Product Classification based on SVM and PHOG Descriptor

He Zhang[†] and Zijun Sha[††],

Graduate School of Advanced Mathematics and Human Mechanisms, University of Toyama
Gofuku 930 - 8555, Toyama - shi, Japan

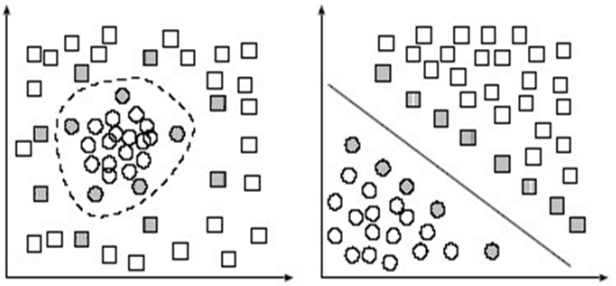
Summary
It is a great need to classify the numerous product into some categories automatically. In this paper, we adopt SVM classifier combined with PHOG (Pyramid of Histograms of Orientation Gradients) descriptors to implement product–image classification. The support vector machine maps the input vectors into a high-dimension space, in which a max margin supper hyper plane is set up to classify the samples and PHOG can flexibly represent the spatial layout of local image shape. Experimental results showed the effectiveness of the proposed algorithm.
Key words: Product classification, SVM, PHOG, Kernel, LIBSVM

1. Introduction

For the efficiency of product retrieve, it is of necessity to classify the numerous product into some categories automatically`, such as clothes, household appliances, office equipment . Each category can be classified into many sub-categories. For example, shoes, bags, T-shirts are different from the types of textile and clothing products. In this paper, SVMs (Support Vector Machines) and PHOG (Pyramid of Histograms of Orientation Gradients) descriptor are employed to implement the product classification. As supervised learning models[1], Support Vector Machines (SVMs) exhibits standout learning capability. Support vector machines can efficiently perform a non-linear classification using the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces. PHOG is an excellent image global shape descriptor, which consists of a histogram of orientation gradients over each image subregion at each resolution level[4]. The process will be described in detail as blow.

2. Support vector machine

The classification can be seen as one kind of machine learning task. In the recent last two decades, Support Vector Machine(SVMs) have become an popular supervised tool for machine learning community.
The support vector machine maps the input vectors into a high-dimension space, in which a max margin supper hyper plane is set up to classify the samples. Figure 1 illustrates the process.



(a) Original space (b) feature space(RBF mapping)

Figure 1 Separable classification with RBF kernel functions in the original space and feature

The general processes of C-SVM[1] are explained as below.

(1) Assuming the training dataset D is given as:

$$D=\{(x_i,y_i)\,|\,x_i\in R^d,\ y_i\in\{-1,1\}\},\ i=1,2...m$$

Where y_i is the label of training sample x_i, m is the total number of training samples.

(2) In the training phrase, select appropriate kernel function $k(x_i,x_j)$ and penal parameter C>0;

(3) Construct and solve the convex programming problem:

$$\min_{\alpha}\quad -\sum_{i=1}^{m}\alpha_i+\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j k(x_i,x_j),$$

$$s.t.\quad \sum_{i=1}^{m}\alpha_i y_i=0,$$

$$0\le\alpha_i\le C,\ i=1,2...m;$$

Where $k(x_i,x_j)=\left\langle\phi(x_i),\phi(x_j)\right\rangle$. The solution

Manuscript received September 5, 2013
Manuscript revised September 20, 2013

$\alpha^* = (\alpha_1^*, ...\alpha_m^*)^T$ are always sparse, i.e. there are small number of non-zero coefficients, the corresponding training samples are defined as the support vector machines.

