Universe Scientific Publishing

DOI:10.18686/ahe.v8i6.13533

Construction of a Multi-source Data Analysis Framework Based on Artificial Intelligence

Hui Zheng

Zhangzhou Vocational Institute of Technology, Zhangzhou, Fujian, 363000

Abstract: Building a multi-source heterogeneous data framework for artificial intelligence, comprehensively understanding the current development status of artificial intelligence play a key role in showcasing the development trend of artificial intelligence and promoting disruptive development in multiple industries. On the basis of defining the artificial intelligence data resource system, the article selects journal articles, core patents, and industry reports as clue data to reveal the development of artificial intelligence, and constructs a multi-source data analysis framework based on artificial intelligence. At the same time, this paper puts forward policy recommendations to strengthen the key technologies of artificial intelligence, improve the technological innovation level of enterprises, and promote the international artificial intelligence technology linkage, so as to provide reference for improving the development level of artificial intelligence in China.

Keywords: Artificial intelligence; Multi-source data analysis; D-S evidence reasoning method

1. Introduction

As an important driving force for a new round of scientific and technological revolution and industrial change, artificial intelligence is a new science covering robotics, image recognition, language recognition, machine learning, natural language processing, computer vision, etc. Under the strong promotion of big data and algorithms, artificial intelligence has now penetrated into all sectors of society, various fields and industries, and gradually formed the artificial intelligence industry, which has become a new engine for promoting high-quality social and economic development. The development and extension of artificial intelligence can provide technical support for the development of industries, help related industries achieve transformative and subversive changes, and promote human society from the industrial age to the intelligent age. In view of this, this paper will break through the qualitative research paradigm and use multi-source data including journal papers, core patents and industry reports as research methods to explore the basic research, technological innovation and global scientific and technological layout of artificial intelligence, and put forward relevant countermeasures and suggestions.

2. Multi source data fusion theory

Multi-source data analysis refers to the process of integrating data from multi-dimensional data sources in various ways to form consistent, accurate and rich data sets. Multi-source data analysis provides more comprehensive information than a single data source, and the analysis process is more scientific. At present, multi-source data analysis mainly includes three algorithms, namely physics, parameter, and cognitive model. Among them, the D-S evidence theory inference algorithm is an inference theory of parameter class, which is a generalized extension of classical probabilistic inference, and it can effectively handle the uncertainty of research problems. Therefore, this article selects the D-S evidence reasoning algorithm (Dempster Shafer) to analyze and construct a framework for artificial intelligence.

3. Construction of a multi-source data analysis framework based on artificial intelligence

3.1 Research ideas and methods

Considering that different research samples reflect different focuses of research subjects under multi-source data, this article will use multi-source data to study artificial intelligence, so as to analyze artificial intelligence in a more comprehensive and detailed way.

Specifically, this paper takes journal papers, core patents and industry reports as research samples for this study, analyzes the development trend of artificial intelligence in the international scope by means of econometric analysis, patent analysis and content analysis, and puts forward relevant policy suggestions. In the process of research, through the analysis of high-quality journals and papers, we can better reflect the trends and achievements of basic research on artificial intelligence at home and abroad. As the key carrier of artificial intelligence technology innovation, core patents are an important representation of artificial intelligence core technology, which can provide reference for clarifying the key direction of global core technology R&D and competition in artificial intelligence. Based on the collection and analysis of industry reports, it can better infer and forecast the development prospects of the related industries and markets of artificial intelligence industry, and provide a reference basis for accurately reflecting the investment and financing, industry scale and competition situation of artificial intelligence.

3.2 Data source

Firstly, research data related to journal articles is mainly obtained from the WOS database platform. The WOS platform includes first-class academic literature covering humanities, social sciences, arts, and other fields, providing rich data materials for this study and making artificial intelligence related research more comprehensive and scientific. The article covers the period from 2009 to 2018 and a total of 156431 articles were retrieved. Due to the download limit of WOS, this article ultimately downloaded and integrated literature in units of 500 files per year. Secondly, the core patent data comes from the Derwent Innovation index database, and the data is selected by the combination of INNOGRAPHY patent strength and citation index method. After obtaining the relevant data of core patents, the research data are divided into core patent data, important patent data and general patent data according to the importance weight. In the process of research, this paper mainly takes core patent data as the main data reference of this research. Finally, the industry report data mainly comes from statistical agencies, industry associations, market research institutions, corporate financial reports and Internet data.

3.3 Multi source data preprocessing based on D-S evidence theory

First of all, due to some differences in the number of journal papers, the number of core patents and industry reports in the process of describing the development of artificial intelligence, it is necessary to evaluate the consistency of the multi-source data of these three dimensions. Based on this, under the condition of not changing the relative relationship between years, this paper uses the normalization treatment method to calculate, and the specific formula is as follows:

$$NY = \frac{Y}{Y\max} \tag{1}$$

In equation (1), NY refers to the normalized data for each year, Y represents data for each year; $Y\max$ represents the absolute value of data for each year.

