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

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

2019

Wang, K., Shi, X., Goh, A. P. X., & Qian, S. (2019). A machine learning based study on pedestrian movement dynamics under emergency evacuation. *Fire Safety Journal*, 106, 163-176. doi:10.1016/j.firesaf.2019.04.008

<https://hdl.handle.net/10356/143390>

<https://doi.org/10.1016/j.firesaf.2019.04.008>

© 2019 Elsevier Ltd. All rights reserved. This paper was published in *Fire Safety Journal* and is made available with permission of Elsevier Ltd.

Downloaded on 26 Apr 2023 18:45:06 SGT

Manuscript Details

Manuscript number	FISJ_2018_335
Title	A machine-learning based study on pedestrian movement dynamics under emergency evacuation
Article type	Research Paper

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, the movement pattern and its influencing factors are still not well understood. In this study, we explore the potential of adopting machine learning methods to study evacuees' stepwise movement dynamics based on an emergency evacuation experiment. The behavior patterns were categorized through 2-step cluster analysis and principal influencing components were identified through principal component analysis. The relationship between these behavior patterns and principal components were investigated using different modeling methods, namely multinomial logit model (MLM), Decision Tree (DT), Support Vector Machine (SVM) and K-Nearest Neighbor (KNN). Results show that: (1) Distance to the target exits has the most pronounced effect on evacuee's behavior patterns. (2) Surrounding evacuees' actions also have a significant and complex influence on a single evacuee's behavior pattern. (3) Machine learning methods (DT, SVM, and KNN) outperform MLM in terms of prediction accuracy. DT, SVM, and KNN are most advantageous in interpretability, prediction accuracy, and processing speed respectively. These findings can provide useful insights for further development of emergency movement dynamics and the emergency evacuation modeling.

Keywords	Emergency evacuation; Movement dynamics; Machine learning; Behavior pattern.
Corresponding Author	Shunzhi Qian
Corresponding Author's Institution	Nanyang Technological University
Order of Authors	KE WANG, Xiupeng Shi, Shunzhi Qian
Suggested reviewers	Long Shi, Jianhong YE, Milad Haghani

Submission Files Included in this PDF

File Name [File Type]

Cover Letter.pdf [Cover Letter]

Highlights.pdf [Highlights]

Manuscript.pdf [Manuscript File]

Conflict of Interest .pdf [Conflict of Interest]

To view all the submission files, including those not included in the PDF, click on the manuscript title on your EVISE Homepage, then click 'Download zip file'.

Research Data Related to this Submission

There are no linked research data sets for this submission. The following reason is given:
Data will be made available on request

Cover letter

Dear editors,

We are pleased to submit a manuscript titled “A machine-learning based study on pedestrian movement dynamics under emergency evacuation”, for consideration as a research paper by *Fire Safety Journal*.

This study analyzes evacuees’ stepwise movement dynamics under emergency evacuation with a combination of machine learning methods. The significance of this study lies in two aspects. Theoretically, the empirical facts of single evacuee’s speed changing behavior are summarized, which can provide insight into group movement dynamics and thus facilitate evacuation management. Technically, machine learning methods demonstrate better prediction performance compared with the commonly used multinomial logit model. This gives rise to new methodology for evacuation behavior analysis. We believe these findings will be of interest to the readers in the areas of human behaviors, evacuation modeling, and safety design.

We confirm that this manuscript is original and has not been published before, nor is it currently under consideration for publication elsewhere. All the authors listed have contributed sufficiently to the study. We have included acknowledgements and funding sources in the manuscript. There is no conflict of interest existing.

Thank you for your consideration of this manuscript and look forward to hearing from you.

Sincerely,

Ke Wang, Xiupeng Shi, Shunzhi Qian

Highlights

- A bias-less analytical methodology is proposed to study movement dynamics from individual perspective.
- Evacuees' behavior patterns during evacuation and the principal influencing components are extracted.
- The classification performance of three machine learning methods and multinomial logit model is compared.
- Distance to target exits and surrounding evacuees' actions can affect one's evacuation speed.
- Surrounding evacuees' direction change action negatively influences one's evacuation speed.

1 **A machine-learning based study on pedestrian movement dynamics**
2 **under emergency evacuation**

3
4 Ke Wang, Xiupeng Shi, Shunzhi Qian*

5

6

7

8

9

10

11

12

13

14

15

16

17

18

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

21

22 * Corresponding author

23 Email address: szqian@ntu.edu.sg

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, the movement pattern and its influencing factors are still not well understood. In this study, we explore the potential of adopting machine learning methods to study evacuees' stepwise movement dynamics based on an emergency evacuation experiment. The behavior patterns were categorized through 2-step cluster analysis and principal influencing components were identified through principal component analysis. The relationship between these behavior patterns and principal components were investigated using different modeling methods, namely multinomial logit model (MLM), Decision Tree (DT), Support Vector Machine (SVM) and K-Nearest Neighbor (KNN). Results show that: (1) Distance to the target exits has the most pronounced effect on evacuee's behavior patterns. (2) Surrounding evacuees' actions also have a significant and complex influence on a single evacuee's behavior pattern. (3) Machine learning methods (DT, SVM, and KNN) outperform MLM in terms of prediction accuracy. DT, SVM, and KNN are most advantageous in interpretability, prediction accuracy, and processing speed respectively. These findings can provide useful insights for further development of emergency movement dynamics and the emergency evacuation modeling.

Keywords

Emergency evacuation; Movement dynamics; Machine learning; Behavior pattern.

