Optimizing Video Search Reranking Via Minimum Incremental Information Loss

[*] Yuan Liu [†], Tao Mei [‡], Xiuqing Wu [†], Xian-Sheng Hua [‡]
[†] Department of Elec. Eng. & Info. Sci., University of Sci. & Tech. of China, Hefei, 230027, P.R.China
[‡] Microsoft Research Asia, Beijing, 100190, P.R.China
yuanliu.ustc@gmail.com, wuxq@ustc.edu.cn, {tmei,xshua}@microsoft.com

ABSTRACT

This paper is concerned with video search reranking — the task of reordering the initial ranked documents (video shots) to improve the search performance — in an optimization framework. Conventional supervised reranking approaches empirically convert the reranking as a classification problem in which each document is determined relevant or not, followed by reordering the documents according to the confidence scores of classification. We argue that reranking is essentially an optimization problem in which the ranked list is globally optimal if any two arbitrary documents from the list are correctly ranked in terms of relevance, rather than simply classifying a document into relevant or not. Therefore, we propose in this paper to directly optimize video search reranking from a novel viewpoint of information theory, that is, to identify an optimal set of correctly-ranked document pairs which maximally preserves the relevant information and simultaneously carries the irrelevant information as little as possible. The final reranked list is then directly recovered from this optimal set of pairs. Under the framework, we further propose an effective algorithm, called *minimum incremental information loss* (MIIL) reranking, to solve the optimization problem more practically. We conducted comprehensive experiments on automatic video search task over TRECVID 2005-2007 benchmarks, and showed significant and consistent improvements over the text search baseline and other reranking approaches.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Retrieval models

General Terms

Algorithms, Performance, Experimentation.

[*]This work was performed when Yuan Liu was visiting Microsoft Research Asia as a research intern.

MIR'08, Vancouver, BC, Canada.

Keywords

Video search, reranking, mutual information, optimization.

1. INTRODUCTION

With the daunting volumes of online video data, video search has become one of the key techniques for visual content management. Due to the great success of text document retrieval, most of existing video search systems rely on textual information which is usually obtained from automatic speech recognition (ASR). However, the video relevance cannot be merely judged by text-based approaches as the text transcripts are usually noisy and even unavailable [1] [2].

To address the heavy reliance on text-based search techniques, video search reranking has received increasing attention, which is defined as reordering the ranked documents (video shots) based on the initial search results or some auxiliary knowledge, targeting a higher probability that any two arbitrary documents from the list are ranked correctly in terms of relevance.

Conventional reranking approaches to video search have proceeded along two dimensions: (1) mining the meaningful information from the initial ranked list to perform *unsupervised* reranking [2] [3] [4] [5], and (2) leveraging the auxiliary knowledge to reorder the samples in a *supervised* way [6] [7]. Typical *unsupervised* approaches focus on mining the relevant or irrelevant information in the initial search results which is usually obtained from the text-based search systems. However, without any clues in terms of relevance, it is difficult for unsupervised approaches to understand the query meaning exactly, especially for some ambiguous textual queries. For example, it is difficult to determine the user's meaning of the query keyword "train", i.e., "exercise" or "a kind of vehicle."

On the other hand, *supervised* reranking approaches aim to better understand the query by leveraging some auxiliary knowledge. Most current supervised approaches empirically convert the reranking as a binary classification problem to determine whether a sample is relevant or not, then reorder the samples according to the confidence scores of classification. However, although some systems have obtained good search performance, it is known that an optimal classification performance cannot guarantee optimal search precision [8], usually evaluated by average precision (AP) [9]. Using an example shown in Figure 1, reranked list A and B is obtained by the two reranking approaches which aim to optimize the classification accuracy and ranking precision, respectively. The reranked list A has higher "best accuracy" [8] gain in terms of classification but yields a suboptimal ranking precision gain. Considering the pairs constructed by a relevant and an irrelevant sample, there are only 12 pairs correctly ranked in A, while 15 pairs are correctly ranked in B. Therefore, we can conclude that B is better than A from the perspective of search reranking, and search reranking viewed as the classification problem cannot obtain the globally optimal reranking.

To summarize, conventional reranking approaches fail to give a global optimization to guarantee the highest probability that arbitrary pair of samples is correctly ranked in terms of relevance. To address this problem, we propose in this work a *supervised reranking framework* in which reranking is formulated as an optimization problem aiming to maximize the number of pairs correctly ranked.

We solve the optimization problem from a novel viewpoint of information theory. First, we learn the relevant and irrelevant information from the query examples. Second, a set of *sample pairs* is identified to maximally preserve the relevant information and simultaneously carry the irrelevant information as little as possible. Then the final reranked list is directly recovered from this optimal set of pairs. In order to solve the optimization problem more practically, we propose an effective algorithmic solution, called *minimum incremental information loss* (MIIL) reranking, which iteratively selects one pair at each round from the initial ranked list to keep minimum information loss.

