

Large scale hotel resource retrieval algorithm based on characteristic threshold extraction

Min Fang,
Department of Hotel Management, Tourism College of Zhejiang,
Hangzhou, 310000
China

Received: June 25, 2021. Revised: December 3, 2021. Accepted: January 2, 2022. Published: January 3, 2022.

Abstract—At present, the hotel resource retrieval algorithm has the problem of low retrieval efficiency, low accuracy, low security and high energy consumption, and this study proposes a large scale hotel resource retrieval algorithm based on characteristic threshold extraction. In the large-scale hotel resource data, the mass sequence is decomposed into periodic component, trend component, random error component and burst component. Different components are extracted, the singular point detection is realized by the extraction results, and the abnormal data in the hotel resource data are obtained. Based on the attribute of hotel resource data, the data similarity is processed with variable window, the total similarity of data is obtained, and the abnormal detection of redundant resource data is realized. The abnormal data detection results and redundant data detection results are substituted into the space-time filter, and the data processing is completed. The retrieval problem is identified, and the data processing results are replaced in the hotel resource retrieval based on the characteristic threshold extraction to achieve the normalization of data source and rule knowledge. The characteristic threshold and retrieval strategy are determined, and data fusion reasoning is carried out. After repeated iteration, effective solutions are obtained. The effective solution is fused to get the best retrieval result. Experimental results showed that the algorithm has higher retrieval accuracy, efficiency and security coefficient, and the average search energy consumption is 56n J/bit.

Keywords—Data sets, data processing, feature threshold extraction, hotel resources, retrieval method.

I. INTRODUCTION

HOTEL resources can find the location of the hotel through web or mobile phone retrieval, integrate the hotel information of various places, and provide users with the search service of hotel information and location or destination surrounding information [1]. Through the hotel information

search service system, the hotel manager can realize the online service of the hotel management, and can make the local users subscribe to the room through the network, widening the sales channel of the hotel room [2]. It can be seen that hotel resource retrieval is indispensable to hotels or guests, which has attracted the attention and research of relevant personnel.

Liu and Liu put forward a resource retrieval algorithm based on Lagrange [3]. Experimental results show that the algorithm can effectively increase the accuracy and retrieval efficiency of multimedia resource classification and retrieval, but it has higher energy consumption. Hu *et al.* put forward a cloud based resource discovery algorithm based on cloud peer to peer network [4]. The algorithm realizes the fast location of multi attribute cloud resources. It does not produce larger query delay with the increase of network node and type dimension, but the retrieval precision is low. Jin *et al.* put forward a resource search algorithm based on cuckoo search and multi priority list dynamic programming [5]. This algorithm can get higher platform resource utilization rate, but the retrieval efficiency is low. Zhao *et al.* put forward a resource scheduling algorithm based on improved quantum particle swarm optimization [6]. The algorithm greatly saves the task completion time, but does not detect and remove abnormal data in the network, resulting in poor security of the algorithm.

II. LARGE SCALE HOTEL RESOURCE RETRIEVAL

In the current large-scale hotel resource retrieval security is poor, because of this, the feature threshold extraction algorithm is a local feature detection algorithm, can be used to detect and describe the local features in the image, quickly realize the detection of abnormal data. Therefore, this paper introduces the feature threshold extraction algorithm into this field, and designs a hotel resource retrieval algorithm based on feature threshold extraction to identify retrieval problems and replace them with data processing results, so as to realize the standardization of data source and rule knowledge. The feature threshold and retrieval strategy are determined, and the data fusion reasoning is carried out. After repeated iterations, an effective solution is obtained. The effective solutions are fused

to get the best retrieval results.

A. Anomaly Detection and Cleaning of Hotel Resource Data

The time series of the hotel resource data can be divided into four components, periodic component, trend component, burst component and random error component, and each component has different characteristics. First, the four components are extracted, and the singular points are detected according to the feature extraction results. Then, the four components are combined to detect outliers. Fig. 1 is the flow chart of the hotel resource anomaly detection.

