Image Emotional Semantic Retrieval Based on ELM

Peile Zhang, Min Yao, Shenzhang Lai
College of computer science & Technology
Zhejiang University
Hangzhou 310027, P.R. China
zhangpeile/myao@zju.edu.cn

Shenzhang Lai
College of computer science & Technology
Zhejiang University
Hangzhou 310027, P.R. China

Peile Zhang,
College of computer science & Technology
Zhejiang University
Hangzhou 310027, P.R. China

Junfei Zhuo
The Computer Centre
Zhejiang University
Hangzhou 310027, P.R. China
zhuojf@huahai.net

Abstract—Image emotional semantic retrieval is one of the important subjects in information science. This paper discusses the basis of emotional cognitive, the dimension of emotional expression and the establishment of scientific emotion space, and proposes a kind of improved extreme learning machine for image emotion semantic retrieval. Finally, a kind of prototype system for image emotion semantic retrieval has been developed. The experimental results show that prototype system for image emotion semantic retrieval is effective.

Keywords-image; emotion; image retrieval; semantic retrieval; ELM

I. INTRODUCTION

With the advent of the multimedia age, the picture as the main medium of information transmission has become a major tool for people to express emotion, conduct social activities. However, with the explosive growth of the images, how to retrieval images efficiently is an urgent problem to be solved.

The traditional image retrieval methods are mainly based on the underlying features of the images, which search images by text-based or content-based markers without considering the semantic information of the images and having bigger difference in semantics with human understanding. In other words, there is a huge semantic gap (Semantic Gap) between the underlying visual characteristics of the images and human search for images by means of image semantics [1-3], unable to meet the demand for people o retrieve images with semantics. In order to eliminate semantic gap in semantic image retrieval, semantic-based image retrieval (SBIR for short) has been proposed. Semantics is the text abstract to image by human, different semantics levels stands for different complexity about image content description in image retrieval [4].

Emotion is the most difficult level in semantic description. Unlike the underlying visual features, image emotional semantics is the closer semantic level with human emotion recognition for images. This paper forms emotional space by adjective pairs, and builds mapping between image emotional semantic and the underlying characteristics of the images. At the same time, a kind of improved Extreme Learning Machine is used to execute image emotional semantic classification and retrieval. Experiment results shown its better retrieval accuracy and robustness.

II. IMAGE EMOTIONAL SEMANTIC RETRIEVAL

Image emotional semantic retrieval mainly covers following three key aspects: 1) extraction and dimensionality reduction of image visual features; 2) emotional semantic description and its relation with image features; 3)classification and retrieval of machine learning based Image emotional semantic. Many previous research has focused on the first aspect, thus, this thesis will mainly analyze the other two.

A. Emotional Semantic Description

1) Choice on Adjective Space

Studies have shown that human emotional variation intensity is continuous [5]. Every single position of emotional space represents a specific type of emotion, therefore, if directly applying the whole emotional space to describe the emotion of image, which will result in numerous types of emotion. Moreover, each of those emotions is similar with the others and accompanied by certain degree of redundancy. Thus, it will fail to describe certain type of human emotion visually. For this reason, this thesis puts forward the idea of building an adjective-based emotion space, by selecting orthogonal and representative adjectives, to construct a coarse-grained space to express emotions.

Based on the combination of Plutchic emotional space and PAD dimensional model of emotion, this thesis selects 12 pairs of emotional adjectives with representative and orthogonality. Under certain emotional dimension, each pair of adjectives becomes the bipolar pair, and thus forms the corresponding emotional space with 12 dimensions. This method can not only effectively condense the emotional complexity of continuous emotional space, but

also describe human emotions in a more complete way. 12 pairs of adjectives are shown in Table I:

TABLE I. ADJECTIVE EMOTIONAL SPACES

passionate-indifferent	comfortable-uncomfortable
warm-cold	harmony-conflicted
excited-quiet	exquisite-coarse
orderly-disorderly	serene-noisy
mild-pungent	clear-vague
bright-dark	terrified-calm

2) Establishment of Training Image Library

The core part of image emotional semantic retrieval is the training of classifier, which gives the classifier an ability to recognize human emotion. The training result of classifier depends on the completeness of the training image library, the more accurate mapping between image and emotion; the training result will be more favorable. This paper selects 1288 images of landscape, characters and buildings, based on the adjective emotional space in table I, those images are labeled with different emotional tag though user emotional labeling experiment.

