

COVID-19公共突发事件数据分析及预测系统

线雨琪 严超群 师闯闯 付江龙

河北建筑工程学院

[摘要]一场突如其来的公共突发事件打乱了人们的正常生活。事件反反复复，形势依然不容乐观。利用人工智能和大数据技术最大程度减少经济代价并实现精准防疫，对未来的常态化突发事件工作具有重要意义。

[关键词] Python; echarts; 公共突发事件; 预测

[DOI] 10.12252/j.issn.2096-6261.2021.09.1928

本文将对此系统利用flask、echarts搭建公共突发事件数据分析及预测系统，实现全球地图数据空间呈现化；利用fbprophet、echarts感知和预测公共突发事件发展趋势与关键节点，并将其进行可视化体现；利用python爬取权威信息并实时更新进行分析，感知突发事件发展趋势，通过分析后得出详细的功能介绍。用户直接可以看到实时数据并且可以获取利用各国家各地区的公共突发事件人数时间序列，预测的未来几天的患病人数的信息，一站式地理空间数据和分析系统，以支持前沿研究和及时决策。

一、引言

(一) 研究背景

在我们对抗公共突发事件时，产生了空前未有的庞大数据量的突发事件数据，而这些数据繁多，运用大数据分析技术和相关方法能够帮助政府有关部门监控事件发展、分发应急救援物资、追踪疾病传染源，使政府的有关部门可以更好地开展公共突发事件防控。然而，与公共突发事件数据有关的权威数据来源非常分散，缺乏统一的整合平台。其次，现有平台大多只展示数据，数据种类非常单一，而且缺乏对公共突发事件数据的分析，这就导致用户无法直观、全面的了解。

(二) 研究概述

本文将设计一个COVID-19公共突发事件数据分析及预测系统，主要工作就是利用flask、echarts搭建公共突发事件数据分析及预测系统实现全球地图数据空间呈现化；利用fbprophet、echarts感知和预测公共突发事件发展趋势与关键节点，并将其进行可视化体现；利用python爬取权威信息并实时更新进行分析，感知事件的发展趋势。

二、功能原理

(一) echarts图表库

对于数据的交互性、在终端上的性能、和对大量数据的处理也随着人们的需求越来越高，开发人员逐渐重视起来。开发人员使用echarts可以更轻松、更灵活的对数据进行可视化展示。

(二) Vscode

Visual Studio Code是一个来自Microsoft的轻量级代码编辑器。它有多个特点，例如：开源；支持多语言开发；轻量级、插件丰富；可扩展性强；人性化，宇宙级。它具有一个丰富的插件生态系统来支持其他语言。

(三) python

1. requests工具。Python自带的urllib2模块具有我们需要的功能但是它使用起来不是很方便。Requests自称为“人类的HTTP”，这使得它使用起来更简单。requests可以实现urllib2能实现的所有功能。Requests工具在COVID-19公共突发事件数据分析及预测系统中，主要用于请求url以爬取事件数据，调用API获取事件数据。

2. NumPy。NumPy是两个英文单词Numerical和Python的缩写。它是Python语言的一个非常强大的第三方类库，主要功能为计算多维数组，其次，也可以用于矩阵的计算。具有很多数学相关计算库以支持专门数组相关操作的计算。它的开放源代码是由许多合作者开发和维护的，计算速度非常快，非常实用。

3. Pandas。大量的数据，对于人来说是很难从中提取出有用数据的。它可以使人们方便的去分析大量的数据。基于统计学的专业知识，它包含了很多数据处理的方法和工具，以及大量的数据分析方法。例如金融、数理研究领域。它实现了复杂的数据分析逻辑，Excel等工具完全比不上它，不需要我们做

重复性工作来处理大量数据。它对接了多种数据可视化库来进行数据的可视化展示。

4. prophet。Prophet是Facebook基于pystan开源的一款时间序列预测库。它具有以下优点：（1）准确而快速。（2）全自动。对混乱的数据进行合理的预测不需要人工。（3）可调整的预测。可以通过调参来调整预测值。（4）在R或Python中可用。

三、功能实现

(一) 系统功能介绍

通过这个系统，用户直接可以看到实时数据并且可以获取利用各国家各地区的公共突发事件人数时间序列，预测的未来几天的患病人数的信息，一站式地理空间数据和分析系统，以支持前沿研究和及时决策。1. 数据处理模块。疫情开始至今产生了大量数据，本系统使用了霍普金斯大学公共突发事件数据仓库，使用git工具与本地同步。2. 数据预测模块。选用了fbprophet库进行预测。为了确保预测准确性，开发过程中花费了大量时间进行调参。3. 绘图模块。该模块使用了echarts图表库，实现公共突发事件全球地图数据空间呈现化，将公共突发事件历史人数信息在三维地球仪上进行柱状图标注。

(二) 系统工作流程图

系统从丁香园、霍普金斯大学数据仓库开始，先进行数据清洗，将数据归一化处理，然后将公共突发事件的实时信息在前端进行可视化展示。根据时间序列信息预测未来5天事件发展，并进行可视化展示。

(三) 环境搭建

1. Vscode环境搭建。2. python3环境安装配置。

(四) 编码实现系统呈现页面

平台的主体界面是一个可以交互的世界地图，平台主要包括四个功能：1. 公共突发事件数据展示：我们使用语义相似度+人工匹配的方法，对爬取的多个数据源的数据进行了整合，并在地图上加以显示。2. 公共突发事件展示：我们收集相关的政府政策、新闻报道、学术研究，开发匹配算法将事件与地区坐标匹配，最终以气泡的形式显示在地图上。3. 公共突发事件数据预测：我们根据国家/地区历史公共突发事件数据，在SEIR模型内的基础上引入大众防控指数，并结合机器学习方法预测未来公共突发事件趋势。4. 风险指数评估：地区的风险指数反映了该地区内公共突发事件的严重程度，我们利用机器学习方法，综合考虑多维度数据评估地区风险指数。此外，平台会实时更新所有数据，并且支持按照时间轴进行历史数据的查看，并可以按照时间自动播放

结论：

通过对实时新闻资料、公共突发事件数据的爬取以及对数据的清洗、处理及展示，根据已有数据进行分析、预测分析未来趋势、搭建后端并且实现提前分析所有涉及的数据，公共突发事件海量信息的获取和确定不同类型的图表的使用场景的水平能力得以提升。

参考文献：

[1] 付江浩. 医疗保险时间序列数据统计分析[D]. 东北师范大学, 2019.

基金项目：项目名称：COVID-19疫情数据分析及预测系统
项目成员：线雨琪、严超群、师闯闯
指导教师：付江龙