

doi:10.13582/j.cnki.1672-7835.2014.04.001

■ 毛泽东研究

毛泽东与中国工业化理论和政策的开创^①

——基于数据挖掘的新发现

冷兆松

(中国社会科学院当代中国研究所,北京100009)

摘要:数据挖掘技术在历史文献研究中的创新性应用,使得总路线宣传提纲和“一五”计划等旧文献释放出新的史料价值——发现它们是标志毛泽东对中国工业化理论和政策创新的重要文献,有9个方面的开创性贡献,至今仍具有启发和指导意义。本案例表明,数据挖掘技术运用于历史文献研究,可以实现方法创新和知识创新。

关键词:毛泽东;中国工业化;总路线宣传提纲;“一五”计划;数据挖掘

中图分类号:A84;F420 **文献标志码:**A **文章编号:**1672-7835(2014)04-0001-06

Mao Zedong and China's Industrialization Theory and Policy: The New Discovery Based on Data Mining

LENG Zhao-song

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

Abstract: Through the innovative application of data mining technology into the research of historical literature, the two old documents, i. e. both the publicity outline of the general line, and the “first five-year” plan have released new historical values. Marked as Mao Zedong's innovation of the important literature on China's industrialization theory and policy, they have made nine pioneering contributions, which are still of enlightenments and guiding significances. This case shows that data mining technology, applied in the study of historical literature, can realize both method innovation and knowledge innovation.

Key words: Mao Zedong; Chinese industrialization; the publicity outline of the general line; the “first five-year” plan; data mining

一 国史党史历史文献挖掘的步骤、流程和工具

(一) 数据挖掘技术

关于数据挖掘,有多种解释,最常见的有基于数据库的挖掘、基于应用数据的挖掘、基于信息集合的挖掘3种概念。中国工程院院士、系统工程与管理科学专家王众托认为,数据挖掘是现代信息技术和现代数学、系统科学等思想、方法集成后形成的一种新型技术,以挖掘深度为标准可以划分为2种类型:一

① 收稿日期:2013-11-08

作者简介:冷兆松(1963-),男,山东烟台人,博士,研究员,主要从事当代中国经济研究。

是浅层次的在线分析,比如收款机对数据的即时处理;二是深层次的知识获取,这种数据挖掘技术包括统计、分类、聚类、回归分析、特征化、变化和偏差分析等多种类型^[1]。这里只简要介绍本文涉及的分

类、聚类分析技术。

分类,就是以某种共同点为划分标准,依据预先设定的分类模式,把一大堆数据里不同的对象划分成不同的类。比如汽车经销商,以喜好为标准、依据多种分类模式,可以对客户作出多种分类:喜好小排量汽车的客户、喜好大排量汽车的客户、喜好跑车的客户、喜好越野车的客户、喜好轿车的客户,等等。

聚类,与分类相比较的一个重要特点是,通过聚类分析把数据划分成的组或类不是预先定义好的。聚类分析技术,要通过对数据的分析,看哪些数据在一块聚堆儿,使得同组内数据对象的相似性最大化,聚出的堆就形成类^{[1],[2]}。

世界范围内数目庞大的成功案例清楚地表明,包括聚类分析在内的数据挖掘技术,具有非常神奇的功能,可以让数据“发声”,使我们发现很多以前完全没有意识到的重要联系^{[3]19}。在本文中,总路线宣传提纲和“一五”计划与毛泽东之间重要内在联系的发现,就是聚类分析运用于历史文献研究的成果。

(二)数据挖掘技术与历史文献研究

数据挖掘技术是基于计算机实现的,而计算机只能处理由0,1组合的代码串和结构化的数字信息。而历史文献是非结构化的文本信息,需要经过文本转换才适用数据挖掘。专注于文本信息的数据挖掘,被称为文本挖掘。文本挖掘技术可直接适用于历史文献研究。

在中国,传统中医文献的文本挖掘进展较好,取得了不少研究成果和技术专利。但是,国史党史的文本挖掘,研究成果极其罕见,技术专利更是空白。

韶山毛泽东同志纪念馆冯瑛在论文中指出,数字化研究方法在毛泽东研究领域的应用还是盲点。引入文本挖掘等数字化研究方法,拓展毛泽东思想研究新渠道,是急需解决的重大课题^[4]。

