

# 2006—2016 年管理科学与工程研究热点主题研究 ——基于 LDA 概率主题模型分析\*

● 陈植元 杨海霞<sup>2</sup> 王先甲<sup>3</sup>

(1, 2, 3 武汉大学经济与管理学院 武汉 430072)

**【摘要】**文章旨在通过对管理科学与工程领域的科技文献(限中文文献)进行统计,运用数据挖掘方法,探讨该领域近十年研究主题的变化规律,从而揭示我国管理科学与工程学科研究发展趋势。以2006—2016年管理科学与工程领域12本期刊中21267篇中文科技文献为样本库,运用概率主题模型LDA对科技文献进行文本挖掘,生成得到20个研究主题,并将样本文献按照主题概率进行分类分析。得到主要结论:5个主题呈现上升趋势,其中主题“运营管理”、“调度交通”、“公司治理”类别下的文献数量占比较大;3个主题呈现下降趋势。文章初步挖掘并分析了管理科学与工程领域的研究主题,有助于管理科学与工程领域的研究者了解该领域的主题热点及变化趋势。

**【关键词】**LDA 文本挖掘 热点主题 文档聚类

中图分类号: G350 文献标识码: A

## 1. 引言

管理科学与工程是一门综合自然科学、工程科学与社会科学的交叉学科,其研究成果对社会、经济、工程等方面的管理问题提供理论与实践上的指导,对社会与经济发展起到重要作用。在管理理念和技术发展迅速的当今社会,管理科学与工程已经成为我国自然科学领域和社会科学领域不可或缺的核心学科。而科研文献作为研究成果的重要载体,不但承担了学术思想交流和传播的功能,也体现了每个阶段的研究主题和热点。因此,对管理科学与工程类文献进行主题研究,可以加深对该学科研究主题的理解和研究热点的追踪,

---

\* 基金项目: 国家自然科学基金“集中与分散决策模式下的随机动态双边匹配策略研究”(项目批准号: 71871166)和“复杂网络上演化博弈合作形成机理与控制策略”(项目批准号: 71871171)。本文为武汉大学自主科研项目(人文社会科学)研究成果,得到“中央高校基本科研业务费专项资金”资助。

通讯作者: 王先甲, E-mail: wangxj@whu.edu.cn。

为科研工作者提供一个俯瞰管理科学与工程领域的视角。

文章以国家自然科学基金委员会管理学部网站上公布的《管理科学重要学术期刊表》中与管理科学与工程相关的12个期刊在2006—2016年期间刊载的21267篇论文为样本,首先对文献的分布情况进行描述性统计,其次借助概率主题模型LDA(Latent Dirichlet Allocation)(Blei et al., 2003; Blei, 2012)对文献的内容进行文本建模,结合困惑度和专家判断选定主题模型的最优主题数量,获得20个研究主题;并根据LDA模型的输出结果进行主题的趋势分析和文献的聚类分析。文章的结果可以帮助科研人员初步了解我国管理科学与工程研究领域近10年研究主题的分布情况;分析研究主题的趋势变化,可以对研究热点的选择提供进一步指导。

