A Dynamic Case-based Reasoning System for Responding to Infectious Disease Outbreaks

Jinli Duan[a,*], Zhibin Lin[b], Feng Jiao[c], Yixian Jiang[d], Kexing Chen[e]

[a] College of Modern Management, Yango University, Fuzhou, China

[b] Durham University Business School, Mill Hill Lane, Durham DH1 3LB, United Kingdom

[c] INTO Newcastle University, Newcastle University, Newcastle upon Tyne, NE1 7RU, United Kingdom

[d] Public Health School of Fujian Medical University, Fuzhou, China

[e] Fuzhou Center for Disease Control and Prevention, Fuzhou, China

Abstract

Infectious diseases are a global public health problem, which requires timely and effective responses. This study proposes a novel model that contributes to the development of such responses. First, the problem scenario features of infectious disease emergency scenarios are extracted, and the problem scenario is structurally described. A Markov model is adopted to analyze the scenario evolution of the infectious disease outbreaks. Then, a dynamic case-based reasoning model is built. Different matching algorithms are designed for crisp symbols, crisp numbers, interval numbers, and fuzzy linguistic variables. The similarity between the target scenario and various historical scenarios is calculated. Finally, an optimized dynamic emergency decision guide is provided. An experiment is conducted to test the validity and feasibility of the proposed method. The results suggest that the model can

* Corresponding author.

E-mail address: 78308776@qq.com (J. Duan), zhibin.lin@durham.ac.uk (Z. Lin), feng.jiao@newcastle.ac.uk (F. Jiao), 371466977@qq.com (Y. Jiang), kexing6855@yeah.net (K. Chen).

realistically simulate the process of infectious disease outbreaks and quickly match the recorded scenarios to generate effective and real-time responses.

Keywords: Markov model, Case-based reasoning, Emergency response, Infectious disease, Covid-19, SARS.

1. Introduction

Since the beginning of the 21st century, public health emergencies have occurred frequently all over the world, such as SARS in 2003, human-induced avian influenza in 2004, H1N1 influenza in 2009, and currently the COVID-19 pandemic. These major infectious disease outbreaks not only seriously endanger human health and life, but also social and economic development. With the acceleration of urbanization, modernization, and globalization, the continuous emergence of new infectious diseases poses great threats to humankind. Public health decision-makers face scenarios that go beyond their experience, because new infectious disease outbreaks are highly uncertain, and the evolution could be very complex. Therefore, timely and effective responses to major infectious diseases are of critical importance, since once the best chance of mitigation is missed, an outbreak may become uncontrollable.

Emergency decision-making is usually based on case-based reasoning (CBR). The CBR systems incorporate a knowledge base that encapsulates experiences which are a set of problems previously solved within the system. They are used to solve new problems by recalling information representing previous similar situations and by reusing information and knowledge from these situations. In recent years, several researchers have attempted to combine CBR with other intelligent algorithms and have achieved significant breakthroughs in certain areas such as clinical medicine (Ebrahimian et al., 2018; Szyszkowicz et al., 2018) and public health (Araz & Jehn, 2013; Exum et al., 2018).

Previous studies have focused on improving dynamic emergency decision-making, given the high uncertain evolution of infectious disease emergencies. Jiang et al. (2011) proposed an emergency response solution adjusted according to the emergency effect, loss, and cost of each stage of emergency. They suggested adjusting emergency plans according to a comparison between decision-makers' psychological reference points and the real-time loss. Wang and

Wang (2013) built a sequential dynamic game model, where the dynamic game is carried out according to the time sequence of the factors influencing the evolution of the emergencies to provide effective emergency measures. Yang and Xu (2012) generated emergency plans according to a dynamic game between "scenario" and "solution" under the framework of a dynamic game model. Jie et al. (2005) revised the emergency decision-making according to the development of the situation. Chen et al. (2010) suggested the selection of a scheme with the highest comprehensive evaluation value as an emergency plan according to expected profit and loss, to deal with emergencies.

The above studies and others in the extant literature provide a basis for exploring the spread and evolution of major infectious diseases. However, there are at least two major deficiencies that have yet to be addressed: (1) The literature lacks a clear definition and structural description of infectious disease events as a multi-attribute set; (2) Currently, it is difficult or impossible to develop effective dynamic emergency measures, due to the lack of simulations of the structure and dynamic process of epidemic transmission.

This study thus aims to narrow the above two research gaps by proposing a novel model. First, we propose a structured description framework, extract the attributes of infectious disease emergencies, and divide them into a set of event types, key attributes, environmental attributes, and hazard assessment attributes. Second, we use a Markov model in addition to the general CBR model, to analyze the dynamic evolution process of each whole infectious disease outbreak event randomly, and generate effective emergency decision-making aids. We conduct an experiment to test our proposed model using the COVID-19 pandemic as the target case and selected eight related infectious disease emergencies between 2000 and 2019 in China as historical cases.

The remainder of this paper is structured as follows. Section 2 analyzes scenario evolution in different stages of infectious diseases to extract the attributes of the problem

scenario. Section 3 presents a model based on a Markov chain and the CBR method. Section 4 applies the methodology through an experiment, and presents the computational results. Finally, Section 5 concludes and suggests some possible future research directions.

2. Structural description of an infectious disease emergency

2.1. Attributes extraction of infectious diseases emergency

An infectious disease emergency is a set of multi-attribute events, which is difficult to describe in a single index. As such, we first attempt to extract the internal and external attributes of events to establish a more comprehensive attribute set for the problem scenarios to analyze the evolution process, which lays the foundation for our proposed decision-making model (Liu et al., 2013). These attributes are abstracted into a set of four attributes according to the extent to which they affect the evolution of events: event type, key attribute, environmental attribute, and hazard assessment attribute. The problem scenario attributes for infectious disease emergencies are shown in Table 1.

