ICIC Express Letters
Part B: Applications
Volume **8**, Number **4**, April **2017**

ICIC International ©2017 ISSN 2185-2766
pp. **703–709**

RESEARCH ON TEMPORAL EXPRESSION RECOGNITION: A CASE STUDY OF UYGHUR

Alim Murat[1,2,3], Xiao Li[1,2], Tonghai Jiang[1,2], Yating Yang[1,2,*]
Xi Zhou[1,2] and Lei Wang[1,2]

[1]Xinjiang Technical Institute of Physics & Chemistry, Chinese Academy of Science
No. 40-1, Beijing South Road, Urumqi 830011, P. R. China
{ alim.murat; xiaoli }@ms.xjb.ac.cn; *Corresponding author: yangyt@ms.xjb.ac.cn

[2]Xinjiang Key Laboratory of Minority Speech and Language Information Processing
Urumqi 830011, P. R. China

[3]University of Chinese Academy of Science
No. 19A, Yuquan Road, Shijingshan District, Beijing 100049, P. R. China

Received October 2016; accepted January 2017

Abstract. *Temporal Expression Recognition (TER) was proven to help natural language tasks like information extraction and question answering to obtain a higher performance. In this paper, we present a method for extraction of temporal expression from Uyghur text based on machine learning technique – conditional random field (CRF). A CRF classifier is trained with human annotated data, and different feature sets are combined and used in order to achieve the best performance on this task. Experimental results show that the best performing model gave an F1-Measure of 82.71 when all features are fully integrated. We also show the effect of word stemming on the performance. After word stemming, the results obtained improve to F1-Measure of 83.19.*
Keywords: Temporal expression recognition, Uyghur, Conditional random field, Word stemming

1. **Introduction.** Extraction of temporal expressions (timexes) from an input text is considered a very important step in several natural language processing tasks, namely, information extraction (IE), and summarization [1]. In question answering (QA), temporal information always answers the "when" kind of questions and is considered as a fundamental work to this task. Many work has been done and achieved the desired temporal annotation in English, Italian, Spanish, German and Chinese [1].

Particularly, for English, the current state-of-the-art temporal expression extraction achieves around 90% F1-Measure for identifying the timexes, and around 81% F1-Measure for normalizing the value of timexes. Unfortunately, there is a dearth of such approaches and systems, which tags documents in TimeML standard, for Uyghur language. To our knowledge, so far there has been no research pointed out to adopt the CRF model for TER in Uyghur texts. Therefore, the main idea of this work in this paper is to carry out experiments to analyze the performance of CRF model for the Uyghur TER task. We take into consideration the peculiarities of Uyghur and compare the results with different feature set combinations. However, the contributions of this work aim to build a basis for contemporary research in areas of question answering, event relation and chronology development in Uyghur.

In this work, the development of a CRF based statistical approach for recognition of timexes in Uyghur is proposed. The proposed approach exploits CRF classifier using five different feature sets sequentially, in order to achieve the best recognition performance on the human annotated data set.

A good amount of successful research has been done in temporal expression annotation in the previous literature. Strötgen and Gertz [2] focus specifically on a multilingual approach to temporal annotation and explain a simple technique to parallel the English temporal tagging program for other languages. In their work, the developed time tagger system is shown to have an F-Measure of 0.962 with identification and 0.832 when considering values of identified expressions.

Another system, MayoTime [3] was developed and adapted from publicly available Temporal Tagger, Heideltime. This system applies the TIMEX2 tagging to Italian and English texts and was shown to have an F-Measure of 0.926 with identification and 0.872 when considering values of identified expressions. Another system which generates TIMEX2 annotations in English text is known as the SUTime [4] which is a temporal tagger for recognizing and normalizing temporal, and it is available as a Java and is a part of the Standford CoreNLP pipline. SUTime is a deterministic rule-based system designed for extensibility.

HeidelTime [5], developed at University of Heidelberg works using a rule base and employs regular expression matching for extraction and normalization of temporal expressions. HeidelTime is shown to have an F1-Measure of 0.86 in temporal expression identification.

