

2020 年第七届中国可视化与可视分析大会

数据可视分析挑战赛

(ChinaVis Data Challenge 2020)

作品说明文档

参赛队名称：**重庆大学-冯丹**

作品名称：新冠病毒疫情潜在影响与次生灾害可视分析系统

新冠作品主题关键词：新冠疫情、多源数据、潜在影响、次生灾害、防范

团队成员：

- 冯丹，重庆大学，1846561704@qq.com，队长
- 月小琪，重庆大学，2550770153@qq.com
- 孙德晟，重庆大学，1621224528@qq.com
- 刘黎，重庆大学，1097670026@qq.com
- 唐豪，重庆大学，1963024305@qq.com
- 胡海波，重庆大学，haibo.hu@cqu.edu.cn，指导老师

团队成员是否与报名表一致（是或否）：是

是否学生队（是或否）：是

使用的分析工具或开发工具（如果使用了自己研发的软件或工具请具体说明）：

- Echarts、D3.js、Python、Express、Vue.js、Node.js
- VSCode、Pycharm、Excel、SVN、Git

共计耗费时间（人天）：60 人天

本次比赛结束后，我们是否可以在网络上公布该文档与相关视频（是或否）：是

一、 作品简介

1.1 作品主题

疫情潜在影响与次生灾害分析：利用可视化与可视分析技术，关联多源数据（比如：经济指标、物资采购、线上教育等），评估疫情对国家的经济发展、人民出行、教育培训和国民心理等方方面面带来的影响，防控复工复产困难、物资供需失衡、民众心理创伤等疫情次生灾害的发生。

1.2 问题场景

自 2019 年 12 月新冠疫情发生以来，已经蔓延至全国及世界各地。全国采取了有史以来最为严格的封城措施阻断疫情传播，然而新冠病毒较强的传染性、致病性、及隐匿性给疫情及其次生灾害的防控带来了巨大的挑战。疫情对国家的经济发展、人民出行、教育培训和国民心理等诸多方面产生了较大影响。分析及评估新冠疫情所产生的影响，防控次生灾害的发生具有重要意义。本系统以此为背景，主要工作如下：

- 关联多源数据，利用可视化的方式直观有效地展示疫情对国家的经济发展、企业生产、教育培训和国民心理等方面的具体影响
- 对多源数据进行可视分析，评估疫情带来的潜在影响
- 根据可视分析和评估的结果，提出可行的、具有参考意义的防控措施

1.3 目标用户

1. 关注疫情影响的普通群众；
2. 需要分析及评估疫情影响并制定相关防控措施的政府部门；
3. 需要基于疫情影响做出合适决策，制定相关策略的企业与机构。

1.4 应用价值

本系统利用可视化与可视分析技术，充分结合疫情、政策、经济、交通、教育、心理等多源数据，清晰展示了疫情对各个方面带来的影响，并给出了制定相关的防控措施的建議。

1. 疫情、防控政策等权威数据有助于分析疫情的发展态势及政府的相关防控举措，帮助政府部门评估政策措施的有效性，改进已有政策并制定新的政策。
2. 对经济、交通数据的分析展示了疫情对国民经济、企业生产等产生的宏观影响，有助于针对性地制定合适的经济应对策略，确保企业正常运行，经济稳定增长。

3. 对教育数据的分析不仅展示了疫情对各级学校复学以及重要考试的影响，同时反映了对线上教育机构发展的促进和对线下教育机构的冲击。为各种学校及教育机构改进自身教学模式提供了参考。
4. 对心理数据的分析展示了武汉人民心理状态的变化趋势和相关原因，帮助相关部门在控制疫情的同时制定合理措施，以保证疫情重灾区民众的心理健康。

二、 数据介绍

为充分分析疫情对国家各领域所造成的影响，现从疫情、经济、交通、教育、心理以及武汉政策措施六个方面介绍系统所使用的数据。

2.1 疫情数据

1. 数据来源及内容介绍

疫情数据从GitHub开源爬虫项目2019新型冠状病毒（COVID-19/2019-nCoV）疫情状况的时间序列数据仓库(<https://github.com/BlankerL/DXY-COVID-19-Crawler>)中获取，包括从1月20日至5月31日丁香园的公开数据，数据格式为csv，主要字段包括日期(Date)、地区(string)、新增确诊人数(int)、治愈人数(int)、死亡人数(int)以及累计新增(int)、治愈(int)和死亡人数(int)统计。

2. 数据评价

数据源于丁香园实时更新，其中湖北省的数据曾有部分存在数据异常，原因为国家卫健委发布第六版新冠肺炎诊疗方案后诊断标准被修改，出现了确诊病例核减的情况，因此根据湖北省发布的官方数据修改为正常数据。

2.2 政策数据

1. 数据来源及内容介绍

- 消费券发放数据：采集于深圳“本地宝”微信公众号，原始数据整理为csv格式，包括6月份之前的所有信息，主要字段包括地区(string)、时间(Date)、金额(float)以及类型(string)。
- 武汉政策措施数据：主要统计于武汉发布官方微博，其中1月19日至2月17日的数据来自石墨文档(<https://shimo.im/sheets/HdDdQrXvPV8XhWGY/MODOC>)，后续三月数据从官方微博爬取，并进行人工标注过滤及分类，数据主要包括发布日期(Date)、所属类别(string)以及具体政策措施(string)三个字段。

2. 数据评价

发放消费券是各地（尤其疫情重灾区及经济发达地区）为刺激经济更快更好复苏而推行的经济类政策，从深圳“本地宝”微信公众号人工采集，该数据来源于官方发布，准确可靠。

武汉市的措施数据由官方发布中获取，包括政策和行动，前者一般有具体名称，包括政策、规定、通知、公告、倡议等，如《武汉市新型肺炎防控指挥部通告（第 1 号）》；后者为具体行动，如“来京人员手机登记报到小程序上线”。官方微博发布的信息相对实时性高且全面，因此将其作为主要信息来源。

2.3 经济数据

1. 数据来源及内容介绍

- 宏观经济相关数据：整理于国家统计局官方发布，原始数据格式为 csv，包括 2019 年 4 月至 2020 年 4 月的累计增长速率数据，主要字段包括日期 (Date)、消费零售总额增长率(float)、工业增长率(float)、进出口增速(float)等主要经济指标（为排除春运影响,原数据中 1 月与 2 月统一合并为 2 月数据）。宏观经济数据用于反映疫情期间经济的整体走势，同时细分类别数据用于反映各行业受疫情影响的不同波动；
- 代表物资采购指数：获取于阿里指数平台，数据格式为 json，选取了一次性医用耗材、消毒产品和面部防护等三项主要医疗物资的供应量和采购指数。

2. 数据评价

宏观经济数据采集于国家统计局公布数据，数据严谨权威，通过社会消费品零售总额增长率、工业增长率、进出口增速等主要经济指标可以总体反映疫情期间国家整体的经济发展趋势；同时，针对上述每一个宏观经济指标还统计了该指标下各行业的变化数据，如消费零售总额数据可细分为汽车类、通讯类、日用品类、烟酒类等，利用整体指标和各行业细分类数据从宏观和微观层面全面分析疫情对国家经济的影响。

除具有参考价值的经济指标外，还选取了与疫情密切相关医疗物资采购数据进行分析。数据获取于阿里指数，作为了解电子商务平台市场动向的数据分析平台，该网站数据由阿里巴巴每日运营的基本数据统计得出，数据完整准确，通过对供应指数以及采购指数的整理，可以反映出一定时间内相关物资的市场供求情况。

2.4 交通数据

1. 数据来源及内容介绍

- 迁徙数据：包括 2019 年 1 月至 4 月和 2020 年 1 月至 4 月的国内整体迁徙指数(float)以及 2020 年 1 月至 4 月各省份的迁出指数(float)。

- 各省市邮政快递增速数据：统计于国家邮政局每月公布的邮政行业运行情况报告，包括 2019 年 12 月至 2020 年 4 月的增长速率(float)。
- 公路、水运、铁路、民航以及城市客运量数据：统计于国家交通运输部，包括 2019 年 12 月至 2020 年 4 月各类别的客运量(longint)以及货运量(longint)。

2. 数据评价

人口迁徙数据爬取于百度地图迁徙大数据，该数据源于百度地图获取数亿部手机的定位信息，数据可信度高，能够反映出时间段内人群迁徙的轨迹以及强度信息，判断疫情对人们出行的影响，并且结合不同时间的迁徙规模指数可以在一定程度上判断各省市企业复工的情况。

