Visual Reranking with Local Learning Consistency

Xinmei Tian[1,⋆], Linjun Yang[2], Xiuqing Wu[1], and Xian-Sheng Hua[2]

[1] University of Science and Technology of China, Hefei, Anhui, 230027, China
[2] Microsoft Research Asia, 100190, Beijing, China
xinmei@mail.ustc.edu.cn, linjuny@microsoft.com,
xqwu@ustc.edu.cn, xshua@microsoft.com

Abstract. The graph-based reranking methods have been proven effective in image and video search. The basic assumption behind them is the ranking score consistency, *i.e.*, neighboring nodes (visually similar images or video shots) in a graph having close ranking scores, which is modeled through a regularizer term. The existing reranking methods utilise pair-wise regularizers, *e.g.*, the Laplacian regularizer and the normalized Laplacian regularizer, to estimate the consistency over the graph from the pair-wise perspective by requiring the scores to be close for pairs of samples. However, since the consistency is a term defined over the whole set of neighboring samples, it is characterized by the local structure of the neighboring samples, *i.e.*, the multiple-wise relations among the neighbors. The pair-wise regularizers fail to capture the desired property of consistency since they treat the neighboring samples independently. To tackle this problem, in this paper, we propose to use local learning regularizer to model the multiple-wise consistency, by formulating the consistent score estimation over a local area into a learning problem. Experiments on the TRECVID benchmark dataset and a real Web image dataset demonstrate the superiority of the local learning regularizer in visual reranking.

Keywords: Image/video search, visual reranking, regularizer, local learning.

1 Introduction

Most of the frequently-used image/video search engines build on text-based search by using the associated textual information, such as URL, surrounding text from the Web pages, speech transcript, *etc.* However, it is commonly observed that the text-based search results are not satisfying enough. Irrelevant or little-relevant images are often returned as top results, which is caused by the mismatching between images and its surrounding texts, the low ability to discriminate images' relevance levels with only textual information and so on.

⋆ This work was performed when Xinmei Tian was visiting Microsoft Research Asia as a research intern.

S. Boll et al. (Eds.): MMM 2010, LNCS 5916, pp. 163–173, 2010.

To tackle these difficulties, visual reranking is proposed to improve the text-based search result by incorporating visual information. Up to now, many reranking methods have been proposed, *e.g.*, classification-based [8,13], clustering-based [3] and graph-based [4,6,7,10]. Among them, the graph-based methods have drawn increasing attention recently and have already shown promising results on image and video search [6,10]. In such methods, a graph is constructed with the samples, *i.e.*, images or video shots, as the nodes and the edges between them being weighted by their visual similarity. In [4,6,7], reranking is formulated as random walk over the graph and the ranking scores are propagated through the edges. The stationary probability of the random walk process is used as the reranked score directly.

The basic assumption behind the graph-based reranking methods is that: neighboring nodes in the graph should have similar ranking scores, *i.e.*, the ranking score consistency (or smoothness) over the graph topology. By modeling this assumption explicitly, a general reranking framework, *i.e.*, Bayesian reranking, is proposed in [10]. Bayesian reranking maximizes the ranking score consistency while minimizes the ranking distance, which represents the disagreement between the reranked result and the initial text-based ranking result. As illustrated in [10], the other graph-based reranking methods [4,6,7] can also be unified into this framework.

The ranking score consistency over the graph is represented through a regularizer term, which plays a crucial role in the graph-based reranking methods. In [10], the widely used regularizers in graph-based classification and video annotation, *i.e.*, Laplacian regularizer and the normalized Laplacian regularizer, are directly adopted. In [4,6,7], a variant of normalized Laplacian regularizer is utilised, as discussed in [10].

However, all the regularizers mentioned above model the ranking score consistency by approximating from the pair-wise perspective. Specifically, for each sample, a set of pairs is formed between it and each of its neighbors, and then, the overall consistency is measured by aggregating the individual consistency over each pair. In fact, a sample's consistency on a local area of the graph is multiple-wise instead of pair-wise, since the consistency is a term defined over the whole neighboring samples instead of over each pair of them. In other words, the consistency is characterized by the structure of the neighboring set of data. The pair-wise regularizers treats each neighbors independently with the multiple-wise relationship among them unconsidered. From this point of view, the consistency approximated with pair-wise regularizers is not satisfactory enough.

