Nearest-Neighbor Classification Using Unlabeled Data for Real World Image Application

Shuhui Wang[1] Qingming Huang[1, 2] Shuqiang Jiang[1] Qi Tian[3]

[1]Key Lab of Intell. Info. Process., Inst. of Comput. Tech., CAS, Beijing, 100190, China
{shwang,sqjiang}@jdl.ac.cn

[2]Graduate University, Chinese Academy of Sciences Beijing, 100049, China
qmhuang@jdl.ac.cn

[3]Dept. of Computer Science, Univ. of Texas at San Antonio, TX78249, U.S.A.
qitian@cs.utsa.edu

ABSTRACT

Currently, Nearest-Neighbor approaches (NN) have been widely applied to real world image data mining. These approaches have the following three disadvantages: (i) the performance is inferior on small datasets; (ii) the performance of approximated nearest neighbor search will degrade for data with high dimensions; (iii) they are heavily dependent on the chosen feature and distance measure. To overcome these intrinsic weaknesses, we propose a novel Nearest-Neighbor method, which improves the original NN approaches from three aspects. Firstly, we propose a novel neighborhood similarity measure, where the similarity between test images and labeled images in the database is calculated jointly by the original image-to-image similarity and the average similarity of their neighboring unlabeled data. Secondly, we adopt the kernelized locality sensitive hashing to effectively conduct the nearest neighbor search for high dimensional data. Finally, to enhance the robustness of the method on different genres of images, we propose to fuse the discrimination power of different features by considering all the retrieved nearest neighbors via hashing systems using different features/kernels. Experimental result shows the advantage over traditional Nearest-Neighbor methods using the labeled data only. Even when the ratio of labeled data is very small, our method could also achieve remarkable results, thanks to the help of unlabeled data and multiple features.

Categories and Subject Descriptors

H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing; I.2.6 [**Artificial Intelligence**]: Learning; I.4.8 [**Image Processing and Computer Vision**]: Scene Analysis

General Terms

Algorithms, Experimentation.

Keywords

image classification, Nearest-Neighbor methods, neighborhood similarity measure, kernelized locality sensitive hashing.

MM'10, October 25–29, 2010, Firenze, Italy.

1. INTRODUCTION

Automatic image classification has drawn considerable attention during the past few decades. The significant endeavors made in the research community have resulted in many novel and effective approaches. For typical dataset such as Caltech-101, the classification accuracies of state-of-the-art methods have been improved from 20% to almost 90% during the past few years [2,9].

Among the existing approaches, a well studied paradigm for image classification is learning based approaches, which requires an intensive training step for classifier models (For example, SVM [4], Boosting [12] and Distance Metric Learning [11]). Another is totally data driven, which requires no training step on model parameters. The most common data driven approach is Nearest-Neighbor Classification (NN), which classifies an image by the class of its most similar images in the database.

Compared with the learning based approaches, data driven approaches have several advantages: (i) no training and learning step is required; (ii) no over-fitting issues should be considered; (iii) they can naturally handle thousands of image classes and millions of images. Furthermore, when the number of labeled images in the database is large enough, the error rate of Nearest-Neighbor approaches converges to the optimal Bayes error rate [1], which provides a theoretical foundation for NN methods.

However, Nearest-Neighbor approaches usually achieve inferior performance than learning based approaches in many scenarios. To bridge the performance gap between NN approaches and learning based approaches, a lot of studies have been conducted from different aspects. Firstly, from distance metric aspect [2, 5, 8], Boiman *et al*. [2] claimed that the previously used image-to-image distance will lead to the degradation of NN approaches and proposed an image-to-class distance measure. Friedman [5] proposed a new local similarity measure based on kernel methods and recursive partitioning techniques. Another similarity measure was proposed in [8] to incorporate the invariance of translations and scaling. From database size aspect [3, 8], Torralba *et al*. [8] found that with extremely large tiny image database, *i.e.*, 80 millions, NN could work significantly well for image annotation, although the tags of the images are very noisy. Deng *et al*. [3] constructed a large scale database with human labeled ground truth. NN approach is more likely to achieve good performance on this large scale database. Next, from image analysis aspect, Boiman *et al*. [2] showed that feature quantization will reduce the discrimination power of local features, which will lead to inferior performance of NN methods. Finally, from learning aspect, Zhang *et al*. [13] combined the efficiency of NN and the effectiveness of SVM.

