

认知科学视角下对数字史学的透视

孙 辉

(中国社会科学院当代中国研究所, 北京 100009)

摘要: [目的/意义] 数字史学作为一种新的史学研究形态, 在学界颇受争议。本文从认知科学的视角透视数字史学, 旨在客观理性地看待它, 推动史学在创新的轨道上健康发展。[方法/过程] 采用文献调研法对历史学界关于数字史学的观点进行梳理; 采用认知科学关于记忆、心智计算的理论对传统史学研究和数字史学研究的认知特点进行分析, 并归纳出研究流程和注意要点。[结果/结论] 数字史学的特征是认知过程中部分记忆和计算的高度外化; 数字史学改造了传统史学研究的认知手段、认知方法和认知过程, 最终影响认知能力和认知结果; 在数字史学研究中, 史学功底训练、直觉思维和论证驱动依然重要; 本科阶段不宜开展数字史学研究; 本文还对数字史学项目的开展提出建议。

关键词: 数字史学 认知科学 数字人文

分类号: G250

DOI: 10.31193/SSAP.J.ISSN.2096-6695.2020.01.06

0 引言: 来自史学领域的声音

近代以来, 人文学科的研究一直受到自然科学和社会科学的影响, 呈现出一种以强调“实证”为特征的“科学化趋势”。在史学领域, 20世纪五六十年代发源于美国的“新经济史”研究, 开启了计量史学的先河。20世纪末, “数据驱动的科研发现和基于研究数据的知识管理, 已经成为许多领域的重要基石, 包括人文社会科学领域”^[1]。“数字人文”受到人文学科研究者的极大关注。数字史学是数字人文在史学领域的应用, 是计量史学的升级版, 它在计量指标分析的基础上增加了词频分析、社会网络分析、知识图谱、地理信息系统分析等信息科学方法的应用, 数字史学能否成为一种研究范式, 抑或遭遇20世纪70年代计量史学曾经受到的打击, 我们需要倾听来自史学领域的声音。

年轻学者尤其是有境外学习经历的学者比较认同并乐于在研究中使用数字史学的方法; 年长的学者尤其是传统的历史学者对数字史学的认识基本停留在史料数字化的层次, 且对新工具和新方法持冷静或保守的态度, 甚至担心数字史学会把年轻学者引入歧途。

[作者简介] 孙辉 (ORCID:0000-0003-0103-7107), 女, 研究馆员, 博士, 研究方向为信息组织、数字人文, Email: sunh@iccs.cn。

积极的观点认为,通过数据采集、数据储存、数据分析、数据管理为大范围、中时段、长时段定量分析提供了现实的可能^[2];数字史学能实现长时段巨篇幅文献的遥读;把零散的史料依照一定规则重新排列组合以后,往往会产生豁然开朗的感悟,“能够发现靠传统文献阅读无法发现的隐藏在历史资料堆中的史实”;具备“验证、修正、创新”三重功能^[3],是史学研究“现代化”、“科学化”的一个重要途径,有助于史学学科由“解释型学术”走向“求是型学术”^[4]。另外在文献学方面,数字史学在古籍自动校勘、自动编纂与自动注释^[5],自动断句标点、自动分词标引^[6]等方面发挥着特殊作用。

对数字史学持冷静态度的学者,将数字史学方法与传统史学治学方法相比,认为存在如下问题。

(1) 数字化史料的可靠性问题。数字化史料解决了传统史学“皓首穷经”的问题,但数字化史料毕竟不是原生资料,在加工过程中容易出现错字、句读错误,这些错误会影响史料的检索获取;另外,在此基础上用检索来的资料进行分析加工,其研究成果的准确度有待商榷。

(2) 史料检索带来若干问题。傅斯年所说的“史料即史学”不能演变成“检索即史学”。其一,检索结果为片段集合,没有镶嵌在整本书或整个版面中的片段丢掉了很多信息,“文字少的段落离开语境以后意义不明”^[7]。其二,全文检索造成大量漏检和误检,这个问题在古籍检索中尤为突出。“在中国古代历史文献的记载中,文本字面含义与史实之间常常存在错位,这就使得事情变得更为复杂了。”^[8]其三,以材料检索代替读书的便捷路径是否可靠。王国维的“读书得间”论认为,“宜由细心苦读以发现问题,不宜悬问题以觅材料”,“那些出乎研究者预设范围之外、不包含所谓关键词的‘隐性史料’,往往深藏于史籍的字里行间,深藏在我们熟视无睹的史事中”^[9],贸然以问题为导向仅利用检索的结果开展研究,造成“研究者预先设定的思路往往成为搜寻和筛选数据的藩篱而不自知”^[10],“如此完成的论文,其思想厚度就会大打折扣”,“论文数量激增,但学术思想贫乏,真正的问题突破无从谈起,而快餐式成果大量累积。”^[11]如果说检索带来的前两个问题,学者可以通过改变检索策略或阅读原文得到改善,那么第三个问题则被很多学者认为很严重。学者认为“读书得间”,尤其在史学训练的初级阶段,不可跨越。^[8]

