

Extreme Learning Machine with Gaussian Kernel Based Relevance Feedback Scheme for Image Retrieval

Lijuan Duan, Shuai Dong, Song Cui and Wei Ma

Abstract As for the huge gap between the low-level image features and high-level semantics, content-based image retrieval still could not receive a satisfying result by now. Since the special request of the relevance feedback, making full use of the rare number of labeled data and numerous unlabeled data is an ideal way. Because ELM has excellent classification accuracy and processing time, and high accuracy and fast speed are the key factors to evaluate the relevance feedback performances. In this paper, we proposed an Extreme learning Machine with Gaussian kernel Based Relevance Feedback scheme for image retrieval, to overcome the above limitations, our method uses three component classifiers to form a strong learner by learning different features extracted from the hand-marking data, then we use it to label the image database automatically. From the experiments we can see the use of the ELM with kernel have high classification accuracy, the processing time get largely decreased at the same time. Thus, it improves the efficiency of entire relevance feedback system. The experiments results show that the proposed algorithm is significantly effective.

Keywords Image retrieval · Relevance feedback · Extreme learning machine · Unlabeled data

L. Duan (✉) · S. Dong · S. Cui · W. Ma

Beijing Key Laboratory of Trusted Computing, College of Computer Science and Technology, Beijing University of Technology, Beijing 100124, China
e-mail: ljduan@bjut.edu.cn

S. Dong
e-mail: dongse@emails.bjut.edu.cn

S. Cui
e-mail: cuisong@emails.bjut.edu.cn

W. Ma
e-mail: mawei@bjut.edu.cn

L. Duan · S. Dong · S. Cui · W. Ma

Beijing Key Laboratory on Integration and Analysis of Large-Scale Stream Data, College of Computer Science and Technology, Beijing University of Technology, Beijing 100124, China

1 Introduction

In recent years, with the explosive increase in volume of digital images, content-based image retrieval technique is getting popular, lots of systems had been developed in the decades, including QBIC, Photobook, MARS, PicHunter and others [1, 2]. In a CBIR system, the system fails to be so perceptive to the user's intention that cannot return the satisfactory results, which mostly owing to the huge gap between the high-level semantic concepts and the low-level image features. In that way, relevance feedback methods were proposed [3]. And this technique had widely applied in vary content-based image retrieval systems [4, 5].

In this paper, we proposed an Extreme learning Machine (with Gaussian kernel) based relevance feedback approach. Firstly, we extract the features of labeled data and use them to train the component ELM classifiers. Secondly, we predict the whole database based on the tri-training method and vote for the result. Thirdly, the new labeled data are used to retrain the component classifiers and get the final learner group.

Cox et al. [6] proposed an interactive relevance feedback scheme based on Bayesian model by optimizing the features' probability distribution. Rui et al. [3] proposed an optimizing scheme for the relevance feedback performance by analyzing the features of positive samples. Zhou et al. [7] combined the semi-supervised method with active learning method by labeling the uncertain samples, which could have the unlabeled-data used.

The rest of this paper is organized as follows. Section 2 describe the method we proposed. Section 2.1 contains the description of our ELM with kernel based relevance feedback system. Section 3 describes the experiments and the performance of our scheme. Section 4 concludes this paper.

2 Method

2.1 Extreme Learning Machine with Gaussian Kernel

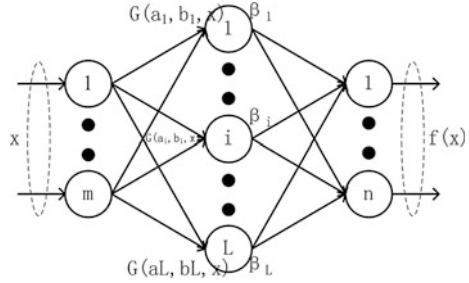
Extreme learning machine (ELM) was first proposed by Huang et al. [8]. ELM works for generalized single-hidden layer feed forward networks (SLFNs) [9, 10]. ELM algorithm tends to provide better generalization performance at extremely fast learning speed. Structure of the SLFNs [10] is shown as Fig. 1.

