Research on The Information Processing Mode of Minority Traditional Culture Under the Background of Big Data

Qian Li

qwe1570414@163.com

Jiangxi Vocational and Technical College of industry and trade, Nanchang, Jiangxi, 330038, China

Abstract. The solution of problems such as "semantic gap", long time, low efficiency and non-standardization will help to apply the research results of multiple fields to multiple fields in multiple fields and provide new ideas for multiple fields. By using the techniques of text data, image description, data mining, data collection and screening, data processing and mining, data output and feedback, it has completed the transmission, protection, sharing and re-creation of traditional national culture information. By using these materials to sort and arrange, the effective storage, retrieval and sharing of Chinese folk traditional culture materials are realized [1].

Keywords: minority traditional culture; Information processing model; Image description; Text data;

1 Introduction

Data mining of national traditional information is a process of analysis, thinking, interpretation and value generation of its image information and text information. Video information is a kind of video information recorded in the form of family life, school education, daily life, religious festivals, etc[2]. "Characters" refers to the contents of traditional Chinese culture, including printed materials such as literature, books and local Chronicles[3], as well as online reviews, descriptions and self-impressions. With the rapid development of modern information technology, the traditional culture of our country contains more and more information, however, by the restriction of the information processing object, at present only a little part needs to be manually completed. In addition, in the process of the expression of text and pictures and other information, there will be some deviation due to the subjective reasons of people, which will have an adverse effect on the result[4].

2 Current situation and existing problems of information processing of ethnic minorities' traditional culture

2.1 Current situation of information processing of ethnic minorities' traditional culture

Collecting and saving national historical documents is an important part of the study of Chinese historical documents. "Static collection" is the use of high definition scanning equipment, such

as digital cameras, scanners, etc., to collect text and text data; Dynamic collection is the collection of opera, dance and other sound and image materials through photography, photography and other technologies. "Production and storage" refers to the use of multimedia information processing technology or image generation software, audio, video, images, texts and other categories, and save them in different forms, to provide data support for related scientific research [5].

2.2 Problems in information processing of ethnic minorities' traditional culture

In terms of processing the information of ethnic minority traditional cultural images, there are the following aspects: First, the texture, shape and color of ethnic costumes can be extracted by the fusion of multiple attributes, so as to fully express the complex temporal and spatial distribution information in the images, thus improving the accuracy of image analysis. Second, through the study and improvement of the deep network of traditional clothing, the high level semantic information of traditional clothing can be efficiently searched. Third, three types of convolutional neural networks such as AlexNet, VGGNet and ResNet are used respectively to extract features in the image by using Canny operator, LBP and AlexNet, etc., and then depth information is used to process features in the image. Finally, the traditional image recognition method is used to process the image in the image to improve the accuracy of the image recognition. Fourthly, the algorithm based on image segmentation and gray-scale is used to preprocess the Gauss geometric features, and by improving the Gauss geometric feature extraction algorithm, it is used in the classification of SVM to realize the recognition of Gauss geometric features [6].

3 Establish the information processing mode of national traditional culture

Text digitization refers to converting narrative sentences in text and natural language into data and processing them so as to overcome the difficult problems caused by the complexity of text expression. The realization process includes: keyword extraction and lexical frequency data processing. In terms of keyword extraction, this paper adopts the method of combining TF-IDF algorithm and search function. IDF method extracts keywords, removes common words, retains keywords, and encodes them. The find function looks for the keyword you want within a specified range. Word frequency statistics is a statistic used to count the number of occurrences of a word in a corpus. There are many kinds of statistical charts, including bar chart, fan chart, line chart, network chart, stem and leaf statistical chart and histogram. And extract the unnecessary important content from the massive text, can extract the important content from the massive text[7]. Word frequency statistics can clearly show some key information about the region, which can effectively deal with the huge workload brought by the complexity of language expression and the difference of information sources and recording methods. This efficient, fast and accurate data method can lay the foundation for Chinese character data processing and processing. See Figure 1, 2 [8].
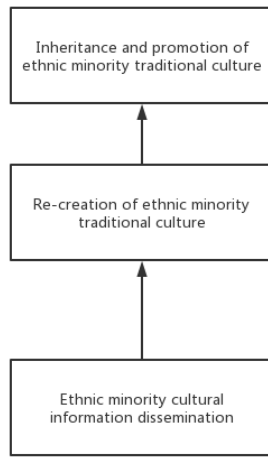
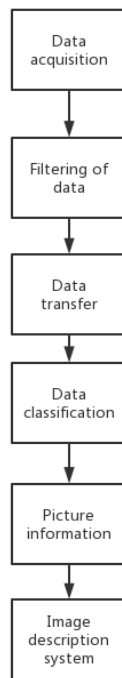**Fig. 1** Information processing model of ethnicminority traditional culture

