

文章编号:1003-207(2014)zk-0056-07

# 管理科学与工程领域的创新轨迹研究

——基于 TOP 期刊的文献计量和文本挖掘视角

李思志<sup>1</sup>, 李佳骏<sup>2</sup>, 李艳红<sup>2</sup>

(1. 上海财经大学商学院, 上海 200433; 2. 上海财经大学信息管理与工程学院, 上海 200433)

**摘要:**以管理科学与工程领域 2006—2013 年在 10 种国际顶级期刊上发表的 5779 篇文章为基础, 通过引文分析和共引分析两种角度, 结合特征提取方法, 剖析了管理科学与工程领域研究的科研产出分布情况, 并通过知识图谱的方法刻画和探索了管理科学与工程研究领域的创新轨迹。

**关键词:**特征抽取; 引文分析; 共引分析; 知识图谱; 创新轨迹

**中图分类号:**TP391.1      **文献标识码:**A

## 1 引言

管理科学与工程是一门研究解决社会、经济、工程等方面的管理问题, 具有丰富内涵和广泛外延的综合性学科。刻画和探索管理科学与工程研究领域的创新轨迹, 对于更好地把握学科发展趋势, 探索、归纳、甚至预测该领域的热点问题, 更好地分配科研能力, 推进管理科学与工程学科的发展具有重要意义。

管理科学与工程学科发展轨迹问题与管理科学与工程热点问题存在较为密切的关系, 轨迹绘制的一般方法是在热点发现的基础上对整理出的热点问题按时间先后进行排序。国内学者李群霞、张群<sup>[2]</sup>统计了 1994—2004 年国内外管理科学与工程学科各研究领域发表论文的数量, 并对各领域的研究内容、热点、发展趋势进行了分析和比较, 对比了国内外研究情况的异同。张玲玲等<sup>[3]</sup>以国外 1994—2004 年被 SCI、SSCI 收录的 117 种期刊、国内 1994—2003 年 77 种中文期刊、ISTP 国际会议索引数据库 2000—2004 年会议收录中发表的管理科学与工程领域的论文作为基础数据源, 利用文献计量分析方法分析了国内外管理科学与工程学科各领域发展的历史、现状以及未来的研究热点, 并重点就

国内外热点的形成原因以及热点内容的异同进行了深入研究。学者王虎<sup>[4]</sup>对近十年来在各个领域发表的文章进行了关键词的检索, 并通过对这些检索的文献进行阅读和比较, 研究了在这十年的时间内国内外的管理科学与工程研究的基本现状和特征, 并对国内外研究的热点问题进行了比较分析。尚未见到国外相关文献。现有管理科学与工程学科发展轨迹研究的相关工作的特点和局限是: 具有少量十年前的热点分析工作, 近年的极少; 聚焦热点发现和趋势研究, 未侧重轨迹绘制和分析; 从管理科学与工程学科各研究领域出发, 分析细致, 同时也受限于领域划分。使用文献计量学的方法, 研究手段单一, 没有兼顾文本挖掘的方法。

文本挖掘作为数据挖掘领域的热点与前沿, 在处理非结构化文本方面具有得天独厚的优势。国外学者 Pons—Porrata<sup>[6]</sup>、Tseng<sup>[7]</sup>、Guo Ying<sup>[8]</sup> 等人运用文本挖掘方法在标题发现、专利分析、创新路径预测方面, 意味着将文本挖掘方法运用于学科科研热点发现和发展轨迹描绘在理论上可行。我国学者章以金<sup>[9]</sup>、彭华民<sup>[10]</sup>等综合运用文本挖掘和文献计量学方法研究学科科研热点和发展轨迹, 在信息管理与信息系统、中国社会福利等领域进行了成功的实践, 这种思路虽然尚未运用到管理科学与工程领域的创新轨迹研究, 但为本文的研究开拓了视野。

本文在弥补当前研究工作不足的基础上, 综合运用文本挖掘与文献计量的方法, 以管理科学与工程领域 2006—2013 年在 10 种国际顶级期刊上发表的 5779 篇文章为基础, 对管理科学与工程领域的热点问题挖掘, 并将其按时间顺序排列, 以期对该

