International Journal of Computer Science and Engineering (IJCSE)
ISSN(P): 2278–9960; ISSN(E): 2278–9979
Vol. 9, Issue 1, Dec–Jan 2020; 1–10

International Academy of Science, Engineering and Technology
Connecting Researchers; Nurturing Innovations

FUZZY GMDH-TYPE METHOD AND ITS APPLICATION IN BOTTLE-NECK DIAGNOSIS OF REGIONAL ECONOMIC SYSTEM

Zheng Shi[1] Zheng Wen[2] & Jin Xia[3]

[1,3]Research Scholar, Dalian University for Minorities, School of Economic and Management, Liaoning Province Dalian Development Zone, China

[2]Research Scholar, Northeastern University in Qinhuangdao, Hebei Province, Qinhuangdao, China

ABSTRACT

The purpose of this paper is to recognize Bottle-neck Diagnosis in Regional Economic System. This paper presents a new self-organizing data mining method applied to Bottle-neck Diagnosis in Regional Economic System. The result shows that the new method is more effective in diagnosis compared with GMDH method. Its algorithm is relatively simple. It relies on the expert's destination of the inputs and outputs from lots of data instead of the models. The comparative results between new method and the GMDH method show that the accuracy of this new method is similar as GMDH method, however the additional precondition can be found. This is the prominent characteristic of the new method. Thus, its application in diagnosis in the fields of Management Science is more promising.

KEYWORDS: *Fuzzy GMDH-Type Method, Bottle-Neck Diagnosis, Self-Organizing Data Mining, Regional Economic System, GMDH-Type Neural Networks Algorithm*

Article History

Received: 09 Nov 2019* | *Revised: 26 Nov 2019* | *Accepted: 30 Nov 2019

INTRODUCTION

Data mining technology has received great attention by experts at home and abroad in recent years, because it can mine and learn valuable and implicit knowledge from a large number of data[1]. It also has been widely used in the field of fault diagnosis [2].

GMDH-Type Neural Networks is an algorithm first proposed by Kondo, a Japanese scholar. This algorithm has been widely used in economic management and social system prediction because it can objectively determine the hierarchy and the number of hidden layers in the neural network and avoid the subjectivity of data partitioning in GMDH[3].

The core idea of GMDH-Type Neural Networks is to identify the non-linear system model by continuously sifting the combined model through the external criteria (accuracy criteria or compatibility criteria) of the neural network and GMDH algorithm [4]. If this idea is applied to the diagnosis of the bottleneck in the regional economic system, the basic law of the relationship between the most direct bottleneck of regional economic development and the current situation of regional economy can be discovered. However, if this method is used without improvement, the relationship obtained is a deterministic causal relationship, and the information used is deterministic information, which is often far from the diagnostic facts of regional economic bottlenecks. To improve this situation, we can adopt the method of NF-GMDH network, which integrates GMDH network with fuzzy logic, so that GMDH network can use not only data information, but also language information. Each layer of neurons in NF-GMDH is a fuzzy model. But after the input of each layer is determined, the condition part of the rule is determined. The parameters of the fuzzy model are estimated by HPM (Hybrid Projection Method) method, which neurons are retained and deleted by using external criteria, and the output of the selected neurons constitutes the output of the intermediate neurons[6-9]. When the test data criteria of each layer are[1] no longer reduced, the optimal fuzzy model is obtained. The method of NF-GMDH solves the problems of extracting fuzzy rules and utilizing linguistic information, but it does not solve the problem of testing rules, so it is impossible to mine other data and complete the diagnosis process[5].

Because the diagnosis of regional economic bottleneck focuses on the diagnosis of management problems, the fact in the diagnosis process is the uncertainty knowledge that people can not fully understand. The NF-GMDH method can not effectively solve the above problems. We propose a new algorithm: Fuzzy GMDH-Type method. This method can not only make full use of all kinds of information obtained in the diagnosis process, but also ensure the ability of the method to extract uncertainty rules.

