

基于数字人文视角的“情感—时空”模型探析

姜育彦, 李雅茹*

(南京大学金陵学院信息科学与工程学院, 南京 210089)

摘要: [目的 / 意义] 数字工具、技术的出现扩大了人文、艺术和社会科学领域对知识的传统理解, 重新定义了人文科学、社会科学、艺术学和自然科学之间的界限。[方法 / 过程] 笔者借助文本智能分词和情感极性分析等技术, 以提高方法的科学性和数据支持的量化程度。对从《唐宋名家词选》中选取的 48 首苏轼词作按创作时间排序后进行文本情感分析。通过情感极性分析技术判断作者的情感变迁, 基于相关的定量历史数据库 (提供历史地理信息支持), 建立作品内含情感和时空背景两者之间的关系模型。[结果 / 结论] 此模型的建立以期建立古典文学作品知识库、深入研究传统文化、考证古人创作时间、人物历史以及所相关的历史事件和地方史志追溯提供新的支持。

关键词: 数字人文; 史学研究; 情感分析; 中文分词

中图分类号: G258.6

文献标识码: A

文章编号: 1002-1248 (2020) 06-0023-11

引用本文: 姜育彦, 李雅茹. 基于数字人文视角的“情感—时空”模型探析[J]. 农业图书情报学报, 2020, 32(6): 23-33.

An Analysis of the "Emotion Space Time" Model from the Perspective of Digital Humanity

JIANG Yuyan, LI Yaru*

(School of information engineering, Jinling college, Nanjing University, Nanjing 210089)

Abstract: [Purpose / Significance] The emergence of digital tools and technologies has expanded the traditional understanding of knowledge in the fields of humanities, arts and social sciences, and redefined the boundaries between humanities, social sciences, arts and natural sciences. [Method / Process] Based on the special research situation of "Su Ci", this paper introduces the research of text polarity emotion analysis into historiography, in order to enhance the scientificity of methods and the quantifiable degree of data support. By digitizing the paper data (such as OCR) to generate data files that are convenient for computer use, and importing them into the emotion analysis system, the

收稿日期: 2020-03-16

基金项目: 江苏高校哲学社会科学一般项目“自我调节理论视角下求职者社会化信息搜寻行为影响机理及对策研究” (项目编号: 2019SJA0152)

作者简介: 姜育彦 (1999-), ORCID: 0000-0002-8353-5179, 男, 本科, 研究方向: 自然语言处理。

通讯作者: 李雅茹 (1985-), 女, 硕士, 讲师, 研究方向: 数据分析, Email: 651887786@qq.com。

corresponding relationship of "emotion age" is generated based on the existing historical data obtained by the same method. Based on the relevant quantitative historical database (providing the support of Historical Geographic Information), the model of "emotion time space" is established. [Results / Conclusions] The establishment of this model is expected to provide new support for the establishing the knowledge base of classical literature works, the in-depth study of traditional culture, the textual research of the creation time of the ancients, the history of the characters, the related historical events and the local historical records.

Keywords: Digital Humanities; historical research; emotional analysis; Chinese word segmentation

1 引言

自罗伯托·布萨将计算机技术成功应用于人文学科领域后^[1], 数字人文研究便逐渐影响文学、历史、文学艺术等领域, 形成了强有力的时代浪潮。领域的碰撞使得人文科学、社会科学和自然科学的界限变得模糊, 多学科交叉的研究开始越来越多地涌现出来。在技术的不断推进下, 传统文化领域也受到数字技术的影响, 开始促进“数字”与“人文”有机结合。中国五千年的不曾间断的历史传承, 也为数字人文研究提供了充分的选择。

由于在传统研究中史料、文学作品都具备专业性和特殊性。如何将文字和特定的历史时空衔接、融合是一大难点^[2]。数字人文引领的风潮使得中国历史的研究充满了无限的活力, 是历史学研究前所未有的机遇。笔者尝试以情感角度作为新的历史研究窗口, 基于量化历史数据库的语义检索、信息关联、可视化生成等功能, 研究苏词背后的历史信息, 拉近历史地理学与计算机科学之间的距离, 以及在量化历史数据库信息补充、多维度建设和开放利用共享等方面的进一步探索。

2 相关研究

2.1 数字与文学的结合性

在近年探讨数字与文学结合的相关研究中, 大多着力于古典文学数字文献化的研究和基于已有数据库的深入探索两大方面 (如图 1 所示)^[3]分别阐述如下。

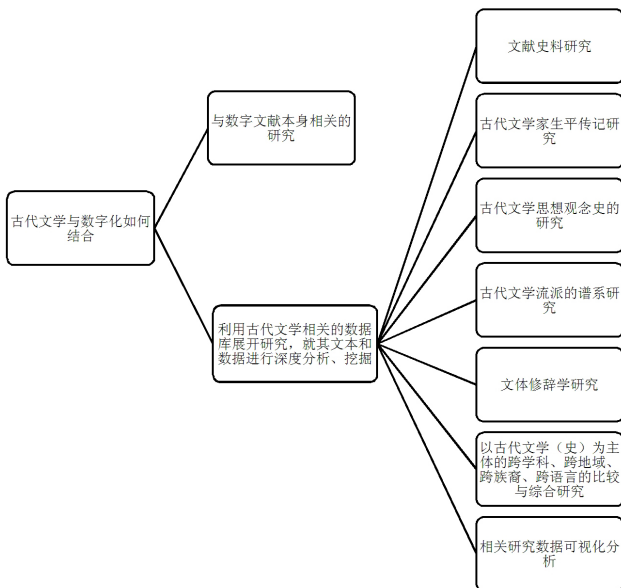


图 1 数字与古代文学结合形式分类图

2.1.1 古典文学数字文献化

(1) 资源形式: 高新陵将纸质文献和数字文献的整合分为 3 个层次, 并逐一分述比较进而比较两种载体资源整合体系的学科倾向和人文价值追求, 分析研究它们的不足与互补^[4]。

