XLTU: A Cross-Lingual Model in Temporal Expression Extraction for Uyghur

Yifei Liang[1,2], Lanying Li[3], Rui Liu[1,2], Ahtam Ahmat[1,2], and Lei Jiang[1,2(✉)]

[1] Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China
{liangyifei,liurui3221,aihetanmuaihemaiti,jianglei}@iie.ac.cn
[2] School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China
[3] Civil Aviation Flight University Of China Xinjin Flight College, China
lly010214@163.com

Abstract. Temporal expression extraction (TEE) plays a crucial role in natural language processing (NLP) tasks, enabling the capture of temporal information for downstream tasks such as logical reasoning and information retrieval. However, current TEE research mainly focuses on resource-rich languages like English, leaving a gap for minor languages (e.g. Uyghur) in research. To address these issues, we create an English-Uyghur cross-lingual dataset specifically for the task of temporal expression extraction in Uyghur. Besides, considering the unique characteristics of Uyghur, we propose XLTU, a **C**ross-**L**ingual model in **T**emporal expression extraction for **U**yghur, and utilize multi-task learning to help transfer the knowledge from English to Uyghur. We compare XLTU with different models on our dataset, and the results demonstrate that our model XLTU achieves the SOTA results on various evaluation metrics. We make our code and dataset publicly available[1].

Keywords: temporal expression extraction · Uyghur · cross-lingual · multi-task learning.

1 Introduction

Currently, temporal expression extraction (TEE) is an important NLP task [1], which specifically refers to detecting expressions about time such as date, duration, etc. This task has wide-ranging applications in downstream tasks, including question answering [2], information retrieval [3], and causal reasoning [4]. In the past, the work of TEE mainly relies on rule-based approaches [5,6], while the current focus has shifted towards leveraging deep learning techniques [7,8,9]. However, the field of TEE for minor languages still lacks sufficient research and development, indicating a noticeable scarcity in this area. Due to the scarcity of annotated datasets for minor languages, it shows the suboptimal performance of deep learning methods in these languages.

✉ Corresponding author
[1] https://github.com/lyfcsdo2011/XLTU

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63751-3_11

In this study, we focus on addressing TEE in Uyghur, a language with distinctive characteristics that set it apart from more widely used languages. Most languages, used for pretraining (e.g. mBERT [10], XLM-R [11]), are read and written from left to right. However, Uyghur is read and written in the opposite direction, as depicted in Figure 1. Additionally, the vocabulary of Uyghur significantly differs from that of European languages. When applying pre-trained cross-lingual models to the Uyghur language, these differences will lead to substantial deviations in feature learning and knowledge transfer, because the models cannot obtain Uyghur language features well.

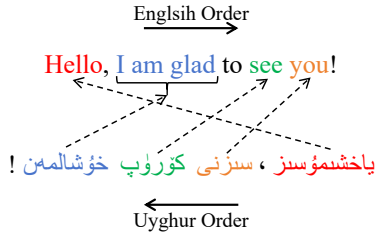


Fig. 1. The order difference between English and Uyghur. English follows a left-to-right pattern, while Uyghur follows a right-to-left pattern.

In order to address these challenges, it is crucial to expand the high-quality datasets and improve the performance of TEE methods in Uyghur. In our study, we create a high-quality English-Uyghur cross-lingual dataset specifically for TEE in Uyghur. This dataset allows us to transfer the knowledge from English to Uyghur. Besides, considering the unique characteristics of Uyghur, we propose XLTU: a **Cross-L**ingual model in **T**emporal expression extraction for **U**yghur, a method based on a pre-trained model, and utilize multi-task learning (MTL) [12] to facilitate transfer of English knowledge to Uyghur in TEE (as shown in Figure 2).

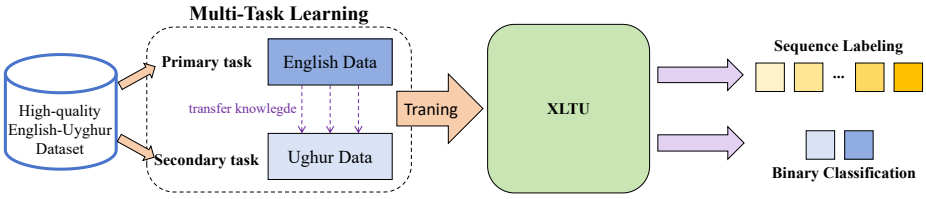


Fig. 2. Overview diagram of our work. We create a high-quality English-Uyghur dataset for TEE (the left). Besides, we propose XLTU, and utilize multi-task learning to train the model. The primary task is formulated as a sequence labeling task, and the secondary task is formulated as a binary classification task.

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63751-3_11

Our model involves two tasks: a primary task and a secondary task. In the primary task, we train the model using existing annotated English TEE data. This helps the model learn the explicit knowledge and understand the structure of temporal expressions in English. In the secondary task, we map the annotated English TEE data samples to Uyghur. This process allows us to obtain sentence-level labels (containing one or more time expressions) based on the original token-level labels. In a weakly supervised manner, we transfer the implicit knowledge learned in the target language by teaching the model to detect whether the target language contains temporal expressions.