**收稿日期:**2014-06-30; **修订日期:**2014-07-28

**基金项目:**上海市自然科学基金资助项目(13ZR1413400); 上海市金融信息技术研究重点实验室开放课题资助项目(2013110933); 上海财经大学‘211 工程’四期重点学科建设项目资助项目

**作者简介:**李思志(1978—), 女(汉族), 黑龙江人, 上海财经大学商学院讲师, 博士, 研究方向: 管理科学、决策支持。

领域近 8 年的创新轨迹进行绘制和分析。

## 2 相关知识和工具

(1)引文分析 是利用各种数学方法和统计学的方法进行比较、归纳、抽象和概括,对科技期刊、科技文献及其所包含的著者信息、机构信息、国家来源等文献记录间的引用和被引用关系进行分析,以揭示其数量特征和内在规律的一种计量方法。目前,引文分析理论已经在图书情报领域、科学计量学领域广泛应用,主要是因为引文信息易于获得,引文分析方法在使用上限制极少,对研究者的先决条件要求也极少。

(2)共引分析 是指两篇文献同时作为另外一篇文献的参考文献这一特殊情况。具有同引关系的文献在研究主题上有一定程度的相关性,因此,可根据共引频次或共引强度来测度文献在研究内容方面的相似度,并通过这种联系形成文献的共引网络,结合图论知识,文献共引网络中节点所处的位置就反映了研究内容上的相似程度。

(3)特征抽取 文本的特征项的选取是文本挖掘的一个基本问题,它从文本中抽取出的特征词进行量化用以代表文本信息。从非结构的原始文本转化为结构化的计算机可以识别处理的信息,需要对文本内容进行科学的抽象,并建立相关数学模型,用以描述和代替文本,从而使计算机能够通过通过对这种模型的计算和操作来实现对文本的识别。

(4)Citespace 是一款基于 JAVA 平台,应用于科学文献识别并显示科学发展新趋势和新动态的软件,由美国德雷塞尔大学信息可视化专家陈超美博士编写。Citespace 软件以共引文献为基础,将某一研究领域的科技文献及其参考文献定义为一种映射关系,同时引入了 pathfinder 算法<sup>[5]</sup>,优于以往所采用的精简网络的方法。利用 CiteSpace 可以找到某一科学领域的研究进展和当前的研究前沿,及其对应的知识基础。

(5)文本挖掘工具 WinPython 是由 Python (x,y)的作者开发的基于 Python 的集成开发环境,其中内置了 numpy 和 spyder 等科学计算包,功能比较全面。为了能够实现文本特征提取这一功能,在此基础上引入了 nltk 包用以处理自然语言,引入了 gensim 包用以实现 TF-IDF 技术。

(6)知识图谱 这是一种通过描述知识资源来反映对象发展历程和结构关系,从而生成可视化的图形的科研方法。知识图谱,将图论、数学、信息学

等理论与可视化技术相融合,将对象发展背后所隐藏的时间顺序和复杂的知识关联通过可视图形展现出来。

## 3 研究数据和方法

学者梅世强<sup>[1]</sup>提出“管理科学与工程是为了实现管理目标,应用工程技术科学、数学、系统科学及社会科学知识,对人员、物资、设备、技术、能源和信息等组成的各种系统进行设计、评价、决策、改进、实施和控制的一门学科。”从这一定义出发,结合国内外相关研究文献,我们认为国外学科分类中的 MS/OR 领域与我国所定义的管理科学与工程领域口径基本吻合。

本文的数据均来源于 Web of Science 网站。Web of Science 是美国信息技术情报所 (ISI) 成立的大型综合性期刊引文索引数据库,由美国 Thomson Scientific (汤姆森科技信息集团) 基于 WEB 开发,共包括 8,000 多种世界范围内最有影响力的、经过同行专家评审的高质量的期刊。我们在 2006—2013 年 ISI 提供的期刊引证报告 (JCR) 中抽取 2012 年报告中 MS/OR 方向的排名前 10 的期刊作为本次研究的数据源,它们分别是 Management Science、Journal of Operations Management、Mathematical Programming、Manufacturing & Service Operations Management、Operations Research、Mathematics of Operations Research、Transportation Science、Transportation Research Part B、INFORMS Journal on Computing 和 Production and Operations Management。这些期刊被公认为是管理科学与工程领域的高水平期刊,拥有较高的学术影响力,基本可以展现管理科学与工程近年来的研究方向,较充分体现出国际管理科学与工程领域的主要研究内容。对于 Manufacturing & Service Operations Management、Transportation Research Part B 两本刊名与网站检索规范相冲突的期刊,采用该期刊的数据对象唯一标示符 (DOI) 作为检索的关键词,其他 8 部期刊以刊名作为检索关键词。