(3) 历史是否可以量化。尽管量化历史学家陈志武认为:“有许多表面看上去不能量化的研究课题,其实只要我们发挥想象力,还是可以找到代理变量,或者通过创造性研究设计把不容易量化的研究变得可以量化。”^[12]但部分量化研究存在如下问题。一是数据库统计口径或平台功能设计不合理,有些情感计算、社会网络计算方法往往抹平了事物之间复杂丰富的关系,“根据错误理念所建立的数据库系统,虽看似科学谨严,实则南辕北辙”^[9]。二是见数不见人,史学没有体现人文关怀。传统学界这样评论“计量史学”的代表作《背负十字架的年代》(*Time on the Cross: The Economics of American Negro Slavery*): 计量史学的专家能够精准地算出每个美国黑奴遭受奴隶主鞭打的次数,但是他们从来没有去关注暴力与奴役状态对男女奴隶的生活与心灵带来的影响^[13]。

显然,对于数字史学这个命题,学界并没有形成共识。历史研究的本质是人的认知活动^[14]。在历史研究过程中,构建知识体系、爬梳史料、发现问题、还原史实、发现规律、写作论文等都是历史学者在历史领域的认知行为。数字史学改造了传统史学研究的认知手段、认知方法和认知过程,最终影响认知能力和认知结果。本文从认知科学的视角透视数字史学,认为数字史学的主要特征是,利用机器海量存储和高速计算的特点将史学研究过程中部分认知高度外化;描述数字

史学的认知流程, 并从人脑和机器的差异入手分析传统史学研究和数字史学研究的认知特点; 提出数字史学研究中需要注意的问题, 并对数字史学项目的开展提出建议。本文旨在客观理性地看待数字史学这种新的研究状态, 推动史学在创新的轨道上健康发展。需要说明的是, 认知科学的核心难点是人类对脑的结构和功能原理了解还远远不够, 认知科学的一些理论有待脑科学的进一步验证, 但部分已通过实验方法、逻辑方法、内省方法检验的理论或“前理论”, 对于透视数字史学这种新的研究形态, 仍然具有一定意义。

1 数字史学: 研究过程中部分认知的高度外化

认知科学是研究人类感知和思维信息处理过程的科学, 包括从感觉的输入到复杂问题求解, 从人类个体到人类社会的智能活动, 以及人类智能和机器智能的性质。^[15] 认知科学和计算机科学的发展是相互启迪、相互影响的。认知科学中的信息加工理论来自“计算机隐喻”(computer metaphor), 它认为, 人类认知过程的本质就是信息加工过程, 是人脑对环境输入的信息进行编码、储存、提取和操作的过程。人也被称为信息的有限能力加工者。^[16] “计算机隐喻”一方面使计算机成为探索人类心智功能最为可行的模型; 另一方面通过硬件和软件模拟人类认知, 研究认知的形式化表征和算法、进行计算机建模, 即开展人工智能, 人工智能是人类认知的拓展。可以这么认为, 计算机科学启发并产生了基于信息加工理论(最初表现为符号主义计算理论)的认知心理学; 认知神经科学尤其是神经元细胞和神经网络的发现促进了人工智能领域的联结主义计算理论(人工神经网络理论)的产生; 人工智能的新发展又拓宽了认知心理学和神经科学的研究思路。比如, 联结主义计算理论被用来描述人的感知, 符号主义计算理论被用来描述人的高级思维; 深度学习的理论源头是联结主义计算理论, 知识图谱的理论源头就是符号主义计算理论。

认知科学与计算机科学的相互影响和不断发展的结果, 就是人脑认知的不断外化。如同麦克卢汉所说的“媒介是人的延伸”^[17], 数字史学的特征就是史学研究中的记忆高度外化和计算高度外化。

1.1 作为历史源头的记忆的外化

“记忆是历史的源头。”^[18] 古人留下的历史文献是记忆的外化结果; 今人认知历史, 其本质是一个对故人遗留记忆的获取、整理、辨识和解释过程。博闻强记是传统史学研究的基本素质。认知科学认为, 记忆就是脑神经细胞之间的相互呼叫作用。记忆包括感觉记忆、工作记忆和长时记忆, 分别类似计算机的寄存器、内存和外存(如图1所示)。其特点有二: 一是长时记忆“健忘”, 需要“学而时习之”; 二是工作记忆容量有限, 一次只能处理大约7个单元的信息。

