T1>, <T2>* and *<T3>* to them. The conversion event trigger question from above would return the answer *conversion* which we assign the special token *<T4>*. When looking for the themes of the discovered conversion trigger, we would annotate the special tokens *<T2>* before and after the word *PEP* with *B* and *I*. Having extracted the arguments we extract the simple conversion event *<E1>: conversion<T4> theme<T2> product<T3>* and add the event tagging token *<E1>* to our input context. A fully modified context with special tagging tokens for our example after having extracted the full event structure would look like: *<E2Cause> <T1> pyruvate kinase <T1> <E2Regulation> <T5> catalyzes <T5> the <E2Theme> <E1Conversion> <T4> conversion <T4> of <E1Theme> <T2> PEP <T2> to <E1Product> <T3> pyruvate <T3>*.

The parameters in the multi-turn question answering are shared and jointly trained but the extracted answers from one turn still affect subsequent answers of downstream turns. To alleviate this problem, we could try to improve the model performance by including a reinforcement learning approach as described by Li et al. in 2019.

For implementation, we use Li et al.'s implementation as our starting point. We further rely on Huggingface's Transformers [Wolf et al., 2019] and BioBERT [Lee et al., 2019]. BioBERT provides weights pretrained on biomedical texts from the PubMed corpus.

References

Jari Björne. *Biomedical Event Extraction with Machine Learning*. PhD thesis, University of Turku, 2014.

Yubo Chen, Liheng Xu, Kang Liu, Daojian Zeng, and Jun Zhao. Event extraction via dynamic multi-pooling convolutional neural networks. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 167–176, 2015.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

Ronen Feldman and James Sanger. *The text mining handbook: advanced approaches in analyzing unstructured data*. Cambridge university press, 2007.

Xiaocheng Feng, Lifu Huang, Duyu Tang, Heng Ji, Bing Qin, and Ting Liu. A language-independent neural network for event detection. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 66–71, 2016.

Graciela H Gonzalez, Tasnia Tahsin, Britton C Goodale, Anna C Greene, and Casey S Greene. Recent advances and emerging applications in text and data mining for biomedical discovery. *Briefings in bioinformatics*, 17(1):33–42, 2015.

Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 09 2019. ISSN 1367-4803. doi: 10.1093/bioinformatics/btz682. URL `https://doi.org/10.1093/bioinformatics/btz682`.

Qi Li, Heng Ji, and Liang Huang. Joint event extraction via structured prediction with global features. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 73–82, 2013.

Xiaoya Li, Fan Yin, Zijun Sun, Xiayu Li, Arianna Yuan, Duo Chai, Mingxin Zhou, and Jiwei Li. Entity-relation extraction as multi-turn question answering. *arXiv preprint arXiv:1905.05529*, 2019.

Bryan McCann, Nitish Shirish Keskar, Caiming Xiong, and Richard Socher. The natural language decathlon: Multitask learning as question answering. *arXiv preprint arXiv:1806.08730*, 2018.

Makoto Miwa and Sophia Ananiadou. Nactem eventmine for bionlp 2013 cg and pc tasks. In *Proceedings of the BioNLP Shared Task 2013 Workshop*, pages 94–98, 2013.

Makoto Miwa and Mohit Bansal. End-to-end relation extraction using lstms on sequences and tree structures. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1105–1116, 2016.

Makoto Miwa, Rune Sætre, Jin-Dong Kim, and Jun'ichi Tsujii. Event extraction with complex event classification using rich features. *Journal of bioinformatics and computational biology*, 8(01):131–146, 2010.

Thien Huu Nguyen, Kyunghyun Cho, and Ralph Grishman. Joint event extraction via recurrent neural networks. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 300–309, 2016.

Tomoko Ohta, Sampo Pyysalo, Rafal Rak, Andrew Rowley, Hong-Woo Chun, Sung-Jae Jung, Sung-Pil Choi, Sophia Ananiadou, and Jun'ichi Tsujii. Overview of the pathway curation (pc) task of bionlp shared task 2013. In *Proceedings of the BioNLP Shared Task 2013 Workshop*, pages 67–75, 2013.

Marco A Valenzuela-Escárcega, Gus Hahn-Powell, Mihai Surdeanu, and Thomas Hicks. A domain-independent rule-based framework for event extraction. In *Proceedings of ACL-IJCNLP 2015 System Demonstrations*, pages 127–132, 2015.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008, 2017.

Christopher Walker, Stephanie Strassel, Julie Medero, and Kazuaki Maeda. Ace 2005 multilingual training corpus. *Linguistic Data Consortium, Philadelphia*, 57, 2006.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, et al. Transformers: State-of-the-art natural language processing. *arXiv preprint arXiv:1910.03771*, 2019.

Yukun Zhu, Ryan Kiros, Rich Zemel, Ruslan Salakhutdinov, Raquel Urtasun, Antonio Torralba, and Sanja Fidler. Aligning books and movies: Towards story-like visual explanations by watching movies and reading books. In *Proceedings of the IEEE international conference on computer vision*, pages 19–27, 2015.