(2) 应用实践: 林泽明等对徽州的历史与自然环境、村落、建筑、宗族社会、商派、菜系、科技、医学、理学、朴学、人物、教育、工艺、文书档案、民俗、方言、戏曲、版画、篆刻等信息进行遴选, 构建徽州水系文化地图特色数据库, 有助于徽州文化的广泛传播^[5]。

(3) 技术手段: 郑永晓提出数字文献学是以利用计算机和数据库技术从事古籍整理和文献研究, 对中国古典文学、哲学史学的发展产生重要影响^[6]。郭利敏等尝试将卷积神经网络用于数字人文古籍汉字的元数

据加工,将古籍汉字识别问题转换为卷积神经网络的分类问题,在缺乏训练集的情况下通过数据生成技术构建训练集进行模型训练,用于古籍汉字的识别^[7]。尹晓琳认为在大数据时代语境下,少数民族民间文学资料信息化平台的建设至关重要,通过数据库可以实现资源共享,便于研究与传承民间文化^[8]。欧阳剑提出,以大规模中国古籍文本为研究对象,采用大数据研究理念,对古籍进行处理、分析、计算、预测,创建了一个以语言学、历史文献学、历史地理学等人文学科研究为主的古籍实时统计分析平台,可辅助研究者在大量的古籍文献中发现新的模式、现象、趋势^[9]。王洁认为利用大数据研究与传播中国传统文化是大势所趋,并阐述其目前已经过了数字典藏与数字人文两个阶段的发展并提出在坚持大数据思维,继续积极建设专业、完备、可以加值的数据库的同时,须警惕科技对学术研究与人文思维的负面影响,并且不能忽视相关法律法规的建设^[10]。

2.1.2 基于已有数据库的深入探索

(1) 资源组织:赵思渊以文献学研究为基础,建立特定的元数据结构,提供交叉导航、数据统计等功能,在帮助其他研究者获取所需文献的基础上探索新的研究方向^[11]。姚天泓以张学良史料数据资源为例,基于 CIDOC-CRM 概念框架模型,从数据描述规范、语义标注等角度进行史料数据资源的语义化知识组织研究,从而使研究者通过数字人文研究方法和技术手段转变对数据资源的研究视角,得出超出一般传统阅读、理解、辨析处理的能力范畴,获取不同于传统人文社会科学研究的观察、领悟深度和范围^[12]。毕强、刘健运用聚类分析、语义相似度计算、协同过滤推荐算法等方法,提出了基于领域本体的数字文献资源聚合及服务推荐的方法和途径,为优化其资源再组织结构 and 提升知识服务能力提供有价值的参考与指导^[13]。夏翠娟提出重新认识家谱文献价值的思路和深度开发家谱文献内容的方法及路径并以上海图书馆家谱知识服务平台相关内容的为例,验证了所提出的思路、方法的可行性^[14]。

(2) 应用实践:朱春洁通过对知网、万方等数据

库收录的流贬研究论文、专著进行统计,结合关键词互现网络系统发现这些成果的分布与研究特点^[15]。杨海慈等基于中国历代人物传记资料库(CBDB)数据,构建宋代学术师承本体,并对数据库中的文人关系数据进行语义化转换,在此基础上利用知识图谱的原理和方法描述宋代文人的学术师承关系,开发“宋代学术语义网络”平台展示知识图谱的知识架构和数据内容,为相关历史研究者和爱好者提供知识发现和探索的数据入口^[16]。蔡迎春构建了一个包括原版民国时期文献目录、新版民国时期文献书目和新版民国时期文献子目的目录型“数据平台”,同时探讨名称规范档、GIS技术和可视化技术在“数据平台”中的应用与实践^[17]。严承希从数字人文视角出发,借助符号分析方法对哈佛大学“中国历代人物资料库”进行实证探索与可视化分析,结合已有的史学问题和相关观点,从宋代政治整体网络分布特征、核心人物的地位与结构拓扑以及不同时期宋代政治网络的时序政治关系演化模式3个层次,进行逐一分析与讨论,为研究宋代党争政治格局提供了一种新的思考方式,同时也展现了数字人文技术在人文历史领域研究中的可行性与巨大潜能^[18]。严程以数字人文手段处理顾太清与沈善宝等10余位诗友的往来诗作,将文献中涉及的人物、时间和事件信息转换成变动的人际网络。可以发现以顾太清为中心的女性诗人交游群体在诗社存续期间的两次重大结构变化。在此基础上,析出诗社的建立、波折、聚散等时间节点,破解这一清代满汉融合闺秀诗社的兴衰变迁之谜^[19]。

(3) 方法探索:陈诚研究了图书馆古籍文献专题数据库的构建技术及规范;着重从建立统一的古籍长期保存标准、人才培养机制、数字化共享等方面,提出建立图书馆古籍文献长期保存机制的策略^[20]。袁悦等基于已完成人工校验和机器自动标注的《左传》与《国语》构成的训练和测试语料,以南京师范大学先秦词性标记集为主、以北京大学、中国科学院计算技术研究所和教育部词性标记集为辅,共形成3种不同大小的新标记集,通过条件随机场以及添加特征模板比较这3种词性标记集合在同一语料上进行实体抽取结

果的差异性,并基于该结果探索进一步提升训练结果^[21]。胡以涛等以农业文化遗产保护和研究为例,选择元数据(DC)规范提炼核心元数据集,并根据农业文化遗产特点进行元素扩展,采用RDF/XML描述方案,构建一套完整的人文对象数字资源元数据标准。其实践表明元数据方法的采用有利于对数字人文对象的科学管理与研究应用^[22]。王涛运用数字工具对人物志的数据挖掘,发现了德意志历史人物群体的隐含问题,甚至有悖于常识的结果,揭示出德意志主流历史学家对“德意志人”的想象构建。通过对大量德意志人物诞生地与逝世地的社会网络分析构成了历史人物的“死亡地图”,清晰显示了德意志重要历史名城由南部向北部扩散的过程^[23]。

