

doi:10.19920/j.cnki.jmsc.2021.06.008

基于文献计量的大数据管理决策研究热点分析^①

伍之昂¹, 赵新元², 黄 宾³, 吴 刚^{4*}

(1. 南京审计大学信息工程学院, 南京 211815; 2. 中山大学管理学院, 广州 510275;
3. 浙江水利水电学院经济与管理学院, 杭州 310018;
4. 国家自然科学基金委员会管理科学部, 北京 100085)

摘要: 传统的模型驱动的决策范式正在演变成大数据驱动的决策新范式, 国家自然科学基金委员会于2015年启动了“大数据驱动的管理与决策研究”重大研究计划, 以推动我国管理学领域大数据的基础研究. 本文以重大研究计划资助的114个项目的文本信息, 及2012年~2019年在51种管理类重要国际期刊上发表的近6000篇论文为基础, 利用文本挖掘与文献计量方法, 从词粒度和主题粒度展开比较分析, 并对部分以大数据作为关键词的文献按研究方向进行分类分析. 对比分析结果清晰地呈现出大数据驱动的研究在管理学领域的研究态势和创新脉络; 剖析国内学者与国际领域主流关注的焦点与异同; 揭示出大数据驱动的管理与决策方向上研究热点及其演化趋势, 以及未来值得深入研究的一些科学问题.

关键词: 大数据; 大数据驱动的管理与决策; 文献计量分析; 主题模型

中图分类号: C931.6 **文献标识码:** A **文章编号:** 1007-9807(2021)06-0117-10

0 引 言

大数据已经成为国内外学术界、产业界持续关注的热点领域, 同时也受到世界各国的广泛关注, 国家层面的大数据战略相继被推出^[1]. 2014年大数据首次写入中国中央政府工作报告; 2015年党的十八届五中全会提出实施国家大数据战略. 自从2008年和2011年Nature和Science杂志相继推出大数据专刊以来, 大数据吸引了众多不同学科研究者的目光, 如物理学、信息科学、管理学、经济学、生物医学等. Ahmad等^[2]文献计量结果表明, 2008~2017期间各类期刊共刊载了近35000篇有关大数据的论文, 中国学者发文数量仅次于美国列第二位. 国家自然科学基金委员会(NSFC)于2015年9月启动了“大数据驱动的管理与决策研究”重大研究计划(简称NSFC大数据

重大研究计划), 旨在充分发挥多学科交叉合作研究的优势, 围绕学科领域趋势、理论应用特点, 注重基础性、前瞻性和交叉性研究创新^[3]. NSFC大数据重大研究计划已资助项目114项, 并吸引160余家国内高校、研究机构和企业, 以及30多家境外高校的研究人员参与其中.

本文旨在利用文本挖掘及文献计量等方法, 对NSFC大数据重大研究计划已资助项目信息、及管理领域国际重要学术期刊中的相关文献, 展开统计分析、主题挖掘、及项目与文献的对比性分析, 以期分析大数据驱动的管理与决策相关研究在整个管理领域的所处地位与发展趋势, 对比分析NSFC大数据重大研究计划侧重资助的四大研究方向及4个领域情境与国际上热点的一致性和差异性, 为优化学科方向优先资助领域、以及国内学者捕捉研究动态和研究议题提供决策参考. 毋

^① 收稿日期: 2020-07-06; 修订日期: 2020-07-14.

基金项目: 国家自然科学基金资助项目(72072091; 71571093; 71872191); 教育部人文社会科学研究规划基金项目(18YJA630151); 广东省自然科学基金资助项目(2018A030313502; 2021A1515011978).

通讯作者: 吴 刚(1977—), 男, 河北唐山人, 博士, 研究员, Email: wugang@nscf.gov.cn

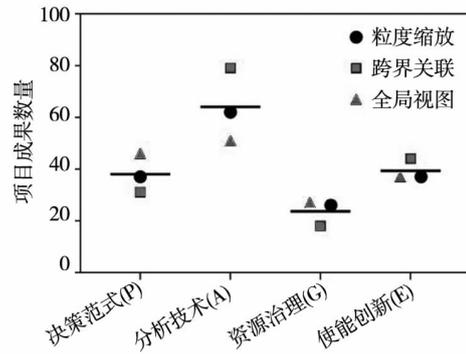
庸讳言,文献计量方法已经被应用于分析大数据研究领域的文献. Ahmad 等^[2]对 2008 ~ 2017 10 年间大数据研究文献的刊文数量和引用数量做了系统性分析,但是其刊文多的期刊多来自于信息科学领域. 还有很多研究立足于特定的应用领域,如商务^[4]、生物医学和医疗^[5]、供应链管理^[6,7]等,围绕大数据对该领域的影响及大数据在该领域内的研究动态和应用等方面展开分析综述. 相比于上述工作,本文立足于管理文献分析大数据管理决策方向的研究动态,同时,本文引入 NSFC 大数据重大研究计划已资助项目信息,作为折射国内学者在大数据管理决策方向上的研究缩影.

