Advances in Computer, Signals and Systems (2022)
Clausius Scientific Press, Canada

DOI: 10.23977/acss.2022.060301
ISSN 2371-8838 Vol. 6 Num. 3

Study on the Influence of Exit Door Characteristics on Evacuation Process based on XGBoost Regression and Pathfinder Simulation

Minghui Fan

College of Engineering, Tibet University, Lhasa, 850000, China
pluto116@163.com

Keywords: XGBoost, simulation, exit door characteristics, evacuation and escape

Abstract: In recent years, the emergency evacuation of personnel has received increasing attention, and how to achieve the rapid and safe evacuation of personnel has become a problem to be solved. In this paper, the model is established by XGBoost and combined with Pathfinder software simulation, the data centering of the finding is processed to get two exit data. Then regression analysis is performed by SPSSPRO software to get the characteristic importance plot of door width and evacuation rate and the prediction plot of test data. Similarly, the influence of visibility on the evacuation process is analyzed to provide some guidance for the crowd in the evacuation escape process.

1. Introduction

In recent years, along with the dramatic increase in population in modern cities, the number and scale of large-scale gatherings have increased, bringing not only many potential dangers to public places, but also a series of safety accidents to many occasions, which cannot be separated from the emergency evacuation problem [1]. Many experiments and simulations have been conducted to better understand the behavior of crowds in various situations. These studies have shown that crowd evacuation through doors of limited width or narrow passages is of concern. The experiments and related data also illustrate that the actual effectiveness of evacuation can be affected by factors such as the width of the door, the composition of the crowd, and the location of the exit [2]. And because factors such as the realism of the environment and background, crowd density, and sample size differ from the actual situation, the results of simulations through simulators deviate from those of experiments conducted with real people, and the results of a large number of studies are currently controversial and contradictory.

In this paper, the width of the door and visibility are taken as the main variables, combined with the principles of proximity and minimization for analysis, and the XGBoost-based personnel evacuation model is established and simulated by Pathfinder software. Finally, the effects of visibility and door size on the evacuation process are illustrated by the regression analysis results and simulation.

2. Acquisition of data and assumptions

The data in this paper were obtained from the 2022 Certification Cup (SPSSPRO) Mathematical Modelling Web Challenge A. The following assumptions were made to facilitate model building: (a) there are no barriers in the square room, and the maximum number of people does not exceed the maximum capacity of the room, and there is only one door to evacuate the room; (b) the people in the square room are aware of the basic conditions of the room; (c) the people in the room and the hall anxiety changes with a certain pattern; (d) visibility in the hall has a linear relationship with the level of anxiety of the personnel; (e) there is no influence of unusual factors and unusual events in the model and the experiment.

3. Personnel evacuation model based on XGBoost algorithm

3.1 Introduction to the model

XGBoost algorithm is a class of synthetic algorithms that combine basis functions and weights to form a good fit to the data. Due to the advantages of XGBoost model such as strong generalization ability, high scalability, and fast computing speed, it is popular in the fields of statistics, data mining, and machine learning [3].

3.2 Model building

Exit doors and visibility are variables, and in order to control the variables for qualitative analysis, this paper uses SPSSPRO software to center the data, and after preserving the consistency of the variables, two identical doors are obtained and then analyzed. For simplicity of analysis, the width of the exit door, anxiety level, age mean, total number of people in the hall and the number of people at the exit, six variables are analyzed here. For a data set containing n m-dimensional entries, the XGBoost model can be expressed as

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), f_k \in F(i = 1,2,\cdots,n) \tag{1}$$

$$F = \{f(x) = w_{q(x)}\}(q: R^m \rightarrow \{1,2,\cdots,T\}, w \in R^T) \tag{2}$$

$F$is the set of CART decision tree structures,q is the tree structure of the sample mapping to the leaf nodes, and T is the number of leaf nodes, w is the real number fraction of leaf nodes. The objective function of the XGBoost model can be divided into an error function term and a model complexity function term Ω.L The objective function of the XGBoost model can be divided into an error function term and a model complexity function term Ω. The objective function can be written as

$$O_{bj} = L + \Omega \tag{3}$$

$$L = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{4}$$

$$\Omega = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \tag{5}$$

In the above equation, the γT is the L_1 regular term, the $\frac{1}{2\lambda\sum_{j=1}^{T}w_j^2}$ is the L_2 the regular term.

