

Surveillance Study on Virus Infection

Jiayuan Liu¹, Chuwei Zhou², Xuan Fang³, Zhe Zhang⁴, Yuqi Bai⁵

City College of Dalian University of Technology Dalian, Liaoning 116000

Abstract: According to the hidden characteristics of different attribute groups and viruses, the entropy weight method of MATLAB is used to determine the index weight and comprehensive score of different industries, and objectively evaluate through charts and model results, and design a reasonable monitoring scheme. Question 1: Through quantitative analysis, MATLAB simulation and entropy weight method to calculate the weight and comprehensive score, evaluate the transmission mode, different groups and virus concealment, and take a certain city as an example for modeling analysis, so as to determine the goal of finding virus carriers as soon as possible with the as small number of people tested as possible. Through scientific analysis and model calculation, improve the effect of epidemic monitoring and control, and find the virus carriers as soon as possible.

Keywords: Virus carrier; Monitoring scheme; Entropy weight method

1. Foreword

Surveillance research on viral infection is an important public health effort to grasp and monitor the spread of different viruses in time and help prevent and control the spread of disease. Virus infection is widespread, rapid and concealed, posing a serious threat to human health and social stability. Therefore, the establishment of an effective viral infection surveillance system is crucial for the early detection of outbreaks, appropriate measures and public health protection. In the surveillance study of viral infection, pathogen surveillance, new pathogen discovery, data analysis and model building need to be considered. Surveillance study of viral infection is a complex and urgent task that requires the integrated application of medicine, epidemiology, data science and public health. Effective surveillance research can help with early warning and response to outbreaks, and ensure public health and social security.

2. Model hypothesis

2.1 Model assumptions

1. Suppose there is no one hiding it up.
2. Suppose that the probability of daily virus infection is equal.
3. Suppose that there is an undetected virus carrier in a population.
4. It is assumed that the virus does not vary during the surveillance period.
5. It is assumed that the errors between the data used in this model and the data from other sources are ignored, with no overall impact.

3. Symbol description

Table 3.1 Conformity and meanings used in this article

symbol	meaning
X_i	The average number of contacts per day in the i th industry
T	Day a Maximum viral incubation period
A	Number of viral nucleic acid monitoring times
N	Total number of tests
t	Total time of detection The earliest number associated with positivity

4. Model building and solution

4.1 Problem Establishment

A test protocol for a population requires analysis of the population and mode of virus transmission. According to global data, the virus is mainly transmitted through human-to-human transmission, and only people who meet at the same time and place are at risk of infection. Therefore, it is crucial to select representative individuals, who have access to most of the group members and timely reflect the infection. Over time, viruses are hidden and difficult to detect early on, so the frequency of detection is also critical. The groups are classified according to industries, and the average number of contacts, complexity, contact time and close contact possibility of each industry are considered, and the weight is determined according to the relevant data information and objective evaluation method, and the comprehensive quality evaluation model is established to determine the testing scheme by scoring.

4.2 Model building

4.2.1 Data preprocessing

Data source: The modeling data are all from the Statistical Yearbook of the number of employees in Liaoning Provincial Bureau of Statistics. We chose Dalian City for research.

Step 1: Download the data source through Liaoning Provincial Bureau of Statistics-Liaoning Provincial Statistical Yearbook, and save and sort out the situation of different individual personnel in Dalian, Liaoning Province by the end of 2020.

Step 2: Screen and integrate the individual data of different industries in Dalian city, Liaoning Province, make the derived data into tables and import them into MATLAB for simulation, and the total score is obtained by using the entropy weight method.

4.2.2 Basic steps for determining the weight by the entropy method

Introduction of entropy weight method

Entropy weight method is a mathematical method to calculate a comprehensive index based on the information provided by each factor. As an objective comprehensive weight-setting method, it mainly determines the weight according to the amount of information transmitted by each indicator to the decision maker. According to the basic principle of information theory, information is a measure of system order; and entropy is a measure of system disorder. Therefore, system entropy can be used to reflect the information provided to the decision maker, and system entropy can be obtained through the entropy weight method.