1 数据抽取及统计性描述

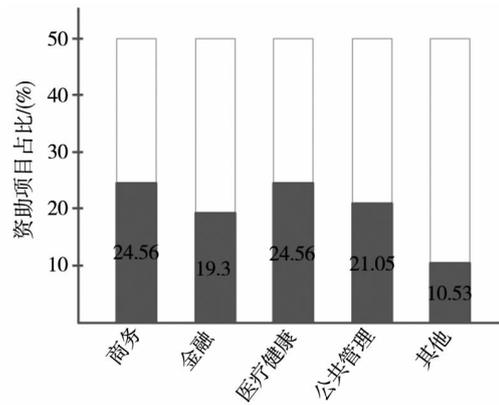
本文的数据包含两部分. 第一部分数据是 NSFC 大数据重大研究计划从 2015 年 ~ 2018 年资助的项目数据,本文抽取出每个项目的中英文标题、关键词、摘要等内容描述信息. 重大研究计划以全景式 PAGE 框架作为总体思路框架^[3],主要围绕 4 个领域情境,沿着四大研究方向展开创新研究,同时每个研究方向分为粒度缩放、跨界关联、全局视图 3 个层次,构成 PAGE 4 × 3 的要素矩阵. 图 1(a)展示了所有项目产出的代表性成果在 PAGE 4 × 3 要素矩阵上的分布情况,成果所属的类别由课题组提供. 可以看出,分析技术方向的成果最多,紧随其后的是使能创新方向,即关注大数据能力带动的价值创造. 从图 2(b)给出的领域情境分布可以看出,绝大部分资助项目以 PAGE 框架要求的领域场景展开研究,少量项目涉及到农业、电力、航天等其他场景,总体上看,资助项目在 PAGE 框架四大场景上的分布较为均衡.

第二部分数据则来源于管理领域国际重要学术期刊中与“大数据驱动的管理与决策研究”相

关的文献,通过合并 FT Rank 50 和 UTD 24 两个期刊列表,获得 51 本重要学术期刊. 然后根据全景式 PAGE 框架 4 个研究方向及 4 个领域情境,从项目数据的英文摘要和英文关键词中抽取出不低于 10% 的项目申请书出现的词汇,得到如表 1 所示的大数据驱动管理与决策领域的术语种子. 在 Web of Science 数据库中,检索 2012 年 ~ 2019 年时间段内 51 本期刊中标题、摘要、关键词包含术语种子的论文,获得共计 5 937 篇文章作为文献数据集.



(a) 研究方向分布
(a) Distribution on research directions



(b) 领域场景分布
(b) Distribution on application scenarios

图 1 “大数据驱动的管理与决策研究”重大研究计划项目分布
Fig. 1 Distribution of projects funded by the big-data-driven managerial decision-making research program

表 1 大数据驱动管理与决策领域术语种子

Table 1 Term seeds in the field of big-data-driven managerial decision-making

类型	术语种子
通用	big data, data driven, decision making, big data analytics
理论范式(P)	panoramic decision, panoramic modeling, decisioning paradigm
分析技术(A)	data fusion, correlation analysis, deep/machine learning, knowledge network, social network(s), social media
资源治理(G)	data governance, data privacy, data sharing, data asset
使能创新(E)	value creation, innovation
领域情境	healthcare, public management, marketing, finance regulation, transportation/transport

注: 部分术语种子包含词义相近的词汇,如 big-data, decision maker, big data analysis

图 2 展示了文献数据集的年际分布, 管理学期刊中大数据驱动的管理决策相关的文献年数量较多, 且总体上呈现出上升趋势. 尽管 2012 年 ~ 2015 年“大数据”关键词在论文中出现次数还极少, 与大数据相关的术语却已经开始大量出现, 如社会网络、数据驱动、数据分析等, 这说明早期的研究广泛开始使用数据分析方法. 2016 年开始, 随着“大数据”这一术语被学术界接受, “大数据”方始遍及于文献. 表 2 列出了刊载大数据驱动管理与决策领域论文最多的前 20 种期刊, 可以看出, 这些期刊覆盖了很多管理学细分领域, 如创新管理、战略管理、信息系统、市场营销、运筹管理等.

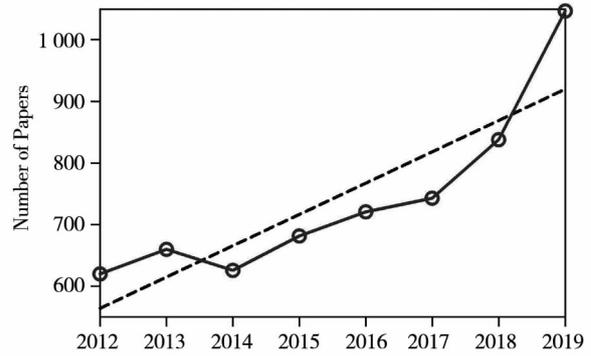


图 2 大数据驱动管理与决策领域文献的年际分布
Fig. 2 Interannual changes in the number of published papers related to big-data-driven managerial decision-making

表 2 管理学重要期刊中载文量最多的前 20 种期刊

Table 2 Top 20 journals with the largest number of published papers in management science

Rank	期刊	载文量	比例	Rank	期刊	载文量	比例
1	Journal of Business Ethics	653	28.43%	11	Journal of Management Information Systems	140	58.33%
2	Research Policy	382	39.30%	12	Information Systems Research	116	31.96%
3	Strategic Management Journal	336	42.59%	13	Journal of Applied Psychology	115	18.25%
4	Management Science	293	21.12%	14	Journal of Marketing Research	106	25.60%
5	Organization Science	232	39.39%	15	Operations Research	106	15.30%
6	Academy of Management Journal	168	30.94%	16	Organization Studies	104	24.53%
7	Production and Operations Management	162	19.98%	17	Journal of Management Studies	103	36.52%
8	Journal of Management	157	34.66%	18	Journal of Consumer Research	99	21.52%
9	Organizational Behavior and Human Decision Processes	155	39.95%	19	Journal of Business Venturing	95	31.88%
10	MIS Quarterly	141	38.84%	20	Entrepreneurship Theory and Practice	95	31.15%

注：“比例”指载文量占该期刊在 2012 年 ~ 2019 年刊文总数量的比值.

