2019-2020学年春季学期

《人工智能进展》(0869EY01)

课程报告

成绩

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 学 号 | | 18120255 | 学 院 | 计算机工程与科学学院 | |
| 姓 名 | | 姚施越 | 手工签名 |  | |
| 报告题目 | | 针对新冠肺炎的大数据分析 | | | |
| 课程报告成绩50% | 选题：有前沿性和应用前景。（20%） | | | |  |
| 报告主要部分：1、对所研讨领域的相关研究工作进行较全面的综述，相关研究要做到查新、查全，有论述有评价；2、报告主要内容叙述条理清楚、内容详细；3、有相关技术的案例分析，鼓励对研讨内容的动手实践； 4、案例分析不直接粘贴所有代码；5、代码有注释或说明；6、鼓励有原创性工作（技术或算法）。（60%） | | | |  |
| 书写格式：书写规范、用词正确、无明显的错别字，图表、代码清晰规范，格式协调、效果好，参考文献书写（不少于10篇参考文献，英文文献不少于2篇）、引用规范合理。（20%） | | | |  |
| **报告评语：**  教师签名：  日期： 年 月 日 | | | | | |

针对新冠肺炎的大数据分析

姚施越

(18120255，计算机工程与科学学院)

**摘要：**2019年末，新型冠状病毒肺炎疫情突然来袭，给国家治理体系和治理能力带来了严峻考验。此次疫情中，大数据为打赢疫情防控阻击战提供了有力支撑。本文将讨论大数据在疫情防控中的应用，首先介绍其国内外研究现状；其次，针对疫情态势预测以及疫情地图可视化，阐述本人所做的工作；最后，在文末指出目前大数据在疫情防控应用中的问题并给出解决方案。

**关键词：**大数据；模型预测；数据可视化；疫情防控；复工复产

# 引言

2019年末在武汉爆发的新型冠状病毒肺炎传播迅猛，已被世卫组织定为“国际关注的突发公共卫生事件”，各省市也陆续通过启动重大突发公共卫生事件一级响应来控制人口流动。习近平总书记在中央全面深化改革委员会时也强调了大数据在疫情防控中的作用[1]。面对此次突如其来的疫情，我们也能看到，大数据凭借其成熟的技术和产业、强大的数据采集和分析能力，结合互联网、云计算、区块链、物联网等新兴技术，为疫情追踪、区域防控、数据发布、模型预测等提供了技术支撑，在疫情防控和复工复产中发挥了巨大的作用[2]。

本文将以大数据在疫情防控中的应用为主体展开讨论，首先简要介绍该领域国内外研究现状；其次，对于疫情态势预测，使用爬取的疫情数据，实现简单的SIR传染病模型，并以图例直观展示；接下来，阐述疫情地图的具体实现，利用Echarts.js库进行开发，针对疫情发展的态势，以三维柱状图、三维色块图、二维条形图等图表形式，从全球、国内、省份三个维度对确诊病例的数量及变化趋势进行可视化。最后，指出目前大数据与疫情结合的问题，同时给出应对措施。

# 国内外研究现状

## **2.1 相关研究综述**

考虑到学术期刊是各类科研项目成果发表的主要形式，本文以中国知网数据库收录的文献为依据，对大数据在疫情防控中的应用的研究现状进行梳理[3]。截至2020年6月21日，在中国知网以“疫情”和“大数据”作为关键词以篇名进行跨库精确检索，共得到研究文献612篇（文献最早出现于2013年）。相关研究内容主要集中于以下3个方面。

### 2.1.1 回溯及预测疫情扩散趋势

利用来自腾讯位置、百度迁移、各省统计局、卫健委的大数据，结合对数线性回归模型，估算疫情爆发初期，湖北省人口流动的情况[4][5]。同时，对迁出人群作分类分析，并利用Pearson相关系数等对防疫措施作评价。使用SIR、SEIR模型[6]与确诊人数作拟合，观察传染系数R0在武汉实施交通管制前后的变化，分析政府举措的有效性[7]。此外，也有以国内2019航空出境信息，将城市之间流动性作为主要考虑因素，做流行病预测模型相关方面的研究[8]。

### 2.1.2 疫情防控中的网络舆情

选取2020年1月23日舆情大规模爆发至3月20日舆情处于消退期官方媒体、微博、微信公众号三大信息平台发布的共计20 余万字的相关网络文本，分析网络舆情形成过程及内在机理。在此基础上对当下各方应对网络舆情工作存在的不足之处进行分析[9]。

### 2.1.3 助力社会治理和政府决策

针对疫情防控过程中，对大数据助力复工复产[10]以及公共危机管理[11]的显著作用作综述性研究，也指出大数据在应用和治理体系方面的问题。

## **2.2 对相关研究的评价**

总体来看，学界对大数据在疫情中的应用开展了丰富深刻的研究，为本研究提供了充足的借鉴。但同时也存在一些不足，例如针对疫情趋势预测，缺少具体的模型演示；对于疫情可视化相关的研究较少。本文也将主要针对这两个部分进行研究，对个人的研究成果进行阐释。

# 基于SIR模型的疫情态势预测

## **3.1 概述**

基于疫情高危人群相关数据，结合疫情新增确诊、疑似、死亡、治愈病例数，借助传播动力学模型、动态感染模型、回归模型等大数据分析模型和实践技术，不仅可以分析展示发病热力分布和密切接触者的风险热力分布，还可以进行疫情峰值拐点等大态势预判。

利用深度学习等新兴人工智能技术，联合出行轨迹流动信息、社交信息、消费数据、暴露接触史等大量数据进行科学建模，可以根据病患确诊顺序和密切接触人员等信息定位时空碰撞点，进而推算出疾病传播路径，为传染病溯源分析提供理论依据。

本部分将实现简单的SIR模型用于做预测，同时使用爬取自丁香医生等平台的真实数据，以图表形式直观展示疫情中各部分人群的变化情况，包括真实值与预测值的对比图。

## **3.2 SIR模型的介绍**

### 3.2.1 模型流程图



图1 SIR模型流程图

### 3.2.2 具体含义

假设一个区域内的总人数为；

易感者（Susceptible），用表示其人数，易感者按照概率转化为感染者；

感染者（Infectious）人数用来表示，感染者按照概率转化为痊愈者；

痊愈（Recovered）的人数用来表示[12]。

### 微分方程形式

该模型的具体推演过程如下：

（1）有个感染者每天会接触个人，这些易感者有的概率会变为感染者，易感者占总人数的比例为，将以上所有量乘在一起再加上负号，便是每天易感者转为感染者的病例数。

（2）感染者每天增加人数由上式求得，同时个感染者有的概率会变为痊愈者，把这一项减去。

（3）痊愈的增加人数已由上式求得。

我们看下其微分方程形式

### 迭代形式

## **3.3 SIR模型的实现**

### 3.3.1 问题描述-- 如何利用python实现SIR模型

引入numpy和scipy库用于建立微分方程和数学计算，pandas库用于读入包含数据的csv文件，matplotlib库用于绘制图例。

在确定了库之后，如何实现？

### 3.3.2 问题的分析与算法设计

求解分析：使用此前3.2.3中微分方程的形式建立微分方程，具体的实现，将在下文详细展开。

模型的参数设定部分代码如下：.

