Comparison of Chinese Character Correct and Error Classifier for Overseas Students Based on Handwriting Motion Characteristics

Jun Zhang and Xiwen Zhang*

1College of Advanced Chinese Training, Beijing Language and Culture University, Beijing, China; 2College of Information and Science, Beijing Language and Culture University, Beijing, China
Email: 1zhangjun@blcu.edu.cn, 2*zxw@blcu.edu.cn

Abstract: The quality of Chinese characters written by foreign students is reflected not only in the result of writing, but also in the writing movement. 25 characteristics of handwriting movement were selected from 39 parameters, such as time, space, movement and pressure. Based on the feature of handwriting motion, 6 classifiers, such as Decision Tree and Artificial Neural Network, are selected to compare the performance of them. Experiments show that the accuracy of DT algorithm is the best regardless of whether the target characters are known or not. Handwriting movement features are effective representations of the quality of Chinese characters written by foreign students.

1. Introduction

With the rapid development of international Chinese education, more and more foreign students take Chinese proficiency test, especially computer-based test. How to evaluate the quality of Chinese characters handwritten in real time and quickly is a key technical problem to be solved.

The basis of evaluating the correctness and error of Chinese characters handwritten is the legal norms of relevant characters. The objects of evaluating include strokes (such as stroke order, stroke shape, stroke direction, etc.), radical, whole characters (such as structural layout, etc.). Generally speaking, the research paradigm of quality evaluation is: collecting writing samples - selecting handwriting features - evaluating writing quality. There are two kinds of writing features extracted: one is Chinese character type/shape feature [2-5]; the other is Chinese character image feature [6-7]. By comparing the characters handwritten and template characters, we can evaluate the correctness of writing. In addition to strictly referring to the legal writing standards for comparison, some studies [8] also consider that in the actual writing practice, in order to pursue the beauty of calligraphy, the strokes and components will be alienated and deformed, and put forward the correctness standards suitable for different situations. According to the researcher, that evaluation scheme can be applied to the evaluation of Chinese character handwritten by foreign students, but no specific experimental work has been found until now. The above researches are based on the premise of known "template characters", that is, the target characters are known before the evaluation.

According to the related theory of handwriting output [9], handwriting is a complex fine motor skill, which involves many cognitive abilities such as motor control, vision space perception and language ability [10]. It requires not only the motor control of hand muscles or small muscle groups, but also the combination of psychological activities such as sensory perception and attention to complete specific tasks. Different handwriting qualities reflect the different physiological and psychological cognitive
states of writers, and present regular writing movement characteristics in the writing process, such as writing track length, speed, acceleration, etc. At present, almost all the research all the world aims to reveal the characteristics of handwriting movement of some kinds of patients such as agraphia [11], Parkinson’s [12], writing spasm [13], hyperactivity disorder [14], etc., and to develop clinical diagnosis and intervention treatment tools.

In view of the "specificity" of handwriting motion features of specific population groups, some researchers have found a new way to evaluate the quality of handwriting by extracting kinematic features (such as speed, acceleration, etc.) or dynamic features (such as stroke pressure, etc.) Guan [15] proposed that there are 19 physiological and cognitive behavioral factors affecting the writing quality of Chinese characters. The results showed that the most important factor affecting the handwriting quality of primary school students in lower grades is the degree of familiarity with orthography and pen strength. With the popularization of digital tablet and digital ink technology, it is possible to obtain more detailed handwriting motion feature data [16]. Zhang [17] uses Anoto digital pen to select 25 kinds of writing motion features from four dimensions of time, space, motion and power, and uses artificial neural network (ANN) algorithm as classifier to classify writing correctness and errors under the condition of unknown "template characters". The results showed that the classification accuracy of a single ANN classifier is not ideal. After using the lifting algorithm, the accuracy is more than 90%. The movement features are effective representations of the quality of Chinese characters handwritten by foreign students.

The above research paradigm has a high accuracy in classifying the correctness and errors through the shape or image features of handwriting characters. However, as mentioned before, these researches are based on the premise of known "template characters", that is, the written target characters before the evaluation, which belongs to "post evaluation". We can break through the limitation of known template words and realize the real-time "blind evaluation" in the writing process under the condition of unknown target words through the handwriting movement features. As a part of a series of studies, this study will continue to compare the performance of a variety of classifiers, and finally choose the best classifier based on the writing motion features, which is suitable for the quality evaluation of Chinese characters handwritten by foreign students.

2. Method

2.1 Datum

In this study, an experimental class was selected from 15 beginning level Chinese classes of one college in Beijing Language and Culture University as a fixed source of data acquisition. The 404 types of Chinese characters (15542 tokens) were collected in the learning stage of Chinese characters from the first three volumes of the textbook ‘The Road to Success’ [18], which is used by BLCU. There were 22 writers, including 14 male and 8 female students, who come from 12 different non-Chinese cultural countries.

