数字人文视域下的古文献文本标注与可视化研究*

——以《左传》知识库为例

□李斌　王璐　陈小荷　王东波

摘要　在数字人文研究范式下，传统的以电子化和全文检索为基础的古籍研究模式已难以满足历史学、文献学、语言学等学科深度研究的需要。古籍文本特别是史书所记载的词语、时间、地点、人物、事件等要素都需要结构化的历史人文数据库，从而实现历史要素的定量分析与可视化。文章以古汉语自动分析技术为基础，结合人工标注和校对，以实体标注方法解决历史人物的同名异指和异名同指问题，对史学名著《左传》进行了词语切分、词性、时间、人物ID、地点GIS信息标注，进而实现了热点人物、人物关系网、人物游历轨迹与距离等量化统计与可视化，为古籍文本的内容标注、结构化人文知识库建设提供新的研究路径。最后，讨论了知识库进一步的完善方案与应用场景。

关键词　数字人文　《左传》　实体标注　数据库　古文信息处理
分类号　TP391.1　G250.7　H087
DOI　10.16603/j.issn1002－1027.2020.05.011

1　引言

我国古典文献历史悠久、卷帙浩繁，为研究了解各个时期的人物、地理、民情、社会状况等提供了宝贵翔实的第一手资料。然而，古籍善本往往存在重藏轻用、不易获取且文本内容晦涩难懂等问题。传统的古文献研究主要依靠资深学者的人工整理和博闻强识，但能够通读古籍且能融会贯通、应用于学术与社会生活的专家屈指可数，这些都严重影响了古籍文献内容的开发与利用。

本文以先秦历史典籍《左传》为研究对象，尝试了数字人文知识库的构建，在分词与词类标注的基础上，深入标注了人名、地名信息，使用可视化技术，建成了在线的查询网站，进而对词语、人物等要素的时空分布、人物游历距离等进行计量分析，力图为文史和语言的量化研究提供新的视角、数据标注方法与技术解决方案。《左传》乃研究先秦史不可或缺的典籍。无论建国初期对中国历史分期和社会性质的辩论，抑或近年来的先秦社会史的研究，《左传》均是研究者最基本的史料依据。许倬云早在60年代就尝试量化《左传》所记载人物的社会背景，从而探讨先秦和秦汉的社会流动状况[1]。随着计算技术的不断提高，使得当代研究者在进行同类研究时可以更全面、多角度地把握和分析史料，《左传》知识库的建立亦以此为目标。

2　研究现状

在古籍文本的数字人文数据库构建方面，国内外已经有了一些重要的研究工作。古籍文本的电子化已经取得了许多重大成果，如《文渊阁四库全书》《四部丛刊》《中国基本古籍库》《国学宝典》等全文检索数据库以及元数据加工方法[2]。但这些数据库还缺乏内容信息的深度标注，只能满足基于“字”的全文检索，无法挖掘和统计出更多信息。下面主要从分词词性、命名实体标注、地理信息标注三个方面，介绍较为深入的文本标注与数据库建设的必要性以及学界的研究进展。

* 国家社会科学基金重大项目“基于《汉学引得丛刊》的典籍知识库构建及人文计算研究”（编号：15ZDB127）、国家社会科学基金“中文抽象语义库的构建及自动分析研究”（编号：18BYY127）、江苏省高校优势学科建设工程资助项目的研究成果之一。
通讯作者：李斌，ORCID：0000-0002-7328-9947，邮箱：libin.njnu@gmail.com。

(1)分词与词性标注。古籍文本中没有词界，只有进行词语的切分，才能实现古籍文本基于“词”的检索。词性的信息，不仅包含名词、动词，还可以区分出人名、地名、时间等要素。古籍的分词与词性的标注，已经有了一些计算机自动分析的技术，形成了一些标注语料库。如台湾“中央研究院”的自动分析工具和古代汉语语料库、南京师范大学自动分析工具和先秦[3]及中古汉语标注语料库[4]。这些语料库主要服务于语言研究，产生了许多词汇方面的研究成果，同时也为古代典籍的数字人文知识库的建立奠定了良好的基础。

(2)命名实体信息。古籍文本中有大量的命名实体(人名、地名、官职、年号、器物、文书等)，对文史研究来说价值巨大，但在现有古籍电子文本上难以直接检索和获取，因此需要进行人工标注，从而进行后续的自动分析。哈佛大学费正清中国研究中心主持建设的“中国历代人物传记数据库”(China Biographical Database，CBDB)[5]，从唐代之后的大量古籍文本和工具书中，人工整理了36万条历史人物的姓氏、官职、族谱等传记资料，建设了可供动态查询的人物关系数据库。该数据库超乎传统的史料全文数据库，对史料就历史人物特性加以分析，便于进行传记学、社会网络层面的研究。英文数据方面也有一些自动处理的尝试，如埃尔森(Elson D K)等通过对60部19世纪英国小说和期刊进行社会网络挖掘，构建出人物社会网络关系[6]。

