Fire Safety Journal 106 (2019) 163–176

Contents lists available at ScienceDirect

Fire Safety Journal

journal homepage: www.elsevier.com/locate/firesaf

A machine learning based study on pedestrian movement dynamics under emergency evacuation

Ke Wang, Xiupeng Shi, Algena Pei Xuan Goh, Shunzhi Qian*

School of Civil and Environmental Engineering, Nanyang Technological University, 639798, Singapore

ARTICLE INFO

Keywords:
Emergency evacuation
Movement dynamics
Machine learning
Movement pattern

ABSTRACT

Knowledge of evacuees' movement dynamics is crucial to building safety design and evacuation management. Although it is recognized that stepwise movement is the fundamental element to construct the whole evacuation process, movement pattern and its influencing factors are still not well understood. In this study, we explored the potential of adopting machine learning methods to study evacuees' stepwise movement[1] dynamics based on two videos of quasi-emergency evacuation experiments. The movement patterns were categorized through Two-step Cluster Analysis and principal influencing factors were identified through Principal Component Analysis. The relationship between the movement patterns and the principal components were investigated using different modeling methods: traditional method (Multinomial Logit Model, MLM) and machine learning methods (Decision Tree, Support Vector Machine, K-Nearest Neighbor, and Artificial Neural Network). Results from two experimental videos showed reasonable consistency and the main findings are: (1) Distance to the target exit has the most pronounced effect on a single evacuee's stepwise movement pattern. (2) Surrounding evacuees' actions also have significant and complex influence on a single evacuee's stepwise movement pattern. (3) MLM showed comparable prediction accuracy with machine learning methods when the scenario is simple. The superiority of machine learning became apparent when the scenario was more complex, with a maximum enhancement of 13.25% in prediction accuracy. Each machine learning method demonstrated distinct features and advantages in different aspects.

1. Introduction

Emergency evacuation is gaining increasing attention due to the high frequency of emergency situations and the severity of the consequences. Numerous incidents have been reported regarding overcrowding and crushing during emergency situations. They occurred in sport stadiums (*e.g.* 28 people killed in Cairo Egypt due to confrontation of fans with police in 2015), schools (*e.g.* 6 pupils killed and 25 injured in Kunming, China in 2014, which was triggered by the falling of a mattress in the corridor), social gathering places (*e.g.* 21 people killed in a nightclub incident in Chicago, USA in 2003, which was triggered by club security using pepper spray to break up a fight), shopping centers (*e.g.* 9 people injured in a stampede after sounds of popping balloons mistaken for gunfire in Orlando, USA in 2016) and other places. It has been observed that in many cases, injuries or deaths of the victims are actually caused by the crowd's non-adaptive behaviors, such as stampeding, pushing, knocking and trampling on others, rather than the actual cause (such as fire or explosion) [1,2]. This highlights the necessity and significance of a better understanding of evacuees' movement dynamics, which can be used to facilitate building safety design and evacuation management and planning.

Evacuees' movement dynamics can be viewed as the result of hierarchical decision-making processes, *i.e.* strategical-level (whether or not to initiate evacuation), tactical-level (route choice), and operational-level decision making process (stepwise movement) [3,4]. To characterize evacuation processes, a few physical models were developed, such as Social Force Model and Cellular Automata (CA) Model. These models were subsequently improved by taking into account some other factors such as the availability of instructions [5,6], evacuees' familiarity with exits [7], degree of collision avoidance [8] *etc.* To guide the setting of behavioral rules in these models, analysis of observational or experimental data appears to be a prerequisite [9–11]. On the other hand, statistical models such as logit models can help with the extraction of behavioral rules, by defining the influencing factors on evacuations and by describing the extents of their influence. Lovreglio et al. studied exit choice with the consideration of heterogeneity of decision makers, and found the significant influence brought by other decision makers [12]; A further study done by the same group showed that

* Corresponding author.
E-mail address: szqian@ntu.edu.sg (S. Qian).

https://doi.org/10.1016/j.firesaf.2019.04.008
Received 10 September 2018; Received in revised form 1 March 2019; Accepted 17 April 2019
Available online 24 April 2019

Table 1
Details of the experiment videos.

	Scenario number	Number of evacuees	Available exits	Exit widths
Video 1	9	143	1,2,3,4	w1 = w2 = w3 = w4 = 50 cm
Video 2	19	142	2,3,5	w2 = w3 = w5 = 50 cm

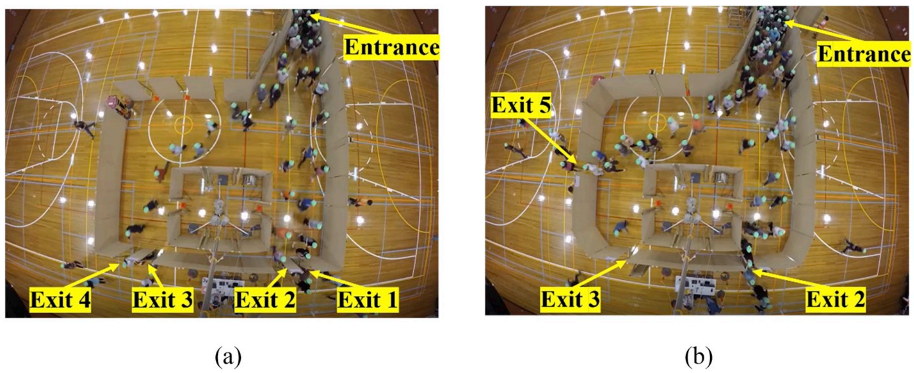

Fig. 1. Spatial layout of evacuation experiments in (a) Video 1 and (b) Video 2.

herding behavior in exit choice can be influenced by both environmental variables and personal ones, such as number of people close to the nearest exit, height of the decision maker *etc.* [13]; Aleksandrov et al. studied evacuees' choices between lifts and stairs and concluded that navigation signs and large number of people in stairs prompt evacuees to use lifts [14]. An experimental study carried out by Cao et al. subjected the evacuees to various visibility conditions. The numerical analysis showed that limited visibility would increase the likelihood for evacuees to follow neighbors or to help each other [15].

Different from strategic level and tactical level decision-making studies, stepwise movement dynamics are mostly analyzed in an aggregated way. At the operational level, evacuees make instantaneous decisions such as acceleration or stopping to avoid collision. This may affect evacuees' escaping characteristics and thus the whole evacuation process [16]. In many studies, evacuation velocity has been documented at group level, such as the average walking speed for a certain age or gender group [17–19]. Some researchers have further studied the influence of staircase height [20], visibility conditions [21], pedestrian load [22] *etc.* on individual speed, whereas such "individual speed" is an averaged value over a long range or the entire trajectory. These aggregated analyses may not be able to fully reflect microscopic evacuation phenomena. Therefore, following the idea of conducting short-range evacuation dynamics analysis [23], this study is focused on the rules governing short-range evacuation movement, with a wider consideration of environmental features, such as evacuees' internal features, features of the exits, and the behaviors of the surrounding evacuees.

Apart from the conventional statistical analysis techniques, machine learning is a promising approach for big data analysis. It is well-developed in computer science area and has been mostly applied in economics and medical science for its superiority in extracting the hidden rules from real-life data [24]. There have been studies using machine learning methods in human movement analyses. Support Vector Machine was used to identify walking conditions for hemiparetic persons [25]. Principal Component Analysis was applied to quantify human upright standing with some "principal movements" [26]. Neural Network was adopted to analyze one's horizontal and vertical normal walking velocity in relation to surrounding twenty-five people's vertical and horizontal velocities [27]. Cluster analysis was utilized to deduce police officers' patrol patterns [28]. However, there is a scarcity of studies that explore the potential and compare the performance of different machine learning methods in emergency evacuation studies.

This study aims to gain insight into evacuees' stepwise movement and key factors governing their movement patterns through a combination of machine learning methods based on two experimental video records. The rest of the paper is organized as follows: Section 2 describes the experimental setup and data collection methods; Section 3 introduces the general research methodology and the basics of the various methods; Sections 4 and 5 are respectively results from the analyses, and discussions; Section 6 summarizes the conclusions of the study.

2. Data collection

2.1. Data source

This study is based on two video records of quasi-emergency evacuation experiments carried out by Haghani and Sarvi. The aim of their experiments was to extract underlying rules that govern exit wayfinding of evacuees in multi-exit places. Around 150 participants were requested to enter the designated area and escape from the exits as fast as possible. The evacuation experiment was conducted for 20 times with different combinations of design parameters (Appendix A). The researchers tested the influence of spatial distance to exit, level of congestion around exits, others' exit choices *etc.* on exit selection [29]. In this study, we intended to study evacuation dynamics on a more microscopic level (operational level), *i.e.* the stepwise movement patterns and their influencing factors. Two available videos, Video 1 and Video 2 (details shown in Table 1 and Fig. 1.), were used for this study.

Supplementary video related to this article can be found at https://doi.org/10.1016/j.firesaf.2019.04.008

2.2. Coordinate mapping and trajectory extraction

A two-step mapping approach was adopted to map from the pixel coordinates to world coordinates, as large lens distortion was observed due to the wide angle of the GoPro Hero 4 Black Camera, placed at 8 m height to capture the evacuation process from a top-down view. Firstly, lens distortion was removed automatically by Adobe Photoshop CS6 with its embedded function. Subsequently, automatic linear modeling was applied to transform from pixel coordinates to world coordinates via Matlab 2018a. The border point (0 −500) has the largest cumulative mapping error, which is 28.343 cm. This results in a maximum error of 0.138 cm/pixel (= 28.343 cm/205 pixel).

Stepwise positions of each evacuee were tracked for further variable extraction. Although some programs claim the capability of recognizing objects and extracting trajectories automatically, despite multiple trials they were unable to give satisfactory results in this case due to high mobility and high density of the evacuees. Hence, a semi-automated extraction process was adopted, *i.e.* coordinates of the evacuees' heads (center of the hats) were obtained manually, and then automatically transformed from pixel coordinates to world coordinates using the automatic linear model mentioned before.