The main contributions of this paper are: 1) We create a high-quality English-Uyghur cross-lingual dataset for TEE multi-task in minor language Uyghur. 2) We propose XLTU utilizing multi-task learning for TEE in Uyghur. 3) We show that XLTU can effectively promote the learning of Uyghur language in TEE, and achieves SOTA results on our dataset.

2 Related Work

Although TEE is very important in NLP, there are limited studies on this task, particularly for languages with limited data resources. Most existing research in this field has primarily focused on resource-rich languages like English. Currently, there are two main types of technologies used for TEE.

One is a rule/pattern-based method. HeidelTime [5] is the best-performing method so far and covers more than ten languages. It is driven by a carefully tuned set of rules. This approach is later extended to additional languages via HeidelTime-auto [13], which exploits language-independent processing and rules. Other methods, such as SynTime [6], SUTIME [14], and PTime [15], utilize heuristic rule-based approaches and pattern-learning techniques.

The second type of approach for TEE involves deep learning methods, and this is also a current major research direction. For instance, [16] proposes an RNN-based model, while [7] utilizes BERT with linear classifiers. [8] feeds mBERT embeddings into BiLSTM with CRF layers and outperforms HeidelTime-auto in four languages. [9] proposes a framework based on pre-trained models and learns in a multi-task manner. However, compared to other tasks, the performances of the deep learning-based methods reported are inferior in cross-lingual TEE. This is highly attributed to the lack of annotated datasets for minority languages. In our work, we propose XLTU to make the model learn cross-lingual features much better. Besides, we create a high-quality cross-lingual dataset to make up for the insufficient data available for minor languages.

Moreover, applying the label projection method can better solve the problem of lack of data in TEE. TMP [17] is originally proposed for cross-lingual named entity recognition (NER) [18], projecting English data in IOB (which means Inside Outside Beginning) tagging format [19] using machine translation, orthographic and phonetic similarities to other languages of the package. [9] proposes a MTL framework to transfer temporal knowledge of source languages into target languages.

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63751-3_11

In the early stage, an important motivation for MTL is to alleviate the problem of data sparsity in machine learning. When the *big data* era emerges, multi-task learning is more effective which utilizes more data from different learning tasks than single-task learning. [12] proposes a model called MT-DNN which combines multi-task learning and language model pre-training for language representation learning. MulT [20] is an end-to-end multitask learning Transformer [21] framework to simultaneously learn multiple high-level vision tasks. DeMT [22] is a novel MTL model that combines both merits of deformable CNN and query-based Transformer for multi-task learning of dense prediction.

3 English-Uyghur Cross-Lingual TEE Dataset

3.1 Temporal Expression Types

ISO-TimeML [23] has already presented the TEE dataset annotation guideline, there are four types of temporal expressions, i.e., *Date*, *Time*, *Duration*, and *Set*. *Date* refers to a calendar date, usually a day or a larger unit of time. *Time* refers to a time of day, with a granularity smaller than a day. *Duration* refers to an expression that clearly describes a period of time. *Set* refers to a regular set of time of recurrence. An intuitive representation can be seen in Table 1.

Table 1. Temporal expressions of four types. Definitions of types Seeing 3.2.

Please pay attention, I will see you next $\underbrace{\text{Friday}}_{\text{Date}}$, have a good rest.
The warrants may be exercised until $\underbrace{\text{90 days}}_{\text{Duration}}$ after their issue date.
I persist in exercising $\underbrace{\text{every day}}_{\text{Set}}$ after work to keep a healthy body.
I have a stomach ache, I need to go to hospital on $\underbrace{\text{Friday}}_{\text{Time}}$ morning.

3.2 Dataset Structure

For the English dataset, following [9], we collect TE3 [1], Wikiwars [24] and Tweets [6]. As for the Uyghur datasets, a part of them is obtained through machine translation of English TE3 [1] and Tweets [6]. We also employ web crawling techniques to collect additional data, which is then carefully cleaned and filtered to ensure high-quality data for manual labeling. According to the multi-tasks we have designed, the primary task takes the form of cross-lingual sequence labeling, which includes the Named Entity Recognition (NER) [18] task. Meanwhile, the secondary task is designed as a binary classification task.

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63751-3_11

For the primary task, the training data consists of the whole English dataset in the NER format. The test data consists of Uyghur data that has been manually annotated in the NER format and is used to evaluate the cross-lingual capabilities of our model by predicting temporal expressions in Uyghur. In total, we annotate 22,726 pieces of Uyghur data for this task, including 330 pieces labeled as *Date*, 100 pieces as *Duration*, 33 pieces as *Set*, and 40 pieces as *Time* (as shown in Table 2).