我们利用 Citespace 从 Web of Science 下载了所选期刊研究时间段内 (2006—2013 年) 的全部文献记录 5779 条。每条文献记录包括文献标题、摘要、关键词、作者来源、出版物、地址与引用文献等信息。

本文的研究思路和方法是:以 2006—2013 年发表在国际管理科学与工程 TOP 期刊上的学术论文为研究数据,依据引文分析的理论,对文献的科研产

出进行统计分析。利用文献计量学中的共引分析方法建立引文网络,形成文献记录间的拓扑结构,通过聚类分析方法,把分析对象之间错综复杂的共引网状关系简化为数目相对较少的若干类群之间的关系并表示出来,在此基础上分析研究对象所代表的学科及文献的结构和特点,来获知学科的热点领域和研究方向。最后,选取知识图谱的方法呈现研究路径。

#### 4 研究过程

##### 4.1 科研产出分析

###### 4.1.1 高产作者分析

表 1 列出了前 9 名作者的产量,研究时段的最早发表年份和通过文本分析识别关键词生成的研究热点系数(Burst)。在收集的数据来看,发表论文最多的是 Gilbert Laporte,2006—2013 年间发表论文 29 篇;在产量排名前 10 位的作者中唯一具有比较高的研究热点系数的是 Aleda V. Roth,这表明该作者在 2006—2013 年间对管理科学与工程领域的热点问题研究做出了突出的贡献。

表 1 高产作者统计表

产量	热点	作者	最早发表年份
29		Gilbert Laporte	2006
26		Ward Whitt	2007
26		Hai Yang	2010
25		Dimitris Bertsimas	2010
22		Carlos F. Daganzo	2006
21		Yanfeng Ouyang	2010
21		David SimchiLevi	2006
19	3.05	Aleda V. Roth	2007
19		Milind Dawande	2007

###### 4.1.2 高产研究机构分析

研究文献产量排名前 18 位的研究机构如表 2 所示。明显看出,美国的研究机构占绝对优势,并包揽前 10 位,其他研究机构仅有来自中国香港的一所院校位列榜上。从表中数据可以发现前 18 名的机构全部位于英语普及、现代工业和服务业发达的地区。说明在管理科学与工程领域的顶级国际期刊上发表研究成果,语言的优势是重要的一方面。数据还显示,管理科学与工程的研究机构多为高等院校,企业及其他单位占少数,说明高等学校是管理科学与工程学术研究的主要阵地。

表 2 高产机构统计表

产量	机构	起始年份	产量 2	机构 2	起始年份 2
214	MIT	2006	107	Carnegie Mellon	2006
185	Georgia Inst	2006	106	Univ Illinois	2006
174	Northwestern Univ	2006	101	Cornell Univ	2006
169	Columbia Univ	2006	96	Univ Minnesota	2006
168	Univ Calif Berkeley	2006	95	NYU	2006
147	Stanford Univ	2006	86	Univ N Carolina	2008
120	Univ Michigan	2006	85	Univ Texas Austin	2008
118	Univ Penn	2006	85	Hong Kong Univ Sci & Technol	2006
116	Duke Univ	2006	84	Penn State Univ	2006

###### 4.1.3 高产国家或地区

各个国家或地区在管理科学与工程领域的研究状况如表 3 所示。美国的科研产出占绝对优势地位,中国的产量位居第三,整体数量较为领先。

排名前 18 位的大多数是经济和教育基础较好,现代工业和服务业发达的国家和地区,表明这些地区是近些年管理科学与工程领域发展和研究的前沿阵地。

表 3 高产地区统计表

产量	国家/地区	起始年份	产量 2	国家地区/2	起始年份 2
3633	USA	2006	120	ISRAEL	2006
474	canada	2006	110	AUSTRALIA	2006
402	PEOPLES R CHINA	2006	100	BELGIUM	2006
282	FRANCE	2006	87	SWITZERLAND	2006
270	ENGLAND	2006	77	TURKEY	2006
234	GERMANY	2006	75	JAPAN	2006
219	NETHERLANDS	2006	67	CHILE	2006
189	ITALY	2006	55	INDIA	2006
145	SINGAPORE	2006	49	SOUTH KOREA	2006