With L hidden nodes in output layer, the output function of SLFNs can be expressed by:

$$f_L(\mathbf{x}) = \sum_{i=1}^L \beta_i g_i(\mathbf{x}) = \sum_{i=1}^L i \beta_i G(\mathbf{a}_i, \mathbf{b}_i, \mathbf{x}). \quad (1)$$

The function also can be written as $f(\mathbf{x}) = \mathbf{h}(\mathbf{x})\boldsymbol{\beta}$ where $\boldsymbol{\beta} = [\beta_1, \beta_2, \dots, \beta_L]$ is the vector of the output weights between the hidden layer of L neurons and the

Fig. 1 Single-hidden layer feed forward networks



output neuron and $\mathbf{h}(x) = [h_1(x), h_2(x), \dots, h_L(x)]$ is the output vector of the hidden layer with respect to the input x , which maps the data from input space to the ELM feature space [11].

In particular, L equals to N , which is rare condition because L is far smaller than N in actual problem, that is to say, that there is error between the output value and the actual value. So, the most important thing is to find least-squares solution β of the linear system.

$$\mathbf{H}\beta = \mathbf{T}, \quad \beta = \mathbf{H}^\dagger \mathbf{T}. \tag{2}$$

where \mathbf{H}^\dagger is the Moore-Penrose Generalized inverse of matrix \mathbf{H} [12, 13], $\mathbf{H}^\dagger = (\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'$ or $\mathbf{H}'(\mathbf{H}\mathbf{H}')^{-1}$, depending on the singularity of $\mathbf{H}'\mathbf{H}$ or $\mathbf{H}\mathbf{H}'$.

In the newly developed kernel ELM, it's getting more stable to introduce a positive coefficient into the learning system. If $\mathbf{H}'\mathbf{H}$ is nonsingular, the coefficient $1/\lambda$ is added to the diagonal of $\mathbf{H}'\mathbf{H}$ in the calculation of the output weights β . After that, $\beta = \mathbf{H}'(\mathbf{I}/\lambda + \mathbf{H}\mathbf{H}')^{-1} \mathbf{T}$, the corresponding function of the regularized ELM is:

$$f(\mathbf{x}) = \mathbf{h}(\mathbf{x})\beta = \mathbf{h}(\mathbf{x})\mathbf{H}' \left(\frac{1}{\lambda} \mathbf{I} + \mathbf{H}\mathbf{H}' \right)^{-1} \mathbf{T}. \tag{3}$$

Huang et al. [11] shown that ELM with a kernel matrix can be defined as follows. Let $\mathbf{\Omega}_{ELM} = \mathbf{H}\mathbf{H}': \mathbf{\Omega}_{ELM_{i,j}} = \mathbf{h}(\mathbf{x}_i)\mathbf{h}(\mathbf{x}_j) = K(\mathbf{x}_i, \mathbf{x}_j)$. The output function can be written as:

$$f(\mathbf{x}) = \mathbf{h}(\mathbf{x})\mathbf{H}' \left(\frac{1}{\lambda} \mathbf{I} + \mathbf{H}\mathbf{H}' \right)^{-1} \mathbf{T} = \begin{bmatrix} K(\mathbf{x}, \mathbf{x}_1) \\ \vdots \\ K(\mathbf{x}, \mathbf{x}_N) \end{bmatrix}' \left(\frac{1}{\lambda} \mathbf{I} + \mathbf{\Omega}_{ELM} \right)^{-1} \mathbf{T}. \tag{4}$$

The hidden layer feature mapping $\mathbf{h}(\mathbf{x})$ need not to be known, and instead its corresponding kernel $K(\mathbf{u}, \mathbf{v})$ can be computed. In this way, the Gaussian kernel is used, $K(\mathbf{u}, \mathbf{v}) = \exp(-\gamma\|\mathbf{u} - \mathbf{v}\|_2)$ [14].

2.2 Algorithm Description

Due to the special requirement of relevance feedback, short processing time and high accuracy are the key points. Our method shown the promising effect.

Step 1: Label the data.

In the image retrieval system we proposed, let L denote the labeled data set, U denote the unlabeled data set. While $L = P \cup N$, where P denote the labeled positive samples and N denote the labeled negative samples.

Step 2: Feature extraction and train the component classifiers.

Inspired by the co-training paradigm [15], the labeled data used as the data of first train, and extract three different features to train the component classifier to reduce the variance of unstable procedures, which leading to improved prediction [16].

Step 3: Vote.

