

决策驱动型省级政务大数据治理技术架构及实现

孟凡^{1,2}, 杨群力¹, 高阳²

1. 江苏省战略与发展研究中心, 江苏 南京 210036;

2. 计算机软件新技术国家重点实验室(南京大学), 江苏 南京 210093

摘要

新时期政务数据在形态、产生机制等方面发生了巨大转变, 呈现模态多样、体量巨大、结构复杂、时序演变等特征, 给传统治理带来了全新的挑战。针对此问题, 以省级政务大数据治理为切入点, 分析了传统省级政务大数据治理的技术架构和瓶颈。围绕政府决策场景提出了一种决策驱动型省级政务大数据治理技术架构——凤翔模型, 分析其内涵概念、设计原则和总体架构, 并对其可行性进行论证。针对凤翔模型技术架构中的4项关键业务问题和关键技术问题, 进一步提出针对性的技术实现思路, 为各省市后续治理项目建设以及相关技术路线研究选型提供参考。

关键词

政务大数据; 数据治理; 决策; 复杂结构

中图分类号: TP391

文献标志码: A

doi: 10.11959/j.issn.2096-0271.2023015

Technical architecture and implementation of decision-driven provincial government big data governance

MENG Fan^{1,2}, YANG Qunli¹, GAO Yang²

1. Jiangsu Strategy and Development Research Center, Nanjing 210036, China

2. State Key Laboratory for Novel Software Technology at Nanjing University, Nanjing 210093, China

Abstract

In the new era, government data has undergone tremendous changes in terms of form and generation mechanism, presenting multi-modality, huge volume, complex structure, and time series evolution, which has brought new challenges to traditional governance. This paper takes provincial government big data governance as the starting point, and mainly includes three aspects: (1) analyzed the technical framework and bottlenecks of traditional provincial government big data governance; (2) based on the government decision-making scenario, proposed a decision-driven provincial government big data governance technology architecture (called Fengxiang model), introduced its components, design principles, overall architecture, and demonstrated its feasibility; (3) aiming at the four key business

issues and key technical issues in the technical framework of Fengxiang model, further proposed targeted technical implementation ideas to provide a reference for the construction of follow-up governance projects in various provinces and cities and the research and selection of related technical routes.

Key words

government big data, data governance, decision making, complex structure

0 引言

大数据、人工智能等新一代信息技术已对人类经济社会的发展产生了深刻影响和变革。党中央、国务院围绕如何充分利用新一代前沿技术赋能政府决策科学化、公共服务便捷化和行政监管智能化相继出台多项政策文件。党中央、国务院对加快推进全国一体化政务大数据体系建设进行了重要部署,发布了《全国一体化政务大数据体系建设指南》,要求各省各地积极开展政务大数据体系相关体制机制和应用服务创新,增强数字政府效能,为推进国家治理体系和治理能力现代化提供有力支撑。面对跨部门、跨区域、跨层级的信息来源,新时期政务数据的形态和产生机制已发生深刻变化,形成了多源、异质、时变、隐匿的复杂政务大数据^[1-5]。因此,提出切实有效的数据治理架构和方法是构建高质量政务大数据的基础保障,更是为各级政府部门提供决策参考的必要前提。

传统意义上,数据治理通常被定义为一种对数据质量控制的规程^[6-7]。由于传感、计算和存储能力的迅猛发展,新时期政务大数据存在模态多样、体量巨大、结构复杂、时序演变等特点,给传统数据治理带来了全新挑战。因此,相比于传统治理方式,对政务大数据的治理不但需要考虑到数据时效性、合规性、安全性等基本指

标,更要以政府业务工作目标为驱动,直面前跨部门、跨地域、跨层级数据在归集、融合、校验等过程中的各项困难与挑战,进而实现为精准化社会治理提供决策支撑的核心目标。

本文以省级政务大数据治理的现状分析为切入点,首先,分析了传统政务大数据治理的主要技术架构和瓶颈;其次,提出了一种决策驱动型省级政务大数据治理技术架构(以下简称凤翔模型),分析其内涵定义、设计原则和总体架构,并对其可行性进行论证;最后,结合凤翔模型技术架构中的关键业务问题和技术问题,进一步提出技术实现思路,为各省市后续项目建设以及相关技术路线选型研究提供参考。

1 省级政务大数据治理现状

1.1 数据治理通用定义

根据国家标准《信息技术大数据术语》(GB/T 35295-2017)中的内容,数据治理被定义为对数据进行处置、格式化和规范化的过程,涉及对数据全生命周期的管理,无论数据处于静态、动态、未完成状态,还是交易状态^[8]。各行业各领域依据以上数据治理的通用定义进行针对性研究和扩展。本节主要分析省级政府部门对政务大数据进行治理的过程中采用的传统技术架构和存在的瓶颈。

1.2 传统政务大数据治理技术架构

在数据治理通用定义的基础上,传统政务大数据治理主要是指对政府各部门各行业产生的政务数据,通过制定自上而下的大数据管理制度和流程,对数据进行采集、清洗、加工、共享和管理,实现对政府各部门“数据烟囱”的横向融合^[9-12]。

从技术流程角度而言,传统政务大数据治理主要关注管控数据的合规性、安全性、完整性和准确性等方面,以此建立高质量政务数据库,为各部门在市场监管、行政决策和公共服务等方面提供数据基础。本文从数据治理工作的开展过程进行划分,总结出传统政务大数据治理的两个阶段,如图1所示。

