Journal of Computer and Communications, 2021, 9, 143-152
https://www.scirp.org/journal/jcc
ISSN Online: 2327-5227
ISSN Print: 2327-5219

Teaching Machines to Read and Comprehend Tibetan Text

Yuan Sun[1,2]*, Sisi Liu[1,2], Chaofan Chen[1,2], Zhengcuo Dan[1,2], Xiaobing Zhao[1,2]

[1]School of Information Engineering, Minzu University of China, Beijing, China
[2]Minority Languages Branch, National Language Resource and Monitoring Research Center, Beijing, China
Email: *tracy.yuan.sun@gmail.com

How to cite this paper: Sun, Y., Liu, S.S., Chen, C.F., Dan, Z.C. and Zhao, X.B. (2021) Teaching Machines to Read and Comprehend Tibetan Text. *Journal of Computer and Communications*, **9**, 143-152.
https://doi.org/10.4236/jcc.2021.99011

Received: July 1, 2021
Accepted: September 27, 2021
Published: September 30, 2021

Abstract

Teaching machine to comprehend a passage and answer corresponding questions, the machine reading comprehension (MRC) has attracted much attention in current years. However, most models are designed to finish English or Chinese MRC task, Considering lack of MRC dataset, the low-resource languages MRC tasks, such as Tibetan, it is hard to get high performance. To solve this problem, this paper constructs a span-style Tibetan MRC dataset named TibetanQA and proposes a hierarchical attention network model for Tibetan MRC task which includes word-level attention and re-read attention. And the experiments prove the effectiveness of our model.

Keywords

Machine Reading Comprehension, Hierarchical Attention, Dataset

1. Introduction

Machine reading comprehension (MRC) aims to teach machines to read and understand human language text. The task of machine reading comprehension asks the machine to read texts such as an article or a story, then answer some questions related to the text. The questions can be designed to query the aspects that human care about. Based on the answer form, MRC is simply categorized into four tasks: cloze tests, multiple choices, span extraction and free answering. In recent years, many Chinese and English machine reading comprehension datasets have emerged, such as: SQuAD [1], MCTest [2], MS-MARCO [3], Du-Reader Dataset [4] etc. Following these datasets, many models have been proposed, such as S-Net [5], AS Reader [6], IA Reader [7] etc. And they achieved great performance. However, for low-resource language machine reading comprehension such as Tibetan, it is rarely mentioned. The main reasons are follows: 1) Lacking large-scale open Tibetan MRC datasets, the relevant experiments cannot be carried out. This is also the main factor that hinders the development of Tibetan MRC. 2) Compared to English MRC, word segmentation tools for Tibetan are under developing. The wrong word segmentation results will lead to semantic ambiguity, which will be propagated to downstream tasks. 3) For low-resource MRC tasks, it is difficult to achieve good performance on small-scale dataset. Therefore, it needs the MRC model to strengthen its understanding.

To address these issues, this paper proposes an end-to-end model for Tibetan MRC. In order to reduce the error propagation caused by word segmentation, the model incorporates syllable-level information. In addition, to enhance the ability of model understanding, we adopt a hierarchical attention structure. In summary, our contributions are as follows:

- In order to solve the problem of lacking Tibetan MRC corpus, we construct a high-quality Tibetan MRC dataset named TibetanQA (The Tibetan Question Answering dataset), which covers multi-domain knowledge and is constructed by crowdsourcing.
- To solve the segmentation errors, we combine syllable and word embedding, so that the model can learn the more complex information in Tibetan.
- To reduce the impact of long text paragraph information that is irrelevant to the question, this paper uses a word-level attention mechanism to focus on the key words of the answer. To enhance the understanding ability of model, this paper adopts a hierarchical attention network, which includes word-level attention and re-read attention to provide clues to answer the question.

2. Related Work

Machine reading comprehension is an important step in natural language processing from perceptual text to understand text. In the early times, lacking large-scale datasets, most of MRC system are rule-based or statistical models. In the next decades, researchers begin to focus on MRC dataset construction. They treat machine reading comprehension as a problem with supervised learning and use manual annotation to construct question-answer pairs. Hermann *et al.* propose a blank-filling English machine reading comprehension dataset CNN & Daily mail [8]. Hill *et al.* release the Children's Book Test dataset [9], this dataset is only a simple shallow semantic understanding and do not involve deep reasoning. To settle this problem, Laid *et al.* publish the RACE dataset in 2017 [10]. This dataset pays more attention to reasoning ability. For span extraction MRC, Rajpurkar *et al.* collect a large-scale dataset named Stanford Question Answering Dataset (SQuAD) with highly quality.