各省市快递业务数据由国家邮政局公布，通过快递业务量的增减率可以看出各省市的物流以及交通出行受疫情的影响程度。

国内整体客运货运数据整理于国家交通运输部，包括公路、水运、铁路、航空以及城市客运数据，用以反映疫情期间各类出行方式受影响的程度。

2.5 教育数据

1. 数据来源及内容介绍

- 移动应用增长榜数据：七麦数据平台统计的 2019 年 12 月至 2020 年 4 月的全平台移动应用下载增速排行数据，数据主要字段包括应用名(string)、排名(int)、分类(string)以及所属公司名(string)。
- 各省复学时间：包含疫情期间各省市发布的各年级复学时间(string)以及中高考时间(string)，数据爬取于各省市官方账号最新发布的信息。
- 教育机构股价数据：获取于老虎证券平台，包括新东方、好未来等主要线上教育机构在 2019 年 5 月至 2020 年 5 月期间的每周股票收盘价，数据字段主要包括日期(Date)、类别(string)以及收盘价(float)，原始数据为 json 格式。
- 线下教育机构数据：来源于中国民办教育协会发布的《疫情期间培训教育行业状况的调研报告》，从中提取了受疫情影响的线下教育机构对上半年营收的预期，以及应对资金短缺、经营困难的策略。

2. 数据评价

移动应用增长榜数据从国内专业的移动应用数据分析平台七麦数据获取，由每月下载量增速进行排行，由疫情期间移动应用下载量的变化，可以用以判断教育模式在疫情期间的转变，并在一定程度反映出疫情期间复学总体趋势。

各省复学时间源于各省市教育部门官博，清晰地体现各地教育随疫情演变而做出的调整，反映疫情对各地开学、复学、中高考的影响。

教育机构股价数据选取主流的教育机构进行数据整理，根据选取的典型机构，通过每周股票收盘价判断各机构在疫情前后的发展趋势，用以分析线上教育受疫情的影响。

线下教育机构数据源于行业协会的调研报告，具有整体性、广泛性、准确性，能够较为全面地反映线下教育机构在疫情下的总体状况，评估疫情对线下教育的影响。

2.6 心理数据

1. 数据来源及内容介绍

- 央视新闻热门微博数据：为 2020 年 1 月至 5 月的央视新闻官方账号(微博 ID 为“央视新闻”)所发布的微博中评论数量超过五千的所有微博数据，原始数据以 csv 格式存储，包括微博文本(string)、发布时间(Date)、转发数(int)、评论数(int)、点赞数(int)。
- 武汉日记超话数据，爬取了微博中从 1 月 20 日至 5 月在武汉日记超级话题下的所有原创微博数据，原始数据以 csv 格式存储，包括发布用户昵称(string)、微博文本(string)、发布时间(Date)、转发数(string)、评论数(string)、点赞数(string)。

2. 数据评价

央视新闻热门微博数据通过从央视新闻官方微博筛选讨论量较高、热度较大的微博来反映疫情期间民众的关注点变化，具有代表性。

武汉日记超话数据是在武汉日记超级话题下筛选出的疫情期间的原创微博信息，对于分析处于疫情重灾区的武汉人民的心理变化情况具有参考价值。

上述数据汇总结果如表 1 所示：

表 1 数据来源于数据字段汇总表

数据集名称	来源	主要字段
全球疫情数 (2.20-5.31)	github 开源项目： https://github.com/BlankerL/DXY-COVID-19-Crawler	地区、日期 (天)、累计确诊、治愈、死亡、新增确诊、治愈、死亡人数
宏观经济数据 (2019.2-2020.4)	国家统计局	类目、日期 (月)、增长率
各省市消费券发放 详情	各省市“本地宝”微信公众号	地区、日期、金额、类别
代表物资采购指数 (2019.6-2020.5)	阿里指数（数据分析平台）	类目、日期 (月)、采购指数、供应指数
各省市快递数据 (2019.1-2020.4)	国家邮政局	地区、日期 (月)、快递业务 增率

客运货运民航数据 (2019. 1-2020. 4)	国家交通运输部	类目、日期 (月)、客运指 标、货运指标
整体迁徙指数 (2019, 2020)	百度地图迁徙大数据平台	日期(天)、迁徙 指数
各省市迁徙数据 (1. 1-4. 26)	百度地图迁徙大数据平台	地区、日期 (天)、迁出指 数、迁入指数
移动应用增长榜 (2019. 12-2020. 4)	七麦数据网站（移动产品商业分析平台）	日期(月)、应用 名、排名、所属 类别、所属公司 名
各省复学时间	各省官方微博	地区、日期 (天)、毕业/非毕 业年级复学时 间、中高考时间
主要线上教育机构 股价数据 (2019. 5-2020. 4)	老虎证券（互联网券商）	日期(周)、机构 名、股票收盘价
线下教育机构数据	疫情期间培训教育行业状况的调研报告	上半年预期营 收、应对经营困 难的措施、及所 占比重
央视新闻热门微博 (1. 1-5. 31)	新浪微博央视新闻官方账号	日期(天)、高频 关键词
武汉日记超话微博 数据 (1. 24-5. 30)	新浪微博	日期(天)、高频 关键词、关键词 正负评分
武汉政策数据 (1. 17-2. 17)	石墨文档：新冠肺炎的举措汇总 https://shimo.im/sheets/HdDdQrXvPV8XhWGY/MODOC	日期(天)、举措 信息、类别
武汉政策数据 (2. 18-5. 20)	新浪微博武汉发布政府官方账号	日期(天)、举措 信息、类别

三、 分析任务与可视分析总体流程

3.1 分析任务

根据所需解决的问题，总结出主要任务共包括两个方面：

1. 评估疫情影响

评估疫情对各个方面带来的影响，这个任务需要搜集大量数据进行详细的分析并展示；首先为全方面地评估疫情带来的影响，整理了包括经济、教育、国民心理和交通出行等各方面的数据。对搜集到的数据进行筛选、清洗等预处理，找到各个方面的代表性数据和数据具体属性。随后使用图表可视化所选数据，展示结果，以便向用户清晰展示疫情带来的具体影响。

2. 防控疫情次生灾害的发生

这个任务则需要将融合多源数据进行关联分析，挖掘深层信息，以便进行分析。对于防控疫情次生灾害的发生，需要找出次生灾害发生的原因和带来的影响，并根据分析的结果找出相应的解决方案。对于这个方面，单纯从某一个数据来进行分析无法得出结论。需要对已找到的数据进行关联分析，通过这些多源数据之间的联系和差别剖析出更深层次的信息来进行分析。

3.2 可视分析总体流程

本系统的可视分析具体流程如图 1 所示。



图 1 可视分析总体流程

1. 明确主题

本系统旨在展示疫情对国家各方面的影响范围与程度，根据主题确定可视化设计中需要使用的数据和视图。

2. 采集数据

根据系统主题，确定需要疫情、政策、经济、交通出行、教育培训和心理共六个方面的数据，这些数据需要从不同平台中获取，采用不同的方式获取。

3. 数据处理与变换

数据可视化的基础是数据表示与变换，为了实现有效的可视化、分析和记录，输入数据需要从原始状态到一种便于处理的结构化数据表示形式。从各种数据源获取的原始数据大多是“脏数据”，不能直接用于信息展示，需要进行数据清洗、数据集成与数据变换等处理，去除数据噪声，完善数据，最后将数据格式转换为图表需要的格式。

4. 可视化映射

将原始数据处理完成并转换为结构化表示形式后，根据数据特征和要表达的信息选择多种图表、设计图表元素，将数据以一种直观、易于理解和操纵的方式呈现出来。

5. 交互设计

通过图表之间的交互，能够标记感兴趣的数据对象，提供观察数据的不同视角，方便查询和跟踪变化情况。

6. 分析推理

分析推理挖掘出深度信息是可视化系统的最终目标，通过视图展示的直观信息和彼此之间的交互，洞悉疫情对经济、交通出行、教育培训和心理等方面的影响范围和影响程度，探索出疫情带来的次生灾害，从而有针对性地给出防控措施的建议。

四、 数据处理与算法模型

4.1 数据清洗

从各种途径获取的原始数据大多存在数据内涵不一致、格式不统一，数据冗余、不完整、纬度高、错误或异常等问题，直接使用这些“脏数据”会带来低质量的分析结果，需要进行数据清洗。

1. 疫情数据

缺失值处理：疫情数据获取于 GitHub 的开源项目，首先使用 python 语言校验数据的完整性。本项目从数据行角度判断原始数据集是否存在缺失，判断每行是否包含缺失值，并标注缺失行。由于官方每日都会公布具体的数据，因此，对于这些缺失数据，本系统从官方发布的通告中进一步查找存在缺失数据当日的数据。

异常值处理：缺失值处理完成之后，为保证疫情数据准确性，本项目获取 GitHub 上的开源项目的数据后，又自行爬取了网易新闻疫情地图所展示的数据并进行对比，针对其中存在出入的数据，通过官方发布进行核实订正，确保数据的准确性。

2. 政策数据

数据整合：原始政策数据来自于石墨文档整理以及微博爬取，两类数据原始格式存在较大差异，故对其进行数据整合。

1 月 19 日至 2 月 17 日的政策数据来源与石墨文档，首先将其重新划分为综合、公共交通与社区、交通、教育、企业、医疗以及心理七个分类。后续政策数据由武汉发布官方微博获取，其中会有重复发布的情况，处理时对此进行了人工清洗，并且根据信息中包含的主要关键词将其进行类别划分，保留政策发布日期以及发布的具体内容，最终将 1 月 19 日至 5 月 20 日的政策数据统一划分并整理为可读数据。