从上述研究中可以发现,通过数字化手段进行“全景”式的人文研究,可以彰显数字和文学之间紧密的结合性,加强各研究领域间的沟通与合作,避免人文、社会科学研究走进死胡同。

2.2 新研究维度的史学依据

2.2.1 时空维度

如司马迁所说的“通古今之变,成一家之言”,“历史”这一概念本身指的就是对一切“时间”中发生变迁现象的研究。其中时间问题是历史学家,尤其是年鉴学派历史学家躲避不开也最为关心的问题^[24]。费尔南·布罗代尔指出:历史学家无法摆脱时间问题^[25]。雅克勒·高夫则进一步阐明:时间是历史的基本材料^[26]。可见传统的史学研究主要注重时间维度。

中南民族大学王兆鹏教授于2012年开始主持《唐宋文学年表信息平台建设》^[27]项目,形成一套唐宋文学编年史的平台。这一平台首次提出了“系地”的理论支撑点,将时间、人、活动和作品构建起庞大的知识体系并付诸实践。这种基于大数据的历史趋势研究所得结果是传统的研究范式无法实现^[28]。这样的文学地理学研究以立体的方式呈现人群的活动,展现一个时期内历史变动大趋势,为相关学者从不同角度宏观、微观的研究探索提供了助力^[29]。在打开文学创作的空间维度大门的同时,却又缺失“历史个体—特定地

点—人物群体—历史事件”的多维度延伸。

2.2.2 情感维度

上世纪20年代,法国社会历史学家哈布瓦赫在《论集体记忆》中首次提出了“集体记忆”,认为:(1)集体记忆被容纳在一个由人们构成的聚合体中;(2)集体记忆在历史群体中存续并且以其基础中汲取力量;(3)集体记忆不容纳非群体成员中的个体;(4)每一个集体记忆都需要得到在时空被界定的群体的支持^[30];(5)“时空”载体是“记忆”以及“集体记忆”赖以产生的温床;(6)记忆是立足于现在的对过去的重构^[31]。到了上世纪80年代,另一位社会历史学家皮埃尔·诺拉对此进行深度“挖掘”,进行对“记忆”与“场所”的深入研究。这标志着记忆场所理论得以系统建构并进入到体系化、应用化阶段^[32]。通过历史记忆研究方法的发展历程可以得出:集体记忆是构建人物时空关系的重要历史研究手段。古典文学作品作为个体记忆的文字载体,在流传的过程中逐渐将个体记忆扩散为集体记忆,使得个体记忆突破时间、空间的限制,得以延伸,在更广泛的时空范围内共享。而情感作为个体记忆的重要组成部分可以作为数字人文视角下的历史研究突破口,通过情感角度对古典文学作品的深入研究也有利于构建历史事件和人际网络之间的桥梁(如图2所示),使深藏在历史深处的集体记忆得以量化地呈现给研究人员。

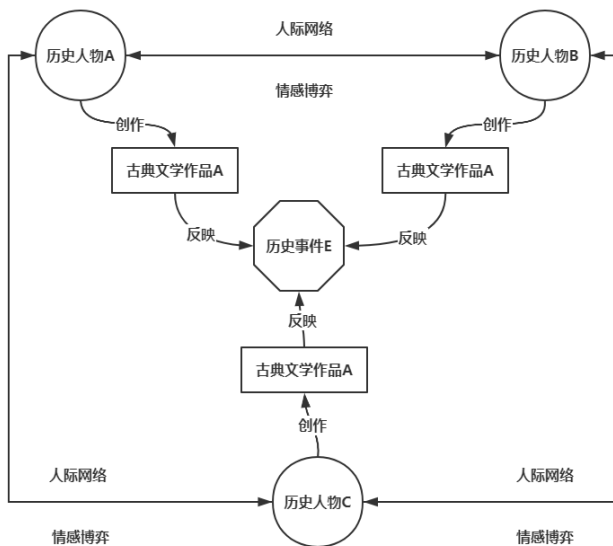


图2 深入研究模型设想

3 情感角度入手的苏词历史信息研究模式设计

3.1 “情感—时空”生成模型设计

由于纸质文本中历史信息的组织模式，往往是离散分布而不是连续罗列的。这些信息的处理、组织和解释需要考虑来源材料相对应的学科规范和研究方法论。考虑到能够使生成数据让机器理解并按一定标准进行有选择地表示和描述，并广泛应用于可视化和数据挖掘。建立数据库的过程需要格外重视人文数据的完整性、可计算性、可重用性、可发现性和可用性^[3]。

基于“苏词”这一特殊研究情景，笔者通过将文本极性情感分析研究引入史学，以增强方法上的科学性和数据支撑上的可量化度，并建立如下模型（如图3所示）。通过将纸本资料进行数字化处理（如OCR）生成便于计算机利用的数据文件，将其导入至情感分析系统中，以同样方法处理获得的现有历史资料为依据，生成“情感—年代”对应关系，在相关量化历史数据库（提供历史地理信息支撑）的基础上，以历史事件为纽带构建“事件—时间”、“事件—空间”、“事件—情感”、“事件—人物”的关联数据组合，建立“情感—时空”模型。

3.2 基于“情感—时空”模型的知识管理模式设计

考虑到古典文学作品在经历文本图像化、图像数字化，形成半结构化或者结构化的数据库后，用户的需求也全面升级，使用深度也随之大幅提升。这就需要研究者整合关联知识、结构化数据资源、配备调用工具。基于从情感角度研究“苏词”相关历史的特殊情景，知识管理模式应满足以下需求（如图4所示）：

(1) 采用合适的数据库模型。数据模型需要考虑到与现有的历史地理数据库、各种相关开放关联信息数据库、数字人文项目的开放数据平台（如中国历代人物传记资料库（CBDB）、唐宋文学编年地图平台、上海图书馆家谱知识库等）等相兼容，以支持不同团队研究成果的融合，同时要具备良好的可扩展性便于后续研究进一步的修改和拓展。