The whole database as known as U is put into the combined-classifier-group. From the Fig. 2 we can see the framework of voting procedure. Each component classifier will give its predicted label of i th trail of U data set, and the output is set as Y_{in} . All the trails are fed into m number of sub-ELM. Finally, a class label set ψ can be got.

Step 4: Label the unlabeled data automatically.

Let L_{temp} denote the whole database with labels. L_{temp} is used to retrain the component classifiers in order to level up the classification ability. From the Fig. 3, the dataset U has been classified by updating the classifier group.

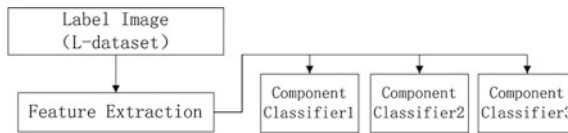


Fig. 2 The framework of voting

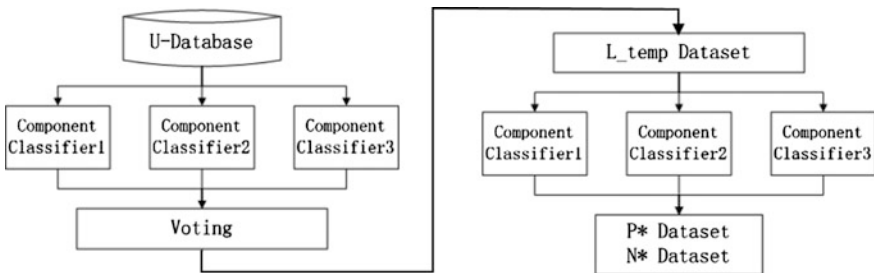


Fig. 3 The framework of labeling the unlabeled data

For the i th trail, P_i is the total number of the component classifiers which predict the class of i th trail to be 0; Q_i is the total number of the component classifiers which predict the class of i th trail to be 1. The sum of P_i and Q_i is 3. The vote function can be shown as bellow:

$$\text{vote}(\Psi) = \begin{bmatrix} Y_1 \\ \vdots \\ Y_M \end{bmatrix}_{M \times 1}, Y_i = \lfloor \frac{Q_i}{P_i} \rfloor \tag{5}$$

Step 5: Update the classifiers.

As the result, let P^* denote the positive dataset, N^* denote the negative dataset. In each iteration of the feedback, the whole result list is returned by the ascend order of the similarity rank.

3 Experiments

In this section, we firstly introduce the image database our experiments performed on. Secondly we describe the methods of the feature extraction. Then we illustrate the structure of our relevance feedback system. Finally, we compare the performance of our methods with other methods in the aspect of the average-AP values and the aspect of processing time.

3.1 Image Database

We perform our experiments on the COREL photo gallery, which contains 1000 images into 10 categories, and 6000 images into 60 categories. The size of every image in the database is 256×384 or 384×256 . A ground truth of the image database is needed to evaluate the performance of our experiment. Thus, the natural categories of the COREL photo gallery are used as the semantic categories, and we define that images belong to the same category are relevant, otherwise, are irrelevant.

3.2 Feature Selection

For the image retrieval system with relevance feedback, use 3 image features are used: SIFT, color and LBP.

Based on the former work in our laboratory, using Bag-of-Features model to extract two image features which based on SIFT and color, respectively. The color

information is the most informative feature because of its robustness with respect to scaling, rotation, perspective, and occlusion.

Local Binary Pattern as known as LBP which is an effective image descriptor, it has been used as the texture feature of the image.

3.3 Image Retrieval System

For the image relevance feedback, the image retrieval system is required to return the images which are the most semantically relevant. Figure 4 shows our system has three main parts. The feedback part is the key function in the system.

- Retrieval part: Extract three features which mentioned above of every image in the image database. Computing the Euclidean distance as the similarity between the query and each image of the database, and return the top 20 images as the retrieval result.
- Feedback part: Let the user choose and label the negative images, extract the image features, the images that displayed and without labeled are set positive. Training classifier groups for the each kind of features respectively, to label the unlabeled database automatically. Updating the classifier groups by using the new-labeled data, as the result, the final classifier is using for classifying the unlabeled database.
- Display part: Displaying the retrieval result for the retrieval operation, displaying the feedback result of the each iteration for the relevance feedback operation.

