Comprehensive Transportation Logistics Network Level Layout Based on Principal Component Factor and Cluster Analysis

Jingrong Zhang

School of Management Engineering, Zhengzhou University, Zhengzhou Henan, China
Corresponding author: 13676995513@139.com

Abstract. Comprehensive Transportation Logistics Network (CTLN) acts as a crucial prop and fundamental carrier for regional economic and social development. Firstly, an index system for evaluating the development of regional Comprehensive Transportation Logistics (CTL) nodes is established; then regional CTLN nodes are ranked according to their importance by the method of Principal Component Analysis(PCA), and main factors affecting the development of regional CTL nodes are analyzed by applying factor analysis, and regional CTL nodes are classified according to their feature similarities by applying cluster analysis; and then level structure of constructing regional CTLN is proposed. Finally, combined with geographic locations of different nodes, level layout model of CTLN of the whole region is obtained. Taking Henan province this region as an instance, level layout model of hub-and-spoke CTLN taking Zhengzhou at its core is proposed after analysis, providing a reference basis for constructing CTLN in whole province scientifically and reasonably.

1 Introduction

Transportation logistics remains a vital precursor and basis for the development of modern logistics, also a key link and crucial component for the development of regional logistics, as well as acts as the significant driving force for developing regional manufacturing industry and commercial circulation business. And transportation logistics network is an important prop and fundamental carrier for the development of regional economy and society and for joint development of logistics industry and manufacturing industry. With constant development and integration of modern logistics and transportation, as well as with deepening supply-side structural reform in transportation logistics field, transportation logistics is striding to an advanced stage of CTL where various modes of transportation are balanced, cooperated according to their functions and integrated organically, meanwhile transportation logistics network during construction is approaching to an advanced stage of seamless, interconnected and reasonably arranged CTLN.

In recent years, studies on transportation logistics network have been conducted by many scholars from various respects, and many achievements have been gained. For instance, Shi Jin et al. [1] have discussed integrated construction of cross-border subregional logistics transportation, Chen Wenqiang et al. [2] have probed into reasonable pricing of toll roads in logistics transportation network by applying game theory approach; He Qichao et al. [3-5] have inquired into influences to regional logistics network planning from the views of carbon tax constraint and transportation cost; Han Shilian et al. [6-8] have studied route planning, scheduling and optimizing issues concerning transportation logistics network; Zhang Shixiang et al. [9-11] have studied the construction and cases of hub-and-spoke logistics network respectively.

However, throughout existing study achievements, documents concerning transportation logistics network layout from the respect of comprehensive transportation are not retrieved. Given this, the present paper proposes nodes ranking of regional CTL according to their importance by the method of PCA, then main factors affecting the development of regional CTL nodes are analyzed by the method of factor analysis, and regional CTL nodes are classified according to their similarities by the method of cluster analysis. Thereafter, based on network nodes importance ranking, influencing factors analysis and similarity classification, level structure of regional CTLN is proposed, then combined with geographical location of various nodes, level layout model of CTLN of the whole region is obtained, providing a reference basis for constructing regional CTLN system scientifically and reasonably.

2 CTL evaluation index system

To measure CTL development level in different nodes, and in line with index selecting principles of interpretability, accessibility, comparability, comprehensiveness and completeness, an evaluation index system is established after consulting abundant documents, see Table 1.

Table 1. CTL evaluation index system.

Primary index	Secondary index
Node Comprehensive Development Level A1	Total Output Value B1
	Per Capita Total Output Value B2
	Fixed Asset Investment B3
CTL Market Size A2	Total Society Freight Volume B4
	Freight Turnover B5
	Total Retail Sales of Social Consumption B6
	Total Export Import Volume B7
	Added Value of Industries above a Designated Scale B8
	Business Volume of Postal and Telecommunication Services B9
	Volume of Telecommunication Services B10
CTL Facility Level A3	Length of Highways on Operation B11
	Railway Lines Transit Condition B12
	Inland Navigation Condition B13
	Air Cargo Loading and Unloading Capacity B14
	National Highway Freight Hub Condition
	National Railway Logistics Base Condition

Wherein, node comprehensive development level (A1): reflects economic and social foundation of CTL development in different nodes, including total output value (B1), per capita total output value (B2) and fixed asset investment (B3).

