Research on the Application of Data Mining Technology Based on Rough Set in the Development of Bio Health Industry Cluster

Shu Chen[1], Haojie Liu[2,a]

1School of Management,Wuhan University of Technology, Wuhan,Hubei,430070,China
2School of Information Engineering,Wuhan University of Technology, Wuhan,Hubei,430070,China
E-mails: anderson1020@163.com; Tel: +86 151-7253-6071

Corresponding author is Haojie Liu

Abstract
In this paper, the author mainly studied the application of data mining technology based on rough set in the development of bio health industry cluster. The 21st century is the era of life science; biotech has become the key point of modern S & T research and development, and is playing an increasingly important role in the improvement of human health, living environment, and the increase of agricultural and industrial production and quality. Successful cases of the developed countries manifests that, for enterprises of the biotech industry, industry cluster development model, which is featured by the closeness in geographical space and the cooperation and division of labor between companies, has facilitated the rapid development of local biotech industry to a large extent. This paper proposes strategic suggestions on further strengthening the effect of cluster independent innovation to facilitate independent innovation of Wuhan biotech industry cluster.
Keywords: DATA MINING TECHNOLOGY; ROUGH SET; DEVELOPMENT; BIO HEALTH INDUSTRY CLUSTER

1. Introduction

With economic development and the process of economic globalization speeding up, the barriers of resource elements between different countries is becoming lower, the competition on industrial development between developed countries and developing countries is becoming increasing. Because the economic development of any country is based on the industrial development, the research on the patterns of the industrial development has very important practical significance.

The 21st century, when humans have solved the problem of food and clothing, urbanization, informationization unprecedented flourishing, human material civilization and spirit civilization has been to a certain height, humans began to slowly consider healthcare, gene mystery and green environmental protection, clean energy etc. Aspects of the problem, "the ecological economy", "sustainable development" issues were frequently discussed. Life science, in the current situation, becomes the greatest concern, cutting-edge research field. Bio-technology industry as the most brilliant life science, is a bright pearl from medicine, agriculture, energy, environmental protection, food, chemical, resources, health care and other fields, improve human survival environment, maintaining human's physical and mental health. Can foresee, Bio-technology industry will sweeping the

globe, to meet the current human society deeper demand, transforming the lifestyle of the human society and economic development model. This will lift after IT after the revolution of new industrial revolution wave. Bio-industry clusters, as creatures of industrial development core carrier, the intensity of their competitiveness, researchers pay more and more attention.

Chen's paper [1], starting from the industrial cluster theory of industrial cluster theory, industrial clusters' competitiveness theory, industrial cluster theory of competitive strategy and other related literature in order to compare exhaustive review, Richard as with distinctive features, accord with Bio-industry cluster theory research of skeleton. Based on the related theory of Bio-technology industry cluster are constructed, Bio-industry cluster constitutional indexes system model, evaluation model and other related model, introduces global three regions of typical Bio-industry cluster and China national Bio-technology industry base, and on the basis of China Bio-technology industry cluster respectively the international competitiveness and China's own Bio-industry cluster development level the factor analysis, attempts to find the key factors affecting success, China Bio-industry cluster development level and Hubei Bio-technology development of industrial clusters prospects etc. the answer to the problem.

The rules of the industrial structure development are including the generality and diversity of the development of industrial structure, the balanced and unbalanced development of the industrial structure, the industry amalgamation and the industrial separation, the rationalization and optimization of the industrial structure. As to the single industrial development rules are including the industrial life cycle rule and industrial cluster rule. The single industrial trends are including the internationalization and ecological trends.

There are two problems on the industrial development, one is the overall characterization of the industrial development, and another is the function and the promotion mechanisms. So we can define that the development mode of the industrial development is the release of the rule and characteristics of industrial development under its promotion agent. Then, as to empirical study, the dissertation analyzed the development mode of China's automobile components industry within this research framework. After analyses the rules and trend of development of China's automobile components industry, this paper released the promotion mechanisms and agents, finally, the paper give the evaluation the performance of development mode of China's biotechnology health industry, and give the policy suggestions as well.

2. The Current State of Bio Health Industry Cluster in Hubei

The new technology revolution got booming development with modern life science and biotechnology in the 1980s when the IT revolution attains the quickest development. The history of world is in a great changing process, that is, from the age of physics and chemistry to the age of life science, from the industry revolution to the life science revolution. The bio-industry will become the most important strategic industry of the world sustaining development of economy and society in 21 centuries. It has become the pivot of every country's economic development, especially of big countries, to accelerate bio-industry development, promote its competitor power and occupy the highest point in a new turn international competition. In modern history, China has missed several opportunities of developing with the world technology revolution, leaving many pities and lessons. Nowadays, the world biotechnology is in the beginning stage of large-scale industrialization, which leaves both the developed and the developing countries facing a new selection and the undeveloped countries a historic opportunity to get a spanning development.

