

基于书写过程监测的汉字书写自动教学方法

谢建斌,戴 永,李文涛

(湘潭大学智能计算与信息处理教育部重点实验室,湖南 湘潭 411105)

摘 要:为及时纠正低龄用户在汉字书写过程中产生的典型错误,提出一种改进的汉字书写自动教学方法。构建由主导笔顺和约束笔顺组成的教学知识,通过表格化方法实现基于教学知识的用户书写过程监测。设计笔向环节跟踪笔画,利用传统汉字笔画错误识别方法分析笔画关系,建立错交、错离机制监督逆跨书写,并设置部件定制标志指导部件关系。实验结果表明,改进方法不仅能实现笔画、笔画关系、错交、错离、部件关系有错即改的自动教学,而且监测正确率达 93% 以上。

关键词:书写过程;汉字书写;自动教学;错交与错离;部件定制

中文引用格式:谢建斌,戴 永,李文涛. 基于书写过程监测的汉字书写自动教学方法[J]. 计算机工程,2016,42(1):187-192.

英文引用格式:Xie Jianbin,Dai Yong, Li Wentao. Automatic Teaching Method of Chinese Character Writing Based on Writing Process Monitoring[J]. Computer Engineering,2016,42(1):187-192.

Automatic Teaching Method of Chinese Character Writing Based on Writing Process Monitoring

XIE Jianbin, DAI Yong, LI Wentao

(Key Laboratory of Intelligent Computing and Information Processing, Ministry of Education, Xiangtan University, Xiangtan, Hunan 411105, China)

【Abstract】In order to correct the classical errors made by young users in writing Chinese character, this paper puts forward an improved automatic teaching method for writing Chinese character. This method constructs teaching knowledge composed by dominant and binding stroke order, uses the teaching knowledge to monitor and manages the writing process through tabulation. It designs the link chain of strokes to track the stroke, uses traditional ways of identifying stroke errors to analyze the stroke relations, builds the mechanism of wrong intersections and separation to monitor the reverse writing and sets the part marks to instruct the part relations. Experimental result shows that the improved method can realize the automatic error-correct teaching of strokes, stroke relations, wrong intersections and separation, and parts relations. The correct rate of monitoring is up to 93% .

【Key words】writing process; Chinese character writing; automatic teaching; wrong intersection and wrong separation; part customization

DOI:10.3969/j.issn.1000-3428.2016.01.033

1 概述

计算机汉字书写教学方法可分为描红和自由式书写两大类。当前人们主要研究自由式书写自动教学,文献[1-2]研究中文字手写教学系统的自动错误检测算法,但未能给出部件分析策略。文献[3]提出一种联机手写汉字书写学习支持服务系统评价方法。文献[4]将汉字智能工具中的书写错误识别技术应

用于汉字书写教学。现有汉字书写教学方法基本采用整字分析法,即先记录笔画,待整字笔画书写完毕再进行分析与质量评价。该方法存在不能及时制止用户的错误书写行为,允许用户一错再错的问题。文字书写教学系统主要面向低龄用户,低龄时代的认知与行为内容很容易被铭记,教学过程应尽可能多地使低龄学生面对正确知识,做到有错即改。基于这一认识,本文提出基于书写过程监测的汉字书

基金项目:湖南省教育厅基金资助项目(13C914);湖南省“十二五”重点学科建设基金资助项目。

作者简介:谢建斌(1987-),男,硕士研究生,主研方向为知识处理、人工智能;戴 永,教授;李文涛,硕士研究生。

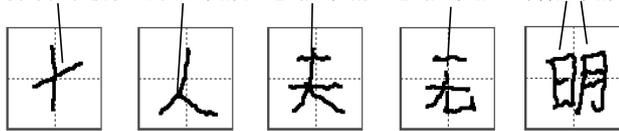
收稿日期:2014-12-18 **修回日期:**2015-01-13 **E-mail:**261491428@qq.com

写实时自动教学方法。教学知识由笔画编码、笔画关系、部件编码、错交笔顺、错离笔顺等子字段组成。将教学知识监测原理表格化,建立以直向和折向为主要内容的笔画整体形态分析环节结构。应用已有方法分析笔画关系,逆跨非紧邻笔画分析错交、错离关系,设置关系尾码 x 分析部件关系,遇错则实时向用户提出教学意见。

