A Survey On Visual Search Reranking

Nishana S S
Dept. Computer Science & Engineering
SCT College of Engineering
Trivandrum,India
nishanalilu@gmail.com

Subu Surendran
Associate Professor, Dept. Computer Science & Engineering
SCT College of Engineering
Trivandrum,India
subusurendran@gmail.com

Abstract—Due to the explosive growth of online video data and images , visual search is becoming an important area of research. Most existing approaches used text based image retrieval which is not so efficient. To precisely specify the visual documents, Visual search reranking is used. Visual search reranking is the rearrangement of visual documents based on initial search results or some external knowledge inorder to make the search efficient. Here we are making a survey of three different reranking methods 1) Reranking via Random walk over document level context graph 2) Reranking via Minimum Incremental Information Loss and 3) Reranking via Pairwise Learning and make a comparative study of it.

Keywords—Visual search, Reranking, Context graph, Pairwise learning, Optimization, Mutual information

I. INTRODUCTION

Nowadays due to the continuous growth of online video data, image and video retrieval have becoming an active area of research. Most existing approaches for image retrieval is based on the text based approach. To correctly specify an image an alternative method is needed. This has lead to the idea of visual search reranking. Visual search reranking means rearranging the visual documents based on the initial search result or some prior ideas or knowledge extracted from it so as to increase the search precision.

Based on how the knowledge is extracted there are three different reranking methods – self-reranking[1], example-reranking[2] and crowd reranking[7]. Self-reranking can be considered as an unsupervised learning in which reranking is based on some initial search results or patterns or characters identified from it. No external knowledge is used here. Example reranking is based on query example provided by users and crowd reranking is based on crowd sourcing knowledge extracted from web. In self rereanking ambiguity problem arises since it depends entirely on the text. The ambiguity problem can be solved by providing examples. Ambiguity problem arises in crowd reranking too.

In this paper we are discussing about three different types of reranking that uses different techniques - Reranking via Random walk over document level context graph[1], Reranking via Minimum Incremental Information Loss (MIIL)[2] and Reranking via Pairwise Learning[3]. Of these random walk reranking is a self reranking method. MILL is an example reraking method and pairwise learning utilize both example reranking and crowd reranking. We go through these three papers and study what techniques are used for reranking and find how efficient each method is.

The rest of the paper is organized as follows. Section II introduces the different types of reranking methods, section III presents a comparitive study of the three methods and section IV concludes the paper.

II. TYPES OF RERANKING

A. Reranking via random walk over document level context graph

This method perform reranking as a random walk over document level context graph. Context graph is a graph with nodes represents documents and edges between them represents the multimodal contextual similarity[6] between two documents.Assume that we have N nodes which represents the video stories. The N nodes are the N documents obtained in the initial search results. The graph traversal is initialized from one node and based on the multimodal similarity between the documents and the original text scores in the initial search result , it traverse to the next node. To govern the transition of the random walk, a transition matrix [Pij] is used. Pij represents the probability of transition from one node to other. At each instance, calculate the state probability of each node.The state probability at time instance k is denoted as $x_k=[P_{(k)}(i)]$.

Consider two nodes i and j. s(i) and s(j) represents the initial text search scores. Pij is the probability of reaching from node i to node j. Mj represents edges connected to node j.

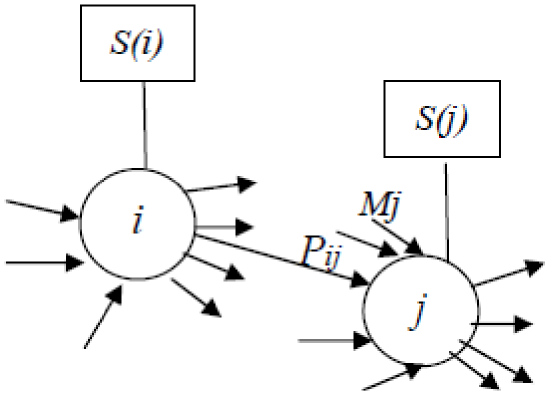

Figure 1: Example of a context graph for random walk

The state probability of node j at time instance k is defined as follows:

$$x_{(k)}(j) = \alpha \sum_{i\in Mj} x(k-1)\,(i)Pij + (1-\alpha)\,S(j)$$

where $\alpha \in [0,1]$ linearly weights two terms. Thus the state probability of the node j at time instance k depends on the state probability of all nodes incident on the node j at time instance k – 1 and the initial text score.