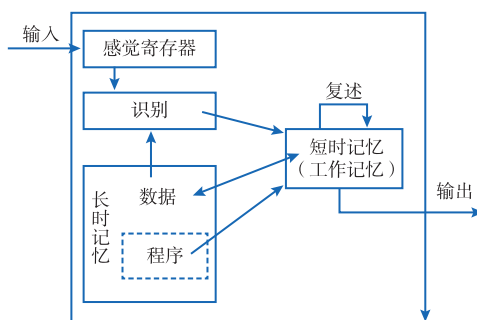


图1 以短时记忆(工作记忆)为中心的系统各成分的协调活动^[19]

由于人类的记忆能力有限，人类记忆的外化过程一直没有停止过。人类与动物心理的区别是从语言认知开始的，语言是记忆外化的第一步。人类先有语言后有文字，在“书写的历史”出现之前，人类通过口耳相传将历史记忆保留下来，如《荷马史诗》《格萨尔王》等。“好记性不如烂笔头”，文字的发明又进一步将记忆外化。之后，人类又发明各种记录载体（纸张、磁带、磁盘等）和复制手段（印刷、数字化、互联网），人类记忆外化的手段越来越多。

在数字史学研究中，数据库和知识库将部分记忆高度外化。这种高度外化的记忆对于历史学者来说有几方面优点：一是克服人类“遗忘”的缺点，可以永久保存；二是突破个人记忆的限度，理论上，数据库可以保存整个史学领域的记忆；三是不同于人类记忆必须经过三阶段过程（感觉记忆—工作记忆—长时记忆），历史学者无须查找和阅读大量史料，随时随地拥有海量史料；四是通过外化记忆，人类将工作记忆的一部分转移，可提高人脑的思维效率，在数字史学中，通过思维导图、地理信息分析、知识图谱等可视化手段，人类将一部情景记忆（或心理表象）转移到机器上，从而将更多的认知资源用于计算，可大大提高思维的敏捷度，比如，约翰·斯诺霍乱的地图帮助人们找到了传染源。

记忆高度外化中的信息组织也类似人类认知的信息组织模式。比如，由于人的工作记忆容量有限，在思维中处理的是概念、命题等浓缩的语义记忆，因此，人类的高级认知离不开概念、命题的提炼。在史料数据库中，用于标引的主题概念、分类号等浓缩信息，缩短了记忆单元，便于计算。再如，认知心理学根据内省实验方法，提出语义记忆的扩散激活模型（见图2），认为语义记忆中的许多概念节点是联结在一起的。^[19]在知识库构建中，基于语义的细粒度知识单元及其关联就是一种借鉴，图3为笔者参与项目“中华人民共和国国史知识库”片段，这种组织方式便于知识漫游式检索和知识发现。

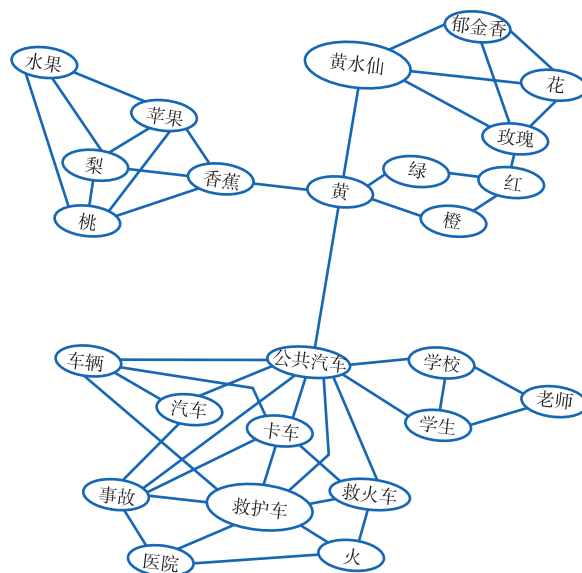


图2 扩散激活模型提出的语义记忆网络片段^[19]

1.2 作为历史研究的心智计算的外化

在认知科学看来, 历史研究是一种心智计算。人脑基于自身的生理特征, 在心智计算方面具有以下特点。(1) 人类的感知通道有限, 信息存入记忆需要占用意识资源足够长的时间。因此当一件事进入意识, 另外一件事必须等待, “注意”受自身意识和已有知识经验的影响。(2) 基于生存目的, 人类需要在最短的时间获得认知, 人脑擅长基于统计的并行方式的模糊计算, 人脑能从背影快速识别某个熟悉的人, 但错别字常常从眼前滑过。(3) 逻辑思维是进化后期的产物, 是人类特有的高级思维活动, 受工作记忆容量的影响, 人脑在此类计算方面采用的是产生式串行计算方式, 人类的逻辑思维速度有限。(4) 受生理因素影响, 神经元经过一段时间的活动后就受到抑制, 人脑容易疲劳, 人脑不善于处理时段大数据。

