基于 SEIDR 的疫情数据分析及精准防控研究

摘要

从新冠疫情最初的爆发，到如今的疫情常态化，全国各地陆续出现了不同程度的新冠病毒感染情况。疫情数据是反映各地疫情形势的重要信息来源，主要包括人员信息、场所信息、个人自查上报信息、场所扫码信息、核酸采样检测信息、疫苗接种信息等。通过分析这些数据，能协助我们有效实行疫情精准防控，有利于进行人员分类管理、传播途径追踪、疫情追溯、控制疫情蔓延等。因此本文围绕疫情数据，运用 MySQL 数据库实现人员分类管理，并结合 SIR 等传染病传播模型,实现对接种疫苗对病毒传播指数影响的分析。同时，尝试利用 DBSCAN 等聚类方法以及尝试调整损失函数对不同风险场所进行分类运用 DBSCAN 和进行流行病学特征分析，结合格兰杰因果检验，对疫情的精准防控进行进一步优化。主要研究内容和成果包括以下几部分：

一、人员分类管理：首先基于赛题附件所给的表格，设计 E-R 图，创建数据库；然后从时间和空间两个维度，追踪阳性个体的出行轨迹，以识别可能存在的时空伴随者，后依照国务院发布的《新型冠状病毒肺炎防控方案（第九版）》（下称防控九版）判断密接者；其次，得到密接者信息后，以防控九版中次密接者识别的方法，可进一步识别次密接者。

二、接种疫苗对病毒传播影响分析：首先依据核酸采样检测信息表和疫苗接种信息表等得出接种过疫苗的人中核酸阳性的概率及每日累计确诊人数；其次，采用优化算法，基于总体人口和接种疫苗者的感染数据，对 SIR 模型中的传染率进行参数判别，得出传染率和β 恢复率 γ；然后引入 SEIR 模型和 SEIDR 模型，并考虑接种疫苗因素修改原有数学模型，分别得出无疫苗情况和接种不同针数疫苗的情况下易感者、暴露者/潜伏者、感染者、死亡者和康复者人数随时间的变化趋势，基于已有数据建立仿真建模；最后计算各个时间基本再生数，对比未接种疫苗和接种疫苗情况下基本再生数R_0的大小和变化趋势，以此量化分析接种疫苗对病毒传播的影响。

三、重点管控场所的分析确定：首先，通过确定各场所人流量及阳性人员流动量，建立特征；其次运用 DBSCAN 和损失函数进行聚类分析，进而得到各场所病例的分布，划分高中低风险区，为地方政府科学地划分疫情风险等级，明确分级分类的防控策略提供科学依据；最后考虑时间序列分析，使用格兰杰因果检验，对中高风险区进行两两因果检验，据此分析两个区域疫情爆发有无因果关系，分析区域疫情的空间传播性。

四、实现精准防控：为实现精准防控和人员管理，除了附件中的疫情数据，还需要采集一个城市人员的流入和流出信息；通过采集到的数据，计算扩散比（DR）和人口流动比率（MR）；构建目标城市各点坐标的邻接矩阵，采用

Moran 系数和 Geary 系数进行全域空间相关性检验，并进行局部自相关性检验，以此弥补整体检验对局部特征的敏感程度；最后利用加权法可以算出某一地点在某特定时间段传播风险指数R_s。

关键词：疫情防控、精准防控、SEIDR 模型、DBSCAN 聚类、格兰杰因果检验，局域空间相关性检验

目录

一、 绪论

1.1 背景与意义

新冠疫情是指由一种名为 SARS-CoV-2 的病毒引起的全球性传染病爆发，也被称为 COVID-19。这种病毒最初在中国湖北省武汉市被发现，并在短时间内迅速传播到全球各地。由于其高度传染性和致命性，新冠病毒很快成为全球卫生紧急状态。世界卫生组织于 2020 年 3 月 11 日宣布全球进入新冠病毒大流行状态。

病毒的传播速度非常快，通过飞沫传播，感染者可在无症状情况下传播病毒，从而使得防控工作变得更加困难。为了控制疫情的蔓延，各国政府采取了各种措施，包括全面隔离、封锁城市、关闭学校和商业机构、限制公共交通和旅行等。

随着科技的不断进步，大数据分析技术越来越成为疫情精准防控的重要工具。大数据分析能够收集和处理大量的数据，为防控措施的制定和实施提供有力支持。

不断适应新冠病毒变异新特点和疫情防控新形势，科学精准优化细化防控举措，是打好这场攻坚战的必答题。各地贯彻落实习近平总书记 11 月 10 日主持召开的中共中央政治局常委会会议精神，按照第九版疫情防控方案和二十条优化措施要求，就核酸检测、风险区域管控等出台相应举措，力求提升科学精准防控水平。国务院联防联控机制综合组 11 月 21 日公布了《新冠肺炎疫情防控核酸检测实施办法》等 4 个文件，明确提出，在流行病学调查基础上，根据疫情发生地区人口规模大小、感染来源是否明确、是否存在社区传播风险及传播链是否清晰等因素综合研判，根据风险大小，按照分级分类的原则，确定检测人群的范围、频次和先后顺序。

响应提升科学精准防控水平的新趋势，在疫情爆发期间，，大数据分析技术被广泛应用于病例追踪和防控策略的制定。通过分析患者的行程轨迹、病情数据、社交媒体信息等，大数据分析可以帮助疾控部门迅速定位阳性病例和密切接触者，并实施有针对性的隔离和检测措施，从而有效地控制疫情的扩散。此外，大数据分析还可以为疫情预测和预警提供风险量化评估，以及细化优化精准政策的科学

支持。通过对历史数据和当前趋势的分析，大数据分析可以预测疫情的发展趋势，并及时向政府和公众发出警报，提醒大家采取必要的防护措施。除此之外，大数据分析还可以在疫情后期进行复盘和总结，帮助政府和疾控部门更好地总结经验和教训，为未来疫情防控提供参考。

总的来说，大数据分析技术为疫情精准防控提供了高效快捷的手段，它在在疫情防控中发挥越来越重要的作用，为保障公共卫生和人民健康作出很大的贡献。

1.2 工作及思路

疫情数据是反映各地疫情形势的重要信息来源，通过分析这些数据，能协助我们有效实行疫情精准防控，有利于进行人员分类管理、传播途径追踪、疫情研判、控制疫情蔓延等。

在大数据环境下，本文围绕疫情数据，运用 MySQL 数据库实现人员分类管理，并结合 SIR 等传染病传播模型，计算基本再生数，实现对接种疫苗对病毒传播指数影响的分析。同时，运用 DBSCAN 和损失函数进行聚类分析，结合格兰杰因果检验，对场所进行分类，筛选出需要进行重点管控的区域。最后，结合人员流入流出信息，实现全局相关性检验和局部自检验，并计算传播风险指数，更有效更精准地实现疫情防控和人员管理。