Followed these large-scale datasets, some important research based on deep learning methods have broken out for MRC. The Match-LSTM model is proposed by Wang *et al.* [11]. They adopt Long Short-Term Memory (LSTM) [12] to encode the question and passage respectively, and then introduce the attention-based weighted representation of question in the LSTM unit. Subsequently, to capture long-term dependencies between words within a passage, the team of Microsoft proposed R-Net model [13]. Cui *et al.* propose the Attention-Over-Attention Reader model [14].

Different from the previous work, Seo *et al.* propose the BiDAF model [15] which adopts two directional attentions. Xiong *et al.* propose a DCN model [16] that uses an interactive attention mechanism to capture the interaction between a problem and a paragraph.

The above models based on single-layer attention have the problem of weak semantic interaction ability between the capture problems and paragraphs due to the small number of attention layers and shallow network depth. To solve this problem, a series of recent works have enhanced the model by stacking several attention layers [17]. Huang *et al.* propose Fusion-Net [18]. The model uses a fully perceptual multilayer attention architecture to obtain the complete information in the problem and integrate it into the paragraph representation. Wang *et al.* [19] propose a multi-granular hierarchical attention fusion network to calculate the attention distribution at different granularities, and then perform hierarchical semantic fusion. Their experiments prove that multiple layers of attention interaction can achieve better performance. Tan *et al.* [20] propose an extraction-generative model. They use RNN and attention mechanism to construct question and context representations, then they use seq2seq to generate answers based on key information.

3. Dataset Construction

Considering lack of Tibetan machine reading comprehension dataset, this paper constructs a span-style Tibetan machine reading comprehension dataset named TibetanQA. This process is mainly divided into three stages: passages collection, questions collection, and answer verification.

3.1. Passage Collection

We obtain a large amount of text information from the Yunzang website. In order to improve the quality of the TibetanQA, the articles cover a wide range of topics, including nature, culture, education, geography, history, life, society, art, person, science, sports and technology. In addition, we have deleted noise information in articles, such as images, tables, and website links, and discarded article shorter than 100 characters, finally, 763 articles are selected to the dataset.

3.2. Question Construction

In order to collect questions effectively, we develop a QA collection web application, the students whose native language is Tibetan are invited to use this application. For each passage in the article, they first need to select a segment of text or a span in the article as the answer, and then write the question in their own language into the input field. Students are tasked with asking and answering up to 5 questions on contents of one article. The answer must be part of the paragraph. When they finish an article, the system will automatically assign the next article to them. To construct a more challenging corpus, we conduct a short-term training to guide them how to provide effective and challenging questions. For each student, we will first teach them how to ask and answer questions, and then use a small amount of data to test them, only students with an accuracy rate of 90% can do the following work. We don't impose restrictions on the form of questions and encourage them to ask questions in their own language.

3.3. Answer Verification

In order to further improve the quality of the dataset, we invite another group of Tibetan students to check the dataset after obtaining the initial dataset. They select the valid QA pairs, discard the incomplete answers or questions, and strip away the question with incorrect grammar. In the end, we construct 10,881 question and answer pairs. To better train our model, we organize TibetanQA into json format, and add a unique ID to each question answer pair (see **Table 1**). Finally, these question answer pairs are partitioned at random into a training set, development set and test set (see **Table 2**).

4. Model Details

In this section, we introduce our model in detail.

4.1. Data Preprocessing

Different from English, Tibetan is a Pinyin language. The smallest unit of word is syllable. And some syllables can indicate the some meaningful "case". The "case" in Tibetan refers to a type of function syllable that distinguishes between words and explains the role of the word in a phrase or sentence.

Table 1. An example in tibetan MRC corpus.

