课题编号：

**重庆市人工智能技术创新重大主题专项**

**重点研发项目**

多源异构数据融合与知识发现水务决策平台应用示范

**实 施 方 案**

**课题实施方案编制组**

**二〇一八年一月**

目 录

[一、项目背景 4](#_Toc503974716)

[（一）立项的背景 4](#_Toc503974717)

[（二）项目拟解决问题 5](#_Toc503974718)

[（三）现有工作基础与优势 5](#_Toc503974719)

[1．现有工作基础 5](#_Toc503974720)

[2．项目研发优势 8](#_Toc503974721)

[（四）项目申请单位及参与单位研究基础 9](#_Toc503974722)

[1．项目申请单位—重庆水务集团 9](#_Toc503974723)

[2．项目参与单位—重庆邮电大学 10](#_Toc503974724)

[3．项目参与单位—重庆远通电子 12](#_Toc503974725)

[二、研究内容 13](#_Toc503974726)

[（一）主要研究内容 13](#_Toc503974727)

[（二）关键技术 14](#_Toc503974728)

[1．基于人工智能的水务领域决策支持系统平台 14](#_Toc503974729)

[2．多源异构数据知识特征提取与融合 14](#_Toc503974730)

[3．动态数据挖掘 15](#_Toc503974731)

[4．异构数据多属性决策支持知识发现 15](#_Toc503974732)

[5．基于迁移学习的规则优化 15](#_Toc503974733)

[三、技术路线 16](#_Toc503974734)

[（一）水务数据治理分析 16](#_Toc503974735)

[（二）基于最大间隔超平面的增强特征提取 18](#_Toc503974736)

[（三）面向关联聚类知识的的决策级融合 20](#_Toc503974737)

[（四）不可靠数据鲁棒置信规则分类 27](#_Toc503974738)

[（五）基于信息粒度模型的动态属性约减 37](#_Toc503974739)

[（六）异构数据多属性决策支持发现方法 40](#_Toc503974740)

[（七）基于语义解释的快速数据粒化 43](#_Toc503974741)

[（八）基于马尔科夫逻辑网的关联规则迁移学习 49](#_Toc503974742)

[（九）水务决策平台软件开发 54](#_Toc503974743)

[（十）水务决策平台应用示范 55](#_Toc503974744)

[四、考核指标 55](#_Toc503974745)

[五、实施方案 56](#_Toc503974746)

[（一）课题负责人 56](#_Toc503974747)

[（二）任务分工图 57](#_Toc503974748)

[（三）进度管理图 59](#_Toc503974749)

[（四）组织架构图 61](#_Toc503974750)

[六、管理制度 62](#_Toc503974751)

[（一）项目管理制度 62](#_Toc503974752)

[（二）经费管理办法 62](#_Toc503974753)

[七、保障措施 66](#_Toc503974754)

[（一）技术的风险分析与对策 66](#_Toc503974755)

[（二）市场风险与对策 67](#_Toc503974756)

[（三）管理风险分析与对策 68](#_Toc503974757)

[八、专家名单 69](#_Toc503974758)

**多源异构数据融合与知识发现水务决策平台**

**应用示范项目实施方案**

# 一、项目背景

## （一）立项的背景

随着云计算、大数据、物联网、移动应用、人工智能等新技术迅猛发展和深入应用，城市信息化发展正酝酿着重大变革和新的突破，由数字化城市向智慧化发展已成为必然趋势。在我国智慧城市已经上升到国家的经济、科技战略层面，水务管理是城市管理的重要组成部分，智慧水务是智慧城市建设的必然延伸。“十三五”期间，水务企业面临着适应经济新常态的重大挑战，面对“生态文明建设”、“国企改革”等政策推进力度增强、“云物移大智”技术应用日益广泛、提高运营决策效率需求更加急迫的发展环境，信息化必然成为水务企业转型发展的有力工具，水务信息化建设势在必行。

作为近年来的新蓝海，智慧水务已成为我国传统水务领域转型升级的必由之路。智慧水务是一个提出只有数年的新概念，其背景是“智慧地球”、“智慧城市”等概念提出之后，具体到水务工作的一次目标升华。简单说，智慧水务通过数采仪、无线网络、水质水压表等在线监测设备实时感知城市供排水系统的运行状态，并采用可视化的方式有机整合水务管理部门与供排水设施，形成“城市水务物联网”，并可将海量水务信息进行及时分析与处理，并做出相应的处理结果辅助决策建议，以更加精细和动态的方式管理水务系统的整个生产、管理和服务流程，从而达到“智慧”的状态。然而现有的“智慧水务”方面建设，受相关技术的影响，其效果并没有达到智慧的程度。

本项目结合大数据、人工智能等领域最前沿的理论技术，根据目前重庆市水务方面实际需求，响应国家人工智能2.0战略，构建研发本项目。本项目具备很好的技术领先优势，知识发现是人工智能领域实现强人工智能的基础，因此本项目无论是从实际产业应用方面，还是智慧水务建设的核心技术层面，均能很好体现项目自身的先进性。

（上一段话替换为：）本项目结合大数据、人工智能等领域最前沿的理论技术，根据目前重庆市水务方面实际需求，响应国家人工智能2.0战略，构建研发。本项目的核心理论之一是知识发现，而知识发现是实现强人工智能的基础，因此本项目无论是从智慧水务产业建设应用方面，还是水务行业人工智能核心自主知识产权研究方面，均能很好体现自身的先进性。

## （二）项目拟解决问题

1．知识发现和决策推理在大数据环境下，针对特定行业，需要新的更为有效的理论方法。

（上一段话替换为：）知识发现和决策推理在大数据环境下，针对水务行业的应用，需要新的更能体现先进性和智能化的理论算法。

2．传统方法在利用原决策属性约简结果和信息粒度降低算法计算复杂度方面，效果不理想。

（上一段话替换为：）传统方法利用原决策属性约简结果和信息粒度等相关方法，在降低计算复杂度方面，效果不够理想。

3．FRBCS语义模型存在不灵活性、在处理复杂的分类问题时，FRBCS的分类正确率常常会受到数据粒度和数据维度的限制；

4．目前的知识融合方法理论应用到具体行业领域时，往往难以达到预期效果。

5．传统的方法对自然数据集进行特征提取时，效率难以保证。

6．现有水务系统的海量数据尚未得到准确有效的应用。

（上一段话替换为：）现有水务系统的海量数据，在改善水务系统业务效率和准确性方面，应用效果不够明显。

## （三）现有工作基础与优势

### 1．现有工作基础

本项目基于生产SCADA调度中心、客户服务中心、营销系统、能效分析系统等，集合建设重庆水务集团多源异构全产业数据的处理和立体存储，构建全产业数据构建抉择平台，为企业提供全局性多维立体信息管理体系，同时为企业快速抉择、全局抉择提供多视角信息基础，提高应急事件抉择响应效率，提升平台的智能化和高效性。

以《全国水利信息化规划(金水工程规划)》及《全国水务局信息化规划》为原则。确定三个平台(网络平台、数据平台、应用平台)、三个层次(决策层、调度层、操作层)、三大行业(水利、供水、排水)为总体框架。在三大平台中，网络平台是载体，数据平台是基础，应用平台是核心。在三个层次中，决策层对应领导和办公室，调度层对应管理单位，操作层则对应市具体的水务工程、营业、服务、管网络等信息运营。

首先，建立具有统一标准、分层次交换和共享、具有数据更新维护机制的数据平台，包括建设标准体系、行业基础数据库和核心数据仓库库、数据更新维护机制等的大数据系统平台。信息釆集与交换平台是调度运行管理系统的基础，既包括各类信息从采集、传输到加工、存储和管理的各个环节，也包括自动采集、人工上报、外系统数据接入的各类信息；其次，数据资源管理平台的主要作用是满足海量数据的存储管理要求；整合系统资源，避免或减少重复建设，降低数据管理成本；整合数据资源，保证数据的完整性和一致性。通过数据的容灾备份，保证数据的安全性；数据资源管理平台主体是构建多维立体信息架构，主要由各类数据库、数据库管理系统及数据备份系统三部分组成；最后，应用支撑平台是连接数据中心和业务应用的桥梁，其作用是实现资源的有效共享和应用系统的互连互通，为应用系统的功能实现提供技术支持、多种服务及运行环境，是实现应用系统之间、应用系统与其他平台之间进行信息交换、传输、共享的核心。应用支撑平台主要包括：数据处理、权限、日志、数据査询、图形、报表、综合报警、综合展示、跨区同步等服务或组件。

重庆水务集团数据中心是公司建设重心之一，随着水务集团信息化建设发展需要，水务集团已投资1000万元建成水务集团数据中心机房和数据灾备机房，构建了具有云计算能力的硬件基础平台。在信息系统方面，集团先后投资建设了水务集团水厂自控系统、客户服务系统、营业管理系统、用户工程管理系统、管网GIS、DMA分区管理系统、生产报表管理系统、视频监控系统等，形成了水务集团业务大数据汇集的重要基础。

在信息化基础建设方面，水务集团的两大机房及主要办公区域已建立高可用双专线的网络结构；已设立了重要服务器区、协同办公区、财务核心；目前整体内部网络安全加固是 2015 年设计实施完成。并于2016 年底完成了集团与二级单位的直连线路升级改造（ SDH > MSTP）, 实现了与散布于重庆主城及各区县近50多家下级单位建立了生产视频专线、协同办公专线、财务及生产数据专线；由于集团业务管控的需求，对网络结构主要分为三大块（工业生产监控、协同办公和财务管）。各水务集团三级单位通过专线和无线等物联网技术实现了所辖管网、厂外加压泵站、居民二次供水站房以及生产厂区的信息联通。

重庆水务集团现有的数据中心和在信息化建设方面的努力，为本项目的顺利实施提供了优良的基础条件。

### 2．项目研发优势

目前水务行业全产业链数据异常庞大，包括自来水生产、管网供应、客户服务、营业销售及其管理；排水污水处理、环保监测、排水管网安全及管理；营销系统、管网GIS系统、客户服务系统、生产实时在线监控系统等多个子系统，各个子系统之间的数据在数据粒度、维度和结构等方面的差异，导致数据之间的统一分析、处理存在诸多困难。

