



Multi-Task Learning Based Traditional Mongolian Words Recognition

Hongxi Wei, Hui Zhang, Jing Zhang, Kexin Liu

School of Computer Science, Inner Mongolia University, Hohhot, China

Provincial Key Laboratory of Mongolian Information Processing Technology, Hohhot, China

National and Local Joint Engineering Research Center of Mongolian Information Processing Technology, Hohhot, China



Abstract

In this paper, a multi-task learning framework has been proposed for solving and improving traditional Mongolian words recognition. To be specific, a sequence-to-sequence model with attention mechanism was utilized to accomplish the task of recognition. Therein, the attention mechanism is designed to fulfill the task of glyph segmentation during the process of recognition. Although the glyph segmentation is an implicit operation, the information of glyph segmentation can be integrated into the process of recognition. After that, the two tasks can be accomplished simultaneously under the framework of multi-task learning. By this way, adjacent image frames can be decoded into a glyph more precisely, which results in improving not only the performance of words recognition but also the accuracy of character segmentation. Experimental results demonstrate that the proposed multi-task learning based scheme outperforms the conventional glyph segmentation-based method and various segmentation-free (i.e. holistic recognition) methods.

Introduction

Optical character recognition (OCR) plays an important role in the procedure of digitizing printed books or documents. However, OCR is still a challenging task for traditional Mongolian. The corresponding letters within one traditional Mongolian word are conglutinated together in vertical direction. Letters have initial, medial or final visual forms according to their positions within a word. Traditional Mongolian has a very special writing system. Its writing order is from top to bottom and its column order is from left to right. The traditional Mongolian is also a kind of agglutinative language, so the amount of its vocabularies is about hundreds of thousands.

Aiming at the above characteristics, glyph segmentation based approaches^{[1][2][3]} were firstly proposed. However, for certain fonts, such as Hawang font, it is difficult to obtain individual glyphs accurately from Mongolian words using the segmentation algorithm. As for as the printed Mongolian words recognition, an end-to-end model based on bidirectional LSTM (Bi-LSTM) and connectionist temporal classification (CTC) has been proposed in^[5]. In this manner, the operation of glyph segmentation can be substituted by the frame segmentation. However, the size of frames is fixed, which leads to one frame may contain part of a glyph or multiple glyphs.

Inspired by multi-task learning (MTL), We proposed a sequence-to-sequence model with attention mechanism, which can fulfill the tasks of glyph segmentation and printed Mongolian words recognition simultaneously.

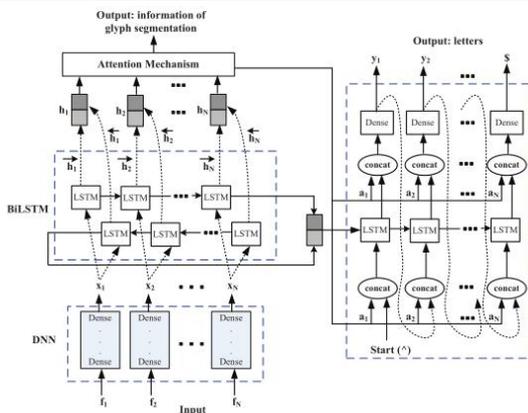


Figure 1. The model structure of the proposed MTL.

Methodology

For the task of the printed Mongolian words recognition, a conventional sequence-to-sequence model with attention mechanism has been employed. It consists of three components: an encoder, a decoder and attention mechanism between encoder and decoder. The attention mechanism can fulfill the task of glyph segmentation which can be regarded as an auxiliary task improving the performance of the printed Mongolian words recognition.

Experimental Results

The length of the frame is an important factor affecting the recognition accuracy. From Fig. 2, we can see that the accuracy is increasing when the length of frame is decreased. The best accuracy on the testing set can be attained to 99.87% and the appropriate length of frame is set to 2 pixels.

We can see that the proposed model of MTL is superior to not only LSTM-CTC but also Seq2seq on the four types of fonts in Table 1. By analyzing the experimental results, we can conclude that the performance has been improved through integrating the information of glyph segmentation into the attention mechanism and the implicit operation helps the main task to complete much better.

The performance of glyph segmentation has been compared with a conventional algorithm presented in^{[3], [4]}. The results listed in Table 2 show that the attention mechanism significantly outperforms the conventional algorithm on the four types of fonts. It concludes that the attention mechanism is greater than the conventional algorithm.

Table 1. The comparative results between the proposed mtl and baselines on the four fonts (accuracy).

Method	LSTM-CTC	Seq2seq	Our proposed
Bai Font	87.96%	99.08%	99.92%
Bao Font	86.56%	99.01%	99.84%
Hawang Font	86.12%	99.00%	99.76%
Biaoti Font	88.12%	99.03%	99.96%
Avg	87.19%	99.03%	99.87%

Table 2. The comparative results between the proposed mtl and the glyph segmentation algorithm on the four fonts (IoU).

Method	Glyph-segmentation	Our proposed
Bai Font	37.97%	78.30%
Bao Font	34.72%	82.98%
Hawang Font	28.73%	75.15%
Biaoti Font	40.65%	82.66%
Avg	35.52%	79.77%

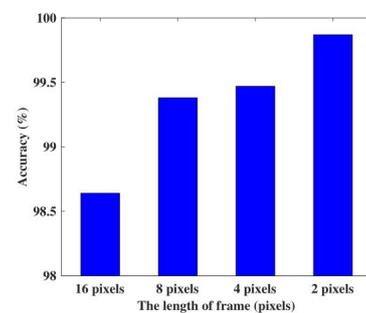


Figure 2. The comparative results of various lengths of frame.

Conclusions

- The performance of recognition of the proposed model of MTL outperforms LSTM-CTC and Seq2seq.
- The performance of glyph segmentation of the proposed model of MTL is also significantly superior to the conventional segmentation algorithm.
- By the above comparisons, we can draw a conclusion that the two tasks in multi-task framework can promote each other.

Contact

Jing Zhang
School of Computer Science, Inner Mongolia University, Hohhot, China
Email: cszjing@163.com
Phone: 18748100097

References

1. H. Wei and G. Gao, "Machine-printed traditional Mongolian characters recognition using BP neural networks," in Proceedings of 2009 International Conference on Computational Intelligence and Software Engineering (CISE'09), 2009, pp. 1-7.
2. H. Hu, H. Wei, and Z. Liu, "The CNN based machine-printed traditional Mongolian characters recognition," in Proceedings of the 36th Chinese Control Conference (CCC'17), 2017, pp. 3937-3941.
3. L. Peng, C. Liu, X. Ding, J. Jin and et al., "Multi-font printed Mongolian document recognition system," International Journal on Document Analysis and Recognition (IJ DAR), vol. 13, no. 2, pp. 93-106, 2010.
4. G. Gao, W. Li, H. Hou, and Z. Li, "Multi-agent based recognition system of printed Mongolian characters," in Proceedings of Active Media Technology, 2003, pp. 376-381.
5. W. Wang, H. Wei, and H. Zhang, "End-to-End Model Based on Bidirectional LSTM and CTC for Segmentation-free Traditional Mongolian Recognition," in Proceedings of the 38th Chinese Control Conference (CCC'19), 2019, pp. 8723-8727.