Secondly, considering the differences in the peak growth rates of the number of papers, the number of core patents, and industry reports, it is necessary to eliminate the time misalignment and data mutation phenomena between multiple sources of data. Based on this, this article uses smoothing processing to correlate the years of multi-source data samples. The specific calculation method is as follows:

$$\overline{K} = \frac{W^a K}{\|W\|_1} \tag{2}$$

In equation (2), W is the weight vector; K represents the data vector covered by the smooth window; $\|W\|_1$ represents the 1-norm of the vector.

Thirdly, there are two kinds of fluctuation phenomena in the development and evolution of artificial intelligence: trend increase and trend decrease. Therefore, it is necessary to measure the basic hypothesis space constituted by artificial intelligence. The specific calculation formula is as follows:

$$S = \{s_1 = \{\text{Trend increase}\},\ s_2 = \{\text{Trend reduction}\}\} \tag{3}$$

At the same time, multi-source data constitutes the observation space, and the relevant formulas are as follows:

$$\begin{aligned} &P = \{A_1, A_2, A_3\} \\ &A_1 = \{\text{Relative growth rate of journal article length}\} \\ &A_2 = \{\text{Relative growth rate of patent quantity}\} \\ &A_3 = \{\text{Relative growth rate of industry reports}\} \end{aligned} \tag{4}$$

In equation (4), S refers to the basic hypothesis space; P is the observation space; A_1、A_2、A_3 respectively represent the relative growth rates of journal article length, the relative growth rate of patent quantity and the relative growth rate of industry reports.

On the basis of the above multi-source data processing, this paper combines the different judgment evidence existing in multidimensional data according to the D-S synthesis formula. The specific synthesis method is as follows:

$$m(h) = m_1 \oplus \cdots \oplus m_N = \frac{1}{\wedge} \sum_{\bigcap_{i=1}^{3} h_i = h} \prod_{j=1}^{N} m_j(h_i), h \subseteq S \tag{5}$$

$$\wedge = \sum_{\bigcap_{i=1}^{3} h_i \neq \otimes} \prod_{j=1}^{N} m_j(h_i) \tag{6}$$

In equations (5) and (6), m is the resultant mass function, which refers to the evaluation weights of each hypothesis.

4. Multi-source data difference analysis based on artificial intelligence

Firstly, correlation analysis based on the number of papers. Through the statistical analysis of the country distribution of artificial intelligence sample literature entries, it is found that the proportion of first authors in China and the United States is 58% of the global high-quality journal papers, which occupies a greater advantage compared with other countries. According to the annual distribution statistics of sample literatures on artificial intelligence, high-quality papers in the field of artificial intelligence worldwide have shown an increasing trend in recent years, with a growth rate of 16.23% during 2009-2018. This shows that high-level basic research on artificial intelligence is growing rapidly. Secondly, relevant analysis of core patents. After conducting country distribution statistics on the sample data of core patents in artificial intelligence, it was found that the United States holds a dominant position in the number of applications for core patents in artificial intelligence, accounting for 67% of the global patent count. Additionally, the United States is a global leader in machine learning, speech recognition, and image recognition in artificial intelligence, with a development speed far exceeding other countries. Japan ranked second in the development of artificial intelligence core patents, accounting for 17% of the global total. China ranks third in the world, not far behind Japan. It is worth mentioning that machine learning, speech recognition and image recognition as the core technologies of artificial intelligence, the United States and Japan firmly grasp these three core technologies, and occupy the technical high ground in the world. Thirdly, the relevant analysis of industry reports. According to the 2019 Global Artificial Intelligence Industry Data Report, the number of active artificial intelligence companies worldwide is 5,386, of which the United States ranks first, China ranks second, followed by the United Kingdom, Canada and India. It can be inferred that North America, Asia, and Europe are becoming the global artificial intelligence leading regions.

5. Policy recommendations

First of all, China should increase support for artificial intelligence innovation, focus on the research and development of cutting-edge core technologies in the intelligent industry, tackle high-end new technologies such as advanced machine learning, accelerate the breakthrough of basic key technologies in artificial intelligence, deeply activate the potential of artificial intelligence, and release the empowering dividend of artificial intelligence for industrial upgrading. In addition, all regions should continue to strengthen the overall planning of artificial intelligence applications, promote the deep integration of artificial intelligence and related industries, and promote the intelligent upgrading of the whole industrial chain, so as to accelerate the promotion of new industrialization. Secondly, it is necessary to strengthen the construction of computing infrastructure and industrial data platforms, continuously expand the supply of public services such as computing power and data resources, promote the research and development of core technologies in areas such as robot vision perception and key components, and promote interdisciplinary integration to cultivate more compound innovative talents in the field of artificial intelligence. Finally, China should strengthen international cooperation and technological linkage on a global scale, accelerate the construction of transnational artificial intelligence technology exchange channels, establish a global artificial intelligence technology open innovation ecosystem, and improve the connectivity of artificial intelligence technology innovation and artificial intelligence innovation benefits.

References:

[1] Fang Yanqiong, Tang Shengwei, Gu Bochuan, et al. Data Fusion Based on Binary Identifier and Improved D-S Theory [J]. Journal of Electrical Engineering, 2019, 16 (03) : 184-191.

[2] Li Li, Cui Leilei, Wu Xinnian, et al. Research on Situation of Artificial Intelligence Industry from the Perspective of a Comprehensive Chain Based on Multi Source Data[J]. Science and Technology Management Research, 201, 41 (21) : 100-111.