

摘要

本文构建灰色马尔科夫模型，预测 2019 年全国感染该疾病的发病人数和死亡人数。再通过该预测模型，预测各个人群与地区的发病率特征。将发病率特征带入 topsis 模型，得到防控排名前 3 位的重点区域和重点人群。

针对问题一，在传统灰色预测模型的基础上，附加马尔科夫链校正预测结果误差，得到灰色马尔科夫模型，预测 2019 年传染病数据。本组首先建立传统 GM(1,1) 模型，预测分析 2004-2019 年该流行病的发病人数和死亡人数。再通过马尔科夫模型，由 2004-2016 年的数据模拟残差在各个区间的分布，计算 2017-2019 年预测残差的期望值。最后将预测结果与残差期望做差，校正传统灰色预测的固有偏差。求得 2019 年患病人数 740960 人、死亡人数 2211 人。

针对问题二，基于问题一模型，预测各人群、各地区的患病情况特征，再通过 topsis 法选出各人群、各地区排名前三的重点防控对象。首先将附件中各个省市与各个职业的发病人数、死亡人数数据，带入问题一的灰色马尔科夫模型中。预测出 2019 年发病人数、死亡人数，发病人数增长率、死亡人数增长率和死亡率。并将上述五个指标带入 topsis 综合评价模型，分别赋予权重 0.15、0.3、0.15、0.3 和 0.1，，计算各人群、各地区的综合评价得分，选取排序前三位的省市——新疆、西藏和青海和排序前三位的职业——农民、家政家务和退休人员分别作为重点防控区域和重点防控人群。

针对问题三，基于问题二中的模型，延用遗传算法，并调整泊口仿真模型，增加摧毁泊口事件，求解出泊口被摧毁对应的最佳装载方案。首先调整泊口仿真模型，当运行时间等于攻击时间时，停止被摧毁港口正在进行的任务，并将任务进度清零，并在之后的检索列表中剔除被摧毁泊口。重新运行遗传算法以寻找港口被摧毁对应的最佳装载方案。最优方案为军用登陆舰全部调用，民用船调用方案为 2 万吨级滚装船 18 艘，3 万吨级滚装船 3 艘，集装箱船 1 艘，杂货船 1 艘和客船 1 艘。装载时间为 20 小时，共使用船舰 365 艘。

本文中所提到的模型优点主要有两点：一、使用沙石算法与泊口仿真模型，使得军舰与泊口的利用率达到较高水平，最终方案转载时间少，使用船支数量少；二、利用遗传算法优化装载方案，鲁棒性强，全局搜索能力强。

关键词：沙石算法 遗传算法 泊口仿真模型 部队装载优化模型

目录

一、 问题重述	3
1.1 问题背景	3
1.2 问题概述	3
二、 模型假设	3
三、 符号说明	4
四、 问题一模型的建立与求解	5
4.1 问题描述与分析	5
4.2 模型的建立	5
4.2.1 灰度预测 GM(1,1)	5
4.2.2 马尔科夫模型校正	6
4.2.3 预测结果评价指标	7
4.3 灰色马尔可夫模型的求解	7
4.4 结果分析	10
五、 问题二模型的建立与求解	10
5.1 问题描述与分析	10
5.2 模型的建立	10
5.2.1 预测评价指标计算	10
5.2.2 topsis 法	11
5.3 模型的求解	12
5.4 结果分析	14
六、 模型的评价	15
6.1 模型的优点	15
6.2 模型的缺点	15
6.3 模型改进	15
附录 A 代码	17
A.1 问题一沙石算法-python 源代码	17

一、问题重述

1.1 问题背景

随着全球化的进程，人类活动范围日益扩大，人群流动频繁，传染病可在大范围内迅速传播，是对人类社会存在威胁的公共卫生问题。在疾病控制实际工作中，疾病的发病与流行趋势分析是极其重要的一环，科学、准确的分析能对卫生行政部分制定疾病预防与控制策略产生重要的影响，传染病早期预警将大大降低传染病的社会经济危害。

为了提高某传染病疫情和突发公共卫生事件报告的质量和时效，加强对全国感染病人的诊断、治疗和督导管理，卫生部建立了全国监管机制，及时通报相关病情和相关数据，并通过对疫情数据的动态分析，建立该传染病防治工作督导检查、防治效果评价和制定防治对策和策略，控制并逐渐消灭该传染病。构建预测模型从早期探测到传染病的爆发并及时预警，采取应对措施，是目前传染病防控的重要手段，具有重要的实际意义。

