Multi-Task Learning Based Traditional Mongolian Words Recognition

Hongxi Wei, Hui Zhang, Jing Zhang
School of Computer Science
Inner Mongolia University
Hohhot, China
cswhx@imu.edu.cn
Outline

• Introduction
• The Proposed Method
• Experimental Results
• Conclusion
Outline

• Introduction
• The Proposed Method
• Experimental Results
• Conclusion
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 because of its special characteristics.
Introduction (Cont.)

- **Characteristics** of traditional Mongolian
  - The letters 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**.
  - The amount of its **vocabularies** is huge about **hundreds of thousands**.
Introduction (Cont.)

- **Characteristics** of traditional Mongolian
Introduction (Cont.)

- Printed Mongolian words recognition
  - Glyph segmentation
    - Problem: For certain fonts, such as Hawang font, it is difficult to obtain individual glyphs accurately from Mongolian words using the segmentation algorithm.
  - Frame segmentation.
    - Bi-LSTM + CTC\textsuperscript{[20]}
    - Problem: The size of frames is fixed, which leads to one frame may contain part of a glyph or multiple glyphs.
Introduction (Cont.)

• Printed Mongolian words recognition
  ■ Glyph segmentation
    ▶ Problem: For certain fonts, such as Hawang font, it is difficult to obtain individual glyphs accurately from Mongolian words using the segmentation algorithm.
  ■ Frame segmentation
    ▶ Inspired by muli-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.
Outline

• Introduction
• Related Work
• The Proposed Method
• Experimental Results
• Conclusion
Related Work

• MTL is considered as an important paradigm that aims at improving the generalization performance of a task using other related tasks [21].
  - **Soft parameter sharing**
    - Different task has its own model with its own parameters. The models are tied together either by information sharing or by requiring parameters to be similar.
    - Defect: not flexible enough or computationally expensive.
  - **Hard parameter sharing**
    - The hidden layers are shared by all tasks and several task-specific output layers are allowed to keep.
    - It greatly reduces the risk of over-fitting. It is easier to build and computationally less complex.
Related Work (Cont.)

- The main task is to recognize the printed Mongolian words using a sequence-to-sequence model with attention mechanism.
- Therein, the attention mechanism is designed to fulfill the task of glyph segmentation, which is considered as another task.
Outline

• Introduction
• The Proposed Method
• Experimental Results
• Conclusion
The Proposed Method

• The Model Structure of MTL
The Proposed Method

• The Model Structure of MTL
  - Encoder
    - **DNN**: To extract features from input sequences of word images.
    - **BiLSTM**: To generate a list of hidden vectors for encoding the input sequences.
  - Decoder
    - **LSTM + softmax**
  - Attention Mechanism
    - a feed forward neural network with a hidden layer
    - **softmax** layer
The Proposed Method (Cont.)

- **Input** of Model
  - step1: Each Mongolian word image should be scaled into the same size in advance.
  - step2: Each scaled Mongolian word image is divided into a quantity of frames with equal size along writing direction. Therein, the adjacent two frames are overlapped half a frame.
The Proposed Method (Cont.)

- **Output** of Model
- We have defined a list of glyphs with annotations to solve the problem that **many** visual forms may map to **one** letter.
The Proposed Method (Cont.)

- **Output** of Model
  - To be able to obtain the *information of glyph segmentation automatically*, an algorithm based on glyph splicing and dynamic time warping (DTW) has been proposed.
Outline

• Introduction
• The Proposed Method
• Experimental Results
• Conclusion
Experimental Results

• Dataset
  • A Mongolian dictionary has been collected, which consists of 202,500 vocabularies.
    ➢ word images: 810,000 (202,500 x 4 fonts); 300dpi
    ➢ label
      • glyphs
      • glyph segmentation
  • Training Set: 800,000(200,000 x 4 fonts)
  • Testing Set: 10,000 (2,500 x 4 fonts)

• Evaluation metric
  • The accuracy of recognition & IoU(intersection over unit)
Experimental Results (Cont.)

- The Impact of Fram Length

![Bar chart showing the impact of frame length on accuracy. Accuracy increases as the frame length decreases from 16 pixels to 2 pixels.]
Experimental Results (Cont.)

• Related Methods for Comparison

<table>
<thead>
<tr>
<th>Method</th>
<th>LSTM-CTC</th>
<th>Seq2seq</th>
<th>Our proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bai Font</td>
<td>87.96%</td>
<td>99.08%</td>
<td>99.92%</td>
</tr>
<tr>
<td>Bao Font</td>
<td>86.56%</td>
<td>99.01%</td>
<td>99.84%</td>
</tr>
<tr>
<td>Hawang Font</td>
<td>86.12%</td>
<td>99.00%</td>
<td>99.76%</td>
</tr>
<tr>
<td>Biaoti Font</td>
<td>88.12%</td>
<td>99.03%</td>
<td>99.96%</td>
</tr>
<tr>
<td>Avg</td>
<td>87.19%</td>
<td>99.03%</td>
<td>99.87%</td>
</tr>
</tbody>
</table>

• 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.
Experimental Results (Cont.)

• Related Methods for Comparison

<table>
<thead>
<tr>
<th>Method</th>
<th>Glyph-segmentation</th>
<th>Our proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bai Font</td>
<td>37.97%</td>
<td>78.30%</td>
</tr>
<tr>
<td>Bao Font</td>
<td>34.72%</td>
<td>82.98%</td>
</tr>
<tr>
<td>Hawang Font</td>
<td>28.73%</td>
<td>75.15%</td>
</tr>
<tr>
<td>Biaoti Font</td>
<td>40.65%</td>
<td>82.66%</td>
</tr>
<tr>
<td>Avg</td>
<td>35.52%</td>
<td>79.77%</td>
</tr>
</tbody>
</table>

• We can see that the attention mechanism significantly outperforms the conventional algorithm on the four types of fonts.
Experimental Results (Cont.)

- Related Methods for Comparison

The results of glyph segmentation by attention mechanism.

(a) The real areas of glyphs in one Mongolian word.

(b) The prediction areas of glyphs by attention mechanism.
Outline

• Introduction
• The Proposed Method
• Experimental Results
• Conclusion
Conclusion

- 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**.
Questions & Answer

Thank you for listening!