Reinforcement learning for safe evacuation time of fire in Hong Kong-Zhuhai-Macau immersed tube tunnel

Kun Tian and Shuping Jiang

College of Civil Engineering, Chongqing Jiaotong University, Chongqing, People's Republic of China; China Merchants Chongqing Communication Technology Research & Design Institute CO., LTD, Chongqing, People's Republic of China

ABSTRACT
In this paper, authors mainly study the laws of safe evacuation time based on reinforcement learning when fire breaks out in the immersed tunnel. In case of fire, time is life. When the people in the tunnel begin to escape, they will instinctively choose the best path they believed in. It is bound to cause congestion and increase the overall escape time. Therefore, the authors designed the reinforcement learning (RL) scheme with multiple escape routes to seek the Nash equilibrium. In each iteration, they update their escape strategy on the basis of the previous outcome. Since the minimum overall time is the objective function, the result tends to converge. In this paper, the author carried out a fire test with a heat release rate of 50 WM. Therefore, total number of people trapped in the high-temperature hazardous area under the condition of traffic jams is 158. Finally, the minimum safe evacuation time of personnel is calculated as 110.5 s through the reinforcement learning model. This paper will provide scientific support for long offshore immersed tube tunnel fire evacuation and emergency evacuation decision-making system.

ARTICLE HISTORY
Received 28 February 2018
Accepted 6 August 2018

KEYWORDS
reinforcement learning; immersed tunnel; fire; safe evacuation

1. Introduction
In recent years, with the establishment of the international economic zone and the new exploitation of global marine resources, an increasing number of transcontinental and transnational passages have been constructed. Immersed tube tunnel plays an important role, as it has many advantages, for instance, it does not occupy the navigation channel, can be used under any weather circumstances, has little impacts on the environment and makes full use of the underground space, etc.. Gradually, people are changing their traditional concept that ‘coming across the water, making a bridge’. The immersed tunnel is under the river and deep sea, which brings convenience as well as difficulties of evacuation when fire disasters occur. The fire has became one of the main disasters which occur frequently in tunnels. And the deep depth of the immersed tube tunnel makes the escape from vehicle fire extremely difficult. Through the analysis of a large number of accident cases, we found that no effective measures were taken to prevent the possible crowd crowding before the accident and there was no timely and effective evacuation in case of fire. When fire and other accidents happen, the clock is ticking, how to choose the optimal path to achieve minimum evacuation time becomes an important issue.

Researchers around the world have made some progress in the study of personnel evacuation. From the perspective of modelling mechanism, the evacuation behaviour model is divided into three types (Hu, Liu, Wang, & Cheng, 2012). (1) Fluid and Particle Models (FSEG, 2003). The representative is the social forces model of Helbing, Farkas, and Vicsek (2000), which can be used to explain the phenomenon of social psychology and physical force. It’s also a continuous model. (2) Matrix-based Models. These are all discrete models. Because the construction area is discretized into the cell in the matrix based system, there are some boundary conditions. In addition, the decentralized approach relies on the user’s technology. (3) Emergency Models. The representative is the Legion emergency model (2004) and other models for emergency evacuation of people (Epstein & Axtell, 1996; Johnson, 2001; Waldrop, 1992). In these models, however, individual behaviour is often oversimplified. With the development of computer technology and the research of people’s behaviour and psychology, researchers begin using computers to directly simulate the movement of people in buildings. A large number of evacuation models have been established. So far, evacuation simulation research has been carried out in two directions: based on proxy simulation and non-proxy simulation. At the
same time, a large number of commercial simulation software has emerged, such as EGRESS, FDS, EXODUS (2003), SIMULEX (2000) and so on.

(1) AGENT-BASED INDIVIDUAL DECISION-MAKING MODELS

There are three types of research in this area: (A) Function Relations. Main representative is Bo and Yong-gang (2010) proposed LWDPSO (Linear Weight Decreasing Particle Swarm Optimization) algorithm, which can be used to determine the next agent’s position and speed. And Peizhong, Xin, and Tao (2011) established the interaction model between people and fire that can be described by three functional relationships. (B) Rules. Sharad Sharma used fuzzy logic in 2009 to reveal the impact of individual emotions on the behaviour of the evacuation process. And in 2010, the decision rules based on fuzzy logic were further developed. (C) Functional relations and Rules. Fangqin and Aizhu (2008) established some corresponding rules to guide the individual export route by combining with FDS (fire Dynamic System). Shi, Ren, and Chen (2009) establish a series of behavioural decision rules for individual individuals after analysing the relationship between individual, individual and environment, individual and fire.

