CORE Metadata, citation and similar papers at core.ac.uk
Provided by International Institute for Science, Technology and Education (IISTE): E-Journals

Developing Country Studies www.iiste.org
ISSN 2224-607X (Paper) ISSN 2225-0565 (Online)
Vol.6, No.5, 2016

# Exploring the Technology Input and Economy Output in Chinese National Innovation Demonstration Zone Based on Rough Set Theory

Xiangze Xiao[1*] Nan Zhou[2]
1.School of Public Affairs, University of Science and Technology of China, No.96 Jin Zhai Road, Hefei 230026, China
2.School of Law and Politics, Tianjin University of Science & Technology, No.1038 Dagu Nan Road, Tianjin 300222, China

**Abstract**
The previous study about the relationship between technology input and economy output was mainly concentrated on their linear or functional formulation, while little on the data independencies between them. This study explored the data independencies between technology input and economy output of Chinese National Innovation Demonstration Zone based on Rough Sets Theory, for the purpose of conducting a new way to understand the unstructured relation between technology input and economy output, as well as to promoting the effective combination of technology and economy output of Chinese National Innovation Demonstration Zone, which was the most important part of national innovation system of China. The Rough Set Theory was applied to analyze the 8 Chinese National Innovation Demonstration Zone's technology input and economy output data from 2007 to 2014. The result demonstrated that: (1) of the economy output indicators, ratio of technical income to total income, ratio of net profit to total income and export were not combined effectively with the technology input, while total income, technical income, net profit and taxes submitted had been combined with the technology input very significantly; (2) of the technology input indicators, all of them had shown the linkage with economy output indicators significantly, and expenditure on R&D activities was the most important one; (3) an two factor theory effect might existed between the technology input and economy output, senior and middle level professional qualifications, personal engaged in R&D activities and expenditure on R&D activities were the hygiene factors, ratio of expenditure on R&D activities to personal engaged in R&D activities and ratio of expenditure on R&D activities to total income were motivation factors.
**Keywords:** Chinese National Innovation Demonstration Zone, Technology input indicators, Economy output indicators, Combination effectiveness, Rough Set Theory

## 1. Introduction

Chinese National Innovation Demonstration Zone (CNIDZ), which were selected from the High-tech Zone, are settled to promote self-innovation, economy development (Xiong Xi, et.al., 2016), and the effective combination of technology and economy of China. The CNDIZ administrators were authorized to carry out a series of pilot experiment science & technology policy to achieve these goal. Until the end of 2014, China has settled 8 CNDIZs, which were Zhongguancun in Beijing (ZGC), Zhangjiang (ZJ) in Shanghai, Donghu (DH) in Wuhan, Shenzhen (SZ) in Guangdong, Sunan (SN) in Jiangsu, Changzhutan (CZT) in Hunan, Hewubeng (HWB) in Anhui and Binhai (BH) in Tianjin (Zhou Hongyu, 2015). As most research of CNIDZ were normative and focused on the selection standard and pilot experiment policy, which brought a lack of empirical study, especially the evaluation of combination effectiveness of technology and economy of 8 CNDIZs, which were the performance of those pilot experiment science &technology policy. Therefore, we had the question which would be explored in this paper.

In this paper, the rough set theory (RST) was applied to try to research the combination effectiveness of technology and economy of 8 CNIDZs, based on the data of their technology input indicators and economy output indicators from 2007 to 2014. The questions would be answered preliminary, which was the combination effectiveness between the technology input indicators (including technology talent input indicators and technology expenditure input indicators) and the economy output indicators (including economy benefit output indicators, technology benefit output indicators and society benefit output indicators) of the 8 CNIDZs.

The remainder of this paper was organized as follows: Sect. 2 was the literature review, Sect. 3 briefly reviews the used method—Rough Set Theory, the data collection and pre-processing, Sect. 4 used the empirical data by RST to analyze the combination effectiveness of technology input indicators and economy output indicators, Sect. 5 provided the conclusion of this paper.

## 2. Literature review

There were still not many existed research papers about the CNIDZ, and most of them were conducted about the following four issues. The first was how to select the CNIDZ from the High-tech Zone (XIE Jialong, et.al., 2013a; XIE Jialong, et.al., 2013b). The second was how to set up the operating framework of CNIDZ (Hu Shuhua, et.al., 2011). The third was about the pilot experiment science &technology policy of CNIDZ, such as the incentive policy of equity and dividends (Guo Rong, et.al., 2013), the S&T policy and financial policy (Dai Lijuan, et.al. 2013), the incentive policy of industrialization of S&T achievements (Xu Wangqiang, et.al., 2012) and how to make full use of these policy effectively (Wu Ke, et.al., 2012). The fourth was about the technology transfer (Jiang Pengju, et.al., 2014) and policy transfer (Huang Y, et.al., 2013) of CNIDZ. Summarizing all these papers above, we could see that there was still little empirical study.

As the CNIDZs were selected from the High-tech zone, we also have reviewed the research papers about High-tech zone, which were conducted about the relationship between technology input and economy output, and most of these papers could be classified into two categories. The first was multi-attributes evaluation analysis (Hu Shuhua, et.al., 2012; Xie Jialong, et.al., 2012), in which the indicators were not classified as variable (cause) or independent variable (effect). The second was cause & effect analysis, in which every indicator was classified as variable (cause) or independent variable (effect), and there were different kinds of methods to explorer these causal relationship, such as data envelopment analysis (DEA) (Liu Hua, et.al., 2011; Xie Ziyuan, 2011), stochastic frontier analysis (SFA) (Wu Peng, et.al., 2010; Zhou Jiao, et.al., 2014) and structural equation model (SEM) (Ou Guangjun, et.al., 2013; Wu Shu'e, et.al., 2012) and so on. We thought that, to explorer the relationship between technology input and economy output, cause & effect analysis was better than multi-attributes evaluation analysis, as the former could distinguish technology input (cause) and economy output (effect). Although cause & effect analysis was better, there was some assumptions in both DEA and SEM methods, and linear relationship between cause and effect was one of them. The linear relationship assumption has been criticized by some research (Balconi M, et.al., 2010) and there was some nonlinear relationship explorer normative research (Feng Feng, et.al., 2013) and empirical research (Gao X, et.al., 2009; Gao X, et.al., 2010), which were different to the linear assumption. On the other way, in SFA model (Han Jing,2010; Tang Dexiang, et.al., 2008), the relationship between variables and independent variables were nonlinear, too. Thus, we thought that nonlinear relationship assumption was better than linear relationship assumption, as in the former, we could find more kinds of differences, which including the linear relationship, also. And then we found that there were different kinds of nonlinear relationship assumption, and which kind should we chosen?

