A study of region division based on spatial units fusion of clustering algorithm

Yu Wang*

School of Transportation and Logistics, Southwest Jiaotong University, Chengdu 610031, Sichuan, China

School of Traffic and Transportation Engineering, Dalian Jiaotong University, Dalian 116028, Liaoning, China

Received 1 March 2014, www.cmnt.lv

Abstract

Region division is the foundation of regional socio-economic development planning. Traditional regional division is only built upon the information of attributes in each spatial unit. However, the spatial relationships and their spatial interaction between units are ignored. In the study, based on the spatial linkage theory, the principle and steps of spatial units fusion of clustering algorithm is proposed. According the spatial range of Chinese railway as a case, determine the spatial units and the linkage model was built based on the proximity and attribute characters, the experiments of regional division for Chinese railway are achieved. The experiment shows the results are highly accordant with the real situations and have proved the feasibility of the algorithm.

Keywords: spatial linkage theory, region division, spatial units fusion of clustering algorithm, Chinese railway

1 Introduction

Regional planning is an important part of regional socio-economic development planning. Region division is the foundation of regional planning, and the division foundation in current research is classified into two types. One take the economic attribute as the main basis [1,2], as economic belt [3], economic region [4] and metropolitan area [5]. But the academe has not come to an agreement with it, the reason is the principles and methods are differrent and can also affect the regional planning establishment and implementation. Another take the spatial units attribute as the main basis [6]. But this method only consider the individual attribute data and the relationship are ignored in practice.

Based on the spatial linkage theory [7], starting with the relationship with the contiguous spatial units, the fusion of clustering algorithm [8,9] were studied and so as to realize the region division.

2 Spatial units fusion of clustering algorithm

2.1 ALGORITHM PRINCIPLE

Within the selected spatial range, the spatial set can be formed under the proper spatial units are selected, by computing the units correlation data, the correlation coefficient set is determined, fused with the two adjacent spatial units which has maximum correlation coefficient, connect the units in one direction according to the units attribute until there are not any external connection by the fusion units, so the units which have fused cluster a class. According to this algorithm, all the spatial units are clustered several classes in the spatial range.

2.2 ALGORITHM STEPS

Step 1: Determination of the spatial range *s*.

Step 2: Determination of the spatial units s_i:

$S = \{S_1, S_2, ..., S_i\}$.

Step 3: Determination of the model of the relation coefficient *f*.

Step 4: Take a spatial unit as the starting point S_1, computed the relation coefficient f_{1j} with the neighboring spatial units, fused with the two neighboring spatial units which has maximum relation coefficient max f_{1j}.

Step 5: Determination of the neighboring spatial units in one direction connection T_{1j}. The units attribute data is used to represent the (Q_1, Q_2), determine the connect direction T_{1j} by comparing Q_1 and Q_2, if $Q_1 / Q_j > 1$, $T_{1j} = 1$, otherwise $T_{1j} = 0$.

Thus, the two contiguous units have only one direction connection.

Step 6: In the same way, by computing as step4 and step5 until there are not any external connection by the fusion units, so the units which have fused cluster the first class C_1.

Step 7: According to this algorithm, all the spatial units are clustered several classes, $S = (C_1, C_2, ..., C_k)$.

* ***Corresponding author's*** e-mail: wangyudjtu@163.com

3 An example of region division of Chinese railway

3.1 DETERMINE THE SPATIAL RANGE

Regional railway planning is the effective link to the "Medium and Long-term Railway Network Planning" [10], so the spatial range of region division of railway transportation should be carried out within the scope of national railway space (except Hong Kong, Macau and Taiwan).

3.2 DETERMINE THE SPATIAL UNITS

Chinese railway transport mainly presented as inter-provincial transport relationship, and the volume of freight flow mostly completed between 1,000~2,000 *km* [11]. So the spatial units selection are based on the provinces, municipalities and autonomous regions. Considering the uniqueness of the adjacent space, fused the only adjacent units to each other at first and the number of spatial units are 28. As shown in Table 1.