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 occur 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 actually 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 crowd's panic and other 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 to better understand 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.* strategic level (whether to initiate evacuation), tactical level (route choice) and operational level decision making process (stepwise movement) [3]. Among them, evacuees' stepwise movement is the primary element to construct the whole evacuation process. Substantial effort has been put into simulating evacuees' movement during evacuation; two most representative models proposed are Social Force (SF) Model and Cellular Automata (CA) Model. SF Model regards evacuees' movement as the physical effect of distance based forces, namely 'attraction force from the target', 'repulsive force from other pedestrian and barrier' due to territorial effect [4], 'attraction force from some other pedestrians' [5]. CA Model simulates evacuees' movement by defining the transition probabilities between discrete space [6]. With the development of the social behavior theory and psychological science, more factors

76 derived from theories have been incorporated into the movement dynamics simulation like the
77 availability of instruction [7,8], familiarity of the exits [9], degree of collision-avoidance [10],
78 and panic level [11,12]. Aforementioned studies establish the principles of the movement
79 dynamics, *i.e.* the movement pattern of each step and the rules governing stepwise movement.
80 However, one thing to be noted is that there is no single commonly accepted theory that can
81 explain the movement dynamics comprehensively [13,14]; some theories are even
82 contradictory such as Panic Theory [15] versus Decision-Making Theory [16]. To solve this
83 problem, more robust empirical knowledge of movement dynamics is needed to enhance the
84 understanding and thus to facilitate evacuation simulation [13,17].

85 Among several data gathering methods, controlled laboratory experiments with human
86 subjects are a most ideal way to extract individual-level movement dynamics due to its
87 environmental realism and wide topic inclusion range [17]. The advancements in photography
88 facilities and image processing techniques support reliable video analysis of the recorded
89 experiments. Some researchers have performed the experiments to explore the relationship
90 between several design factors and movement dynamics. Since all the factors are predefined
91 based on the researchers' knowledge, it is more likely to answer 'how much' the design factor
92 can influence the movement, instead of 'what' factors can influence evacuees' movement,
93 while the latter one is more fundamental in developing behavior theory. To answer the 'what'
94 question, machine learning is a promising approach. It is well-developed in computer science
95 area and has been mostly applied in economics and medical science for its superiority in
96 extracting the hidden rules from real-life data [18]. There have been studies using machine
97 learning methods in human movement analyses. Support Vector Machine was used to identify
98 walking conditions for hemiparetic persons [19]. Principal component analysis was applied to
99 quantifying human upright standing with some 'principal movements' [20]. Neural Network
100 was adopted to analyze one's horizontal and vertical normal walking speed in relation to
101 surrounding twenty-five people's vertical and horizontal speeds [21]. Cluster analysis was
102 utilized to deduce police officers' patrol patterns [22]. However, to the best of our knowledge,

there are few studies that analyze stepwise movement dynamics (movement patterns and behavior rules) under emergency evacuation. Moreover, there is a scarcity of studies that explore the potential and compare the performance of different machine learning methods in emergency evacuation analysis.

This study aims to gain insight into evacuees' movement dynamics of their stepwise movement and key components governing their movement patterns through a combination of machine learning methods based on an experimental video record. The rest of the paper is organized as follows: Section 2 describes the experimental setup and data collection methods; Section 3 introduces the basics of various methods and research methodology; Sections 4 and 5 are 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 a video record of emergency evacuation experiments carried out by Sarvi and Haghani, Center for Disaster Management and Public Safety, University of Melbourne. Around 150 participants were requested to escape from the exits as fast as possible. The evacuation experiment was conducted for 20 times with different combinations of design parameters. Permission has been obtained for us to use Run 19 for stepwise movement analysis (Table 1). More detailed information can be found from their paper [23].

2.2. Mapping and validation of coordinates

A two-step mapping approach was adopted to map from the pixel coordinates to world coordinates, as large lens distortion was observed due to wide angle of GoPro Hero 4 Black Camera, placed at 8 m height to capture the evacuation process from a top-down view. Firstly, positive radial distortion was removed aided by Adobe Photoshop CS6. Comparison of processed figure and original figure is shown in Fig. 1. Subsequently, rotation matrix and translation matrix were applied to transform from camera's coordinates to world coordinates.

The maximum error for nineteen reference points is 27.458 cm and 28.343 cm for x-axis (horizontal axis) and y-axis (vertical axis), respectively.

Table 1

Specification of design parameters in Run19 [23].

Run number	Number of evacuees	Available exits	Exit widths
19	143	2,3,5	w2=w3=w5=50 cm

Fig. 1. A snapshot of raw footage in Run 19: (a) before and (b) after lens distortion removal.

2.3. Extraction of variables

Stepwise positions of evacuees were identified for further variable extraction. Although some software claims the capability of recognizing objects and extracting trajectory automatically, it turns out that these programs cannot give satisfactory results in this case due to high mobility and density of evacuees. Hence, a semi-automated identification process was adopted, *i.e.* coordinates of evacuees' heads (center of the hats) were obtained manually, which were then automatically transformed from pixel coordinates to world coordinates using the algorithm derived in section 2.2. Each evacuee's stepwise position is denoted by dots in the world coordinates as shown in Fig. 2.

Fig. 2. Stepwise position of evacuee No.1-No.143 for the entire evacuation process.

To capture the details of evacuees' movement, twenty-five variables were derived based on their stepwise position (Appendix A). These variables can be divided into behavior-related variables and environment-related ones, where the former describe an evacuees' speed and acceleration rate, and the latter describe evacuees' internal environment (gender in this study) and external environment (such as density in the surrounding). Due to the scope of the experiment, we are unable to obtain other internal features of evacuees based on visual analysis of the video except for gender. Environment-related variables represent features of exits, features of surrounding evacuees *etc.* The aim of this process is to extract as many variables as possible to minimize potential omission of information.