The above mentioned rule based systems use a large set of hand crafted rules. To minimize this dependence on a rule set, a machine learning approach was proposed by Ahn et al. [6]. This new architecture replaced the rule base by a set of machine learned classifiers to achieve the desired temporal expression annotation in English. Their system, TimexTag is shown to have an F1-Measure of greater than 0.8. ATT system developed by [7] used big windows and rich syntactic and semantic features for the TempEval time expression and event segmentation and classification tasks. It uses wide variety of features, like lexical, part of speech, dependency and constituency parse and semantic roles. ATT system achieved an F1-Measure of 0.85 in SemEval 2013 sub-task of TER.

The major contribution of this work is the Uyghur TER system that has been developed for extraction and annotation of timexes in Uyghur text. Uyghur TER can be very useful in many ways namely serving as a basic component for question answering system, helping the extraction of important features of temporal tagged text.

We begin by describing temporal expression linguistically in Section 2. A machine-learning approach based on feature learning from a human-annotated text is presented and explained in detail in Section 3. Evaluation results and analysis of our approach that learns the various features is reported in Section 4. Finally, conclusion and suggestions for further studies are presented in Section 5.

2. **Temporal Expression in Uyghur.**

From a general viewpoint, the Uyghur TER task aims at carrying out the following two basic goals.

The detection of the existing timexes in given Uyghur raw text: to determine a boundary and extent of text fragments, which are composed of one or more word units, indicating a proper timex in the given Uyghur text. So given a document D, words w in D, w must be ascertained whether or not inside of a timexes.

Classifying the detected timexes: to classify the recognized Uyghur timexes as one of the temporal expression classes, which is presented in the TimeML annotation standard and briefly shown in Table 1. In certain document D, there should be a mapping named I: $t \to \chi$, set t as the detected timexes in D, in which $x \in X$.

Uyghur is a typical inflectional language; therefore, an Uyghur word can be formed as the following components:

$$Word = prefix(ex) + lemma + suffix(es)$$

Table 1. Temporal expression classes in Uyghur

Class	Example	Example (Eng.)
DATE	2016-يىلى 3-ئاينىڭ 23 كۈنى؛ جۈمە	March 23, 2016; Friday
TIME	ئۈچكە ئون مىنۇت؛ سائەت بەشتە	Ten minutes to three; At five
PERIOD	2 ئاي؛ 48 سائەت	2 months; 48 hours
FREQUENCY	ھەپتىدە ئىككى قېتىم؛ يىلدا بىر	Twice a week; once a year

Generally, the prefixes are articles, prepositions or conjunctions, while the suffixes can be objects or personal and possessive anaphora [8]. From the linguistics perspective, Uyghur's inflectional characteristics make Uyghur text, compared to other languages which are of fewer morphology variations, sparser and hence most NLP tasks in Uyghur are more challenging. However, concerning the recognition of timexes in Uyghur, it is noted that the identification of Uygur timexes depends most on the word and phrases that are seen in the text. However, both words and phrases can occur in different forms, and as a result, a large amount of training corpus is so significant for high accuracy performance in this task. According to the statistics presented in [9], there are word-stems about 40000 and 289 word-forming affixes, which can basically build approximately 120 million words in the form of stem + suffixes.

In order to reduce data sparseness in Uyghur texts, following two solutions are needed.

1) **Word stemming:** referred to as a process ignoring all affixes added to a lemma to express the exact denotation. This solution is useful for many applications like information retrieval and question answering, due to the fact that those articles, prepositions, and conjunctions are considered as stop words. The realization of this solution was presented in [10].
2) **Word segmentation:** involves the separation between different granularities of a word based on the space character. As a result, this solution is more suitable for NLP tasks that need to have the various word morphemes, like word sense disambiguation and named entity recognition (NER). As for Uyghur segmentation, the tool has been developed varying in its application. In this paper, we thus use our word splitter.

In our evaluation, we have employed the first resolution to reduce data sparseness and showed the results obtained, respectively, by before and after the word stemming. The results will be drawn in the experiment section.

