

新冠肺炎背景下对大流行病的分类界定研究

邱纪元¹, 王富帅², 朴凤贤^{3*}

¹沈阳航空航天大学航空宇航学院, 辽宁 沈阳

²沈阳航空航天大学材料与工程学院, 辽宁 沈阳

³沈阳航空航天大学理学院, 辽宁 沈阳

Email: 1281883102@qq.com, 963356506@qq.com, *540145468@qq.com

收稿日期: 2020年9月29日; 录用日期: 2020年10月14日; 发布日期: 2020年10月21日

摘要

2020年3月12日, 世界卫生组织宣布新型冠状病毒肺炎是一种大流行病。如何对流行性传染病有关大流行病或流行病进行界定一直是备受关注的问题。本文通过选用历史上已被确认为大流行病或流行病的传染病, 依据大流行病严重程度评估框架建立分类体系, 利用主成分分析法提取主成分, 并建立系统聚类与二分类Logistic回归分析模型分别对于这些传染病进行定性定量分类研究, 为流行性传染病的界定提供一种思路。

关键词

新冠肺炎, 定性定量, 分类研究

Classification and Definition of Pandemic in the Context of COVID-19

Jiyuan Qiu¹, Fushuai Wang², Fengxian Piao^{3*}

¹Aerospace Academy, Shenyang Aerospace University, Shenyang Liaoning

²Faculty of Materials and Engineering, Shenyang Aerospace University, Shenyang Liaoning

³Faculty of Science, Shenyang Aerospace University, Shenyang Liaoning

Email: 1281883102@qq.com, 963356506@qq.com, *540145468@qq.com

Received: Sep. 29th, 2020; accepted: Oct. 14th, 2020; published: Oct. 21st, 2020

Abstract

On March 12, 2020, the World Health Organization (WHO) declared COVID-19 a pandemic. How to

*通讯作者。

define pandemics or epidemics related to diseases has been a matter of great concern. This paper selects infectious diseases that have historically been identified as pandemics or epidemics and establishes a classification system based on the framework for assessing the severity of pandemics, principal component analysis is used to extract principal components, and establish systematic clustering and dichotomous Logistic regression analysis model for the qualitative and quantitative classification of these infectious diseases. It provides a way of thinking for the definition of epidemic infectious diseases.

Keywords

COVID-19, Qualitative and Quantitative, Classification

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在漫长的人类历史发展过程中，人类曾遭遇过数不清的传染病的侵扰。2020年初，新型冠状病毒引发的病毒性肺炎席卷全球。3月12日，世界卫生组织(World Health Organization)宣布，新型冠状病毒肺炎(COVID-19)是一种大流行病(Pandemic)。关于流行性传染病的术语，“Pandemic”、“Epidemic”和“Endemic”之间的界定条件是较为模糊的，即便是一些医学专家也会较为频繁地错误使用，这是因为随着时间的推移，对于某一种传染病的传播流程或多或少的都会产生影响，造成了这些术语的定义也是时常变动的。因此，建立一个合理的流行性传染病级别的分类模型，对于世界卫生组织及时地向世界公布信息，降低传染病带来的损失就显得尤为重要。

2. 文献综述

对于流行病这一词语，最早可追溯于 Homer 的《Odyssey》中，该词后来由 Hippocrates 赋予了医学意义[1]。到17世纪初，“流行病(Epidemic)”和“地方病(Endemic)”这两个词用于指代人口水平发生疾病的对比条件，“流行病(Epidemic)”发病状况广泛，“地方病(Endemic)”发病率低[2]。亚特兰大疾病控制中心对于“流行病(Epidemic)”做出大致如下的定义：在特定地区或特定人群中发生的疾病、伤害或其他健康状况比预期的状况要严重。通常情况下，这些病例被推定为有共同原因或以某种方式相互关联[3]。Porta 和 Miquel 对大流行病(Pandemic)的定义是：一种跨越国际边界的大规模流行病，通常影响到全世界范围的人民[4]。William C. Shiel Jr.对于“地方病(Endemic)”的定义是：这种疾病传播过程具有特定人口，环境或区域的特征，但频率相对较低[5]。