We thought that maybe we should rethink what did those nonlinear relationship assumptions mean, as well as the linear relationship assumption. And we considered it as a kind of dependency relationship, which might be different in different research, and should be investigated one by one. So we thought that rough set theory could be an appropriate approach to explorer this question and found out the dependency relationship. RST was proposed by Pawlak in 1982(Pawlak Z, 1982) and it was an effective technique for discovery of data dependencies (Pawlak Z, et.al., 1995) and it has been comprehensive explained and reviews by Slowinski in 1992(Slowinski R, 1992). RST has been applied in many kinds of questions, such as manager competency (Wu W W, et.al., 2005), personal investment portfolios (Shyng J Y, et.al., 2010), as well as R&D innovation performance (Wang C H, et.al., 2000), which was similar and related to our research. In RST, each attribute was classified as condition attribute (cause) and decision attribute (effect), and this technique would attempt to induce, extract and measure decision rules related the condition attributes and decision attributes. And in our research, the technology input indicators of the CNIDZs would be condition attributes, the economy output indicators would be decision attributes, the decision rules related them would measure the combination effectiveness between technology input and economy output of CNIDZs. It was necessary to transform different types of original data to nominal data, these transformations were mostly conducted by similarity relation (Slowinski R, et.al., 2000; Vanderpooten D, 1997), which we thought that the univariate hierarchical clustering approach might be more appropriate, as the latter could handle different categories by different scale of distances, while remaining the follow of approximation rules.

## 3. Method and data

*3.1 Rough Set Theory*

Rough set theory could be described as an information system, which could be represented by $S = <U, Q, V, f>$, where $U$ denoted the set of objects (such as cases, observations), which was finite and non-empty, $Q$ denoted the set of attributes(such as variables, features), which describing the objects, it could be divided into two subsets $C$ and $D$, where $C$ denoted the set of condition attributes and $D$ denotes the set of decision attributes, and $Q = C \bigcup D$, $C \bigcap D = \varnothing$, $V = \bigcup_{\alpha \in Q} V_\alpha$, in which $V_\alpha$ denoted the set of value of attribute $\alpha$, and $f : U \times Q \to V$ was an function, in which for $\forall x \in U$ and $\forall \alpha \in Q$, we had $f(x,a) \in V_a$.

An important premise of RST was that the ability to classify meant knowledge, and the indiscernibility relation would be unfavorable to make the classification. For $\forall P \subseteq Q (P \neq \varnothing)$, the indiscernibility relation was defined as $I_P = \{(x,y) \in U \times U : f(x,q) = f(y,q), \forall q \in P\}$. As we could see, the $I_P$ was a partition of $U$ by $P$. Based on the indiscernibility relation, we could define the upper approximation and low approximation. The low approximation $\underline{P}(X)$ was defined as the set of objects in which all of its elements could be certainly classified as belonging to class $X$ by the set of attributes $P$, $\underline{P}(X) = \{x \in U : I_P(x) \subseteq X, X \subseteq U\}$. And the upper approximation $\bar{P}(\mathrm{X})$ defined as the set of objects in which any of its elements could be possibly classified as belonging to $X$ by $P$, $\bar{P}(X) = \bigcup_{x \in X} I_P(x), X \subseteq U$. The boundary of $X$ by $P$ was defined as $B_{nP}(X) = \bar{P}(X) - \underline{P}(X)$, which contained the objects that couldn't be certainly classified as belonging to $X$ by $P$. So we could estimate the accuracy of classification of $X$ by $P$ using $\alpha_P(\mathrm{X}) = card(\underline{P}(X)) / card(\bar{P}(X))$, in which $card(\mathrm{X})$ denoted the count of elements in set $\mathrm{X}$ and if $\alpha_P(\mathrm{X})$ =1, $\mathrm{X}$ is definable in $U$ by $P$, if $\alpha_P(\mathrm{X}) < 1$, we said it was undefinable by $P$, and a rough set .

Given $A \subseteq Q$, if another attributes sets $B \subset A$ and $I(B) = I(A)$, and there was no subset of $B$ that $C \subset B$ and $I(C) = I(B)$, we said that $B$ was an reduct for $A$, which meant the latter could be redundant to former without losing essential classificatory information and the latter could be definable with the minimum number of attributes by the former. Let the set of all reducts for $A$ be denoted as $Red(A)$, and the core of $A$ was the intersection of them, which meant $Core(A) = \bigcap Red(A)$, which was the set of the most important attributes.

As the attributes set $Q$ was divided as the condition attributes set $C$ and decision attributes set $D$, we could got the decision rules, just like this kind of formation $if\ a_1 = v_1, a_2 = v_2, ..., a_n = v_n, then\ d = d_k$, in which $Q = \{a_1, a_1, ...a_n\}$, $D = \{d\}$. As in this kind of decision rules, there were maybe some reduct condition attributes, it was necessary to do the reduction, which means the procedure of derivation and captures of decision rules. An induced decision rule could be expressed as $if\ a_1^{'} = v_1^{'}, a_2^{'} = v_2^{'}, ..., a_l^{'} = v_l^{'}, then\ d = d_k^{'}$, $\{a_1^{'}, a_2^{'}, ..., a_l^{'}\} \subseteq \{a_1, a_2, ..., a_n\}$. The decision rules reflected a relationship between a set of condition attributes and a decision attribute, and it could be extracted based on upper and low approximation from the former kind of decision rules. The reduction could be conducted by Boolean reasoning, and the extracted decision rules could be vividly expressed as the decision flow graph. As the reduction computation was very complex, it was managed usually by some software, such as Rosetta.

