Cross-lingual Structure Transfer for Relation and Event Extraction

Ananya Subburathinam[1]*, Di Lu[1]* , Heng Ji[2],
Jonathan May[3], Shih-Fu Chang[4], Avirup Sil[5], Clare Voss[6]
[1]Computer Science Department, Rensselaer Polytechnic Institute
[2]Computer Science Department, University of Illinois at Urbana-Champaign
[3]University of Southern California Information Sciences Institute
[4]Electrical Engineering Department, Columbia University
[5]IBM T.J. Watson Research Center
[6]Army Research Laboratory
hengji@illinois.edu

Abstract

The identification of complex semantic structures such as events and entity relations, already a challenging Information Extraction task, is doubly difficult from sources written in under-resourced and under-annotated languages. We investigate the suitability of cross-lingual structure transfer techniques for these tasks. We exploit relation- and event-relevant language-universal features, leveraging both symbolic (including part-of-speech and dependency path) and distributional (including type representation and contextualized representation) information. By representing all entity mentions, event triggers, and contexts into this complex and structured multilingual common space, using graph convolutional networks, we can train a relation or event extractor from source language annotations and apply it to the target language. Extensive experiments on cross-lingual relation and event transfer among English, Chinese, and Arabic demonstrate that our approach achieves performance comparable to state-of-the-art supervised models trained on up to 3,000 manually annotated mentions: up to 62.6% F-score for Relation Extraction, and 63.1% F-score for Event Argument Role Labeling. The event argument role labeling model transferred from English to Chinese achieves similar performance as the model trained from Chinese. We thus find that language-universal symbolic and distributional representations are complementary for cross-lingual structure transfer.

1 Introduction

Advanced Information Extraction (IE) tasks entail predicting structures, such as relations between entities, and events involving entities. Given a pair of entity mentions, *Relation Extraction* aims to identify the relation between the mentions and classify it by predefined type. *Event Extraction* aims to identify event triggers and their arguments in
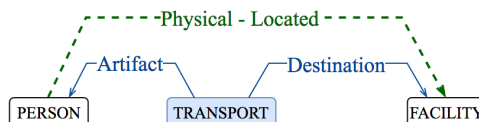
Figure 1: Information Network example constructed by Event (blue) and Relation (green dashed) Extraction.

unstructured texts and classify them respectively by predefined types and roles. As Figure 1 illustrates, both tasks entail predicting an information network (Li et al., 2014) for each sentence, where the entity mentions and event triggers are nodes, and the relations and event-argument links are edges labeled with their relation and argument roles, respectively.

There are certain relations and events that are of primary interest to a given community and so are reported predominantly in the low-resource language data sources available to that community. For example, though English language news will occasionally discuss the *Person* Aung San Suu Kyi, the vast majority of *Physical-Located* relations and *Meeting* events involving her are only reported locally in Burmese news, and thus, without Burmese relation and event extraction, a knowledge graph of this person will lack this available information. Unfortunately, publicly-available gold-standard annotations for relation and event extraction exist for only a few languages (Doddington et al., 2004; Getman et al., 2018), and Burmese is not among them. Compared to other IE tasks such as name tagging, the annotations for Relation and Event Extraction are also more costly to obtain, because they are structured and require a rich label space.

Recent research (Lin et al., 2017) has found that relational facts are typically expressed by identifiable patterns within languages and has shown
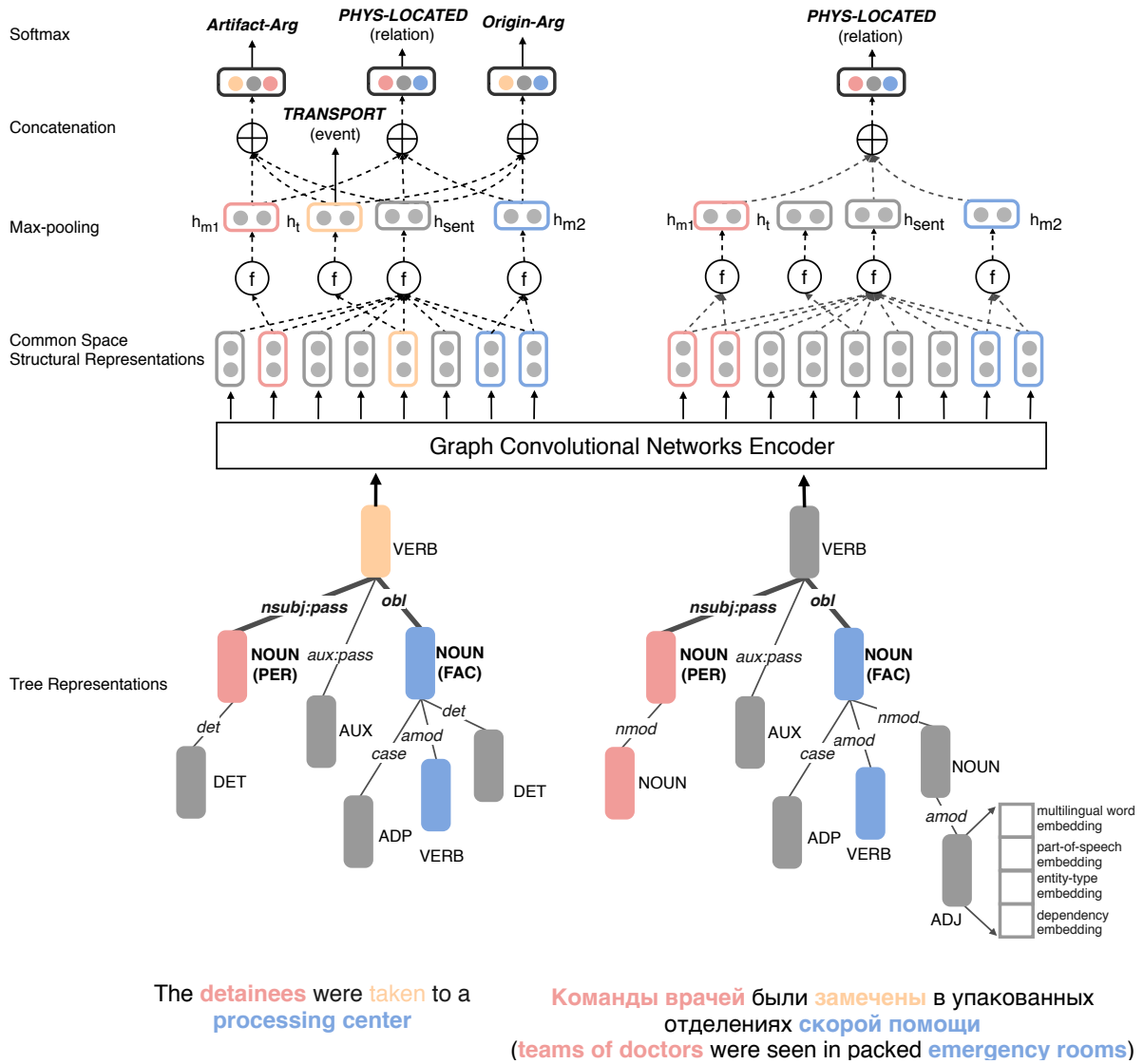
Figure 2: Multilingual common semantic space and cross-lingual structure transfer.

