Expert Systems With Applications 255 (2024) 124760

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

MT4CrossOIE: Multi-stage tuning for cross-lingual open information extraction

Tongliang Li [a,1], Zixiang Wang [b,1], Linzheng Chai [b], Jian Yang [b,*], Jiaqi Bai [b], Yuwei Yin [c], Jiaheng Liu [b], Hongcheng Guo [b], Liqun Yang [b], Hebboul Zine el-abidine [b], Zhoujun Li [b]

[a] Computer School, Beijing Information Science and Technology University, Beijing, China
[b] State Key Lab of Software Development Environment, Beihang University, Beijing, China
[c] The University of Hong Kong, Hong Kong

ARTICLE INFO

Keywords:
Information extraction
Cross-lingual transfer
Disentangled training
Multi-lingual training
Mixture of LoRA

ABSTRACT

Cross-lingual open information extraction aims to extract structured information from raw text across multiple languages. Previous work uses a shared cross-lingual pre-trained model to handle the different languages but underuses the potential of the language-specific representation. In this paper, we propose an effective multi-stage tuning framework called MT4CrossOIE, designed for enhancing cross-lingual open information extraction by injecting language-specific knowledge into the shared model. Specifically, the cross-lingual pre-trained model is first tuned in a shared semantic space (e.g., embedding matrix) in the fixed encoder and then other components are optimized in the second stage. After enough training, we freeze the pre-trained model and tune the multiple extra low-rank language-specific modules using mixture of LoRAs for model-based cross-lingual transfer. In addition, we leverage two-stage prompting to encourage the large language model (LLM) to annotate the multi-lingual raw data for data-based cross-lingual transfer. The model is trained with multi-lingual objectives on our proposed dataset `OpenIE4++` by combining the model-based and data-based transfer techniques. Experimental results on various benchmarks emphasize the importance of aggregating multiple plug-in-and-play language-specific modules and demonstrate the effectiveness of MT4CrossOIE in cross-lingual OIE.[2]

1. Introduction

Open information extraction (OIE) aims to extract key structured data from an arbitrary domain text in the form of predicates (usually verbals or verbal phrases) and their corresponding arguments (Niklaus, Cetto, Freitas, & Handschuh, 2018), without pre-defined relation schemas. Considering the sentence *"Joe Biden became the US president in the year 2021"*, three tuples are expected to extract by OIE systems: *<Joe Biden; became; the US president>*, *<Joe Biden; became the US president; in the year 2021>* and *<Joe Biden; became; the US president; in the year 2021>*. Due to the domain independence and scalability, OIE provides powerful help for downstream tasks like question answering (Bhutani, Suhara, Tan, Halevy, & Jagadish, 2019; Mausam, 2016), summarization (Rahat & Talebpour, 2018), and knowledge graph completion (Choi, Lee, & Lee, 2023).

Most of the existing methods are highly dependent on the labeled data and do not perform well in low-resource languages. Multi2OIE (Ro, Lee, & Kang, 2020) is the first neural-based method to tackle OIE task in multiple languages and achieves a satisfactory performance on cross-lingual transfer. Large language models (Brown et al., 2020; OpenAI, 2023; Touvron et al., 2023) have exhibited extraordinary abilities and have been widely applied to various tasks, including OIE (Jeronymo et al., 2023; Wan et al., 2023), and other tasks (Lin et al., 2023; Yin, Yang, Yang, & Liu, 2023). Despite the success of the existing advances in OIE, the following limitations have not been fully investigated yet: (1) Fine-tuning the entire language model may result in its previously learned knowledge being forgotten due to the catastrophic forgetting (Yang, Ding et al., 2022). (2) Another limitation is the lack of robustness in handling low-resource languages and cannot tackle different languages simultaneously (Boggia et al., 2023).

* Corresponding author.
E-mail addresses: tonyliangli@bistu.edu.cn (T. Li), wangzixiang@buaa.edu.cn (Z. Wang), challenging@buaa.edu.cn (L. Chai), jiaya@buaa.edu.cn (J. Yang), bjq@buaa.edu.cn (J. Bai), yuweiyin@hku.hk (Y. Yin), liujiaheng@buaa.edu.cn (J. Liu), hongchengguo@buaa.edu.cn (H. Guo), lqyang@buaa.edu.cn (L. Yang), z.hebboul@buaa.edu.cn (H. Zine el-abidine), lizj@buaa.edu.cn (Z. Li).

[1] These authors contributed equally.
[2] https://github.com/CSJianYang/Multilingual-Multimodal-NLP

https://doi.org/10.1016/j.eswa.2024.124760
Received 22 September 2023; Received in revised form 18 May 2024; Accepted 8 July 2024
Available online 14 July 2024

Generally, most previous works adopt a shared cross-lingual pre-trained model to handle the different languages but underscore the potential of the language-specific representation.

In this paper, we propose a multi-stage tuning framework for CrossOIE to encourage knowledge sharing among different languages. Inspired by the previous work (He, Liu, Gao, & Chen, 2021; Liu, Niehues, Cross, Guzmán, & Li, 2021), the word embedding of the cross-lingual pre-trained model is tuned to align the representations of different languages by freezing the encoder in the first stage while other components are adjusted in the second stage. Given the strong OIE model, we add the mixture-of-LoRAs (mLoRA) into the fixed model for different languages, where all languages depend on the same backbone model, and adjust the output by combining different low-rank adapters. The multilingual encoder including mBERT and XLM-R pre-trained on massively multilingual corpora are used to initialize the model without mLoRA, where mBERT and XLM-R bring fundamental multilingual capability for our proposed model. Besides, we leverage the large language model (LLM) as the cross-lingual annotator to label the multi-lingual raw corpora originating from the English data. Finally, we combine model-based and data-based transfer in our framework to improve the performance of the cross-lingual OIE.

Our contributions are summarized as follows:

- We propose a multi-stage tuning framework called MT4CrossOIE for CrossOIE that combines model-based and data-based methods to transfer knowledge into the pre-trained model, which has the superiority in extracting the tuples across different languages.
- We build a new multi-lingual corpus called `OpenIE4++`, which consists of the original English data and their counterparts of other languages via the large language model with the cross-lingual prompt.
- We conduct extensive experiments on multiple languages (English, Arabic, Chinese, German, Spanish, and Portuguese) and the results demonstrate that MT4CrossOIE outperforms baseline models on most languages of different benchmarks. Finally, we perform an extensive analysis and reveal the nature of OIE in different languages.

2. Cross-lingual OIE

Given the source information extraction model Θ_{IE}^{src} only trained on the source information extraction dataset and the target raw sentence $x = (x_1, \dots, x_m)$ with m words, the zero-shot cross-lingual information extraction aims to identify potential arguments and focuses on extracting predicates among different arguments mentioned in the raw text. Then, we can obtain a list of tuples $T = \{T_1, \dots, T_N\}$, where $T_i = (a_i^1, p_i, a_i^2, \dots, a_i^q)$ is the ith tuple, p_i denotes the predicate in T_i and a_i^j is the jth argument of p_i. The a_i^1 is considered as the subject and $a_i^2, \dots, a_i^q$ are objects associated with T_i. The problem definition of zero-shot cross-lingual open information extraction (OIE) is described as:

$$P(T|X) = \prod_{i=1}^{N} P(T_i|x; \Theta_{IE}^{src}) \tag{1}$$

where the tuples T are derived from the target raw sentence x. T_i is the ith tuple. The source language has annotated labels but the target corpora have no accessible handcrafted labels. $P(T|x)$ represents the predicted distributions of labels. The source information extraction model Θ_{IE}^{src} trained on the source annotated corpus is expected to be evaluated on the target language without any labeled dataset. For example, given the Germany sentence "Hofmann wurde in Salt Lake City, Utah, geboren.", CrossOIE task can infer tuples <*Hofmann; wurde geboren; in Salt Lake City*> and <*Hofmann; wurde; geboren in Utah*> on English training data and without any relation schema pre-defined. Multi2OIE (Ro et al., 2020) aims to extract facts in a sentence without relying on pre-defined schemas. It aims to be more flexible and capable of handling a wider range of textual input. In this work, we propose to unify the model-based transfer from the cross-lingual pre-trained model and data-based transfer with machine translation to transfer knowledge from the source language to the target language.

3. Methodology

In this section, we propose the multi-stage fine-tuning method for cross-lingual OIE as shown in Fig. 1, where we align the semantic representation in the first stage and adjust other model components in the second stage by disentangled tuning. Next, we introduce the mixture-of-LoRAs to compose the language-specific representations for prediction. Furthermore, we trigger the cross-lingual generalization ability of a large language model using the cross-lingual prompt to construct the multi-lingual corpora `OpenIE4++` to further augment the cross-lingual transfer. Generally, we aims at distilling the knowledge from the large language model.

3.1. Backbone model

Given the input sentence $x = \{x_1, \dots, x_n\}$ of language L_k, our backbone model first predicts the predicate tagset $t^p = \{t_1^p, \dots, t_n^p\}$ with a predicate head, and then outputs the argument tagset $t^a = \{t_1^a, \dots, t_1^a\}$ with an argument head. Following the previous work (Ro et al., 2020), we use BIO (Beginning-Inside-Outside) sequence-labeling scheme to tag the predicates and arguments in a sentence. The backbone is a two-step n-ary extraction which is first extracting all predicates and then the arguments associated:

$$P(T^p, T^a|x) = P(T^p|x)P(T^a|x, T^p) \tag{2}$$

where T^p and T^a have the same length and OIE is regarded as an n-ary extraction task. In our work, we both leverage the mBERT and XLM-R to initialize the model to encourage multilingual agreement among different languages.

Multi-head attention. Given the input embedding X, we project X into Q as the query, K as the key, and V as the value in the self-attention module to extract the representations:

$$X_{attn} = \Big\Vert_{h=1}^{H} \mathrm{SF}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{3}$$

where $\mathrm{SF}(\cdot)$ denotes the softmax function, and $\Vert_{h=1}^{H}$ is the feature concatenation of the H attention heads. The input X is projected into $Q = W^q X, K = W^k X, V = W^v X$ with the learned matrix W^q, W^k, W^v. After the self-attention module, other standard operations (e.g. feed-forward network) are used. Finally, we obtain the representations $H = \{h_1, \dots, h_n\}$ and $H \in \mathbb{R}^{n \times d_k}$.

