

AutoTutor 背后的技术启思与人文眷注

——访美国孟菲斯大学智能导学系统专家亚瑟·格雷泽教授

本刊特邀记者 刘凯 王韶 隆舟 屈静

[摘要] 亚瑟·格雷泽(Arthur C. Graesser)是美国孟菲斯大学心理学系和智能系统研究院教授,牛津大学教育系荣誉研究员,先后获得杰出科学贡献奖(文本与语言学会)、心理学应用教育和培养杰出贡献奖(美国心理学会)、孟菲斯大学终身成就奖、Harold W. McGraw 学习科学奖等,与同事合作开发了 AutoTutor、DeepTutor、Electronix Tutor、PAL3 及 Coh-Metrix 等知名教学软件,研究兴趣为认知科学、言语处理和学习科学。本文系格雷泽教授 2019 年 10 月来华参加“计算行为科学研讨会——智适应教学系统的模型和分析”会议期间接受访谈整理而成。访谈以 AutoTutor 为切入点,探讨与之有关的教师培训、情感、协作学习、深度学习、GIFT 及 MOOC 等问题。

我国改革开放迄今已四十余年,但是与美国相比,为何原创总是乏善可陈? AutoTutor 背后的技术启思与人文眷注——格雷泽教授的心路分享也许能给我们“精神引航”。

此次访谈恰逢格雷泽教授退休之际,借学术回望之机管窥其学术思想内涵,希望能对学界同仁有所启发。

[关键词] 智能导学系统;深度学习;人机对话;情绪;自动导师;通用智能导学框架;慕课

[中图分类号] G434 **[文献标识码]** A **[文章编号]** 1007-2179(2020)02-0004-09



记者:格雷泽教授,您好。很高兴您能接受我们的学术专访。您在教育技术学领域,尤其在智能导学系统(ITS)研究领域,可谓名闻遐迩。今天访谈的主题主要围绕三方面展开:一是基于个人视角概述您的研究工作;二是请您回应几个热点问题,因为您的经验和见解也许能对中国同仁有帮助和启发;三是请您简评自己的学术生涯。

格雷泽教授:在全球范围内,学校都在应对数字技术进课堂的挑战,但教师通常没有作好准备,这个

挑战异常严峻。虽然好于传统教学的解决方案不少,但只要教师没有理解、接受或使用,那么一切都将枉然。所以,教师培训是重中之重。此外,情绪管理、小组探究或互动指导等有益于教学的非技术方法也很多,但教师仍缺乏足够的训练。教师很少接受如何引导和利用学生情绪或协作解决问题的培训。他们不了解信息技术不足为奇,但我们面对的挑战是巨大的。

中国也许可以解决这些问题,通过师资培训或职业发展灵活、有针对性地提升教师能力。但是,美

[收稿日期]2020-01-21 **[修回日期]**2020-02-26 **[DOI 编码]**10.13966/j.cnki.kfjyyj.2020.02.001

[基金项目]国家自然科学基金项目“面向青少年网络适应的个性化信息服务优化方法研究”(71974072);教育部人文社会科学基金项目“‘职业仓’:基于招聘大数据的询证模型构建及其应用研究”(20YJC880056);湖南省教育厅科学研究项目“智能导学系统调节师范生问题解决中的困惑情绪研究”(18B491)。

[作者简介]刘凯,博士,硕士生导师,渤海大学教育科学学院,研究方向:机器教育、通用人工智能(ccnulk@mail.ccnu.edu.cn);王韶,硕士研究生,渤海大学教育科学学院,研究方向:教育数据挖掘、智能导学系统(916730573@qq.com);隆舟,博士,怀化学院教育科学学院,研究方向:智慧教育(jojolong@foxmail.com);屈静(通讯作者),博士研究生,北京大学教育学院,研究方向:智能导学系统、教育经济学(lucyqujing@pku.edu.cn)。

国的发展步子非常缓慢,原因可能是:一,教师工资普遍偏低,教育职业缺乏吸引力。相对而言,中国教师的平均收入更可观;二,很多教师虽然接受过培训,但培训内容未涉及信息技术。为了学习新知或提升新技能,他们需要通过进修不断充电;三,学生是社交媒体、电视、电影院、计算机和数字游戏的“原住民”,师生代际差异也是难点之一。

再谈谈情绪,这是极为重要的问题。人非草木孰能无情,但直到最近十年,人们才真正了解如何应对学习者情绪。过去,教师没有受过指导如何处理自己的情绪,也没有正确地关注情绪。他们了解气愤、欺凌以及兴奋和烦恼,但不明白困惑、挫败和无聊。比如,我和同事悉尼(Sydney)及德梅罗(Demerol)对比研究学习者主观情绪感受与教师和专家观测之间的差异,经常发现一个问题:询问学生感受时,他们会说“我觉得很无聊”,但教师却认为学生此时正积极参与。实际上,半数情况下,学习者感到无聊时,教师都以为他们正在积极思考。其原因部分在于,人们不会将无聊写在脸上。我们发现,无聊、集中思想或不无聊也不集中思想时,人们的表情没有显著差别。所以,教师观察学生的面部表情,并不总能起作用。困惑具有显著表情标记——好似一头雾水。事实上,困惑和惊奇情绪一看便知,但挫败感不易察觉。由于教师没有受过如何阅读情绪的训练,他们通常只能觉察生气、愤怒和悲伤等生存性情绪,却不知如何辨识与学习密切相关的情绪。