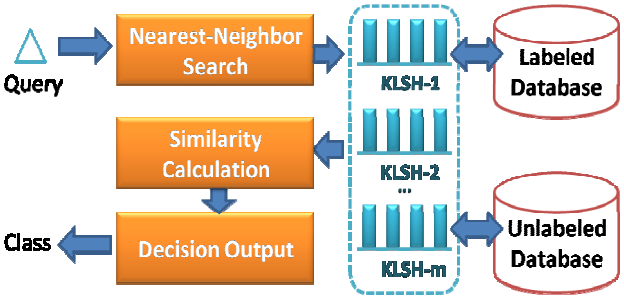

Figure 1: Flow chart of our Nearest-Neighbor classification system.

Local kernel alignment of nearest neighbors was proposed by Lin *et al*. [7] to combine the discrimination power of multiple kernels. State-of-the-art performance was achieved by [13] and [7] with significant reduction of computation cost compared with the traditional global learning approaches.

In our study, we claim that the performance of NN will degrade under three circumstances. Firstly, NN approaches are weak on small size database, because the original image-to-image distance is sensitive to noise. When the size of labeled data is small, noise and small image variation such as translation and scaling would easily influence the local neighborhood structure. One feasible way to alleviate this problem is to use the unlabeled data, as we can obtain huge amount of unlabeled data without much effort. Since the unlabeled data could be used to describe the distribution in the feature space, the neighborhood similarity between test images and labeled images in the database is calculated with both the original similarity and the average similarity of their neighboring unlabeled data. Compared with the original image-to-image distance, the neighborhood similarity encodes the local density information, which provides more stable and noise-free description of the similarity among images.

Secondly, NN approaches degrade on high dimensional space, for the data distribution is very sparse. The "kernel trick" used in many previous approaches is a good way to overcome this problem. Therefore, we adopt a kernelized locality sensitive hashing (KLSH) [6] to construct an approximate nearest neighbor search system, where the hash function is formed by using a small subset of the whole database. The complexity of KLSH is almost the same as the original LSH, which implies that KLSH inherits the efficiency of LSH.

Finally, NN approaches are very sensitive to the feature and the distance measure. Different feature and distance measure provide different description for different genres of images. For example, color feature is good at describing plant images while it is not suitable for object images. The limitation of using one feature only constrains NN approaches to apply on general image classification tasks. To overcome this issue, we use a set of features instead of one, and all the neighbors returned by the system using different features contribute to the final decision output. This is similar to the ensemble method in spirit, where outputs from distinguished classifiers are combined.

In general, we propose a new Nearest-Neighbor image classification method. The key contributions are presented in three aspects: (1) we propose a neighborhood similarity measure for Nearest-Neighbor methods that calculates the image similarity jointly by the original image-to-image similarity and average similarity of their neighboring unlabeled data; (2) we use nearest neighbors retrieved by a set of KLSH systems using different features/kernels to form the final decision output of an unknown sample; (3) we construct a practical system that is able to perform real world image classification.

The rest of this paper is organized as follows: Section 2 describes our method. In Section 3 we conduct extensive experiment to evaluate our method. Section 4 gives some conclusions and discusses the future work.

2. METHOD

2.1 Overview

We propose a Nearest-Neighbor method as described in Figure 1. Firstly, a set of kernelized LSH systems are constructed both on labeled and unlabeled data using different kernels as introduced in Section 3.2. A query image is fed into the system and all the nearest neighbors found by different KLSH are returned. Then the similarity between query image and the returned nearest labeled data is calculated using the method introduced in Section 3.1. The final decision output is determined by the scheme introduced in Section 3.3. We describe each part in details.

2.2 Neighborhood Similarity Measure

Generally, a better similarity measure should be able to encode the local density and manifold information. The motivation of the neighborhood similarity measure is to use the unlabeled data to approximate the true data distribution in the unknown kernel space. Our method is based on the following assumptions:

(a) Data points with similar local density are likely to be more similar than data points with different local density.

(b) The similarity among data points on dense manifolds tend to be larger than data points on sparse manifolds.

These two assumptions are consistent with the manifold assumption [14] and cluster assumption [10, 14] used in many semi-supervised literature. Specifically, given a fixed feature representation $\Phi_{ori}(x)$ which corresponds to a certain kernel, we represent a sample x by the linear combination of its own implicit representation with respect to a certain kernel K and its average of neighboring representation as:

$$\Phi_{Nbd}(x) = \alpha\Phi_{ori}(x) + \frac{(1-\alpha)}{|Nbd(x)|}\sum_{x'\in Nbd(x)}\Phi_{ori}(x') \quad (1)$$

where $Nbd(*)$ denotes the neighborhood unlabeled sample set.

