Research on the Competitiveness of Talents in Eastern China based on Data Mining

Yanyan Chen
School of Business, Hohai University, 211100, Nanjing, China
811699795@qq.com

Abstract. On the basis of related research, we constructed an evaluation index system of talent competitiveness. Values are assigned to the indices with the use of the entropy weight method, so that reflecting the characteristics of competitiveness between different evaluation objects. In order to include more uncertainties of raw data in the evaluation process, the evaluation results were expressed by intuitionistic fuzzy sets and the data output was aggregated accordingly. By using the intuitionistic fuzzy information-based data mining comprehensive evaluation method, a comparative research was conducted on different levels of talent competitiveness in eastern China. Meanwhile, some targeted advices were offered with respect to promoting the competitiveness levels of talents in eastern China.

1. Introduction
The Internet is both a tool and a way of thinking. The "Internet Plus" action plan is helping develops a new type of operation. It has become a development priority for China to deepen the implementation of innovation-driven development strategy. Nowadays, as the Internet is more and more leading the innovation-driven development, the country has entered a new stage of innovation, transformation and development. According to national statistics and forecasting, the Internet will lift China’s GDP by 0.3%-1.0% from 2013 to 2025. This means that in the next decade or more, the contribution of the Internet to GDP will gain a new height as a crucial engine of economic progress (Ca0&Zou, 2014).

Technology is the primary productive force, and its most active and pioneering component is talents—the first resource and the pillar of scientific and technological innovation. As China's economic center, the eastern region has a high level of science and education, with massive talents being attracted here. The economic and social growth in eastern China has an important role to play in leading and promoting national economy. Therefore, we researched into talent competitiveness in eastern China in the Internet age as well as the difference between talents fostering at the regional level. Our research findings are of great practical significance to the implementation of regional measures to comprehensively uplift all-round talent competitiveness.

2. Overview of the research area
Eastern China lies in the eastern margin of Asian Continent, on the west coast of the Pacific Ocean. It covers the northeastern areas in China, the eastern margin of North China, East China, the southeastern areas in China and the southern and eastern sea areas including Bohai Sea, Yellow Sea, East China Sea and South China Sea. There are several provinces and municipalities in eastern China,
such as Beijing Municipality, Tianjin Municipality, Shanghai Municipality, Hebei province, Jiangsu province, Zhejiang province, Fujian province, Shandong province, Guangdong province, and Hainan province. The resource-rich eastern areas have regional superiority with respect to science and education, by encompassing the economic, political and cultural centers of China. Over 50% talents are accumulated in the area, as the major driving source of national innovation.

3. Construction of the evaluation index system and evaluation method

3.1. The principle of evaluation index system construction

(1) Scientificity. Evaluation indices should be selected in accordance with the nature and characteristics of talent competitiveness in the internet context; subjective judgment should be reduced or avoided as far as possible, so that reaching the goal of evaluation and comparison.

(2) Operability. Quantitative or easy-to-quantify evaluation indices are a preference. Based on the analysis of factors of talent competitiveness in eastern China, we take full advantage of the statistical indexes released by the government statistics department and corresponding index data, in order to ensure data authenticity.

(3) Dynamicity. The evaluation index system constructed in the paper should fit for the requirements of the time, and keep up with the latest research achievements at home and abroad. The research reports from international organizations and policy documents formulated by Chinese government departments should be the source of new evaluation indices free for dynamic update.

3.2. Construction of evaluation index system

Based on the principle of talent competitiveness evaluation index system construction at the regional level as well as related studies (Zhu et al., 2012; Li & Yang, 2013), we stratified and methodized the regional talent competitiveness evaluation index system into a three-layer hierarchic structures. The objective is to identify the evaluation indices themselves and their relationships in the internet age. Specifically speaking, the three layers are objective layer, factor layer and index layer, as listed in Table 1.

3.3. Determination of the intuitionistic fuzzy information evaluation method based on data mining

In the context of the Internet era, the research issue of talent competitiveness evaluation in eastern areas of China can be treated as a comprehensive evaluation project of city-varying talent competitiveness in the same set of evaluation criteria. The intuitionistic fuzzy information based on data mining evaluation method is used here to research on the project.