而且,教师不仅阅读情绪有困难,回应情绪也存在不足。事实上,我曾在这个问题上申请过专利。以困惑为例,我们都希望学生能够自主思考,也希望学生感到困惑,因为困惑能促进深度思考,但教师的想法是“既然他们感到困惑,我最好把问题变得简单一点,或者直接公布答案吧”。教师希望摆脱困惑局面,但这绝非处理困惑情绪的最佳方法,反倒应希望学生盘桓在一种理想的困惑中。不过,应对持续的困惑并不容易。如果学生正在汲取知识,我们希望他们保持困惑,因为他们可能由此作出有益的推理及解决问题的尝试。当他们攻克难题,会非常欣喜。如果能力不足,就可能放弃。尽管我们对此已有所知,但世界上几乎没有教师受过相应的培训,索性可以编制电脑程序完成这项任务,或通过 Auto-

Tutor 系统训练人类教师。

记者: AutoTutor 系统能帮助教师开展教学?

格雷泽教授: 是的。我们可以通过数字技术训练教师。例如,系统展示不同情绪的学生,然后让教师预测相应情绪,以此练习情绪识别。同样,数字技术也可以训练教师回应情绪。运用数字技术助力职业发展是个全新的领域,尽管它还处于起步阶段,但很有发展潜力。

在团队中,个人提议不会总被他人采纳,所以困惑和失望情绪很常见。这也是小组学习常有负面效果的原因。例如,皮埃尔·迪伦巴克(Pierre Dillenbourg)曾记录团队情绪变化轨迹。在团队成立初期,成员间交流常带来积极影响,大家因此相互了解,一切都很和谐。然而,当团队目标确定后,成员间可能对同一问题提出不同的解决方案,而方案的复杂性和视角的多样性往往很难用某种简单的办法解决。当观点冲突严重时,团队就会放弃尝试甚至解散。不过,如果成员间最终达成共识,此时又会带来积极影响。

因此,情绪和团队活动非常重要。我们应该如何制定社交互动的规范和策略?比如,团队活动中,我们不直接舍弃个人提议,而是吸收其中的恰当内容,这样建议者就会觉得自己作了贡献。试想,如果某人分享一个为之骄傲的好主意,那么在得到反馈“这主意真是太糟糕了!我们才不会这么做”与“我喜欢你提议中的第三步,应该把它考虑进去”之间,当事人的感受会有天壤之别。即使该倡议不是最终方案,只要可能是其中的一部分,上述规范和策略就能促进团队协作。该领域前景广阔,毕竟全世界都想知道如何才能实现有效的团队合作,但相关研究少得可怜。常见的例子是,团队的两个成员意见产生分歧时,应该如何应对和处理?遗憾的是,人们没有受过指导该如何解决协作问题,世界上也没有专门教授协作的课程。事实上,学校应当提供相关培训,方法之一便是运用虚拟会话代理,将受训者编入虚拟小组中,尝试解决争端。我们可以预设多种角色,制造多种冲突,观察受训者如何处理,还可以设置旁观者。而这些都能通过虚拟代理加以引导和训练,教师和学生均可使用。这一方向仍是有待探索的处女地。

记者:这需要使用对话分析,您曾提出一对一虚拟对话教学五步骤,在多人情况下是否也有章可循?

格雷泽教授:是的,可以用虚拟代理训练学生或教师如何使用语言开展小组互动。我总结了不同情境的八条产生式规则,这里列举几项。假设团队中有个学生不爱说话,缺乏行动的勇气或主动性,我们就可在团队中增设一位虚拟导师。

其中的规则之一是,如果某人不说话,导师会主动问他“你怎么看?”“你的任务完成了吗?”或“你还有任务没完成吗?”,目的是将其置于任务焦点下。而对那些滔滔不绝的学生,导师可以说“你说得对。不过,我很好奇其他人怎么看?”,从而扩散会话,了解其他成员的想法。假如大家没有专心解决问题,只是在闲聊,这时可以让虚拟代理说“好的,现在请大家回到手头的工作上来”。今天,借助人工智能技术,我们能实时监测某人的发言是否与主题相关。如果持续闲聊,虚拟导师会打断并指出“我不知道这与主题有什么关系”。另一个问题是,一旦有人说脏话或者言辞过激,即发言不当或不专业,系统会提醒“请别说这种话,我们还有正事要做呢”。

当小组成员的意见发生分歧时,解决办法之一是找到分歧的共同点。假设在某方面两人都同意,但其他方面有异议,那么接下来需要明晰的是,赞同与不赞同的内容。其实,只要组员共同厘清这些异同,他们自然而然会形成某种共识。其中有多种对话技巧,都是人工智能和自然语言理解可利用的。我们可以使用虚拟代理灵活驱动,经济又高效。

记者:也就是说,虚拟代理能促进人们之间更好地沟通交流。那么,相对于一对一辅导,团队导学系统是否有所不同?

