Crosslingual Transfer Learning for Relation and Event Extraction via Word Category and Class Alignments

Minh Van Nguyen[1], **Tuan Ngo Nguyen**[1], **Bonan Min**[2], **and Thien Huu Nguyen**[1,3]
[1] Dept. of Computer and Information Science, University of Oregon, Eugene, OR, USA
[2] Raytheon BBN Technologies, USA
[3] VinAI Research, Vietnam
{minhnv,tnguyen,thien}@cs.uoregon.edu,
bonan.min@raytheon.com

Abstract

Previous work on crosslingual Relation and Event Extraction (REE) suffers from the monolingual bias issue due to the training of models on only the source language data. An approach to overcome this issue is to use unlabeled data in the target language to aid the alignment of crosslingual representations, i.e., via fooling a language discriminator. However, as this approach does not condition on class information, a target language example of a class could be incorrectly aligned to a source language example of a different class. To address this issue, we propose a novel crosslingual alignment method that leverages class information of REE tasks for representation learning. In particular, we propose to learn two versions of representation vectors for each class in an REE task based on either source or target language examples. Representation vectors for corresponding classes will then be aligned to achieve class-aware alignment for crosslingual representations. In addition, we propose to further align representation vectors for language-universal word categories (i.e., parts of speech and dependency relations). As such, a novel filtering mechanism is presented to facilitate the learning of word category representations from contextualized representations on input texts based on adversarial learning. We conduct extensive crosslingual experiments with English, Chinese, and Arabic over REE tasks. The results demonstrate the benefits of the proposed method that significantly advances the state-of-the-art performance in these settings.

1 Introduction

Relation and Event Extraction (REE) are important tasks of Information Extraction (IE), whose goal is to extract structured information from unstructured text (Walker et al., 2006). Due to their complexity, annotations for REE tasks are costly and only available in a few languages. Thus, there have been growing interests on crosslingual learning for REE in which a model is trained on a language, i.e., source language, and applied to another language, i.e., target language, where the annotations are not available. Recent approaches for crosslingual REE have mainly employed multilingual word embeddings, e.g., MUSE, (Joulin et al., 2018; Ni and Florian, 2019; Liu et al., 2019; Subburathinam et al., 2019) or multilingual pre-trained language models, e.g., multilingual BERT, (Devlin et al., 2019; M'hamdi et al., 2019; Ahmad et al., 2021; Nguyen and Nguyen, 2021) to learn crosslingual representation vectors for REE.

However, previous work on crosslingual REE suffers from the monolingual bias issue due to the monolingual training of models on only the source language data, leading to non-optimal crosslingual performance. A solution for this issue can resort to language adversarial training (Chen et al., 2019; Huang et al., 2019; Keung et al., 2019; Lange et al., 2020; He et al., 2020) where unlabeled data in the target language is used to aid crosslingual representations via fooling a language discriminator. The underlying principle for this approach is to encourage the closeness of representation vectors for sentences in the source and target languages (i.e., aligning representation vectors). However, a critical drawback of language adversarial training is the failure to condition on classes/types of examples in the alignment process. As such, a target language example of a class could be incorrectly aligned to a source language example of a different class in REE, causing confusion and hindering the performance of the models. The middle sub-figure in Figure 2 demonstrates the class misalignment of representation vectors in crosslingual REE.

To this end, we propose a crosslingual alignment method that explicitly conditions on class information of REE tasks to enhance representation alignment and learning. Our major intuition is that the semantics of the classes in REE tasks (e.g., the event type of *Attack* in event extraction) are gen-

Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, pages 5414–5426
November 7–11, 2021. ©2021 Association for Computational Linguistics

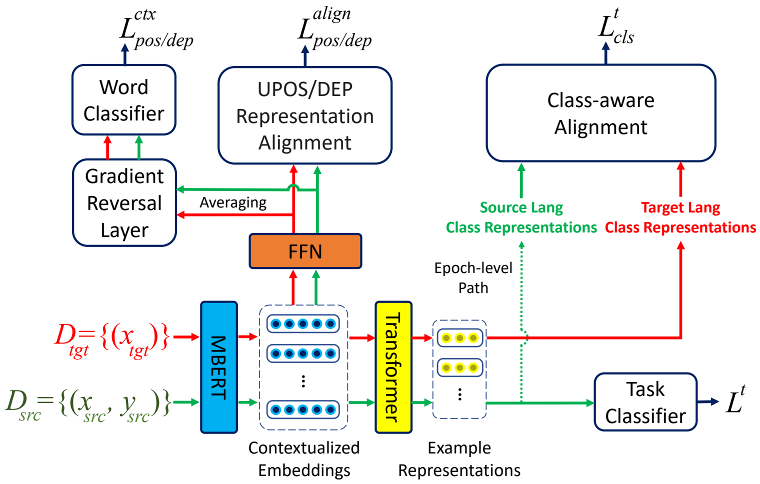

Figure 1: Overall architecture of the proposed models for RE, EAE. For ED, example representations are the contextualized embeddings.

erally invariant across languages that can be leveraged as anchors to bridge representation vectors for examples in different languages. As such, we can obtain two semantic representation vectors for each class in an REE task based on representation vectors of examples in either source or target language. Afterward, the representation vectors of the same class can be regulated to match each other, serving as a mechanism for class-aware crosslingual alignment of representation vectors for source and target examples. To implement this idea, we use multilingual BERT (mBERT) to obtain same-space representations for examples in both source and target languages to facilitate the alignment process. Afterward, the source-language representation vector for a class is computed via representation vectors of source-language examples that belong to the corresponding class. For the target language, as class information is not provided, we seek to compute target-language representation vector for a class by aggregating representation vectors for unlabeled examples, weighted on an estimation of the probabilities for the examples to exhibit the class.

In addition to class semantics, we propose to further exploit universal parts of speech and dependency relations in parsing trees (i.e., word categories) to improve the cross-lingual alignment for representation vectors in REE. As such universal word categories have been consistently annotated for more than 100 languages (Zeman et al., 2020) and can be generated with high accuracy via existing toolkits, e.g., the transformer-based toolkit Trankit for multilingual NLP (Straka, 2018; Qi et al., 2020; Nguyen et al., 2021b), we expect this information to provide helpful anchor knowledge for cross-lingual representation learning. To this end, similar to the class-aware alignment, we propose to align representation vectors of the same universal word categories that are computed using contextualized representations of examples in the source and target languages to further improve the language-independence of representation vectors for REE.

A potential issue with the computation of word category representations via contextualized representations of examples is the preservation of context word information in representations for word categories that might introduce noise and hinder the representation alignment. To address this issue, we propose an adversarial training model that seeks to explicitly filter context information from word category representations. This is achieved by using Gradient Reversal Layer (Ganin and Lempitsky, 2015) to prevent word category representations from being able to recognize the context words in the original examples. We expect that this filtering mechanism can improve the word category pureness of the representations, thus providing appropriate inputs for the alignment process for improved representation learning.

We conduct extensive experiments with different crosslingual settings on English, Chinese, and Arabic for three REE tasks, i.e., Relation Extraction, Event Detection, and Event Argument Extraction. The results demonstrate the benefits of the proposed method that significantly advances the state-of-the-art performance in these settings.

