

基于时间分割的手写输入系统的用户绩效模型^{*}

吴昌旭¹ 杨群会¹ 张侃¹ 胡永革² 杨磊²

(¹中国科学院心理研究所,北京 100101) (²英特尔中国研究院,北京 100000)

摘要 为定量估计与提高基于时间分割的手写输入系统用户绩效,推导了用户绩效的静态与动态模型。实验一检验与修正了这两个模型,并获得整合模型。实验二证明该整合模型能对不同训练水平用户的绩效有较高的解释率。应用该整合模型发现:当识别正确率 RA、输入字数 N 为定值时,修改时间 T、单字手写时间 WT、分割时间 ST、识别时间 R 这四个因素对任务完成时间 D 的影响大小次序为 $WT = ST = R > T$;当这四个因素为定值时,RA 每提高 1%,D 至少节省 1000ms。

关键词 数学心理模型,手写汉字输入系统,绩效模型,时间分割。

分类号 B849:TB18

1 引言

由于中文字符的符号特性与便携式电子设备的发展,手写输入成为当今人机交互的新的手段与研究热点^[1~4]。研究表明,手写输入系统的人机交互绩效是这些系统成功的关键^[5,6]。而分析与提高这一交互绩效的基础是建立用户绩效模型。本研究将对基于时间分割(通过字与字之间的时间间隔来分割汉字)的手写系统的用户绩效进行建模。

理论上,系统绩效往往包括任务完成时间与错误率。Casali 和 Williges(1990)对语音输入系统的研究将任务完成时间与未要求修改的错误量作为绩效的指标^[7]。在本研究中,因为被试必须修改所有错误识别的汉字,错误识别的字数越多,被试完成整个任务的时间就越长,所以任务完成时间本身已包含了错误量这一因素,因而在本研究中仅将任务完成时间(D)作为系统绩效的指标。

2 研究构思

本研究首先在假设各因素相互独立、不相互影响的条件下,建立影响基于时间分割的手写系统各因素的静态模型;然后在以往的研究与理论基础上,建立能描述各因素之间动态直接与间接影响的通径分析的理论模型;通过实验一检验静态模型、修正动

态模型;然后合并两模型,目的是通过动态模型中变量间可能的因果关系,减少静态模型中的变量数目;最后,通过实验二,在实际商用软件与不同训练水平的个体的实验条件下,检验这一整合模型与实际绩效的拟合程度。

3 用户绩效模型的推导与建立

3.1 静态用户绩效模型的推导

3.1.1 无错误识别条件下静态用户绩效模型的推导 Ben Shneiderman (1984)曾提出人机交互中系统事件与用户事件环状交互模式^[8]。Steven L T 等人将这一模型进一步深化,提出了存在系统操作时间与用户操作时间的交互模型(见图 1)^[9]。

系统事件包括系统的响应时间等(见图 1)。在本研究中,因为中文汉字输入系统消耗在显示与锁定的时间小于 10ms,故这里不考虑系统这部分响应时间。这里的响应时间包括分割时间(ST, Segmentation Time)与识别时间(R, Recognition Time)。

这一模型中的用户事件包括:(1)用户思考与计划时间:指从用户触发系统交互到有用户开始进行下一个任务的时间间隔;(2)用户延迟时间(UD):指系统开始等待用户输入下一任务到用户实际开始输入下一任务的时间间隔;(3)用户操作时间:用户操作任务的时间,在中文汉字输入系统中对应用户的

收稿日期:2002-12-16

^{*} 英特尔中国研究院(Intel China Research Center)资助;教育部十五规划课题(课题号:FB011067),中国科学院院长基金资助。

通讯作者:张侃, email: zhangk@psych.ac.cn

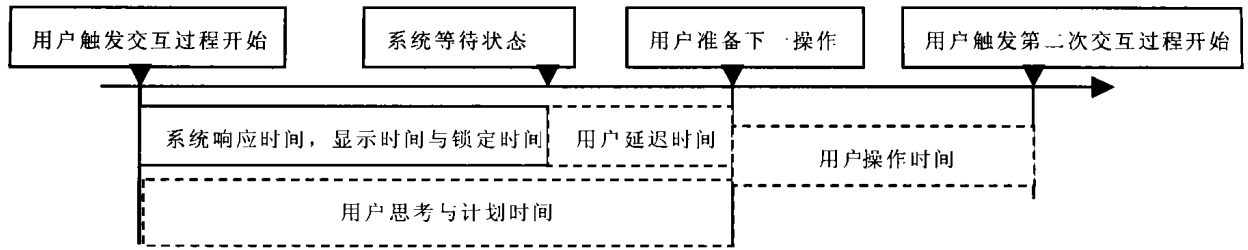


图1 人机交互中的用户事件与系统事件

单字手写时间(WT)。

因此,根据认知心理学中的开窗原理与反应时加法原则^[10],在无识别错误的情况下,用户使用手写系统抄写一个句子(包含N个汉字)的任务完成时间(D),参照图1,可以描述为:

$$D = \sum_{i=1}^N (WT + ST + R + UD) + Error \quad (1)$$

对于用户延迟时间(UD),这一过程实际是用户判读屏幕所呈现的被识别汉字正确与否的过程,根据张侃(1991)提出的汉字键盘输入认知模型^[23],手写输入用户,无论是听写、看写(本研究实验任务)还是想写,读取下一个待输入汉字都会经历知觉加工过程;高定国(1995)对汉字的认知速度研究发现,高频少笔画汉字单字判读(见字读音)反应时均值为439ms,标准差为48ms^[11];因为本研究实验材料也为高频少笔画汉字,所以将这一均值作为用户读取下一个待输入汉字的估计时间。即 $UD = 439ms$ 。