2 分析框架

2.1 关键词的度量

关键词是文献计量分析中常用的分析单位, 也是识别研究热点有效手段. 关键词在文档中出现的次数或频率能直观地表达出该词的重要程度, 除此之外, 本文引入两个统计量: TF-IDF 和 Burst 分值, 以更加精确地衡量关键词在文献或项目中的重要性. 设 D 为文本集合 (如所有项目的文本信息), d_j 为一条文本记录 (如其中一个项目的文本), w_i 为关键词, 则 w_i 的 TF-IDF 值计算方法为

$$TF-IDF(w_i, D) = \ln(C_{w_i, D} + 1) \ln \frac{|D|}{|d_j: w_i \in d_j| + 1} \quad (1)$$

式 (1) 为 TF-IDF 的变例, $C_{w_i, D}$ 表示 w_i 在整个文本集合 D 中出现的次数, 其对数值作为词的出现频率. 另外, 为刻画关键词在某时间段 t 内的突发性, 即识别关键词随时间的演化趋势, 引入信息检索领域的 Burst 分值^[9]作为指标, 其定义为

$$Burst(w_i, D, t) = \frac{|d_{j,t}: w_i \in d_{j,t}|}{|d_j: w_i \in d_j|} - \frac{1}{T} \quad (2)$$

其中 $d_{j,t}$ 为时间段 t 内的文本记录, T 为文本集合跨越的时间总长度. 显然, 正的 Burst 分值越大表示时间段 t 内 w_i 爆发性增长趋势越大. 若将 T 内所有时间分段上的 Burst 分值联系起来观察, 容易看出关键词随时间的演变趋势.

2.2 主题识别方法

潜在狄利克雷分配 (Latent Dirichlet Allocation, LDA) 模型^[10]是最著名且获得应用最多的

隐含语义结构提取方法, LDA 是一个“词—主题—文档”三层结构的贝叶斯概率生成模型, 将每个文档表示为多个主题混合, 而每个主题是固定词表上的一个多项式分布. LDA 模型的基本原理及其参数估计方法遍及于文献, 在此不再赘述. 简言之, LDA 模型将得到两组概率: 1) $p(z_k | d_j)$ 表示文本记录 d_j 在主题 z_k 上的出现概率, 即隶属度; 2) $p(w_i | z_k)$ 表示词典中关键词 w_i 在主题 z_k 上的出现概率. 本文定义主题强度来描述主题的热门程度

$$\text{Strength}(z_k) = \frac{\sum_j p(z_k | d_j)}{\sum_k \sum_j p(z_k | d_j)} \quad (3)$$

式(3)定义的主题强度实际上是经过归一化的主题 z_k 上所有文档的隶属度之和, 主题强度越大说明越多的文档跟该主题的关联性越紧密. 主题数量 K 是 LDA 模型的预设参数, 为选择最优主题数量, 采用文献[10, 11]建议的困惑度(perplexity)指标来确定最优主题数量, 文本集合 D 的困惑度计算公式为

$$\text{Perplexity}(D) = \sum_{d_j \in D} e^{-p_j},$$

$$p_j = \frac{\sum_k \sum_{w_i \in d_j} \ln p(z_k | d_j) p(w_i | z_k)}{|d_j|} \quad (4)$$

困惑度反映了集合 D 中文本属于主题的不确定性, 困惑度值越高说明不确定性越大, 因此, 困惑度值越低说明 LDA 模型在文本集合 D 上的泛化性能越好.

3 分析结果

3.1 项目资助方向和领域场景在文献中的研究趋势

3.1.1 4个资助方向的研究趋势

NSFC 大数据重大研究计划沿着决策范式(P)、分析技术(A)、资源治理(G)、使能创新(E)四大研究方向展开创新研究, 图1(a)给出了资助项目在 PAGE 方向上的分布情况. 本节尝试基于

PAGE 框架对已有文献进行研究方向的分类, 并在 PAGE 方向上与资助项目进行比较分析. 首先从文献数据中抽取出内容与大数据紧密相关的部分文献, 将关键词或摘要中明确提及“大数据”一词的文献抽出, 共计来自于 28 种期刊的 112 篇论文. 图3展示了刊载论文最多的9种期刊上的分布情况. 可以看出, 2012年和2013年的文献几乎都发表在综合杂志 Harvard Business Review (HBR), 这些综述性文章探讨了大数据时代的到来对管理学变革、商务智能、企业价值等方面的潜在影响. 随着时间的推移, 刊载与大数据紧密相关文献的期刊呈现出多样化, 运筹学、信息系统、市场营销和创新管理等领域的权威期刊均开始出现与大数据关联紧密的高水平研究, 这清晰地说明了大数据驱动的研究范式已经被管理学主流学者们所接受.

