

课题编号：2022YFC3801102

密级：公开

国家重点研发计划项目
“建筑与市政公用设施智慧运维理论与方法”

课题二

“建筑与市政公用设施智慧运维性态全息
快速感知与多源异构数据融合”

专题三
“面向多源异构运维数据的融合模型与管理
方法”

开题报告

承担单位：清华大学深圳国际研究生院

专题负责人：胡振中

二零二三年一月

1 前言

1.1 研究背景

（1）数据驱动技术

在建筑产业的生产过程中，基础设施的运营和维护是成本较高、周期较长、辐射范围较广的一个阶段。相较传统的基于人工巡检、纸质资料记载的碎片化、滞后化的运营管理模式，智能技术的引入提升了运维阶段的协同能力和对成本、风险的控制能力，推动建筑管理过程的精细化和全面化。近年来，城市基础设施智能运维技术面向协同性、集成性发展，融合数据驱动、建筑信息模型、数字孪生、扩展现实等热门技术，提供更节能、高效、可持续化的基础设施运营与管控方案。

数据驱动技术基于项目进程、成本、操作等动态数据与建筑三维信息模型的融合趋势，对动态数据的获取、分析、融合与管理技术进行创新，以实现建筑模型信息由静态向动态、兼顾大尺度和小尺度空间的多维度扩展。它是实现运维过程信息化和智能化的基石，也是数字孪生、系统集成与协作等技术的重要支撑，主要体现在：数字孪生技术基于现实世界的物理对象和虚拟世界的数字模型之间数据的自动流动，实现二者的双向交互。在实时数据驱动的基础上，数字孪生技术能够实现多智能体的协同，进一步完成时空多尺度下的模型搭建，从而实现对基础设施性态的自动化评估和预测。系统集成与协作需要融合数据监测和管理系统、模型分析和预测系统、平台化运维系统，对基础设施进行一体化智能管控。在平台化建设中，数据的集成和动态管理服务是系统智能网联架构中的重要组件。由此可见，基于数据驱动技术的运维数据感知与融合系统构成了智慧运维的底层数据基础，也是实现性态指标智能识别、平台智能决策的支撑性技术。

（2）基础设施运维数据

数据的集成是数据驱动技术的重要步骤。然而现阶段，受限于传统运维模式下数据来源渠道、资料整合途径的缺陷，智慧运维过程所需的建筑数据从采集、传输到映射、集成的过程中面临诸多难题，最具代表性的挑战来源于运维数据的

三个典型特征：储存分散性、多源异构性和同步困难性。

运维数据的储存分散性特征源于建筑运维行业中子系统分化复杂，专业性、安全性要求较高，且周期跨度长，种种因素导致运维过程涉及大量专业分包的独立作业，各个子系统独立建设，数据资源难以连通和共享。在这种长期积累形成的行业体系下，数据传输、储存的分散性难以避免，为数据的收集和聚合带来较大困难。目前，我国尚缺少建筑运维领域的数据标准化体系建设，长期分散的各类运维数据无法统一到同一框架下，不利于数据的集成。

运维数据的多源异构性特征源于运维过程各子类数据的采集方式、储存结构、传输协议的巨大差异。传统的运维过程积累了大量文字、图片、音视频形式的资料，以非结构化的数据为主导。伴随着行业的信息化进程，以建筑信息模型（BIM）、地理信息系统（GIS）为代表的建筑模型技术为智慧运维系统提供了一类重要的数据来源。这些信息模型往往基于不同的协议和标准框架，如 IFC 格式常用于建筑单体结构的精细表达，而 CityGML 格式则更适用于全局整体尺度下的区域信息管理。此外，随着智能感知体系的发展，实时监测数据的种类、格式和数量得到极大丰富。无线智能传感器、遥感定位、无人机、激光扫描等技术可以监测到基础设施自身结构、周边环境、运行环境的关键数据，扩展智能感知的时空、对象范围。由此可见，基础设施智慧运维过程涉及模型数据、监测数据、非结构化数据，包含众多数据类型和数据储存、表达格式，造成了运维数据多源异构的特点，进而为数据的集成过程带来较大阻碍。

运维数据的同步困难性特征体现在不同数据源之间、数据源和数字孪生体之间信息关联的难度较大。运维项目各专业领域独立开展作业，专业间信息共享效率较低、体系性较差，数据无法统一；且由于运维阶段众多子系统并未主动关联建筑模型，各类监测、管理、调度数据之间从根源上存在信息壁垒，难以实现同步关联和协同利用，更无法支撑数字孪生体的搭建。数据采集过程协同性差，信息传输通道一致性差、效率低下，导致支撑智能运维系统的数据集成服务很难实现。

（3）运维数据融合模型

针对智慧运维不同数据源在结构、格式、端口、协议等方面的差异，搭建统

一的运维数据融合模型是实现数据集成的必经之路。数据融合模型是预期整合现有的多种运维模型和数据集的语义信息、数据格式、操作模态等内容的关联信息模型。在建立数据标准化体系的基础上，通过特征提取、特征匹配、时空映射等方式将异构数据关联至统一框架下，使他们能够互相调用、交互作用，即可得到异构数据的融合模型。

国内已有一些大型企业推出自主研发的数据融合平台，如腾讯云、阿里云等。腾讯云使用了图扑软件 HT 对 BIM 模型进行轻量化处理和外观优化处理，能够支持 BIM 模型数据与 3D 场景或地图场景的叠加，实现 BIM+GIS 的可视化效果。用户可以从 BIM 客户端上传 IFC 文件，服务器端将在处理后按照 BIM 属性结构将模型展示在客户端；用户编辑后可将选中的 BIM 组件生成 BIM 模型快照或 OBJ 文件（图 1-1），支持 GIS 系统的直接读取。阿里云使用 TSL 语言获取更精确地物理对象描述；对于数据异构问题，阿里云是在数据源接入前独立开发代码进行格式操作，而并非在数据融合过程中进行处理。这些数据融合平台普遍存在数据类型较少，融合过程的同步性和连贯性较差，数据服务接口与数字孪生多智能体之间的信息连通性不好等问题。因此，运维数据融合模型的下一步研发重点在于扩展运维数据种类，开发数据动态匹配的融合机制，搭建可用于支撑数字孪生和智能决策的运维数据服务。

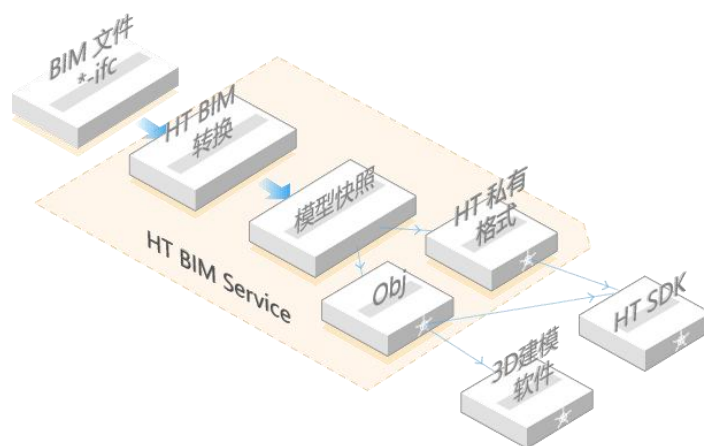


图 1-1 腾讯云“基于 BIM 技术的可视化管理系统”功能流程图

1.2 目标和意义

为解决建筑运维数据储存分散性、多源异构性和同步困难性的问题，实现数据的集成与协同应用，本课题将关注基础设施运维数据的融合模型与管理方法，研究多源异构数据的关联、融合与共享的理论方法，实现支撑智慧运维系统的数据服务平台建设。预期完成运维过程多类监测数据与信息模型的动态匹配，实现时空多尺度下的数据关联融合，构建“云-边-端”协同架构下多源异构运维数据动态融合和管理模型，搭建智慧运维数据服务。

从技术层面，本课题的研发将在多源异构运维数据融合领域实现重大创新突破，为构建多层级关联融合数据库提供理论支撑和标准化数据体系，为实现多源异构运维大数据的集成建立统一编码与关联融合机制，为搭建运维数据动态融合和管理服务提供“云-边-端”协同架构下的解决方案。从项目层面，本课题的实施将有助于构建上下级联合、横向贯通、逻辑一体化的数据资源支撑服务系统，支撑课题三的性能评价与演化预测预警，支撑课题四决策理论制定，是智能运维数据感知和融合系统的重要组成部分。

2 国内外研究综述

2.1 文献检索概述

为了充分了解多源异构运维数据融合的相关研究现状，研究团队对国内外相关文献进行了详细的检索与调研。

首先基于对课题和相关理论的初步了解，提取相关中英文关键词，主要包括：多源异构、数据融合、运维、“Multi-source Heterogeneous”、“Multi-source isomerism”、“Data fusion”、“datafusion”、“information fusion”、“O&M”、“OPS”、“Operations”、“Operation and maintenance”、“DevOps”等。选取的中文数据库主要包括中国知网、万方数据资源系统、维普资讯等；外文数据库主要包括 Web of Science 的核心合集、Engineering Village 的 Compendex，Scopes, Science Direct 等。根据提取的关键词建立检索式，在相关的中外文数据

库中进行检索。对所得到的文献进行统计分析，选择相关度较高的文献摘要进行初步阅读，并根据摘要内容将文献分为主要文献与次要文献。主要文献进行精读，次要文献进行粗读。对参考价值较高的文献进行引文搜索，全面了解其研究现状。

本部分主要展示研究团队在 CNKI、Engineering Village 的 Compendex 和 Web of Science 核心合集中的检索结果。相关调研内容汇总如表 2-1 所示。从检索结果可以看出，目前多源异构运维数据融合的相关文献数量比较少，该领域尚未研究充分，可见本研究具有较大的研究价值与必要性。

表 2-1 文献调研汇总表

数据库	检索跨度	检索式	结果	备注
中国学术期刊 网络出版总库	2013- 2022 年	SU='数据融合'	2.15 万	数据 融合 相关 调研
Engieering Village 的 Compendex	2013- 2022 年	(Data fusion or datafusion or information fusion) WN ALL	105340	
Web of Science 核心合集	2013- 2022 年	ALL=(Data fusion or datafusion or information fusion)	116168	
中国学术期刊 网络出版总库	2013- 2022 年	SU='数据融合'*'多源异构'	333	多源 异构 数据 融合 调研
Engieering Village 的 Compendex	2013- 2022 年	(Data fusion or datafusion or information fusion)and (Multi-source Heterogeneous or Multi-source isomerism)WN ALL	494	
Web of Science 核心合集	2013- 2022 年	ALL=(Data fusion or datafusion or information fusion)) AND ALL=(Multi-source Heterogeneous or Multi-source isomerism)	282	

续表 2-1

中国学术期刊 网络出版总库	2013- 2022 年	SU='数据融合'*'多源异构'*'运维'	6	多源 异构 运维 数据 融合 调研
Engieering Village 的 Compendex	2013- 2022 年	(Data fusion or datafusion or information fusion)and (Multi-source Heterogeneous or Multi-source isomerism)and(O&M or OPS or Operations or （Operation and maintenance） or DevOps)WN ALL	87	
Web of Science 核心合集	2013- 2022 年	ALL=(Data fusion or datafusion or information fusion)AND ALL=(Multi-source Heterogeneous or Multi-source isomerism)AND ALL=(O&M or OPS or Operations or （Operation and maintenance） or DevOps)	29	

为调查该研究领域的发展趋势，研究团队将近十年中每年发表的相关文献数量进行时间图绘制展示，如图 2-1、图 2-2、图 2-3 所示。可以看出，数据融合相关文献数量总体呈现上升趋势，因而可以体现出该研究领域具备较好的发展前景并日益受到学界 的关注。