|  |
| --- |
| 1. #模型的参数 2. beta = 8e-6#易感者转化为感染者的概率 3. gamma = 0.04#感染者转化为痊愈者的概率 4. TS = 1.0#变化时间 5. ND = 100.0#结束时间 6. S0 = 39000#初始易感者人数 7. I0 = 41#初始感染者人数 8. INPUT = [S0, I0, 0.0]#规整化数据，用作微分方程的参数 |

模型的微分方程部分代码如下：.

|  |
| --- |
| 1. # 模型的微分方程 2. **def** diff\_eqs(INP, t): 3. Y = np.zeros((3))#建一个长度为3的数组用于存三步微分方程 4. V = INP 5. **print**(V) 6. #利用SIR模型建立微分方程 7. Y[0] = -beta \* V[0] \* V[1] 8. Y[1] = beta \* V[0] \* V[1] - gamma \* V[1] 9. Y[2] = gamma \* V[1] 10. **return** Y |

主函数中利用SIR模型预测数据绘图部分代码如下：.

|  |
| --- |
| 1. #SIR模型预测数据绘图部分 2. t\_start = 0.0#起始时间 3. t\_end = ND#结束时间 4. t\_inc = TS#变化时间 5. t\_range = np.arange(t\_start, t\_end+t\_inc, t\_inc)#用于创建等差数组，即天数变化 6. RES = spi.odeint(diff\_eqs, INPUT, t\_range)#scipy库中一个数值求解微分方程的函数 7. #利用matplotlib库绘制图例 8. fig = pl.figure() 9. pl.subplot(111) 10. pl.plot(RES[:, 1], "-r", label = "Infectious") 11. pl.plot(RES[:, 0], "-g", label = "Susceptible") 12. pl.plot(RES[:, 2], "-k", label = "Recovered") 13. pl.legend(loc = 0) 14. pl.title("SIR model") 15. pl.xlabel("Time") 16. pl.ylabel("All cases") 17. pl.savefig("预测值.png") 18. **print**("预测最大感染人数:%d 位置:%d" % (RES[:,1].max(), np.argmax(RES[:, 1])))#返回一个numpy数组中最大值的索引值 |

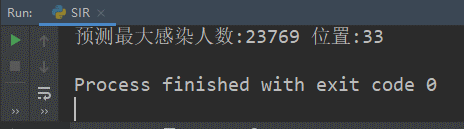


图2 pycharm控制台中输出预测最大感染人数

这表明，感染者的人数在第33天，即在2月14日达到峰值，这与真实数据相契合。

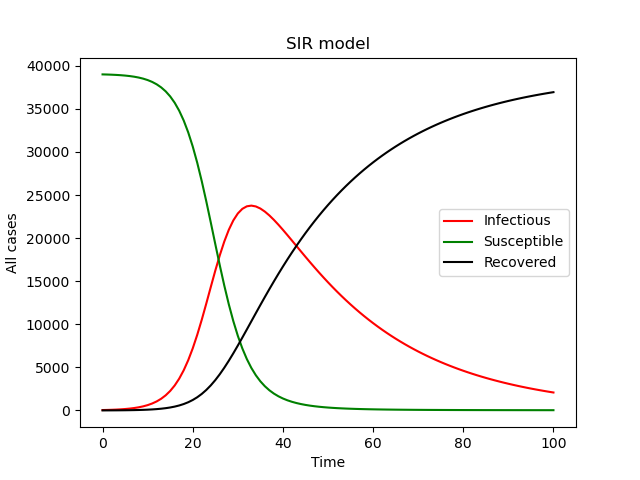


图3 利用SIR模型预测数据绘图

可以看出，感染者的人数是先增加后减少的。

表1 使用到的真实数据的csv文件的格式展示（数据来源：爬取自丁香医生）

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| date | 感染者 | 疑似者 | 死亡 | 治愈 | 累积密切接触 |
| 2020/1/10 | 41 | 0 | 1 | 2 | 739 |
| 2020/1/11 | 41 | 0 | 1 | 6 | 763 |
| 2020/1/12 | 41 | 0 | 1 | 7 | 763 |
| … | … | … | … | … | … |
| 2020/6/18 | 83325 | 7 | 4634 | 78398 | 755832 |
| 2020/6/19 | 83352 | 11 | 4634 | 78410 | 756223 |

主函数中利用读取到的真实数据绘图部分代码如下：.

|  |
| --- |
| 1. # 读取数据 2. data = pd.read\_csv("data.csv", index\_col = ["date"]) 3. data["现有感染者"] = data["感染者"] - data["死亡"] - data["治愈"] 4. **print**(data) 6. # 数据作图，用到了matplotlib库 7. fig = pl.figure() 8. pl.subplot(111) 9. pl.plot(data["现有感染者"], "-r", label = "Infectious") 10. pl.plot(data["疑似者"], "-g", label = "Susceptible") 11. pl.plot(data["死亡"], "-b", label = "Dead") 12. pl.plot(data["治愈"], "-k", label = "Recovered") 13. pl.plot(data["现有感染者"]-data["现有感染者"].shift(1), "-y", label = "Increase")#每日新增 14. pl.legend(loc = 0) 15. pl.title("Real data") 16. pl.xlabel("Time") 17. pl.ylabel("All cases") 18. pl.savefig("真实值.png") |

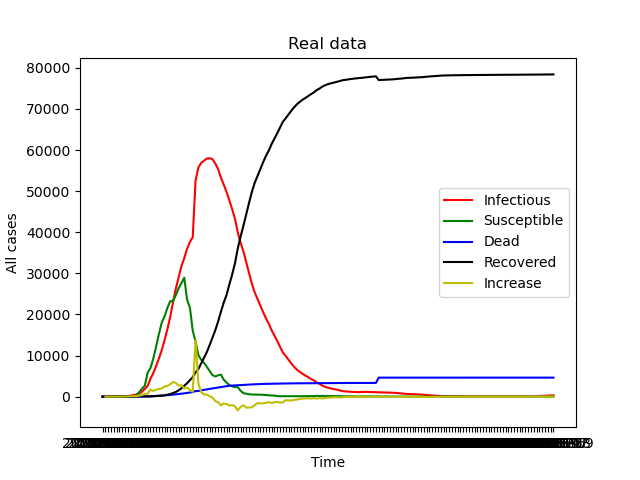


图4 利用读取到的真实数据绘图

可以看出，感染者的变化趋势同样先增加后减少，与SIR模型结果一致。

主函数中将预测数据和真实数据一同绘图部分代码如下：.