that the consistency in patterns observed across languages can be used to improve relation extraction. Inspired by their results, we exploit language-universal features relevant to relation and event argument identification and classification, by way of both symbolic and distributional representations. For example, language-universal POS tagging and universal dependency parsing is available for 76 languages (Nivre et al., 2018), entity extraction is available for 282 languages (Pan et al., 2017), and multi-lingual word embeddings are available for 44 languages (Bojanowski et al., 2017; Joulin et al., 2018). As shown in Figure 2, even for distinct pairs of entity mentions (colored pink and blue, in both English and Russian), the structures share similar language-universal symbolic features, such as a common labeled depen-

dency path, as well as distributional features, such as multilingual word embeddings.

Based on these language-universal representations, we then project all entity mentions, event triggers and their contexts into one multilingual common space. Unlike recent work on multilingual common space construction that makes use of linear mappings (Mikolov et al., 2013; Rothe et al., 2016; Zhang et al., 2016; Lazaridou et al., 2015; Xing et al., 2015; Smith et al., 2017) or canonical correlation analysis (CCA) (Ammar et al., 2016; Faruqui and Dyer, 2014; Lu et al., 2015) to transfer *surface features* across languages, our major innovation is to convert the text data into structured representations derived from universal dependency parses and enhanced with distributional information to capture individual entities as well

as the relations and events involving those entities, so we can *share structural representations* across multiple languages.

Then we construct a novel cross-lingual structure transfer learning framework to project source language (SL) training data and target language (TL) test data into the common semantic space, so that we can train a relation or event extractor from SL annotations and apply the resulting extractor to TL texts. We adopt graph convolutional networks (GCN) (Kipf and Welling, 2017; Marcheggiani and Titov, 2017) to encode graph structures over the input data, applying graph convolution operations to generate entity and word representations in a latent space. In contrast to other encoders such a Tree-LSTM (Tai et al., 2015), GCN can cover more complete contextual information from dependency parses because, for each word, it captures all parse tree neighbors of the word, rather than just the child nodes of the word. Using this shared encoder, we treat the two tasks of relation extraction and event argument role labeling as mappings from the latent space to relation type and to event type and argument role, respectively.

Extensive experiments on cross-lingual relation and event transfer among English, Chinese, and Arabic show that our approach achieves highly promising performance on both tasks.

2 Model

2.1 Overview

Our cross-lingual structure transfer approach (see Figure 2) consists of four steps: (1) Convert each sentence in any language into a language-universal tree structure based on universal dependency parsing. (2) For each node in the tree structure, create a representation from the concatenation of multilingual word embedding, language-universal POS embedding, dependency role embedding and entity-type embedding, so that all sentences, independent of their language, are represented within one shared semantic space. (3) Adopt GCN to generate contextualized word representations by leveraging information from neighbors derived from the dependency parsing tree, and (4) Using this shared semantic space, train relation and event argument extractors with high-resource language training data, and apply the resulting extractors to texts of low-resource languages that do not have any relation or event argument annotations.

2.2 Tree Representations

Most previous approaches regard a sentence as a linear sequence of words, and incorporate language-specific information such as word order. Unlike sequence representations, tree representations such as constituency trees and dependency trees are typically constructed following a combination of syntactic principles and annotation guidelines designed by linguists. The resulting structures, such as the `verb − subject` relation and the `verb − object` relation, are found across languages. In this paper, we choose dependency trees as the sentence representations because the community has made great efforts at developing language-universal dependency parsing resources across 83 languages (Nivre et al., 2016).

We define the dependency-based tree representation for a sentence as $G = (V, E)$, where $V = \{v_1, v_2, ..., v_N\}$ is a set of words, and $E = \{e_1, e_2, ..., e_M\}$ is a set of language-universal syntactic relations. N is the number of words in the sentence and M is the number of dependency relations between words. To make this tree representation language-universal, we first convert each tree node into a vector which is a concatenation of three language-universal representations at word-level: multilingual word embedding, POS embedding (Nivre et al., 2016), entity-type embedding, and dependency relation embedding. More details are reported in Section 3.2.

2.3 GCN Encoder

Structural information is important for relation extraction and event argument role labeling, thus we aim to generate contextualized word representations by leveraging neighbors in dependency trees for each node.

Our GCN encoder is based on the monolingual design by Zhang et al. (2018b). The graphical sentence representation obtained from dependency parsing of a sentence with N tokens is converted into an $N \times N$ adjacency matrix $\boldsymbol{A}$, with added self-connections at each node to help capture information about the current node itself, as in Kipf and Welling (2017). Here, $A_{i,j} = 1$ denotes the presence of a directed edge from node i to node j in the dependency tree. Initially, each node contains distributional information about the ith word, including word embedding $\boldsymbol{x}_i^w$, embeddings for symbolic information including its POS tag $\boldsymbol{x}_i^p$, dependency relation $\boldsymbol{x}_i^d$ and entity type $\boldsymbol{x}_i^e$.

