Using Non-Parametric Quantum Theory to Rank Images

Songhao Zhu[1], Baoyun Wang[1], Yuncai Liu[2]

[1]School of Automation, Nanjing University of Posts and Telecommunications, Nanjing, 210046, China
[2] School of Electronic Information and Electric Engineering, Shanghai Jiao Tong University, Shanghai, 200240, China
{zhush, bywang}@njupt.edu.cn, {homewliu}@sjtt.edu.cn

ABSTRACT

Recently learning to rank has become one of the popular means to create a ranking model for social image search. However, the results of existing approaches are not as satisfactory for the large gap between low-level visual features and high-level semantic concepts, and these sophisticated approaches require a significant amount of parameters tuning to be effective and efficient. In this paper, we propose a novel framework for social image re-ranking based on a non-parametric quantum technique, which re-ranks top retrieved images by considering the inter-relationship between images through the quantum estimation and requires no explicit parameter tuning. The basic idea of the proposed framework is inspired by the photon polarization experiment supporting the theory of quantum measurement. Experimental results conducted on the Flickr dataset demonstrate the effectiveness and efficiency of the proposed framework.
Key words—Social image search, non-parametric approach, quantum measurement, Flickr

1. INTRODUCTION

A common problem of content-based image search is: to achieve semantically related images to a query image, a ranking model needs to be learned with a significant amount of parameters tuning with only a small number of given positive images, such as [1-5]. In spite of encouraging performance has been achieved, the content-based image search problem has not been solved satisfactorily due to the large gap between low-level visual features and high-level semantic concepts, i.e., images of dissimilar semantic content may share some common low-level visual features, while images of similar semantic content may be scattered in the feature space. This motivates our study in this paper, which aims to improve ranking results by a novel non-parametric quantum ranking scheme. Besides, unlike previous approaches that often require a significant amount of parameters tuning to obtain an efficient and effective similarity measure, there is no explicit parameter tuning required in our scheme.

Motivated by the photon polarization experiment in [6], which is one of the important experiments that describes the basic theory of quantum estimation, we propose a novel learning framework named non-parametric quantum estimation based image search. In our framework, the images in the database is considered as photons generated from the light source, and the search process as estimating all the database images by the query polarization filter.

Figure 1 shows the basic idea of the proposed framework. Firstly, to deal with the issue of high computational cost involved in large image databases, a pre-filtering processing is utilized to remove the most irrelevant images while keeping the most relevant ones in the database to an input query. Then, in the first-round of the search, the relevance probability between each remained database image and the input query image is calculated by using a graph-based semi-supervised learning approach. Finally, quantum estimation is adopted to refine the relevance of searched candidates with respect to the top most images.

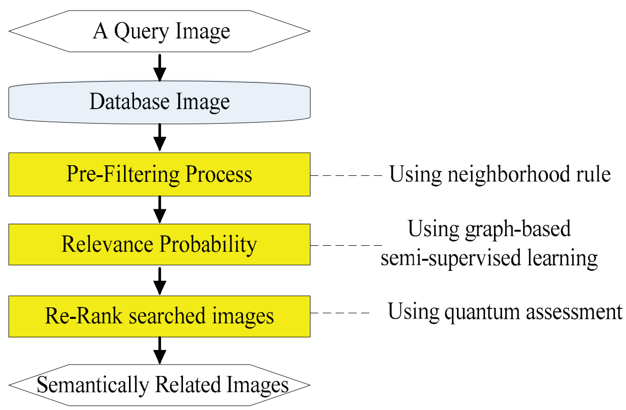


Figure 1. Flowchart of the proposed framework.

The organization of this paper is as follows. Section 2 details the pairwise similarity measurement and the pre-filtering processing. Section 3 and 4 introduce the process of computing the relevance scores of images using a graph-based semi-supervised learning approach and the refinement processing using one important quantum theory, respectively. Experimental results are provided in Section 5, followed by the concluding remarks in Section 6.

2. PAIRWISE SIMILARITY MEASUREMENT AND PRE-FITEERING PROCESSING

In this section, we will first introduce the pairwise similarity measurement by considering the influence of both other images and neighbourhood images. Then, we will analyze the importance and necessity of the pre-filtering processing.