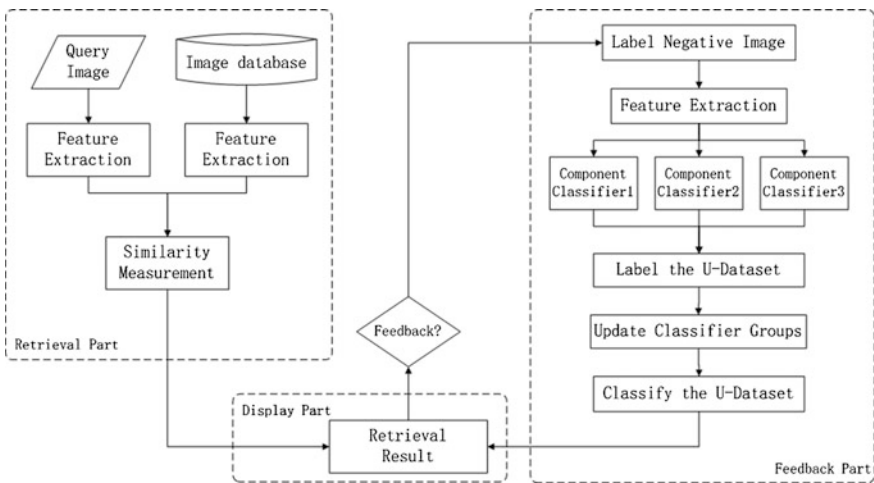


Fig. 4 The framework of image retrieval system with relevance feedback

3.4 Experiment Results

Before the relevance feedback procedure, the top 20 images are result which displayed on the panel. The user interface is shown on the Fig. 5. The image that at the top of left is the query image and the other images on the ‘result panel’ are the retrieval result. User label the ‘negative’, the rest images are labeled ‘positive’ automatically.

From Fig. 6 has shown only 4 images of the retrieval result meet our request. After the first iteration of relevance feedback, the result had notable improvement, the result shown on Figs. 7, 8 and 9 show the feedback results by iteration 1, 2, 3, respectively. The ‘right’ images are getting more, semantically, the result is more and more meets our request. It took us only 3 iterations to have all the results right.

We use the Average Precision (AP) measure. The retrieval result has been optimized by performing our relevance feedback scheme. At the each iteration of relevance feedback, The AP value can be obtained, which is defined as the average of precision value obtained after each relevant image is retrieved. Let \bar{P} denotes the average precision which is obtained at the current iteration, and it is computed by:

$$\bar{P} = \sum_{E_i \in R} \frac{P_i}{|S|} \tag{6}$$

where P_i denotes the precision value obtained after the system retrieves i top-ranked images, E_i is the element of the relevant images set, S is the set of all relevant images that belong to the same category as the query, and $|S|$ denotes the cardinality of S . The AP calculated over all the relevant images can avoid the fluctuation of precision that is usually encountered by the traditional precision measurement.