1.2 问题概述

围绕相关附件和条件要求，研究海运装载行动输送兵力任务的合理安排，依次提出以下问题：

问题一：根据合适的指标建立模型，分析流行病在 2004-2016 年的变化趋势，并预测 2019 年全国感染该病的发病数和死亡数。

问题二：基于 2004-2016 年每隔三年的不同地区的和职业分类的数据，建立疾病传播模型，并预测 2019 年传染病重点防控前 3 名的区域和职业人群。

问题三：结合地区经济发展的相关数据，选择一个角度建立传染病与经济发展相关的模型，并分析结论。

问题四：综合模型结果及分析，给卫生健康委员会相关部门写一封公开信，谈谈对传染病疫情防治的看法和建议。

二、模型假设

- (1) 为保证预测结果精确性，假设题目所给出数据真实可信。
- (2) 假设重点防控的区域和人群中的发病人数和死亡人数的增长率比其基数更加重要。
- (3) 假设在战争中，装载消耗更少时间的优先度高于使用更少的民用船。
- (4) 假设在装载过程中，同一泊口的船舰装载交替时间可以忽略不记。
- (5) 假设港口被摧毁时，由于提前得到信息，港口上的船只与兵力没有损失，只是正在进行的装载工作停止，且进度完全损失。

三、符号说明

符号	说明
X_i	第 i 型装备
s_{xi}	装备 X_i 的占用面积
l_i	装备 X_i 的长
w_i	装备 X_i 的宽
ε_i	装备 X_i 的面积修正系数
p_i	全副武装人员数
η_1	海军舰船的有效面积率
η_2	民用船只的有效面积率
y_k	登陆舰数量
z_n	民用船数量
D	决策向量
S_D	舰载面积系数向量
S_z	民用船有效面积向量
S_y	登陆舰有效面积向量
S_p	每营人口面积向量
A^k	面积规划向量
T_n	编制总部对数列
P_n	舰船面积数列
η	平均有效面积利用率

四、问题一模型的建立与求解

4.1 问题描述与分析

问题一要求，根据附件中 2004 年至 2016 年的流行病相关数据，预测 2019 年全国感染该疾病的发病人数和死亡人数。本组首先选择合适的指标后建立灰色预测模型，预测分析 2004-2019 年该流行病的发病人数和死亡人数。再通过马尔科夫模型，由 2004-2016 年的数据模拟残差在各个区间的分布，计算 2017-2019 年预测残差的期望值。最后将预测结果与残差期望做差，校正传统灰色预测的固有偏差，经过两种模型的有效结合达到科学预流行病的未来发展趋势。其问题一思维流程图如图 1 所示：

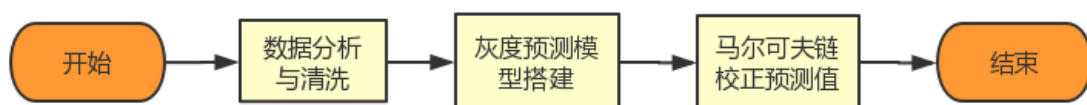


图 1 问题一思维流程图

4.2 模型的建立

4.2.1 灰度预测 GM(1,1)

设 2004-2016 年总发病人数为时间序列：

$$X^{(0)} = [x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(13)]$$

通过一次累加生成 1-AGO 序列：

$$X^{(1)} = [x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(13)]$$

式中： $x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i), k = 1, 2, \dots, 13$ 。

根据 1-AGO 序列建立微分方程为 [1]：

$$\frac{dX^{(1)}}{dt} + aX^{(1)} = u \quad (1)$$

式中： a 称为发展灰度， u 称为内生控制灰度。设 $\hat{\alpha}$ 为待估参数向量，且 $\hat{\alpha} = [a, u]^T$ ，利用最小二乘法求出：

$$\hat{\alpha} = (B^T B)^{-1} B^T Y_n \quad (2)$$

求解方程 (1)，可得第 $k+1$ 年传染病发病数初步预测模型为：

$$\hat{X}(k+1) = [X^{(0)}(1) - \frac{u}{a}]e^{-ak} + \frac{u}{a}, k = 1, 2, \dots, 16 \quad (3)$$

同理将死亡数作为向量 $X^{(0)} = [x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(13)]$ 带入模型可求得 2017-2019 年死亡数灰度预测值。