3. 宏观经济、交通与邮政数据

数据转换：宏观经济数据以及交通邮政数据来源于官方发布，对数据的主要处理为统一整理出相同时间段内数据，按照所需格式转换，将多源数据进行整合。

缺失值处理：迁徙数据爬取过程中，百度地图慧眼团队于 2020 年 5 月 8 日停止迁徙大数据平台的数据更新，导致部分日期迁徙数据存在缺失值。故本系统在进行数据处理的过程删除了不完整的数据，最终保留了 1 月 1 日至 4 月 26 日的完整数据。

4. 心理数据

对于心理类数据，从微博上爬取的数据有两种类型，一类是央视新闻官方微博发的微博，另一类是武汉超话中所有用户发出的微博，针对这两种微博数据，使用的清洗步骤一致，进行具体处理时使用的算法不同。

过滤特定类型的重复文本：原始数据中的重复数据有两种类型，一类是由营销号、官方微博账号发的微博，这类重复数据需要被过滤；另一类重复数据是一般用户发出的相同的正文，如“武汉加油”，这类文本是不同个体的内心真实想法的写照，不能被过滤。

去除重复值分两步进行，首先找出原始数据中的所有重复值，并提取出这些文本的所有数据项。然后，本项目以微博正文的长度作为衡量重复数据类型标准，若重复文本的长度超过一定值，则认为是由营销号和官方微博账号发出还是由一般用户发出，反之为一般用户的微博，本项目根据数据特点，选取阈值为 15。

处理不完整数据：原始数据中包含多个字段，用于分析的字段为微博日期、微博正文，其余字段的确实不会影响分析过程，因此，需要检查的内容为日期和正文。对于有缺失的数据，直接从源数据中删除。

去除非本文数据：原始数据中存在着较多非文本的数据，如一些图片、表情包、颜文字等数据，这些数据属于噪声，会影响处理结果，因此需要将它们剔除。

4.2 数据预处理与算法模型

4.2.1 Pearson Correlation Coefficient(PPC):居民消费价格指数相似

性度量

相似性能够反映受疫情影响结果相似的省份，该系统以各省份的居民消费价格指数增长率为基础，采用 PPC 进行分析，计算存在一定相似度的省份。

协方差能反映两个随机变量之间的相关关系，如式(1)所示。

$$\text{cov}(X, Y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n-1} \quad (1)$$

其中， $X = (x_1, x_2, \dots, x_n)^T$, $Y = (y_1, y_2, \dots, y_n)$ 。

Pearson 相关系数是一种统计学方法，可以定量地衡量变量之间的相关关系，其定义如式(2)所示：

$$r(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var[X]Var[Y]}} \quad (2)$$

其中， $Cov(X, Y)$ 为 X 与 Y 的协方差， $Var[X]$ 为 X 的方差， $Var[Y]$ 为 Y 的方差，该公式剔除了量纲的影响。如果所有的点都分布在一条直线上，那么相关系数就是+1 或-1，这取决于直线的斜率是正的还是负的，如果相关系数等于 0，那么 x 和 y 之间就没有相关性，相关系数取值与相关程度之间的具体关系如表 2 所示：

表 2 系数相关性

相关系数	相关程度
$ r = 0$	不相关
$0 < r \leq 0.3$	微弱相关
$0.3 < r \leq 0.5$	低度相关
$0.5 < r \leq 0.8$	显著相关
$0.8 < r \leq 1$	高度相关
$ r = 1$	完全相关

4.2.2 结巴分词：央视新闻官方微博数据处理

首先对央视新闻官方微博数据进行处理，本项目拟从央视新闻的微博中获取人们关注的内容，进而分析出人们关注点的变化趋势以及导致这些变化产生的原因等深层次信息，所以本项目将预处理之后的数据以周为周期，再按周统计关键词及其词频，处理流程如图 2 所示。

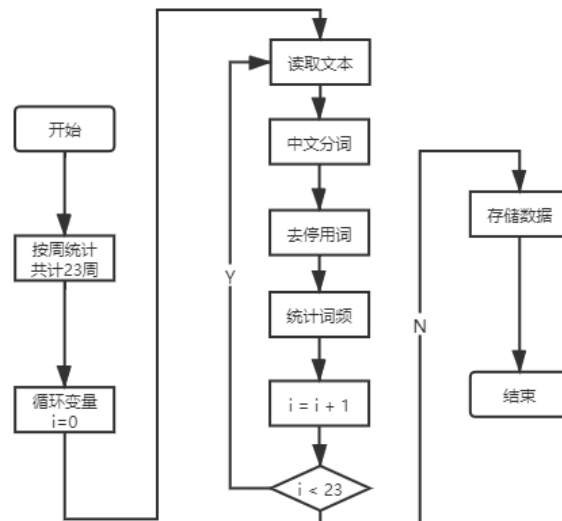


图 2 央视新闻官方微博数据处理流程图

1. 数据分段

对于央视新闻官方微博数据，本项目以时间线展示人们关注的内容，以周为单位，因此首先将与处理所有数据以周为单位进行统计，便于后期处理。

2. 结巴中文分词

● 分词模式

结巴分词支持三种分词模式：全模式、精确模式和搜索引擎模式，其中全模式将句子中所有可以成词的词语都扫描出来，速度非常快，但是不能解决歧义；精确模式将句子最精确地切开，适合文本分析；搜索引擎模式是在精确模式的基础上，对长词再进行切分，提高召回率，适合用于搜索引擎分词。根据数据特点，本项目采用精确模式进行分词。

● 算法介绍

结巴分词涉及的算法包括：

- 基于 Trie 树结构实现高效的词图扫描，生成句子中汉字所有可能成词情况所构成的有向无环图（DAG）；
- 采用动态规划查找最大概率路径，找出基于词频的最大切分组合；
- 对于未登录词，采用基于汉字成词能力的 HMM 模型，使用 Viterbi 算法。

3. 去除停用词

本项目使用的停用词是以哈工大停用词库为基础，并在此基础上根据自身需要添加额外的词汇和符号，包括没有意义的连词、副词、语气词等内容。

4. 统计词频

经过前面两个步骤之后，已经将文本中的数据变成需要的单个词的模式，能够用来统计词频，词频使用 Counter 函数统计，即可统计数据中每个词出现的次数。

对所有数据执行上述步骤后，即得到每周内词频较高的词语，数据存储格式与对应图表要求的数据格式一致，为字典类型：{"value": words, "frequency": frq}，其中 words 表示一个词，frq 表示这个词的词频，词频越高，说明被提及的次数就越多，即关注度就越高。

此时得到的词并不能都可用于表示出一个关注点，还需要通过人工筛选，从所有高频词中选出体现为一个事件、一种情绪、一个人等内容的具有代表性的词汇，使得这些词能够直观具体地表示出关注的内容。

4.2.3 情感词典：武汉超话微博数据处理

本项目使用情感词典获取武汉人民每天的宏观情感极性，进一步展示出人民的心理变化情况。经过预处理后，武汉超话微博数据有效数据共计 41904 条，日期范围为 2020 年 1 月 24 日至 2020 年 5 月 30 日，平均每天 320 条数据。本项目基于这些数据，挑选了合适的情感词典算法，以此为依据统计出人们的情感极性：正面、负面、中性。

1. BosonNLP 情感词典介绍

通过情感分析调用模块调用短评文本对语句进行情感分析，并给出单个文本的情感得分。BosonNLP 是基于微博、新闻、论坛等数据来源构成的情感词典，

主要利用情感词判断一句话或一篇文章的情感倾向，本项目使用 BosonNLP 情感词典作为词汇本体，对其进行扩充，同时使用否定词词典、程度副词词典和停用词词典作为情感分析的基础计算单个文本情感得分的计算，如公式(3)

$$senti_score = (-1)^n \sum \sum v_{ij} w_i sensibility_i \quad (4)$$

其中， v_{ij} 代表第*i*个情感词的第*j*个程度副词， w_i 代表第*i*个情感词的权重， $sensibility_i$ 代表第*i*个情感词， n 代表否定词的个数。由于程度副词*v*取值均为正，所以情感得分的正负取决于两部分，一部分是情感词，如果是积极情感词，那么 $sensibility_i > 0$ ；如果为消极情感词，那么 $sensibility_i < 0$ ；中性情感词的 $sensibility_i$ 接近于 0。另一部分为否定词的个数，由于每条微博的字数与内容没有进行限制，所以该计算方法的取值范围为正无穷到负无穷，得分越高，正面情感倾向越强，得分越低，负面情感倾向越强。

2. 计算情感极性

● 分词并去除停用词

使用 python 语言进行处理，利用自带的 jieba 库对文档分词；这里使用的停用词表与处理央视新闻官网微博使用的停用词表不同，需要过滤掉的内容较少，否定词与程度副词不能被过滤，需要去除代词、连词等内容。

● 构建词语序列

经过上一步的处理，得到分词后的数据，然后将它转化为词典类型，词典的格式为{'key': value}，其中 key 表示一个词或词语，value 表示 key 在分词结果中的索引。