(2) AGENT BASED CROWD EVACUATION MODELING

This representative model includes the social forces model of Masakuni Muramatsu (1999), the lattice gas model of Helbing, Hostikka, and Keski-Rahkonen (2005) and the famous study on human rules of Reynolds (1987). All these models have explained the problem of escape from different angles, but no satisfactory answer has been given on the minimum safety time of personnel evacuation. All of these models have promoted the study of personnel evacuation, but none of them take into account the actual situation of personnel evacuation. Furthermore when individuals evacuate, every decision is random and dynamic. Therefore, these models cannot objectively reflect the real situation of evacuation when fire occurs.

Personnel evacuation is carried out along a timeline. When a fire breaks out, the exit doors near the fire point are will be open, and people in cars are informed of the fire by observing, broadcasting in the tunnel or other warning signals, then they begin to escape. After getting off the car, they start to choose the escape door. As the escape door opens more than one, people usually flee to the nearest exit door. Therefore, it will be crowded on the escape route to the exit door in a very short time, which will greatly affect the overall evacuation time of the personnel. How to achieve a safe escape in the shortest possible time and get the minimum safe time (minimum time to ensure safe escape) becomes a choice optimization problem. However, the existing evacuation strategy for tunnel fire has not been considered. Reinforcement learning in the optimization problem is a mathematical model widely used in other fields, which can solve this problem well.

Reinforcement learning (RL) is an important machine learning method, and it has many applications in intelligent control robot and analysis and prediction (Du, Han, & Li, 2014). Reinforcement learning is developed from the theory of animal learning and parameter disturbance adaptive control. It can be traced back to Pavlov’s conditioned reflex experiment. The term reinforcement learning in computers was first coined by Minsky in the 1950s. It is a special and adaptive way of learning by interacting with the environment and using environmental feedback as input. Since the end of last century, with the breakthrough in basic mathematical research on reinforcement learning, researches and applications of reinforcement learning are increasingly being carried out and has developed into a multidisciplinary interdisciplinary science, including operations research, neural networks, psychology and control engineering. It is one of the most active branches in the field of artificial intelligence. The basic principle is that if an Agent’s behavioural strategy leads to the reward (reinforcement signal) of the environment, then the tendency of the Agent to generate this behaviour strategy will be strengthened. The goal of agent is to find the optimal strategy in each discrete state to obtain the maximum reward. Therefore, it is possible to establish a personnel evacuation model by strengthening the learning model to solve the minimum safe time.

Main work and difficulties of this paper:

(i) The author has established the world’s largest section immersed tube tunnel experimental base, and in which carried out the large-scale fire test to calculate high-temperature hazard range of fire.
(ii) The author conducted an evacuation test in the base to measure the true evacuation speed of personnel.
(iii) It is that first time to apply reinforcement learning model to the field of evacuation of tunnel fire personnel and to establish an evacuation model.

2. Related entity experiment

There is a big difference between the escape of the tunnel personnel and the mountain tunnel. There are two main reasons: (1) Limited escape area. In case fire personnel
escape from the immersed tube tunnel, the final escape area is limited by the area of the artificial island connected to it. The final escape area of the mountain tunnel is anywhere outside the tunnel (including the entire mountain surface), which can be regarded as infinite. (2) More escape directions. The immersed tube tunnel adopts the horizontal smoke exhaust method. In the fire, people can escape from the fire in two directions: upstream and downstream. However, the mountain tunnel adopts the vertical smoke exhaust method, only to escape from the fire upstream.

In order to solve the problem of the minimum safe time of personnel escape in immersed tube tunnel through RL model, the author carried out a large number of experiments in the experimental base of HK-ZH-M, and the experimental base was shown in Figure 1. The HK-ZH-M passage is a magnificent engineering project in the world and has successfully overcome many worldwide problems. The experimental tunnel base is based on the national science and technology support plan to study the disaster prevention and reduction of immersed tube tunnel. The experimental tunnel is located in Fujian province, China, and is constructed according to the section size of HK-ZH-M immersed tunnel. Thousands of experiments have been carried out in the experimental tunnel, and a large number of valuable experimental datum and many real and reliable scientific laws have been obtained. All the physical tests involved in this paper are carried out in this experimental base.

2.1. Experiment on high-temperature hazard range of fire

The cause of death in tunnel fire can be divided into two categories: high temperature and gas (Tian, Chen, Jiang, & Xie, 2015). Because the smoke can be controlled by ventilation when the fire occurs, so this paper only discusses the evacuation situation of the personnel in the danger range of high-temperature effect. The experiment of fire burning is shown in Figure 2. The scale of the fire in the experiment was 50 MW (2 heavy trucks collide with each other), which is the recommended value given in the tunnel fire prevention standard (Borghetti, Derudi, Gandini, Frassoldati, & Tavelli, 2017). The experiment used six 1.5 × 1.5 m oil pools with 93# gasoline as the source of fire, and the thermal heat release rate (HRR) reached 50 MW. Studies (Xu, 2014) have shown that people feel serious discomfort in mouth, nose and esophagus at temperatures above when the temperature is above 50°C. People will appear giddy, and quickly collapsed when gets 95°C. The human existence breathing limit temperature is about 131°C. The physiological function decline at 140°C, and lose entirely when hits 180°C. For breathing, however, the temperature above 60°C was unbearable to the majority of people, which makes fire rescue and indoor personnel escape slow.

