Fractals, Vol. 17, No. 3 (2009) 299–309

AN ITERATING MODEL OF INDUSTRIES DISTRIBUTION IN REAL SPACE

YIMIN SHEN
Computer Engineering Department, Chengdu Electromechanical College
2 Huapaifang Street, Chengdu 610031, Sichuan, P. R. China
shen_yimin@sina.com

YUSHUN FAN
Automation Department, Tsinghua University
Haidian District, Beijing 100084, P. R. China
fanyus@tsinghua.edu.cn

STUART CAMPBELL
TIE Holding NV, Antareslaan 22-24
2132 JE Hoofddorp, The Netherlands
stuart.campbell@tieglobal.com

Received June 27, 2008
Accepted December 6, 2008

Abstract

An industry distribution model provides an important reference point for enterprises to extend business scope, and for governments to make industrial area planning. Current industry development models usually ignore the spatial distribution of enterprises, or build the models in virtual spaces. This paper proposes an iterating Model of Industry development in Real Space (MIRS) based on national division standards. Utilizing a multi-level iterating structure and a random concentration distribution approach, MIRS presents the fractal characteristics of log-normal distribution, self-similarity, and increasing trend of fractal information dimension. The coincidence between MIRS and actual industries is verified by the data collected from three industries in China, this allows MIRS to be an effective approach for predicting industry behaviors and a quantitative reference for decision-makers of companies.

Keywords: Industry Distribution; Real Space; Fractal; Iterating Model.

1. INTRODUCTION

Various enterprises and industries are important elements in the economy of a country. Accordingly, one of the questions that investors, macro economists, and governments are striving to answer is how to explain the economic phenomena (e.g. initializing, developing, recession, fluctuation, trend, and demise, etc.) of industries in different countries and regions as time moves forward.[1] A common and practical approach is to establish an adaptable mathematical model to represent the progress of industrial development. A mathematical model of the real world helps decision makers to obtain the correct information, predict the future trends for when they plan to choose a proper investment or program a scientific layout of national or regional economy.

The first step of industry modeling is a statistical study on a firm-size distribution in various countries. Experience has given pertinent, quantitative results about the distribution of size and growth rates of business companies. For example, Machado and Mata analyzed the size distribution of manufacturing firms in Portugal, and concluded that firm sizes are log-normally distributed.[2] Polèse *et al.* analyzed the location patterns of industries in Spain using census data for the years 1991 and 2001, and revealed that a crowding-out process existed in Spain, fuelling the growth of manufacturing activities in locations in close proximity to metropolitan areas.[3] Pagano and Schivardi found that differences in the size distribution within sectors affected cross-country differences in average firm size.[4] Outside Europe, Gupta *et al.* studied the statistical distribution of firm size for USA and Brazilian publicly traded firms through the Zipf plot technique, and examined its log-normal distribution.[5]

The scale of the above conclusions, however, is not wide enough. Here the word "scale" refers to the extension that a study focuses on and the effective range of the conclusion. For example, the scale of a result is about 10–100 km when a city is focused on in Polèse's work, whereas it is about 1000 km when a country is observed in Gupta's work. Obviously, a law of economy has a wider extension if its scale can be extended. But the current studies on enterprises distribution are applicable to either a couple of regions (say about 10–100 km), or an entire country (say about 1000 km). Therefore, a narrow denotation is expected. In this case, whether each of these results is adaptable to the others' scale or wider scale becomes a challenge.

In the circumstance that these statistic phenomena exist, an important domain of study on industry distribution is to discover the reason for them. Because of the complex behavior of industries in real world, the scholars have to establish simple models in virtual space to recapitulate the procedure of industry development. For example, Lotti and Santarelli developed a kernel density estimator to find out the general firm size distribution of young firms within some industries located in Italy.[6] A convergence towards limit distribution was observed in this work. Cabral and Mata analyzed a comprehensive data set of Portuguese manufacturing firms and proposed a simple theory based on financing constraints to explain the evolution of the firm size distribution.[7] Yang and Tse established the Box–Cox regression model with heteroscedasticity that estimated firm-size distribution of manufacturing firms in Portugal.[8] Fu *et al.* developed a dynamic percolation model which captured some of the features of the economical system, i.e., merging and splitting of business firms, represented as aggregates on a d-dimensional lattice.[9] These approaches can be summarized as iterating the evolution rules set in a virtual space with some initial enterprises to repeat the progress of industry and enterprises.

However, it is hard to map the real space to the virtual spaces on which the above models are based. The gap between real and virtual world hampers the verification of the models, and hence increases the application difficulty. Only when the object of study was limited to a scale narrow enough, can some initial results in real space be extracted. For instance, Pavlínek and Janák examined the regional restructuring of the supplier base of a Czech vehicle manufacturer, and after classifying 232 suppliers, their analysis revealed several trends of the Czech passenger car industry in the past 15 years, especially the relationships between the location of the suppliers and their basic characteristics such as their size, ownership, age and position in the supplier tier.[10] In the typical application domains of enterprises distribution models, e.g. location selection, many applied approaches were qualitative, like checklist methods, analog approaches, regression models, hierarchy processes, etc.[11]

It is worth noticing that concentration, self-similarity, and fractality have been found out in industry distribution. For example, using multiple indicators, Shen revealed a highly unbalanced spatial distribution and an increase in inequality of rural industrial development in China.[12] It was concluded that although the government policies attempted to reduce the inequality, the gap among regions in the period 1989–1994 was widened due to the strong impacts of the concentrating effects of market forces. Warnecke and Huser predicted that one of the essential requirements for production structures with a future is the capacity of entrepreneurial ways of thinking and acting in all areas, therefore each fractal unit itself should be a (little) "Fractal Company".[13] Dong *et al.* introduced a fractal extended enterprises model for collaboration.[14] In his paper, the hierarchical architecture of an extended enterprise can be generated through the recursion and iteration of proposed simple unit, named fractal-agent. The discovery of concentration and self-similarity brings more challenges to the industry distribution modeling since the model has to be in accordance with these new features as well as old ones.