2.1. Pairwise Similarity Measurement

According to [7], there are two basic assumptions used in the graph-based semi-supervised learning algorithms. The idea of the first basic assumption is described as follows: nearby images are likely to have the same label, and the second basic assumption is described as follows: images within the same manifold are likely to have the same label. Based on the first assumption and the experimental results in [8], the similarity between images i and j is here calculated using the metric of Manhattan distance:

$$S_c(i,j) = \exp\left[-\frac{|x_i - x_j|}{\sigma}\right] = \prod_{l=1}^{L} \exp\left[-\frac{|x_i^l - x_j^l|}{\sigma^l}\right] \tag{1}$$

where x_i^l is the l^{th} dimension in the low-level feature vector x_i of image i, L is the feature space dimension, and σ^l is the positive parameter reflecting the scope of different dimensions. According to the second assumption and for the purpose of consistency, the similarity between images i and j is also measured with respect to the Manhattan distance metric:

$$\mathrm{S}_n(i,j) = \exp\left(-\frac{|n_i - n_j|}{\sigma}\right) \tag{2}$$

where n_i and n_j are the measurement of neighbourhood density of image i and j respectively, and n_i is defined as below:

$$n_i = \frac{1}{N_i}\sum_{p=1}^{N_i}\prod_{l=1}^{L} \exp\left[-\frac{|x_i^l - x_p^l|}{\sigma^l}\right] \tag{3}$$

where N_i is the set of neighbours of image i.

Considering both of the above two assumptions simultaneously, the formulation of the pairwise similarity between images i and j can be defined as below:

$$\begin{aligned} S(i,j) &= S_c(i,j) \bullet \mathrm{S}_n(i,j) \\ &= \exp\left[-\frac{|x_i - x_j|}{\sigma}\right] \bullet \exp\left(-\frac{|n_i - n_j|}{\sigma}\right) \end{aligned} \tag{4}$$

where '•' means the Hadamard product. The first term on the right side of equation (4) indicates that the similarity between two images decreases with the increment of their distance in the feature space, and the second one shows that the similarity between two images decreases with respect to the increment of the density difference.

2.2. Pre-filtering Processing

It is known to all that the computational cost renders many state-of-the-art algorithms intractable when facing large image databases. To deal with this issue, an efficient pre-filtering processing is here utilized to preserve the most relevant images whilst filtering out the most irrelevant ones for a given input image.

After analyzing the whole image retrieval process, it can be seen that the pre-filtering processing should simultaneously satisfy the following two criteria: low computational cost and high recall rate. Here, a modified nearest neighbour rule is adopted to implement the pre-filtering processing. Specifically, for a given query image, images in the database are ranked according to the pairwise similarity measurement calculated using equation (4): the larger the value of pairwise similarity measurement for an image, the higher the ranking score it has. After ranking dataset images according to the pairwise similarity value, a specified percentage of images are filtered out. Under the circumstances, the whole computational cost can scale from $O(M^3)$ (here M is the number of database images) to $O(N^3)$ (here N is the number of retained images and $N << M$), and therefore the computational cost is significantly reduced.

3. RELEVANCE SCORE COMPUTATION

For a given query image, a scale should be introduced to describe the relevance between the query image and each retained image in the database. Here, a graph-based semi-supervised learning approach, manifold ranking [7] is utilized to estimate such relevance scales, which contains the following two steps: similarity graph construction and relevance score computation.

3.1. Similarity Graph Construction

Constructing a similarity graph of a query image requires two major steps: the first step is vertexes setting and the second one is vertexes connecting.

Vertexes Setting. For a query image q, suppose $Э=\{1, 2,\ldots, i,\ldots, N\}$ is the set of N images generated using the nearest neighbour rule as describe in section 2.2. Then, the form of a weighted similarity graph G can be formulated as $G=(V,E)$, where V denotes the vertices constructed with $Э$ and the query image q, and all vertices are connected with each other through weighted edge set E.

Vertexes Connecting. The edge set E of a similarity graph G are weighted by a $(N+1)*(N+1)$ affinity matrix W, where element $W(i,i)$ is set to be zero to avoid self-reinforcement and element $W(i,j)$ when $i\neq j$ is constructed with respect to equation (4).

3.2. Relevance Score Computation

The procedure of the relevance score computation consists of two major steps: weighted matrix normalization and relevance score computation.

Weighted Matrix Normalization. The weighted matrix W is normalized using the following equation:

$$S = D^{-\frac{1}{2}} W D^{\frac{1}{2}} \tag{5}$$

where D is a diagonal matrix with entry $D(i,i)$ set to be the sum of the i^{th} row of the weighted matrix W.