Hereinafter, we used the term 'case' to refer to each stepwise movement of the evacuees. Each evacuee's evacuation process was divided into several cases with a constant time interval, *i.e.* 0.2 s. Each case consists of two behavior-related variables and twenty-three 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, *i.e.* the target exit should remain constant for a route. To minimize the influence of exit-changing actions, as well as to maximize the data volume that is applicable to the assumption, only cases targeting Exit 2 were chosen for the study.

3. Methodology

This research intends to identify evacuees' movement dynamics through machine learning methods (Fig. 3). Firstly, behavior-related variables were used to classify behavior patterns through 2-step cluster analysis. Similar cases were grouped together into specific behavior patterns. Secondly, principal component analysis was applied to identify the independent principal components, aiming to sort out the uncorrelated variables and merge interdependent variables. To find out how these principal components influence behavior patterns, four approaches were used for comparison, namely multinomial logit model, Decision Tree, Support Vector Machine, and K-Nearest Neighbor.

Fig. 3. Research framework.

3.1. 2-step cluster analysis

Cluster analysis is an exploratory data mining technique to group homogenous cases. 2-step cluster analysis was chosen because of its superiority over other two traditional cluster methods, K-means cluster analysis and hierarchical cluster analysis. 2-step cluster analysis can deal with large dataset (where the hierarchical cluster analysis is weak) without specifying the number of clusters beforehand (which is required for the K-means cluster analysis).

2-step cluster analysis consists of two steps. Firstly, all the cases are pre-classified into sub-clusters using a sequential clustering method. Cluster Feature (CF) tree is built based on the

181 closeness criteria and distance criteria. Secondly, all the sub-clusters are grouped into final
182 clusters using the agglomerative hierarchical clustering methods [24].

183 Two behavior-related variables, V and A were used for behavior pattern categorization using
184 this 2-step cluster analysis. All the data were randomized and standardized to meet the data
185 requirement. The log-likelihood distance was chosen as the distance matrix and Schwarz's
186 Bayesian Criterion (BIC) was chosen as the clustering criteria.

187 **3.2. Bivariate correlation test and principal component analysis**

188 Spearman's Rank Coefficient was used to measure the relationship between environment-
189 related variables and identified behavior patterns. The variables with Spearman's Rank
190 Coefficient larger than 0.3 or smaller than -0.3 were deemed correlated variables and held for
191 further analysis.

192 Principal component analysis (PCA) is used to merge the correlated environment-related
193 variables and to reduce the component dimension [25]. A set of linearly independent
194 components can be obtained to represent most of the information from a set of possibly
195 correlated variables. Kaiser-Meyer-Olkin (KMO) Measure is used to test data adequacy which
196 indicates the proportion of variance in the variables that might be caused by underlying factors.
197 Bartlett's test of sphericity can be used to examine whether the variables are unrelated or
198 unsuitable for structure detection. Once the above two requirements are met, decomposition
199 method of the covariance matrix is used to obtain the principal components of the dataset [26].

200 **3.3. Modeling methods**

201 **3.3.1. Multinomial logit model**

202 To find the relationship between behavior patterns and principal components, multinomial
203 logit model (MLM) is used as the baseline for comparison, which is a widely used models in
204 pedestrian evacuation analysis. It can describe the relative risk of being with one type of
205 behavior pattern versus being with the reference type. Here, the behavior pattern with the largest

possibility is chosen as the reference type. Based on the multinomial logit model specified in Eqs. (1) - (3), maximum likelihood estimates for coefficient β_{jk} can be obtained, using an iteratively weighted least square algorithm [27].

$$\ln(P_1/P_k) = \alpha_1 + \beta_{11} * PC_1 + \beta_{12} * PC_2 + \dots + \beta_{1p} * PC_p \quad (1)$$

$$\ln(P_2/P_k) = \alpha_2 + \beta_{21} * PC_1 + \beta_{22} * PC_2 + \dots + \beta_{2p} * PC_p \quad (2)$$

$$\ln(P_{k-1}/P_k) = \alpha_{k-1} + \beta_{(k-1)1} * PC_1 + \beta_{(k-1)2} * PC_2 + \dots + \beta_{(k-1)p} * PC_p \quad (3)$$

Where P_j : the possibility of an outcome being with type number j

k: the total number of behavior types

β_{jk} : the coefficient, *i.e.* the effect of the predictor variables (principal components in this study) on the relative risk of being with type j versus reference type k

PC_p : the predictor variables (principal components in this study)

3.3.2. Machine learning methods

Decision Tree (DT) is a flowchart-like graph for decision analysis. The tree structure consists of leaf nodes representing classification types (the type of behavior patterns in this study), internal nodes representing attributes (principal component in this study), and branches representing value of the attributes (value of principal component in this study). To construct the tree, a principal component needs to be chosen at each step to best separate the cases. The ‘best’ here is gauged by the homogeneity of the principal components within the subset, which is measured by Gini’s Diversity Index [28]. To test model flexibility, maximum number of splits is specified at different levels (4, 20, 100) for study.

Support Vector Machine (SVM) can classify cases with the best hyperplane which can achieve the largest margin between any two classes. By mapping low-dimension data into high-dimension one with kernel functions, the cases can be separated efficiently [29]. Four types of kernel functions are used for comparison, 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 predictors. This study adopts the iterative data training process proposed by Kecman *et.al* [30].

K-Nearest Neighbor (KNN) categorizes data points based on their surrounding k points' classes [31]. The matrix for distance and the number of nearest neighbors can determine how the model works. The distance matrix can be Cosine distance (Cosine KNN), Cubic distance (Cubic KNN) or Euclidean distance with specified distance weight (Weighted KNN), and Euclidean distance *etc.* Moreover, the number of nearest neighbors is set to be 1 (Fine KNN), 10 (Medium SVM) and 100 (Coarse KNN) for Euclidean distance in this study.