格雷泽教授:是的,这是个很好的问题,因为两者侧重点不同。我的博士生曾作过研究,对比传统课堂、一对一辅导和小组互动中学生提问的数量。结果发现,小组提问数量最多,超过一对一辅导,远多于课堂——因为课堂几乎为零。颇具讽刺的是,它发生在美国。美国统计数据发现,学生在课堂中每小时提问数量平均仅 0.11 个!换言之,每名学生大概九到十小时才会问一个问题。然而,小组互动

中每小时却高达 20 个。所以,主动性或提问可以当作衡量标准开展研究。这位博士后来做博士后,目前就职于 ACT^①,年薪 15 万美元。很显然,知识就是金钱,前提是能够找到并真正解决问题。

记者:对研究新手而言,这既受启发又相当励志。那么,以 AutoTutor 为例,在一对一辅导中,智能导学系统在教学中只需建立畅通的教学对话即可,但在团队教学中,是不是系统本身也需要参与其中,有角色参与并起到协调作用?

格雷泽教授:对,这就是为什么把它叫做 Tutor 而不是 Mentor 或 Monitor 的原因。在理想状态下,虚拟导师的风格应该不同:有的是微观管理者,催促人们做事;有的负责观察学生、观察互动,在互动没有进展时发言推进讨论。假设某个小组正在协作解决问题或设计某种产品,但突然因某种原因停滞不前了,这时虚拟导师会出面提供信息,帮助小组攻克难关。有人认为这种导师比微观管理者更好,因为学生会得到激励并保持主动。布伦特·摩根(Brent Morgan)等人的文章^②,讨论了不同导师的策略,可供参考。

关于团队教学,我想强调一个重要问题,那就是,现在的问题比过去复杂,需要更深层的理解和学习,一人往往独木难支。因此,采用团队协作促进学习者的深层学习仍是所有人的共同愿景。这有三种实现方式:

一、在个体层面上,通过一对一辅导改善学生的心智模式。这可借助 AutoTutor 或其他智能导学系统完成。

二、在团队层面上,通过协作解决问题或制作物品。比如,极具挑战的“创客运动”,无人能独自完成,甚至连答案都无从知晓。这通常应由不同技能的人组成团队共同探讨,也只有在团队中才能获得深入的解决方案。

三、选择“真正的问题”。所谓“真正的问题”是指能激发人们内在动力开展研究和解决的问题。举例说,假设某学校帮派和暴力问题严重,不论作为整体还是其中的小团体,学生都会主动寻求办法避免陷入帮派纠纷,这就是个“真正的问题”。于是,学生对某件事越有动力,会钻研得越深,越可能探查到更深层的知识。有趣的是,今天的美国人正奋力研

究如何摆脱特朗普(Trump),且已经学习到很多关于政府的知识。

记者:这真是个值得深入探索的领域。在智能导学系统中团队往往以小组出现,那么小组辅导的理论挑战是什么?

格雷泽教授:一是小组如何运作,二是如何提升小组协作质量。我是协作问题解决能力评价专家团主席。我与他人合作撰写了涵盖各种团队理论的论文,作为2015年国际学生评价项目(PISA)的成果。就目前来说,理论还在发展完善中,不过首要问题是如何在协作问题解决上训练学生。

帕特里克·格里芬(Patrick Griffin)是澳大利亚研究者,2007—2015年间,他的研究团队尝试训练学生的协作技能,但由于缺乏经费未能继续。幸运的是,他出版了这方面的著作。

还有斯蒂芬·菲奥里(Stephen Fiore),我们合作在《Nature》子刊发表了讨论训练团队协作解决问题的论文^③。文章讨论的是,有时分组解决问题反而不会得到反馈。比如,学生分组完成项目时,教师拍拍头表扬一句“不错”,但这不是详细的反馈,因为教师不清楚该如何推进合作。

鉴于反馈的重要性,许多人将目光投向智能导学系统,希望它能教导如何成为优秀的合作者。事实上,美国海军已经尝试设计了此类智能导学系统基本框架,形成了通用智能导学框架(Generalized Intelligent Framework for Tutoring,简称GIFT)开源项目^④。关于协作解决问题的基本理论,你们可以参考那篇论文。它囊括了大部分相关文献。但是,对于如何训练协作解决问题的研究仍很少。而这正是我想要做的,我想研究人们还没有研究过的东西,因为那里会有更多的发现。

记者:您已经为智能导学系统研究作出了卓越贡献。我们想知道,当您从学校毕业时,为何选择从事教育技术研究?怎么萌发创建AutoTutor系统的?

