A Survey on Bayesian Visual Reranking

Nightingale.D[1], **Akila Agnes**[2]

[1]Karunya University, School of Computer Science and Technology,
Karunya Nagar, Coimbatore, India
tech.nightingale@gmail.com

[2]Karunya University, School of Computer Science and Technology,
Karunya Nagar, Coimbatore, India
akilaagnes@gmail.com

Abstract: ***Visual re ranking is a method introduced mainly to refine text-based image search results. It utilizes visual information of an image to find the "true" ranking list from the noisy one done by the search based on texts. The process uses both textual and visual information. In this paper, textual and visual information is modeled from the probabilistic perspective visual reranking is in the Bayesian framework, thereby named as Bayesian visual reranking. In this method, the text based information is taken as likelihood, to find the preference strength between re ranked results and text-based search results which is the ranking distance. The visual information of an image is taken as the conditional prior, to indicate the ranking score consistency between the visually similar samples. This process maximizes visual consistency and minimizes the ranking distance. For finding the ranking distance, three ranking distance methods are use . Three different regularizers are studied to find the best results. Extensive experiments are done on text based image search datasets and Bayesian visual reranking proved to be effective.***

Keywords: Visual re ranking, visual consistency, regularizer, ranking distance, Bayesian framework.

1. Introduction

The searches that are implemented in recent days are mostly done by 'query by keyword'. They are made by using various text information of the images like the texts near the images, captions, titles or even speech transcripts. But sometimes these texts do not match with the image as they might mean something else other than the image, thereby making the image searches inefficient. So visual information should also be considered to refine the search results. But using only visual information also has a lot of disadvantages like improper visual features. This leads to the process of visual reranking.

Visual reranking is a mixed process of both the text based image results and the visual features to obtain good performance in image searches. The process can be explained by Fig. 1.1 in which the text query is "cloud", here first a text based search is done and it returns few mismatched results like images 2, 6 which are dissimilar. Then a visual consistency pattern is used in reranking to refine the initial ranking list and thereby we get the images 2, 6 reordered in the last and the related images reordered in the first. This process of reordering the images based on both text based and visual cues is called image search reranking.

In [1], the visual reranking uses the Bayesian framework modelling it in the probabilistic pattern. Therefore for the Bayesian framework we need likelihood and a conditional prior. Here the likelihood is taken as the text features and the conditional prior is taken as the visual features using the visual consistency criteria for the purpose of reranking. Ranking distance is an important factor in visual search reranking, which affects the overall reranking performance significantly but has not been well studied before. Thereby here the pair wise ranking distance is used which experimentally performed in [1] is very successful.

2. Existing Techniques

Several techniques have been put forward for better performance of visual reranking in the recent days. The methods include classification based, clustering based, random walk based, auxiliary method based etc, some of the existing methods for visual reranking mechanism used in visual search are explained below.

2.1 Classification based:

In this method, it simplifies reranking as a classification problem. There are normally three steps

- Select training samples from initial text-based search results
- Train a classifier with selected samples.
- Reorder all samples according to predictions given by the trained classifier.

In the first step, pseudo relevance feedback (PRF) is often utilized. Pseudo Relevance feedback is a method that began from text retrieval. It takes few of the top-ranked documents from the search results done initially as pseudo positive. Instead it uses the images from the query or example video clips as positive samples. The pseudo-negative samples that are found are taken from either the lowest ranked images in the initial result or the database assuming that few images in the database are a match.

In step two, different classifiers such as SVM, boosting and ranking SVM can be adopted. Although the classifiers are of a lot of effect, a lot of training data is demanded for satisfactory performance because of the many parameters to be estimated. Information Bottleneck principle, is applied to find suitable clustering which maximizes the similar list is found

First the text-based search engine returns the images related to the query “Cloud” from textual cues and then the reranking process is applied to refine this result by extracting visual information showing the top few ranked images in the text based search results and the re ranked results respectively.