CTL market size (A2): reflects CTL development level and size in different nodes, including total society freight volume (B4), turnover volume of freight transport (B5), total retail sales of social consumption, total export import volume (B7), added value of industries above a designated scale (B8), business volume of postal and telecommunication services (B9) and volume of telecommunication services (B10).

CTL facility level (A3): reflects material foundation and facility condition of CTL development in various nodes, including length of highways on operation (B11), railway lines transit condition (B12), inland navigation condition (B13), air cargo loading and unloading capacity (B14), national highway freight hub condition (B15) and national railway logistics base condition (B16).

3 CTL nodes ranking according to their importance based on PCA

Though indexes being screened, a certain degree of correlation and information overlapping still exist. While by PCA [12], original indexes can be replaced by fewer comprehensive principal component indexes through dimensionality reduction, where new indexes have reserved most information of original ones and new indexes are independent and uncorrelated with each other, getting issues simplified. Therefore, PCA can be applied in comparing and ranking CTL development level in different nodes to determine different importance degree of various nodes in CTLN.

Supposing, CTL evaluation sample matrix : $X^* = (x^*_{ij})_{m \times n}$, $i = 1,2,\cdots,m$, $j = 1,2,\cdots,n$ is obtained with n indexes of m nodes. Firstly, indexes are standardized through formula (1), then standard evaluation matrix is obtained: $X = (x_{ij})_{m \times n}$

$$x_{ij} = (x^*_{ij} - \overline{x_j^*})/s_j^* \tag{1}$$

Wherein, $\overline{x_j^*}$ and s_j^* presents sample mean of the jth index and sample standard deviation respectively.

Secondly, calculate correlative coefficient matrix $R_{n \times n}$ among indexes and its eigenvalues $\lambda_1 \geq ... \geq \lambda_n \geq 0$ as well as regularized eigenvector e_j;

Then, principal component is obtained

$$Y_j = (y_{ij}) = X e_j \tag{2}$$

Wherein, variance contribution rate of the jth principal component is $\alpha_j = \lambda_j / n$. When accumulated variance contribution rate $\alpha = \sum_{j=1}^{k} \alpha_j$ reaches a certain value (generally ≥85%), it can be regarded that the first k principal components $Y_1 - Y_k$ have embodied information of original indexes comprehensively with fewer indexes.

Then, taking variance contribution rate α_j of the first k components as weight, comprehensive evaluation function is obtained through linear weighing sum

$$z_i = \sum_{q=1}^{k} \alpha_q y_{iq}, \quad q = 1, 2, ..., k \leq n \tag{3}$$

A higher value of z_i indicates higher comprehensive strength and competitiveness, namely a higher importance degree of this city as logistics node among the whole regional CTLN; while a lower value of z_i indicates lower comprehensive strength and competitiveness.

4 Influencing factor analysis of ctl development based on factor analysis

Factor analysis [13] applies a few abstract common factors to express data structure of original indexes through dimensionality reduction and simplification. Compared with PCA, factor analysis is more explanatory, therefore, it is usually conducted on the basis of PCA to study main influencing factors of index variables. Thus factor analysis model further inspecting main factors influencing the development of regional CTL nodes is obtained:

$$Y_j = \mu_j + A_j F_k + \varepsilon_j, \quad j = 1, 2, \cdots, n, \ k \leq n \tag{4}$$

Wherein, μ_j is mean value, A_j is factor loading matrix, F_k is common factor, and ε_j is specific factor.

To make common factors more explanatory, factor matrix is usually rotated. What the present paper applies is the commonest varimax orthogonal rotation.

5 CTL nodes similarity analysis based on cluster analysis

Cluster analysis [14] gradually classifies individuals or groups according to feature similarities of their index variables to obtain classification system that can reflect close or distant relationship among individuals or groups. The first k principal components are adopted to analyze and classify CTL nodes according to their feature similarities, which greatly simplifies the present issue.

Firstly, the first k principal components comprise a new matrix $Y^* = (y^*_{iq})_{m \times k}$ and distances among new sample data of various nodes are calculated according to the formula (5). What the present paper applies is the commonest Euclidean distance. A smaller distance coefficient indicates a higher feature similarity level of CTL development among nodes, namely, a higher similarity level among CTL nodes; While a lower one indicates a lower feature similarity level.

$$d_{iq} = \left[\sum_{q=1}^{k} (y^*_{iq} - y^*_{jq})^2 \right]^{1/2}, \quad q = 1, 2, \cdots, k \leq n \tag{5}$$

6 CTLN level layout

In the present paper, different nodes in CTLN are compared and ranked according to their comprehensive competitiveness and importance by the method of PCA; main factors affecting the development of different nodes in CTLN are analyzed by the method of factor analysis; and different nodes in CTLN are compared and classified according to their similarity degree and similar features by the method of cluster analysis.