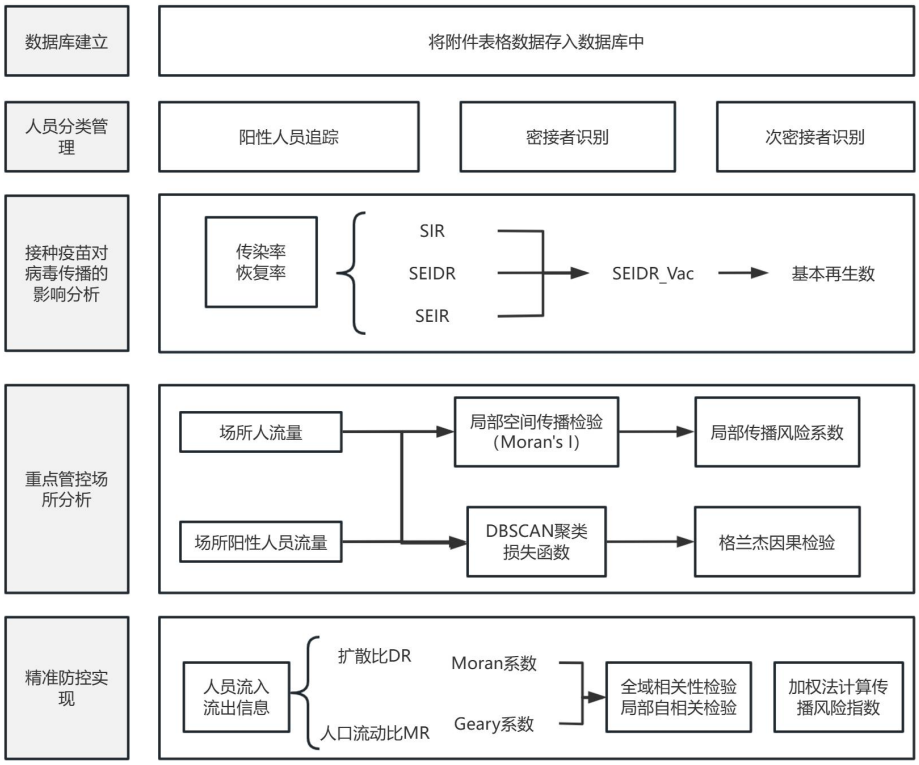


图 1

二、 数据库的建立

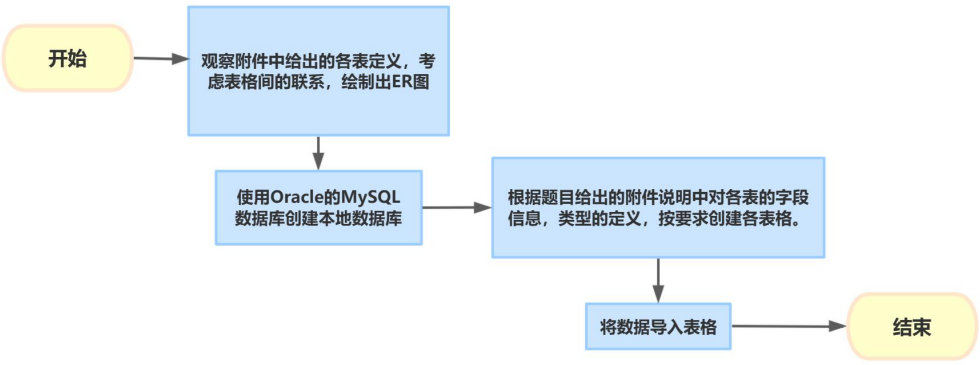


图 2 数据库建立流程

对于数据库的建立，我们第一步需要观察附件给出的各表定义，考虑表格间的联系，并绘制出 E-R 图，如下图所示：

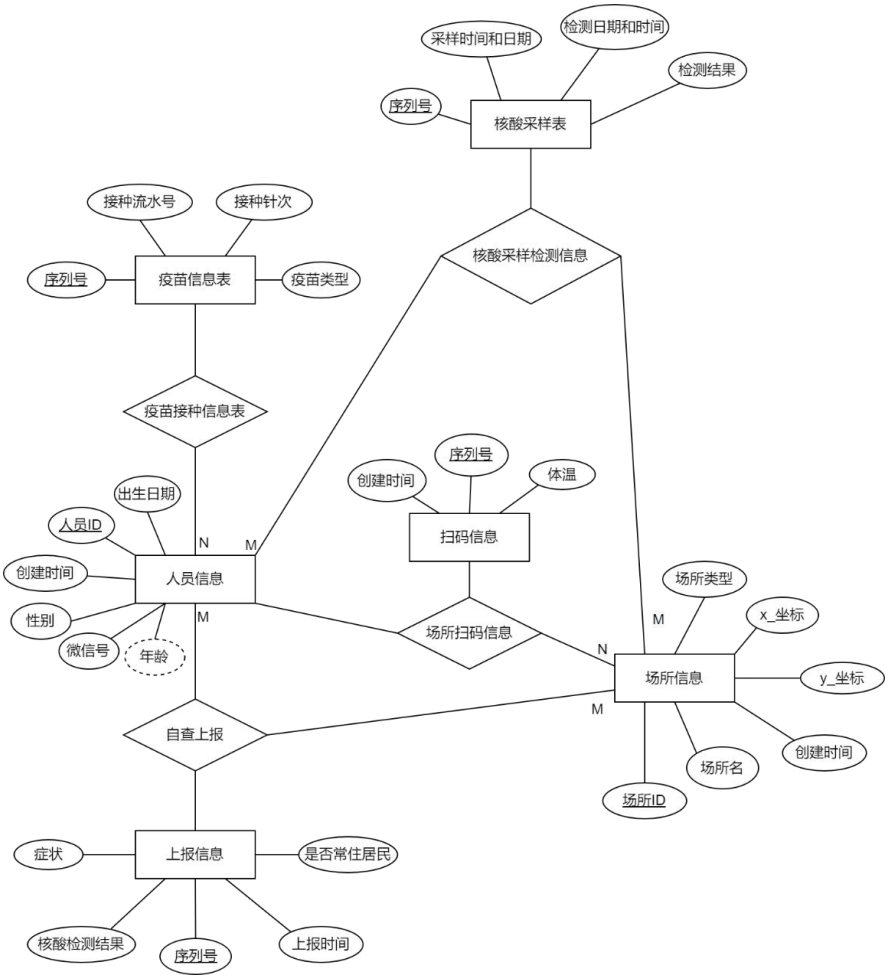


图 3 E-R 模型

第二步，我们需要创建本地的数据库，在本次项目中我们选择使用 Oracle 的 MySQL 数据库进行数据库的创建。

第三步，我们需要根据题目给出的附件对各表的字段信息及类型的定义，按要求创建各表格。在创建表格的过程中应注意表格间的依赖关系对表格创建先后顺序的影响。在本题给出的数据中，人员信息表和场所信息表应该优先建立。换言之，先创建其余表格将导致表格间依赖关系缺失，例如核酸采样检测信息表中的人员 ID 依赖于人员信息表的主码。完成人员信息表和场所信息表两张表的建立后，其他表格的创建顺序可随意。其次，尽管这些表格不符合数据库范式规定，我们依然选择直接创建表格并导入数据，因为这样做可以方便我们后续调取数据

第四步，我们将数据导入表格。

三、 针对问题一的解决方案

3.1 问题分析及思路

本文的主要任务是基于核酸检测结果，追踪阳性个体的出行轨迹，以识别可能存在的密接者。考虑到出行轨迹的两个关键维度，即时间和空间，我们的追踪方法也分别从这两个维度出发。通过核酸检测信息表中的数据，我们能够获取阳性个体的 user_id（人员 ID）、grid_point_id（场所标识码）以及 cysj（采样时间和日期）。再结合场所码扫码信息表，我们可以推测出可能存在的密接者。

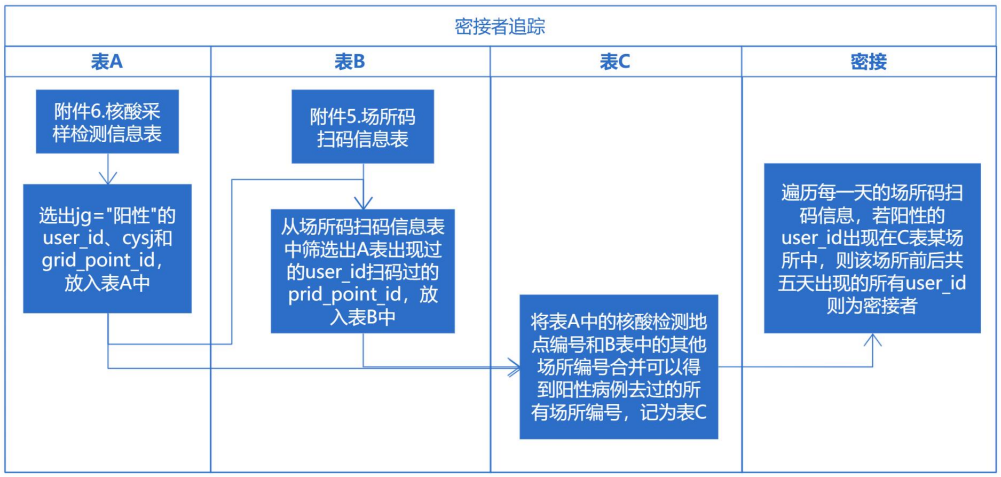


图 4

3.2 密接者追踪

3.2.1 阳性人员筛选

我们依据核酸采样检测信息表（附件 6）中的 jg（检测结果）属性，筛选出检测结果为阳性的个体。我们将 jg="阳性"的 user_id（人员 ID）、cysj（采样日期和时间）以及 grid_point_id（场所 ID）存入表 A，形成阳性个体信息表。同时我们也按照 10 天为一个周期绘制了阳性人员分布散点图，以下展示的是 11 月 1 号到 11 月 10 号的阳性人员分布散点图：

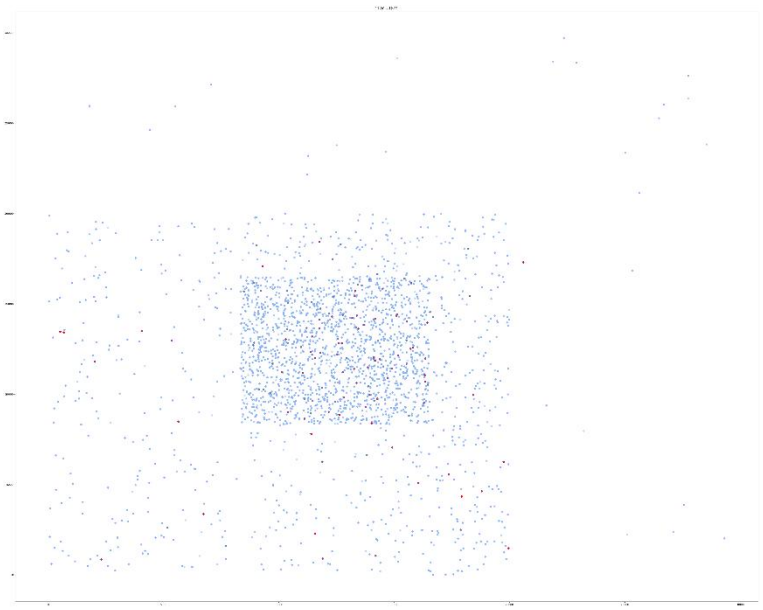


图 5

3.2.2 基于时间及场所的密接者查找筛选

密接者的定义源自《新型冠状病毒肺炎防控方案》（第九版）：在疑似病例和确诊病例症状出现前 2 天开始，或无症状感染者标本采样前 2 天开始，与其有近距离接触但未采取有效防护的人员被视为密接者。对于通过多次核酸检测方式（如高风险职业人群的定期核酸检测）发现的病例，其密切接触者的判定时限为从最后一次核酸检测阴性采样当天起至隔离管控前。因此，我们将与阳性感染者在同一场所出现在某一时间点前后五天内的个体，定义为时空伴随者。

首先，我们需要从空间上确定阳性个体涉及的场所。通过对核酸采样检测信息表的筛选，我们已获得一部分阳性个体去过的场所信息，即阳性个体去过的核酸检测点。然而，这些场所信息并不全面，因此，我们需结合场所码扫码信息表，获取阳性个体去过的其他场地信息，并将其存入表 B。最后，将表 A 和表 B 合并，得到阳性个体去过的所有场地信息，存入表 C。

接下来，我们需要从时间角度确定潜在密接者。根据疾病防控方案中对密接者的定义，我们将与阳性感染者在某一时间点的前后五天内，在同一场所出现的人判定为密接者。因此，我们需要确定阳性感染者的核酸检测时间和场所信息，并筛选出时间和地点与阳性感染者相符的人。为此，我们将数据按每 5 天一段进

行筛选，以涵盖前后两天共五天的时间范围。最终，通过以上筛选步骤，我们可以得出所有密接者的 user_id。

对于得到的密接者信息表，我们截取了部分数据进行展示，如下图所示：

密接日期	密接者ID	密接场所	阳性人员ID
2022/10/3 20:23	12675	公交车41	11733
2022/10/7 15:30	98234	地铁站10	6028
2022/10/9 0:53	47101	居民小区13	6028
2022/10/18 16:39	85414	居民小区8	85124
2022/11/1 11:58	98468	医院21	53472
2022/11/4 19:09	49074	居民小区11	27962
2022/11/11 9:16	62251	居民校区30	69764
2022/11/18 14:20	46317	公交车27	61537