针对上述多个子系统之间数据异构、多源的问题，由重庆水务集团承担、重庆邮电大学合作、重庆远通电子技术开发有限公司参与，充分发挥产学研模式的优势，整合现有行业资源，构建研发多源异构数据融合与知识发现水务决策平台。

本平台充分利用大数据技术特点，结合新一代信息化建设技术，和主流的互联网思维模式，通过基于顶层设计框架，借用大数据原理将各信息系统联接打通，消除“信息孤岛”和“信息荒岛”，让决策管理层可以及时便利了解整体业务运营情况，并对各已经稳定运行的业务系统进行新一代信息技术与业务需求的补充，逐步将人工智能技术切实运用到实际行业中。

重庆水务集团作为国内水处理行业的上市企业和龙头企业，信息化建设和应用成果对全国水行业企业可以起到指导和借鉴作用，对行业管理和技术标准建立具有重要意义。人工智能技术在重庆水务集团的率先应用，为全国水行业的信息化建设提供标准化的参考示范。

## （四）项目申请单位及参与单位研究基础

与本课题相关的课题实施申请单位及主要参与单位有：重庆水务集团股份有限公司、重庆邮电大学、重庆远通电子技术开发有限公司。

### 1．项目申请单位—重庆水务集团

重庆水务集团成立于2001 年，2010 年在上交所整体上市，目前是国内最大的、具有完整产业链、国有控股的专业水务上市公司，专注于城市供水和污水处理的投资、建设和运营管理，公司及公司的合营联营企业拥有重庆市主城区94％的自来水市场和全市94％的污水处理市场。截至2016年末，公司已投入运营的自来水制水系统42 套，日供水能力316 万立方米；污水处理厂65座，日污水处理能力285 万立方米。现有16 家全资及控股子公司和6 家合营、联营公司，业态涉及环境综合治理、基础设施咨询与建设等领域。公司在供排水领域具有投资建设运营一体化、供排水一体化、厂网一体化、服务保障一体化等竞争优势，形成了集约化、标准化、专业化的运营体系，已经成长为国内领先的的专业化水务投资及运营管理公司，并凭借丰富的行业经验和强大的运营实力，实现了全天候安全优质供排水，推动了社会经济和环境资源的可持续发展。

近年来，重庆水务集团主持了水专项“三峡库区城市污水处理厂功能提升与污泥处理处置技术研究与综合示范”课题，参与了水专项“山地丘陵城市供水安全保障共性技术研究与示范”等课题，获得的国拨经费2000 余万元。与重庆大学、同济大学、中科院生态中心、北京排水集团等单位建立了合作关系，加强了产学研结合，解决了企业存在的实际问题，形成了“三峡库区城市污水厂典型工艺功能提升和低成本改造”、“三峡库区城市污水厂污泥多元化处理处置”、“山地城市多级加压供水系统分区分级与安全调控”等关键技术，先后获得了重庆市国企贡献奖、重庆市科技进步奖，部分研究成果应用在库区污水处理厂、自来水厂。目前，集团总资产已超过200亿元，不仅具有解决配套资金的实力，而且具有较强的技术力量，可以整合资源，建立协调机制，有利于本课题工程示范项目的顺利开展。

### 2．项目参与单位—重庆邮电大学

重庆邮电大学是国家布点设立交重庆建设的几所邮电高校之一，是工业和信息化部与重庆市建的一所以信息科学技术为特色和优势，在邮电通信行业、信息产业领域，在重庆乃至全国具体重要影响和地位的高等院校。学校以发展邮电通信事业、振兴名族信息产为使命，是“中国数字通信的发源地”。学校立足信息行业，主动服务地方经济社会发展，不断探索产学研结合新模式，努力构建开放办学大平台。学校是国家TD-SCDMA产业联盟的唯一高校成员，与大唐电信、北大、清华、中科院计算所等单位联合组建新一代无线移动通信战略合作联盟，与中国科学院、中国社会科院、中国电子科技集团等共同组建院所科研基地，在与中国电信、中国移动、中国联通、化为、中兴等通信企业长期合作的同时，还与长安、四联、重钢等企业建立了校企合作基地，与重庆北部新区等共同组建了人才培养及成果转化基地，与重庆多个区县建立校地战略合作关系。学校还积极开展国际合作与交流，与新加坡、韩国、俄罗斯、欧盟等国家和地区的高校联合组建了多个合作教学科研机构，与加拿大、美国、英国、意大利等国高校开展联合人才培养项目。学校先后随办了“全球移动通信大会”、“工业无线通信国际大会”、光互联网技术国际会议”等有影响的国际会议。

目前，重庆邮电大学已与北京大学、总参五十四所、公安部十一局网络安全研究中心、公安部第三研究所、公安部网络安全重点实验室、国家863反病毒中心等国内知名院校和科研单位开展了长期项目技术合作，并已取得了较好的成果，为本项目研究内容的深度和广度提供了充分保障。我校与国内知名信息安全企业重庆爱思网安信息技术有限公司、北京北信源自动化技术有限公司等建立了良好的校企合作关系，为本项目科研成果转化，走“产、学、研、用”相结合的道路打好了基础。

重庆邮电大学实验室理论研究与工程实践相结合，承担国家科技支撑重大专项课题、国家自然科学基金、重庆市科技重点项目等20余项；获得国家科技奖励 1项，省部级奖 6 项，其中，2009年、2013年牵头获得重庆市科技进步一等奖，2010年牵头获得国家科技进步二等奖，2015年牵头获得重庆市科技进步二等奖；发表论文100余篇，其中包括在IEEE国际顶级期刊SCI/EI论文60余篇；申请发明专利 50余项，其中授权 20余项；出版著作 10余部，承办国际会议2次。以上基础为本项目研究内容的深化和提升创造了必要条件。

### 3．项目参与单位—重庆远通电子

重庆远通电子技术开发有限公司信息管理部门及其下辖专业从事水务领域自动化及信息化设计、研发和运维的团队，技术领先、经验丰富、敢于拼搏、具有创新精神，历经近20年发展，积累了丰富雄厚的技术研发能力，并围绕智慧水务领域的发展需求建立了较大规模的专业研发基地，构建起了包括科研、设计、生产、运维等全产业链的智慧水务服务体系，在城市供排水、节能环保行业的自动化领域处于国内领先地位。

重庆远通电子技术开发有限公司在自动化工程领域具有全方位的开发集成能力：熟悉多家世界顶尖自控产品的各种应用开发集成，具备MS .Net、Java、Oracle、SQL Server等软件平台的企业级开发能力，同时能提供对主流工业控制系统的工程集成；团队具有《电子工程承包资质》、《环境污染治理资质》、《安防系统建设资质》等多种专业资质；近年来，在水务智能装备的研发方面，获得一批有代表性的专利，在智慧水务体系及其应用系统的研究、开发与实践方面，也获得了一系列的软件著作权。团队完成了一批城市供水、城市排水、节能环保等领域内重点工程的设计与系统集成。典型工程在重庆、江苏、吉林、湖北、云南、四川、青海等地信息化建设。

目前，团队在工业大数据、机器学习技术方面持续加大资金与技术投入，并初步获得了一些技术，在智慧水务相关技术研发方面也取得了切实进步。深入研究了大数据技术、机器学习对智慧水务的支撑体系，搭建了原型体系平台，进行进一步的研究探索；实践并提炼了以大数据、机器学习为核心的数据科学在工业应用中快速有效应用的工程实践方法——敏捷数据科学方法，为大数据支撑的智慧水务体系的切实推进奠定了方法论基础。

# 二、研究内容

## （一）主要研究内容

1．构建包括水务部门日常综合业务、水务信息智能决策系统等在内的综合性业务示范平台，涵盖针对水务部门相关业务的运转、监管、维护，水务部门业务相关的决策支持，现有业务数据的智能分析，不同部门之间的信息共享，多部门数据的统一分析，智能挖掘等功能。

2．针对现有历史数据及系统日常运行所产生的海量数据，结合大数据和人工智能领域前沿技术，构建多源异构数据知识特征提取与融合方法，将多源异构数据，转化为统一知识表达语义描述，进而对上述知识表述进行特征提取，在此基础上完成知识特征的融合，以达到针对多源异构数据的归一化规则表述。

3．传统的机器学习假设训练数据与测试数据具有相同的分布，而水务系统业务数据的实际情况并不能满足这一假设，因此造成数据分析的效率大幅降低。人工智能技术能够有效地在相似的领域或任务之间进行信息的共享、迁移，使传统的从零开始的学习变成可积累的学习，具有智能化、成本低、效率高等优点，与本项目采用的动态数据挖掘能够很好的结合，提高系统数据分析效率的同时，对知识发现的深度提供一定的增益效果。

4．该项目采用自主研发和国际先进技术相结合的技术手段，研发多源异构数据融合与知识发现水务决策平台,依托重庆水务集团有限公司和重庆远通电子技术开发有限公司进行应用示范。

## （二）关键技术

### 1．基于人工智能的水务领域决策支持系统平台

针对水务领域相关的实际业务特点，充分利用大数据、人工智能技术和科学有效的物联网前端采集技术在数据采集上的优势，获取不同尺度上的多源异构数据，包括但不限于总体供水数据、城镇居民日常用水数据、水质监测数据、气象数据、环境数据、突发性水域数据等。进而，通过搭建决策支持系统，刻画水务领域在不同尺度上和不同时空的智能业务支持综合平台。

### 2．多源异构数据知识特征提取与融合

在大数据环境下，多源异构知识的融合为研究者从众多分散、异构的数据源和知识源中挖掘出隐含的、有价值的和尚未被发现的信息和知识提供了非常有效的手段和方法。针对目前知识融合方法的不足，在对大数据环境下的异构知识融合方法进行深入研究的基础上，将已有的数据融合算法合理地移植到知识融合中，设计并构造大数据环境下的多源异构知识融合算法。依据知识源粒度的动态选择，改进知识源分解与合并算法，以获得合适粒度大小的知识源集合和尽可能真实可靠的知识。

### 3．动态数据挖掘

构造一种基于信息粒度的动态属性约简模型，详细分析当决策表中出现新属性动态増加时，信息粒度的増量式计算方法；在此基础上，利用信息粒度作为启发信息设计动态求解属性约简算法，以期有效利用原决策表的属性约简结果和信息粒度，降低算法的计算复杂度，并使得约简结果具有较好的传承性。

