A New Method to Generate Semantic Templates Based on Multilayer Perceptron

Ya-Li Qi, Guo-Shan Zhang and Ye-Li Li

Abstract Content-based image retrieval pays more attention to reducing semantic gap. Semantic template is a promising method for reducing semantic gap, and consists of mapping between high-level and low-level visual features. The work presented here proposes a semantic template method via multilayer perceptron, which has three layers: an input layer, a hidden layer, and an output layer. In the proposed method, the pixel features of an interesting region are selected as input features, the features weights are originally designed randomly with random seeds, and softmax is selected as the activation function. Experiments show the proposed method has high accuracy for image retrieval, and the accuracy can be improved by adding samples to train the MLP (Multilayer perceptron) classifier until a relative stable state is achieved.

Keywords Content-based image retrieval (CBIR) · Semantic gap · Multilayer perceptron (MLP) · Semantic template

1 Introduction

In the early 1990s, the content-based image retrieval (CBIR) system was introduced [1]. The largest challenge to effective image retrieval using CBIR is a semantic gap between low-level feature and high-level semantics, which prevents its wide

Y.-L. Qi · G.-S. Zhang
School of Electrical Engineering and Automation, Tianjin University, Tianjin, China
e-mail: zhanggs@tju.edu.cn

Y.-L. Qi (✉) · Y.-L. Li
The Computer Department, Beijing Institute of Graphic Communication, Beijing, China
e-mail: qyl@bigc.edu.cn

Y.-L. Li
e-mail: liyl@bigc.edu.cn

A. Hussain (ed.), *Electronics, Communications and Networks V*,
Lecture Notes in Electrical Engineering 382, DOI 10.1007/978-981-10-0740-8_5

application as a solution [2, 3]. In the past few decades, many algorithms have been proposed to reduce semantic gap. These methods can be classified into five categories [2]: object ontology, machine learning, relevance feedback (RF), semantic template (ST), textual information, and visual content, for Web image retrieval. In real image retrieval systems, a synthetic approach based on the five methods can be used in CBIR.

ST is not yet as widely used as the other techniques, but it is a promising approach in CBIR [1]. ST constructs a bridge between semantic features and visual features to reduce semantic gap. It is usually defined as a 'representative' concept, and calculated from a set of sample images [4, 5].

Smith and Li use composite region templates (CRTs) to decode image semantics [4], which defines the region as a semantic description, and describes the prototypal spatial arrangements of the semantic in the images. Chang et al. use semantic visual templates (SVT) to map low-level visual features to high-level semantics for video retrieval, which depends on the user's feedback [6]. Zhang et al. generate ST automatically in the process of RF [7]. In addition, wordnet is also used to construct a network of ST [8]. In the retrieval process based on the ST, once the user proposes a query concept, the system can find a corresponding semantic template. It uses the centroid feature and the corresponding weight to find similar images. Based on this method, Zhang et al. propose an algorithm to generate weak semantics [9]. Sarwar et al. propose a retrieval system using a corpus of natural scene images via imparting human cognition [10]. Hu et al. directly use Gradient Field HOG (GF-HOG) and Bag of Visual Words (BoVW) to define semantic template [11]. In addition in recent years, some scholars use annotation methods to extract the image semantic and generate semantic templates [12–14]. These annotation methods usually retrieve images according to the similarity distance of the semantic templates.

Here we propose an algorithm to generate semantic templates based on multilayer perceptron (MLP). The algorithm uses the pixel features of interesting regions as input data and generates a random weight vector with random seed for the hidden layer of MLP. To achieve highly accurate classification, it selects softmax as an activation function for output layer of MLP.

2 MLP

MLP is an artificial neural network model based on feedforward, which maps input data to a series of appropriate outputs [15]. It consists of multiple layers with multiple nodes, and each layer is fully connects to the next one (see Fig. 1). MLP uses backpropagation to train the network, which is a supervised learning technique. It can solve the non-linear problems which cannot be solved by single-layer perceptron.

