TSINGHUA SCIENCE AND TECHNOLOGY
ISSN 1007-0214 67/67 pp418-422
Volume 13, Number S1, October 2008

Urban Fire Risk Clustering Method Based on Fire Statistics*

WU Lizhi (吴立志)**, REN Aizhu (任爱珠)

Institute of Engineering Disaster Prevention and Mitigation, Tsinghua University, Beijing 100084, China

Abstract: Fire statistics and fire analysis have become important ways for us to understand the law of fire, prevent the occurrence of fire, and improve the ability to control fire. According to existing fire statistics, the weighted fire risk calculating method characterized by the number of fire occurrence, direct economic losses, and fire casualties was put forward. On the basis of this method, meanwhile having improved K-mean clustering arithmetic, this paper established fire risk K-mean clustering model, which could better resolve the automatic classifying problems towards fire risk. Fire risk cluster should be classified by the absolute distance of the target instead of the relative distance in the traditional cluster arithmetic. Finally, for applying the established model, this paper carried out fire risk clustering on fire statistics from January 2000 to December 2004 of Shenyang in China. This research would provide technical support for urban fire management.

Key words: urban fire; fire statistics; fire risk; K-mean clustering

Introduction

In modern society, fire statistics and fire analysis have become key ways for us to understand the law of fire, prevent the occurrence of fire, and improve the ability to control fire. By means of analyzing fire statistics, we could correctly summarize and master the law of fire, then advance the development of fire service.

So far, many scholars have studied fire statistics. For example, by graph, Wang Haihui et al.[1] analyzed the fire statistics from the middle of 1980s to 1994 in China, including distribution of direct property loss and catastrophic fire accidents per year; distribution of fire direct property loss, casualties, and injured per month; distribution by fire occurrence places; and distribution by fire causes etc. Yang Lizhong et al.[2], Holborna P G et al.[3], Tommy Rosengerg[4], and others also researched on fire statistics by applying the ways of graph analysis or mathematical statistics, and drew some useful conclusions.

Zheng Shuangzhong[5] pointed out that urban fire is featured by abruptness, chain reaction, complexity, difficult disposal etc. Fire risk is the result of interaction of Category I fire hazard sources and Category II fire hazard sources, and its degree is greatly influenced by fire loss and severity. The degree of fire risk could be embodied to a certain extent by the loss of historical statistics, and urban comprehensive fire risk situation could be assessed on the basis of statistical result in combination with present situation of urban fire. Focus on urban fire statistics, this paper studied the automatic classifying method toward fire risks.

1 Urban Fire Risk Measurement Method Based on Fire Statistics

Urban fire risk was the compositive embodiment of fire occurrence frequency and loss. The mainly factors should include the number of fire occurrence, direct economic loss, indirect loss, casualties, injured,

Received: 2008-05-30

∗ Supported by the National Key Technologies Research and Development Program of China during the 10th Five-Year Plan (No. 2001BA803B02-02)

∗∗ To whom correspondence should be addressed.
E-mail: wulizhi119@sohu.com; Tel: 86-0316-2067614

damaged area, environmental loss, and political influence etc. While from actual situation of Chinese fire statistics, number of fire occurrence, direct economic loss, casualties were quite comprehensive. Data of damaged area is seriously missing. There were neither statistics of fire indirect economic loss nor environmental loss etc. Therefore, the number of fire occurrence, direct economic loss, and casualties were regarded as the basis for fire risk assessment in this paper. With regard to other influencing factors, we could employ translation method to take the influence degree towards fire risk into consideration.

The number of fire occurrence, direct economic loss, and casualties has different importance in fire risk assessment. They should be allocated correspondingly normalized weight. Weights serve to express the importance of each criteria compared with the others. This paper defined urban fire risk measurement method based on fire statistics as

$$R_i = \{(C_i, L_i, D_i), (W_\mathrm{C}, W_\mathrm{L}, W_\mathrm{D}) \mid C_i, L_i, D_i \in \Omega\} \quad (1)$$

where R_i is representing urban fire risk in No. i period of time; C_i is representing the number of urban fire occurrence in No. i period of time; L_i is representing urban fire direct economic loss in No. i period of time; D_i is representing the urban fire casualties in No. i period of time; Ω is representing urban serial fire statistics set; W_C, W_L, and W_D are respectively representing the weight of fire occurrence number, fire loss, and casualties.

Figure 1 is the fire risk acceptable curved chart characterized by the number of fire occurrence, director economic loss and casualties. The fire risk under the curved chart is obviously acceptable, while the fire risk above it is not. It is supposed to take measures to decrease fire risk.