Table 1. Problem scenario attributes of infectious disease emergency

Attributes	Meanings of the attributes	Formats of attribute values
Event Type (ET)	Gradual change type or drastic change type (TP)	Crisp symbol
Key Attributes (KA)	Infectivity (IC)	Fuzzy linguistic variable
	Transmission Routes (TR)	Crisp symbol
	Fatality Rate (FR)	Crisp number
	Susceptible Populations (SP)	Crisp symbol
	Weather Characteristics (WC)	Crisp symbol
Environmental Attributes (EA)	Population Mobility (PM)	Fuzzy linguistic variable
	Population Density (PD)	Interval number
Hazard Assessment Attributes (HAA)	Number of Infected People (NIP)	Crisp number
	Death Toll (DT)	Crisp number
	Scope of Influence (SI)	Fuzzy linguistic variable

2.2. The state space transmission of infectious disease emergency

We divide the process of infectious disease transmission into eight state spaces (shown in Table 2), based on the characteristics of the transmission and evolution processes of infectious disease emergencies.

Table 2. State space of infectious disease emergency transmission

State of infectious disease emergency transmission	Description
T_1	The outbreak of infectious diseases
T_2	No restrictions on population flow
T_3	The latent period turned to the onset period
T_4	No isolation and treatment timely
T_5	Virus detection and screening
T_6	Isolation and treatment
T_7	Recovery with immunity and death
T_8	The end of the infectious disease crisis and the treatment of the aftermath

2.3. Analysis of infectious disease emergency evolution

The spread and evolution process of infectious diseases is analyzed, based on the structural description of infectious disease emergencies and the analysis of related attributes of related events. When an infectious disease outbreak begins, the pathogen can be expected to spread according to a certain path: vulnerable populations become the epidemic diffusion power within the stochastic system model; while potential patients and the sick patients imported from outside the outbreak area become the diffusion power outside the system; because of the flow of populations, exposed individuals get sick and cause rapid spread of the epidemic situation. If there is no timely isolation and effective treatment, a large number of people may die, leading to panic among the public and endangering social security.

The evolution process of the spread of infectious diseases is a very complex system, and Markov models are suitable simulation modeling tools for complex systems. The state

transition matrix in Markov models is a transformation process between the internal states of the system. We therefore adopt a Markov model to systematically analyze the evolution of infectious diseases. This approach allows a CBR model for the dynamic management of infectious disease emergencies. We cite the conceptual model created by Qiao et al. (2016) which contributes a prototype of the Markov Chain model in simulating the outbreak of infectious disease. Based on the characteristics and relationships of transmission states, an evolution process model for an infectious disease emergency can be depicted in Figure 1, which presents the evolution process from the outbreak to death.

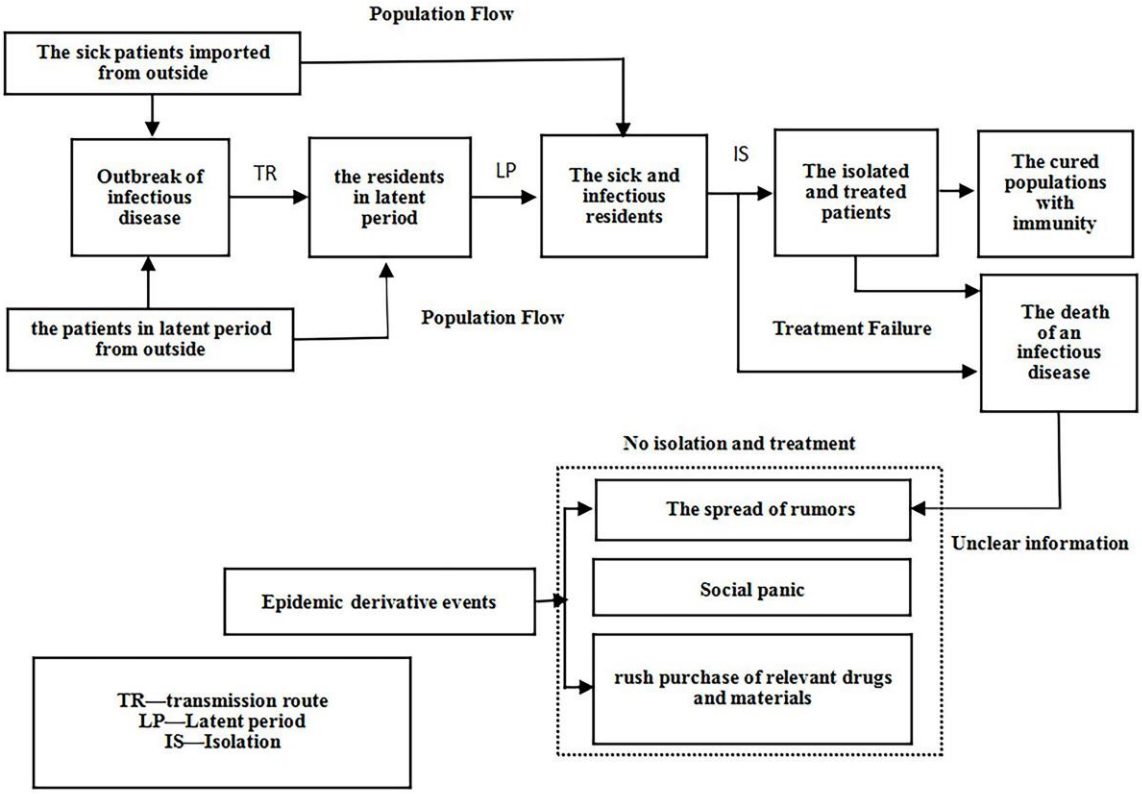

Figure 1. Evolution process of infectious disease transmission

3. The model

The proposed model consists of four stages. First, a Markov model is used to simulate the evolution process of infectious disease emergencies. Second, the similarity between current and historical scenarios is calculated. Third, similar case sets are retrieved, and the specific

emergency measures are modified according to the local medical conditions and the supply of relevant materials. Fourth, the next stage is predicted by the Markov chain of state space transmission, and the similarity between the predicted and historical scenarios is calculated. The corresponding remedial and revision measures are determined, based on the degree of similarity and the threshold.

