

收益管理中基于大数据仓库的需求无约束估计: 框架与挑战

郭鹏

贵阳学院经济管理学院, 贵州 贵阳 550005

摘要

现有需求无约束估计方法均为基于公司内部数据仓库中所获需求信息而开发,在当前基于大数据分析的激烈竞争市场环境中,无法满足收益管理系统日益增长的实时需求预测和优化决策分析需要。为了实时、动态地同时获取并分析内部和外部数据资源中有关每位顾客的无约束需求数据,包括结构化和非结构化的信息,提出了以面向收益管理需求无约束估计为主题的大数据仓库框架,并据此讨论了无约束需求知识挖掘以及需求无约束估计商务智能分析工具开发应用过程中面临的各项挑战。

关键词

收益管理;需求预测;无约束估计;数据仓库;大数据;商务智能;情感分析

中图分类号:N945.23, N945.24 文献标识码:A doi: 10.11959/j.issn.2096-0271.2017031

Demand unconstraining estimation based on big data warehouse in revenue management systems: framework and challenges

GUO Peng

School of Economics and Management, Guiyang University, Guiyang 550005, China

Abstract

In order to achieve and exam the unconstrained demand data related to each customer in real time and dynamically from both internal and external data resources, including structured and unstructured information, a big data warehouse framework for demand unconstraining estimation in revenue management was proposed. After that, the challenges in unconstrained demand knowledge mining, as well as the process of development and application of business intelligence analysis tools for demand unconstraining estimation were discussed.

Key words

revenue management, demand forecasting, unconstraining estimation, data warehouse, big data, business intelligence, sentiment analysis

1 引言

收益管理(revenue management, RM)理论诞生至今,研究者对其进行了多种定义,其中被引用最多的是:“以最佳的顾客和经营者满意度,在正确的时间以正确的价格将正确的产品销售给正确的顾客^[1]。”RM的目标是通过对比易逝性产品存量的优化控制和管理来最大化公司的盈利能力,其整个决策过程依赖于顾客需求数据。但是,在RM优化策略实施过程中,未能成功预订到偏好产品的顾客的需求会带来“溢出(spill)”和“再现(recapture)”效应。因此,在对未来需求进行预测之前,需要对历史顾客真实需求进行无约束估计(unconstraining estimation)^[2-4]。事实上,需求无约束估计不仅是RM需求预测的核心问题之一^[5],在大数据背景下的企业竞争情报深度学习过程中,它更是一项充满挑战的基础性工作,其对顾客历史真实需求的准确估计决定着预测效果的好坏,并直接影响整个收益管理系统(revenue management system, RMS)商务智能(business intelligence, BI)实施的成败。

商务分析的开展需要基于大数据集合的高级分析技术,为使RMS能够更好地与如今急速发展的基于机器学习的BI相结合, RM智能商务分析应该包含如下3方面的内容:大数据的收集与管理、基于数据描述和预测的规律性分析以及智能决策^[6,7]。同时,数量本身并不能代表大数据,还应具有其他一些特点,比如产生数据的高速性、多样性和复杂性等^[8]。与大数据相关的各种概念被描述为一类技术,它们能够完成信息系统科学研究中一条最基本的原则:在正确的时间以正确的数量、质量将正确的信息提供给正确的接收方。因此,基于大数据的BI有助于提高RMS中的需求预测

效果,并为RM需求无约束估计方法创新和方式转变提供了一次前所未有的机会^[9-11]。

除了收集关于顾客需求的数据外, RMS中的需求无约束估计引擎还需采集市场份额、销售渠道、代理费、价格、客座/入住/租车率等维度的数据。张建城等人^[12]对适合我国国情和市场特色的RM预测系统的总体设计进行了探讨;冯兴杰^[13]、王红^[14]、李丽娟等人^[15]根据我国民航RM的实际应用分析,分别对RMS数据仓库(data warehouse, DW)的设计和构建进行了研究;满青珊等人^[16]设计了我国区域航空公司RM的DW总体架构,以舱位销售和成本管理为主题建立了RM决策支持系统。

但是,在移动互联网爆发式增长、在线旅行社(online travel agency, OTA)带来的巨大冲击并造成自身生存空间岌岌可危的新形势下,开发和利用内部数据资源不足以保证RMS应用企业在市场中的竞争优势。上述RM研究中的DW主要是由发生在组织内部的交易和查询信息组成;同时,现有的需求无约束估计方法均是基于公司内部DW中的需求信息而开发的^[2-4],在描述、计算和分析过程中并未考虑来自于外部数据源中的数据,这造成了现有方法在大数据背景下的应用局限,无法满足RMS实时预测、机器学习、智能优化决策和可视化分析的需要。因此,在RM实践中有必要转向考虑大数据仓库(big data warehouse, BDW)的开发和应用,它是由内部和外部数据集合共同构成的^[17]。本文首先提出了一种以面向RM需求无约束估计为主题的BDW框架,然后据此对RM需求知识挖掘以及需求无约束估计BI开发应用过程中所面临的挑战进行了探讨。需要指出的是,本文所提RM需求无约束估计BDW框架是一般化的,可适用于其他行业和研究领域,本文仅以航空客运、酒店和汽车租赁等RM应用行业为例展开具体说明。

2 问题描述

2.1 收益管理中的需求无约束估计问题

(1) 历史顾客真实需求信息的缺失

在航空客运、酒店、汽车租赁和体育娱乐等RM应用行业中,为了在RM实践中引导顾客购买具有高利润率的产品,公司通常会对产品的可用性进行控制,并在整个优化决策过程中始终平衡好以下两方面的工作:一方面,在产品资源丰富的情况下对低价格产品的销售;另一方面,在预售期结束以前对高价格产品的预留保护。任何上述类似的策略都高度依赖于需求预测,而且必然会面临由“受截尾需求(censored demand)”引起的历史顾客真实需求信息缺失现象:当产品预订总量达到了“预订限制(booking limit)”数量后,随后到来的历史真实需求被系统“驳回”,并受到了拒绝。这部分被拒绝的历史顾客可能最终会放弃购买,并成为“溢出需求(spilled demand)”;也可能继续选择购买价格更高(buy-up)或更便宜(buy-down)的可预订产品,并成为“再现需求(recaptured demand)”“升舱需求(upgrading demand)”或“子舱位变更需求(upselling demand)”。

在上述任何一种情况下,预售系统记录和剩余可用替代产品的可观察需求都不能反映历史顾客的第一选择(first choice),系统内的需求信息具有不完备性。观察到的“受截尾需求”是由历史顾客重新预订(re-book)、重复问询(duplicate)、潜在的取消预订(would-be cancellation)和已预定而未登机/到店(no-shows)行为共同造成的,不能代表顾客的“初始需求(primary demand)”。

基于“受截尾需求”进行预测必将导致错误的动态定价、存量控制和超售等决策结果,直接造成公司总收入的“螺旋式下降(spiral-down)”现象^[18]。

(2) 需求无约束估计和无约束需求预测
为避免上述情况的发生,可在在预测未来顾客需求以前,对历史乘客“初始需求”分布参数的估计过程称为“需求无约束估计(demand unconstraining estimation)”,也可称为需求的非限化估计(det truncating estimation)、溢出分析(spill analysis)和截尾数据分析(censored data analysis),整个过程使用的方法被称为“无约束估计方法(unconstraining method)”^[2-4, 19, 20]。

