2007

Deriving semantics for image clustering from accumulated user feedbacks

Yanhua Chen
Manjeet Rege
Ming Dong
Farshad Fotouhi

Follow this and additional works at: http://scholarworks.rit.edu/other

Recommended Citation

Deriving Semantics for Image Clustering from Accumulated User Feedbacks, Proceedings of ACM Multimedia 2007. Held in Augsburg, Germany: 24-29 September 2007. This is a digitized copy derived from an ACM-copyrighted work. ACM did not prepare this copy and does not guarantee that it is an accurate copy of the originally published work.

This Conference Proceeding is brought to you for free and open access by RIT Scholar Works. It has been accepted for inclusion in Presentations and other scholarship by an authorized administrator of RIT Scholar Works. For more information, please contact ritscholarworks@rit.edu.
Deriving Semantics for Image Clustering from Accumulated User Feedbacks

Yanhua Chen Manjeet Rege Ming Dong Farshad Fotouhi
Machine Vision Pattern Recognition Lab Database and Multimedia Systems Group
Department of Computer Science, Wayne State University
Detroit, MI 48202, USA
{chenyanh.rege, mdong, fotouhi}@wayne.edu

ABSTRACT
Image clustering solely based on visual features without any knowledge or background information suffers from the problem of semantic gap. In this paper, we propose SS-NMF: a semi-supervised non-negative matrix factorization framework for image clustering. Accumulated relevance feedback in a CBIR system is treated as user provided supervision for guiding the image clustering. We consider the set of positive images in the feedback as constraints on the clustering specifying that the images “must” be clustered together. Similarly, negative images provide constraints specifying that they “cannot” be clustered along with the positive images. Through an iterative algorithm, we perform symmetric non-negative tri-factorization of the image-image similarity matrix to infer the clustering. Theoretically, we prove the correctness of SS-NMF by showing that the algorithm is guaranteed to converge. Through experiments conducted on general purpose image datasets, we demonstrate the superior performance of SS-NMF for clustering images effectively.

Categories and Subject Descriptors
1.5.3 [Pattern Recognition]: Clustering-algorithms.

General Terms
Algorithms, Performance, Experimentation.

1. INTRODUCTION
Advancements in digital technology over the last decade has resulted in an exponential growth in the number of images being produced in applications such as medical image databases, criminal suspect tracking, travel image gallery, and personal or family picture collections. This phenomenon created a need to develop automatic techniques for organizing these image collections in order to support user-friendly browsing. Consequently, image clustering has received extensive attention in multimedia and content-based image retrieval (CBIR) research.

There has been considerable work done on unsupervised approaches for image clustering [1–3]. These methods primarily perform image clustering by analyzing the visual features extracted from the images. However, current state-of-the-art visual features are unable to represent image content on a semantic level. As a result, these approaches are susceptible to the problem of semantic gap. That is, images dissimilar semantically but similar in the visual feature space may get clustered together. On the other hand, semantically similar images may get clustered separately. Hence, it is highly desired to incorporate semantic information into the image clustering process.

Relevance feedback mechanism, originally developed for text document retrieval, has served as a remedy to the semantic gap problem in CBIR to an extent. Its basic idea is to have the users provide feedback by evaluating the query results. Typically, in a CBIR system, a user marks a few images as relevant (or positive) and non-relevant (or negative). By keeping the users in the loop during retrieval, a CBIR system is able to adjust query at every iteration for subjectivity in judgement by learning from the positive and negative examples provided by the users. On the other hand, the accumulated user feedback logs can be made use of for long-term learning in image databases [4, 5].

In this paper, we propose a non-negative matrix factorization (NMF) [6, 7] based framework to incorporate semantics into image clustering. Under the proposed semi-supervised NMF (SS-NMF) methodology, accumulated user feedback logs in a CBIR system are utilized to derive semantics for clustering the images. All the images marked in the feedback are viewed as a form of user provided supervision. The set of positive images are looked upon as constraints specifying that they “must” be clustered together. Similarly, the set of negative images provide the constraints that they “cannot” be clustered together with the positive images. Note that the user usually marks a few images in his/her feedback with most of the images being left unmarked. We show that supervision derived from the few images marked in the feedback logs can greatly enhance the image clustering results. We derive an iterative algorithm to perform symmetric non-negative tri-factorization of the image-image similarity matrix. The correctness of the algorithm is proved by showing that the algorithm is guaranteed to converge. Experiments performed on general purpose image datasets demonstrate the effectiveness of the proposed work.