When training the model optimally using the training data, it is necessary to keep the original model unchanged and add a new function f to the model so that the objective function is reduced as much as possible by the following process[4].

$$\hat{y}_i^{(0)} = 0 \tag{6}$$

$$\hat{y}_i^{(1)} = \hat{y}_i^{(0)} + f_1(x_i) \tag{7}$$

$$\hat{y}_i^{(2)} = \hat{y}_i^{(1)} + f_2(x_i) \tag{8}$$

$$\cdots\cdots$$

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + f_t(x_i) \tag{9}$$

In the above equation, the $\hat{y}_i^{(t)}$ is the first t time model prediction, the $f_t(x_i)$ is the t new function added for the second time, at which time the objective function is expressed as

$$Obj^{(t)} = \sum_{i=1}^{n}\left(y^i - \left(\hat{y}_i^{(t-1)} + f_t(x_i)\right)\right)^2 + \Omega \tag{10}$$

In the XGBoost algorithm, a second-order Taylor expansion is performed to obtain an approximate objective function in order to quickly find the parameter that minimizes the objective function.

$$Obj^{(t)} \approx \sum_{i=1}^{n}[\left(y^i - \hat{y}_i^{(t-1)}\right)^2 + 2(y^i - \hat{y}_i^{(t-1)})f_t(x_i) - h_i f^2{}_t(x_i)] + \Omega \tag{11}$$

When the constant term is removed, it is known that the objective function is only related to the first- and second-order derivatives of the error function. At this point, the objective function is expressed as

$$Obj^{(t)} \approx \sum_{i=1}^{n}\left[g_i w_{q(x_i)} + \frac{1}{2}h_i w_{q(x_i)}^2\right] + \gamma T + \frac{1}{2}\sum_{j=1}^{T} w_j^2 \tag{12}$$

$$bj^{(t)} \approx \sum_{i=1}^{n}\left[(\sum_{i\in I_j} g_i)w_i + \frac{1}{2}(\sum_{i\in I_j} h_i + \lambda)w_j^2\right] + \gamma T \tag{13}$$

If the structural part of the tree q is known, the objective function can be used to find the optimal w_j^* and get the optimal objective function value. Its essence can be categorized as a quadratic minimax solution problem. The solution yields.

$$w_j^* = \frac{-\sum_{i\in I_j} g_i}{\sum_{i\in I_j} h_i + \lambda} \tag{14}$$

4. Evacuation simulation based on Pathfinder software

The software simulation method can predict the evacuation time in advance during the peak and low peak periods, and better find the optimal evacuation route plan, so as to provide scientific and correct guidance for the crowd in evacuation.

Pathfinder is a new, simple, intuitive and easy-to-use intelligent emergency evacuation assessment system developed by Thunderhead engineening, Inc. It utilizes technology from the fields of computer graphics simulation and game characters, and provides a graphical virtual walkthrough of each individual movement in multiple groups [5]. For simulation, Pathfinder is a simple, intuitive, and easy-to-use new intelligent emergency evacuation and escape assessment system. Therefore, in this paper, Pathfinder software is used to simulate the evacuation process in different situations by setting the width of 0m, the width of 50m, and the total area of 2500m^2 square hall, as shown in Figure 1, and there are two doors and the clarity of the room can be adjusted.