一般来讲,可将通过无约束估计得到的需求数据称为“无约束需求数据(unconstrained demand data)”,而基于它的预测过程被称为“无约束需求预测(demand forecasting based on unconstraining)”^[21]。已有研究表明,即使在需求水平很低的情况下,“无约束需求数据”对公司收入的影响也达到了3.5%^[22],同时,无约束估计过程对公司收入的改进率为2%~12%^[23]。RM需求无约束估计方法的发展和应用已有近40年的历史(1977—2017年),按照采用的技术手段可将其分为3类:基础法(basic method)、数理统计法(statistical method)和选择模型(choice-based model)法,参考文献[2-4]分别对现有方法的研究进行了综述。在考虑竞争和顾客选择行为的情况下,不同无约束估计方法对公司收入贡献率的差别可达到2%~15%^[24]。

Pöhl S^[25]在其研究中提到,在将预订系统中可观察到的历史预订数据输入RMS以前,应该做“无约束化”处理,如图1所示。无约束估计可以缩小RMS需要的“无约束需求数据”和系统观察到的“受

截尾需求数据”之间的差距,其主要功能是为优化决策顺序实施过程中的预测模型提供能反映真实情况的需求信息,它可被看作对历史顾客真实需求的“预测”。Weatherford L R^[26]对RMS优化策略顺序实施过程中一个完整的预测系统应包括的具体步骤进行了描述,如图2所示,对截尾需求数据的无约束估计是无约束需求预测过程的关键。

2.2 基于大数据仓库的收益管理需求无约束估计

(1) 大数据背景下的收益管理商务智能数据收集的维度、规模和活性是传统RM决策平台与大数据平台成功融合的重要战略资产和创新驱动力,而获取、转化处理以及挖掘需求数据价值的的能力也是大数据时代RM实践企业面临的战略性重大课题。正是意识到了上述机会,越来越多的RMS实践企业都将大数据作为“数据基因重组”的基础。

BI是通过对新机会的识别,并基于深刻学习实施有效战略的一种方法,它为业务提供能带来市场优势和长期稳定性的竞争情报^[27],是将企业中现有数据转化为知识,并据此为业务经营决策的制定提供支撑的系统架构^[28]。一般来讲,BI系统由如下一系列技术、概念和阶段构成:数据集成、DW、在线联机分析处理(online analytical processing, OLAP)多维数据集、数据挖掘方法和分析工具。

为了商业分析的目的,公司依靠将数据转化为有意义可用知识的一系列大数据技术和工具,基于竞争情报的BI便可被开发和获取,这些技术和工具包括推荐系统、社交网络分析、新产品分析、定价策略分析和情感分析等^[29]。在大数据时代,RMS实践企业需要通过交流、购买等方

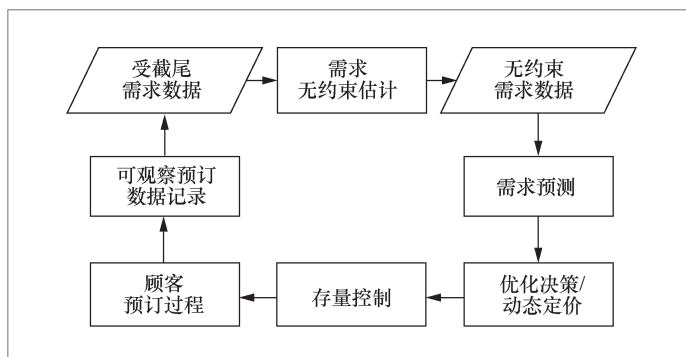


图1 收益管理系统动态决策过程^[25]

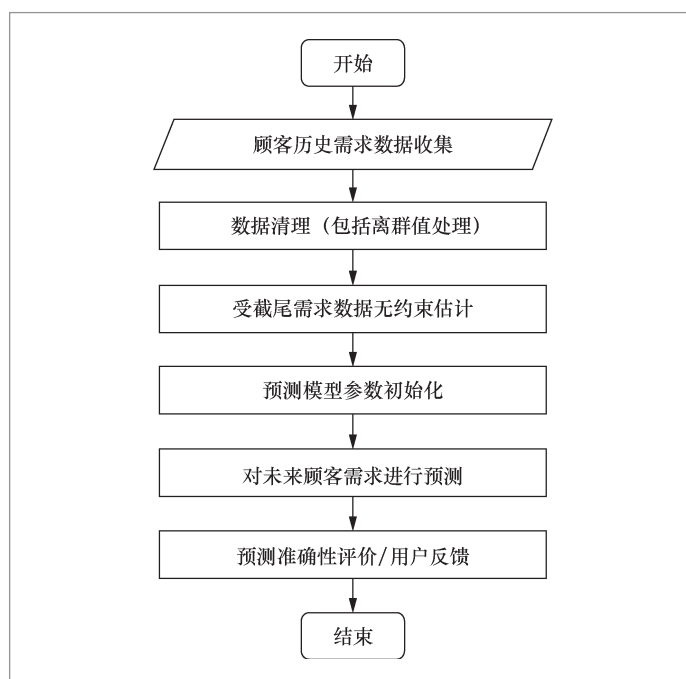


图2 收益管理“无约束需求预测”过程^[26]

式,从同行业其他公司或大数据提供商等渠道掌握行业内或不同行业的直接或跨界需求数据,进行相互补充、印证、后验和预测,并在此过程中强化淡旺季营销,形成新的商业模式和盈利增长点^[30]。

(2) 大数据仓库与实时需求无约束估计

BDW的概念主要是指抽取、整合以及存储来自于各种数据源的大量数据的行为,其可能同时包括结构化(structured)和非结构化(unstructured)的数据^[31,32]。

传统的RMS站在目标市场细分的角度比较产品剩余存量和需求预测值,并依靠预测和估计得到的需求概率值进行决策。而基于BDW的RMS需求无约束估计BI能够实时和动态地访问有关每位顾客的信息,它们是伴随与预订直接或间接相关的在线活动而动态产生的,通过与其他潜在的未来顾客预订请求信息结合,BDW就有可能支持实时需求无约束估计和预测。当具有对每位顾客的真实需求、偏好和支付意愿进行即时评估的能力时,真实的目标和最有效的优化决策就能够达成。

上述机会可用经济理论来解释,即BDW将支持淘汰目前行业中得到普遍应用、低效率的三级价格歧视模型(third-degree price discrimination model),并转变到使用一级价格歧视模型(first-degree price discrimination model)或完全价格歧视模型,公司能够依靠后者获取绝大部分的消费者剩余^[11]。因此,本文所提基于BDW的RM需求无约束估计框架将有助于提高企业在大数据环境中的BI实时决策竞争力。