2 汉字书写结构分析

汉字的基本笔画类型有 28 种^[5],笔画的结构特点表现为:(1)无闭环笔径;(2)以直线笔径为主(含变向结构的笔径,从折点分割形成的笔段仍是结构稳定的直线笔径);(3)笔迹的行笔方向规范(除钩笔产生反向笔迹外,其余笔迹均为非反向地从左至右,从上到下);(4)关键点产生的笔迹形态少且规范(除端点与左、右钩极点外,其余关键点均由右折和左下折产生,前者占 11 种笔画,后者占 10 种)。汉字的笔画关系有 7 类 169 种^[6],在书写过程中表现出来的特点为:(1)紧邻笔画相同,字不同则笔画搭配的空间位置不同(如“十”与“正”);(2)紧邻笔画不同,字不同但笔画搭配的空间位置可能相同(如“不”与“正”);(3)笔画空间位置种类繁多、微分区域复杂、教学精确度要求高。关于部件及其关系,不同应用目标有不同的描述与处理方法,本文以笔画关系是否可计算为依据描述部件关系,特点为:(1)2 个部件空间位置关系不受各部件所含笔画数目影响(如“你”与“引”);(2)部件是笔画书写过程中的定制形态,与传统的偏旁部首无必然关联,但传统部首关系起主导作用。受主观和客观因素影响,如身心不够成熟、坐姿不端正、握笔不稳、书写屏材质不同等,初学汉字的人容易出现书写错误,错误形态多种多样,图 1 是一位一年级学生在触摸屏上书写汉字的 5 种典型错误截图。汉字书写教学内容不但要把好笔画及紧邻笔画关系指导关,更应着重于错交、错离及部件关系的监督与告诫。

横画书写过斜 T 字交关系偏下 笔画关系错交 笔画关系错离 部首关系偏下



(a)笔画错误 (b)笔画关系错误 (c)错交错误 (d)错离错误 (e)部件错误

图 1 触摸屏上的汉字书写样本

3 教学知识构建

教学知识由主导笔顺和约束笔顺组成,一字一条。主导笔顺指完成该汉字书写、人们公认的笔画顺序;约束笔顺指人们在按主导笔顺进行笔画书写时,当前写出笔画不但要满足与紧邻已写笔画的空

间关系,还应同时满足与其他笔画的空间关系。约束笔顺分为错交与错离 2 类,错交指当前写出笔画与非紧邻已写笔画形成错误的十字交关系;错离指当前写出笔画与非紧邻已写笔画未形成教学知识规定的空间关系。文字 W 书写过程教学知识 $Q(W)$ 表示为:

$$Q(W) = (w_{11}w_{12}\cdots w_{1n_1})(w_{21}w_{22}\cdots w_{2n_2})\cdots \\ (w_{c1}w_{c2}\cdots w_{c n_c})(r_{11}^*r_{12}^*\cdots r_{1(n_1-1)}^*r_{11}^{\rho_1}) \\ (r_{21}^*r_{22}^*\cdots r_{2(n_2-1)}^*r_{22}^{\rho_2})\cdots (r_{c1}^*r_{c2}^*\cdots r_{c(n_c-1)}^*\Phi) \\ |(W_{EL}, W_{EC}) = [w_{11}w_{12}\cdots w_{1n_1}w_{21}w_{22}\cdots w_{2n_2}\cdots \\ w_{c1}w_{c2}\cdots w_{c n_c}][r_{11}^*r_{12}^*\cdots r_{1(n_1-1)}^*r_{11}^{\rho_1}r_{21}^*r_{22}^*\cdots r_{2(n_2-1)}^* \\ r_{22}^{\rho_2}\cdots r_{c1}^*r_{c2}^*\cdots r_{c(n_c-1)}^*\Phi][(w_{i_1}, w_{j_1}) \\ (w_{i_2}, w_{j_2})\cdots (w_{i_m}, w_{j_m})][(w_{u_1}, w_{v_1}, r_{q_1}^*, \xi_1) \\ (w_{u_2}, w_{v_2}, r_{q_2}^*, \xi_2)\cdots (w_{u_q}, w_{v_q}, r_{q_q}^*, \xi_q)] \quad (1)$$