In the first 20 years of 21 century, our country will continue to push the industrialization progress and to set up the well-to-do society. Of course, to complete the above targets is an arduous mission in the country for owning so large population. Therefore, we should persist in holding scientific-developing view, changing the way of economy increase and looking for a new way of industrialization with Chinese character. This new way, ensuring sustaining development, depends on the enhancement of technology and economic benefit, redaction of resources consumption and environment pollution, and the sufficient utilization of human resources. The characteristics of bio-industry can satisfy the above requests to a large extent. Therefore, it is a historic strategic selection for China's peaceful rising to develop bio-industry and bio-economy strongly thus makes it a pillar strategic industry in our country, the key-point of which is to cultivate and promote our country's competitor power of bio-industry. The fast fierce development and day by day vigorous world competition of bio-industry make it become the focus of the domestic and international government, academic circles, industry field and even the whole world. The need of practice pushes the development of theories, so there is the necessity and urgency of putting forward and re-

searching a competitor power of bio-industry theory. New result by index checking shows that the related research hasn't been done, so this thesis is a kind of initial research. Surrounding the topic of competitor power of bio-industry, this thesis carried on extensively thorough actual investigation (in Changjiang triangle, Zhujiang triangle, Beijng and Tianjin et al), abundant literature search (life science, biotechnology, economics, management, industrial economics, environmental science and futurology et al) and careful data sorting and literature overview.

Kang's paper [2] describes and deduction, contrast and analogy, qualitative and quantitative, theoretical research and empirical research method of combining the, in the system elaborated Bio-industry cluster based on the theory of Bio-technology ascension in Hubei province, seeking the industrial clusters' competitiveness of strategic planning, in order to achieve drive national Bio-technology industry cluster competitiveness enhancement.

Wang's paper [3] mainly deals with the following questions: What is the innovation mechanism of industry cluster? What is the current situation of Hubei biotech industry cluster and independent innovation? Has the biotech industry cluster in Hubei facilitated industry's independent innovation? How to evaluate its effect? How does the industrial innovation of Hubei biotech industry cluster play in role? What could be the strategic suggestions? His paper explains the evolution of Industry Cluster theory, the development of independent innovation theory and the research perspective of industry cluster innovation and also makes special analysis on the mechanism of industry cluster innovation.

Jong [4] introduces the quantitative method adopted in this paper. It first describes indicators and methods of identifying industry geographic concentration, industrial clustering and industrial cluster; and then discusses the pros and cons of the common indicators for evaluating technical innovation and cluster independent innovation and gives an introduction to the statistical method adopted in the empirical part. His paper is also an analysis of the current status of biotech industry cluster and independent innovation in Hubei. It first discusses the current situation of the development of Hubei biotech industry, and then analyzes the trend of biotech industry cluster; finally it describes the development of industrial innovation of biotech industry in Hubei.

Miller [5] analyzes the effect industry cluster has on the industry innovation of Hubei biotech industry. It analyzes the impact of clusters on independent innovation of Hubei biotech industry. It points out in particular that the bio-industry cluster has provided a platform to carry out knowledge-intensive service activities, which stimulates outsourcing services, and thus promoted the innovation level of biotech industry in Hubei. Then it makes quantitative analysis on the position Hubei biotech industry cluster and innovation level takes in the country based on statistical data. It also makes evaluation by using the 17 indicators that reflect industry innovation capacity. Finally, it comes to the conclusion that the development of Hubei Wuhan biotech cluster has played a positive role in industry innovation and has statistical meanings. It introduces the developing overview of biotech industry in Wuhan Guanggu, and then proposes 6 assumptions on the internal mechanism of cluster independent innovation based on the micro- mechanism of promoting cluster innovation-- Collaborative Learning Theory in the second chapter. It tests the working path of industry cluster independent innovation at the micro-level through surveys and interviews on the innovation of Wuhan biotech industry cluster. Under the premise of the conclusions in previous chapters and in combination of the problems existed in industry innovation and proposes strategic suggestions on further strengthening the effect of cluster independent innovation to facilitate independent innovation of Hubei biotech industry cluster.

3. Rough Set Theory and the Algorithm

In mathematical forms, complete information systems are special case of incomplete information systems containing null attribute values; incomplete information systems containing null attribute values are special case of incomplete information systems containing some objects with partial known attribute values; and they are special set valued information systems.

In complete information systems, each object has unique known attribute value with respect to every attribute. So, we can define a discernibility relation on the universe, define lower and upper operators by indiscernibility classes, and discuss attribute reduce and decision rules optimization. In incomplete information systems, attribute values of some objects with respect to certain attributes are unknown or partial known, and in set-valued information systems, objects may have multiple attribute values with respect to certain attributes, so it is difficult to define an equivalent relation. On one hand, incomplete information systems can be transfixed to complete information systems, but this may cause the information losing; on the other hand, tolerant rough set model can be used to deal with incomplete information systems [7]. This gives new approaches using tolerant

rough set model to discuss attribute reduce and optimal decision rules acquisition in incomplete information systems and set-valued information systems.

