Construction and Application of Investment Value Model for Hydrogen and Fuel Cell Industries Based on Clustering Algorithm

Xudong Li[1*], Qian Liu[2]

{lixudong2019@catarc.ac.cn[1*], liuqian2014@catarc.ac.cn[2]}

China Auto Information Technology (Tianjin) Co., Ltd, No. 68, East Xianfeng Road, Dongli District, Tianjin, China

Abstract. As one of the pillar industries of the national economy, automobiles have gradually become a hot field pursued by local governments and industrial capital in recent years, driven by technologies such as electrification, intelligence, networking, and low-carbon. At the same time, the automotive industry has the characteristics of wide technological coverage and strong synergy with related industries. Therefore, in order to better develop the automotive industry, it is necessary to not only deeply analyze the development prospects of various technological routes, but also combine various industrial resource elements, fully explore the value of industrial data, and use the cutting-edge theory of data mining to provide more scientific and quantitative investment value evaluation methods and algorithm models for government departments and industrial capital. Further elucidating the development laws of high-tech in the automotive industry, providing scientific basis for policy formulation, investment attraction, and investment layout.Under the promotion of China's "dual carbon" strategy, the hydrogen and fuel cell industries have received widespread attention from society and have become key areas for government investment and financial capital. This article constructs a cluster analysis model for the hydrogen energy industry, comprehensively evaluates the current development characteristics of the hydrogen energy industry, and provides theoretical support for subsequent policy formulation and industrial investment.

Keywords: Clustering Algorithm; Industrial Investment; Industrial Clustering; Hydrogen Energy and Fuel Cell Industry.

1 Introduction

The automobile industry is an important pillar industry in China, and automobile consumption is an important force driving domestic market demand. Automobile consumption accounts for about 10% of the total retail sales of consumer goods in society and accounts for over 4% of China's GDP. Considering the enormous driving effect of the automotive industry on industries such as energy, transportation, steel, aluminum, petrochemical, and intelligent chips, as well as the large number of job opportunities created, the significance of the automotive industry for China's economic and social development is even more important.

At present, the global new round of technological changes represented by new generation information technology, new materials, big data, artificial intelligence, green energy, etc., have a profound impact on the development process of the automotive industry, presenting a new

ICEMME 2023, November 17-19, Beijing, People's Republic of China

DOI 10.4108/eai.17-11-2023.2342712

look of collaborative and integrated development of multiple technological routes[1]. The automotive industry is accelerating its evolution towards electrification, intelligence, and networking. Driven by technological innovation, new industries, patterns, and formats will emerge.

The automotive industry is a typical high-tech industry, with high research and development investment, a large proportion of researchers, and the use of cutting-edge technology to produce high-tech products[2]. To develop high-tech industries, more resource elements are required, and the investment in funds, manpower, and time is also enormous. There are also more uncertain factors, low predictability, and greater risks involved. Therefore, if government departments and financial capital blindly invest in new technologies and fields in the automotive industry, it will lead to an overly loose industrial layout, which is not conducive to the sustained and healthy development of the industry.

Therefore, government departments and financial institutions urgently need to accurately grasp the development status, potential, and investment value of high-tech in the automotive industry, in order to better carry out high-quality project investment attraction, investment layout, and other work. Using data mining technology to analyze and model high-tech and related economic data in the automotive industry, bridging the barriers between different types of data, and providing theoretical and methodological support for scientific decision-making, has strong practical significance.

Hydrogen energy can be widely applied in fields such as energy, transportation, and industry. The hydrogen energy industry spans multiple fields such as energy, materials, and equipment manufacturing. It can effectively drive the transformation and upgrading of traditional industries, as well as generate new industrial chains. It is not only an effective path to achieve carbon neutrality, but also provides a new track for national and local economic development and the cultivation of strategic emerging industries. Fuel cell vehicles have gradually become a hot field of industrial investment. In 2022, the cumulative production and sales of hydrogen fuel cell vehicles increased by 105.4% and 112.8% year-on-year, respectively, showing a rapid expansion trend in overall application. This article will use clustering algorithms to construct an analysis model for the hydrogen and fuel cell vehicle industry, and comprehensively evaluate the current development status and future trends of China's hydrogen and fuel cell vehicle industry.