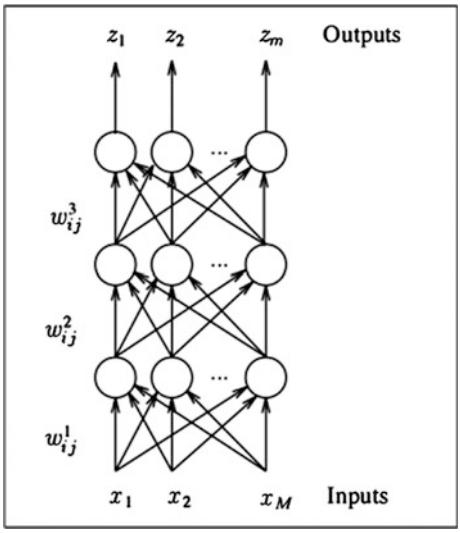

Fig. 1 Multilayer perceptron

2.1 Layers Structure

The MLP usually consists of an input layer and an output layer. The output layer usually has one or more hidden layers, which include some nonlinearly-activating nodes. Each node in the front layer connects with a weight W_{ij} to every node in the following layer (Fig. 1).

2.2 Activation Function

Each neural node (except input nodes) has an activation function. There are three types of activation functions: linear, logistic and softmax. Linear activation functions are only copied the data by treating node, which is generally used for regression problems, and not suited for classification problems. Logistic activation functions are usually used for multiple classification problems, which generate multiple independent logical attributes as outputs. Softmax activation functions are usually used for common classification problems. The output of this type function is mutually exclusive classes [15].

2.3 Learning Through Back-Propagation

MLP is a supervised learning model based on back-propagation. After a piece of data is processed, the perceptrons change the connection weights by comparing the

difference between the error in the output and the expected result. The output layer weights are modified via the activation function derivation, and the hidden layer weights are changed according to the activation function back-propagation.

3 Experiment

3.1 Generating Semantic Template Based on MLP

To train the MLP classifier with three layers, we selected an interesting region to represent the semantic template features, and used the pixel feature of the interesting region to train MLP classifier. The corresponding output is the semantic perception.

- S1: For the input layer, the pixel feature is the color feature of the pixel, and the pixel feature is used to describe the interesting region as an object. As such, the feature vectors are color feature vectors of pixels in the region of interest. The feature vectors need to be normalized. To do so, the initial feature vectors need substract the mean of the training vectors, and divide the standard deviation of the training vectors components. Therefore, the processed feature vectors have a mean of 0 and a standard deviation of 1.
- S2: For the hidden layer, the weights are initialized randomly. To ensure the reproducible results, a random seed may be passed to the random number generator. When the training results have a relatively large error, the model may select a different value for random seed and retrain the MLP to achieve a smaller error.
- S3: For the output layer, the activation function selects softmax.
- S4: In the retrieval process, we segment the image by a watershed algorithm, which segments the objects form background in image following reference [16]. Then input sub region into MLP classifier. If there is relevant semantic perception in the image, the MLP classifier will classify this image to the relative semantic class as retrieval result.

3.2 Experimental Results

All experiments in presented here were performed on a Lenovo workstation with Intel Core2.5 GHz CPU and Halcon11. The experimental images were selected from http://image.baidu.com/.

In our experiments, we generated two semantic templates: one is ‘orange’, the other is ‘lemon’. We selected a region of interest for each object as inputs for the semantic template classifier (see Fig. 2). Then train the MLP classifier for semantics: orange and lemon respectively.

We used the pixel feature of the region of interest to train MLP classifier. Then we normalized the feature vectors and designated 42 as the random seed, which is based on experience [15]. The region rounded by the orange line represents the orange semantic, and the region rounded by the yellow line represents the lemon semantic. There are two types error of the result: one is the orange image is classified to lemon semantic, or the lemon image is classified to orange semantic; the other is orange image or lemon image is classified into two semantic classes. The results of the classification by MPL are shown in Fig. 3.

For this experiment, we selected two samples to train MLP. One is orange, the other is lemon. The classifying data are 100 images, and the number of correct results is 66. To improve the accuracy of this algorithm, we added the samples to train semantic template classifier. One experiment selected four samples to train MLP, two orange, and two lemons. The other experiment selected six samples to train MLP-three orange samples and three lemon samples. We have performed the experiments with a larger number of samples. We selected eight, ten and twenty samples to train the algorithm. As Table 1 shows, the accuracy of the classifier can be improved via adding samples to train the MLP classifier until achieving a relative stable state, and the accuracy of classification is not significantly improved by increasing the size of training set.