3.1. Markov model

Markov models have been widely used in various fields, such as medicine, engineering, and management for the transformation and evolution of stochastic states. For an infectious disease transmission network, many epidemic processes in the network are realized by random contact between connection states. Markov models are well suited for the prediction and inference in the setting of state space models because of their fast and reliable numerical performances. Several studies have applied Markov models to the task of simulating the evolution of infectious diseases (Bartolucci et al., 2014; Blasone et al., 2008; Haeussler et al., 2018; Ludkovski & Lin, 2013). For instance, Gómez et al. (2010) used the Markov chain method to propose a discrete-time formula to solve the problem of contact-based epidemic transmission. O'Neill et al. (2000) explored how to use the Markov chain and Monte Carlo through modeling and analyzing measles and influenza outbreak data. Didelot et al. (2017) applied a reversible jump Monte-Carlo Markov Chain to both simulated data and real data from an outbreak of tuberculosis in a public health environment during real-time outbreak investigations. Evan (2019) proposed a discrete time Markov Chain model which is applied to approximate the probability of disease extinction in an epidemic model. Endo et al. (2019) established the particle Markov-chain Monte Carlo (PMCMC) model to efficiently explore high-dimensional parameter space using time-series data to simulate the evolution of an infectious disease outbreak. Tada et al. (2020) used Markov chain models to analyze liver disease in patients with hepatitis C virus state transition probability matrices.

Markov model is a stochastic process that moves in a sequence of phases through a set of states and has a memory of only one state. This means that the probability of entering a certain state in a certain phase is not necessarily independent of previous states, only depends on the nearest state in the previous phase. This property is known as the Markov property. The Markov chain can be used to simulate spread paths and trends. The application universality, discretization and randomness of the Markov model meet the simulation requirements of infectious disease transmission state transition in our study.

In the historical case base, the state of infectious disease emergency is denoted as T_n *(n=1,2,...)*. Suppose that the probability from the state of T_i to T_j is $p_{i,j}$. According to the process of infectious disease emergency, the state space can be divided into *n* states, where the first state is an outbreak of infectious diseases, and the last state is the end of infectious disease crisis and the treatment of aftermath.

The transition rate between two adjacent states is $p_{i,j}$. The infectious disease emergency state transition matrix can be written as,

$$p = \begin{pmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,n} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,n} \\ \cdots & \cdots & \cdots & p_{3,n} \\ p_{n,1} & p_{n,2} & \cdots & p_{n,n} \end{pmatrix} \tag{1}$$

There is a Markov Chain $\langle T, p, \alpha \rangle$, where T is the state space, p is the probability of transition from one state to another, and a is the initial state. After determining the range of T, there are many estimation methods for p. We use the maximum likelihood estimation (MLE) estimation p.

$$p_{ij} = \frac{C(T_i, T_j)}{\sum_{r=1}^{n} C(T_i, T_r)} \tag{2}$$

where $C(T_i, T_j)$ represents the number of transition times from the state T_i to the state T_j in the historical case base. According to the historical case base, the infectious disease state space transmission Markov chain is shown in Figure 2.

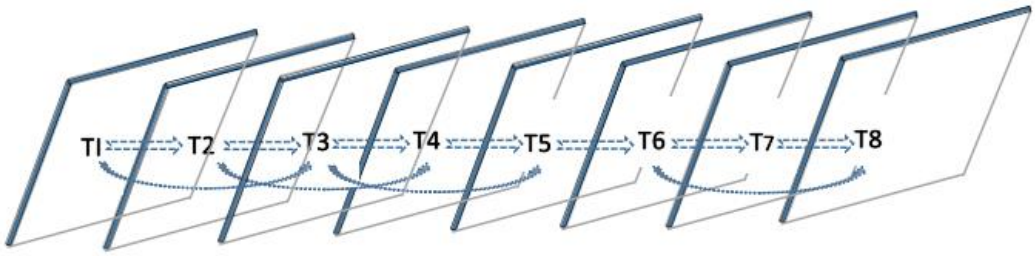

Figure 2. Markov chain of state space transmission in infectious diseases emergency

Specifically, based on the prototype in Figure 1, an eight-state scenario simulating major infectious disease outbreaks is developed. This simulation presents the basic transformation relationship between states which is the process chain from T1 to T2, from T2 to T3, from T3 to T4, from T5 to T6, T6 to T7, and from T7 to T8, respectively. According to the actual transformation process of 8 states in the historical case base, Figure 2 is obtained. For example, there are two possible transformations of state T1, one is to T2, and the other is to T3.

According to the infectious disease state space transmission Markov chain, the state transition matrix is obtained. Specifically, based on the formula (2) of the state transmission matrix in the Markov Chain model, if T_1 transits to T_2, and T_1 transits to T_3, the possibility of T_1 transiting to T_2 is $p_{12} = \frac{1}{2}$, and the possibility of T_1 to T_3 is $p_{13} = \frac{1}{2}$. In this regard, Equation (3) can be obtained as the following.

$$p = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{3}$$

3.2. Calculating the similarity

The case base is composed of four types of data: crisp symbols, crisp numbers, interval numbers, and fuzzy linguistic variables.

3.2.1. Local similarity

There are four common formats of attribute values of a problem scenario, i.e., crisp symbols, crisp numbers, interval numbers, and fuzzy linguistic variables. The four formats of attribute values are expounded as follows (Fan et al., 2014).

