「蓼食う PC も好きずき」ニューラルネットワークを用いた手書きの美しさの学習

ホァン　ハン[1,a)]　アランニャ クラウス[1,b)]

概要：美しい文字は読みやすく、理解しやすいと思われる。しかし、文字認識システムにおいて、この理論が成り立つかどうかは疑問である。更に、日本語の仮名は中国語の漢字が簡易化された後の産物なので、書き手の中国語経験の有無が日本語を書くことに影響を与えるかについても知る必要がある。最後に、文字認識システムが人間のように文字の美しさを判定できるかどうかについても研究の余地がある。その三点を確認するため、深層ニューラルネットワークを用いて機械学習による文字認識システムを実装し、中国語の習得経験のある留学生とない留学生の手書きデータを集め、実験を行った。その結果、人間が綺麗と判断した文字が、文字認識システムで認識されやすいとは限らないことが確認できた。又、中国語の習得経験がある人の方が、日本語の文字が美しい確率が高いことが分かった。参考に、機械学習による文字の美しさの判定について、現在までに確認できた結果を示す。

Beauty is on the chip of the beholder: Teaching a neural network to recognize beautiful handwriting

Huang Han[1,a)]　Claus Aranha[1,b)]

Abstract: It is said that a property of beautiful words is that they are easier to be read and understood, but we wonder if this is the same for machines. Furthermore, Japanese characters such as Hiragana come from simplified Chinese characters, so we want to know if having learned Chinese would affect the handwriting of Japanese kana or not. At last, we wonder if the machine could learn the properties of beautiful handwriting. We set up a neural network system where the machine learns to recognize hiragana and katakana from a standard database, and we collect handwriting samples from many foreign Japanese learners, with and without Chinese background. By using this neural network system, our results indicate that beautiful words (as judged by humans) are not necessarily easier to be recognized by a machine. Also, we note that Chinese language background does not affect the machine-recognition ratio, but plays a big role on the beauty of the words (as judged by humans). Finally, we show some preliminary results for the ability of the machine to learn the concept of beautiful handwriting.

1. Introduction

Nowadays, more and more people use a computer to write instead of relying on words written by hand. Because of this, having beautiful handwriting might be more impressive because it is rarer. However, it is hard to define whether a person's calligraphy is beautiful or not, since this is a highly subjective evaluation. Thus, we wondered if there are any common properties of beautiful words.

Currently there are some smartphone applications[*1]that claim to judge the beauty of one's handwriting, but these applications have some limitations. First, they require the user to write on the phone's interface, instead of judging a word written on paper. This means that the user is writing on an interface which is not as natural as paper which is currently more used for

[1] 筑波大学情報科学類
Tsukuba University, College of Information Sciences
[a)] amanotaiga3665@gmail.com
[b)] caranha@cs.tsukuba.ac.jp

© 2017 Information Processing Society of Japan

handwriting in daily life.

Second, the most used method for judging beauty in these applications is to take the x, y coordinates of the words drawn by the user, and compare these with a standard font. However, humans take into account things other than purely the distance between two lines, judging also the curvature, consistency and many other things.

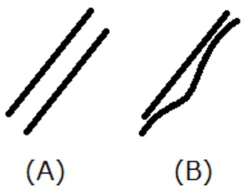

Fig. 1: どちらのラインが似てますか？（A）がカーブにより評価され、（B）が距離により評価されます。

Fig. 1 Which pairs of lines are more similar? (A) Judges by curvature, while (B) judges by distance

If we only compare the distance, then as illustrated in Figure 1, the two lines of (B) would be "better" than (A). However, humans are sensitive to curvature. Because the criteria used by humans are complicated, we didn 't manage to find a definite one. Instead, we tried to make a computer learn the personal evaluation and predict if a given handwriting sample is beautiful or not.

We set up a neural network system where the machine learns to recognize hiragana from a standard database, and we collected handwriting samples from many foreign Japanese learners, with and without Chinese background. By using this neural network system, our results indicate that beautiful words (as judged by humans) are not necessarily easier to be recognized by a machine.