Regarding the secondary task, the training data comprises Uyghur sentences obtained through machine translation from English. These sentences are manually labeled for classification.

Table 2. The statistics of the English-Uyghur cross-lingual datasets.

Language	Dataset	Domain	Expressions	Dates	Times	Durations	Sets	Tokens
English	TE3 [1]	News	1830	1471	34	291	34	124592
	Wikiwar [24]	Narrative	2634	2634	0	0	0	
	Tweet [6]	Utterance	1128	717	173	200	38	
Uyghur	Our work	Websites	503	330	40	100	33	22726

4 Proposed Model

TEE is formalized as a sequence labeling task. Inspired by [8,9,12,25], the architecture of our model is shown in Figure 3.

4.1 Pre-trained Multilingual Model

Considering the limited availability of resources for the Uyghur language, we utilize the base XLM-Roberta model (XLM-R) [11] as the backbone. XLM-R is a state-of-the-art multilingual model and outperforms other models in various cross-lingual tasks. One of the main advantages of XLM-R is its extensive training on a wide range of languages and datasets, this gives XLM-R a larger vocabulary to learn and adapt to the characteristics of Uyghur words. It has also introduced three training targets to further enhance its performance in cross-lingual tasks. The pre-trained multilingual model consists of lexicon and Transformer encoder layers, as shown in Figure 3. The backbone of the model is shared across all the MTL tasks during both the training and testing phases.

4.2 TextCNN

TextCNN [25] has already demonstrated that Convolutional Neural Networks (CNN) [26] can be effectively applied to text processing tasks, yielding impressive results. CNNs can be combined with pre-trained language models to further extract informative features for downstream tasks (e.g. classification),

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63751-3_11

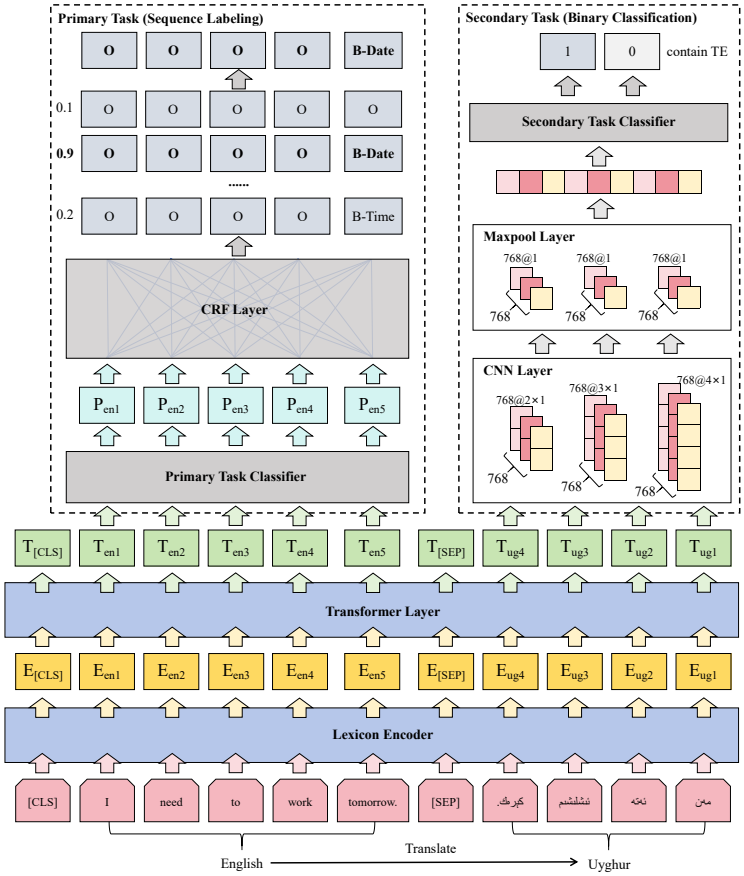


Fig. 3. Model structure of XLTU. It shows how our model transfers knowledge from English to Uyghur through the primary and the secondary tasks.

as seen in models like BERT-CNN [27]. Considering the specific characteristics of the Uyghur language, we have introduced a CNN neural network after the pre-trained language model in our architecture. This allows us to leverage the last hidden state of the language model to extract additional and right-to-left features, which are then utilized in the secondary task. This CNN component enhances the model's ability to capture relevant information and improve performance on the given task, as shown in Figure 3.

4.3 Conditional Random Fields

Conditional Random Field (CRF) [28] is widely used in sequence labeling tasks. It has been proven to enhance the performance of sequence labeling models and address issues such as mismatched predicted labels or labels that do not start with the 'B' label, such as 'O O B-Date I-Time' or 'O O I-Date I-Date'.

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63751-3_11

Figure 3 demonstrates that the undirected graphical structure of CRF enables the model to learn the context relationship between each token in both directions. Therefore, CRF helps our model better capture the contextual and right-to-left information and make more accurate predictions.