Based on the above-mentioned analysis to importance ranking, influencing factor analysis and similarity classification of network nodes, corresponding level and category of various nodes in CTLN can be proposed, therefore network level structure of regional CTLN can be determined. Then combined with geographical locations of different nodes, level layout model of CTLN in the whole region can be determined.

7 Case analysis: henan province CTLN level layout

Henan province is a vital comprehensive transportation hub in China, thus there is the old saying "who gains the Central Plains gains the world" since ancient times. In national strategic planning, one important position of Henan is "national important comprehensive transportation hub and modern logistics center" [15]. At present, economic and social development in this province is increasingly accelerated, and the level of opening up rises gradually. The implementation of "One Belt One Road" strategy, the construction of Central Plains economic zone, as well as gradually deepening implementation of Henan free trade zone, Zhengzhou airport comprehensive experimental area and other national strategies all have put forward new requirements to the construction of CTLN in Henan province.

7.1 Related data

Latest data of this study (2015), are mainly from two ways:

(1) From statistical material. The following data can be obtained from Henan Province Statistic Yearbook and statistical bulletin of each city respectively: total output value (B1), per capita total output value (B2), fixed asset investment (B3) in node comprehensive development level (A1); total society freight volume (B4), turnover volume of freight transport (B5), total retail sales of social consumption (B6), total export import volume(B7), added value of industries above a designated scale (B8), business volume of postal and telecommunication services (B9) and volume of telecommunication services (B10) in CTL market size (A2); as well as length of highways on operation (B11), inland navigation condition (B13), air cargo loading and unloading capacity (B14) in CTL facility level (A3).

(2) Based on expert scoring. The three qualitative indexes: railway lines transit condition (B12), national highway freight hub condition (B15) and national railway logistics base condition (B16) can be determined by expert scoring on highway and railway infrastructure construction according to specialty statistic materials of each city and "National Highway Transportation Hub Layout Planning" from former Ministry of Transportation and "Railway Logistics Base Layout Planning & 2015-2017 Construction Planning" from China Railway Corporation. In expert scoring, interval values from 0-10 is adopted, for instance, Zhengzhou is crossed by "米"shaped main line railway, therefore its score is 10; city has no railway, then the score is 0; similarly the scores of other cities can be obtained by the same method.

7.2 CTL nodes ranking according to their importance

Firstly, data to be analyzed is standardized, and then corresponding eigenvalue and variance contribution rate of each principal component are obtained by applying principal component operation in SPSS, see Table 2.

Table 2. Corresponding eigenvalue and variance contribution rate of each principal component.

Principal component	Eigenvalue	Variance contribution rate α_j	Cumulative variance contribution rate α
Y1	9.910	61.939	61.939
Y2	2.903	18.143	80.081
Y3	1.004	6.276	86.357
Y4	0.601	3.758	90.115
Y5	0.529	3.307	93.423
Y6	0.408	2.551	95.974
Y7	0.221	1.382	97.356
Y8	0.181	1.134	98.490
Y9	0.128	0.802	99.292
Y10	0.067	0.419	99.711
Y11	0.028	0.176	99.887
Y12	0.009	0.058	99.945
Y13~Y16	≈0	≈0	100

To reduce information loss as much as possible and to ensure the following operation close to initial condition to the greatest extent, the first 6 principal components are selected, then the accumulated variance contribution rate is $\alpha = 95.97\%$, namely the 6 principal components have reflected the information of original indexes with the precision of 95.97%. Then obtain

$$z = 0.619Y1 + 0.181Y2 + 0.063Y3 + 0.038Y4 + 0.033Y5 + 0.026Y6 \quad (6)$$

Table 3 presents nodes comprehensive ranking according to their importance in Henan CTLN. Since centralized standardized processing method is adopted, there exist plus z_i values and minus z_i values.

Table 3. CTL nodes ranking according to their importance.