In this context, this paper tackles the problem of industry modeling in real space to bridge the gap in theoretical model and natural environment. An iterating Model of Industry development in Real Space (MIRS) is presented in this paper with the following features:

(1) based on real space;
(2) the results of the model are valid in a scale much wider than the literatures;
(3) as a quantitative approach, this model can be iterated easily and verified quantitatively with data in reality; and
(4) it is able to recapitulate the development procedure of industry, in a sense of coincidence in statistics, self-similarity, and fractality.

The rest of the paper is organized as follows: First, Sec. 2 presents the iterating model and describes the stochastic selection approach of districts in detail. Then, the features of the model are analyzed in Sec. 3 to discover its rules and theorem about its distribution, self-similarity, and fractal trend. Section 4 verifies the model by applying it to medical, electronics, and chemical industries in China. Finally, Sec. 5 concludes this paper.

2. ENTERPRISE DISTRIBUTION MODEL IN REAL SPACE

The MIRS proposed in this paper consists of two parts, i.e.

(1) Hierarchical Iterating Framework (HIF) that establishes an iteration structure for industry development in the multi-level space; and
(2) Self-Concentration Distribution (SCD) that upgrades uniform distribution with a simple linear term to realize HIF.

Not directed against any special industries, HIF provides a mechanism that is widely applicable. Deterministic or random approaches need to be plugged in to HIF to make it operative. SCD is such a random approach that can be integrated with HIF to generate verisimilar features.

In this section, the two parts are demonstrated in detail.

2.1. Hierarchical Iterating Framework (HIF)

There are many kinds of virtual spaces where the enterprises exist, interact, and develop. Which are chosen is based on how the scholars imagine the model. However, the space where the enterprises really live in can only be the surface of the earth. Obviously, it is much more significant to study the trend of enterprises distribution and development in the real world than in the virtual one.

We can assume, or oppose, there exists some scale-free and cross-country laws of enterprise distribution. However, they are hard to be verified effectively by real data because of the various approaches of collecting data among countries. Therefore, quantitative studies have to be conducted only within the territory of each country.

In fact, each country has its own standard of administrative divisions, which defines the hierarchical region structure, for example, the scales of divisions are region, province, city, and county in China.[15] Now we formulate the hierarchical structure of divisions.

Assuming there are L layers of divisions in a country. The layer l has M_l divisions: $A_{l1}, A_{l2}, \ldots, A_{lM_l}$ $(l = 1, 2, \ldots, L)$. Each division A_{li} of the layer l $(l = 1, 2, \ldots, L-1)$ consists of m_{li} divisions belonging to the layer $l+1 : A_{l+1,j}$, denoted by $A_{l+1,j} \subset A_{li}$.

In order to keep the coincidence of denotations, let A_0 be the entire country, $M_0 = 1$, and m_{01} denotes the quantity of divisions of the top level.

Consider an industry, and assume that N_{01} enterprises of the industry, $C_1, C_2, \ldots, C_{N_{01}}$, have been opened for business and been distributed in the divisions by time T. Let N_{li} denote the Quantity of Enterprises (QoE) in division A_{li}, then it is obvious that:

$$\sum_{A_{l+1,j} \subset A_{li}} N_{l+1,j} = N_{li}$$
$$(l = 0, 1, \ldots, L-1; i = 1, 2, \ldots, M_l).$$

And at time T, the next enterprise of the industry, $C_{N_{01}+1}$, is going to choose a proper level L division (usually a county) to initialize its business. A Hierarchical Iterating Framework (HIF) is developed here to simulate the behavior of the enterprise when it makes the decision of locating its headquarters.

HIS is divided into L steps, corresponding to the L layers of the divisions:

(1) firstly, the enterprise $C_{N_{01}+1}$ selects a division A_{1i_1} ($i_1 \in \{1, 2, \ldots, M_1\}$) from the first layer divisions $A_{11}, A_{12}, \ldots, A_{1M_1}$;
(2) then, it selects a division A_{2i_2} ($i_2 \in \{1, 2, \ldots, M_2\}$) from the subordinate divisions, $\{A_{2i} | A_{2i} \subset A_{1i_1}, i = 1, 2, \ldots, M_2\}$, of the division A_{1i_1} selected in the previous step;
(3) and so on, until the L-layer division A_{Li_L} ($i_L \in \{1, 2, \ldots, M_L\}$) is selected.

The nested divisions A_{li_l} ($l = 1, 2, \ldots, L$) gained in the above process is the registration address of the enterprise $C_{N_{01}+1}$.