4.2.2 马尔科夫模型校正

利用马尔科夫模型对 GM(1,1) 预测误差项的状态及状态概率进行预估，并利用预测状态的期望值对 GM(1,1) 预测值进行修正 [2]。用 2004-2016 年预测数据与真实数据残差进行状态划分，设残差序列为：

$$\varepsilon = [\varepsilon(1), \varepsilon(2), \dots, \varepsilon(13)]$$

最大残差绝对值为 $\delta_{max} = \max_{1 \leq i \leq 13} |\varepsilon(i)|$ ，将预测误差化均分为三个状态。令 $\lambda = \frac{\delta_{max}}{6}$ 。状态分别为 $E_1 : (-3\lambda, -\lambda)$ 、 $E_2 : (-\lambda, \lambda)$ 和 $E_3 : (\lambda, 3\lambda)$ 。其中初始状态概率向量计算公式为：

$$\begin{cases} t_0 = [p_{E1}, p_{E2}, p_{E3}] \\ p_{Ek} = \frac{n_{Ek}}{13} \end{cases} \quad (4)$$

式中： n_{Ek} 是状态 E_k 在 2004-2016 年内出现的次数，以状态 E_k 出现的频率代替其出现的概率 p_{Ek} 。且构建状态转移矩阵为：

$$P = \begin{pmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{pmatrix}$$

式中： P_{ij} 是由状态 E_i 经过一个时期转移到 E_j 的转移概率。

即马尔科夫模型可表示为：

$$t_{k+1} = t_k \cdot P \quad (5)$$

设状态区间的中间值分别为 \bar{E}_1 、 \bar{E}_2 和 \bar{E}_3 ，即第 k 年 GM(1,1) 的误差期望为：

$$\eta = [p_{E1} \quad p_{E2} \quad p_{E3}] \cdot \begin{bmatrix} \bar{E}_1 \\ \bar{E}_2 \\ \bar{E}_3 \end{bmatrix} \quad (6)$$

当第 k 年的患病人数的 GM(1,1) 预测值为 $\hat{x}(k)$ 时，修正后的灰色马尔可夫组合预测模型 $\bar{x}(k)$ 可以记作：

$$\bar{x}(k) = \hat{x}(k) - \eta \quad (7)$$

4.2.3 预测结果评价指标

均方根误差 (RMSE)、平均相位误差绝对值 (MAPE) 和纳什效率系数 (NSE) 三者是常用来衡量预测结果的指标。RMSE 能评价患病人数和死亡人数中高值的预测结果，其计算公式为：

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y_i^*)^2}$$

均方根误差越小，表明模型可靠性越高，结果越准确。

MAPE 用来评价预测数据中平稳部分的预测结果，其计算公式为：

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - y_i^*}{y_i} \right| \times 100\%$$

MAPE 所求值为绝对值，是一个相对指标，当两个 MAPE 值进行比较时，值越小的说明模型可靠性越高。

NSE 可以用来评价模型的预测能力，其计算公式如下：

$$\text{NSE} = 1 - \frac{\sum_{i=1}^n (y_i - y_i^*)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

求得 NSE 值越接近 1，表示模型质量越好，模型可信度越高。接近 0，表示模拟结果接近观测值的平均水平，即总体结果可信，但模拟误差较大。远远小于 0，则模型是不可信的。

4.3 灰色马尔可夫模型的求解

通过 GM(1,1) 计算 2004-2016 年发病人数预测值得到灰度预测解如下：

$$\hat{X}(k+1) = -2527359e^{-0.037k} + 3497638, k = 1, 2, \dots, 16 \quad (8)$$

其误差状态区间如表 1 所示：

表 1 发病人数状态区间划分

状态	E_1	E_2	E_3
残差区间	$(-66389, -22130]$	$(-22130, 22130]$	$(22130, 66389]$

根据误差区间范围，将 2004-2016 年发病人数预测值归类于误差区间如表 2 所示：

表 2 发病人数误差状态区间

年份	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
状态区间	E_2	E_2	E_1	E_2	E_3	E_2	E_1	E_1	E_2	E_2	E_2	E_2	E_2

由此求得初始状态概率向量 t_0 , 转移矩阵 P 为:

$$t'_0 = [3/13, 9/13, 1/13]$$

$$P' = \begin{pmatrix} 1/3 & 2/3 & 0 \\ 1/4 & 5/8 & 1/8 \\ 0 & 1 & 0 \end{pmatrix} \quad (9)$$

