A Novel Scoring Method for Stroke Order Based on Choquet Integral with Fuzzy Measure

GUEY-SHYA CHEN
Graduate Institute of Educational Measurement
Taichung University
No. 140, Ming-Sheng Rd., Taichung, 40306
TAIWAN

YU-DU JHENG
Graduate Institute of Educational Measurement
Taichung University
No. 140, Ming-Sheng Rd., Taichung, 40306
TAIWAN

HSIANG-CHUAN LIU
Department of Bioinformatics
Asia University
No. 500, Lioufeng Rd., Wufeng , Taichung County, 41354
TAIWAN

SHIH-YAT CHEN
Department of Mathematics Education
Taichung University
No. 140, Ming-Sheng Rd., Taichung, 40306
TAIWAN

Abstract: - The purpose of this research is to develop a novel algorithm based on Choquet integral with fuzzy measure for scoring stroke order of Chinese Character writing. The learning of stroke order is very important for a fixed stroke writing character system therefore this research based on the theory of "five indexing system of Chinese characters" and combine the features of strokes develops a computer-based stroke order assessment system to real-time examine learners' stroke order and score their performances. Except traditional additive scoring model, an non-additive scoring approach based on Choquet integral with λ -measure and L -measure is proposed. A real data set with 65 samples is examined and experiment results show that the performance of scoring model based on Choquet integral with fuzzy measures outperform the performances of traditional scoring methods.

Key-Words: - stroke order, computer-based assessment, five indexing systems of Chinese characters, λ -measure, L -measure, Choquet integral

1 Introduction

The learning of stroke order is very important for a fixed stroke writing character system, for example Chinese characters and there is an average of 12 strokes in each traditional Chinese character. Many teaching tools for Chinese characters writing are developed [1], [2] but few researchers have interest in developing system for the assessment of writing

The work described in this paper is sponsored by the National Science Council under Grant NSC 95-2520-S-142-003-MY3.

Chinese characters. Therefore we develop a computer-based assessment for stroke order [3] and include its scoring model.

Instead of typing on a keyboard, students can input the characters through a hand writing computer input device then the system records the movement of each stroke and adapt to the writing speed for different students. By breaking each character into different strokes and then fetching important features of each stroke, an algorithm for recognizing stroke order and 4 scoring methods are developed.

It is complex to score the stroke order because many various latent factors are embedded in the interaction between strokes for each character. A traditional linear additive scoring model is simple but it does not fit the need. It does not consider different characters have the same partial part, for example characters "十" and "土" both have the same "十" part. Therefore a fuzzy scoring model based on Choquet Integral is proposed in this research. Four different scoring methods are examined with a real data set.

This paper was organized as followings. Feature extraction and stoke order recognition algorithm were introduced in section 2. In section 3, fuzzy measures were reviewed and Choquet integral was described in section 4. Four scoring methods were described in section 5. Experiment and result were described in section 6 and final section is for conclusions.

2 Stroke Order Recognition and Feature Extraction

The stroke order recognition algorithm is based on both the theory of "five indexing system of Chinese characters" proposed by Chen Lifu [4], and the rules of stroke order published by the Ministry of Education in Taiwan [5].

2.1 Five Indexing System of Chinese Characters [3], [4]

The theory of 'five indexing systems of Chinese characters', which uses quadrant as a way to distinguish different strokes is shown in Figure 1. There are only four quadrants in this approach so it may lead to unreliable outputs. In this research, we use slope to analyze strokes characteristics.

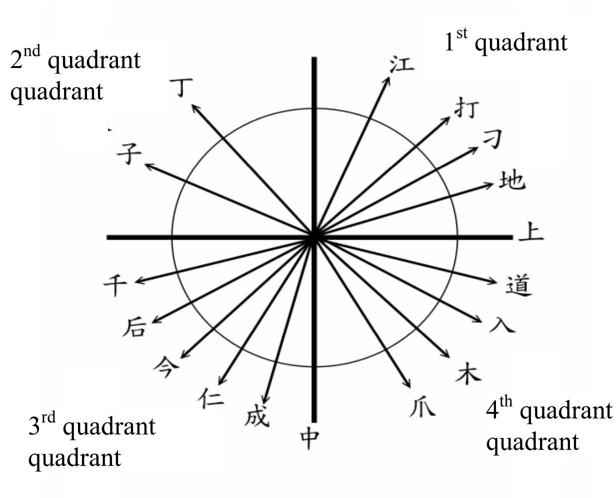

Figure 1 Basic stroke diagram of 'five indexing systems of Chinese characters'

2.2 Basic Stroke Detection [3]

Based on "five indexing system of Chinese characters", we use slope instead of quadrants in this research. The system takes diă stroke and nà stroke as nà stroke because of the similarity between the two slopes and it omits huà stroke because of its specialty. The six categories of basic strokes used in Chinese characters writing assessment system are shown in Table1.

