Analysis and research of digital economy based on the background of big data

Zequn Yuan1,*

1 School of Electric Power, North University of China

Abstract. At present, there are two main approaches to measure the development level of digital economy at home and abroad: one is the direct method, which directly estimates the scale of digital economy in a region according to the defined measurement range; the other is to establish an evaluation index system of digital economy, which measures the development level of digital economy from multiple dimensions and calculates the comprehensive score of digital economy development level according to the weighted sum of each dimension. This paper adopts the second method to construct the evaluation index system of digital economy based on the connotation of digital economy from five dimensions to objectively reflect the development level of digital economy in each province. This paper builds on deep learning big data to study and analyze the digital economy, making a breakthrough in a new research area.

1. Introduction

The digital economy refers to an economic system in which digital technologies are widely used and consequently bring about fundamental changes in the overall economic environment and economic activities [1]. The digital economy is also a new socio-political and economic system in which both information and business activities are digitized [2]. Transactions between businesses, consumers and governments via the Web are growing rapidly. The digital economy focuses on the production, distribution, and sale of goods and services that all rely on digital technology. In this paper, the evaluation index system is constructed from multiple dimensions, and the comprehensive score of digital economy development level is calculated by weighting the sum of the weights of each index [3]. The methods of determining indicator weights are broadly divided into two categories: one is subjective assignment, such as network analysis, expert scoring, etc.; the other is objective assignment, such as entropy value method, factor analysis, etc. The subjective weighting method is based on individual subjective opinions and lacks an objective basis for determining weights; the objective weighting method relies on sample data and follows objective facts in determining weights [4].


2.1 Entropy method

In this paper, the entropy value method is used to measure the comprehensive development level of the digital economy, and the calculation steps are as follows:

Step 1: Construct the initial data matrix, \( X = \begin{bmatrix} x_{11} & \cdots & x_{1u} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mu} \end{bmatrix} \), where \( x_{ij} \) indicates the value of the first sample and the first evaluation index.

Step 2: When the value of the indicator is larger, the data normalization process is selected \( x_{ij}' = \frac{x_{ij} - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \), when the value of the indicator is smaller, the data normalization process is selected \( x_{ij}' = \frac{x_{\text{max}} - x_{ij}}{x_{\text{max}} - x_{\text{min}}} \), where \( x_j \) is the jth indicator, \( x_{\text{min}} \) is the minimum value of the jth indicator and \( x_{\text{max}} \) is the maximum value of the jth indicator.

Step 3: Calculate the weight of the value of the year indicator under the jth indicator, \( y_{ij} = \frac{x_{ij}}{\sum_{i=1}^{m} x_{ij}} \), get the weight matrix of the data \( Y = \{y_{ij}\}_{m \times n} \).

Step 4: Calculate the information entropy value of the jth term \( e_j = -K \sum_{i=1}^{m} y_{ij} \ln y_{ij} \) where \( K \) is a constant, \( K = \frac{1}{\ln m} \).

Step 5: Information utility value of a metric, \( d_j = 1 - e_j \).
Step 6: Calculate the weight of the jth indicator \( w_j = \frac{d_j}{\sum_{i=1}^{n} d_j} \)

Step 7: The weighted sum formula calculates the composite score of the sample \( U = \sum_{i=1}^{n} y_i w_j \).

2.2 Indicator selection and data sources

(1) Digital infrastructure level; (2) Digital industrialization level; (3) Industry digitization level; (4) Digital technology innovation level; (5) Digital financial development level.

2.3 Measurement results

Both traditional machine learning and deep learning require appropriate datasets for model training, validation and testing [5]. The dataset consists of modulated signals that have been correlated to generate signals that require a good understanding of the signal, noise interference, and channel model, which are then generated by the scientific computing software Matlab [6]. In this paper, we introduce the signal model, data processing and data generation process, and list and explain the parameter settings.

2.4 Signal Model

The signal model in this work can be expressed in equation 1:

\[
y(k) = h(k) \ast x(k) + n(k)
\]

where \( y(k) \) represents the received baseband unknown signal; \( x(k) \) represents the generated baseband modulated signal; \( n(k) \) represents the interference from additive Gaussian white noise; and \( h(k) \) is the complex channel environment interference, i.e., the Rayleigh fading channel accompanied by the Doppler shift.

