Improving Cross-Lingual Transfer for Event Argument Extraction with Language-Universal Sentence Structures

Minh Van Nguyen and Thien Huu Nguyen
Department of Computer and Information Science
University of Oregon, Eugene, Oregon, USA
{minhnv,thien}@cs.uoregon.edu

Abstract

We study the problem of Cross-lingual Event Argument Extraction (CEAE). The task aims to predict argument roles of entity mentions for events in text, whose language is different from the language that a predictive model has been trained on. Previous work on CEAE has shown the cross-lingual benefits of universal dependency trees in capturing shared syntactic structures of sentences across languages. In particular, this work exploits the existence of the syntactic connections between the words in the dependency trees as the anchor knowledge to transfer the representation learning across languages for CEAE models (i.e., via graph convolutional neural networks – GCNs). In this paper, we introduce two novel sources of language-independent information for CEAE models based on the semantic similarity and the universal dependency relations of the word pairs in different languages. We propose to use the two sources of information to produce shared sentence structures to bridge the gap between languages and improve the cross-lingual performance of the CEAE models. Extensive experiments are conducted with Arabic, Chinese, and English to demonstrate the effectiveness of the proposed method for CEAE.

1 Introduction

Event Argument Extraction (EAE) aims to classify argument roles of entity mentions for events in text. For example, given the sentence *"He **died** of injuries from a grenade attack by a fellow soldier"*, the task requires systems to identify the entity mention "*a fellow soldier*" as the *Agent* of the event *Die*, which is triggered by the verb "***died***". EAE is an important component of event extraction (EE) that has been extensively studied with different approaches (Ji and Grishman, 2008; Liao and Grishman, 2011a; Li et al., 2014; Nguyen and Grishman, 2015b; Nguyen et al., 2016; Nguyen and Grishman, 2018; Liu et al., 2018; Zhang et al., 2019b; Wang et al., 2019). Cross-lingual Event Argument Extraction (CEAE) is an instance of EAE that considers the setting where test languages (i.e., target languages) are different from training languages (i.e., source languages). The goal is to transfer knowledge in source languages, where data is abundant, to low-resource target languages. The previous work on CEAE (Subburathinam et al., 2019) has shown the existence of shared syntactic structures of sentences across languages, which are useful for cross-lingual transfer. In particular, with the multilingual word embeddings, Subburathinam et al. (2019) develop a model based on Graph Convolutional Networks (GCNs) (Kipf and Welling, 2017; Zhang et al., 2018), which operates on universal dependency trees to capture the shared structures.

Notably, the use of the dependency trees of the sentences for GCNs in (Subburathinam et al., 2019) essentially treats the existence of the syntactic connections between the words in the universal dependency trees as the language-universal knowledge that can be exploited to bridge the gap between languages for EAE. In (Subburathinam et al., 2019), such syntactic connection existences are formalized via the adjacency matrices $A^{dep} = \{a^{dep}_{ij}\}_{i,j=1..N}$ of the dependency trees (i.e., N is the number of words in the input sentence and $a^{dep}_{ij} = 1$ if the i-th and j-th words are connected in the dependency tree) that would be consumed by GCNs for representation learning. We call A^{dep} the *syntax-based structures* of the sentences for convenience (as a^{dep}_{ij} is based on the syntactic connection of the words).

As such, in this work, we introduce two novel sources of information as the language anchors, which are complementary to the syntactic connections A^{dep}, to enable GCNs to learn better language-general representations for EAE. The first source of information relies on the semantic similarities of the pairs of words in the input

Proceedings of the Sixth Arabic Natural Language Processing Workshop, pages 237–243
Kyiv, Ukraine (Virtual), April 19, 2021.

sentence to induce the *semantic-based structures* $A^{sem} = \{a_{ij}^{sem}\}_{i,j=1..N}$ for GCNs. The rationale for such semantic-based structures is that despite the vocabulary differences between languages, the semantic similarity of the words is a language-invariant concept and can be leveraged to enhance the cross-lingual knowledge transfer for EAE. In this work, we rely on the multilingual representation vectors of the words to facilitate such semantic similarity computation for a_{ij}^{sem} in different languages. For the second source of information, we propose to employ the syntactic dependency relations between the words (e.g., *nsubj*, *conj*) in the dependency trees to obtain the *relation-based structures* $A^{rel} = \{a_{ij}^{rel}\}_{i,j=1..N}$ for EAE with GCNs. Specifically, $A^{rel} = \{a_{ij}^{rel}\}_{i,j=1..N}$ is an extension of A^{dep} that further considers the natures (i.e., the relations) of the syntactic connections between the words (i.e., instead of using only the existence of the connections as in A^{dep}). Similar to the semantic-based structures, we argue that the syntactic dependency relations from the universal dependency trees are also language-independent and can be helpful for our CEAE problem. To this end, we employ the embeddings of the dependency relations to compute the relation-based structure scores a_{ij}^{rel}. Note that all the structures A^{dep}, A^{sem}, and A^{rel} are fed into GCN models for representation learning in this work. Finally, we conduct extensive experiments to demonstrate the benefits of the proposed sentence structures, leading to the state-of-the-art performance for CEAE with Arabic, Chinese, and English as the experiment languages. To our knowledge, this is the first work to examine semantic-based and relation-based structures for EAE.

