Non-negative Matrix Factorization for Multimodal Image Retrieval

Fabio A. González PhD

Bioingenium Research Group
Computer Systems and Industrial Engineering Department
Universidad Nacional de Colombia
Knowledge Discovery and Webmining Lab
CECS, The U of L

Graduate Seminar, Spring 2010
CECS Department, The University of Louisville
Outline

1. The Problem
   - Content-based image retrieval
   - Semantic image retrieval
   - Multimodal image retrieval

2. Non-Negative Matrix Factorization
   - The Netflix Prize
   - Non-negative matrix factorization
   - NMF vs SVD

3. NMF for Multimodal Retrieval
   - Semantic space
   - Multimodal clustering
   - Image annotation
   - Multimodal retrieval
Outline

1. The Problem
   - Content-based image retrieval
   - Semantic image retrieval
   - Multimodal image retrieval

2. Non-Negative Matrix Factorization
   - The Netflix Prize
   - Non-negative matrix factorization
   - NMF vs SVD

3. NMF for Multimodal Retrieval
   - Semantic space
   - Multimodal clustering
   - Image annotation
   - Multimodal retrieval
Outline

1 The Problem
   • Content-based image retrieval
     • Semantic image retrieval
     • Multimodal image retrieval

2 Non-Negative Matrix Factorization
   • The Netflix Prize
   • Non-negative matrix factorization
   • NMF vs SVD

3 NMF for Multimodal Retrieval
   • Semantic space
   • Multimodal clustering
   • Image annotation
   • Multimodal retrieval
Content-Based Image Retrieval
Low-level vs. High-level

Semantic Gap

High Level

Low Level

F. González (UNAL)
1 The Problem
   - Content-based image retrieval
   - Semantic image retrieval
   - Multimodal image retrieval

2 Non-Negative Matrix Factorization
   - The Netflix Prize
   - Non-negative matrix factorization
   - NMF vs SVD

3 NMF for Multimodal Retrieval
   - Semantic space
   - Multimodal clustering
   - Image annotation
   - Multimodal retrieval
Semantic Annotation using ML

Glandulas ecrinas
Anexos piloebaceos

Cambio quístico
Infiltrado epidermis
Nódulo, emplazada, hendidura
Recall vs Precision Graph - Semantic models

Histogram Intersection
MetaFeatures
Identity Kernel
Sobel

F. González (UNAL)
Disadvantages

- Requires a training set with expert annotations, so it is a costly process.
Disadvantages

- Requires a training set with expert annotations, so it is a costly process
- Does not scale for large semantic vocabularies
Disadvantages

- Requires a training set with expert annotations, so it is a costly process
- Does not scale for large semantic vocabularies
- The mapping from visual features to annotations may lose the visual richness
1 The Problem
   - Content-based image retrieval
   - Semantic image retrieval
   - Multimodal image retrieval

2 Non-Negative Matrix Factorization
   - The Netflix Prize
   - Non-negative matrix factorization
   - NMF vs SVD

3 NMF for Multimodal Retrieval
   - Semantic space
   - Multimodal clustering
   - Image annotation
   - Multimodal retrieval
Multimodality

Text and images come naturally together in many documents

- Academic papers, books
- Newspapers, web pages
- Medical cases
Multimodal Retrieval

- Unstructured text associated to images may be used as semantic annotations
  - Images and texts are complimentary information units
  - Take advantage of interactions between both data modalities
Multimodal Retrieval

- Unstructured text associated to images may be used as semantic annotations
  - Images and texts are complimentary information units
  - Take advantage of interactions between both data modalities

- Problems:
  - Text associated to images is not structured
  - Unclear relationships between keywords and visual patterns
  - Possible presence of noise in both data modalities
Multimodal Retrieval

- Unstructured text associated to images may be used as semantic annotations
  - Images and texts are complimentary information units
  - Take advantage of interactions between both data modalities

- Problems:
  - Text associated to images is not structured
  - Unclear relationships between keywords and visual patterns
  - Possible presence of noise in both data modalities

- Retrieval scenarios:
  - Cross-modal:
    - find images based on a text query
    - find text based on an image query (image annotation)
  - Visual retrieval based on a visual query
The Problem
- Content-based image retrieval
- Semantic image retrieval
- Multimodal image retrieval

Non-Negative Matrix Factorization
- The Netflix Prize
- Non-negative matrix factorization
- NMF vs SVD

NMF for Multimodal Retrieval
- Semantic space
- Multimodal clustering
- Image annotation
- Multimodal retrieval
1 The Problem
   - Content-based image retrieval
   - Semantic image retrieval
   - Multimodal image retrieval