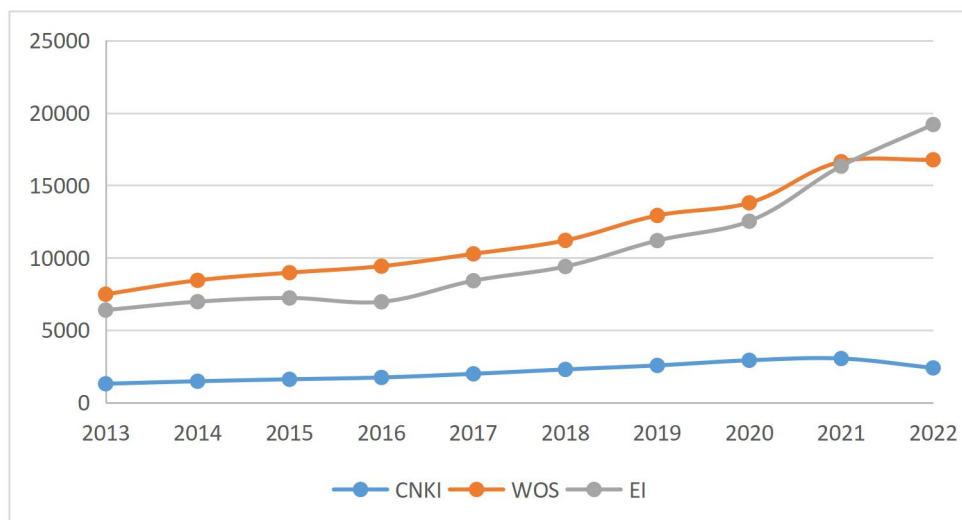


图 2-1 数据融合调研

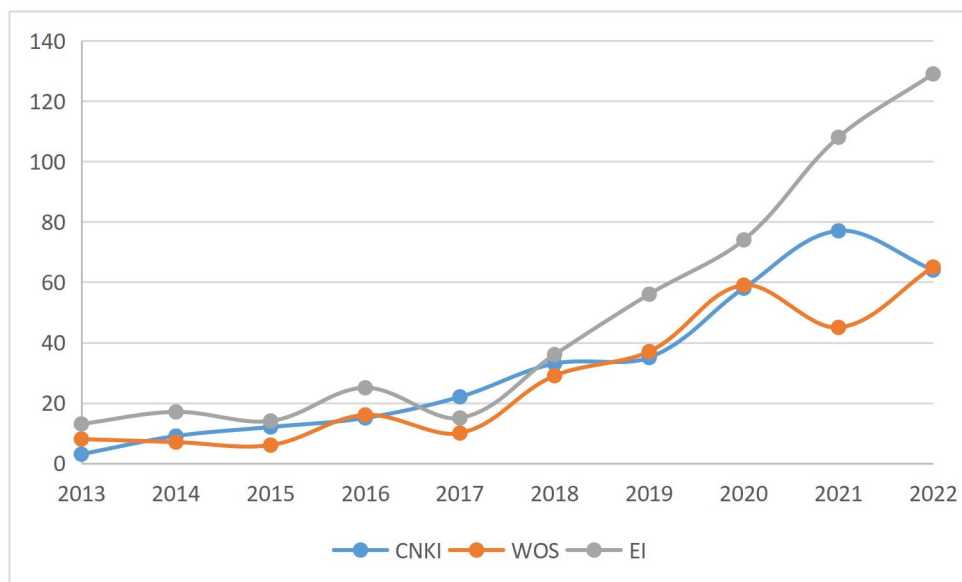


图 2-2 多源异构数据融合调研

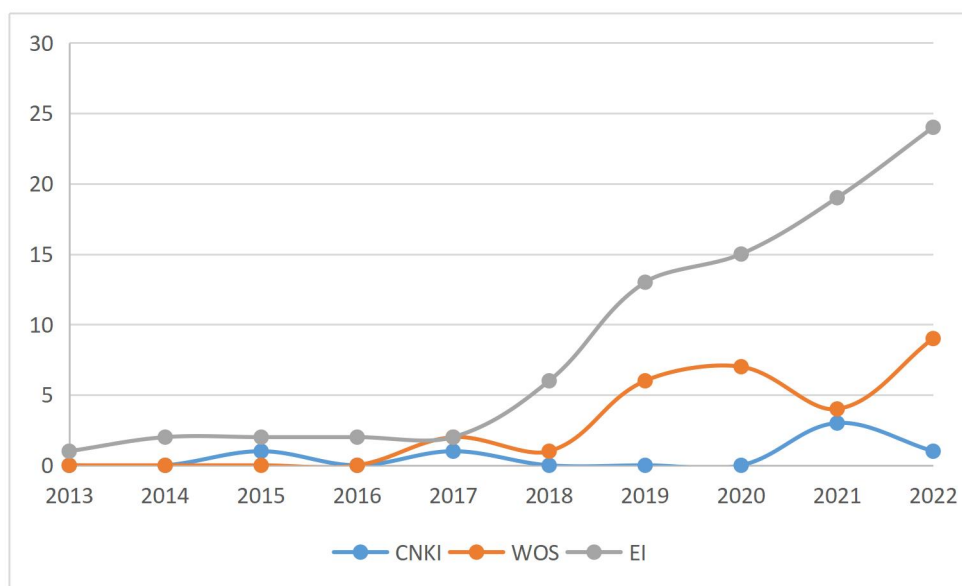


图 2-3 多源异构运维数据融合调研

2.2 相关研究综述

数据融合技术开始是起源于国外，20 世纪 70 年代，美国就对多传感器的数据融合进行了相关研究。该技术首先是应用在军事领域，主要是对从多传感器中获取的多源数据进行处理并据此做出自动决策。后来逐渐扩展到民生领域，包括工业控制、车辆识别、智慧医疗诊断、交通控制等。

在学术研究方面，20 世纪 80 年代，美国组建了数据融合的专家组，专门针对数据融合技术进行讨论，包括组织相关的学术会议，目前比较著名的学术会议包括 SPIE 传感器融合年会^[1]、美国三军信息融合年会以及国际信息融合学会 (ISIF)^[2]等，这些会议不断地进行着数据融合方面的技术探讨，并产出了许多技术进步，形成了许多该领域内的专业文献与数据融合的统一参考框架。其中 Waltz、Llinas^[3]等人出版的《多传感器数据融合》和 Hall^[4]等人出版的《多传感器数据融合数学基础》等出版物都产生了重大影响。

在数据融合模型的建立方面，1986 年美国国防部实验室理事联合会 JDL (Joint Directors of Laboratories) 首先建立了关于数据融合的专业术语检索词典和基本系统模型^[5]，并在 1988 年做出进一步完善。JDL 是目前数据融合领域最被广泛使用和改进的模型。后来 Bowman^[6]等人提出了数据融合树的概念来对此做出拓展，

将融合问题划分为结点，每个概念结点都包括数据关联、相关、评估等功能。以此为基础，Bosse E^[7]提出了一种建模和仿真的方法来实现一个数据融合系统的实际应用设计。

我国多源异构数据融合的研究起步较晚。20 世纪 80 年代末期，我国才有相关的研究论文产出。而随着各方对于数据融合技术的重视，90 年代国内涌现出来了大量关于多源异构数据融合的专著与研究论文，比如我国首批关于多传感器数据融合理论的著作《数据融合理论与应用》^[8]，为后人的研究提供了充足的理论支持。进入二十一世纪以后，国内先后召开了两次信息融合研讨会，在促进数据融合发展方面起到了一定的作用，一些著作包括《多传感器融合及其应用》^[9]《多源信息融合》^[10]相继出版。国内相关学者也积极研究数据融合技术应用在不同领域。例如：段建斌^[11]运用模糊神经网络对多源数据融合的技术，得到实时监测瓦斯变化情况的效果；北京大学的化柏林、李广建^[12]等在大数据环境下针对竞争情报进行数据融合研究，并重构了多源异构数据融合的处理流程以及算法体系。

另外在算法研究方面，国内学者也提出了一些数据融合的先进算法，何友^[13]在文献中提出一种有无反馈式航迹融合方程，调节局部节点的估计精度；文成林^[14]在文献中提出一种基于数据融合的分布式预测算法。随着人工智能技术的发展和计算机硬件性能的提高，国内学者开始将神经网络和模糊理论应用到数据融合技术当中，比如荣健^[15]提出的将自适应模糊神经网络和卡尔曼滤波器应用到目标跟踪系统中的融合算法，有效地通过自适应调整跟踪参数，防止目标丢失；张雨浓^[16]提出一种结合样条逼近理论和神经网络原理的融合算法，通过权值修正迭代尽快确定权值，达到更快的运算速度和计算精度，王海颖^[17]通过形成联系度矩阵并扩维等方法，提出了一种新的基于集对分析联系度的多源数据融合方法，这些著作和算法的提出都比较具有突破性的进步。

2.3 综述小结

国内外研究综述表明，近几十年来，多源异构数据融合技术已经得到了学界的广泛关注与研究，不断有新的方法被提出以及经典理论的创新应用等，包括贝叶斯推理、D-S 证据理论、模糊集理论、神经网络等，但就运维领域而言，相关

研究文献还比较少，研究也不充分。通过分析考虑多源异构运维数据的一般性与独特性，综合构建多源异构运维数据模型，对于提升数据运维服务水平具有显著意义。

3 研究内容与创新点

3.1 研究内容

(1) 建立多源异构运维数据融合模型

综合考虑异构数据源的差异性、完整性、时效性等特性，全面深入分析运维数据的组成、特征、内涵、形式和应用模式，研究多源异构运维数据的语义、逻辑和框架体系，基于协同国际和国内数据标准的原则和统一方法，定义标准化数据描述与表达方法，建立融合数据大纲，从而建立多源异构运维数据融合模型。

(2) 提出基于数据动态匹配的融合机制与算法

针对运维过程语义约束的图形图像、点云、BIM 模型等多维度静态数据，以及航拍视频、定点监控数据和高维信号等大容量动态数据，研究以上时变异构数据在指征统一对象时的特征表达、转换对齐、视觉匹配方法，综合考虑多端采集、边缘计算及动态调度的耦合效应，从时空关联特性、对象识别和模式匹配算法等方面，提出多源异构数据的融合机制与方法，在“云-边-端”协同框架下实现多源异构运维数据的动态融合。

(3) 搭建基于数据融合的智能运维数据服务

面向多源海量模型数据和多端应用需求，研究基于云计算平台的数据动态存储与管理技术、图形图像轻量化技术，研发基于数据中台和微服务架构的数据接口，实现集成高效融合算法、目标识别算法的后台计算服务，作为“云边端”协同架构下的“大脑”，支撑智能的运维数据挖掘，提高运维数据的应用价值。

3.2 拟解决的关键科学和技术问题

(1) 重大科学问题：建筑与市政公用设施多源异构运维大数据的统一编码与关联融合机制

建筑与市政公用设施智慧运维数据具有多源异构的特征，其规模也成指数型

增长趋势，造成了信息壁垒和信息孤岛；数据之间无法关联融合，造成数据冗余歧义；传统的数据集中管理方法效率低下，制约了数据驱动工程管理的应用和发展。因此，本课题拟研究数据统一编码的理论和方法，考虑数据传输效率、传输安全、检索精准度等要素，提出多源异构运维大数据的分布式边缘计算与管理體系架构。在此基础上，从数据采集、传输、处理和应用层面，深入解析工程数据内在关联机制，定义多源异构数据的高效融合运算法则，以及阐述数据全面互联互通的内在机理，支撑多源异构数据的关联与融合，是本课题一个关键科学问题。

(2) 关键技术问题：基于“云-边-端”协同框架的多源异构运维数据动态匹配和融合技术

基于多源异构数据统一编码方法，本课题拟解决分布式边缘框架下实现数据动态匹配的方法，构建时空多尺度下的运维信息融合模型。考虑不同数据源在结构、语义、逻辑、框架体系方面的差异，研究并实现时变异构数据在指向统一目标和过程时的特征表达、转换对齐、视觉匹配方法，实现数据融合过程中的信息关联建模，提升数据融合和表达过程的一致性和连贯性；考虑“云-边-端”架构多端点的数据采集、边缘计算、动态调度耦合效应，从时空关联的角度设计动态匹配算法，构建动态融合的运维信息模型，是本课题的一个关键技术问题。

3.3 创新点

根据国内外研究现状，现阶段对建筑与市政公用设施运维产生大量数据进行感知、融合、存储具诸多挑战：针对运维数据的多来源、异构型问题，多源异构运维数据的融合模型尚无深入研究，缺乏理论支撑。面向多源异构运维数据的融合模型与管理方法，本课题将从理论研究到技术研发作出如下创新：

(1) 根据多源异构数据的结构、语义、逻辑、框架体系，定义基础设施运维标准化数据体系，构建运维大数据统一编码机制和分布式边缘计算与管理體系架构，定义包含 5 种以上多源异构运维数据的融合模型的统一表达和描述方式，作为本研究的理论支撑和运维数据融合模型领域的理论突破。

(2) 针对运维过程语义约束的多维度、大容量的静态和动态数据，提出基于数据动态匹配的融合机制与方法，在“云-边-端”协同框架下实现数据的关联

建模。预期完成 5 种以上运维数据的动态融合，达到行业领先水平。

(3) 搭建基于数据融合的智能运维数据服务，基于云计算平台的数据动态存储与管理技术、图形图像轻量化技术，实现集成高效融合算法、目标识别算法等后台计算服务，支撑智能的运维数据挖掘，提高运维数据的应用价值。该技术预期获得多项专利、软件著作权、学术论文等形式的知识产权。

4 研究方法与技术路线

4.1 研究方法

4.1.1 基于数据标准化的运维数据模型构建方法

依据本体建模理论，参考建筑信息模型及物联网领域常用的 IFC、CityGML、SSN、IoT-Lite 等数据描述与表达体系，将多源异构运维数据的表征方法统一至同一框架体系下，扩展标准化体系的语义与逻辑系统，构建基于本体框架的数据融合模型。

(1) 基于本体建模的数据模型表达方法

对建筑运维阶段数据的组织形式、组成成分、结构特征、语义内涵、应用模式等进行总结，而后采用本体建模的方法，基于<类，属性，实例>的本体逻辑结构，构建数据描述的标准化体系。该体系用于描述数据融合模型的上层框架，采用 OWL (Ontology Web Language) 建模语言，包含个体、属性、类三个组分。个体是类的实例，属性描述两个个体之间的逻辑和关系，类则是个体实例的集合。通过子类和父类之间关系的构建，来构造体系中的层次结构。该模型基于 RDF (Resource Description Framework) 资源描述框架，以 XML 或 N-Triples、Turtle 等序列化方法进行传输和储存。其中包括 `rdf:Class`（用于定义类）、`rdf:domain`（用于描述属性类别）、`rdf:subClassOf`（用于描述类的父类）、`rdf:subProperty`（用于描述属性的父属性）等 RDF 词汇，以及 `owl:equivalentClass`（定义相同的类）、`owl:sameAs`（定义同一实体）等本体映射词汇。