2 Related Work

EAE and EE have been extensively studied for English in the monolingual context of Event Extraction, featuring both the traditional machine learning models (Patwardhan and Riloff, 2009; Liao and Grishman, 2011b; Li et al., 2013; Yang and Mitchell, 2016) and the recent advanced deep learning models (Chen et al., 2015; Sha et al., 2018; Wang et al., 2019; Zhang et al., 2019a; Nguyen and Nguyen, 2019; Lai and Nguyen, 2019; Lai et al., 2020; Pouran Ben Veyseh et al., 2020). Only a few works have considered cross-lingual learning for EAE (Chen and Ji, 2009; Hsi et al., 2016; Subburathinam et al., 2019).

Cross-lingual transfer learning has also been examined for the other related tasks of EAE, including multilingual relation extraction (Kim et al., 2010; Qian et al., 2014; Faruqui and Kumar, 2015; Lin et al., 2017; Zou et al., 2018; Wang et al., 2018) and semantic role labeling (Mulcaire et al., 2018, 2019; Liu et al., 2019). However, none of these works explores edge-based attention GCN as we do.

Finally, our work is also related to the recent text structure models for other NLP tasks, including relation extraction (Sahu et al., 2019; Tran et al., 2020), event factuality prediction (Veyseh et al., 2019), and text summarization (Balachandran et al., 2020).

3 Model

We formalize EAE as a multi-class classification problem. Let $W = w_1, w_2, ..., w_N$ be a sentence (of N words) with w_t as the trigger word and w_a as the argument candidate (i.e., an entity mention) ($1 \leq t, a \leq N$). The goal of EAE is to predict the role y^* of w_a for the event triggered by w_t.

Following (Subburathinam et al., 2019), we use the UDPipe toolkit (Straka and Straková, 2017) to obtain the universal dependency tree for W, the part of speech (POS) tags and BIO entity type tags for the words in W. For convenience, let R be the set of universal dependency relations and E be the matrix for the embedding vectors of such dependency relations where E_r indicates the embedding vector for $r \in R$.

In the first step for cross-lingual EAE, each word w_i in W is represented by the concatenation vector x_i of three language-universal embedding vectors: $x_i = [x_i^w, x_i^p, x_i^e, x_i^d]$ where x_i^w is the multilingual word embedding for w_i from MUSE (Joulin et al., 2018), x_i^p is the embedding vector for the POS tag of $w_i \in W$, x_i^e is the embedding vector for the entity type tag of w_i, and x_i^d is the embedding vector E_{r_i} for the dependency relation r_i between w_i and its governor. The POS tag and entity type tag embeddings are initialized randomly and learned via training. After this step, the input sentence W would be transformed into a sequence of representation vectors $X = x_1, x_2, \ldots, x_N$. As presented in the introduction, our CEAE model involves three major sentence structures (i.e., A^{dep}, A^{sem}, and A^{rel}) that would be consumed by the GCN models to perform CEAE. We describe these components in detail below.

Models	Cross-lingual (training/test languages)									Monolingual		
	en/ch	en/ar	ch/en	ch/ar	ar/en	ar/ch	en+ch/ar	en+ar/ch	ch+ar/en	en	ch	ar
GCN	**59.0**	61.8	51.6	60.6	43.1	50.1	63.1	**60.1**	51.9	63.9	59.3	64.0
RSGCN	58.4	**62.9**	**53.9**	**63.3**	**48.0**	**52.6**	**64.0**	59.1	**55.5**	**64.9**	**63.7**	**66.9**

Table 1: F1 scores of the models on the test data.