2.3. Variable selection

To capture the details of the evacuees' movement, a few variables were derived based on the evacuees' stepwise positions. These variables can be divided into two types: movement-related variables and environment-related ones (Appendix B). Movement-related variables are those reflecting an evacuee's stepwise movement pattern, such as velocity and acceleration rate. Environment-related variables are internal and external environmental features that may influence an evacuee's movement pattern, such as features of the exits (density and flowrate of evacuees *etc.*) [11,30,31], the evacuee's position (distance to exits, distance to obstacles *etc.*) [19,29], the local density [30], relative velocity with surrounding evacuees [27], surrounding evacuees' behaviors [13], the evacuee's intrinsic features (gender *etc.*) [11], velocity heterogeneity in the surroundings, and phase of evacuation. Most of these environment-related variables were adopted based on previous studies in which they have been shown to impact evacuation dynamics, in terms of the exit choice or the velocity. This study further tested whether these variables can influence stepwise movement patterns.

Hereinafter, we use the term "case" to refer to each stepwise movement of evacuees. Each evacuee's evacuation process was divided into several cases with a constant time interval of 0.2 s. Each case consists of movement-related variables and environment-related ones. It should be noted that this research focuses on operational-level movement decision analysis, which is assumed to be nested under tactical-level decision. In other words, in our study the target exit should remain constant for an evacuee during the evacuation. To minimize the influence of exit-changing actions, as well as to maximize the data volume that is applicable to the assumption, cases targeting Exit 1 in Video 1 and cases targeting Exit 2 in Video 2 were chosen for the study.

3. Methodology

This research follows an analytical procedure as shown in Fig. 2. Firstly, movement-related variables were used to classify movement patterns through Two-step Cluster Analysis. Similar cases were grouped together into specific movement patterns. Secondly, Principal Component Analysis was applied to identify the independent principal components, thereby discarding the irrelevant variables and merging interdependent variables. To find out how these principal components influence movement patterns, five approaches were used for comparison, namely Multinomial Logit Model, Decision Tree, Support Vector Machine, K-Nearest Neighbor, and Artificial Neural Network. All the data were normalized to the range of [0,1] and the case sequence was randomized to satisfy data requirement.

3.1. Two-step cluster analysis

To categorize stepwise movement patterns, cluster analysis was used. It can group cases into different clusters, so that the cases within one cluster are similar to one another and distinct from cases in other clusters [32], as shown in Fig. 3(a). The proximity between any two data points is gauged by distance measure (or similarity criteria), such as Euclidean distance, Manhattan distance, Log-likelihood distance *etc.* Normally, the clustering can produce a set of solutions, and the optimal one is selected based on clustering criteria such as Schwarz's Bayesian Criterion (BIC).

Two-step Cluster Analysis was chosen for this study because of its superiority over two other traditional cluster methods, K-means cluster analysis and hierarchical cluster analysis. Two-step Cluster Analysis can deal with large datasets (where the hierarchical cluster analysis is weak) without the number of clusters specified beforehand (which is required for the K-means cluster analysis). For more detailed information, please refer to Martinez et al.'s work [33]. In this study, the evacuees' stepwise movement velocity (V) and acceleration rate (A) were used for movement pattern categorization. Log-likelihood distance was chosen as the distance matrix and BIC was chosen as the clustering criteria.

3.2. Bivariate correlation test and Principal Component Analysis

To remove variables irrelevant to the identified movement patterns, Spearman's Rank Coefficient was used as a criterion. It can measure the relation between environment-related variables and the identified movement patterns. The variables with absolute Spearman's Rank

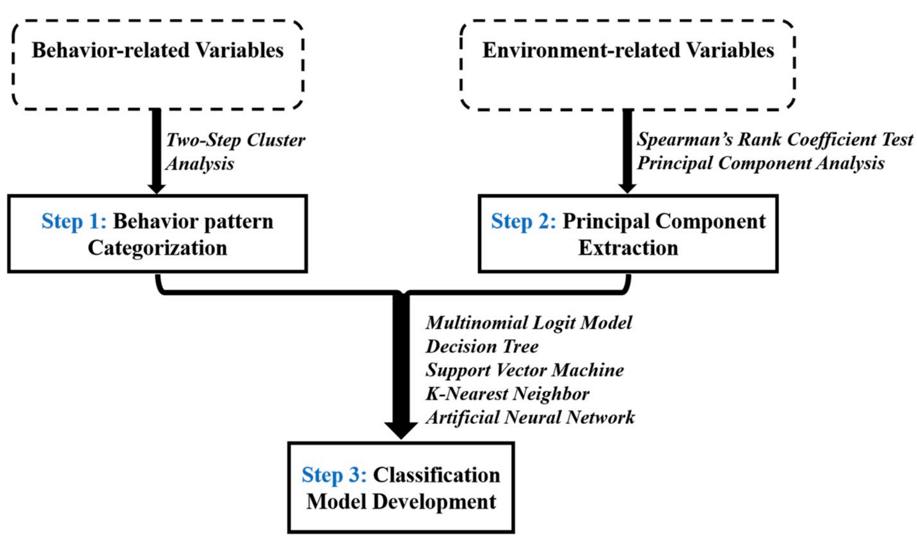

Fig. 2. Analytical procedure.

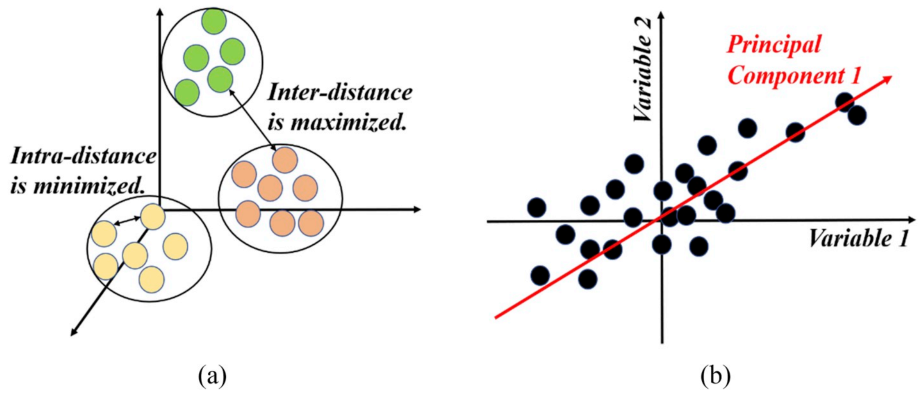

Fig. 3. Graphic representation of (a) Cluster Analysis and (b) Principal Component Analysis.

Coefficient larger than 0.3 were deemed relevant variables and held for further analysis.

Subsequently, Principal Component Analysis (PCA) was adopted to decrease variable dimension and remove the interdependency among variables. PCA is a dimension-reduction tool to transform possibly correlated variables into a smaller number of independent variables, known as principal components (PCs) [34]. These independent principal components can represent most of the information that the original variables contain. The direction with the largest variance represents the "most important" component as assumed by PCA, because all the data can be best separated in this dimension. In Fig. 3(b), the two-dimension data can be projected onto Principal Component 1 to achieve dimension reduction. Detailed mathematical illustration can be found in Smith's paper [35].

3.3. Modeling methods

All the cases (movement patterns and their corresponding principal components) were divided into two sets: training set and testing set. The training set was used to train the model (determining the coefficients for the Multinomial Logit Model, and parameters for the four machine learning models). The testing set was used to test the classification performance of the trained model. The predicted results from the testing set were then compared with the real experiment data, thus the final prediction accuracy was obtained.

3.3.1. Multinomial Logit Model

To find out the relationship between movement patterns and principal components, Multinomial Logit Model (MLM), a traditionally used model in evacuation studies, was used as the baseline for comparison. It can predict the relative possibility of an individual case belonging to a certain movement pattern as specified in Eqs. (1)–(3). More detailed information on the calculation of coefficients β_{jp} can be found in McCullagh and Nelder's paper [36].

$$In(P_{MP1}/P_{MPk}) = \alpha_1 + \beta_{11}{}^*PC_1 + \beta_{12}{}^*PC_2 + \ldots + \beta_{1p}{}^*PC_p \tag{1}$$

$$In(P_{MP2}/P_{MPk}) = \alpha_2 + \beta_{21}{}^*PC_1 + \beta_{22}{}^*PC_2 + \ldots + \beta_{2p}{}^*PC_p \tag{2}$$

$$In(P_{MP(k-1)}/P_{MPk}) = \alpha_{k-1} + \beta_{21}{}^*PC_1 + \beta_{(k-1)2}{}^*PC_2 + \ldots + \beta_{(k-1)p}{}^*PC_p \tag{3}$$

where P_{MPk}: the possibility of an outcome belonging to Movement Pattern k.

k: total types of movement patterns

β_{jp}: the coefficient, *i.e.* the effect of the Principal Components p on the relative risk of belonging to Movement Pattern j versus Movement Pattern k

PC_p: Principal Component p

3.3.2. Decision tree

Decision Tree (DT) was also applied to classify movement patterns based on the principal components. The tree structure consists of leaf nodes representing classification types (types of movement patterns in this study), internal nodes representing a test on an attribute, and branches representing the outcome of a test. The classification process is illustrated in Fig. 4(a).

To construct the tree, all the attributes are tried at each internal node. The attribute with the best classification performance is selected as the final attribute for this internal node. This process runs iteratively until it satisfies the classification requirement. The total number of internal nodes is called the number of splits.

3.3.3. Support Vector Machine

Support Vector Machine (SVM) classifies cases by finding the best hyperplane to separate them. The input variables (principal components) can form an n-dimensional space, where n equals the number of principal components. The distance between this hyperplane and the closest data point is called the margin. The hyperplane with the largest margin would be selected as the final model, such as Hyperplane 2 in Fig. 4(b).

Kernel function is a commonly used technique to increase computational efficiency, when dealing with complicated classification problems such as non-linear problems [37]. Four types of kernel functions are used in this study, namely linear kernel, quadratic kernel, cubic kernel, and Gaussian kernel. The scale of Gaussian kernel is set to be sqrt(p)/4, sqrt(p), sqrt(p)*4 for Fine Gaussian SVM, Medium Gaussian SVM and Coarse Gaussian SVM respectively, where the p is the number of principal components.