In order to classify emotions, we need to use some adverbs to extend emotion, making it more rich and specific. Quantification of this method can be describe as a decimal between 0 and 1, for example, for "warm - cold", very warm, relatively warm, not cold not warm, relatively cold, very cold, these five degrees can be denoted as 0, 0.25, 0.5, 0.75, 1respectively. This method is used to label the emotion of Images.

In the process of labeling image by users, it is very hard to avoid subjectivity because of users from different areas may have different understandings of a same image. So deferent area users are selected, each user marks the emotion of 100 images in the image library; each image labeled with a decimal between 0 and 1, stores the label data, and finally summarizes the experiment data to obtain an average emotional result to represent the emotion of image. Then according to the average emotional result to classify image emotion, for example, when there is an image in the "warm - cold" emotional space, its average emotional result is 0.87, we can consider this image belongs to lukewarm emotion; while the average emotional result is 0.12, it belongs to passionate emotion. By using this method we can get the final emotional image library.

B. ELM based Image Emotional Semantic Classification

1) Introduction of Extreme Learning Machine

Given detects of single hidden layer feed-forward neural network algorithm, it is unable to meet the massive image data application environment. This paper uses a relatively new single hidden layer feed-forward neural network algorithm called extreme learning machine[6] algorithm for image features classification and image retrieval.

Steps of Extreme Learning Machine algorithm:

For a given set of training sample for ELM input layer, $S=\{(x_i,t_i)\,|\,x_i\in R^n, t_i\in R^m\}, i=1,2,...,N$, where N is the total number of training samples, $\tilde{N}$ is the number of nodes of hidden layer, g(x) is the activation function, xi is the i-th sample of input layer, t_i is the corresponding output classify vector of the i-th input sample.

a) Randomly generated connection weights between the input layer and the hidden layer wi and the threshold value of hidden layer nodes bi, where $i=1,2,...N$ *;*

b) Calculate the hidden layer output matrix H according to the formula $H\beta=T$ *, where H:*

$$H=\begin{bmatrix} g(w_1\cdot x_1+b_1) & \cdots & g(w_1\cdot x_1+b_1) \\ \vdots & \ddots & \vdots \\ g(w_1\cdot x_1+b_1) & \cdots & g(w_1\cdot x_1+b_1) \end{bmatrix},$$

$$\beta=[\beta_1,\beta_2,...,\beta_{\tilde{N}}]^T$$

c) According to the formula $H\beta=T$ *,* β *can be calculated as* $\beta=H^{-1}T$.

Unlike normal SLFNs learning algorithm, ELM algorithm does not need to iterate W and b, which respectively is the connection weights between the input layer and the hidden layer, and the threshold value of hidden layer nodes, to get the optimal solution. When the algorithm starts, W and b are randomly generated [7], β can be calculated from $H\beta=T$, then gets a set of parameters, which has been proven to be affective to classification problem. ELM does not need to iterative, so that achieves a significantly improved efficiency.

In practice, the number of training samples is far greater than the number of nodes in the hidden layer, H is rectangular matrix, there is not necessarily exists a set of parameters (W, β , b) such that (2.1) and (2.2) are equivalent:

$$\left\|H(\hat{W},\hat{b})\hat{\beta}-T\right\|=\min_{w,b,\beta}\left\|H(W,b)\beta-T\right\| \quad (1)$$

$$E=\sum_{j=1}^{N}\left\|\sum_{i=1}^{\tilde{N}}g(w_i\cdot x_j+b_i)\beta_i-t_j\right\|^2 \quad (2)$$

For the formula (2.2), namely the cost function to obtain the optimal solution, the traditional methods must execute iterative approximation, while ELM algorithm does not require iterative, all its need is just simply specify the appropriate W and b, least squares solution $\hat{\beta}$ can be obtained as:

$$\left\|H\hat{\beta}-T\right\|=\min_{\beta}\left\|H\beta-T\right\| \quad (3)$$

$$\text{Namely}\quad \hat{\beta}=H^{-1}T=(H^TH)^{-1}H^TH \quad (4)$$

Where A is a generalized inverse matrix of H. Huang Guang Bin [8] proves as long as the activation function g(x) infinitely differentiable, then W and b needn't to update, you can find the least-squares solution of $\hat{\beta}$.

Due to the randomness of ELM parameters, different performances of the different parameters are mainly reflected in the choice of excitation function and threshold. Many scholars have conducted in-depth research in this area, Huang Guan bin proposed ELM and kernel ELM, which are used to solve input problem in complex space; The training samples of original ELM are provided one-time, all the samples should input into the network at the

same time, so after the training, the ELM cannot be changed. Based on this deficiency, Liang [9] proposed Online Sequential ELM, in which training samples can be inputted batch by batch. This algorithm effectively compensates the shortcomings of the original ELM that cannot add new training samples, making the system more time-sensitive.