In this paper, we propose to use a local learning regularizer to model the desired multiple-wise consistency. Specifically, for each sample, instead of calculating the consistency with each of its neighboring samples individually, the local learning based regularizer considers the consistency with all of its neighboring samples simultaneously. In this regularizer, a local model is firstly trained for each sample with its neighbors and then this model is used to predict this sample's consistent ranking score. Finally, by minimizing the difference between the objective ranking score and this locally predicted one, the desired multiple-wise consistency over the graph is guaranteed.

The rest of this paper is organized as follows. In Section 2, we will briefly introduce the graph-based visual reranking method. The proposed local learning regularized reranking method and its solution are detailed in Section 3. Experimental results are presented and analysed in Section 4. The parameter sensitivity analyses are presented in Section 5, followed by the conclusion in Section 6.

2 Graph-Based Visual Reranking

Given N samples (images or video shots) $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N\}$ and their initial text search result represented by a ranking score vector $\bar{\mathbf{r}} = [\bar{r}_1, \bar{r}_2, \cdots, \bar{r}_N]^T$ where $\bar{r}_i$ is the ranking score of $\mathbf{x}_i$. A larger ranking score indicates the sample is more relevant and thus should be ranked higher.

Bayesian reranking [10], a general graph-based visual reranking framework is proposed recently. In this framework, a graph $\mathcal{G}$ is constructed with nodes being the samples while weights on edges being the similarity between the corresponding samples. Specifically, the weight w_{ij} on the edge between nodes $\mathbf{x}_i$ and $\mathbf{x}_j$ is computed using Gaussian kernel $w_{ij} = \exp(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2})$, where σ is the scaling parameter. Then, the reranking is formulated into an optimization problem and the optimal reranking score vector $\mathbf{r} = [r_1, r_2, \cdots, r_N]^T$ is derived by minimizing the following energy function,

$$E(\mathbf{r}) = Reg(\mathcal{G}, \mathbf{r}) + c \times Dist(\mathbf{r}, \bar{\mathbf{r}}). \tag{1}$$

Here, the first term, termed regularizer term, penalizes the ranking score inconsistency over the graph topology while the second term, termed ranking distance term, penalizes the derivation of the reranked result from the initial ranking. c is a trade-off parameter which balances the influence of the two terms. We will briefly introduce the two terms respectively in the following.

2.1 Ranking Distance

The ranking distance term is measured by the disagreement between the two ranking score vectors $\mathbf{r}$ and $\bar{\mathbf{r}}$. The preference strength distance is proposed in [10]:

$$Dist(\mathbf{r}, \bar{\mathbf{r}}) = \frac{1}{2}\sum\nolimits_{i,j} dist_{ij} = \frac{1}{2}\sum\nolimits_{i,j,i \neq j} (1 - \frac{r_i - r_j}{\bar{r}_i - \bar{r}_j})^2, \tag{2}$$

where $dist_{ij}$ denotes pair $(\mathbf{x}_i, \mathbf{x}_j)$'s distance which is measured by the two samples' preference strength difference before and after reranking. The preference strength which means the score difference of two samples, *i.e.* $r_i - r_j$ for pair $(\mathbf{x}_i, \mathbf{x}_j)$, reflects the degree of $\mathbf{x}_i$ being ranked before $\mathbf{x}_j$. The preference strength distance has been proven to be effective [10], and will be directly utilised in this paper.

2.2 Regularizer

The regularizer term aims to model the ranking score consistency over the graph topology. The widely used regularizers in semi-supervised classification and video annotation, *i.e.*, Laplacian regularizer [15] and normalized Laplacian regularizer [14], are directly adopted in [10].

