Deriving Semantics for Image Clustering from Accumulated User Feedbacks

Machine Vision & Pattern Recognition Lab.

Yanhua Chen, Manjeet Rege, Ming Dong and Farshad Fotouhi
{chenyanh, rege, mdong, fotouhi}@wayne.edu
Department of Computer Science, Wayne State University, MI, USA

Wayne State University Department of Computer Science — World-Class Education in the Real World™

Introduction

Image clustering solely based on visual features without any knowledge or background information suffers from the problem of semantic gap. We propose **SS-NMF**: a **Semi-Supervised Non-negative Matrix Factorization** framework for image clustering. Accumulated relevance feedback in a content-based image retrieval (CBIR) system is treated as user provided supervision for guiding the image clustering. We show that supervision derived from the few images marked in the feedback logs can greatly enhance the image clustering results.

NMF (Non-negative Matrix Factorization)

1. NMF was initially proposed for "parts-of-whole" decomposition, and later extended to a general framework for data clustering. It can model widely varying data distributions and can accomplish both hard and soft clustering simultaneously.

2. We perform symmetric non-negative tri-factorization of image-image similarity matrix $A = X^T X \in R^{n \times n}$ (where $X \in R^{m \times n}$ is feature-image matrix) to do image clustering:

$$A \approx GSG^T$$

where $G \in R^{n \times k}$ is the cluster indicator matrix, $S \in R^{k \times k}$ is the cluster centroid matrix that gives a compact $k \times k$ representation of X, with k being the number of clusters.

Clustering with Accumulated User Feedbacks

Relevance feedback logs (RF_i):

1. F_i^+: denotes the set of positive images marked in the feedback
2. F_i^-: denotes the set of negative images marked in the feedback

Define set of pairwise constraints:

1. Must-Link constraints C_{ML}: pair of images $(i_i, i_j) \in C_{ML}$ indicating i_i and i_j must be clustered together *iff* $(i_i, i_j) \in RF_h$, $1 \le h \le k$ (total number of logs)
2. Cannot-Link constraints C_{CL}: pair of images $(i_i, i_j) \in C_{CL}$ indicating i_i and i_j must be clustered together *iff* one of the two images $\in F_h^+$, while the other $\in F_h^-$, $1 \le h \le k$ (total number of logs)

SS-NMF Clustering

Objective function of SS-NMF

$$J_{SS-NMF} = \min_{S \ge 0, G \ge 0} \left\| \tilde{A} - GSG^T \right\|^2 = \min_{S \ge 0, G \ge 0} \left\| (A - W_{reward} + W_{penalty}) - GSG^T \right\|^2$$

where $\tilde{A}$ is affinity or similarity matrix A with constraints $W_{reward} = \{w_{ij} \mid (i_i, i_j) \in C_{ML}, s.t. y_i = y_j\}$ and $W_{penalty} = \{w_{ij} \mid (i_i, i_j) \in C_{CL}, s.t. y_i = y_j\}$,

w_{ij} is the penalty cost for violating a constraint between images i_i and i_j, and y_i is the cluster label of i_i

Updating rules

We propose an iterative procedure for the minimization of objective function where we update one factor while fixing the others:

$$S_{ih} \leftarrow S_{ih} \frac{(G^T \tilde{A} G)_{ih}}{(G^T G S G^T G)_{ih}} \qquad G_{ih} \leftarrow G_{ih} \frac{(\tilde{A} G S)_{ih}}{(G S G^T G S)_{ih}}$$

SS-NMF Algorithm Correctness and Convergence

Correctness Proof

1. Introduce the Lagrangian multipliers λ_1 and λ_2 to minimize the lagrangian function,

$$L(S, G, \lambda_1, \lambda_2) = \min_{S \ge 0, G \ge 0} \left\| \tilde{A} - GSG^T \right\|^2 - Tr(\lambda_1 S^T) - Tr(\lambda_2 G^T)$$

2. Based on Kullback-Leibler complementarily condition, we can compute the gradient descent of $\partial J / \partial S$ while fixing G. We can then successively update S which will converge to a local minima of the problem.
3. Similarly, given S, we can update G to make $\partial J / \partial G$ monotonically decreasing which will converge to a local minima of the problem.
4. S and G should update alternatively.

Convergence Proof

1. Assuming $L(S, S')$ is an auxiliary function of $J(S)$ if $L(S, S') \ge J(S)$ and $L(S, S) = J(S)$, we minimize a lower bound, set $S^{(t+1)} = \arg\min_S L(S, S^{(t)})$, then $J(S^{(t)}) = L(S^{(t)}, S^{(t)}) \ge L(S^{(t+1)}, S^{(t)}) \ge J(S^{(t+1)})$. Thus $J(S)$ is monotonically decreasing and is bounded from up.
2. Similarly, assuming $L(G, G')$ is an auxiliary function of $J(G)$ if $L(G, G) = J(G)$ and $L(G, G') \ge J(G)$, we minimize a lower bound, set $G^{(t+1)} = \arg\min_G L(G, G^{(t)})$, then $J(S^{(t)}) = L(S^{(t)}, S^{(t)}) \ge L(S^{(t+1)}, S^{(t)}) \ge J(S^{(t+1)})$. Thus $J(G)$ is monotonically decreasing and is bounded from up.