2 Problem Statement

We study cross-lingual transfer learning for three REE tasks as defined in the ACE 2005 dataset (Walker et al., 2006), i.e., Relation Extraction (RE), Event Detection (ED), and Event Argument Extraction (EAE). Given two entity mentions in an input sentence, the goal of RE is to determine the semantic relationship between the mentions according to predefined relation types/classes (e.g., *Employment*). For ED, its purpose is to identify event triggers, which can be verbs/normalization with one or multiple words, that express occurrences of events of predefined types (e.g., *Attack*). Finally, given an event trigger and an entity mention, EAE aims to predict the role (e.g., *Victim*) that the entity mention plays in the corresponding event. Note that, we have a special type *None* to indicate non-relation, non-trigger, or non-argument for RE, ED, and EAE respectively.

For further discussion, let $D_{src} = \{(x_{src}, y_{src})\}$ ($|D_{src}| = N_{src}$) be the labeled training set in the source language. As such, for ED, x_{src} is an input sentence and y_{src} serves as the golden sequence tag (using BIO) for the words in x_{src}. For RE and EAE, x_{src} involves an input sentence along with indexes of the given trigger word and entity mentions while y_{src} represents the golden relation type or argument role for the input. We also assume access to an unlabeled dataset $D_{tgt} = \{(x_{tgt})\}$ ($|D_{tgt}| = N_{tgt}$) in the target language where x_{tgt} consists of similar information as x_{src} for the corresponding task.

3 Baseline Methods

To prepare for our cross-lingual representation alignment techniques for REE, we first describe the baseline models explored in this work.

3.1 Using Source Language Data Only

In this section, we present two baselines that train models based only on labeled data in the source language. These baselines are the current state-of-the-art (SOTA) models for crosslingual transfer learning for ED, RE, and EAE on the ACE 2005 dataset (Walker et al., 2006).

BERTCRF (M'hamdi et al., 2019): This is the current SOTA model for crosslingual ED. Given an input sentence $\mathbf{w} = [w_1, w_2, \ldots, w_n]$ with n words (in x_{src}), the model first sends $\mathbf{w}$ to the mBERT encoder to obtain a sequence of contextualized representations $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_n]$ where $\mathbf{z}_k$ is the representation for each $w_k \in \mathbf{w}$, computed as the average of its word-piece representations returned by the last layer of mBERT. The ED task is then done by performing sequence labeling over the words in $\mathbf{w}$ where each word is assigned with a BIO tag to capture boundaries and event types of event triggers in $\mathbf{w}$. In particular, the final representation vector for trigger prediction $\mathbf{r}^{ED}_{src,k}$ is directly formed from the word representation $\mathbf{z}_k$ (i.e., $\mathbf{r}^{ED}_{src,k} = \mathbf{z}_k$). Afterward, this prediction representation is fed into a feed-forward network FFN^{ED} to obtain a score vector that exhibits the likelihoods for w_k to receive possible BIO tags for the predefined event types: $\mathbf{s}^{ED}_{src,k} = \text{FFN}^{ED}(\mathbf{r}^{ED}_{src,k})\ \forall 1 \leq k \leq n$.

Next, the score vectors are sent to a Conditional Random Field (CRF) layer to learn the inter-dependencies between the tags and obtain conditional probability for possible tag sequences $P^{ED}(.|\mathbf{w} = x_{src})$. The negative log-likelihood of the golden tag sequence y_{src} is then used to train the model:

$$L^{ED} = -\sum_{(x_{src}, y_{src}) \in D_{src}} \log(P^{ED}(y_{src}|x_{src})) \quad (1)$$

Finally, Viterbi decoding is employed to perform prediction in inference time.

GATE (Ahmad et al., 2021): This is the current SOTA model for crosslingual RE and EAE on the ACE 2005 dataset. Given an input sentence $\mathbf{w}$ in x_{src}, this model uses the same encoding step with mBERT in BERTCRF to obtain the contextualized representation $\mathbf{z}_k$ for each $w_k \in \mathbf{w}$. Afterward, an overall word representation vector $\mathbf{v}_k$ for w_k is formed by the concatenation: $\mathbf{v}_k = [\mathbf{z}_k; \mathbf{z}^{pos}_k; \mathbf{z}^{dep}_k]$ where $\mathbf{z}^{pos}_k$ and $\mathbf{z}^{dep}_k$ are the embeddings of the universal part of speech and the dependency relation for w_k. Here, the dependency relation for a word is obtained by retrieving the dependency relation between the word and its governor in the dependency tree. For RE, given two entity mentions, the sequence of vectors $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n]$ is then passed to a Transformer layer (Vaswani et al., 2017) along with a syntax-based attention mask to compute a final representation vector $\mathbf{r}^{RE}_{src}$ for relation prediction over the input x_{src}. Afterward, a score vector for the possible relations is computed via a feed-forward network FFN^{RE}: $\mathbf{s}^{RE}_{src} = \text{FFN}^{RE}(\mathbf{r}^{RE}_{src})$.

The score vector $\mathbf{s}^{RE}_{src}$ is then sent to a softmax layer to obtain a distribution over possible relation types for x_{src}: $P^{RE}(.|x_{src})$. Finally, to train the model, we minimize the standard negative log-likelihood of the golden label y_{src}:

$$L^{RE} = -\sum_{(x_{src}, y_{src}) \in D_{src}} \log(P^{RE}(y_{src}|x_{src})) \quad (2)$$

For EAE, given an event trigger and an entity mention, we follow the same steps above for RE to compute the representation vector for role prediction $\mathbf{r}^{EAE}_{src}$, the score vector $\mathbf{s}^{EAE}_{src}$, and the negative log-likelihood for optimization L^{EAE}.

Finally, for convenience, let $\mathbf{r}^{ED}_{tgt,k}$, $\mathbf{r}^{RE}_{tgt}$, and $\mathbf{r}^{EAE}_{tgt}$ be the final representation vectors for x_{tgt} in the unlabeled data of target language. We also have $\mathbf{s}^{ED}_{tgt,k}$, $\mathbf{s}^{RE}_{tgt}$, and $\mathbf{s}^{EAE}_{tgt}$ for the likelihood score vectors for examples in the target language. These vectors are computed in the same way as their source language counterparts in this section.

3.2 Using Unlabeled Target Language Data

To avoid the monolingual bias in the cross-lingual methods for REE in Section 3.1, our work aims to exploit unlabeled data in the target language to improve the cross-lingual representations for REE. This section presents the typical approaches for leveraging unlabeled target language data for cross-lingual transfer learning in NLP, offering additional baselines for our proposed model later.

Language Adversarial Training (LADV): To leverage unlabeled data in the target language, this method introduces a language discriminator that receives representation vectors for input sentences and predicts the language identity (i.e., source or target) of the sentences (Chen et al., 2019; Huang et al., 2019; Keung et al., 2019; Cao et al., 2020). As such, given an REE task $t \in \{ED, RE, EAE\}$, the method seeks to jointly train a model for t (i.e., those described in Section 3.1) and the language discriminator so that the induced representation vectors for t can contain necessary information for the predictions in t and be language-agnostic to better transfer knowledge across languages at the same time.

