

基于 LDA 模型的新能源汽车政策文本量化分析

赵公民, 吕京芹, 武勇杰

(中北大学 经济与管理学院, 太原 030051)

摘要:运用潜在狄利克雷分配模型(LDA)进行主题挖掘,以国家和中部六省两个层级的 144 个新能源汽车相关政策作为研究对象,从高频词汇、主题强度及年度变化等方面进行量化分析。结果表明国家层级主题主要是节能发展、动力电池回收等,从绿色发展的整体角度推动发展;中部六省层级的主题主要是设施建设和财政补贴等,更注重从基础保障方面推进;新能源汽车产业政策体系较为完善,但是层级间政策存在互动性不足、忽视传导渗透作用等问题。

关键词:主题挖掘;LDA;新能源汽车;政策文本

中图分类号:F426.471 **文献标志码:**A **文章编号:**1671-1807(2021)01-0049-07

随着环境污染与资源消耗问题日益严重,作为环保产业的新能源汽车受到广泛关注。国家十分重视新能源汽车的发展与规划,自 2001 年“十五”时期的电动汽车专项课题开始关注新能源汽车^[1]。2010 年开始大力召开相关会议开展新能源汽车推广工作,并颁布具体行业政策。《中国制造 2025》报告中,新能源汽车的发展已经上升为国家战略^[2]。2011 年新能源汽车销量仅为 8 259 辆,2019 达到 120.6 万辆,可见新能源汽车产业的快速发展与国家产业支持政策密不可分。因此,政策文本的量化研究对于新能源汽车的发展以及未来政策制定具有重要意义。

国内外专家学者们对政策文本的研究涵盖众多领域。国外学者 Kitson 等^[3]收集 1990—2010 年近 20 年间的卫生政策进行研究,确定以患者为中心的护理核心要素;Talamini 与 Dewes^[4]研究巴西有关政策文本与学术文献,分析液体生物燃料政策效果与科研成果之间的差异;Jansa 等^[5]收集了 1984—2014 年间美国 50 个州 12 项政策的法案文本,研究时间推移政策文本的全面性变化。国内学者曾婧婧^[6]以 2004—2014 年间泛珠三角区域合作政策文本为研究对象,发现区域发展的差异性以及政府政策制定具有倾向性;黄萃等^[7]以 1978—2013 年出台的 1 298 份科技文献政策为研究对象,发现政府间合作关系的发展趋势;苏飞与刘红^[8]以省级政府数据开放政策为研究对

象,发现开放工作中的不足并提出建议;谭玉等^[9]对 1998—2018 年间出台的 59 篇大学生创新创业政策文本进行分析,总结出创新政策的特征;魏萍等^[10]从实践、空间等维度研究湖北省 294 份“一带一路”建设政策,发现主体的扩散与聚焦点。

杨慧与杨建林^[11]将政策文本量化分析方法分为文献计量、社会网络分析、内容分析以及文本挖掘等类型。在政策文本的研究过程中,文献计量方法应用领域十分广泛,刘颖琦等^[12]运用文献计量的方法分析中国新能源汽车产业联盟在知识转移与技术创新等方面与国外的差距;郑征等^[13]运用文献计量方法从年代分布、研究主题等角度分析离子液体领域的发展态势;潘丹等^[14]综合运用内容分析法和文献计量法研究农村环境管理政策演变特征。社会网络分析方法一般与文献计量方法结合使用,李英等^[15]结合文献计量方法与社会网络分析方法,以 2005—2015 年间的 360 篇文献为研究对象,探讨国际绿色车辆路径的发展态势;赵佳丽等^[16]综合运用文献计量与社会网络分析法研究晋商活动的地理特征及山西的地理优势。关于内容分析法的应用,王薇与刘云^[17]运用内容分析法对国家出台的 47 项新能源汽车政策从基本政策工具与价值链进行分析;谢倩等^[18]运用内容分析法研究乡村教师支持政策。文本挖掘则可以深度分析文本的潜在含义,