Predicate and argument extraction. The sentence representations H are then fed into a predicate prediction head that consists of a feed-forward network and a softmax layer to classify each token into a predefined predicate tag. We obtain the predicted tags $t^p = \{t^1, \dots, t^n\}$ and the cross-entropy loss $\mathcal{L}_p$ is optimized for predicate extraction.

After predicting the predicate tags, we sum the average representation of the predicted predicate h^p and each word representation H, which are then fed into an argument extractor comprised of N_2 multi-head attention blocks as in Eq. (3). Finally, the output of the multi-head attention block is fed into the argument classifier.

3.2. Disentangled tuning

Let $\Theta = \{\theta_p, \theta_w, \theta_b, \theta_c\}$ denote all model parameters, where θ_p is the position embedding, θ_w is the word embedding, θ_b is the pre-trained model, and θ_c is the predicate and argument classifier. For a token at position i in a sequence, we represent it using two vectors, H_i and P_i, which represent its word and position embedding, respectively. The

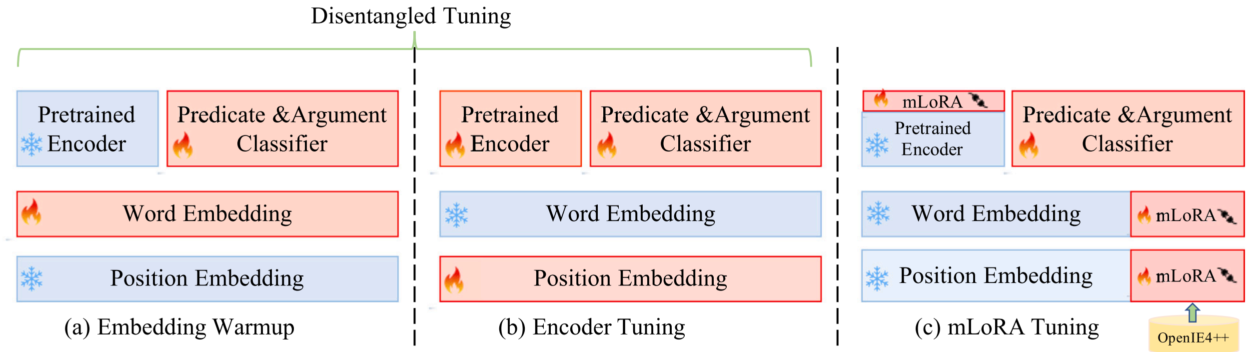


Fig. 1. The training sketch of mT4CrossOIE, where the blue ice icon indicates parameter-frozen modules while the red fire icon denotes trainable ones. In the first stage (a) and second stage (b), we align the semantic representation by disentangled tuning. In the third stage (c), we introduce the mixture of LoRAs to compose the language-specific representations for prediction. Additionally, we construct the multi-lingual corpora `OpenIE4++` to trigger the cross-lingual generalization ability.

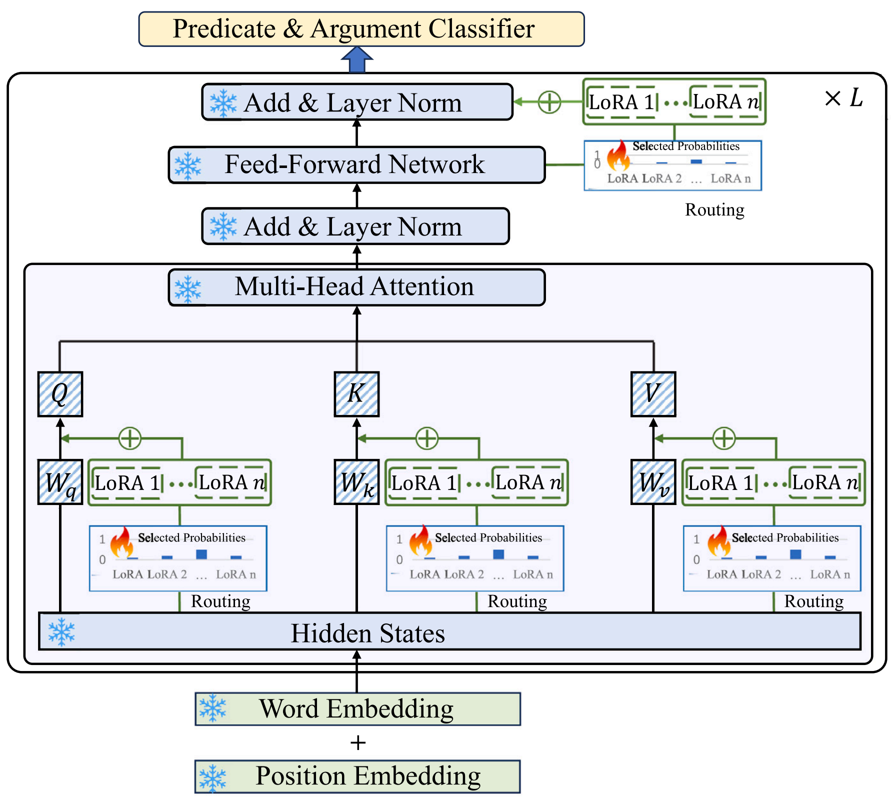


Fig. 2. The overview of mT4CrossOIE. We calculate the selection probabilities of all LoRA adapters and choose the top-k LoRA experts obeying the probability values. Selection probabilities are determined by the hidden state of each layer.

calculation of the cross-attention score $A_{i,j} = \frac{(H_i+P_i)(H_j+P_j)^T}{\sqrt{d}}$ between tokens i and j can nearly be decomposed into four parts as (omitting scaling factor $\sqrt{d}$):

$$A_{i,j} = \overbrace{H_i H_j^T + \underbrace{H_i P_j^T + H_j P_i^T}_{}}^{\text{(1) Content-based Terms}} \underbrace{+ P_i P_j^T}_{\text{(2) Position-based Terms}} \tag{4}$$

where $A_{i,j}$ denotes the attention score between position i and j. The content-based term (1) $H_i H_j^T + H_i P_j^T + H_j P_i^T$ optimizes the word embedding while the position-based term (2) $H_i P_j^T + H_j P_i^T + P_i P_j^T$ relates to the position embedding. Here, we emphasize the importance of disentangled tuning, which can effectively improve the cross-lingual capability by separately aligning the word embeddings and position embeddings in different languages.

Zero-shot inference depends on the cross-lingual generalizability of the pre-trained model to conditions unseen in training. In the context of zero-shot OIE, the input should ideally be encoded into a language-agnostic representation. Inspired by the previous work (He et al., 2021; Liu et al., 2021), we propose a disentangled tuning strategy to relax the constraint between the word and position information. As shown in Fig. 1, we only tune the content-based term in the first stage by only tuning the word embedding parameters (θ_w) and classifier (θ_c):

$$\mathcal{L}^{(1)} = -\mathbb{E}[\log P(T|x; \Theta_1 = \{\theta_w, \theta_c\})] \tag{5}$$

where x is the input sentence and T are the extracted tuples.

Then, other components continued to be tuned $\Theta_2 = \{\theta_p, \theta_b, \theta_c\}$ to optimize the position-based term (2) $H_i P_j^T + H_j P_i^T + P_i P_j^T$ by freezing the word embedding matrix:

$$\mathcal{L}^{(2)} = -\mathbb{E}[\log P(T|x; \Theta_2 = \{\theta_p, \theta_b, \theta_c\})] \tag{6}$$

where x is the input sentence and T are the extracted tuples.

3.3. Mixture-of-LoRAs for CrossOIE

LoRA (Hu et al., 2022) is a tuning technique in LLMs, which enables efficient and flexible transfer by introducing task-specific modifications to a fixed pre-trained model. We can observe from Fig. 2 that the mixture-of-LoRAs (mLoRA) is used on different languages for the cross-lingual transfer, where a group of LoRA adapters is lightweight compared to the pre-trained model. The adapters with a low-rank down-project matrix and up-project matrix can be directly inserted into the pre-trained embedding, attention, and feed-forward network. Given the source sentence $x = \{x_1, \dots, x_n\}$ of n tokens and a group of T_m LoRA experts, we use mLoRA to learn the language-sensitive representations for the same task of different languages:

$$h_a^{L_i} = \mathcal{A}_{\theta_{g(L_i)}}(h^{L_i}) \tag{7}$$

where $g(L_i)$ are selected LoRA experts derived from the language representations. $\mathcal{A}(\cdot)$ denotes the LoRA adapter module and $\theta = \{\theta_1, \dots, \theta_T\}$ denotes the adapter pool. $\mathcal{A}_{\theta_{g(L_i)}}$ is calculated by:

$$\mathcal{A}_{\theta_{g(L_i)}}(h^{L_i}) = W h^{L_i} + \sum_{A_t, B_t \in S(e)} \lambda \Delta W h^{L_i} \tag{8}$$

where W is the fixed weight. $\Delta W = BA$ is denoted by a low-rank decomposition ($A \in \mathbb{R}^{d \times r} \wedge B \in \mathbb{R}^{r \times d} \wedge r \ll d$). The matrices A and B are initialized by a random Gaussian distribution and zero. λ is the scaling factor and r is the inner dimension. $S(e)$ denotes the subset from the language-specific modules.

In Eq. (8), all experts only require fine-tuning a small number of language-specific parameters instead of all parameters of the pre-trained model. Thus, we can simultaneously train multiple experts for different languages, which all share the same freezing pre-trained parameters. We use multiple adapters from the selected subset to maximize the transfer of knowledge across languages:

$$g(L_i) = \mathtt{TopK}\left(\frac{exp(\alpha_j^{L_i})}{\sum_{t=1}^{T} exp(\alpha_t^{L_i})}\right) \tag{9}$$

where $\mathtt{TopK}(\cdot)$ is the selection function, where we calculate the selection probabilities of all LoRA adapters and choose the top-k LoRA experts obeying the probability distribution. $\alpha_j^{L_i}$ is a scalar from the representations of language L_i (We use the hidden state of the special token [CLS] of each layer). $S(e) = \{(A_k, B_k)\}_{k=1}^{K}$ and $\alpha_j^{L_i}$ is used to incorporate the different experts.