(三)国史党史历史文献挖掘的步骤、流程和工具

数据挖掘,是大数据时代知识发现过程的关键步骤。知识发现一般包括3个基本步骤,一是输入和整理原始数据,二是进行数据挖掘,三是对数据挖掘成果进行解释、评价和表达。

在数据挖掘过程中,通常会用到数据仓库这个工具。数据仓库,是在数据库的基础上、不满足于对数据的一般处理而产生的一个工具。数据库,通常情况下对数据进行特定处理之后就完成使命了;数据仓库则不同,它要不断积累数据,并且不断加以分析和处理^[1]。

在国史党史的历史文献挖掘中,数据仓库是一个很有用的工具。我们以毛泽东与中国工业化为例,概要说明国史党史历史文献挖掘的步骤、流程和工具(图1)。

第1步,以联机检索、搜索引擎为信息抽取器,通过资料分类,建立多个专题数据库(A,B,C)。(1)通过联机检索的方式,对中国知网、人大报刊复印资料、人民日报、中共中央文献等所有可能涉及毛泽东与中国工业化的数据库,实施多个关键词的全面检索,最大限度地抽取全部相关资料。通过百度、读秀等搜索引擎,对国际互联网、图书馆等实施多个关键词的全面搜索,力求抽取全部重要资料。(2)以专题或关键词为分类依据,把抽取的资料重组到各个对应数据库。对于包括多个专题或关键词的文献资料,可以切块分割之后,再分别放到相关数据库。

第2步,以数据库A、数据库B、数据库C为基础,建立“数据仓库1”,并进行分类和聚类分析。根据分析结果,通常还需要进行第二次甚至多次定向资料搜索(即定向重复第1步的各项工作),不断丰富和完善数据仓库,形成“数据仓库2”甚或“数据仓库3”等,并据以实施第二次甚或多次分类和聚类分析。这是知识发现的关键一步。

第3步,对数据挖掘成果进行解释、评价和表达。本文将毛泽东与中国工业化文献挖掘成果的解释、评价和表达,作为第二、三部分,予以重点展示。

需要指出的是,在具体实施过程中,上述每个步骤都包含了多个环节、细节、专业技巧,操作过程极其繁复,不便表述和图示。

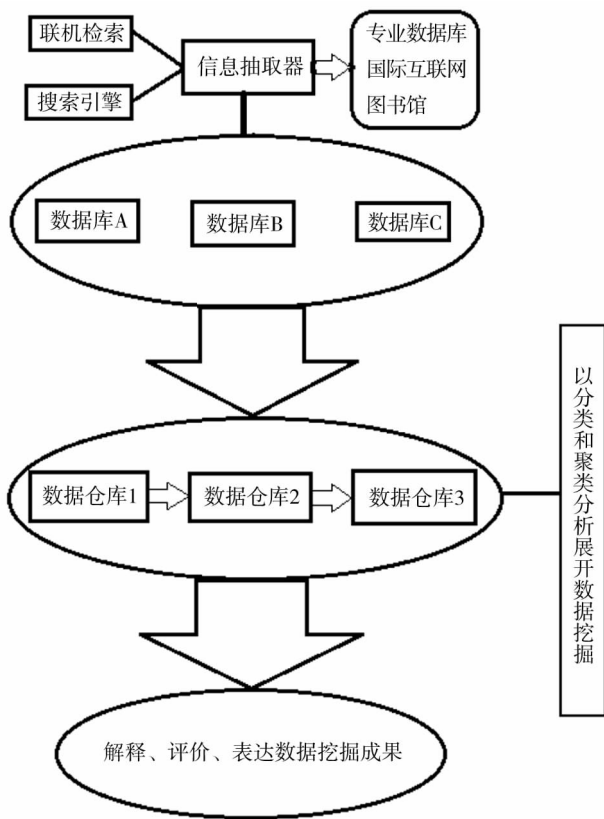


图 1 国史党史历史文献挖掘步骤、流程、工具

二 旧文献释放出新价值——标志毛泽东对中国工业化理论和政策创新的重要文献

本部分是对数据挖掘成果的表达。

通过数据挖掘技术在历史文献研究中的创新性应用我们发现,1953年12月中央宣传部制发的总路线宣传提纲^{[5]596}和1955年7月一届全国人大二次会议通过的第一个五年计划^{[6]350},这两篇非个人署名的重要文献与毛泽东的关系非同寻常。

