Supervised Hashing with Jointly Learning Embedding and Quantization

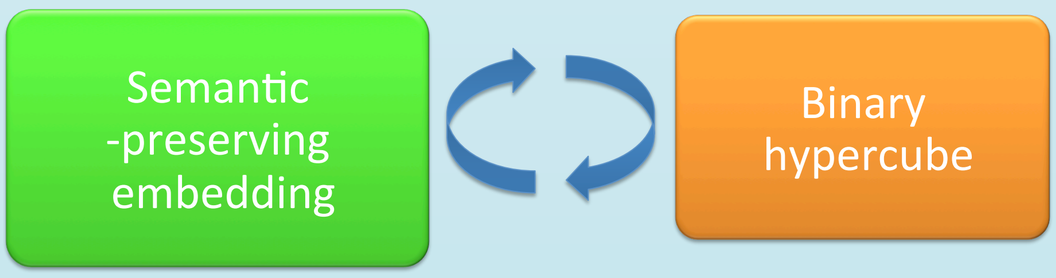
Hao Zhu[1], Xiang Xiang[2], Feng Wang[2], Trac D.Tran[3]

[1]JD Finance, Beijing, China
[2]Dept. of Computer Science, Johns Hopkins University
[3]Dept. of ECE, Johns Hopkins University

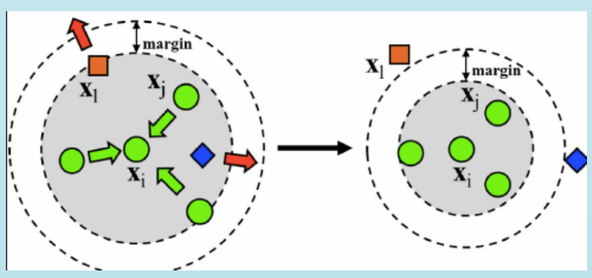
Introduction

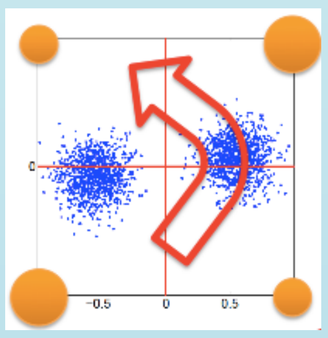
Compared with unsupervised hashing, supervised hashing commonly illustrates better accuracy in many real applications by leveraging semantic (label) information. However, it is tough to solve the supervised hashing problem directly because it is essentially a discrete optimization problem. Some other works try to solve the discrete optimization problem directly using binary quadratic programming, but they are typically too complicated and time-consuming while some supervised hashing methods have to solve a relaxed continuous optimization problem by dropping the discrete constraints. However, these methods typically suffer from poor performance due to the errors caused by the relaxation manner. In this paper based on the general two-step framework: learning binary embedded codes and learning hash functions, we propose a new method to solve the problem introduced by relaxing the cost function. Inspired by the property of rotation invariance of learning embedding features, our method tries to jointly learn similarity-preserving representation and rotation transformation for better quantization alternatively. In experiments, our method shows significant improvement. Compared with the methods based on discrete optimization our methods obtains the competitive performance and even achieves the state-of-the-art performance in some image retrieval applications.

Our Solution



metric learning + quantization loss





Using Iterative Quantization is a direct way for unsupervised hashing methods but only a few works discuss its application in supervised hashing method (e.g. CCA-ITQ). So far as we know, no works discuss the application of using ITQ in hashing methods based on embedding approaches. In the empirical analysis, we found the ITQ cannot improve the performance of two-stepped methods as simple as the CCA-ITQ. Thus based on the property of rotation in- variance of embedding techniques, we propose a novel way to jointly learn representation and quantization based on the two-stepped framework.