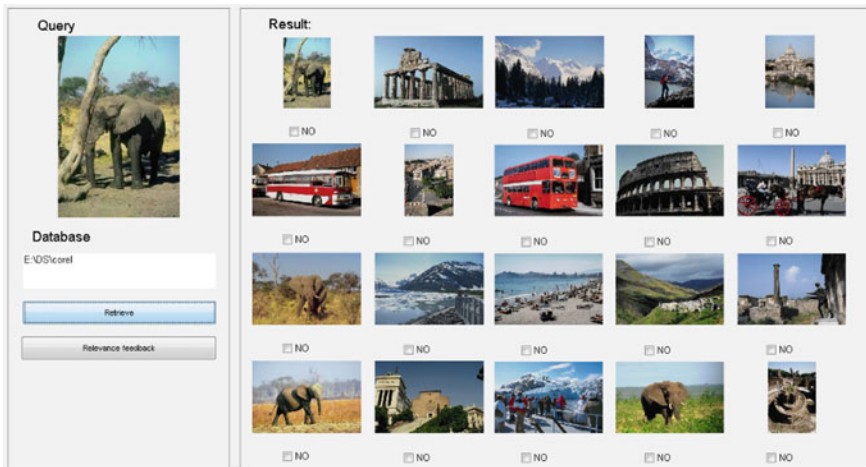


Fig. 5 The user interface and the retrieval result

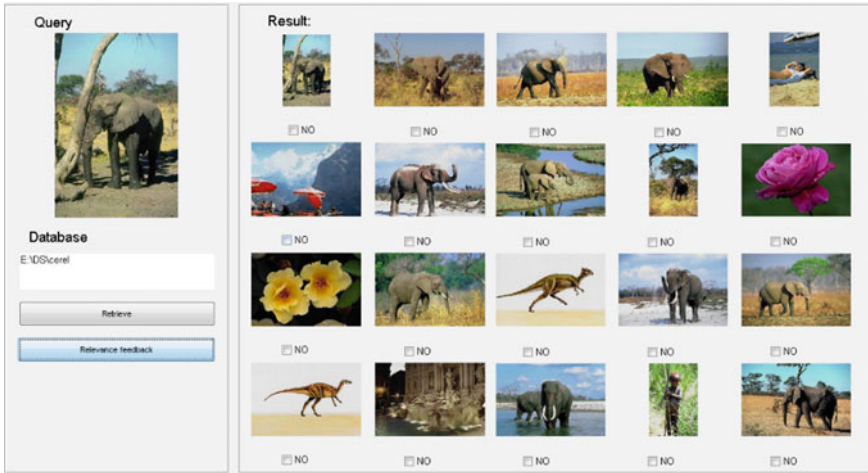


Fig. 6 The result of first relevance feedback iteration

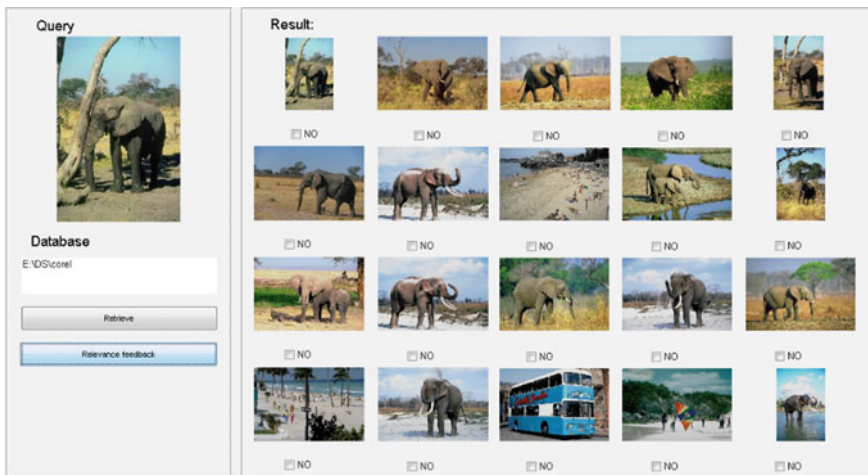


Fig. 7 The result of second relevance feedback iteration

We report the feasibility and utility of our algorithm, and compare it with five feedback methods. Figures 9 and 10 show the average-AP value of different method.

According to the uncertainty of the ELM which brought by the random parameters, 3 ELM are used classifiers to get a stable component-classifier, respectively. Thus, 9 elm classifiers are used to compose the component-classifiers. Because of the randomness of the ELM, the experiments performed 20 times to get the stable data. From the figures we can see that the result of 9-elm-classifiers

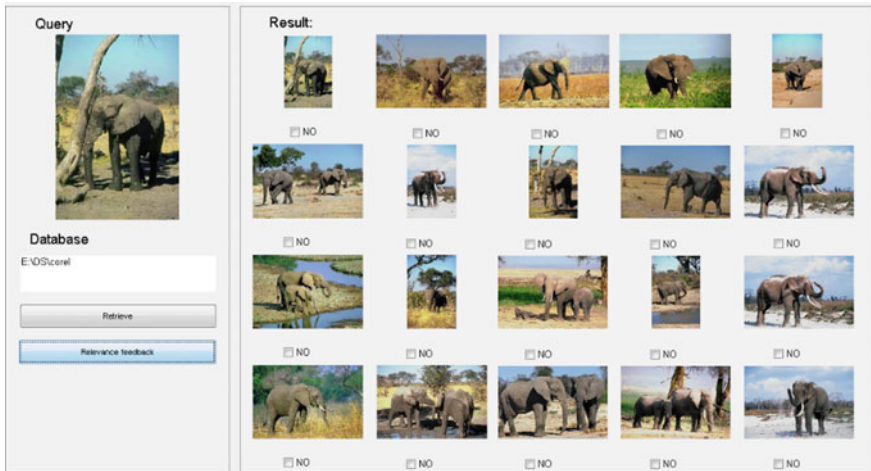
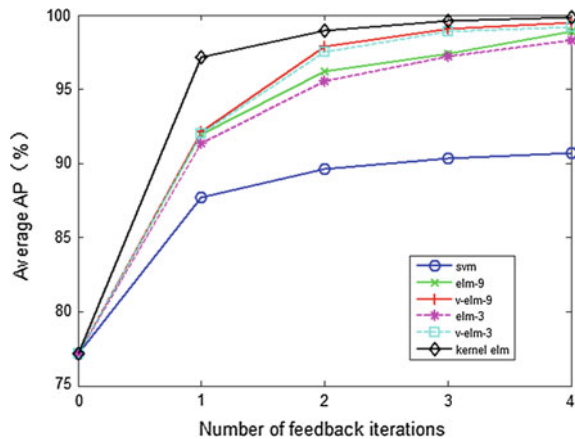