格雷泽教授:非常好的问题。其实,一切都源于兴趣。在整个职业生涯中,我一直对交流和对话感兴趣。我发现在辅导过程中,师生的知识结构之间实际上鸿沟很大。我十分好奇人们是如何解决这个问题的,事实上人们对此知之甚少。因此,我的研究

领域之一便是指向(direction giving)。比方说,大家都有过向陌生人问路的经历,路人尽管十分友善并详细地告知答案,但你仍可能在一大堆热情回复中迷茫不已。很显然,这样的交流并不顺畅。有时他们还会追问“你明白了吗?”,而你鼓起勇气回答“是的,非常感谢”,然后目送他走远后赶忙找下个人打听。这是现实生活中存在巨大沟通鸿沟的典型案例。教学是另一个案例。也有人研究医生与病人之间的互动。病人去看医生时,医生总是使用成串晦涩难懂的专业术语,让病人不知所云。

于是,我特别想探索这一过程,了解人际鸿沟前人们是如何沟通的。所以,我对辅导过程特别感兴趣,分析了许多众所周知的教学过程、面部表情和动作等。然后,我发现教师们所做的也没什么特别之处,加之我是学习认知科学出身,有计算机科学背景,所以我打赌计算机也能做到。

实际上,我读大学时是神奇的六十年代,反叛精神特别强,我们抗议越战,我当时还留了长发,是个嬉皮士。我主修过八个专业,但没有全部获得学位。一开始我学习数学和计算机科学,但这与社会学相关不大。我便去学哲学,又学经济学,接着又学英语、人类学和语言学。经过广泛涉猎后,最后我进入心理学。因此,我主修心理学,但辅修数学、计算机科学、哲学和语言学等。后来这些正是认知科学的基础,也是认知科学的主要研究领域。回过头看,只要跟着这些问题走,最终都会殊途同归。

有了这些基础后,再加上我喜欢创造,所以每次分析各类行为时,我总会思考这些人的行为能否通过计算机实现,并对此充满信心。恰好在这一时期,出现了语义分析技术,并引发了计算语言学的变革。早期分析人类教师时,胡祥恩^⑤教授曾跟我一起工作,我们合作论文之一就叫量化会话心理学^⑥。我们提交了申请书,这成为后续研究的起点。就这样,我进入教育领域,开启了研究生涯。

虚拟代理出现很早,第一个代理名叫“巴迪(Baldi)”。那时,各种想法不断涌现,我开始对创造新东西感兴趣,但创造需要渊博的知识。比如,汽车半路抛锚,得去哪儿修呢?是去汽车制造厂,还是汽车经销商?有些人能滔滔不绝地讲个不停“这车五秒内能达到45公里时速”,这只是对汽车进行测量,并没有制造和维修能力。重点是,现在的心理学

家对各种事物开展实验——通常也只是测量而非构建事物。比起测量者,创造者对事物的内部机制理解更深。

记者: AutoTutor 已形成了家族,您设计 AutoTutor 的愿景是什么?

格雷泽教授:在美国,非常多的成年人阅读水平低到无法找到合适的工作,所以我想让有需要的人能免费使用。我们跟“专业读写”(Pro-literacy)机构合作,花了很长时间建立训练理解能力的导学系统,其中包含 35 个小课程模块,每个时长约半小时,目的是帮助阅读障碍的成年人。这是其中的愿景之一,若能帮助成千上万人更好地阅读,我将十分欣慰。

还是用数据说话。在美国大学生中,阅读能力未达到八年级的占 38%,数学能力严重不足的占 40%,形势相当严峻——他们就这样进入大学,学不好便面临退学。实际上,我们可以对理解和阅读水平低的大学生进行实验,无须费力招募校外被试。如果将范围拓宽,有这类问题的成年人有多少?答案是五千万。那么其中又有多少人去读写中心求助呢?答案是 3%。也就是说,五千万美国人中仅一百五十万人试图寻求帮助,更多人甚至不知道这个机构。因此,我希望 AutoTutor 能够让更多人受益,这就是我的初衷。

记者: 与人类教师相比,他们喜欢用这种方式提高自己的能力吗?