Let at a particular instance T some nodes converge to node j and further on any instance T+1,T+2,....,∞, there is no more nodes to converge to node j. Then we can say that the state probability of the node j at time instance T is the stationary probability of that node.

$$x_{(T)}(j) = \alpha \sum_{i\in Mj} x(T)\,(i)Pij + (1-\alpha)\,S(j)$$

Thus the stationary probability of all nodes are calculated and this stationary probability is taken as the new text relevance score. Based on these scores the documents are reranked.

B Reranking via Minimum Incremental Information Loss(MIIL)

This video search reranking involves two parts- Learning and Reranking. In the learning process several query examples are provided for each textual query.This is the first pairwise approach for visual search reranking. The example images are paired with samples randomly selected from initial search result. The objective of learning is to find out the relevant and irrelevant information. Concept detection[9] is processed on this example pairs to form the relevant and irrelevant information.

In the reranking process based on the relevant and irrelevant information an optimal pair set is obtained by an optimization based technique. Assume we have an initial ranked list X = {xi| i=0,1,....N-1} where x represents the samples. We first convert the samples into a pair set T={tij| i,j = 0,1,2......N-1}. That means every sample in the initial ranked list is paired with every other samples in the list. If the sample order is considered , we get the pair set

T'={tij| i,j = 0,1,2......N-1;xi>xj}. xi > xj means xi is ranked higher than xj in the initial ranked list. To find out the optimal pairset this technique use the basic idea of mutual information. Mutual information measures the amount of information that one variable contains the other[10].

Let t be an element of the pair set T', we have to find out the Mutual information between the element t and the relevant(Y+) and irrelevant(Y-) information as defined:

$$MI(t,Y) = P(t).\sum_{y\in Y} P(y/t)\,log\frac{P(y/t)}{p(y)}$$

Where P(t) denotes the prior probability and $P(y/t)$ denotes the posterior probability of the pair samples.

To perform reranking, MIIL use the optimization technique which maximizes the mutual information between the pairs and relevant information and at the same time minimizes the mutual information[10] between the pair and irrelevant information. This can be approached by maximizing the Weighted Difference D(T').

$$D(T') = \sum_{t\in T'} MI(t,Y+) - \lambda MI(t,Y-)$$

where λ represents the trade of between the preservation of relevant information and l oss of irrelevant information.

MILL reranking utilize the idea of lossy information compression theory. It views reranking as denoising problem, where noise is the incompressible part in the data and the relevant information forms the compressible part[11]. Here the best possible pair is selected at each round. The best possible pair is the one which maximizes the weighted difference. Then minus all other pairs formed by atleast one of the elements of the optimal pairset. Thus a new pair set is formed. Then at each round, map the selected pair to the new ranked list. At the ith round the elements of the pairs are located at rank i and rank N-i+1 where N represents the number of elements in the initial ranked list.

C Reranking via Pairwise Learning

Pairwise learning utilizes the two methods- example reranking and crowd reranking. Like MIIL reranking it also has two parts Learning and Reranking. First fed the textual query and get the initial ranked list. Then this query is fed to the web search engine to get a set of image search result.

In the learning process, some query examples are provided and by using these we filter the web search results and obtain the clean web examples by visual similarity. Let Q = $\{q_i \mid i = 1,2,....,K\}$ denote the query example set provided by the user, and K is the number of query examples. The final web example set is derived by

$$\varepsilon = \{ e_i \min(1<=j<=k) \| q_j - e_i \| < T_h \}$$

where T_h is the fixed threshold estimated by the average of distances between each query example pair and it is defined as follows:

$$T_h = \frac{\sum_{i,j=1}^{k} ||qj - ei||}{k'}$$

Where k' is the number of query example pairs.

