Adaptively Transfer Category-classifier for Handwritten Chinese Character Recognition

Yongchun Zhu[1, 2], Fuzhen Zhuang[1, 2], Jingyuan Yang[3], Xi Yang[4], Qing He[1, 2]

1. Key Lab of Intelligent Information Processing of Chinese Academy of Sciences (CAS) , Institute of Computing Technology, CAS, Beijing, China

2. University of Chinese Academy of Sciences, Beijing, China

3. George Mason University

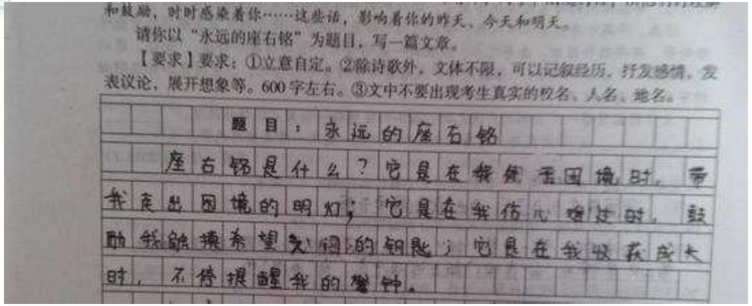
4. Sunny Education Inc., Beijing 100102, China

Motivation

Handwritten Chinese Character Recognition (HCCR) plays an important role in real-world applications.

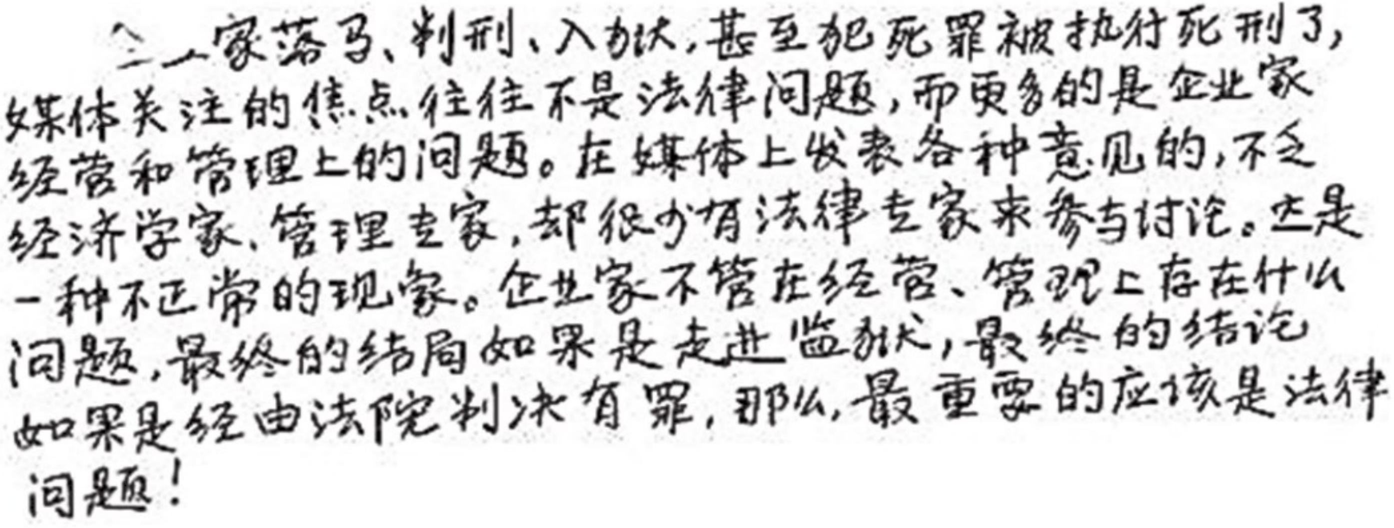
Bank Check Recognition

Intelligence Education



Motivation

Most research and competition of HCCR are focus on some standard data sets. The performance of machine on the data sets has surpassed humans.


家落马、判刑、入狱，甚至把死罪被执行死刑了，
媒体关注的焦点往往不是法律问题，而更多的是企业家
经营和管理上的问题。在媒体上发表各种意见的，不乏
经济学家、管理专家，却很少有法律专家来参与讨论。这是
一种不正常的现象。企业家不管在经营、管理上存在什么
问题，最终的结局如果是走进监狱，最终的结论
如果是经由法院判决有罪，那么，最重要的应该是法律
问题！

Motivation

Assume that we want to make an **Intelligence Education software** which need to recognize the handwritten Chinese characters of middle school students. However, we do not have enough labeled data. How do we do it?