得到由灰色预测与马尔科夫校正后预测解如图 2 所示:

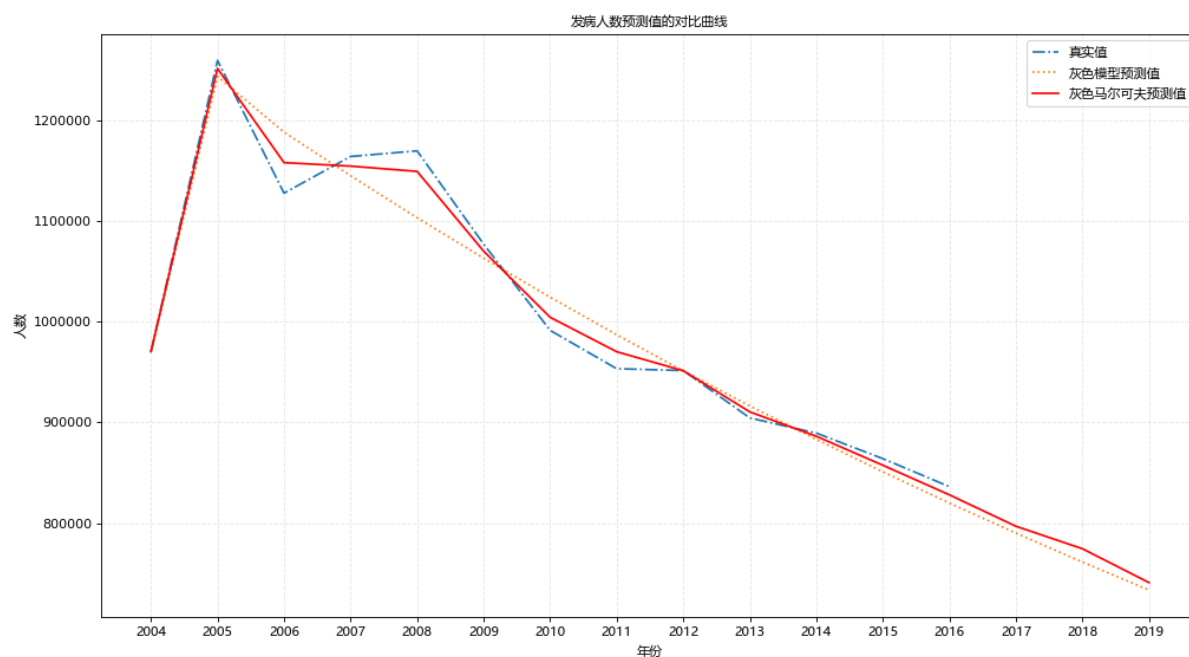


图 2 发病人数预测对比曲线图

同理, 计算 2004-2016 年死亡人数预测值得到灰度预测解如下:

$$\hat{X}(k+1) = -92315e^{-ak} + 93750, k = 1, 2, \dots, 16 \quad (10)$$

其误差状态区间如表 3 所示:

表 3 死亡数状态区间划分

状态	E_1	E_2	E_3
残差区间	$(-684, -228]$	$(-228, 228)$	$(228, 684]$

将死亡人数预测值归类于误差区间如表 4 所示：

表 4 死亡人数误差状态区间

年份	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
状态区间	E_2	E_2	E_2	E_3	E_1	E_3	E_2	E_2	E_2	E_2	E_1	E_2	E_2

求得初始状态概率向量 t'_0 , 转移矩阵 P' 为:

$$t'_0 = [2/13, 9/13, 2/13]$$

$$P' = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/8 & 3/4 & 1/8 \\ 1/2 & 1/2 & 0 \end{pmatrix} \quad (11)$$