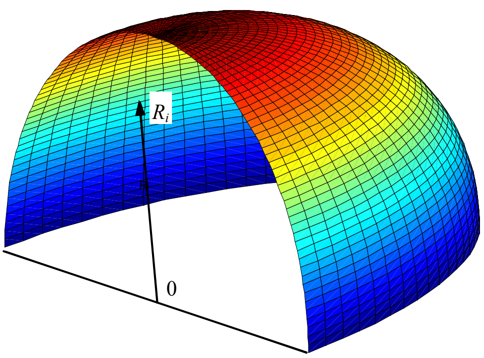


Fig. 1 Urban fire risk acceptable curved chart

2 Urban Fire Risk Clustering Method

As for urban fire risk, this paper studied its dynamic classifying method by means of clustering analysis. Clustering analysis is an important and widely-applied technology in data analysis. Clustering means that data object is divided into several classes or clusters. The objects in the same cluster possess higher similarity, while there is much difference among the objects in different clusters. Similarity is calculated by describing attribute of the object[6].

There are many clustering arithmetic, which usually involve partitioning method, density-based method, grid-density method, and model-based method etc.[7] We should choose appropriate method on the basis of data type, clustering intention and specific application requirements etc. The intention of analyzing urban fire statistics is to find out the distribution law of fire risk among fire statistics. Therefore, this paper finally selected Partitioning Method. For the sake of vividly portraying fire risk distribution, the number of clusters was defined 5 and K-mean clustering was employed here.

In traditional clustering arithmetic, the distance of p dimension object $x_i = (x_{i,1}, \cdots, x_{i,p})$ usually is expressed in Minkowski distance as

$$d_q(x_i, x_j) = \| x_i - x_j \| = \left[\sum_{k=1}^{p} | x_{ik} - x_{jk} |^q \right]^{1/q} \quad (2)$$

Every factor characterizing fire risk was of internal correlation, when it reacted on fire risk. Accordingly, the equation of dissimilarity distance between fire risks should be redesigned. Taking the connotation of fire risk into account and combining the idea of Euclid distance, we improved the dissimilarity distance equation of traditional clustering arithmetic in this paper, and defined the equation of fire risk as

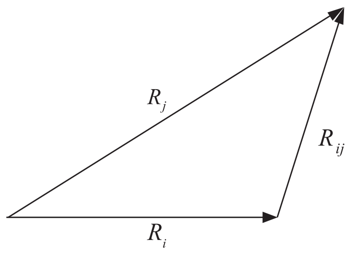
$$R_i(C_i, L_i, D_i) = \sqrt{W_\mathrm{C} C_i^2 + W_\mathrm{L} L_i^2 + W_\mathrm{D} D_i^2} \quad (3)$$

Here, W_C, W_L, and W_D represent weights. The values of them could be calculated according the actual situation.

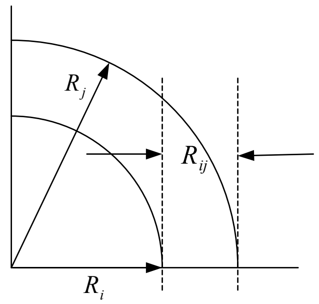
Then, the dissimilarity distance of urban fire risk between time T_i and time T_j is given as

$$d(R_i, R_j) = | R_i - R_j | \quad (4)$$

Figure 2 shows the difference on the calculating methods of dissimilarity distance of traditional clustering arithmetic and fire risk clustering arithmetic.



(a) Distance between traditional objects



(b) Distance between objects for fire risk

Fig. 2 The calculating method of two distance

In accordance with the above analysis, this paper adopted K-mean theory and method to perform clustering analysis towards urban fire risk. The core idea of K-mean method is that data are partitioned to different clusters through iteration, so as to achieve the minimization of target function and then to make the generated clusters closer and more independent. First of all, k objects were selected randomly as the centroids of initial k clusters, then other objects were allocated to the nearest clusters according to the distance from every centroid, therefore newly-formed centroid could be calculated. This iteratively re-positioning process was repeated continuously till reaching the minimization of target function[8].

Supposed p is data object, c_i is the mean of cluster C_i, then the usual target function is squared error criterion function as

$$E = \sum_{i=1}^{k} \sum_{p \in C_i} \| p - c_i \|^2 \tag{5}$$

The distance of this target function is called Euclid distance.

