

星汉窃渺——书信网络中蕴藏的人际关系挖掘*

李惠^{1,2} 侯君明³ 陈涛¹ 朱庆华² 刘炜¹

(1 上海图书馆 2 南京大学信息管理学院 3 中华书局·古联数字)

摘要 中国古代名人的书信,内容广博,记录了历史人物对社会世态的观察体验、对为官处世的经验心得,以及内心世界的描述披露。研究古代名人书信,可以帮助后人加深对历史事件的理解,补充史料细节,更加全面地认识历史人物的复杂性和多面性。由于语料规模和获取渠道的限制,目前名人书信研究多局限于个人所写所寄的书信。本文基于已提出的古代书信网络模型,设计相关网络度量方法,构建书信关联网络模型,旨在探索个人书信集里蕴含的群体人际关联。本文将此模型和度量方法运用于小规模数据集上开展试验,并加以史料佐证,希望能够为古代文本的知识发现提供新的思路和参考。

关键词 书信网络 自我中心 主题关联度 书信关联网络

DOI: 10.13663/j.cnki.lj.2020.05.011

Latent Relations in Ego-centric Correspondence Network

Li Hui^{1,2}, Hou Junming³, Chen Tao^{1,2}, Zhu Qinghua², Liu Wei¹

(1 Shanghai Library; 2 School of Information Management, Nanjing University; 3 Zhonghua Book Company · Gulian Media Tech Co., Ltd.)

Abstract Historic correspondence, in the form of letters, contains significant information of historic figures such as their notes on the society, fluctuations in career, and life experiences. These letters can help us expand knowledge beyond recorded historic events, supplement the details of historic literature, and understand the complexity and versatility of these historic figures more comprehensively. However, due to the limited scale of and access to historic correspondence, most correspondence research is stuck on the level of individual letters. Different from previous researches, we, on the basis of our historical correspondence network model, introduce two indicators, namely, topic relevance and topic similarity, to explore the interpersonal/community relationships embedded in the individual ego-centric correspondence network. We conduct experiments on the dataset of an individual letter collection, and correlate the corresponding experimental results with certain historic events. This can be a valuable reference and guide for researchers doing similar projects in future.

Keywords Correspondence network, Ego-centric, Topic relevance, Correspondence relatedness network

0 引言

近年来,开展了很多古文资源的数字化工程的探索,如上海图书馆的中国家谱知识服务平台^[1]和全国报刊索引^[2],一方面是对古文的加工和再现^[3],是古籍整理中必不可少的组成部分;另一方面是批量古文信息处理的基础和前提,推进实现针对古文献的数据挖掘^[4]。来自语言文字学、计算机科学、图书情报学等诸

多学科的学者纷纷投身于古文信息处理领域,开展基于不同语言单位(如字、词、句、段落、篇章等)的多种研究,如命名实体标注、事件

* 本文系上海市超级博士后“数字人文中名人档案知识网络的深度建模和多维分析”和国家社会科学基金项目“数字人文中图像文本资源的语义化建设与开放图谱构建研究”(项目编号:19BTQ024)的研究成果之一。

通信作者:李惠, E-mail: lh9743@126.com

抽取、网络分析等。

在数字化的古文资源中,古代名人的书信文献也是不可或缺的一部分。数字化的古代书信,主要由元数据和书信正文两部分组成。书信的元数据一般包括收信人和寄信人姓名、写信日期、写信地址以及寄信地址5个元素。古代名人书信主要可分为3类^[5]:一是亲人之间的家书,较为真实地记录了人物的成长轨迹和精神世界;二是友人交流的私人书信,较为详细地记载了人物的社会往来和兴趣爱好等信息;三是君臣、同僚往来的公文书信,较为审慎地描述了人物的政治抱负、价值取向和思想动向。

古代名人书信有两个主要特征:时效性和单一性。名人书信多具有鲜明的时代特色,与特定历史事件息息相关。而书信的收录者在整理的过程中,往往聚焦于名人个体单方向传递的书信^[6]。换言之,由于年代、环境、家庭等因素,档案收集的多为名人个人所写所寄的存稿,而未收录他人所写的书信。因而国内外数字人文学者运用社会网络分析(social network analysis)的相关工具,以可视化的效果来呈现书信中的人物关系,多限制于单向的名人-通信人关系,形成的多为自我中心(ego-centric)的书信网络,对通信人之间的关系,即他人(alter)之间的关系却较少剖析。