针对基础设施智慧运维的场景，数据融合模型的辐射范围包括 BIM、GIS、传感、点云、遥感等至少 5 种多源异构数据类型，以及超高层建筑、道路、桥梁、

隧道、交通枢纽等应用场景类型，如图 4-1 所示。预期的本体构建过程是在参考建筑行业数据标准中的既有本体体系的基础上，针对多源异构数据的内涵和特征进行本体体系的手动构建，还可结合对行业标准、规范文本信息的挖掘，进行本体体系的局部自动扩展。

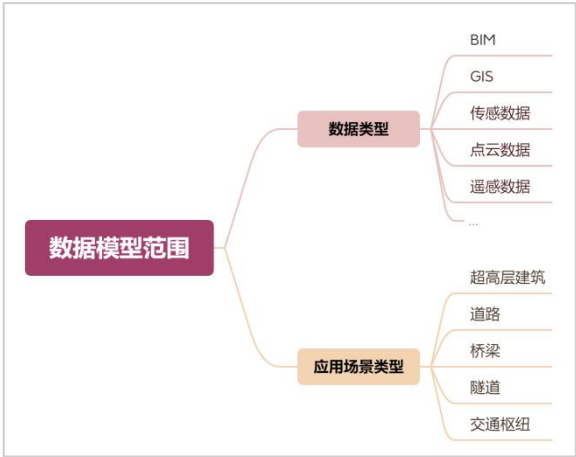


图 4-1 数据模型的范围描述

(2) 基于实体-关系框架的数据模型语义构建方法

基于运维数据的本体体系，融合建筑行业、物联网产业既有数据标准框架，构建数据融合模型的语义框架。计划参考的数据标准框架包括 IFC、CityGML、SSN 等，其通用的语义逻辑是基于实体-关系的描述方式，以下进行展开叙述。

IFC (Industry Foundation Classes) 是建筑行业国际通用的数据标准体系，也是 BIM 数据的描述框架，基本结构分为资源层、核心层、交互层和领域层。它的基础表达方式是基于 EXPRESS 的实体关系模型，由实体-关系连接而成，如图 4-2 所示。该体系采用继承型层次结构描述实体和关系对象，信息类别覆盖构件信息(IfcElement)、几何信息(IfcRepresentation)、材质信息(IfcMaterial)、空间关联信息(IfcRelConnects Element)等。图 4-3 展示了该体系继承型实体定义的一个示例，即从根节点逐层继承，直至定义建筑构件中梁的描述类(IfcBeamType)。

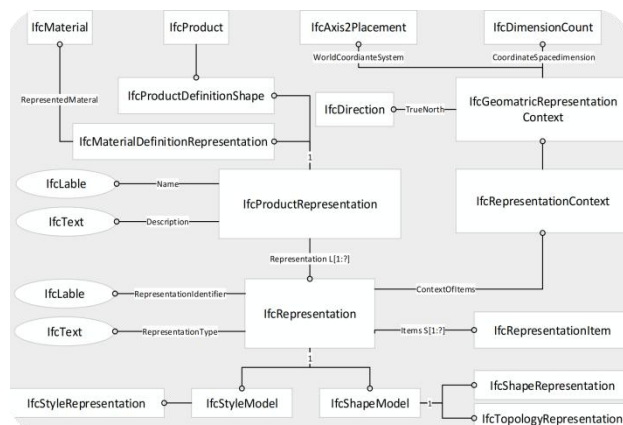


图 4-2 EXPRESS 模型示意图

Entity inheritance

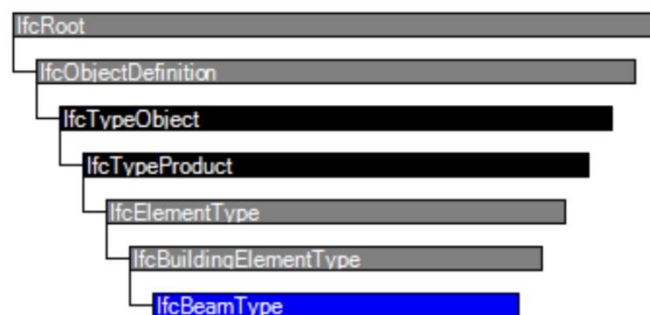


图 4-3 基于继承性结构的实体定义方式示例图

CityGML 是 OGC 城市地理标记语言的编码标准,采用基于 XML 的类和关系模型进行表达,其基本组成是基于继承型层次结构的实体和关系对象。图 4-4 展示了 CityGML 中桥梁模型的局部 UML 图,图中可见该标准基于依赖、实现、组成、关联、泛化等关系类型,定义了实体间的连接。

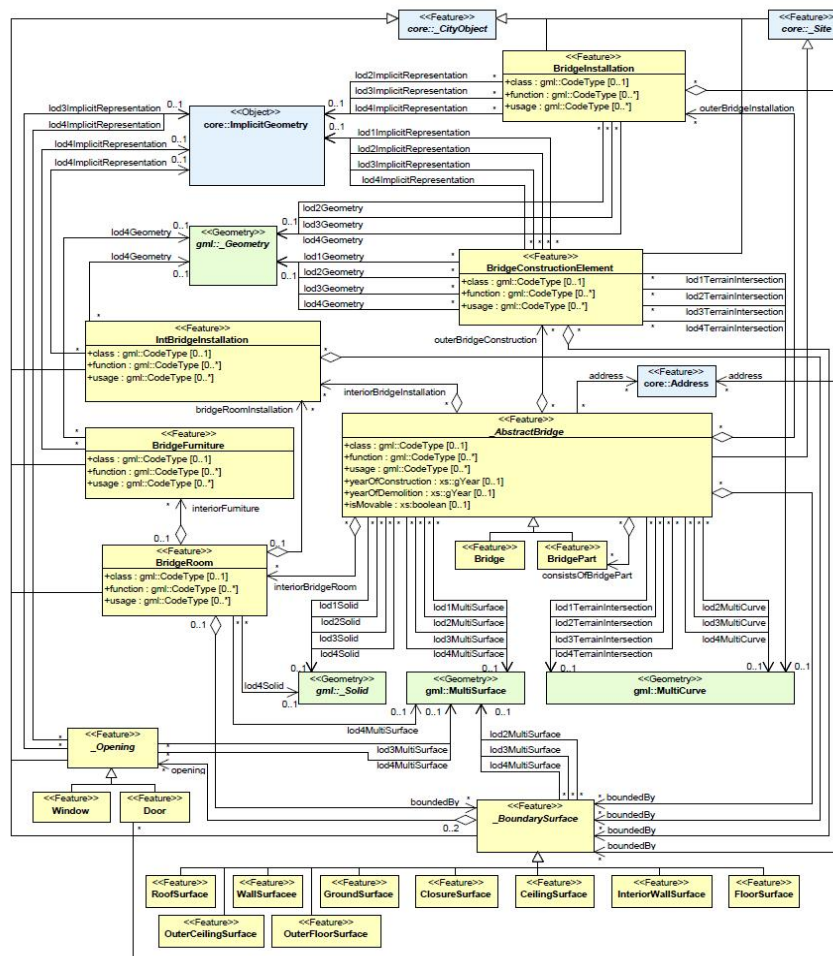


图 4-4 CityGML 桥梁模型部分 UML 图

SSN (Semantic Sensor Network) 是物联网技术领域的本体框架，适用于传感数据的信息表达，它基于本体建模的方法，通过本体和映射方法的模块化及模块的切割，定义类和关系模型。图 4-5 为 SSN 文档中的类和关系示意图，其中绿色部分为轻量化的 SOSA 本体框架。其中类和关系被封装为 Deployment, System, Procedure, Result 等八个模块，可用于表达传感器信息 (Sensor)、观测信息 (Observation)、被观测属性信息 (observedProperty) 等多类信息。

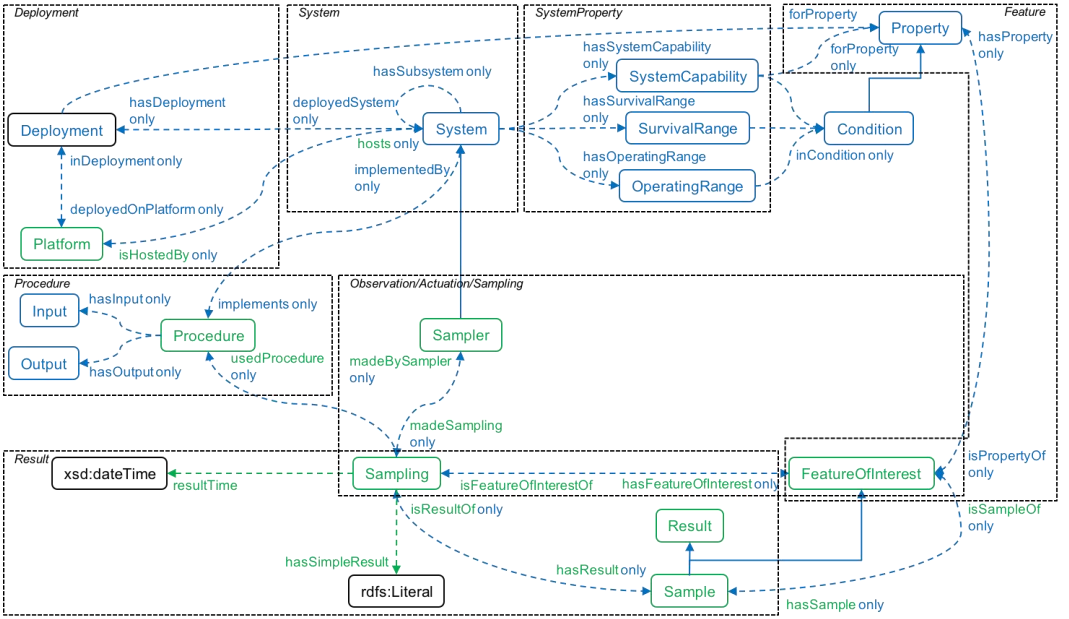


图 4-5 SSN/SOSA 的类和关系示意图

(3) 基于通用类扩展的数据模型语义扩展方法

在构建数据融合模型语义框架的基础上，考虑如何基于该框架进行语义扩展。一种方式是基于通用类扩展，即通过继承语义框架中实体和关系的表达范式，描述新定义的对象，用 EXPRESS 语言进行表达，实现实体集、属性集的扩展。建筑领域现有较为成熟的技术是基于 IFC 标准的信息模型语义扩展技术，将异构数据的表达统一至 IFC 框架。例如基于 IFC 标准的继承型关系类型扩展传感器监测信息的表达方式，如图 4-6 所示。其中定义了用以描述传感器属性信息的实体对象，定义了这些实体的属性类型、数据类型等，形成传感数据的描述框架。又如构件表观损伤信息的 IFC 标准扩展，基于 IFC 既有框架定义了桥梁病害信息的描述方式，实现病害对象的扩展和表达，如图 4-7 所示^[19]。

名称	属性类型	数据类型	定义
SensorType	IfcPropertySingleValue	IfcLabel	表示传感器芯片型号
SensorRange	IfcPropertyBoundedValue	IfcReal	表示传感器测量数据范围
SensorAccuracy	IfcPropertySingleValue	IfcReal	表示传感器测量精度
SensorInterval	IfcPropertySingleValue	IfcDuration	表示传感器数据采集间隔
SensorOutDataType	IfcPropertyEnumeratedValue	IfcLabel	表示传感器输出数据类型
SensorAlarmRange	IfcPropertyBoundedValue	IfcReal	表示传感器采集数据报警阈值

图 4-6 监测信息的 IFC 格式对象扩展

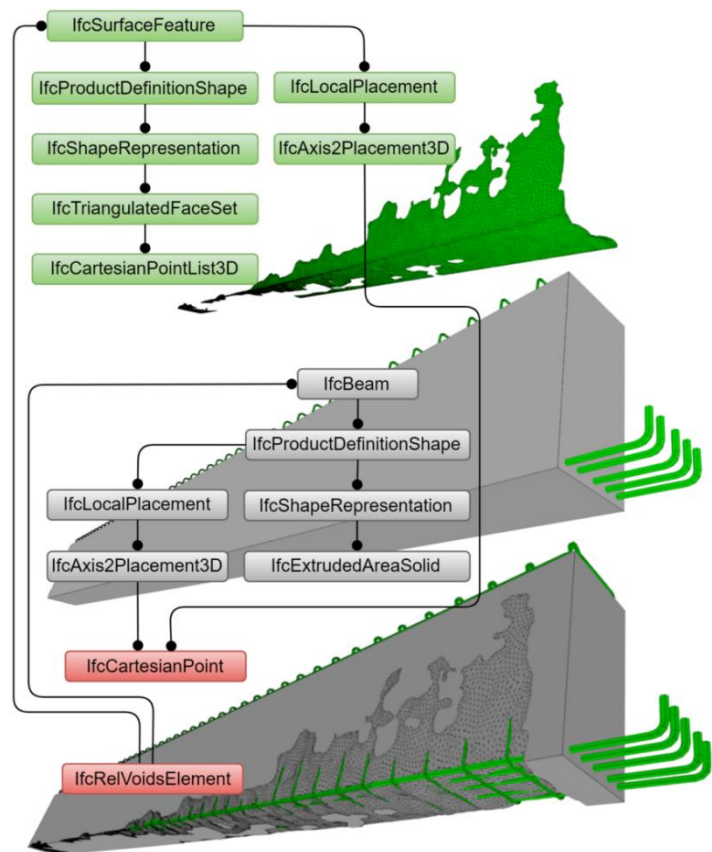


图 4-7 桥梁病害信息的 IFC 标准扩展与表达^[19]

