

A New Approach for Hierarchical Dividing to Passenger Nodes in Passenger Dedicated Line

Chanchan Zhao^{***}, Feng Liu^{*}, and Xiaowei Hai^{***}

Abstract

China possesses a passenger dedicated line system of large scale, passenger flow intensity with uneven distribution, and passenger nodes with complicated relations. Consequently, the significance of passenger nodes shall be considered and the dissimilarity of passenger nodes shall be analyzed in compiling passenger train operation and conducting transportation allocation. For this purpose, the passenger nodes need to be hierarchically divided. Targeting at problems such as hierarchical dividing process vulnerable to subjective factors and local optimum in the current research, we propose a clustering approach based on self-organizing map (SOM) and k-means, and then, harnessing the new approach, hierarchical dividing of passenger dedicated line passenger nodes is effectuated. Specifically, objective passenger nodes parameters are selected and SOM is used to give a preliminary passenger nodes clustering firstly; secondly, Davies–Bouldin index is used to determine the number of clusters of the passenger nodes; and thirdly, k-means is used to conduct accurate clustering, thus getting the hierarchical dividing of passenger nodes. Through example analysis, the feasibility and rationality of the algorithm was proved.

Keywords

Hierarchical Dividing, K-Means, Passenger Nodes, Passenger Dedicated line, Self-Organizing Map

1. Introduction

The first real passenger dedicated line of China is Beijing-Tianjin Intercity Railway launched on August 1, 2008. Recent years saw a rapid development of passenger dedicated lines in China; by the end of 2016, China is operating over 22,000 km passenger dedicated lines, topping in the world. The *Middle and Long Term Railway Network Plan* [1] has pointed out that China would augment the scale to 38,000 km as of 2025. Under such a context, an even more extensive passenger nodes network would shape, whose natural and social properties would exert huge impact to passengers since these nodes are closely linked to arriving, departing and transferring passengers and relative railway technologies.

Furthermore, most stations in passenger dedicated line are well developed cities or hubs with huge passenger flow, so a reasonable hierarchical dividing of passenger nodes is of great practical significance to the transportation productivity layout and passenger traffic allocation.

In common hierarchical dividing approaches, the mostly chosen parameters are average daily

※ This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Manuscript received February 22, 2017; first revision June 13, 2017; accepted June 21, 2017.

Corresponding Author: Chanchan Zhao (cczhao@imut.edu.cn)

^{*} School of Computer and Information Technology, Beijing Jiaotong University, Beijing, China (cczhao@imut.edu.cn, fliu@bjtu.edu.cn)

^{**} College of Information Engineering, Inner Mongolia University of Technology, Hohhot, China (cczhao@imut.edu.cn)

^{***} School of Economics and Management, Inner Mongolia University of Technology, Hohhot, China (haixiaowei2006@126.com)

passenger volume, number of average daily originating trains, population, geographical location, accessibility, gross regional domestic product, city administrative level, etc., among which some are subjective and easy to be influenced by subjective factors during data-measurement. In addition, when using k-means algorithm for clustering, the local optimum is easily caused because only the objective function is calculated. What's more, manual judgment is required to determine the number of clusters and thus individual experience has a great effect on the clustering result.

Based on analysis and summary over existing hierarchical dividing approaches and targeted on their defectives, we propose a new approach on the self-organizing map (SOM) and k-means basis to complete hierarchical dividing of passenger nodes in passenger dedicated line. Firstly, objective passenger nodes parameters are selected and SOM is used to give a preliminary passenger nodes clustering and to obtain the correspondent best matching unit (BMU), preventing the effect of individual subjective factors; secondly, the Davies–Bouldin index (DBI) is used to determine the k value, number of clusters of the passenger nodes, evading the local optimum due to only calculating objective function in the traditional k-means algorithm; and thirdly, BMU and k value are inputted into k-means algorithm to conduct accurate clustering, thus getting the hierarchical dividing of passenger nodes in passenger dedicated line.

The rest of this paper is organized as follows. Section 2 summarizes related work. Section 3 is the review of the SOM and k-means algorithm, followed by the proposed approach in Section 4. Section 5 shows the simulation experimental results, and Section 6 concludes the paper with summary and future research direction.

2. Related Work

At present, there is no uniform method for the hierarchical dividing of railway passenger nodes. In the actual scene, the division of the railway passenger nodes is mainly according to the number of the passenger technical operations handled by the passenger station and qualitatively refers to the politics, economy, culture, and transportation layout in the place where the station is located in. Then the nodes are divided into the top-class station, first-class station, second-class station, third-class station, fourth-class station, and fifth-class station.

For the study of the hierarchical dividing of railway passenger nodes, based on the establishment of importance degree evaluation system of passenger nodes, some scholars used rough set, analytic hierarchy process and principal component analysis to calculate the importance degree of each passenger node as well as set the threshold to classify the passenger nodes. This method mainly relies on experience. And its division has strong subjective intention and low efficiency [2].

