

分类号	
学校代码	10700
学号	2210420043

西安理工大学

# 硕士学位论文

(学术学位)

基于大模型的洪涝灾害防御数字化孪生系统的研究与实现

何果

学科门类: 工学

一级学科: 水利工程

二级学科: 水文学及水资源

指导教师: 罗军刚

申请日期: 2024年6月

# 论文题目：基于大模型的洪涝灾害防御数字化孪生系统的研究与实现

学科名称：水文学及水资源

研究生：何 果

签名：\_\_\_\_\_

指导教师：罗军刚 教授

签名：\_\_\_\_\_

## 摘 要

建设洪涝灾害防御工程旨在应对、预防和减轻洪涝灾害对人类社会、环境和经济的影响。在洪涝灾害频发、管理需求迫切以及计算机水平发展迅速的新时代背景下，研究基于新型信息技术的洪涝灾害防御系统对提高业务人员的管理水平以及决策能力具有重要意义。但是目前的洪涝灾害防御系统存在着应用功能分散、衔接困难、计算过程不够透明以及缺少宏观模拟等问题。因此，本论文以甘肃省临洮县为研究对象，将数字孪生技术引入洪涝灾害防御系统，并提出基于大模型驱动专业模型和业务过程的应用模式，结合目前常用于洪涝灾害防御的模型，设计并开发了一套洪涝灾害防御系统，在实现智能化交互的过程中，发挥行业信息化的引领作用，逐步推广新的应用范式。论文的主要成果如下：

(1) 对洪涝灾害防御系统的架构进行研究并对其功能进行设计。通过对系统应用需求的分析，对系统的技术体系进行研究设计，在网络地理信息系统、可视化展示以及大模型等技术手段和组件化业务的应用理念支撑下，提出了基于大模型驱动的洪涝灾害防御数字化孪生系统的架构，并从整体视角实现对系统功能的设计。

(2) 基于组件化思维的专业模型集成体系研究。基于模型结构和逻辑上的特性，利用组件思维将洪水预报模型、洪水演进模型、防洪调度模型以及参数优选算法划分为相互独立的基础模块，以标准化接口的方式将模块按照输入输出约束进行拼接，构建中粒度模型，形成应用组件库，实现专业模型的集成开发，解决模型开发效率低，代码复用率差的问题，并为创建人机结合的大模型平台打下基础。

(3) 研究大模型驱动业务知识-数据-模型-系统的应用模式。基于大模型，通过对知识融合方法的研究，采用 RAG (Retrieval Augmented Generation) 策略将多源异构信息利用大模型进行统一管理，构建一个业务“知识场”。同时，通过对智能体 (agent) 的研究，以构建洪涝防御业务推理链和大模型业务库的方式，提出大模型驱动洪涝数据和模型的应用模式，并从系统设计的角度，探索大模型与业务过程的融合方法，改变传统业务系统应用模式，为创建人机语言交互的智能平台提供思路。

(4) 开发洪涝灾害防御数字孪生平台。基于系统的架构，利用 DEM、遥感影像、三维模型等数据实现孪生实体的可视化构建；采用 Vue、Spring Boot 等开发技术进行前端和 Java 服务端的开发；利用 Numpy、Pandas 等 Python 框架实现对专业模型端以及大模型应用端的开发。通过集成实时洪水预报、透明化预演、智能化预演等功能实现洪涝灾害防御的业务应用。

**关键词：**洪涝灾害防御；系统设计；数字孪生；组件化；大模型

**Title: RESEARCH AND IMPLEMENTATION OF DIGITAL TWIN SYSTEM FOR FLOOD DISASTER PREVENTION BASED ON LARGE LANGUAGE MODEL**

**Major: Hydrology and Water Resources**

**Name: Guo HE**

**Signature: \_\_\_\_\_**

**Supervisor: Prof.Jungang LUO**

**Signature: \_\_\_\_\_**

**Abstract**

The construction of flood control and disaster defense engineering aims to address, prevent, and mitigate the impact of flood disasters on human society, the environment, and the economy. Against the backdrop of frequent flood disasters, urgent management needs, and rapid development in computer technology, researching flood defense systems based on new information technologies is significant in improving the management and decision-making capabilities of professionals. However, current flood defense systems suffer from issues such as dispersed application functions, difficulty in integration, lack of transparency in computational processes, and a shortage of macro-level simulation. Therefore, this thesis takes Lin Tao County, Gansu Province as the research object, introduces digital twin technology into flood defense systems, and proposes an application model based on large language models driving professional models and business processes. Combining with commonly used models for flood defense, a flood defense system is designed and developed, gradually promoting new application paradigms while achieving intelligent interaction and serving as a guide for industry informatization. The main achievements of the thesis are as follows:

(1) Research on the architecture of flood defense systems and design of their functions. Through the analysis of system application requirements, the technical system of the system is researched and designed. Supported by technologies such as network geographic information systems, visualization displays, and large language models, as well as the concept of componentized business, an architecture for a digital twin system for flood defense driven by large language models is proposed, and the design of system functions is achieved from a holistic perspective.

(2) Research on a componentized professional model integration system. Based on the

characteristics of model structure and logic, flood forecasting models, flood evolution models, flood control scheduling models, and parameter optimization algorithms are divided into mutually independent basic modules using component thinking. These modules are spliced according to input-output constraints in a standardized interface manner to build medium-grained models, form application component libraries, realize integrated development of professional models, and solve the problems of low model development efficiency and poor code reuse. This also lays the foundation for creating a large language model platform combining human and machine.

(3) Research on the application pattern of large language models driving business knowledge-data-model-system. Based on large language models, through the study of knowledge fusion methods, the RAG (Retrieval Augmented Generation) strategy is adopted to unify and manage multi-source heterogeneous information using large models, constructing a business "knowledge space". At the same time, through research on intelligent agents, a flood defense business reasoning chain and a large model business repository are proposed to build a model-driven flood data and model application pattern, and from the perspective of system design, explore the integration method of large models with business processes, change the traditional business system application pattern, and provide ideas for creating intelligent platforms for human-machine language interaction.

(4) Development of a flood defense digital twin platform. Based on the system architecture, visualization construction of twin entities is achieved using DEM, remote sensing images, 3D models, etc. Front-end and Java server development are carried out using development technologies such as Vue, Spring Boot, etc. Development of professional model ends and large language model application ends is realized using Python frameworks such as Numpy, Pandas, etc. Flood defense business applications are implemented through the integration of real-time flood forecasting, transparent simulation, and intelligent simulation functions.

**Key words:** Flood disaster defense; System design; Digital twin; Modularization; Large language model

# 目 录

1 绪论.....	1
1.1 研究背景与意义.....	1
1.1.1 研究背景.....	1
1.1.2 研究意义.....	1
1.2 国内外研究现状.....	2
1.2.1 洪涝灾害防御系统.....	2
1.2.2 数字孪生应用.....	3
1.2.3 大模型应用.....	3
1.3 存在的问题.....	4
1.4 主要研究内容.....	4
1.5 技术路线.....	5
1.6 研究区域概况.....	6
2 洪涝灾害防御系统架构研究与设计 .....	7
2.1 应用需求分析.....	7
2.1.1 业务需求分析.....	7
2.1.2 非业务需求分析.....	8
2.2 技术体系研究.....	8
2.2.1 数据可视化.....	8
2.2.2 网络地理信息系统.....	10
2.2.3 大模型技术.....	11
2.2.4 虚拟建模技术.....	12
2.3 系统架构设计.....	12
2.3.1 系统功能架构设计.....	12
2.3.2 系统技术架构设计.....	13
2.4 业务功能设计.....	14
2.4.1 系统业务划分.....	14
2.4.2 数据管理服务.....	15
2.4.3 洪涝防御“四预”服务 .....	15
2.5 非业务功能设计.....	19
2.5.1 用户权限设计.....	19
2.5.2 数据库设计.....	20
2.6 本章小结.....	21
3 洪涝灾害防御专业模型集成研究 .....	22
3.1 多模型集成体系.....	22

3.2 洪水预报模型.....	22
3.2.1 新安江模型.....	22
3.2.2 GR4J 水文模型.....	24
3.2.3 预报模型组件化.....	24
3.3 参数优选算法.....	28
3.3.1 参数优选方法.....	28
3.3.2 精度评定标准.....	30
3.3.3 参数优选组件化.....	30
3.4 洪水演进模型.....	31
3.4.1 马斯京根模型.....	31
3.4.2 一维水动力学模型.....	32
3.4.3 演进模型组件化.....	33
3.5 防洪调度模型.....	38
3.5.1 调度模型构建.....	38
3.5.2 调度模型组件化.....	39
3.6 模型组件库建立.....	40
3.7 本章小结.....	41
4 基于大模型的洪涝业务应用模式研究 .....	42
4.1 大模型与洪涝知识融合研究 .....	42
4.1.1 大模型知识融合方法.....	42
4.1.2 知识向量化处理.....	43
4.1.3 知识应用策略.....	45
4.1.4 洪涝知识库构建.....	47
4.2 大模型与洪涝数据及模型融合研究 .....	48
4.2.1 大模型 Agent 概述 .....	48
4.2.2 大模型与洪涝业务交互策略.....	49
4.2.3 洪涝业务推理链构建.....	50
4.2.4 多策略下意图识别增强.....	51
4.2.5 大模型驱动洪涝数据及模型实现.....	53
4.3 大模型与业务过程融合研究 .....	54
4.4 本章小结.....	55
5 洪涝灾害防御数字孪生平台的构建与应用 .....	56
5.1 数字孪生场景构建.....	56
5.1.1 流域地形可视化.....	56
5.1.2 流域地物模型可视化.....	56
5.1.3 流域水系可视化.....	57
5.2 前端架构搭建.....	58
5.3 后端架构搭建.....	59

5.3.1 数据服务端搭建.....	59
5.3.2 业务算法模型端搭建.....	61
5.3.3 大模型端搭建.....	62
5.4 洪涝灾害防御业务应用.....	63
5.4.1 数据管理.....	63
5.4.2 综合监控面板.....	66
5.4.3 实时洪水预报.....	66
5.4.4 过程透明化预演.....	68
5.4.5 智能化预案.....	70
5.5 本章小结.....	71
6 结论与展望.....	72
6.1 结论.....	72
6.2 展望.....	72
参考文献.....	74



# 1 绪论

## 1.1 研究背景与意义

### 1.1.1 研究背景

《2023 年中国水利发展报告》对 2023 年的水利工作提出了分析和展望，以推动新阶段水利高质量发展，加快数字孪生建设，部署水旱灾害防御为重点工作<sup>[1]</sup>。在“十四五”规划和 2035 年远景目标规划中，明确提出了七大数字经济重点产业和十大数字化应用场景。数字流域作为水利现代化建设的前沿概念，不断取得突破性的创新，先后提出了“数字黄河”、“数字黑河”、“数字长江”等数字流域建设设想，并取得了显著的成果，水利事业正在向着智慧化的方向前进。

随着数字技术的发展，人工智能、大数据、虚拟仿真等技术的广泛应用，许多传统行业正在逐步迈入信息化和智能化的时代，跨界和融合成为行业发展的新能源。人工智能（AI）作为当前产业革命的关键动力，在赋能传统行业升级方面成效显著。近年来，大模型（LLM）已经成为人工智能研究和应用的热门话题，其展示了卓越的语言理解、生成和知识推理能力，实现了对生产力的大幅度提升。尽管 LLM 已经广泛应用于医疗保健、金融、教育、自动化生产等领域，但在水利行业几乎没有涉及，如何将最先进的 AI 技术应用于水利行业，应用于什么场景是一个需要考虑的问题。

洪涝灾害作为全球最常见的自然灾害之一，对人类社会和经济造成了巨大的影响。随着气候变化的不断加剧，洪涝事件的频率和严重性呈上升趋势。这使得洪涝灾害的防御成为当今社会的紧迫问题。从最早的基于经验和工程模型到现在的平台管控，针对洪涝灾害防御的能力有了大幅度的提高。但目前的系统技术体系和防御方法仍有很大的提升空间，一是业务人员信息化水平通常相对滞后，而系统应用存在功能分散、衔接困难，常常会出现系统已开发完毕，但是业务人员不会使用的问题；二是目前很多洪涝灾害防御系统主要停滞在图表数据展示和二维可视化的层面，在宏观模拟方面存在着不足，且对于预案在线化、知识化程度不够，业务人员仍需要查阅纸质文本，不仅容易漏，而且效率低。因此，如何更新洪涝灾害防御技术体系和方法，实现智慧化管理，减少业务人员学习成本，使未曾接触过信息化的人都能根据自身需求完成业务，是另一个需要解决的问题。

### 1.1.2 研究意义

随着物联网、大数据、云计算、人工智能在各行各业的发展融合不断加深，孤立的水利信息化相对滞后。在深入剖析存在的问题和应用特点的基础上，明确了新时代洪涝灾害防御系统需要解决的问题，水利信息化发展需要跨界与融合、集成应用模式需要集成创新。

面对所提到的两个问题，基于现有的信息化技术体系，跨模态跨领域的数据和知识深度学习模型成为解决这一问题的关键途径，于是本文提出将数字孪生技术和大模型与洪涝

灾害防御系统相结合，通过对多模态数据的统一管理，利用数字孪生技术和数据可视化技术，实现更直观的洪涝灾害场景模拟，便于业务人员更直观的把握洪涝情况，为公众提供更可靠的洪涝灾害防护。同时，引入大模型，将与洪涝业务相关的方法和模型集成化，改变传统模型库和知识库的应用体系，利用大模型结合可视化设计和交互技术，提供更易用的界面交互方式，实现对用户输入的自动解析，减少业务人员使用系统的难度。在实现智能化交互的过程中，发挥行业信息化的引领作用，逐步推广新的应用范式。

## 1.2 国内外研究现状

### 1.2.1 洪涝灾害防御系统

洪水是世界上最频繁和最具破坏性的自然灾害之一。面对人口增长、城市化加速以及全球气候变化的挑战，洪涝灾害的频繁发生对民众安全与城市经济构成了巨大风险。越来越多的人开始关注洪涝、干旱等灾害的系统性综合治理。根据发生洪涝灾害的区域不同，可以将洪涝灾害防御系统分为流域性的洪涝灾害防御系统以及城市洪涝防御系统。

洪涝灾害防御系统在流域层面通常以融合水文模型和调度模型来构建，通过整合水文、气象和水库等多源数据，依托河网水系构建综合性的流域防洪体系。例如，李勇<sup>[2]</sup>等通过集成预报技术，利用水文流量数据，成功研发了湄江河流域的洪水监测与预警系统，从而提升了该流域对洪水监测与预警的整体效能。李宗花<sup>[3]</sup>等针对洪泽湖流域的洪涝灾害，开发了包含数据库、仿真模拟和辅助决策三个层面的模型框架，并据此建立了洪泽湖流域的洪涝灾害预警系统。朱建英<sup>[4]</sup>等以江苏省淮河流域为研究背景，阐述了水旱灾害防御调度系统的理念，并就如何提升调度效能提出了具体建议。杨芳<sup>[5]</sup>等通过采用一二维水动力模型、水库群多目标联合调度模型等专业模型，运用水情、雨情等数据，探讨了水利专业模型在珠江流域水旱灾害防御中的应用。

随着城市内涝问题的频繁出现，众多城市开始致力于洪涝灾害防御的探索，并在这一过程中取得了显著成果。Zaalberg<sup>[6]</sup>等运用交互式 3D 技术，对城市洪涝灾害进行虚拟仿真，讨论了在虚拟仿真中传达现实世界洪水风险对受威胁居民的影响。郑晓阳<sup>[7]</sup>等在 GIS 支持下建立了上海市防汛决策支持系统，成功实现对上海市洪涝灾害的精准管控。Macchione<sup>[8]</sup>等采用二维水力建模方式，通过洪水模拟得出三维可视化结果，全面评估了城市洪涝灾害的系统情况。韩刚<sup>[9]</sup>等以深圳市河湾流域为研究对象，以洪涝模型为基础，构建并实施了一套依托三维数字城市技术的洪涝预警调度系统，实现了对城市洪涝过程的精细模拟和风险评价。

近些年来，各国在洪涝灾害防御系统的建立和设计方面积累了许多理论和实践经验，通过引入信息化和智能化设计，提升了业务人员的决策能力，从而有效保障了人民的人身和财产安全。但从文献数量来看，大多数系统仍然局限在数据展示和二维可视化的层面，对于引入三维模拟仍处于探索阶段。

### 1.2.2 数字孪生应用

数字孪生 (Digital Twin) 是近年来备受关注的研究领域, 国内外学者在这一领域展开了广泛且深入的研究。其概念最早可以追溯到 2003 年, 当时 Grieves<sup>[10]</sup>教授首次提出并将其称为“物理产品的虚拟数字表示”。到 2016 年, 他与 John Vickers 共同提出了“数字孪生”的概念<sup>[11]</sup>。时至今日, 这个术语已被研究人员广泛接受, 可以将其定义为通过互联网技术建立虚实交互的接口, 将物理空间中设备的各种属性在虚拟空间中进行完整映射。

早期的数字孪生技术主要应用于航空航天领域, 美国空军将数字孪生技术引入飞行器的生产工业流程, 实现了在飞行器生产上效率和质量的显著提升<sup>[12]</sup>。随着计算机水平逐渐提高, 数字孪生技术逐渐渗透到各行各业。Boje<sup>[13]</sup>等人讨论了数字孪生技术在建筑信息建模 (BIM) 中的概念和应用, 并提出了未来需要研究的技术路线。Wang<sup>[14]</sup>等人提出了一种构建海上风力机的数字孪生框架, 实现对风力机支撑结构的实时监测、故障诊断和运行优化, 为未来风力机支撑结构的可靠性分析提供有益的应用前景。Onile<sup>[15]</sup>等人总结了数字孪生技术在能源领域的应用前景以及目前遇到的问题。Milton<sup>[16]</sup>等人提出了一种基于实时概率的数字孪生在线分析方法, 并成功应用于电子转化器的监测和诊断问题。

伴随着“数字地球”的提出<sup>[17]</sup>, 开创了全球“智慧”建设的潮流。为了应对气候变化和人类活动增加带来的流域管理挑战, 数字地球概念催生了“数字流域”的概念, 数字孪生技术开始应用于流域建设。张秋文<sup>[18]</sup>等研究了数字流域整体框架构建, 分析了数字流域的理论基础和关键技术。王忠静<sup>[19]</sup>等在数字流域的研究基础上, 进一步阐述了数字化流域的理念, 并建立了以数字化流域为基础的水资源规划数字化平台架构。自 2021 年以来, 在国家的大力推动下, 数字孪生流域成为全社会的热点<sup>[20]</sup>。王凯<sup>[21]</sup>等提出了淮河流域防洪“四预”数字孪生建设的总体框架、实施路径和布局策略。李卫政<sup>[22]</sup>等以甬江流域水资源管理与调配“四预”流程为切入点, 构建水资源管理与调配业务应用系统, 最终实现水资源的智能优化调配与管理。此外, 在黄河流域<sup>[23]</sup>、长江流域<sup>[24]</sup>等地也引入数字孪生技术展开了数字孪生流域的探索, 并取得了显著成果。

### 1.2.3 大模型应用

在过去的二十年中, 语言建模作为理解和生成语言的主要方法得到了广泛研究。随着 2022 年底基于 GPT-3.5<sup>[25]</sup>的 ChatGPT 发布, 包含千亿甚至更多参数的大型语言模型 (LLM) 逐渐引起了社会的关注, 相关研究迅速增多。在这一趋势下, 基于多样化架构的大模型不断涌现。例如, 谷歌的研究团队采用了 GPT 风格的单向自回归模型结构, 成功训练了一个拥有 5400 亿参数的语言模型 PaLM<sup>[26]</sup>。Touvron<sup>[27]</sup>等人基于 transformer 结构, 采用 decoder-only 的方式, 构建一个从 7B 到 65B 参数的基础语言模型集合 LLaMA。国内百度团队提出了兼顾自然语言理解和生成的框架 ERNIE<sup>[28]</sup>, 使得经过训练的模型可以通过样本学习或微调的方式针对自然语言理解和生成任务进行定制。清华大学与智谱 AI 共同研制了双向密集预训练语言模型 GLM<sup>[29]</sup>, 以自回归空白填充的方式实现对模型的训练。

目前, LLM 广泛应用于医疗保健、教育、金融等行业, 为其带来了巨大的生产力, 在医学领域方面, Tang<sup>[30]</sup>等人通过利用 GPT 生成大量带标签的高质量合成数据, 为 LLM 在临床文本挖掘中的适用性提供了解决方案。Nov<sup>[31]</sup>等人在医疗咨询问题上, 通过将患者问题输入到 ChatGPT 的方式构建聊天机器人, 结果表明 ChatGPT 对患者问题的回答与提供者的回答几乎没有区别。Yang<sup>[32]</sup>等人对 LLM 在心理健康分析和情绪推理方面进行全面评估, 显示其在可解释的心理健康分析方面有巨大潜力。Singhal<sup>[33]</sup>等人提出了 MultiMedQA 框架和 Med-PaLM 模型, 通过结合现有开放问答数据集, 在美国医疗许可考试中取得了专家级的表现, 并在回答消费者的医疗问题时得到医生的认可。在教育领域, Malinka<sup>[34]</sup>等人通过搜集考试、编程作业和学期论文等方面的数据, 评估了 ChatGPT 对大学教育的影响, 发现其可作为辅助工具, 用于讨论问题解决或加速学习过程。Susnjak<sup>[35]</sup>等人评估了 ChatGPT 执行高级认知任务的能力, 结果显示 GPT 能够生成逻辑一致的跨学科答案。在金融领域, Araci<sup>[36]</sup>等人引入了 FinBERT 模型用于处理金融领域的 NLP 任务, 结果显示在两个金融情绪分析数据集上, 各项测量指标都有所改善。

关于大模型在实际业务应用的落地, 各行各业都在积极探索, 试图将其引入到本行业中, 以提升行业的生产力。尽管大模型已经在多个领域进行尝试, 但在水利行业中却几乎没有涉足。

### 1.3 存在的问题

目前信息化系统开发和智慧水利建设仍存在着一些问题和挑战。

(1) 传统信息系统设计下的关键信息获取效率需进一步提高。

目前技术体系下的智慧水利信息化系统都存在多级菜单、众多按钮、复杂界面的问题, 这样就导致业务人员学习成本加大, 很难快速掌握系统使用, 有时为了找一个界面的内容需要耗费不少的时间和精力。智慧水利项目中涉及众多业务系统及各种各样的数据, 工作人员很难在短时间内全部掌握并快速获取当前最关心的问题答案。

(2) “四预”应用过程中数据-模型-知识融合程度不足。

支撑“四预”应用依赖于数据、模型计算、专家经验和业务规则的紧密结合。传统的应用建设过程中知识和数据的串联仍需要由人通过线下讨论的方式逐步对齐和深入, 数字孪生的应用目标难以通过信息系统开发呈现自动化、智能化的效果。

### 1.4 主要研究内容

本文以甘肃省临洮县为研究对象, 针对现有洪涝灾害防御系统的不足, 旨在建立一个能够实现三维可视化展示的洪涝灾害防御系统, 通过对大模型的研究, 对系统的应用模式进行改造, 并利用数字孪生技术和网站开发技术对系统进行实现。本文的研究内容主要包括以下四部分。

(1) 系统研究与设计。对当前洪涝灾害防御系统现状及存在问题进行分析, 从实际应用需求出发, 分析系统的技术体系, 构建出系统的总体架构, 并结合系统的架构, 对系统

的功能体系进行设计。

(2) 专业模型集成研究。基于模型结构和逻辑上的特性，分析了洪水预报模型、洪水演进模型、防洪调度模型以及参数优选算法的基础模型与方法，通过基础方法构建中粒度模型的方式，利用组件技术，将基础方法封装为组件，按照输入输出约束进行组件拼接，构建模型实例，实现了专业模型的集成开发。

(3) 基于大模型的洪涝业务应用模式研究。通过对大模型技术体系的研究，采用 RAG 策略实现针对洪涝业务的知识库构建，完成洪涝知识与大模型的融合，实现大模型驱动洪涝知识的应用方式。同时，通过对大模型 Agent 的研究，通过构建洪涝业务推理链、大模型业务库以及意图识别策略的方式，实现大模型驱动洪涝数据和模型，并通过对系统运行方式的设计，提出大模型驱动业务系统的应用模式。

(4) 系统开发与应用。通过对核心理论的研究，依据洪涝灾害防御系统的设计需求，利用孪生可视化、软件开发等技术实现孪生场景的搭建以及系统前后端的开发，并对平台的界面设计以及业务应用进行介绍。

### 1.5 技术路线

本文从洪涝灾害防御系统现状问题出发，探究大模型和数字孪生技术在洪涝灾害防御系统中的应用。构建了基于大模型驱动的洪涝灾害防御系统框架，根据洪涝灾害业务，设计各个模块功能，并以数字流域为底板，实现大模型驱动专业小模型，驱动业务系统的应用流程。最后，依据需求对系统进行开发实现。论文的技术路线如图 1-1 所示。

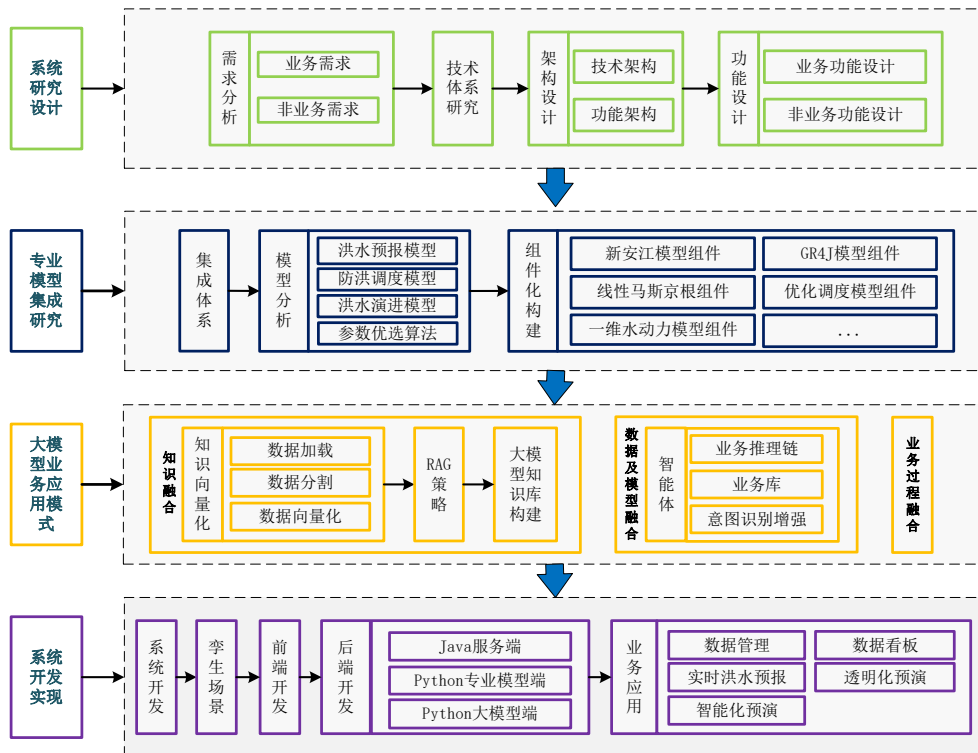


图 1-1 研究技术路线图

Fig.1-1 Research technology roadmap

## 1.6 研究区域概况

### (1) 地理信息

临洮县位于甘肃省中部、定西市西部，地理坐标为东经 103°29′~104°19′，北纬 35°03′42″~35°56′46″。地处黄土高原与青藏高原的交汇地带全县总面积为 2851 平方千米。地势较为复杂，属于陇中黄土高原丘陵沟壑区，地势东南高西北低，分布区域地形变化差异大<sup>[37]</sup>。洮河穿城而过，其支流众多，散流于各地，位于临洮县内的主要一级支流包括漫坝河、东峪沟、红道峪沟、大碧河、上营沟、中辅沟等。图 1-2 为临洮县流域水系分布图。

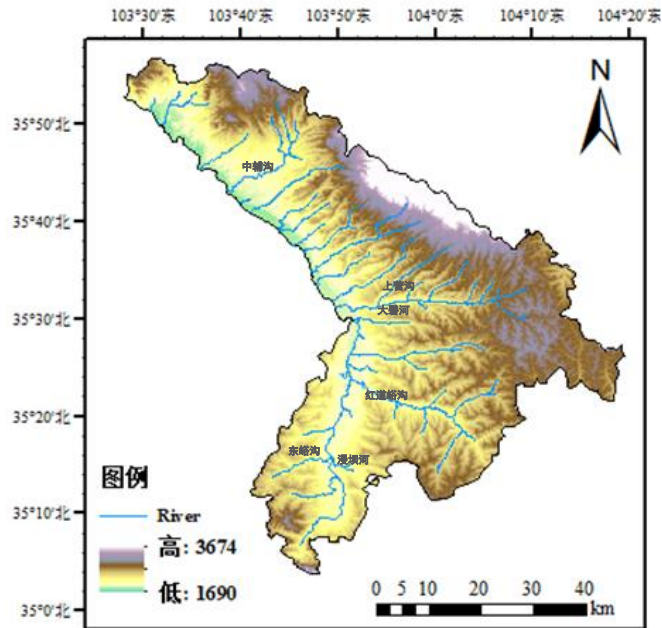


图 1-2 研究区域水系分布图

Fig.1-2 River system distribution maps of the study area

### (2) 水文气候特征

洮河是黄河上游最大的支流，从海甸峡水库入临洮，平均径流为  $138\text{m}^3/\text{s}$ ，最大为  $1790\text{m}^3/\text{s}$ ，最小为  $17.15\text{m}^3/\text{s}$ 。洮河径流主要来源于大气降水，以雨水补给为主。春汛期为 3 月中旬到 5 月，夏汛时期为 6 月到 9 月，其中 7-9 月径流量占全年比例最大，约为 43.5%<sup>[38]</sup>。临洮县属温带大陆性气候，四季分明，年平均气温为  $7^\circ\text{C}$ （最高气温  $34.6^\circ\text{C}$ ，最低气温  $-29.5^\circ\text{C}$ ）。无霜期为 80-190 天，年平均降雨量在 317-760mm 之间，蒸发量为 1400mm 以上。年内降雨分布不均，70% 以上的降水集中在 7、8、9 三个月<sup>[39]</sup>。

## 2 洪涝灾害防御系统架构研究与设计

洪涝灾害防御系统的开发旨在通过科学技术手段和有效管理策略，减少洪涝灾害对人类、社会、环境和经济的影响。传统洪涝灾害系统存在功能分散、衔接困难以及业务人员学习成本较大的问题。为了解决这些难题，本章从系统的应用需求出发，探讨系统的技术体系，对系统的架构以及业务功能进行研究和设计。

### 2.1 应用需求分析

#### 2.1.1 业务需求分析

系统首先应具备直观的监控面板，能够对雨情、水情等监测信息的变化过程进行直观的展示，便于业务人员能够及时了解到当前流域情况以及未来的变化趋势。其次为了提高业务过程的可靠性，每个业务计算的中间过程应该是透明、可调节的，计算结果应该是孪生可视化的，便于业务人员利用其经验对计算过程进行修正，得到更为准确的计算结果。最后，系统应确保业务人员能够全覆盖使用，而不需要专业培训。这意味着系统既要保留传统菜单式的操作方式，又要引入语言控制业务的方式，以减轻业务人员的使用难度。根据需求，洪涝灾害防御系统业务可分为基础数据管理和洪涝“四预”业务应用，具体包括：