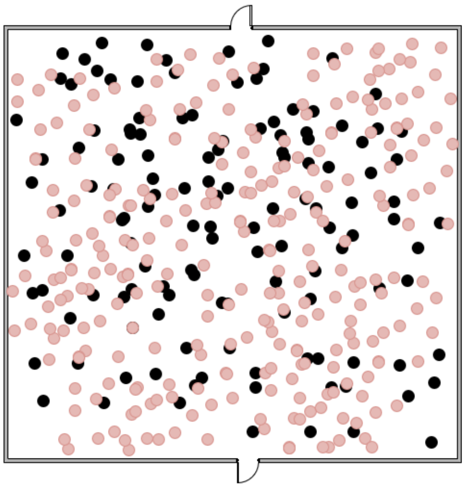

Figure 1. Simulation of a hall situation.

4.1 Evacuation of people simulation

In the software, a random distribution of personnel composition, as shown in Table 1, is used to create groupings for simulation.

Table 1. Composition of personnel.

Category	Children	Teenagers	Youth	Middle-aged	Elderly
Age status/year	1-14	14-18	18-35	35-60	>60
Traveling speed/(m/s)	0.95	1.51	1.45	1.39	1.01

4.2 Change the width of the door for simulation

The width of the door is expanded to 2d from the original d width, and then the evacuation simulation is performed, as shown in Figure 2.

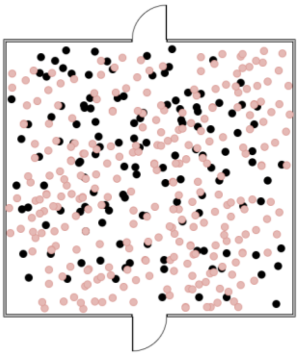

Figure 2. Changing the width of the door for simulation.

4.3 Simulation by changing the visibility of the hall.

The reduced visibility in the lobby will directly affect the anxiety level of the crowd [6] and also indirectly cause misjudgment of the door location (i.e., the red dot in the figure) as shown in Figure 3.

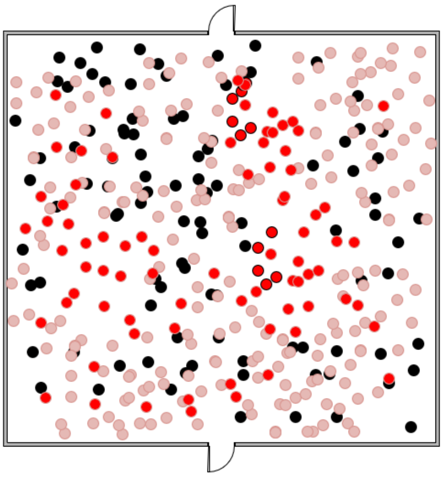

Figure 3. Changing the visibility of the hall for simulation.

5. Model solving

5.1 Centralized processing of data

Data centering means that after subtracting the mean from each score, the centered data (centered score) is obtained. In general, the predictor variables that need to be entered to study the moderating effect are the interaction term (XZ) and the individual variables (X and Z). However, since the interaction term is directly derived from the product of the variables, the covariance of the predictor variables can become large. Moreover, when the variables are all quantitative, data centering is an essential process.

To this end, the data collected were data-centered through SPSSPRO software for variables such as width of exit door, anxiety level, disaster level, age mean, and headcount flow, respectively [7].

5.2 Topic information extraction

For the expected time, the principles are "the principle of proximity understanding" and "the principle of less understanding". The principle of proximity is to understand the distance between the coordinates of each person in the hall and the two exits, the shortest is the optimal solution; the principle of less understanding, in the dynamic process, the remaining people in the hall to make a judgment on the two exits, that is, the judgment of the number of people less than the exit is the optimal evacuation plan. The above two principles for the main choice of program.

For visibility understanding, when the visibility in the hall is reduced, it can directly cause the anxiety level of the personnel in the hall, and to some extent, misjudge the correct location and direction of the exit. So here, we convert the constraint of low visibility into a misjudgment of the exit location, and increasing the level of anxiety can directly lead to a misjudgment of the exit. Thus, reduced visibility is equated to some extent with increased anxiety.

In this paper, we only consider the hall exit located at the opposite side position and then evacuate as shown in Figure 4. The evacuation process in this case will have a double-arc distribution.