Relevance Score Computation. The computation of relevance score of each image in $Э$ is implemented by an iterative process:

$$c(t+1) = \alpha S c(t) + \beta y \tag{6}$$

where y is a vector with entry $y(i)$=1 if the i^{th} image is a query image and $y(i)$=0 if the i^{th} image is a database image, and $c(0)=y$. Furthermore, α is in [0, 1) and $\alpha+\beta=1$. The i^{th} element of $(N+1)*1$ vector c is the candidate relevance value of the corresponding database image for a query image q.

4. THE BASIC ALGORITHM

Once the initial relevance scores of remaining database images for a query image are achieved, a non-parametric quantum ranking algorithm is utilized to re-rank candidate results by taking into account the inter-relationship between images through the "quantum estimation". This section contains the following two parts: quantum estimation explanation and proposed approach description.

4.1. Explanation of Quantum Estimation

The basic idea of quantum estimation can be perfectly explained through the photon polarization experiment. The polarization state of a photon can be modelled using a unit vector with appropriate direction. More specifically, the quantum state of any photon's polarization can be estimated using a linear combination:

$$\phi = \mu\,|\uparrow\rangle + \nu\,|\rightarrow\rangle \tag{7}$$

where "$|\uparrow\rangle$"and "$|\rightarrow\rangle$" are vertical polarization and horizontal polarization respectively, and μ and υ are complex numbers with $|\mu|^2 + |\upsilon|^2 = 1$.

The quantum estimation of a photon's polarization state transforms the state into the corresponding orthonormal basis vector, namely the probability of the state is represented as the squared magnitude of the amplitude in the direction of the corresponding basis vector. For example, a state Φ formulated as equation (7) is estimated using "$|\rightarrow\rangle$"with probability $|\upsilon|^2$ and "$|\uparrow\rangle$"with probability $|\mu|^2$. After the estimation of "$|\rightarrow\rangle$", the state Φ will fail to $\upsilon|\rightarrow\rangle$. Similarly, after the estimation of "$|\uparrow\rangle$", the state Φ will fail to $\mu|\uparrow\rangle$. Details please refer to [6].

4.2. Description of Proposed Approach

The initial quantum state of any image i is defined as:

$$\phi_i = \mu_i\,|1\rangle + \nu_i\,|0\rangle \tag{8}$$

where $|\mu_i|^2 + |\upsilon_i|^2 = 1$, the state "$|1\rangle$" means the relevance basis, and the state "$|0\rangle$" means the irrelevance basis with respect to the given query q. In the first-round retrieval, only the relevance basis "$|1\rangle$" is utilized to estimate the initial relevance of any image i and achieve the relevance probability $|\mu_i|^2$. Here, $|\mu_i|^2 = p(i|q)$ is the relevance probability obtained using the manifold-tanking learning approach as described in section 3.

For the purpose of re-ranking the images, the top measurement (t-measurement for short) is here introduced to estimate any image i with respect to the topmost images, where the intuition of this top-measurement is that usually the topmost images are more likely to be relevant than other ones. Similar to the definition of image i, the quantum state of a top image ti can be defined as follows:

$$\phi_{ti} = \mu_{ti}\,|1\rangle + \nu_{ti}\,|0\rangle \tag{9}$$

The state of image i after the t- measurement Φ_{ti} is the interested state and formulated using the following equation:

$$\phi_i = \mu_i\,|1\rangle + \nu_i\,|0\rangle = \gamma\phi_{ti} + \eta\phi_{ti}^{-1} \tag{10}$$

where Φ_{ti}^{-1} is the orthonormal form of Φ_{ti}, and γ and η are complex numbers. After replacing Φ_{ti}^{-1} with the corresponding form, equation (10) can be rewritten as:

$$\begin{aligned}\phi_i &= \mu_i\,|1\rangle + \nu_i\,|0\rangle = \gamma\phi_{ti} + \eta\phi_{ti}^{-1} \\ &= \gamma\left(\mu_{ti}\,|1\rangle + \nu_{ti}\,|0\rangle\right) + \eta\left(-\nu_{ti}\,|1\rangle + \mu_{ti}\,|0\rangle\right) \\ &= \left(\gamma\mu_{ti} - \eta\nu_{ti}\right)|1\rangle \ + \ \left(\gamma\nu_{ti} + \eta\mu_{ti}\right)|0\rangle\end{aligned} \tag{11}$$