|  |
| --- |
| 1. # 将预测值与真实值画到一起 2. fig = pl.figure() 3. pl.subplot(111) 4. pl.plot(RES[:, 1], "-r", label = "Infectious") 5. pl.plot(data["现有感染者"], "o", label = "Real") 6. pl.plot(data["现有感染者"]-data["现有感染者"].shift(1), "-y", label = "increase") 7. pl.legend(loc = 0) 8. pl.title("SIR model") 9. pl.xlabel("Time") 10. pl.ylabel("All cases") 11. pl.savefig("预测值与真实值.png") |

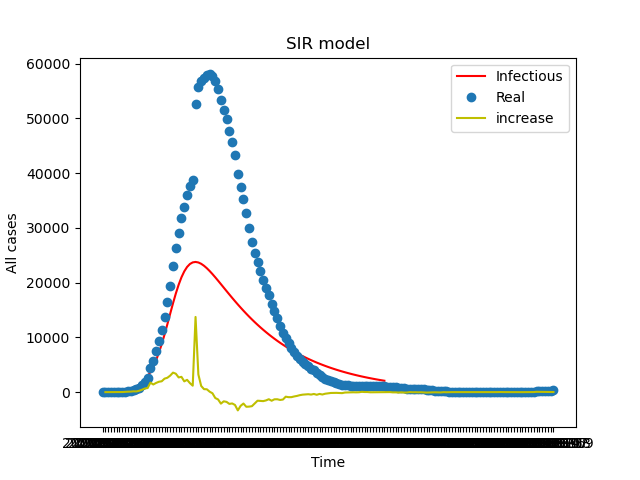


图5 利用预测数据和真实数据一同绘图

可以看出，感染者的变化趋势基本是一致的，真实数据的峰值更高。

## **3.4 问题与展望**

本部分使用到的是SIR模型，是因为获取到的真实数据包括确诊、易感、死亡、治愈人数，符合SIR模型。实际情况更加复杂，易感染人群在一开始会经历潜伏期，一段时间之后才出现症状，因此我们可以考虑在SIR模型的基础上引入潜伏者，潜伏者有一定概率转化为感染者。此外，本模型非常简单，考虑的因素还远远不够，可以引入防控强度、医疗资源、人口流动、复工情况和境外输入等数据。

# 疫情数据可视化

## **4.1 概述**

大数据可视分析是指在大数据自动分析挖掘方法的同时,利用支持信息可视化的用户界面以及支持分析过程的人机交互方式与技术,有效融合计算机的计算能力和人的认知能力,以获得对于大规模复杂数据集的洞察力[13]。

在这次抗击疫情过程中，公众非常关心疫情发展趋势，疫情防控部门希望不断总结经验教训，评估现有措施的有效性。疫情地图便是一个很好的方式。本项目将针对疫情发展的态势，以三维柱状图、三维色块图、二维条形图等图表形式，从全球、国内、省份三个维度对确诊病例的数量及变化趋势进行可视化。

## **4.2 项目介绍**

### 4.2.1 功能流程图

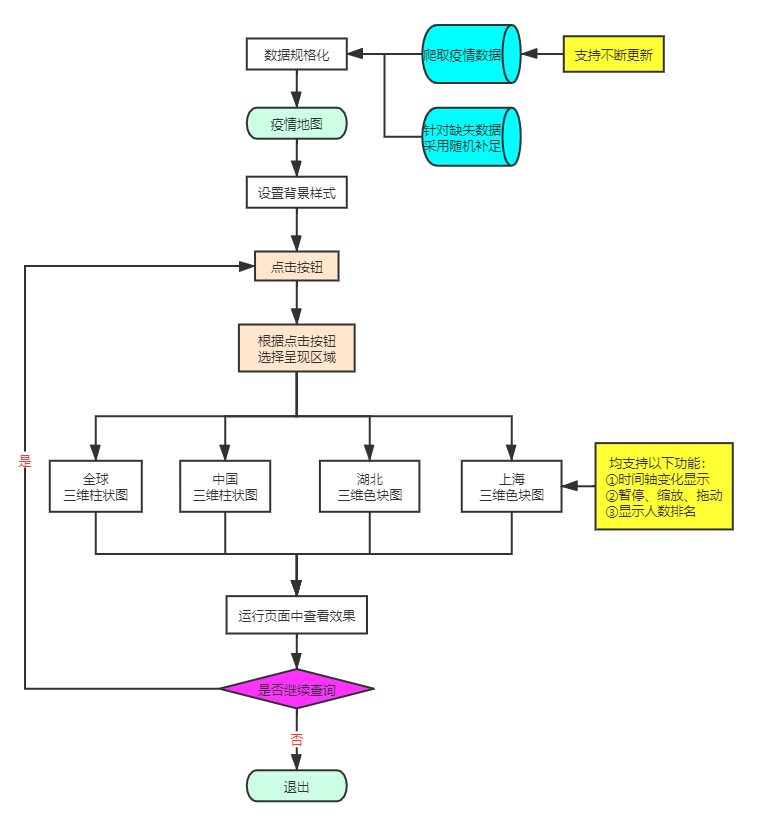


图6 功能流程图

### 4.2.2 开发工具及环境说明

开发语言：Html+JavaScript

开发工具：Visual Studio Code

浏览器：Chrome

插件：Live Server

库：Echarts.js

## **4.3 项目实现**

### 4.3.1 问题描述-- 全国疫情呈现方式及如何实现全国各省份疫情三维柱状图动态效果

考虑全国疫情的呈现方式，用二维/三维，动态/静态，柱状图/折线图？

在确定了呈现方式后，如何实现？

### 4.3.2 问题的分析与算法设计

求解分析：对于疫情地图的呈现方式，原先尝试的是全国范围二维形式以及柱状图、折线图等形式，随后，出于动态显示效果的直观和三维的立体效果，在之前二维的基础之上，对部分代码进行重新实现。至于具体的实现，将在下文详细展开。

世界地图类似，只是将map换成world，下文中不再详述。

### 4.3.3 关键实现技术（或代码）

全国各省份疫情三维柱状图动态效果代码如下：.