Experiments show that the temperature distribution at the ground height H is 1.5, 2.5, 3.5, 4.5, 5.5m and 6.5 m, respectively, as shown in Figure 3. The results show that when HRR is 50 MW, in close to the fire area, from the height of 1.5 m (personal health level), the range above 60°C is about 25 m of fire upstream and downstream. Therefore, this paper only considers the minimum safety time of the people within a 50 m range.

2.2. Experiment of personnel escape velocity

2.2.1. Experimental design

The SFPE Handbook of Fire Protection Engineering, developed by The National Fire Protection Association, points

![Figure 1](image1.png)

*Figure 1.* The surface of the experimental base of HK-ZH-M immersed tunnel.
out that the walking speed of the personnel in the uncrowded normal conditions is 1.5 m/s. But in the event of a fire, people are anxious to evacuate from fire, the speed of escape will increase. However, in the event of an escape, the escape velocity of the personnel will decrease due to the obstacle on the walking path (such as the irre-movable cars or the unsmooth overhaul road). Therefore, in order to obtain the escape velocity of the people in the tunnel under the fire scene, the author carried out the personnel speed test in the experimental tunnel, and the escape plan was shown in Figure 4.

The total number of bus seats for the evacuation is 51 (including one for drivers and co-drivers), and the size of the vehicle is 11.48 m long, 2.49 m wide, 3.54 m high.

The following are the characteristics of the sample:

(1) Age distribution. The age distribution of the test is shown in Table 1.

The standard axis of the age is $\varphi = 26$. Therefore, this test can be used as a support for people aged between 25 and 30 when the fire occurs in the tunnel. It can

Table 1. Age distribution table of evacuation personnel.

| Age       | Under 25 | 25–30 | 30–35 | 35–40 | 45–50 | Above 50 | Total |
|-----------|----------|-------|-------|-------|-------|----------|-------|
| Amount    | 2        | 24    | 12    | 7     | 2     | 1        | 48    |

Figure 2. 50 MW gasoline fire (93#) combustion experiment.

Figure 3. Temperature distribution of 50 MW gasoline fire tunnel.

Figure 4. Evacuation plan of fire personnel.
also be used as a reference for other age groups to evacuate.

(2) Gender distribution. There are 41 male samples and 7 female samples.

(3) Health. Before the experiment, the authors investigated the physical condition of each participant. The result is 42 good, 6 excellent, no history of heart disease, no disabled people.

(4) Fire experience. In the survey before the experiment, the participants had no fire experience and did not know the relevant knowledge of fire escape (Figure 5).

2.2.2. Analysis of experimental results

Ratio flow (refers to the number of people passing through the unit width in unit time, people/(m*s) at the exit of the channel). The movement behaviour of personnel is affected by three important parameters: velocity, flow and density, and the relationship between the values and parameters of these three parameters determine the movement behaviour of the personnel. Through the video data of the escape experiment, the authors extracted and calculated these parameters and obtained the parameters of the personnel movement in the real fire conditions.

The area $A_{enter}$ near the exit is the measurement area. There is an overhaul road before exit, and the speed is reduced when the personnel steps on the road. During the escape, the personnel began to decelerate in the range of about 1.2 m before the overhaul road, and the width of the overhaul road is 0.75 m. Therefore, the area $A_{enter}$ of the entrance can be calculated.

$$A_{enter} = \frac{1}{2} \pi R^2 = \frac{1}{2} \times 3.14 \times (1.2 \text{ m} + 0.75 \text{ m})^2 \approx 6 \text{ m}^2$$

By the data recorded by the camera, the author selects a video data with a duration of 30 s, and the crowd movement is relatively continuous during this time period. Through the analysis of the evacuation behaviour in this period, the author gets personnel density $\rho_i$ (the density of the initial moment of the $i$ seconds) of each moment (in seconds) in area $A_{enter}$, and $\rho = 1.92 \text{ people/m}^2$.

During the evacuation process during the 30 seconds of extraction, initially, the personnel density is small, the personnel escape velocity is larger. As the personnel density increases, the movement of the population in the exit area tends to be stable, with a decrease in speed and fluctuation. The ratio flow – density is given in SFPE – relation:

$$q = 1.4\rho - 0.3724\rho^2$$

Calculate the ratio flow at the exit door. $q = 1.31 \text{ people/(m} \times \text{s)}$

Average speed. Calculation formula of average velocity:

$$v_a = \frac{\sum v_i}{\Delta T}$$

The escape velocities of the samples are shown in Figure 6.

