The Thirty-Second AAAI Conference
on Artificial Intelligence (AAAI-18)

Enhancing RNN Based OCR by Transductive Transfer Learning from Text to Images

Yang He, Jingling Yuan, Lin Li
Wuhan University of Technology
yanghe@whut.edu.cn, yjl@whut.edu.cn, cathylilin@whut.edu.cn

Abstract

This paper presents a novel approach for optical character recognition (OCR) on acceleration and to avoid underfitting by text. Previously proposed OCR models typically take much time in the training phase and require large amount of labelled data to avoid underfitting. In contrast, our method does not require such condition. This is a challenging task related to transferring the character sequential relationship from text to OCR. We build a model based on transductive transfer learning to achieve domain adaptation from text to image. We thoroughly evaluate our approach on different datasets, including a general one and a relatively small one. We also compare the performance of our model with the general OCR model on different circumstances. We show that (1) our approach accelerates the training phase 20-30% on time cost; and (2) our approach can avoid underfitting while model is trained on a small dataset.

Introduction

As a major application of pattern recognition and machine learning, optical character recognition (OCR) is widely used for converting text from visual documents into digitally text to facilitate document management for search and information retrieval. In this work, we concentrate on a basic sequence-to-sequence OCR model (Breuel et al. 2013), called DL model for short, aiming at designing a method that generalizes well to different model structures. Although existing advanced approaches about DL model has been incessantly come up with, the gist of these models remain constant, because additional networks just change the size of data DL model handles.

In this model, there is a typical phenomenon in training. Before the accuracy of OCR model rapidly increases, there lies a period, we called it the latent period, at the beginning of training that model’s accuracy stays nearly constant, occupying much time.

In addition, there lies a passive relationship between the volume of training data and the length of latent period. While trained on a huge dataset, the latent period has suppressed by the overwhelming amount of data. With the decrease of training data, the latent period lags and model's accuracy wanes. When it comes to a small dataset whose training data are smaller than its test data, DL model will fall into underfitting, and its accuracy fluctuates round a low level.

To efface two problems mentioned above, we build a model based on transductive transfer learning, called TTL model for short. Our hypothesis of this idea is that, the image is a counterpart of corresponding text in high dimensional space, namely image is an embodiment presentation of text in 2-dimension and text is a projection of image towards a lower dimension. Like dimensionality reduction, no matter what dimension data will be projected to, the relationship between data remains constant. In OCR, the relationship is character sequential relationship, an order probability between continuous characters. With transfer learning, the shared relationship can be transferred from text to images.

In this paper, we consider to combine these two problems together, as they are both resulted from the capacity of dataset. We design a novel approach to eradicate the two problems mentioned above. Our main contributions are:

1) Acceleration of training phase.
2) Avoiding underfitting in a small dataset.

Our approach

Our approach follows the combination architecture of a teacher network, matrix extension and a student network. In the teacher network, given an input sequence of sentences, the corresponding sequence of one-hot encoding is fed as the input of the teacher network with the output of probability matrix over characters. Then we extend these probability matrices to the same size of images by the rule of repeating tensors of matrix for certain times, being fed as the input of the student network. After that, parameters of the student network will be transferred to OCR model composed of RNNs with LSTM blocks and full connected networks with softmax activation, and then trained on image data. At this time, it was believed that the character sequential relationship has been transferred from text to the OCR model, ameliorating model's performance. Furthermore, the training of student network on extended matrices is 10% faster than the training of OCR model on images. The network architecture is illustrated in Figure 1.

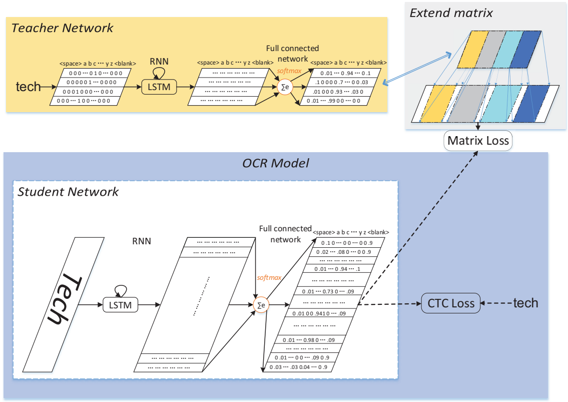

Figure 1. The Brief Architecture of Our Approach

Experimental Results

Acceleration

To support our approach, we trained 5 models on the same dataset: one was DL model without pre-training, and others were TTL model pre-trained for different times on extended matrices. After trained on images for 60 iterations, it turned out that the latent period of these TTL models are curtailed for 20-30% on time cost before its accuracy increases.

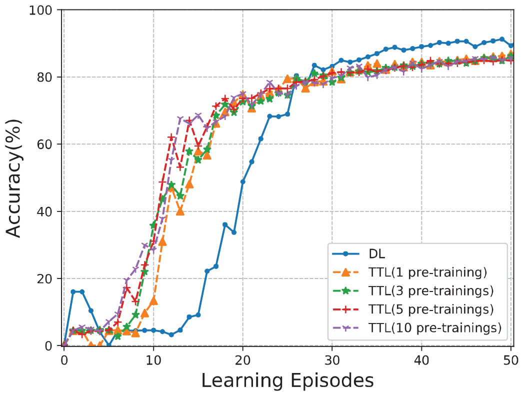

Figure 2. Results of Acceleration of Different Models

Underfitting

The accuracy of sequence-to-sequence OCR model is positive with the capacity of dataset. To emphasize this problem, we built a dataset whose training data are 500 images relatively small to a 1000 images test data. We firstly pre-trained TTL model on extended matrix for 50 iterations, and then trained TTL model on images for 100 iterations. And DL model is trained on images for 150 iterations. In addition, we delay the start point of TTL model curve to the point of 50 iterations. The trainings of two models are shown in Figure 3. DL model is always under a low accuracy, approximately 20%. Meanwhile, pre-trained model's accuracy can aggrandize as if it was trained on a common dataset, and finally it reaches quite higher than DL model, over 80%.

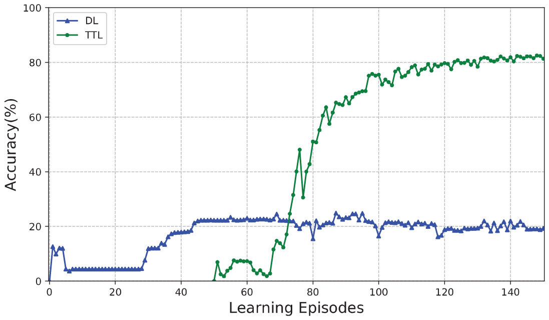

Figure 3. Results of Performance of Different Models

Conclusions and Future Work

The results presented in this paper show that our approach based on transductive transfer learning indeed can be used for the improvement of sequence-to-sequence OCR model both in acceleration and underfitting. While accuracy of our approach has waned, acceleration has been proved effective. In addition, this work can be extended and improved in future in many directions, such as, more delicate rule on matrix extension, higher accurate teacher network and multi-linguistic OCR model.

Acknowledgements

This work was supported by projects of 15BGL048, 2015BAA072, NSFC 61303029 and WUT: 2017II39GX.

References

Breuel, T. M., Ul-Hasan, A., Al-Azawi, M. A., & Shafait, F. (2013, August). High-performance OCR for printed English and Fraktur using LSTM networks. In *Document Analysis and Recognition (ICDAR), 2013 12th International Conference on* (pp. 683-687). IEEE.

Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, *22*(10), 1345-1359.