For the attributes of crisp symbols, the attribute values are the kind of enumeration values, there are no quantitative relationships among the attribute values. In other words, these kinds of attribute values cannot be compared. In this case, the attribute similarities of the problem scenario between the historical cases and the target case can be evaluated by judging whether the attribute values are equivalent to each other. Given the above analysis,

Let $Sim_k(S_0, S_i)$ denote the attribute similarity of problem scenario between the historical case S_i and target case S_0 about attribute k, then the calculation formula of the similarity is given by,

$$Sim_k(S_0, S_i) = \begin{cases} 1, & f_{ik} = f_{0k} \\ 0, & f_{ik} \neq f_{0k} \end{cases} \tag{4}$$

For the factors of crisp numbers, we calculate the similarity of factors utilizing distance, and the formula is,

$$Sim_k(S_0, S_i) = \exp\left(-\frac{\sqrt{(f_{0k} - f_{ik})^2}}{\max\left(\sqrt{(f_{0k} - f_{ik})^2}\right)} \right) \tag{5}$$

The $\max\left(\sqrt{(f_{0k} - f_{ik})^2}\right)$ represents the maximum value of the difference between the attributes value of the target scenario and the value of the factors of the historical case scenario.

For the factors of interval numbers, the attribute value of the target scenario is donated as $\left[f_{0k}^l, f_{0k}^u \right]$, and the attribute value of the historical case scenario is donated as $\left[f_{ik}^l, f_{ik}^u \right]$. The calculation of the similarity is given by,

$$Sim_k(S_0, S_i) = \exp\left[-\frac{\sqrt{\left(f_{0k}^l - f_{ik}^l\right)^2 + (f_{0k}^u - f_{ik}^u)^2}}{\max\left(\sqrt{\left(f_{0k}^l - f_{ik}^l\right)^2 + (f_{0k}^u - f_{ik}^u)^2}\right)} \right] \tag{6}$$

For the factors of fuzzy linguistic type, the linguistic variable can be represented by the triangular fuzzy number. The attribute value of the target scenario is denoted as $\tilde{f}_{0k} = \left(f_{0k}^a, f_{0k}^b, f_{0k}^c \right)$, and the attribute value of the historical case scenario is donated as $\tilde{f}_{ik} = \left(f_{ik}^a, f_{ik}^b, f_{ik}^c \right)$. The calculation of the similarity is given by:

$$Sim_k(S_0, S_i) = \exp\left[-\frac{\sqrt{\left(f_{0k}^a - f_{ik}^a\right)^2 + (f_{0k}^b - f_{ik}^b)^2 + (f_{0k}^c - f_{ik}^c)^2}}{\max\left(\sqrt{\left(f_{0k}^a - f_{ik}^a\right)^2 + (f_{0k}^b - f_{ik}^b)^2 + (f_{0k}^c - f_{ik}^c)^2}\right)} \right] \tag{7}$$

3.2.2. Global similarity

In this study, an attribute weight determination algorithm based on coverage is adopted. Suppose U specifies a case base, A is the attribute set of all cases in the case base, S_i is the i^{th} case in the case base, $i = 1, 2, \cdots, n$, and q_j is the j^{th} attribute in the case attribute set $j = 1, 2, \cdots, m$.

The following is the definition of case coverage. The coverage of the case S_i represents a set of cases that meet this condition $Sim(S_i, S_j) \geq \alpha$, where α is the similarity threshold.

$$Cover(S_i) = \left\{ S_j \mid Sim(S_i, S_j) \geq \alpha, S_j \in U \right\} \tag{8}$$

where $Sim(S_i, S_j) = \dfrac{\sum_{k=1}^{m} \omega_k Sim_k(S_i, S_j)}{\sum_{k=1}^{m} \omega_k}$, ω_k is the attribute weight to be determined, α is the similarity threshold. The steps of the attribute weight determination algorithm based on coverage are given by,

Step (1). Assign initial values $\omega_k = \dfrac{1}{m}$ to all case attribute weights in the case base U, $\sum_{k=1}^{m} \omega_k = 1$, m is the number of the attributes.

Step (2). Calculate the coverage C_{0i} of S_i, when S_i includes all attributes. Let $k = 1$;

Step (3). Remove the attribute q_k in the attribute set and recalculate the coverage C_{ki} of S_i;

Step (4). If $k \geq m$, go to the next step; otherwise, let $k = k + 1$, return Step (3);

Step (5). If $i \geq n$, go to the next step; otherwise, let $i = i + 1$, return Step (2);

Step (6). Adjust the weight ω_k of attribute q_k, according to the influence degree of each attribute in different cases on the average case coverage;

$$\omega_k' = \frac{\left|\frac{1}{n}\sum_{i=1}^{n} C_{0i} - \frac{1}{n}\sum_{i=1}^{n} C_{ki}\right|}{\frac{1}{n}\sum_{i=1}^{n} C_{0i}}, \quad k=1,2,..,m \tag{9}$$

Step (7). Normalize the weight ω_k' to get the new weight ω'' of each attribute;

$$\omega_k'' = \frac{\omega_k'}{\sum_{k=1}^{m} \omega_k'}, \quad k = 1,2,...,m \tag{10}$$

The idea of the algorithm based on coverage is to assign more weight to those case attributes which have an important impact on the retrieval effect, that is, when the attribute is removed, if the number of similar cases retrieved is very different from the number of cases retrieved before dropping, the attribute will be given greater weight (Chan et al., 2019; Humensky et al., 2020).

3.3. Revising corresponding solutions

The consequences of the outbreak of infectious diseases are very serious, facing the loss of people's lives and health. Therefore, the corresponding prevention and control measures must be operable and effective. The solution based on a similar historical scenario directly to the target scenario does not take into account the medical and public health conditions in the outbreak area and the actual effect of the measures from similar cases set (Choi et al., 2020; Wang et al., 2020). So, the revisions need to be made by public health experts.