2. SEMI-SUPERVISED NON-NEGATIVE MATRIX FACTORIZATION MODEL FOR IMAGE CLUSTERING

CBIR systems typically represent the entire image collection using a feature-image matrix $X \in \mathbb{R}^{m \times n}$ where columns index the
3. ALGORITHM DERIVATION

While constraint violations are kept to a minimum.

and re-written as:

\[ \begin{align*}
D_y \text{ and penalty cost for violating a constraint between images} \\
\text{(locally) minimized according to the given distortion measure} \ D
\end{align*} \]

Information embedded in accumulated relevance feedback logs contain semantics that can be incorporated into image clustering. Let the total number of logs collected be \( k \). Then for a particular log \( RF_i \), let \( F_i^+ \) and \( F_i^- \) denote the set of positive and negative images marked in this feedback. From this, we can obtain two constraints that can be imposed on clustering the images. First, all images in \( F_i^+ \) are semantically related and must be clustered together. Second, none of the images in \( F_i^- \) should be clustered with any of the images in \( F_i^+ \).

We now define the set of pairwise must-link constraints \( C_{ML} \) and cannot-link constraints \( C_{CL} \) on the images as follows:

- Pair of images \((i, j)\) in \( C_{ML}\) indicating that \( i \) and \( j \) must be clustered together
- Pair of images \((i, j)\) in \( C_{CL}\) indicating that \( i \) and \( j \) cannot be clustered together

The constraints are accompanied by associated violation cost matrix \( W \). An entry \( w_{ij} \) in this matrix denotes the cost of violating the constraint between images \( i \) and \( j \), if such a constraint exists, that is, either \((i, j)\) in \( C_{ML} \) or \((i, j)\) in \( C_{CL} \). The model relies on a distortion measure \( D : R^n \rightarrow R \), to compute distance between images. For a given \( k \), the goal is to partition the set of images into disjoint clusters \( \{X_i\}_{i=1}^k \), such that the total distortion between the images and the corresponding cluster representatives is (locally) minimized according to the given distortion measure \( D \), while constraint violations are kept to a minimum.

3. ALGORITHM DERIVATION

We define the objective function of SS-NMF as follows:

\[ J_{SS-NMF} = \min_{S \geq 0, G \geq 0} \| A - \text{GSTM} \|^2 \]  

We propose an iterative procedure for the minimization of equation (2) where we update one factor while fixing the others. The updating rules are,

\[ S_{ih} \leftarrow S_{ih} \frac{(G^T AG)_{ih}}{(G^T GSG^T G)_{ih}} \]  

\[ G_{ih} \leftarrow G_{ih} \frac{(AGS)_{ih}}{(GSG^T GS)_{ih}} \]

The matrices \( S \) and \( G \) are initialized with non-negative values.

4. PROOF OF ALGORITHM CORRECTNESS AND CONVERGENCE

We now prove the theoretical correctness and convergence of SS-NMF. Motivated by [9], we render the proof based on optimization theory, auxiliary function and several matrix inequalities.

4.1 Correctness

1. Following the standard theory of constrained optimization, we introduce the Lagrangian multipliers \( \lambda_1 \) and \( \lambda_2 \) to minimize the lagrangian function,

\[ L(S, G, \lambda_1, \lambda_2) = \min_{S \geq 0, G \geq 0} \| A - \text{GSTM} \|^2 \]

\[ -\text{Tr}(\lambda_1 S^T) - \text{Tr}(\lambda_2 G^T) \]

2. Based on the Kullback-Leibler complementarity condition, we can compute the gradient descent of \( \frac{\partial L}{\partial S} \) 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 \( \frac{\partial J}{\partial G} \) monotonically decreasing which will converge to a local minima of the problem.

4. \( S \) and \( G \) should update alternatively.