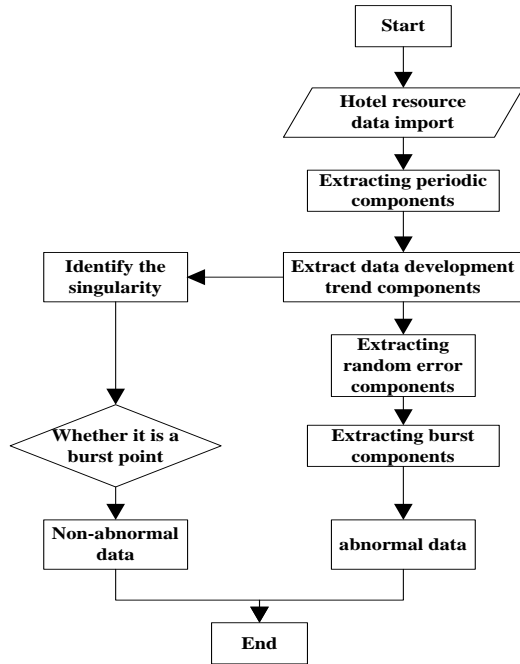


Fig. 1 Anomaly detection of hotel resources

According to Fig. 1, periodic components will be extracted from hotel resource data. The extraction of cyclical components of hotel resource data is helpful to detect outliers in hotel resource data. The feature extraction of periodic components in hotel resource data time series is to determine the period of time series. The general methods to determine the period are Fourier transform, wavelet transform, difference calculation and so on [7], [8]. Based on the characteristics of time series and the method of large data processing, a new method is used to determine the cycle L' . In this paper, we use (1) to calculate the time series $X' = \{x_1, x_2, \dots, x_m\}$ and get the matrix A' . Among them, the expression of A' is:

$$A' = \begin{bmatrix} x_2 - x_1 & x_3 - x_2 & \cdots & x_{n-1} - x_{n-2} \\ x_3 - x_1 & x_4 - x_2 & \cdots & x_n - x_{n-2} \\ \vdots & \vdots & \vdots & \vdots \\ x_m - x_1 & x_{m+1} - x_2 & \cdots & x_{n+m-3} - x_{n-2} \end{bmatrix} \quad (1)$$

Each line of the matrix A' is linearly fitted. The parameters are recorded as $(a'_1, a'_2, \dots, a'_{m-1})$ and $(b'_1, b'_2, \dots, b'_{m-1})$ respectively. Each row of the A' 's subscript is replaced in the corresponding linear equation $Y' = a'N' + b'$, and the

corresponding A'' is obtained. The expression of A'' is:

$$A'' = a'N' + b' = \begin{bmatrix} a'_1 \\ a'_2 \\ \vdots \\ a'_{m-1} \end{bmatrix} \cdot \begin{bmatrix} 1 & 2 & \cdots & n-2 \\ 1 & 2 & \cdots & n-2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 2 & \cdots & n-2 \end{bmatrix} + \begin{bmatrix} b'_1 \\ b'_2 \\ \vdots \\ b'_{m-1} \end{bmatrix} \quad (2)$$

Based on the above calculation, least squares method is used to calculate the minimum error per row of A' and A'' . The number of lines with the minimum error is L' :

$$L' = \begin{bmatrix} (x_2 - x_1 - a'_1 - b'_1) + \cdots + (x_{n-1} - x_{n-2} - (n-2)a'_1 - b'_1)^2 \\ (x_3 - x_1 - a'_2 - b'_2) + \cdots + (x_n - x_{n-2} - (n-2)a'_2 - b'_2)^2 \\ \vdots \\ (x_m - x_1 - a'_m - b'_{m-1}) + \cdots + (x_{m+n-3} - x_{n-2} - (n-2)a'_{m-1} - b'_{m-1})^2 \end{bmatrix} \quad (3)$$