### 4．异构数据多属性决策支持知识发现

针对异构数据，结合概念格理论，研究基于概念格的知识发现方法。具体地，研究异构形式背景及其概念格，通过异构形式背景定义异构决策形式背景，进一步在异构决策形式背景上讨论规则提取问题，研究挖掘非冗余决策规则算法，进而构造异构数据环境下，数据多属性决策支持的知识发现算法。

### 5．基于迁移学习的规则优化

针对共享知识是规则、结构和逻辑等关联规则的情况，研究基于马尔可夫逻辑网的关联规则迁移学习方法。首先利用伪对数似然函数，将源领域中马尔可夫逻辑网表示的知识迁移到目标领域中，建立两个领域之间的关联；通过对源领域进行自诊断、结构更新和目标领域搜索新子句，来优化映射得到的结构，进而适应目标领域的学习。

# 三、技术路线

## （一）水务数据治理分析

建立水务决策平台首要的任务就是数据分析，建立信息的交互需要规范的主数据，推动数据标准的建立，以确保数据的唯一性。主数据标准包括：数据编码标准化、数据项名称标准化、数据字典的建立、单据报表格式设计。具体的数据编码以业务需求和主数据特点为基础，选择适当的表现形式。编码结构规范如图1所示。

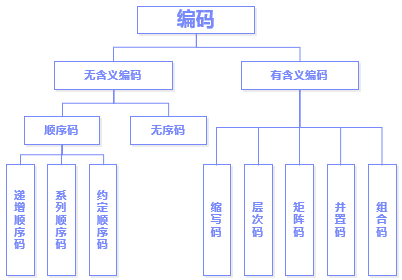


图1 主数据编码规范

未来建议主数据分别在各自的属主系统进行管理，系统间数据共享通过系统间的集成实现，未来逐步转变到混合模式。

主数据管理的技术实现上方式是通过统一的主数据管理平台，利用数据发布中间件将主数据从主数据管理平台发布到各业务系统中。技术平台选择后可以在其中对主数据进行统一管理，实现主数据的长效管理机制的支撑。管理功能包括：主数据管控、主数据稽核、版本管理。系统在数据采集、存储、共享和加工方面的具体建设内容如表1所示。

表1 系统在数据采集、存储、共享和加工方面的具体建设内容

| **领域** | **建设内容** |
| --- | --- |
| 采集 | * 数据可以分为结构化数据、非结构化数据、地理数据、实时数据； * 通过信息系统实现结构化数据（水质检验数据、项目数据、财务数据、人力资源数据、物资数据、设备台帐数据等）、部分非结构化数据的采集（如照片、视频、图档等）； * 通过物联网手段进行结构化数据（如物资盘点数据）、部分非结构化数据的自动采集（如通过高拍仪获取客户数据）、实时数据采集（如水质监控数据、智能水表数据）、地理数据采集（如通过GPS获取位置数据）。 |
| 存储 | * 通过Oracle数据库存储结构化数据、非结构化数据、地理数据、实时数据，新建系统不再使用SQL Server数据库，非结构化数据通过各应用系统进行采集和存储，非结构化数据可以互相调用和查询； * 非结构化数据量（如项目竣工图纸、施工图纸）增长到一定程度，可考虑采用专业的ECM（企业内容管理）工具（如FileNet）来进行存储和管理； * 实时数据量增长到一定程度，可考虑采用专业数据库（如PI数据库）进行存储和管理。 |
| 共享 | * 根据应用系统集成需求，通过应用系统进行数据的共享，如财务管理系统需要获取人力资源管理系统的薪酬数据； * 在数据共享之前，需要建立业务协同机制，明确相关的业务流程、边界、规则，在理顺业务协同机制后再开展数据共享； * 在数据共享之前，需要保证源头数据的准确性、完整性、及时性； * 借助于云计算的BaaS/DaaS理念，利用专业的集成平台进行数据的共享。 |
| 加工 | * 通过各应用系统进行数据的简单加工，如数据的分类统计和分析，如按组织、按时间（年、月、周、日）、按类型（供水、排水）； * 通过定期抽取各应用系统的数据建立数据仓库，进行数据的综合分析和加工，定义业务场景（如设备维护策略、项目后评估）进行专题分析； * 复杂业务领域（如调度）可引进专业模型（如水力模型）进行学习和训练，进行数据的深入加工； * 数据加工初期可利用专业BI平台（如BQ、Cognos）,在中后期可利用大数据平台（如Hadoop）提升数据加工的效率和敏捷性。 |

## （二）基于最大间隔超平面的增强特征提取

传统上，将特征提取分为线性特征提取技术与非线性特征提取技术。主成分分析(Principal Component Analysis, PCA)，线性辨别分析与因素分析等属于线性特征提取技术。松弛最大方差展开算法(Relaxed Maximum Variance Unfolding, RMVU)，核模糊主成分分析算法(Kernel Fuzzy PCA, KFPCA),自联想多层自动编码器(Autoassociative Multilayer Perceptron, AMLP)等属于非线性特征提取技术。为了对自然数据集进行更好的特征提取，我们拟通过对线性特征提取技术KPCA与MLP特征提取算法进行拓展，研究增强的特征提取方法：采用最大分割超平面，将样本映射到由一组间隔最大化且两两正交的超平面的法线所张成的子空间中，实现输入样本的特征提取。再用提取出的特征训练处具有差异性的基分类器。

设定,是训练数据，是输入样本)，是相应的类标签。假设是相互独立，同分布随机变量。通过求解下列优化问题(1)得到最大间隔超平面：

(1)

其中是要求解的超平面，是映射函数，是正则化参数，用来权衡误分类代价。

通过引入拉格朗日乘子与，并应用拉格朗乘子方法，可得到问题(1)的对偶式。

 (2)

对应于k的特征提取为：

 (3)

其中的第个元素是1，其余的元素为0。基于KKT条件，原始问题的最优解可由最优值与表示，即

 (4)

第s次迭代所获得的最优解。

 (5)

其中，，核函数,因此（3）式是的线性组合。基于（2）与（3），则可得到

 (6)

则仍是的线性组合，通过归纳可得出，算法的每次迭代，所求的的权值都是的线性组合。

给定问题（1）的参数，得到最优解与，则第一个提取的特征。转换样本数据，即将样本数据投影到同的正交空间中。为简单起见，假设已被规范化，即，则映射后的数据为

 (7)

用表示矩阵，应用（5）求解出的参数为的的解为，则第二个提取出的特征为,其中。

根据所需要提取的特征维数，重复上述特征提取的过程，直到提取出指定数量的特征为止。

## （三）面向关联聚类知识的的决策级融合

面向关联聚类知识的信息资源决策级融合目的在于以广义知识的内容属性和分类知识的结构属性为依据，根据资源属性特征把具有依赖或时序关系的知识对象关联在一起，进行知识聚类，将数据级融合生成的知识网络以及概念级融合生成的分类知识组织体系结构中具有潜在关联关系的知识资源以关联聚类知识形式呈现出来，为用户决策提供所需的知识。时序关联分析主要用于识别具有先后次序的不同资源之间的关联性，根据数据之间的逻辑关系，维度及语义等信息，可以发现彼此之间信息发布的先后顺序，追踪信息的来源等。

**时序（sequence）**：将承载相应信息资源的数据集记为，将对应的时间参数记为，则称为时序。

**时序关联（sequential correlation）**：若时序为，时序集合记为，对于任意时序和满足，那么时序DTi和DTj为时序关联关系，若时序DTi出现在时序DTj时间间隔之后，则记两者之间的时序关联关系为：

(8)

**序列（series）**：信息资源集合称为一个资源的序列，因此资源序列 S 由某些数据根据对应时间先后排列组成的序列，因此 S 可以记为下列表达式

(9)

**序列支持度（Support degree）**：对于某个时间序列库，某个时间序列 S 的序列支持度是指含有 S 的全部信息序列占有的比重：

(10)

其中|DT |表示信息资源库DT的信息序列数量。

**序列置信度（Sequence confidence）**：假设任意两个数据集序列分别为和，并且两者之间满足序列相互关联关系那么关联序列和的序列置信度可以表示为：

(11)

对于时间关联序列s和t，如果时间序列s出现，时间序列t以如下概率出现：  
 (12)

因此，对于数据集时间序列和，若其关联规则为，那么其序列置信度可以表达为：

(13)

式中表示数据集合的第i个时间序列，为数据集合 的第i个时间序列；代表数据资源集合的序列支持度，相应的序列置信度计算式表示为：

(14)

综上可知，以上所给出的数据集时间序列支持度、时间序列置信度、时间序列支持度向量、时间序列置信度向量等参量，满足如下关系式：

(15)

**频繁度向量（Frequency vector）**：假设数据集的时间序列在数据集合中发生频繁度为，那么相应的数据集时序关联相关频繁度向量表示为：

(16)

实现信息资源时序关联融合要解决以下两个关键问题：一是频繁度向量FV与信息资源频繁集合L的计算；二是依据信息资源集合L产生时序关联融合方法，通过FV产生向量集合SV和CV。其中，前者针对时序频繁度信息资源集合L和时序向量时序频繁度信息资源集合FV的运算属于时序关联融合的核心环节，其运算结果直接关系政府网站信息资源的时序关联融合效果。后者通常采用现有的频繁度向量生成函数获得，相对比较简便。