图 6

对于密接者，我们也绘制了散点图以直观地展现它的分布，下图所示的是 11 月 1 号到 11 月 10 号的密接者分布散点图：

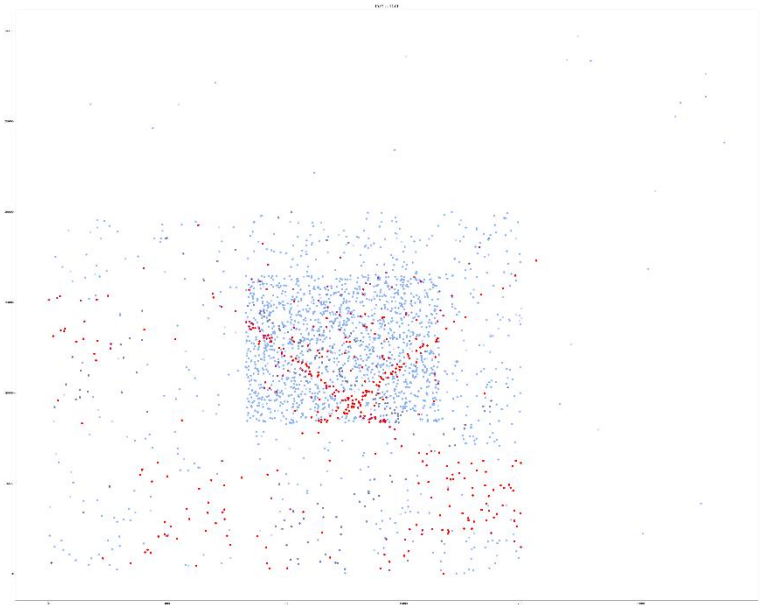


图 7

四、 针对问题二的解决方案

4.1 问题分析及思路

题目主要是要求我们依据问题 1 的结果，通过密接者的出行时间与场所追踪次密接者。对于这个问题我们跟第一问的思路基本一致，主要还是从时间和空间两个维度进行考虑。

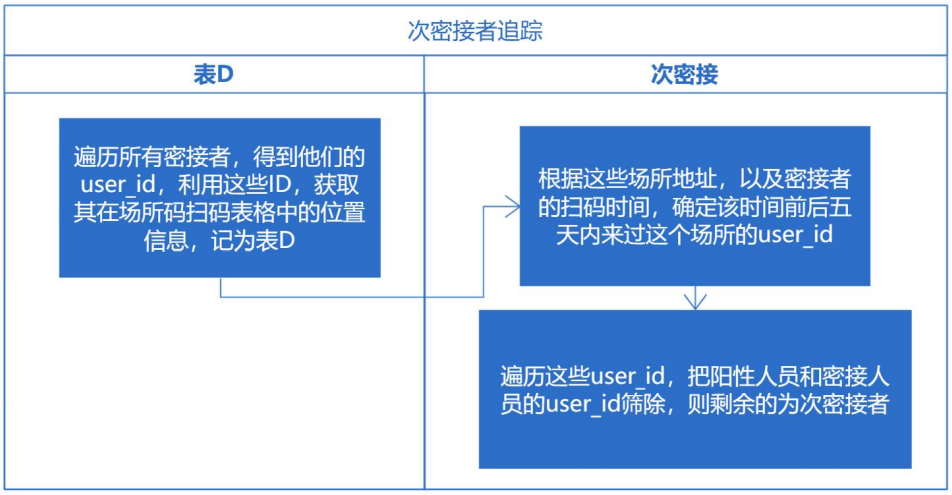


图 8

4.2 次密接者追踪

4.2.1 次密接者场所信息

通过问题 1 所得出的密接者结果，我们可以得到密接者的 user_id，通过这些 ID，我们可以得到密接者去过的场所位置信息，记为表 D。

4.2.2 次密接者时间信息

根据这些场所的 ID，以及密接者的核酸检测或扫场所码的信息，我们可以筛选出这些场所前后各两天共五天内所有来过的人的 ID。

4.2.3 次密接者筛选

将以上得出的所有去过相关场所的 user_id，除去阳性人员和密接者的 user_id，我们就可以得出次密接者的 user_id。

对于得到的次密接者信息表，我们截取了部分数据进行展示，如下图所示：

次密接者ID	次密接日期	次密接场所ID	密接者ID
54513	2022/10/1 15:30	居民小区38	10
56854	2022/10/6 2:42	居民小区38	10019
87351	2022/10/12 16:26	影剧院40	10020
71690	2022/10/21 15:45	公交车7	10099
93690	2022/10/31 14:56	公共浴室46	10147
68761	2022/11/3 2:55	公交车34	10126
99411	2022/11/12 13:02	游泳场馆46	10161
92933	2022/11/21 9:13	地铁站21	10371

图 9

五、 针对问题三的解决方案

5.1 问题分析及思路

接种疫苗可以让人体免疫系统产生对特定病原体的防御能力，以便在接触到这些病原体时，可以迅速识别并攻击它们。即使接触到病毒，接种疫苗的人也不容易被感染。这就减少了病毒在人群中的传播。同时，接种疫苗也可以减轻症状，具体表现为重症率，死亡率的显著下降，治愈率的提高，从而降低病毒在人群中传播的风险。所以综上所述，接种疫苗会对病毒传播指数产生一定的影响。

为充分量化分析接种疫苗对病毒传播指数的影响，本文需要通过附件中所给出的核酸采样检测信息表和疫苗接种信息表等得出接种过疫苗的人中核酸阳性的概率及每日累计确诊人数，并结合优化版的 SIR 模型，得出接种不同针数疫苗的人群和未接种疫苗的人群随着时间变化的传染率等描述疾病传播的量化数据。

为解决以上问题，本章首先采用优化算法，对 SIR 模型中的传染率进行参数估计，得出传染率β和恢复率γ。然后我们引入 SEIR 模型和 SEIDR 模型，分别得出无疫苗情况和接种不同针数疫苗的情况下易感者、暴露者/潜伏者、感染者、死亡者和康复者人数随时间的变化趋势。同时计算基本再生数（Basic Reproduction Number），对比未接种疫苗和接种疫苗情况下基本感染数的大小和时间变化趋势。

5.2 贝叶斯——疫苗影响模型

针对接种疫苗对病毒传播指数影响的分析，我们最初采用了贝叶斯概率模型。

在这里我们假设居民 60 日内出现核酸阳性为事件 A，居民接种疫苗为事件 B。将附件的数据进行统计计算我们得出$p(A) = 0.01$，$p(B) = 0.8$，$p(B|A) = 0.5$。

通过以上数据我们得出：

在接种疫苗的前提下居民感染病毒的概率为

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)} = \frac{0.5 * 0.01}{0.8} = 0.00625$$

在未接种疫苗的前提下居民感染病毒的概率为

$$p(A|\bar{B}) = \frac{p(\bar{B}|A)p(A)}{p(\bar{B})} = \frac{0.5*0.01}{0.2} = 0.025。$$

通过比较以上两种情况的概率，我们可以看出接种疫苗对病毒有一定的遏制作用，能降低感染率，同时也能降低传播风险。但贝叶斯模型也有较多缺点，如下所示：

• 贝叶斯模型中是基于特征条件独立假设进行分类，因此当数据集的特征存在关联时，分类效果不佳。

• 需要事先假设特征的先验分布，如果假设与真实情况不太相符，那么模型效果也会受到影响。类别的先验分布一般基于训练数据进行计算。当训练数据没有代表性，不能表征真实数据的情况时，会产生较多误分类，同时对拥有时间序列的数据处理效果不好。

基于以上贝叶斯模型的缺点，我们采用了传染病模型对疫苗影响进行量化，以此来优化接种疫苗对病毒传播影响的分析，具体流程将在下文中阐述。

5.3 SIR 模型理论

本章对 SIR 模型进行了进一步优化。SIR 模型是常见的一种描述传染病传播的数学模型，其基本假设是将人群分为易感人群（Susceptible，指未得病者，但缺乏免疫能力，与染病者接触后容易受到感染）、感染人群（Infective，指染上传染病的人，可传播给易感人群）、移除人群（Removed，被移出系统的人，因病

愈或者死亡的人，这部分人不再参与感染和被感染的过程）。其中三类人群的转换关系如下图所示。

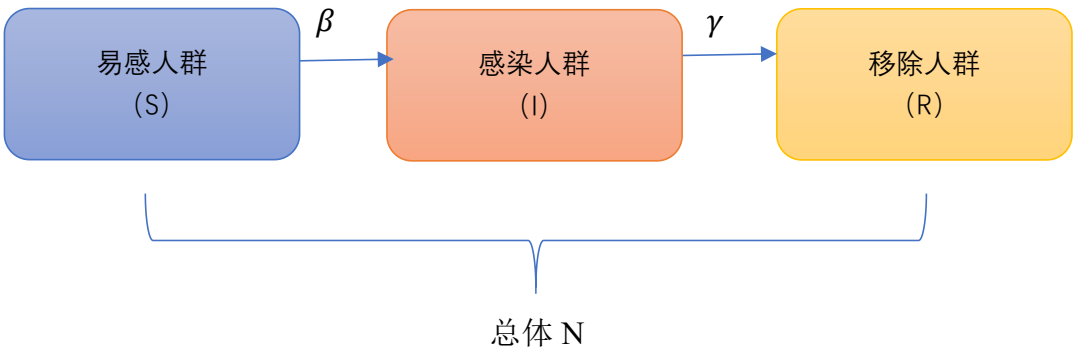


图 10

易感人群与感染人群接触时被感染的概率为传染率β。传染率反映了疾病传播的强度，传染率越大则易感人群和感染人群接触后被传染的可能性越大。假设感染人群以固定平均速率恢复或者死亡，则恢复率γ取决于感染的平均持续时间。此处用三类人群的首字母 S、I、R 分别表示三类人群的数量，N 表示总人数[1]。那么三类人群的数量随时间的动态变化规则可以用以下常微分方程组表示：

$$\begin{cases} \frac{dS}{dt} = -\beta\frac{IS}{N}(1) \\ \frac{dI}{dt} = \beta\frac{IS}{N} - \gamma I(2) \\ \frac{dR}{dt} = \gamma I(3) \end{cases}$$

5.4 采用优化算法对传染率进行参数辨识

通过上面的介绍我们知道 SIR 模型实际上是采用动力学模型（三个常微分方程）对三类人群随时间变化的过程进行建模，采用传染率和恢复率来量化描述疾病传染和疾病被治愈等行为。很重要的一点是只有获得准确的动力学模型参数才有可能建立一个相对精确的模型。所以我们采用 SIR 模型对新型肺炎传播进行建模的主要问题就是确定出以下参数

- 传染率β和恢复率γ

• 易感人群（Susceptible）初值，感染人群（Infective）初值，移除人群（Removed）初值。

对于这些参数的确定，我们首先根据核酸检测信息和疫苗接种信息，得出2022年10月3日到2022年12月1日每天累计确诊人数的信息。同时我们也要确定γ的取值。γ为恢复率，因为新冠肺炎病毒的恢复期大约为14天，因此对于未接种疫苗的人来说，$\gamma = \frac{1}{14} \approx 0.07$。由于接种过疫苗的人抵抗力会强一点，病愈所花费的时间也会更短一点。据调查统计，接种疫苗的人恢复期大约是6天，所以对于接种疫苗的人群，$\gamma = \frac{1}{6} \approx 0.17$。

同时我们再引入一个新的参数$nContact$。$nContact$表示感染者接触的未感染着人数，其会因为政府管控措施、地区密集度等因素而改变。此处使用任务一的数据，即阳性人员人均密接人数作为$nContact$的值。由问题一可知，密接人次为90203，感染人次为1032，密接排查时长为3天。通过计算我们可以得出每日密接者的平均数为$Contact_avg = 90203/(1032 * 3) = 29.14$。但由于我们对于密接者的定义较于疾病传播学的日均接触人数$nContact$有差别，所以我们通过查阅资料，将$nContact$假设为5。