## 2. 相关研究现状

为了更好地把握管理科学领域的发展状态,国内已有许多学者进行相关研究。刘作仪和徐贤浩(2009)统计了2003—2008年国家自然科学基金资助的管理科学与工程学科项目资助情况及发表SCI论文情况,并提供了政策建议。张玲玲等(2005)使用文献计量的方法研究了1994—2004年SCI、SSCI收录的117种期刊、国内的1994—2004年77种中文期刊,及国际会议的论文,获得了管理科学与工程在此时间段内的研究热点。缪园等(2007)以管理科学与工程学科代码为基础,收集国家自然科学基金资助的项目作为样本数据,构造评价模型,对管理科学与工程的研究热点进行非线性评价,并对发展趋势进行预测。刘作仪等(2012)使用2001—2010年Web of Science数据库收集的4262篇论文,运用文献计量的方法,分析了管理与运筹学领域的研究势态和热点。陈传明和刘海建(2009)分析了2005—2006年间CSSCI中管理学论文的关键词,分析研究热点与趋势。李华峰和袁勤俭(2017)研究了2006—2015年国家自然科学基金对管理科学的资助情况。国内也有学者对管理科学与工程领域的一些新兴的主题及研究热点进行分析,如行为运作管理(刘作仪和查勇,2009)、港口管理与运营(王帆等,2017),医疗运作管理(杜少甫等2013)。上述研究主要是通过国家自然科学基金立项数据,或者是通过关键词,或者是通过文献数量统计的方法,少有研究使用基于文本挖掘技术的主题模型方法来分析科技文献。随着时间跨度的增加和期刊数量的增加,手工统计文献数量和关键词聚类分析的方法已然无法胜任。

本文使用的LDA模型由Blei等人在2003年提出,是一种概率主题模型。它可以将文档集合中每篇文档的主题按照概率分布的形式给出。LDA模型也是一种无监督机器学习算法,在训练的时候不需要手工标注训练集,仅仅是文档集合与指定主题的数量即可。而LDA的最大优点,就是对于每个主题均可找出一些词语来描述它。2004年Griffiths和Steyvers(2004)应用LDA方法分析PNAS(Proceedings of the National Academy of Sciences)期刊文献摘要的主题及主题变化趋势。随后,LDA模型帮助多个领域的研究者提取领域内主题或进行主题分析,如计算语言学领域(Hall et al., 2008; 杨海霞等2016)、文献计量学领域(Wu et al., 2010)、图书情报管理领域(关鹏等,2016; 王曰芬等,2016)、会计学领域(Fang et al., 2018)、经济学领域(Piepenbrink &

Nurmammadov, 2015) 等。

以上研究均表明，LDA 模型在挖掘文献主题方面效果较好。故笔者借用概率主题模型 LDA，对我国 2006—2016 年管理科学与工程领域 12 本期刊刊载的 21267 论文进行分析，从而掌握近十年来该领域的研究热点及其变化趋势。

### 3. 数据与实验

#### 3.1 数据来源

文章根据国家自然科学基金委员会管理学部网站上公布的《管理科学重要学术期刊表》为基准，选取 A 类期刊中与管理科学与工程领域紧密相关的 12 个期刊（其中排名不分先后：《系统工程理论与实践》、《管理世界》、《中国管理科学》、《系统工程学报》、《管理评论》、《管理工程学报》、《南开管理评论》、《管理科学》、《预测》、《管理科学学报》、《系统科学学报》和《运筹与管理》，获取期刊文献。所有数据均来源于中国知网的文献输出功能，时间跨度为 2006—2016 年共计 11 年。对收集到的文献集合进行整理：首先删除新闻、征文通知等非科技文献，其次删除标题、关键词、摘要不完整的文献，最后获得 21267 条文献记录。经过数据整理后，最后选取的期刊及篇数按年份统计如表 1 所示。

表 1 论文数据按年份综述表

年份 期刊	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
《系统工程理论与实践》	256	304	292	286	301	354	349	401	412	356	321
《管理世界》	198	310	316	326	288	275	266	254	250	257	223
《中国管理科学》	147	136	300	156	157	146	291	265	345	379	239
《系统工程学报》	115	111	122	121	130	115	104	96	87	85	82
《管理评论》	123	113	112	197	193	227	247	217	222	249	276
《管理工程学报》	127	124	124	150	126	138	105	113	111	110	114
《南开管理评论》	101	99	92	119	106	101	98	97	97	92	103
《管理科学》	82	76	86	79	77	73	69	64	73	69	74
《预测》	91	85	82	82	82	86	82	85	81	78	72
《管理科学学报》	71	64	97	91	122	105	106	94	99	99	114
《运筹与管理》	190	186	174	172	173	185	219	214	212	223	231
《系统管理学报》	110	149	122	124	111	110	115	133	121	118	138

