ISSN 2616-5775 Vol. 5, Issue 14: 44-47, DOI: 10.25236/AJCIS.2022.051407

Classification of graphic office information in intelligent office automation system

Xiaoqian Zhang

Computer Simulation, Beijing, China
13681310771@163.com

Abstract: *Due to the low F-Measure value of traditional methods in practical application, the classification effect of graphic office information is not good, so a classification method of graphic office information in intelligent office automation system is proposed. Through the semantic association of images and texts, the features of image and text office information are extracted, the weight of information features is determined, and the similarity between image and text office information is calculated according to the features, so as to achieve information classification. The experiment proves that the F-Measure value of the design method is high, and it has a good application prospect in the field of image and text office information classification.*

Keywords: *Intelligent office automation system; Graphic office information; Classification; F-Measure value; Weight*

1. Introduction

Graphic and text office information classification is one of the core functions of intelligent office automation system, through the automatic classification of information, so that the mass of graphic and text office information can be stored and searched orderly, to provide system users with office convenience, so that the system gives full play to the function of intelligent office automation. With the continuous development of big data, graphic and text office information has the characteristics of large quantity and variety, so the classification of graphic and text office information has a certain difficulty. Domestic research on the classification of graphic and text office information started relatively late, the existing technology and theory are not mature enough, information classification is still in the preliminary exploration stage, and there is still a big gap in technical level compared with foreign countries. Although the classification of graphic and text office information has attracted the attention and concern of the research field in recent years, relevant scholars and experts have carried out a series of studies and proposed some methods and ideas, but the existing methods are not ideal in the practical application of the classification of graphic and text office information, the F-Measure value is low, and the traditional methods have been unable to meet the requirements of information classification in the accuracy. This paper puts forward the classification method of graphic office information in intelligent office automation system.

2. Image text semantic association

The classification of graphic and text office information is based on graphic and text semantic features. Therefore, graphic and text office information features are extracted through graphic and text semantic association to provide a basis for information classification. On the basis of considering the sentence level features of the text, the global features of the text are associated with the regional features of the image to highlight the image text information features. The following figure is the schematic diagram of image text semantic association.

ISSN 2616-5775 Vol. 5, Issue 14: 44-47, DOI: 10.25236/AJCIS.2022.051407

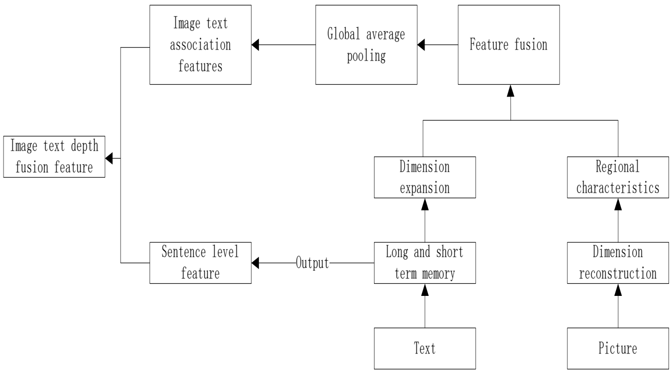

Figure 1: Schematic Diagram of Image Text Semantic Association

As shown in Figure 1, suppose that the given image text pair is (T,K), where T is image information; K is text information [1]. The word embedding technology is used to embed each word in text information K into a word vector with dimension h, input the word vector into the long-term and short-term memory network, and extract the global features of text information K, which is expressed by the formula:

$$U = L(k_1, k_2, \cdots, k_n) \tag{1}$$

Where, U represents the global feature of text information K; L stands for long and short term memory network; k represents word vector [2]. Feature T of image information is extracted by BP neural network, which can be expressed as:

$$P = S(t) \in t^{w \times q \times x} \tag{2}$$

Where, P represents the image information T feature; S represents BP neural network; t is the image vector; w is the image width; q is the image length; x represents the number of images [3]. The extracted text information features and image features are spliced to obtain the deep fusion features of image and text office information:

$$D = Concat(U, P) \tag{3}$$

Where, D represents the feature of deep integration of graphic and text office information; $Concat$ is the loss function. The above formula is used to extract the features of the graphic office information to be classified.

3. Calculation of feature weight

The importance of information features to the classification of graphic office information varies. In order to ensure the classification accuracy of graphic office information, weight is used to show the importance of each type of features to the classification of graphic office information. Set Y is established from the information features extracted above. Set Y is composed of m features. The weight of each feature is calculated using the weighting method. The calculation formula is:

$$\begin{cases} \varpi_m = z * H \\ z = \dfrac{C}{\sum C} \end{cases} \tag{4}$$

Where, ϖ_m represents the weight value of the m feature in feature set Y; z represents the importance of features in graphic office information. H represents the frequency of features in the graphic office information; C represents the importance coefficient of features in graphic office information [4]. The above formula is used to calculate the weight of each feature in the feature set Y.

