

HchMER: 混合人机智能的手写数学表达式识别方法^{*}

康文惠^{1,2}, 黄进¹, 仝青山^{1,2}, 付强^{1,2}, 田丰¹, 戴国忠¹

¹(人机交互北京市重点实验室(中国科学院 软件研究所), 北京 100190)

²(中国科学院大学 计算机科学与技术学院, 北京 100049)

通信作者: 黄进, E-mail: huangjin@iscas.ac.cn



摘要: 随着人工智能和端到端识别方法在手写数学表达式识别上的应用, 数学表达式识别准确率有了明显的提高. 然而, 与公开数据集上的测试不同, 实际应用中人的参与为识别算法引入了更多的不确定因素, 如个性化的笔画信息、包含歧义的手写字符, 以及不确定的公式结构等, 影响了识别算法的性能. 为此, 提出了一种混合人机智能的手写数学表达式识别方法 HchMER. 该方法借助手写数学公式识别算法、知识库和用户反馈, 增强机器对用户输入的数学表达式的理解, 从而提升手写数学表达式的编辑速度和准确率. 为了验证 HchMER 的有效性, 将其分别与 MyScript Math Recognition (MyScript) 算法, 以及一个成熟的商用产品“Microsoft Ink Equation” (InkEquation) 进行了比较. 实验结果表明, HchMER 在准确率上较 MyScript 和 InkEquation 分别提高了 23.2% 和 26.51%. 在平均完成时间上, HchMER 比 MyScript 增加了 44.46% (9.6 s), 但是比 InkEquation 降低了 11.48% (4.05 s). 同时, 被试在问卷调查和半结构化访谈中对 HchMER 给予了肯定.

关键词: 人机混合智能; 草图理解; 在线手写数学表达式识别; 用户界面

中图法分类号: TP391

中文引用格式: 康文惠, 黄进, 仝青山, 付强, 田丰, 戴国忠. HchMER: 混合人机智能的手写数学表达式识别方法. 软件学报, 2025, 36(2): 915-938. <http://www.jos.org.cn/1000-9825/7169.htm>

英文引用格式: Kang WH, Huang J, Tong QS, Fu Q, Tian F, Dai GZ. HchMER: Hybrid Human-machine Intelligence Method for Handwritten Mathematical Expression Recognition. Ruan Jian Xue Bao/Journal of Software, 2025, 36(2): 915-938 (in Chinese). <http://www.jos.org.cn/1000-9825/7169.htm>

HchMER: Hybrid Human-machine Intelligence Method for Handwritten Mathematical Expression Recognition

KANG Wen-Hui^{1,2}, HUANG Jin¹, TONG Qing-Shan^{1,2}, FU Qiang^{1,2}, TIAN Feng¹, DAI Guo-Zhong¹

¹(Beijing Key Laboratory of Human-computer Interaction (Institute of Software, Chinese Academy of Sciences), Beijing 100190, China)

²(School of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: With the application of artificial intelligence (AI) and end-to-end recognition methods in handwritten mathematical expression recognition, there has been a significant improvement in recognition accuracy. However, in contrast to tests on public datasets, real-world applications involving human input introduce more uncertain factors into recognition algorithms in practice. Factors such as personalized stroke information, ambiguous handwritten characters, and uncertain formula structures can significantly impact the performance of the recognition method. To address these challenges, HchMER, a hybrid human-machine intelligence method for handwritten mathematical expression recognition, is proposed. HchMER combines handwritten mathematical formula recognition algorithms, knowledge bases, and user feedback to enhance the machine's comprehension of user-input mathematical expressions, thereby improving the editing speed and accuracy of handwritten mathematical expressions. To assess the effectiveness of HchMER, it is compared with MyScript Math

* 基金项目: 国家自然科学基金 (62172397); 中国科学院青年创新促进会项目 (20200113)

收稿时间: 2023-09-07; 修改时间: 2023-12-07, 2024-02-05; 采用时间: 2024-02-24; jos 在线出版时间: 2024-06-14

CNKI 网络首发时间: 2024-06-19

Recognition (MyScript) and a mature commercial product named “Microsoft Ink Equation” (InkEquation). Results show that HchMER outperformed MyScript and InkEquation in accuracy by 23.2% and 26.51%, respectively. In terms of average completion time, HchMER exceeded MyScript by 44.46% (9.6 s) but fell short of InkEquation by 11.48% (4.05 s). Furthermore, participants affirm HchMER in a questionnaire survey and semi-structured interviews.

Key words: human-machine hybrid intelligence; sketch understanding; online handwritten mathematical expression recognition; user interface

手写数学表达式识别在网络教学、科学研究和办公自动化等不同领域中有着广泛应用. 与符号调色板类公式编辑器 (如 MathType)、数学描述语言 (如 LaTeX, MathML) 相比, 手写数学表达式避免了在符号调色板上繁琐而耗时的选择和大量标记语言的记忆, 是一种更自然高效的输入方法, 尤其是在移动设备或大屏幕上^[1-4]. 然而, 由于手写数学表达式本身具有的特点, 如二维非线性结构、包含特殊字符, 手写字符的歧义性和模糊性等, 使得在实际环境中进行识别成为一项极具挑战性的任务^[5,6].

人为不确定因素, 如因人而异的手写字符、连笔或潦草书写, 以及书写字符不包含与当前数据集中等, 为手写数学表达式识别在实际应用中带来更大的挑战. 针对这些挑战, 研究人员从 3 个层面进行了探讨和研究: 第 1 类方法是深剖识别算法. 研究人员从经典三阶段结构化方法到端到端编码解码的识别方法, 手写数学表达式的识别准确率有了显著性提高^[7-9]. 不幸的是, 目前最先进的识别方法在手写数学表达式识别与排版公式检测竞赛所提供的数据集上只能达到 80% 左右的准确率^[10-12], 它与实际应用中可接受的准确率之间仍然存在着巨大的鸿沟^[13]. 第 2 类方法是约束识别内容, 主要表现为利用交互式用户界面或人为介入的方法来协助手写数学表达式的识别. 这类方法的关键在于如何借助规范化书写或人为修改笔画的方式服务识别算法, 且需要花费更长的时间来完成数学表达式的书写任务. 如在手写数学表达式识别引入人的参与来修正公式结构中不确定笔画和结构补笔^[14], 借助用户书写字符间空间关系的自动检测来生成子表达式周围的可见边界框^[15]. 第 3 类方法是干预识别结果, 主要表现为对识别结果的再编辑. 例如, 以文本方式直接编辑识别结果的数学描述语言 (LaTeX, MathML), 以“所见即所得”的符号调色板^[16]直接对标准排版的数学表达式进行编辑. 但是, 第 1 类方法因公开数据集难以体现用户书写内容的多样性而使得识别方法在实际应用中表现欠佳, 尤其在用户潦草书写和连笔的情况下. 第 2 类方法尽管引入约束条件或人为修改笔画, 但是其目的在于优化识别算法的输入信息, 而非输出结果. 它能够在花费较多时间的情况下解决部分识别问题, 可仍然无法解决因识别算法自身缺陷引起的识别错误问题. 最后一类方法能够纠正识别错误的字符和结构, 但是存在效率低下、结构修改困难和对新用户不友好等问题. 因此, 如何在用户参与的非实验环境下为用户提供快速准确的手写数学表达式结果是亟需解决的问题.

鉴于当前在实际使用中手写数学表达式识别存在的问题, 本文提出了一种混合人机智能的手写数学表达式识别方法 HchMER, 以促进人为修改识别错误的情况下快速准确地获取手写数学表达式. 该方法以手写错误的直观提示、错误字符的快速推荐和错误结构的高效编辑为目的, 借助手写数学公式识别算法、公式数据集和数学的组合逻辑和规则, 以及人对错误识别结果的反馈信息来混合人和机器的对于手写数学表达式的理解. 为了验证 HchMER 的有效性, 我们进行了两项对比实验. 一项是 HchMER 与可人为修改手写信息的基础方法 MyScript Math Recognition (MyScript) 的对比, 另一项是 HchMER 与可结合符号调色板编辑的成熟商用产品 Microsoft Ink Equation (InkEquation) 的对比. 实验结果表明, HchMER 能够在设定的任务上获得 94.25% 的准确率, 比 MyScript 和 InkEquation 分别提高了 23.2% 和 26.51%. 在单个公式的平均完成时间上比 InkEquation 降低了 11.48% (4.05 s), 但比基础方法增加了 44.46% (9.6 s). 在用户主观评价中, 被试在偏好、效率和准确度等方面给予了肯定, 这也与定量结果相一致. 本文的主要贡献有:

(1) 以用户在实际应用中可人为修改识别错误为基础, 结合手写错误的直观提示、错误字符的快速推荐和错误结构的高效编辑, 提出了一种混合人机智能的手写数学表达式识别算法 HchMER.

(2) 通过两项对比实验发现, HchMER 在完成时间上略长于仅仅依赖笔画修改的基础方法 MyScript 外, 在准确率上均优于另外两种方法.

(3) 通过与 MyScript 进行对比实验发现, 在针对人为参与的 MyScript (AI 算法) 应用中, 提供有效的纠错方法

是不可或缺的. 同时, 与 InkEquation 的对比实验说明设计高效的纠错技术是提升手写数学公式识别方法的重要途径.

1 相关研究

为了阐述手写数学表达式识别研究的最新进展, 我们从手写数学表达式识别、人为因素导致的识别错误, 以及人机混合智能这 3 个方面进行了文献调研.

1.1 手写数学表达式识别

手写公式识别一直是草图识别方面比较活跃的研究领域, 识别过程可以看作将手写公式笔画翻译成计算机可以使用形式化序列, 如 LaTeX 或 MathML 等数学描述语言^[17]. 近年来, 随着笔和触摸设备的快速发展, 手写公式识别研究吸引了更多学者的兴趣, 且存在相关的国际会议“ICFHR 手写数学表达式离线识别和发现竞赛”^[18]以供研究人员交流学习.

手写公式识别方法主要分为经典三阶段结构化方法和基于端到端的深度学习方法. 基于结构化的方法将公式识别划分为 3 个子任务, 分别是笔画分割、字符识别和结构解析^[19]. 在这类识别中, 可以按顺序依次完成 3 个子任务, 即前一个子任务的输出是后一个子任务的输入; 也可以结合上下文语义信息全局解决识别任务, 即在构建表达式结构的同时完成分割任务. 不过顺序解决的方法容易累积错误, 难以在识别的整个过程中借助全局语义信息^[20,21], 而全局解决的方法存在时间复杂度高, 需要定义有效的搜索策略等问题^[22]. 基于端到端的深度学习方法是借助深度神经网络和编码解码框架构建的方法, 在识别过程中将笔画分割-字符识别-结构解析 3 个过程融为一体, 直接将手写笔画转换为数学描述语言^[8]. 对于这类识别方法, 研究人员主要从解码编码模块、注意力模型、数据编码方式等角度入手解决手写公式识别难的问题^[23,24]. 例如, 在编码解码模块层面, Bian 等人^[23]在原有的从左到右编码解码结构下, 添加了从右到左的解码器来促进双向学习. 数据编码上常用的改进方式有 LaTeX 标记序列转换为解析树^[25]和基于符号关系的序列表示方法^[26]等. 对注意力模型而言, 该模型用以解决符号大小不同、注意力分布不精确和注意力漂移等问题^[27,28]. 常用的注意力模型有注意多尺度模型^[28]、位置矫正注意模型^[29]和 multi-head attention^[30]等. 不过, 基于端到端的方法依赖于数据集的大小, 且存在笔画和 LaTeX 序列的对齐、缺乏语法信息的整体表征等问题, 尤其是无法生成字符与笔画的对应关系, 难以解释识别的具体机制.

除了从算法层面提升公式识别能力外, 部分研究者另辟蹊径从笔式用户界面的角度来提高手写数学公式的识别^[14,15]. 如 Math Boxes^[15], 在检测到公式符号间 (包括上标、下标和分数) 的空间关系时, 将自动生成子表达式周围的可见边界框, 且数学框中包含的子表达式可以通过直接向其给定边界添加新项来扩展. 另一方面, 也存在借助人在回路的方法来解决手写公式识别问题的研究^[14], 其核心是在公式识别的不同阶段引入人的参与来提升算法的识别能力.

仅仅从手写数学表达式识别结果的准确率来说, 目前最先进的识别方法在手写数学表达式识别与排版公式检测竞赛所提供的数据集上只能达到 80% 左右的准确率^[11,12,31-33], 与可接受的准确率 (>91%)^[13]之间仍然存在着巨大的鸿沟^[14], 尤其是接近可接受准确率最后的 10%, 仅仅依靠识别算法来缩小差距可能是一项极具挑战性的任务.

1.2 人为因素导致的识别错误

尽管手写数学表达式的识别算法随着深度神经网络和大数据的发展, 从经典三阶段结构化方法演变到基于端到端深度神经网络的方法, 但是仅仅依赖识别算法仍然无法解决手写数学公式在实际应用中的全部问题: 1) 手写笔画自身带来的识别难问题^[14,15], 包括不同风格的用户书写、潦草书写、连笔书写等带来的识别问题; 2) 数学表达式中相似字和特殊字符因人为书写导致的歧义问题^[25,27], 如 x 和 X 大小写相似、 z 和 2 的书写相似等; 3) 数学表达式中的结构因书写 (空间位置关系) 不确定导致的识别难问题^[34], 包括上下标结构、根式、分式等非线性结构, 以及嵌套结构的识别问题; 4) 数据集中不包含用户书写字符的问题^[10]. 当前以 CROHME 数据集为数学表达式识别算法的数据源, 但是该数据集包含的数学字符不到 120 个, 远远小于日常需求中的字符量, 缺少如集合间的操作符号、推导符号等.