When Laplacian regularizer is adopted,

$$Reg_{Lap}(\mathcal{G}, \mathbf{r}) = \mathbf{r}^T \mathbf{L} \mathbf{r} = \frac{1}{2} \sum\nolimits_{i,j} w_{ij} (r_i - r_j)^2, \tag{3}$$

where $\mathbf{L} = \mathbf{D} - \mathbf{W}$ is the Laplacian matrix. $\mathbf{D} = Diag(\mathbf{d})$ is a degree matrix with $\mathbf{d} = [d_1, d_2, \cdots, d_N]^T$ and $d_i = \sum_j w_{ij}$.

When normalized Laplacian regularizer is adopted:

$$Reg_{NLap}(\mathcal{G}, \mathbf{r}) = \mathbf{r}^T \mathbf{L}_n \mathbf{r} = \frac{1}{2} \sum\nolimits_{i,j} w_{ij} (\frac{r_i}{\sqrt{d_i}} - \frac{r_j}{\sqrt{d_j}})^2, \tag{4}$$

where $\mathbf{L}_n = \mathbf{I} - \mathbf{D}^{-\frac{1}{2}} \mathbf{W} \mathbf{D}^{-\frac{1}{2}}$. $\mathbf{I}$ is the unit matrix, $\mathbf{W}$ and $\mathbf{D}$ are the same as in Laplacian matrix.

However, both regularizers model the ranking score consistency pair-wisely and have less ability to capture the multiple-wise property of the ranking score consistency. As will be discussed in the next section, local learning based regularizer models the multiple-wise consistency by formulating the score estimation into a learning problem without any heuristic assumptions.

3 Visual Reranking with Local Learning Regularizer

3.1 Local Learning Regularizer

With the ranking score consistency assumption, the desired property of $\mathbf{r}$ is that: for each sample $\mathbf{x}_i$, its ranking score r_i should be consistent with its neighboring samples' scores simultaneously. In Eq. (3) and Eq. (4), this multiple-wise consistency is approximated by accumulating $\mathbf{x}_i$'s consistency with each of its neighboring samples individually.

To reveal the intrinsic multiple-wise consistency, we tackle this problem from the local learning perspective. If a sample's ranking score can be estimated from its neighbors, this sample's multiple-wise consistency with its neighbors is guaranteed. From this point of view, we can model the ranking score consistency from the machine learning perspective. Specifically, for sample $\mathbf{x}_i$, we first learn the desirably consistent score $\hat{r}_i$ from its neighboring samples. By requiring the objective r_i be close to this predicted value $\hat{r}_i$, the multiple-wise consistency is guaranteed. The details will be discussed in the following.

For each sample $\mathbf{x}_i$, a local model $o_i(\cdot)$ is trained locally with the data $\{(\mathbf{x}_j, r_j)\}_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_i)}$, where $\mathcal{N}(\mathbf{x}_i)$ denotes the set of $\mathbf{x}_i$'s neighboring samples. $\mathbf{x}_i$'s ranking score can be predicted by this learned model. Then the regularizer term is derived by aggregating the local model's prediction loss on each sample:

$$Reg_{Local}(\mathcal{G}, \mathbf{r}) = \sum\nolimits_i (r_i - o_i(\mathbf{x}_i))^2. \tag{5}$$

The task of the local model $o_i(\cdot)$ is to predict sample $\mathbf{x}_i$'s ranking score r_i from its neighboring samples accurately. Many approaches can be used as the local model. A linear one is adopted in [12]. However, due to the complexity of the real-world images, it is hard to predict the scores accurately by using simple linear model. To handle this difficulty, we propose to use a local kernel model. Since it is apparently a regression problem, the kernel ridge regression statistical model [2], which is well-known and simple to implement, is adopted in this paper.

In kernel ridge regression, by using a kernel mapping $\phi(\cdot)$ operating from input space $\mathcal{X}$ to a kernel space $\mathcal{F}$, $\phi : \mathbf{x} \in \mathcal{X} \mapsto \Phi(\mathbf{x}) \in \mathcal{F}$, the dependencies between $\mathcal{N}(\mathbf{x}_i)$ and its score vector $\mathbf{r}_i = [r_j]^T_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_i)}$ is modeled as:

$$o_i(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x}). \tag{6}$$

Its cost function is:

$$\sum\nolimits_{j, \mathbf{x}_j \in \mathcal{N}(\mathbf{x}_i)} (r_j - \mathbf{w}^T \phi(\mathbf{x}_j))^2 + \lambda \left\| \mathbf{w} \right\|^2, \tag{7}$$

λ is a coefficient to balance the capacity and complexity of this model.