315

We represent this initial representation $\boldsymbol{h}_i^0$ as follows:

$$\boldsymbol{h}_i^{(0)} = \boldsymbol{x}_i^w \oplus \boldsymbol{x}_i^p \oplus \boldsymbol{x}_i^d \oplus \boldsymbol{x}_i^e$$

Then, at the kth layer of convolution, the hidden representation is derived from the representations of its neighbors at the $(k-1)^{th}$ layer. Thus, the hidden representation at the kth layer for the ith node is computed as:

$$\boldsymbol{h}_i^{(k)} = \text{ReLU}\left(\sum_{j=0}^{N} \frac{\boldsymbol{A_{ij}} \boldsymbol{W}^{(k)} \boldsymbol{h}_j^{(k-1)}}{d_i + b^{(k)}} \right)$$

where $\boldsymbol{W}$ is the weight vector, b is the bias term, and d_i represents the degree of the ith node. The denominator in the equation denotes the normalization factor to neutralize the negative impact of node degree (Zhang et al., 2018b). The final hidden representation of each node after the kth layer is the encoding of each word $\boldsymbol{h}_i^{(k)}$ in our language universal common space, and incorporates information about neighbors up to k hops away in the dependency tree.

2.4 Application on Relation Extraction

The GCN encoder generates the final hidden representation, $h_i^{(k)}$, for each of the n nodes. We perform max-pooling over these final node representations to obtain a single vector representation for the sentence, h^s. Similar to previous work (Zhang et al., 2018b), we also use the following recipe to obtain relation type classification results for each mention pair in a sentence: (1) max-pooling over the final representations of the nodes representing entity mentions, to get a single vector representation for each mention in a pair under consideration, $\boldsymbol{h}^{m_1}$ and $\boldsymbol{h}^{m_2}$, (2) a concatenation of three results of max-pooling: ($[\boldsymbol{h}^{m_1}; \boldsymbol{h}^s; \boldsymbol{h}^{m_2}]$) to combine contextual sentence information with entity mention information (Santoro et al., 2017; Lee et al., 2017; Zhang et al., 2018b), (3) a linear layer to generate a combined representation of these concatenated results, and (4) a Softmax output layer for relation type classification. The objective function used here is as follows:

$$L^r = \sum_{i=1}^{N} \sum_{j=1}^{L_i} y_{ij} \; \log(\sigma(\boldsymbol{U}^r \cdot [\boldsymbol{h}_i^{m1}; \boldsymbol{h}_{ij}^s; \boldsymbol{h}_j^{m2}]))$$

(1)

where $\boldsymbol{U}^r$ is a weight matrix.

2.5 Application on Event Argument Role Labeling

Event argument role labeling distinguishes arguments from non-arguments and classifies arguments by argument role. To label the role of an argument candidate (x_j^a) for an event trigger (x_i^t), we first generate the sentence representation $\boldsymbol{h}^s$, argument candidate representation $\boldsymbol{h}^a$, and trigger representation $\boldsymbol{h}^t$ by applying pooling on sentence, argument candidate x_j^a and event trigger x_i^t respectively, which is the same as that for relation extraction. The mapping function from the latent space to argument roles is composed of a concatenation operation ($[\boldsymbol{h}^t; \boldsymbol{h}^s; \boldsymbol{h}^a]$), a linear layer ($\boldsymbol{U}^a$) and a Softmax output layer:

$$L^a = \sum_{i=1}^{N} \sum_{j=1}^{L_i} y_{ij} \; \log(\sigma(\boldsymbol{U}^a \cdot [\boldsymbol{h}_i^t; \boldsymbol{h}_{ij}^s; \boldsymbol{h}_j^a])) \quad (2)$$

where N is the number of event mentions, L_i is the number of argument candidates for i^{th} event mention and σ is the Sigmoid function.

3 Experiments

3.1 Data and Evaluation Metrics

	Relation Mentions	Event Mentions	Event Arguments
English	8,738	5,349	9,793
Chinese	9,317	3,333	8,032
Arabic	4,731	2,270	4,975

Table 1: Data statistics.

We choose the Automatic Content Extraction (ACE) 2005 corpus (Walker et al., 2006) for our experiments because it contains the most comprehensive gold-standard relation and event annotations for three distinct languages: English, Chinese and Arabic (see Table 1). Our target ontology includes the 7 entity types, 18 relation subtypes and 33 event subtypes defined in ACE. We randomly choose 80% of the corpus for training, 10% for development and 10% for blind test. We downsample the negative training instances by limiting the number of negative samples to be no more than the number of positive samples for each document.

For data preprocessing, we apply the Stanford CoreNLP toolkit (Manning et al., 2014) for Chinese word segmentation and English tokenization, and the API provided by UDPipe (Straka and

Straková, 2017) for Arabic tokenization. We use UDPipe[1] for POS tagging and dependency parsing for all three languages.

We follow the following criteria in previous work (Ji and Grishman, 2008; Li et al., 2013; Li and Ji, 2014) for evaluation:

- A relation mention is considered correct if its relation type is correct, and the head offsets of the two related entity mention arguments are both correct.

- An event argument is correctly labeled if its event type, offsets, and role label match any of the reference argument mentions.

3.2 Training Details

We use the multilingual word embeddings released by (Joulin et al., 2018) which are based on aligning monolingual embeddings learned by fastText (Bojanowski et al., 2017) from Wikipedia. We adopt the universal POS tag set (17 word categories) and 37 categories of dependency relations defined by the *Universal Dependencies* program (Nivre et al., 2016), and seven entity types defined in ACE: person, geopolitical entity, organization, facility, location, weapon and vehicle. Table 2 shows the hyperparameters of our model.

Hyperparameter	Value
word embedding size	300
POS embedding size	30
entity embedding size	30
dependency relation embedding size	30
hidden dimension size	200
dropout	0.5
number of layers	2
pooling function	max pooling
mlp layers	2
learning rate	0.1
learning rate decay	0.9
batch size	50
optimization	SGD

Table 2: GCN hyperparameters.

3.3 Overall Performance

In order to fully analyze the cross-lingual learning capabilities of our framework, we assess its

performance by applying models trained with various combinations of training and test data from these three languages, as shown in Tables 3 and 4. We can see that the results are promising. For both tasks, the models trained from English are best, followed by Chinese, and then Arabic. We find that extraction task performance degrades as the accuracy of language-dependent tools (for sentence segmentation, POS tagging, dependency parsing) degrades.