根据表 1，我们可以对数据进行描述性统计：首先进行发文量的统计。2006—2016 年

时间段，样本库中各个期刊的管理科学与工程类文献占比如图 1 所示。占比较大的依次是《系统工程理论与实践》(17%)、《管理世界》(15%)、中国管理科学(12%)、管理评论(10%)、运筹与管理(10%)。这五个期刊的发文数量相对其他期刊较大。而其他期刊发文量相对较少的部分原因是双月刊(如《管理科学》和《预测》)或单期刊文量较少(如《管理科学学报》)的缘故。

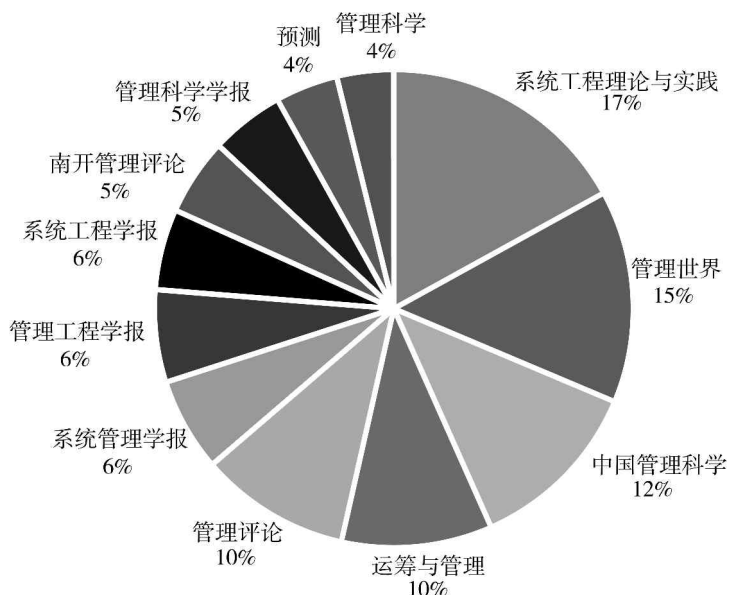


图 1 2006—2016 年各个期刊的发文总量占比

其次，我们按照年度发文数量统计出于 2006—2016 年的变化情况如图 2 所示。我们可以看到，发文总量的趋势是逐年上涨，2008 年和 2012 年达到区域性峰值，而 2010 年总发文量有所回落。2016 年发文数量有所下降是因为截止论文撰写之时，有些期刊论文还未被知网检索导致。

### 3.2 文献数据预处理与 LDA 的参数设置

此处对 LDA 模型进行简单介绍。LDA 模型的原理即：一篇文档中的每个词出现的原因，都是作者通过概率选择了主题，并从该主题中以概率选择了该词。因此在某文档中出现该词的概率为：

$$\Pr(\text{词语} | \text{文档}) = \sum_{\text{主题}} \Pr(\text{词语} | \text{主题}) \cdot \Pr(\text{主题} | \text{文档}) \quad (1)$$

其中最左边的概率是使用每个词语出现的频数进行估计，而右边  $\Pr(\text{主题} | \text{文档})$  矩阵将是下文的表 2；而表 3 将展示估计出的  $\Pr(\text{词语} | \text{主题})$  矩阵。LDA 模型就是估计出公式 (1) 右侧的两个矩阵，接下来我们阐述具体的实现步骤。

首先将每篇文献的中文标题、中文关键词和中文摘要分别合并，从而得到 21267 个文档。我们将此 21267 篇文档定义为研究样本库。由于研究样本是中文文献，故首先需要将