##### (1) 基础数据管理

洪涝灾害防御系统的核心在于对各类数据的有效整合，系统应包含各方面的数据，如：实时监测数据、监测站点数据、水库数据、历史数据以及各类业务数据。数据应具备统一管理，保证数据的存储效率及检索速度，在应对多用户同时操作数据时做到分布式处理、高并发读写的要求。

##### (2) 洪涝“四预”业务应用

洪涝灾害防御系统的“四预”业务主要包括洪涝的预报、预警、预演及预案。

1) 预报：针对洪涝的预报功能主要包括降雨预报和洪水预报。降雨预报能够依据远程的降雨接口展示最近 7 天的降雨情况。洪水预报能够实现在遥感监测设备的基础上，获取相关数据，运行产汇流模型以及河道演进模型，对指定断面的洪水过程进行预报，获得该断面的洪水三要素。

2) 预警：预警分为实时预警、预报预警、预演预警三个方面。实时预警依据获得的实时数据去判断是否超过警戒线，超过警戒后通知该部分的负责人，并在 GIS 地图上闪烁预警。预报预警能够实现在预报过程中，对预报结果达到警戒线的地区进行预警展示。预演预警则是在预演过程中，当预演洪水对防护区造成危害时给出预警。

3) 预演：预演分为洪水预演和调度预演。洪水预演是在确定暴雨中心和防护对象的前提下，能够对沿途过程中每个中间断面进行洪水预报，得到洪水频率，并根据洪水频率和历史淹没数据给出洪水淹没范围，判断是否对防护区造成影响。调度预演则是在确定水库

来水的前提下，对水库的调度算法或下泄流量进行调节，得到不同下泄流量对下游各断面的影响，便于制定同时保障大坝以及防洪对象安全的洪水泄流计划。

4) 预案：能够做到预案在线化，支持语言交流方式获取预案内容，并能够实现根据洪水信息智能推荐预案。

### 2.1.2 非业务需求分析

除了满足业务需求外，确保系统的安全性和稳定性也至关重要。以下是对系统的安全性和数据库的非业务需求进行的分析。

#### (1) 安全需求分析

洪涝灾害防御系统的设计需要在权限管理、数据加密、系统日志等方面进行安全体系建设，防止不同业务的人员操作不属于自身业务部分的功能，造成数据混乱、数据泄露等问题，确保系统的稳定运作。

1) 权限管理：实施角色分配，确保不同用户只能访问其相关业务功能。例如，普通用户只能查看数据而不能编辑或删除，而超级管理员具有所有权限，并可进行权限分配。

2) 数据加密：对用户密码和敏感数据进行加密处理，以防止未经授权的访问和数据泄露。

3) 系统日志：实现系统平台的自动记录，包括用户操作、访问记录和系统错误。有助于及时排查问题，提高系统的可维护性。

#### (2) 数据库需求

对于普通的设备数据可采用 SQL 类数据库进行存储，加快数据的检索效率，而对于需要经常查询而不常更新的业务类数据，应采用 NOSQL 形式的缓存类数据库进行存储，减少系统的运行压力。

## 2.2 技术体系研究

基于以上的需求分析，本节在需求的基础上对系统的技术体系进行研究和分析。

### 2.2.1 数据可视化

数据可视化是一种前端技术，将数据以图形、图表等可视化的方式呈现在系统中，并允许用户使用不同的数据源进行自定义分析。数据可视化技术主要包括数据通讯技术以及数据展示技术。

#### (1) 数据通讯

系统中包含大量的数据，要做到这些数据的可视化展示，必须先做到与数据进行交互。不论是实时数据还是历史数据，产生数据的第一时间都会存储于数据库中，所以数据的交互归根结底就是前端与数据库的通讯，也就是前端与后端的通讯。实现前后端通讯的流程如图 2-1 所示。



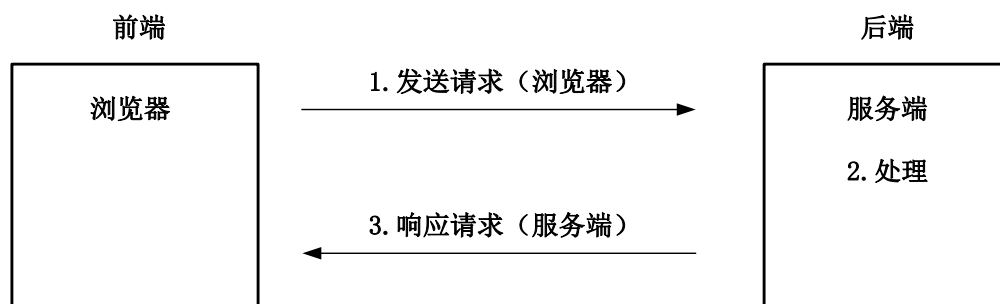


图 2-1 前后端通讯流程图

Fig.2-1 Front and rear end communication flowchart

在 Web 开发中，前后端通讯是实现用户与服务器数据交互的重要组成部分，以下是一些常见的数据通讯技术：

1) **AJAX (Asynchronous JavaScript and XML)**: AJAX 是一种用于在不重新加载整个页面的情况下，异步地与服务器进行数据交互的技术<sup>[40]</sup>。通常使用 JavaScript 发送异步请求，可以获取或发送数据，然后通过 DOM 操作更新页面。

2) **HTTP (Hypertext Transfer Protocol)**: HTTP 是一种基于请求-响应模型的协议，用于在客户端和服务器之间传输数据。前端通过发送 HTTP 请求到后端，后端响应并返回相应的数据。

3) **Messaging Queue**: 消息队列是一种异步通信机制，常见的消息队列包括 RabbitMQ 和 Kafka 等。其原理是通过前后端发布和订阅消息的形式进行数据交换，通常被用于解耦前后端的通信，实现可靠的消息传递。

4) **WebSocket**: 是 Web 浏览器和服务器之间的一种全双工通信协议，它允许客户端和服务器之间建立持久的连接，并实现实时的双向通信<sup>[41]</sup>。相较于传统的 HTTP 请求，WebSocket 提供了更低的延迟和更高的性能，特别适用于需要实时通信的场景。

在系统业务中，气象、云图等信息主要通过调用第三方的在线接口集成到系统中，对于这类只需要查询、不需要反馈的数据，可采用直接发送 HTTP 请求或者通过 AJAX 进行处理。对于业务计算，需要调取后端业务逻辑，考虑到系统的运行效率以及用户体验，计算过程最好是异步执行，且能够实现不刷新页面就能实现交互。因此，对于业务计算上的数据通讯，采用 AJAX 是最佳选择。对于调取专业模型的通讯，则需要引入消息队列进行通讯，如果专业模型的计算时间过长，会导致前后端通讯超时，为了解决这个问题，引入消息队列可以有效地进行异步通讯，确保系统的稳定性和效率。而在大模型的交互模块中，需要让大模型依据用户的语义，实现通过语言调取业务的目的。为了实现这一需求，前端与后端的接口需要保持一个长连接，一旦后端感知到大模型的反馈，应能够迅速向前端发出响应，以调取业务场景。在这个方面，WebSocket 则是最佳选择，能够提供实时的双向通讯，确保语言模型与系统的高效互动。

通过以上的分析，系统中的通讯技术主要采用 AJAX、WebSocket 以及消息队列中的

RabbitMQ。

## (2) 数据展示

在浏览器中的 Web 前端页面，常用的可视化方法大致分为四种：HTML/ CSS、SVG、Canvas 以及 WebGL。其特点见表 2-1。

表 2-1 常用可视化方法特点

Tab.2-1 Characteristics of common visualization methods

方法名称	优点	缺点	应用
HTML/CSS	易学习和使用，适用于简单的页面布局和基本样式	不适合处理复杂的图形	网页布局、简单图标、基本样式
SVG	矢量图形，可无损缩放，适合中等规模的图形	对于大规模数据的可视化，性能有限	图表、地图、中等规模矢量图形
Canvas	性能较高，特别适合处理大量的图形数据	不支持 DOM 元素，不便于交互	复杂动画、游戏、绘图应用
WebGL	提供硬件加速，性能最高，适合高性能的 3D 图形	对于简单的可视化，使用 WebGL 可能过于庞大	高性能的 3D 图形应用

针对以上数据可视化方法，目前有很多三方工具和库，实现了对可视化方法的二次开发，表 2-2 为常用数据展示工具库的对比分析。

表 2-2 常用数据展示工具库对比

Tab.2-2 Common data display tool library comparison

方法	优点	缺点	应用
ECharts	提供丰富的图表类型和样式，支持 SVG 和 Canvas 渲染，适用于大数据集	定制性相对较弱	常见的数据可视化需求
D3	极高的灵活性和定制性，可实现几乎任何类型的数据可视化	使用较为复杂	复杂、高度定制的数据可视化需求
vis	轻量级，提供了易用的网络图和时间轴可视化组件	社区相对较小	绘制关系图，网络拓扑图
Nivo	支持 React 生态系统，可轻松集成到 React 应用中	部分图表类型相对较少	快速实现现成图表需求，尤其是在 React 应用中

系统中的数据展示主要包括数据变化过程的展示、数据比例的展示以及流域拓扑图的展示。在数据看板中实现数据的快速渲染以及在业务过程中实现数据内容的在线调节是十分重要的。ECharts 作为目前使用最为广泛的数据可视化工具库，支持大数据集的渲染，并且集成了 SVG 和 Canvas，支持通过拖动图表改变数据的需求，因此采用 ECharts 作为系统可视化渲染的主要工具。而 Vis 对拓扑图绘制有很好的支持度，因此选择它作为流域拓扑图绘制及可视化展示的主要工具。对于简单数据的排列展示，则采用原生的 HTML/CSS 进行轻量化展示。

### 2.2.2 网络地理信息系统

网络地理信息系统 (WebGIS) 指的是网页端的地理信息系统 (GIS)，简单来说就是将地理空间信息以地图的方式在网页端展现出来，实现可视化和空间分析功能。为了更好的

实现对临洮县内的洮河流域进行孪生构建，需要采用支持三维地图和三维模型渲染的 WebGIS 引擎。目前，支持针对不同应用场景进行三维渲染和二次开发的 WebGIS 框架主要包括 Cesium 和 Mapbox 两种，其特点对比如表 2-3 所示。

表 2-3 WebGIS 框架特点对比

Tab.2-3 Comparison of WebGIS framework features

	Cesium	Mapbox
3D 地图支持	适用于构建虚拟地球和三维地理信息应用	整体而言在 3D 方面的特性不如 Cesium
数据源和图层	支持多种数据源，包括 3D Tiles、GeoJSON、KML 等	提供自家的矢量地图瓦片服务，支持添加用户自定义的矢量和栅格图层
性能	适用于处理大规模、复杂的 3D 地理信息数据。	在处理大规模 3D 数据时，可能相对 Cesium 稍显不足
应用场景	适用于需要强调全球 3D 可视化的应用	适用于强调地图样式、个性化地图的应用

考虑到本系统对三维孪生场景的需求，Cesium 相比于 Mapbox 在 3D 地图支持度更高，可以实现交互式 Web 开发，达到动态地理信息数据的共享，且具有更强的三维渲染能力，更适应于构建三维孪生场景，因此，本文选择 Cesium 作为渲染孪生流域的基座。

### 2.2.3 大模型技术

大模型（LLM）<sup>[42-43]</sup>是“人工智能预训练大语言模型”的简称，指的是包含数千亿甚至更多参数的语言模型。通过模拟人类的创作过程，使模型学习海量文本数据，从而实现生成具有实用价值的文本、图像、音频等内容。在本文的系统设计中，LLM 的作用在于对临洮县洪涝灾害业务泛在知识的采集和梳理，构建支持人机交互的一体化孪生平台，提高流域监管与调度决策的便利性，实现各类水利业务流转的自动化与智能化，减少业务人员对于新技术的学习成本和日常的重复性工作。

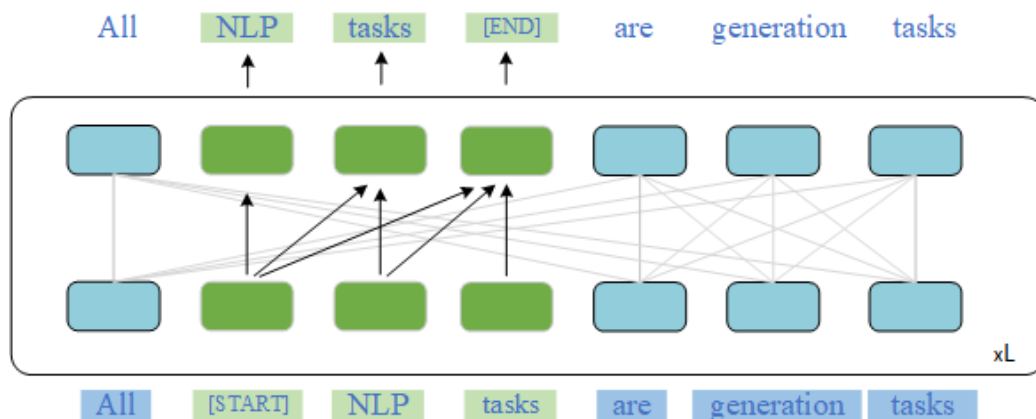


图 2-2 GLM 模型结构

Fig.2-2 GLM model structure

随着技术的发展，市面上涌现出大量的大模型，包括 GPT、Llama、ChatGLM。考虑到洪涝灾害防御系统的特定应用场景，需要大模型对中文有较好的支持能力，同时为减少开发成本，需要对语义理解及函数调用有较好的支持。而 ChatGLM4 是一种基于 GLM-4 的

大模型，如图 2-2 所示，采用了清华大学 KEG 实验室提出的 GLM 模型结构，能够处理大量的自然语言文本数据，对中英文都有较强的支持度，且在性能上与 GPT4 相近。于是本文选择采用双语模型 ChatGLM4 作为系统开发的核心大模型，以提供精准的语义分析和模型驱动功能。

#### 2.2.4 虚拟建模技术

三维地图引擎能够实现流域底图的孪生展示，并提供一些基础图形的绘制功能，如墙体、曲面等。但在业务场景中涉及水库、防护区等模型时，为了实现较好的孪生效果，需要采用虚拟实体建模技术对其进行建模处理。虚拟建模技术是实现孪生可视化的重要组成部分，是一种使用计算机生成的模型来模拟和呈现现实世界或想象的场景、对象或系统的技术。其优点在于真实感较强，能够通过构建三维模型实体来实现对现实环境中物体的映射。常见的虚拟建模技术主要包括三维软件建模和倾斜摄影建模。

三维软件建模主要应用于游戏或工业设计领域，具有高建模精度和高自由度的特点。能够实现较为精细的实体映射。但这类方法的建模流程相对较复杂，学习成本高，时间花费长，因此不太适用于大规模的建模。倾斜摄影建模主要用于建设施工、地图制作等场景，适用于大范围的建模区域，学习成本较低。可以真实还原地表建筑物的特征，主要应用于构建面模型，但其交互感较差。

对于本系统而言，临洮县内水库数量较少，但对水库的建模仿真需求较高，因此水库的建模可采用三维软件建模的方式。而对于防护区来说，模型范围较大，人工建模的工作量和时间成本过高，因此采用倾斜摄影的方式进行建模。

### 2.3 系统架构设计

本节主要基于前文对系统的需求分析以及技术系统的研究，对洪涝灾害防御系统的功能架构和技术架构进行研究设计。

#### 2.3.1 系统功能架构设计

系统的功能体系结构可分为三个层次，即业务应用层、孪生模拟层和语言大模型层，如图 2-3 所示。业务应用主要围绕洪涝灾害防御的“四预”展开。一方面，它负责对系统数据进行管理和展示；另一方面，在“四预”功能的支持下，实现对流域信息的实时监控、洪水预报、优化调度等，以提升洪涝灾害防御的治理保护能力。孪生模拟层主要是围绕业务对模型算力、算据和孪生场景的模拟提供支持，提高系统的可信度。语言大模型层则通过语言交流的形式实现业务文本类知识的查询，调取专业模型、调取业务场景等功能，改变业务系统的应用模式，以降低系统的学习成本和操作难度，提高系统的便利性。

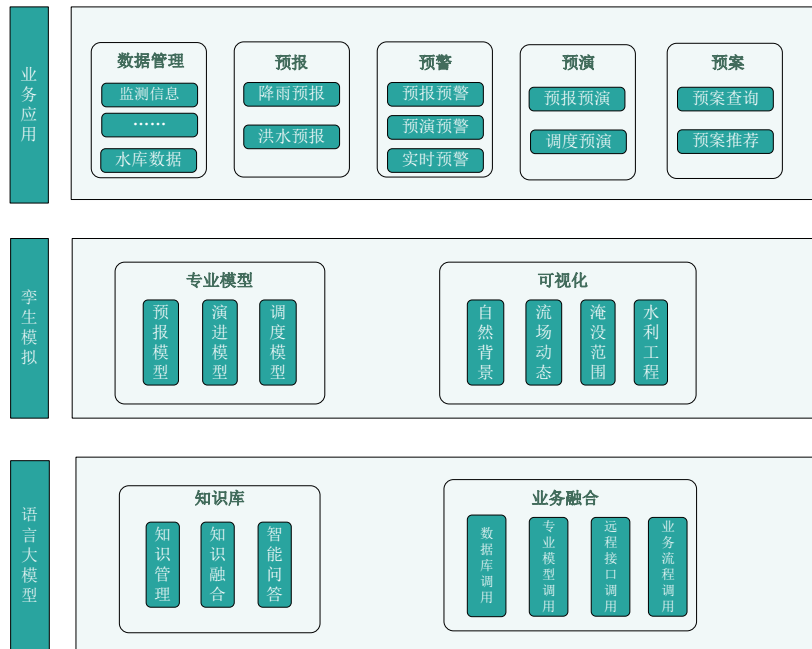


图 2-3 系统功能架构图

Fig.2-3 System function architecture diagram

### 2.3.2 系统技术架构设计

系统技术架构设计是融合多项技术进行系统开发的重要步骤，有效保证了系统开发的稳定性和可用性。基于需求分析和技术体系研究的结果，系统的技术架构可以分为前端、服务端、专业模型端和大模型端。具体而言，前端主要涉及数据的可视化展示、孪生场景的展示以及三维模型的呈现；服务端主要包括数据的存储、传输、业务逻辑计算以及调用专业模型端的算法；专业模型端主要涉及系统业务中模型算法的计算；而大模型端则通过语义交流的方式调用业务方法。系统的技术架构图 2-4 所示。

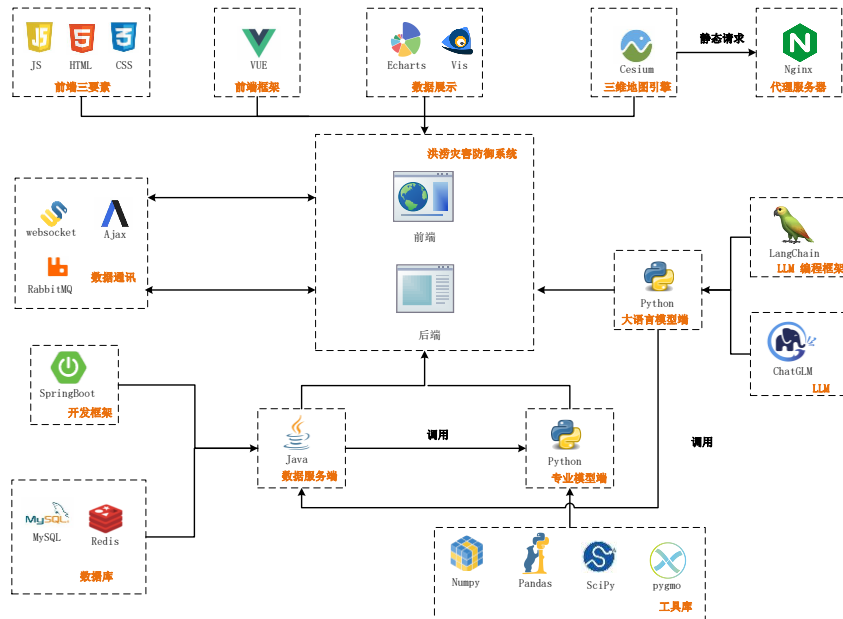


图 2-4 系统技术架构图

Fig.2-4 System technical architecture diagram

## 2.4 业务功能设计

本节在洪涝灾害防御业务需求分析和系统架构的基础上，对系统的业务进行研究。首先，对整体业务进行划分，随后对这些业务功能展开详细设计分析，并对核心流程进行合理规划。

### 2.4.1 系统业务划分

洪涝灾害防御系统的业务涵盖面广泛且复杂。因此，本节按照洪涝防御“四预”的流程对系统功能进行了划分。运用组件化思想，将业务划分为数据管理和洪涝防御“四预”服务两个主要业务模块。这些业务模块又涵盖各自的多个分支的业务逻辑，业务总体划分如图 2-5 所示。

数据管理主要用于管理流域站点、水库、知识库等关键信息。预报服务功能包括降雨预报和洪水预报，降雨预报依赖气象服务器提供的接口，对降雨情况进行预报；洪水预报则基于预报模型，对流域关键断面进行预报计算。灾害预警模块主要根据监测和计算数据与警戒线的大小关系，对产生的突发情况向相关人员进行预警通知。预演服务模块侧重于针对降雨引发的洪水和调度产生的下泄流量，去分析其对防护区的影响，提供计算结果和场景模拟。预案服务模块功能侧重于针对洪水情况或预演结果，为特定断面提供预案查询和方案推荐服务。

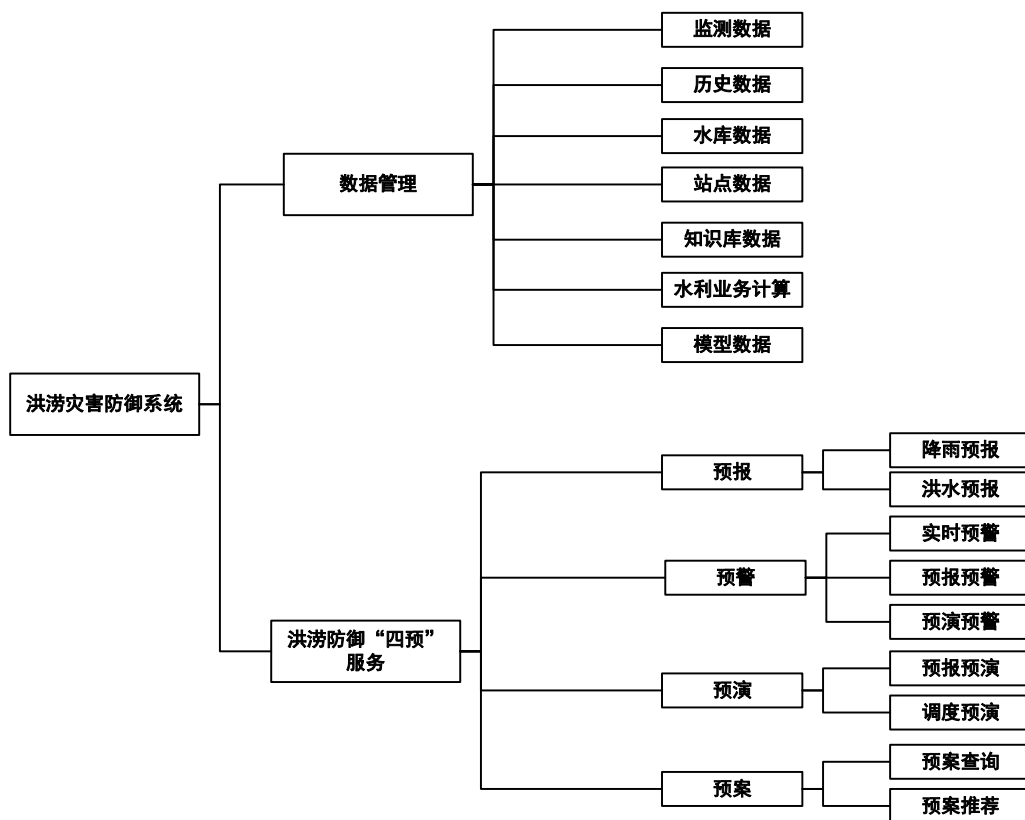


图 2-5 业务划分结构

Fig.2-5 Business division structure

### 2.4.2 数据管理服务

数据管理在洪涝灾害防御系统中扮演基础数据服务提供者的角色，利用统计的数据进行计算和展示。其业务范围主要包括监测数据管理、历史数据管理、水库数据管理、站点数据管理、模型组件数据管理、知识库数据管理以及水利业务计算。业务执行形式涵盖传统式的菜单点击和大模型问答两种方式。以下是相应业务功能的详细描述：

1) 监测数据管理。监测数据包括水库、水文站以及雨量站的监测数据。水库监测数据管理用于查询水库的监控视频、入库流量过程、出库流量过程以及水位库容变化过程。水文站监测数据管理用于查询水文站监测的河道断面的水位、流量变化过程。雨量站监测数据管理则是用于查询雨量站监测的降雨量的变化过程。

2) 历史数据管理。包括对历史预报方案及产生结果的管理，以及对历史预演结果的管理，以便业务人员能够参考历史方案应对类似情况。

3) 水库数据管理。用于管理洮河流域内水库的基本情况，包括水库的名称、位置、管理单位等信息。

4) 站点数据管理。用于管理洮河流域内水文站和流量站等站点的名称、经纬度以及其他基本信息，实现在三维地图上的展示。

5) 模型组件数据管理。用于管理系统内预报、演进、调度等模型组件的基本信息，包括模型组件的名称、存储位置等。

6) 知识库数据管理。主要提供知识库数据的分类以及新增、修改、查看、删除和导出业务。其数据类型主要为预案、方案等文件，上传后通过文本分割以及向量化处理将其存储于向量数据库中。

7) 水利业务计算。主要对系统内的基础信息提供计算服务，包括但不限于水库时段入库流量计算、雨量站时段降雨量计算。

在 Java 后端，对以上业务通过面向对象的形式将数据类型抽象为实体类 (Entity)，实体类包括水库基本情况、监测信息、站点信息等基础信息的属性构建。功能类则通过编写对应实体类的实现类 (ServiceImpl) 来实现。

### 2.4.3 洪涝防御“四预”服务

洪涝防御“四预”业务主要包括预报、预警、预演及预案。通过集成雨情、水情等信息，以洪水预报、水库优化调度、洪水演进模拟等模型为基础，实现流域雨水汛情的动态监控、宏观模拟以及模型算法的统一管理。按照“降雨-产流-汇流-演进”<sup>[44]</sup>、“支流-干流-断面”、“来水-过程-防护-调度”、“洪量-对象-方案-组织”四个关键链条，结合洪涝“四预”四个环节，运用大模型、数字孪生、组件化等技术方法，模拟分析河流断面的洪水过程以及关键防护对象的淹没情况，实现交互调整调度方案优选以及预案内容智能推荐。

#### (1) 预报服务

洪涝的预报功能包括降雨预报和洪水预报。其中洪水预报又分为流域洪水实时滚动预

报和针对特点断面的洪水预报。

在降雨预报方面，系统通过接入气象网提供的服务接口，输入位置和时间，获取未来降雨预报数据。随后，利用前端可视化工具对数据进行渲染展示，以获得降雨变化趋势。

流域洪水实时滚动预报则通过整合雨情、区域等信息，将临洮县内的流域划分为多个子流域。通过产汇流模型计算每个子流域的流量过程，将得到的流量过程传入演进模型进行洪水演进模拟。此过程中，通过观察支流向干流洪水演进的情况，更新洮河干流的洪水过程。同时，通过设定滚动频率，系统在每个滚动节点结束后利用最新实测的洪水过程进行修正，以实现滚动预报的效果。通过孪生流域漫游的形式，进行滚动预报业务，实时滚动预报业务的操作流程见图 2-6。

针对特定断面的洪水预报，首先利用断面上游的实测洪水数据，通过演进模型将其演进至目标断面，获取目标断面的实测洪水过程。然后，利用预报模型在此实测洪水的基础上，对后续洪水过程进行预报。随着实测数据的更新，预报模型会利用最新的数据进行重新计算，以修正预报过程。

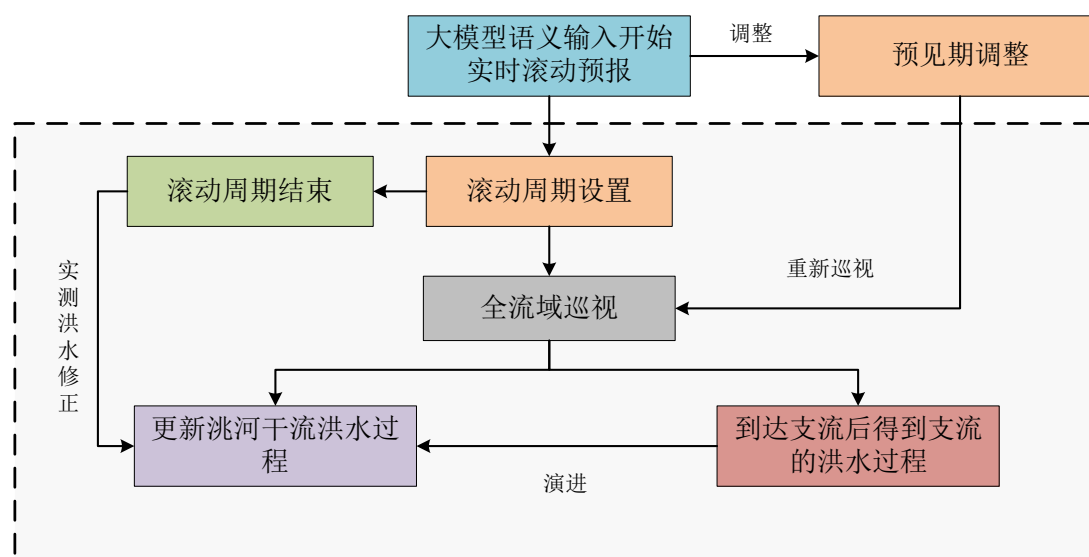


图 2-6 滚动预报业务流程

Fig.2-6 Rolling forecast business process

预报业务流程在系统中有两种使用方式：一是采用传统的菜单点击方式进行预报业务，二是通过与大模型进行交流，利用语义识别的方式调取 Java 后端的接口，Java 后端再调取 Python 端的模型算法，获取算法结果后进行前端孪生模拟及结果展示。

为了更好地实现预报业务的模块化，业务被抽象成 Java 实体类和功能类，其抽象模型包括：1) 降雨预报结果类，其属性包括地理位置坐标、时间以及降雨量等属性。2) 洪水预报模型算法实体类，包括模型算法所需的参数。3) 模型算法输入的抽象类，属性为包括预报模型的共有的输入属性，雨情信息，预见期，经纬度信息等。4) 模型算法输入的抽象实现类型，继承自模型算法的输入抽象类，新增了特定算法的输入参数。5) 计算结果的输出实体类，包括洪水过程、洪峰流量和洪峰时间等属性。6) 流程功能类，主要用于对输入



数据的处理以及调取 Python 模型，以实现业务流程。

### (2) 预警服务

预警服务包括实时预警、预报预警和预演预警。实时预警主要针对雨量站、水文站等位置的监测数据进行实时监测。系统结合监测站点的监测指标数据，自动判断是否存在超过阈值的情况。一旦监测的数据超过设定的阈值，系统将在三维地图上以警告闪烁的方式标示，并通过短信通知相应位置的负责人。用户点击查看时，系统将展示数据的历史变化过程以及阈值红线。

