A risk assessment system of COVID-19 based on Bayesian inference

Jie Wei¹, Yiqiang Li¹ and Yufeng Nie*¹

¹ School of Mathematics and Statistics, Northwestern Polytechnical University, Xi’an, Shaanxi 710129, China

*Corresponding author’s e-mail: yfnie@nwpu.edu.cn

Abstract. The novel coronavirus disease (COVID-19) has now spread to most countries in the world. Preventing and controlling the risk of the coronavirus disease has rapidly become a major concern. A risk assessment system of novel coronavirus disease is proposed based on Bayesian inference in this paper. The system includes multiple handheld terminals and a cloud processing centre. The handheld terminal measures, records, and uploads the individual’s physical information (e.g., body temperature, cough) and GPS information of the terminal. We establish a Bayesian diagnosis network to deduce the risk probability related to the individual’s detection information. The cloud obtains the individual’s detection information and positions in last 14 days, and estimates the epidemic risk probability using Bayesian inference. This probability can be helpful for relevant institutions to judge the individual’s risk levels and corresponding measures. This risk assessment system, which assesses the COVID-19 risk of subjects dynamically, can not only assist and guide the normalization of epidemic prevention and control in relevant institutions, but also assist in epidemiological case tracing.

1. Introduction

In December 2019, the novel coronavirus disease (COVID-19) was first reported in Wuhan City of China [1]. Unfortunately, the epidemic has spread out globally and rapidly, which has brought great challenges to the world’s public health security. According to data from the World Health Organization on May 29, there have been more than 5.77 million confirmed cases of new coronavirus pneumonia worldwide, and more than 360,000 deaths have been caused. More than 200 countries and regions have been affected by the epidemic.

As there is still no specific antiviral agents and vaccines available to treat this new infection, preventing person-to-person transmission measures, such as keeping suitable social distance, family quarantine, and even locking down entire cities to restrict the flow of people, have so far become the main, if not only, choice for many countries [2]. However, these measures are still not sufficient to stop the rapid spread of this coronavirus at a global scale. The epidemic has caused varying degrees of impact on the economies of many countries. How to maintain normal production and living order, ensure smooth transportation and normal flow of personnel, rectify and standardize the order of the market economy, and improve safety in production is the current focus of many countries.

There has been some research on diagnosing Coronavirus Disease 2019 (COVID-19) using Artificial Intelligence. In [3-5], systems to diagnose COVID-19 that utilized conventional neural network (CNN) were developed. However, the system neural network is difficult to interpret, requires large data to train, time and capital intensive and utilizes a lot of memory to execute the network. In [6], an expert system for diagnosing COVID-19 was designed using Clips and Delphi expert system
languages. The system results showcased high detection accuracy, but failed to diagnose respiratory tract diseases with overlapping symptoms with COVID-19. In [7], Bayesian Belief Network (BBN) was utilized in diagnosing COVID-19 with its symptoms. The system had an accuracy of 99.38% in predicting COVID-19 without the influence of the overlapping symptoms.

There has been some research on prediction and risk assessment. In [8], a hybrid SRIMA-WBF model was proposed to deal with the real-time forecasts of the daily COVID-19 cases in five different countries. The risk of COVID-19 was also assessed in [8], and some strategies are affirmed because of their obvious effectiveness. In [9], a method is proposed to dynamically assess the infection risk of ships based a data-driven and machine learning approach.

In order to control the spread of the epidemic, this paper proposes a risk assessment system of COVID-19 based on Bayesian inference. The system estimates the individual’s epidemic risk probability using Bayesian inference, which is based on the individual’s detection information and positions in last 14 days.

After the epidemic progresses to a stable stage, the epidemic prevention work will become normal for practical production and living considerations, which requires that the staff's daily epidemic prevention work should be simplified as much as possible, and the results should be as accurate as possible. The system can collect personal information and whereabouts using a terminal, upload and evaluate personal risks in real time, which can simplify the process. The epidemic prevention and control work is normalized, which will inevitably generate a large amount of data. The evaluation system can evaluate the risk levels of different regions and communities based on big data, guide the accurate prevention and control work, which can help maintain normal production and living order.