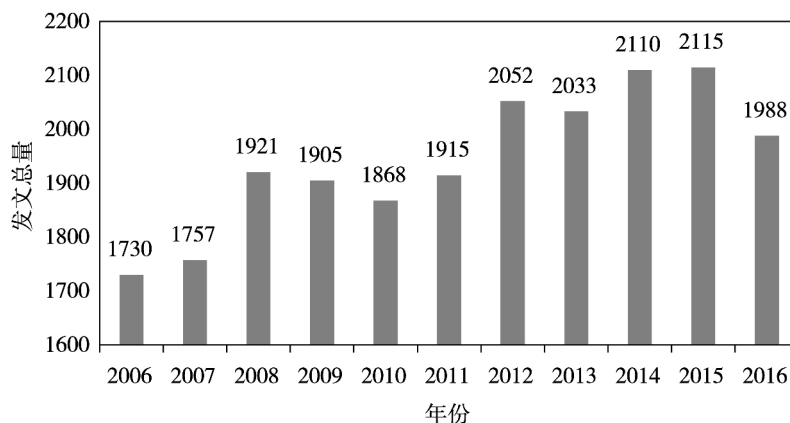


图2 2006—2016年所列12本期刊年度发文总量变化趋势

中文文献的文本进行分词，如将“资本市场的暴涨暴跌挑动投资者的敏感神经，造成市场极端波动和危机传染”进行中文分词后，得到“资本市场-的-暴涨-暴跌-挑动-投资者-的-敏感神经-，-造成-市场-极端-波动-和-危机-传染”。而在中文文本中，一段话的不同分割将会产生不同的词语含义，故让机器能自动识别语义并进行分割也是本文的困难所在。为了得到较好的分词效果，笔者借助R语言中的RWordseg包，通过人工补充词库的方法进行循环调整，从而构建了管科科学与工程领域的文献专业术语词库。

具体步骤为：(1) 建立专家词库，共3072个词（供应链、纳什均衡，等等）；(2) 在基础停顿词1208个词的基础上，增加文献分析特有的停顿词，形成一个共5621个词的停顿词库（本文，文中，文章，表明，等等）；(3) 借助R语言的Rwordseg包，对文本进行分词，并去除停用词，得到21267行分词后的文档；(4) 借助R语言的tm包将文本语料化，形成21267个文档词矩阵。

在得到文档-词矩阵后，借助R软件自带的topicmodels包，运行LDA模型。在使用该模型前，我们需要确定模型的最优主题数。根据topicmodels包中推荐参数，设置文档-主题分布 $\theta$ 的参数 $\alpha = 0.1$ ，主题-词分布 $\phi$ 的参数设置为 $\beta = 0.1$ ，迭代次数为2000次，并且选择Gibbs采样的方法估计模型的后验参数。我们将主题数 $K$ 从5~50依次进行试验，发现 $K$ 在15~25之间，主题的可识别度较高。最后结合专家判断，发现主题数为20的主题模型，能较好地涵盖管理科学与工程领域的研究内容。

LDA模型的运行结果主要会输出两个矩阵：首先是26217篇文献的主题分布 $\theta_{ij}$ 和20个主题的词项分布 $\varphi_{j,v}$ 。其中 $\theta_{ij}$ 表示文献 $i$ 中主题 $j$ 的概率， $\varphi_{j,v}$ 表示主题 $j$ 中词 $v$ 的概率。

### 3.3 实验结果与分析

#### 3.3.1 文献-主题分布与主题-词分布

根据LDA模型的实验结果，得到了21267篇文献的主题分布 $\theta_{ij}$ 和20个主题的词项分布 $\varphi_{j,v}$ ，如表2和表3所示。例如 $\theta_{1,2} = 0.0242$ 表示文献1属于主题2的概率是0.0242。根据表1的结果，我们可以进行主题强度分析，也可以对文档进行分类；根据表3的高概