The above problem can be solved well by clustering analysis. Wang and Lyu [2] adopted clustering method to classify railway passenger transport nodes. Firstly, it clustered property variables by hierarchical clustering, and clustered passenger transport nodes samples by affinity propagation algorithms according to the simplified nodes indexes. Finally, it analyzed three clustering effectiveness indexes contained CH, KL and IGP indexes to the clustering consequence. Gao et al. [3] do a k-means clustering analysis of key nodes and edges in Beijing subway network. Based on the k-means clustering, stations (nodes) and intervals (edges) in a subway network were grouped by three metrics: two basic

topological properties (degree and betweenness), and their roles in transporting people (passenger volume). Xu and Qin [4] build important evaluation indexes for urban rail transit network. Besides degrees and betweenness in complex network, they add PageRank value and passenger flow indicators in consideration of the urban rail transit network topology model to better evaluate relationships between stations in urban rail transit network and offer theoretical guarantee for risks. They provide factor analysis method to calculate the importance degree of each station in the networks, which is illustrated by the case of Beijing Railway. Zhou et al. [5] analyzes the significance and methods of high-speed passenger railway nodes classification and designs high-speed rail train line stops program based on the classification. Finally, they take Beijing-Guangzhou high-speed railway as an example to make an empirical study. Park and Lee [6] study classification of subway stations and trip behavior of subway passengers through partitioning the graph of the subway system into roughly equal groups. They propose a heuristic algorithm to partition the subway graph. In the experimental results, they illustrate the subway stations and edges in each group by color on a map and analyze the trip behavior of subway passengers by the group origin-destination matrix.

The above studies indicates that, to a certain extent, some thoughts of hierarchical dividing of passenger nodes are worth learning and reference, yet some approaches are unable to reflect objective conditions since they select some subjective parameters and are highly influenced by human factors during data-measurement. In addition, manual judgment is required to determine the number of clusters in current studies and thus individual experience has a great effect on the clustering result. Therefore, in this paper objective property data is used as the basis for hierarchical dividing of passenger nodes in passenger line. DBI is used to determine the number of clusters of the passenger nodes. During the clustering, moreover, local optimum resulted from calculating only the objective function in traditional k-means algorithm will be prevented.

3. Review of the Self-Organizing Map and K-Means Algorithm

3.1 Self-Organizing Map

SOM refers to a guideless clustering algorithm brought by Kohonen [7], a neural expert from University of Helsinki, Finland. SOM simulates human brain where neural cells have different “division of labor” in different areas, i.e., response characteristics are different in different brain zones, a process completed automatically. SOM divides input mode sets via seeking the optimum reference vector sets.

SOM consists of two layers: an input and a radial layer. The input layer is used to receive input data, and there is no link among internal nodes; number of nodes in radial layer corresponds to the number of mode space dimensions after mapping; each and every nerve cell in the radial layer links to and mutually motivates its neighborhood. After training, different nodes in radial layer represent different dividing modes. Fig. 1 outlines the general structure of SOM.

SOM has been widely used in network detection [8], environmental monitoring [9-11], maritime research [12-14], and so on.

The SOM algorithm proposed by Kohonen [7] is shown as Algorithm 1.

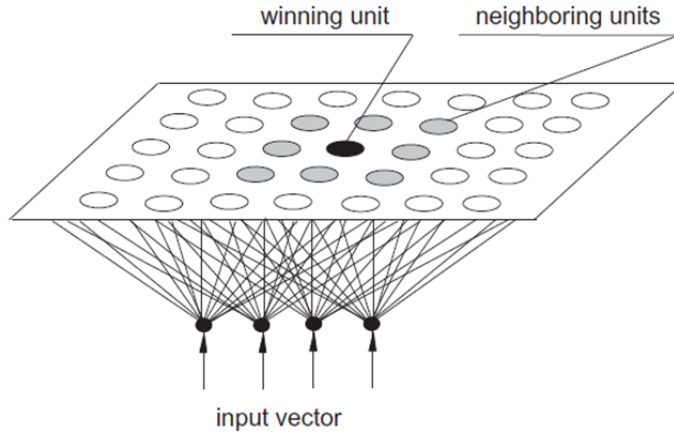


Fig. 1. Schematic structure of SOM.

Algorithm 1: Self-Organizing Map (SOM)

Input: sample vectors, iteration number N

Output: BMU

Workflow:

Step 1 Initialize sample vectors and weight vector $\omega_j(0)$, $n = 0$

Step 2 Extract the connection vector $x(n)$ from the sample space with a certain probability

Step 3 Calculate the Euclidean distance between $x(n)$ and each neuron to obtain the winning neuron $i(x(n))$

Step 4 Modify the weights of neurons

Step 5 $n++$, Repeat steps 2–4 until stopping criterion ($n > N$) is met.

In step 3, the winning neuron is calculated according to Eq. (1):

$$i(x(n)) = \arg \min_j \|x(n) - \omega_j(n)\|, j = 1, 2, \dots, l \quad (1)$$

In step 4, the modified weights of neurons are calculated according Eqs. (2)–(4):

$$\omega_j(n+1) = \omega_j(n) + \eta(n) h_{j,i(x)}(n) (x(n) - \omega_j(n)) \quad (2)$$

$$h_{j,i(x)}(n) = \exp\left(-\frac{d_{j,i}^2}{2\sigma^2(n)}\right) \quad (3)$$

$$d_{j,i}^2 = \|r_j - r_i\|^2 \quad (4)$$

where $\eta(n)$ is learning rate; $h_{j,i(x)}(n)$ is neighborhood function, which follows the Gauss probability distribution; r_j is the spatial position of excitatory neurons j ; $\sigma(n)$ is the width of the topological neighborhood function.