针对前 3 个问题, 研究人员主要通过修改笔画和提供识别候选集^[19]来解决, 但是这种解决方法并不完备, 主

要原因在于用户的书写习惯和风格并不容易改变,且现有的数据集无法囊括各种风格的书写字符和公式;其次是在测试集上取得较高准确率的识别算法并不能完美迁移到实际应用中,即不完备知识下训练的算法在未知或者从未见过的手写公式识别任务上表现不佳;另外,用户在使用识别算法,尤其是使用端到端方法时,识别结果会随着持续书写或者修改已有笔画而出现不可理解的识别结果,比如在修改错误识别字符对应的笔画时,结果中部分已识别正确的公式也会发生改变.针对上述的4个问题,也有研究采用对识别结果的再次编辑来解决,如基于数学描述语言(LaTeX, MathML)的编辑和基于调色板用户界面的编辑(如 Microsoft Office, MathType).通常情况下,基于调色板用户界面的编辑方式以自上而下的方式进行,使得对已输入公式的结构修改变得困难^[13].另外,该编辑方式需要用户浏览符号和模板菜单,在编辑字符时需要在键盘和鼠标之间来回切换,对新手用户不友好^[34-37].

总的来说,目前针对数学表达式识别结果的纠正主要通过修改笔画和对识别结果再编辑进行的,但是笔画的修改只能解决部分识别问题,而识别结果的再编辑是以人的记忆能力为驱动,主要体现在符号调色板中手动查找繁琐的模块/字符或对特定标记字符/语句的记忆上.从结果来看,这几种方法对识别准确率的提升有限.从交互效率来看,基于数学描述语言或符号调色板的方法操作繁琐、修改公式结构困难,且对新用户不友好.为此,我们借助人机混合智能的思想,结合人为参与、错误字符/结构与公式之间的关联、历史数据集等信息设计了一种高效的识别方法以便于快速准确地获得手写数学表达式标准化结果.

1.3 人机混合智能

随着人工智能性能的快速提升,机器已从工具的定位过渡到可以合作的伙伴,即赋予机器类似人类的社会认知技能和/或类似人类的化身^[38].而在这个转变过程中,人与机器在完成任务时的分工也在发生着变化.另外,仅仅依赖于人工计算和机器自动化难以完全解决实际应用中遇到的问题,如人工效率低下,或机器预测的准确率不足以满足实际需求等.为此,研究人员引入人机混合智能的方法来弥补机器能力和人们不同需求之间的社会技术差距^[39,40].

人机混合智能的核心思想是通过人类智能和机器智能的互补提升,达到比仅仅依赖人工或机器完成任务时更好的效果^[41].从增强智能的主体划分,可以分为人在回路的方法和机器在回路的方法,其中,人在回路的方法主要用于增强机器的智能,以机器智能为核心;机器在回路的方法主要借助机器辅助来增强人的智能,以人为核心.相较而言,人在回路的研究占主导优势,这种方法允许人类直接影响人工智能的工作,从而改进人工智能单独工作时可能导致的错误和失败,或借助人的反馈来循环改进系统^[42,43].如 So^[44]提出了一种将人融入敏捷产品开发循环的过程框架,利用生成的用户反馈来训练机器学习模型并指导后续的设计过程,其优势在于借助用户反馈的定量评估取代定性用户测试;Lee 等人^[45]结合数据驱动模型与专家的运动学特征知识来评估中风患者的运动质量.在整个过程中,治疗师可以回顾分析结果,以便更好地了解患者的表现,并提供反馈以调整个性化康复评估模型.

不同于人在回路的方法,机器在回路的方法主要优势在于利用人工智能优势来弥补人在感知、认知和行为层面的不足.在感知和行为层面,研究人员主要是借助智能硬件设备、算法或远程辅助来提升用户的感知或运动能力^[46,47].而在认知层面,更多的研究聚焦于如何借助机器智能来提升用户的工作效率、决策能力和创造能力等方面,如 Clark 等人^[48]在故事创作和标语设计上引入了机器在回路的方法,Gerlach 等人^[49]借助机器在回路的方法用以支持用户在维基百科实体链接系统中的编辑,Qosim 等人^[50]在合成氨的过程中介入机器学习方法用以异常检测,李洋等人^[51]将人机混合引入到了笔式 CAD 概念设计中去.然而,不管是人在回路还是机器在回路的混合智能,目前的研究更多聚焦于具体问题的具体解决方案上,即在具体问题中讨论如何混合人和机器的智能.

在手写数学表达式识别方面,除了人在回路方法的讨论外^[14],如何将人和机器两种智能混合到数学表达式上的研究很少.为此,我们提出了一种通过混合人和机器在处理数学表达式方面的优势来提高机器识别效率的方法——HchMER,该方法与人在回路方法的区别在于,我们更关注人在参与过程中如何结合可能出现的错误为用户提供可行高效的纠错方式,以及如何结合人机混合智能的思想来解析用户的纠错意图.

2 HchMER 方法

针对当前识别算法在实际应用中存在的笔画错误、识别字符错误和识别结构错误问题,我们设计了 HchMER 方法,用于手写错误的直观提示、错误字符的快速推荐和错误结构的高效编辑. 接下来,我们将从 HchMER 的基本框架出发,详细描述其设计细节.

2.1 HchMER 基本框架

如图 1 所示, HchMER 的基本框架包含一个知识库和 5 个模块, 模块分别是: 1) 笔画输入模块, 主要用于用户输入笔画和对笔画的修改 (清空、删除、撤销、恢复); 2) 识别模块, 负责对笔画输入模块的笔画进行识别, 并向用户和人机混合模块提供识别结果; 3) 用户纠错模块, 主要是为用户提供公式的纠错方式. 在该模块中, 用户可以根据机器提供的纠错方式进行手写公式再处理. 需要说明的是, 该模块并不能改变识别算法的性能, 也不根据用户行为对识别算法的参数进行调整; 4) 纠错任务理解模块. 该模块结合用户纠错模块提供的纠错信息, 负责整合用户传递的信息, 理解用户表达的纠错意图, 向人机混合模块传递用户信息和唤醒人机混合模块中对应的处理任务; 5) 人机混合模块, 负责 3 类任务的完成. 图 1 中, 在输出不正确 (否) 的情况下, 虚线表示用户可以使用 HchMER 提供的交互技术进行纠错; 实线表示 HchMER 可以再次通过识别和错误提示向用户提供反馈. 第 1 类任务是结合数据库的先验知识和识别结果, 用以提示当前识别可能存在的错误; 后两类任务主要负责错误字符的快速推荐和错误结构的高效编辑, 其过程为在接受纠错任务理解模块的唤醒后, 结合用户信息、机器对识别结果和识别错误的解析信息, 进行修正输出公式. 知识库主要为人机混合模块提供手写数学公式信息和数学公式相关的专家知识. 在使用 HchMER 方法时, 用户和机器可以经过多粒度多轮次的混合, 进而提升人和机器对公式的认知.

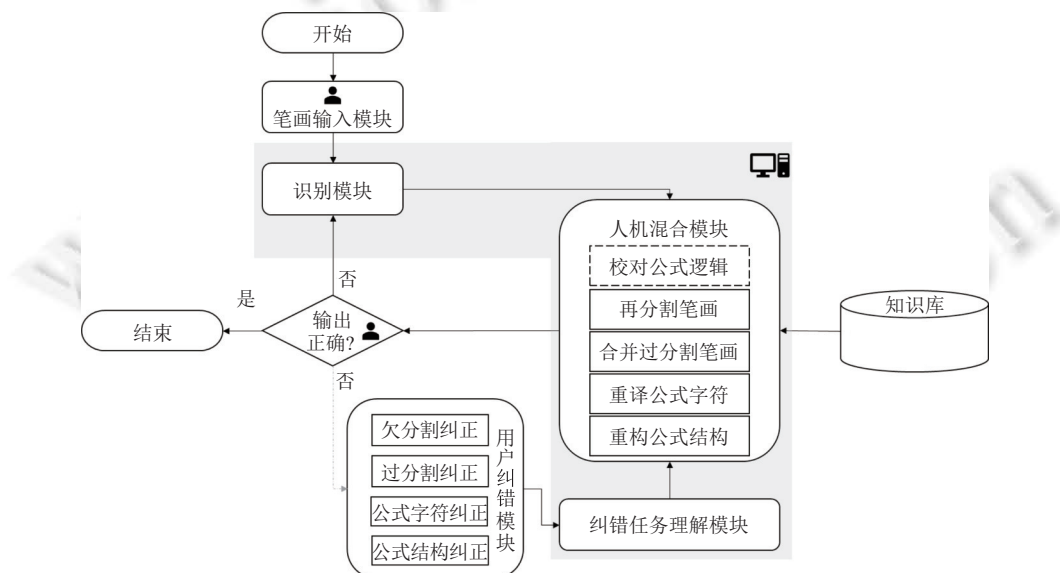


图 1 HchMER 的基本框架

在整个框架中, 人与机器智能的混合存在于两个环路. 一个环路是机在回路 (人-机-人), 即用户借助笔画输入模块输入公式信息后, 机器经过识别和人机混合模块的错误提示, 向用户输出机器认为正确的数学公式. 在用户认为输出不正确的情况下, 机器可再次通过识别和错误提示向用户提供反馈; 另一个是人在回路 (机-人-机), 主要是人机混合模块向用户提供输出后, 用户通过用户纠错模块提供的纠错方式向机器反馈信息, 以提升机器输出更准确的数学公式. 在机器难以理解用户意图的情况下, 通过向用户提供当前错误的反馈信息, 以获得用户的交互信息, 从而优化人机混合模块的输出性能.

最后, 在整个框架中, 除了用户自主修改的笔画外, 我们假定用户输入的笔画信息都准确的. 这意味着, 在用户

纠错模块提供的交互方式并不是通过修改笔画信息来完成纠错的. 这种假定的目的在于促使机器智能去适应用户的书写习惯和风格, 而不是用户去适应机器的识别能力和识别条件.

2.2 知识库

我们的知识库主要包含 CROHME 比赛数据集, 公式内字符间的组合逻辑和规则. CROHME 比赛数据集是广泛应用于在线手写识别的数据集^[10,18,52], 包含 10061 个手写数学表达式和 101 个数学字符^[18], 主要用于识别算法的训练和作为历史数据来获取公式内字符和结构之间的关系. 公式内字符间的组合逻辑和规则来源于专家知识, 如“操作符不能直接和上下标(幂/底)组合”“双目操作符连接两个操作数”等, 主要用于对识别结果的逻辑校正. 在 HchMER 中, 这些组合逻辑和规则经由人工将非结构化专家知识转化为 LaTeX 公式结果中字符间的关系所得.

另外, 我们定义了一些术语.

- (1) 笔画块: 包含在同一个边界框内的笔画被称为笔画块, 主要服务于用户对笔画, 或笔画对应字符的选择.
- (2) 结构关系: 表示公式中字符与相邻字符之间的空间关系, 包括 overscript (hat)、superscript、inline、subscript 和 underscript.
- (3) 字符属性: 为了更好地缩小纠错范围, 在我们的实验中将公式字符的属性划分为操作符、操作数和定界符, 其中操作数又分为希腊字母、普通字符 (a-z).
- (4) 纠错动作: 指代用户与数学公式笔画交互时所采取的动作, 主要包括单击、双击、拉框和移动.
- (5) 纠错信息: 在我们实验中, 纠错信息主要包括公式笔画、笔画块、字符等.

2.3 笔画输入模块

笔画输入模块以用户的手写笔画为输入, 以数学公式笔画的数字化信息为输出向识别模块传递可识别数据. 该模块主要目的是方便用户自由书写和修改数学公式笔画. 为了区分用户纠错模块提供的交互方式, 我们将用户对数学公式笔画的修改纳入到笔画输入模块. 对数学公式笔画的修改主要包括清空、撤销、恢复和删除. 需要说明的是, 在笔画输入模块中对数学公式笔画的修改, 并不以迎合后续的识别算法为目的, 也不以后续错误结果的纠正为目的. 这种对笔画的修改, 主要针对用户的笔误、漏写和自主修改 (因得到错误识别反馈后对错误对应笔画的修改, 我们称为被动修改).

2.4 识别模块

识别模块以笔画输入模块提供的可识别数据为输入, 以识别算法为驱动, 向人机混合模块提供输入数学公式的识别结果. 在该模块中, 我们采用 MyScript Math Recognition (MyScript) (<https://www.myscript.com/ai/>) 作为基础识别方法. 从识别逻辑来说, MyScript 在同一水平上同时处理数学公式的分割、识别、语法和语义分析, 以保证识别过程中上下文语义的连续性. MyScript 的主要组成为: 1) 在分割阶段, MyScript 结合公式笔画的在线和离线特征, 采用字符分类器确定字符分割的组成和概率; 2) 在识别阶段, MyScript 采用多层感知机 MLP 进行字符识别; 3) 在语法和语义分析阶段, MyScript 设计了与公式笔画空间关系相关的语法规则, 如图 2 中一个分数规则定义了分子、分数线和分母之间的垂直关系, 而后采用统计语言模型进行结构解析和识别结果的生成.