Based on the expression in Equation (1), the similarity of query x and the labeled examples y are calculated by the weighted averaging of the original kernel value $K_{ori}(x,y)$ and the neighborhood kernel value as:

$$K_{Nbd}(x,y) = \alpha K_{ori}(x,y) + (1-\alpha)\frac{\sum K_{ori}(x',y')}{|Nbd(x)||Nbd(y)|} \quad (2)$$
$$x'\in Nbd(x),\ y'\in Nbd(y),\ x',y'\in U$$

where $K_{ori}(x,y) = \langle\Phi_{ori}(x),\Phi_{ori}(y)\rangle$, U denotes the unlabeled data. α is the weight parameter. We set $\alpha = 0.5$ empirically. The modified feature representation is similar to the cluster center in the convex hull formed by the linear combination of feature representation from the data itself and the neighborhood unlabeled samples. Under this representation, different clusters rather than different samples become better separated from each other. Compared with the image-to-image similarity, the neighborhood similarity measure provides better discrimination power for a set of samples instead of only one sample. Therefore, it is more robust to noise and small image variations.

2.3 Kernelized Locality Sensitive Hashing

The intuitive of kernelized LSH is to perform LSH in an unknown high dimensional kernel space. With similar theorem used in Kernel PCA [6], the hashing function is written as:

$$h(\phi(x)) = sign\left(\sum_{i=1}^{P} \mathbf{w}(i) K(x_i, x)\right) \quad (3)$$

where $\phi(x)$ denotes the unknown representation in high dimensional space. The weight vector $\mathbf{w}$ is calculated as described in Equation (4):

$$\mathbf{w} = K^{-1/2}\left(\frac{1}{T}\mathbf{e}_s - \frac{1}{P}\mathbf{e}\right), K = USU^T, K^{-1/2} = US^{-1/2}U^T \quad (4)$$

where K is the kernel matrix of the randomly chosen P items of the whole database, and usually P is very small (we set P=300 for our experiments unless special statement is given) compared with the whole data size. $\mathbf{e}$ is a vector with P ones and $\mathbf{e}_s$ is an indicator vector for a subset S of the P items where:

$$\mathbf{e}_s(i) = \begin{cases} 1, & if\ i \in S \\ 0, & else \end{cases}, i = 1,..P \quad (5)$$

The size of S is T. In this paper we set T = 30. A set of hash functions could be obtained by randomly choosing the subset S. In this paper, we use 5 kinds of kernels to generate five KLSH systems as described in Table 1, and we generate 128 hash functions for each KLSH.

2.4 Decision Output

Based on the KLSH system, we firstly identify those nearest neighbor samples from the database. For each KLSH system, the nearest neighbor samples are those examples whose hash codes are in the most similar buckets with the query samples. Since we must guarantee that both the nearest labeled and unlabeled data can be chosen, we respectively retrieve the labeled and unlabeled data in the nearest buckets, considering the distribution difference of the hash code representation. The fact that the chosen labeled and unlabeled data are neighboring samples to each other can be guaranteed by triangle inequality.

Suppose we have obtained $(N_L^1,...,N_L^M)$ nearest labeled samples and $(N_U^1,...,N_U^M)$ nearest unlabeled data by MKLSH system, then the set of the nearest labeled samples and unlabeled samples are:

$$N_L = unique(N_L^1 \cup ... \cup N_L^M). \quad N_U = unique(N_U^1 \cup ... \cup N_U^M) \quad (6)$$

If all the returned labeled samples come from the same class, the decision output is directly assigned with this class index. Otherwise, it is calculated as follows:

$$C = \arg\max_Q \frac{1}{|N_{L,Q}|} \sum_{j=1}^{|N_{L,Q}|} \sum_{m=1}^{M} w^m K_{Nbd}^m(x_q, x_j^Q) \quad (7)$$

$$where: \quad x_j^Q \in N_{L,Q}, \quad w^1 + ... + w^m = 1$$

where K_{Nbd}^m denotes m^{th} neighborhood similarity measure. $N_{L,Q}$ denotes the set of Q^{th} class samples in N_L. w^m denotes the weight of m^{th} feature/kernel determined by experiment on the validation set, which is a subset of unlabeled data with human labeled ground truth. The whole procedure is described in Algorithm 1.