对于流行病的界定条件，国内的相关研究较国外的较少。国外的学者们提出了一种大流行病严重程度评估框架(PSAF)，它使用象限来评估大流行病的传播率和临床严重程度，并将其合并为一个总体影响来对于一种传染病是否为大流行来进行评估[6]。传染病的传播率可以通过可获得的数据来进行衡量，这些数据包括地区感染率，感染人数以及感染的地区数等[7]。临床严重程度也可以通过多种措施进行计算，包括病死率，传染病的医学机理等[8]。在2014年，美国疾病控制和预防中心(CDC)将PSAF作为其官方大流行严重程度评估工具[9]。同时，它也是CDC在COVID-19大流行时国家大流行战略中列出的官方评估工具[10]。

3. 研究目的及意义

本文的研究目的是在新型冠状病毒肺炎大流行病的背景下，基于历史上已经被定义过为大流行病或流行病的传染病，建立合理的这二者的分类体系以及模型，从而对于大流行病和流行病进行定性和定量分类分析，为未来对于某一种传染病的界定提供一种研究方向。

4. 流行病和大流行病的分类体系的构建

4.1. 分类指标的选取

根据国外现有的研究方案，即大流行严重程度评估框架(Pandemic Severity Assessment Framework)中提出的两个主要因素来确定大流行病的影响。首先第一个主要因素是临床严重程度，即有关传染病感染能力的严重程度。第二个主要因素是传输率，即某种传染病是多么容易的进行人传人传播。基于以上两个主要因素，选取如下指标构建评价体系(表 1)：

Table 1. Epidemic and pandemic classification systems

表 1. 流行病和大流行病的分类体系

分类体系	主要因素	分类指标
流行病和大流行病的分类体系	临床严重程度	病死率
		表面存活时间
		潜伏期
		谷歌点击搜索量
	传输率	基本传染数
		全球感染人数
		传染病流行期
		影响国家地区数量

4.2. 分类指标的解释

4.2.1. 临床严重程度

- 病死率(Casefatality rate, 简称 CFR): 表示在一定时期内, 因患某种疾病死亡的人或动物数量占患病人或动物总数的比例, 该指标用于衡量疾病所造成的严重性评估。
- 表面存活时间(Survival time on surfaces): 表示病毒离开宿主细胞在介质表面可以存活的时间, 在这段时间内, 它可以感染一个新的宿主, 该指标可以衡量疾病难以被消灭的程度。
- 潜伏期(Incubation period): 是一般疾病在发展过程中的一定阶段, 这一阶段是从致病刺激物侵入机体或机体发生作用起, 到机体出现反应或开始呈现症状时止, 该指标可以衡量疾病难以被发现的程度。
- 谷歌点击搜索量(Google hits): 是指在谷歌搜索引擎上, 对于某一种疾病的搜索点击量, 该指标可以衡量该疾病对于民众的严重影响程度。

4.2.2. 传输率

- 基本传染数(Basic reproduction number, 简称 R_0): 是指在流行病学上, 指在没有外力介入, 同时所有人都没有免疫力的情况下, 一个感染到某种传染病的人, 会把疾病传染给其他多少个人的平均数, 该指标可以衡量该流行病难以控制的程度, R_0 的数字越大, 表示流行病越难以控制。
- 全球感染人数(Global infections): 是指在某一时间内, 某种已知病毒或未知病毒的全局感染人数。该指标可以衡量疾病的传输能力。
- 传染病流行期(Epidemic period): 是指在某一特定时间内, 动物的发病量或发病率超过正常水平, 该时期发病数量大, 波及范围广。
- 影响国家或地区数量(Number of countries or regions affected): 是指某种疾病在发病过程中的所到达的国家或地区数, 该指标可以衡量该疾病的传输能力。

5. 研究方法与步骤

5.1. 基于主成分分析的变量特征提取

5.1.1. 模型的基本思想简介

主成分分析是采取一种数学降维的方法, 其所要做的就是设法将原来众多具有一定相关性的变量, 重新组合成一组新的相互无关的综合变量来代替原来的变量[11]。本文中利用主成分分析法对于流行病和大流行病的分类体系中的分类指标进行重新组合, 从而获得新的综合变量, 即完成变量的特征提取。

5.1.2. 模型的建立

首先对于所取得的原始数据进行标准化处理, 假设样本的观测数据矩阵如下所示:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \quad (1)$$

因此, 可以按照如下的方法对于原始数据进行标准化处理:

$$x_{ij}^* = \frac{x_{ij} - \bar{x}_j}{\sqrt{\text{Var}(x_j)}} \quad (i=1, 2, \dots, n; j=1, 2, \dots, p) \quad (2)$$