(3)在古代地理信息方面，也有着许多标注与分析工作，特别是将地理信息系统(Geographic Information System，GIS)技术用于历史地理研究，将历史上的地理要素转移到以现代地理坐标系统为参照系的电子地图上，有助于发掘隐藏在数据背后的潜在信息和传统研究难以发现的现象[7]。英国艺术与人文研究委员会的数据库“赫斯提亚”(Hestia)[8]，通过对古希腊历史学家希罗多德(Herodotus)的《历史》中提到的地点及其地理特征的描述进行信息抽取和可视化，提供了可以深入挖掘文本的可视化工具。英国历史地理信息系统(Great Britain Historical GIS，GBHGIS)[9]涵盖了19世纪以来英国行政边界的变迁，整合时空属性，提供了社会、经济和人口变迁的分析方法。美国国家历史地理信息系统(U.S. National Historical Geographic Information System，NHGIS)[10]是一个为历史人口研究而创建的历史地理信息系统，整合了美国从1790—2000年的人口统计数据。在国内，复旦大学历史地理研究中心的中国历史地理信息系统(China Historical GIS，CHGIS)[11]，建立了一套中国历史时期逐年连续变化、开放的基础地理信息数据库，提供了GIS基础数据平台。这些平台的构建，为古籍文本的地理信息标注作出了良好的探索。夏翠娟还介绍了历史地理数据在图书馆数字人文项目中的其他应用[12]。

古代文学研究领域也引入了命名实体标注和数据库技术，对部分诗词的作者、写作时间与地点、作者关系等信息进行标注，如“唐宋文学编年地图”[13]、“历代诗人地域分布”[14]。台湾法鼓山的佛典数据库也正在标注重要人物、地点信息[15]。基于这些数据库，已经得出了许多重要的统计数据成果。不过，这些数据库只标注了重要人物和地点，如果能够完整地、穷尽地标注历史文本中命名实体的详细信息，则能够得出更多有益的统计结果。而且，这种数据库一次建立，便可反复使用，避免了各单位重复开发，可减少人力物力的浪费。

中文古籍数字化建设工作在经过多年发展后，累积了大量古籍文本数字化资源的成果。国际上数字人文研究已经大量使用了多种计算分析技术，而国内古籍数字化研究由于人才培养不足、相关标准规范缺乏等原因，综合运用多种技术手段进行古籍数字化深度开发的实践尚为少见。古籍数字化是一项艰巨而复杂的工程，联合新兴的数字人文研究成果，才能促进古籍数字化在深度开发上有所作为。本文旨在基于《左传》文本，探索历史文献的深度加工与分析方法，使用文本信息标注、GIS等技术建立知识库，开发检索平台，实现古籍文献的深度加工和可视化检索，服务于文史研究的多种计量分析需要。

3 《左传》历史人文数据库的建设

《左传》作为第一部编年体通史，文史研究价值极高。《左传》记载了东周前期自鲁隐公元年(公元前722年)起，迄于鲁哀公二十七年(公元前468年)的254年间各国政治、经济、军事、外交、文化方面的重要事件和重要人物。《左传》现有的全文检索、人物、典故等词典及研究论著丰富，但碎片化严重，往往需要学者多年的悉心积累。如果能将已有的研究成果整合为知识库，完成词语切分和时间、人物、地

名的实体标注，开发可视化检索平台，可以为历史人文研究提供知识储备与服务。

3.1 数据来源

原始数据来自《左传》的白文部分，全文 18 万字，使用南京师范大学制定的古汉语分词与词性标注规范和自动分析工具，人工校对了分词和词性信息的先秦《左传》语料库[16]。根据《十三经注疏》《中国历史地图集》和《春秋左传词典》等工具书，给正文中的每个人名、地名都指定了唯一的 ID 编号，填写了每个命名实体的信息。

3.2 内容标注与数据库架构

在《左传》语料库的基础上进行了二次标注。首先，对地名标注了每个地点的今地名和 GIS 坐标；其次，根据《左传》自身的纪年体例，将书中的历史纪年对应到公元纪年，并标明每个段落的历史年份；最后，给人物表中的每个人物增设了人物主名。由此，形成了《左传》数字人文数据库的构建，完成了词类标注基础上的历史时间、地点、人物信息的全面标注，可以为语言学、历史人文方面提供重要的数据资源。《左传》数据库包括人名表、人物表、时间人物表、人物同现表、地名表、人物地名表共计六张数据表，具体字段与结构如图 1 所示。其中标注“*”的字段为《左传》语料库已经标注的数据[16]，其余字段为本文所补充的数据。