(2) 采用标准化的编码格式。数据的编码格式与数据使用、共享的便捷性紧密关联。标准化的数据编码格式有助于数据在不同系统平台（Linux、Windows、Android、IOS）之间的传输、开发和应用。

(3) 实施地名规范控制和时间片数据序列。根据用户需求的时间片段、地点范围，使得每一个历史阶段上曾经出现过的地名，都有唯一的 URI 并可以在现今的地图数据上被标识，提供关联数据的索引和简述。

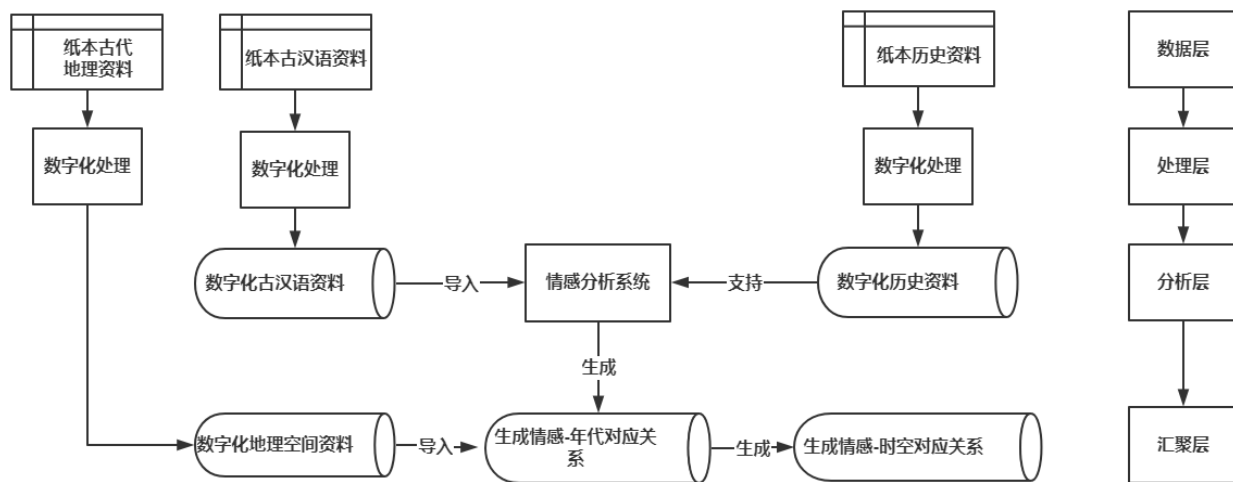


图3 “情感—时空”对应模型

当用户提供地名（或特定 URL）和一定的时间区间后，返回关于该地名存续的时空范围、过往名称、行政归属，以及有关联的历史人物、历史事件、文学作品。

(4) 基于 API 提供开放数据支持。Web 是互联网的主要载体，基于 WEB 服务的 API 接口可以提供自由的数据访问环境，是互联网环境下数据应用、开放分享、协同建设的常规选择。

4 情感角度入手的苏词历史信息研究实证

4.1 苏词的历史研究意义

宋代的士大夫们也在取得了显赫的政治的同时形成了崇高的社会责任感，开始以国家主人翁的心态去追逐自己的人生追求^[34]。但大多数人经历残酷政治打击后，都退回到了自己的内心世界，为个体生命寻求精神上的寄托。这样的人生历程使得作品大多包含着作者兼济天下和独善其身并存的精神理念，显得“平淡

见真纯，质朴显自然”。使得在士大夫这一群体中，“入世—出世”成了绝大多数人生命历程的主体。在这一阶段，佛教以其独特的形式逐渐渗透于士大夫阶层，使得宋代士大夫兼具随遇而安的旷达与喜悦与儒家的积极入世精神^[35]。

以苏轼为代表的士大夫群体，以儒学思想主导人生的基本方向，以佛老思想客观通达地观察问题，在经受政治上的挫折和生活中的磨难的同时，对生活抱有热情，坚持着对人生及美好事物的积极追求，寻觅着心灵的净土^[36]。所以苏轼一生的几起几落，不仅带给他旺盛的创作活力，更赋予了其作品饱满丰富的情感，把宋代诗歌的个性推向了新的高度。他用自身的创作实践表明：词是“无事不可写，无意不可入”的。他的作品充分表达了当时的社会面貌和现实生活，揭示了社会发展的风貌，可以帮助人文学者回到“历史现场”，具有较高的史学研究意义。

