JOURNAL OF INFORMATION PROCESSING SYSTEMS **JIPS**

Wireless Network Health Information Retrieval Method Based on Data Mining Algorithm

Xiaoguang Guo[*]

Abstract

In order to improve the low accuracy of traditional wireless network health information retrieval methods, a wireless network health information retrieval method is designed based on data mining algorithm. The invalid health information stored in wireless network is filtered by data mapping, and the health information is clustered by data mining algorithm. On this basis, the high-frequency words of health information are classified to realize wireless network health information retrieval. The experimental results show that exactitude of design way is significantly higher than that of the traditional method, which can solve the problem of low accuracy of the traditional wireless network health information retrieval method.

Keywords

Data Mining Algorithm, Health Information, Retrieval Method, Wireless Network

1. Introduction

Elliott [1] believes that health information refers to all the information related to medical or health care, including medical knowledge, health knowledge and patient health service information. At present, wireless network has gradually become an important channel for people to obtain health information. Therefore, the health information on the wireless network presents an explosive growth trend, and the phenomenon of "information technology is difficult to find" appears [1]. Due to the diversified development of various health information content in wireless network, for example, health information contains a large number of text, audio, video, graphics, images, and other elements, and the display forms are also different. Nowadays, people's attention to health issues has expanded from disease treatment to disease prevention and self-care, and the demand for health information has increased [2].

Retrieval behavior as the intermediary between individual users and wireless network health information, in-depth research on it will help to accurately identify user needs and provide health information that meets the needs for user retrieval [3].

A data mining algorithm is a set of heuristics and calculations to create a data mining model based on data. In order to create a model, a data mining algorithm need to analyze data [4]. Data mining algorithms can analyze this valuable, applied, and analyzed information and knowledge processes hidden in many incomplete random combinations of data [5,6].

The Library, Tianjin College of University of Science and Technology Beijing, Tianjin, China (mzek92@163.com)

2. Data Mining Algorithm

There are 10 sub-algorithms in data mining algorithm. In this paper, two algorithms which can be applied to wireless network health information retrieval are described in detail, which are K-means clustering algorithm and BP neural network algorithm. The data mining algorithm flow based on the above two algorithms is shown in Fig. 1 [7,8].

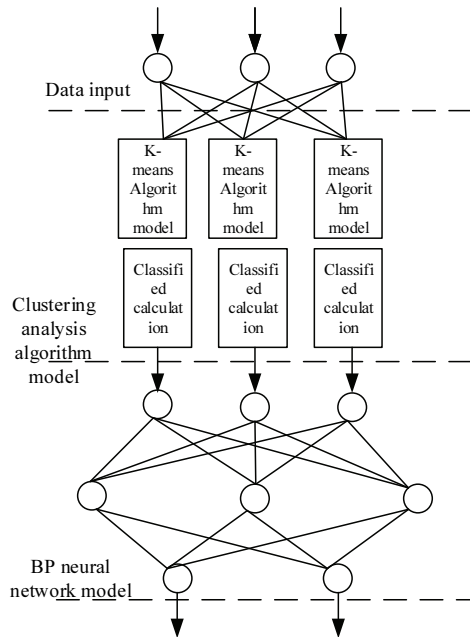The data mining algorithm flow based on the two algorithms is shown in Fig. 1.

Fig. 1. Data mining algorithm flow.

3. Wireless Network Health Information Retrieval Method based on Data Mining Algorithm

In this paper, we design a wireless network health information retrieval method based on data discovery algorithm, which can obtain a large number of health information data from the wireless network, and transform the obtained health information data into well-structured data [9,10].

3.1 Filtering Wireless Network Health Information

In the process of wireless network health information retrieval, there must be a large number of invalid information to interfere with the retrieval results, so this paper must filter the wireless network health information in advance [11]. After getting the vector, in order to filter some invalid information, we need to perform the vector mapping operation. The mapping process of wireless network health information is shown in Fig. 2.