Differentiating Eq. (7) w.r.t. $\mathbf{w}$ and then equating it to zero, we obtain:

$$\mathbf{w} = \Phi_i (\Phi_i^T \Phi_i + \lambda \mathbf{I})^{-1} \mathbf{r}_i,$$

where Φ_i denotes matrix $[\phi(\mathbf{x}_j)]^T$ for $\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_i)$. Then, for $\mathbf{x}_i$, the score predicted by its local model $o_i(\cdot)$ is:

$$o_i(\mathbf{x}_i) = \mathbf{w}^T \phi(\mathbf{x}_i) = \mathbf{k}^T (\lambda \mathbf{I} + \mathbf{K})^{-1} \mathbf{r}_i = \beta_i^T \mathbf{r}_i, \tag{8}$$

where $\mathbf{k}$ is a vector with $k_j = \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_{t_j}) = k(\mathbf{x}_i, \mathbf{x}_{t_j})$, and $\mathbf{K}$ is a matrix with $k_{mn} = \phi(\mathbf{x}_{t_m})^T \phi(\mathbf{x}_{t_n}) = k(\mathbf{x}_{t_m}, \mathbf{x}_{t_n})$, $j, m, n = 1, 2, \cdots, |\mathcal{N}(\mathbf{x}_i)|$ and $\mathbf{x}_{t_j}, \mathbf{x}_{t_m}, \mathbf{x}_{t_n} \in \mathcal{N}(\mathbf{x}_i)$ with t_p is the subscript of the pth sample in $\mathcal{N}(\mathbf{x}_i)$. As for kernel based methods, we only need to define the kernel function k without defining $\phi(\cdot)$ explicitly. The Gaussian kernel is adopted as the kernel function in this paper.

The local learning regularizer is formulated as:

$$\begin{aligned} Reg_{Local}(\mathcal{G}, \mathbf{r}) &= \sum\nolimits_i (r_i - o_i(\mathbf{x}_i))^2 \\ &= \sum\nolimits_i (r_i - \beta_i^T \mathbf{r}_i)^2 = \mathbf{r}^T \mathbf{R}_{Local} \mathbf{r} \end{aligned} \tag{9}$$

$\beta_i^T = \mathbf{k}^T (\lambda \mathbf{I} + \mathbf{K})^{-1}$. $\mathbf{R}_{Local} = (\mathbf{I} - \mathbf{B})^T (\mathbf{I} - \mathbf{B})$ is the local learning regularization matrix and $\mathbf{B} = [b_{ij}]_{N \times N}$ where b_{ij} equals the corresponding element of β_i if $\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_i)$, otherwise $b_{ij} = 0$.

3.2 Solution

With the local learning regularizer in Eq. (9) as the regularizer term and the preference strength distance in Eq. (2) as the distance term respectively, the objective function of the local learning regularized Bayesian reranking is formulated as:

$$
\begin{aligned}
E(\mathbf{r}) &= Reg_{Local}(\mathcal{G}, \mathbf{r}) + c \times Dist(\mathbf{r}, \bar{\mathbf{r}}) \\
&= \mathbf{r}^T \mathbf{R}_{Local} \mathbf{r} + c \times \frac{1}{2} \sum\nolimits_{i,j,i \neq j} (1 - \frac{r_i - r_j}{\bar{r}_i - \bar{r}_j})^2.
\end{aligned} \tag{10}
$$