3.4. Dynamic prediction of the next scenario

The 8- stage model of infectious disease outbreak is obtained based on the simplification of Figure 1; thus, Table 2 and Figure 2 are presented to illustrate the transformation relationship between these 8 stages. Based on the Markov Chain of state space transmission in infectious diseases in Figure 2, we get the state transition matrix, that is Equation 3. With the probability

of transition, we can predict the next scenario S_0', $S_0' = \sum_{i=1}^{8} p_{0i}.S_i$. Then proceed to the following steps to adjust and update the next prevention and control measures. The specific steps are as follows:

Step (1). Calculate the similarity $Sim(S_0', S_i)$, $i \in n$ between the predicted next scenario S_0' and the historical scenario base case;

Step (2). Retrieve a similar scenario S_k, $k \in n$ in the historical case base to the scenario S_0';

Step (3). If $k = i$, it shows that the current corresponding solution can completely control and cope with the situation at scenario S_0' without any compensatory remedies.

If $k \neq i$, and $Sim(S_0', S_k) - Sim(S_0, S_k) \leq \delta$, it indicates that the decision-maker thinks that the current scheme is acceptable and still valid for the current scenario. Generally, the value δ is given by the experience of decision-makers, the range of similarity value and the cost of the adjustment scheme. The smaller δ is, the higher the requirement of similarity is.

If $k \neq i$ and $Sim(S_0', S_k) - Sim(S_0, S_k) > \delta$, it indicates that according to the experience of decision-makers, the current corresponding solution cannot control the current scenario, stop the current scheme and execute the new scheme corresponding to the scenario S_k which is the most similar to the scenario S_0'.

4. Experiment

Eight infectious diseases that occurred in China with great impact were selected as the historical cases, and COVID-19 was taken as the target case to verify the application of the proposed method.

4.1. The cases

To ensure the timeliness of the case, we used COVID-19 as the target case and selected eight related infectious disease emergencies between 2000 and 2019 in China as historical cases. There are hepatitis A in Shanghai in 1988, SARS in 2003, Influenza A (H1N1) in 2004, Influenza A (H1N1) in 2009, Hand Foot and Mouth Disease in 2009, Influenza A (H5N1) in 2011, Influenza A (H7N9) in 2013, and Influenza A (H7N9) in 2017. We apply actual data from 9 outbreaks of infectious diseases in China using Equations (4), (5), (6), and (7) to calculate the similarity of 11 attributes between COVID-19 and 8 historical cases of infectious diseases, then Table 3 is obtained. The actual data is collected from China Health Statistical Yearbook and National Population Health Data Centre, China. This part of the information can be accessed through the following link (The Data Center of China Public Health Science, 2021): https://www.phsciencedata.cn/Share/edtShareNew.jsp?id=39205.

Using the two cases of the SARS in 2003 and Influenza A (H1N1) in 2009, the general framework of structural description of major infectious diseases is obtained through case verification based on event introduction, event sorting and application of this structured description method.

The SARS incident is a global epidemic of infectious diseases, which first broke out in Shunde, Guangdong Province on November 16, 2002, and rapidly spread out to Southeast Asia and other parts of the world. The epidemic was gradually eliminated by the middle of 2003. The nature and grade of SARS are mainly affected by the source of infection, infectivity, transmission route, mortality rate, and characteristics of the susceptible population. The scenario attributes, such as information disclosure, population mobility, the scope of influence, population density, and prevention and control measures affect whether the incident continues to escalate. Consequently, ‘SARS’ is structured as {gradual change type}, {SARS virus infected persons, extremely infectious, close contact infection (close air transmission), the mortality rate was 9.30%}, and the susceptible population was the contact patients (e.g. medical

personnel, family members of patients)}, {poor information disclosure}, {strong population mobility (especially during the Spring Festival transportation)}, {prevention and control measures are disinfection and wearing protective equipment}, {the weather from cold to hot makes the virus more easily spread}, {the scope of the impact is global}, and {the population density in severe epidemic areas is large}.

Influenza A (H1N1) is an acute respiratory infectious disease caused by a new H1N1 virus, which is highly infectious. In March 2009, a large number of people died after the outbreak of "human infection with swine influenza" in Mexico. Then it spread rapidly to many regions and countries around the world till August 2010, when the World Health Organization announced that the pandemic of Influenza A (H1N1) had ended. The type A H1N1 event is a gradual change event. Its nature and risk level are mainly affected by the source of infection, infectivity, transmission route and mortality rate, and the characteristic of the susceptible population. The scenario attributes, such as information disclosure, population mobility, the scope of influence, population density, and prevention and control measures affect whether the incident will continue to escalate. These attributes are used as the evaluation indicators of the loss caused by the incident. 'Influenza A (H1N1)' is structured as {gradual change}, {type A H1N1 virus infection (asymptomatic infection also has infectivity)}, {highly infectious}, {close contact infection (mainly through droplets or aerosols through respiratory tract)}, {mortality rate 6.77%}, {susceptible groups are the contact of patients (such as family members of patients, medical staff}, {the population mobility was very high (especially on holidays)}, {the prevention and control measures were disinfection, wearing protective equipment and vaccine development and vaccination}, {the incidence was higher in winter and spring}, {the impact range was global}, and {the population density in severe epidemic areas was high}.

4.2. The computation result of similarity

The computation processes and results are presented below. We first obtained the results of attribute similarity.

Step (1). For the attribute in the format of crisp symbols, i.e., TP, TR, SP, the attribute similarities $Sim_{TP}(S_0, S_i)$, $Sim_{TR}(S_0, S_i)$, $Sim_{TP}(S_0, S_i)$ are calculated using Equation (4), and the computation results are shown in Table 3.

Step (2). For the attribute in the format of crisp numbers, i.e., FR, NIP, DT, the attribute similarities $Sim_{FR}(S_0, S_i)$, $Sim_{NIP}(S_0, S_i)$, $Sim_{DT}(S_0, S_i)$ are calculated using Equation (5), and the computation results are shown in Table 3.

Step (3). For the attribute in the format of interval numbers, i.e., PD, the attribute similarity $Sim_{PD}(S_0, S_i)$ is calculated using Equation (6), and the computation results are shown in Table 3.