We project the language representation e^{L_i} of language L_i into the LoRA expert distribution using the learned matrix $W_a \in \mathbb{R}^{d \times T}$, where d is the hidden size and T_m is the number of experts. The weight of LoRA expert $\alpha_j^{L_i}$ is calculated by:

$$\alpha^{L_i} = e^{L_i} W_a \tag{10}$$

where $\alpha = \{\alpha_1, \dots, \alpha_T\}$. For all modules of the pre-trained model, we leverage the mixture-of-LoRAs to learn the language-sensitive representations for the different input sentences by activating top-k experts.

3.4. Multi-lingual training

LLM as cross-lingual annotator. Large language models (LLMs) equipped with a growing arsenal of prompt-based methods offer the powerful off-the-shelf few-shot capability to the cross-lingual NLP task. To facilitate the generalizability of the cross-lingual model, we design the cross-lingual prompt $P = \{p_1, p_2\}$ to trigger the potential of LLM. The cross-lingual annotation problem can be decomposed into translation procedure and OIE annotation. We use the prompt p_1 for translation and the prompt p_2 for OIE annotation as:

$$P(y, T|X) = P(y|x, p_1)P(T|y, p_2) \tag{11}$$

where y is the target translation of X and T is the corresponding extracted tuples. The source sentence is translated into the target sentence with the first prompt p_1 and then extracted into multiple tuples with the second prompt. Table 1 shows the detailed chain-of-thought (Wei et al., 2022) prompt for cross-lingual annotation using the large language model.

Multi-lingual training objective. Given the supervised corpus D_{L_i}, we expand the source corpus to the multi-lingual corpora $D = \{D_{L_1}, \dots, D_{L_K}\}$ of K language using LLM. The training objective of the CrossOIE can be described as:

$$\mathcal{L}_m = -\frac{1}{K}\sum_{i=1}^{K} \mathbb{E}_{x,T \in D_{L_i}} \log(T|x) \tag{12}$$

where x and T are input sentences and extracted tuples from the multi-lingual corpora.

Multi-lingual training schedule. Due to an adequate training in first and second stage with the high-resource corpus (English), a simple concatenate with all language data training will lead to poor performance because of the imbalanced data of low-resource languages. Following the previous study (Aharoni, Johnson, & Firat, 2019; Wang, Zhai, & Hassan, 2020), we employ a temperature-based sampling mechanism to randomly sample OpenIE4++ data in the third stage training. Concretely, we calculate the ratio of each language data:

$$p_i = \frac{|L_i|}{\sum_{j=1}^{L} |L_j|} \tag{13}$$

where $|L_i|$ is the number of ith language. Then we adopt the batch-balance method by randomly sampling the sentence in different languages according to a multinomial distribution with probabilities $\{q_1, q_2, \dots, q_N\}$:

$$q_i = \frac{p_i^{\frac{1}{\tau}}}{\sum_{j=1}^{L} p_i^{\frac{1}{\tau}}} \tag{14}$$

where N is the number of languages, and τ is the temperature.

4. Experiments

4.1. Experimental setup

Datasets.

- Our training data is the same as that used in Ro et al. (2020) and Zhan and Zhao (2020) for the first and second stages. This English training dataset was bootstrapped from extractions of the OpenIE4 system (Mausam, 2016). It contains n-ary extractions, enabling model evaluation on both binary and n-ary extraction benchmarks. We also randomly select 42k annotated sentences from the original training data and triggered the large language model (gpt-3.5-turbo) to obtain the labeled dataset based on our prompts. The dataset contains 5 languages: Arabic, Chinese, German, Portuguese, and Spanish. The new annotated dataset coupled with original English data aggregates the OpenIE4++. Table 2 lists the statistics. The table shows the average token number of each language after tokenization. Especially, for the language Arabic, we first generate the 10K samples, but find that the fine-tuned model with 10K Arabic samples leads to the worse performance of other languages (But two experiments still have similar average performance of all languages: Δperformance $<$ 0.1). Therefore, we select 2K samples for Arabic training data to reduce the total training data size for efficiency (We will also add the 10K samples into the released dataset considering the expensive labeling cost.).

Table 1
Prompts and their usage for the large language model (`gpt-3.5-turbo`). `[X]` is the source sentence of English and `[Y]` is the translated sentence of,target language `[L]`. `[S1]`, `[R1]`, and `[O1]` denote the subject, relation, and object of the target translated sentence: `[Y1]`, where we provide the example consisting of the target sentence `[Y1]` and its extraction results (`[S1]`,`[R1]`,`[O1]`) for the few-shot extraction.

Task	Prompt
p_1: Translate `[X]` to `[Y]`	You are a translator. Please translate the following English text into the `[L]`: `[X]`
p_2: Annotate `[Y]`	You are an Information Extraction expert. The following are the extraction results of `[Y1]`, which are represented by Subject, Relation, and Object: `[S1]`, `[R1]`, `[O1]` Please refer to the extraction results above, extracting a triple that corresponds Subject, Relation, and Object from the translated sentence: `[Y]`. Note that the subject, relation, and object must originate from the continuous segment of the sentence. The output format must be the same as the sample above.

Table 2
Statistics of sentence number, tuple number, maximum sentence length, minimum sentence length, and average sentence length in `OpenIE4++`.

Statistics	Ar	De	Es	Pt	Zh
#Sent.	10,000	10,000	10,000	10,000	10,000
#Tuples	14,960	15,339	16,842	17,033	15,871
Max_len	61	63	68	75	110
Min_len	3	4	4	4	5
Avg_len	19.3	21.9	24.8	24.0	35.4

- Re-OIE2016 (Zhan & Zhao, 2020) is a more accurate English n-ary extraction benchmark that is manually re-annotated the entire OIE2016. Spanish and Portuguese versions of Re-OIE2016 are extended by Ro et al. (2020), with the same number of sentences and tuples for each language.
- CaRB (Bhardwaj, Aggarwal, & Mausam, 2019) is an English n-ary extraction benchmark which is a crowd-sourced re-annotated dataset based on dev and test splits of OIE2016. It has higher coverage and quality of the reference extractions compared to most of the OIE benchmarks.
- BenchIE (Gashteovski et al., 2021) is a multi-lingual OIE benchmark for binary extraction evaluation in English, Chinese, and German. Unlike other datasets, BenchIE is an exhaustive fact-based benchmark that includes fact synsets, Each synset is a set of all acceptable surface forms of the same fact. In other words, the gold standard into account the informational equivalence of extractions, which makes evaluation more comprehensive.

Evaluation metrics. We evaluated each OIE system using the F1-score. The F1-score is a balanced assessment of the model, combining precision and recall into a single measure. We used the evaluation code provided with each benchmark. Allowing the extractions to be slightly different from the gold tuples, as there are no restrictions on the elements of open extractions.

For the CaRB evaluation, we utilize their *tuple match* which is a stricter token-level matching scorer for a rigorous evaluation. It matches predicted predicates with golden predicates and predicted arguments with golden arguments respectively. Since the *lexical match* evaluation has numerous shortcomings (Bhardwaj et al., 2019), we also use *tuple match* matching criterion on the multi-lingual Re-OIE2016. BenchIE provides a fact-level matching scorer which takes the informational equivalence of extractions into account by exactly matching extracted triple with the corresponding gold fact synset (i.e., the same fact with different surface forms).

Implementation details. We train the model for 1 epoch in each stage. The batch size is set to 128 in the first and second stages, and 64 in the third stage. In the third stage training, the temperature is set to 100 to ensure a relative balanced ratio for sampling different languages in a batch. The maximum sentence length is set to 100. The number of experts in each mLoRA is set to 6, which is equivalent to the number of languages. The LoRA rank is set to 64. We use AdamW (Loshchilov & Hutter, 2019) as our optimizer with an initial learning rate of 3e−5. For the cross-lingual encoder, we use the multi-lingual BERT (mBERT) (Devlin, Chang, Lee, & Toutanova, 2019). The model is trained on a single NVIDIA Tesla V100 (32 GB). We choose the top-4 LoRA experts based on the best average F1 score at the inference stage. To further verify the effectiveness of our model designs, we experiment on a variant initialized with XLM-R. For training settings, the learning rate is set to 6e−5 for the first training stage, the rest of the settings remain the same as the mBERT version.

4.2. Baselines

We compare our model with both English and multi-lingual baselines. For the evaluation of the English datasets, we use non-neural systems: Stanford (Angeli, Premkumar, & Manning, 2015), ClausIE (Corro & Gemulla, 2013), MinIE (Gashteovski, Gemulla, & Corro, 2017) and neural models: RnnOIE (Stanovsky, Michael, Zettlemoyer, & Dagan, 2018), SpanOIE (Zhan & Zhao, 2020), IMoJIE (Kolluru, Aggarwal, Rathore, Mausam and Chakrabarti, 2020), CIGL (Kolluru, Adlakha, Aggarwal, Mausam and Chakrabarti, 2020), OpenIE6 (Kolluru, Adlakha et al., 2020), Multi2OIE (Ro et al., 2020), and GPT-3.5.[3]

For the evaluation of multiple languages, Multi2OIE (Ro et al., 2020) is used as neural-network-based baselines. Rule-based systems like ClausIE and MinIE cannot be used for languages other than English. We use ArgOE (Gamallo & García, 2015) and PredPatt (White et al., 2016) as rule-based baselines, which are only two multi-lingual systems. GPT-3.5, an iteration of the Generative Pre-trained Transformer models developed by OpenAI, represents a significant leap in the field of artificial intelligence, particularly in natural language processing (NLP). To verify the effectiveness of our method, we compare our method with GPT-3.5 (`gpt-3.5-turbo`) as the large language model and the Multi2OIE initialized with XLM-R (Conneau et al., 2020). We feed the prompt "Please extract a triple that corresponds Subject, Relation, and Object from the given sentence: {sentence}. Note that the subject, relation, and object must originate from the continuous segment of the sentence". to GPT-3.5 for zero-shot extractions.