名人的通信圈里虽有名垂至今的大人物,但更多的是鲜有记载的小知识分子、小宦宦。这些小人物,虽然“名不见经传”,但他们之间的人际网络,才是当时社会的人情写照,是更全面更客观的历史佐证。本文旨在抛砖引玉,突破书信数据源的限制,尝试从名人书信的自我中心网络出发,挖掘书信中隐含的群体人脉网络,开展特定历史时期知识分子人脉关系的综合探索,为数字人文研究提供新的线索和思路。

1 相关研究

根据书信体文献的语料库规模,关于书信网络的研究可分为个人书信研究和群体书信研究。个人书信网络,主要包含某一特定人物所写和所收的书信集,属于星形结构(star

structure)的自我中心网络。网络的中心就是该人物,但是其余通信人之间的书信未涉及。群体书信网络涵盖了同时代其他人物的书信往来,属于社会中心网络(socio-centric network),亦可看作是无数张个人书信网络的集合。

国内数字人文领域,从计算分析的角度,对数字化的古代文献,开展文本分析或者社会网络挖掘的研究为数并不多,特定的书信研究更是稀少^[7]。国外的书信研究大多数聚焦于知名学者的个人书信网络^[8-10],分析特定学者的学术思想交流情况^[11]。目前也有呼吁^[12-14],旨在整合不同项目的数据资源,突破个人书信集的限制,探索群体社会网络的概貌。然而书信真正的保存情况,就不同的项目和书信作者的身份而言,存在着很大的差异。例如,上层社会的男性知识分子的书信保存条件,就比女性、异乡人、手艺人等的书信保存条件良好很多。再者,将分散在不同的图书馆、档案馆、私人收藏者或机构等的书信文献,整合并数字化,非一日之功。因而短期内想在大规模的书信语料库上开展探索和分析,还是比较困难的。

David Blei^[15]从统计建模的角度,将话题(topic,也被译作主题)定义为一组固定词汇的概率分布,即在相同语境下共现频繁的词群。他将主题模型(topic modeling)定义为一种能够从大规模文本语料中自动挖掘出隐含主题信息的概率模型。Latent Dirichlet Allocation(LDA)^[16]作为最经典最常用的主题模型之一,融合了贝叶斯统计和优化算法,将文本看作是一组主题的多项式分布,而主题是一组词汇的多项式分布。

话题模型作为大规模文本的话题导航工具,已经广泛应用于数字人文研究中,主要可分为共时和历时研究两类。共时研究关注某一特定历史时期的话题,如Meeks^[17]运用话题模型工具分析了50篇关于数字人文的文章,将结果可视化为本话题网络。通过话题网络,他们可以观察文本、话题、词语如何相互关联。相比共时研究,历时研究更侧重于话题的动态性,如Nelson^[18]通过观察话题随时间的变化,来分析美国南北战争期间Richmond市的

政治社交环境; Mimno^[19]在24本期刊中观察话题随着时间的变化,以及期刊之间的差异性随时间的改变。

Keli Du^[20]调查了数字人文大会(Digital Humanities,简称DH)有关话题模型应用的53篇文章,发现这些文章多侧重于话题概览和可视化效果,但对于模型应用的技术细节,如语料库的预处理,如何选取主题数量,如何去衡量主题的质量等并没有涉及,因而相关研究的可靠性和可复制性存在争议。考虑到这些问题,本文的研究并不侧重具体话题的内容,而是基于话题分布来计算通信者之间的相似性,从而挖掘古代书信个人网络中隐藏的人物关系和群体关系。

2 书信关联网络

一封书信可以表示为一个六元组集合 $l = (S, R, t, l_s, l_r, c)$ 。 S 表示寄信人集合, R 表示收件人集合。 $t \in T$ 指代写信日期。 $l_s \in L$ 和 $l_r \in L$ 分别指代寄信和收信地点。每封信的内容 $c \in C$ 可以表示成一组词序列 $\{w_1, w_2, \dots, w_n | w \in W\}$,其中 w_n 指代文本中的 n 个词。

信中的一个话题 θ ,定义为组词汇的概率分布 $\{p(w|\theta)\}_{w \in W}$,且 $\sum_{w \in W} p(w|\theta) = 1$ 。我们假定所有书信文本中有 k 个主题,且所有主题的集合为 Z 。那么一封信 c 中的话题分布可以定义为 $\{p(\theta|c)\}_{\theta \in Z}$ 。我们用 $n_{w,c}$ 指代信 c 中出现词 w 的频次,而 c 中 w 出现在topic θ 下的频次为 $n_{w,c,\theta}$, c 中topic θ 下所有词的数量为 $n_{c,\theta} := \sum_w n_{w,c,\theta} = \sum_w n_{w,c} p(\theta|c)$ 。

