

目 次

论文

平台企业的灵活积累机制

——一个合约理论的视角 滕 飞(1)

信息逆差:代码权力膨胀与公众权力让渡 施颖婕(21)

大语言模型能用来给主题模型中的主题编码吗?

——兼论大语言模型在社会科学领域应用的前景

..... 李 代 张博伦 周浥莽(33)

数字时代的科技公司与国家治理

——一个综合性的分析框架 钟瑞雪(60)

新闻大数据视角下测度区域一体化水平的新路径

——以长三角地区为例 李 军 董方杰(84)

解密数字社会中的幸福密码

——不同互联网使用程度下社会信任感和社会公平感对主观幸福感的

影响机制研究 郭媛媛 张 腾(103)

研究报告

- 数字时代安全科技价值报告 陆菲菲 秦 昊 张友红 吕 鹏(133)
- 问题触发的算法模型响应机制探索 许正军 袁 岳(153)

译文

数字健康的社会技术伦理

——生命伦理学进路的批判和延伸

..... 詹姆斯·肖 约瑟夫·多尼娅 著 何 丽 译(166)

书评

数字资本主义和智能技术的当代批判

——评《过度智能》 周银知(186)

访谈

通过人类行动思考机器智能

——专访人类学家露西·萨奇曼 露西·萨奇曼 胡万亨(198)

CONTENTS

THESIS

The Flexible Accumulation Mechanism of Platform Enterprises: A Contract Theory Perspective	Teng Fei(1)
“Information Deficit”: Code Power Inflation and Public Power Cession ...	Shi Yingjie(21)
Can LLM Label Topics from A Topic Model: The Future of LLM in Social Science	Li Dai, Zhang Bolun, Zhou Yimang(33)
Technology Companies and National Governance in the Digital Age: A Comprehensive Analytical Framework	Zhong Ruixue(60)
A New Approach to Measuring Regional Integration from the Perspective of Big Data in News: A Case Study of the Yangtze River Delta Region ...	Li Jun, Dong Fangjie(84)
Decrypting the Happiness Code in the Digital Society: A Study on the Influence Mechanism of Social Trust and Fairness Perception on Subjective Well-Being under Different Internet Usage Levels	Guo Yuanyuan, Zhang Teng(103)

RESEARCH REPORTS

Report on the Value of Security Technology in the Digital Age	Lu Feifei, Qin Hao, Zhang Youhong, Lyu Peng(133)
Exploration of Algorithmic Model Response Mechanisms for Problem Triggering	Xu Zhengjun, Yuan Yue(153)

TRANSLATED TEXT

The Sociotechnical Ethics of Digital Health: A Critique and Extension of Approaches from

Bioethics written by J. Shaw, J. Donia; trans. by He Li(166)

BOOK REVIEW

A Contemporary Critique of Digital Capitalism and Smart Technologies: Review of *Too*

Smart Zhou Yinzhi(186)

INTERVIEW

Thinking Machine Intelligence Through Human Actions: An Interview with Anthropologist

L. Suchman L. Suchman, Hu Wanheng(198)

问题触发的算法模型响应机制探索

许正军 袁 岳*

摘要:随着数据成为各行业发展的基础性要素和战略性资源,针对问题的分析研判正由经验判断向数据分析转变,由此产生了通过数据分析提出问题解决对策的数据赋能的主要范式。本文以问题触发的算法模型为响应机制,以扎根理论的研究方法为思路,以原发问题为质性研究材料,以算法库为智能中枢,以实现从问题浮现到问题识别、问题提炼和问题响应全过程的自动化和智能化,为公共服务和商业智能领域真正实现以问题为导向发挥数据的基础要素和战略资源作用提供了一种新的思路探索。

关键词:扎根理论 数据赋能 问题 算法模型

随着数字技术的飞速发展及其在各行业的应用与渗透,数据正在不断产生和汇聚,已成为各行业发展的基础性要素和战略性资源。以问题为导向,“用数据说话,用数据决策,用数据管理,用数据创新”已成为各行业数字化转型和高质量发展的主要抓手。针对问题的分析研判也正由经验判断向数据分析转变,数据赋能正在浸润各行各业。