The rest of the paper is organized as follows. In Section 2, we discuss the development of the overall model in assessing the infection risks. Sections 3 provides the conclusions and outlines the findings and further work.

2. Overall Framework for risk assessment of COVID-19

2.1. Bayesian Network
The Bayesian network is a directed acyclic graph model, which uses probabilities to illustrate the conditional admissibility that exists between routers on the graph. Bayesian networks (BNs) play a central role in a wide range of automated reasoning applications. Bayesian inference is a statistical inference method in which Bayes’ theorem is used to update the probability of a hypothesis when additional evidence or information becomes available, and it is an important technique in statistics, and especially in mathematical statistics. Bayesian network is based on the Bayes theorem which relies on probability.

The Bayes’ theorem is represented in the mathematical equation below:

\[ P(a|b) = \frac{P(b|a)P(a)}{P(b)} \]  

(1)

Where, \( P(a) \) is the probability of event “a” happening without any information about event “b”. It is called the “Prior”.

\( P(a|b) \) is the conditional probability of event “a” happening given that event “b” has already occurred. It is otherwise called the “Posterior”.

\( P(b|a) \) is the conditional probability of event “b” happening given that event “a” has already occurred. It is called the “Likelihood”.

\( P(b) \) is the probability of event “b” happening without any information about event “a”. It is called the “Marginal Likelihood”.

Bayesian inference is an important inference method in artificial intelligence, especially in probability prediction. Bayesian inference is the process of using Bayes’ theorem to derive attributes about populations or probability distributions from data. Bayesian reasoning does not need to store a lot of data all the time, but constantly updates the previous probabilities based on observations. In fact,
the Bayesian framework updates the probability in real time while processing the data, which has obvious benefits when dealing with big data.

2.2. Overall Framework

The main procedure used in this system is described in Figure 1. As shown, a risk assessment system of COVID-19 is composed of multiple handheld terminal devices and a cloud processing centre. The handheld terminal measures, records, and uploads the individual’s physical information (e.g., body temperature, coughing) and GPS information of the terminal. The cloud obtains individual’s detection information and positions in last 14 days, and estimates the individual’s epidemic risk probability using Bayesian inference.

![Figure 1. The main procedure of the risk assessment system.](image)

The system records the individual’s physical through the handheld terminals, records the individual’s recent whereabouts information, and assesses the individual’s current epidemic risk in the cloud processing centre. In addition, the system can assist in epidemiological investigations and provide basic data for the formulation and improvement of future vaccination and epidemic prevention and control strategies.

The overall Bayesian Inference Network for risk assessing is described in Figure 2. The basis of reasoning mainly includes three layers: the epidemic situation in the current region, the physical information of the individual, and the place where the individual has recently visited. The epidemic situation in the region should be derived from official data. The key data is the total population, the number of confirmed and suspected COVID-19 population in the region. In addition, according to the diagnosed and suspected population distribution, the risk probabilities of different fragmented regions can be obtained. Pay attention to privacy violations here. The physical information of the individual only considers the records of the current detection time. The Bayesian diagnostic network (Figure 3) was used to infer the probability of risk from the individual’s physical information. The positions of the individual in the last N days that stored in the cloud and the information of the COVID-19 risk location in the region should be integrated to calculate the risk probability.
3. Methods for risk assessment

3.1. risk assessment based on physical information
According to the diagnosis network in [7] and the quick detection, we establish the Bayesian diagnosis network as shown in Fig. 3. In [10], it was stated that the symptoms of this disease are fever, cough, sore throat, tiredness, aches, sneezing, nasal congestion, malaise, diarrhoea, septic shock, pneumonia and difficulty in breathing. In a quick detection, the information we can obtain quickly and conveniently includes: body temperature, which can be measured by contact or non-contact infrared equipment, cough and sneezing, which can be obtained by observation. For further requirement, the inspector can obtain information such as whether the individual has a tiredness, dyspnoea or sore throat through consultation.