The correct extraction of periodic component features is the precondition for feature extraction of trend components [9], [10]. The periodic L' is extracted by the periodic component features, and the time series X' is divided into matrix B' by periodic L' , in which $x_{M'}$ is the last data of the time series $x_{M'}$, and the data after $x_{M'}$ are all empty $N'A$. Among them, the expression of matrix B' is:

$$B' = \begin{bmatrix} x_1 & x_2 & \cdots & x_i & \cdots & x_{L'} \\ x_{L'} & x_{L'+2} & \cdots & x_{L'+2} & \cdots & x_{2L'} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{M' \times L'+1} & x_{M' \times L'+1} & \cdots & x_{M'} & N'A & \cdots \end{bmatrix} \quad (4)$$

The matrix B' has the characteristic coordinates of two directions, and all points in the same line represent all the sets of time points in the same period, and all points in the same column represent a set of points in the same position in different periods. Each column of matrix B' is sequentially removed and a total of L' time series are obtained, respectively, as $L'_1 \sim L'_L$. Among them, each L'_i sequence has M' data (the data after $x_{M'}$ are all null values $N'A$).

All points in sequence L'_i are the set of points in the same position within the sequence X' , which excludes the impact of burst components and random errors, and has the same distribution characteristic in theory [11], [12]. If the trend component of the sequence L'_i is fixed, then all points in the L'_i are on the same horizontal line, and all points outside the line are considered singularities. If the trend component of the sequence L'_i is distributed according to a certain law, the point which is not distributed according to the law is regarded as a singular point; For example, the trend component of the L'_i sequence is random distribution. It is necessary to find out the random distribution range $[\min, \max]$, and all points beyond this range are singular points.

Through the above analysis, we can get the feature expression of trend component.

$$F(e') = L'_i \cdot y'_{[\min, \max]} \quad (5)$$

Among them, $F(e')$ represents the trend component eigenvalues of hotel resource data.

The random error component R' is extracted by the characteristic value $F(e')$ of the trend component of the hotel resource data, and the distribution function of the random error component is determined according to the actual situation of the time series. The specific process is as follows:

Each row of matrix B' is extracted successively, and a total of M' time series are obtained. They are respectively recorded as $M'_1 \sim M'_{M'}$, and each M' sequence has L' data. Statistical analysis of sequence X' results in the distribution model $X' \sim F(x)$ of sequence X' . For all sequence M' , a new set of sequence $M''_1, M''_2, \dots, M''_{M'}$ can be obtained according to the $F(x)$ distribution model, the newly generated sequences are sorted from small to large, and then the elements from the original sequence are sorted from small to large and decrease in sequence with the elements in the sequence $M''_1, M''_2, \dots, M''_{M'}$, and the M' random error sequence is recorded as $R'(i)$. According to $R'(i)$, the random error distribution model of each sequence mm can be obtained as $F(e')_i$. Among them, the expression of $F(e')_i$ is:

$$F(e')_i = \sum p_i(r) \quad (6)$$

In the formula, $p_i(r)$ represents the probability of occurrence of random errors of sequence M' .

According to all $R'(i)$, the random error distribution model $F(x)_e$ of sequence X' can be obtained:

$$F(x)_e = \sum p(r) \quad (7)$$

Among them, $p(r)$ represents the probability of random error occurrence of sequence X' . The result R' (7) is the distribution range of random error of sequence X' , which is very useful for judging whether singularity is abnormal. The random error component R' can be obtained by using (6) and (7):

$$R' = \frac{F(x)_e}{F(e')_i} \quad (8)$$

The feature of burst component in hotel resource data is the key step to judge whether the singular point is abnormal. According to the calculation of the random error component R' , the eigenvalues of the burst point components in the hotel resource data can be obtained.

$$B'(x) = R' \cdot (sum(i) - sum(i)') \quad (9)$$

$B'(x)$ represents the eigenvalue of the burst point component in the hotel resource data, $sum(i)$ represents the

sum of the difference square between the sequence X' and the M' , and the $sum(i)'$ represents the random error of the M' .

After obtaining the characteristic value $B'(x)$ of the burst point component in the hotel resource data, we determine whether the singular point e' belongs to the $B'(x)$ or random error component R' range. If so, the singular point is not abnormal data; Otherwise, singular points are abnormal data, which is obtained by analyzing the properties of singular points. Processing them can improve the processing effect of abnormal data in hotel resource data and the accuracy of experimental comparison.