|  |
| --- |
| 1. <button onclick="changetoXinzeng()">每日新增</button> 2. <button onclick="changetoZhiyu()">累计治愈</button> 3. <button onclick="changetoXianyou()">现有确诊</button> 4. // <button>标签，后续要用到，用以切换至其他数据来显示 5. // 全国各省份疫情3D柱状图动态效果 6. option = { 7. baseOption: { 8. timeline: { 9. // 利用timeline组件实现动态效果 10. axisType: 'category',// 类目轴，适用于离散的类目数据。 11. autoPlay: true,// 设置自动播放 12. playInterval: 500,// 切换间隔设定为500毫秒 13. symbolSize: 12,// timeline标记的大小 14. left: '5%', 15. // timeline组件离容器左侧的距离，'5%' 是相对于容器高宽的百分比 16. right: '5%', 17. // timeline组件离容器右侧的距离，'5%' 是相对于容器高宽的百分比 18. bottom: '0%', 19. // timeline组件离容器下侧的距离，'0%' 是相对于容器高宽的百分比 20. width: '90%',// 文字块的宽度，表示的是所在文本块的 contentWidth（即不包含文本块的 padding）的百分之多少 21. data: days,// 数据为给定的天数数组 22. tooltip: { 23. formatter: days// 时间轴的提示框组件 24. }, 25. }, 26. visualMap: [{ 27. // 视觉映射组件 28. type: 'continuous',// 连续型 29. seriesIndex: 0, 30. text: ['确诊人数'],// 标题文本 31. calculable: true, 32. // 最大值 33. max:100000,// 最大值 34. inRange: { 35. // 三种颜色按次序呈现为条状 36. color: ['#87aa66', '#eba438', '#d94d4c'] 37. } 38. } 39. ], 40. xAxis: { 41. // 直角坐标系 grid 中的 x 轴 42. min: 0, 43. max: 4000, 44. show: false// 不显示 x 轴 45. }, 46. yAxis: [{ 47. // 直角坐标系 grid 中的 y 轴 48. inverse: true,// 是反向坐标轴，即横轴为y轴 49. offset: '2',// 配置 offset 属性防止同个位置多个 Y 轴的重叠 50. 'type': 'category',// 类目轴 51. data: '', 52. nameTextStyle: { 53. color: '#fff' 54. }, 55. axisLabel: { 56. // 坐标轴刻度标签 57. textStyle: { 58. // 字体样式 59. fontSize: 14, 60. color: '#000000', 61. }, 62. interval: 0 63. }, 64. }], 65. }, 66. animationDurationUpdate: 1000,// 图例翻页时的动画时长1000毫秒 67. animationEasingUpdate: 'quinticInOut',// 数据更新动画的缓动效果 68. options: [] 69. }; 70. // 柱状图数据设定（data、days、province分别为给定的矩阵形式的确诊人数数据、日期和省份名称） 71. **for** (**var** n = 0; n < days.length; n++) { 72. **var** res = [];//初始化最后用于呈现的数据数组 73. **for** (**var** j = 0; j < data[n].length; j++) { 74. **var** geoCoord = geoCoordMap[province[j]]; 75. **if** (geoCoord) { 76. // 由于湖北省的疫情较为严重，确诊人数较多，使用柱状图呈现时会出现过高的情况，其他省份则会几乎没有，故对其特殊处理，使得其高度下降，呈现的效果更好。 77. **if** (j == 0) { 78. res.push({ 79. // 名称为省份名称 80. name: province[j], 81. // concat() 是js中用于连接两个或多个数组的方法，这里用作省份经纬度和确诊人数的连接 82. value: geoCoord.concat(data1[n][j]) 83. }); 84. } 85. **else** 86. { 87. res.push({ 88. name: province[j], 89. value: geoCoord.concat(data[n][j]) 90. }); 91. } 92. } 93. } 94. // 对数据进行排序，以便后续的省份人数排名呈现 95. res.sort(function(a, b) { 96. **return** b.value - a.value; 97. }).slice(0, 6); 98. res.sort(function(a, b) { 99. **return** a.value - b.value; 100. }); 101. option.options.push({ 102. // 页面标题 103. title: [{ 104. text: days[n] + '新型冠状病毒全国感染人数', 105. textStyle: { 106. color: '#2D3E53', 107. fontSize: 28 108. }, 109. left: 20, 110. top: 20, 111. } 112. ], 113. tooltip: { 114. show: true, 115. }, 116. // 使用3D坐标系组件 117. geo3D: { 118. map: 'china',// 使用中国地图 119. roam: true,// 支持缩放 120. }, 121. series: [{ 122. name: '确诊人数', 123. type: "bar3D",// 使用3D柱状图 124. coordinateSystem: 'geo3D',// 使用之前所给的地理坐标系 125. barSize: 1, //柱子粗细 126. shading: 'lambert', 127. opacity: 1, 128. bevelSize:0.3, 129. label: { 130. formatter: '{b}'// 标签的呈现内容 131. }, 132. data: res// 数据采用之前所设置的data数组 133. },] 134. }); 135. } 136. // 使用给定的选项设置图表以呈现 137. **if** (option && typeof option === "object") { 138. myChart.setOption(option, true); 139. } 140. **function** changetoXianyou(){ 141. // 函数：切换至现有确诊 142. // 仅改变了data数组，不在此列出 143. // 使用给定的选项设置图表以呈现 144. // 这样就实现了切换效果 145. **if** (option && typeof option === "object") { 146. myChart.setOption(option, true); 147. } 148. } |

