Improving Image Classification Quality Via Dissimilarity Measure In Non-Euclidean Spaces

Lingling Chen, Songhao Zhu

School of Automatic, Nanjing University of Post and Telecommunications, Nanjing, 210046, China
zhush@njupt.edu.cn

Abstract

This paper proposes an image classification scheme by learning the dissimilarity measure in non-Euclidean spaces. Specifically, the dissimilarity representations of samples from a pseudo-Euclidean space are first constructed; then, the dissimilarity increment distribution information of each category is achieved with respect to the high-order statistics of triplet-neighbor points for each image; finally, a maximum a posteriori algorithm fused with the Gaussian Mixture Model and triplet-dissimilarity increments distribution is utilized to estimate the relevance of each image category with each input image. Experimental results conducted on a general image database demonstrate the effectiveness of the proposed scheme.

Introduction

In the past years, many techniques have been proposed to address the problem of image classification[1-8]. There are two key assumptions in these algorithmic techniques: the first assumption is that images in the database are usually distributed in the Euclidean space, and the second one is that the dissimilarity-based matching is based on the pairwise measure. However, the first assumption is not always valid and poses a challenge, since image data are usually represented as distributions over semantic words feature space and lie in complex non-Euclidean manifolds. The second assumption, pairwise measure based dissimilarity-based matching, can only be realized without any error if and only if the original set of dissimilarities are Euclidean themselves.

To deal with the above two problems, this paper presents a novel

image classification scheme. The basic idea of the proposed scheme is described as follows. To describe the dissimilarity between images represented as a bag of words, dissimilarity representation in the modified pseudo-Euclidean spaces is first provided. Then, a measure of the high-order statistics of the triplets of neighbor points within each input new image data is constructed to estimate the dissimilarity increments distribution of each category. Finally, a maximum a posteriori algorithm is utilized to calculate the likelihood of each image in the database to achieve images similar to an input new image.

Dissimilarity Representation

Dissimilarity Space

A pseudo-Euclidean space, E can be constructed based on the pseudo-inner product of two real vectors **x** and **z**:

$$\langle \mathbf{x},\ \mathbf{z} \rangle_{\mathrm{E}} = \mathbf{x}^{T} \gamma_{rs} \mathbf{z} \tag{1}$$

where

$$\gamma_{rs} = [\mathbf{I}_{r \times r} \quad \mathbf{0};\ \ \mathbf{0} \ \ \text{-}\mathbf{I}_{s \times s}] \tag{2}$$

If vector x consists of an ordered pair of two real vectors: $\mathbf{x}=(\mathbf{x}^{+}, \mathbf{x}^{-})$ where x_i^{+} and x_i^{-} are the components of x^{+} and x^{-} respectively, then the form of Equ.(2) can be rewritten as follows:

$$\langle \mathbf{x},\ \mathbf{z} \rangle_{\mathrm{E}} = \sum\nolimits_{i=1}^{r} x_i^{+} z_i^{+} - \sum\nolimits_{j=1}^{s} x_j^{-} z_j^{-} \tag{3}$$

Suppose there are r largest positive eigenvalues and s negative eigenvalues with the largest absolute value for Gram matrix G constructed with elements $G_{ij}= x_i^{T}x_j$, and all of the eigenvalues are modified by adding a constant to ensure both positive and negative eigenvalues all become positive. The added constant is denoted as $1.5|t|$, where 't' is the negative eigenvalue with the maximum absolute value.

Dissimilarity Measure

To deal with the problem of dimension curse and computation complexity, a prefiltering processing is utilized to preserve the k most similar images whilst filtering out the most dissimilar images to each input image. From the point of view of system performance,

the prefiltering processing should simultaneously satisfy the following two criterions: low computational cost and high recall rate. Here, a modified nearest neighbor rule is utilized to perform the prefiltering processing, and k is chosen as 200 [9]. After such a prefiltering processing, the feature dimension and computational complexity can be reduced significantly. With these k most relevant images, the distances from an image to its k most similar images are calculated and utilized as dissimilarity representation.

