Non-negative Matrix Factorization for Multimodal Image Retrieval

Fabio A. González PhD

Machine Learning 2015-II
Universidad Nacional de Colombia

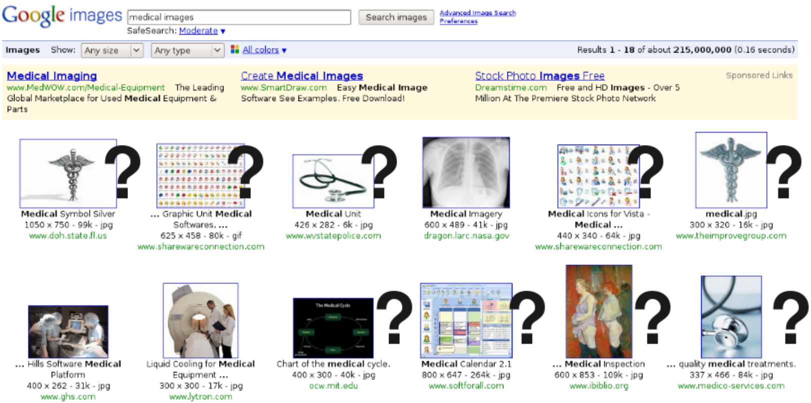
Outline

1. The Problem
   - Content-based image retrieval
   - Semantic image retrieval
   - Multimodal image retrieval
2. Matrix Factorization
   - The Netflix Prize
   - Non-negative matrix factorization
   - NMF vs SVD
3. NMF for Multimodal Learning
   - Semantic space
   - Multimodal clustering
   - Image annotation
   - Multimodal retrieval
   - Application to histology images

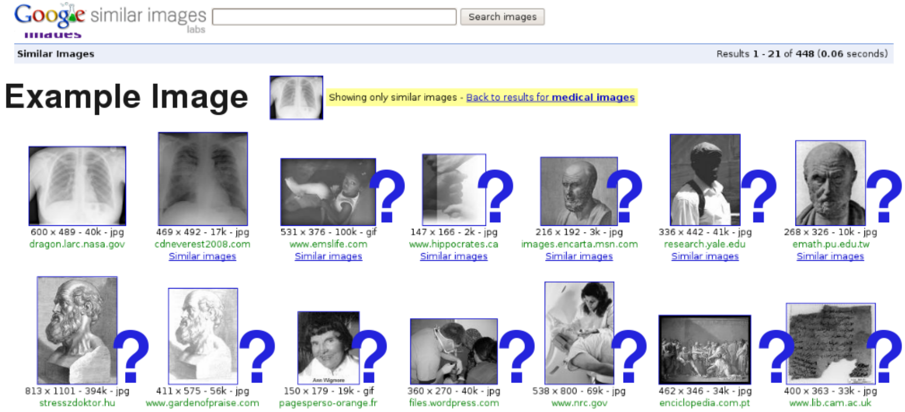
Outline

Outline

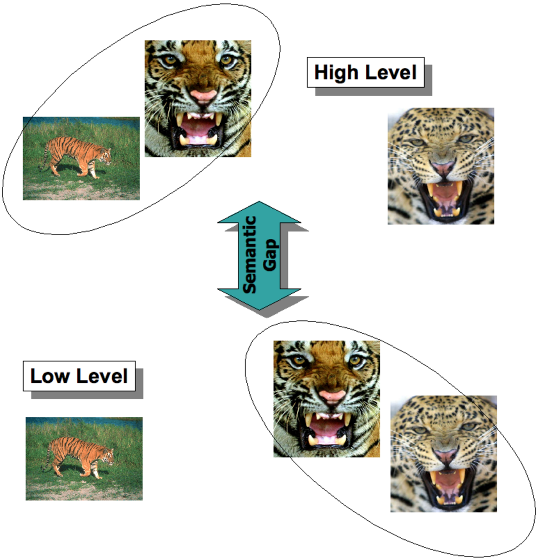
1. The Problem
   - Content-based image retrieval
   - Semantic image retrieval
   - Multimodal image retrieval
2. Matrix Factorization
   - The Netflix Prize
   - Non-negative matrix factorization
   - NMF vs SVD
3. NMF for Multimodal Learning
   - Semantic space
   - Multimodal clustering
   - Image annotation
   - Multimodal retrieval
   - Application to histology images