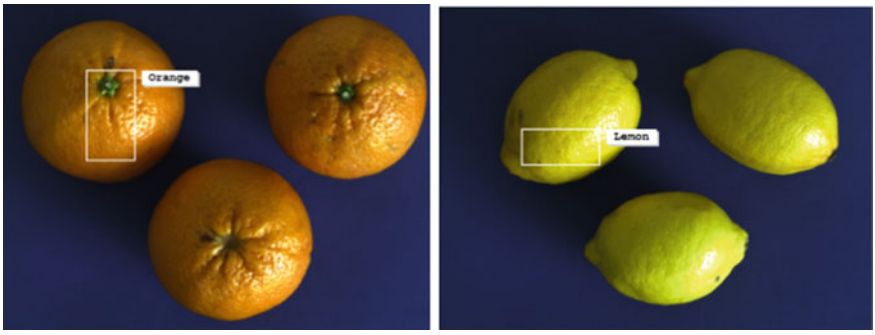

Fig. 2 Select interesting region of object as input for orange and lemon semantic

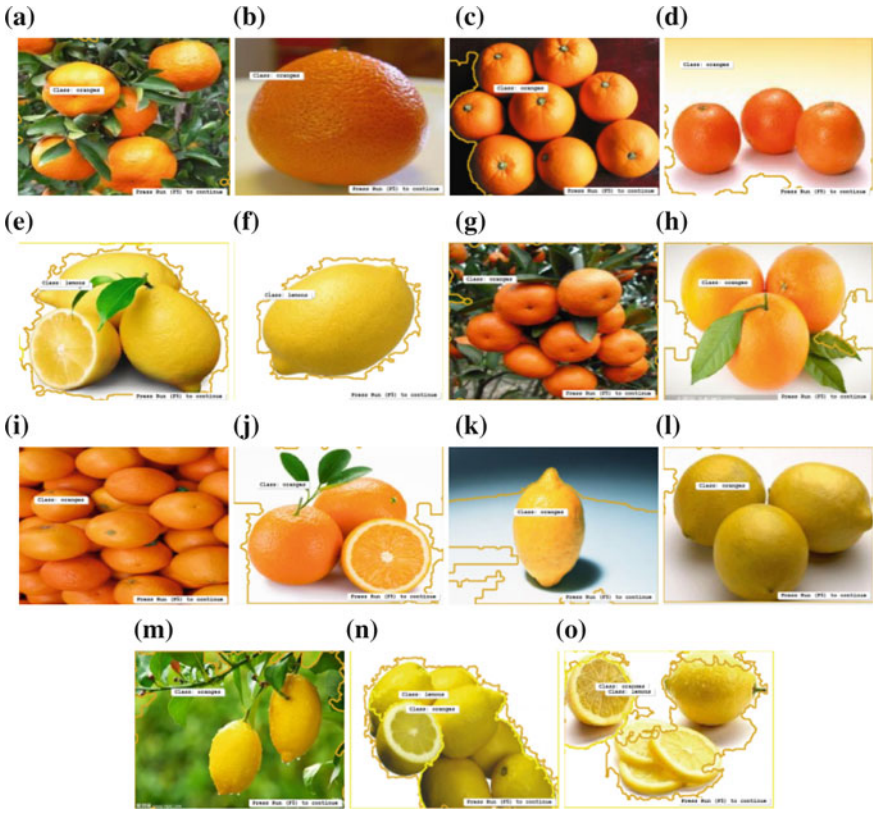

Fig. 3 The results of the classification based on the semantic template classifier. *1* The sub figures (**a**–**j**) show the correct classifying results; *2* The sub figures (**k**–**m**) show the error classifying results; *3* The sub figures (**n**, **o**) show the error classifying into two class result

