Visual Analysis of Multi-factor Association on Inland Waterway Accident

Cong Lu [1], Shu Gao [2]

[1]Hubei Collaborative Innovation Center for Early Warning and Emergency Response Technology, Wuhan, 430070, P.R.China;

[2] School of Computer Science&Technology，Wuhan University of Technology, Wuhan 430063，China

Keyword: Shipping accident; Parallel coordinate; Visualization; clustering; Multiple factors.

ABSTRACT

In order to explore the relationship between the multi factors of the inland river shipping accident and understand the law of the accident, the data of Yangtze River trunk line from recent years are collected and analyzed. First, we extracted the key factors of the shipping accidents and preprocessed the data; then, we studied the data clustering method, and displayed the clustering results by visually projecting them in parallel coordinates; finally, we used the human-computer interaction technology and the multi view collaborative visualization method to dig into the correlation among the factors which cause the inland shipping accidents. The results showed that, it provided a new method for maritime regulators to effectively analyze the law and cause mechanism of the inland shipping accidents, and avoid the risks of them.

1. INSTRUCTION

Water transportation is a risky industry. In recent years, with the gradual development of large-scale, specialization and high-speed inland ships, the potential safety hazards and waterway accidents are increasing, resulting in irreversible economic losses and casualties. Therefore, it is significant to analyze the accident of inland river shipping and its occurrence rules [1], to formulate relevant safety measures, preventing accidents and reducing the occurrence of similar accidents [2].

Because of the high dimension of data, it is difficult to comprehensively reveal the intrinsic relationship among the factors, and cannot reveal the hidden rules. The visual analysis method [3, 4] combines the strong perception facing with visual information and the advantage of the computer's analysis and calculation ability, which comprehensively utilizes the human-computer interaction technology and the visualization interface, to understand the information behind the data and knowledge more intuitively and efficiently.

At present, researches have been carried out on shipping accidents and visualization techniques [2, 5]. However, these methods mainly use the graphs, histograms, pie charts or rose charts to statistically analyze two or three factors. It is lack of multi-factor relevance of the visual exploration, and there is no provision of human-computer interaction, which cannot integrate of expert experience and cognitive ability in the analysis process.

In this paper, according to the characteristics of inland waterway accidents, it uses multidimensional visualization technology [6] - parallel coordinate, to study data characteristics and the relevance of the related factors of shipping accidents in multi-level, multi-angle and multi-granularity, in low-dimensional visual space, to facilitate the maritime supervisors, in order to quickly and effectively dis-cover the rules of ship accidents and excavate the internal relations of the factors of accidents.

The general idea of this paper is as follows: Firstly, the factors related to shipping accidents are extracted. Secondly, these factors are normalized on the two-dimensional plane and expressed by parallel axes of different dimensions. Thirdly, by data clustering and differentiation with different colors, it

visualizes the similarity of shipping accidents in low-dimensional space. Finally, it reveals the inherent correlation between the factors of inland river shipping accidents.

2. MULTI-FACTOR DATA PREPROCESSING OF INLAND WATERWAY ACCIDENT

2.1 Multi-Factor Extraction

The risk factors of ship accidents are usually extracted from the four aspects[7, 8] of man-ship-loop-pipe. Based on the former work, this paper filters out some unimportant or difficult to obtain the value of the ship accident. The 14 typical factors are the crew experience, crew seaworthiness, ship size, ship speed, age, visibility, water level, wind, traffic volume, VTS supervision degree, in and out port management (IOM), the time of the accident, the location of the accident, the type of accident and the cause of the accident, as shown in Figure 1.

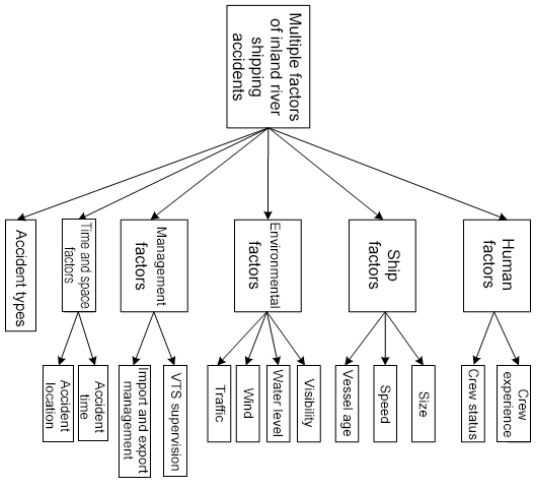

Figure 1. Extraction of typical factors of shipping accident

After data collection and cleaning, it respectively deals with them according to data types:

1) The directly quantified factors are illustrated by numerical representation, stored after corresponding data conversion, such as the location of the accident, the Yangtze River waterway (from Yunnan to the Yangtze River water rich to the sea, a total length of 2838km). In order to facilitate the visualization, the location of the accident is mapped to the range of 0 ~ 2838, with the upper range of 0 ~ 1074, the middle interval of 1075 ~ 1975, and the lower interval of 1976.