(4) Choose the α_j^* with the value of $(0,C)$, calculate:

$$b^* = y_j - \sum_{i=1}^{m} y_i \alpha_i^* k(x_i, x_j);$$

(5) For the test sample x, the final classification decision function is

$$f(x) = \mathrm{sgn}(\sum_{i=1}^{m} \alpha_i^* y_i k(x_i, x) + b^*)$$

A kernel function k(x,y)is a function that for all $\in$X satisfies

$$k(x,y) = < \emptyset(x), \emptyset(z) >$$

Where ∅ is a mapping from X to an feature space F:

$$\emptyset: x \rightarrow \emptyset(x) \epsilon F$$

Below is some common kernel functions for image classification in practice.

(1). Linear Kernel

If $\phi(x) = x$, we get the linear kernel:

$$k\ (x,z) = x^T z$$

(2). Radial basis function kernel

Radial basis function kernels are built based on Euclidean distance, with one adjustable parameter, i.e. σ.

$$k\ (x,z) = \exp(-\|x-z\|^2) / 2\sigma^2$$

Small value of σ allow kernel classifiers to fit any labels, corresponding to the large value of d in the polynomial kernel. while large value of σ gradually reduce the kernel to a constant function, which make it impossible to learn any meaningful classifier.

(3). Histogram intersection kernel [2]

$$k_{HI}(h,h') = \sum_i \min(h_i, h_i')$$

(4). Chi-Square Kernel[3]

The Chi-Square kernel comes from the Chi-Square distribution.

$$k(x,z) = \exp(-\|x-z\|^2) / 2\sigma^2$$

Where:

$$\chi^2(x,z) = \sum_i \frac{(x_i - z_i)^2)}{x_i + z_i}$$

3. Image descriptor

In this part, an image is represented with its local shape (distribution over edge orientations within a region) and its spatial layout(tiling the image into regions at multiple resolutions). The PHOG (Pyramid of Histograms of Orientation Gradients) descriptor[4] consists of a histogram of orientation gradients over each image subregion at each resolution level.

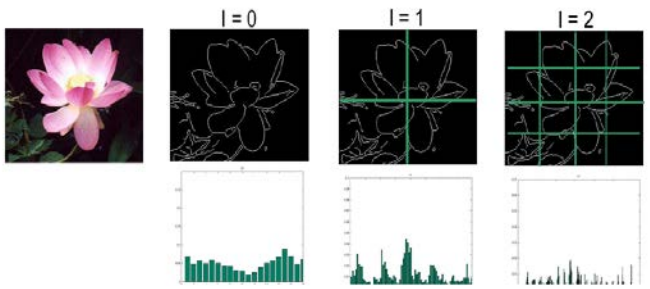
2.1 Local shape

Local shape can be described by a histogram of edge orientations (quantized into K bins) within an image subregion. The edge orientations are quantized into K bins, each of which represents the number of edges which have a certain angular range orientations. The contribution of each edge is weighted by its magnitude.

2.2 Spatial layout

The PHOG image descriptor is a concatenation of all the HOG[5] vectors, each of which is computed for each grid cell at each pyramid. Consequently, level 0 is represented by K-bin histogram, level 1 is represented by a 4K-bin histogram, etc, and the final PHOG descriptor of the entire image is a vector with dimensionality $K\sum_{l\epsilon L} 4^l$ .For example, for levels up to L= 1 and K= 30 bins it will be a 150-vector. To prevent overfitting, we limit the number of levels to L= 3 in our implementation. More to the point, the PHOG is normalized to sum to unity to ensures that images with more edges are not weighted more strongly than others. The diagrams shown in Figure 3 depict the PHOG descriptor at each level.

Figure 3 PHOG descriptor at each level(0~2)



4. Experiment and Result

Figure 4 Product images of ten categories

4.1 Experiment set

We adopt PHOG descriptors combined with SVM to implement product–image classification. All the experiments in this paper were implemented on the computer with Intel Pentium 2.00GHz CPU, 3GB RAM, Windows XP operation system and MATLAB2010a. The product images of ten categories were mainly collected from MSN shopping web site. Figure 3 illustrates some samples of the product images.

We adopted LIBSVM package [1] as SVM classifier implementation. and use the kernels defined in section 2. Multi-class classification is done with one-versus-one method. For the one-versus-one approach, classification is done by a max-wins voting strategy, in which every classifier assigns the sample to one of the two classes, then the vote for the assigned class is increased by one vote, and finally the class with the most votes determines the instance classification[6].