3.3.4. K-nearest neighbor

K-Nearest Neighbor (KNN) categorizes cases based on their nearest K points' classes [38]. K is the number of surrounding neighbors. A single case belongs to the class that the majority of its K neighbors belong to. For example, the black person in Fig. 4(c) belongs to blue persons' class if K equals 3, whereas it belongs to grey persons' class if K equals 10. Each case is determined by this way in an n-dimensional space where n is the number of attributes (principal components in this study).

The distance between data points can be calculated by many matrixes, such as Cosine distance, Cubic distance, Euclidean distance. Moreover, the value of K, *i.e.* number of nearest neighbors can vary according to the need. Here, a combination of different distance matrixes and different K values were set, namely Cosine KNN, Cubic KNN, Weighted KNN (Euclidean distance with specified distance weight), Fine KNN (Euclidean distance with K being equal to 1), Medium KNN (Euclidean distance with K being equal to 10), and Coarse KNN (Euclidean distance with K being equal to 100).

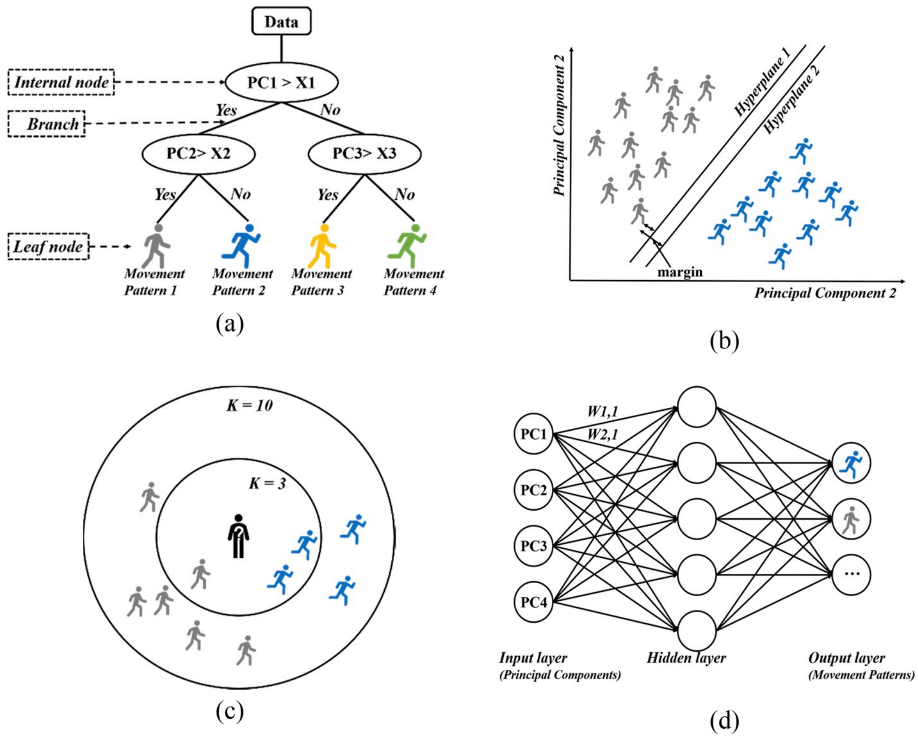

Fig. 4. Graphic representation of four machine learning methods with virtual examples for: (a) Decision Tree, (b) Support Vector Machine, (c) K-Nearest Neighbor, and (d) Artificial Neural Network. In all the figures, different colors of agents indicate different movement patterns. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.).

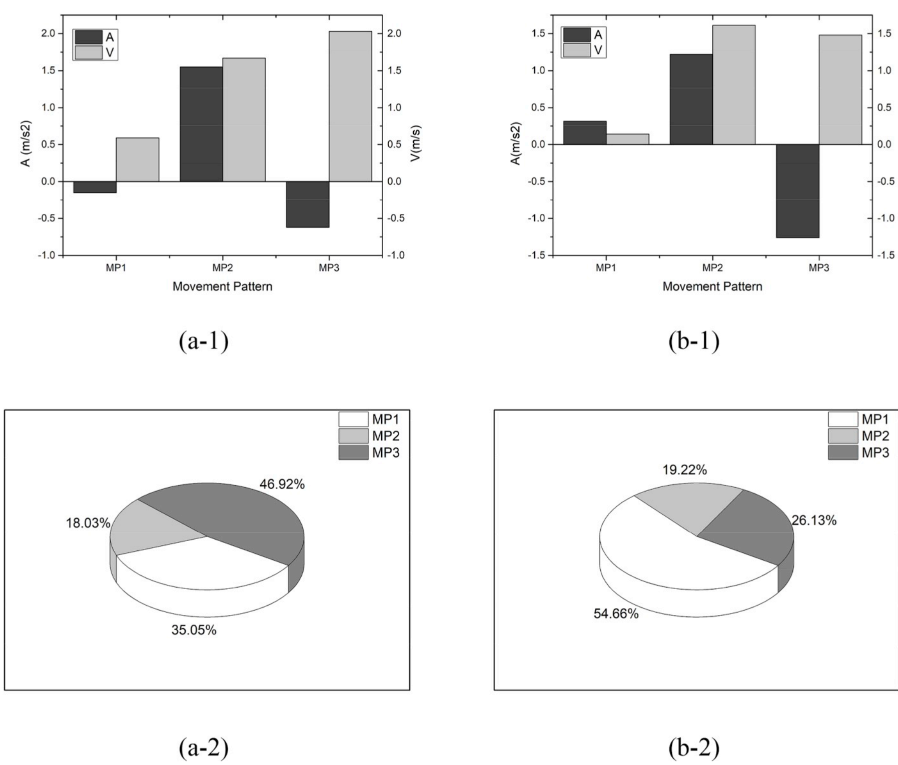

Fig. 5. Cluster centers of three movement patterns in (a-1) Video 1 and (b-1) Video 2. Proportion of cases belonging to each of the three movement patterns in (a-2) Video 1 and (b-2) Video 2.

3.3.5. Artificial Neural Network

Artificial Neutral Network (ANN) is a computation model inspired by biological systems such as the brain. It imitates the human thinking process with many layers of interconnected neurons [39]. Neural networks consist of input layers, hidden layers, and output layers, representing respectively the input (principal components), hidden rules, and output (movement patterns), as shown in Fig. 5(d). Information from the preceding layer can be delivered to the following layer. Weights of each layer (*e.g.* $W_{1,1}$ *and* $W_{2,1}$) are adjusted according to each layer's classification performance. This process goes iteratively to achieve a satisfactory accuracy.

Here a basic type of neural network was applied, *i.e.* the two-layer feed-forward network, with an output layer and a hidden layer. The number of neurons in the hidden layer was set to be 5, 10, 15, and 20 respectively.

4. Result and analysis

We performed analysis on two videos. Observably, evacuation process is more complicated in Video 2, as there are fewer available exits and the distribution of exits is asymmetric. Moreover, there are more right-angled egress paths in Video 2 due to the location of Exit 5, which works less efficiently compared with straight paths during collective panic egress [40].

Results showed that in both videos, the cases were categorized into three movement patterns. The influencing factors on these movement patterns were identified: three common principal components were identified in both videos, with an additional one for Video 2. The reason for one more principal component identified in Video 2 may be caused by the higher complexity of the scenario in Video 2, which will be discussed in Section 4.2.2. The performance of the five classification models for both videos showed reasonable consistency. The analytical procedures for Videos 1 and 2 are shown in Sections 4.1 and 4.2 respectively.

4.1. Analysis of Video 1

4.1.1. Movement pattern categorization

Three clusters were identified from Two-step Cluster Analysis as shown in Fig. 5(a). 35.05% of the cases belong to Movement Pattern 1, "Constant low velocity MP". In 18.03% cases, evacuees accelerate (mean = 1.55 m/s2) on the basis of a high velocity (mean = 1.67 m/s), which are categorized into MP2, "High velocity with acceleration MP". As for the remaining 46.92% cases, evacuees decelerate (mean = −0.62 m/s2) on the basis of a high velocity (mean = 2.03 m/s), demonstrating the so-called "High velocity with deceleration MP" (MP3).

4.1.2. Principal component extraction

To remove irrelevant variables, Spearman's Rank Coefficient was calculated between "Type of Movement Pattern" and environment-related variables as the criterion. Thirteen variables were kept for further PCA as shown in Appendix C. It was found that evacuees tend to be more sensitive to other's acceleration action than the deceleration action, because "V_Accele_Sur" (number of accelerating actions in the surroundings) was kept for PCA whereas "V_Decele_Sur" (number of decelerating actions in the surroundings) was discarded after Spearman's Rank Coefficient test.

In the PCA process, three Principal Components (PCs) were extracted to represent the majority (86.279%) of the information conveyed by the thirteen variables (Appendix D). As shown in Table 2, PC1 mainly contains distance information, *e.g.* "Dist_E1" (0.960), "Dist_E2" (0.967), "Dist_E3" (0.977), "Dist_E4" (0.977), "Dist_Obstacle" (0.968), "Ind_Time" (−0.903), and "V_Accele_Sur" (0.719). PC2 represents one's velocity competitiveness, as it relates closely with "Sum_ΔV_Sur" (0.964), "Ave_ΔV_Sur" (0.947), and "Num_Faster" (−0.785). PC3 depicts direction changing actions in the surroundings, as it related closely with "D_Constant_Sur" (−0.967).

4.1.3. Training and testing of classification models

To find out the rules governing evacuees' movement dynamics, *i.e.* to see how these identified principal components can influence movement patterns, Multinomial Logit Model and machine learning methods, namely Decision Tree, Support Vector Machine, K-Nearest Neighbor, and Artificial Neural Network, were applied for comparison.