(4) 基于转换对齐的数据模型语义扩展方法

另一种语义扩展方法是基于数据标准的转换对齐，即通过义信息映射、本体消歧、标准整合等方法，统一同一对象在语义结构、时空表达、数据存储、信息模型、接口协议等方面的规定，实现不同语义框架的融合。这一过程的实现一般基于空间关系、几何描述、操作模态、语义层次等特征的转换对齐和信息关联，也可基于已开发的扩展机制。在建筑领域，这一方法常用于 IFC 和 CityGML 予以框架的融合。

4.1.2 基于特征提取与数据融合的运维数据处理方法

针对传感数据、点云数据、图像数据等多种运维数据类型，先对数据进行预处理，提取数据特征和语义，再在同类异构的数据间实现数据融合，生成同构数据集的数据处理方法。

(1) 基于特征提取算法的数据集处理方法

对于传感数据、点云数据、图像数据等运维数据类型，分别采用信号处理、点云建模、图像处理等方法，对数据进行预处理与特征提取。

① 信号处理

信号处理方法用于对传感数据进行处理。傅里叶变换（FT）是基础的信号分析和处理手段，将复杂的信号转换为三角函数的线性组合形式，从而将信号由时域分析变换为频域分析。依据信号的周期性和连续性特征，不同种类的时域信号对应不同类型的傅里叶变换形式。这里以连续时间傅里叶级数（CFS）为例给出傅里叶变换的公式，如下所示。

$$f(t) = a_0 + \sum_{k=1}^{\infty} (a_n \cos k\Omega t + b_n \sin k\Omega t)$$

$$a_0 = \frac{1}{T} \int_{-T/2}^{T/2} f(t) dt$$

$$a_k = 2 \cdot \frac{1}{T} \int_{-T/2}^{T/2} f(t) \cos n\Omega t dt$$

$$b_k = 2 \cdot \frac{1}{T} \int_{-T/2}^{T/2} f(t) \sin n\Omega t dt$$

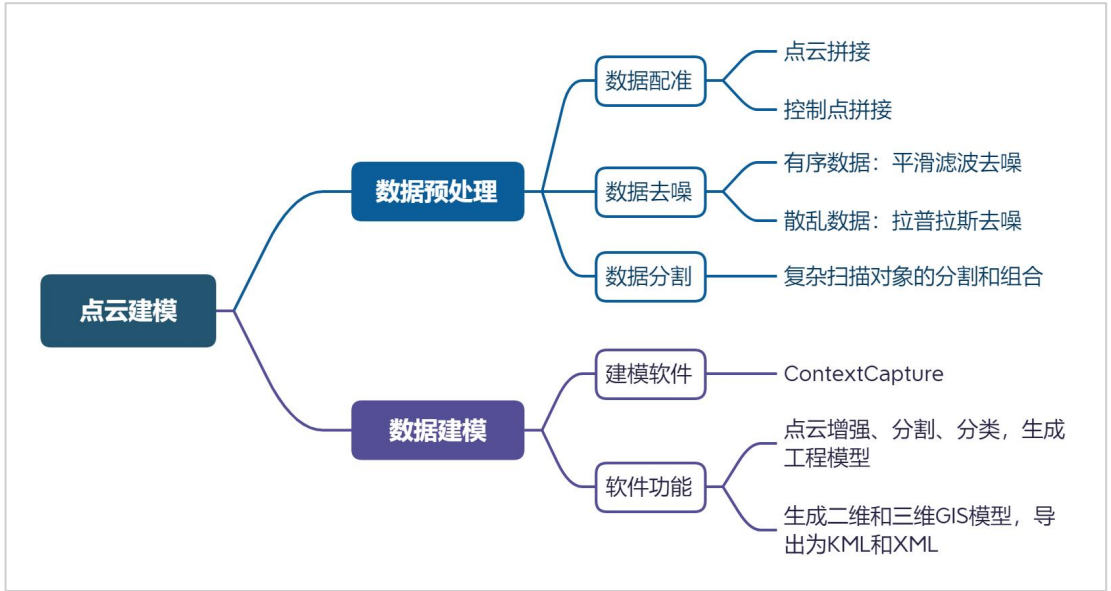
在工程应用中，快速傅里叶变换（FFT）是常用的方法，能将信号分解的时间复杂度降为 $O(n \log n)$ 。但傅里叶变换的问题在于：仅能获取信号频率分解结果，而无法获知这些频率存在的时间点，因此傅里叶变换擅长处理平稳信号，而无法较好地处理非平稳（时变）信号。为处理传感数据的时变特性，需要用到小波变换、Hilbert_Huang 变换等方法。

小波变换（Wavelet Transform, WT）将无限长的三角函数基转换成有限长的会衰减的小波基，在获取信号的频率分布的基础上，还能得知不同频率出现的具体时刻，因而在时域和频域上都有相当高的分辨率。Hilbert_Huang 变换（HHT）先对信号进行经验模态分解（EMD），然后将分解出的每个 IMF 分量进行 Hilbert 变换，同样能够较好地捕捉非平稳信号的局部特征。这两种方法共同成为传感数据时频域分析的重要手段。

② 点云建模

点云建模方法是通过挖掘点云数据的几何特性，利用特征提取算法针对点云的点、线、面特征进行三维重建，用以表达构筑物的复杂结构和细节问题。本课

题计划采用激光测量方法获取点云数据，利用数据中的三维坐标和激光反射强度进行构筑物几何形态的建模。通用的建模方法如图 4-8 所示。首先通过数据配准、去噪、分割，对点云数据进行预处理；然后借助通用建模软件 ContextCapture 对影像进行增强、分割、分类，生成工程模型，进一步可导出为 GIS 模型。



在通用方法的基础上，还需要结合构筑物特点，从点云模型中提取目标对象的特征和语义。基础性方法是基于点云平面的特征提取方法，即通过人工构造曲率、高程等特征的判别规则，提取建筑杆件，适用于特征明显、易于分割的构筑物，如门、窗等。而对于不易直接提取的目标对象，如裂缝、表皮损害等，可利用邻域搜索、聚类等机器学习方法进行特征边界点提取。进一步，对于不易于分割的目标，可利用基于深度学习的语义分割网络，对构筑物的细部结构实现划分，提取对象的特征和语义。

③ 图像处理

图像处理方法是对建筑运维过程涉及的遥感数据、影像数据进行解析，通过算法设计实现图像变换、编码、压缩、增强、复原、分割等功能，用以将图像数据转化为运维模型可识别的对象，方便图像数据与运维模型的信息整合。

图像变换过程包括几何变换、尺度变换、频域变换等。几何变换涉及平移、旋转、镜像、转置等；尺度变换涉及图像缩放及最近邻插值、线性插值、双三次

插值等插值算法；空间域与频域间变换是针对图像阵列较大的问题，利用傅立叶变换、沃尔什变换、离散余弦变换等间接处理技术，将空间域的处理转换为频域处理，不仅可减少计算量，而且可获得更有效的处理，例如傅立叶变换可在频域中进行数字滤波处理。

图像增强是在不考虑图像降质的前提下，突出图像中所感兴趣的部分。例如强化图像高频分量，可使图像中物体轮廓清晰，细节明显；强化低频分量可减少图像中噪声影响。采用的降噪方法包括中值滤波、高斯滤波、邻域平均法等。如图 4-9 所示即为海底构筑物的图像增强实例。另外可以使用图像（边缘）锐化方法，对轮廓边缘进行增强，例如梯度锐化，Roberts 算子、Laplace 算子、Sobel 算子等，便于目标提取。该技术适用于构筑物实景照片的特征增强。

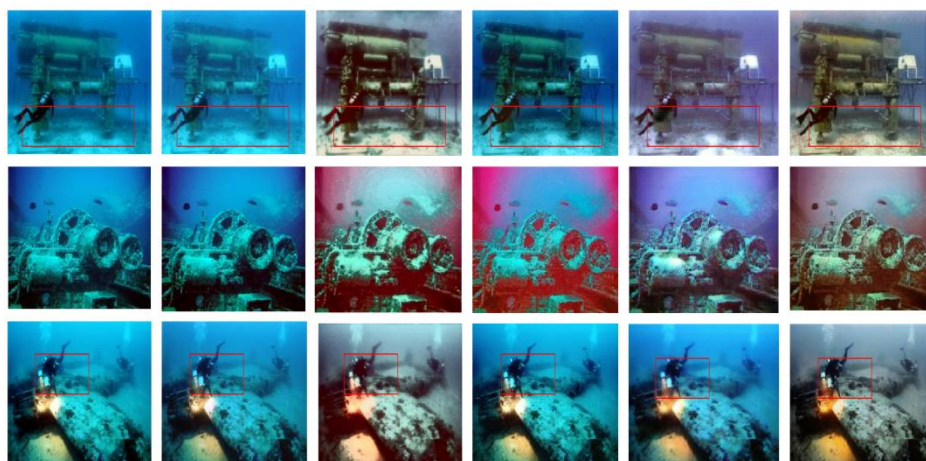


图 4-9 图像增强示意图

进一步，需要对图像进行分割，提取目标对象的边缘轮廓。可采用阈值分割（固定阈值分割、最优/OTSU 阈值分割、自适应阈值分割）、基于边界分割（Canny 边缘检测、轮廓提取、边界跟踪）、基于区域分割（区域生长、区域归并与分裂、聚类分割）等方法。常见用途是遥感影像的处理，例如基于遥感图像的海岸线提取（图 4-10）。

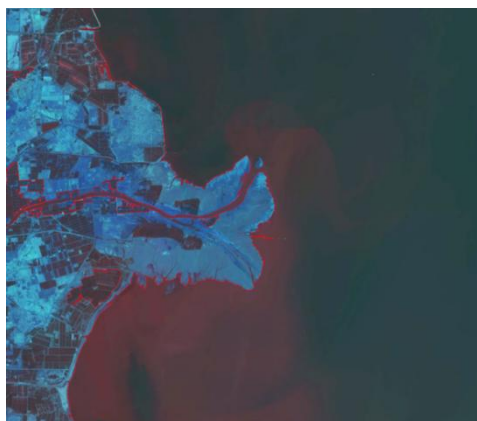


图 4-10 图像分割示意图

在此基础上，对图像特征进行提取，结合图像的几何特征（位置与方向、周长、面积、长轴与短轴、距离）、形状特征（矩形度、圆形度、不变矩、偏心率、多边形描述、曲线描述）、颜色特征（颜色直方图、颜色矩）等对目标对象进行识别，提取图片信息。此外，采用模板匹配、图像分类、缺陷识别等领域的既有算法，可实现图像特征的识别和语义的对应。模板匹配包括轮廓匹配、归一化积相关灰度匹配、不变矩匹配、最小均方误差匹配等方法；图像分类（识别）属于模式识别的范畴，主要过程是将图像经过某些预处理后，进行图像分割和特征提取，从而进行判决分类；表面缺陷的目标识别包括贝叶斯分类、K 最近邻(KNN)、人工神经网络(ANN)、支持向量机(SVM)、K-means 等常见算法。

（2）基于数据融合算法的同构数据集合成策略

从多来源所收集的运维数据在不同的应用场景下需要采用不同的融合层次，具体可分为：数据级数据融合（像素级数据融合）、特征级数据融合、决策级数据融合。

数据级融合是最低层次的数据融合，该层次的数据融合的任务是直接处理收集到的原始数据，在各种系统捕获的原始信息没有经过预处理之前就进行数据的综合和分析。该级别的数据融合由于数据量往往很大，数据的传输和处理都需要很大的系统开销，实时性差，但是这种融合层次的数据通常不会由于个别异常数据而造成融合结果有太大的偏差，因为数据级融合尽可能多的保留了融合数据所反映的所有信息。

数据级数据融合通常用于多源图像融合分析、图像识别和理解、同质雷达波

形的信号处理。

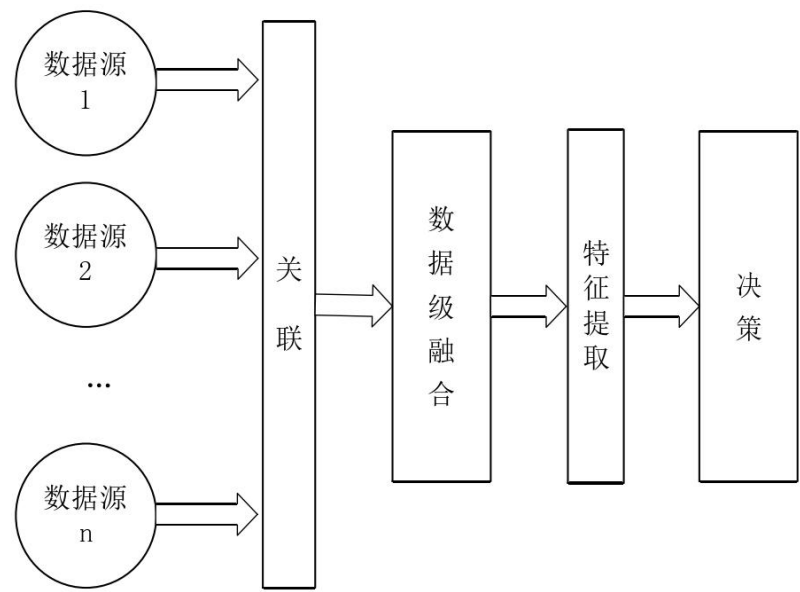


图 4-11 数据级融合过程图

特征级融合属于数据融合的中间层次，该层次的数据融合能有效避免数据级融合中大量数据处理对数据融合系统带来较高的处理成本，又能解决决策级融合对于识别目标的特征细节上的忽略。特征级融合由于数据的传输量较小所以对数据的传输带宽要求也比较低。但是由于在传感器数据的传输的过程中会受到各种各样的干扰导致特征级数据融合的可靠性受到影响。特征级数据融合的优点也很突出，那就是实现了大量传感器数据的压缩，这样的压缩数据占用的通信带宽较低有利于融合系统的实时融合任务的处理。在特征级数据融合阶段所提取的关于融合目标的特征信息直接与最终的决策级数据融合阶段的决策有着很大的关系，特征级数据融合的最终结果能够最大程度地给决策分析所需的关于融合目标的特征信息。