After the address decision of $C_{N_{01}+1}$, the quantity of the division A_{li} becomes:

$$N'_{li} = N_{li} + 1(i = i_l)$$
$$(l = 1, 2, \ldots, L; i = 1, 2, \ldots, M_i) \qquad (1)$$

Here, the format $1(B)$ is the indicator function of Boolean condition B, i.e.,

$$1(B) = \begin{cases} 1 & B = \text{True} \\ 0 & B = \text{False} \end{cases}.$$

Repeating the iterating process of HIF while N_{01} increases from 0 to any larger number or even infinity, the enterprises of an industry appear gradually and distribute in the real map of the country territory.

In the real world, determining the address of a new company may be a continuous process. It is simplified to a discrete stochastic process in HIF. The reason that we can do it in this way is that the enterprise can not take part in the industry and affect the other parties until the investors make their decision finally. Only after the enterprise starts to produce its products or provide its service, can the date of establishment be certain.

2.2. Random Concentration Distribution (SCD) Approach

However, HIF discussed above only provides the framework of the iterating model, and the algorithm of how to select the division in each layer is not given yet. A random distribution approach is proposed in this section.

Assuming the first $l-1$ steps have been completed, i.e., $A_{1,i_1}, A_{2,i_2}, \ldots, A_{l-1,i_{l-1}}$ are determined ($l = 1, 2, \ldots, L$).

Under this condition, let the probability of division A_{li} to be selected at the lth step isp_{li}. Obviously we have $\sum_{i=1}^{M_l} p_{li} = 1$.

If each of the l-layer divisions has the same probability to be chosen, then the chance that A_{li} is selected is:

$$p_{li} = \frac{1}{m_{l-1,i_{l-1}}} 1(A_{li} \subset A_{l-1,i_{l-1}})$$
$$(i = 1, 2, \ldots, M_l), \qquad (2)$$

with $i_0 = 1$.

However, Eq. (2) is untenable in the real world because the enterprise does not choose from the candidate divisions in a constant probability. In fact, resources (e.g., natural environment, human resources, material supplement, product market, etc.) are necessary or beneficial to the industry. In addition, the enterprise has to consider some other factors like competition in the local area, convenience of transportation, or economy of taxation policy. Therefore, during the process of the selection, the enterprises prefer the areas with advantages of the resources and/or factors that benefit the industry, i.e. the divisions with more Powers of Concentration (PoC). The administrative authorities (governments, local councils, etc.) keep making their local environment better to improve the PoCs with the object to attract more enterprises.

Because of the complexity of the relationship among the PoCs, it is very hard and unnecessary to establish the mathematical equations about the actions among them, similarly with the case of other economic problems.

But the result of the co-action of the PoCs between a division and an industry can be observed from a single variable, i.e. the QoE of the industry in the division. Obviously, an area with more enterprises of an industry is more attractive to the industry than those with fewer enterprises. Therefore, it is reasonable to assume the emergence probability of the next enterprise is positive related with the QoE in the area.

On this basis, a proportional part is added in the uniform distribution illustrated in Eq. (2) to reflect the positive relation:

$$p_{li} = \frac{\frac{1}{m_{l-1,i_{l-1}}} + \alpha N_{li}}{1 + \alpha N_{l-1,i_{l-1}}} 1(A_{li} \subset A_{l-1,i_{l-1}}) \quad (i = 1, 2, \ldots, M_l). \qquad (3)$$

Here α is a constant determined by the industry, and it characterizes to what degree the selection process is dependent on the PoCs of the areas. αN_{li} is the proportional term reflecting the relation between division selection and PoCs. The denominator $1 + \alpha N_{l-1,i_{l-1}}$ keeps the sum of probabilities equal to 1.

The distribution with linear term, as defined in Eq. (3) is called Self-Concentration Distribution (SCD). Obviously, when $\alpha = 0$, the SCD degenerates to the uniform distribution as defined in Eq. (2).

In the initial cycles of the iterating, the uniform distribution term plays a leading role since N_{li} is small. In this stage, the coming enterprises are well-distributed comparatively. And then, because of the randomicity of the stochastic process, discrepancy in QoE appears gradually among divisions. Also, the term αN_{li} has more and more important effect on the behavior.

SCD constructs a kind of random model with parity-breaking. Here parity means all the divisions are on an equal footing with others when the model is initialized because that N_{li} is equal to 0. And parity-breaking reflects that some fortunate divisions are developed more quickly and concentrated randomly to be centers of the industry.

It can be seen that no higher-order polynomial, exponential function, or other forms with more complexity is introduced in the model. Only a simple linear proportion term is used in Eq. (3) because this paper is the first attempt to create an iterating model of enterprise distribution in real space. A simple model can establish a benchmark in this domain of study, which can be taken as the base of further works. Moreover, even such a simple model is able to reproduce the main features of industry distribution during its development, as illustrated by the simulations in the following sections.

3. FEATURES OF THE MODEL

3.1. Stochastic Features

In order to find out the model behaviors as well as to enable the verification of its efficiency, the real space of the study in this paper is limited to the Chinese mainland (except of the three special administrative regions or provinces with special economic and political environment, i.e. Hongkong, Macao, and Taiwan). By reference to the national standard for the administrative divisions of the People's Republic of China,[15] 4-layer hierarchical divisions are constructed in this paper:

Region-level divisions (regions): There are $M_1 = 6$ regions in the space: A_1 — Northern China; A_2 — Northern East China; A_3 — Eastern China; A_4 — Southern China; A_5 — Southern West China; and A_6 — Northern West China.