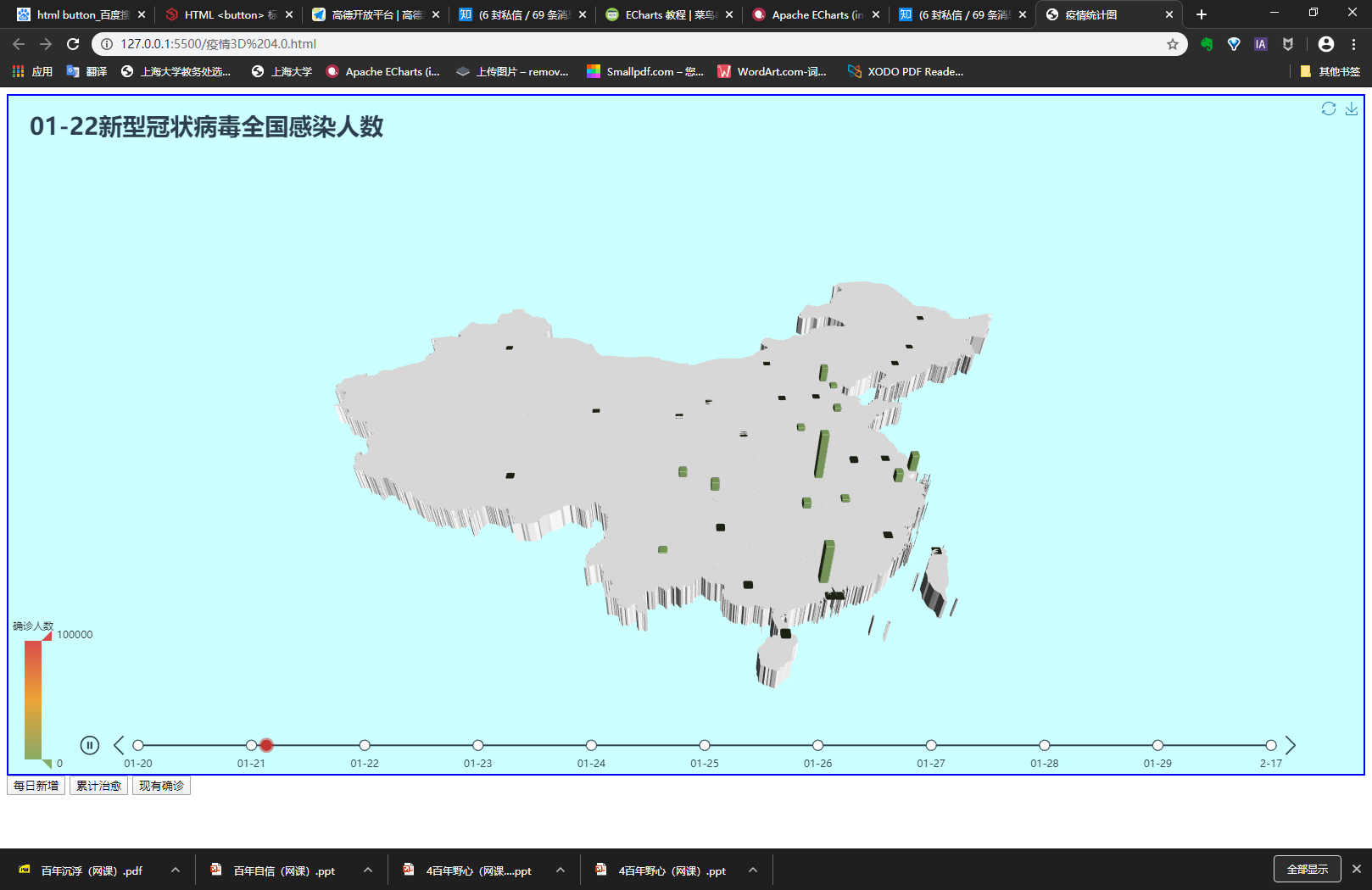


图7 全国各省份疫情三维柱状图动态效果图

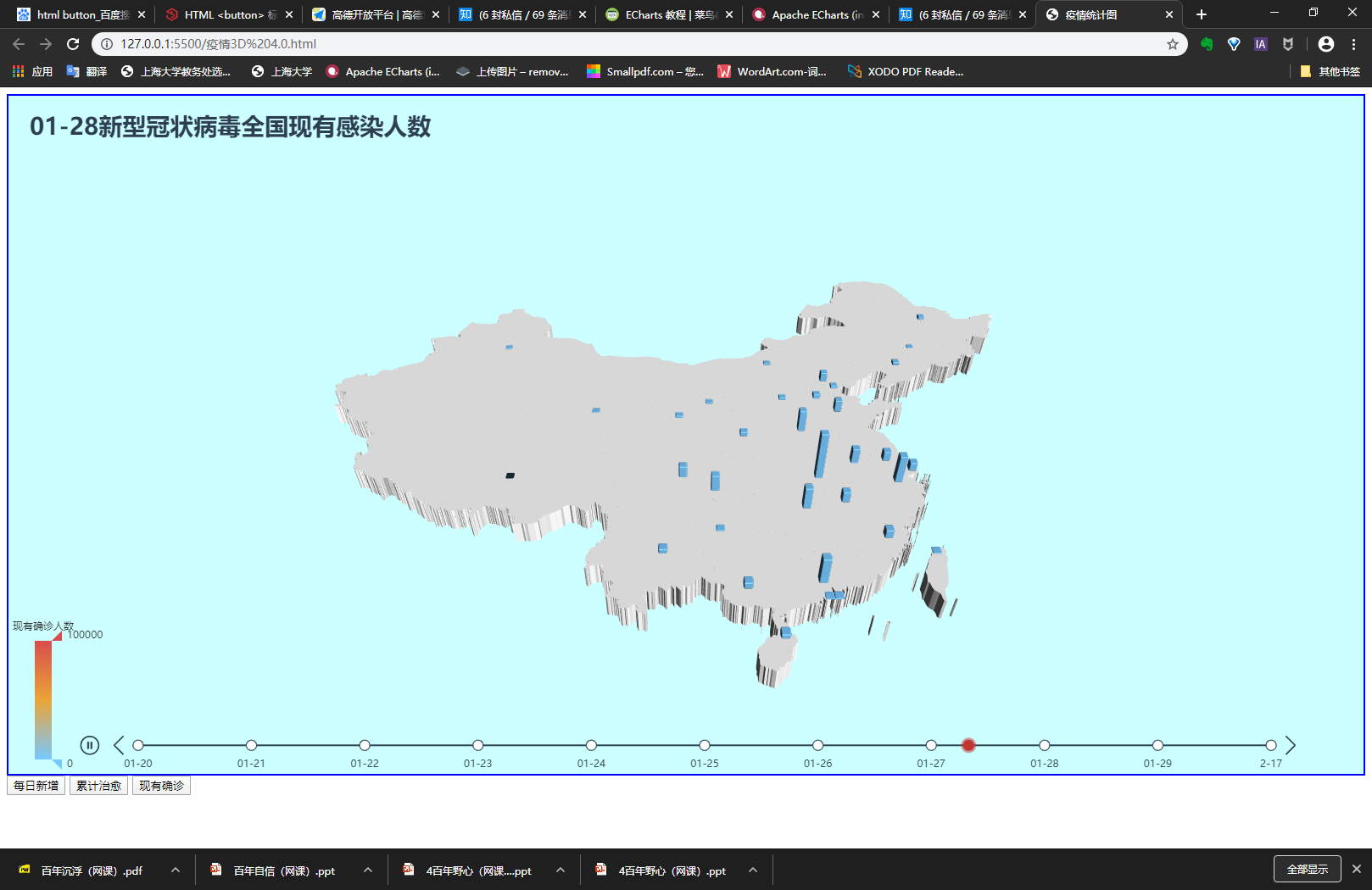


图8 全国各省份疫情三维柱状图动态效果图（现有确诊）

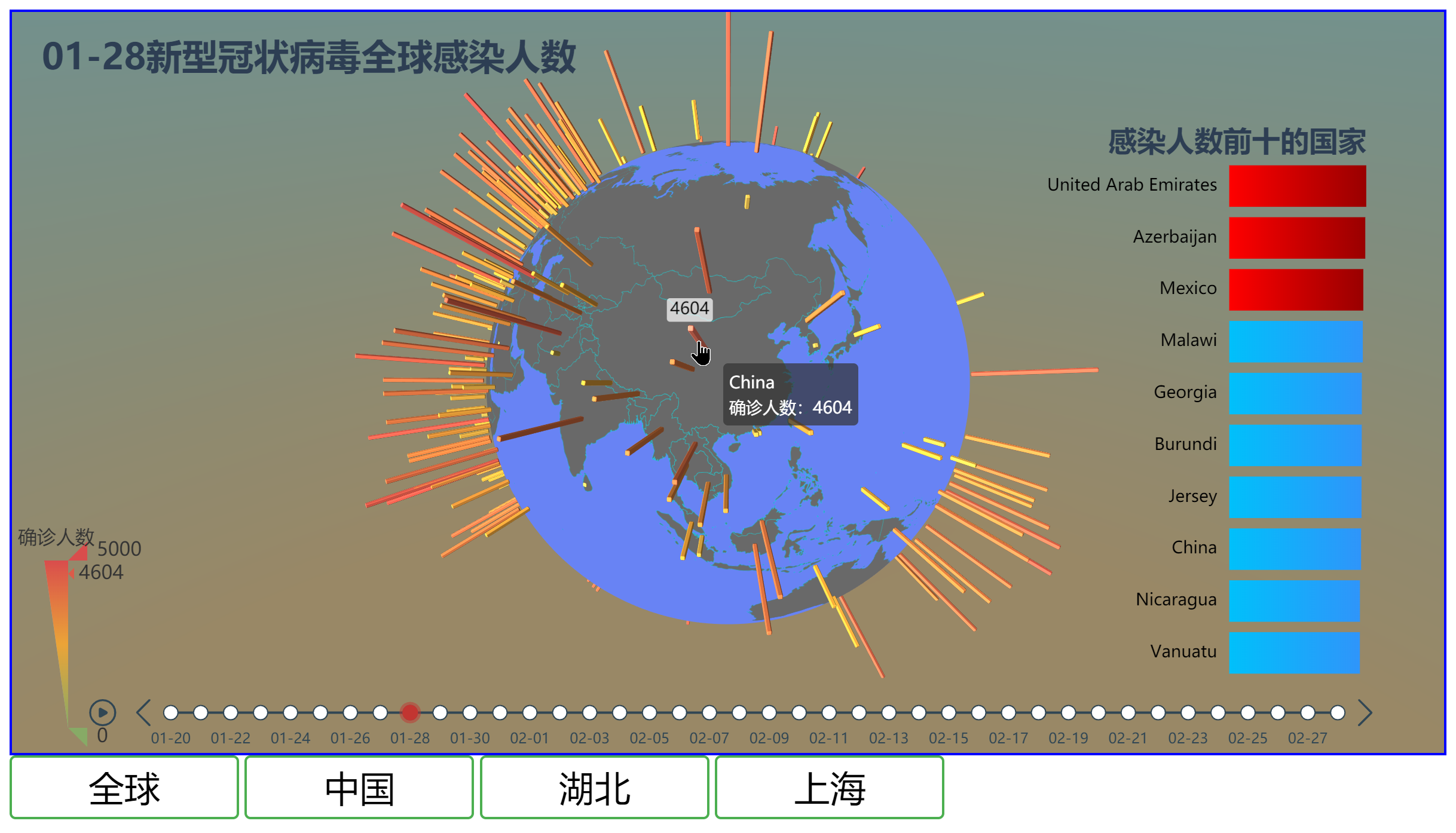


图9 全球疫情三维柱状图动态效果图（数据非真实）

### 4.3.4 问题描述-- 省份疫情呈现方式、如何实现省份疫情效果

考虑上海、湖北疫情的呈现方式，用二维/三维，动态/静态，柱状图/折线图？

在确定了呈现方式后，如何实现？

接下来是，如何实现省份、全国及全球的切换部分？

### 4.3.5 问题的分析与算法设计

对于省份疫情地图的呈现方式，仍沿用全国疫情的呈现，使用三维形式，代码的逻辑主体基本沿用，各区的差异则通过色块呈现，对这一部分进行了改动。

### 4.3.6 关键实现技术（或代码）

点击跳转按钮及显示上海市疫情代码如下：.

|  |
| --- |
| 1. // 利用鼠标点击事件，通过点击的按钮获取其名称进行判断并跳转相应的地图的函数 2. <button class = "button1" onclick="showEarth()">全球</button> 3. <button class = "button1" onclick="showChina()">中国</button> 4. <button class = "button1" onclick="showHubei()">湖北</button> 5. <button class = "button1" onclick="showShanghai()">上海</button> 6. // 显示上海疫情片段代码 7. **function** showShanghai(){ 8. **var** globel = echarts.init(document.getElementById('globel')); 9. // JSON文件(地图数据)路径 10. **var** uploadedDataURL = "shanghai.json";// shanghai.json 11. // 显示加载动画效果,可以在加载数据前手动调用该接口显示加载动画，在数据加载完成后调用 hideLoading 隐藏加载动画。 12. globel.showLoading(); 13. // 引入JSON文件 14. $.getJSON(uploadedDataURL, **function**(geoJson) { 15. // 注册地图名字（shanghai）和数据(geoJson) 16. echarts.registerMap('shanghai', geoJson); 17. // 隐藏动画加载效果。 18. globel.hideLoading(); 19. // 图表配置项后续类似，对这些部分不再展示 20. **if** (option && typeof option === "object") { 21. globel.setOption(option, true); 22. } 23. }); 24. } |