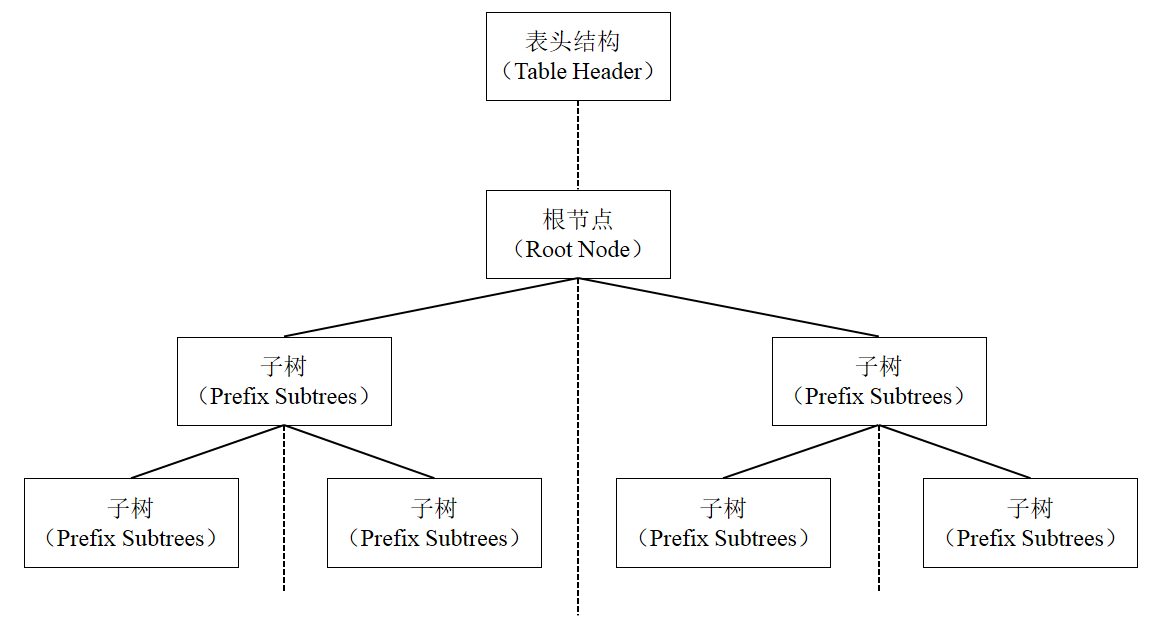


图2 时序关联融合算法IFPT树状结构

对于频繁度向量FV与信息资源频繁集合L的计算，目前主要有两种类型算法，一种算法在运行的过程中往往需要产生大量时序频繁度信息资源集合备选集，而且要求反复迭代扫描相关信息资源库，因此其运行效率低，时间和空间复杂度有待进一步优化。第二种类型算法在运行过程中可以产生时序序列支持度向量，该算法较第一种方法更为简单可行，但效率依然不高。为更好地提升算法运行效率，研究面向数据集时序关联融合算法，即 IFPT（Improved Frequent Pattern Tree）的时序关联融合算法，可实现对强密度大容量数据信息资源的时序关联融合。

算法 IFPT 树状结构包括1个时序频繁度信息资源集合表头结构（Table Header）、1个根节点（Root Node）、多个含前缀的子树（Prefix Subtrees）组成，相应的时序融合算法 IFPT 树状结构如图2所示。

IFPT2

图3 时序关联融合算法IFPT的流程

含前缀的子树节点包括结点地址（ADD）、时序支持度累加值（TSV: Timing Support Value）、时序频繁度信息资源集合名称（TF: Timing Frequency）和时序频繁度向量（FS: Frequency Sequence）4个子域组成。其中，节点地址ADD标识下一个同类型时序频繁度信息资源集合，可为NULL；时序支持度累加值为子树中含有该时序频繁度信息资源集合的全部时序序列集合数量；时序频繁度信息资源集合名词一般采用自定义形式；时序频繁度向量中包含的元素数量为n，任何元素是该子树上信息子集包含该时序频繁度信息集合数量。

与传统FPG等相关算法相比，IFPT算法采用了运算和保存时序频繁度向量FV及时序频繁度信息资源集合SFS的方法。

时序关联融合 IFPT 算法步骤如下：

输入：信息资源数据集合和数据子集

输出：时序频繁度信息资源集合L与相应时序频繁度向量 FV，s

方法：

IFPT-Construction，,

构建一个长为“0”的时序频繁度信息资源集合，设定其时序支持度为 TS 的资源事务数量，其时序频繁度向量为的资源事务数。

按顺序扫描，获得时序频繁度信息资源集合F、时序支持度累加值与其时序频繁度向量 FV，根据时序支持度累加值逆序排序，产生时序频繁度信息资源集合列表，记为L；构建IFPT树的根节点T，记为“空”。按顺序对中的所有信息资源时序序列依次运行；选择该信息资源时序序列中出现在L中的所有时序频繁度信息资源集合，将其依据L的顺序排序，记作 ，其中ps是第1个元素，PS为剩余元素。运行 add\_tr([], T, i)；运行函数add\_tr(, T, i)以下相应操作：若T有子节点N，同时 NS.ITname=ps.ITname。则 N 的时序支持度累加值提高 l，N 的时序频繁度矢量中的对应i 个元素fNi增加1，或者构建一个新节点N，同时设定其时序支持度为l，频繁度矢量中第i 个元素fNi是l，剩余元素是 0，同时其上一级节点连接指向 N，其节点连接指向同样名词的下一节点。若参数PS非空则循环调用add\_tr(, T, i)。

Process IFPT-Increase(Tr,α)

For each 的时序频繁度信息资源集合表头结构

do {

；

Output 频繁度模型,,；

；

构建的条件模型相关条件 ；

If 包含 single path PS then {

For each PATH PS 上节点集合 do {

Output 频繁度模式 ，

；

For each c (FV)中的元素 c (FV) do

c (FVi) min FV I of node in ；

}

}

else

if then {

(Lt, s, FV) IOFP-Increase(Tree,)；

tL←L∪L；

}

}

return L with corresponding and s

在 IFPT-Increase 方法中，实现函数 IFPT-built 和 IFPT-Increase 属于其核心任务，其时序频繁度矢量 FV 在时序频繁度信息资源集合产生过程中实现计算和保存，避免反复查询数据库。

## （四）不可靠数据鲁棒置信规则分类

在人工智能领域，模糊规则分类系统(Fuzzy Rule-Based Classification System, FRBCS)也是一种常用的分类器设计方法。作为一种基于计算智能的分类方法，FRBCS是基于易于用户理解的语义模型进行分类，因此其分类过程具有很好的直观性。然而，任何事物都有其两面性。由于语义模型的不灵活性，在处理复杂的分类问题时，FRBCS的分类正确率常常会受到一定的限制。例如，当输入输出映射关系比较复杂时，简单的语义逻辑变量很难对这种复杂的非线性映射关系进行精细的刻画。此外，由于模糊规则结构对于数据噪声缺少鲁棒性，在如战场目标识别等恶劣的工作环境中，常常会导致大量错误规则的产生，从而大大影响了其分类性能。实际上，在复杂的分类问题中，模糊不确定性往往与其它多种不确定性共存。由于模糊集理论能很好地处理模糊信息，因此研究模糊理论集与置信规则理论之间的关系，对复杂分类问题的性能提升具有重要意义。

为了处理在实际应用中常常存在的数据噪声，在置信规则库构建和置信推理方法设计中均进行了有针对性的处理。一方面，置信规则库中每个规则的结论部分是通过融合所有分配给其相对应的前提模糊区域的样本得到的，这一处理方式能有效地降低含噪声的数据对于规则结论部分的影响。另一方面，在置信推理方法中输入样本的分类结果是通过融合所有由该样本激活的规则的结论部分得到的，因此，即使在恶劣的工作环境下产生了一些不可靠的规则，通过这一步的融合处理可进一步降低误分类的风险。

考虑到置信函数理论在不确定信息建模和推理方面的优势，我们对传统的FRBCS在置信函数框架下进行扩展，提出一种置信规则分类系统(BRBCS)。如图4所示，所提出的BRBCS主要包括两大部分：置信规则库(BRB)和置信推理方法(BRM)。BRB构建了特征空间与类别空间之间的输入输出关系，BRM提供了基于构建的BRB对输入样本进行分类的推理机制。首先引入用于数据分类的置信规则结构；其次，基于这一置信规则结构，提出基于训练数据自动生成BRB的方法；最后，设计鲁棒的推理方法实现对输入样本的分类。



图4 置信规则分类系统（BRBCS）结构图

**结论置信分布：**对于一个复杂的分类问题，一个规则的结论部分可以用置信分布的形式描述，即将规则的结果指派给具有不同置信度的多个类别。对于一个M个类别的分类问题，每一个类别的置信度为，则规则的结论部分可以表示为如下的置信分布形式：。

**特特征权重：**在实际的分类问题中，不同的特征在决定分类结果时可能有很大的差异性。因此，需要给每一个特征指派一个权重用来表征这种重要性的差异。

将上面两个因素引入到模糊规则结构中，对于一个M个类别，P个特征的分类问题，可以得到如下用于数据分类的置信规则结构：

**置信规则:**

**如果是且…且是,则结果为**

**,规则权重为，属性权重为**

其中是第个规则分配给类别的置信度。在置信规则结构中，规则的结论部分可以是不完整的，即，剩余置信代表的是该规则对于结果的全局不确定度。规则权重描述了规则的可信度。特征权重描述了不同特征在决定分类结果时的重要性差异。

使用BRBCS处理分类问题的首要任务是基于训练数据构建描述特征空间与类别空间之间的输入输出关系的置信规则库。每一个置信规则由四部分组成，分别为前提部分，结论部分，规则权重以及特征权重。由于置信规则的前提部分与模糊规则的前提部分一样，因此可以基于同样的方法构建。

1．规则置信度生成

在置信规则结构中，结论部分是采用置信分布的形式描述的。在本节中，基于置信函数理论，提出如下的结论置信度生成方法。

与模糊规则中的结论部分生成方法类似，首先，需要计算每一个训练样本与不同前提部分的匹配度。在FRBCS中，直接指定具有最大匹配度的训练样本的类别作为该规则的结论部分。然而，在训练数据中存在噪声的情形下，这种结论部分生成方法具有很大的风险。在BRBCS中，我们对所有分配给该规则前提模糊区域的训练样本进行融合，从而生成具有置信分布的结论部分。

辨识框架是基于置信函数理论进行建模和推理的基础。对于该分类问题，类别集合可以看作辨识框架。令代表分配给前提模糊区域的训练样本集。对于任一训练样本，其类别可以看作是支持为相对应规则结论部分的一个证据。然而，考虑到训练样本的不确定性，这个证据并非具有完全的可信度。在置信函数框架下，这个证据可以解释为只将一部分置信（由匹配度表示）分配给。另外，由于可这个证据并没有指向其它任一类别，因此将剩余置信分配给辨识框架，用来表征全局不确定性。基于以上分析，可将这个证据用下面的mass函数来表示：

(17)