收稿日期:2020-08-12

基金项目:山西科技厅软科学重点项目(2019042012-1);山西省哲学社会科学规划项目(1910900032MZ)。

作者简介:赵公民(1970—),男,山西闻喜人,中北大学经济管理学院,院长,教授,博士,研究方向:服务创新和能源发展;吕京芹(1994—),女,山东烟台人,中北大学经济与管理学院,硕士研究生,研究方向:成果转化;武勇杰(1987—),男,山西晋中人,中北大学经济与管理学院,讲师,博士,研究方向:区域经济和产业经济。

张永安与周怡园^[19]采用文本挖掘的方法和 PMC 指数模型法对新能源汽车政策进行量化评价;朱青等^[20]运用文本挖掘和分类方法,以医疗诉讼文本为研究对象,设计出辅助判决的决策支持系统。

通过阅读政策文本量化分析相关文献,发现文本挖掘方法能够凭借技术理性和机器效率对政策文本的潜在信息进行挖掘,并为其他政策文本量化分析方法提供支撑。但政策文本挖掘类的文献数量较少,而且主要聚焦于聚类、分类的研究视角。语义视角的政策研究补充是未来文本挖掘技术的焦点,而主题挖掘在语义研究视角方面具有特殊功能。因此,本文以 2010—2019 年国家、中部六省两个层级的新能源汽车相关政策作为研究对象,运用主题挖掘技术对政策的高频词、主题强度及年度变化进行量化分析,深入挖掘政策文本的潜在信息。

1 数据来源与研究方法

1.1 数据来源

新能源汽车政策文本包括国家、中部六省两个层级,国家层级政策文本主要从中国政府网、国家发展和改革委员会、国家能源局等门户网站进行搜集,中部六省的政策文本主要从相应的省市级的能源局、经信委、发展和改革委员会等相关门户网站进行搜集,并结合“北大法宝”(url:http://www.pkulaw.cn/)的搜索结果对政策文本进行整理。整理得到国家层级的政策文本 45 个,中部六省层级的政策文本共 99 个。

1.2 研究方法

20 世纪 90 年代政策文本挖掘相关研究逐渐得

到发展,研究对象从原始的非结构化文本发展至网页类的半结构化文本再至文献、报告等非结构化文本。信息技术的飞速发展对我们搜集信息数据的方式产生了影响,R 语言的机器学习算法对于文本挖掘研究具有重要意义。主题挖掘能够深入文本内部,文档的主题研究以及分类、聚类研究都需要语义视角的补充,因此大规模语料库的多文档主题挖掘成为热点研究对象。主题挖掘的常用手段为主题模型,主要有潜在语义索引(Latent Semantic Index ,LSI)、概率性潜在语义索引(Probabilistic Latent Semantic Index ,PLSI)和潜在狄利克雷分配模型(Latent Dirichlet Allocation ,LDA)等^[21],由于 LDA 模型的主题分布具有预测训练集与非训练集文档和词,渐渐得到学者们的认可并给予完善,使其成为研究非结构化多文本的有效工具。

2 主题挖掘模型构建

政策文本的特点是文本较长、语义完整且术语较多,与京东、淘宝等购物网站的评价类文本不同,政策文本不包含情感态度和个性差别。而且,政策文本与文献也有区别,其不具有文献中的总分结构,且语言表达内容与领域和类别的不同有关。由于政策文本的特性,要对政策文本进行全文挖掘,并将词层面和语义层面结合起来对文本内容进行主题分析。本文在学者们的研究基础之上,结合新能源汽车政策的相关特点,构建了主题挖掘模型,模型流程图如图 1 所示。模型分为文本获取、文本预处理、N-gram 模型、高频词汇可视化及 LDA 主题模型几个部分。