1) The standardization of evaluation indicators

In order to eliminate the difference in non-uniform evaluation index dimensions, we necessarily standardize raw data. If the index value of the target variable \( a_j \) in the evaluation index \( c_j \) is \( x_{ij} \), the data normalization method of \( c_j \) is:

For positive indicators, the normalized formula is

\[
X'_{ij} = \frac{x_{ij} - \min x_j}{\max x_j - \min x_j}
\]  

(1)

For negative indicators, the normalized formula is

\[
X'_{ij} = \frac{\max x_j - x_{ij}}{\max x_j - \min x_j}
\]  

(2)
If the evaluation values of different targets are deemed to obey the normal distribution, we may have \( \max x_j = \bar{x}_j + 3\sigma_j, \min x_j = \bar{x}_j - 3\sigma_j \), where \( \bar{x}_j \) and \( \sigma_j \) are the respective mean value and standard deviation of each evaluation index. In the case that the two cycles of values are considered, the standardized value of \( c_j \) related \( a_i \) may be more than one. We let the smaller one is \( \mu_{ij} \), and the bigger one is \( \mu^\theta_{ij} \).

In the intuitionistic fuzzy value, the membership degree \( \mu_{ij} \) is an expression of the worst performance of the evaluation target in the worst criteria. The fuzziness index \( \pi_{ij} = 1 - (\mu_{ij} - (\nu_{ij})) \) reflects the fuzzy degree of \( \mu_{ij} \), which means that the actual performance falls into the interval of \( [\mu_{ij}, 1-\nu_{ij}] = [\mu_{ij}^l, \mu_{ij}^u] \), where \( \mu_{ij}^l \) and \( \mu_{ij}^u \) are the respective upper boundary and lower boundary of the interval obtained in the process of standardizing evaluation index data (Chen et al., 2014). Therefore, we have \( \mu_{ij} = \mu_{ij}^l, \nu_{ij} = 1 - \mu_{ij}^u \). Accordingly, we acquire the intuitionistic fuzzy evaluation data of evaluation indices.

Table 1. The talent competitiveness evaluation index system in eastern China

| Objective layer | Factor layer | Weight | Index layer | Weight | Unit |
|-----------------|--------------|--------|-------------|--------|------|
| Talent resource | Technical personnel number per ten thousand people | 0.101 | people |
| Talent resource | The overall number of scientific and technological staff | 0.083 | 10^5 thousand people |
| Talent resource | The growth rate of scientific and technological staff | 0.071 | % |
| Talent resource | The number of universities, colleges, and science and technology institutes | 0.063 | - |
| Talent resource | The number of students in colleges and universities per 10,000 population | 0.051 | people |
| Talent input | Investment in research and development | 0.102 | 100 million yuan |
| Talent input | The ratio of investment in research and development to GDP | 0.081 | % |
| Talent input | The sum of appropriation expenditure on science and technology activities | 0.063 | 100 million yuan |
| Talent input | Education appropriation input | 0.081 | 100 million yuan |
| Talent output | The number of accepted patent application | 0.094 | Piece |
| Talent output | The number of science and technology papers | 0.064 | - |
| Talent output | The volume of transaction of technical contracts | 0.085 | 100 million yuan |
| Talent output | The added value of high-tech industry | 0.061 | 100 million yuan |
2) Weight assignment of evaluation indices

The relative importance of a random evaluation indicator \( c_j \) in the evaluation index set can be expressed by the weight \( W_j \). As said before, the importance is directly controlled by the competitiveness situations among different schemes. Such being the case, we adopt the following entropy weight method to calculate the situation-based index weight.

Assuming that there are \( n \) evaluation indexes in \( m \) to-be-evaluated schemes with unknown weight, we thus obtain a multi-index evaluation matrix \( X = (x_{ij})_{m \times n} \)

\[
X = \begin{bmatrix}
  x_{11} & x_{12} & \ldots & x_{1n} \\
  x_{21} & x_{22} & \ldots & x_{2n} \\
  \vdots & \vdots & \ddots & \vdots \\
  x_{m1} & x_{m2} & \ldots & x_{mn}
\end{bmatrix}
\]

Where \( x_{ij} \) is the original value of the \( i \)th evaluation index of the \( j \)th to-be-evaluated scheme.