接下来主要是如何得出准确的传染率β。为了方便进行参数计算，我们认为在疾病传播早期有$S \approx N$，因为传播早期患病人数较少，所以可以近似认为所有人都是易感人群，将这个假设代入上述的（2）式中可得

$$\frac{dI}{dt} = (\beta - \gamma)I \qquad (4)$$

易知该微分方程的通解为：

$$I(t) = Ce^{(\beta-\gamma)t} \qquad (5)$$

由$I(t=0) = 1$可得$C = 1$，代入式（5）中可得

$$I(t) = e^{(\beta-\gamma)t} \qquad (6)$$

由此可以构建如下参数辨识问题：

决策变量：传染率β

目标函数：$min \sum_{t \in T} (e^{(\beta-\gamma)t} - \hat{I}(t))^2$

其中$\hat{I}$为实际的患病人数，T为时间集合，以天为单位[2]。这里的$\hat{I}$我们是根

据核酸检测信息和疫苗接种信息得出的，通过计算 2022 年 10 月 3 日到 2022 年 12 月 1 日每天累计确诊人数，我们就可以得到$\hat{I}$随时间变化的值。通过求解上述优化问题即可得到新冠肺炎的传染率$\beta = 0.1959$，并且$Infection_probability = 0.0392$。其中$Infection_probability$为感染概率，它与$\beta$有如下关系式：$\beta = nContact \times Infection\_probability$。

假设总人数$N = 10000$，则通过以上公式和参数可以得到如下的变化趋势图。

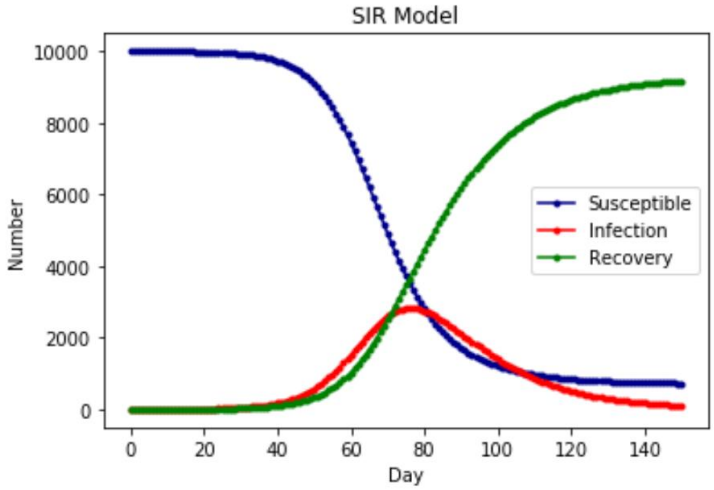


图 10

5.5 未接种疫苗无死亡模型

在未接种疫苗并且不考虑死亡的情况下，我们采用了 SEIR 模型。

$$\begin{cases} dS/dt = -r\beta(I+E)S/N \\ dE/dt = r\beta(I+E)S/N - \alpha E \\ dI/dt = \alpha E - \gamma I \\ dR/dt = \gamma I \end{cases}$$

其中α为接触者向感染者的转化速率，取$\alpha = \frac{1}{7}$，因为潜伏期为 7 天[3]。将上述 SIR 模型中的其他参数代入这个模型，可以得出以下趋势图。

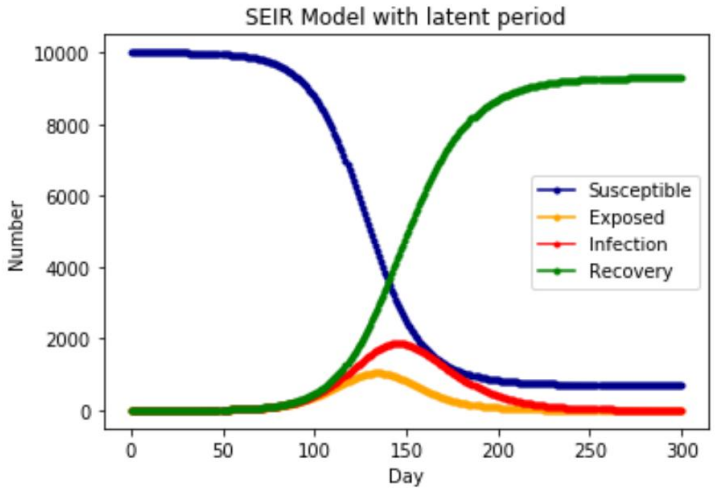
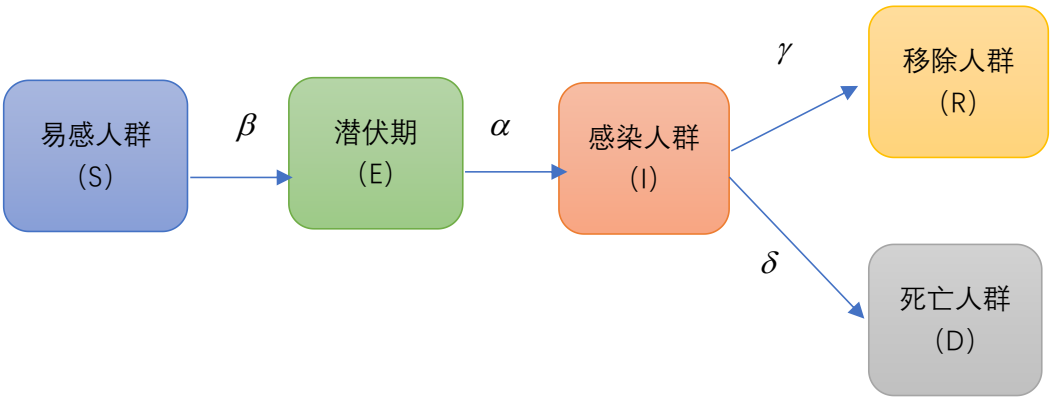


图 11

5.6 未接种疫苗有死亡模型

假设患病者（Infectious）在患病期间有概率康复或者死亡，设定死亡率为δ。则在患病期间，患者可以转化为康复者（Recovered）和死亡者（Dead），则延伸为 SEIDR 模型，其微分公式如下：

$$\begin{cases} dS/dt = -r\beta(I+E)S/N \\ dE/dt = r\beta(I+E)S/N - \alpha E \\ dI/dt = \alpha E - \gamma I - \delta I \\ dD/dt = \delta I \\ dR/dt = \gamma I \end{cases}$$



我们了解到新冠病毒感染者的死亡率大约是 0.005[4]，将其代入上面的方程，得到如下图所示的结果。

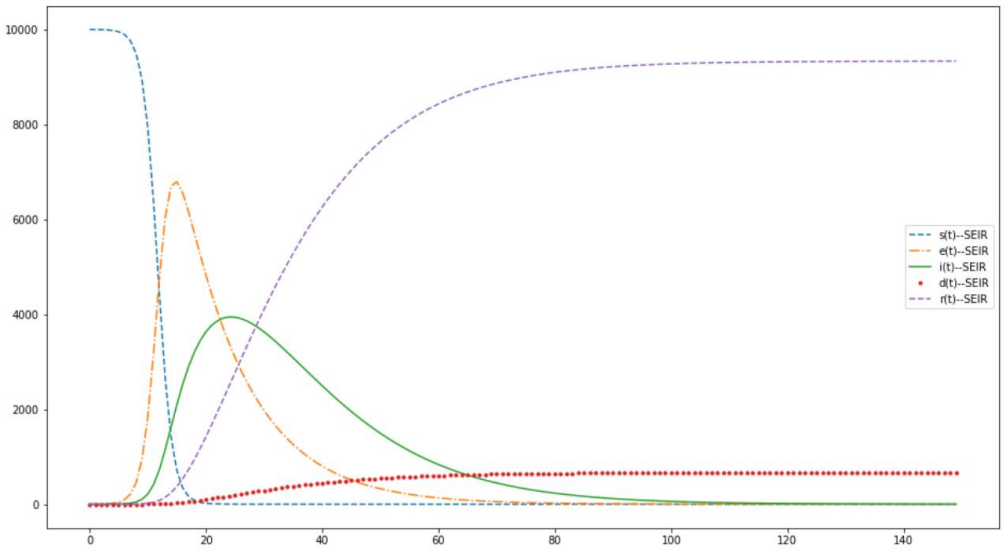


图 12

5.7 接种疫苗模型

接种疫苗可以让接种疫苗的人不容易被感染，减少了病毒在人群中的传播。同时，接种疫苗也可以减轻症状和疾病的严重程度，从而降低病毒在人群中传播的风险。所以，接种疫苗会对病毒传播指数产生一定的影响。引入疫苗的接种影响因素，接种疫苗可以直接反映在 SEIDR 中的感染率、死亡率、康复率中，影响传染病的致死率数值情况。具体的影响可以基本分为以下四种情况进行进一步讨论：

（1）一剂疫苗，在传染病发生时期即存在疫苗，在传染病爆发的同时开始接种疫苗。

（2）一剂疫苗，传染病发生时并无疫苗，疫苗相对于传染病传播具有延迟性，在传染病开始爆发 T 日后开始接种疫苗。

（3）二剂疫苗，在传染病发生时期即存在疫苗，在传染病爆发的同时开始接种第一剂疫苗；在两针疫苗的间隔期后开始接种第二剂疫苗。

（4）二剂疫苗，传染病发生时并无疫苗，疫苗相对于传染病传播具有延迟性，在传染病开始爆发 T 日后开始接种第一剂疫苗；在两针疫苗的间隔期后开始接种第二剂疫苗。

引入这几种情况之后，接种疫苗有死亡模型的基本微分公式如下：

$$\begin{cases} dS/dt = \sum -r\beta_i (I+E)/N + \sum v_j S_i, i \in (0,1,2), j \in (1,2) \\ dE/dt = \sum r\beta_i (I+E)/N + \sum \alpha E_i, i \in (0,1,2) \\ dI/dt = \sum (\alpha E_i - \gamma_i I_i - \delta_i I_i), i \in (0,1,2) \\ dD/dt = \sum \delta_i I_i, i \in (0,1,2) \\ dR/dt = \sum \gamma_i I_i, i \in (0,1,2) \end{cases}$$