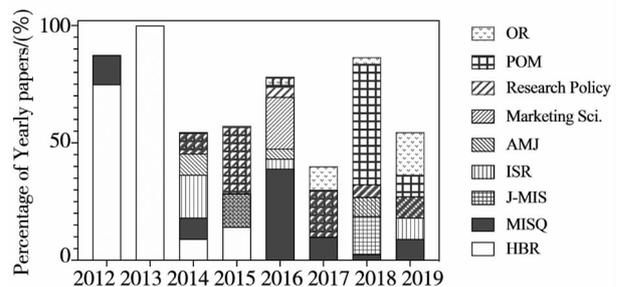


图3 以“大数据”为关键词的文献在期刊上的分布

Fig. 3 Distribution of the published papers in journals with big data as the key word

进一步对上述 112 篇论文按照 PAGE 4 个方向进行了人工分类, 每篇论文仅被分到 1 个方向, 大部分综述论文探讨了大数据对管理学及其若干细分领域的潜在影响、带来的范式转变与挑战性问题, 因此将综述论文划分到决策范式(P)方向, 编委会撰写的介绍性文章(如大数据专刊的介绍)分到其他类别. 此外, 将着重立足于大数据分析模型、方法和技术上创新的论文分到分析技术(A)方向, 而将探讨大数据在不同应用领域情境中应用创新的论文分到使能创新(E)方向, 还有一些探讨大数据共享与治理的论文被分到资源治理(G)方向. 图4展示了年度论文在 PAGE 4 个方向上的分布. 首先, 决策范式(P)方向上的论文占

了很大比例,尤其是 2012 年至 2014 年,经过对 P 方向上论文内容的观察,发现早期的研究主要从宏观角度探析大数据对管理学整体及几个细分领域的研究范式的转变^[4, 13]. 受此影响,越来越多管理学细分领域的学者们开始探讨大数据对各自子领域的影响和范式转变,包括供应链管理^[7]、库存管理^[14]、服务创新^[15]、医疗改革^[12, 15]、运筹管理^[17, 18]等. 其次,在分析技术(A)与使能创新(E)两个方向上的文献在近年来逐渐成为主流,而资源治理(G)方向上的高水平研究相对比较匮乏,这种趋势与图 1(a)所示的资助项目在 PAGE 方向上的分布吻合. 上述结果清晰地反映出大数据相关研究在管理学上的发展脉络,有关决策范式(P)的研究不断细化和深入,在大数据驱动的研究范式逐渐获得学者的认同后,大数据分析技术及应用创新的高水平研究开始涌现.

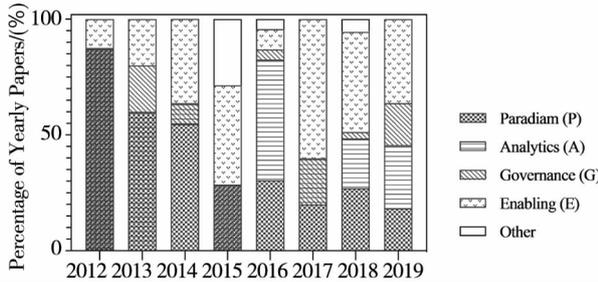


图 4 以“大数据”为关键词的文献在 PAGE 框架上的分布
Fig. 4 Distribution of the published papers in PAGE framework with big data as the key word

3.1.2 4 个资助领域场景的研究趋势

如第 1 节所述,NSFC 大数据重大研究计划鼓励围绕 PAGE 框架设定的四大领域场景展开研究,即商务、金融、医疗健康、公共管理. 本节试图分析围绕这 4 个领域场景的研究在文献中的演变趋势. 为每个领域场景设定 10 个出现频次最高的种子词汇^②,统计文献中每年包含每个领域场景种子词汇的篇数,然后计算种子词汇集合的 Burst 分值,图 5 给出了实验结果. 总体上看,四大领域场景相关的研究均不同程度地有所上升. 其中,商务和公共管理领域的研究热度变化相对偏小,这说明从

2012 年至今,这两个领域一直是管理学研究关注的重要场景. 同时,医疗健康领域的研究增长趋势最为明显,正在渐渐成为管理学中的重要交叉领域. 由图 1(b),NSFC 大数据重大研究计划所资助的项目中医疗健康在四个场景中列第一位,这说明优先资助领域与国际学术界关注热点较若画一.

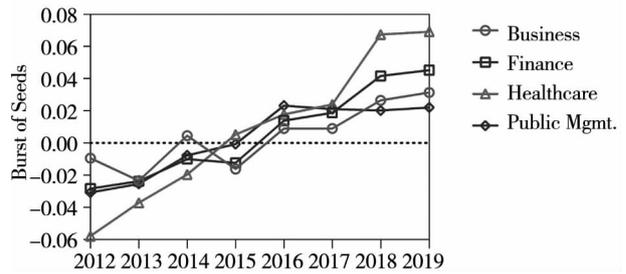


图 5 文献中四大领域场景相关研究的变化趋势
Fig. 5 Trends of the research on four application scenarios

3.2 项目数据和文献数据在关键词及其形成的主题上的对比分析

本节将项目申请书中文标题、中文摘要、中文关键词作为项目文本库,用“结巴”中文分词工具进行分词,去除停用词后获得 2 252 个词汇. 与此相对应,将论文标题、摘要和关键词作为英文文献文本库,采用 NLTK 自然语言处理工具进行英文分词,去除停用词,并利用 NLTK 提取词原型(Lemmatization),获得 7 670 个词汇. 值得注意的是,在分词过程中,把大量术语作为预设的人工词库,避免重要术语被分词割裂,比如数据驱动、网络舆情、HR-management、social network 等.