Also, we note that Chinese language background does not affect the machine-recognition ratio, but plays a big role on the beauty of the words (as judged by humans). Finally, we find it is possible for a computer to learn the properties of beautiful words.

2. Related Works

Regarding the recognition of Japanese characters, Charlie Tsai [1] achieved 96.1% accuracy on Japanese Character Hiragana by using Convolutional neural network. Because Japanese recognition was only part of our research, we did not try to design a brand new neural network. Instead, we used Tsai's work as a base, and tested several small variants based on the best model in his report. After finding one that had some improvement, we used it to train the recognition and beauty evaluation.

[*1] 美文字判定,Gloding Inc.

表 1: ETL Character Dataset の詳細

Table 1 Details of the ETL Character Dataset

Set	Size,	Classes	Writers	Examples
ETL-7	64,63	48	200	9600
ETL-8	128,127	71	160	11360

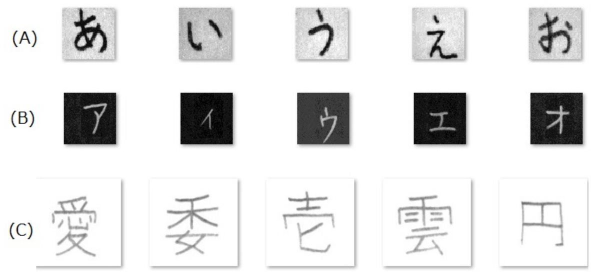

Fig. 2: (A) ひらがな, (B) カタカナ, (C) 漢字

Fig. 2 (A) hiragana, (B) katakana , (C) is kanji

3. Dataset

We used the ETC Character Database, produced by the Electrotechnical Laboratory [3], as our training data. The ETL database contains hand-written and machine-printed numerals, symbols, Latin alphabets and Japanese characters. Each dataset has different numbers of writers (Figure 3), classes, and size (Table 1). Although there are three categories of characters in Japanese (illustrated in Figure 2), we focused on Hiragana in this research. Also, we chose to combine ETL-7 and ETL-8 in order to have more data and used it to train our neural network. Because we didn ' t use handakuten and dakuten, we only picked up 46 classes in both datasets

We performed the following preprocessing steps on the data to increase the contrast for the evaluation function (Figure 4): Extract the files from ETL dataset to get the images of Hiragana characters. Resize them into the same size (32*32), do brightness enhance, and then crop and center the images. To generalize the result of neural network, we added zoom (0.2) when preprocessing.

Fig. 3: 協力者のサンプル

Fig. 3 A sample of writers

Fig. 4: 画像前処理

Fig. 4 preprocessing

4. Experiment

We collected handwriting data data from 46 students and 2 teachers in Tsukuba University, including 20 Not-

© 2017 Information Processing Society of Japan

Japanese without Chinese background, 17 Not-Japanese with Chinese background and 11 Native Japanese speakers. We asked each participant to write the 20 character sentence “あおきようたさんはしめいてはいされました” (Figure 5), taken from a sample news paragraph[*2].

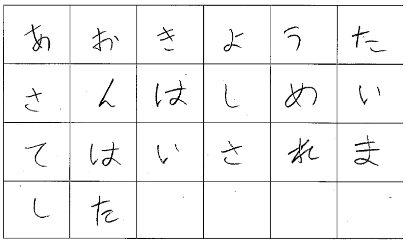

Fig. 5: 記入欄の例。
Fig. 5 Sample of writing sheet.

After we collected the handwriting sample, 30 people chosen randomly from the participants evaluated the beauty of the samples. Each person judged 10 sample sheets, so we had 6 evaluations for each sheet.

To simplify our analysis we used only a two-level judging score. Each person would evaluate each sheet as “beautiful” or “not beautiful”. This score would be given for the entire sheet, instead of one score for each character.

For classification, we classified the handwriting sheet into beautiful only when it got more than half positive judgment. The The result showed that 37 sheets are not beautiful, 11 sheets were beautiful. We transformed the data into computer by scanning.