4.4 Cross-Lingual Knowledge Transfer based on MTL

Our model is capable of transferring knowledge from English to Uyghur. To facilitate explicit and implicit knowledge transfer, we have designed the primary and secondary tasks on top of the backbone. The primary task focuses on explicitly encoded time expressions in English. It is formulated as a sequence labeling task and utilizes the training data of English to train the backbone network, which includes the primary task classifier and CRF layer. The architecture is illustrated in the top of left corner of Figure 3. In the primary task, We incorporate two different loss functions $\mathcal{L}_t$ and $\mathcal{L}_{crf}$:

$$\mathcal{L}_t = -\sum_{i=1}^{b}\sum_{j=1}^{m_i} \mathbb{1}(y_{ij}, c) log(softmax(W_1 \cdot x)), \tag{1}$$

$$\mathcal{L}_{crf} = -logP(Y|X;\theta), \tag{2}$$

where $\mathcal{L}_t$ represents the loss between the labels directly predicted by the backbone outputs and the ground-truth labels. $\mathcal{L}_{crf}$ represents the loss between the predicted labels after the backbone outputs pass through CRF and the ground-truth labels. b is the total number of input sequences and m_i is the length of the i_{th} sequence. $x \in \mathbb{R}^d$ is the embedding of the j_{th} token in the i_{th} sequence of output by the backbone model. d is its dimension. $c = argmax(W_1 \cdot x)$ is the predicted label for each token while y_{ij} is the ground-truth label of each token. $W_1 \in \mathbb{R}^{|c| \times d}$ is the classifier parameters of the primary task. $|c|$ is the total number of unique ground-truth labels. $\mathbb{1}(,)$ is 1 if two are equal, 0 otherwise. Y is the set of ground-truth labels while X it the set of predicted labels, θ is the parameters of backbone model. $P(Y|X;\theta)$ represents the probability of labeling sequence Y under the condition of given input sequence X, which can be formulated as:

$$P(Y|X;\theta) = \frac{1}{Z(X;\theta)} \exp(\sum_{i=1}^{n}\sum_{j=1}^{k} \theta_j f_j(y_{i-1}, y_i, x_i)), \tag{3}$$

where $Z(X;\theta)$ is the normalization factor, $f_j(y_{i-1}, y_i, x_i)$ is the characteristic function of CRF, θ_j is the weight corresponding to the characteristic function. And the final target of the primary task is to minimize the $\mathcal{L}_{sl}$:

$$\mathcal{L}_{sl} = \alpha \cdot \mathcal{L}_{crf} + \beta \cdot \mathcal{L}_t, \tag{4}$$

where α and β are the weight ratios corresponding to the two losses.

After finishing the primary task, our backbone has already learned the explicit knowledge of TEE. So the secondary task implicitly captures the linguistic

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63751-3_11

features of temporal expressions in Uyghur with the explicit knowledge learned in the primary task. It is formulated as a binary classification task. The input for this task is the Uyghur sequences, and the labels are sentence-level classification labels (as mentioned earlier). In this task, the language features of the last hidden state of the model are further extracted using a CNN, which helps in classifying the sequences. The secondary task enables the model to learn the features of temporal expressions in the target Uyghur language, implicitly. This is a weakly-supervised task and requires no token-level labels for each Uyghur token. The manually annotated token-level labels from the Uyghur datasets are used to evaluate the cross-lingual capability of the model after training. The ultimate objective of the secondary task is to minimize the $\mathcal{L}_{bc}$:

$$\mathcal{L}_{bc} = -\sum_{i=1}^{b} \mathbb{1}(y_i', c_i') log(softmax(W_2 \cdot x')), \tag{5}$$

where $x' \in \mathbb{R}^d$ is the sequence embedding output of CNN by passing the outputs of the backbone model to it. $c' = argmax(W_2 \cdot x')$ is the predicted sequence label of ith sequence while y_i' is the ground-truth sequence label. $W_2 \in \mathbb{R}^{2 \times d}$ is the classifier parameters of secondary task. Then We train our model concurrently by multi-task learning.

4.5 Data Format

We provide an illustrative example in Figure 4 to demonstrate how knowledge is transferred from English to Uyghur in our model. In the primary task, the model extracts the explicit features of temporal expressions in English through a sequence labeling task. In the secondary task, the model takes the English translations of X_1 and X_2 as input. Y_3 and Y_4 indicate whether the sequences contain temporal expressions. The value of 1 indicates the presence of temporal expressions, while the value of 0 indicates their absence. These labels can be inferred from the labels Y_1 and Y_2 obtained in the primary task.

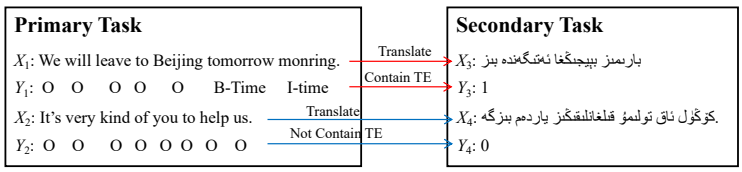


Fig. 4. An illustrative training example. In primary task, X_1 and X_2 are the inputs, while Y_1 and Y_2 are the corresponding output labels. In secondary task, X_3 and X_4 are the Uyghur translation of X_1 and X_2, while the outputs Y_3 and Y_4 can be inferred from Y_1 and Y_2 whether the sequences contain temporal expressions.