The above analysis shows that ELM can randomly set the parameters of the network and calculates the other parameter when used to solve classification problem, in this way can highly improve the training efficiency. However because of the randomness of ELM parameters, the parameters have not been optimized, which may reduce the training ability of ELM, in this paper, we propose a modified ELM to improve the accuracy of the original ELM.

2) The Modification of ELM and Performance Analysis

a) Modification of Extreme based on Genetic Algorithm

As mentioned above，the randomness of parameters of original ELM may reduce the ability of classification, we propose a new algorithm called GA-ELM, which combines ELM with genetic algorithm. This algorithm is used to classify image in this paper. The main process of this algorithm is as follow: Get 50 difficult ELM parameters by using original ELM, use genetic algorithm to modify these 50 parameters, choose 10 best ELM parameters and form a combined ELM.

Genetic Algorithm (GA) [10], a heuristic search method simulating biological evolution, is widely used to solve optimization and search problems. Genetic algorithm is a kind of evolutionary algorithm, which mainly simulates crossover, mutation and selection of chromosome in evolution of biological communities, and selects the better offspring. The initial value of GA is chromosomal sequence of community, which is binary sequence coding from the solution of practical problem, individuals of initial group are randomly selected; In the process of GA, fitness coefficient is used to indicate the degree of adaption to the environment of individual, if the fitness coefficient is larger than standard value set before the algorithm started, we can stop the algorithm, and select best individual from the population, which can be decoded to get the solution of practical problem; otherwise, we continue the algorithm until the fitness coefficient larger the standard value. According to the initial crossover and mutation probability, individual is selected to operate chromosome changing. Crossover is a process that parental chromosomes compose to form a new chromosome, which may be a better chromosome and making the population evaluate toward the direction of optimal solution. Mutation, a process that part of individual chromosome has changed, maintains the biological diversity of the population. By crossover and mutation, a new generator is produced, and then repeats the calculation of fitness coefficient as mentioned above, until reach the times of evolution, the algorithm will stop.

The genetic algorithm is applied to optimize the parameters of ELM, if ELM hidden layer nodes are $\tilde{N}$, then the connection weight of input layer and hidden layer W and the hidden layer node threshold value b can be united into the initial population of individuals as $e=\left[\omega_{11},\cdots,\omega_{1\tilde{N}},\cdots,\omega_{n1},\cdots,\omega_{n\tilde{N}},b_1,b_2,\cdots,b_{\tilde{N}}\right]$, Wherein ω_{ij} and b_i is a random number between [-1,1].50 ELM parameters form an initial population, and the calculation of accommodation coefficient is based on the standard deviation of individual chromosomes and training samples, as shown in (2.5) as follows:

$$F=\sqrt{\frac{\sum_{j=1}^{N}\left\|\sum_{i=1}^{\tilde{N}}\beta_i g(w_i\cdot x_j+b_i)-t_j\right\|_2^2}{N}} \quad (5)$$

According to the accommodation coefficient to choose individual, do crossover and mutation probability, the last generation produces suitable alternative individual to form combined ELM.

ELM parameter individual, which optimized by using genetic algorithm, more lager of its accommodation coefficient shows lower training standard deviation of corresponding ELM, according to which we will be able to get better performance of individuals. However, for a particular ELM, its performance not only determined by its accommodation coefficient, but also the weight of norm of corresponding ELM, i.e. related to $\|\beta\|$, smaller the weights of norm , more superior performance of ELM. Through the above analysis, this paper choose individual accommodation coefficient and minimum paradigms standard in the selection of individuals, these two parameters are calculated normalized respectively, and calculate the sum, as (2.6) shows, sort the result in descending order, and select top ten ELM to form the combined ELM.

$$C_{Num}=\max_{Num}\left\{\frac{f_i}{\sum_{j=1}^{M}f_i}+\frac{\|\beta\|}{\sum_{j=1}^{M}\|\beta\|}\right\} \quad (6)$$

Where Num represents the number of selected ELM, in this paper Num equals 10, C_{Num} represents the selected ELM, f_i represents the accommodation coefficient of the i-th ELM, $\|\beta\|$ represents the weight of norm of i-th ELM, where i equals 1, 2, ..., M, M is the number of initial ELM, in this paper M equals 50.By saving the parameters of the selected ten ELMs, a combined ELM can be generated to retrieval image.