(1) Additive Gaussian white noise

Additive White Gaussian Noise (AWGN) refers to a noise signal in which the spectral components follow a uniform distribution and the amplitude follows a Gaussian distribution [7]. The term "white" refers to the constant power spectrum, that is, the power spectral density is a constant at all frequencies, which can be expressed by equation 2:

\[
P_n(f) = \frac{n_0}{2} (W/Hz) - \infty < f < +\infty
\]

where \( n_0 \) is the normal number; \( P_n(f) \) curve is shown in Figure 1.

![Fig. 1. Power Spectral Density](image)

(2) Rayleigh fading channel

Rayleigh fading (Rayleigh Fading) is a wireless communication channel, due to the signal for multipath propagation to reach the receiving point at the field strength from different propagation paths, each path delay time is different, and the superposition of each directional component wave, and the resulting standing wave field strength, thus forming a signal fast fading called Rayleigh fading, and its envelope obeys the Rayleigh distribution. The Rayleigh distribution is a smooth narrowband Gaussian process with mean 0 and variance \( \sigma^2 \). Its probability density expression is shown in equation 3:

\[
f(x) = \frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right), x \geq 0
\]

where \( \sigma^2 \) is the variance, and comparing with equation 3, it can be seen that the mean value \( a \) is 0; when \( \sigma^2=1 \), the \( f(x) \) curve is shown in Figure 2.

![Fig. 2. Rayleigh distribution curve](image)

(3) Doppler shift

Doppler Shift (Doppler Shift) is when the mobile station moves at a constant rate along a certain direction, due to the propagation distance difference, it will cause the phase and frequency change, usually this change is called Doppler Shift. It reveals the law that the properties of waves change in motion [8]. When the motion is in front of the source, the wave is compressed, the wavelength becomes shorter and the frequency becomes higher (blue shift blue shift); when the motion is behind the source, the opposite effect occurs, the wavelength becomes longer and the frequency becomes lower (red shift red shift) [9]. The higher the speed of the wave source, the greater the effect produced.

The classical Jake Doppler spectrum is used to study the performance of the system at Doppler shifts with the expression shown in Equation 4:

\[
E(f) = \frac{1}{\pi f_{max}} \left[ 1 - \left( \frac{f}{f_{max}} \right)^2 \right] \mid f < f_{max}
\]

where \( c \) is the speed of light, \( \lambda \) is the wavelength, and \( f_{max} \) represents the maximum Doppler shift, which can be expressed by Equation 5:

\[
f_{max} = \frac{v_f c}{\lambda}
\]

where \( f_c \) represents the carrier frequency, which is taken as 2 GHz in this experiment.

(4) Analysis of the time evolution characteristics of the level of development of the digital economy
In order to study the temporal evolution characteristics of the digital economy development level, the temporal trends of the nation and the three regions are plotted respectively according to the above digital economy development level measurement results, as shown in Figure 3 below.

![Fig. 3. Economic development level trend chart](image)

As can be seen from Figure 3, the average value of digital economy development water in 30 provincial regions nationwide rose from 0.1319 to 0.3048 from 2011 to 2019, with a growth rate of 131.06%, showing a rapid upward trend in general. The water average of digital economy development in the eastern, central and western regions also shows a rapid upward trend. The average value of digital economy development water in the eastern region is much higher than the national average, rising from 0.1990 to 0.4239 in 2011, with a growth rate of 113.02%; the average value of digital economy development water in the central region is slightly lower than the national average, rising from 0.0999 to 0.2536 in 2011, with a growth rate of 153.85%; the average value of digital economy development water in the western region is lower than the national average, rising from 0.0881 to 0.2230 in 2011, with a growth rate of 153.12%. It can be seen that, from the level of development of the digital economy, the east > central > west, the eastern development advantage is huge; from the digital economy development growth rate, central > west > east, the central and western regions are struggling to catch up stage, and the digital divide between the east is expected to narrow. Next, this paper will analyze in detail the changing trends of digital economy development levels in 30 provincial regions.