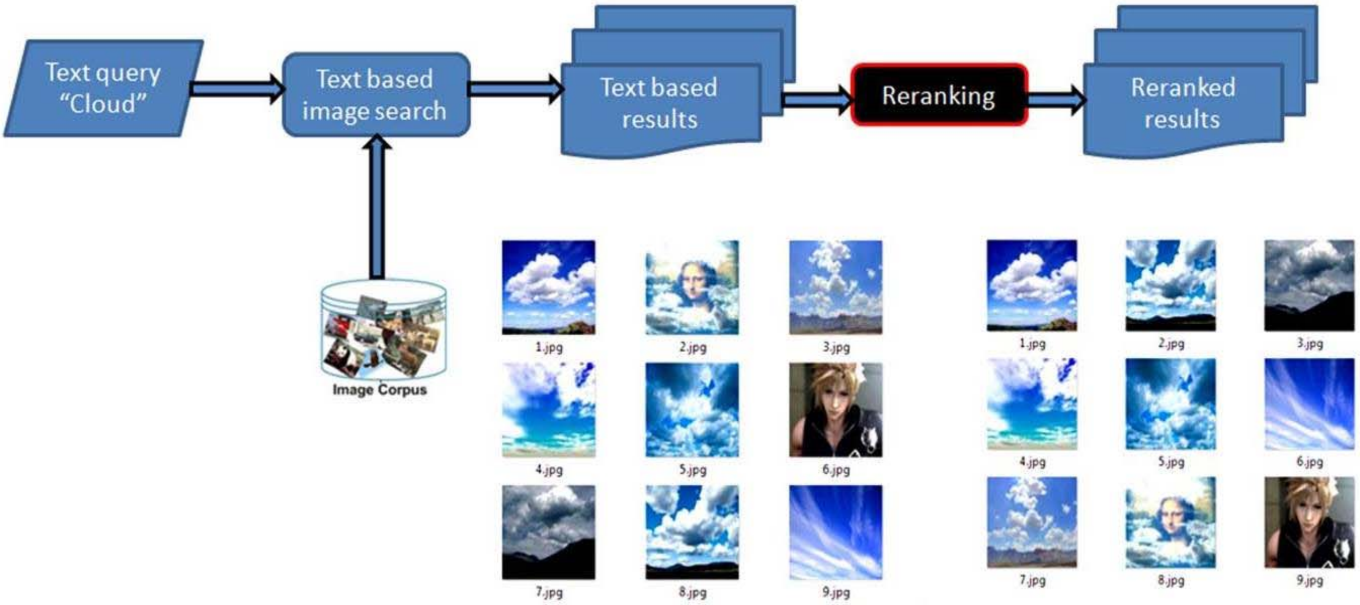

Figure 1: Illustration of visual re ranking

2.2 Clustering based

The second method is clustering based. In the paper (Hsu.W. H., 2007), each image is assigned a soft pseudo label based on the initial text search result, and then the information between its clusters and the labels. Re ranked by arranging the clusters based on the cluster conditional property firstly and then sorting the samples within a cluster based on the closely related feature density. This method is very effective on named-person based queries while it is limited to those queries which have very specific similar characteristics.

2.3 Random walk based

The third category is random walk based. A graph is drawn with images as the nodes and the edges between them are calculated by visual similarity. Then, reranking is derived as random walk over the graph and the ranking is defined from the edges. To find the text search result, a dongle node is appended to each image with fixed value to the initial text ranking score. In this paper also random walk method is unified with the Bayesian visual reranking framework.

2.4 Auxiliary knowledge based

There are also other methods which incorporate auxiliary knowledge as depicted in fig 2.1, including the following;

- Face detection: For high-level feature extraction, the benefit of unlabeled data by semi-supervised learning methods, including adaptive semi-supervised learning with kernel density estimation, manifold ranking, and transductive graph. Moreover, we fusion is done in two different levels: modality level and model level. For rushes exploitation, the duplicate content based on ordinal video signature is detected. Then classification (i.e. classifying each sub-shot into static, pan, tilt, zoom, rotation, or object motion in terms of camera motion) is performed.

- Query Example: A query, consisting of a text description plus images or video is posed against a video collection, and relevant shots are to be retrieved. This system accomplishes this by using the retrieval results of multiple retrieval agents. The overall system can be decomposed into several agents, including a text -oriented retrieval agent, which is responsible for finding the text in the speech transcripts, a video-information oriented agent which is responsible for searching the ‘manually’ provided movie abstracts and titles) and a basic nearest neighbor image matching agent which can be combined with classification-based pseudo-relevance feedback (PRF). The motivation of the classifier based PRF approach is to improve the image retrieval performance by feeding back relevance estimates based on the initial search results into a classifier and then refining the retrieval result using the classification output. To address the issue of comparability between retrieval scores produced by different types of agents, the retrieval scores of these agents are converted into posterior probabilities in an attempt to create normalized output scores.