City	Zhengzhou	Luoyang	Nanyang	Zhoukou	Xinyang	Xinxiang	Shangqiu	Anyang	Jiaozuo
Score	6.86	1.74	1.66	0.62	0.61	0.22	0.18	-0.25	-0.4
Rank	1	2	3	4	5	6	7	8	9
City	Xuchang	Zhumadian	Luohe	Pingdingshan	Kaifeng	Puyang	Sanmenxia	Hebi	Jiyuan
Score	-0.7	-0.71	-0.85	-0.89	-1.2	-1.42	-1.43	-1.9	-2.13
Rank	10	11	12	13	14	15	16	17	18

According to Table 3, CTL nodes according to their importance in this whole province can be sorted into 3 levels as shown in Table 4. Wherein, the score of Zhenghou is far ahead, belonging to the first level; Luoyang, Nanyang, Zhoukou, Xinyang, Xinxiang, Shangqiu as CTL nodes with general importance fall into the second level (z>0); An'yang, Jiaozuo, Xuchang, Zhumadian, Luohe, Pingdingshan, Kaifeng, Puyang, Sanmenxia, Hebi and Jiyuan as CTL nodes with lower importance fall into the third level (z<0).

Table 4. CTL nodes importance level.

Level	CTL node
The first level	Zhenghou
The second level	Luoyang, Nanyang, Zhoukou, Xinyang, Xinxiang, Shangqiu
The third level	An'yang, Jiaozuo, Xuchang, Zhumadian, Luohe, Pingdingshan, Kaifeng, Puyang, Sanmenxia, Hebi and Jiyuan

7.3 Main Influencing factors of CTL nodes development

Since common factors usually realize their explanation to original indexes through conceptual abstraction, it is inadvisable to adopt too many common factors, and generally accumulated variance contribution rate is $\alpha \geq 80\%$. On the basis of the above-mentioned PCA, factor loading matrix is orthogonally rotated by maximum variance, and then factor loading and common factor variance after orthogonal rotation are obtained as shown in Table 5. To make analysis more convenient, the first 3 common factors are extracted, and the factor variance contribution rate reaches 92%, satisfying analysis requirements well [12,13].

In Table 5, factor loading matrix has provided correlation coefficient of each index on the first 3 common factors. A greater correlation coefficient indicates greater importance and influence of this index on certain common factor, while communality describes the contribution of all common factors make to the total variance of index variables. Meanwhile, communalities of most variables in Table 5 are high after extracting common factors, implying that most information is reserved while original variables are rotated to factor variables.

Table 5. Factor loading matrix after orthogonal rotation.

Index	B1	B3	B6	B7	B8	B9	B10	B11
Factor 1	**0.944**	**0.953**	**0.951**	**0.887**	**0.914**	**0.971**	**0.955**	-0.002
Factor 2	-	-0.219	-0.278	0.299	-0.025	-0.114	-0.203	**-0.945**
Factor 3	0.180	0.097	0.062	0.098	0.217	-0.001	0.144	0.247
Communality	**0.974**	**0.966**	**0.985**	**0.885**	**0.883**	**0.956**	**0.974**	**0.953**
Index	B12	B14	B15	B4	B5	B2	B13	B16
Factor 1	0.132	0.233	0.127	0.427	0.059	0.763	-0.002	0.749
Factor 2	-	**-0.907**	**-0.932**	-0.323	-0.182	-0.383	-0.945	-0.451
Factor 3	0.139	0.271	-0.029	**0.868**	**0.956**	0.247	0.247	0.264
Communality	**0.962**	**0.950**	**0.886**	**0.877**	**0.950**	**0.790**	**0.953**	**0.834**

Note：KMO=0.776>0.5，Sig.=0.000<0.05。

From Table 5, it can be known that there are 3 principal factors (namely common factors) affecting development comprehensive competitiveness of regional CTL nodes. The first common factor is determined by total output value (B1), fixed asset investment (B3), total retail sales of social consumption (B6), total export import volume (B7), added value of industries above a designated scale (B8), business volume of postal and telecommunication services (B9), volume of telecommunication services (B10), which can be defined as economic development factors. The second common factor is determined by length of highways on operation (B11), railway lines transit condition (B12), air cargo loading and unloading capacity (B14), which can be defined as comprehensive transportation infrastructure factors. The third common factor is determined by national highway freight hub condition (B15) and freight turnover (B5), which

can be defined as CTL supply and demand factors, and these two indexes not only reflect the supply and demand in CTL development, but directly present CTL development.