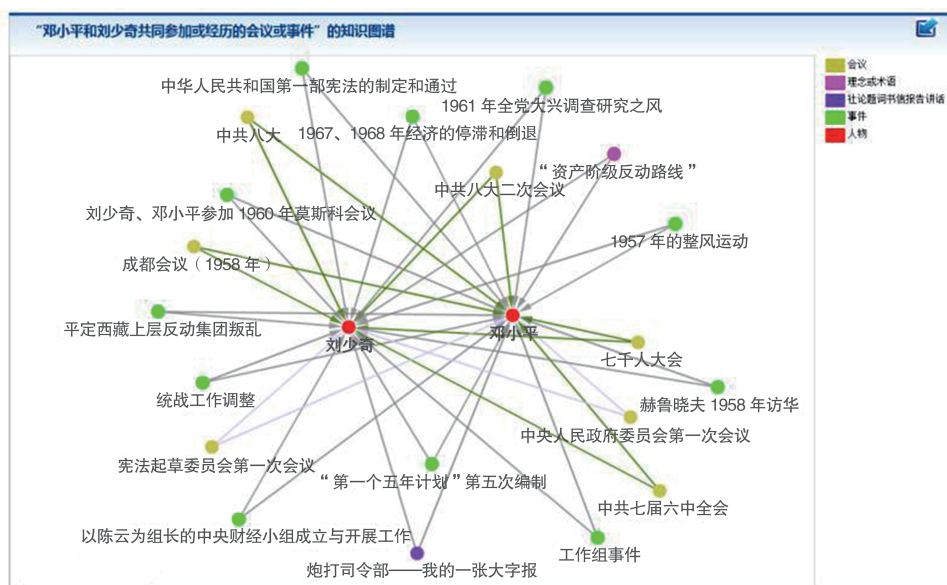


图3 “中华人民共和国国史知识库”片段^[20]

计算机处理逻辑问题的能力比人类要强, 理论上讲, 所有可以形式化的计算都可以交给计算机。在人类文明进程中, 人类通过发明算盘、计算机等将这部分串行计算的任务外包。在数字史学领域, 学者将研究过程中的一部分逻辑思维外包, 这些外包, 有的是提高了计算效率, 有的是扩展了计算类型, 延伸了人脑的计算能力。

■利用统计工具处理时段大数据, 可迅速获得统计图表和分析结果。

■按认知心理学观点, 人类决策的过程就是从目标单元到结果单元的搜索过程。计算机可以处理比人脑复杂的搜索, 数字史学通过对知识图谱的路径检索, 可以辅助找到一些历史发展脉络, 比如两个人物之间的关系, 两个事件之间的关联。

■不受人脑感知能力限制, 利用计算机自动比对技术, 帮助古籍版本自动校勘、自动编纂。

■计算机在一定规则下自动分类、自动聚类等, 能帮助人类迅速定位到相关历史文献或相关历史要素。

此外,受认知科学的联结主义理论影响,基于人工神经网络的机器学习技术不断发展,人工智能在基于统计的模糊运算方面取得重大突破,自然语言处理能力不断增强,利用计算机自动识别技术,可帮助古籍文字的自动识别。

1.3 部分认知高度外化的影响

理论上,随着数字史学的进一步发展,“历史学也完成了一个转变,即从传统的事实性史学,转变为思想性史学,历史学研究的目标,也将从弄清历史事实,转变为历史解读”^[11]。真正达到司马迁所说的“究天人之际,通古今之变,成一家之言”的境界,成为一门“求是型”学术。

有学者认为,数字史学有三重功能:验证、修正、创新。^[3]对于“验证”功能,传统史学研究者认为没有意义。笔者认为,所谓“验证”,就是证明某种“总体印象”“感悟”,这些“总体印象”“感悟”,是历史学者阅读大量文献后经直觉思维在大脑中形成的隐性知识,这种隐性知识是大脑多次记忆和计算的结果。也就是说,每一次阅读相关知识都刺激和加深了大脑中的神经回路,最后自然得出某种“总体印象”。在传统史学的论证中,一般采用典型案例给予证明。对于大的空间和时间尺度的海量文献,个人的认知能力有限,个人这种隐性知识无法分享,借助数字史学的记忆和计算外化,将部分史学要素形式化、指标化并进行计算,确实可以验证某些“总体印象”,这种量化验证,在要素简单的史学命题中是有意义的。

2 数字史学的认知流程及其注意要点

2.1 数字史学重构史学认知流程

记忆外化改变了传统史学流程中文献阅读和检索方式;计算外化改变了分析材料的方式,基于长时段大样本数据分析,可显性量化地反映各历史要素之间的关联。这种史学研究中智力资源的拓展改变了传统史学的研究流程。在《论互联网时代的历史学》一文中,作者提出当代史学研究的三种形态^[11],笔者在此基础上,描述了数字史学对传统史学的流程重构(见图4)。