Table 1 Six basic strokes in the system

number	stroke name	shape	corresponding angle
1	héng	一	0°~ 25°
2	tī	㇀	25°~ 80°
3	gōu	㇔	90°~ 150°
4	piě	丿	195°~ 255°
5	shù	丨	265°~ 275°
6	nà	㇏	295°~ 335°

Athough diă stroke and nà stroke are considered as the same category, the system will still category them as diă stroke or nà stroke in the output result according to their length. Basically, every stroke can be broken up into these six basic strokes categories this research has proposed.

2.3 Compound Stroke Processing

After distinguishing the six basic strokes, the system uses combination to deal with the other complex strokes. Example: When the system detects stroke "∟

", it automatically breaks it as two of the basic strokes shù and héng, therefore naming it shù-héng. Using this method, we can break up the 23 compound strokes, announced by the ministry of education, into all of the six basic strokes. All the compound strokes and their corresponding basic strokes are shown in table 2.

2.4 Overlapping/Complementary Interaction

There are some overlapping interactions among tradition Chinese characters, for example "大" and "天" both two characters have the same part of "大". And there are many same parts which are "交", "木", "人" among characters "校", "佼", "休".

The writing order of character "川" is from left-hand side to the right-hand side. The order of "仁" is from left-hand side to the right-hand side then from top to bottom. There are many writing order rules for traditional Chinese characters. For example from outside to inside for character "月", stroke shù after stroke héng for character "士", stroke nà after stroke piě for character "入". There are some latent complementary interactions within ordering rules among characters.

Table 2 23 Compound strokes

nuumber	shape	compound stroke	example
1		héng-gōu	也、家
2		héng-piě	友、又
3		héng-shù	五、真
4		héng- piě-héng-gōu	吃、乙
5		héng-shù-gōu	的、雨
6		héng-shù-héng	朵、凹
7		héng-shù-nà-gōu	九、熱
8		héng-piě-héng-shù-gōu	乃、秀
9		shù–héng	叫、忙
10		shù-tī	長、滾
11		shù-gōu	小、可
12		shù-piě	青、麼
13		shù-héng-gōu	北、包
14		shù-héng-shù	吳、誤
15		shù-héng-shù-gōu	第、號
16		piě-tī	紅、級
17		piě-héng	錄、碌
18		piě-nà	經、安
19		nà	襄、裏
20		nà-nà	不、美
21		nà-héng-gōu	心、想
22		nà-shù-gōu	手、狗
23		nà-nà-gōu	我、戲

3 Fuzzy Measure

Two well known fuzzy measures, λ -measure, L -measure, are reviewed as follows.

3.1 Fuzzy Measures [6], [7], [8]

A fuzzy measure μ on a finite set X is a set function $g_\mu : 2^X \to [0,1]$ satisfying the following axioms:

(i) $g_\mu(\phi) = 0$, $g_\mu(X) = 1$ (boundary conditions)(1)

(ii) $A \subseteq B \Rightarrow g_\mu(A) \le g_\mu(B)$ (monotonicity) (2)

3.2 Singleton Measure

The singleton measure s of a fuzzy measure μ on a finite set X is a function $s : X \to [0,1]$ satisfying

$$s(x) = \mu(\{x\}), x \in X \tag{3}$$

3.3 λ -measure [9]

For given singleton measure s, let $\lambda \in (-1,\infty)$, a λ -measure, g_λ is a fuzzy measure on a finite set X , $|X| = n$, satisfying

(i) $A, B \in 2^X, A \cap B = \phi, A \cup B \ne X$

$$\Rightarrow g_\lambda(A \cup B) = g_\lambda(A) + g_\lambda(B) + \lambda g_\lambda(A) g_\lambda(B) \tag{4}$$

(ii) $$\prod_{i=1}^{n}[1 + \lambda s(x_i)] = \lambda + 1 > 0, s(x_i) = g_\lambda(\{x_i\}) \tag{5}$$

Note that since $\lambda \in (-1,\infty)$, the above equation of variable λ with degree n has only one solution, and the solution is always not a closed form.