由此,从数据到问题,即通过数据分析提出问题解决对策,便成为数据赋能的主要范式。在此范式下,人们往往在针对场景问题的数据采集、数据分析和分析、可视化监测、趋势判断和预警预测等工程技术方面倾注大量的投入,而对问题的本原,即对问题产生的根源、内容、逻辑、性质和类别以及如何捕捉、识别、提炼和响应等方面往往缺少足够的关注。也就是说,在数据赋能应用过程中,缺乏对问题浮现、问题识别、问题提炼和问题响应等业务环节足够的渗透与智能投射,导致以问题为导向的数据赋能在上述业务环节存在“梯度消失”。

以问题为导向的数据赋能应贯穿从问题浮现到问题解决的全过程,包括问

* 许正军,零点有数。袁岳,零点有数。

题浮现、问题识别、问题提炼和问题响应(包括响应问题、确立解决方案及后续解决执行)等各环节。亦即,在数据赋能应用过程中,应从上述全链条角度,以问题数据为材料,以算法模型为智能引擎,实现数据智能在各环节的渗透与投射,消除数据赋能各环节的“梯度消失”。其中,算法模型及其响应机制至为关键。

为此,本文以扎根理论(grounded theory)基于原发问题提炼阐发的社会学研究方法,提出问题触发的算法模型响应机制构建思路,旨在通过算法模型的构建、应用和迭代,建立从问题浮现到问题识别、问题提炼和问题响应的全过程数智化响应机制。

一、总体思路

众所周知,数据承载着信息,信息蕴藏着知识,知识可以赋能智慧应用以创造价值。随着数字技术的飞速发展及其在各行各业的深入应用,各类场景问题可通过数字化技术进行数字化记录,形成问题数据。对这些问题数据进行收集(形成问题库)、处理和分析,挖掘其中的信息和知识,为问题研判、问题响应和问题解决提供决策依据,构成了以问题为导向的数据赋能的基本逻辑。

但在实际过程中,上述大多只停留在数据采集、数据处理和分析、可视化监测、趋势判断和预警预测等工程技术方面(图 1),对于问题的本原,如问题产生的根源、内容、逻辑、性质和类别以及如何捕捉、识别、提炼和响应等方面缺乏足够的考虑,即在数据业务化层面做得不够充分。

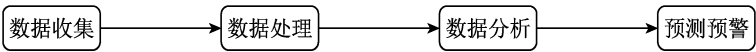


图 1 从数据采集到预测预警的示意图

问题触发的算法模型响应机制将立足于从原发问题的本原探究出发,从问题浮现、问题识别、问题提炼和问题响应等各环节着手,以问题数据为材料,以算法模型为智能引擎,建立从问题浮现到问题识别、问题提炼和问题响应的全过程数智化响应机制(图 2),实现“数据赋能”贯穿从问题浮现到问题解决的全