Fig.2 Information processing model of ethnic minority traditional culture

Firstly, the WBS technology is used to classify the collected images of traditional ethnic life scenes and Chinese text information to achieve data fusion with similar attributes or consistent

features. By classifying the data, it can make the processing of the data more specific and more operable, so as to make the results more targeted and systematic.You can see this in Figure 3.
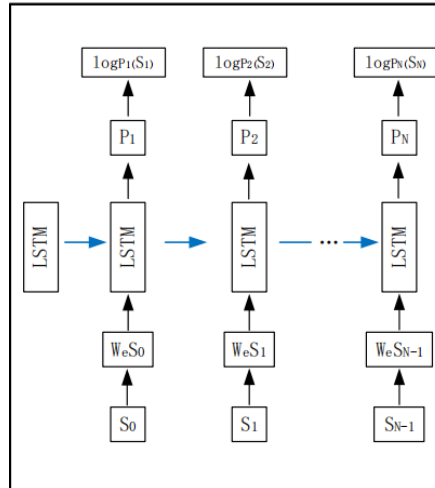

Fig. 3. Image information processing of minority traditional history

4 Application of information processing mode in national traditional culture

After the secondary screening by experts and scholars, the teaching scene pictures of the Mongolian ethnic group conforming to the subject of this paper are selected and imported into the computer

First, obtain relevant literature related to Mongolian education, Mongolian education and Mongolian ethnic education from 2015 to 2021 by using the database of China Knowledge Net, select 115 articles related to Mongolian ethnic education according to the content and source, and list them as the content of this study; The second is to select the chapters about the minority culture from the General History of the Mongolian Nationality, the Mongolian Nationality and the Mongolian Nationality in Jilin Province. The third is to carry out field investigation in Mongolian areas, using cameras, airships and other equipment to shoot the required image materials; Fourth, the search of Mongolian traditional culture images, using the "crawler" technology to collect a large number of Mongolian traditional culture and educational images, and saved in the.jpg database. Firstly, the image quality, whether it belongs to the traditional Mongolian education scene and role occupying were selected as the selection criteria, and the images were initially screened, and 5 Mongolian students were surveyed by questionnaire.

After the secondary screening by experts and scholars, the Mongolian education scene which is consistent with the title of the paper is finally selected and input into the computer.

Table 1. Word frequency statistics of Mongolian traditional education keywords

Serialnumber	Keywords	Wordfrequency	Serialnumber	Keywords	Wordfrequency
0	Mongols	156	7	important	51
1	parents	152	8	unity	44
2	education	134	9	elder	42
3	love	81	10	respect	32
4	Mongoliannationality	77	11	Goodfaith	31
5	Familyeducation	61	12	respect	30
6	Attachimportanceto	61	13	reputation	30

Using the index library in python, A lower triangular A-font is created in Table 1 with a size of 24x24. The main diagonal element is the word frequency of the keyword, and all the elements on the right are 0[9] . In A sentence, if a keyword m and a keyword appear in a sentence, then this number is accumulated by 1 to calculate the co-occurrence frequency of each keyword and assign it to the corresponding element on the left, resulting in a co-occurrence matrix A. On the basis of co-occurrence matrix A[10] , the results of dividing each element in A by the same diagonal unit are assigned to B respectively, and it is concluded that matrix B represents the relevance between each key word. Finally, the visualization of A and B is realized by using matshow function in matplotlib library. Figure 4 and 5 show the co-occurrence matrix and correlation matrix of high-frequency keywords in traditional Mongolian teaching. The numbers 0,1,2,23 in the figure correspond to keywords such as "Mongolian", "parents", "education" and "waste" respectively[11] .
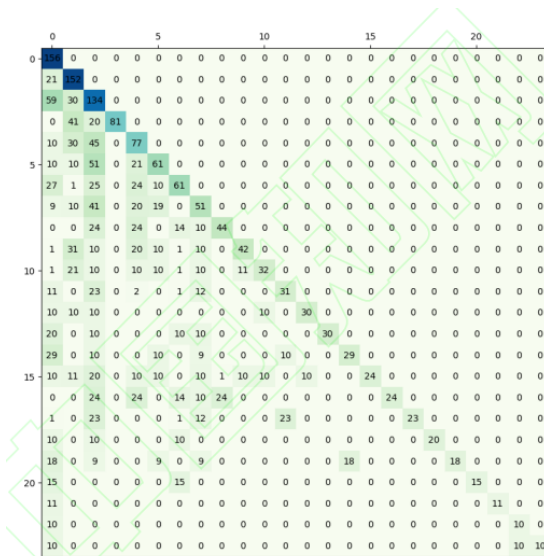