Google images
medical images
Search images
Advanced Image Search
Preferences
SafeSearch: Moderate
Images Show: Any size Any type All colors
Results 1 - 18 of about 215,000,000 (0.16 seconds)
Medical Imaging
www.MedWOW.com/Medical-Equipment The Leading Global Marketplace for Used Medical Equipment & Parts
Create Medical Images
www.SmartDraw.com Easy Medical Image Software See Examples. Free Download!
Stock Photo Images Free
Dreamstime.com Free and HD Images - Over 5 Million At The Premiere Stock Photo Network
Sponsored Links
Medical Symbol Silver
1050 x 750 - 99k - jpg
www.doh.state.fl.us
... Graphic Unit Medical Softwares, ...
625 x 458 - 80k - gif
www.sharewareconnection.com
Medical Unit
426 x 282 - 6k - jpg
www.wvstatepolice.com
Medical Imagery
600 x 489 - 41k - jpg
dragon.larc.nasa.gov
Medical Icons for Vista - Medical ...
440 x 340 - 64k - jpg
www.sharewareconnection.com
medical.jpg
300 x 320 - 16k - jpg
www.theimprovegroup.com
... Hills Software Medical Platform
400 x 262 - 31k - jpg
www.ghs.com
Liquid Cooling for Medical Equipment ...
300 x 300 - 17k - jpg
www.lytron.com
Chart of the medical cycle.
400 x 300 - 40k - jpg
ocw.mit.edu
Medical Calendar 2.1
800 x 647 - 264k - jpg
www.softforall.com
... Medical Inspection
600 x 853 - 109k - jpg
www.ibiblio.org
... quality medical treatments.
337 x 466 - 84k - jpg
www.medico-services.com
?

Query by Visual Example



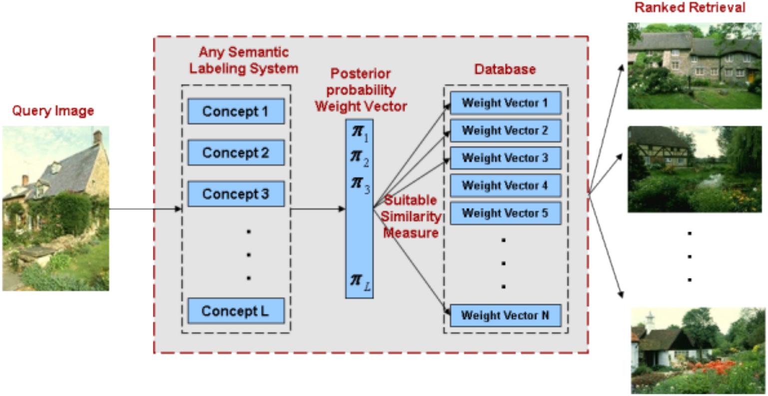
Low-level vs. High-level



Outline

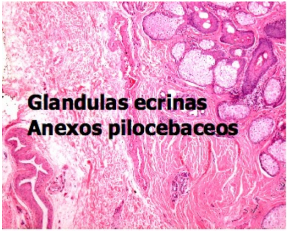
1. The Problem
   - Content-based image retrieval
   - Semantic image retrieval
   - Multimodal image retrieval
2. Matrix Factorization
   - The Netflix Prize
   - Non-negative matrix factorization
   - NMF vs SVD
3. NMF for Multimodal Learning
   - Semantic space
   - Multimodal clustering
   - Image annotation
   - Multimodal retrieval
   - Application to histology images

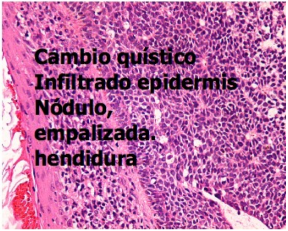
Semantic Annotation using ML

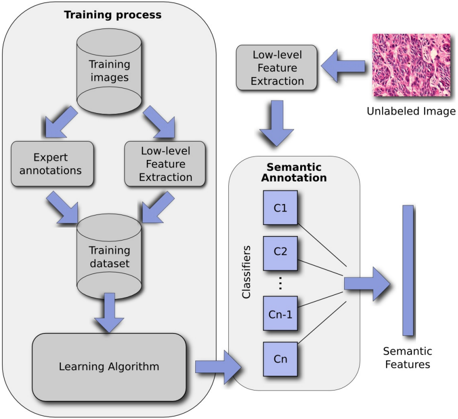


Source: Nuno Vasconcelos, UCSD, http://www.svcl.ucsd.edu/projects/qbse/

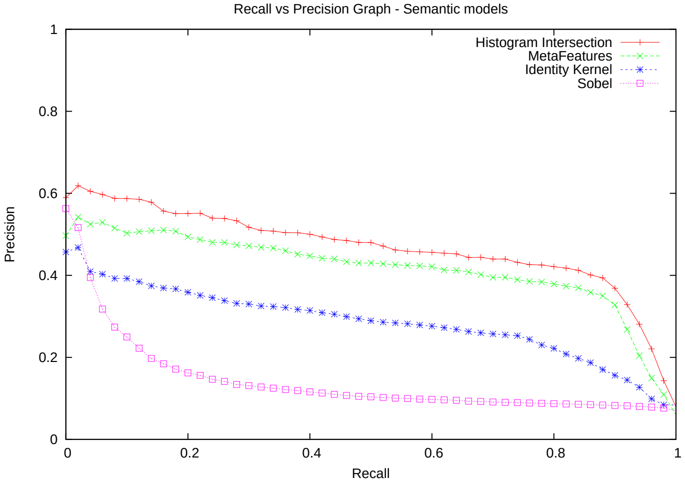
An Example (1)


