

## Research on The Use of Big data in Anorectal Branch

Niu Mingliao<sup>1,2</sup>, Liu Yuyan<sup>1,2</sup>, Zhen Huanhuan<sup>1,2</sup> and Xi Zuowu<sup>1,2</sup>

<sup>1</sup> Henan Province Hospital of TCM, 450002.Zhengzhou, Henan, China

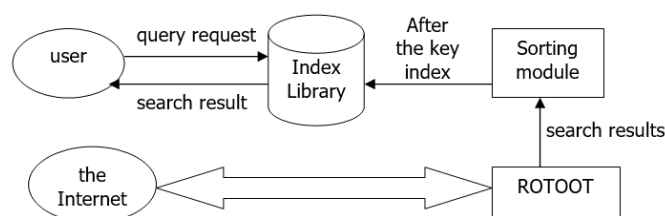
<sup>2</sup> The Second Affiliated Hospital of Henan University of Traditional Chinese Medicine, 450002.Zhengzhou, Henan, China

**Abstract:** The era of big data to the community more than just data becomes large, increased resources, more is to change the mindset and the consequent data mining technology continue to make breakthroughs, the rapid development of the ability to use the data. In this context, all the valuable data will soon be refined and used. This paper studies the big data in Hospital anorectal branch. Focused on the image segmentation and recognition as well as the establishment of a system of self-diagnosis and treatment advice on this basis. And comparative analysis of the system and physician clinical diagnostic results show that the system is rich in content, to show the diversity, adequate diagnostic index has further expanded its scope of application value.

**Keywords:** Image thinning, algorithm, Medical Imaging, edge analysis.

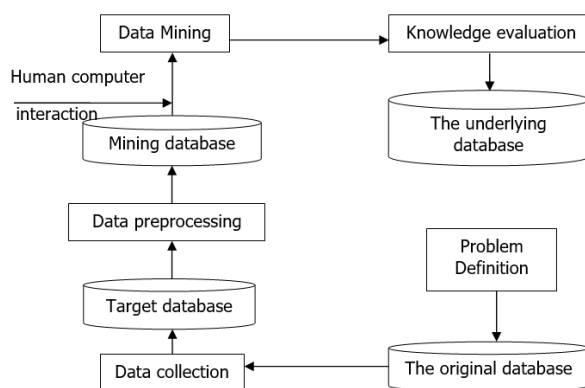
### 1 Big data and Data Mining

Big data general by the Internet module, Robot module, order module, the index database and query service thereof. Internet module is a module integrated information generated is a big information system. Robot modules are found in the Internet database documentation module. Sorting module is to extract information by keyword by the line sort, to sort the results can be fed to the data in order to back the use of data mining technology on-line information through data mining and extraction. Indexing module is to get the sort of information input to facilitate the user's query[1]. Discover service module is user-friendly query and use. Information Retrieval System database extract basic operating modes is shown in Fig1.



**Figure 1.** Extract Database Information Retrieval System Schematic Diagram of The Basic Work

The main goal of data mining is to focus on the acquisition of knowledge from big data, which determines that it is simply impossible to put a black box data can be obtained in a closed process user information knowledge. But there are many steps connected to each other, repeating the process of human-computer interaction.



**Figure 2.**Data Mining Technology

Data mining process generally includes problem definition, data collection, data preprocessing, data mining, knowledge evaluation process.

(1) Define the problem: The problem is usually defined by the data mining mutual cooperation among professionals, is scheduled for the purpose of data mining, data mining which determines the meaning and effect.

