Machine Learning for Image Classification

A

Yu-Jin Zhang
Department of Electronic Engineering, Tsinghua University, China

INTRODUCTION

Machine learning is a powerful tool for pattern classification. The goal of machine learning is to develop methods that can automatically detect patterns in data, and then to use the uncovered patterns to predict future data or other outcomes of interest. Machine learning is thus closely related to the fields of statistics and data mining, but differs slightly in terms of its emphasis and terminology (Murphy, 2012). Machine learning uses the theory of statistics in building mathematical models, because the core task is making inference from a sample. A model may be predictive to make predictions in the feature, or descriptive to gain knowledge from data, or both (Alpaydin, 2010).

Recent year, image classification has become quite a significant topic in image engineering (Zhang, 2013a) with the requirements from action recognition (Zheng et al., 2012), emotional image retrieval (Li & Zhang, 2010), scene categorization (Liu & Zhang, 2011) and behavior understanding (Zhang, 2013b), *etc*. Image classification aims at associating different images with some semantic labels to represent the image contents abstractly. To achieve this goal, various machine learning and pattern recognition techniques could be used (Bishop, 2006).

Among many potential techniques adopted in image classification, the technique that uses dictionary learned by sparse coding has achieved competitive performance recently. Sparse coding is capable of reducing the reconstruction error in transforming low-level descriptors into compact mid-level features. However, dictionary learned only by sparse coding does not have the ability to distinguish different classes, and it is not the optimum dictionary for the classification task.

In this article, some current techniques for image classification and their existing problems are reviewed, some research directions are discussed, and some progress works, especially a novel discriminant dictionary learning method combining linear discriminant analysis with sparse coding and an integrated dictionary learning technique on manifold are introduced. Their performances are compared with other techniques and quite satisfactory results have been obtained.

BACKGROUND

Currently, the typical framework adopted by the majority of existing image classification systems is discriminative model (Grauman, 2005; Lazebnik, 2006; van Gemert, 2010; Wang, 2010; Yang, 2009). Initially, bag of words model (also called codebook or codeword, *i.e.* dictionary), which treats an image as a collection of "Visual Words," is the most commonly used method in image classification (Grauman, 2005). Although it achieves some satisfactory results, bag of words model has two drawbacks. One is that the spatial information for classification is lost because of unordered "Visual Words," thus severely limiting the classification performance. The other is that each feature only corresponds to one word, so this hard decision will cause too large reconstruction error. For the former, the spatial pyramid matching method proposed has achieved remarkable success, and thus becomes an indispensable step for image classification (Lazebnik, 2006). For the latter, in order to solve the visual word ambiguity, various approaches have been suggested, such as kernel-codebook (van Gemert, 2010), locality-constrained linear coding which utilizes the linear combination of *N*-neighborhood bases to represent features (Wang, 2010). Furthermore, sparse coding based dictionary learning, which represents features by the sparse linear combination of several bases, is proposed (Yang, 2009), and achieved state-of-the-art performance.

However, the reconstruction error criterion takes effect mainly in measuring the mapping expression when transforming low-level descriptors into compact mid-level features. For the classification task, merely

DOI: 10.4018/978-1-4666-5888-2.ch021

abasing the reconstruction error is far from enough. The optimum dictionary should have the ability to distinguish different classes. Hence, discriminative information is incorporated by minimizing the loss of mutual information between features and labels during the quantization step (Lazebnik, 2008). Besides, a training approach using discriminative dictionary based on sparse coding, with respect to the sparse codes rather than the pooling results is proposed (Mairal, 2008), so it requires each code to be labeled, and ignores global image statistics.

DISCRIMINANT SPARSE CODING

To obtain a more discriminative dictionary while preserving its descriptive abilities, linear discriminant analysis (LDA) is combined with sparse coding (Liu, 2012). To be specific, the fisher linear discriminant analysis information is embedded as regularization term into the original objective function of the sparse coding method to effectively reduce the within-class scatter as well as increasing the between-class scatter.

Framework

The framework of sparse coding-based image classification include five major steps (Yang, 2009): Feature extraction, Image coding, Dictionary learning, Spatial pyramid matching (SPM) pooling, and Classifier training. Figure 1 shows the flow chart of the detailed framework.

Notations and Formulae

Let $X \in R^{D \times N}$ is the local descriptors randomly extracted from training images for learning dictionary, D is the dimension of X, and N is the number of samples in X. For classification task, let C represents the number of classes, N_0 represents the number of images extracted from each class for learning dictionary and training classifier, M represents the number of features randomly extracted from each image for learning dictionary. Then it has $N = C \times N_0 \times M$. $B \in R^{D \times K}$ is the dictionary, K is the size of the dictionary, $S \in R^{K \times N}$ is the local descriptors' codes under given dictionary B.

For the convenience of the solution ensuing discriminant sparse coding (DSC), additional non-negativity constraint to the sparse codes is imposed. Thus, sparse coding can be formulated as the minimization of the following objective function:

$$\begin{aligned} f(\mathbf{B},\mathbf{S}) &= \left\|\mathbf{X}-\mathbf{B}\mathbf{S}\right\|_F^2 + 2\alpha\left\|\mathbf{S}\right\|_1 \\ s.t.\ \mathbf{S} &\geq 0, \left\|\mathbf{B}_{\bullet i}\right\|_2 = 1,\ \forall i = 1,2,\cdots,K \end{aligned} \tag{1}$$

Here and in the following, $B_{\bullet n}$ and $B_{k\bullet}$ denote the n-th column and k-th row vectors of matrix B, respectively;

Figure 1. Framework for sparse coding-based image classification