The optimal solution $\mathbf{r}^*$ is obtained by minimizing Eq. (10). Denote $\alpha_{ij} = 1/(\bar{r}_i - \bar{r}_j)$, then we can get

$$
\begin{aligned}
\mathbf{r}^* &= \arg\min_{\mathbf{r}} \mathbf{r}^T \mathbf{R}_{Local} \mathbf{r} + c \times \frac{1}{2} \sum\nolimits_{i,j,i \neq j} (1 - \alpha_{ij}(r_i - r_j))^2 \\
&= \arg\min_{\mathbf{r}} \mathbf{r}^T \mathbf{R}_{Local} \mathbf{r} + c \times \frac{1}{2} \sum\nolimits_{i,j,i \neq j} \alpha_{ij}^2 (r_i - r_j)^2 \\
&\quad - c \sum\nolimits_{i,j,i \neq j} \alpha_{ij}(r_i - r_j) \\
&= \arg\min_{\mathbf{r}} \mathbf{r}^T \mathbf{R}_{Local} \mathbf{r} + c \times \mathbf{r}^T \mathbf{L}_A \mathbf{r} - c \sum\nolimits_{i,j,i \neq j} \alpha_{ij}(r_i - r_j) \\
&= \arg\min_{\mathbf{r}} \mathbf{r}^T (\mathbf{R}_{Local} + c\mathbf{L}_A) \mathbf{r} - \tilde{\mathbf{c}}^T \mathbf{r}
\end{aligned} \tag{11}
$$

where $\tilde{\mathbf{c}} = 2c(\mathbf{Ae})$ with $\mathbf{e}$ is a vector with all elements equal 1 and $\mathbf{A} = [\alpha_{ij}]_{N \times N}$ is an anti-symmetric matrix. $\mathbf{L}_A$ is a Laplacian regularization matrix defined over the graph $\mathcal{G}_A$. $\mathcal{G}_A$ has the same structure with $\mathcal{G}$ but the weight between nodes $\mathbf{x}_i$ and $\mathbf{x}_j$ is $|\alpha_{ij}|$ instead of w_{ij}.

Differentiating (11) w.r.t $\mathbf{r}$ and then equating it to zero, it gives:

$$
\begin{aligned}
2(\mathbf{R}_{Local} + c\mathbf{L}_A)\mathbf{r}^* &= \tilde{\mathbf{c}} \\
\mathbf{r}^* &= \frac{1}{2}(\mathbf{R}_{Local} + c\mathbf{L}_A)^{-1} \tilde{\mathbf{c}}
\end{aligned} \tag{12}
$$

Then a closed-form solution for local learning regularized reranking is obtained. The solutions for Bayesian reranking with Laplacian regularizer and normalized Laplacian regularizer have the same form with Eq. (12) only by replacing the $\mathbf{R}_{Local}$ with $\mathbf{L}$ and $\mathbf{L}_n$ respectively.

4 Experiments

4.1 Experimental Setting

To evaluate the effectiveness of the proposed local learning regularizer for reranking, we conducted experiments on two datasets. One is the widely used video search benchmark, *i.e.*, TRECVID 2005-2007 test set [11], which consists of 508 videos and 143,392 shots totally. Its text search baseline is obtained based on the Okapi BM-25 formula [9] using ASR/MT transcripts at shot level. For each of the 72 queries, 24 for each year, the top 1400 shots returned by the search system are used as the initial text search result.

The other dataset is a real-world Web image dataset collected from Google[1]. We selected 29 queries from a commercial image search query log as well as

[1] http://images.google.com/

popular tags of Flickr[2]. For each query, at most top 1000 images returned by Google image search are collected. Thus, there are 24,676 images in total. The initial search result given by Google is regarded as the text search baseline.

For both datasets, the visual feature used is 225-dimentional block-wise colour moments extracted over 5*5 fixed grid partitions. The K-Nearest Neighbor is adopted to find the neighboring samples.

For TRECVID 2005-2007, the video shot's relevance is provided by NIST [11] on two levels, *i.e.*, "Relevant" or "Irrelevant". The most often used performance measure for this dataset is the non-interpolated Average Precision (AP) [1], which is also adopted in this paper. We average the APs over all the 24 queries in each year to get the Mean Average Precision (MAP) to measure the overall performance.