本文将书信网络模型,定义为重边的超图 $H = (V, E)$ ^[21],节点集合 V 代表收发信人,边集合 E 代表人物间的通信往来。每条边 $e = \langle H_e, T_e, i \rangle$ 包含一组节点, i 指代每封信的索引, $H_e \subseteq V$ 指代有向边的边头,即每封信的收信人, $T_e \subseteq V$ 指代边尾,即寄信人,且 $H_e \cap T_e = \emptyset$ 。

每条边都赋予了相应的属性 $(d, t, l_s, l_r, c, \{\theta_1, \theta_2, \dots, \theta_k\}) | d \in \mathcal{N}, t \in T, l_s \in L, l_r \in L, aw \in V, c \in C, \theta \in Z$ 。为了区分相同的通信人之间可能存在的多封书信往来,我们用字母 d 来表示每封信的 id 。考虑到一封信可能有多

个寄信人,但只有一个是写信作者。我们定义函数 $\text{top}: E \rightarrow \{\theta_1, \theta_2, \dots, \theta_k\}$ 来标注超图中的每条边每封信的话题分布。

尽管获得语料的渠道有限制,书信网络多为个人规模,我们仍想探索网络中潜在的群体人脉关系,换言之,如果两位通信人之间的书信,话题相似度达到一定阈值,那么这两人在网络中被看作是关联的。于是本文基于书信网络模型,计算话题关联度和相似度,构建书信关联网络,意在通过个人网络的书信话题关联,推测群体的社交网络,“窥一斑而知全豹”。

话题关联度。为了调查通信人相关的书信里涉及的各种话题,本文设计了话题关联度。给定书信网络 H ,节点 v (任意通信者)的话题关联度定义为他参与的所有书信中的话题分布集合 $\{p(\theta|v)\}_{\theta \in Z}$ 。我们用 $\tau(v)$ 指代所有与节点 v 相连的边集合。于是节点 v 和话题 θ 的关联度,可以计算为通信人所有书信中话题 θ 下的词语数量占总词数的比例:

$$\begin{aligned} p(\theta|v) &:= \frac{\sum_{e \in \tau(v)} n_{\theta,e}}{\sum_{e \in \tau(v)} \sum_{w \in c = \text{con}(e)} n_{w,c}} \\ &= \frac{\sum_{e \in \tau(v)} \sum_{c = \text{con}(e)} n_{c,\theta}}{\sum_{e \in \tau(v)} \sum_{w \in c = \text{con}(e)} n_{w,c}}, \end{aligned}$$

$n_{\theta,e}$ 指代附加到边 e 上的topic θ 下的所有词语,而 $\sum_{w \in c = \text{con}(e)} n_{w,c}$ 指代文本 c 中的所有词语。

话题相似度。为了比较通信人在书信话题上的异同,本文采用Kullback-Leibler Divergence距离(KL)^[22]和Jensen-Shannon距离(JS)^[23]计算话题关联度之间的相似性。 KL 距离衡量给定变量的两种概率 p 和 q 分布之间的差异情况,计算方法如下所示:

$$KL(p, q) := \sum_{x \in X} p(x) \ln \frac{p(x)}{q(x)},$$

但是 KL 距离不满足对称性,于是对称的Jensen-Shannon(JS)距离作为 KL 距离的改进,解决了非对称的问题,计算方法如下所示:

$$JS(p, q) := \frac{1}{2} \left[KL\left(p, \frac{p+q}{2}\right) + KL\left(q, \frac{p+q}{2}\right) \right].$$

书信关联网络。基于上述的两种计算方法，本文设计了如下的邻接矩阵 $M(V \times V)$ 。

$$M_{ij} = \begin{cases} 1 & \exists v_i, v_j \in V, JS(p(\theta|v_i), p(\theta|v_j)) \geq \eta \\ 0 & \text{otherwise} \end{cases}$$

v_i 和 v_j 对应书信网络中的任意两个节点，如果这两个节点的话题相似度达到一定阈值，那么项 $M_{ij} = 1$ ，反之为 0。当项 M_{ij} 值为 1，就在书信关联网络中 v_i 和 v_j 之间添加一条边。基于邻接矩阵 M ，本文将书信关联网络定义为无向图，如下： $G_h = (V_h, E_h)$ ，节点 $V_h \subseteq V$ 指代通信人，边