Basic steps to determining the entropy method:

Select n samples and m indicators, then is the value of the j index of the i sample ($i=1,2,\dots, n; j=1,2,\dots, m$);

Normalization of indicators: homogenization of heterogeneous indicators

Because the measurement units of each index are not unified, so before using them to calculate the comprehensive indicators, they should be standardized first, that is, the absolute value of the index into the relative value, and make, so as to solve the problem of homogenization of different quality index values. Moreover, because the values of positive index (very large index) and negative index (very small index) represent different meanings (the higher the positive index, the better, the lower the negative index, the better), therefore, we use different algorithms for data standardization processing for high and low indicators. The specific methods are as follows:

forward pointer:

$$x'_{ij} = \frac{x_{ij} - \min\{x_{ij}, \dots, x_{nj}\}}{\max\{x_{1j}, \dots, x_{nj}\} - \min\{x_{1j}, \dots, x_{nj}\}} \quad (4.1)$$

Negative indicators:

$$x'_{ij} = \frac{\max\{x_{1j}, \dots, x_{nj}\} - x_{ij}}{\max\{x_{1j}, \dots, x_{nj}\} - \min\{x_{ij}, \dots, x_{nj}\}} \quad (4.2)$$

x'_{ij} Is the value of the j th index of the i th sample ($i=1, 2, \dots, n; j=1, 2, \dots, m$) 。 For convenience, the normalized data still count as x_{ij} 。 Other types of indicators are shown in the figure below, and there is a corresponding processing code in the running program.

3. Calculate the proportion of the i th sample under the j th index:

$$p_{ij} = \frac{x_{ij}}{\sum_{i=1}^n x_{ij}} \quad i = 1, 2, \dots, n; j = 1, 2, \dots, m \quad (4.3)$$

4.The entropy value of the j-th-term index is calculated:

$$e_j = -k \sum_{i=1}^n p_{ij} \ln(p_{ij}) \quad \text{其中 } k = \frac{1}{\ln(n)} > 0, e_j \geq 0 \quad (4.4)$$

5.Information entropy redundancy was calculated:

$$d_j = 1 - e_j \quad (4.5)$$

6.Calculate the weights of each index:

$$w_j = \frac{d_j}{\sum_{j=1}^m d_j} \quad (4.6)$$

7. The composite score was calculated for each sample:

$$s_i = \sum_{j=1}^m w_j \cdot p_{ij} \quad (4.7)$$

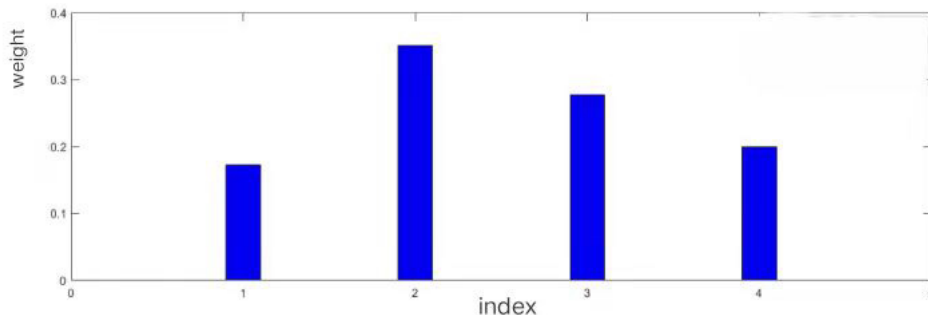
4.2.3 Model solution

Introduction of each indicator:

- (1) The average number of daily contacts in different industries X_{1i} (2) The complexity of the average day X_{2i}
 (3) Mean residence time (i. e., probability of being infected) X_{3i} (4) The possibility of close contact X_{4i}

Then, the above found data is imported into MATLAB for simulation, and the entropy weight method is used to calculate the weight and comprehensive score, so as to determine the final monitoring scheme.

4.2.4The various index weight of the influencing factors



- 1:Average number of contacts per day; 2:The complexity of daily contact with people
 3:Residence time; 4:Possibility of close contact

Figure 4.2 Weight diagram of each indicator

The entropy weight method is used for simulation, as shown in Figure 4.2. It can be seen that the average daily number of contact index accounts for 17.25% of the weight, the average daily complexity index accounts for 35.07% of the weight, the residence time index accounts for 27.73% of the weight, and the close contact possibility index accounts for 19.95% of the weight.

4.3 Interpretation of result

Through the entropy weight method model chart analysis, the proportion of higher index have an average daily contact with complexity and stay time, but in the city (6 million) of different industries, healthcare, public transportation, accommodation and catering industry, education industry influence on the virus, as a result, we need to focus on monitoring the several industries to achieve the effect of controlling the outbreak.

References:

- [1]Mareni, R.W. et al. (2019). Huning virus pathogenesis and viral replication. *Viruses* 11 (4), 319.
 [2]Parks, et al., (2020). Ebola-associated virus replication and ferret disease. *Cell-host and microbial* 27(2),324-334.e4.
 [3]Morrens, D.M., Taubenber, J.K. (2019). Influenza Cataclysm, 1918. *New England Journal of Medicine* 381 (19), 1857-1859.