3 框架的设计

当今社会,信息、创造力和知识在任何决策的制定和执行过程中都扮演着重要角色。为了应对全球化竞争,设计和制定允许收集、组织和处理内部与外部需求信息的机制显得尤其重要。出于对大数据时代下RM实时需求预测与优化决策的考虑,以差异化的方式对收集到的需求信息进行无约束估计,这在RMS实践中的所有层面都是十分重要的。这样做的目的在于减少RM决策制定过程中的不确定性,并跟踪优化模型顺序实施过程中涉及的那些最为敏感的参数。图3是本文所提面向以RM需求无约束估

计为主题的BDW框架设计,对其设计机理和实施阶段的描述可分为以下6个方面。

(1) 需求数据抽取设计

这一阶段主要是指自动地从数据源中收集需求相关信息,这些需求数据既包括内部计算机预售系统、离港系统、常客系统、全球分销系统(global distribution system, GDS)和中央预订系统(central reservation system, CRS)中的旅客实名记录(passenger name records, PNR)、用户查询日志记录、顾客个人资料、消费者心理资料和忠诚度类型等信息,还包括从外部网络环境中获取的数据。本文设计的大数据仓库框架不将论述重点放在如何从公司内部数据库中获取需求数据(具体可参见参考文献[12-16]的研究),而仅针对从外部网络资源中提取需求相关信息的工作进行探讨。

自从互联网被发明以来,服务行业中的大部分交易都可通过电子化手段完成,比如,如今的旅客在进行与旅行相关的所有活动时,都会留下电子化的“足迹”。围绕顾客预订选择行为和享受服务的整个过程,可将RM应用行业的需求数据分为预订数据、过程数据和反馈数据。其中,预订数据包括顾客“初始需求”在企业官网、OTA网站和搜索引擎上的选择行为数据,反映了整个市场的需求状况和顾客偏好信息;过程数据包括顾客登机、入住、驾驶租赁车辆时产生的常规和个性化数据;反馈数据包括顾客通过各种渠道的点评和反馈信息。

(2) 需求数据存储设计

在完成上一阶段的工作后,需要高速存储大量且多样的数据。本文将用于存储来自于内部和外部数据源需求信息的数据库称为主数据库(primary database, PDB),它可被用于同时存储结构化和非结构化的信息。为了分析历史顾客的需求

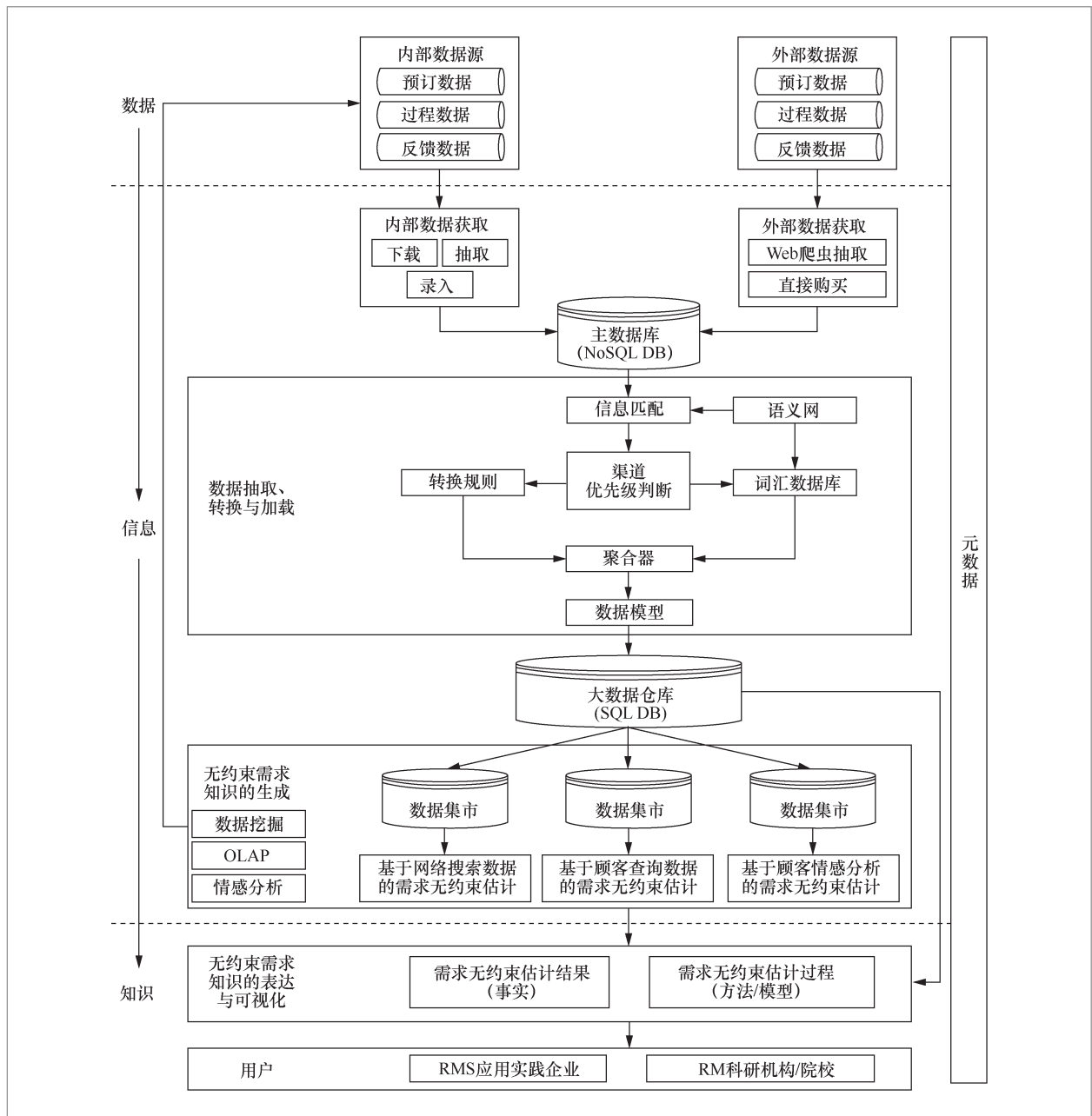


图3 基于大数据仓库的收益管理需求无约束估计框架

表现或预测未来需求的趋势，需求数据可以从PDB中获取，并聚集成一些单元化的信息。同时，该PDB也是用来创建和进行时间序列分析的基础。

(3) 需求数据集成设计

由于不同数据源中数据格式上的差异，需要采用不同的技术从PDB中提取

(extraction)、转换(transformation)和加载(load)相关信息，完成数据的ETL阶段。通过对数据源的识别，在提取过程中选择和获取相关数据。在需求数据转换过程中，需考虑PDB中所有相关数据的清理、标准化以及集成等任务，以支持分析工具。数据加载过程主要是指对需求数据进

行存储和更新。需求数据集成中的关键任务是异构数据源需求信息的提取,结构化和半结构化的需求信息通常通过语言、语义或基于约束的信息集成技术提取,而非结构化需求信息则由基于统计语言模型或自然语言处理的包装类以及文本挖掘方法获取^[33,34]。