3. Measurement of the Level of Quality Development of the Manufacturing Industry

3.1 Super Efficiency SBM

Data envelopment analysis (DEA) is a linear programming method that evaluates the relative effectiveness of decision units based on inputs and outputs. Tone proposed the super-efficient SBM model considering the non-desired outputs. The super-efficient SBM model not only deals with the non-desired outputs more appropriately, but also allows further comparisons to be made in an efficient decision unit. In the super-efficient SBM model, the efficiency value of the decision unit is not limited by $[0,1]$, so that the decision unit with efficiency value of 1 can be measured more accurately and the efficiency evaluation problem can be optimized. The super-efficient SBM model is constructed as follows:

Suppose the productivity of $n$ decision making units DMUs is measured, denoted as:

$$DMU_j(i = 1, 2, ..., n)$$

Each DMU has $m$ inputs, denoted as:

$$x_i(i = 1, 2, ..., m)$$

$q$ outputs, where the desired output is denoted as:

$$y_r(r = 1, 2, ..., q_1)$$

The non-expected output is noted as:

$$b_t(t = 1, 2, ..., q_2)$$

### 3.2 Indicator selection and data sources

1. Selection of indicators

Referring to the available literature and given the availability of data, the input and output indicators for manufacturing are shown in Table 1 below.

<table>
<thead>
<tr>
<th>Tier 1 Indicators</th>
<th>Secondary indicators</th>
<th>Tertiary indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input Indicators</td>
<td>Labor input</td>
<td>Average number of workers in manufacturing industry (10,000)</td>
</tr>
<tr>
<td></td>
<td>Capital investment</td>
<td>Manufacturing net fixed assets (billion yuan)</td>
</tr>
<tr>
<td></td>
<td>Energy input</td>
<td>Total industrial energy consumption (tons of standard coal)</td>
</tr>
<tr>
<td>Output Indicators</td>
<td>Expected output</td>
<td>Operating income of manufacturing industry (billion yuan)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Industrial wastewater discharge (million tons)</td>
</tr>
<tr>
<td></td>
<td>Non-desired outputs</td>
<td>Industrial sulfur dioxide emissions (million tons)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Industrial solid waste generation (million tons)</td>
</tr>
</tbody>
</table>

As shown in the table above, the input indicators in this paper include labor input, capital input and energy input, among which labor input is expressed by the average number of workers in manufacturing industry in each province; capital input is expressed by the net value of fixed assets in manufacturing industry, which can
4. CONCLUSION

This paper measures the level of digital economy development and the level of high-quality development of manufacturing in 30 provinces, cities and autonomous regions in China from 2011 to 2019, respectively, to quantitatively evaluate the development of each province and city in China and lay the foundation for the empirical analysis in Chapter 5.

For the measurement of digital economy development level, the evaluation index system of digital economy is constructed from five aspects: digital economy infrastructure level, digital industrialization level, industry digitization level, digital technology innovation level and digital financial development level, and the digital economy development level is measured by the entropy value method and analyzed based on the measurement results, and it is found that the digital economy development level of 30 provinces and cities in China has been improved from 2011 to 2019, and the digital economy development level shows a decreasing distribution pattern in the east and west, and the growth rate is faster in the regions lagging behind in digital economy development, and the digital divide is expected to be further reduced [10].

References

1. Li Qing, Dang Yangxiang. Research on the construction of curriculum system of big data and accounting in higher education institutions under the background of digital economy [J]. Old brand marketing, 2023(06):175-177.


7. Tang Lei, Gong Chen. Research on the construction of school-based big data center in higher education institutions under the background of digital economy--Sichuan Culture Industry Vocational College as an example [J]. Modern Marketing (Upper Journal), 2022(12):124-
126. DOI: 10.19921/j.cnki.1009-2994.2022-12-0124-042.


10. Li, Bingxiang, Ren, Hanxiao, Yin, Huiru, Guan, Lin. Recognition, measurement and presentation disclosure of big data assets in the context of digital economy[J]. Finance and Accounting.