Proceedings of the 2012 International Conference on Machine Learning and Cybernetics, Xian, 15-17 July, 2012

GENERATING DIVERSE AND RELEVANT IMAGE SEARCHING RESULTS WITH DIVRANK

JING LI, ZHONG JI*, JING ZHANG, YU-TING SU

School of Electronic Information Engineering
Tianjin University, Tianjin, PR China
Email: jizhong@ tju.edu.cn

Abstract:

Nowadays, web-scale image search engines (e.g. Google Image Search, Bing Image Search) rely almost purely on surrounding text features. This leads to ambiguous and noisy results. Moreover, most of existing ranking methods for image search often return results according to its relevance with the query, leaving its diversity aside. In order to address these problems, in this paper, we proposed a GDRID (Generating Diverse and Relevant Image results with DivRank) visual search reranking algorithm, which extends the DivRank algorithm to enhance the diversity as well as relevance of the initial search results with visual information. DivRank is based on a reinforced random walk in an information network which can automatically balance the prestige and diversity of the top ranked vertices in a principle way. We evaluate GDRID by using empirical experiments on two popular image data sets, MSRA-MM and Bing reranking data sets. Experimental results outperform existing network-based ranking methods in terms of enhancing diversity in prestige.

Keywords:

Visual search reranking; DivRank; Relevance reranking; Diversity reranking.

1. Introduction

As World-Wide-Web grows in an exploding rate, the amount of multimedia data has been increasing explosively [1]-[2], [26]. Search engines become into indispensable tools to users. Results obtained by using keyword matching are often irrelevant because the text around the images does not always reflect image content [5]. In order to rerank the text-based search results, visual search reranking technique begins to draw a lot of researchers' attention. Based on original text search results, image reranking reorders the output result through digging the intrinsic relationship of data or drawing on external knowledge to generate a new order list which would better meet the user search needs.

The basic idea of reranking is extracting visual information from the original images results and adjusting the primitive orders. Generally, the methods can be divided into three categories: classification-based method [6], clustering-based method [7]-[8], and graph theory-based method [9] - [11]. PRF (Pseudo Relevance Feedback) [6] hypothesis that most of correct outcomes (also known as positive samples) distribute in the front part of the primitive search results while error outcomes (negative samples) are in the rear part. The authors then use the samples to train a sorting function which later would be used to rerank the original results. Hsu et al. [8] applied the information bottleneck theory and got each class's conditional probability after clustered the outcome returned from the text based search. The class with larger probability means it is more relevant to the query while the one with smaller probability would be considered as noise. Hsu et al. [9] introduced graph theory in order to solve reranking problem, and used random walk to update each sample's relevance score which would decide every sample's final ranking position.

Generally speaking, ranking includes relevance ranking and diversity ranking. Relevance reranking is regarded as the bedrock of information retrieval. It is reasonable to assume that the effective of the system would be maximized if the returned results to each query are in the order of decreasing probability of relevance. However, the necessity of diversity may not seem so intuitive comparing to the relevance, but its importance has been long acknowledged in the information retrieval [13]-[14]. In most of cases, users may not be able to precisely and exhaustively describe what they need. Therefore it might be more helpful if the results cover as many aspects as possible. For example, when a user provides an ambiguity query such as "Paris", he/she may refer to different topics, like a celebrity, a city name, etc. Images that returned by querying one keyword usually can't contain all aspects, and the structure of returned image set

978-1-4673-1487-9/12/$31.00 ©2012 IEEE

usually is much more complicated than expected. Take the query "Paris" for instance, it may return different topics, such as Paris Hilton (an famous American socialite), Paris Eiffel Tower, Paris' map, Paris Notre dame, which are illustrated in figure 1. Since the relevance scores are only determined by their relevance to the query, and of course would miss some sub-topics inevitably. In some other cases, the users can not fully describe what they need, and usually just provide simple words. For example, when a user only provides a simple query "ball", he/she may actually want to find a colorful baseball. In this case, a diverse image set is much more needed than just a normal set of images. However these two measures including relevance and diversity are often difficult to maximize simultaneously. In this paper, we extend a text reranking technique which tries to cope with these two aspects and translate this principle into images. The key idea of our approach is the conjunction of image similarity estimation and a vertex-reinforced random walk hypothesis, idea borrowed from Mei's work [15]. The basic idea is that the transition would be reinforced by the number of previous visits to that state.