最后我们可以得出如下图所示的变化趋势图。

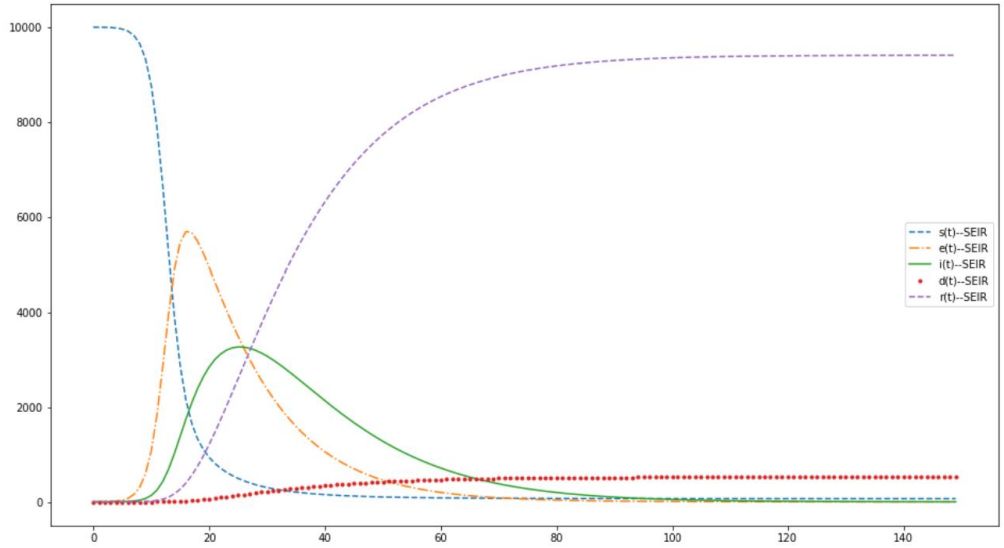


图 13

5.8 接种疫苗和未接种疫苗情况下各参数比较

通过上述的未接种疫苗有死亡模型和接种疫苗模型，我们可以将其中国的感染情况和死亡情况进行对比。其结果如下图所示：

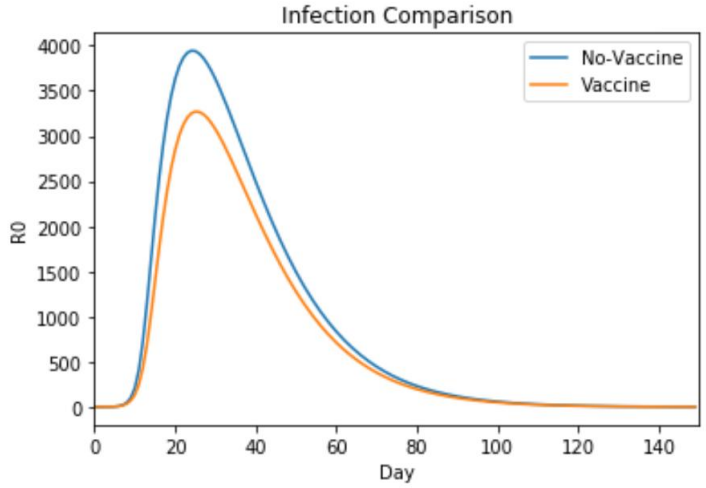


图 14

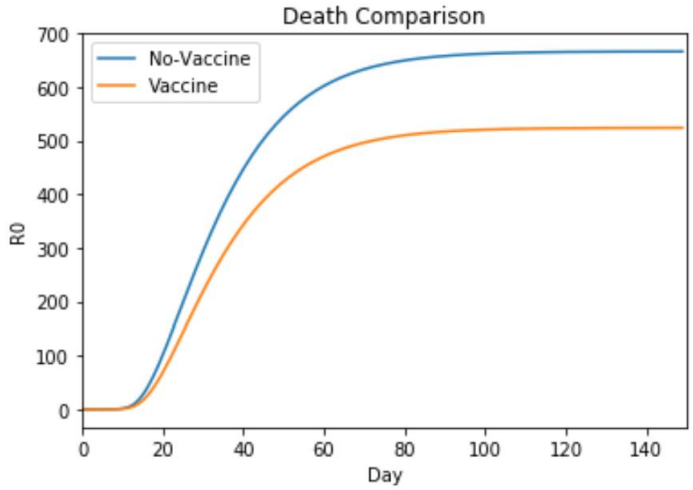


图 15

通过上述图片可以看出，未接种疫苗的情况下，感染和死亡的人数较于接种疫苗的情况更多，感染人数峰值由 3000 到 4000，死亡人数由 500 上升至 700。由此可见，接种疫苗是能有效降低感染率和死亡率的。

接下来，我们引入一个新冠病毒基本传染数R_0。R_0基本传染数是指在无外力介入，同时所有人不具备对该疾病免疫力的情况下，一名病毒携带者，会把疾病传染给其他多少个人的平均数，基本传染数是衡量疾病传染性的一个重要指标。若$R_0 < 1$ 则传染病将会逐渐消失，若$R_0 > 1$ 传染病会以指数方式散布，成为流行病。

在这个问题中，我们分别计算了未接种疫苗的情况下R_0随时间变化的趋势以

及接种疫苗的情况下R_0随时间的变化趋势，有如下的式子：

$$R_0(t) = \frac{\beta S(t)}{N\gamma}$$

得到的趋势图如下图所示：

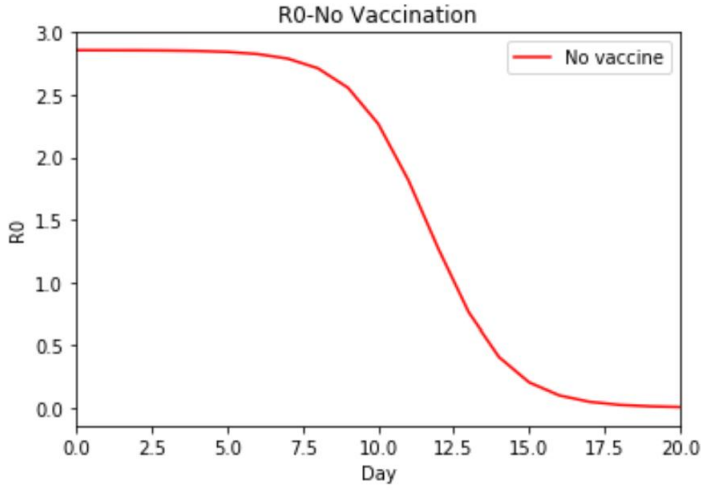


图 16

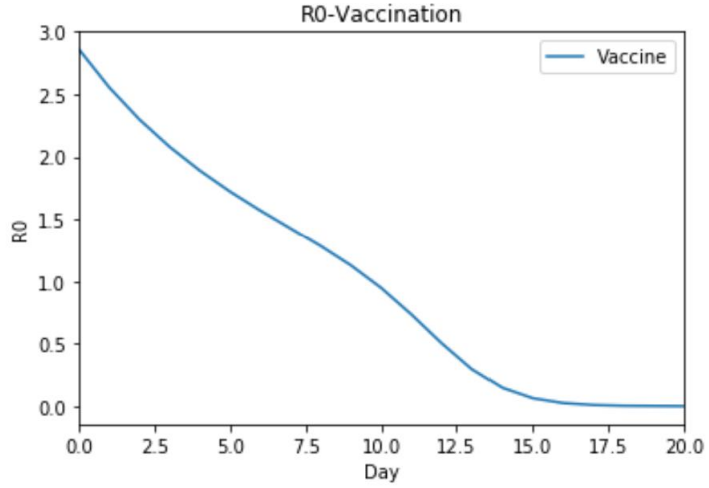


图 17

我们将两种情况下的R_0进行了对比，会发现接种疫苗的情况下R_0下降的趋势更快，并且比未接种疫苗的情况更早降到 1 以下。所以由此可以看出接种疫苗对于抑制疫情的大范围传播有着积极的作用。

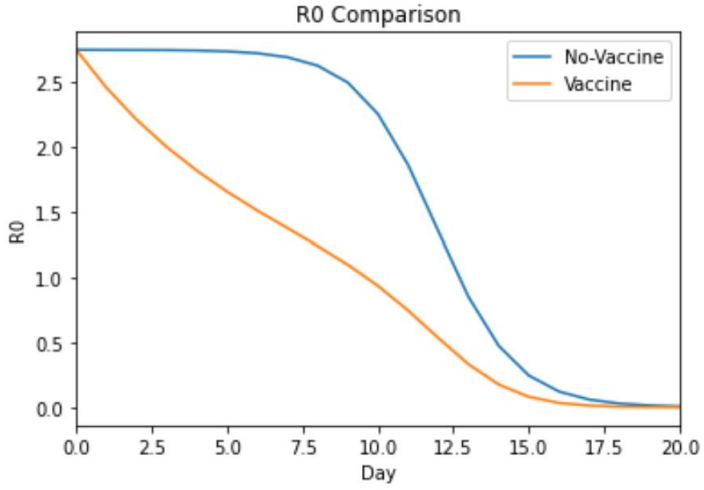


图 18

以上的模型运用的输入数据都是疫情爆发初期数据的理想化数据，接下来我们再将附件所给的数据代入此模型中，验证接种疫苗对于病毒传播指数影响情况。将数据代入后我们得到如下图所示的R_0变化情况，可以看出与上述理想化模型的变化趋势是基本一致的。

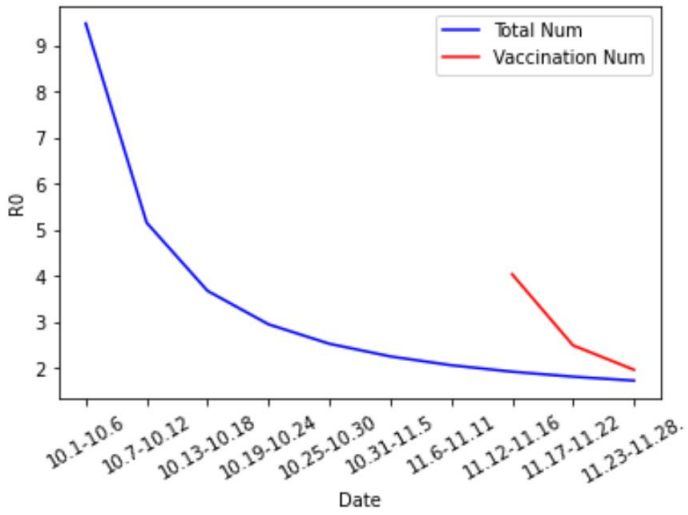


图 19

六、 针对问题四的解决方案

6.1 问题分析及思路

根据阳性人员数量及辐射范围，可以对某场所疫情防控风险进行评估，进而确定需要重点管控的场所。本章通过确定各场所人流量及阳性人员流动量，建立

特征，然后运用 DBSCAN 和损失函数进行聚类分析，进而得到各场所病例的分布，划分高中低风险区，为地方政府科学地划分疫情风险等级，明确分级分类的防控策略提供科学依据。我们将中高风险区纳入需要重点管控的场所。

由于疫情具有传播性，且在不同场所之间的传播有时间滞后性。于是我们考虑时间序列分析，使用格兰杰因果检验，对中高风险区进行两两因果检验，据此分析两个区域疫情爆发有无因果关系，分析区域疫情的空间传播性。

6.2 数据预处理及特征建立

首先我们需要筛选出阳性患者的核酸检测记录。通过附件 6 核酸采样检测信息表，将 jg 属性为阴性和未知的核酸检测记录丢弃，即为阳性患者的核酸检测记录。我们对于某一场所有出现阳性人员的定义为，如果来过此场所的人在来这个场所后的两天内核酸检测结果为阳性，则该场所有阳性人员出现。所以接下来我们遍历每一天的数据集，找出在某天的两天后核酸检测结果为阳性的人员的 ID，并获取目标时间范围点。