所谓传统史学研究形态,是指近代以来所形成的以问题意识为导向,以阅读文献、爬梳材料、历史考据为基础,以分析归纳、抽象概括、历史评价为主要研究方式的史学研究活动。这也是在计算机技术普及之前,史学研究的一般程序^[11]。其认知流程如图4所示。数字史学对其中的④~⑥进行改造,增加资料来源,延伸计算能力,计算结果驱动史学发现,数据驱动增强逻辑思维能力;⑤~②表示,在计算过程中可能发现与原有问题毫不相干的新问题(这种纯粹的数据驱动,在现有的数字史学成果中还未发现)。图4中流程不完全是自上而下的,如④~⑦,存在不断反复酝酿的过程,甚至在写作过程中,学者都可能有新的观点产生。因为写作的过程就是一个逻辑思维外化为文字的过程,在此过程中,心理表象和心理逻辑中不明确或不能自洽的问题就会暴露,此时,研究会进入新的境界。

无论是传统史学还是数字史学,其研究过程都是隐性知识转换成显性知识的过程;在研究过程中,直觉思维和逻辑思维并举,并最终表现为具有较强逻辑性的学术论文;史学专业训练、问题意识是两种研究流程都不可或缺的环节;论证驱动是一种逻辑思维方式,常常表现为典型案例的归纳,传统史学和数字史学都离不开论证驱动,数字史学的计算结果要与相关史料相互印证补

充,“将数据驱动与论证驱动结合起来,以真正的问题意识为指引,而不是为了卖弄数字工具的噱头,才能带来更专业的成果呈现” [13]。

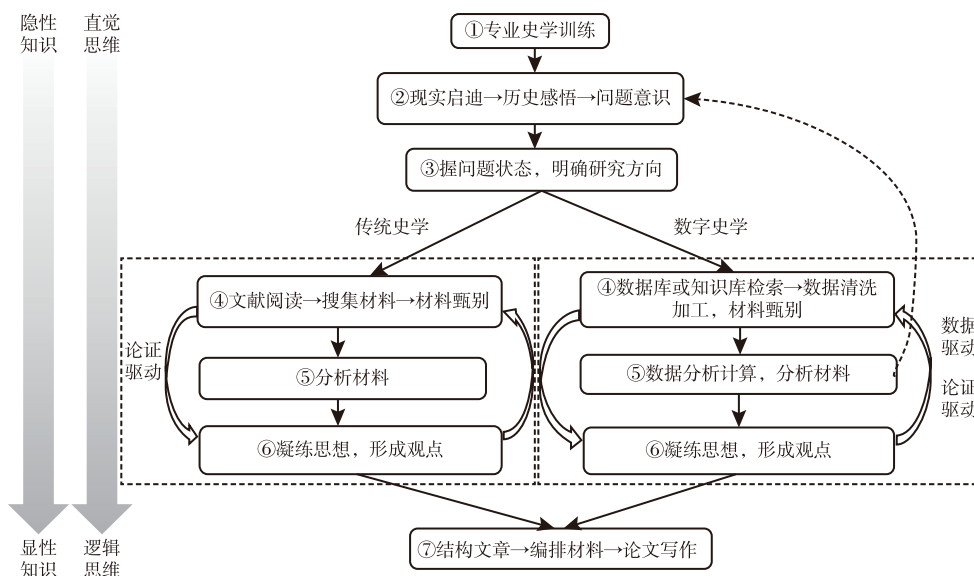


图 4 数字史学对传统史学的认知流程进行重构

在史学领域,当前数字史学“叫好不卖座”,有学者认为:“在‘传统’的以全文阅读为基础的研究中,成熟的学者往往可以通过对不同案例的全面考量,判断其典型性,才举为例证。这一过程尽管看似主观,事实上可能反而更接近史实。但是根据这些记载建立起来的数据库,看似精确,实际是只能将复杂、立体的描述性历史文本转化成平面的数据,抹杀了它们相互间的差异性,这样一来,数据库统计还能够在多大程度上反映历史现象的普遍性,不免令人存疑。” [8] 笔者认为,究其原因,是数字史学的研究流程在实际应用中出现了差错,要么丢掉认知流程中的重要环节,要么夸大了机器存储和计算的作用。本文将在分析人脑与机器差异的基础上提出在数字史学研究中要注意的一些问题。