得到由灰色预测与马尔科夫校正后预测解如图 3 所示：

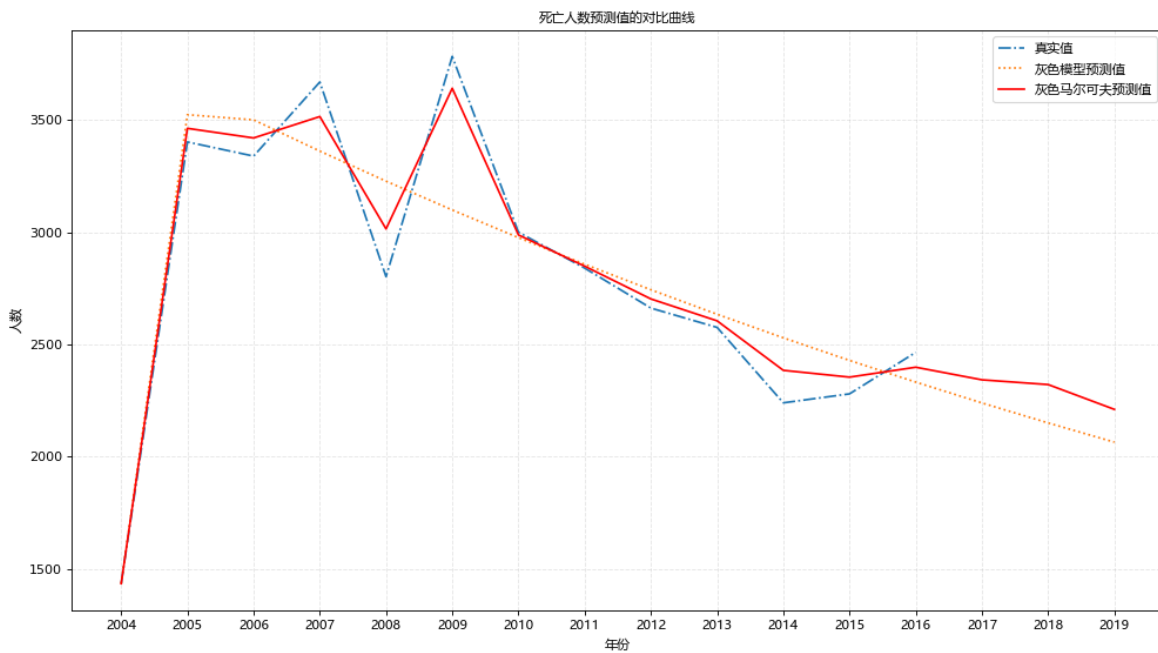


图 3 死亡人数预测对比曲线图

4.4 结果分析

由图 2 和图 3 中的预测解曲线直观对比可知, 由马尔科夫模型校正后的预测值相较于普通灰色预测值与实际值的拟合度更高, 波动性一致, 且灰色模型预测值是一条相对平滑的曲线, 不能反应实际值的波动。两种模型预测指标如表 5 所示:

表 5 预测结果检验

检验参数	RMSE	MAPE	NSE
传统灰色预测数值 (患病数)	30040.04	0.0213	0.9455
灰色马尔科夫预测数值 (患病数)	12838.64	0.0095	0.9900
传统灰色预测数值 (死亡数)	265.88	0.0628	0.8178
灰色马尔科夫预测数值 (死亡数)	101.13	0.0273	0.9736

从上述的预测结果可以得出: 利用灰色马尔科夫模型修正后求出的患病人数与死亡人数的均方根误差值 RMSE, 其值都小于传统灰色模型预测值 RMSE 值, 表明校正后结果可靠性更高。且修正后模型 MPAE 值较于传统模型更接近 0, NSE 值更接近 1, 说明改进后的灰色马尔科夫模型的拟合程度更高, 预测效果更好, 适用于传染病发病数和死亡数的短期预测。

五、 问题二模型的建立与求解

5.1 问题描述与分析

问题二本质是一个预测与综合评价问题, 题目要求结合不同地区和职业分类统计的数据, 预测 2019 年传染病防控排名前 3 位的重点区域和重点人群。首先将附件中不同地区与不同职业的发病人数、死亡人数数据, 带入问题一的灰色马尔科夫模型中。预测出 2019 年发病人数、死亡人数, 并计算出每个地区和每种职业对应的发病人数增长率、死亡人数增长率和死亡率。并将上述五个指标带入 topsis 综合评价模型, 分别选取排序前三位的省市和人群作为重点防控区域和人群。

5.2 模型的建立

5.2.1 预测评价指标计算

将附件中 20 个职业 2004 年, 2007 年, 2010 年, 2013 年和 2016 年的患病人数数据 $X_i = [x_{i1}, x_{i2}, x_{i3}, x_{i4}, x_{i5}] (i = 1, 2, \dots, 20)$, 带入问题一灰色马尔科夫模型, 得到预测结