预报预警指的是系统结合河道、堤防等工程指标数据，当进行预报计算时，判断支流、断面的超警超限情况，同时结合防御能力进行判别分析。如果预报得到的洪水过程超过设定的红线，系统将在页面上显示预警信息，同时进行语言警告，并通过短信形式通知相关负责人。

预演预警主要是在预演环节中，针对预演的结果给出预警。具体而言，如果预演得到的洪水频率过大，导致的淹没范围对防护区域产生影响，业务人员可根据此次预演情况，主动向防护对象区域的负责人发送警告通知。整个预警流程如图 2-7 所示。

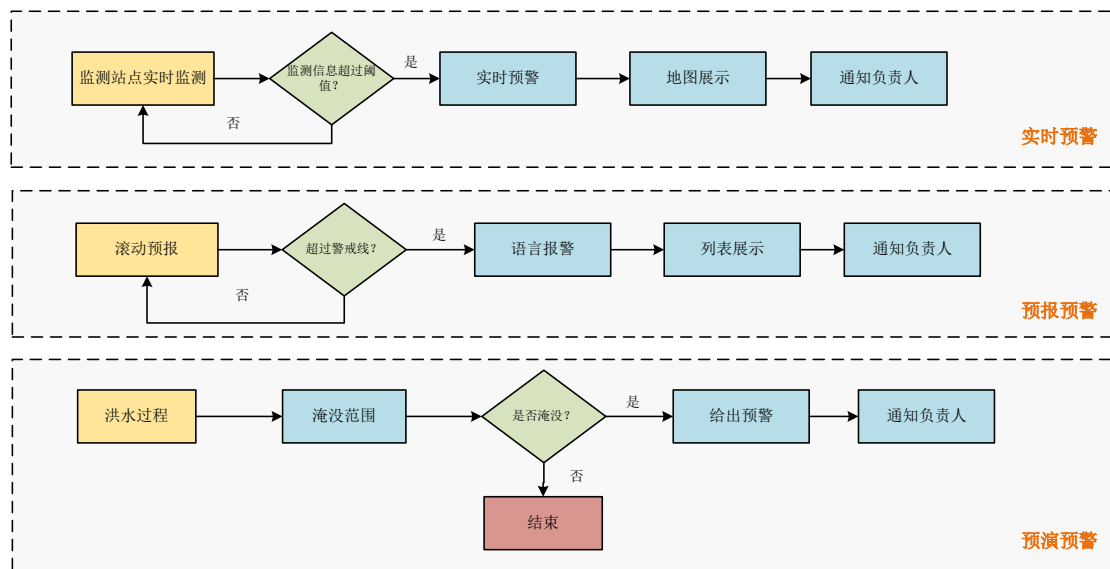


图 2-7 预警业务流程图

Fig.2-7 Early warning service flow chart

将预警服务抽象为 Java 实体类和功能类，其抽象模型包括：1) 预警对象实体类，包括预警对象类型、预警指标、预警警戒线范围、备注等属性。2) 通知对象实体类，包括预警类型、预警对象、预警信息、通知人、负责人、备注等属性。3) 业务功能类型，主要实现是否达到警戒线的判断，并与负责人进行通讯。

### (3) 预演服务

基于可视化平台和流域数字化场景，运用三维可视化技术，实现临洮县内洮河流域的数字化场景构建。将预演服务划分为预报预演和调度预演两类，其中预报预演中包括了部分调度预演业务。其业务流程执行方式同样为两种，传统的菜单式和大模型驱动式。

预报预演业务流程如图 2-8 所示，对于大模型驱动式（传统菜单式的同理）的业务流程，描述如下：

- 1) 业务人员向大模型输入类似“某河流某时段某场降雨对某防护区产生的影响”等问题，其中包含了具体的位置信息、降雨信息和防护对象信息。
- 2) 三维引擎按照起点和终点、支流和干流的顺序对暴雨中心和防护对象过程中的所有关键断面进行漫游。
- 3) 到达每个断面时，系统会提供此次降雨对该断面产生影响，在计算的过程中，模型算法及参数都可以通过人工交互进行调整。
- 4) 如果在中间过程中涉及到水库，到达水库断面时，系统会给出入库流量，用户可以选择调度方式，得到相应的出库流量过程，同样，算法参数和出库流量过程也可通过人工交互进行调整。
- 5) 到达防护对象时，基于计算得到的洪水过程，系统给出洪水的三要素以及淹没范围。
- 6) 根据淹没范围，系统给出此次降雨对防护对象区域的影响。

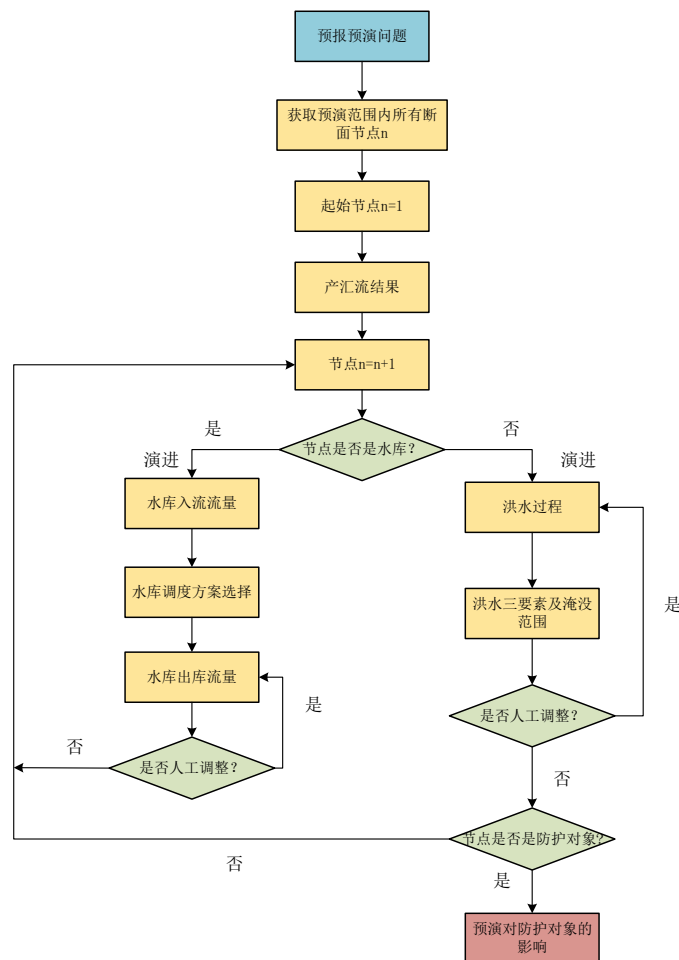


图 2-8 预报预演业务流程图

Fig. 2-8 Forecast preview business flow chart

调度预演主要是模拟不同调度方案下水库调度对下游的影响，大模型驱动式（传统菜

单式类似)的业务流程描述如下:

1) 向大模型输入希望进行调度预演的水库信息。

2) 设定水库的入库流量过程以及相关的防护对象,调整水库的调度方案。

3) 孪生漫游的形式,对水库下泄流量进行演进计算,得到关键断面初始流量与水库下泄流量叠加后流量过程。

4) 到达防护区时,给出此次水库下泄方案对防护对象的影响。

#### (4) 预案服务

在预案方面,充分利用已有的洪水防洪预案、水利工程运行方案、调度业务规则和其他相关方案等已有知识,建立了大模型驱动的行业知识库。业务应用场景主要包括两个方面:一是在预警时或预演结束后,根据洪水过程及目标位置,系统能够智能地从预案知识库中推荐合适的预案,并提供解决方案。二是业务人员主动向大模型询问知识库中的内容时,系统能够提供相应的答案,并给出文件的来源。

## 2.5 非业务功能设计

洪涝灾害防御系统除了核心业务功能之外,还需要包含必要的非业务功能,以确保系统的安全及稳定。

### 2.5.1 用户权限设计

权限校验在系统中扮演着非常关键的角色,尤其是涉及大量数据和对部分数据需要保密的情况下。为了确保系统的稳定性和安全性,对用户进行分区与隔离,并进行权限分配是至关重要的。系统要求用户通过认证后方可进入并使用,不同权限的用户只能操作其权限范围内的功能。

#### (1) 功能权限

本系统选择了 Spring Security+OAuth2.0 作为洪涝灾害防御系统的安全框架,用于用户的认证和授权。认证过程涉及验证用户是否具有访问系统的权限,而授权则确认用户能够访问的资源范围。具体流程包括用户输入用户名和密码,系统利用 BCrypt<sup>[45]</sup>算法对密码进行加密,然后将加密后的用户输入密码与数据库存储的密码进行比较,只有在用户没有被禁止访问且密码匹配的情况下,系统才允许用户进入,并展示用户权限下可操作的功能页面。

#### (2) 数据权限

数据权限涉及不同角色拥有的数据范围。系统根据实际应用环境,按照用户部门职级别权限由大到小的顺序,将部门数据分为全公司、一级部门、二级部门以及特别个体四个层级。在所有需要数据权限控制的表中,引入部门 ID 字段,并利用 MyBatis 的拦截功能根据部门 ID 下的权限进行数据筛选。

为满足系统对数据权限的需求,采用对不同用户分配不同角色的方式,基于角色的差异,对系统数据和内容进行控制,建立项目-角色-权限的数据管理模式。用户的操作通常涉

及新增、查询、删除、编辑四种，这些操作对应于前端的路由和后端的接口访问。前端利用 Vue 框架，根据用户权限呈现可访问的页面和操作按钮。后端通过整合 Spring Security，通过 @PreAuthorize 注解对函数接口的访问权限进行控制，前后端共同鉴权，确保系统的安全性。

### 2.5.2 数据库设计

数据库设计是实现数据持久化的一个非常关键的环节。在本文系统中，涉及的数据可以概括为实时监测数据、历史数据以及业务数据等。目前，数据库按照数据结构可分为关系型数据库（RDBMS）和非关系型数据库（NoSQL）两种<sup>[46]</sup>。在关系型数据库中，常用的包括 MySQL、SQL Server 和 Oracle 等。MySQL 是开源软件，提供免费的社区版，具有轻量级的特性，适用于需要高性能和较小资源开销的应用。因此，在本系统中选择了 MySQL 作为关系型数据库。而对于常用的业务数据，可以采用非关系型数据库进行存储。为了提高系统的运行效率，选择 Redis 作为缓存数据库，用于缓存需要经常查询而不经常更新的数据。

#### （1）数据库逻辑设计

鉴于系统内涉及大量关键数据，为了追踪业务人员的操作历史，包括创建、删除、修改等行为，数据库需要记录操作痕迹。此外，为了保留系统中的初始数据，系统的数据库选择以逻辑删除的方式，在重要数据表的结构中添加通用字段，如表 2-4 所示。

表 2-4 数据库表结构通用字段设计

Tab. 2-4 Database table structure common field design

字段名称	字段类型	允许为空	字段说明
create_by	varchar	否	创建人
update_by	varchar	否	修改人
delete_by	varchar	是	删除人
create_time	datetime	否	创建时间
update_time	datetime	否	修改时间
delete_time	datetime	是	删除时间
del_flag	char	否	删除标识

系统中每个服务功能都建立了独立的表结构，通过关联表结构主键的方式来建立表与表之间的关系。鉴于系统的性质，系统需要频繁查询降雨量、河道流量等信息。为了优化对经常查询的实时数据的访问，系统采用了设置多个数据索引的策略，以使用户能够更有效地进行查询工作，例如水文站编码、雨量站编码等。

#### （2）缓存数据库设计

在面对高并发数据读写时，使用基于 InnoDB 存储引擎的 MySQL 数据库可能会遇到性能瓶颈<sup>[47]</sup>。由于洪涝灾害防御系统涉及的数据量庞大，随着数据和访问强度的增加，数据库的压力也会相应增加。在频繁的读写压力下，数据库的工作效率难以保证。为了解决这一问题，系统需要建立数据缓存，减少额外的数据库读写操作。

本文选择了 Redis 作为系统的缓存数据库，并采用了 Spring 框架中的 Spring Cache，以支持系统对数据的缓存需求。系统利用 Spring Cache 提供的注解，以面向切片的方式来最大限度地降低代码耦合性，从而实现系统数据的缓存操作，缓存数据执行流程如图 2-9 所示。

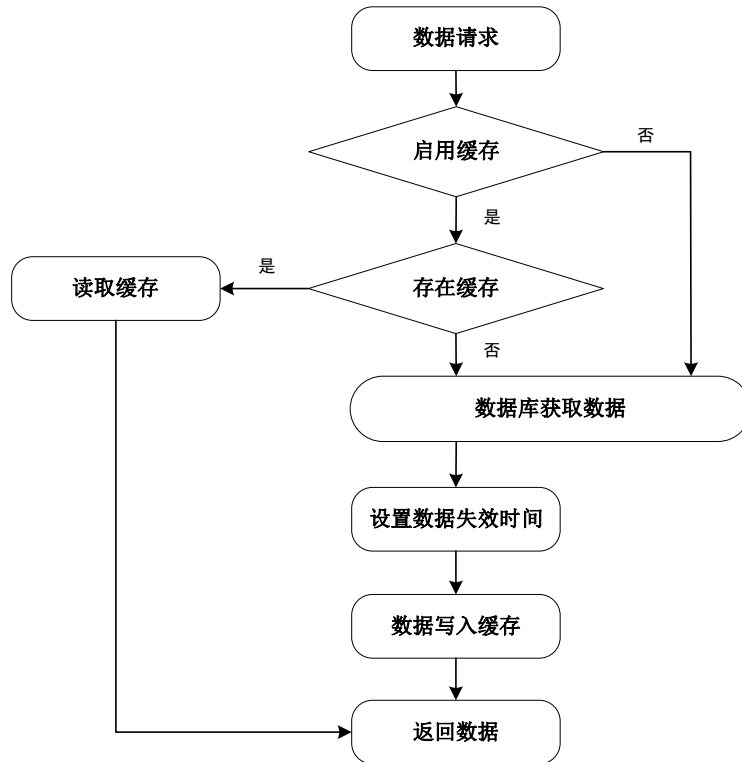


图 2-9 缓存数据执行流程

Fig. 2-9 Cache data execution flow

在执行数据查询操作时，系统首先会判断是否开启了缓存功能。若启用了缓存，则会在缓存中查询是否存在对应的数据。如果存在，则直接从缓存中读取数据并返回给用户；如果缓存中没有对应的数据，则会进入数据库进行相应的数据查询，将查询的结果保存到缓存数据库中，并将数据返回给用户。通过以上操作，系统能够大幅度减轻数据库的访问压力，从而提高数据的访问效率。

## 2.6 本章小结

本章旨在探讨基于大模型驱动的洪涝灾害数字化系统的架构设计。首先，对洪涝灾害防御系统的应用需求进行了分析，在需求分析的基础上，进行了对数据可视化、网络地理信息系统等技术体系的分析和选择，以明确系统的功能架构和技术架构，确定系统的总体设计。最后，针对系统的业务和非业务功能进行了分析设计，为系统的开发和功能实现提供了支撑。

### 3 洪涝灾害防御专业模型集成研究

洪涝灾害防御系统的核心在于处理洪水的“四预”业务，涉及的模型主要包括水文预报模型、洪水演进模型以及水库防洪调度模型，因此，本章主要以多模型集成的角度分析这些模型的集成关系，并依据多模型集成体系，对专业模型进行组件化拆分，以实现模型的集成开发。

#### 3.1 多模型集成体系

目前，应用于洪涝灾害防御的专业模型众多，包括洪水预报、洪水演进、防洪调度模型等。以洪水预报模型为例，每个模型都具有自身的适应条件，即使在同一流域，不同的下垫面条件也会导致模型精度的差异，若每个模型都独立开发，将会导致系统代码冗余、复用率低等问题。为了解决这一问题，本章提出了多模型集成体系，如图 3-1 所示。在模型应用中，根据每个模型的特点，可将其划分为多个基础模块。以基础模块为最小粒度，进行组件化处理，并建立模型方法组件库。通过标准化接口约束基础模块组件的输入和输出，按照业务逻辑关键将各个基础模块进行集成，构建成中粒度的集成模型，形成模型方法数据库，以便模型的使用与管理。

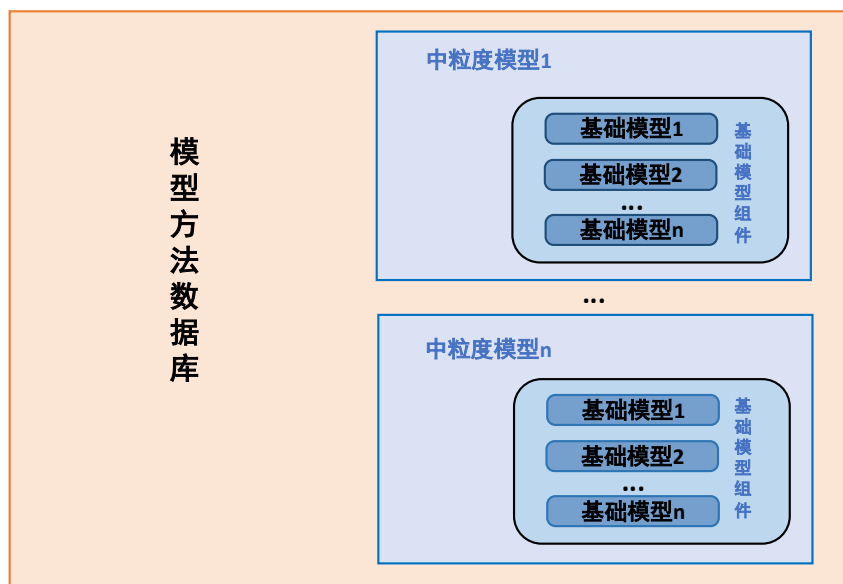


图 3-1 集成体系示意

Fig. 3-1 Integrated architecture schematic

#### 3.2 洪水预报模型

##### 3.2.1 新安江模型

新安江模型是一种用于洪水预报和径流模拟的集总式水文模型。20 世纪 70 年代，赵人俊等经过多年研究提出了新安江模型，是我国第一个完整的流域水文模型，主要应用于湿

润和半湿润地区<sup>[48-49]</sup>。新安江模型的整体流程如图 3-2 所示，由蒸散发计算、水源划分、产流计算、汇流计算四个部分组成，以实测降雨  $P$  和潜在蒸发能力  $EM$  为输入，流域出口断面总流量  $Q$  和流域蒸散发  $E$  为输出。模型主要涉及的参数及其取值范围见表 3-1。

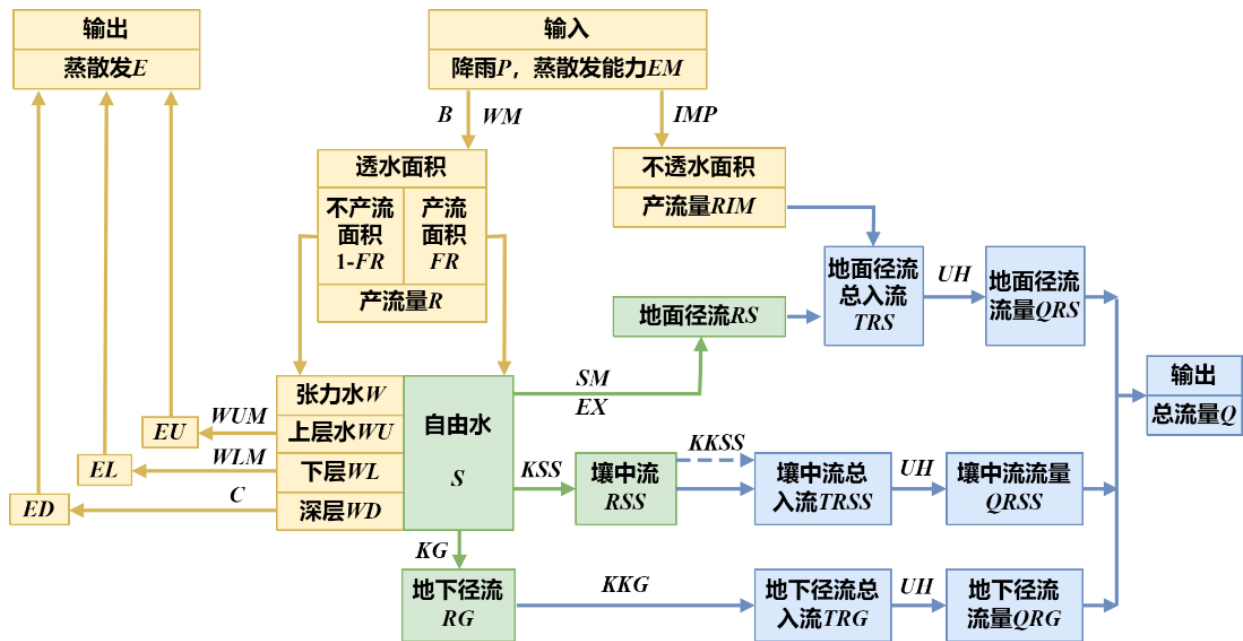


图 3-2 新安江（三水源）模型原理图

Fig. 3-2 Flowchart of Xin'anjiang model (three sources)

表 3-1 新安江模型的参数及取值范围<sup>[50-52]</sup>

Tab. 3-1 Parameters and their sampled ranges of Xin'anjiang model

参数类别	参数含义	取值范围
蒸散发计算	K (蒸散发能力折算系数)	0.2~1.5
	WUM (上层张力水容量)	5~20
	WLM (下层张力水容量)	60~90
	C (深层蒸散发系数)	0.08~0.18
产流计算	WM (流域平均蓄水容量)	80~170
	B (蓄水容量曲线的指数)	0.1~0.4
	IMP (流域不透水面积比)	0.01~0.05
	SM (自由水蓄水容量)	10~50
分水源计算	EX (自由水蓄水容量曲线指数)	0.5~2
	KG (地下水出流系数)	0.01~0.70
	KSS (壤中流出流系数)	0.01~0.70
	KKSS (壤中流消退系数)	0.5~0.9
汇流计算	KKG (地下径流消退系数)	0.950~0.998
	KE (单位河段马斯京根参数 K)	/
	XE (单位河段马斯京根参数 X)	/
	UH (单元流域上的地面径流单位线)	/

### 3.2.2 GR4J 水文模型

GR4J 是一种集总式概念性水文模型，由法国学者 Perrin<sup>[53]</sup>等于 2003 年在 GR3J 模型的基础上进行改进得到。通过使用两个非线性水库进行产汇流计算，其中第一个水库为产流水库，第二个水库为汇流水库，模型的结构如图 3-3 所示。模型主要包括  $x_1$ 、 $x_2$ 、 $x_3$ 、 $x_4$  四个参数，根据 Perrin 多年资料的验证<sup>[54]</sup>，GR4J 参数 80% 的概率置信区间如表 3-2 所示。

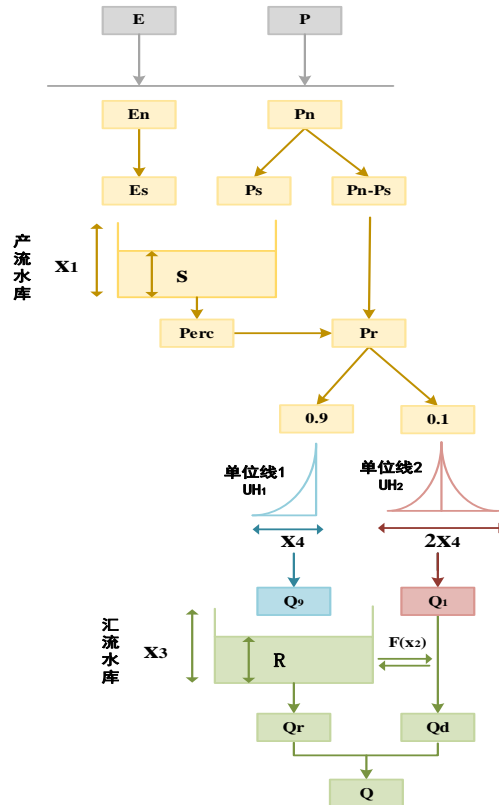


图 3-3 GR4J 模型结构

Fig. 3-3 Model structure of the GR4J

表 3-2 GR4J 参数的 80% 置信区间

Tab. 3-2 The 80% confidence interval for the GR4J parameter

参数类别	参数含义	中间值	范围区间
$x_1/mm$	产流水库最大蓄水量	350	100~1200
$x_2/mm$	地下水交换系数	0	-5~3
$x_3/mm$	汇流水库最大蓄水量	90	20~300
$x_4/d$	单位线汇流时间	1.7	1.1~2.9

### 3.2.3 预报模型组件化

根据水源划分方式，新安江模型主要包括二水源新安江模型和三水源新安江模型两种类型，均包括蒸散发计算、产流计算、水源划分和汇流计算四个阶段。而 GR4J 水文模型主要包括产流阶段和汇流阶段。因此，可按照模型结构和逻辑中的共性，将预报模型组件库可划分为蒸散发计算、产流计算、水源划分和汇流计算四大类。

新安江模型的内部结构较为复杂，采用了三层蒸散发模式，将下垫面分为上层、下层



和深层，以满足不同蒸散发需求；产流阶段根据降水和田间持水度的关系以蓄满产流的形式产生径流，结束后进行水源划分；河网汇流一般采用线性水库进行计算，而河道汇流则通常采用马斯京根模型。相比之下，GR4J 水文模型较为简单，主要采用一层蒸发模式，以蓄满产流的形式使用非线性水库计算产流量，汇流阶段则采用时段单位线进行计算，鉴于不同径流成分的差异，产流量中的 90%采用单位线 UH1 演算，10%采用单位线 UH2 演算，前者需要经过汇流水库的再次调节，后者直接汇集到流域出口断面<sup>[55]</sup>。

基于以上分析，系统的预报模型组件库主要可分为蒸散发计算组件、产流计算组件、水源划分组件、坡地汇流组件和河网汇流组件。蒸散发计算组件可细分为一层、二层和三层蒸散发计算；产流计算组件主要为蓄满产流；水源划分组件可划分为二水源和三水源划分；坡地汇流组件包括单位线汇流、线性水库汇流和非线性水库汇流；河网汇流组件主要为马斯京根汇流。预报模型组件库的结构如图 3-4 所示。

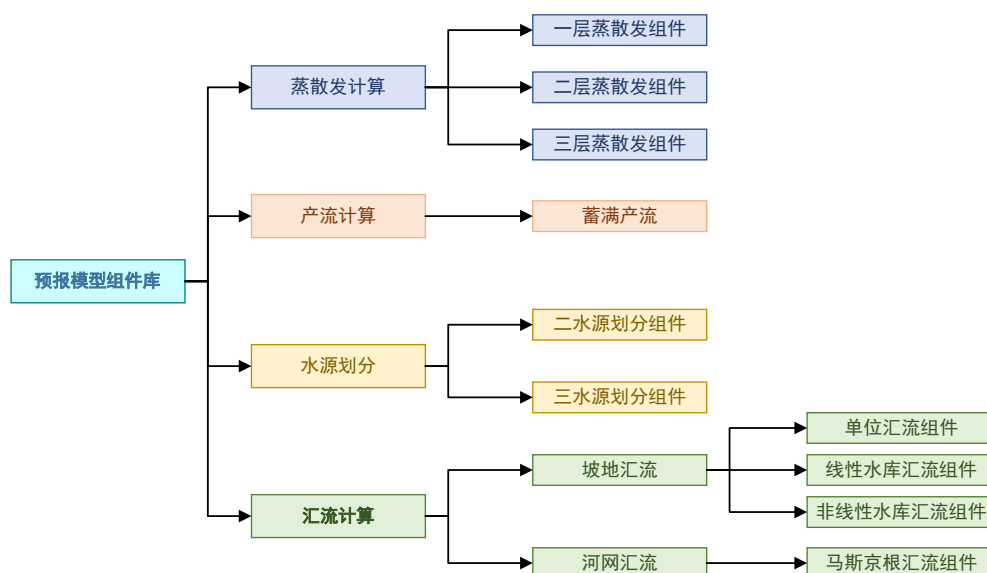


图 3-4 预报模型组件库结构

Fig. 3-4 Forecast model component library structure

### (1) 组件接口标准化

在模型计算过程中，数据传递构成了连接组件的纽带，因此需要确定组件和所需数据之间的对应关系，以规范化组件的输入和输出。为此，组件之间采用 JSON 实体作为数据接口，实体中包括开始状态、结束状态以及组件的输入输出格式。组件的开发方式遵循统一的开发标准，通过构建父类组件定义抽象方法，然后以继承父类组件的方式构建子类实例，从而实现具体方法。下面对预报模型组件库的五种组件进行介绍。

蒸散发计算模块由一层次、二层次和三层次蒸散发组件组成，这些组件的共同特点是处理流域内的降雨数据，并输出相应的蒸发量；产流计算模块采用蓄满产流方法，接收降雨序列数据作为输入，生成流域的径流深序列；水源划分组件接受出流系数等信息作为输入，出流信息作为输出；汇流计算模块包括坡地汇流和河网汇流两部分，其中坡地汇流又

分为单位线汇流、线性水库汇流和非线性水库汇流，其功能是将流域径流深序列转化为出口断面的流量；河网汇流则使用马斯京根法，其输入为出口断面流量，输出为河道断面的流量。组件接口的标准化的方式使得模型计算过程更加规范化，有助于提高模型的可重用性和易用性。

(2) 新安江模型组件化

按照新安江模型的原理，以三水源新安江模型为例，其组件构成如图 3-5 所示。

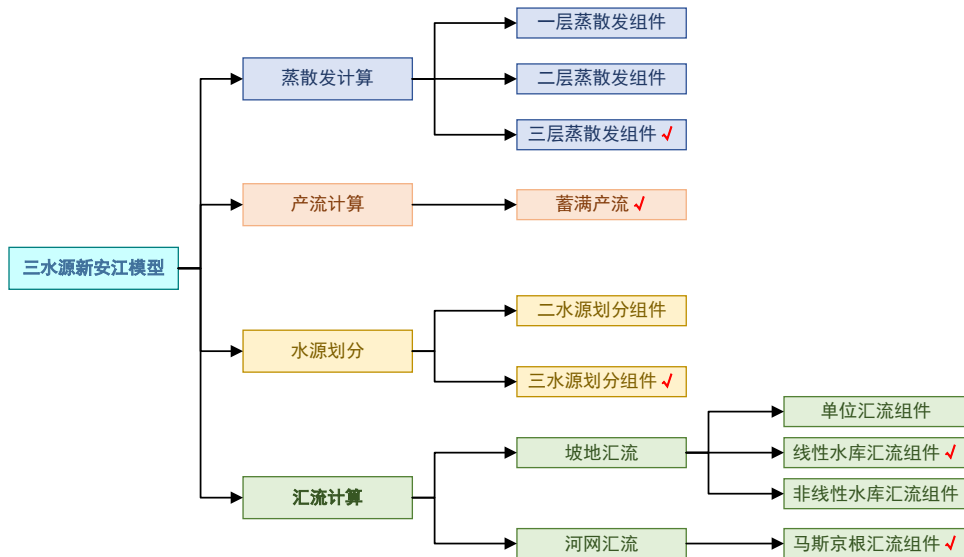


图 3-5 新安江（三水源）模型组件构成

Fig. 3-5 Composition of Xin'anjiang model component (three sources)