Based on the information given in [1], which introduced clinical features of patients infected with COVID-19 in Wuhan, China, we set up the diagnosis network shown in Figure 3. Based on the established Bayesian network, the epidemiological risk probability of the subjects can be inferenced.
Assuming that the individual’s detection information collected by the terminal inspector includes fever, cough, sneezing, aches, tiredness, dyspnoea, and sore throat. Denote it as \( \text{info} = \{ \text{fever, cough, sneezing, aches, tiredness, dyspnoea, sore throat} \} \), in which each variable is a Boolean.

According to the individual’s detection information \( \text{info} = \{ \text{fever, cough, ...} \} \), according to the Bayesian diagnosis network in Figure 3, the probability \( p_0 \) of epidemic risk based on the individual’s detection information can be calculated according to the Bayesian diagnosis network.

According to the data and analysis given in [1], we range the probability of fever on condition of COVID-19 \( p(\text{fever} | \text{COVID} - 19) \) from \([0.9, 1.0]\), the probability of cough on condition of COVID-19 \( p(\text{cough} | \text{COVID} - 19) \) from \([0.6, 0.8]\); the probability of tiredness on condition of COVID-19 \( p(\text{tiredness} | \text{COVID} - 19) \) from \([0.4, 0.5]\); the probability of dyspnoea on condition of COVID-19 \( p(\text{dyspnoea} | \text{COVID} - 19) \) from \([0.5, 0.6]\); the probability of sneezing, aches, and sore throat on condition of COVID-19 ranges from \([0.01, 0.1]\). The probability of COVID-19 \( p(\text{COVID} - 19) \) is the overall disease risk probability of the region, and the value is calculated by the official data. The probability of fever \( p(\text{fever}) \) is the probability of fever in the region, which can be calculated based on local medical statistics. Similar to the probabilities \( p(\text{cough}), p(\text{sneezing}), p(\text{aches}), p(\text{tiredness}), p(\text{dyspnoea}) \) and \( p(\text{sore throat}) \).

The probability of COVID-19 \( p(\text{COVID} - 19) \) should be calculated based on local official epidemic prevention data. The formula is as follows:

\[
\frac{N_{\text{COVID} - 19}}{N_\Omega} 
\]

Where \( N_\Omega \) is the number of people detected in the region, and \( N_{\text{COVID} - 19} \) is the number of COVID-19 diagnosed and suspected.

The probabilities \( p(\text{fever}), p(\text{cough}) \) can be updated iteratively. The update formula is as follows:

\[
p(\text{fever}) = \frac{N_{\text{fever}}}{N_\Omega} \quad (3) 
\]

\[
p(\text{cough}) = \frac{N_{\text{cough}}}{N_\Omega} \quad (4) 
\]

Where \( N_\Omega \) is the total number of people detected in the region; \( N_{\text{fever}} \) is the number of fevers detected in the region; \( N_{\text{cough}} \) is the number of coughs detected in the region.

According to the above information, the risk probability of COVID-19 in the case of fever and cough can be obtained. Calculated as follows:

\[
p(\text{COVID} - 19 | \text{fever}) = \frac{p(\text{fever} | \text{COVID} - 19)p(\text{COVID} - 19)}{p(\text{fever})} \quad (5) 
\]

\[
p(\text{COVID} - 19 | \text{cough}) = \frac{p(\text{cough} | \text{COVID} - 19)p(\text{COVID} - 19)}{p(\text{cough})} \quad (6) 
\]

Similar to the probabilities \( p(\text{COVID} - 19 | \text{sneezing}), p(\text{COVID} - 19 | \text{aches}), p(\text{COVID} - 19 | \text{tiredness}), p(\text{COVID} - 19 | \text{dyspnoea}) \) and \( p(\text{COVID} - 19 | \text{sore throat}) \).