To implement this method, we first obtain a representation vector for each input sentence in the source and target language data by feeding it into mBERT to obtain word representation vectors $[\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_n]$ as in BERTCRF. Following (Keung et al., 2019), the average of such word vectors is used as the representation for the sentence in this baseline. For convenience, let $\mathbf{a}_{src}$ and $\mathbf{a}_{tgt}$ be the sentence representation vectors for the input sentences in x_{src} and x_{tgt} respectively. Also, let f^t_{lng} be the language discriminator for task t (implemented by a feed-forward network with a sigmoid activation in the end). In the next step, the representation vector $\mathbf{a}_*$ ($* \in \{src, tgt\}$) for each sentence is sent to f^t_{lng} to obtain a probability $p_* = f^t_{lng}(\mathbf{a}_*)$, indicating the likelihood that the input sentence belongs to the source language. Treating source and target language sentences as positive and negative examples, the loss for the discriminator L^{disc} is then computed via the negative log-likelihood: $L^{disc} = -\sum_{x_{src} \in D_{src}} \log(p_{x_{src}}) - \sum_{x_{tgt} \in D_{tgt}} \log(1 - p_{x_{tgt}})$. The overall joint loss to train the model for t with LADV is thus: $L = L^{task} + L^{disc}$. Note that as LADV aims to prevent the language discriminator from recognizing the language identity from sentence representation vectors, we insert the Gradient Reversal Layer (GRL) (Ganin and Lempitsky, 2015) between $\mathbf{a}_*$ and f^{task}_{lng} to reverse the gradients during the backward pass from L^{disc}. Overall, fooling the language discriminator in LADV with GRL eliminates language-specific features to improve generalization across languages for t.

mBERT Finetuning (FMBERT): Recently, it has been shown that fine-tuning multilingual pretrained language models on unlabeled data of the target language can improve the crosslingual performance for NLP tasks (Pfeiffer et al., 2020). Motivated by such prior work, this baseline exploits the unlabeled data in the target language for cross-lingual representation learning by fine-tuning mBERT on the data using mask language modeling (MLM) (Devlin et al., 2019). Afterward, the fine-tuned mBERT model is utilized in the encoders for the baseline models for REE tasks in Section 3.1.

4 Proposed Method

4.1 Class-based Alignment

An overview for the proposed model is shown in Figure 1. As described in the introduction, to avoid the potential cross-class alignment of representation vectors in the source and target language, this section presents a novel method for crosslingual representation alignment in REE where class information of tasks is explicitly employed to improve the alignment process. In particular, due to the language-universal nature of the semantics of the classes for an REE task, semantic representation vectors for a class should match each other no matter if they are computed with data from the source or target language. To this end, we seek to obtain two versions of representation vectors for each class in an REE task. One version is based on representations of examples for the source language while the other version employs representations from target language examples. The two representation versions will then be matched to achieve cross-lingual representation alignment for REE.

As such, let l be a class in an REE task t (e.g., l is a BIO tag for event types in ED). We compute the source-language representation $\mathbf{c}^t_{src,l}$ for l via the average of representation vectors for examples with label l in D_{src}. In particular, for $t = RE$ or EAE, we have:

$$\mathbf{c}^t_{src,l} = \frac{1}{N^l_{src}} \sum_{(x_{src}, y_{src})} \mathbb{1}[y_{src} = l]\mathbf{r}^t_{src} \quad (3)$$

Similarly, for $t = ED$:

$$\mathbf{c}^{ED}_{src,l} = \frac{1}{N^l_{src}} \sum_{(x_{src}, y_{src})} \sum_{k=1}^{|x_{src}|} \mathbb{1}[y_{src,k} = l]\mathbf{r}^{ED}_{src,k} \quad (4)$$

Here, $\mathbb{1}$ is the indicator function, and N^l_{src} is the number of examples (for RE and EAE) or words (for ED) in D_{src} that are annotated with label l.

In the target language, as the golden labels y_{tgt} for the examples x_{tgt} are not provided, we propose to obtain a target-language representation $\mathbf{c}^t_{tgt,l}$ by aggregating representation vectors for all examples $x_{tgt} \in D_{tgt}$. Probability estimations for examples or words to belong to class l are used as the weights for the aggregation. In particular, we obtain the probability estimations by sending the score vectors $\mathbf{s}^{ED}_{tgt,k}$, $\mathbf{s}^{RE}_{tgt}$, and $\mathbf{s}^{EAE}_{tgt}$ to a softmax layer: $\hat{\mathbf{y}}^{ED}_{tgt,k} = \text{softmax}(\mathbf{s}^{ED}_{tgt,k})$, and $\hat{\mathbf{y}}^t_{tgt} = \text{softmax}(\mathbf{s}^t_{tgt})$ (for $t = RE$ or EAE). As such, we obtain the target-language representation for l via the weighted sum of $\mathbf{r}^t_{tgt}$ (for RE and EAE):

$$\mathbf{c}^t_{tgt,l} = \frac{\sum_{x_{tgt} \in D_{tgt}} \hat{\mathbf{y}}^t_{tgt,l}\mathbf{r}^t_{tgt}}{\sum_{x_{tgt} \in D_{tgt}} \hat{\mathbf{y}}^t_{tgt,l}} \quad (5)$$

Similarly, for ED:

$$\mathbf{c}^{ED}_{tgt,l} = \frac{\sum_{x_{tgt} \in D_{tgt}} \sum_{k=1}^{|x_{tgt}|} \hat{\mathbf{y}}^{ED}_{tgt,k,l}\mathbf{r}^{ED}_{tgt,k}}{\sum_{x_{tgt} \in D_{tgt}} \sum_{k=1}^{|x_{tgt}|} \hat{\mathbf{y}}^{ED}_{tgt,k,l}} \quad (6)$$

where $\hat{\mathbf{y}}^t_{tgt,l}$ and $\hat{\mathbf{y}}^{ED}_{tgt,k,l}$ represent the likelihood score for class l in vectors $\hat{\mathbf{y}}^t_{tgt}$ and $\hat{\mathbf{y}}^{ED}_{tgt,k}$ respectively. The alignment for the representations of class l is then achieved by minimizing the negative cosine similarity of the source- and target-language vectors (i.e., for task t):

$$L^t_{cls} = -\sum_l \text{cosine}(\mathbf{c}^t_{src,l}, \mathbf{c}^t_{tgt,l}) \quad (7)$$

Adaptive Coefficient: In our implementation, we compute the source-language representations $\mathbf{c}^t_{src,l}$ for l after each training epoch while the target-language representations $\mathbf{c}^t_{tgt,l}$ are obtained for in each training minibatch. The current parameters of the models are utilized to perform such calculation. As such, the quality of the representation vectors for classes might vary along the training process of the models. In particular, later epochs might correspond to better model parameters, thus leading to more reliable class representations. To this end, we propose to apply an adaptive coefficient λ_{cls} for the class alignment loss L^t_{cls} so its impact is gradually increasing along the training: $\lambda_{cls} = \frac{2}{1+\exp(-e/E)} - 1$ where E and e are the total and current numbers of training epochs, respectively. Note that λ_{cls} is small in the early training stages and gradually increase in the process.

4.2 Word Category-based Alignment

We further exploit universal parts of speech (UPOS) and dependency relations as the language-agnostic knowledge to align crosslingual representations for REE. To achieve a fair comparison with prior work (Subburathinam et al., 2019; Ahmad et al., 2021), we employ the UDPipe toolkit (Straka and Straková, 2017) to obtain parts of speech and dependency relations for the sentences. Due to their similarity, we will only describe the UPOS-based alignment process and the dependency-based alignment can be done in the same way.