According to Fig. 2, when mapping input, the original data stored in the wireless network and the mapping vector calculated by the master node, and the output is the mapped data [12]. The mapping vector is equivalent to a switch. Only columns with a value of 1 will be output, and other columns will be filtered. The filtered data is equivalent to the function of where statement. Whenever a line of data meets the search conditions, it will send out a logic high signal, otherwise it will send out a logic low signal [13].
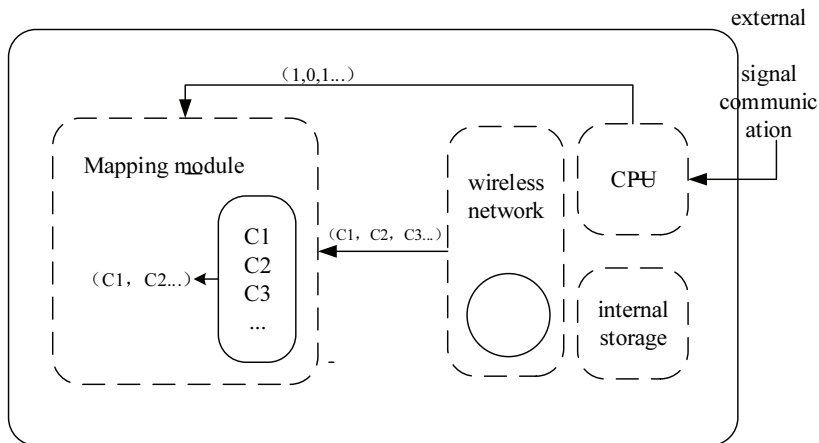

Fig. 2. Wireless network health information mapping.

3.2 Clustering Analysis of Wireless Network Health Information based on Data Mining Algorithm

Considering that the filtered wireless network health information is relatively scattered, there are some difficulties in the retrieval process [14-16]. Therefore, based on data mining algorithm, this paper uses K-means to cluster wireless network health information [17]. In this process, firstly, the K-means minimization error function of data mining algorithm is used to classify and calculate the filtered wireless network health information [18].

In the first step, K centers are randomly selected from different wireless network health information, the corresponding preprocessing is performed to get the purer health information. It's described as follows:

$$D_k = \sum_{i=1}^n d(C_i, C_i). \tag{1}$$

In the formula, D_k represents the constant for a given data set; n represents the dimension of wireless network health information.

In the second step, each different wireless network health information data point is configured to the nearest center point from the data point, and divided into sample cluster points. The points of different sample data clusters are divided into the sample cluster represented by the center which is closer to them, that is, the center which is closer to the center of the initial cluster is divided into one class. In this step, the distance formula is introduced:

$$d(x,y)^2 = \sum_{i=1}^{n}(x_i - y_i)^2. \tag{2}$$

In formula (1), $d(x,y)$ is Euclidean distance; n represents the dimension of wireless network health information; x, y represents health information of heterogeneous wireless networks [19]. According to the average value of each cluster sample wireless network health information object, the distance between each object and these central objects is calculated, and the corresponding objects are re divided according to the minimum distance, the average value of the distance from each point in each class to the central point of this class is recalculated, and each health information is allocated to its nearest central point, The center point of each wireless network health information data point in the sample cluster is used to represent the center point of the sample cluster [20]. On the basis of difference of parameter data and the data center of these clustering information can be calculated again according to the center point of different clustering information, and the minimum data calculated each time is composed of matrix D. then there is the formula (2):

$$D = \begin{cases} x11, x12 \ldots \ldots x1n \\ x21, x22 \ldots \ldots x2n \\ \ldots \ldots \\ xk1, xk2 \ldots \ldots xkn \end{cases}. \tag{3}$$

In the formula, x is the set of minimum values. According to the minimum distance, the corresponding wireless network health information is re divided.