### 4.2 共引网络聚类分析

利用 CiteSpace 软件对文献进行共被引分析,采用一种适于多元、分时、动态的复杂网络分析的知识图谱技术,能够对文献进行关键词共现分析、期刊引用分析、作者共引分析、引文共引分析。分析设置中,以文章的参考文献作为节点构建了共引网络,时间单位(time slice)选择为 1 年,并运用 pathfinder 算法修剪网络后,对各节点进行聚类。

从本次聚类的结果来看,共聚为 11 类。指标 Modularity Q 为 0.6774 表明本次操作比较合理地将数据聚成了 11 个松散耦合的类别;而指标 Mean Silhouette 为 0.9034 则表明各聚成的类中的均匀性(homogeneity)非常好。综合两项结果来看本次聚类是有效的。在聚类的过程中,Citespace 生成了各聚类中的高频关键词,如表 4 所示。

表 4 各类主要关键词

编号	(热点指数)主要关键词
#0	(5.59) global facility network; (5.59) risk-pooling; (5.59) global operations management; (5.59) assemble-to-prdr systems; (5.59) arrival sequence
#1	(5.59) b-spline; (5.26) nonparametric estimation; (5.26) mixed logit; (4.24) mixed logit-models; (4.24) stated choice
#2	(3.27) operating room; (2.64) block time; (2.64) health care operations; (2.64) operating-room time; (2.64) surgical scheduling
#3	(7.88) preference reversals; (7.88) self-regulation; (7.88) production lines; (7.88) dynamic decision-making; (6.93) random choice
#4	(11.02) investor; (9.11) weighs; (9.11) attention; (8.73) usergenerated content; (8.27) stock-market
#5	(11.77) robust solutions; (9.73) sensor network localization; (7.88) stable set polytope; (7.88) multistage minimax; (7.88) stochasticitygap
#6	(15.32) structural equation model; (12.25) interorganizational relationships; (11.51) partner selection; (11.51) test statistics; (11.22) environmental uncertainty
#7	(10.91) stochastic games; (8.62) ticket resale; (7.88) rational expectations equilibrium; (7.88) affine dynamic-models; (6.99) magazine subscriptions
#8	(9.29) investor; (7.38) weights; (7.38) attention; (7.38) games-group decision; (6.99) user-generated content
#9	(11.22) exponential servers; (7.21) service-level differentiation; (6.93) contact rates; (6.93) many-server queues; (6.93) steady-state analysis
#10	(11.1) cell transmission model; (10.56) riemann problem; (10.17) fundamental diagram; (9.73) roads; (9.22) traffic oscillations

在共被引网络聚类分析中,按共引产生的时间顺序和文献内容关键词分析,可将研究领域划分为若干类别,每一类别代表管理科学与工程领域的一个知识群。从本次聚类结果来看,共产生 11 个知识群,表明在 2006—2013 年之间,管理科学与工程领域出现了 11 种明确的研究区域,在同一研究区域内,研究问题之间存在紧密联系,而在不同的研究区域内,学者们研究的问题存在着较

大的差异。

### 4.3 文本特征抽取

为了提高文本特征抽取的精度,我们首先利用 CiteSpace 对各年度的文献记录进行了聚类处理(以文章的参考文献作为节点),并从各类中提取出该类中主要的研究方向,之后通过分词、去除停用词、标记词性,依据 TF-IDF 技术进行特征抽取四大步骤整理出各年度研究的高频关键词。

表 5 经文本挖掘处理得出的高频关键词

高频关键词	年份	高频关键词	年份
"-0.545 * "enropy maximization"	+	-	
0.545 * "circlarcity"	+	-	
0.272 * "coordinatinginventory control"	+	-	
0.272 * "inventory control"	+	-	
0.272 * "fixedorderingcost"	+	-	2006
0.179 * "confirmatory facor-analysis"	+	-	
0.179 * "structuralequationmodel"	+		
0.156 * "stablestet"	+	0.156 * "polytope"	
0.156 * "qraph"			
		0.458 * "queuingtheory" + 0.458 * "spillback" + 0.458 * "time" + 0.45 * "delay" + 0.396 * "highway" + 0.028 * "network" + 0.022 * "acyclicity"	2010
		0.022 * "userequilibriumtrafficassignment" + 0.022 * "simplicialdecomposition" + 0.014 * "enumeration"	

(续表 5)