果:

$$X_i = [x_{i1}, x_{i2}, x_{i3}, x_{i4}, x_{i5}, x_{i6}] (i = 1, 2, \dots, 20) \quad (12)$$

同理, 将死亡人数数据 $X'_i = [x'_{i1}, x'_{i2}, x'_{i3}, x'_{i4}, x'_{i5}] (i = 1, 2, \dots, 20)$ 带入第一问模型, 得到预测结果:

$$X'_i = [x'_{i1}, x'_{i2}, x'_{i3}, x'_{i4}, x'_{i5}, x'_{i6}] (i = 1, 2, \dots, 20) \quad (13)$$

则预测中 2019 年发病人数增长率 δ_i 、死亡人数增长率 δ'_i 和患病人口死亡率 η_i 分别为:

$$\delta_i = \frac{x_{i6} - x_{i5}}{x_{i5}} \times 100\% \quad (14)$$

$$\delta'_i = \frac{x'_{i6} - x'_{i5}}{x'_{i5}} \times 100\% \quad (15)$$

$$\eta_i = \frac{x'_{i6}}{x_{i6}} \times 100\% \quad (16)$$

得到各个人群的决策属性向量:

$$d_i = [x_{i6}, x'_{i6}, \delta_i, \delta'_i, \eta_i] \quad (17)$$

同理将全国各省即直辖市各年患病人口与死亡人口数据带入模型可得各个地区的决策属性向量 $d'_i (i = 1, 2, \dots, 31)$ 。

5.2.2 topsis 法

设职业多属性决策矩阵 $A = (a_{ij})_{20 \times 5}$ 可表示为:

$$A = \begin{bmatrix} d_1 \\ d_2 \\ \cdot \\ \cdot \\ \cdot \\ d_{20} \end{bmatrix} \quad (18)$$

将 A 标准化可得规范化决策矩阵 $B = (b_{ij})_{20 \times 5}$, 其中:

$$b_{ij} = a_{ij} / \sqrt{\sum_{i=1}^{20} a_{ij}^2}, i = 1, 2, \dots, 20; j = 1, 2, \dots, 5 \quad (19)$$

假定患病人数和死亡人数增长率高的区域更需要重点防控, 构造权重向量:

$$W = [0.15, 0.3, 0.15, 0.3, 0.1] \quad (20)$$

即可求得加权规范矩阵为 $C = (c_{ij})_{20 \times 5}$:

$$c_{ij} = w_j \cdot b_{ij}, i = 1, 2, \dots, 20; j = 1, 2, \dots, 5 \quad (21)$$

即可求得正理想解 $C^* = [c_1^*, c_2^*, c_3^*, c_4^*, c_5^*]$ 和负理想解 $C^0 = [c_1^0, c_2^0, c_3^0, c_4^0, c_5^0]$, 其中:

$$c_j^* = \max_i(c_{ij}), j = 1, 2, \dots, 5 \quad (22)$$

$$c_j^0 = \min_i(c_{ij}), j = 1, 2, \dots, 5 \quad (23)$$

计算各职业属性决策向量到正理想解和负理想解的距离。备选职业属性决策向量 d_i 到正理想解的距离为与到负理想解的距离为:

$$s_j^* = \sqrt{\sum_{j=1}^n (c_{ij} - c_j^*)^2}, i = 1, 2, \dots, 20 \quad (24)$$

$$s_j^0 = \sqrt{\sum_{j=1}^n (c_{ij} - c_j^0)^2}, i = 1, 2, \dots, 20 \quad (25)$$

计算各方案的综合评价指数:

$$f_i^* = s_j^0 / (s_j^0 + s_j^*), i = 1, 2, \dots, 20 \quad (26)$$

按 f_i^* 由大到小排列求得重点防控人群次序。同理将各地区决策向量 $d'_i (i = 1, 2, \dots, 31)$ 带入 topsis 模型可得重点防控地区次序。

5.3 模型的求解

将各职业患病人数与死亡人数的数据, 带入灰度马尔科夫模型得到预测值如表 6、7 所示:

表 6 各职业发病数预测结果

职业	2004	2007	2010	2013	2016	2019
幼托儿童	2239	1416.433	552.277	218.444	86.402	34.175
散居儿童	6179	3759.599	2454.325	1638.953	1094.463	730.862
学生	61578	65186.262	50766.106	40071.014	31629.098	24965.674
...
家政、家务及待业	54916	67490.081	78063.992	91181.696	106503.671	124400.317
其他	34028	54221.751	40402.927	29919.473	22156.188	16407.263