按照组件之间输入输出的需求关系，将新安江模型各个模块按照业务逻辑进行组装，如图 3-6 所示，通过上层组件向下层组件传递参数的方式，将整个模型进行串联构建。

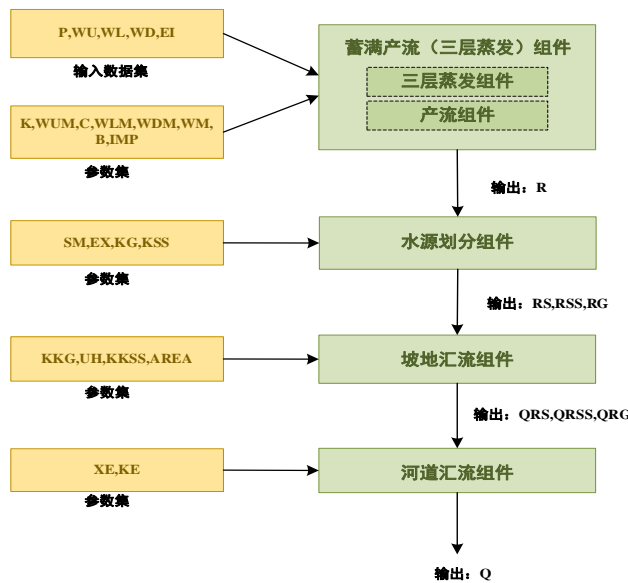


图 3-6 新安江模型（三水源）各组件关系

Fig. 3-6 Relationship between components of Xin'anjiang model (three water sources)

每个组件均采用 Python 技术进行独立开发和封装，通过标准化的接口实现组件间的便捷联通，新安江模型总共要实现八个实例组件，组件的名称、类名和功能如表 3-3 所示。

表 3-3 新安江模型各组件实现

Tab. 3-3 Implementation of each component of Xin'anjiang model

组件名称	类名	功能
一层蒸散发组件	one_evap.py	一层模式下蒸散发计算
二层蒸散发组件	two_evap.py	二层模式下蒸散发计算
三层蒸散发组件	three_evap.py	三层模式下蒸散发计算
蓄满产流组件	runoff_under.py	蓄满产流的方式计算产流
二水源划分组件	two_sour_div.py	进行二水源划分
三水源划分组件	three_sour_div.py	进行三水源划分
线性水库汇流组件	xxsk_flow.py	线性水库方式计算河网汇流量
马斯京根汇流组件	muskigum_flow.py	马斯京根方式计算河道汇流量

### (3) GR4J 水文模型组件化

GR4J 水文模型组件构成如图 3-7 所示，主要为一层蒸散发组件、蓄满产流组件、单位线汇流组件以及非线性水库汇流组件。通过向一层蒸散发组件输入有效降雨 $P$ 和蒸发能力 $E$ 获得蒸发量 $E_n$ 和产流水库降水量 $P_s$ ；随后将模型参数 $x_1$ 和降雨蒸发数据传入到产流组件得到产流结果 $P_r$ ；然后产流结果 $P_r$ 、参数 $x_4$ 和时间 $t$ 传入单位线汇流计算组件获得单位线汇流结果 $Q_9$ 和 $Q_1$ ；最后将单位线汇流结果和 $x_2$ 、 $x_3$ 传入非线性水库汇流组件得到流域出口断面总流量 $Q$ ，整个 GR4J 水文模型组件间逻辑关系如图 3-8 所示。

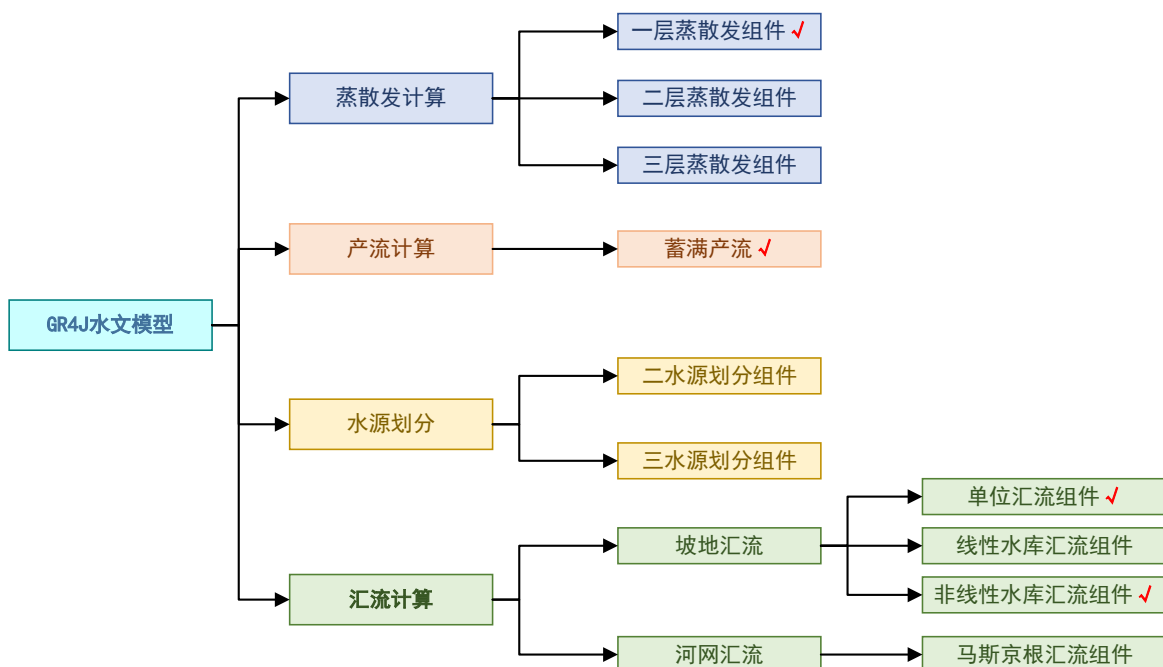


图 3-7 GR4J 水文模型组件构成

Fig. 3-7 Component composition of GR4J hydrologic model

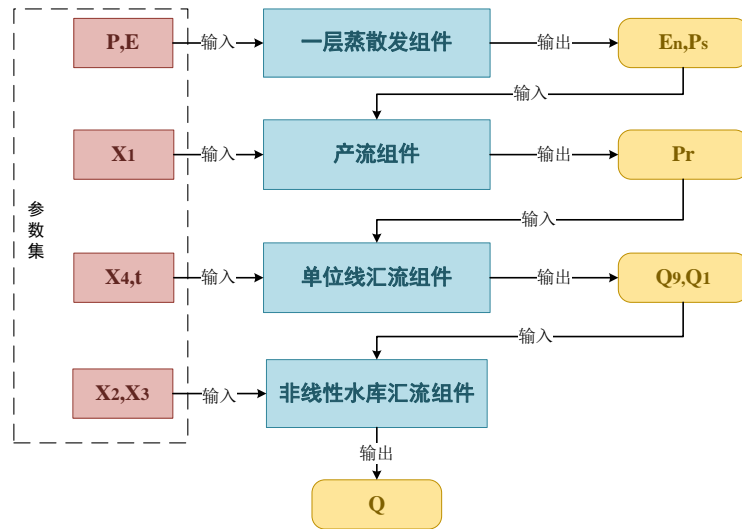


图 3-8 GR4J 模型各组件关系

Fig. 3-8 GR4J model component relationships

GR4J 水文模型总共要实现四个实例组件，其中蒸散发组件和产流组件在新安江模型组件里以介绍，表 3-4 主要展示额外的组件信息。

表 3-4 GR4J 水文模型各组件实现

Tab. 3-4 Implementation of each component of GR4J model

组件名称	类名	功能
单位线汇流组件	uh_flow.py	单位线方式计算汇流量
非线性水库汇流组件	un_xxsk_flow.py	非线性水库方式计算汇流量

### 3.3 参数优选算法

#### 3.3.1 参数优选方法

以水文模型参数优选为例，常用的自动寻优算法有很多，其中包括传统的方法如网格搜索法和牛顿法，这些算法的核心是通过比较目标函数值的增减来决定优化路径，进而逐步引导搜索至最优解。在处理线性和低维度模型时，这些传统方法表现较为成熟，取得了一定的成功。然而，在面对高度非线性和参数维度较高的水文模型时，它们往往无法取得令人满意的结果。为了提高效率和精度，研究者们逐渐引入了一些新颖的自动寻优算法。这些算法采用更灵活、智能的策略，能够更好地应对复杂的水文模型参数优选问题。

##### (1) NSGA-II 优化算法

目前 NSGA-II (Non-dominated Sorting Genetic Algorithm II) [56-58] 算法在水文模型参数优选取得了许多研究成果。NSGA-II 是一种多目标优化算法，是由 Deb [59] 于 2002 年在 NSGA 的基础上进行改进得到的，其特点包括采用非支配排序和拥挤度算子以有效评估和维护多个优秀解，实现均匀分布的 Pareto 前沿，同时具有简单的参数设置和高效的进化算子，使其在实践中广泛应用于多领域的问题求解，其算法流程如图 3-9 所示。

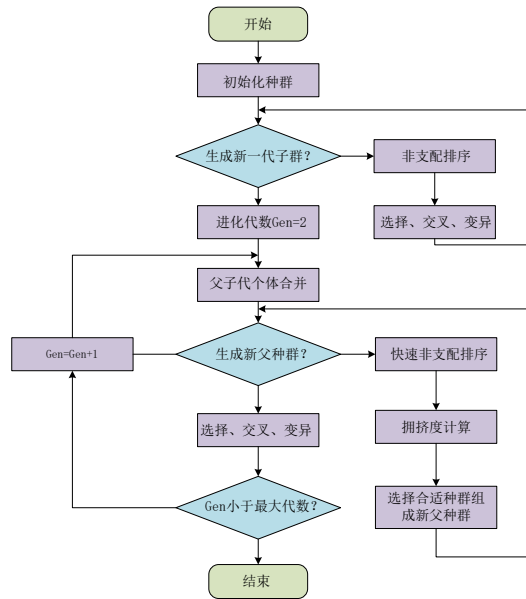


图 3-9 NSGA-II 算法流程图

Fig. 3-9 Flowchart of NSGA-II algorithm

(2) 粒子群算法

粒子群优化算法 (Particle Swarm Optimization Algorithm) 是由 Kennedy 和 Eberhart<sup>[60]</sup> 于 1995 年提出的一种集群智能优化算法, 灵感源自于观察鸟群寻找食物的过程。在 PSO 中, 候选解被表示为群体中的粒子, 每个粒子根据自身经验和邻居粒子的信息来调整自己的位置和速度。通过不断的迭代更新, 粒子向着目标函数值更优的方向移动, 并逐步收敛于全局最优解。PSO 算法具有简单易实现、全局寻优能力强的特点, 适用于解决各种优化问题。PSO 的算法流程如图 3-10 所示。

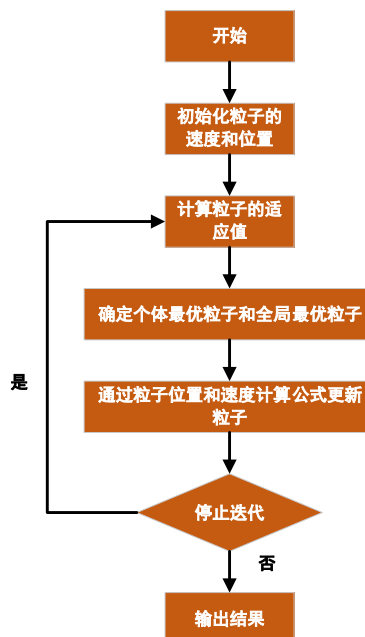


图 3-10 PSO 算法流程图

Fig. 3-10 Flowchart of PSO algorithm

### 3.3.2 精度评定标准

就水文预报而言，单目标的参数优化仅着眼于水文过程中的单一特性，无法充分挖掘其过程中不同特征，为了确保模型参数能够更全面的反应流域特征，本文选用以下四个目标作为洪水预报模型参数寻优的目标函数：

纳什效率系数（NSE）：

$$NSE = 1 - \frac{\sum_{i=1}^N (q_{ij,sim} - q_{ij,obs})^2}{\sum_{i=1}^N (q_{ij,obs} - \bar{q}_{i,obs})^2} \quad (3-1)$$

径流深误差（RRD）：

$$RRD = \frac{r_{i,sim} - r_{i,obs}}{r_{i,obs}} \times 100\% \quad (3-2)$$

洪峰出现时间误差（PKT）：

$$PKT = t_{i,sim}^{peak} - t_{i,obs}^{peak} \quad (3-3)$$

洪峰流量误差（RPE）：

$$RPE = \frac{q_{i,sim}^{peak} - q_{i,obs}^{peak}}{q_{i,obs}^{peak}} \times 100\% \quad (3-4)$$

式中，N 为洪水总数； $q_{ij,sim}$ 和 $q_{ij,obs}$ 分别为第*i*场洪水洪峰流量预测值和实际值； $\bar{q}_{i,obs}$ 为第*i*场洪水实测流量平均值； $r_{i,sim}$ 和 $r_{i,obs}$ 分别为第*i*场洪水的模拟径流深和实测径流深。 $t_{i,sim}^{peak}$ 和 $t_{i,obs}^{peak}$ 分别为第*i*场洪水洪峰出现时间的模拟值和实测值； $q_{i,sim}^{peak}$ 和 $q_{i,obs}^{peak}$ 分别为第*i*场洪水洪峰流量预测值和实测值。

参照《水文情报预报规范》，其中 NSE 大于 0.7 为合格，其值越高模型参数率定的效果越好；RRD 小于 20%为合格，其值越低越好；PET 小于 30%为合格，其值越低越好；RPE 小于 20%为合格，其值越低越好。

### 3.3.3 参数优选组件化

由于参数优选算法的算法机制不同，算法之间不存在复用性，故可按照优化算法的使用类型，通过控制算法输入的目标函数和约束条件，采用标准化接口的方式，将每种算法封装为独立的组件，然后将评价指标封装为精度评定组件，其实现如表 3-5 所示。当想修改和增加参数优选算法和评定指标以适配新模型时，如果采用已有优选算法，只需继承父类抽象，实现具体实例即可，后续演进模型和调度模型参数优选算法皆是如此实现。

表 3-5 参数优选各组件的实现

Tab. 3-5 Implementation of component for parameters calibrate

组件名称	类名	功能
NSGA-II 算法组件	nsga_2.py	参数优选
PSO 算法组件	pso.py	参数优选
精度评定组件	accuracy_eva.py	精度评定

### 3.4 洪水演进模型

#### 3.4.1 马斯京根模型

##### (1) 线性马斯京根模型

马斯京根模型是 G.T.McCarthy 于 1938 年提出的一种洪水流量演算的方法, 能够根据河道的上游水情推求河道下游的水情, 其基本表达形式为:

水量平衡方程:

$$\frac{dW}{dt} = I - Q \quad (3-5)$$

式中:  $W$  为河段槽蓄量 ( $\text{h} \cdot \text{m}^3/\text{s}$ );  $t$  为时间 ( $\text{h}$ );  $I$  为河段入流量 ( $\text{m}^3/\text{s}$ );  $Q$  为河段出流量 ( $\text{m}^3/\text{s}$ )。

蓄泄方程:

$$W = K[xI + (1 - x)Q] = KQ' \quad (3-6)$$

式中:  $Q'$  表示蓄泄流量 ( $\text{m}^3/\text{s}$ ),  $x$  为流量比重系数,  $K$  表示槽蓄系数 ( $\text{h}$ )。

将公式 (3-5) 和公式 (3-6) 进行差分求和得到离散的差分解为:

$$Q_2 = C_0 I_2 + C_1 I_1 + C_2 Q_1 \quad (3-7)$$

其中:

$$\begin{cases} C_0 = \frac{0.5\Delta t - Kx}{0.5\Delta t + K - Kx} \\ C_1 = \frac{0.5\Delta t + Kx}{0.5\Delta t + K - Kx} \\ C_2 = \frac{-0.5\Delta t + K - Kx}{0.5\Delta t + K - Kx} \end{cases} \quad (3-8)$$

且满足:

$$C_0 + C_1 + C_2 = 1 \quad (3-9)$$

公式 (3-8) 和 (3-9) 共同组成了线性马斯京根模型的公式, 其中  $I_1$ ,  $I_2$  表示时段的初末的入流量,  $Q_1$  表示河道时段初始流量, 通过确定模型的参数  $K$  和  $x$ , 以及演算时段  $\Delta t$ , 就可以计算出流量演算系数  $C_0$ 、 $C_1$ 、 $C_2$ , 以获得下断面流量  $Q_2$ 。

##### (2) 非线性马斯京根模型

实际洪水演进过程较为复杂, 而线性马斯京根模型假定流量随时间线性变化。在河道比降较小的情况下, 随着水位上升, 水流速度迅速增加, 流量并不呈现线性变化, 这就导致采用线性马斯京根模型会产生较大的误差。故 Gill<sup>[61]</sup> 等人提出了非线性马斯京根模型。其非线性槽蓄方程如下:

模型 1:

$$W_t = K[xI_t + (1 - x)Q]^m \quad (3-10)$$

模型 2:

$$W_t = K[xI_t^m + (1 - x)Q^m] \quad (3-11)$$

模型 3:

$$W_t = K[xI_t^p + (1-x)Q^q] \quad (3-12)$$

其中， $m$ 表示槽蓄非线性指数。

由式(3-10)、(3-11)以及(3-12)可分别得到三个非线性马斯京根演进方程，如下所示：

模型 1:

$$Q_t = \left(\frac{1}{1-x}\right) \left(\frac{W_t}{K}\right)^{\frac{1}{m}} - \left(\frac{x}{1-x}\right) I_t \quad (3-13)$$

模型 2:

$$Q_t = \left[ \left(\frac{1}{1-x}\right) \left(\frac{W_t}{K}\right) - \left(\frac{x}{1-x}\right) I_t^m \right]^{\frac{1}{m}} \quad (3-14)$$

模型 3:

$$Q_t = \left[ \left(\frac{1}{1-x}\right) \left(\frac{W_t}{K}\right) - \left(\frac{x}{1-x}\right) I_t^p \right]^{\frac{1}{q}} \quad (3-15)$$

从以上三个方程模型，只要确定其参数 $K$ ， $x$ ， $m$ ， $p$ 和 $q$ 的最优解，就可以利用非线性马斯京根模型得到较为准确的洪水演进结果。

### 3.4.2 一维水动力学模型

本文采用由 Bartos<sup>[62]</sup>等人开发的 pipedream 软件工具包提供的水动力学求解器进行一维水动力学模型的构建，水动力学求解器的计算原理图 3-11 所示。

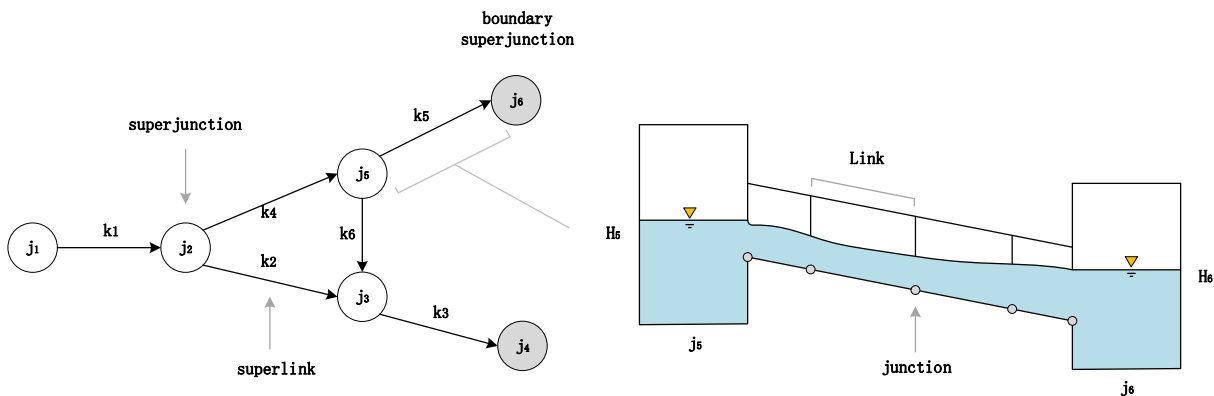


图 3-11 水动力学求解器的计算原理图

Fig. 3-11 Calculation schematic diagram of hydrodynamics solver

水动力学求解器是基于—维非恒定流的圣维南（Saint-Venant）方程进行开发实现的，其组成包括质量守恒和动量守恒定理，具体表达式如下：

质量守恒连续性方程：

$$\frac{\partial A}{\partial t} + \frac{\partial Q}{\partial x} = q_{in} \quad (3-16)$$



动量守恒方程:

$$\frac{\partial Q}{\partial t} + \frac{\partial}{\partial x}(Qu) + gA \left( \frac{\partial h}{\partial x} - S_0 + S_f + S_L \right) = 0 \quad (3-17)$$

式中:  $A$ 为过水断面面积 ( $\text{m}^2$ );  $Q$ 为通过断面的流量 ( $\text{m}^3/\text{s}$ );  $u$ 为水流的平均速度 ( $\text{m/s}$ );  $h$ 为水深 ( $\text{m}$ );  $x$ 为沿河道的距离 ( $\text{m}$ );  $q_{in}$ 为单位宽度的横向流入 ( $\text{m}^3/\text{s}$ );  $S_0$ 、 $S_f$ 以及 $S_L$ 分别表示河床比降、摩擦水头损失比降和局部水头损失比降。

Pipedream 通过采用 SUPERLINK 方法<sup>[63]</sup>来求解上述方程, 采用交错网格的方式, 对每个结合点应用连续方程 (索引为 $Ik$ ), 而对每个链接应用动量方程 (索引为 $ik$ ), 使用向后隐式欧拉法将质量守恒连续性方程离散为:

$$Q_{ik}^{t+\Delta t} - Q_{i-1k}^{t+\Delta t} + \left( \frac{B_{ik}\Delta x_{ik}}{2} + \frac{B_{i-1k}\Delta x_{i-1k}}{2} + A_{s,ik} \right) \cdot \frac{h_{ik}^{t+\Delta t} - h_{ik}^t}{\Delta t} = Q_{in,ik} \quad (3-18)$$

将动量方程离散为:

$$\begin{aligned} (Q_{ik}^{t+\Delta t} - Q_{ik}^t) \frac{\Delta x_{ik}}{\Delta t} + u_{i+1k} Q_{i+1k}^{t+\Delta t} - u_{ik} Q_{ik}^{t+\Delta t} + gA_{ik}(h_{i+1k}^{t+\Delta t} - h_{ik}^{t+\Delta t}) \\ - gA_{ik}S_{0,ik}\Delta x_{ik} + gA_{ik}(S_{f,ik} - S_{L,ik})\Delta x = 0 \end{aligned} \quad (3-19)$$

式中:  $B$ 为链接的顶部宽度,  $A_s$ 为交汇点的表面积,  $Q_{in}$ 为外部流量输入。每个 superlink 的边界条件由上游和下游的 superjunction 提供, 假设在 superlink 的进口和出口处流量呈堰状流动:

$$Q = CA\sqrt{2g\Delta H} \quad (3-20)$$

式中:  $C$ 进出口流量系数;  $\Delta H$ 为 superjunction 与相邻的 superlink 边界连接的水头差。

通过将 Saint-Venant 方程的解嵌入到隐式线性方程组中, 水动力模型在每个时间步长同时求解所有未知数。

### 3.4.3 演进模型组件化

马斯京根模型在河道洪水演进中扮演着重要角色, 运用槽蓄方程和水量平衡方程对划分的河段进行连续演算, 其计算阶段包括参数率定、洪水演进和洪水叠加三个部分, 故可将其拆解为这三大类组件。马斯京根模型又分为线性和非线性两种, 非线性模型相比线性模型采用非线性槽蓄方程以及非线性演进方程获得计算结果, 因此, 洪水演进部分可继续划分为线性演进组件以及非线性演进组件。

一维水动力模型则是通过单元划分和河道断面概化, 利用 Saint-Venant 方程进行洪水演进计算。其关键计算步骤在于流域的概化以及水动力计算, 因此, 可将其拆解为流域拓扑和水动力两个大类组件。

基于以上分析, 系统的演进模型组件库主要包括洪水演进组件、洪水叠加组件、流域拓扑组件、一维水动力组件以及参数率定组件, 其结构如图 3-12 所示。

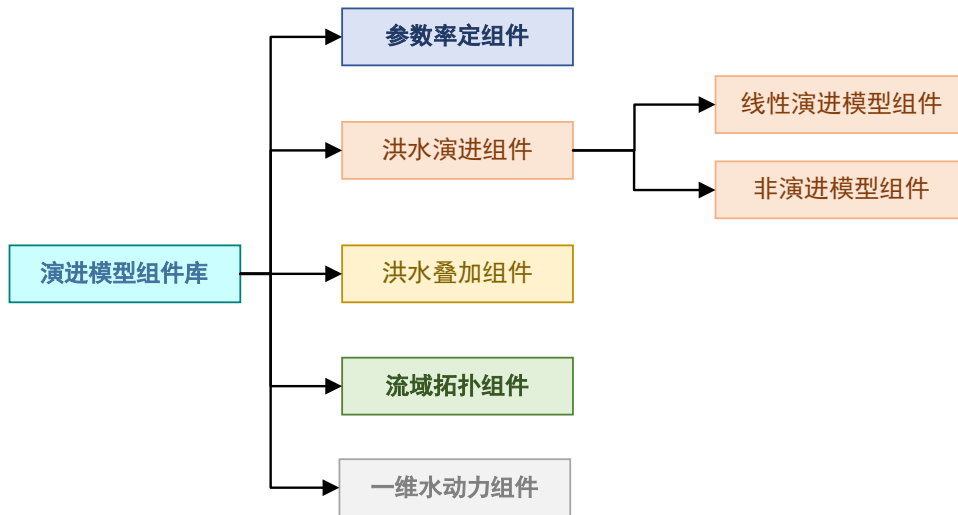


图 3-12 演进模型组件库结构

Fig. 3-12 Routing model component library structure

(1) 线性马斯京根模型组件化

线性马斯京根模型结构较为简单，其组件划分如图 3-13 所示。

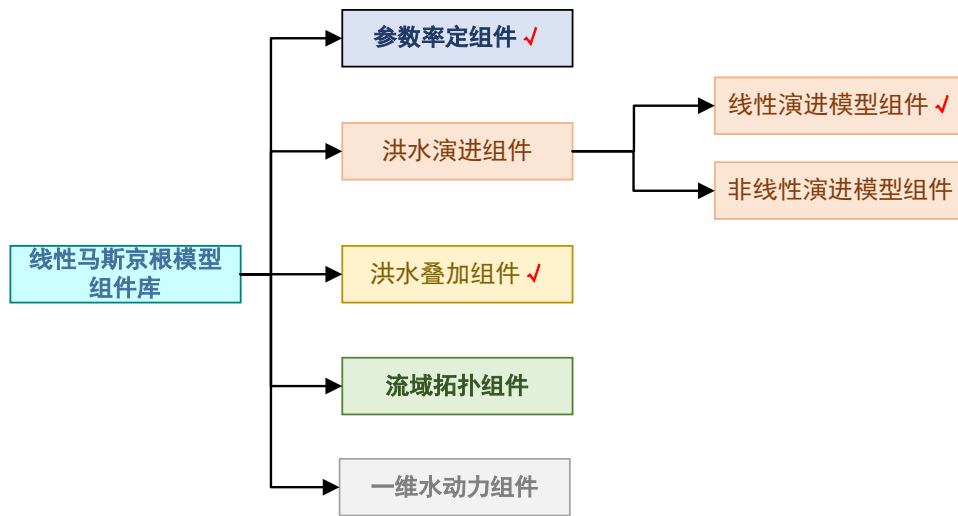


图 3-13 线性马斯京根模型组件划分

Fig. 3-13 Linear Muskingen model component partitioning

在线性马斯京根组件中，通过向参数率定组件输入时段内河段入流量 $I$ 和出流量 $Q$ ，率定得到线性马斯京根模型参数 $K$ 和 $x$ ，进而求出演进后的蓄泄流量 $Q'$ ，最后利用叠加组件传入蓄泄流量 $Q'$ 得到实际出流量 $Q$ 。按照组件的输入和输出将各个模块按照业务逻辑关键进行组装，如图 3-14 所示。组件的类名和功能的说明如表 3-6 所示。

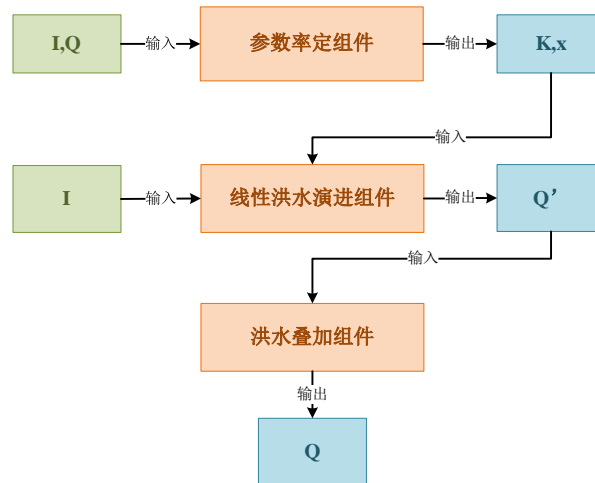


图 3-14 线性马斯京根模型各组件关系

Fig. 3-14 Linear Muskingum model component relationships

表 3-6 线性马斯京根模型各组件的实现

Tab. 3-6 Realization of each component of linear Muskingum mode

组件名称	类名	功能
线性洪水演进组件	linear_flood_routing.py	进行洪水演进计算
洪水叠加组件	flood_overlay.py	洪水过程叠加
参数率定组件	flood_routing_pc.py	计算模型参数

(2) 非线性马斯京根模型组件化

非线性和线性马斯京根模型，都可以划分为洪水演进、参数率定以及洪水叠加，并且洪水叠加组件可复用线性马斯京根模型的组件，不同之处在于非线性马斯京根模型在洪水演进过程中选择非线性洪水演进组件进行运算，如图 3-15 所示。额外组件的类名和功能的说明如表 3-7 所示。

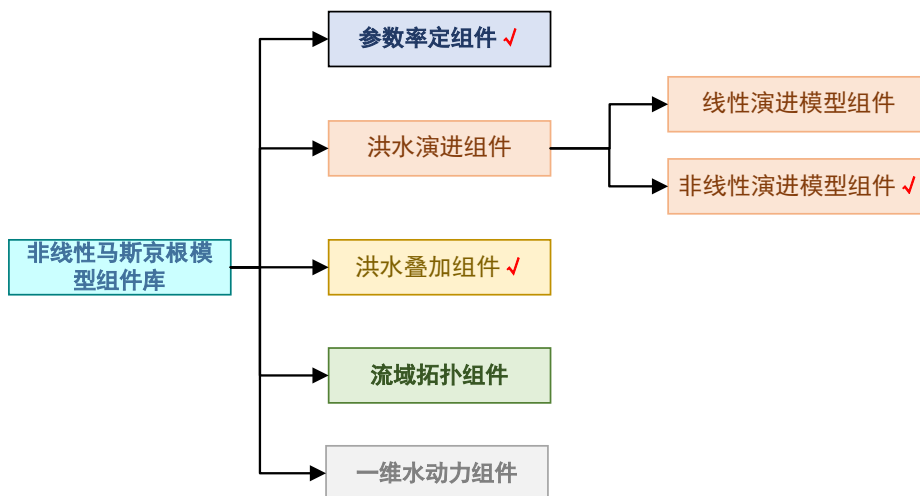


图 3-15 非线性马斯京根模型组件划分

Fig. 3-15 Component partition of nonlinear Muskingum model

表 3-7 非线性马斯京根模型各组件的实现

Tab. 3-7 Realization of each component of nonlinear Muskingum model

组件名称	类名	功能
非线性洪水演进组件	nonlinear_flood_routing.py	进行洪水演进计算
参数率定组件	nonlinear_flood_routing_pc.py	计算模型参数