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63751-3_11

5 Experiments

5.1 Dataset

We utilize our English-Uyghur cross-lingual dataset. The dataset statistics are presented in Table 2. For the Uyghur language, we utilize the entire Uyghur dataset for test. For the English language, we utilize the entire English dataset including three separate datasets for training.

5.2 Baselines

To evaluate the performance of our model, we compare it with several popular deep learning methods, specifically focusing on cross-lingual models. We compare our model to the following models:

- mBERT [10]: This model is based on the multilingual BERT architecture.
- XLM-R [11]: We compare our model with the base and large versions of the vanilla XLM-Roberta model. XLM-R is a transformer-based language model specifically designed for cross-lingual tasks.
- XLTime [9]: We compare our model with three different variations of XLTime, which is a cross-lingual temporal expression extraction model. XLTime has shown promising results in capturing temporal expressions across multiple languages by using the MTL method.

5.3 Evaluation Approaches and Metrics

Following the previous research [1,9], we evaluate our model using two different approaches and measure the performance using F1-score, precision, and recall. The first approach is in *strict match* [1] evaluation, where all tokens of a temporal expression must be correctly identified for it to be considered as correctly extracted. This means that the predicted labels should match the ground-truth labels exactly in terms of both the sequence and the type of the temporal expression. For example, if the ground-truth labels are 'O O B-Set', any other prediction, such as 'O O B-Date', would be considered completely wrong. This evaluation approach is referred to as *with type*.

The second approach called *without type*, takes a more lenient approach. In this evaluation, as long as the labels of a temporal expression are predicted, regardless of whether the types match, it will be considered as correct. For example, if the ground-truth labels are 'O O B-Set', a prediction of labels 'O O B-Date' would be counted as correct. This approach focuses on capturing the presence of temporal expressions rather than matching their specific types.

5.4 Experiment Details

We adopt the base of the XLM-Roberta model (XLM-R) as our backbone which consists of 12 layers, 12 heads, and 270M parameters. We set the embedding

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63751-3_11

dimension d as 768 to be consistent. We set batch size as 4 and dropout ratio as 0.2. We employ the AdamW as our optimizer with a learning rate of $7e^{-6}$ and a warm-up proportion of 0.5. We set the values of α and β as 0.2 and 0.8 in (4). For the CNN layer, we use a filter size of (2, 3, 4), and the kernel size is determined by the dimensions (k, d), with k corresponding to the filter size. we include a ReLU layer as an activation function following the backbone. We train all models for 50 epochs and select the best model for prediction. In order to meet the setting requirements of sequence labeling, the dataset is annotated and designed in the IOB2 format. We train all models on 8×NVIDIA Geforce RTX 3090 GPU.

5.5 Experiment Results

We evaluate our model and baselines on our dataset, employing two evaluation approaches as shown in Table 3. We observe that:

1) In both approaches, XLTU outperforms other models in terms of F1-score, recall, and precision, and achieves the SOTA results.

2) mBERT and XLTime-mBERT perform poorly in both approaches on our dataset. This is probably because their structures are not suitable for extracting features from the Uyghur language. Unlike XLM [29], mBERT simply replaces the training corpus of BERT with multilingual datasets. Although it provides shallow transfer [30] benefits for languages with vocabulary overlap, it may not be helpful for Uyghur with a completely different vocabulary.

3) Comparing XLTime-mBERT with mBERT or XLTime-XLMRbase with XLMR-base, we know that MTL does help the model to transfer knowledge. However, for XLMR-large and XLTime-XLMRlarge, MTL may have a negative impact. The large number of parameters in XLMR-large, combined with the relatively small size of our English-Uyghur dataset, may lead to overfitting during training when MTL is introduced.

4) Comparing to XLTime-XLMRbase, it shows that the introduction of CRF and CNN improves the model's perception of temporal expressions, as well as specific temporal expression label categories, enabling more accurate recognition.

5) We note that the performance of all models is not particularly high (the best F1-score is 0.66). This could be attributed to the characteristics of Uyghur itself and limited dataset. As the language characteristics of English and Uyghur are quite different, the model cannot fully capture the language characteristics of Uyghur through knowledge transfer from English. Nevertheless, compared to other models, our model still demonstrates superior cross-lingual capabilities.