表 7 各职业死亡数预测结果

职业	2004	2007	2010	2013	2016	2019
幼托儿童	2	1.858	1.012	0.571	0.323	0.182
散居儿童	9	14.193	10.118	7.270	5.223	3.753
学生	22	38.123	30.824	24.931	20.164	16.309
...
家政、家务及待业	133	244.727	271.116	301.405	335.077	372.512
其他	46	143.799	101.315	70.427	48.956	34.031

解得正理想解与负理想解分别为：

$$C^* = [-0.1856, -0.1847, -0.0809, -0.1163, -0.0929] \quad (27)$$

$$C^0 = [0.0066, 0.0048, 0.0653, 0.0261, 0.0331] \quad (28)$$

即求得 20 种职业综合评价指数如图 4 所示：

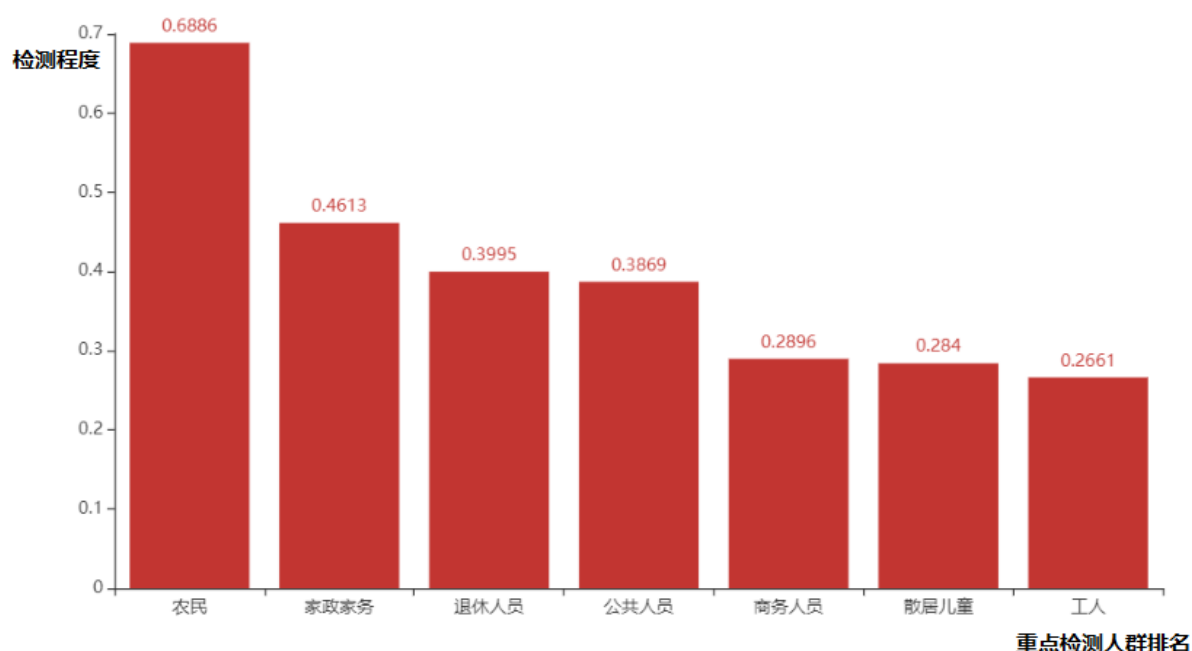


图 4 人群综合评价指数

由图 4 可知防控排名前 3 位的职业分别为农民、家政家务和退休人员。

同理将各地区患病人数与死亡人数预测值带入模型，计算正理想解与负理想解分别为：

$$C^* = [-0.0355, -0.0567, 0, 0, 0] \quad (29)$$

$$C^0 = [0.0504, 0.0704, 0.1210, 0.2310, 0.0742] \quad (30)$$

即求得 31 个地区综合评价指数如图 5 所示：

不同区域TOPSIS打分情况

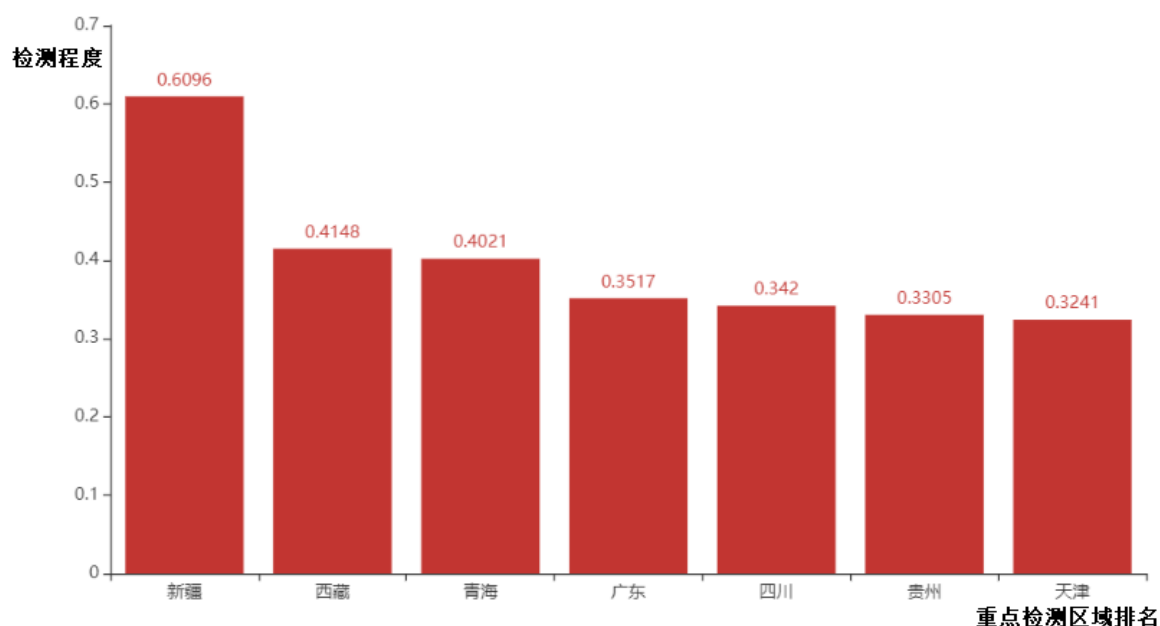


图 5 地区综合评价指数

由图 5 可知防控排名前 3 位的地区分别为新疆，西藏和青海。

5.4 结果分析

农民、家政家务和退休人员的发病人数、死亡人数灰色马尔科夫预测曲线如下图所示：

由图可知家政家务从业人员发病人数和死亡人数都处于上升趋势，且上升率较高，属于重点防控人群。而农民发病人数和死亡人数都虽处于下降趋势但其患病人口基数较大导致其综合评价指数较高，属于重点防控人群。而退休人口发病人数和死亡人数的增长率和基数都不大，但其发病后死亡较高，也属于重点防控人群。