Using English as training data, our cross-lingual transfer approach achieves similar performance on Chinese event argument role labeling (59.0%), compared to the model trained from Chinese annotations (59.3%), which is much higher than the best reported English-to-Chinese transfer result on event argument role labeling (47.7%) (Hsi et al., 2016).

We also show polyglot results for event argument role labeling in Table 4, by combing the training data from multiple languages. We observe that our model benefits from the combination of training data of multiple languages. The polyglot transfer learning does not provide further gains to relation extraction because the model converges quickly on a small amount of training data.

Test Train	English	Chinese	Arabic
English	68.2	**42.5**	**58.7**
Chinese	**62.6**	69.4	54.0
Arabic	58.6	35.2	67.4

Table 3: Relation Extraction: overall performance (F-score %) using perfect entity mentions.

Test Train	English	Chinese	Arabic
English	63.9	**59.0**	**61.8**
Chinese	**51.6**	59.3	60.6
Arabic	43.1	50.1	64.0
English + Chinese	–	–	**63.1**
English + Arabic	–	**60.1**	–
Chinese + Arabic	**51.9**	–	–

Table 4: Event Argument Role Labeling results (F1 %) with perfect event triggers.

---

[1]Universal Dependencies 2.3 models: english-ewt-ud-2.3-181115.udpipe, chinese-gsd-ud-2.3-181115.udpipe, arabic-padt-ud-2.3-181115.udpipe

Feature	Chinese	Arabic
Multilingual Word Embedding	40.2	53.8
+ POS Embedding	40.1	54.9
+ Entity Embedding	40.8	56.9
+ Dependency Relation Embedding	**42.5**	**58.7**

Table 5: Relation Extraction: Contribution of adding each language-universal feature consecutively for relation extraction on Arabic and Chinese test sets, using the model trained on English (F1 scores (%)).

3.4 Ablation Study

Tables 5 and 6 show the impact of each feature category. We can see that all categories help except Chinese POS features for Relation Extraction and Arabic POS features for Event Argument Role Labeling.

Many Chinese word segmentation errors occur on tokens involved in names. For example, "总统(NOUN, president) 萨姆(PROPN, Samuel) ·(PUNCT) 努乔马(PROPN, Nujoma)" is mistakenly tagged as "总(NOUN) 统萨姆(PROPN) ·(PUNCT) 努乔(NOUN) 马(NOUN)".

In Arabic, sometimes one word is a combination of *Noun* and *Verb*. For example, the single word "الإسرائيلي" means *"Israeli conflict"* in English, including both a trigger and an argument, which are not separated by our tokenizer. In contrast, two entity mentions are unlikely to be combined into one word in Arabic, thus Relation Extraction does not suffer from tokenization errors and corresponding POS features.

Feature	Chinese	Arabic
Multilingual Word Embedding	52.4	56.6
+ POS Embedding	57.1	37.8
+ Entity Embedding	58.3	41.4
+ Dependency Relation Embedding	**59.0**	**61.8**

Table 6: Event Extraction: Contributions of adding each language-universal feature consecutively for event argument role labeling on Arabic and Chinese test datasets, using a model trained on English (F1 scores (%)).

3.5 Comparison with Supervised Approach

We also compare the results with supervised monolingual models trained from manual annotations in the same language. Figures 3 and 4 show the learning curves of these supervised models. For event argument role labeling, we can see that without using any annotations for the target language, our approach achieves performance comparable to the supervised models trained from

more than 3,000 manually annotated event argument mentions, which equal to approximately 1,326 Chinese sentences and 1,141 Arabic sentences based on the statistics of ACE data.

Our model performs particularly well on relation types or argument roles that require deep understanding of wide contexts involving complex syntactic and semantic structures, such as *PART-WHOLE:Artifact*, *PART-WHOLE:Geographical*, *GEN-AFF:Org-Location* and *ORG-AFF:Employment* relations, and *Injure: Victim* argument role. Despite only having 14 training instances, our model achieves near 100% F-score on *PART-WHOLE:Artifact* relations when transferred from English to Chinese. Our model achieves 86% F1 score on *PART-WHOLE:Geographical* relations when transferred from English to Arabic, and 73% and 79% F1 scores on *GEN-AFF:Org-Location* relations when transferred from English to Chinese and to Arabic, respectively. In an *Injure* event, a *Person* can either be an *Agent* or a *Victim*. Surface lexical embedding features are often not sufficient to disambiguate them. Our model is effective at transferring structural information such as dependency relations between words, and obtains 72.97% F1 score on labeling *Injure: Victim* when transferred from English to Chinese, and 75.43 % from English to Arabic.

In addition, our model achieves very high performance on event argument roles for which entity type is a strong indicator. For example, a *weapon* is much more likely to play as an *Instrument* rather than a *Target* in an *Attack*. Our model achieves 89.9% F1 score on *Attack: Instrument* and 91.4% F1 score on *Personnel: POSITION* argument roles when transferred from English to Chinese.

3.6 Using System Extracted Name Mentions

Target Language	F1 score
Chinese	56.9
Arabic	60.1

Table 7: Event Argument Role Labeling results (F1 %) on Chinese and Arabic using English as training data (with system generated entity mentions)

Table 7 shows the results of event argument role labeling on Chinese and Arabic entity mentions automatically extracted by Stanford CoreNLP instead of manually annotated mentions. The system extracted entity mentions introduce noise and
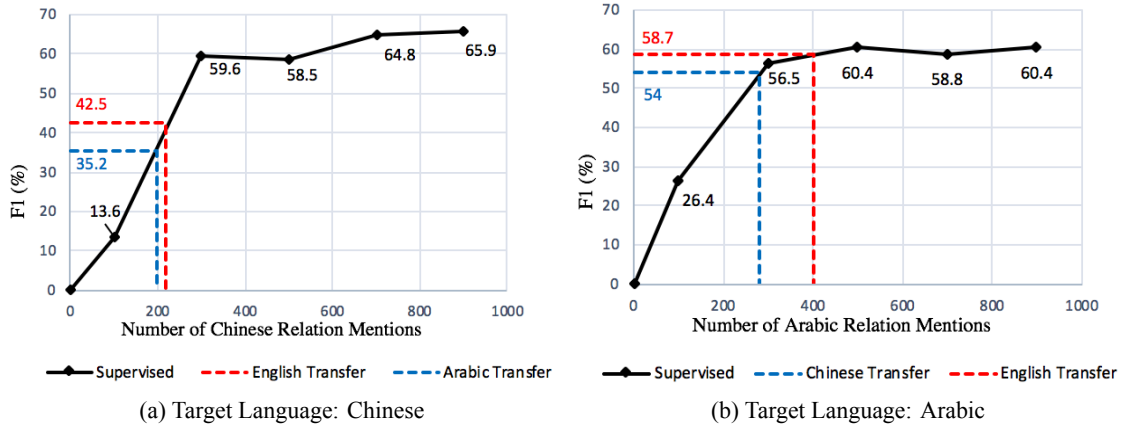
(a) Target Language: Chinese   (b) Target Language: Arabic