Finally, the probability \( p_{\text{detection}} \) of the epidemic risk based on the individual’s detection information can be obtained.

\[
p_{\text{detection}} = p(\text{COVID} - 19 | \text{info}) \quad (7) 
\]

The probability \( p_{\text{detection}} \) of the epidemic risk can be obtained in the CPT, or we can calculate it using the formula as follows. According to the relevance of symptoms, we sort the information in \( \text{info} \), as \( \{ \text{fever, cough, dyspnoea, tiredness, sneezing, aches, sore throat} \} \), and denote the symptoms as \( \text{sym}_i (1 \leq i \leq 7) \).

\[
p(\text{COVID} - 19 | \text{info}) = \frac{p(\text{COVID} - 19 | \text{sym}_1)}{p(\text{COVID} - 19 | \text{sym}_1)} \quad (8) 
\]

Denote the risk probability of COVID-19 on condition of the first two symptoms as:
\[ p(\text{COVID} - 19 | \text{fever}_{\text{info}}, \text{cough}_{\text{info}}) = p(\text{COVID} - 19 | \text{sym}_1, \text{sym}_2) = (1 - p(\text{COVID} - 19 | \text{sym}_1))p(\text{COVID} - 19 | \text{sym}_2) \quad (9) \]

And so on, then

\[ p_{\text{detection}} = p(\text{COVID} - 19 | \text{info}) = p(\text{COVID} - 19 | \text{fever}_{\text{info}}) + (1 - p(\text{COVID} - 19 | \text{fever}_{\text{info}}))p(\text{COVID} - 19 | \text{cough}_{\text{info}}) + \cdots \]

\[ = p(\text{COVID} - 19 | \text{sym}_1) + (1 - p(\text{COVID} - 19 | \text{sym}_1))p(\text{COVID} - 19 | \text{sym}_2) + \cdots \quad (10) \]

3.2. Risk assessment based on position information

The next step is to obtain the risk probability for the whereabouts of individual. The GPS location of the current terminal is uploaded to the cloud and identified as the individual’s current whereabouts point \( m_0 \). Combined with the individual’s past whereabouts on the cloud \( \{m_1, m_2, \cdots \} \) within 14 days, the whereabouts of the individual in the last 14 days \( \{m_{10}, m_{11}, m_{12}, \cdots \} \) is obtained.

The position data mainly takes the positions over the past 14 days into account, because the generally accepted incubation period of COVID-19 is 14 days.

Assuming that the risk locations \( \{a_1, a_2, \cdots, a_n\} \), the safety distances \( \{l_1, l_2, \cdots, l_n\} \) and risk weights \( \{w_1, w_2, \cdots, w_n\} \) in the region have been calculated using the official data. For the location \( m_i \), calculate the probability \( p_{ij} = f(m_i, a_j, w_j) \) of the individual’s position \( m_i \) and the risk point \( a_j \) in the region.

\[ f(m_i, a_j, w_j) = \begin{cases} 
0, & \text{where } l(m_i, a_j) \geq l_j \\
\frac{c_i \cdot w_j}{\ln(l(m_i, a_j)^{-1})}, & \text{where } l(m_i, a_j) < l_j
\end{cases} \quad (11) \]

Here, \( c_i \) is a coefficient to reflect the attenuation of the time series’ impact on the current risk, and decreases as \( i \) increases. Here, we take \( c_0 = 1, c_i = \frac{1}{i} \).