The rest of the paper is organized as follows. We review related work on video search reranking in Section 2. Section 3 gives detailed descriptions about the proposed reranking framework, including the problem formulation and algorithmic solution. Experimental results are reported in Section 4, followed by concluding remarks in Section 5.

2. RELATED WORK

In this section, we will review the representative video search reranking approaches. As aforementioned, from the perspective of learning methods used for reranking, conventional approaches can be categorized into *unsupervised* and *supervised*.

The most straightforward *unsupervised* approaches are to select pseudo-labeling samples based on the pseudo relevance feedback (PRF) framework [2] [5], and then build the reranking models to reorder the initial search results. The key to this kind of reranking methods is to select the pseudo-labeling samples from the initial search results to mine the relevant or irrelevant information. For example, Kennedy *et al.* [5] took the pseudo-positives and pseudo-negatives to discover the related concepts. Then, the scores of related concepts are used as features in Support Vector Machine (SVM) to build classifiers. A recent work [2] proposed a pseudo preference feedback (PPF) based reranking approach taking different degrees of relevance into account. Specifically, it discovered an optimal set of pseudo preference pairs at first, which are then used as training samples to learn the reranking models. Unlike the proposed reranking approach, PPF-based approach tried to obtain several optimal pairs to instruct the succeeding reranking, rather than optimize the ranked list directly.

An alternative kind of *unsupervised* method is developed with the assumption that video documents with similar visual features tend to have the similar relevance [3] [4]. Therefore, more efforts are devoted to discovering recurrent patterns, i.e., patterns with high occurrence frequency between

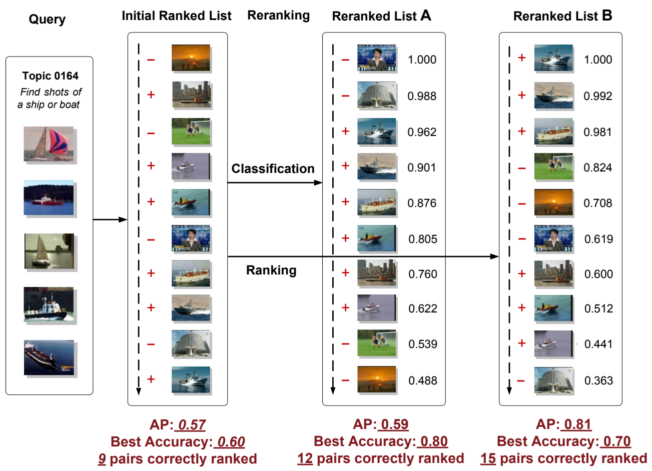

Figure 1: Suboptimal reranking vs. globally optimal reranking. Reranked list A and B is obtained by the two reranking approaches which optimize the classification accuracy and ranking precision, respectively. The sign "+" and "-" located on the left of the ranked lists denotes the ground truth of shots, and the decimal fractions on the right are relevance scores obtained by the two approaches. According to the evaluation below the ranked list, we can find that reranking which optimizes for classification accuracy is not directly concerned with the ranking performance (AP and correctly-ranked pairs).

shots in the search results. For example, Hsu *et al.* [3] directly discovered the recurrent patterns based on IB-based clustering and took the clusters with high cluster conditional probability as the relevant recurrent patterns, which would be ranked higher; otherwise, the clusters with low cluster conditional probability were taken as noises although they are also frequently recurrent. Recently, Hsu *et al.* [4] formulate the reranking solution as a random walk over the context graph, where stories are nodes and the edges between them are weighted by multimodal contextual similarities. Thus samples with recurrent patterns are linked compactly and thus the samples ranked lower can be picked up in such a way.

On the other hand, current *supervised* reranking approaches [6] [7] usually treat the query examples as "positive" and sample the low-ranked samples in the initial search results as "pseudo-negative." Then, a set of visual features are extracted to build a new search model and produce the visual-based search results. Such visual-based search results can be fused with the initial text-based search results. These supervised reranking approaches reduce the ranking/reranking into a binary classification problem which treats the relevant samples as "positive" and irrelevant samples as "negative." As these approaches make little use of the initial ranked list, they are more like fusion of text-based and visual-based search and thus achieve limited success.