Fig. 8 The result of third relevance feedback iteration

Fig. 9 The average-AP values of different method performed on 10 categories



method is very close to the result of 3-elm-classifiers method, so 3-elm-classifiers method is better in order to reduce the unnecessary processing time. The average-AP values of the kernel ELM method are superior for the 4 iterations. And the final iteration our method perform the most perfect result than the other methods.

Figures 11 and 12 show the processing time of various methods getting short as the number of feedback iterations increases. And they show our methods based on ELM with Gaussian kernel have enormous advantage in each iteration whatever on 10 categories or 60 categories which is the most important factor to the relevance feedback. Eventually, from the result of accuracy and time, the effectiveness of the ELM with Gaussian kernel based relevance feedback scheme is utilized.

Fig. 10 The average-AP values of different method performed on 60 categories

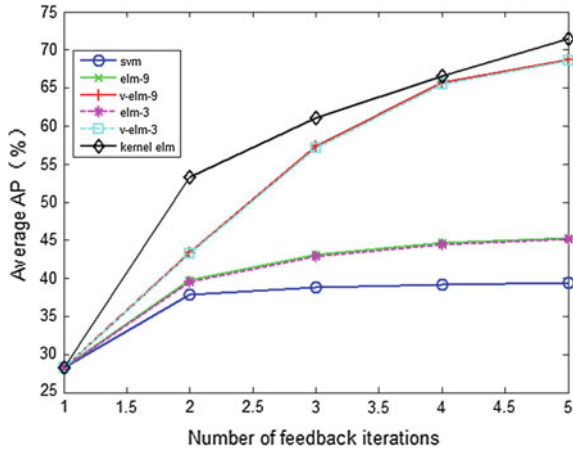


Fig. 11 The processing times of various methods performed on 10 categories

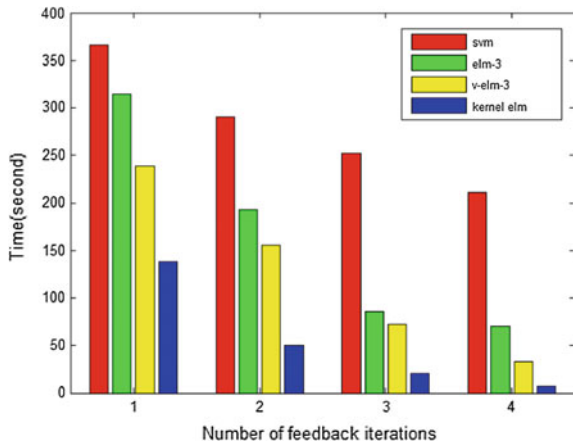
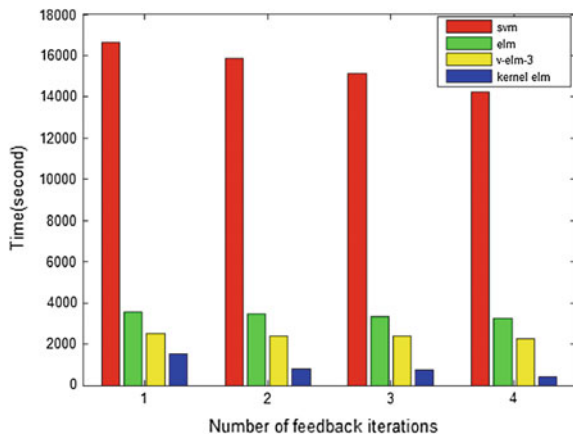