4. Result and analysis

4.1. Behavior pattern categorization

Three clusters were identified from 2-step cluster analysis as shown in Fig. 4. Over half of the cases (54.62%) belong to the Behavior Pattern (BP) 1, 'Constant low velocity' BP, *i.e.* evacuees with low velocity (mean = 0.32 m/s, SE = 0.01 m/s) and low acceleration rate (mean = 0.14 m/s², SE = 0.02 m/s²). In 19.20% cases, evacuees are accelerating (mean = 1.22 m/s², SE = 0.05 m/s²) with a high velocity (mean = 1.61 m/s, SE = 0.03 m/s), which is called BP2, 'High velocity with acceleration' BP. As for the remaining 26.18% cases, evacuees decelerate (mean = -1.26 m/s², SE = 0.03 m/s²) with a high velocity (mean = 1.48 m/s, SE = 0.03 m/s), the so-called 'High velocity with deceleration' BP (BP3).

Fig. 4. Cluster composition and cluster centers of three behavior patterns.

4.2. Extraction of principal components

To remove the uncorrelated variables, Spearman's Rank Coefficient was tested between 'Type of Behavior Pattern' and remaining twenty-three environment-related variables. Thirteen variables were kept for further PCA as shown in Appendix B. It is found that evacuees tend to be more sensitive to other's acceleration action than the deceleration action, therefore 'VChange_3' (number of accelerating actions in the surrounding) was kept for PCA whereas 'VChange_1' (number of decelerating actions in the surrounding) was discarded.

Notably, the discarded factors are not absolutely unimportant in the emergency evacuation, as evacuation study is highly context-dependent [17]. 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 speeds during evacuation [32] which is not confirmed 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 high-density evacuees. Evacuee density near the target exit was also found to influence one's route choice [33], so as individual and overall crowd speed [34]. Nevertheless, the density near Exit 2 was rather high most of the times, which is not well controlled in this study. To tackle this unavoidable issue encountered by most of the video-based studies, more careful data gathering work is needed.

The KMO test ($KMO = 0.617$) and Bartlett's Test of Sphericity ($Sig. = 0.000$) indicates data adequacy and suitability for dimension reduction. Four principal component (PC)s were extracted to represent the majority (83.075%) of the information conveyed by the thirteen variables (Appendix C). As shown in Table 2, PC1 mainly contains distance information, *e.g.* distance to Exit 2 (0.941), distance to Exit 3 (0.963), distance to Exit 5 (0.763), distance to nearest wall (0.903), the phase of evacuation (-0.702), and surrounding evacuees' acceleration action (0.622). PC2 represents one's speed competitiveness, as it relates closely with 'Sum_ΔV' (0.943), 'Ave_ΔV' (0.925), and 'No_Faster' (-0.860). PC3 describes speed heterogeneity along specific direction while PC4 depicts direction change actions in one's surrounding.

Table 2

Rotated component matrix by principal component analysis of environment-related variables.

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		0.177	0.136
Dist_Wall	Distance from current position to nearest wall	0.903	0.105	0.100	0.181
Dist_E5	Distance from current position to Exit 5	0.763	0.134	-0.233	0.316
Ind_T	Indicator to depict the phase of route (cumulative time divided by total time spent in a route)	-0.702		-0.414	0.102
VChange_3	No. of evacuee with obvious acceleration action in one's surrounding	0.622		0.287	-0.123
Sum_ΔV	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	
No_Faster	No. of evacuee with higher speed in one's surrounding		-0.860	0.228	-0.220
Sum_Abs_ΔV	Overall speed heterogeneity in one's surrounding	0.123	0.186	0.905	0.187
DChange_1	No. of evacuee maintaining their direction in one's surrounding	0.404	-0.128	0.748	-0.152
Ave_Abs_ΔV	Average speed heterogeneity in one's surrounding		0.278	0.724	0.535
DChange_2	No. of evacuee with direction change action in one's surrounding	-0.331		-0.139	-0.829

4.3. Movement rule generation

To find the rules governing evacuees' movement dynamics, *i.e.* to see how these identified PCs can influence BPs, we applied machine learning methods, namely Decision Tree, Support Vector Machine and K-Nearest Neighbor, along with multinomial logit model for comparison.

4.3.1. Results from multinomial logit model

Base on the model specified in Section 3.3.1, 2000 cases were randomly selected as the training set to obtain the coefficient estimates as shown in Table 3 and Eqs. (4) – (6). A final prediction accuracy of 64.71% was obtained for the rest 578 testing cases.

Table 3

Multinomial logit estimates (reference category: Behavior Pattern 1), where BP1, BP2, BP3 stands for ‘Constant low velocity’ BP, ‘High velocity with acceleration’ BP, ‘High velocity with deceleration’ BP respectively.