332 types of characters without incorrect token and 404 types with incorrect token have 5378 and 10164 tokens in the database, respectively, which constitute the “correct database” and the “incorrect database”. 21 types have just less than two tokens, in order to ensure the equalization of samples, in this experiment, 310 correct types of character were sampled from the “correct database” by stratified random sampling, and 2 correct tokens were selected for each character type randomly. Matching to 310 types from the “correct database”, 620 incorrect tokens are also extracted from the “incorrect database”, which together constitute a "training sample". Repeat the above method to implement non-return sampling and get "test samples".

The research uses digital ink technology, "Anoto" digital pen system and matching writing paper to collect the data of the writers' handwriting behavior. In the handwriting movement, the process from stroking to lifting is a stroke-step (SS). When a writer writes Chinese characters in a standardized stroke form according to the stroke order criterion, the SS corresponds to the stroke. If the stroke sequence or the stroke form is out of order, there is no corresponding relationship between the two.
2.2 Features
Written motion analysis includes kinematics analysis and dynamics analysis. Among them, kinematic analysis parameters include time variables, space variables and motion parameters, and dynamic analysis is generally written pressure [19]. Danna J.et al. [20] summarized the variables and their frequency of use in 42 related studies from 1993 to 2012. In our study, 14 features with frequency of use more than twice were selected by referring to the statistical results. According to the characteristics of Chinese characters, 25 features were put forward, a total of 39 features, specifically divided into four categories: spatial information, time information, motion information and dynamic information. According to the method of Tsanas et al [21], 25 features which are significantly related to the predictive classification variable ("correct word"/ "CW") at the level of 0.05 according to Pearson coefficient are selected, as shown in table 1.

| Spatial information               |
|----------------------------------|
| FPL                              | the frequency of pen lifting |
| ASSL                             | average stroke step length   |
| MinSSL                           | minimum stroke step length   |
| RMinSSL                          | the ratio of minimum stroke step length to total handwriting |

| Time information                 |
|----------------------------------|
| GT                               | global time                   |
| PT                               | on-paper time                 |
| RPAT                             | the ratio of on-paper time to in-air time |
| ASST                             | average stroke step time      |
| AAT                              | average in-air time among stroke steps |
| MaxSST/ MinSST                   | Maximum/ minimum stroke step time |
| RMinSST                          | the ratio of minimum stroke step time to on-paper time |

| Motion information               |
|----------------------------------|
| GV/ GA                           | global velocity/ acceleration |
| PV/ PA                           | on-paper velocity/ acceleration |
| MaxSSV/ MinSSV                   | Maximum/ minimum stroke step velocity |
| GV-NCV/GA-NCV                    | global velocity / acceleration NCV |
| MaxSSV-NCV/MaxSSA-NCV            | maximum stroke step velocity/ acceleration NCV |
| ASSV-NCV/ ASSA-NCV               | average stroke step velocity/ acceleration NCV |

| Dynamical information            |
|----------------------------------|
| MinSSP                           | minimum stroke step pressure   |

2.3 Classifies
Six widely used classifiers such as decision tree, artificial neural network and Bayesian network are selected for comparison.

1) Decision tree (DT)
Decision tree [22] is a supervised machine learning algorithm, which uses tree structure to classify objects. According to the feature, the node subtree is created recursively, the node records the feature, and the leaf node is the category. Decision tree is one of the most widely used classification and prediction algorithms because of its excellent data analysis ability and intuitive results display. C5.0 decision tree algorithm is selected in our experiment, which is improved by ID3 algorithm and has milestone significance.

2) Artificial neural network (ANN)
ANN [22] simulates human brain thinking, which is divided into input layer, middle layer and output layer. The middle layer may contain several hidden layers. The number of layers and the number of neurons in each layer determine the complexity of the network. Through repeated learning
and training of known information, the network can deal with the relationship between information and analog input and output by gradually adjusting the weights of neuron connections. Neural network technology has obvious advantages in dealing with fuzzy data, random data and nonlinear data. In our experiment, multi-layer perceptron algorithm is selected, which is the most commonly used structure of artificial neural network.

3) Bayesian network (BN)
BN [22] is one of the most effective theoretical models in the field of uncertain knowledge expression and reasoning. A Bayesian network is a directed acyclic graph. Nodes represent random variables. The directed edge between nodes represents the relationship between nodes. Conditional probability is used to express the relationship strength. Prior probability is used to express the information without parent nodes. The uncertainty is described by Bayes probability, and the utility function is used to select the optimal decision which maximizes the expected utility, so as to realize the reasoning of uncertainty.

