Research on Key Information Retrieval Method of Complex Network Based on Artificial Intelligence

Bozhong Liu(✉)

School of Electronic and Information Engineering, Guang'an Vocational Technical College, Guang'an, China
liubozhong77@163.com

Abstract. Aiming at the problem of poor retrieval accuracy and slow retrieval speed of information retrieval method based on hyperlink, a key information retrieval method based on artificial intelligence was proposed. The method was mainly divided into three steps, and each step was completed with the help of artificial intelligence. First, file information was preprocessed (information processing and information filtering), then keywords were extracted from information content, and finally semantic similarity calculation and semantic information matching were conducted to complete key information retrieval in complex networks. The results showed that the accuracy of key information retrieval method of complex network based on artificial intelligence was improved by 2.27% and the speed of retrieval was improved by 3.06 s.

Keywords: Artificial intelligence · Complex networks · Information retrieval · Keywords extraction · Semantic similarity · Semantic information matching

1 Introduction

With the passage of time, China's enterprises and institutions, the party and government departments of the number of files is increasing. In order to facilitate the preservation and access, the paper version of file information is converted into electronic version, and then stored in the department or organization's computer network system. However, these personal network systems are mostly self-organization, self-similarity, attractor, small world, scale-free, namely complex network [1]. The structure of complex network is complex, the connection mode is disordered and the nodes are various. Therefore, when people need to retrieve the information or data they need in a complex network, they will not only be slow, but also may be wrong. Therefore, many key information retrieval methods come into being in order to improve the efficiency of information retrieval. Hyperlink based information retrieval method is the most widely used, but with the emergence of more and more disordered archival data, the retrieval method is also unable to meet the actual needs of enterprises, institutions, party and government organs and other departments, and the problem of poor accuracy and slow speed of retrieval starts to appear [2]. The key information retrieval method based on artificial intelligence is proposed. The application of artificial intelligence in the key

G. Gui and L. Yun (Eds.): ADHIP 2019, LNICST 301, pp. 263–270, 2019.
https://doi.org/10.1007/978-3-030-36402-1_28

information retrieval of complex networks is one of the main research directions in the field of information retrieval. In this study, the file information was preprocessed by artificial intelligence first, then the key words were extracted by artificial intelligence, and finally the semantic similarity calculation and semantic information matching were conducted by artificial intelligence to complete the key information retrieval in the complex network. To verify the effectiveness of the method, a comparative experiment was carried out. The results show that compared with hyperlink based information retrieval method, the retrieval accuracy is improved by 4% and the retrieval speed is improved by 2.3 s

2 Information Retrieval Under Artificial Intelligence

Artificial intelligence, or AI for short, is a branch of computer science, mainly used in robot manufacturing, language recognition, image recognition, natural language processing and so on. Artificial intelligence can replace human brain to complete heavy scientific and engineering calculation, and can do it faster and more accurately than human brain [3]. The artificial intelligence technology is applied to the key information retrieval, that is, intelligent retrieval. It is based on the needs of users, and searches out the parts required by users in the large and complex network [4]. The process of intelligent information retrieval method is shown in Fig. 1.

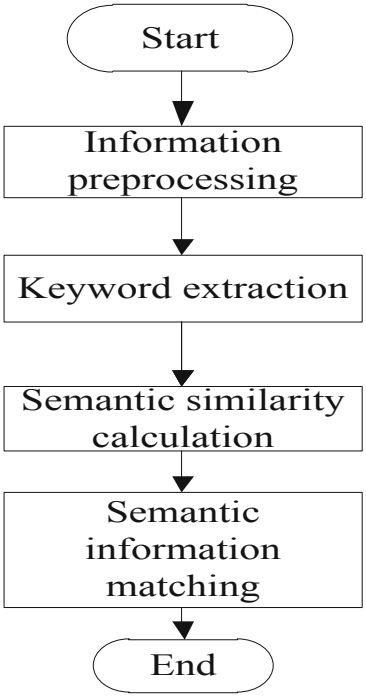

Fig. 1. Process of intelligent information retrieval method

2.1 Information Preprocessing

Information is stored at random. The information stored in the complex network will become confused over time and flow of system data, and there is a risk of being mixed in by other information. The format is mixed and cannot be read. Therefore, it is necessary to use artificial intelligence to preprocess information. Information preprocessing is a complex work, which needs two steps (information processing and information filtering) [5]. Only when the information is sorted out well, can the efficiency of subsequent key information retrieval be accelerated to achieve the goal of this research.

2.1.1 Information Processing

Information processing is mainly aimed at the problem of information format and content. The information processing process is shown in Fig. 2.