接下来, 计算样本间的相关系数矩阵, 则对于标准化后的相关系数矩阵为:

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & r_{pp} \end{bmatrix} \quad (3)$$

其中,

$$r_{ij} = \frac{\sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)}{\sqrt{\sum_{k=1}^n (x_{ki} - \bar{x}_i)^2} \sqrt{\sum_{k=1}^n (x_{kj} - \bar{x}_j)^2}} \quad (n > 1) \quad (4)$$

利用得出的相关系数矩阵 \mathbf{R} , 计算其对应的特征值 $(\lambda_1, \lambda_2, \dots, \lambda_p)$ 和相应的特征向量:

$$\mathbf{a}_i = (a_{i1}, a_{i2}, \dots, a_{ip}) \quad (i=1, 2, \dots, p) \quad (5)$$

接着,计算主成分的贡献率,这里的贡献率指的是某个主成分的特征值所占全部特征值合计的比重:主成分的贡献率计算公式为:

$$\frac{\lambda_i}{\sum_{k=1}^p \lambda_k} (i=1,2,\dots,p) \quad (6)$$

累积贡献率计算公式为:

$$\frac{\sum_{k=1}^i \lambda_k}{\sum_{k=1}^p \lambda_k} (i=1,2,\dots,p) \quad (7)$$

计算主成分载荷,其计算公式为:

$$l_{ij} = p(z_i, z_j) = \sqrt{\lambda_i} e_{ij} (i, j=1,2,\dots,p) \quad (8)$$

得到各主成分载荷后,按照如下公式进行进一步计算,得到各主成分的得分:

$$\begin{cases} z_1 = l_{11}x_1 + l_{12}x_2 + \dots + l_{1p}x_p \\ z_2 = l_{21}x_1 + l_{22}x_2 + \dots + l_{2p}x_p \\ \vdots \\ z_m = l_{m1}x_1 + l_{m2}x_2 + \dots + l_{mp}x_p \end{cases} \quad (9)$$

利用原始数据经过处理后的得到各主成分得分后,便可选择这些贡献率大的主成分,作为流行病和大流行病的分类依据变量。

5.2. 基于系统聚类分析的定性分类处理

5.2.1. 模型的基本思想简介

系统聚类分析的基本思想是首先将每个样本单独聚为一类,再对样本间的“距离”进行确认,最后在依据样本间的距离进行有层次的聚类,即距离近的先聚类,距离远的后聚类,所以这种聚类方法又叫做层次聚类,不断地对此过程进行重复,直到每种类型都聚集到合适的类中[12]。本研究利用主成分分析法提取出的主成分,将所收集到的 n 种历史上已被确定的流行病或大流行病进行层次聚类,即完成这 n 种传染病的定性分类处理。

5.2.2. 模型的建立

首先将所获取的 n 个样本单独定义成 n 类,确定样本距离的计算公式,在本文中,选取了欧氏距离进行研究,其计算公式为:

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (10)$$

其中, d_{ij} 为样本 x_i 与 x_j 之间的距离。

利用上述公式,将距离最近的两个样本合成一类,剩余样本合成另一类,此时为 $n-1$ 类,重复操作,直到样本最后被聚为一类。经历上述操作,便可以得到样本传染病定性分类后的结果。

5.3. 基于二分类 Logistic 回归分析的定量分类处理

5.3.1. 模型的基本思想简介

二分类 Logistic 回归分析又被称作二项 Logistic 分析,其可以非常有效的对于数据进行二分类处理

[13]。同时，其对于结果的解释比较容易理解，利用二分类 Logistic 回归，将所收集到的传染病样本进行定量分类处理。

5.3.2. 模型的建立

设 P 为某种传染病是流行病的概率, P 的取值范围为 $(0, 1)$, 则 $1-P$ 表示该种疾病是大流行病的概率。首先对于 P 进行 Logit 转换, 即对 $\frac{P}{1-P}$ 取自然对数 $\ln \frac{P}{1-P}$ 作为因变量, 建立线性回归方程:

$$\ln \frac{P}{1-P} = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p \quad (11)$$

以上即为二分类 Logistic 模型, 根据上式, 我们便可以得到:

$$P = \frac{\exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p)}{1 + \exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p)} \quad (12)$$