Province-level divisions (provinces): Each region consists of several provinces, municipalities directly under the Central Government, autonomous regions, and special administrative regions, etc. They are collectively called "provinces" in this paper for convenience. There are $M_2 = 31$ provinces in the space, such as Beijing, Sichuan, Zhejiang, and so on.

City-level divisions (cities): A province is composed of some cities, districts, autonomous districts, and districts of the municipalities directly under the Central Government, etc. A general designation, "city", is used hereafter for all the city-level divisions. There are $M_3 = 368$ cities in the space, e.g. the downtown area of Beijing, Chengdu of Sichuan province, Taizhou of Zhejiang province, etc.

County-level divisions (counties): A city can be split into some counties, autonomous counties, and districts of cities, etc. The divisions in this level are all called "counties" hereafter. There are $M_4 = 3,120$ counties in the space, e.g. Haidian District of Beijing, Hi-tech Zone of Chengdu, Xianju county of Taizhou, etc.

Under the above multi-level division structure, MIRS with various α is iterated, and the distribution of QoE in the county-level divisions is observed. Figure 1 shows the relationship between QoE intervals and number of distributions whose QoE is

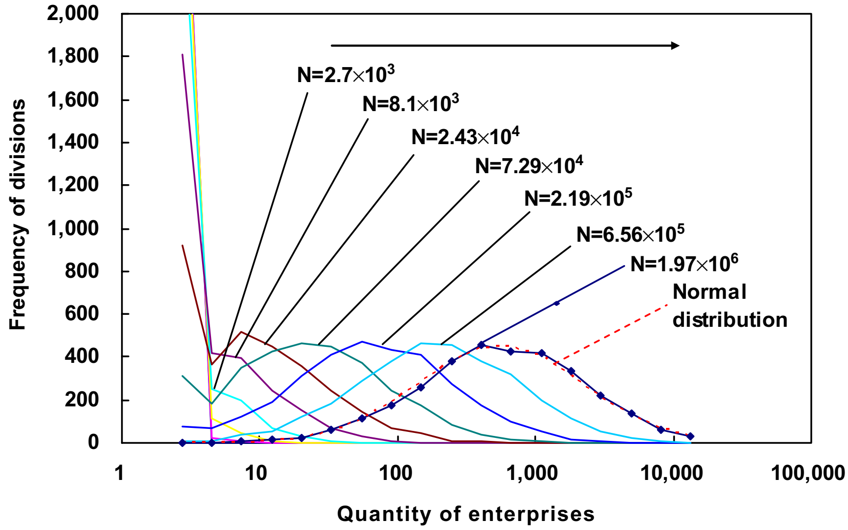


Fig. 1 Trends of the frequency of divisions with related QoE. (The horizontal coordinate axis is logarithmic.)

counted up to be in the intervals, with $\alpha = 0.3846$. Each curve in Fig. 1 represents the statistics when N reaches $3^k * 100$ $(k = 0, 1, \ldots, 9)$.

It can be observed from Fig. 1 that MIRS tends towards log-normal distribution. χ^2 test with 16 degrees of freedom is performed to measure the deviation of the MIRS after about two million $(1.9683{*}10^6)$ times of iteration from log-normal distribution. The test result, 17.17, is smaller than the P-value, 32.00, at the 1% significance level. Therefore, it can be accepted that the industry generated by MIRS after full iteration is log-normally distributed.

The parameter α affects the distribution significantly. The χ^2 test results after two million times iterations under various α are shown in two Fig. 2.

It can be concluded from Fig. 2 that when $\alpha = 0$, the χ^2 test value after two million iterations is still much larger than the P-value of the test, which indicates that there is a great deviation of the case from log-normal distribution. The larger the parameter α is, the more similar MIRS with the log-normal distribution.

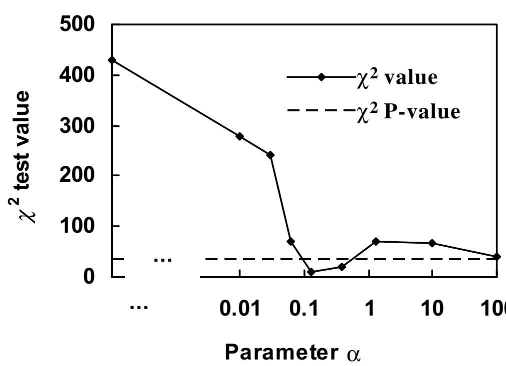


Fig. 2 Relationship between parameter α and χ^2 test value. (The horizontal coordinate axis is logarithmic.)

This phenomenon indicates that it is the proportional term added in Eq. (3) that brings the log-normal feature to MIRS.

3.2. Self-Similarity

According to the principle of Fractal Geometry, iterating system often generates sets with self-similarity, though usually in the sense of statistics.[16] In this section, the self-similarity of MIRS and the scale where the self-similarity exists are revealed.