非线性马斯京根组件间的衔接与线性马斯京根模型类似，不同之处在于在参数率定组件中需要按照非线性洪水演进组件所需参数进行率定。

### (3) 一维水动力模型

一维水动力模型采用的是 pipedream 软件工具包提供的水动力学求解器进行构建，通过对其原理的介绍，其关键步骤在于流域拓扑节点（superjunction）的构建以及节点间的关系由节点间的连线（surperlinks）确定，其包含主要数据内容如表 3-8 所示。其中流域拓扑节点主要由流域关键断面以及支干流交会口确定，如图 3-16 所示，依据断面数据（www.openstreetmap.org），可将临洮县内的洮河流域划分为 31 个节点，其中 8 个节点为支干流的交会口，其余为转折节点。

表 3-8 superjunction 以及 surperlinks 数据内容描述

Tab. 3-8 Description of superjunction and surperlinks parameters

参数类别	参数名称	单位	参数含义
surperlinks	id		连线编号
	name		连线名称
	n		曼宁糙率系数
	dx	m	连线长度
	Q_0	m <sup>3</sup>	初始流量
	h_0	m	内部结点的初始深度
	sj_0		上游节点编号
	sj_1		下游节点编号
	shape(g1,g2,g3)		横截面信息
superjunction	id		节点编号
	name		节点名称
	z_inv	m	节点底部标高
	h_0	m	节点初始深度
	bc		节点是否为边界节点
	max_depth	m	节点处允许的最大深度

故依据其原理，一维水动力学模型组件构成如图 3-17 所示，通过向流域拓扑组件输入关键节点划分数据以及节点间的断面数据（节点关系、断面长度、断面宽度、河床深度、河床糙率，初始流量）构建流域拓扑关系，得到 superjunction 以及 superlink 数据。再将得到的节点数据同入流量数据 I、模型构建参数以及预见期 t，输入到模型计算组件即得到出流量数据 Q，按照组件的输入和输出将各个模块按照业务逻辑关键进行组装，如图 3-18 所示。其包含主要数据内容如表 3-9 所示。

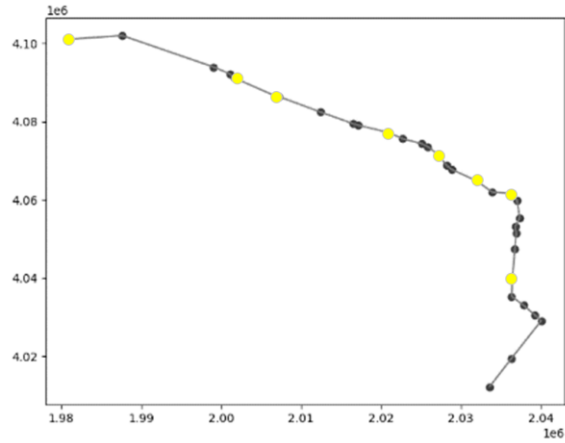


图 3-16 临洮县内洮河流域节点概化

Fig. 3-16 Node generalization of Tao River basin in Lintao County

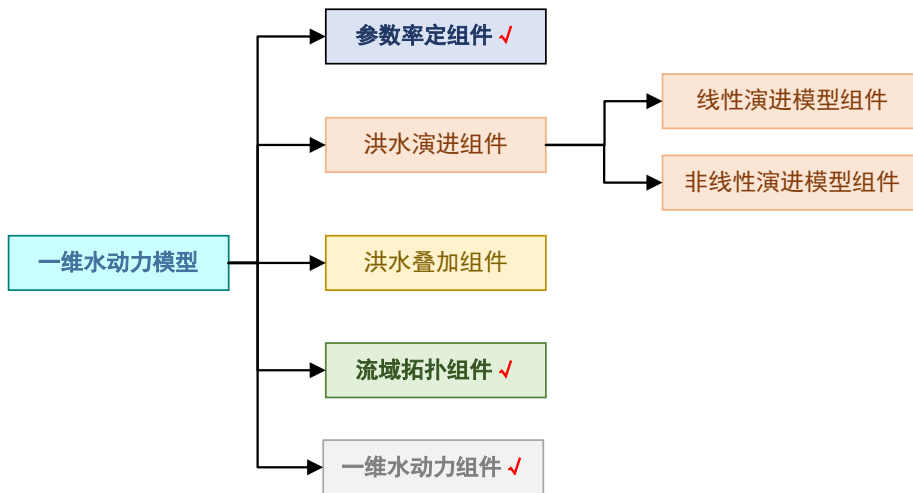


图 3-17 一维水动力学模型组件划分

Fig. 3-17 Component partition of one-dimensional hydrodynamic model

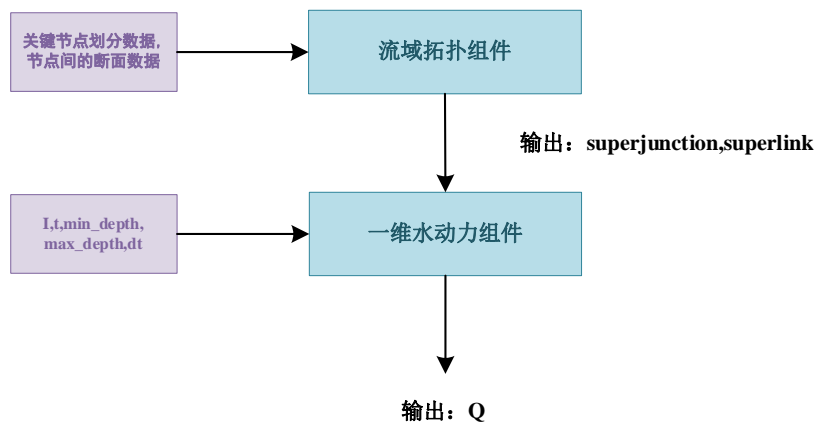


图 3-18 一维水动力学模型各组件关系

Fig. 3-18 Relationship between components of one-dimensional hydrodynamic model

表 3-9 一维水动力模型各组件的实现

Tab. 3-9 Realization of each component of one-dimensional hydrodynamic model

组件名称	类名	功能
流域拓组件	basin_topology.py	进行流域概化
一维水动力组件	one_hydro.py	进行一维水动力计算

### 3.5 防洪调度模型

#### 3.5.1 调度模型构建

防洪调度的主要任务是蓄洪滞洪，利用水库的调节，消减上游的洪水过程，使水库下泄流量不超过下游断面的安全流量。防洪调度模型主要包括常规经验调度以及优化调度两种，常规调度方式虽仍活跃在实际生产环境中，但在数学模型不断兴起的今天，其缺陷越发突出，而目前 NSGA-II 算法在水库优化调度方面<sup>[64-66]</sup>取得了许多研究成功，鉴于此本文主要采用优化调度的方式，引入 NSGA-II 算法对水库进行调度模型的构建，整个调度流程图 3-19 所示。

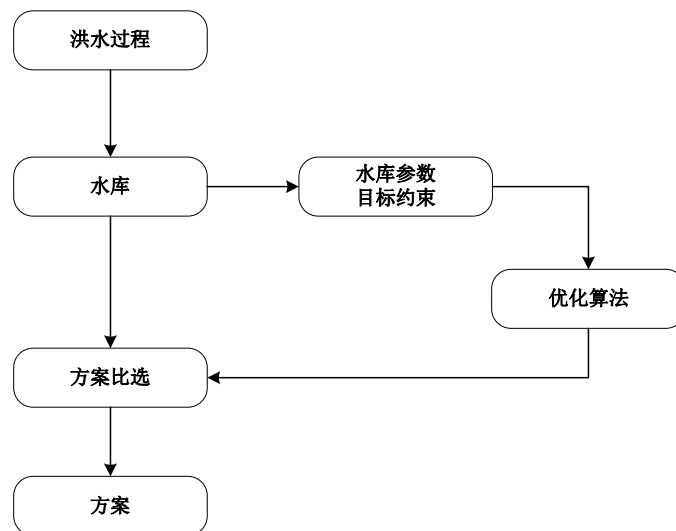


图 3-19 防洪调度流程

Fig. 3-19 Flood control dispatching process

目前，最大削峰准则在水库防洪优化调度中被广泛应用，其核心理念是将入库洪峰流量的减少作为防洪调度最优性的判别指标，通过充分利用防洪库容，最小化水库出库流量（或防洪断面过流量）的最大值<sup>[65]</sup>。本文以最大削峰准则为标准，建立水库防洪优化调度模型。

目标函数为：

$$F_1 = \min \left( \sum_{t=1}^m q_t^2 \right) \quad (3-21)$$

$$F_2 = \min \left( \sum_{t=1}^m h_t^2 \right) \quad (3-22)$$

其中： $m$ 为时段数； $q_t$ 为 $t$ 时刻的水库下泄量； $h_t$ 为 $t$ 时刻的水库水位。

约束条件为：

(1) 水库水量平衡约束

$$V_{t+1} = V_t + (Q_t - q_t) \cdot \Delta t \quad (3-23)$$

其中： $V_t$ 为 $t$ 时段末水库蓄水量； $Q_t$ 为 $t$ 时段内的平均入库流量； $q_t$ 为 $t$ 时段内的平均下泄流量； $\Delta t$ 为计算时间步长。

(2) 水库水位约束

$$Z_{min} \leq Z_t \leq Z_{max} \quad (3-24)$$

其中： $Z_{min}$ 为水库允许的最低水位； $Z_{max}$ 为水库允许最高水位； $Z_t$ 为时段内的水库水位。

(3) 水库库容约束

$$V_{min} \leq V_t \leq V_{max} \quad (3-25)$$

其中： $V_{min}$ 为水库允许的最低库容； $V_{max}$ 为水库允许最高库容； $V_t$ 为时段内的水库库容。

(4) 水库下泄流量约束

$$q_{min} \leq q_t \leq q_{max} \quad (3-26)$$

其中： $q_{min}$ 为水库允许的最小下泄流量； $q_{max}$ 为水库允许最高库容； $q_t$ 为时段内的水库下泄流量。

(5) 下泄流量波动约束

$$|q_t - q(t-1)| \leq \Delta q_m \quad (3-27)$$

其中： $|q_t - q(t-1)|$ 为相邻时段内下泄流量的波动值； $\Delta q_m$ 为允许的波动值。

利用最大削峰准则，在保证各水库库容最大利用的同时，确保水库下泄流量处于安全范围内，并使其最小化，从而得到水库的出库流量过程。

### 3.5.2 调度模型组件化

构造防洪调度模型需要水库特征参数、水库库容曲线、水位流量曲线等数据。因此，本文按照构造调度模型所需条件，将用于调度数据处理的部分划分为水库特征参数组件和特征曲线插值组件。同时，调度算法部分被划分为参数率定组件和 NSGA-II 优化调度组件，如图 3-20 所示，其组件实现如表 3-10。若需要增加调度模型的优化算法，只需添加相应组件即可。

表 3-10 防洪调度各组件实现

Tab. 3-10 Implementation of flood control dispatching components

组件名称	类名	功能
水库特征参数组件	reservoir_parameter.py	获取水库特征参数
特征曲线插值组件	reservoir_curve.py	获取库容、流量曲线
NSGA-II 优化调度组件	nsga2_opt.py	NSGA-II 优化调度
参数率定组件	reservoir_pc.py	进行调度模型参数率定

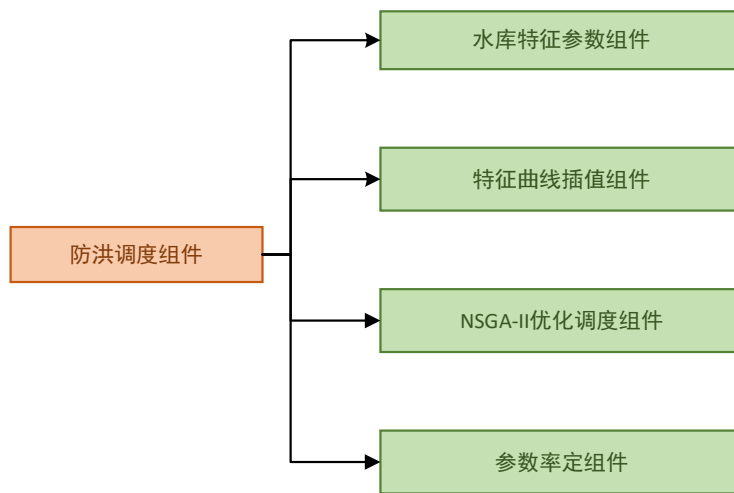


图 3-20 防洪调度模型组件划分

Fig. 3-20 Partition of flood control dispatching model components

### 3.6 模型组件库建立

模型组件库的作用在于实现模型组件之间的共享，使得一个组件可以被多个模型所复用，从而提高模型的可复用性，减少重复开发的工作量。模型组件库的建立依赖于接口的标准化处理，如图 3-21 所示，主要是通过构建最顶层组件的抽象类，其余组件继承抽象类，实现具体的算法实例来实现。通过统一所有组件的继承关系，建立了统一访问的机制，新增模型时，只需按照统一的接口标准，继承其类型下的抽象类，实现具体算法即可完成组件构建，上传至组件库，而无需修改现有组件库中的任何组件。

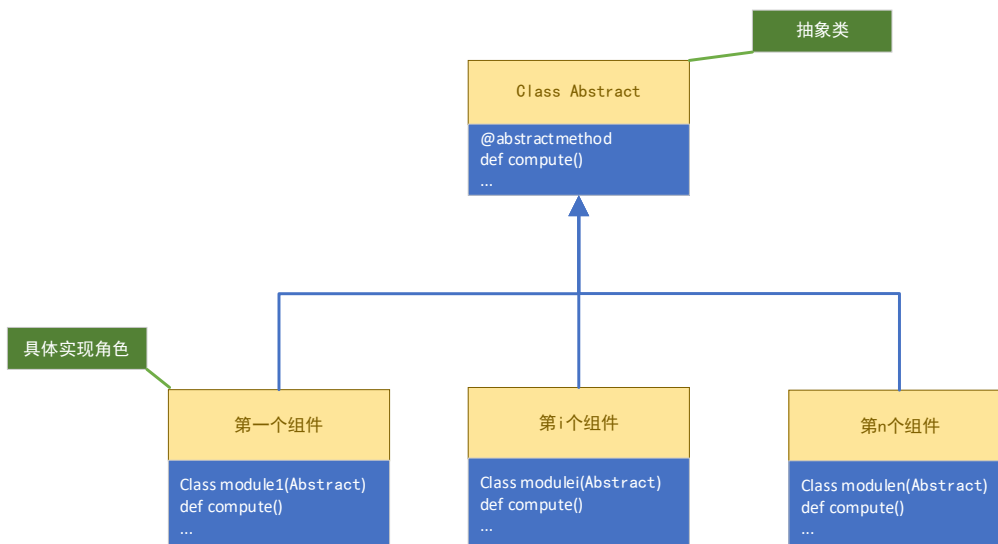


图 3-21 组件标准接口设计图

Fig. 3-21 Component standard interface design drawing

通过模型组件库，可以实现模型的统一管理和高效利用，模型组件库的组件公用情况如表 3-11 所示。



表 3-11 组件库组件公用情况  
Tab. 3-11 Component library component public condition

模型类型	组件	公用模型
洪水预报	一层蒸散发组件	新安江模型, GR4J 模型
	二层蒸散发组件	新安江模型
	三层蒸散发组件	新安江模型, GR4J 模型
	蓄满产流组件	新安江模型, GR4J 模型
	二水源划分组件	新安江模型
	三水源划分组件	新安江模型
	线性水库组件	新安江模型
	非线性水库组件	GR4J 模型
	单位线汇流组件	GR4J 模型
	马斯京根组件	新安江模型
洪水演进	洪水演进组件	线性和非线性马斯京根
	洪水叠加组件	线性和非线性马斯京根
	流域拓扑组件	一维水动模型
	一维水动力组件	一维水动模型
参数优选	NSGA-II 算法组件	\
	PSO 算法组件	\
	精度评定组件	\
防洪调度	水库特征参数组件	调度组件
	特征曲线插值组件	调度组件
	NSGA-II 优化调度组件	NSGA-II 优化调度

### 3.7 本章小结

本章针对组件化思维下的专业模型集成体系进行了研究。从专业模型的结构和逻辑关系入手, 分析了洪水预报模型、洪水演进模型、防洪调度模型以及参数优选算法的基础模型与方法, 包括新安江模型、GR4J 水文模型、马斯京根模型等。然后在基础方法构建中粒度模型的体系下, 将涉及的模型算法组件化, 通过标准化接口的方式使组件按照业务逻辑进行拼接, 实现专业模型的集成开发, 构建模型组件库, 为后续大模型调用专业模型提供支撑。

## 4 基于大模型的洪涝业务应用模式研究

### 4.1 大模型与洪涝知识融合研究

支撑洪涝灾害防御的系统应用建设在一定程度上对经验知识的积累，依赖于专家经验和业务规则的紧密结合。传统的应用在规则制度在线化、知识化方面存在不足，业务人员在工作过程中仍需查阅纸质文档，知识的串联需要由人通过线下讨论的方式逐步对齐和深入，不仅容易漏，而且效率低。为解决这些问题，本节将研究大模型与洪涝知识的融合，以提高系统的智能化水平，促进知识的有效串联和应用，从而提升工作效率和准确性。

#### 4.1.1 大模型知识融合方法

在洪涝灾害防御业务中，洪涝知识主要包括业务规则、调度规程、应急预案、水库运行方式、运行数据等信息。与传统应用相比，大模型驱动这些知识具有显著优势，因为大模型本身已经学习了大量文本数据，从而具备更强的上下文理解能力，能够更准确地理解用户的问题。同时，大模型能够处理包括文本和图像在内的多模态数据，使其能够构建更加全面的知识集，提供更加多样化的信息和服务。目前基于大模型的知识融合方式主要有三种：微调（Fine-Tuning）、基于提示工程（Prompt Engineering）以及检索增强生成（Retrieval Augmented Generation）。

Fine-Tuning 是一种通过调整大模型参数来适应特定知识的方法，特别适合于那些需要高度专业化的任务，但它的成功训练依赖于高质量的数据集和精心设计的训练方式<sup>[67]</sup>，而且微调的结果可能并不总是符合预期<sup>[68]</sup>。基于 Prompt 的方式则是将特定信息以上下文的形式作为输入提供给大模型，从而在大模型内部建立短期记忆，这种方式可以确保回答的正确性和精度，但大模型一次性处理的文本量（token）是有限的，如果知识数据规模较大，超出这一限制，其可行性和效率可能会受到影响。

鉴于 Fine-Tuning 和 Prompt 方式各自存在的局限性。本文选择采用检索增强生成的方式来实现大模型与洪涝知识的融合。RAG 是一种通过集成外部知识库来增强大模型的范式<sup>[69]</sup>，其具体定义可以从实例应用中总结得到，图 4-1 展示了 RAG 应用的工作流程，当用户向大模型提出一个关于预案中预警标准的问题时，受限于其预训练数据，大模型无法直接给出准确的回答，而 RAG 通过从外部知识库中检索到相关文档，弥补了这一知识空缺，它将检索到的相关内容和用户最初的问题合并成一个提示，从而引导大模型基于这个提示给出一个较为满意的回答。这种方法有效地利用了外部知识库中的信息，提升了大模型对特定问题的回答能力。

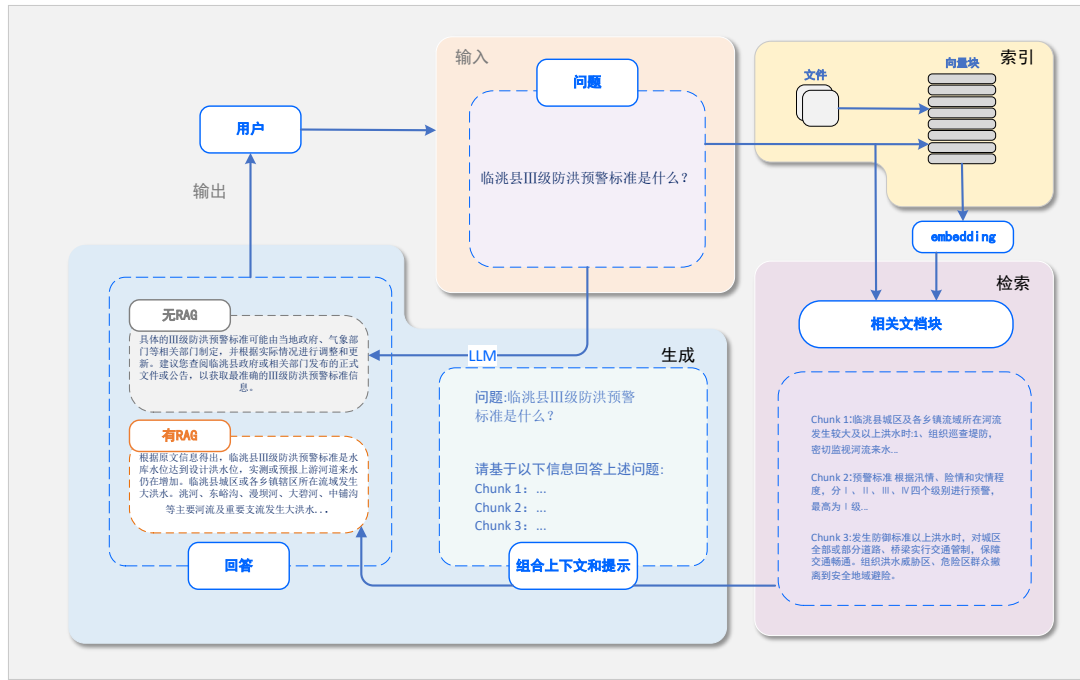


图 4-1 RAG 问答流程实例

Fig. 4-1 A representative instance of the RAG process applied to question answering

#### 4.1.2 知识向量化处理

RAG 应用流程的完整实现首先涉及对需要检索的知识数据进行预处理。在洪涝灾害防御业务中，知识数据通常以 Word、PDF、CSV 和 TXT 等格式存储，有时还包括图片数据。为了实现对这些多模态数据的高效管理和检索，需要将它们数字化并存储在向量数据库中，具体流程如图 4-2 所示，步骤包括数据加载、数据分割、数据向量化以及向量存储。

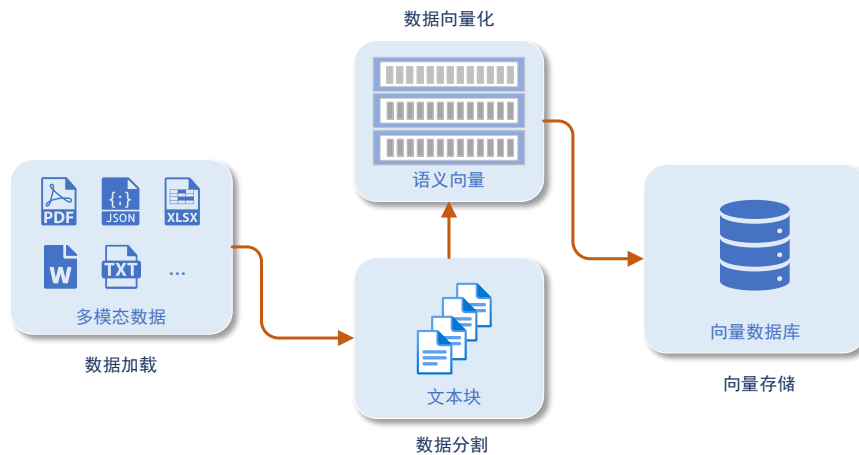


图 4-2 知识数据处理流程

Fig. 4-2 Knowledge data processing flow

LangChain 是一个专门用于构建语言模型驱动应用程序的框架<sup>[70]</sup>，它提供了一系列工具、组件和接口，以简化创建由大模型支持的应用程序的过程。本文主要基于 LangChain 框架，完成知识数据的预处理工作，具体步骤如下：

### (1) 数据加载。

包含洪涝知识的文件通常由多种数据类型构成，为了加载这些不同类型的数据，需要将处理为同一个范式。LangChain 框架为不同类型的数据提供了相应的加载器（Loader）。对于常见的文本类型如 CSV、TXT、JSON 等，可以直接使用对应的 Loader 进行加载，对于 Word、PDF、Markdown 等包含文字、图片和表格的非结构化数据，可以通过继承 LangChain 提供的 UnstructuredFileLoader 接口，自定义编写 Loader，以处理 PDF 文件为例，数据加载后的结果如图 4-3 所示。最终目标是将不同类型的数据都转换为文本类型，以便于统一数据的类型。

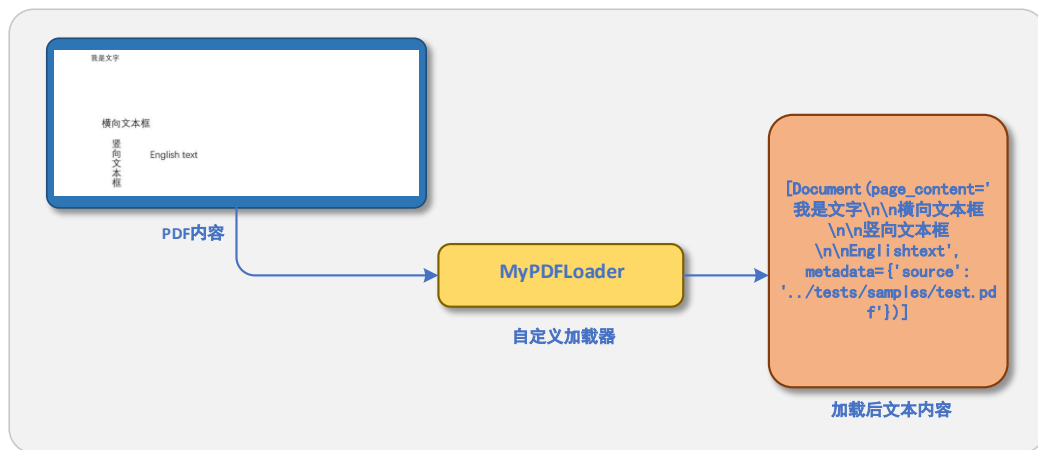


图 4-3 PDF 数据加载示意图

Fig. 4-3 PDF data loading diagram

### (2) 文本分割。

加载后的文本数据可能包含十页甚至数百页的内容，因此，需要按照一定尺寸将文档内容划分成小块，使后续更容易对数据进行向量化处理。文本分割主要考虑两个因素，一是语义完整性对整体的检索效果的影响，二是向量化模型的 Tokens 限制情况。为了保证中文环境下的语义情况，通常采用句号、问号等常用断句符号作为分割符合，以提高分割块语义的完整性。同时根据向量化模型的 Tokens 限制，选择固定文本分割的长度。

LangChain 提供了多种文本分割工具，RecursiveCharacterTextSplitter 是其推荐使用的文本分割器，可以通过指定换行符来分割文本。本文以继承该函数，重写函数方法的方式，将中文常用的断句符号设置为分割点，以正则表达式匹配的方式进行文本切割，以提高对中文文本的解析能力，然后通过设置文本块状尺寸大小（chunk\_size）和重叠字符串长度（chunk\_overlap），减少语义信息的损失，其中 chunk\_size 指的是被切割的字符串的最大长度，chunk\_overlap 指切割时，前后两段字符串重叠的字符数量，本文经过多次测试，选择将 chunk\_size 大小默认设置为 250，chunk\_overlap 设置为 50，以适应更多文本的处理需求。

### (3) 数据向量化。

数据向量化通常通过嵌入（embedding）技术实现。Embedding 是一种将文本编码为低维连续向量的技术<sup>[71]</sup>，通过考虑文本上下文、语法模式和语义相关性，使相似语义的文本

数据在向量空间中更加接近。目前常用的 Embedding 技术有很多，本文选用“bge-large-zh”作为 embedding 的基础模型。在使用 embedding 模型对文本进行编码时，首先会使用分词器（Tokenizer）对文本进行分词，基础分词方式为一个汉字即对应一个字符，然后将分词后的字符序列转化为对应的向量表示，如图 4-4 所示。在分词过程中，会遇到专业领域词汇不在词汇表中的问题，为了提高向量的语义完整性，本文采用构建关键词目录的方式对“bge-large-zh”模型进行二次训练，改变模型的分词器分词方式，使其按照关键词进行分词，以提高后续向量检索时的命中率，优化后的向量化示意图 4-5。

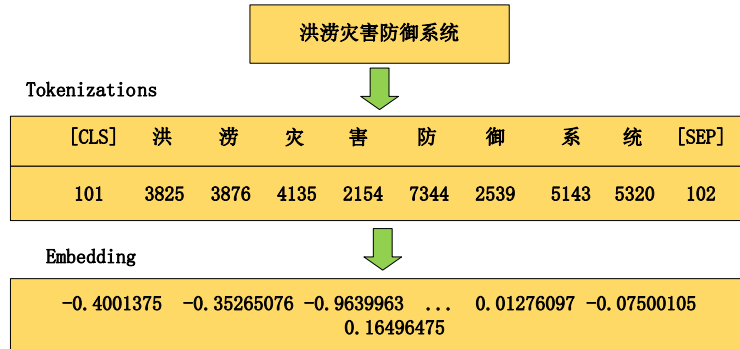


图 4-4 优化前的文本向量化编码

Fig. 4-4 Vectorized text coding before optimization

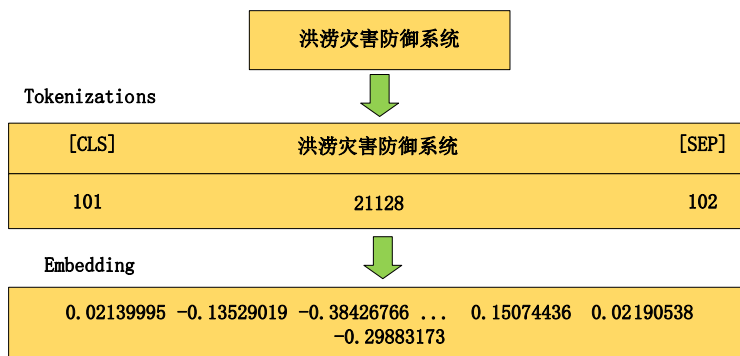


图 4-5 优化后的文本向量化编码

Fig. 4-5 Optimized text vectorization coding

(4) 向量存储。考虑到知识库数据可能庞大，采用传统的迭代方式进行检索可能耗费大量时间。因此，本文引入了 FAISS<sup>[72]</sup>（Facebook AI Similarity Search）向量数据库，以加速搜索过程。FAISS 是由 Facebook 开发的一款高性能相似性搜索开源库，专注于处理大规模高维向量数据，其采用倒排文件和乘积量化等高效的索引结构，能够显著提高向量相似性搜索的速度。

#### 4.1.3 知识应用策略

RAG 流程实现的第二阶段为对知识数据的应用，当用户向大模型提问时，通过高效的检索方式，召回与提问最相关的知识，然后将这些知识以上下文的形式注入到 Prompt 当中，大模型依据 Prompt 提示，生成相应答案。而所谓的 Prompt<sup>[73]</sup>工程指的是在使用大模型进