Algorithm 1: The proposed Nearest-Neighbor procedure

Initial setting:

A labeled dataset with N_c classes of images; an unlabeled dataset. M features / kernels.

Procedure:

1. Build M kernelized LSH systems on both labeled data and unlabeled data for each features / kernels.
2. **For** each test query sample q:
   (a) Obtain $(N_L^1,...,N_L^M)$ nearest labeled samples and $(N_U^1,...,N_U^M)$ nearest unlabeled samples from the database retrieved by M KLSH system.
   (b) **If** all the returned labeled samples come from one class
   return *Class index* for q
   else
   Calculate the similarity between q and the labeled data using Equation (2).
   End
   (c) **Return** the final decision output by Equation (7).

end For

3. EXPERIMENTS

The detail of experiment setup is shown in Table 1. We use the Caltech-256 as the labeled dataset, and download more than 250K unlabeled images from the web as the unlabeled dataset. We run all the experiments on a desktop computer, which does not need a lot of computational resources. Five features or kernels describing different property of images are used, including texture [15], color, Bag-of-Words [15] and global feature [8].

We randomly select 5, 10, 15, 20, 25 and 30 labeled samples from each class of Caltech-256 as the labeled data, and randomly select another 25 samples for each class as the test data. We repeatedly choose the training and testing samples for ten times, and all the results reported in this paper are average and standard deviation of results on these ten different and independent data separations. All the web data is used as the unlabeled data.

When the unlabeled data is not used, our method is equal to the traditional Nearest-Neighbor approach. We treat this as the baseline method. We implement the baseline method in 3 versions, where NN-1, NN-3, NN-5 denotes the baseline methods using 1, 3 and 5 kernels, respectively. For NN-1, the kernel (1) in Table 1 is used, and for NN-3, the kernels (1)-(3) are used. We denote our approach by UNN-1, UNN-3, and UNN-5, using the same 1, 3 and 5 kernels respectively as the baseline methods. For all the methods, we use a 64-bit hash function for each kernel, and set the number of nearest neighbors for neighborhood similarity measure as 10 for both performance and efficiency consideration.

Table 1: Experiment setup description

Data sources
● Caltech-256: image dataset with 256 objects. Size: 30607. ● Web image data: collected from flickr.com with 256 object categories. The dataset is downloaded using the same class names as Caltech-256. Size: 251921.

Environment
OS: Windows XP; Computer: Lenovo ThinkCenter M6000t desktop; CPU: Intel(R) Core2 Duo E7500 @3.00GHz; Memory: 4.0G RAM; Programming platform: Matlab R2009.

Features and similarity measures used in this paper
(1) 3 level PHOG-180 with chi2 + Gaussian kernel. (2) 8 × 8 Color Moment with RBF kernels. (3) Gist descriptor with chi2 + Gaussian kernel. (4) 3 level spatial pyramid kernel with dense SIFT feature (visual vocabulary size: 500) (5) 3 level PHOG-360 with chi2 + Gaussian kernel.

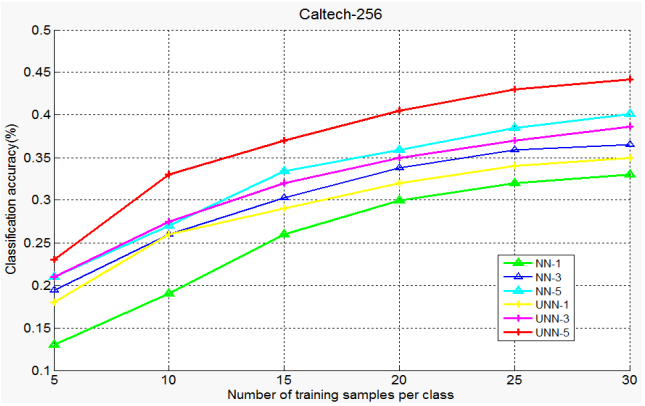

Figure 2: Classification accuracy curves.

For all the methods, the nearest neighbors of either labeled or unlabeled data are retrieved by the samples in the top 2 nearest hash buckets. Some experiment results and the performance curves are demonstrated in Table 2 and Figure 2, respectively.

In Figure 2, compared with NN-1, NN-3, and NN-5, the average classification accuracy is improved when the unlabeled data is incorporated, as can be seen from the performance curves of UNN-1, UNN-3, and UNN-5. From Table 2 we can see that, our methods, UNN-1, UNN-3 and UNN-5 outperform NN-1, NN-3 and NN-5 respectively, both on the average and standard deviation. Our method (UNN-5) is also comparable to and even outperforms the approach in [2].