The iterative quantization is able to control the space distribution of embedded features without changing distances between features. Thus we can find a good embedding with minimizing quantization loss.

$$L = -logP(B|S) \qquad \theta_{ij} = \frac{1}{2}u_i u_j^T$$
$$= \sum \log(1 + e^{-s_{ij}\theta_{ij}}) + \sum_{i=1}^{n} \|sign(u_i) - Ru_i\|_2^2$$

Therefore, we can minimize U and R in an alternating procedure. Please note this procedure need many iterations. Fixing U , optimize R. This reduces to the Orthogonal Procrustes Problem (OPP):

$$\sum_{i=1}^{n} \|sign(u_i) - Ru_i\|_2^2 \;\; s.t. R^T R = I_c$$

where an optimum of OPP can be obtained as same as the method in ITQ, and then U is updated as RU. Fixing R, optimize U. Since directly optimizing the whole B and U is impossible when the dataset is too large because it would be very time-consuming, so a simple strategy more feasible is to optimize each row of U at a time with its other rows fixed.

Experiments

We compare our model with several baselines on three widely used benchmark datasets: CIFAR-10, CIFAR-100, and NUS-WIDE. The CIFAR-10 dataset consists of 60,000 32×32 color images which are categorized into ten classes (6000 images per class). It is a single-label dataset in which each image belongs to one of the ten categories. The CIFAR-100 dataset is just like the CIFAR-10, except it has 100 classes containing 600 images each. The NUS- WIDE dataset has nearly 270,000 images collected from the Internet. It is a multi-label dataset in which each image is annotated with one or multiple class labels from 81 classes. Following, There also exist some images without any label, which are not suitable for our evaluation. After removing those images without any label, we get 209,347 images for our experiment. We consider two images to be semantically similar if they share at least one common label. Otherwise, they are semantically dissimilar.

To be independent of deep feature's representation power, this paper has not touched upon deep features. Instead, conventional hand-crafted features such as GIST are used for all hashing methods during evaluation. We represent each image in CIFAR-10 and CIFAR-100 by a 512-dimensional GIST vector. We represent each image in NUS-WIDE by an 1134 dimensional low-level feature vector, including 64-D color histogram, 144 -D color correlogram, 73-D edge direction histogram, 128-D wavelet texture, 225-D block-wise color moments and 500-D bag of words based on SIFT descriptions.

Table 1. Results of MAP on CIFAR 10

Method	8-bits	16-bits	32-bits	64-bits
Ours (JLEQ)	0.5153	0.5811	0.6147	0.6353
LFH	0.2908	0.4098	0.5446	0.6038
SDH	0.2642	0.3994	0.4145	0.4346
TSH	0.2365	0.3080	0.3455	0.3663
KSH	0.2334	0.2662	0.2923	0.3128
SPLH	0.1588	0.1635	0.1701	0.1730
COSDISH	0.4986	0.5768	0.6191	0.6371
FastH	0.4230	0.5216	0.5970	0.6446

Table 2. Results of MAP on NUS-WIDE

Method	8-bits	16-bits	32-bits	64-bits
Ours (JLEQ)	0.5892	0.6134	0.6023	0.5940
SDH	0.4739	0.4674	0.4908	0.4944
LFH	0.5437	0.5929	0.6025	0.6136
TSH	0.4593	0.4784	0.4857	0.4955
KSH	0.4275	0.4546	0.4645	0.4688
SPLH	0.3769	0.4077	0.4147	0.4071
COSDISH	0.5454	0.5940	0.6218	0.6329
FastH	0.5014	0.5296	0.5541	0.5736

References

Gong, Y., Lazebnik, S., Gordo, A., & Perronnin, F. (2013). Iterative quantization: A procrustean approach to learning binary codes for large-scale image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(12), 2916-2929.

Lin, G., Shen, C., Suter, D., & Van Den Hengel, A. (2013). A general two-step approach to learning-based hashing. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 2552-2559).

Zhang, P., Zhang, W., Li, W. J., & Guo, M. (2014, July). Supervised hashing with latent factor models. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval* (pp. 173-182). ACM.