Through the above analysis and calculation, the abnormal data in hotel resource data can be obtained, and then the data cleaning method is used to realize the cleaning of abnormal data and redundant data.

The single data attribute of the selected hotel resources is too dependent on attributes, which directly affects the efficiency and accuracy of redundant data. Selecting all data attributes will take a long time on the one hand. On the other hand, some attributes will have a negative effect on the accuracy of redundant data [13], [14]. In order to solve the above problems, the space and punctuation symbols are used as segmentation symbols, the characters corresponding to the character strings are divided into words, the words are sorted, and then the sequence of characters is sorted. The sorting results are put in a variable window for similarity processing. When processing, we need to set three parameters: the minimum window W'_{\min} , the maximum window W'_{\max} , and the minimum threshold T'_{\min} . Set the window's initial value to W'_{\min} , match the hotel resource data r_i'' and the records in the window. Using the 3-gram method to split the string, the similarity between r_i'' and the record in the window is calculated.

$$same_i = sim(r_i'' \oplus r_i''') \quad (10)$$

Among them, r_i'' represents the record in the window, \oplus represents the matching symbol, and $same_i$ represents the similarity between the r_i'' and the r_i''' in the window, and uses a two-dimensional array to store the similarity values.

When matching to $r_i'' W'_{\min}$, if $sim(r_i'' \oplus r_i''') > T'_{\min}$, the window W' will expand, and r_i'' will continue to match the next record. When the similarity is $same < T'_{\min}$ or window $W' > W'_{\max}$, the match ends.

Through the above calculation, the total similarity is calculated, and the contribution of different fields in the total similarity record to the record features is not consistent. According to the importance of attributes, attribute ω_i of attribute C'_i is given. Thus, the total similarity value $Tsim(Z)$ can be obtained.

$$Tsim(Z) = \frac{\sum \omega_i \cdot same_i}{\sum_{i=1} \omega_i} \quad (11)$$

Among them, $\omega_1 + \omega_2 + \dots + \omega_j = 1$ can only be compared with fields when the corresponding values of the two records on the i attribute are not empty. Set the similarity threshold μ_i , if $Tsim(Z) > \mu_i$, it means two data duplication.

The anomaly detection results and redundant data are introduced into the space-time Buron filter to complete the hotel resource data processing. The hotel resource data set filtered by filters can be expressed as:

$$D = \frac{TSBF}{(Tsim(Z) + B'(x))} \cdot \lambda \quad (12)$$

Among them, D represents the hotel resource data set filtered by the filter, and λ represents the filtration coefficient.

The above is mainly for the cleaning of abnormal data and redundant data in large hotel resources. In this process, we first detect the abnormal data and redundant data in the hotel resource data, and then use spatiotemporal Buron filter to filter the abnormal data and redundant data to get the processed data.

B. Large Scale Hotel Resource Retrieval Based on Characteristic Threshold Extraction

Because the characteristic threshold extraction can analyze the table or view data in the data source, remove irrelevant redundant data, extract the effective data from the database, and convert it into its own recognizable format, it has strong retrieval advantages. Therefore, according to the cleaning results of abnormal data and redundant data in 2.1, this paper adopts the characteristic threshold extraction to realize the retrieval of hotel resources. The specific steps are as follows.

Data representation and problem recognition: in the fusion of unitary space θ data source $D(N'', Q')$, the case library CB and rule base RB are generated. The target problem is identified, the target problem set C_q is set up, the case unitary matrix θ' , the regular unitary matrix S'' , the effective solution x^* of the hotel resource retrieval and the transfer function matrix $T(X)$ are emptied.

Fusion of unitary space mapping and normalization: uniform normalization of case data and rule knowledge V_{ij} in data source $D(N'', Q')$:

$$V_{ij}^D = \frac{(V_{ij} - \min V_{ij})}{(\max V_{ij} - \min V_{ij})} \quad (13)$$

Among them, V_{ij}^D represents normalization results, $\min V_{ij}$ represents minimum rule knowledge, and $\max V_{ij}$ represents minimum rule knowledge.