3. Task Definition and Approach.

3.1. **Uyghur TER as sequence labeling task.** A temporal expression like some date and time expressions have been deemed to be named entities and have been listed in the scope of NER systems. Generally, NER can be represented as a supervised tagging problem [11]; for this reason, we cast Uyghur timexes recognition as a sequence labeling task. We assumed that an input sequence of token $T_1^n = t_1 t_2 t_3 \cdots t_n$, the Uyghur TER problem is to create a label sequence $L_1^n = l_1 l_2 l_3 \cdots l_n$, in which l_i either belongs to the set of predefined Uyghur timex classes or is not temporal expression. The general label sequence L_1^n shows the highest probability of occurring for the token sequence T_1^n between all potential label sequences. This can be written as:

$$\hat{L}_1^n = \arg\max\{Pr(L_1^n | T_1^n)\}$$

By virtue of chunking method, we use IOB2[1] labeling scheme [12] to tag our corpus. In this scheme, each sentence contains a word at the beginning followed by its IOB label.

[1]IOB2 representations: the beginning of a TE (B), inside of a TE (I), outside of a TE (O) and sometimes the E is used with the last.

The label encodes the Uyghur timexes and discriminates whether the current token is inside or outside of temporal expression. We illustrate this for the sentence "*He will come back home on October 15.*" which has a timexes in the following.

Table 2. IOB tagging scheme in Uyghur

Tokens	IOB tags	Tokens (Eng.)
ئۇ	O	He
ئۆيگه	O	Home
10	B-timex	
-	I-timex	October
ئاينىڭ	I-timex	
15	I-timex	
-	I-timex	15
كۈنى	I-timex	
قايتىپ	O	Back
كىلىدۇ	O	Will come
.	O	.

3.2. **CRF based approach.** In this task, Uyghur TER is used to train model applying well-known machine learning technique: conditional random fields (CRF). This technique is popular among the related work and shows good performance for the proposed task.

CRF [11] is a probabilistic framework for the undirected graphical model to segment and tag sequence data. In this model, we assume that X is a random variable over the token sequence to be tagged, and Y is a random variable over the corresponding label sequence. CRF model intends to find the label Y which maximizes the conditional probability $P(Y|X)$ for a token sequence x. The CRF model is feature-specific model in which features gain binary values; the feature functions would be expressed in the following:

$$f_k(y,x) = \begin{cases} 1, & \text{if } x = \text{دۈشەنبە}\,(Monday) \text{ and } y = B-Timex \\ 0, & \text{else} \end{cases}$$

The CRF model can be seen as a generalization of Maximum Entropy and Hidden Markov Model that defines a conditional probability distribution taking the following form:

$$p(y|x) = \frac{1}{Z(x)} \exp\left(\sum_{k=1}^{K} \lambda_k \cdot f_k(y,x)\right)$$

where K is the number of features and λ_i indicate the weights given to the feature in the training, and $Z(x)$ is a normalization factor taking the following form:

$$Z(x) = \sum_{y \in Y} \exp\left(\sum_{k=1}^{K} \lambda_k \cdot f_k(y,x)\right)$$

Features are used to represent those human annotated data by a form of the vector in CRF model. In this experiment, the standard CRF's classifier requires the quality of features, which can directly impact on task performance. Hence, simple lexical, character features, which can easily be derived from the surface forms words, are used in this paper. We also used a context window of two words to the left and right. A set of feature templates, which are selected in Uyghur TER task is shown in the following table.

Each feature is capable of denoting some information regarding Uyghur temporal expression. The word feature helps in using the context information while inferring the tag of token or word. The POS tag feature implies whether the current word is a temporal

Table 3. List of features used in the experiment

Type	Features
Word (WF)	U01:%x[-1,0]; U05:%x[-1,0]/%x[0,0]
POS	U13:%x[1,1]; U21:%x[-1,1]/%x[0,1]/%x[1,1]
Number (NF)	U32:%x[0,2]; U35:%x[-2,2]/%x[-1,2]
Character (CF)	U47:%x[0,3]; U49:%x[2,3]
Dictionary (DF)	U62:%x[0,4]; U70:%x[-2,4]/%x[-1,4]/%x[0,4]

noun. Once this is a temporal noun, it is very likely to be a part of the temporal expression. Additionally, number and character features are very useful to discriminate a timexes composing a digit or special symbolic character. Dictionary feature is used for detecting unique and language-specific time nouns in existing corpus.