4) Support vector machine (SVM)
SVM [22] is a kind of data mining method developed on the basis of statistical learning theory. It is a kind of generalized linear classifier which classifies data according to the supervised learning method. It has many advantages in solving the small sample, nonlinear and high-dimensional binary classification problems. In this algorithm, the observation data is regarded as m points in the n-dimensional feature space. By finding a hyperplane in the feature space, the two types of samples are classified separately.

5) Binomial logistic regression (BLR)
BLR [23] is a kind of generalized linear model. It is a multivariate analysis method to study the relationship between binomial observation results and some influencing factors. Through the maximum likelihood estimation, the logistic classifier obtains a set of weights, which directly reflect the size and direction of the action of independent variables. After the weights and test data are added linearly, they are classified according to the sigmoid function.

6) k-nearest neighbor classifier (KNN)
KNN classifier [24] calculates the distance between the unknown samples and all the known samples by taking the samples of all the known classes as the reference, selects the K known samples closest to the unknown samples, and classifies the unknown samples and the k nearest samples into one class according to the voting rule that the minority obeys the majority. Therefore, the algorithm does not conduct data modeling, the "lazy learning method" is to be carried out when entering the test data.

2.4 Result
On the platform of SPSS modeler 18.0, this research uses "C5.0", "neural network", "Bayesian network", "SVM", "logistic" and "KNN" nodes to compare the classification methods. Except for "logistic" and "KNN", all algorithm nodes are system default settings. "Logistic" adopts "forward stepping method"; KNN "selects balance speed and accuracy, automatically calculates K value (3-5), European distance.

| Table 2. Classification accuracy without template characters |
|-----------------------------------------------------------|
| sample          | training sample | test sample |
| DT              | 80.48           | 74.35       |
| ANN             | 68.79           | 68.87       |
| BN              | 68.15           | 66.37       |
| SVM             | 68.39           | 68.06       |
| BLR             | 68.63           | 68.63       |
| KNN             | 74.84           | 72.02       |

All of the above classifiers take about 10s, and the operation speed is not much different. It can be
seen from table 2 that in unknown writing template characters, the average accuracy rate of evaluating the correctness and error of Chinese characters written by foreign students through Chinese characters writing movement features is 69.4%, and the accuracy rate of DT is the highest, reaching 74.75 in the test sample; the next is KNN classifier, and the third is ANN classifier, reaching 72.02% and 68.87 respectively. Although the classification accuracy of nearly 75% is not ideal, the study can still show that the movement features of Chinese character handwriting is an effective representation of the writing quality of foreign students. Based on the written motion features, the decision tree classifier has the best classification effect.

In order to improve the accuracy of classification, 34 kinds of Chinese character shape related features, such as number of strokes, number of substrokes, number of components, Chinese character structure, types of strokes, types of substrokes, types of components and so on, were introduced in the experiment, which together with 39 kinds of writing movement features formed a feature set. Through SPSS modeler 18.0 feature selection algorithms, 25 motion features and one shape feature (the ratio of stroke number to stroke-step number) are screened out. The specific settings of the six classifiers are the same as above.

Table 3. Classification accuracy with template characters

|     | Training sample (performance) | Test sample |
|-----|------------------------------|-------------|
| DT  | 81.85(1.7%)                  | 79.76(7.3%) |
| ANN | 74.11(7.7%)                  | 72.74(5.6%) |
| BN  | 72.66(6.6%)                  | 70.65(6.4%) |
| SVM | 68.31(-0.1%)                 | 67.98(-0.1%)|
| BLR | 66.13(-3.6%)                 | 65.81(-1.3%)|
| KNN | 76.13(1.7%)                  | 73.63(2.2%) |

It can be seen from table 3 that after the known number of strokes of writing target words, i.e. adding the "step to draw ratio" feature, the classification performance of DT, ANN, BN and KNN has been improved, among which the accuracy of DT performance has increased by 7.3% to 79.76; BN and Ann have increased by 6.4% and 5.6% respectively, but in terms of overall accuracy, KNN's classification performance is still next to DT, higher than BN and ANN, to 73.6%.

3. Discussion
In view of the evaluation of the correct and wrong writing of Chinese characters by foreign students, this paper starts with the writing process, selects 25 kinds of writing movement characteristics of Chinese characters, selects six classifiers DT, ANN, BN, SVM, binomial logistic regression and KNN, and "blind evaluation" under the condition of unknown writing target words. The test results show that the classification performance of decision tree is the best, and the writing motion feature is an effective representation of writing quality. After adding a glyph feature, the classification performance of decision tree is improved the most, and the accuracy is also higher than other classifiers, about 80%.