(4) 大数据仓库设计

在以上阶段完成后,PDB中不同数据源的异构需求数据被映射到统一的数据格式,并被存储到一个二级数据库(secondary database, SDB)中,它包含了与顾客需求相关的所有数据,本文将该SDB称为BDW。为了能够精确支持RM需求无约束估计BI分析,该BDW允许大数据分析工具的开发和应用,其中包括数据挖掘、OLAP、情感分析(sentiment analysis)以及意见挖掘(opinion mining)。在考虑适合存储、分析和信息处理的数据模型基础上,选择与RM需求无约束估计工作最相关的顾客需求信息进行描述、预测和优化。在某些情况下,企业决策者也可考虑使用与某个部门或某项活动普遍相关的数据子集,其所构成的数据库被称为数据集市(data mart, DM),它既可依赖于BDW,也可独立于其之外。这样,企业便可通过考虑信息结构多维模型,并使用DM来满足RM需求无约束估计BI分析的需要。

(5) 无约束需求知识的生成

如前所述,RM预测和优化决策所需的无约束需求知识可主要通过数据挖掘、OLAP和情感分析等方法获取。数据挖掘方法包括人工智能算法在历史数据中发现需求知识的各种应用,同时对RMS实践中的不同策略实施结果进行预测^[35]。数据挖掘任务可分为如下两个部分:描述和预测。一方面,描述任务被认为是规则的认识,这些规则被用来特征化历史需求数

据,它同样包含一些技术,比如聚类和汇总。另一方面,预测任务则是指恰当地确定能够定义一个变量行为的新模型,它可被用来估计未来的变量值。需要指出的是,这里的预测任务是为RM需求无约束估计BI分析服务的,并且可与OLAP多维数据集一同用来探索BDW或者DM中的信息,它可站在不同的商业角度进行无约束需求信息挖掘。另外,还需对顾客情感方面的内容进行检测和处理,而情感分析也可被用来帮助解决由数据到知识的转换问题^[36-38]。

数据挖掘、OLAP和情感分析层面上由浅至深的转换、描述和分析等工作的顺利开展,为基于BDW的RM需求无约束估计、预测和优化决策分析体系的建立提供了依据和保障。这同时也涉及从注重数据的“精确化”到承认“不精确”、从追溯因果关系到依靠相关关系的大数据应用理念的转变,以求需求无约束估计结果更加接近RM实践中的真实情况。

(6) 无约束需求知识的应用

从以上各阶段获得的无约束需求数据、信息和知识形成了体量巨大的大数据链条。同时,将非结构化数据转化为结构化数据,使文本、图片、视频及其他媒介数据有效地流动、归类和整合,并最终形成顾客行为的画像数据库,包括不同粒度下的顾客无约束需求量、选择性概率和产品忠诚度等量化指标,实现信息呈现方式的可视化^[39],提高操作、分析和决策效率,规避运作风险,这些始终是大数据应用的技术重点^[40]。这些需求分析工具允许对那些生成企业管理报告或仪表板的无约束需求数据进行分析调查,这些报告和仪表盘可能以图形、文字和表格的形式来呈现,其中可包括从数据挖掘任务、交互式查询、关键绩效指标、多维数据集、平衡记分卡、需求无约束估计和预测方法等方面得到的各种结果。

4 面临的挑战

为实现大数据背景下RM需求无约束估计BI系统的应用要求,在上述所有阶段中采用适当的方法是非常重要的,基于本文所提BDW框架,对需求数据进行提取、存储、集成和应用时,还面临着如下3个方面的挑战。

4.1 外部网络资源中需求信息的提取

如前所述,提取外部网络资源的顾客需求数据对RM需求无约束估计具有重要意义,包括顾客的搜索方式、对各种营销策略的反应、购买竞争对手产品所支付的价格、预订或购买后对产品的评价等数据。同时,可将外部数据源分为结构化数据和非结构化数据,前者包括交易数据、问卷和评分数据等,而后者则可能是自由文本形式的网络评论以及包括网络视频数据在内的Web 2.0时代的各种丰富信息内容。RM应用行业市场营销人员需要获取大量的此类外部需求信息进行无约束估计、预测和制定优化决策,并通过使用BI分析技术促进自身产品以最优的价格卖给正确的顾客。

在基于网络的RM应用行业商业模式发展早期,外部的行业需求数据是可以从公司官方网站和OTA网站上直接获取的,为了促进产品销售,开放和公开这些数据是符合数据提供方利益的。但现如今,这一情况却迅速地发生着变化,基于互联网的外部资源信息不再免费,并且不再容易收集。一方面,电商网站不愿承担由于数据开放而使其服务器过载的后果,这将导致其正常的在线预售功能受到影响,它们试图通过签订基于应用程序编程接口

(application programming interface, API)的商业合同来确保自身的利益,这些API能使客户访问其网站引擎,并获取规定的数据库,但通常情况下,此类API商业合同并不便宜。另一方面,仅靠人工作业方式实时地从外部网络资源中提取大量数据几乎是不可能完成的任务。因此,上述两方面的因素为自动化信息提取技术的诞生创造了条件^[41]。

当网页的HTML结构保持不变时,开发能够从特定网站中自动分析语法和提取数据的机理并不困难。目前,已有针对利用Web爬虫(Web crawler)从外部网络信息系统中提取数据的一系列研究,Ferrara E等人^[42]对其应用和技术做了文献综述,其中仅有一小部分文献以B2C为主题进行了研究(可参见Ghobadi A等人^[43]的研究)。另外,如果网站管理员或程序员决定改变文档对象模型(document object model, DOM)树结构或标签属性,比如在开发过程中使用JavaScript^①和Ajax^②技术来执行与电子商务有关的交互活动,以方便人们交流,并可根据以往的输入将信息展示给顾客,在此情况下,就必须对数据的提取机理进行重新定义或调整,而该情况经常会发生在需要从中获取RM顾客需求信息的电商网站上。国内外已有针对上述问题的一系列研究成果(如Martins D等人^[41]、Baumgartner R等人^[44]、周岳骞等人^③、时永坤^[45]的研究),采用的Web爬虫工具包括WebDriver^④、XPath^⑤等。尽管如此,基于上述工具的数据提取不是一个完全无监督的过程,如果页面布局发生变化,还需要依靠人工监督协助完成。在未来,有必要结合RMS实施实践,开发适用于RM需求无约束估计的改进版Web爬虫,以便实现对页面布局修改的自动检测和“理解”以及在网络需求大数据提取的自动