这样,基于公式(1),如果 Average(WT)表示平均单字手写时间,用户在无错误识别的情况下完成整句句子(包含N个汉字)抄写的交互绩效可进一步描述为:

$$D = N(Average(WT) + ST + R + UD_{stable}) = N(Average(WT) + ST + R + 439) + Error \quad (2)$$

3.1.2 有错误识别条件下静态用户绩效模型的推导 由于无法完全正确识别用户所有的手写汉字(尤其在连笔与草体书写情况下),因此必须对错误识别条件下静态交互绩效模型加以研究。

在无错误识别的基础上,添加修改错误的过程(如图2)。这一过程包括五个子过程:过程 按“回格键”;过程 重新手写原字(手写汉字);过程 , 与输入过程相同。为便于通过实验验证研究模型的正确性,设 T(如图2)表示从屏幕呈现错误识别汉字到用户开始第二遍手写原汉字之间的时间间隔。

据此分析,可以将这一条件下的手写汉字过程分解为两大部分:

- (1) 手写输入部分: 手写输入 N 个汉字耗时: $N(WT + ST + R + UD)$ (3)
- (2) 修改部分: 设用户输入 N 个汉字并及时修改,其中一共进行 x 次修改操作 ($x \leq N, x \geq 0$)

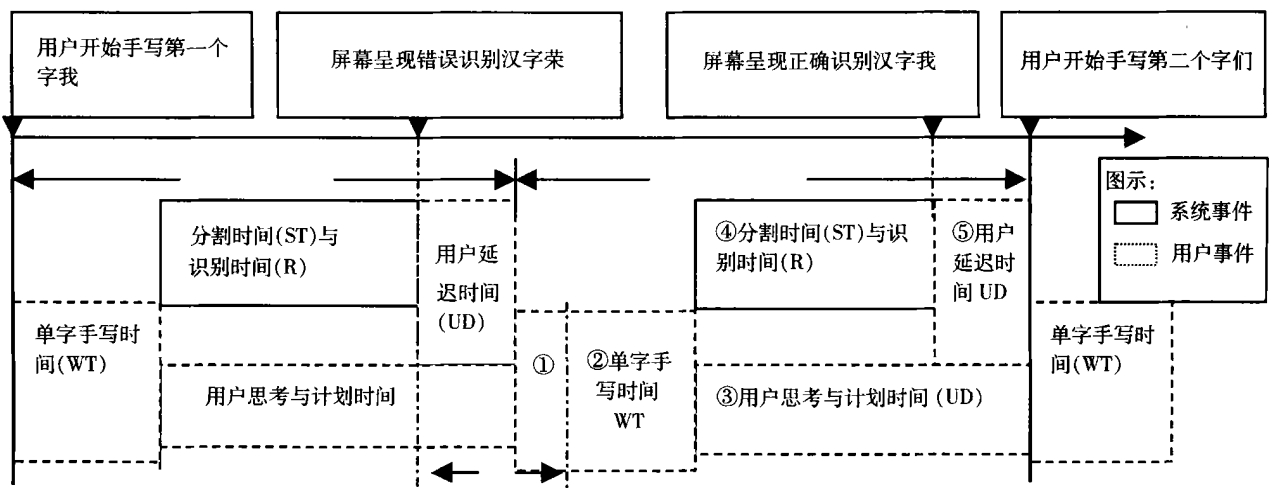


图2 基于时间分割的手写汉字输入系统用户事件与系统事件

$$RA = N / (N + x) \Rightarrow x = N[(1/RA) - 1]$$

其中一共会进行 $N[(1/RA) - 1]$ 次修改操作

由图 2, 每次修改过程耗时为 $(T - UD) + WT + ST + R + UD$

N 个汉字并及时修改, x 次修改过程总耗时为:

$$N[(1/RA) - 1][(T - UD) + WT + ST + R + UD] \quad (4)$$

有识别错误条件下, 合并(3), (4)式, 整个任务完成时间为:

$$\text{Duration} = N(WT + ST + R + UD) + N[(1/RA) - 1][(T - UD) + WT + ST + R + UD] + \text{Error} \quad (5)$$

$$= N(WT + ST + R + 439) + N[(1/RA) - 1][(T - 439) + WT + ST + R + 439] + \text{Error} \quad (6)$$

3.2 动态用户绩效模型的建立

已有对手写输入系统及其它人机系统的研究表明, 分割时间、识别正确率、修改时间、单字手写时间四个因素之间存在着动态的直接与间接的因果影响关系:

其一, 随着响应时间的变化, 用户会及时调整自己的工作风格^[12]。用户会加快或放慢手写速度, 其单字手写时间也可能会随之缩短或增长(图 3 中路径 A);