上式中, β_0 表示常数项, $\beta_i (i=1, 2, \dots, p)$ 表示回归系数, x_i 表示第 i 个变量矩阵[14]。

基于以上处理, 便可以得到样本疾病定量分析后的结果。

5.4. 研究步骤

本文中的模型建立与数据处理是建立在 SPSS 24.0 版本上的。以下是本文对于所收集的 32 种历史上已被确定的流行病或大流行病进行分类判别的步骤:

第一步: 将对应收集好的 32 种疾病的数据指标利用 SPSS 软件中的“降维”操作处理, 对于结果选取累积贡献率大于 85% 的 n 个主成分, 并计算对应的主成分得分。

第二步: 利用第一步所得到的主成分, 利用 SPSS 软件中的“分类”功能, 将所收集的 32 种疾病聚类合适后为止。

第三步: 利用第一步所得到的主成分, 选取共线性较弱的主成分, 利用 SPSS 软件中的“回归”功能, 经过实验选取最优分类结果, 即为分类结果正确率较高的方法为止。

经过以上三个步骤, 我们便可以得到较优的对于某种疾病是流行病或大流行病的分类结果, 最后, 我们利用已经被判定为大流行病的新型冠状病毒肺炎对于模型的合理性进行验证。

6. 数据的来源及预处理

本文中所选用的疾病以及数据来源于世界卫生组织(World Health Organization)、美国疾病预防控制中心(Centers for Disease Control and Prevention)以及 Knoema 数据库等。根据前文解释定义的 8 项指标, 本文收集了已被定义为流行病或者大流行病的 32 种疾病的数据。

由于某些传染病的数据指标较难以统计, 因此我们对于部分统计量的数据以估计量代替样本数据的方法作为实验数据进行分析。

7. 实证分析

首先我们对于原始数据利用主成分分析法进行分析, 通过 KMO 和巴特利特的检验后, KMO 的值为 0.851, 接近于 1, 说明所使用的原始数据适合提取主成分。同时巴特利特球形度检验中的 Sig 值为 0.000, 小于显著性水平 0.05, 拒绝原假设, 说明变量间存在相关关系。同时, 利用 SPSS 软件做出原始数据的碎石图如图 1 所示。

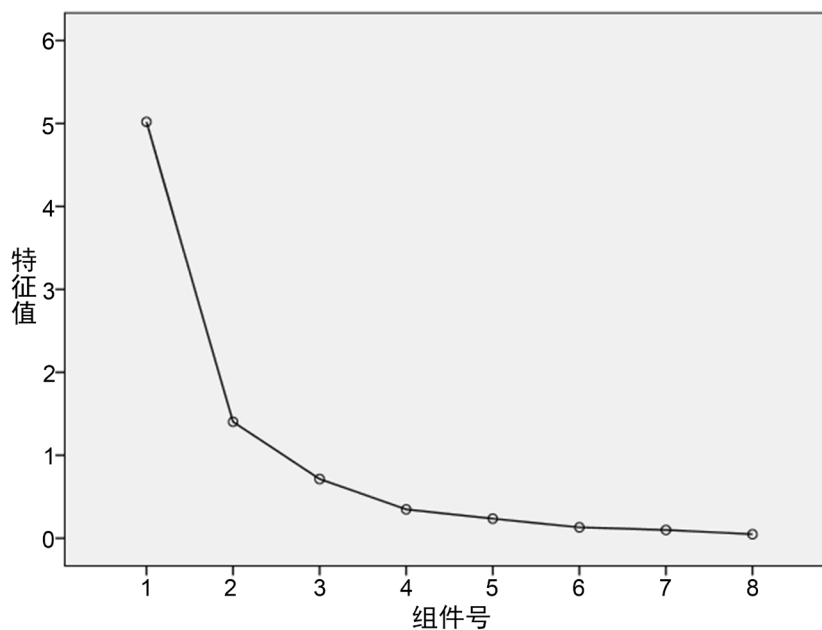


Figure 1. Gravel map
图 1. 碎石图

通过碎石图，我们可以看到前三个主成分后，特征值点的走势逐渐趋于平缓，因此，我们可以选取前三个成分作为主成分进行分析。

Table 2. The total variance of the interpretation
表 2. 总方差解释

成分	总计	初始特征值 方差百分比	累积%	总计	提取载荷平方和 方差百分比	累积%	总计	旋转载荷平方和 方差百分比	总计
1	5.020	62.754	62.754	5.020	62.754	62.754	3.811	47.642	47.642
2	1.403	17.541	80.294	1.403	17.541	80.294	2.286	28.574	76.216
3	0.714	8.924	89.218	0.714	8.924	89.218	1.040	13.002	89.218
4	0.348	4.350	93.568						
5	0.236	2.945	96.513						
6	0.131	1.640	98.153						
7	0.100	1.250	99.403						
8	0.048	0.597	100.000						