2.2 数字史学研究中的注意要点

2.2.1 史学训练不可或缺

人脑和计算机在物质基础和结构设计上存在根本差异,在人脑中,记忆不是简单的存储,还包含有计算、简约和抽象的取舍过程,记忆和计算总是同时发生。因此无论是语义记忆还是图像记忆,本质上都是统计记忆,一般来说越是长期的、大量的、反复的记忆越难遗忘,脑中留下的记忆实际上是脑认知函数和脑遗忘函数的卷积。 [21] 我们当前的计算机是冯·诺依曼架构,其计算和存储分离,采用存储程序式计算,程序也是预先设定好的,无法根据外界的变化和需求的变化进行自我演化。我们的大脑却是一个出色的、能够长时间稳定工作的通用智能系统,不仅能举一反三,处理视觉、听觉、语言、学习、推理、决策、规划等各类问题,还可以在学习和发育过程中不断自适应和进化。至今为止,还没有任何一个通用智能系统能接近人类水平。

传统史学学者强调“读书得间”“读书破万卷”“皓首穷经”式的全文阅读，这种阅读就是一个记忆与计算同时发生的认知过程，是一个知识内化于心的过程。阅读，就其本质而言，是人类就自身对主客观世界的认识结果进行编码、记叙表达的意义再生成活动。^[22]对文献的研读和揣摩，在增加历史知识的同时，也刷新了大脑对这问题的认识。承载记忆的神经回路不断发生变化，有的增加，有的增强；增加的有的是事实，有的是对历史规律的认识。通过爬梳史料和“细读”，学者不断修改大脑中知识体系，真正在脑中积累大量知识，大脑认识问题的能力不断增强，这有利于提高对史料和历史规律的敏感度，从而对辨别史实、具备观察问题的整体性眼光起到重要作用。建构主义理论也认为，人的认知结构是人通过对客观现实的逐步适应建构起来的。认知结构是主客体相互作用的结果，它既不是客体的简单复写，也不是主体预先构成或天赋，而是主体在与客体的相互作用过程中，通过同化与顺应而构建起来的。史学训练就是一种历史思考力的建构。

数据库中的史料虽然很多，但对于某一阶段的学者来说，这种史料是静态的，是与入脑在物质上是分离的，这些史料对于人脑来说，没有记忆也就没有计算，知识和规则无从积累。通过简单的“检索式阅读”或数据抽取分析，进行“遥读”，将记忆与计算过多外包给机器，不利于历史学者思考力的培养。过分的让渡记忆和计算能力，历史学者容易困在数据库的“信息茧房”之中，降低历史研究的敏感度，丧失历史思考力。历史思考力是历史学者的核心价值所在，“数字史学”的主体必须是人。

2.2.2 问题意识离不开直觉思维

如果说扎实的史学功底是开展史学研究的基础，那么，问题意识则是史学研究的起点。只有当一个问题真正的问题时，史学研究才能取得成功，只有问题是新颖的，史学研究才可指是独创性的。关于问题意识，可以用认知科学中的原型理论来解释。原型理论认为，在记忆中贮存的不是与外部模式有一对一关系的模板，而是原型。原型是一个类别或范畴的所有个体的概括表征，反映一类客体具有的基本特征。当刺激与某一原型有着最近似的匹配，即可得到识别。例如，人类在感知外部世界时，一般认为采用基于鬼域模型的直觉思维进行识别。如果发现不匹配的地方，则发现问题。

直觉思维是一种基于人脑已有知识进行的统计学意义上的多次模拟计算。在传统史学研究中，提出一个问题实际上是察觉到了某种隐藏的东西，是得到某种“暗示”，这种“暗示”的第一次出现应该是直觉思维的结果。因为模式识别上出现不匹配，所以“问题”出现。马克斯·韦伯指出：“历史学家的‘得体感’或‘直觉’，而非一般化和对于‘规则’的意识揭示了因果联结。”^[23]历史学是一门人文科学，面对的是复杂的人类系统，在面向复杂系统的判断中，人类的直觉思维能力往往超过逻辑思维能力。

随着数字史学出现，数据驱动的逻辑思维是否会提出问题？目前情况下，人工智能不具备“意识”，不可能主动提出问题。所谓数据驱动下的问题意识，可能在两种情况下发生。一是研究者将发现问题的规则告诉机器，比如设计某种算法，请机器给出知识图谱中关联度最高的历史要素；二是研究者观察各种图表数据后，再次用直觉思维发现问题，像魏格纳通过观察地图发现“大陆漂移学说”一样。

2.2.3 数据驱动和论证驱动要统一

数字史学提倡用数据驱动，数据驱动强调大样本全样本，是基于严密规则串行的计算，形式化程度高，这是自然科学采用的研究方法。单纯的数据驱动的史学作品更像是技术报告，其对数

字背后的原因探讨不够深入, 往往被认为是“只见数字不见人”。论证驱动采用典型材料枚举式的证明方法, 论证手段为文字描述, 形式化程度低, 但语言可读性高, 历史代入感强, 体现人文特质。所谓典型事例, 是研究者的心智计算后, 获得较大权重的一些知识。单纯的数据驱动将记忆和计算都交给机器, 存在以下问题。

