

Feature-Filtering Data-Mining Algorithm for Urban Waterlogging Path Optimization in Extreme Weather

Wei Zhong^{1a}, Youdong Zhang^{1b}

School of management, Tianjin university of technology, China

^a zhongwei@tjut.edu.cn, ^b 728990038@qq.com

Keywords: city waterlogging; feature selection; risk aversion; data mining; municipal management

Abstract: Good personnel hedge scheme which can reduce the negative effects of urban waterlogging disaster is one of the effective ways and scientific means to improve the level of urban municipal management. Strong randomness is the characteristics of stormwater runoff route under city extreme weather, this paper analyzes the characteristics and takes the dull phenomenon of traditional hedging model into account under random emergencies in extreme weather. An urban waterlogging personnel hedge mechanism is proposed based on feature selection data mining route decision making model. The flow path of urban waterlogging region is added into the model. Clustering method is used to optimize the extreme weather urban waterlogging personnel hedge path characteristics. Then the optimal data mining decision support analysis model is constructed and a personnel hedge path of urban waterlogging disaster is concluded.

Introduction

One of the most important tasks in the urban flood disaster contingency plan is to realize the scientific and orderly evacuation of the personnel in the area of the waterlogged disaster, which refers to all people, vehicles and other related objects in the waterlogged area that need to be evacuated. Many scholars at home and abroad have done a lot of research and practical work in disaster risk prevention^{[1]-[3]}, but they don't consider that the urban waterlogging disasters occurred in extreme weather have huge rainwater runoff in short time, high random line and random change of flow direction. The traditional escape route model based on constructed mitigation model to calculate the hedge path is vulnerable to randomly change and makes the calculation results inaccurate. In view of the above problems, this paper proposes a personnel hedge path selection model based on feature screening data mining analysis in extreme weather urban waterlogging. First urban waterlogging area path characteristics data are selected and optimized based on clustering method, then a data mining optimization model is constructed and the most reasonable hedge path is obtained. The algorithm can improve the accuracy and scientific of decision in extreme weather.

Analysis on Personnel Hedging Path Demand in City Waterlogging Disaster and Decision Principle

Analysis on the demand and related theories of the decision making for personnel safe path

Urban waterlog disaster is a natural disaster occurred in city whose infrastructure construction is not scientific and oversized. Urban waterlogging disaster usually brings the affected city damages of traffic, electric power, communication, drainage facilities and citizens and has negative influence on the reasonable and orderly operation of city. Based on the above analysis, the relevant departments of the city must prepare ahead of time corresponding contingency plans in order to reduce urban waterlogging disasters caused by the loss. Waterlogging warning response mechanism on traffic is a very important part of the city waterlogging disaster emergency plan, when the city waterlogging disaster occurs, evacuating the people in the waterlogging disaster area of scientifically and reasonably is a key step to reduce the city waterlogging disaster losses, constructing a scientific and reasonable urban waterlogging disaster personnel hedge path mechanism is an effective way to reduce urban waterlogging disaster losses and to ensure the rational operation of urban.

Traditional personnel hedge path planning algorithm usually selects the optimal staffing hedge path through the establishment of a decision model, and then complete the waterlogging personnel hedge path decision. The decision making model of the traditional personnel safe haven path can be established using the following formula:

$$R = \frac{2S(\log_2(n) + U_i)^2}{j \sqrt{i^2 N + 2P + 1}} \quad (1)$$

In the formula (1), s representatives hedge path and the distance between the minimum value, n represents the number of all hedge path, U_i denotes the with a hedge path, N represents urban waterlog disaster area submerged path number, P represents the number of original path used in urban waterlogging disaster areas and ϕ represents the residual coefficients of waterlogging mobile route. Through the mathematical formula, a city waterlogging personnel hedge path planning model can be built and then the shortest hedge path to the affected people can be calculated. And then, by using the following formula to calculate the reasonable parameters of the change to hedge can be calculated by fomular (2), in order to make the safe path scientific.