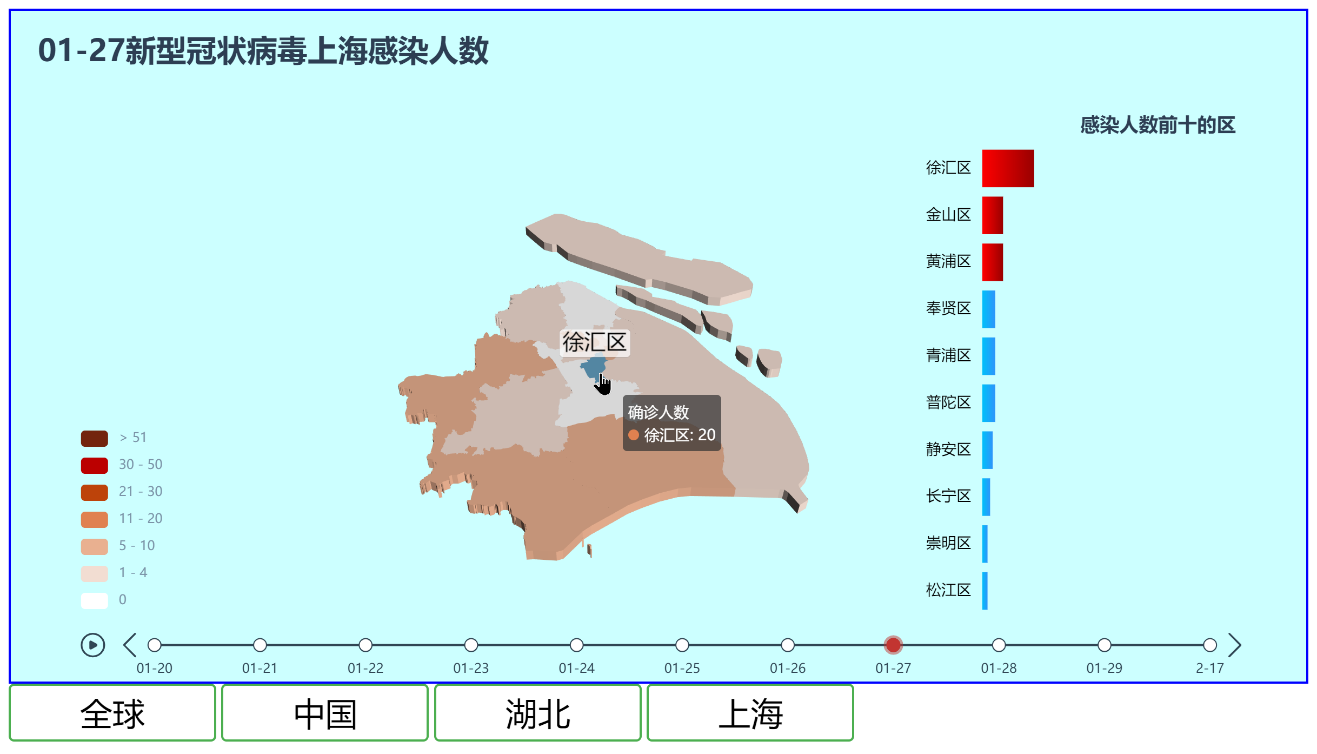


图10 上海疫情三维色块图动态效果图（数据非真实）

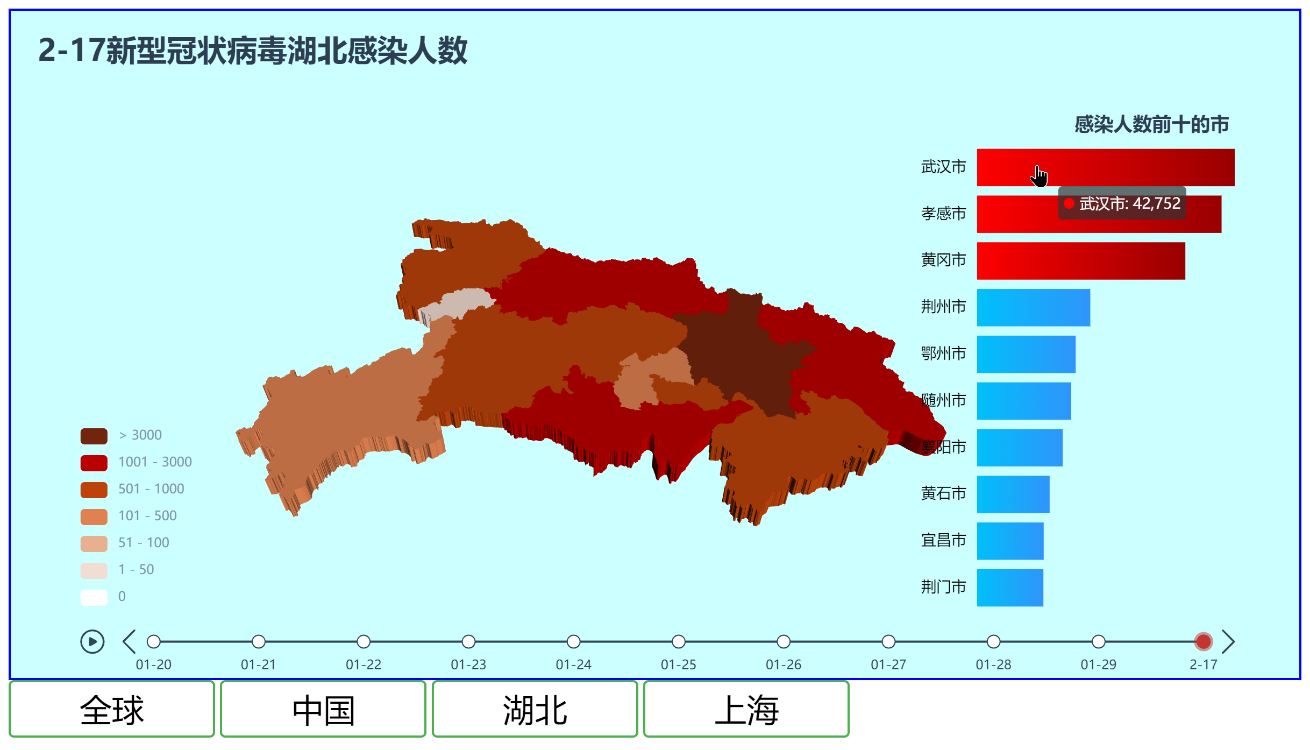


图11 湖北疫情三维色块图动态效果图

## **4.4 问题与展望**

本项目的数据来源部分是丁香医生（实时数据）、各地卫健委（历史数据），利用爬虫手段解析其发布的新闻公告来获得当天和当地各市的疫情数据，后续考虑到工作量较大，因此对于缺少的数据采用随机数进行补足。

未来将考虑实现查看周边病例的涉及地点和活动场所。此外，结合疫情地图，分析政府联防联控的手段的有效程度和媒体对不同地区的病例追踪及疫情解读。

# 总结

当前，大数据在疫情研判、防控部署、人员监测上发挥了重要作用，为产业协同、线上招商、复工复产提供了有力支持。本文主要对疫情态势预测和疫情地图可视化两大领域作了介绍。基于以上讨论，我们可以发现，面对疫情防控和复工复产的严峻考验，我国大数据应用治理体系仍暴露出一些问题。