For the shape implementation, we use Canny edge detector to extract edge contours and use a 5*5 Sobel mask compute the orientation gradients(ranging from 0 to 180]). The HOG descriptor is discretized into K (ranging between 5 and 50) orientation bins. The vote from each contour point depends on its gradient magnitude. The pyramid level is set to 3.

Classification accuracy is the common assessment of classification performances. In order to get objective classification results, all the experiments employed cross-validation methods, in which all the samples are divided into two parts, 70% for the training and 30% for the test. All the experiments are repeated N times and the average accuracy is calculated as the final result. Average classification accuracy rate calculated using the following formula:

$$Average \quad accuracy = \frac{1}{N}\sum_{i=1}^{N}\frac{R_i}{S_i}$$

Where N indicates the number of experiments, R_i is the number of images which are correct classified in the i th experiment, S_i is the total number of image in the i th experiment.

4.2 Experiment results and analyses

Table 1 illustrated the classification accuracy variation with the number of training images per category and kernel functions. The experimental results indicate:

(1) The average accuracies increase with the training samples. However the accuracy is moving towards stabilization as the training number reaches 30.

(2) Chi-square kernel performs best, followed by histogram intersection kernel, RBF kernel, linear kernel perform worst.

Table 1 Average classification accuracy variation with training samples and kernel functions(L=3)(%)

	5	15	30	50
Linear kernel(C=1500)	74.5	80.0	81.8	82.0
RBF kernel(C=1500, g=0.07)	82.4	88.3	91.3	91.5
Histogram intersection kernel	85.2	90.6	94.1	94.3
Chi-square kernel	85.5	90.8	94.2	94.6

5. Conclusion

In this paper, we adopt SVM classifier combined with PHOG descriptors to implement product–image classification. Experimental results showed the effectiveness of the proposed algorithm. As a effective descriptor, PHOG can flexibly represent the spatial layout of local image shape. However, PHOG is not a kind of universal descriptor, and the product categories we test is far from t practical application. It is a prosperous direction to combine PHOG and other complimentary descriptor (such as appearance descriptor, colour descriptor, texture) with (multiple effective kernel[7] function based) SVM classifier to implement large-category product image classification.

References

[1] Chang, C. and C. Lin, 2011. LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2(3): 1-39.

[2] Barla A, Odone F and Verri A. Histogram intersection kernel for image classification. International Conference on Image Processing, ICIP[C]. Italy 2003,3:513-516.

[3] Jianguo, Z., et al. 2006. Local Features and Kernels for Classification of Texture and Object Categories: A Comprehensive Study. Conference on Computer Vision and Pattern Recognition Workshop, 213-238.

[4] Bosch A, Zisserman A and Munoz X. Representing shape with a spatial pyramid kernel[C]. Proceedings of the 6th ACM international conference on Image and video retrieval.NK, USA, 2007:401-408.

[5] Dalal N and Triggs B. Histograms of oriented gradients for human detection. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 2005, Vol 1:886-893.

[6] (http://en.wikipedia.org/wiki/Support_vector_machine)

[7] Siddiquie, B., S. Vitaladevuni, and L. Davis, Combining Multiple Kernels for Efficient Image Classification[C]. Workshop on Applications of Computer Vision (WACV), 2009:1-8.

He Zhang received the B.S. degree from ChangChun University, Ji Lin, China, in 2005 and the M.S. degree from University of Toyama, Toyama, Japan in 2011 Now, he is working toward the D.E. degree at University of Toyama, Toyama, Japan. His main research interests are multiple-valued logic, image recognition.

Zijun Sha received the B.S. degree from Hunan University of science and technology, Hunan, China in 2008 and the M.S. degree from University of Toyama, Toyama, Japan in 2012. Now, he is working toward the D.E. degree at University of Toyama, Toyama, Japan. His main research interests are multiple-valued logic, artificial neural networks and information security