As such, we utilize an embedding table U (initialized randomly) to capture representation vectors for the possible UPOS, serving as an anchor knowledge across languages. Next, to facilitate the UPOS-based representation alignment, we compute additional representation vectors for UPOS based on representation vectors of examples in both source and target languages. In particular, for each word w_k in an input sentence $\mathbf{w}$ (from x_{src} or x_{tgt}), we send its contextualized representation $\mathbf{z}_k$ from mBERT into a feed-forward network FFN^{UPOS} to produce a representation vector $\mathbf{q}_k$ for the UPOS w^{pos}_k of $w_k \in \mathbf{w}$: $\mathbf{q}_k = \text{FFN}^{UPOS}(\mathbf{z}_k)$. Afterward, to leverage the language-universal of U, we propose to match $\mathbf{q}_k$ to the embedding vector of w^{pos}_k in U for $\mathbf{q}_k$ in both source and target language data. In other words, induced representation vectors in the source and target languages are both matched to the anchor knowledge U, providing a mechanism to align source and target representations.

To match $\mathbf{q}_k$ and U, we seek to maximize the similarity between $\mathbf{q}_k$ and the embedding of w_k^{pos} in U while minimizing $\mathbf{q}_k$'s similarities with embeddings of other UPOS at the same time. To implement this idea, we utilize the following function for minimization:

$$L_{pos}^{align} = \sum_{\mathbf{w}\in D, w_k \in \mathbf{w}} \log(\sum_{u\in O} e^{\mathbf{q}_k U[u] - \mathbf{q}_k U[w_k^{pos}]}) \quad (8)$$

where $D = D_{src} \cup D_{tgt}$, O is the set of possible UPOS, and $U[u]$ is the embedding of u in U.

Context Information Filtering: Note that L_{pos}^{align} is also the negative log-likelihood for a feed-forward classifier that uses U as the weight matrix and $\mathbf{q}_k$ as the input vector to predict the UPOS w_k^{pos} for w_k. As such, minimizing L_{pos}^{align} also serves to retain relevant information for UPOS prediction in the representation vector $\mathbf{q}_k$. However, due to the direct computation of $\mathbf{q}_k$ from the contextualized representation $\mathbf{z}_k$, it is possible that $\mathbf{q}_k$ still preserves context information from the input sentence $\mathbf{w}$. This might introduce noise into $\mathbf{q}_k$ as ideally, we expect $\mathbf{q}_k$ to focus only on information about UPOS. As such, to improve the quality of $\mathbf{q}_k$ for representation alignment, we propose to explicitly filter context information from vectors $\mathbf{q}_k$. Our main idea is to ensure that $\mathbf{q}_k$ cannot be used to recover the context words in $\mathbf{w}$. To achieve this goal, we first obtain an aggregated vector for the UPOS representation vectors in the input sentence $\mathbf{w}$: $\overline{\mathbf{q}} = \frac{1}{n}\sum_{k=1}^{n} \mathbf{q}_k$. The resulting vector is then fed into a Gradient Reversal Layer (GRL) (Ganin and Lempitsky, 2015), followed by a word classifier (i.e., a feed-forward network FFN^{ctx} with a softmax layer in the end) to compute a probability distribution over the words in our vocabulary: $\hat{\mathbf{y}}^{ctx} = \text{softmax}(\text{FFN}^{ctx}(\text{GRL}(\overline{\mathbf{q}})))$. Finally, to filter the context information from $\mathbf{q}_k$, we minimize the negative log-likelihood of the context words w_k in the input sentence $\mathbf{w}$:

$$L_{pos}^{ctx} = -\sum_{\mathbf{w}\in D_{src}\cup D_{tgt}} \sum_{w_k \in \mathbf{w}} \log(\hat{\mathbf{y}}^{ctx}[w_k]) \quad (9)$$

where $\hat{\mathbf{y}}^{ctx}[w_k]$ is the probability for word w_i in the distribution $\hat{\mathbf{y}}^{ctx}$. Note that while the minimization of the negative log-likelihood generally encourages input representations to reveal information about the prediction outputs (i.e., context words in our case), the introduction of GRL in L_{pos}^{ctx} reverses this process to discourage the context information in $\overline{\mathbf{q}}$, thus purifying $\mathbf{q}_k$ to focus on UPOS knowledge and facilitating the representation alignment.

In the next steps for universal dependency relations, we follow the same procedure for L_{pos}^{align} and L_{pos}^{ctx} to obtain the losses L_{dep}^{align} and L_{dep}^{ctx} respectively for minimization. For convenience, let $L_{pos} = L_{pos}^{align} + L_{pos}^{ctx}$ and $L_{dep} = L_{dep}^{align} + L_{dep}^{ctx}$. In summary, the overall loss function to train our models for a task $t \in \{ED, RE, EAE\}$ with both class and word category alignment is thus: $L^{main} = L^t + \lambda_{cls} L_{cls}^t + \lambda_{pos} L_{pos} + \lambda_{dep} L_{dep}$ where λ_{cls} is the adaptive coefficient, and λ_{pos} and λ_{dep} are trade-off parameters.

Language	Data	RE (**#rels**)	ED (**#trgs**)	EAE (**#args**)
English	Train	4,974	4,420	7,018
	Dev	626	505	877
	Test	620	424	878
Chinese	Train	4,767	2,213	5,931
	Dev	572	111	741
	Test	605	197	742
Arabic	Train	2,918	1,986	3,959
	Dev	357	112	495
	Test	378	169	495

Table 1: Statistics of the multilingual datasets for ED, RE, and EAE in ACE 2005. **#rels**, **#trgs** and **#args** represent the numbers of relations, event triggers, and event arguments respectively.

5 Experiments

Datasets and Hyper-parameters: Following previous work (M'hamdi et al., 2019; Subburathinam et al., 2019; Ahmad et al., 2021), we use the multilingual dataset ACE 2005 (Walker et al., 2006) to evaluate REE models in this work. ACE 2005 annotate documents for entity mentions, event triggers, relations, and arguments in English (EN), Chinese (ZH) and Arabic (AR). We apply the same data split and preprocessing for ACE 2005 as prior work (M'hamdi et al., 2019; Ahmad et al., 2021) for a fair comparison. Overall, there are 18 relation types, 33 event types, and 35 argument roles in this dataset. For each of the language (i.e., English, Chinese and Arabic) and task (i.e., ED, RE, and EAE), the data split provides training, development, and test data. In our cross-lingual transfer learning experiments, the models will be trained on the training data of one language (the source) and evaluated on the test data of another language (the target). The unlabeled data for the target language is obtained by removing the labels from its training data. The statistics of the ACE 2005 dataset for the three tasks are shown in Table 1.

Model	Even Argument Extraction						Relation Extraction					
	EN ZH	EN AR	ZH EN	ZH AR	AR EN	AR ZH	EN ZH	EN AR	ZH EN	ZH AR	AR EN	AR ZH
GATE	63.2	68.5	59.3	69.2	53.9	57.8	55.1	66.8	71.5	61.2	69.0	54.3
GATE+LADV	63.9	67.7	60.3	68.6	55.8	57.8	56.8	64.2	70.2	61.6	68.9	54.8
GATE+FMBERT	63.7	68.7	59.3	**69.3**	54.6	58.1	55.8	66.9	71.8	61.7	69.2	54.9
GATE+CCCAR	**65.5**	**69.4**	**62.0**	**69.3**	**57.5**	**59.1**	**58.1**	**67.9**	**72.0**	**63.5**	**70.5**	**57.7**

Table 2: Performance (F1 scores) of models on test data for EAE and RE in six crosslingual settings. Each column corresponds to one setting where source languages are written above target languages. Underlined numbers designate settings where the proposed model is significantly better than other models with $p < 0.01$.