ISSN 2616-5775 Vol. 5, Issue 14: 44-47, DOI: 10.25236/AJCIS.2022.051407

4. Information classification

On the basis of the above, the feature vector of the extracted graphic and text office information is integrated to represent the graphic and text office information from multiple dimensions, which can be expressed as:

$$G(D)=\left[\frac{b}{\|b\|}\cdot\frac{\alpha}{\|\alpha\|}\right]^{\varepsilon} \tag{5}$$

In the formula, $G(D)$ represents the integrated graphic office information feature vector; b represents the short - and long-term memory network information vector representation of the atlas information; α represents the representation of graphic and text feature vectors extracted based on BP neural network; ε represents the transpose operation of the matrix; $\|\ \|$ stands for normalization operation. Compare each graphic office information feature with other information features, and calculate the similarity of the two graphic information. The calculation formula is:

$$\kappa=\sum_{m}\frac{\varpi_i G(D)_i}{\varpi_j G(D)_j} \tag{6}$$

Where, κ represents the similarity between graphic office information i and graphic office information j; $G(D)_i$ D represents the i characteristics of graphic office information; $G(D)_j$ represents the j feature of graphic office information [5]. Use the above formula to calculate the similarity of two information, and set a threshold value according to the actual situation. If the similarity is less than the threshold value, it means that the two image and text information do not belong to the same category. If the similarity is greater than the threshold value, it means that the two image and text information belong to the same category, and it is divided into the same category of data, thus realizing the classification of image and text office information.

5. Experimental demonstration

In order to test the feasibility of the proposed graphic and text office information classification idea, a graphic and text office packet is selected as the experimental object. The packet contains 20,000 pictures and 5.26GB text information, with a total of 6.5GB of data. The images include face images, car images, table images, etc. Text data includes 7 kinds of payment information, personal basic information, shopping information, etc. The data base is large and various, which meets the experimental requirements. This design method is used to classify the graphic and text office information in the data packet, and two traditional methods are selected as the comparison, the two traditional methods are based on the block chain and the multi-layer semantic fusion respectively. The following are represented by traditional methods 1 and 2. According to the above process, semantic association, feature extraction, feature fusion, weight calculation and information classification of graphic information are carried out. In the experiment, F1 value is taken as the accuracy evaluation index of the three methods. F1 value is F-Measure, which is the comprehensive evaluation index of information classification method and can reflect the classification accuracy. The larger F1 value is, the higher the classification accuracy of text and text office information is.

$$F1=\frac{2\times T\times P}{T+P} \tag{7}$$

Where, T represents the recall rate; P stands for accuracy. In the experiment, the amount of graphic office information is used as a variable, and 0.5 information is used as the base number. 1GB data is added for each classification. The above formula is used to calculate the F1 values of three methods under different information amounts. The experimental data is recorded using spreadsheets. The specific data is shown in the following table 1.

ISSN 2616-5775 Vol. 5, Issue 14: 44-47, DOI: 10.25236/AJCIS.2022.051407

Table 1: Comparison of F1 Values of Three Methods under Different Information Volumes

Sample information/GB	Design method	Traditional method 1	Traditional method 2
0.5	0.98	0.74	0.77
1.5	0.97	0.68	0.71
2.5	0.97	0.65	0.63
3.5	0.96	0.62	0.56
4.5	0.95	0.58	0.51
5.5	0.95	0.55	0.47

It can be seen from the data in the above table that the F1 value of the design method is relatively high. Although it will gradually decrease with the increase of the amount of graphic office information, the decrease is relatively small, and can be basically controlled above 0.95. When the amount of graphic office information is 5.5GB, the F1 value is 0.95; The F1 value of the two traditional methods is relatively low. The F1 value will decrease significantly with the increase of the amount of graphic and text office information. When the amount of graphic and text office information is 5.5GB, the F1 values of traditional method 1 and traditional method 2 are 0.55 and 0.47 respectively, which are far lower than the design method. The design uses short-term memory network and BP neural network to classify and extract image and text information features, and dig out graphic and text features in depth, It provides an accurate basis for information classification. Therefore, the experimental results prove that the design method has absolute advantages in precision, and is more suitable for the classification of graphic office information in intelligent office automation system than the traditional methods.

6. Conclusion

As an important functional module of intelligent office automation system, the classification effect directly affects the performance of the system. This time, aiming at the shortcomings and defects of the traditional methods, a new idea of graphic and text office information classification is proposed, which realizes the optimization and innovation of the traditional methods. It has good practical significance to improve the classification accuracy of graphic office information. Due to the limited time of this research, the proposed method has not been widely applied and operated in practice, and there may be some shortcomings in some aspects. In the future, we will carry out a deep exploration of this topic to provide a strong theoretical support for the classification of graphic and text office information.

References

[1] Li Feng, Liu Lin, Fan Hongyan. Research on The Classification and Coding Technology of Information Model for Automobile Industrial Building Construction Based on OmniClass [J]. China Engineering Consulting,2021(11):94-98.

[2] Chen Xuanyi, Sun Xiaolei, Wu Zhenggang, et al. Research and Application of BIM Based Electromechanical Component Information Classification Coding Technology in Exhibition projects [J]. Installation, 2021(09):65-67.

[3] Zhang Li, Ma Jing. A text classification method combining word statistical characteristics and semantic information [J]. Computer Engineering & Science, 2021, 43(07):1308-1315.

[4] Wang Zhen, Zhang Haiqing, Peng Li, Et al.Information Extraction and Classification Method of Medical Data based on Singular Value Decomposition [J]. Journal of Chengdu University of Information Technology, 2020, 35(05):537-541.

[5] Zhang Dong-Dong, Peng Dun-Lu. Ent-Bert: Entity Relation Classification Model Combining BERT and Entity Information [J]. Journal of Chinese Computer Systems, 2020, 41(12):2557-2562.