Models	en/ch	en/ar	ch/en	ch/ar	ar/en	ar/ch
RSGCN	**55.3**	**63.3**	**56.0**	**63.8**	**50.5**	**53.8**
-A^{rel}	53.1	60.0	52.6	61.8	49.6	52.0
-A^{sem}	53.5	61.5	54.0	62.6	48.7	53.1
-A^{sem}-A^{rel}	53.8	61.4	52.3	59.4	47.6	51.9

Table 2: F1 scores of the models on the development data.

3.1 Language-Universal Sentence Structures

A sentence structure in this work refers to a real-valued matrix $A = \{a_{ij}\}_{i,j=1..N}$ capturing the levels of interactions/dependencies between the pairs of words in W. In particular, the score $a_{ij} \in A$ represents the contribution that the representation vector of w_j would provide for the representation vector computation of w_i in W according to some information source/perspective (e.g., syntax or semantic). Our goal in this work is to obtain the sentence structures for W based on the language-independent knowledge (thus called language-universal sentence structures) to enable cross-lingual representation learning for EAE. As such, three types of sentence structures are utilized in this work:

(i) Syntax-based Sentence Structures (denoted by $A^{dep} = \{a_{ij}^{dep}\}_{i,j=1..N}$): This structure is inherited from (Subburathinam et al., 2019) to capture the syntactic connections of the words in the dependency tree T of W. In particular, $a_{ij}^{dep} = 1$ only if w_i and w_j are connected in T.

(ii) Semantic-based Sentence Structures (denoted by $A^{sem} = \{a_{ij}^{sem}\}_{i,j=1..N}$): As mentioned in the introduction, this type of structures aims to leverage the semantic similarities between pairs of words as the universal knowledge across languages for CEAE. In this work, we use the multilingual word embedding vectors x_i^w to capture the semantic representations of the words w_i. The semantic-based structure score a_{ij}^{sem} is then computed by: $a_{ij}^{sem} = \tanh\left(u^\top(x_i^w \odot x_j^w)\right)$ where $\odot$ is the element-wise product and u is a learnable vector.

(iii) Relation-based Sentence Structures (denoted by $A^{rel} = \{a_{ij}^{rel}\}_{i,j=1..N}$): The syntax-based structures A^{dep} only consider the syntactic connections of the words to generate the structure scores. In this work, we note that the dependency relations (e.g., *nsubj*, *conj*) between the words in the universal dependency trees are also the language-independent concepts. To this end, we propose to further exploit such dependency relations to obtain the relation-based structure scores a_{ij}^{rel} for CEAE: $a_{ij}^{rel} = \tanh(v^\top E_{r_{ij}})$ if w_i and w_j are connected in T and 0 otherwise (here r_{ij} is the dependency relation between w_i and w_j in T). Here, v is a learnable vector.

Note that we normalize A_{dep} via the neighbor sizes of the words, and A_{sem}, A_{rel} via the softmax function to ensure that the weights corresponding to a word w_i (i.e., a_{ij}^* for $j = 1..N$) sum to 1.

3.2 Graph Convolutional Neural Networks

In order to exploit the aforementioned sentence structures for representation learning for CEAE, we propose to first combine the structures via a linear combination, leading to a richer structure $A = \{a_{ij}\}_{i,j=1..N}$:

$$A = \gamma_1 A^{dep} + \gamma_2 A^{sem} + (1 - \gamma_1 - \gamma_2) A^{rel} \quad (1)$$

Afterward, we follow (Subburathinam et al., 2019) to feed A into a GCN model for representation learning. In particular, the GCN model in this work involves L layers. The representation vector h_i^l for the word $w_i \in W$ at the l-th layer ($1 \leq l \leq L$) is computed via: $h_i^l = \text{ReLU}(\sum_{j=1}^N a_{ij}(W^l h_j^{l-1} + b^l))$ where h_i^0 is set to x_i for all $1 \leq i \leq N$, and W^l and b^l are the learnable weight matrices and bias vectors at the l-th layer.