5. Method

5.1 Convolution Neural Network

Convolutional Neural Networks are inspired by biological structure, which contains a large amount of cells. These cells are found to be activate the sub-region of the visual field. Inspired from this study, the neurons in a convolutional layer connect to the sub-regions of the layers before that layer instead of being fully-connected as in other types of neural networks. Therefore, convolutional layer could have relatively lower cost in computing comparing to other neural network.

Using non-linear Relu activation function, which is $\max(0, x)$, increased representational power of the model. Using pooling helped reduce the parameters and avoid over-fitting. Dropout is another way to avoid over-fitting by keeping some active neuron, while others are randomly turned off. Softmax function is a function often used as a classifier. This function could show probabilities of K different possible outcomes, and add up to 1.

5.2 Network Structure

Relu activation function was used at every layer. First, we had 32 convolutional layer in 3by3 and stride 1. After pooling, there was 64 convolutional layer in the same size 3by3 and stride 1. After these convolutional layers, we had a flatten layer and the first fully-connected layer to combine all characteristics learned by layers before that. At last, we had a softmax function to classify the result.

5.3 Training for hiragana recognition

Following the idea described by Tsai [1], we used the Adam optimizer [2] using its default parameters, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-8}$, and 16 for batch size.

Due to time constraint, we had 70 epoch and took average of running 5 times, 60% of data for training set, 20% for test set, and 20% for validation set. After testing several models in 7-layer or 8-layer, the result showed that 7-layer with fully-connected layer (512) had the highest accuracy, so we used this 7-layer neural network.

We used this structure to both train hiragana recognition and the beauty of characters. As shown in table 2, test accuracy was 99.02% for hiragana recognition.

5.4 Training for beauty

Starting from the supposition that there are some common properties of beautiful handwriting, we trained the network on the full samples, and not on individual characters. That is, we had 37*20=740 characters that were not beautiful and 11*20=220 characters were beautiful.

However, there were more not beautiful samples than beautiful ones, so we duplicate the beautiful samples to make both classes the same amount in a balanced dataset. Thus, we also had 37*20=740 characters which were judged to be beautiful.

Using almost the same neural network as for recognition, only changed the dense layer to 256 neurons to avoid over-fitting. With 100 epoch and 5 repetitions, we had 80% for test accuracy (Figure 6).

6. Result

There are three graphs in this paragraph. First is background vs. accuracy (Figure 7), which means how often the by pre-trained model could correctly identify a character. The second graph (Figure 8) is background vs. the beauty evaluation (by people). The third (Figure 9) is

[*2] 【福田文彦画像】郵便局員を逮捕！客の口座から無断で２９００万円！, http://masakichi0628.com/fukuda-fumihiko-41-settou

表 2: 試したニューラルネットワーク構造とその正答率

Table 2 Different configurations tried and their test accuracy

M7-1	M7-2	M7-3	M8_1	M8_2	M8_3	M8_4
Conv32	Conv32	Conv64	Conv32	Conv32	Conv32	Conv32
Conv32	Conv32	Conv64	Conv32	Conv32	Conv32	Max pool
Max pool						Conv4
Conv64	Conv64	Conv128	Conv64	Conv64	Conv64	Max pool
Conv64	Conv64	Conv128	Conv64	Max pool		Conv128
Max pool				Conv128	Conv128	Conv128
Fc256	Fc512	Fc1024	Fc512	Max pool		
			Fc512	Fc512	Fc1024	Fc1024
Accuracy for each network structure						
0.98017	0.99027	0.98385	0.98854	0.98628	0.98645	0.98090

accuracy vs. beauty evaluation, which shows the relationship between beautiful handwriting and ease for machine to recognize it or not. To compare the significance between the differences in the first two graphs, we use Student's t test with a threshold of 5% for p.

In graph of Background vs. accuracy, it is clear that the accuracy of Japanese-Native is much higher than Not-Japanese, however, there is no significant difference between foreigners who have Chinese background and those who don't (p-value bigger than 0.3).