其中, w, r 依次表示笔画与关系; r^*, ρ 依次表示笔画关系与部件关系。设文字 W 由 k 条笔画组成,即 $W = \{w_1, w_2, \dots, w_k\}$ 。式(1)中每对方括弧中的内容为一个字段,表明 $Q(W)$ 由 4 个字段组成,第 1 字段、第 2 字段(笔画与主导笔顺笔画关系字段)是按定制部件划分的子主导笔顺, w_{11} 是 $w_1, w_{c n_c}$ 是 w_k , 其余类推, $n_1 + n_2 + \dots + n_c = k, w_{11}$ 是书写过程跟踪的开始, $w_{c n_c}$ 是书写过程跟踪的结束, Φ 为主导笔顺结束标志;第 3 字段为错交笔顺字段,用 W_{EL} 表示;第 4 字段为错离笔顺字段,用 W_{EC} 表示; W_{EL}, W_{EC} 不受部件影响,只作用于已写笔画。对于 $w_i, w_j, i, j \in \{1, 2, \dots, k\}$, 若 $i = j + 1$, 称 w_i, w_j 为主导笔顺书写过程中的紧邻笔画;若 $i > j + 1$, 且 w_i, w_j 为预测的错误交叉关系, 则称 (w_i, w_j) 为错交笔顺元素;若 $i > j + 1$, 且预测的 w_i, w_j 会偏离应具备的 r^*, ξ 为偏离度分析系数或畸变系数, 则称 (w_i, w_j, r^*, ξ) 为错离笔顺元素。

4 汉字书写自动教学方法

笔迹信息的采集与前置处理采用文献[7]提供的方法。用 P 表示实时采集到的笔迹点二维坐标向量, $P = [p_1, p_2, \dots, p_n] = [(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)]$, X 方向的坐标序列向量为 $P_x = [x_1, x_2, \dots, x_n]$ 、 Y 方向的坐标序列向量为 $P_y = [y_1, y_2, \dots, y_n]$ 。通过 $S[k, L]$ 保存书写正确笔画的笔迹点,用以分析各类关系。

4.1 $Q(W)$ 监测机制及笔画与笔画关系书写教学

表 1 给出 $Q(W)$ 对用户汉字书写过程的监测原理。先写笔画为第 1 条~第 $(k-1)$ 条,后写笔画为第 2 条~第 k 条。每条笔画采用 2 种下标表示,即按所属部件序号确定下标和按整体笔画数目、笔序确定下标(见括号中笔画符号), $k = n_1 + n_2 + \dots$

+ n_c 。用户选择被练习文字,系统的知识库提供被练习文字书写过程的知识模板。用户每写完一条笔画,即起笔信息触发系统进入笔画、笔画关系、错交关系、错离关系等分析。同一部件中的后写笔画与紧邻的先写笔画为主导笔顺,两笔画之间通过相应的子关系分析模块决策,如表中 w_{12} 和 w_{13} 为第 1 部件的第 2 笔画、第 3 笔画,两笔画之间关系由 r_{12}^* 子模

块分析。 w_{1n_1}, w_{21} 分别为相邻部件的前部件最后一条笔画和紧邻后部件第 1 条笔画,遇此种情况,表明书写过程进入换部件书写状态。 W_{EL}, W_{EC} 分析与部件无关,所有后写笔画都有可能与非紧邻主导笔顺的先写笔画产生错交、错离,表中 W_{EL}, W_{EC} 排序越靠后,分析工作量越大,但后写笔画与非紧邻主导笔顺笔画不一定存在 W_{EL}, W_{EC} 。