通过总方差解释表 2，前三个主成分的累积贡献率可以达到 89.218%，大于 85%，说明选取前三个主成分便包含了原始数据 89.218% 的信息，选取前三个主成分便可以说明问题。得到每种传染病的三个主成分后，绘制其两两分类散点图如图 2 所示。

由图 2 可知，FAC1_1 与 FAC2_1 之间的共线性较弱，且对于结果的分类情况较好，因此，选用 FAC1_1 与 FAC2_1 作为变量进行系统聚类分析，利用 SPSS 软件做出如图 3 所示谱系图。

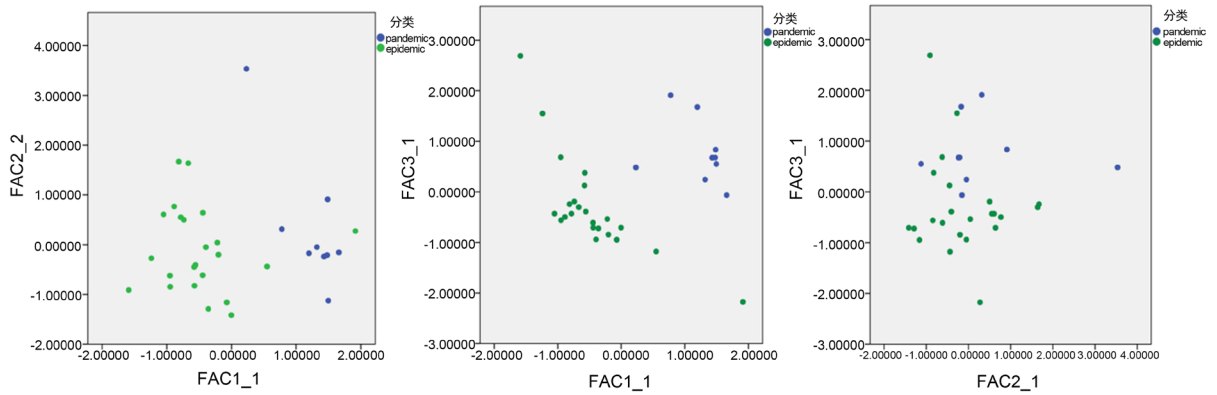


Figure 2. Classification scatter plot
图 2. 分类散点图

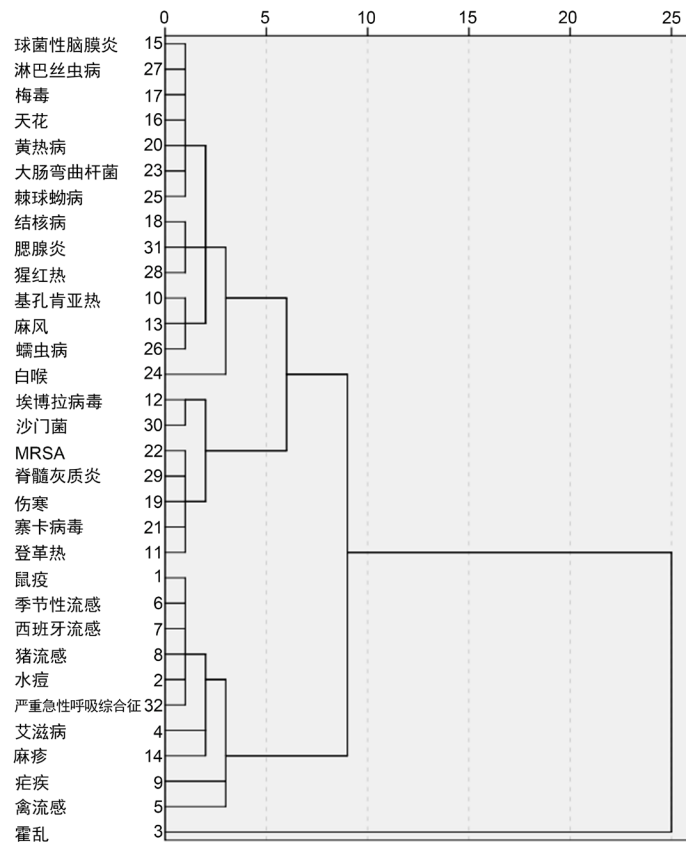


Figure 3. Pedigree map
图 3. 谱系图

根据谱系图的分类结果，可以看到，除去单独为一类的霍乱后，其余的传染病可以归为以下两大类：

- 第一大类传染病(10项)：鼠疫、季节性流感、西班牙流感、猪流感、水痘、严重急性呼吸综合征、艾滋病、麻疹、疟疾、禽流感。
- 第二大类传染病(21项)：球菌性脑膜炎、淋巴丝虫病、梅毒、天花、黄热病、大肠弯曲杆菌、棘球蚴病、结核病、腮腺炎、猩红热、基孔肯亚热、麻风、蠕虫病、白喉、埃博拉病毒、沙门菌、MRSA、脊髓灰质炎、伤寒、寨卡病毒、登革热。

其中,对于分类结果,第一大类别传染病中,大流行病占比 80%,流行病占比 20%;第二大类别中,大流行病占比 4.8%,流行病占比 95.2%。由此,可以说明对于所收集的 32 种传染病类别的定性分析结果较好。

由于前文分析 FAC1_1 与 FAC2_1 的共线性较弱,所以选用这两个变量作为解释变量,利用 SPSS 软件的回归功能建立二分类 Logistic 模型进行传染病分类的定量分析,在分别使用向前步进(似然比)与向后步进(似然比)后,得到如下的模型拟合度对比表:

Table 3. Comparison of model fit

表 3. 模型拟合度对比表

指标	方法	
	向前: LR	向后: LR