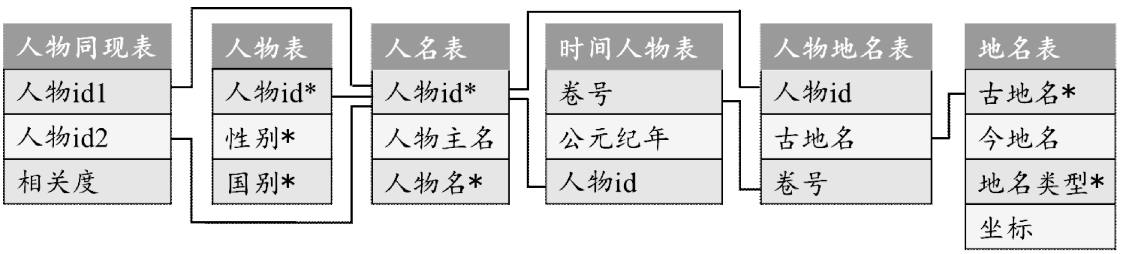

图 1 《左传》数据库的结构

3.3 人物信息标注

《左传》中一人之名号，往往有多种，需根据各种注疏文献和工具书进行辨识，使用唯一的 ID 号来标定。据笔者统计《左传》共有 2406 个人物，平均每个人物有 1.78 个名字，所以使用 ID 编号是十分有必要的。人物的不同名称的分布情况如图 2 所示，其中有四成的人物不止一个名称，37%的人物有 2—5 个名称，3%的人物有 6—10 个名称。名称最多的人物是晋国大夫范武子，《左传》记载他的名称有 10 个之多，可见异名同指现象在标注 ID 之后才易于统计分析。

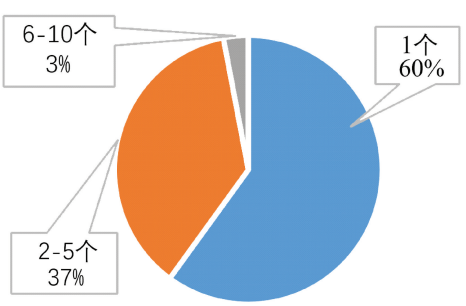

图 2 《左传》人物的不同名称的数量分布

《左传》中同名异指现象也不少见。方朝晖指出“《左传》姓名相同之名称，或因地而异人，或因时而异人，或时同地同名同，而人仍有异者”[17]。同名异指是指字形完全相同的人名在具体的上下文中代表两个或者两个以上的人物。笔者统计了《左传》人名对应人物数分布情况，其中 90%的人名都只对应一个人物，余下 10%的人名存在重名，至多一个人名为 16 个人物共用。

一个人物往往具有多种名字和称呼，在后续的计量分析与可视化中会带来许多不便，需要选择一个作为人物主要的称呼。因此，我们参考《春秋左传词典》[18]、《春秋左传人物谱》[17]等文献资料，人工标注了每个人物最具代表性的、为学界所熟知的“人物主名”。不过，“人物主名”不一定出自“人物名”，举例如表 1。秦穆公在《左传》中的名字有“穆、穆公、秦伯、秦伯任好、秦君、秦穆”，人工给出的人物主名“秦穆公”并不在其中。

表 1 人名表示例

人物 ID	人物主名	人物名	性别	国别
1302	秦穆公	穆 穆公 秦伯 秦伯任好 秦君 秦穆	男	秦

在此基础上，建立了时间人物表、人物同现表。《左传》为编年体史书，按年号分卷，由“卷号”即可获得年份时间，进而确定“公元纪年”。“人物 ID”在某卷出现，即视作该人物在该年份活动。

3.4 地名信息标注

地名数据处理包括标注地名表的“今地名”“坐标”两个字段。参考《中国历史地图集》[19]、CHGIS[11]等资料工具，人工标注了《左传》文本中1066个古地名对应的今地名。利用百度地图API[20]，解析今地名地址获得地名对应坐标经纬度数据，然后进行人工校核。与此同时，还标注了地名的类型，如诸侯国名、地名、河流、山脉等，举例见表2。

表2 地名表示例

地名ID	地名	类型	今天所在地	百度GIS坐标
21	莒	诸侯国名	山东省日照市莒县	118.848454 35.587102

4 《左传》检索系统与时空地图平台

4.1 基于实体ID的检索框架

使用Web开发技术，建设了《左传》检索系统与时空地图平台[21]，已上线提供检索服务。其结构功能如图3所示。平台除了全文检索功能外，基于底层的数字人文知识库，还提供人物、地名、人物关系等多种查询方式。这些实体的查询主要依赖于数据库中的实体ID。例如，人物查询需依据“人物ID”来进行，获得“人物ID”的操作流程如图4所示。