其二, Frankish 和 Hull (1995) 认为, 一般而言, 用户在操作过程中能根据手写系统给予的反馈对手写系统形成心理表征^[13]。因此, 如果用户发现正确识别率降低, 那么他们可能会知觉到系统的识别能

力不够强, 因而放慢手写速度来适应系统的变化(图 3 中路径 B);

其三, 根据动作控制理论^[14], 对于一些基本的动作反应, 如简单反应时、动作速度、手指灵活程度等, 存在着一种较微弱的一般动作能力 (general motor ability)。即这些简单的动作反应之间会有一定程度上的相互影响。因为手写过程主要与个体的动作速度与手指灵活程度有关, 所以在一定程度上可以推论, 随着单字手写时间的缩短, 动作速度与手指灵活程度的提高, 被试的简单反应时与动作速度也会有相应的降低与提高。而简单反应时与动作速度又是修改过程的主要部分, 前者对应于用户延迟时间, 后者对应按“回格键”操作。因此, 随着单字手写时间的降低修改时间 (T) 也会有所下降(图 3 中路径 C);

其四, 基于 Steven (1992) 的人机交互模型(见图 1)与手写系统的用户事件与系统事件模型(见图 2), 分割时间(路径 D)、修改时间(路径 E)、及单字手写时间(路径 G)构成了任务完成时间的主要部分。这些时间的延长, 将直接导致任务完成时间的延长。Mackenzie 和 Chang (1999) 的实验表明, 用户的单字手写时间影响了文本输入的速度^[1](图 3 中路径 G)。另外, 在考虑了单位修改时间(每次修改的时间长度)对绩效的影响后, 修改发生的频率(识别正确率)的高低也影响了交互绩效(路径 F)。

综上所述, 因素之间动态的影响模型可描述为图 3 所示。

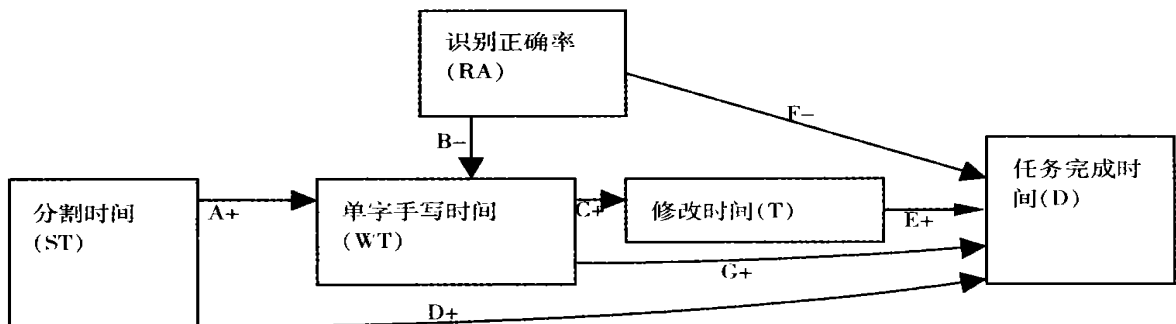


图 3 预测的因素之间动态绩效交互模型

注: + 表示两因素之间是线性递增函数; - 表示两因素之间是线性递减函数。

4 实验一 对静态与动态绩效模型的验证与修正

4.1 实验方法

4.1.1 被试 选取 12 名 19~21 岁大学二年级学

生参加实验。所有被试矫正视力正常, 右利手。所有被试以前未使用过任何手写输入设备与系统, 来自非统计与心理专业, 不熟悉统计学知识。对被试使用计算机与文本编辑系统的已有经验与使用时间进行记录。

4.1.2 实验设备 主机:内存 523,080 KB RAM, 主频 650 兆,显示器刷新频率 80Hz, 分辨率 1024 × 768VGA; 手写硬件设备:“MengQue Pen” Model: ET-0405-U 输入系统(手写板、识别系统)一套; 实验手写界面与手写速度测试软件;采用 IntelPen 汉字识别软件及其为实验专门开发的手写速度测试插件。

4.1.3 实验材料 在心理学百科全书上随机抽取心理统计学的术语 18 条,略加修改后保证字数相同(每句 24 个汉字),总笔画数基本一致(均值 179 笔/句,标准差 11 笔/句),均为高频少笔画汉字。通过预实验挑选出 6 句,在将单字手写时间作为协变量时,10 名非心理或统计专业的三年级大学生对这 6 句话的抄写时间无显著差异。

4.1.4 实验设计与程序 由于采用回归分析与路径分析,而非方差分析,实验只采用被试内单因素设计。自变量为 6 个水平时间长度的汉字分割时间,其余因素为软件纪录(其中有路径分析的内源变量与外源变量)。实验材料及自变量的呈现采用拉丁方顺序。

自变量为 6 个水平时间长度的汉字分割时间:190ms, 440ms, 740ms, 1040ms, 1340ms, 1640ms。最小值(190ms)由预实验确定;最大值(1640ms)参考了 Yormans (1983)的研究结果,用户能接受的系统的最长的等待时间为 2 秒^[12]。