In light of the great difference between both the dimension and the order of magnitude of different evaluation indexes, the first step is to normalize the raw data, whose approaches are seen in (1) and (2). Next, we calculate the information entropy \( E_i \) of the \( i \)th evaluation index

\[
E_i = -1/\ln(m) \sum_{j=1}^{m} \frac{q_{ij}}{q_j} \ln\left(\frac{q_{ij}}{q_j}\right) \quad (i=1,2,\ldots,m; \ j=1,2,\ldots,n)
\]

(3)

Where \( q_{ij} \) is the normalized value of the raw index data of \( x_{ij} \), \( q_j \) is the sum of normalized value of evaluation indices in the year of \( j \). We let \( \frac{q_{ij}}{q_j} \ln\left(\frac{q_{ij}}{q_j}\right) = 0 \) when \( \frac{q_{ij}}{q_j} = 0 \).

Finally, according to the entropy theory, the information entropy \( E_i \) of the \( i \)th evaluation index is obtained and used to calculate the weight of the \( i \)th index:

\[
Q_i = (1 - E_i)/(n - \sum_{i=1}^{n} E_i) \quad (i=1,2,\ldots,n)
\]

(4)

Where \( Q_i \) is the weight of the \( i \)th evaluation index, \( E_i \) is the information entropy of the \( i \)th evaluation index, \( n \) is the number of evaluation indices. Also, we have \( \sum_{i=1}^{n} Q_i = 1 \), \( Q_i \in [0,1] \). The entropy weight method based index weight is listed in Table 1.

3) Collection and comparison of evaluation information

In the evaluation process, we collect the intuitionistic fuzzy evaluation results of different evaluation indices according to the calculation result of intuitionistic fuzzy sets. The formula of a pair of arbitrary intuitionistic fuzzy sets \( \tilde{B} \) and \( \tilde{Q} \) as well as the real number \( \beta \geq 0 \) is

\[
\tilde{B} + \tilde{Q} = \left\{ \left( x, \mu_B(x) + \mu_Q(x) - \mu_B(x)\mu_Q(x), \nu_B(x)\nu_Q(x) \right) \bigg| x \in X \right\}
\]

(5)
\[ \tilde{B} \tilde{Q} = \left\{ \left( x, \mu_B(x) - \mu_Q(x), \nu_B(x) - \nu_Q(x) \right) \mid x \in X \right\} \]  

(6)

\[ \beta \tilde{B} = \left\{ \left( x, 1 - (\mu_B(x))^\beta, (\nu_B(x))^\beta \right) \mid x \in X \right\} \]  

(7)

\[ \tilde{B}^\theta = \left\{ \left( x, (\mu_B(x))^\theta, 1 - (1 - \nu_B(x))^\theta \right) \mid x \in X \right\} \]  

(8)

Thus, the target \( a_i \)'s intuitionistic fuzzy evaluation result obtained in different criteria is \( \tilde{d}_{ij} = \left( \mu_{y_j}, \nu_{y_j} \right) \) \( (j = 1, \ldots, n) \). According to the index weight vector \( W = (w_1, \ldots, w_n)^T \), the comprehensive evaluation result can be expressed by the intuitionistic fuzzy weighed mean operator \( M_w \):

\[ \tilde{a}_i = M_w(\tilde{d}_{i1}, \ldots, \tilde{d}_{in}) = \sum_{j=1}^{n} (w_j \tilde{d}_{ij}) = \left( 1 - \prod_{j=1}^{n} (1 - \mu_{y_j})^{w_j}, \prod_{j=1}^{n} (\nu_{y_j})^{w_j} \right) \]  

(9)

Accordingly, the overall and factor-layer-level intuitionistic fuzzy evaluation result of all indices is:

The overall value \( \tilde{a} = \left( 1 - \prod_{j=1}^{13} (1 - \mu_{y_j})^{w_j}, \prod_{j=1}^{13} (\nu_{y_j})^{w_j} \right), i = 1 \ldots 10 \);  

(10)

The talent resource value;

\[ \tilde{a}^r = \left( 1 - \prod_{j=1}^{5} (1 - \mu_{y_j})^{w_j}, \prod_{j=1}^{5} (\nu_{y_j})^{w_j} \right), w_j = w_j / \sum_{j=1}^{5} w_j , i = 1 \ldots 10 \]  

(11)

The talent input value;

\[ \tilde{a}^i = \left( 1 - \prod_{j=4}^{4} (1 - \mu_{y_j})^{w_j}, \prod_{j=4}^{4} (\nu_{y_j})^{w_j} \right), w_j = w_j / \sum_{j=1}^{4} w_j , i = 1 \ldots 10 \]  

(12)

The talent output value;

\[ \tilde{a}^o = \left( 1 - \prod_{j=4}^{4} (1 - \mu_{y_j})^{w_j}, \prod_{j=4}^{4} (\nu_{y_j})^{w_j} \right), w_j = w_j / \sum_{j=1}^{4} w_j , i = 1 \ldots 10 \]  

(13)

With the comprehensive evaluation result of various schemes, we realize the comparative analysis of talent attraction in different regions (Hong & Choi, 2000; Puello et al., 2016). In comparing the volume of two intuitionistic fuzzy sets, we define the score function \( \Delta \) and the precise function \( \sigma \) of an arbitrary intuitionistic fuzzy set \( \tilde{A} = \left( \mu_\tilde{A}, \nu_\tilde{A} \right) \) as:

\[ \Delta(\tilde{A}) = \mu_\tilde{A} - \nu_\tilde{A} \in [-1, 1], \sigma(\tilde{A}) = \mu_\tilde{A} + \nu_\tilde{A} = 1 - \pi_\tilde{A} \in [0, 1] \]

Thus, for arbitrary intuitionistic fuzzy sets \( \tilde{A} = \left( \mu_\tilde{A}, \nu_\tilde{A} \right) \) and \( \tilde{B} = \left( \mu_\tilde{B}, \nu_\tilde{B} \right) \)

If \( \Delta(\tilde{A}) > \Delta(\tilde{B}) \), we have \( \tilde{A} > \tilde{B} \).
If $\Delta(\tilde{A}) = \Delta(\tilde{B})$, we have three sub-results:

(a) If $\sigma(\tilde{A}) > \sigma(\tilde{B})$, we have $\tilde{A} > \tilde{B}$;
(b) If $\sigma(\tilde{A}) = \sigma(\tilde{B})$, we have $\tilde{A} = \tilde{B}$;
(c) If $\sigma(\tilde{A}) < \sigma(\tilde{B})$, we have $\tilde{A} < \tilde{B}$.

4. Evaluation results and analysis

According to the formulas (1) - (13) of the intuitionistic fuzzy data mining evaluation method, in combination with the raw value of each evaluation index in 2014 and 2015, we obtain the eastern China talent competitiveness evaluation result in 2015 in the Internet age.