3.4 L -measure [10]

For given singleton measure s, A L-measure, g_L, is a fuzzy measure on a finite set X, $|X| = n$, satisfying

(i) $L \in [0, \infty)$

(ii) $s(x) = g_L(\{x\}), x \in X$ (6)

(iii) $^\forall A \subset X, n - |A| + (|A|-1)L > 0$

$$\Rightarrow g_L(A) = \max_{x \in A}[s(x)] + \frac{(|A|-1)L\sum_{x\in A} s(x)}{[n-|A|+(|A|-1)L]\sum_{x\in X} s(x)}\left[1 - \max_{x\in A}[s(x)]\right] \quad (7)$$

Note that (i) Since $L \in [0, \infty)$, L-measure have infinitely many solutions with closed form, (ii) When $L = 0$，L-measure is just a P-measure, (iii) The L-measure is more sensitive than the λ-measure.

4 Scoring Model for Stroke Order

There are four scoring models for stroke order discussed in this paper. The first two methods are traditional additive scoring models and the others are based on fuzzy scoring models.

4.1 Traditional Additive Scoring Model

The first method is generally used in a traditional paper-pen based character writing test. If all the strokes in a characters are correct, give one point otherwise there is zero point. For example, there are four characters “小”, “生”, “同”, “找” in a writing test and 4-point is obtained if a student can write the stroke order of these four characters correctly.

Because there are different number of strokes in each Chinese character a polytomous model is used for scoring each character. There are 3-stroke for “小”, 5-stroke for “生”, 6-stroke for “同”, and 7-stroke for “找” and 21-point is obtained if a student can write the stroke order of these four characters correctly. But a student write “小”and “同”correctly and write 2-stroke of “生” and 3-stroke of “找” correctly, 14-point is obtained (3+2+6+3=14).

4.2 Fuzzy Scoring Model Based on Choquet Integral

Let μ be a fuzzy measure on a finite set X. The Choquet integral [11] of $f: X \to R_+$ with respect to μ is denoted by

$$\int_C f d\mu = \sum_{j=1}^{n}\left[f\left(x_{(j)}\right) - f\left(x_{(j-1)}\right)\right] g_\mu\left(A_{(j)}\right) \quad (8)$$

where $f\left(x_{(0)}\right) = 0$, $f\left(x_{(j)}\right)$ indicates that the indices have been permuted so that

$$0 \le f\left(x_{(1)}\right) \le f\left(x_{(2)}\right) \le \ldots \le f\left(x_{(n)}\right) \quad (9)$$

$$A_{(i)} = \left\{x_{(i)}, x_{(i+1)}, \ldots, x_{(n)}\right\}, A_{(n+1)} = \phi \quad (10)$$

And

(i) if $\mu = \lambda$, Equation (8) is a Choquet integral base on λ-measure scoring model.

(ii) if $\mu = L$, Equation (8) is a Choquet integral base on L-measure scoring model.

The following example is used to explain our proposed scoring method based on Choquet integral with L-measure. We suppose that $f(x_1) = 6, f(x_2) = 7, f(x_3) = 8, f(x_4) = 9$, are known and their weight measure value are as followed.

$g(\{x_1\}) = 0.1$, $g(\{x_2\}) = 0.2$, $g(\{x_3\}) = 0.3$, $g(\{x_4\}) = 0.4$

And from the definition of fuzzy measure, it is known

$g(\phi) = 0$，$g(\{x_1, x_2, x_3, x_4\}) = 1$

Considering fuzzy measure of a join event, we obtain

$$g_L(A_1) = 1 \quad g_L(A_2) = 0.6818 \quad g_L(A_3) = 0.3714 \quad g_L(A_4) = 0.1 \quad (12)$$

Therefore the value of Choquet integral is

$$C_L(f) = f(x_1)g_L(A_1) + (f(x_2) - f(x_1))g_L(A_2) + (f(x_3) - f(x_2))g_L(A_3) + (f(x_4) - f(x_3))g_L(A_4) \quad (13)$$

$= 6\times1 + (7-6)\times0.6818 + (8-7)\times0.3714 + (9-8)\times0.1$
$= 7.1532$

5 Experiment and Result

A real data set with 65 samples from a primary school in Taiwan and a Chinese school in Australia including 16 independent variables, examination scores of 16 Chinese characters[1], and one dependent

