Image Retrieval with Multinomial Relevance Feedback

Dorota Głowacka and John Shawe-Taylor
Department of Computer Science, University College London
Gower Street, London WC1E 6BT, UK
d.glowacka@cs.ucl.ac.uk, jst@cs.ucl.ac.uk

1 Introduction

We consider content-based image retrieval when the user is unable to specify the required content through tags or other explicit properties of the images. In this type of scenario the system must extract information from the user through limited feedback. We consider a protocol that operates through a sequence of rounds in each of which a set of k images is displayed and the user must indicate which is closest to their target. Performance is assessed by (1) the number of rounds needed before the user is presented with the target image or an image that is among the t nearest neighbours of the target in the database; (2) the average distance from the target of the k images presented to the user at each iteration. While this problem has been studied before, we present 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.

We tested our algorithm in simulations and in a real-life scenario using a database of 40,000 images from the Tiny Images Dataset [2]. Building on earlier work [1] we assume a polynomial user model with the probabilities of response proportional to a function of the distance to the target. We use this model in our simulations. In the real-life experiments, the users were presented with a random image from the database and asked to search for it. At each iteration, the user was presented with 10 images and asked to click on the image most similar to the one they were looking for. The search terminated when the user was presented with the target image. Experimental results show the new approach compares favourably with previous work.

2 The algorithm

We describe an exploration strategy which is an attempt to solve the search problem. The algorithm maintains weights $m(\mathbf{x})$ on the images $\mathbf{x}$ in the database $\mathscr{D}$ and calculates the images to be presented to the user according to these weights.

Let us begin with the definition of Dirichlet distribution. Let $\Theta = \{\theta_1, \ldots, \theta_n\}$ be a multinomial probability distribution on the discrete space $\mathscr{X} = \{\mathscr{X}_1, \ldots, \mathscr{X}_n\}$ with $\mathbf{x}$ a random variable in the space $\mathscr{X}$. The Dirichlet distribution on Θ is given by the following formula: $P(\Theta \mid \alpha, M) = \frac{\Gamma(\alpha)}{\prod_{i=1}^{n}\Gamma(\alpha m_i)} \prod_{i=1}^{n} \theta_i^{\alpha m_i - 1}$, where $M = \{m_1, \ldots, m_n\}$ is the base measure defined on $\mathscr{X}$ and is the mean value of Θ, and α is a precision parameter that specifies how concentrated the distribution is around M. α can be regarded as the number of (pseudo-) measurements observed to obtain M. We refer to this distribution as $Dir(\alpha m_1, \ldots, \alpha m_n) = Dir(\alpha M)$

Consider a possibly continuous input space $\mathscr{X}$. A Dirichlet Process (DP) on $\mathscr{X}$ is a distribution over distributions on $\mathscr{X}$ with samples being measures on $\mathscr{X}$. For a random distribution G to be distributed according to a DP, its marginal distributions on partitions of the input space have to be Dirichlet distributed. G is DP distributed with base measure M and precision parameter α, written $G \sim DP(\alpha, M)$ if $(G(\mathscr{X}_1), \ldots, G(\mathscr{X}_n)) \sim Dir(\alpha M(\mathscr{X}_1), \ldots, \alpha M(\mathscr{X}_n))$ for every measurable partition $\mathscr{X}_1, \ldots, \mathscr{X}_n$ of $\mathscr{X}$.

We are interested in using the Dirichlet distribution to learn a distribution from observations based on a multinomial noise model. If the target distribution on $\mathscr{X}_1, \ldots, \mathscr{X}_n$ is $\mu = (\mu_1, \ldots, \mu_n)$, we will observe $\mathscr{X}_i$ with probability μ_i. Let $z_1, \ldots, z_l$ s.t. $z_i \in \{\mathscr{X}_1, \ldots, \mathscr{X}_n\}$ be a sequence of independent draws from μ. We are interested in the posterior distribution G given the observations of $z_1, \ldots, z_l$. Let $\mathscr{X}_1, \mathscr{X}_2, \ldots, \mathscr{X}_n$ be a finite measurable partition of $\mathscr{X}$, and let n_k be the number of observed values in $\mathscr{X}_k$, that is $n_k = \mid \{i : z_i = \mathscr{X}_k, i = 1, \ldots, l\} \mid$. Then, $(G(\mathscr{X}_1), \ldots, G(\mathscr{X}_n)) \sim Dir(\alpha M(\mathscr{X}_1) + n_1, \ldots, \alpha M(\mathscr{X}_n) + n_n)$ The posterior DP has updated precision parameter $\alpha^* = \alpha + l$ and the base measure $M^* = \frac{\alpha M + \sum_{i=1}^{n} n_i l_i}{\alpha + l}$, where l_i is a unit vector with entry 1 in the i^{th} component. The posterior base measure distribution is a weighted average between the prior base measures M and the empirical distribution $\frac{\sum_{i=1}^{n} n_i l_i}{n}$. The weight associated with the prior base distribution is proportional to α, while the empirical distribution has weight proportional to the number of observations n. Thus, as the number of observations grows, $n \gg \alpha$, the posterior is dominated by the empirical distribution.