Tabel 4 shows that XLTU performs better in predicting *Date* and *Set* labels. This discrepancy can be attributed to the complexity of the labeled data and the unequal number of labels. Most English data in these labels are simple, so does the corresponding Uyghur data. For example, 'tomorrow' is labeled as 'B-Date', while more complex expression like 'March 15' is labeled as 'B-Date I-Date'. On the other hand, the English data structure for *Duration* and *Time* labels is more complicated. These labels often consist of more than three words. In contrast, the corresponding Uyghur data typically consists of a few words, and the number

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63751-3_11

Table 3. Results of multilingual TEE on English-Uyghur cross-lingual dataset for two approaches. Number with bold is the optimal result, number with underline is the suboptimal result.

	w/type			w/o type		
Model	F1-score	Precision	Recall	F1-score	Precision	Recall
mBERT	0.14	0.39	0.08	0.14	0.53	0.08
XLMR-base	0.34	0.41	0.30	0.42	0.51	0.36
XLMR-large	<u>0.52</u>	<u>0.52</u>	<u>0.54</u>	0.56	0.55	0.58
XLTime-mBERT	0.23	0.23	0.25	0.32	0.35	0.29
XLTime-XLMRbase	0.50	**0.53**	0.52	<u>0.62</u>	<u>0.59</u>	<u>0.66</u>
XLTime-XLMRlarge	0.42	0.40	0.53	0.57	0.51	0.65
XLTU(Ours)	**0.54**	**0.53**	**0.59**	**0.66**	**0.64**	**0.67**

Table 4. Evaluation details of our model for all labels.

w/type of XLTU				
Label	F1	Precision	Recall	Support
Date	0.63	0.56	0.72	330
Duration	0.28	0.40	0.21	100
Set	0.59	0.64	0.55	33
Time	0.46	0.50	0.43	40

of labeled words does not always match the English labeled data (as shown in Figure 5).

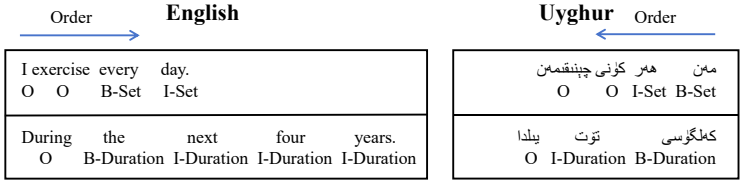


Fig. 5. Examples of labeled data in our English-Uyghur dataset. Not all labels are aligned one by one like *Duration* labels.

5.6 Ablation Study

To examine the effectiveness of the components in our model, we conducted an ablation study by removing the CRF and CNN layers. Table 5 illustrates that our model without CRF layer or CNN layer will degrade the performance on TEE task. According to the results, we know that CRF helps our model to learn contextual and right-to-left features of TEE and make more accurate predictions, while CNN can extract additional and right-to-left features of Uyghur.

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63751-3_11

Table 5. Results of Ablation Study. '-CRF' means our model without CRF layer. '-CNN' means our model without CNN layer.

	w/type			w/o type		
Model	F1	Precision	Recall	F1	Precision	Recall
XLTU	**0.54**	**0.53**	**0.59**	**0.66**	**0.64**	**0.67**
-CRF	0.38	0.39	0.43	0.53	0.52	0.53
-CNN	0.35	0.46	0.32	0.56	0.52	0.60
-CRF & -CNN	0.34	0.41	0.30	0.42	0.51	0.36

Table 6. Results of Comparison Experiments. 'G_i' means XLTU with ith parameter group of (α, β), such as 'G_4' represents 4th group of $(\alpha = 0.4, \beta = 0.6)$.

	w/type of XLTU			w/o type of XLTU		
Group	F1	Precision	Recall	F1	Precision	Recall
G_0	0.38	0.39	0.43	0.53	0.52	0.53
G_1	0.40	0.44	0.49	0.58	0.56	0.60
G_2	**0.54**	**0.53**	**0.59**	**0.66**	**0.64**	**0.67**
G_3	0.45	0.46	0.51	0.65	**0.64**	0.65
G_4	0.42	0.46	0.48	0.57	0.57	0.57
G_5	0.49	0.49	0.54	0.60	0.58	0.61
G_6	0.44	0.50	0.46	0.59	0.63	0.55
G_7	0.47	0.50	0.46	0.64	0.63	0.65
G_8	0.50	0.49	0.53	0.62	0.61	0.62
G_9	0.43	0.44	0.47	0.59	0.60	0.57
G_{10}	0.45	0.47	0.48	0.55	0.55	0.54