Glandulas ecrinas
Anexos pilocebaceos


Cambio quístico
Infiltrado epidermis
Nódulo,
empalizada,
hendidura


Training process
Training images
Expert annotations
Low-level Feature Extraction
Training dataset
Learning Algorithm
Low-level Feature Extraction
Unlabeled Image
Semantic Annotation
Classifiers
C1
C2
Cn-1
Cn
Semantic Features

An Example (2)



Recall vs Precision Graph - Semantic models
Histogram Intersection
MetaFeatures
Identity Kernel
Sobel
Precision
Recall
0
0.2
0.4
0.6
0.8
1

Disadvantages

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

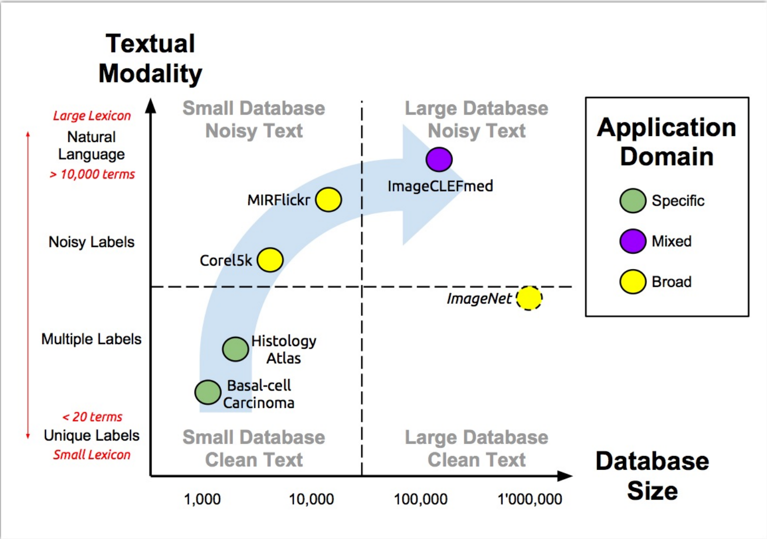
Disadvantages

- 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

Scalability



Outline

1. The Problem
   - Content-based image retrieval
   - Semantic image retrieval
   - Multimodal image retrieval
2. Matrix Factorization
   - The Netflix Prize
   - Non-negative matrix factorization
   - NMF vs SVD
3. NMF for Multimodal Learning
   - Semantic space
   - Multimodal clustering
   - Image annotation
   - Multimodal retrieval
   - Application to histology images

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

Outline

Outline

The Netflix Competition

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

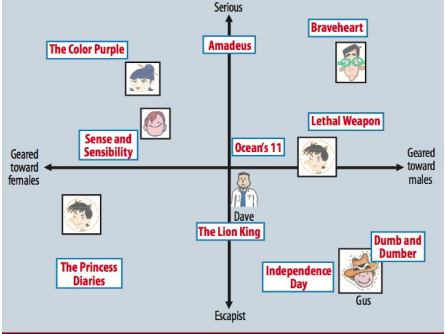


Figure 2. A simplified illustration of the latent factor approach, which characterizes both users and movies using two axes—male versus female and serious versus escapist.

Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," Computer, vol. 42, no. 8, pp. 30-37, August 2009

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%

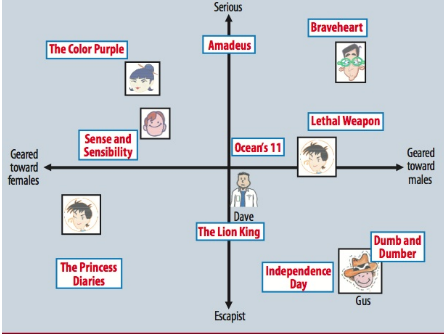


Figure 2. A simplified illustration of the latent factor approach, which characterizes both users and movies using two axes—male versus female and serious versus escapist.

Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," Computer, vol. 42, no. 8, pp. 30-37, August 2009

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

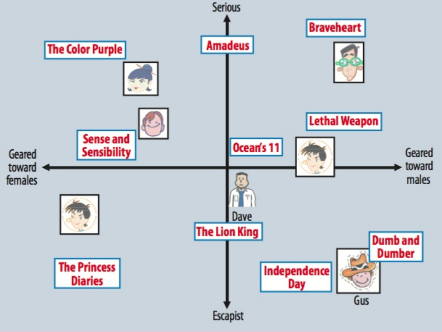


Figure 2. A simplified illustration of the latent factor approach, which characterizes both users and movies using two axes—male versus female and serious versus escapist.

Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," Computer, vol. 42, no. 8, pp. 30-37, August 2009

- 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 Matrix Factorization to build a latent-factor representation of users and movies

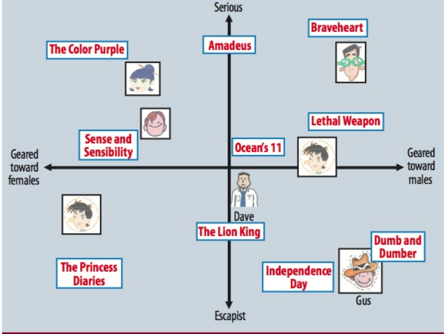


Figure 2. A simplified illustration of the latent factor approach, which characterizes both users and movies using two axes—male versus female and serious versus escapist.

Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," Computer, vol. 42, no. 8, pp. 30-37, August 2009

The Latent-Factor Model

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

$$R \approx QP$$

$$Q, P \geq 0$$

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

Outline

Non-negative Matrix Factorization

- Problem: to find a factorization

$$X_{n\times m} = W_{n\times r} H_{r\times m}$$

- Optimization problem:

$$\begin{array}{cc} \min_{A,B} & ||X - WH||^2 \\ \text{s.t.} & W, H \geq 0 \end{array}$$

- $|| \cdot ||$ is the Frobenius norm
- It is a non-convex optimization problem
- Solution alternatives:
  - Gradient descendent methods
  - Multiplicative updating rules

Multiplicative Rules

- Optimization problem:

$$\begin{array}{cc} \min_{W,H} & ||X - WH||^2 \\ \text{s.t.} & W, H \geq 0 \end{array}$$

- Incremental optimization:

$$H_{a\mu} \leftarrow H_{a\mu} \frac{(W^T X)_{a\mu}}{(W^T W H)_{a\mu}}$$

$$W_{ia} \leftarrow W_{ia} \frac{(X H^T)_{\alpha\mu}}{(W H H^T)_{\alpha\mu}}$$

Divergence Optimization

- Optimization problem:

$$
\begin{array}{ll}
\min_{W,H} & D(X|WH) = \sum_{ij} \left(X_{ij} \log \frac{X_{ij}}{(WH)_{ij}} - X_{ij} + (WH)_{ij} \right) \\
\text{s.t.} & W, H \geq 0
\end{array}
$$

- 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

PCA and SVD

- Problem:

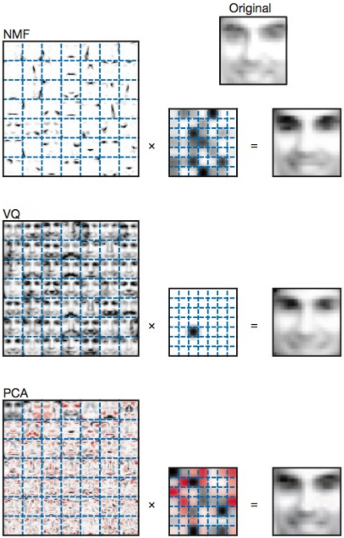
$$X_{n\times m} = W_{n\times r}H_{r\times m}$$

- 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.



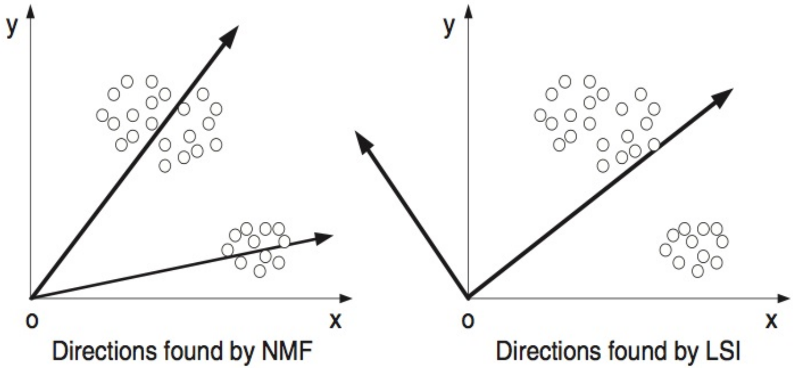
D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," Nature, vol. 401, no. 6755, pp. 788-791, October 1999