3.2 K-means Algorithm

The k-means algorithm [15] is a classic clustering method that divides the data set in k classes of similar data on the basis of the Euclidean distance criterion. The process starts with k vectors, randomly selected from the data set and used as temporary cluster centroids. Afterwards, the algorithm calculates the distances between the centroids and all the vectors of the data set and associates each vector to its nearest centroid. After all data have been assigned, the new cluster centroids are calculated according to Eq. (5):

$$c_i = \frac{1}{m_i} \sum_{j=1}^{m_i} x_{ij} \quad (5)$$

where c_i is the centroid of cluster C_i ; m_i is the number of x_j data gathered in cluster C_i .

The process iterates until there is no further change in the centroids' position. The k-means algorithm proposed by MacQueen [15] is shown as Algorithm 2.

Algorithm 2: K-means Algorithm

Input: data set, clustering number k

Output: clustering results

Workflow:

Step 1 Randomly select k cluster centers

Step 2 Calculate the distance from each object to each cluster center and assign the object to the nearest cluster

Step 3 Recalculate the k cluster centers

Step 4 Repeat steps 2-3 until criterion function converged

4. The Proposed Approach

In this paper we adopt a two-level approach, based on the combined use of SOM and k-means clustering algorithm. Specifically, SOM is used to give a preliminary passenger nodes clustering and to obtain the corresponding BMU firstly, preventing the effect of individual subjective factors; secondly, DBI index is used to determine the k value, number of clusters of the passenger nodes, evading the local optimum due to only calculating objective function in the traditional k-means algorithm; and thirdly, BMU and k value are inputted into k-means algorithm to conduct accurate clustering. This proposed approach will be used to classify the passenger nodes later. Its workflow is shown as follows:

Step 1. Data preprocessing. Preprocessing methods include log-transformation [16,17], standard normalization [18,19], and so on. In this paper, we adopt histogram equalization method.

Step 2. Data conversion. To convert the preprocessed data into a two-dimensional vector for SOM networks, as illustrated in Eq. (6).

$$M = \begin{bmatrix} x_{11} & x_{11} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots \\ x_{m1} & \dots & \dots & x_{mn} \end{bmatrix} \quad (6)$$

Step 3. SOM training. The Gaussian function batch training mode is used to train the SOM neural network to get the mapping neurons.

Step 4. Clustering number determination. DBI [20] is used to determine the number of clusters. The number of clusters corresponding to the minimum value of DBI is the final clustering number. DBI value is calculated according to Eq. (7):

$$DBI_{optimum} = \arg \min_C \left\{ \frac{1}{C} \sum_{k=1}^{k=C} \max_{l \neq k} \left\{ \frac{S(k) + S(l)}{D(k, l)} \right\} \right\} \quad (7)$$

where C is the number of clusters; k is the cluster k ; $S(k)$ is the average distance of all the elements of the cluster to their cluster center; $D(k, l)$ is the distance between clusters centers.

Step 5. Clustering. The k value of k-means algorithm is determined by DBI value, and k-means algorithm is used to classify the weights of neurons.

The procedure is shown in Fig. 2.

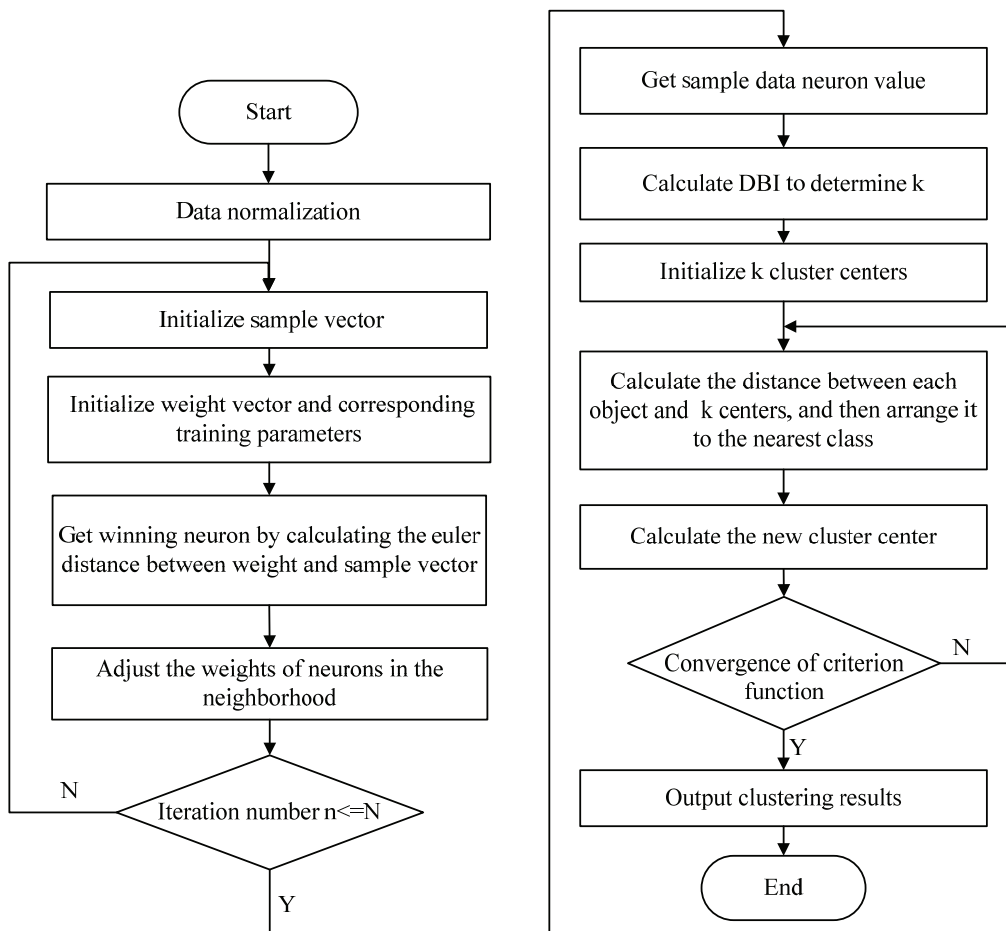


Fig. 2. Flowchart of the proposed approach.