4. **Experiments and Results.**

4.1. **Corpus.**

In Uyghur, we have no standard datasets that enable our results to be compared with state-of-the-art experimental results. However, we used the human-annotated dataset which consists of 100 news article from the semi-annual daily half-hour broadcast of "CCTV News" and "Xinjiang News" in Uyghur. The preprocessing of dataset includes cleaning HTML tags and converting the text into TimeML document format. As shown in Table 4, during the training section only 75 news articles are used to learn our CRF model. The rest of the 25 articles are used to test the performance.

Table 4. Corpus used in the experiment

Corpus	#of docs	#of tokens	#timexes
Training	75	25539	625
Evaluation	25	9549	233

4.2. **Evaluation metrics.**

The performance of the proposed Uyghur TER task is evaluated based on the criteria used in the ACE TERN-2004 evaluation. Two standard measures, Precision (P) and Recall (R) are used for evaluation in this work, where Precision is the measure of the number of Uyghur TEs correctly identified over the number of TEs identified and Recall is the measure of a number of Uyghur TEs correctly identified over an actual number of Uyghur TEs. F-Measure which is the harmonic mean of Precision and Recall

$$F = \frac{(\beta^2 + 1)PR}{\beta^2(P + R)}$$

When $\beta = 1$, F-Measure is called F1-Measure or simply F1-Score.

4.3. **Results and analysis.**

We have used the corpus (mentioned in the previous section) to evaluate the proposed approach in Uyghur TER task based on CRF++[2] implementation and explore the impact of feature selecting on performance. In the CRF model, proper feature selection can lead the performance to the optimal level. However, this possible optimal result usually involves all features to be traversed according to the number of its factorial combination. The overall recognition result has been presented by using the official TERN scorer. Table 5 highlights the obtained experimental results from the discussed approach.

As can be seen from Table 5, the difference in performance between the baseline (that only uses the word token) and final feature combinations is 12.88% absolute in Precision, 31.36% absolute in Recall and 22.64% absolute in F1-Score. This clearly indicates that

[2]http://crfpp.sourceforge.net/

TABLE 5. Uyghur TER performance at different feature combinations

Features	Precision	Recall	F1-Measure
WF	76.15	51.08	60.07
WF+POS	77.82	63.75	68.4
WF+POS+NF	80.33	65.95	70.32
WF+POS+NF+CF	83.02	74.55	75.03
WF+POS+NF+CF+DF	89.03	82.44	82.71

the more features there are, the better the performance is and proves the significance of feature engineering in this task. Although each feature merely marginally contributes to performance, it surprisingly becomes significant when all feature contributions are combined. However, the result gained by the performance of complete feature combinations shows that Uyghur TER has seen the very optimal score based on all feature sets.

The impact of word stemming. In a previous literature review [13], we have found that the error rate caused by the complex morphology of the typical inflectional language was taken into consideration. This type of error is mainly due to the data sparseness in agglutinative language morphology. In this paper, we have carried out experiments before and after stemming the word in the dataset. In Table 6 we present the results obtained with the original dataset and the results obtained after the word stemming, based on the same optimal feature-set acquired in the above experiment.

TABLE 6. Uyghur TER performance before and after the word stemming

	Precision	Recall	F1-Measure
Before stemmed (BS)	89.03	82.44	82.71
After stemmed (AS)	90.12	83.35	83.19

As is shown in Table 6, a clear comparison between BS and AS indicates that the result obtained by AS is relatively better than BS. The main reason is that it can offset the problem of data sparsity induced by the inflection of word in morphologically rich languages, like Uyghur, Kazak, and Kirghiz.