2 Non-Negative Matrix Factorization
   - The Netflix Prize
   - Non-negative matrix factorization
   - NMF vs SVD

3 NMF for Multimodal Retrieval
   - Semantic space
   - Multimodal clustering
   - Image annotation
   - Multimodal retrieval
The Netflix Competition

- Problem: prediction of user ratings for films (collaborative filtering)

The Netflix Competition

- Problem: prediction of user ratings for films (collaborative filtering)
- Prize: $1 Million to the first algorithm able to improve Netflix own algorithm results by at least 10%

The Netflix Competition

- Problem: prediction of user ratings for films (collaborative filtering)
- Prize: $1 Million to the first algorithm able to improve Netflix own algorithm results by at least 10%
- Won on Sept/21/2009 by BellKor’s Pragmatic Chaos

The Netflix Competition

- Problem: prediction of user ratings for films (collaborative filtering)
- Prize: $1 Million to the first algorithm able to improve Netflix own algorithm results by at least 10%
- Won on Sept/21/2009 by BellKor’s Pragmatic Chaos
- Their approach used Non-negative Matrix Factorization (NMF) to build a latent-factor representation of users and movies

### The Latent-Factor Model

$$
\begin{align*}
\text{users} & \quad \begin{bmatrix}
  r_{11} & \cdots & r_{1m} \\
  \vdots & \ddots & \vdots \\
  r_{n1} & \cdots & r_{nm}
\end{bmatrix} = \quad \text{users} & \quad \begin{bmatrix}
  q_{11} & \cdots & q_{1f} \\
  \vdots & \ddots & \vdots \\
  q_{n1} & \cdots & q_{nf}
\end{bmatrix} \times \quad \text{movies} & \quad \begin{bmatrix}
  p_{11} & \cdots & p_{1m} \\
  \vdots & \ddots & \vdots \\
  p_{f1} & \cdots & p_{fm}
\end{bmatrix}
\end{align*}
$$

$$R \approx QP$$

$$Q, P \geq 0$$

- $n \approx 5 \times 10^5$
- $m \approx 1.7 \times 10^4$
- $|\{(i, j) | r_{ij} \neq 0\}| \approx 10^8$
- $f \leq 200$
Outline

1. The Problem
   - Content-based image retrieval
   - Semantic image retrieval
   - Multimodal image retrieval

2. Non-Negative Matrix Factorization
   - The Netflix Prize
   - Non-negative matrix factorization
   - NMF vs SVD

3. NMF for Multimodal Retrieval
   - Semantic space
   - Multimodal clustering
   - Image annotation
   - Multimodal retrieval
Non-negative Matrix Factorization

- Problem: to find a factorization

\[ X_{n \times m} = WH_{r \times m} \]

- Optimization problem:

\[
\min_{A, B} \quad \|X - WH\|^2 \\
\text{s.t.} \quad W, H \geq 0
\]

- \( \| \cdot \| \) is the Frobenius norm

- It is a non-convex optimization problem

- Solution alternatives:
  - Gradient descent methods
  - Multiplicative updating rules
Optimization problem:

\[
\min_{A,B} \quad \|X - WH\|^2 \\
\text{s.t.} \quad W, H \geq 0
\]

Incremental optimization:

\[
H_{a\mu} \leftarrow H_{a\mu} \frac{(W^T X)_{a\mu}}{(W^T WH)_{a\mu}} \\
W_{ia} \leftarrow W_{ia} \frac{(XH^T)_{\alpha\mu}}{(WHH^T)_{\alpha\mu}}
\]
Optimization problem:

\[
\min_{A,B} \quad D(X|WH) = \sum_{ij} \left( X_{ij} \log \frac{X_{ij}}{(WH)_{ij}} - X_{ij} + (WH)_{ij} \right)
\]

s.t. \quad W, H \geq 0

Multiplicative Rules:

\[
H_{a\mu} \leftarrow H_{a\mu} \frac{\sum_i W_{ia} X_{i\mu}/(WH)_{i\mu}}{\sum_i W_{ia}}
\]

\[
W_{ia} \leftarrow W_{ia} \frac{\sum_\mu H_{a\mu} X_{i\mu}/(WH)_{i\mu}}{\sum_\mu H_{a\mu}}
\]
Outline

1 The Problem
   - Content-based image retrieval
   - Semantic image retrieval
   - Multimodal image retrieval

2 Non-Negative Matrix Factorization
   - The Netflix Prize
   - Non-negative matrix factorization
   - NMF vs SVD

3 NMF for Multimodal Retrieval
   - Semantic space
   - Multimodal clustering
   - Image annotation
   - Multimodal retrieval
PCA and SVD

- Problem:
  \[ X_{n \times m} = W_{n \times r} H_{r \times m} \]

- Principal Component Analysis (PCA)
  \[ X = U \Sigma V \]
  \[ W = U \Sigma^{\frac{1}{2}}, H = \Sigma^{\frac{1}{2}} V \]

- PCA = SVD keeping the 'best' Eigenvectors
- Columns of U are orthonormal
- There is not restriction on sign.