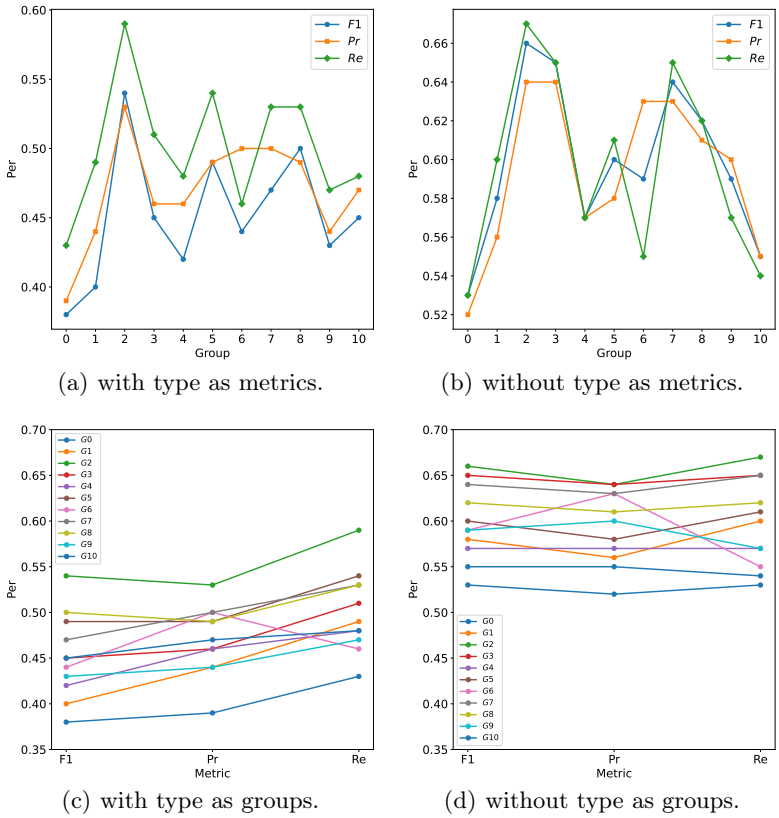
Based on these results, CRF layer plays a more significant role in the *without type* evaluation approach which helps our model better learn the English TEE features in the primary task and then be transferred in the secondary task, while the CNN layer has a greater impact in the *with type* approach. This enables the model to effectively identify the types of labels and avoid mistaking them for other types. On the other hand, the CNN layer aids in extracting Uyghur-specific features that are important for classification which enables the model to effectively identify the presence of temporal expressions.

5.7 Comparison Experiment

To investigate the impact of the two loss weight ratios, α and β, in the model in (4), We conduct additional comparison experiments. We perform 11 sets of experiments, varying the values of (α, β) from $(0, 1.0)$, $(0.1, 0.9)$, ..., to $(0.9, 0.1)$, $(1.0, 0)$. We group these experiments into G_0 to G_{10} for easy reference. We evaluate the results separately using the *with type* and *without type* evaluation approaches and visualize the experimental results based on both individual metrics and grouped results, as shown in Figure 6.

From Table 6 we observe that $G_2(\alpha = 0.2, \beta = 0.8)$ as mentioned in Section 5.4, achieves the SOTA result in both the *with type* and *without type* evaluation

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63751-3_11



(a) with type as metrics. (b) without type as metrics.

(c) with type as groups. (d) without type as groups.

Fig. 6. Visualization of Comparison Experimental Results. We visualize them separately by metrics and by group for the two evaluation approaches, all of the Y-axis represent the scores. We can see that $G_2(\alpha = 0.2, \beta = 0.8)$ performs the best in terms of results.

approaches. Analyzing Figure 6(a) and 6(b), we notice that the model performs better when the weight ratio of the CRF loss, α, is either larger or smaller (e.g., 0.2 or 0.7), but its performance is relatively poor when the ratio is close to half (e.g., 0.4 or 0.6). These findings validate our choice of setting α and β as 0.2 and 0.8 in the experiment.

6 Conclusion

We create an English-Uyghur cross-lingual dataset for temporal expression extraction tasks in Uyghur. By carefully considering the unique characteristics of Uyghur, we propose XLTU and utilize multi-task learning to help transfer the knowledge from English to Uyghur in TEE. We compare XLTU with different

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63751-3_11

models, and the results demonstrate that our model XLTU achieves the SOTA results on various evaluation metrics.

In our future work, we will seek an effective method for data augmentation to expand our high-quality dataset. And we will also try to apply it to Uyghur social platforms for public opinion analysis or others.

7 Acknowledgement

This work is supported by Xinjiang Uygur Autonomous Region Key Research and Development Program (No.2022B03010).