$E_h \subseteq V_h \times V_h$ 指代通信人之间的话题关联关系。

3 实验分析

本文选取岳麓书社 2011 年出版的《曾国藩全集》^[24] 中曾国藩的个人书信集（家书除外）开展实验。《曾国藩全集》共 31 册，其中书信就有 12 册，可见书信在曾氏传世文字中的分量。曾国藩书信的内容丰富，反映了名人性格的复杂性和多面性；且收信人范围广泛，为我们认识晚清政局提供了重要的史料和参考^[25]，对清末历史研究大有裨益。

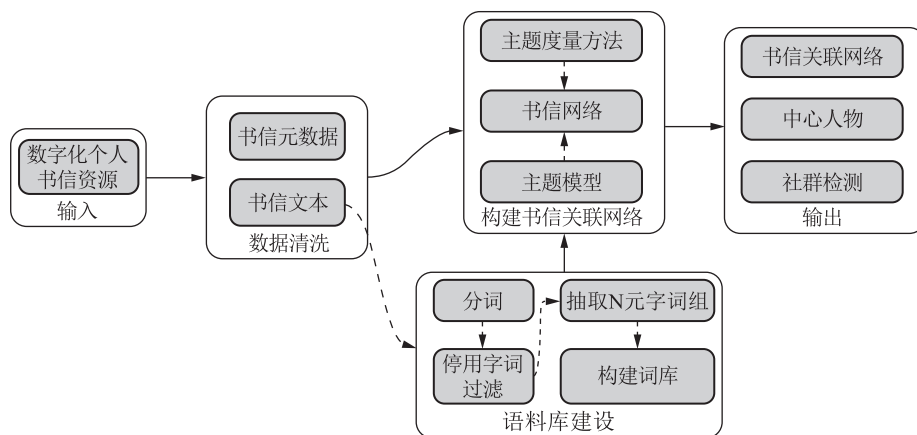


图 1 构建书信关联网络的实验流程

本次实验共选取 6 146 封书信，涉及 165 位通信者（曾国藩除外），其中每位通信者至少是 10 封曾国藩书信的收件人，保证实验数据的充分性，实验的具体流程，如图 1 所示。虽然现代汉语分词已取得较大进展，但是古籍文本分词由于受到古代汉语词汇特征、语义、语法等限制，始终没有形成一种行之有效的方法^[26]。本文不拘泥于分词限制，采用 N 元语法(n-gram)^[27] 结合中文分词工具 jiebaR^[28]、自建古汉语词典和停用词词典对书信开展一到二元字词组抽取工作。在此基础上，通过 TF-IDF (Term Frequency-Inverse Document Frequency)^[29] 算法再次过滤文本中常见而意义不大的字词组，预处理之后的一到二元字词组共 2 287 个。

本文综合选取 3 种方法，即最大似然估计^[30]、话题之间的距离计算^[31] 和话题之间的连贯性^[32] 的方法，确定本次实验所有语料中的话题数为

表 1 曾国藩的个人书信网络和书信关联网络的基本属性

书信网络基本属性	个人书信网络	书信关联网络
节点数	1 004	165
边数	1 896	2 128
直径	2	3
信件数	14 231	6 146
中心人物	曾国藩	郑元璧

38 个，生成曾国藩书信的话题模型，计算不同收信人之间的话题关联度和相似度（本次实验的话题相似度阈值定为 0.45），构建书信关联网络。

书信关联网络。本次实验的书信关联网络包含节点 165 个，边 2 128 条，如表 1 所示。可以看出，书信关联网络，相比于曾国藩的个

人书信网络(网络的具体构建细节可参见文献[7])而言,虽然信件数明显减少,但边数和直径都有增加,而且中心人物由曾国藩变为郑元璧。郑元璧其人,系庚子进士,曾任陕西道御史,刚正不阿,清廉如水^[33];也曾是曾国藩幕府机构湖南东征筹饷局的主要负责人之一,与其他幕僚一起筹备粮饷装备,组建湘军^[34]。为官之外,他也很重视书院教育,曾在多所书院任教监管,并捐书捐银。他的家风优良,曾“一门两举三进士”。曾国藩在与他的书信中,涉及地方官事务、后勤保障、政治军事等话题。本文推断良好的家世、丰富的工作经历和推崇教育的治学理念,让他在诸多话题上和其他通信人(尤其是幕府同僚)有明显关联,也是符合历史史实的。

表2 书信关联网络和个人书信网络中,中心度计量结果排名前10的通信人

书信关联网络	个人书信网络
郑元璧	胡林翼
孔祥珂	彭玉麟
吴文镛	左宗棠
张澂山	李鸿章
吴安谷	乔松年
联福	杨岳斌
王鑫	吴坤修
董似穀	李续宜
俊达	官文
奕訢	李瀚章