Step (4). For the attribute in the format of fuzzy linguistic variables, i.e., IC, PM, SI, the fuzzy linguistic variables are transformed into the triangular fuzzy number $\tilde{f}_{0k} = \left(f_{0k}^{a}, f_{0k}^{b}, f_{0k}^{c} \right)$, And the attribute similarities $Sim_{IC}(S_0, S_i)$, $Sim_{PM}(S_0, S_i)$, $Sim_{SI}(S_0, S_i)$ are calculated using Equation (7) the computation results are shown in Table 3.

Using the attribute weight determination algorithm based on coverage, the global similarity, $Sim(S_0, S_i)$, $i = 1, 2, ..., 8$, is then obtained, $Sim(S_0, S_1) = 0.4256$, $Sim(S_0, S_2) = 0.7461$,

Table 3. The computation results of attribute similarity

Attribute similarity	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8
$Sim_1(S_0, S_i)$	1	1	1	0	0	0	1	1
$Sim_2(S_0, S_i)$	0.3679	0.8219	0.6013	0.6013	0.3561	0.6615	0.7342	0.7342
$Sim_3(S_0, S_i)$	0	1	1	1	0	1	1	1
$Sim_4(S_0, S_i)$	0.3104	0.8134	0.6329	0.6329	0.3076	0.6859	0.7092	0.7092
$Sim_5(S_0, S_i)$	0.5143	0.8517	0.8073	0.8073	0.3278	0.8148	0.8269	0.8269
$Sim_6(S_0, S_i)$	0	0	1	1	0	1	1	1
$Sim_7(S_0, S_i)$	0.3053	0.8762	0.8231	0.8411	0.4674	0.8297	0.8409	0.8365
$Sim_8(S_0, S_i)$	0.4291	0.8614	0.8128	0.8079	0.4291	0.7245	0.7883	0.8126
$Sim_9(S_0, S_i)$	0.5392	0.4783	0.1459	0.7623	0.7623	0.1482	0.1537	0.1345
$Sim_{10}(S_0, S_i)$	0.1244	0.5689	0.1452	0.1503	0.1503	0.1274	0.1132	0.1126
$Sim_{11}(S_0, S_i)$	0.8357	0.6345	0.4362	0.7462	0.7462	0.5124	0.5287	0.5261

$Sim(S_0, S_3) = 0.5372$, $Sim(S_0, S_4) = 0.5145$, $Sim(S_0, S_5) = 0.5295$, $Sim(S_0, S_6) = 0.5873$, $Sim(S_0, S_7) = 0.5504$, $Sim(S_0, S_8) = 0.5423$. According to the similarity results obtained above, combined with the actual situation, this study sets the similarity threshold $\partial = 0.55$. Therefore, S_2, S_6, and S_7, meet the threshold condition, and constitute a historical scenario similarity set $\{S_2, S_6, S_7\}$.

5. Experiment results

According to the global similarity, similar case sets are retrieved, and the specific emergency measures are modified according to the local medical conditions and the supply of rescue materials. Then, according to the Markov chain to predict the next stage of the scenario, the similarity between the predicted scenario and the historical scenario is calculated. According to the size between the similarity and the threshold, the corresponding remedial and revision measures are determined. The experimental results include the case similarity in the current stage, the calculation result of the Markov transfer matrix and the case similarity calculation results in the next stage.

5.1. Results of infectious disease emergency solution revision

Appropriate prevention and control measures can be generated from the emergency solutions in the historical case similarity set. Alternative solutions can be formed through public health expert evaluation of the emergency measures from a historical scenario similarity set. On the one hand, they take into account the medical conditions and public health conditions in the epidemic area, that is, the operability of emergency measures. On the other hand, considering the actual effect of the measures from a similar case set, the measures with good actual effect are adopted, and the measures with poor actual effect are modified or deleted.

We distinguish two types of prevention and control measures from the historical scenario similarity set $\{S_2, S_6, S_7\}$. There are five prevention and control measures for SARS to achieve the goal of blocking transmission and reducing health hazards, i.e., early detection, early report, early investigation, early isolation, and early treatment. The decision-maker strictly manages the sources of infection to achieve the goal of gradually blocking transmission.

“Early detection” refers to the timely and effective management and control of the most important sources of infection, by improving the sensitivity of medical staff, early detection of suspected cases, and rapid detection and diagnosis.

“Early report” means that suspected patients and confirmed cases need to be reported to health or disease control departments within the specified time, to start investigation and treatment as soon as possible.

“Early investigation” means that after receiving reports of confirmed patients, suspected patients and positive detection data, the disease control department needs to send epidemiological investigators to conduct an in-depth and detailed investigation and analyze the exposure of patients before and after the disease and their contacts after the disease, to find the source of infection and all related close contacts.

"Early isolation" means that all confirmed cases and all suspected cases should be isolated for treatment, and all close contacts should be isolated for medical observation.

"Early treatment" refers to the effective symptomatic, supportive treatment and antiviral treatment to prevent mild illness from developing into severe disease. Severe cases are given full medical care and society strives to reduce the mortality rate. At the same time, this approach can also eliminate patients as sources of infection. Key measures in the strategy of managing pandemic respiratory diseases include priority treatment of severe cases, observation of mild patients at home, and taking measures to increase social distance, such as forbidding or reducing large-scale gatherings, suspension of school, suspension of work, and even declaring a state of emergency, curfew and so on.

Under the recommendation of the prevention and control measures for SARS and influenza, and through the experts' consideration of the constraints on the medical and public health conditions in the epidemic area and the actual effect of SARS and influenza, prevention and control measures are formed at this stage.

5.2. Prediction results of the next scenario and solution

According to the Markov model computation steps, the results of the state transition matrix are obtained.