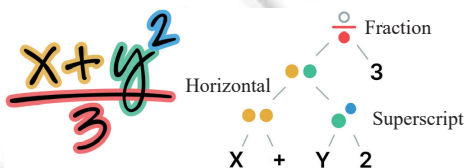


图 2 分式语法^[53]

本文基础方法选择 MyScript 的原因在于: 1) MyScript 的识别准确率表现优异. MyScript 在 2019 年在线手写数学表达式识别竞赛中取得了 79.15% 的识别准确率 (训练集为 CHROME 2014, 测试集合为 Test 2012, Test

2013)^[10,18] 与 2019 年在线手写数学表达式识别竞赛中的识别准确率相比^[52], MyScript 的公式识别准确率排名第 3, 且与第 1 的识别准确率相差 0.58% (排名第 1 的 USTC-iFLYTEK 识别准确率为 80.73%)。2) MyScript 仍属于传统三阶段的结构化识别算法, 能够在分割、识别和语义分析阶段分别引入人的智能。像 USTC-iFLYTEK 这类基于端到端识别算法, 在算法层面引入人的智能 (纠错信息) 是目前难以实现的。3) MyScript 的已有识别结果并不会因新笔画的输入而发生大幅度改变, 最可能引起的变化是与新输入笔画相关联的小部分笔画的识别结果。然而, 基于端到端的识别算法并不能确保持续正确识别与新输入笔画无关的笔画。4) MyScript 算法对运行环境的依赖比较弱, 能够为用户提供多种运行环境下的识别服务。5) MyScript 是目前开源在线手写识别方法中识别准确率最高的算法之一, 能够保证我们方法的可复现和重复性研究。

2.5 用户纠错模块

用户纠错模块主要以用户感知识别结果为基础, 为用户提供人为参与识别算法的交互技术。鉴于当前识别算法仅提供错误字符的识别候选集用以纠正错误识别的字符, 没有融合数学表达式上下文信息, 如错误字符与其关联字符间的关系、历史数据, 用户对该错误字符的操作等, 以及识别算法无法解决字符对应笔画错误分割导致的错误识别问题, 我们提出了 3 种错误字符快速推荐的交互技术, 分别为过/欠分割纠正和公式字符纠正。针对公式识别结果中结构修改难问题, 我们设计了一种高效的公式结构纠正技术。需要说明的是, 在结构修改中也会使用到过/欠分割纠正技术。

接下来, 我们将详细描述这 4 种纠错技术, 其中, 在纠错时, 笔画块的边界框通过长击背景来显示。

2.5.1 过/欠分割纠正

为解决识别模块中基础算法难以准确分割数学公式字符笔画的问题, 我们提出了一种字符笔画再分割的交互手段。它能够引入人的智能用以对机器欠分割错误的纠正和对过分割笔画的合并。需要说明的是, 字符笔画的分割错误包括笔画欠分割和笔画过分割两种, 其中欠分割是指将不属于某一字符 (如字符 A) 的笔画划分到了它的笔画组成中, 而过分割是指将属于某一个字符的笔画分割到了其他字符的笔画组成中去。在我们方法中, 欠分割纠正以长击欠分割笔画的笔画块触发, 从而向机器传递笔画分割错误的信息, 促使人机混合智能模块再分割 (见第 2.7.2 节)。过分割合并以拉框的方式解决笔画过分割的错误, 用以将过分割的两个或两个以上的笔画块重新组合为一个 (见第 2.7.2 节)。例如, 当“k”的笔画被过分割为“l”和“<”的笔画时, 过分割提示可以通过拉框的方式将两个过分割的笔画重新组合为“k”的笔画, 然后提供于人机混合模块和识别模块进行识别和字符的快速推荐。另外, 欠分割提示和过分割合并能够在字符笔画分割中交叉使用。

2.5.2 公式字符纠正

公式字符纠正的目的在于修改识别结果中错误的字符。常见的字符识别错误有相似字符、大小写字符和特殊字符等字符的识别错误。在字符纠正中, 我们提出了一种自适应的软键盘字符扩充技术 (见第 2.7.2 节)。用户能够以单击错误字符笔画的方式来触发软键盘的弹出。如果键盘内提供的字符不正确, 用户可以通过在软键盘的边界滑动进行自适应替换和扩充键盘内的字符。这种自适应的扩充技术避免了用户在笔输入和键盘输入之间来回切换, 也能够减少用户在软键盘中定位和输入正确字符的时间。

单击错误字符笔画而不是识别结果中字符, 主要是因为识别结果中字符可能与手写字符数不匹配, 甚至存在识别模块忽略选中笔画的可能性。

2.5.3 公式结构纠正

复杂的公式结构是导致手写公式识别错误的原因, 也是导致识别结果纠错和再编辑困难和耗时的主要原因。因此, 为了解决识别结果中结构错误的问题, 特别是上下标结构的错误, 我们提出了一种公式结构纠正的交互手段。

在结构纠正中, 用户可以在不对原有笔画更改的情况下, 通过控制错误结构内字符笔画的指向来直接修改识别结果。与修改笔画后再次借助识别算法的识别来纠错不同, 我们的方法主要聚焦于直接根据用户的修改意图进行识别结果的精准操控 (见第 2.6、2.7 节)。这种通过笔画直接操作识别结果的优势在于: (1) 避免了再次识别后仍然无法解决结构识别不准确的问题; (2) 能够更好地理解用户修改的意图, 避免再次修改与错误字符相关联的结

构. 例如, 避免如图 3 中 MathDeck 在进行结构修改后, 还需要修改与之关联字符“x”的结构, 因为它并没有考虑与“r”关联的“x”, 而是以面板的形式弹出“x”; (3) 保证公式识别结果与笔画信息的一致性, 避免直接对识别结果进行交互式编辑的繁琐和耗时, 尤其是针对需要插入字符, 或处理本应该包含在同一水平结构内的字符但在识别时刻分到其他结构的纠正.

图 3 MathDeck 结构修改示例

2.6 纠错任务理解模块

纠错任务理解模块以用户的纠错行为 (纠错动作和纠错信息) 为信息源, 结合已有的识别结果来推断用户的纠错意图, 而后与人机混合模块协同工作来执行相应的纠错任务. 简单来说, 纠错任务理解主要用于将来自用户纠错模块的用户纠错行为映射到人机混合模块的具体任务, 其形式化表达为: 给定用户的纠错动作集 A 和纠错信息集 E , 人机混合模块在用户以纠错动作 a_i 处理纠错信息 e_i 后, 执行的交互任务 t' 为:

$$t' = \arg \max p(t_i | a_i, e_i), a_i \in A, e_i \in E, t_i \in T, t' \in T,$$

其中, T 表示人机混合模块可执行的任务集, t_i 表示任务集 T 中的任务, $p(t_i | a_i, e_i)$ 表示在纠错动作 a_i 和纠错信息 e_i 的条件下执行任务 t_i 的概率. 纠错动作集 A 主要包含如单击、双击和拉框等纠错动作, 纠错信息集 E 主要包含笔画的边界框、笔画信息和笔画对应的识别字符等.

从纠错是否确定的角度出发, 即 $p(t_i | a_i, e_i)$, 可以分为确定性纠错和不确定性纠错. 确定性纠错是明确的 ($p=100\%$), 如在字符笔画再分割中, 长击欠分割笔画的边界框所表达的纠错属于确定性纠错, 旨在向人机混合模块传递边界框内的笔画分割错误. 同理, 不确定性纠错是不明确的 ($0 < p < 100\%$), 需要结合历史信息和经验数据后从多个可能性纠错任务中推断出当前所表达的纠错. 如字符结构纠错中, 通过笔的上下左右控制结构是非确定纠错, 因为仅通过字符的移动是无法确定该字符与相近邻字符间的公式结构, 还需要依赖对已有结构错误类型的判断、与近邻字符组合的结构逻辑和整个公式的逻辑关系来推断正确的公式结构.

对于 $p(t_i | a_i, e_i)$ 而言, 它用于结合经验数据中纠错信息、纠错动作和执行的任务之间的关联关系, 以及当前的纠错动作和信息来推断可能被执行的任务, 即:

$$p(t_i | a_i, e_i) = \frac{p(t_i, a_i, e_i)}{p(a_i, e_i)},$$

其中, $p(t_i, a_i, e_i)$ 表示用户面对纠错信息 e_i 的情况下, 以纠错动作 a_i 触发交互任务 t_i 的联合概率, $p(a_i, e_i)$ 表示用户执行纠错动作 a_i 与纠错信息 e_i 交互的联合概率. 这两者的联合概率均从历史数据 $\langle A, E, T \rangle$ 三元组中计算而来.

2.7 人机混合模块

人机混合模块以人的反馈信息为中心, 结合知识库中先验知识、机器对识别结果的理解和纠错任务理解模块提供的纠错任务信息, 对识别结果进行再处理. 该模块包含 5 个交互任务 $T = \{t_1, t_2, \dots, t_5\}$, 分别为: 1) 没有人直接参与的任务, 即错误提示 t_1 . 该任务结合知识库内已有经验和数学规则, 直接对识别结果进行校对处理; 2) 存在人直接参与的任务, 即再分割笔画 t_2 、合并过分割笔画 t_3 、重译公式字符 t_4 , 和重构公式结构 t_5 , 分别服务于错误提示 (t_1), 快速推荐 (t_2, t_3, t_4) 和结构编辑 (t_2, t_5). 这些任务需要经过用户纠错模块接收人的反馈信息、纠错任务理解模块对反馈信息的加工处理, 以及结合知识库已有的经验和规则来纠正识别结果. 接下来, 我们将从 3 个层面详细描述 5 个交互任务的具体实现.

2.7.1 错误提示

错误提示 t_1 主要用于结合公式内字符间的组合逻辑和规则校对识别模块输出公式中所包含的字符、结构和字符间关系, 若存在错误则向用户提示. 这种校对是通过从头到尾遍历公式内关联字符的组合关系进行的, 其形式化表达为: 给定公式字符间的组合规则集 R 和识别模块输出的公式 f , 依次遍历公式 f 内关联字符间 s_i, s_j 的组

合关系是否符合 R 中的某一规则 r . 具体来说, 校对结果 Q 表示为:

$$Q = \prod_{i=0}^{|f|} \{p(C(s_i, s_j) \models r_k)\}, s_i \in f, s_j \in f, r_k \in R, 0 < i, j < |f|, 0 < k < |R|,$$

其中, $C(s_i, s_j)$ 表示公式内字符 s_i 和 s_j 间的组合关系, $\models$ 表示满足符号, $|f|$ 表示公式包含字符的数量, $|R|$ 表示组合规则集中规则的个数. Q 表示整个公式的校对结果, 其中 $Q=1$ 表示校对正确, $Q=0$ 表示校对错误.

针对校对结果 Q , 如果 $Q=0$, 人机混合模块向识别模块发出再次识别的请求. 当识别结果的校对连续两次都难以正确的情况下, 人机混合模块直接向用户反馈错误结果, 同时指出出现错误的位置和错误类型. 在 $Q=1$ 时, 人机混合模块直接向用户反馈识别结果.

2.7.2 快速推荐

快速推荐主要分为结合人为甄别笔画分割错误的推荐和直接对识别结果中错误字符的推荐. 其中, 人为甄别笔画分割错误的推荐需要对笔画进行再处理 (t_2, t_3), 而后进行二次识别和推荐. 直接对识别结果中错误字符的推荐无需处理笔画信息, 只关注笔画对应字符与所在公式间的上下文关系.

(1) 再分割笔画

再分割笔画 t_2 是指将本属于不同字符却划归于一个字符的笔画集合 O_u 进行再分割的过程, 其中输入是用户长击获得的欠分割笔画块信息, 输出是笔画的重新分割假设. 具体分割为: 结合欠分割笔画集合 O_u 内笔画间的最近/最远距离、属于同一字符的笔画最大距离阈值、笔画可见性、欠分割笔画数量 N , 和最多笔画数量 $L_{\max}$ (单个字符可能包含的最多笔画数量) 来构造再分割笔画的可分割假设, 其中笔画可见性是指两个笔画的最近点之间的连线未被其他笔画穿过, 而两个笔画的最近/最远距离与最大距离阈值的关系用于判断这两个笔画能否被划分到同一个分割假设中.

借鉴文献 [7] 中分割假设构造方法, 本文的分割假设也以无向图进行构造, 其原理为: 以笔画为节点, 以可见且可以划分到同一分割假设的笔画之间的连线为边构造无向图 G , 则 G 就是一个分割假设. 对于分割假设 G 而言, G 所包含的连通子图 $B = \{b_i | b_i \in O, |b_i| \leq L_{\max}\}$ 是可被接受的分割假设, 而非连通子图则是不能被接受的分割假设, 其中 O 为公式的所有笔画集合, b_i 为集合内的某一个笔画. 最后, 欠分割笔画集合被划为多个连通子图, 且在划分过程中受限于欠分割笔画数量 N 和形成数学符号的概率 $p(B_i)$, 形式化表达为:

$$O_u \Rightarrow \{B_0, \dots, B_n\},$$

$$\text{s.t.} \quad \begin{cases} \sum_{i=0}^n |B_i| = N \\ p(B_i) = p_{\text{GMM}}(c=1 | g(B_i)) \\ p(B_i) > k \end{cases},$$

其中, n 代表连通子图的数量, $|B_i|$ 表示当前分割假设下所包含的笔画数量, $p(B_i)$ 表示不同笔画分割假设形成数学符号的概率, $c=1$ 表示可接受分割假设 B_i 的笔画能够形成数学符号, GMM 和 $g(B_i)$ 与文献 [7] 中一致, 其中 GMM 为高斯混合模型, $g(B_i)$ 是一个分割假设的 4 维向量特征, 分别为平均水平距离、平均垂直偏移, 平均大小差异和两个笔画最近邻点的平均距离. k 为一个常数, 在本文中被设置为 0.5.

最后, 再分割笔画模块将新构建的分割假设依次反馈到识别模块进行重新识别, 而人机混合模块在得到重新识别的结果后, 根据笔画与字符间的关系、字符在公式中的位置、字符与相邻字符间的关系进行结果重组, 并经过逻辑校对后提供输出结果.

(2) 合并过分割笔画

与再分割笔画相反, 合并过分割笔画 t_3 是将本属于同一字符的笔画却划归到多个字符的笔画集合 O_{ov} 进行合并的过程, 其输入为拉框获得的过分割笔画, 输出也是笔画的重新分割假设. 与再分割笔画的不同之处在于 O_{ov} 的组成, O_{ov} 由多个笔画块组成:

$$O_{ov} = \{O_1, \dots, O_m\},$$

其中, m 为用户拉框选中笔画块的数量. 合并过分割笔画的原理在于将用户拉框选择的过分割笔画进行合并为一个笔画集合, 然后结合笔画间的相互关系 (与再分割笔画一致) 进行构造分割假设. 最后, 与再分割笔画处理一致,

即将新构建的分割假设反馈到识别模块,而后由人机混合模块组合处理后向用户输出.