Figure 3: Relation Extraction: Comparison between supervised monolingual GCN model and cross-lingual transfer learning.

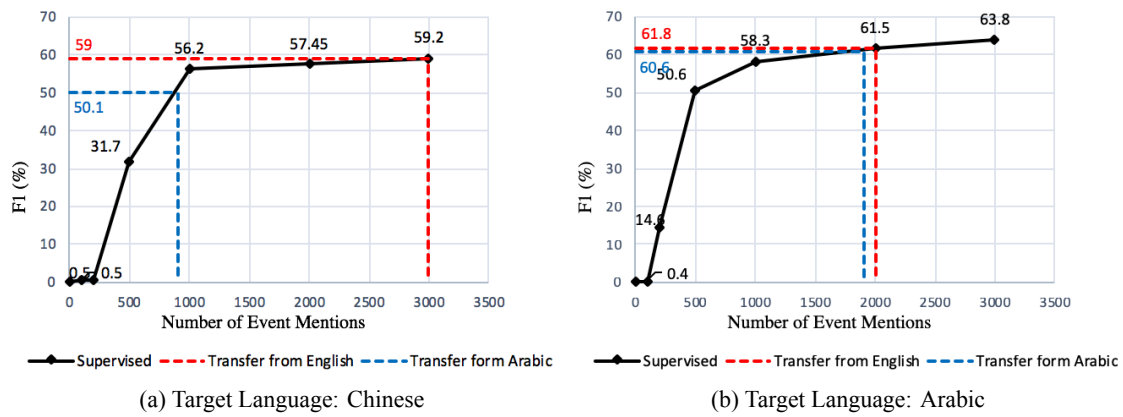(a) Target Language: Chinese   (b) Target Language: Arabic

Figure 4: Event Argument Role Labeling: Comparison between supervised monolingual GCN model and cross-lingual transfer learning.

thus decrease the performance of the model, but the overall results are still promising.

4 Related Work

A large number of supervised machine learning techniques have been used for English event extraction, including traditional techniques based on symbolic features (Ji and Grishman, 2008; Liao and Grishman, 2011), joint inference models (Li et al., 2014; Yang and Mitchell, 2016), and recently with neural networks (Nguyen and Grishman, 2015a; Nguyen et al., 2016; Chen et al., 2015; Nguyen and Grishman, 2018; Liu et al., 2018b; Lu and Nguyen, 2018; Liu et al., 2018a; Zhang et al., 2018a, 2019). English relation extraction in the early days also followed supervised paradigms (Li and Ji, 2014; Zeng et al., 2014; Nguyen and Grishman, 2015b; Miwa and Bansal, 2016; Pawar et al., 2017; Bekoulis et al., 2018; Wang et al., 2018b). Recent efforts have attempted to reduce annotation costs using distant supervision (Mintz et al., 2009; Surdeanu et al., 2012; Min et al., 2013; Angeli et al., 2014; Zeng et al., 2015; Quirk and Poon, 2017; Qin et al., 2018; Wang et al., 2018a) through knowledge bases (KBs), where entities and static relations are plentiful. Distant supervision is less applicable for the task of event extraction because very few dynamic events are included in KBs. These approaches, however, incorporate language-specific characteristics and thus are costly in requiring substantial amount of annotations to adapt to a new language (Chen and Vincent, 2012; Blessing and Schütze, 2012; Li et al., 2012; Danilova et al., 2014; Agerri et al., 2016; Hsi et al., 2016; Feng et al., 2016).

Regardless of the recent successes in applying cross-lingual transfer learning to sequence labeling tasks, such as name tagging (e.g., (Mayhew et al., 2017; Lin et al., 2018; Huang et al., 2019)), only limited work has explored cross-lingual relation and event structure transfer. Most previous efforts working with cross-lingual structure trans-

fer rely on bilingual dictionaries (Hsi et al., 2016), parallel data (Chen and Ji, 2009; Kim et al., 2010; Qian et al., 2014) or machine translation (Faruqui and Kumar, 2015; Zou et al., 2018). Recent methods (Lin et al., 2017; Wang et al., 2018b) aggregate consistent patterns and complementary information across languages to enhance Relation Extraction, but they do so exploiting only distributional representations.

Event extraction shares with Semantic Role Labeling (SRL) the task of assigning to each event argument its event role label, in the process of completing other tasks to extract the full event structure (assigning event types to predicates and more specific roles to arguments). Cross-lingual transfer has been very successful for SRL. Early attempts relied on word alignment (Van der Plas et al., 2011) or bilingual dictionaries (Kozhevnikov and Titov, 2013). Recent work incorporates universal dependencies (Prazák and Konopík, 2017) or multilingual word embeddings for Polyglot SRL (Mulcaire et al., 2018). Liu et al. (2019) and Mulcaire et al. (2019) exploit multi-lingual contextualized word embedding for SRL and other Polyglot NLP tasks including dependency parsing and name tagging. To the best of our knowledge, our work is the first to construct a cross-lingual structure transfer framework that combines language-universal symbolic representations and distributional representations for relation and event extraction over texts written in a language without any training data.

GCN has been successfully applied to several individual monolingual NLP tasks, including relation extraction (Zhang et al., 2018b), event detection (Nguyen and Grishman, 2018), SRL (Marcheggiani and Titov, 2017) and sentence classification (Yao et al., 2019). We apply GCN to construct multi-lingual structural representations for cross-lingual transfer learning.