7.4 CTL nodes similarity analysis

New sample matrix is formed by the obtained 6 principal component indexes, and then cluster analysis is conducted by applying Ward cluster approach according to Euclidean distance in SPSS. When Euclidean distance is 20, cities in Henan province as CTL nodes can be divided into 3 categories based on Euclidean distance similarity, as shown in Figure 1.

Wherein, the first category only includes Zhengzhou; the second category includes Zhoukou, Xinyang, Luohe, Shangqiu, Xinxiang, Nanyang and Luoyang; the third category includes Jiaozuo, Jiyuan, Sanmenxia, An'yang, Pingdingshan, Zhumadian, Kaifeng, Puyang, Xuchang and Hebi.
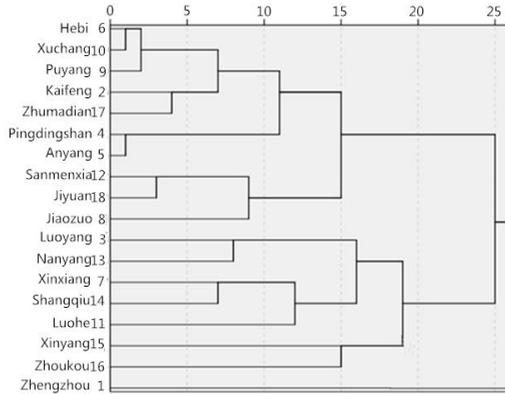

Fig.1. CTL nodes cluster figure.

Table 6. CTL nodes similarity category.

Category	CTL node
The first category	Zhengzhou
The second category	Luoyang, Nanyang, Xinxiang, Shangqiu, Luohe, Xinyang, Zhoukou
The third category	Jiaozuo, Jiyuan, Sanmenxia, An'yang, Pingdingshan, Zhumadian, Kaifeng, Puyang, Xuchang, Hebi

(1) Zhengzhou, the provincial capital of Henan province, is the political, economic and cultural center of this province, as well as the comprehensive transportation and communication hub equipped with national highway, railway, aviation and communication, and also is one of the 21 national logistics node city. Many industrial and commercial enterprises cluster in Zhengzhou, therefore there are large number and size of transportation logistics enterprises. At present, Zhengzhou is being constructed as a national central city and international logistics center, whose strategic position in future will be more prominent. Therefore, only Zhengzhou falls into the first category.

(2) Luoyang, Nanyang, Xinxiang, Shangqiu, Luohe, Xinyang and Zhoukou possess favorable geographical location and transportation condition and are crossed by "十" shaped national main line railway and high-grade highway, belonging to national highway transportation junctions planned by former Ministry of Transportation. And all of these cities enjoy fine economic and social development and have solid industrial base and dynamic social consumption. Except Luohe, these cities possess large administrative areas; Xinyang, Zhoukou and Luohe also have water transportation. Therefore, these cities are sorted into the same category.

(3) Though Jiaozuo, Jiyuan, Sanmenxia, An'yang, Pingdingshan, Zhumadian, Kaifeng, Puyang, Xuchang and Hebi also are crossed by national main line railway and high-grade highway, yet the shape generally is "一"; and they are far away from Zhengzhou, except Kaifeng and Xuchang, some of them locate in the border of two provinces; these cities have not so large area and population. Therefore location and transportation advantages of these cities are faded, and transportation logistics radiating capacity is lowered. But these cities have their own features in economic and industrial development. Thus these cities are sorted into the same category.

7.5 Henan CTLN level layout

Based on the above mentioned analysis, it can be seen that: there are many similarities in analysis results in nodes importance ranking (Table 4) and nodes similarity clustering (Table 6). Except the divergence in recognition of "Luohe" in the second level (the second category), other levels (categories) are the same. The difference in Table 4 and Table 6 is caused by different study emphasis placed by the two methods. PCA lays its emphasis on the study of comprehensive competitiveness and importance of each node in CTL development, while cluster analysis places its study emphasis on feature similarity among nodes in CTL development.