The average velocity of the overall survival of personnel in the export area is calculated as $v_a = 2.7 \text{ m/s}$. But in this experiment, visibility is very good. In the case of fire,
the visibility in the tunnel is reduced due to the shading effect of the smoke. Considering the effect of the reduction coefficient on the evacuation speed, the evacuation speed of the tunnel is taken as \( v' = 1.62 \text{ m/s} \). At the same time, from fire information is obtained to the total evacuation, the time cost \( t_1 = 27 \text{ s} \). A total of 27 s will be spent to evacuate the bus.

2.3. Total number of people trapped in high-temperature hazardous area

The experimental tunnel was built with the HK-ZH-M immersed tube tunnel. Therefore, the proportion of vehicles in the tunnel is selected according to the traffic volume prediction results of the basic quota scheme from The feasibility study report of the Hong Kong-Zhuhai-Macau bridge project. In 2035, the traffic volume is expected to reach 90,000 pcu (Passenger Car Unit), including 21,250 pcu for private cars, 4250 pcu for tour buses, 14,950 pcu for container trucks, and 8759 pcu for ordinary goods. According to the foregoing, the high-temperature danger zone is 50 m. Considering the severity of traffic congestion during a fire, according to the calculation, the number of people needing to be evacuated in the high-temperature hazardous area is \( N = 158 \), in the case of 50 MW fire in the tunnel.

3. Model establishment

Through the support of the above work, a mathematical model is established in this section to solve the problem of minimum safety time for personnel evacuation.

3.1. Establishment of fire evacuation objective function

When there is a fire in an immersed tunnel, the location of the fire is random. In order to ensure the security of the whole system, at the same time, according to the most unfavourable principle of the working condition, it can be considered that the fire point is close to an escape door. Therefore, when the vehicles in the tunnel catches fire and forms a fire, it is believed that the people in the fire hazard zone get off and converge at a point (origin) after time. Then they select the path to start the escape. Along with the longitudinal of the tunnel, the vehicles on one side of the fire (defined as the downstream direction) have been driven away. There are fewer obstacles on the escape path, and the available capacity \( C_e \) is larger. On the other side of the fire, which is defined as upstream, the traffic is blocked. The vehicle cannot be moved and the available capacity \( C_e \) on the escape route is smaller. When every trapped person begins to escape, they face the choice of an escape route. Through the study on the rule of human escape behaviour (Zhou Jian, 2014), in a disaster, the human mind is reduced. After choosing a path and starting to escape, few individuals can change paths through the surrounding environment. In addition, the width of the tunnel (determines the length of the escape path) is relatively small, so it is believed that when the personnel chooses the path to escape, they will no longer change other paths. After \( t_2 \) time, all the people in the fire danger zone managed to escape to the cross-channel. Think that all the people converged on a point destination (D) in the horizontal channel. In order to obtain the minimum time for safe escape for all personnel, it is necessary to ensure that the crowding level on each path is equal.

Now, it transforms the tunnel fire evacuation problem into an O–D (origin–destination) path selection problem with reinforcement learning.

For the O–D model of personnel escape (PE), the number of escape doors is denoted as the escape route nodes recorded \( E \), and the number of escape routes can be recorded as \( L \). So this PE O–D model can be represented by a directed graph \( G(E, L) \), it is quite clear that \( L \subseteq E \times E \). The number of people in the fire hazard zone is recorded \( N \), \( N = \{1, 2, \ldots, N\} \). Everyone has a particular path to escape, and these paths form a set \((o_p, d_p) \subseteq R \times R \). The escape route chosen by the escape crew \( k \) in the \( i \) times is denoted as \( r_{ik} \), and \( R = \{r_{ik}\} \subseteq N, k = 1, 2, \ldots, R_k \). The probability of choosing \( r_{ik} \) route to escape is \( p_{ik} \). The time function of escaping can be defined as \( T_r(f_1), f_1 = 1, 2, \ldots, n \), thus, \( T_r(\cdot) \) is the BPR function that we know, i.e.