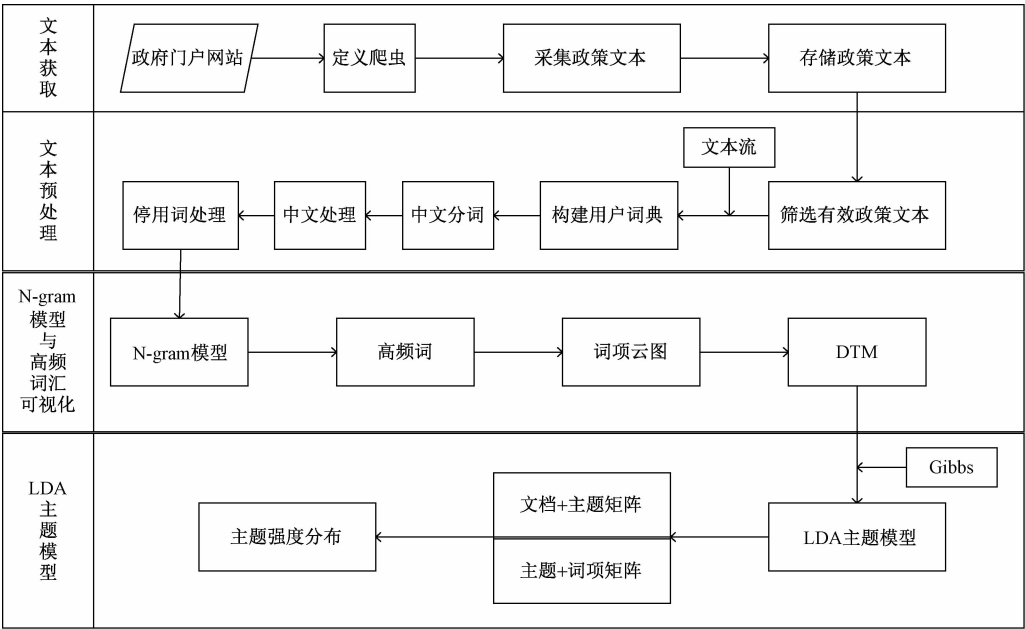


图 1 新能源汽车政策主题挖掘模型示意图

中部六省高频词汇的频率比国家高出两倍多说明各省市的出台政策的关注点较为集中。国家出台的新能源汽车政策,从高频词汇的频率方面来看动力蓄电池产业发展、节能发展和基础设施建设是国家着重强调的方面。中部六省是在国家政策的基础上根据地方特点制定的相关政策,更加注重充电设施建设、补贴补助等方面。

3.2 主题强度分析及可视化

主题强度是用来衡量国家和中部六省的每个主题相对分量,利用 LDA 模型中的文档—主题矩阵和主题—词矩阵,得到每篇文档与每个主题和每个主题与每个词的后验概率。主题强度公式如下:

$$p_k = \frac{\sum_i^n \theta_{ki}}{N}$$
 (3)

其中, N 代表文档的数量; θ_{ki} 代表第 k 个主题在第 i 个文档中的概率; p_k 代表第 k 个主题的主题强度。排名前 20 的主题强度如表 2 所示。由表 2 数据可知,经过软件运算所得到的主体强度分布情况与人工计算结果基本一致。选取强度数值最大的三个作为热门主题,最后确定国家主题(3,13,18),中部六省主题(5,10,20)。表 3 和表 5 分别为国家和中部六省两个层级主题相关度较高的特征词,与人工判断与语料查询相结合得出主题标题。表 4、6 则为两个层级的热门主题年度变化。

表 2 主题强度分布

编号	国家	中部六省	编号	国家	中部六省
1	0.061 815	0.026 310	11	0.030 480	0.070 208
2	0.039 698	0.047 577	12	0.045 337	0.027 588
3	0.074 267	0.039 870	13	0.073 197	0.061 164
4	0.038 728	0.029 102	14	0.042 122	0.032 915
5	0.034 208	0.090 294	15	0.031 922	0.035 960
6	0.048 817	0.048 090	16	0.054 905	0.048 563
7	0.052 598	0.057 020	17	0.050 698	0.040 632
8	0.056 557	0.042 183	18	0.085 433	0.051 917
9	0.036 891	0.052 811	19	0.053 590	0.046 320
10	0.037 073	0.074 137	20	0.051 665	0.077 339

2014 年 Carson 和 Kenneth^[24]提出了一种主题可视化方法:LDAvis。LDAvis 以特征词和主题的关联程度选择表示主题的特征词,而且 LDAvis 可视化图可以帮助人们从整体的视角观察各个主题之间的关系。如图 3、图 4 分别是国家、中部六省的 LDAvis 可视化图,其中每个圆代表一个主题,主题