The Proposed Algorithm

Higher-Order Statistics
Let Y be a set of patterns, and (y_i, y_j, y_k) be a triplet of nearest neighbors belonging to Y, where y_j is the nearest neighbor of y_i and y_k is the nearest neighbor of y_j, different from y_i. According to the introduction in literature [10], the dissimilarity diffusion between these patterns is formulated using the following equation:

$$dif_i = \left| d\left(y_i, y_j\right) - d\left(y_j, y_k\right) \right| \tag{4}$$

Different from the pairwise distance measure between two points, the measure as shown in Equ.(4) is a distance among a triplet of points. Thus, this distance measure describes the information of higher-order statistics and dissimilarity of the data.

In literature [11], under the hypothesis that the data follow a Gaussian distribution, $p(dif_i \mid \lambda_j)$, the dissimilarity diffusion distribution of a category c_j with the dissimilarity diffusion dif_i can be formulated as follows:

$$p\left(dif_i \middle| \lambda_j\right) = \frac{\pi \xi^2}{4\lambda_j^2} \times dif_i \times \exp\left(-\frac{\pi \xi^2}{4\lambda_j^2} dif_i^2\right) + \left(\frac{4\lambda_j^2}{\pi \xi^2} - dif_i^2\right) \times \exp\left(-\frac{\pi \xi^2}{8\lambda_j^2} dif_i^2\right) \times \text{ercf}\left(\frac{\pi \xi}{\sqrt{2}\lambda_j} dif_i\right) \tag{5}$$

where λ_j is the mean value of the Gaussian distribution of the category c_j, ercf($\cdot$) is the complementary error function, and ξ=sqrt(2).

High Order Statistics Based Maximum A Posterior
Let $\{y_i, c_i, dif_i\}$ be the diffusion information of a feature vector y_i in the database, where c_i is the category label of y_i, dif_i is the set of

diffusions obtained with respect to all triplets of points containing $\mathbf{y}_i$, and both $\mathbf{y}_i$ and dif_i are conditionally independent of c_i. Under the hypothesis that each category c_i has its single statistical model with an associated diffusion parameter $\square_i$, the above-discussed dissimilarity diffusion distribution can be seen as the high-order statistics of data.

The prior probability of the category c_j, $p(c_j)$ is formulated as follows:

$$p(c_j) = \frac{|c_j|}{N} \tag{6}$$

where $|c_j|$ is the number of points of the class j, and N is the total number of the data.

The formulation of the likelihood probability of the category c_j, $p(\mathbf{y}_i, dif_i \mid c_j)$ is here described using the following equation:

$$p(y_i, dif_i \mid c_j) = p(y_i \mid c_j) p(dif_i \mid c_j) \tag{7}$$

where $p(\mathbf{y}_i \mid c_j)$ is the class-conditional density of the vector $\mathbf{y}_i$, and $p(dif_i \mid c_j)$ is the class-conditional density of the set of diffusions where $\mathbf{y}_i$ belongs.

Under the hypothesis of Gaussian distribution of the data, the formula for computing $p(\mathbf{y}_i \mid c_j)$ can be written as:

$$p(y_i \mid c_j) = \sum_{k=1}^{K} \varphi_k \, p(y_i \mid \Sigma_k, \mu_k) \tag{8}$$

where K is the number of Gaussian components of category c_j, φ_k is the weight of each Gaussian component, and $p(\mathbf{y}_i \mid \Sigma_k, \mu_k)$ is the Gaussian distribution. These three parameters φ_k, Σ_k, and μ_k can be here achieved based on the idea in [12].

The following formula expression can be used to calculate the value of p(difi|cj):

$$p(dif_i \mid c_j) = \frac{1}{L} \sum_{l=1}^{L} p(dif_i^l \mid c_j) \tag{9}$$

where L is the number of diffusions of the set dif_i, dif_i^l is the nth diffusion in the set dif_i, and $p(dif_i \mid c_j)$ equates to $p(dif_i \mid \lambda_j)$ as given in Equ.(5).

Based on the above discussed prior probability p(cj) and likelihood probability p(yi, difi | cj) of the category cj, the derivation procedure of the used maximum a posterior of the category cj, p(cj |yi, difi) can be described as below:

$$p\left(c_j \middle| y_i, dif_i\right) = \frac{p\left(y_i, dif_i \middle| c_j\right) p\left(c_j\right)}{p\left(y_i, dif_i\right)} = p\left(y_i, dif_i \middle| c_j\right) p\left(c_j\right) = p\left(y_i \middle| c_j\right) p\left(dif_i \middle| c_j\right) p\left(c_j\right)$$

(10)

where $p(\mathbf{y}_i, dif_i)$ is a constant computed over the data.