表 1 教学知识 $Q(W)$ 对汉字书写过程的监测原理

后写笔画	先写笔画												
	w_{11} (w_1)	w_{12} (w_2)	...	$w_{1(n_1-1)}$ (w_{n_1-1})	w_{1n_1} (w_{n_1})	w_{21} (w_{n_1+1})	...	w_{C1} (w_{k-n_c+1})	w_{C2} (w_{k-n_c+2})	w_{C3} (w_{k-n_c+3})	...	$w_{C(n_c-2)}$ (w_{k-2})	$w_{C(n_c-1)}$ (w_{k-1})
w_{12} (w_2)	r_{11}^*												
w_{13} (w_3)	W_{EL} W_{EC}	r_{12}^*											
...										
w_{1n_1} (w_{n_1})	W_{EL} W_{EC}	W_{EL} W_{EC}	...	$r_{1(n_1-1)}^*$									
w_{21} (w_{n_1+1})	W_{EL} W_{EC}	W_{EL} W_{EC}	...	W_{EL} W_{EC}	$r_{11}^{\rho 1}$								
w_{22} (w_{n_1+2})	W_{EL} W_{EC}	W_{EL} W_{EC}	...	W_{EL} W_{EC}	W_{EL} W_{EC}	r_{21}^*							
...						
w_{C2} (w_{k-n_c+2})	W_{EL} W_{EC}	W_{EL} W_{EC}	...	W_{EL} W_{EC}	W_{EL} W_{EC}	W_{EL} W_{EC}	...	r_{C1}^*					
w_{C3} (w_{k-n_c+3})	W_{EL} W_{EC}	W_{EL} W_{EC}	...	W_{EL} W_{EC}	W_{EL} W_{EC}	W_{EL} W_{EC}	...	W_{EL} W_{EC}	r_{C2}^*				
w_{C4} (w_{k-n_c+4})	W_{EL} W_{EC}	W_{EL} W_{EC}	...	W_{EL} W_{EC}	W_{EL} W_{EC}	W_{EL} W_{EC}	...	W_{EL} W_{EC}	W_{EL} W_{EC}	r_{C3}^*			
...
w_{Cn_c} (w_k)	W_{EL} W_{EC}	W_{EL} W_{EC}	...	W_{EL} W_{EC}	W_{EL} W_{EC}	W_{EL} W_{EC}	...	W_{EL} W_{EC}	W_{EL} W_{EC}	W_{EL} W_{EC}	...	W_{EL} W_{EC}	$r_{C(n_c-1)}^*$

笔画书写过程分析采用子模块制,28 种笔画分别对应 28 个分析模块,模块编号分别为 01,02, ..., 28。根据汉字笔画结构特点,监测内容定为笔迹方向和笔画整体结构正误。笔迹方向分为直向和折向两大类,直向集合选择{左→右,上→下,右→左下斜,左→右下斜,左→右上斜},折向集合元素选择{左钩折,左折,下折,右折,左上钩折,右上钩折}。2 个集合的元素联合编码的集合为{00,01, ..., 11, 12},其中,“00”,“12”分别为起点、终点编码;“01”对应“左→右”;“11”对应“右上钩折”;其余类推。直向分析内容为起始与末尾笔段方向,笔画整体结构分析主要进行折向次数计算及折点引申笔段方向的正确性分析。针对不同的错误属性向用户提供对应的指导意见。28 种汉字笔画书写过程整体形态的主要分析环节结构如表 2 所示。

表 2 笔画整体形态的分析环节结构

笔画	编码	分析环节结构	笔画	编码	分析环节结构
、	01,02	04 OR 03	ㄥ	29,30	01070907
一	03,04	01	冫	31,32	0108
...
冫	27,28	0106	ㄥ	55,56	010809

在表 2 中,折向编码后加“B”表示为该折点的第 2 种形态,以 δ 表示笔画规模粒度,取 $\delta = 2$,即笔画取短、长 2 个粒度,因此,每条笔画对应连续的 2 个编码,奇数编码对应短笔画,偶数编码对应长笔画,“01,02”笔画对应“01”号分析模块,“03,04”笔画对应“02”号分析模块,依次直至“55,56”笔画对应“28”号分析模块。

汉字笔画书写过程的基本分析步骤为:(1)提取当前应检测的 w_{ij} ($i = 1, 2, \dots, c, j = 1, 2, \dots, n_i$);

(2)依据所设笔画规模粒度 δ 值计算 w_{ij} 对应的分析模块编号;(3)进入 w_{ij} 分析子模块;(4)获取 P ;(5)预处理 P ;(6)分析起始与末尾笔段方向的正确性;(7)计算整体笔画的折向元素数目,分析数目的正确性;(8)依据分析环节结构有序分析折向的正确性;(9)所有分析环节正确,将 P 按序存入 $S[k,L]$ 。

点笔画、小粒度且实写笔迹点数目符合给定要求等不进行起始与末尾笔段方向分析。任何分析环节出错通过文字、图案或语音等方式给用户提出对应的指导意见。

设置左右、上下、十字交等七大类笔画关系,按3_3制空间划分, r^* 中的 $*$ = $B_1B_2B_3$, B_1 为大类编码(占2位), B_2 为大区统一编码(占2位), B_3 为小区编码(占1位),169种子关系,建立七大类对应的分析子模块。主导笔顺分析时按 Q 中 $r_{i(j-1)}^*$,即按第 i 部件的第 j 条与第 $j-1$ 条 w 之间的笔画关系编码进入分析模块。笔画关系分析方法与手写汉字识别技术中的笔画空间关系识别方法相似^[8]。 w_{i1} ($2 \leq i < c$)写完不进行笔画关系分析,笔画分析正确之后记录 w_{i1} 。当第 j 条与第 $j-1$ 条 w 之间的关系不满足 $r_{i(j-1)}^*$ 时,系统给用户提出相关指导意见。