(1) 阶段一: 数据治理规划阶段

该阶段主要针对数据治理实施工作进行前期规划和设计,具体包括针对待归集数据制定数据资源目录标准、制定信息安全管理规范、制定数据清洗规则以及制定

数据质量评估指标。

(2) 阶段二: 数据治理实施阶段

该阶段主要根据规划阶段确定的各项治理规则和措施进行技术层面的研发和实施。省级政务大数据治理平台通常包括4个部分:围绕已归集的政务数据进行归集配置、数据清洗、主题加工以及与外部系统的交换共享。

1.3 传统政务大数据治理瓶颈

本文在总结传统政务大数据治理技术架构的基础上,进一步分析其中存在的4个主要瓶颈,具体如下。

(1) 目标导向性弱

传统政务大数据治理过程与治理后的预期业务目标联系不紧密,缺乏能够针对具体政务应用场景的治理过程管理。传统数据治理过程虽然也在强调对数据质量的管控和提升,但往往以追求通用的数据质量评估标准为手段,与最终行政业务目标的耦合度不高,对实际工作的指导性往往不明显。

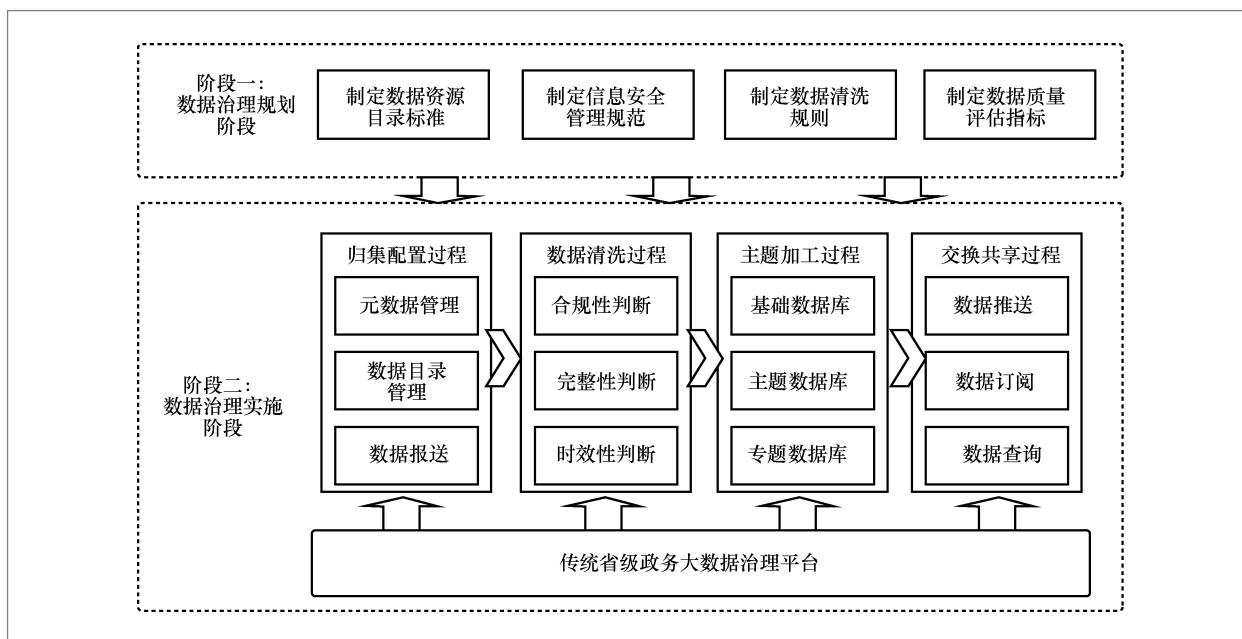


图1 传统政务数据治理流程

(2) 专家依赖度高

传统政务大数据治理过程在制定治理规则时,往往需要领域专家深度参与治理前期和实施过程中的各阶段,甚至需要他们制定出具体的数据治理措施和评价指标。总体而言,该过程对专家的依赖度较高,需投入较多行政资源对数据治理过程进行定制化。

(3) 技术拓展性差

因为传统政务大数据治理过程缺乏目标导向性并对专家依赖程度高,所以治理方在依据制定的数据治理措施进行系统功能研发和实施后,若治理流程中出现一些局部调整,从信息化系统层面而言,相关功能将面临较大改造甚至重构,进一步提高了各级政府部门人力、物力和财力投入。

(4) 过程协同性低

面对新时期复杂的政务大数据,数据涉及的市场主体等行政管理对象的经营等活动过程和行为被记录得更充分,这使数据治理工作的实施主体由原来的单一主体向多主体协作转变,即一个数据治理流程往往需要多方共同参与。具体而言,如数据清洗环节,从不合规数据的发现到反馈再到修正是一个多方参与的闭环流程,参与方往往包括数据使用方(发现疑问数据)、数据治理方(反馈疑问情况)、数据源头方(核实处理疑问反馈)等。这种多方协同治理的业务模式,对政务大数据治理协同机制提出了更高的要求。