2 Research theory and methods

2.1 Basic theory

A large concentration of companies in a specific industry in a specific space can promote technology and knowledge spillovers between companies in that industry, facilitate the diffusion and innovation of knowledge and technology, and ultimately translate into economic growth. Industrial agglomeration can bring at least three benefits to cluster development: ① the sharing effect of intermediate inputs, that is, the sharing of infrastructure, thereby enjoying discounted prices for intermediate products (energy, transportation, third-party services, etc.); ② Knowledge spillover effect, namely the knowledge diffusion effect caused by closer social networks; ③ The pool effect of labor, which means sharing the talent market and making it easier to recruit various types of talents with high industry matching.

Based on the concept of Agglomeration and the theory of externalities (MAR) proposed by German economist Alfred Weber (1929), British economist Alfred Marshall, and American economist Kenneth J. Arrow, industrial agglomeration has multiple promoting effects on promoting industrial development and has been validated in many industrial fields.

The essence of clustering analysis method is to use computers to automatically divide objects into different groups based on certain standards, with objects in each group having similar attributes or approximate relationships with each other. Cluster analysis can be seen as an unsupervised pattern classification method that only relies on the attributes of the object itself to distinguish the similarity between objects [3].

2.2 Clustering algorithm

Clustering has a wide range of applications in fields such as commerce, industry, and finance, and there are also many algorithms, usually including density based clustering algorithms, system clustering, K-means clustering, model based self-organizing mapping neural networks (SOM), etc. [4].

This article adopts the SOM model, As shown in Figure 1, which is composed of an input layer and a competition layer. The layers are fully connected with corresponding weights, and the neurons in the competition layer are laterally connected to form a two-dimensional plane. Finally, the neurons adjust their weights through competition and cooperation, searching for the best matching neurons of the sample, and gathering similar data among similar neurons [5].

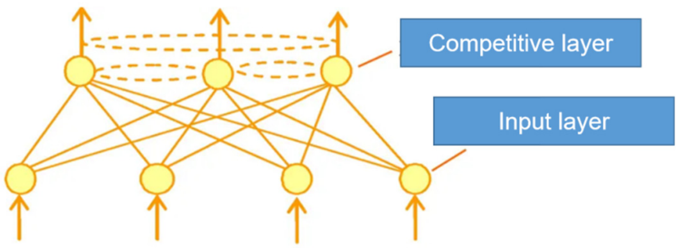

Fig. 1. Schematic diagram of SOM model.

The classic SOM algorithm process is as follows:

(1) Randomly select samples to initialize the connection weights between the input layer and the competition layer.

(2) Sequentially read the indicator sample x and use the Euclidean distance formula to find the winning neuron, As shown in formula (1).

$$i(x) = \arg \min_{j} \| x(n) - w_j \|, j = 1, 2, 3 \ldots, l \tag{1}$$

(3) Update the weights of the winning neuron and neighboring neurons based on the neighborhood function using the following formula, As shown in formula (2).

$$w_j(n+1) = w_j(n) + \eta(n) h_{j,i(x)} (x(n) - w_j(n)), j \in h_{j,i(x)} \tag{2}$$

Among them, 0 (n) 1 is the learning rate parameter, and $h_{j,\ i(x)}$ (n) represents the neighborhood function, As shown in formula (3) and formula (4).

$$\eta(n) = \eta(0)\ (1 - \frac{n}{N}) \tag{3}$$

$$h_{j,i(x)}(n) = \exp\ (- \frac{\| j - i(x) \|^2}{2\delta(n)^2}) \tag{4}$$

(4) Repeat the second iteration step until the network weights converge and the clustering results are output.

Due to the lack of original labels in clustering analysis, accuracy and other indicators cannot be used to evaluate the effectiveness of the algorithm. Purity, Dunn index, contour coefficient, and other indicators need to be used to evaluate the effectiveness of clustering analysis.