$$h = \frac{R \log_2 n - U_i}{i + N^2 + 2p} \quad (2)$$

The Disadvantages of Traditional Algorithm in the Extreme Weather

City waterlogging hazards are caused by the collection of urban storm water runoff, one of its effects is the collection has constraints on the reasonable operation in the municipal traffic. Traditional algorithm usually considers the influence of stormwater runoff on traffic when the urban waterlogging occurs, the research object is the selection of personnel hedge path during the collection process of stormwater when the urban waterlogging occurs. But the city waterlogging usually occurs in extreme weather (according to the National Office of "flood control manual" provisions, when the cumulative rainfall of 24 hours is more than 50 mm it can be defined as extreme weather), which is characterized by a huge amount of precipitation in a short period of time, runoff formed similar to flood, the action lines have more random than other cases. The shortcomings of the traditional algorithm is it takes the city rainwater collection process as a complete continuous process to analyze personnel hedge path selection, without considering the characteristics of city rainwater runoff under extreme weather, according to the formula (1) the following conclusion can be obtained, if the random of city rainwater runoff flow path increases, it will cause waterlogging area flooded path N and the number of all hedge path number n error. When the decision model allows an error, the error is to the calculated residual ϕ , but when the error of N and n is large enough, ϕ will be too large. And then it leads to the deviation of the model R, and the robustness is decreased. At the same time, according to the formula (2) that if the deviation of R model is too large, it will reduce the rationality of personnel hedge path η and influence the accuracy of path selection

Optimization Model of City Waterlogging Disaster Personnel Hedge Path during Extreme Weather

The Screening Characteristics in City Waterlogging Area

When the feature of urban waterlog disaster area path is rich, the hedge path feature vector $F=\{u_1, u_2, \dots, u_p\}$ can be used to mark in this region. The complexity of city waterlogging action path calculation will increase with the number of personnel hedge path feature P increases. Therefore, in the selection of the feature vector, it should calculate the representative feature vector as possible, so as to reduce the volume of the characteristics, and improve the accuracy and scientific of results when reduce the workload.

This paper analyzes from the view of association between urban waterlogging area regional characteristics and personnel hedge path, the coefficient of correlation between the article i hedging path characteristic component and the city waterlogging area can be calculated with formula (3):

$$d_i = \frac{p \times (U_i - 2)^2}{(p^2 - i)(p^2 + i)}. \quad (3)$$

Among them, P is the number of personnel hedge path feature, δ_i is the coefficient of association between rule i hedge path feature component and the city waterlogging area. Clustering method is used to optimize the characteristics of the safe haven route, to improve the accuracy of calculational model, and to reduce the workload of calculation. Formula (4) is used to optimize the characteristics of the personnel hedge path.

$$g = \frac{tU_i \times \lg(t^2 - 1) \times i}{\sqrt{\sum_{i=1}^p (U_i^2 + N)^2}}. \quad (4)$$

Among them, t says the extreme weather city waterlogging duration. γ says clustering parameters.

Formula (5) is used to calculate the similarity between the characteristics of the hedge path in the personnel after cluster optimization:

$$\text{sim}(U_i, U_j) = \frac{\sum_{i=1}^p d_i \times d_j}{\sqrt{(\sum_{i=1}^p d_i)^2 (\sum_{j=1}^p d_j)^2}} \quad (5)$$

According to the above process, we can analyze and select out the characteristics of urban waterlogging area path, and the cluster analysis method can be used to optimize and select personnel hedge path characteristics, and then a accurate basic data of the urban waterlogging personnel hedge path optimization model can be established.

The Optimization Model of City Waterlogging Data Mining

The main flow of urban waterlogging data mining optimization model is mapping the urban waterlogging personnel hedge path into feature space by establishing a mapping relationship, and then use the principle of normalization to work correspondingly.