For the Web image dataset, each image's relevance degree is judged by three participants on four levels, *i.e.*, "Excellent", "Good", "Fair" and "Irrelevant", and its final relevance degree is set as the middle one among the scores given by the three judges. The performance on this dataset is measured by the Normalized Discounted Cumulated Gain(NDCG) [5], which is a measure commonly used in information retrieval when there are more than two relevance levels.

4.2 Experimental Results and Analysis

We compare the local learning regularizer with the other two popular ones, *i.e.*, Laplacian regularizer and normalized Laplacian regularizer, in the Bayesian reranking framework. Bayesian reranking with the three different regularizers are denoted as LocalReg, LapReg and NLapReg respectively. We also compare LocalReg with another well-known graph-based method - VisualRank [6], which also measures the consistency pair-wisely, as discussed in [10]. The parameters are globally set for each method respectively.

The overall performance on the two datasets are summarized in Table 1 and Fig. 1 respectively. We can see that LocalReg outperforms LapReg, NLapReg as well as VisualRank over the two datasets consistently and stably. This result

Table 1. MAP comparison for different reranking methods on TRECVID 2005-2007

Method	TRECVID 2005		TRECVID 2006		TRECVID 2007	
	MAP	Gain	MAP	Gain	MAP	Gain
Baseline	0.0441	-	0.0381	-	0.0306	-
VisualRank	0.0506	14.74%	0.0401	5.25%	0.0333	8.82%
LapReg	0.0487	10.43%	0.0461	21.00%	0.0465	51.96%
NLapReg	0.0534	21.09%	0.0434	13.91%	0.0471	53.92%
LocalReg	**0.0583**	**32.20%**	**0.0497**	**30.45%**	**0.0485**	**58.50%**

[2] http://www.flickr.com/

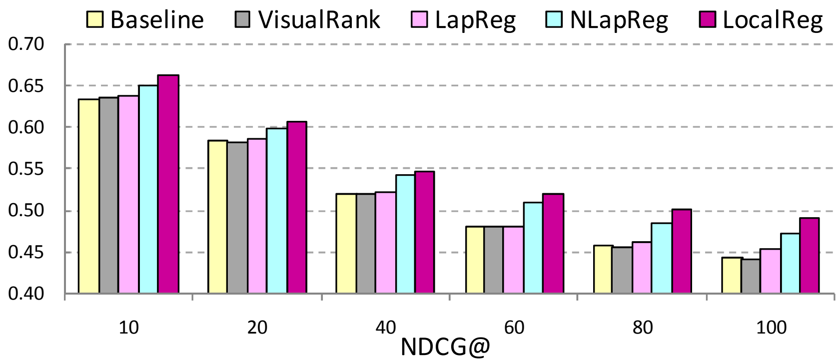

Fig. 1. NDCG comparison for different reranking methods on Web image dataset

demonstrates that pair-wise approximation of the consistency over the graph is not precise enough. The local learning based regularizer captures the multiple-wise consistency and thus provides more satisfactory results.

Besides the overall performance, we also investigate the effectiveness of LocalReg over each query, as shown in Fig. 2. Here, due to the space limitation, we only take TRECVID 2005 for illustration. From Fig. 2, we can see that most of the queries benefit from LocalReg after reranking and some queries show significant gains, such as query 152[3], 156[4], 169[5] and 171[6]. By further comparing LocalReg with other three reranking methods, we find that although the other ones achieve high improvements on certain queries, *e.g.*, query 154[7] for