2 凤翔模型

2.1 模型内涵

基于前文对传统政务大数据治理过程的分析,为进一步解决其面临的4个主要瓶颈,本文提出了一种决策驱动型省级政

务大数据治理模型,简称为凤翔模型。该模型可归纳为“三一二五”治理架构,即以支撑和服务政府宏观区域态势、中观行业监管、微观行政对象审批三方面决策场景为核心驱动力,在明确一个待决策目标任务的基础上,通过建立基础资源管理机制和行业动态知识管理两项基础工作,经过决策量化、表示学习、在线任务管理、检验分析、质量评估五大核心环节,共同构成了面向省级政务大数据治理的总体技术架构。

2.2 设计原则

本文在传统数据治理的基础上,提出决策驱动型政务大数据治理架构的四大基本设计原则,具体包括数据治理过程的决策导向性、任务可量化性、领域参考性、高效集约性,强调了数据治理需要与实际的政务决策场景相结合,避免陷入为了治理而治理的情况。

(1) 决策导向性

明确数据治理的业务目标是开展治理工作的第一步,其中政务数据治理的核心价值之一是服务于政府社会治理和公共服务工作。具体而言,本文提出的数据治理模型是为了满足具体政府决策工作而开展的针对性工作,数据治理本身并非目标,而是手段。

(2) 任务可量化性

在明确了待决策目标的基础上,治理各环节需能够依据给出的决策业务目标,进一步明确具体的决策任务,并对其进行量化表示和形式化定义,以便为后续具体的治理环节建模提供基础。

(3) 领域参考性

由于政府各行业行政部门有自身的业务规则 and 标准,这些业务规则 and 标准是各部门经过长期实践积累的宝贵经验。因此,在开展数据治理的过程中,需充分结合政务领域

的专业知识,为最终的决策目标提供助力。

(4) 高效集约性

决策目标往往具有时效性要求,因此治理过程需在规定时间内准确完成,即围绕最终待决策的业务目标,要求各治理环节及时、准确完成。同时,省级政务大数据治理平台需采用集约建设模式,由省级数据治理部门集约建设,减少各地各部门行政资源的重复投入。

2.3 总体架构

针对以上定义和设计基本原则,本文提出的凤翔模型主要包括四部分:一是“三类”决策场景(凤首);二是“一个”决策目标任务(凤体);三是“两项”核心基础(凤翼);四是“五大”核心环节(凤尾)。具体过程如图2所示。

(1) 凤首:“三类”决策场景

模型首先明确了政务决策的3类场景:一是宏观区域态势决策,具体是指政府部门对本辖区内的特定区域,从宏观态势上对经济走向等方面进行分析和预警,为部门宏观政策制定提供决策参考;二是中观行业监管决策,具体是指从政府部门所属条线角度出发,通过数据治理为本行业内的市场主体监管提供决策参考;三是微观行政审批决策,具体是指本部门以法定行权审批为依据,对数据进行治理后,为行政相对人进行具体审批工作时提供决策参考。

(2) 凤体:“一个”决策目标任务

模型中决策目标是指以待决策场景为依据,进一步明确决策目标中具体决策任务。从政府决策目标角度而言,可将决策目标进一步抽象为3类任务情况:一是数据预测任务,如对待决策目标或事件计算出具体的量化得分;二是分类任务,如对待决策目标或事件计算出是否符合要求的二元判断;三是决策任务,即基于现有部门的

制度文件,利用在实践中不断交互形成的政务数据资源,在分析挖掘的基础上及时动态更新施政策略,实现对现有制度文件的优化和完善。

(3) 凤翼:“两项”核心基础

模型中定义了治理过程的两项基础:一是政务大数据基础资源管理,包括数据目录管理、元数据管理、数据采集管理和信息安全管理;二是政务领域动态知识图谱管理,主要是指在决策中要充分利用已有的行业知识,包括行业动态知识图谱构建、政务知识图谱信息提取、动态演化特征表示和政务监督信息利用等内容。

(4) 凤尾:“五大”主要环节

模型进一步定义了治理的五大主要过程。一是决策行为量化管理,主要以待决策目标为依据,对决策任务进行形式化管理;二是表示学习管理,主要针对现有复杂政务大数据进行有效表示学习,为后续任务提供基础;三是治理任务在线动态管理,主要针对具体的治理任务进行效率管理,提高治理任务的运行质效,包括解决如何面对治理任务的演化特性、多任务间调度策略优化、单任务内并行化策略学习等问题;四是治理检验分析,主要在以上步骤的基础上,对数据执行具体的治理措施,包括语法规则校验、语义重复性检验、语义异常检验等方面;五是治理质量评估管理,主要针对已治理完毕的数据进行质量评估,判断是否达到既定的质量目标,具体包括评估指标定义、评估规则库构建、执行评估过程和评估结果解释等方面。

综上,凤翔模型的“三一二五”治理过程具有以下3个方面的主要创新:

- 为政务大数据治理过程明确了目标导向性,创新地将政府决策与数据治理过程管理相融合,增加了政务大数据治理的业务价值;
- 为政务大数据治理过程明确了形式化

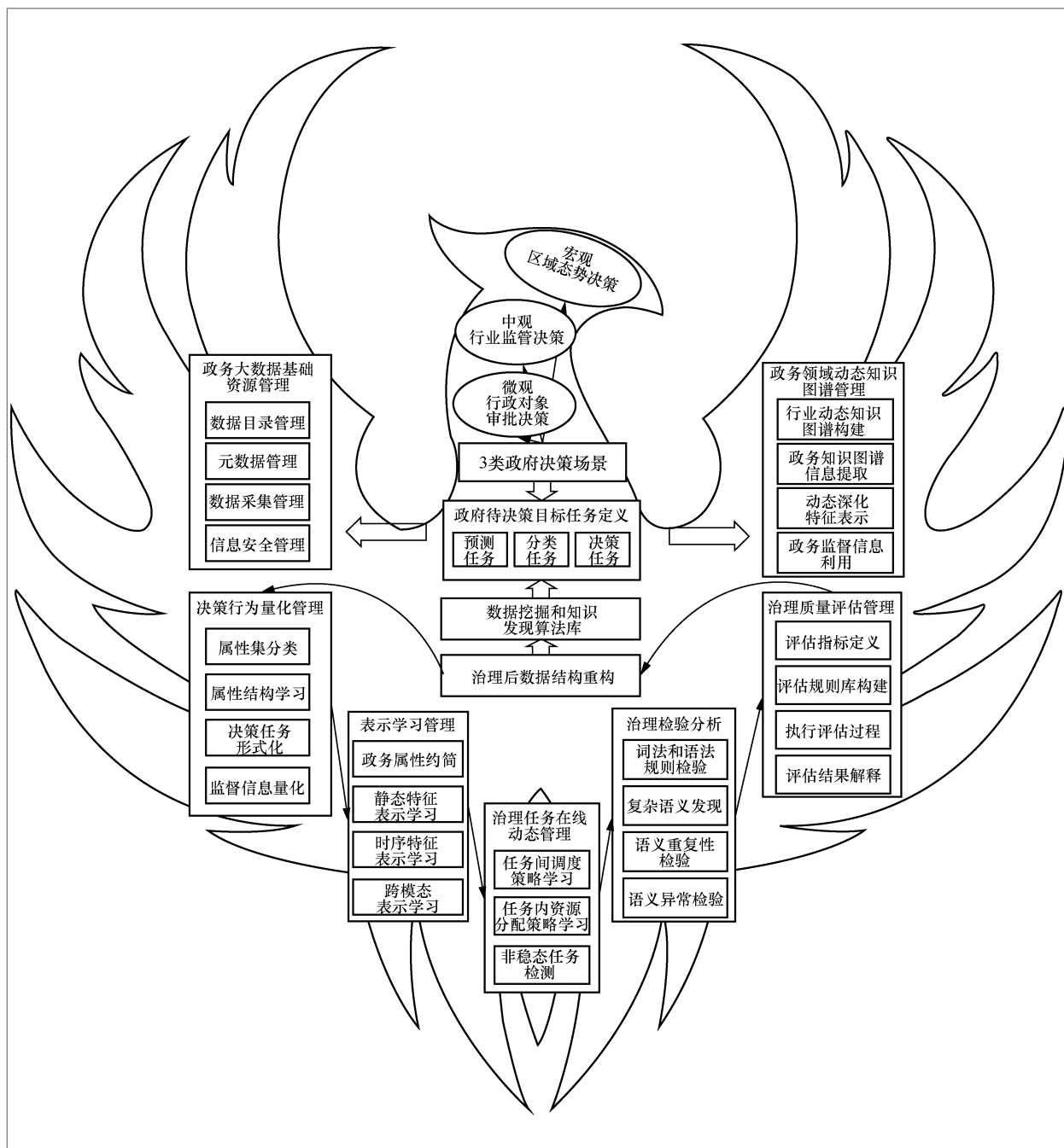


图2 凤翔模型：决策驱动型省级政务大数据治理模型架构

的任务可量化管理机制，从方法学角度提升了决策驱动型政务大治理过程的可行性；

- 为政务大数据治理过程引入了人工智能和大数据等前沿技术方法，进一步从技术路线上加强了政务大数据治理整体技

术水平和质量。

2.4 模型论证分析

从凤翔模型内在层级结构上分析，其

自上而下包含4个主要部分：凤首明确了3类决策场景，确定了数据治理的终极业务目标，使数据治理的业务导向性清晰明确；凤体依据数据治理业务目标进行形式化定义，将业务目标转化为可实施的技术语言，增加了模型的可操作性；凤翼明确了执行数据治理具体任务前的两项基础工作，为增加数据治理的质量提供助力；最后，凤尾明确了5项治理的具体过程，确定了数据治理的技术实现细节。通过以上4个层次，将政务大数据治理工作从高层次的业务目标逐步细化到具体的技术实施细节，保障了模型的可行性和有效性。

3 凤翔模型关键技术

引入人工智能和机器学习等前沿新技术是实现政务大数据高水平治理和创新发展的关键驱动力。本文以提出的凤翔模

型为例，对其中的关键业务和技术问题进行分析 and 梳理。

3.1 复杂政务大数据表示技术

就政务大数据本身而言，新时期政府数据资源的形态已发生了根本性变化：传统以离线、静态为主的政务数据转向在线动态、多模态演变的政务大数据形态；从数据产生机制来看，则由传统线下一来源转向线上线下融合产生，数据的复杂性明显增加。因此，面对动态、多变、异构的复杂政务大数据，对其进行有效表示是开展治理工作的基础。