TABLE 1 Spatial units of region division of Chinese railway

Region code	Region name	spatial unit code	Region code	Region name	spatial unit code
110000	Beijing	S_1	410000	Henan	S_{15}
120000	Tianjin	S_1	420000	Hubei	S_{16}
130000	Hebei	S_1	430000	Hunan	S_{17}
140000	Shanxi	S_2	440000	Guangdong	S_{18}
150000	Mongolia West	S_3	450000	Guangxi	S_{19}
150000	Mongolia East	S_4	460000	Hainan	S_{18}
210000	Liaoning	S_5	500000	Chongqing	S_{20}
220000	Jilin	S_6	510000	Sichuan	S_{20}
230000	Heilongjiang	S_7	520000	Guizhou	S_{21}
310000	Shanghai	S_8	530000	Yunnan	S_{22}
320000	Jiangsu	S_9	540000	Tibet	S_{23}
330000	Zhejiang	S_{10}	610000	Shaanxi	S_{24}
340000	Anhui	S_{11}	620000	Gansu	S_{25}
350000	Fujian	S_{12}	630000	Qinghai	S_{26}
360000	Jiangxi	S_{13}	640000	Ningxia	S_{27}
370000	Shandong	S_{14}	650000	Xingjian	S_{28}

3.2 SPATIAL UNITS CORRELATION MODEL

Spatial units correlation model was built from the method of spatial transport linkage theory [7], according to the data of $O-D$ matrix and considering the bidirectional transport data transfer. The model can be used to evaluate the correlation strength with the adjacent spatial units, as shown by Equation (1):

$$f_{ij} = a_i \frac{O_{ij}}{O_i} + a_i \frac{D_{ij}}{D_i} + a_j \frac{D_{ij}}{O_j} + a_j \frac{O_{ij}}{D_j}, (i, j = 1, 2, ..., n), \quad (1)$$

where f_{ij} is the correlation coefficient, O_{ij} is the freight flow from unit i to j, D_{ij} is the freight flow from unit j to i, O_i and O_j is the total output freight flow with unit i and j, D_i and D_j is the total input freight flow with unit i and j, a_i and a_j is the standardization coefficient, n is the number of spatial units.

Regional division of Chinese railway has based on the analysis on the freight flow in china. a_i and a_j were obtained from "China Statistical Yearbook for Regional Economy-2013" [12], O_{ij}, D_{ij}, O_i, O_j, D_i, D_j were obtained from "China Transportation Yearbook 2013"[13].

3.3 CALCULATION PROCESS

Take S_1 as the starting point, search for the adjacent spatial units as S_2, S_3, S_5, S_{14}, S_{15} and calculate the f_{1j}. As shown in Table 2.

TABLE 2 The f_{1j} between the adjacent spatial units

Variable / Adjacent units	a_1	O_{1j}/O_1	D_{1j}/D_1	a_j	D_{1j}/O_j	O_{1j}/D_j	f_{1j}
S_1-S_2	0.96	0.03	0.53	0.88	0.49	0.05	1.00
S_1-S_3	0.96	0.01	0.19	1.58	0.29	0.02	0.68
S_1-S_5	0.96	0.01	0.01	1.41	0.04	0.02	0.10
S_1-S_{14}	0.96	0.07	0.04	1.37	0.11	0.05	0.33
S_1-S_{15}	0.96	0.02	0.01	0.81	0.01	0.03	0.02

From Table 2, we note that S_1 and S_2 have the maximum correlation coefficient, $\max f_{1j} = 1.00$. Determine S_1 and S_2 in one direction connection, the exchanges of freight volume can be regarded as (Q_1, Q_2), as $Q_i = O_{ij}$, $Q_j = D_{ij}$, O_{12}/D_{12}=0.02 < 1, so the only direction connection between S_1 and S_2 was $S_1 \leftarrow S_2$.