Original ELM image using a single network to predict the emotional category, since a single ELM's performance has randomness and uncertainty, we propose using a combined ELM to forecast, using the max probability output of the combined ELM as the final prediction. The prediction process for the combined ELM: For a test sample image (x, t), using the i-th ELM of the combined ELM to predict, its corresponding input values are w^i and bi, the prediction result is o^i; Use ten trained ELM of the combined ELM to predict the emotional classification of testing image, get ten predict result, summarize the ten

result obtains O_{ens} , calculate the proportional share of each result, such as (2.7) as follows:

$$O_{ens} = \frac{1}{Num}\sum_{i=}^{Num} o_i \quad (7)$$

Select the max proportional share emotional classification from O_{ens} , as seen in (2.8), we can get the classification result of the testing image.

$$O_{final} = \max\left\{ \overset{i}{O}_{ens} \right\} \quad (8)$$

Where $\overset{i}{O}_{ens}$ is the i-th classification of O_{ens} , $i = 1,2,...Num$.

b) Performance analysis of modified ELM

To verify the performance of the combined ELM, we compare it to the performance of original ELM and SVM. The training set is some images of satellites and letters, its related parameters as shown in table II, choose $g(x) = \frac{1}{e^{-x}+1}$ as activation function.

TABLE II. PARAMETERS OF TRAINING DATA SET

Dataset	Training Capacity	Testing Capacity	Number of Hidden layer nodes
Satellite	1500	410	100
Letter	12000	4000	200

TABLE III. TRAINING PERFORMANCE COMPARISON OF MODIFIED ELM WITH ORIGINAL ELM AND SVM

Dataset	Algorithm	Training time(s)	Training accuracy(%)
Satellite	Modified ELM	786.42	87.31
	Original ELM	0.76	84.69
	SVM	6268.36	79.68
Letter	Modified ELM	1889.38	80.07
	Original ELM	2.16	78.65
	SVM	48865.33	73.85

Through 20 times experiment, we obtain such training result shown in Table III, which is the average result of 20 experiments. From the result, we can see that although the training time of modified ELM is much longer than original ELM and SVM, due to modified ELM needs multiple times of training and to modify the parameter by using genetic algorithm, generally it needs much shorter training time than SVM. In terms of training accuracy, modified ELM is much higher than original ELM and SVM on both dataset, which proved that the improved ELM can improve training accuracy rate.

To verify the testing performance of the combined ELM, we test the combined ELM based on the dataset shown in Table II, and compare it to the performance of original ELM and SVM. Experiments conducted 20 times, the average result is shown in Table IV.

TABLE IV. TESTING PERFORMANCE COMPARISON OF MODIFIED ELM WITH ORIGINAL ELM AND SVM

Dataset	Algorithm	Testing accuracy (%)	Testing mean square error (%)
Satellite	Modified ELM	89.24	0.163
	Original ELM	86.69	0.186
	SVM	83.28	0.218
Letter	Modified ELM	82.96	0.189
	Original ELM	79.63	0.233
	SVM	77.47	0.692

The above table shows that modified ELM algorithm has advantages in testing accuracy when comparing with original ELM and SVM algorithm. Meanwhile, it has much lower mean square error of testing accuracy, which shows that modified ELM has a stable performance and significantly improve the robustness.

III. PROTOTYPE SYSTEM

Based on the core algorithms described in the previous section, we designed and implemented prototype system for image emotional semantic retrieval based on the improved ELM. On the prototype system, the user can through two ways, namely text input retrieval and image input retrieval to obtain the required emotional images.

In text input retrieval mode, the user input hope emotion semantic text, the system matches the image library created by training, finds out the similar image with user input on emotion semantics in the image library and, and returns the retrieved results to the user. Its advantage is fast and accurate. Users can retrieve the exact emotional semantic image from the system.

In image Input retrieval, users upload to retrieve images through the user interface module. The system extracts the feature vector from the input images, executes emotion recognition by trained ELM, maps it to a similar emotional semantic classification, and returns the similar emotional semantic images to users. Image input approach has the advantages of more intuitive and convenient. Users do not enter clear semantic keywords, only by uploading an image, the system automatically complete the emotion recognition and returns the retrieval results, realizing intelligent interactions between the system and human. Therefore, image input retrieval can overcome the semantic gap between image features and the high-level emotional bottom effectively.