4.1.3.1. Multinomial Logit Model. Base on the model specified in Section 3.3.1, 700 cases were randomly selected as the training set to obtain the coefficient estimates as shown in Table 3 and Eq. (4) – (6). A final prediction accuracy of 79.27% was obtained using the rest 193 testing cases.

$$In(P(MP2)/P(MP1)) = -0.3572 + 3.0350*PC1 + 1.6885*PC2 + 1.4449*PC3 \tag{4}$$

$$In(P(MP3)/P(MP1)) = 1.0027 + 1.9621*PC1 + 2.5238*PC2 + 1.8114*PC3 \tag{5}$$

$$In(P(MP2)/P(MP3)) = -1.3599 + 1.0729*PC1 - 0.8353*PC2 - 0.3665*PC3 \tag{6}$$

Some features of an evacuee's velocity can be identified through the analysis of the coefficients in Eqs. (4) and (5). All the PCs have very pronounced effect in distinguishing MP2 and MP1, similarly for the differentiation between MP3 and MP1. The consistence of positive signs of coefficients in Eqs. (4) and (5) indicates that larger PC1, PC2, and PC3 can increase the possibility of movement pattern falling into MP2 and MP3, both of which exhibit "higher velocity" compared with MP1. Tracing back to the aggregated meaning of PCs as shown in Table 2, higher PC1 means a larger distance from the current position to the target exit and obstacles, an earlier phase in a route, and more surrounding evacuees taking acceleration actions. Hence the positive

Table 2
Rotated component matrix by principal component analysis of environment-related variables[a]. The numbers represent the relation between variables and principal components. The numbers with large absolute values are marked in bold.

Variable	Definition	Principal Component		
		1	2	3
Dist_E3	Distance from current position to Exit 3	**.977**	.115	.139
Dist_E4	Distance from current position to Exit 4	**.977**	.113	.137
Dist_Obstacle	Distance from current position to nearest obstacle	**.968**		
Dist_E2	Distance from current position to Exit 2	**.967**	.161	.160
Dist_E1	Distance from current position to Exit 1	**.960**	.182	.164
Ind_Time	Indicator to depict the phase of route (cumulative time divided by total time spent in a route)	**-.903**	-.287	
V_Accele_Sur	No. of evacuees with obvious acceleration action in one's surroundings	**.719**		
Sum_ΔV_Sur	Overall velocity competitiveness compared with one's surrounding evacuees	.139	**.964**	
Ave_ΔV_Sur	Average velocity competitiveness compared with one's surrounding evacuees	.162	**.947**	
Num_faster	No. of evacuees with higher velocity in one's surroundings		**-.785**	-.553
D_Constant_Sur	No. of evacuees maintaining their direction in one's surroundings			**-.967**
V_Constant_Sur	No. of evacuees maintaining velocity in one's surroundings	-.212		**-.943**
D_Change_Sur	No. of evacuees with direction changing action in one's surroundings	-.270	-.228	-.326

[a] KMO measure of sampling adequacy = 0.673, indicating the adequacy of sample size; Bartlett's Test of Sphericity: Sig. = 0.000, indicating the suitability for dimension reduction.

Table 3
Multinomial logit estimates (reference category: Movement Pattern 1), where MP1, MP2, MP3 stands for "Constant low velocity" MP, "High velocity with acceleration" MP, "High velocity with deceleration" MP respectively.

Parameter	Video 1(Video 1)		Video 2 (Video 2)	
	Coefficient estimates (MP2/MP1)	Coefficient estimates (MP3/MP1)	Coefficient estimates (MP2/MP1)	Coefficient estimates (MP3/MP1)
Intercept terms	−0.3572 (−1.3816) *	1.0027 (5.0403) **	−1.1709 (−10.0342)***	−0.3208 (−3.6668) **
PC1	3.0350 (10.9034)***	1.9621 (8.1593)***	2.5576 (19.8200)***	1.8262 (15.1224)***
PC2	1.6885 (6.0640)***	2.5238 (10.4809)***	1.5250 (12.1096)***	1.8423 (15.4619)***
PC3	1.4449 (5.9470)***	1.8114 (9.6619)***	1.6291 (14.3748)***	1.3682 (14.8161)***
PC4			1.3981 (12.4267)***	1.2853 (13.3068)***

***, **, * indicate the statistical significance at 99%, 95% and 70% level respectively. The number in the brackets reflects the t-statistic of each estimate.

impact of PC1 on higher velocity is intuitive: at the beginning ("an earlier phase of a route" and "a larger distance from target exit"), evacuees tend to run fast to take a favorable position. Furthermore, "a larger distance from obstacles" creates an obstruction-free environment for the evacuees and "more evacuees accelerating" is also a positive stimulus because an individual evacuee would more likely adopt a higher velocity to avoid falling behind [41]. Besides PC1, the positive effect of PC2 on higher velocity is also self-evident as higher PC2 means stronger velocity competitiveness compared with one's surrounding evacuees. Larger PC3 can give rise to a higher velocity because fewer evacuees' changing their directions can be viewed as a strong affirmation for the current exit choice as supported by "follow-the-majority" theory [42], so that the evacuees would not hesitate or wander but head directly to their target.

Apart from velocity, acceleration and deceleration action also follow certain rules as demonstrated by Eq. (6). The possibility of any movement pattern being classified as MP2 or MP3 is compared in Eq. (6), where the two MPs are characterized both by a relatively high velocity but by opposite trends in velocity changing. It is clear that PC1 has greater influence on the acceleration behavior than other PCs. Higher PC1 can increase the occurrence rate of acceleration actions, *i.e.* the chance of a case falling into MP2. Reasons are as follows: at the beginning of an evacuation, evacuees tend to accelerate to evacuate as fast as possible. Moreover, under the pressure of acceleration actions in the surroundings, one is more likely to accelerate to catch up with others. As for PC2, stronger velocity competitiveness will decrease the likelihood of acceleration. This may be because one has reached the upper limit of velocity under this situation, and/or he/she prefers to stay close with his/her group.

4.1.3.2. Machine learning methods. Both machine learning methods and MLM showed comparable prediction accuracy (Table 4). Among the three subtypes within the DT model, Medium Tree gave the highest prediction accuracy (78.10%). Most of the SVMs had a slightly better classification performance than DTs, nevertheless they were more computationally expensive. Medium Gaussian SVM (81.50%) showed the highest prediction accuracy among all the SVMs, followed by Quadratic SVM (81.10%) and Cubic SVM (80.40%). However, training time varied greatly among SVMs, and Gaussian Kernel was more computationally efficient than Linear Kernel, Quadratic Kernel and Cubic Kernel in multiclass classification. KNNs were more efficient in calculation efficiency compared with other methods, but still with comparable accuracy. Weighted KNN outperformed other KNNs in terms of both prediction accuracy (79.80%) and training time (0.44 s). The prediction accuracies and training times for all the KNNs did not show much difference, which proved the robustness of this method. The prediction accuracy and computation efficiency of ANNs were relatively high and stable. Model with highest prediction accuracy was ANN with 10 hidden neurons (81.41%).

4.2. Analysis of video 2

4.2.1. Movement pattern categorization

Similar to the categorization for Video 1, three clusters were identified as shown in Fig. 6(b). Over half of the cases (54.62%) belong to the Movement pattern (MP) 1, "Constant low velocity" MP, *i.e.* evacuees with low velocity (mean = 0.32 m/s) and low acceleration rate (mean = 0.14 m/s2). In 19.20% cases, evacuees accelerate (mean = 1.22 m/s2) on the basis of a high velocity (mean = 1.61 m/s), which is called MP2, "High velocity with acceleration" MP. As for the remaining 26.18% cases, evacuees decelerate (mean = −1.26 m/s2) with a high velocity (mean = 1.48 m/s), the so-called "High velocity with deceleration" MP (MP3).

4.2.2. Principal component extraction

Thirteen variables were kept for PCA analysis after Spearman's Rank Coefficient Test (Appendix C). Most of the variables are consistent with those for Video 1. "Sum_Abs_ΔV_Sur" (overall velocity heterogeneity in one's surroundings) and "Ave_Abs_ΔV_Sur" (average velocity heterogeneity in one's surroundings) were additionally picked in this video, indicating the velocity heterogeneity in the surroundings may be an important factor. In contrast, number of surrounding evacuees' "maintaining constant velocity" ("V_Constant_Sur") was not kept for Video 2. This suggests that "V_Constant_Sur" might be a less significant factor in describing surrounding evacuees' velocity change action than number of surrounding evacuees' "taking acceleration action" ("V_Accele_Sur"), which was kept for both videos.

Notably, the discarded factors in both two videos should not be deemed unimportant in real emergency evacuations, as evacuation study is highly context-dependent [10]. This study is based on a small-geometric scale, slightly high-density, indoor emergency evacuation scenario. Some researchers concluded that males and females ("Ind_Gender") have different walking velocities during evacuations [43], which is not supported by our study. It is plausible that the male participants were not able to take full advantage of their physical strengths with the short evacuation distance and the high density of evacuees. Evacuee density near the target exit was also found to influence one's route choice [44], so as the individual and the overall crowd velocity [45]. Nevertheless, the density near exits was rather high during most of the time, which is not well controlled in this study. To tackle this unavoidable issue encountered by most of the video-based studies, additional careful data gathering work is needed.

In the PCA process, four principal component (PC)s were extracted to represent 83.075% of the information (Appendix E). As shown in Table 5, the first three principal components convey similar meanings as those in Video 1. The fourth principal component is related closely with the sum and the average value of the velocity differences between the evacuee and his/her surrounding evacuees. Compared to that in Video 1, the evacuation process in Video 2 is more complicated. Route changing behavior, detouring behavior, and overtaking behavior are more frequently found in Video 2. In other words, the evacuation went less smoothly. This macro chaos is reflected in a micro factor, PC 4, the

Table 4
Modeling accuracy and training time for three machine learning models. The training time was the processing time by the Windows 10 Enterprise 64-bit, Intel (R) Xeon(R) CPU E5-1630v3.