4.2 Convergence

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

5. ADVANTAGES OF SS-NMF

We now discuss the advantages of SS-NMF over other well-known semi-supervised clustering algorithms, i.e., semi-supervised kernel k-means (SS-KK) [10] and semi-supervised spectral clustering with normalized cuts (SS-SNC) [11]. In SS-KK, the cluster indicator \( G \) is required to be orthogonal, meaning that each row of \( G \) has only one non-zero element, implying that each image belongs to only 1 cluster. In SS-NMF, \( G \) is near-orthogonal leading to soft clustering results, i.e., each image can fractionally belong to more than 1 cluster. This is particularly helpful for knowledge discovery in image clustering. Note that the user constraints might...
not be consistent across different feedbacks for the same image. The semantics of an image is usually imprecise, and depends on the users’ interpretation. For example, a pair of images can have a must-link constraint between them because of one feedback and a cannot-link constraint due to another. Hence, an image can belong to more than one semantic category. With the soft clustering capability in SS-NMF, we can discover semantics between images and different clusters, that otherwise would not have been possible with the traditional hard clustering approach.

With regards to computational efficiency, SS-NMF uses an efficient iterative algorithm instead of solving a computationally expensive constrained eigen decomposition problem as in SS-SNC. The time complexity of SS-NMF is $O((tkn^2)$ where $k$ is the number of clusters, $n$ is the number of images, and $t$ is the number of iterations. In fact, the time complexity is similar to that of the classical SS-KK clustering algorithm. However, compared to SS-KK, SS-NMF algorithm is simple as it involves some basic matrix operations and hence can be easily deployed over a distributed computing environment when dealing with large image repositories. Another advantage in favor of SS-NMF is that a partial answer can be obtained at intermediate stages of the solution by specifying a fixed number of iterations.

6. EXPERIMENTS

6.1 Datasets and Evaluation Methodology

![Sample images from the image categories used.](image)

In order to evaluate the proposed algorithm, we simulated the user feedback logs by generating the pairwise constraints on 5 image categories from the Corel image database. Each of the constraints were generated by randomly selecting pairs of images. 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 (must-link). On the other hand, if they belong to different categories, the minimum weight in the similarity matrix is used for the constraint (cannot-link). The image categories used in our experiments were Owls, Roses, Elephants and Horses. We refer to the categories using the first alphabet in the figures and tables as O, R, L, E, and H, respectively. Each category consists of 100 images. For the image features, we adopted the HSV space and performed principal component analysis (PCA) along H, S and V dimensions separately. The image was then projected in the eigen vector space to get weights along the principal components. A feature vector for each image was formed by concatenating weights along the three dimensions. For the experiments, we formed different combinations of image categories by mixing different categories together. To evaluate, we have used the accuracy metric AC defined as:

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

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. Since, an iterative algorithm is not guaranteed to find the global minimum, it is beneficial to run the algorithm several times with different initial values and choose one trial with a minimal objective value. In reality, usually a few number of trials are sufficient.

6.2 Results

We first perform comparison of 3 popular unsupervised image clustering methods: kernel k-means (KK), spectral normalized cuts (SNC), and NMF, with the proposed algorithm. For SS-NMF, we provided only 3% pairwise constraints out of a total of $\binom{2}{n}$ possible image pairs, where $n$ denotes the total number of images in the experiment. For KK, we used a polynomial kernel $K = (1 + X^T_X)^p$ with $p = 1$. In Table 1, we compare the algorithms using AC values. Amongst the unsupervised methods, NMF proves to be the best one. However, the accuracy of all the unsupervised methods greatly deteriorates and is unable to produce meaningful results on datasets having more than 2 clusters. This is because, these methods rely only on the visual features of the images to perform the clustering. On the other hand, the superior performance of SS-NMF is evident across all the image datasets. From this experiment, we can infer that in general a semi-supervised method can greatly enhance the image clustering results by benefitting from the user provided knowledge. Moreover, SS-NMF is able to generate significantly better results by quickly learning from few pairwise constraints provided. Table 2 demonstrates the performance of SS-NMF when varying amounts of pairwise constraints were available a priori. We report the results in terms of the confusion matrix $C$ and the cluster centroid matrix $S$. As the available prior knowledge increases from 0% to 5%, we can make the following two key observations. Firstly, the confusion matrices $C$ tend to become perfectly diagonal indicating higher clustering accuracy. Second observation pertains to the cluster centroid matrix $S$ which represents the similarity or distance between the clusters. Increasing values of the diagonal elements of $S$ indicate higher inter-cluster similarities. As expected, when the amount of prior knowledge available is more, the performance of the algorithm clearly gets better.