In summary, although both *unsupervised* and *supervised* conventional reranking approaches are effective, they only indirectly or partially optimize the ranked list and thus fail to guarantee the global optimization of reranking. For example, IB-based reranking [3] neglects the rank relation between the samples in different clusters, as it ranks the clusters at first and then ranks the samples within each cluster. PRF-based reranking [6] cast the reranking as a classifica-

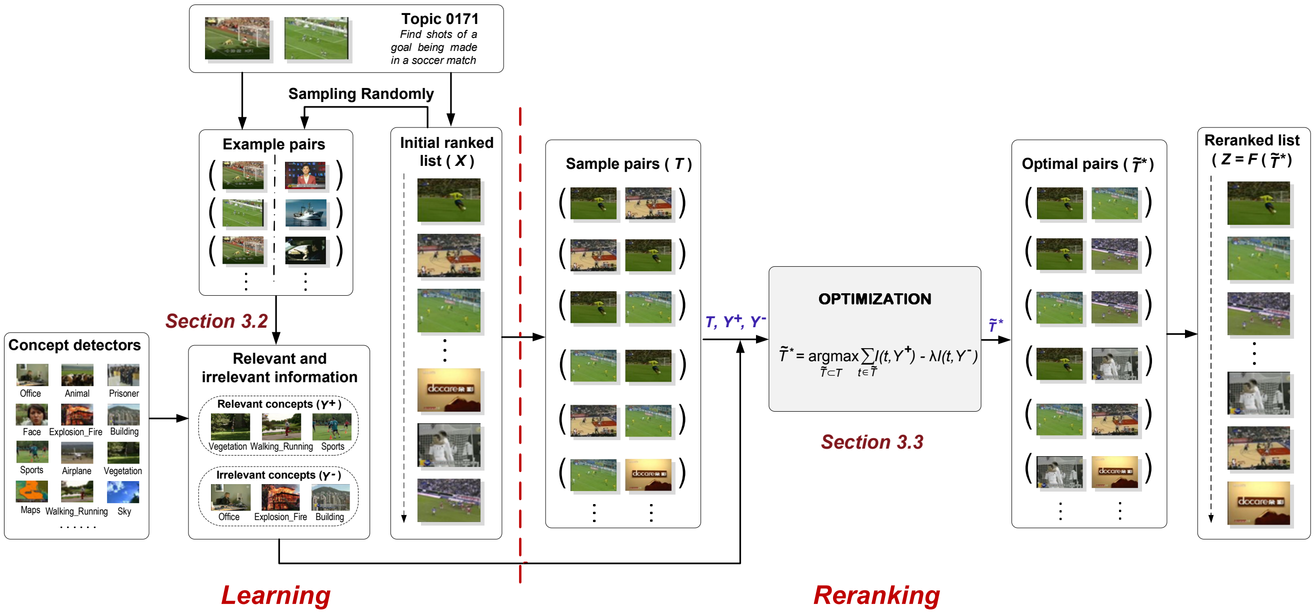

Figure 2: The framework of the proposed reranking approach.

tion problem, which leads to suboptimal reranking performance due to the gap between classification accuracy and ranking precision as analysis above. To address these issues, we propose to directly optimize the ranked list by discovering an optimal set of sample pairs ranked correctly, which will be described in the next section.

3. OPTIMIZATION OF VIDEO SEARCH RERANKING

We will give the formulation and implementation of the proposed optimization-based video search reranking in the following sections. Specifically, we introduce the framework and the problem formulation in Section 3.1. In Section 3.2, we present how we discover the relevant and irrelevant information from the query examples. Based on this information, we will propose an algorithmic solution, called minimum incremental information loss (MIIL) reranking, in Section 3.3.

3.1 Problem Formulation

The video search reranking framework includes two components, i.e., *learning* and *reranking*. The framework of the proposed reranking is illustrated in Figure 2.

In the *learning* process, several query examples, usually images or video shots, for each textual query are provided. The objective of learning is to mine the relevant (Y^+) and irrelevant information (Y^-), which is expressed by a set of predefined visual concepts [10] [11]. First, a set of example pairs are constructed by the query examples and shots randomly sampled in the initial search results. Then the concept detection is processed in these example pairs. The details of the feature extraction and concept detection can be found in [11]. With the concept detection scores, we can determine which concepts are relevant or irrelevant to a given query. We will give more details in Section 3.2.

In the *reranking* process, based on the relevant and irrelevant concepts, an optimal reranked list is obtained by an optimization-based method. We first convert the samples in initial ranked list $X = \{x_i|i = 0, \ldots, N-1\}$ to a pair set $T = \{t_{ij}|i, j = 0, \ldots, N-1; i \neq j\}$, without considering the sample order. Thus the size of pair set is $|T| = N \times (N-1)$. While considering the sample order, each ranked list corresponds to a pair set $\tilde{T} = \{t_{ij}|i, j = 0, \ldots, N-1; x_i \succ x_j\}$, where "$\succ$" denotes the rank relationship between a pair of samples and "$x_i \succ x_j$" indicates x_i is ranked higher than x_j (i.e., x_i is more relevant to the query than x_j). Clearly, $\tilde{T} \subset T$ and the size of $\tilde{T}$, $|\tilde{T}| = |T|/2$. For N initial search results, there are $N!$ possible reranked lists. Thus the major objective can be formulated to find the optimal pair set $\tilde{T}^*$ among the $N!$ possible pair sets.