In the decision rule $if\ \alpha, then\ \beta$, we could also rewrite it as $\alpha \to \beta$. Usually, We used $support(\alpha)$ denoted the count of objects which were within $\alpha$, and $support(\alpha \wedge \beta)$ denoted the count of objects which were within $\alpha$ and $\beta$ simultaneously. The accuracy of the decision rules $\alpha \to \beta$, which meant the possibility if one object was within $\alpha$, it was within $\beta$ also, could be denoted as $accuracy(\alpha \to \beta) = support(\alpha \wedge \beta) / support(\alpha)$. The coverage of the decision rules $\alpha \to \beta$, which meant the possibility if one object was within $\beta$, it was within $\alpha$ also, denoted as $coverage(\alpha \to \beta) = support(\alpha \wedge \beta) / support(\beta)$.

*3.2 Data collection and pre-processing*

In this paper, we have collected the data of technology input indicators and economy output indicators of 8 CNIDZs (Zhongguancun in Beijing (ZGC), Zhangjiang (ZJ) in Shanghai, Donghu (DH) in Wuhan, Shenzhen (SZ) in Guangdong, Sunan (SN) in Jiangsu, Changzhutan (CZT) in Hunan, Hewubeng (HWB) in Anhui and Binhai (BH) in Tianjin) from 2007 to 2014. We had chosen 5 condition attributes and 7 decision attributes.

The 5 condition attributes were senior and middle level professional qualifications ($a_1$), personal engaged in R&D activities ($a_2$), expenditure on R&D activities ($a_3$), ratio of expenditure on R&D activities to personal engaged in R&D activities ($a_4$), ratio of expenditure on R&D activities to total income ($a_5$). $a_1$ and $a_2$ were the technology talent input indicators, as they were both conducted with personal. $a_3$, $a_4$, $a_5$ were the technology expenditure input indicators, as they were all conducted with expenditure.

The 7 decision attributes were total income ($d_1$), technical income ($d_2$), ratio of technical income to total income ($d_3$), net profit ($d_4$), ratio of net profit to total income ($d_5$), taxes submitted ($d_6$) and export($d_7$). $d_1$, $d_4$, $d_5$ were the economy benefit output indicators, as they were conducted with total income and net profit. $d_2$, $d_3$ were the technology benefit output indicators, as they were both conducted with technical income. $d_6$, $d_7$ were the society benefit output indicators, as they were conducted with taxes and export, which belonged to the whole society .

All the data was from the China Torch Statistical Yearbook from 2008 to 2015. And the table 1 was the descriptive statistics of the condition attributes and decision attributes.

To apply the data to RST, we needed do the data pre-processing firstly. To get the normal data from the interval data, the univariate hierarchical clustering approach was applied to 5 condition attributes and 7 decision attributes one by one. And the results were given by table 2. From it, we could see that $a_2$, $a_3$, $a_4$ and $d_7$ were clustered to 3 sub-intervals and the other attributes (including other condition attributes and decision attributes) were clustered to 2 sub-intervals. The detailed sub-intervals of condition attributes and decision attributes were given by the table 2.

## 4. Results analysis

As the causal relationship from condition attributes to decision attributes was not conducted immediately, the decision attributes had time delay compared with condition attributes in the same objects (meant one CHIDZ at one year). Making the input-output analysis as references, we also considered the time delay as one year, which meant, for an example, the condition attributes data in 2009 would be corresponded to the decision attributes in 2010. After this processing, the count of objects was changed from 64 to 56, 8 CNIDZs and 7 years (from year 2008 to year 2014).

Then, we used Rosetta (version 1.4.41) to do the RST computation, and the decision rules were extracted from it. As we had 7 decision attributes, we needed to do the RST computation for seven times, and got the decision rules for every decision attribute one by one. To avoid the contingency data, we had ignored the decision rules which had only one support. Also, in this study, these decision rules, such as $\alpha \rightarrow \beta$, in which there were uncertain values of $\beta$, were also ignored, as they were not certain decision rules.

When the decision attribute was total income ($d_1$), the decision rules was given by the table 3, and the decision flow graph was given by the Fig. 1.1 in Fig.1.There were 3 reducts, $\{a_1, a_3, a_5\}$, $\{a_2, a_3, a_5\}$ and $\{a_3, a_5\}$, and the core was the intersection of these reducts, which was $\{a_3, a_5\}$. The support of $d_1 = 1$ was 48, the support of $d_1 = 2$ was 2, and the total suppose of $d_1$ was 50. From table 3 and fig. 1.1, we could find that: (1) if the senior and middle level professional qualifications was in the low level, or the personal engaged in R&D activities and expenditure on R&D activities were both in the low level, the total income would be in the low level. (2) while the senior and middle level professional qualifications were not in the low level, or the personal engaged in R&D activities and expenditure on R&D activities was not in the low level, the total income would be maybe in the low level still. (3) if the total income was in the high level, the expenditure on R&D activities should be in the high level and the ratio of expenditure on R&D activities to total income should be in the low level at the same time, and the reverse was true, too.

When the decision attribute was technical income ($d_2$), the decision rules was given by the table 4, and the decision flow graph was given by the Fig. 1.2 in Fig.1 There was only one reducts $\{a_2, a_3, a_4, a_5\}$, and it was the core also. The support of $d_2 = 1$ was 36, the support of $d_2 = 2$ was 4, and the total suppose of $d_2$ was 40. From table 4 and Fig. 1.2, we could find that: (1) if technical income was in the low level, the senior and middle level professional qualifications and the personal engaged in R&D activities should be in the low level; if technique income was in the high level, the senior and middle level professional qualifications and the personal engaged in R&D activities should be in the high level; and when the personal engaged in R&D activities was in the medium level, the technical income could be in the low level, or the high level. (2) if technical income was in the high level, the ratio of expenditure on R&D activities to personal engaged in R&D activities should be in the low level; while when the technical income was in the low level, the ratio of expenditure on R&D activities to personal engaged in R&D activities could be in the low level, or the medium level, or the high level. (3) if technical income was in the low level, the ratio of expenditure on R&D activities to total income should be in the low level; while when technical income was in the high level, the ratio of expenditure on R&D activities to total income could be in the low level, or the high level.