Fig. 12 The processing times of various methods performed on 60 categories



4 Conclusions

Recently, Extreme Learning Machine has been widely applied in relevance feedback. It's an important and efficient way to improve the performance of image retrieval. Most of the advantages of ELM are suitable for the relevance feedback, such as: short processing time, high accuracy of classification, and good generalize ability, etc. However, the conventional feedback relevance schemes could not give considerations to the both accuracy and speed. To combine ELM with the semi-supervised method, we can overcome the limitation of the conventional problems. The experiments performed on the Corel-Photo gallery shows that our new method can have an excellent performance.

Acknowledgement This research is partially sponsored by Natural Science Foundation of China (Nos. 61003105, 61175115 and 61370113), the Importation and Development of High-Caliber Talents Project of Beijing Municipal Institutions (CIT&TCD201304035), Jing-Hua Talents Project of Beijing University of Technology (2014-JH-L06), and Ri-Xin Talents Project of Beijing University of Technology (2014-RX-L06), and the International Communication Ability Development Plan for Young Teachers of Beijing University of Technology (No. 2014-16).

References

1. Meng, W., Hao, L., Tao, D., et al.: Multimodal graph-based reranking for web image search. *IEEE Trans. Image Process.* **21**(11), 4649–4661 (2012)
2. Hu, W., Xie, N., Li, L., et al.: A survey on visual content-based video indexing and retrieval. *IEEE Trans. Syst. Man Cybernet. Part C Appl. Rev.* **41**(6), 797–819 (2011)
3. Rui, Y., Huang, T.S., Ortega, M., et al.: Relevance feedback: a power tool for interactive content-based image retrieval. *IEEE Trans. Circuits Syst. Video Technol.* **8**(5), 644–655 (1998)
4. Auer, P., Leung, A.P.: Relevance feedback models for content-based image retrieval. *Stud. Comput. Intell.* **346**, 59–79 (2011)
5. Wang, M., Ni, B., Hua, X.S., et al.: Assistive tagging: a survey of multimedia tagging with human-computer joint exploration. *ACM Comput. Surv.* **44**(4), 1173–1184 (2012)
6. Cox, I.J., Miller, M.L., Minka, T.P., et al.: An optimized interaction strategy for bayesian relevance feedback. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'98)*, pp. 553–558 (1998)
7. Zhi-Hua, Z., Ke-Jia, C., Yuan J.: Exploiting unlabeled data in content-based image retrieval. In: *Proceedings of ECML-04, 15th European Conference on Machine Learning*, pp. 525–536 (2004)
8. Huang, G.B., Zhu, Q.Y., Siew, C.K.: Extreme learning machine: a new learning scheme of feedforward neural networks. *Proc. Int. Joint Conf. Neural Netw.* **2**, 985–990
9. Huang, G.B., Zhu, Q.Y., Siew, C.K.: Extreme learning machine: theory and applications. *Neurocomputing* **70**, 489–501 (2006)
10. Huang, G.B., Wang, D.H., Lan, Y.: Extreme learning machines: a survey. *Int. J. Mach. Learn. Cybernet.* **2**(2), 107–122 (2011)
11. Huang, G.B., Zhou, H., Ding, X., et al.: Extreme learning machine for regression and multiclass classification. *IEEE Trans. Syst. Man Cybernet. Part B Cybernet.* **42**(2), 513–529 (2012). A Publication of the IEEE Systems Man and Cybernetics Society
12. Serre, D.: *Matrices: theory and applications.* *Mathematics* **32**, xvi, 221 (2002)

13. Dwyer, P.S., Rao, C.R., Mitra, S.K., et al.: Generalized inverse of matrices and its applications. *J. Am. Stat. Assoc.* (1971)
14. Huang, G.B.: An insight into extreme learning machines: random neurons, random features and kernels. *Cognit. Comput.* **6**(3), 376–390 (2014)
15. Blum, A., Mitchell, T.: Combining labeled and unlabeled data with co-training. In: *Colt Proceedings of the Workshop on Computational Learning Theory*, pp. 92–100 (1998)
16. Breiman, L.: Bagging predictors. *Mach. Learn.* **24**(2), 123–140 (1996)