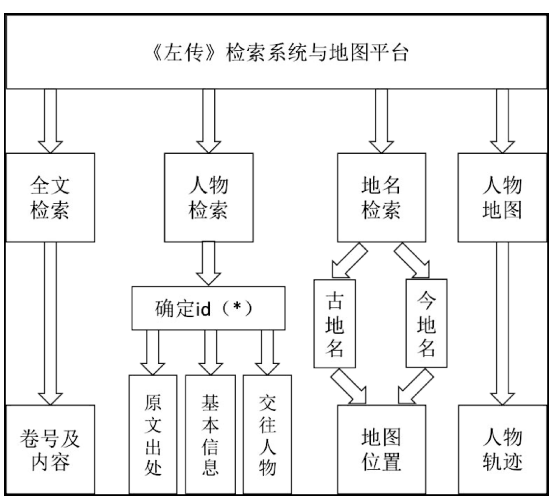

图3 《左传》检索系统结构

4.2 人物检索

人物检索页面，供用户查询人物的基本信息、上下文信息和人物关系。如图5所示，以“秦穆公”为例，输入“穆公”，得到所有名字包含“穆公”的人物列表，用户根据事先人工标注的“人物主名”以及“性别”“国别”信息，得以确定其ID为1302。在人物检索页面，可查得秦穆公以各种称呼出现的33个上下文，以及与他交往的人物28位，这些人物会按出现频率的百分比由高到低排列。相比传统的利用人物别名直接扩展查询来说，这样可以更准确地查找到人物出现的上下文。而相关人物的列表，也更直观地展示出人物的社交情况。

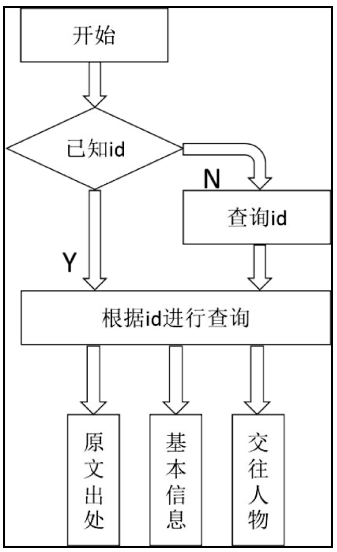

图4 人物查询流程

4.3 地名检索

地名检索页面利用百度地图API技术开发了《左传》古今地名的查询功能，能够满足用户的两种查询需求，实时在地图上绘制地点的红色标志。图6(左)为古地名“鲁”的查询结果，图6(右)为今地名“山东”查询结果，可以看出，“齐鲁大地”山东省的覆盖面积要大于鲁国。

4.4 人物地图——人地同现轨迹图

人物的历时轨迹，也是文史研究的重点之一。历史人物的游历路径，在传统研究中往往需要考证大量的史料，多用文字描述人物所到之处，不是很直观。我们采用了近似计算和可视化方法，根据人物和地点在文本中的同现信息(在一个句子中同时出现)、文本的公元纪年和地名信息，可以生成人物的历时轨迹图。以“楚昭王”为例，图7直观显示了楚昭王一生行经10地的活动轨迹，每个坐标点标注说明了其游历当地具体的时间和地名，让文本内容可视化地呈现在用户面前。

平台还支持查询多个人物在电子地图上的轨迹，可用于查看人物轨迹是否重合。以晋惠公、晋文

图 5　人物检索示例

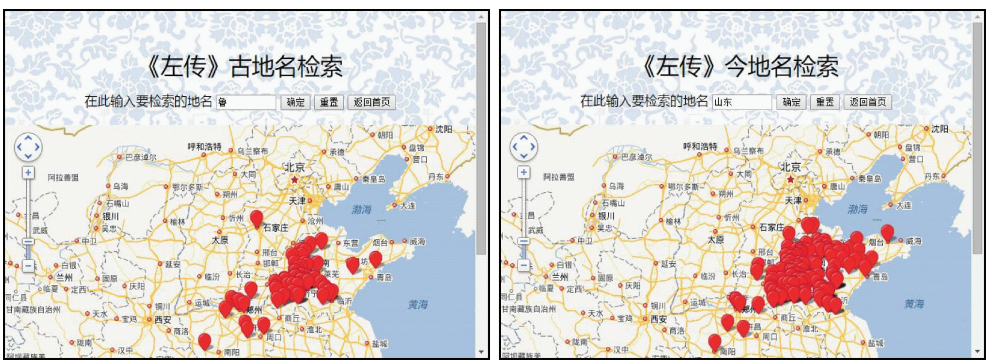

图 6　地名检索示例(“鲁国”和“山东”包含的古地点)