The main task in each cycle of MIRS iteration is determining the division series $A_{1i_1}, A_{2i_2}, \ldots, A_{Li_L}$ progressively. The disorder after a completed cycle can be implicated by the Shannon Entropies[17]:

$$E_l = -\sum_{i=1}^{M_l} p_{li} \ln p_{li} \quad (l = 1, 2, \ldots, L). \tag{4}$$

Let the diameter of the whole space be 1, then the average diameter of an l-level division is:

$$\delta_l = M_l^{1/2}.$$

MIRS is iterated in the space, and its entropy after a definite series of iteration times $(N = 3^k \times 100, k = 0, 1, \ldots, 9)$ is calculated. The relationship between minus entropy $-E_l$ and diameter δ_l (under logarithmic horizontal axis) is drawn in Fig. 3 ($\alpha = 0.3846$).

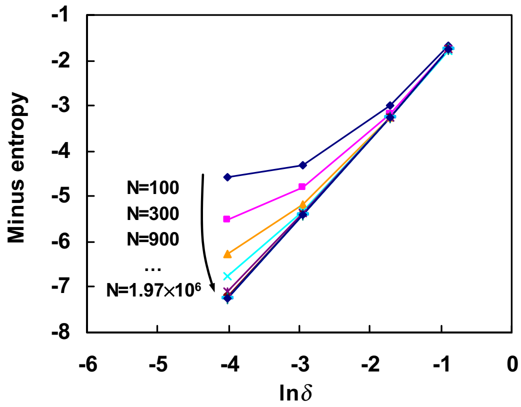


Fig. 3 Relationship between minus entropy $-E_l$ and diameter δ_l.

Figure 3 shows that $-E_l$ tends to have a linear relationship with ln δ_l along with the iteration, i.e.

$$-E_l = D\ln\delta_l + r.$$

Here D and r are l independent constants. According to the theory of Fractal Geometry,[18] this linear relation between $-E_l$ and ln δ_l means that the self-similarity of industry exists in the interval between δ_L and δ_1, and the constant D is the current fractal information dimension. The dimension has an increasing trend along with the iteration process as a whole, as viewed in Fig. 4. But the increase slows down and there seems to be an indication of convergence. The next section tries to discover the character of the convergence.

3.3. Trend of Fractality

Limited with the two-dimensional surface of the earth, the dimension of industry calculated above will no longer exceed 2. But 2 is not the limit of the convergence, as proved in Theorem 1.

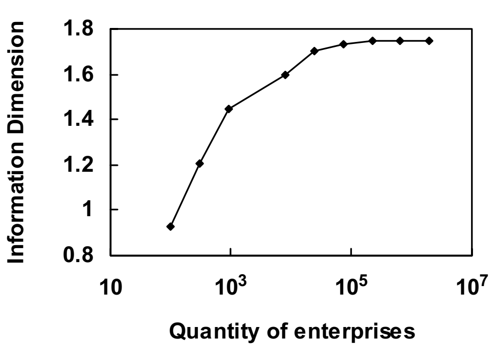


Fig. 4 Trend of information dimension along with the iteration process. (The horizontal axis is logarithmic.)

Theorem 1. *The information dimension of MIRS does not converge to* 2.

Proof. (counterevidence) Let E_{li} be the ith term in Eq. (4), i.e.

$$E_{li} = -p_{li}\ln p_{li} \quad (l = 1, 2, \ldots, L; i = 1, 2, \ldots, M_l). \tag{5}$$

After each iteration of MIRS, the term E_{li} turns to be E'_{li}. If $A_{li} \not\subset A_{l-1,i_{l-1}}$, then $E'_{li} = E_{li}$; Or else, if $A_{li} \subset A_{l-1,i_{l-1}}$, then the mean value of E'_{li} is

$$E(E'_{li}) = E(-p_{li}\ln p_{li}) \quad (i = 1, 2, \ldots, M_l).$$

Substituting Eqs. (1) and (3) into the equation above, we have

$$\begin{aligned} E(E'_{li}) = &-\frac{\frac{1}{m_{i-1,i_{l-1}}} + \alpha N_{li}}{1 + \alpha N_{i-1,i_{l-1}}}\frac{N_{li}+1}{N_{l-1,i_{l-1}}+1} \\ &\times \ln\frac{N_{li}+1}{N_{l-1,i_{l-1}}+1} - \left(1 - \frac{\frac{1}{m_{i-1,i_{l-1}}} + \alpha N_{li}}{1+\alpha N_{i-1,i_{l-1}}}\right) \\ &\times \frac{N_{li}}{N_{l-1,i_{l-1}}+1}\ln\frac{N_{li}}{N_{l-1,i_{l-1}}+1}. \end{aligned} \tag{6}$$

Assuming the current distribution is uniform, then,

$$N_{li} = \frac{N_{l-1,i_{l-1}}}{m_{l-1,i_{l-1}}} \quad (i = 1, 2, \ldots, m_{l-1,i_{l-1}}). \tag{7}$$

Substituting Eq. (7) to Eq. (6), we have

$$\begin{aligned} E(E'_{li}) = &\frac{\ln m_{l-1,i_{l-1}}}{m_{l-1,i_{l-1}}} - \frac{1}{m^2_{l-1,i_{l-1}}(N_{l-1,i_{l-1}}+1)} \\ &\times \Bigg[(N_{l-1,i_{l-1}} + m_{l-1,i_{l-1}}) \\ &\times \ln\left(1 + \frac{m_{l-1,i_{l-1}}}{N_{l-1,i_{l-1}}}\right) - S_i(N_{l-1,i_{l-1}}+1) \\ &\times \ln\left(1+\frac{1}{N_{l-1,i_{l-1}}}\right)\Bigg]. \end{aligned} \tag{8}$$