Figure 1. Images relevant to query "Paris".

The paper is organized as follows. We provide a short review on the related work in Section 2. In Section 3, we make a brief description of the GDRID algorithm. Experimental results are presented in section 4 and section 5 concludes the paper.

2. Related work

Image reranking can be performed using textual information associated with images, visual description or a combination of the two. Grangier et al. [16] proposed a kernel-based discriminate model on the image reranking. The authors used the training data to train the classifier. But built the training data and labeled them is a time-consuming and labor-intensive work. Jing et al. [4] proposed VisualRank and applied the classic PageRank algorithm to the image ranking. [17]-[19] are based on multi-modal and consider both visual content features and text information. Schroff et al. [17] first used image's text information such as image's ATL tag, image's file name to train a Bayesian posterior probability model for sorting, and then used visual features to train a SVM classifier to further optimize. Introducing relevance degree information in dimensionality reduction technique, Ji et al. [18] proposed a Rank-CCA algorithm for image reranking, which only employed low-level and low-dimensional visual features, and achieved state-of-the-art performance. Richter et al. [19] built a image ranking model which used manually annotation and visual content features to compute the similarity which would be used to form a graph. Then they ranked the images by using random walk algorithm in the graph.

It has been acknowledged that diversity also plays important role in information retrieval [20]-[21]. For example, Li et al. [21] proposed Dual-Rank method to improve web image retrieval results. The authors introduced a graph model to model the relationship between the images, visual features and image quality, and transform it into a constrained multi-objective optimization problem. Then Dual-Rank use both Inter-cluster Ranking which ranks within the cluster and Intro-cluster Ranking which ranks between the clusters. However, it is very hard to let the computer recognize the structure of the cluster and accurately estimate the clustered number. Under some circumstances, even artificially determined the number of the clustering is not an easy job.

The methods relying on text information such as users' annotation are performing not well, and consider the fact that the annotation usually is subjective and fuzziness. Researchers begin to focus on improving the annotation's quality by analyzing and mining the visual content. [22]-[23] used the visual content features to learn the image annotation and the image correlation. In [22], the authors used the nearest neighbor voting strategy to determine which tag should be related to which image. In [23], Liu et al. first built a graph with images related to one query, and then estimate the image similarity by using graph knowledge. Wang et al. [24] performed an excellent work in balancing diversity and relevance in image reranking. They proved that relevance reranking can be regarded as the process of optimizing the mathematical expectation of the conventional Average Precision (AP) measure. Besides, they also proved that diversity reranking can be viewed as an optimization process to a new Average Diversity Precision (ADP) measure. Ji et al. [3] diversified the image relevance reranking with absorbing random walks, which had been proved to be effective.

3. Image reranking with the proposed GDRID algorithm

DivRank is based on a similar reinforced random walk known as vertex-reinforced random walk in an information network which balances both the prestige and diversity of the top ranked vertices in a principle way. It initially applied in text documents summarization. We borrow the idea proposed in [15], whose main idea is unlike the time-homogeneous random walks that remains constant over time, and we believe the transition probability to one state from others is reinforced by the number of previews visits to that state so as to form a rich get richer phenomenon and "absorb" the score of neighbors, thus encouraging diversity.

3.1. DivRank modal

Given a query q, text-based image search engine would produce a list of search results. We take the top n images as the candidate set $S = \{s_1, s_2, \ldots, s_n\}$, S would cover a certain amount of sub-topics relevant to q in usual case, i.e. $T = \{t_1, t_2, \ldots, t_n\}$. Let $G = (V, E)$ be a weighted graph built by candidate images. V refers to a set of candidate images and E refers to the ties between images. We define the weight of an edge using $w(u, v)$, and use the image similarities to represent $w(u, v)$. Since the G is an undirected graph, therefore $w(u, v)$ must not take any negative value.