《宋代文学通论》、《宋词文化学研究》、《宋词与民俗》等相关文献资料，将宋词置于特定的历史时期、文化空间中进行研究分析，还原文字其中的历史信息。

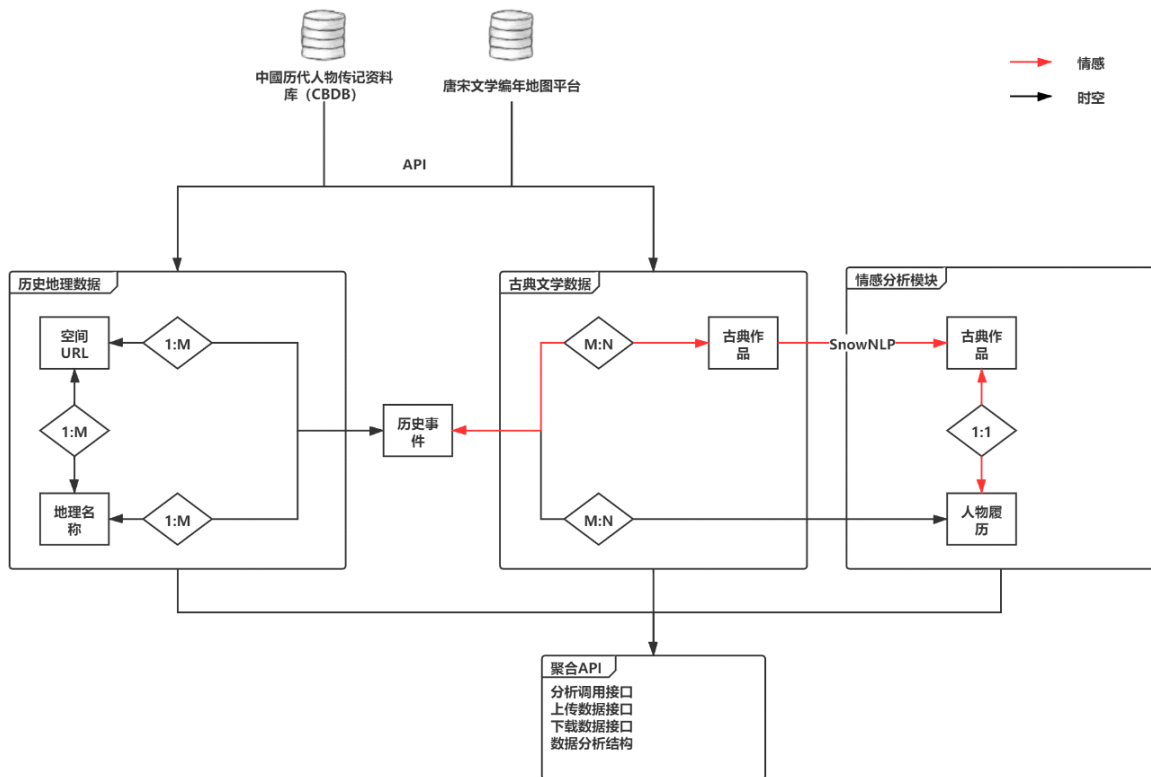


图 4 知识管理模式结构图

将宋词文本中富含的历史、文化、社会风俗的信息挖掘出来,进行整理、归纳和总结,从而揭示某个特定时期历史人物的生活群像和文化心理^[37]。而笔者选择充分利用“苏词”作为研究材料,尝试从情感分析角度探索文本所蕴含的丰富历史信息并试图从已复现的历史情境中探究历史细节、追溯人物真相,探析多角度感受历史发展脉络的实践可行性。

4.2 苏词“情感—时空”关系实证

后现代史学理念中最具冲击力和颠覆性的观点,在于指明历史写作的“虚构”性质。由于文本中的历史信息是离散而不是连续的,这些信息的处理、组织和解释需要考虑学科规范和研究方法。通过将大数据分析引入史学研究,能够增强方法上的科学性和数据

支撑上的可量化度,这样有助于提高历史研究的确切性和可靠性^[38]。

笔者以词学大师龙榆生先生的《唐宋名家词选》^[39]作为数据来源,共选取48首。通过唐宋文学编年地图^[27]API输入词名查询,获取作品创作年代。基于SnowNLP^[40]的自训练模型进行情感判断。通过将相关数据导入模型处理,得出48首词作的情感极性数据(如表1所示),并绘制成图。参考苏轼在历史记载中的生活轨迹(如表2所示)比较两张表可以发现作品的情感极性值和苏轼人生轨迹的变化幅度基本一致(排除3.2.3—3.2.5影响),符合2.1中设计的情感年代对应关系。相关数据说明如下:

(1)表1中, $y \geq 0.5$ 时表示作者当年呈现积极表现, $y \leq 0.5$ 时表示作者当年呈现消极表现。

表1 文本情感参数

作品	年代	情感参数	情感极性	作品	年代	情感参数	情感极性
《江城子·湖上与张先同赋》	1072年	0.999 993 9	积极	《醉翁操·琅然》	1082年	1.000 000 0	积极
《行香子·过七里瀨》	1073年	0.999 999 9	积极	《洞仙歌·冰肌玉骨》	1082年	0.074 657 2	消极
《永遇乐·长忆别时》	1074年	0.999 999 9	积极	《满庭芳·三十三年》	1082年	0.742 778 8	积极
《少年游·去年相送》	1074年	0.001 398 4	消极	《哨遍·为米折腰》	1082年	1.000 000 0	积极
《虞美人·有美堂赠述古》	1074年	0.999 967 3	积极	《念奴娇·中秋》	1082年	1.000 000 0	积极
《南乡子·送述古》	1074年	0.991 905 8	积极	《念奴娇·赤壁怀古》	1082年	0.793 407 5	积极
《蝶恋花·密州上元》	1075年	0.061 806 1	消极	《满江红·寄鄂州朱使君寿昌》	1082年	0.840 158 3	积极
《江城子·十年生死两茫茫》	1075年	0.000 009 4	消极	《南歌子·雨暗初疑夜》	1082年	0.003 611 3	消极
《水调歌头·明月几时有》	1076年	0.000 853 5	消极	《浣溪沙·山下兰芽短浸溪》	1082年	1.000 000 0	积极
《阳关曲·中秋作》	1077年	0.091 840 7	消极	《水龙吟·小舟横截春江》	1082年	0.002 041 6	消极
《洞仙歌·江南腊尽》	1077年	0.001 889 2	消极	《定风波·莫听穿林打叶声》	1082年	1.000 000 0	积极
《永遇乐·明月如霜》	1078年	0.006 236 4	消极	《临江仙·夜饮东坡醒复醉》	1083年	0.010 155 9	消极
《浣溪沙·软草平莎过雨新》	1078年	1.000 000 0	积极	《水调歌头·落日绣帘卷》	1083年	1.000 000 0	积极
《浣溪沙·麻叶层层荷叶光》	1078年	1.000 000 0	积极	《浣溪沙·细雨斜风作小寒》	1084年	0.999 972 5	积极
《浣溪沙·旋抹红妆看使君》	1078年	0.999 991 3	积极	《水龙吟·露寒烟冷蒹葭老》	1084年	0.822 133 5	积极
《浣溪沙·照日深红暖见鱼》	1078年	0.999 999 6	积极	《定风波·常羨人间琢玉郎》	1086年	1.000 000 0	积极
《浣溪沙·簌簌衣巾落枣花》	1078年	0.999 999 2	积极	《点绛唇·不用悲秋》	1090年	0.999 990 0	积极
《卜算子·黄州定慧院寓居作》	1080年	0.000 449 5	消极	《减字木兰花·空床响琢》	1091年	0.999 999 9	积极
《南乡子·霜降水痕收》	1081年	0.234 004 8	消极	《木兰花令·次欧公西湖韵》	1091年	0.004 590 3	消极
《浣溪沙·醉梦昏昏晓未苏》	1081年	0.999 997 7	积极	《浣溪沙·门外东风雪洒裾》	1091年	0.999 993 4	积极
《浣溪沙·覆块青青麦未苏》	1081年	0.999 958 5	积极	《八声甘州·有情风万里卷潮来》	1091年	1.000 000 0	积极
《水调歌头·昵昵儿女语》	1081年	0.000 890 2	消极	《青玉案·三枕上吴中路》	1093年	1.000 000 0	积极
《水龙吟·次韵章质夫杨花词》	1081年	0.004 029 7	消极	《贺新郎·乳燕飞华屋》	1095年	0.250 149 4	消极
《南歌子·雨暗初疑夜》	1082年	0.003 611 3	消极	《鹧鸪天·林断山明竹隐墙》	1100年	0.076 606 9	消极