Then γ and η are the solutions of the following linear equations respectively:

$$\begin{cases}\gamma\mu_{ti} - \eta\nu_{ti} = \mu_i \\ \gamma\nu_{ti} + \eta\mu_{ti} = \nu_i\end{cases} \tag{12}$$

Recalling that $|\mu_{ti}|^2 + |\upsilon_{ti}|^2 = 1$, therefore the solution of γ is:

$$|\gamma| = |\mu_i\mu_{ti} + \nu_i\nu_{ti}| \tag{13}$$

Since the retrieval process is here viewed as estimating the quantum state of all the images with the query polarization filter, the quantum state Φ_i of the i^{th} image will fail to the direction of Φ_{ti} after the t-measure:

$$\phi_i^t = \gamma\phi_{ti} = \gamma\mu_{ti}\,|1\rangle + \gamma\nu_{ti}\,|0\rangle \tag{14}$$

where Φ_i^{t} is the state vector of image i after the t-measurement.

To achieve the revised relevance probability of image i with respect to the query image q, the current state of image i Φ_i^{t} is estimated by failing to the relevance basis $|1\rangle$:

$$p\left(i|ti, q\right) = |\gamma\mu_{ti}|^2 \tag{15}$$

By the formulation above, it can be seen that $|\gamma|$ is 1 when $i = ti$ according to the following derivation process:

$$|\gamma| = |\mu_i\mu_{ti} + \nu_i\nu_{ti}| = |\mu_{ti}{}^2 + \nu_{ti}{}^2| = 1 \tag{16}$$

and

$$p\left(i|ti, q\right) = |\gamma\mu_{ti}|^2 = |\mu_{ti}|^2 = p\left(ti|q\right) \tag{17}$$

As it can be seen from the above equation that the relevance probability of the topmost image ti is unchanged after the t-measure.

According to the above derivation process, the formualtion of the revised relevance probability of image i with respect to the query image q can be formulated as follows:

$$p\left(i|q\right) = \sum_{ti\in\Gamma}\left[p\left(i|ti, q\right) * s\left(i, ti\right)\right] \tag{18}$$

where Γ is a set of the topmost images, and $s(i,ti)$ is the weight of the corresponding top measurement. In the proposed framework, equation (18) is considered as the basis of image re-ranking.

5. EXPERIMENTAL RESULTS

5.1. Experiment Design

To evaluate the effectiveness of the proposed scheme and compare it with state-of-the-art approaches, a general purpose image database, the 68,200 Corel dataset are utilized as the ground truth database. This image dataset contains 682 semantic classes, where each class contains 100 images on the same topic, such as ancestor, beach, bus, church, desert, forest, and golf. The database is split into these two parts: 30,000 images as the training set and the remaining 38,200 as the queries.

To describe the reprehensive visual content of images, the extracted feature vector is a 428-dimensional vector, including 225-dimensional block-wise color moment, 128-dimensional wavelet texture and 75-dimensional edge distribution histogram.

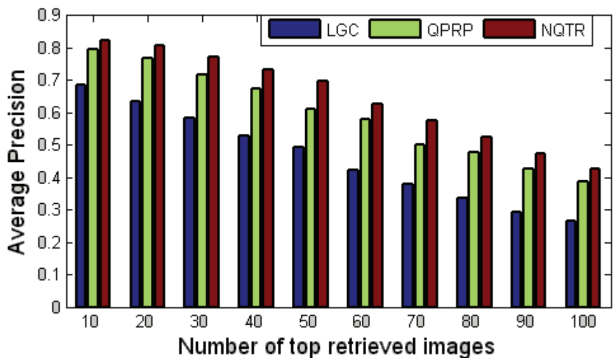
To evaluate the performance of the proposed retrieval quality improvement scheme, precision and recall on top m retrieved images of each query image are first calculated respectively, and then these two measure scores of all the query images are averaged as the overall performance evaluation measure:

$$Precision\Theta m = \frac{|r_{correct}|}{|r_{system}|},\ Recall\Theta m = \frac{|r_{correct}|}{|r_{ground}|} \quad (19)$$

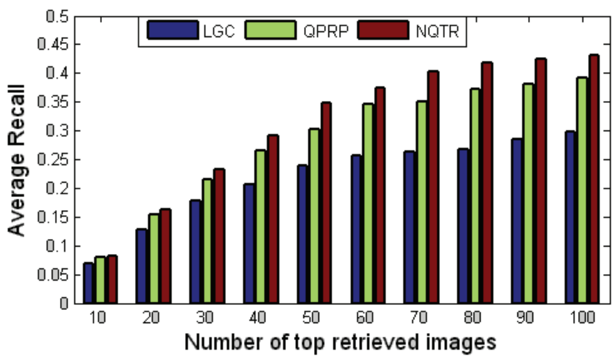
where $|r_{ground}|$ denotes the number of ground truth images of one query image q, and $|r_{correct}|$ is the number of correct images in the retrieved images $|r_{system}|$ of one query image q.