过程。

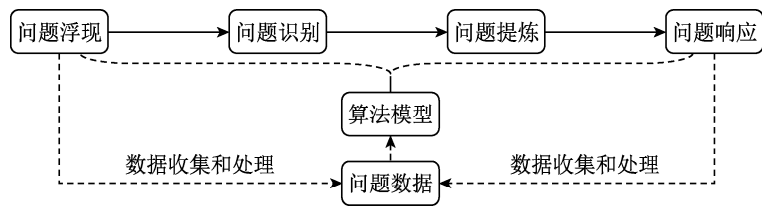


图 2 问题触发的算法模型响应机制总体思路

从图 2 可见,问题浮现是整个机制的起点,问题响应是整个机制的终点,二者都是问题数据的来源——前者是问题的初始来源,后者是问题解决成效的反馈,并且构成了一个问题从浮现到解决的闭环。通过对问题浮现和问题响应的数据收集和处理,形成问题数据库,就形成了问题触发的算法模型响应机制构建之源,这也是问题“业务数据化”的过程。问题识别和问题提炼是问题“业务数据化”的核心,是从问题浮现到问题响应的智慧转换器,而智慧之源来自算法模型。算法模型封装了多种基础算法和针对各类垂直应用、特定任务解决能力的智能引擎,这些算法和引擎通过对问题数据的分析和挖掘,高效、精准地识别问题的来源、特征、内容、类别和其他相关属性(如紧急事件、突发事件、热点事件或其他难点痛点堵点事件等),并根据问题识别的结果辨析出那些符合实际的可触及、可控制、可解决的问题,同时提炼出问题中的相关概念和命题,结合历史经验(对策库)和一些标杆做法,提炼出相应的问题解决理论体系和响应对策,从而实现“数据赋能”贯穿从问题浮现到问题解决的全过程。

二、问题浮现

问题浮现是整个机制形成的起点,也是初始问题数据收集的来源。无疑,了解问题浮现的起因与渠道至为关键。

不论是公共服务领域,还是商业领域,问题浮现的起因大多来自需求侧的各类诉求,包括投诉、表扬、建议、咨询、求助与评价等,也有来自第三方的调查或对供给侧的服务监督过程中的主动发现。问题浮现的渠道也呈现多源性特点。

以公共服务为例,问题的浮现有来自企业和群众的主动反馈,如通过 12345 政务服务便民热线,或通过政务服务好差评平台,或通过线下政务大厅专设窗口(如“不办事”窗口);也有来自大调研或第三方调查过程中的主动发现,如通过结构化或半结构化问卷的发放与问题采集;还有来自政府部门在重点事件关注或执法过程中的发现,或是如“双随机、一公开”执法监督过程中的主动发现;也包括对相关互联网办事平台与其他互联网媒体监测过程中的问题捕获;等等。问题的内容可能涉及方方面面,如政务服务、营商环境、市场监管、生态环保、公共管理、公共安全、公共服务等。问题的焦点可能涉及一些长期以来的难点、堵点和痛点,也可能涉及一些突发、并发或热点事件,抑或涉及政府部门某些重点关注的事件,等等。

不同于以往的仅从各业务系统收集数据并试图从数据中挖掘问题,深入了解与分析将有助于多渠道、多角度捕获问题的浮现,特别是对于潜在的隐性问题的发现,有助于拓展问题数据的收集渠道(如大调研、第三方调查、访谈、会议协商、检测记录等),提升数据收集的质量。同时,对问题浮现的深入了解将增强对问题浮现机理的全面掌握和深入了解,为后续的问题识别、问题提炼和问题响应创造良好条件。事实上,问题浮现的渠道、主题、内容和主体之间的矩阵图谱构成了具象化的问题浮现的机理特征(图 3)。