In graph of Background vs. Beauty, the result shows that the beauty scores of foreigners of Chinese background are the same as Japanese (p-value bigger than 0.4),and both are much higher than foreigners without doesn ' t

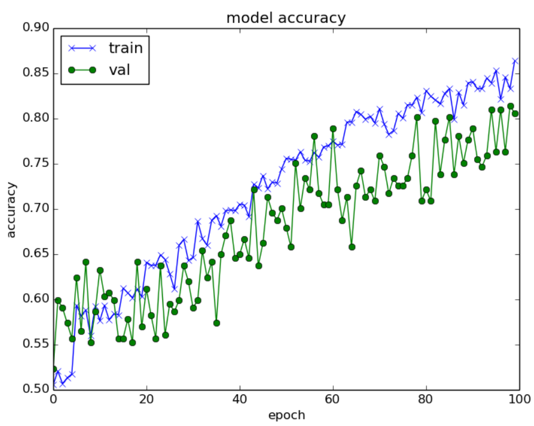

Fig. 6: 学習グラフの一つ（全 5 回）

Fig. 6 One of the training curves (out of 5 total)

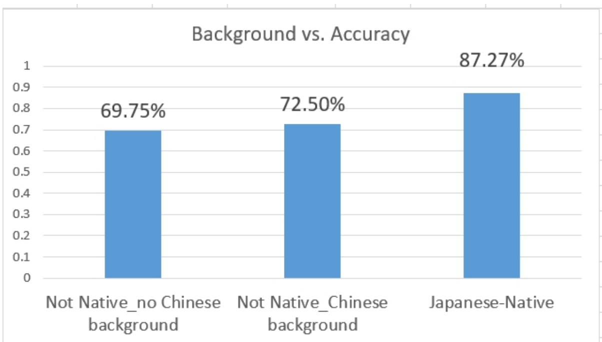

Fig. 7: 言語教育と正答率

Fig. 7 Language Background vs. Accuracy

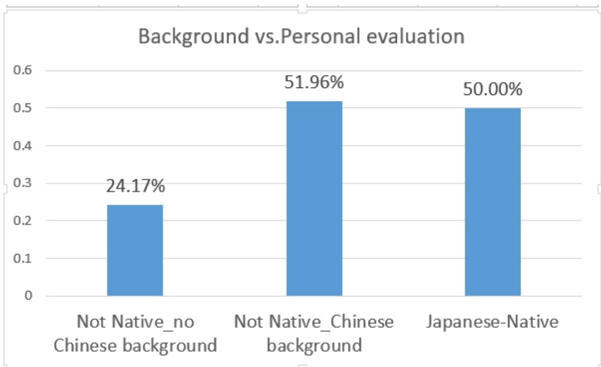

Fig. 8: 言語教育と綺麗の評価

Fig. 8 Language Background vs. Personal evaluation

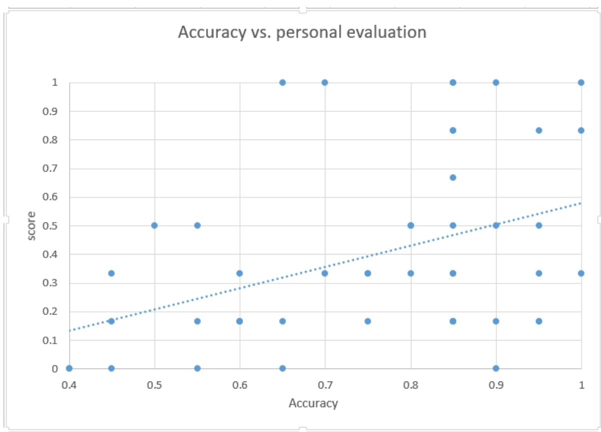

Fig. 9: 正答率と綺麗評価

Fig. 9 Accuracy vs. Beauty

© 2017 Information Processing Society of Japan

have Chinese background.

Also, the accuracy for all handwriting was around 72%, which is much lower than 98.9% obtained for the ETL character dataset. The correlation between Accuracy and Beauty score is 0.4.

7. Discussion of the Results

The correlation between accuracy and beauty score is not really high but still shows some positive relationship. Although it is said that one of the properties of a beautiful handwrite is that it is easy to be recognized, it seems that we cannot say the same for the machine.