Before the iteration, we have

$$E_{li} = -\frac{1}{m_{l-1,i_{l-1}}}\ln\frac{1}{m_{l-1,i_{l-1}}} = \frac{\ln m_{l-1,i_{l-1}}}{m_{l-1,i_{l-1}}}.$$

So the first term of Eq. (8) is exactly E_{li}, then

$$E(E'_{li}) - E_{li} = \frac{1}{m^2_{l-1,i_{l-1}}(N_{l-1,i_{l-1}}+1)} \times \left[(N_{l-1,i_{l-1}} + m_{l-1,i_{l-1}}) \times \ln\left(1+\frac{m_{l-1,i_{l-1}}}{N_{l-1,i_{l-1}}}\right) - S_i(N_{l-1,i_{l-1}}+1) \times \ln\left(1+\frac{1}{N_{l-1,i_{l-1}}}\right)\right]. \quad (9)$$

According to the Taylor series of the function $\ln(1+x)$ about point 1, $\ln(1+x) = \sum_{n=1}^{\infty}(-1)^{n-1}\frac{x^n}{n}$, the two logarithmic terms in Eq. (9) are expanded to their Taylor series about point 1. After high-order terms merged,

$$E(E'_{li}) - E_{li} = -\frac{(m_{l-1,i_{l-1}}-1)}{2m_{l-1,i_{l-1}}N_{l-1,i_{l-1}}(N_{l-1,i_{l-1}}+1)} + o(N^{-2}_{l-1,i_{l-1}}).$$

Because $m_{l-1,i_{l-1}}$ is a constant larger than 1 and determined by the national standard, when $N_{l-1,i_{l-1}}$ is large enough, we have

$$E(E'_{li}) - E_{li} < 0. \quad (10)$$

The inequality (10) is summarized from $i = 1$ to $m_{l-1,i_{l-1}}$,

$$E(E'_l) - E_l < 0;$$

i.e. the mean value of the entropy decreases.

Uniform distribution has the maximum entropy, so it cannot be the limit distribution. But if the dimension converges to 2, the limit distribution is the uniform distribution in the surface. This brings a contradiction.

Therefore, Theorem 1 holds by counterevidence. □

It can be indicated from the theorem that if MIRS is proved to reflect the development process of industry correctly, then the industry will not homogenized gradually as time goes on and enterprises appears continuously, as might be expected. The industry will maintain its status below an upper bound of dimension less than 2, and present an "order in disorder" and self-similarity appearance in the scale.

4. MODEL VERIFICATION

Now we take the statistics of three industries in China to verify whether MIRS tallies with the actual situation.

4.1. Statistic Test

A database[19] is used for the verification of MIRS. The database stores the information of category, company name, registration month, division code, business scope, etc., of millions of enterprises from 1951 to 1996. Three industries, i.e. chemical industry with 303,809 enterprises, medical industry with 72,036 enterprises, and electronics industry with 24,850 enterprises, are chosen for the verification in this paper. The total quantities of enterprises of the three industries are shown in Fig. 5. Since the vertical axis of the figure is logarithmic, the industries seem to keep improving in a geometric progression during the period.

In order to test the statistic similarity between MIRS and actual data, the QoEs of each county-level division until 1996 are calculated based on the division code of every enterprise. The QoE distribution on the counties are χ^2 tested, compared with log-normal distribution. The test is executed for all the three industries, and the results are listed in Table 1. (The P-value $\chi^2_{0.01}(5) = 15.09$).

As shown in Table 1, the statistic significance of similarity between QoE distribution and log-normal distribution can be observed for the mature industries like chemical and medical industries, while the developing industry like electronics industry is still on its way to the distribution limit.

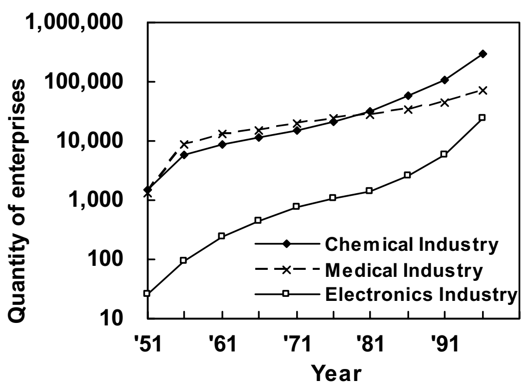


Fig. 5 Quantity of enterprises of the three industries. (The vertical axis is logarithmic.)

Table 1 χ^2 **Test Result of the Actual Data.**

Industry	Test Value	Result
Chemical industry	12.50	Accepted
Medical industry	10.74	Accepted
Electronics industry	25.86	Rejected

This phenomenon coincides with the feature of MIRS which was demonstrated in Fig. 1.

4.2. Self-Similarity Test

The Shannon Entropy of all the division levels is calculated, and the F test is applied to test the linearity between minus entropy and $\ln\delta$. The test results (i.e. 2191, 3072, and 836 for the three industries, respectively) are all markedly higher than the P-value, 98.5, at the 1% significance level. Therefore, it can be determined at the 1% significance level that the enterprises of the three industries are fractal in a scale interval between 55 km and 1265 km, about 1.36 orders of magnitude, according to the definition of information dimension.

Furthermore, the relationship between minus entropy and $\ln\delta$ by the end of every five years are observed, based on the registration month of each enterprise in the database. The result of medical industry is shown in Fig. 6 as an example.