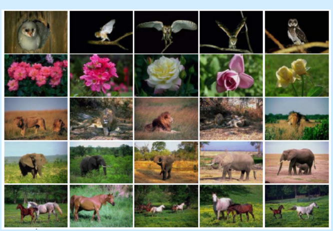
Advantages of SS-NMF

	SS-KK (Semi-Supervised Kernel K-means)	SS-SNC (Semi-Supervised Spectral Normalized Cuts)	SS-NMF (Semi-Supervised Non-negative Matrix Factorization)
Clustering Indicator	•Hard clustering •Exact orthogonal	•The derived latent semantic space to be orthogonal. •No direct relationship between the singular vectors and the clusters	•Soft clustering •Map the images into non-negative latent semantic space which may not be orthogonal •Cluster label can be determined by the axis with the largest projection value
Time Complexity	•Iterative algorithm	•Solving a computationally expensive constrained eigen decomposition	•Efficient iterative algorithm •Simple based on basic matrix computation and easily deployed over a distributed computing environment when dealing with large image repositories. •Obtain partial answer at intermediate stages of the solution by specifying a fixed number of iterations

Experiment Datasets and Evaluation Methodology

Experiment setup

1. Datasets: entire image database consists of 1,500 images with 300 images in each category, we randomly select 100 images from 5 categories which are *Owls (O)*, *Roses (R)*, *Lions (L)*, *Elephants (E)* and *Horses (H)* to form different combinations of image categories.
   - Must-link: If both the images happen to belong to the same category in the ground truth, the constraint is assigned maximum weight in the image-image similarity matrix;
   - Cannot-link: If both images belong to different categories, the minimum weight in the similarity matrix is used for the constraint.



Up Figure : Sample images from the image categories used

2. Evaluation accuracy metric:

$$AC = \frac{\sum_{i=1}^{n} \delta(y_i, \hat{y}_i)}{n}$$

where $\hat{y}_i$ is the estimated cluster label assigned to an image, y_i is ground truth label, and n denotes the total number of images in the experiment. $\delta(y_i, \hat{y}_i)$ is the delta function that equals one if $\hat{y}_i = y_i$ else its zero.

Experiment Results

Experiment 1:

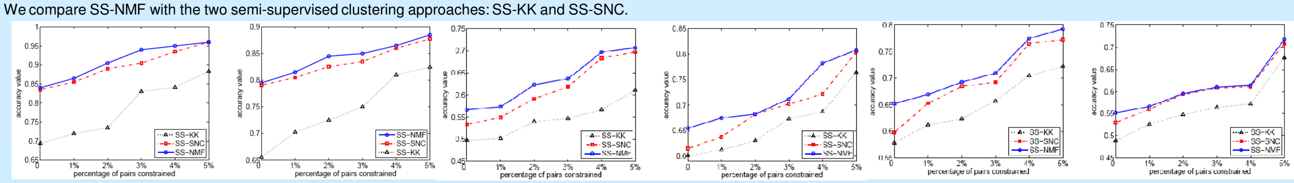
We perform comparison of three popular unsupervised image clustering methods: KK (Kernel K-means), SNC (Spectral Normalized Cuts), and NMF, with SS-NMF.

Right Table : Comparison of image clustering accuracy between KK, SNC, NMF and , SS-NMF with only 3% pairwise constraints on the images. It shows that SS-NMF consistently outperforms other well-established unsupervised image clustering methods.

	O-R	L-H	R-L	O-R-L	O-R-L-E	O-L-E-H
KK	0.6933	0.6553	0.8600	0.6750	0.6012	0.5775
SNC	0.8300	0.7900	0.8750	0.7092	0.6150	0.5975
NMF	0.8400	0.7950	0.8950	0.7167	0.6550	0.6525
SS-NMF	**0.9400**	**0.8500**	**0.9300**	**0.8833**	**0.7125**	**0.7095**

Experiment 2:

We compare SS-NMF with the two semi-supervised clustering approaches: SS-KK and SS-SNC.



From Left to Right Figures : Comparison of image clustering accuracy between SS-KK, SS-SNC, and SS-NMF for different percentages of images pairs constrained (O-R, L-H, L-E-H, O-R-L-E, O-L-E-H, and O-R-L-E-H). It shows that SS-NMF consistently outperforms other semi-supervised image clustering methods.

ACM Multimedia 2007, Germany