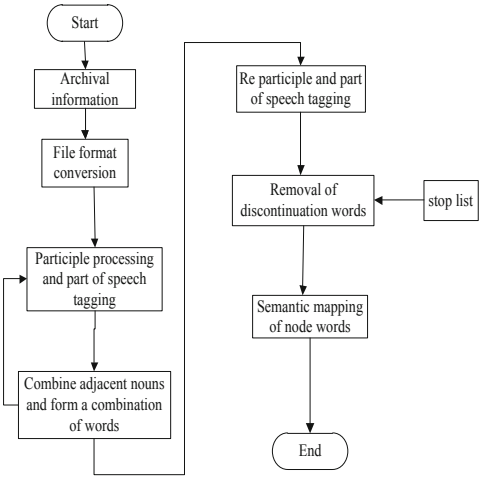

Fig. 2. Information preprocessing

2.1.2 Information Filtering

After the completion of the information processing, the information obtained is still confused, only the internal or format of the simple correction and processing. Further treatment is therefore required. This process is actually a complex set of language rules. These language rules are incorporated into information filtering software, and in the ai information retrieval system [6]. The artificial intelligence will collect all the above preliminarily processed file information, then label the existing garbage information and send it to the isolation area. The information in the isolation area will be tested and processed again. If there is a problem, it will be isolated, and if there is no problem, it will be allowed to pass. Then, with the above unannotated information, it is sorted out according to the language rules in the artificial intelligence input in advance, and finally stored in the database in an orderly way [7].

2.2 Keywords Extraction

When users search for archival information, the key words of the content to be searched must be entered in the human-computer interaction interface. Therefore, in order to improve the efficiency of information retrieval, keyword extraction of information content stored in complex networks is needed to speed up subsequent information matching and complete information retrieval [8]. The powerful computing power of artificial intelligence is used to calculate the importance of each node (words in the information), and then the nodes are sorted by importance, and a certain number of keywords are extracted according to the set requirements. The specific process is shown in Fig. 3.

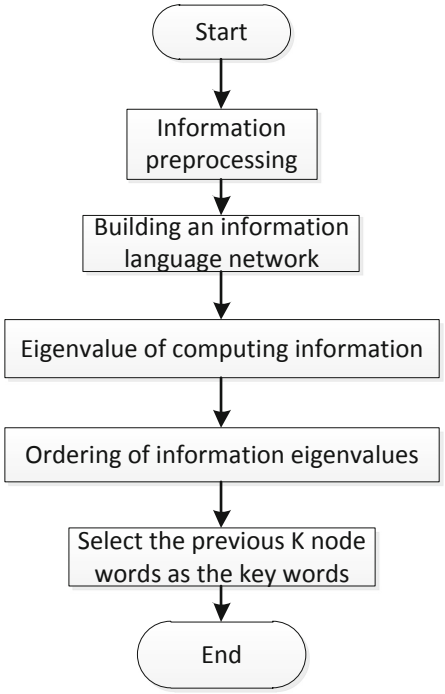

Fig. 3. Keyword extraction process

The formula in step 3 is as follows:

$$\xi = \frac{a \cdot \mathrm{B}_i}{\sum_{i=1} \mathrm{B}_i} + \frac{(1+a) \cdot C_i}{2} \tag{1}$$

In the formula, ξ is the eigenvalue; B_i is the importance of the node; C_i is the aggregation coefficient of nodes i; a is an adjustable parameter.

2.3 Semantic Similarity Calculation and Matching

After the user enters the search keywords, the ai will search similar keywords according to the keyword semantics. This keyword is extracted according to the information content in the above chapters, and then the user will present the file information content listed according to the importance in the human-computer interaction interface. In this link, there are mainly two steps: semantic similarity calculation and semantic information matching [9].

2.3.1 Semantic Similarity Calculation

Suppose the semantic similarity of the keywords E1 and E2 is, and the calculation formula is:

$$sim(E1, E2) = \alpha \cdot simA(E1, E2) + \beta \cdot simB(E1, E2) + \gamma \cdot simC(E1, E2) \quad (2)$$

$$\alpha + \beta + \gamma = 1 \quad (3)$$

The formula, $simA(E1, E2)$ is the semantic similarity of distance; $simB(E1, E2)$ is the semantic similarity based on content; $simC(E1, E2)$ is the semantic similarity based on attributes.

2.3.2 Semantic Information Matching

After the similarity calculation of keywords is completed, semantic information matching should be conducted according to the similarity. In the hyperlink based information retrieval method, the polysemy of a word cannot be distinguished, and the artificial intelligence can easily make up for this shortcoming [10]. It will, like the human brain, automatically identify and match the semantic information. For example, users use hyperlinked information retrieval methods to retrieve information about financial aspects. After entering the key word "finance" through the search interface, this method, because it completely ignores the semantic information of the search term, only USES the word "finance" to search mechanically, and finally retrieves only the meaning of the word "finance", or the related information with the word "finance". And after the use of artificial intelligence for retrieval, to "financial" a more precise definition for the term, "financial" contains "tax", "debt", "spending" and other subsystems, so the use of artificial intelligence can not only provide users with the word "finance" keywords information, also can retrieve the relevant information, to satisfy the user may need to know about the "financial" related subclass information [11–13].

3 Contrast Experiment

In order to verify the effectiveness of the key information retrieval method of complex network based on artificial intelligence, a comparative experiment was conducted. The accuracy and speed of retrieval are taken as indicators to judge the effectiveness of the method. It is assumed that there are 1,500 files in a company's file information storage system, which are divided into five types: employee personal information files (800), business files (200), financial files (250), personnel flow files (200), and company development history records (50).