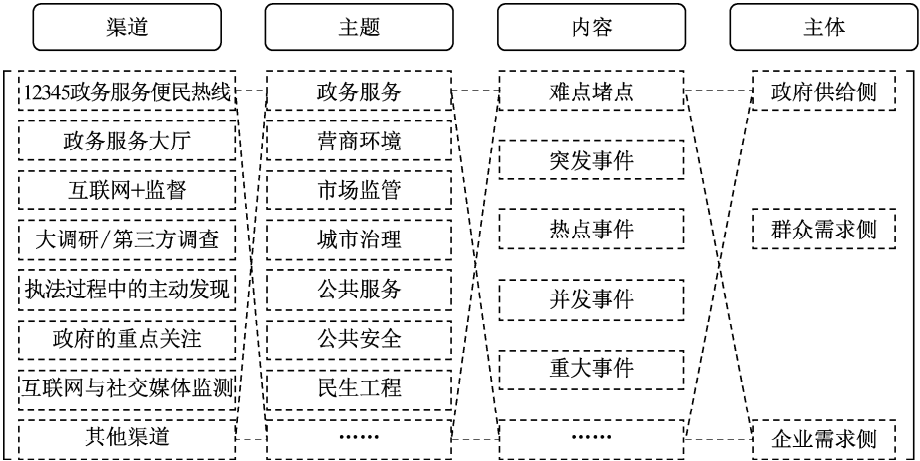


图 3 问题的渠道、主题、内容和主体之间的矩阵图谱

三、问题识别

问题识别在于挖掘问题的特征和识别问题的类型,这将采用扎根理论质性研究方法来实现。首先,对问题数据进行详细的分类和标签化处理,识别问题的基本概念和特征(简称“初始编码”);其次,将这些概念和特征进行关联,分析它们之间的逻辑联系、条件关系和因果机制,找出它们的主题范畴(简称“主轴编码”);最后,确定一个核心主题,围绕这个主题整合主题范畴和概念,构建问题的核心结构(简称“选择性编码”)。上述编码可挖掘问题的特征和识别问题的类型。

以公共服务领域中的营商环境问题为例。如前所述,营商环境问题来源众多:有来自政务数据资源库,也有来自企业访谈、营商小组讨论、相关文件资料等渠道,还有来自 12345 政务服务便民热线、政务好差评、投资服务中心、招商过程中的观察记录,以及其他政务服务业务系统等。基于对营商环境问题浮现机理的理解进行多源数据收集和处理,形成营商环境问题数据库,以此作为问题识别的质性研究材料。

首先,寻找并标记营商环境问题数据中供需双方关于营商环境提升和优化的初始概念与特征内容。例如,从企业需求侧可能会发现如“希望更加公平的市场环境”“司法程序复杂”“项目货款拖欠严重”“招工难”等众多与营商环境需求相关的诉求概念与特征内容,从政策供给侧可能会发现如“关于进一步加强公共资源交易监管的指导意见”“知识产权保护”“信用监管”“人才培养与人才引进政策”等众多与营商环境提升相关的诉求概念与特征内容。

其次,对这些初始概念或特征内容进行分类和关联,归并或提炼到每一个主题范畴中。例如,将“希望更加公平的市场环境”和“关于进一步加强公共资源交易监管的指导意见”归并到“市场环境”这个主题,将“司法程序复杂”和“知识产权保护”归并到“法治环境”这个主题,将“项目货款拖欠严重”和“信用监管”归并到“信用环境”这个主题,将“招工难”和“人才培养与人才引进政策”归并到“要素环境”这个主题,等等。

最后,确定一个核心主题,整合各个主题范畴和概念。例如,确定“营商环

境优化和提升行动”这个核心主题,将“市场环境”“法治环境”“信用环境”“要素环境”及其他主题、概念进行整合,形成针对营商环境优化和提升行动的基本理论框架(图 4)。

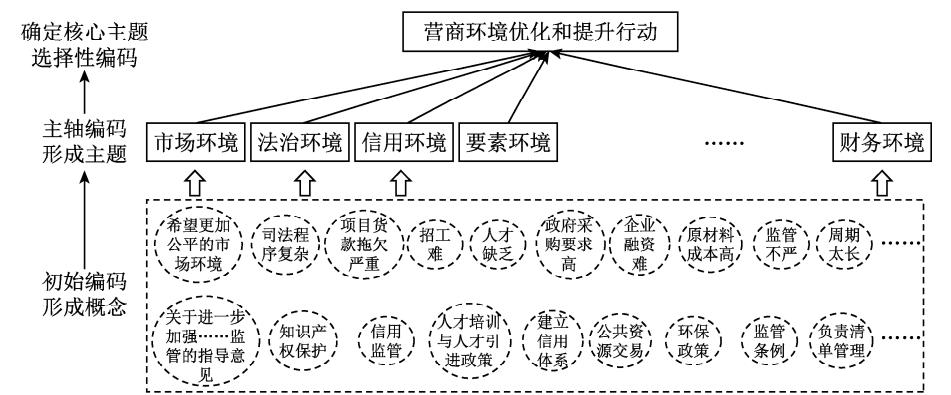


图 4 针对营商环境问题的编码与识别 (1)