的文档数目以圆的大小表示,圆越大表示主题的文档越多,各个主题的关联程度以圆的距离表示,距离越近代表关联程度越大。

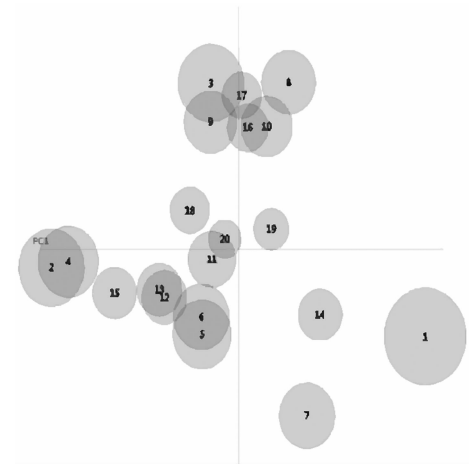


图 3 国家层级 LDAvis 可视图

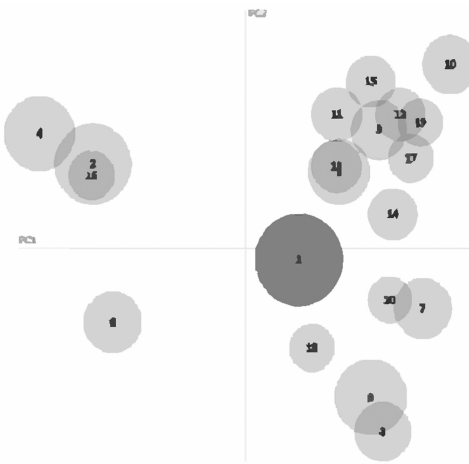


图 4 中部六省层级 LDAvis 可视图

3.2.1 国家主题强度分析

由表 3 可知,3 节能发展、13 动力蓄电池回收、18 基础设施建设是国家热门主题。伴随着环保问题越来越受到公众关注,集低污染、高智能于一体的新能源汽车成为汽车产业发展的重点方向。在国家大力发展新能源汽车的基础之上,动力蓄电池作为其关键部件也呈现快速增长状态,由于电池充电与放电数量的增加呈现的衰减性,电池的寿命要远远小于汽车的寿命,因此电池的回收再利用是需要引起关注的重点问题。主题 18 基础设施建设, Sierzchula 等^[25]对 30 多个国家的新能源汽车数据进行实证分析,发现充电基础设施的建设是新能源汽车销售的关键因素。

表 3 国家热门主题词及相关度

相关主题	主题含义		主题含义		主题含义	
	特征词	相关度	特征词	相关度	特征词	相关度
国家热门主题	3 节能发展		13 动力电池回收		18 基础设施建设	
	汽车	0.140 508	蓄电池	0.112 112	充电	0.094 969
	新能源	0.093 146	动力	0.107 144	基础设施	0.083 827
	节能	0.059 938	利用	0.061 505	建设	0.049 066
	发展	0.049 594	回收	0.054 053	发展	0.015 197
	技术	0.033 263	企业	0.049 396	电站	0.013 414
	汽车产业	0.025 097	废旧	0.047 223	配建	0.012 968
	研究	0.022 375	生产	0.023 937	停车场	0.012 523
	零部件	0.016 386	综合利用	0.021 143	网络	0.012 523
	相关	0.014 753	拆解	0.014 313	城市	0.011 186
	混合	0.014 209	再生	0.013 692	电动汽车	0.010 740

表 4 国家热门主题年份变化表

年份	主题 3 节能发展	主题 13 动力电池回收	主题 18 基础设施建设
2010	0.069 922	0.007 468	0.014 484
2011	0.218 191	0.014 041	0.006 452
2012	0.036 957	0.012 374	0.010 508
2013	0.032 207	0.008 976	0.034 319
2014	0.044 617	0.011 988	0.031 636
2015	0.037 355	0.011 376	0.126 682
2016	0.015 720	0.135 910	0.078 727
2017	0.020 729	0.137 222	0.079 384
2018	0.099 060	0.148 991	0.068 495
2019	0.025 601	0.005 128	0.064 468