Model Type	Model Subtype	Video 1 (Run 9)		Video 2 (Run 19)	
		Accuracy	Training Time/s	Accuracy	Training Time/s
DT	Fine Tree	74.80%	0.81	72.70%	0.80
	Medium Tree	**78.10%**	0.60	**74.60%**	3.14
	Coarse Tree	72.00%	0.51	72.00%	3.02
SVM	Linear SVM	78.60%	2.20	75.30%	9.73
	Quadratic SVM	81.10%	5.51	76.40%	181.52
	Cubic SVM	80.40%	41.48	74.30%	387.50
	Fine Gaussian SVM	80.10%	1.98	75.40%	7.16
	Medium Gaussian SVM	**81.50%**	2.46	**77.70%**	7.96
	Coarse Gaussian SVM	78.90%	2.40	75.80%	8.93
KNN	Fine KNN	76.40%	1.13	71.50%	3.88
	Medium KNN	79.50%	0.88	**76.60%**	3.74
	Coarse KNN	77.80%	0.75	76.50%	3.69
	Cosine KNN	78.90%	0.65	75.10%	3.64
	Cubic KNN	79.20%	0.53	76.40%	3.96
	Weighted KNN	**79.80%**	0.44	76.00%	3.52
ANN	Two-layer feed-forward network (5 hidden neurons)	80.85%	1.11	76.45%	1.54
	Two-layer feed-forward network (10 hidden neurons)	**81.41%**	1.34	77.12%	1.56
	Two-layer feed-forward network (15 hidden neurons)	81.01%	1.27	77.24%	1.78
	Two-layer feed-forward network (20 hidden neurons)	81.23%	1.64	**77.96%**	1.68
MLM		79.27%	Not applicable	64.71%	Not applicable

velocity heterogeneity in one's surroundings.

4.2.3. Training and testing of classification models

4.2.3.1. Multinomial Logit Model. In Video 2, 2000 cases were randomly selected as the training set to obtain the coefficient estimates as shown in Table 3 and Eq. (7) – (9). A final prediction accuracy of 64.71% was obtained using the rest 578 testing cases.

$$In(P(MP2)/P(MP1)) = -1.1709 + 2.5576{*}PC1 + 1.5250{*}PC2 + 1.6291{*}PC3 + 1.3981{*}PC4 \tag{7}$$

$$In(P(MP3)/P(MP1)) = -0.3208 + 1.8262{*}PC1 + 1.8423{*}PC2 + 1.3682{*}PC3 + 1.2853{*}PC4 \tag{8}$$

$$In(P(MP2)/P(MP3)) = -0.8501 + 0.7314{*}PC1 - 0.3173{*}PC2 + 0.2609{*}PC3 + 0.1128{*}PC4 \tag{9}$$

The results show that the coefficients are in concordance with those derived from Video 1. All the PCs show pronounced effect in distinguishing MP2/MP3 versus MP1, while only PC1 works well in distinguishing MP2 and MP3. As mentioned previously, PC4 stands for the velocity heterogeneity. PC4 is positively correlated with the chance of a case falling into MP2 or MP3 (both high velocity but with acceleration or deceleration). A possible explanation is that a higher PC4 represents a more dispersed velocity distribution. This is more likely to occur in an evacuation scenario with less hindrance such that the evacuees can run faster than in a jammed situation. This speculation needs to be further tested on more evacuation scenarios with different complexities.

4.2.3.2. Machine learning methods

Different from results obtained from Video 1, machine learning methods demonstrated higher prediction accuracy of up to 13.25% compared with Multinomial Logit Model. Medium Tree (74.60%). Medium Gaussian SVM (77.70%), Medium KNN (76.60%), and ANN with 20 hidden neurons (77.96%) showed the best prediction performance among their respective types.

Medium Tree was plotted to gain some insight into the relationship between MPs and PCs (Fig. 6). PC1 was identified as the first node to extend the branches with a threshold value of 0.468. PC2 came afterwards to further differentiate the cases. This indicates that PC1 and PC2

Table 5
Rotated component matrix by principal component analysis of environment-related variables. The numbers represent the relation between variables and principal components. The numbers with large absolute values are marked in bold.

Variables	Definitions	Component			
		1	2	3	4
Dist_E3	Distance from current position to Exit 3	0.963		0.206	
Dist_E2	Distance from current position to Exit 2	0.941		0.136	0.177
Dist_Obstacle	Distance from current position to nearest obstacle	0.903	0.105	0.181	0.100
Dist_E5	Distance from current position to Exit 5	0.763	0.134	0.316	−0.233
Ind_T	Indicator to depict the phase of route (cumulative time divided by total time spent in a route)	**−0.702**		0.102	−0.414
V_Accele_Sur	No. of evacuees with obvious acceleration action in one's surroundings	**0.622**		−0.123	0.287
Sum_ΔV_Sur	Overall speed competitiveness compared with one's surrounding evacuees	0.114	**0.943**		0.220
Ave _ΔV	Average speed competitiveness compared with one's surrounding evacuees	0.132	**0.925**		0.209
Num_Faster	No. of evacuee with higher speed in one's surroundings		**−0.860**	−0.220	0.228
Sum_Abs_ΔV_Sur	Overall speed heterogeneity in one's surroundings	0.123	0.186	0.187	**0.905**
D_Constant_Sur	No. of evacuees maintaining their direction in one's surroundings	0.404	−0.128	−0.152	**0.748**
Ave_ Abs_ΔV	Average speed heterogeneity in one's surroundings		0.278	0.535	**0.724**
D_Change_Sur	No. of evacuees with direction changing action in one's surroundings	−0.331		**−0.829**	−0.139

*KMO measure of sampling adequacy = 0.617, indicating the adequacy of sample size; Bartlett's Test of Sphericity: Sig. = 0.000, indicating the suitability for dimension reduction.

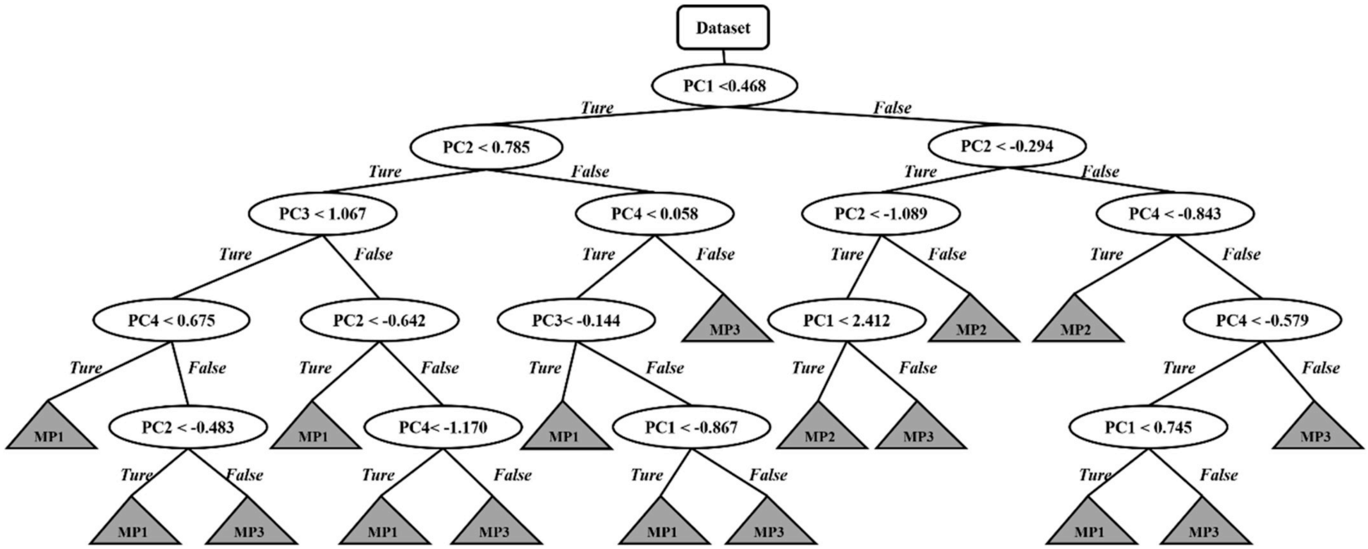

Fig. 6. The flow-chart representation of Medium Decision Tree Model for Video 2.

are the two most important PCs to classify the three MPs, which is consistent with the results shown in Table 3. PC2, PC3, and PC4 are then used as the classification criteria in the third step, followed by a messier classification process in the fourth step, suggesting that PC3 and PC4 are not as powerful as PC1 and PC2 in the differentiation of MPs.

4. Discussion

5.1. Comparison of the models used for MP classification

To the best of the authors' knowledge, this study is one of the few studies exploring the potential of DT, SVM, KNN, and ANN in the analysis of movement dynamics under emergency evacuations. Noticeable enhancement in the prediction accuracy compared with MLM proves the feasibility and the capability of machine learning methods in movement dynamics analysis. Slight variation in the prediction accuracy can be viewed as evidence of the robustness of the movement pattern categorization and the principal component extraction process. It can be further concluded that for both videos, ANN shows the best prediction performance in terms of prediction accuracy and stability. SVM has relatively high prediction accuracy whereas large variation in the training time. KNN has slightly lower prediction accuracy than SVM. The training time for KNN varies with data complexity: when the data are relatively simple highest training speed is achieved among all the methods (Video 1), while the speed is intermediate when the scenario is more complex (Video 2). As for DT, although it does not show superiority in accuracy or training speed, it is the most transparent and understandable method. This is because it has an interpretable classification process as shown in Fig. 6. These results largely agree with the conclusions from recent studies [46,47].

Compared to the aforementioned machine learning methods, the traditional model, MLM is nevertheless superior in its transparency and compatibility. MLM is comprehensible as it has a transparent processing process and a clear statistical meaning, whereas most machine learning methods function like black boxes. This is one of the reasons why MLM has been widely used in decision-making studies [48,49]. Moreover, the compatibility with other simulation models like CA model makes this model an ideal choice for depicting the decision-making process in evacuation simulations.

5.2. Theoretical and practical implications of the study

By analyzing influence of the environment on stepwise movement patterns of evacuees, this study can contribute to the understanding of group evacuation dynamics and provide insight for evacuation management. Being in the earlier phase of an evacuation and a stronger acceleration tendency in the surroundings can be stimuli for an evacuee's acceleration decision. This indicates that the movement patterns of evacuees are more diverse at the early phase of an evacuation, thus more measures should be taken to maintain order and avoid chaos at this phase. Also, a similar analytical procedure can be adopted in studies of collision avoidance behaviors, overtaking behaviors, exit selection *etc.*

Moreover, machine learning based approaches can better reflect the reality, as less prior knowledge is required for the data processing. Unsupervised cluster analysis procedure can classify movement patterns based on data structure, rather than researchers' preconception on the number of clusters. Principal Component Analysis can extract independent principal components from a large number of possibly correlated variables. In this sense, the bias induced by the variable selection and the computational risk induced by the inter-dependency among variables can be reduced.