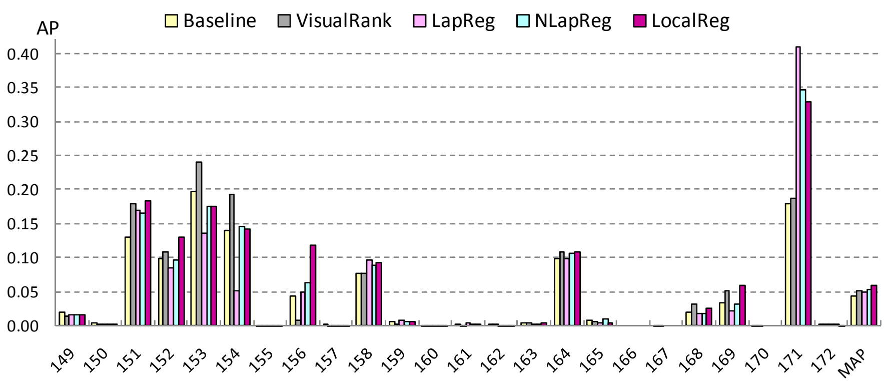

Fig. 2. Performance comparison over each query on TRECVID 2005

[3] Find shots of Hu Jintao.
[4] Find shots of tennis players on the court.
[5] Find shots of George Bush entering or leaving a vehicle.
[6] Find shots of a goal being made in a soccer match.
[7] Find shots of Mahmoud Abbas.

VisualRank, query 171 for LapReg and NLapReg, they show dramatic performance decrease on other queries, *e.g.*, query156 for VisualRank and query 154 for LapReg. In contrast, LocalReg presents stable performance improvements on most of queries with slight performance decrease on few ones. This phenomenon further demonstrates the superiority of the multiple-wise consistency derived from local learning regularizer.

5 Parameter Sensitivity

In the local learning regularizer, there are two important parameters, *i.e.*, the number of nearest neighbors K for local model learning and the trade-off parameter λ in Eq. (7) of kernel ridge regression. In this section, we will analyse the sensitivity of LocalReg to them, by taking experiments conducted on the TRECVID dataset as illustration.

5.1 The Number of Nearest Neighbors K

The number of nearest neighbors K plays an important role in local learning regularizer. With too few neighbors, over-fitting is likely to occur and thus the learned model gives poor prediction for the centroid sample. A larger K can ensure more samples are involved for local model training. However, the samples which are far away from the centroid sample have less to do with it and thus gives little contribution to the local model. Besides, more noises will be introduced at the same time when enlarging the training size. A proper K is required to achieve a good performance.

The MAP-K curves are presented in Fig. 3. As illustrated, larger Ks are preferred on both TRECVID 2005 and 2007, while on TRECVID 2007 a small K gives better performance. As analysed from the data, the average numbers of relevant samples across queries are 41, 55 and 24 for TRECVID 2005 - 2007 respectively. We can observe that the optimal K is roughly proportional to the average number of relevant samples. TRECVID 2007 has the fewest relevant samples, thus, a smaller training set is preferred to prevent too many noises are involved. This observation can provide a rough guideline in setting K in the future practical application.

5.2 The Trade-Off Parameter λ

We also investigated the influence of the trade-off parameter λ in kernel ridge regression. Figure 4 shows the performance of LocalReg with different λ in terms of MAP on TRECVID 2005-2007.

From the figures, we can see that, compared with the text search baseline, the performance is improved with variant λs consistently. For further observation, we can find that moderate, larger and smaller λs are preferred on the three years respectively. The reason is that, on TRECVID 2007, the local learning problem is the hardest (with the lowest baseline and the fewest relevant samples) thus a