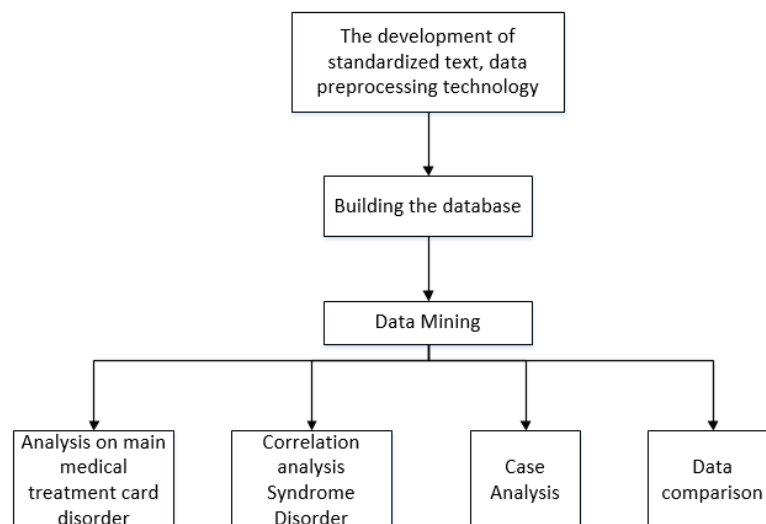
(2) Data Collection: Data collection is to find business-related data and user information, which includes the format of the data, the data to be representative, and other relevant data of three parts.

(3) Data preprocessing: data preprocessing is to extract relevant data from the target database and then processing the data through the line to remove redundancy, eliminate noise, make up the defect data. So that the data integrity to achieve.

(4) Data mining, data mining is already on the pre-processed data through the line of data mining, information extraction, and related points.

(5) Evaluation of knowledge, knowledge that is appraisal of the value of information extraction processing will be evaluated to determine the Rules of its information, so the data is fed to the base knowledge base, and eventually to a graphical user interface in the form of user-submitted information.

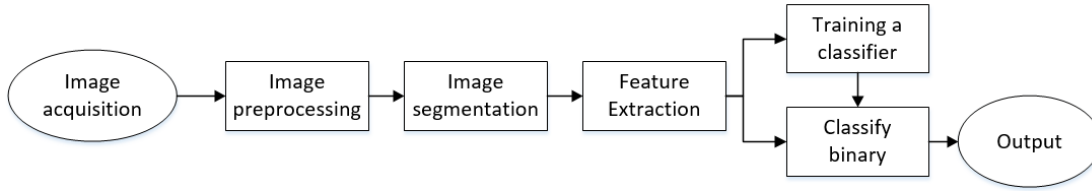
Because of the diversity of information on the Internet determines our diversity of information extraction mining technology. Overall we extract information mining technology based on mining objects are generally divided into information content mining, structure mining and mining information access information content. Among them, the information content including text mining and data mining and multimedia information mining two kinds[2]. Technical information on mining structure also includes hyperlinks control, internal control structure, URL control. The main technical route used as shown in Figure 3.



**Figure 3.** Technology Roadmap

## 2 Big data Applications in Image Recognition Anorectal

Image Recognition born in the 1920s, with the rise of the 1950s artificial intelligence in the 1960s, the rapid development of image recognition as a new discipline. A variety of principles and techniques of the subject being studied to give attention to other disciplines, increasingly used in many fields[3]. With the development of information technology and pattern recognition, image recognition is the first image processing, and automatically classify the images by machine or computer. The image preprocessing good input recognition system, trained and final classification, the output is a good target classification, image recognition process shown in Figure 4.



**Figure 4.** Image Recognition Process

Overall data feature calculation method  $pr_c$ ,  $c$  is a given transaction clustering,  $\mu$  is a significant threshold of production, so that the overall transaction clustering  $c$  calculated using the value of  $pr_c$  general characteristics are as follows:

$$pr_c = \{p \mid weight(p, pr_c) \geq \mu\} \quad (1)$$

$$weight(p, pr_c) = \frac{1}{|c|} \sum_{i \in c} w(p, t) \quad (2)$$

Wherein,  $c$  is a given transaction clustering,  $\mu$  is a significant threshold yield.

Depth-first search is a more search in the early. Its purpose is to achieve the HTML file does not contain any hyperlinks. Can not go forward until the search continued along the hyperlink on the HTML file and then returns to an HTML file, then continue to select the hyperlink in the file[4]. When no other hyperlinks can choose is stand for the search has ended.