Based on the state transition matrix, we can predict the next stage scenario S_0', $S_0' = \sum_{i=1}^{8} p_{0i}.S_i$ and calculate the similarity $Sim(S_0', S_i)$ and according to the comparison results between the similarity $Sim(S_0', S_i)$ and the similarity $Sim(S_0, S_i)$, we can decide whether we need to take new measures to prevent and control the spread situation. The result is obtained: $Sim(S_0', S_1) = 0.3175$, $Sim(S_0', S_2) = 0.6818$, $Sim(S_0', S_3) = 0.5023$, $Sim(S_0', S_4) = 0.4588$ $Sim(S_0', S_5) = 0.4724$, $Sim(S_0', S_6) = 0.6179$, $Sim(S_0', S_7) = 0.5792$, $Sim(S_0', S_4) = 0.4736$. Let the similarity threshold $\partial = 0.55$. Still,

S_2, S_6, and S_7, meet the threshold condition, and constitute a historical scenario similar set $\{S_2, S_6, S_7\}$. It shows that the current corresponding solution can completely control and cope with the situation at the scenario S_0' without any compensatory remedies.

6. Conclusions

The development of infectious disease outbreaks is often unpredictable and complex, which requires dynamic emergency responses. In this study, we extracted the features of infectious disease emergencies and analyzes infectious disease emergency scenario evolution. We then proposed a dynamic model based on a Markov model and CBR to analyze scenario evolution in eight stages using a probability matrix. We designed various matching algorithms for crisp symbols, crisp numbers, interval numbers, and fuzzy linguistic variables, and calculated the similarity between the target scenario and historical scenarios. Finally, we provide various alternatives for dynamic emergency decisions through the comparison of the similarity degree of various scenarios.

We empirically tested the proposed model with an example of the infectious disease emergency of COVID-19. The results showed that our model can better realize dynamic CBR to address high uncertainty and complex evolution. It can simulate the infectious disease outbreak process more realistically, and quickly match scenarios to generate effective and real-time emergency decision-making aids to help control the spread of infectious diseases and reduce losses.

Most outbreaks of infectious diseases are a type of sudden and dynamic incidents, which is normally under the impact of a range of uncertain factors. The outbreak of infectious diseases can become more dynamic with the diffusion and migration of infectious sources, therefore data should be treated dynamically and in real-time rather than statically. The model developed in this study allows decision-makers to input real-time data such as the number of

infections and mortality, which contributes to an updating prediction on the development of epidemic outbreaks. All predictions about any incident of an infectious disease outbreak can be stored in a dynamic way, which consists of 8 stages of observations and validation on epidemic development. These stages are connected and organized within a sequence of events, and refer to the entire incident of an infectious disease outbreak. Any incident can be compared with a target incident based on similarity to provide possible preventions of infectious disease in the specific stage.

Moreover, the model is also able to predict attribute data about further infectious disease development based on the subsequent stage simulated by the Markov chain model. The attribute data achieved in prediction can be assessed for its degree of similarity with the threshold determined by public health experts. If the degree of similarity is within the threshold determined by public health experts, it refers to the current prevention strategy that is still effective to control the epidemic. If the degree of similarity exceeds this threshold, a new round of adjustments shall be made to the current prevention and control measures.

This study has certain limitations. First, most of our historical cases are infectious diseases with the respiratory tract as the route of transmission, it would be a challenge to extend our model to all the major infectious disease emergencies. Future research could set the case problem attribute of the model as dynamic, and design the case attribute according to the characteristics of different infectious diseases, and test the homogeneity between problem space and solution space. Second, our model does not consider the effect of infectious disease control measures adopted in the historical cases. Future studies should consider the effectiveness of emergency measures in historical cases in the model, for example, including only cases with effective measures in the historical case base. As emergency measures are often constrained by medical relief resources, an important development direction is to

consider the level of medical relief resources in the historical case database to improve the operability and effectiveness of the solutions.

CRediT authorship contribution statement

Jinli Duan: Conceptualization, Writing - original draft. **Zhibin Lin:** Supervision, Writing - review & editing. **Feng Jiao:** Investigation, Writing & editing. **Yixian Jiang:** Formal analysis. **Kexing Chen:** Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported by the Innovation Strategy Research Project of Fujian Province under grant NO. 2020R0090.

References

Araz, O. M., & Jehn, M. (2013). Improving public health emergency preparedness through enhanced decision-making environments: A simulation and survey based evaluation. *Technological Forecasting and Social Change, 80*(9), 1775-1781. https://doi.org/10.1016/j.techfore.2012.09.018

Bartolucci, F., Farcomeni, A., & Pennoni, F. (2014). Rejoinder on: Latent Markov models: a review of a general framework for the analysis of longitudinal data with covariates. *TEST, 23*(3), 484-486. https://doi.org/10.1007/s11749-014-0393-3

Blasone, R.-S., Vrugt, J. A., Madsen, H., Rosbjerg, D., Robinson, B. A., & Zyvoloski, G. A. (2008). Generalized likelihood uncertainty estimation (GLUE) using adaptive Markov Chain Monte Carlo sampling. *Advances in Water Resources, 31*(4), 630-648. https://doi.org/10.1016/j.advwatres.2007.12.003

Chan, E. Y. Y., Huang, Z., Hung, K. K. C., Chan, G. K. W., Lam, H. C. Y., Lo, E. S. K., & Yeung, M. P. S. (2019). Health Emergency Disaster Risk Management of Public Transport Systems: A Population-Based Study after the 2017 Subway Fire in Hong Kong, China. *International Journal of Environmental Research and Public Health, 16*(2), 228-239. https://doi.org/10.3390/ijerph16020228

Chen, X., Wang, Y., Wu, L.-y., Yan, G.-y., & Zhu, W. (2010). Emergency decision model with multiple stages, multiple objectives, and multidivisional cooperation. *Systems Engineering - Theory & Practice, 3*-(11) , 1977-1985. https://doi.org/10.12011/1000-6788(2010)11-1977