3.2. Reranking with DivRank

The general form of DivRank is described as below:

$$p_T(u,v) = (1-\lambda)\cdot p^*(v) + \lambda \cdot \frac{p_0(u,v)\cdot N_T(v)}{D_T(u)} \quad (1)$$

$$D_T(u) = \sum_{v\in V} p_0(u,v) N_T(v) \quad (2)$$

where $\boldsymbol{P}_0 = \beta\boldsymbol{P} + (1-\beta)\boldsymbol{I}$, $0\le\beta\le1$. $P_T(u, v)$ is the transition probability from state u to state v at time T. $p^*(v)$ represents the prior preference distribution of visiting vertex v. The images ranked higher in the original are set with larger probability, all the probabilities satisfy $p^*(v) \ge 0$, $\sum_{v\in V} p^*(v) = 1$. $P_0(u,v)$ is the primitive transition matrix prior to any reinforcement, even can be considered as in a regular time-homogeneous random walk, and the transition probabilities will be boosted by the certain amount of visits to every vertex. $\boldsymbol{P}$ is the organic transition matrix acquired from the adjacent relationship of a weighted network. $\boldsymbol{I}$ is an identity matrix to forge self-links, which in $\boldsymbol{P}_0$ help to prevent the vertices from losing the profit already acquired during the reinforcement. $N_T(v)$ records the visiting times of corresponding object which acts as the reinforcing factor during the random walk process. Matrix $D_T(u)$ is to re-normalize $\boldsymbol{P}_0\boldsymbol{N}_T$ into a transition matrix $\boldsymbol{P}_T$ and to make sure the process will eventually converge. The reinforced random walk defined above by Equation 1 converges to a stationary distribution $f(v)$ And $\sum_{v\in V} f(v) = 1$. Which

$$f(v) = \sum_{v\in V} p_t(u,v) f(u), \forall t \ge T. \quad (3)$$

Random walk ranks items according to their stationary distributions, therefore the center group would dominant the top ranked items. However, Random walks doesn't show any sign of diversity, on the contrary, in DivRank, nodes with a higher degree will get a higher weight, which in turn results in a larger accumulative N_T. And the nodes already have a high N_T tend to get an even higher weight. Also the self-link to a vertex makes sure that even if all the neighbors shrink, the vertex could still be large as long as N_T is large so that it can present good diversity.

The complete DivRank in GDRID algorithm is given in figure 2.

4. Experiment and results

In this section, we evaluate the effectiveness of GDRID We demonstrate the experiment on an image database MSRA-MM_V1.0 [12], which covers 68 groups of images and another image database Bing Reranking[1] collected from the Bing Image search engine using 120 query keywords in July, 2010. Our task is to find the most relevant and diverse images from the image network and database when given a query.

Input: $\boldsymbol{W}$, $\boldsymbol{r}$, λ
Output: a list L of n items according to diversity and relevance

---

1: Set $L = \varphi$.

// Calculate the primitive transition matrix $p_T(u,v)$.

2: Build the transition matrix P_T from the organic transition matrix P_0 acquired from the Graph.

3 Initialize the stationary distribution $\pi(v)$ and set it to $1/N$, N is the number of the total image.

4: while ($\pi(v)$ doesn't converge).

5: Compute the number of times N_T the walker has visited node v up to time T.

6: Calculate the transition matrix $P_T(u,v)$ at time T by using the N_T.

7: End while and return $\pi(v)$.

Figure 2. DivRank in GDRID algorithm

We estimated similarity between images by using the corresponding feature vectors which represent the two images. Then the distance between two images can be defined as below:

$$d(i,j) = \exp(-\frac{1}{N}\sum_{k=1}^{N} d_k(i,j)/\beta) \tag{4}$$

$$d_k(i,j) = \left\| \frac{f_{ik}-\mu_k}{\sigma_k} - \frac{f_{jk}-\mu_k}{\sigma_k} \right\|^2 \tag{5}$$

where k represents the kth dimensional feature, N is the feature dimension, β is a constant number, μ_k and σ_k are its mean and variance respectively. We adopt ADP (Average Diverse Precision) which proposed and proved to be effective in [24] to evaluate the performance.