5. Performance Evaluation

5.1 Simulation Setup

Four parameters are selected as key influence factors for hierarchical dividing of passenger nodes. They are average daily passenger volume (PassengerF), average daily originating train number (TrainN), population quantity (Population), and gross regional product (GRP). To conduct the simulation, we first normalize each parameter. The comparison before and after normalization of each parameter is shown in Figs. 3–6.

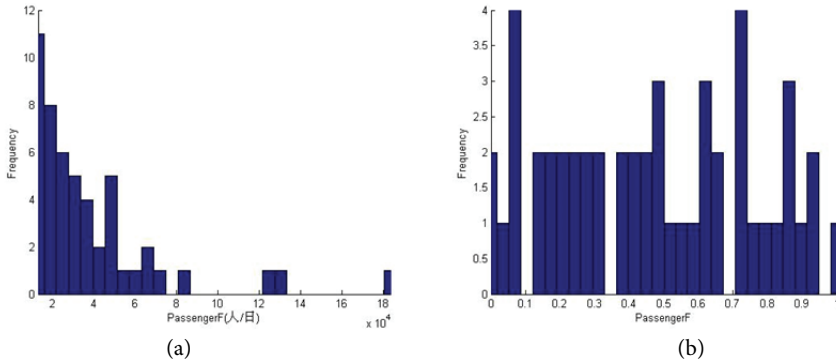


Fig. 3. PassengerF histogram. (a) Before normalization and (b) after normalization.

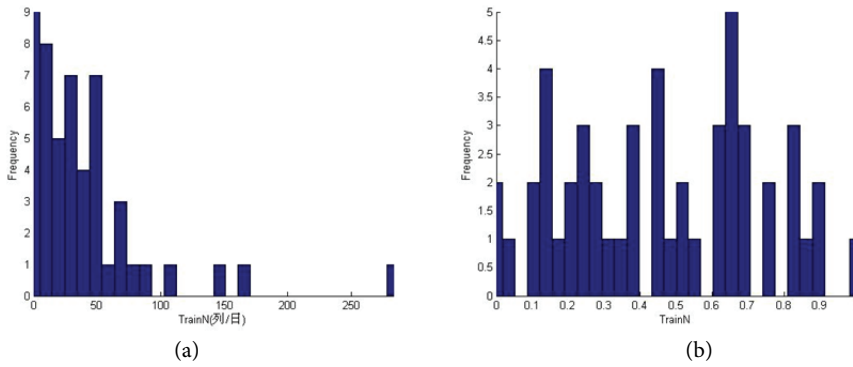


Fig. 4. TrainN histogram. (a) Before normalization and (b) after normalization.

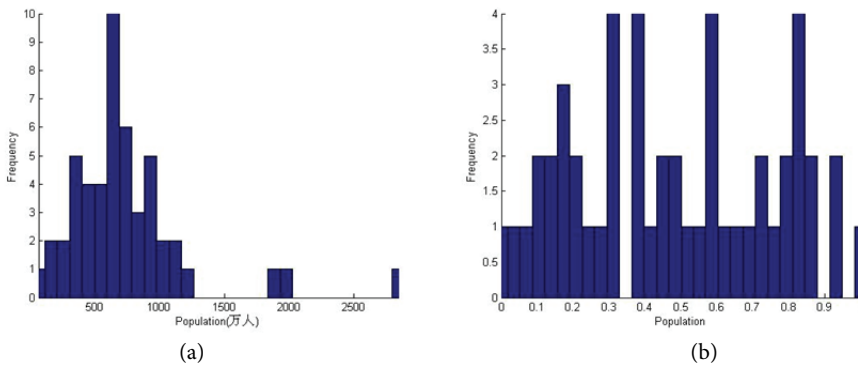


Fig. 5. Population histogram. (a) Before normalization and (b) after normalization.

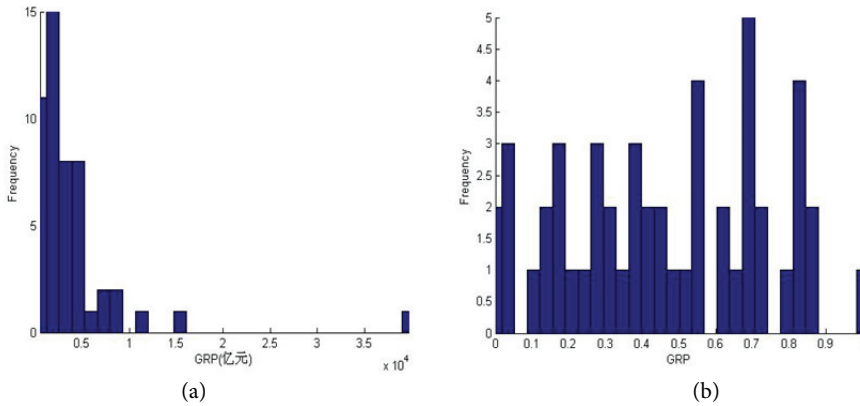


Fig. 6. Gross regional product (GRP). (a) Before normalization and (b) after normalization.