率主题词，我们可以给每个主题进行命名，同时进行分析。

表 2 管理科学与工程类文献的主题分布

$\theta_{ij}$	1	2	3	4	...	16	17	18	19	20
1	0.0242	0.0242	0.0825	0.0339	...	0.0242	0.1116	0.0242	0.0242	0.0242
2	0.0263	0.0263	0.0263	0.0473	...	0.0578	0.0789	0.0263	0.0578	0.0263
3	0.0268	0.0376	0.0483	0.0268	...	0.0268	0.2741	0.0376	0.0268	0.0483
...	...	...	...	...	...	...	...	...	...	...
21265	0.0925	0.0259	0.0185	0.0185	...	0.0185	0.0259	0.0555	0.0259	0.0185
21266	0.0324	0.0324	0.0324	0.0324	...	0.1103	0.0454	0.0454	0.0324	0.0324
21267	0.0294	0.0294	0.0294	0.0294	...	0.0529	0.0294	0.0294	0.0294	0.1235

表 3 管理科学与工程类文献主题的高频词

主题名称	主题相关的 8 个高频词							
组织行为	绩效	组织	关系	员工	团队	领导	公平	实证
管理案例	管理	案例	中国	文化	持续	发展	借鉴	国际
网络分析	网络	结构	演化	仿真	集群	特征	复杂	节点
宏观经济	中国	区域	地区	城市	经济增长	收入	差异	农村
战略管理	企业	投资	战略	创业	企业绩效	制造	家族	关系
公司治理	公司	上市	显著	股权	公司治理	控制	样本	并购
数理统计	模型	随机	参数	分布	概率	线性	数值	贝叶斯
投资交易	交易	市场	波动	投资者	资产	股票	回归	基金
市场营销	服务	质量	顾客	价值	品牌	购买	营销	意愿
应急对策	动态	模型	应急	安全	政策	突发事件	周期	危机
运营管理	供应链	需求	零售商	决策	供应商	契约	协调	库存
知识管理	能力	知识	关系	协同	联盟	整合	外包	生态
公共政策	环境	社会	政府	治理	公共政策	污染	监管	保护
调度交通	算法	优化	规划	约束	路径	调度	遗传算法	运输
系统仿真	系统	预测	控制	数据	交通	仿真	建模	可靠性
决策理论	评价	决策	指标	模糊	属性	聚类	排序	偏好
金融风险	风险	银行	信用	金融	Var	资金	贷款	Copula

续表

主题名称	主题相关的 8 个高频词							
数据包络	技术	效率	产业	投入	要素	DEA	能源	行业
博弈论	策略	市场	博弈	竞争	价格	均衡	定价	纳什均衡
机制设计	信息	机制	激励	拍卖	效用	委托	契约	谈判

表 3 列出了各个主题的前 8 个高概率主题词。观察主题的此项分布情况，发现主题内部的词高度相关。例如，主题“运营管理”中高频词：供应链、需求、零售商、决策、供应商、契约和库存，均与“运营管理”紧密相关。而“供应链”的频率排序在“库存”之前，也体现了如今对供应链管理的研究已经逐渐从库存管理转入供应链管理的趋势。当然这些词之间的概率差异也非常小，排序也仅仅起一种参考价值。但这也从一方面表明 LDA 模型在提取管理科学与工程类文献的潜在主题方面是有效的。

### 3.3.2 主题强度分析

主题强度是描述主题在某一段时间内的热门程度，本文用  $\theta$  表示。在某一时间段内某个主题的文献数量越多，说明该主题的主题强度越高，即可以认为是热点主题。本文以时间“年份”作为变量，将文档-主题分布  $\theta$  按年计算，得出各个主题的主题强度分布情况。根据主题每年的强度大小，构建自回归模型，找出主题强度上升的研究主题及主题强度下降的研究主题。发现在 99% 的置信水平下，20 个主题中有 5 个研究主题呈现上升趋势（见图 3），3 个研究主题呈现下降趋势（见图 4）；其余的 12 个研究主题的趋势变化不明显。