实验的具体程序:(1)介绍手写系统的使用,被试熟悉手写输入与修改的操作;(2)修改操作的练习:要求被试将左手食指放在计算机的退格键上,当出现错误识别的汉字时,立即按退格键将它抹去然后修改;当被试按退格键修改的时间小于 2000ms 时,进入下一实验阶段;(3)指导语的呈现:要求被试以最快速度抄写屏幕上方的句子,抄写中可使用自己平时的书写字体,出现识别错误后,立即修改;(4)正式实验:采用被试内设计,每个被试共经历 6 次尝试(trials),分割时间的每个水平按拉丁方排序各出现一次。被试每抄写完一句就休息 3 分钟。

4.2 实验结果

4.2.1 对静态绩效模型的验证 因为实验一中 $N = 24$, 所以待验证的静态模型为:

$$24(WT + ST + R + 439) + 24[(1/RA) - 1][(T - 439) + WT + ST + R + 439] + \text{Error}$$

(1) 使用线性回归考察模型估计的任务完成时间对实际任务完成时间的解释量

将 x -轴作为模型估计的时间, y -轴作为实际

任务完成时间,建立线性回归,考察模型估计的任务完成时间对实际任务完成时间的解释量(见图 4)。

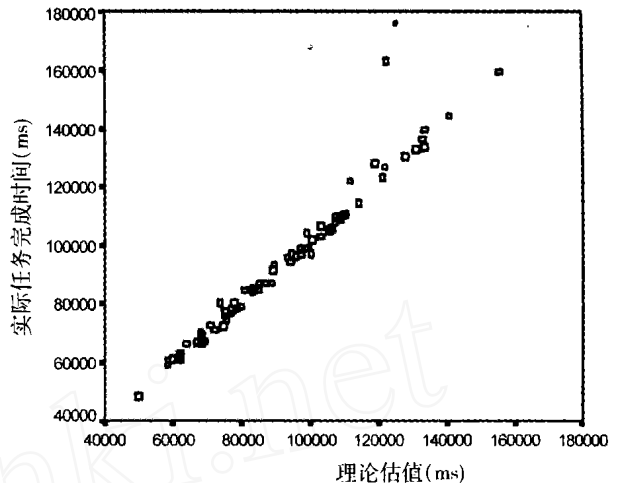


图 4 实际任务完成时间与模型估计时间的散点图

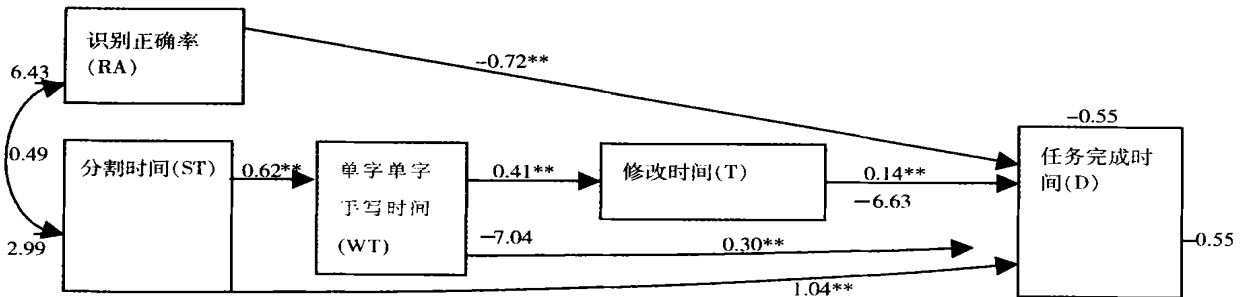
实际任务完成时间与模型估计时间之间存在着显著的线性相关(非标准化回归系数 $B = 1.07$, $t = 40.37$, $p < 0.001$);并且 $\text{adjusted } R^2 = 0.96$ 说明模型估计时间能够解释 96% 的实际任务完成时间的变异。而回归方程的常数项可作为原静态模型的常数修正项,得:

$$D = 24(WT + ST + R + 439) + 24[(1/RA) - 1][(T - 439) + WT + ST + R + 439] - 4973 \quad (7)$$

考察这一模型对实际数据的解释率是否随汉字认知速度均值(用户延迟时间 UD)的变化而出现较大变化:当用户延迟时间在两个标准差范围($439 \pm 48 \times 2$)ms 波动时,adjusted R^2 未有变化($\text{user delay} = 535\text{ms}$: adjusted $R^2 = 0.96$; $\text{user delay} = 343$, adjusted $R^2 = 0.96$)。

(2) 实际任务完成时间与模型估计的任务完成时间的差异检验

因为实际任务完成时间与模型估计的任务完成时间均服从正态分布(Kolmogorov - Smirnov Value = 0.092 for observed duration, $df = 69$, $p > 0.05$; Kolmogorov - Smirnov Value = 0.082 for observed duration, $df = 72$, $p > 0.05$),故采用独立样本的 t 检验考察实际任务完成时间是否与模型估计的任务完成时间有显著差异: $t = 0.415$ ($df = 139$) ($p = 0.679 > 0.05$) 及 Levene's Test for Equality of Variances ($F = 0.576$, $p > 0.05$),表明实际任务完成时间与模型估计的任务完成时间无显著差异。