实际过程中,可根据需要进一步将主题范畴细分为多级子主题范畴。例如,将“市场环境”进一步细分为“市场准入”“市场监管”等,将“法治环境”进一步细分为“依法监管”“法律诉讼”等,将“信用环境”进一步细分为“社会信用”“信用监管”等(图 5)。

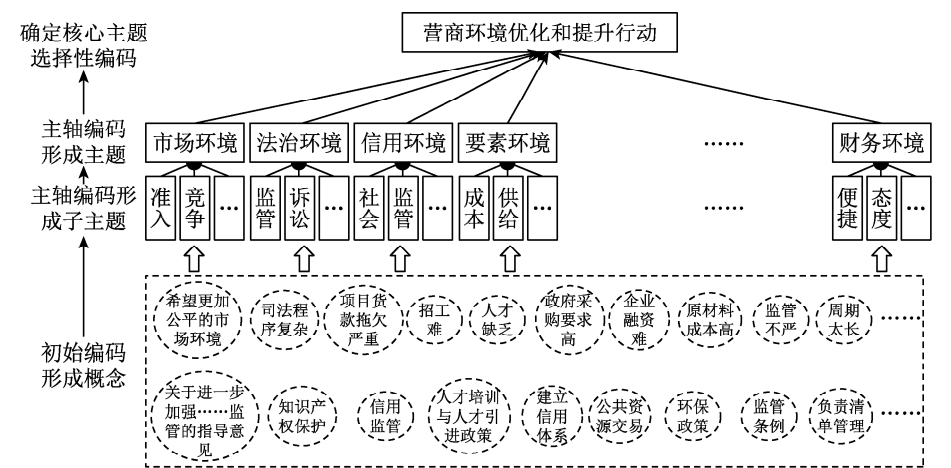


图 5 针对营商环境问题的编码与识别 (2)

考虑到与监测和评价指标的描述相一致,在问题识别过程中,可借助相关监测与评价的指标名来匹配编码过程中的相关概念和主题范畴。针对营商环境问题,可参照营商环境评价指标体系来进行。

四、问题提炼

在问题识别形成概念和主题基本理论框架的基础上,进一步对这些概念进行关联和匹配,实现问题提炼,为下一步的问题响应提供建议与理论指导。问题提炼主要涉及三个方面的核心内容:提炼概念中的异常情况(形成类似分类算法中的负例数据集);辨析异常情况的可触及性、可控性和可操作性,并结合其他特征变量分类排序形成问题序列(相当于对负例数据集进行等级分类和排序);进一步对问题进行深度挖掘,形成新的概念和命题,构建问题新的理论范式,并结合经验对策和一些标杆做法生成问题响应的对策与建议(相当于对负例数据进行分析挖掘后形成的对策与建议)。从技术角度而言,这三个核心内容构成了问题提炼器的三大核心功能(图 6)。

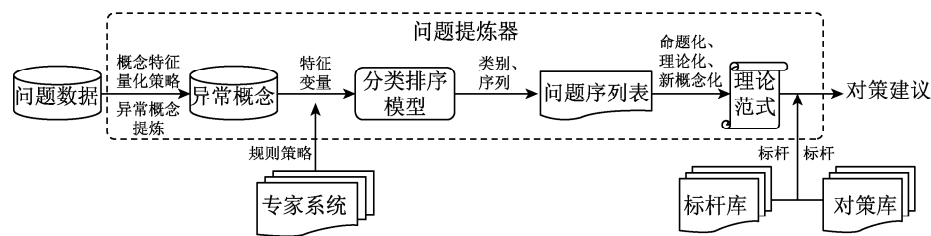


图 6 问题提炼器的三大核心功能

仍以营商环境问题为例。企业侧的需求和政府侧的供给是营商环境问题的两个方面,因而在问题识别的基础上,可对概念进行供需之间的关联和匹配,以发现供需之间哪些概念具有一致性,哪些概念存在部分差异,哪些概念存在显著冲突和矛盾(对于新出现的概念,技术上可视为此类),等等。例如,企业需求侧的“希望更加公平的市场环境”与政府供给侧的“关于进一步加强公共资源交易监管的指导意见”之间可能是一致的,“司法程序复杂”与“知识产权保护”之间可能存在部分差异,而“项目货款拖欠严重”与“信用监管”、“招工难”与“人才