隐藏的人物。本文采用中心度的网络度量方法,衡量书信关联网络中节点(人物)在网络中的重要性,具体计算为与该节点相接的边数,如表2所示。有意思的是,关联网络中排名前10的人物,与曾国藩个人书信网络的人物排名,截然不同。比起左宗棠、李鸿章等为大众熟知的晚清政治家,这些人物乍看有些陌生,但细究起来,不可小觑。孔祥珂是孔子嫡孙,家世显赫^[35],并曾参与抗击捻军;吴文镛不仅是曾国藩的恩师,而且是湘军将领胡林翼和江忠源的举荐者^[36];董似穀是亦官亦商的晚清首富盛宣怀的老丈人^[37],书法作品流传至

今;恭亲王奕訢是洋务运动的主要领导者^[38]。这些人物相比曾国藩个人书信网络中排名靠前的人物,毫不逊色,因而出现在书信关联网络的中心,也是适宜和恰当的。

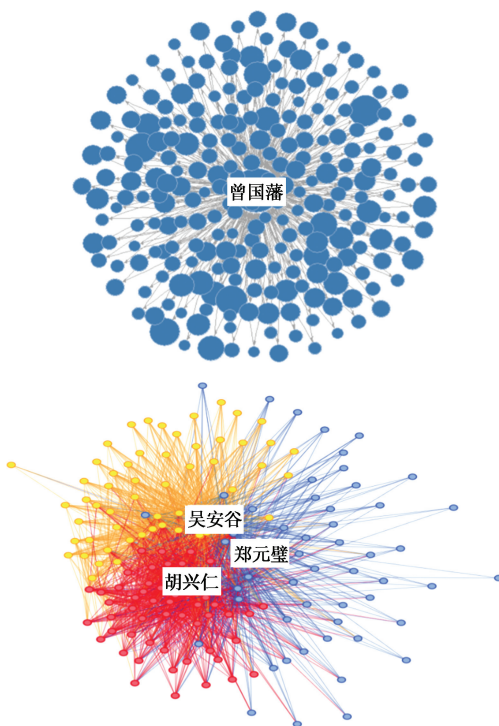


图2 曾国藩的个人书信网络和书信关联网络

群星璀璨的社群。为了发现网络中紧密连接的节点社群(“朋友圈”),本文采用社群发现的Louvain算法,探索书信关联网络中隐含的社群结构。Louvain算法基于模块度(Modularity)的相关概念,通过两步迭代,即模块度优化和社群聚合,最大化整体图的模块度^[39],该算法效率高,准确率也有保证^[40]。据此,本次实验的书信关联网络,最终可分为3个社群,如图2所示。相比上面图示曾国藩的个人网络而言,书信关联网络不再是“一家独大”的星形结构,而是“三足鼎立”的社群关系。这3个社群并不相互孤立,群内外皆有连接。每个社群中,中心度最高的通信人均在图2中用黑字标出,在书信关联网络中,我们用不同颜色代表不同的社群,并在图3中展示了3个社群各自的特色话题下的高频词汇。

最大的社群包含 56 人(图中蓝色指代, 郑元璧所在群), 该社群政治家居多, 如奕訢、毛鸿宾、丁宝桢等; 该社群的特色话题, 既包含如“金陵”“江南”“沪”“直隶”“畿辅”等地名术语, 也包含如“委员”“藏事”“吏治”“报销”“章程”等时政术语, 着重于各地地方官的行政事务、官员委任以及信息传递, 这与该社群中政治家居多的现象吻合。红色指代的胡兴仁所在群, 成员多为湘系人物, 如刘蓉、郭嵩焘、唐训方等人皆为湘军幕府核心成员。该社群的特色话题包含如“贼”“捻匪”“逆”等对起义军的蔑称, 和如“会剿”“夹击”“调度”等军事作战词汇, 侧重于对起义军的镇压, 与社群中湘军居多的情况吻合。最小的社群包含 54 人(黄色指代, 吴安谷所在群), 该社群的特色话题, 不仅涉及民生, 包含如“秋禾”“播种”“二麦”“抗旱”等农耕术语和“春夏”“秋

节”“甘霖”等季节和气候术语, 也涉及军事, 包含如“贼”“逆”等对起义军的蔑称, 与吴安谷江西按察使和督粮道的身份相符, 也和群里的众多人物, 如钟谦钧、左宗棠、刘铭传等的生平事迹是关联的。同时, 我们注意到该社群中许多成员不仅在朝为官, 也是当时酷爱诗书的文人雅士, 如吴嘉宾、张曜孙、耆龄等, 虽然吴安谷相关的史料并不多, 我们基于社群, 推测他可能也对文学有浓厚兴趣, 于是查阅相关文献, 发现他确实曾经编过一本吴煊的《读史笔记》^[41], 因而我们的推测是合理的。