Label enough data to train the model. ✖

Use extra HCCR data sets directly.

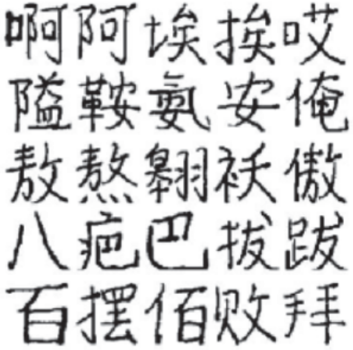
Motivation

The fonts are diverse.

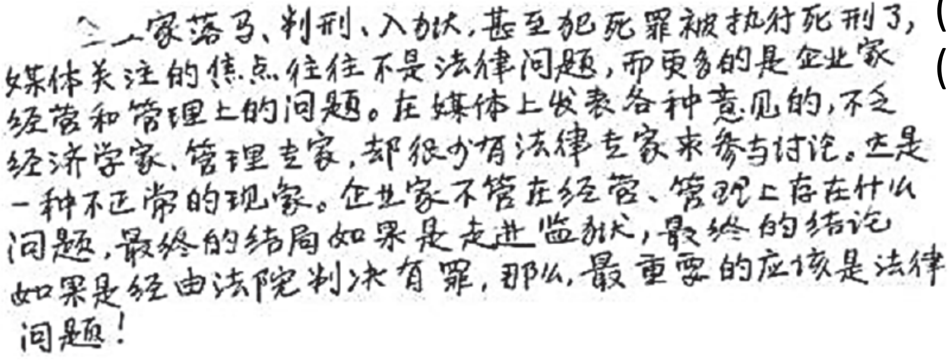
(a)Clear

(b)Middle

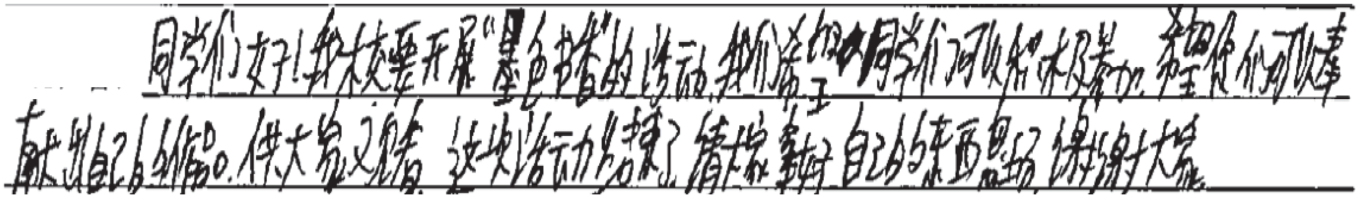
(c)Messy



(a) HCL2000



(b) CASIA-HWDB1.1



(c) MSS-HCC

Motivation

Assume that we want to make a **Intelligence Education software** which need to recognize the handwritten Chinese characters of middle school students. However, we do not have enough labeled data. How do we do it?

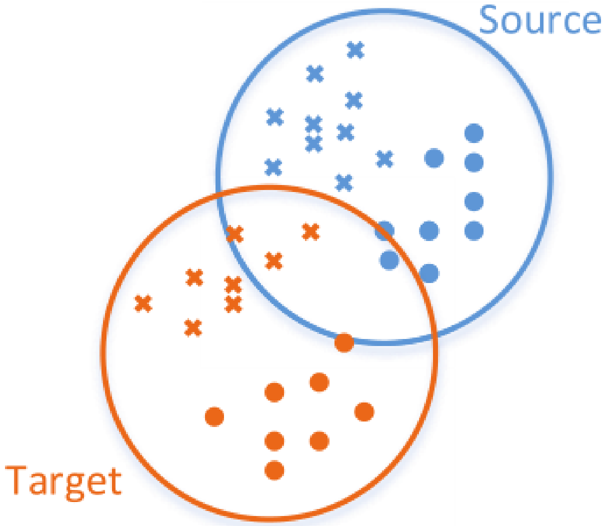
Label enough data to train the model. ✖

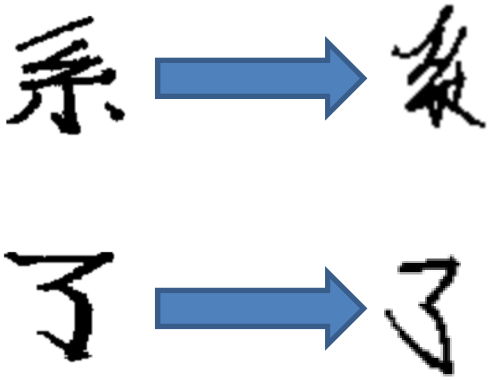
Use extra HCCR data sets directly. ✖

How can we use extra HCCR data sets to help real-world applications?