First set (y_i, z_i) ($i=1,2,\dots,p$) as urban waterlogging personnel hedge path feature space point and the y_i is the data source of the model. z_i is the output source of the model and p is the number of features.

Formula (6) is used to express the nonlinear regression function in the ideal case of a safe haven path:

$$g(y) = x \cdot y + c \quad (6)$$

Among them, x and c are the relevant parameters of the safe haven path.

By using equation (7), the nonlinear regression function is transformed into a more optimal function:

$$j = \min \frac{1}{2} xU_i + \sum_{i=1}^p (d_i + d_i^*). \quad (7)$$

At the same time, the optimization function needs to meet the following conditions:

$$\begin{cases} z_i - x \cdot y - D \cdot c \leq v + d_i \\ x \cdot y - c - D \cdot z_i \leq v + d_i^* \end{cases} \quad (8)$$

Among them, d_i and d_i^* is the relaxation parameter. v is the control parameter of mining city waterlogging data optimization model. D is the penalty parameter.

The value coefficient of personnel hedge path characteristics can be calculated by formula (9).

$$x = \sum_{i=1}^p (b_i - b_i^*)k(y_i). \quad (9)$$

Among them, b_i and b_i^* are the control factors, but they also need to meet the conditions in formula (10):

$$\sum_{i=1}^p b_i = \sum_{i=1}^p b_i^* \quad 1 < b, b < d \quad (10)$$

Using the following formula, we can draw the optimal model of data mining by formula (11):

$$g(y) = \sum_{i=1}^p (b_i - b_i^*) l(y_i, y) + c \quad (11)$$

Among them, $l(y_i, y)$ is used to calculate the correlation coefficient of the escape route in the data mining, and it can be calculated by formula (12):

$$l(y, y_i) = \exp\left|\frac{y - y_i}{s_i}\right|^2 \quad (12)$$

The calculation model of mining city waterlogging personnel hedge path data can be obtained as formula (13):

$$g(y) = \sum_{i=1}^p (b_i - b_i^*) \exp\left|\frac{y - y_i}{s_i}\right|^2 + c \quad (13)$$

Among them, parameter σ_i and c need to be optimized.

Through the above analysis it is the personnel hedge path optimization decision support model under extreme conditions in urban waterlogging area, path characteristics selected by clustering method are used in the model for screening data mining computations, and then the reasonable waterlogging hedge path is calculated to improve the accuracy of choosing personnel hedge path under the extreme weather, at the same time, strengthen the City corresponding early warning capability.

Conclusion

This paper proposes a personnel hedge path selection model based on feature selection data mining in extreme weather during city waterlogging. First urban waterlogging area path characteristics data are selected and optimized based on clustering method, then the optimal data mining decision support analysis model is constructed and a personnel hedge path of urban waterlogging disaster is concluded. It can give advice to personnel hedge work during urban waterlog disaster under extreme weather.

Acknowledgements

Sponsored by Ministry of education, humanities and Social Science Fund Project - oriented plan corresponding mechanism of urban water logging disaster scenarios simulation and model deduction (14YJC630211)

References

- [1] Zuo Xiaobing, Liu Jingang. Research and Simulation of Flood Escape Routes Decision-Making Method. *Computer Simulation*, 2013, 30(4): 299- 302. (in Chinese)
- [2] He Ming, Chen Guohua, Du Youxiang, Weng Xi, Jiang Jun. Simulation Method for Personnel Evacuation under Fire Based on Dijkstra. *China Safety Science Journal*, 2010, 20(12): 46-51 (in Chinese)
- [3] Ma Jiyang. Study on the Traffic Emergency Evacuation Model of Chengdu Based on the Spatial Diffusion of Rainstorm-waterlogging. *Sichuan Normal University Master Thesis*, 2014 (in Chinese).
- [4] Gong Xusheng. Human-machine Cooperation Path Planning Based on Three Dimensional Terrain. *National University of Defense Technology master thesis*, 2010