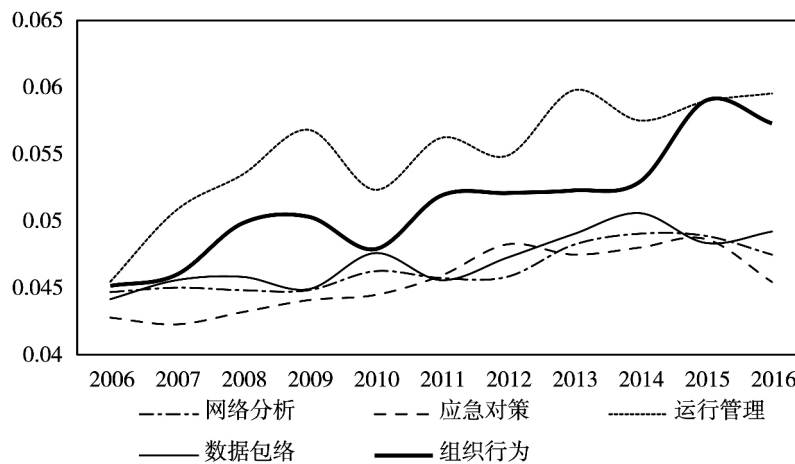


图 3 呈上升趋势的主题

根据观察上升的研究主题，可以解释为管理科学随着时间的推移，人们越来越注重量化模型在实际背景中的应用。主题“数理统计”热度下降的原因是管理科学领域越来越注重实践，而纯粹的统计优化和运筹理论研究热度降低；系统仿真的热度降低，可以被认

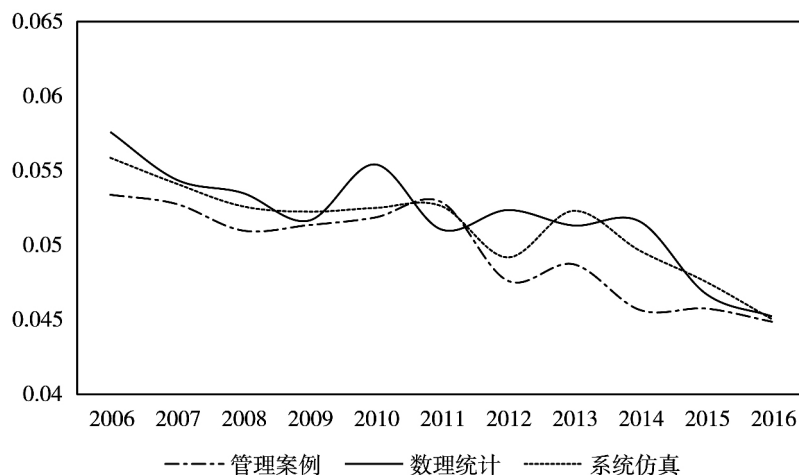


图4 呈下降趋势的主题

为是这些期刊对量化理论的结果需求量在增加；而“管理案例”的发文数量降低，本文对此的解释是，由于此类文章对内容的理论与实践结合要求非常高，其稿件数量也有下降的趋势，而这方面的稿件数量的降低也被各期刊的编辑所洞察。此处需要注意，主题强度的下降并不表明该主题被研究者所忽视，而只是在统计意义上，主题文献出现的热度相比较而言，有下降的趋势。为了更进一步细化观察不同时期的研究主题强度差异，我们将2006—2016年划分为两个时间窗口：2006—2010年与2011—2016年。表4是不同时间窗口期下的热点主题及其强度。

表4 不同时期的高强度主题

2006—2016		2006—2010		2011—2015	
主题	强度	主题	强度	主题	强度
调度交通	0.0582	调度交通	0.0593	运营管理	0.0578
运营管理	0.0550	决策理论	0.0548	调度交通	0.0574
决策理论	0.0538	数理统计	0.0545	组织行为	0.0542
公司治理	0.0530	公司治理	0.0541	决策理论	0.0529
宏观经济	0.0524	宏观经济	0.0531	公司治理	0.0520