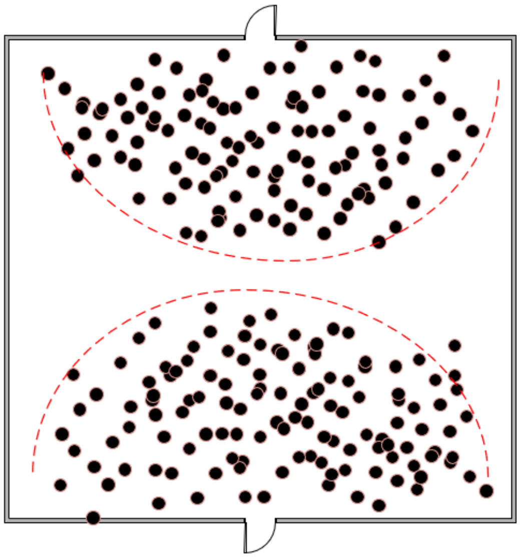

Figure 4. Double arc evacuation distribution.

5.3 Solution of XGBoost regression model

In this paper, the XGBoost regression model is built by the training set data. After that, the feature importance is calculated by the established XGBoost. The established XGBoost regression model is applied to the training and testing data to get the model evaluation results. However, because XGBoost has randomness, the result of each operation is not the same, if this training model is saved, the subsequent data can be directly uploaded to this training model for calculation of prediction [8]. In which the model is solved based on eight indicators of doorway number flow, anxiety level, age mean, hall number, doorway width, centralized value of doorway width, expanded anxiety level, and expanded doorway width, where the model parameters are shown in Table 2.

Table 2. Model parameters.

Parameter Name	Parameter Value
Training time	0.174s
Data Slicing	0.8
Data shuffle	Yes
Cross-validation	No
Base Learners	gbtree
Number of base learners	100
Learning Rate	0.1
L1 canonical term	0
L2 canonical term	1
Sample Collection Sampling Rate	1
Tree feature sampling rate	1
Node feature sampling rate	1
Minimum weights of samples in leaf nodes	0
Maximum depth of the tree	10

The characteristic importance ratios of each indicator obtained after SPSSPRO run are shown in Figure 5 and Figure 6.

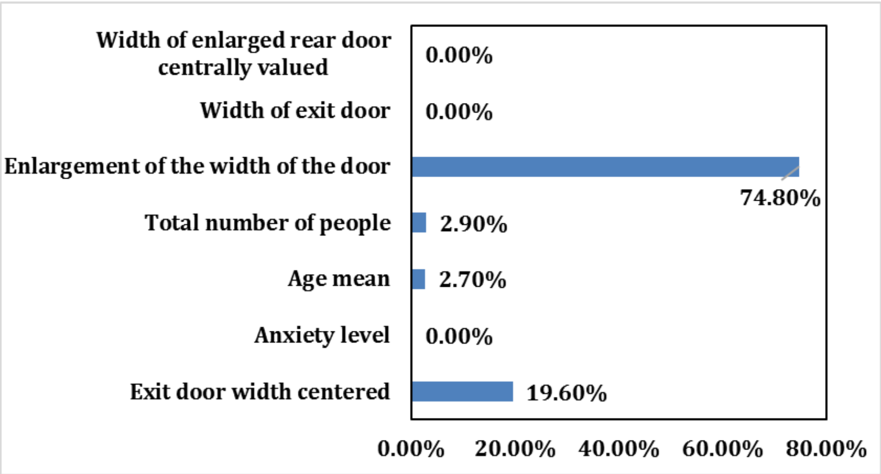

Figure 5. Changing the characteristic importance ratio of door width.

As can be seen from the proportion in the figure, the highest proportion of the characteristic importance of changing the door width reaches 74.80%, followed by the proportion of the characteristic importance of two doors reaching 19.60%, followed by the total number of people in the hall and the mean age of the number of people ranked next with nearly 3%. These can indicate that the door width is positively correlated with the evacuation process, and the larger the door width, the faster the evacuation process.