历史是否都可以计算? 当前, 对于情感计算、思想观念计算, 多采用词频统计或词间关系分析的方法。词语确实表达某种情感或观念, 但人类的情感或观点不仅仅表现为词语, 况且, 简单地通过某几个词语的分析, 只能提供某种线索, 不能给出结论。

计算机是否能穷尽历史知识和规则? 在人工智能领域, 存在封闭世界和开放世界的理论。机器计算时面对的是一个封闭的世界, 而人类大脑计算时面对是一个开放世界。所谓封闭世界, 就是知识和规则都是确定的。尽管数据库或知识库中保存了大量历史记忆, 但确实无法穷尽历史知识和规则。如前所述, 历史学者在长期的学术训练之后, 获得大量知识, 包括隐性知识和显性知识。虽然数据库和知识库保存了学术共同体的知识, 但其保存的只是显性知识, 不包括大量的在学者脑中的隐性知识, 治学经验、辨识眼光、判断力等“计算规则”, 更不可能得到体现。

计算工具的选择是否合理? 关于计算工具的选择方面, 很多数字史学研究者存在盲区。比如, 知识图谱和社会关系网络的可视化形态类似, 但并不是同一种技术。知识图谱反映实体之间的复杂关系; 社会关系网络一般反映单一关系, 用于社群计算。CBDB 从本质上看, 是一个知识图谱, 反映的是人物之间复杂的社会关系(如亲属关系、师承关系), 如果使用不当, 将复杂关系扁平化并进行社会网络计算, 就会得出不合常理的结论。

因此, 单纯的数据驱动并不可取, 论证驱动中要辅以数据驱动, 数据驱动之后要回到论证驱动。

此外, 史学作品追求复原“历史本体”, 强调历史代入感。对读者来说, 描述性文章能在大脑中建立一种情景记忆。在情景记忆中, 心理表象与语词在心理操作中进行双重编码, 二者建立对应关系, 心理表象是产生代入感的原因。

3 结语: 一些建议

从认知科学角度, 数字史学研究要想取得“靠谱”的研究成果, 必须具备三个要素: “靠谱的人, 靠谱的记忆, 靠谱的计算”。针对业界的一些现象, 建议如下。

3.1 不宜过早开展数字史学研究

本科阶段是学业生涯的早期, 是知识结构建立和思考力培养的重要阶段, 本科生在这个阶段可以了解一下数字史学研究方法, 但不提倡开展数字史学研究。此外, 急功近利式的数字人文研究, 纯属自废武功。

3.2 不要盲目使用数字史学研究方法

不是所有的史学议题都适合使用数字史学方法, 经济史、社会史等偏社会科学的领域, 可以尝试; 数字史学研究过程中, 要充分了解数据来源、工具原理; 不能把记忆和计算全部交给机器, 传统史学方法不能丢。

3.3 不要盲目开展数字史学项目

数字史学是个新事物, 自然受到各方关注。近年来, 图书馆学界和数字出版界都认为这是一

个新的服务增长点,做了不少项目。应该说,数字史学领域的大多数项目是图书馆学或出版商在做。图书馆学界和数字出版界在开展数字史学项目时,建议如下。

■外化记忆要“求善保真”。要重视“数字善本”^[9],数字化的文本不能出错。

■外化记忆要完整。作品编排和版面组织也包含大量信息,要尊重原始文本,保留原始文本的完整信息并可检索。

■外化记忆的信息组织方式要科学。语义知识库的细粒度信息组织方式与人脑信息组织方式接近,在其碎片化加工过程中,一定要有领域专家的智慧投入,一定要符合学者的认知习惯。

■外化记忆要重视常识知识库构建。将常识甚至相关图像补充进知识库,这一点在知识图谱领域已经得到共识。

■要处理好信息检索与推送的关系,防止“信息茧房”出现。

■防止平台成为“计算黑箱”,机器的思维过程一定要公开透明。

■平台中计算方法和呈现方法要科学,比如合理使用知识图谱和社会关系网络技术,GIS分析中的底图要与历史时期匹配等,要合理配置模型、分词词典和算法等。

■平台的数据和功能要可以订制,用户可以自由组合数据和计算,这一点对服务商和用户都提出很高要求,相关培训一点要跟上。CBDB单机版开放所有数据结构、数据和程序代码,便于用户理解数据和使用平台;其与“中文在线”合作后,如何延续这种公开透明和自由订制,是这个项目商业化过程中的难点。