UUID	01d5c760-a4a3-13e5-e9b8-4a03057ce3d2
Passage	ཕྱུ་སྐྱེད་ཟས་སྣོད་མང་ཆེ་བ་ནི་བསྐྱར་སྐྱེས་ཀྱི་སྤོས་འཁྱིག་གི་རྒྱུ་ཆ་ཡིན་པ་དང་། དེ་ཡི་ནང་དུ་རྒྱུ་ཆ་གསར་པ་ཉུང་ཤས་འདུས་ཡོད་ཅིང་། ཡང་ཏ་ཤིག་ཕྱེ་མ་ཚད་ངེས་ཅན་ཞིག་བསྡེབས་ནས་རྒྱུན་པར་མི་རྣམས་ཀྱིས་བཀོལ་སྤྱོད་བྱེད་བཞིན་པའི་ཕྱུ་སྐྱེད་ཟས་སྣོད་དེ་གྲུབ་པ་རེད།… The materials of foam plastics are mostly recycled plastics. It usually contains a small amount of new materials and a certain amount of talc powder. The foam plastics are widely used as tableware by people…
Question	ཕྱུ་སྐྱེད་ཟས་སྣོད་མང་ཆེ་བ་ནི་གང་གི་རྒྱུ་ཆ་ཡིན། What are the materials of most foam plastics?
Answer	བསྐྱར་སྐྱེས་ཀྱི་སྤོས་འཁྱིག་ Recycled plastics

Table 2. Tibetan MRC dataset statistics.

	Train Set	Dev Set	Test Set	Total
paragraphs	4229	332	478	5039
Questions	7113	500	600	8213

In fact, there are many syllables in Tibetan can provide some key information for MRC task as the "case" do. Therefore, it is necessary to embed the syllable information in the encoding layer. On the other hand, the embedding of syllables can reduce the semantic ambiguity caused by incorrect word segmentation. Based on the above considerations, this paper combines syllables and words information. Next, we will introduce word-level and syllable-level Tibetan text pre-processing in our experiments.

- Syllable-level preprocessing: It is easy to split the syllables, because there is delimiter between the syllables. With the help of delimiter-".", we can separate the syllables.
- Word-level preprocessing: Each word is composed of difference syllables, which is difficult to spilt word in sentences. For word-level segmentation, we use Tibetan word segmentation tools [21].

Finally, the specific format is as shown in **Table 3**.

4.2. Input Embedding Layer

With strong grammatical rules, Tibetan is made up of syllables, and syllables are the smallest unit of Tibetan. It is noteworthy that some syllables can contain information, such as reference, subordination, gender, etc. This information will help to predict the correct answer. Therefore, at the input encoding layer, we also embed the syllables into word represent, which can extract more information for the network.

Suppose there is a question and a passage, And they can be present as: $Q=\{q_1,q_2,q_3,\cdots,q_n\}$ and $P=\{p_1,p_2,p_3,\cdots,p_m\}$, we turn them into syllable-level embedding ($\{s_1^q,s_2^q,s_3^q,\cdots,s_n^q\}$ and $\{s_1^p,s_2^p,s_3^p,\cdots,s_m^p\}$) and word-level embedding ($\{w_1^q,w_2^q,w_3^q,\cdots,w_n^q\}$ and $\{w_1^p,w_2^p,w_3^p,\cdots,w_m^p\}$) respectively. We use a pre-trained model to encode question and passage, each word token is encoded into a 100-dimensional vector with fastext through a lookup manner. As for syllable-level encoding, we use a bi-direction long short-term memory neural network (BiLSTM) and use the final state as the syllable-level token. Finally, we fuse two vectors of different levels through a two-layer highway network, and the final passage and questions are finally coded as: $\{M_t^q\}_{t=1}^n$ and $\{M_t^p\}_{t=1}^m$.

4.3. Word-Level Attention

Just as people participate in a reading comprehension test, people will read the questions firstly, then start to briefly read the passage, mark the words relevant question, and pay more attention on the keywords. Finally, they will search for the correct answer. Inspired by this, we propose word-level attention. We

Table 3. Data preprocessing sample.

Original sentence	ཐུ་སྐྱིད་ཟས་སྡོད་མང་ཆེ་བ་ནི་བསྐྱར་སྐྱེས་ཀྱི་སྤོས་འཁྱིག…
Syllable-level	ཐུ /སྐྱིད /ཟས /སྡོད /མང /ཆེ /བ /ནི /བསྐྱར /སྐྱེས /ཀྱི /སྤོས /འཁྱིག…
Word-level	ཐུ་སྐྱིད་/ཟས་སྡོད་/མང་ཆེ་བ་/ ནི་/བསྐྱར་སྐྱེས་/ཀྱི་/ སྤོས/ འཁྱིག…

perform word-level attention to calculate the importance of each word in the passage to the question. Similarly, assuming that the passage word-level embedding is $\{M_t^p\}_{t=1}^m$ and the question word-level embedding is $\{M_t^q\}_{t=1}^n$. The attention vector of each word in the passage is calculated by the Equation (1).