特征级数据融合常用的算法有：加权平均法、卡尔曼滤波法、人工神经网络等。

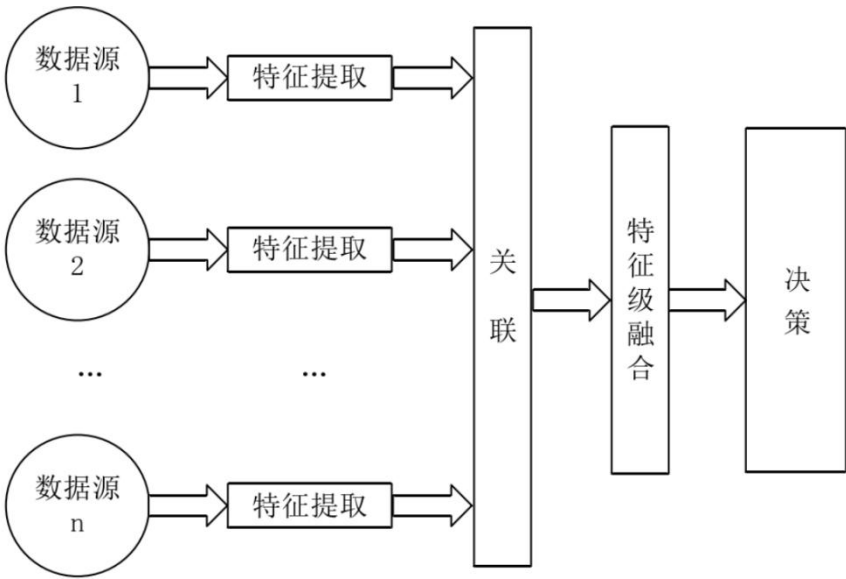


图 4-12 特征级融合过程图

决策级融合属于高层次的数据融合，该层次的数据融合将对特征级融合的结果进行进一步的融合以便于对指挥决策提供依据。根据决策级融合阶段的输出是一个综合而又全面的决策结果，在理论上这个综合而又全面的决策结果比任何单源信息的决策更加精确和有效。决策级融合的优点是具有很高的灵活性，由于决策级融合的抽象程度更高，对信息的压缩提炼更加彻底，所需输出的数据量很小，融合系统对数据的传输带宽要求要比数据级融合要低得多；决策级数据融合由于对反映真实融合场景和监测融合目标的不同维度的多源异构数据进行有效的融合，所以决策级数据融合的容错性很强，能对错误数据做出正确的反馈；通信容量小，分布式多传感器数据融合采集信息多样，抗干扰能力强。

决策级数据融合算法主要有：贝叶斯推理，模糊推理、D-S 证据理论。

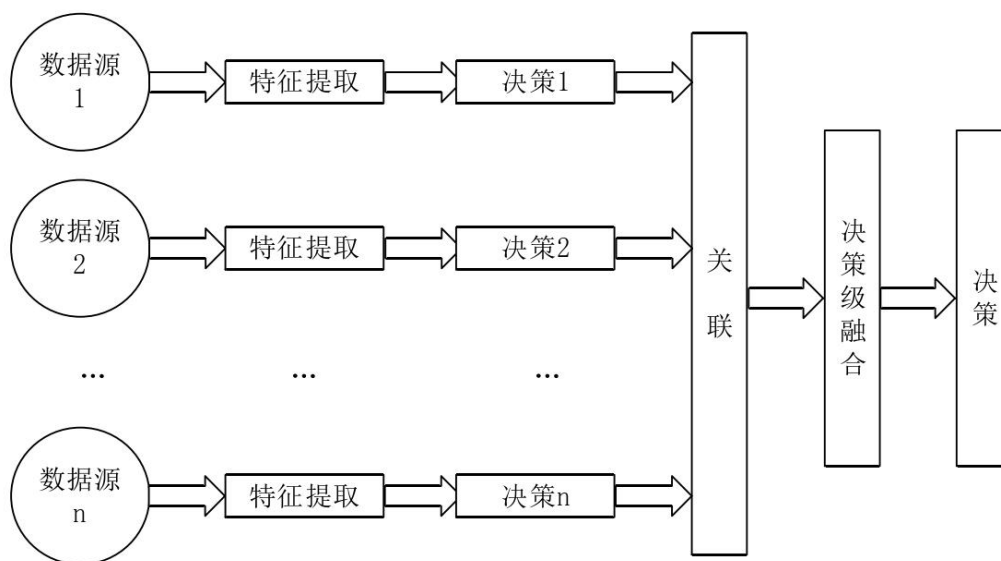


图 4-13 决策级融合过程图

本课题计划采用的数据融合算法包括：

① 加权平均算法

加权平均数是数据级融合中最简单易行的方法，在数据级融合中应用较为广泛，该方法将数据源所提供的一组有冗余信息的数据赋予加权系数后做加权平均的处理：用 w_i 代表赋予数据源 i 的权重，用 t_{ij} 表示数据源 i 对决策 j 的支持度，那么 $\sum w_i t_{ij}$ 就是综合的各个证据对各决策的支持程度，得到的结果即为数据融合的结果，这种方法简单直观但不够精确，且在权重的确定上，会受一些主观因素影响。

② 贝叶斯推理

贝叶斯推理用于多源数据融合时，其基本的原理是利用概率原则组合传感器信息并将不确定性通过概率表示出来，然后计算在给定多源数据这一条件下，某个假设为真的后验概率。贝叶斯方法要求系统的决策 $A_1, A_2, A_3, \dots, A_m$ 之间是相互独立的，对于数据源给出的属性（证据） $B_1, B_2, B_3, \dots, B_n$ ，计算每一个证据在各假设为真的条件下的概率 $P(B_i|A_j)$ ，从而可以得到 n 个证据下的联合条件概率：

$$P(B_1, B_2, \dots, B_n|A_j) = P(B_1|A_j) \cdot P(B_2|A_j) \dots P(B_n|A_j)$$

利用贝叶斯公式，可以得到：

$$P(A_j|B_1, B_2, \dots, B_n) = \frac{P(B_1, B_2, \dots, B_n|A_j) \cdot P(A_j)}{P(B_1, B_2, \dots, B_n)}$$

在实际应用中,通常选用上式计算出来的最大结果所对应的目标类型为最后的判决结果,这种判决法则称为最大后验概率(MAP)判决准则。

贝叶斯推理相较于传统的推理方法主要有以下几个优势:首先当给定某一证据时,它能提供一种计算某一假设事件概率的方法;其次,贝叶斯公式允许嵌入一些关于假设事件发生的先验知识;最后,贝叶斯推理允许使用主观概率,这些主观概率包括假设事件发生的先验概率,以及在假设事件发生的条件下出现证据的概率(似然函数)。贝叶斯推理的最后这个特性使它能适用于多传感器的数据融合,因为它不需要各传感器的概率密度函数。

③ 卡尔曼滤波法

卡尔曼滤波法用于动态环境中多传感器信息的实时融合,其算法核心是计算各传感器数据之间的加权平均值,其中,权值与测量方差成反比。在实际应用中,通过调节各传感器的方差值来改变权值,从而得到更可靠的结果。

卡尔曼滤波融合算法计算模型如下:

$$\begin{cases} X_{k+1} = \Phi X_k + \omega_k \\ Z_k = HX_k + v_k \end{cases}$$

其中, X 是系统状态矢量, Φ 代表状态转移矩阵, w 是系统噪声, Z 是观测矢量, v 是观测噪声, H 是系统观测矩阵。采用最小方差估计方法根据测量值 Z 估计系统状态矢量 X 的 Kalman 滤波方程,状态更新包括时间更新和测量更新两个部分。测量更新过程的数据修正以本次测量值和前一次预估值之间的误差为基准,更新本轮预估值。

卡尔曼滤波法最适用于的情况是:线性模型系统且与传感器之间的误差刚好是高斯白噪音模型,在这种情况下卡尔曼滤波法能够提供最优估计。

④ D-S 证据理论

下图给出了 Dempster-Shafer 数据融合过程的框图。在此框图中每个传感器都有一组可观察量,这些可观察量都体现了物体及他们所在环境的某些信息。各传感器对这些可观察量再利用各分类算法(传感器级融合)进行分类。这里对每

个传感器 k ($k=1, \dots, N$) 所能得到的知识就是对物体类型的判别 (用 O_i 表示, $i=1, \dots, n$) 赋予一个 0 到 1 之间的概率分配 $m_k(O_i)$, 这个概率分配值反映了对该判决的确信程度。概率分配值越接近于 1, 说明该判决越有明确的证据支持, 从而对物体类型的不确定程度就越低。然后各传感器的各概率分配值通过 Dempster 规则融合, 从而再选出某种假设, 使该假设能被在各传感器上已经得到的绝大多数证据所支持。

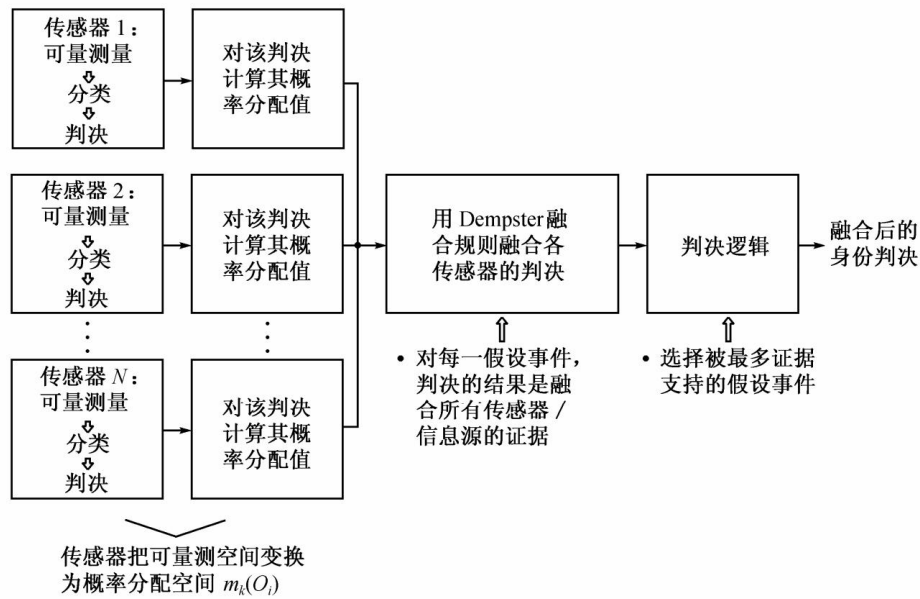


图 4-14 Dempster-Shafer 数据融合过程

在 DS 证据理论中, 由互不相容的基本命题 (假定) 组成的完备集合称为识别框架, 表示对某一问题的所有可能答案, 但其中只有一个答案是正确的。该框架的子集称为命题。分配给各命题的信任程度称为基本概率分配 (BPA, 也称 m 函数), $m(A)$ 为基本可信数, 反映着对 A 的信度大小。

识别框架 Θ 上的函数 $m(x)$ 是 $2^\Theta \rightarrow [0,1]$ 的映射, A 为 2^Θ 的任一子集, 记作 $A \subseteq 2^\Theta$, 如果函数 $m(x)$ 满足条件:

$$\begin{cases} m(\phi) = 0 \\ \sum_{A \subseteq 2^\Theta} m(A) = 1 \end{cases}$$