-2 对数似然	15.449	12.663
考克斯-斯奈尔 R 方	0.506	0.547
内戈尔科 R 方	0.728	0.787

表 3 的作用是检验模型拟合度的高低。-2 对数似然的值越小,表示模型的拟合程度越好;考克斯-斯奈尔 R 方和内戈尔科 R 方代替了线性回归中的 R 方,这两个指标的数值越接近于 1,说明模型最终的拟合程度越好。由表 3 得到的结果,向后: LR 方法的-2 对数似然值较向前: LR 方法小,而考克斯-斯奈尔 R 方和内戈尔科 R 方更接近于 1,所以,使用向后: LR 方法使得模型的拟合程度提高。

Table 4. H-L test randomness table

表 4. H-L 检验随机性表

步骤 1	分类 = Pandemic		分类 = Epidemic		总计
	实测	期望	实测	期望	
1	2	2.862	1	0.138	3
2	3	2.590	0	0.410	3
3	3	2.340	0	0.660	3
4	1	0.874	2	2.126	3
5	0	0.126	3	2.874	3
6	0	0.084	3	2.916	3
7	0	0.050	3	2.950	3
8	0	0.035	3	2.965	3
9	0	0.025	3	2.975	3
10	0	0.013	5	4.987	5

表 4 为 H-L 检验随机性表,从表中所得到的结果可以看到,分类 = Pandemic 与分类 = Epidemic 这两种情况下的实测值与观测值大致相等,此种结果便可认为模型的分类型准确率较高,即模型的拟合程度较优。由此,便可得到模型的拟合结果如表 5 所示。

Table 5. Fitting results
表 5. 拟合结果

	B	标准误差	瓦尔德	自由度	显著性	Exp(B)	Exp(B)的置信区间	
							下限	上限
FAC1_1	-3.086	1.062	8.440	1	0.004	0.046	0.006	0.366
FAC2_1	-0.990	0.731	1.834	1	0.176	0.372	0.089	1.557
常量	2.401	1.137	4.457	1	0.035	11.029		