Table 1: Comparison of image clustering accuracy between KK, SNC, NMF and, SS-NMF with only 3% pairwise constraints on the images.

<table>
<thead>
<tr>
<th></th>
<th>O-R</th>
<th>L-H</th>
<th>R-L</th>
<th>O-R-L</th>
<th>O-R-L-E</th>
<th>O-L-E-H</th>
</tr>
</thead>
<tbody>
<tr>
<td>KK</td>
<td>0.6933</td>
<td>0.6553</td>
<td>0.8600</td>
<td>0.6730</td>
<td>0.6012</td>
<td>0.5775</td>
</tr>
<tr>
<td>SNC</td>
<td>0.8300</td>
<td>0.7900</td>
<td>0.8750</td>
<td>0.7092</td>
<td>0.6150</td>
<td>0.5975</td>
</tr>
<tr>
<td>NMF</td>
<td>0.8400</td>
<td>0.7959</td>
<td>0.8900</td>
<td>0.7450</td>
<td>0.6550</td>
<td>0.6550</td>
</tr>
<tr>
<td>SS-NMF</td>
<td>0.9400</td>
<td>0.8500</td>
<td>0.9300</td>
<td>0.8833</td>
<td>0.7125</td>
<td>0.7095</td>
</tr>
</tbody>
</table>

We now compare SS-NMF with the two semi-supervised clustering approaches: SS-KK and SS-SNC. As before, for SS-KK, a polynomial kernel was used. In Figure 2, we plot the AC values against increasing percentage of pairwise constraints available, for different combinations of the image categories. On the whole, all the three algorithms perform better as the percentage of pairwise constraints increases. SS-KK is unable to achieve decent cluster-


Table 2: The comparison of confusion matrix C and cluster centroid matrix S of SS-NMF for different percentages of image pairs constrained.

<table>
<thead>
<tr>
<th>% of constraints</th>
<th>O-R data set</th>
<th>L-E-H data set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>O-R</td>
<td>L-E-H</td>
</tr>
<tr>
<td></td>
<td>SS-KK</td>
<td>SS-SNC</td>
</tr>
<tr>
<td>0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>81</td>
<td>13</td>
</tr>
<tr>
<td>S</td>
<td>45.659</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>48.837</td>
<td>0</td>
</tr>
<tr>
<td>S</td>
<td>0</td>
<td>48.837</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>50.797</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>84</td>
<td>11</td>
</tr>
<tr>
<td>S</td>
<td>46.029</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>48.837</td>
</tr>
<tr>
<td>S</td>
<td>0</td>
<td>48.837</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>55.531</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>89</td>
<td>1</td>
</tr>
<tr>
<td>S</td>
<td>47.011</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>48.931</td>
</tr>
<tr>
<td>S</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>94</td>
<td>2</td>
</tr>
<tr>
<td>S</td>
<td>48.289</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>49.491</td>
</tr>
<tr>
<td>S</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 2: Comparison of image clustering accuracy between SS-KK, SS-SNC, and SS-NMF for different percentages of image pairs constrained (a) O-R, (b) L-H, (c) R-L, (d) O-R-L, (e) L-E-H, (f) O-R-L-E, (g) O-L-E-H and (h) O-R-L-E-H

7. CONCLUSION

We presented SS-NMF: a semi-supervised approach for image clustering based on non-negative matrix factorization. In the proposed framework, accumulated relevance feedback in a CBIR system is treated as user provided supervision for incorporating semantics into the image clustering process. We derived an iterative algorithm to perform symmetric tri-factorization of the image-image similarity matrix, and proved the correctness of the algorithm. Empirically, we showed that SS-NMF outperforms other well-established unsupervised and semi-supervised image clustering methods.

8. REFERENCES