Mutual information (MI) between two variables is a basic concept which roots in information theory and essentially measures the amount of information that one variable contains the other [12]. It is natural to introduce MI to evaluate how well the discovered relevant/irrelevant concepts indicate the rank relationship existing in the pairs. Let t be an element of the pair set $\tilde{T}$. MI between t and Y(Y^+ or Y^-), $I(t, Y)$, is defined by Kullback-Leibler distance based on the joint distribution $p(t, Y)$ and the distribution associated to the complete independence $p(t) \times p(Y)$ [12], i.e.,

$$\begin{aligned} I(t, Y) &= p(t, Y) \log \tfrac{p(t,Y)}{p(t) \times p(Y)} \\ &= \textstyle\sum_{y \in Y} p(t, y) \log \tfrac{p(t,y)}{p(t) \times p(y)} \\ &= p(t) \textstyle\sum_{y \in Y} p(y|t) \log \tfrac{p(y|t)}{p(y)} \end{aligned} \quad (1)$$

Thus, the mutual information between the pair set $\tilde{T}$ and concept set Y, is given by $I(\tilde{T}, Y) = \sum_{t \in \tilde{T}} I(t, Y)$.

Accordingly,we propose to define the reranking criterion which maximizes the MI between the pairs and relevant information while simultaneously minimizes the MI between the pairs and irrelevant information. This dual optimization task can be approached by maximizing the weighted difference $L(\tilde{T})$:

$$\begin{aligned} L(\tilde{T}) &= \textstyle\sum_{t \in \tilde{T}} L(t) \\ &= \textstyle\sum_{t \in \tilde{T}} I(t, Y^+) - \lambda I(t, Y^-) \end{aligned} \quad (2)$$

where $L(t) = I(t, Y^+) - \lambda I(t, Y^-)$. λ determines the tradeoff between preservation of the relevant information Y^+ and loss of the irrelevant information Y^-. Thus the reranking criterion is given by:

$$\tilde{T}^* = \arg\max_{\tilde{T} \subset T} L(\tilde{T}) \tag{3}$$

When the optimal pair set $\tilde{T}^*$ is obtained, a round robin criterion is introduced to obtain the final reranked list $Z = F(T^*)$. Specifically, if t_{ij} (i.e., $x_i \succ x_j$) is an element of the optimal pair set $\tilde{T}^*$, we assign a vote to x_i. Conversely, the prediction $x_j \succ x_i$ would be considered as a vote for x_j. Then add all the votes assigned to each sample, and the samples are finally ranked in descending order of the sum of the votes they are assigned.

To solve the above optimization problem defined in Equation (3), the posterior probability of pair samples $p(y|t_{ij})$ and the prior distribution of pair samples $p(t_{ij})$ is required. We introduce a mapping function $f : X, Y \rightarrow \Re$, which provides a measurement to evaluate the relevance of individual samples based on the discovered relevant and irrelevant concepts. Specifically, $f(x_i, y) \succ f(x_j, y)$ indicates that the model asserts that $x_i \succ x_j$ about the concept y. We define the posterior probability of pair samples $p(y|t_{ij})$ using a logistic function:

$$p(y|t_{ij}) = \frac{1}{1 + e^{-m \times [f(x_i, y) - f(x_j, y)]}} \tag{4}$$

We can see that $p(y|t_{ij})$ is approaching to 1 when $f(x_i, y)$ is much larger than $f(x_j, y)$, while it is approaching to 0 when $f(x_i, y)$ is much lower than $f(x_j, y)$. When $f(x_i, y)$ is equal to $f(x_j, y)$, $p(y|t_{ij}) = 0.5$, it denotes that x_i and x_j is vague to determine which one is more relevant by learning from the query examples. m determines the confidence of information learned from query examples.

We also consider the initial model $g : X \rightarrow \Re$ which indicates the information provided by the initial ranked list. Hence, the prior distribution $p(t_{ij})$ is estimated as:

$$p(t_{ij}) = \frac{1}{1 + e^{-n \times [g(x_i) - g(x_j)]}} \tag{5}$$

where n determines the confidence of initial search results.

The above two probabilities of sample pair correspond to the information obtained from the query examples and initial search results, which is expressed by $f(x, y)$ and $g(x)$, respectively. In this paper, they are defined as:

$$f(x, y) = \textit{detection score of sample } x \textit{ for concept } y \tag{6}$$

$$g(x) = \textit{order of sample } x \textit{ in initial ranked list} \tag{7}$$

The prior distribution of concept y, $p(y)$, is estimated by the distribution of training data of concept detections [11].

3.2 Discovery of Relevant and Irrelevant Information

S. Johnson have asserted that example is always more efficacious than precept [13]. It is reasonable to use a few query examples to gain much higher performance. In this paper, we mine the relevant and irrelevant information from the query examples provided by the user. In general, shots which contain the same concepts have the similar relevance to a certain query. Motivated by this observation, we view the concepts as the relevant/irrelevant information.