When the decision attribute was ratio of technical income to total income ($d_3$), the decision rules was given by the table 5, and the decision flow graph was given by the Fig. 1.3 in Fig.1. There was only one reducts $\{a_2, a_3, a_4, a_5\}$, and it was the core also. The support of $d_3 = 1$ was 0, the support of $d_3 = 2$ was 7, and the total suppose of $d_3$ was 7. As the percentage of total support to total objects was too low (7/56=12.5%), we thought that ratio of technical income to total income was not an important decision attribute to the decide system, and would do no more detailed analysis.

When the decision attribute was net profit ($d_4$), the decision rules was given by the table 5, and the decision flow graph was given by the Fig. 1.4 in Fig.1 There were 3 reduct, $\{a_1, a_2, a_3, a_4\}, \{a_1, a_3, a_4, a_5\}, \{a_2, a_3, a_4, a_5\}$ and the core was $\{a_3, a_4\}$. The support of $d_4 = 1$ was 44, the support of $d_4 = 2$ was 2, and the total suppose of $d_4$ was 46. From the table 6 and Fig. 1.4, we could find that: (1) if net profit was in the low level, the senior and middle level professional qualifications should be in the low level, or the medium level; if net profit was in the high level, the senior and middle level professional qualifications should be in the medium level, or the high level. (2) if net profit was in the low level, the personal engaged in R&D activities should be in the low or medium level, and the expenditure on R&D activities should be in the low or medium level, too; if net profit was in the high level, the personal engaged in R&D activities should be in the high level, and the expenditure on R&D activities should be in the high level, too. (3) if net profit was in the high level, ratio of expenditure on R&D activities to personal engaged in R&D activities should be in the low level, and ratio of expenditure on R&D activities to total income should be in the low level, too; if net profit was in the low level, ratio of expenditure on R&D activities to personal engaged in R&D activities could be in the low level, or the medium level, or the high level, and ratio of expenditure on R&D activities to total income should be in the low level, or the high level.

When the decision attribute was ratio of net profit to total income ($d_5$), the decision rules was given by the table 7, and the decision flow graph was given by the Fig.1.5 in Fig.1. There was only one reducts $\{a_2, a_3, a_4, a_5\}$, and it was the core also. The support of $d_5 = 1$ was 0, the support of $d_5 = 2$ was 12, and the total suppose of $d_5$ was 12. As the percentage of total support to total objects was too low (12/56=21.43%), we thought that ratio of net profit to total income was not an important decision attribute and would do no more detailed analysis.

When the decision attribute was taxes submitted ($d_6$), the decision rules was given by the table 8, and the decision flow graph was given by the Fig. 1.6 in Fig.1. we could found that the decision rules of taxes submitted ($d_6$) were as same as the decide rules of net profit ($d_4$), and we would not repeat the analysis, as they were all the same.

When the decision attribute was export ($d_7$), the decision rules was given by the table 9, and the decision flow graph was given by the Fig.1.7 in Fig.1. There was 2 reducts, $\{a_1, a_3, a_4, a_5\}$ and $\{a_2, a_3, a_4, a_5\}$, and the core was $\{a_3, a_4, a_5\}$. The support of $d_7 = 1$ was 9, the support of $d_7 = 2$ was 2, the support of $d_7 = 3$ was 2, and the total suppose of $d_7$ was 13. As the percentage of total support to total objects was too low (13/56=23.21%), we thought that export was not an important decision attribute and would do no more detailed analysis.

Summary above, we could come to the conclusion that: (1) to the 7 decision attributes, ratio of technical income to total income, ratio of net profit to total income and export were not as important as other decision attributes, as their decision rules' total support were too low, total income, technical income, net profit and taxes submitted were the important decision attributes, and the decision rules of net profit were as same as the decision rules of taxes submitted.(2) to the 5 condition attributes, all of them had been the reducts' element, $\{a_3\}$ was the intersection of all the core of each decision attribute, which meant it was the most important condition attribute. (3) when the decision attribute was total income, or technical income, or net profit, or export, if senior and middle level professional qualifications, personal engaged in R&D activities and expenditure on R&D activities were in the low or medium level, the decision attribute mostly was in the low level; however if senior and middle level professional qualifications, personal engaged in R&D activities and expenditure on R&D activities were in the medium or high level, the decision attribute could not be in the high level, it still needed the ratio of expenditure on R&D activities to personal engaged in R&D activities in the medium or high level and the ratio of expenditure on R&D activities to total income in the high level. This could be seen as an transformed effect of the two factor theory, senior and middle level professional qualifications, personal engaged in R&D activities and expenditure on R&D activities were the hygiene factors, if they were in the low level, the decision attribute would be in the low level, ratio of expenditure on R&D activities to personal engaged in R&D activities and the ratio of expenditure on R&D activities to total income were motivation factor, only when they reach the medium or high level, the decision attribute could reach the high level.

## 5. Conclusion

In this paper, we had used the rough set theory to explore the combination effectiveness of technology input and economy output of 8 Chinese National Innovation Demonstration Zone, based on the data of their technology input indicators and economy output indicators from 2007 to 2014. From the research, we could got the conclusions as follows: (1) of the economy output indicators, ratio of technical income to total income, ratio of net profit to total income and export were not combined effectively with the technology input, while total income, technical income, net profit and taxes submitted had been combined with the technology very significantly.(2) of the technology input indicators, all of them had shown the linkage with economy output indicators significantly, and expenditure on R&D activities was the most important one. (3) There was maybe a transformed effect of the two factor theory between the technology input and economy output, senior and middle level professional qualifications, personal engaged in R&D activities and expenditure on R&D activities were the hygiene factors, and ratio of expenditure on R&D activities to personal engaged in R&D activities and the ratio of expenditure on R&D activities to total income were motivation factor.