The case unitary matrix θ' is obtained by using orthogonal discriminant method in case set.

$$\theta' = \begin{bmatrix} r_{11} & r_{21} & r_{31} \\ r_{12} & r_{22} & r_{32} \\ r_{13} & r_{23} & r_{33} \end{bmatrix} \quad (14)$$

Based on the decision tree rules and the rule unitary matrix S'' , the fusion unitary space θ of fusion reasoning is formed:

$$\theta = V_{ij}^D \cdot \theta' \cdot S'' \quad (15)$$

According to the calculation by (15), the characteristic threshold and retrieval strategy are determined. Using $T(X)$ to calculate the characteristic threshold vector σ of hotel resource data:

$$\sigma = \log \theta \cdot T(X) \quad (16)$$

After getting the characteristic threshold vector σ of the hotel resource data, the relationship between the target problem C_q and the meta knowledge in the data knowledge base is defined in the θ , and then the fusion reasoning strategy is formulated.

In spatial θ , the reasoning is repeated and iterated until the effective solution is obtained. Among them, the effective solution process is as follows:

First, the knowledge database is retrieved, and the knowledge database is identified, classified and optimized in turn. The three threshold values are the matching degree $S(a)$, the rule granularity \mathcal{G} and the inference confidence CF , respectively. The expression of the three is as follows:

$$S(a) = \sigma \cdot C_q \quad (17)$$

$$\mathcal{G} = S(a) \cdot \sigma \quad (18)$$

$$CF = \mathcal{G} \cdot \sigma \quad (19)$$

According to the actual value of $S(a)$, \mathcal{G} and CF , the corresponding strategy is implemented, and the decision fusion confidence $CF(x^*)$ of effective solution x^* is obtained:

$$CF(x^*) = \sum_{X'_1 \cap X'_2 = x^*} CF(X'_1) \cdot CF(X'_2) \quad (20)$$

The fusion reasoning control is based on the relevant feedback information, and the retrieval reasoning process is activated repeatedly, and the results are refined until the effective solution is finally obtained. The final expression of effective solution is as follows:

$$best(x^*) = \inf_{\sigma=1} [CF(X'_i)] \cdot \zeta \quad (21)$$

So far, large hotel resource retrieval based on feature threshold extraction can be realized, and the method design of this paper can be completed.

III. EXPERIMENTAL RESULTS AND ANALYSIS

In order to verify the overall performance of the large-scale hotel resource retrieval algorithm based on feature threshold extraction, the multimedia cloud resource classification retrieval method based on Lagrange algorithm and the multi-attribute cloud resource retrieval algorithm based on cloud P2P network are more popular at present [3], [4].

Therefore, this paper selects the above methods to compare with this method, the experiment is configured to 2.20GHz i7-4702MQ CPU, 12 GB memory, 1 TB hard disk, and the Windows 10 system. MATLAB experiment box is used to implement the proposed algorithm.

The experimental results are as follows:

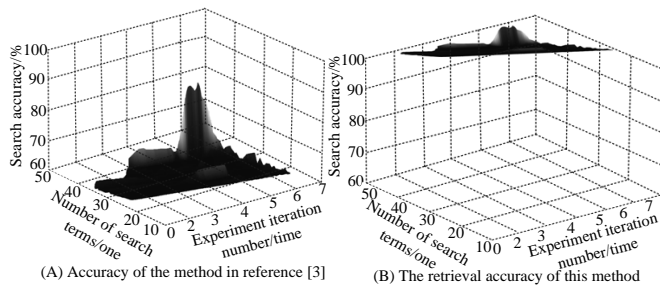


Fig. 2 Comparison of retrieval accuracy of different methods

It can be seen from Fig. 2 that the highest accuracy of the method in [3] is 75%, and the highest accuracy of the method in this paper is 98%. The results show that the design method in this paper has certain advantages. In order to further verify the effect of the design method in this paper, continue to compare the retrieval time of different methods, and the comparison results are shown in Fig. 3.