Algorithm 1 MIIL reranking algorithm

Input:
⋆ Initial ranked list: X, $N^{(0)} = |X|$.
⋆ Parameter: λ.
⋆ Concept set: $Y = Y^+ \cup Y^-$.
Output:
⋆ Reranked list: Z.

1: **Initialization**
  Construct sample pair set $T^{(0)}$ with samples in X.
  Estimate $p(y)$ according to the training set of concept detection for every $y \in Y$.
  For $i, j = 1, 2 \ldots N^{(0)}$
    Compute $p(y|t_{ij})$ for every $y \in Y$ as Equation (4).
    Compute $p(t_{ij})$ as Equation (5).
    Compute $L(t_{ij})$ according to Equation (2).
  End for
  The number of loop: $K = ceil(N^{(0)}/2)$.
2: **Main Loop**
  For $i, j = 1, 2 \ldots K$
    Find $t^{(i)} = \arg\max_{t \in T^{(i-1)}} \{I(t, Y^+) - \lambda I(t, Y^-)\}$.
    Add the two samples x_{i1} and x_{i2} in pair $t^{(i)}$ into the new ranked list Z: x_{i1} at the rank $N^{(0)} - i + 1$ while x_{i2} at the rank i, assuming $x_{i2} \succ x_{i1}$.
    Remove the pair constructed by x_{i1} or x_{i2} from $T^{(i-1)}$, and update $T^{(i)}$.
  End for

The mining process is similar to the optimization of search reranking. We sample shots from the initial ranked list randomly and use them to form the example pairs with the query examples. Let Y be concept set, Q be query pair set. The optimal relevant concept set Y^{+^*} and irrelevant concept set Y^{-^*} are obtained as follows:

$$Y^{+^*} = \arg\max_{Y^+ \subset Y} \sum\nolimits_{y^+ \in Y^+, q \in Q} I(q, y^+) \tag{8}$$

$$Y^{-^*} = \arg\min_{Y^- \subset Y} \sum\nolimits_{y^- \in Y^-, q \in Q} I(q, y^-) \tag{9}$$

where we assume that q is binary and $p(q)$ is either 0 or 1. We fix the number of relevant and irrelevant concepts as K to be the constraint of the two functions.

We resolve the two functions by ranking the concepts $y \in Y$ in terms of $I(q, y)$ values. Then we select top K concepts as relevant ones and below K concepts as irrelevant ones.

3.3 Minimum Incremental Information Loss (MIIL) Reranking Algorithm

As aforementioned, there are $N!$ possible reranked lists for the given initial search results with the size of N. Therefore, it is impractical to compare with each other when N is extremely large. In this section, we will present a practical and effective algorithm to solve this optimization problem.

3.3.1 MIIL Reranking

Inspired by the lossy information compression theory, we view reranking as "denoising" problem, such that "noise" is defined as the incompressible part in the data while the compressible part defines the meaningful information bearing signal [14]. In this algorithm, we select "the best possible pair" at each round, and "the best possible pair" is viewed as the compressed data which preserves the most relevant information while excludes the most irrelevant information.

The algorithm starts with a set of pair samples constructed by the initial search results. Let $N^{(0)} = |X|$ be the size of the initial search results, thus there are $M^{(0)} = N^{(0)} \times (N^{(0)} - 1)$ pair samples to form initial pair set $T^{(0)}$. At each round we select an optimal pair t and minus the pairs which are constructed by at least one element in t from $T^{(0)}$. Let $t^{(i)}$ be the selected "the best possible pair" at the i^{th} round, T^i be the current pair sample set and $T^{(i+1)}$ denote the new pair sample set after the minus of several pairs in $T^{(i)}$. The reranking can be formulated as:

$$t^{(i+1)} = \underset{t \in T^{(i)}}{\arg\max} \{I(t, Y^+) - \lambda I(t, Y^-)\} \quad (10)$$

After pair selecting at each round, we map the selected pair into the new ranked list. At the i^{th} round, the two samples of the selected pair are located at the rank i and rank $N^{(0)} - i + 1$ in the new ranked list Z. The pseudo code for the algorithm is given in Algorithm 1.

The criterion presented in Equation (10) is equivalent to finding the pair which has least information loss:

$$t^{(i+1)} = \underset{t \in T^{(i)}}{\arg\min} \{I(T^{(i)}, Y) - [I(t, Y^+) - \lambda I(t, Y^-)]\} \quad (11)$$

where $I(T^{(i)}, Y) = I(T^{(i)}, Y^+) - \lambda I(T^{(i)}, Y^-)$, it is a constant for each $t \in T^{(i)}$. Obviously, the information loss is incremental with rounds increasing. Thus the reranking is formulated as finding most confidential pairs via *minimum incremental information loss* (MIIL).