表2 已有历史信息

年代	年号	履历	官阶	人生变化
1070年	熙宁三年	在京任监官告院	从六品	-
1071年	熙宁四年	在京任监官告院兼判尚书祠部 以太常博士直史馆通判杭州	从六品 从五品	积极 积极
1075年	熙宁八年	密州太守	从五品	消极
1077年	熙宁十年	徐州太守	从五品	消极
1079年	元丰二年	潮州太守	从五品	消极
1080年	元丰三年	贬黄州团练副使	从八品	消极
1084年	元丰七年	谪移汝州团练副使	从八品	消极
1085年	元丰八年	任朝奉郎, 知登州, 在京任礼部郎中	从六品	积极
1086年	元祐元年	在京为中书舍人, 翰林学士知制造	正三品	积极
1087年	元祐二年	在京为翰林学士, 复任侍读学士	正三品	积极
1088年	元祐三年	在京任翰林学士, 知制造兼侍读, 权知礼部贡举	正三品	积极
1089年	元祐四年	以龙图阁学士知杭州	正三品	积极
1091年	元祐七年	吏部尚书以龙图阁学士知颍州 以龙图阁学士知扬州 以端明殿学士兼翰林侍读学士任礼部尚书 以兵部尚书、龙图阁学士除兼侍读	正二品 正三品 正二品 正二品	积极 消极 积极 积极
1093年	元祐八年	以端明殿学士、翰林侍读学士知定州	正三品	消极
1094年	元祐九年	贬宁远军节度使, 惠州安置	从八品	消极
1097年	绍圣四年	贬琼州别驾, 昌化军安置	正九品	消极
1100年	元符三年	舒州团练副使, 永州安置 复任朝奉郎提举成都玉局观	从八品 正六品	积极 消极

(2) 表2中, 以苏轼官场仕途升迁为衡量因素, 1~18为从九品至正一品。考虑到宋代官员的官职由职事官、散官、勋官组成, 数据优先考虑职事官阶, 在职事官阶不明时则以散官官阶为准。

(3) 在表1、表2中1075年—1080年所对应的情感数据呈消极表现而官阶数据未波动, 是因为考虑到京官与外迁的影响修正。

(4) 在表1、表2中1084年所对应的情感数据呈积极表现而官阶数据下降是由于在苏轼奉诏赴汝州就任途中幼儿夭折且路费已尽。苏轼便上书朝廷, 请求暂时不去汝州, 先到常州居住, 后被批准。他在常州居住, 既无饥寒之忧, 又可享美景之乐, 而且远离了京城政治的纷争, 能与家人、众多朋友朝夕相处。

(5) 在表1、表2中1093年情感数据呈积极表现而官阶数据下降, 是作者为送客而作。为赞赏苏坚的

故乡, 称许苏坚的才华, 写其对家乡的思念, 祝贺他的归乡。所以纵然作者自身境遇不佳, 但是该作仍饱含对友人的美好祝愿和羡慕之情。

(6) 在表1、表2中1100年所对应的情感数据呈消极表现而官阶数据上升, 是因为考虑到宫观使大多为无实职虚衔的影响修正。

4.3 基于地理信息、“情感—时间”关系构造“情感—时空”关系

空间信息与情感信息之间很难构造直接性的关联, 但可以基于4.2中得出的“情感—时间”关系和开放使用的历史地理数据库(如图4所示), 可以通过“事件—空间”、“事件—情感”的信息关联模式(如图5所示), 构造“情感—时空”关系。



图4 历史空间信息的图像化展示

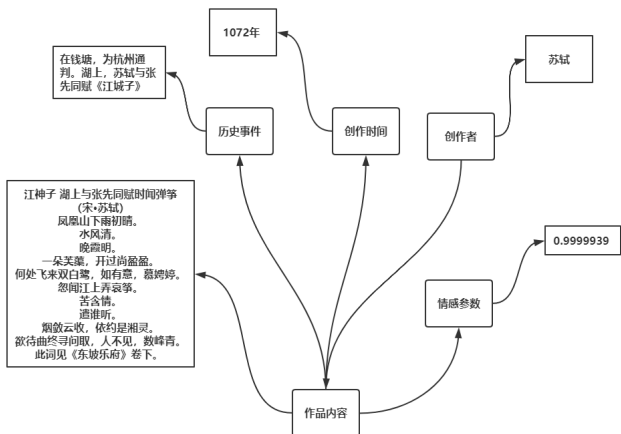


图5 “事件—时间”、“事件—空间”、“事件—情感”、“事件—人物”聚合展示

5 情感分析方法价值分析

5.1 在保护历史的同时更好地利用

数字人文可以通过深挖文献, 创造应用增值, 给使用者提供多领域、深层次的传统文化知识。相关使用者可以根据自己的需求, 用检索的方式, 从海量序列化、规则化的数字典藏文献中搜寻所需的资料, 然后将这些资料整合成一个能满足专业需求的资料集, 很好地解决了历史文献(尤其是古籍善本)利用的空间限制和文物保护问题^[41]。