Second, we wanted to know whether Chinese background would affect the recognition and beauty scores or not. The results showed that the recognition accuracy between foreign students with Chinese background and without Chinese background are almost the same. However, the beauty score had significant differences. The beauty score of foreigners who had Chinese background were almost the same as Japanese. Since Japanese Characters were simplified from Chinese Kanji, those who had Chinese background were more likely to grasp the balance of Japanese character. However, this still doesn ' t mean that it is easier for the computer to recognize.

Third, from the results of beauty classification on the neural network (Table 3), it showed that the machine was good at predicting not beautiful samples. However, it would often misclassify beautiful samples as not beauriful mistakenly. Figure 10 shows some of these examples.

Lastly, from classification report(average of running 5 times), it shows that machine are good at predicting not beautiful ones, and seldom misclassified not beautiful ones into beautiful ones, however, machine made more mistake on misclassified beautiful ones into not beautiful ones. Figure 10 are some example of misclassified images.

表 3: 機械学習による綺麗の認識結果

Table 3 Results for machine beauty classification

	precision	Recall
Not beautiful	0.91	0.64
Beautiful	0.72	0.94

8. Conclusions

Due to time constraint, we gave beauty scores to full character sheets, and not individual characters in this experiment. However, a person might write beautiful " A " but not beautiful " B ". Thus, we want to judge by character in the future.

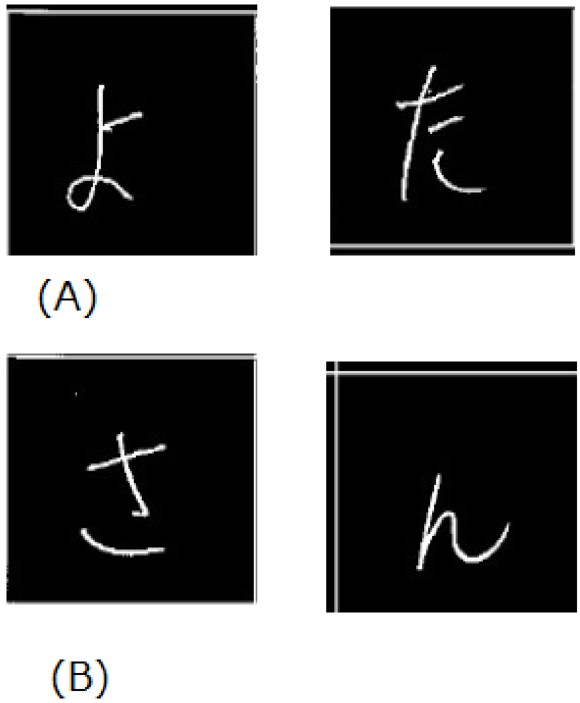

Fig. 10: 認識誤解の例.
(A) 機械が美しいと判断され、人間が美しいと判断されなかった。
(B) 人間が美しいと判断され、機械が美しいと判断されなかった。

Fig. 10 Example of misclassification. (A) Judged beautiful by the machine, but not beautiful by human. (B) Judged not beautiful by machine, but beautiful by human.

Furthermore, considering the judgment were given by people of different nationalities, 6 scores might not be enough because the aesthetics of handwriting are different from country to country.

Also, small dataset was one of the problems in this research, we only had 20*48=960 characters in total. However, we had more than ten thousand images for training recognition.

参考文献

[1] Charlie Tsai, *Recognizing Handwritten Japanese Characters Using Deep Convolutional Neural Networks*, Technical Report, Stanford University. (2016)

[2] Diederik P. Kingma and Jimmy Lei Ba, *Adam: A Method for Stochastic Optimization* Proceedings on the International Conference on Learning Representations – ICLR doi:abs/1412.6980. (2014)

[3] Electrotechnical Laboratory, Japanese Technical Committee for Optical Character Recognition, *ETL Character Database*, 1973-1984.

[4] Yuhao Zhang, *Deep Convolutional Network for Handwritten Chinese Character Recognition.* Technical Report, Stanford University. (2015)