$$S_u = V^T \tan(W_u^Q M_i^q + W_u^P M_j^p) \quad (1)$$

where W_u^Q and W_u^P is a trainable weight matrix, and S_u presents the similarity matrix. Next, we will normalize S_u, in which every row will be normalized by a softmax function, shown in the Equation (2).

$$a_u \propto \exp(S_u) \quad (2)$$

To determinate which words in passage are helpful to answer the question, the query-to-context word-level attention. To determinate which words in passage are helpful to answer the question, the query-to-context word-level attention A_i^p is shown as the Equation (3).

$$A_i^p = \sum a_u M_j^q \quad (3)$$

Finally, we will use Bi-LSTM to obtain the sentence-pair representation V_t^p. And the notation is shown as the Equation (4).

$$V_t^p = BiLSTM(V_{t-1}^p, [A_t^p, M_t^p]) \quad (4)$$

4.4. Re-Read Attention

The word-level attention layer is a shallow attention calculation. To enhance the attention, we adopt a high-level attention to consider which sentence contains the correct span of answer. Therefore, we introduce the "re-read attention". Re-read attention aims to calculate the attention between the passage and question on sentence level. Before we calculate the attention, we need to understand the question. Namely, for each token in question, we employ BiLSTM to generate a higher level of question embedding y_i^q. The notation is shown as the Equation (5).

$$y_i^q = BiLSTM(y_{i-1}^q, [s_i^q, w_i^q]) \quad (5)$$

where y_{i-1}^q presents the previous hidden vector, s_i^q is syllable-level after input embedding layer and w_i^q is the output of word-level attention layer.

Based on the understanding of the question, similarly, we perform the re-read attention, and the calculation equations are (6)-(8).

$$S_v = V^T \tan(W_v^Q y_i^q + W_v^P V_j^p) \quad (6)$$

$$a_v \propto \exp(S_v) \quad (7)$$

$$A_i^p = \sum a_v y_i^q \quad (8)$$

where S_v is the similarity matrix between passage and question semantic embedding, y_i^q is question embedding vector, V_j^p is the output of word-level attention layer.

Finally, we use BiLSTM to encode the output of re-read attention layer. The final vector is coded as, shown in the Equation (9).

$$K_t^p = BiLSTM(K_{t-1}^p, [A_t^p, y_t^q]) \tag{9}$$

4.5. Output Layer

The main goal of this layer is to predict the starting position of the answer. At this level we use a softmax layer to achieve. This layer will predict the probability of each position in the give passage to be the start or end of the answer. And it can be described as the Equations (10) and (11).

$$p^{start} = softmax(W_1 K^p) \tag{10}$$

$$p^{end} = softmax(W_2 K^p) \tag{11}$$

where W_1 and W_2 are training parameters, p^{start}, p^{end} are the start and end position of answer.

5. Experimental Result and Analysis

5.1. Dataset and Evaluation

We conduct some experiments on SQuAD, SQuAD (8K) and TibetanQA. **Table 4** shows the statics of datasets.

To evaluate the effect of the model, this paper uses two common evaluation methods EM and F1. EM is the percentage of the predicted answer in the dataset that is the same as the true answer. F1 is the average word coverage between the predicted answer and the true answer in the dataset.

5.2. Experiments on Different Models

Before the experiment, we would like to introduce our baseline models. They are foundation but have achieved great performance in English MRC task.

Considering that there are no syllables in English, we remove the syllable embedding in our model on SQuAD. Next, we conduct some experiments on SQuAD, SQuAD (8K) our datasets. All models use fasttext embedding and are implemented by us, and the results of experiment are as **Table 5**.

Table 4. Answer types with proportion statistics.

Dataset	Train		Test	
	paragraphs	question	paragraphs	questions
SQuAD	17,007	68,758	1889	18,841
SQuAD (8k)	1552	7027	216	1037
TibetanQA	4229	7113	478	600

SQuAD: The Stanford Question Answering Dataset (SQuAD) is a new challenge reading comprehension dataset. It was construct by crowdsourcing and published in 2015. SQuAD (8K): This is a dataset consisting of about 8000 question and answer pairs, which are randomly selected from the SQuAD dataset. TibetanQA: The dataset is constructed by us and it uses a manual construction method. We collected 5039 texts of knowledge entities in various fields on the Tibetan encyclopedia website and manually constructed 8213 question answer pairs.