Table 2: Accuracy with 30 training images per class

Methods	Performance
Boiman *et al*. [2]	≈ 42%
NN-1	33.0 ± 2.1%
NN-3	36.5 ± 1.75%
NN-5	40.1 ± 1.4%
UNN-1	35.0 ± 1.1%
UNN-3	38.6 ± 0.76%
UNN-5	**44.4 ± 0.42%**

Next, we conduct experiment to show how the setting of the number of the nearest neighbors affects the performance and the computational time required for testing. The method we use for experiment is UNN-5. The results are shown in Table 3. We can see that the best accuracy is achieved when the neighboring sample number is about 10. However, the average test time for each test sample increases significantly when the size of neighbors increases. If the neighborhood size is very large, say 100, the computational cost will be much larger. Meanwhile, the neighborhood similarity will be over-smoothed, and it will fail to describe the local data distribution. Therefore, for both efficiency and effectiveness, we set the neighboring size as 10.

Table 3: Accuracy and time for different neighborhood size

#Neighbors	1	3	5	10	15	20
Accuracy (%)	40.3	42.5	44.2	44.4	44.3	42.3
Avg time (s)	7	10	14	29	48	75

4. CONCLUSION

We propose a new Nearest-Neighbor classification method in this paper. Our contribution includes three aspects: (1) we propose a novel neighborhood similarity measure which encodes the local density information by using the unlabeled data; (2) we effectively combine the discrimination power of different features/kernels; (3) we have designed an image classification system based on nearest neighbor method and multiple kernelized LSH. Our method provides promising classification result comparing with the traditional Nearest-Neighbor approaches. Future study will be focusing on developing more robust local learning model based on approximated nearest neighbor search system.

5. ACKNOWLEDGEMENT

This work was supported in part by National Natural Science Foundation of China: 60773136 and 60833006, in part by National Basic Research Program of China (973 Program): 2009CB320906, and in part by Beijing Natural Science Foundation: 4092042. This work was also supported in part by Akiira Media Systems, Inc. for Dr. Qi Tian.

6. REFERENCES

[1] C. M. Bishop. Pattern Recognition and Machine Learning. Springer-Verlag New York, Inc., Secaucus, NJ, 2006.

[2] O. Boiman, E. Shechtman, and M. Irani. In defense of nearest-neighbor based image classification. In CVPR, 2008.

[3] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, L Fei-Fei. ImageNet: a large-scale hierarchical image database. In CVPR, 2009.

[4] J. Fan, Y. Gao, and H. Luo. Multi-level annotation of natural scenes using dominant image components and semantic concepts. In ACM Multimedia, 2004.

[5] J. H. Friedman. Flexible metric nearest neighbor classification. Technical report, 1994.

[6] B. Kulis, K. Grauman. Kernelized Locality Sensitive Hashing for Scalable Image Search. In ICCV, 2009.

[7] Y. Lin, T. Liu, C. Fuh. Local ensemble kernel learning for object category recognition. In CVPR, 2007

[8] A. Torralba, R. Fergus, W. T. Freeman. 80 Million Tiny Images: A Large Data Set for Nonparametric Object and Scene Recognition. In PAMI, 30 (11): 1958 - 1970, 2008.

[9] M. Varma and D. Ray. Learning the Discriminative Power-invariance Trade-off. In ICCV, 2007

[10] J. Weston, C. Leslie, E. Le, D. Zhou, A. Alisseeff, and W. S. Noble. Semi-supervised protein classification using cluster kernels. In Bioinformatics, 21 (15): 3241 – 3247, 2005.

[11] L. Wu, S. C. H. Hoi, R. Jin, J. Zhu, and N. Yu. Distance Metric Learning from Uncertain Side Information with Application to Automated Photo Tagging. In ACM Multimedia, 2009.

[12] R. Yan, J. Tesic, J. R. Smith. Model-shared Subspace Boosting for Multi-label Classification. In ACM SIGKDD, 2007.

[13] H. Zhang, A. C. Berg, M. Maire and J. Malik. SVM-KNN: Discriminative Nearest Neighbor Classification for Visual Category Recognition. In CVPR, 2006.

[14] X. Zhu. Semi-supervised Learning Literature Survey. Technical Report 1530, University of Wisconsin - Madison, 2006.

[15] A. Bosch, A. Zisserman, and X. Munoz. Representing shape with a spatial pyramid kernel. In CIVR, 2007.