Evaluation and Analysis of Electronic Information Industry Clustering Level Based on Soft Subspace Clustering Algorithm

YongTong Chen*
China West Normal University, China, 637000

*Corresponding author e-mail: Yongtong Chen@qq.com

Abstract. Based on the soft subspace clustering, this paper constructs an industry clustering index which can be used to test the significance, and then calculates and analyzes the clustering level of China's electronic information industry. The results show that the clustering phenomenon of China's electronic information industry is very obvious, and the clustering index shows an inverted U-shaped trend of rapid rise and slow decline in the whole research range. In terms of the distribution of the gathering areas, they are mainly concentrated in the eastern coastal areas such as Guangdong, Jiangsu and Shanghai, and the concentration degree is constantly increasing.

Keywords: Electronic Information Industry, Kurtosis, Skewness, Aggregation Index

1. Introduction
Since the middle of the 20th century, the information technology industry has grown rapidly, which has become an important part of the national competitive advantage, and the geographical aggregation characteristics of the electronic information industry have attracted more and more attention of many countries and regions. The successful experience of Silicon Valley[1-2], Hsinchu Industrial Park and other electronic information industry clusters in the United States confirms that the formation of industrial clusters plays a very important role in the electronic information industry. On one hand, industrial clusters can create more employment opportunities; on the other hand, industrial clusters can improve the division of labor in the electronic information industry, which will have a positive impact on labor productivity. In addition [3-4], industrial agglomeration can also bring many advantages, such as symbiosis effect, location effect, synergy effect, structural effect, etc., which has an important impact on the development of electronic information industry. Therefore, the regional coordinated
development of electronic information industry. At present, domestic scholars mainly use concentration, Gini coefficient, Hoover coefficient, eg coefficient to study the aggregation level of industry, which is different from the above research perspectives. This paper will build the industry aggregation index which can be used for significance test from the angle of kurtosis and skewness, and then measure\(^5\)\(^6\).

2. Soft subspace clustering algorithm

Soft subspace clustering is a kind of extremely unbalanced distribution, which is the expression of breaking distribution symmetry. Normal distribution refers to the symmetry of sample distribution, and its mean value is equal to the median value. When the sample distribution is biased, there is a gap between the mean value and the median value, resulting in tailing. Skewness is defined as:

\[
S = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^3} / \delta^3
\]  

(1)

In the above formula, \(n\) is the sample capacity as well as the average value, and \(\delta\) is the standard deviation. The greater the absolute value is, the more asymmetric the distribution will be. The \(s\) of normal distribution is 0; when \(s > 0\), the distribution is biased to the right; when \(s < 0\), the larger negative \(s\) shows a long tail to the left. The skewness cannot only indicate the magnitude of the skew normal quantity, but also indicate the direction of the skew degree. Generally, the value changes between -3 and -3, where 0 represents the symmetrical distribution such as the distribution of the center of gravity, +3 represents the extreme right skew state, and -3 represents the extreme left skew state.

Skewness indicates the symmetry degree of sample frequency distribution in electronic information industry, while kurtosis indicates the tip degree of sample distribution. Compared with the normal distribution, the peak distribution refers to the tip distribution higher than the normal distribution, and the low peak distribution refers to the distribution flatter than the normal distribution. From the numerical point of view, kurtosis is generally used to express the kurtosis of the top of frequency distribution. According to experience, the kurtosis coefficient of normal distribution is 3. Kurtosis is defined as:

\[
K = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^4} / \delta^4
\]  

(2)

Compared with the normal distribution, \(K\) measures the weight or flatness of the distribution. When \(k > 3\), the two sides of the normal distribution are steeper than the normal distribution, which is called peak state.

Skewness (s) and kurtosis (k) are statistics used to test the normal distribution of samples. According to the definition of \(\chi^2\) distribution, the variables of independent distribution obey the standard normal distribution, that is, the mean value is zero, the variance is 1, then their square and obey the \(\chi^2\) distribution. After standardized treatment of S and K, it can be concluded that:
\[ \left| \frac{S}{\sqrt{6/n}} \right|^2 + \left| \frac{K - 3}{\sqrt{24/n}} \right|^2 \geq \chi^2(2) \]  

(3)

Because \[ \left| \frac{S}{\sqrt{6/n}} \right|^2 - \left| \frac{K - 3}{\sqrt{24/n}} \right|^2 \geq 0, \] from above:

\[ \left| \frac{S}{\sqrt{6/n}} \right|^2 + \left| \frac{K - 3}{\sqrt{24/n}} \right|^2 \geq \frac{n}{6} S(K - 3) \]  

(4)

Since the \( \chi^2 \) distribution is only positive, in the above formula, we define the clustering strength index of soft subspace as:

\[ \gamma_{ks} = \frac{n}{6} \left| S(K - 3) \right| \]  

(5)

In the above formula, when \( s = 0 \) or \( K \approx 3 \), there must be no soft subspace clustering. When the skewness is large and the kurtosis is small, the clustering intensity of soft subspace is not necessarily large. That is to say, soft subspace clustering is a coupling effect of lateral deviation and longitudinal stretching.

3. Empirical evaluation and analysis

Using Eviews software, we can calculate the skewness from 1997 to 2007, the right skewness occurs in different degrees, indicating that large-scale and small frequency samples may appear at the right end of the distribution. From the value of skewness in 1997-2002, the skewness fluctuated around 2. It began to rise from 2003 until it reached the maximum value of 3.72 in 2002, and then began to decline, but the decline was small. It can be seen that although in China presents obvious soft subspace clustering, the level of soft subspace clustering in different years presents a trend of fluctuation.