Table 5. Experimental result of different models on three datasets.

No	Model	SQuAD		SQuAD (8k)		Our dataset	
		EM	F1	EM	F1	EM	F1
1	R-Net	70.3	78.3	61.2	69.3	51.6	60.1
2	BiDAF	67.8	76.6	60.8	68.5	52.1	62.2
3	QANet	72.6	80.4	**65.1**	74.2	53.5	61.5
4	Ti-Reader	**73.1**	**81.2**	64.9	**75.8**	53.8	63.1
5	Ti-Reader + syllable					**63.4**	**71.3**
6	Ti-Reader + syllable-WA					59.3	67.4
7	Ti-reader + syllable-RA					57.3	66.5

R-Net: This model was proposed by Microsoft Research Asia Team (Wang, Yang & Zhou, 2017). They pay more attention to the interaction between questions and passage through a gate attention-base network. BiDAF: The BiDAF model was proposed by Seo et al (Seo, Kembhavi, Farhadi & Hajishirzi, 2016). Different from R-Net, the Bi-DAF model adopted two directions interaction layer. They didn't use the self-matching as R-Net did but calculated two attentions about query-to-context and context-to-query. QANet: This model was proposed by Adam *et al.* [22]. They combine local convolution with global self-attention and achieved better performance on SQuAD dataset. What deserves to be mentioned the most is they improve their model by data augments. For a better comparison, we remove the data enhancement in the next experiments. Ti-Reader: This is our model, which including a hierarchical attention networks.

It can be found our model have a better performance on three difference datasets. For the SQuAD, our model achieves 73.1% and 81.2% on EM and F1. Compared with BiDAF, our model increases 4.6% on F1. For the SQuAD (8K), the EM reaches 64.9% and F1 reaches 75.8%. Compared with R-Net, our model increases 3.7% on EM and 6.5% on F1. Compared with BiDAF, our model increases 4.1%, 7.3% on EM and F1. Compare with QANet, our model shows an improvemesF1. Thus, we can see that our model performs better on the SQuAD (8K). For our dataset, we can find our model is superior to other models. The Ti-Reader achieves 53.8% on EM and 63.1% on F1. And when we include the syllable embedding, the difference is +9.6% on EM and +8.2% on F1.

Additionally, we explore the following two kinds of attention mechanisms: word-level attention and re-read attention. The experiment shows the performance of the model is decreased. The EM value is decreased by 3.1% and the F1 value is decreased by 3.9% when removing word-level attention. The result illustrates the word-level attention mechanism can dynamically assign the weight of each word, so that the model can focus on those valuable words and improve the performance of the model. The Re-read attention mechanism is an interaction between the passage and question. It can be found that the EM of the model has decreased by 5.1%, and F1 value has decreased by 4.8% when remove the re-read attention.

6. Conclusions

In this paper, we proposed the Ti-Reader model for Tibetan reading comprehension. The model uses hierarchical attention mechanism, including word-level attention and re-read attention. At the same time, we conduct some extra experiments, and prove their effectiveness. Compared with two classic English MRC models, BiDAF and R-Net, the experiments show that our model has more advantages for Tibetan MRC. However, there are still some wrong answers.

In the future, we will continue to improve the accuracy of the model's prediction answers and design lighter models.

Acknowledgements

This work is supported by National Nature Science Foundation (No. 61972436).

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Rajpurkar, P., Zhang, J., Lopyrev, K. and Liang, P. (2016) Squad: 100,000+ Questions for Machine Comprehension of Text. *Proc. EMNLP*, Austin, TX, 1-10. https://doi.org/10.18653/v1/D16-1264

[2] Richardson, J.C., Christopher, M. and Erin, R. (2013) Mctest: A Challenge Dataset for the Open-Domain Machine Comprehension of Text. *Proc. EMNLP*, Seattle, Washington, 193-203.

[3] Nguyen, T., Rosenberg, M., *et al.* (2017) Ms Marco: A Human Generated Machine Reading Comprehension Dataset. *Proc ICLR*, Toulon, France, Louisiana, USA, 1-10.

