**大城市暴雨积水点多源数据实时采集系统**

**工**

**作**

**报**

**告**

编制日期：2024年 4月

目 录

[第一章 项目建设概况 1](#_Toc164707645)

[1.1. 项目简介 1](#_Toc164707646)

[1.2. 项目目标和任务 2](#_Toc164707647)

[第二章 项目建设内容 3](#_Toc164707648)

[第三章 项目过程管理 4](#_Toc164707649)

[3.1. 进度管理 4](#_Toc164707650)

[3.2. 人员组织管理 4](#_Toc164707651)

[3.3. 项目质量管理 4](#_Toc164707652)

[第四章 项目建设成果 5](#_Toc164707653)

[4.1. 项目完成情况 5](#_Toc164707654)

[4.2. 项目试运行应用情况 6](#_Toc164707655)

[4.3. 项目提交文件 6](#_Toc164707656)

[第五章 尚待改进的工作及建议 7](#_Toc164707657)

# 项目建设概况

## 项目简介

全球气候变化和快速城市化打破了城市降水—汇水—排水原有的平衡, 加剧了中国城市洪涝问题, 造成了巨大的生命和财产损失。因此, 亟须探索精确、高效的对洪涝灾害事件进行监测分析方法, 提高城市防洪抗灾能力, 降低灾害损失影响。

已经有相当多的研究对洪涝灾害事件进行监测分析，诸如分析受灾范围与受灾人口，集成遥感与深度学习的方法提取积水深度等。然而传统的传感器网络提供的数据时空分辨率不足，且卫星对地表观测容易受到云层的遮挡，致使访问周期变长。与此同时，城市洪涝渍水的特点是影响时间较短，这也对数据的时间分辨率（密度）产生较高的需求。

以多源数据为代表的社会感知方法是物理感知网络的一种重要补充，主要是指时空标记的大数据以及基于这种大数据的方法和应用，着重于对社会经济领域，在自然灾害的监测与预警中同样发挥着重要的作用，是一种快速、低成本、高效的监测方式。社会感知数据相较于传感网络有着数据海量，覆盖面积广，观测密度高且可以详细的记录人类行为模式的优点。然而社会感知方法获取的海量的观测数据有着严重的噪声问题。社会感知的数据来源多样，结构不一致且有效信息密度很低。以社交媒体数据为代表的社会感知数据，大量有价值的信息淹没在大量无关消息之中。对自然灾害的描述信息，多以文本描述形式存在。这些不规则的信息无疑为科学研究带来了一定的影响，凭借人工进行筛选费时费力且不具有时效性。

所以我们的系统利用网络爬虫等技术实现针对微博、抖音等渠道的大城市暴雨积水点多源数据实时获取，使用预训练的UIE模型对获取的数据进行处理，提取如积水点，时间，积水深度等有效信息，形成北京市等大城市近5年积水点数据集。

## 项目目标和任务

利用网络爬虫等技术实现针对微博、抖音等渠道的大城市暴雨积水点多源数据实时获取，使用预训练的UIE模型对获取的数据进行处理，提取如积水点、时间和积水深度等有效信息，形成北京市等大城市近5年积水点数据集。

主要有以下内容：

1. 完成从抖音、微博等渠道实现大城市暴雨积水点多源数据获取；

2. 完成对多源数据获取系统实时部署；

3. 完成对获取数据的去重、辩伪、时域对齐等数据清洗；

4. 完成对数据清洗系统实时部署；

5. 形成大城市历史积水点数据集；

6. 涉及的大城市数量不低于10个；

7. 历史积水点时间序列数据集不低于5年；

8. 实现业务运行日志、监控与报警。

# 项目建设内容

大城市暴雨积水点多源数据实时采集系统采用C/S架构，分为平台层，服务层和客户端。其中平台层包括服务器、网络等基础设施，还有数据库等中间件；服务层则负责实现API接口和日志服务等基本功能模块；客户端则通过使用PyQt5和ECharts可视化显示。

客户端主要包括三大模块：显示模块、属性设置模块和功能模块。

显示模块：通过饼状图、柱状图和地图可视化显示区域，将显示通过爬虫收集整理得到的每个城市经5年的历史积水点数据集。

属性设置模块：共有城市选取模块，采集渠道设置模块和采集属性设置模块。城市选取模块中可以选取超大型城市和大型城市，并且可以通过配置文件增加或删除城市。采集渠道设置模块中可以选取搜狐、微博、腾讯和抖音等渠道。采集属性设置模块中，共有定时采集模块，实时数据采集模块和历史数据采集模块。