● 对词语序列结果分类，获取情感词、否定词、程度副词

从情感词典、否定词词典、和程度副词词典中获取数据，对上一步得到的词语序列进行分类，得到的数据格式为{key: 'value'}，其中 key 表示词的索引，value 表示程度值，即词语的分值。

● 计算情感得分

计算得分的具体步骤如下：

- 找出所有情感词的下标，构建新词组；
- 新词组构建方法：该情感词与前一情感词之间的否定词及程度副词与该情感词之和（第一个情感词前至句首）；
- 计算：程度副词的程度值与情感词的情感值相乘，每增加一个否定词，使该式符号取反；
- 句子累加。

执行完上述步骤后，就可得到一段文本的情感得分。通过对每一条语句执行上述步骤，即可得到所有语句的情感极性。

● 分段统计词频

得到用户每天的情感极性分布情况只是从宏观角度进行展示，还需要从微观角度获取细节部分。

在计算情感得分的步骤中，能够得到每一段文本的具体得分数值，根据情感得分总分布趋势，本项目将每周中用户的情感分为 11 个等级，包括 5 种正面倾向等级，5 中负面情感等级和中性情感。正面情感得分区间为 $(0, 20]$ 、 $(20, 40]$ 、 $(40, 60]$ 、 $(60, 80]$ 、 $(80, +\infty)$ ；负面情感得分区间为 $(-\infty, -80]$ 、 $[-80, -40]$ 、 $[-40, -20]$ 、 $[-20, -10]$ 、 $[-10, 0)$ ；中性等级情感得分为 0。

依据情感等级，将一周内的所有数据分为 11 组，代表 11 种不同等级的情感倾向。然后使用 4.2.2 节中处理央视新闻官方微博数据的方法统计每种情感倾向中的高频关键词。这些关键词是用户关注点的具体表现，从而能够结合情感极性的变化情况，确定他们情感变化的原因，即是说，如果在某一时期内他们的情感普遍偏向负面，就可以根据在该阶段内的关键词找出导致他们产生负面情绪的原因。下图展示了处理武汉超话微博单元数据的总体步骤：

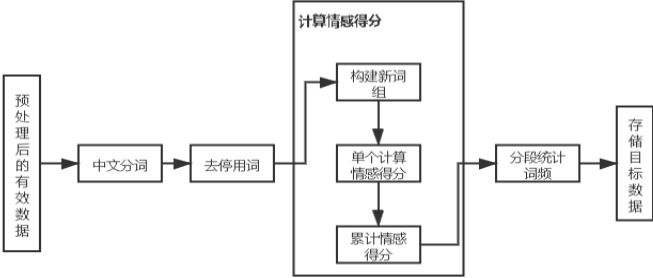


图 3 处理武汉超话微博数据流程

五、 可视化与交互流程设计

为完成本系统的分析任务，需要综合使用多种可视化图表进行展示与分析。对于分析任务一，要求设计的图表能够直观展示出各类别数据，让用户能够清晰地了解疫情带来的影响。对于分析任务二，则需要该作品进行多视图关联展示，用户可以通过视图的交互和关联挖掘到更深层次的信息，进一步得出解决方案。基于这两种目标，本系统对于不同模块的信息展示设计了不同的可视化视图与交互方式。

5.1 视图设计方案

针对疫情、经济、教育、心理、交通、政策等六个主要方面，系统根据所用数据特征分别选择合适的映射方式，设计合理的视图、布局与色彩搭配。



图 4 系统整体布局图

5.1.1 疫情发展态势版块视图设计

疫情数据具有时间维度与地理空间维度，精确地表明了各地区每日内新增确诊人数、治愈人数、死亡人数以及累计新增、治愈和死亡人数统计等多种统计数据。为清楚展示疫情的地理空间维度，本系统使用可切换视角的平面地图（图 4-b），该地图能够从宏观层面展示疫情的总体分布区域，便于对感兴趣的区域进行详细的分析；

同时，设计面积图（图 4-a）具体展示每个区域中的各项数据，图形中面积的大小能够反映出当地数据的具体数值，色彩用以区别不同的数据指标，该面积图能够与地图联动，直观呈现出各地疫情的影响程度与变化趋势。

5.1.2 政策版块视图设计

对于政策版块，本系统使用了消费券发放信息和武汉的政策措施数据。为了展示各地区消费券发放的具体情况，使用矩状树图（图 4-c）来进行对比展示。

在矩阵树图中，以矩形之间的包含关系，展示各地区之间的层次结构关系；以矩形面积大小对比，直观展现不同地区发放的消费金额的大小对比关系。矩阵树图支持下钻操作，能层次化地展现具体的数据信息，避免一次性显示左右数据而导致视图杂乱。

该作品采用日历图和饼图详细展示武汉政策措施，具体展示内容为每天不同类型政策的发布数量。由于每项政策与日期联系紧密，因此选用日历图作为主要展示方式。在此基础上，采用饼图表示不同类型的数据，既能直观表示出当天发布的政策类别，也能突出每种类别的数量关系。

5.1.3 经济版块视图设计

对于经济版块，本系统选用国家统计局官方发布的宏观经济数据和阿里指数平台发布的代表性物资采购指数。代表性物资采购指数使用折线图（图 4-e）

进行展示，折线的变化趋势直观反映这些物资采购数量的变化情况，辅以颜色表示物资类别，图表简洁直观。

为了对三大产业的增速以及各 GDP 占比数值进行对比，本系统采用堆叠柱状图和折线图的复合图（图 4-f）从多层次多角度来进行展示。其中，堆叠柱状图展示各产业的具体 GDP 数值，折线图展示各产业的增速，堆叠柱状图在纵向清晰地比较三大产业的 GDP 数值，在横向呈现每一产业的 GDP 变化趋势。

同时，本系统使用柱状正负图（图 4-h）和散点正负图（图 5）展示国民经济主要指标的增长率以及这些指标的详细数据。在柱状正负图中，横向展示时间维度，使用颜色属性区分不同年份；纵向展示不同类别指标的具体增长率，正负值在异侧，使得各指标的差异一目了然。散点正负图可以同时展示更多数据指标的变化，便于按行列进行对比，空间利用率高，通过对柱状正负图进行交互出现，用于展示某一经济指标下具体产业的增长量。其中，散点的大小用以表示增长量数值的大小，散点的颜色区分增长的正负，横向对比能够得到一个产业的生产值受疫情影响的程度，纵向对比得到各产业受疫情影响的不同程度。



图 5 散点正负图

最后，本系统使用关系图（图 4-g）展示各省份的经济发展态势相似性。关系图能够在具有一定相似度的省份建立连接，使得省份之间的关系更加清晰。

5.1.4 教育版块视图设计



图 6 教育板块视图设计

对于教育版块，本系统统计了移动应用增长榜数据，各省的复学时间和考试变化时间，以及主要线上教育机构的股价数据和线下教育机构的基本情况。

对于移动应用增长数据，本系统使用排行榜（图 6-b）进行展示，排行榜可以直观体现各 APP 下载量增长，又可以方便用户直接获取信息。并采用了折线图（图 6-c）和柱状图（图 7）来分别对主要线上教育机构的股价数据和线下教育机构的基本情况进行展示，可以清晰的展示出疫情对线上教育机构的促进和对线下教育机构的冲击。

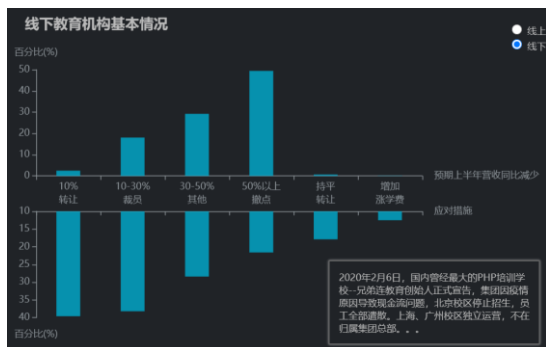


图 7 线下教育机构基本情况

最后，本系统采用了热力图展示各省各年级的复学时间和考试变化时间。该数据具有时间信息、地理空间信息以及教育阶段信息，因此用横轴表示省份，纵轴表示时间，颜色表示不同教育阶段，纵向呈现一个省发布各教育阶段的复学安排的时间，横向对比各省发布相关通知的时间先后关系。一定程度上折射出这些省份防控疫情的有效程度。