[1] 三竹同土交小生母回米找白木也送山

variable, the overall score of character writing tests at school is applied to evaluate the performances of two traditional additive scoring models and two Fuzzy Choquet integral scoring models based on λ -measure, L -measure (L =5, L =10) to compute the mean square error (MSE) of the dependent variable. The formula of MSE is

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2 \tag{14}$$

Responding the ratio of the credit hour for 16 Chinese characters, all of the fuzzy measures about the independent variables are assigned the same singleton measures as follows,

$$\left\{g_\mu\left(\{x_1\}\right), g_\mu\left(\{x_2\}\right), \ldots, g_\mu\left(\{x_{16}\}\right)\right\}$$
$$= \{0.04, 0.08, 0.08, 0.04, 0.08, 0.04, 0.06, 0.06, 0.08, 0.08, 0.09, 0.06, 0.05, 0.04, 0.09, 0.04\},$$
$$\mu = \lambda,\ L=5,\ L=10. \tag{15}$$

When the singleton measures of any fuzzy measure μ are given, all the event measures of μ can be found, and then, the Choquet integral based on μ can be computed. The fuzzy measures L, has infinitely many solutions of fuzzy measure can be selecting and we choose the values as $L = 5$ ， $L = 10$, and exploit proposed four forecasting models computing their MSE value. The experimental results of all forecasting models are listed in Table 3.

In Table 3, we found the following results,

(i) Choquet Fuzzy scoring model outperforms tradition additive scoring model.

(ii) Choquet Fuzzy scoring using L -measure outperforms model using λ -measure.

(iii) Choquet Fuzzy scoring model based on L -measure with L =5 and L =10 have similar performance.

(iv) Among traditional scoring the performance of polytomous model is better than that of dichotomous model.

Table3 MSE of four scoring models

scoring model	approach	MSE
traditional additive	dichotomous model	14.8618
	polytomous model	5.2336
Choquet integral	λ - measure	3.9290
	L - measure(L=5)	3.4555
	L - measure(L=10)	3.3438

6 Conclusion and Future Works

In this research, we develop a computer-based assessment system and propose a new scoring model based on Choquet integral with fuzzy measure for stroke order of Chinese character writing. An educational data experiment is conducted for comparing the performance of 4 approaches: 2 traditional additive approaches and 2 fuzzy scoring approaches based on Choquet integral with λ -measure and L -measure. Both fuzzy scoring models have better performance and traditional dichotomous model has the worst result.

We will continue to improve the performance of stroke order scoring method using other fuzzy measures. Furthermore, the system can also be adapted to the character set of other countries which have fixed strokes for each character.

References:

[1] N. Takesue, K. Mochida, A. Kitadai, M. Nakagawa, A handwriting-based Kanji learning system enabling teachers to designate evaluation points, *ISPJ SIG Technical Report*, Vol.2005, No.15, 2005, pp. 15-22.

[2] K. T. Sun and C. I. Wang, An intelligent tutoring system for teaching the stroke rrder of Chinese Characters, *The Sixth International Conference for Advancement of Computing in Education*, Beijing, China, 1998.

[3] G.-S. Chen, Y.-D. Jheng and L.-F. Lin, Computer-based assessment for the stroke order of Chinese characters writing, *The Second International Conference on Innovative Computing, Information and Control* (ICICIC2007) , Kumamoto, Japan on September 5-7, 2007.

[4] L.F. Chen, *Five Indexing System of Chinese Character*, Research center of Chinese, Taipei, 1971.

[5] Ministry of Education, *The Stroke Order Manual for Common Standard Chinese Characters*, 1999.

[6] G. A. Shafer, *Mathematical Theory of Evidence*, Princeton University Press, Princeton, New Jersey, 1976.

[7] Z. Wang and G. J. Klir, *Fuzzy measure theory*, Plenum Press, New York, 1992.

[8] A. P. Dempster, Upper and lower probabilities induced by multi-valued mapping. Annals of Mathematical Statistics, 38, 325-339. 1967.

[9] M. Sugeno, *Theory of fuzzy integrals and its applications*, PhD thesis, Tokyo Institute of Technology, Tokyo, Japan, 1974.

[10] H.-C. Liu, Y.-D. Jheng, W.-C. Lin, and G.-S. Chen, A novel fuzzy Mmeasure and its Choquet integral regression model, *International Conference on Machine Learning and Cybernetics* (ICMLC2007), 2007.

[11] G. Choquet, Theory of capacities, *Annales de l'Institut Fourier*, Vol.5, 1953, pp. 131-295.