① <https://www.javascript.com/>

② <http://api.jquery.com/category/ajax/>

③ <http://www.paper.edu.cn/releasepaper/content/201601-181>

④ <http://www.w3.org/TR/webdriver>

⑤ <http://www.w3.org/TR/xpath>

化、精确度和速度等方面的平衡^[46,47]。

4.2 结构化和非结构化需求信息的集成和存储

本文所提BDW框架的第一阶段主要针对使用Web爬虫工具从特定网页中进行数据提取,虽然不是所有被提取到的需求数据都能被用到RM需求无约束估计模型中,但这一过程却容易产生用于被存储的大量非结构化原始动态需求数据。本文所提框架的目标是使用多个Web爬虫工具从多个电子商务网站中同时提取数据,由于电商网站经常会添加和删除新字段,且每个网站提供者会使用不同的方式对相同的产品进行描述,提取过程针对不同网站的不同产品可能会得到不同的信息,因此,几乎不可能针对一系列的电商网站专门设计并维持一种具有合理结构的关系数据库(relational database, RDB)架构^[43]。为了存储这些信息,需要依靠动态存储数据库技术来实现, NoSQL数据库就是选择之一^[48]。由于从不同网站中提取到的数据不具有相同的设计,其所具有的非模式化属性在非结构化动态需求数据存储时就显得非常重要^[49]。

一旦搜集到来自于不同网络渠道中的非结构化数据, RM需求无约束估计BDW框架中的ETL过程就具有更高规模的复杂性。相同含义的数据会基于不同的名称被分类,这增加了整理提取需求信息的必要性,以保证存储在BDW中的信息对于需求无约束估计工作的一致性和可靠性。因此,有必要对PDB中的需求数据进行清理和转化,并在这之后将经过整理的需求数据加载到BDW中。这是BDW框架设计过程中面临的另一项挑战。

为了建立BDW,首先需要获取数据在PDB中的具体存储位置,以适当的方式

加载到整理程序的内存中。其次,如果需求信息是字符串格式,比如名称、日期、全球定位系统(global positioning system, GPS)坐标和价格等,有必要做一些转换工作,使来自于不同数据源需求信息的存储格式具有一致性。同时,还应设计一种算法将对使用Web爬虫从不同渠道中提取到的需求信息进行匹配,以避免由于PDB文件中ID信息分配机制而造成的对相同需求信息的重复存储和冲突问题。在对来自于不同渠道的信息进行取舍时,应事先确定各种渠道的优先级。如果是顾客评论类的自由文本需求信息,则需要借助不断更新的数据词典(data dictionary)、词汇数据库(lexical database)和语义网(semantic Web),根据上下文的语境解释其具体含义^[50-52]。由于每条评论代表一个不同的观点,并可能来自于不同的个体,因此,对评论类的需求信息不应进行配对与合并,而仅将其进行格式转换后存储^[17]。在不可能将出现在NoSQL数据库中的信息转化为结构化信息的情况下,在BDW中有必要考虑语义的概念,比如可以借助WordNet[®]工具对文本或语料库进行语义分析,以便进行一个适当的数据处理和转换过程,并且在之后将信息以一种适当的结构进行存储。

在完成上述工作后,为了建立需求信息之间的逻辑关系,需要建立它们之间的实体关系(entity-relationship, E-R)模型,将非结构化信息转化为结构化信息。比如,可转化为关系数据库模型(relational database model, RDBM),并遵守一定的数据存储顺序和路由规则,将PDB中的数据以“多对多”以及类似的关系存储到BDW的表格中。因此,为了便于RM需求无约束估计BI系统的最后使用, BDW应该被设计为具有规范格式和存储规则的关系数据库。最后,除了整合从不同数据源中收集到的信息以外,一旦

⑥
<http://wordnet.princeton.edu>

与RM需求无约束估计相关的历史数据和信息发生变化,有必要对它们进行实时的更新与维护。

4.3 收益管理需求无约束估计商务智能应用

作为大数据背景下需求竞争情报评估的一项基础性工作,在能够获取外部数据源需求信息的基础上,RM需求无约束估计BI应用分析面临着如下新的挑战。

(1) 基于网络搜索数据的需求无约束估计

随着大数据相关领域的发展,针对网络搜索数据与社会经济活动之间相关性的研究也随之广泛开展。其中,来自国内外的系列研究均验证了网络搜索数据对旅游需求量的时空动态变化的解释能力,比如基于谷歌和百度指数等建立旅游需求的预测模型^[53-61]。但是,现有研究仅证明了网络搜索数据与历史可观察旅客需求量之间的相关性,并未涉及旅客网络搜索数据与历史无约束需求量之间的相关关系,即未从RM需求无约束估计的角度进行研究,而无约束需求数据是任何形式需求预测的数据基础。因此,为了提高基于此类方法进行预测的准确性和有效性,有必要对基于网络搜索数据的需求无约束估计进行研究,并建立相应的计量经济无约束估计模型。

(2) 基于顾客查询数据的需求无约束估计

由本文所提BDW框架可知,从外部数据源获取的需求信息既可以是Web爬虫或网络机器人从网页中提取的顾客网络查询记录,也可以是从第三方大数据提供商(比如GDS和CRS)获取的顾客查询日志。相较于从内部数据源获取的实时客流数据,上述两类顾客查询数据更具预知性。周岳骞^[62]对航空客运行业的研究表明,相

较于传统的基于历史客运量需求模型,在对突发性事件需求预测方面,基于GDS查询日志数据的需求指数模型在准确性和敏感性方面都表现出很大优势。今后的发展趋势应在充分考虑顾客选择行为的基础上,结合顾客选择模型探索基于顾客查询数据的需求无约束估计方法,在提高精度的基础上进行实时的RM需求无约束估计和预测。

(3) 基于顾客情感分析的需求无约束估计

如前所述,情感分析或意见挖掘技术可能是本文设计框架中最大的挑战之一,它包括自然语言处理、计算语言学和文本挖掘,是一组处理意见数据,并试图从中获取有价值信息的一类技术^[63]。语义是PDB文本字段内容信息挖掘的关键之一,同时,还需考虑顾客的评价数据以及怎样将情感分析应用到与组织相关的产品情报中。这些都是意见挖掘过程中具有挑战性的主题,也是很难实现的部分,在所设计的框架中主要体现在将PDB中的信息存储到BDW的阶段,而后者对基于顾客情感分析的RM需求无约束估计的影响会更大。如何通过电商平台和社交网络顾客评价中的情感信息与真实的历史顾客需求建立联系,即如何判断顾客的积极和消极观点对最终顾客预定行为的影响,并在此基础上,结合机器学习和人工智能算法进行实时的需求无约束估计,是大数据RM今后的研究趋势之一。

5 结束语

随着信息技术的发展和成功应用,如今的互联网时代被称作大数据时代,其应用核心是大数据分析,这是一个令人激动的新研究领域。管理学领域在大数据背景

下的变革主要体现在公共管理、工商管理和管理工程3个方面^[64]。运筹与决策科学是大数据BI的一个核心概念,作为运筹学(Operations Research, OR)中快速发展的重要分支之一,同时也是管理科学、计算机技术等在实际应用中最为成功的学科之一,RM基于对市场动态的理解,建立相应的策略,预测和引导顾客的行为,并从固定的资源中获取最大化的收入和利润。实施RM需求无约束估计、预测和最优决策所需的数据量是巨大的,并且具有实时、高速、多样和复杂等特性,这恰恰说明了采用目前被称为BDW概念的必要性。