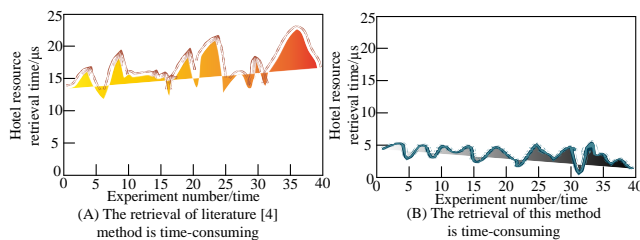


Fig. 3 Retrieval time of different methods

As can be seen from Fig. 3, the method in [4] takes a long time, which shows that the retrieval efficiency of the algorithm is low. The average retrieval time of the design method in this paper is short, which shows that the algorithm is effective. Experimental results show that the algorithm has high retrieval efficiency. Then comprehensively compare the retrieval energy consumption of the three methods, and the results are shown in Fig. 4.

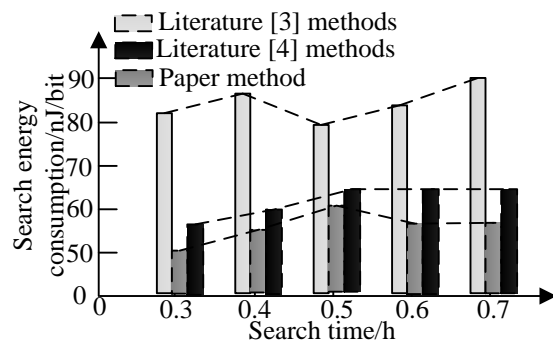


Fig. 4 Comparison of retrieval energy consumption of different methods

According to Fig. 4, compared with the other two methods, the retrieval energy consumption of this method is less, which can use less energy while ensuring the retrieval accuracy. Finally, the retrieval security of the three methods is compared,

and the results are shown in Fig. 5.

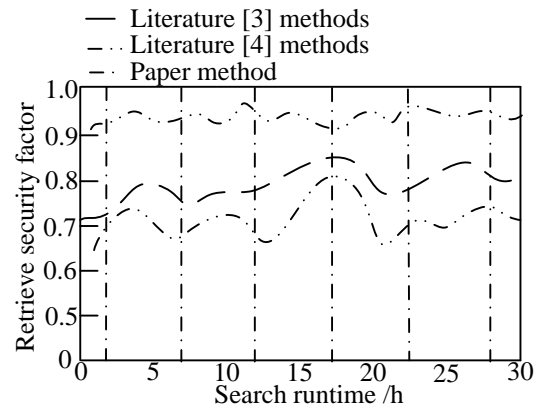


Fig. 5 Safety comparison of different methods

According to Fig. 5, the other two methods can not effectively remove abnormal data, resulting in low scheduling safety factor. The security of this method is high, with an average of more than 90%, which realizes the dual security of resource retrieval security, and can make the algorithm more secure.

According to the above analysis, this method is superior to the existing algorithms in terms of retrieval accuracy, retrieval efficiency, retrieval energy consumption and retrieval security.

IV. CONCLUSION

There are many kinds of hotel resources, so it is very difficult to achieve certain retrieval accuracy. At present, there are many problems in the related algorithms of information data retrieval, and a large-scale hotel resource retrieval algorithm based on characteristic threshold extraction is proposed. In order to improve the retrieval precision, efficiency and security by cleaning the abnormal data and redundant data in the hotel resource data, the retrieval energy consumption is reduced, and the retrieval security factor is introduced to further improve the security of resource retrieval. However, this algorithm does not classify the hotel resource data in detail, which may lead to the deviation of the retrieval results. Therefore, it needs to be tested in more data types in the future to improve the detection performance of the algorithm.