The third step is to determine whether to carry out iterative calculation until all big data values are no longer assigned or have reached the maximum number of iterations. The fourth step is to cluster K health information data points in wireless network space. It's described as follows:

$$E = \sum_{j=1}^{m} \frac{z_j}{z} E_j. \tag{4}$$

In the formula, m represents the numbers of clusters; z_j represents the size of cluster j; z represents the total numbers of data; E_j represents the entrophy of cluster j.

3.3 High Frequency Word Frequency Classification of Wireless Network Health Information

After clustering analysis of wireless network health information based on data mining algorithm, because there are too many items involved in the actual retrieval process, it is not realistic to do correlation analysis on each search term. In order to improve its accuracy, it is necessary to count the frequency of high-frequency words of wireless network health [17]. The following table shows information.

According to Table 1, the top-10 high-frequency words of wireless network health are listed. In the retrieval process, for each retrieval instruction, we can get thousands of result URL lists, which can be sorted according to the relevance with the search words. Each result URL has a serial number, which represents the position of the URL in the return list [18]. Based on this, every 10 URLs and their summary constitute the result information of the page, so as to obtain the URL click location frequency distribution, as shown in Table 2.

Table 1. Statistical results of high frequency words of wireless network health

Serial No.	High frequency words	Word frequency
1	hospital	8,029
2	Pregnant	6,946
3	treatment	5,321
4	symptom	4,829
5	effect	4,220
6	pregnant woman	317
7	healthy	2,968
8	baby	2,439
9	medical care	2,268
10	Lose weight	2,108

Table 2. URL click location frequency distribution

Serial No.	Frequency	Percentage (%)
1	59,870	29.73
2	36,265	18.01
3	25,041	12.43
4	18,138	9.01
5	14,799	7.35
6	11,841	5.88
7	10,146	5.04
8	8,809	4.37
9	8,159	4.05
10	8,332	4.14
Total	201,400	100

According to the frequency distribution of URL click location, the word frequency classification of wireless network health high-frequency words is realized. Table 3 shows classification of high-frequency words of wireless network health.

Table 3. Classification of high-frequency words of wireless network health

No.	Theme	Related high-frequency words	Word frequency count
1	Disease	Menstruation, surgery, capsule, AIDS, ovulation, cough, hypertension, diabetes, have a fever, hemorrhagic fever, etc.	44,832
2	Healthcare	Health, massage, traditional Chinese medicine, folk prescription, nutrition, diet, Taijiquan, fitness, etc.	29,088
3	Mother and baby	Pregnancy, pregnant women, infants, babies, childbirth, milk powder, abortion, contraception, etc.	23,915
4	Medical institution	Hospitals, traditional Chinese medicine hospitals, health bureaus, pharmacies, etc.	14,198
5	Cosmetic surgery	Beauty, facial mask, plastic surgery, whitening, skin care products, breast enhancement, weight loss, etc.	8,707
6	Other	Doctors, pharmacists, physical examination, family planning, equipment, outpatient, registration, etc.	3,110

3.4 Realize Health Information Retrieval in Wireless Network

Based on the above classification results of wireless network health high-frequency words, the wireless network health information retrieval is realized based on specific retrieval behavior types [19].

Table 4 shows the types and descriptions of health information retrieval behavior. The specific implementation process is as follows: firstly, the health information of wireless network is input into BP neural network, and according to the described health information retrieval behavior type, the weights and thresholds in neural network are constantly adjusted to gradually approach the required results, so as to minimize the output error [20].

$$\Delta\omega = \eta O_K^P (1 - O_K^P)(t_K^P - O_K^P) O_i^P. \tag{5}$$

In this stage, users make comprehensive judgment, evaluation, and decision-making on health information, select the required health information, and then perform further sharing behavior. In this way, the wireless network health information retrieval is realized [21].