3.2.1 项目数据和文献数据关键词对比

对所有词汇计算 TF-IDF 统计量,以表征词汇在数据集上的重要性,图 6 词云图展示了两个数据集上 TF-IDF 值最高的 200 个关键词. 可以从中观察出资助项目关键词更加聚焦于大数据融合分析技术,及大数据在医疗健康、金融等领域的应用,而文献研究范围则比较广泛,很多传统的管理学研究方向仍然占据主导地位,如决策支持、企业管理、创新管理、组织管理、心理伦理等.

② 限于篇幅,仅列举医疗健康领域的种子词汇: healthcare、life、medical、patients、disease、treatments、hospital、clinic、drug、electronic medical records.



(a) 项目词云图 (b) 文献词云图
 (a) Word cloud of projects (b) Word cloud of literature

图6 项目和文献数据中词云图对比

Fig. 6 Comparison of the word clouds corresponding to projects and literature

大数据时代的到来,数据驱动的方法、大数据技术是否正在逐渐渗透到管理学领域的经典研究呢?为此,作者观察文献数据集中近3年来呈现上升及下降趋势的关键词,以每个关键词近3年的Burst分值之和作为统计量,由式(2)Burst分值定义,若Burst分值之和为正,说明该关键词近3年增长趋势明显;反之,Burst分值之和为负,则该关键词下降趋势明显.图7分别给出了Burst分值和最大和最小的前10个关键词,可以看到,近3年来与大数据分析方法与系统相关的关键词呈现出很明显的增长,如Platform、Machine Learning、Big Data、Deep Learning等.反之,一些传统的研究热点在近3年来明显下降,如与“优化”相关的关键词Heuristics和Complexity,与信息系统相关的关键词IT和IS等.

3.2.2 资助项目的主题识别与分析

利用LDA模型对NSFC大数据重大研究计划资助的114个项目进行主题抽取,将主题数量K从5增加到50,从中挑出困惑度值最低的情况作为最

优K取值,此时主题数量K=10.图8左侧给出了10个主题强度分布,右侧列出了5个强度最大主题中概率排在前10的关键词作为主题含义的代表.具体地,主题3主要包含了大数据分析相关的一些通用词汇,包含于大部分资助项目的题目、关键词和摘要之中,因此该主题强度明显高于其他主题.主题6主要是社会网络相关的内容,社会网络分析已经成为管理学的重要研究方法和视角,社交媒体蕴含着丰富的时空、行为、文本等大数据,大数据分析技术的不断成熟推动了社会网络分析与管理学各分支中的深度融合.主题2主要是商务场景的研究内容,并涉及到系统平台的研制及应用.主题5侧重于大数据治理体系,从技术上涵盖大数据管理与存储、共享与开放、安全与隐私保护、质量管理等多个方面,这是继大数据分析之后的又一个热点,也是NSFC大数据重大研究计划优先资助领域之一.主题8主要是医疗健康场景的研究内容,由第1节所述,该领域场景是获得资助最多的场景之一,因此其对应的主题也比较活跃.

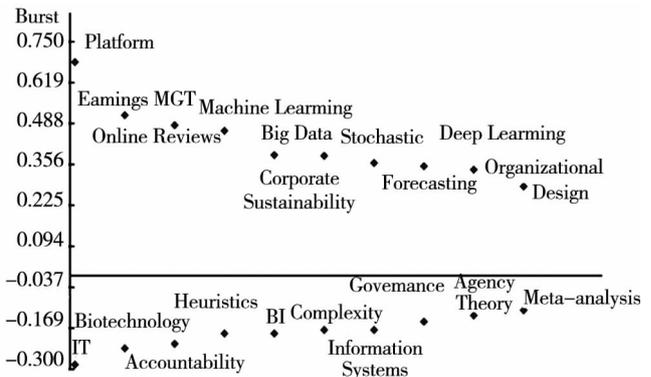


图7 文献中近3年上升和下降趋势最快的词汇

Fig. 7 Words with the fastest rising and decreasing trends in the literature over the past three years

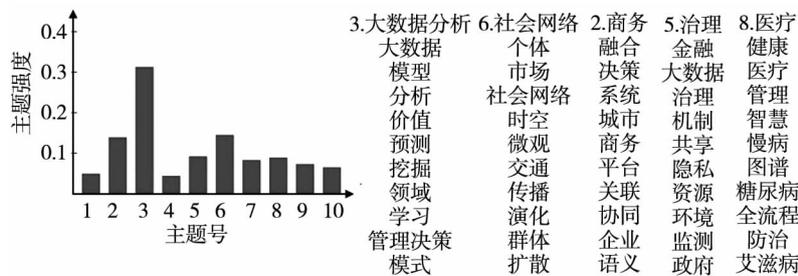


图8 项目数据集上主题识别结果

Fig. 8 Topics extracted from project data set

3.2.3 文献数据主题识别与分析

本节展示文献数据集上主题识别的结果,由于文献数据跨度为 8 年,为了展示主题的演化,将每两年的数据作为一个阶段,共分为 4 个阶段.在每个阶段的文献数据上,利用 LDA 模型进行主题识别,同样根据式(4)困惑度最低确定主题数量 K,并从中挑选出强度最高的 5 个主题,列出每个主题中概率最大的 5 个关键词及其概率分布,图 9 展示了实验结果.可以从中观察到以下几个现象:

1) 与数据驱动的管理决策紧密相关的文献

中,以盈利性组织和企业经营活动为研究对象的工商管理持续受到密切关注,包括阶段 1 的主题 29、主题 10 和主题 0,阶段 2 的主题 5 和主题 0,阶段 3 的主题 21、主题 24、主题 3 和主题 4,及阶段 4 的主题 10、主题 11 和主题 15,这些主题关键词涵盖了组织管理、战略管理、客户行为、经济金融、企业社会责任等内容;同时,以政府和非盈利机构为管理主体的公共管理稳定地成为 2012 ~ 2017 年度 3 个阶段强度排名前列的主题,如阶段 1 主题 17、阶段 2 主题 4 及阶段 3 主题 19.

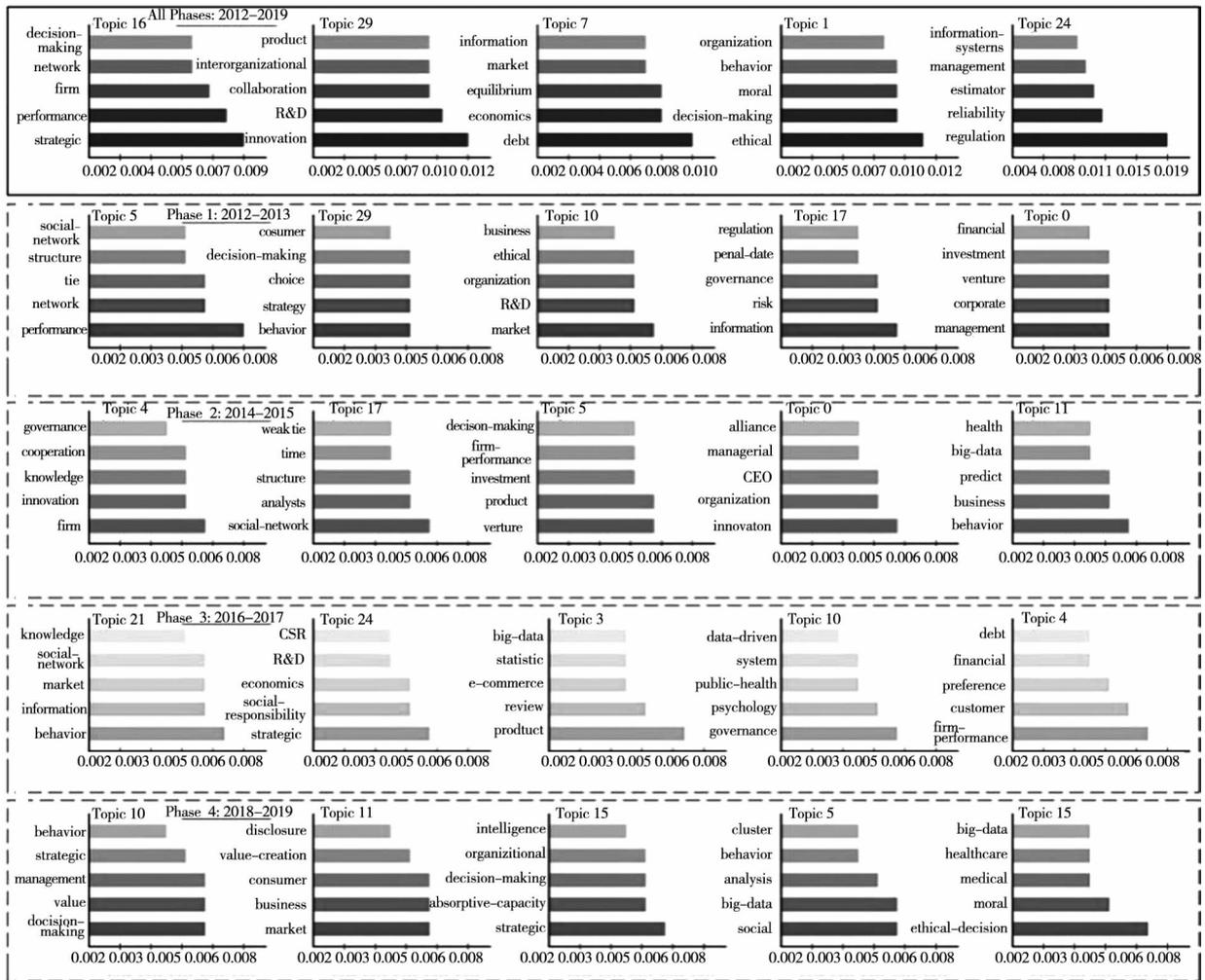


图 9 文献数据集上主题识别结果

Fig. 9 Topics extracted from bibliometric data set

2) 与大数据驱动相关的研究从 2014 年开始成为学术界的关注焦点,阶段 2 关键词首次出现“大数据”,与“行为”、“商务”、“预测”、“健康”一起构成主题 11;阶段 3 主题 3 则包括“大数据”、

“电子商务”、“商品评论”等关键词,主题 19 包括了“数据驱动”、“公共健康”、“心理学”、“治理”等关键词;阶段 4 主题 15 包括“大数据”、“医疗健康”、“道德”、“伦理决策”等关键词;这说明大