大数据分析能够帮助航空、酒店和汽车租赁等RM应用行业更好地了解自身在竞争市场中的优势和劣势,考虑到基于网络的现代服务市场环境变化迅速,公司所有者需要实时的可视化业务信息来检测竞争者情况,并需要对遇到的问题进行自动化的实时解答,在此情况下,BDW就是一个必要的信息系统。本文针对RM需求无约束估计问题提出了相应的BDW框架,从不同外部数据源中高速搜集到大量信息,连同组织内部的数据一起构成了BDW。本文的研究能帮助RM应用行业在精细化地洞察顾客历史真实购买行为的基础上,对市场需求情况进行实时无约束估计和预测,发现细微的模式和关联,据此匹配顾客需求,优化产品的设计、成本结构和维护监控,并提供个性化、差异化服务,快速地响应市场,开展精准营销,形成大数据生产力。

与开发一个信息系统所需要的概念类似,如今开发和应用一个BDW也已成为了可能,但是这需要传统数据库与外部数据资源的融合,因此有必要考虑一些新技术,比如Web爬虫、NoSQL数据库、数据词典、词汇数据库、语义网络和情感分析等。另一方面,由于需求无约束估计问题在RM理论与实践扮演着重要角色,在大数据

背景下面临着如何基于网络搜索和查询数据以及情感分析等信息对顾客历史“初始需求”进行实时精细化估计的挑战。基于这些挑战而开发的RM大数据分析工具,其对BI的成功实施和应用具有重要的支撑作用,符合当今企业和社会的发展趋势。

在有缺货情况的库存管理、存在替代产品的零售业生产规划、无线通信行业顾客需求估计与存量计划以及基于需求预测的现代旅游服务供应链管理等情形的研究领域,如何准确估计顾客需求“溢出”和“再现”效应的问题同样受到了普遍关注^[40, 61, 65-67]。因此,本文所提基于BDW的RM需求无约束估计框架不仅适用于传统RM应用行业,其运行机理还可被一般化到上述其他行业中。另外,由于RM需求无约束估计在本质上属于寿命数据(lifetime data)问题处理的范畴^[68-70],本文所提框架在未来衍生出的新方法也可被应用于需要广泛收集数据的生命可靠性测试、医学临床实验、疾病预测和医疗保健等领域。

参考文献:

- [1] CROSS R G. Revenue management: hard-core tactics for market domination [M]. New York: The Broadway Books, 1997.
- [2] GUO P, XIAO B C, LI J. Unconstraining methods in revenue management systems: research overview and prospects[J]. Advances in Operations Research, 2012, 2012(270910): 1-23.
- [3] AZADEH S S, MARCOTTE P, SAVARD G. A taxonomy of demand uncensoring methods in revenue management[J]. Journal of Revenue and Pricing Management, 2014, 13(6): 440-456.
- [4] WEATHERFORD L. The history of unconstraining models in revenue management[J]. Journal of Revenue and Pricing Management, 2016, 15(3): 222-228.

- [5] WEATHERFORD L. The history of forecasting models in revenue management[J]. *Journal of Revenue and Pricing Management*, 2016, 15(3): 212-221.
- [6] 黄宜华. 大数据机器学习系统研究进展[J]. *大数据*, 2015, 1(1): 2015004.
HUANG Y H. Research progress on big data machine learning system [J]. *Big Data Research*, 2015, 1(1): 2015004.
- [7] 孟小峰, 慈祥. 大数据管理: 概念、技术与挑战[J]. *计算机研究与发展*, 2013, 50(1): 146-169.
MENG X F, CI X. Big data management: concepts, techniques and challenges[J]. *Journal of Computer Research and Development*, 2013, 50(1): 146-169.
- [8] 李国杰, 程学旗. 大数据研究: 未来科技及经济社会发展的重大战略领域——大数据的研究现状与科学思考[J]. *中国科学院院刊*, 2012, 27(6): 5-15.
LI G J, CHENG X Q. Research status and scientific thinking of big data[J]. *Bulletin of Chinese Academy of Sciences*, 2012, 27(6): 5-15.
- [9] SCHERMANN M, KRCCMAR H, HEMSEN H, et al. Big data - an interdisciplinary opportunity for information systems research[J]. *Business & Information Systems Engineering*, 2014, 6(5): 261-266.
- [10] 徐宗本, 冯芷艳, 郭迅华, 等. 大数据驱动的管理与决策前沿课题[J]. *管理世界*, 2014, 30(11): 158-163.
XU Z B, FENG Z Y, GUO X H, et al. Leading issues in the big data driven management and decision making[J]. *Management World*, 2014, 30(11): 158-163.
- [11] XUAN L W, HEO C Y, SCHWARTZ Z, et al. Revenue management: progress, challenges, and research prospects[J]. *Journal of Travel & Tourism Marketing*, 2015, 32(7): 797-811.
- [12] 张建城, 朱金福. 航空公司收益管理预测系统的总体设计[J]. *中国科技信息*, 2004, 15(24): 107.
ZHANG J C, ZHU J F. The overall design of airline revenue management forecasting system[J]. *China Science and Technology Information*, 2004, 15(24): 107.
- [13] 冯兴杰. 基于数据仓库的收益管理系统结构设计[J]. *中国民航学院学报*, 2003, 21(S2): 122-126.
FENG X J. Structure design of revenue management system based on data warehouse[J]. *Journal of Civil Aviation University of China*, 2003, 21(S2): 122-126.
- [14] 王红. 航空公司客运收益管理系统数据仓库的设计[J]. *计算机应用与软件*, 2004, 21(6): 49-50.
WANG H. The design of data warehousing of the civil aviation revenue management system[J]. *Computer Applications and Software*, 2004, 21(6): 49-50.
- [15] 李丽娟, 夏洪山. 收益管理系统数据仓库模型建立研究[J]. *江苏航空*, 2007, 26(3): 5-7.
LI L J, XIA H S. The data warehousing modeling of revenue management system[J]. *Jiangsu Aviation*, 2007, 26(3): 5-7.
- [16] 满青珊, 张金隆, 韦司滢, 等. 区域航空收益管理数据仓库设计与实现[J]. *情报杂志*, 2009, 28(12): 122-126.
MAN Q S, ZHANG J L, WEI S Y, et al. Design and implement of regional airline revenue management data warehouse[J]. *Journal of Intelligence*, 2009, 28(12): 122-126.
- [17] RAMOS C, CORREIA M, RODRIGUES J, et al. Big data warehouse framework for smart revenue management[C]// *Naun International Conference on Management, Marketing, Tourism, Retail, Finance and Computer Applications*, June 10, 2015, Tenerife, Canary Islands, Spain. [S.l.:s.n.], 2015: 13-22.
- [18] COOPER W L, HOMEM-DE-MELLO T, KLEYWEGT A J. Models of the spiral-down effect in revenue management[J]. *Operations Research*, 2006, 54(5): 968-987.
- [19] VULCANO G, RYZIN G V, RATLIFF R. Estimating primary demand for substitutable products from sales transaction data[J]. *Operations Research*, 2012, 60(2): 313-334.