5.1.5 心理版块视图设计

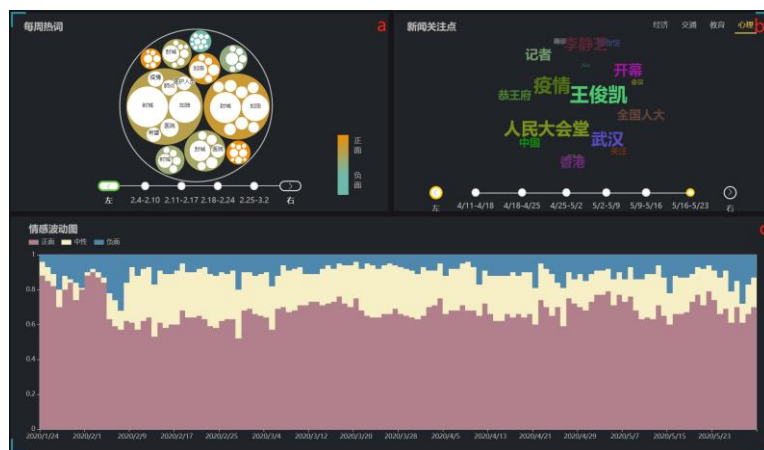


图 8 心理版块视图设计

对于心理版块，本系统采用了央视新闻热门微博数据和微博超话下的武汉日记关键词数据。为了向用户直观的展示群众关注的所有重点内容，本系统设计 3 种图表从不同角度和关注点展示数据。

首先使用 3D 动态词云（图 8-b）展示样式新闻热门微博数据的主要关注点，能够从不同维度展现数据：时间、词汇本身、词汇的颜色。由于获取的原始数据已经是关注度较高的内容，因此只需要展示出这些关注点的具体内容，而无需区分这些关注点的被关注程度的强弱。

然后，本系统设计堆叠柱状图（图 8-c）和打包图（图 8-a）从多个层面展示武汉超话微博数据。其中，堆叠柱状图用以展示疫情期间民众情感极性的波动情况，该图形成的曲线既能呈现民众情感随疫情发展而变化的整体趋势，也能表示出三种情感极性的精确占比数值。打包图用于辅助堆叠柱状图分析情感极性的变化原因。打包图以周为单位，通过视觉映射表示不同情感等级，利用圆圈之间的嵌套关系表示情感等级与关注点的对应关系，外层圆的大小表示了每种等级的占比，里层圆的大小表示群众对各个关键词的关注度，圆的数量体现了讨论热度。该图通过层级之间的关系，既能够直观呈现每个阶段内不同情感等级下的关注点，这些关注点是引起民众心理状态变化的主要原因；也能从时间维度上展示出民众参与讨论的积极性的变化情况。

5.1.6 交通版块视图设计



图 9 交通版块视图设计

对于交通版块，该作品采用了各省的迁徙数据，邮政快递数据，公路、水运、铁路以及民航的客运货运数据以及城市客运数据。

本系统使用折线图（图 9-a 和图 9-b）来展示公路、水运、铁路、民航以及城市客运这些交通运行数据。通过纵向对比直观展示出不同类别的交通出行方式受疫情影响的严重程度，并且可以从时间和类别两种视角进行对比。

对于邮政快递受疫情影响前后的业务量增速，本系统采用平行坐标系（图 9-c）来展示。纵向表示省份，横向表示时间，不同省份曲线采用相同色彩，避免使用太多颜色造成视图的混乱和过度修饰。

迁徙数据是时空数据，因此本项目使用平面地图（图 10）展示迁出情况，使用折线图展示整体迁徙情况。在地图中，以颜色属性展示迁出到不同目的地的人数占比情况，颜色越深，表明人数相对越多；通过设置时间，能够对前期和后期的数据进行详细对比。



图 10 迁徙情况

5.2 系统交互流程设计

本系统通过分析国内疫情发展态势，将系统整体分为ABC三个模块，通过关联分析，分别从经济、交通、教育、心理四个方面评估了疫情带来的影响。



图 11 系统整体视图

1. 模块 A

模块 A 展示了国外疫情、国内疫情以及国内迁徙指数。

在国外疫情、国内疫情这两个模块，用户可以通过与地图上的散点进行交互选择要查看的具体地区，同时疫情态势面积图会随之更新，展示该地区的详细疫情数据。

在国内迁徙指数模块中，用户能够通过对日期和地区的选择查看当日该地区到全国各地的迁出指数，反应出该地区人员流动情况。此外通过视觉映射组件可以对迁出地区的热度进行筛选，用于具体查看一定热度范围内的迁出目的地。

2. 模块 B

同时，用户可以在 B 模块查看消费券和武汉政策的相关信息。在消费券模块，用户可以通过点击矩形树图中某一区域来查看当地发放的消费券详情。

在武汉政策模块，饼图中不同扇形区域的颜色表示不同类型的政策，用户可以通过点击日历图中饼图中的一个扇形区域来查看某种类型政策的具体内容，选择某一扇形区域后，右边将会以文本形式展示出政策的详细内容，滑动日历框可以查看不同月份对应的政策发布情况，根据扇形区域颜色的种类变化可以判断不同时期政策措施针对的重点。

3. 模块 C

在 C 模块，用户可以分别查看经济、交通、教育和心理方面的数据。

在经济板块，用户可以通过悬浮来查看各类走势图的详细信息。正负柱状图通过居民消费指数、工业增长率等宏观指标反映国家宏观经济变化走势，点击正负柱状图中相应的标签名可以查看该经济指标下各行业的具体变化情况。

在 CPI 相似度关系图中，用户可以通过鼠标悬浮的方式查看与某个省份 CPI 走势相似度最高的三个省份。

在交通板块，用户可以通过鼠标悬浮来查看不同交通方式的客运货运量以及快递业务变化量。

在教育板块，用户可以通过鼠标悬浮查看折线图中主要线上教育机构的股票每周收盘价，切换可查看时线下教育机构的基本情况图；在排行榜中用户可以选择月份了解不同月份 app 下载的排行信息。在复学时间热力图中，点击相应省份复学时间色块，模块 A 中国内疫情图可同步切换为对应省份在该时间段的疫情变化走势，帮助协同分析疫情对于复学时间的影响，而直接选中省份标签则可查看各省的复学安排。

在心理板块，通过选择 3D 词云球中的不同时间段可以查看各相应时期的高讨论量热词，分析民众疫情前后的关注点变化。结合打包图和情绪柱状堆叠图，可以关联分析某一时间段内正负情绪的占比差异以及相应情绪下的关键词。对于民众情绪变化的起伏点，结合关键词进行深入分析可以发现产生相应积极或消极情绪的关键因素，在分析过程中可以据此挖掘民众心理健康的防范措施。

六、 实验\案例\场景分析

本系统对疫情对国家的经济、交通运输、教育培训和国民心理产生的影响进行了定量的分析，并针对各个方面产生的影响给出了相关的应对策略。

如图 12 所示，国内的疫情在二月中旬达到了顶峰，但在二月下旬便得到了较为明显的遏制和缓解。

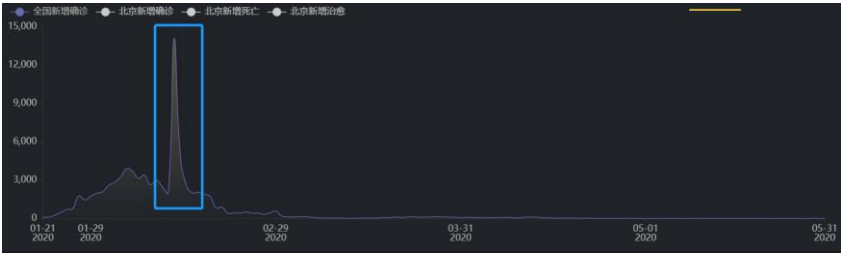


图 12 国内新增确诊人数

如图 13 所示，世界疫情在 2 月份左右开始出现病例，但此时增长趋势较为缓慢，在 3 月末，确诊人数增长幅度出现较大提高，并呈线性增长趋势，可以判断世界的疫情在 3 月中旬爆发之后一直未能得到一个良好的遏制。

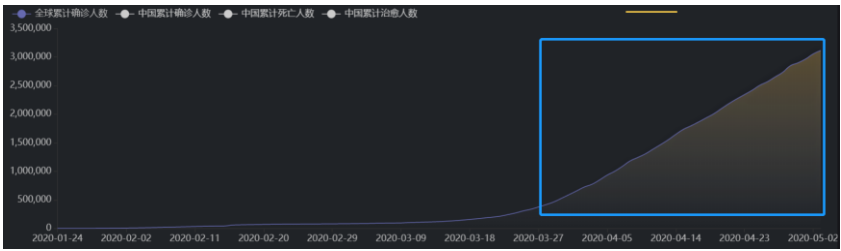


图 13 世界疫情累计确诊人数

6.1 经济模块场景分析

在经济方面，本系统对代表性物资采购、GDP 数值和增速以及国民经济主要指标增长率进行了可视化展示与分析。

如图 14 所示，对于代表性物资采购数据，可以明显发现，在疫情发生前期这些代表性物资的采购数量大幅升高，而其供应量却出现下降趋势，主要是前期疫情扩散较迅速，人们对防护产品的需求量迅速增长，但是由于此时企业还未全面复工，生产速度跟不上需求量的增长，导致物资缺乏。在 3 月中旬及以后，供应量保持上升趋势，一是因为部分企业已经复工，二是加大了生产防护用品的规模，但是在四月中旬之后，由于疫情的缓解，这些物资的采购量出现了迅速下滑的状态，但是其供应量在逐步上升。对于这些应急性物资而言，政府需要做好提前的储备，以防止出现疫情初期的供不应求的状态。同时也要做好市场的规范化，防止在疫情之后出现大量不合格物资涌入市场。