The size of the SOM is determined as $[5 \times 8]$, and then the map will be trained with the input variables to self-organize the 50 input vectors.

Number of clusters is determined automatically by DBI value, as shown in Fig. 7. It's 4 which is corresponding to the minimum DBI value.

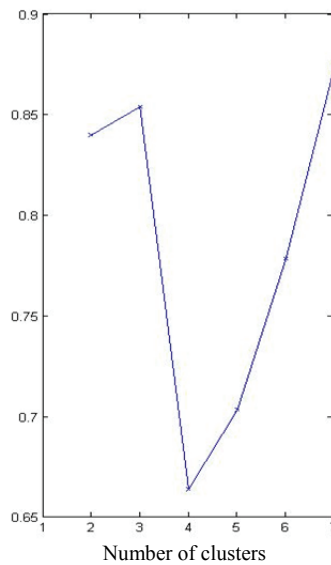


Fig. 7. DBI value.

5.2 Results and Analysis

The component planes of Fig. 8 can be used for correlation hunting between parameter [21]. The color shades of SOM neurons for PassengerF and TrainN exhibit a similar increase from the upper left part to the lower right, which reveal that average daily passenger volume and average daily originating train number have a strong correlation. Similarly, TrainN and Population are also positively correlated; however, no clear correlation with any other parameter is emergent for GRP.

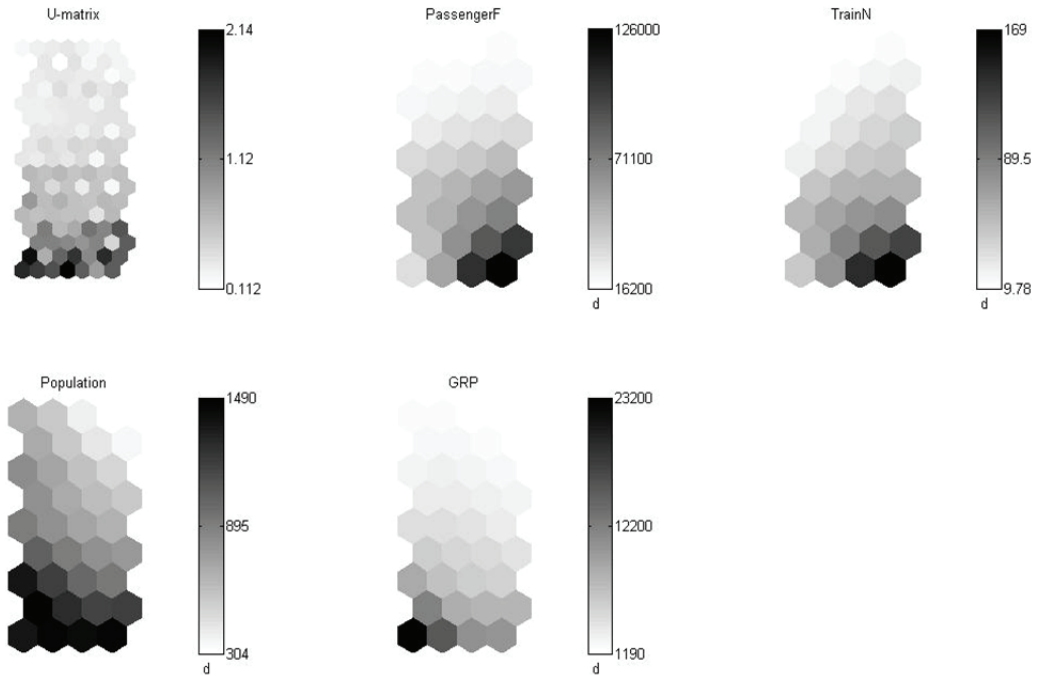


Fig. 8. U-matrix and component plane of sample data.

The result of simulation experiment show that the passenger nodes in passenger dedicated line are divided into four grades. Clustering result is shown in Table 1; clustering result analysis is shown in Table 2; and average data of each cluster is shown in Table 3.

The following conclusions can be drawn by Tables 1–3:

- The mean values of four parameters of Level-1 passenger nodes are all higher than those of the rest three levels. Passenger nodes in this level reflect huge passenger volume, great originating flow, large population and highly-developed economy. Passenger nodes of Level-1 should be critical to the whole railway network, serving as the nerve center of the entire railway system and linking multiple important passenger lines.
- The mean values of four parameters of Level-2 passenger nodes are all lower than those of Level-1 passenger nodes, but higher than the rest two levels. These passenger nodes reflect relatively big passenger volume, huge departing flow, large population and developed economy.
- The mean values of four parameters of Level-3 passenger nodes are all lower than those of Level-1 and Level-2 passenger nodes. Level-3 passenger nodes are higher than Level-4 passenger nodes with respect to the mean values of daily passenger volume, daily originating train quantity and gross regional domestic product, but lower in terms of mean population. These passenger nodes are located in cities of medium scale population and economy.
- As to passenger nodes of Level-4, they are only higher than Level-3 passenger nodes but still lower than Level-1 and Level-2 in terms of mean population, and lower than those of the other three levels in mean values of all the other three parameters. These passenger nodes are generally less prominent with weak attraction and passenger distribution capacity.