5.2. Experiment Results

The performances of three approaches, including the approach proposed by Zhou *et al.* [7] (LGC for short), Zuccon *et al.* [9] (QPRP for short), and the proposed approach, are illustrated in figure 2, where the LGC is adopted as the baseline approach.



(a) Precision values comparison of top @m.



(b) Recall values comparison of top @m.

Figure 2: Performance comparison of different approaches.

As it is shown in figure 2, both QPRP and ours achieve significant and suitable improvement over LGC, and the proposed approach outperforms the QPRP. The experimental results confirm the proposed scheme from the following two aspects: i) by taking into account both the structure assumption and the neighborhood assumptions, the proposed method can better describe the pairwise similarity measure between images than traditional methods; ii) by taking into account the inter-relationship between the topmost images and other database images, the proposed method can better infer the probality of relevance between the query and database images than other methods.

6. CONCLUSION

In this paper, a novel image retrieval framework based on the non-parametric quantum theory is proposed to narrow the gap between low-level features and high-level semantics, where the idea is inspired by a recently developed quantum probability theory [6, 9]. The theory is initially proposed to infer the inter-relationship between documents by the so-called quantum interference. In the proposed image retrieval framework, this theory is adopted to explore the inter-relationship between the topmost images and other database images, where the pairwise similarity measure between images is calculated based on the structure assumption and neighborhood assumptions. Experiments conducted on a general purpose image database consisting of 68,200 Corel images have demonstrated that the performance of image retrieval can be improved significantly by incorporating the top measurement.

Since feature selection is an open issue and has great impact on the experimental results, our feature work will introduce other features into the scheme to further improve the system performance. Our further work also includes conducting experiments on other large image databases, and comparing with other quantum inspired schemes. Furthermore, the relevance feedback is another research direction in our further work.

ACKNOWLEDGMENTS

This work is supported by RFDP of 20103223120003, NSFJ of BK2011758 and BK2010077, PAPD, CNSF of 60972045 and 61071089, SKLNST.

REFERENCE

[1] H. Sahbi, J. Audibert, and R. Keriven. Graph-Cut Transducers for Relevance Feedback in Content Based Image Retrieval. Proceeding of IEEE Conference on Computer Vision, pp:1-8, 2007.

[2] M. Wang, B. Liu, and X. Hua. Accessible image search. Proceeding of ACM Conference on Multimedia, pp:291-300, 2009.

[3] F. Perronnin, Y. Liu, J. Sánchez, and H. Poirier. Large-scale image retrieval with compressed Fisher vectors. Proceeding of IEEE Conference on Computer Vision and Pattern Recognition, pp:3384-3391, 2010.

[4] J. Zhuang and S. Hoi. Non-parametric Kernel Ranking Approach for Social Image Retrieval. Proceeding of ACM Conference on Image and Video Retrieval, pp:26–33, 2010.

[5] C. Beecks, I. Assent, and T. Seidl. Content-Based Multimedia Retrieval in the Presence of Unknown User Preferences. Proceeding of IEEE Conference on Multimedia Modeling, pp:140-150, 2011.

[6] E. Rieffel and W. Polak. An introduction to quantum computing for non-physicists. Computing Surveys, 2000, 32 (3): 300-335.

[7] D. Zhou, O. Bousquet, N. Thomas, W. Jason, and S. Bernhard. Learning with Local and Global Consistency. Proceeding of IEEE Conference on Neural Information Processing Systems, pp: 321-328, 2003.

[8] M. Kokare, B. Chatterji, and P. Biswas. Comparison of similarity metrics for texture image retrieval. Proceeding of IEEE Conference on Convergent Technologies for Asia-Pacific Region, pp: 571-575, 2003.

[9] G. Zuccon and L. Azzopardi. Using the quantum probability ranking principle to rank interdependent documents. Proceeding of European Conference on Information Retrieval, pp: 357-369, 2010.