References

1. Naushad UzZaman, Hector Llorens, Leon Derczynski, James Allen, Marc Verhagen, and James Pustejovsky. 2013. Semeval-2013 task 1: Tempeval-3: Evaluating time expressions, events, and temporal relations. In Second Joint Conference on Lexical and Computational Semantics, Volume 2: Proceedings of SemEval 2013, pages 1-9.
2. Eunsol Choi, He He, Mohit Iyyer, Mark Yatskar, Wentau Yih, Yejin Choi, Percy Liang, and Luke Zettlemoyer. 2018. Quac: Question answering in context. In Proceedings of EMNLP 2021.
3. Bhaskar Mitra, Nick Craswell, et al. 2018. An introduction to neural information retrieval. Now Foundations and Trends.
4. Amir Feder, Katherine A. Keith, Emaad Manzoor, Reid Pryzant, Dhanya Sridhar, Zach Wood-Doughty, Jacob Eisenstein, Justin Grimmer, Roi Reichart, Margaret E. Roberts, Brandon M. Stewart, Victor Veitch, and Diyi Yang. 2021. Causal inference in natural language processing: Estimation, prediction, interpretation and beyond.
5. Jannik Strötgen and Michael Gertz. 2013. Multilingual and cross-domain temporal tagging. Language Resources and Evaluation, 47(2):269–298.
6. Xiaoshi Zhong, Aixin Sun, and Erik Cambria. 2017. Time expression analysis and recognition using syntactic token types and general heuristic rules. In Proceedings of ACL 2017, pages 420–429.
7. Sanxing Chen, Guoxin Wang, and Börje Karlsson. 2019. Exploring word representations on time expression recognition. Technical report, Tech. rep., Microsoft Research Asia.
8. Lukas Lange, Anastasiia Iurshina, Heike Adel, and Jannik Strötgen. 2020. Adversarial alignment of multilingual models for extracting temporal expressions from text. In Proceedings of Workshop on Representation Learning for NLP at ACL 2020, pages 103-109.
9. Yuwei Cao, William Groves, Tanay Kumar Saha, Joel R. Tetreault, Alex Jaimes, Hao Peng, Philip S. Yu. 2022. XLTime: A Cross-Lingual Knowledge Transfer Framework for Temporal Expression Extraction. In Findings of NAACL 2022.
10. Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of NAACL-HLT 2019, pages 4171–4186.
11. Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. In Proceedings of ACL 2020, pages 8440–8451.

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63751-3_11

12. Xiaodong Liu, Pengcheng He, Weizhu Chen, and Jianfeng Gao. 2019a. Multi-task deep neural networks for natural language understanding. In Proceedings of ACL 2019, pages 4487–4496.
13. Jannik Strötgen and Michael Gertz. 2015. A baseline temporal tagger for all languages. In Proceedings of EMNLP 2015, pages 541–547.
14. Angel X Chang and Christopher D Manning. 2012. Sutime: A library for recognizing and normalizing time expressions. In Lrec, volume 3735, page 3740.
15. Wentao Ding, Guanji Gao, Linfeng Shi, and Yuzhong Qu. 2019. A pattern-based approach to recognizing time expressions. In Proceedings of AAAI 2019, volume 33, pages 6335–6342.
16. Egoitz Laparra, Dongfang Xu, and Steven Bethard. 2018. From characters to time intervals: New paradigms for evaluation and neural parsing of time normalizations. Transactions of the Association for Computational Linguistics, 6:343–356.
17. Alankar Jain, Bhargavi Paranjape, and Zachary C. Lipton. 2019. Entity projection via machine translation for cross-lingual NER. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pages 1083–1092, Hong Kong, China. Association for Computational Linguistics.
18. Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, and Chris Dyer. 2016. Neural architectures for named entity recognition. In Proceedings of NAACL-HLT 2016, pages 260270.
19. Lance A Ramshaw and Mitchell P Marcus. 1999. Text chunking using transformation-based learning. In Natural language processing using very large corpora, pages 157–176. Springer.
20. Deblina Bhattacharjee, Tong Zhang, Sabine Susstrunk and Mathieu Salzmann. 2022. MulT: An End-to-End Multitask Learning Transformer. CVPR2022.
21. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin. 2017. Attention Is All You Need. NeurIPS.
22. Yangyang Xu, Yibo Yang, Lefei Zhang. 2023. Deformable Mixer Transformer for Multi-Task Learning of Dense Prediction. AAAI2023.
23. James Pustejovsky, Kiyong Lee, Harry Bunt, and Laurent Romary. 2010. Isotimeml: An international standard for semantic annotation. In LREC, volume 10, pages 394–397.
24. Pawel Mazur and Robert Dale. 2010. Wikiwars: A new corpus for research on temporal expressions. In Proceedings of EMNLP 2010, pages 913–922.
25. Yoon Kim. 2014. Convolutional Neural Networks for Sentence Classification. EMNLP.
26. Yann LeCun, Leon Bottou, Yoshua Bengio, Patrick Haffner. 1998. Gradient-Based Learning Applied to Document Recognition. In Proceeding of the IEEE 1998.
27. X Lu, B Ni. 2019. BERT-CNN: a Hierarchical Patent Classifier Based on a Pre-Trained Language Model. DOI:10.48550/arXiv.1911.06241.
28. Lafferty J , Mccallum A , Pereira F C N . 2002. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In proceedings of ICML .DOI:10.1109/ICIP.2012.6466940.
29. Alexis Conneau, Guillaume Lample. 2019. Cross-lingual Language Model Pretraining. 33rd Conference on Neural Information Processing Systems of NeurIPS 2019.
30. Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, Jennifer Wortman Vaughan. 2010. A theory of learning from different domains. Machine Learning 2010, 79(1-2):151-175.DOI:10.1007/s10994-009-5152-4.

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63751-3_11