- Concept Detection: Here automatic multimodal fusion for video search is done by by employing not only textual and visual features, but also semantic and conceptual similarity between video shots to re-rank the search results. It develops an approach to video search which not only can avoid the dependency on specific query characteristics, training data and human interference, but also can leverage textual relevancy, semantic concept relevancy, and visual similarity in a novel fashion. It would smooth the multimodal information sources in an implicit yet “soft” graph-based propagation way instead of an explicit and “hard” linear aggregation. it requires no involvement of human effort as the relevance of video shots to a given topic is propagated through the multiple graphs automatically. Furthermore, the fusion across textual, visual and semantic conceptual information is implemented in a graph-based iterative style, which combines the information from multimodalities in a natural and sound way. Though the incorporation of auxiliary knowledge leads to the performance improvement; it is not a general treatment. They suffer from either limited applicability to the specific queries (face detection), user interfaces (query example), or the limited detection performance and small vocabulary size (concept detection). And thereby they cannot be used for a process that demands efficiency. They are good with name-person based queries but limited when it comes to applicability.

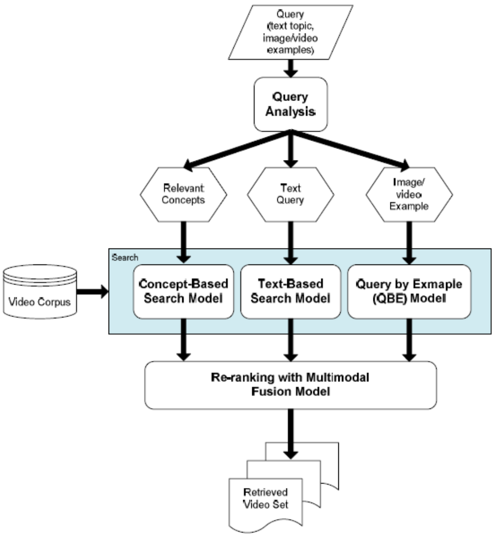
Query (text topic, image/video examples)
Query Analysis
Relevant Concepts
Text Query
Image/ video Example
Search
Video Corpus
Concept-Based Search Model
Text-Based Search Model
Query by Exmaple (QBE) Model
Re-ranking with Multimodal Fusion Model
Retrieved Video Set

Figure 2: Auxiliary knowledge based methods

2.5 Ranking Methods

The likelihood of the Bayesian framework [1] is modeled via ranking distance. This process estimates the disagreement between the ranking lists before and after reranking. It is a crucial factor which significantly affects reranking performance.

Point wise approach: Main idea is to sort the problem is transformed into a multi-class classification or regression problem. Multi-class classification example: Suppose the query the query and its related documentation set is: {d1, d2... dn}. So first the characteristics of n Solo pair are extracted: (query, di) and expressed as a feature vector. Correlation between the query and di as a label, the label classification: {of Perfect, Excellent, Good, Fair, Bad,}, a total of five categories.

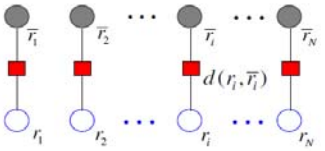

Figure 3: The factor graph representation of point-wise ranking distance in which the ranking distance is computed by summing each sample’s distance**.**

Thus, for a query and its set of documents, we can form the n training instances. With the training examples, we can use any of the multi-class classifier learning, such as maximum entropy, the SVM. Point wise is relatively simple, with no formal start. Such an assumption implicit in the Point wise methods: absolute correlation assumptions, it assumes that the correlation is query-independent, query-independent. In other words, as long as the (query, document), such as the "perfect", they are placed in the same category, that is, belonging to the same instance of the class, regardless of what query is. Practice, however, the correlation is not a query-independent. Very common query and its related documents, their may be higher the tf among a very rare queries and one of its related documentation. This will result in the training data is inconsistent; it is difficult to achieve good results. Forecast document for the same category can not make a sort.

Pair wise approach: In the pair wise approach, the learning task is formalized as classification of object pairs into two categories (correctly ranked and incorrectly ranked). The approach is employed by using the SVM techniques to build the classification model. The method is referred to as Ranking SVM. They employed Cross Entropy as loss function and Gradient Descent as algorithm to train a Neural Network model. Learning to rank, particularly the pair wise approach, has been successively applied to information retrieval. Pair wise approach applied Ranking SVM to document retrieval. A method of deriving document pairs for training, from users’ clicks-through data was developed.