则 $m(x)$ 就是 2^Θ 上的基本概率分配函数, 也称为质量函数或 mass 函数。

$m(A)$ 为命题 A 的基本概率分配函数, 表示某个证据对命题 A 的信任程度, 称为基本置信度, 空集 ϕ 的基本置信度为 0。

信任函数 $\text{Bel}(A)$ 表示对命题 A 的信任程度。信任函数（Belief Function）：集合 $A \subseteq 2^\Theta$ 是识别框架 Θ 的任一子集， A 中全部子集对应的基本置信度之和称为信任函数 $\text{Bel}(A)$ ，即：

$$\text{Bel}: 2^\Theta \rightarrow [0,1]$$

$$\text{Bel}(A) = \sum_{B \subseteq A} m(B) = 1$$

似然函数 $\text{Pl}(A)$ 表示对命题 A 非假的信任程度，也即对 A 似乎可能成立的不确定性度量。对于识别框架 Θ ，存在映射 $2^\Theta \rightarrow [0,1]$ ， A 为识别框架内的任一子集，似然函数 $\text{Pl}(A)$ 的意义是：对 A 的“非假信任度”，即对 A 可能成立的不确定性度，是对“ A 非假”这一命题的信任程度，因此似然函数也被称为不可驳斥函数或上限函数，此时有：

$$\text{Pl}(A) = \sum_{B \cap A \neq \emptyset} m(B) = 1 - \text{Bel}(\bar{A})$$

实际上， $[\text{Bel}(A), \text{Pl}(A)]$ 表示 A 的不确定区间， $[0, \text{Bel}(A)]$ 表示命题 A 支持证据区间， $[0, \text{Pl}(A)]$ 表示命题 A 的置信区间， $[\text{Pl}(A), 1]$ 表示命题 A 的拒绝证据区间。设 m_1 和 m_2 是由两个独立的证据源（传感器）导出的基本概率分配函数，则 Dempster 组合规则可以计算这两个证据共同作用产生的反映融合信息的新的基本概率分配函数。

⑤ 模糊推理

模糊集理论（FST, Fuzzy Set Theory）为处理现实世界中的不确定性、不精确性和复杂性提供了可替代框架。任何理论都有其研究对象，对于某个问题的有关议题的集合称为该问题的论域。对于论域 U 上的一个集合 A ，对于 $\forall u \in U$ ，定义：

$$C_A(u) = \begin{cases} 1 & \text{当 } u \in A \\ 0 & \text{当 } u \notin A \end{cases}$$

则称 $C(u)$ 为 A 的特征函数，特征函数 $C_A(u)$ 在 $u=u_0$ 处的值 $C(u_0)$ 称为 u_0 对 A 的隶属度。隶属度越接近 0，表示 u_0 属于集合 A 的程度越低。

在论域 U 中， μ_A 是将任何 $u \in U$ 映射为 $[0,1]$ 上某个值的函数，即：

$$\mu_A: U \rightarrow [0,1]$$

$$u \rightarrow \mu_A(u)$$

那么 μ_A 就是论域 U 上的隶属函数, $A = \{ \mu_A(u) | u \in U \}$ 称为论域 U 上的一个模糊集, μ_A 为 u 对模糊集 A 的隶属度。模糊集合完全由其隶属函数确定, 即一个模糊集合与其隶属函数是等价的。

模糊集理论在精确性解的问题上不如传统方法, 但是当系统高度线性化或者是在应用场景下不能得到充分代表其操作的模型时, 模糊控制就能够显示出非常高的鲁棒性。因为在这样的情况下, 模糊集理论可以用简单而又判断力强的规则来代替那些不适当的模型。

⑥ 神经网络

神经网络可以对复杂的非线性映射进行模拟, 且由于具备很高的运算速度、联想能力、适应性、容错能力和自组织能力等特点, 神经网络能够很好地适应多源数据融合的处理要求。BP (Back Propagation) 神经网络是目前使用最普遍的一种神经网络, 它能够对输入的样本进行学习, 采用梯度搜索技术, 从输入层开始, 经隐含层单元, 最终传向输出层。BP 神经网络结构模型如下图所示:

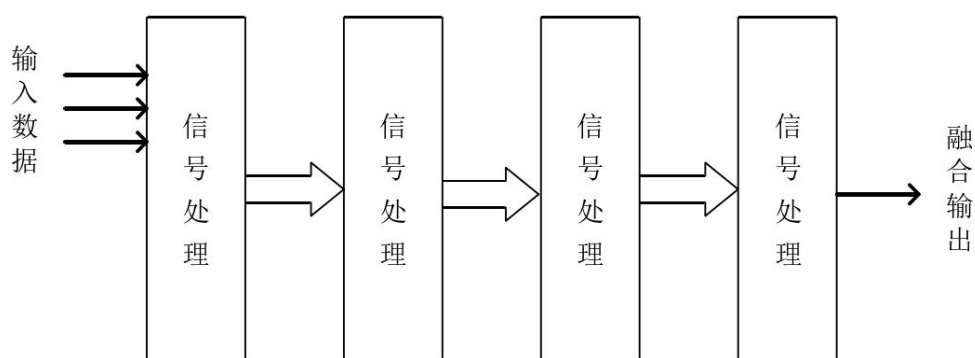


图 4-15 神经网络在数据融合中的应用

神经网络的自学习能力, 可以根据输入的样本之间的相似性制定一定的规则, 也可以通过自学习来得到不确定推理机制, 由此, 神经网络可以在数据融合领域得到应用。

4.1.3 基于特征匹配与语义融合的异构数据关联融合方法

针对建筑运维过程中多模态异构数据集,采用特征匹配和语义融合的方法实现数据的关联融合。

(1) 基于特征匹配的多源异构数据集融合方法

① 特征向量映射与集成

特征向量的映射与集成是以向量形式描述各个特征,并将各特征融入统一的高维向量,实现特征之间的串联的方法。这种方法能够将异构数据映射至一个统一的向量空间,在此空间集成数据中的各类特征。当信息的维度不大、信息量较少、各个特征的独立性较强时,该方法较为适用,串联的信息可读性较好。而当信息维度较高、来源多样时,该方法容易造成特征的过拟合,也容易忽略部分特征。在应用时,该方法可用来处理属性特征、颜色特征等小维度信息的融合。

② 特征对齐方法

特征对齐是对于异构数据集中描述同一对象的元素或分支寻找对应关系,通过格式对齐实现异构数据集之间的特征匹配的方法。该方法适用于统一编码的特征信息,可用于处理运维数据空间维度和时间维度的对齐。结合数据集的坐标特点,可能用到以卷积神经网络为代表的机器学习方法,以实现特征的自动对齐。

③ 机器学习与深度学习方法

机器学习与深度学习方法包括分类器模型、生成模型等。分类器模型包括SVM、神经网络等算法。一个分类器可以由多模态特征向量构建,用于训练同维度的向量。生成模型以贝叶斯网络为代表,可以从多维度向量中进行学习,能够处理可变长的向量。

(2) 基于语义融合的多源异构数据集融合方法

① 基于自动编码和自推理的语义生成方法

由于数据融合模型定义是一种表达范式,语义覆盖力有限,因而随着数据量的扩充,语义信息需要有自动生成的潜力。自动编码和自推理是两种支持语义自动生成的方法。自动编码是基于共享语义框架中的表达范式进行统一编码,实现

实例对象自动生成的方法。目前该方法常用于文本信息挖掘，采用监督学习的方式，预先构造标记过的输入信息，加入模型后进行训练，如图 4-16 所示。本课题考虑将该方法的使用场景迁移至数据模型的语义系统，将标记数据替换为符合语义描述逻辑的语义扩展样本，实现有约束的语义自动编码。

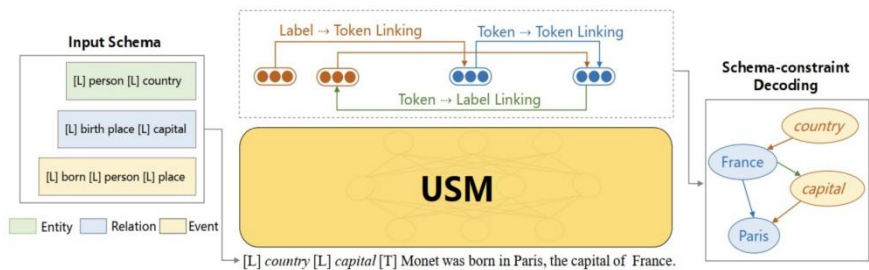


图 4-16 基于监督学习的语义自动编码方式示意图

自推理方法是将 RDF 框架中的语义元素视作形式逻辑的组成，通过语义间的关系特征，将多源数据中挖掘出的语义信息补充至数据模型的语义框架，实现语义扩充的方法。该方法常借助于语义系统中的层状关系，进行层状自推理。

② 基于语义相似性的异构数据互补方法

多源异构数据集中存在可以互相补充的语义信息，利用不同对象间潜在的语义相关性（或相似度）融合不同的数据集，可以形成更为完整的语义体系，从而实现数据集间的信息互补。深度语义匹配模型 DSSM（Deep Structured Semantic Models）是一种计算语义相似度的前沿算法，由输入层、表示层、匹配层构成。如图 4-17 所示，输入层将语义映射到向量空间，经过表示层中含有多个隐层的 DNN 神经网络，输出一个低维（128 维）语义向量。匹配层中利用余弦相似度 $R(Q, D)$ 表示语义样本间的相似度，再通过 softmax 函数将语义相似度转换为后验概率 $P(D^+|Q)$ 。在训练阶段，通过极大似然估计最小化损失函数，将残差在表示层中反向传播，最终通过随机梯度下降使模型收敛。^[20]如下所示为该过程中相似度、后验概率、损失函数的计算公式。

$$R(Q, D) = \text{cosine}(y_Q, y_D) = \frac{y_Q^T y_D}{\|y_Q\| \cdot \|y_D\|}$$

$$P(D^+ | Q) = \frac{\exp(\gamma R(Q, D^+))}{\sum_{D'} \exp(\gamma R(Q, D'))}$$

$$\text{Loss} = -\log \prod_{(Q, D^+)} P(D^+ | Q)$$

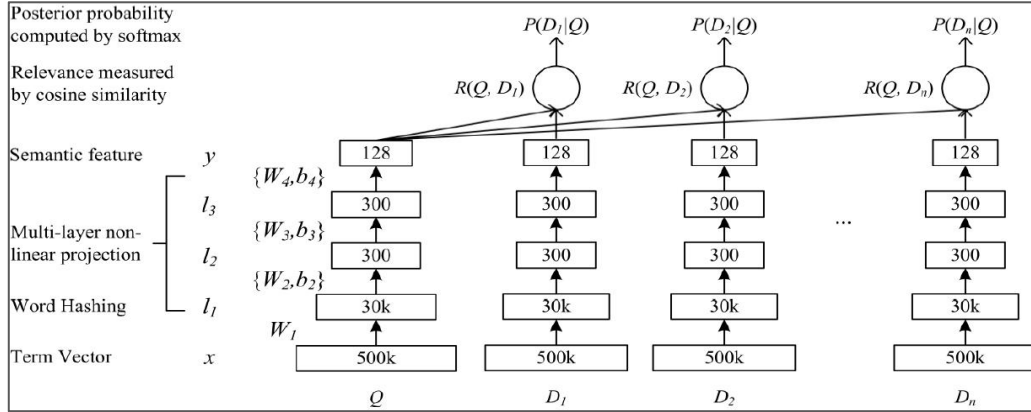


图 4-17 排序相似性经典模型 DSSM 示意图^[20]

③ 基于多层语义匹配的异构数据迁移融合算法

基于多层语义匹配的异构数据迁移融合算法是另一种语义融合的算法。采用深度学习方法，通过多层非线性转化，更好地抽象数据高层语义特征，促进相似语义融合，有效消解异构数据集间的语义偏差对迁移融合结果的影响。该算法设计了多层语义统一深度网络架构，在每层中对异构数据进行深度语义匹配，并利用顶层输出特征对网络进行整体相关耦合。^[21]通过优化模型匹配的目标函数，得到异构数据私有深度匹配网络和模态高层语义共享空间；在共享空间内完成有标签源域模态知识到目标模态无标签任务的迁移学习。该算法可用于文本迁移分类，具有较好的鲁棒性。

4.1.4 基于分布式框架的运维数据整合和模型加载策略

针对运维数据融合模型，研究分布式框架下的数据整合和模型加载策略，从提升数据管理效率的角度提供数据服务的解决方案。

(1) 分布式架构下的数据融合策略

在分布式结构中，每一个源节点在单独处理后，会将结果发送至融合节点，

在此处进行局部的数据融合。因此需要配置若干中间融合节点，以便处理接收到的其他节点的信息。考虑分布式结构中数据节点的处理方法，对融合节点的布局、配置和节点处数据融合的处理流程进行设计，用分布图、流程图等方式进行表达。

（2）数据动态储存与管理技术

在分布式框架下，对于分布在不同地点的各个节点，通过网络实现节点间的通信和数据传输，客户在使用时无需关注储存的节点，只需像使用本地文件一样对数据进行管理和储存。考虑通过集群应用、分布式系统等功能，设计实现云储存和设备协同技术，搭建对外提供数据存储和业务访问功能的数据服务系统。具体涉及数据融合模型的文件传输协议、文件储存操作、数据库管理、数据加速处理方法等内容的设计。

（3）数据模型轻量化处理方法

针对数据模型的储存轻量化、传输轻量化、显示轻量化技术进行研究。存储轻量化方法采用模型的网格简化技术，考虑重要性和原网格数量大小，进行三角形网格简化，对模型中部分节点和边进行折叠。传输轻量化方法考虑几何信息、属性信息的轻量化，传输前进行数据压缩，在应用端对模型文件进行解压，缓解网络传输压力。显示轻量化方法考虑采用渐进网格技术、背面剔除基数、深度检测技术、细节层次技术、渐变面法向量等技术，在满足模型表达尺度和可视化效果要求的前提下，尽量较少模型显示过程中的资源占用。

（4）云-边-端层次加载策略

针对数据融合模型文件较大、数据传输效率低、场景加载较慢等问题，考虑设计“云-边-端”框架下的模型层次加载策略，按照几何、语义、用户的相关性构建加载优先级顺序，划分云、边、端层的加载任务。云层主要负责模型的转换与拆分、数据的储存与同步；边层负责数据同步、计算优先级、用户相关性统计等；端层负责渲染、交互、缓存等功能。