Next step is to find out the concept relatedness to the given query. This can be found out by two methods.

First method is by using a set of pretrained concept detectors , concept detection[9] is performed on the example set and find out the confidence score of each web example and the concept. Second method is by utilizing the text associated with the web examples. Here Google Distance[4][5] is used to measure two textual words.

By combining the two methods the concept relatedness of the given query to the given concept is obtained as given:

$$Y_i = \frac{\lambda}{K} \sum_{k=1}^{K} C\,f(ek, cj) + \frac{(1-\lambda)}{J} \sum_{i=1}^{J} GD(wi, cj)$$

where $C\,f(ek, cj)$ is the confidence score of the concept cj of the web example ek obtained from the pre-trained concept detectors. λ ($0 \leq \lambda \leq 1$) is a parameter to tune the contribution of concept detectors and surrounding text.

In the reranking process, initial ranked list is converted to a pairset in which all documents are paired with all other documents in the initial list but the ranking order is preserved. That means if the pair set is t(xi,xj), xi is ranked higher than xj in the initial ranked list. Reranking is formulized as an optimization problem which minimizes the three energy functions-Ranking Distance, Knowledge Distance and Smooth Distance.

$$E(r) = \alpha Dr(r,r') + \beta Dk(r,y) + \gamma Ds(r)$$

Where $r' = [r1',r2',....,rM']^T$ and $r = [r1,r2,...,rM]^T$ are the initial and reranked pairwise ordinal score. That is if r_i' denotes the initial ordinal score for the pair $t_i(xm,xn)$,then

$$r_i' = \frac{n-m}{M}$$

where M denotes the number of pair set.

Ranking distance specifies that initial ranking order should be preserved. It is calculated as:

$$Dr_1(r, r') = \sum_{i=1}^{M} (ri - ri')^2$$

Knowledge distance specifies that the reranked pairs should be consistent with the learnt knowledge. Smooth distance specifies that if two pairs have similar characteristics, their corresponding ordinal score should be very close. Knowledge and smooth distance can be defined by using a set of pretrained concept detectors. First we represent the pairset as a matrix F = fij where fij denotes the relatedness of the pair ti(xm,xn) to the jth concept cj.

$$f_{ij} = C'f(t_i, c_j) = \frac{1}{1 + e^{-(C\,f(xm,cj) - Cf(xn,cj))}}$$

the knowledge distance is defined as:

$$Dk(r,y) = \sum_{i=1}^{M} \sum_{j=1}^{L} f_{ij} y_j r_i$$

Let $s = [s_1, s_2, ..., s_M]$ denote the vector with entries

$S_i = \sum_{j=1}^{L} f_{ij} y_j$.s can be viewed as the approximate cosine similarity between the concept-based representation of document Pair f_{ij} and the learned concept relatedness of the given query y_j , since f and y belong to the range of [0,1].

Smooth distance is defined as follows:

$$Ds\ (\mathrm{r}) = \sum_{i,j=1}^{M} wij(ri - rj)^2$$

where w_{ij} is the similarity between ti an tj.

The optimization is proposed as a function which minimizes these distances.

$$\min_r\{\alpha \mathrm{Dr1}(\mathrm{r}, \mathrm{r}') + \beta \mathrm{D}k\ (\mathrm{r}, \mathrm{y}) + \gamma \mathrm{Ds}(\mathrm{r})\} = \min_r\{\alpha \sum_{i=1}^{M}(ri - ri')^2 - \beta \sum_{i=1}^{M}\sum_{j=1}^{L} fijyjri + \gamma\sum_{i,j=1}^{M} wij(ri - rj)^2 \}$$