其中。

对于集合中所有的其它训练样本，基于同样的方式，可以构建一系列由其类别和匹配度决定的mass函数。为了获得前提部分所对应规则的结论部分，可将这些mass函数基于Dempster组合规则进行融合。如公式(17)所示，每个mass函数都是简单mass函数，即除了辨识框架外只包含一个焦元。针对这种特殊结构的msss函数的融合问题，Dempster组合规则的计算量可以极大降低，并可以写成如下的解析形式：

(18)

其中代表的是集合中所有类别为的样本构成的集合。总冲突值由下式进行计算：

(19)

因此，规则的结论部分中各类别的置信度为

(20)

其中表示未分配给任一单独类别的置信度。

在FRBCS中，规则的结论部分由与其前提部分具有最大匹配度的一个训练样本决定。而在BRBCS中，规则的结论部分融合了所有分配给该规则前提模糊区域的训练样本的类别信息，并且根据匹配度量化了不同训练样本在融合中的贡献值。因此，这种基于置信函数理论的规则结论生成方法能很好地降低数据噪声对分类结果的影响。

2．规则权重生成

在数据挖掘领域，人们经常使用可信度 (Confidence) 和支持度(Support)这两个度量对关联规则 (Association Rule)进行评估。置信规则Rq可以写成如下的关联规则形式： =>。与标准的关联规则相比，置信规则的特殊性在于其前提部分用模糊的逻辑表示，而结论部分用置信分布表示。

可信度是对关联规则的准确性的衡量。对于置信规则而言，其结论部分是通过融合所有来源于前提模糊区域所包含的训练样本的证据获得的。那么，如果这些证据之间具有较大的冲突（例如，这些证据分配最大的置信度给不同的类别），那么结论部分就具有较低的准确性。在置信函数框架下，有多种度量证据之间冲突的模型。在这里，我们采用简单方便的Dempster组合规则中的冲突因子来度量。基于上述分析，置信度规则Rq的可信度可以表示为

 (21)

其中为平均冲突因子，由下式计算：

 (22)

式中代表的是分配给前提模糊区域的训练样本数目。

与可信度相对应的支持度衡量的是关联规则的代表性。对于置信规则的生成而言，一共有N个训练样本可用。但对于某一规则而言，十几只用到了分配给其前提模糊区域的那一部分训练样本。因此，规则的支持度可以表示为

(23)

综上所述，可信度和支持度从两个不同的角度刻画了置信规则的权重，因此应该综合考虑这两个因素。一方面，如果规则Rq具有较高的可信度但是较低的支持度（例如，只有一个训练样本分配给前提模糊区域），规则的权重应该受到折扣，因为这种情况下规则的结论部分很容易受到数据噪声的影响；另一方面，如果规则Rq具有较高的支持度但是较低的可信度（例如，大量的训练样本分配给前提模糊区域，但是它们的类别具有很大分歧），规则的权重同样应该收到折扣，因为这种情况下前提模糊区域所对应的很可能是类别重叠区域。考虑到可信度和支持度均与置信规则的权重成正比例关系，这里我们采用二者的乘积运算来刻画置信规则的权重，即

 (24)

经过归一化处理，可以得到如下的置信度规则权重：

 (25)

3．特征权重生成

在置信规则结构中，特征权重反映了前提特征对于规则结论部分影响的相对程度。换句话说，具有较高权重的前提特征对于规则结论部分应该有更大的影响。

因此，特征权重可以用前提特征与结论部分之间的相关性来刻画。在本节中，我们基于相关性分析，提出如下的特征权重生成方法。

假设对于特征有个模糊划分。基于这 个模糊划分，将前面构造的具有个置信规则的规则库划分为个子规则库 ，，其中每个子规则库 由前提特征取值为模糊划分 的所有置信规则组成，即

 （26）

然后，对于每个子规则库，应用加权平均算子对其包含的所有规则的结论部分进行融合：

 （27）

其中和为由公式

 （28）

计算所得的规则的结论置信度；为由公式

 （29）

计算所得的规则的权重。

我们基于前提特征取不同值时相对应的结论部分的变化量来表征二者间的相关度(Correlation Factor)。当的取值从变为，相应结论部分的变化为：

 （30）

我们将定义为前提特征与结论部分之间的相关度，即。

基于同样的方法，我们可以得到所有个特征与结论部分之间的相关度

，。最终，特征 的权重可以由归一化的相关度来表示，即

 （31）

4．置信度推理方法

在FRBCS中，对于输入样本的分类采用的是单优胜((Single-Winner)推理方法。然而，当训练样本中含有大量噪声时，有可能产生一些不可靠的规则，这时采用单优胜推理方法会有很大的风险。在本节中，我们提出一种鲁棒的置信推理方法((BRM)，其主要思想是首先计算输入样本与被该样本激活的规则之间的关联度，然后将关联度看作证据的可靠性，对所有激活规则的结论部分在置信框架下进行融合。

**关联度计算**

令为待分类的输入样本。首先计算该输入样本与构建的置信规则库中每一个规则的匹配度：

 (32)

其中为前提模糊集对应的隶属度函数，为第个特征的权重。

令为置信规则库中个规则的集合，令表示所有被输入样本激活的规则的集合，即：

 (33)

被激活规则的结论部分对于输入样本分类的有效性由两个因素决定:匹配度和规则权重。其中，匹配度反映的是输入样本落入该规则的前提模糊区域的可能性;而规则权重则刻画了该规则结论部分的可靠性。我们将匹配度和规则权重的联合作用称为关联度(Association Degree)，计算如下:

 (34)

**置信框架下推理**

在上一节中，我们给出了输入样本与被激活规则的结论部分之间的关联度计算方法。本质上，关联度表征的是被激活规则的结论部分对于输入样本分类的有效性。因此，在对所有被激活规则的结论部分进行融合时需要考虑关联度的差异。在置信框架下，通常使用Shafer折扣运算对不完全可靠的证据进行处理。将(34)计算所得的关联度看作证据的可靠度，对各被激活规则的结论部分可进行如下折扣处理:

 (35)

 (36)

基于所有个被激活规则，应用上面公式可以得到个mass函数，。

为了得到输入样本y的分类结果，需要将上面经过折扣处理的L个mass函数使用Dempster规则进行融合。考虑到待组合的每个mass函数的焦元均为单元素集合(辨识框架几除外)，Dempster组合规则计算复杂度可以降为线性复杂度，并可以写成如下的递归形式:

(37)



 (38)

 (39)

其中代表前个mass函数的融合结果。一递归的初始值设置如下: ，。当递归标志达到时，便得到了最终的融合结果，以及。

基于上面融合后的mass函数，可以使用最大pignistic概率规则做出最终决策，即输入样本被指派为具有最大pignistic概率的类别。

## （五）基于信息粒度模型的动态属性约减

属性约简是数据挖掘、知识发现和计算智能等领域中一种重要的数据预处理方法，属性约简可对高维属性数据进行属性降维。即通过消除数据中的无关和冗余属性，可以剔除噪音数据并压缩数据的规模，改善数据分类的性能，其主要思想是保持数据分类能力不变的条件下，达到删除不相关和冗余属性的目的，主要优势在于它不依赖于先验知识，只依赖于数据本身。

决策表的属性往往是动态变化的，即系统中存在新属性的增加或原属性的删除。当在决策表中增加一个属性集P，首先根据信息粒度计算方法，需首先计算其对应的，并计算其对应的信息粒度。然后可知决策表中的核属性是必不可少的，为此需计算属性集中有无核属性存在，如核属性集，则需将其加入属性约简中，计算对应的划分

(40)



易知，若此时的信息粒度并不满足属性集C下的粒度度量，则需分别计算其余属性在核属性加入的前提下的信息粒度，即对于，计算其决策表上对应的划分，由此可计算其对应的信息粒度；直到新加入的属性满足属性约简定义。

决策表，对于，当时，则条件属性为核属性，同时记决策表S的核属性集为。

决策表，若为决策表的属性约简。则满足：

1．；

2．对于，使得。

通过上述增量式计算方法，我们可有效利用在条件属性C下对样本集的划分和在原属性约简下对样本集的划分，无需重新求解其在决策表下的划分，有效节省了计算时间，提高了算法的计算效率。面向决策表的动态属性约简算法的设计思想：首先，计算在新属性集P在原和的划分计算出新的划分，由此计算决策表正区域；然后，计算属性集下新的信息粒度，根据信息粒度中求解出新对象集P中核属性，并利用核属性对原的信息粒度进行动态更新；最后，通过添加重要的属性到属性约简中，直到满足信息粒度下的属性约简结果。具体的算法描述如下:

算法：基于信息粒度的动态属性约简算法（Dynamic attribute reduction algorithm based on information granularity算法ARIG）

输入：决策表，原属性约简，增加的新属性集P。原决策表的划分和；

输出：新的属性约简。

步骤1. 初始化属性集P中的核属性集，新的属性约简；

步骤2. 对于属性集P，利用计数排序方法求解下的条件等价类；

步骤3. 计算决策表的正区域，负区域；

步骤4. 计算决策表在下的信息粒度；

步骤5. 对于，分别计算信息粒度；如果，则新增的属性a为核属性，即；

步骤6. 对于新的核属性，计算其在决策表上的等价类：

并计算属性集的信息粒度；



步骤7. 如果。则，输出新的属性约简，算法结束；否则跳转至步骤8；

步骤8. 对于，一次计算其的等价类

，由其计算其对应的信息粒度为；然后根据条件属性的重要度量，选择信息粒度的最大值所对应的属性存入中，即；

步骤9. 如果信息粒度，则输出新的属性约简，算法结束；否则跳转至步骤8。

## （六）异构数据多属性决策支持发现方法

决策形式背景是通过一对形式背景构造的，这样做的主要目的是通过在形式背景上引入决策属性来进行具体的决策分析。截至目前，决策形式背景的研究已取得初步成果，但是其也面临着一系列的挑战。比如，在应对大数据环境时缺乏可行性与有效性，因为现有的方法都是假设数据是同构的，即属性取值构成的偏序两两相同，然而大数据环境下的数据往往是异构的，即属性取值构成的偏序不尽相同，甚至完全不同。在这种情况下，现有的处理方法将失效，无法给决策形式背景提供决策分析支持。

受此启发，基于概念格来研究异构决策形式背景的知识来发现问题。首先，定义异构形式背景，并在此基础上提出异构决策形式背景；然后，分析异构形式背景的概念格构造，并讨论异构决策形式背景的规则提取问题；最后，给出挖掘非冗余决策规则的有效算法。