Model	Event Detection					
	EN ZH	EN AR	ZH EN	ZH AR	AR EN	AR ZH
BERTCRF	68.5	30.9	-	-	-	-
BERTCRF+LADV	70.0	33.5	41.2	20.3	37.2	55.6
BERTCRF+FMBERT	69.4	33.4	42.9	20.0	36.5	56.3
BERTCRF+CCCAR	**72.1**	**42.7**	**45.8**	**20.7**	**40.7**	**59.8**

Table 3: Performance (F1 scores) on test data for ED in six crosslingual settings. Each column corresponds to one setting where source languages are written above target languages. "-" indicates results that are not reported in the original work. Underlined numbers designate settings where the proposed model is significantly better than other models with $p < 0.01$.

We use the same hyper-parameters for BERTCRF and GATE as provided by previous work (M'hamdi et al., 2019; Ahmad et al., 2021). Specific hyper-parameters for our model are tuned on the development data. In particular, we use two layers for the feed forward networks with 50 hidden units for the layers, 50 dimensions for the UPOS and dependency embeddings, and 0.1 for the parameters λ_{pos} and λ_{dep}. For the baseline FMBERT, we utilize the *huggingface* library to finetune mBERT on unlabeled target data with MLM for $100,000$ steps (i.e., batch size of 64 and learning rate of $5e$-5).

Performance Comparison: We compare the proposed crosslingual method for REE on two groups of baselines. The first group involve models that only use source language data for training, i.e., BERTCRF and GATE. These are current SOTA methods for crosslingual ED, RE, and EAE. The second baseline groups additionally employ unlabeled data in the target language to support crosslingual representation learning in REE, i.e., LADV and FMBERT. Our proposed method also leverages unlabeled data in the target language, called CCCAR for class- and word category-based crosslingual alignment of representations. Note that LADV, FMBERT, and CCCAR should be applied on top of a source-only method (i.e., BERTCRF and GATE) to form a complete model.

Tables 3 and 2 show the test data performance of the models for the three REE tasks in six crosslingual settings (i.e., with different pairs of languages for the source and target). It is clear from the tables that the proposed method CCCAR consistently outperforms other methods in all crosslingual settings for the three REE tasks. In particular, for EAE, CCCAR substantially improves the baseline model GATE (i.e., the current SOTA) by 1.9% on average while those improvement for LADV and FMBERT are only 0.45% and 0.38%. The same trend can be seen for RE and ED where CCCAR on average improves the baselines by 1.97% for the former and 7.7% for the latter. These results clearly demonstrate the effectiveness of the proposed method, highlighting the benefits of the class- and word category-based alignment for crosslingual REE.

Model	English → Chinese			English → Arabic		
	RE	ED	EAE	RE	ED	EAE
CCCAR	**58.1**	**72.1**	**65.5**	**67.9**	**42.7**	**69.4**
- Class Align.	56.6	69.9	63.6	66.9	38.8	68.9
- Adaptive Coeff.	57.4	71.5	64.7	67.3	41.3	69.2
- UPOS Align.	57.9	71.4	65.1	66.9	40.4	69.3
- Dep Align.	57.8	71.7	64.7	67.1	41.5	68.9
- Word Cat Align.	57.0	70.9	64.4	67.0	40.0	68.7
- Context Filtering	57.6	71.2	64.9	67.4	41.6	69.0

Table 4: Performance (F1 scores) of models. In the row for the proposed model CCCAR, we use BERTCRF as the base model for ED, and GATE as the base model for RE and EAE.

Ablation Study: This section conducts an ablation study to understand the contribution of each designed component in the proposed crosslingual alignment method CCCAR. In particular, we examine the performance of the following ablated models: (i) **- Class Align.**: this model excludes the class-based alignment component (i.e., the loss L^t_{cls}) from CCCAR; (ii) **- Adaptive Coeff.**: instead of using the adaptive coefficient λ_{cls} for the class-based alignment loss L^t_{cls}, this model utilizes a fixed value (i.e., 0.2 as tuned on development data) for λ_{cls}; (iii) **- UPOS Align.**: this model eliminates the UPOS-based alignment component (i.e., the losses L^{align}_{pos} and L^{ctx}_{pos}) from CCCAR; (iv) **-**

Dep Align.: the alignment component based on dependency relations (i.e., the losses L_{dep}^{align} and L_{dep}^{ctx}) is not utilized in this model; (v) **- Word Cat Align.**: this model removes both UPOS-based and dependency-based alignment from CCCAR (i.e., excluding L_{pos} and L_{dep}); and (vi) **- Context Filtering**: the word context filtering for the representation vectors of UPOS and dependency relations (with GRL) is not employed in this model (i.e., eliminating the losses L_{pos}^{ctx} and L_{dep}^{ctx}).

Table 4 presents the test data performance of the models in the English-to-Chinese and English-to-Arabic settings for the three REE tasks. As can be seen, removing any component of the proposed model would hurt the performance significantly across different settings and tasks, thus clearly illustrating the benefits of the designed components for CCCAR. The performance of the models drops the most when the class-based alignment is excluded, further demonstrating the importance of class-aware alignment for crosslingual REE.

Source-language Data Usage: Previous experiments show that using unlabeled data in the target language to align representation vectors in CCCAR can improve the performance for the source-only baselines for REE. In this section, we seek to understand how much labeled data in the source language can be saved if unlabeled data in the target language is employed with CCCAR for an REE task. In particular, we are interested in the portion of source language data that, once combined with unlabeled target language data via CCCAR, can produce similar performance as the source-only baseline trained on full source language data. To this end, we show the learning curves of the source-only and CCCAR-augmented models for REE tasks when the size of the source-language training data varies. Figure 3 show the curves for the English-to-Chinese setting. As can be seen, the proposed CCCAR method with unlabeled target data only needs to use approximately 60% of the source-language training data for RE and EAE to achieve comparable performance with the source-only baselines on full source language data. This portion for ED is less than 80%. These results thus suggests an additional benefit of CCCAR to significantly reduce necessary data annotation for the source language based on unlabeled target language data in crosslingual learning for REE.

Alignment Effect of the Proposed Method: As discussed earlier, a major issue for LADV is that it might align representations of examples with different classes in the crosslingual setting. CCCAR can address this issue as it explicitly relies on class information for representation alignment. To demonstrate these arguments, Figure 2 uses the t-Distributed Stochastic Neighbor Embedding (t-SNE) (Van der Maaten and Hinton, 2008) to visualize the example representations induced by GATE, the LADV baseline GATE+LADV, and the proposed GATE+CCCAR. This visualization is done over 4,000 randomly selected examples for the top 5 frequent classes in EAE. Here, examples are sampled from training data for both source and target languages in the English-to-Chinese setting. As can be seen, in the source-only model GATE, representations for examples from the source language are quite separate from those in the target language. The representation alignment in GATE+LADV can address this issue by pushing representations from both languages closer. However, representations for examples with different classes are unexpectedly aligned in GATE+LADV, causing suboptimal representations for crosslingual settings. Finally, due to the explicit condition on class information for alignment, GATE+CCCAR can match representations for both languages while avoiding the cross-class alignment to improve crosslingual performance for REE.