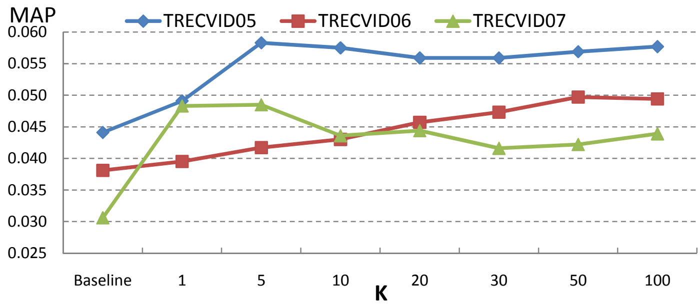

Fig. 3. The performance of LocalReg with different K on TRECVID dataset

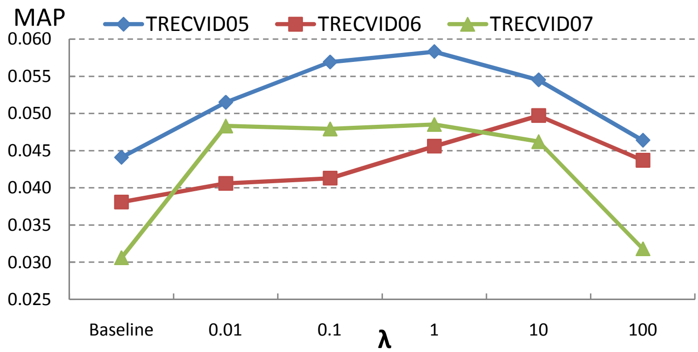

Fig. 4. The performance of LocalReg with different λ on TRECVID dataset

complex model is required with small λ. For TRECVID 2006, it has the most relevant samples and the local learning problem is the easiest. A simple model is effective enough and there for a larger λ is preferred. On TRECVID 2005, although it has the highest baseline, its relevant samples are fewer than that of TRECVID 2006. As a consequence, a moderate λ is optimal for it.

6 Conclusion

By investigating the effect of the regularizer term in graph-based visual reranking, an effective regularization approach, *i.e.*, local learning regularizer, is adopted in this paper from the local learning perspective. The local learning regularizer models the multiple-wise consistency by formulating the score estimation into a learning problem without any heuristic assumptions. Extensive experiments conducted on both TRECVID dataset and real Web image dataset demonstrate the superiority of the local learning regularizer.

References

1. Trec-10 proceddings appendix on common evaluation measures, http://trec.nist.gov/pubs/trec10/appendices/measures.pdf
2. Cristianini, N., Shawe-Taylor, J.: An introduction to support vector machines and other kernel-based learning methods. Cambridge University, Cambridge (2000)
3. Hsu, W.H., Kennedy, L.S., Chang, S.-F.: Video search reranking via information bottleneck principle. ACM Multimedia, 35–44 (2006)
4. Hsu, W.H., Kennedy, L.S., Chang, S.-F.: Video search reranking through random walk over document-level context graph. ACM Multimedia, 971–980 (2007)
5. Järvelin, K., Kekäläinen, J.: Ir evaluation methods for retrieving highly relevant documents. In: SIGIR, pp. 41–48 (2000)
6. Jing, Y., Baluja, S.: Visualrank: Applying pagerank to large-scale image search. IEEE Transactions on PAMI 30, 1877–1890 (2008)
7. Liu, J., Lai, W., Hua, X.-S., Huang, Y., Li, S.: Video search re-ranking via multi-graph propagation. ACM Multimeida, 208–217 (2007)
8. Liu, Y., Mei, T., Hua, X.-S., Tang, J., Wu, X., Li, S.: Learning to video search rerank via pseudo preference feedback. In: ICME, pp. 297–300 (2008)
9. Robertson, S.E., Walker, S., Hancock-Beaulieu, M., Gatford, M., Payne, A.: Simple, proven approaches to text retrieval. Cambridge University Computer Laboratory Technical Report TR356 (1997)
10. Tian, X., Yang, L., Wang, J., Yang, Y., Wu, X., Hua, X.-S.: Bayesian video search reranking. ACM Multimedia, 131–140 (2008)
11. TRECVID. TREC Video Retrieval Evaluation, http://www-nlpir.nist.gov/projects/trecvid/
12. Wu, M., Schölkopf, B.: Transductive classification via local learning regularization. In: 11th International Conference on AISTATS, pp. 624–631 (2007)
13. Yan, R., Hauptmann, A.G., Jin, R.: Multimedia search with pseudo-relevance feedback. In: Bakker, E.M., Lew, M., Huang, T.S., Sebe, N., Zhou, X.S. (eds.) CIVR 2003. LNCS, vol. 2728, pp. 238–247. Springer, Heidelberg (2003)
14. Zhou, D., Bousquet, O., Lal, T.N., Weston, J., Schölkopf, B.: Learning with local and global consistency. In: NIPS (2003)
15. Zhu, X., Ghahramani, Z., Lafferty, J.D.: Semi-supervised learning using gaussian fields and harmonic functions. In: ICML, pp. 912–919 (2003)