图 7　人物地图检索示例(“楚昭王”)

公为例，历史上曾身为晋公子的夷吾（惠公）、重耳（文公）遭陷逃难，后几经磨难返回故土重新称王，从地图检索的结果（图 8）可以窥见一二。通过多条人物轨迹是否重合以及具体重合的时间点，能够推算出历史人物是否曾经相会过。

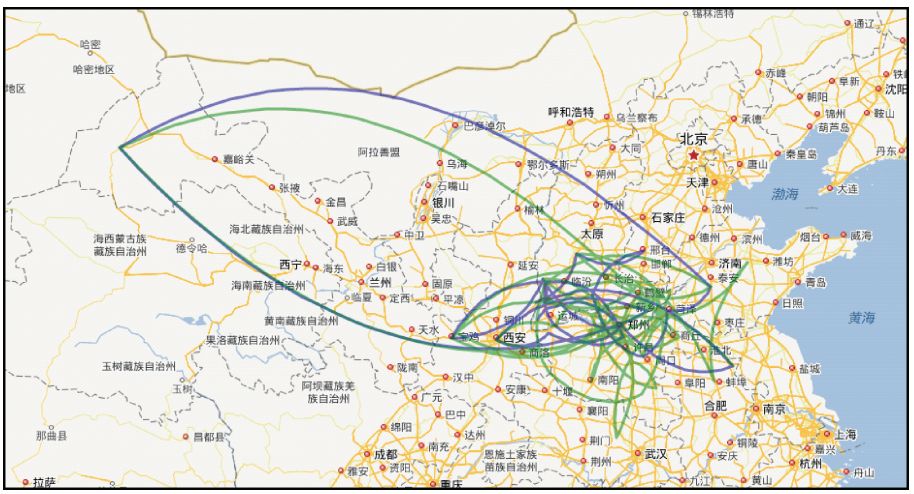

图 8　双人轨迹对比图示例（蓝线为晋惠公轨迹、绿线为晋文公轨迹）

5　计量分析

利用《左传》历史人文数据库和检索系统，可以进行丰富的计量分析工作，得到传统的全文检索数据库难以获得的统计数据。在人物方面，2406 个人物共有 4284 个人名。其中，2262 个为男性，144 个为女性。可以看出这是一部男性主导的历史。在地名方面，共有 995 个地理实体，其中 680 个为诸侯国辖内的地点，还有 19 条河流、11 座山等。值得注意的是，仍有 222 个地名暂时无法考证出具体所在地，例如“阪高”“陘隰”等。这样穷尽的标注工作，可以列出很多疑难问题，为历史研究者快速发现问题提供线索。下文从最热人名、地名的频次及同现等方面展开计量分析。

5.1　重名统计——最热春秋人名

基于《左传》人名信息的详细标注，解决了同名异指和异名同指的问题。通过统计人名数据库，可找出《左传》高频重名的人名，如表 3 所示。

表 3 中所列人名首字为氏，他们所属的诸侯国多为大国，尾字“侯”“子”“伯”均为诸侯王的爵位封号，“氏＋爵位”的结构是春秋人名的典型结构，这种结构的人名又是当时的诸侯国君，不同诸侯国的国君均可用此称号，所以成为热度最高的重名。表 3 的统计结果表明，传统的基于字符串检索是难以区分这些人物的，而在计算机进行古籍人名自动识别和消歧时，应重点关注这些频率较高的重名结构，并制定相应的识别规则。

表 3　《左传》高频重名表

人物名称	人物数（个）	频次（次）
晋侯	16	263
卫侯	15	106
齐侯	15	207
楚子	13	176
郑伯	12	163

5.2　实体同现统计

人物与人物、人物与地名的同现数据，可以用来近似计算人物关系和人物所去过的地方。对《左传》文本中每个句子内的“人物 ID－人物 ID”同现情况进行统计，共得到 12166 条信息。对“人名 ID－地名 ID”的同现情况进行统计，共有 6640 条同现信息。每一条信息都含有同现频次，可以近似地作为两个实体的相关度。比如人物同现次数越多，则两个人物的相关度越高。在此基础上，统计出人物交往的广度和强度，得出人物社交排行榜。

（1）最“广交”人物

用人物同现次数来近似地估计人物之间的交往关系，同现人物越多，交际也就越广。表 4 列出了“广交”的十大人物，最“广交”《左传》人物前三名为：晋文公 99 个同现人物，晋悼公 85 个，范宣子（晋国法家人物）71 个。由此可以看出晋国君臣在春秋时代的地位。