数据正在渗透到管理学的众多分支方向。

3) 社交网络分析在 2012 年~2019 年的文献数据中持续成为热门主题,大量研究从不同层面(如个体、团队、组织等),围绕社会性结构以及其作用机制对管理学现象影响而展开,这些现象包括领导力、谏言行为、市场营销、广告投放、以及边界拓展等。

4 结束语

本文对 NSFC 大数据重大研究计划资助项目信息、以及管理学重要国际期刊中大数据相关的文献展开了比较分析,得到以下几个结论:

(1) 大数据驱动管理与决策相关的研究已经成为管理学重要期刊中持续增长的热点议题之一。大数据决策范式、分析技术、及大数据应用相关的研究在管理学期刊已经占有重要的一席之地,而且大数据处理技术正在从传统的数据挖掘和数据分析向深度智能化迈进。大数据驱动的研究范式,呈现出一种解决问题的新思路。

(2) 大数据驱动管理与决策相关的研究更多地围绕盈利性组织和企业经营活动等微观层面展开,政府层面的大数据治理相关的研究发展滞后。面向盈利性组织和企业展开的研究属于全景式 PAGE 框架中的“使能创新”方向,其本质是如何利用企业内外部大数据、推动商务和企业绩效的提升。

(3) 本文沿着全景式 PAGE 框架梳理了明确提及“大数据”的文献,从中观察到管理学重要国际期刊早期主要探讨大数据时代的来临对管理学

研究的变革性影响与随之而来的挑战性问题,这类有关管理决策范式转变机理的文献推动了后续研究在两个分支上的快速发展。其一,大量管理学子领域开始讨论大数据引起的管理决策范式转变和挑战性议题,供应链管理、服务创新、运筹学等,这说明大数据驱动的研究范式深度渗透到管理学研究。其二,大数据驱动的应用研究日渐丰富多样,即大数据“使能创新”获得研究者的高度关注,有别于传统的模型驱动范式,由“应用+数据”驱动的研究范式变得越来越流行。

(4) 从 4 个典型应用场景角度,医疗健康是 NSFC 大数据重大研究计划资助最多的领域场景,医疗健康相关的文献从 2012 年开始呈现飞速上升趋势。可利用的医疗大数据来源于病历卡、处方记录、传感器监测记录、影像信息、在线评论文本等,相比于传统医疗实验数据,医疗大数据获取成本低、易脱敏、样本量大、可支撑的变量丰富,这极大地促进了大数据驱动的研究和应用在医疗健康领域的崛起。

本文的研究不可避免地存在着局限性。首先,在文献数据源上,在管理学领域虽然具有代表性,但是覆盖面依然不广,例如医学领域的学者们在综合性期刊与医学类期刊上发表大数据相关的成果屡见不鲜,本文的文献数据未能涵盖类似细分领域的动态和成果。其次,项目申请书和论文的参考文献未纳入分析范围,共引网络分析是文献计量的重要手段,能刻画文献之间的关联、及国内外学者之间的关联,有助于得出更加丰富详实的结论。

参考文献:

- [1] 冯芷艳, 郭迅华, 曾大军, 等. 大数据背景下商务管理研究若干前沿课题[J]. 管理科学学报, 2013, 16(1): 1-9. Feng Zhiyan, Guo Xunhua, Zeng Dajun, et al. On the research frontiers of business management in the context of big data [J]. Journal of Management Sciences in China, 2013, 16(1): 1-9. (in Chinese)
- [2] Ahmad I, Ahmed G, Shah S A A, et al. A decade of big data literature: Analysis of trends in light of bibliometrics[J]. The Journal of Supercomputing, 2018: 1-17.
- [3] 陈国青, 吴刚, 顾远东, 等. 管理决策情境下大数据驱动的研究和应用挑战——范式转变与研究方向[J]. 管理科学学报, 2018, 21(7): 1-10. Chen Guoqing, Wu Gang, Gu Yuandong, et al. The challenges for big data driven research and applications in the context of managerial decision-making: Paradigm shift and research directions[J]. Journal of Management Sciences in China, 2018, 21(7): 1-10. (in Chinese)
- [4] Chen H, Chiang R H L, Storey V C. Business intelligence and analytics: From big data to big impact[J]. MIS Quarterly, 2012, 36(4): 1165-1188.