5. **Conclusion.** We have introduced a language independent statistical approach for extraction of temporal expressions occurring in Uyghur natural language texts. The proposed task was evaluated on the human annotated corpus which is a dataset as a golden standard. In the experiment, we used CRF classifier to recognize Uyghur timexes and also verified the optimal feature-sets for better performance. However, this optimal feature combination helped train classifiers for automatically recognizing Uyghur timexes, with reasonable success (82.71% F1-Measure). We also investigated the impact of word stemming on the performance. By word stemming, the results obtained improved to 83.19% F1-Measure.

We plan to expand the more temporally annotated corpus for Uyghur and also perform detailed error analysis, based on the use of additional features (e.g., from WordNet and Syntactic trees). We will employ feature selection in a principal manner for further improvement on performance. Further, we will consider the automatic extraction of temporal resolutions rules, using the state-of-the-art timexes tagger [5].

Acknowledgment. This work in this paper is supported by the Young Creative Sci-Tech Talents Cultivation Project of Xinjiang Uyghur Autonomous Region (2014711006, 2014721 032), the Natural Science Foundation of Xinjiang (2015211B034), the Xinjiang Key Laboratory Fund under Grant No. 2015KL031 and the Strategic Priority Research Program of the Chinese Academy of Science under Grant No. XDA06030400, the West

Light Foundation of the Chinese Academy of Sciences under Grant (YBXM-2014-04, 2015-XBQN-B-10). We sincerely thank the anonymous reviewers for their thorough reviewing and valuable suggestions.

REFERENCES

[1] R. Campos, G. Dias, A. M. Jorge and A. Jatowt, Survey of temporal information retrieval and related applications, *ACM Computing Surveys*, vol.47, no.2, 2015.

[2] J. Strötgen and M. Gertz, Multilingual and cross-domain temporal tagging, *Language Resources and Evaluation*, vol.47, no.2, pp.269-298, 2013.

[3] S. Sohn, K. B. Wagholikar, D. Li, S. R. Jonnalagadda, C. Tao, R. K. Elayavilli and H. Liu, Comprehensive temporal information detection from clinical text: Medical events, time, and TLINK identification, *Journal of the American Medical Informatics Association*, vol.20, no.5, pp.836-842, 2013.

[4] A. X. Chang and C. D. Manning, SUTime: A library for recognizing and normalizing time expressions, *LREC*, pp.3735-3740, 2012.

[5] J. Strötgen and M. Gertz, Heideltime: High quality rule-based extraction and normalization of temporal expressions, *Proc. of the 5th International Workshop on Semantic Evaluation*, pp.321-324, 2010.

[6] D. Ahn, J. Rantwijk and M. Rijke, A cascaded machine learning approach to interpreting temporal expressions, *Proc. of Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT)*, 2007.

[7] H. Jung and A. Stent, ATT1: Temporal annotation using big windows and rich syntactic and semantic features, *The 2nd Joint Conference on Lexical and Computational Semantics*, vol.2, pp.20-24, 2013.

[8] H. Tömür, *Modern Uyghur Grammer*, The Minority Press, 1987.

[9] H. J. Xue, X. H. Dong, L. Wang, O. Turghun and T. H. Jiang, Unsupervised Uyghur word segmentation method based on affix corpus, *Computer Engineering and Design*, vol.32, no.9, pp.3191-3194, 2011.

[10] A. M. Azragul and Y. Abaydula, Based on the morphological analysis of the modern Uyghur noun stems recognition research, *Journal of Chinese Information Processing*, vol.29, no.6, pp.208-212, 2015.

[11] J. Lafferty, A. McCallum and F. Pereira, Conditional random fields: Probabilistic models for segmenting and labeling sequence data, *Proc. of the 18th International Conference on Machine Learning*, vol.1, pp.282-289, 2001.

[12] S. Maiti, U. Garain, A. Dhar and S. De, A novel method for performance evaluation of text chunking, *Language Resources and Evaluation*, vol.49, no.1, pp.215-226, 2015.

[13] Y. Benajiba and P. Rosso, Arabic named entity recognition using conditional random fields, *Proc. of Workshop on HLT & NLP within the Arabic World, LREC*, vol.8, pp.143-153, 2008.