Evaluation of dimensionality reduction methods for image auto-annotation

Hideki Nakayama¹
nakayama@isi.imi.i.u-tokyo.ac.jp
Tatsuya Harada^{1,2}
harada@isi.imi.i.u-tokyo.ac.jp
Yasuo Kuniyoshi¹
kuniyosh@isi.imi.i.u-tokyo.ac.jp

¹ Graduate School of Information Science and Technology The University of Tokyo Tokyo, Japan

² PRESTO, JST

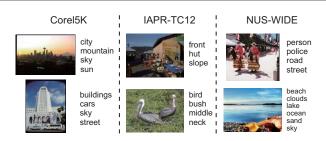


Figure 1: Examples of annotations from benchmark datasets.

Image auto-annotation [3, 4] is one of the most important challenges in computer vision. The goal of this task is to predict multiple keywords for generic images captured in real environments (Fig. 1). Recent studies have shown that a simple nearest neighbor-like approach is quite successful. However, while this approach is successful from the viewpoint of annotation accuracy, its computational costs, in terms of both complexity and memory use, tend to be high due to the size of the training datasets.

In this study, we focus on a fundamental problem: given a visual feature representation, how far can we go using simple linear dimensionality reduction methods to compress the semantic distance of images? Using the sample labels for supervision, new Euclidean distance metrics are embedded in a small-dimensional subspace. We perform extensive comparisons of several methods using various datasets and visual features, to consider under which circumstances these methods are effective. We also show how they can be applied effectively to image annotation.

Compared dimensionality reduction methods

Suppose we have a p-dimensional image feature \mathbf{x} , and a q-dimensional label feature \mathbf{y} . Our objective is to obtain a new d-dimensional small vector (d << p), whose distance metric could be the L2 distance.

- (1) **PCA, PCAW** (whitened principal components)
- (2) **PLS**, **nPLS**: Partial Least Squares finds linear transformations $\mathbf{s} = V_x^T(\mathbf{x} \bar{\mathbf{x}})$ and $\mathbf{t} = V_y^T(\mathbf{y} \bar{\mathbf{y}})$ that maximize the covariance between new values \mathbf{s} and \mathbf{t} . We also test PLS after normalizing the variances of original feature elements (nPLS).
- (3) **CCA**: Canonical Correlation Analysis is intimately related to PLS. Whereas PLS finds the projections that maximize the covariance between the two new values, CCA finds those that maximize the correlation.
- (4) **CCD**: Previously, we proposed the Canonical Contextual Distance [5], which is based on the probabilistic CCA (PCCA) [1]. Using the probabilistic structure, we can obtain a hidden latent space, in which two canonical spaces (image side and label side) are integrated. The distance between a query and a training sample can be computed by means of KL-divergence in the latent space. Moreover, we can further utilize training labels for actual distance computation (Fig. 2).

Datasets and features

We used three challenging datasets: **Corel5K** [3], **IAPR-TC12** [4], and **NUS-WIDE** [2]. For Corel5K and IAPR-TC12, we tested four visual features: 1) SIFT bag-of-words (BoW), 2) Hue BoW, 3) GIST, and 4) HSV color histogram. For NUS-WIDE, we tested: 1) edge histogram, 2) color correlogram, 3) grid color moment, and 4) SIFT BoW. To provide baselines, we computed various base distances for each feature (e.g., χ^2 distance etc.). To cope with non-linearity, we also tested embedding original metrics via kernel PCA, using a small number of training samples for kernelization. Regarding label features, we used a binary vector indicating the presence of each word.

Experiments

To annotate query images, we simply used the k nearest neighbor method with a linear search. For evaluation, we followed the methodology of previous works [3]. First, we computed word-specific recall and precision.

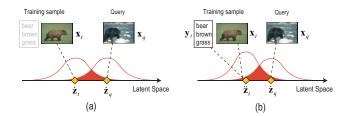


Figure 2: Illustration of CCD. Estimation of distance between a query and training sample: (a) from the *x*-view only (CCD1); and (b) considering both the *x* and *y*-views (CCD2).

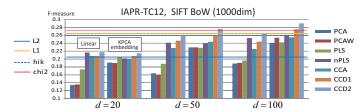


Figure 3: Example of annotation performance comparison.

These metrics were averaged over all the testing words to obtain Mean-Recall (MR) and Mean-Precision (MP). Because of the trade-off of these two scores, we evaluated the total performance using the F-measure.

Figure 3 shows a fraction of our experimental results. In many cases, nPLS and CCD show superior performance. However, it is sometimes difficult for simple linear methods to compete with the original domain-specific metrics, in terms of accuracy. In such cases, KPCA embedding works effectively and substantially improves performance, although a small fraction of training samples was used for kernelization (*n*=300). Another observation is that the performance of the CCA family is often ordered CCD2 > CCD1 > CCA, which indicates the importance of considering the *y*-view for distance computation. While simple CCA is not always effective and is sometimes outperformed by PCA or PCAW, we observe that CCD2 consistently outperforms these methods.

Conclusion

We investigated and compared several linear dimensionality reduction methods for non-parametric image annotation. Obtaining powerful small codes in a scalable manner is a crucial issue in implementing large-scale image annotation systems. Linear methods enable training in linear time and are suitable for this purpose. Using the semantic information provided by multiple labels, we can obtain a small-dimensional latent subspace reflecting the semantic distance of instances. Moreover, the superior performance of 2-view CCD indicates the importance of using label information explicitly in actual distance computation.

- [1] F. R. Bach and M. I. Jordan. A probabilistic interpretation of canonical correlation analysis. Technical Report 688, Department of Statistics, University of California, Berkeley, 2005.
- [2] T.-S. Chua, J. Tang, R. Hong, H. Li, Z. Luo, and Y.-T. Zheng. NUS-WIDE: a real-world web image database from National University of Singapore. In *Proc. ACM CIVR*, 2009.
- [3] P. Duygulu, K. Barnard, N. de Freitas, and D. Forsyth. Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. In *Proc. ECCV*, 2002.
- [4] A. Makadia, V. Pavlovic, and S. Kumar. A new baseline for image annotation. In *Proc. ECCV*, 2008.
- [5] H. Nakayama, T. Harada, and Y. Kuniyoshi. Canonical contextual distance for large-scale image annotation and retrieval. In *Proc. ACM* workshop LS-MMRM, 2009.