Supervised non-negative matrix factorization based latent semantic image indexing

Dong Liang (梁 栋), Jie Yang (杨 杰), and Yuchou Chang (常宇畴)

Institute of Image Processing and Pattern Recognition, Shanghai Jiaotong University, Shanghai 200240

Received November 2, 2005

A novel latent semantic indexing (LSI) approach for content-based image retrieval is presented in this paper. Firstly, an extension of non-negative matrix factorization (NMF) to supervised initialization is discussed. Then, supervised NMF is used in LSI to find the relationships between low-level features and high-level semantics. The retrieved results are compared with other approaches and a good performance is obtained.

OCIS codes: 100.0100, 100.5010, 150.0150.

Latent semantic indexing (LSI) is a powerful tool to search the mapping from low-level features to high-level semantics in image retrieval[1]. In traditional LSI, singular value decomposition (SVD) is used to create a lower dimensional semantic space. However, it is difficult to interpret negative components in decomposed matrices considering they represent feature frequency. On the other hand, the latent semantic space derived by SVD is orthogonal, which implies that all query topics are orthogonal. In practice, it is quite common that the high-level semantics comprising an image collection are not completely independent of each other[2].

Recently, an unsupervised technique non-negative matrix factorization (NMF) was introduced[3] for searching a reduced representation of global data. In our paper, a supervised NMF (SNMF) is proposed. Furthermore, SNMF is used in LSI as an alternative of SVD to find the relationships between low-level features and high-level semantics. The motivation of introducing NMF to LSI for image retrieval is that image data and features are non-negative and NMF is based on non-negative restrictions. Moreover, the latent semantic space derived by NMF does not need to be orthogonal, and each image is guaranteed to take only non-negative values in all the latent semantic directions. It means that each axis in the space derived by the NMF has a much more straightforward correspondence with each semantic class than in the space derived by SVD[2]. Lastly, NMF is advantageous for applications involving large matrices because NMF computation is based on the simple iterative algorithm.

NMF is an outstanding method to decompose a non-negative matrix into two non-negative matrices as $V \approx WH$. The dimensions of basis matrix W and encoding matrix H are $n \times r$ and $r \times m$ respectively. Rank r is chosen to satisfy $(n+m)r < nm$.

In order to demonstrate the difference between NMF and SVD, we apply NMF and SVD to one data set. This dataset consists of two clusters, and each cluster contains ten samples. NMF is used with $r = 2$. We plot the dataset and the semantic directions found by NMF and SVD in Fig. 1, where the difference between NMF and SVD is well represented. Firstly, NMF does not require the derived lower dimensional space to be orthogonal, and it guarantees that each sample takes only non-negative values in all the semantic directions. While in the SVD space, each sample may take negative values in some of the directions. Secondly, each direction corresponds to a semantic class in the NMF space, and all the samples belonging to the same semantic class spread along the same direction. While the orthogonal requirement by SVD makes the derived latent semantic directions less likely to correspond to each of the clusters and does not provide a direct indication of the data partitions[2]. From the analysis above, in addition to the non-negativity, another property of NMF is that the columns of W tend to represent clusters of semantically relative elements as used in the semantic analysis of a corpus of encyclopedia articles[3]. Therefore, W describes a hidden reduced semantic space and H contains the semantic features of the samples in data matrix V.

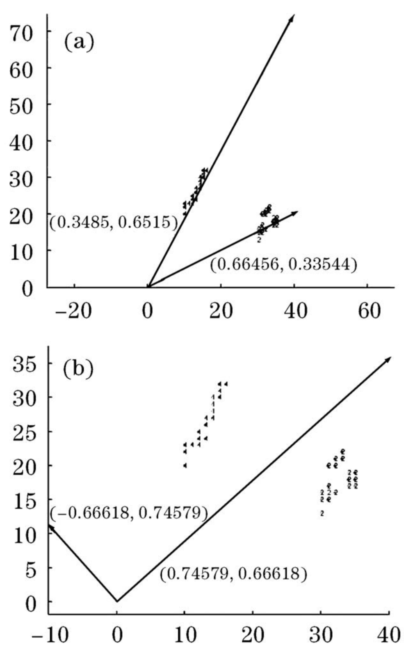

Fig. 1. The dataset and semantic directions found by NMF (a) and by SVD (b).