Take S_2 as the starting point, search for the adjacent spatial units as S_1, S_3, S_{15}, S_{24} and calculate the f_{2j}. As shown in Table 3.

TABLE 3 The f_{2j} between the adjacent spatial units

Adjacent units \ Variable	a_2	O_{2j} / O_2	D_{2j} / D_2	a_j	D_{2j} / O_j	O_{2j} / D_j	f_{2j}
$S_2 - S_1$	0.88	0.49	0.05	0.96	0.03	0.53	1.00
$S_2 - S_3$	0.88	0.01	0.01	1.58	0.01	0.02	0.06
$S_2 - S_{15}$	0.88	0.04	0.01	0.82	0.01	0.20	0.22
$S_2 - S_{24}$	0.88	0.01	0.02	0.90	0.02	0.04	0.07

From Table 3, we note that S_2 and S_1 have the maximum correlation coefficient, $\max f_{2j} = 1.00$. Determine S_1 and S_2 in one direction connection, as O_{21} / D_{21} =64.3 >1, so the only direction connection between S_2 and S_1 was $S_2 \to S_1$. Take S_3 as the starting point, search for the adjacent spatial units as S_1, S_2, S_{24}, S_{25} and calculate the f_{3j}. As shown in Table 4. From Table 4, we note that S_3 and S_1 have the maximum correlation coefficient, $\max f_{3j} = 0.68$.

TABLE 4 The f_{3j} between the adjacent spatial units

Adjacent units \ Variable	a_3	O_{3j} / O_3	D_{3j} / D_3	a_j	D_{3j} / O_j	O_{3j} / D_j	f_{3j}
$S_3 - S_1$	1.58	0.29	0.02	0.96	0.01	0.19	0.68
$S_3 - S_2$	1.58	0.01	0.02	0.88	0.01	0.01	0.06
$S_3 - S_{24}$	1.58	0.01	0.01	0.90	0.01	0.02	0.03
$S_3 - S_{25}$	1.58	0.01	0.01	0.54	0.01	0.03	0.04

Determine S_3 and S_1 in one direction connection, as O_{31} / D_{31}=57.6 >1, so the only direction connection between S_3 and S_1 was $S_3 \to S_1$.

Take S_{14} as the starting point, search for the adjacent spatial units as S_1, S_9, S_5 and calculate the f_{14j}. As shown in Table 5.

TABLE 5 The f_{14j} between the adjacent spatial units

Adjacent units \ Variable	a_{14}	O_{14j} / O_{14}	D_{14j} / D_{14}	a_j	D_{14j} / O_j	O_{14j} / D_j	f_{14j}
$S_{14} - S_1$	1.37	0.11	0.05	0.96	0.07	0.04	0.33
$S_{14} - S_9$	1.37	0.03	0.01	1.76	0.04	0.06	0.24
$S_{14} - S_{15}$	1.37	0.10	0.02	0.82	0.04	0.13	0.30

From Table 5, we note that S_{14} and S_1 have the maximum correlation coefficient, $\max f_{14j}$ =0.33. Determine S_{14} and S_1 in one direction connection, as O_{141} / D_{114} =1.72 >1, so the one direction connection between S_3 and S_1 was $S_{14} \to S_1$. There are not any external connection by the unit S_1, so the units S_1, S_2, S_3, S_{14} which have fused cluster the first class $C_1 = (S_1, S_2, S_3, S_{14})$, the fused path was $S_2 \to S_1; S_3 \to S_1; S_{14} \to S_1$.

As the calculation process is same, the other class calculation is not repeated, the fused cluster steps as shown in Table 6.