Table 1 The results of the classify based on the semantic template classifier

The number of the samples to train MLP		2 (%)	4 (%)	6 (%)
The accuracy of the classifying		66	74	83
Error of classifying	Error classify to one class	26	20	15
	Classify to two classes	8	6	2

4 Conclusion

In this work we propose a semantic template method based on MLP to reduce the semantic gap for CBIR. The proposed method uses pixel features of a chosen region of interest, uses random weights and softmax to improve the classifier efficiency. Our experimental results show our method has a good efficiency to classify images

based on semantic. In addition, experiments shown that the accuracy of the classifier can be improved via adding samples to train the MLP classifier until attaining a relative stable state, and to achieve the stable state only need small amount of training samples. The proposed method can be applied into classification problem with two classes.

Acknowledgements The research thanks to Beijing University Youth Excellence Program (YETP1467) and BIGC Key Project (EA201411, Eb201507, KM201510015011).

References

1. Eakins, J.P., Graham, M.E.: Content-based image retrieval: a report to the JISC technology applications programme. Technical report. Institute for Image Data Research, University of Northumbria, Newcastle (1999)
2. Liu, Y., Zhang, D., Lu, G., Ma, W.: A Survey of content-based image retrieval with high-level semantics. Pattern Recogn. **40**, 262–282 (2007)
3. Li, Z., Shi, Z., Zhao, W., Li, Z., Tang, Z.: Learning semantic concepts from image database with hybrid generative/discriminative approach. Eng. Appl. Artif. Intell. **26**, 2143–2152 (2013)
4. Smith, J.R., Li, C.S.: Decoding image semantics using composite region templates. In: IEEE workshop on content-based access of image and video libraries (CBAIVL-98), pp. 9–13, IEEE, Piscataway (1998)
5. Denman, S., Halstead, M., Fookes, C., Sridharan, S.: Searching for people using semantic soft biometric descriptions. Pattern Recogn. Lett. **68**, 306–315 (2015)
6. Chang, S.F., Chen, W., Sundaram, H.: Semantic visual templates: linking visual features to semantics. In: International Conference on Image Processing (ICIP), Workshop on Content Based Video Search and Retrieval. Chicago, Illinois. 10, 531–534 (1998)
7. Zhang, Y., Liu, X., Pan, Y.: Apply semantic template to support content-based image retrieval. In: Proceeding of the SPIE Storage and Retrieval for Media Database. San Jose, CA, USA. 3972(12), 442–449 (2000)
8. Miller, G., Beckwith, R., Fellbaum, C., Gross, D., Miller, K.: Introduction to wordnet: an on-line lexical database. Int. J. Lexicography. **3**, 235–244 (1990)
9. Zhang, C., Liu, J., Tian, Q., Liang, C., Huang, Q.M.: Beyond visual features: a weak semantic image representation using exemplar classifiers for classification. Neurocomputing. **120**, 318–324 (2013)
10. Sarwar, S., Qayyum, Z.U., Majeed, S.: Ontology based image retrieval framework using qualitative semantic image descriptions. Procedia Comput. Sci. **22**, 285–294 (2013)
11. Hu, R., Collomosse, J.: A performance evaluation of gradient field HOG descriptor for sketch based image retrieval. Comput. Vis. Image Underst. **117**, 790–806 (2013)
12. Zarchi, M.S., Monadjemi, A., Jamshidi, K.: A concept-based model for image retrieval systems. Comput. Electr. Eng. **46**, 303–313 (2015)
13. Kurtz, C., Beaulieu, C.F., Napel, S., Rubin, D.L.: A hierarchical knowledge-based approach for retrieving similar medical images described with semantic annotations. J. Biomed. Inform. **49**, 227–244 (2014)
14. Ortiz, A., Gorriz, J.M., Ramirez, J., Salas-Gonzalez, D.: Improving MR brain image segmentation using self-organising mapsand entropy-gradient clustering. Inf. Sci. **262**, 117–136 (2014)
15. Wikipedia. http://en.wikipedia.org/wiki/Multilayer_perception
16. Vincent, L., Soille, P.: Watersheds in digital spaces: an efficient algorithm based on immersion simulations. IEEE Trans. Pattern Anal. Mach. Intell. **13**, 583–598 (1991)