Problem that one might encounter in application of standard NMF is random initialization because the solution of NMF is always a local optimal solution. However, the entirely random initialization is time consuming in an average sense because it provides a random starting point for searching a parts-based representation of global data. Compared with random initialization, structured initialization[4] provides a most restrictive starting point and cannot pull the factorization out once the algorithm plunges into a local minimum. Maybe a supervised but flexible initialization is more appropriate. The idea is to initialize W with representative vectors of each known semantic class in learning stage and randomly initialize H for less restrictive. Concept vector[5] is used to represent each semantic cluster and initialize each column of W with three reasons. Firstly, concept vector is non-negative; secondly, based on the Cauchy-Schwarz inequality, in an average sense, all of the vectors in a cluster are closer to the concept vector than any other vector, as measured by cosine similarity; thirdly, for a given cluster, the only basis feature obtained by NMF when r is configured to be 1 equals concept vector of given cluster.

We design an experiment to explore the progression of basis images in iterating for three initializations. ORL face database is used because the progression of W can be better visualized by seeing the changes made in the basis faces, where W consists of 40 basis faces. Figure 2 shows the progression of 40 basis images after 0, 10, and 100 iterations. The faces on the top, middle and bottom rows are those obtained using random initialization, supervised initialization, and structured initialization, respectively. We can see that supervised initialization has a head start at beginning to emphasize, and extract facial components compared with other initializations. Conversely, we cannot see any emphasized parts in basis faces for structured initialization after 100 iterations. Figure 2 indicates that structured initialization maybe violate the original idea of NMF and supervised initialization maybe more appropriate. Another advantage benefited from the supervised initialization is that when initialize W with concept vectors, the lower rank r is determined simultaneously equaling the number of semantic categories in the database.

Here, we give a description of SNMF-based latent semantic image indexing approach for image retrieval assuming k semantic categories in image set.

Learning Stage: Firstly, low-level features of images in the database are extracted to construct a feature-image matrix V^p (say of size $n \times m$). Then, concept vector of each semantic category is calculated to form concept vector matrix $C = \{c_1, c_2, \cdots, c_k\}$ (say of size $n \times k$). Lastly, SNMF is used to factorize V^p into W^P andH^P, where W^P is initialized by C and H^P is random initialized.

Retrieval Stage: For query image q, low-level features V^q are projected into W^P and the latent semantic feature H^q is extracted using NMF-projecting algorithm[6]. After measuring the similarity between semantic features of the query image and each image in the database, some top images are shown to the user.

The image database used consists of 1906 images of 74 semantic categories. Hue-saturation combined histogram feature[7] is used because each dimension represents

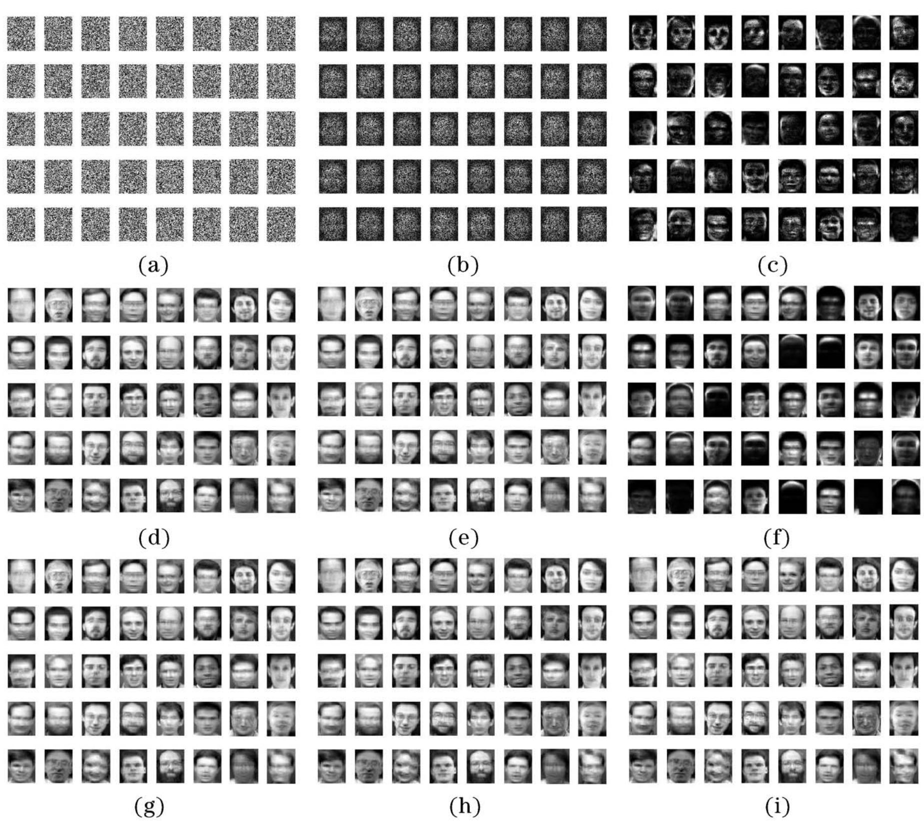

Fig. 2. Progress for the random initialization after 0, 10 and 100 iterations respectively (a)—(c), for the supervised initialization (d)—(f), and for the structured initialization (g)—(i).