The ADP is defined below:

$$ADP(\tau, D) = \frac{1}{R}\sum_{j=2}^{n} y(\tau(j))Div(\tau(j)) \times \left(\frac{1}{j}\sum_{k=2}^{j} y(\tau(k))Div(\tau(k))\right) + a(1) \tag{6}$$

$$Div(\tau(k)) = \min_{1\le t<k}(1 - d(\tau(t),\tau(k))) \tag{7}$$

$$a(1) = \begin{cases} 1 & if \quad y(\tau(1)) = 1 \\ 0 & if \quad y(\tau(1)) = 0 \end{cases} \tag{8}$$

where $D=\{x_1,x_2,...,x_n\}$ represents a collection of images. $y(x_i)$denote the binary relevance label of x_i, it is 1 if the image at the reranked position i is relevant to the query, otherwise, it is 0 .τ *is* the ordering of images, R represents the number of true relevant images in the D. $d(\tau(t), \tau(k))$ measures the distance between images t and k. Comparing to AP(Average Precision), ADP takes both relevance and diversity into consideration by adding the diversity factor $(Div(\tau(k)))$.

Different from the probability r defined in [24], we extend the relevance score function as follows:

$$r(t) = \frac{h(t)}{\sum_{t=1}^{n} h(t)} \tag{9}$$

$$h(t) = \frac{2e^{-(t-1)/z}}{1+e^{-(t-1)/z}} \tag{10}$$

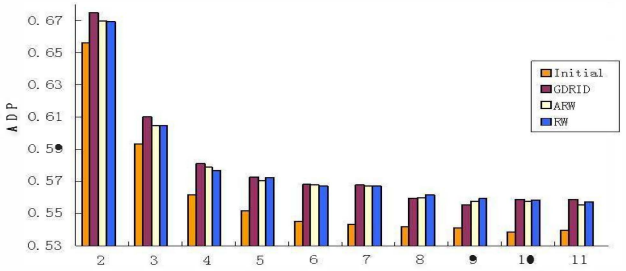
where t is the initial position number, z can take 50,100,200. Here we set it to 50 and the sum of it is 1.

We compare our method with personalized PageRank [25] with a simple prior. We perform reranking tasks on two image databases.

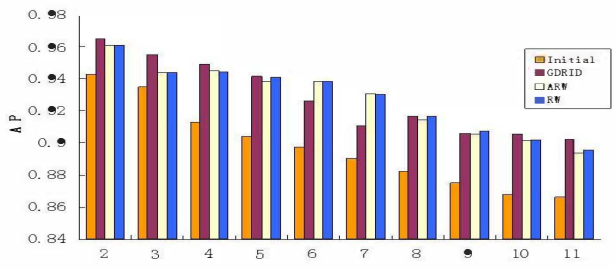
4.1. MSRA-MM_V1.0 dataset

Our first experiment considers an image database MSRA-MM_V1.0. The dataset contains two sub-datasets, i.e., an image dataset and a video dataset that are collected from Microsoft Live search. The image part contains about 1 million images from 1165 queries and video part contains 23 thousands of videos. Figure 3 illustrates the average ADP and AP performances of 68 queries in different depth (the number of top images). Since the ADP and AP in depth 1 are too high comparing to the others, therefore, we observe the depth from 2 to 11. We can find that GDRID outperforms ARW (grasshopper) [25] and RW (random walk) [9] both on relevance (as shown in AP, the higher the better) and diversity (as shown in ADP, the higher the better). Figure 4 illustrates each query's ADP and AP comparisons in depth 10 (we take 10 query results for example). We can see that the GDRID perform much better than ARW and RW both in diversity and relevance.

[1]http://mmlab.ie.cuhk.edu.hk/CUHKSR/Dataset.htm

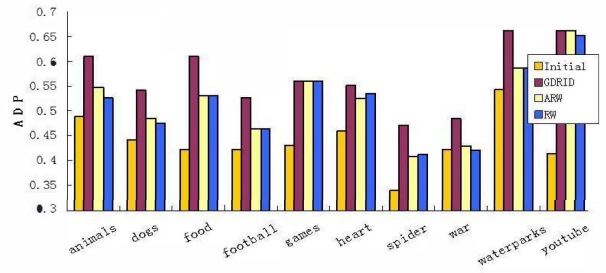


(a) ADP results

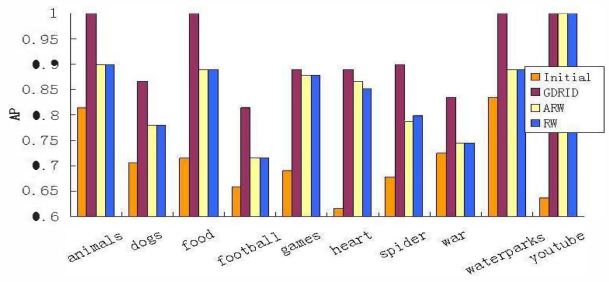


(b) AP results

Figure 3. ADP and AP performance comparisons in different depth with MSRA-MM dataset