Latent Semantic Indexing (LSI)

$$\begin{array}{cc} & \text{documents} \\ \text{terms} & \underbrace{\begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{bmatrix}}_{X} \end{array} = \begin{array}{cc} & \text{factors} \\ \text{terms} & \underbrace{\begin{bmatrix} w_{11} & \cdots & w_{1r} \\ \vdots & \ddots & \vdots \\ w_{n1} & \cdots & w_{nr} \end{bmatrix}}_{W} \end{array} \times \begin{array}{cc} & \text{documents} \\ \text{factors} & \underbrace{\begin{bmatrix} h_{11} & \cdots & h_{1m} \\ \vdots & \ddots & \vdots \\ h_{r1} & \cdots & h_{rm} \end{bmatrix}}_{H} \end{array}$$

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

NMF vs LSI



W. Xu, X. Liu, and Y. Gong, "Document clustering based on non-negative matrix factorization," in SIGIR '03: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval. New York, NY, USA: ACM, 2003, pp. 267-273

Outline

Neurocomputing 76 (2012) 50–60

Contents lists available at SciVerse ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Multimodal representation, indexing, automated annotation and retrieval of image collections via non-negative matrix factorization

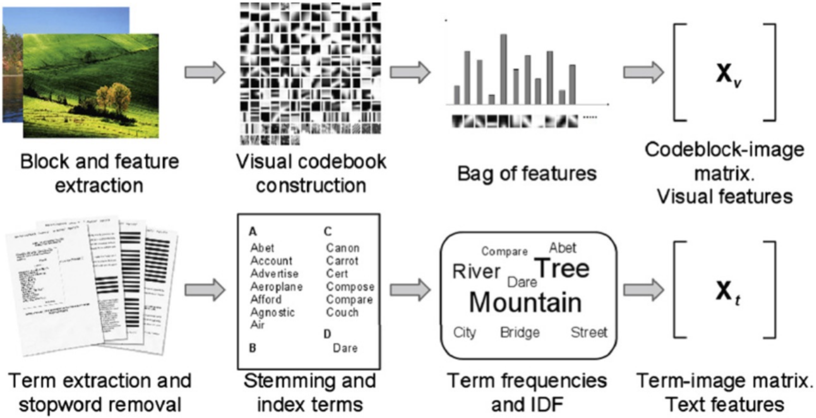
Juan C. Caicedo [a], Jaafar BenAbdallah [b], Fabio A. González [a,*], Olfa Nasraoui [b]

[a] Computer Systems and Industrial Engineering Department, National University of Colombia, Cra 30 45 - 03, Ciudad Universitaria, Edif. 453, Of. 114. Bogotá, Colombia
[b] Department of Computer Engineering and Computer Science, University of Louisville, Louisville KY, USA

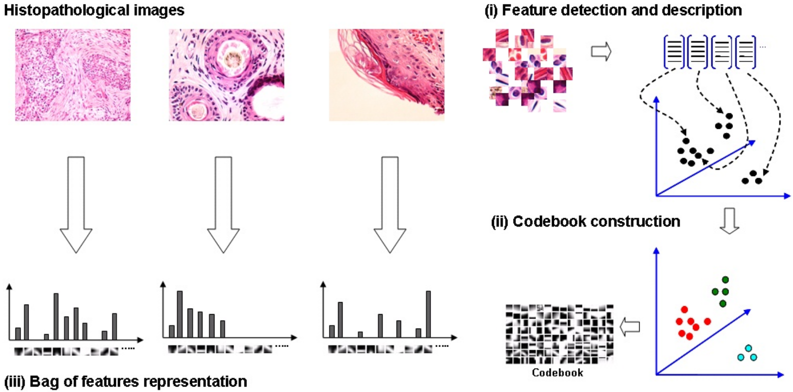
Outline

1. The Problem
   - Content-based image retrieval
   - Semantic image retrieval
   - Multimodal image retrieval
2. Matrix Factorization
   - The Netflix Prize
   - Non-negative matrix factorization
   - NMF vs SVD
3. NMF for Multimodal Learning
   - Semantic space
   - Multimodal clustering
   - Image annotation
   - Multimodal retrieval
   - Application to histology images

Multimodal Representation



Bag-of-Features Image Representation



Semantic Space: Multimodal Latent Indexing (I)

- Objects are described by terms in a textual vocabulary and a visual vocabulary

Semantic Space: Multimodal Latent Indexing (I)