格雷泽教授:是的,大多数人喜欢使用 AutoTutor。首先,人们在它面前不会感到尴尬,可以轻松地展示和审视自己的不足;其次,系统备有教师和学生两种代理角色,有时我们会安置学生代理与真人比赛环节,但设定人不会输给机器。有趣的是,如果人的表现不好,学生代理会被“降级”到低一级学段。这类系统能够帮助学生培养自信等品质,这是我正在做的事。

我还想创建另一个系统,希望和优秀游戏团队合作,开发一组对情绪高度敏感的代理系统。之所以选择游戏方向,是因为思考常会受挫,人们生性不喜欢多思考。但是,虚拟代理和比赛或许能把任务变成娱乐而引人入胜。人们花时间努力通关,甚至

有时会感到沮丧,直到成功完成任务。我希望用虚拟代理达到这种效果,也想和公司合作开发出色的学习游戏,但这样的公司不好找。我们曾尝试与培生教育集团(Pearson Education Group)合作,开发一款寄予厚望的商业游戏。该游戏面向高中生和大学生,用于训练批判性思维,讲述火星人降临地球,以扩散伪科学的伎俩妄图侵占地球。学习者需要努力成为联邦科学局的探员,找到并将火星人缉捕。但他们只有掌握科学方法,才能在宣传材料中鉴别出伪科学。然而,五年前培生教育集团遭遇经济问题,只得中途放弃所有高端科技项目。

记者: 真是太可惜了,这属于“黑天鹅”因素,完全不可控制。不过,游戏中有很多可控因素,最可能影响学习效果的因素是哪些呢?

格雷泽教授:成功的游戏的确有利于生成浅层知识,但不适用于深层知识学习。人们很难直接在游戏中深度探究。据我所知,所有成功的商业游戏,都无法提供深层知识,主要原因是思维上的阻隔和非连续性。

不过,如果能尝试找到正确的游戏设计方法,或许可以克服。比如,有些人——尽管只是少数——喜欢有深度的、严肃的游戏,他们甚至可能花上几十个小时沉浸于此。如果能让更多的人参与其中,那么人们便可能进入并享受深度学习、系统性思维和因果推断的乐趣,之后还会对浅层内容感到无聊。截至目前,我只见过游戏偶然或短暂地达到深度学习的效果。至于其中的要素,可参考我、胡祥恩和鲍勃(Robert A. Sottilare)合作撰写的论文^⑦。在游戏环境下,使用通用智能导学框架和深度学习的可能性,奥秘在于反馈和选择。

再聊一个哥特式的想法。假定任务为引导孩子进入深度学习,可他们大多很顽皮并爱搞破坏。所以,如果游戏主题是“如何炸毁一幢建筑”,他们一定会长时间非常用心地研究这个问题。又比如,“如何给一个村庄下毒,在规定时间内使死亡率最高?”,有些人还喜欢玩欺骗和密谋类的游戏。既然男孩和女孩都会经历这个阶段,为何不借机使之与深度学习相关联?

还有个值得思考的问题:为什么在游戏中人们对积极的事情没那么感兴趣呢?为什么孩子们的学

习会偏好哥特式元素? 这种现象在心智中如何发展起来的? 我敢打赌中国也会有这种情况, 肯定有孩子一拿到玩具就想拆解, 即使拆开无法装回原样。请注意, 他们正自发努力弄清事情背后的运作方式, 这就是深层知识的萌芽! 但有多少家长和教师不是在斥责声中扼杀了孩子们这种深度思考和大胆探索之途呢?

记者:不止孩子, 成年人仍有这种冲动。感谢您的启发, 这是个非常深刻的问题。您如何看待深层知识和深度学习之间的关系?

格雷泽教授:深层知识不仅需要系统性思考, 还需要因果互动。通常, 某个事物被视为包含众多要素的整体系统, 那么附带效应和意外事件会使实际方案变复杂。

例如, 人行横道刚被使用时, 事故率陡增三倍。原因在于, 行人认为自己被允许过马路, 可司机不十分清楚这一点, 不少司机直接忽视人行横道而引发交通事故, 这属于意料外的后果。直至十年后情况才发生逆转, 人行横道切实降低了事故频率。这需要以“有多少司机会注意人行横道?”“有多少行人会十分小心?”等反思思维去思考。当习惯于系统性思考后, 就会意识到一个简单的解决办法很可能存在纰漏, 因为背后的整体机制不简单。

现实问题根本不像建立数学模型求得最优解一样容易, 折中权衡极具挑战。比如, 美国新奥尔良市曾多次遭遇飓风袭击, 城市受损严重。起初, 人们计划建造超大堤坝挡住洪水, 事实证明此事绝非那么简单, 因为只要水位漫过堤坝, 就是一场灭顶之灾。如果换种思考方式, 种植一些植物群落, 随机排列在海岸线上而非把墙建得越来越高, 就能使阻挡洪水的屏障功用发挥到最大化, 这才是真正稳妥的解决方案。同理, 在政治领域中, 特朗普想建造“美式长城”, 结果这个愚蠢提议让更多的人越过了边境。

记者:因此在您看来, 为了提升学生深度学习的能力, 对话是必需的?