随着能源匮乏以及气候问题越来越严重,新能源汽车等绿色产业的发展成为有效解决问题的方式之一,受到广泛关注。2010 年国家开始系统性的出台相关政策推动新能源汽车的发展。蓄电池作为新能源汽车的核心部件,其使用寿命是 5~8 年。

随着新能源汽车销量增加,蓄电池的回收利用成了需要关注的对象,2016—2018 年主题强度显著上升是由于国家出台了 6 项蓄电池回收利用相关政策。主题 18 在 2015 年主题强度最高,是因为国家于 2014 年首次在政策中系统地提出基础设施规划。

3.2.2 中部六省主题强度分析

由表 5 可知,5 新能源汽车发展、10 公共设施建设、20 财政补贴是中部六省热门主题。中部六省主要是从推进公共设施建设和财政补贴两方面推进新能源汽车的发展。例如:建设充电设施智能服务平台,在高速公路服务区、住宅小区等地安装充电设施,在合规的基础之上优先公共充电设施建设用地审批,为新能源汽车建立专门的停车位并在停车费用方面给予一定的优惠。对企业进行补贴,促进其进行技术创新;颁布销售补贴、免税等利好政策,引导消费者购买新能源汽车;并结合省情首先从政府等公共机构做起,对传统燃油公交车的价格补助做出了调整。

表 5 中部六省热门主题词及相关度

相关主题	主题含义		主题含义		主题含义	
	特征词	相关度	特征词	相关度	特征词	相关度
中部六省热门主题	5 新能源汽车发展		10 公共设施建设		20 财政补贴	
	新能源	0.054 586	充电	0.129 870	新能源	0.052 151
	汽车	0.042 547	设施	0.100 466	汽车	0.049 568
	建设	0.036 227	建设	0.050 613	企业	0.038 297
	单位	0.026 696	运营	0.028 971	补贴	0.037 358
	发展	0.025 994	停车场	0.027 329	生产	0.031 723
	责任	0.023 987	公共	0.025 837	销售	0.025 148
	设施	0.013 052	规划	0.020 762	车辆	0.024 444
	企业	0.012 249	管理	0.017 478	推广应用	0.021 861
	工业	0.011 246	新建	0.014 791	产品	0.019 982
	城乡	0.010 745	社会	0.013 597	财政补贴	0.016 930

由表 6 可知,2010 年至 2014 年热门主题强度变化可以分析出中部六省新能源汽车政策的滞后性。主题 5 和主题 20 在 2016 年强度都达到最大,主题 10 在 2016 年和 2017 年增加幅度也较大,是因为新能源汽车在政策引导下迅速发展的同时也受骗补和政策补贴落地问题的影响,国家在 2016 年出台十余项政策进行宏观调控,各省市也积极出台政策推动新能源汽车稳健发展。然而 2019 年 3 月新能源汽车补贴标准大幅调整,地方补贴退出,主题 20 的强度在 2019 年降低到 0.033 272。毋庸置疑,财政补贴的退坡会导致购置成本增加,然而国家政策的引导是为了促进技术的发展。

表 6 中部六省热门主题年份变化表

年份	主题 5 新能源汽车发展	主题 10 公共设施建设	主题 20 财政补贴
2010	0.000 000	0.000 000	0.000 000
2011	0.000 000	0.000 000	0.000 000
2012	0.011 173	0.011 403	0.039 575
2013	0.000 000	0.000 000	0.000 000
2014	0.019 588	0.040 772	0.058 488
2015	0.022 647	0.039 226	0.099 049
2016	0.209 038	0.089 237	0.120 023
2017	0.088 337	0.092 367	0.041 873
2018	0.027 030	0.065 022	0.084 809
2019	0.017 872	0.066 781	0.033 272

4 结论与启示

4.1 结论

本文基于 LDA 主题挖掘模型,对新能源汽车政策文本进行量化分析,得到高频词汇、主题强度以及年度变化趋势。主要得到以下结论:

1)国家层级和中部六省层级的政策中公共基础设施建设主体强度都较高,说明两个层级都认为基础设施的建设与完善是新能源汽车发展的必要条件。只有基础设施的建设与新能源汽车发展“两条腿走路”才会走得更稳健、长远。

2)国家层级与中部六省层级相比而言,国家层级在政策制定时不仅注重新能源汽车的发展,同时积极落实绿色发展理念,而中部六省层级更重视基础保障等具体措施对于新能源汽车发展的推动作用,大力建设充电桩、停车场的基础设施;生产过程的企业技术补贴到销售过程的客户购买补贴;从基础设施建设和财政补贴方面推动新能源汽车发展。

3)总体而言,政策特点与新能源汽车产业政策的要求基本一致,政策体系较为完善,但是层级间

政策存在互动性不足、忽视传导渗透作用等问题。如中部六省的蓄电池回收强度较低,与国家主体之间存在差异,可能导致执行效果不理想。

4.2 启示

通过对国家和中部六省两个层面的政策文本进行量化分析,总结本次主题挖掘的成果,对新能源汽车的未来发展提出以下几点启示:

1)重视新能源汽车产业政策的系统性,加强层级间的沟通能力。一是政策制定时充分研究产业价值链的系统发展,要统筹新能源汽车的研发、生产、购置、使用及回收等环节的系统性政策,而不是针对某一环节的政策。二是要根据地区发展的实际情况,因地制宜地制定产业政策,层级间积极沟通,根据政策实施效果不断进行反馈调整,保证政策有效执行。

2)尊重市场规律,实现“市场+政策”双轮驱动。一是加强政策意见收集,在制定、出台政策前要广泛调研,征求各部门、企业、用户的意见,确保政策执行实施过程的有效性。例如补贴退坡和免征购置税等政策渐变式退出,可以促进技术进步和创新发展,释放市场的竞争力。二是发挥市场的主体地位,鼓励企业创新商业模式,拓宽营销渠道,如发展“融资租赁”、“分时租赁”、“汽车+”等新业态。

3)协调地方政策实施,夯实层级间协同基础。一是各地方要有合作共赢的心态,发挥区位优势,携手推动中部地区新能源汽车产业的发展。中部六省的位置邻近,技术具有外溢性,所以各省可以在蓄电池回收技术、充电设施地址规划等领域加强合作。二是地方产业政策在实施过程加强互动,能够减少上下层级间的协同难度,国家和各省市之间协同推进新能源汽车的发展,能更快提升新能源汽车的竞争力,实现从“汽车大国”到“汽车强国”的飞跃。

参考文献

[1] 国家 863 计划电动汽车重大专项已启动[J]. 汽车研究与开发,2001(5):3.

[2] 中华人民共和国中央人民政府. 中国制造 2025[EB/OL]. (2015—05—19)[2018—05—08]. http://www.gov.cn/zhengce/content/2015-05/19/content_9784.htm.

[3] KITSON A, MARSHALL A, BASSETT K, et al. What are the core elements of patient-centred care? a narrative review and synthesis of the literature from health policy, medicine and nursing [J]. Journal of Advanced Nursing, 2013, 69(1):4—15.