Difference between Fuzzy GMDH-Type Method and GMDH Method

Fuzzy GMDH-Type method firstly fuzzifies the data, extracts the fuzzy rules, and then tests the rules as the rules for the diagnosis of regional economic bottlenecks through neural network. Fuzzy GMDH-Type method can extract the uncertain fuzzy rules. The concrete steps include: dividing the original data into training set and test set. The training set is used to estimate the weights of the neural network, and the test data is used to define the membership function and organize the structure of the neural network. After training, the parameters of the fuzzy model are obtained by using HPM method. According to the Balance of Variable Criteria, the reserved neurons are determined as the output of the intermediate neurons. When the criterion value is no longer reduced, the fuzzy model can be obtained. In the training set, n fuzzy rules can be created for regional economic bottlenecks, which can be used for reasoning.

In contrast, GMDH algorithm needs to construct GMDH input and output model first[10-14]. In self-organizing data mining, prior knowledge can be directly used to select reference functions and external criteria. Generally, self-organizing data mining uses general K-G polynomials as reference functions. However, when possessing prior knowledge of the system (domain expert knowledge), it can be directly used to construct specific reference functions reflecting system knowledge. According to the author's previous research, GMDH method has interaction among variables in the application of regional economic bottleneck diagnosis system: therefore, the model structure is as follows:

$$y = a_0 + f_1(x_1) + f_2(x_2, x_3) \tag{1}$$

Application of Fuzzy GMDH-TYPE Method in Regional Economic Diagnosis

Based on the regional economic data of China in 2003, the bottleneck model of regional economy was established by using GMDH method. The variables related to regional development bottlenecks were selected through software i.e. KnowledgeMiner.

[1] ROC (receiver operation charachateristic) refers to the operating characteristic curve of the subject, which can give consideration to both sensitivity and specificity to evaluate comprehensively the recognition performance of the classifier. As a quantitative index, the area under ROC curve can directly and effectively help to optimize classification thresholds and compare the performance of different classifiers.

The input variables are:

X1 - Illiteracy Rate X2 - Transaction Volume X3 in Technology Market - Number of Telephone Users X4 at the End of the Year - Number of Internet Users

X5-Railway Operation Mileage X6-Ratio of Grade Highway X7-Highway Mileage X8-Total Import and Export X9-Total Retail of Social Commodities X10-Government Management Ability X11-Relation between Government and **Enterprise X12-Regional Openness**

The input dependent variable is:

Y-Regional Competitiveness

Our reasoning method is that the coefficient of an independent variable is the largest, which shows that the independent variable is the bottleneck of the development of the region. By using Knowledge Miner software to screen the data, the optimal model is obtained.

$$Y = 0.0166216Z_{11} + 0.0185553Z_{12} + 0.0716154 \quad (2)$$

$$Z_{11} = -1.73994X_1 - 0.0617410X_{11} + 0.713224 \quad (3)$$

$$Z_{12} = 0.000286295X_5 - 0.0000144063X_7 + 0.152434 \quad (4)$$

Among them, the sum of squares of prediction error (PESS) is 0.7352, the average absolute percentage error is 17.56%, the approximate heteroscedasticity is 0.6424, and the determination coefficient is () 0.3576. The output variable is X13 (i.e. regional competitiveness). Relevant input variables are X1 (illiteracy rate), X11 (government-enterprise relationship), X5 (railway operating mileage) and X7 (road mileage). Their ability to eliminate model errors is 33%, 28%, 17% and 22%, respectively. The three optimal models of GMDH algorithm mining are illustrated as follows:

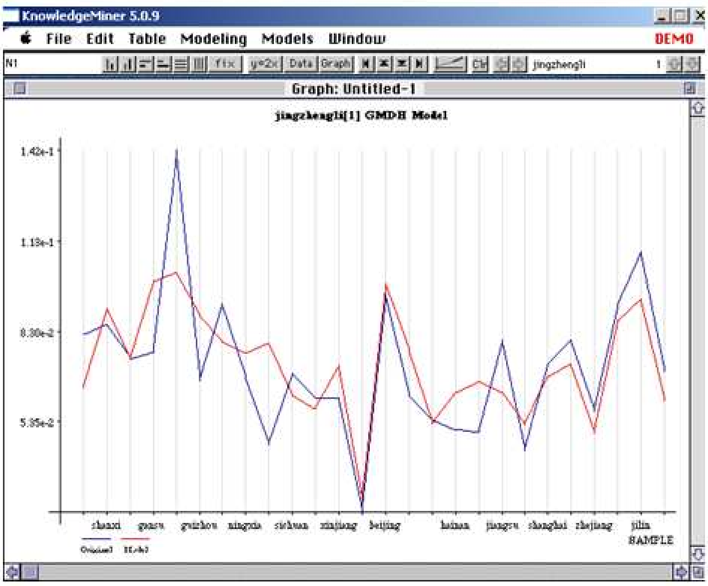

Figure 1: Model 1 Obtained by GMDH Algorithm.

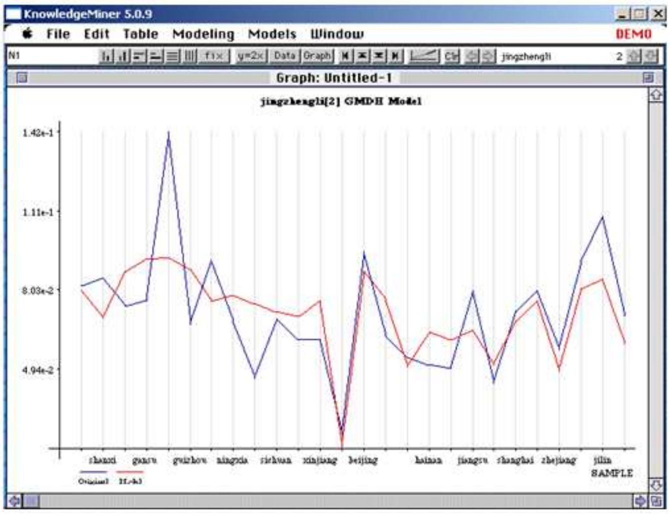

Figure 2: Model 2 Obtained by GMDH Algorithm.

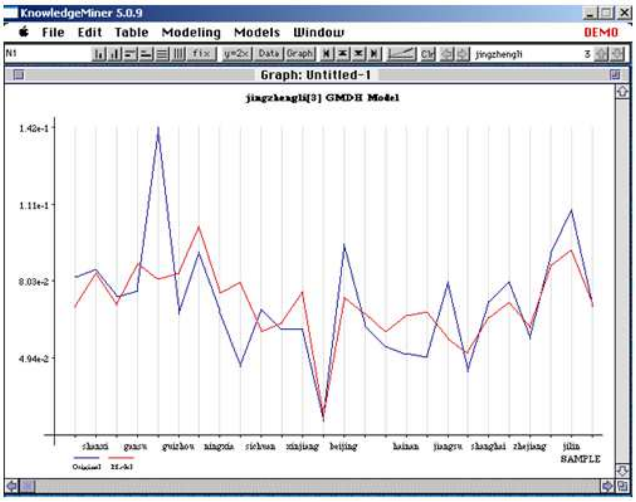

Figure 3: Model 3 Obtained by GMDH Algorithm.