5 Conclusions and Future Work

We show how cross-lingual relation and event argument structural representations may be transferred between languages without any training data for the target language, and conclude that language-universal symbolic and distributional representations are complementary for cross-lingual structure transfer. In the future we will explore more language-universal representations such as visual features from topically-related images and videos and external background knowl-edge.

References

Rodrigo Agerri, Itziar Aldabe, Egoitz Laparra, German Rigau Claramunt, Antske Fokkens, Paul Huijgen, Rubén Izquierdo Beviá, Marieke van Erp, Piek Vossen, Anne-Lyse Minard, et al. 2016. Multilingual event detection using the newsreader pipelines. In *Proceedings of International Conference on Language Resources and Evaluation 2016*.

Waleed Ammar, George Mulcaire, Yulia Tsvetkov, Guillaume Lample, Chris Dyer, and Noah A Smith. 2016. Massively multilingual word embeddings. *arXiv preprint arXiv:1602.01925*.

Gabor Angeli, Julie Tibshirani, Jean Wu, and Christopher D Manning. 2014. Combining distant and partial supervision for relation extraction. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*.

Giannis Bekoulis, Johannes Deleu, Thomas Demeester, and Chris Develder. 2018. Adversarial training for multi-context joint entity and relation extraction. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2830–2836, Brussels, Belgium. Association for Computational Linguistics.

Andre Blessing and Hinrich Schütze. 2012. Crosslingual distant supervision for extracting relations of different complexity. In *Proceedings of the 21st ACM international conference on Information and knowledge management*.

Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*.

Chen Chen and Ng Vincent. 2012. Joint modeling for Chinese event extraction with rich linguistic features. In *Proceedings of the 24th International Conference on Computational Linguistics*.

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 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*.

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

Vera Danilova, Mikhail Alexandrov, and Xavier Blanco. 2014. A survey of multilingual event extraction from text. In *Proceedings of the International Conference on Applications of Natural Language to Data Bases/Information Systems*.

George R Doddington, Alexis Mitchell, Mark A Przybocki, Lance A Ramshaw, Stephanie M Strassel, and Ralph M Weischedel. 2004. The automatic content extraction (ACE) program-tasks, data, and evaluation. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation*.

Manaal Faruqui and Chris Dyer. 2014. Improving vector space word representations using multilingual correlation. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*.

Manaal Faruqui and Shankar Kumar. 2015. Multilingual open relation extraction using cross-lingual projection. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.

Xiaocheng Feng, Lifu Huang, Duyu Tang, Heng Ji, Bing Qin, and Ting Liu. 2016. 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)*.

Jeremy Getman, Joe Ellis, Stephanie Strassel, Zhiyi Song, and Jennifer Tracey. 2018. Laying the groundwork for knowledge base population: Nine years of linguistic resources for TAC KBP. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018)*.

Andrew Hsi, Yiming Yang, Jaime Carbonell, and Ruochen Xu. 2016. Leveraging multilingual training for limited resource event extraction. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*.

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, Volume 1 (Long and Short Papers)*.

Heng Ji and Ralph Grishman. 2008. Refining event extraction through cross-document inference. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*.

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*.

Seokhwan Kim, Minwoo Jeong, Jonghoon Lee, and Gary Geunbae Lee. 2010. A cross-lingual annotation projection approach for relation detection. In *Proceedings of the 23rd International Conference on Computational Linguistics*.

Thomas N. Kipf and Max Welling. 2017. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*.

Mikhail Kozhevnikov and Ivan Titov. 2013. Cross-lingual transfer of semantic role labeling models. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*.

Angeliki Lazaridou, Georgiana Dinu, and Marco Baroni. 2015. Hubness and pollution: Delving into cross-space mapping for zero-shot learning. 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)*.

Kenton Lee, Luheng He, Mike Lewis, and Luke Zettlemoyer. 2017. End-to-end neural coreference resolution. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*.

Peifeng Li, Guodong Zhou, Qiaoming Zhu, and Libin Hou. 2012. Employing compositional semantics and discourse consistency in chinese event extraction. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*.

Qi Li and Heng Ji. 2014. Incremental joint extraction of entity mentions and relations. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*.

Qi Li, Heng Ji, and Liang Huang. 2013. 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)*.

Qi Li, Heng Ji, HONG Yu, and Sujian Li. 2014. Constructing information networks using one single model. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*.

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 2011*.

Yankai Lin, Zhiyuan Liu, and Maosong Sun. 2017. Neural relation extraction with multi-lingual attention. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*.

Ying Lin, Shengqi Yang, Veselin Stoyanov, and Heng Ji. 2018. A multi-lingual multi-task architecture for low-resource sequence labeling. In *Proc. The 56th Annual Meeting of the Association for Computational Linguistics*.

Jian Liu, Yubo Chen, Kang Liu, and Jun Zhao. 2018a. Event detection via gated multilingual attention mechanism. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*.

Nelson F. Liu, Matt Gardner, Yonatan Belinkov, Matthew E. Peters, and Noah A. Smith. 2019. Linguistic knowledge and transferability of contextual representations. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics*.

Xiao Liu, Zhunchen Luo, and Heyan Huang. 2018b. Jointly multiple events extraction via attention-based graph information aggregation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*.

Ang Lu, Weiran Wang, Mohit Bansal, Kevin Gimpel, and Karen Livescu. 2015. Deep multilingual correlation for improved word embeddings. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.

Weiyi Lu and Thien Huu Nguyen. 2018. Similar but not the same: Word sense disambiguation improves event detection via neural representation matching. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*.

Christopher Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven Bethard, and David McClosky. 2014. The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd annual meeting of the Association for Computational Linguistics: system demonstrations*.

Diego Marcheggiani and Ivan Titov. 2017. Encoding sentences with graph convolutional networks for semantic role labeling. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*.

Stephen Mayhew, Chen-Tse Tsai, and Dan Roth. 2017. Cheap translation for cross-lingual named entity recognition. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*.

Tomas Mikolov, Quoc V Le, and Ilya Sutskever. 2013. Exploiting similarities among languages for machine translation. *arXiv preprint arXiv:1309.4168*.