Similar to Fig. 3, the minus entropy and $\ln\delta$ of the actual industries trends to linear relationship gradually, as shown in Fig. 6. Therefore, MIRS reproduces the development process of self-similarity of industry distribution very well.

4.3. Fractal Trend Test

The information dimension of the three industries by the end of each years are calculated from 1951 to 1996. At the same time, according to the quantity of enterprises registered every 5 years, MIRS with $\alpha = 0.125$, 0.385 and 1.333, respectively for the industries are iterated to compute their dimensions. The results are shown in Fig. 7.

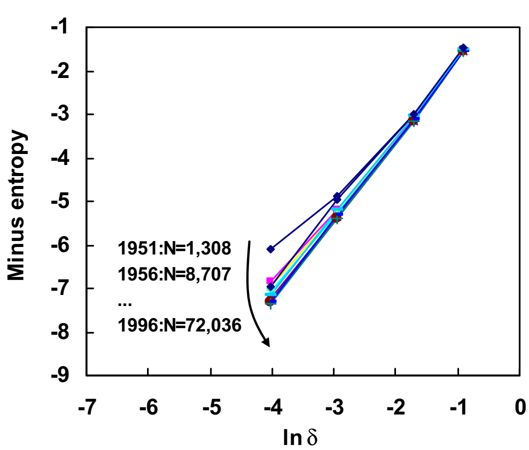
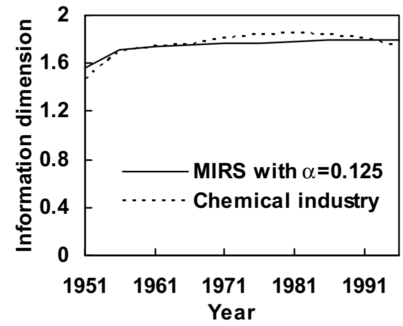
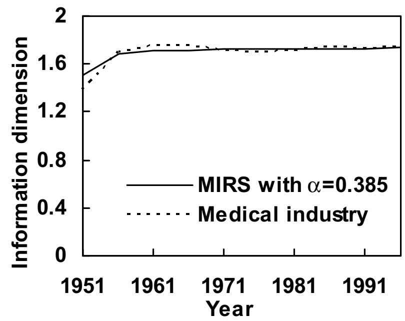


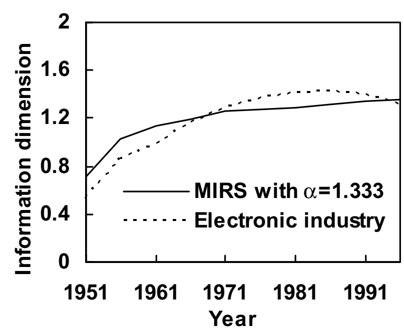
Fig. 6 The relationship between minus entropy and $\ln\delta$ of medical industry from 1951 to 1996.



(a) Comparing MIRS with chemical industry.



(b) Comparing MIRS with medical industry.



(c) Comparing MIRS with electronics industry.

Fig. 7 Comparison of the trends of the information dimension between MIRS and the industries.

It can be observed from Fig. 7 that

(1) during the 45 years from 1951 to 1996, the dimensions of MIRS have an increasing trend, similar with the three industries;
(2) after some years of increasing values, the dimension of MIRS slows down, and presents an convergence appearance at a range less than 1.8, other than tending towards 2, which coincides with Theorem 1; and
(3) by choosing realistic parameter α, MIRS reproduces the dimension process of chemical and medical industries.

The deviation amplitude between MIRS and electronics industry in Fig. 7(c) is higher than the other two industries. As a relatively new industry, the electronics industry is still in the process of becoming more and more stable. It is a probable error that gives rise to the deviation. Even though, the entire trend and convergence situation of MIRS are similar with the actual electronics industry.

Another phenomenon needs attention is that the information dimension of medical industry has gone through a full fluctuation period. It intersects downward the curve of MIRS around year 1971, as shown in Fig. 7(b). On the other hand, the first cross-down happened at the other two industries appeared after 20 years, as shown in Figs. 7(a) and 7(c). It indicates that medical industry seems to have "phase position" far in advance of the others. Therefore, if the downward intersection were taken as a base of prediction, it would have been concluded in 1971 that medical industry had matured, while the others would still expect a development by leaps and bounds.

The prediction could only be verified by the actual situation after five or ten years. In Fig. 5, it can be seen that a turning point appeared around 1981. The three industries had similar speed of development before that year, but then were separated into two types. Chemical and electronics industries made much bigger strides forward than the medical industry.

5. CONCLUSIONS AND FUTURE WORK

It is well known that enterprises live in the real world marked with longitude and latitude. But current studies are established in the virtual spaces to dodge the complicacy in the real space. It is uncertain whether there exists a mapping between the two kinds of spaces, which keeps the topological features like continuity, fractality, etc. So the models built in virtual spaces are difficult to apply even though this kind of models could be similar to reality in some statistical features, like log-normal distribution.

The first contribution of this paper is that MIRS establishes an industry development model in the real geographic space, instead of virtual one, which creates favorable conditions for data collection, result verification, and model improvement.

Upgrading uniform distribution with a linear term, MIRS behaves in a similar manner with true industries. The other contribution of this paper is that MIRS can be regarded as a good start and a benchmark of future works to improve the model, because of its simple form and quantitative results.