行自然语言处理任务时，设计和调整输入的提示，来引导大模型生成预期输出的文本，详细内容将在 4.2 节进行介绍。

知识数据的应用流程主要包括三个环节：数据检索、二次重排序以及大模型生成，如图 4-6 所示。

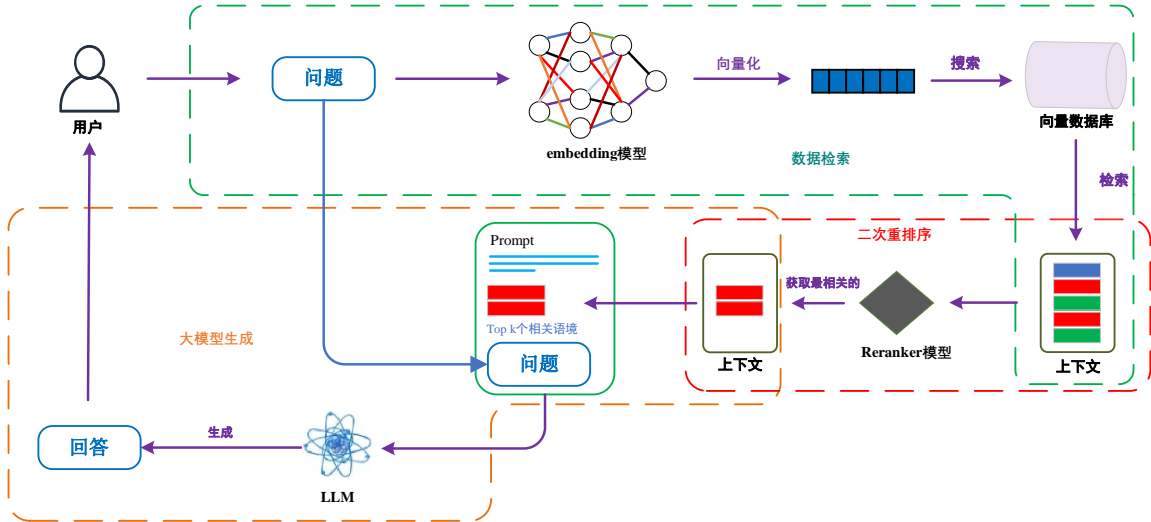


图 4-6 知识数据的应用流程

Fig. 4-6 Application flow of knowledge data

(1) 数据检索。数据检索主要采用相似性检索的方式，当用户提交问题时，问题会通过 embedding 进行向量化编码，然后通过计算问题向量与向量数据库中索引结构存储的向量之间的余弦相似度，获得相似度最高的向量，作为问题上下文，具体步骤如图 4-6 绿色虚线框所示，其中余弦相似度计算公式为：

$$\cos(v, w) = \frac{b \cdot q}{|b||q|} = \frac{\sum_{i=1}^L b_i q_i}{\sqrt{\sum_{i=1}^L b_i^2} \sqrt{\sum_{i=1}^L q_i^2}} \quad (4-1)$$

式中， $b$ 表示向量库中存储的向量， $q$ 表示用户的问题向量， $L$ 表示向量的长度。

(2) 二次重排序。当从向量数据库中检索到多个上下文时，这些上下文与用户问题的相关性可能不同，一些问题可能非常相关（图 4-6 中红色小块），而另一些相关程度可能不够（图 4-6 中绿色和蓝色小块），故本文选择在数据检索后完成后加入“bge-reranker-large”重排序模型，以评估检索得到的上下文与用户问题的相关性，优先将相关性最大的上下文作为提示生成答案。

(3) 大模型生成。在从 FAISS 向量库中检索到相关的上下文信息后，为了确保大模型能够利用这些信息生成准确的回答，需要将这些上下文信息注入到 Prompt 模板中，如图 4-7 所示。以确保大模型在生成回答时能够拥有足够的背景知识，从而提高回答的准确性和相关性。

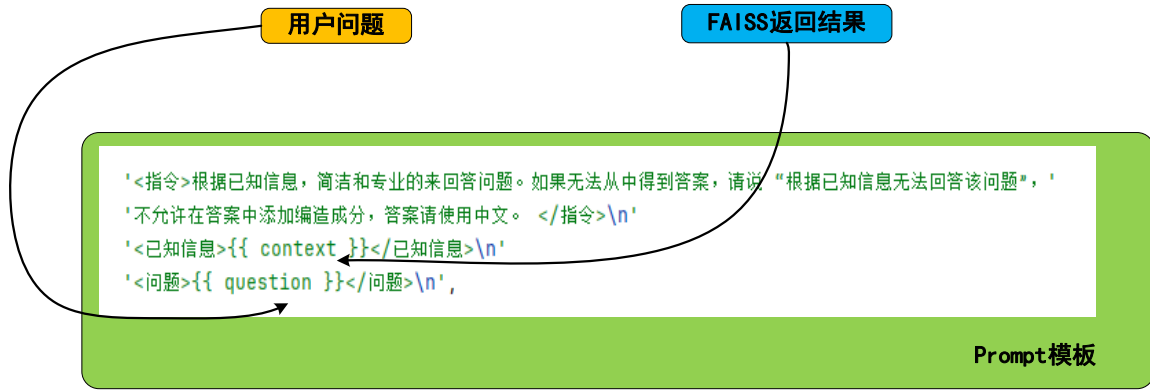


图 4-7 上下文信息注入 prompt 模板

Fig. 4-7 The context information is injected into the prompt template

#### 4.1.4 洪涝知识库构建

基于以上原理，本文建立预报方案库、专家经验库以及调度规则库，将洪涝知识数据利用大模型进行集成融合。

(1) 预报方案库。预报方案库的重点是洮河干支流关键断面以及水库等水利工程。内容包括流域防洪预案、应急抢险预案、工程的调度预案等，实现文字预案的结构化、数字化和模块化，实现预案快速便捷管理。

(2) 专家经验库。专家经验库的重点是对洪峰流量预报经验、闸坝工程调度经验、水位流量曲线经验等专家经验以及专家对历史洪水处理的决策、安全事件的处理等决策方法进行建设。实现历史洪水处理、安全事件处理等经验决策的数字化、结构化，融合当前水文气象信息、处理方案信息，为业务决策提高经验支持。

(3) 业务规则库。业务规则库主要包括业务法律法规、规章制度、技术标准等，通过对业务文档的处理，形成一套可组合应用的结构化规则集，以支撑业务场景的规则适配，应对不断变化的规则需求，规范和约束水利业务的管理行为。

问题	水库水位超过汛限及以上水位时该怎么办？	临洮县 II 级防洪预警标准是什么？
无知识库 LLM	应严格按照国家法律法规和应急预案操作，确保各项行动符合国家防汛抗灾的总体要求。同时，要充分发挥中国特色社会主义制度优势，统一指挥、协调联动，形成救灾救灾的强大合力。	根据提供的参考信息，没有直接提及临洮县 II 级防洪预警标准的具体细节。通常，防洪预警级别会根据防洪工程的重要性、可能造成的社会经济损失、危及的人口数量等因素来确定。
有知识库 LLM	<pre>{   "page_content": "水库水位超过汛限及以上水位，实测或预报上游河道来水持续增加时：\n1、立即开闸泄水，严密监视水情，加密组织巡查大坝、泄水建筑物，组织下游风险区群众做好撤离准备，\n2、调度抢险物资，组织抢险队伍上坝待命。必要时在坝顶抢筑子堰，防止洪水漫顶。",   "metadata": {     "source": "临洮县防汛抗洪应急预案.docx"   },   "type": "Document",   "score": 0.4573060870170593 }</pre>	<pre>{   "page_content": "II 级防洪预警：水库水位达到设计洪水水位，实测或预报上游河道来水仍在增加。\\n临洮县城区或各乡镇辖区所在流域发生大洪水。\\n洮河、东峪沟、漫坝河、大碧河、中铺沟等主要河流及重要支流发生大洪水。\\n洪水围困15~29人，并威胁其生命安全，或者死亡与失踪10~29人，或者影响500名以上、1000名以下群众正常生活。",   "metadata": {     "source": "临洮县防汛抗洪应急预案.docx"   },   "type": "Document",   "score": 0.28669804334640503 }</pre>

图 4-8 有无知识库大模型回答对比

Fig. 4-8 There is no knowledge base large model answer comparison

为了验证大模型驱动的知识在问答任务中的表现，以预报方案库为例，展示了二个专

业技术问题的问答记录，并与普通的大模型进行比较，如图 4-8 所示。由于问题内容的专业性，普通的大模型缺乏这方面的领域知识，无法获得准确的结果。相比之下，进行知识融合的大模型能够更为准确地进行回答。

## 4.2 大模型与洪涝数据及模型融合研究

### 4.2.1 大模型 Agent 概述

大模型一般只拥有它被训练的知识，而这种知识是固定的，无法获取外部的数据。而在用户与大模型的交互中，交流的场景及需求是多种多样的，固定的知识使得大模型在推理过程中容易与外界处于“断层”的状态。为了解决这种“断层”状态，针对不同的任务，大模型可能需要借助各种不同的工具或数据源。对于一般的咨询类问题，大模型可以依靠自身的知识储备以及外挂知识库的方法来提供答案。但当用户需要查询特定行业的实时数据时，大模型则无法依靠自身的数据来回答。

而随着大模型技术的不断提升，人们不再满足简单的语言交流，想要直接告诉大模型一个最终目标，然后它自己去理解任务、分解任务、处理任务，再把最终的结果返回给用户，因此基于大模型的 Agent<sup>[74]</sup>被提出。所谓 Agent，是指基于人工智能技术开发的程序或系统，能够执行特定任务、解决问题，或模拟人类智能行为的实体。简单来说 Agent 就是用来与外部环境进行交互，以得到模型内部没有训练过的数据。

Agent 决策流程主要包括三部分：感知、规划和行动<sup>[74]</sup>，如图 4-9 所示，为一个精简的 Agent 决策流程。

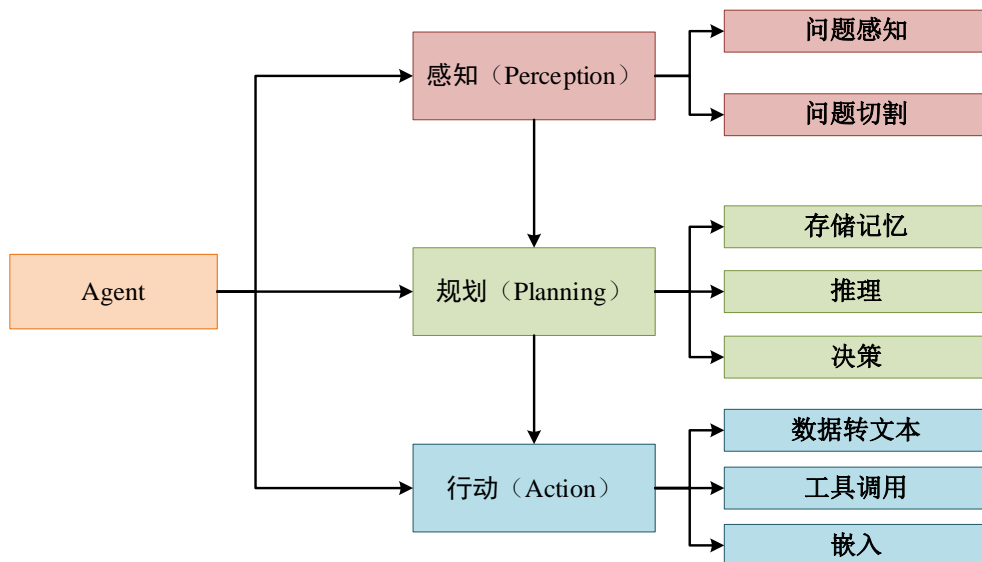


图 4-9 Agent 决策流程图

Fig. 4-9 Agent decision flow chart

(1) 感知 (Perception)。主要负责从用户的问题中去提取关键部分并进行问题的处理。人类都是依靠五官去感受周边的环境，使我们能够感知世界并与之互动。同样对于大模型



的 Agent 来说，这种感知也十分重要，对用户的问题进行感知，有助于 Agent 更好的理解问题背景，做出明确的决策。

(2) 规划 (Planning)。主要是用作存储上下文记忆，以及对处理后的问题进行推理及决策。规划阶段是 Agent 的核心部分，如同人的大脑能够处理各种信息，产生各种思想，控制不同的行为。为了确保数据的有效处理，在接受到感知模块处理过的信息后，规划模块会以摘要、向量或其他数据结构的形式记住 Agent 过去的观察、想法和行为，以备决策使用。

(3) 行动 (Action)。主要是将规划决策后的结果利用工具进行执行，并将工具执行结果返回给用户。通过外部工具的使用，达到与外部环境交互的目的。

#### 4.2.2 大模型与洪涝业务交互策略

尽管大模型拥有庞大的知识储备和强大的涌现能力，但在实际应用中，我们仍会明显感受到模型的局限性，例如无法获取最新信息、只能提供文字建议而不能直接解决问题。这些问题在一定程度上限制了大模型在洪涝业务中的应用潜力。为了克服这些局限，大模型与洪涝业务的交互需要进一步扩展，实现通过大模型获取雨晴、水情等实时数据，并利用这些数据驱动算法模型，获得计算结果。

为了实现这一交互目标，本文通过创建工具函数的方式，利用 Agent 的感知、行动和规划能力，让大模型不再仅依赖于自身的数据库知识，而是可以通过附加一个业务库的方式来响应用户需求。当用户提问时，大模型可以检索这个业务库，根据用户的实际需求调用相应函数获取运行结果，并基于这些结果给出回答，其流程如图 4-10 所示。

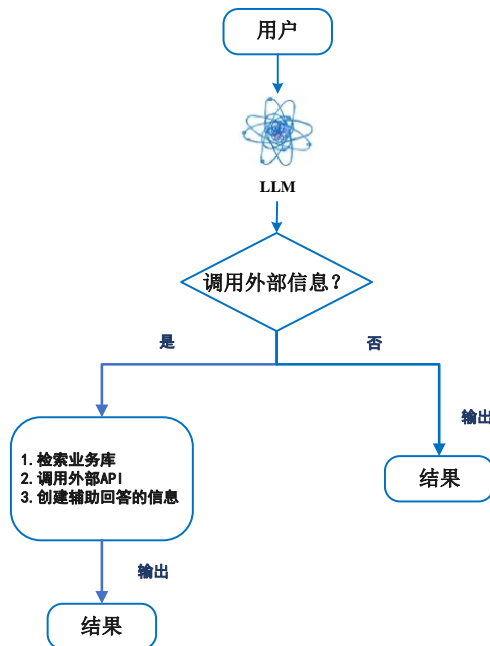


图 4-10 大模型交互流程

Fig. 4-10 LLM interaction process

大模型与洪涝业务的进一步交互主要场景有两个：一是与数据库进行交互，以获取实

时数据；二是与专业模型进行交互，通过调用这些模型来获得业务的计算结果。因此，业务库主要包括两类函数：数据库交互的函数以及调用专业模型的函数。

(1) 数据库交互函数。这些函数的核心功能是利用大模型的语义理解能力和编程技能，根据用户的需求生成相应 SQL 语句，实现与数据库进行交互。由于大模型本身具有强大的编程能力，特别是在编写 SQL 语句方面，于是本文采用给予大模型数据库表结构关系的方法，将数据库里的每张表，按照基本解释以及字段说明进行组织，然后将所有表的描述信息合并为一个文件，为大模型提供数据库背景的提示，加强其 SQL 语句编写的准确性。随后，通过调用 pymysql 工具库的方式执行大模型生成的 SQL 语句，实现与数据库的交互。

(2) 调用专业模型的函数。这些函数主要依赖于大模型对用户问题的拆解能力，需要根据用户意图调取相应的水利专业模型获取计算结果。为了实现这一功能，本文利用第四章组件化后的专业模型，使用 gRPC 技术将其转换为可以通过参数传递进行调用的外部服务，通过意图识别和数据库查询获得模型组件的参数及输入，然后利用 LangChain 框架下的 SequentialChain 将模型组件按照逻辑关系进行串联计算，以实现驱动模型算法。当用户的需求需要调取专业模型进行解决时，如果不指定采用的模型算法，则优先采用存储于数据库中的默认模型类别及参数进行实现。

通过构建业务库的方式，可进一步提高了大模型在处理洪涝业务时的灵活性和实用性，使其能够更好地适应复杂多变的洪涝管理需求。

#### 4.2.3 洪涝业务推理链构建

洪涝业务问题常常涉及复杂的逻辑关系，大模型能否根据语境中的提示，挖掘在背后隐藏的逻辑关系完全取决于大模型的推理能力。为了增强大模型对业务背景的理解和用户问题的把握，研究如何向模型传递现有的业务知识至关重要。

提示工程 (Prompt) 在此过程中扮演着桥梁的角色，通过连接用户输入和模型输出，帮助 Agent 感知和规划用户的问题。目前比较常见的提示工程方法主要包括 Few-shot<sup>[75]</sup>、Few-shot-Cot<sup>[76]</sup>以及 Least-to-Most<sup>[77]</sup> (LtM)。其中 Few-shot 通过提供相似问题的问答对，让模型参考学习，从而提升其解决类似问题的推理能力。相比于 Few-shot，Few-shot-Cot 还需要给出获得问题结果的推导过程。而 LtM 的方法则是先让大模型去分解问题，通过逐一解决子问题的方式来获得解决原始问题的答案，以实际问题为例，如图 4-11 所示，执行流程包括：

Step1: 将原始问题分解为多个顺序子问题；

Step2: 解决第一个子问题：“获取漫坝河实时降雨情况”，得到漫坝河实时降雨序列；

Step3: 将第一个子问题的问题和答案传递给第二个子问题：“获取中铺沟断面和漫坝河的位置关系”，得到漫坝河和中铺沟断面的位置拓扑关系；

Step4: 以此类推，逐一解决子问题，得到最终问题的答案。

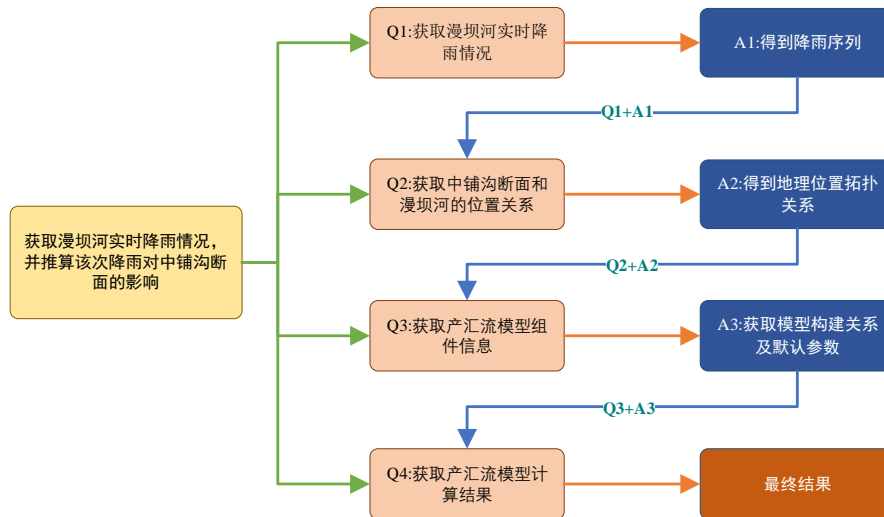


图 4-11 LTM 方式执行流程

Fig. 4-11 The flow is executed in LTM mode

为了使大模型对洪涝灾害防御的业务体系更加了解，本文采用了外挂业务推理链的方式，以提高模型对业务的理解和任务的拆解能力。考虑到大模型本身的海量训练样本，外部挂载的文档无论内容多少，都无法动摇其原始的知识体系。因此外部文档的核心作用是引导模型重视某部分信息、同时补充模型本身不具备的信息。换句话说，外部业务推理链的创建是为了给模型说明规则，同时提供一些“启发”。

基于上述原则，本文采用 Few-shot 提示构建洪涝灾害防御业务的推理链，利用 LTM 逐一解决子问题的方式，解决原始的业务需求。灾害防御业务的推理链主要包括三部分：项目背景知识、业务推理链示例以及大模型身份设定。项目背景知识指的是告诉大模型洪涝灾害防御系统应用的背景、核心功能、运行方式等；身份设定则是赋予大模型角色，告诉其负责的任务以及推理链的执行方式。推理链示例主要以问答对的方式利用 Few-shot 引导大模型进行中间问题拆解。例如“获取漫坝河实时降雨情况，并推算该次降雨对中铺沟断面的影响”这个问题可以拆解为“获取漫坝河实时降雨情况”、“获取中铺沟断面和漫坝河的位置关系”、“获取产汇流模型组件信息”以及“获取产汇流模型计算结果”四个子问题，通过构建多个此类拆解问题，构建推理链示例，强化大模型对于业务的理解和任务的拆解能力。

#### 4.2.4 多策略下意图识别增强

用户意图对齐是大模型普适性的最根本保证，特别是在智能体（Agent）的开发过程，业务库调用的精确性取决于能否准确识别用户的意图。当大模型内置的函数调用机制无法满足使用需求时，则需要探索其他方法来增强函数调用的稳定性，其中核心的思路就是能否将此前完全在大模型内部自动完成的外部函数选取工作来手动完成，即通过一些其他方法来先确定当前需求要使用的外部函数，进而手动创建函数调用过程。

基于以上思路，本文采用有监督学习的方法，融合三种意图识别的技术，以提升了用

户意图分类的精确度保证函数调用的准确度，包括：基于 Few-shot 的函数调用意图识别、基于 Embedding 的机器学习意图识别以及基于训练文本匹配的意图识别。

(1) 基于 Few-shot 的函数调用意图识别。利用大模型本身的函数调用机制，将业务问题以 Few-shot 的方式，作为大模型进行函数调用的提示。

(2) 基于 Embedding 的机器学习意图识别。其具体实现流程为先积累大量用户问题，然后手动对其进行标注，需要注明在当前用户问题需要调用哪个外部函数（或者无需调用外部函数）来进行回答。当完成用户需求文本数据标注之后，接下来对其进行 Embedding 编码，并将编码结果视作数值型特性，利用 LGBM、XGB 等机器学习模型进行建模预测，预测当前用户需求属于哪一类需求，并由此判断回答当前用户问题需要调用哪个外部函数。

(3) 基于训练文本匹配的意图识别。其同样需要累计大量手动标记后的用户问题，通过对用户问题划分训练集和测试集，然后围绕训练集和测试集进行 Embedding 编码，将测试集文本和训练集文本进行余弦相似度的计算，找到与测试样本最匹配的训练集文本，根据训练集文本上的标注，确定用户的意图。

基于 Few-shot 的函数调用意图识别更加依靠于大模型本身的推理能力，Few-shot 只是进一步激活大模型的能力，在用户意图风格区别较大时效果最佳。基于 Embedding 的机器学习意图识别，依赖于 Embedding 词向量的有效性和机器学习模型本身的能力，在大样本下效果最佳。而基于训练文本匹配的意图识别，经测试，方法性能不如基于 Embedding 的机器学习意图，但更适用于融合新标签的场景中。而限于各类方法的缺陷，单一的意图识别很难保证较高的准确性，于是本文采用融合以上三种意图识别增强的方法，利用投票法（Voting）判断问题最终意图识别结果，获得结果后将用户意图类别作为用户原问题的提示，再次将问题传给大模型，如果顺利执行则返回结果，否则提醒用户重新组织语言，如图 4-12 所示，具体步骤为：

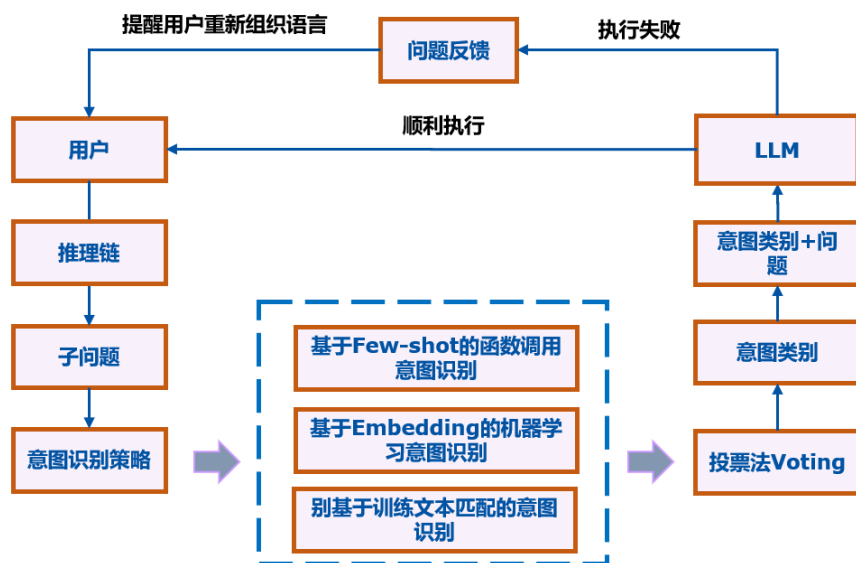


图 4-12 意图识别策略

Fig. 4-12 Intention recognition strategy

- Step1: 借助 Few-shot 的函数调用创建一个预测结果;
- Step2: 借助基于 Embedding 的机器学习意图识别创建 2-3 个预测结果;
- Step3: 借助基于训练文本匹配的意图识别创建 1-3 个预测结果;
- Step4: 综合上述预测结果, 根据投票法判断最终意图类别;
- Step5: 将用户意图类别作为用户原问题的提示输入给大模型;
- Step6: 顺利执行则返回结果, 否则就提醒用户重新组织语言描述问题。

通过引入意图识别增强策略, 确保在 Agent 的运行过程中, 能够准确地对执行外部函数的调用, 保证运行的稳定性和准确性。

#### 4.2.5 大模型驱动洪涝数据及模型实现

大模型驱动洪涝数据及模型的实现是以提示工程技术为基础, 通过 Agent 工具函数的编写, 调取并串联模型组件库的组件、获取并展示数据库里的数据, 具体的实现流程如图 4-13。

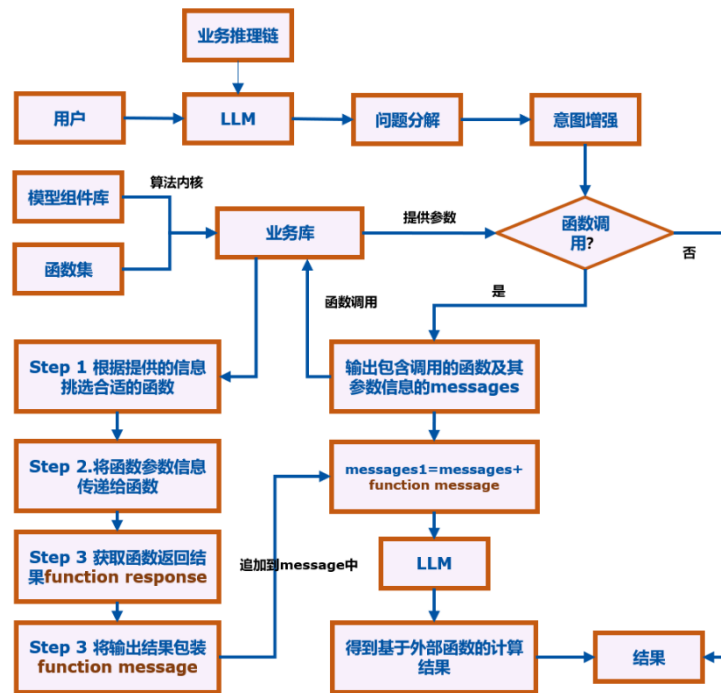


图 4-13 大模型驱动外部函数实现流程

Fig. 4-13 Large language model drives the external function implementation process

首先通过业务推理链分解用户原始问题, 得到子问题后利用意图识别增强技术解析用户的意图类别, 判断是否需要执行函数调用策略, 如果需要调用函数, 则大模型返回解析出的函数名称及参数, 然后通过执行函数库里相应的函数, 将计算结果以上下文的形式 (message) 传递给大模型, 使大模型了解函数的计算结果, 重新组织语言回答用户的问题, 大模型调取洪涝数据及模型示例如图 4-14 所示。

利用 Agent 技术, 将专业模型组件和外部交互工具封装成供大模型调用的工具函数,

以构建业务库的方式，使算法模型的使用方式不再局限于传统的应用模式，语言交互调用的方式能够为算法模型的应用带来更多的场景，为驱动洪涝灾害防御系统带来更大便利。

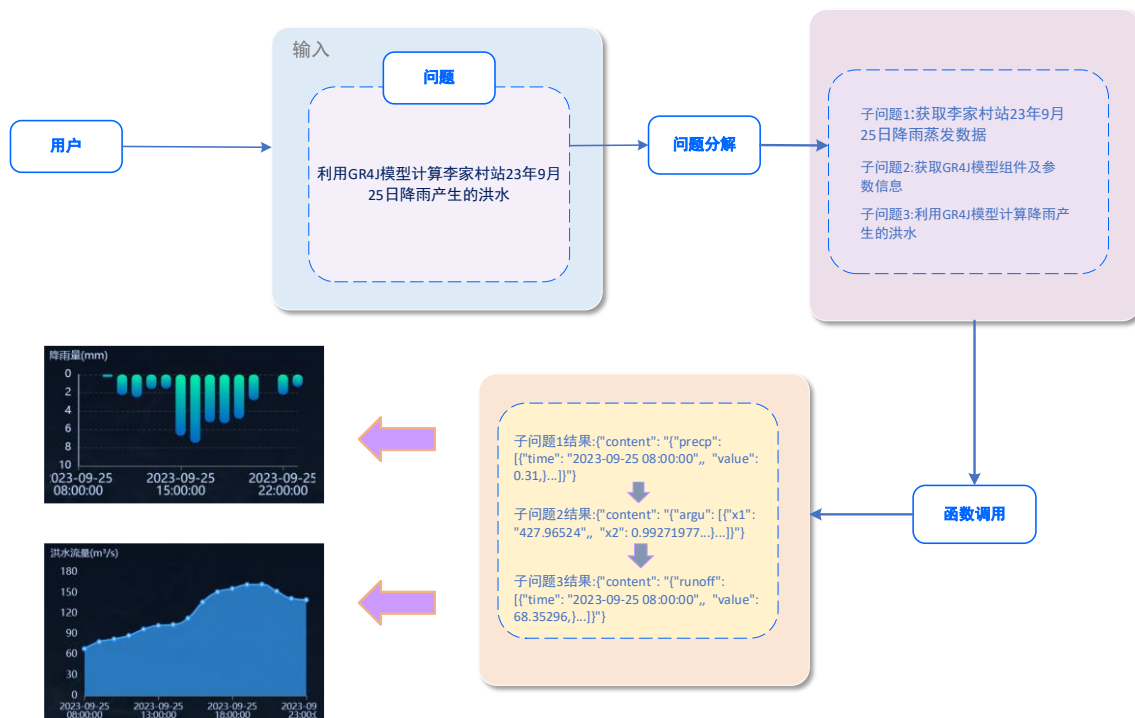


图 4-14 大模型调取洪涝数据及模型示例

Fig. 4-14 Large language model to retrieve flood data and model examples

### 4.3 大模型与业务过程融合研究

大模型与洪涝业务融合的目标在于实现与业务过程融合，而与业务过程融合在于利用大模型控制业务系统的运行过程。业务系统的运行是依赖于系统后端方法的调用以及前端的过程展示，如果能够让大模型根据用户语义调取后端的接口，再根据后端执行的方法调取前端场景，那么就可以实现通过大模型直接操控系统的目的，简化业务人员对于系统的学习成本。

实现大模型驱动业务系统的流程如图 4-15 所示，其本质过程与实现大模型调用外部函数的过程相似。通过语义解析理解用户问题，判断是否需要调用操作系统来解决问题。如果需要，就调用编写的服务接口。相较于普通的外部函数调用，引入了调用接口后，后端向前端发送响应的步骤。具体而言，一旦后端完成接口调用，将利用 **WebSocket** 技术向前端发送响应。前端会按照匹配响应规则的方式调用相应的前端方法，以实现基于业务场景的展示。