In the next step, an overall representation vector V is computed based on hidden vectors in the last layer of the GCN model via: $V = [h_a^L, h_t^L, \text{max\_pooling}(h_1^L, \ldots, h_N^L)]$. This vector is sent into an one-layer feed-forward network to

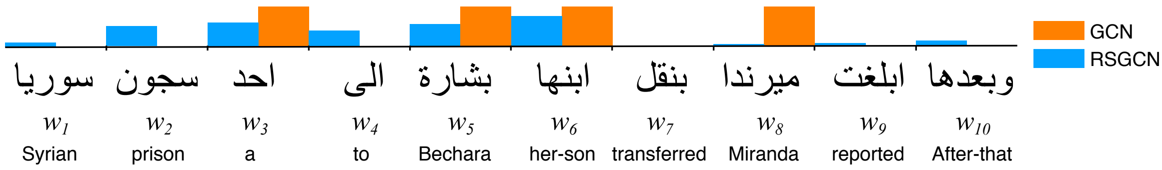

Figure 1: Visualization of the weights of the words with respect to the trigger word w_7 (i.e., a_{7j} and a_{7j}^{dep}).

estimate the distribution $P(.|W, w_a, w_t)$ over all the possible argument roles for our CEAE problem. Finally, the negative log-likelihood $L = -\log P(y^*|W, w_a, w_t)$ is used as the loss function to train the models in this work.

4 Experiments

4.1 Dataset and Hyper-parameters

Following (Subburathinam et al., 2019), we evaluate the models in this work using the multilingual ACE 2005 dataset where the EAE data is provided for three languages, i.e., Arabic (*ar*), Chinese (*ch*), and English (*en*). We use the preprocessed data and the train/dev/test split provided by (Subburathinam et al., 2019) to ensure a fair comparison. The development data is used to finetune hyper-parameters. In particular, we use 30 dimensions for the POS and entity type embeddings, 30 dimensions for the dependency relation embeddings (that are initialized randomly and updated during training), 200 dimensions for the hidden vector of GCN, 2 layers for the GCN model, the bath size of 50 for mini-batching, 0.5 for the learning rate for the SGD optimizer, and 0.9 for the learning rate decay. For the novel sentence structures, we observe that $\gamma_1 = 0.6$ and $\gamma_2 = 0.1$ leads to the best performance of the proposed model on the development data.

4.2 Comparison with the State of the Art

This section compares the proposed model (called **RSGCN** – Rich Sentence Structure-based GCN) with the GCN-based model in (Subburathinam et al., 2019) (called **GCN**). In particular, the models are trained on the training datasets for one or two of the three language (i.e., *en*, *ch* and *ar*) that are then evaluated on the test datasets for the other languages. Table 1 reports the performance of the models. As we can see, the proposed model RSGCN significantly outperforms GCN on seven over nine cross-lingual settings, and interestingly also on all the monolingual settings with substantial performance gaps ($p < 0.01$). This clearly demonstrates the advantages of the proposed semantic-based and relation-based structures for CEAE.

4.3 Ablation Study

To assess the contributions of the novel semantic-based (A^{sem}) and relation-based (A^{rel}) structures in this work, we exclude each of them from RSGCN and evaluate the performance of the remaining model on development data. This ablation study is conducted over six different cross-lingual settings, i.e., six choices for different source and target language.

Table 2 shows that both structures A^{sem} and A^{rel} are necessary for the proposed model as removing any of them would hurt the performance across different language pairs. We also observe that for most language pairs (e.g., *en/ar*, *ch/en*), excluding A^{rel} would lead to a deeper performance drop than those for removing A^{sem}, thus demonstrating the more importance of A^{rel} over A^{sem}. We attribute this to the fact that the A^{rel} structure is based on explicit structural information (i.e., dependency relations) which could be more valuable for the structure transfer.

4.4 Analysis

To understand the effect of the proposed structures, we analyze the examples from the development data of the setting *en/ar* that RSGCN makes correct predictions but GCN does not. Among others, we find that the proposed sentence structures help RSGCN assign more appropriate structure scores for the words for better representation learning. Consider Figure 1 as an example where a rough translation for the sentence is *"After that, Miranda was informed that her son Bechara had been transferred to a Syrian prison"*. In this example, the word w_7 (i.e., *"transferred"*) is the trigger of the event *"Movement-Transport"*, and the word w_6 (i.e., *"her son"*) is the argument with the role *"Artifact"*.

As shown in the figure, both RSGCN and GCN assign the highest score for the most important word w_6 (i.e., *"her son"*). However, in addition to that, GCN also considers the words w_5 (i.e., *"Bechara"*), w_3 (i.e., *"a"*), and w_8 (i.e., *"Miranda"*) as equally important as w_6. This is problematic as the irrelevant words w_3 and w_5 for argument role identification might introduce noise into the representation vectors. Even worse, the high score for w_8 might cause the model to incorrectly predict *"Miranda"* as the argument in this case. To this end, the proposed sentence structures help RSGCN to mitigate such issues by re-distributing the scores so *"her son"* can have the highest score and the confusing word *"Miranda"* is almost canceled (with the nearly zero score), eventually leading to the success of RSGCN on this example.