This study helped to drawn attention on rule-based decision-making in analyzing the combination effectiveness between technology input and economy output, and had some new findings, which would be favorable for a better understanding of this issue, especially the two factor theory effect. As an empirical study, this paper surely had the space to be improved further, and which could concentrate on the two factor theory effect, to find more support and findings. The following research may use DRSA (dominance-based rough set approach) to explore the detail information about the two factor theory effect between technology input and economy output in Chinese National Innovation Demonstration Zone, on the other hand, based on the results of rough set theory, it may be feasible to define the weights of technology input indicators and economy output indicators, and make the integrated evaluation of innovation ability of Chinese National Innovation Demonstration Zone.

## References

Balconi M, Brusoni S, Orsenigo L. (2010). In defence of the linear model: An essay. *Research Policy*, 39(1): 1-13.

Dai Lijuan, Zhang Yi (2013). Research on science and technology financial innovation of Wuhan East Lake national innovation demonstration zone. *Science and Technology Management Research*, 33(6): 4-8.

Feng Feng, Xiao Xiangze, Zhang Leiyong. (2013). Study on the classification and network construction of symbiosis in Industry-University-Research cooperation. *Science of Science and Management of S&T*, 34(2): 3-11.

Gao X, Guan J. (2009). A scale-independent analysis of the performance of the Chinese innovation system. *Journal of Informetrics*, 2009, 3(4): 321-331.

Gao X, Guo X, Sylvan K J, et al. (2010). The Chinese innovation system during economic transition: A scale-independent view. *Journal of Informetrics*, 2010, 4(4): 618-628.

Guo Rong, Xue Wei, et al (2013). Evaluation of the Innovation Policy for National Innovation Demonstration Zone. *Forum on Chinese Science and Technology*, 11: 11-15.

Han Jing. (2010). An empirical analysis on China's High-technology industry innovation efficiency based on SFA. *Studies in Science of Science*, 28(3): 467-472.

Hu Shuhua, Wang Song, Deng Hengjin. (2011). Study on the development of national independent innovation demonstration park based on Four-Three structure model. *Science & Technology Progress and Policy*,28(9): 29-32.

Hu Shuhua, Xie Jia-long, et al. (2012). Research on the competitiveness of national high-tech zones based on multi-step fuzzy comprehensive evaluation. *Soft Science*, 25(5): 53-56.

Huang Y, Audretsch D B, Hewitt M. (2013). Chinese technology transfer policy: the case of the national independent innovation demonstration zone of East Lake. *The Journal of Technology Transfer*, 38(6): 828-835.

Jiang Pengju, Wei Ping, et al. (2014). Countermeasures to international technology transfer of south Jiangsu independent innovation demonstration zone. *Journal of Changzhou University (social science edition)*, 15(2): 60-63.

Liu Hua, Lu Yajuan. (2011). Calculation of dynamic efficiency for nation-level high-tech industrial development zones. *Science and Technology & Economy*, 24(1): 46-50.

Ou Guangjun, Sun Qian. (2013) Exploratory analysis of the High-tech zone collaborative innovation system's construction and influencing factors. *Journal of Industrial Technological Economics*, 3: 97-104.

Pawlak Z, Grzymala-Busse J, Slowinski R, et al. (1995). Rough sets. *Communications of the ACM*, 38(11): 88-95.

Pawlak Z. (1982). Rough set. *International Journal of Computer and Information Sciences*, 11(1), 341-356.

Shyng J Y, Shieh H M, Tzeng G H. (2010). An integration method combining Rough Set Theory with formal concept analysis for personal investment portfolios. *Knowledge-Based Systems*, 23(6): 586-597.

Slowinski R, Vanderpooten D. (2000). A generalized definition of rough approximations based on similarity. *IEEE Transactions on Knowledge & Data Engineering*, 2: 331-336.

Slowinski R.1992. Intelligent Decision Support: Handbook of Applications and Advances of Rough Sets Theory. Dordrecht: Kluwer Academic Publishers.

Tang Dexiang, Li Jingwen, Meng Weidong. (2008). The region difference analysis and the path dependence of R&D impact on the technique efficiency. *Science Research Management*, 29(2): 115-121."

Vanderpooten D. (1997). Similarity relation as a basis for rough approximations. *Advances in machine intelligence and soft computing*, 4: 17-33.

Wang C H, Chin Y C, Tzeng G H. (2000). Mining the R&D innovation performance processes for high-tech firms based on rough set theory. *Technovation*, 30(7): 447-458.

Wu Ke, Wang Xia. (2012). Research on relative policy and application of Zhangjiang national independent innovation demonstration zone. *Science & Technology Progress and Policy*, 29(12): 98-102.

Wu Peng, Yu Yongze. (2010). Research on the output-based efficiency of national level high tech industrial development districts and its impact factor. *Journal of Xi'an Jiaotong university (social science edition)*, 5: 16-23.

Wu Shu'e, Wang Yubao. (2012). Empirical analysis of the relationship between firms' S&T input and output performance in science park. *Forum on Chinese Science and Technology*, 1: 74-80.

Wu W W, Lee Y T, Tzeng G H. (2005). Simplifying the manager competency model by using the rough set approach//*Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing.* Springer Berlin Heidelberg, 484-494.

Xie Jialong, Hu Shuhua, Jiang Yuanyuan. (2012). Research on the competitiveness spatial differentiation of national Hi-Tech zones based on catastrophe progression method. *Science of Science and Management of S&T*, 32(12): 101-108.

XIE Jialong, HU Shuhua. (2013). Four capacities selection system of national demonstration zone and its application. *Studies in Science of Science*, 31(9): 1412-1421.

XIE Jialong, HU Shuhua. (2013). Design and application of the selection system for national indigenous innovation demonstration zone. *Chinese Soft Science*, 8: 67-79.

Xie Ziyuan. (2011). The factors influencing innovation efficiency of national level new and high technology development zones in China. *Science Research Management*, 32(11): 52-58.

Xiong Xi, Wei Xiao. (2016). An Innovation Capacity Evaluation of National Innovation Demonstration Zone. *Economic Geography*, 36(1): 33-38.