其中主要包括3个方面的关键业务问题：一是对离线政务数据进行合理呈现；二是对具有时序特性的政务数据进行合理呈现；三是对文本、图像、视频等政务数据进行关联。**图3**给出了面向复杂政务大数据表示的3类主要问题以及

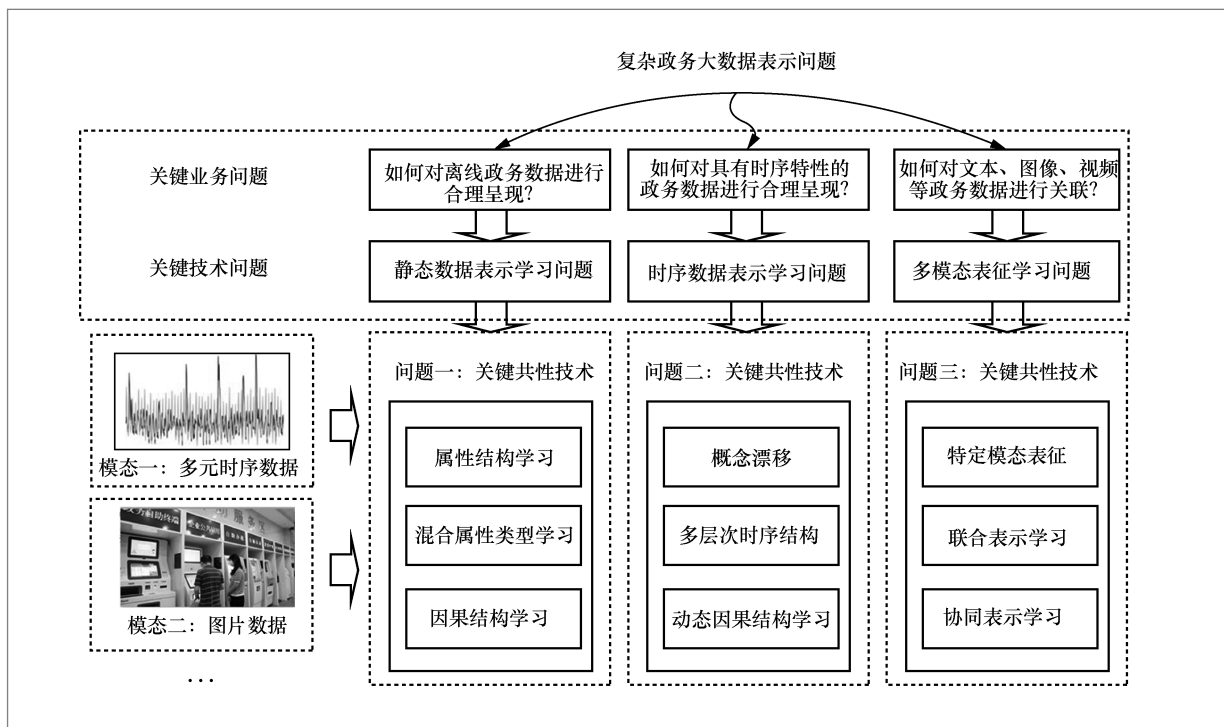


图3 复杂政务大数据表示共性关键技术

对应的共性关键技术。

具体而言,针对静态政务数据,可利用属性结构学习技术、混合属性类型学习技术和因果结构学习技术捕捉数据之间的潜在结构,实现对原始特征空间知识表示学习过程的强化;针对时序政务数据,可利用概念漂移技术捕捉数据中的分布变化,利用多层次时序结构捕捉数据中时序层次结构,利用因果结构学习技术推断数据中属性的因果结构,实现对数据中动态特性的刻画;针对多模态政务数据,应充分挖掘各模态之间的关系,并利用联合表示学习和协同表示学习技术在联合特征空间中学习跨模态表征,强化对不同模态数据之间关系的捕捉。

3.2 政务领域动态知识图谱构建技术

在确定待决策目标任务后,需充分利用行业内已有的领域先验知识和积累的

丰富实践经验,并将其融入决策分析的过程。相比传统静态知识图谱,由于政务相关领域的知识和规则在不断优化和完善,其具有较强的动态特性。因此,构建高质量的行业动态知识图谱是实现了对现有领域先验知识表示和推理的重要手段和载体。

其中主要包括3个方面的关键业务问题:一是如何提取各行政部门已有的领域知识;二是如何对提取的政务领域知识进行呈现;三是如何对已构建的领域知识进行组织。**图4**给出了行业动态知识图谱构建3类主要问题以及对应的共性关键技术。

具体而言,针对时序图谱信息提取问题,可采用动态实体抽取技术、动态关系抽取技术和实体关系对齐技术对复杂政务数据中的动态特性进行结构化;针对时序结构再表示问题,可利用时序知识图谱构建技术中的时序快照和复用机制,对时序实体和关系进行再表示;在以上基础上,可进一步研究制定政务时序图谱的学