表 4　十大最“广交”人物

ID	人物名称	同现人物数(个)
1207	晋文公	99
1505	晋悼公	85
964	范宣子	71
1386	晋景公	70
1824	楚庄王	65
297	齐桓公	62
1993	郑文公	61
1966	晋厉公	57
1008	羊舌肸	56
689	楚共王	55

(2)最“密交”人物

人物之间同现次数越多，则很可能是关系特别密切，或者在后世眼中同等重要。表 5 给出了十大“密交”人物：公子宋和公子归生，伯舆和王叔陈生，均同现 29 次，子叔声伯和施孝叔同现 28 次。后羿和寒浞的儿子寒浇、豷等人并非春秋人物，皆出自夏朝时期从“太康失国”到“少康中兴”的前后百年之间，《左传》所记臣子谏言多引此典，以喻兴衰，从而使后羿之典中的人物频频同现。此外，还可以使用互信息、卡方等统计量，进一步计算出人物之间的关系强度，观察中低频次的人物关系。

表 5　十对最高频同现人物

ID1	人物 1	ID2	人物 2	同现频次(次)
314	公子宋	154	公子归生	29
898	伯舆	245	王叔陈生	29
135	子叔声伯	1251	施孝叔	28
792	后羿	2136	寒浇	28
1207	晋文公	1302	秦穆公	27
1966	晋厉公	1234	郤至	25
792	后羿	1786	寒浞	23
1966	晋厉公	2368	栾书	23
282	荀偃	964	范宣子	22
792	后羿	2316	豷	21

(3)春秋霸国的验证

春秋霸国的归属各个版本向来不尽相同，墨子和荀子认为五霸归属齐、晋、楚、吴和越，东汉班固归结五霸主为齐桓、晋文、秦穆、宋襄、楚庄，历史学家翦伯赞在《先秦史》中指出五霸当属齐、晋、楚、秦、郑[22]。笔者通过三个角度进行数据统计，给出关于“春秋霸国”量化的参考结果：①基于《左传》标注文本，结合“地名表”信息，统计了诸侯国在文本中的出现频次，表 6 给出了出现频次最高诸侯国前五名；②基于“人物地名表”对于人物相关地名的数据，表 7 给出了诸侯国相关人物频次最高的前五名；③基于“人物表”中人物的国别信息，表 8 给出了诸侯国所属人物数量最多的前五名。

表 6　诸侯国频次表

诸侯国	频次(次)
晋	816
楚	634
郑	518
齐	358
宋	325

表 7　诸侯国相关人物频次表

诸侯国	相关人物频次(次)
楚	5190
宋	2384
秦	2074
晋	1984
齐	1946

表 8　诸侯国所属人物数表

诸侯国	所属人物数(个)
晋	1405
楚	948
郑	648
齐	602
鲁	391

统计结果显示，晋、楚、齐三国在三张榜单上均榜上有名，亦为各家说法所共有，霸主之名可谓实至名归。另外，表中上榜的诸侯国里除了鲁国，均曾被人列为霸国，这与《左传》出于鲁国史官之手有关，所以鲁国的人物较多。

5.3　人物游历地点、距离统计

春秋时期的交通设施与今日不可同日而语，古人的一生能游历多少地方、旅行多远距离，我们并不清楚，在古书中也没有直接的记载。借助数据库技术，则可以近似地计算出人物可能游历过的地点和行程长度。由于“人物地名表”包含了人物相关的所有地名，“时间人物表”包含人物和地名同现的时间，两相结合可以得到人物相关的地名序列，即人物活动轨迹，进而计算出人物活动距离。以下按照沿两点之间直线距离、沿当代道路步行距离两种方式对每个人物活动距离进行测算。

首先，根据地球曲面上两点近似距离公式[23]，求得两地之间直线距离，求总和可得每个人物的活动距离总长。

$$\mathrm{Dis}=111.199*\sqrt[2]{(\varphi_1-\varphi_2)^2+(\lambda_1-\lambda_2)^2*\cos\left(\frac{\varphi_1+\varphi_2}{2}\right)^2}$$

（λ 为经度，φ 为纬度）

其次，调用百度地图的步行距离接口，计算活动距离。计算结果发现，在具有人地同现信息的 1604 个人物中，人均直线距离 1029.98 公里，沿道路距离 1308.26 公里，活动地点 3.97 个。如表 9 所示，活动距离最长、同时也是活动地点最多的前十位春秋人物中，七位为君王，三位为大夫、将军。活动距离最长的春秋人物是周武王，行经 2 万余公里，可见这些人物征伐与游历的历史。这种近似计算的方法，虽然不是特别精确，却可以快速地获得古人可能的行程长度，如果加上时间信息，还可以推算出当时的交通设施状况和行进速度。