| Table 2. The talent competitiveness evaluation result in 2015 in eastern China |
|----------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|                               | Comprehensive Value | Talent Resource | Talent Input | Talent Output |
|                               | $\mu$ | $\nu$ | $\Delta$ | $\mu$ | $\nu$ | $\Delta$ | $\mu$ | $\nu$ | $\Delta$ | $\mu$ | $\nu$ | $\Delta$ |
| **Beijing**                   | 0.60 | 0.38 | 0.21 | 0.55 | 0.40 | 0.15 | 0.50 | 0.31 | 0.18 | 0.53 | 0.38 | 0.14 |
|                              | 8   | 9   | 9   | 3   | 2   | 3   | 2   | 6   | 6   | 1   | 3   | 8   |
| **Tianjin**                   | 0.54 | 0.39 | 0.15 | 0.49 | 0.41 | 0.08 | 0.36 | 0.30 | 0.06 | 0.49 | 0.40 | 0.09 |
|                              | 5   | 1   | 4   | 8   | 3   | 5   | 9   | 1   | 8   | 8   | 4   | 4   |
| **Hebei**                     | 0.49 | 0.45 | 0.03 | 0.44 | 0.49 | -   | 0.45 | 0.40 | 0.05 | 0.47 | 0.43 | 0.04 |
|                              | 5   | 8   | 7   | 7   | 7   | 1   | 4   | 8   | 1   | 7   | 9   | 2   |
| **Shanghai**                  | 0.61 | 0.36 | 0.24 | 0.53 | 0.43 | 0.10 | 0.51 | 0.37 | 0.13 | 0.57 | 0.39 | 0.18 |
|                              | 4   | 7   | 7   | 8   | 6   | 2   | 1   | 2   | 9   | 1   | 1   | 0   |
| **Jiangsu**                   | 0.59 | 0.40 | 0.19 | 0.46 | 0.34 | 0.12 | 0.45 | 0.33 | 0.11 | 0.50 | 0.37 | 0.12 |
|                              | 7   | 0   | 7   | 7   | 2   | 5   | 5   | 7   | 9   | 8   | 1   | 8   |
| **Zhejiang**                  | 0.56 | 0.40 | 0.16 | 0.45 | 0.44 | 0.01 | 0.51 | 0.38 | 0.12 | 0.50 | 0.38 | 0.11 |
|                              | 2   | 2   | 0   | 5   | 1   | 4   | 6   | 7   | 9   | 2   | 0   | 2   |
| **Fujian**                    | 0.54 | 0.44 | 0.10 | 0.41 | 0.47 | -   | 0.43 | 0.33 | 0.09 | 0.41 | 0.37 | 0.03 |
|                              | 8   | 0   | 8   | 7   | 1   | 4   | 0   | 8   | 2   | 2   | 6   | 6   |
| **Shandong**                  | 0.56 | 0.40 | 0.15 | 0.49 | 0.37 | 0.11 | 0.47 | 0.36 | 0.11 | 0.48 | 0.35 | 0.13 |
|                              | 2   | 7   | 2   | 3   | 4   | 9   | 3   | 0   | 3   | 5   | 4   | 1   |
| **Guangdong**                 | 0.57 | 0.46 | 0.11 | 0.42 | 0.32 | 0.10 | 0.54 | 0.40 | 0.14 | 0.48 | 0.36 | 0.11 |
|                              | 4   | 2   | 2   | 6   | 4   | 2   | 6   | 2   | 1   | 7   | 4   | 1   |
| **Hainan**                    | 0.47 | 0.44 | 0.03 | 0.40 | 0.46 | -   | 0.47 | 0.42 | 0.05 | 0.46 | 0.42 | 0.03 |
|                              | 7   | 2   | 5   | 7   | 2   | 0   | 5   | 9   | 9   | 0   | 1   | 3   |

As shown in Table 2 and Figure 1, the overall talent competitiveness is strongest in Shanghai, followed by Beijing, Jiangsu, Shandong and Tianjin. The contributing factors to intensified talent competitiveness include geological advantage, massive universities colleges institutes, well-developed scientific and education resources, talent aggregation, and the governmental support of huge investment in science and education development. In this way, these regions witness the initial driving force of scientific and technological talents to promote economic development and transformation. As a comparison, regions with the weakest and second last weakest talent competitiveness are Hainan and Hebei, respectively. The major cause to the less satisfactory situation is little science and technology activities investment and the small number of universities, colleges, science and technology institutes.
in these areas. The power of science and technology innovation to promote the development of local high-tech and economy should be strengthened further.

![Figure 1. The talent competitiveness comprehensive evaluation result](image)

From the figure 2 below, we can see that in terms of human resources, the eastern region of Beijing, Jiangsu, Shandong, Shanghai and Guangdong are the biggest pool of scientific and educational resources, topping the list of the number of universities, colleges and science and technology institutes as well as the percentage of in-school college students in ten thousand population. It lays solid foundation for the enhancement of local talent competitiveness. The scores of talent resource in Hainan, Fujian and Hebei are negative, hitting the bottom of the score list. The major reason is the small number of universities, colleges and science and technology institutes as well as inadequate talent attraction power.