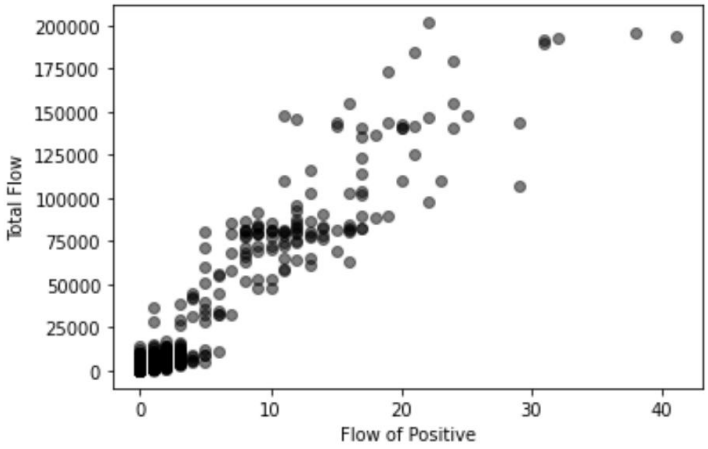
根据以上对数据的处理，我们可以得一个包含场所编号、人流量和阳性人员流动量的表格。其中，我们将特定场所的人数设为特征 2，将特定场所出现阳性人员数量设为特征 3。

6.3 DBSCAN 聚类

基于密度的噪声应用空间聚类(DBSCAN)是一种无监督的 ML 聚类算法。无监督的意思是它不使用预先标记的目标来聚类数据点。聚类是指试图将相似的数据点分组到人工确定的组或簇中。它可以替代 KMeans 和层次聚类等流行的聚类算法。

较于 KMeans 聚类算法，DBSCAN 聚类不要求指定集群的数量，避免了异常值，并且在任意形状和大小的集群中都有较好的效果。它没有质心，聚类簇是通过将相邻的点链接在一起的过程形成的。

通过 DBSCAN 聚类我们得到如下图所示的聚类结果噪声点图。




```
CV-1 score: 0.9137
CV-2 score: 0.9172
CV-3 score: 0.9046
CV-4 score: 0.9185
CV-5 score: 0.9140
```

图 20

由上图可以看出噪音点过多，聚类效果不佳，所以我们选择更换方法，具体做法将在下一小节详细阐述。

6.4 损失函数分类模型

首先我们将上文提及的$feature_2$（场所人数：人次）和$feature_3$（阳性人数：人次）合并成一个数据集。然后我们通过如下损失函数计算每个点的损失函数值：

$$Loss \quad Function = \frac{1}{Flow\_of\_Positive^2 + (0.0002 * Total\_of\_Flow)^2}$$（人次^-2）

我们将损失函数值为最大的 2%的点归为高风险地区，将损失函数值为最大的 4%的点归为中高风险地区。对于中风险地区，我们就将中高风险地区的点除去高风险地区的点，即为中风险地区。分类结果图如下所示：

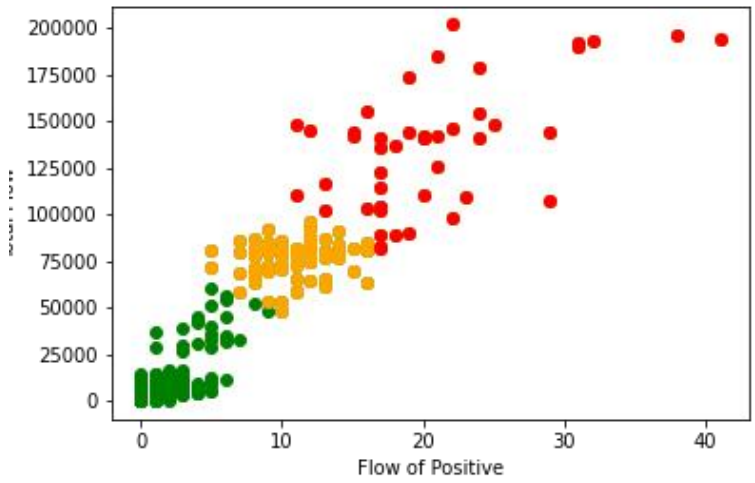


图 21

其中绿色为低风险场所，黄色为中风险场所，红色为高风险场所。得到高风险区有 45 个，中风险区有 73 个。以下选取部分例子展示
高风险区和中风险区的结果。

	grid_point_id	name	point_type	x_coordinate	y_coordinate	create_time	累计阳性人次	流动人口
101	102	地铁站2	交通	15179.91	12137.15	2022/7/5 16:27	17	101861
107	108	居民小区2	住宅	16331.03	11048.13	2021/1/23 21:56	15	141651
166	167	居民小区3	住宅	13308.25	15707.57	2020/7/10 2:35	17	141085
225	226	居民小区4	住宅	15713.85	12514.70	2020/4/27 8:36	11	148307
284	285	居民小区5	住宅	16441.30	13954.42	2020/7/11 14:27	12	145440
337	338	地铁站6	交通	11568.11	11986.94	2021/1/31 17:31	23	109774
454	455	公交车8	交通	14382.78	11931.00	2022/5/3 3:43	11	110036
461	462	居民小区8	住宅	16345.34	10696.84	2020/11/16 23:35	20	140738
514	515	地铁站9	交通	13938.52	9955.34	2021/6/25 6:16	19	90125
632	633	地铁站11	交通	13181.36	14217.89	2021/10/2 2:40	18	88652

图 22

	grid_point_id	name	point_type	x_coordinate	y_coordinate	create_time	累计阳性人次	流动人口
48	49	居民小区1	住宅	14098.83	11960.41	2021/8/9 7:48	9	79936
101	102	地铁站2	交通	15179.91	12137.15	2022/7/5 16:27	17	101861
107	108	居民小区2	住宅	16331.03	11048.13	2021/1/23 21:56	15	141651
166	167	居民小区3	住宅	13308.25	15707.57	2020/7/10 2:35	17	141085
218	219	公交车4	交通	11808.33	14733.07	2021/12/24 3:26	11	65512
...	...	...	...	...	...	...	...	...
2815	2816	地铁站48	交通	9044.77	18237.79	2020/6/3 11:06	8	63119
2821	2822	居民小区48	住宅	11767.34	18431.86	2021/11/30 8:52	24	141053
2874	2875	地铁站49	交通	12332.49	17478.13	2021/10/23 20:07	16	63676
2880	2881	居民小区49	住宅	9290.16	17076.51	2022/5/1 2:34	20	140608
2939	2940	居民小区50	住宅	20603.80	17298.19	2021/8/6 2:52	21	184953

图 23

6.5 基于 Local Moran's I 空间自关性的传播系数量化

对于城市不同场所间的人员流动，我们在考虑其相关性时，构建目标城市中各点坐标的邻接矩阵，可以采用 Moran 系数和 Geary 系数进行全域空间相关性检验 [REF: 刘 勇，杨东阳，董冠鹏，等：南省新冠肺炎疫情时空扩散特征与人口流动风险评估]

$$Moran'sI = \frac{n\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i-\bar{x})(x_j-\bar{x})}{(\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij})\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

$$(i \neq j)$$

式中：n表示研究的目标城市地标个数；ω_{ij} 为空间权重矩阵。由于我们的数据规模是城市规模，属于小范围内的疫情传播分析，因此空间权重矩阵中的数值，我们采用两点的地理坐标 i, j 的距离的倒数；为各标点的累计阳性流量；$\bar{x}$ 为所有标点的阳性流量均值。 Moran 系数的范围应该在[-1, 1]。当系数小于 0 时认为空间范围内的数据具有较为显著的负自相关性，为正则代表具有空间正自相关。

全域空间相关性检验反映出目标城市整体空间分布情况，难免会损失部分局部特征等信息，为了提高模型准确度，我们需要对部分地标进行局部自相关检验，以此弥补整体检验对局部特征的敏感程度。

$$Local\ Moran'sI = \frac{n(x_i - \bar{x})\sum_{j=1}^{m} w_{ij}(x_j - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2},\ (i \neq j)$$

此处我们对中高风险区进行分析。设定邻域半径为 1000，以数据中的 x,y 坐标作为空间权重矩阵的输入，得出相应的 Moran 系数。

随后使用激活函数 Sigmoid 函数转换成局部传播风险系数

$$S(x) = \frac{1}{1 + e^{-x}}$$

风险系数接近 0.5，说明该场所传播情况与邻域的中高风险场所传播情况的自关性较低；风险系数接近 0，说明该场所传播情况与邻域的中高风险场所传播情况呈较高负自关性；风险系数接近 1，说明该场所传播情况与邻域的中高风险场所传播情况呈较高正自关性。下图展示的是五个中高风险地区风险系数：

	grid_point_id	x_coordinate	y_coordinate	累计阳性人次	风险系数
0	49	14098.83	11960.41	9	0.499607
1	102	15179.91	12137.15	17	0.499999
2	108	16331.03	11048.13	15	0.499999
3	167	13308.25	15707.57	17	0.499999
4	219	11808.33	14733.07	11	0.499997

图 24

6.6 基于时间序列的不同场所疫情爆发因果检验模型

格兰杰检验方法是基于时间序列的自回归模型（AR 模型），通过比较包含和不包含前一时间点其他序列信息的模型残差方差大小，来判断序列之间是否存在因果关系。若在包含了变量 X、Y 的过去信息的条件下，对变量 Y 的预测效果要优于只单独由 Y 的过去信息对 Y 进行的预测效果，即变量 X 有助于解释变量 Y 的将来变化，则认为变量 X 是引致变量 Y 的格兰杰原因。

在疫情爆发的场所中，格兰杰因果检验可以用于分析不同场所之间疫情爆发的时间相关性。假设我们有多个场所的疫情数据，可以将其转化为时间序列，通过格兰杰因果检验来分析场所之间是否存在因果关系。例如，如果我们发现在场

所 A 疫情爆发前，场所 B 的疫情数据对其具有显著的因果影响，那么我们可以认为场所 B 是疫情爆发的源头，并及时采取措施遏制疫情的扩散。

在此我们将中高风险场所两两进行格兰杰因果检验，运用 grangercausalitytests(data, maxlag=3)函数计算格兰杰因果检验，这里我们设置最大滞后期为 3。滞后期即为回归问题多项式的阶数，它的作用是指明采用过去多少个点对当前点的点求解回归问题。

在格兰杰因果检验中，我们通常会计算出一个 F 值和一个 p 值。F 值表示检验统计量的值，p 值则表示检验的显著性水平。如果 p 值小于显著性水平，我们就可以认为价格序列对另一个序列有显著的因果影响。这里的显著性水平我们设置为 ，表示当没有实际差异是得出存在差异，会有 5%的风险。也就是说我们有 95%的把握认为是正确的。

我们使用累计阳性人数作为拟合数据，遍历所有中高风险场所，得到有因果性的场所。但是由于我们使用的拟合数据是累计的数据，所以可能会导致有些场所数据在某段时间波动较小。然而波动较小的数据可能会造成过拟合，这些过拟合的置信水平（$p-value$）一般都远小于 0.001。所以我们现在将 $0.001 \leq p-value \leq 0.05$ 的场所认为是具有强相关性的。