4.2 技术路线

本课题的技术路线分为四个部分，如图 4-18 所示。

第一部分是运维数据的标准化模型构建。根据数据组成、特征、内涵、形式和应用模式，建立数据模型表达的标准化体系；参考建筑领域、物联网领域既有数据标准，定义数据融合模型程式化、标准化语义描述框架；在此基础上开发语义扩展的方式，形成数据融合的标准化模型。

第二部分是同构数据集生成策略。针对传感数据、点云数据、图像数据等不具备标准化语义体系的运维数据类型，采用信号处理、点云建模、图像处理等方法对数据进行预加工，提取数据的结构和语义特征；而后在同类异构数据集间进行数据级、特征级数据融合，生成同构数据集。

第三部分是异构数据集关联融合策略。针对不同模态的运维数据，采用特征匹配和语义匹配的方式对异构数据进行整合。特征匹配针对可以用向量表示的数据特征，通过特征向量映射、特征对齐、机器学习等方法在特征层级对数据进行融合。语义匹配针对将异构数据的语义自动整合进数据标准化模型的过程，考虑自编码和自推理、语义相似性计算、语义迁移融合等方法。

第四部分是分布式框架下的数据服务搭建。借助专题二开发的分布式架构，采用分布式数据融合、数据动态储存管理等技术，对各类节点及云端处的数据融合、储存、管理方法进行设计；并利用模型轻量化技术、层次加载技术，对“云-边-端”协同框架下运维数据融合模型的加载方式进行设计。二者结合，实现更高效的数据动态存储、管理与可视化，提升运维信息模型的应用价值。

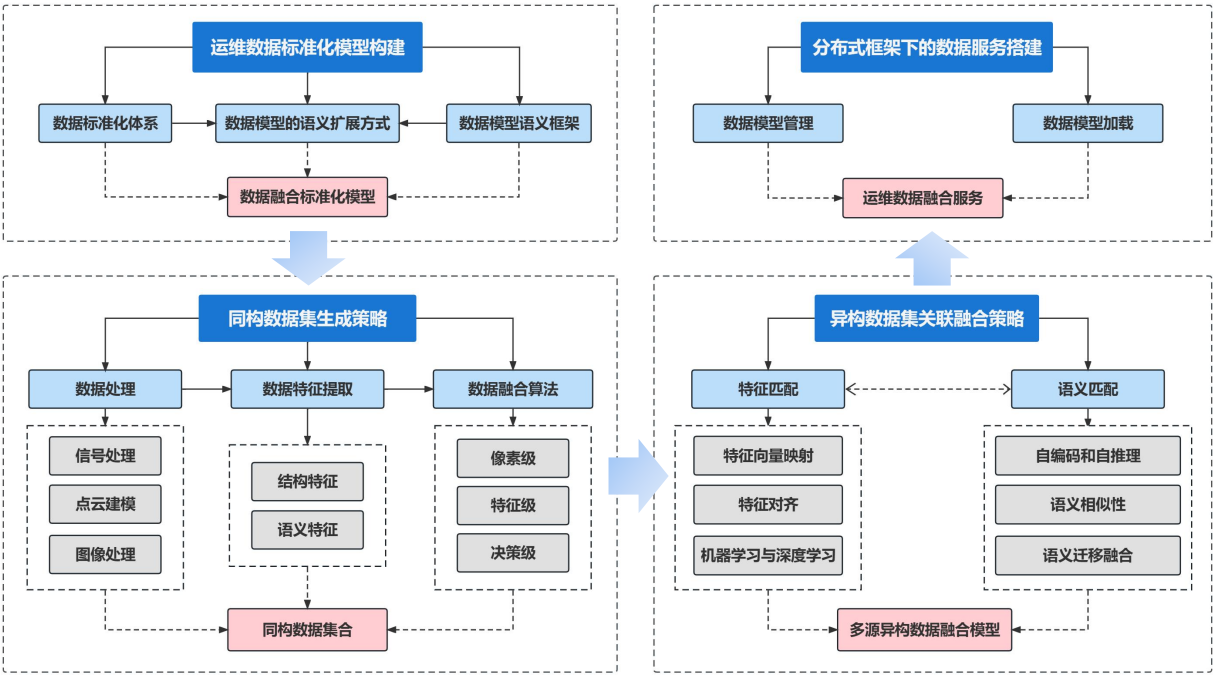


图 4-18 课题二子课题三总体技术路线图

5 基础研究与研发进展

5.1 研究基础

在土木工程信息研究领域，课题组通过集成 BIM、物联网、人工智能等技术，结合我国重大工程建设，面向市政工程、大型公建、道路桥梁、轨道交通等目标对象，在其理论研究、技术攻关、标准编制、软硬件研发及应用方面，以及基于 BIM 的数字化运维信息理论体系、工程性能模拟与分析技术、工程数据挖掘与分析技术等领域取得了国内外有广泛影响力的科研成果。

课题组 10 多年来一直以信息模型技术为核心，集成物联网、互联网+、人工智能等技术，结合我国重大工程建设，面向滨海智慧城市、住宅建筑、大型公建、道路桥梁、轨道交通等目标对象，开展了海洋与土木工程信息技术有关科学问题和工程方法的研究，在发展基础科学研究、研发自主产权平台、促进行业技术进步等方面有所突破与创新。本专题是胡振中老师主持的国家自然科学基金面上项目“基于 BIM 的大型公共建筑运维期安全管理技术研究”（编号：51478249）、面上项目“信息驱动下既有建筑物/群多尺度性能模拟与分析技术研究”（编号：

51778336)以及参与的多个其他科研项目中关于信息模型研究及成果基础上的进一步深化和拓展,具有延续性和创新性。

课题组已有的相关研究基础主要包括以下几个方面:

(1) 完善了基于 BIM 的数字化施工与运维信息理论体系

课题组提出了集产品模型、过程模型及决策模型于一体的 BIM 结构体系,以及建筑信息子模型(Sub-BIM)理论,建立了多层次信息模型结构和分阶段递进式 BIM 建模方法,完善了面向数字化施工和运维管理的 BIM 数据标准,研发了数据交互接口,搭建了相应的集成管理技术体系。在此基础上,研发了基于 BIM 的云平台架构,实现了海量异构工程信息的智能感知、云存储、PB 级数据管理和高效分析应用,支持各专业多参与方的信息共享和协同工作。本创新成果突破了传统工程信息管理和应用方式低效落后的瓶颈,为 BIM 模型构建、共享和管理提供了理论、标准和框架支撑,解决了 BIM 应用中面临的海量信息高效存储与管理的问题,从而完善了数字化施工和运维管理的信息技术体系。

(2) 研发了基于 BIM 的数字化智能运维管理技术

课题组研究了基于国际 IFC 标准的 BIM 扩展机制并开发了自主产权的数据接口,实现了项目竣工信息模型的电子化集成交付,支持数字化的运维期日常管理、维护维修管理和应急管理;研究了非关系型监测信息的存储和表达机制,以及面向多协议的监测和传感技术,实现了动态监测数据与 BIM 静态数据的集成;研究了基于人工智能的运维数据挖掘技术,实现了海量运维数据的关联关系、因果关系的识别与归纳,从而对设备运行状况进行分析、比较和预警预报。本创新成果突破了传统运维管理缺乏技术手段和软件工具的瓶颈,提高了运维过程智能化管理水平,形成了面向数字化运维的管理技术体系。

(3) 建立面向大型公共建筑施工和运维管理的信息模型

课题组基于所提出的 BIM 结构体系和 Sub-BIM 理论,曾建立面向大型公共建筑施工和运维管理的信息模型,以及基于多尺度 BIM 的数据引擎,如下图所示。该模型能为多源异构运维数据的融合模型的建立提供理论与模型基础。

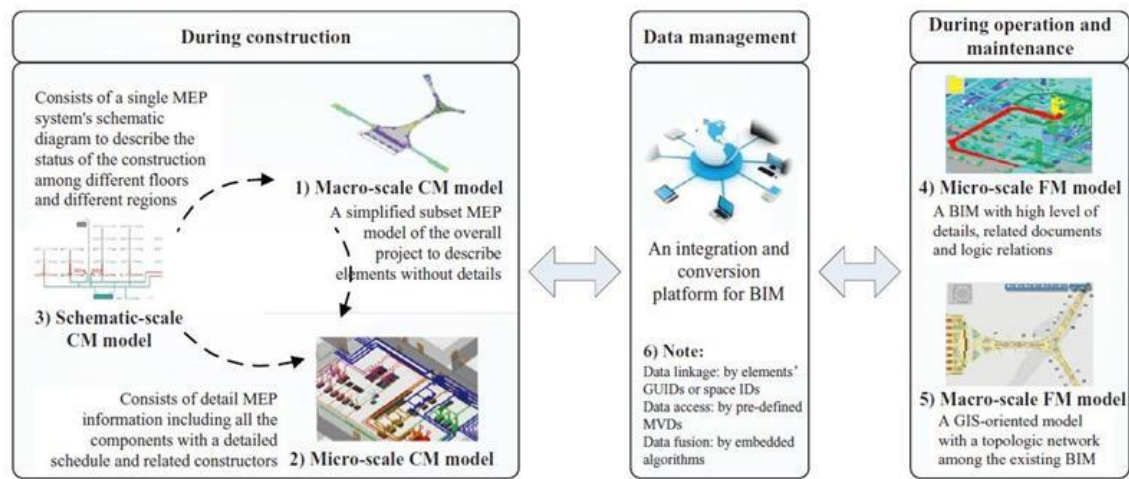


图 5-1 运维信息模型及信息引擎

(4) 研发基于 BIM 的大数据管理平台与 BIM 应用系统

课题组曾研发了基于 BIM 的大数据管理平台以及多个 BIM 应用系统，如下图所示，并在二十余个大型工程中得以成功应用。这些系统能处理海量的、精细化的建筑单体信息，平台承载数据量可超过 50 万个构件，超过 1000 万个三角形，以及超过 100G 的建筑属性信息。这些大型平台与系统的研发，为本项目搭建基于数据融合的智能运维数据服务提供了技术和经验支持。

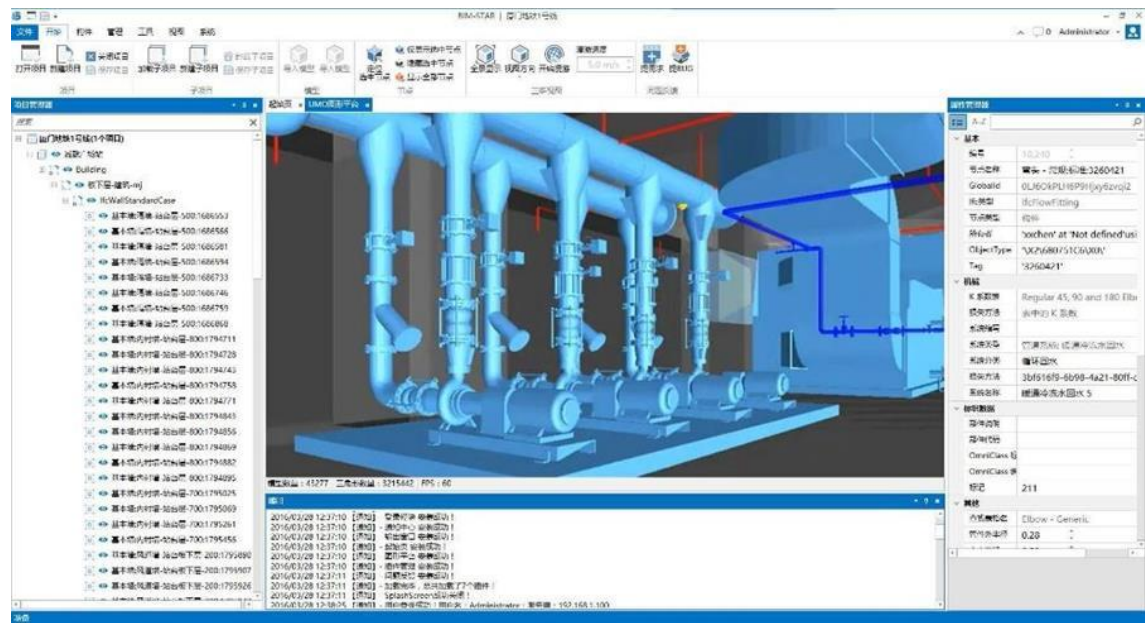


图 5-2 面向机电设备运维的 BIM 系统

课题组已有的相关研究成果主要包括以下几个方面：

(1) 与项目相关的科技计划

- 1) 主持“十三五”国家重点研发计划课题：乡村住宅设计与建造技术应用信息平台（2018YFD1100900），经费 80 万元，执行时间：2018.12-2022.12，在研。
- 2) 主持国家自然科学基金项目（面上）：信息驱动下既有建筑物/群多尺度性能模拟与分析技术研究（51778336），经费 60 万元，执行时间：2018.01-2021.12，结题。
- 3) 主持“十三五”国家重点研发计划子课题：基于能耗监测系统的建筑及其机电系统标准化大数据提取技术研究（2017YFC0704200），经费 37.82 万元，执行时间：2017.07-2020.06，结题。
- 4) 主持国家自然科学基金项目（面上）：基于 BIM 的大型公共建筑运维期安全管理技术研究（51478249），经费 85 万元，执行时间：2015.01-2018.12，结题。

(2) 与项目相关的部分代表性研究论文

- 1) L.T. Wu#, J.R. Lin, S. Leng#, J.L. Li, **Z.Z. Hu***. Rule-based information extraction for mechanical-electrical-plumbing-specific semantic web[J]. *Automation in Construction*, 2022, 135:104108.
- 2) **Z.Z. Hu***, S. Leng#, J.R. Lin, S.W. Li, Y.Q. Xiao#. Knowledge Extraction and Discovery Based on BIM: A Critical Review and Future Directions[J]. *Archives of Computational Methods in Engineering*. 2021, 1-22
- 3) **Z.Z. Hu***, S. Yuan#, C. Benghi, J.P. Zhang, X.Y. Zhang#, D. Li, M. Kassem. Geometric optimization of building information models in MEP projects: Algorithms and techniques for improving storage, transmission and display[J]. *Automation in Construction*. 2019, 107: 102941.
- 4) Y.Y. Zhang#, **Z.Z. Hu**, J.R. Lin*, J.P. Zhang. Linking data model and formula to automate KPI calculation for building performance benchmarking[J]. *Energy Reports*, 2021, 7, 1326-1337.
- 5) S. Yuan#, **Z.Z. Hu**, J.R. Lin, Y.Y. Zhang*. A Framework for the Automatic Integration and Diagnosis of Building Energy Consumption Data[J]. *Sensors*, 2021, 21(4): 1395.
- 6) S. Leng#, J.R. Lin, **Z.Z. Hu*** and X.S. Shen. A hybrid data mining method for tunnel engineering based on real-time monitoring data from tunnel boring machines[J]. *IEEE Access*,

2020, 8: 90430-90449.