5.2 深度挖掘古典文学的价值所在

古典文学的内容与时代背景是紧密结合的, 所以古人作品中含有着大量历史信息, 具有高度的研究价值。邵炳军根据现有的文献资料与考古发掘的材料, 梳理了“二王并立”时期的史实线索和文化背景^[42]。韩国河, 惠夕平, 闫广宇等在3S技术的支撑下, 基于《国语》和《史记》中“桓公寄孀”的记载, 对官庄城址周边地区进行了系统调查和测绘^[43]。这对研究考证特定历史时期人物的生平事迹及其文学作品的创作年代均具有重要意义。

5.3 社会记忆的建立

数字人文作为数字信息技术与人文社会科学深入

融合的产物, 在信息资源的开发利用方面具有明显优势。以徽州文书研究^[44]为例, 相关学者通过大量保存的长期族谱等文献史料和建筑等非文献史料, 深入探索特定时间、地点的社会全貌。作为“历史的新同盟”, 数字人文将带来新的历史眼光, 让历史阐释带来更多可能性的呈现, 在做到尊重数字方法的客观性并结合历史语境的想象力^[44]的前提下, 为社会记忆的构建提供有益的启示, 在人类文明传承中起到重要的作用。

6 情感分析方法实践中的问题

6.1 诗词的创作特点带来的分析难度

清初的王夫之关于“情景互”的观点, 叶燮关于“形依情, 情附形”的观点, 虽然已为境界说中的本质论奠定了基础。最终清代学者王国维在《人间词话》中说“一切景语皆情语”最系统地阐述了艺术境界中“景”与“情”的关系。此处以苏轼本人的诗作《新居》为例, 全诗却无一字直接写新筑之室, 而是通过其他的事物描写自己的心情。

朝阳入北林, 竹树散疏影。
短篱寻丈间, 寄我无穷境。
旧居无一席, 逐客犹遭屏。
结茅得兹地, 翳翳村巷永。
数朝风雨凉, 畦菊发新颖。

俯仰可卒岁，何必谋二顷。

所以在未建立古汉语意象相对应的“实体 (Entity) —类型 (type) —属性 (property)”的本体三元组数据时，“朝阳”、“竹树”、“短篱”这一类没有明确情感倾向的数据往往在机械化的处理中忽略的不同场景下的本体的离子型^[45]。

6.2 缺少专业高效的语料处理工具

现有的自然语言处理系统大多建立在现代汉语乃至外语的数据来源训练而成。使用该类工具研究的结果使一部分古汉语独有的词汇和表达形式丧失在模型精准性的影响下，造成结果的一定程度的偏离。清代诗人赵翼在《题遗山诗》写道：“国家不幸诗家幸，赋到沧桑句便工。”王朝的成败兴衰是一剂强有力的催化，很大程度上影响了个人的情感寄托，导致文人往往成就在动乱的年代。如南唐后主李煜，“四十年来家国，三千里地山河。凤阁龙楼连霄汉，玉树琼枝作烟萝，几曾识干戈？”又如文天祥一句“人生自古谁无死，留取丹青照汗青”留名于史。国破家亡之际的诗人不管是愤慨、悲凉、无奈，何种的情感爆发是活在安逸之中的人们往往无法经历到的。所以在样本训练时积极的素材过少的情况下，如何更精确地区分情绪极性仍是一个值得深入探讨的问题。

6.3 脱离序列文本体系造成的信息缺失

历史文献碎片化的形态决定了其每一个零散个体的研究价值通常要置于一个以时间和空间来划分“文献群”为单位展开的整体研究环境中才能被发现^[46]。所以随着研究的深入，对单件作品的分析，常常不能满足研究的需要，往往需要转入以或时间、或地域、或社会网络、或多数者融合为关联的文献序列，进行多文献综合分析的研究，有待于在本模式基础上进一步开发前后语段极性转折分析工具。

7 总结

通过基于前人研究的思考，笔者提出了在以历史

人物情绪介入历史时空研究的设想，并证实该探索角度的理论依据。通过自然语言处理（文本智能分词、情感极性分析）的技术手段，以苏轼的48首词为蓝本进行研究，所得结果在排除额外影响后与其真实生活经历一致。由此得出情感分析是可以作为介入历史的研究手段。通过这一手段，可以在时间信息、地理信息、人物信息都缺乏时中通过作品信息这一维度来进行未知问题的设想，或是只是缺少某一种信息是予以补充和判断。在王兆鹏教授主导建设的《唐宋文学年表信息平台》的“时间—地点”的二维结构之外加以情感倾向的补充，构建“情感—时空”关系，可以对历史时空进行更完善，更细腻的探索。在文章进行的研究过程中，提出了情感分析方法相比传统史学研究范式的价值，与实践应用中的不足，其训练所需的“语料—情感倾向”知识库，有待后续研究进一步完善，并在后续研究过程中尝试基于唐宋文学年表信息平台接口的历史人物群像感情分析工具的开发。

参考文献：

- [1] Busa R. The annals of humanities computing: the index thomisticus[J]. Computers and the Humanities, 1980, 14(2): 83-90.
- [2] 梁晨. 量化数据库：“数字人文”推动历史研究之关键[J]. 江海学刊, 2017(2): 162-164+239.
- [3] 王贺. “数字人文”如何与现代文学研究结合[J]. 现代中文学刊, 2019(1): 78-80.
- [4] 高新陵. 纸质文献资源与数字文献资源的整合比较研究[J]. 情报科学, 2012, 30(10): 1487-1491.
- [5] 林泽明, 胡迟, 晏实江. 徽州水系文化地图特色数据库构建探析[J]. 大学图书情报学刊, 2019, 37(5): 16-19.
- [6] 郑永晓. 传承与超越: 数字文献学的未来发展刍议——兼论日本文献数字化对我国之启示[J]. 中国比较文学, 2019(4): 2-13.
- [7] 郭利敏, 葛亮, 刘悦如. 卷积神经网络在古籍汉字识别中的应用实践[J]. 图书馆论坛, 2019, 39(10): 142-148.
- [8] 尹晓琳. 少数民族民间文学资料信息化建设探析[J]. 图书馆学研究, 2016(21): 41-44.
- [9] 欧阳剑. 面向数字人文研究的大规模古籍文本可视化分析与挖掘[J]. 中国图书馆学报, 2016, 42(2): 66-80.