表 9 《左传》人物游历地点数和距离

名次	人物名称	地点数（个）	直线距离（公里）	步行道路距离（公里）
1	周武王	48	19300	23514
2	崔杼	33	16018	20021
3	晋文公	47	15964	19757
4	楚庄王	36	14907	18808
5	范宣子	35	14498	17391
6	郑文公	39	14362	17835
7	秦穆公	33	13860	16033
8	知武子	36	13828	16930
9	晋景公	38	12594	15913
10	周文王	33	11996	15296

5.4 多事之秋——实体历时分布

以时间为轴线，根据文本中实体的密度，可以发掘古代文本中重要的时间。相比篇幅的长短，命名实体所出现的频次密度可能更能反映某一时段的重要性。根据《左传》每个段落的时间信息，对应到公元纪年后，可以得到文献中记载的公元前 722－468 年之间的人物、地点出现次数的曲线。从图 9 可以看出，人物和地点的出现高度相关，人名略多于地名，且在不同时间上差异巨大。人物曲线在公元前 542 年前后更是达到最高峰 274 次，地点曲线在公元前 548 年到达顶峰 176 次。这种历时分析可以快速而清晰地定位出历史上的多事之秋。

综上，借助数据库工具和 GIS 技术，用定量的视角探究《左传》中的人物和地点，形成春秋时期的历史人物时空，统计所得信息和数据，也可看作是对古籍文本本身的增值或补充。希望通过基于“词”的、多角度的检索与计量分析，提供新的研究思路、方法和工具，推进对于历史文献的研究。

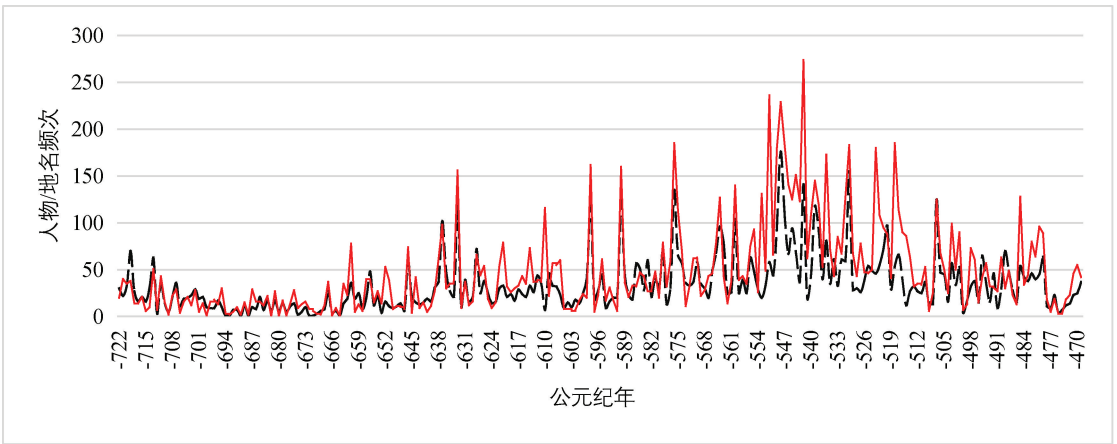

图 9 人物（红色实线）和地名（黑色虚线）出现频次的历时分布

6 结语

我国历史文献丰富，在古籍纸质资源较为丰富的今天，全文检索已广泛应用，但如何综合运用多种现代数字化技术，进行古籍内容的深度加工与分析，是古典文献传承与研究的重要议题。文章在数字人文和人文计算的研究范式下，尝试构建了史学名著《左传》的历史人文数据库，主要进行了以下工作：

（1）建立了数字人文知识库，标注了人物主名、今地名以及 GIS 数据，创建实体同现表，建成包括人名表、人物表、时间人物表、人物同现表、地名表、人物地名表六张数据表的数据库。

（2）开发了可视化检索系统，基于数据库开发包括全文检索功能在内的线上检索系统，利用百度地图，结合人物、空间、时间信息搭建了人物地图平台。

（3）计量分析，通过数据库内容，以定量视角探究《左传》中的人物、地点数据，发掘出热点人物、人物关系、旅行距离等重要信息。在整体上探索和论证了数字人文技术对于历史文献加工的可行性与基本研究形态。这种新型历史人文数据库的建立，不仅对《十三经》等经典文献的深度数字化有参考价值，对于古籍整理研究工作也是一种新的尝试，能够把古籍的多种信息以数据库的形式存储和展现出来，解决传统的大长编或卡片检索效率低、媒介单一等问题，还可以通过数据分析更好地研究文献内容和多种文献之间的关系，成为古籍研究的参考和辅助工具。此外，通过数据对接，还可以应用于对图书馆、博物馆的馆藏善本古籍和藏品进行文献、实物与地理信息的互联互通。