Now, the key information retrieval method of complex network based on artificial intelligence and information retrieval method based on hyperlink are used to retrieve the information of 15,000 files. The results are shown in Table 1.

Table 1. Search results of 15,000 documents by two methods

		Key information retrieval method for complex networks based on artificial intelligence			Information retrieval method based on hyperlink		
Category	The number of all file letters that belong to this type	The number of correctly retrieved/share	Accuracy rate/%	Average accuracy/%	The number of correctly retrieved/share	Accuracy rate/%	Average accuracy/%
1	800	790	98.75	98.95	795	99.38	96.68
2	200	200	100		198	99	
3	250	250	100		250	100	
4	200	200	100		190	95	
5	50	48	96		45	90	

As can be seen from Table 1, the average accuracy rate is 98.95% when 1,500 files are retrieved intelligently using the key information retrieval method of complex network based on artificial intelligence. The average accuracy rate was only 96.68% when 1500 files were retrieved by hyperlink based information retrieval method. The former was 2.27% higher than the latter.

Now to verify the difference in retrieval speed between the two methods. The two methods were used to search 250 financial files. The retrieval results are shown in Table 2.

Table 2. Compare the speed of information retrieval between the two methods

Retrieval times	Key information retrieval method for complex networks based on artificial intelligence/s	Information retrieval method based on hyperlink/s
1	2.3	5.4
2	2.5	5.6
3	2.4	5.4
4	2.3	5.8
5	2.6	5.2
Average time consuming/s	2.42	5.48

As can be seen from Table 2, the average time spent in retrieving 250 financial files is 2.42 s when using the key information retrieval method of complex network based on artificial intelligence. The average time spent in retrieving 250 financial files was 5.48 s using hyperlink based information retrieval method. Compared with the latter, the former is 3.06 s shorter than the latter.

All in all, the key information retrieval method based on artificial intelligence is superior to the information retrieval method based on hyperlink in terms of both retrieval accuracy and retrieval speed.

4 Conclusions

To sum up, archival information, as an important basis for the development of enterprises, institutions, party and government departments, and its storage mode has gradually begun to develop toward informatization, networking and intellectualization. However, due to the fact that the archival information is mostly stored in the complex network with personal nature, it will be difficult to find information when users want to look up certain information, and it will be difficult to retrieve it for a while. Even if it is retrieved, the retrieved content is not accurate. A key information retrieval method based on artificial intelligence is proposed. The experimental results show that the retrieval time of this method is shorter than that of the traditional method, and the accuracy of information retrieval is higher, and the average accuracy is 98.95%. This method not only improves the accuracy of information retrieval, but also improves the efficiency of information retrieval, which provides great convenience for file information retrieval.

References

1. Zanganeh, M.Y., Hariri, N.: The role of emotional aspects in the information retrieval from the web. Online Inf. Rev. **42**(4), 520–534 (2018)
2. Ou, S., Tang, Z., Su, J.: Construction and usage of terminology services for information retrieval. J. Libr. Sci. China **42**(2), 32–51 (2016)
3. Wang, L.: Text information retrieval algorithm simulation analysis under massive data. Comput. Simul. **33**(4), 429–432 (2016)
4. Liu, J., Cao, S.: Analysis on the information retrieval experience: a perspective of flow theory. Libr. Inf. Serv. (8), 67–73 (2017)
5. Cai, J., Tang, Y.: A new randomized Kaczmarz based kernel canonical correlation analysis algorithm with applications to information retrieval. Neural Netw. Official J. Int. Neural Netw. Soc. **98**, 178 (2017)
6. Huang, Y.Y., Li, Y.: Research on cognitive information retrieval model based on the "scarcity theory". J. Intell. **35**(11), 136–140 (2016)
7. Ashwin, K.T.K., Thomas, J.P., Parepally, S.: An efficient and secure information retrieval framework for content centric networks. J. Parallel Distrib. Comput. **104**, 223–233 (2017)
8. Kralj, J., Robnik-Šikonja, M., Lavrač, N.: HINMINE: heterogeneous information network mining with information retrieval heuristics. J. Intell. Inf. Syst. **50**, 29–61 (2017)
9. Cheng, Y., Lai, M.: Research on the information retrieval model based on D-S theory. Libr. Inf. Serv. **61**(21), 5–12 (2017)
10. Liu, P., Feng, M., Ming, L.: Practical k-agents search algorithm towards information retrieval in complex networks. World Wide Web **1**, 1–21 (2018)

11. Craswell, N., Croft, W.B., Rijke, M.D., et al.: Neural information retrieval: introduction to the special issue. Inf. Retrieval J. **21**(2–3), 1–4 (2017)
12. Kishore, C.B.D.J., Reddy, T.B.: An efficient approach for land record classification and information retrieval in data warehouse. Int. J. Comput. Appl. 1–10 (2018)
13. Min, Y., Duan, J., Huang, M.: Research on information retrieval and intelligent fusion based on cloud computing environment. Modern Electron. Tech. (6), 162–164 (2018)