Xu Wangqiang, Jin Man, Zhang Hongfang. (2012). Study on incentive policy of major scientific and technological achievements Industrialization of Wuhan East Lake National Innovation Demonstration Zone. *Science & Technology Progress and Policy*, 29(6): 100-103.

Zhou Hongyu. (2015). Comparative study on the innovation ability of National Innovation Demonstration Zone. Science & Technology Progress and Policy, 32(22), 34-39.

Zhou Jiao, Zhao Min. (2014). Research on innovation efficiency and influencing factors of national High-tech industrial development zones. *Science and Technology Management Research*, 34(10): 1-6.

Table 1. Descriptive statistics of condition attributes and decision attributes

| Attributes | | Unit | Mean | St.D |
|---|---|---|---|---|
| $a_1$ | Senior and middle level professional qualifications | person | 74720 | 66747 |
| $a_2$ | Personal engaged in R&D activities | person | 120952 | 102249 |
| $a_3$ | Expenditure on R&D activities | 1000 yuan | 14091066 | 10706459 |
| $a_4$ | Ratio of expenditure on R&D activities to personal engaged in R&D activities | 1000 yuan / person | 124.96 | 35.19 |
| $a_5$ | Ratio of expenditure on R&D activities to total income | - | 0.02 | 0.01 |
| $d_1$ | Total income | 1000 yuan | 674202050 | 667585953 |
| $d_2$ | Technical income | 1000 yuan | 65463297 | 96775994 |
| $d_3$ | Ratio of technical income to total income | - | 0.08 | 0.05 |
| $d_4$ | Net profit | 1000 yuan | 47081490 | 45031997 |
| $d_5$ | Ratio of net profit to total income | - | 0.07 | 0.02 |
| $d_6$ | Taxes submitted | 1000 yuan | 35663883 | 34547100 |
| $d_7$ | Export | 1000 USD | 13690986 | 11002546 |

Table 2. Sub-intervals and their codes for condition attributes and decision attributes

| Attributes | Intervals / number of intervals | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| $a_1$ | (0, 88417] | (88417,+∞) | |
| $a_2$ | (0, 103698] | (103698, 167327] | (167327,+∞) |
| $a_3$ | (0, 10588116] | (10588116, 26480985] | (26480985,+∞) |
| $a_4$ | (0, 122.42] | (122.42, 149.75] | (149.75,+∞) |
| $a_5$ | (0, 0.024] | (0.024,+∞) | |
| $d_1$ | (0, 1299508956] | (1299508956,+∞) | |
| $d_2$ | (0, 71041221] | (71047221,+∞) | |
| $d_3$ | (0, 0.065] | (0.065,+∞) | |
| $d_4$ | (0, 72892799] | (72892799,+∞) | |
| $d_5$ | (0, 0.018] | (0.018,+∞) | |
| $d_6$ | (0, 59567645] | (59567645,+∞) | |
| $d_7$ | (0, 12390277] | (12390277, 26170509] | (26170509,+∞) |

Table 3. Decision rules when the decision attribute was total income ( $d_1$ )

| Decision attribute part | Reduct | Condition attributes part | Number of matching objects | % of objects covered |
|---|---|---|---|---|
| $d_1=1$ | Total objects | 50 | 48 | 96.00% |
| | $\{a_1,a_3,a_5\}$ | $a_1=1$ | 45 | 93.75% |
| | | $a_1=2,a_3=2,a_5=2$ | 3 | 6.25% |
| | $\{a_2,a_3,a_5\}$ | $a_2=1,a_3=1$ | 32 | 66.67% |
| | | $a_2=2,a_3=2$ | 8 | 16.67% |
| | | $a_2=1,a_3=2,a_5=2$ | 5 | 10.42% |
| | | $a_2=2,a_3=2,a_5=2$ | 2 | 4.17% |
| $d_1=2$ | Total objects | 6 | 2 | 33.33% |
| | $\{a_3,a_5\}$ | $a_3=3,a_5=1$ | 2 | 100.00% |