the frequency of each component emerging. Average-retrieval-rank[8], which is a variation on recall, is used as evaluation criterion. The reason for selecting average-retrieval-rank as criterion is that it is independent of the number of returned images. For the same database, the smaller the average-retrieval-rank is, the more effective the approach is.

Firstly, we design an experiment to evaluate the effectiveness of standard NMF-based LSI approach, where SVD-based LSI approach and the approach without LSI are used as the baselines. Figure 3 shows average-retrieval-rank, as a function of the reduced dimensions when the number of iterations is fixed to 200. As seen from Fig. 3, standard NMF-based LSI approach performs better than other approaches. Note that no special initialization for NMF involved in this experiment.

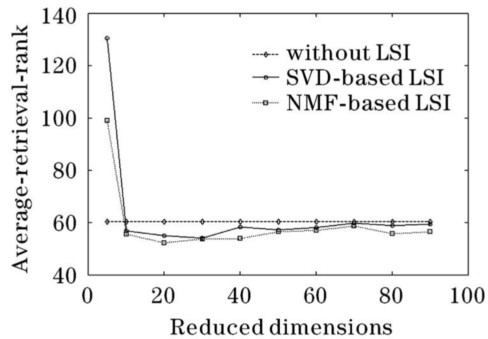

Fig. 3. Average-retrieval-rank as a function of reduced dimensions (200 iterations).

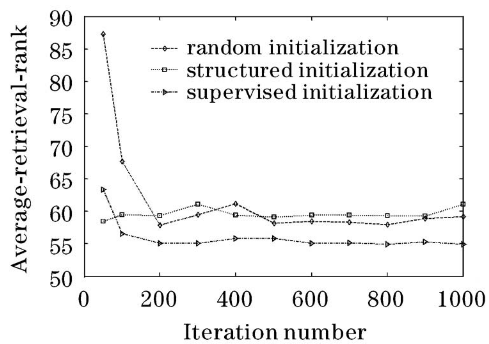

Fig. 4. Average-retrieval-rank as a function of number of iterations (rank r was fixed to 74).

Another experiment is designed to evaluate three initializations. Figure 4 shows the average-retrieval-rank, as a function of the iteration number when the reduced dimension is fixed to 74. We can see that the performance of supervised initialization is superior to that of structured initialization and random initialization. The integrated performances on two experiments indicate that the approach of supervised NMF is the better choice for LSI than others.

In conclusion, an extension of NMF to supervised initialization is proposed. Supervised NMF provides a restrictive and flexible starting point of searching reduced representation of global data. In addition, SNMF is used in LSI as an alternative of SVD to extract the latent semantic structure of images. Experimental results show that the proposed approach performs better than other approaches.

This work was supported by the Key Technologies R&D Program of Shanghai under Grant No. 03DZ19320. D. Liang's e-mail address is rallip@sjtu.edu.cn.

References

1. Z. Pečenović, *Image Retrieval Using Latent Semantic Indexing* M. Sc. Thesis (Swiss Federal Institute of Technology, Lausanne, 1997).
2. W. Xu, X. Liu, and Y. Gong, *SIGIR Forum* (Toronto, Canada, 2003) p.267.
3. D. D. Lee and H. S. Seung, Nature **401,** 788 (1999).
4. S. Wild, J. Curry, and A. Dougherty, Pattern Recognition **37,** 2217 (2004).
5. I. S. Dhillon and D. S. Modha, Machine Learning **42,** 143 (2001).
6. D. Guillamet and J. Vitria, in *Proceedings of 11th International Conference on Image Analysis and Processing* 256 (2001).
7. R. Zhao and W. I. Grosky, Pattern Recognition **35,** 593 (2002).
8. S.-O. Shim and T.-S. Choi, in *Proceedings of IEEE International Conference on Image Processing* 493 (2003).