Parameter	Coefficient estimates (BP2/BP1)	Coefficient estimates (BP3/BP1)
Intercept terms	-1.1709 (-10.0342) ***	-0.3208 (-3.6668) ***
PC1	2.5576 (19.8200) ***	1.8262 (15.1224) ***
PC2	1.5250 (12.1096) ***	1.8423 (15.4619) ***
PC3	1.3981 (12.4267) ***	1.2853 (13.3068) ***
PC4	1.6291 (14.3748) ***	1.3682 (14.8161) ***

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

$$\ln(P(BP2)/P(BP1)) = -1.1709 + 2.5576 * PC1 + 1.5250 * PC2 + 1.3981 * PC3 + 1.6291 * PC4 \quad (4)$$

$$\ln(P(BP3)/P(BP1)) = -0.3208 + 1.8262 * PC1 + 1.8423 * PC2 + 1.2853 * PC3 + 1.3682 * PC4 \quad (5)$$

$$\ln(P(BP2)/P(BP3)) = -0.8501 + 0.7314 * PC1 - 0.3173 * PC2 + 0.1128 * PC3 + 0.2609 * PC4 \quad (6)$$

Some features of one’s speed can be identified through the analysis of the coefficients in Eqs. (4) - (5). All the PCs have very pronounced effect in distinguishing BP1 and BP2, similarly for the differentiation between BP1 and BP3. The consistence of positive signs of coefficients in Eqs. (4) - (5) indicates that larger PC1, PC2, PC3, and PC4 can increase the possibility of behavior pattern falling into BP2 and BP3, both of which exhibit ‘higher velocity’ compared with BP1. Tracing back to the aggregated meaning of PCs as shown in Table 2, higher PC1 means larger distance from current position to target exit and barrier, earlier phase in a route, and more surrounding evacuees taking acceleration actions. Hence positive impact from PC1 on higher velocity is quite intuitive: at the beginning (‘earlier phase of a route’ and ‘larger distance from target exit’), evacuees tend to run fast to take a favorable position. Furthermore, ‘larger distance from barrier’ creates an obstruction-free environment for evacuees and ‘more evacuees accelerating’ is also a positive stimulus because one would more likely adopt a higher speed to avoid falling behind [35].

Besides PC1, positive effect of PC2 on higher velocity is also self-evident as higher PC2 means stronger speed competitiveness compared with one’s surroundings. Higher PC3 suggests larger speed heterogeneity and less direction changes in one’s surroundings. A tentative

explanation can be that higher PC3 represents a dispersed velocity distribution, which indirectly indicates an unhindered walking environment (less space constraint). In this case, people can run faster than the jammed situations. Larger PC4 can give rise to a higher velocity because less evacuees changing their directions can be viewed as a strong affirmation for the current exit choice as supported by ‘follow-the-majority’ theory [36], so that evacuees would not hesitate or wander but head directly to their target.

Apart from speed, acceleration and deceleration action also follow certain rules as demonstrated by Eq. (6). The possibility of any behavior pattern being classified as BP2 and BP3 is compared in Eq. (6), where both BPs are characterized by a relatively high speed but opposite trends in speed changing. It is clear that PC1 has far greater influence on the acceleration behavior than other PCs. Higher PC1 can increase the occurrence rate of acceleration actions, *i.e.* BP2. Reasons are as follows: at the beginning of 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 speed competitiveness will decrease the likelihood of acceleration. This may be because one has reached the upper limit of velocity under this situation, and also wants to stay close with the group. The rest PCs are not so effective in distinguishing BP2 and BP3 as the previous two judging from the absolute values of the coefficients.

Scatter plots of BPs against PCs were generated to visualize the differentiation effect of the four PCs. Based on the partition of the 3 BPs indicated by different colors, PCs with better separative effect can be identified. BP1 (blue dots) and BP2 (red dots) are better separated in Fig. 5(a) than in Fig. 5(d). Similarly, Fig. 5(b) shows better separation result between BP1 (blue dots) and BP3 (yellow dots) than that in Fig. 5(e). This can corroborate the greater significance of PC1 and PC2 in differentiating between BP2/BP3 and BP1. Nevertheless, both the combination of PC1 versus PC2, and PC3 versus PC4 do not show satisfactory separation result in Figs. 5(c) and 5(f), indicating BP2 and BP3 are difficult to differentiate. These conclusions are in good consistency with those obtained from Eqs. (4) - (6).

Fig. 5. Scatter plot of behavior patterns against different combinations of principal components. The blue, red and yellow dots represent BP1 ('Constant low velocity'), BP2 ('High velocity with acceleration'), and BP3 ('High velocity with deceleration') respectively.

4.3.2. Results from machine learning methods

The results from machine learning methods show higher prediction accuracy compared with that from multinomial logit model (Table 4). Among the three subtypes within the DT model, Medium Tree gives the highest prediction accuracy (74.60%) with a training time of 11.336 s. Most of the SVMs have a slightly better classification performance than DTs, nevertheless they are more computationally expensive. Medium Gaussian SVM (77.70%) shows the highest prediction accuracy among all the SVMs, followed by Quadratic SVM (76.40%) and Coarse Gaussian SVM (75.80%). However, overall training time varies greatly, which indicates Gaussian Kernel is more computationally efficient than Linear Kernel, Quadratic Kernel and Cubic Kernel in multiclass classification. KNNs are more efficient in calculation efficiency compared with other methods, but still with comparable accuracy. Medium KNN outperforms other KNNs in terms of both prediction accuracy (76.60%) and training time (3.5303 s). The prediction accuracies and training times for all the KNNs do not show much difference except for Fine KNN, which proves the robustness of this method.

Table 4

Modeling accuracy and training time for three machine learning models. The training time was the processing time by the Windows 10 Enterprise Intel(R) Core i5-8250U.

Model Type	Model Subtype	Accuracy	Training Time/s
Decision Tree	Fine Tree	72.70%	12.864
	Medium Tree	74.60%	11.336
	Coarse Tree	72.00%	10.98
SVM	Linear SVM	75.30%	31.044
	Quadratic SVM	76.40%	436.04
	Cubic SVM	74.30%	1148.3
	Fine Gaussian SVM	75.40%	15.546
	Medium Gaussian SVM	77.70%	17.811
	Coarse Gaussian SVM	75.80%	21.145
KNN	Fine KNN	71.50%	4.2232
	Medium KNN	76.60%	3.5303
	Coarse KNN	76.50%	4.5577
	Cosine KNN	75.10%	8.7978
	Cubic KNN	76.40%	4.3737
	Weighted KNN	76.00%	4.2232
MLM		64.71%	Not applicable

Furthermore, Medium Trees was plotted to gain some insight into the relationship between BPs and PCs (Fig. 6), as DT is capable of displaying the processing steps. PC1 is identified as the first node to extend the branches with a threshold value of 0.467922. PC2 comes afterwards to further differentiate the cases. This indicates that PC1 and PC2 are the two most important PCs to classify the three BPs, which is consistent with the results obtained in Section 4.3.1. 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 BPs.