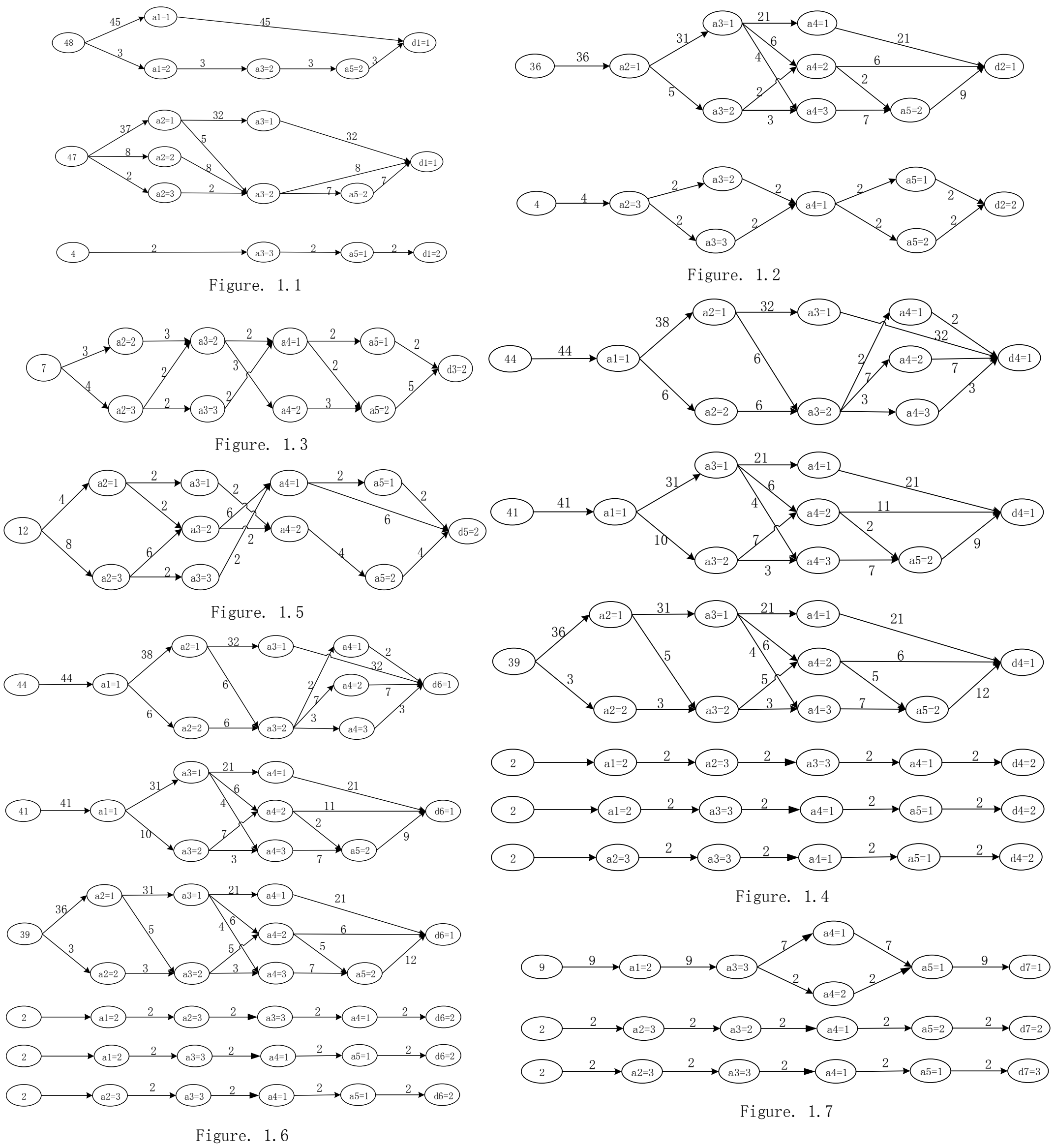


Figure 1. The decision flow graph

(Figure 1.1 was the decision flow graph when decision attribute was total income; Figure 1.2 was the decision flow graph when decision attribute was technical income; Figure 1.3 was the decision flow graph when decision attribute was ratio of technical income to total income; Figure 1.4 was the decision flow graph when decision attribute was net profit; Figure 1.5 was the decision flow graph when decision attribute was ratio of net profit to total income; Figure 1.6 was the decision flow graph when decision attribute was taxes submitted; Figure 1.7was the decision flow graph when decision attribute was export.)

Table 4. Decision rules when the decision attribute was technical income ( $d_2$ )

| Decision attribute part | Reduct | Condition attributes part | Number of matching objects | % of objects covered |
|---|---|---|---|---|
| $d_2=1$ | Total objects | 45 | 36 | 80.00% |
| | $\{a_2,a_3,a_4,a_5\}$ | $a_2=1,a_3=1,a_4=1$ | 21 | 58.33% |
| | | $a_2=1,a_3=1,a_4=2$ | 6 | 16.67% |
| | | $a_2=1,a_3=1,a_4=2$ | 4 | 11.11% |
| | | $a_2=1,a_3=2,a_4=3,a_5=2$ | 3 | 8.33% |
| | | $a_2=1,a_3=2,a_4=3,a_5=2$ | 2 | 5.56% |
| $d_2=2$ | Total objects | 11 | 4 | 33.33% |
| | $\{a_2,a_3,a_4,a_5\}$ | $a_2=3,a_3=2,a_4=1,a_5=2$ | 2 | 50.00% |
| | | $a_2=3,a_3=3,a_4=1,a_5=1$ | 2 | 50.00% |

Table 5. Decision rules when the decision attribute was ratio of technical income to total income ( $d_3$ )

| Decision attribute part | Reduct | Condition attributes part | Number of matching objects | % of objects covered |
|---|---|---|---|---|
| $d_3=2$ | Total objects | 32 | 7 | 21.88% |
| | $\{a_2,a_3,a_4,a_5\}$ | $a_2=2,a_3=2,a_4=2,a_5=2$ | 3 | 42.86% |
| | | $a_2=3,a_3=3,a_4=1,a_5=1$ | 2 | 28.57% |
| | | $a_2=3,a_3=2,a_4=1,a_5=2$ | 2 | 28.57% |

 

Table 6. Decision rules when the decision attribute was net profit ( $d_4$ )

| Decision attribute part | Reduct | Condition attributes part | Number of matching objects | % of objects covered |
|---|---|---|---|---|
| $d_4 = 1$ | Total objects | 48 | 44 | 91.67% |
| | $\{a_1, a_2, a_3, a_4\}$ | $a_1 = 1, a_2 = 1, a_3 = 1$ | 32 | 72.72% |
| | | $a_1 = 1, a_2 = 2, a_3 = 2, a_4 = 2$ | 4 | 9.09% |
| | | $a_1 = 1, a_2 = 1, a_3 = 2, a_4 = 2$ | 3 | 6.82% |
| | | $a_1 = 1, a_2 = 1, a_3 = 2, a_4 = 3$ | 3 | 6.82% |
| | | $a_1 = 1, a_2 = 2, a_3 = 2, a_4 = 1$ | 2 | 4.54% |
| | $\{a_1, a_3, a_4, a_5\}$ | $a_1 = 1, a_3 = 1, a_4 = 1$ | 21 | 47.73% |
| | | $a_1 = 1, a_3 = 2, a_4 = 2$ | 7 | 15.91% |
| | | $a_1 = 1, a_3 = 1, a_4 = 2$ | 6 | 13.64% |
| | | $a_1 = 1, a_3 = 1, a_4 = 3, a_5 = 2$ | 4 | 9.09% |
| | | $a_1 = 1, a_3 = 2, a_4 = 3, a_5 = 2$ | 3 | 6.82% |
| | $\{a_2, a_3, a_4, a_5\}$ | $a_2 = 1, a_3 = 1, a_4 = 1$ | 21 | 47.73% |
| | | $a_2 = 1, a_3 = 1, a_4 = 2, a_5 = 1$ | 4 | 9.09% |
| | | $a_2 = 1, a_3 = 1, a_4 = 3, a_5 = 2$ | 4 | 9.09% |
| | | $a_2 = 2, a_3 = 2, a_4 = 2, a_5 = 2$ | 3 | 6.82% |
| | | $a_2 = 1, a_3 = 2, a_4 = 2, a_5 = 2$ | 2 | 4.54% |
| | | $a_2 = 1, a_3 = 1, a_4 = 2, a_5 = 2$ | 2 | 4.54% |
| $d_4 = 2$ | Total objects | 8 | 2 | 25.00% |
| | $\{a_1, a_2, a_3, a_4\}$ | $a_1 = 2, a_2 = 3, a_3 = 3, a_4 = 1$ | 2 | 100.00% |
| | $\{a_1, a_3, a_4, a_5\}$ | $a_1 = 2, a_3 = 3, a_4 = 1, a_5 = 1$ | 2 | 100.00% |
| | $\{a_2, a_3, a_4, a_5\}$ | $a_2 = 3, a_3 = 3, a_4 = 1, a_5 = 1$ | 2 | 100.00% |