受限于历史资料的缺漏和有限的人力，研究尚存在以下问题和不足，需要在未来的工作中继续完善：

（1）人物、人物关系、GIS 信息的数据标注，需要结合最新的考古成果不断改进，例如对地名进行更多的属性标注（行政级别、起始年、终结年等）。这需要开发在线纠错功能，让更多的专家学者参与到标注工作中来。

（2）人物同现和人地同现还只是近似计算，将来也可以考虑对人物关系进行更细致的分类，例如朋友、亲属、同僚等等。

（3）可视化检索的呈现方式仍有许多改进的空间。例如，可以同图书馆和博物馆合作，将数据库互联互通，检索结果呈现文献和文物的图片等馆藏信息；反之也可以通过文物查询文献知识库。

（4）希望能与历史人文、汉语史、文化、考古等多个学科联合研究，得出更为深刻、更有价值的结论。

致谢：感谢邓国亮老师和多位匿名审稿人对论文的修改建议，感谢朱福耘、李晓炜、曹艺凡、王雨非同学的数据标注工作。

参考文献

1 许倬云.中国古代社会史——春秋战国时期的社会流动[M].桂林：广西师范大学出版社，2006.
2 夏翠娟，林海青，刘炜.面向循证实践的中文古籍数据模型研究与设计[J].中国图书馆学报，2017，43(6)：16－34.
3 陈小荷，冯敏萱，徐润华，等.先秦文献信息处理[M].北京：世界图书出版公司北京公司，2013.
4 董志翘.为中古汉语研究夯实基础——“中古汉语研究型语料库”建设琐议[J].燕山大学学报(哲学社会科学版)，2011(1)：1－6.
5 中国历代人物传记数据库管理委员会.中国历代人物传记数据库项目(China Biographical Database，CBDB)[EB/0L].[2019－05－05].https://projects.iq.harvard.edu/chinesecbdb.
6 Elson D K，Dames N，Mckeown K R. Extracting social networks from literary fiction[C]//Proceedings of the 48th annual meeting of the association for computational linguistics.Sweden：ACL，2010：138－147.
7 陈刚.“数字人文”与历史地理信息化研究[J].南京社会科学，2014(3)：136－142.
8 The Open University.Hestia[EB/0L].[2019－05－05].https://hestia.open.ac.uk/.
9 University of Portsmouth. Great Britain historical geographic information system (GBHGIS)[EB/OL].[2019－05－05]. http://www.geog.port.ac.uk/hist－bound/project_rep/proj_gbhgis.htm.
10 Minnesota population center. The national historical geographic information system (NHGIS)[EB/OL].[2019－05－05]. http://www.nhgis.org.
11 复旦大学历史地理研究中心.中国历史地理信息系统(China historical geographic information system，CHGIS)[EB/OL].[2019－05－05].http://www.fas.harvard.edu/～chgis.
12 夏翠娟.中国历史地理数据在图书馆数字人文项目中的开放应用研究[J].中国图书馆学报，2017，43(2)：40－53.
13 王兆鹏，搜韵网.唐宋文学编年地图[EB/OL].[2019－05－05].https://sou-yun.com/poetlifemap.html.
14 搜韵网.历代诗人地域分布[EB/OL].[2019－05－05].https://sou－yun.com/IndexByMap.aspx.
15 中国台湾电子佛典协会. 法鼓山佛典数据库(CBETA)[EB/OL].[2019－05－05]. http://www.cbeta.org/.
16 陈小荷，李斌，冯敏萱，等. 先秦《左传》语料库[EB/OL].[2019－05－05].https://catalog.ldc.upenn.edu/LDC2017T14.
17 方朝晖.春秋左传人物谱[M].济南：齐鲁书社，2001.
18 杨伯峻，徐提.春秋左传词典[R].北京：中华书局，1988.
19 谭其骧.中国历史地图集(第一册)[R].北京：中国地图出版社，1982.
20 百度.百度地图 API[EB/0L].[2019－05－05].http://api.map.baidu.com/lbsapi/cloud/index.htm.
21 南京师范大学文学院.《左传》检索平台[EB/0L].[2019－05－05].http://langsphere.com/zzsk/.
22 陈筱芳.“春秋五霸”质疑与四霸之成功[J].西南民族大学学报(人文社会科学版)，1992(5)：83－88.
23 韩忠民.知经纬度计算两点精确距离[J].科技传播，2011(11)：196－197.

作者单位：李斌、陈小荷，南京师范大学文学院，江苏南京，210097
王璐，北京大学对外汉语教育学院，北京，100871
王东波，南京农业大学信息科学技术学院，江苏南京，210093

收稿日期：2019 年 5 月 29 日
修回日期：2020 年 8 月 20 日
（责任编辑：关志英）

（转第 90 页）