我们从风险区选择了 10 个区域进行相关性可视化，颜色越浅代表格兰杰相关性越高，如下图所示：

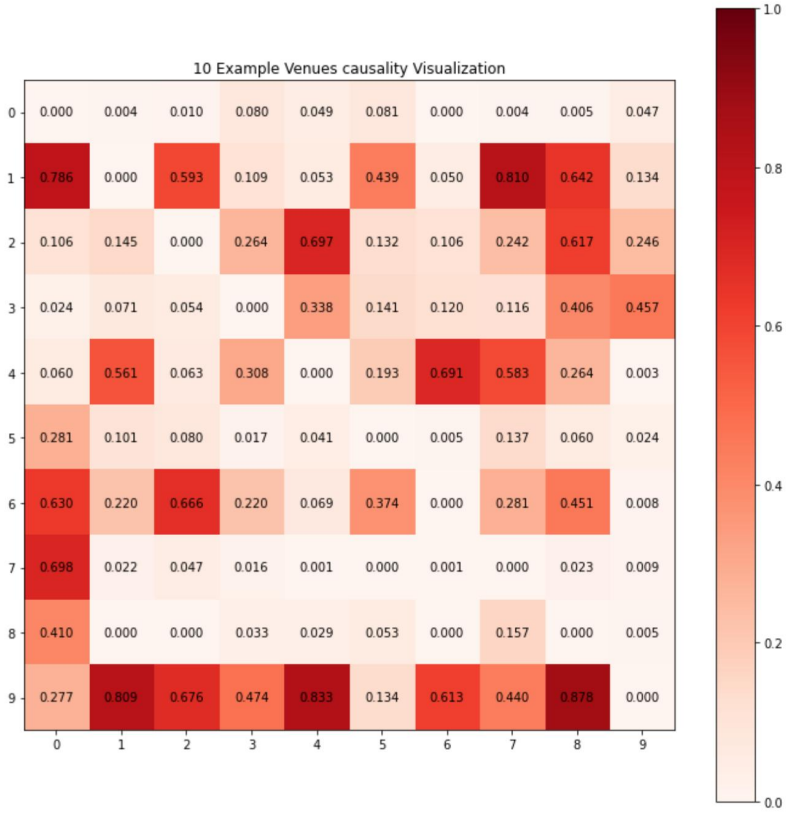


图 25

6.7 重点管控区域的确定

对于重点管控区域的确定，我们主要从阳性人员数量及辐射范围两方面进行确定。我们将上述分类中的高风险场所以及根据格兰杰检验得出的有从传播因果关系的区域进行重点管控。

在不同的疫情形势下，我们也可以依据不同的原则进行重点区域管控。在常态化管控中，我们主要从阳性人员数量方面进行重点管控的界定，仅对高风险区进行重点防控，中风险地区是否进行重点防控视情况而定。当出现少量分散式疫情时我们主要从疫情辐射范围进行衡量，有传播因果关系的区域需进行重点管控，即当关联元组中的一个场所发生疫情，则可以考虑对另一个关联场所重点防控。

七、 针对问题五的解决方案

为实现精准地进行疫情防控和人员管理，你我们认为还需采集：一个城市人员的流动信息和地表间的邻接信息，模型的处理和计算思路大致如下：

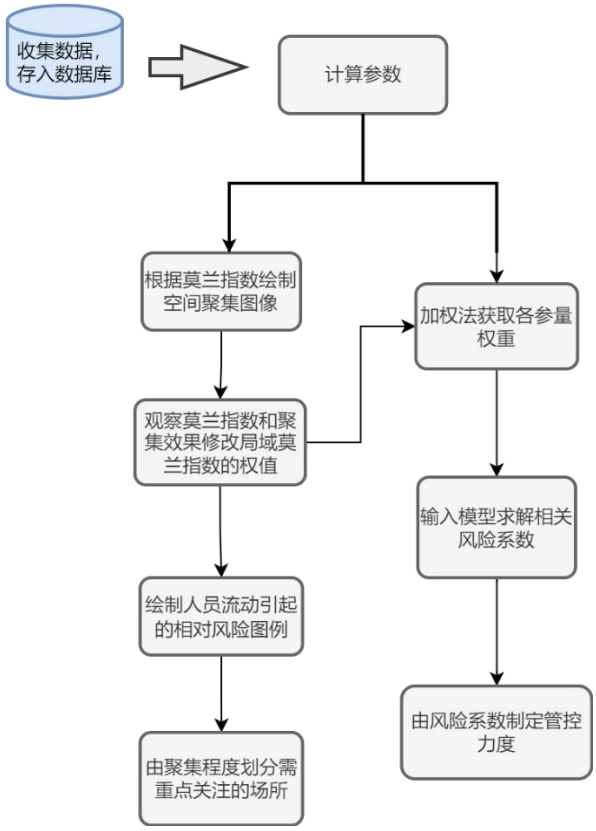


图 26

7.1 模型准备（数据准备）

根据国务院联防联控机制综合组发布的《新型冠状病毒肺炎防控方案（第九版）》科学划分风险管控区域部分的阐述，疫情低、中、高风险区域划分主要取决于（地区人员活动，传播风险等指标）。若要实现疫情的（更加）精准的防控，就需要对某一区域的传播风险实现精确的预估。人口流动是疫情扩散的关键因素，也是影响一个地区传播风险指数的重要特征。在我们现有数据中，个人自查上报信息表中虽然可以得到常住居民的信息，但是对于流动人口的数据所知甚少。非常住居民上报的坐标意义含糊，无法确定该人是流入还是流出；酸检测结果字段也并非上报的必填信息，在数据集中存在少量空值。采集一个城市确诊人员的流入流出信息，可以通过数学模型计算出某地属于输入性或是扩散性新冠传播风险

区，并根据计算结果，对输入性和扩散性传播区域施行特定的管控，以达到精准防控的目的。

对于采集到的数据，定义确诊病例中有其他城市旅居史携带病毒入省的病例是输入性，当地被感染型病例为扩散性。利用采集到的数据，计算扩散比（Deffusion Ratio，DR）和人口流动比率（Mobility Ratio, MR）

DR = 累计扩散数量 / 累计输入性数量

MR = 人口流动数量 / 常住人口的比重

7.2 模型假设

模型接收的是城市中部分地标的流动人员信息（带有流入/流出标签），输出某点的在不同时间段（与输入时间段保持一致）的风险传播指数。模型假设数据是准确理想并且认为同一时间内的扩散和输入过程时相互独立.

拥有了流动人员数据和点之间的邻接关系，利用第三问建立的 SEIDR(SIR, SEIR, SEIDR)疫情传播模型，对流入/流出人员比例较高的地点中划分易感人群，感染者和已康复的人群，通过计算这三者在某一时间段内的比例，调整对该场所的管控力度

对于城市不同场所间的人员流动，我们在考虑其相关性时，构建目标城市中各点坐标的邻接矩阵，可以采用 Moran 系数和 Geary 系数进行全域空间相关性检验[5]

$$Moran'sI = \frac{n\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i-\bar{x})(x_j-\bar{x})}{(\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij})\sum_{i=1}^{n}(x_i-\bar{x})^2} \quad (1)$$

$$(i \neq j)$$

$$\text{Geary} = (\text{n}-1)\Sigma_{\text{i}=1}^{n}\Sigma_{\text{j}=1}^{\text{n}}\omega_{\text{ij}} * \frac{\left(\text{x}_{\text{i}}-\text{x}_{\text{j}}\right)^2}{[2\text{n} * \Sigma_{i=1}^{\text{n}}(x_i-\bar{x})^2]} \quad (2)$$

式中：n表示研究的目标城市地标个数；ω_{ij} 为空间权重矩阵。由于我们的数据规模是城市规模，属于小范围内的疫情传播分析，因此空间权重矩阵中的数值，我们采用两点的地理坐标 i, j 的距离的倒数；为各标点的累计确诊人数；x 为所有标点的确诊数量均值。 Moran 系数的范围应该在[-1, 1]。当系数小于 0 时认为空间范围内的数据具有较为显著的负自相关性，为正则代表具有空间正自相

关。

相较于莫兰指数，Geary 系数则能够反映出数据在空间的分散程度，适用于大范围规模数据。由于我们的数据规模是单一城市范围内的，Moran 指数理论上更适用于我们的模型。但我们仍可使用 Geary 系数来辅助验证模型的结果。Geary 系数的取值范围在[0, 2]，当值小于 1 时则说明空间内数据具有正相关性，大于 1 则具为负相关。[6]

全域空间相关性检验反映出目标城市整体空间分布情况，难免会损失部分局部特征等信息，为了提高模型准确度，我们需要对部分地标进行局部自相关检验，以此弥补整体检验对局部特征的敏感程度。

$$LocalMoran'sI = \frac{n(x_i-\bar{x})\sum_{j=1}^{m} w_{ij}(x_j-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \quad (3)$$

$$(i \neq j)$$

我们规定莫兰指数 $|M\_I| \leq 0.5$ 时，局域莫兰指数：$M\_L = e^{\frac{1}{|M_I/M_L|}(M_I-M_L)}$，当莫兰指数绝对值不小于 0.5 时，不修改计算得到的局域莫兰指数。

$$M_L = T(M_I) = \begin{cases} e^{\frac{1}{|M_I/M_L|}(M_I-M_L)} & |M\_I| \leq 0.5 \\ M_L & else \end{cases} \quad (4)$$

除了通过某一标点内的确诊数量来判断其扩散风险指数外，我们还能通过确诊病例的流入流出数据评估其外部传播的风险：

$$RR_i = \begin{cases} I \times (1+P) & P \geq 0 \end{cases}$$

$$RR_0 = \begin{cases} I \times (1-P) & P \leq 0 \end{cases} \quad (5)$$

公式中 I 为发病率；P 为人口流动比率；RR_i为人口流入对本地所致的相对风险系数；RR_0为人口流出对外部所致的相对风险系数。

7.3 模型求解

首先利用数据集，取出目标时间段和地点，将空间数据进行标准化，然后计算第二部分指出的 4 个系数——莫兰指数：M_I，局域莫兰指数：M_L，Geary 系数：GC，人口流入相对风险系数：RR_i，人口流出相对风险系数：RR_0。根据地标邻接数据建立邻接矩阵M_0。使用莫兰指数公式计算每个地理位置的莫兰指数，从而评估疫情在地理空间上的聚集程度。如果某个地理位置的莫兰指数较高，说明疫情在该地区存在聚集现象，需要加强该地区的管控措施，如加强人员管控、加强检疫等。