Bonan Min, Ralph Grishman, Li Wan, Chang Wang, and David Gondek. 2013. Distant supervision for relation extraction with an incomplete knowledge base. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.

Mike Mintz, Steven Bills, Rion Snow, and Dan Jurafsky. 2009. Distant supervision for relation extraction without labeled data. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*.

Makoto Miwa and Mohit Bansal. 2016. 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)*.

Phoebe Mulcaire, Jungo Kasai, and Noah A. Smith. 2019. Polyglot contextual representations improve crosslingual transfer. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*.

Phoebe Mulcaire, Swabha Swayamdipta, and Noah A. Smith. 2018. Polyglot semantic role labeling. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*.

Thien Huu Nguyen, Kyunghyun Cho, and Ralph Grishman. 2016. 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*.

Thien Huu Nguyen and Ralph Grishman. 2015a. Event detection and domain adaptation with 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 2: Short Papers)*.

Thien Huu Nguyen and Ralph Grishman. 2015b. Relation extraction: Perspective from convolutional neural networks. In *Proceedings of North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.

Thien Huu Nguyen and Ralph Grishman. 2018. Graph convolutional networks with argument-aware pooling for event detection. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

Joakim Nivre, Mitchell Abrams, Željko Agić, Lars Ahrenberg, Lene Antonsen, Katya Aplonova, Maria Jesus Aranzabe, Gashaw Arutie, Masayuki Asahara, Luma Ateyah, Mohammed Attia, Aitziber Atutxa, Liesbeth Augustinus, Elena Badmaeva, Miguel Ballesteros, Esha Banerjee, Sebastian Bank, Verginica Barbu Mititelu, Victoria Basmov, John Bauer, Sandra Bellato, Kepa Bengoetxea, Yevgeni Berzak, Irshad Ahmad Bhat, Riyaz Ahmad Bhat, Erica Biagetti, Eckhard Bick, Rogier Blokland, Victoria Bobicev, Carl Börstell, Cristina Bosco, Gosse Bouma, Sam Bowman, Adriane Boyd, Aljoscha Burchardt, Marie Candito, Bernard Caron, Gauthier Caron, Gülşen Cebiroğlu Eryiğit, Flavio Massimiliano Cecchini, Giuseppe G. A. Celano, Slavomír Čéplö, Savas Cetin, Fabricio Chalub, Jinho Choi, Yongseok Cho, Jayeol Chun, Silvie Cinková, Aurélie Collomb, Çağrı Çöltekin, Miriam Connor, Marine Courtin, Elizabeth Davidson, Marie-Catherine de Marneffe, Valeria de Paiva, Arantza Diaz de Ilarraza, Carly Dickerson, Peter Dirix, Kaja Dobrovoljc, Timothy Dozat, Kira Droganova, Puneet Dwivedi, Marhaba Eli, Ali Elkahky, Binyam Ephrem, Tomaž Erjavec, Aline Etienne, Richárd Farkas, Hector Fernandez Alcalde, Jennifer Foster, Cláudia Freitas, Katarína Gajdošová, Daniel Galbraith, Marcos Garcia, Moa Gärdenfors, Sebastian Garza, Kim Gerdes, Filip Ginter, Iakes Goenaga, Koldo Gojenola, Memduh Gökırmak, Yoav Goldberg, Xavier Gómez Guinovart, Berta Gonzáles Saavedra, Matias Grioni, Normunds Grūzītis, Bruno Guillaume, Céline Guillot-Barbance, Nizar Habash, Jan Hajič, Jan Hajič jr., Linh Hà Mỹ, Na-Rae Han, Kim Harris, Dag Haug, Barbora Hladká, Jaroslava Hlaváčová, Florinel Hociung, Petter Hohle, Jena Hwang, Radu Ion, Elena Irimia, Ọlájídé Ishola, Tomáš Jelínek, Anders Johannsen, Fredrik Jørgensen, Hüner Kaşıkara, Sylvain Kahane, Hiroshi Kanayama, Jenna Kanerva, Boris Katz, Tolga Kayadelen, Jessica Kenney, Václava Kettnerová, Jesse Kirchner, Kamil Kopacewicz, Natalia Kotsyba, Simon Krek, Sookyoung Kwak, Veronika Laippala, Lorenzo Lambertino, Lucia Lam, Tatiana Lando, Septina Dian Larasati, Alexei Lavrentiev, John Lee, Phương Lê Hồng, Alessandro Lenci, Saran Lertpradit, Herman Leung, Cheuk Ying Li, Josie Li, Keying Li, KyungTae Lim, Nikola Ljubešić, Olga Loginova, 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, Yuji Matsumoto, Ryan McDonald, Gustavo Mendonça, Niko Miekka, Margarita Misirpashayeva, Anna Missilä, Cătălin Mititelu, Yusuke Miyao, Simonetta Montemagni, Amir More, Laura Moreno Romero, Keiko Sophie Mori, Shinsuke Mori, Bjartur Mortensen, Bohdan Moskalevskyi, Kadri Muischnek, Yugo Murawaki, Kaili Müürisep, Pinkey Nainwani, Juan Ignacio Navarro Horñiacek, Anna Nedoluzhko, Gunta Nešpore-Bērzkalne, Lương Nguyễn Thị, Huyền Nguyễn Thị Minh, Vitaly Nikolaev, Rattima Nitisaroj, Hanna Nurmi, Stina Ojala, Adédayọ̀ Olúòkun, Mai Omura, Petya Osenova, Robert Östling, Lilja Øvrelid, Niko Partanen, Elena Pascual, Marco Passarotti, Agnieszka Patejuk, Guilherme Paulino-Passos, Siyao Peng, Cenel-Augusto Perez, Guy Perrier, Slav Petrov, Jussi Piitulainen, Emily Pitler, Barbara Plank, Thierry Poibeau, Martin Popel, Lauma Pretkalniņa, Sophie Prévost, Prokopis Prokopidis, Adam Przepiórkowski, Tiina Puolakainen, Sampo Pyysalo, Andriela Rääbis, Alexandre Rademaker, Loganathan Ramasamy, Taraka Rama, Carlos Ramisch, Vinit Ravishankar, Livy Real, Siva Reddy, Georg Rehm, Michael Rießler, Larissa Rinaldi, Laura Rituma, Luisa Rocha, Mykhailo Romanenko, Rudolf Rosa, Davide Rovati, Valentin Ro☐ca, Olga Rudina, Jack Rueter, Shoval Sadde, Benoît Sagot, Shadi Saleh, Tanja Samardžić, Stephanie Samson, Manuela Sanguinetti, Baiba Saulīte, Yanin Sawanakunanon, Nathan Schneider, Sebastian Schuster, Djamé Seddah, Wolfgang Seeker, Mojgan Seraji, Mo Shen, Atsuko Shimada, Muh Shohibussirri, Dmitry Sichinava, Natalia Silveira, Maria Simi, Radu Simionescu, Katalin Simkó, Mária Šimková, Kiril Simov, Aaron Smith, Isabela Soares-Bastos, Carolyn Spadine, Antonio Stella, Milan Straka, Jana Strnadová, Alane Suhr, Umut Sulubacak, Zsolt Szántó, Dima Taji, Yuta Takahashi, Takaaki Tanaka, Isabelle Tellier, Trond Trosterud, Anna Trukhina, Reut Tsarfaty, Francis Tyers, Sumire Uematsu, Zdeňka Urešová, Larraitz Uria, Hans Uszkoreit, Sowmya Vajjala, Daniel van Niekerk, Gertjan van Noord, Viktor Varga, Eric Villemonte de la Clergerie, Veronika Vincze, Lars Wallin, Jing Xian Wang, Jonathan North Washington, Seyi Williams, Mats Wirén, Tsegay Woldemariam, Tak-sum Wong, Chunxiao Yan, Marat M. Yavrumyan, Zhuoran Yu, Zdeněk Žabokrtský, Amir Zeldes, Daniel Zeman, Manying Zhang, and Hanzhi Zhu. 2018. Universal dependencies 2.3. LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.