3 Establishment of indicator system

3.1 Principles for constructing an evaluation index system

The construction of an evaluation system generally requires following three main principles: ① The selected indicators need to have systematic characteristics, that is, each indicator is independent of each other but internally connected, and the logical relationship is progressive layer by layer, from macro to micro, or from shallow to deep, each indicator can form a whole, ensuring that the main body can be evaluated from all aspects; ② Indicators need to be representative, that is, they can highly summarize the subject or be highly related to the subject to be evaluated, ensuring that the main information can be well preserved and accurately quantified in cases where there are relatively few indicators; ③ Indicators have scientific nature, that is, each indicator can objectively evaluate the subject, the information represented cannot be duplicated, and the number of selected indicators cannot be too large or too small.

3.2 Construction of hydrogen energy and fuel cell clustering model

The hydrogen energy and fuel cell industry as a whole can be divided into three parts: upstream, midstream, and downstream, as shown in Figure 2. The upstream mainly includes hydrogen production methods such as fossil fuel hydrogen production, industrial by-product hydrogen production, and electrolytic water hydrogen production; The midstream mainly includes the storage, transportation, and refueling of hydrogen gas; Downstream is the utilization of hydrogen energy, including fuel cells, hydrogen electric coupling, hydrogen combustion, and other methods.

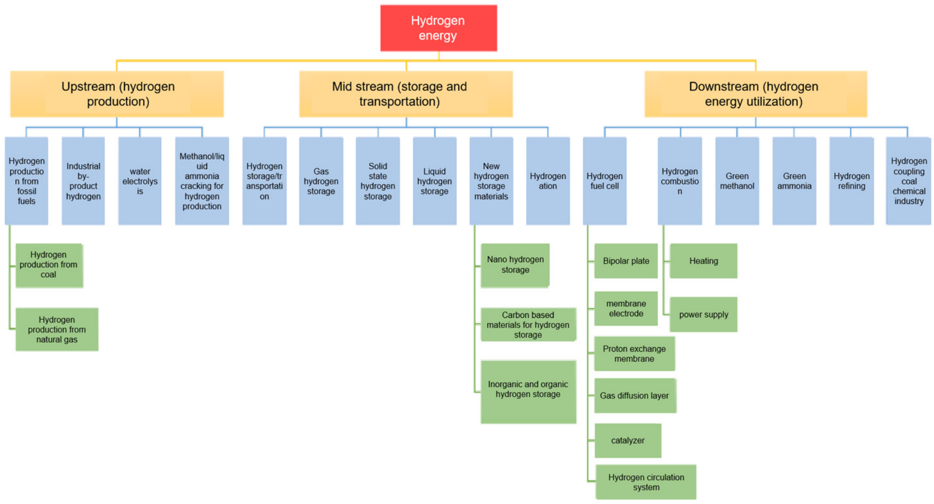

Fig. 2. Graph of Hydrogen Energy and Fuel Cell Industry Chain.

To objectively and accurately analyze the development level of hydrogen energy and fuel cell industries in various regions of China, it is necessary to grasp the development status of the industry from multiple dimensions, levels, and aspects. By using big data technology, we focus on collecting two types of data: industrial information and capital information. Establish a hydrogen and fuel cell industry analysis model covering secondary indicators and 8 data categories as shown in Table 1.

Table 1. Analysis model for hydrogen energy and fuel cell industry.

Primary indicators	Secondary indicators
industry information	business information
	patent information
	social security information
capital information	financing Information
	market value distribution
	specialized, refined, and innovative enterprises
	coopetition
	financial health

Based on the definition of the hydrogen energy industry chain, it covers 1273 hydrogen related enterprises, including 86 listed companies. The statistics of the regional distribution of enterprises are shown in Table 2.

Table 2. Distribution of hydrogen and fuel cell enterprises.

Area	Enterprise scale (10000 yuan)	Number of enterprises
Beijing City	175000000	126
Shanghai City	24100000	121
Jiangsu Province	5140000	192
Shandong Province	5590000	108
Guangdong Province	4460000	146
Zhejiang Province	8310000	92
Inner Mongolia Autonomous Region	3760000	27
Hebei Province	4540000	45
Shaanxi Province	4550000	43
Sichuan Province	2950000	60
Liaoning Province	4810000	38
Hubei province	1750000	41
Anhui Province	1140000	31
Henan Province	1270000	36
Chongqing City	827000	20
Xinjiang Uygur Autonomous Region	2480000	11
Fujian Province	2600000	17
Hunan Province	273000	21
Ningxia Hui Autonomous Region	809000	6
Guizhou Province	633000	5
Heilongjiang Province	727000	9
Tianjin City	619000	18
Jilin Province	9080000	11
Shanxi Province	4150000	23
Gansu Province	4510000	4
Yunnan Province	382000	5
Jiangxi Province	283000	9
Guangxi Zhuang Autonomous Region	152000	5
Qinghai Province	90300	2
Hainan Province	3000	1