Finally, this study provides a cost-effective big data analysis approach for further studies on evacuation movement rule extraction. This analytical procedure is applicable to the cases where the relationship between typical behaviors and their influencing factors remains to be explored. The “behaviors” include all possible actions an evacuee may take during emergency evacuations, such as strategic level behaviors (initiating the evacuation action *etc.*), tactical level behaviors (exit choice, exit-changing action *etc.*), and operational level behaviors (acceleration, collision avoidance action *etc.*). Fig. 7 is a graphic representation of application of the proposed method in the analysis of “exit-changing behavior”. Starting from defining the variables describing evacuation behaviors and their possible influencing variables, the proposed approach can reveal their relationship by the learning of data. With the growth of data sharing, the value of big data analysis via adoption of machine learning methods, would be increasingly appreciated by researchers in the area of evacuation study.

5.3. Limitations of the study and future work

In this study, one single machine learning method was applied at one time, whereas each of them demonstrated its superiority over

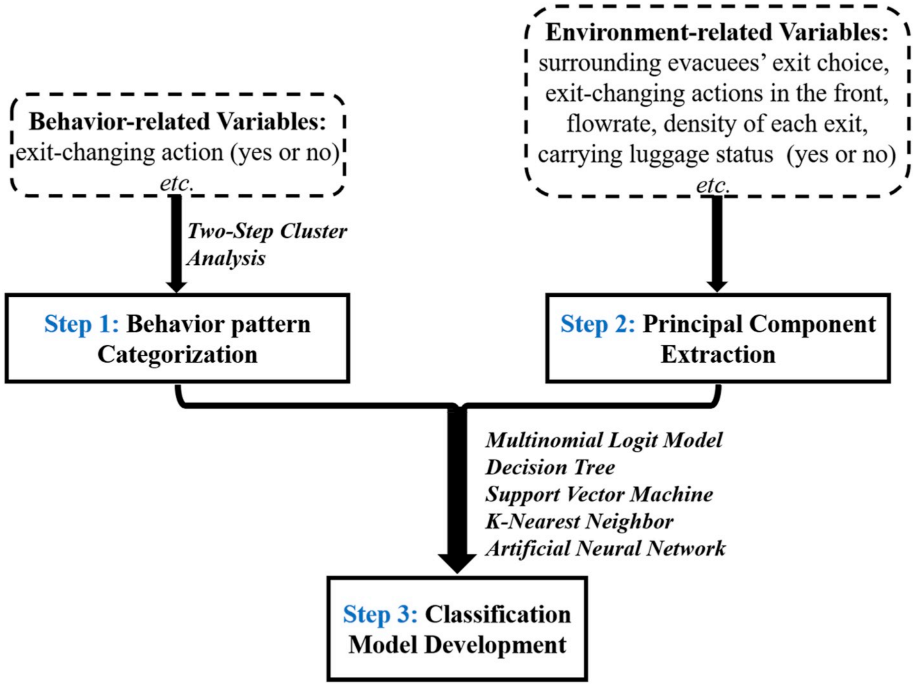

Fig. 7. A schematic diagram of application of the proposed method in the analysis of "exit-changing behavior".

others in different aspects. More complicated network structures and other algorithms for parameter learning such as backpropagation can be attempted since the basic ANN performed well in our study. Another direction is towards the application of decision tree-based ensemble learning methods, as they have been shown to give better prediction performance compared with single learning methods [50]. Bagging, Boosting, and Stacking ensemble learning methods can be compared to obtain an optimal model with higher interpretability, accuracy, training velocity, and stability.

Internal features of the evacuees such as age, risk perception, sense of direction *etc.* have not been taken into our consideration. The hypothesis that stronger velocity competitiveness will decrease the likelihood of acceleration has not been validated either. To address these problems, questionnaires and corresponding psychological experiments are needed to gather information on the evacuees' personal traits and cognitive processes. Future work will focus on the influence of evacuees' internal features on evacuation movement dynamics.

The error introduced in the evacuee trajectory tracing process may also influence the accuracy of this study. To improve the data extracting efficiency and accuracy, a semi-automatic approach (automatic tracking with manual validation/optimization) could be used as an alternative for small-scale evacuation analysis. For large-scale evacuation analysis, WIFI-based approach and radar-based approach may also provide solutions to this problem [51,52].

Finally, this study is based on a quasi-emergency scenario, to which a superior alternative would be an actual emergency evacuation. Also, space configuration of more diversity can further corroborate the applicability of the proposed method and the generality of the conclusions. Future study will evaluate the efficacy of machine learning methods/models in the analysis of actual emergency evacuation. The results obtained from the two kinds of scenarios can be used for comparison to study the behavioral differences under quasi-emergency and actual emergency situations.

6. Conclusion

Using existing videos, this study aims to gain insight into evacuees' stepwise movement dynamics through machine learning methods. The results from both videos are consistent and the major conclusions are shown below:

(1) Evacuees' stepwise movement is complex and is hard to be precisely predicted. The prediction accuracy in this study reached as high as 81.50%. To improve the accuracy, more intrinsic features of evacuees need to be considered in future study along with improvement in other aspects, such as application of ensemble learning methods and trajectory positing tools.
(2) Several rules of movement dynamics were deduced: (a) distances to the target and to obstacles are the most important aspects that influence an evacuee's stepwise movement pattern. When an evacuee is at the early stage of the route, and far from the exits or obstacles, he/she is more likely to accelerate. (b) The influence of surrounding evacuees' actions on a single evacuee's movement pattern is significant. Evacuees are more sensitive to others' acceleration actions than deceleration actions. Surrounding evacuees' changes in direction can be viewed as a negative stimulus to one's velocity.
(3) Machine learning methods outperformed MLM in terms of prediction accuracy with a maximum enhancement of 13.25% when the scenario was more complex, whereas the improvement was not obvious under the simpler scenario. Among the methods we used, ANN has the highest prediction accuracy and stability. SVM can also achieve high prediction accuracy despite long training time. KNN demonstrates an acceptable prediction performance, which is related to data complexity, and DT is advantageous in its interpretability.

Acknowledgments

This study is conducted under the financial support from NTU (Nanyang Technological University) Research Scholarship. The authors thank Mr. Sarvi's research group (University of Melbourne) for their generous sharing of the video records. We also thank Mr. Yushu Chen (National University of Singapore) for critical discussion and comments on the manuscript.

Appendix

Appendix A
Details of the trail runs of experiments [29].

Scenario number	Number of evacuees*	Available exits	Exit widths
1	75	1,2,3,4	w1 = w2 = w3 = w4 = 100 cm
2	75	1,2,3,4	w1 = w2 = w3 = w4 = 100 cm
3	75	1,3,4	w1 = w3 = w4 = 100 cm
4	75	1,3,4	w1 = w3 = w4 = 100 cm
5	75	1,2,4	w1 = w2 = w4 = 100 cm
6	75	1,2,4	w1 = w2 = w4 = 100 cm
7	150	1,2,3,4	w1 = w2 = w3 = w4 = 100 cm
8	150	1,2,3,4	w1 = w2 = w3 = w4 = 100 cm
9	150 (Video 1)	1,2,3,4	w1 = w2 = w3 = w4 = 50 cm
10	150	1,2,3,4	w1 = w2 = w3 = w4 = 50 cm
11	150	1,3,4	w1 = w3 = w4 = 50 cm
12	150	1,2,3	w1 = w2 = w3 = 50 cm
13	150	2,3	w2 = w3 = 50 cm
14	150	1,4	w1 = w4 = 50 cm
15	150	2,5	w2 = w5 = 50 cm
16	150	2,5	w2 = w5 = 50 cm
17	150	2,5	w2 = 50 cm, w5 = 100 cm
18	150	2.5	w2 = 100 cm, w5 = 50 cm
19	150 (Video 2)	2,3,5	w2 = w3 = w5 = 50 cm
20	150	2,3,5	w2 = w5 = 50 cm, w3 = 100 cm

*The number of evacuees in each run may be slightly different from the reported number in the actual experiment.

Appendix B
Details of the variables.

Type	Subtype	Variable	Type of variable	Definition	Calculating methods
Movement-related variables	Evacuee j's movement	$V(i,j)$	Scale (m/s)	Velocity of evacuee j at time step i	Euclidean distance between adjacent steps divided by time interval (0.2 s)
		$A(i,j)$	Scale (m/s^2)	Acceleration rate of evacuee j at time step i	Velocity difference between adjacent steps divided by time interval (0.2 s)
Environment-related variables	Proximity to exits [29]	$Dist\_E1(i,j)$	Scale (m)	Distance from evacuee j to Exit 1 at time step i	Euclidean distance from evacuee j to Exit 1 at time step i
		$Dist\_E2(i,j)$	Scale (m)	Distance from evacuee j to Exit 2 at time step i	Euclidean distance from evacuee j to Exit 2 at time step i
		$Dist\_E3(i,j)$	Scale (m)	Distance from evacuee j to Exit 3 at time step i	Euclidean distance from evacuee j to Exit 3 at time step i
		$Dist\_E4(i,j)$	Scale (m)	Distance from evacuee j to Exit 4 at time step i	Euclidean distance from evacuee j to Exit 4 at time step i
		$Dist\_E5(i,j)$	Scale (m)	Distance from evacuee j to Exit 5 at time step i	Euclidean distance from evacuee j to Exit 5 at time step i
	Density of exits [11]	$Densi\_E1(i)$	Scale (No. of people)	Density of Exit 1 at time step i	No of evacuee in the semicircle of Exit 1 (radius = 1.5 m) at time step i
		$Densi\_E2(i)$	Scale (No. of people)	Density of Exit 2 at time step i	No of evacuee in the semicircle of Exit 2 (radius = 1.5 m) at time step i
		$Densi\_E3(i)$	Scale (No. of people)	Density of Exit 3 at time step i	No of evacuee in the semicircle of Exit 3 (radius = 1.5 m) at time step i
		$Densi\_E4(i)$	Scale (No. of people)	Density of Exit 4 at time step i	No of evacuee in the semicircle of Exit 4 (radius = 1.5 m) at time step i
		$Densi\_E5(i)$	Scale (No. of people)	Density of Exit 5 at time step i	No of evacuee in the semicircle of Exit 5 (radius = 1.5 m) at time step i

