Image Co-clustering with Multi-modality Features and User Feedbacks

Yanhua Chen, Ming Dong
{chenyanh, mdong}@wayne.edu
Machine Vision and Pattern Recognition Lab
Department of Computer Science,
Wayne State University, MI, USA

Wanggen Wan
{wanwg}@staff.shu.edu.cn
School of Communication and Information Engineering,
Shanghai University, Shanghai, P.R.China

WAYNE STATE UNIVERSITY World-Class Education in the Real World™ Warriors

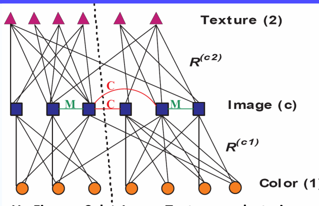
上海大学 Shanghai University

Introduction

In Content-based Image Retrieval (CBIR) research, advance technology that fuses the heterogeneous information into image clustering has drawn extensive attention recently . However, using multiple features for co-clustering images without any user feedbacks is a challenging problem. We propose **SS-NMF**: a **Semi-Supervised Non-negative Matrix Factorization** framework to incorporate user feedbacks into image co-clustering. SS-NMF improved the quality of image co-clustering by learning distance metrics based on user feedbacks.

SS-NMF (Non-negative Matrix Factorization) Co-clustering Model

1. NMF was initially proposed for "parts-of-whole" decomposition, and later extended to general framework for data clustering.
2. Given a **Heterogeneous Relational Data (HRD)** set with a central data type (i.e., image) and l (1≤p≤l) feature modalities (i.e., color, texture), the goal of SS-NMF is to simultaneously cluster images into k_c disjoint clusters and l features into k_p disjoint clusters.



Up Figure : Color-Image-Texture co-clustering with constraints

NMF tri-factorization: Model HRD as a set of related matrices by using $R^{(cp)}$ to represent the relation between an image data and a feature modality. The task of co-clustering can be formulated as an optimization problem as,

$$J = \min_{G^{(c)}\ge 0, G^{(p)}\ge 0, S^{(cp)}\ge 0} \sum_{1\le p\le l} \left\| R^{(cp)} - G^{(c)} S^{(cp)} G^{(p)} \right\|^2 \qquad (1)$$

where $G^{(c)}$ and $G^{(p)}$ are the cluster indicator matrices for images and each feature modality respectively, $S^{(cp)}$ is the cluster association matrix which gives the relation between images and each feature modality.

Define user feedbacks: In SS-NMF, all the images marked in the user feedbacks are viewed as a form of user provided supervision. A user marks a few images as relevant (or positive) and non-relevant (or negative).

1. Must-Link constraints $M=\{(x_i,x_j)\}$: positive images x_i and x_j are labeled as belonging to the same cluster
2. Cannot-Link constraints $C=\{(x_i,x_j)\}$: negative images x_i and x_j are labeled as belonging to the different cluster

SS-NMF Co-clustering Algorithm

Distance Metric Learning

In order to accomplish image co-clustering with user feedbacks, we learn a new distance metric $L^{(cp)}$ (p=1, 2) over each feature modality of the Euclidean distance $d(x_i^{(cp)}, x_j^{(cp)})$ such that $(x_i^{(cp)}, x_j^{(cp)})$ belong to M are moved closer to each other while $(x_i^{(cp)}, x_j^{(cp)})$ belong to C are moved further away.

Let $R^{(c1)}$ denotes the image-color matrix and $R^{(c2)}$ denotes the image-texture matrix, the distance metric can be learned by the following optimization problem,

$$\max g(L^{(cp)}) = \frac{\sum_{(x_i^{(cp)},x_j^{(cp)})\in C} \left\| x_i^{(cp)}, x_j^{(cp)} \right\|_{L^{(cp)}}}{\sum_{(x_i^{(cp)},x_j^{(cp)})\in M} \left\| x_i^{(cp)}, x_j^{(cp)} \right\|_{L^{(cp)}}} \qquad (2)$$

This maximization problem is equivalent to the generalized **Semi-supervised Discriminate Analysis (SS-LDA)** problem.

Modality Selection

Each feature modalities can play different roles in the grouping images, then introduce a factor a=[$\alpha^{(c1)}$, $\alpha^{(c2)}$] to denote the relative importance of color and texture.

Note that the modality selection and distance metric learning are strongly dependent and must be achieved simultaneously so that the image clustering results can be globally optimized.

Updating rules

We propose an iterative procedure for the minimization of objective function on new distance metrics $R^{(c1)}$ and $R^{(c2)}$ where we update one factor while fixing the others:

$$G_{ih}^{(1)} \leftarrow G_{ih}^{(1)} \frac{\left(S^{(c1)^T} G^{(c)^T} \tilde{R}^{(c1)}\right)_{ih}}{\left(S^{(c1)^T} G^{(c)^T} G^{(c)} S^{(c1)} G^{(1)}\right)_{ih}} \qquad (3)$$

$$G_{ih}^{(c)} \leftarrow G_{ih}^{(c)} \frac{\left(\tilde{R}^{(c1)} G^{(1)^T} S^{(c1)^T}\right)_{ih} + \left(\tilde{R}^{(c2)} G^{(2)^T} S^{(c2)^T}\right)_{ih}}{\left(G^{(c)} S^{(c1)} G^{(1)} G^{(1)^T} S^{(c1)^T}\right)_{ih} + \left(G^{(c)} S^{(c2)} G^{(2)} G^{(2)^T} S^{(c2)^T}\right)_{ih}} \qquad (4)$$