Table 1. Clustering result

Passenger node	Cluster	Passenger node	Cluster
S1	1	S26	3
S2	1	S27	3
S3	1	S28	3
S4	1	S29	3
S5	2	S30	3
S6	2	S31	3
S7	2	S32	3
S8	1	S33	3
S9	1	S34	4
S10	2	S35	4
S11	1	S36	4
S12	1	S37	4
S13	2	S38	4
S14	2	S39	4
S15	2	S40	4
S16	2	S41	4
S17	2	S42	4
S18	2	S43	4
S19	2	S44	4
S20	3	S45	4
S21	3	S46	4
S22	3	S47	4
S23	3	S48	4
S24	3	S49	4
S25	3	S50	4

Table 2. Clustering result analysis

Cluster	Passenger node	Sample size	Proportion (%)	Cumulative proportion (%)
1	S1,S2,S3,S4,S8,S9,S11,S12	8	16	16
2	S5,S6,S7,S10,S13,S14,S15,S16,S17,S18,S19	11	22	38
3	S20,S21,S22,S23,S24,S25,S26,S27,S28, S29,S30,S31,S32,S33	14	28	66
4	S34,S35,S36,S37,S38,S39,S40,S41,S42, S43,S44,S45,S46,S47,S48,S49,S50	17	34	100

Table 3. Average data of each cluster

Cluster	Average daily passenger volume (people/day)	Average daily originating train number (trains/day)	Population quantity (ten thousand people)	Gross regional product (billion yuan)
1	91,894	121	1,156	8,207
2	49,329	43	999	3,984
3	26,133	25	531	2,221
4	14,894	12	575	1,742

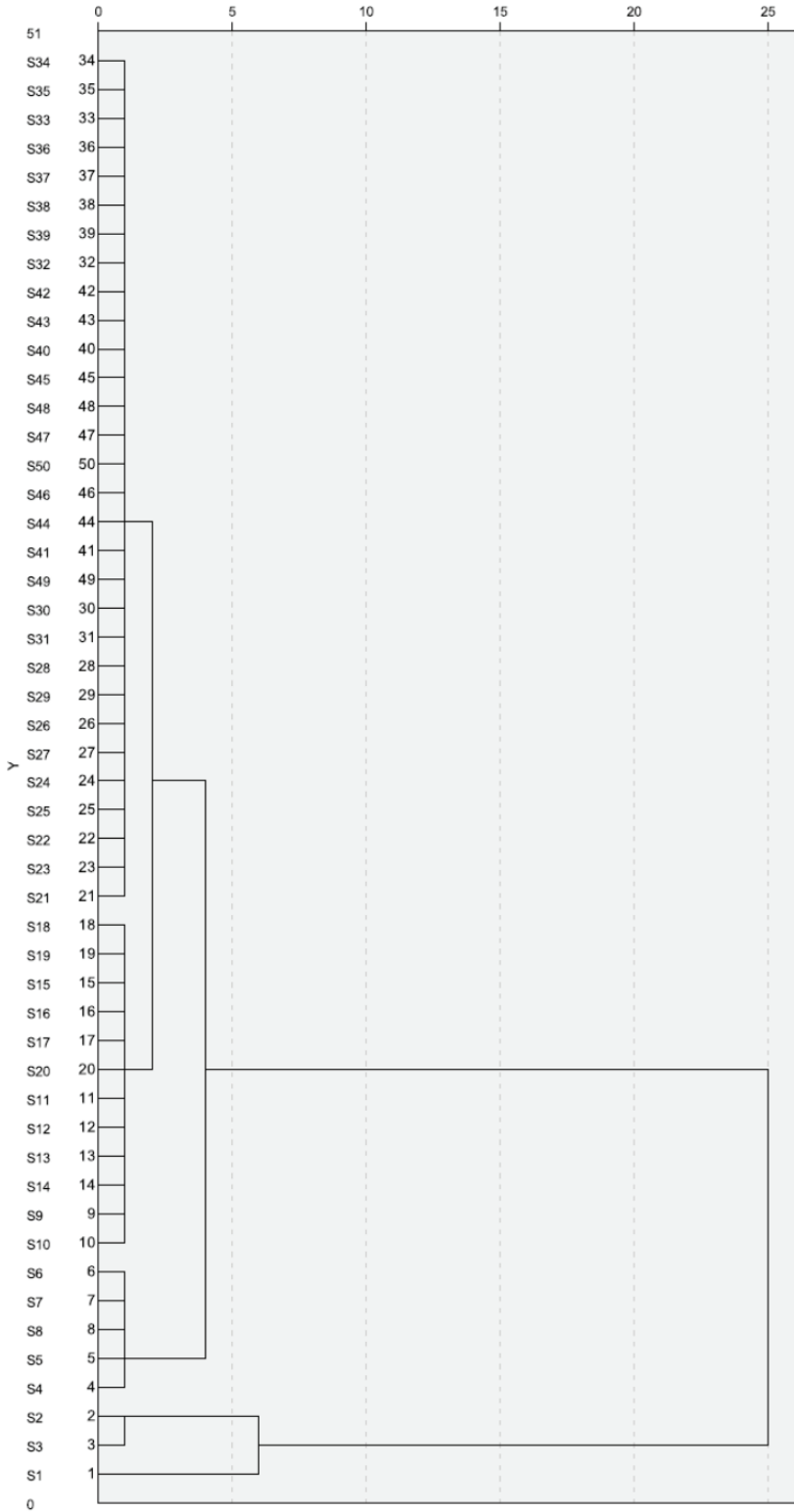


Fig. 9. System clustering result.