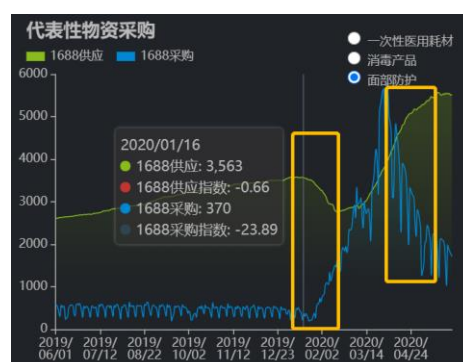


图 14 代表性物资采购

而对三大产业和国内生产总值 (GDP) 数值及增速，如图 15 所示，国内的 GDP 数值在 2000 年至 2019 年期间一直处于增长状态。在 2003 年受到非典和 2008 年国内经济危机、汶川地震的影响，GDP 的增速虽然放缓，但 GDP 依旧保持低速增长的趋势。但是由于此次疫情的影响，在 2020 年第一季度，三大产业和 GDP 的增速在 20 年内首次出现负值，总量出现下降。第二产业因为它的工作性质，无法通过网络代替等原因所受冲击最大，第一产业由于是刚需受冲击最小，第三产业通过产业网络化减少了部分损失相对表现不错。由于全国性的停工停产和严格的出行、复产管制措施，第二产业在第一阶段所受冲击最大，国内部分供应链的暂时断裂，更是加剧的制造业在疫情中所受的冲击。因此，应该积极采取应对措施，在防控疫情的同时，稳定经济发展，防止对经济产生重大影响。

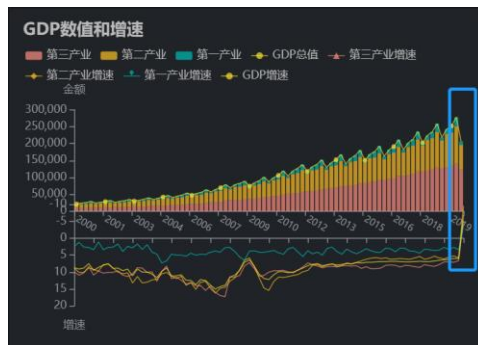


图 15 GDP 数值和增速

各省居民消费价格指数(CPI)如图 16 所示，箭头指向表明与被选中省份有较强相似度的城市。本系统对主要城市进行了分析，发现受疫情影响越严重的地区，其 CPI 的相关性改变越严重。如湖北地区，在疫情发生前，其 CPI 的变化趋势和北京，江苏等一些较为发达的城市相似。但在疫情发生后，其 CPI 的变化趋势则和宁夏、青海等一些较为落后的城市相似，而在这次疫情对这些地区的影响程度相对较低，表明此次疫情对湖北地区的经济影响程度较重。



图 16 疫情前（右）、后（左）与湖北具有强相似度的省份

对于经济的增长，各主要省市都采取发放消费券的形式刺激经济复苏。图 17 结合各省市的疫情发展趋势，在疫情重灾区或者经济发达地区，其消费券发放力度更大；而越早发放消费券的地方，那里的疫情就越先被控制住。

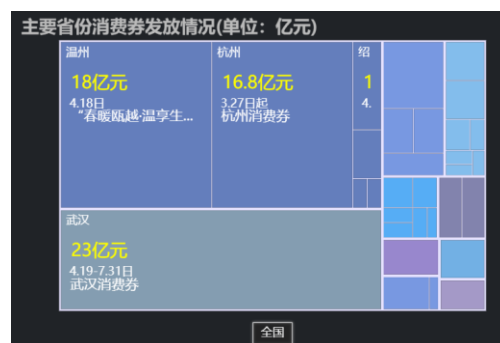


图 17 主要省份消费券发放

图 18 展示了疫情前后国民经济主要指标增长率，从图中可以看出在疫情发展最严重的 2 月份，各项指标增长率都基本为负值，并且其中固定资产投资额和房地产投资的负增长值最大。在三、四月份国内疫情逐渐得到控制时，这些主要指标都有小幅度的回升。



图 18 国民经济主要指标增长率

同时，通过对各个指标内的具体产业进行对比，可以发现在绝大部分产业产值降低的情况下，烟草制品业的利润依旧在稳步上升，如图 19 所示。经分析，出现该情况的原因在于疫情期间人民的娱乐活动减少，烟草的消费量小幅度提升。为了减少疫情对于经济的影响，国家需要根据不同地区不同情况采取不同的政策，对复工复产以及人民消费进行刺激，加快经济复苏的脚步。



图 19 详细指标增长量

6.2 交通模块场景分析

对于各省的迁徙数据，本系统采用了迁徙地图和迁徙折线图来进行展示。根据迁徙地图（图 20），在武汉封城前期，其主要的迁出地为重庆，南阳，信阳等地区。结合各地疫情发展趋势进一步分析可以发现，这些地区均是疫情重灾区。通过该现象，本系统建议疫情区应加强人员迁出的管控，迁出地也应采取更全面的防范措施对迁入人员进行检测、隔离、跟踪随访。



图 20 湖北 1.22 日迁徙数据

图 21 展示了国内迁徙整体数据，可以发现和去年同期相比，在疫情爆发之后，国内人员的迁徙量明显减少。在 1 月 23 日武汉封城之后，国内的迁徙数量呈现了大幅度的下滑。而 2 月中旬开始，迁徙人数逐步小幅度回升，经过分析，迁徙数量的增加的原因是部分地区开始复工复产。在 4.4 日清明节当天，迁徙人数首次超过去年同期迁徙人数。

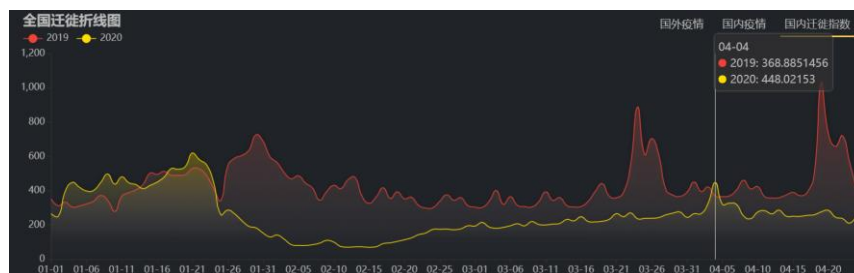


图 21 全国迁徙指数图

图 22 展示了公路、水运、铁路、民航以及城市客运在疫情前后的数据。可以发现在 2020 年的 2 月份，各类交通的客运量和货运量都呈现了大幅度的下滑。客运量的下滑较为剧烈，货运量的下滑则较为缓慢。其原因是在疫情发生后，大量的抗疫物资需要运输，而人员的流动则受到了较大的限制。3 月份之后由于疫情相对缓解，客运量与货运量都开始出现小幅度的回升，但公路货运量回升最快，抗疫物资仍主要依靠公路运输。

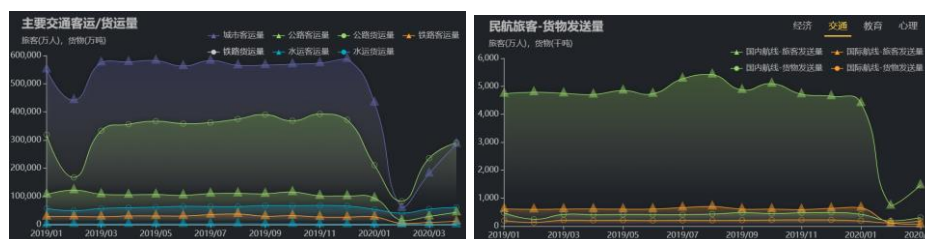


图 22 主要交通客运/货运量（左）和民航旅客-货物发送量（右）

各省的邮政快递数据如图 23 所示，在疫情严重的地区由于管控措施更严格，其快递业务在疫情期间的增速下降更快，比如湖北快递增速从 2020 年 1 月开始，均为负增长，为所有省市最低。

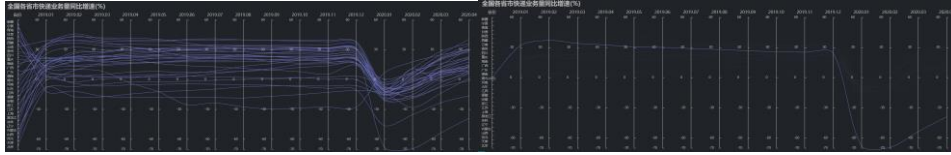


图 23 各省市快递业务同比增速（左）和湖北快递业务同比增速（右）

但是该项目发现，在图 24 中，疫情期间宁夏的快递业务的增速反而在升高。经过分析后，其原因有两点：宁夏在疫情期间的管控严格，其受到疫情的影响并不严重；宁夏的政府为了保证抗疫物资的准时送达，提出了相关政策来保障本省快递行业有序稳定运行。