5 Conclusion and Future Work

We introduce two novel sentence structures for cross-lingual EAE with GCNs based on the semantic similarity and the universal dependency relations of the words in the input sentences. The experiments demonstrate the benefits of the proposed sentence structures that lead to the state-of-the-art performance for different experiments scenarios for CEAE. In the future, we plan to apply the proposed model to other related tasks, e.g., cross-lingual relation extraction (Veyseh et al., 2020). In addition, motivated by the recent introduction of high-performance multilingual NLP toolkits, e.g., Trankit (Nguyen et al., 2021), we expect to extend our work to other languages to better demonstrate the benefits of the proposed models. Finally, we will also explore the performance of our models when recent pre-trained multilingual language models, e.g., multilingual BERT (Devlin et al., 2019), are employed to encode input texts for different languages.

Acknowledgments

This research has been supported by the Army Research Office (ARO) grant W911NF-17-S-0002. This research is also based upon work supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via IARPA Contract No. 2019-19051600006 under the Better Extraction from Text Towards Enhanced Retrieval (BETTER) Program. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ARO, ODNI, IARPA, the Department of Defense, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein. This document does not contain technology or technical data controlled under either the U.S. International Traffic in Arms Regulations or the U.S. Export Administration Regulations.

References

Vidhisha Balachandran, Artidoro Pagnoni, Jay Yoon Lee, Dheeraj Rajagopal, Jaime Carbonell, and Yulia Tsvetkov. 2020. Structsum: Incorporating latent and explicit sentence dependencies for single document summarization. In *arXiv preprint arXiv:2003.00576*.

Yubo Chen, Liheng Xu, Kang Liu, Daojian Zeng, and Jun Zhao. 2015. Event extraction via dynamic multi-pooling convolutional neural networks. In *ACL*.

Zheng Chen and Heng Ji. 2009. Can one language bootstrap the other: a case study on event extraction. In *Workshop on Semi-Supervised Learning for Natural Language Processing*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *ACL*.

Manaal Faruqui and Shankar Kumar. 2015. Multilingual open relation extraction using cross-lingual projection. In *NAACL-HLT*.

Andrew Hsi, Yiming Yang, Jaime Carbonell, and Ruochen Xu. 2016. Leveraging multilingual training for limited resource event extraction. In *COLING*.

Heng Ji and Ralph Grishman. 2008. Refining event extraction through cross-document inference. In *ACL*.

Armand Joulin, Piotr Bojanowski, Tomas Mikolov, Hervé Jégou, and Edouard Grave. 2018. Loss in translation: Learning bilingual word mapping with a retrieval criterion. In *EMNLP*.

Seokhwan Kim, Minwoo Jeong, Jonghoon Lee, and Gary Geunbae Lee. 2010. A cross-lingual annotation projection approach for relation detection. In *COLING*.

Thomas N. Kipf and Max Welling. 2017. Semi-supervised classification with graph convolutional networks. In *ICLR*.

Viet Dac Lai and Thien Huu Nguyen. 2019. Extending event detection to new types with learning from keywords. In *Proceedings of the 5th Workshop on Noisy User-generated Text (W-NUT) at EMNLP 2019*.

Viet Dac Lai, Tuan Ngo Nguyen, and Thien Huu Nguyen. 2020. Event detection: Gate diversity and syntactic importance scores for graph convolution neural networks. In *EMNLP*.

Qi Li, Heng Ji, Yu Hong, and Sujian Li. 2014. Constructing information networks using one single model. In *EMNLP*.

Qi Li, Heng Ji, and Liang Huang. 2013. Joint event extraction via structured prediction with global features. In *ACL*.

Shasha Liao and Ralph Grishman. 2011a. Acquiring topic features to improve event extraction: in preselected and balanced collections. In *RANLP*.

Shasha Liao and Ralph Grishman. 2011b. Acquiring topic features to improve event extraction: in preselected and balanced collections. In *RANLP*.

Yankai Lin, Zhiyuan Liu, and Maosong Sun. 2017. Neural relation extraction with multi-lingual attention. In *ACL*.

Nelson F. Liu, Matt Gardner, Yonatan Belinkov, Matthew E. Peters, and Noah A. Smith. 2019. Linguistic knowledge and transferability of contextual representations. In *NAACL-HLT*.

Xiao Liu, Zhunchen Luo, and Heyan Huang. 2018. Jointly multiple events extraction via attention-based graph information aggregation. In *EMNLP*.