Type	Subtype	Variable	Type of variable	Definition	Calculating methods
Environment-related variables	Flowrate around exits [30,31]	$FlowRate\_E1(i)$	Scale (No. of people)	Flow rate of Exit 1 at time step i	No of evacuee who passed the gate of Exit 1 at time step i
		$FlowRate\_E2(i)$	Scale (No. of people)	Flow rate of Exit 2 at time step i	No of evacuee who passed the gate of Exit 2 at time step i
		$FlowRate\_E3(i)$	Scale (No. of people)	Flow rate of Exit 3 at time step i	No of evacuee who passed the gate of Exit 3 at time step i
		$FlowRate\_E4(i)$	Scale (No. of people)	Flow rate of Exit 4 at time step i	No of evacuee who passed the gate of Exit 4 at time step i
		$FlowRate\_E5(i)$	Scale (No. of people)	Flow rate of Exit 5 at time step i	No of evacuee who passed the gate of Exit 5 at time step i
	Proximity to obstacle [19]	$Dist\_Obstacle(i, j)$	Scale (m)	Distance from evacuee j to nearest obstacle at time step i	Euclidean distance from evacuee j to nearest obstacle at time step i
	Local density [30]	$Densi\_Sur(i, j)$	Scale (No. of people)	Density of evacuee j's surroundings at time step i	No of evacuees in the surrounding circle of evacuee j at time step i. Circle Radius: 2.5 m [27].
	Velocity heterogeneity in one's surroundings	$Sum\_Abs\_\Delta V\_Sur(i, j)$	Scale (m/s^2)	Overall velocity heterogeneity in evacuee j's surrounding area at time step i	Sum of absolute velocity difference of evacuee j and his surrounding evacuees at time step i
		$Ave\_Abs\_\Delta V\_Sur(i, j)$	Scale (m/s^2)	Average velocity heterogeneity in evacuee j's surrounding area at time step i	Average of absolute velocity difference of evacuee j and his surrounding evacuees at time step i
	Gender [11]	$Ind\_Gender(j)$	Nominal	Identified gender of evacuee j	0: Male; 1: Female; 2: Ambiguous (hard to judge)

Type	Subtype	Variable	Type of variable	Definition	Calculating methods
Environment-related variables	Relative velocity with surrounding evacuees [27]	$Sum\_\Delta V\_Sur(i, j)$	Scale (m/s^2)	Overall velocity competitiveness of evacuee j among his surrounding evacuees at time step i	Sum of velocity difference of evacuee j and his surrounding evacuees at time step i
		$Ave\_\Delta V\_Sur(i, j)$	Scale (m/s^2)	Average velocity competitiveness of evacuee j among his surrounding evacuees at time step i	Average of velocity difference of evacuee j and his surrounding evacuees at time step i
		$No\_Faster\_Sur(i, j)$	Scale (No of people)	Rankings of evacuee j's velocity among his surrounding evacuees at time step i	No. of surrounding evacuees who has higher velocity than evacuee j at time step i
	Surrounding evacuees' behavior (velocity change) [13]	$V\_Decele\_Sur(i, j)$	Scale (No of people)	No. of evacuee with obvious deceleration action in evacuee j's surroundings at time step i	No. of evacuee with $A(i, k) < -1.5m/s2$ in evacuee j's surroundings at time step i
		$V\_Constant\_Sur(i, j)$	Scale (No of people)	No. of evacuee maintaining constant velocity action in evacuee j's surroundings at time step i	No. of evacuee with $\lvert A(i, k)\rvert \le 1.5m/s2$ in evacuee j's surroundings at time step i
		$V\_Accele\_Sur(i, j)$	Scale (No of people)	No. of evacuee with obvious acceleration action in evacuee j's surroundings at time step i	No. of evacuee with $A(i, k) > 1.5m/s2$ in evacuee j's surroundings at time step i
	Surrounding evacuees' behavior (direction change) [13]	$D\_Constant\_Sur(i, j)$	Scale (No of people)	No. of evacuee maintaining constant direction action in evacuee j's surroundings at time step i	No. of evacuee who maintains the same direction in evacuee j's surroundings at time step i
		$D\_Change\_Sur(i, j)$	Scale (No of people)	No. of evacuee with direction change action in evacuee j's surroundings at time step i	No. of evacuee who changes his direction in evacuee j's surroundings at time step i
	Evacuation phase	$Ind\_T(i, j)$	Scale	Indicator to depict the phase of evacuee j's route to a specific exit	Cumulative time divided by the total time in a specific route of evacuee j at time step i

Appendix C

Variables with absolute value of Spearman's Rank coefficient larger than 0.3

		Video 1 (Run 9)			Video 2 (Run 19)		
		Type of Movement Pattern			Type of Movement Pattern		
		Correlation Coefficient	Sig. (2 tailed)	N	Correlation Coefficient	Sig. (2 tailed)	N
Proximity to exits	Dist_E1	0.787**	.000	893			
	Dist_E2	0.781**	.000	893	0.532**	.000	2578
	Dist_E3	0.754**	.000	893	0.545**	.000	2578
	Dist_E4	0.750**	.000	893			
	Dist_E5				0.318**	.000	2578
Proximity to obstacle	Dist_Obstacle	0.646**	.000	893	0.486**	.000	2578
Relative velocity with surrounding evacuees	Sum_ΔV_Sur	0.425**	.000	893	0.436**	.000	2578
	Ave_ ΔV	0.417**	.000	893	0.434**	.000	2578
	Num_Faster	0.384**	.000	893	0.321**	.000	2578
	Sum_Abs_ΔV_Sur				0.386**	.000	2578
	Ave_Abs_ΔV_Sur				0.469**	.000	2578
Surrounding evacuees' velocity change	V_Constant_Sur	0.444**	.000	893			
	V_Accele_Sur	0.522**	.000	893	0.372**	.000	2578
Surrounding evacuees' direction change	D_Constant_Sur	0.302**	.000	893	0.308**	.000	2578
	D_Change_Sur	0.406**	.000	893	0.481**	.000	2578
Evacuation phase	Ind_T	0.731**	.000	893	0.412**	.000	2578

**Correlation is significant at the 0.01 level (2-tailed).

Appendix D
Result of principal component analysis: total variance explained (Video 1)

Total Variance Explained

Component	Initial Eigenvalues			Extracted Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6.888	52.988	52.988	6.888	52.988	52.988	6.202	47.709	47.709
2	2.525	19.422	72.410	2.525	19.422	72.410	2.664	20.492	68.201
3	1.803	13.869	86.279	1.803	13.869	86.279	2.350	18.078	86.279
4	.828	6.367	92.646						
5	.588	4.523	97.170						
6	.142	1.095	98.264						
7	.113	.867	99.131						
8	.054	.413	99.544						
9	.033	.252	99.796						
10	.023	.180	99.977						
11	.003	.022	99.999						
12	.000	.001	100.000						
13	3.344E-6	2.572E-5	100.000						

Appendix E
Result of principal component analysis: total variance explained (Video 2).

Total Variance Explained

Component	Initial Eigenvalues			Extracted Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.262	40.475	40.475	5.262	40.475	40.475	4.416	33.968	33.968
2	2.606	20.044	60.520	2.606	20.044	60.520	2.655	20.420	54.388
3	1.091	8.391	68.910	1.091	8.391	68.910	1.308	10.059	64.447
4	1.841	14.165	83.075	1.841	14.165	83.075	2.422	18.628	83.075
5	.660	5.077	88.152						
6	.564	4.335	92.487						
7	.414	3.184	95.671						
8	.209	1.611	97.282						
9	.174	1.336	98.618						
10	.079	.605	99.223						
11	.069	1.841	14.165						
12	.032	.243	99.996						
13	.001	.004	100.000						

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.firesaf.2019.04.008.

References

[1] X. Pan, C.S. Han, K. Dauber, K.H. Law, Human and social behavior in computational modeling and analysis of egress, Autom. ConStruct. (2006), https://doi.org/10.1016/j.autcon.2005.06.006.
[2] J. Wang, M. Chen, W. Yan, Y. Zhi, Z. Wang, A utility threshold model of herding-panic behavior in evacuation under emergencies based on complex network theory, Simulation (2017), https://doi.org/10.1177/0037549716678659.
[3] S.P. Hoogendoorn, P.H.L. Bovy, Pedestrian route-choice and activity scheduling theory and models, Transp. Res. Part B Methodol. (2004), https://doi.org/10.1016/S0191-2615(03)00007-9.
[4] R. Lovreglio, E. Ronchi, D. Nilsson, A model of the decision-making process during pre-evacuation, Fire Saf. J. (2015), https://doi.org/10.1016/j.firesaf.2015.07.001.
[5] X. Wang, W. Guo, X. Zheng, Information guiding effect of evacuation assistants in a two-channel segregation process using multi-information communication field model, Saf. Sci. (2016), https://doi.org/10.1016/j.ssci.2016.04.005.
[6] S. Cao, W. Song, W. Lv, Modeling pedestrian evacuation with guiders based on a multi-grid model, Phys. Lett. Sect. A Gen. At. Solid State Phys. (2016), https://doi.org/10.1016/j.physleta.2015.11.028.
[7] D. Li, B. Han, Behavioral effect on pedestrian evacuation simulation using cellular automata, Saf. Sci. (2015), https://doi.org/10.1016/j.ssci.2015.07.003.
[8] J. Tanimoto, A. Hagishima, Y. Tanaka, Study of bottleneck effect at an emergency evacuation exit using cellular automata model, mean field approximation analysis, and game theory, Phys. Stat. Mech. Appl. Phys. Stat. Mech. Appl. (2010), https://doi.org/10.1016/j.physa.2010.08.032.
[9] X. Yang, Z. Wu, Y. Li, Difference between real-life escape panic and mimic exercises in simulated situation with implications to the statistical physics models of emergency evacuation: the 2008 Wenchuan earthquake, Phys. Stat. Mech. Appl. Phys. Stat. Mech. Appl. (2011), https://doi.org/10.1016/j.physa.2010.10.019.
[10] M. Haghani, M. Sarvi, Crowd behaviour and motion: empirical methods, Transp. Res. Part B Methodol. (2018), https://doi.org/10.1016/j.trb.2017.06.017.
[11] M.J. Hurley, D. Gottuk, J.R. Hall, K. Harada, E. Kuligowski, M. Puchovsky, J. Torero, JjM. Watts, C. Wieczorek, SFPE Handbook of Fire Protection Engineering, fifth ed., (2016), https://doi.org/10.1007/978-1-4939-2565-0.
[12] R. Lovreglio, D. Borri, L. Dell'Olio, A. Ibeas, A discrete choice model based on random utilities for exit choice in emergency evacuations, Saf. Sci. (2014), https://doi.org/10.1016/j.ssci.2013.10.004.
[13] R. Lovreglio, A. Fonzone, L. dell'Olio, D. Borri, A study of herding behaviour in exit choice during emergencies based on random utility theory, Saf. Sci. (2016), https://doi.org/10.1016/j.ssci.2015.10.015.
[14] M. Aleksandrov, A. Rajabifard, M. Kalantari, R. Lovreglio, V.A. González, People choice modelling for evacuation of tall buildings, Fire Technol. 54 (2018) 1171–1193.
[15] S. Cao, L. Fu, P. Wang, G. Zeng, W. Song, Experimental and modeling study on evacuation under good and limited visibility in a supermarket, Fire Saf. J. 102 (2018) 27–36, https://doi.org/10.1016/j.firesaf.2018.10.003.
[16] M.M. Ishaque, R.B. Noland, Behavioural issues in pedestrian speed choice and street