4 Empirical analysis

4.1 Cluster analysis of hydrogen and fuel cell industries

The clustering coefficient is an indicator used to reflect the degree of tightness and connectivity among enterprises within a cluster. From a national perspective, as shown in Figure 3, the average clustering coefficient of the hydrogen energy industry is 0.615, indicating that China has initially formed a hydrogen energy industry cluster.

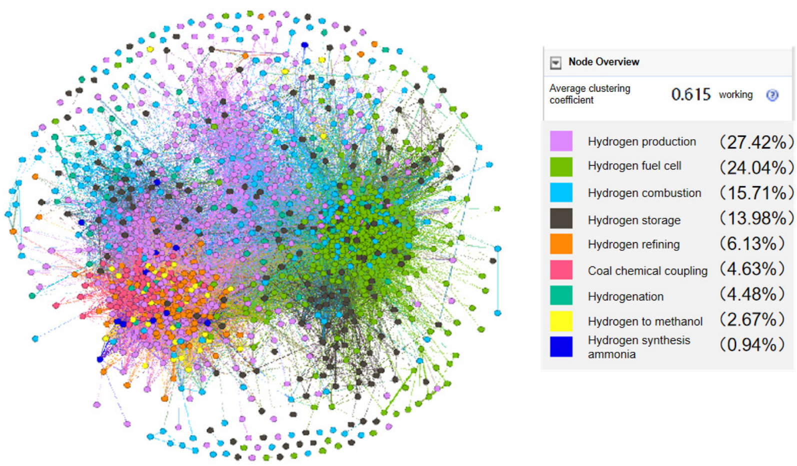

Fig. 3. Schematic diagram of national hydrogen and fuel cell industry clustering.

From the perspective of provinces, as shown in Figure 4, Beijing, Shanghai, and Jiangsu provinces rank among the top three in the national hydrogen and fuel cell cluster, with strong development momentum.

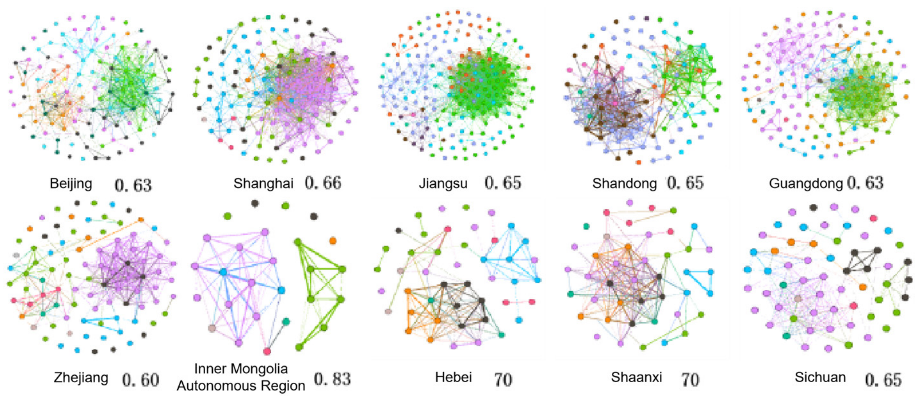

Fig. 4. Schematic diagram of hydrogen energy and fuel cell clustering in various provinces.

4.2 Analysis of financing and patent situation in the hydrogen and fuel cell industries

From the financing and patent situation of hydrogen energy and fuel cell processes, it has the following characteristics:

① The "hydrogenation" process has both high capital concentration and patent concentration TOP nodes, indicating that the development of the track is mainly led by leading enterprises.

② The "green ammonia" link is a technology driven node (with patent concentration higher than capital concentration), and the development of enterprises on this track relies on technology support.