4.2 错交与错离关系教学

由 $Q(W)$ 监测表结构可知,通过 $S[k,L]$ 进行两错(错交、错离)分析有2种方法:(1) W 的 k 条笔画全部写完,然后按照 $Q(W)$ 监测原理进行分析,提出相应的教学意见,该方法类似文献[1-2]方法;(2)从 w_3 开始每写完一条笔画就进行相应的两错分析,遇到当前笔画与已写笔画出现两错状态,立即提出教学意见,并根据教学严格度对用户后续书写行为进行约束,最严格度为禁止继续书写后续笔画。第(2)种方法可实现即错即改,避免一错再错或不知所错之处,适用于低龄用户学习书写汉字。本文系统实现第(2)种方法。两错分析与部件无关,按 $W = \{w_1, w_2, \dots, w_k\}$ 中元素进行排序, $Q(W)$ 表中为非部件分类的笔画标识符。两错分析从 w_3 开始,第1轮分析 w_3 与 w_1 之间是否有预测的 W_{EL} , W_{EC} 元素存在,如果有,依据 W_{EL} , W_{EC} 元素结构进入两错分析,未出现两错进入 w_4 的书写监测,发现错误则提出相应的教学意见。依次检测直至 w_k 。由此可知, w_3 只需进行一轮 W_{EL} , W_{EC} 元素分析,而 w_k 需进行 $k-2$ 轮 W_{EL} , W_{EC} 元素分析。虽不是每一轮 W_{EL} , W_{EC} 都会存在分析元素,但总体来说笔画越多、结构越复杂的文字,越写到后面,系统工作量越大。尽管如此,由于检测与分析环节是随书过程有序递进的,因此程序复杂度不会增加,也不会影响教学的针对性及意见的客观性。

两错分析的具体内容仍是笔画关系分析,在查询到 (w_i, w_j) 或 (w_i, w_j, r^*, ξ) 后同样可借鉴文献[9]中方法实现分析与教学。

4.3 部件关系教学

部件关系设置三大类,分别是左右、上下、半包围等。左右、上下按3_3制空间划分,共有22种子关系;半包围设置7种子关系。建立三大类分析子模块。各部件的子主导笔顺跟踪结束,说明用户书写笔画结构、笔画关系正确,且无错交、错离现象,最后分析内容是借助主导笔顺跟踪生成的 $S[k,L]$ 和关系字段中 r_i^p 分析写出 W 所具有的部件关系的正确性。 r_i^p 含3个字段,即 $\rho = U_1U_2x$,其中, U_1 为部件类码,通常接续笔画关系类赋码; U_2 为空间细分码; x 表示当前写出的部件将与其后面的 x 个部件形成 U_1U_2x 所标识的部件关系。当在关系字段中找到 U_1U_2x 时,基本分析方法为:(1) $x=1$,后续紧邻部件为同层部件,进行紧邻部件关系分析;(2) $x>1$,进入新层部件的首部件,根据该层各部件的 U_1U_2x ,对当前写完的2个紧邻部件进行关系分析;继续在该层进行同层部件关系分析,该层部件分析完成后,将该层部件作为一个整体与父部件进行关系分析。不管是同层,还是父子层,只要关系分析结果不满足 U_1U_2 ,系统向用户提供教学意见。

4.4 教学过程主体结构

以多文种融合文字书写教学系统^[10]中汉字书写教学为例。汉字书写自动教学过程分为2个环节:(1)练习文字准备,包括系统获取练习文字所处的文种、年级、课号及字序等编码参数,练习方式参数,练习格式参数,每字书写次数参数等,按照获得的参数在隐喻文字练习簿进行备写场景布局;(2)根据文种、年级、课号及字序等编码参数获取当前被练习文字的文字书写过程教学知识 $Q(W)$,在当前书写格式中利用 $Q(W)$ 监测与督促用户书写汉字的过程。