培训与人才引进政策”之间可能存在显著冲突和矛盾(图7)。

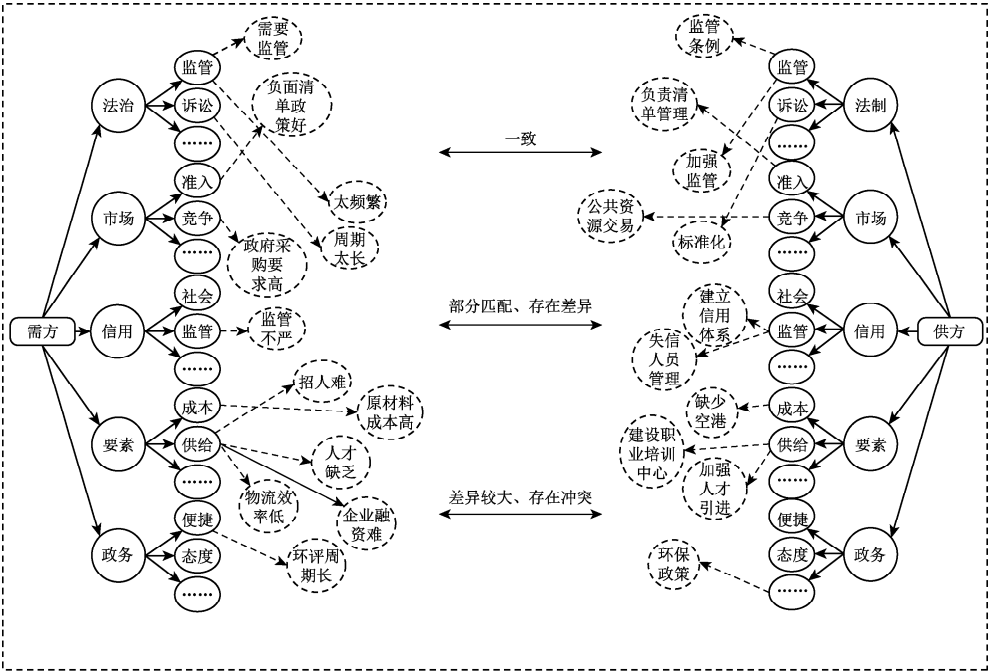


图7 营商环境问题中的异常概念提炼示意图

其中,一致性说明营商环境的提升行动是满足需求的,存在部分差异的说明营商环境仍需要进一步完善和优化,而存在冲突和矛盾的或新出现的概念,则需要重点关注。

鉴于并非所有的异常概念都具有合理性和可解性,因而问题提炼的第二个核心内容则是进一步辨析问题的可触及性、可控性和可操作性,并根据轻重缓急与其他相关特征变量(如差异程度、紧急程度、突发/并发/热点、社会影响、合理性、可解性、难易程度等),通过分类排序模型生成问题序列。这部分工作需要借助有关规则和专家系统来进行。

问题提炼的第三个核心内容将是进一步对问题进行深度挖掘,形成新的概念和命题,构建问题新的理论范式,并结合经验对策和一些标杆做法生成问题响应的对策与建议。具体而言,通过进一步对初始问题进行深度挖掘,从初始问题的表述中提炼概念、实体、关系、属性、事件、特征等关键要素及其之间的逻辑关

系、因果机制和知识图谱,构建一系列新的命题范畴。同时,通过重新挖掘初始问题的关键逻辑和规则,进一步提炼、综合和整合不同表述的同类概念,形成一系列新的概念表述;结合前述的异常分析和问题序列,构建新的核心主题。新的概念表述、新的命题和新的核心主题组合形成了初始问题的新的理论范式。在此基础上,再结合经验对策和一些标杆做法生成问题响应的对策与建议。