1953年9月24日党在过渡时期的总路线首次正式向全国公布之时,中央宣传部应毛泽东的要求起草总路线宣传提纲,六易其稿都没有达到毛泽东的要求,到第七稿时他才比较满意,毛泽东作了重要修改和增补之后最终定稿。1953年12月13日,他写信给主持起草工作的胡乔木,对修改和审定情况作了简要说明^{[7]265-266}。总路线宣传提纲立项、起草、修改、定稿的全过程,都是在毛泽东主导下进行的,并经他亲笔修改后定稿。因此,宣传提纲实际上是由毛泽东主导和领衔完成的中共中央的集体之作,直接而充分地表达了毛泽东的思想观点。

为筹划国家工业化,早在国民经济恢复时期,我国第一个五年计划的制定工作,在周恩来、陈云、李富春主持下就启动了,毛泽东高度重视并全程指导,仔细审阅了一稿又一稿的计划草案。《五年计划轮廓草案》完成后,毛泽东委托周恩来和陈云等人于1952年8月带到苏联,商请苏共中央和苏联政府提出意见和建议。9月24日,周恩来和陈云回国的当天,毛泽东主持召开了中央书记处会议听取同苏联商谈的情况汇报,讨论“一五”计划的各项重大问题。正是在这次会议上,毛泽东首次提出了过渡时期总路线的初步构想,并将其确定为“一五”计划的根本指导思想。1954年11月3日,毛泽东等中央领导人到达广州,集中审议“一五”计划草案,毛泽东在这里一住就是四个星期^{[7]269-307}。^[8]。审定完成的“一五”计划,于1955年7月30日获得一届全国人大二次会议通过,并且作为国家工业化的良好开端载入史册。总之,“一五”计划的起草、制定、定稿、审议的全过程,都是在毛泽东主导下完成的。因此,“一五”计划实际上是由毛泽东主导和领衔完成的中共中央的集体之作,直接而充分地表达了毛泽东的思想

观点。

总之,以数据仓库为工具、通过分类和聚类分析海量的历史文献资料,或者说,通过数据挖掘技术在历史文献研究中的创新性应用,使得总路线宣传提纲和“一五”计划这两篇旧文献显现出新的重要史料价值——二者都是由毛泽东主导和领衔完成的,与毛泽东之间具有密不可分的内在联系。

学术界还没有意识到这一点。为了便于理解,不妨假设给这两篇文献署名,依据上述考证和分析获得的实际历史情况,则:总路线宣传提纲的主编是毛泽东,执行主编是胡乔木;“一五”计划的主编是毛泽东,执行主编是周恩来、陈云、李富春。通过这样的署名假设,这两篇文献与毛泽东之间的重要联系,就一目了然了。

三 总路线宣传提纲和“一五”计划对中国工业化理论和政策的开创性贡献

本部分是对数据挖掘成果的解释和评价。

由毛泽东主导和领衔完成的总路线宣传提纲和“一五”计划,是开创中国社会主义工业化理论和政策的经典文献,宣传提纲着重于理论创新,“一五”计划着重于政策创新。它们赋予中国社会主义工业化理论和政策9个方面的开创性贡献。

第一,作出关于中国旧有工业水平和性质的基本判断。

中国旧有工业是十分落后的。钢铁工业基础极其薄弱,没有真正的机器制造工业,没有现代国防工业。1949年,使用机器的工业的产值约占工农业生产总值的17%,农业及其副业、个体手工业和工场手工业的产值约占83%。

中国旧有工业是带有殖民地半殖民地性质的,有轻(工业)缺重(工业)的畸形产业结构从根本上缺失独立自主的基础和能力。主体是轻工业,只有零星的重工业,这又主要是帝国主义国家在中国的修理工厂(轮船、铁路等)和为帝国主义国家提供原料和半成品的矿山和工厂。

中国旧有工业的主体,是资本主义工业。1949年,资本主义私营工业占工业总产值63%,国营工业占34%,合作社经营和公私合营工业共约占3%^{[5]604-605}。

新中国成立之初,或者说过渡时期开始之时,我国工业的基本状况就是这样。轻工业是穷困落后的,重工业是几近空白的,这就是我们一直以来所说的1949年新中国所继承旧有工业的“一穷二白”。

第二,作出关于三年恢复时期后中国工业的水平和性质的基本判断。

经过1950年到1952年3年的经济恢复时期,现代化工业获得了显著的增长,国营工业在现代化工业中获得了优势地位,资本主义工业仍然占据相当大的比重。1952年,现代工业占工农业总值的比重约为28%,农业、个体手工业和工场手工业占工农业总产值的比重约为72%;在工业总产值的所有制构成中,国营工业约占51%,资本主义工业约占40%,合作社经营、公私合营工业合计约占9%。