格雷泽教授:对, 很有必要。我赞同乔治·米勒 (George Miller) 的观点, 这位神人不仅对认知心理学、心理语言学和认知神经科学的建立贡献巨大, 还将乔姆斯基 (Chomsky) 的生成语法理论引入心理

学, 甚至参与了 WordNet 项目, 不过最知名的一定是神奇数字“ 7 ± 2 ”。他认为语言是最不模糊的交流方式, 人们通常从语言中获取的信息和判断比其他交流形式都高级。很多人说有研究表明人们的主要交流方式是面部表情和肢体语言, 这太偏激了, 这些结论通常来自实验室中简单互动的图片或视频, 因此生态效度很低。假如你想向我传达复杂信息, 比如你打算明年春夏之交的某个时间来孟菲斯拜访我, 如何用肢体语言表达呢? 同样, 若想解释事物的问题出在哪里, 不用言语表达, 沟通将变得非常困难。

回到“指路”的例子。如果向别人问路, 98% 的人会用手指着目标方向, 甚至自己都没有意识到就指了指, 此时范围已明显缩小。反之, 如果他们只用语言回答“向南或向东南方向走”, 就不易理解。这时, 手势往往能迅速传达准确信息。因此, 手势和面部表情有时比语言交流效果好。

然而, 现实生活中人们需要整合很多信息, 特别是在假设的情况下, 必须使用语言。动物不讨论假设世界, 有人也许会说它们也有这种迹象, 其实没有, 只有人类才会讨论假设的世界。与他人交流时, 提及的多为别的时间和地点, 针对此时此地的信息不多。我敢说你们的谈话内容有很大部分不是关于假设的就是关于其他时间和地点的。在谈论假设事件时, 离开语言通常很难做到。因此, 我坚信理解深层知识通常需要涉及清晰的辨别和解释, 这也正是语言的看家本领。不然, 用非言语的方式交流深层知识, 那就太痛苦, 会不得不中断的。

记者:为了更好地学习深层知识, 学生需要更多的深度解释, 创设对话环境是一种理想方式。不过, 对话与学生的认知紧密联系。对严肃游戏的研究, 有认知和情绪两种视角。近年来, 为何越来越多的研究者开始对情绪感兴趣?

格雷泽教授:实际上, 概念学习是认知的, 但情绪更能激发人们的动机和注意力, 二者都为学习所必需。进一步来说, 使人保持学习的动力是什么呢? 就是情绪。我刚完成一篇论文^⑧, 对“情绪在学习环境中扮演的角色”进行梳理, 指出情绪对学习而言好似黏合体验的胶水。在日常生活中, 情绪同样非常重要。假如早上一起床心情就不好, 这会影响我

们大半天或一整天。多年来,情绪在心理学研究中一直居次席,现在转向情绪研究正是为了掌控动机——让学生保持动力来源、获得足够关注和毅力坚持学习。如果动机缺乏,成功便遥不可及。

记者:我们非常希望深入了解和使用 AutoTutor,也很好奇它有什么缺点?

格雷泽教授:正所谓“瓜瓜苦蒂,物无全美”。对这个问题的详细阐述,可参考 2016 年的文章^⑨。此外,一篇刚发表的论文^⑩,也指出了其中的不足。我把 AutoTutor 的缺点大致归纳如下:

第一个实际上是人类自身的弱点,与 AutoTutor 无关。比如,当学生回答问题时,他只需说出一两个关键词就能获得人类教师的认可。但 AutoTutor 却期望学生说出详细的信息,这会让学生抵触。很自然,学生在自动化辅导环境下也会有这种期望,于是 AutoTutor 可能会惹到他们。这是非常现实的矛盾,但并非不能解决。

第二个缺点是 AutoTutor 无法妥善处理学生的问题,毕竟设计者不可能考虑到所有潜在问题,考虑到的问题也不可能提供所有答案。因此,无法直接回答学生的提问时,AutoTutor 会使用语言技巧来回避。比如,它会说“哦,这是个好问题,请稍后再问我”或者“这是个好问题,你怎么看?”

第三个缺点是 AutoTutor 不会复述。复述是很重要的教学技巧,即教师根据学生发言重新组织语言加以表述。假如你说“哦,AutoTutor,有问题”。如果我需要重述你的话,我会说“你是想知道 AutoTutor 应用是否存在问题吗?”虽然这只是重复,却是用更完善的语言加以组织和表达。优秀教师在课堂上也许会这么做,尤其在不明确学生的表述时,可以说“乔治正在思考语言和交流之间的关系,你们对此有什么看法?”这种做法一箭双雕——既能带动学生,又能串联起其他同学的问题。我们希望 AutoTutor 也能做到,但目前还达不到。事实上,很多人类教师也不擅长复述,需要接受培训。匹兹堡大学专门开设会话课程,指导教师如何在课堂上与学生更好地交谈,复述便是其中的常见方法。