Motivation

Transfer learning!





Problem Definition

Source domain $\mathcal{D}_s = \{x_i^{(s)}, y_i^{(s)}\}|_{i=1}^{n_s}$

Target domain $\mathcal{D}_t = \mathcal{D}_t^L \cup \mathcal{D}_t^U = \{x_i^{(t)}, y_i^{(t)}\}|_{i=1}^{n_t^L} \cup \{x_i^{(t)}\}|_{i=1}^{n_t^U}$

Model

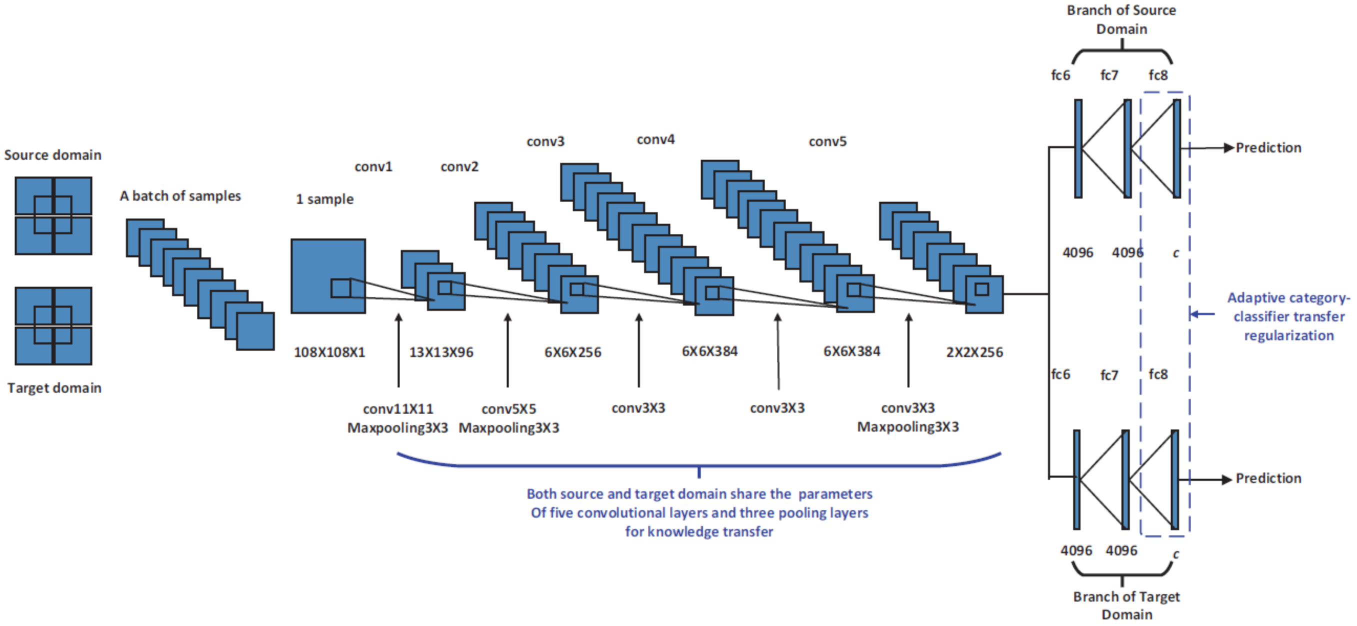


Fig. 2. The network structure of ATC-HCCR.

$$\mathcal{L} = J_r(\mathcal{D}_s \cup \mathcal{D}_t^L, y) + \Omega(W_s^8, W_t^8), \;\; \Omega(W_s^8, W_t^8) = \lambda \sum_{i=1}^{c} \gamma_i \cdot |\theta_i^{(s)} - \theta_i^{(t)}|.$$

Lambda is a trade-off parameter, while gamma is for transferring categoty-classifier between the source and target domains.