- 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

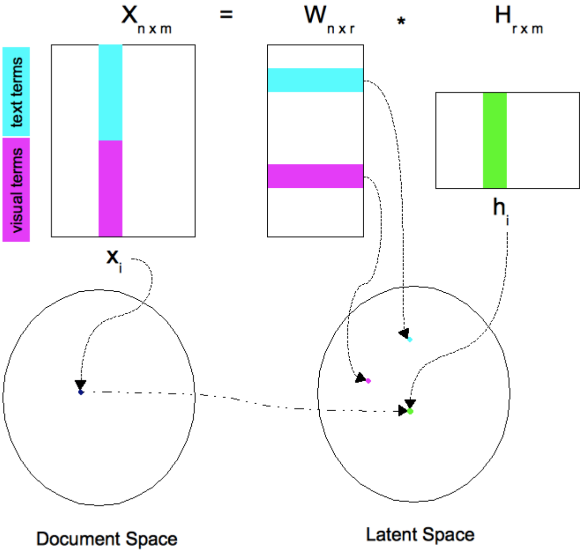
Semantic Space: Multimodal Latent Indexing (I)

- 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

Semantic Space: Multimodal Latent Indexing (I)

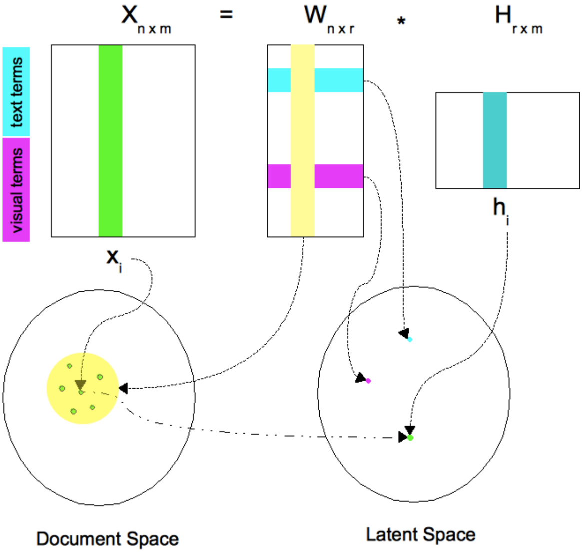
- Objects are described by terms in a textual vocabulary and a visual vocabulary
- Objects are mapped to a latent (semantic) space where objects are represented by a set of latent factors
- NMF is used to build the latent representation
- Three main tasks:
  - Multimodal clustering
  - Automatic image annotation
  - Image retrieval

Semantic Space: Multimodal Latent Indexing (II)