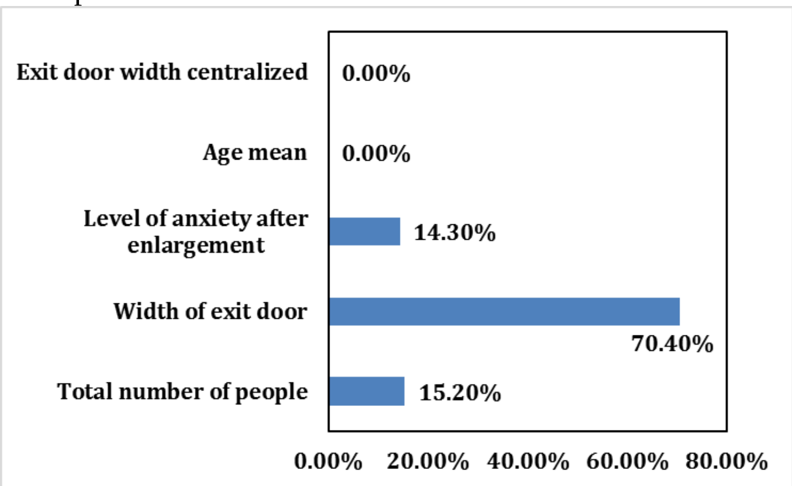

Figure 6. The proportion of importance of characteristics that change the level of anxiety.

Since the reduced visibility in the hall is relatively difficult to simulate, this paper converts the reduced visibility into increased anxiety to study the effect on the evacuation process [9]. The proportion of the characteristic importance of the increased anxiety level is 14.30%, which is relatively high. It shows that visibility is positively related to evacuation process, and the lower the visibility, the slower the evacuation process.

The evaluation about the XGBoost regression model can be obtained by the test, as shown in Table 3.

Table 3. Model evaluation.

	MSE	RMSE	MAE	MAPE	R^2
Training set	0.266	0.516	0.365	0.182	0.991
Test set	0.276	0.525	0.364	0.183	0.991

The prediction evaluation metrics of the cross-validation set, training set and test set are shown specifically in the above table, and the prediction effectiveness of XGBoost is measured by quantitative metrics. Among them, the evaluation metrics of the cross-validation set can be used to continuously adjust the hyperparameters to obtain a reliable and stable model.

Compare the true and predicted values for changing the door width and changing the visibility, as shown in Figure 7 and Figure 8.

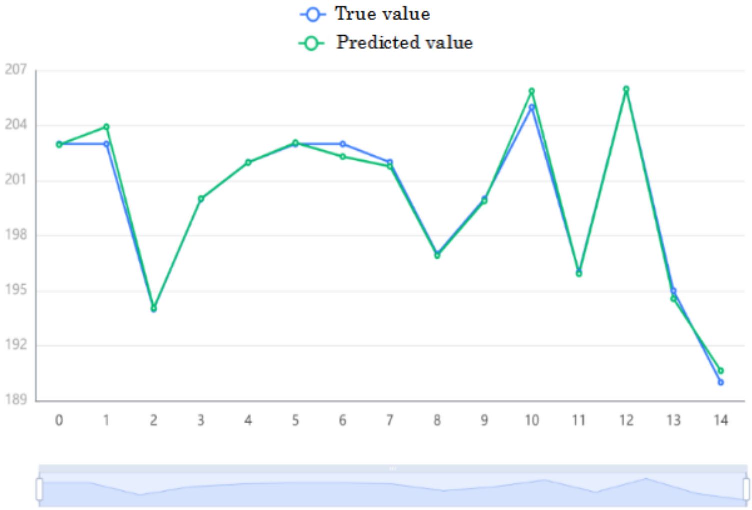

Figure 7. True and predicted scenario for changing door width.