Model

Algorithm 1 Transfer Learning with Adaptively Transfer Category-classifier

Input: Given one source domain $\mathcal{D}_s = \{x_i^{(s)}, y_i^{(s)}\}|_{i=1}^{n_s}$, and one target domain $\mathcal{D}_t = \mathcal{D}_t^L \cup \mathcal{D}_t^U = \{x_i^{(t)}, y_i^{(t)}\}|_{i=1}^{n_t^L} \cup \{x_i^{(t)}\}|_{i=1}^{n_t^U}$, trade-off parameters λ and weights γ, the number of nodes in full connected layer and label layer, k and c.
Output: Results of x_i belongs to the vector of probability for each category.

1. Use both $\mathcal{D}_s$ and $\mathcal{D}_t^L$ to train AlexNet.
2. Use the parameters in Step1's model to initialize ATC-HCCR shown in Figure 2.
3. Choose a batch of instances from $\mathcal{D}_s$ or $\mathcal{D}_t^L$ as input.
4. Use AdamOptimizer with loss function Eq. (5) to update all variables.
5. Continue Step3 and Step4 until the algorithm converges.
6. Input $\mathcal{D}_t^U$ and get the vector of probability for each category that x_i belongs to.

Experiments

Table 2. The statistics of three data sets.

	HCL2000	CASIA-HWDB1.1	MSS-HCC
#category	3,755	3,755	27
#instance	3,755,000	1,126,500	5,920

HCL2000 and CASIA-HWDB1.1 are standard HCCR data sets, while MSS-HCC collected by ourselves is written by middle school students. MSS-HCC is written much in messy.

The original data is shown as above. We do not consider the split algorithm and we manually select clear pictures to form the MSS-HCC data set.

Experiments

Observations:

- The performance improves with the increasing values of sampling radio of target domain data as labeled data.
- Applying transfer learning for tackling HCCR problems is important.
- Our model ATC-HCCR achieves the best results over all baselines.

Table 3. The performance (%) comparison on three data sets among AlexNet-HCCR, preDNN and ATC-HCCR.

	HCL2000 → CASIA-HWDB1.1						Mean
	1.67%	3.33%	5%	6.67%	8.33%	10%	
AlexNet-HCCR(s)	63.30	63.31	63.40	63.48	63.53	63.41	63.41
AlexNet-HCCR(t)	30.83	61.64	79.52	78.04	81.01	81.85	68.82
AlexNet-HCCR(s+t)	73.07	76.78	79.52	81.13	82.24	82.08	79.14
preDNN	73.01	76.89	79.37	81.47	82.47	83.56	79.46
ATC-HCCR	76.79	79.78	82.37	84.13	85.08	85.06	**82.20**
	HCL2000 → MSS-HCC						Mean
	5%	10%	15%	20%	25%	30%	
AlexNet-HCCR(s)	61.49	63.18	62.61	62.75	63.92	64.30	63.04
AlexNet-HCCR(t)	66.44	82.83	89.77	90.96	92.57	93.00	85.93
AlexNet-HCCR(s+t)	86.31	88.95	91.02	91.55	92.22	94.76	90.80
preDNN	86.93	90.69	92.61	93.45	93.90	94.61	92.03
ATC-HCCR	87.76	91.12	93.24	93.71	94.57	94.88	**92.55**
	CASIA-HWDB1.1 → MSS-HCC						Mean
	5%	10%	15%	20%	25%	30%	
AlexNet-HCCR(s)	76.01	78.38	78.27	78.12	78.38	77.87	77.84
AlexNet-HCCR(t)	66.44	82.83	89.77	90.96	92.57	93.00	85.93
AlexNet-HCCR(s+t)	89.48	91.38	92.27	93.67	94.21	94.98	92.67
preDNN	89.19	92.64	92.74	93.58	94.61	94.98	92.96
ATC-HCCR	90.98	93.14	93.80	94.55	94.68	95.29	**93.74**

Conclusion

- As there is little work about transfer learning for HCCR, based on Alexnet, we propose a new network framework by adaptively transferring category-classifier for HCCR problems.
- We also collect a small set of much more challenging HCCR data, and finally conduct experiments on three data sets to demonstrate the effectiveness of our model.

Thanks Q & A

Name: Yongchun Zhu | Email: zhuyongchun18s@ict.ac.cn