- [10] 王洁.大数据思维与数字人文的加值应用——传统文化数据库发展的新趋势[J].图书馆理论与实践,2018(5):104-108.
- [11] 赵思渊.地方历史文献的数字化、数据化与文本挖掘:以《中国地方历史文献数据库》为例[J].清史研究,2016(4):26-35.
- [12] 姚天泓,陈艳梅,刘革,鲁超.基于 CIDOC-CRM 的数字人文史料资源语义化知识组织研究——以张学良史料资源为例[J].图书馆学刊,2019,41(7):35-43.
- [13] 毕强,刘健.基于领域本体的数字文献资源聚合及服务推荐方法研究[J].情报学报,2017,36(5):452-460.
- [14] 夏翠娟.文化记忆视域下家谱文献价值的再认识和内容的深开发[J].图书与情报,2019(5):49-57.
- [15] 朱春洁.数字人文视角下中国古代流贬研究文献可视化分析[J].图书馆,2020(1):29-35.
- [16] 杨海慈,王军.宋代学术师承知识图谱的构建与可视化[J].数据分析与知识发现,2019,3(6):109-116.
- [17] 蔡迎春.民国时期文献目录数据平台的构建与实践[J].图书馆论坛,2019,39(3):121-127.
- [18] 严承希,王军.数字人文视角:基于符号分析法的宋代政治网络可视化研究[J].中国图书馆学报,2018,44(5):87-103.
- [19] 严程.顾太清交游网络分析视野下“秋红吟社”变迁考[J].山东社会科学,2018(7):64-69.
- [20] 陈诚.数字人文时代图书馆古籍文献资源的数字化长期保存机制探讨[J].中国中医药图书情报杂志,2019,43(3):6-9.
- [21] 袁悦,王东波,黄水清,李斌.不同词性标记集在典籍实体抽取上的差异性探究[J].数据分析与知识发现,2019,3(3):57-65.
- [22] 胡以涛,惠富平.元数据方法在数字人文视域下的应用探索——以农业文化遗产为例[J].图书馆,2019(1):82-87.
- [23] 王涛.数字人文框架下《德意志人物志》的群像描绘与类型分析[J].历史研究,2018(5):148-166+192.
- [24] 雅克·勒高夫.圣路易[M].商务印书馆:北京,2002:14.
- [25] 布罗代尔.论历史[M].北京大学出版社,2008:51.
- [26] 勒高夫.历史与记忆[M].中国人民大学出版社:北京,2010.
- [27] 唐宋文学编年地图平台[EB/OL]<https://sou-yun.cn/PoetLifeMap.aspx>
- [28] 王兆鹏,郑永晓,刘京臣.借器之势,出道之新——“数字人文”浪潮下的古典文学研究三人谈[J].文艺研究,2019(9):79-88.
- [29] 梅新林.“新文学地理学”学术体系之建构[J].浙江社会科学,2017(7):112-125+159.
- [30] 莫里斯·哈布瓦赫.论集体记忆[M].毕然,郭金华译.上海:上海人民出版社,2002:39-40.
- [31] 杰罗姆·特鲁克.对场所的记忆和记忆的场所:集体记忆的哈布瓦赫式社会——民族志学研究[J].国际社会科学杂志,2012(4):33-46.
- [32] 皮埃尔·诺拉.记忆之场——法国国民意识的文化社会史[M]黄艳红,译.南京:南京大学出版社,2015(8).
- [33] 欧阳剑,彭松林,李臻.数字人文背景下图书馆人文数据组织与重构[J].图书情报工作,2019,63(11):15-24.
- [34] 高卫红.苏轼人生思想探析[J].郑州大学学报(哲学社会科学版),2009,42(4):50-51.
- [35] 谢建忠.论佛教哲学与苏轼的“人生如梦”思想[J].西南民族学院学报(哲学社会科学版),2000(6):54-58.
- [36] 高卫红.苏轼人生思想探析[J].郑州大学学报(哲学社会科学版),2009,42(4):50-51.
- [37] 袁瑾.民俗视野中的宋词创作[J].兰州学刊,2009(9):181-184.
- [38] 李剑鸣.大数据时代的世界史研究[J].史学月刊,2018(9):12-16.
- [39] 龙榆生.《唐宋名家词选》[EB/OL].[2020-01-24].<http://longyusheng.org/ci/sushi/index.html>.
- [40] SnowNLP: Simplified Chinese Text Processing [EB/OL].[2020-02-15].<https://github.com/isnowfy/snownlp>.
- [41] 王洁.大数据思维与数字人文的加值应用——传统文化数据库发展的新趋势[J].图书馆理论与实践,2018(5):104-108.
- [42] 邵炳军.两周之际三次“二王并立”史实索隐——周“二王并立”时期诗歌创作历史文化背景研究之一[J].社会科学战线,2001(2):134-140.
- [43] 韩国河,惠夕平,闫广宇,王书成,赵艺博,刘伟伟,樊明亮.河南荥阳官庄城址周边系统调查与初步收获[J].中原文物,2015(4):4-9+63.
- [44] 杨文.数字人文视阈下的社会记忆构建研究[J].情报资料工作,2019,40(5):38-45.
- [45] 董慧,陈文樵,罗忆,姜赢,王菲,周义刚,余思伟.数字档案本体分子论及其应用研究[J].中国档案,2010(3):57-60.
- [46] 梁灿兴.可获得性论的文献及相关概念[J].图书馆,2002(1):9-15.