TABLE 6 The fused cluster steps of the class

Class	Step	Starting unit	Maximum relation unit	max f_{ij}	T_{ij}
C_1	1	S_1	S_2	1.01	$S_1 \to S_2$
	2	S_2	S_1	1.01	$S_2 \to S_1$
	3	S_3	S_1	0.68	$S_3 \to S_1$
	4	S_{14}	S_1	0.33	$S_{14} \to S_1$
C_2	1	S_4	S_5	0.66	$S_4 \to S_5$
	2	S_5	S_6	0.73	$S_5 \leftarrow S_6$
	3	S_6	S_5	0.73	$S_6 \to S_5$
	4	S_7	S_6	0.53	$S_7 \to S_6$
C_3	1	S_8	S_{10}	0.51	$S_8 \leftarrow S_{10}$
	2	S_9	S_8	0.48	$S_9 \to S_8$
	3	S_{10}	S_8	0.51	$S_{10} \to S_8$
	4	S_{11}	S_9	0.68	$S_{11} \to S_9$
C_4	1	S_{12}	S_{13}	0.66	$S_{12} \to S_{13}$
	2	S_{13}	S_{12}	0.66	$S_{13} \leftarrow S_{12}$
C_5	1	S_{15}	S_{16}	0.24	$S_{15} \leftarrow S_{16}$
	2	S_{16}	S_{15}	0.24	$S_{16} \to S_{15}$
	3	S_{24}	S_{16}	0.23	$S_{24} \to S_{16}$
C_6	1	S_{17}	S_{18}	0.81	$S_{17} \leftarrow S_{18}$
	2	S_{18}	S_{17}	0.81	$S_{18} \to S_{17}$
C_7	1	S_{19}	S_{21}	0.54	$S_{19} \to S_{21}$
	2	S_{20}	S_{21}	0.67	$S_{20} \leftarrow S_{21}$
	3	S_{21}	S_{20}	0.67	$S_{21} \to S_{20}$
	4	S_{22}	S_{20}	0.43	$S_{22} \to S_{20}$
C_8	1	S_{23}	S_{26}	0.11	$S_{23} \leftarrow S_{26}$
	2	S_{25}	S_{27}	0.29	$S_{25} \leftarrow S_{27}$
	3	S_{26}	S_{25}	0.13	$S_{26} \leftarrow S_{25}$
	4	S_{27}	S_{25}	0.29	$S_{27} \to S_{25}$
	5	S_{28}	S_{25}	0.1	$S_{28} \to S_{25}$

The transport relation with spatial units has its own characteristics and the characteristics distributed on the geospatial form the different transportation regions.Shown in Table 7.

TABLE 7 Name and composition of transportation regions

No.	Code	Name of region	Composition of region
1	C_1	Northern	Hebei, Beijing, Tianjin, Shanxi, Mongolia West, Shandong
2	C_2	Northeast	Heilongjiang, Jilin, Liaoning, Mongolia East
3	C_3	East	Shanghai, Jiangsu, Anhui, Zhejiang
4	C_4	Southeast	Fujian, Jiangxi
5	C_5	Central	Shaanxi, Henan, Hubei
6	C_6	Southern	Guangdong, Hainan, Hunan
7	C_7	Southwest	Guangxi, Guizhou, Sichuan, Chongqing, Yunnan
8	C_8	Northwest	Gansu, Ingria, Xingjian, Qinghai, Tibet

4 Conclusion

Traditional regional division model for spatial units is only built upon the information of attributes in every unit. However, the spatial relationships and their spatial interacttion between units are not considered sufficiently. In this study, based on scale-space theory, the spatial units cluster fusion algorithm is proposed so that the elements of spatial relationship between units could be integrated besides considering the information of attributes. By this approach, considered both of the spatial proximity and attribute similarity at certain spatial range, the spatial units could be clustered into one class if their connective direction is the same and achieve the regional division.

This paper uses the spatial transport linkage theory to construct the spatial units correlation model for railway freight transportation regional division. According to the data of $O-D$ matrix with the adjacent spatial units and considering economic parameters as limitation, the regional division is divided 8 parts of railway freight transporttation in china. As Northern, Northeast, East, Southeast, Central, Southern, Southwest and Northwest. As shown in Figure 1. The results are highly accordant with the real situations and can offer spatial background for regional railway freight transportation planning and development.

