Sifting through images with Multinomial Relevance Feedback

Dorota Głowacka, Alan Medlar and John Shawe-Taylor

University College London

December 8, 2010
Problem - content-based image retrieval when the user is unable to specify the required content through tags or other properties of the images.

The system must extract information from the user through limited feedback.

Solution - a protocol that operates through a sequence of rounds in each of which a set of images is displayed and the user must indicate which is closest to their target.

A novel approach that makes use of the Dirichlet distribution as the conjugate prior to the multinomial distribution in order to model the system’s knowledge about the expected responses to the images.
The search engine supports the user in finding an image matching her query.

The search engine calculates a set of images $x_{i,1}, \ldots, x_{i,k}$ and presents them to the user.

If one of the images matches the user’s query, then the search terminates.

Otherwise the user chooses one of the images $x^*_i$ as most relevant according to a distribution $D\{x^*_i = x_{i,j} \mid x_{i,1}, \ldots, x_{i,k}; t\}$, where $t$ denotes the ideal target image.
Comparative Feedback

Following Auer & Leung 2009, we assume the following probability measure of choosing image $x_{i,j}$:

$$D\{x_i^* = x_{i,j} \mid x_{i,1}, \ldots, x_{i,k}; t\} = (1 - \alpha) \frac{S(x_{i,j}, t)}{\sum_{j=1}^{k} S(x_{i,j}, t)} + \frac{\alpha}{k}$$

The similarity measure $S(\cdot, \cdot)$ is given by

$$S(x, t) = \exp\{-ad(x, t)^2\}$$
The Dirichlet Sampling Search Algorithm

- Problem - model our knowledge about the user’s interests.
- The user feedback can be viewed as a multinomial distribution.
- A natural choice for the model is its conjugate prior.
- Algorithm is based on the Dirichlet Process:

\[
P(\Theta | \alpha, M) = \frac{\Gamma(\alpha)}{\prod_{i=1}^{n} \Gamma(\alpha m_i)} \prod_{i=1}^{n} \theta_i^{\alpha m_i - 1}
\]

where \( M = \{m_1, m_2, \ldots, m_n\} \) is the base measure and is the mean value of \( \Theta \), and \( \alpha \) is a precision parameter.
The posterior has updated precision parameter

\[ \alpha^* = \alpha + 1 \]

and the base measure

\[ M^* = \frac{\alpha M + \frac{1}{n_i} X_i}{\alpha + 1} \]

where \( X_i \) is the partition of size \( n_i \).
Trade-off between exploration and exploitation.

- Draw $k$ samples from the posterior distribution and select the image with the highest probability.

- Images with higher weights $m$ are more likely to be relevant and thus more likely to be presented to the user.
Experimental setup

- The data - the VOC2007 dataset with 23 categories and 9963 images.
- Each image is annotated by a bounding box and the feature value for an object class is the size (as calculated from the bounding box) of the largest object from this class in the image.
- If no object from a particular class is present, then the feature value is 0.
- We set $k = 2, 5, 10$ so that only 2, 5, 10 images are presented to the user in each iteration.
- All the results are averaged over 1000 searches for randomly selected target images from the dataset.
PicHunter (Cox et al. 2000) uses Bayes’ rule to predict the user’s target image.

The system maintains a set of probabilities $p_{1:n}$ for every image $x_{1:n}$ in a dataset. Initially, all the probabilities $p_i = \frac{1}{n}$.

After each iteration, the system estimates the probability that image $x_i$ is the user’s target image.

The probabilities are updated as in $p_i = p_i \times G(d(x_i, s_m))$, where $d(x_i, s_m) = \| x_i - s_m \|$ is the distance between $x_i$ and the image $s_m$ selected by the user in iteration $m$, and $G$ is defined as:

$$G = \frac{\exp(-d(x_i, s_m))/\sigma}{\sum_{j=1}^{n} \exp(-d(x_j, s_m))/\sigma}$$  \hspace{1cm} (1)
Auer & Leung Algorithm

- Weighting scheme that demotes less relevant images by a discount factor $\beta$
- Initially, the weights of all images are set to 1.
- At each iteration, a set of images are presented to the user who chooses one of the images as most relevant.
- If the search has not terminated, all the images presented to the user have their weights set to 0.
- All the images far from the image selected by the user are demoted by the discount factor $\beta$. 
Table: Comparison of the performance the AL algorithm, the DS algorithm and *PicHunter* PH as the value of $k$ increases.

<table>
<thead>
<tr>
<th>Target Size</th>
<th>k = 2</th>
<th></th>
<th>k = 5</th>
<th></th>
<th>k = 10</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AL</td>
<td>DS</td>
<td>PH</td>
<td>AL</td>
<td>DS</td>
<td>PH</td>
</tr>
<tr>
<td>1</td>
<td>845</td>
<td>330</td>
<td>1188</td>
<td>431</td>
<td>123</td>
<td>448</td>
</tr>
<tr>
<td>5</td>
<td>448</td>
<td>99</td>
<td>917</td>
<td>92</td>
<td>46</td>
<td>356</td>
</tr>
<tr>
<td>10</td>
<td>219</td>
<td>60</td>
<td>733</td>
<td>51</td>
<td>28</td>
<td>172</td>
</tr>
</tbody>
</table>
Sparse data representation

- With large datasets, calculating the distances of all images from the $k$ images presented to the user is expensive.
- We produce a small dataset of images selected randomly from the large dataset.
- At each iteration we replace images with the lowest probability with new images selected from the large dataset.
Figure: (a) The average distance from the target of the images shown to the user in 30 iterations of the DS algorithm with and without sparse data representation; (b) The distance of the image closest to the target in each iteration.
Real users experiments

Welcome to Image Search

Find the image below.

Which algorithm to run: [default]

How many images to show each iteration: [6 or 3]

Let's get started.

Image Search

Hover your mouse pointer over the closest image to the kind of thing you are searching for.

Click on "Next" to see similar images or "This one!" to terminate your search.

If none of the images is suitable, then you can ignore this round.

Alternatively, you could restart your search.

Dorota Główacka, Alan Medlar and John Shawe-Taylor

Sifting through images with Multinomial Relevance Feedback
Real users experiments

- 10 users performed 4 searches using the DS algorithm, the AL algorithm as well as a random search (40 searches altogether)

- 6 images presented to the user at each iteration

- Users instructed to terminate the search when they found the target image or after 50 iterations of the search algorithm

- Average number of iterations to find the target image using the DS algorithm was 29

- Average number of iterations to find the target image using the AL algorithm and random search was 47 and 48, respectively
Real users experiments

Figure: (a) The average distance from the target of the images shown to the user in the first 10 iterations of the DS algorithm, the AL algorithm and random search; (b) The distance of the image closest to the target in each iteration.
Conclusions

- A new approach to content-based image retrieval based on multinomial relevance feedback.

- The model suggests an algorithm for generating images for presentation that trades exploration and exploitation.

- The model allows to make predictions about the scaling of the algorithm and convergence properties.

- The experiments confirm that the new approach outperforms earlier work using a more heuristic strategy.