图5 动态绩效模型的路径分析图 (** $p < 0.01$)

4.2.2 对动态绩效模型的验证与修正 因为研究的各个因素之间存在着直接与间接的因果关系,但又不存在潜变量(对各因素的因素分析 $KMO < 0.5$; Bartlett Test of Sphericity $p < 0.001$)^[15],所以我们通过路径分析的方法验证与修正动态绩效模型^[16]。Guerrier 和 Manivannan (1999)已在人机工程领域成功地采用这种研究方法^[17]。这一方法最大的优点在于它能够同时考察多个因素直接与间接的因果关系,在统计控制某些难以用实验控制的变量的情况下,考察量之间的关系^[15]。对原理论的动态绩效模型的验证与修正见图5,在修正过程中,根据 Lisrel 8.3 给出的一个修正建议,将识别正确率与单字手写时间之间不显著的路径(原假设中路径B)筛去。

这一路径模型的整体拟合值 Goodness of Fit Index (GFI) 为 0.981, adjusted Goodness of Fit Index (AGFI) 为 0.904, Root Mean Square Error of Approximation (RMSEA) = 0.0376 < 0.05, 以及 P-Value for Test of Close Fit (RMSEA < 0.05) = 0.423 > 0.05, 表明整体上模型对数据的拟合度较好^[15]。但模型的 Parsimony Goodness of Fit Index (PGFI) = 0.196; Parsimony Normed Fit Index (PNFI) = 0.295 说明模型在描述变量的关系上还不够简洁。

相应的非标准化回归方程为:

$$T = 0.816WT - 51.94 \quad (\text{regression coefficient: } p < 0.001, R^2 = 0.224) \quad (8)$$

$$WT = 0.123ST + 1377.38 \quad (\text{regression coefficient: } p < 0.01, R^2 = 0.06) \quad (9)$$

$$D = -90784RA + 30.24ST + 39.53WT + 2.798T + 70649.44 \quad (\text{regression coefficient: } p < 0.001, R^2 = 0.939) \quad (10)$$

4.3 实验一 讨论

静态绩效模型能够很好地解释实际任务的完成时间(解释率达到 96%),并且使用不同的汉字认知

速度均值(343 ~ 535ms)对模型的解释率影响不大。由于高频汉字占据了 99% 的日常交流的文本内容^[18],这也表明静态绩效模型能够有效地运用到各种日常文本的手写输入系统中。但是,在实际手写过程中,这些因素在相互影响并发生动态的变化,因此可以利用它们的相互影响与决定关系(动态模型),减少静态模型中估计实际任务完成时间的所需的预测变量个数。

筛去一条路径后的原动态绩效模型能较好地拟合实验一的数据(AGFI = 0.904)。尽管它对数据的整体拟合程度较高,但是,(1)它无法替代静态模型,因为动态模型在实验前只是对各因素间关系的定性描述,实验后得到的回归方程也只是一个经验模型;以往对手写识别系统及语音识别系统建立的绩效模型一般也都是采用实验数据的直接线性回归^[1,19,20],通过经验数据建立的回归方程作为绩效模型。它们的缺陷就在于实验数据的得出必须基于一个特定的界面,也就不能任意修改其中的某些变元与常量来推广到其它界面中。通过理论建立的静态模型能够在一定程度上克服这一缺点(实验二的数据支持了这一论断,见本文 6.1)。(2)从回归方程(9)看,单字手写时间(WT)与分割时间(ST)之间的线性回归虽然显著,但分割时间对单字手写时间的解释率非常小($R^2 = 0.06$),表明单字手写时间这一变量在一定程度上独立于分割时间。(3)识别正确率与单字手写时间之间的路径(path B)未被数据支持,表明单字手写时间也独立于识别正确率。综合(2)、(3),由于模型中没有其它变量解释单字手写时间的变化,因此可以认为单字手写时间在这一模型中是一个外源变量;这一结果与 Wolf (1990)对手写系统的研究有一致之处,他也发现识别系统的特性改变并不会影响用户的手写风格(writing styles)^[21],也就在程度上说明用户的单字手写时间主要由个体差异决定,与系统识别正确率关系不大。

为达到减少静态模型中预测变量个数,动态模型中公式(8)整合到静态绩效模型公式(7):

$$D = 24 (WT + ST + R + 439) + 24 [(1/RA) - 1] [(T - 439) + WT + ST + R + 439] - 4973 \quad (7)$$

$$T = 0.816WT - 51.94 \quad (8)$$

$$\Rightarrow D = 24 (WT + ST + R + 439) + 24 [(1/RA) - 1] [(0.816WT - 51.94 - 439) + WT + ST + R + 439] - 4973$$

将模型推广到输入 N 个汉字的条件,而且修正项以 $(N/24)$ 增加:

$$D = N (WT + ST + R + 439) + N [(1/RA) - 1] [(0.816WT - 51.94 - 439) + WT + ST + R + 439]^{3/4} (N/24) 4973 \quad (11)$$