(3) 重译公式字符

字符识别错误是手写数学公式识别中常见的错误之一. 重译公式字符 t_4 是针对这类错误而设计的, 主要解决如何快速编辑和纠正字符识别中常见字符识别错误, 包括相似字符、大小写和特殊字符这 3 类识别错误. 针对以往以全键盘方式提供字符选择时用户查找繁琐和耗时问题, 以及以识别候选集为基准提供的字符选择列表不可扩充的问题, 我们提出了一种自适应的软键盘字符扩充技术. 该技术结合当前识别错误字符 s_e 的属性 l 、 s_e 的相似字符集合 Sim 、 s_e 前后关联字符集合 S_r 、 s_e 与 S_r 的组合方式 C 和历史数据有选择地提供候选集 S , 并以软键盘方式向用户显示. 历史数据包含当前错误字符的高频修改字符, 关联字符组合, 关联字符与高频字符组合.

$$S = \bigcup_{i=0}^n s_i,$$

$$\text{s.t.} \begin{cases} \{p(s_i|l, C(s_i, S_r)) > bias, s_i \in S_e \cup Sim \\ C(s_i, S_r) = C(s_e, S_r) \\ l \in L \end{cases},$$

其中, S_e 是指 s_e 的可选高频字符集, L 表示公式字符的属性集, $bias$ 表示字符 s_i 可接受的概率, 由专家经验决定.

$$S_e = \bigcup_{j=0}^n s_j, p(s_j|C(S_r, s_j)), s_j \in S_h,$$

其中, s_j 指代一个高频字符, S_h 是 s_e 的所有高频修改字符集合. 当用户选择正确结果后, 重译公式字符子模块会替换原有公式中的错误结果, 并向用户输出.

2.7.3 结构编辑

结构编辑需要再分割笔画技术 t_2 和重构公式结构 t_5 的支撑, 其中重构公式结构 t_5 是针对识别结果中错误结构的编辑设计的, 主要通过将错误结构内的字符指向其正确结构的边界框或边界框的横向延展来实现. 如图 4 所示, 其中 A 表示任意字符, 1-6 分别表示与 A 之间的结构关系, 依次是 overscript, superscript, inline (Right), subscript, underscript, inline (Left). 每个字符均有 6 个虚拟边界框, 即结构关系集 ST 包含 6 种结构, 用于相邻字符结构的调整. 在指向过程中, 若输入公式中没有表达关联字符的某些结构关系, 即不存在对应边界框内的笔画, 则 t_5 会自动生成相应的虚拟边界框.

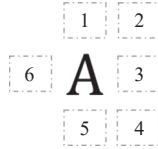


图 4 字符的虚拟边界框

在重构公式结构中, 我们认为除了结构错误外字符本身是完全正确的, 只需调整错误结构内字符与相邻字符的结构关系即可实现结构的重构. 如果结构内字符需要进行分割, 再分割笔画技术 t_2 也可用于分割其对应的笔画. 对于整个重构过程, 错误结构内字符 S_{st} 的正确结构重构依赖于 S_{st} 移动距离和移动方向、移动时相邻字符 S_{adj} 的选择 (水平位置的前一个或后一个)、移动终点与相邻字符虚拟边界框的关系和 S_{st} 与相邻字符的历史结构数据.

$$St(S_{st}, S_{adj}) = \arg \max_{st \in ST} p(st | (dis(S_{st}, S_{adj}), Dir(S_{st}, S_{adj}), R(S_{st}, S_{adj}), St'(S_{st}, S_{adj}))), R(S_{st}, S_{adj}) \in ST, St'(S_{st}, S_{adj}) \in ST,$$

其中, S_{st} 和 S_{adj} 由用户决定, $dis(S_{st}, S_{adj})$ 表示 S_{st} 的移动距离, $Dir(S_{st}, S_{adj})$ 表示 S_{st} 的移动方向, $R(S_{st}, S_{adj})$ 表示移动终点与相邻字符虚拟边界框的关系, $St'(S_{st}, S_{adj})$ 表示 S_{st} 和 S_{adj} 在历史数据中出现频率最高的结构关系.

当用户完成结构内字符调整后, t_5 会纠正公式中的存在错误结构, 并向用户输出.

3 用户研究

为了验证我们方法在人为参与情况下的识别效率, 我们进行了两项比较实验: 1) 与仅能够通过人为修改书写

笔画的 MyScript 进行对比, 用于说明在手写数学公式识别中, 除笔画修改外引入人为参与来处理识别结果的必要性; 2) 与能够编辑识别结果的 InkEquation 进行对比, 用于验证基于理解用户纠错行为的混合方法比依赖符号调色板编辑的方法更高效. 在与 InkEquation 的对比实验中, 我们并不关注底层识别算法的具体实现, 而是聚焦于引入笔画修改和识别结果处理这两类人为参与后识别方法的整体表现. 两项实验中, 我们选择了 20 个数学表达式作为测试任务, 并从准确度、完成时间和主观评价方面进行了比较.

3.1 实验设备和用户界面

(1) 实验设备. 我们的实验评估是在 ThinkPad X1 Yoga gen6 上进行的, 它配备了第 11 代 Intel® Core™ i7-1165G7 CPU (2.80 GHz), 16 GB 内存和分辨率为 1920×1200 的 14 英寸可翻转触摸显示器, 运行专业版微软 Windows 10 操作系统. 该设备支持笔和 10 点触控交互. 在实验过程中, 被试可以根据自身习惯来调节屏幕倾斜度, 以便他们可以舒适地坐在桌面前工作.

(2) 用户界面. 实验评估时使用的用户界面如图 5 所示, 其中图 5 (a) 为 HchMER 和 MyScript 方法的用户界面, 由输入区 (底部)、笔画编辑工具区 (中部) 和公式输出区 (顶部) 这 3 部分组成. 输入区主要用于输入公式和通过对公式笔画的笔交互操作来编辑识别结果, 笔画编辑工具区主要为用户提供对笔画撤销、恢复、清除、擦除 4 种编辑方法, 以及对任务完成时间的控件按钮 (开始和下一个). 公式输出区用于识别结果和用户纠错结果的输出, 输出结果为基于 LaTeX 数学描述语言渲染的标准化排版数学公式. 对于 InkEquation 数学公式输入, 我们使用默认的用户界面, 如图 5 (b) 所示. 同时, 为了更好地记录用户使用 InkEquation 时耗费的时间, 我们设计了用于 InkEquation 计时的界面, 主要包含计时开始、计时结束、下一个、重置和保存功能.



图 5 用户界面

3.2 被试

20 名被试 (11 名男性, 9 名女性) 和 1 名主试同意并参与了实验, 年龄为 22–54 岁 (平均值 $M=31.1$, 标准差 $SD=9.02$). 这些被试来自于该研究团队以外的人员, 招募于当地大学. 除了参与本次实验外, 被试与研究团队没有任何社会关系. 在实验开始之前, 对被试进行了背景调查. 总体而言, 所有人均熟悉触摸设备和笔输入设备, 以及 InkEquation 提供的公式墨迹输入和符号调色板输入工具. 另外, 被试主要从事于软件开发、用户界面和教育等方面的工作和学习.

3.3 过程

实验过程主要从 3 个方面展开, 分别为: (1) 被试在主试的指导下熟悉 3 种方法; (2) 被试分别使用这 3 种方法完成数学表达式的书写任务 (详见第 3.4 节); (3) 被试需要填写一份后问卷调查表, 并就使用方法完成任务时的见解进行半结构化访谈. 后问卷调查表, 见表 1, 主要用于收集被试在准确性、效率、满意度、偏好和操作难易程度等方面的主观评价 (以 5 分制方式, 其中难易程度中分值越大表示越容易操作)^[14,15,54]. 半结构化访谈以当前手写数学表达式的识别算法在哪些方面无法满足使用需求、是否需要为识别算法设计更好的纠错技术、3 种方法在使用过程中的问题, 以及 HchMER 还应该在哪些方面改进为主要访谈问题进行的.

表 1 后问卷调查表 (3 种方法对应的问卷分别填写)

编号	您是否认同以下结论? 评分结果中1表示非常不认同, 5表示非常认同
Q1	您觉得通过InkEquation\HchMER\MyScript能够准确获得手写数学表达式
Q2	您觉得InkEquation\HchMER\MyScript在手写数学表达式识别方面的效率非常高
Q3	您对InkEquation\HchMER\MyScript很满意
Q4	您很喜欢InkEquation\HchMER\MyScript方法
Q5	您觉得InkEquation\HchMER\MyScript很容易操作

3.4 任务和数据收集

在数学表达式的书写任务中, 被试被要求书写表 2 内的 20 个数学表达式 3 次, 每种方法 1 次. 对于每个被试而言, 先使用哪种方法的顺序是交替进行的. 在任务过程中, 被试可以通过方法各自提供的纠错技术进行修改笔画或错误的识别结果, 包括对书写笔画的擦除, InkEquation 对识别结果的编辑, 和使用本文提出纠错方法等. 为了更好地对比 3 种方法, 我们要求被试尽可能在 1 min 内完成每个公式输入和纠错. 需要说明的是, 在这 1 min 内我们并不限制用户的修改方式和修改次数. 如果在 1 min 内已经完成公式输入或者超过 1 min 仍未完成, 被试需要跳过该任务并继续下一个任务. 这样设计的目的在于我们能够通过固定时间内完成公式的数量和准确度, 来更好地衡量 3 种方法的公式输入性能.

在整个实验过程中, 我们记录了 MyScript 方法, HchMER 方法和 InkEquation 在每个数学公式上的识别结果和完成时间, 以及被试完成任务后对 3 种方法的主观评估数据.

4 实验结果

结合用户研究中设计的具体任务, 我们通过分析 3 种方法在准确率、完成时间, 以及被试的主观评价方面的具体表现, 说明 HchMER 在人为参与情况下实现手写数学表达式识别的优越性.

4.1 评价指标及分析方法

我们采用准确率 (Accuracy) 和完成时间 (time to completion, TTC) 评价 3 种方法的具体表现^[14,52]. 准确率是指被试在完成 20 个公式书写任务后, 表达式完全正确的数量与总表达式数量 (=20) 的比值. 在这里, 表达式完全正确是指每个任务的输出结果与其对应的标准结果完全一致. 完成时间也以被试为中心, 由被试使用某种方法完成任务后每个公式的全过程时间的均值计算, 其中每个公式的全过程时间为单个公式从手写输入开始到结束的全部时间, 包括手写输入、识别和纠错时间. 如果某个公式的完成时间超过了 1 min, 则完成时间按 1 min 计算. 在我们收集到的数据中, 仅有 60/1200 条数据时长超过了 1 min.

我们将收集到的数据整理为一个数据集, 其数据量为 3 种方法×20 个公式任务×20 名被试×2 类评价指标. 在后续的统计分析中, 方法为自变量, 2 类评价指标为因变量. 对于因变量准确率和完成时间, 我们以被试为单位分别计算, 即每位被试在使用一种方法完成任务后对应一个准确率和一个完成时间, 其数据集合并为 3 种方法×20 名被试.

对两个因变量的分析, 我们按照相同的处理分别进行, 具体步骤为: 1) 使用 Kolmogorov-Smirnov 检验 ($\alpha = 0.05$) 进行正态性检验; 2) 如果数据呈现正态分布, 我们采用 t 检验进行显著性分析. 如果数据呈现非正态分布, 将采用威尔科克森符号秩检验进行显著性分析. 除了以准确率和完成时间评估 3 种方法外, 我们还对 3 种方法进行了主观性评估, 其统计分析的步骤与两个因变量的处理方法一致.

4.2 与 MyScript 的对比结果

HchMER 与 MyScript 在表达式平均准确率上的对比如图 6(a) 所示, MyScript 的平均准确率为 76.5% (SD=0.086), HchMER 的平均准确率为 94.25% (SD=0.09). 相比于 MyScript, HchMER 在平均准确率上提升了 23.2%. 由于两种方法的准确率数据呈现非正态分布, 我们采用威尔科克森符号秩检验进行了显著性分析, 其检验

结果表明两种方法在准确率上存在显著性差异 ($z=-3.194, p<0.05$). 两种方法在平均完成时间上的对比结果如图 6(b) 所示, MyScript 的平均完成时间为 21.61 s (SD=6.82), HchMER 为 31.21 s (SD=6.73). 相较于 MyScript, HchMER 在单个公式上多花费了大约 10 s 的完成时间 (增加了 44.46%). 由于平均完成时间呈现正态分布, 我们选择采用 t 检验来进行显著性分析. 检验结果表明 MyScript 和 HchMER 在平均完成时间上存在显著性差异 ($t=7.917, p<0.05$). 在不考虑笔画纠错占用时间的情况下, HchMER 纠错的时间占比为 30.7%. 除了识别时间外, MyScript 的交互时间在笔画的重写和修改上, 而 HchMER 还包括采用我们提出的纠错技术来完善和修正结果. 总体来说, 在同样引入人修改时, HchMER 提出的方法比在现有先进的识别方法上进行人为笔画纠错表现得更好, 也进一步阐述了在识别方法的基础上设计更好的后编辑方法是必要的.