- [5] Luo J, Wu M, Gopukumar D, et al. Big data application in biomedical research and health care: A literature review[J]. *Biomedical Informatics Insights*, 2016, 8: BII. S31559.
- [6] Nguyen T, Li Z, Spiegler V, et al. Big data analytics in supply chain management: A state-of-the-art literature review[J]. *Computers & Operations Research*, 2018, 98: 254 – 264.
- [7] Helmuth C A, Craighead C W, et al. Supply chain management research: Key elements of study design and statistical testing[J]. *Journal of Operations Management*, 2015, 36: 178 – 186.
- [8] 周青, 顾远东, 吴刚. 创新管理研究热点的国际比较与学科资助方向——国家自然科学基金项目管理视角的思考[J]. *经济管理*, 2017, (12): 190 – 201.
Zhou Qing, Gu Yuandong, Wu Gang. A comparison of the hot topics on innovation management at home and abroad for the suggestion of the priorities for research funding: From the perspective of project management of NSFC[J]. *Business Management Journal*, 2017, (12): 190 – 201. (in Chinese)
- [9] Hoonlor A, Szymanski B K, Zaki M J. Trends in computer science research[J]. *Communications of the ACM*, 2013, 56(10): 74.
- [10] Blei D M, Ng A Y, Jordan M I. Latent Dirichlet allocation[J]. *Journal of Machine Learning Research*, 2003, 3(Jan): 993 – 1022.
- [11] Mathew G, Agrawal A, Menzies T. Finding trends in software research[J]. *IEEE Transactions on Software Engineering*, Early Access, DOI: 10.1109/TSE. 2018.2870388. 2018.
- [12] Kohli R, Tan S S L. Electronic health records: How can IS researchers contribute to transforming healthcare? [J]. *Management Information Systems Quarterly*, 2016, 40(3): 553 – 573.
- [13] Agarwal R, Dhar V. Big data, data science, and analytics: The opportunity and challenge for IS research[J]. *Information Systems Research*, 2014: 443 – 448.
- [14] Bertsimas D, Kallus N, Hussain A. Inventory management in the era of big data[J]. *Production and Operations Management*, 2016, 25(12): 2006 – 2009.
- [15] Lehrer C, Wieneke A, vomBrocke J, et al. How big data analytics enables service innovation: Materiality, affordance, and the individualization of service[J]. *Journal of Management Information Systems*, 2018, 35(2): 424 – 460.
- [16] Hopp W J, Li J, Wang G. Big data and the precision medicine revolution[J]. *Production and Operations Management*, 2018, 27(9): 1647 – 1664.
- [17] Cohen M C. Big data and service operations[J]. *Production and Operations Management*, 2018, 27(9): 1709 – 1723.
- [18] 卞亦文, 闫欣, 杨列勋. 社会学习视角下运营管理决策研究[J]. *管理科学学报*, 2019, 22(5): 18 – 30.
Bian Yiwen, Yan Xin, Yang Liexun. Operations management decision issues from the social learning perspective[J]. *Journal of Management Sciences in China*, 2019, 22(5): 18 – 30. (in Chinese)

Status and trends in big-data-driven managerial decision-making on bibliometric

WU Zhi-ang¹, ZHAO Xin-yuan², HUANG Bin³, WU Gang^{4*}

1. School of Information Engineering, Nanjing Audit University, Nanjing 211815, China;

2. Business School, Sun Yat-sen University, Guangzhou 510275, China;

3. College of Economics and Management, Zhejiang University of Water Resources and Electric Power, Hangzhou 310018, China;

4. Department of Management Sciences, National Natural Science Foundation of China, Beijing 100085, China

Abstract: Big data and data science are emerging as a new paradigm and tool in management research. In such contexts, the National Natural Science Foundation of China (NSFC) has launched a major research program since 2015: Big Data driven Management and Decision (BDMD), so as to promote China's fundamental research on big data. Text mining and bibliometric analysis are applied to studying the textual information of 114 granted projects under the BDMD major research program as well as nearly 6 000 articles from 51 leading

international journals over the last 8 years. Our comparative analysis is performed at both the keyword-level and topic-level. The paper also attempts to examine those literatures explicitly claiming big data along with the PAGE framework and by classifying them into different categories. Our research highlights the dynamic BDMD research landscape, with its focus constantly moving to new challenges due to new technological developments. Our findings are able to reveal the current research interests and the revolution trends in the field of BDMD. Furthermore, the comparative analysis between NSFC grants data and bibliometric data provides a robust roadmap for future research in the BDMD field.

Key words: big-data; big data driven management and decision; bibliometric analysis; topic model

(上接第 56 页)

表 A2 指数收益率与“净购买压力”指标回归结果(50ETF 价格作为标的证券价格)

Table A2 Regression results for testing index return and net buy pressure(using 50ETF' s price as underlying price)

被解释变量: R_{t+1}	(1)	(2)	(3)
$NBP1$	1.127 *** (13.054)		1.069 *** (3.294)
$DNBP$		9.091 *** (12.631)	0.502 (0.186)
R_t	-0.276 *** (-58.954)	-0.275 *** (-58.855)	-0.276 *** (-58.944)
截距	2.19E-05 (1.043)	2.17E-05 (1.031)	2.19E-05 (1.043)
调整 R^2	0.078 3	0.078 1	0.078 3

注: 使用 50ETF 价格作为隐含波动率、希腊值等计算中输入的标的证券价格, 下同。

表 A3 波动率指数与“净购买压力”指标回归结果(50ETF 价格作为标的证券价格)

Table A3 Regression results for testing volatility index and net buy pressure(using 50ETF' s price as underlying price)

被解释变量: $\Delta iVIX_{t+1}$	(1)	(2)	(3)	(4)
$NBP2$	1.626 *** (7.283)			1.699 *** (3.459)
$GNBP$		6.963 *** (3.754)		-9.193 *** (-3.078)
$VNBP$			7.557 *** (7.437)	4.388 *** (2.438)
$\Delta iVIX_t$	-0.185 *** (-43.990)	-0.183 *** (-43.648)	-0.185 *** (-44.020)	-0.186 *** (-44.149)
截距	-6.191e-03 *** (-2.983)	-8.542e-03 *** (-4.171)	-6.869e-03 *** (-3.349)	-6.185e-03 *** (-2.981)
调整 R^2	0.045 6	0.044 7	0.045 6	0.045 9