4.3. Main results

Our model initialized by a cross-lingual pre-trained model first tunes the word embedding and then after enough training, we freeze the pre-trained model and tune the multiple extra low-rank language-specific modules using mixture-of-LoRAs for model-based cross-lingual transfer. In addition, we leverage two-stage prompting to encourage the large language model (LLM) to annotate the multi-lingual raw data for data-based cross-lingual transfer. The model is trained with multi-lingual objectives on our proposed dataset OpenIE4++ by combining the model-based and data-based transfer techniques.

4.3.1. English

We compare our model with several unsupervised and supervised baselines on CaRB and BenchIE English benchmarks in Table 3. Compared to rule-based and neural-based models, MT4CrossOIE achieves a relatively high F1 score on CaRB n-ary extractions. Constrained-IGL (CIGL) is an individual component in OpenIE6, which achieves the highest performance among all prior models but can only use

[3] https://platform.openai.com/docs/models.

Table 3
MT4CrossOIE performance comparison with baseline models on CaRB n-ary and BenchIE binary English extraction benchmarks. DT refers to disentangled tuning.

Models	CaRB			BenchIE		
	F1	PREC.	REC.	F1	PREC.	REC.
ClausIE	44.9	–	–	33.9	50.3	25.6
MinIE	41.9	–	–	33.7	42.9	27.8
Stanford	23.9	–	–	13.0	11.1	15.7
RnnOIE	46.7	55.6	40.2	13.0	37.3	7.8
SpanOIE	49.4	60.9	41.6	–	–	–
IMoJIE	53.5	–	–	–	–	–
CIGL	**54.0**	–	–	–	–	–
OpenIE6	52.7	–	–	25.4	31.1	**21.4**
GPT-3.5	38.2	53.1	38.2	20.2	34.7	14.2
Multi2OIE	51.9	59.5	**45.9**	23.8	37.7	17.4
Multi2OIE (XLM-R)	51.5	63.0	43.6	26.0	42.2	18.7
MT4CrossOIE	51.8	**65.8**	42.7	29.1	50.0	20.5
- DT	50.9	63.0	42.7	23.2	37.5	16.7
- mLoRA	51.6	65.6	42.5	28.6	48.8	20.0
- OpenIE4++	51.3	64.9	42.4	**29.3**	**50.5**	20.7
MT4CrossOIE (XLM-R)	51.7	64.7	43.1	26.7	45.6	18.9

English-specific constraints in training. The performance gap between MT4CrossOIE and the multi-lingual baseline Multi2OIE[4] is minimal on CaRB n-ary extraction. Since CaRB's evaluation scheme penalizes long extraction in the precision calculation, however, it may cause high recall just simply adding words in extraction. We notice that even though MT4CrossOIE cannot reach the highest F1 score on CaRB, it yields the highest precision score, which is more convincing in this evaluation scheme. MT4CrossOIE performs best compared to other neural-based baselines on BenchIE binary extraction. Even though rule-based systems like ClausIE and MinIE outperform all neural systems, they cannot be used for non-English languages. Similar to the result on CaRB, the high performance on BenchIE is attributed to the high precision.

Generally speaking, compared to the large language model, our fine-tuned cross-lingual OIE model with lightweight parameters, enables faster deployment and operation which can be essential for real-time applications or environments with limited computational resources. As shown in Table 3, we can observe that our method still beats the large language model. The generated result from GPT3.5 cannot be successfully parsed into correct formats leading to worse performance. We also observe that the XLM-R variant of MT4CrossOIE outperforms the Muliti2OIE baseline initialized with XLM-R, which indicates our model design is not limited to certain cross-lingual pre-trained encoders.

4.3.2. Multi-lingual

In Table 4, we compare our model with multi-lingual baselines on Re-OIE2016 multi-lingual version benchmark that is proposed by Ro et al. (2020). MT4CrossOIE outperforms the other baselines in all languages and yields the highest F1 values, which has demonstrated the excellent cross-lingual abilities of our framework. Specifically, MT4CrossOIE outperforms Multi2OIE by 0.2%, 1.6%, and 0.8% in English, Portuguese, and Spanish, respectively. CoNLL04. We perform a significant test with the best baseline suggesting that the performance is statistically significant ($p < 0.05$). Meanwhile, our method is stable with all F1 standard deviations (no more than 0.5). We report the average result and the standard deviation (scores in brackets) when re-training with different seeds. The average results and standard deviations for English, Spanish, and Portuguese are 69.7 (± 0.3), 60.7 (± 0.3), 60.8 (± 0.5), respectively. The superiority of our framework is attributed to its high precision, which is more reliable since CaRB evaluation rewards long extractions with much higher recall scores

[4] The results reported in Ro et al. (2020) are based on the English version and we reproduce it by loading the officially released multi-lingual version checkpoint for a fair comparison in this study.

Table 4
The models are tested using CaRB's evaluation scheme *tuple match* for rigorous evaluation on the multi-lingual Re-OIE2016. The results are cited from Ro et al. (2020), which only reported binary extraction performance due to the baseline systems being binary extractors.

Language	System	F1	PREC.	REC.
En	ArgOE	43.4	56.6	35.2
	PredPatt	53.1	53.9	52.3
	GPT-3.5	66.2	64.2	68.3
	Multi2OIE	69.3	66.9	**71.7**
	Multi2OIE (XLM-R)	**69.6**	70.2	69.0
	MT4CrossOIE	69.5	**73.4**	66.0
	- mLoRA	69.3	72.7	66.2
	- OpenIE4++	69.1	72.4	66.1
	MT4CrossOIE (XLM-R)	69.0	71.9	66.2
Pt	ArgOE	38.3	46.3	32.7
	PredPatt	42.9	43.6	42.3
	GPT-3.5	56.2	52.6	60.4
	Multi2OIE	59.1	56.1	**62.5**
	Multi2OIE (XLM-R)	60.5	62.8	58.5
	MT4CrossOIE	**60.7**	**63.5**	58.2
	- mLoRA	57.6	60.4	55.0
	- OpenIE4++	55.0	61.0	50.1
	MT4CrossOIE (XLM-R)	60.5	63.8	57.6
Es	ArgOE	39.4	48.0	33.4
	PredPatt	44.3	44.8	43.8
	GPT-3.5	42.9	44.3	41.2
	Multi2OIE	60.2	59.1	**61.2**
	Multi2OIE (XLM-R)	60.0	63.7	56.7
	MT4CrossOIE	**61.0**	**65.0**	57.5
	- mLoRA	57.9	61.4	54.70
	- OpenIE4++	53.8	61.0	48.1
	MT4CrossOIE (XLM-R)	60.0	64.5	56.1

as we discussed in Section 4.3.1. As shown in Table 4, we conduct ablation study by removing OpenIE4++ and mLoRA separately, which both leads to severe performance degradation. Therefore, we combine the model transfer mLoRA and data transfer OpenIE4++ for the best performance.

In Table 5, we compare MT4CrossOIE with the multi-lingual neural-based model on the BenchIE non-English datasets. Similar to the method proposed in 3.4, we triggered gpt-3.5-turbo to annotate 100 sentences from BenchIE-English data to Arabic. Then we amended all the incorrect triples manually with the help of a native Arabic speaker. From Table 5, MT4CrossOIE outperforms Multi2OIE in all languages. Chinese and Arabic have a significant improvement over the baseline model. We can observe that both multi-lingual models perform significantly worse in German. There are many German verb stems and their separable prefixes appear in sentences. The decoding method of both models used a BIO tagging scheme that identifies continuous phrases, which are always absent in predicates, resulting in an extremely low recall.

The XLM-R-based models only obtained comparable results with mBERT-based models, indicating that while the performance of mBERT is usually low compared to these models on multi-lingual translation tasks, the insight seems not applicable to the cross-lingual Open Information Extraction task. We find that the performance of Multi2OIE (XLM-R) on Chinese is extremely low. By carefully examining the extraction results, we discover that a large number of arguments from the relation triples cannot be correctly tagged, leading to extraction failures. We also observe that our method significantly outperforms the GPT-3.5 large language model which adopts few-shot prompting. This indicates that training "small" models with our method is superior than employing prompt engineering on "large" models for the cross-lingual Open Information Extraction task.

Table 5
The performance of multi-lingual neural-based OIE models on BenchIE-multi-lingual binary extraction. The results of Multi2OIE are reproduced in multi-lingual settings. DT refers to disentangled tuning.

	Zh			De			Ar		
	F1	PREC.	REC.	F1	PREC.	REC.	F1	PREC.	REC.
GPT-3.5	14.9	22.1	11.3	3.5	6.5	2.4	8.3	20.4	5.2
Multi2OIE	9.0	11.0	7.6	3.3	5.7	2.3	4.4	12.5	2.7
Multi2OIE (XLM-R)	0.6	4.2	0.3	3.2	5.8	2.2	6.0	16.4	3.7
MT4CrossOIE	**16.0**	**23.1**	**12.3**	4.4	8.5	2.9	**11.0**	**23.8**	**7.2**
- DT	12.0	18.2	9.0	3.6	7.1	2.4	8.8	20.8	5.5
- mLoRA	14.9	20.9	11.6	4.1	8.0	2.8	7.0	18.3	4.3
- OpenIE4++	14.9	21.1	11.6	**5.0**	**11.6**	**3.2**	4.4	19.7	2.5
MT4CrossOIE (XLM-R)	11.7	21.5	8.0	4.1	7.6	2.9	9.4	22.8	5.6

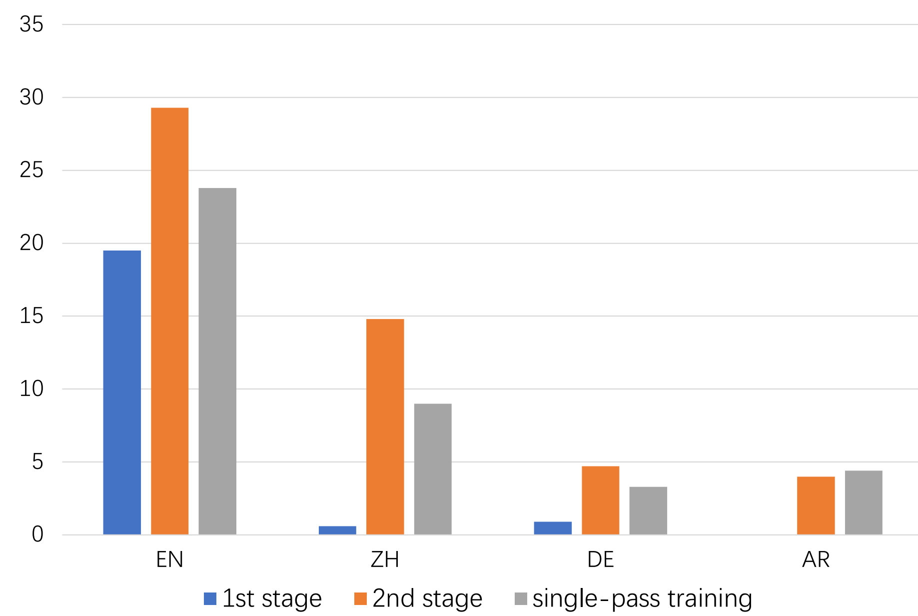


Fig. 3. The comparison among different disentangled tuning stages and one pass training strategy.