Using the analytic hierarchy process to analyze the development of the depth-first search of the network strategy, and judged the matrix factors for weighing the relative weights of the criteria. The calculated weight has plus method, the root method and power method. In this paper we used method of plus to determine the matrix, each column is normalized corresponding weight. Each column normalized approximate their corresponding weights for non-consistency of judgment matrix, then strike arithmetic mean of these  $n$ -th column vector as a last weights. The specific formula is:

$$w_i = \frac{1}{n} \sum_{j=1}^n \frac{a_{ij}}{\sum_{k=1}^n a_{kj}} \quad (3)$$

In layers Sort we should judgment the matrix consistency test. Under normal circumstances, this nature does not require judgment matrix. But from the human understanding of the rule. A correct judgment matrix in order of importance are certain laws of logic. For example. If A is greater than B, B than the C important from Logically so A should be more important than C. If pairwise comparisons A than C important result of the violation of consistency criteria judgment matrix. It is unreasonable in the logic[5].

Therefore, in order to satisfy the consistency of a substantially we should do conformance testing in practice. Only through testing to judgment matrix is reasonable we will continued to analyze the results in logical[6].

(1) consistency index

$$C.I. = \frac{\lambda_{\max} - n}{n - 1} \quad (4)$$

(2) random index

According to the judgment matrix's different order of investigation under the table, and get an average random consistency index RI.

At first we assumed breadth search in the first  $k-1$  layer  $m$  elements relative to the total target weight  $w^{(k-1)} = (w_1^{(k-1)}, w_2^{(k-1)}, \dots, w_n^{(k-1)})$ . The  $k$ -th layer and the  $n$ -th element of the  $j$ -th element of a single sort

weights for the upper layer (layer  $k$ ) is  $p_j^{(k)} = (p_{1j}^{(k)}, p_{2j}^{(k)}, \dots, p_{nj}^{(k)})$ . Weight from  $j$  disposable elements is zero. suppose  $p^{(k)} = (p_1^{(k)}, p_2^{(k)}, \dots, p_n^{(k)})$ , When the search reaches the  $k$ -th layer when the sorting of the

elements on the first k-1 layer elements, the k-th layer of the total ranking of the elements for the total target as follows:

$$w^{(k)} = (w_1^{(k)}, w_2^{(k)}, \dots, w_n^{(k)})^T = p^{(k)} w^{(k-1)} \quad (5)$$

$$\text{or } w_i^{(k)} = \sum_{j=1}^m p_{ij}^{(k)} w_j^{(k-1)} \quad ; \quad i=1,2,\dots,n \quad (6)$$

Similarly, the need to test the consistency of the total sort results. Assumed to have been calculated as a criterion for the k-1 layer of the j-th element of the C.I.  $j^{(k)}$ , R.I. and  $j^{(k)}$ . C.R,  $j = 1, 2, \dots, m$ , the K-layer test indicators is

$$C.I.j^{(k)} = (C.I.1^{(k)}, C.I.2^{(k)}, \dots, C.I.n^{(k)}) w^{(k-1)} \quad (7)$$

$$R.I.j^{(k)} = (R.I.1^{(k)}, R.I.2^{(k)}, \dots, R.I.n^{(k)}) w^{(k-1)} \quad (8)$$

$$C.R.^{(k)} = \frac{C.I.^{(k)}}{R.I.^{(k)}} \quad (9)$$

When  $C.R.(k) < 0.1$ , We believe that the overall consistency of the judgment matrix is acceptable. According to the above derivation we can be determined that complex network of breadth-first search strategy in the the horizontal search will search range narrowed and improve accuracy in vertical search[7]. As the breadth-first search strategies able to the search path shrink to a minimum. So it has required for today's web search shrinking efficiency.

Due to the fluid network is a very complex network in the actual production. So it is must to seek a relatively efficient search method. Therefore, when we use the breadth-first search strategy also need to maximize the advantages. Using the analytic hierarchy process we known that with network complexity development the objective function moves up (except  $r=0$ ) [8]. Examine the breadth-first search strategy extremum will help us to establish the limits of search efficiency function. we got that:

$$\begin{aligned} \overline{\Omega}^h(r) &= K^h e^{\sigma^h [\rho^h (\tau^A + \tau^h) - \mu^h (\tau^A + \tau^h)] r} \\ &+ (1 - K^h) e^{-\sigma^h [(1 - \mu^h) \rho^h \tau^h - \Omega_r^h(0) / \sigma^h] r} \end{aligned} \quad (10)$$