在大模型与业务过程的融合中，采用了以大模型为主导的系统驱动方式，为业务运行提供了高效而智能的支持。通过与系统后端的深度集成，大模型能够精准地调取后端接口方法，实现对操作系统的直接操控。大模型驱动业务系统的优势在于显著降低了业务人员的学习成本，通过大模型的引导，业务人员无需研究复杂的系统运行机制，而是可以专注于业务本身，实现业务人员与系统之间的高效沟通和协同。

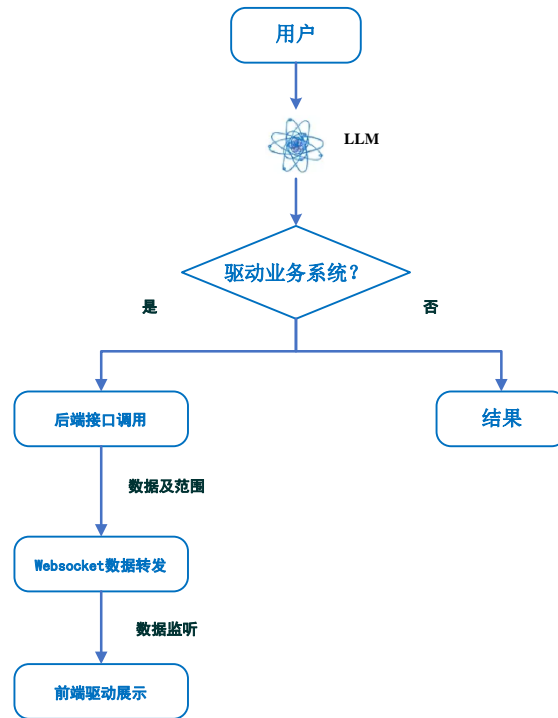


图 4-15 大模型驱动业务系统实现流程

Fig. 4-15 Large language model drives the implementation process of business system

#### 4.4 本章小结

本章针对大模型驱动下的洪涝业务应用模式展开研究。首先，通过探讨大模型知识融合方法，将知识数据向量化存储，并采用 RAG 策略实现对预案、方案等知识数据的在线管理。其次，研究了大模型如何与实时数据和专业模型实现融合，通过构建业务推理链和工具函数库，利用意图识别增强技术，使大模型能够驱动专业模型，实现更高效的交互。最后，对大模型与业务过程的融合进行了详细讨论，并提出了实现大模型驱动业务系统的具体方法。

## 5 洪涝灾害防御数字孪生平台的构建与应用

根据前几章的理论基础，本章介绍了系统的孪生场景搭建、前端设计和后端服务的搭建，并对系统应用开展研究，以进一步展示系统的功能和应用价值。

### 5.1 数字孪生场景构建

#### 5.1.1 流域地形可视化

流域地形渲染是三维场景模拟的核心，它使得山川河流的起伏得以直观展现，为三维数字孪生环境提供了基础支撑。实现三维流域地形的可视化需要将 DEM 数据与遥感影像数据相结合，并在 WebGIS 平台上进行渲染。本系统使用的 DEM 数据和遥感影像数据来源于中国科学院计算机网络信息中心地理空间数据云平台（www.gscloud.cn），利用 WebGL 的开源框架 Cesium 来实现地形的三维可视化。

在可视化遥感影像数据时，首先通过 Nginx 服务器将影像瓦片按照等级（z）和坐标编号（x，y）暴露为 URL 服务，然后利用 Cesium 提供的 UrlTemplateImageryProvider 接口解析服务地址实现。对于高程数据，同样需要通过 Nginx 服务器将 DEM 数据的存储位置暴露为 URL 服务，然后利用 Cesium 提供的 CesiumTerrainProvider 接口解析地址实现。流域地形可视化效果示例如图 5-1 所示。

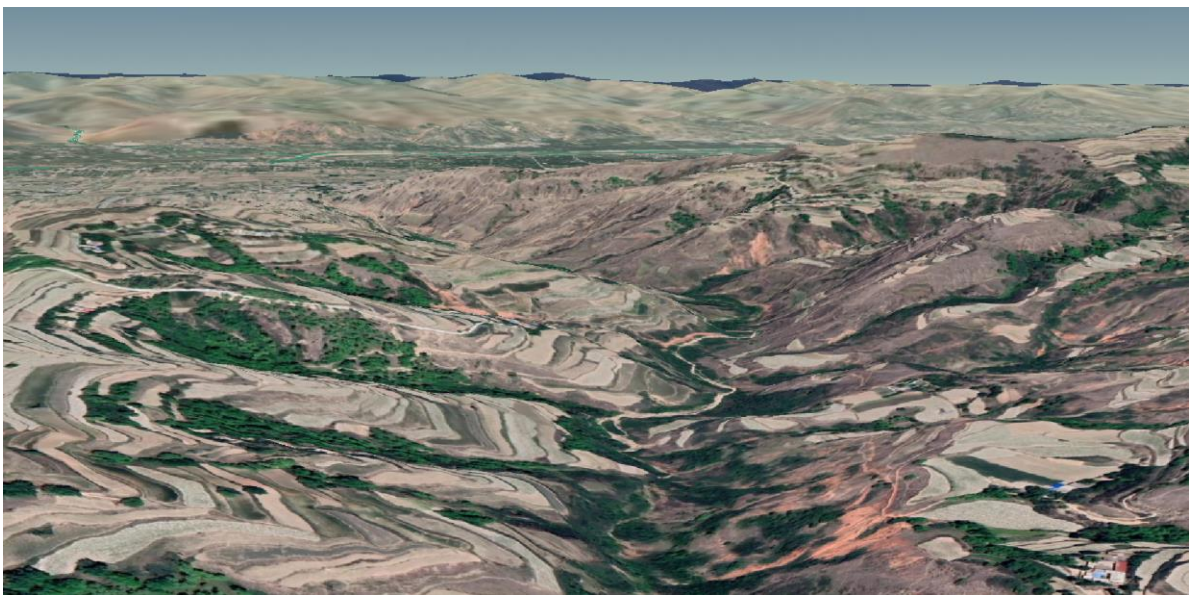


图 5-1 三维地形可视化示例图

Fig. 5-13D terrain visualization example map

#### 5.1.2 流域地物模型可视化

流域地物模型的可视化是指在三维地形上对地物模型进行渲染和展示。本系统涉及的模型主要包括水坝模型和防护区楼房模型，数据来源于倾斜摄影建模和三维软件建模，倾斜摄影建模的原始数据来自无人机航拍，通过导入建模软件，使用计算机视觉以及三维重



建技术进行图像处理，生成真实的三维实景模型。三维软件建模方面，本文采用导入 CAD 的方式实现水库模型的建模处理，通过捕捉 CAD 图纸线条，进行线条的拉伸变换，并对拉伸后的线条进行闭合处理，实现立体水库模型的构建。

为了在三维地形上渲染这些模型，首先利用 Nginx 服务器将模型存储路径作为服务地址公开，然后利用 Cesium 提供的 Cesium3DTileset 接口解析模型服务地址。最终处理后的水库三维模型可视化效果如图 5-2 所示，防护区楼房的三维模型可视化效果如图 5-3 所示。



图 5-2 水库三维模型可视化

Fig. 5-2 Visualization of reservoir 3D model



图 5-3 防护区楼房三维可视化

Fig. 5-3 3D visualization of buildings in protected areas

### 5.1.3 流域水系可视化

在构建孪生流域的过程中，河流水系数据的准确可视化是展示支干流关系、实现流域水网展示和洪水演进模拟等关键环节的基础。本文使用的河流水系数据来源于对临洮县 DEM 高程数据的提取，然后利用 GeometryInstance 接口对河流数据进行实例化，并以图元

(primitives) 的形式将实例化后的数据添加到三维视图中，通过改变图元的材质 (Material) 属性对材质的颜色、纹理背景以及水流震动速度等属性进行调整，以实现河流水系动态效果的真实模拟。河流水系可视化示意如图 5-4 所示。



图 5-4 河流水系可视化

Fig. 5-4 River system visualization

## 5.2 前端架构搭建

系统前端搭建主要包括页面布局设计、视图切换、信息展示以及与后端算法的有效通信。在前端开发中，本文采用了 Vue3 作为主要语言框架，使用 JavaScript 来编写前端业务逻辑，并利用 Cesium 作为三维地图的渲染引擎。为了管理前端产生的临时数据，采用 Pinia 作为状态管理中间件，同时使用 Vue Router 来控制页面跳转。在用户界面 (UI) 设计方面，系统结合了 Element plus 和 HTML/CSS 技术。为了实现数据可视化，系统采用了 Apache ECharts 和 vis.js 两种技术，以提供丰富的数据可视化展示。在网络数据交互方面，采用 Axios 和 WebSocket 与后端数据进行交互。系统的登录页面如图 5-5 所示。

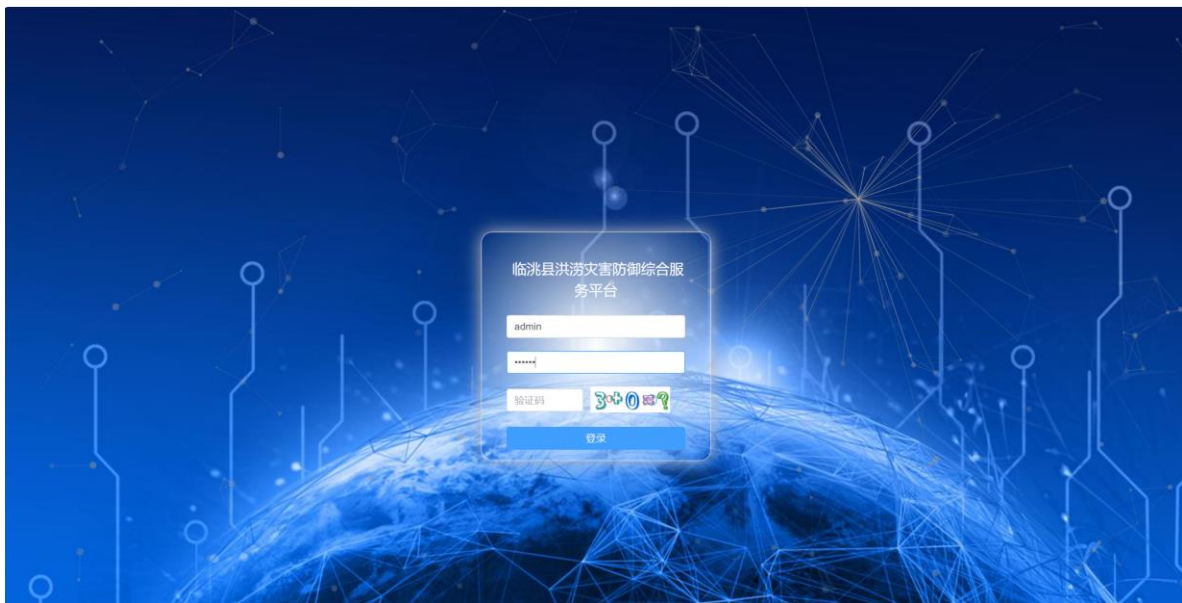


图 5-5 系统登录页面

Fig. 5-5 System login page

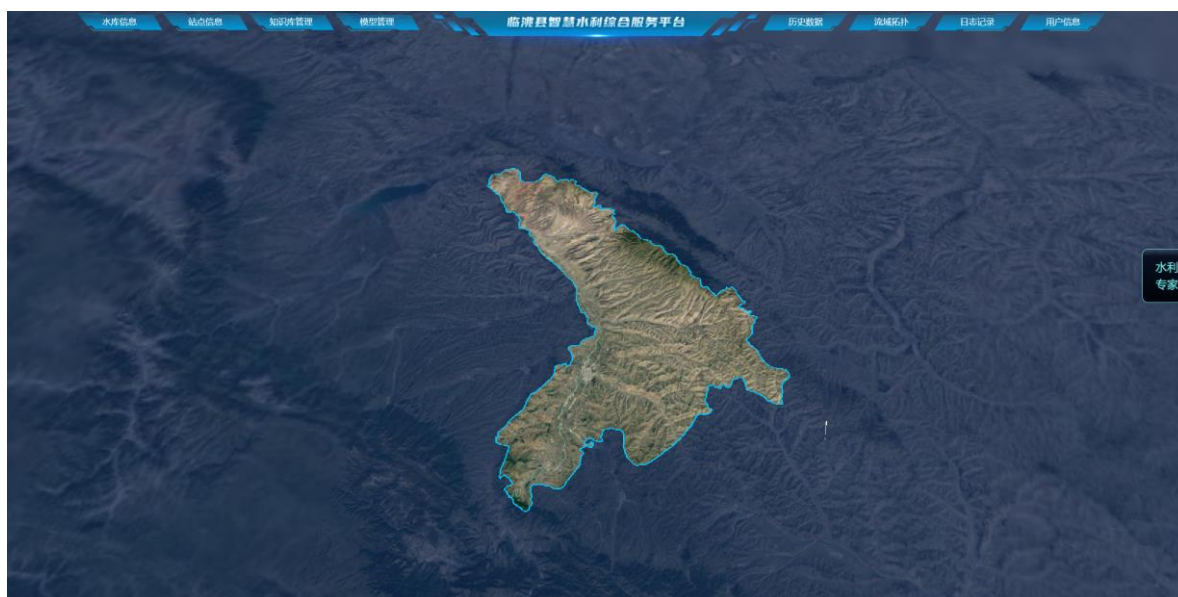


图 5-6 系统底板展示

Fig. 5-6 System backboard display

系统的登录页面采用提交表单的方式进行用户验证，包括用户名、密码和验证码的校验。成功登录后，用户将被重定向到系统主页面；若登录失败，系统将显示错误信息。如图 5-6 所示，前端的整体布局包括顶部的操作菜单、右侧的与大模型交互的接口，其余部分用于展示孪生地图。操作菜单中包含系统孪生场景中水库、站点信息管理、模型数据管理、知识库管理、系统历史数据管理、流域拓扑图管理、日志记录和用户信息管理等选项。每个菜单项都是一个 Vue 组件，组件内部可以包含其他子组件，通过组件的切换来实现前端内容的展示。

### 5.3 后端架构搭建

系统后端开发分为三个主要部分：数据服务端、专业模型端以及大模型端。数据服务端和专业模型端通过 gRPC 进行后端数据的通信，所有后端服务都以暴露端口请求的形式供前端或其他服务调用。数据服务端负责处理业务流程计算和数据库交互，而专业模型端则以组件形式封装算法模型，以便于数据服务端的调用。大模型端负责对用户语义进行识别，调用相应的工具类，执行数据服务端提供的接口方法，并将结果响应给前端，以实现业务目的。

#### 5.3.1 数据服务端搭建

数据服务端使用 Java 语言编写，基于 Spring Boot 框架，将业务代码划分为四个层次：Entity 实体层用于封装数据库表中的数据；Mapper 持久层负责与数据库表的交互；Service 业务层负责业务模块的逻辑设计；Controller 控制层控制业务流程。数据服务端的目录结构如图 5-7 所示。系统采用 Maven 作为项目管理工具，确保系统开发的统一标准，避免 jar 包版本冲突，如图 5-8 所示。

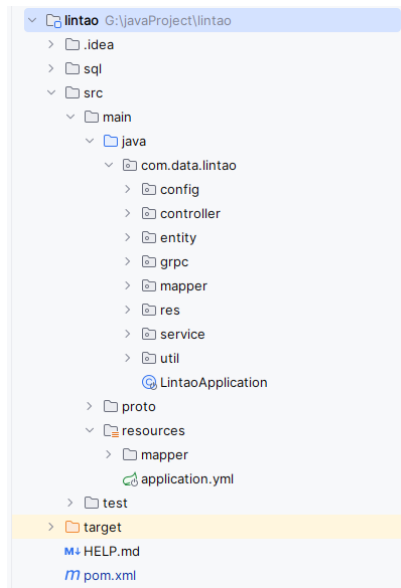


图 5-7 数据服务端目录结构

Fig. 5-7 Data server directory structure

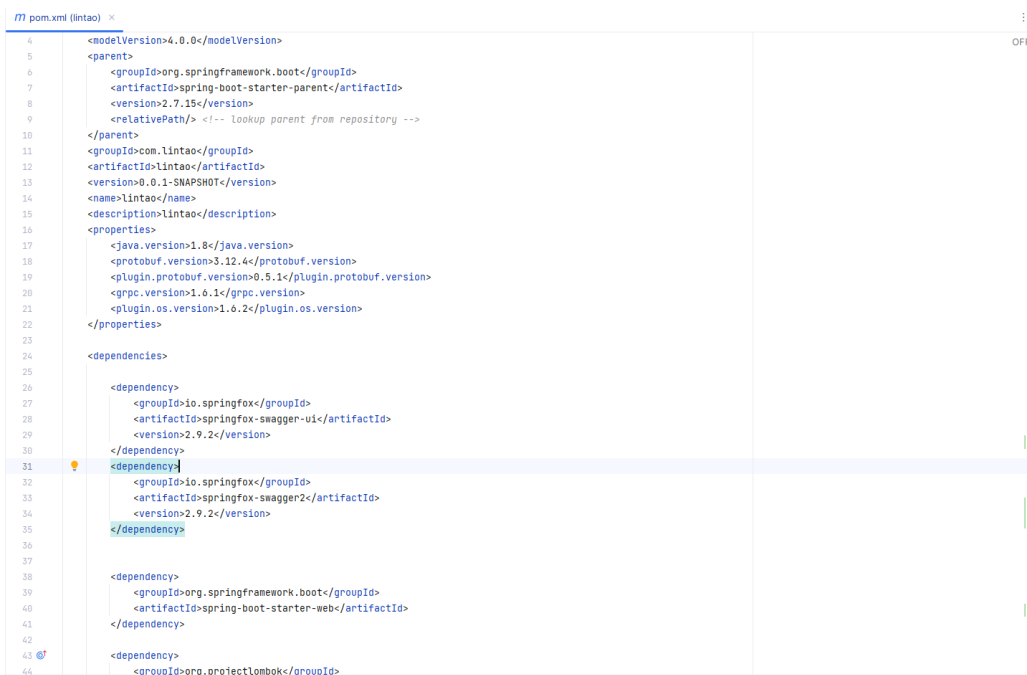


图 5-8 maven 项目管理界面

Fig. 5-8 maven project management interface

Java 后端接口采用 RESTful 风格开发，将业务操作分为创建、删除、获取和修改四类，这些操作与 HTTP 协议的 POST、DELETE、PUT、GET 请求相对应。在 Controller 层，通过使用 @PostMapping、@DeleteMapping、@PutMapping 和 @GetMapping 注解，添加相应的请求访问规则。请求数据的传输采用 JSON 格式，以提高网络传输效率<sup>[78]</sup>。为了详细描述系统前后端通信接口的地址和详情，系统采用 Swagger 框架，以明确接口的接受参数。Swagger 接口详情信息如图 5-9 所示。



图 5-9 Swagger 接口展示界面

Fig. 5-9 Swagger Interface Displays the interface

### 5.3.2 业务算法模型端搭建

在业务算法模型端，采用 Python 语言进行开发，旨在为系统提供强大的模型计算能力。将模型算法与服务端分离，不仅降低了服务部署和维护的成本，而且 Python 语言的丰富内置运算框架也显著减少了编码的复杂性，并提升了运算效率。项目目录结构如图 5-10 所示。

业务算法模型端主要依赖于 Numpy、Pandas、Scipy 等基础框架和 pymoo 算法库。通过将模型算法的输入和输出封装为 Protobuf 的数据交换格式，以 RPC 通讯的方式，将每个模型组件发布为服务接口，定义模型接口的数据类型，实现与 Java 端进行通讯，模型输入参数封装如图 5-11 所示。达到通过调用 Java 端的 RESTful 接口，调取相应的模型服务。

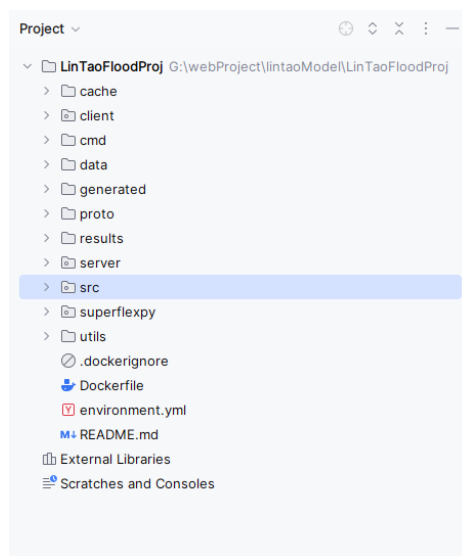


图 5-10 Python 模型端目录结构

Fig. 5-10 Python model-side directory structure

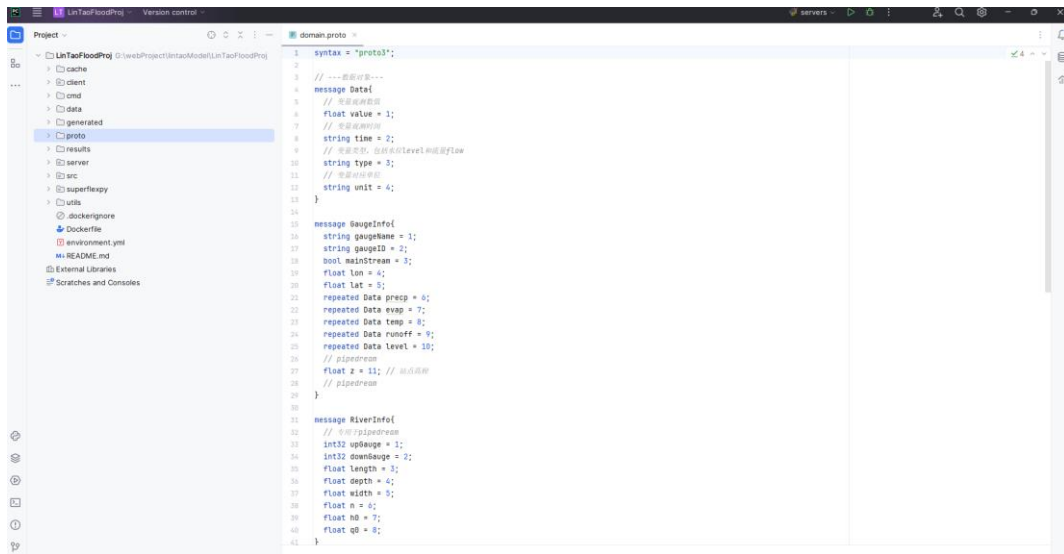


图 5-11 模型输入参数封装示例

Fig. 5-11 Model input parameter encapsulation example

### 5.3.3 大模型端搭建

大模型板块同样采用 Python 语言进行开发，主要是因为目前主流的大模型都是基于 Python 语言构建的，同时模型内部的函数接口设计也是依据 Python 语言开发的。在系统构建中，本文以国产模型 chatglm4 为基础，利用 langchain 框架提供的 retrieval 服务进行知识库的开发。借助 chatglm4 模型内部提供的函数调用形式，结合 pandas、numpy、pymysql 等基础框架，实现 Agent 服务的搭建，开发页面如图 5-12 所示。通过 FastAPI 将大模型板块提供的模型对话、知识库管理等功能以 Web API 的形式，提供公共的服务接口，如图 5-13 所示，以供用户使用。

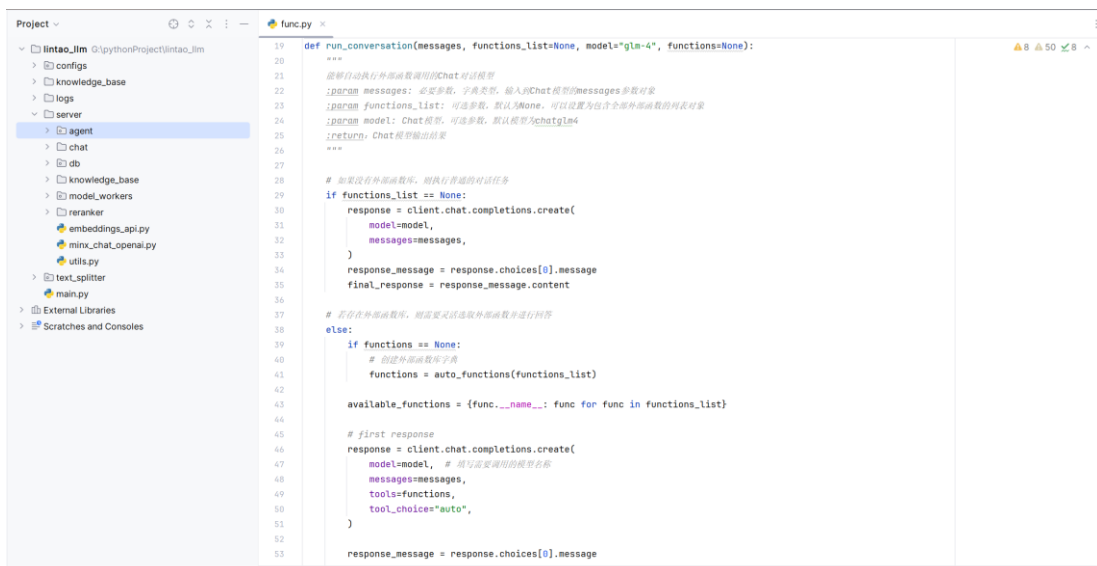


图 5-12 大模型服务端开发页面

Fig. 5-12 Large model server side development page

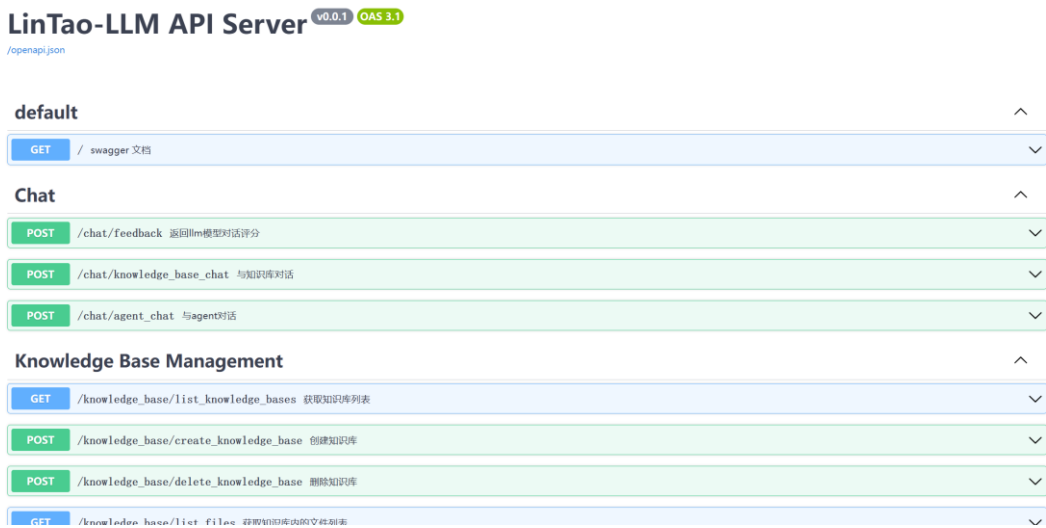


图 5-13 大模型服务端接口展示

Fig. 5-13 Large model server interface presentation

## 5.4 洪涝灾害防御业务应用

### 5.4.1 数据管理

数据管理是确保系统内各类信息有效组织和访问的关键环节，负责对水库信息、测站信息、模型数据、历史数据、流域拓扑节点数据以及知识库数据进行集中管理，每个数据管理模块都配备了独立的管理界面。

#### (1) 水库信息管理

水库的信息管理主要是对临洮县内水库基本信息、三维模型、监测数据的影子设备地址以及其他一些水库的信息进行管理。用户可以通过新增水库的信息，实现三维 GIS 地图上新增水库的孪生展示，在监测面板上查看水库的监测信息，也可通过点击水库对应监测数据，查看历史的监测信息，便于业务人员对水库信息的管理，页面展示如图 5-14。



图 5-14 水库信息管理界面

Fig. 5-14 Reservoir information management interface

## (2) 测站信息管理

测站的信息管理包括对临洮县内的水文站、雨量站、流量站等监测站点的信息进行管理。主要包括站点的名称、位置、监测数据源、监测数据阈值等信息。用户可通过新增站点的经纬度、站点的基本信息以及站点监测设备影子地址，实现在三维 GIS 上新增站点的信息，点击后展示监测设备最新的监测信息，站点信息管理界面如图 5-15，站点监测信息展示如图 5-16。

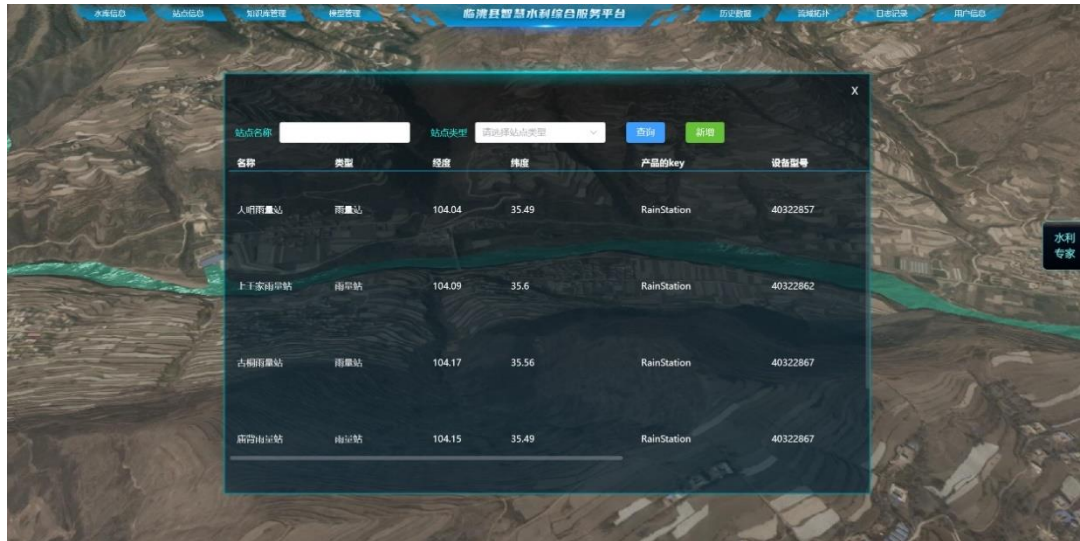


图 5-15 站点信息管理界面

Fig. 5-15 Site information management interface



图 5-16 站点监测信息展示界面

Fig. 5-16 Site monitoring information display interface

## (3) 模型数据管理

模型数据管理主要包括对系统内现有模型的数据进行展示，包括模型的名称、模型类型、模型路径以及模型说明，如图 5-17 所示。





图 5-17 模型管理界面

Fig. 5-17 Model management interface

(4) 历史数据管理

历史数据管理包括系统的在不同情况下的预演记录、方案发布记录以及问题处理记录。便于业务人员遇到问题时能够参考以往成功的解决方案，实现高效准确的应对风险。

(5) 流域拓扑节点数据管理

流域拓扑节点数据管理专注于临洮县洮河流域的关键节点信息，进行拓扑结构的维护。允许用户通过拖拽方式新增节点信息或修改节点位置，从而实现对流域基本信息的直观理解，并为后续的预演环节提供了数据支持，使得用户能够获得预演中间断面的信息，确保了数据来源的准确性和预演流程的可追溯性，流域拓扑节点数据管理界面如图 5-18。