We call this optimization problem as difference pairwise reranking (*DP-reranking*)[3]. We can obtain the solution as follows:

$$\mathrm{r} = \frac{1}{2}(\alpha\mathrm{I} + 2\mathrm{c}\Delta')^{-1}(2\alpha\mathrm{r}' + \beta\mathrm{s}')$$

where I is an identity matrix whose diagonal elements are 1 and the others are 0. $\Delta = D - W$, where $W = [w_{ij}]_{N \times N}$ and D is a diagonal matrix with its (n- n)-element $d_{ii} = \sum_{j=1}^{N} wij.$ r''and s'are obtained by replacing the last element of r' and s with zero, respectively. Δ' is obtained by replacing the last row of Δ with $[0,0,...,1]_{1 \times N}$. Letting $\mathrm{v} = \frac{1}{2}(\alpha\mathrm{I} + 2\mathrm{c}\Delta')^{-1}$, we can see that (14) consists of two parts, i.e.2αvr', and βvs. They correspond to the initial search results and learnt concept relatedness, respectively, and both are smoothed by each other. Therefore, the reranked list can be viewed as the combination of the initial search results and the learnt external knowledge[3].

The final step is to recover the reranked list. It can be obtained by Round Robin method. Round robin reranking first assigns the reranked ordinal score to the first element of each pair and the second element is assigned the value 0. All the scores assigned to the same element is added together. According to this score the documents are reordered in decreasing order of their scores.

III COMPARITIVE ANALYSIS

Here we analyse three different techniques for reranking visual images. From this survey we understand the importance of reranking. Reranking is important since it is an efficient method to search and retrieve visual documents(images and videos). Earlier approaches uses classification performance as the optimization objective. It says whether a document is relevant or not. Bt it cant provide an optimal ranked list.

A. Method used

- Random walk reranking is an example of self reranking. That is it doesn't s use any external knowledge. It extract features from the initial ranked list and based on the similarity between the documents.
- Minimum Incremental Information Loss reranking is an example reranking method. Here Query examples are provided for each textual query and reranking is based on these examples provided by the user.
- Pairwise learning uses both example reranking and crowd reranking. The textual query is fed to the web search engine to obtain a collection of web images. User provide some examples to the given query and this examples are used to filer the web images to get clean web examples.

B. Problems

- Since random walk doesn't use any external knowledge and depends entirely on the initial result obtained by a text based approach, ambiguity problem arises. That is if we are searching for Tiger, Tiger may be an animal or it may be a biscuit. Thus the result obtained is a combination of these two. In this case the system cannot determine what the user is really search for.
- In MIIL learning Examples are provided by the user. So based on these examples the system can determine what the user is actually searching for. Using the example set and the initial result, Example pairs are formed and concept detection is processed on this example pairs to obtain the relevant and irrelevant information. Thus the ambiguity problem is solved here. Here the problem is that the user is no able to provide sufficient amount of examples. For an efficient searching model, large amount of training data is essential.
- Since Pairwise Learning uses both crowd reranking and example reranking, it solves the ambiguity problem and also large collection of images is obtained by crowd reanking, thus by solves the problem of limitation of example set needed. Crowd reranking alone taken arises the problem of ambiguity. That is why a combination of the two mehods are used.

C. Techniques used

- In Random walk reranking, reranking is performed as a random walk over the document level context graph. Multimodal contextual similarity is used to gain control over the random walk or the graph traversal. The stationary state probability of the document is taken as the text relevant score.

- MIIL learning uses optimization technique to find an optimal pair set. Mutual information theory is used here to define the reranking criteria. It proposes the criteria which maximizes the MI between the pairs and relevant information and minimizes MI between the pairs and irrelevant information .
- Pairwise reranking also uses optimization technique. Here the basic idea is to minimize the energy function or the three distance, Ranking distance, Knowledge distance and Smooth distance.