Wherein,  $K^h$  is a constant determined by the under equation:

$$K^h = \frac{(1 - \mu^h) \rho^h \tau^h}{(1 - \mu^h) \rho^h \tau^h + \rho^h (\tau^A + \tau^h) - \mu^h (\tau^A + \tau^h) - \Omega_r^h(0) / \sigma^h} \quad (11)$$

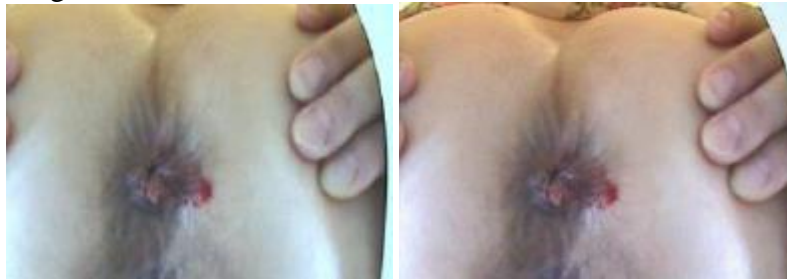
At the same time,  $\Omega_r^h(0)$  is the Derivative function of  $d\Omega^h(0) / dr = \sigma^h [(1 - \mu^h) \tau^A - \mu^h (\tau^M + \rho^h \tau^h)]$ .

$\overline{\Omega}^h(r) = K^h e^{\sigma^h [\rho^h (\tau^A + \tau^h) - \mu^h (\tau^A + \tau^h)] r} + (1 - K^h) e^{-\sigma^h [(1 - \mu^h) \rho^h \tau^h - \Omega_r^h(0) / \sigma^h] r}$  is changed in trend of the decision by the first exponential term. If the first index is negative. So for all  $r > 0$ ,  $\overline{\Omega}^h(r)$  is decreasing. Moreover, every search to meet the conditions of single-center system has been stable. However, if a search for once the index is positive the ingle-center system will be broken when N takes a value within a certain range[9].

As N increases, the stability of the single-center system depends on non-black hole[10]. For any one search h has  $\mu^M (\tau^A + \tau^M) / (\tau^A + \tau^h) \geq \rho^h$  in the middle column. Therefore, for all of the search range, whether N has much, the function of the limit of the search strategy is less than 1. In other words, the regardless of N how much the stability of the single-center system will not be broken. So we got that the breadth-first search strategy search rate can still be increased. Then conclusion that:

Assume that  $\mu^A \tau^A \leq \mu^M \tau^M$ , Given any research value h and g ( $h \leq \tilde{H}$ ,  $g \leq \tilde{H}$ ). If  $\{\rho^h > \rho^g \text{ and } \tau^h \geq \tau^g\}$  or  $\{\rho^h \geq \rho^g \text{ and } \tau^h > \tau^g\}$ , Then  $\tilde{N}^h < \tilde{N}^g$ ,  $\tilde{r}^g < \tilde{r}^h$  and  $\tilde{\theta}^h > \tilde{\theta}^g$ . In that  $\tilde{\theta}^h = |d\Omega^h(0) / dr|$

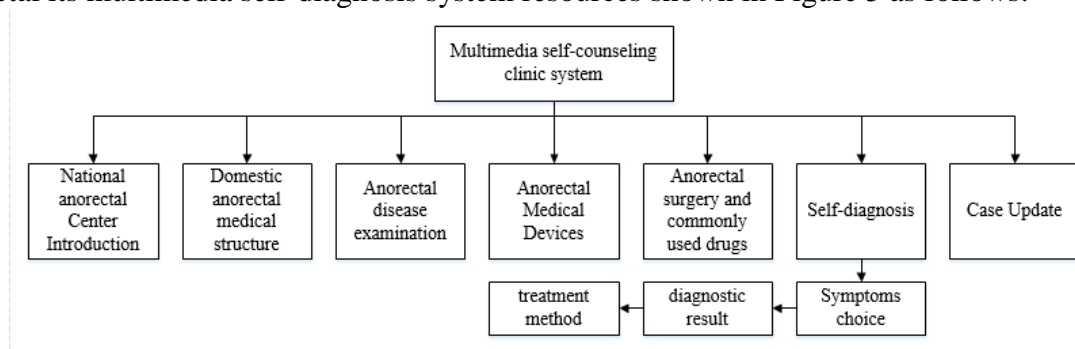
Therefore, according to the theory of the above algorithm in Matlab software simulation, the following figure first deputy to original image. Second Vice picture image after image recognition algorithm. shown in Figure 5.