5.3 Comparative Performance Analysis

In order to compare our proposed clustering approach with other clustering methods, we apply the same data to the following different algorithms: system clustering and traditional k-means algorithm. Through the comparative performance analysis, the effectiveness and rationality of our newly proposed clustering approach is proved.

5.3.1 Compared with system clustering

We use the system clustering method to cluster 50 passenger nodes, and the result is shown in Fig. 9.

It can be seen from Fig. 9 that the sample data can be divided into three or four categories. If they are divided into three categories, the first category only includes S1, the second category includes S2 and S3 as well as the rest are the third category. If they are divided into four categories, the first category only includes S1, the second category includes S2 and S3, the third category includes S4, S5, S6, S7, and S8 as well as the rest are the fourth category.

If the method is used for clustering, it is necessary to manually determine the final number of clusters, which makes the personal experience to a certain extent have a significant impact on the results. Compared with this method, the DBI in the clustering method proposed in this paper can directly work out the number of optimal clusters is 4. Furthermore, the algorithm of system clustering is based on regression analysis that is essentially linear correlation analysis. So some errors are inevitable. The clustering method of this paper is based on the classification of the whole sample data, which can reflect the essence of the problem more truly. In conclusion, our proposed approach is superior to system clustering.

5.3.2 Compared with traditional k-means algorithm

We use the traditional k-means algorithm to cluster 50 passenger nodes, and the results are shown in Tables 4 and 5.

Table 4. Number of cases in each cluster

Cluster 1	1.000
Cluster 2	2.000
Cluster 3	31.000
Cluster 4	16.000
Effective	50.000
Defective	0.000

From the clustering results in Tables 4 and 5, it can be seen that the clustering effect is not very satisfactory. This clustering method brings 50 passenger nodes into two categories: the third and the fourth. The data in these two categories account for 94% of the total. In addition, the clustering number in the traditional k-means is entered manually. And the clustering number in this paper is determined by the DBI, which avoids the fact that the k-means algorithm only calculates the objective function to lead to a local optimal situation. In conclusion, our proposed approach is superior to traditional k-means algorithm.

Table 5. Traditional k-means clustering result

Case number	Passenger node	Cluster	Distance	Case number	Passenger node	Cluster	Distance
1	S1	1	0.000	26	S26	3	5579.675
2	S2	2	3206.297	27	S27	3	5680.704
3	S3	2	3206.297	28	S28	3	5110.310
4	S4	4	33440.595	29	S29	3	4133.863
5	S5	4	18448.326	30	S30	3	3321.081
6	S6	4	14069.658	31	S31	3	2011.108
7	S7	4	13404.683	32	S32	3	1092.909
8	S8	4	10096.357	33	S33	3	1947.974
9	S9	4	3416.392	34	S34	3	2372.474
10	S10	4	1567.185	35	S35	3	2540.183
11	S11	4	6102.071	36	S36	3	3064.050
12	S12	4	6461.678	37	S37	3	3659.759
13	S13	4	6187.772	38	S38	3	3845.323
14	S14	4	6206.689	39	S39	3	3827.404
15	S15	4	11113.626	40	S40	3	4971.550
16	S16	4	12785.420	41	S41	3	5220.099
17	S17	4	14271.046	42	S42	3	5355.394
18	S18	4	15410.351	43	S43	3	5683.571
19	S19	4	15518.065	44	S44	3	6209.003
20	S20	3	15367.264	45	S45	3	6360.139
21	S21	3	11645.563	46	S46	3	6566.281
22	S22	3	10591.104	47	S47	3	6440.974
23	S23	3	9172.625	48	S48	3	6525.897
24	S24	3	8192.050	49	S49	3	7467.720
25	S25	3	8117.674	50	S50	3	7100.199

6. Conclusion

A train operation plan based on a reasonable hierarchical dividing of passenger nodes could help to satisfy passenger flow better, and enlarge the competitive edge of passenger dedicated line market. In this work, we present a new approach for the hierarchical dividing of passenger nodes based on SOM and k-means algorithm. It eliminates the individual influence and local optimum above mentioned existing in traditional hierarchical dividing process, helps railway authorities better seize the significance of passenger nodes, and guides them, in certain ways, in compiling train operation plan and conducting transportation allocation. Nevertheless, since the main purpose of this paper lies in verifying the reasonability and effectively of the proposed approach, the index parameters selected for hierarchical dividing of passenger nodes still require to be modified in accordance with situations of specific cities; Furthermore, appropriate readjustment should be conducted for cluster results on the basis of qualitative analyses. And these above problems still need to be further studied.

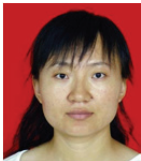
Acknowledgement

This work is supported by the Natural Science Foundation of Inner Mongolia (No. 2016MS0706 and 2017MS0702), the Institution of Higher Learning Science Research Project of Inner Mongolia (No. NJZY078), and the Science Research Project of Inner Mongolia University of Technology (No. ZD201522).