汉字书写自动教书步骤具体如下:

Step1 将 $Q(W)$ 的4个字段分别存于对应的向量数组中。

Step2 获取 P ,并利用文献[8]方法进行前置处理。

Step3 利用笔画向量数组元素进行笔画书写过程教学,将书写正确笔画的 P 有序存入 $S[k,L]$ 。

Step4 利用笔画关系向量数组元素及 $S[k,L]$ 进行笔画关系书写过程教学。

Step5 当笔画关系向量数组中出现 r_i^p 时,不分析当前笔画与前一笔画的笔画关系。

Step6 利用两错向量数组元素及 $S[k,L]$ 进行两错分析与教学。

Step7 若 k 条笔画未书写完,则转Step2;否则继续。

Step8 利用 $S[k,L]$ 进行部件关系的书写教学。

Step9 本次作业未写完,取下一个 W 的 $Q(W)$ 并转Step1。

Step10 结束。

以上教学步骤正确,说明用户对练习文字的骨架书写基本正确,文字书写的质量由文献[11]方法进行专门评价。

5 实验及效果分析

本文实验平台与文献[7,9,11]所述实验平台相同。采用文献[6]提供的笔画、笔画关系、部件关系编码,取 $\delta = 2, 28$ 条笔画产生 56 个编码,编码范围为 01 ~ 59;笔画关系编码范围为 600 ~ 64122;部件关系编码范围为 7000x ~ 7212x。

图 2 为“话”字的书写自动教学示例。

$$\begin{aligned}
 Q(\text{话}) &= (01,50)(07,04,06)(05,31,03) \\
 &\quad (610,70002)(610,622,71001) \\
 &\quad (63100,64122,0000) | ((3,1)(4,2) \\
 &\quad (5,3)(8,2))((8,6,63122,3)) \\
 &= [01,50,07,04,06,05,31,03][610, \\
 &\quad 70002,610,622,71001,63100,64122, \\
 &\quad 0000][(3,1)(4,2)(5,3)(8,2)] \\
 &\quad [(8,6,63122,3)] \quad (2)
 \end{aligned}$$

图 2(a)为练习内容取自文献[12]下册的第一课内容,图 2(b)为选中“话”字进行练习的截图,图 2(c)为“话”字点阵图,表 3 为“话”书写正确后的 $S[k,L]$ 。

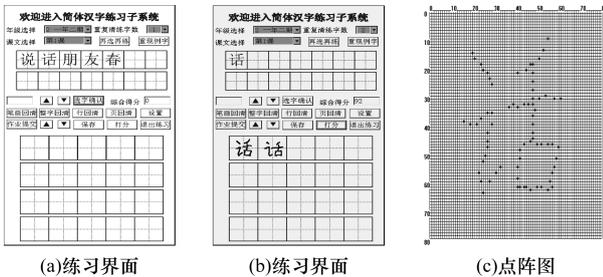


图 2 “话”字书写示例

表 3 “话”字笔画容器 $S[k,L]$

笔画	笔画坐标序列
$S[1]$	(20,15) (21,17) (24,19) (25,21) (27,22) (28,24) (29,26)
$S[2]$	(16,39) (19,40) (22,40) (24,38) (27,36) (29,36) (29,40) (27,43) (27,46) (27,48) (26,51) (25,53) (23,57) (24,60) (25,64) (28,60) (31,57) (33,55)
$S[3]$	(55,10) (53,14) (51,17) (48,19) (45,21) (41,22)
$S[4]$	(37,34) (39,33) (41,34) (44,33) (46,33) (48,33) (51,31) (54,30) (58,31) (61,31)
$S[5]$	(47,19) (48,22) (48,25) (48,28) (48,31) (48,33) (49,35) (48,37) (48,40) (48,43) (48,45)
$S[6]$	(49,42) (52,41) (55,41) (59,41) (62,44)
$S[7]$	(44,47) (47,46) (50,47) (52,47) (55,47) (57,47) (60,48) (59,52) (58,55) (57,57) (56,60)
$S[8]$	(42,62) (45,62) (47,63) (50,62) (52,62) (55,63) (57,62)