社群中像吴安谷这样史料相对较少的人物还有很多, 但通过我们构建的书信关联网络, 在曾国藩个人书信的基础上, 探索了这些人物之间隐藏的话题关联关系, 并得到史实的佐证, 说明了该方法在相关历史研究中的合理性、有效性和可应用性。

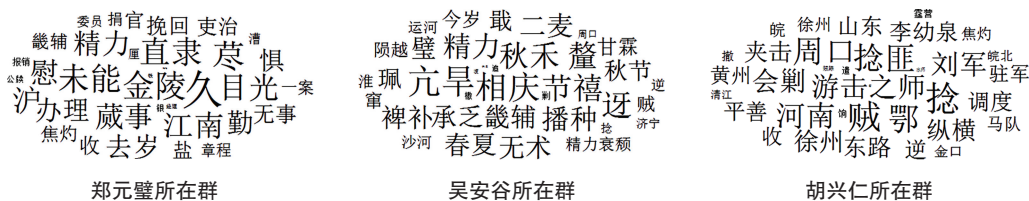


图3 书信关联网络中3个特色话题词群

4 结论

中国古代名人书信, 记录了许多历史人物的传奇人生和转瞬即逝的历史时刻^[42]。书信内容丰富, 涉及人生哲理、政治氛围、社会环境、工作生活等诸多方面, 既具有史料、文学、文物等多方面的研究价值^[43], 又有正式出版物中难觅的名人内心世界的写照。本文基于名人个人书信网络, 结合统计模型, 计算人物间的话题关联度和相似度, 构建书信关联网络, 探索自我书信网络中蕴含的群体关系, 并在小规模数据集上开展实验, 形象地展示了特定历史时期的知识分子人脉关联网络。本文的主要贡献列举如下:

(1) 相比于先前书信研究多从个人书信网络的角度开展, 本文突破语料规模的限制, 在个人书信中挖掘隐含的群体关联网络, 并用史

料证明该方法的有效性和可应用性。该方法不仅可以帮助学者挖掘书信中的隐含关系, 而且可以对史料不足的人物考证提供线索和参考。

(2) 考虑到古代汉语分词的困难和公开数据的不足, 本文实验中建设的晚清书信词库和相关停用词表, 会在整理之后免费公开, 供研究者使用, 为古代汉语的信息处理提供数据支持。

本文的研究作为更大范围研究的一个预设, 在下一步的工作中, 我们将整合更多资源和更多数据, 如著作关系(批校题跋)、缙绅录、宗族谱系等, 开展晚清政坛社会关系的综合探索, 为数字人文提供新的研究视角。

(本文数据链接地址: <http://hdl.handle.net/20.500.12304/10375>)