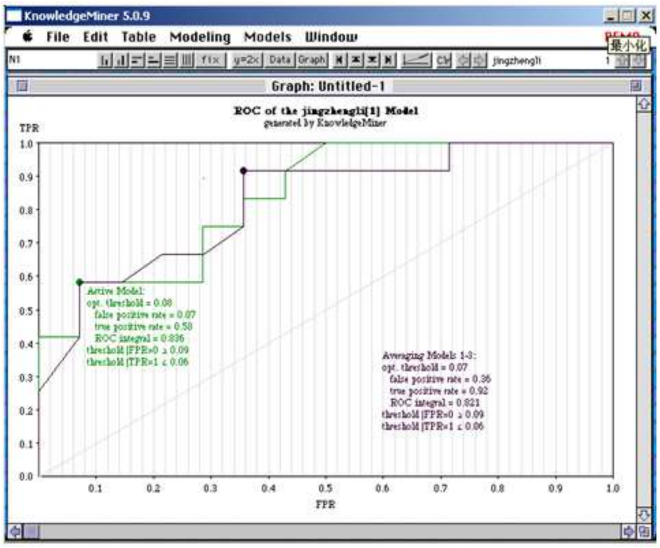

Figure 4: Roc of GMDH Algorithm.

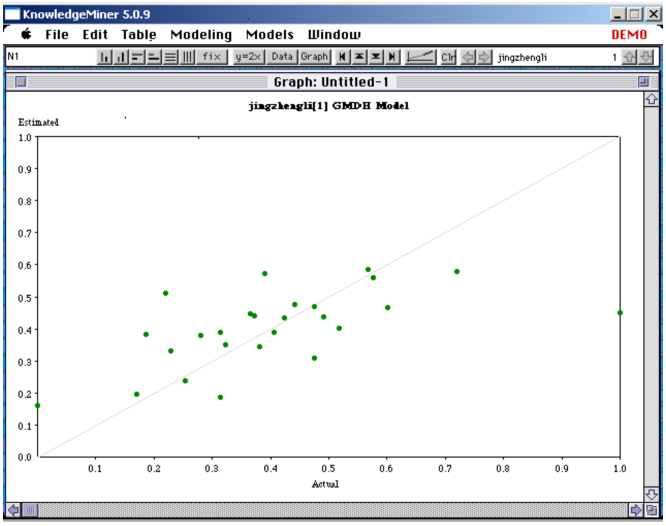

Figure 5: Model Scatter Plot of GMDH Algorithm.

From the model data and Fig. 1-3, it can be found that the model obtained by this algorithm can be basically used for further rule extraction. The rule is that if the number of railway mileage (high) is low, the regional competitiveness will be high (low). It shows that the railway mileage in 2003 is a major bottleneck of regional competitiveness.

Then the data from 2001 to 2005 are tested by artificial neural network. All data samples that can be interpreted by the above rules are deleted from the whole data set. The deleted data set is used as training data sample set to construct a diagnostic neural network model. Several diagnostic models can be constructed by repeatedly mining data of different years based on the above methods. After the construction of the diagnosis model is completed, the data set formed from 1995 to 2005 is tested. When the error of diagnosis results is less, than a certain threshold value, the data samples are deleted from the data set, which indicates that the fault information hidden in the deleted large amount of data can be expressed by the diagnosis model, while a small amount of data that has not been deleted can be stored in the data set after the noise is removed by the user interaction test in case base. This case knowledge expresses special bottleneck knowledge which is different from general diagnostic knowledge model. Binary tree organization should be carried out according to its abnormal characteristics in order to facilitate case retrieval for case-based bottleneck diagnosis.

In further work, the data processing function of KnowledgeMiner can be used to fuzzify the data. Using the command of creating a fuzzy input-output model, the following rules can be obtained after data processing in 2003. If the illiteracy rate is not high or low, the railway transportation mileage is high (low), the regional competitiveness is high (low). The total absolute error of the rule is 1.70, the average absolute error is 6.18%, and the approximate heteroscedasticity is 13.1815. The specific model diagram is shown in Figure 7-9.