- [20] RYZIN G V, VULCANO G. A market discovery algorithm to estimate a general class of nonparametric choice models[J]. *Management Science*, 2015, 61(2): 281-300.
- [21] 罗利, 萧柏春. 收入管理理论的研究现状及发展前景[J]. *管理科学学报*, 2004, 7(5): 75-83.
LUO L, XIAO B C. Revenue management: state-of-the-art and future prospects [J]. *Journal of Management Sciences in China*, 2004, 7(5): 75-83.
- [22] SKWAREK D K. Revenue and traffic impacts of alternative detruncation methods[C]// AGIFORS Reservations and Yield Management Study Group, October 11, 1996, Zurich, Switzerland. [S.l.:s.n.], 1996.
- [23] WEATHERFORD L R, PÖLT S. Better unconstraining of airline demand data in revenue management systems for improved forecast accuracy and greater revenues[J]. *Journal of Revenue and Pricing Management*, 2002, 1(3): 234-254.
- [24] WEATHERFORD L R. Improved revenues from various unconstraining methods in a passenger origin-destination simulator (PODS) environment with semi-restricted fares [J]. *Journal of Revenue and Pricing Management*, 2013, 12(1): 60-82.
- [25] PÖLT S. From bookings to demand: the process of unconstraining [C]// AGIFORS Reservations and Yield Management Study Group, October 5, 2000, New York, NY, USA. [S.l.:s.n.], 2000.
- [26] WEATHERFORD L R. Unconstraining methods [C]// AGIFORS Reservations and Yield Management Study Group, October 5, 2000, New York, NY, USA. [S.l.:s.n.], 2000.
- [27] RUD O P. Business intelligence success factors tools for aligning your business in the global economy[M]. Hoboken N J: Wiley & Sons, 2009.
- [28] 乔向杰, 张凌云. 知识型旅游目的地管理平台框架及其构建[J]. *人文地理*, 2014, 138(4): 104-110.
QIAO X J, ZHANG L Y. Framework and construction of a knowledgeable destination[J]. *Human Geography*, 2014, 138(4): 104-110.
- [29] 程学旗, 靳小龙, 王元卓, 等. 大数据系统和分析技术综述[J]. *软件学报*, 2014, 25(9): 1889-1908.
CHENG X Q, JIN X L, WANG Y Z, et al. Survey on big data system and analytic technology[J]. *Journal of Software*, 2014, 25(9): 1889-1908.
- [30] 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 Z Y, GUO X H, ZENG D J, 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.
- [31] MOHANTY S, JAGADEESH M, SRIVATSA H. Big data imperatives: enterprise 'big data' warehouse, 'BI' implementations and analytics[M]. California: Apress, 2013.
- [32] TRIA F D, LEFONS E, TANGORRA F. Big data warehouse automatic design methodology[J]. *Big Data Management, Technologies, and Applications*, 2014(2): 115-149.
- [33] FUCHS M, ABADZHIEV A, SVENSSON B, et al. A knowledge destination framework for tourism sustainability: a business intelligence application from Sweden [J]. *Tourism*, 2013, 6(2): 121-148.
- [34] FUCHS M, HÖPKEN W, LEXHAGEN M. Big data analytics for knowledge generation in tourism destinations - a case from Sweden[J]. *Journal of Destination Marketing & Management*, 2014, 3(4): 198-209.
- [35] 李涛, 曾春秋, 周武柏, 等. 大数据时代的数据挖掘——从应用的角度看大数据挖掘[J]. *大数据*, 2015, 1(1): 2015041.
LI T, ZENG C Q, ZHOU W B, et al. Data mining in the era of big data: from the application perspective [J]. *Big Data Research*, 2015, 1(1): 2015041.
- [36] 张紫琼, 叶强, 李一军. 互联网商品评论情感分析研究综述[J]. *管理科学学报*, 2010, 13(6): 84-96.
ZHANG Z Q, YE Q, LI Y J. Literature

- review on sentiment analysis of online product reviews [J]. Journal of Management Sciences in China, 2010, 13(6): 84-96.
- [37] 徐健. 基于网络用户情感分析的预测方法研究[J]. 中国图书馆学报, 2013, 39(3): 96-107.
XU J. Research on predicting methods based on network user sentiment analysis[J]. Journal of Library Science in China, 2013, 39(3): 96-107.
- [38] 涂海丽, 唐晓波. 基于在线评论的游客情感分析模型构建[J]. 现代情报, 2016, 36(4): 70-77.
TU H L, TANG X B. Tourist sentiment analysis model building based on online reviews [J]. Journal of Modern Information, 2016, 36(4): 70-77.
- [39] 任磊, 杜一, 马帅, 等. 大数据可视分析综述[J]. 软件学报, 2014, 25(9): 1909-1936.
REN L, DU Y, MA S, et al. Visual analytics towards big data [J]. Journal of Software, 2014, 25(9): 1909-1936.
- [40] 陈涛, 李佼. 基于大数据的旅游服务供应链管理研究[J]. 电子政务, 2013, 132(12): 32-40.
CHEN T, LI J. Study on the tourism service supply chain management based on big data [J]. E-Government, 2013, 132(12): 32-40.
- [41] MARTINS D, LAM R, RODRIGUES J, et al. A Web crawler framework for revenue management[C]//The 14th Int. Conf. on Artificial Intelligence, Knowledge Engineering and Data Bases (AIKED '15), January 10, 2015, Tenerife, Canary Islands, Spain. [S.l.:s.n.], 2015: 88-97.
- [42] FERRARA E, DE MEO P, FIUMARA G, et al. Web data extraction, applications and techniques: a survey [J]. Knowledge-Based Systems, 2014, 70(C): 301-323.
- [43] GHOBADI A, RAHGOZAR M. An ontology based semantic extraction approach for B2C e-commerce [J]. The International Arab Journal of Information Technology, 2011, 8(2): 163-170.
- [44] BAUMGARTNER R, CERESNA M, LEDERMULLER G. Deep web navigation in web data extraction [C]// International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce, November 28-30, 2005, Vienna, Austria. New Jersey: IEEE Press, 2005: 698-703.
- [45] 时永坤. 基于WebDriver的定向网络爬虫设计与实现[J]. 软件, 2016, 37(9): 94-97.
SHI Y K. Design and implement of WebDriver based crawler [J]. Computer Engineering & Software, 2016, 37(9): 94-97.
- [46] 王元卓, 靳小龙, 程学旗. 网络大数据: 现状与展望[J]. 计算机学报, 2013, 36(6): 1125-1138.
WANG Y Z, JIN X L, CHENG X Q. Network big data: present and future[J]. Chinese Journal of Computers, 2013, 36(6): 1125-1138.
- [47] QIU T, YANG T. Automatic information extraction from e-commerce web sites [C]//The International Conference on E-Business and E-Government, ICEE 2010, May 7-9, 2010, Guangzhou, China. New Jersey: IEEE Press, 2010: 1399-1402.
- [48] 申德荣, 于戈, 王习特, 等. 支持大数据管理的NoSQL 系统研究综述[J]. 软件学报, 2013, 24(8): 1786-1803.
SHEN D R, YU G, WANG X T, et al. Survey on NoSQL for management of big data[J]. Journal of Software, 2013, 24(8): 1786-1803.
- [49] REDMOND E, WILSON J R. Seven databases in seven weeks: a guide to modern databases and the NoSQL movement [M]. Dallas: Pragmatic Bookshelf, 2012.
- [50] 程学旗, 兰艳艳. 网络大数据的文本内容分析[J]. 大数据, 2015, 1(1): 2015029.
CHENG X Q, LAN Y Y. Text content analysis for web big data [J]. Big Data Research, 2015, 1(1): 2015029.
- [51] 袁书寒, 向阳, 鄂世嘉. 基于特征学习的文本大数据内容理解及其发展趋势[J]. 大数据, 2015, 1(1): 2015030.
YUAN S H, XIANG Y, E S J. Text big data content understanding and development