本课要求学生掌握 5 个生字,每字重复练习次数为 2 次。“话”字书写监督过程:书写完第 1 笔“讠”,取 $Q(\text{话})$ 笔画字段中第 1 个笔画代码,即“01”除 2 四舍五入取整,进入以该值为编码的“讠”笔画分析子模块,如果书写不符合要求,子模块输出相应的指导意见,反之笔迹点数据存入 $S[k,L]$ 第 1 行,系统进入第 2 笔监督;第 2 笔书写结束经分析正确,笔迹点数据存入 $S[k,L]$ 第 2 行,根据 $Q(\text{话})$ 的笔画关系字段的第 1 个笔画关系代码“610”进入到以 R610 为标识的笔画关系分析子模块,如果笔画关系不符合要求,输出相应的指导意见,反之系统进入第 3 笔监督;第 3 笔书写结束经分析正确,笔迹点数据存入 $S[k,L]$ 第 3 行,因为 $Q(\text{话})$ 的笔画关系字段的第 2 个笔画关系代码“70002”为部件关系代码,所以系统不分析笔画关系;从第 3 笔开始,系统进行两错分析,分析第 3 笔与第 1 笔之间是否有预测的错交、错离元素存在,根据 $Q(\text{话})$ 的错交关系字段的第 1 对错交序偶进行错交关系分析,第 3 条笔画与第 1 条笔画没有相交,因此没有形成错交关系; $Q(\text{话})$ 的错离关系字段没有第 3 笔画,第 1 笔画的错离预测四元标识结构,系统进入第 4 笔监督,如此直至最后一条笔画。“话”字由“讠”、“干”、“口”3 个部件组成,在关系字段中找到 $r_1^{p1}(70002)$,因为 $x > 1$,先不进行部件分析,进入新的部件层,找到 $r_2^{p2}(71001)$,此时 $x = 1$,利用上下最值法分析前后 2 个紧邻的部件关系,部件“干”和部件“口”满足全幅上下关系,新层部件分析完成后,将 2 个子部件“干”和“口”作为一个整体与父部件“讠”进行部件关系分析。方法为左右最值法,2 个部件满足全幅左右关系,整字教学指导结束,整字书写正确,并给出了单字的综合得分“92”。

图 3 是一年级学生的实际书写教学效果截图,每字重复练习次数为 1 次。分别是笔画错误、笔画关系错误、笔画错交错误、笔画错离错误、部件错误的示例,针对每种错误给出相应的指导意见。图 3(b)为图 3(a)中“九”字的点阵示意图,图 3(a)中前 2 个字书写练习正确,第 3 个“九”字的第 2 条笔画“乙”的关键点应为 3,而截图所写笔画的关键点数为 2,因此系统给出“横折弯钩书写错误,请重新书写”的提示。图 3(c) ~ 图 3(f)为 4 类错误及系统给出指导意见的实例。弹出文字窗口显示指导意见的同时,通过语音同步播放^[12]。实验文字来自义务教育课程标准实验教科书语文一年级上下册^[12],由 5 位不同的低龄用户书写,每个用户随机选取其中的 20 个汉字书写,每个字书写 3 次,共 300 个实验样本。表 4 给出了笔画、笔画关系、错交、错离、部件关系的监测正确率。

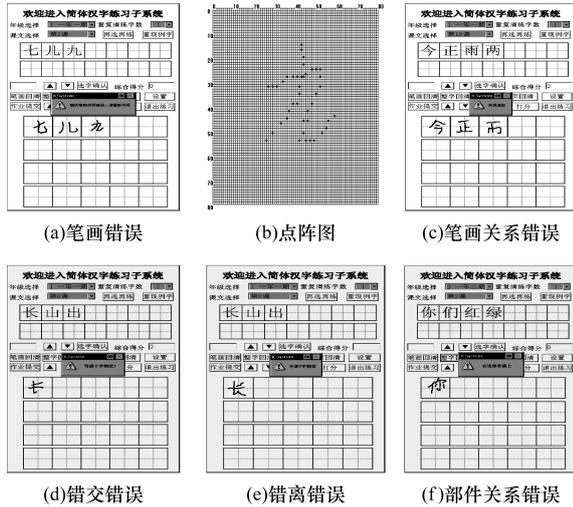


图 3 小学一年级学生实际书写效果

表 4 汉字书写教学的监测正确率 %

用户	笔画	笔画关系	错交	错离	部件关系
1	98.1	96.9	94.5	93.5	96.8
2	97.6	97.3	93.7	94.7	97.4
3	98.4	95.7	95.1	94.1	96.7
4	98.3	97.4	94.6	93.6	97.1
5	97.2	96.7	94.2	95.2	96.2