③ Hydrogen storage, hydrogen production, hydrogen fuel cells, and hydrogen refining are capital driven industrial nodes (with higher capital concentration than patent concentration). The development of enterprises on the track relies on capital input, and the market has begun to enter a relatively balanced stage.

As shown in Table 3 and Figure 5, the concentration of the "hydrogen coupled coal chemical industry" (dual low TOP node) industry is relatively small, the enterprise base is large, and the market tends to be fragmented.

Table 3. Analysis on the Growth Rate of Hydrogen Energy and Fuel Cell Financing.

Hydrogen energy industry nodes	Growth rate
Hydrogen production	14.45%
Fuel cell	6.52%
Hydrogen coupling coal chemical industry	0.00%
Hydrogen combustion	13.27%
Hydrogen refining	0.00%
Green methanol	7.28%
Green ammonia	0.00%
Hydrogenation	69.08%
Hydrogen Storage	118.84%

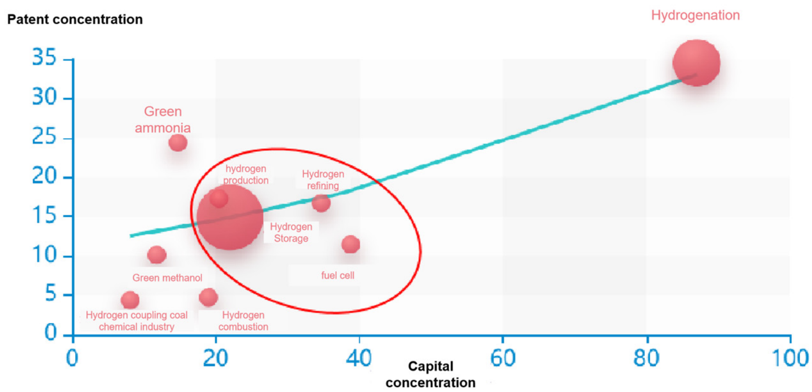

Fig. 5. Hydrogen and Fuel Cell Capital and Patent Concentration Matrix.

5 Conclusions

From the results of cluster analysis, it can also be seen that the clustering coefficients of Beijing, Shanghai, Jiangsu, Shandong, and Guangdong provinces currently rank in the top five, becoming a hot topic for local governments to attract investment and financial institutions to invest in projects.

Analyzing the reasons, considering that the hydrogen and fuel cell industries are still in the early stages of policy driven development, the rapid development of the above regions mainly benefits from policy support. In August 2021, the five ministries and commissions issued the "Notice on Starting the Demonstration and Application of Fuel Cell Vehicles" as a subsidy policy for the hydrogen energy industry. Currently, the country has formed five major urban clusters, including Beijing Tianjin Hebei, Shanghai, and Guangdong Province. Guangdong Province, Beijing, Shanghai, and other cities have simultaneously introduced local funding subsidy policies that are synchronized with national subsidies, effectively promoting the development of local hydrogen and fuel cell industries.

The clustering analysis model in this article can also be applied to other investment hotspots in the automotive industry, such as autonomous driving, intelligent cockpit, and automotive grade chips, providing quantitative analysis results for government industrial layout, financial institution investment, and helping them make more scientific decisions.

References

[1] Qiao Yingjun, Yan Jianlin, Zhong Zhihua, et al.: Research on the Transformation and Upgrading of China's Automotive Industry [J]. China Engineering Science, 2019, 21 (3): 6 DOI: 10 15302/J-SSCAE-2019 03. 001.

[2] J. Hou, J. C. Chen, H.F. Song,et al.: Are Non-R&D Innovation Activities Actually Effective for Innovation Sustainability Empirical Study from Chinese High-Tech Industry[J].MDPI,2018,11(1).

[3]atthias Carnein,Heike Trautmann:Optimizing Data Stream Representation: An Extensive Survey on Stream Clustering Algorithms,[J] Business & Information Systems Engineering. Volume 61 , Issue 3 . 2019. PP 277-297.

[4]Mashhoodi Bardia,Muñoz Unceta Pablo:Escalating environmental inequalities in larger European regions: A data mining,[J] Data in Brief. Volume 51 , Issue . 2023.

[5] Wang Zhigang.: Research on the Application of Artificial Neural Networks in Macroeconomic Forecasting [J]. Modern Economic Information, 2018 (01): 18.