表 2 评估使用的数学表达式

编号	表达式	编号	表达式
E1	$n\log(n)$	E11	$\log_2 8 + \log_3 9 + \log_4 16$
E2	$\log_2 8 = 3$	E12	$\frac{\tan \alpha - \tan \beta}{1 + \tan \alpha \tan \beta}$
E3	$\tan z = \frac{\sin z}{\cos z}$	E13	$\pi \int_c^d \{g(y)\}^2 dy$
E4	$\cos \theta_1 + i \sin \theta_1$	E14	$\lim_{n \rightarrow \infty} t_{3n} = \frac{1}{2} \log 2$
E5	$\int_a^b \frac{\sqrt{\pi}}{2} dx$	E15	$\lim_{x \rightarrow \frac{\pi}{2} + 0} \tan x = -\infty$
E6	$\sin^2 \theta + \cos^2 \theta = 1$	E16	$\log z = \log r + i(\theta + 2n\pi)$
E7	$\cos x + i \sin x = e^{ix}$	E17	$\cos(a+b) = \cos a \cos b - \sin a \sin b$
E8	$e = \sum_{k=0}^{\infty} \frac{1}{k!}$	E18	$\lim_{x \rightarrow \infty} \int_0^x e^{-y^2} dy = \frac{\sqrt{\pi}}{2}$
E9	$f \sum_{i=2n+3m}^{10} ix$	E19	$a(n) = \sum_{k=1}^n (-1)^{n-k} k!$
E10	$\sum_{\log 3}^0 \frac{1}{e^t + 1} dt$	E20	$\lim_{n \rightarrow \infty} \frac{1}{n} \sum f\left(\frac{r}{n}\right) = \int_0^1 f(x) dx$

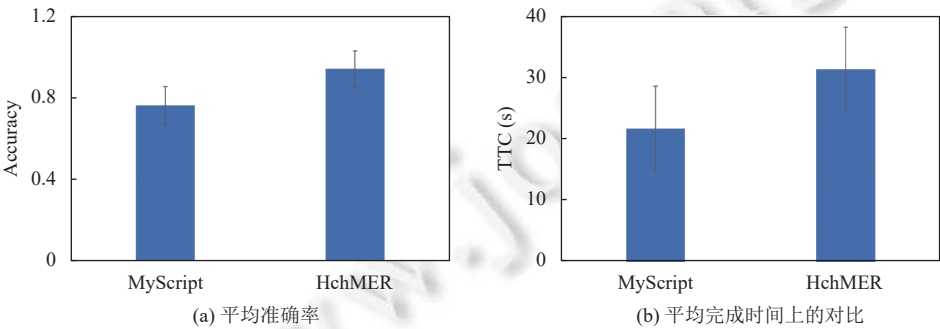


图 6 MyScript 和 HchMER 在平均准确率和平均完成时间上的对比

4.3 与 InkEquation 的对比结果

由 HchMER 与 MyScript 的比较结果得知, 即便 MyScript 能够通过多次笔画纠错来提高识别结果的准确率, 但 MyScript 在任务上的平均准确率仍然不如 HchMER. 为了更进一步验证 HchMER 在人为参与情况下准确率和

完成时间上的具体表现,我们也对比了它与 InkEquation 在具体任务上的表现. InkEquation 与 HchMER 类似,它能够在手写数学表达式识别的基础上结合笔画纠错和基于符号调色板的识别结果再编辑,帮助用户获得更准确的结果.如图 7(a) 所示, InkEquation 的平均准确率为 74.5% (SD=0.108), HchMER 的平均准确率为 94.25% (SD=0.09).相较于 InkEquation, HchMER 在平均准确率上提升了 26.51%. 另外,由于两种方法在准确率上的数据呈现非正态分布,我们采用威尔科克森符号秩检验进行了显著性分析,其结果表明两种方法在准确率上存在显著性差异 ($z=-3.391, p<0.05$). 两种方法在平均完成时间对比结果如图 7(b) 所示, InkEquation 的平均完成时间为 35.26 s (SD=7.26), HchMER 为 31.21 s (SD=6.73). 相较于 InkEquation, HchMER 在平均完成时间上降低了 11.48%. 另外,平均完成时间呈现正态分布,但是其 t 检验结果表明两种方法在平均完成时间上不存在显著性差异 ($t=1.626, p=0.120$). 在与 InkEquation 的对比实验中发现, HchMER 在准确率和完成时间上均优于 InkEquation. 对比结果突出了 HchMER 在人为参与的数学表达式识别方面的优势,也说明了在人为参与情况下提供高效纠错技术的必要性. 由于每个公式的准确率和完成时间与每个被试的计算方式不同. 为此,我们在附录 A 详细说明了每个公式的准确率和完成时间计算方法,以及表 2 所列 20 个数学表达式在 3 种方法上的平均准确率和平均完成时间,见图 A1 和图 A2.

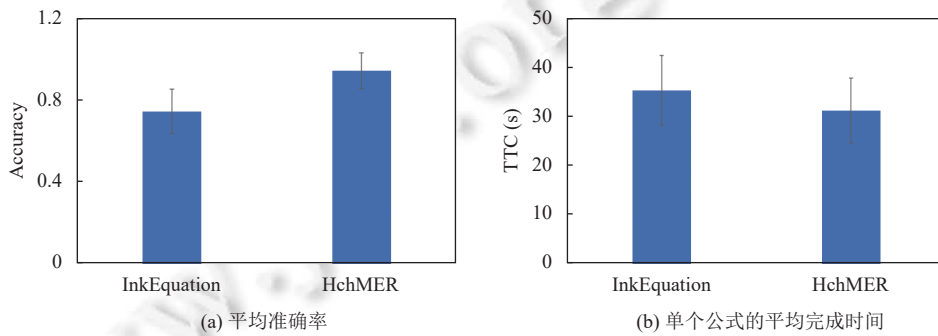


图 7 InkEquation 和 HchMER 在平均准确率和单个公式的平均完成时间上的对比

4.4 用户主观评价

除了在准确率和完成时间上进行比较外,用户的主观评价也用于验证我们方法的优劣性. 3 种方法在准确性、效率、偏好、操作难易程度和满意度上的表现如图 8 所示. 结果显示, HchMER 在主观评价的 5 个方面均优于 InkEquation. 除了在操作难易程度上略低于 MyScript 外,其他 4 个方面也优于 MyScript. 由于主观评价数据中部分数据并不完全服从正态分布,为此,我们采用威尔科克森符号秩进行两两显著性分析. 检验结果表明: MyScript 与 HchMER 两种方法在准确率 ($z=-3.346, p<0.05$)、效率 ($z=-3.286, p<0.05$), 和喜好 ($z=-2.495, p<0.05$) 上存在显著性差异,但在满意度 ($z=-1.532, p=0.073$) 和操作难易程度 ($z=-1.35, p=0.121$) 上不存在显著性差异; InkEquation 与 HchMER 方法在操作难易程度上不存在显著性差异 ($z=-0.293, p=0.769$) 外,其他方面均存在显著性差异 ($p<0.05$),其中 z 值分别为 $z=-2.966$ (准确度)、 $z=-3.355$ (效率)、 $z=-2.913$ (偏好) 和 $z=-2.701$ (满意度). 在主观性评价中,我们发现即便在操作难易程度上略逊色于 MyScript (一种只依赖于笔画纠正的手写数学表达式识别方法),被试还是给予 HchMER 方法积极肯定的评价.

另外,从半结构化访谈中发现被试对 HchMER 提出的重译字符和结构编辑给予了很好的评价. HchMER 被认为能够满足他们在实际应用中的需求,如 P4 在访谈说“我很容易忘记在 InkEquation 的哪个菜单下找到像‘ θ ’这样的字符”,以及 P6 认为“MyScript 这样的方法很好,但是它在识别‘log’‘sin’等字符时效果很差”. 这些观点从侧面印证了为识别结果提供快速准确的纠错方式的必要性. 另外, P7 在访谈中“我感觉 HchMER 在识别错误后给我的修改方法很好,不像 InkEquation 还需要在几个菜单中来回选择”,以及 P12 觉得 HchMER 提供的方法很有趣,表示“我觉得 HchMER 在修改错误结构时比较好玩,比 InkEquation 好. 使用 InkEquation 时,我有时候不知道怎么修改错误的结构”. 这些访谈结果也说明 HchMER 能够高效地帮助被试纠正识别错误,主要原因在于 HchMER 能够快

速提供候选集合用于查找和选择,从而避免了像 InkEquation 这样的公式编辑器需要在提供的字符调色板中多次选择才能找到正确结果.此外,HchMER 结合错误结果、被试纠错行为、知识库等,能够通过理解用户的纠错行为来提供字符和结构的高效编辑,进而简化了操作流程.当然,也有被试反馈了我们方法可能存在的问题,如 P6 提到在使用 HchMER 时遇到的困扰:“在使用 HchMER 时,我有时会忘记应该单击还是拉框来纠正错误,希望以后可以提供一些简单易记的操作方式”.在这方面,我们主要借鉴了笔式用户界面操作中的部分交互方式^[55],在后续的研究中我们会为用户提供一些更加直观的手势或纠错动作.总体来说,HchMER 在准确性、效率、满意度、偏好和难易程度方面获得了更好的主观反馈.

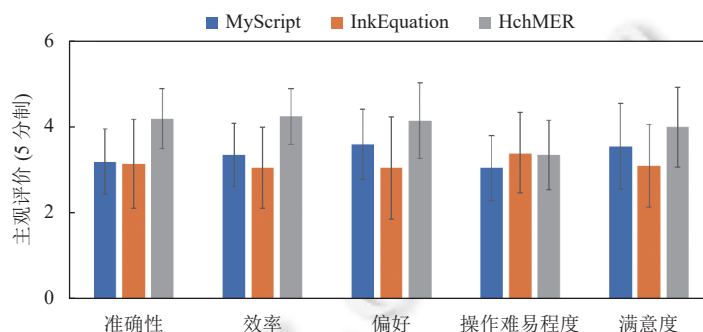


图 8 3 种方法的主观评价对比,其中操作难易程度越高表示越容易操作

最后,被试针对方法提出了宝贵的建议,如“希望能够支持多行公式的识别和编辑”“在与公式信息交互时,可以引入笔手势来缓解单击/长击之间的误操作”,以及“可以自行移动和缩放书写面板,同时也能对公式识别结果的一些细节进行修改”等.基于这些宝贵建议,我们将会在后继工作中继续着力于提升数学公式识别效率和准确率的研究.

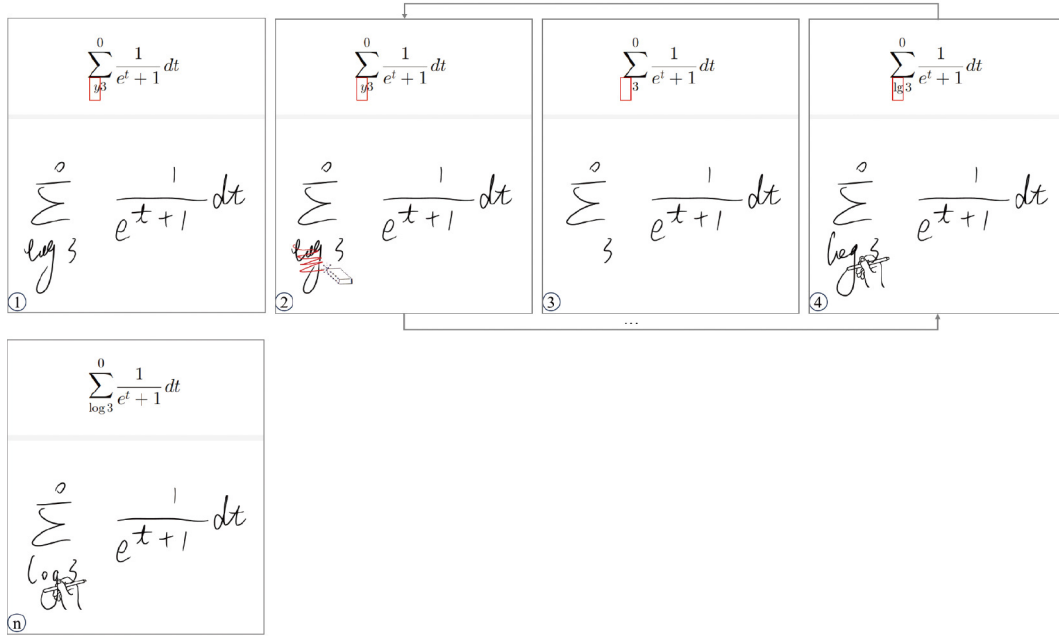
5 讨 论

针对如何在非实验环境中人为参与的情况下快速准确地获得手写数学表达式标准结果的问题,我们以手写错误的直观提示、错误字符的快速推荐和错误结构的高效编辑为目的,设计了一种混合人工智能的手写数学表达式识别算法 HchMER.该算法从用户纠错行为(动作+信息)出发,借助 4 种交互技术和用户纠错意图的理解,结合知识库中先验知识、机器对识别结果的理解,以及对应的纠错信息来混合人和机器的智能,进而快速准确地获取手写数学表达式的标准结果.从两项对比实验发现:HchMER 在准确率上均优于其他两种方法,且其准确率在用户可接受准确率的范围内^[13,14],而在完成时间上优于 InkEquation,略弱于 MyScript.在用户主观评价中发现被试认为 HchMER 在 5 个方面均优于 InkEquation.除操作难易程度外,其他 4 个方面均优于 MyScript.尽管 HchMER 在人为参与的使用中能够准确高效地获取手写公式结果,但是我们在用户实验过程中也发现一些需要讨论的问题.

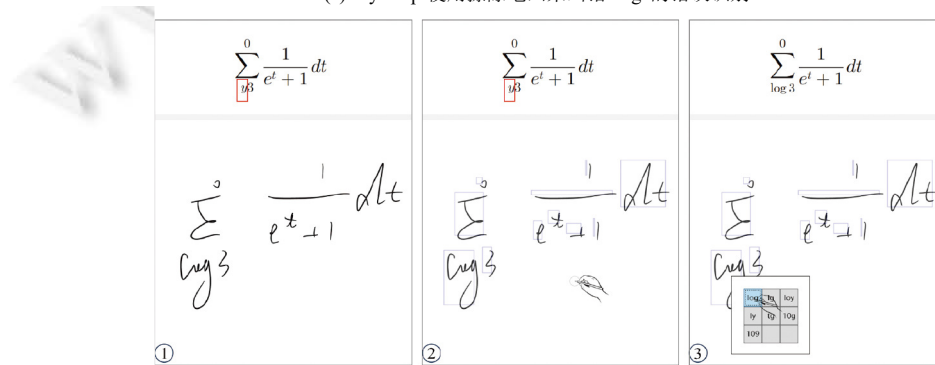
5.1 HchMER 与 MyScript

从两种方法在准确率的表现来看,尽管 MyScript 的识别能力在现有的识别算法中名列前茅,但是在面向相似字符、上下标结构和潦草书写时仍然无法完全正确识别,即使被试多次使用笔画纠错技术进行重写笔画,也是如此.相比于 MyScript, HchMER 结合识别结果、字符识别候选集、关联字符和历史数据,为用户提供更全面的纠错信息和备选项.在上下标结构上, HchMER 在不修改笔画信息的前提下,能够将结构错误的字符移动到其正确结构的边界框或边界框的横向延展位置,避免了算法的二次识别和可能的识别错误.图 9 为 HchMER 和 MyScript 纠正错误字符的示例,其中 MyScript 在面对识别错误时,只能通过擦除和重写笔画后再次识别的方式来获得更准确的结果,而 HchMER 可以通过点击错误识别字符对应的笔画,从推荐的备选集中快速获得准确结

果, 如图 9(b) 中“log”被识别为“y”时, 点击“log”的笔画来快速获得“log”. 另外, HchMER 和 MyScript 在结构错误时的纠错示例如图 10 所示, MyScript 仍只能通过擦除和重写字符来纠错, 而 HchMER 仅需对错误结构内字符重新指向正确位置即可. 如图 10(b), 由“2”在“log”的下标位置, 但“log”在错误结构中不存在下标结构. 为此, HchMER 会根据“log”的结构关系自动生成虚拟边界框以供纠错. 通过对准确率的量化比较可以说明, 针对识别算法可能产生的错误结果设计纠错或编辑技术是必要的.