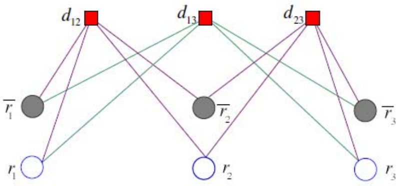

Figure 4: The factor graph representation of pair wise ranking distance in which the ranking distance is computed by summing each pair’s distance

List wise approach: In the list wise approach, instead of using object pairs as instances, list of objects. These algorithms try to directly optimize the value of one of the above evaluation measures, averaged over all queries in the training data. This is difficult because most evaluation measures are not continuous functions with respect to ranking model's parameters, and so continuous approximations or bounds on evaluation measures have to be used. Instances in learning are used. The key issue for the list wise approach is to define list wise loss function.

2.6 Regularizers

Regularizers are concepts from the machine learning field. In visual reranking are mainly for finding the visual consistency between images or videos. The various regularizers used regularly are

Laplacian regularizers: Here first a graph G is constructed with nodes being the samples and similar samples are linked by edges. If two samples x^i and x^j are linked, the weight w^{ij} on the edge between them is calculated by using Gaussian radial basis function kernel

$w^{ij} = \exp \{-\| x^i - x^j \|^2 / 2\sigma^2$, where σ is the scaling parameter. Else if the two samples are not connected, $w^{ij} = 0$. Here the regularizer is defined by

$$\psi^i(r, X) = \frac{1}{2} \Sigma\, w^{ij} \, (r^i - r^j)^2 \qquad (1)$$

It approximates the visual consistency of x^i from the pair wise perspective by accumulating the weighed score difference between x^i and each of its neighbors x^j .

Normalized Laplacian Regularizer: Here also first a graph G is constructed with nodes being the samples and similar samples are linked by edges. If two samples x^i and x^j are linked, the weight w^{ij} on the edge between them is calculated by using Gaussian radial basis function kernel which is given by $w^{ij} = \exp \{-\| x^i - x^j \|^2 / 2\sigma^2$, where σ is the scaling parameter. Else if the two samples are not connected, $w^{ij} = 0$. Here the regularizer is defined by

$$\psi^i(r, X) = \frac{1}{2} \sum_j w^{ij} \left(\frac{r_i}{\sqrt{d_i}} - \frac{r_j}{\sqrt{d_j}} \right)^2 \qquad (2)$$

From the above regularizers it is clear that both Laplacian and normalized Laplacian regularizers approximate the ranking score consistency for each sample pair-wisely and have less ability to capture the multiple wise ranking score consistency.

Local Learning Regularizer: Local learning regularizer [1] models the multiple-wise consistency by formulating the score estimation as a learning problem without heuristic assumptions. The consistency over a local area means that each sample has strong correlation with its neighbors. In other words, each sample's labeling information is partially embedded in its neighbors. Therefore, if we can deduce a sample's label from its neighbors precisely, this sample is regarded as locally consistent. The local learning regularizer is developed in such manner. For a sample, instead of calculating the consistency with each of its neighbors individually, the local learning regularizer considers the consistency with all of its neighboring samples simultaneously. In this regularizer, a local model is first trained for each sample with its neighbors and then used to predict its consistent ranking score. Finally, by minimizing the difference between the target ranking score and this locally predicted one, the desired multiple-wise consistency is guaranteed.

With the visual consistency assumption, the desired property of r, is that: for each sample x^i and its neighbors, their ranking scores on G should be smooth enough. Smoothness is a term defined over the whole neighbor set, instead of over each of the samples separately. 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, the multiple-wise consistency is guaranteed. From this point of view, we model the ranking score consistency from the machine.

3. Comparison of different approaches

Several works have addressed the visual reranking process in visual search process. The various regularizers and the ranking methods are used for the comparison process. Other methods like the auxiliary methods, classification based method and information bottleneck principle are also tried out for better results. The below table gives an overview of all methods used in the visual search process with their merits and demerits.