此外,还有个问题,AutoTutor 情绪反馈系统目前仍不理想,不过我们正在改善。例如,假设学生回答非常精彩,我们希望系统能够给出与之对应的情绪

反馈——用热切而非平时那种平淡的语气称赞学生。人们一直希望设计出情绪反馈得当的虚拟导师,用细致的情感、热情的或者怀疑的语气进行反馈。很多人类教师可以做到,但 AutoTutor 还做不到。

记者:由于 AutoTutor 已被整合到 GIFT 中,请简单介绍一下两者之间的关系。

格雷泽教授:GIFT 既是理论框架,也是工程框架,有一套开源的软件系统,任何人都可以借助 GIFT 免费快速搭建导学系统。现在,世界上活跃的智能导学系统有数十个。比如,MathTutor 是个辅导数学的智能导学系统,美国军方也有用于训练射击术的智能导学系统。GIFT 对所有这些智能导学系统进行诠释,AutoTutor 只是其中涉及自然语言交互的一例。

记者:AutoTutor 基于对话思想而设计,GIFT 和 AutoTutor 之间是否存在架构冲突呢?

格雷泽教授:不,没有实质性的冲突。AutoTutor 和史蒂夫·里特(Steve Ritter)的 Cognitive Tutors for Mathematics,后者面向数学,前者面向自然语言。但当学生陷入僵局时,系统都会给出提示。假如学生仍不明白,才会告知答案。此外,其他导学系统通常也遵从“线索-提示-结论”循环,不会直接被训练成“填鸭机”。这是 GIFT 基本原则决定的——不只是讲课或直接给出答案,而是提出问题并给出提示,鼓励学生主动解决问题。

记者:MOOC 在中国非常受欢迎,您如何看待 MOOC 和 AutoTutor 之间的关系,二者能否融合?

格雷泽教授:今天,“融合”指的是不同技术的协同工作。八年前,我们讨论过 MOOC 与 AutoTutor 的结合,曾设计了基于 MOOC 的 AutoTutor 模型。其实,我们既然有 MOOC,又有会话代理,没理由不能融合。当时本想深入探究,却因忙于其他事务而搁浅,但我仍觉得二者融合十分必要。

在 MOOC 模式下,要么观看视频,要么阅读文本或完成练习,学生不知道自己是否真正理解。因为人的阅读理解能力并不完美,学生对材料的实际理解程度和自认为理解程度之间的相关性仅 0.27。

这其实是缺乏反馈所致。这看似复杂,其实很简单,不需要懂得很多技术知识就能做到。比如,学生正在学习 MOOC 视频课程,中途暂停并弹出对话框“如果按 1-5 评分,请您就本材料的理解程度打分”。学生给出评分,系统继续提问“现在,请你概括观看的内容并输入”。不少学生开始都自认为理解到位,却往往在概括环节哑口无言。这是个非常简单的操作,只需暂停并概括总结即可,其间仅需采集概括的内容,甚至无须给出任何反馈。这种方法有两方面优势:一,可以提高内在认知标准和理解标准,学生将意识到“我可能没有完全理解学习材料,因为我什么也说不出来”;二,助力学生深入内化信息。最终,理解和概括能力都将获得提升。

记者: AutoTutor 有良好的过程控制,其效果可以细化至个体,但 MOOC 不同。

格雷泽教授: 是的,就 AutoTutor 控制 MOOC 教学进度而言,技术上没有任何问题。Moodle^① 就是理想的平台,胡祥恩教授已经将很多课程以 Moodle 的方式进行部署,创作插件提供多种控制方式,可对视频或多页材料的目标位置设置断点,实现学生评分及收集总结等操作,全程无需编程、易掌握。其实,MOOC 本身也有聊天互动的功能,但使用者寥寥无几。人们可以开发一个插件运行虚拟代理,让它站出来说“等一会,你理解了吗?你学习得怎么样?能替我总结一下刚刚学习的内容吗?”使之更富社交性。还可设置积分,让你认为另一个学生正与你一起学习,如果你能提供帮助,就可以加分。

记者: 请畅想一下,如果您有足够的财富,您会选择哪种人工智能技术开发更好的 AutoTutor 呢?

格雷泽教授: 首先跳入脑海的念头是,一定要做与众不同的事。以物联网为例,世界上有各式各样的设备,人们并不熟悉它们的功能,也不知道该如何使用。现在,也请你们畅想一下,假如设备能够像 AutoTutor 一样与人类互动,可以向使用者介绍自己该有多好。在你拿到一个新设备却窘于不知如何使用时,它便开口道“嗨,你了解我吗?”,而不是像现在,你拿着一个不会说话的新轮胎,心中痛苦地嘀咕着“我该怎么办?”然而,未来的孩子们或许会惊讶道“我的自行车竟然是个哑巴!”

记者: 太酷了,也许下一个谷歌和脸书就在这里。谢谢您,最后还有三个小问题,请您用一句话回答。第一个小问题是,在整个职业生涯中,您最高兴的事情是什么?