Table 7. Decision rules when the decision attribute was ratio of net profit to total income ( $d_5$ )

| Decision attribute part | Reduct | Condition attributes part | Number of matching objects | % of objects covered |
|---|---|---|---|---|
| $d_5 = 2$ | Total objects | 42 | 12 | 28.57% |
| | $\{a_2, a_3, a_4, a_5\}$ | $a_2 = 3, a_3 = 2, a_4 = 1, a_5 = 1$ | 4 | 33.33% |
| | | $a_2 = 1, a_3 = 1, a_4 = 2, a_5 = 2$ | 2 | 16.67% |
| | | $a_2 = 1, a_3 = 2, a_4 = 2, a_5 = 2$ | 2 | 16.67% |
| | | $a_2 = 3, a_3 = 2, a_4 = 1, a_5 = 2$ | 2 | 16.67% |
| | | $a_2 = 3, a_3 = 3, a_4 = 1, a_5 = 1$ | 2 | 16.67% |

Table 8. Decision rules when the decision attribute was taxes submitted ( $d_6$ )

| Decision attribute part | Reduct | Condition attributes part | Number of matching objects | % of objects covered |
|---|---|---|---|---|
| $d_6 = 1$ | Total objects | 48 | 44 | 91.67% |
| | $\{a_1, a_2, a_3, a_4\}$ | $a_1 = 1, a_2 = 1, a_3 = 1$ | 32 | 72.72% |
| | | $a_1 = 1, a_2 = 2, a_3 = 2, a_4 = 2$ | 4 | 9.09% |
| | | $a_1 = 1, a_2 = 1, a_3 = 2, a_4 = 2$ | 3 | 6.82% |
| | | $a_1 = 1, a_2 = 1, a_3 = 2, a_4 = 3$ | 3 | 6.82% |
| | | $a_1 = 1, a_2 = 2, a_3 = 2, a_4 = 1$ | 2 | 4.54% |
| | $\{a_1, a_3, a_4, a_5\}$ | $a_1 = 1, a_3 = 1, a_4 = 1$ | 21 | 47.73% |
| | | $a_1 = 1, a_3 = 2, a_4 = 2$ | 7 | 15.91% |
| | | $a_1 = 1, a_3 = 1, a_4 = 2$ | 6 | 13.64% |
| | | $a_1 = 1, a_3 = 1, a_4 = 3, a_5 = 2$ | 4 | 9.09% |
| | | $a_1 = 1, a_3 = 2, a_4 = 3, a_5 = 2$ | 3 | 6.82% |
| | $\{a_2, a_3, a_4, a_5\}$ | $a_2 = 1, a_3 = 1, a_4 = 1$ | 21 | 47.73% |
| | | $a_2 = 1, a_3 = 1, a_4 = 2, a_5 = 1$ | 4 | 9.09% |
| | | $a_2 = 1, a_3 = 1, a_4 = 3, a_5 = 2$ | 4 | 9.09% |
| | | $a_2 = 2, a_3 = 2, a_4 = 2, a_5 = 2$ | 3 | 6.82% |
| | | $a_2 = 1, a_3 = 2, a_4 = 2, a_5 = 2$ | 2 | 4.54% |
| | | $a_2 = 1, a_3 = 1, a_4 = 2, a_5 = 2$ | 2 | 4.54% |
| $d_6 = 2$ | Total objects | 8 | 2 | 25.00% |
| | $\{a_1, a_2, a_3, a_4\}$ | $a_1 = 2, a_2 = 3, a_3 = 3, a_4 = 1$ | 2 | 100.00% |
| | $\{a_1, a_3, a_4, a_5\}$ | $a_1 = 2, a_3 = 3, a_4 = 1, a_5 = 1$ | 2 | 100.00% |
| | $\{a_2, a_3, a_4, a_5\}$ | $a_2 = 3, a_3 = 3, a_4 = 1, a_5 = 1$ | 2 | 100.00% |

Table 9. Decision rules when the decision attribute was export ( $d_7$ )

| Decision attribute part | Reduct | Condition attributes part | Number of matching objects | % of objects covered |
|---|---|---|---|---|
| $d_7 = 1$ | Total objects | 31 | 9 | 29.03% |
| | $\{a_1, a_3, a_4, a_5\}$ | $a_1 = 2, a_3 = 3, a_4 = 1, a_5 = 1$ | 7 | 77.78% |
| | | $a_1 = 2, a_3 = 3, a_4 = 2, a_5 = 1$ | 2 | 22.22% |
| $d_7 = 2$ | Total objects | 19 | 2 | 10.53% |
| | $\{a_2, a_3, a_4, a_5\}$ | $a_2 = 3, a_3 = 2, a_4 = 1, a_5 = 2$ | 2 | 100.00% |
| $d_7 = 3$ | Total objects | 6 | 2 | 33.33% |
| | $\{a_2, a_3, a_4, a_5\}$ | $a_2 = 3, a_3 = 3, a_4 = 1, a_5 = 1$ | 2 | 100.00% |