**Table 1. Skewness and kurtosis of electronic information industry distribution in 1997-2007**

| Particular year | skewness | kurtosis | Particular year | skewness | kurtosis |
|-----------------|----------|----------|-----------------|----------|----------|
| 1997            | 2.31     | 8.16     | 1999            | 3.47     | 16.13    |
| 1998            | 2.27     | 8.30     | 2000            | 3.24     | 14.38    |
| 1999            | 2.28     | 8.60     | 2001            | 3.48     | 16.10    |
| 2000            | 1.98     | 6.48     | 2002            | 3.72     | 17.80    |
| 2001            | 2.57     | 9.38     | 2003            | 3.59     | 16.42    |
| 2002            | 2.86     | 11.05    | 2004            | 3.33     | 14.18    |
| 2003            | 3.10     | 13.23    | 2005            | 3.32     | 14.13    |
| 2004            | 3.35     | 15.28    | 2006            | 3.31     | 14.16    |
| 2007            | 3.54     | 16.77    | 2007            | 3.13     | 12.64    |

Using Eviews software, we can calculate the kurtosis of the output value in 1997-2007, as shown in Table 1. On the whole, the peak of electronic information industry began to rise in 1997, and reached its peak in 1997, with a leap in 2000. Then it decreased slightly from 1999 to 2002 and began to rise again, reaching the maximum value of 17.80. Since 2003, the peak degree of electronic information industry has been decreasing year by year, but it is higher than the lowest level of 6.48 in 2000.
Therefore, from the kurtosis point of view, China's electronic information industry has shown an obvious trend of soft subspace clustering, but there are fluctuations in different years.

Information system audit takes the information system of an organization such as an enterprise or a government as the object of the audit, comprehensive monitors and evaluates whether it can effectively and reliably achieve the organization's strategic goals beginning with aspects of information asset security, data integrity, and system effectiveness and efficiency through modern auditing theory and IT management theory, and puts forward detailed recommendations for improving and strengthening the organization's control of the information system. Different from the above general network security model system, considering the actual requirements of the security system, we propose a network security model suitable for the security system. As shown in Figure 1, the model includes communication module, environment request module, monitoring module and local policy library.

Agent manager

Signal communication

book
land
Strategy
slightly
library

Monitor

Environment request

equipment

**Figure 1. Information security risk analysis structure**

The communication module is the interface layer of information security risk analysis, which is responsible for transmitting various information, including: receiving control instructions from the network security model construction; feeding back the local network security model information to the network security model construction; exchanging information with the local monitoring module. The monitoring module is the core of the network security model, Responsible for the implementation of the network security model task.

The soft subspace clustering index of China's electronic information industry has been operating at a low level, mainly because this stage is the initial stage of the rise of China's high-tech industry, and the gap between regions is relatively small; from 2001 to 2002, the soft subspace clustering index of China's electronic information industry continued to rise, rapidly growing to 252.08 in 1997, The central and western areas are relatively slow in development. In 1997-2000, the aggregation index of
the electronic information industry in China decreased slightly, mainly was restricted by the financial crisis in 1997; During 2002-2007, the aggregation index of declined slowly, mainly because the central and western regions gradually paid attention in this stage, and paid attention to the transfer of electronic information industry in the eastern coastal areas, which made the electronic information industry in the central and Western Regions also get considerable development. As of 2007, the aggregation index is 155.93, which is in the level of moderate aggregation. Looking at the change trend of aggregation index from 1997 to 2007, it can be seen that the aggregation index of China's electronic information industry shows an inverted U-shaped curve trend of rapid rise first and then slow decline.

4. Conclusions
This paper constructs the industry aggregation index from the angle of skewness and kurtosis. On this basis, it evaluates and analyzes the aggregation level in 1997-2007, and draws the following conclusions: first, on the whole, the aggregation index value is greater than 18.42 in the whole research range, which indicates that the aggregation level of China's electronic information industry is relatively high; the phenomenon of industrial agglomeration is very obvious. Second, from the change trend of the aggregation index, the aggregation index of China's electronic information industry has been kept at a high level from 1997 to 2007, and from the change trend of each year, it shows the change trend of the inverted U-shaped curve, which first rises rapidly, then falls slowly.

References
[1] Gülçin Büyükozkan, & Gizem Çiçti. (2012). A combined fuzzy ahp and fuzzy topsis based strategic analysis of electronic service quality in healthcare industry. Expert Systems with Applications, 39(3), 2341-2354.
[2] J. Saltz, M.A. Serva, & R. Heckman. (2013). The get immersion experience: a new model for leveraging the synergies between industry and academia. Journal of Information Systems Education, 24(2), 121-131.
[3] Shigehiko Tanaka, Takayuki Fujiwara, Manabu Hashimoto, Takuma Funahashi, & Hiroyasu Koshimizu. (2011). Evaluation of jpeg blockiness by the fast analysis of information on local image frequency. Ieej Transactions on Industry Applications, 131(4), 600-607.
[4] Shaoyong Yang, Angus Bryant, Philip Mawby, Dawei Xiang, & Peter Tavner. (2011). An industry-based survey of reliability in power electronic converters. IEEE Transactions on Industry Applications, 47(3), 1441-1451.
[5] Abraham Gebregergis, & Pragasen Pillay. (2010). Implementation of fuel cell emulation on dsp and dspace controllers in the design of power electronic converters. IEEE Transactions on Industry Applications, 46(1), 285-294.
[6] S. D. J. Weier, Mohsin A. Shafi, & R. A. McMahon. (2010). Precision calorimetry for the accurate measurement of losses in power electronic devices. IEEE Transactions on Industry Applications, 46(1), 278-284.