(a) MyScript使用擦除笔画来纠错“log”的错误识别



(b) HchMER使用快速推荐方法来实现“log”的纠错

图 9 HchMER 和 MyScript 在字符错误时的纠错示例

另外, 尽管给定时间为 1 min, 但是我们发现被试在多次修改笔画信息后, 如果 MyScript 方法仍然无法准确识别, 他们会跳到下一个任务. HchMER 在时间占用上与 MyScript 的不同之处在于, HchMER 并不花费二次识别时间, 即再次识别错误结果的修正笔画时间, 而是通过已识别结果, 关联字符和历史数据直接推理可能的纠错结果, 这使得 HchMER 在纠错时可以花费更少的时间来获得更准确的结果. 总体来说, 与仅依赖于识别和笔画纠错的 MyScript 相比, HchMER 在准确率上表现更佳, 也能够满足用户的基本需求 (>91%), 但是在完成时间上略有不足.

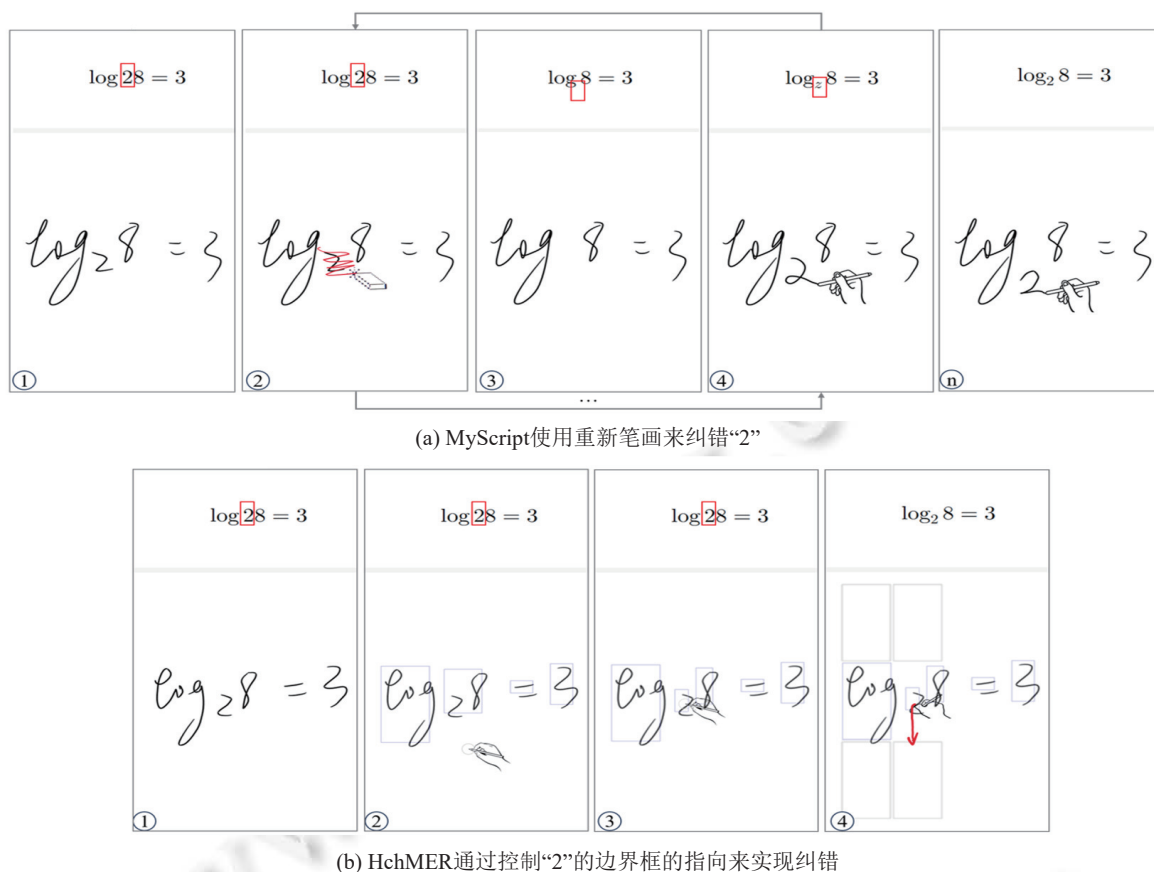


图 10 HchMER 和 MyScript 在处理结构错误时的示例

5.2 HchMER 与 InkEquation

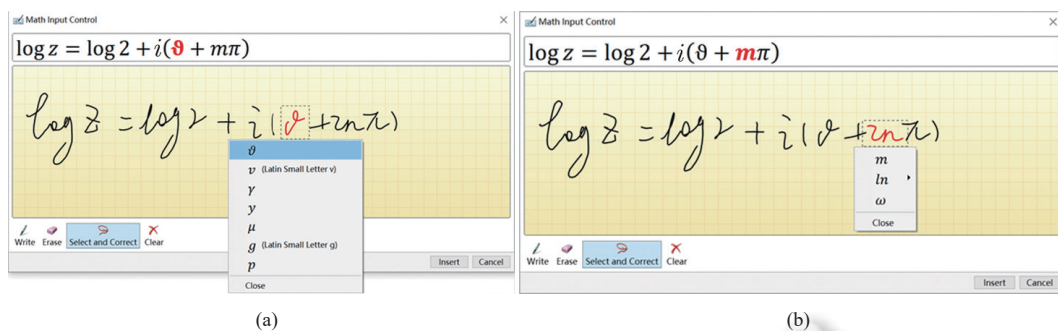
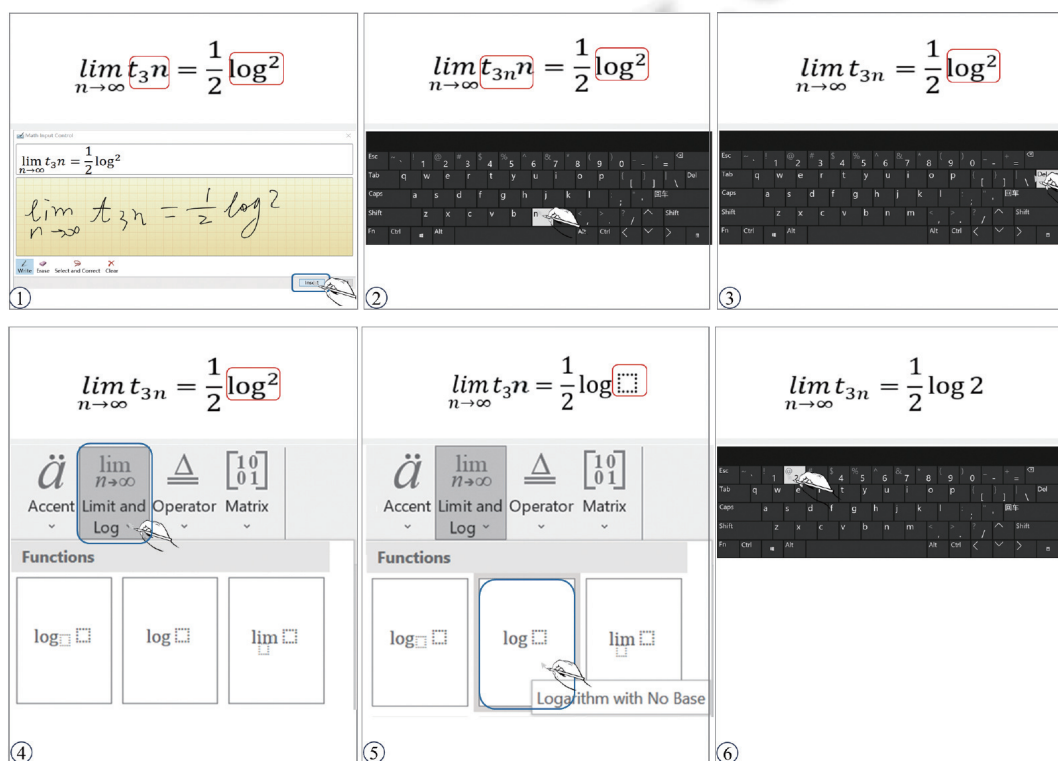
对单个公式而言,我们发现两种方法的手写识别部分均在包含如“log”“sin”等多字符组成的公式符号,如“ θ ”“ α ”等特殊字符,如“9”“z”等相似字符上识别不佳,尤其是这些字符与上下标结构组合时表现更甚。图 11 为两种方法在纠正 $\log z = \log r + i(\theta + 2n\pi)$ 中字符“ θ ”“ r ”“ $2n$ ”和“log”的示意图,其中图 11(a)为 InkEquation 的纠错过程,在不考虑多次重写公式笔画的情况下(即使多次重写也不能保证得到正确结果),可以通过调色板输入或键盘输入来获取错误字符的正确结果。HchMER 的纠错过程如图 11(b)所示,它能够通过单击错误字符的笔画来获取错误字符的推荐结果。尽管 InkEquation 也可以为错误字符提供推荐结果,如图 12 所示,但与 HchMER 不同,它无法结合已识别结果,关联字符和历史数据,而仅依赖识别算法来实现推荐,而这种推荐有时也无法获得准确的结果。在结构纠错上, HchMER 能够通过控制错误结构内字符笔画的指向来快速纠正,例如,图 13(b)中“ $\log^2$ ”纠正为“ $\log 2$ ”时仅需将“2”的边界框指向“log”的边界框即可。但是, InkEquation 对错误结构的纠正特别繁琐,如图 13(a)所示, InkEquation 需要在提供的模板中通过鼠标和键盘多次操作才能完成。

从完成时间来看, InkEquation 在键盘可输入的直线型字符 (inline structure) 上更加快捷,如 E1 和 E17, 被试并不会从符号调色板中的三角/对数函数中选择“sin”“cos”和“log”,但是它在解决两个字符之间的错误结构时很耗时和繁琐,即便如 e^x 向 e^{ix} 这类简单结构的修改也是如此。相反, HchMER 在面向特殊字符和公式结构时更加高效,避免了在模板中选择结构和需要的特殊字符,也避免了用户在鼠标键盘之间的频繁切换。另外我们也发现, HchMER 方法在 E11–E20 这 10 个公式上花费的时间远少于 InkEquation 方法 (减少 15% 以上),这也说明我们方法在长度较长,包含多个复杂结构的公式上表现更优异。

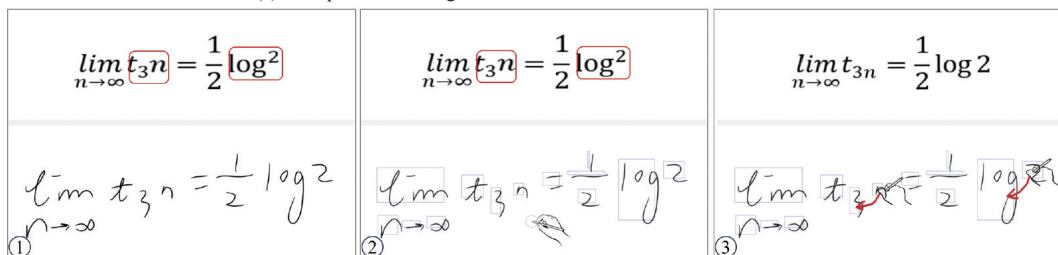


(b) HchMER纠正“ θ ”的过程, “log”和“ $2n$ ”的纠正和HchMER步骤3一致

图 11 InkEquation 和 HchMER 在纠正 $\log z = \log r + i(\theta + 2n\pi)$ 中字符错误的示意图

图 12 InkEquation 为 $\log z = \log r + i(\theta + 2n\pi)$ 中“ θ ”和“ $2n$ ”提供的候选集

(a) InkEquation 纠正“log”和“2”，以及之间“3”和“n”之间的结构错误

(b) HchMER 从“ t_{3n} ”纠正为“ t_{3n} ”，从“ $\log^2$ ”纠正为“ $\log 2$ ”的过程图 13 InkEquation 和 HchMER 在纠正 $\lim_{n \rightarrow \infty} t_{3n} = \frac{1}{2} \log 2$ 中结构错误的示意图

5.3 HchMER 框架与实验设计

在 HchMER 框架中,我们以 MyScript 为识别模块的基础算法,主要原因在于其与当前相关研究中第一的手写数学公式识别算法性能相差无几的同时,能够以开源方式为研究人员提供再利用和优化的机会.与文献 [52] 中 MyScript 方法在测试集上取得的平均准确率 (79.15%) 相比,尽管在我们实验中用户可以通过笔画纠错技术对手写数学表达式进行修改,但 MyScript 的平均准确率仍然降低了 3.35%. 其主要原因在于不同用户的手写风格差异巨大,而且测试集无法囊括每个用户的手写样本,进而导致实际应用中 MyScript 的准确率比测试集上的准确率低.需要说明的是,识别模块在 HchMER 框架是独立的,仅为混合模块提供公式的识别结果,能够被替换为其他手写数学公式识别方法.为了更好地引入人的参与,HchMER 框架的识别模块目前主要针对“笔画分割—字符识别—结构解析”这类经典三阶段结构化方法.在后续的研究中,我们计划尝试为基于端到端的手写数学公式识别方法设计快速纠错和编辑的技术.由于我们提出的用户纠错技术是面向识别模块的输出,其性能可能在不同基础方法上会有所不同.从识别结果的准确程度来看,当识别结果表现不佳时,我们的方法能够帮助用户在字符和结构上快速纠正以获得更准确的结果,而当识别结果表现良好时,用户使用我们方法的频率会降低,或者无需使用提出的纠错技术.在用户纠错模块中,我们使用笔式用户界面设计中的部分交互方式^[55],如单击、长击和拉框等,但是这种方式在针对同一纠错信息的多任务处理时存在一定的局限性,可能会增加纠错任务理解的不确定性.这种不确定性是由可用于交互的动作(有时被称为交互原语)过少引起的,在以后的工作中,可以引入易操作的交互组件或通用笔手势以促进用户的精确交互,或引入笔+触摸交互^[56]的方式.

