Mining query logs with topic models

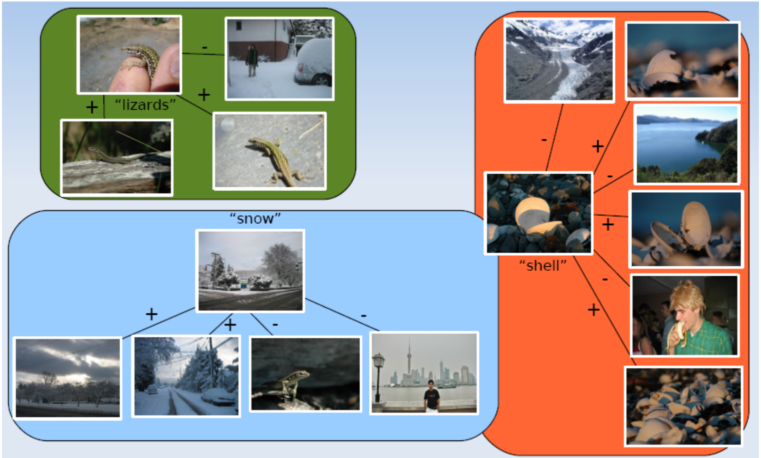
Donn Morrison

Computer Vision and Multimedia Laboratory, University of Geneva

August 31, 2009

Objectives

- Modelling user interaction (relevance feedback)
- Improve image retrieval through indexing
- Incremental image annotation



Nature of the data (relevance feedback)

- Query-by-example paradigm
- User refines query by marking positive (+1) and negative (-1) examples from results
- Query is refined until the search terminates (successfully or not)

At any time, we have a collection of M images and N queries. The collection of relevance judgements can be represented as a matrix $\mathcal{R}$ of co-occurrences:

$$
\begin{array}{c c}
 & \text{Sessions} \\
\text{Images} & \left[\begin{array}{c|cccc}
 & q_1 & q_2 & \dots & q_N \\
\hline
d_1 & 1 & -1 & \dots & 1 \\
d_2 & -1 & 0 & \dots & -1 \\
d_3 & 1 & -1 & \dots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
d_M & 1 & 0 & \dots & -1
\end{array}\right]
\end{array}
$$

Topic modelling

- Goal: explain observed co-occurrences by estimating linear combinations of hidden factors
- Text modelling: Underlying topics are said to generate word observations in text documents
- In our case, the hidden factors are the users' intent during search as well as concepts or objects expressed in the images of the collection

Non-negative matrix factorisation (NMF)

Seek an approximation $\mathcal{R} \approx WH$ such that the Frobenius norm $||\mathcal{R} - WH||_F$ is minimised. We iterate update steps:

$$H_{cj} \leftarrow H_{cj} \frac{\sum_i \frac{W_{ic}\mathcal{R}_{ij}}{(WH)_{ij}}}{\sum_i W_{ic}} \tag{1}$$

$$W_{ic} \leftarrow W_{ic} \frac{\sum_j \frac{H_{cj}\mathcal{R}_{ij}}{(WH)_{ij}}}{\sum_j H_{cj}} \tag{2}$$

where W is the image-topic matrix and H is the topic-query matrix (Lee and Seung, 1999).

NMF constrains values in the co-occurrence matrix to be $>= 0$, so we scale our RF data ($\mathcal{R}_{ij}$) into this range:

$$\begin{aligned} -1 &\rightarrow 0 \\ 0 &\rightarrow 0.5 \\ 1 &\rightarrow 1 \end{aligned}$$

which can be loosely interpreted as the probability of an image d_i being relevant to a query q_j.

Singular value decomposition (SVD)

Any matrix $\mathcal{R}$ can be rewritten in the form:

$$\mathcal{R} = U\Sigma V^T \tag{3}$$

where U are the left singular vectors, Σ are the singular values (square roots of the eigenvalues), and V^T are the right singular vectors.

A rank-k approximation to $\mathcal{R}$ can be achieved by retaining the k largest singular values in Σ:

$$\mathcal{R}_k = U_k \Sigma_k V_k^T \tag{4}$$

Orthonormality constraint:

$$\begin{aligned} U^T U = V^T V = I \\ ||U|| = ||V|| = 1 \end{aligned} \tag{5}$$

User Relevance Model (URM)

Extension of probabilistic latent semantic analysis (PLSA) (Hofmann, 1999)

- Generative, probabilistic model
- Documents and queries assumed to be generated from the same *concept-space*

Relevance judgement generation:

- generate a query with probability $P(q)$
- select latent concept with probability $P(c|q)$
- select a document with probability $P(d)$
- generate a relevance judgement $P(r|d, c)$