但是,(8)式的 WT 对 T 的解释率偏低 ($R^2 = 0.224$),使用这一方程替代静态模型中的参数 T 仍需要实验进行考察;另外,由于实验一修正了静态与动态模型,因此有必要对修正与合并后的模型再加以实验检验,考察它在实际识别软件与不同训练水平的用户绩效数据的拟合程度;最后,还需要考察推导、整合后的模型(11)是否比从单纯从实验数据获得的经验回归方程(10)更能有效地解释另一实验的数据。

5 实验二

5.1 实验方法

5.1.1 实验设计与被试 采用单因素被试间设计。自变量为被试的不同训练水平:(1)无训练组(Naive user):选取从未使用过手写输入系统的大学生7人,实验中无训练过程。(2)有训练组。按训练的新近程度,又分为:(a)过期训练组(Outdated-experienced user):6名12个月前曾参加过实验一的被试,在这12个月中未再次使用手写系统;(b)新近训练组(Fresh-experienced user):选取从未使用过手写输入系统的大学生6人,在实验二中接受一定的手写输入练习(对于无训练组与有训练组的划分的有效性见本实验结果5.2.2)。这三组被试年龄在19~24岁,矫正视力正常,均是右利手,来自非统计与心理专业,不熟悉统计学知识,实验过程中均采用右利手进行实验。

5.1.2 实验设备 主机:联想昭阳8100Cfe笔记本电脑一台,主频PIII 750Hz,内存128M,14.1英寸TFT LCD液晶显示器,分辨率1024×768 XGA。手写硬件设备:汉王99手写板、手写笔

一套,型号EM III 6045, Power Source:5VDC。通过串行接口与主机相连。商用手写界面与软件设置:采用Microsoft拼音2.0自带的手写识别系统;通过界面设置固定它的汉字分割时间(ST)为1秒,加上它的识别时间(R)为100ms;整个(ST+R)时间为1100ms。要求被试只在左框内书写。

手写速度测试软件与方法:采用研究一中的可纪录手写速度的IntelPen汉字识别软件。为了避免手写字的大小对手写速度测试的影响,将印有微软手写界面的透明胶片粘贴在屏幕上,要求被试仅在胶片所印的界面的左框中进行手写输入。

5.1.3 实验程序与实验材料 在训练正式实验前首先测量被试的手写速度(实验材料为实验一中材料的第一句话),之后立即进入不同的实验程序:(1)无训练组:立即进行正式实验,介绍实验的手写系统后(无被试尝试),要求被试以最快速度抄写屏幕上方的句子,抄写中可使用自己平时的书写字体,出现识别错误后,立即按退格键将它抹去后重写;为避免正式实验本身的练习效应,正式实验只使用材料的第二句话,且只写一遍;由主试纪录被试在书写过程中的错误及完成任务时间。(2)过期训练组:实验程序同无训练组。(3)新近训练组:立即进行手写输入训练,介绍实验的手写系统后,要求他们尝试输入一段含有46个汉字的段落,直至每个被试不发生因操作不熟练而引起的出错(训练时间控制在3~5分钟),进入正式实验,其正式实验的材料与程序同前两组。

5.2 实验结果

5.2.1 三组被试实际任务完成时间与模型估计时间的比较 因为组内被试数量较少,故采用非参数检验与相关分析考察整合模型对实验数据的拟合程度(见表1)。

5.2.2 三组不同训练水平的被试的组间比较 单因素方差分析表明,三组不同训练水平的被试在完成任务时间上显著差异, $F(2,18) = 3.791, p < 0.05$;进一步的简单效应分析(Post Hoc LSD)发现差异主要来源于无训练组在任务完成时间上明显长于其它两组($p < 0.05$)。这也说明本实验开始对被试的分组是有效的。

计算被试在完成任务时间与模型估计时间之间的差值,将这一差值作为因变量,单因素方差分析表明,三组不同训练水平的被试在这一差值上没有显著差异, $F(2,18) = 1.778, p > 0.05$;说明模型解释不同训练水平用户的完成任务时间上有一致性。

表 1 三组被试实际任务完成时间与模型估计时间的比较

组别	描述统计结果						实际任务完成时间与模型估计时间的比较			
	实际任务完成时间		正确识别率		单字手写时间		非参数检验 (Wald-Wolfowitz Test)		Pearson 相关	
	<i>M</i> (ms)	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i> (ms/字)	<i>SD</i>	<i>Z</i> 值	<i>p</i> (1-tailed)	相关系数 <i>R</i>	决定系数 <i>R</i> ²
无训练组	151286	44271	0.592	0.22	1814	280	1.39	0.922 > 0.05	0.98 **	0.96
有训练组										
过期训练组	107833	16666	0.824	0.10	1607	229	-1.51	0.067 > 0.05	0.91 *	0.82
新近训练组	132769	41000	0.692	0.19	1784	252	0	0.522 > 0.05	0.86 *	0.73

注: ** $p < 0.01$; * $p < 0.05$

5.3 实验二 讨论

非参数统计与 Pearson 相关分析的结果表明整合模型(11)仍能在常用商业识别软件中应用,而且对不同训练水平的用户绩效有较高的解释能力($R^2 > 0.7$)。尤其对于无训练组,该模型对实际数据的解释率达到了 96%。