Joakim Nivre, Marie-Catherine De Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajic, Christopher D Manning, Ryan T McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, et al. 2016. Universal dependencies v1: A multilingual treebank collection. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation*.

Xiaoman Pan, Boliang Zhang, Jonathan May, Joel Nothman, Kevin Knight, and Heng Ji. 2017. Cross-lingual name tagging and linking for 282 languages. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*.

323

Sachin Pawar, Pushpak Bhattacharyya, and Girish K. Palshikar. 2017. End-to-end relation extraction using neural networks and Markov logic networks. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*.

Lonneke Van der Plas, Paola Merlo, and James Henderson. 2011. Scaling up automatic cross-lingual semantic role annotation. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*.

Ondrej Prazák and Miloslav Konopík. 2017. Cross-lingual SRL based upon universal dependencies. In *Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017*.

Longhua Qian, Haotian Hui, Ya'nan Hu, Guodong Zhou, and Qiaoming Zhu. 2014. Bilingual active learning for relation classification via pseudo parallel corpora. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*.

Pengda Qin, Weiran Xu, and William Yang Wang. 2018. Dsgan: Generative adversarial training for distant supervision relation extraction. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*.

Chris Quirk and Hoifung Poon. 2017. Distant supervision for relation extraction beyond the sentence boundary. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*.

Sascha Rothe, Sebastian Ebert, and Hinrich Schütze. 2016. Ultradense word embeddings by orthogonal transformation. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.

Adam Santoro, David Raposo, David G Barrett, Mateusz Malinowski, Razvan Pascanu, Peter Battaglia, and Timothy Lillicrap. 2017. A simple neural network module for relational reasoning. In *Proceedings of Advances in Neural Information Processing Systems*.

Samuel L Smith, David HP Turban, Steven Hamblin, and Nils Y Hammerla. 2017. Offline bilingual word vectors, orthogonal transformations and the inverted softmax. *arXiv preprint arXiv:1702.03859*.

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*.

Mihai Surdeanu, Julie Tibshirani, Ramesh Nallapati, and Christopher D. Manning. 2012. Multi-instance multi-label learning for relation extraction. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*.

Kai Sheng Tai, Richard Socher, and Christopher D. Manning. 2015. Improved semantic representations from tree-structured long short-term memory 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)*.

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

Guanying Wang, Wen Zhang, Ruoxu Wang, Yalin Zhou, Xi Chen, Wei Zhang, Hai Zhu, and Huajun Chen. 2018a. Label-free distant supervision for relation extraction via knowledge graph embedding. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*.

Xiaozhi Wang, Xu Han, Yankai Lin, Zhiyuan Liu, and Maosong Sun. 2018b. Adversarial multi-lingual neural relation extraction. In *Proceedings of the 27th International Conference on Computational Linguistics*.

Chao Xing, Dong Wang, Chao Liu, and Yiye Lin. 2015. Normalized word embedding and orthogonal transform for bilingual word translation. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.

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*.

Liang Yao, Chengsheng Mao, and Yuan Luo. 2019. Graph convolutional networks for text classification. In *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence*.

Daojian Zeng, Kang Liu, Yubo Chen, and Jun Zhao. 2015. Distant supervision for relation extraction via piecewise convolutional neural networks. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*.

Daojian Zeng, Kang Liu, Siwei Lai, Guangyou Zhou, and Jun Zhao. 2014. Relation classification via convolutional deep neural network. In *Proceedings of the 25th International Conference on Computational Linguistics*.

Jingli Zhang, Wenxuan Zhou, Yu Hong, Jianmin Yao, and Min Zhang. 2018a. Using entity relation to improve event detection via attention mechanism. In *Proceedings of NLPCC 2018*.

Tongtao Zhang, Heng Ji, and Avirup Sil. 2019. Joint entity and event extraction with generative adversarial imitation learning. *Data Intelligence Vol 1 (2)*.

Yuan Zhang, David Gaddy, Regina Barzilay, and Tommi S Jaakkola. 2016. Ten pairs to tag-multilingual pos tagging via coarse mapping between embeddings. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.

Yuhao Zhang, Peng Qi, and Christopher D. Manning. 2018b. Graph convolution over pruned dependency trees improves relation extraction. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*.

Bowei Zou, Zengzhuang Xu, Yu Hong, and Guodong Zhou. 2018. Adversarial feature adaptation for cross-lingual relation classification. In *Proceedings of the 27th International Conference on Computational Linguistics*.