[4] Wei, H., Kai, L., Jing, L., *et al.* (2018) DuReader: A Chinese Machine Reading Comprehension Dataset from Real-World Applications. *Proc ACL*, Melbourne, Australia, 37-46.

[5] Tan, C.Q., Wei, F.R., Nan, Y., *et al.* (2018) S-Net: From Answer Extraction to Answer Synthesis for Machine Reading Comprehension. *Proc. AAAI*, 5940-5947.

[6] Kadlec, R., Schmid, M., *et al.* (2016) Text Understanding with the Attention Sum Reader Network. *Proc ACL*, Berlin, Germany, 908-918. https://doi.org/10.18653/v1/P16-1086

[7] Alessandro, S., Philip, B., Adam, T. and Yoshua, B. (2016) Iterative Alternating Neural Attention for Machine Reading. arXiv preprint arXiv:1606.02245

[8] Hermann, K.M., Kočiský, T., Grefenstette, E., Espeholt, L., Kay, W., Suleyman, M. and Blunsom, P. (2015) Teaching Machines to Read and Comprehend. *Advances in Neural Information Processing Systems*, 1693-1701.

[9] Hill, F., Bordes, A., Chopra, S. and Weston, J. (2015) The Goldilocks Principle: Reading Children's Books with Explicit Memory Representations. arXiv preprint arXiv:1511.02301

[10] Lai, G., Xie, Q., Liu, H., Yang, Y. and Hovy, E. (2017) Race: Large-Scale Reading Comprehension Dataset from Examinations. arXiv preprint arXiv:1704.04683 https://doi.org/10.18653/v1/D17-1082

[11] Wang, S. and Jiang, J. (2016) Machine Comprehension Using Match-Lstm and Answer Pointer. arXiv preprint arXiv:1608.07905

[12] Hochreiter, S. and Schmidhuber, J. (1997) Long Short-Term Memory. *Neural Computation*, **9**, 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

[13] Wang, W., Yang, N., Wei, F., Chang, B. and Zhou, M. (2017) Gated Self-Matching Networks for Reading Comprehension and Question Answering. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, Vol. 1: Long Papers, 189-198. https://doi.org/10.18653/v1/P17-1018

[14] Cui, Y., Chen, Z., Wei, S., Wang, S., Liu, T. and Hu, G. (2016) Attention-Over-Attention Neural Networks for Reading Comprehension. arXiv preprint arXiv:1607.04423 https://doi.org/10.18653/v1/P17-1055

[15] Seo, M., Kembhavi, A., Farhadi, A. and Hajishirzi, H. (2016) Bidirectional Attention Flow for Machine Comprehension. arXiv preprint arXiv:1611.01603

[16] Xiong, C., Zhong, V. and Socher, R. (2016) Dynamic Coattention Networks for Question Answering. arXiv preprint arXiv:1611.01604

[17] Yin, J., Zhao, W.X. and Li, X.M. (2017) Type-Aware Question Answering over Knowledge Base with Attention-Based Tree-Structured Neural Networks. *Journal of Computer Science and Technology*, **32**, 805-813. https://doi.org/10.1007/s11390-017-1761-8

[18] Huang, H.Y., Zhu, C., Shen, Y. and Chen, W. (2017) Fusionnet: Fusing via Fully-Aware Attention with Application to Machine Comprehension. arXiv preprint arXiv:1711.07341

[19] Wang, W., Yan, M. and Wu, C. (2018) Multi-Granularity Hierarchical Attention Fusion Networks for Reading Comprehension and Question Answering. arXiv preprint arXiv:1811.11934 https://doi.org/10.18653/v1/P18-1158

[20] Tan, C., Wei, F., Yang, N., Du, B., Lv, W. and Zhou, M. (2017) S-Net: From Answer Extraction to Answer Generation for Machine Reading Comprehension. arXiv preprint arXiv:1706.04815 https://doi.org/10.1007/978-3-319-99495-6_8

[21] Long, C.J., Liu, H.D., Nuo, M.H., Wu, J., Mohammad and Le, Q.V. (2018) Tibetan Pos Tagging Based on Syllable Tagging. *Chinese Info. Process*, **29**, 211-216. (In Chinese)

[22] Yu, A.W., Dohan, D., Luong, M.T., Zhao, R., Chen, K., Norouzi, M. and Le, Q.V. (2018) Qanet: Combining Local Convolution with Global Self-Attention for Reading Comprehension. arXiv preprint arXiv:1804.09541