Our principal findings towards the three industries and MIRS are as follows.

(1) The quantity of enterprises of county-level divisions tends to be log-normal distributed;
(2) The entropy of industry is self-similar in the range from 55 km to 1265 km, and the self-similarity becomes more and more significant as time goes by;
(3) The information dimension of industry presents an increasing but convergent trend in a whole;
(4) MIRS tallies with the process of the three industries during the past 45 years in the above features; and
(5) The limit of industry information dimension will not be 2, according to MIRS.

The approaches proposed in this paper can be applied to the following scenarios.

(1) MIRS model for various industries can be established by choosing proper parameters to obtain the features and trends of self-similarity and concentration. Then the results can be taken by organizations like logistics companies, traffic control departments or raw material plants, as the basis of planning.
(2) The information of the related industries generated during the iteration of MIRS can be taken as references for investors and decision-makers of companies when they are investing in an industry or opening new branches for their current company.
(3) The information dimension and MIRS model can be compared to observe the dimension-fluctuation around MIRS. The intersection between them can be used to judge the current phase of the industry and predict the future trend.

The following domains are proposed for future research:

First, since MIRS has been verified in this paper to be coincident with the three industries in China, it is worth for it to be applied in more industries and more countries to discover additional laws of industries, as well as to improve the MIRS;

Then, though Theorem 1 indicates that there might be a limit of dimension other than 2, this paper does not conclude the expression of this limit.

If there is an upper bound less than 2 for most or all industries, it will become a constant with extremely important economic and technical value. Therefore, finding out the upper bound of the dimension can be the subject of future research.

ACKNOWLEDGMENTS

This work was supported in part by the National Hi Technology Research and Development Program of China under Grant 2006AA04Z166, and by the National Natural Science Foundation of China under Grant 60674080. Yimin Shen would also like to thank Hamzeh Sheikhhasan, Xiaofeng Di, Jian Zhou, Bo Liu, and Xiaoyan Bing for their comments which led to improvements of this paper.

REFERENCES

1. N. G. Mankiw, *Macroeconomics*, 5th edn. (Worth Publishes, New York, 2003).
2. J. A. F. Machado and J. Mata, Box-Cox quantile regression and the distribution of firm sizes, *J. Appl. Econ.* **15** (2000) 253–274.
3. M. Polèse, F. Rubiera-Morollón and R. Shearmur, Observing regularities in location patterns an analysis of the spatial distribution of economic activity in Spain, *Eur. Urban Regional Stud.* **14**(2) (2007) 157–180.
4. P. Pagano and F. Schivardi, Firm size distribution and growth, *Scand. J. Econ.* **105**(2) (2003) 255–274.
5. H. M. Gupta, J. R. Campanha, D. R. Aguiar, G. A. Queiroz and C. G. Rahej, Gradually truncated log-normal in USA publicly traded firm size distribution, *Phys. A* **375** (2007) 643–650.
6. F. Lotti and E. Santarelli, Industry dynamics and the distribution of firm size: a nonparametric approach, *S. Econ. J.* **70**(3) (2004) 443–466.
7. L. M. B. Cabral and J. Mata, On the evolution of the firm size distribution: facts and theory, *Am. Econ. Rev.* **93**(4) (2003) 1075–1090.
8. Z. Yang and Y. K. Tse, Modelling firm-size distribution using box-cox heteroscedastic regression, *J. Appl. Econ.* **21** (2006) 641–653.
9. D. Fu, S. V. Buldyrev, M. A. Salinger and H. E. Stanley, Percolation model for growth rates of aggregates and its application for business firm growth, *Phys. Rev. E* **74** (2006) 036–118.
10. P. Pavlínek and L. Janák, Regional restructuring of the Škoda Auto supplier network in the Czech Republic, *Eur. Urban Regional Stud.* **14**(2) (2007) 133–155.
11. R. J. Kuo, S. C. Chi and S. S. Kao, A decision support system for selecting convenience store location through integration of fuzzy AHP and artificial neural network, *Comput. Ind.* **47** (2002) 199–214.
12. X. Shen, Spatial inequality of rural industrial development in China, 1989–1994, *J. Rural Stud.* **15**(2) (1998) 179–199.
13. H. J. Warnecke and M. Huser, Lean production, *Int. J. Prod. Econ.* **41** (1995) 37–43.
14. H. Dong, D. Liu, Y. Zhao and Y. Chen, A novel approach of networked manufacturing collaboration: fractal web-based extended enterprise, *Int. J. Adv. Manuf. Tech.* **26**(11–12) (2004) 1436–1442.
15. Reference to standard: Codes for the administrative divisions of the People's Republic of China GB/T 2260–1999 (Chinese Standard Press, Beijing, 1999).
16. B. B. Mandelbort, *The Fractal Geometry of Nature* (Freeman, New York, 1983).
17. C. E. Shannon, A mathematical theory of communication, *Bell Syst. Tech. J.* **27**(7) (1948) 379–423.
18. K. Falconer, *Fractal Geometry* (John Wiley & Sons, New York, 1990) p. xxi.
19. Guangzhou Dongsheng Xuri Science and Technology Development Company, Ltd., Database of Chinese Enterprises and Institutions (Zhuhai Special Zone Audio-Video Production Center, Zhuhai, 2000).