References

- [1] L. Feng, Y. Kuang, and X. Fu, "Energy-efficient network cooperation joint resource configuration in multi-RAT heterogeneous cognitive radio networks," *Electron Lett*, vol. 52, no. 16, pp. 1414-1416, 2016.
- [2] P. Yi, H. Ding, and B. Ramamurthy, "Budget-optimized network-aware joint resource allocation in Grids/Clouds over optical networks," *J Lightwave Technol*, vol. 34, no. 16, pp. 3890-3900, 2016.
- [3] P. Liu, and C. Liu, "Classification retrieval method for multimedia cloud resources based on Lagrange algorithm," *Journal of Shenyang University of Technology*, vol. 39, no. 4, pp. 433-437, 2017.
- [4] K. Hu, S. P. Chen, and Y. H. Pu, "Multi-attribute cloud resources range-query algorithm based on hierarchical P2P

- cloud network,” *Application Research of Computers*, vol. 33, no. 6, pp. 1822-1825, 2016.
- [5] Y. D. Jin, H. H. Yang, and P. F. Duan, “Approach to operation task and platform resource matching based on CS and MPLDS,” *Computer Simulation*, vol. 34, no. 2, pp. 5-9, 2017.
- [6] Y. Zhao, X. B. Hui, and J. H. Xu, “Cloud computing resource scheduling strategy based on improved RAQPSO algorithm research,” *Journal of Air Force Engineering University (Natural Science Edition)*, vol. 17, no. 6, pp. 70-75, 2016.
- [7] B. Xiong, and K. Hofmann, “Binary search algorithm for adaptive impedance matching network,” *Electronics Letters*, vol. 52, no. 9, pp. 714-716, 2016.
- [8] B. Wahn, J. Schwandt, and M. Krueger, “Multisensory teamwork: using a tactile or an auditory display to exchange gaze information improves performance in joint visual search,” *Ergonomics*, vol. 59, no. 6, pp. 781-795, 2016.
- [9] D. Gupta, and K. Berberich, “Diversifying search results using time: An information retrieval method for historians,” *Pattern Recognition*, vol. 11, no. 2, pp. 133-142, 2016.
- [10] T. Y. Xu, H. R. Ren, and G. B. Zhang, “The construction and implementation of image retrieval framework for digital library - Based on the non sampling contourlet transform,” *Modern Information*, vol. 37, no. 6, pp. 55-60, 2017.
- [11] A. X. Zhao, “Design and implementation of system for specific data search in massive question banks,” *Modern Electronics Technique*, vol. 39, no. 20, pp. 49-52, 2016.
- [12] M. Hu, M. H. Dai, and H. Chen, “Personalized retrieval method based on user profile,” *Application Research of Computers*, vol. 33, no. 2, pp. 417-420, 2016.
- [13] Y. Z. Zhang, and J. J. Li, “Research on movie resources ontology construction based on social tags,” *Library and Information Service*, vol. 60, no. 12, pp. 130-138, 2016.
- [14] P. Yang, Y. K. Li, and Y. X. Hu, “Remote sensing image retrieval based on tolerance granular computing theory,” *Remote Sensing for Land & Resources*, vol. 29, no. 4, pp. 43-47, 2017.

Management from Hong Kong Polytechnic University in 2006. She is currently serving as the director of the Cooperative Development Division at Tourism College of Zhejiang. Her research areas mainly include: the sustainable development of tourism and rural revitalization through tourism. In the past five years, she has completed more than 20 research projects as the Principal Investigator and published more than 10 academic papers.

Author Contribution:

This paper studies the large-scale hotel resource retrieval algorithm based on feature threshold extraction. Min Fang proposed the algorithm, extracted different components, detected singular points according to the extraction results, obtained the abnormal data in the hotel resource data, standardized the data source and rule knowledge, determined the feature threshold and retrieval strategy, conducted data fusion reasoning, obtained the best retrieval results and conducted experimental analysis, she is a contributor to all the work of this manuscript without any disputes.

Creative Commons Attribution License 4.0 (Attribution 4.0 International, CC BY 4.0)

This article is published under the terms of the Creative Commons Attribution License 4.0

https://creativecommons.org/licenses/by/4.0/deed.en_US



Min Fang, female, born in October 1977, and she is an associate professor. She graduated from Hangzhou College of Commerce with a Bachelor of Management majoring in Tourism Management in 2003, and received the Master of Science (MSC) in Hotel and Tourism