[4] TALAMINI E, DEWES H. The macro-environment for liq-

- uid Biofuels in Brazilian science and public policies[J]. Science and Public Policy, 2012, 39(1): 13—29.
- [5] JANS A J, HANSEN E, GRAY V. Copy and paste lawmaking: legislative professionalism and policy reinvention in the States[J]. American Politics Research, 2019, 47(4): 739—767.
- [6] 曾婧婧. 泛珠三角区域合作政策文本量化分析: 2004—2014 [J]. 中国行政管理, 2015(7): 110—116.
- [7] 黄萃, 任弢, 李江, 等. 责任与利益: 基于政策文献量化分析的中国科技创新政策府际合作关系演进研究[J]. 管理世界, 2015(12): 68—81.
- [8] 苏飞, 刘红. 中国省级政府数据开放政策量化研究[J]. 科技促进发展, 2019, 15(2): 127—132.
- [9] 谭玉, 李明雪, 吴晓旺. 大学生创新创业政策的变迁和支持研究——基于 59 篇大学生创新创业政策文本的分析[J]. 现代教育技术, 2019, 29(5): 112—118.
- [10] 魏萍, 张紫馨, 李青樵. 地方政府融入“一带一路”建设的政策创新扩散研究——基于湖北 294 份政策文本的量化分析[J]. 情报杂志, 2020, 39(4): 82—89, 116.
- [11] 杨慧, 杨建林. 融合 LDA 模型的政策文本量化分析——基于国际气候领域的实证[J]. 现代情报, 2016, 36(5): 71—81.
- [12] 刘颖琦, 王静宇, Kokko Ari. 产业联盟中知识转移、技术创新对中国新能源汽车产业发展的影响[J]. 中国软科学, 2016(5): 1—11.
- [13] 郑征, 霍锋, 王瑜. 基于文献计量的离子液体领域研究现状及发展趋势[J]. 过程工程学报, 2019, 19(5): 890—899.
- [14] 潘丹, 唐静, 杨佳庆, 等. 1978—2018 年中国农村环境管理政策演进特征——基于 206 份政策文本的量化分析[J]. 中国农业大学学报, 2020, 25(6): 210—222.
- [15] 李英, 李惠, 成琪. 基于文献计量和知识图谱的国际绿色车辆路径问题研究发展分析[J]. 中国管理科学, 2016, 24(S1): 206—216.
- [16] 赵佳丽, 程光辉, 周丽丽. “一带一路”视角下明清晋商活动的地理特征统计分析[J]. 经济问题, 2018(5): 116—123.
- [17] 王薇, 刘云. 基于内容分析法的我国新能源汽车产业发展政策分析[J]. 科研管理, 2017, 38(S1): 581—591.
- [18] 谢倩, 王子成, 周明星. 新中国成立 70 年乡村教师支持政策文本量化分析——基于政策工具视角[J]. 现代教育管理, 2020(4): 61—67.
- [19] 张永安, 周怡园. 新能源汽车补贴政策工具挖掘及量化评价[J]. 中国人口·资源与环境, 2017, 27(10): 188—197.
- [20] 朱青, 卫柯臻, 丁兰琳, 等. 基于文本挖掘和自动分类的法院裁判决策支持系统设计[J]. 中国管理科学, 2018, 26(1): 170—178.
- [21] DEERWESTER S, DUMAIS T, LANDAUER G, et al. Indexing by latent semantic analysis[J]. Journal of the American Society of Information Science, 1990, 41(6): 391—407.
- [22] SUEN C Y. N-Gram statistics for natural language understanding and text processing [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1979(2): 164—172
- [23] BLEI DM, NG A Y, JORDAN M I. Latent dirichlet allocation[J]. Journal of Machine Learning Research, 2003, 3(4/5): 993—1022.
- [24] CARSON S, KENNETH E S. LDavis: a method for visualizing and interpreting topics [J]. Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces, 2014: 63—70.
- [25] SIERZCHULA W, BAKKER S, MAAT K, et al. The influence of financial incentives and other socio-economic factors on electric vehicle adoption [J]. Energy Policy, 2014, 68: 183—194.

Quantitative Analysis of New Energy Vehicle Policy Text Based on LDA Model

ZHAO Gong-min, LV Jing-qin, WU Yong-jie

(School of Economics and Management, NUC, Taiyuan 030051, China)

Abstract: Latent Dirichlet Allocation model is used to mine the theme, 144 new energy vehicle related policies at two levels of the state and six provinces in the central region are taken as the research objects. And a quantitative analysis is made in terms of high-frequency vocabulary, theme intensity and annual changes of hot themes. The results show that the national level themes are mainly energy-saving development, power battery recovery, etc.; the themes of the six central provinces are mainly public facilities construction and financial subsidies; the national level promotes the development from the overall perspective of green development and energy conservation and environmental protection, while the central six provinces pay more attention to the promotion of financial subsidies; the new energy vehicle industry policy system is relatively perfect, but the hierarchical politics There are some problems such as lack of interaction and neglect of conduction and penetration.

Key words: theme mining; LDA; new energy vehicles; policy text