6 Related Work

REE has been extensively studied for English, featuring traditional machine learning methods (Patwardhan and Riloff, 2009; Liao and Grishman, 2011; Li et al., 2013; Yang and Mitchell, 2016) and advanced deep learning models (Nguyen and Grishman, 2015; Chen et al., 2015; Nguyen et al., 2016a; Nguyen and Grishman, 2018; Wang et al., 2019; Zhang et al., 2019; Sahu et al., 2019; Veyseh et al., 2020b,a,c; Lin et al., 2020; Nguyen et al., 2021a). Recently, several works have considered cross-lingual transfer learning for three REE tasks (Ni and Florian, 2019; Liu et al., 2019; Subburathinam et al., 2019) where multilingual pre-trained language models (e.g., mBERT) have been proved as an important encoding component (Ahmad et al., 2021; Nguyen and Nguyen, 2021).

However, a fundamental limitation of existing crosslingual models for REE is the monolingual bias due to the sole reliance on source language data for training. In other NLP tasks, LADV has been explored to address this issue by leveraging

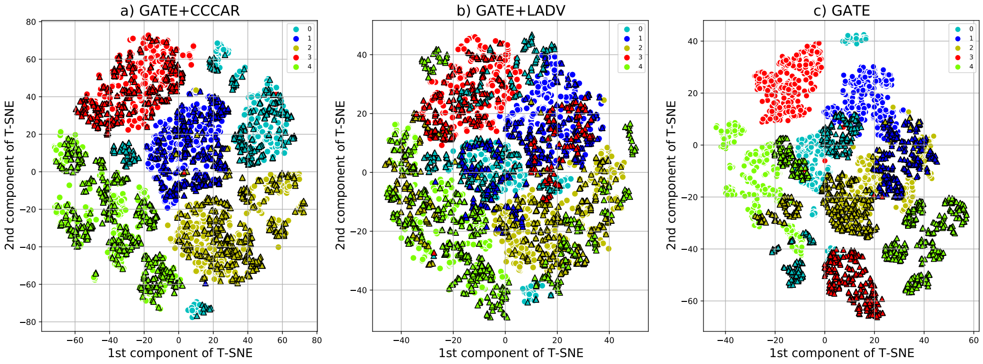

Figure 2: T-SNE visualizations for the representations of 4,000 randomly selected examples from English (i.e., source language) and Chinese (i.e., target language) data. Circles and triangles represent English and Chinese examples respectively. Colors represent different classes in EAE. GATE+CCCAR shows induced representation vectors from our proposed model.

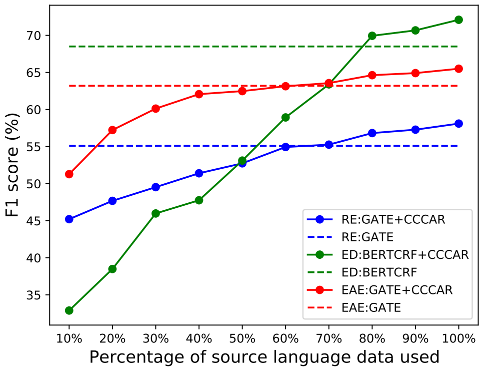

Figure 3: Performance on test data of the models in the English-to-Chinese setting. Dash lines represent the performance of the source-only baselines using 100% of the source-language training data.

unlabeled data in the target language to perform crosslingual representation alignment (Chen et al., 2019; Huang et al., 2019; Lange et al., 2020; Cao et al., 2020; He et al., 2020). Unfortunately, LADV suffers from the cross-class alignment issue, making it less optimal for crosslingual REE. Finally, we note that language-universal representation learning is related to domain adaption research where models seek to learn domain-invariant representations (Ganin and Lempitsky, 2015; Fu et al., 2017; Adel et al., 2017; Xie et al., 2018; Cicek and Soatto, 2019; Tang et al., 2020; Ngo et al., 2021).

7 Conclusions

We present a novel method for crosslingual transfer learning for REE that leverages unlabeled data in the target language to support language-universal representation learning. Our method exploits class semantics in REE tasks and universal word categories (i.e., UPOS and dependency relations) as bridges to align representation vectors across languages. In our method, representation vectors for classes and word categories are computed via contextualized representations of examples to implement representation matching for crosslingual alignment. Extensive experiments show that the proposed method achieves SOTA performance for three REE tasks in different crosslingual settings. In the future, we plan to extend our methods to related problems in IE (e.g., coreference resolution).

Acknowledgments

This research has been supported by the Army Research Office (ARO) grant W911NF-21-1-0112 and the NSF grant CNS-1747798 to the IUCRC Center for Big Learning. 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

Tameem Adel, Han Zhao, and Alexander Wong. 2017. Unsupervised domain adaptation with a relaxed covariate shift assumption. In *Proceedings of the Association for the Advancement of Artificial Intelligence (AAAI)*.

Wasi Uddin Ahmad, Nanyun Peng, and Kai-Wei Chang. 2021. Gate: Graph attention transformer encoder for cross-lingual relation and event extraction. In *Proceedings of the Association for the Advancement of Artificial Intelligence (AAAI)*.

Yue Cao, Hui Liu, and Xiaojun Wan. 2020. Jointly learning to align and summarize for neural cross-lingual summarization. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*.

Xilun Chen, Ahmed Hassan Awadallah, Hany Hassan, Wei Wang, and Claire Cardie. 2019. Multi-source cross-lingual model transfer: Learning what to share. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL)*.

Yubo Chen, Liheng Xu, Kang Liu, Daojian Zeng, and Jun Zhao. 2015. Event extraction via dynamic multi-pooling convolutional neural networks. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*.

Safa Cicek and Stefano Soatto. 2019. Unsupervised domain adaptation via regularized conditional alignment. In *Proceedings of the International Conference on Computer Vision (ICCV)*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*.

Lisheng Fu, Thien Huu Nguyen, Bonan Min, and Ralph Grishman. 2017. Domain adaptation for relation extraction with domain adversarial neural network. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (IJCNLP)*.

Yaroslav Ganin and Victor Lempitsky. 2015. Unsupervised domain adaptation by backpropagation. In *Proceedings of the International Conference on Machine Learning (ICML)*.

Keqing He, Yuanmeng Yan, and Weiran Xu. 2020. Adversarial cross-lingual transfer learning for slot tagging of low-resource languages. In *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*.

Lifu Huang, Heng Ji, and Jonathan May. 2019. Cross-lingual multi-level adversarial transfer to enhance low-resource name tagging. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*.

Armand Joulin, Piotr Bojanowski, Tomas Mikolov, Hervé Jégou, and Edouard Grave. 2018. Loss in translation: Learning bilingual word mapping with a retrieval criterion. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Phillip Keung, Yichao Lu, and Vikas Bhardwaj. 2019. Adversarial learning with contextual embeddings for zero-resource cross-lingual classification and NER. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*.

Lukas Lange, Anastasiia Iurshina, Heike Adel, and Jannik Strötgen. 2020. Adversarial alignment of multilingual models for extracting temporal expressions from text. *arXiv preprint arXiv:2005.09392*.

Qi Li, Heng Ji, and Liang Huang. 2013. Joint event extraction via structured prediction with global features. In *Proceedings of the 51th Annual Meeting of the Association for Computational Linguistics (ACL)*.

Shasha Liao and Ralph Grishman. 2011. Acquiring topic features to improve event extraction: in pre-selected and balanced collections. In *Proceedings of the International Conference Recent Advances in Natural Language Processing (RANLP)*.

Ying Lin, Heng Ji, Fei Huang, and Lingfei Wu. 2020. A joint neural model for information extraction with global features. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*.