- trend based on feature learning [J]. *Big Data Research*, 2015, 1(1): 2015030.
- [52] ZHENG X, SCHWARTZ Z, GERDES J H, et al. What can big data and text analytics tell us about hotel guest experience and satisfaction [J]. *International Journal of Hospitality Management*, 2015, 44(44): 120-130.
- [53] PAN B, WU D C G, SONG H. Forecasting hotel room demand using search engine data[J]. *Journal of Hospitality and Tourism Technology*, 2012, 3(3): 196-210.
- [54] YANG Y, PAN B, SONG H. Predicting hotel demand using destination marketing organization's web traffic data [J]. *Journal of Travel Research*, 2014, 53(4): 433-447.
- [55] YANG X, PAN B, EVANS J A, et al. Forecasting Chinese tourist volume with search engine data [J]. *Tourism Management*, 2015, 46: 386-397.
- [56] BANGWAYO-SKEETE P F, SKEETE R W. Can Google data improve the forecasting performance of tourist arrivals? mixed-data sampling approach[J]. *Tourism Management*, 2015, 46: 454-464.
- [57] 马丽君, 孙根年, 黄云玛, 等. 城市国内客流量与游客网络关注度时空相关分析[J]. *经济地理*, 2011, 31(4): 680-685.
- MA L J, SUN G N, HUANG Y M, et al. A correlative analysis on the relationship between domestic tourists and network attention [J]. *Economic Geography*, 2011, 31(4): 680-685.
- [58] 龙茂兴, 孙根年, 马丽君, 等. 区域旅游网络关注度与客流量时空动态比较分析——以四川为例[J]. *地域研究与开发*, 2011, 30(3): 93-97.
- LONG M X, SUN G N, MA L J, et al. An analysis on the variation between the degree of consumer attention of travel network and tourist flow in regional tourism: a case of Sichuan province [J]. *Areal Research and Development*, 2011, 30(3): 93-97.
- [59] 黄先开, 张丽峰, 丁于思. 百度指数与景区游客量的关系及预测研究——以北京故宫为例[J]. *旅游学刊*, 2013, 28(11): 93-100.
- HUANG X K, ZHANG L F, DING Y S. Study on the predictive and relationship between tourist attractions and the Baidu index: a case study of the forbidden city [J]. *Tourism Tribune*, 2013, 28(11): 93-100.
- [60] 王炼, 贾建民. 基于网络信息搜索的旅游需求预测——来自黄金周的证据[J]. *系统管理学报*, 2014, 23(3): 362-368.
- WANG L, JIA J M. Forecasting of tourism demand in the golden week with online information search [J]. *Journal of Systems & Management*, 2014, 23(3): 362-368.
- [61] 陈涛, 刘庆龙. 智慧旅游背景下的大数据应用研究: 以旅游需求预测为例 [J]. *电子政务*, 2015, 153(9): 6-13.
- CHEN T, LIU Q L. Research on the application of big data in the context of intelligent tourism: a case study of tourism demand forecasting [J]. *E-Government*, 2015, 153(9): 6-13.
- [62] 周岳骞. 面向民航需求建模的用户查询日志大数据分析方法与实现[D]. 北京: 北京交通大学, 2016.
- ZHOU Y Q. Research on query log data of civil aviation for passenger demand modeling[D]. Beijing: Beijing Jiaotong University, 2016.
- [63] MARRESE-TAYLOR E, VELÁSQUEZ J D, BRAVO-MARQUEZ F. A novel deterministic approach for aspect-based opinion mining in tourism products reviews [J]. *Expert Systems with Applications*, 2014, 41(17): 7764-7775.
- [64] 吴启迪. 大数据环境下管理科学领域相关问题探讨[J]. *科技促进发展*, 2014, 10(1): 9-14.
- WU Q D. A discussion on management science issues under the big data environment [J]. *Science & Technology for Development*, 2014, 10(1): 9-14.
- [65] JAIN A, RUDI N, WANG T. Demand estimation and ordering under censoring: stock-out timing is almost all you need [J]. *Operations Research*, 2015, 63(1): 134-150.
- [66] FISHER M, VAIDYANATHAN R. A

- demand estimation procedure for retail assortment optimization with results from implementations [J]. *Management Science*, 2014, 60(10): 2401-2415.
- [67] MA X, DENG T, LAN B. Demand estimation and assortment planning in wireless communications [J]. *Journal of Systems Science and Systems Engineering*, 2016, 25(4): 398-423.
- [68] LIU P H, SMITH S, ORKIN E B, et al. Estimating unconstrained hotel demand based on censored booking data[J]. *Journal of Revenue and Pricing Management*, 2002, 1(2): 121-138.
- [69] LAWLESS J F. Statistical models and methods for lifetime data [M]. New York: John Wiley & Sons, 2003.
- [70] KLEIN J P, MOESCHBERGER M L. Survival analysis [M]. New York: Springer, 2005.

作者简介



郭鹏(1984-),男,博士,贵阳学院经济管理学院讲师,主要研究方向为收益管理理论与方法、需求无约束估计和预测、系统仿真优化、大数据分析。2006年起至今,从事有关收益管理、系统仿真优化方面的研究,作为第一作者在《Advances in Operations Research》《系统工程理论与实践》《系统科学与数学》《数理统计与管理》《计算机仿真》等期刊上发表十余篇论文,主持和参与了多项课题;2010—2014年参与完成了国家自然科学基金委员会重大项目(No.71090402);2015年主持并完成了贵阳市科协软科学研究项目(No.2015B23),同年主持了国家自然科学基金一般项目(No.15BGL198);2016年主持了贵州省教育厅高校人文社会科学研究自筹项目(No.2016ZC021)、贵阳市科学技术协会软科学研究项目(No.2016A01)和贵阳学院院级科研项目(No.GYXY[2016]24)。

收稿日期:2016-11-29

基金项目:国家自然科学基金资助项目(No.15BGL198);贵州省教育厅高校人文社会科学研究基金资助项目(No.2016ZC021);贵阳市科学技术协会软科学研究基金资助项目(No.2016A01)

Foundation Items: The National Social Science Foundation of China (No.15BGL198), The Colleges & Universities Humanities and Social Sciences Research Project of Educational Commission of Guizhou Province of China (No.2016ZC021), Soft Science Research Project of Guiyang Association of Science and Technology (No.2016A01)