其中，我们可以看到研究主题“调度交通”在2006—2016年一直是管理科学与工程领域研究较为热门的主题，依次为“运营管理”、“决策理论”、“公司治理”、“宏观经济”。可以看到随着时间的推移组织行为的研究也逐渐成为管理类期刊中的热点。

### 3.3.3 主题论文数量统计

对每篇文档的主题概率进行排序，按照每篇文档的最大概率进行主题归类，即：若文献*i*中主题*j*的概率最大，那么将该文献归类为主题*j*的类别下。按照此规则，每篇文章



都可以找到最接近的主题。表 5 是统计结果，我们可以看到“运营管理”、“调度交通”、“公司治理”是占比前三的文献。

表 5 主题论文数量统计

主题	文献篇数	百分比
组织行为	1142	5.37
管理案例	821	3.86
网络分析	771	3.63
宏观经济	1373	6.46
战略管理	913	4.29
公司治理	1506	7.08
数理统计	878	4.13
投资交易	1331	6.26
市场营销	1093	5.14
应急对策	609	2.86
运营管理	1760	8.28
知识管理	759	3.57
公共政策	728	3.42
调度交通	1683	7.91
系统仿真	1178	5.54
决策理论	1276	6.00
金融风险	972	4.57
数据包络	978	4.60
博弈论	780	3.66
机制设计	716	3.37
合计	21267	100

其中，“运营管理”占比 8.28%，“调度交通”占比 7.91%，“公司治理”占比 7.08%。而占比最少的是“应急对策”（占比 2.86%）。可见“运营管理”发文数量是“应急对策”发文数量的将近 3 倍。表 5 从数量角度展示了研究主题之间的热度关系。

#### 4. 结语

文章结合描述性统计分析和文本挖掘方法对管理科学与工程领域的发文分布和研究主

题及研究趋势进行相应分析,结合模型困惑度和专家经验判断确定 LDA 模型的最优主题数目;针对从中国知网下载的 21267 篇管理科学与工程文献挖掘识别出 20 个主题,通过自回归模型分析,得到 5 个趋势上升的研究主题和 3 个趋势下降的研究主题;对每篇文献,根据文献的主题概率值,找出各个文档的最高概率主题,对文档进行类别划分,了解每个主题的发文数量的占比。

结果表明,管理科学与工程类期刊的发文趋势可以归纳为以下三点:首先,研究主题热点朝着量化理论分析及应用背景紧密结合的方向发展;其次,组织行为的研究成为热点;最后,数据包络分析、网络分析和应急对策分析三个主题是十年来上升趋势较快的新兴主题。

当然,本文的研究依然存在不足之处:(1) 本文所选取的期刊仅是国家自然科学基金委列举的重点期刊,并没有考虑国内管理科学与工程领域学者在其他相关类别的优秀期刊及国外优秀期刊上发表的论文文献。对此,我们只是为该研究提供了一种思路,未来的研究可以进一步扩充样本容量,进行主题分析,从而更准确地了解我国国内管理科学与工程领域的发展状态;(2) LDA 文本挖掘模型对中文分词识别能力的缺陷无法完全去除,中文分词的准确性影响文本建模的效果。(3) 本文仅仅从时间角度对文献的热度和趋势进行分析,而实际可以将论文中的作者数量、作者单位、作者受资助情况、作者地域等更多的外部信息融入研究中,或许可以得到更丰富、有趣的结论。