\[
T_r(f_1) = t_0 \left(1 + \alpha \left( \frac{f_1}{C_e} \right) \right) \tag{4}
\]

where \( t_0 \) represents the time that the person is using the path \( r \) in free case, \( C_e \) is the number of people that can be accommodated without congestion on path \( r \), \( \alpha \) and \( \beta \) is
the two adjustment coefficients, they are determined by the given escape environment, the value of β is greater than 3. And T_r(\cdot) is differentiable, and the first order function is positively related to f_\ell. Then we can see that for the escape crew k, k = 1, 2, ..., N, here are two factors influencing the escape time: (1) the escape route p_k of self-selection; (2) escape paths p_{-k} chosen by others. So, the total number of people on the escape route r is

\[ f_r(p_k, p_{-k}) = \sum_{k=1}^{n} \varepsilon'_k \]  

with \[ \varepsilon'_k = \begin{cases} 1, & \text{if } k \in p_k \\ 0, & \text{otherwise} \end{cases} \]

Therefore, for the survivor k, when he knows the probability p_k of choosing the path r and probabilities p_{-k} of other people choosing the same way, his total escape time is T_r(p_k, p_{-k}) can be expressed mathematically.

\[ T_k(p_k, p_{-k}) = \sum_{k \in p_k} c_k(f_k(p_k, p_{-k})), k = 1, 2, \ldots, n \]  

Obviously, when this kind of crowding problem is analysed with the theory proposed by Rosenthal (1973), we can easily get this mathematical model. Therefore, the solution of PE problem becomes a mathematical programming (MP) solution. As follows:

\[ \min T(f) = \sum_{e \in E} \int_{0}^{T_e} r_e(s) ds \]  

\[ \sum_{k=1}^{R_k} \varepsilon_{k,k} = 1 \]  

\[ \varepsilon_{k,k} \geq 0 \]

where \( f = (f_1, f_2, \ldots, f_N)^T, f_e = \sum_{k=1}^{R_k} \sum_{i=1}^{R_e} f_{k,i} e_{k,i}, k = 1, 2, \ldots, N, i = 1, 2, \ldots, R_e \).  

### 3.2. Learning automata design

The symbols in this paper are defined as follows:

\[ \{ \Xi, U, \mathcal{I}, \{ \xi_n \}, \{ u_n \}, \{ p_n \}, T \} \]

defines:

(i) \( \Xi \) denotes the input bounded set of the automaton.

(ii) \( U \) is the set \{u(1), u(2), \ldots, u(N)\} of the automaton actions.

(iii) \( \mathcal{I} = (\Omega, F, P) \) is a probability space.

(iv) \( \{ \xi_n \} \) denotes a sequence of automaton inputs (environment response, \( \xi_n \in \Xi \)) provided by the environment in a binary (P-model environment) or continuous (S-model environment) form.

(v) \( \{ u_n \} \) denotes a sequence of actions (automaton outputs).

(vi) \( p_n = [p_n(1), p_n(2), \ldots, p_n(N)]^T \) is the probability distribution at time n \( p_n(i) = P(\omega : u_n = u(i)/F_{n-1}) \) and \( \sum_{i=1}^{N} p_n(i) = 1, \forall n \) with \( F_n = \sigma(\xi_1, u_1, p_1; \ldots; \xi_n, u_n, p_n) \) is the \( \sigma \)-algebra generated by the corresponding events (\( F_n \in F \)).

(vii) \( c_n = [c_n(1), c_n(1), \ldots, c_n(N)]^T \) is the conditional mathematical expectation vector depends the environment responses (at time n).

(viii) \( T \) represents the reinforcement scheme (updating scheme) which changes the probability vector \( p_n \) to \( p_{n+1} \):

\[ p_{n+1} = p_n + \gamma_n T_n(p_n; \{ \xi_t \}_{t=1}^{n}; \{ u_t \}_{t=1}^{n}) \]  

\[ p_{1(i)} > 0 \quad \forall i = 1, \ldots, N \]

with \( \gamma_n \) is a scalar correction factor and the vector.

\[ T_n(\bullet) = [T_n^1(\bullet), \ldots, T_n^N(\bullet)]^T \]

satisfies the following conditions (for preserving probability measure).

#### 3.3. Learning automata design

After considering the feature that connected in a feedback loop to the random medium (environment) of learning automata, the author chooses the reinforcement schemes (updating schemes) which are the mechanisms used to change the probability vector \( p_n \) to \( p_{n+1} \).

(a) Nonprojectional Reinforcement Schemes

\[ p_{n+1} = p_n + \gamma_n T_n(p_n; \{ \xi_t \}_{t=1}^{n}; \{ u_t \}_{t=1}^{n}) \]  

\[ p_{1(i)} > 0 \quad \forall i = 1, \ldots, N \]

where \( \gamma_n \) is a scalar correction factor, the vector \( T_n(\bullet) = [T_n^1(\bullet), \ldots, T_n^N(\bullet)]^T \) satisfies the following conditions.