4. Acknowledgments
This research project is supported by Science Foundation of Beijing Language and Culture University (supported by "the Fundamental Research Funds for the Central Universities") (18ZDJ05).

5. References
[1] Zhang J 2015. The Construction of error type system in Chinese character writing behavior of overseas students, Overseas Chinese Education, 2: 269-275.
[2] Xia W and Jin L 2008. A Method for layout evaluation of online handwritten Chinese character quality based on template, National Symposium on Pattern Recognition, Beijing: China Automation Society, China Image and Graphics Society, 354-359.
[3] Wang Q, Dai Y, Fan L and Sun G 2013. Fuzzy Analysis Method for quality of handwritten Chinese Characters, Computer Engineering and Applications, 49 (21): 180-185.
Bai X, Jiang J, Deng H and Li Y 2015. Experimental research on the stroke of handwritten Chinese character identification methods based on similarity. Research and exploration in laboratory, 32 (12): 132-136, 167.

Han R, An W, Xun E and Li Q 2016. Real-time grading judgment for stroke quality in Chinese character handwriting. Journal of computer applications, 36 (S1): 281-285,314.

Liu Y 2000. The Research in Auto Judge on hand mapping typeface. Journal of institute of Surveying and Mapping, 17 (1): 49-52.

Qi H, Chen F, Zhuang L, Chen B 2008. A size-unrestricted method for Chinese character writing structure assessment. Journal of Zhengzhou University (Natural Science Edition), 03: 59-62.

Jiang J, Wu J.Y, Han Q, Li Y 2019. Implementation and Effects Test of Comprehensive Scheme for Correctness Evaluation of Handwritten Chinese Characters, E-education Research, 09:50-57.

Van Galen G P. 1991. Handwriting: Issues for a psychomotor theory [J]. Human movement science, 10(2-3):165-191.

Cummings J. 1985. Clinical Neuropsychiatry, New York : Grune and Stratton.

Falk, T. H., Tam, C., Schellnus, H., & Chau, T.2011. On the development of a computer-based handwriting assessment tool to objectively quantify handwriting proficiency in children [J]. Computer Methods and Programs in Biomedicine, 104(3), 102–111.

Van Gemmert, A. W. A., Teulings, H.-L., & Stelmach, G. E. 2001. Parkinsonian patients reduce their stroke size with increased processing demands [J]. Brain and Cognition, 47, 504–512.

Baur, B., Schenk, T., Fu’holzer, W., Scheuerecker, J., Marquardt, C., Kerkhoff, G., et al. 2006 Modified pen grip in the treatment of writer’s cramp [J]. Human Movement Science, 25, 464–473.

Shen, I.-H., Lee, T.-Y., & Chen, C.-L. 2012. Handwriting performance and underlying factors in children with attention deficit hyperactivity disorder. Research in Developmental Disabilities, 33, 1301–1309.

Guan Q 2013. Define Writing Disabilities and Evaluate Writing Quality, Chinese Journal of Special Education, 02:51-56.

Mathew T, Abhishek L, Pramod K.P. 2017. Handwriting Analysis in Parkinson’s Disease: Current Status and Future Directions, Published online 1 November 2017 in Wiley InterScience (www.interscience.wiley.com).

Zhang, J. 2019. Automatic Quality Evaluation of Chinese Character Handwriting by Foreign Students Based on Handwriting Movement Characteristics. CSAE 2019: Proceedings of the 3rd International Conference on Computer Science and Application Engineering. 1-6.

Qiu J, 2008. Road to Success, Beijing: Beijing Language and Culture University Press.

Tucha, O., Tucha, L., & Lange, K. W. 2008. Graphonomics, automaticity and handwriting assessment: Graphonomics, automaticity and handwriting assessment. Literacy, 42(3), 145–155.

Danna,J.,Paz- Villagrá’n, V., Velay,J. 2013. Signal-to-Noise velocity peaks difference: A new method for evaluating the handwriting movement fluency in children with dysgraphia Research in Developmental Disabilities, 34:4375–4384.

Tsanas,A., Little,M., P. McSharry,P., Ramig,L., 2011. Nonlinear speceanalysis algorithms mapped to a standard metric achieveclinical useful quantification of average Parkinson’sdisease symptom severity, J. R. Soc. Interface, 08:842–855.

Xue,W. 2014. Data mining based on SPSS modeler (Second Edition), Beijing: China Renmin University Press.

Wang, J. C. 2001. Logistic regression model method and application, Beijing: Higher education press.

Zhang, H.B., Zhou, W.Z. 2019. IBM SPSS modeler 18.0 authoritative data mining guide, Beijing: People's post and Telecommunications Pres