- 7) Y.W. Zhou#, **Z.Z. Hu**, J.R. Lin*, J.P. Zhang. A review on 3D spatial data analytics for building information models[J]. *Archives of Computational Methods in Engineering*. 2020, 27: 1449–1463.
- 8) J.R. Lin, J.P. Zhang, X.Y. Zhang#, **Z.Z. Hu***. Automating closed-loop structural safety management for bridge construction through multisource data integration[J]. *Advances in Engineering Software*. 2019, 128: 152-168.

(3) 与项目相关的知识产权

表 5-1 与项目有关的知识产权情况

序号	专利名称	类型	专利号
1	基于浓度分析的水体污染扩散模拟预测方法和装置	发明专利	ZL202111240610.5
2	基于粒子运动的水体污染扩散模拟预测方法和装置	发明专利	ZL202111232257.6
3	基于词典匹配的实体标注方法、模块及装置	发明专利	ZL202011079331.0
4	基于 IFC 标准的建筑信息模型数据集成与交换引擎装置和方法	发明专利	ZL201110023930.5
5	海洋环境信息建模系统 V1.0	软件著作权	2021SR0772916
6	建筑群风环境与污染物扩散模拟系统 V1.0	软件著作权	2021SR0772915
7	建筑及其机电系统动态标准化大数据提取软件 V1.0	软件著作权	2020SR0267715
8	基于 BIM 的大型公共建筑运维期安全管理平台 V1.0	软件著作权	2018SR729088
9	基于 BIM 的时变结构安全性能模拟和分析系统 V1.0	软件著作权	2017SR088117
10	基于 BIM 的建筑设备运维管理系统 V1.0	软件著作权	2017SR046923
11	基于 BIM 的建设工程文件归档管理系统 V1.0	软件著作权	2016SR123940
12	基于 BIM 的机电设备设施管理系统 V2.0	软件著作权	2016SR057011

续表 5-1

13	基于 BIM 的结构分析模型转换系统 V1.0	软件著作权	2013SR13876 2
14	基于 BIM 的建筑工厂化（设计、加工、仓储） 管理系统 V1.0	软件著作权	2013SR12164 2
15	基于 BIM 的机电设备智能管理系统 V1.0	软件著作权	2012SR11669 3
16	基于 BIM 的建筑工程 4D 施工安全与冲突分析 系统 V1.0	软件著作权	2011SR03105 9
17	基于建筑信息模型的工程项目 4D 施工动态管 理系统 V1.0	软件著作权	2010SR07039 0
18	基于 IFC 标准的建筑工程 4D 施工管理系统 V1.0	软件著作权	2006SRBJ2011

5.2 研究团队

表 5-2 团队人员基本情况表

序号	姓名	性别	出生日期	证件类型	证件号码	专业技术职称	职务	最高学位	专业	投入本专题的全时工作时间 (人月)	人员分类代码	在专题中分担的任务	是否有工资性收入	工作单位
1	胡振中	男	1983-11-28	身份证	440301198311281331	副高级	院长助理	博士	土木工程	24	课题负责人	专题三负责人	是	清华大学深圳国际研究生院海洋工程研究院
2	伍震	男	1991-06-09	身份证	431122199106090039	其他	研究助理	硕士	企业管理	30	其他研究人员	多源异构数据融合服务平台设计	是	清华大学深圳国际研究生院海洋工程研究院
3	袁静	女	1989-01-19	身份证	440301198901190923	其他	研究助理	学士	工程管理	30	其他研究人员	多源异构数据融合服务平台设计	是	清华大学深圳国际研究生院海洋工程研究院
4	安芃	女	1999-12-24	身份证	1201051	其他	无	学士	土木工	30	其他研究	多源异构数	否	清华大学深圳国际研究生院海洋

					9991224 2420				程		人员	据融合技术 与平台研发		工程研究院
5	刘毅	男	1999-04-14	身份证	3607211 9990414 721X	其他	无	学士	土木工程	30	其他研究 人员	多源异构数 据融合技术 研发	否	清华大学深圳国际研究生院海洋 工程研究院
6	李一林	男	1998-10-19	身份证	1308271 9981019 3638	其他	无	学士	土木水 利	30	其他研究 人员	多源异构数 据融合技术 研发	否	清华大学深圳国际研究生院海洋 工程研究院
7	刘龙祥	男	2000-06-18	身份证	3709232 0000618 4511	其他	无	学士	土木水 利	30	其他研究 人员	多源异构数 据融合技术 与平台研发	否	清华大学深圳国际研究生院海洋 工程研究院

6 考核指标与工作计划

表 6-1 子课题考核指标表

子课题	预期成果 类型	考核指标				考核方式（方 法）及评价手段
		指标 名称	立项时已有 指标值/状态	中期指标值/ 状态	完成时指标值/状 态	
面向多源 异构运维 数据的融 合模型与 管理方法	<input checked="" type="checkbox"/> 新理论 <input type="checkbox"/> 新原理 <input type="checkbox"/> 新产品 <input checked="" type="checkbox"/> 新技术 <input checked="" type="checkbox"/> 新方法 <input type="checkbox"/> 关键部件 <input type="checkbox"/> 数据库 <input type="checkbox"/> 软件 <input type="checkbox"/> 应用解决方案 <input type="checkbox"/> 实验装置/系统 <input type="checkbox"/> 临床指南/规范 <input type="checkbox"/> 工程工艺 <input type="checkbox"/> 标准 <input type="checkbox"/> 论文 <input type="checkbox"/> 发明专利 <input type="checkbox"/> 其他____	3.1 基于多源异构运维数据语义逻辑特征与数据标准化的异构融合模型与共享机制	无	a) 提出运维数据标准, 建立融合模型; b) 3 种以上异构运维数据的动态融合, 准确度大于 95%;	a) 针对不少于 5 类典型与市政公用设施, 实现 5 种以上异构运维数据的融合; b) 实现 5 种以上异构运维数据的动态融合, 数据融合准确度大于 95%	a) 发表学术论文 3 篇; b) 申请或授权软件著作权 2 项; c) 技术报告 1 项, 专家评审证明; d) 培养 1 名博士研究生、1 名硕士研究生
		3.2 基于数据动态匹配的“云-边-端”协同架构下多源异构运维数据动态融合管理技术	无	提出基于数据融合的智能运维数据服务方案	多源异构运维数据融合与共享平台	

表 6-2 子课题工作计划

计划完成时间(月)	研究任务	考核指标	成果形式
2023 年 1 月 至 2023 年 6 月	(1) 典型市政公用设施的运维数据标准化体系研究 (2) 多源异构运维数据融合模型构建	数据融合标准化体系	阶段性研究报告
2023 年 7 月 至 2023 年 12 月	(1) 静态多维度数据转换对齐方法研究 (2) 至少 3 类运维数据的融合方法与融合效果检验	3 类运维数据的融合和效果检验	申请专利 1 项；投稿或发表 1 篇论文；培养研究生 1 名
2024 年 1 月 至 2024 年 6 月	(1) 动态数据转换对齐方法研究 (2) “云-边-端”协同框架下的数据融合与加载方法研究 (3) “云-边-端”协同框架下至少 3 种运维数据的动态融合	3 类以上异构运维数据的动态融合，数据融合准确度大于 95%	投稿或发表 1 篇论文；专题中期报告 1 份
2024 年 7 月 至 2024 年 12 月	(1) 至少 5 类运维异构数据的融合方法研究 (2) 数据融合方法在 5 类典型基础设施工程数据集中的检验	至少 5 类运维数据的融合和效果检验，覆盖至少 5 类典型市政公用设施	开发完成相关模块
2025 年 1 月 至 2025 年 6 月	(1) “云-边-端”协同框架下至少 5 种运维数据的动态融合 (2) 数据动态融合方法在 5 类典型基础设施的工程数据集中的检验	至少 5 类运维数据的动态融合管理，覆盖至少 5 类典型市政公用设施，数据融合准确度大于 95%	投稿或发表论文 1 篇，培养研究生 1 名

计划完成时间(月)	研究任务	考核指标	成果形式
2025 年 7 月 至 2025 年 12 月	(1) 基于云计算平台的数据动态存储与管理技术、图形图像轻量化技术研发 (2) 搭建集成高效融合算法的后台计算服务系统	可支撑智能运维数据挖掘的运维数据融合服务	开发完成仿真基础模块； 申请或授权软件著作权 1 项； 投稿或发表论文 1 篇； 专题研究报告构成课题研究报告

参考文献

- [1] Pfeiffermann E, Braeuninger H W, Bihler E, et al. SPIE's International Symposium on Optical Science, Engineering, and Instrumentation[J]. 1999: 184–191.
- [2] Waltz E, Linas J. Multisensor Data Fusion. Boston: Artech House, 1990.
- [3] Hall L D. Mathematical Techniques in Multisensor Data Fusion. Norwood, MA: Artech House, 1992.
- [4] White F E. Data fusion lexicon[R]. JOINT DIRECTORS OF LABS WASHINGTON DC, 1991.
- [5] Steinberg A N, Bowman C L, White F E. Revisions to the JDL data fusion model[C]//Sensor Fusion: Architectures, Algorithms, and Applications III. International Society for Optics and Photonics, 1999, 3719: 430-442.
- [6] Llinas J, Bowman C, Rogova G, et al. Revisiting the JDL data fusion model II[R]. SPACE AND NAVAL WARFARE SYSTEMS COMMAND SAN DIEGO CA, 2014.
- [7] Bossé é, Roy J, Paradis S. Modeling and simulation in support of the design of a data fusion system[J]. Information Fusion, 2000, 1(2): 77-87.
- [8] 康耀红. 数据融合理论与应用[M]. 西安电子科技大学出版社, 1997.
- [9] 杨万海. 多传感器数据融合及其应用[M]. 西北电子科技大学出版社, 2004.
- [10] 韩崇昭. 多源信息融合[M]. 清华大学出版社, 2006.
- [11] 段建斌. 无线传感器网络和多源信息融合的瓦斯监测系统思路[J]. 山东煤炭科技, 2017(07): 69-70+73.
- [12] 化柏林, 李广建. 大数据环境下的多源融合型竞争情报研究[J]. 情报理论与实践, 038(4): 1-5.
- [13] 何友, 陆大, 彭应宁, 等. 带反馈信息的分布多传感器航迹融合算法[J]. 电子与信息学报, 2000, 22(5): 705-714.
- [14] 文成林, 吕冰, 葛泉波. 一种基于分步式滤波的数据融合算法[D]. , 2004.
- [15] 荣健, 乔文钊. 基于模糊神经系统的多传感器数据融合算法[J]. 电子科技大学学报, 2010, 39(3): 376-378+ 424.

- [16] 张雨浓, 杨逸文, 肖秀春, 等. 样条神经网络的权值直接确定法[J]. 系统工程与电子技术, 2010, 31(11): 2685-2687.
- [17] 王海颖. 多源数据关联与融合算法研究[D]. 江南大学, 2016.
- [18] Yao J T P, Brown C B, Fu K S, et al. NAFIP-1. Panel discussion on introduction of fuzzy sets to undergraduate engineering and science curricula[J]. International Journal of Man-Machine Studies, 1983, 19(1): 5–7.
- [19] Duan I , Stojanovic V , Trapp M , et al. Bridge damage: Detection, IFC-based semantic enrichment and visualization[J]. Automation in Construction, 2020, 112.
- [20] Huang P S , He X , Gao J , et al. Learning deep structured semantic models for web search using clickthrough data[C]// Proceedings of the 22nd ACM international conference on Conference on information & knowledge management. ACM, 2013.
- [21] 赵亮. 多模态数据融合算法研究[D]. 大连理工大学.