References

- [1] ChinaDaily, "Medium- and long-term railway network plan," 2016 [Online]. Available: http://www.chinadaily.com.cn/opinion/2016-07/01/content_25925793.htm.
- [2] W. X. Wang and H. X. Lyu, "Classification of railway passenger transport nodes based on affinity propagation cluster," *Application Research of Computers*, vol. 33, no. 10, pp. 2926-2928, 2016.
- [3] B. Gao, Y. Qin, X. M. Xiao and L. X. Zhu, "K-means clustering analysis of key nodes and edges in Beijing subway network," *Journal of Transportation Systems Engineering and Information Technology*, vol. 14, no. 3, pp. 207-213, 2014.
- [4] Y. Z. Xu and Y. Qin, "Factor analysis of key nodes in urban rail network," in *Proceedings of IEEE International Conference on Intelligent Transportation Engineering*, Singapore, 2016, pp. 27-31.
- [5] P. F. Zhou, B. M. Han, and Q. Zhang, "High-speed railway passenger node classification method and train stops scheme," *Applied Mechanics and Materials*, vol. 505-506, pp. 632-636, 2014.
- [6] J. S. Park and K. Lee, "Classification of the Seoul Metropolitan Subway Stations using graph partitioning," *Journal of the Economic Geographical Society of Korea*, vol. 15, no. 3, pp. 343-357, 2012.
- [7] T. Kohonen, "Self-organizing map," *Proceedings of the IEEE*, vol. 78, no. 9, pp. 1464-1480, 1990.
- [8] F. Wang, B. L. Xu, Y. W. Qian, Y. M. Dai and Z. Q. Wang, "Anomaly Detection Model Based on Hybrid Classifiers," *Journal of System Simulation*, vol. 24, no. 2, pp. 854-858, Feb. 2012.
- [9] Y. H. Jin, A. Kawamura, S. C. Park, N. Nakagawa, H. Amaguchi, and J. Olsson, "Spatiotemporal classification of environmental monitoring data in the Yeongsan River basin, Korea, using self-organizing maps," *Journal of Environmental Monitoring*, vol. 13, no. 10, pp. 2886-2894, 2011.
- [10] M. Alvarez-Guerra, C. Gonzalez-Pinuela, A. Andres, B. Galan, and J. R. Viguri, "Assessment of self-organizing map artificial neural networks for the classification of sediment quality," *Environment International*, vol. 34, no. 6, pp. 782-790, 2008.
- [11] K. Nishiyama, S. Endo, K. Jinno, C. B. Uvo, J. Olsson, and R. Berndtsson, "Identification of typical synoptic patterns causing heavy rainfall in the rainy season in Japan by a self-organizing map," *Atmospheric Research*, vol. 83, no. 2-4, pp. 185-200, 2007.
- [12] V. S. Lobo, "Application of self-organizing maps to the maritime environment," in *Information Fusion and Geographical Information Systems*. Heidelberg: Springer, 2009, pp. 19-36.
- [13] M. Liukkonen, E. Havia, H. Leinonen, and Y. Hiltunen, "Quality-oriented optimization of wave soldering process by using self-organizing maps," *Applied Soft Computing*, vol. 11, no. 1, pp. 214-220, 2011.
- [14] J. C. Creput, A. Hajjam, A. Koukam, and O. Kuhn, "Self-organizing maps in population based metaheuristic to the dynamic vehicle routing problem," *Journal of Combinatorial Optimization*, vol. 24, no. 4, pp. 437-458, 2012.
- [15] J. B. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, CA, 1967, pp. 281-297.

- [16] L. Zhang, M. Scholz, A. Mustafa, and R. Harrington, "Application of the self-organizing map as a prediction tool for an integrated constructed wetland agroecosystem treating agricultural runoff," *Bioresource Technology*, vol. 100, no. 2, pp. 559-565, 2009.
- [17] D. Bedoya, V. Novotny, and E. S. Manolagos, "Instream and offstream environmental conditions and stream biotic integrity: importance of scale and site similarities for learning and prediction," *Ecological Modelling*, vol. 220, no. 19, pp. 2393-2406, 2009.
- [18] S. Greco, R. Slowinski, and I. Szczech, "Properties of rule interestingness measures and alternative approaches to normalization of measures," *Information Sciences*, vol. 216, pp. 1-16, 2012.
- [19] H. L. Garcia and I. M. Gonzalez, "Self-organizing map and clustering for wastewater treatment monitoring," *Engineering Applications of Artificial Intelligence*, vol. 17, no. 3, pp. 215-225, 2004.
- [20] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 1, no. 2, pp. 224-227, 1979.
- [21] A. Hentati, A. Kawamura, H. Amaguchi, and Y. Iseri, "Evaluation of sedimentation vulnerability at small hillside reservoirs in the semi-arid region of Tunisia using the self-organizing map," *Geomorphology*, vol. 122, no. 1-2, pp. 56-64, 2010.



Chanchan Zhao <https://orcid.org/0000-0002-5441-4597>

She received the M.E. in Computer Science from Taiyuan University of Technology in 2007. She is pursuing the Ph.D. degree in School of Computer and Information Technology, Beijing Jiaotong University, Beijing, China. Her current research interests include optimization theory and method, wireless sensor networks, information fusion and interoperability technology.



Feng Liu <https://orcid.org/0000-0002-1777-7874>

He received the Ph.D. degree in School of Information, Renmin University of China in 2010. He is now a Professor in School of Computer and Information Technology, Beijing Jiaotong University, Beijing, China. His research interest include computer software, communications software and network management.



Xiaowei Hai <https://orcid.org/0000-0003-3406-7150>

He received the Ph.D. degree in School of Economics and Management, Beijing Jiaotong University in 2014. He is now an associate professor in Management College, Inner Mongolia University of Technology, Hohhot, China. His current research interests include information management, software engineering and high-speed railway technology.