虽然3年恢复时期现代化工业取得了很大进步,我国仍然是贫穷落后的农业国,拖拉机、汽车、飞机都不能自行制造,没有现代国防工业。1952年,钢产量平均每人约2.4公斤,只相当于苏联的1/63,棉布产量平均每人可得约9米,相当于苏联的1/3强^{[5]605}。

第三,第一次全面定义了“中国社会主义工业化”的科学内涵。

以“三个使”全面定义了中国社会主义工业化3个方面的科学内涵。一是使我国拥有强大的重工业以独立自主地装备各个工业部门,二是使现代化工业在国民生产总值中占绝对优势,三是使社会主义工业成为我国唯一的工业^{[5]605}。第一、二项是从生产力角度下定义,第三项是从生产关系角度下定义。所以,中国社会主义工业化的内涵,既包括生产力的规定性,也包括生产关系的规定性,是生产力和生产关系的内在统一。

从生产力和生产关系两个方面规定“中国社会主义工业化”的内涵,符合毛泽东一贯的思想方法,也是马克思主义的基本思想方法。

第四,把社会主义工业化,明确为中国工业化的本质特点。

中国工业化首先必须发展国营工业,并发展国营交通运输业、国营商业和合作社商业,对农业、手工业和资本主义工商业进行社会主义改造,确保社会主义经济成分的比重不断增长,“使社会主义工业成为我国整个国民经济发展的起决定作用的领导力量”^{[5]603,608-609}。

第五,以“四个就可以”,说明实现社会主义工业化的重要意义;以“四个可以不”,批判不重视工业化的种种错误思想。

“四个就可以”——“实现国家的社会主义工业化,就可以促进农业和交通运输业的现代化,就可以建立和巩固现代化的国防,就可以保证逐步完成非社会主义经济成分的改造……就可以大大发展社会主义的商业,大大加强工农联盟,并且大大提高国家的经济财政力量和人民的收入,使全体人民的物质和文化生活水平可以有把握地、不断地提高。”^{[5]605-606}

“四个可以不”——“要批判一切违反总路线的错误思想,如认为我国可以不要工业化、可以不忙工业化、可以降低工业化速度、可以不以发展重工业为中心,认为工业化对农民和一般人民不利,认为有了苏联的援助,我国的工业化就不重要等等思想。”^{[5]612}

第六,以“一个只有”、“四个如果”,说明我国工业化的启动阶段必须以重工业为中心的基本依据。

只有建立了重工业,才能装备工业、运输业、农业等国民经济的各行各业。如果不建立重工业,就没有现代化国防,帝国主义一定还要欺侮我们。如果不建立重工业,运输业还会停留在破旧状态,轻工业也不能满足人民需要。如果不建立重工业,现有轻工业将会因为得不到新装备的后续补充和持续改进而变得破旧不堪。如果不发展机器工业和化学工业,农民将得不到新式农具和农业机器,得不到更多更好的化肥,农业合作化和农产品增加这2个方面都必将遇到困难。“所有这一切都说明国家社会主义工业化的中心必须是发展重工业。”^{[5]606-607}

“一五”计划期间是我国工业化的启动阶段,根据我国当时的基本国情以及重工业严重缺失的畸形产业结构,以重工业为中心和重点是完全正确的。经过“一五”计划重点和优先发展重工业战略的实施,我国有轻(工业)缺重(工业)的畸形工业结构得到了初步的修正,因此,1957年毛泽东不失时机地提出“一手抓工业,一手抓农业”两手抓的思想^{[9]1005},在1959年至1960年期间毛泽东提出和发展了在优先发展重工业的条件下发展工业与发展农业同时并举的思想^{[10]121-123}。

第七,提出以“两个优先发展”的工业政策,谋求“两个显著增长”的发展目标。

规定“两个优先发展”的工业政策。在“一五”期间,必须优先发展重工业,以求得建立巩固的国防、满足人民需要和对国民经济实现社会主义改造的物质基础;必须优先发展社会主义经济,在这个原则下对各种经济成分实行统筹兼顾^{[6]352}。

谋求“两个显著增长”的发展目标。通过“一五”计划的实施和完成,“不仅将使我国的国民经济有巨大的发展,现代工业在工业农业总产值中所占的比重有显著的增长,而且将变更各种社会经济成分之间的关系,即社会主义经济成分在国民经济中会有很大的增长,而其他经济成分则相应地缩小其原来在国民经济中所占的比重”^{[6]357}。