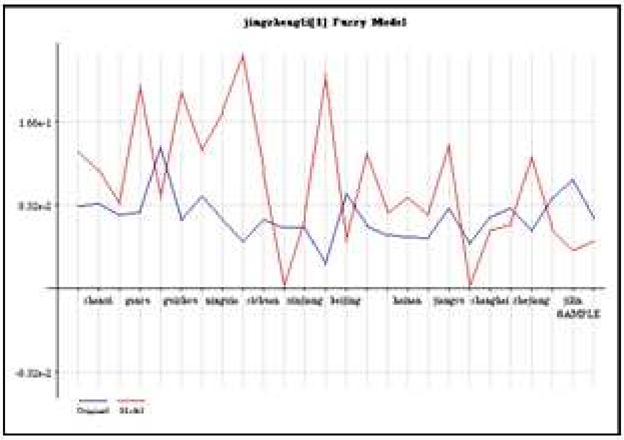

Figure 7: Optimal Model 1 Obtained by GMDH Algorithm.

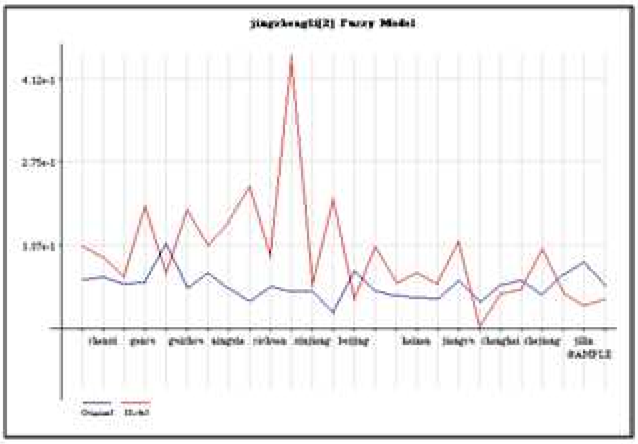

Figure 8: Optimal Model 2 Obtained by Fuzzy GMDH Algorithm.

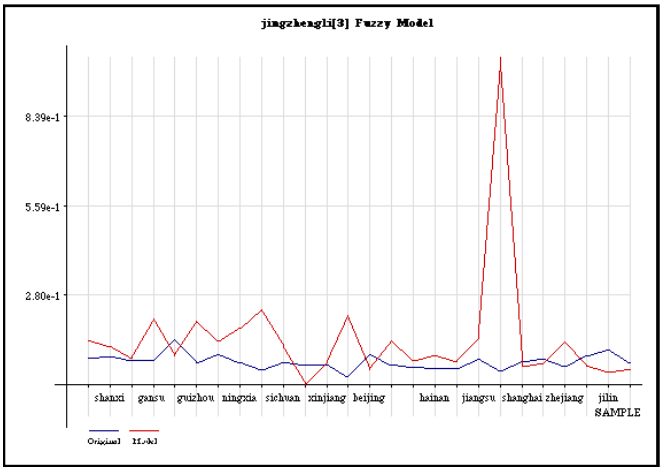

Figure 9: Optimal Model 3 Obtained by Fuzzy GMDH Algorithm.

Similarly, artificial neural network is used to test the data from 2001 to 2005, and several diagnostic models are constructed. The other steps are the same.

CONCLUSIONS

In this paper, a new self-organizing data mining algorithm, Fuzzy GMDH-Type method, is proposed and successfully applied to the diagnosis of regional economic bottlenecks. This method solves the problem of identifying regional economic bottlenecks. By comparing with the classical GMDH algorithm, it is found that the recognition accuracy of this method is high and new preconditions are added.

This study provides a basic operation scheme for the diagnosis of regional economic bottlenecks. This scheme will combine the subsequent work with the reverse reasoning model to form a set of diagnostic procedures for the diagnosis of regional economic bottlenecks, thus laying a relatively solid foundation for the scientific evaluation, monitoring and diagnosis of regional economy.