Two kinds of lower and upper approximate operators are defined by use of maximal tolerant classes, respectively. By computing minimal disjunctive forms of discernibility function of maximal tolerant classes, optimal credible decision rules can be derived. In Ref. [8], attribute descriptors were introduced, a kind of lower and upper approximate operators were defend, and the concept of redact of the attribute descriptor was proposed, but its computation approach was not given. Incomplete information systems containing objects with partial known attribute values are studied by attribute descriptors. We define a new kind of lower and upper approximate operators, construct discernibility functions of attribute descriptors to compute redacts of attribute descriptors and induce optimal decision rules. We define so called GS-redact, DS-redact, G-redact and D-redact to evaluate the significance of attributes. The basic algorithm is shown as following:

$$h_{4ws} - h_{4w} = v(p_{4wb} - p_{4w}) \tag{1}$$

As is known to all, efficiency of the cluster is expressed in formula 2.

$$\frac{h_{4ws} - h_{4w}}{h_{4wb} - h_{4w}} = \eta_p \tag{2}$$

Formula 3 is deduced by formula 1 and formula 2.

$$h_{4wb} - h_{4w} = \frac{v(p_{4wb} - p_{4w})}{\eta_p} \tag{3}$$

Then we have:

$$P_p = \frac{\rho g Q H}{1000 \eta_p} \tag{4}$$

The total performance can be obtained from formula 5.

$$P = P_0 + \Delta P' - P_g + \rho g \Delta Z \tag{5}$$

Then, equation (2) can be expressed as the equaiton (6):

$$\frac{P_{01}}{P_{00}} = \frac{G_{01}}{G_{00}} \tag{6}$$

As seen from equation (4)-(5), according to the balance theory, formula 7 is formed as the following equations.

$$D_x(h_x - h_{xc})\eta_m' = (h_{4wb} - h_{4w})D_0$$

$$\text{or } D_x = \frac{h_{4wb} - h_{4w}}{h_x - h_{xc}} \bullet \frac{D_0}{\eta_m'} \tag{7}$$

Then we can get:

$$D_x = \frac{v(p_{4wb} - p_{4w})}{(h_x - h_{xc})\eta_p} \frac{D_0}{\eta_m'} \tag{8}$$

$$v = f(p, T) \tag{9}$$

Set-valued information systems are discussed and Set-valued information systems are generalization of single valued information systems. In different practice problems, set-valued information systems can be interpreted conjunctively or disjunctively. We use maximal tolerant classes to classify objects in the universe, define two kinds of lower and upper approximate operators respectively and define original and optimal decision rules. We also define the concept of relative redact of maximal tolerant class, and construct its discernibility function, to compute optimal decision rules by Boolean resealing techniques.

4. Results and Discussion

The biological industry is the21st century most important strategy industry. It is changing people's life variously in the energy, food, medicine, environmental protection and so on. And the high-end talent is the core factor that promotes the healthy and steady development of the biological industry. The realistic situation is unoptimistic. The bio-industry cluster is the core carrier of a biological industry development to promote the core competitiveness of enterprise. It should be discussed of high-tech industry and analyzes the development of Hubei province biological characteristics of industrial clusters. A creative model is constructed for the bio-industry clusters in Hubei and investigations should be done.

This paper, however, tries to construct a assessment indicator system from the he application of data mining technology based on rough set in the development of bio health industry cluster which is shown in Table1. Table 2 shows inspection model that is the name and meaning of each variable. The findings of Table 3 show that the empirical results are in consistency with the research hypothesis of this paper. Table 4 shows regression results of the five industries panel data model and corresponding statistical test results. By comparing Table 4 with Table 3, it is found that in the previous asset value-added rate and the data mining model based on rough set, the symbols of regression coefficients in all formats are identical, while only the degree of significance and values of the coefficients are different.

Conclusions

In this paper, the author mainly studied the application of data mining technology based on rough set in the development of bio health industry cluster.