Latent Semantic Indexing (LSI)

Documents are represented by the frequency of keywords (terms)
Uses SVD to find the factorization
Factors = semantic concepts
Columns of $W$ are orthonormal
Outline

1 The Problem
   - Content-based image retrieval
   - Semantic image retrieval
   - Multimodal image retrieval

2 Non-Negative Matrix Factorization
   - The Netflix Prize
   - Non-negative matrix factorization
   - NMF vs SVD

3 NMF for Multimodal Retrieval
   - Semantic space
   - Multimodal clustering
   - Image annotation
   - Multimodal retrieval
1 The Problem
   - Content-based image retrieval
   - Semantic image retrieval
   - Multimodal image retrieval

2 Non-Negative Matrix Factorization
   - The Netflix Prize
   - Non-negative matrix factorization
   - NMF vs SVD

3 NMF for Multimodal Retrieval
   - Semantic space
   - Multimodal clustering
   - Image annotation
   - Multimodal retrieval
Objects are described by terms in a textual vocabulary and a visual vocabulary.
• Objects are described by terms in a textual vocabulary and a visual vocabulary
• Objects are mapped to a latent (semantic) space where objects are represented by a set of latent factors
Objects are described by terms in a textual vocabulary and a visual vocabulary.

Objects are mapped to a latent (semantic) space where objects are represented by a set of latent factors.

NMF is used to build the latent representation.
Objects are described by terms in a textual vocabulary and a visual vocabulary.

Objects are mapped to a latent (semantic) space where objects are represented by a set of latent factors.

NMF is used to build the latent representation.

Three main tasks:

- Multimodal clustering
- Automatic image annotation
- Image retrieval
Semantic Space: Multimodal Latent Indexing (II)

\[ \mathbf{X}_{n \times m} = \mathbf{W}_{n \times r} \ast \mathbf{H}_{r \times m} \]

- **Document Space**
- **Latent Space**

- **Text terms**
- **Visual terms**
Bag-of-Features Image Representation

(i) Feature detection and description

(ii) Codebook construction

(iii) Bag of features representation
1 The Problem
- Content-based image retrieval
- Semantic image retrieval
- Multimodal image retrieval

2 Non-Negative Matrix Factorization
- The Netflix Prize
- Non-negative matrix factorization
- NMF vs SVD

3 NMF for Multimodal Retrieval
- Semantic space
- Multimodal clustering
- Image annotation
- Multimodal retrieval
Multimodal Clustering

\[ X_{n \times m} = W_{n \times r} \ast H_{r \times m} \]

- Document Space
- Latent Space

visual terms

text terms

F. González (UNAL)

CECS Grad Sem 34 / 49
Dual Multimodal Clustering

\[ X_{n \times m} = W_{n \times r} \ast H_{r \times m} \]

- **Document Space**
- **Latent Space**
- **Term Space**

Visual terms vs. text terms
Does It Work?

- Clustering Corel data set: 1000 images, 25 categories
- Input matrix: 2500 vectors of dimension 1000
- Clustering performance comparison: $K$-means and two NMF variants:

<table>
<thead>
<tr>
<th></th>
<th>$K$-means</th>
<th>NMF (Frobenius norm)</th>
<th>NMF (KL divergence)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.2288</td>
<td>0.2768</td>
<td>0.2905</td>
</tr>
</tbody>
</table>

*results are average of 10 runs
Outline

1. The Problem
   - Content-based image retrieval
   - Semantic image retrieval
   - Multimodal image retrieval

2. Non-Negative Matrix Factorization
   - The Netflix Prize
   - Non-negative matrix factorization
   - NMF vs SVD

3. NMF for Multimodal Retrieval
   - Semantic space
   - Multimodal clustering
   - Image annotation
   - Multimodal retrieval
1. Apply NMF to training data