**Figure 5-1.** The Original Image      **Figure 5-2.** Processed Images.

### 3 Big data Applications in The Diagnosis And Treatment of Anorectal Buffet Advisory System

Self consulting medical system is a surgery computer technology big data and anorectal departments of organic combination of product, through the clearing database integration of different diseases patients to statistical analysis to provide an accurate diagnosis and determine effective medical institutions and specialist physicians treatment programs, but also provide scientific guidance and counseling services for patients and their families. Anorectal is unique to humans a common disease with high incidence, a wide variety of clinical symptoms characteristic of complex, serious problems with the patient's physical and mental health[11]. Self-diagnosis system will be applied to the treatment of anorectal areas, can effectively play to literacy, disease prevention and treatment. Anorectal its multimedia self-diagnosis system resources shown in Figure 5 as follows:



**Figure 6.** Multimedia Self-counseling Clinic System

### 4 Conclusion

By studying big data applications in image recognition and anorectal big data on self-help advice anorectal treatment system. Of anorectal disease Automatic identification of a certain amount of research, especially in image segmentation anorectal disease, disorders of the image feature extraction, feature data and be able to automatically trained to identify aspects of classification algorithms, and so do a lot of work. In the course of the study, to appreciate the anorectal disorders automatic identification technology does have some usefulness in the diagnosis of medical terms. Further advice on self-treatment systems, the study has been completed anorectal disease multimedia self-counseling clinic system implanted in the hospital Vaccine Immunol system to better serve patients, greatly improving the human services level of the hospital. I believe that with further development and refinement of such systems will be increasingly applied to the process of universal treatment and disease prevention knowledge of diseases, in order to improve digital and network of medical information, to enhance the level of service to the hospital contribution.

## References

1. Mallat S, Zhang Z. Matching pursuit with time frequency dictionaries. IEEE Trans, on Signal Processing, 1993; 41 (12) : 3397—3415
2. Smith S. Overview of fMRI analysis. British Journal of Radiology, 2014, 77(2):167-175.
3. RosaM. Ventura F i. Image codingwith matching pursuit. M S thesis, UPC2EPFL, 2001.
4. Antoniadis A., Bigot J., Sapatinas T. Wavelet estimators in nonparametric regression: a comparative simulation study. Journal of Statistical Software, 2001, pages 1-83.
5. Achim A., Bezerianos A. Tsakalides P Novel Bayesian Multiscale Method for Speckle Removal in Medical Ultrasound Images. IEEE Transactions on Medical Imaging, 2001, 20(8):772-783.
6. Ma Liming, Li Yanting, Huang Zhaojia, other medical image processing technology [J] Modern Hospital Journal, 2007,7 (11): 15-19
7. Chen Shaoming. Development strategy anorectal subjects [C] .2012 Medical Frontiers - anorectal Branch of the Chinese Medical Association tenth Four national anorectal symposium Selected Papers, 2012.
8. Donoho DL, Elad M. Op timally Sparse Rep resentation in General Dictionaries viaMinimization[J ]. PNAS, 2003, 100 (5) : 2197- 2202.
9. Yong Cao, Xiaoguang Yue, Fei Xiong and Youjie Zhao. The Software Reliability Model Using Hybrid Model of Fractals and BP Neural Network. IETI Transactions on Computers. 2015.1(1) :11-21
10. Qiao Zhihong, Wu Qing Qiao, Rain peak filtering algorithm based on direction of the fingerprint image enhancement Computer Engineering and Applications. 2004 (4): 87-89
11. A Ross, A K Jain, J Z Qian. Information fusion in biometrics. The third international conference on audio and video based biometric person authentication. 2012(4):354-360