文献[2]提出用属性关系图进行汉字书写教学,节点表示文字的笔画,边表示文字的笔画关系,K-means 聚类算法对笔画空间关系进行细化处理。这种描述空间关系的方法考虑了笔画间的相对距离,使得笔画空间关系有了更详细的定义,可全面地描述文字结构。使用细化的空间关系,能指出人们难以发现的笔画之间存在的关系错误。该方法针对文字的笔画空间关系错误进行分析,而且监测正确率指标符合实用要求,但该方法是在整字书写完毕才指出错误所在,无法让用户在第一时间得知错误及见错就改。图 4 表明,本文方法在笔画与笔画关系的监测正确率方面优于文献[2]方法,而且其他 3 项仅为本文方法所具有。

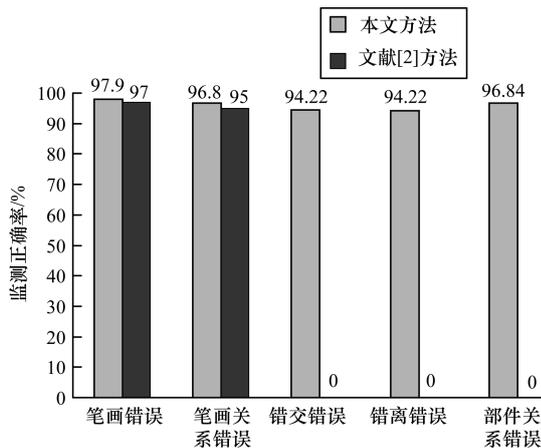


图 4 监测正确率对比

6 结束语

本文从汉字笔画、笔画关系、部件关系在书写过程中的特点出发,提出实现见错就改的教学方法并取得如下成果:(1)给出面向汉字书写过程自动教学的教学知识结构,并利用表格化方法建立实现机制;(2)设计笔向环节链跟踪笔画,建立错交、错离机制监督逆跨书写及部件定制标志指导部件关系;(3)应用传统方法实现笔画关系分析。本文方法可推广到其他文种,如英文、韩文等,从而实现外文文字书写的自动教学。下一步工作将利用动态规划方法对自动教学系统进行优化,进一步提高监测正确率。

参考文献

- [1] Hu Zhihui, Xu Yun, Huang Liusheng. A Chinese Handwriting Education System with Automatic Error Detection [J]. Journal of Software, 2009, 4 (2): 101-107.
- [2] 胡智慧. 汉字智能工具中的书写错误识别技术研究与应用[D]. 合肥:中国科学技术大学,2010.
- [3] Yamaguchi T, Muranaka N, Tokumaru M. Evaluation of Online Handwritten Characters for Penmanship Learning Support System[C]//Proceedings of HCI'11. Orlando, USA; Human-Computer Interaction, 2011:121-130.
- [4] Tan C K. An Algorithm for Online Strokes Verification of Chinese Characters Using Discrete Features [C]// Proceedings of the 8th International Workshop on Frontiers in Handwriting Recognition. Washington D. C., USA: IEEE Press, 2002:339-344.
- [5] 中国社会科学院语言研究所. 新华字典[M]. 11 版. 北京:商务印书馆,2011.
- [6] 吴佑寿,丁晓青. 汉字识别——原理、方法与实现[M]. 北京:高等教育出版社,1992.
- [7] 戴 永,张维静,孙广武. 基于文字书写教学的笔迹噪声处理[J]. 计算机工程与应用, 2014, 50 (14): 164-167.
- [8] Liu C L, Jaeger S, Nakagawa M. Online Recognition of Chinese Characters: The State-of-the-art[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26(2): 198-213.
- [9] 戴 永,刘任任,王求真,等. 可联网交互的多功能规定格式习字系统及方法:中国, ZL201010149767. 2[P]. 2010-09-01.
- [10] 喻世东,戴 永,王求真,等. 适用于文字书写教学教室系统的嵌入式局域网协议[J]. 计算机工程, 2014, 40(9): 284-290.
- [11] 孙广武,戴 永,喻世东,等. 音素关联的多文种语音融合编码方法[J]. 计算机工程与应用, 2013, 49(19): 217-221.
- [12] 课程教材研究所,小学语文课程教材研究开发中心. 义务教育课程标准实验教科书语文一年级[M]. 北京:人民教育出版社,2004.