现实应用中遇到的形式背景的属性取值形成的序不尽相同，甚至完全不同，这类数据集称为异构形式背景，其定义描述如下：

一个异构形式背景可以表示成四元组，其中是对象集，是属性集，中属性的取值构成的序两两不同，是对象在属性下所有取值组成的集合，是上的关系，表示对象在属性下取值为。

一个异构决策形式背景可以表示为七元组，其中和为异构形式背景且通常地和分别为异构决策形式背景的条件属性集和决策属性集。

设四元组为异构形式背景，在对象集和向量上定义如下运算：

 （41）

 （42）

给定两个偏序集和，称：和g： 是偏序集之间的伽罗瓦连接，如果它们满足以下性质：

1）；

2）；

3）。

设为异构形式背景，，，若且，则称为异构形式背景的概念，其中为概念的外延，为概念的内涵。

其异构形式背景的概念格构造算法如下：

算法１：异构形式背景的概念格构造算法（给出构造的步骤）

输入：一个异构形式背景

输出：形式背景的所有概念

Step1 初始化：；

Step2 ；

Step3 从中任选一个元素；

Step4 如果，转Step6，否则转Step5；

Step5 如果对于任意都不成立，则；

Step6 如果，则，并返回Step3；

Step7 如果对于任意都不成立，则；

Step8 如果，则，并返回Step2；

Step9 如果则；

Step10 输出。

异构决策形式背景的规则提取

设为异构决策形式背景的子背景，则在中是非冗余的，当且仅当且，

其中表示映射定义在异构形式背景上。

算法2：异构决策形式背景的非冗余决策规则提取算法

输入：一个异构决策形式背景

输出：的所有非冗余规则

Step1 初始化：；

Step2 调用算法１构造异构形式背景和的概念格，分别记为；

Step3 对于中的每个元素，如果且，则；

Step4 输出。

由于算法2调用了算法1，因此其时间复杂度也是指数级的。

## （七）基于语义解释的快速数据粒化

应用粒计算技术将数据粒化可以提高后续分析与处理的计算效率。如果过多地注重数据的细节会降低计算效率，甚至抓不住关键问题。故此采用粒计算将数据分解成若干个部分，即信息粒，使得信息粒的数量为传统方法可以处理的规模是十分必要的。由于粒化程度不同，会将数据转为不同的粒层，各个粒层的信息粒具有不同的粒度。粒层的确定直接影响后续分析与处理的效率与性能，故此粒计算研究的关键为如何设计粒计算模型以及如何粒化。

粒计算己经成功应用于以下几个方面：1)应用于不确定性数据，通过分析粒和粒之间关系，从不同粒度层次上处理问题，分析数据的内部结构；2)应用于海量数据和大数据，利用粒计算方法选择合适的粒度，降低求解问题的复杂度，从而得到较好的解决方案，提高处理效率；3)粒计算模型将需要处理的各种数据转化为知识粒的形式，通过粒计算不仅可以将数学性质相似的数据转化为粒的形式，并且可以发现紧凑的易于理解的概念，是知识发现的一种有效途径.4)粒计算已经成为分析和设计智能系统的有力工具。

然而目前信息粒化仍然面临着很多挑战，其中包括：面向海量数据或者大数据如何实现快速粒化，如何形成具有语义解释的信息粒。本文提出的快速数据粒化方法就是基于语义解释基础上的，并且易于对该模型进行改进以处理复杂问题。

首先，为了控制算法的复杂度和语义解释的复杂程度，对属性进行筛选。其次，对每个样本进行描述，即赋予每个样本一个易于理解的语义解释，这个语义解释对应于模糊规则前件，模糊规则后件为标号。再次，典范样本描述被挑选出来作为某个信息粒的描述，也就是某一数据粒的语义解释。最后，应用这些挑选出的典范样本描述对所有样本进行标号完成数据粒化。

采用符号表示待粒化数据，假定大小为的矩阵。设表示的第个属性(变量)，对应于的第列。的第行对应于样本，表示样本在第个属性下的值。

1．无监督的属性选择

Ferreira和Figueiredo于2012年提出了一种无监督的属性选择方法RFS，该方法的假设条件是:方差越大的属性就包含越多的信息，属性也就越重要，相反方差越小的属性包含的信息越少，属性也就越次要。

RFS方法的主要思想为:给定所选择属性方差的累积和占总方差累积和的比重，选择尽可能少的属性。具体步骤为:首先将每个属性离散化，设属性被离散为个可能值。其次，计算每个属性的方差,求出，并且将按照值大小进行排序得到,使得。最后选择合适的使得:

 (43)

这里为参数，，用来控制所选择属性个数。由于无监督的属性选择方法RFS需要对数据离散化，就存在离散化算法常常面临的两个问题。一方面将每个属性给出最优的离散化需要考虑到优化问题，比如离散值的个数问题，间断点的选取问题等。另一方面，数据离散化之后无疑会丢失信息，影响属性选择。故此对无监督的属性选择方法RFS进行改进，将数据的离散化步骤去掉，也就是说，的计算方法更改为称改进的方法为mRFS，算法的伪代码见表2。

表2 改进的RFS算法（mRFS）

|  |
| --- |
| Aigorithm:改进的RFS算法 |
| **Input:** |
| X: 数据集; |
| L: ，参数L用于控制挑选的属性个数。 |
| **Output:** |
| : 得到的属性约减的数据集; |
| FeatKeep: 所挑选属性的标号。 |
| **Begin:** |
| 1. 计算每个属性的方差，; |
| 2. 根据方差值的大小进行降序排列，得到; |
| 3. 计算使得: ; |
| 4. 将前个方差值大的属性标号放入数组FeatKeep中; |
| 5. 通过数组FeatKeep构建，仅保留个方差值大的属性; |
| 6. 返回，FeatKeep。 |

2．模糊数

本文采用模糊数构造样本的模糊描述，与第二章模糊数不同的是，参数的确定方式不同。最主要的区别在于，大多数针对的是分类问题，也就说可以利用类标号，然而数据粒化却没有类标号可以采用，即无监督的模糊数。

针对的第个属性，为了简便依然用表示挑选出来的属性个数。属性为上的个参数确定方法采用EIB方法，具体如下:

首先，将数据的第列按照值的大小由小到大进行排列，并且设最小值为，最大值为。

其次，将属性的样本空间进行划分，划分为区间长度相等的个区间，区间长度为。

最后，计算，这里。

个参数定义为区间内样本均值这里，是指数据X的第列。这样参数的取值由原始数据和参数确定，能够有效地减少参数个数。

3．样本描述

模糊IF-THEN规则是一种很直观且易于理解的知识形式，这里采用模糊IF-THEN规则的前件作为样本的模糊描述，规则的后件为标号。假定每个属性上都定义了个模糊数，则样本空间被划分为个模糊子空间。每个样本的模糊描述也可以认为是该样本所在的模糊子空间，具体定义方式如下:

令是数据集上的模糊数集合，样本的模糊描述定义为:

 (44)

对于每个模糊数集合，表示将集合A中所有模糊数进行模糊逻辑合取运算，即“and”运算。样本的模糊描述对应于一个模糊子空间，也对应于一个模糊集，可以根据实际问题调整样本模糊描述的定义方式，使得算法更容易根据应用领域的需要进行修改。对于样本，表示属于模糊描述包含模糊数的平均隶属度:

 (45)

样本描述对应于模糊规则的前件，后件为数据粒标号。

对于两个样本，假设样本描述分别为和，他们之间的相似性定义为:

 (46)

两个样本描述和之间的相似性在区间[0,1]上取值，相反的两个样本描述和之间的反相似性为:

 (47)

显然，对于任意给定的两个样本描述和之间的反相似性也在区间[0,1]上取值，并且容易验证满足:







4．模糊概念的评价标准

若每个属性上都生成K模糊数，则样本空间被划分为模糊子空间，设表示全体模糊子空间的集合，即。表示集合的幂集(集合的代数)，，为若干个模糊子空间的集合，表示一个模糊概念。

下面给出模糊概念的评价标准，，模糊概念的评价标准由以下公式给出：

 (48)

这里，表示所包含模糊子空间的个数，表示样本集包含的样本个数，表示在空间下的余集，和的计算由以下公式给出:

 (49)

 (50)

这里为样本描述，表示模糊数的平均隶属度。

## （八）基于马尔科夫逻辑网的关联规则迁移学习

通过对从源领域中得到的与当前目标领域相关的源领域模型的迁移来完成精确有效的结构学习，将迁移视为优化结构，通过算法来诊断源MLN(Markov Logic Network)结构中不正确部分并加以优化，提高学习的精确性。从源领域到目标领域迁移MLN结构的算法主要分为两个步骤：首先，将源领域中的MLN结构与目标领域进行映射，建立两个领域之间的关联；然后，对映射得到的结构进行优化以适应目标领域的指标。

**结构映射**

结构映射采用类型一致的映射概念，源领域到目标领域的映射表示子句中源子句和目标子句之间的关系，这种关系定义了两个谓词间的映射类型。如果映射是类型一致的，那么在一个子句中，源领域的一个类型映射到目标领域至多是一个类型，而且此目标类型必须由子句中所有谓词组成的源类型映射构成。如果源领域中每个给定子句中的谓词都能映射到兼容的目标谓词或空谓词，那么这个映射就是合理的。所谓空谓词，就是从映射合理的子句中清除所有文字的源谓词。如果两个谓词拥有相同的参数个数，且参数的类型能和当前约束兼容，则它们兼容。例如，若当前的类型约束条件为空，源谓词Publication(title，person)如果要和目标谓词Gender(person，gend)兼容，就必须加入类型约束title->person和person->gend到当前的类型约束，并且之后所有子句中的谓词映射都必须遵守这两个约束。在这个合理的映射关系建立以后，可以通过计算MLN的WPLL(Weighted Pseudo Log-likelihood)来估计迁移谓词，这时MLN仅由这些迁移子句组成。结构映射独立地映射每个源子句，通过利用计算目标数据的WPLL分数估计所有可能的类型一致映射，不断重复这个过程，就可以得到每个源子句的最优谓词映射，也就是WPLL值最大的迁移子句。