crossing behaviour: a review, Transport Rev. 28 (2008) 61–85.
[17] J.H.T. Lam, J.K.K. Yuen, E.W.M. Lee, R.Y.Y. Lee, Experimental study on upward movement in a high-rise building, Saf. Sci. (2014), https://doi.org/10.1016/j.ssci.2014.07.011.
[18] T. Kretz, A. Grünebohm, A. Kessel, H. Klüpfel, T. Meyer-König, M. Schreckenberg, Upstairs walking speed distributions on a long stairway, Saf. Sci. (2008), https://doi.org/10.1016/j.ssci.2006.10.001.
[19] J. Ma, W. guo Song, Z. ming Fang, S. ming Lo, G. xuan Liao, Experimental study on microscopic moving characteristics of pedestrians in built corridor based on digital image processing, Build. Environ. (2010), https://doi.org/10.1016/j.buildenv.2010.03.015.
[20] M. Delin, J. Norén, E. Ronchi, K. Kuklane, A. Halder, K. Fridolf, Ascending stair evacuation: walking speed as a function of height, Fire Mater. (2017), https://doi.org/10.1002/fam.2410.
[21] G.Y. Jeon, J.Y. Kim, W.H. Hong, G. Augenbroe, Evacuation performance of individuals in different visibility conditions, Build. Environ. (2011), https://doi.org/10.1016/j.buildenv.2010.11.010.
[22] J. Chen, R. Liu, J. Wang, Y. Chen, Experimental influence of pedestrian load on individual and group evacuation speed in staircases, Fire Technol. (2017), https://doi.org/10.1007/s10694-017-0655-1.
[23] G. Antonini, M. Bierlaire, M. Weber, Discrete choice models of pedestrian walking behavior, Transp. Res. Part B Methodol. (2006), https://doi.org/10.1016/j.trb.2005.09.006.
[24] L. Mussone, M. Bassani, P. Masci, Analysis of factors affecting the severity of crashes in urban road intersections, Accid. Anal. Prev. 103 (2017) 112–122.
[25] H. yin Lau, K. yu Tong, H. Zhu, Support vector machine for classification of walking conditions of persons after stroke with dropped foot, Hum. Mov. Sci. (2009), https://doi.org/10.1016/j.humov.2008.12.003.
[26] P.A. Federolf, A novel approach to study human posture control: “Principal movements” obtained from a principal component analysis of kinematic marker data, J. Biomech. (2016), https://doi.org/10.1016/j.jbiomech.2015.12.030.
[27] Y. Ma, E.W.M. Lee, R.K.K. Yuen, An artificial intelligence-based approach for simulating pedestrian movement, IEEE Trans. Intell. Transp. Syst. (2016), https://doi.org/10.1109/TITS.2016.2542843.
[28] J. Shen, T. Cheng, Clustering analysis of officer'S behaviours in London police foot patrol activities, Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. (2015), https://doi.org/10.5194/isprsannals-II-4-W2-143-2015.
[29] M. Haghani, M. Sarvi, Stated and revealed exit choices of pedestrian crowd evacuees, Transp. Res. Part B Methodol. (2017), https://doi.org/10.1016/j.trb.2016.10.019.
[30] T. Rinne, K. Tillander, Peter Grönberg, Data collection and analysis of evacuation situations, VTT Tied. - Valt. Tek. Tutk. Tutkimuksia (2010) 24–25.
[31] L.D. Vanumu, K. Ramachandra Rao, G. Tiwari, Fundamental diagrams of pedestrian flow characteristics: a review, Eur. Transp. Res. Rev. (2017), https://doi.org/10.1007/s12544-017-0264-6.
[32] P.-N. Tan, M. Steinbach, V. Kumar, Data mining cluster analysis: basic concepts and algorithms, Introd. to Data Min. (2013) 2.
[33] I.N. Martínez, J.M. Morán, F.J. Peña, Two-step cluster procedure after principal component analysis identifies sperm subpopulations in canine ejaculates and its relation to cryoresistance, J. Androl. 27 (2006) 596–603.
[34] Y.-H. Chang, H.-H. Yang, Cabin safety and emergency evacuation: passenger experience of flight CI-120 accident, Accid. Anal. Prev. 43 (2011) 1049–1055.
[35] L.I. Smith, A Tutorial on Principal Components Analysis, (2002).
[36] P. McCullagh, J.A. Nelder, Generalized Linear Models, CRC press, 1989.
[37] N. Cristianini, J. Shawe-Taylor, An Introduction to Support Vector Machines and other kernel based learning methods, AI Mag. (2000) 26.
[38] T. Cover, P. Hart, Nearest neighbor pattern classification, IEEE Trans. Inf. Theory (1967), https://doi.org/10.1109/TIT.1967.1053964.
[39] S.M. Lo, M. Liu, P.H. Zhang, R.K.K. Yuen, An artificial neural-network based predictive model for pre-evacuation human response in domestic building fire, Fire Technol. (2009), https://doi.org/10.1007/s10694-008-0064-6.
[40] C. Dias, M. Sarvi, N. Shiwakoti, M. Burd, Emergency egress through angled escape routes: combining experiments with biological entities and pedestrian crowd simulation, Proceedings of the 91st Transportation Research Board Annual Meeting (TRB 2012), 2012, pp. 1–12 Washington D.C., United States.
[41] Z. Fu, L. Luo, Y. Yang, Y. Zhuang, P. Zhang, L. Yang, H. Yang, J. Ma, K. Zhu, Y. Li, Effect of speed matching on fundamental diagram of pedestrian flow, Phys. Stat. Mech. Appl. Phys. Stat. Mech. Appl. (2016), https://doi.org/10.1016/j.physa.2016.03.060.
[42] M. Haghani, M. Sarvi, Human exit choice in crowded built environments: investigating underlying behavioural differences between normal egress and emergency evacuations, Fire Saf. J. (2016), https://doi.org/10.1016/j.firesaf.2016.07.003.
[43] J. Ye, X. Chen, N. Jian, Impact analysis of human factors on pedestrian traffic characteristics, Fire Saf. J. (2012), https://doi.org/10.1016/j.firesaf.2012.05.003.
[44] M. Haghani, M. Sarvi, Social dynamics in emergency evacuations: disentangling crowd's attraction and repulsion effects, Phys. Stat. Mech. Appl. Phys. Stat. Mech. Appl. 475 (2017) 24–34.
[45] L. Shi, Q. Xie, X. Cheng, L. Chen, Y. Zhou, R. Zhang, Developing a database for emergency evacuation model, Build. Environ. (2009), https://doi.org/10.1016/j.buildenv.2008.11.008.
[46] M. Fernández-Delgado, E. Cernadas, S. Barro, D. Amorim, D. Amorim Fernández-Delgado, Do we need hundreds of classifiers to solve real world classification problems? J. Mach. Learn. Res. (2014), https://doi.org/10.1016/j.csda.2008.10.033.
[47] Y. Liu, J.W. Bi, Z.P. Fan, Multi-class sentiment classification: the experimental comparisons of feature selection and machine learning algorithms, Expert Syst. Appl. (2017), https://doi.org/10.1016/j.eswa.2017.03.042.
[48] L. Chen, T.Q. Tang, H.J. Huang, J.J. Wu, Z. Song, Modeling pedestrian flow accounting for collision avoidance during evacuation, Simulat. Model. Pract. Theor. (2018), https://doi.org/10.1016/j.simpat.2017.12.011.
[49] M.M. Shuaib, Incorporating intelligence for typical evacuation under the threat of fire spreading, Saf. Sci. 106 (2018) 1–9.
[50] J.-S. Chou, C.-F. Tsai, A.-D. Pham, Y.-H. Lu, Machine learning in concrete strength simulations: multi-nation data analytics, Constr. Build. Mater. 73 (2014) 771–780.
[51] M. Kotaru, K. Joshi, D. Bharadia, S. Katti, Spotfi: decimeter level localization using wifi, ACM SIGCOMM Comput. Commun. Rev, ACM, 2015, pp. 269–282.
[52] N. Feng, Y. Zhang, Y. Xu, Y. Li, Robust laser radar-based robot localization using UFIR filtering, 5th Int. Conf. Information, Cybern. Comput, Soc. Syst., IEEE, 2018, pp. 473–477 2018.