(a) ADP results



(b) AP results

Figure 4. Example of ADP and AP performance comparisons in depth@10

4.2. Bing dataset

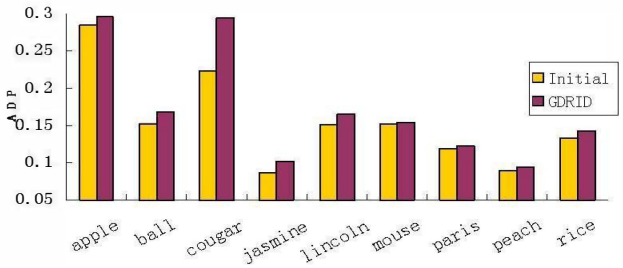
Our next experiment considers Bing data set. It includes 120,000 labeled images of around 1500 categories retrieved by the Bing Image Search using 120 query keywords. In this experiment, since the dataset provide no features, we extract BOW (bag of visual words) as the image visual features. Figure 5 illustrates the top 16 results of query "Pairs" about ARW and GDRID. We can see that both the ARW and GDRID present diversity results. Figure 6 illustrates the ADP and AP performance in BING data sets, which shows that GDRID performs better in diversity and relevance than the original. We set β to 0.2 and λ to 0.9 in both experiments.

(a) Initial ranking results

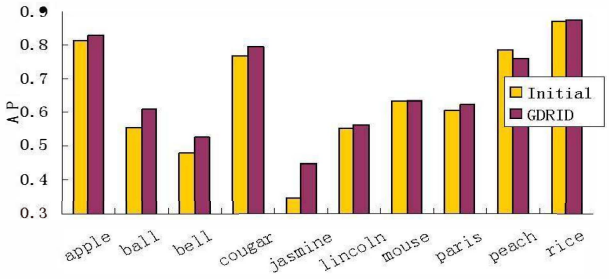
(b) ARW reranking results

(c) GDRID reranking results

Figure 5. Top 16 results of query "Paris" before and after reranking



(a) ADP results



(b) AP results

Figure 6. ADP and AP comparisons in BING data sets

5. Conclusions

In this paper, we present a novel image reranking approach which simultaneously takes both relevance and diversity into consideration. We extend vertex-reinforced random walk to image domain, utilizing visual features to reorder the original image search results, which has been proved to be very effective.

The reasonable next step is to extract more effective visual information such as the information in the saliency region of the image rather than the whole area. And then we could build a more accurate graph modal to uncover the intrinsic relationship of each image in a certain image data set.

Acknowledgements

The paper is supported by the National Natural Science Foundation of China (No. 61170239), the Innovation Foundation of Tianjin University (No.60302019).

References

[1] M. Wang, X.S. Hua, J.H. Tang, R.C. Hong, "Beyond Distance Measurement: Constructing Neighborhood Similarity for Video Annotation", IEEE Transactions on Multimedia, vol. 11, no. 3, pp. 465-476, 2009.

[2] M. Wang, X.S. Hua, R.C. Hong, et al., "Unified Video Annotation Via Multi-Graph Learning", IEEE Transactions on Circuits and Systems for Video Technology, vol. 19, no. 5, pp. 733-746, 2009.

[3] Z. Ji, Y.T. Su, X.J. Qu, Y.W. Pang, "Diversifying the image relevance reranking with absorbing random walk", The sixth International Conference on Image and Graphics, pp. 981-986, Hefei, China, 2011.

[4] Y.S. Jing, S. Baluja, "VisualRank: Applying Pagerank to Large-scale Image Search", IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1877-1890, 2008.

[5] L. Kennedy, M. Naaman, "Generating Diverse and Representative Image Search Results for Landmarks", International World Wide Web Conference, Beijing, 2008.