4.4. Ablation study

We assess the benefits of each design in our proposed framework, including disentangled training, mLoRA, and the `OpenIE4++` dataset.

Effectiveness of disentangled tuning. We use a high-quality multi-lingual dataset BenchIE to explore the effect of the disentangled tuning strategy. From Fig. 3, we observe that our model reaches higher performance on multiple languages after disentangled tuning compared to Multi2OIE which is tuned on a single pass. It is apparent that our framework truly keeps knowledge from being forgotten from the big picture. Since we tune the different parts in English training data, the language and task features are learned adequately in English even without prior knowledge in the first stage, while a big gap in the first and second stages of the other three languages. The results in Chinese and German demonstrate the model's satisfactory zero-shot performance even though non-English data is not available in the training stages. However, the results in Arabic seem slightly lower, even with no performance in the first stage. We suppose that English, Chinese, and German are subject-verb-object languages, while Arabic is a verb-subject-object language, and their subjects or objects can be expressed as part of the verb, resulting in low performance. Such interference may hurt model performance in a certain language during our disentangled tuning. Moreover, we examine the performance when directly training with mLoRA without the first and second stages. Results can be found from the "MT4CrossOIE -DT" ablation setting in Table 3. The F1 dropped 0.9% and 5.9% respectively for both English datasets in 3, as well as 4.0%, 0.8%, and 2.2% for Chinese, Germany and Arabic in 5, demonstrating the effectiveness of our disentangled tuning design on multi-lingual datasets.

Effectiveness of mLoRA. To investigate how mLoRA influence the overall performance, we conduct an ablation study on the third stage. From Tables 3 and 5, we observe that all components are helpful for the proposed method. In particular, there is an evident performance drop across all languages when removing the mLoRA from our proposed method, which indicates the effectiveness of the mLoRA for improving the capability of the model-based cross-lingual transfer. It also demonstrates that different experts can provide diverse knowledge to enrich the limited language-specific representation.

Effectiveness of `OpenIE4++`. Moreover, without the help of `OpenIE4++`, training mLoRA only with the raw English corpus (OpenIE4) will cause a distinct performance drop in other languages, which demonstrates the effectiveness of the data-based cross-lingual transfer. This indicates that training with a mixture of multiple languages contributes to improving the low-resource language representation, especially in non-Indo-European language families (e.g., Chinese and Arabic). Interestingly, we observe that German has a small performance enhancement without a German training corpus. We assume that the limited German training corpus did not provide much help due to their sophisticated language feature applied to the BIO scheme and the interference of the other languages can hurt the model performance on German. Besides, English and German have many similarities as they are in the same language family, which also benefits the cross-lingual transfer. We notice that performance on BenchIE English also has a minimal improvement compared to MT4CrossOIE, while a slight drop on

Table 6
Cases of baseline and our method. We only select four annotations that are most similar to the output as golden references. [·] denotes an optional item. The two arguments and the corresponding predicate are separated by –>. Error parts are highlighted in red-colored font, while correct parts are in blue-colored.

Sent. #1	Sligo town then became an incorporated municipal borough with a Royal Charter issued by the British King James I in 1613/14.
Gold	[a] Royal Charter –> issued by –> [the] British King [a] Royal Charter –> issued –> by [the] British King [a] Royal Charter –> issued by –> [the] [British] [King] James I [a] Royal Charter –> issued –> by [the] [British] [King] James I
Multi2OIE	A Royal Charter –> issued –> by the British King James I in 1613/14
MT4CrossOIE	A Royal Charter –> issued –> by the British King James
Sent. #2	It hosts the "Zomercarnaval ", the second largest Caribbean carnival in Europe, originally called the Antillean carnival.
Gold	It –> hosts –> [the] ["] Zomercarnaval ["] It –> hosts –> [the] "Zomercarnaval" It –> hosts –> [the] [second largest] Caribbean carnival [in Europe] It –> hosts –> [the] Antillean carnival
Multi2OIE	It –> hosts –> the " Zomercarnaval " the second largest Caribbean carnival in Europe
MT4CrossOIE	It –> hosts –> the " Zomercarnaval
Sent. #3	The Anti-Monitor began to siphon the positive matter of New York City to create his Antimatter waves.
Gold	[The] Anti-Monitor –> [began to] siphon –> [the] positive matter [of New York City] [The] Anti-Monitor –> began –> to siphon [the] positive matter [of New York City] [The] Anti-Monitor –> began to –> siphon [the] positive matter [of New York City] [The] Anti-Monitor –> began to siphon –> [the] positive matter [of New York City]
Multi2OIE	The Anti-Monitor –> began to siphon the positive matter of New –> York City
MT4CrossOIE	The Anti-Monitor –> began to siphon –> the positive matter of New York City
Sent. #4	Salomon Brothers says, " We believe the real estate properties would trade at a discount ... after the realty unit is spun off ....
Gold	Salomon Brothers –> says –> [,] We believe [the] real estate properties would trade at [a] discount [...] after [the] realty unit is spun off [...] Salomon Brothers –> says –> [,] We believe [the] real estate properties would trade at [a] discount Salomon Brothers –> says –> [,] ["] We believe [the] real estate properties would trade at [a] discount [...] after [the] realty unit is spun off [...] Salomon Brothers –> says –> [,] ["] We believe [the] real estate properties would trade at [a] discount
Multi2OIE	Salomon Brothers –> says –> We believe the real estate properties would trade at a discount ... after the realty unit is spun
MT4CrossOIE	Salomon Brothers –> says –> We believe the real estate properties would trade at a discount ... after the realty unit is spun off

CaRB. We suppose the performance decline is mainly caused by model overfitting on CaRB. Our full model has achieved a great balance.

4.5. Case study

To provide an in-depth analysis of the cases, we examine the extraction outputs of our proposed MT4CrossOIE and the Multi2OIE baseline from several random samples on the BenchIE English benchmark, as shown in Table 6. The following samples exemplify four advantages of MT4CrossOIE. We summarize the four major superiorities of our framework in these samples:

Concise extraction. As shown in Sent. #1, our framework concisely extracts the target triplet, while the baseline model appends unnecessary date information at the end of the sentence.

Coreference resolution. In Sent. #2, our method is capable of identifying the coreference. Instead, the baseline model extracts an appositive clause, making the extraction redundant and confusing.

Named entity recognition. In Sent. #3, the named entity "New York City" is correctly recognized by our method. However, the baseline model fails to recognize it as a whole.

Grammatical correctness. In Sent. #4, the preposition "off" is missing in the baseline model extraction, causing a grammatical error, while our framework adds it accurately.

4.6. Discussion

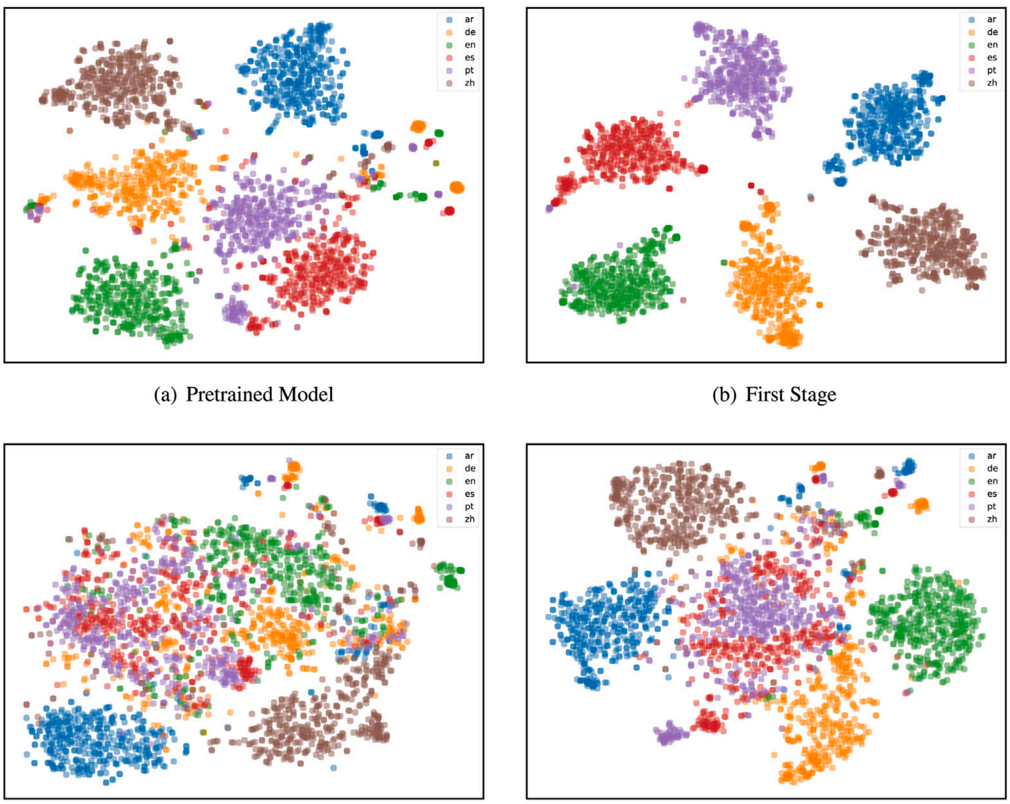
Quality of `OpenIE4++`. To construct the multilingual corpora `OpenIE4++`, we adopt the large language model to translate the English labeled to other languages. To verify the quality of `OpenIE4++`, we randomly select 500 samples from the training corpora and the correctness of each sample. The evaluate criteria is set as {"very bad", "bad", "neutral", "good", "very good"}. We recruited 2 volunteers to score these samples, where more than 98% samples are classified into "neutral", "good", and "very good". Therefore, our created multilingual corpora `OpenIE4++` are high-quality enough to effectively improve the performance of cross-lingual open information extraction.