3.3.2 Computational Complexity

The running time of MIIL algorithm can be split into two parts. The first part is the initialization, which requires $O(n)$ time, where $n = N^{(0)}$. The second part computes the optimal pair t at each round, which requires $O(K \times |T^{(i)}|)$ time. Though in the worse case this is $O(n^3)$, the number of $T^{(i)}$, $|T^{(i)}|$, is decreasing with the rounds increasing. In practical software developing, we can save the results of the difference of mutual information, i.e., $I(t, Y^+) - \lambda I(t, Y^-)$ for each t in $T^{(0)}$ at the first round. Thus the succeeding rounds need no computing any mutual information and sorting. Therefore, the algorithm actually requires $O(n^2)$ time.

Compared with the original formulation of finding the best ranking list among the $N!$ possible solutions, which requires $O(N!)$ time at least, the MIIL algorithm is much more practical and efficient.

4. EXPERIMENTS

We evaluated our video search reranking approach along several dimensions. First, we will report the discovery of relevant and irrelevant concepts; Then, we will compare the proposed reranking approach with the three representative reranking methods. Finally, we will analyze the influence of the parameters mentioned in the optimization functions on the search performance.

4.1 Experimental Settings

We have conducted the experiments on automatic search task over the benchmark TRECVID 2005-2007 test sets, (tv2005, tv2006 and tv2007) [9]. All the videos are segmented into shots provided by NIST, and then the three collections contain 45766, 79484, and 18142 shots, respectively. In each year, 24 query topics with several image/video examples are provided with the ground truth of relevance. For each query, video search systems are required to return a ranked list of up to $1,000$ shots.

For each video, the speech transcript was obtained through automatic speech recognition (ASR), as well as machine translation (MT) for the non-English sources. The text transcripts related to each shot are determined by the shot boundaries. The text-based search results were mainly developed with Okapi BM25 [11] [17] ranking function, which was used as the baseline in the following experiments.

4.2 Evaluation on the Relevant and Irrelevant Information

We conducted the discovery of relevant/irrelevant information based on a series of concept detectors. Specifically, we used the lexicon of 39 concepts defined in LSCOM-Lite [15], and fixed the number of relevant and irrelevant concepts $K = 6$. We list some query topics and select three relevant and irrelevant concepts in Table 1. It can be found that many concepts selected as relevant/irrelevant information of the queries are consistent with human understanding. For example, the three concepts *Walking_Running*, *Vegetation* and *Sports* were detected as relevant information, while *Building*, *Explosion_Fire* and *Office* as irrelevant information for the query "*Topic 0171: Find shots of a goal being made in a soccer match.*" Based on the relevant and irrelevant information, the text baseline of this query was improved by 90% in terms of AP in the following reranking (AP: 0.1779→0.3390).

4.3 Evaluation on Reranking

To demonstrate the effectiveness of the proposed MIIL reranking approach, we compared it with the following three representative reranking approaches: (1) Context reranking [4], (2) PPF-based reranking [2] and (3) NPRF-based reranking [6]. Among them, context reranking is a representative of unsupervised reranking. PPF-based reranking is the first reranking approach which considers the different relevance of samples pairwisely, and NPRF-based reranking is a representative supervised reranking, which also uses the query examples to learn the semantic meaning of query. We choose the parameters in the three approaches which achieve the best reranking performance based on our text baseline. Note that both Context reranking and PPF-based reranking approaches used low-level visual features, such as color moment and texture. We also tried to use features in concept space to perform the reranking and found that the low-level visual features perform better. The features used in NPRF-based reranking approaches are scores of a set of concept detection.

The proposed reranking belongs to supervised reranking since it learns the relevant and irrelevant information from the query examples. The parameter λ in Equation (2) is set to 0.3 empirically, m in Equation (4) and n in Equation (5) is set to 1.0 simply. The effect of these parameters selection will be discussed later.

Table 2 shows the comparison of the four reranking approaches mainly measured by MAP (mean average precision), which is widely accepted to measure the system performance in information retrieval and search [9]. We can see that MAP of each year is significantly improved over the text baseline and the improvement over the text baseline is higher than the other reranking approaches. To avoid bias of MAP, we eliminate search topic *0219* ("*Find shots*

Table 1: Experimental results for relevant and irrelevant concept discovery.

Query	Relevant Concepts	Irrelevant Concepts
0153 Tony Blair	Face, Government-Leader, Person	Vegetation, Animal, Snow
0171 a goal being made in a soccer match	Sports, Vegetation, Walking_Running	Building, Explosion_Fire, Office
0187 one or more helicopters in flight	Waterscape_Waterfront, Outdoor, Sky	Flag-US, Government-Leader, Office
0196 scenes with snow	Snow, Outdoor, Sky	Office, Charts, Sports
0199 a person walking or riding a bicycle	Walking_Running, Outdoor, Road	Vegetation, Office, Computer_TV-screen
0204 a street protest or parade	People-Marching, Crowd, Walking_Running	Computer_TV-screen, Meeting, Office
0212 a boat moves past	Sky, Waterscape_Waterfront, Boat_Ship	Truck, People-Marching, Car

Table 2: Comparisons of the four reranking approaches over TRECVID 2005-2007 test set.