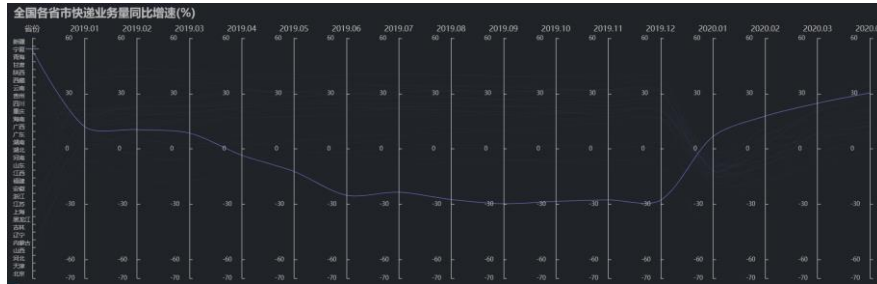


图 24 宁夏快递业务同比增速

疫情期间，全国各地的迁徙量减少以及各类交通货运量和客运量的减少是必须的。国家通过交通管制和社区封禁等活动减少人民出行，以达到防控疫情的目的，实践证明，这些举措有效地控制了疫情的扩散，使得能够在一个月左右的时间取得抗疫成果。而对于疫情爆发之后，对于疫情爆发地的管控和疫情爆发地重点迁出地区的管控也要更加严格，防止这些地区的疫情太过严重而得不到及时的控制。

6.3 教育模块场景分析

对于教育方面，本系统统计了移动应用增长榜数据、各省的复学时间和考试变化时间、主要上市线上教育机构的股价数据以及线下教育机构的调研数据。

移动应用增长榜数据如图 25 所示，钉钉等学习软件在今年的 3 月份飙升到了第一名。经过分析发现，钉钉等学习类软件热度飙升是由于国内的各类学校广泛开展了线上学习，学生为了进行线上学习开始下载这些软件。但是这个 APP 排行榜也反映了一个问题，各个学校对于线上教育都是首次使用，会出现各式各样的问题。诸如线上教育软件不统一，同一个学校可能会使用多个软件；教师和学生软件的使用不熟悉等现象。为了防止这种情况的出现，该项目建议各个学校可以在正常时期适当开展线上教育，以便在特殊时期学生与老师都能很好地适应从线下教育转换为线上教育。

热门App排行榜 2019-12			热门App排行榜 2020-03		
应用排名	应用名称	类别	应用排名	应用名称	类别
1	抖音短视频	摄影与录像	1	钉钉	商务
2	拼多多	购物	2	抖音短视频	摄影与录像
3	快手	摄影与录像	3	拼多多	购物

图 25 疫情前热门 APP 下载排行榜（左）和疫情期间

线上教育机构的股价数据如图 26 所示，股票收盘价高低往往反映出市场对某只股的关注程度，可以发现这些教育机构的股票在图中竖线之前，整体处于缓慢上升的状态。而疫情的发生后，这些线上教育机构股票的关注度明显上升。出现这个情况的原因，是由于线下教育被叫停，各个学校也开展了线上教育，使这些线上教育机构得到了迅速的发展。因此，可以认为，疫情从一定程度上推动了线上教育企业的发展。

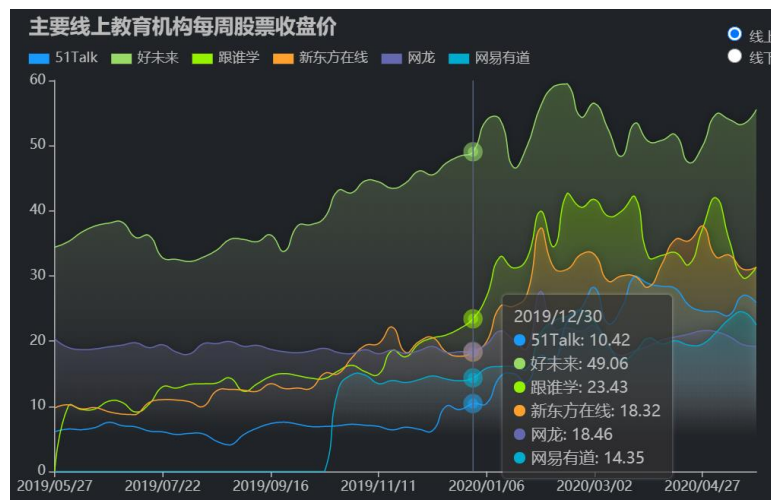


图 26 主要线上教育机构每周股票价收盘价

相比于线上教育，线下教育呈现了明显的相反的结果。如图 27 所示，线下教育的预期营收基本都有一定程度的降低，其中近一半的机构表示预期营收同比下降 50%以上。同时，为了应对疫情带来的影响，采取增资、裁员等应对措施线下培训机构都各占 40%左右，而裁员也会对就业形势产生较大的负面影响。这种情况，是由于疫情原因导致线下教育机构招生困难、开学困难而导致。比如曾经国内最大的线下 PHP 培训机构兄弟连教育因疫情导致资金问题而宣告‘破产’。因此建议相关部门对线下教育采取一定的减租减税措施，降低线下机构贷款门槛等政策扶持线下教育。而线下教育机构自身也需要改变自身经营理念，创新教育模式，以此来提升机构抗风险能力。

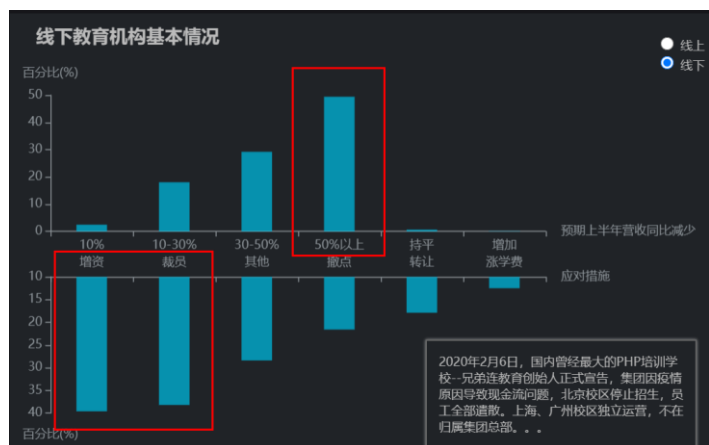


图 27 线下教育机构基本情况

各省市的各个年级的复学时间和考试变化时间如图 28 所示。可以看出，各地都是采取分阶段开学措施，毕业年级最先复学。受疫情影响程度较低的地区，其开学时间更早，如青海，宁夏等地区。而受疫情影响较严重的地区，其开学时间则较晚，如北京，湖北等地区。

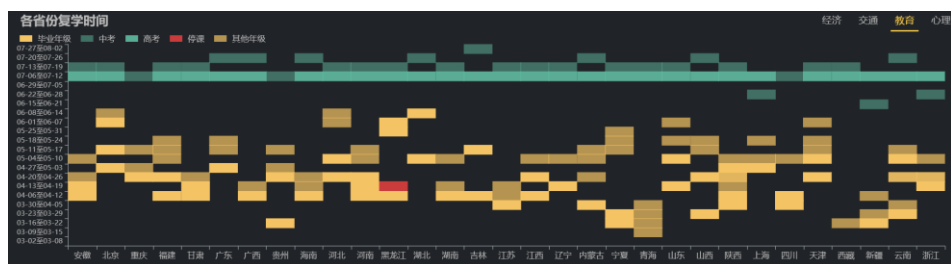


图 28 各省的复学时间和考试变化时间

同时，图中红色矩形块表明黑龙江地区在开学后又重新停课。点击节点，具体查看黑龙江地区的疫情趋势，如图 29 所示。可以发现黑龙江地区的疫情在三月末到四月这段时间内出现反弹现象，从而导致了黑龙江地区的开学后又停课。出现疫情反弹的原因是由于黑龙江地区在疫情得到控制后，对于境外输入的管控松懈导致的。因此，在疫情期间，各个地区应当一直保持严格管控，在加强市区人民出行的同时对于境外人员的输入也要严格检控，防止出现疫情的反弹现象。

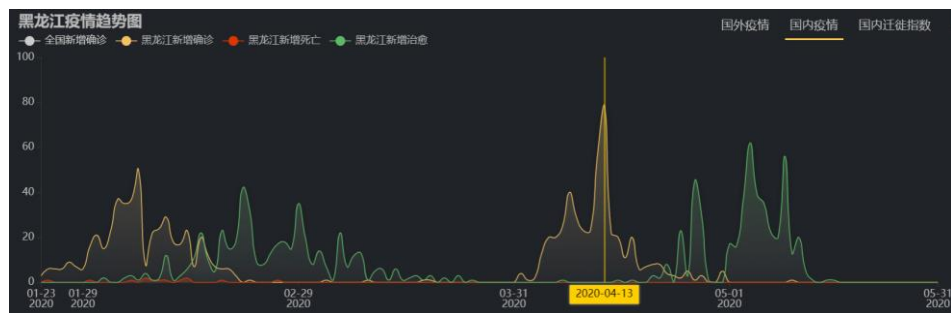


图 29 黑龙江疫情趋势图

全国高考、各省市中考也相比往年推迟一个月左右。经分析，推迟这两场考试的主要原因有三点：（1）疫情虽然得到有效控制，但部分地区仍有偶

发病例的出现，大量学生聚集考试仍然不利于疫情的防控；（2）由于我国目前教育资源分布不均，城乡差距明显。农村及贫困家庭学生无法在疫情期间接受或者只能低质量接受线上教育，这相对于城市学生来说存在客观上的差距。因此中考高考推迟不仅是考试时间的后延，更是促进教育公平的体现；（3）线上教育虽然有所发展，但学校老师的线上教学经验仍相对不足，教学质量低，无法在短期内达到线下教育的效果。

