Deriving Semantics for Image Clustering from Accumulated User Feedbacks

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

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 $A = X'X \in R^{m \times m}$ (where $X \in R^{m \times n}$ is feature-image matrix) to do image clustering:

   $A = GS^T$

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

SS-NMF Clustering

Objective function of SS-NMF

$J = \sum_{ij} \beta_{ij} ||x_i - W_{ij} - W_{ij}^T||_2^2 + \lambda \sum_{ij} g_{ij} ||W_{ij}||_F$

where $A$ is affinity or similarity matrix $A$ with constraints $\sum_{ij} w_{ij} = 1 \forall i,j \in C,$ $\sum_{i} w_{ij} = n_j \forall j \in C,$ $\sum_{j} w_{ij} = n_i \forall i \in C,$ $w_{ij}$ is the penalty cost for violating a constraint between images $i$ and $j$, and $g_{ij}$ is the cluster label of $i$.

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

1. $x_i \leftarrow \frac{\sum_j w_{ij} g_{ij} x_j}{\sum_j w_{ij} g_{ij}}$
2. $G \leftarrow \frac{\sum_i w_{ij} g_{ij} x_i}{\sum_i w_{ij} g_{ij}}$

SS-NMF Algorithm Correctness and Convergence

Correctness Proof
1. Introduce the Lagrangian multipliers $\lambda_i$ and $\mu_j$ to minimize the lagrangian function,

   $L = \sum_{ij} \beta_{ij} ||x_i - W_{ij} - W_{ij}^T||_2^2 + \lambda \sum_{ij} g_{ij} ||W_{ij}||_F$

2. Based on Kullback-Leibler complementarily condition, we can compute the gradient descent of $\beta_{ij}$ 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 $\sum_i g_{ij}$ monotonically decreasing which will converge to a local minima of the problem.

4. $S$ and $G$ should update alternatively.

Convergence Proof
1. Assuming $J(S, G)$ is an auxiliary function of $J(S)$ and $J(G)$, we minimize a lower bound, set $\delta S_i = -\delta J(S, G)/\delta S_i$, then $J(S) = \sum_i ||x_i - W_{ii}||_2^2 + \lambda \sum_i g_{ii} ||W_{ii}||_F$.

2. Similarly, assuming $J(G, G)$ is an auxiliary function of $J(G)$ and $J(G, G)$, we minimize a lower bound, set $\delta G_j = -\delta J(G, G)/\delta G_j$, then $J(G) = \sum_j ||x_j - W_{jj}||_2^2 + \lambda \sum_j g_{jj} ||W_{jj}||_F$.

Advantages of SS-NMF

- The derived latent semantic space to be orthogonal.
- No direct relationship between the individuals and the clusters.
- 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.

2. Evaluation accuracy metric:

   $AC = \sum_i \delta y_i \delta \hat{y}_i$

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

Experiment Results

Experiment 1:
We perform comparison of three popular unsupervised image clustering methods: K-mean, Spectral Normalized Cuts, and NMF, with SS-NMF.

Right Table: Comparison of image clustering accuracy between SS-KK, SS-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.

Experiment 2:
We compare SS-NMF with the two semi-supervised clustering approaches: SS-KK and SS-SNC.