Table 1. The evaluation index system of corporate bio health cluster responsibility

Responsible for all the stakeholders	The evaluation index	Calculation formula
responsibility for creditors	quick ratio	(current assets- inventory) / current liabilities
Responsibility for shareholders	profit rate of net asset	net profit/{ (beginning owners' equity + ending owners' equity) ÷ 2 }
responsibility for employees	Employee wage growth rate	(total wages in current period - total wages in previous period) / total wages in previous period
responsibility for customers	growth rate of prime business	(prime business revenue in current period - prime business revenue in previous period) / prime business revenue in previous period
	cost rate of prime business	prime business cost/ prime business revenue
responsibility for providers	velocity of accounts payable	prime business cost + ending inventory - beginning inventory / average balance of accounts payable
	cash accounts payable rate	net cash flow from operating / accounts payable
responsibility for government	taxes on asset	total corporate tax / average total assets
responsibility for community	employment growth rate	(number of employees in current period - number of employees in previous period)/number of employees in previous period
	public donation expenditure rate	donation expenditure / current sales

Table 2. Inspection model : Name and Meaning of each variable

Variable nature	variable name	variable symbol	calculation method
explanatory variable	The previous asset value-added rate	VAPAit-1	(Employees' salary + Interests + Tax + Dividend + Retained earnings)/Average total assets,
explained variable	Corporate social responsibility	CSRit	Synthesizing an indicator with the help of all the indicators and the weight given to each stakeholder can reflect corporate social responsibilities

Table 3. Regression results of the overall sample panel data model and corresponding statistical test results

Explanatory variables	Coefficient estimates ß	**P values**	Adjusted R^2	F statistics
VAPAit-1	0.667	0.001	0.181	9.125

Note: *, **and*** represent the significant correlation when the value is 10%, 5% and 1% respectively. The table shows the results after heteroscedastic amen.

Table 4. Regression results of the five industries panel data model and corresponding statistical test results

	public undertaking	commerce	real estate	industry	integrated industry
Coefficient estimates ß	0.182	0.248	0.322	0.287	0.341
P values	0.000	0.002	0.001	0.002	0.003
Adjusted R^2	0.391	0.189	0.265	0.136	0.124
F statistics	16.463	11.125	13.254	9.825	7.213

Note: *, **and*** represent the significant correlation when the value is 10%, 5% and 1% respectively. The table shows the results after heteroscedastic amen**.**

The 21st century is the era of life science; biotech has become the key point of modern S & T research and development, and is playing an increasingly important role in the improvement of human health, living environment, and the increase of agricultural and industrial production and quality.

Successful cases of the developed countries manifests that, for enterprises of the biotech industry, industry cluster development model, which is featured by the closeness in geographical space and the cooperation and division of labor between companies, has facilitated the rapid development of local biotech

industry to a large extent. The creative points of this paper are: (1) Qualitative analysis on the impact of industry cluster on industry innovation of Hubei biotech industry cluster. (2) Quantitative analysis on the position Hubei biotech industry cluster and innovation level takes in the country based on statistical data of the yearbook. Quantitative analysis on the impact industry cluster has on industry innovation and comes to the conclusion that the development of Hubei Wuhan biotech cluster has played a positive role in industrial innovation and has statistical meanings. (3) Carry out surveys on Hubei Wuhan biotech industry cluster to test the working paths of the core of industry cluster independent innovation--Collaborative Learning. (4) Proposes strategic suggestions on promoting independent innovation of Hubei biotech industry cluster. This paper proposes strategic suggestions on further strengthening the effect of cluster independent innovation to facilitate independent innovation of Wuhan biotech industry cluster.

Acknowledgements

This work is supported by the Key Project of Guangxi Social Sciences, China (No.gxsk201424), the Education Science fund of the Education Department of Guangxi, China (No.2014JGA268), and Guangxi Office for Education Sciences Planning, China (No.2013C108).

References

1. Jui-Kuei Chen, I-Shuo Chen. TQM measurement model for the biotechnology industry in Taiwan. Expert Systems With Applications, 2008, pp. 365-374.
2. Kyung-Nam Kang, Yoon-Sik Lee. What affects the innovation performance of small and medium-sized enterprises (SMEs) in the biotechnology industry? An empirical study on Korean biotech SMEs. Biotechnology Letters, 2008, pp. 3010-3018.
3. Kai Wang, Jin Hong, Dora Marinova, Liang Zhu. Evolution and governance of the biotechnology and pharmaceutical industry of China. Mathematics and Computers in Simulation, 2008, pp. 799-812.
4. Simcha Jong, Kremena Slavova. When publications lead to products: The open science conundrum in new product development. Research Policy, 2013, pp. 64-77.
5. Miller J A, Nagarajan V. The impact of biotechnology on the chemical industry in the 21st century. Trends in Biotechnology, 2000, pp. 185-194.
6. Smaglik Paul, Frantz Simon. The biotechnology sector's craving for commercial expertise is fuelling a cultural exchange with drug industry. Nature, 2003, pp. 422-431:.
7. DeLamarter John. Biotechnology partnerships--medicine for an ailing industry?. Nature Biotechnology, 2003, pp. 218-234.
8. D. McDonagha, A. Brusebergb, C. Haslamc. Visual product evaluation: exploring users' emotional relationships with products. Applied Ergonomics, 2002, 33, p.p.231-240.