Data Set	Text Baseline	Context		PPF		NPRF		MIIL	
		MAP	Gain	MAP	Gain	MAP	Gain	MAP	Gain
tv2005	0.0447	0.0494	+10.5%	0.0462	+3.4%	0.0464	+3.8%	0.0639	**+43.0%**
tv2006	0.0382	0.0419	+9.7%	0.0421	+10.2%	0.0462	+20.9%	0.0622	**+62.8%**
tv2007	0.0307	0.0318	+3.6%	0.0315	+2.6%	0.0297	-3.3%	0.0178	-42.0%
tv2007 (without *Topic 0219*)	0.0140	0.0147	+5.0%	0.0124	-11.4%	0.0156	+11.4%	0.0182	**+30.0%**

Table 3: Comparisons of the four reranking approaches over TRECVID 2005-2007 test set. #Improve means how many topics gain a performance improvement over the text baseline in terms of MAP. #Best means how many topics gain the best performance among the four reranking approaches in terms of MAP.

Data Set	#Improve				#Best			
	Context	PPF	NPRF	MIIL	Context	PPF	NPRF	MIIL
tv2005	12	11	17	**20**	3	2	7	**15**
tv2006	14	13	18	**21**	1	4	4	**18**
tv2007	14	13	**17**	15	3	2	7	**12**
Sum	40	37	52	**56**	7	8	18	**45**

that contain the Cook character in the Klokhuis series.") in TRECVID 2007. Considering the NPRF-based reranking and MIIL reranking used the concept based features, the correct related concepts ("people", "face") to topic *0219* in our concept lexicon are too general. Thus the two reranking approaches degrade the text baseline for this topic. In the future work, we will enlarge the size of concept lexicon to resolve this problem. From Table 3, among all the 72 search topics, we can see that MIIL reranking improves the text baseline for 56 search topics (78%), and it is best for 45 search topics (63%).

The two supervised reranking approaches, i.e., NPRF-based reranking and MIIL reranking, work better than the two other unsupervised reranking approaches. Although it is somewhat unfair to compare supervised approaches with the completely unsupervised approaches, we can see that it is quite reasonable to obtain a significant improvement with only a few query examples. In fact, reranking contains two components, one is to mine the information from the initial ranked list, and the other is to seek help from the auxiliary knowledge for directing the ranking optimization. In the next sub-section, we will prove that it is important to use auxiliary information to understand the uses' queries better.

4.4 Evaluation on Parameter Sensibility

In the proposed reranking approach, there are three parameters: λ adjusts the weight of relevant and irrelevant information; m and n denotes the importance of the auxiliary knowledge and initial ranking information, respectively. In our previous experiments, we empirically set them. In this sub-section, we try to analyze how these parameters influence the reranking performance.

It is known that the TRECVID 2007 topics are almost all generic, and the generic topics may be more dependent on the visual information than the specific which usually score high on text-based search performance [9] [16]. Moreover, the TRECVID 2007 topics reflect a deliberate emphasis on events, which lead to low performance of conventional text-based search baseline as well as the succeeding reranking. Note that we also eliminate search topic *0219* to analyze the parameter sensibility.

4.4.1 λ

To clarify the effect of λ in Equation (2), we illustrate the performance curves with respect to λ for different data sets in Figure 3. From the figure we can see that λ at the maximum is about 0.3. This proves that it is reasonable to consider both relevant and irrelevant information.

In this experiment, the information is expressed by a set of concepts. Due to the small size of the concept lexicon, it is more difficult to select the absolutely irrelevant concepts. Thus the reranking runs with $\lambda < 1$ outperform with $\lambda > 1$ at most time.

4.4.2 m

As aforementioned, our proposed reranking is developed by combining the auxiliary knowledge in Equation (4) and the initial ranking information in Equation (5). Either m or n can be used to investigate the tradeoff between initial ranking information and auxiliary knowledge. In this sub-section, we present how m impacts the performance.

The MAP with variant m is illustrated in Figure 4. We can see that the best performance is obtained when m is around 1.0. When m is equal to 0, the performance is

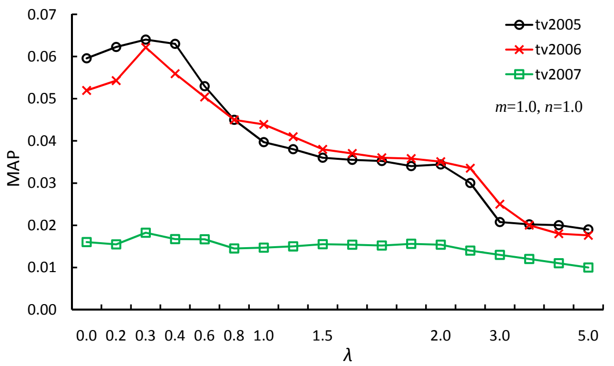

Figure 3: MAP with variant λ.