另外,模型对实际数据的解释率随着训练的新近程度的增加而下降。其原因可能在于人对动作技能的学习存在着较大的个体差异^[22],训练过程中个体对训练学习的个体差异带来的数据变异的增大可能是导致解释率下降的主要原因。

6 对整合模型的进一步解释与应用

6.1 整合模型与简单回归方程(实验一)的比较

同样对于实验二的数据,利用(11)与实验二所有三组被试的数据的 Pearson 相关系数达到 0.959 ($p < 0.001$),决定系数 $R^2 = 0.92$;而利用实验一获得的简单回归方程(10)与实验二所有三组被试的数据的 Pearson 相关系数为 0.804 ($p < 0.001$),决定系数 $R^2 = 0.64$ 。说明经推导、整合后的模型比单纯从实验一数据获得的经验公式更能有效地解释实际变异,也证明了根据理论进行数学建模的意义。

6.2 当 RA, N 为定值时, T, WT, ST, R 对用户绩效的影响大小比较

模型(11)简化后得:

$$D = N[(1/RA)(T + WT + ST + R) + 439 - T] - (N/24) * 4973 = N\{[(1/RA) - 1]T + (1/RA)(WT + ST + R + 439) - 4973/24\} \quad (12)$$

所以,当 RA 为定值时, T, WT, ST, R 与 D 呈线性增函数关系。因为 T, WT, ST, R 单位相同(均为 ms),所以它们各自的参数 $[(1/RA) - 1]$, $(1/RA)$, $(1/RA)$, $(1/RA)$ 代表了这些因素对因变量的权重(相当于回归方程中的标准化回归系数)。又因为 $[(1/RA) - 1] < (1/RA)$,所以这四个因素对

因变量的影响大小次序为 $WT = ST = R > T$;因此,缩短系统的识别时间(R)与减小系统的分割时间及缩短用户的手写单字时间,将对用户绩效产生相同大小的影响,而且这三个因素的改善比提高用户的修改速度更能有效地提高用户完成任务的绩效。

6.3 当 T, WT, ST, R, N 为定值时, RA 与用户任务完成时间的关系

$$(1 - 14) = > D = N\{[(1/RA)(T + WT + ST + R) + 439 - T] - 4973/24\} \quad (13)$$

所以, RA 与用户任务完成时间呈倒数模型,即 RA 对因变量的影响随着 RA 值的增大而减小。从数值上看, RA 每提高 1%,任务完成时间至少节省 1000 ms。

7 总结

(1)本研究推导与建立了基于时间分割的手写输入系统用户绩效的静态模型(6)与动态模型(图 3),这两个模型在实验一中得到了检验与修正,并获得了两者的整合模型(11)。

(2)实验二进一步证明:该整合模型能在常用商业识别软件中应用,并对不同训练水平的用户绩效的有较高的解释能力($R^2 > 0.7$)。

(3)应用该整合模型发现:当识别正确率 RA、输入字数 N 为定值时,修改时间 T、单字手写时间 WT、分割时间 ST、识别时间 R 这四个因素与任务完成时间 D 呈线性增函数关系,影响大小次序为 $WT = ST = R > T$;当 T、WT、ST、R、N 为定值时,识别正确率 RA 与任务完成时间 D 呈倒数模型。识别正确率每提高 1%,任务完成时间至少节省 1000ms。

参 考 文 献

- 1 I. Scott M, Larry Ch. A Performance Comparison of Two Handwriting Recognizers. *Interacting with Computers*, 1999, 11 (3): 283 ~ 297