The effect of LoRA rank. Given a limited memory budget, what is the optimal combination of rank r for our mLoRA module in the top-k strategy? The results of the effect of r and top-k on model performance are presented in Table 7. Note that mLoRA performs competitively with a relatively big r (e.g., rank=32 and rank=64). We argue that increasing r to a certain degree does cover a more meaningful subspace. We notice that rank=2 also achieves a considerable performance. This is more desirable if considering the model capacity. We also find that top-k is correlated with the rank of the LoRA module. The combination of $r = 64$ and $k = 4$ gains the best performance.

Cross-lingual representation of multi-stage training. We visualize the average sentence representations in Fig. 4 using t-SNE (Maaten & Hinton, 2008). (a) We load pre-trained parameters of mBERT without any fine-tuning to observe the sentence representations directly across different languages. The language representations are scattered in space after prior pre-training in 104 languages. (b) From the observation of our first stage, the language representations have become even more scattered. There is an evident distinction among languages. (c) After tuning in the second stage, all language representations are mixed together. The languages are aligned in a shared space, where the similar semantic representations of languages are close in the same position area. The model obtains the benefits of shared parameters after the disentangled tuning. (d) We observe that the language representations are slightly scattered again in the third stage. That indicates the languages perceive mLoRA in the third stage, which means language features are well-distinguished after obtaining independent parameters while retaining the benefits of shared parameters.

Table 7
Performance of MT4CrossOIE with different LoRA ranks on benchmarks. We choose the best top-k value for each rank. The bold font **Total** denotes the sum of F1 scores across all datasets.

	CaRB	BenchIE				Re-OIE2016			**Total**
	En	En	Zh	De	Ar	En	Pt	Es	
rank = 1 (k = 1)	51.58	29.22	**17.12**	3.85	9.80	69.41	61.21	60.61	302.79
rank = 2 (k = 3)	51.43	**29.31**	16.62	3.70	10.46	69.41	61.02	61.14	303.09
rank = 4 (k = 3)	51.52	28.99	16.23	3.56	9.78	69.31	**61.27**	60.93	301.59
rank = 8 (k = 6)	51.53	29.23	16.38	3.71	9.45	69.43	61.03	61.10	301.86
rank = 16 (k = 6)	51.58	28.81	16.66	3.97	9.80	69.46	61.06	**61.26**	302.60
rank = 32 (k = 2)	51.70	29.08	15.69	4.10	**11.37**	69.23	61.05	61.26	303.49
rank = 64 (k = 2)	**51.80**	29.11	15.88	**4.38**	11.04	69.44	60.76	60.95	303.36
rank = 64 (k = 4, ours)	51.78	29.10	16.02	**4.38**	11.04	**69.48**	60.71	61.03	**303.54**
rank = 64 (k = 6)	51.77	28.99	15.99	**4.38**	11.04	**69.48**	60.45	60.87	302.97
rank = 128 (k = 2)	**51.80**	28.81	16.37	**4.38**	10.49	69.40	60.49	60.85	302.60



(a) Pretrained Model (b) First Stage

(c) Second Stage (d) Third Stage

Fig. 4. t-SNE (Maaten & Hinton, 2008) visualization of the average sentence representations in the `OpenIE4++` dataset for multi-stage training strategy. (a) are initial representations of cross-lingual pre-trained models. (b) are features after the first stage. (c) are features after the second stage. (d) are features after the third stage.

5. Related work

Open information extraction. Open Information Extraction (OIE) is a task that extracts a set of n-ary relation tuples from an arbitrary domain text (Niklaus et al., 2018). OIE systems have two main categories: (I) Unsupervised rule-based approaches, which perform extractions with dependency parsers and PoS taggers based on fine-grained rules or handcrafted features (Corro & Gemulla, 2013; Fader, Soderland, & Etzioni, 2011; Gashteovski et al., 2017; Guo et al., 2023; Lauscher, Song, & Gashteovski, 2019; Mausam, Schmitz, Soderland, Bart, & Etzioni, 2012). Most recent OIE approaches are usually based on neural networks (Liu et al., 2020) which are built as different supervised learning models. Neural solutions become popular and achieved considerable improvement due to the large-scale OIE benchmarks (Bhardwaj et al., 2019; Stanovsky & Dagan, 2016; Zhan & Zhao, 2020). (II) Supervised neural OIE models, which handle OIE tasks by utilizing sequence labeling models to tag each token as a role label in a sentence (Jia, Shijia, Ding, Chen, & Xiang, 2022; Ro et al., 2020; Roy, Park, Lee, & Pan, 2019; Sarhan & Spruit, 2019; Stanovsky & Dagan, 2016; Stanovsky et al., 2018), using span-based models to directly predict whether a span-level phrase is a predicate or an argument instead of a BIO tag in a token-level (Zhan & Zhao, 2020), or performing an encode-decode schema to produce extraction tuples as a sequence step by step using sequence generation models (Cui, Wei, & Zhou, 2018; Kolluru, Aggarwal et al., 2020). OIE models can be utilized to build knowledge graphs (Song, Zhang et al., 2023) or extract key information for other downstream tasks (Li et al., 2023; Song et al., 2023). In this paper, we view OIE as a sequence labeling problem and build up MT4CrossOIE.

Cross-lingual NLP tasks. Cross-lingual tasks include various NLP tasks involved in multiple languages (Chi et al., 2020; Devlin et al., 2019; Guo et al., 2022; Liu, Yu, Peng, Sun, & Li, 2022), such as cross-lingual pre-training (Conneau et al., 2020; Conneau & Lample, 2019; Yang, Ding et al., 2022; Yang, Ma, Huang, Zhang, Dong, Huang, Muzio, Singhal, Hassan, Song, & Wei, 2021; Yang et al., 2020), cross-lingual named entity recognition (Yang, Huang et al., 2022; Zhou et al., 2022), and cross-lingual summarization (Bhattacharjee et al., 2023), and multilingual translation (Lu, Huang, Zhang, Wei, & Lam, 2024; Tan et al., 2019; Yang, Guo, Yin, Bai, Wang, Liu, Liang, Chai, Yang, & Li, 2024; Yang, Yin, Ma, Zhang, Li et al., 2022; Yang, Yin, Ma, Zhang, Wu et al., 2022; Yang, Yin, Yang, Ma, Huang, Zhang, Wei, & Li, 2023). Cross-lingual transfer (the process of leveraging knowledge and resources from one language to another) plays a pivotal role in cross-lingual tasks. This approach not only saves resources but also helps overcome the data scarcity problem in low-resource languages. Most of the previous studies (Ro et al., 2020) cannot be easily extended to the cross-lingual scenario of the OIE task, and thus our method is proposed to leverage the multi-stage training gradually distill the source language knowledge to other languages.

6. Conclusion

In this paper, we propose MT4CrossOIE, a multistage tuning framework for cross-lingual open information extraction, which injects language-specific knowledge into the shared model. Moreover, we devise a novel data augmentation strategy, which leverages the chain-of-thought prompt to encourage the large language model annotating the multi-lingual raw data for data-based cross-lingual transfer. Experimental results demonstrate that our approach outperforms the previous state-of-the-art approaches by a significant margin. Further analysis demonstrates our model effectively obtains language-agnostic representations in the shared parameters and language-specific knowledge in the mixture-of-LoRAs to reduce the gap among different languages.

CRediT authorship contribution statement

Tongliang Li: Formal analysis, Supervision, Writing – review & editing. **Zixiang Wang:** Conceptualization, Methodology, Formal analysis, Software, Writing – original draft. **Linzheng Chai:** Methodology, Writing – review & editing. **Jian Yang:** Software, Writing – review & editing. **Jiaqi Bai:** Writing – review & editing. **Yuwei Yin:** Writing – review & editing. **Jiaheng Liu:** Writing – review & editing. **Hongcheng Guo:** Writing – review & editing. **Liqun Yang:** Writing – review & editing. **Hebboul Zine el-abidine:** Writing – review & editing. **Zhoujun Li:** Supervision, Writing – review & editing, Validation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The dataset has been made public.

References

Aharoni, R., Johnson, M., & Firat, O. (2019). Massively multilingual neural machine translation. In J. Burstein, C. Doran, & T. Solorio (Eds.), *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)* (pp. 3874–3884). Minneapolis, Minnesota: Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/N19-1388, URL: https://aclanthology.org/N19-1388.

Angeli, G., Premkumar, M. J. J., & Manning, C. D. (2015). Leveraging linguistic structure for open domain information extraction. In *Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing of the Asian federation of natural language processing, ACL 2015, July 26-31, 2015, Beijing, China, volume 1: long papers* (pp. 344–354). The Association for Computer Linguistics, http://dx.doi.org/10.3115/v1/p15-1034.

Bhardwaj, S., Aggarwal, S., & Mausam (2019). CaRB: A crowdsourced benchmark for open IE. In K. Inui, J. Jiang, V. Ng, & X. Wan (Eds.), *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019* (pp. 6261–6266). Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/D19-1651.

Bhattacharjee, A., Hasan, T., Ahmad, W. U., Li, Y., Kang, Y., & Shahriyar, R. (2023). CrossSum: Beyond english-centric cross-lingual summarization for 1, 500+ language pairs. In A. Rogers, J. L. Boyd-Graber, & N. Okazaki (Eds.), *Proceedings of the 61st annual meeting of the association for computational linguistics (volume 1: long papers), ACL 2023, Toronto, Canada, July 9-14, 2023* (pp. 2541–2564). Association for Computational Linguistics, URL: https://aclanthology.org/2023.acl-long.143.

Bhutani, N., Suhara, Y., Tan, W., Halevy, A. Y., & Jagadish, H. V. (2019). Open information extraction from question-answer pairs. In J. Burstein, C. Doran, & T. Solorio (Eds.), *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, volume 1 (long and short papers)* (pp. 2294–2305). Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/n19-1239.