2) It quantitatively evaluates qualitative factors and stores them by discretization, such as collision, stranding, rock-struck, touching, wave damage, fire / explosion, wind disaster, self-sink, operational pollution accidents and other accidents according to the "Water Traffic Accident Statistical Measures" [9]. It can discretize different types of incidentsinto different values, establishthe relevant mapping tableand store it.

2.2 Multivariate Data Standardization

Because ofthe datadifferences of various factors, it is hard to analyze the multi-dimensional data. So, it is necessary to standardize the extracted data. Let the source data matrix be:

$$\text{factors} = \begin{bmatrix} f_{11} & f_{12} \Lambda & f_{1m} \\ f_{21} & f_{22} \Lambda & f_{2m} \\ \text{M} & \text{M} & \text{M} \\ f_{n1} & f_{n2} \Lambda & f_{nm} \end{bmatrix}$$

The matrix rownis the number of shipping accident samples; and thematrix column mis the data dimension, that is, the number of correlation factors. The elements in the source data matrix are normalized using equation (1).

$$a_{i,j} = L * \frac{f_{i,j} - \min(j)}{\max(j) - \min(j)} \qquad (1)$$

Where L is the value of the coordinate axis, j is the minimum value of the matrix j, j is the maximum value of the matrix, after the conversion after the value.

3. CLUSTERING ANALYSIS OF SHIPPING ACCIDENTS BASED ON PARALLEL COORDINATES

3.1 Parallel Coordinate

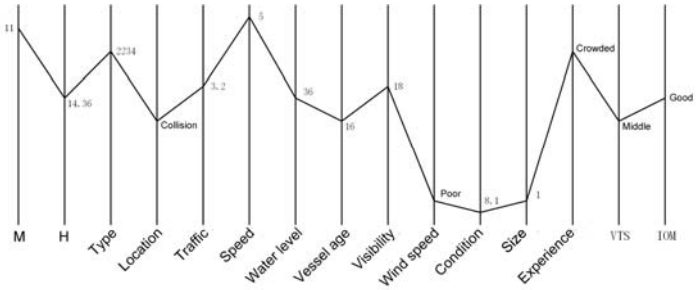

Figure 2. A 15-dimensional parallel coordinate plot of a shipping accident data

The parallel coordinate [10] is a m-dimensional data attribute represented by m parallel axes on a two-dimensional plane. Each axis represents an attribute dimension, and the value range on the axis corresponds to the minimum and maximum values of the corresponding attribute. Therefore, each m-dimensional data item can obtain m points on m parallel coordinate axes according to the value of each attribute, and connect these m points in turn to form a fold line. So every m-dimensional data item can be represented in a m-dimensional parallel coordinate system by a polyline, and similar data items have a similar tendency toward a polyline. Thus, the data relationship between each attribute will become very intuitive, convenient for people to observe the data, understanding and understanding. Figure 2 shows a parallel plot of shipping accident data, where the time of occurrence of the accident has two axes, one that the accident occurred in the month, and the second said the time of the accident the day.

3.2 Clustering Analysis For Shipping Accidents

In order to avoid the visual chaos caused by a large amount of data on the parallel coordinate graph [11], this paper studies the data clustering method, merges the similar accident data items into different clusters, and uses different colors to distinguish clusters. Thus reducing the confusion caused by folding line stacking phenomenon, while strengthening the user's understanding of the results of clustering.

3.2.1 CALCULATION METHOD OF ACCIDENT SIMILARITY DISTANCE

Generally speaking, the occurrence of inland river shipping accidents has certain similarity. In order to describe the similarity better, the rule of shipping accident can be used to cluster the shipping accident data. An m-dimensional vector is used to represent a shipping accident in which each dimension of the vector corresponds to a factor, so the vector defining an accident is:*Accident = (a_1, a_2, …, a_m).*

The similarity distance between the accident s and the accident t can be calculated by the formula (2).