D. Method used to retrieve the reranked list

- The stationary probability of the documents is used as the text relevant scores in random walk. Based on the higher score the document appears higher in the reranked list.
- MIIL Learning use the lossy information compression theory to retrieve the reranked list. In this algorithm the best possible pair is selected at each round and this best possible pair is viewed as the compressed data which preserves the most relevant information while excludes the most irrelevant information. All pairs formed by the elements of these pairs are removed and the elements are mapped to the reranked list.
- Pairwise learning uses Round robin method to retrieve the reranked list. Here the ordinal score is assigned to the first element of each pair and the second element is assigned zero. Then it sum up all the scores assigned to the same element and retrieve the ranked list in descending order of their scores.

Table 1: Comparison of the three reranking methods

	Random Walk	MIIL Reranking	Pair wise Learning
Method	Self Reranking	Example Reranking	Example reranking & crowd reranking
Problem	Ambiguity	Limited example provided by user	-
Approach	-	Pairwise	pairwise
Technique	Random walk based on multimodal contextual similarity	Optimization	Optimization
Optimization Technique	-	Mutual Information	Minimization of ranking distance, knowledge distance, smooth distance
Retrieval of reranked list	Descending order of their stationary probability	Lossy Information Compression Theory	Round Robin method

IV. CONCLUSION

In this paper we discuss about three reranking methods- Random walk over document level context graph, Reranking via MIIL and Reranking via Pair wise Learning.Of this random walk, is a self reranking method which is not so efficient since it has the ambiguity problem. The second method MIIL is better than Random walk since it uses the example provided by the user. Till it has the problem of the limitation of examples provided by the user. For an efficient reranking method large collection of training samples is needed. Thus crowd reranking comes forward which uses a collection of web images. But here the problem is ambiguity because the web search is based on textual query.

So to overcome these problem a new reranking is proposed which is the Pair wise learning. It utilizes both example reranking and crowd reranking. Thus the ambiguity and limitation of examples are solved. From the analysis we can say that Pair wise learning is the best reranking method.

V. REFERENCES

[1] W. Hsu, L.Kennedy, and S.-F. Chang, "Video search reranking through random walk over document-level context graph," in Proc. ACM Int. Conf. Multimedia, 2007, pp. 971–980.

[2] Y. Liu, T. Mei, X. Wu, and X.-S. Hua, "Optimizing video search reranking via minimum incremental information loss," in Proc. ACM Int. Workshop Multimedia Information Retrieval, 2008, pp. 253–259.

[3] Yuan Liu and Tao Mei, Member, IEEE "Optimizing Visual Search Reranking via Pairwise Learning" IEEE transaction on multimedia, vol. 13 , no .2 April.2011.

[4] R. Cilibrasi and P. Vitanyi, "The Google similarity distance," IEEE Trans. Knowl. Data Eng., vol. 19, no. 3, pp. 370–383, Mar. 2007.

[5] R. Fergus, L. Fei-Fei, P. Perona, and A. Zisserman, "Learning object categories from Google's image search," in Proc. IEEE Int. Conf. Computer Vision, 2005, pp. 1816–1823.

[6] L. Kennedy and S.-F. Chang, "A reranking approach for context-based concept fusion in video indexing and retrieval," in Proc. ACMInt. Conf. Image and Video Retrieval, 2007.

[7] Y. Liu, T. Mei, and X.-S. Hua, "Crowdreranking: Exploring multiple search engines for visual search reranking," in Proc. ACM Special Interest Group on Information Retrieval, 2009, pp. 500–507.

[8] D. Cossock and T. Zhang, "Statistical analysis of Bayes optimal subset ranking," IEEE Trans. Inf. Theory, vol. 54, no. 11, pp. 5140–5154, 2008.

[9] T. Mei, X.-S. Hua, W. Lai, L. Yang, Z.-J. Zha, Y. Liu et al. MSRA-USTC-SJTU at TRECVID 2007: high-level feature extraction and search. In *NIST TRECVID Workshop*, 2007.

[10] T.M. Cover, and T.A. Thomas. The elements of information theory. Plenum Press, New York, 1991.

[11] S.G. Chang, B. Yu, and M. Vetterli. Adaptive Wavelet Thresholding for Image Denoising and Compression. *IEEE Transactions on Image Processing* ,9(9): 2000.