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

5. Discussion

5.1. Comparison of the models used for BP classification

To the best of the authors' knowledge, this study is one of the few studies exploring the potential of DT, SVM, and KNN in the analysis of movement dynamics under emergency evacuation. Noticeable enhancement in the prediction accuracy compared with multinomial logit model 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 behavior pattern categorization and principal component extraction process. Among the three types of machine learning methods, SVM shows the highest prediction accuracy in general. KNN surpasses the other two methods with regard to training speed and DT has the best transparency (*i.e.* easy-to-understand processing steps). These results agree with the conclusions from recent studies [37,38].

Compared to the aforementioned machine learning methods, multinomial logit model is nevertheless superior in its transparency and compatibility. Multinomial logit model is comprehensible as it has a transparent processing process and clear statistical meaning, whereas most machine learning methods function like black boxes. This is one of the reasons why

multinomial logit model has been widely used in decision-making studies [39,40]. Moreover, the compatibility with other simulation models like CA model makes this model an ideal choice for depicting the decision-making process in evacuation simulation.

5.2. Theoretical and practical implications of the study

By analyzing the influence of the environment on speed and speed-changing behaviors 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 stronger acceleration tendency in the surrounding can be stimuli for evacuees' acceleration decision. This indicates the whole group may show more diverse dynamic patterns at the early phase of the evacuation, thus more measures should be taken to maintain the order and avoid chaos at this phase. Similar analytical procedure can be adopted in the study of collision avoidance behaviors. These individual-based analyses can potentially offer new insight on the group evacuation dynamics such as non-adaptive behaviors.

Moreover, machine learning based approaches can better reflect the reality, as less prior knowledge is required for data processing. Unsupervised cluster analysis procedure can classify behavior 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, bias induced by the variable selection and the computational risk induced by the inter-dependency among variables can be reduced.

Finally, this study provided a cost-effective approach for the analysis of evacuation movement dynamics with limited available resources. As known to all, high cost involved in evacuation experiments is a large hindrance for many researchers. By mining data from a publicized experiment video, this study digs into a deeper level of evacuation dynamics analysis (operational level) compared with the original study (tactical level) [23]. The re-extraction and

re-analysis process adopted in this study can also be applied to other video records to extract deeper level information.

5.3. Limitations of the study and future work

In this study, single machine learning method is applied at one time, whereas each of them demonstrates its superiority over others in different aspects. Future study aims to attempt decision tree-based ensemble learning methods, as they show better prediction performance compared with single learning methods [41]. Bagging, Boosting, and Stacking ensemble learning methods will be compared to obtain an optimal model with high transparency, high accuracy, and high training speed.

Internal features of evacuees such as risk perception, sense of direction *etc.* have not been taken into consideration. The assumption that stronger speed 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 tackle this problem, a powerful trajectory tracing program for high-density and high-mobility situation is in great need. The advancement of global positioning system with mobile phones can also provide solutions to this problem, which may be incorporated into future experimental study.

Finally, this study is based on a quasi-emergency scenario, to which a superior alternative would be actual emergency evacuation. Future study will evaluate the efficacy of the machine learning methods/models in the analysis of actual emergency evacuation. The results obtained from both 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 prove the feasibility of this approach in emergency evacuation study, offering a cost-effective method for operational-level decision-making study. Major conclusions are summarized as follows:

(1) Evacuees' stepwise movement is complex and is hard to be precisely predicted. The prediction accuracy in this study reaches 77.7%. To improve the accuracy, more intrinsic features of evacuees need to be considered in future study along with improvement in other aspects, such as the application of ensemble learning methods and trajectory tracing tool.

(2) Several rules of movement dynamics are deduced: (a) distances to the target and to barriers are the most important aspect that influences an evacuee's velocity and acceleration rate. When the evacuee is at the early stage of the route, and far from exits or barriers, he/she is more likely to accelerate. (b) The influence of surrounding evacuees' actions on a single evacuee's behavior pattern is significant. Evacuees are more sensitive to others' acceleration actions than deceleration action. Surrounding evacuees' changes in direction can be viewed as a negative stimulus to one's velocity.

(3) Machine learning methods outperform multinomial logit model in terms of prediction accuracy. DT, SVM, and KNN are most advantageous in interpretability (transparency), prediction accuracy, and processing speed respectively.

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 A Details of the variables.

Type	Subtype	Variable	Type of variable	Definition	Calculating methods	Note
Behavior-related variables	Evacuee j's movement	$V(i, j)$	Scale (m/s)	Speed of evacuee j at time step i	Euclidean distance between adjacent steps divided by time interval (0.2s)	
		$A(i, j)$	Scale (m/s ²)	Acceleration rate of evacuee j at time step i	Speed difference between adjacent steps divided by time interval (0.2s)	
Environment-related variables	Proximity to exits	$Dist_E2(i, j)$	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	
		$Dist_E5(i, j)$	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	
	Features of exits	$Densi_E2(i)$	Scale (No. of people)	Density of Exit 2 at time step i	No of evacuee in the semicircle of Exit 2 (radius = 1.5m) 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 = 1.5m) 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 = 1.5m) 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	
		$FlowRate_E5(i)$	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	