2.1 Application to the image selection problem

Let $\mathscr{D}$ be a dataset of n images $(\mathbf{x}_i)_{i=1,\ldots,n}$. Let $M = \{m_1, \ldots, m_n\}$ be the base measure defined on $\mathscr{D}$. Initially, we set $m_i = \frac{1}{n}$ for $i = 1, \ldots, n$. Let $\mathbf{x}_i^* \in \{\mathbf{x}_{i,1}, \ldots, \mathbf{x}_{i,k}\}$ be the image chosen by the user at iteration i from among the k presented images $\{\mathbf{x}_{i,1}, \ldots, \mathbf{x}_{i,k}\}$. We suppress the index i to simplify the exposition. We need to define the user model that defines how the image $\mathbf{x}_i^*$ is chosen. The simplest possibility is that we view the set of images $\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ as partitioning the complete space of images into sets $\mathscr{X}_1, \ldots, \mathscr{X}_k$ with $\mathscr{X}_j = \{\mathbf{x} : d(\mathbf{x}_j, \mathbf{x}) < d(\mathbf{x}_{j'}, \mathbf{x}), j' \neq j\}$. In the event of ties a random assignment to one of the minimum distance partitions is made. Now, if we take the "true" response probability as $m_i^* \propto f(d(\mathbf{x}_i, \mathbf{t}))$ where f is a monotonically decreasing function and $\mathbf{t}$ is the target image, then the user model should choose partition $\mathscr{X}_j$ with probability $P(\mathscr{X}_j) = \sum_{i:\mathbf{x}_i \in \mathscr{X}_j} m_i^*$ When we consider the user's choice, we need to update the base measures and the precision parameter. The standard posterior update tells us how to do this for the partition probabilities, but not for the single images. Since we are not able to distinguish between the images in a partition, the natural choice is to update

if $x_i \in \mathscr{X}_j$ **then**
  $m_i \leftarrow \frac{\alpha m_i + |\mathscr{X}_j|^{-1}}{\alpha + 1}$
else
  $m_i \leftarrow \frac{\alpha m_i}{\alpha + 1}$
end if
$\alpha \leftarrow \alpha + 1$

This choice is the maximum entropy update consistent with the required update on the current partition. Note that at each iteration the partition of the images will change, but that the update will correctly compute the posterior measure over the complete set of images.

The final ingredient in the search algorithm is the way in which the images to be presented to the user should be chosen at each iteration. This involves a trade-off between presenting images that appear promising based on best current estimates of the mean given by the posterior measure $(m_i)_{i=1,\ldots,n}$ (exploitation) and trying areas where our current estimate could be too pessimistic (exploration). The strategy we adopt to solve this problem is to draw k samples from the posterior distribution and select the image $\mathbf{x}_j$ that has the highest probability in the jth sample, $j = 1, \ldots, k$. Note that we should not use the individual m_j of the images when drawing these samples since the image selected will be a proxy for the approximately n/k images in its partition and it is the partition that we wish to choose. This problem is overcome by multiplying each of the αm_i by n/k before drawing each sample. We summarise the selection algorithm as

for $\mathtt{j} = 1, \ldots, k$ **do**
  $\mathtt{r} \leftarrow \mathtt{randg}(m * \alpha * n/k)$
  $[\mathtt{val}, \mathtt{i}] \leftarrow \max(\mathtt{r})$
  $\mathtt{ind}[\mathtt{j}] \leftarrow \mathtt{i}$
end for
Output: Array `ind` gives indices of selected images

One way of viewing this strategy is selecting images to display with probability proportional to the probability that the partition they define contains the target. This algorithm randomly selects images from the dataset according to a sample from the posterior. Images with highest weights in each sample are presented to the user.

3 Sparse data representation

When the image dataset is very large, calculating the distances of all the data points from the k images displayed at each iteration of the search is computationally expensive. We therefore propose a revised version of our algorithm, where we work with a sparse representation of $\mathscr{D}$, thus reducing the number of calculations required in each step of the search algorithm.

Assuming that we have an image dataset $\mathscr{D}$ of size n, initially we produce a small dataset $\mathscr{A} = \{y_1, \ldots, y_l : y_i \in \mathscr{D} : l \ll n\}$. At each iteration, we replace f data points with the lowest base measures with new data points $\{y_1, \ldots, y_f : y_j \in \mathscr{D} \wedge y_j \notin \mathscr{A}\}$. After updating the dataset $\mathscr{A}$, we calculate the base measure m_j of a newly added point y_j:

for $i = 1, \ldots, l$ **do**
  $b_i = d(y_i, y_j)$
end for
$[val, ind] = min(b)$
$m_j = m_{ind}$

That is, the base measure of a newly added point takes the value of the base measure of the point in $\mathscr{A}$ closest to it.

The intuition behind this algorithm is that in each iteration we remove data points with the lowest weights and consequently the lowest probability of being in the proximity of the target image. Thus, although the dataset $\mathscr{A}$ is much smaller than the original image dataset $\mathscr{D}$, as the search progresses, $\mathscr{A}$ contains more and more images close to the target image with very few images of little interest to the user. The results of the experiments indicate that even such a simple data representation allows the algorithm to find the target image relatively fast.

References

[1] P. Auer, and A.P. Leung, "Relevance feedback models for content-based image retrieval", in *Multimedia Analysis, Processing and Communications*. New York, Springer, 2009.

[2] A. Torralba, R. Fergus, W.T. Freeman, "80 million tiny images: a large data set for nonparametric object and scene recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30(11), pp. 1958 – 1970, 2008.

Topic: other applications
Preference: oral/poster