正像有的学者所说,回顾史学进步的各个关键点,技术层面的进步,似乎都并没有形成显著的决定性的推进力。也许计算机技术与此前的学术手段的进步有很多不同,但是要断言这种技术提升会在很大程度上改变史学的学术走向或说形成革命性的积极影响,除了开发想象力之外,还需要实践的证明^[24]。数字史学任重道远,冷静客观看待,小心进行探索,方能行稳致远。

【参考文献】

- [1] HEY T, TANSLEY S, TOLLE K. 第四范式:数据密集型科学发现[M]. 潘教峰, 张晓林, 等, 译. 北京: 科学出版社, 2012.
- [2] 姜义华. 大数据催生史学大变革[N]. 中国社会科学报, 2015-04-29(B5).
- [3] 邱伟云. 验证、修正、创新:数字史学方法的三重功能[J]. 南京大学学报(哲学·人文科学·社会科学), 2019, 56(2): 87-90.
- [4] 梁晨, 董浩, 李中清. 量化数据库与历史研究[J]. 历史研究, 2015(2): 113-128+191-192.
- [5] 常娥. 古籍计算机自动校勘、自动编纂与自动注释研究[M]. 芜湖: 安徽师范大学出版社, 2013.
- [6] 黄建年. 古籍计算机自动断句标点与自动分词标引研究[M]. 芜湖: 安徽师范大学出版社, 2013.
- [7] 王文涛. 信息时代的文献阅读和史料检索[J]. 史学月刊, 2015(1): 11-14.
- [8] 包伟民. 数字人文及其对历史学的新挑战[J]. 史学月刊, 2018(9): 5-12.
- [9] 陈爽. 回归传统:浅谈数字化时代的史料处理与运用[J]. 史学月刊, 2015(1): 14-17.
- [10] 胡优静. 历史学数字资源利用的误区及其应对[J]. 史学月刊, 2017(5): 26-29.
- [11] 李振宏. 论互联网时代的历史学[J]. 史学月刊, 2016(11): 97-113+137.
- [12] 陈志武. 量化历史研究的过去与未来[J]. 社会科学文摘, 2017(3): 88-91.
- [13] 王涛. “数字史学”:现状、问题与展望[J]. 江海学刊, 2017(2): 172-176.

- [14] 王旭东. 历史研究信息化应用模式刍议 [J]. 史学理论研究, 2006(4):136-147.
- [15] [美] R.M. 哈尼什. 心智、大脑与计算机: 认知科学创立史导论 [M]. 王森, 李鹏鑫, 译. 杭州: 浙江大学出版社, 2010.
- [16] [美] 加洛蒂 (K. M. Galotti). 认知心理学: 认知科学与你的生活 (原书第 5 版) [M]. 吴国宏等, 译. 北京: 机械工业出版社, 2015.
- [17] [加] 埃里克·麦克卢汉. 理解媒介: 论人的延伸 [M]. 何道宽, 译. 南京: 译林出版社, 2011.
- [18] 杨庆峰. 历史数字化、认知与记忆 [J]. 江海学刊, 2017(2):154-161+239.
- [19] 陈永明, 罗永东. 现代认知心理学: 人的信息加工 [M]. 北京: 团结出版社, 1989.
- [20] 孙辉, 王颖, 张智雄. 基于工具书语料的国史知识库构建和检索 [J]. 现代情报, 2016, 36(1):64-73.
- [21] 李德毅. 脑认知的形式化——从研发机器驾驶脑谈开去 [J]. 科技导报, 2015, 33(24):125.
- [22] 韩震. 阅读的本质 [N]. 光明日报, 2020-02-10(15).
- [23] 马克斯·韦伯. 社会科学方法论 [M]. 韩水法、莫茜, 译. 北京: 中央编译出版社, 1998:123.
- [24] 王子今. “史识”与计算机“利器” [J]. 史学月刊, 2015(1):9-11.

Observing Digital History from the Perspective of Cognitive Science

SUN Hui

(Institute of Contemporary China Studies, Chinese Academy of Social Science, Beijing 100009, China)

Abstract: [**Purpose/significance**] As a new historical research form, digital history is controversial in academia. Observing it from the perspective of cognitive science, the article aims to treat it objectively and rationally. [**Method/process**] The literature survey method is used to sort out the viewpoints in the history field. The cognitive science's theories of memory and mental computing are used to analyze the cognitive characteristics of traditional historical research and digital historical research, and to summarize the research process and key points. [**Result/conclusion**] Digital history is characterized by a high degree of externalization of some memories and calculations in the cognitive process. It has changed the cognitive means, cognitive methods, and cognitive processes of traditional historical research, and ultimately affected cognitive abilities and cognitive results. Historical skills training, intuitive thinking and argumentation drive are still important. It is not appropriate to conduct digital history research at the undergraduate level. This article also puts forward suggestions for the development of digital history projects.

Keywords: Digital history; Cognitive science; Digital humanities

(本文责编: 孔青青)