Jian Liu, Yubo Chen, Kang Liu, and Jun Zhao. 2019. Neural cross-lingual event detection with minimal parallel resources. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*.

Meryem M'hamdi, Marjorie Freedman, and Jonathan May. 2019. Contextualized cross-lingual event trigger extraction with minimal resources. In *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*.

Nghia Trung Ngo, Duy Phung, and Thien Huu Nguyen. 2021. Unsupervised domain adaptation for event detection using domain-specific adapters. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*.

Minh Van Nguyen, Viet Dac Lai, and Thien Huu Nguyen. 2021a. Cross-task instance representation interactions and label dependencies for joint information extraction with graph convolutional networks. In *Proceedings of the Conference of the*

North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT).

Minh Van Nguyen, Viet Dac Lai, Amir Pouran Ben Veyseh, and Thien Huu Nguyen. 2021b. Trankit: A light-weight transformer-based toolkit for multilingual natural language processing. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*.

Minh Van Nguyen and Thien Huu Nguyen. 2021. Improving cross-lingual transfer for event argument extraction with language-universal sentence structures. In *Proceedings of the 6th Arabic Natural Language Processing Workshop at EACL 2021 (WANLP@EACL 2021)*.

Thien Huu Nguyen, Kyunghyun Cho, and Ralph Grishman. 2016a. Joint event extraction via recurrent neural networks. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*.

Thien Huu Nguyen and Ralph Grishman. 2015. Event detection and domain adaptation with convolutional neural networks. In *The 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (ACL-IJCNLP)*.

Thien Huu Nguyen and Ralph Grishman. 2018. Graph convolutional networks with argument-aware pooling for event detection. In *Proceedings of the Association for the Advancement of Artificial Intelligence (AAAI)*.

Jian Ni and Radu Florian. 2019. Neural cross-lingual relation extraction based on bilingual word embedding mapping. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*.

Siddharth Patwardhan and Ellen Riloff. 2009. A unified model of phrasal and sentential evidence for information extraction. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*.

Jonas Pfeiffer, Ivan Vulić, Iryna Gurevych, and Sebastian Ruder. 2020. MAD-X: An Adapter-Based Framework for Multi-Task Cross-Lingual Transfer. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D Manning. 2020. Stanza: A python natural language processing toolkit for many human languages. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*.

Sunil Kumar Sahu, Fenia Christopoulou, Makoto Miwa, and Sophia Ananiadou. 2019. Inter-sentence relation extraction with document-level graph convolutional neural network. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*.

Milan Straka. 2018. Udpipe 2.0 prototype at conll 2018 ud shared task. In *Proceedings of the Conference on Computational Natural Language Learning (CoNLL)*.

Milan Straka and Jana Straková. 2017. Tokenizing, POS tagging, lemmatizing and parsing UD 2.0 with UDPipe. In *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*.

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 *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*.

Hui Tang, Ke Chen, and Kui Jia. 2020. Unsupervised domain adaptation via structurally regularized deep clustering. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*.

Laurens Van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-sne. *Journal of machine learning research*, 9(11).

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Proceedings of the Conference on Neural Information Processing Systems (NeurIPS)*.

Amir Pouran Ben Veyseh, Franck Dernoncourt, Dejing Dou, and Thien Huu Nguyen. 2020a. Exploiting the syntax-model consistency for neural relation extraction. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*.

Amir Pouran Ben Veyseh, Franck Dernoncourt, My Thai, Dejing Dou, and Thien Huu Nguyen. 2020b. Multi-view consistency for relation extraction via mutual information and structure prediction. In *Proceedings of the Association for the Advancement of Artificial Intelligence (AAAI)*.

Amir Pouran Ben Veyseh, Tuan Ngo Nguyen, and Thien Huu Nguyen. 2020c. Graph transformer networks with syntactic and semantic structures for event argument extraction. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing: Findings (EMNLP)*.

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

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 *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*.

Shaoan Xie, Zibin Zheng, Liang Chen, and Chuan Chen. 2018. Learning semantic representations for unsupervised domain adaptation. In *Proceedings of the International Conference on Machine Learning (ICML)*.

Bishan Yang and Tom M. Mitchell. 2016. Joint extraction of events and entities within a document context. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*.