参考文献

- [1] 中国国家谱知识服务平台[EB/OL]. [2019-09-02]. <http://jiapu.library.sh.cn/#/>.
- [2] 全国报刊索引[EB/OL]. [2019-09-02]. <http://www.cnbkys.com/>.
- [3] 毛建军. 古籍数字化概念的形成过程探析[J]. 科技情报开发与经济, 2006, 16 (22): 160-162.
- [4] 黄水清, 毛东波. 古文信息处理研究的现状及趋势[J]. 图书情报工作, 2017, 61 (12): 43-49.
- [5] 张鹏立. 浅谈中国古代书信[J]. 河北省社会主义学院学报, 2014 (2).
- [6] 吕瑞花, 韩晶晶, 韩露. 基于元数据的科技名人档案编目[J]. 科技导报, 2013, 31 (14): 64-69.
- [7] 李惠, 侯君明. 古代书信体文献的社交网络模型[J]. 南京师范大学文学院学报, 2018(3): 164-172.
- [8] Darwin Correspondence Project[EB/OL]. [2019-09-02]. <http://www.darwinproject.ac.uk/>.
- [9] Thomas Bodley [EB/OL]. [2019-09-02]. <http://www.livesandletters.ac.uk/projects/diplomatic-correspondence-thomas-bodley-1585-1597>.
- [10] The Newton Project[EB/OL]. [2019-09-02]. <http://www.newtonproject.ox.ac.uk>.
- [11] Lightman B. Correspondence Networks [M]. A Companion to the History of Science. John Wiley & Sons, Hoboken. 2016.
- [12] Circulation of Knowledge and Learned Practices in the 17th-century Dutch Republic[EB/OL]. [2019-09-02]. <http://ckcc.huysgens.knaw.nl/>.
- [13] Early Modern Letters Online[EB/OL]. [2019-09-02]. <http://www.livesandletters.ac.uk>.
- [14] Mapping the Republic of Letters.[2019-09-02]. <http://republicofletters.stanford.edu/>.
- [15] Blei D M, Ng A Y, Jordan M I. Latent Dirichlet Allocation[J]. Journal of Machine Learning Research, 2003, 3 (1): 993-1022.
- [16] Blei D M. Probabilistic Topic Models[J]. Communications of the ACM, 2012, 55 (4): 77-84.
- [17] Comprehending the Digital Humanities[EB/OL]. [2019-09-02]. <https://dhs.stanford.edu/comprehending-the-digital-humanities/>.
- [18] Nelson R K. Mining the Dispatch[EB/OL]. [2019-09-02]. <https://dsl.richmond.edu/dispatch/>.
- [19] Mimno D. Computational Historiography: Data Mining in a Century of Classics Journals[J]. Journal on Computing and Cultural Heritage, 2012, 5 (1): 1-19.
- [20] Keli Du. A Survey on LDA Topic Modeling in Digital Humanities[EB/OL]. [2019-09-02]. <https://dev.clariah.nl/files/dh2019/boa/0326.html>.
- [21] Li H. Social Network Extraction and Exploration of Historic Correspondences [D]. Heidelberg: Heidelberg University, 2018.
- [22] Kullback S, Leibler R A. On Information and Sufficiency[J]. Annals of Mathematical Statistics, 1951(22): 79-86.
- [23] Lin J. Divergence Measures based on the Shannon Entropy[J]. IEEE Transactions on Information Theory, 1991, 33 (1): 145-151.
- [24] 曾国藩. 曾国藩全集[M]. 长沙: 岳麓书社, 2011.
- [25] 唐浩明. 唐浩明评点曾国藩书信[M]. 长沙: 岳麓书社, 2011: 1-6.
- [26] 李筱瑜. 基于新词发现与词典信息的古籍文本分词研究[J]. 软件导刊, 2019, 18 (4): 60-63.
- [27] Lin Y, Michel J B, Aiden E L, et al. Syntactic Annotations for the Google Books Ngram Corpus[C]. Proceedings of the Association for Computational Linguistics (System Demonstrations), 2012: 169-174.
- [28] Chinese Text Segmentation with R[EB/OL]. [2019-09-02]. <https://github.com/qinwf/jiebaR>.
- [29] Ramos J. Using TF-IDF to Determine Word Relevance in Document Queries[C]. Proceedings of the First Instructional Conference on Machine Learning, 2003, 242: 133-142.
- [30] Griffiths T L, Steyvers M. Finding Scientific Topics[J]. Proceedings of the National Academy of Sciences 101 (Suppl_1) 2004: 5228-5235.
- [31] Juan C, Tian X, Tao L J, et al. Density-based Method for Adaptive LDA Model Selection[J]. Journal Neurocomputing. 2009, 72 (7-9): 1775-1781.
- [32] Röder M, Both A, Hinneburg A. Exploring the Space of Topic Coherence Measures[C]. Proceedings of the Eighth ACM International Conference on Web Search and Data Mining. 2015: 399-408.
- [33] 陈松溪. 关于郑振铎的家世[J]. 固原师专学报 (社会科学), 1997 (2): 1-6.
- [34] 龙方成. 试论曾国藩的后勤保障思想[J]. 湖湘论坛, 1994 (6): 54-57.
- [35] 孙经超. 清代衍圣公行政职权研究[D]. 曲阜: 曲阜师范大学, 2018.
- [36] 王丽娇. 晚清疆臣吴文镛[D]. 长春: 东北师范大学, 2014.
- [37] 全群旺. 青果巷与“中国商父”盛宣怀[J]. 档案与建设, 2017: 40-41.
- [38] 董明月. 洋务运动对中国近代化的影响[D]. 长春: 吉林大学, 2009.
- [39] Blondel V D, Guillaume J L, Lambiotte R, et al. Fast Unfolding of Communities in Large Networks[J]. Journal of Statistical Mechanics: Theory and Experiment, 2008 (10): 10008.
- [40] Albert-László Barabási and Márton Pósfai. Network Science[M]. Cambridge: Cambridge University Press, 2016: 355-375.
- [41] 吴焯; 吴安谷. 读史笔记[EB/OL]. [2019-12-20]. <https://chinese-cat.lib.cam.ac.uk/mulu/fb85.html>.