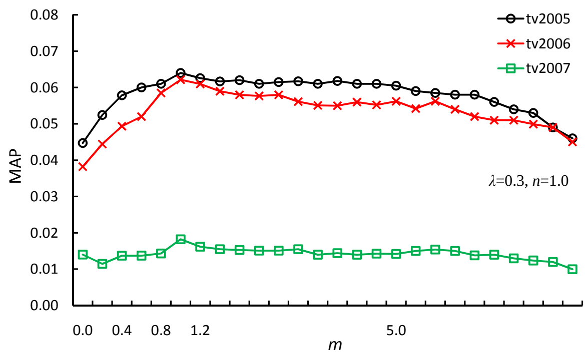

Figure 4: MAP with variant m.

the same as the text baseline, without any improvement. When m goes to infinity, the reranking process relies almost entirely on the auxiliary knowledge ignoring the text baseline, thus the reranking reduces to query-by-example (QBE) problem and the performance degrades. From the above observations, we can find that both initial ranking information and auxiliary knowledge play important roles in the video search reranking.

5. CONCLUSIONS

In this paper, we have presented a novel optimization-based framework for video search reranking, by directly optimizing the entire ranked list rather than individual samples. The proposed approach is general and the weighted difference of mutual information to be optimized can be replaced by the other score functions which can express the relationship between the pairs and relevant/irrelevant information. We also proposed the effective MIIL reranking algorithm, which can improve time complexity of the algorithm and make it more practical. Experiments conducted on the TRECVID 2005-2007 dataset have demonstrated that the proposed reranking approach markedly outperforms the text baselines, as well as the existing reranking approaches.

In the future, we intend to deeply study the selection of relevant and irrelevant concepts. Although the mutual information is an effective measure of the dependence between relevant/irrelevant information and sample pairs, it neglects the semantic relationship between the concepts. In order to deal with complex queries with multiple related concepts, it is an interesting and promising research topic to investigate the semantic linkage between the concepts in concept-query mapping problem.

6. REFERENCES

[1] J. Adcock, M. Cooper, A. Girgensohn, and L. Wilcox. Interactive Video Search Using Multilevel Indexing. In *International Conference on Image and Video Retrieval*, 2005.

[2] Y. Liu, T. Mei, X.-S. Hua, J. Tang, X. Wu, and S. Li. Learning to video search rerank via pseudo preference feedback. In *IEEE International Conference on Multimedia and Expo*, 2008.

[3] W. Hsu, L. Kennedy, and S,-F Chang. Video search reranking via information bottleneck principle. In *ACM International Conference on Multimedia*, 2006.

[4] W. Hsu, L. Kennedy, and S,-F Chang. Video search reranking through random walk over document-level context graph. In *ACM International Conference on Multimedia*, 2007.

[5] L. Kennedy, and S.-F Chang. A reranking approach for context-based concept fusion in video indexing and retrieval. In *International Conference on Image and Video Retrieval*, 2007.

[6] R. Yan, A. Hauptmann, and R. Jin. Multimedia search with pseudo-relevance feedback. In *International Conference on Image and Video Retrieval*, 2003.

[7] J. Tešić, A. Natsev, L. Xie, and J.-R. Smith. Data Modeling Strategies for Imbalanced Learning in Visual Search. In *IEEE International Conference on Multimedia and Expo*, 2007.

[8] Y. Yue, T. Finley, F. Radlinski, and T. Joachims. A Support Vector Method for Optimizing Average Precision. In *ACM SIGIR*, 2007.

[9] TRECVID. http://www-nlpir.nist.gov/projects/trecvid/.

[10] A. Yanagawa, S.-F. Chang, L. Kennedy, and W. Hsu. Columbia university's baseline detectors for 374 lscom semantic visual concepts. *Columbia University ADVENT technical report 222-2006-8, Mar. 20*, 2007.

[11] 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.

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

[13] M.M. Zloof. Query by example. In Proceedings of *National Compute Conference*. AFIPS Press, 44:431-438, 1975.

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

[15] M.R. Naphade, L. Kennedy, J.R. Kender, S.-F. Chang, and J.R. Smith. A Light Scale Concept Ontology for Multimedia Understanding for TRECVID 2005. *IBM Research Technical Report*, 2005.

[16] A. F. Smeaton, P. Over, W. Kraaij. Evaluation campaigns and TRECVid. In *ACM International Conference on Multimedia Information Retrieval*, 2006.

[17] S.-E. Robertson, K.-Spärck Jones. Simple, Proven Approaches to Text Retrieval. *Cambridge University Computer Laboratory Technical Report TR356* , 1997.