新疆，西藏和青海的灰色马尔科夫预测曲线如下图所示：

由图可知新疆，西藏和青海的发病人数和死亡人数都处于整体上升趋势，尤其是新疆还有相对于其它二者较大的患病人口基数，都属于重点防控区域。

六、模型的评价

6.1 模型的优点

- (1) 将多目标优化转换成背包优化问题，自主设计沙石算法，使得船舰的面积利用率高，对船只数量的需求量少。
- (2) 使用遗传算法，具有很强的全局搜索能力和鲁棒性，运算时间远小于全遍历算法。

6.2 模型的缺点

遗传算法初始解由卡特蒙洛法随机生成，每次搜索结果不完全相同，可能引起结果的偏差，需要多搜索几次选择才能得到最优结果。

6.3 模型改进

可使用改进的生命遗传算法，加强算法的局部搜索能力，解决算法早熟的问题。

参考文献

- [1] Saad Ahmed Javed,Sifeng Liu. Correction to: Predicting the research output/growth of selected countries: application of Even GM (1, 1) and NDGM models[J]. Scientometrics,2019,120(3).
- [2] 李立欣, 文海东, 许健开. 基于灰色马尔可夫模型的能源消耗预测 [J]. 中国科技信息,2018(15):74-75.
- [3] Yawen Wang,Zhongzhou Shen,Yu Jiang. Analyzing maternal mortality rate in rural China by Grey-Markov model[J]. Medicine,2019,98(6).
- [4] Saad Ahmed Javed,Sifeng Liu. Correction to: Predicting the research output/growth of selected countries: application of Even GM (1, 1) and NDGM models[J]. Scientometrics,2019,120(3).

附录 A 代码

A.1 问题一沙石算法-python 源代码