Type	Subtype	Variable	Type of variable	Definition	Calculating methods	Note
Environment-related variables	Proximity to barrier	$Dist_Wall(i, j)$	Scale (m)	Distance from evacuee j to nearest wall at time step i	Euclidean distance from evacuee j to nearest wall at time step i	
	Features of surrounding	$Densi_Sur(i, j)$	Scale (No. of people)	Density of evacuee j's surrounding at time step i	No of evacuees in the surrounding circle of evacuee j at time step i	Circle center: evacuee j's location at time step i; Circle Radius: 2.5 m [21].
		$Sum_Abs_ΔV(i, j)$	Scale (m/s ²)	Overall speed heterogeneity in evacuee j's surrounding area at time step i	Sum of absolute speed difference of evacuee j and his surrounding evacuees at time step i	
		$Sum_ΔV(i, j)$	Scale (m/s ²)	Overall speed competitiveness of evacuee j among his surrounding evacuees at time step i	Sum of speed difference of evacuee j and his surrounding evacuees at time step i	
		$Ave_Abs_ΔV(i, j)$	Scale (m/s ²)	Average speed heterogeneity in evacuee j's surrounding area at time step i	Average of absolute speed difference of evacuee j and his surrounding evacuees at time step i	
		$Ave_ΔV(i, j)$	Scale (m/s ²)	Average speed competitiveness of evacuee j among his surrounding evacuees at time step i	Average of speed difference of evacuee j and his surrounding evacuees at time step i	
		$No_Faster(i, j)$	Scale (No of people)	Rankings of evacuee j's speed among his surrounding evacuees at time step i	No. of surrounding evacuees who has higher speed than evacuee j at time step i	
		$VChange_1(i, j)$	Scale (No of people)	No. of evacuee with obvious deceleration action in evacuee j's surrounding at time step i	No. of evacuee with $A(i, k) < -1.5m/s^2$ in evacuee j's surrounding at time step i	
		$VChange_2(i, j)$	Scale (No of people)	No. of evacuee maintaining constant speed action in evacuee j's surrounding at time step i	No. of evacuee with $ A(i, k) \leq 1.5m/s^2$ in evacuee j's surrounding at time step i	
		$VChange_3(i, j)$	Scale (No of people)	No. of evacuee with obvious acceleration action in evacuee j's surrounding at time step i	No. of evacuee with $A(i, k) > 1.5m/s^2$ in evacuee j's surrounding at time step i	

Type	Subtype	Variable	Type of variable	Definition	Calculating methods	Note
Environment-related variables	Features of surrounding	$DChange_1(i, j)$	Scale (No of people)	No. of evacuee maintaining constant direction action in evacuee j's surrounding at time step i	No. of evacuee who maintains the same direction in evacuee j's surrounding at time step i	
		$DChange_2(i, j)$	Scale (No of people)	No. of evacuee with direction change action in evacuee j's surrounding at time step i	No. of evacuee who changes his direction in evacuee j's surrounding at time step i	
	Gender indicator	$Ind_Gender(j)$	Nominal	Identified gender of evacuee j	0: Male; 1: Female; 2: Ambiguous (hard to judge)	Judged from clothes, shoulder width, hair <i>etc.</i> Ambiguous type is not used.
	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	If one changed his exit, then it is deemed a new route.
Intermediate variables	Target indicator	$Target_E(i, j) *$	Nominal	Target exit of evacuee j at time step i		Judged by head direction and body rotation action [23].
	Variables used to describe Evacuee j's surrounding evacuees	$V_Change(i, j) *$	Nominal	Indicator to depict whether there is a large speed change	Comparing $A(i, j)$ with threshold value	1: $A(i, j) < -1.5 \left(\frac{m}{s^2}\right)$; 2: $ A(i, j) \leq 1.5 (m/s^2)$; 3: $A(i, j) > 1.5 (m/s^2)$.
		$D(i, j) *$	Nominal	Direction of evacuee j at time step i	The orientation of evacuee j between adjacent steps.	1: West; 2: East; 3: South; 4: North.
		$D_Change(i, j) *$	Nominal	Direction change action of evacuee j at time step i	Comparing direction of evacuee j at adjacent steps	1: no change; 2: change.

Variables marked with * are intermediate variables for the calculation of other variables

Appendix B Bivariate Correlation Test results for the variables with Spearman's Rank Coefficient larger than 0.30 or smaller than -0.30.

	Type of Behavior Pattern		
	Correlation Coefficient	Sig. (2 tailed)	N
Dist_E2	0.532**	.000	2578
Dist_E3	0.545**	.000	2578
Dist_E5	0.318**	.000	2578
Sum_Abs_ΔV	0.386**	.000	2578
Sum_ΔV	0.436**	.000	2578
No_Faster	-0.321**	.000	2578
Ave_Abs_ΔV	0.469**	.000	2578
Ave_ΔV	0.434**	.000	2578
VChange_3	0.372**	.000	2578
DChange_1	0.308**	.000	2578
DChange_2	-0.481**	.000	2578
Dist_Wall	0.486**	.000	2578
Ind_T	-0.412**	.000	2578

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

Appendix C Result of principal component analysis: total variance explained.