Experiments

To evaluate the performance of the proposed image classification scheme, a general purpose image database containing 50,000 Corel dataset is used as the ground truth database. This dataset is manually subdivided into 500 semantic categories, and each image category consists of 100 images of the same topic. Each image in the whole database is used as an input new image, and the average of the classification results over all the 50,000 images is used to evaluate the system performance.

Feature selection has been an active research topic for many decades due to the fact that feature selection might have a great impact on the final retrieval results. In the current implementation, the extracted feature vector is a 428-dimensional vector, including 225-dimensional block-wise color moment, 128-dimensional wavelet texture and 75-dimensional edge distribution histogram.

Table 1: Performance comparison of average precision.

Method	CD	SPCA	The proposed method
Average precision	0.54	0.59	0.66

Precision of all images is averaged and adopted as the final performance evaluation measure. The average precision of all these three methods are illustrated in Table 1, from which it can be clearly seen that the proposed method outperforms the covariance discriminative based method (CD for short) [3] and separable principal components analysis based method (SPCA for short) [8]. This improvement can be analyzed from the following two aspects. One is that the pair-wise measurement between two points cannot

sufficiently reflect the contribution from one image to another one in the color difference histogram based method and the manifold-ranking based method, while our method implements the 'anisotropic' contribution by taking into account the high-order dissimilarity among a triplet of points. Another is that exploring the properties of images in non-Euclidean space also helps to enhance the accuracy of image retrieval.

Conclusions

In this paper, we propose a novel image classification scheme by exploring the dissimilarity measure in the non-Euclidean space. The proposed method can be interpreted as a classifier that forces a category to have the same increment structure, even though the mean and covariance of Gaussian components within a category are distinct. Experimental evaluation on the Corel dataset shows that the most competitive performance of the proposed method compared to the covariance discriminative based method and separable principal components analysis based method.

Acknowledgments

References

[1]   Z. Lu, L. Wang, and J. Wen. Direct Semantic Analysis for Social Image Classification. Proceedings of AAAI Conference on Artificial Intelligence, 2014: 1258-1264.

[2]   Y. Zhang, Z., and L. Davis. Learning Structured Low-Rank Representations for Image Classification. IEEE Conference on Computer Vision and Pattern Recognition, 2013: 676-683.

[3]   R. Wang, H. Guo, L. Davis, and Q. Dai. Covariance discriminative learning: A natural and efficient approach to image set classification. Proceedings of the

IEEE Conference on Computer Vision and Pattern Recognition, 2012: 2496-2503.

[4]   B. Fernando, É. Fromont, and T. Tuytelaars. Effective Use of Frequent Item set Mining for Image Classification. Proceedings of the European Conference on Computer Vision, 2012: 214-227.

[5]   S. Zhong, Y. Liu, and Y. Liu. Bilinear deep learning for image classification. Proceedings of the ACM Conference on Multimedia 2011: 343-352.

[6]   S. Khan and C. Barnes. Using residual vector quantization for image content classification. Proceedings of the IEEE Conference on Acoustics, Speech, and Signal Processing, 2011: 1041-1044.

[7]   D. Parikh. Recognizing jumbled images: The role of local and global information in image classification. Proceedings of the IEEE Conference on Computer Vision, 2011: 519-526.

[8]   Y. Xi and P. Ramadge. Separable PCA for image classification. Proceedings of the IEEE Conference on Acoustics, Speech, and Signal Processing, 2009: 1805-1808.

[9]   S. Zhu and Y. Liu. Improving image retrieval effectiveness via Random Walk with Restart. Optical Engineering, 2008, 47(11): 1-8.

[10] A. Fred and J. Leitao. A new cluster isolation criterion based on dissimilarity increments. Pattern Analysis and Machine Intelligence, 2003, 25(8): 944-958.

[11] H. Aidos and A. Fred. Statistical modeling of dissimilarity increments for d-dimensional data: Application in partitional clustering. Pattern Recognition, 2012, 45(9): 3061-3071.

[12] M. Figueiredo and A. Jain. Unsupervised learning of finite mixture models. Pattern Analysis and Machine Intelligence, 2002, 24(3): 381–396.