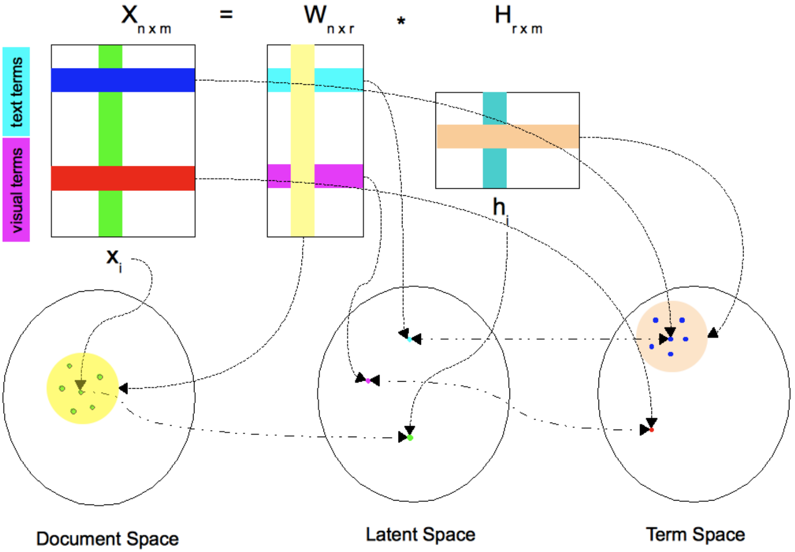


Outline

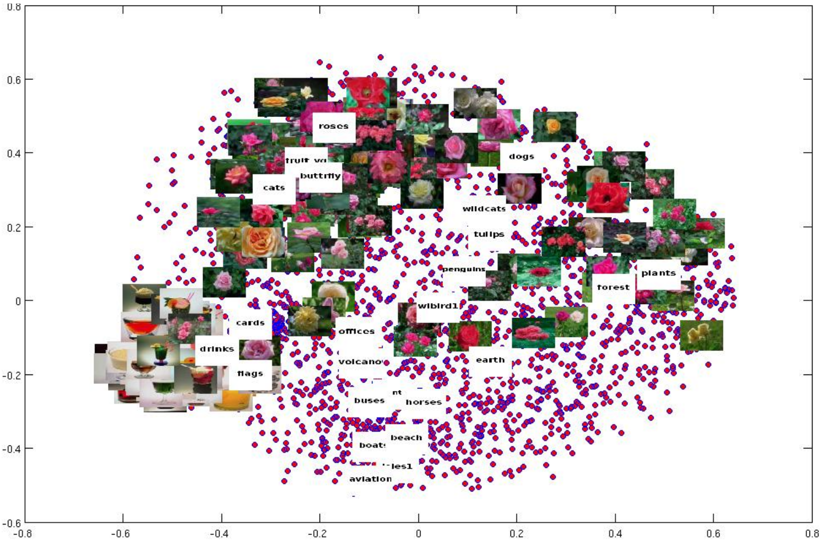
Multimodal Clustering



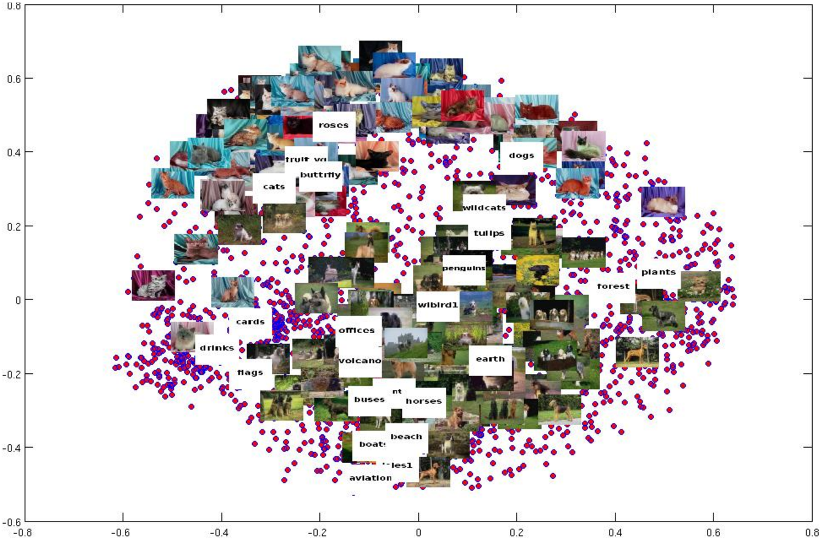
Dual Multimodal Clustering



Semantic Space Visualization (I)



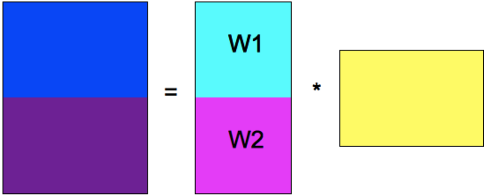
Semantic Space Visualization (II)



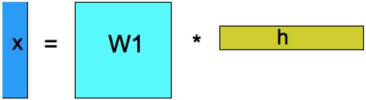
Outline

Image Annotation

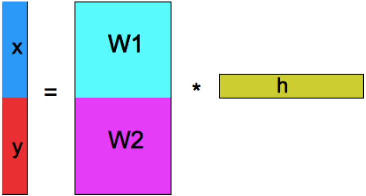
1. Apply NMF to training data



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]*



Annotation Results

Collection	Query	Visual	Multimodal	Ground Truth
Corel 5k		water **sky** **plane** tree	**plane** **jet** clouds **sky**	sky plane jet
Corel 5k		**water** tree people grass	**buildings** **water** people sun	water sky buildings
MIRFlickr		male **people** **structures** people_r1 male_r1 **female** sky transport animals car	**plant_life** **flower** **flower_r1** indoor sky **structures** **people** animals **female** female_r1	female flower flower_r1 people plant_life structures
MIRFlickr		**plant_life** structures animals **tree** flower transport **water** male people dog	**plant_life** **tree** **sky** tree_r1 structures **water** **river** **clouds** lake sea	clouds plant_life river river_r1 sky tree water

Outline

Image 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

Image Retrieval

- 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

Image Retrieval

- 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

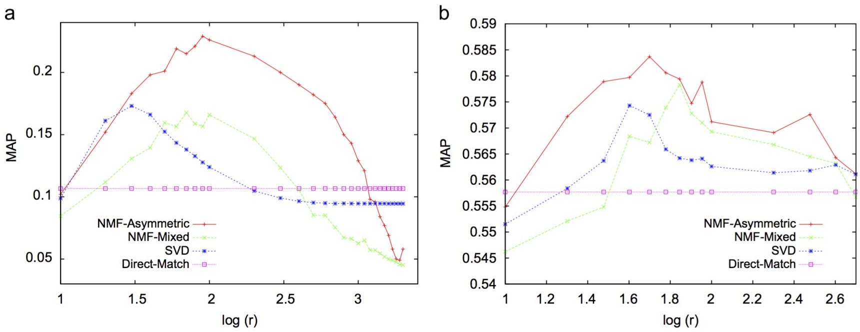
Image Retrieval

- 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

Retrieval Performance

Model	Corel 5k		MIRFlickr	
	MAP	**Gain** (%)	**MAP**	**Gain** (%)
Direct matching	0.1071	N/A	0.5577	N/A
SVD mixed	0.1780	66.2	0.5743	2.98
NMF mixed	0.1727	61.2	0.5783	3.67
NMF asymmetric	0.2369	121.2	0.5837	4.67