表 5 是模型的拟合最终结果, 从表中可以得到 FAC1_1 和常量的差异显著性 P 值均小于 0.05, 具有统计学意义, 因此, 最终的二分类 Logistic 模型拟合方程为:

$$\text{Logit}(P) = 2.401 - 3.086 * \text{FAC1}_1 \quad (13)$$

对于拟合得到的模型, 可以对于未来某种传染病是否为大流行进行预测。即当预测值小于 0.5 时, 则该传染病为大流行病; 当预测值大于 0.5 时, 则该传染病为流行病; 选用新型冠状病毒肺炎的数据带入模型验证, 经历上述步骤, 计算得到的对应 FAC1_1 的值为 0.92177, 带入到式(13)中, 得到如下结果:

$$\text{Logit}(P) = 2.401 - 3.086 * 0.92177 = -1.1073 < 0.5$$

由上述计算结果, 依据结论, 可以判断新型冠状病毒肺炎属于大流行病。

8. 结语

本文通过选用历史上已被确认为大流行病或流行病的传染病, 根据大流行严重程度评估框架构建分类指标进行数据统计, 利用主成分分析法提取主成分, 分别建立系统聚类模型进行定性分类和二分类 Logistic 回归分析模型进行定量分类, 最后利用新型冠状病毒肺炎进行结果的验证。本文模型对于未来某种传染病的分类界定提供了一种思路, 有利于医疗机构可以更好地应对疫情, 从而降低因传染病带来的损失。

参考文献

- [1] Martin, P.M. and Martin-Granel, E. (2006) 2,500-Year Evolution of the Term Epidemic. *Emerging Infectious Diseases*, **12**, 976-980. <https://doi.org/10.3201/eid1206.051263>
- [2] Lodge, T. (1603) A Treatise of the Plague: Containing the Nature, Signes, and Accidents of the Same, with the Certain and Absolute Cure of the Fevers, Botches and Carbuncles that Raigne in These Times. Edward White, London.
- [3] (2012) Principles of Epidemiology. Third Edition, Centers for Disease Control and Prevention, Atlanta, Georgia.
- [4] Porta, Miquel, Ed. (2008) Dictionary of Epidemiology. Oxford University Press, Oxford, 179. <https://doi.org/10.1590/S1135-57272008000400008>
- [5] Shiel Jr., W.C. (2018) Medical Definition of Endemic. <https://www.medicinenet.com/script/main/art.asp?articlekey=3234>
- [6] Roos, R. (2017) New CDC Guidelines on Flu Pandemic Measures Reflect 2009 Lessons. Center for Infectious Disease Research and Policy.
- [7] Reed, C., Biggerstaff, M., Finelli, L., Koonin, L.M., Beauvais, D., Uzicanin, A., Plummer, A., Bresee, J., Redd, S.C. and Jernigan, D.B. (2013) Novel Framework for Assessing Epidemiologic Effects of Influenza Epidemics and Pandemics. *Emerging Infectious Diseases*, **19**, 85-91. <https://doi.org/10.3201/eid1901.120124>
- [8] Qualls, N., Levitt, A., Kanade, N., Wright-Jegede, N., Dopson, S., Biggerstaff, M., Reed, C. and Uzicanin, A. (2017) Community Mitigation Guidelines to Prevent Pandemic Influenza, United States, 2017. Morbidity and Mortality Weekly Report. Center for Surveillance, Epidemiology, and Laboratory Services, Centers for Disease Control and Prevention. 66(RR-1), 1-34. <https://doi.org/10.15585/mmwr.rr6601a1>

-
- [9] Holloway, R., Rasmussen, S.A., Zaza, S., Cox, N.J. and Jernigan, D.B. (2014) Updated Preparedness and Response Framework for Influenza Pandemics. *Morbidity and Mortality Weekly Report*. Center for Surveillance, Epidemiology, and Laboratory Services, Centers for Disease Control and Prevention. 63 (RR-6), 1-18.
- [10] Pandemic Severity Assessment Framework (PSAF) (2016) National Pandemic Strategy. Centers for Disease Control and Prevention.
- [11] 卓金武, 王鸿钧. MATLAB 数学建模方法与实践[M]. 第三版. 北京: 北京航空航天大学出版社, 2018.
- [12] 赵姗姗. 基于 SPSS 中系统聚类的 CPI 分析[D]: [硕士学位论文]. 新乡: 河南师范大学, 2013.
- [13] 谭中明, 谢坤, 丁国平. P2P 网贷信用风险判别研究——基于二项 Logistic 分类模型[J]. 海南金融, 2018(8): 4-12.
- [14] 赵沛. 二分类 Logistic 回归模型对上市公司财务状况的预测效度研究[D]: [硕士学位论文]. 南宁: 广西大学, 2015.