功能模块：功能模块中，共有采集模块，模型处理模块和历史数据模块。采集到的数据将会显示在系统的相应区域。点击开始处理，即可将采集到的数据上传至服务端，在服务端将调用所训练的信息处理模型进行关键字段的提取，如日期，城市，地点等信息。并将提取后的信息保存至数据库，在保存过程中将会对数据进行去重、清洗等操作。提取后的信息将传回客户端，客户端将在系统的右下角区域显示处理后的结果，点击保存即可更新显示模块，并将信息保存在本地。

服务端使用FastDeploy AI推理部署工具进行快速化部署UIE模型，并使用Docker容器快速部署环境。使用Flask提供API接口，使用FastDeploy进行模型推理，并使用MySQL进行数据持久化。

# 项目过程管理

## 进度管理

1. 2023年10月，签订合同。
2. 2023年11月，完成项目实施方案的编制工作。
3. 2023年12月，完成需求调研与需求确认的工作。
4. 2023年12月，完成系统详细设计工作。
5. 2024年1月，完成大城市暴雨积水点多源数据实时采集处理系统的系统功能研发。
6. 2024年2月，完成UIE模型训练。
7. 2024年3月，完成UIE模型部署服务。
8. 2024年3月，完成大城市暴雨积水点多源数据实时采集处理系统的测试和系统集成联调测试工作。
9. 2024年4月，系统进入试运行阶段。
10. 2024年5月，完成项目验收。

## 人员组织管理

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 序号 | 姓名 | 职务/职称 | 学历 | 专业资格 | 在本项目中担任的工作 | 工作经历 | 人员分工 |
| 1 | 曹贵猛 |  | 硕士 |  | 爬虫功能的实现，对数据进行预处理，添加相应预警功能 |  |  |
| 2 | 邵天诚 |  | 硕士 |  | 模型部署、文档编写、系统功能实现和完善、数据持久化 |  |  |
| 3 | 桂晶晶 |  | 硕士 |  | 对天津、济南、合肥三个城市的天气数据的数据预处理和数据标注工作。 |  |  |
| 4 | 张雨晴 |  | 硕士 |  | 对北京、南京、广州三个城市的天气数据的数据预处理和数据标注工作。 |  |  |
| 5 | 陈创慧 |  | 硕士 |  | 对武汉、深圳、杭州三个城市的天气数据的数据预处理和数据标注工作。 |  |  |
| 6 | 朱舒婷 |  | 硕士 |  | 对成都、东莞、西安、重庆四个城市的天气数据的数据预处理和数据标注工作。 |  |  |
|  |  |  |  |  |  |  |  |

## 项目质量管理

**问题监控和跟踪**

实施过程前期产生的需求规格说明书、系统设计说明书、测试用例等文档是后期编码、测试的主要依据和输入，这些文档的质量直接决定了软件系统的好坏、系统返工的多寡以及客户满意度。因而对这些文档的评审尤为重要，评审的目的在于在交付给下游开发或测试时及早发现问题，修正错误，以免问题和错误在系统中的蔓延。

文档评审采用同行评审会议的方式进行。评审过程中，主要从以下几方面考察文档的质量：

可读性。主要从文档是否符合模板规范、逻辑结构层次是否清晰明确、文字表达是否无歧义等方面判断；

完整性。主要从文档是否完全满足要求，是否已覆盖所有的功能点等方面判断；

一致性。主要判断文档表述是否前后不一、是否有矛盾等；

技术可行性。主要判断目前的技术框架是否支持，是否有类似的经验，是否有技术风险等。

**技术评审**

技术评审包括项目技术框架的选型评审、具体某个模块的技术实现方式评审等。技术框架的评审目的是为了在进入大规模编码开发前确认选择何种技术框架、判断现有的技术框架是否满足项目功能和性能需求、框架是否足够稳定以及可能存在的风险等，具体某个模块的技术实现方式评审目的是为了保证选择的实现方式目前来说是最优的、可以推广到其他模块使用的。技术评审通过评审会议的方式进行。

# 项目建设成果

## 项目完成情况

截止目前已完成了项目合同规定的全部建设内容，具体提交的成果如下：

完成大城市暴雨积水点多源数据实时采集系统1套

完成大城市历史积水点数据库建设

完成UIE模型推理部署服务系统1套

## 项目试运行应用情况

试运行期间，通过该系统，对抖音、微博、搜狐和腾讯网等渠道的大城市暴雨积水点多源数据实时获取，使用预训练的UIE模型对获取的数据进行处理，提取如积水点、时间和积水深度等有效信息，形成北京市等大城市近5年积水点数据集，试运行系统运行良好，符合项目验收标准。

## 项目提交文件

已按照要求提交相关验收文档

# 尚待改进的工作及建议

无。