（一）数据来源有限，采集手段落后

在大数据时代，公众的信息需求发生了变化，公众既希望获知一个地方总体概况的统计数据，也希望看到相关病人个体的数据。而根据目前政府公开的信息和企业案例中提供的数据，绝大部分地区都还难以满足这些要求。一些政府部门和基层组织在采集疫情相关数据时仍然采用手工作坊式的人海战术。在采集时通常采用手工填表、电话问询等方式，这既给基层工作人员增加负担，也无法保证数据的真实性。

（二）数据缺乏治理，数据质量堪忧

从疫情分析的数据来看，目前公开渠道获取的数据不规范，数据口径、数据统计时间等维度不统一，难以统一录入数据库，为分析带来很大困难。数据源混杂重复、数据质量不高的情况十分普遍，数据可信性有待提升。

（三）隐私保护不足，存在合规风险

基于疫情防控这一公共利益的切实需要，可以对特殊人群的特定信息进行披露，但应当遵循合规原则，公开内容应确实必要，符合疫情防控目的，公开手段应合理适当，尽可能将负面影响降到最低。为了做好新型冠状病毒感染肺炎疫情联防联控中的个人信息保护，积极利用包括个人信息在内的大数据支撑联防联控工作，中央网络安全和信息化委员会办公室[14]。

（四）数据流通不足，数据孤岛严重

大数据的核心在于互联互通。只有融会了多源数据，才能发挥大数据的价值。在疫情防控中我们也看到，数据分散割裂、聚而不通、通而不用的问题。政府、企业、机构等不同来源间数据共享程度较低，医疗、交通、水电、通信、互联网等不同行业间数据整合不够，全国性的疫情信息数据库有待建立和完善[15]。

针对以上问题，我给出个人的一些想法如下。

（一）注重积累，充分发挥数据价值

加强数据收集和监测能力，确保基础数据真实准确、更新及时、具有权威性，加强区域和部门间数据整合和共享。同时，统一数据标准和口径非常关键。要明确定义数据的来源、数据的格式和数据管理的要求。此外，还要规范数据的处理过程。在保证数据真实性前提下，充分利用现有大数据平台、大数据建设成果，及时有效的实现数据的汇聚融合。

（二）保障隐私，确保数据合法合规

首先明确收集和使用相关个人信息的执行机构，以及相关工作的流程规范，确保收集的信息仅用于疫情防控和疾病防治。其次，在数据共享的过程中，明确各部门与相关单位的共享范围与使用权限，采取匿名化、去标识化、访问控制等措施对数据共享行为进行约束。另外，是在对社会发布疫情信息时，应以保护当事人个人信息为前提，适度公开相关脱敏信息。最后是强化基层信息搜集工作人员的保密意识，从源头把控，杜绝个人信息泄露。

（三）推动技术，拓宽技术应用场景

一是鼓励医疗机构、医疗科研单位应用人工智能技术，优化算法和算力，助力在线诊断、病毒基因测序和疫苗/药物等研发攻关。二是鼓励云计算、大数据、人工智能等相关企业加强联合攻关，在疫情发现、分析、预警、防治等方面创新应用，提升疫情防控治疗成效。

**致谢** 首先，非常感谢刘老师一学期的辛勤付出，在课程学习中，我的收获良多，对于人工智能的各领域有了更清晰的认识。其次，在准备研讨的过程之中我遇到了不少疑难，也非常感谢组内的其他同学对于我疑惑的耐心细致的解答。最后，在本篇报告中，也难免会有不少纰漏和不严密的地方，敬请刘老师批评指正。

参 考 文 献

[1] 中国信通院. 《疫情防控中的数据与智能应用研究报告（1.0版）》. [EB/OL],http://www.caict.ac.cn/kxyj/qwfb/ztbg/202003/t20200303\_275553.htm 2010,3,3

[2] 王庆德,吴欣哲.大数据助力疫情防控和复工复产——基于新冠肺炎疫情[J].中国物价,2020(05):9-11.

[3] 邢鹏飞,李鑫鑫.重大疫情防控中网络舆情形成机制及引导策略研究——基于新冠肺炎疫情期间网络舆情文本的质性分析[J/OL].情报杂志:1-9[2020-06-22].http://kns.cnki.net/kcms/detail/61.1167.G3.20200422.1511.002.html.

[4] 向云波,王圣云.新冠肺炎疫情扩散与人口流动的空间关系及其对中国城市公共卫生分类治理启示[J/OL].热带地理:1-18[2020-06-22].https://doi.org/10.13284/j.cnki.rddl.003247.

[5] 刘张,千家乐,杜云艳,王楠,易嘉伟,孙晔然,马廷,裴韬,周成虎.基于多源时空大数据的区际迁徙人群多层次空间分布估算模型——以COVID-19疫情期间自武汉迁出人群为例[J].地球信息科学学报,2020,22(02):147-160.

[6] Zoltan Neufeld,Hamid Khataee,Andras Czirok. Targeted adaptive isolation strategy for COVID-19 pandemic[J]. Infectious Disease Modelling,2020,5.

[7] 赵序茅,李欣海,聂常虹.基于大数据回溯新冠肺炎的扩散趋势及中国对疫情的控制研究[J].中国科学院院刊,2020,35(03):248-255.

[8] Joseph T Wu,Kathy Leung,Gabriel M Leung. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study[J]. The Lancet,2020,395(10225).

[9] 邢鹏飞,李鑫鑫.重大疫情防控中网络舆情形成机制及引导策略研究——基于新冠肺炎疫情期间网络舆情文本的质性分析[J/OL].情报杂志:1-9[2020-06-22].http://kns.cnki.net/kcms/detail/61.1167.G3.20200422.1511.002.html.

[10] 王庆德,吴欣哲.大数据助力疫情防控和复工复产——基于新冠肺炎疫情[J].中国物价,2020(05):9-11.

[11] 李传军.运用大数据技术提升公共危机应对能力——以抗击新冠肺炎疫情为例[J].前线,2020(03):21-24.

[12] WIKIPEDIA. Compartmental models in epidemiology. [EB/OL], https://en.wikipedia.org/wiki/Compartmental\_models\_in\_epidemiology#The\_SIR\_model

[13] 任磊,杜一,马帅,张小龙,戴国忠.大数据可视分析综述[J].软件学报,2014,25(09):1909-1936.

[14] 中国信通院. 《疫情防控中的数据与智能应用研究报告（1.0版）》. [EB/OL],http://www.caict.ac.cn/kxyj/qwfb/ztbg/202003/t20200303\_275553.htm 2010,3,3

[15] 王庆德,吴欣哲.大数据助力疫情防控和复工复产——基于新冠肺炎疫情[J].中国物价,2020(05):9-11.