Table 4. Types and descriptions of health information retrieval behavior

Serial No.	Behavior type	Behavior description
1	Search entry selection behavior	Selecting initial search entry
2	-	Replacing search entry
3	Retrievable construction line	Selecting the initial search formula
4	-	Reconstructing retrieval
5	Search result browse line	Browse search results
6	-	Browse search results
7	-	Click on the web link
8	-	Back to page
9	Search result selection row	Evaluation page
10	-	Sharing information

4. Experimental Results

4.1 Analysis of Experimental Results

Table 5 shows a comparison of experimental data. According to the experimental results, the data mining algorithm is able to retrieve the wireless network health information accurately, and its accuracy is obviously much higher than that of the control group. These experimental results show that the wireless network health information retrieval method designed in this paper has practical significance, and it is necessary to promote the use.

5. Conclusion

In this paper, data mining algorithm is used to realize the classification and calculation of big data. In the whole process of data mining and calculation, it not only improves the ability of data processing, but

Table 5. Comparison of experimental data

Serial No.	Retrieval time (s)	Retrieve data	Accuracy (%)	
			Experimental group	Control group
1	20	580	93.5	82.5
2	20	580	93.6	82.5
3	20	580	93.4	82.4
4	20	580	93.5	82.0
5	20	580	93.7	82.4
6	20	580	93.6	82.2
7	20	580	93.8	82.3
8	20	580	93.6	82.4
9	20	580	93.5	82.6
10	20	580	93.7	82.2

also increases the speed of data calculation. It also enables users to improve the accuracy of data more accurately from the clustered big data, making it easy for users to grasp the law and connotation of data from the vast database, so as to realize the wireless network health information retrieval quickly and accurately. Up to now, no one in China has used this study. A real data mining in this study can show the user's health information retrieval behavior pattern from a macro perspective. Because the research is still in the initial stage, there are many deficiencies. From the perspective of research methods, the log method cannot connect the query with specific users, and can't count the relationship between demographic characteristics and retrieval behavior; from the implementation of the research steps, we cannot guarantee that the high-frequency word frequency classification can filter out all the query items related to health information. The follow-up research will focus on the changes of health information retrieval behavior of Internet users over time, in order to predict the future trend of health information retrieval, and provide further help for the improvement of search engine retrieval system and related website designers.

Acknowledgement

References

[1] B. J. Elliott, "A content analysis of the health information provided in women's weekly magazines," *Health Libraries Review*, vol. 11, no. 2, pp. 96-103, 1994.

[2] M. A. Senouci, H. Senouci, M. R. Senouci, N. Ferdosian, and A. Mellouk, "Flow/Interface association for multi-connectivity in heterogeneous wireless networks: e-Health case," *Ad Hoc Networks*, vol. 94, article no. 101942, 2019. https://doi.org/10.1016/j.adhoc.2019.101942

[3] Y. K. Ever, "Secure-anonymous user authentication scheme for e-healthcare application using wireless medical sensor networks," *IEEE Systems Journal*, vol. 13, no. 1, pp. 456-467, 2019.

[4] X. Zuo, Z. Chen, L. Dong, J. Chang, and B. Hou, "Power information network intrusion detection based on data mining algorithm," *The Journal of Supercomputing*, vol. 76, no. 7, pp. 5521-5539, 2020.

[5] M. A. Hossain, R. Ferdousi, S. A. Hossain, M. F. Alhamid, and A. El Saddik, "A novel framework for recommending data mining algorithm in dynamic IoT environment," *IEEE Access*, vol. 8, pp. 157333-157345, 2020.

[6] C. Fiarni, E. M. Sipayung, and S. Maemunah, "Analysis and prediction of diabetes complication disease using data mining algorithm," *Procedia Computer Science*, vol. 161, pp. 449-457, 2019.