## ◎ 参考文献

- [1] 陈传明, 刘海建. 2005—2006 年我国管理学的研究热点——基于 CSSCI 关键词的分析 [J]. 管理学报, 2009 (2).
- [2] 杜少甫, 谢金贵, 刘作仪. 医疗运作管理: 新兴研究热点及其进展 [J]. 管理科学学报, 2013 (8).
- [3] 关鹏, 王曰芬, 傅柱. 不同语料下基于 LDA 主题模型的科学文献主题抽取效果分析 [J]. 图书情报工作, 2016, 60 (2).
- [4] 李华锋, 袁勤俭. 基于 2006—2015 年 NSFC 立项数据的管理科学资助分析 [J]. 科技管理研究, 2017 (6).
- [5] 刘作仪, 吴登生, 李建平. 2001—2010 年我国管理与运筹学研究态势的计量分析——基于 Web of Science 数据. 北京理工大学学报: 社会科学版, 2012, 14 (1).
- [6] 刘作仪, 徐贤浩. 管理科学与工程学科基金项目资助和 SCI 论文发表情况分析 [J]. 管理学报, 2009 (8).
- [7] 刘作仪, 查勇. 行为运作管理: 一个正在显现的研究领域 [J]. 管理科学学报, 2009 (4).
- [8] 缪园, 张伟倩, 李媛. 国内管理科学与工程研究热点以及发展趋势——近年国家自然科学基金资助项目的非线性分析 [J]. 科学学与科学技术管理, 2007 (10).
- [9] 王帆, 黄锦佳, 刘作仪. 港口管理与运营: 新兴研究热点及其进展 [J]. 管理科学学报, 2017 (5).

- [10] 王曰芬, 傅柱, 陈必坤. 采用 LDA 主题模型的国内知识流研究结构探讨: 以学科分类主题抽取为视角 [J]. 现代图书情报技术, 2016 (4).
- [11] 杨海霞, 高宝俊, 孙含林. 基于 LDA 挖掘计算机科学文献的研究主题 [J]. 现代图书情报技术, 2016, 32 (11).
- [12] 张玲玲, 房勇, 杨涛, 张超, 李若筠, 刘作仪, 杨晓光, 汪寿阳. 管理科学与工程热点研究领域的文献计量分析 [J]. 管理学报, 2005, 2 (4).
- [13] Blei, D. M., Ng, A. Y., Jordan, M. I. Latent Dirichlet allocation [J]. *Journal of Machine Learning Research*, 2003 (3).
- [14] Blei, D. M. Probabilistic topic models [J]. *Communications of the ACM*, 2012, 55 (4).
- [15] Griffiths, T. L., Steyvers, M. Finding scientific topics [J]. *Proceedings of the National Academy of Sciences*, 2004, 101 (1).
- [16] Piepenbrink, A., Nurmammadov, E. Topics in the literature of transition economies and emerging markets [J]. *Scientometrics*, 2015, 2012 (3).
- [17] Fang, D., Yang, H., Gao, B., et al. Discovering research topics from library electronic references using latent Dirichlet allocation [J]. *Library Hi Tech*, 2018, 36 (3).
- [18] Wu, H., Wang, M., Feng, J. et al. Research topic evolution in “bioinformatics” [C]. In: Proceedings of the 4th international conference on bioinformatics and biomedical engineering (ICBBE). IEEE, 2010.

### **Research Topic Evolution of Management Science and Engineering over 2006—2016 with LDA Probabilistic Topic Model**

Chen Zhiyuan<sup>1</sup> Yang Haixia<sup>2</sup> Wang Xianjia<sup>3</sup>

(1, 2, 3 Economics and Management School of Wuhan University, Wuhan, 430072)

**Abstract:** This paper uses Probabilistic Latent Dirichlet Allocation (LDA) topic model to show the research topic evolution in the field of management science and engineering over the last 11 years. The data consists of 21267 papers from 12 top journals in the area of management science and engineering. The main results of LDA topic model are as follows. First, we find 20 topics is an appropriate number of topics to cover this research field. Second, we find 5 topics have increasing trend and 3 topics have decreasing trend. We hope our findings could help the scholars to overview the field of management science and engineering in a bird’s view.

**Key words:** Latent Dirichlet Allocation (LDA); Text mining; Hot topic; Document cluster

专业主编: 许明辉