2. Find latent representation $h$ of a visual vector $x$, $x = W1 * h$

3. Multiply $h$ by $W$ to get the multimodal vector $[x, y]$
Does it Work?

- pLSA has shown good performance for image annotation

| TABLE 3 |
|------------------|-----------------|-----------------|-----------------|-----------------|
| mAP Values (in Percent) for the Six Methods |
| When Combinations of HS and SIFT Features Are Used |

<table>
<thead>
<tr>
<th>Method</th>
<th>Blobs</th>
<th>HS</th>
<th>SIFT</th>
<th>HS+SIFT</th>
</tr>
</thead>
<tbody>
<tr>
<td>propagation</td>
<td>7.8 (0.7)</td>
<td>9.0 (0.2)</td>
<td>9.4 (1.0)</td>
<td>13.1 (0.5)</td>
</tr>
<tr>
<td>CMRM [15]</td>
<td>11.5 (1.1)</td>
<td>10.7 (1.1)</td>
<td>7.9 (0.5)</td>
<td>13.4 (1.0)</td>
</tr>
<tr>
<td>SVD-COS [26]</td>
<td>12.9 (1.1)</td>
<td>12.9 (0.8)</td>
<td>10.7 (0.7)</td>
<td>16.6 (1.1)</td>
</tr>
<tr>
<td>PLSA-MIXED</td>
<td>5.8 (0.8)</td>
<td>10.2 (0.8)</td>
<td>7.5 (0.6)</td>
<td>11.9 (1.3)</td>
</tr>
<tr>
<td>PLSA-FEATURES</td>
<td>8.2 (0.7)</td>
<td>11.2 (1.0)</td>
<td>10.1 (0.8)</td>
<td>14.0 (1.3)</td>
</tr>
<tr>
<td>PLSA-WORDS</td>
<td>11.0 (0.9)</td>
<td>13.3 (1.0)</td>
<td>11.8 (1.1)</td>
<td>19.1 (1.2)</td>
</tr>
</tbody>
</table>

- NMF with KL divergence was shown to be equivalent to pLSA

Outline

1. The Problem
   - Content-based image retrieval
   - Semantic image retrieval
   - Multimodal image retrieval

2. Non-Negative Matrix Factorization
   - The Netflix Prize
   - Non-negative matrix factorization
   - NMF vs SVD

3. NMF for Multimodal Retrieval
   - Semantic space
   - Multimodal clustering
   - Image annotation
   - Multimodal retrieval
Scenario: visual retrieval based on a visual query
Image Retrieval

- Scenario: visual retrieval based on a visual query
- Images are represented in a latent space using NMF
Scenario: visual retrieval based on a visual query
Images are represented in a latent space using NMF
Some images in the database may not have text content associated
Scenario: visual retrieval based on a visual query
Images are represented in a latent space using NMF
Some images in the database may not have text content associated
The image query is projected the latent space
Scenario: visual retrieval based on a visual query
Images are represented in a latent space using NMF
Some images in the database may not have text content associated
The image query is projected the latent space
Image are retrieved according to their latent space similarity
Corel Images

- A subset of 2,500 images extracted from the Corel Database.
- 25 categories with 100 images each.
Bag of Features

Corel Images

- Image content representation using parts of images
- Blocks are extracted from each image and the SIFT descriptor is computed
- A dictionary of visual patterns is built using k-means
- A histogram counting the occurrence of each codeblock is constructed for each image
This approach outperforms direct image matching by almost 4X
## Some specific performance measures

<table>
<thead>
<tr>
<th>Measure</th>
<th>NMF Search</th>
<th>P1</th>
<th>Direct Search</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>0.4400</td>
<td></td>
<td>0.5280</td>
</tr>
<tr>
<td>P20</td>
<td>0.4508</td>
<td></td>
<td>0.3820</td>
</tr>
<tr>
<td>P50</td>
<td>0.4474</td>
<td></td>
<td>0.3160</td>
</tr>
<tr>
<td>R10</td>
<td>0.0471</td>
<td></td>
<td>0.0446</td>
</tr>
<tr>
<td>R20</td>
<td>0.0949</td>
<td></td>
<td>0.0804</td>
</tr>
<tr>
<td>R50</td>
<td>0.2355</td>
<td></td>
<td>0.1663</td>
</tr>
<tr>
<td>MAP</td>
<td>0.4825</td>
<td></td>
<td>0.1342</td>
</tr>
</tbody>
</table>
Semantic Space Visualization (I)

NMF for MM IR
Ongoing Work

- Bigger data sets: ImageClefMed, Corel5000, Flickr
- Different image low-level features
- Kernel NMF
- Incremental NMF
Thanks!

fagonzalezo@unal.edu.co

http://dis.unal.edu.co/~fgonza