根据莫兰指数的定义，当莫兰指数绝对值接近 1 时，说明这个区域整体在观测量上具有较强（正、负）相关性，此时我们就需要再使用局域莫兰指数来评判局部的特征，探索集聚出现的范围和位置，即定位区域整体中某一个需要进行防控的场所。因此对于某一点在某特定时间段传播风险指数R_s的计算需要根据M_I的值进行调整来降低莫兰指数绝对值较小时，局部莫兰指数受到这种弱自相关性的影响发生的特征损失程度。但是，我们不能舍弃局部莫兰指数。即使在莫兰指数绝对值很小时，局部莫兰指数仍然可以揭示空间上的局部模式和异质性。

第二步，利用加权法对 4 个系数标准化：首先取出数据的极大值，并计算每一个数据以此极大值为量纲的数值。这样的权值计算方法能够最大限度的保留源数据的分布:

$$\mathrm{W_i} = \frac{|\mathrm{x_0}|}{\max\ |\mathrm{x}_{ij}|}\ (\mathrm{i}\ =\ 1,2,\ldots,\mathrm{n};\ \mathrm{j}\ =\ 1,2,\ldots,\mathrm{m})$$

根据加权法得到模型求解函数各关键参数的权值，计算R_s:

$$loss = 1 \div \frac{\omega_{\mathrm{i}}\mathrm{M_i} + \omega_{\mathrm{l}}\mathrm{M_l} + \omega_{\mathrm{g}}\mathrm{G_g} + \mu_{\mathrm{i}}\mathrm{RR_i} + \mu_0 RR_0}{\mathrm{E_1M_i} + \mathrm{E_2M_l} + \mathrm{E_3G_g} + \mathrm{E_4RR_i} + \mathrm{E_5}RR_0}$$

$$\mathrm{R_s} = f(T(M_I), M_I, GC, RR_i, RR_0),\quad argmin\{loss\}$$

其中 $\omega_{\mathrm{iM_i}}, \omega_{\mathrm{lM_l}}, \omega_{\mathrm{gG_g}}$ 是通过全域 Moran 相关系数、局部 local Moran 相关系数和 Geary 相关系数三者加权计算得出，计算过程中应注意 Geary 系数；μ_{iRR_i}，μ_{0RR_0}则是通过流入对本地风险和流出对外部风险系数二者加权。分母则是对所有参量加权平均。 $\mathrm{R_s}$ 值越大，代表该时间段内该点传播新冠的风险越大（越容易产生阳性确诊、密接或次密接人员）。$\mathrm{R_s}$ 的取值范围应在[0, 1]内。通过定义

加权后的系数值（取值范围为（0，1））的倒数为 loss，找寻 loss 最小的过程便是找寻某一个地点其加权系数值最大的过程，加权系数值越大说明在给定数据集的情况下，该地发生疫情传播的风险越大。

根据模型输出每个地点在某一时间段的传播风险指数，我们以此能够设计出不同风险段的防控措施。根据多个时间段的风险指数对比，我们还能够从中得出在前一时间段内提出的防控措施的效果并进行调整，实现空间和时间上的精准防控。

7.4 模型评价

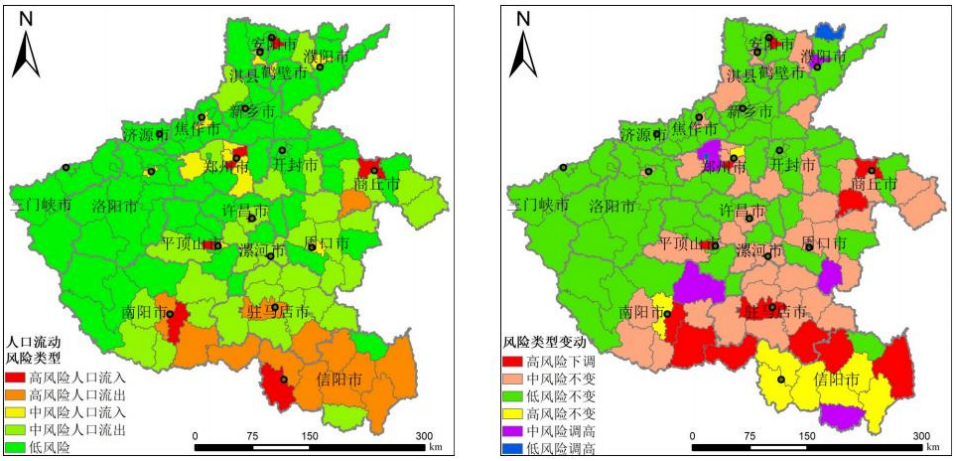
相较于原有模型，该模型考虑地理信息，通过输入额外收集的人员流动信息和流动地标间的数据，分析疫情在不同地区之间的传播程度和风险分布情况。该模型中的莫兰指数可以寻找疫情高发区域：通过计算莫兰指数，可以确定某个地区疫情是否呈现出空间聚集性，从而确定疫情高发区域；预测疫情传播趋势：莫兰指数可以帮助我们分析疫情在不同地区之间的空间关联性，从而预测疫情传播的趋势和方向；制定疫情防控策略：根据莫兰指数的结果，可以有针对性地制定疫情防控策略，比如在疫情高发区域加强监测和隔离措施，以减缓疫情的传播速度。

该模型的确定也很明显——流入流出数据的准确采集是困难的，信息量大，信息准确度不够。此外该模型的关键是数据集的空间相关性。而相关性的检验主要信息来自莫兰指数。对于本题所给出的数据，此模型应能较好反映出城市中各场所传播风险。但对于地理较大的区域，莫兰指数可能会出现误判。因为莫兰指数强调的是空间上的相似性，在地理较大的区域，莫兰指数的计算结果是不够准确的。最后，单独靠莫兰指数，仅仅讨论地理位置之间的相似性是行不通的。本题建立此模型的数据基础还有人员出行数据，如场所扫码信息表等。而人口密度、交通条件等因素均未被莫兰指数纳入考虑。

7.5 防控效果

该模型在题目所给数据的设定下，额外获取地标邻接信息数据和流动数据，可以实现空间和时间上的精准防控——通过 Moran 等系数绘制聚集图像，得到重点管控场所的候选；通过模型计算传播风险，量化疫情防控的程度；以时间为单位，可以追溯不同时期采取的防控措施的有效程度，及时调整给出的新的调控方案。

在[5]中则考虑了莫兰系数，从病例输入、扩散比、总量三个维度评估了春节前后河南省的疫情风险。由于数据集完整性更强，评估效果较为可靠，及时且较为准确地评估了疫情动态，并兼顾人口流动风险进行动态疫情风险划分，实现了多时段，大范围精准防控，为当时其他省份城市的疫情防控提供了思路和经验。



《河南省新冠肺炎疫情时空扩散特征与人口流动风险评估》中不同时段河南省疫苗传播风险评估图示

类似地，在[6]一文中，将流入流出的人员信息进一步细化，实现了多级行政区域的划分与防控。

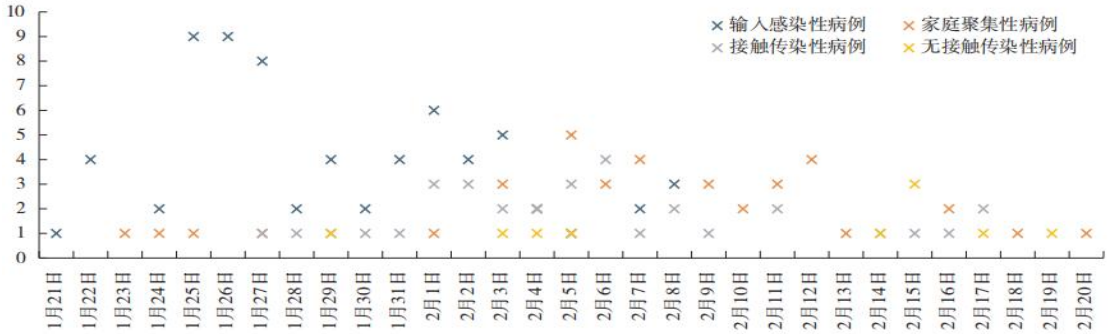


图 27

我们的根据额外采集数据建立的模型中用到了多个关键参量系数，其系数的权值设定参考自[7]中疫情分线指标五大维度的划分思路和[8]中对于不同数据来源、不同数据内容设置数据集间权重。这些设置方法使得他们克服了模型需要大量大规模数据时莫兰指数显著性水平下降的问题，从实践角度验证了及时、准确地进行风险评估的重要性。

维度	子项指标	数据来源
确诊病例流入风险	(1)本市在确诊患者城市流动网络中的中心度 (2)本市在确诊患者城市流动网络中的连接度	腾讯微信确诊患者同行程查询工具平台
湖北人群流入风险	(3)武汉地区流入本市人群概率强度 (4)湖北地区流入本市人群概率强度	百度地图迁徙数据平台
医疗卫生资源风险	(5)本市每万人执业医师数量 (6)本市每万人医院数量 (7)本市每万人定点发热门诊数量 (8)本市每万人百度搜索医疗物资的频度	第六次人口普查数据、《中国统计年鉴 2018》、百度搜索指数平台
前期疫情累积风险	(9)本市每万人确诊人数 (10)本市首例病患确诊以来的时间 (11)本市新冠肺炎传染 R0 再生数	国家和省市卫健委网站
综合发展水平风险	(12)本市人均国内生产总值(代理变量)	《中国统计年鉴 2018》

遗憾的是，由于我们难以寻找到与赛题数据有关的流动人员数据和场所邻接信息，我们无法从实践出发验证模型的准确度和鲁棒性，希望日后能够通过某省某市的大量数据进行模型的后续优化，同时也是我们完善模型的关键步骤。

参考文献

[1] 薛依琳. 基于 SIR 模型对新冠肺炎疫情基本再生数的研究[D].10.27439/d.cnki.gybdu.2022.001290

[2] CoVstat, Modelling the spread of Covid19 in Italy using a revised version of the SIR model[J].2020.05

[3] 曹盛力,冯沛华,时朋朋. 修正 SEIR 传染病动力学模型应用于湖北省 2019

冠状病毒病(COVID-19) 疫情预测和评估.[J]浙江大学学报.2020.04

[4] HuaMD, 新冠肺炎（COVID-19）的死亡率[OL].2022.11.07

[5]刘勇,杨东阳,董冠鹏等.河南省新冠肺炎疫情时空扩散特征与人口流动风险评估——基于 1243 例病例报告的分析[J]. 经济地理,2020,40(03):24-32.DOI:10.15957/j.cnki.jjdl.2020.03.004.

[6] 赵宏波,魏甲晨,王爽等.大城市新冠肺炎疫情风险评估与精准防控对策——以郑州市为例[J]. 经济地理,2020,40(04):103-109+124. DOI:10.15957/j. cnki.jjdl. 2020.04.012.

[7] 陈云松,陈步伟,句国栋等.突发重大疫情下城市系统风险量化评估方法[J].西安交通大学学报(社会科学版),2020,40(04):33-41. DOI:10.15896/j. xjtuskxb.202004004.

[8] 毛漪扬,孙华君,杜汐.天津市基层医疗卫生机构疫情防控风险评估与控制研究[J].中国医疗管理科学,2022,12(02):84-88.