\[ \sum_{i=1}^{N} T_n^i(\bullet) = 0, \forall n \]  

\[ p_{n(i)} + \gamma_n T_n^i(\bullet) \in [0, 1], \forall n, \forall i = 1, \ldots, N \]  

and \( \{ \xi_t \}_{t=1}^{n} \) is the sequence of environment responses \( \xi_n \in R^1 \) which will be constructed on the basis of the available data in different cases (observations), i.e.

\[ \xi_n = \xi_n(\{ y_t \}_{t=1}^{n}) \]

For the non-projective schemes, it is always assumed that the environment response belongs to the unit closed line segment, i.e.

\[ \xi_n \in [0, 1] \]
(b) Projective Reinforcement Schemes

\[ p_{n+1} = \pi_n \{ p_n + \gamma_n T_n (p_n; \xi_1, \ldots, \xi_n; [u_t]_{t=1}^{N}) \} \]

\[ p_1(i) > 0 \quad \forall i = 1, \ldots, N \]

where \( \pi_n \{ \cdot \} \) is the projection operator onto the \( \varepsilon_n \)-simplex, \( S^n \) defined as follows \( S^n \equiv \{ p = (p_1, \ldots, p_n) : p(i) \geq \varepsilon_n \geq 0, \sum_{i=1}^{N} p(i) = 1 \} \) (for these schemas conditions (10) and (11) are not obligatory).

In this paper, the author adopts the Varshavskii–Vorontsova reinforcement schemes. The \( \Phi(p_n) \) function can be expressed as:

\[ \Phi(p_n) = cte - \frac{1}{2} p_n(i)^2 - \frac{1}{3} p_n(i)^3 \quad (12) \]

The loss function \( \Phi_n \) associated with the learning automaton is usually given by

\[ \Phi_n = \frac{1}{n} \sum_{t=1}^{n} \xi_t \quad (13) \]

It is a useful measure for judging the behaviour of a learning automaton. If a stochastic automata to minimize the loss of its function (to find the best control behaviour \( x(\alpha) \)), then it will automatically solve a discrete set of corresponding constraint random optimization problem.

Now let’s consider a random static environment, and the response \( \xi_n \) is characterized by the following two properties:

(H1) The conditional mathematical expectations of the environment responses exist and are stationary, i.e.

\[ E \{ (\xi_n - c_n(i))^2/F_{n-1} \wedge u_n = u(i) \} = a^2(i), \forall i = 1, \ldots, N \quad (14) \]

(H2) The conditional variances of the environment responses are bounded, i.e.

\[ E \{ (\xi_n - c_n(i))^2/F_{n-1} \wedge u_n = u(i) \} = a^2(i), \forall i = 1, \ldots, N \quad (15) \]

\[ \max_{i} \sigma_i^2(i) := \sigma^2 < \infty \quad (16) \]

where \( F_n \subseteq F \) is the \( \sigma \)-algebra generated by the corresponding process, i.e.

\[ F_n := \sigma (\xi_1, p_1, u_1; \ldots; \xi_n, p_n, u_n) \quad (17) \]

3.4. Varshavskii-VoRontsova reinforcement scheme

Consider now the reinforcement scheme, described in

\[ p_{n+1} = p_n + \gamma_n \cdot p_n^T \cdot e(u_n) \cdot (1 - 2 \xi_n) [e(u_n) - p_n] \quad (18) \]

with \( \gamma_n \in [0, 1], \quad \xi_n \in [0, 1], \quad p_1(i) = 1/N, \quad e(u_n) = (0, \ldots, 0, 1, 0, \ldots, 0)^T, \quad u_n = u(i), e_N = (1, \ldots, 1)^T \in R^N \)

In spite of its nonlinear characteristics, the convergence analysis of the scheme can be accomplished on the basis of the Lyapunov approach using the Lyapunov function i.e.

\[ W_n := \frac{1 - p(\alpha)}{p(\alpha)} \quad (19) \]

4. Numerical experiments

4.1. Experiment configuration

In order to verify the correctness of the aforementioned model, the author sets the same boundary conditions for people to escape as the HK-ZH-M immersed tunnel. The physical experiment is shown in Figure 7. The O–D model obtained through mathematical transformation is shown

![Figure 7. The actual scene of escape of the trapped persons.](image-url)
In this function, \( t_0 \) denotes the time taken by different paths for people to be free. This paper focuses on the problem of congestion, so this time can be set to simple values. In this paper, \( t_0 = [1, 1, 1] \), \( C_e \) denotes the capacity of the route. If you exceed this capacity, the path will be crowded. In this paper, according to the length of the escape route, and considering the occurrence of fire, the vehicle in the upper reaches of the fire is blocked, and the vehicle downstream of the fire has escaped from the fire area according to the evacuation information. Therefore, the path capacity is set as \( C_e = [20, 25, 40, 35] \). \( \alpha \) and \( \beta \) are two parameters. The value of this paper is \( \alpha = 0.35 \), and \( \beta = 3.5 \). Therefore, the evacuation model is calculated under the boundary condition of the total number of evacuation \( N = 158 \). The number of iterations is \( i = 800 \).