3.2.2 THE IMPROVED K-MEANS ALGORITHM

K-Means clustering algorithm [12] is a classical algorithm based on partitioning clustering problem, but it has two problems, one is the number of clusters K is not easy to be determined, the other is the initial K clusters The center is not easy to choose (different initial clustering centers will produce different clustering results). For the above two problems, the use of algorithm 1 improved.

Algorithm 1: Clustering algorithm for shipping accident analysis

Input: Normalized ship accident data matrix.

Output: K clustering results.

Step1: Use the hierarchical clustering algorithm to determine the number of clusters clustered K;

Step2: Using the data-based distribution density, select K data points farthest from each other as the initial cluster center (centroid);

Step3: Calculate the similarity distance of all data points to K cluster centers and classify them into the cluster of the nearest cluster center.

Step4: Recalculate the clustering centers of K classes.

Step5: Step3 and step4 iteration until the cluster center no longer changes, the cluster results output

In the visualization of parallel coordinate, if the line graph is presented, there will be such problems as line density and repeated stacking, which leads to the difficulty of analysis. This paper uses the parallel coordinate curve [13] as the visualization. The results of clustering on parallel coordinates are shown in Fig. Users can visually distinguish the clustering results from different colors. It can be seen that there are four clusters of different colors in the graphs. For different clusters, the rules of accident occurrence can be analyzed. For example, most green clusters occur at midnight after the dry season and stranded and reefed. Traffic is usually not large, mostly in the water level below 6 meters.

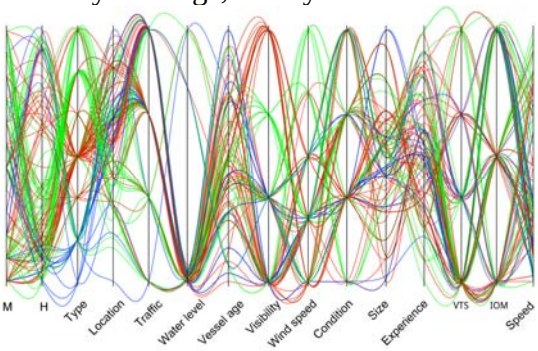

Figure 3. Clustering display on parallel coordinates

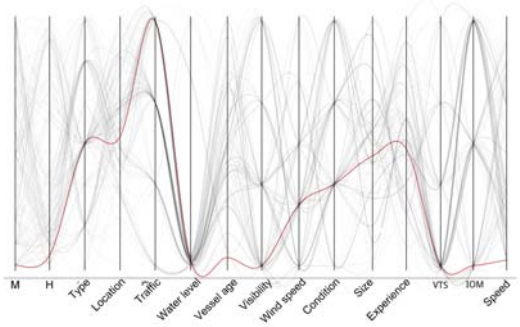

Figure 4. Application of brush technology in parallel coordinates

4. VISUAL ANALYSIS OF INLAND SHIPPING ACCIDENT BASED ON INTERACTIVE TECHNOLOGY

4.1 Brush Technology

The brush technique can highlight a subset of the data of interest to the user. In parallel coordinates, through the brush technique can highlight the selected curve highlighted, not to be selected for the dilution of the curve. Through the brush technology, you can from many accidents, the visual observation of a shipping accident in the relationship between the various factors, as shown in Figure 4.

4.2 TechnologyOf Axis Exchange

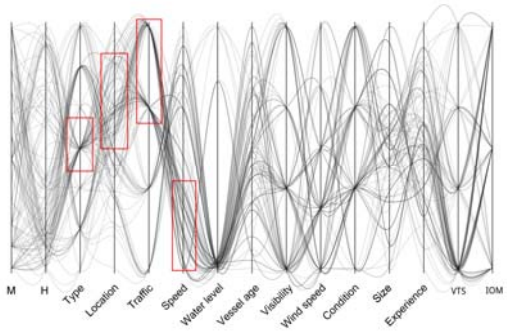

Figure 5. Application of axial exchange technology

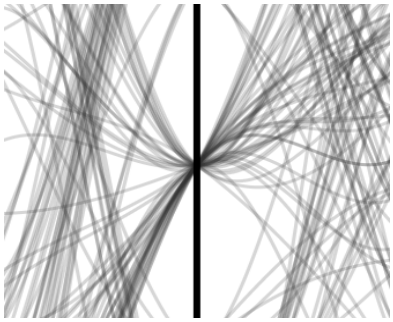

Figure 6. Applications of dimensional zoom technology

The exchange of axes [14] can better illustrate the relationship between the factors. By expert experience, the more closely related axis exchange to the adjacent position, from the curve of the direction, you can visually observe the relationship between adjacent factors.