Boggia, M., Grönroos, S., Loppi, N. A., Mickus, T., Raganato, A., Tiedemann, J., et al. (2023). Dozens of translation directions or millions of shared parameters? comparing two types of multilinguality in modular machine translation. In T. Alumäe, M. Fishel (Eds.), *Proceedings of the 24th nordic conference on computational linguistics, NoDaLiDa 2023, Tórshavn, Faroe Islands, May 22-24, 2023* (pp. 238–247). University of Tartu Library, URL: https://aclanthology.org/2023.nodalida-1.24.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., et al. (2020). Language models are few-shot learners. In H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, H. Lin (Eds.), *Advances in neural information processing systems 33: annual conference on neural information processing systems 2020, NeurIPS 2020, December 6-12, 2020, Virtual*. URL: https://proceedings.neurips.cc/paper/2020/hash/1457c0d6bfcb4967418bfb8ac142f64a-Abstract.html.

Chi, Z., Dong, L., Wei, F., Wang, W., Mao, X., & Huang, H. (2020). Cross-lingual natural language generation via pre-training. In *The thirty-fourth AAAI conference on artificial intelligence, AAAI 2020, the thirty-second innovative applications of artificial intelligence conference, IAAI 2020, the tenth AAAI symposium on educational advances in artificial intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020* (pp. 7570–7577). AAAI Press, URL: https://ojs.aaai.org/index.php/AAAI/article/view/6256.

Choi, H., Lee, H., & Lee, M. (2023). Optimal knowledge component extracting model for knowledge-concept graph completion in education. *IEEE Access*, *11*, 15002–15013. http://dx.doi.org/10.1109/ACCESS.2023.3244614.

Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., et al. (2020). Unsupervised cross-lingual representation learning at scale. In D. Jurafsky, J. Chai, N. Schluter, J. R. Tetreault (Eds.), *Proceedings of the 58th annual meeting of the association for computational linguistics, ACL 2020, online, July 5-10, 2020* (pp. 8440–8451). Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/2020.acl-main.747.

Conneau, A., & Lample, G. (2019). Cross-lingual language model pretraining. In H. M. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. B. Fox, & R. Garnett (Eds.), *Advances in neural information processing systems 32: annual conference on neural information processing systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada* (pp. 7057–7067). URL: https://proceedings.neurips.cc/paper/2019/hash/c04c19c2c2474dbf5f7ac4372c5b9af1-Abstract.html.

Corro, L. D., & Gemulla, R. (2013). ClausIE: clause-based open information extraction. In D. Schwabe, V. A. F. Almeida, H. Glaser, R. Baeza-Yates, & S. B. Moon (Eds.), *22nd international world wide web conference, WWW '13, rio de Janeiro, Brazil, May 13-17, 2013* (pp. 355–366). International World Wide Web Conferences Steering Committee / ACM, http://dx.doi.org/10.1145/2488388.2488420.

Cui, L., Wei, F., & Zhou, M. (2018). Neural open information extraction. In I. Gurevych, & Y. Miyao (Eds.), *Proceedings of the 56th annual meeting of the association for computational linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, volume 2: short papers* (pp. 407–413). Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/P18-2065, URL: https://aclanthology.org/P18-2065/.

Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2019). BERT: pre-training of deep bidirectional transformers for language understanding. In J. Burstein, C. Doran, & T. Solorio (Eds.), *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, volume 1 (long and short papers)* (pp. 4171–4186). Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/n19-1423.

Fader, A., Soderland, S., & Etzioni, O. (2011). Identifying relations for open information extraction. In *Proceedings of the 2011 conference on empirical methods in natural language processing, EMNLP 2011, 27-31 July 2011, John McIntyre conference centre, Edinburgh, UK, a meeting of SIGDAT, a special interest group of the ACL* (pp. 1535–1545). ACL, URL: https://aclanthology.org/D11-1142/.

Gamallo, P., & García, M. (2015). Multilingual open information extraction. In F. C. Pereira, P. Machado, E. Costa, & A. Cardoso (Eds.), *Lecture notes in computer science*: vol. 9273, *Progress in artificial intelligence - 17th portuguese conference on artificial intelligence, EPIA 2015, Coimbra, Portugal, September 8-11, 2015. proceedings* (pp. 711–722). Springer, http://dx.doi.org/10.1007/978-3-319-23485-4_72.

Gashteovski, K., Gemulla, R., & Corro, L. D. (2017). MinIE: Minimizing facts in open information extraction. In M. Palmer, R. Hwa, & S. Riedel (Eds.), *Proceedings of the 2017 conference on empirical methods in natural language processing, EMNLP 2017, Copenhagen, Denmark, September 9-11, 2017* (pp. 2630–2640). Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/d17-1278.

Gashteovski, K., Yu, M., Kotnis, B., Lawrence, C., Glavas, G., & Niepert, M. (2021). BenchIE: Open information extraction evaluation based on facts, not tokens. CoRR abs/2109.06850. arXiv:2109.06850. URL: https://arxiv.org/abs/2109.06850.

Guo, H., Guo, Y., Yang, J., Liu, J., Li, Z., Zheng, T., et al. (2023). LogLG: Weakly supervised log anomaly detection via log-event graph construction. In *International conference on database systems for advanced applications* (pp. 490–501). Springer.

Guo, H., Liu, J., Huang, H., Yang, J., Li, Z., Zhang, D., et al. (2022). LVP-M3: Language-aware visual prompt for multilingual multimodal machine translation. In *Proceedings of the 2022 conference on empirical methods in natural language processing* (pp. 2862–2872).

He, P., Liu, X., Gao, J., & Chen, W. (2021). Deberta: decoding-enhanced bert with disentangled attention. In *9th international conference on learning representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, URL: https://openreview.net/forum?id=XPZIaotutsD.

Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., et al. (2022). Lora: Low-rank adaptation of large language models. In *The tenth international conference on learning representations, ICLR 2022, Virtual Event, April 25-29, 2022*. OpenReview.net.

Jeronymo, V., Bonifacio, L. H., Abonizio, H., Fadaee, M., de Alencar Lotufo, R., Zavrel, J., et al. (2023). Inpars-v2: Large language models as efficient dataset generators for information retrieval. http://dx.doi.org/10.48550/arXiv.2301.01820, CoRR abs/2301.01820. arXiv:2301.01820.

Jia, S., Shijia, E., Ding, L., Chen, X., & Xiang, Y. (2022). Hybrid neural tagging model for open relation extraction. *Expert Systems with Applications*, *200*, Article 116951. http://dx.doi.org/10.1016/j.eswa.2022.116951.

Kolluru, K., Adlakha, V., Aggarwal, S., Mausam, & Chakrabarti, S. (2020). OpenIE6: Iterative grid labeling and coordination analysis for open information extraction. In B. Webber, T. Cohn, Y. He, & Y. Liu (Eds.), *Proceedings of the 2020 conference on empirical methods in natural language processing, EMNLP 2020, online, November 16-20, 2020* (pp. 3748–3761). Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/2020.emnlp-main.306.

Kolluru, K., Aggarwal, S., Rathore, V., Mausam, & Chakrabarti, S. (2020). IMoJIE: Iterative memory-based joint open information extraction. In D. Jurafsky, J. Chai, N. Schluter, & J. R. Tetreault (Eds.), *Proceedings of the 58th annual meeting of the association for computational linguistics, ACL 2020, online, July 5-10, 2020* (pp. 5871–5886). Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/2020.acl-main.521.

Lauscher, A., Song, Y., & Gashteovski, K. (2019). MinScIE: Citation-centered open information extraction. In M. Bonn, D. Wu, J. S. Downie, & A. Martaus (Eds.), *19th ACM/IEEE joint conference on digital libraries, JCDL 2019, Champaign, IL, USA, June 2-6, 2019* (pp. 386–387). IEEE, http://dx.doi.org/10.1109/JCDL.2019.00083.

Li, S., Zhuang, S., Song, W., Zhang, X., Chen, H., & Hao, A. (2023). Sequential texts driven cohesive motions synthesis with natural transitions. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 9498–9508).

Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., et al. (2023). Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, *379*(6637), 1123–1130.

Liu, D., Niehues, J., Cross, J., Guzmán, F., & Li, X. (2021). Improving zero-shot translation by disentangling positional information. In C. Zong, F. Xia, W. Li, & R. Navigli (Eds.), *Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing, ACL/IJCNLP 2021, (volume 1: long papers), virtual event, August 1-6, 2021* (pp. 1259–1273). Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/2021.acl-long.101.

Liu, J., Yu, T., Peng, H., Sun, M., & Li, P. (2022). Cross-lingual cross-modal consolidation for effective multilingual video corpus moment retrieval. In *Findings of the association for computational linguistics: NAACL 2022* (pp. 1854–1862).

Liu, J., Zhou, S., Wu, Y., Chen, K., Ouyang, W., & Xu, D. (2020). Block proposal neural architecture search. *IEEE Transactions on Image Processing*, *30*, 15–25.

Loshchilov, I., & Hutter, F. (2019). Decoupled weight decay regularization. In *7th international conference on learning representations, ICLR 2019, New Orleans, La, USA, May 6-9, 2019*. OpenReview.net, URL: https://openreview.net/forum?id=Bkg6RiCqY7.

Lu, H., Huang, H., Zhang, D., Wei, F., & Lam, W. (2024). Revamping multilingual agreement bidirectionally via switched back-translation for multilingual neural machine translation. In Y. Graham, & M. Purver (Eds.), *Findings of the Association for Computational Linguistics: EACL 2024, St. Julian's, Malta, March 17-22, 2024* (pp. 264–275). Association for Computational Linguistics, https://aclanthology.org/2024.findings-eacl.19.

Maaten, L. v. d., & Hinton, G. (2008). Visualizing data using t-SNE. *JMLR*, *9*, 2579–2605.

Mausam (2016). Open information extraction systems and downstream applications. In S. Kambhampati (Ed.), *Proceedings of the twenty-fifth international joint conference on artificial intelligence, IJCAI 2016, New York, NY, USA, 9-15 July 2016* (pp. 4074–4077). IJCAI/AAAI Press, URL: http://www.ijcai.org/Abstract/16/604.