K-mean clustering arithmetic of urban fire risk was designed on the basis of the above idea. Inputting: (1) attribute values of n fire risks $R_i(C_i, L_i, D_i)$ during period of T (such as one hour), where $i = t_1, \ldots, t_n$; (2) the number of expected clusters k (in this paper $k = 5$); (3) the weight of each fire risk attribute (W_C, W_L, W_D). Outputting: k clusters $C_i, i = 1, \ldots, k$, which makes the target function minimized.

The fire risk clustering process as following:

(1) Calculating the value of fire risk R_i at any moment t_i in accordance with Formula (3);

(2) Selecting k objects as the centroid of initial cluster;

(3) Calculating the distance between object R_i and centroid C_j of every cluster in accordance with Formula (2);

(4) Recalculating the mean value of every new cluster;

(5) Repeating the above steps (3) and (4) until the centroid of cluster unchanged.

3 Practical Example

This paper utilized the fire statistics of Shenyang in China during 2000-2004 as a study case and made one hour as one basic time unit to partition fire statistics of these five years. Number of fire occurrence, direct economic loss, and casualties were the attributes of fire risk.

Normalization processing was performed towards the attributes of fire risk. Namely, real value was divided by maximum value to get current observed value, as

$$\bar{A}_i = \mathrm{std}(A_i) = \frac{A_i}{\max\limits_{t \in T}(A_t)} \quad i = 1, 2, ..., T \tag{6}$$

$$A_i \in \{(C_i, L_i, D_i) \mid C_i, L_i, D_i \in \Omega\}$$

In terms of the analytical hierarchy process (AHP)[9], the weight of each attributes was defined: $W_C = 0.2$; $W_L = 0.4$; $W_D = 0.4$.

The equation of calculating fire risk within time i is as Eq. (3) and the equation of risk dissimilarity between time i and time j is as Eq. (4).

The number of clustering k defined 5 in this paper.

Table 1 shows the fire statistics and fire risk. C_i, L_i, and D_i respectively represents the normalized data of number of fire occurrence, direct economic loss, and casualties. R_i refers to the fire risk index, which was gained by Eq. (3).

Table 1 Fire statistics and fire risk during 24 hours

Time of occurrence	Number of fire	Economic losses	Casualties	C_i	L_i	D_i	R_i
0	691	3 314 824	8	0.5011	0.5630	0.4444	0.5060
1	565	4 034 410	18	0.4097	0.6852	1.0000	0.7883
2	469	2 153 404	6	0.3401	0.3657	0.3333	0.3480
3	401	2 161 266	17	0.2908	0.3671	0.9444	0.6539
4	327	5 394 205	10	0.2371	0.9162	0.5556	0.6859
5	290	1 886 356	11	0.2103	0.3204	0.6111	0.4464
6	289	3 319 378	2	0.2096	0.5638	0.1111	0.3753
7	291	2 503 516	9	0.2110	0.4252	0.5000	0.4257
8	440	2 923 837	5	0.3191	0.4966	0.2778	0.3871
9	533	2 909 961	4	0.3865	0.4942	0.2222	0.3838
10	672	3 253 422	8	0.4873	0.5526	0.4444	0.4986
11	780	3 969 178	8	0.5656	0.6741	0.4444	0.5699
12	802	4 064 368	6	0.5816	0.6903	0.3333	0.5502
13	822	4 078 158	9	0.5961	0.6926	0.5000	0.6025
14	754	3 305 304	6	0.5468	0.5614	0.3333	0.4799
15	651	3 677 014	9	0.4721	0.6245	0.5000	0.5482
16	631	3 866 831	5	0.4576	0.6567	0.2778	0.4952
17	752	5 887 877	5	0.5453	1.0000	0.2778	0.7002
18	1132	5 645 555	8	0.8209	0.9588	0.4444	0.7626
19	1249	5 532 935	6	0.9057	0.9397	0.3333	0.7495
20	1379	5 645 462	5	1.0000	0.9588	0.2778	0.7737
21	1164	4 443 056	2	0.8441	0.7546	0.1111	0.6125
22	1002	3 870 795	5	0.7266	0.6574	0.2778	0.5562
23	822	3 188 417	6	0.5961	0.5415	0.3333	0.4825

Figure 3 is the risk clustering results of 24 hours.