For example, the ship accident type axis, the accident location axis, the traffic volume axis and the ship speed axis exchange to the adjacent location, as shown in Figure 5 can be observed in the red box at a large concentration, can find the collision accident Mostly in the middle and lower reaches of the Yangtze River traffic flow channel, at the same time, can be found in traffic volume and speed of the ship is inversely proportional to the relationship between the greater the traffic volume, the lower the ship speed.

4.3 Technology Of Dimensional Control And Enlargement

Dimensional control technology can reduce the display complexity of parallel coordinates, only show the user interested in the dimension of factors of shipping accident factors, so as to better observe and analyze the data. If the user needs to parse a certain part of a dimension in parallel coordinates, the local coordinate axes can be displayed globally through the dimension-enlargement technique, and the relationship between the data nodes can be analyzed better, as shown in Figure 6.

5. CONCLUSION

In this paper, the technique of parallel coordinate is applied to multivariate analysis of shipping accident. The multidimensional data clustering method and man-machine interaction method on parallel coordinate are studied. The multi-view collaborative visualization method is used to analyze shipping accident. The research results provide a new method for the maritime supervision department to analyze the rule and cause mechanism of inland river shipping accident and avoid the risk.

6 REFERENCES

1. Zhang Jinfen,YanXinping,ChenXianqiao, et al. 2012. Analysis of traffic accidents on the Yangtze River trunk line based on probability distribution [J]. Navigation of China 35(3): 81-84.
2. DENG Yibin, CHU Guanquan, CHEN Jiayuan, et al. 2015. Statistical analysis and safety countermeasures on ferry accidents in inland river [J]. Navigation of China 38(2):74-78.
3. XU Wuxiong, CHU Xiumin, LIU Xinglong. 2015. Advances of maritime traffic information visualization techniques [J]. Navigation of China 38(1): 34-38.
4. REN Lei, DU Yi, MA Shuai, et al. 2014 Visual analysis towards big data [J]. Journal of Software (9): 1909-1936.
5. WU Nai-ping. 2010. Fully understanding some characteristics of inland river ship collision accidents [J]. Navigation of China 33(4): 79-84.
6. GintautasDzemyda. 2015. Multidimensional data visualization - methods and applications[J]. Springer Ebooks51(2):121-124.
7. Zhang D, Yan X, Yang Z, et al. 2011. Application of Formal Safety Assessment to Navigational Risk Evaluation of Yangtze River[C]. ASME 2011, International Conference on Ocean, Offshore and Arctic Engineering, June 19th - 24th 2011. Rotterdam:847-854.
8. CHEN Shu - zhe. 2012. Study on response mechanism of inland river maritime accidents risk catastrophe to seafarer-ship-environment disadjust [D]. Doctoral Dissertation of Wuhan University of Technology.
9. Ministry of Transport of the People 's Republic of China. 2015. Statistical methods for maritime traffic accidents [J]. Gazette of the State Council of the People’s Republic of China, (1): 44-47.
10. Heinrich J, Broeksema B. 2015. Big Data Visual Analytics with Parallel Coordinates[C].Big Data Visual Analytics. IEEE.
11. Peng W, Ward M O, Rundensteiner E A. Clutter Reduction in Multi-Dimensional Data Visualization Using Dimension Reordering[C]. 10th Annual IEEE Symposium on Information Visualization (InfoVis 2004). OCT 10-12, 2004, Austin, TX:89-96.
12. Patel V R, Mehta R G. Modified k-Means Clustering Algo-rithm[M]. Computational Intelligence and Information Technology. Springer Berlin Heidelberg, 2011:307-312.
13. Graham M, Kennedy J. Using Curves to Enhance Parallel Coordinate Visualisations[C].International Conference on Information Visualization, 2003. IV 2003. Proceedings. 2003:10.
14. Lu L F, Huang M L, Zhang J. 2016. Two Axes Re-ordering Methods in Parallel Coordinates Plots[J]. Journal of Visual Languages & Computing32(1):3-12.
15. CHEN Yi,CAI Jin-feng,SHI Yao-bin, et al. 2013. Coordinated visual analytics method based on multiple views with parallel coordinates [J]. Journal of System Simulation 25(1): 81-86.