User Relevance Model (URM)

Joint probability (co-occurrence observations) $P(r, d, q)$ is defined as:

$$\mathcal{R} = P(r, d, q) = P(q)P(d)P(r|d, q), \tag{6}$$

where

$$P(r|d, q) = \sum_{c \in \mathcal{C}} P(r|d, c)P(c|q). \tag{7}$$

Following Bayes rule, we can rewrite the joint probability as:

$$P(r, d, q) = \sum_{c \in \mathcal{C}} P(c)P(q|c)P(d)P(r|d, c). \tag{8}$$

"Fit" of latent variables to observed data measured using log-likelihood:

$$\mathcal{L} = \sum_{d \in \mathcal{D}} \sum_{q \in \mathcal{Q}} \sum_{r} n(r, d, q) log P(r, d, q). \tag{9}$$

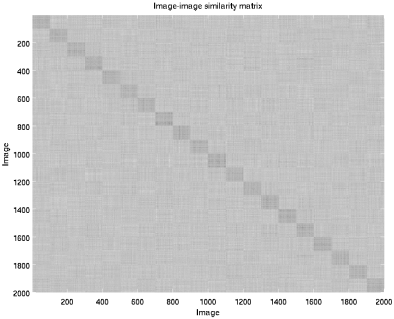
Expectation-maximisation used to converge on a maximum $\mathcal{L}$.

Document similarity with topic models

- Models lend themselves to item/attribute similarity
- We can use these similarity graphs to propagate meta-data and index images

Image similarity using dot product:

- NMF: WW^T
- SVD: UU^T
- URM: $P(r|d,c)P(r|d,c)^T$



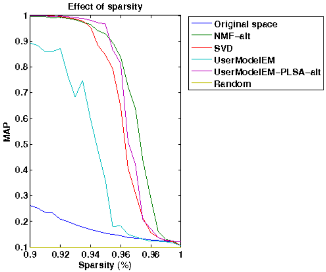
Experiments

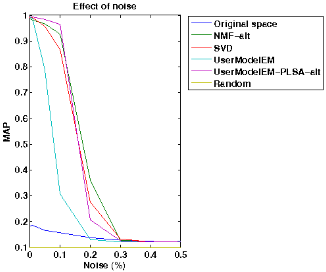
Corel image collection, 1000 images, 10 categories, 100 images per category, 3-5 annotations per image.

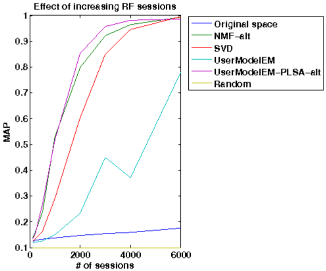
- Sparsity 95%
- Noise 10%
- 3000 artificial query sessions
- 10 latent variables

Image similarity experiments

Accuracy measured using *mean average precision*: each image used as a query; ranked list of most similar images yields a score closer to 1 the more the relevant images are ranked first (an indication of images clustered over latent topics)







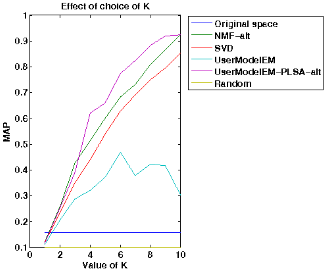


Image annotation experiments

For each unannotated image, rank top-I similar images and select w tags from the pool of W total tags.

Formally:

We repeat a draw $t_{1..w} \sim \mathcal{U}[1, W]$ (without replacement) for each unannotated image where w equals the desired number of annotations ($w = 4$).

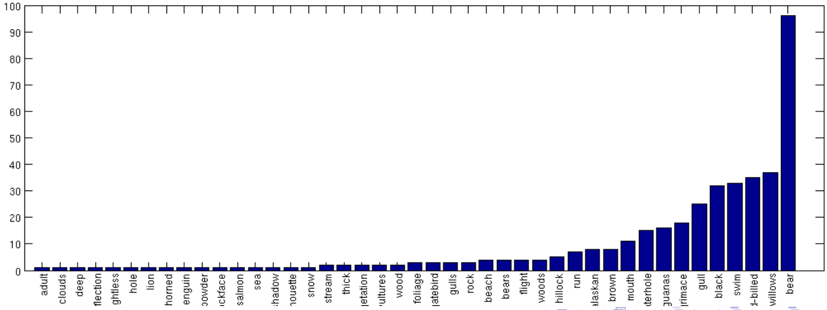
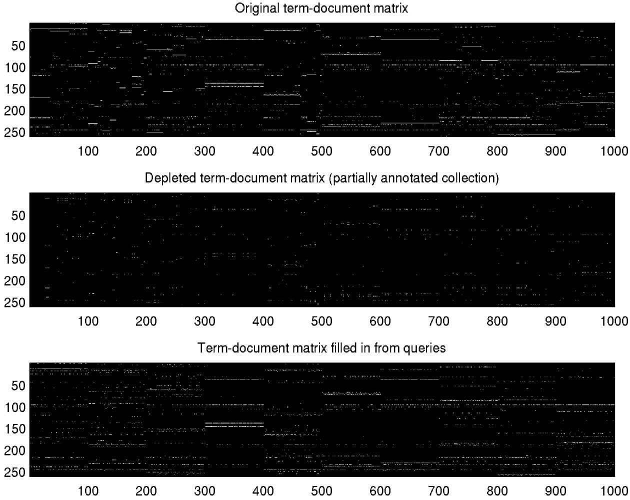