For the location \( m_i \), calculate the Epidemic risk probability \( p_i \) as follows.

\[ p_i = \sum_j f(m_i, a_j, w_j) \quad (12) \]

For the individual, the position epidemic risk probability \( p \) can be calculated as follows:

\[ p_{\text{position}} = \sum_i \sum_j f(m_i, a_j, w_j) \quad (13) \]

3.3. The total probability

Combining the signs and whereabouts of the subject, we give the total probability of epidemic risk \( p \) as follows.

\[ p = p_{\text{detection}} + p_{\text{position}} \quad (14) \]

3.4. Some consideration in the system

According to official epidemic data and medical advice, the risk locations \( \{a_1, a_2, \cdots, a_n\} \) and the safety distances \( \{l_1, l_2, \cdots, l_n\} \) and risk weights \( \{w_1, w_2, \cdots, w_n\} \) in the region should be given carefully and updated regularly.

In order to effectively prevent a resurgence in the pandemic and resume work and production smoothly, the Alibaba Group has developed a tracing health QR (Quick Response) code system to identify different degrees of infection risks based on people’s daily activities and movements. Combined with QR, the probability can divide the risk into three levels, corresponding to different epidemic prevention strategies.
4. Conclusion
Preventing and controlling the increasing severe risk of COVID-19 has currently become one of the main concerns of many countries when taking measures to protect their citizens and to restart the economy. A risk assessment system of novel coronavirus is proposed based on Bayesian inference in this paper. The system collects the physical information of the subject based on the handheld terminal, and estimates the epidemic risk probability using Bayesian inference. This risk assessment system, which assesses the COVID-19 risk of subjects dynamically, can not only assist and guide the normalization of epidemic prevention and control in relevant institutions, but also assist in epidemiological case tracing.

However, there are still many directions to improve the approach both at the data and methodological levels. The region’s risk locations and the safety distances, which should be given carefully and updated regularly, can be also inferenced by the system. But the method should be rebuilt. Posteriors in the system should be learned and updated by big data.

Acknowledgments
The authors gratefully appreciate the support by National Natural Science Foundation of China (grant number 11971386).

References
[1] Huang C, Wang Y, Li X, et al. (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The lancet, 395(10223): 497-506.
[2] Shaw R, Kim Y, Hua J. (2020) Governance, technology and citizen behavior in pandemic: Lessons from COVID-19 in East Asia. Progress in disaster science, 2020: 100090.
[3] Butt C, Gill J, Chun D, et al. (2020) Deep learning system to screen coronavirus disease 2019 pneumonia. Applied Intelligence, 2020: 1.
[4] Narin A, Kaya C, Pamuk Z. (2020) Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks. arXiv preprint arXiv:2003.10849.
[5] Jin C, Chen W, Cao Y, et al. (2020) Development and Evaluation of an AI System for COVID-19 Diagnosis[J]. medRxiv.
[6] Salman F M, Abu-Naser S S. (2020) Expert System for COVID-19 Diagnosis. International Journal of Academic Information Systems Research (IJAISR), ISSN: 2643-9026, Vol. 4, Issue 3, March – 2020, Pages: 1-13.
[7] Osarumwense A S, Osayamen O K. (2020) A CoronaVirus Disease-2019 Prediction Model Based on Bayesian Belief Network.
[8] Chakraborty T, Ghosh I. (2020) Real-time forecasts and risk assessment of novel coronavirus (COVID-19) cases: A data-driven analysis. Chaos, Solitons & Fractals, 2020: 109850.
[9] Wang Z, Yao M, Meng C, et al. (2020) Risk Assessment of the Overseas Imported COVID-19 of Ocean-Going Ships Based on AIS and Infection Data. ISPRS International Journal of Geo-Information, 2020, 9(6): 351.
[10] Cascella, M., Rajnik, M., Cuomo, A., Dulebohn, S.C. and Napoli, R.D. (2020): Features, Evaluation and Treatment Coronavirus (COVID-19). StatPearls Publishing, Treasure Island, FL; 2020. Retrieved 6th April 2020 from URL: https://www.ncbi.nlm.nih.gov/books/NBK554776/.