Table 1: Comparison of Existing Techniques

Feature	*Merits*	*Demerits*
Information Bottleneck Principle	Good performance on name-person based queries	Limited to queries which have significant duplicate characteristics
Classification based approach	Very effective in data retrieval.	Complexity in designing, since sufficient training data are demanded since a lot of parameters are needed.
Face detection(auxiliary based)	Great improvement in performance.	Limited applicability to specific queries
Laplacian regularizer, Normalized laplacian regularizer	Visual consistency is pair wise making it easier to define.	Since there is no multiple wise consistency, the consistency on a local are is not defined accurately.
Point wise ranking distance.	It is the most simplest and direct way to measure ranking distance between two score lists.	Fails to capture disagreement between score lists

4. Conclusion

The Bayesian framework is very effective since it increases the visual consistency and reduces the ranking distance. Therefore many methods are done for both conditional prior who represents the visual consistency and likelihood which represents the ranking distance. The method used above likes the various regularizers and the ranking methods are being studied extensively in the recent days. But by defining and modeling new regularizers and using the proper ranking methods efficient results can be achieved.

References

[1] X. Tian, L. Yang, J. Wang, Y. Yang, X. Wu, and X.-S. Hua, “Bayesian video search reranking,” in Proc. ACM Int. Conf. Multimedia, 2008,pp. 131–140.

[2] H. S. Chang, S. Sull, and S. U. Lee, “Efficient video indexing scheme for content-base retrieval,” IEEE Trans. Circuits Syst. Video Technol., vol. 9, no. 8, pp. 1269–1279, Dec. 1999.

[3] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, “Content-based image retrieval at the end of the early years,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, no. 12, pp. 1349–1380, Dec. 2000.

[4] W. H. Hsu, L. S. Kennedy, and S.-F. Chang, “Video search reranking via information bottleneck principle,” in Proc. ACM Int. Conf. Multimedia, 2006, pp. 35–44.

[5] W. H. Hsu, L. S. 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

[6] J.Liu, W. lai, X.-S Hua, Y. Huang and S. LiVideo Search Re-Ranking via Multi-Graph Propagation in Proc. ACM Int. Conf. Multimedia, 2007, pp. 208–217

[7] Z. Cao, T. Qin, T.-Y. Liu, M.-F. Tsai, and H. Li. Learning to rank: from pair wise approach to list wise approach. in ICML, pages 129–136, 2007.

[8] X.Tia, l.yang, J.Wang, y.Yang, X.Wu and X.-S.Hua.”Bayesian Video Search reranking,” in Proc. ACM Int. Conf. Multimedia, 2007, pp 131-140

[9] Y. Liu, T. Mei, J. Tang, X. Wu, and X.-S. Hua. Learning to video search rerank via pseudo preference feedback. In ICME, 2008.

[10] R. Yan, A. G. Hauptmann, and R. Jin. Multimedia search with pseudo-relevance feedback. in CIVR, pages 238–247, 2003.

[11] R. Yan and A. G. Hauptmann. Co-retrieval: A boosted reranking approach for video retrieval. in CIVR, pages 60–69, 2004.

[12] T. Mei, X.-S. Hua, W. Lai, L. Yang, Z.-J. Zha, Y. Liu, Z. Gu, G.-J. Qi, M. Wang, J.Tang, X. Yuan, Z. Lu, and J. Liu. Msra-ustc-sjtu at trecvid 2007: High-level feature extraction and search. In TREC Video Retrieval Evaluation Online Proceedings, 2007.

[13] A. Natsev, A. Haubold, J. Tesic, L. Xie, and R. Yan. Semantic concept-based query expansion and re-ranking for multimedia retrieval. In ACM Multimedia, pages 991–1000, 2007.

[14] X. Zhu, Z. Ghahramani, and J. D. Lafferty. Semi-supervised learning using Gaussian fields and harmonic functions. In ICML, pages 912–919, 2003.

[15] R. Herbrich, T. Graepel, and K. Obermayer. Large margin rank boundaries for ordinal regression. In Advances in Large Margin Classifiers, pages 115–132. MIT Press, Cambridge, MA, 2000.

[16] S. E. Robertson, S. Walker, M. Hancock-Beaulieu, M. Gatford, and A. Payne. Simple, proven approaches to text retrieval. In Cambridge University Computer Laboratory Technical Report TR356, 1997.

Author Profile

Nightingale D received her B.Tech degree from Kings Engineering College, India in 2011. She is currently pursuing her Masters from Karunya University, India.