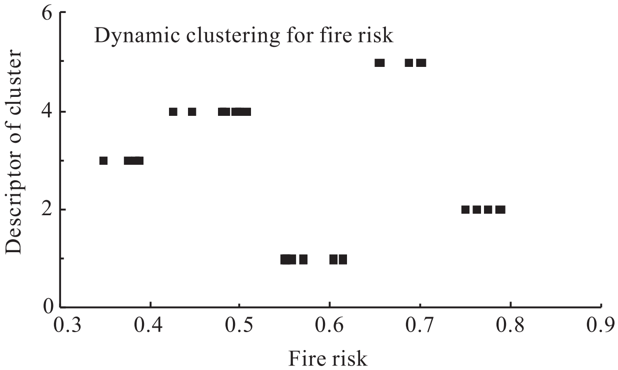


Fig. 3 Fire risk clustering during 24 hours

Table 2 shows the fire risk clustering detailed result during 24 hours. Table 3 reflects the number of observed values (fire risk during time i) in each cluster during the final clustering. Table 4 interprets the central value of each final cluster.

4 Conclusions and Future work

Urban fire statistics potentially involve time distribution law of fire risk. Taking the number of fire occurrence, direct economic loss, and casualties into account, this paper put forward urban fire risk measurement method based on fire statistics, and performed the

Table 2 Fire risk clustering detailed result during 24 hours

Time	Clustering No.	Risk value	Time	Clustering No.	Risk value
0	4	0.5060	12	1	0.5502
1	2	0.7883	13	1	0.6025
2	3	0.3480	14	4	0.4799
3	5	0.6539	15	1	0.5482
4	5	0.6859	16	4	0.4952
5	4	0.4464	17	5	0.7002
6	3	0.3753	18	2	0.7626
7	4	0.4257	19	2	0.7495
8	3	0.3871	20	2	0.7737
9	3	0.3838	21	1	0.6125
10	4	0.4986	22	1	0.5562
11	1	0.5699	23	4	0.4825

Table 3 The number of observed values in each cluster

Cluster No.	Number	Effective value	Missing value
1	6.000		
2	4.000		
3	4.000		
4	7.000		
5	3.000		
		24.000	
			0.000

Table 4 The central value of final clusters

	Cluster				
	1	2	3	4	5
Risk	0.5733	0.7685	0.3736	0.4763	0.6800

clustering analysis towards fire statistics with K-mean clustering method. Fire risks characterized by the number of fire occurrence, direct economic loss, and casualties, even though with the same (or similar) attributes might be divided into different clusters if in light of traditional clustering equation of dissimilarity among objects. Therefore, this paper established urban fire risk model based on fire statistics, furthermore, defined the equation of dissimilarity to achieve automatic classifying and discovering of risks based on fire statistics.

This paper carried out the case study of fire risk clustering within different time periods. With regard to spatial domain, fire risk clustering could be achieved so long as to partition geographical areas under certain rules and select fire statistics within different areas randomly. Similarly, it was applicable to the number of fire occurrence, direct economic loss, casualties, and other single indexes.

On fire risk, only the number of three indexes, i.e., fire occurrence, direct economic loss, and casualties, was considered in this paper, therefore certain limitation existed. As for other influencing factors, the translation method could be absolutely applied to add their influence degree towards fire risk into classifying method of urban fire risk.

Acknowledgments

The authors would like to thank Dr. BAO Hailong for his valuable suggestions and Mr. HAN Xiaopeng in Fire Department of the Ministry of Public Security of the People's Republic of China for his assistance in proving us with fire statistics.

References

[1] Wang Haihui, Fan Weicheng. Progress and problems of fire protection in China. *Fire Safety Journal*, 1997, **28**: 191-205.

[2] Yang Lizhong, Zhou Xiaodong, Deng Zhihua, et al. Fire situation and fire characteristic analysis based on fire statistics of China. *Fire Safety Journal*, 2002, **37**: 785-802.

[3] Holborn P G, Nolan P F, Golt J. An analysis of fatal unintentional dwelling fires investigated by London Fire Brigade between 1996 and 2000. *Fire Safety Journal*, 2003, **38**: 1-42.

[4] Rosenberg T. Statistics for fire prevention in Sweden. *Fire Safety Journal*, 1999, **33**: 283-294.

[5] Zheng Shuangzhong. Research on city fire risk evaluation [Dissertation]. Shenyang: Northeast University, 2003. (in Chinese)

[6] Han Jiawei, Kamber M. Fan Ming, Meng Xiaofeng translation. The Concept and Technology of Data Mine. Beijing: Machine Industry Press, 2004: 223-259.

[7] Zhu Ming. Data Discovery. Hefei: University of Science & Technology of China Press, 2002: 137-143.

[8] Li Xiongfei, Li Jun. Data Mine and Knowledge Discovery. Beijing: Higher Education Press, 2003: 93-101.

[9] Saaty T L, Vargas L G. Models, Methods, Concepts & Applications of the Analytic Hierarchy Process. Boston: Kluwer Academic Publishers, 2001: 1-13.