At first, all the evacuees have a random choice of four routes. Every time after evacuation, the escape time is different because the capacity of each escape route is different. After learning, the next option would be to choose the shorter escape route. Ultimately each person chooses the probability of each path to be a particular value after numerous iterations.

### 4.2. Results analysis

The author calculated the above model through MATLAB software and obtained the following results:

The results show that as the number of iterations increases, the number of people choosing each route is different. Figure 9 represents the number of people on the path R1, R2, R3, and R4. Several diagrams show the same pattern. Namely when the iterations are relatively small, the number of people selected on each route varies greatly. When the iterations are large, the
number of people on each route tends to be stable. This is because when the sample (the escape crew) selects the path on the first time (donated as $n_1$), it is a random choice. Because each path has a different capacity, each person has a different time to escape. And after the escape, all the samples receive this information. Therefore, the next time (donated as $n_2$) escape, the sample will judge from the last (donated as $n_1$) escape result and choose the route (donated as $r^*$) with a smaller escape time. However, when most samples choose $r^*$, it leads to a dramatic increase in congestion and makes escape time (denoted as $t^*$) increased. So, when the sample selects in the $n_3$ time, it will choose another route and get another time on the basis of the second result. So again and again, each sample is learning in the RL model with the number of iterations. Since the minimum safe escape time is the target function for each sample, it has to make sure that each route has the same level of congestion. Therefore, the number of people on each path tends to a definite value, and the final result converges.

The results show that, as the number of iterations increases, the probability of the first, second, third, and four samples to choose the path of the path tends to a fixed value. In other words, with the increase of the number of iterations, different samples will eventually choose a certain path. The reasons for this result have already been explained in the previous article. The results show that the number of people at each exit tends to be fixed when the number of iterations is large. This is in line with the foregoing conclusion. But the result in the picture is oscillating within a certain range, this is due to the small amount of data and the number of iterations (Figure 10).

4.3. The minimum safe escape time

It can be seen from the numerical experiment that the number of samples in four is $N'$, and $N' = [27, 32, 53, 46]$. Let us say that the escape samples get out from the vehicle after $t_1$, and $t_1 = 27s$. Then they start to escape to the four EXIT doors by the choice given by RL. The escape velocity is $v_e$, and $v_e = 1.62m/s$. After time $t_2 = L/v_e$, they arrive
exit entrance area. From the second chapter, we know that the ratio flow of the exit is \( q = 1.31 \text{ persons/m} \cdot \text{s} \). The flow rate of each exit is \( Q = 1.965 \text{ persons/s} \). Through time – flow equation:

\[
T = \frac{N}{Q}
\]

After calculation, people get off and escape to safety spending at least the time \( t_3, t_3 = \max T = 26.5 \text{s} \). Therefore, the minimum safe escape time of the HK-ZH-M immersed tube tunnel is \( t = t_1 + t_2 + t_3 = 110.5 \text{s} \).

5. Conclusion and discussion

5.1. Summary

Based on the mathematical model of RL, the paper studied the evacuation laws of fire in the tunnel. And through the physical experiment and numerical calculation of fire escape, and the important parameters of personnel escape is obtained. After analysing the law of personnel escape, a mathematical algorithm for minimum safe escape time is established, and the minimum safe escape time can be calculated. This is of great significance to the whole project operation disaster prevention.

5.2. Discussion

In this paper, the people in the high-temperature range were selected to conduct an escape study. But in the event of a fire, there are people in the non-hot zone. The cause of death turns from heat to smoke. Therefore, it is necessary to study the law of escape under the influence of toxic gases.

In this paper, the RL model is selected to calculate the evacuation of people. But in real life, the people in the vehicle will have some family members, relatives, friends, classmates and other social relations. In the escape, this group of people often escapes to the same place. So we must take it into account the important influence of social relations. This is the direction of the author’s future research.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