- 2 Richard C D, James L, Jason A B, et al. A Framework for Sharing Handwritten Notes Demonstrations. Proceedings of the ACM Symposium on User Interface Software and Technology. New York, USA: ACM Press, 1998. 119 ~ 120
- 3 Heiko S, Tng T H, Gareth L. Beyond Translation: Approaches to Interactive Products for Chinese Consumers. International Journal of Human - Computer Interaction, 2001, 13 (1): 41 ~ 51
- 4 Cheng Y F. Skills and theory for Inputting Chinese Characters into computer with keyboard (in Chinese). Chinese Institute of Information Beijing, Beijing: China, 1996. 156 ~ 212
(陈一凡. 汉字键盘输入技术与理论基础. 北京:北京信息学院, 1996. 156 ~ 212)
- 5 Clive F, Richard H, Pam M. Recognition Accuracy and User Acceptance of Pen Interfaces Papers: Pen Interfaces. Proceedings of ACM CHI '95 Conference on Human Factors in Computing Systems. New York, USA: ACM Press/ Addison - Wesley Publishing Co, 1995. 503 ~ 510
- 6 Roy S K. VRUSE - a computerized diagnostic tool: for usability evaluation of virtual; synthetic environment systems. Applied Ergonomics, 1999, 30(1): 11 ~ 25
- 7 Sherry P C, Robert D D, Beverly H W. The Effects of Recognition Accuracy and Vocabulary Size of a Speech Recognition System on Task Performance and User Acceptance Communication: Auditory and Vocal Communication. Proceedings of the Human Factors Society 32nd Annual Meeting, Santa Monica, CA: Human Factors Society, 1988. 232 ~ 236
- 8 Ben Sh. Response Time and Display in Human Performance with computer. Computing Surveys, 1984, 16 (3): 265 ~ 285
- 9 Steven L T, Alexander I R. A Performance Model of system Delay and User Strategy Selection [C]. Proceedings of ACM CHI '92 Conference on Human Factors in Computing Systems. New York, USA: ACM Press, 1992. 295 ~ 305.
- 10 Wang S, Wang A Sh. Cognitive Psychology (in Chinese). Beijing: China, Peking University Press, 1992. 5 ~ 65
(王甦,王安圣. 认知心理学. 北京:北京大学出版社, 1992. 5 ~ 65)
- 11 Gao D G, Zhong Y P. The Influence of the frequency of use on the speed of recognition in the Chinese characters (in Chinese). Psychological Science, 1995, 18: 225 ~ 228
(高定国, 钟毅平. 字频影响汉字的认知速度的实验研究. 心理科学, 1995, 18: 225 ~ 228)
- 12 Ben S. Response Time and Display Rate. In: Ben ed. Designing the User Interface Strategies for effective Human Computer Interaction. Boston, USA: Addison - Wesley Publishing Company, 1997. 205 ~ 223.
- 13 Clive F, Hull R. Pam Morgan Recognition Accuracy and User Acceptance of Pen Interfaces Papers: Pen Interfaces. Proceedings of ACM CHI '95 Conference on Human Factors in Computing Systems. New York, USA: ACM Press, 1995. 503 ~ 510
- 14 Richard A S. Principle of Motor Control and Movement Accuracy. In: Richard A Schmidt ed. Motor Learning and Performance. Human Kinetics Books Champaign Illinois, 1991. 101 ~ 141
- 15 Guo Z G. The Statistical Research Methods in Social Sciences (in Chinese). Chinese People 's University Publisher, Beijing: China, 1999. 8 ~ 171
(郭志刚. 社会科学中的统计研究方法. 北京:中国人民大学出版社, 1999. 8 ~ 171)
- 16 Joreskog, Sorbom. LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language. Chicago: Scientific Software International, 1989. 3 ~ 120
- 17 Guerrier J H, Manivannan P, Nair S N. The role of working memory, field dependence, visual search, and reaction time in the left turn performance of older female drivers. Applied Ergonomics, 1999, 30(2): 101 ~ 119
- 18 The Frequency of Usage Dictionary for Chinese Characters (in Chinese). Beijing Language Research Institute, 1986
(现代汉语频率辞典. 北京:北京语言研究所, 1986)
- 19 Spine T M, Beverly H W, Maynard J F. An Economical Approach to Modeling Speech Recognition Accuracy. International Journal of Man - Machine Studies, 1984, 21(3): 191 ~ 202
- 20 Sherry P C, Beverly H W, Robert D D. Effects of recognition accuracy and vocabulary size of a speech recognition system on task performance and user acceptance. Human - Factors, 1990, 32 (2): 183 ~ 196
- 21 Wolf C G. Understanding Handwriting Recognition from the User 's Perspective Computer Systems: Modeling. Proceedings of the Human Factors Society 34th Annual Meeting, Santa Monica, CA: Human Factors Society, 1990. 249 ~ 253
- 22 Gerry M, Davies D R. Differences in ability and performance, in Human performance: Cognition, Stress and Individual Differences. Psychology Press, Taylor & Francis Group, 2001. 255 ~ 259
- 23 Zhang K, Chen Y F. A Cognitive Model of Typing Chinese into Computers (in Chinese). Journal of Chinese Informatics, 1992, 5 (4): 13 ~ 19 (张侃, 陈一凡. 汉字键盘输入的认知模型. 中文信息学报, 1992, 5(4): 13 ~ 19)

HUMAN PERFORMANCE MODELING IN TEMPORARY SEGMENTATION CHINESE HANDWRITING RECOGNIZER

Wu Changxu , Yang Qunhui , Zhang Kan

(Institute of Psychology , Chinese Academy of Sciences , Beijing , 100101 China)

Hu Yongge , Yang Lei

(Intel Research China Center , Beijing , 10000 China)

Abstract

To predict and improve human performance in temporary segmentation Chinese handwriting recognizer , static and dynamic performance models were set up. Experiment 1 verified and modified these two models , and an integrated model ($D = N (WT + ST + R + 439) + N[(1/RA) - 1] [(0.816WT - 51.94 - 439) + WT + ST + R + 439] (N/24)4973$) was obtained. Experiment 2 further indicated that the integrated model could also account for the variance of the human performance data in another handwriting recognizer satisfactorily ($R^2 > 0.7$). The application of the integrated model suggested that : (1) When the recognition accuracy (RA) and the number of characters inputted (N) were constants , the order of influence of the four factors (repairing time (T) , average writing time for each character (WT) , segmentation time (ST) , recognition time (R)) was $WT = ST = R > T$; (2) When these four factors were constants , 1 % improvement of recognition accuracy will save the task completion time at least 1000 ms.

Key words mathematics psychological modeling , Chinese handwriting recognizer , human performance modeling , segmentation time.