Semantic Space Dimension



Outline

Histology Image Retrieval

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

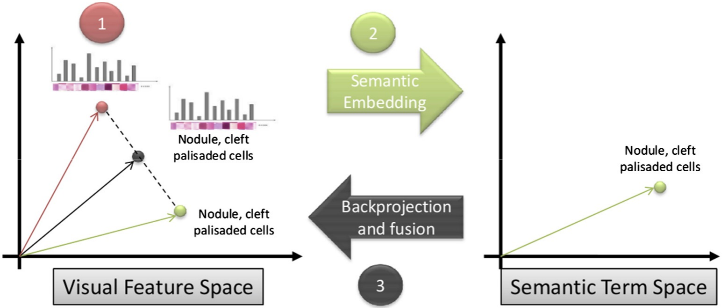
journal homepage: www.elsevier.com/locate/yjbin

Histology image search using multimodal fusion

Juan C. Caicedo [a,*,1], Jorge A. Vanegas [b], Fabian Páez [b], Fabio A. González [b]

[a] *University of Illinois at Urbana–Champaign, Sieble Center for Computer Science, 201 N Goodwin Ave, Urbana, IL 61801, USA*
[b] *MindLab Research Laboratory, Universidad Nacional de Colombia, Bogotá, Colombia*

Semantic Back Projection

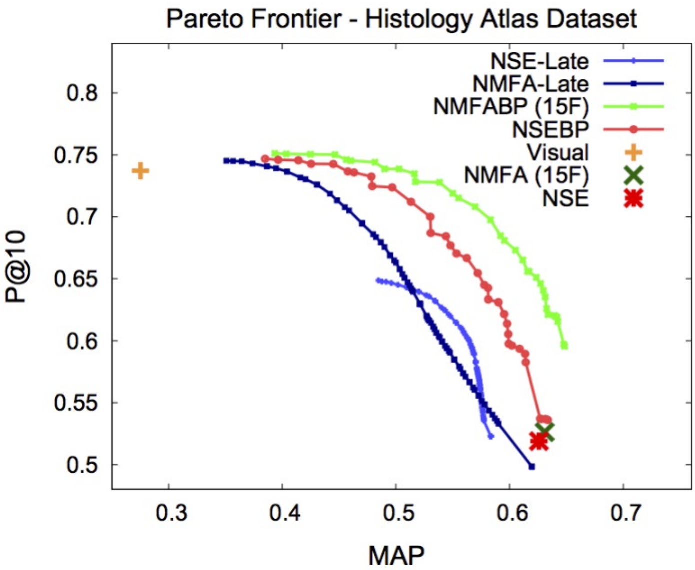


Performance

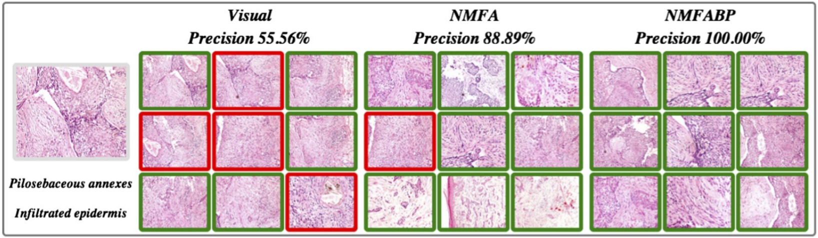
Table 2
Retrieval performance of semantic strategies compared to visual matching. Numbers in bold indicate the best result obtained in each data set with respect to P@10 or MAP. Results show the trade-off between MAP and P@10 on all image collections. Semantic search produces superior MAP while visual search is a strong baseline for early precision. Chance performance refers to the expected performance of a random ranking strategy, and it is significantly lower than visual and semantic search.

Method	Cervical Cancer		Basal-cell C.		Histology Atlas	
	P@10	MAP	P@10	MAP	P@10	MAP
Visual matching (baseline)	**0.5904**	0.5214	**0.4360**	0.2928	**0.7372**	0.2751
Latent embedding (NMFA)	0.5067	0.6591	0.3176	**0.4947**	0.5263	**0.6309**
Direct embedding (NSE)	0.5414	**0.6970**	0.2543	0.4317	0.5230	0.6113
Chance performance	0.4623	0.4681	0.2183	0.1806	0.0978	0.1008

Performance



Retrieval Performance Example



Thanks!

fagonzalezo@unal.edu.co

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