高频关键词	年份	高频关键词	年份
0.801 * "complexity"+0.341 * "utility" +		0.305 * "chainmanagement" +	
0.274 * "dynamicpickup" +		0.305 * "structuralanalysis" +	
0.274 * "timetrafficinformation" +		0.305 * "allianceformation" +	
1.63 * "sharedknowledge"+0.163 * "exploitation" +	2007	0.305 * "centrality"+0.305 * "information"+	2011
0.128 * "verticalintegration" 0.067 * "model" +		0.305 * "innovation" +	
0.045 * "algorithm" +		0.305 * "longitudinalanalysis" +	
0.043 * "supplychainmanagement"		0.305 * "networkstructure" +	
		0.305 * "performance" +	
		0.305 * "socialnetworkanalysis"	
		"-0.468 * "stochastic-approximation" + -	
0.465 * "competition"+0.465 * "contractchoice"+		0.468 * "armedbanditproblem" + -	
0.465 * "capacityinvestment"+0.465 * "callcenter"+		0.468 * "maximun-likelihood-estimation"+-	
0.365 * "pricing"+0.017 * "flexibility" +	2008	0.36 * "information"+-0.319 * "demand"+-	2012
0.017 * "inventory" + 0.008 * "commitment" +		0.104 * "interactionaljustice" + -	
0.008 * "equilibriumanalysis"+0.008 * "cost"		0.104 * "organizationaljustice" + -	
		0.104 * "relationshipcommitment" + -	
		0.098 * "justice" + -	
		0.064 * "fimperformance" "	
0.598 * "lineselectionproblem" +		0.258 * "overconfidence" +	
0.598 * "productline" + 0.525 * "behavior" +		0.258 * "behavioralexperiment" +	
0.049 * "revenuemaanaement" + 0.049 * "demand"+		0.258 * "expert" + 0.258 * "expertise" +	
0.025 * "(" + 0.025 * "demandsubstitution" +	2009	0.258 * "firm" + 0.258 * "intuition" +	2013
0.025 * "choicemodel" +		0.258 * "judgment" +	
0.025 * "mutinomiallogitmodel" +		0.258 * "riskassessment" +	
0.025 * "competition"		0.258 * "representativedesign" +	
		0.258 * "process"	

### 5 创新轨迹分析

将利用文本特征抽取方法得到的高频关键词按相对出现频率排列,手动去除部分特殊停用词,聚类后整理出各年度主要的研究方向,得出管理科学与工程领域的科研创新轨迹,如下图 1、2 所示。

利用科学知识图谱、共引分析发现,过去 8 年中管理科学与工程的研究轨迹沿时间线主要经过了三个阶段:

第一阶段(2007 初—2007 末):准备期。#10 类位于这一阶段,呈现出总体分散,局部集中的趋势,以聚类 #10 为代表的 cell transmission model、riemann problem、fundamental diagram、roads、traff-ic oscillations 等高频关键词成为该时期研究的热点

问题。

第二阶段(2008 初—2009 年末):高产期。从图像上看,本阶段是管理科学与工程领域研究发展的重要时期,聚类呈现出总体集中的趋势,类与类之间的关系密切,相互交错。只有 #0、#2、#1、#3 类稍显突兀,与其他类相距较远,且周围存在较多孤立点。本时期是管理科学与工程领域研究的爆发高潮,在几个热点问题方面涌现出大量关联紧密的高水平研究,从表中可以看出,大多数爆发指数较高的高频关键词都集中在这个时期。

第三阶段(2010 年以后):散布期。总体而言这一阶段的研究方向极为分散,可能是由于在经历了前一个时期的爆发之后,原热点问题的相关研究已趋完善,学者们纷纷投入到新领域的开拓,故而难以判断出该时段管理科学与工程领域的大体研究趋势。