在实验设计阶段,我们以包含多个结构且常见的公式为实验对象.尽管取得了较好的评估结果,但是任务中的 20 个公式并不能包含所有公式的特性.为此,在后续的工作中我们会继续增加实验任务的多样性,进而为我们方法提供更有力的数据支撑.至于任务设计中为什么要求被试必须在 1 min 内完成单个公式的书写,主要原因在于:(1) 结合手写公式识别和编辑相关的文献^[14,15],以及使用调色板输入方法发现,单个公式输入的平均完成时间基本上不超过 1 min; 2) 在不限时时间的情况下,用户能够花费更多的时间来纠错和编辑以使最终结果完全正确,但是这样只能通过完成时间来对比两种方法,并不能突出使用方法的高效性.在实验评估阶段,我们设计的两项对比实验侧重于比较 HchMER 与在不同阶段引入人为参与的方法间的整体性能,但是我们并没有结合 HchMER 中错误提示、字符纠错和结构编辑进行逐一验证这 3 种技术各自对人为参与情况下识别能力的提高,在后续工作中我们可以做消融实验来分析这 3 类纠错技术对手写公式识别的影响.

不同于在公开数据集上训练和测试识别算法,实际应用中用户引入的不确定性因素是数据集无法完全囊括的,这使有必要进行识别结果的进一步处理和编辑.本文结合用户纠错动作、纠错信息,可能的交互任务和历史数据等进行了基础识别算法的再处理,但是并不能枚举所有的用户行为,如误操作、个性化操作等.在后续工作中,我们可以针对某一类或某些固定格式的用户数据设计一种通用的纠错任务理解框架.

6 结 论

随着各种算法在手写数学表达式领域的应用,手写数学表达式的识别率得到了明显的提升.然而,这些方法在实际应用中人为参与的情况下仍然面临着许多难以解决的识别问题.为此,本文从手写错误的直观提示、错误字符的快速推荐和错误结构的高效编辑这 3 个目的出发,提出了一种混合人工智能的手写数学表达式识别方法 HchMER,以方便用户在手写数学表达式时快速准确地获取识别结果.通过两项对比实验发现:HchMER 在准确率上均优于 MyScript 和 InkEquation,分别提高了 23.2% 和 26.51%. 在单个公式的完成时间上,比 MyScript 增加了 44.46% (大约 10 s),比 InkEquation 降低了 11.48% (大约 4 s).同时,被试在主观评价中对 HchMER 给予了肯定.在后续工作中,我们将继续在手写数学表达式效率、准确率、用户体验方面进行研究.

References:

- [1] Anthony L, Yang J, Koedinger KR. Evaluation of multimodal input for entering mathematical equations on the computer. In: Proc. of the 2005 SIGCHI Conf. on Human Factors in Computing Systems (CHI 2005). Portland: ACM, 2005. 1184–1187. [doi: 10.1145/1056808.]

- 1056872]
- [2] He XY, Zhang ZH. GPK: An efficient special symbol input method for keyboards. In: Proc. of the Extended Abstracts of the 2019 CHI Conf. on Human Factors in Computing Systems. Glasgow: ACM, 2019. 1–6. [doi: [10.1145/3290607.3308457](https://doi.org/10.1145/3290607.3308457)]
 - [3] Topol EJ. High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, 2019, 25(1): 44–56. [doi: [10.1038/s41591-018-0300-7](https://doi.org/10.1038/s41591-018-0300-7)]
 - [4] Mangen A, Anda LG, Oxborough GH, Brännick K. Handwriting versus keyboard writing: Effect on word recall. *Journal of Writing Research*, 2015, 7(2): 227–247. [doi: [10.17239/jowr-2015.07.02.1](https://doi.org/10.17239/jowr-2015.07.02.1)]
 - [5] Pal A, Singh KP. R-GRU: Regularized gated recurrent unit for handwritten mathematical expression recognition. *Multimedia Tools and Applications*, 2022, 81(22): 31405–31419. [doi: [10.1007/s11042-022-12889-x](https://doi.org/10.1007/s11042-022-12889-x)]
 - [6] Lin ZH, Li JR, Yang F, Huang SP, Yang X, Lin JM, Yang M. Spatial attention and syntax rule enhanced tree decoder for offline handwritten mathematical expression recognition. In: Proc. of the 2022 Int'l Conf. on Frontiers in Handwriting Recognition. Hyderabad: Springer, 2022. 213–227. [doi: [10.1007/978-3-031-21648-0_15](https://doi.org/10.1007/978-3-031-21648-0_15)]
 - [7] Álvaro F, Sánchez JA, Benedí JM. An integrated grammar-based approach for mathematical expression recognition. *Pattern Recognition*, 2016, 51: 135–147. [doi: [10.1016/j.patcog.2015.09.013](https://doi.org/10.1016/j.patcog.2015.09.013)]
 - [8] Zhang JS, Du J, Dai LR. A GRU-based encoder-decoder approach with attention for online handwritten mathematical expression recognition. In: Proc. of the 14th IAPR Int'l Conf. on Document Analysis and Recognition. Kyoto: IEEE, 2017. 902–907. [doi: [10.1109/ICDAR.2017.152](https://doi.org/10.1109/ICDAR.2017.152)]
 - [9] Mahdavi M, Condon M, Davila K, Zanibbi R. LPGA: Line-of-sight parsing with graph-based attention for math formula recognition. In: Proc. of the 2019 Int'l Conf. on Document Analysis and Recognition. Sydney: IEEE, 2019. 647–654. [doi: [10.1109/ICDAR.2019.00109](https://doi.org/10.1109/ICDAR.2019.00109)]
 - [10] Wang DH, Yin F, Wu JW, Yan YP, Huang ZC, Chen GY, Wang Y, Liu CL. ICFHR 2020 competition on offline recognition and spotting of handwritten mathematical expressions—OffRaSHME. In: Proc. of the 17th Int'l Conf. on Frontiers in Handwriting Recognition. Dortmund: IEEE, 2020. 211–215. [doi: [10.1109/ICFHR2020.2020.00047](https://doi.org/10.1109/ICFHR2020.2020.00047)]
 - [11] Xie YJ, Mouchère H, Simistira Liwicki F, Rakesh S, Saini R, Nakagawa M, Nguyen CT, Truong TN. ICDAR 2023 CROHME: Competition on recognition of handwritten mathematical expressions. In: Proc. of the 17th Int'l Conf. on Document Analysis and Recognition. San José: Springer, 2023. 553–565. [doi: [10.1007/978-3-031-41679-8_33](https://doi.org/10.1007/978-3-031-41679-8_33)]
 - [12] Wang CY, Luo HQ, Rao XQ, Hu W, Bi N, Tan J. Relative position embedding asymmetric Siamese network for offline handwritten mathematical expression recognition. In: Proc. of the 17th Int'l Conf. on Document Analysis and Recognition. San José: Springer, 2023. 117–133. [doi: [10.1007/978-3-031-41676-7_7](https://doi.org/10.1007/978-3-031-41676-7_7)]
 - [13] Anthony L, Yang J, Koedinger KR. A paradigm for handwriting-based intelligent tutors. *Int'l Journal of Human-computer Studies*, 2012, 70(11): 866–887. [doi: [10.1016/j.ijhcs.2012.04.003](https://doi.org/10.1016/j.ijhcs.2012.04.003)]
 - [14] Kang WH, Huang J, Tian F, Fan XM, Liu J, Dai GZ. Human-in-the-loop based online handwriting mathematical expressions recognition. *Journal of Computer-aided Design & Computer Graphics*, 2021, 33(11): 1773–1785, 1794 (in Chinese with English abstract). [doi: [10.3724/SP.J.1089.2021.18796](https://doi.org/10.3724/SP.J.1089.2021.18796)]
 - [15] Taranta EM, LaViola JJ. Math boxes: A pen-based user interface for writing difficult mathematical expressions. In: Proc. of the 20th Int'l Conf. on Intelligent User Interfaces. Atlanta: ACM, 2015. 87–96. [doi: [10.1145/2678025.2701400](https://doi.org/10.1145/2678025.2701400)]
 - [16] Diaz Y, Nishizawa G, Mansouri B, Davila K, Zanibbi R. The MathDeck formula editor: Interactive formula entry combining LaTeX, structure editing, and search. In: Proc. of the 2021 CHI Conf. on Human Factors in Computing Systems. Yokohama: ACM, 2021. 192. [doi: [10.1145/3411763.3451564](https://doi.org/10.1145/3411763.3451564)]
 - [17] Ma SY, Chen CY, Khalajzadeh H, Grundy J. Latexify math: Mathematical formula markup revision to assist collaborative editing in math Q&A sites. *Proc. of the ACM on Human-computer Interaction*, 2021, 5(CSCW2): 403. [doi: [10.1145/3479547](https://doi.org/10.1145/3479547)]
 - [18] Mouchère H, Viard-Gaudin C, Zanibbi R, Garain U. ICFHR 2014 competition on recognition of on-line handwritten mathematical expressions (CROHME 2014). In: Proc. of 2014 14th Int'l Conf. on Frontiers in Handwriting Recognition. Hersonissos: IEEE, 2014. 791–796. [doi: [10.1109/ICFHR.2014.138](https://doi.org/10.1109/ICFHR.2014.138)]
 - [19] Zhelezniakov D, Zaytsev V, Radyvonenko O. Online handwritten mathematical expression recognition and applications: A survey. *IEEE Access*, 2021, 9: 38352–38373. [doi: [10.1109/ACCESS.2021.3063413](https://doi.org/10.1109/ACCESS.2021.3063413)]
 - [20] Le AD, Nakagawa M. Speedup of parsing for recognition of online handwritten mathematical expressions. In: Proc. of the 14th IAPR Int'l Conf. on Document Analysis and Recognition. Kyoto: IEEE, 2017. 896–901. [doi: [10.1109/ICDAR.2017.151](https://doi.org/10.1109/ICDAR.2017.151)]
 - [21] Nguyen CT, Truong TN, Nguyen HT, Nakagawa M. Global context for improving recognition of online handwritten mathematical expressions. In: Proc. of the 16th Int'l Conf. on Document Analysis and Recognition. Lausanne: Springer, 2021. 617–631. [doi: [10.1007/978-3-030-86331-9_40](https://doi.org/10.1007/978-3-030-86331-9_40)]
 - [22] Le AD, Indurkha B, Nakagawa M. Pattern generation strategies for improving recognition of handwritten mathematical expressions. *Pattern Recognition Letters*, 2019, 128: 255–262. [doi: [10.1016/j.patrec.2019.09.002](https://doi.org/10.1016/j.patrec.2019.09.002)]