格雷泽教授: 最开心的时刻是被授予 Harold W. McGraw 学习科学奖,这要感谢胡祥恩教授的帮助。

记者: 第二个小问题,在所有成就中,什么是最成功的?

格雷泽教授: 最成功的事情就是我和同事一起合作创建了 AutoTutor。

记者: 最后一个问题,您最遗憾的事情是什么?

格雷泽教授: 最大的遗憾是没有足够的能力将 AutoTutor 造福于千千万万的人。

所以,我们现在需要与企业合作。其实,到了我这个年龄,并不在乎能否借 AutoTutor 致富,只关心它能否被大众使用。担心侵犯知识产权在我这儿是多余的。我真心希望我的知识产权被大家“侵犯”,也诚挚欢迎大家前来“侵犯”!

[注释]

① ACT 是一家提供美国大学入学考试 (American College Test, 简称 ACT) 服务的非营利性教育培训公司。

② D' Mello S., Jackson T., Craig S., et al. (2008). AutoTutor detects and responds to learners affective and cognitive states [C] // Workshop on emotional and cognitive issues at the international conference on intelligent tutoring systems: 306-308.

③ Fiore, S. M., Arthur, G., & Samuel, G. (2018). Collaborative problem-solving education for the twenty-first-century workforce [J]. Nature Human Behaviour, 2(6):367-369.

④ <https://gifttutoring.org>.

⑤ 胡祥恩教授是美国孟菲斯大学心理学系、电子计算机工程系、计算机科学系教授,智能系统研究院高级研究员,华中师范大学心理学院院长、中国青少年网络心理与行为教育部重点实验室高级研究员,主要研究领域包括教学心理学、实验设计与统计、心理学、人工智能和智能导学系统。

⑥ Graesser, A. C., Swamer, S. S., & Hu, X. (1997). Quantitative discourse psychology [J]. Discourse Processes, 23(3):229-263.

⑦ Graesser, A. C., Hu, X., Nye, B. D., & Sottolare, R. A. (2016). Intelligent tutoring systems, serious games, and the Generalized Intelligent Framework for Tutoring (GIFT) [M]. Using Games and Simu-

lations for Teaching and Assessment. Routledge: 82-104.

⑧ Graesser, A. C. (2019). Emotions are the experiential glue of learning environments in the 21st century[J]. Learning and Instruction, 2019; <https://doi.org/10.1016/j.learninstruc.05.009>.

⑨ Graesser, A. C. (2016). Conversations with AutoTutor help students learn[J]. International Journal of Artificial Intelligence in Education, 26(1):124-132.

⑩ Dowell, N. M. M., Nixon, T. M., & Graesser, A. C. (2019). Group communication analysis: A computational linguistics approach for detecting sociocognitive roles in multiparty interactions[J]. Behavior research methods, 51(3):1007-1041.

⑪ <https://moodle.org>.

(编辑:徐辉富)

The Technical Inspiration and Humanistic Attention behind AutoTutor: An Interview with ITS Expert Professor Arthur Graesser, University of Memphis

Journalist LIU Kai^{1,2}, WANG Shao¹, LONG Zhou³ & QU Jing⁴

(1. Department of Education Science, Bohai University, Jinzhou 121000, China;

2. School of Psychology, Central China Normal University, Wuhan 430079, China;

3. Department of Education Science, Huaihua University, Huaihua 418000, China;

4. Graduate School of Education, Peking University, Beijing 100871, China)

Abstract: Dr. Arthur Graesser is a professor of psychology and intelligent systems at the University of Memphis and is an honorary research fellow in education at the University of Oxford. He has been awarded the Distinguished Scientific Contribution Award from the Society for Text and Discourse, the Distinguished Contributions Award of Applications of Psychology to Education and Training Award from the American Psychological Association, the University of Memphis' First Lifetime Achievement Award, Harold W. McGraw Learning Science Award, etc. He is interested in cognitive science, word processing, and learning science. He also designs and develops well-known teaching software such as AutoTutor, DeepTutor, Electron Tutor, PAL3, and Coh-Matrix with colleagues. This paper is an exclusive interview conducted by professor Graesser during the conference on "Computational Behavioral Science Symposium: Models and Analysis of Intelligent Adaptation Teaching Systems" in mid-October 2019 in China.

Taking AutoTutor as a line-through, this interview involves issues like teacher training, emotion, collaborative learning, deep learning, GIFT and MOOC, etc.

China's economy and education have been developed quickly since reform and opening up 40 years ago. However, compared to the United States, it seems that our education system still lacks originality. We hope this interview may provide insights and guidance on the issue.

This interview coincides with the retirement of Professor Graesser and uses the opportunity of academic review to look at his academic connotation, hoping to inspire colleagues in the academic community.

Key words: intelligent tutoring system; deep learning; human-computer dialogue; emotion; AutoTutor; GIFT; MOOC