Mausam, Schmitz, M., Soderland, S., Bart, R., & Etzioni, O. (2012). Open language learning for information extraction. In J. Tsujii, J. Henderson, & M. Pasca (Eds.), *Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning, EMNLP-CoNLL 2012, July 12-14, 2012, Jeju Island, Korea* (pp. 523–534). ACL, URL: https://aclanthology.org/D12-1048/.

Niklaus, C., Cetto, M., Freitas, A., & Handschuh, S. (2018). A survey on open information extraction. In E. M. Bender, L. Derczynski, & P. Isabelle (Eds.), *Proceedings of the 27th international conference on computational linguistics, COLING 2018, Santa Fe, New Mexico, USA, August 20-26, 2018* (pp. 3866–3878). Association for Computational Linguistics, URL: https://aclanthology.org/C18-1326/.

OpenAI (2023). GPT-4 technical report. arXiv:2303.08774.

Rahat, M., & Talebpour, A. (2018). Open information extraction as an intermediate semantic structure for Persian text summarization. *International Journal on Digital Libraries*, *19*(4), 339–352. http://dx.doi.org/10.1007/s00799-018-0244-z.

Ro, Y., Lee, Y., & Kang, P. (2020). Multi^{2}OIE: Multilingual open information extraction based on multi-head attention with BERT. In T. Cohn, Y. He, & Y. Liu (Eds.), *Findings of the association for computational linguistics: EMNLP 2020, online event, 16-20 November 2020* (pp. 1107–1117). Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/2020.findings-emnlp.99.

Roy, A., Park, Y., Lee, T., & Pan, S. (2019). Supervising unsupervised open information extraction models. In K. Inui, J. Jiang, V. Ng, & X. Wan (Eds.), *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019* (pp. 728–737). Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/D19-1067.

Sarhan, I., & Spruit, M. R. (2019). Contextualized word embeddings in a neural open information extraction model. In E. Métais, F. Meziane, S. Vadera, V. Sugumaran, & M. Saraee (Eds.), *Lecture notes in computer science*: *vol. 11608*, *Natural language processing and information systems - 24th international conference on applications of natural language to information systems, NLDB 2019, Salford, UK, June 26-28, 2019, proceedings* (pp. 359–367). Springer, http://dx.doi.org/10.1007/978-3-030-23281-8_31.

Song, W., Jin, X., Li, S., Chen, C., Hao, A., & Hou, X. (2023). FineStyle: Semantic-aware fine-grained motion style transfer with dual interactive-flow fusion. *IEEE Transactions on Visualization and Computer Graphics*.

Song, W., Zhang, X., Guo, Y., Li, S., Hao, A., & Qin, H. (2023). Automatic generation of 3D scene animation based on dynamic knowledge graphs and contextual encoding. *International Journal of Computer Vision*, *131*(11), 2816–2844.

Stanovsky, G., & Dagan, I. (2016). Creating a large benchmark for open information extraction. In J. Su, X. Carreras, & K. Duh (Eds.), *Proceedings of the 2016 conference on empirical methods in natural language processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016* (pp. 2300–2305). The Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/d16-1252.

Stanovsky, G., Michael, J., Zettlemoyer, L., & Dagan, I. (2018). Supervised open information extraction. In M. A. Walker, H. Ji, & A. Stent (Eds.), *Proceedings of the 2018 conference of the North American chapter of the association for computational linguistics: human language technologies, NAACL-HLT 2018, New Orleans, Louisiana, USA, June 1-6, 2018, volume 1 (long papers)* (pp. 885–895). Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/n18-1081.

Tan, X., Chen, J., He, D., Xia, Y., Qin, T., & Liu, T. (2019). Multilingual neural machine translation with language clustering. In K. Inui, J. Jiang, V. Ng, & X. Wan (Eds.), *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019* (pp. 963–973). Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/D19-1089.

Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., et al. (2023). Llama 2: Open foundation and fine-tuned chat models. http://dx.doi.org/10.48550/arXiv.2307.09288, CoRR abs/2307.09288. arXiv:2307.09288.

Wan, Z., Cheng, F., Mao, Z., Liu, Q., Song, H., Li, J., et al. (2023). GPT-RE: in-context learning for relation extraction using large language models. http://dx.doi.org/10.48550/arXiv.2305.02105, CoRR abs/2305.02105. arXiv:2305.02105.

Wang, Y., Zhai, C., & Hassan, H. (2020). Multi-task learning for multilingual neural machine translation. In B. Webber, T. Cohn, Y. He, & Y. Liu (Eds.), *Proceedings of the 2020 conference on empirical methods in natural language processing* (pp. 1022–1034). Online: Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/2020.emnlp-main.75, URL: https://aclanthology.org/2020.emnlp-main.75.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., et al. (2022). Chain-of-thought prompting elicits reasoning in large language models. In *NeurIPS*. URL: http://papers.nips.cc/paper_files/paper/2022/hash/9d5609613524ecf4f15af0f7b31abca4-Abstract-Conference.html.

White, A. S., Reisinger, D. A., Sakaguchi, K., Vieira, T., Zhang, S., Rudinger, R., et al. (2016). Universal decompositional semantics on universal dependencies. In J. Su, X. Carreras, K. Duh (Eds.), *Proceedings of the 2016 conference on empirical methods in natural language processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016* (pp. 1713–1723). The Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/d16-1177.

Yang, Z., Ding, M., Guo, Y., Lv, Q., & Tang, J. (2022). Parameter-efficient tuning makes a good classification head. In Y. Goldberg, Z. Kozareva, & Y. Zhang (Eds.), *Proceedings of the 2022 conference on empirical methods in natural language processing,* *EMNLP 2022, Abu Dhabi, United Arab Emirates, December 7-11, 2022* (pp. 7576–7586). Association for Computational Linguistics, URL: https://aclanthology.org/2022.emnlp-main.514.

Yang, J., Guo, H., Yin, Y., Bai, J., Wang, B., Liu, J., et al. (2024). M3p: towards multimodal multilingual translation with multimodal prompt. In N. Calzolari, M. Kan, V. Hoste, A. Lenci, S. Sakti, N. Xue (Eds.), *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation, LREC/COLING 2024, 20-25 May, 2024, Torino, Italy* (pp. 10858–10871). ELRA and ICCL, https://aclanthology.org/2024.lrec-main.948.

Yang, J., Huang, S., Ma, S., Yin, Y., Dong, L., Zhang, D., et al. (2022). CROP: zero-shot cross-lingual named entity recognition with multilingual labeled sequence translation. In Y. Goldberg, Z. Kozareva, Y. Zhang (Eds.), *Findings of the association for computational linguistics: EMNLP 2022, Abu Dhabi, United Arab Emirates, December 7-11, 2022* (pp. 486–496). Association for Computational Linguistics, URL: https://aclanthology.org/2022.findings-emnlp.34.

Yang, J., Ma, S., Huang, H., Zhang, D., Dong, L., Huang, S., et al. (2021). Multilingual machine translation systems from microsoft for WMT21 shared task. In L. Barrault, O. Bojar, F. Bougares, R. Chatterjee, M. R. Costa-jussà, C. Federmann, M. Fishel, A. Fraser, M. Freitag, Y. Graham, R. Grundkiewicz, P. Guzman, B. Haddow, M. Huck, A. Jimeno-Yepes, P. Koehn, T. Kocmi, A. Martins, M. Morishita, C. Monz (Eds.), *Proceedings of the Sixth Conference on Machine Translation, WMT@EMNLP 2021, Online Event, November 10-11, 2021* (pp. 446–455). Association for Computational Linguistics, https://aclanthology.org/2021.wmt-1.54.

Yang, J., Ma, S., Zhang, D., Wu, S., Li, Z., & Zhou, M. (2020). Alternating language modeling for cross-lingual pre-training. In *The thirty-fourth AAAI conference on artificial intelligence, AAAI 2020, the thirty-second innovative applications of artificial intelligence conference, IAAI 2020, the tenth AAAI symposium on educational advances in artificial intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020* (pp. 9386–9393). AAAI Press, URL: https://ojs.aaai.org/index.php/AAAI/article/view/6480.

Yang, J., Yin, Y., Ma, S., Zhang, D., Li, Z., & Wei, F. (2022). HLT-MT: high-resource language-specific training for multilingual neural machine translation. http://dx.doi.org/10.48550/arXiv.2207.04906, CoRR abs/2207.04906. arXiv:2207.04906.

Yang, J., Yin, Y., Ma, S., Zhang, D., Wu, S., Guo, H., et al. (2022). UM4: unified multilingual multiple teacher-student model for zero-resource neural machine translation. In L. D. Raedt (Ed.), *Proceedings of the thirty-first international joint conference on artificial intelligence, IJCAI 2022, Vienna, Austria, 23-29 July 2022* (pp. 4454–4460). ijcai.org, http://dx.doi.org/10.24963/ijcai.2022/618.

Yang, J., Yin, Y., Yang, L., Ma, S., Huang, H., Zhang, D., et al. (2023). Gtrans: grouping and fusing transformer layers for neural machine translation. *IEEE ACM Trans. Audio Speech Lang. Process.*, *31*, 1489–1498. http://dx.doi.org/10.1109/TASLP.2022.3221040.

Yin, Y., Yang, Y., Yang, J., & Liu, Q. (2023). FinPT: Financial risk prediction with profile tuning on pretrained foundation models. arXiv preprint arXiv:2308.00065.

Zhan, J., & Zhao, H. (2020). Span model for open information extraction on accurate corpus. In *The thirty-fourth AAAI conference on artificial intelligence, AAAI 2020, the thirty-second innovative applications of artificial intelligence conference, IAAI 2020, the tenth AAAI symposium on educational advances in artificial intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020* (pp. 9523–9530). AAAI Press, URL: https://ojs.aaai.org/index.php/AAAI/article/view/6497.

Zhou, R., Li, X., Bing, L., Cambria, E., Si, L., & Miao, C. (2022). ConNER: Consistency training for cross-lingual named entity recognition. In Y. Goldberg, Z. Kozareva, & Y. Zhang (Eds.), *Proceedings of the 2022 conference on empirical methods in natural language processing, EMNLP 2022, Abu Dhabi, United Arab Emirates, December 7-11, 2022* (pp. 8438–8449). Association for Computational Linguistics, URL: https://aclanthology.org/2022.emnlp-main.577.