Phoebe Mulcaire, Jungo Kasai, and Noah A. Smith. 2019. Polyglot contextual representations improve crosslingual transfer. In *NAACL-HLT*.

Phoebe Mulcaire, Swabha Swayamdipta, and Noah A. Smith. 2018. Polyglot semantic role labeling. In *ACL*.

Minh Nguyen, Viet Lai, Amir Pouran Ben Veyseh, and Thien Huu Nguyen. 2021. Trankit: A light-weight transformer-based toolkit for multilingual natural language processing. In *EACL (System Demonstrations)*.

Thien Huu Nguyen, Kyunghyun Cho, and Ralph Grishman. 2016. Joint event extraction via recurrent neural networks. In *NAACL-HLT*.

Thien Huu Nguyen and Ralph Grishman. 2015b. Event detection and domain adaptation with convolutional neural networks. In *ACL-IJCNLP*.

Thien Huu Nguyen and Ralph Grishman. 2018. Graph convolutional networks with argument-aware pooling for event detection. In *AAAI*.

Trung Minh Nguyen and Thien Huu Nguyen. 2019. One for all: Neural joint modeling of entities and events. In *AAAI*.

Siddharth Patwardhan and Ellen Riloff. 2009. A unified model of phrasal and sentential evidence for information extraction. In *ACL*.

Amir Pouran Ben Veyseh, Tuan Ngo Nguyen, and Thien Huu Nguyen. 2020. Graph transformer networks with syntactic and semantic structures for event argument extraction. In *EMNLP Findings*.

Longhua Qian, Haotian Hui, Yanan Hu, Guodong Zhou, and Qiaoming Zhu. 2014. Bilingual active learning for relation classification via pseudo parallel corpora. In *ACL*.

Sunil Kumar Sahu, Fenia Christopoulou, Makoto Miwa, and Sophia Ananiadou. 2019. Inter-sentence relation extraction with document-level graph convolutional neural network. In *ACL*.

Lei Sha, Feng Qian, Baobao Chang, and Zhifang Sui. 2018. Jointly extracting event triggers and arguments by dependency-bridge rnn and tensor-based argument interaction. In *AAAI*.

Milan Straka and Jana Straková. 2017. Tokenizing, pos tagging, lemmatizing and parsing ud 2.0 with udpipe. In *CoNLL*.

Ananya Subburathinam, Di Lu, Heng Ji, Jonathan May, Shih-Fu Chang, Avirup Sil, and Clare Voss. 2019. Cross-lingual structure transfer for relation and event extraction. In *EMNLP-IJCNLP*.

Hieu Minh Tran, Minh Trung Nguyen, and Thien Huu Nguyen. 2020. The dots have their values: Exploiting the node-edge connections in graph-based neural models for document-level relation extraction. In *EMNLP Findings*.

Amir Pouran Ben Veyseh, Franck Dernoncourt, Dejing Dou, and Thien Huu Nguyen. 2020. Exploiting the syntax-model consistency for neural relation extraction. In *ACL*.

Amir Pouran Ben Veyseh, Thien Huu Nguyen, and Dejing Dou. 2019. Graph based neural networks for event factuality prediction using syntactic and semantic structures. In *ACL*.

Xiaozhi Wang, Xu Han, Yankai Lin, Zhiyuan Liu, and Maosong Sun. 2018. Adversarial multi-lingual neural relation extraction. In *COLING*.

Xiaozhi Wang, Ziqi Wang, Xu Han, Zhiyuan Liu, Juanzi Li, Peng Li, Maosong Sun, Jie Zhou, and Xiang Ren. 2019. Hmeae: Hierarchical modular event argument extraction. In *EMNLP-IJCNLP*.

Bishan Yang and Tom M. Mitchell. 2016. Joint extraction of events and entities within a document context. In *NAACL-HLT*.

Junchi Zhang, Yanxia Qin, Yue Zhang, Mengchi Liu, and Donghong Ji. 2019a. Extracting entities and events as a single task using a transition-based neural model. In *IJCAI*.

Tongtao Zhang, Heng Ji, and Avirup Sil. 2019b. Joint entity and event extraction with generative adversarial imitation learning. In *Data Intelligence*.

Yuhao Zhang, Peng Qi, and Christopher Manning. 2018. Graph convolution over pruned dependency trees improves relation extraction. In *EMNLP*.

Bowei Zou, Zengzhuang Xu, Yu Hong, and Guodong Zhou. 2018. Adversarial feature adaptation for cross-lingual relation classification. In *COLING*.