AEA Technology. (2002). A technical summary of the AEA EGRESS code. Technical Report, AEAT/NOIL/27812001/002(R), Issue 1.
Arnott, R., Palma (de), A., & Lindsey, R. (1991). Does providing information to road drivers reduce congestion? Transportation Research Part A, 25(5), 309–318.
Aziz, F. Z., & Ukkusuri, S. V. (2017). Learning-based traffic signal control algorithms with neighborhood information sharing: An application for sustainable mobility. Journal of Intelligent Transportation Systems, 1–13.
Ben-Elia, E., & Shiftan, Y. (2010). Which road do I take? A learning-based model of route-choice behavior with real-time information. Transportation Research Part A: Policy and Practice, 44(4), 249–264.
Bo, Y., & Yong-gang, W. (2010). A GIS-based simulation for occupant evacuation in an amusement building. 2010 2nd international Asia conference on informatics in control, automation and robotics. 274–277.
Borghetti, F., Derudi, M., Gandini, P., Frassoldati, A., & Tavelli, S. (2017). Tunnel fire testing and modeling: The Morgex North Tunnel Experiment. Springer.
Chen, B., & Cheng, H. H. (2010). A review of the applications of agent technology in traffic and transportation systems. IEEE Transactions on Intelligent Transportation Systems, 11, 485–497.
Du, L. L., Han, L. S., & Li, X. Y. (2014). Distributed coordinated in-vehicle online routing using mixed-strategy congestion game[J]. Transp. Res. Part B: Methodol, 67, 1–17.
El-Tantawy, S., Abdulhai, B., & Abdelgawad, H. (2014). Design of reinforcement learning parameters for seamless application of adaptive traffic signal control. Journal of Intelligent Transportation Systems, 18(3), 227–245.
Epstein, J., & Axtell, R. (1996). Growing artificial societies: Social science from the bottom up. Cambridge: MIT Press.
Fanggin, T. A. N. G., & Aizhu, R. E. N. (2008). Agent-based evacuation model incorporating fire scene and building geometry. Tsinghua Science and Technology, 13(5), 708–714.
Fire Safety Engineering Group. (2003). Building EXODUS: The evacuation model for the building environment.
Gwynne, S., & Galea, E. R. (1997). A review of methodologies and critical appraisal of computer models used in the simulation of evacuation from the built environment. Society of Fire Protection Engineers, 1–93.
Halcrow Group Ltd. Pedroute. (2003). Available from: http://www.halcrow.com/pdf/urban_reg/pedrt_broch.pdf
Helbing, D., Farkas, I., & Vicsek, T. (2000). Simulating dynamical features of escape panic. Nature, 407, 487–490.
Helbing, D., Hostikka, S., & Keski-Rahkonen, O. (2005). A proposal for the goals and new techniques of modeling pedestrian evacuation in fires. The 8th IAFSS symposium, Beijing. 557–567.
Hu, Y., Liu, X., Wang, F., & Cheng, C. (2012). An overview of agent-based evacuation models for building fires. Proceedings of 2012 9th IEEE international conference on networking, sensing and control, Beijing. 2012, 382–386. doi:10.1109/ICNC.2012.6204949.
Johnson, S. (2001). Emergence: The connected lives of ants, brains, cities, and software. New York: Simon & Schuster.
Legion International Ltd. (2004). Legion. Nov. 2004, Available from http://www.legion.biz/
Lu, X., Gao, S., & Ben-Elia, E. (2011). Information impacts on route choice and learning behavior in a congested network: An experimental approach. Transportation Research Record, 22(43), 89–98.
Ministry of Transport of the People’s Republic of China. (2014). Detailed ventilation design rules for highway tunnels JTG/TD70/2-02-2014[S]. China Communications Press.
Muramatsu, M., & Nagatani, T. (1999). Jamming transition in pedestrian counter flow [J]. Physica A: Statistical Mechanics and its Applications, 267, 487–498.
Peizhong, Y., Xin, W., & Tao, L. (2011). Agent-based simulation of fire emergency evacuation with fire and human interaction model. *Safety Science, 49*, 1130–1141.

Ramachandran, G. (1990). Human behavior in fires – a review of research in the United Kingdom. *Fire Technology, 26*(2), 149–155.

Reynolds, C. (1987). Flocks herds and schools a distributed behavioral model. *Computer Graphics, 21*(4), 25–34.

Rosenthal, R. W. (1973). A class of games possessing pure-strategy Nash equilibria. *Internat. J.Game Theory, 2*, 65–67.

Sharma, S., Member, IEEE. (2009). Simulation and modeling of group behavior during emergency evacuation. *Proceedings of Intelligent Agents Conference*, 423–428.

Shi, J., Ren, A., & Chen, C. (2009). Agent-based evacuation model of large public buildings under fire conditions. *Automation in Construction*, 18338–18347.

Tian, K., Chen, D., Jiang, S., & Xie, Y. (2015). Study on reasonable open angle of single smoke vent in immersed tube tunnel. *Technology of Highway and Transport, 4*, 122–126.

Waldrop, M. (1992). *Complexity: The emerging science at the edge of order and chaos*. New York: Simon & Schuster.

Xu, P. (2014). Research on fire smoke movement characteristic in submarine immersed tunnel. Chongqing Jiaotong University.

Zhou, J., Jiang, X., & Chen, D. (2014). Investigation of evacuation behavior in highway tunnel fire. *Fire Science and Technology, 33*(3), 327–329.