**结构优化**

1．自诊断

自诊断的目标是：利用迁移的MLN在目标领域中进行推断，搜寻不成立的子句并查找原因，通过Gibbs采样来完成自诊断。具体而言，就是依次检验目标领域中的每个谓词，将正在测试的当前谓词标记为P\* 。为了完成Gibbs采样，假设P\*是由一系列未知的谓词组成，而用其余所有的闭谓词进行验证。在每一轮采样中，将闭文字X重采样的值加入P\*中，算法考虑所有计算X所需的闭子句集合。对于每个闭子句，将与P\*有关的文字视为结论。表3列出了将目标谓词作为结论，将子句表达为蕴含式的形式。

表3 目标谓词Worker的蕴含式

|  |  |
| --- | --- |
| 子句形式 | 目标谓词Worker的蕴含式 |
| Manager(A) Worker(A)  Proposal(P,A) Worker(A)  WorkedFor(B,A)Worker(A)Worker(A) Proposal(P,A) WorkedFor(A,B) | Manager(A)=>Worker(A)  Proposal(P,A)  => Worker(A)  WorkedFor(B,A)=> Worker(A)  Proposal(P,A) WorkedFor(A,B)  =>Worker(A) |

根据子句相关与否，好坏与否的所有可能的方式，定义X可能的4种情形，C属于这4种情形之一。这些情形考虑条件全部满足，结论都正确的所有可能情况。如果条件都满足，将这些子句标记为相关的，否则为不相关。当且仅当结论是正确的时，表及相关子句为好的；当且仅当结论为错误时，标记不相关子句为好的。假设X =Worker(banny)时，v=ture。

① [相关，好]情形

这种情形下，所有子句条件都满足，结论都正确。例如，C为表3中第一个子句的闭子句加入常量banny，表示为：Manager(banny)=>Worker(banny)。换句话说，仅当X为真实值v时，子句条件能满足，这个值在数据中，这种情形下的子句保持正确。

② [相关，坏]情形

这种情况下，所有子句条件都满足，但是结论不正确。例如，C为Proposal(project1,banny)=> Worker(banny)，相应的子句形式Proposal(Project1,banny) Worker(banny)。换言之，如果X的值为v，数据中的真实值为否定，这个条件才能满足。

③ [不相关，好]情形

这种情况下，所有子句条件不能都满足，因此子句不被激励，即使受到激励，结论也不正确。例如，C表示WorkedFor(jemmy,banny)

=>Worker(banny),相应的子句形式为WorkedFor(jimmy,banny) Worker(banny)。因此，这种情形的子句不论数据中X值是多少，只要X的值为v，则C中和X相关的文字是正确的。

④ [不相关，坏]情形

这种情况下，所有子句条件不能都满足，但是子句一旦被激励，结论正确。

需要指出的是们虽然前面的例子中使用的子句仅包含标记为P\*中的一个文字，但是算法同样适用于含有多文字的子句。这种分类法可以表示为如下公式：

 （51）

仅仅当子句属于[相关,好]这种情形时，X=x的概率才会增加，当子句在[相关,坏]这个情形中，X=x的概率降低，而在另外两种情形中的子句对这个公式没有影响。为了解释这一问题，定义一个与X有关的闭子句集合中的子句girr∈Gx。不论X的真实值是多少，这个girr都能满足条件；又因为不管X的取值为多少，girr(X=x,MBx=m)的取值都为1。式（51）中的Sx（x,m）可被重写成：

 （52）

根据以上推导，可以将式（50）写为：

（53）

由此可以看出，girr和ewirr在计算和定义中可以忽略。然而，如果[不相关，坏]情形中子句除X以外的其他文字被删除，X的值将变得中重要，将会移动至[相关，好]的情形中。类似的，如果在[相关，坏]子句中增加文字，将会忽视X的值，并移动至[不相关，好]的情形中，并且X有正确值的概率不再降低。

2．结构更新

根据自诊断的结果，将太长的子句缩短，并将太短的子句加长。采用过滤束搜索（beam搜索）对子句的集合进行优化。在每一步迭代中，文字增加或减少都在当前的集合中完成，选取其中最优的n个子句，再进行新的迭代。不考虑每一子句中所有文字的加长和缩短，而是仅仅从标记了缩短的子句中移除文字，在标记了加长的子句中添加文字，并利用WPLL进行打分。因此，可以通过限制用于更新的子句数量和限制每个子句的更新类型，来约束搜索空间。

3．搜索新子句

前面两个步骤都是对于源领域迁移学习得到的子句进行更新优化，而无法去挖掘目标领域中特定关系的子句。为了解决这个问题，利用关系寻径（Relational Pathfinding，RPF）方法来搜索目标领域中的新子句。RPF是一种数据驱动方法，用来设计搜索路径以克服局部极值点，考虑将其用于优化阶段，以发现源领域与目标领域之间的特定关系。RPF将关系领域视为图G，常量用节点表示，其任意两个常量的长度至少是2，由边连接。RPF在G中搜索路径，如果找到这个路径，就可以写成闭子句形式。前件是由标记路径上每个边的谓词连接在一起组成的，标记连接两个常量的边的文字称为结论。利用爬山搜索法，尽可能在前件中增加谓词来提高子句性能。

## （九）水务决策平台软件开发

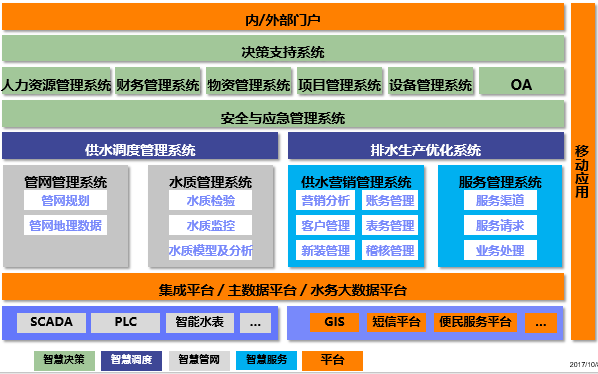
水务决策平台应用架构第一步是完成业务的数字化转型，构建集团统一的应用系统。根据架构设计原则，从界面集成、流程集成、应用集成和数据集成四个层次建立一体化应用平台的集成架构。

图5 水务领域综合业务平台架构

界面集成：通过内外部门户的统一认证、统一鉴权、统一审计和统一界面风格，实现不同应用系统对用户的透明化。

流程集成：通过业务流程引擎支持多部门、跨系统的业务流程，支撑流程化工作模式，提高内部工作效率。流程集成技术架构如图6。

应用集成：通过ESB平台使用标准的服务编排技术整合各系统应用资源，并实现应用系统间的集成。

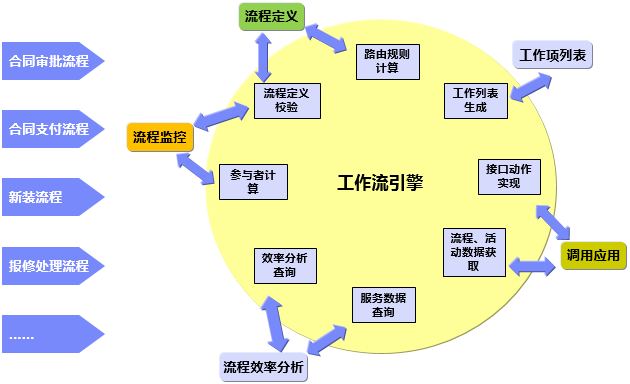


图6 流程集成技术架构

## （十）水务决策平台应用示范

1．应用示范点一：重庆水务集团股份有限公司

2．应用示范点二：重庆远通电子技术开发有限公司（重庆市自来水有限公司）

# 四、考核指标

（一）高质量学术论文4篇。

（二）通过多源异构数据知识特征提取与融合技术、动态数据挖掘等关键技术，研发集视频安全、水质监控、生产管理、应急处理、客户服务等功能为一体的水务决策平台。

（三）完成2个示范工程。

（四）申请相应发明专利6项。

（五）取得软件著作权1项。

# 五、实施方案

## （一）课题负责人

课题负责人庞子山，男，1973年12月出生，毕业于重庆大学城环学院，市政工程专业博士研究生学历，博士学位，正高级（教授级）工程师，1994年12月加入中国共产党。主要工作经历为：1996.07至1997.09在重庆市环境保护局监理总队工作。2003.08至2010.12在重庆市三峡水务有限责任公司工作，历任总工程师、副总经理、总经理。2010.12至2017.07，任重庆渝西水务有限公司执行董事、总经理（法定代表人）、党总支书记，其间：2012年至2017年任重庆市第四次党代会代表、重庆市永川区第十二届人大代表。2017.07至今，任重庆水务集团股份有限公司副总经理、党委委员，兼重庆渝西水务有限公司执行董事、总经理（法定代表人）、党总支书记，期间，2017年至今当选为第五次党代会代表。

该同志于1998年开始从事科研工作，参与了多项污水处理规范工程、解决水污染问题控制目标远景规划、污水处理市场化研究及水环境保护水资源综合利用等课题项目。从事市政工程污水处理工作以来，主持参加科研项目4个，发表期刊论文22篇，作为第一发明人申请实用新型专利13项、发明专利4项。凭借出色的表现，于2006年度被评为“十五”全国建设科技进步先进个人、2009年度被评为重庆“国企贡献奖——先进个人”，2013年度荣获了重庆市“五一”劳动奖章、2017年荣获重庆市“劳动模范”称号。

## （二）任务分工图

针对本项目的最终所要达成的预期目标，宏观层面将其分解为项目实施和核心技术研发两部分，其中项目承担单位负责项目的实施部分，具体包括：基础平台建设、业务逻辑分析、设备购置与维护、系统运行环境搭建、系统部署及测试、示范应用建设、后期运维及市场相关的系列工作；项目合作单位负责项目的核心技术研发部分，具体包括：决策平台研发，多源异构数据融合与知识特征提取方法、知识发现与推理方法、决策评估方法、成果总结、论文专利撰写、核心技术相关的文档材料编写以及相关代码编写、调试等工作，具体分工如图8所示。