五、问题响应

在问题响应环节,借助问题提炼生成的对策与建议,有针对性地制定相应的解决方案,为问题解决提供行动指南。

在制定解决方案之前,首先需要根据问题提炼阶段提炼出的有关问题要素,如问题的内容、性质和主体等,决定是否要进一步与问题主体(问题当事人或其他相关人)进行沟通,以挖掘问题的真实原意;或对照比较是否属于领导重点关注的重大事件之列。同时,根据不同的问题内容、性质和主体,挖掘问题的核心目标是什么,需要投入哪些资源和条件,需要哪些法律法规支撑和保障措施,需要履行哪些决策程序,执行周期会有多长,执行后是否会达到问题的核心目标。将这些决策变量组合在一起进行推演和模拟,当预期效果能达到问题的核心目标时,便可形成解决方案和行动指南。对于达不到预期效果的问题,或是一时找不到好的对策的问题,可以借鉴其他地方的问题解决标杆来确定自己的解决方案。

总之,在问题响应环节,首先需要基于问题提炼的对策与建议,确定问题解决需要的要素条件,形成问题解决的时间路线图,组合不同的条件变量进行推演和模拟,形成有针对性的问题解决方案,这便构建了问题响应和问题解决的基本构架与主要逻辑。

作为闭环,对于问题响应和问题解决的成效需要进一步跟踪,形成问题响应和问题解决的跟踪库。对于成效不高的解决方案,需要进一步修正有关条件变量和时间路线图;对于达到问题核心目标的解决方案,便可形成具有参考价值的对策库,为日后类似问题的解决提供自动响应机制(图8)。

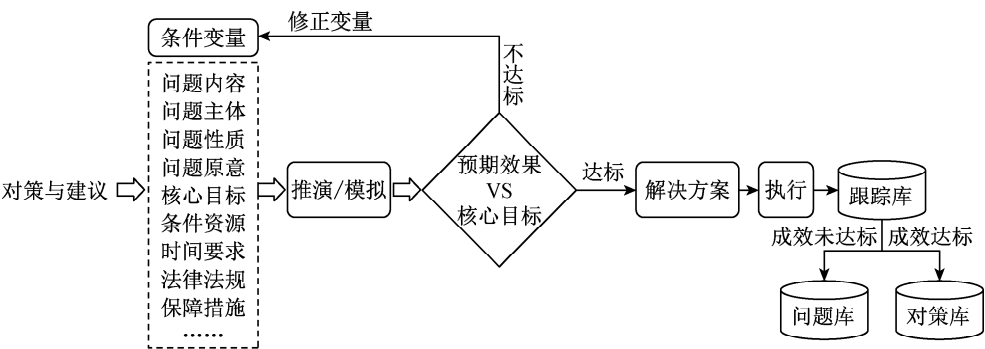


图 8 问题响应过程示意图

六、关键技术

问题触发的算法模型响应机制,仍然离不开数据收集、数据处理和数据分析等一些常用大数据技术的支持,包括数据治理、标准规范和数据安全等方面的技术应用。从前文可知,算法模型为智能引擎,是实现数据赋能贯穿从问题浮现至问题识别、问题提炼和问题响应全过程的关键。因此,本文着重讨论算法模型构建的关键技术。

通常,基于扎根理论的研究可以借助一些成熟的软件工具来实现,如 Nvivo、MAXQDA 和 QDA Miner 等。尽管这些软件在编码方面表现出较强的性能,然而在效率上仍属于“作坊式”定制,在准确度上仍需大量的人工干预。特别是在理论饱和环节,需要重复收集和分析数据,自动化和智能化程度不够。

近年来,机器学习尤其是深度学习技术得到了长足发展。特别是随着 ChatGPT 的出现,以 GPTs 为代表的生成式人工智能正在应用到各行各业中。本文讨论的算法模型关键技术,将全面采用机器学习算法来实现(图 9)。