REFERENCES

1. *Yang jie，Ye Chen Zhou，Huang Xin，Chen NianYi. Data Mining Used in Modeling, Optimization and Fault Diagnosis. COMPUTER INTEGRATED MANUFACTURING SYSTEMS. 2000,10:72-76,81*
2. *T. KONDO. Logistic GMDH - type neural networks and their app lication to the identification of the X - ray filmcharacteristic curve [A]. Proceedings of IEEE Interna tional Conference on Systems, Man and Cybernetics[C].[s. l.]: [s. n.], 1999.*
3. *Narimanzadeh N, Atashkari K, Jamali A, et al. Inverse modelling of multi-objective thermodynamically optimized turbojet engines using GMDH-type neural networks and evolutionary algorithms[J]. Engineering Optimization, 2005, 37(5):437-462.*
4. *Chao P Y, Ferreira P M, Liu C R. Applications of GMDH-type modeling in manufacturing[J]. Journal of Manufacturing Systems, 1988, 7(3):241-253.*
5. *Wang HongYuan，Shi GuoDong，Fu YanWei，Xia DeShen. Data Mining Technique and its Application in Fault Diagnosis.*
6. *He ChangZheng,Zhang Bin,Yu Hai. A Comparative Study of Self-organizing Data Mining and Artificial Neural Networks. System Engineering Theory and Practice.2002, 11: 11-14,50*

JOURNAL OF JIANGSU INSTITUTE OF PETROCHEMICAL TECHNOLOGY.2001, 13(4):42-44

7. *Narimanzadeh N, Darvizeh A, Ahmadzadeh G R. Hybrid genetic design of GMDH-type neural networks using singular value decomposition for modelling and prediction of the explosive cutting process[J]. Proceedings of the Institution of Mechanical Engineers Part B Journal of Engineering Manufacture, 2003, 217(6):779-790.*
8. *Molaabasi H, Khajeh A, Semsani S N, et al. Prediction of Zeolite-Cemented Sand Tensile Strength by GMDH type Neural Network[J]. 2019(12):1-15.*
9. *Eidgahee D R, Rafiean A H, Haddad A. A Novel Formulation for the Compressive Strength of IBP-Based Geopolymer Stabilized Clayey Soils Using ANN and GMDH-NN Approaches[J]. Iranian Journal of Science and Technology, Transactions of Civil Engineering, 2019:1-11.*

10. *Narimanzadeh N, Atashkari K, Jamali A, et al. Inverse modelling of multi-objective thermodynamically optimized turbojet engines using GMDH-type neural networks and evolutionary algorithms[J]. Engineering Optimization, 2005, 37(5):437-462.*

11. *Nariman-Zadeh N, Haghgoo E, Jamali A. Trade-offs in optimization of GMDH-type neural networks for modelling of a complex process[C]// Wseas International Conference on Systems Theory & Scientific Computation. 2006.*

12. *Khalkhali A, Safikhani H. Pareto based multi-objective optimization of a cyclone vortex finder using CFD, GMDH type neural networks and genetic algorithms[J]. Engineering Optimization, 2012, 44(1):105-118.*

13. *Ahmadi B, Nariman-Zadeh N, Jamali A. Path synthesis of four-bar mechanisms using synergy of polynomial neural network and Stackelberg game theory[J]. Engineering Optimization, 2016, 49(6):932-947.*

14. *Sum C C, Lee Y S, Hays J M, et al. Modeling the Effects of a Service Guarantee on Perceived Service Quality Using Alternating Conditional Expectations (ACE)[J]. Decision Sciences, 2010, 33(3):347-384.*

15. *PetrBuryan, Onwubolu G C. Design of enhanced MIA-GMDH learning networks[J]. International Journal of Systems Science, 2011, 42(4):673-693.*

16. *Nourbakhsh A, Safikhani H, Derakhshan S. The comparison of multi-objective particle swarm optimization and NSGA II algorithm: applications in centrifugal pumps[J]. Engineering Optimization, 2011, 43(10):1095-1113.*