Further research is also proposed to improve spatial units correlation model. The data selection should not only reflect the present but also reflect the future trends of the regional development. Economic indicators should be further improved when considering the link between transportation and economic and also considering the impact of macroeconomic policies and geographic and other factors and make the model more comprehensive.

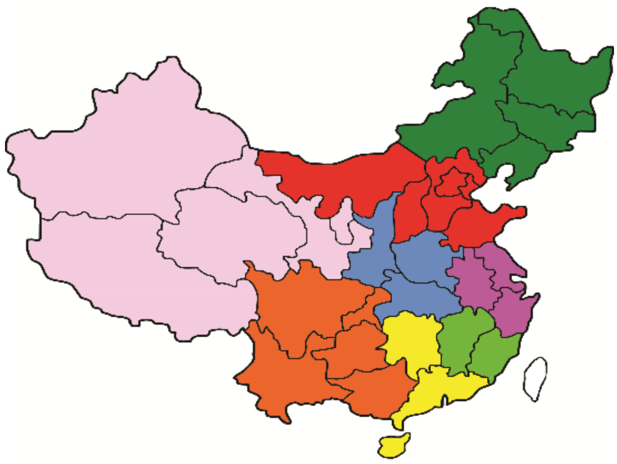

FIGURE 1 The spatial layout of railway freight transportation regional division

Acknowledgments

The project has been supported by National Natural Science Foundation ,China (Grant No. 71173177) and International Exchange Programs, Muroran Institute of Technology, Japan.

References

[1] Fujita M, Krugman P, Venables A J 2001 The spatial economy: cities, regions and international Trade Models of Disorder *MIT Press:* Edinburgh – Cambridge Chapter 6

[2] Capello R 2007 Regional Economics *Routledge Chapman & Hall Press:* London Chapter 4

[3] Zhang Z Z 2010 Evolution and evaluation of the Chinese economic regions division *Journal of Shanxi Finance and Economics University* **13** (2) 89-92 *(in Chinese)*

[4] Bureau of Economic Analysis 2010 *Regional economic measurement division: REIS Regional Economic Information System, 1969-2012"* US

[5] Yang L, Yang K Z 2007 Metropolitan-oriented regionalization of japan and lessons for china *Asia-pacific Economic Review* **33**(2) 113-7*(in Chinese)*

[6] Yuan C W, Wang J W, Lin W X 2006 Division method of transportation regions in china *Journal of Chang'an University* **26**(4) 79-82 *(in Chinese)*

[7] Zhang W C, Jin F J, Rong C H, Tang X F 1994 Analysis on spatial transport linkage *China railway Press:* Beijing Chapter 4

[8] Lindeberg T 1994 Scale-space theory in computer vision *Kluwer Press:* Netherlands Chapter 2

[9] Leung Y, Zhang J, Xu Z 2002 Clustering by Scale-space filtering *IEEE transactions on pattern analysis and machine Intelligence* **22**(12) 1396-410

[10] Ministry of Railways 2008 Medium and Long-Term Railway Network Plan (adjusted in 2008) *China railway Press:* Beijing Chapter 1

[11] Ren M 2007 Analysis on the railway freight exchange characteristics of provincial region *Chinese railways* **23**(10)25-28 *(in Chinese)*

[12] National Bureau of Statistics, The competent national transport ministries Bureau and validation 2013 China Transportation Yearbook *China statistical Press:* Beijing Chapter 4

[13] National Bureau of Statistics 2013 China Statistical Yearbook for Regional Economy – 2013 *China statistical Press:* Beijing Chapter 12

Author

Yu Wang, born in 1979, Dalian, Liaoning, China.

Current position, grades: lecturer at the School of Traffic and Transportation Engineering, Dalian Jiaotong University, China.
University studies: B.E. at Southwest Jiaotong University in China. ME at Southwest Jiaotong University in China.
Scientific interests: traffic and transportation planning and management.
Publications: more than 10 papers.
Experience: teaching experience of 5 years, presided 1 scientific research project.