在问题浮现环节,可以通过“超能交互”、智能语音识别技术及图像识别、图像分割和图生文大模型技术,实现对问题浮现的捕获与问题数据的生成。

在问题识别环节,可以组合运用分词技术(如 N-gram 模型、隐 Markov 模型、最大熵模型和神经网络算法)、TF-IDF 等文本特征选择与向量化模型技术

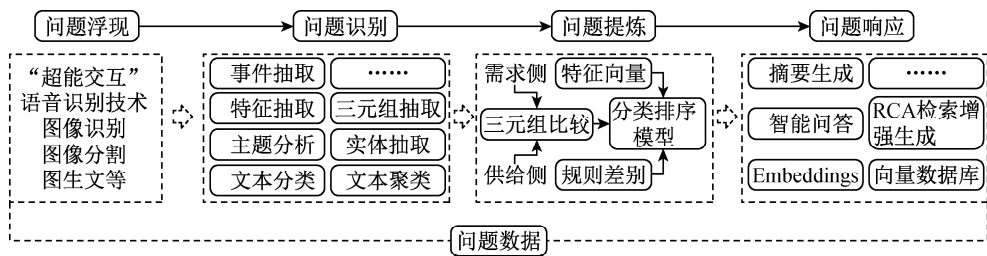


图 9 机器学习算法在问题触发的算法模型响应机制中的应用

选择问题数据中的特征和概念；运用词向量技术（如 Word2Vec）和潜在狄利克雷分布（LDA）主题分析技术，可提取问题数据中的关键词和若干主题，并可标示主题与关键词之间的权重情况，建立主题与关键词之间的关系，识别关键词之间的语义关系和相似度；通过知识智谱技术（如实体抽取、实体对齐、三元组抽取、事件抽取等技术），挖掘各种概念之间的关联。以上技术的运用，再加上文本分类、文本聚类等，为问题识别提供了丰富的技术手段。

在问题提炼环节，通过对实体对齐后的供需三元组属性值进行比较分析，以识别供需之间同一范畴内的概念是否一致，或是否存在差异，是否存在显著冲突和矛盾等情况。对于异常概念，可根据问题的可触及性、可控性和可操作性及其他相关特征变量（如差异程度、紧急程度、突发/并发/热点、社会影响、合理性、可解性、难易程度等），通过分类排序模型生成问题序列；可通过如主成分分析（PCA）或非负矩阵分解（NMF）技术进一步抽取问题数据的特征和主题，形成新的概念和命题，以形成问题的新理论范式。结合历史经验和标杆，利用类 ChatGPT 生成模型生成问题响应的建议和对策。整体上，上述技术的有机组合形成了一个问题提炼器。

在问题响应环节，可通过历史数据训练一些垂直领域大模型，建立问题解决方案中的条件变量、时间路线图与问题解决成效之间的对应关系，并以此大模型作为解决方案的扮演和模拟器，提升问题解决方案制定的科学性。其他如 Embeddings 模型、RGA 检索增强生成和向量数据库技术等，可为问题响应提供智能问答和文档摘要生成能力，助力问题响应过程中的自动化和智能化。

等。对这些渠道的历史数据进行收集,建立 12345 政务服务便民热线行业大模型,打造覆盖 12345 政务服务便民热线各类问题浮现监测、问题识别、问题提炼和问题响应能力的算法模型(库)(模型能力 APIs),并构建 12345 政务服务便民热线问题标准化体系(各类问题指标库)和对策库。

第二是问题提炼器的构建。在整个算法库的构建过程中,作为初始问题新的理论范式构建的重要引擎,问题提炼器在整个算法模型响应机制中起着至关重要的作用。问题提炼器的能力和性能决定了新概念和新命题提炼的质量,也决定了新的理论范式的先进性,从而也决定了对策与建议的科学性和准确性。

第三是针对具体细分场景问题的应用。应用所建立的问题标准化体系、对策库和算法模型(库),监测具体细分场景问题的浮现,通过问题识别、问题提炼(问题解析)和问题响应(推演和模拟),对应到具体的解决方案。问题解决方案的成效将被跟踪记录(跟踪库),作为闭环,不断优化迭代算法模型的应用能力。