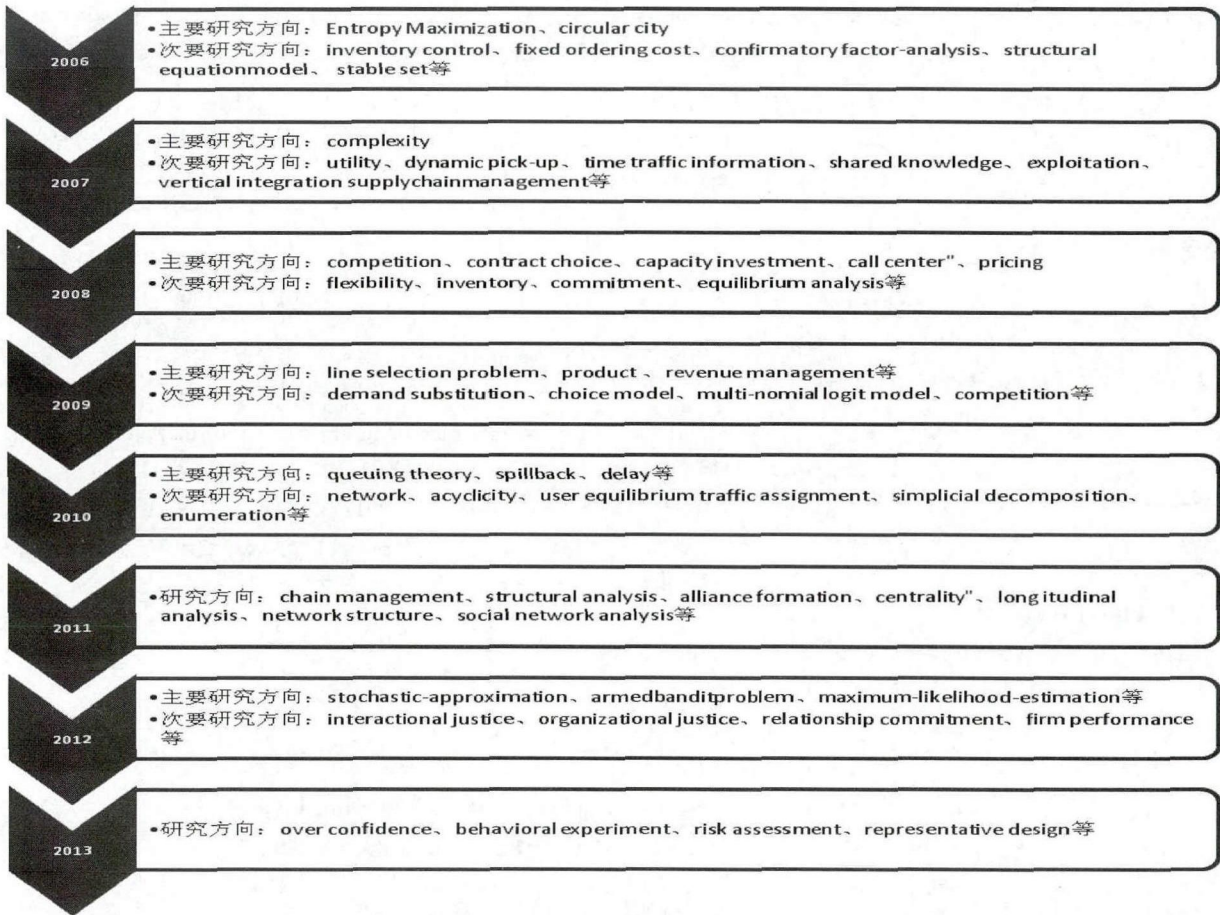


图1 科研创新轨迹1(基于特征抽取)