- [23] Bian XH, Qin B, Xin XZ, Li JW, Su XF, Wang YF. Handwritten mathematical expression recognition via attention aggregation based bi-directional mutual learning. In: Proc. of the 36th AAAI Conf. on Artificial Intelligence. AAAI Press, 2022. [doi: [10.1609/aaai.v36i1.19885](https://doi.org/10.1609/aaai.v36i1.19885)]
- [24] Zhang JS, Du J, Zhang SL, Liu D, Hu YL, Hu JS, Wei S, Dai LR. Watch, attend and parse: An end-to-end neural network based approach to handwritten mathematical expression recognition. Pattern Recognition, 2017, 71: 196–206. [doi: [10.1016/j.patcog.2017.06.017](https://doi.org/10.1016/j.patcog.2017.06.017)]
- [25] Zhang T, Mouchère H, Viard-Gaudin C. A tree-blstm-based recognition system for online handwritten mathematical expressions. Neural Computing and Applications, 2020, 32(9): 4689–4708. [doi: [10.1007/s00521-018-3817-2](https://doi.org/10.1007/s00521-018-3817-2)]
- [26] Truong TN, Ung HQ, Nguyen HT, Nguyen CT, Nakagawa M. Relation-based representation for handwritten mathematical expression recognition. In: Proc. of the 2021 Int'l Conf. on Document Analysis and Recognition. Lausanne: Springer, 2021. 7–19. [doi: [10.1007/978-3-030-86198-8_1](https://doi.org/10.1007/978-3-030-86198-8_1)]
- [27] Li Z, Jin LW, Lai SX, Zhu YC. Improving attention-based handwritten mathematical expression recognition with scale augmentation and drop attention. In: Proc. of the 17th Int'l Conf. on Frontiers in Handwriting Recognition. Dortmund: IEEE, 2020. 175–180. [doi: [10.1109/ICFHR2020.2020.00041](https://doi.org/10.1109/ICFHR2020.2020.00041)]
- [28] Zhang JS, Du J, Dai LR. Multi-scale attention with dense encoder for handwritten mathematical expression recognition. In: Proc. of the 24th Int'l Conf. on Pattern Recognition. Beijing: IEEE, 2018. 2245–2250. [doi: [10.1109/ICPR.2018.8546031](https://doi.org/10.1109/ICPR.2018.8546031)]
- [29] Fu YN, Liu TT, Gao M, Zhou AY. EDSL: An encoder-decoder architecture with symbol-level features for printed mathematical expression recognition. In: Proc. of the 2023 Int'l Conf. on Document Analysis and Recognition. San José: Springer, 2023. [doi: [10.1007/978-3-031-41676-7_8](https://doi.org/10.1007/978-3-031-41676-7_8)]
- [30] Ding HS, Chen K, Huo Q. An encoder-decoder approach to handwritten mathematical expression recognition with multi-head attention and stacked decoder. In: Proc. of the 16th Int'l Conf. on Document Analysis and Recognition. Lausanne: Springer, 2021. 602–616. [doi: [10.1007/978-3-030-86331-9_39](https://doi.org/10.1007/978-3-030-86331-9_39)]
- [31] Fu YN, Cai WY, Gao M, Zhou AY. Symbol location-aware network for improving handwritten mathematical expression recognition. In: Proc. of the 2023 ACM Int'l Conf. on Multimedia Retrieval. Thessaloniki: ACM, 2023. 516–524. [doi: [10.1145/3591106.3592259](https://doi.org/10.1145/3591106.3592259)]
- [32] Zhang XY, Ying H, Tao Y, Xing YL, Feng GH. General category network: Handwritten mathematical expression recognition with coarse-grained recognition task. In: Proc. of the 2023 IEEE Int'l Conf. on Acoustics, Speech and Signal Processing. Rhodes Island: IEEE, 2023. 1–5. [doi: [10.1109/ICASSP49357.2023.10097228](https://doi.org/10.1109/ICASSP49357.2023.10097228)]
- [33] Liu Z, Yuan Y, Ji ZL, Bai JF, Bai X. Semantic graph representation learning for handwritten mathematical expression recognition. arXiv: 2308.10493, 2023.
- [34] Alvaro F, Zanibbi R. A shape-based layout descriptor for classifying spatial relationships in handwritten math. In: Proc. of the 2013 ACM Symp. on Document Engineering. Florence: ACM, 2013. 123–126. [doi: [10.1145/2494266.2494315](https://doi.org/10.1145/2494266.2494315)]
- [35] Hunsinger J, Lang M. A speech understanding module for a multimodal mathematical formula editor. In: Proc. of the 2000 IEEE Int'l Conf. on Acoustics, Speech, and Signal Processing. Istanbul: IEEE, 2000. 2413–2416. [doi: [10.1109/ICASSP.2000.859328](https://doi.org/10.1109/ICASSP.2000.859328)]
- [36] Mansouri B, Zanibbi R, Oard DW. Characterizing searches for mathematical concepts. In: Proc. of the 2019 ACM/IEEE Joint Conf. on Digital Libraries. Champaign: IEEE, 2019. 57–66. [doi: [10.1109/JCDL.2019.00019](https://doi.org/10.1109/JCDL.2019.00019)]
- [37] Nishizawa G, Liu J, Diaz Y, Dmello A, Zhong W, Zanibbi R. MathSeer: A math-aware search interface with intuitive formula editing, reuse, and lookup. In: Proc. of the 42nd European Conf. on Information Retrieval. Lisbon: Springer, 2020. 470–475. [doi: [10.1007/978-3-030-45442-5_60](https://doi.org/10.1007/978-3-030-45442-5_60)]
- [38] Varni G, Pez AM, Mancini M. Get together in the middle-earth: A first step towards hybrid intelligence systems. In: Proc. of the Companion Publication of the 2021 Int'l Conf. on Multimodal Interaction. Montreal: ACM, 2021. 249–253. [doi: [10.1145/3461615.3485413](https://doi.org/10.1145/3461615.3485413)]
- [39] Dellermann D, Ebel P, Söllner M, Leimeister JM. Hybrid intelligence. Business & Information Systems Engineering, 2019, 61(5): 637–643. [doi: [10.1007/s12599-019-00595-2](https://doi.org/10.1007/s12599-019-00595-2)]
- [40] Kamar E. Directions in hybrid intelligence: Complementing AI systems with human intelligence. In: Proc. of the 25th Int'l Joint Conf. on Artificial Intelligence. New York: AAAI Press, 2016. 4070–4073.
- [41] Prakash N, Mathewson KW. Conceptualization and framework of hybrid intelligence systems. arXiv:2012.06161, 2020.
- [42] Johnson M, Vera AH. No AI is an island: The case for teaming intelligence. AI Magazine, 2019, 40(1): 16–28. [doi: [10.1609/aimag.v40i1.2842](https://doi.org/10.1609/aimag.v40i1.2842)]
- [43] Xie Y, Chen M, Kao D, Gao G, Chen X. CheXplain: Enabling physicians to explore and understand data-driven, AI-enabled medical imaging analysis. In: Proc. of the 2020 CHI Conf. on Human Factors in Computing Systems. Honolulu: ACM, 2020. 1–13. [doi: [10.1145/3313831.3376807](https://doi.org/10.1145/3313831.3376807)]
- [44] So C. Human-in-the-loop design cycles—A process framework that integrates design sprints, agile processes, and machine learning with humans. In: Proc. of the 1st Int'l Conf. on Human-computer Interaction. Copenhagen: Springer, 2020. 136–145. [doi: [10.1007/978-3-030-](https://doi.org/10.1007/978-3-030-)]

- 50334-5_9]
- [45] Lee MH, Siewiorek DP, Smailagic A, Bernardino A, Badia SBBI. A human-AI collaborative approach for clinical decision making on rehabilitation assessment. In: Proc. of the 2021 CHI Conf. on Human Factors in Computing Systems. Yokohama: ACM, 2021. 392. [doi: 10.1145/3411764.3445472]
- [46] Tsaknaki V, Cotton K, Karpashevich P, Sanches P. “Feeling the sensor feeling you”: A soma design exploration on sensing non-habitual breathing. In: Proc. of the 2021 CHI Conf. on Human Factors in Computing Systems. Yokohama: ACM, 2021. 266. [doi: 10.1145/3411764.3445628]
- [47] Elvitigala DS, Huber J, Nanayakkara S. Augmented foot: A comprehensive survey of augmented foot interfaces. In: Proc. of the 2021 Augmented Humans Int'l Conf. Rovaniemi: ACM, 2021. 228–239. [doi: 10.1145/3458709.3458958]
- [48] Clark E, Ross AS, Tan CH, Ji YF, Smith NA. Creative writing with a machine in the loop: Case studies on slogans and stories. In: Proc. of the 23rd Int'l Conf. on Intelligent User Interfaces. Tokyo: ACM, 2018. 329–340. [doi: 10.1145/3172944.3172983]
- [49] Gerlach M, Miller M, Ho R, Harlan K, Difallah D. Multilingual entity linking system for Wikipedia with a machine-in-the-loop approach. In: Proc. of the 30th ACM Int'l Conf. on Information & Knowledge Management. ACM, 2021. 3818–3827. [doi: 10.1145/3459637.3481939]
- [50] Qosim H, Zulkarnain. Fault detection system using machine learning on synthesis loop ammonia plant. In: Proc. of the 3rd Asia Pacific Conf. on Research in Industrial and Systems Engineering. Depok: ACM, 2020. 74–80. [doi: 10.1145/3400934.3400950]
- [51] Li Y, Kang WH, Huang J, Yuan XF, Chen XJ, Yu Q, Wu HY, Tian F, Dai GZ. Human-computer hybrid framework based on pen-based interaction for CAD conceptual design. Computer Integrated Manufacturing Systems, 2022, 28(7): 2050–2063 (in Chinese with English abstract). [doi: 10.13196/j.cims.2022.07.011]
- [52] Mahdavi M, Zanibbi R, Mouchere H, Viard-Gaudin C, Garain U. ICDAR 2019 CROHME + TFD: Competition on recognition of handwritten mathematical expressions and typeset formula detection. In: Proc. of the 2019 Int'l Conf. on Document Analysis and Recognition. Sydney: IEEE, 2019. 1533–1538. [doi: 10.1109/ICDAR.2019.00247]
- [53] MyScript. AI, neural networks and handwriting recognition. 2023. <https://www.myscript.com/ai/>
- [54] Zhelezniakov D, Cherniha A, Zaytsev V, Ignatova T, Radyvonenko O, Yakovchuk O. Evaluating new requirements to pen-centric intelligent user interface based on end-to-end mathematical expressions recognition. In: Proc. of the 25th Int'l Conf. on Intelligent User Interfaces. Cagliari: ACM, 2020. 212–220. [doi: 10.1145/3377325.3377482]
- [55] Fan YT, Teng DX, Ma CX, Yang HY, Dai GZ, Wang HA. A model driven development method for pen-based form interface software. Journal of Computer Research and Development, 2012, 49(12): 2671–2685 (in Chinese with English abstract).
- [56] Hinckley K, Yatani K, Pahud M, Coddington N, Rodenhouse J, Wilson A, Benko H, Buxton B. Pen + touch = new tools. In: Proc. of the 23rd Annual ACM Symp. on User Interface Software and Technology. New York: ACM, 2010. 27–36. [doi: 10.1145/1866029.1866036]

附中文参考文献:

- [14] 康文惠, 黄进, 田丰, 范向民, 刘杰, 戴国忠. 人在回路的在线手写数学公式识别方法. 计算机辅助设计与图形学学报, 2021, 33(11): 1773–1785, 1794. [doi: 10.3724/SP.J.1089.2021.18796]
- [51] 李洋, 康文惠, 黄进, 袁晓芳, 陈雪锦, 于茜, 武汇岳, 田丰, 戴国忠. 基于笔式交互的人机混合 CAD 概念设计. 计算机集成制造系统, 2022, 28(7): 2050–2063. [doi: 10.13196/j.cims.2022.07.011]
- [55] 樊银亨, 滕东兴, 马翠霞, 杨海燕, 戴国忠, 王宏安. 一种模型驱动的笔式表单界面软件开发方法. 计算机研究与发展, 2012, 49(12): 2671–2685.

附录 A

由于每个公式的准确率和完成时间与每个被试的计算方式不同. 每个公式的计算方式为: 准确率 (简称 EAccuracy) 是指 20 名被试完成某个公式书写任务后, 表达式完全正确的数量与总表达式数量 (=20, 见正文表 2, 编号是 E1–E20) 的比值. 完成时间 (简称 ETTC) 也是以公式为中心, 由 20 名被试完成某个公式的全过程时间的平均值计算得到.

图 A1 所示 3 种方法在每个公式上的平均准确率. 对于 HchMER 而言, E5 的平均准确率低于 MyScript 和 InkEquation, E1 和 E2 的平均准确率与其他两种方法相接近. 总体而言, HchMER 在 20 个公式上的平均准确率达到 94.25% (SD=0.085).

图 A2 所示 3 种方法在每个公式上的平均完成时间. 从结果来看, MyScript 在完成每个公式时花费的时间最

少, InkEquation 最多. 在笔画数比较少的公式上, HchMER 和 MyScript 所花费的时间更短, 如 E1–E5. 需要说明的是, HchMER 和 MyScript 在每个公式的平均完成时间与每个被试的略有不同. 总体而言, HchMER 在 20 个公式上的平均完成时间为 31.06 s (SD=11.90).

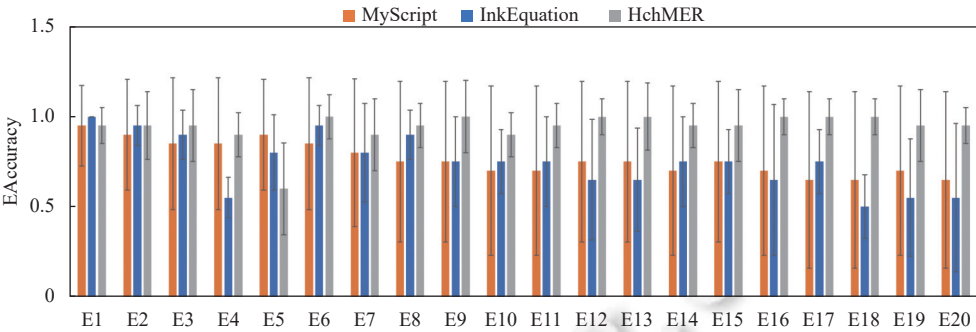


图 A1 表 2 所列 20 个数学表达式在 3 种方法上的平均准确率

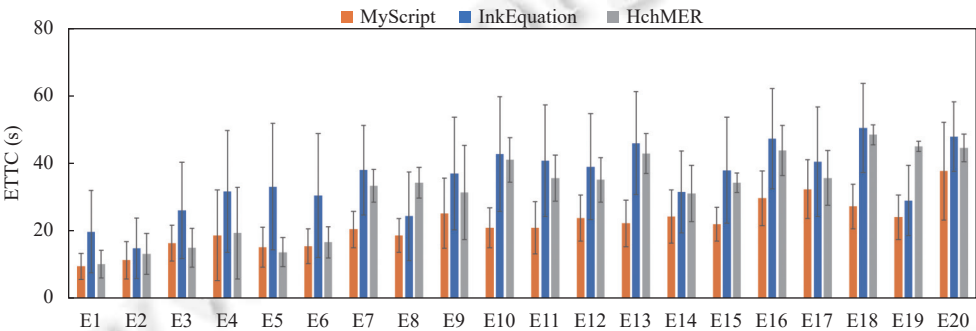


图 A2 表 2 所列 20 个数学表达式在 3 种方法上的平均完成时间



康文惠(1993—), 男, 博士, 特别研究助理, 主要研究领域为人机混合智能, 草图理解, 手写公式识别.



付强(1991—), 男, 博士生, 主要研究领域为人机交互, 草图理解.



黄进(1985—), 男, 博士, 副研究员, CCF 专业会员, 主要研究领域为人机交互, 图形图像处理.



田丰(1976—), 男, 博士, 研究员, 博士生导师, CCF 杰出会员, 主要研究领域为人机交互, 虚拟现实.



仝青山(1983—), 男, 博士, 讲师, 主要研究领域为人机交互, 软件工程.



戴国忠(1944—), 男, 研究员, 博士生导师, CCF 杰出会员, 主要研究领域为人机交互, 计算机图形学.