[6] R. Yan, E. Hauptmann, R. Jin, "Multimedia search with pseudo-relevance feedback", ACM International Conference on Image and Video Retrieval, pp. 238-247, 2003.

[7] N. Ben-Haim, B. Babenko, S. Belongie, "Improving web-based image search via content based clustering." SLAM, New York, 2006.

[8] W.H. Hsu, L.S. Kennedy, S.F. Chang, "Video search reranking via information bottleneck principle", ACM International Conference on Multimedia, pp. 35-44, 2006.

[9] W.H. Hsu, L.S. Kennedy, S.F. Chang. "Video search reranking through random walk over document-level context graph", ACM International Conference on Multimedia, pp. 971-980, 2007.

[10] H. Zitouni, S. Sevil, D. Ozkan, P. Duygulu, "Re-ranking of web image search results using a graph algorithm", IEEE International Conference on Pattern Recognition, pp. 1-4, Dec 2008.

[11] X. Tian, L. Yang, J. Wang, Y. Yang, X. Wu, X.S. Hua, "Bayesian video search reranking", ACM International Conference on Multimedia, pp. 131-140, 2008.

[12] M. Wang, L.J. Yang, X.S. Hua, "MSRA-MM: bridging research and industrial societies for multimedia information retrieval", Microsoft Technical Report (MSR-TR-2009-30), pp. 1-14, 2009.

[13] W. Goffman, "A searching procedure for information retrieval." Information Storage Retrieval, vol. 2, no 2, pp. 73-78, 1964.

[14] B.M. King, E.W. Minium, "Statistical Reasoning in Psychology and Education", Wiley, New York, 2003.

[15] Q. Mei, J. Guo, D. Radev, "Divrank: the interplay of prestige and diversity in information networks", Proceedings of the 16th ACM SIGKDD, pp. 1009-1018, 2010.

[16] D. Grangier and S. Bengio, "A discriminative kernel-based approach to rank images from text queries", IEEE Transactions on Patten Analysis and Machine Intelligence, pp. 1371-1384, 2008.

[17] F. Schroff, A. Criminisi, and A. Zisserman, "Harvesting image databases from the web", IEEE International Conference on Computer Vision, Rio de Janerio, Brazil, 2007.

[18] Z. Ji, P.G. Jing, Y.T. Su, et al., "Rank canonical correlation analysis and its application in visual search

reranking", Signal Processing (2012), http://dx.doi.org/10.1016/j.sigpro.2012.05.006.

[19] F. Richter, S. Romberg, et al, "Multimodal Ranking for Image Search on Community Databases", ACM International conference on Multimedia Information Retrieval, New York, USA, pp. 63-72, 2010.

[20] V.L. ReinierH, G. Lluis, O. Ximena, and V.Z. Roelof, "Visual diversification of image search results", International conference on World Wide Web, pp. 341-350, 2009.

[21] P.J. Li, L. Zhang, J. Ma, "Dual-ranking for Web Image Retrieval", ACM International Conference on Image and Video Retrieval, pp. 166-173, Xi'an, China, 2010.

[22] X. Li, C.G. Snoek, and M. Worring, "Learning tag relevance by neighbor voting for social image retrieval", ACM International Conference on Multimedia Information retrieval, pp. 180-187, New York, USA, 2008.

[23] D. Liu, X.S. Hua, L. Yang, M. Wang, and H.J. Zhang, "Tag ranking", International Conference on World Wide Web, pp. 351-351, April 2009.

[24] M. Wang, K.Y. Yang, X.S. Hua, et al, "Towards a Relevant and Diverse Search of Social Images", IEEE Transactions on Multimedia, 12(8) , pp. 829-842, 2010.

[25] X.J. Zhu, A. Goldberg, J.V. Gael, et al, "Improving diversity in ranking using absorbing random walks", Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics, pp. 1-8, 2007.

[26] M. Wang, X.S. Hua, T. Mei, R.C. Hong, G.J. Qi, Y. Song and L.R. Dai, "Semi-Supervised Kernel Density Estimation for Video Annotation", Computer Vision and Image Understanding, vol. 113, no. 3, pp. 384-396, 2009.