$$G_{ih}^{(2)} \leftarrow G_{ih}^{(2)} \frac{\left(S^{(c2)^T} G^{(c)^T} \tilde{R}^{(c2)}\right)_{ih}}{\left(S^{(c2)^T} G^{(c)^T} G^{(c)} S^{(c2)} G^{(2)}\right)_{ih}} \qquad (5)$$

$$S_{ih}^{(c1)} \leftarrow S_{ih}^{(c1)} \frac{\left(G^{(c)^T} \tilde{R}^{(c1)} G^{(1)^T}\right)_{ih}}{\left(G^{(c)^T} G^{(c)} S^{(c1)} G^{(1)} G^{(1)^T}\right)_{ih}} \qquad (6)$$

$$S_{ih}^{(c2)} \leftarrow S_{ih}^{(c2)} \frac{\left(G^{(c)^T} \tilde{R}^{(c2)} G^{(2)^T}\right)_{ih}}{\left(G^{(c)^T} G^{(c)} S^{(c2)} G^{(2)} G^{(2)^T}\right)_{ih}} \qquad (7)$$

Theoretical Analysis

Correctness and Convergence

1. **Correctness**: If the solution converges based on the updating rules in Equations (3)-(7), the solution satisfies the KKT optimality condition.
2. **Convergence**: If any four of five matrices $G^{(1)}$, $S^{(c1)}$, $G^{(c)}$, $S^{(c2)}$, and $G^{(2)}$ are fixed, Equation (1) deceases monotonically under the updating rules of Equations (3)-(7).

Advantages of SS-NMF

Compare with Spectral Relational Clustering (SRC), NMF co-clustering relaxes orthogonal requirements on the cluster indictor matrices to be near-orthogonal, leading to soft clustering results which imply that each image can fractionally belong to more than one cluster and providing a more intuitive clustering result.

Experiment Datasets and Evaluation Methodology

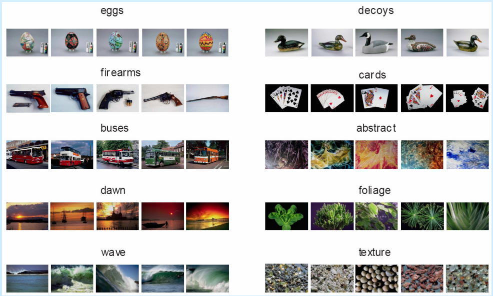
1. **Datasets**:

Name	Data structure	No. of modalities	No. of clusters	No. of images
IT1	eggs,decoys	3	2	200
IT2	dawn,foliage	3	2	200
IT3	decoys,dawn	3	2	200
IT4	decoys,firearms,cards,buses	3	4	400
IT5	abstract,dawn,foliage,waves	3	4	400
IT6	eggs,decoys,dawn,foliage	3	4	400
IT7	eggs,decoys,buses,	3	6	600

Right Table : Data sets for Image co-clustering.

2. **Evaluation accuracy metric**:

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



Up Figure : Image Samples for Co-clustering

Experiment Results

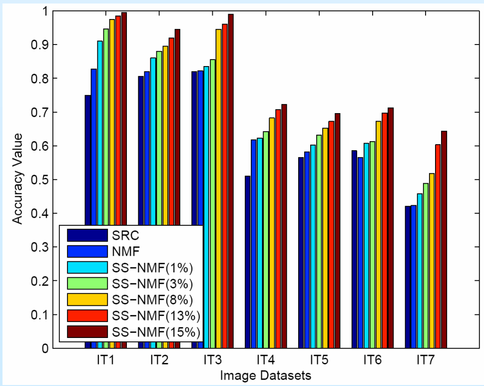
We compare SS-NMF image co-clustering with SRC and NMF methods on:
1. Image Co-clustering Accuracy, 2. Modality Selection, and 3. Time Complexity.

	Co-clustering Accuracy			Modality Importance	
Name	SRC	NMF	SS-NMF	image-color	image-texture
IT1	0.7500	0.8275	0.9950	0.0001	0.2189
IT2	0.8050	0.8200	0.9450	0.1890	0.0002
IT3	0.8200	0.8230	0.9900	0.2188	0.0005
IT4	0.5100	0.6175	0.7225	0.0088	0.2357
IT5	0.5650	0.5810	0.6950	0.3040	0.0002
IT6	0.5850	0.5625	0.7125	0.0001	0.2007
IT7	0.4210	0.4231	0.6433	0.1102	0.0486

Up Table :
Left: Comparison of clustering accuracy between SRC, NMF and SS-NMF with 15% constraints on image co-clustering.
Right: Modality importance of SS-NMF with 3% constraints for image co-clustering: color vs. texture

Method	Time Complexity
SRC	$O(tlkn_cn_p)$
NMF	$O(tl(n_p^3 + kn_cn_p))$
SS-NMF	$O(t(l\max(n_c, n_p)^3 + kn_cn_p))$

Left Table : Comparison of computational speed of SRC, NMF and SS-NMF



Up Figure : Comparison of clustering accuracy between SRC, NMF and SS-NMF with different amount of constraints for image co-clustering

Results show that SS-NMF provides superior performance in terms of high co-clustering accuracy, reasonable modality selection, and efficient computational time .

ACM Multimedia 2009, Beijing, China