APPENDIX

Regional Economic Development of 26 Provinces (Municipalities) randomly selected in 2003

	wenmangrk	jishusccje	dianhuayhs	hulian wswrs	tieluyylc	dengjigl	gululc	jinchuke	xiaofei lsze	zhengfu xiaolv	zhengqi guanxi	kaifang chengd u	jingzh engli
anhui	0.13655	87960	997.9	183.5	2219.7	63374	69560	594781	1331.2	0.13	0.01	5.89	0.082
shanxi	0.119095	168022	671.9	148.8	2892.3	44422	50019	278262	853.2	0.08	13.16	4.59	0.085
chongqing	0.083985	555083	533.4	176.6	718.2	22562	31407	259476	835.5	-0.2	4.13	6.35	0.074
gansu	0.203303	77581	399.8	122.4	2312.5	30947	40293	132714	474.6	-0.16	1.65	3.94	0.076
guangxi	0.08853	41808	638.5	228.6	2738	45284	58451	318675	857.7	-0.23	5.36	5.76	0.142
guizhou	0.196827	17892	332.4	83.1	1900.1	32352	45304	98433	458.8	-0.26	1.82	3.85	0.067
neimenggu	0.136711	108452	422.7	74.9	6202.6	65157	74135	282902	726.8	-0.003	5.18	4.64	0.092
ningxia	0.175595	10047	100.3	33.3	791.4	11770	11916	65323	120.8	-0.15	5.96	3.61	0.068
qinghai	0.234696	8291	76.4	19.5	1091.8	21568	24377	33914	102.7	0.05	9.29	3.73	0.046
sichuan	0.117342	128686	1128.2	424.3	2961.8	75290	112543	563429	2091.1	-0.13	3.98	6.03	0.069
xizang	0.548521	0	27.4	8.6	0	9107	41302	15986	58.3	-0.37	13	2.05	0.061
xinjiang	0.069403	120395	412	117.8	2773.3	64029	83633	476986	421.8	-0.19	6.46	4.12	0.061
yunnan	0.214985	228718	483.2	166.4	2340.3	109300	166133	266913	782.5	-0.12	5.05	4.88	0.024
beijing	0.046094	2653574	683.7	398	1136.1	14139	14453	6850017	1916.7	0.38	8	7.54	0.095
fujian	0.135511	166779	1122.7	318.2	1453.9	42222	54876	3532553	1740.4	0.07	9.66	8.67	0.062
guangdong	0.075469	805730	2567	950.2	2112.5	99733	110253	2.84E+07	5606	0.18	13.62	9.74	0.054
hainan	0.091105	11978	162.8	39.7	221.7	11894	20877	227492	191.6	0.06	15.78	6.07	0.051
hebei	0.073525	67969	1339	289.1	4744	55682	65391	897825	2177.9	0.15	7.26	6.37	0.05
jiangsu	0.144561	765163	2043.9	610.9	1393.6	56300	65565	1.14E+07	3566.5	0.38	11.31	8.13	0.08
shandong	1.0981	525682	2088.8	626.6	3150.5	76170	76266	4463682	3936.5	-0.02	3.5	7.1	0.044
shanghai	0.058787	1427790	733.9	431.6	256.5	6322	6484	1.12E+07	2220.6	0.31	14	8.54	0.072
tianjin	0.063624	420008	360.1	144.6	666.3	9901	10168	2934244	922.3	0.12	5.77	7.02	0.08
zhejiang	0.132247	530353	1656.5	451.2	1249.9	43436	46193	6141081	3157.1	0.31	7.71	9.1	0.057
heilongjiang	0.057846	121165	874.4	226	5483.7	59599	65123	532940	1376.5	0.08	3.06	4.98	0.091
jilin	0.038909	87292	594.6	146.5	3561.8	41362	43779	614841	1110.3	-0.01	0	5.14	0.109
liaoning	0.047403	620200	1278.7	291.5	4173.9	49845	50095	2650917	2330.8	0.1	1.95	6.61	0.07

* The indicators of this paper are the research results of the author in the key projects of the National Social Science Fund, and the regional competitiveness score is the research results of the Liaoning Province Department of Education under the auspices of the author.