Then Henan CTL level structure is obtained as shown in Table 7. Wherein, Zhengzhou is far ahead others as the first level node of CTL; Luoyang, Nanyang, Zhoukou, Xinyang, Xinxiang and Shangqiu possess obvious advantages in regional transportation, economic and social development as the second level nodes of CTL; Though Luohe is similar to Luoyang and other five cities to certain extent, its node importance in CTL is low, thus it together with other cities falls into the third level nodes.

Table 7. CTLN level structure.

Node	CTL node
The first level node	Zhengzhou
The second level node	Luoyang, Nanyang, Zhoukou, Xinyang, Xinxiang, Shangqiu
The third level node	An'yang, Jiaozuo, Xuchang, Zhumadian, Luohe, Pingdingshan, Kaifeng, Puyang, Sanmenxia, Hebi, Jiyuan

Further more, combined with geographical locations of different cities in Henan, hub-and-spoke CTLN layout can be established with Zhengzhou as its axis city, Luoyang, Nanyang, Zhoukou, Xinyang, Xinxiang and Shangqiu as its radiating cities and other cities supporting nodes as shown in Figure 2. Modern CTLN with Henan characteristics where various modes of transportation are closely connected, efficiently linked up and functions are complemented, supporting the rapid economic and social development in Henan province.
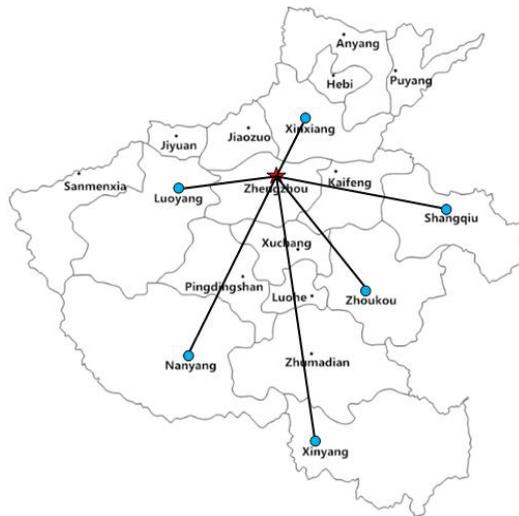

Fig.2. Henan CTLN level layout diagram.

8. Conclusion and Suggestion

(1) The primary factor influencing CTL development is regional economic development level, namely, transportation logistics serves for regional economic development and also is benefited from regional economic development.

(2) Departments of transportation logistics should fully consider various development features of CTL in different areas, and issue corresponding guiding measures to promote joint development and overall uplift of CTL in the whole region.

(3) While constructing CTLN in the whole province, layout and implementation should be conducted in accordance with the three logistics nodes levels: national level, provincial level and regional level. Wherein, Zhengzhou can be positioned as national hub; Luoyang, Nanyang, Zhoukou, Xinyang, Xinxiang and Shangqiu the provincial hubs, other rest cities the regional hubs. Zhengzhou functions as leader, core and growth pole in CTLN in the whole province.

References

1. J. Shi, Y.J. Liu, Econ Geogr. **32**, 98(2012).
2. W.Q. Chen, Y.N. Liu, Log Technol. **33,** 101(2013).
3. Q.C. He, W.P. Luo, L.H. Li, J. RY Sci Eng. **11**, 131(2014).
4. W. Wei, Y.H. Cao, W.D. Cao, Sci Geogr Sin, **29**, 485(2009).
5. X.L. Wang, S.Y. Wang, Pro Geogr, **31**, 1583(2012).
6. S.L. Han, X.W. Liu, Ops Res Manage Sci, **23**, 55(2014).
7. G. Zheng, Q. Li, Bull Sci Technol, **29**, 109(2014).
8. S.G. Song, A.P. Li, L.Y. Xu, J. TJU(nat sci), **38**, 113(2010).
9. S.X. Zhang, J.Z. Huo, Chin J. Manage, **2(S2)**, 194(2005).
10. W. Ding, L. Zhang, J. Li, Chn Soft Sci Mag, **25,** 160(2010).
11. W.B. Li, Y.S. Zhang, Econ Geogr, **31**, 1335(2011).
12. H.M. Lin, Z.F. Du, Stat Res, **30**, 25(2013).
13. J. Yu, W.B. Lu, J. CDU (Nat sci), **22**, 37(2003).
14. X.R. Li, J SD Educ Inst,**14,** 23(2007).
15. National Development and Reform Commission, *Central Plains Economic Zone Planning*(2012).