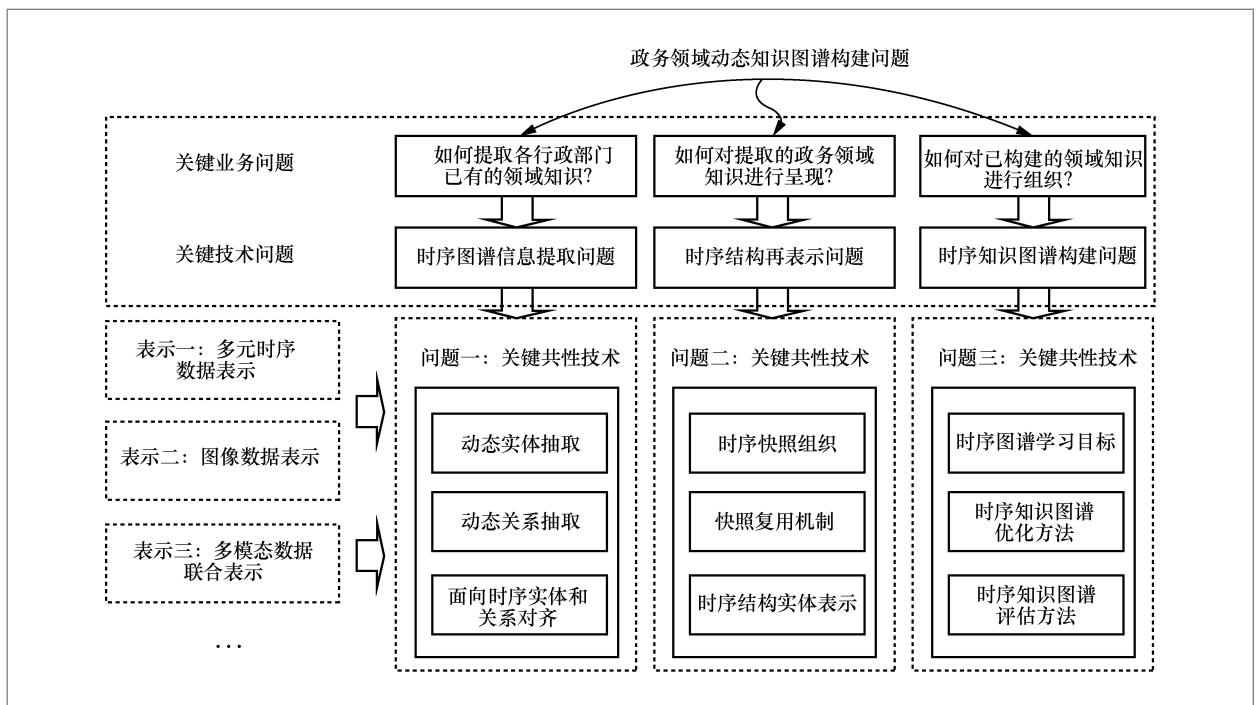


图4 政务领域动态知识图谱构建共性关键技术

习和优化目标,完成政务大数据时序知识图谱的构建。

3.3 治理任务在线动态管理技术

数据治理在线动态管理技术主要是指在新时期一体化政务系统建设背景下,面对规模并发量高、结构复杂的政务大数据,针对各项治理任务,设计更加合理、高效的方法进行资源调度和分配管理,以实现治理过程的质效提升。

其中主要包括3个方面的关键技术问题:一是如何检测治理任务中的动态特性;二是如何实现治理任务间的高效调度和编排;三是如何实现单任务内资源合理分配。图5给出了治理任务实时管理3类主要问题以及对应的共性关键技术。

具体而言,针对非稳态任务建模问题,可利用非稳态环境表示技术、非稳态环境

奖惩技术和非稳态环境目标优化技术实现对非稳态治理任务的刻画;基于非稳态建模,针对多任务集合资源动态调度问题,可利用多智能体规划和合作学习技术,实现对给定资源下任务序列的资源分配;针对单任务内的并行化策略学习问题,可利用元强化学习技术、非稳定性分布式约束优化技术和深度强化学习技术实现对任务内资源的高效利用。

3.4 复杂政务语义检验与分析技术

政务大数据具有多源、异构、演化等复杂特性。面对复杂政务大数据,一方面需要对其进行有效的语义检验,为决策提供高质量的数据基础;另一方面,从数据中发现的语义异常记录也是极具分析价值的信息,可为行政监管提供重要参考。