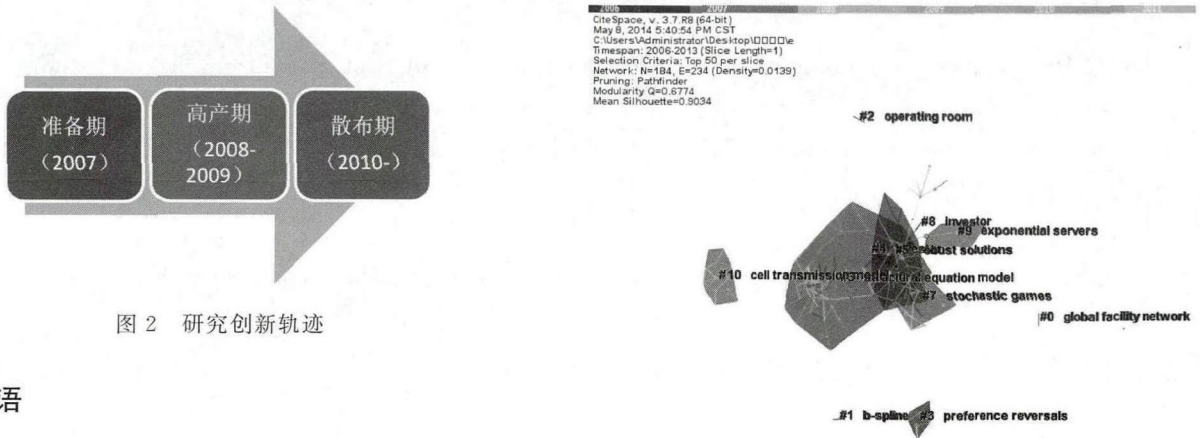


图2 研究创新轨迹

### 6 结语

本文运用文献计量结合文本挖掘的方法对2006—2013年管理科学与工程领域的发展历程进行回顾和分析,获知2006—2013年间管理科学与工程领域出现的11种明确的研究区域,利用科学知识图谱、共引分析将研究创新轨迹划分为准备期、高产期和散布期,结合特征提取等文本挖掘方法揭示近8年来管理科学与工程领域的重要而独特的科研方向,为了解和掌握管理科学与工程的发展与演化过程提供了独特视角。

文献代表着历史,其中蕴涵着一部分的未来趋

图3 创新轨迹2(基于 CiteSpace)

势。本研究是建立在国外2006—2013年被Web of Science收录的10种管理科学与工程领域的国际顶级期刊全部文献记录的大量数据基础之上,研究结果较为客观公正。

本文的研究也存在着一些局限性,首先,在文献的选择上,本文选取的数据源虽然有一定的代表性,但在文献覆盖面和针对性上仍存在一些不足。其次,分析结果的科学性受引文共引网络准确性和合



理性的影响;再次,直接将关键词的词频研究作为判断的依据会有一定的局限性。未来可考虑更深入地结合文本挖掘技术分析文献的研究内容,以得出更丰富的结论,并对停用词表进行优化,这样不仅可以提高特征抽取的速度,同时也可以提升结果的精度。

#### 参考文献:

- [1] 梅世强, 齐二石, 王雪青. 管理科学与工程学科概念及其体系研究[J]. 工业工程, 2006, 9(2): 1-5.
- [2] 李群霞, 张群. 国内外管理科学与工程研究热点及发展趋势比较分析[J]. 中国科技论坛, 2007(6): 126-130.
- [3] 张玲玲, 房勇, 杨涛, 等. 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): 379-385.
- [4] 王虎. 国内外管理科学与工程研究热点及发展趋势比较分析[J]. 现代经济信息, 2012, (22): 001.
- [5] Chen Chaomei, Morris S. Visualizing evolving networks: Minimum spanning trees versus pathfinder networks[C]. Proceedings of IEEE Symposium on Information Visualization, Seattle, Washington, October 19-21, 2003.
- [6] Pons-Porrata A, Berlanga-Llavori R, Ruiz-Shulcloper J. Topic discovery based on text mining techniques[J]. Information processing & management, 2007, 43(3): 752-768.
- [7] Tseng Y H, Lin C J, Lin Y I. Text mining techniques for patent analysis[J]. Information Processing & Management, 2007, 43(5): 1216-1247.
- [8] Guo Ying, Ma Tingting, Porter A L, et al. Text mining of information resources to inform Forecasting Innovation Pathways [J]. Technology Analysis & Strategic Management, 2012, 24(8): 843-861.
- [9] 章以金, 宗乾进, 袁勤俭. 国际管理信息系统研究热点及趋势[J]. 情报杂志, 2013, 32(4): 80-84.
- [10] 彭华民, 万国威. 从沉寂到创新: 中国社会福利30年学术轨迹审视[J]. 东岳论丛, 2010, (8): 145-151.

### Innovation Track of Management Science and Engineering Field— from Bibliometrics and Text Mining Based on Top Journals

LI Si-zhi<sup>1</sup>, LI Jia-jun<sup>2</sup>, LI Yan-hong<sup>2</sup>

(1. Business School, Shanghai University of Finance and Economics, Shanghai 200433, China; 2. School of Information Management and Engineering, Shanghai University of Finance and Economics, Shanghai 200433, China)

**Abstract:** In this paper, 10 kinds of top international journals of management science and engineering are selected as the data source, with downloading all literature records in 2006—2013. Through the analysis of two angles based on output research and co-citation research, the research distribution of management science and engineering by using feature selection is found, and the innovation track by knowledge mapping is explored.

**Key words:** feature selection; citing analysis; co-citation analysis; knowledge mapping; innovation track