Fig. 4. Co-occurrence Matrix of high-frequency keywords for Mongolian traditional education
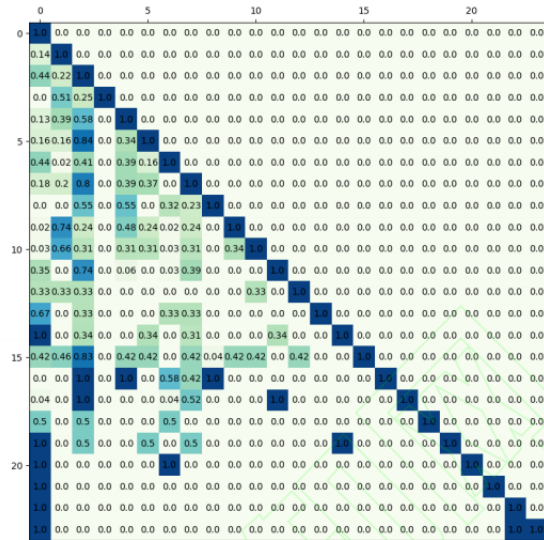
Fig. 5. Correlation Matrix of high-frequency keywords for Mongolian traditional education

5 Conclusion

Under the background of "big data", the traditional methods of national cultural images and texts have been unable to meet the development needs of today's society[12] . Based on the traditional processing method of text information, this paper combines text information with text data to form a working mode of text information processing. The implementation of this project will provide a new way of thinking for the information processing of ethnic traditions, a new, effective and efficient way of information processing for the research of ethnic traditional culture, and a new way of thinking for the development of ethnic culture[13] .

References

[1] Xiao Kun, Lu Yu. Digital Protection of Ethnic Traditional Culture Based on Big Data [J]. Information and Communication,2019(05):177-178.

[2] Gao Huixin. The era of big Data and the protection and inheritance mode of Ethnic minority Traditional culture: A case study of Laochang Township, Xinping County [J]. Drama Home,2018(14):230+247.

[3] Zhao W L. Research and implementation of Ethnic costume image retrieval algorithm based on regional comprehensive matching [D]. Yunnan Normal University,2017.

[4] Wang Jingwu. Research on Museum Information and Intelligent Service Mode based on Internet of Things Technology [J]. Information Science,20,38(11):45-50.

[5] Lv Xinen, Chen Huiling, CAI Zhenao. Research on Personality Behavior Information processing prediction Model in Big Data environment [J]. Information Science,2019,37(10):108-113. (in Chinese)

[6] Li Gang, Zhu Xuefang. Research on Construction of visualization Service model of Tubofile digital Resources based on Semantic Association [J]. Information Science,20,38(07):147-152.

[7] Chen Qiangpu. Research on Deep Neural Network Model for Image Description [D]. Chongqing University,2017.

[8] He Chunyu, Teng Chune. Knowledge ontology construction of intangible cultural heritage -- A case study of the intangible cultural heritage resources of Hezhen Nationality [J]. Information Science, 201,39(04):140-147+156.

[9] Hong Liang, Li Xuesi, Zhou Lina. Field crossing: Application and development trend of data mining [J]. Library and Information Knowledge,2017(04):4+22-32.

[10] Institute of History, Inner Mongolia Academy of Social Sciences. The General History of the Mongolian People [M]. Beijing: Nationalities Publishing House,2001.

[11] Fu Sheng, Yang Fengmo. The Mongolian People [M]. Liaoning: Liaoning Nationalities Publishing House,2009.

[12] Zhao Zhunsheng, Bao Yuwen. Jilin Mongolian [M]. Liaoning: Liaoning Nationalities Publishing House,2016.

[13] Zhang Jinsong. Book Label Association Rule Recognition Method Based on User-generated text [J]. Journal of Information, 201,40(11):182-189.