6.4 心理模块场景分析

在心理模块，本系统使用 3D 动态词云、热词图和情绪波动图三个视图展示两类信息。

1. 新闻关注点：3D 动态词云

在 3D 动态词云中，以周为单位，展示了从 12 月 21 日至 5 月 23 日央视新闻官方微博所发布微博的主题，从中能够得到几个关键的时间点（图 30）。在 12 月份，与新冠肺炎相关的内容没有出现，此时关注的内容主要为跨年；在 1 月初，开始出现“不明肺炎”等词汇，说明疫情已经开始产生一定的影响，但此时讨论的热度仍比较低。1 月 18 日之后，新冠肺炎的讨论量迅速增长，几乎所有的高频关键词都与肺炎相关，即使是在春节期间，关注的焦点也是新冠肺炎。2 月初，“口罩”等词汇持续出现，此前新冠肺炎已经被正式定性，开始强调戴口罩。2 月至 3 月期间的词云显示了疫情在我国扩散的特点：主要地区为湖北武汉，并出现大量聚集性传染以及向其他地区扩散的现象。

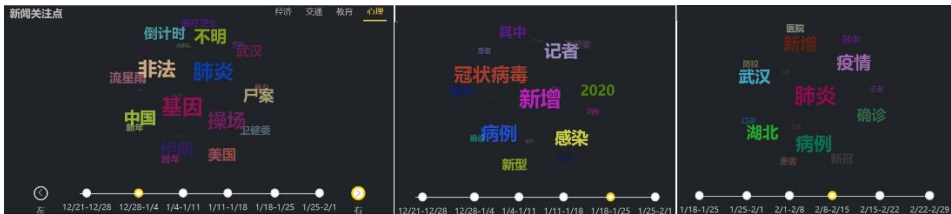


图 30 疫情爆发前（左）、疫情出现初期（中）和疫情最严重时期（右）的关注点

3 月初及以后，外国国家出现的频率逐渐增加，如美国、塞尔维亚、意大利等，结合世界疫情图进行分析，可以看到在 3 月份世界疫情开始出现，符合疫情在国外开始扩散的实际情况。4 月份，我国疫情基本得到控制，此后人们关注点从新冠肺炎转向日常生活。

2. 情感变化趋势：情绪波动图、热词图

通过将情绪波动图与热词图结合分析获取武汉人民情感变化趋势以及影响他们情绪的原因。

情绪波动图（图 31）展示的数据从 1 月 24 日开始，红色表示正面情绪，黄色代表中性情绪，蓝色代表负面情绪，从整体上看，正面情绪占较大比重，中性情绪次之，负面情绪比重最少，这表明虽然疫情带来了武汉人民较大的影响

和损失，但是人们的整体想法仍比较积极。为展示出情绪的变化以及原因，本文选择几个较为突出的时间点进行分析。

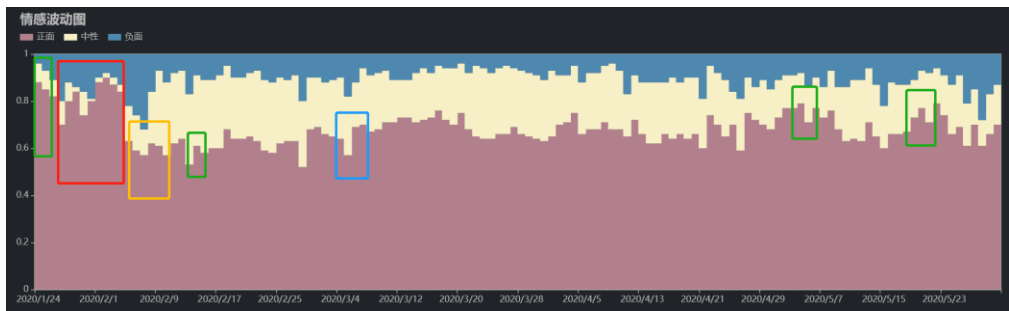


图 31 心理变化趋势

在图中的绿色框线内，从左到右分别是 2.24 日左右、2.14 日左右、3.8 日左右、4.4 日左右以及 5.20 左右，从图中可以看出，在这几日期间，正面情感的增长有较大幅度，而这几日都是节假日期间，这说明节日对人们的情绪调节作用有较强的积极作用，而这一方面是因为节假日固有的让人愉快的特点，另一方面也是由于在这期间网上会举行各种较为积极的活动，也会有传播正能量的节目播放，说明多推广积极正面的活动是调节民众情绪的有效途径。

在第一个红框线内，可以看见积极情绪达到最高点，通过与热词图进行关联分析，能够得到这段期间是火神山与雷神山交付时间；在第一个红框之后，正面情感占比出现一个较大幅度的降低，通过热词图，得到在那个期间内发生了李文良医生事件，群众普遍对这件事情保持较为负面的情绪。

对数据图表进一步分析，在疫情全面爆发之后，相较于疫情之前，人们情绪普遍较低，并在一定值之间波动；在疫情后期，疫情逐渐得到控制，人们的情绪逐渐向正面好转，但是仍低于最初的水平，因此，应重视疫情对人民心理创伤的影响，避免疫情对群众的影响大面积转换为创伤后应激障碍，需要采取一定的措施帮助人们恢复良好的心理状态，可以通过广泛举行积极正面的活动的举措，这种方式的积极作用较为明显。

七、 讨论与总结

7.1 分析与评价

本项目构建了一个完整的可视分析系统，通过关联多源数据，从经济、交通、教育、人民心理四个方面着手，帮助用户直观高效地理解新冠疫情对我国产生的影响，从而防范心理创伤等各项次生灾害的发生。

其中，经济方面，多项指标疫情期间都呈现了负增长的态势，具体表现为国家 GDP 在第一季度出现了 20 年来的第一次负增长，疫情缓和后经济开始逐渐复苏。而代表性物资在疫情初期也出现了供不应求的状态。交通方面，国内的客运量和货运量大幅度下降；各个省市的快递业务均有所下滑。而武汉封城前

期的重点迁出地区在疫情爆发初期均成为了重灾区。教育方面，各省市的复学时间及中考高考时间都有所延迟，且黑龙江地区在第一次开学后由于境外输入影响又重新停课；各类教育 APP 在疫情期间的下载量飙升；线上教育机构在疫情期间得到了发展，但是线下教育机构却受到了严重的冲击。人民心理方面，疫情爆发后，武汉人民的负面情绪相对增多，且大量关注疫情相关信息，基于此，本项目提出了以下几点建议：

①在经济方面，恢复经济发展要循序渐进，采取相关政策时要充分结合疫情态势不能盲目，在保证生活有序的前提下才能更好地进行经济发展；而且对于应急性物资要有提前的储备。

②在交通方面，对于这类传播性极强的病毒，在疫情期间要尽可能的管制群众出行，以便对疫情进行有效的防控。对于迁出指向较高的的地区防控措施应更为严格，此外，在加强市区人民出行的同时对于境外人员的输入也要严格检控，防止出现疫情的反弹现象。

③在教育方面，各个学校可以在正常时期适当开展线上教育，以便在特殊时期师生都能很好地适应从线下教育到线上教育地转换，但进行线上教育前应首先考虑保证教学环境与教学质量；同时政府应当采取合适的措施如减租减税、增加补贴等来保证线下机构的合理运营。

④在民众心理方面，一方面对于疫情情况应该保证公开透明，另一方面，积极的舆论对于民众心理有较好的引导。此外，指定合理的防控措施是保证民众心理有安全感的根本，可以通过举办一些积极的活动来帮助人们恢复良好的心理状态，避免疫情对群众的影响大面积转换为创伤后应激障碍。

7.2 总结与展望

本可视化系统清晰全面地展示了新冠疫情对国民经济、企业生产、教育培训及民众心理等方面产生的影响，分析了产生这些影响的原因，并根据这些原因为用户提供相对有效的预防疫情次生灾害的策略。

但是，本系统采用了较多整体数据，相对而言，对各个方面的分析不够深入。在未来的工作中，将搜集更全面更具体的数据，挖掘更深层次信息，对疫情产生的潜在影响进行更全面综合的评估，为防控疫情次生灾害的防控提供更具价值的参考。