Daniel Zeman, Joakim Nivre, Mitchell Abrams, Elia Ackermann, Noëmi Aepli, Hamid Aghaei, Željko Agić, Amir Ahmadi, Lars Ahrenberg, Chika Kennedy Ajede, Gabrielė Aleksandravičiūtė, Ika Alfina, Lene Antonsen, Katya Aplonova, Angelina Aquino, Carolina Aragon, Maria Jesus Aranzabe, Hórunn Arnardóttir, Gashaw Arutie, Jessica Naraiswari Arwidarasti, Masayuki Asahara, Luma Ateyah, Furkan Atmaca, Mohammed Attia, Aitziber Atutxa, Liesbeth Augustinus, Elena Badmaeva, Keerthana Balasubramani, Miguel Ballesteros, Esha Banerjee, Sebastian Bank, Verginica Barbu Mititelu, Victoria Basmov, Colin Batchelor, John Bauer, Seyyit Talha Bedir, Kepa Bengoetxea, Gözde Berk, Yevgeni Berzak, Irshad Ahmad Bhat, Riyaz Ahmad Bhat, Erica Biagetti, Eckhard Bick, Agnė Bielinskienė, Kristín Bjarnadóttir, Rogier Blokland, Victoria Bobicev, Loïc Boizou, Emanuel Borges Völker, Carl Börstell, Cristina Bosco, Gosse Bouma, Sam Bowman, Adriane Boyd, Kristina Brokaitė, Aljoscha Burchardt, Marie Candito, Bernard Caron, Gauthier Caron, Tatiana Cavalcanti, Gülşen Cebiroğlu Eryiğit, Flavio Massimiliano Cecchini, Giuseppe G. A. Celano, Slavomír Čéplö, Savas Cetin, Özlem Çetinoğlu, Fabricio Chalub, Ethan Chi, Yongseok Cho, Jinho Choi, Jayeol Chun, Alessandra T. Cignarella, Silvie Cinková, Aurélie Collomb, Çağrı Çöltekin, Miriam Connor, Marine Courtin, Elizabeth Davidson, Marie-Catherine de Marneffe, Valeria de Paiva, Mehmet Oguz Derin, Elvis de Souza, Arantza Diaz de Ilarraza, Carly Dickerson, Arawinda Dinakaramani, Bamba Dione, Peter Dirix, Kaja Dobrovoljc, Timothy Dozat, Kira Droganova, Puneet Dwivedi, Hanne Eckhoff, Marhaba Eli, Ali Elkahky, Binyam Ephrem, Olga Erina, Tomaž Erjavec, Aline Etienne, Wograine Evelyn, Sidney Facundes, Richárd Farkas, Marília Fernanda, Hector Fernandez Alcalde, Jennifer Foster, Cláudia Freitas, Kazunori Fujita, Katarína Gajdošová, Daniel Galbraith, Marcos Garcia, Moa Gärdenfors, Sebastian Garza, Fabrício Ferraz Gerardi, Kim Gerdes, Filip Ginter, Iakes Goenaga, Koldo Gojenola, Memduh Gökırmak, Yoav Goldberg, Xavier Gómez Guinovart, Berta González Saavedra, Bernadeta Griciūtė, Matias Grioni, Loïc Grobol, Normunds Grūzītis, Bruno Guillaume, Céline Guillot-Barbance, Tunga Güngör, Nizar Habash, Hinrik Hafsteinsson, Jan Hajič, Jan Hajič jr., Mika Hämäläinen, Linh Hà Mỹ, Na-Rae Han, Muhammad Yudistira Hanifmuti, Sam Hardwick, Kim Harris, Dag Haug, Johannes Heinecke, Oliver Hellwig, Felix Hennig, Barbora Hladká, Jaroslava Hlaváčová, Florinel Hociung, Petter Hohle, Eva Huber, Jena Hwang, Takumi Ikeda, Anton Karl Ingason, Radu Ion, Elena Irimia, Ọlájídé Ishola, Tomáš Jelínek, Anders Johannsen, Hildur Jónsdóttir, Fredrik Jørgensen, Markus Juutinen, Sarveswaran K, Hüner Kaşıkara, Andre Kaasen, Nadezhda Kabaeva, Sylvain Kahane, Hiroshi Kanayama, Jenna Kanerva, Boris Katz, Tolga Kayadelen, Jessica Kenney, Václava Kettnerová, Jesse Kirchner, Elena Klementieva, Arne Köhn, Abdullatif Köksal, Kamil Kopacewicz, Timo Korkiakangas, Natalia Kotsyba, Jolanta Kovalevskaitė, Simon Krek, Parameswari Krishnamurthy, Sookyoung Kwak, Veronika Laippala, Lucia Lam, Lorenzo Lambertino, Tatiana Lando, Septina Dian Larasati, Alexei Lavrentiev, John Lee, Phuong Lê Hồng, Alessandro Lenci, Saran Lertpradit, Herman Leung, Maria Levina, Cheuk Ying Li, Josie Li, Keying Li, Yuan Li, KyungTae Lim, Krister Lindén, Nikola Ljubešić, Olga Loginova, Andry Luthfi, Mikko Luukko, Olga Lyashevskaya, Teresa Lynn, Vivien Macketanz, Aibek Makazhanov, Michael Mandl, Christopher Manning, Ruli Manurung, Cătălina Mărănduc, David Mareček, Katrin Marheinecke, Héctor Martínez Alonso, André Martins, Jan Mašek, Hiroshi Matsuda, Yuji Matsumoto, Ryan McDonald, Sarah McGuinness, Gustavo Mendonça, Niko Miekka, Karina Mischenkova, Margarita Misirpashayeva, Anna Missilä, Cătălin Mititelu, Maria Mitrofan, Yusuke Miyao, AmirHossein Mojiri Foroushani, Amirsaeid Moloodi, Simonetta Montemagni, Amir More, Laura Moreno Romero, Keiko Sophie Mori, Shinsuke Mori, Tomohiko Morioka, Shigeki Moro, Bjartur Mortensen, Bohdan Moskalevskyi, Kadri Muischnek, Robert Munro, Yugo Murawaki, Kaili Müürisep, Pinkey Nainwani, Mariam Nakhlé, Juan Ignacio Navarro Horñiacek, Anna Nedoluzhko, Gunta Nešpore-Bērzkalne, Luong Nguyễn Thị, Huyền Nguyễn Thị Minh, Yoshihiro Nikaido, Vitaly Nikolaev, Rattima Nitisaroj, Alireza Nourian, Hanna Nurmi, Stina Ojala, Atul Kr. Ojha, Adedayo Oluokun, Mai Omura, Emeka Onwuegbuzia, Petya Osenova, Robert Östling, Lilja Øvrelid, Şaziye Betül Özateş, Arzucan Özgür, Balkız Öztürk Başaran, Niko Partanen, Elena Pascual, Marco Passarotti, Agnieszka Patejuk, Guilherme Paulino-Passos, Angelika Peljak-Łapińska, Siyao Peng, Cenel-Augusto Perez, Natalia Perkova, Guy Perrier, Slav Petrov, Daria Petrova, Jason Phelan, Jussi Piitulainen, Tommi A Pirinen, Emily Pitler, Barbara Plank, Thierry Poibeau, Larisa Ponomareva, Martin Popel, Lauma Pretkalniņa, Sophie Prévost, Prokopis Prokopidis, Adam Przepiórkowski, Tiina

Puolakainen, Sampo Pyysalo, Peng Qi, Andriela Rääbis, Alexandre Rademaker, Taraka Rama, Loganathan Ramasamy, Carlos Ramisch, Fam Rashel, Mohammad Sadegh Rasooli, Vinit Ravishankar, Livy Real, Petru Rebeja, Siva Reddy, Georg Rehm, Ivan Riabov, Michael Rießler, Erika Rimkutė, Larissa Rinaldi, Laura Rituma, Luisa Rocha, Eiríkur Rögnvaldsson, Mykhailo Romanenko, Rudolf Rosa, Valentin Roșca, Davide Rovati, Olga Rudina, Jack Rueter, Kristján Rúnarsson, Shoval Sadde, Pegah Safari, Benoît Sagot, Aleksi Sahala, Shadi Saleh, Alessio Salomoni, Tanja Samardžić, Stephanie Samson, Manuela Sanguinetti, Dage Särg, Baiba Saulīte, Yanin Sawanakunanon, Kevin Scannell, Salvatore Scarlata, Nathan Schneider, Sebastian Schuster, Djamé Seddah, Wolfgang Seeker, Mojgan Seraji, Mo Shen, Atsuko Shimada, Hiroyuki Shirasu, Muh Shohibussirri, Dmitry Sichinava, Einar Freyr Sigurdsson, Aline Silveira, Natalia Silveira, Maria Simi, Radu Simionescu, Katalin Simkó, Mária Šimková, Kiril Simov, Maria Skachedubova, Aaron Smith, Isabela Soares-Bastos, Carolyn Spadine, Steinhór Steingrímsson, Antonio Stella, Milan Straka, Emmett Strickland, Jana Strnadová, Alane Suhr, Yogi Lesmana Sulestio, Umut Sulubacak, Shingo Suzuki, Zsolt Szántó, Dima Taji, Yuta Takahashi, Fabio Tamburini, Mary Ann C. Tan, Takaaki Tanaka, Samson Tella, Isabelle Tellier, Guillaume Thomas, Liisi Torga, Marsida Toska, Trond Trosterud, Anna Trukhina, Reut Tsarfaty, Utku Türk, Francis Tyers, Sumire Uematsu, Roman Untilov, Zdeňka Urešová, Larraitz Uria, Hans Uszkoreit, Andrius Utka, Sowmya Vajjala, Daniel van Niekerk, Gertjan van Noord, Viktor Varga, Eric Villemonte de la Clergerie, Veronika Vincze, Aya Wakasa, Joel C. Wallenberg, Lars Wallin, Abigail Walsh, Jing Xian Wang, Jonathan North Washington, Maximilan Wendt, Paul Widmer, Seyi Williams, Mats Wirén, Christian Wittern, Tsegay Woldemariam, Tak-sum Wong, Alina Wróblewska, Mary Yako, Kayo Yamashita, Naoki Yamazaki, Chunxiao Yan, Koichi Yasuoka, Marat M. Yavrumyan, Zhuoran Yu, Zdeněk Žabokrtský, Shorouq Zahra, Amir Zeldes, Hanzhi Zhu, and Anna Zhuravleva. 2020. Universal dependencies 2.7. LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.

Junchi Zhang, Yanxia Qin, Yue Zhang, Mengchi Liu, and Donghong Ji. 2019. Extracting entities and events as a single task using a transition-based neural model. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*.