Image emotional semantic retrieval system based on ELM is developed by modular development. The prototype system consists of image base constructing module, feature extracting module, image emotion matching module and user interface module, the system block diagram is shown as figure 1.

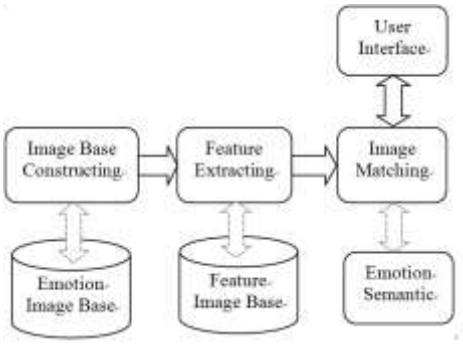

Figure 1. System Block Diagram

A. *Image base constructing module*

The function of the module includes original image preprocessing, the size and format of the image adjusting, image emotion tag labeling by user annotations experiment, and the corresponding image and emotion classification matching index creating by statistical.

B. *Feature extracting module*

The function of the feature extracting module includes extracting image feature, decreasing the dimension of the feature vectors by PCA, storing the final feature vectors in system image base, and building index between image feature vectors and the emotion classification. Since then, the training of the ELM and image retrieval is based on image feature vectors, no longer on the image itself so as to reduce the processing time effectively and improve the efficiency of retrieval.

C. *Image matching module*

According to different choice of emotional input mode, the image emotion matching process will also be different. When a user selects a text as emotional input, the system reads input from the user's emotional keywords, and emotion classification matching in image library. When the user selects a sample image as emotional input, the system will matches the emotion semantics in sample image with images in system library.

D. *User interface module*

The prototype system provides users with a friendly interactive interface, i.e. two different input modes, namely the text input approach and image input approach. When users input image emotional text to be retrieved, such as input "happy", the system executes emotion semantic matching and estimates whether there is the corresponding emotion classification in the system. If successful, it returns the corresponding images to the user. If no corresponding emotion classification is found, users can upload a sample image to search the similar emotional images.

Figure 2 and figure 3 are the results by text search and image search respectively.

Figure 2. Result by text search

Figure 3. Result by image search

IV. CONCLUSION

In this paper, image emotion semantics has been explored based on ELM, including the basis of emotional cognitive, the dimension of emotional expression, the establishment of scientific emotion space, the selection of representative images, feature extraction and dimension reduction, and the improvement of extreme learning machine. The experimental results show that prototype system for image emotion semantic retrieval achieved initial success. The future task is to carry out further research on emotion expression and semantic matching so as to achieve the practical level for image emotion semantic retrieval.

ACKNOWLEDGMENT

This paper is the partial achievement of Project 2013CB329504 supported by National Key Basic Research and Development Program (973 program), and project 2012C21002 supported by Science Technology Department of Zhejiang Province.

REFERENCES

[1] Dorai C, Venkatesh S. Bridging the Semantic Gap with Computational Media Aesthetics. IEEE Multimedia 2003; 10(2):15-17.

[2] Aamir S M, Humaira N. Edge Refinement Method for Content-Based Image Retrieval. IEEE, 1999: 921-924.

[3] Peter E, Christine S. Towards a comprehensive survey of the semantic gap in Visual image retrieval, Lecture notes in computer science, 2003:163-1 68.

[4] Wang N W, Yu Y L, Jiang S M. Image Retrieval by Emotional Semantics: A Study of Emotional Space and Feature Extraction. IEEE, 2006: 3534-3539.

[5] Schlosberg, H. Three dimensions of emotion. Psychological Review, 1954, 61: 81-88.

[6] HUANG G B, ZHU Q Y, SIEW C K. Extreme learning machine: a new learning scheme of feedforward neural networks [J]. Neurocomputing, 2004, 2(2): 985-990.

[7] G-B Huang, L. Chen, and C.-K.Siew, "Universal approximation using incremental feedforward networks with arbitrary input weighs," in Technical Report ICIS/46/2003, (School of Electrical and Electronic Engineering, NanYang Technological University, Singapore), Oct. 2003.

[8] Huang Gang-bin. Learning capability and storage capacity of two-hidden-layer feedforward networks[J]. IEEE Trans on Neural Networks, 2003, 14(2):274-281.

[9] Bartlett PL (1998), the sample complexity of pattern classification with neural networks: the size of the weights is more important than the size of the network. IEEE Trans Info Theory44(2): 525–536.

[10] Goldberg D E. Genetic Algorithms in Search, Optimization and Machine Learning [M]. MA: Addison Wesley, 1989: 1- 83.