![Figure 2. Human resources evaluation results](image)

As seen in Table 2 and Figure 3, in terms of talent input, Beijing obtains the highest score, followed by Guangdong, Shanghai, Zhejiang, Jiangsu and Shandong. The economy in these regions has been among the forefront of our country, with adequate financial resources, and large appropriation in education, science and development. The area of lowest score in talent input is
Hainan, whose economic strength is the worst among the target regions. What is more, the funds for education and science and technology are limited, far less than those in the rest of the regions.

Figure 3. Talent input evaluation results

Figure 4 is the talent output evaluation results. By simultaneously referring to Table 2, we find that the highest level of talent output appears in Shanghai, followed by Beijing, Shandong, Jiangsu, Guangdong and Zhejiang. The regions with the last three levels of talent output are Hebei, Hainan and Fujian, respectively. The contributing factors are large investment in talent cultivation, education, and science and technology, and good environment of talent innovation. Remarkable results of economic development and shift have been achieved after more investment is released to talent innovation.

Figure 4. Talent output evaluation results

5. Conclusion
In this paper, we evaluate the 2015 talent competitiveness in the eastern China in the Internet era. By referring to related research, we establish a talent competitiveness evaluation index system, the data mining evaluation method is adopted on the basis of intuitionistic fuzzy information. The results show that: as a whole, the area of the strongest talent competitiveness is Shanghai, followed by Beijing, Jiangsu, Shandong and Tianjin; the regions with the weakest and second last weakest talent
competitiveness are Hainan and Hebei, respectively; the areas of the most abundant science and education resources are Beijing, Jiangsu, Shandong, Shanghai and Guangdong, and the areas of the poorest science and education resources include Hainan, Fujian and Hebei; in terms of talent input, the highest score is obtained by Beijing, followed by Guangdong, Shanghai, Jiangsu and Shandong, and the lowest score is obtained by Hainan; Shanghai has the highest level of talent output, followed by Beijing, Shandong, Jiangsu, Guangdong and Zhejiang, and the worst level of talent output appears in Fujian.

In the Internet era, in order for talents to better play the important role of promoting the regional economic and social development, the talent development efforts in eastern China should be centered on the core connotation of the Internet, especially guided by Internet thinking. Strong impetus should be provided to talent development system reform and policy innovation, with an overall planning of the construction of talent teams of all types. The advantaged gained by skilled personnel should be highlighted, so that releasing human and capital bonuses and forming the new situation of talent-driven development. Intense endeavors are suggested for the creation of “new environments” of talent development.

References

[1] Caoyang, Zouyunlong. (2014). Analysis between Start-up Education, Creational Education and Entrepreneur Education. Journal of Northeast Normal University (Philosophy and Social Sciences), (2),199-202.

[2] Zhu Anhong, Guo Ruliang, Gao Yan, Kong Weixiu. (2012). Assessment and Comparative Analysis on Central Six Provinces’ Technical Talents Competitiveness. Science and Technology Management Research, (10),66-71.

[3] Li Liangcheng, Yang Guodong. (2013). Regional Competitiveness Evaluation and Analysis of the Scientific and Technological Talents in China. Techno economics & Management Research, (1),24-27.

[4] Chen Wenjun, Yang Wue, He Zhengchu, Zhou Zhenhong. (2014). Comparing Urban Eco-competitiveness of Provincial Capitals in Midwestern China Based on Intuitionistic Fuzzy Information. China Soft Science, (5)151-163.

[5] Hong D H, Choi CH. (2000). Multicriteria Fuzzy Decision-making Problems Based on Vague Set Theory. Fuzzy Sets and Systems, 114 (1),103-113.

[6] Puello, José de J. Jiménez; Feliu, Tomás San; Calvo-Manzano, Jose A. (2016). Una aproximación basada en metamodelado del área de proceso de Validación del CMMI: Un caso de estudio. RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação, 17, 26–40.