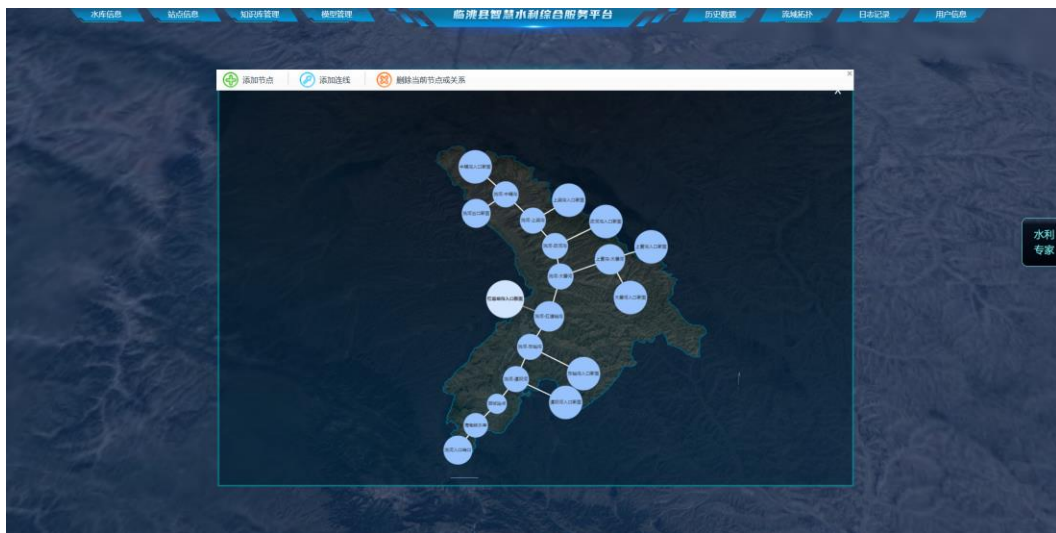


图 5-18 流域拓扑节点数据管理界面

Fig. 5-18 Management page for basin topology node data

(6) 知识库数据

知识库数据管理主要是对大模型知识库里的内容进行管理，提供新增知识分类，管理

分类下知识内容的功能。用户通过在分类下添加知识文件，系统自动根据文件的类型，将文件进行向量化处理，并存储于向量数据库中。

#### 5.4.2 综合监控面板

综合监控面板是实时监测数据的展示平台。通过整合水情、雨情等信息，利用监测设备的在线影子地址，收集实时数据，在已建立的数据底板上展示最新时刻监测点的数据。如图 5-19 所示，左上卡展示重点断面的雨晴情况，包括最大降雨量、流量等信息，并根据降雨量大小给出降雨等级，达到暴雨级别时发出红色警报。左二卡展示基于构建流域下的流量站实时采集的流量信息，以及旱警流量情况。下一卡展示洪水的主要来源以及洪水大小。下二卡展示流量的最大小时降雨量和最大河道流量，并结合天气接口给出未来 7 天的降雨量预测。右下卡则展示临洮县内水库的防洪库容使用情况。



图 5-19 数据看板界面

Fig. 5-19 Data dashboard interface

#### 5.4.3 实时洪水预报

实时洪水预报分为流域滚动预报和关键断面实时洪水预报两个部分。流域滚动预报在用户向大模型输入启动指令后开始，如图 5-20 所示，系统通过接入实时雨情等信息，调用产汇流和洪水演进模型，以预设的滚动周期对临洮县洮河支干流进行洪水预报。如图 5-21 所示，最下方的图表展示了洮河干流的洪水过程线以及支流与干流交汇处的洪水警戒线。当干流洪水超过警戒线时，容易发生倒灌，系统会在左侧提供预警提示。



图 5-20 大模型指令输入示例  
Fig. 5-20 Large model instruction input example

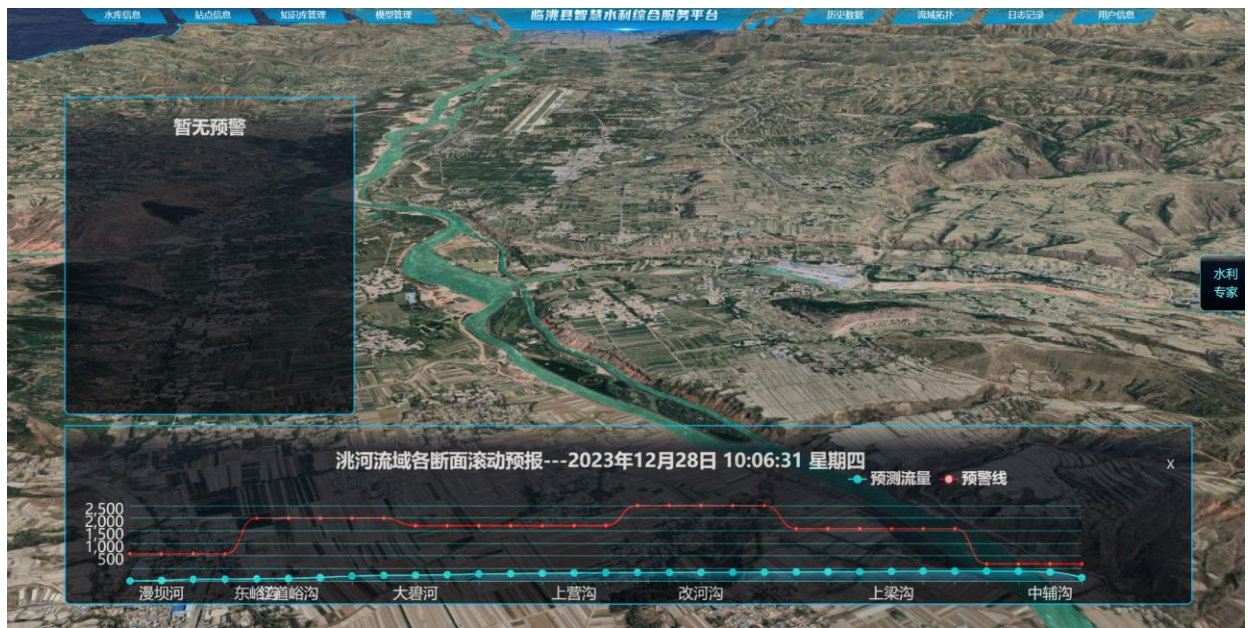


图 5-21 滚动预报界面  
Fig. 5-21 Rolling prediction interface

关键断面实时洪水预报可以通过两种方式进行：一是直接在数据看板上点击关键断面，选择洪水预报功能；二是用户向大模型输入断面洪水预报指令，大模型依据用户语义将判断预报断面，并获得洪水预报结果。如图 5-22 所示，水文站 1 位于水文站 2 的上游，水文站 1 显示流域上游暴雨产生的汇流计算后实时的流量过程线。水文站 2 的红色部分显示了该站的实测流量过程，而绿色虚线部分则展示了模型计算出的洪水过程。



图 5-22 关键断面洪水预报界面  
Fig. 5-22 Flood forecast interface of key sections

#### 5.4.4 过程透明化预演

过程透明化预演分为洪水预演和调度预演两部分。洪水预演可以通过两种方式启动：一是用户在数据看板上点击关键断面，选择洪水预演，并调整降雨情况后开始；二是用户通过语言交互方式告知大模型特定的降雨情况，例如输入“xx 断面发生小时降雨量为 xx 时，会发生什么情况”，随后系统开始洪水预演。系统首先从数据库中获取预演断面的拓扑关系，并将整个预演过程划分为多个子断面，每个子断面代表一个拓扑节点。然后利用产汇流模型计算第一个断面的产汇流结果。用户可以对计算过程进行人为修正，通过改变模型参数或直接改变预报的洪水过程，获得新的结果，如图 5-23 所示。

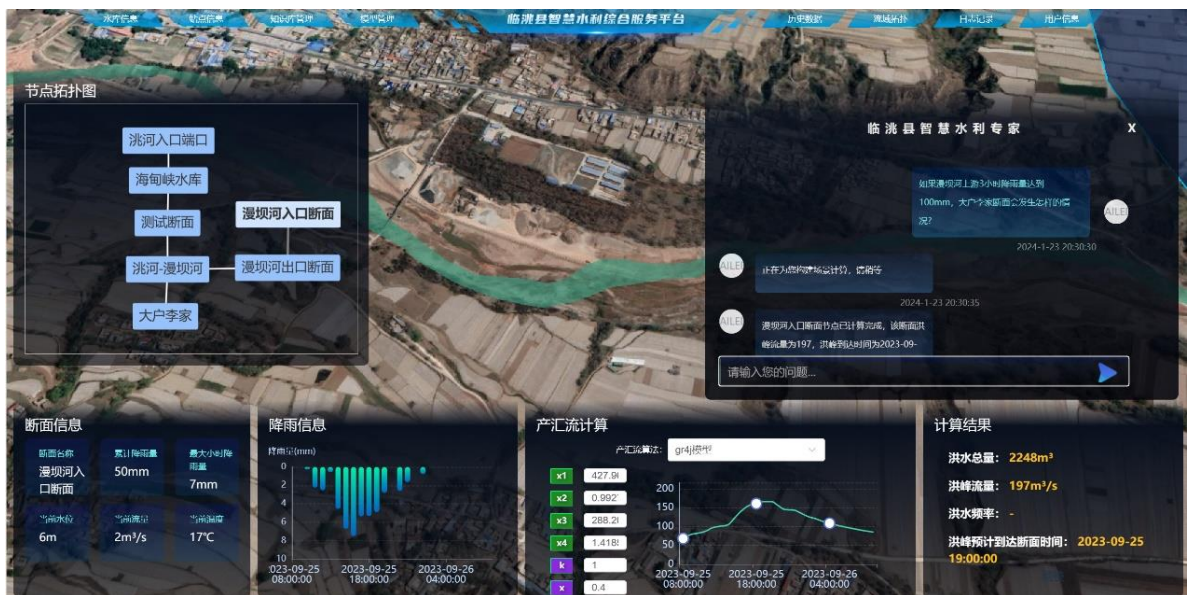


图 5-23 洪水预演（初始节点计算结果）界面展示  
Fig. 5-23 Flood preview (initial node calculation results) interface display

当对第一个断面的修正完成后，用户输入“继续计算”，系统将根据此次修正的洪水过程，以演进的方式计算到下一个断面，如图 5-24。整个预演计算过程透明且可调节，使用户能够清晰地看到预演结果的变化过程。



图 5-24 洪水预演（中间断面）界面展示

Fig. 5-24 Flood preview (middle section) interface display

在洪水预演过程中，当模拟到达水库断面时，会根据上游洪水演进的结果计算出水库的入库流量过程。基于这一入库流量，调用优化算法得到水库的出库流量过程，如图 5-25。在整个调度期间，模型算法的参数都是可以调整的，以便根据实际情况优化洪水管理策略。



图 5-25 洪水预演（水库节点）界面展示

Fig. 5-25 Flood preview (reservoir node) interface display

当洪水预演计算节点到达防护对象时，会根据计算得到的洪水过程线确定洪峰流量的大小，得到洪水频率，从而评估洪水发生的可能性，如图 5-26 所示。此外，还会展示在最大洪峰流量下的淹没范围，帮助决策者理解洪水可能影响的区域，为洪水风险管理和应急响应提供关键信息。

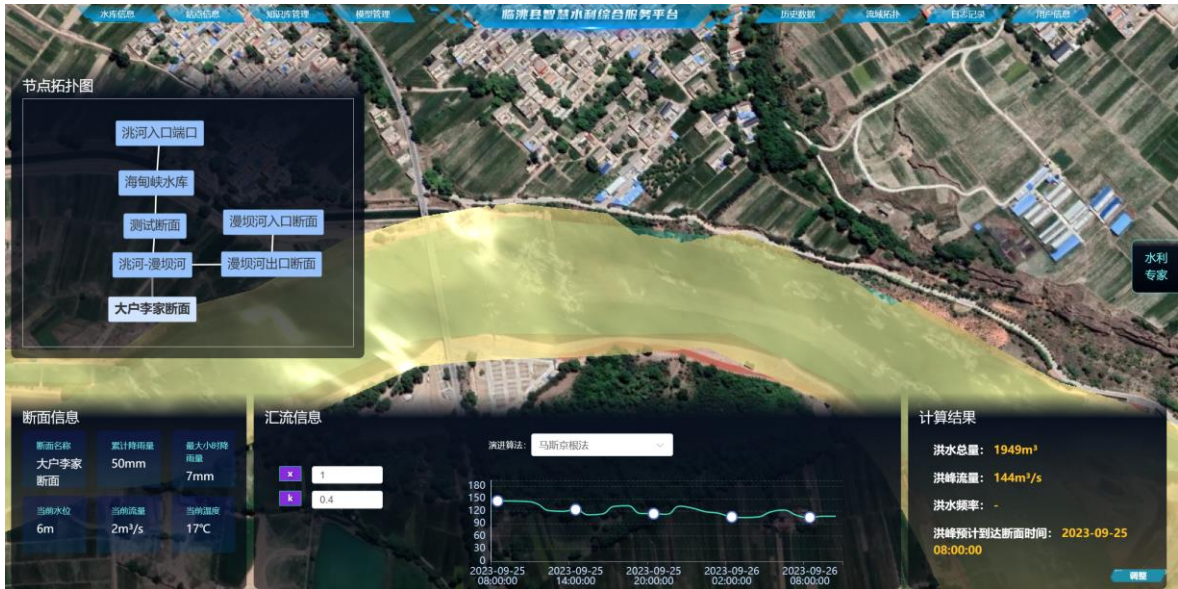


图 5-26 洪水预演（防护断面）界面展示

Fig. 5-26 Flood preview (protection section) interface display

调度预演同样有两种开启方式，一是在数据看板中点击对应水库进入调度预演，二是通过大模型语言交流开启。通过调整入库流量和调度模型参数，获得下泄流量过程，以演进的方式计算此次下泄过程对下游目标的影响，计算过程同样是全透明且可调节的。

#### 5.4.5 智能化预案



图 5-27 预案内容搜索回答示意图

Fig. 5-27 Plan content search answer diagram

通过配置知识库，可以实现与大模型以语言交流的形式查询知识库中的内容。用户可以直接向大模型提出相关问题，然后大模型会从知识库中提取最符合的内容进行回答，如图 5-27 所示。此外，在进行预演环节时，系统会根据最终防护对象的洪水过程，依据洪峰大小，从所配置的知识库中智能推荐相应的预案内容。

### 5.5 本章小结

本章主要对系统的开发细节以及系统的应用进行了介绍。从系统的开发层面出发，介绍了系统的孪生场景搭建、前端页面设计以及系统后端的搭建，其中后端包括 java 数据服务端、python 算法模型端、python 大模型端三个方面，最后对系统的核心功能进行了仿真应用，包括数据管理、综合监测面板、实时洪水预报、过程化预演以及智能化预案，展示了系统开发的合理性。

## 6 结论与展望

### 6.1 结论

洪涝灾害是全球最常见的自然灾害之一，对人类社会和经济造成了巨大的影响，随着气候变化的不断加剧，使得洪涝灾害防御成为当今社会的紧迫问题。目前正在使用的防御系统，虽相比于基于经验和工程模型的计算有了很大提升，但仍有很大的提升空间，依然存在应用功能分散、衔接困难、宏观模拟不足等问题。为了解决这些问题，本文以甘肃省临洮县为研究对象，将大模型和数字孪生技术运用进洪涝灾害防御系统，探究基于大模型的洪涝“四预”流程，并以此为依据对系统进行开发实现，论文研究成果如下：

(1) 提出了一套融合大模型和数字孪生技术的洪涝灾害防御系统架构。目前技术体系下洪涝灾害防御系统，系统存在功能分散、衔接困难、宏观模拟不足等问题，通过对问题的剖析和技术体系的研究，明确了新时代洪涝灾害防御系统应具备的功能和技术架构，为洪涝灾害防御系统的进一步发展提供参考。

(2) 建立了基于组件思维的多模型集成体系。应用于洪涝灾害防御的专业模型众多，每个模型都具有自身的特定条件，但不少模型之间也拥有着共性，在专业模型的集成研究过程中，依据模型内部结构和逻辑上的特征，利用组件思维提取模型算法中的共性，构建相互独立的基础模块，以标准化接口的方式将基础模块按照输入输出约束进行拼接，构建中粒度模型，形成应用组件库，实现专业模型的集成开发，解决模型开发效率低，代码复用率差的问题。

(3) 提出了大模型驱动洪涝知识-数据-模型-系统的应用模式。支撑洪涝“四预”应用依赖于数据、模型计算、专家经验和业务规则的紧密结合。通过对知识融合方法的研究，采用 RAG 策略将多源异构信息利用大模型进行统一管理，成功构建一个业务“知识场”，解决了知识和数据的串联仍需要由人通过线下讨论的方式逐步对齐的问题，提高了知识数据的检索效率。同时，通过对智能体（agent）的研究，以构建洪涝业务推理链、大模型业务库和意图识别增强策略的方式，提出大模型驱动洪涝数据和模型的应用模式，并从系统设计角度，提出了新时代业务系统的驱动方式，简化了模型及系统的应用方式，减少了业务人员额外的学习成本，并为进一步创建人机交互的智能平台提供思路。

(4) 对大模型与数字孪生技术融入洪涝灾害防御系统进行落地实现。通过对理论知识的研究，利用 Java、Python 等语言成功实现了将大模型和数字孪生技术融入洪涝灾害防御系统，并通过集成实时洪水预报、透明化预演、智能化预演等功能实现洪涝灾害防御的业务应用。

### 6.2 展望

本文实现了甘肃省临洮县洪涝灾害防御系统的开发，但由于洪涝灾害防御是一个复杂、



庞大的系统，本文开发的平台还有许多不足之处，还待在今后的工作中进一步补充，具体如下：

(1) 本文只对新安江模型、GR4J 模型、马斯京根模型等进行了划分，但实际洪涝灾害防御涉及的预报模型、演进模型以及调度模型远不止这些。模型的组件越丰富，系统适用性也就越强，丰富模型组件是后续需要完善的功能。

(2) 本文设计开发的洪涝灾害防御系统在实际演示及应用中取得了良好的效果，但由于 WebGL 渲染的原因，在有些孪生场景渲染时会加载缓慢。因此，如何在保证渲染精度的同时平衡系统流程，提高系统的稳定性和流畅性，仍需要进一步研究。

(3) 论文虽采用提出了大模型驱动小模型以及驱动业务系统的应用模式，但该种方式对系统有强耦合性，后续添加模型及业务功能后需要修改给予大模型的提示，以及前端需要展示的页面，如何编排一套通用的业务流程描述标准，构建专用组件库，让大模型根据业务规则生成业务链并编排组件，实现自动生成业务系统，是后续值得考虑的问题。

## 参考文献

- [1] 中华人民共和国水利部编. 2023 中国水利发展报告[M]. 北京: 中国水利水电出版社, 2023.06.
- [2] 李勇, 李登文. 湄江河流域洪涝监测预警系统[J]. 贵州气象, 2007, 31(1): 31-34.
- [3] 李宗花, 叶正伟. 洪泽湖流域洪涝灾害预警系统的构建[J]. 水利水电技术, 2009 (7): 111-113.
- [4] 朱建英, 王正诚, 周春飞, 等. 水旱灾害防御调度系统思维分析[J]. 中国防汛抗旱, 2023, 33(02): 83-86.
- [5] 杨芳, 宋利祥, 李旭东, 等. 珠江流域水旱灾害防御“四预”系统模型研发及应用[J]. 中国水利, 2023, (04): 10-14.
- [6] Zaalberg R, Midden C J H. Living behind dikes: mimicking flooding experiences[J]. Risk Analysis, 2013, 33(5): 866-876.
- [7] 郑晓阳, 胡传廉. 上海市防汛决策支持系统设计[J]. 水利水电科技进展, 2003(01): 25-27.
- [8] Macchione F, Costabile P, Costanzo C, et al. Moving to 3-D flood hazard maps for enhancing risk communication[J]. Environmental modelling & software, 2019, 111: 510-522.
- [9] 韩刚, 王常效, 刘业森, 等. 深圳市洪涝预警调度系统建设方案及实现[J]. 中国防汛抗旱, 2020, 30(11): 14-19.
- [10] Grieves M W. Product lifecycle management: the new paradigm for enterprises[J]. International Journal of Product Development, 2005, 2(1-2): 71-84.
- [11] Grieves M, Vickers J. Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems[J]. Transdisciplinary perspectives on complex systems: New findings and approaches, 2017: 85-113.
- [12] Glaessgen E, Stargel D. The digital twin paradigm for future NASA and US Air Force vehicles[C]//53rd AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics and materials conference 20th AIAA/ASME/AHS adaptive structures conference 14th AIAA. 2012: 1818.
- [13] Boje C, Guerriero A, Kubicki S, et al. Towards a semantic Construction Digital Twin: Directions for future research[J]. Automation in construction, 2020, 114: 103179.
- [14] Wang M, Wang C, Hnydiuk-Stefan A, et al. Recent progress on reliability analysis of offshore wind turbine support structures considering digital twin solutions[J]. Ocean Engineering, 2021, 232: 109168.
- [15] Onile A E, Machlev R, Petlenkov E, et al. Uses of the digital twins concept for energy services, intelligent recommendation systems, and demand side management: A review[J]. Energy Reports, 2021, 7: 997-1015.

- [16]Milton M, De La O C, Ginn H L, et al. Controller-embeddable probabilistic real-time digital twins for power electronic converter diagnostics[J]. IEEE Transactions on Power Electronics, 2020, 35(9): 9850-9864.
- [17]李德仁, 龚健雅, 邵振峰. 从数字地球到智慧地球[J]. 武汉大学学报 (信息科学版), 2010, 35(2): 127-132.
- [18]张秋文, 张勇传, 王乘, 等. 数字流域整体构架及实现策略[J]. 水电能源科学, 2001, 19(3): 4-7.
- [19]王忠静, 李宏益. 数字化流域及其在现代水资源规划中的应用[J]. 水利水电技术, 2004, 35(5): 25-28.
- [20]冶运涛, 蒋云钟, 梁犁丽, 等. 数字孪生流域: 未来流域治理管理的新基建新范式[J]. 水科学进展, 2022, 33(05): 683-704.
- [21]王凯, 徐时进, 刘昌军, 等. 数字孪生淮河防洪“四预”系统关键技术[J]. 水利信息化, 2023, (06): 1-4+31.
- [22]李卫政, 谈娟娟, 张语涵, 等. 数字孪生甬江流域水资源“四预”应用研究[J]. 水利信息化, 2023, (05): 1-6.
- [23]李文学, 寇怀忠. 关于建设数字孪生黄河的思考[J]. 中国防汛抗旱, 2022, 32(2): 27-31.
- [24]程海云. 推进长江数字孪生流域建设的水文实践与思考[J]. 中国水利, 2022, (15): 49-50.
- [25]Brown T, Mann B, Ryder N, et al. Language models are few-shot learners[J]. Advances in neural information processing systems, 2020, 33: 1877-1901.
- [26]Chowdhery A, Narang S, Devlin J, et al. Palm: Scaling language modeling with pathways[J]. Journal of Machine Learning Research, 2023, 24(240): 1-113.
- [27]Touvron H, Lavril T, Izacard G, et al. Llama: Open and efficient foundation language models[J]. arXiv preprint arXiv:2302.13971, 2023.
- [28]Sun Y, Wang S, Feng S, et al. Ernie 3.0: Large-scale knowledge enhanced pre-training for language understanding and generation[J]. arXiv preprint arXiv:2107.02137, 2021.
- [29]Du Z, Qian Y, Liu X, et al. Glm: General language model pretraining with autoregressive blank infilling[J]. arXiv preprint arXiv:2103.10360, 2021.
- [30]Tang R, Han X, Jiang X, et al. Does synthetic data generation of llms help clinical text mining?[J]. arXiv preprint arXiv:2303.04360, 2023.
- [31]Nov O, Singh N, Mann D M. Putting ChatGPT's medical advice to the (Turing) test[J]. medRxiv, 2023: 2023.01. 23.23284735.
- [32]Yang K, Ji S, Zhang T, et al. On the evaluations of chatgpt and emotion-enhanced prompting for mental health analysis[J]. arXiv preprint arXiv:2304.03347, 2023.
- [33]Singhal K, Azizi S, Tu T, et al. Large language models encode clinical knowledge[J]. arXiv preprint arXiv:2212.13138, 2022.

- [34] Malinka K, Peresini M, Firc A, et al. On the educational impact of chatgpt: Is artificial intelligence ready to obtain a university degree?[C]//Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 1. 2023: 47-53.
- [35] Susnjak T. ChatGPT: The end of online exam integrity?[J]. arXiv preprint arXiv:2212.09292, 2022.
- [36] Araci D. Finbert: Financial sentiment analysis with pre-trained language models[J]. arXiv preprint arXiv:1908.10063, 2019.
- [37] 刘桂琴,赵精珍.临洮县 2018 年度气候影响评价[J].安徽农学通报,2019,25(08):146-147.
- [38] 王帅兵,李常斌,杨林山,等.基于标准化降水指数与 Z 指数的洮河流域干旱趋势分析[J].干旱区研究,2015,32(3):565-572.
- [39] 苟斌,张燕.临洮县小流域暴雨洪水调查分析[J].陕西水利,2024,(03):92-94.
- [40] 游丽贞,郭宇春,李纯喜.Ajax 引擎的原理和应用[J].微计算机信息,2006,(06):205-207.
- [41] Pimentel V, Nickerson B G. Communicating and displaying real-time data with websocket[J]. IEEE Internet Computing, 2012, 16(4): 45-53.
- [42] Chen M, Tworek J, Jun H, et al. Evaluating large language models trained on code[J]. arXiv preprint arXiv:2107.03374, 2021.
- [43] Chen Y, Wang R, Jiang H, et al. Exploring the use of large language models for reference-free text quality evaluation: A preliminary empirical study[J]. arXiv preprint arXiv:2304.00723, 2023.
- [44] 王井腾,孙祥鹏,张文明,等.珠江防汛“四预”平台建设及应用[J].中国水利,2022,(22):39-42+20.
- [45] Sriramya P, Karthika R A. Providing password security by salted password hashing using bcrypt algorithm[J]. ARPN journal of engineering and applied sciences, 2015, 10(13): 5551-5556.
- [46] 余姜德,梁本来,冷令.大数据时代背景下《非关系型数据库》课程教学资源库建设探析[J].数字技术与应用,2023,41(10):163-165.
- [47] Cosmina I. Pivotal Certified Professional Spring Developer Exam[J]. Berkeley: Apress, 2017: 435-459.
- [48] 赵人俊编著.流域水文模拟[M].水利电力出版社,1984.
- [49] 包为民.水文预报[M].中国水利水电出版社,2009.
- [50] 杨海霞.新安江模型中初始土壤含水量分布研究[D].中国矿业大学,2019.
- [51] 罗军刚.水利业务信息化及综合集成应用模式研究[D].西安理工大学,2009.
- [52] 张莘莘,童菊秀,廖卫红,等.基于 EnKF 的新安江模型参数和变量同步估计方法[J].南水北调与水利科技,2018,016(002):51-58.
- [53] Perrin C, Michel C, Andréassian V. Improvement of a parsimonious model for streamflow

- simulation[J]. Journal of hydrology, 2003, 279(1-4): 275-289.
- [54]田德存. 基于 GR4J 和 TOPMODEL 的淮河流域水文模型构建与模拟[D].华北水利水电大学,2023.
- [55]邓鹏鑫,王银堂,胡庆芳,等.GR4J 模型在赣江流域日径流模拟中的应用[J].水文,2014,34(02):60-65.
- [56]雒文生,宋星原.洪水预报与调度[M].湖北科学技术出版社,1999.
- [57]郭俊,周建中,邹强,等.水文模型参数多目标优化率定及目标函数组合对优化结果的影响[J].四川大学学报(工程科学版),2011,43(06):58-63.
- [58]黄晓敏,雷晓辉,王宇晖,等.基于 NSGA-II的水文模型参数多目标优化研究[J].人民长江,2012,43(02):16-21.
- [59]Deb K, Pratap A, Agarwal S, et al. A fast and elitist multiobjective genetic algorithm: NSGA-II[J]. IEEE transactions on evolutionary computation, 2002, 6(2): 182-197.
- [60]江燕,刘昌明,胡铁松,等.新安江模型参数优选的改进粒子群算法[J].水利学报,2007,(10):1200-1206.
- [61]Gill M A. Flood routing by the Muskingum method[J]. Journal of hydrology, 1978, 36(3-4): 353-363.
- [62]Bartos M, Kerkez B. Pipedream: An interactive digital twin model for natural and urban drainage systems[J]. Environmental Modelling & Software, 2021, 144: 105120.
- [63]Ji Z. General hydrodynamic model for sewer/channel network systems[J]. Journal of hydraulic engineering, 1998, 124(3): 307-315.
- [64]高玉琴,周桐,马真臻,等.考虑天然水文情势的水库调度图优化[J].水资源保护,2020,36(04):60-67.
- [65]钟平安,邹长国,李伟,等.水库防洪调度分段试算法及应用[J].水利水电科技进展,2003,(06):21-23+56-65.
- [66]Liu D, Huang Q, Yang Y, et al. Bi-objective algorithm based on NSGA-II framework to optimize reservoirs operation[J]. Journal of Hydrology, 2020, 585: 124830.
- [67]Dong G, Yuan H, Lu K, et al. How abilities in large language models are affected by supervised fine-tuning data composition[J]. arXiv preprint arXiv:2310.05492, 2023.
- [68]Zhang X, Zhang X, Yu Y. ChatGLM-6B Fine-Tuning for Cultural and Creative Products Advertising Words[C]//2023 International Conference on Culture-Oriented Science and Technology (CoST). IEEE, 2023: 291-295.
- [69]Gao, Yunfan, et al. "Retrieval-augmented generation for large language models: A survey." arXiv preprint arXiv:2312.10997 (2023).
- [70]Topsakal, Oguzhan, and Tahir Cetin Akinci. "Creating large language model applications utilizing langchain: A primer on developing llm apps fast." International Conference on

Applied Engineering and Natural Sciences. Vol. 1. No. 1. 2023.

- [71] Wang H, Li Y F. Large Language Model Empowered by Domain-Specific Knowledge Base for Industrial Equipment Operation and Maintenance[C]//2023 5th International Conference on System Reliability and Safety Engineering (SRSE). IEEE, 2023: 474-479.
- [72] Johnson J, Douze M, Jégou H. Billion-scale similarity search with gpus[J]. IEEE Transactions on Big Data, 2019, 7(3): 535-547.
- [73] Liu P, Yuan W, Fu J, et al. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing[J]. ACM Computing Surveys, 2023, 55(9): 1-35.
- [74] Xi Z, Chen W, Guo X, et al. The rise and potential of large language model based agents: A survey[J]. arXiv preprint arXiv:2309.07864, 2023.
- [75] Brown, Tom, et al. "Language models are few-shot learners." Advances in neural information processing systems 33 (2020): 1877-1901.
- [76] Wei J, Wang X, Schuurmans D, et al. Chain-of-thought prompting elicits reasoning in large language models[J]. Advances in Neural Information Processing Systems, 2022, 35: 24824-24837.
- [77] Zhou D, Schärli N, Hou L, et al. Least-to-most prompting enables complex reasoning in large language models[J]. arXiv preprint arXiv:2205.10625, 2022.
- [78] Lee Junhee, Anjos Edman, Satti Srinivasa Rao. SJSON: A succinct representation for JSON documents[J]. Information Systems, 2021, 97.