其中主要包括3个方面的关键问题:一

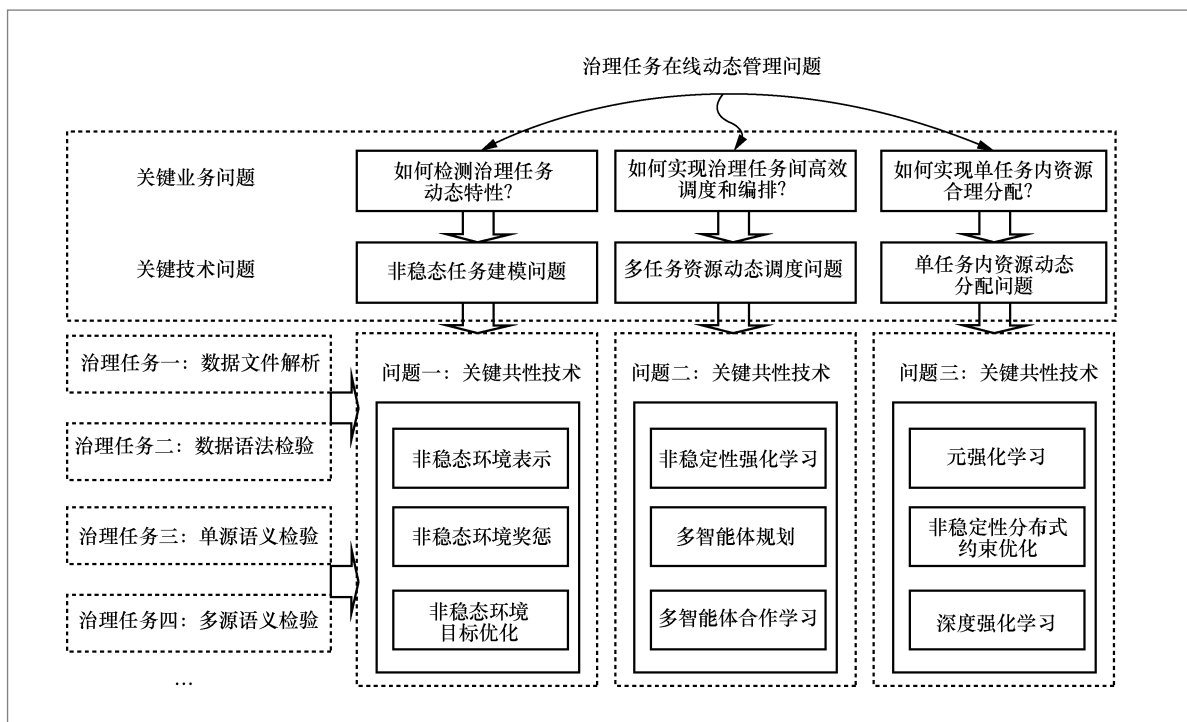


图5 治理任务在线动态管理共性关键技术

是如何发现政务大数据中的语义结构关系；二是如何度量政务数据之间的语义相似性；三是如何找到政务数据内部之间的业务逻辑矛盾。图6给出了复杂政务语义检验与分析3类主要问题以及对应的共性关键技术。

具体而言,针对政务大数据中复杂语义关系发现问题,可利用相关性发现技术、概率图模型技术和结构方程模型技术,实现对政务大数据中语义结构的识别和发现;针对度量政务数据之间的语义相似度问题,可利用二元结构度量技术、图结构度量和混合属性类型度量技术,实现对复杂政务数据中的语义相似度量;针对政务数据异常检验问题,可利用弱监督异常检验技术、无监督异常检验技术和异常可解释技术,实现对政务大数据中业务逻辑矛盾的主动发现。

4 总结与展望

政务数据在形态、产生机制等方面发生了巨大转变,其复杂性显著增加,对政务大数据治理提出了全新要求。本文以省级政务大数据治理为切入点,首先分析了传统省级政务大数据治理的技术架构和瓶颈;其次,围绕政府决策场景提出了一种决策驱动型省级政务大数据治理技术架构(凤翔模型),对其内涵概念、设计原则、总体架构进行分析,并对模型的可行性进行了论证;最后,针对凤翔模型技术架构中的4项关键业务问题和关键技术问题,进一步提出了针对性的技术实现思路,为各省市后续进行治理项目建设以及相关技术路线研究选型提供参考。

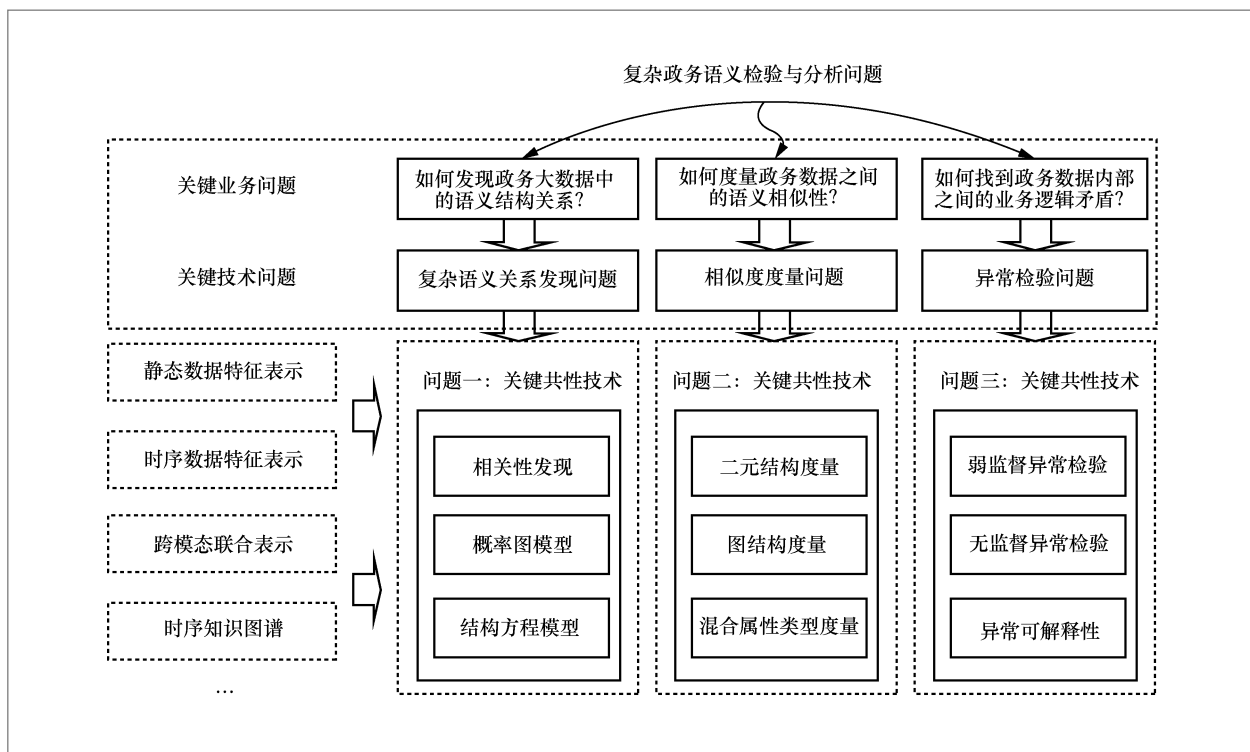


图6 复杂政务语义检验与分析共性关键技术

参考文献:

- [1] 梅宏. 数据治理之论[M]. 北京: 中国人民大学出版社, 2020.
MEI H. The theory of data governance[M]. Beijing: China Renmin University Press, 2020.
- [2] 梅宏. 大数据发展现状与未来趋势[J]. 交通运输研究, 2019, 5(5): 1.
MEI H. Current situation and future trend of big data development. Transportation Research, 2019, 5(5): 1.
- [3] 范灵俊, 洪学海, 黄晔, 等. 政府大数据治理的挑战及对策[J]. 大数据, 2016, 2(3): 27-38.
FAN L J, HONG X H, HUANG C, et al. Challenge and countermeasure of governing government big data[J]. Big Data Research, 2016, 2(3): 27-38.
- [4] 杨孟辉, 杜小勇. 政府大数据治理: 政府管理的新形态[J]. 大数据, 2020, 6(2): 3-18.
YANG M H, DU X Y. Big data governance in governments: a new form of the government administration[J]. Big Data Research, 2020, 6(2): 3-18.
- [5] 代红, 张群, 尹卓. 大数据治理标准体系研究[J]. 大数据, 2019, 5(3): 47-54.
DAI H, ZHANG Q, YIN Z. Study on big data governance standard system[J]. Big Data Research, 2019, 5(3): 47-54.
- [6] 印鉴, 朱怀杰, 余建兴, 等. 大数据治理的全景式框架[J]. 大数据, 2020, 6(2): 19-26.
YIN J, ZHU H J, YU J X, et al. A panoramic framework of big data governance[J]. Big Data Research, 2020, 6(2): 19-26.
- [7] 黄科满, 杜小勇. 数据治理价值链模型与数据基础制度分析[J]. 大数据, 2022, 8(4): 3-16.
HUANG K M, DU X Y. Value chain model of data governance and its application on data governance regulation analysis[J]. Big Data Research, 2022, 8(4): 3-16.
- [8] 信息技术大数据术语: GB/T35295-2017[S]. 北京: 全国信息技术标准化技术委员会, 2017.
Information technology-terminology: GB/T35295-2017[S]. Beijing: National Information Technology Standardization Technical Committee, 2017.
- [9] 杜小勇, 陈跃国, 范举, 等. 数据整理——大数据治理的关键技术[J]. 大数据, 2019, 5(3): 13-22.
DU X Y, CHEN Y G, FAN J, et al. Data wrangling: a key technique of data governance[J]. Big Data Research, 2019, 5(3): 13-22.
- [10] 安小米, 郭明军, 洪学海, 等. 政府大数据治理体系的框架及其实现的有效路径[J]. 大数据, 2019, 5(3): 3-12.
AN X M, GUO M J, HONG X H, et al. Framework of government big data governance system and effective way of implementation[J]. Big Data Research, 2019, 5(3): 3-12.
- [11] 李锋, 周舟. 数据治理与平台型政府建设——大数据驱动政府治理方式变革[J]. 南京大学学报(哲学·人文科学·社会科学), 2021, 58(4): 53-61.
LI F, ZHOU Z. Governing with data and strengthening government of platform kind: improving big-data-driven governance by government[J]. Journal of Nanjing University(Philosophy, Humanities and Social Sciences), 2021, 58(4): 53-61.
- [12] 胡海波. 理解整体性政府数据治理: 政府与社会的互动[J]. 情报杂志, 2021, 40(3): 153-161.
HU H B. Understanding holistic government data governance: interaction between government and society[J]. Journal of Information, 2021, 40(3): 153-161.

作者简介



孟凡(1988-),男,博士,江苏省战略与发展研究中心高级工程师,南京大学计算机科学与技术系推理与学习研究组成员,主要研究方向为大数据、人工智能、电子政务。



杨群力(1972-),男,江苏省战略与发展研究中心副主任、高级工程师,主要研究方向为电子政务、信用体系建设、社会经济发展。



高阳(1972-),男,博士,南京大学计算机科学与技术系副主任、教授、博士生导师、中国人工智能学会会士、理事,主要研究方向为大数据和人工智能。

收稿日期: 2023-02-22

通信作者: 高阳, gaoy@nju.edu.cn

基金项目: 江苏省社会信用体系建设专项(No. JSZC-G2018-393); 计算机软件新技术国家重点实验室(南京大学)开放课题(No. KFKT2022B27)

Foundation Items: Jiangsu Province Social Credit System Construction Special Fund Project (No. JSZC-G2018-393), The State Key Laboratory for Novel Software Technology at Nanjing University Project(No. KFKT2022B27)