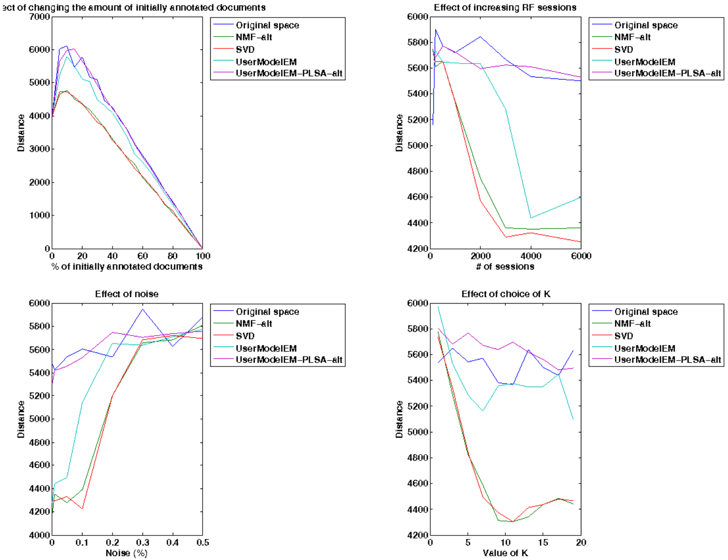
Image annotation experiments



Original vocab size: 253; depleted vocab size: 153; unannotated images: 2

Image annotation experiments

Accuracy measure: Euclidean distance between term-document matrices



Example annotations

(deer, grass, water, white-tailed)

forest, snow, trees, wolf
(grass, shade, trees, wolf)

bear, river, snow
(bear, grizzly, stream, water)

head, lion, mane, rocks
(cats, field, grass, lions)

dust, elephant, sky, water
(bull, elephant, sky, water)

grass, hippo, pair, river
(grass, hippos, wallow, water)

Conclusions

Conclusions

- Introduced a probabilistic User Relevance Model
- Recovery of underlying concepts from documents possible under sparse conditions
- Application to retrieval and image annotation

Future work

- Images with no tags can be brought to the attention of the user in order to ellicit interaction
- Tag quality could be improved by supplementing the RF judgements with low-level feature information (pseudo-relevance feedback)

Thank you

Questions

Thomas Hofmann. Probabilistic latent semantic analysis. In *Proc. of Uncertainty in Artificial Intelligence, UAI'99*, Stockholm, 1999. URL `citeseer.ist.psu.edu/hofmann99probabilistic.html`.

D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755): 788–791, October 1999. ISSN 0028-0836. doi: http://dx.doi.org/10.1038/44565. URL `http://dx.doi.org/10.1038/44565`.

Expectation-maximisation for URM

$$\mathcal{L} = \sum_{d \in \mathcal{D}} \sum_{q \in \mathcal{Q}} \sum_{r} n(r,d,q) log P(r,d,q), \quad n(r,d,q) \in \{0,1\} \tag{10}$$

E-step:

$$P(c|r,d,q) = \frac{P(c)P(q|c)P(r|d,c)}{\sum_{c \in \mathcal{C}} P(c)P(q|c)P(r|d,c)}, \tag{11}$$

M-step:

$$P(q|c) \propto \sum_{d \in \mathcal{D}} \sum_{r} n(r,d,q) P(c|r,d,q), \tag{12}$$

$$P(r|d,c) \propto \sum_{q \in \mathcal{Q}} n(r,d,q) P(c|r,d,q), \tag{13}$$

and

$$P(c) = \frac{\sum_{d \in \mathcal{D}, q \in \mathcal{Q}, r} n(r,d,q) P(c|r,d,q)}{\sum_{d \in \mathcal{D}, q \in \mathcal{Q}, r} n(r,d,q)}. \tag{14}$$