(下转第80页)

- [12] 武汉大学图书馆.关于开展首届武汉大学“书香大使”的通知[EB/OL]. [2019-12-15]. <http://www.lib.whu.edu.cn/news/view.asp?id=8143>.
- [13] 徐文静. 高校图书馆走进中小学生课堂——普林斯顿大学图书馆阅读推广案例分享[J]. 图书馆建设, 2015(4): 54-57.
- [14] 李芙蓉, 周亚. 美国国家层面的图书馆未成年人服务政策与法规探析[J]. 图书馆, 2018(11): 1-8; 16.
- [15] 王平, 毕胜男. 国外图书馆服务中面向未成年人的权利保护体系阐释与思考[J]. 图书馆建设, 2016(9): 29-34; 45.
- [16] 刘菡, 杨乃一, 李思雨, 等. 未成年人阅读、学习与赋能国际研讨会综述[J]. 图书馆建设, 2019(3): 65-73.
- [17] 范并思. 服务活动化: 图书馆服务新趋势[J]. 图书馆学刊, 2017, 39(12): 1-4.
- [18] 许欢, 尚闻一. 美国、欧洲、日本、中国数字素养培养模式发展述评[J]. 图书情报工作, 2017, 61(16): 98-106.
- [19] 吕建强, 许艳丽. 数字素养全球框架研究及其启示[J/OL]. 图书馆建设: 1-11[2019-12-15]. <http://kns.cnki.net/kcms/detail/23.1331.G2.20190807.1123.008.html>.
- [20] 方诗雅, 范并思. 图书馆未成年人服务理念的新发展——《国际图联0-18岁儿童图书馆服务指南》的启示[J]. 图书馆论坛, 2019, 39(9): 93-99.
- [21] 中国图书馆学会. 关于中国图书馆学会等单位共同发起《中国公民信息素养教育提升行动倡议》[EB/OL]. [2019-09-20]. <http://www.lsc.org.cn/contents/1342/14004.html>.
- [22] 洪跃, 付瑶, 杜辉, 等. 国内高校图书馆信息素养教育现状调研分析[J]. 大学图书馆学报, 2016, 34(6): 90-99.
- [23] 赵新华. 全民阅读背景下的图书馆管理与建设探究[J]. 教育现代化, 2019, 6(64): 195-196.
- [24] 张巧娜. 国内高校图书馆阅读推广的经验与启示——基于两届全国高校图书馆阅读推广大赛的比较分析[J]. 新世纪图书馆, 2019(9): 38-42.
- [25] 肖鹏. 重新认识图书馆营销: 概念拓宽与战略升级[J]. 高校图书馆工作, 2014, 34(6): 3-7.

邵剑彬 厦门大学图书馆, 区域研究资料中心主任, 馆员。研究方向: 图书馆管理、读者服务。作者贡献: 论文初稿撰写。E-mail: bblg@xmu.edu.cn 福建厦门 361005

黄国凡 厦门大学图书馆, 副研究馆员。研究方向: 数字图书馆、图书馆营销、图书馆创新服务。作者贡献: 论文修改。福建厦门 361005

(收稿日期: 2019-10-28 修回日期: 2019-12-19)

(上接第92页)

- [42] 叶永烈. 历史的绝笔[M]. 成都: 四川人民出版社, 2016: 1-10.
- [43] 名人书信引发收藏热[EB/OL]. [2019-09-02]. <http://collection.sina.com.cn/gjsb/20080625/14135022288.shtml?from=wap>.

李惠 女, 南京大学信息管理学院, 上海图书馆(上海科学技术情报研究所), 博士后。研究方向: 文本挖掘、社交网络、数字人文。作者贡献: 建模及论文撰写。E-mail: huili@libnet.sh.cn 上海 200031

侯君明 中华书局·古联数字传媒科技有限公司, 数

据库编辑。作者贡献: 数据支持。北京 100073

陈涛 上海图书馆(上海科学技术情报研究所), 博士。研究方向: 关联数据、数字人文、知识图谱。作者贡献: 模型指导与修改。上海 200031

朱庆华 南京大学信息管理学院, 教授。作者贡献: 论文指导与修改。江苏南京 210023

刘炜 上海图书馆(上海科学技术情报研究所), 研究员。作者贡献: 论文指导与修改。研究方向: 数字图书馆、知识组织。上海 200031

(收稿日期: 2019-10-24 修回日期: 2020-01-07)