Component	Total Variance Explained								
	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	1.841	14.165	74.684	1.841	14.165	74.684	2.422	18.628	73.016
4	1.091	8.391	83.075	1.091	8.391	83.075	1.308	10.059	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	.530	99.753						
12	.032	.243	99.996						
13	.001	.004	100.000						

Extraction Method: Principal Component Analysis

References

- [1] X. Pan, C.S. Han, K. Dauber, K.H. Law, Human and social behavior in computational modeling and analysis of egress, *Autom. Constr.* (2006). doi: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). doi: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). doi:10.1016/S0191-2615(03)00007-9.
- [4] A.E. Scheflen, N. Ashcraft, Human territories: How we behave in space-time., (1976).
- [5] D. Helbing, P. Molnár, Social force model for pedestrian dynamics, *Phys. Rev. E.* (1995). doi:10.1103/PhysRevE.51.4282.
- [6] M. Fukui, Y. Ishibashi, Self-organized phase transitions in cellular automaton models for pedestrians, *J. Phys. Soc. Japan.* (1999). doi:10.1143/JPSJ.68.2861.
- [7] 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). doi:10.1016/j.ssci.2016.04.005.
- [8] 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). doi:10.1016/j.physleta.2015.11.028.
- [9] D. Li, B. Han, Behavioral effect on pedestrian evacuation simulation using cellular automata, *Saf. Sci.* (2015). doi:10.1016/j.ssci.2015.07.003.
- [10] 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. A Stat. Mech. Its Appl.* (2010). doi:10.1016/j.physa.2010.08.032.
- [11] D. Helbing, I. Farkas, T. Vicsek, Simulating dynamical features of escape panic, *Nature*. (2000). doi:10.1038/35035023.
- [12] T.Q. Tang, L. Chen, R.Y. Guo, H.Y. Shang, An evacuation model accounting for elementary students' individual properties, *Phys. A Stat. Mech. Its Appl.* (2015). doi:10.1016/j.physa.2015.08.002.
- [13] J.H. Wang, J.H. Sun, Principal aspects regarding to the emergency evacuation of large-scale crowds: A brief review of literatures until 2010, in: *Procedia Eng.*, 2014. doi:10.1016/j.proeng.2014.04.001.
- [14] M. Kobes, I. Helsloot, B. de Vries, J.G. Post, Building safety and human behaviour in fire: A literature review, *Fire Saf. J.* (2010). doi:10.1016/j.firesaf.2009.08.005.
- [15] J.P. Keating, E.F. Loftus, The logic of fire escape, *Psychol. Today*. 15 (1981) 14.
- [16] A. Mintz, Non-adaptive group behavior, *J. Abnorm. Soc. Psychol.* (1951). doi:10.1037/h0063293.
- [17] M. Haghani, M. Sarvi, Crowd behaviour and motion: Empirical methods, *Transp. Res. Part B Methodol.* (2018). doi:10.1016/j.trb.2017.06.017.

- [18] 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.
- [19] 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). doi:10.1016/j.humov.2008.12.003.
- [20] 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). doi:10.1016/j.jbiomech.2015.12.030.
- [21] 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). doi:10.1109/TITS.2016.2542843.
- [22] J. Shen, T. Cheng, Clustering Analysis of Officer’S Behaviours in London Police Foot Patrol Activities, *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* (2015). doi:10.5194/isprsannals-II-4-W2-143-2015.
- [23] M. Haghani, M. Sarvi, Stated and revealed exit choices of pedestrian crowd evacuees, *Transp. Res. Part B Methodol.* (2017). doi:10.1016/j.trb.2016.10.019.
- [24] 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.
- [25] 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.
- [26] S. Wold, K. Esbensen, P. Geladi, Principal component analysis, *Chemom. Intell. Lab. Syst.* 2 (1987) 37–52.
- [27] P. McCullagh, J.A. Nelder, *Generalized linear models*, CRC press, 1989.
- [28] L. Breiman, *Classification and regression trees*, Routledge, 2017.
- [29] N. Cristianini, J. Shawe-Taylor, *An Introduction to Support Vector Machines and other kernel based learning methods*, *Ai Mag.* (2000). doi:citeulike-article-id:114719.
- [30] V. Kecman, T.-M. Huang, M. Vogt, Iterative single data algorithm for training kernel machines from huge data sets: Theory and performance, in: *Support Vector Mach. Theory Appl.*, Springer, 2005: pp. 255–274.
- [31] T. Cover, P. Hart, Nearest neighbor pattern classification, *IEEE Trans. Inf. Theory.* (1967). doi:10.1109/TIT.1967.1053964.
- [32] J. Ye, X. Chen, N. Jian, Impact analysis of human factors on pedestrian traffic characteristics, *Fire Saf. J.* (2012). doi:10.1016/j.firesaf.2012.05.003.
- [33] M. Haghani, M. Sarvi, Social dynamics in emergency evacuations: Disentangling crowd’s attraction and repulsion effects, *Phys. A Stat. Mech. Its Appl.* 475 (2017) 24–34.
- [34] L. Shi, Q. Xie, X. Cheng, L. Chen, Y. Zhou, R. Zhang, Developing a database for emergency evacuation model, *Build. Environ.* (2009). doi:10.1016/j.buildenv.2008.11.008.
- [35] 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. A Stat. Mech. Its Appl.* (2016). doi:10.1016/j.physa.2016.03.060.

- [36] 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). doi:10.1016/j.firesaf.2016.07.003.
- [37] 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). doi:10.1016/j.csda.2008.10.033.
- [38] 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). doi:10.1016/j.eswa.2017.03.042.
- [39] L. Chen, T.Q. Tang, H.J. Huang, J.J. Wu, Z. Song, Modeling pedestrian flow accounting for collision avoidance during evacuation, *Simul. Model. Pract. Theory.* (2018). doi:10.1016/j.simpat.2017.12.011.
- [40] M.M. Shuaib, Incorporating intelligence for typical evacuation under the threat of fire spreading, *Saf. Sci.* 106 (2018) 1–9.
- [41] 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.

Conflict of Interest

We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled “A machine-learning based study on pedestrian movement dynamics under emergency evacuation”.

Sincerely,

Ke Wang, Xiupeng Shi, Shunzhi Qian