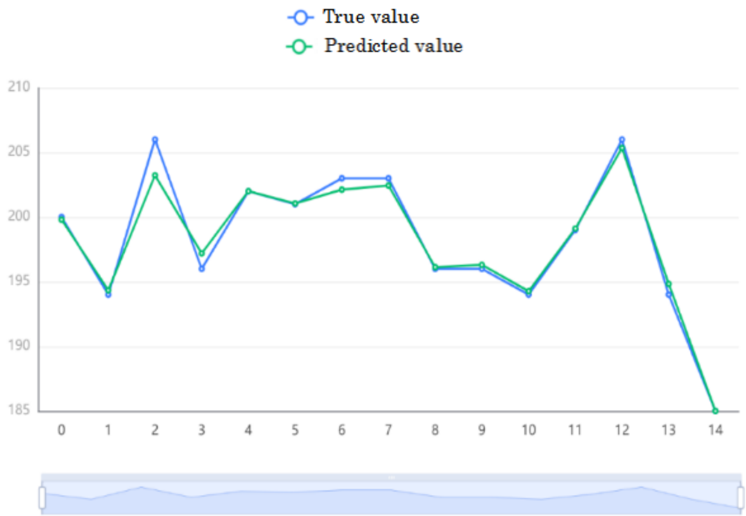

Figure 8. Changing Visibility Real and Predicted.

From the above two real and predicted situations, it can be seen that the XGBoost regression model established in this paper has relatively low error and high practicality. Also, it can be obtained that: the different door widths positively affect the evacuation process, and the visibility level also positively affects the evacuation process of people.

6. Conclusion

In this paper, the model was built by XGBoost and combined with Pathfinder software simulation, the data centre of finding was processed to obtain two exit data. Using SPSSPRO software for regression analysis, the characteristic importance plot of door width versus evacuation rate and the prediction plot of the test data were obtained. The highest characteristic importance proportion of changing door width reached 74.80%, followed by two doors with 19.60%, with the total number of people in the hall and the mean age of the number of people coming in second with close to 3%. Door width was positively correlated with the evacuation process, with the wider the door, the faster the evacuation process. The proportion of characteristic importance accounted for by increasing anxiety levels reached a relatively high 14.30%. This means that visibility is positively related to the evacuation process, and the lower the visibility, the slower the evacuation process [10]. From the model validation, it can be seen that the XGBoost regression model has relatively low error and high practicality. In addition, the difference in door width positively affects the evacuation process, and the visibility also positively affects the evacuation process.

References

[1] Liu, Yixue, Mao, Zhanli, Liao, Mingyu. A metacellular automaton model for crowd evacuation considering follower threshold and subject characteristics[J]. China Science and Technology of Safety Production, 2021, 17(12):156-161.

[2] Zhao LQ, Guo MQ, Tang SHX, Tang JJ. Adaptive crowd evacuation simulation model based on finite rationality constraints[J]. Journal of System Simulation:1-9.

[3] Zheng Wenli. Construction of an ischemic stroke recurrence prediction model based on XGboost algorithm[D]. Zunyi Medical University, 2021.

[4] Sun Yutong. Research on crowd evacuation simulation method integrating density field and social force model[D]. Shandong Normal University, 2021.

[5] Liu Ying, Liu Zimin. Study on crowd evacuation simulation based on Pathfinder for small and medium-sized stadiums[J]. Low Temperature Building Technology, 2021, 43(04):6-10.

[6] Yao Zhenzhen. Research on crowd evacuation simulation method based on reinforcement learning and deep residual network learning[D]. Shandong Normal University, 2020.

[7] Ji Zhonghua. Control and Simulation of Dense Crowd Evacuation[D]. Nanjing University of Posts and Telecommunications, 2019.

[8] Ma Qianqian, Sun Dongxu, Shi Jinming, He Xianying, Zhai Yunkai. Prediction of tumor disease risk in adult population based on support vector machine and XGboost[J]. Chinese general medicine, 2020, 23(12):1486-1491.

[9] Li C. H., Bi Chenggong, Li Z. J. Crowd evacuation model based on improved PSO algorithm[J]. Journal of System Simulation, 2020, 32(06):1000-1008.

[10] Chen Yan. Research on crowd evacuation simulation in hospitals based on Pathfinder[J]. Henan Science and Technology, 2018, (29):107-109.