Choi, H., Cho, W., Kim, M.-H., & Hur, J.-Y. (2020). Public Health Emergency and Crisis Management: Case Study of SARS-CoV-2 Outbreak. *International Journal of Environmental Research and Public Health, 17*(11), 3984-3997. https://doi.org/10.3390/ijerph17113984

Didelot, X., Fraser, C., Gardy, J., & Colijn, C. (2017). Genomic Infectious Disease Epidemiology in Partially Sampled and Ongoing Outbreaks. *Molecular Biology and Evolution, 34*(4), 997-1007. https://doi.org/10.1093/molbev/msw275

Ebrahimian, A., Hashemi-Amrei, S.-H., & Monesan, M. (2018). Exploring Factors Affecting the Emergency Specialists' Decision-Making in Case of Emergencies in Patients. *Critical Care Research & Practice, 18*, 1-7. https://doi.org/10.1016/j.eswa.2010.07.153

Endo, A., Leeuwen, E., Baguelin, M. (2019). Introduction to particle Markov-chain Monte Carlo for disease dynamics modellers. *Epidemics, 29*, 1003-1066. https://doi.org/10.1016/j.epidem.2019.100363

Evan, M. (2019). Local approximation of Markov chains in time and space. *Journal of biological dynamics, 113*(1), 1-23. https://doi.org/10.1080/17513758.2019.1569269

Exum, N. G., Betanzo, E., Schwab, K. J., Chen, T. Y. J., Guikema, S., & Harvey, D. E. (2018). Extreme Precipitation, Public Health Emergencies, and Safe Drinking Water in the USA. *Current Environmental Health Reports, 5*(2), 305-315. https://doi.org/10.1007/s40572-018-0200-5

Fan, Z.-P., Li, Y.-H., Wang, X., & Liu, Y. (2014). Hybrid similarity measure for case retrieval in CBR and its application to emergency response towards gas explosion. *Expert Systems with Applications, 41*(5), 2526-2534. https://doi.org/10.1016/j.eswa.2013.09.051

Gómez, S., Arenas, A., Borge-Holthoefer, J., Meloni, S., & Moreno, Y. (2010). Discrete-time Markov chain approach to contact-based disease spreading in complex networks. *Europhysics Letters, 89*(3), 38009-38014. https://doi.org/10.1209/0295-5075/89/38009

Haeussler, K., den Hout, A. v., & Baio, G. (2018). A dynamic Bayesian Markov model for health economic evaluations of interventions in infectious disease. *BMC Medical Research Methodology, 18*(1), 82-97. https://doi.org/10.1186/s12874-018-0541-7

Humensky, J. L., Abedin, Z., Muhammad, K., McClave, M., Torres, T., DiMaria, E. S., Reilly, M. P., & Pincus, H. A. (2020). Promoting interdisciplinary research to respond to public health crises: The response of the Columbia University CTSA to the opioid crisis. *Journal of Clinical and Translational Science, 4*(1), 22-27. https://doi.org/10.1017/cts.2019.426

Jiang, Y., Fan, Z., & Su, M. (2011). Study on the dynamic adjusting method for emergency decision. *Chinese Journal of Management Science, 19*(5), 104-108. https://en.cnki.com.cn/Article_en/CJFDTotal-ZGGK201105015.htm

Jie, Y., Lei, J., & Hong, C. (2005). Dynamic game analysis of emergency management. *Management Review, 3*, 31-37. https://doi.org/10.14120/j.cnki.cn11-5057/f.2005.03.008

Liu, Y., Fan, Z., & Yuan, Y. (2013). Multi attributes risk decision making method for emergency response. *Operations Research and Management Science, 22*(1), 23-28. https://doi.org/10.3969/j.issn.1007-3221.2013.01.004

Ludkovski, M., & Lin, J. (2013). Sequential Bayesian Inference for Detection and Response to Seasonal Epidemics. *Online Journal of Public Health Informatics, 5*(1). https://doi.org/10.5210/ojphi.v5i1.4570

O'Neill, P. D., Balding, D. J., Becker, N. G., Eerola, M., & Mollison, D. (2000). Analyses of infectious disease data from household outbreaks by Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 49*(4), 517-542. https://doi.org/10.1111/1467-9876.00210

Qiao, X. , Wang, X. , & Xu, F. . (2016). Evolution and decision model of major infectious disease based on Generalized Stochastic Petri Nets. In *2016 13th International Conference on Service Systems and Service Management.* (pp. 1-6). https://doi.org/10.1109/ICSSSM.2016.7538649

Szyszkowicz, M., Kousha, T., Castner, J., & Dales, R. (2018). Air pollution and emergency department visits for respiratory diseases: A multi-city case crossover study. *Environ Res, 163*, 263-269. https://doi.org/10.1016/j.envres.2018.01.043

Tada, T., Toyoda, H., Kumada, T., & Kurisu, A. (2020). Comparison of liver disease state progression in patients with eradication of versus persistent infection with hepatitis C virus: Markov chain analysis. *Journal of viral hepatitis, 28*(3), 538-547. https://doi.org/10.1111/jvh.13444

The Data- Centre of China Public Health Science (2021). *Notifiable infectious Disease.* https://www.phsciencedata.cn/Share/en/data.jsp

Wang, D., Wan, K., & Ma, W. (2020). Emergency decision-making model of environmental emergencies based on case-based reasoning method. *Journal of Environmental Management, 262*, 110382-110391. https://doi.org/10.1016/j.jenvman.2020.110382

Wang, L., & Wang, Y.-M. (2013). Study on the emergency decision method of dynamic reference point based on prospect theory. *Chinese Journal of Management Science, 1*, 132-140. https://doi.org/10.16381/j .cnki .issn1003-207x.2013.s1.062

Yang, J.-j., & Xu, C.-h. (2012). Emergency Decision Engineering Model Based on Sequential Games. *Systems Engineering Procedia, 5*, 276 - 282. https://doi.org/10.1016/j.sepro.2012.04.043