图8 任务分工

## （三）进度管理图



## （四）组织架构图

多源异构数据融合与知识发现水务决策平台应用示范

参与单位

参与单位

申报单位

重庆水务集团

重庆邮电大学

重庆远通电子

课题负责人

秘书小组

示范小组

开发小组

研发小组

数据采集汇集

系统平台搭建

系统应用示范

系统分析研究

软件开发测试

制订

研发计划

项目日常管理

财务过程管理

协调会务管理

制订

研发计划

课题分析

研发

技术交流探讨

图9 课题内部组织架构图

# 六、管理制度

## （一）项目管理制度

1．甲方负责协调示范项目配套建设资金的落实，乙、丙方根据项目建设需要予以配合。

2．项目实施方案、中期检查、任务验收等由乙方负责，甲、丙方配合。财务审计由甲方负责，乙、丙方配合。

3．涉及甲方示范项目考核的技术经济指标由三方共同负责，其中专利、论文、研究生培养、关键技术、标准草案等由乙方负责；涉及甲方研究项目的考核指标由乙方负责完成。涉及甲方开发软件项目的考核指标由丙方负责完成。

4．课题中涉及到甲方的项目研究成果应由甲方和乙方共享，甲方参与项目的相关专利、论文等应双方署名。

5．甲方、乙方和丙方应成立各自的项目协调办公室，三方根据项目进程定期或不定期的开展交流，确保项目的正常开展。

## （二）经费管理办法

1．科研经费为科研专项资金，应专款专用，不得挪用，不得用于与科研项目无关的开支。课题组本着精打细算、厉行节约的原则合理有效地使用经费。

2．重庆水务集团分管财务、科研工作的领导和课题（含子课题）负责人对科研经费的管理和使用负责。

3．财务部门负责科研经费的财务管理和会计核算，并指导课题负责人按照财经法规在其权限范围内使用科研经费。财务部门应严格审核经费审批手续，按照财务制度规范核算，做到账目清楚，核算准确，确保资金的安全和合理使用。

4．课题（含子课题）负责人按照财经法规在其权限范围内使用科研经费，对科研经费使用的真实性、合法性、有效性承担经济与法律责任。

5．集团成立科研项目管理办公室，协助课题（含子课题）负责人合理分配科研经费, 指导课题组按预算书和实施进度开支科研经费。

6．科研经费支出按重庆水务集团的相关财务管理办法执行。课题（含子课题）负责人应严格按照合同约定或者经批准的项目经费预算核定的用途、范围和开支标准使用项目经费，自觉控制经费的各项支出。

7．课题（含子课题）负责人、科研项目管理办公室和其它相关人员应积极接受并配合各级部门和科研经费提供方或其委托的社会中介机构，依据国家有关法规、预算和科研合同对科研经费的管理和使用情况进行的检查监督。

8．科研经费由直接费用和间接费用组成。直接费用是指在项目（课题）实施过程（包括研究、中间试验试制等阶段）中发生的与之直接相关的费用。主要包括设备费、材料费、测试化验加工费、燃料动力费、差旅费等：

（1）设备费：是指在项目（课题）实施过程中购置或试制专用仪器设备，对现有仪器设备进行升级改造，以及租赁使用外单位仪器设备而发生的费用。各项目（课题）承担单位应当严格控制设备购置费支出。对于使用重大专项资金购置的单台/套/件价格在200万元以上的仪器设备，应当按照《中央级新购大型科学仪器设备联合评议工作管理办法（试行）》的有关规定执行。

（2）材料费：是指在项目（课题）实施过程中由于消耗各种必需的原材料、辅助材料等低值易耗品而发生的采购、运输、装卸和整理等费用。

（3）测试化验加工费：是指在项目（课题）实施过程中由于承担单位自身的技术、工艺和设备等条件的限制，必须支付给外单位（包括项目和课题承担单位内部独立经济核算单位）的检验、测试、设计、化验及加工等费用。

（4）燃料动力费：是指在项目（课题）实施过程中相关大型仪器设备、专用科学装置等运行发生的可以单独计量的水、电、气、燃料消耗费用等。

（5）差旅费：是指在项目（课题）实施过程中开展科学实验（试验）、科学考察、业务调研、学术交流等所发生的外埠差旅费、市内交通费用等。差旅费的开支标准应当按照国家有关规定执行。

（6）会议费：是指在项目（课题）实施过程中为组织开展相关的学术研讨、咨询以及协调任务等活动而发生的会议费用。应当参照国家有关规定，严格控制会议数量、规模、开支标准和会期。

（7）国际合作与交流费：是指在项目（课题）实施过程中相关人员出国及外国专家来华工作而发生的费用。国际合作与交流费应当执行国家外事经费管理的有关规定。

（8）出版/文献/信息传播/知识产权事务费：是指在项目（课题）实施过程中，需要支付的出版费、资料费、专用软件购买费、文献检索费、专业通信费、专利申请及其他知识产权事务等费用。

（9）劳务费：是指在项目（课题）实施过程中支付给项目（课题）组成人员中没有工资性收入的相关研发人员（如在校研究生等）和临时聘用人员等的劳务性费用。项目（课题）承担单位聘用的参与重大专项研究任务的优秀高校毕业生在聘用期内所需的劳务性费用和有关社会保险费补助，可以在劳务费中列支。

（10）专家咨询费：是指在项目（课题）实施过程中支付给临时聘请的咨询专家的费用。专家咨询费不得支付给参与课题（含子课题）研究及其管理的工作人员。专家咨询费的开支标准参考《中央财政科研项目专家咨询费管理办法》（财科教〔2017〕128号）。

（11）基本建设费：是指重大专项实施过程中发生的房屋建筑物购建、专用设备购置等基本建设支出，应当单独列示，并参照基本建设财务制度执行。

（12）其他费用：是指在重大专项项目（课题）实施过程中除上述支出项目之外的其他直接相关的支出。其他费用应当在申请预算时单独列示，单独核定。

9．间接费用是指项目（课题）承担单位在组织实施重大专项过程中发生的无法在直接费用中列支的相关费用。间接费用分为“管理费”和“项目（课题）收益”两项。

（1）管理费是指课题实施过程中，对承担单位管理课题、使用本单位现有仪器设备及房屋等，消耗本单位而又无法单独计量的日常水、电、气、燃料等有关费用的补偿。

（2）课题收益是指支付给课题承担单位的收益。课题收益根据课题的特点、承担单位性质等因素核定。

（4）间接费用由课题承担单位统筹使用和管理。间接费用中用于科研人员激励支出的部分，应当在对科研人员进行绩效考核的基础上，结合科研实绩，由所在单位根据国家有关规定统筹安排。

10．科研资金使用中涉及政府采购的，按照国家政府采购有关规定执行。使用科研资金形成的固定资产，按照《企业财务通则》等相关规章制度执行。科研资金形成的知识产权等无形资产的管理，按照国家有关规定执行。科研资金形成的大型科学仪器设备、科学数据、自然科技资源等，在保障有关参与单位合法权益的基础上，按照国家有关规定开放共享，以减少重复浪费，提高资源使用效率。

# 七、保障措施

## （一）技术的风险分析与对策

本课题的主要技术风险有两个方面：一是相关的核心技术研发能否成功取得突破；二是综合示范能够顺利实施并成功示范。

课题组织中将择优选定具有研究基础优势的科研机构承担各相关子课题，联合行业领先的企事业科研设计单位和部门组成产学研一体的联合攻关小组，以保证项目相关核心技术研发能够顺利展开。本项目负责人及承担单位，在城市水务方面，具有深厚的行业经验基础和沉淀，相关业务有多个行业领域专家作为业务流程运转保障；以徐光侠教授为负责人的重庆邮电大学科研团队，在大数据、人工智能等领域具备很强的技术储备，多次承担国家级、省部级重点项目，拥有授权专利十余项。项目实施前，项目负责人和领导小组进行深入讨论研究，同子课题承担单位和部门一起，对技术路线进行严谨设计，制定详细实施方案，多方面征求专家意见并进行严格审核。子课题所涉及的关键技术的开发，都保证具备一定的理论与实践基础。项目实施过程中，加强课题的阶段考核并实行滚动管理。建立课题联络机制，各子课题间加强合作与交流，总结成功经验，以保障研究与示范的顺利进行。成立课题咨询专家组，对课题实施进行咨询与评估。针对关键技术和难点焦点，在课题层面上建立国内外相关领域研究进展的跟踪机制，以保障研发的先进性。通过上述项目管理措施，努力将本项目的技术风险降到最低。

## （二）市场风险与对策

市场的风险主要来自于三个方面，一是市场需求的突然变化，二是成本更低的替代技术出现，三是由于供水安全保障相关产业的自然垄断性质导致的市场失灵。在我国经济持续高速发展的背景下，在可以预见的将来，人们或者各级政府对于供水安全保障的需求不可能下降，只会不断上升，因此第一种市场风险可以忽视。针对潜在替代技术的威胁，本专项一方面要求保证技术突破具备较强的成本有效性和创新性，另一方面致力于形成有效的跟踪机制，以保证研发技术的先进性和适用性与国际先进水平基本同步。

此外，城市供水安全保障技术不是纯粹的市场物品，它具有部分公共物品的属性，因此在某些方面，它需要政府的协调配合；同时，供水产业具有其特殊性，它具有一定的规模效应，属于自然垄断产业，因此市场对相关技术的接受容易受到行政干预的影响。防范此类风险的基本措施，一是需要与地方政府的密切配合和保持沟通，关注并满足地方政府的核心需求；二是重视标准化工作，遏制低劣技术产品的应用。

## （三）管理风险分析与对策

本课题综合性强、涉及范围广，项目实施涉及多个管理部门及多个研究部门，管理复杂，示范课题组织协调难度大。为了减少与避免管理风险，将采取如下措施：

1．成立课题协调小组，承担课题实施的指导、协调和监督管理工作；强调研究与主管部门的工作紧密结合，注重实效；

2．充分发挥咨询专家组的指导与咨询作用。在课题层面上建立关键领域和关键研究环节的数据库，对各子课题的进展进行适时跟踪，并跟踪相关领域的国际进展，以保证指导和协调的及时性和有效性。

3．课题组加强同地方政府部门之间以及工程业主单位的联系与协调，保证课题的顺利实施。

4．严格按照国家经费支出要求进行开支，对课题经费和财务进行科学监控与严格管理。

# 八、专家名单

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 序号 | 姓名 | 职称 | 身份证号 | 联系电话 |
| 1 |  |  |  |  |
| 2 |  |  |  |  |
| 3 |  |  |  |  |
| 4 |  |  |  |  |
| 5 |  |  |  |  |