第八,提出以“两个中心”为工业发展重点,把国民经济推进到现代化的技术轨道上。

“一五”计划规定,不但要以大力发展重工业为中心,而且要以苏联援建的“156项”为“工业建设计划的中心”,以便“创造现代化的技术基础”,“把我国国民经济从技术极端落后的状况推进到现代化技术的轨道上”^{[6]364-365,486}。

第九,作出关于苏联经验、装备和技术援助的评价。

苏联优先发展重工业的先进经验是我们可资借鉴的榜样,与此同时,我们也直接得到了苏联装备的支持,直接得到了苏联专家的技术援助。这三个方面,构成了我国完成社会主义工业化的有利条件^{[5]607-608,611}。

1958年毛泽东总结说,从1956年4月的《论十大关系》开始提出我国自己的建设路线,“认真学习外国的好经验,也一定研究外国的坏经验——引以为戒,这就是我们的路线”^{[12]369-370,380}。1963年、1970

年,毛泽东多次论述和发展了这个观点^{[10]338-339,[11]106}。

以上九个方面,是由毛泽东主导和领衔完成的中共中央的集体之作——总路线宣传提纲和“一五”计划这两篇文献,对中国社会主义工业化理论和政策的开创性贡献。其中的多个方面,至今对我们还有启发和指导意义。篇幅所限,仅举一例。1991年7月1日,江泽民在庆祝中国共产党成立七十周年大会上的讲话中,在谈到正确认识和准确把握社会主义初级阶段基本国情时讲了3个方面,其中的第一点说,由于“中国经济建设是从旧中国一穷二白的基础上起步的”,虽然经济建设经过几十年努力取得了巨大成就,但是仍然难以满足人民群众日益增长的物质文化需要^{[12]151}。这里所说的旧中国一穷二白的经济基础,这个客观的历史判断是在总路线宣传提纲之中就明确了。

四 结 语

从新中国成立到改革开放前,尽管经历了“大跃进”和“文化大革命”两次大折腾,我国工业和经济总体上保持了高速增长,不仅是半殖民地半封建的旧中国所无法比的,也是许多工业化国家所不及的。在几乎一穷二白的基础上,建立起独立的、比较完整的、有相当规模的现代工业体系,迈入工业化初期阶段,实现了中国工业化发展的历史大跨越,为改革开放后工业经济进入腾飞期奠定了物质技术基础。

1956年,毛泽东预言:“二千零一年,也就是进到二十一世纪的时候……中国将变为一个强大的社会主义工业国。”^{[13]156}这位领导中国人民成功开创社会主义工业化道路一代伟人的预言,是可靠的——进入21世纪以来,全世界都在谈论中国的崛起。

在中国工业化成就辉煌之时,我们不能忘记,这是与毛泽东带领中共第一代中央领导集体成功开创中国社会主义工业化的理论与政策密不可分的。

参考文献:

- [1] 王众托. 教育数据挖掘与知识发现[J]. 基础教育课程,2013(Z1):13-14.
- [2] 平金珍,王 茜,于莉莉. 聚类分析在 Web 数据挖掘中的应用研究[J]. 科技信息,2013(19):75.
- [3] (英)维克托·迈尔-舍恩伯格,肯尼思·库克耶. 大数据时代[M]. 盛杨燕,周 涛,译. 杭州:浙江人民出版社,2013.
- [4] 冯 瑛. 毛泽东遗物及生平思想研究的数字化[J]. 毛泽东思想研究,2011(3):78-81.
- [5] 建国以来重要文献选编(第4册)[M]. 北京:中央文献出版社,2011.
- [6] 建国以来重要文献选编(第6册)[M]. 北京:中央文献出版社,2011.
- [7] 逢先知,金冲及. 毛泽东传(1949~1976)[M]. 北京:中央文献出版社,2003.
- [8] 逢先知,李 捷. 毛泽东与过渡时期总路线[J]. 党的文献,2001(4):46-57.
- [9] 毛泽东著作专题摘编[M]. 北京:中央文献出版社,2003.
- [10] 毛泽东文集(第8卷)[M]. 北京:人民出版社,1999.
- [11] 建国以来毛泽东文稿(第13册)[M]. 北京:中央文献出版社,1998.
- [12] 江泽民文选(第1卷)[M]. 北京:人民出版社,2006.
- [13] 毛泽东文集(第7卷)[M]. 北京:人民出版社,1999.

(责任校对 谢宜辰)