[7] D. A. Wood, "Net ecosystem carbon exchange prediction and insightful data mining with an optimized data-matching algorithm," *Ecological Indicators*, vol. 124, article no. 107426, 2021. https://doi.org/10.1016/j.ecolind.2021.107426

[8] X. Shen, X. Fu, and C. Zhou, "A combined algorithm for cleaning abnormal data of wind turbine power curve based on change point grouping algorithm and quartile algorithm," *IEEE Transactions on Sustainable Energy*, vol. 10, no. 1, pp. 46-54, 2019.

[9] A. G. Regino, J. C. dos Reis, R. Bonacin, A. Morshed, and T. Sellis, "Link maintenance for integrity in linked open data evolution: literature survey and open challenges," *Semantic Web*, vol. 12, no. 3, pp. 517-541, 2021.

[10] X. Sun, "Research on time series data mining algorithm based on Bayesian node incremental decision tree," *Cluster Computing*, vol. 22(Suppl 4), pp. 10361-10370, 2019. https://doi.org/10.1007/s10586-017-1358-6

[11] S. M. Tamil and B. Jaison, "Adaptive Lemuria: a progressive future crop prediction algorithm using data mining," *Sustainable Computing: Informatics and Systems*, vol. 31, article no. 100577, 2021. https://doi.org/10.1016/j.suscom.2021.100577

[12] J. K. Sarmah and S. Baruah, "A comparative study on use of data mining algorithm in EDM development," *IOSR Journal of Computer Engineering*, vol. 23, no. 1, pp. 44-48, 2021.

[13] Y. V. Khivintsev, A. V. Kozhevnikov, G. M. Dudko, V. K. Sakharov, Y. A. Filimonov, and A. G. Khitun, "Spin waves in YIG-based networks: logic and signal processing," *Physics of Metals and Metallography*, vol. 120, pp. 1318-1324, 2019.

[14] K. Peng, R. Jiao, J. Dong, and Y. Pi, "A deep belief network based health indicator construction and remaining useful life prediction using improved particle filter," *Neurocomputing*, vol. 361, pp. 19-28, 2019.

[15] P. Santos, A. C. Oliveira, D. Beirao, C. Franclim, M. Magalhaes, L. Couto, L. Sa, and A. Hespanhol, "The Metis project—education for health: a project report," *Universal Access in the Information Society*, vol. 20, no. 2, pp. 417-420, 2021.

[16] L. Abualigah and A. J. Dulaimi, "A novel feature selection method for data mining tasks using hybrid sine cosine algorithm and genetic algorithm," *Cluster Computing*, vol. 24, pp. 2161-2176, 2021.

[17] M. Subramaniam, A. Kathirvel, E. Sabitha, and H. A. Basha, "Modified firefly algorithm and fuzzy c-mean clustering based semantic information retrieval," *Journal of Web Engineering*, vol. 20, no. 1, pp. 33-52, 2021.

[18] F. D. Bortoloti, E. de Oliveira, and P. M. Ciarelli, "Supervised kernel density estimation K-means," *Expert Systems with Applications*, vol. 168, article no. 114350, 2021. https://doi.org/10.1016/j.eswa.2020.114350

[19] P. J. Messino, H. Kharrazi, J. M. Kim, and H. Lehmann, "A method for measuring the effect of certified electronic health record technology on childhood immunization status scores among Medicaid managed care network providers," *Journal of Biomedical Informatics*, vol. 110, article no. 103567, 2020. https://doi.org/10.1016/j.jbi.2020.103567

[20] S. Grossberg and S. Grossberg, "Classical and instrumental learning by neural networks," *Studies of Mind and Brain*. Dordrecht, Netherlands: Springer, 1982, pp. 65-156.

[21] Z. Qian and H. Xia, "Data mining algorithm based on feature weighting," *Journal of Computational Methods in Sciences and Engineering*, vol. 19(S1), pp. 269-276, 2019.

Xiaoguang Guo https://orcid.org/0000-0002-9595-3968

He is a lecture of Tianjin College, University of Science and Technology Beijing. He works in the library of Tianjin College.