Rank canonical correlation analysis and its application in visual search reranking
Zhong Ji*, Peiguang Jing, Yuting Su, Yanwei Pang
School of Electronic Information Engineering, Tianjin University, Tianjin 300072, PR China

Abstract
Ranking relevance degree information is widely utilized in the ranking models of information retrieval applications, such as text and multimedia retrieval, question answering, and visual search reranking. However, existing feature dimensionality reduction methods neglect this kind of valuable potential supervised information. In this paper, we extend the pairwise constraints from the traditional class labels to ranking relevance degrees, and propose a novel dimensionality reduction method called Rank-CCA. Rank-CCA effectively incorporates ranking relevance constraints into standard canonical correlation analysis (CCA) algorithm, and is able to employ the knowledge of both unlabeled and labeled data. In the application of visual search reranking, our proposed method is verified through extensive experimental studies. Experimental results show that Rank-CCA is superior to standard CCA and semi-supervised CCA (Semi-CCA) algorithm, and achieves comparable performance with several state-of-the-art reranking methods while preserving the superiority of low dimensional features.

1. Introduction
With the rapid increase of high dimensional data, dimensionality reduction technique is widely used in many real-world applications, such as data mining, pattern recognition and information retrieval. It aims to transform high dimensional data into a meaningful representation of reduced dimensionality, which is an effective way to mitigate the "curse of dimensionality" and high computational burden and storage cost. Over the past decades, a large number of techniques have been proposed [1–3], from the linear algorithms such as Principal Components Analysis (PCA) and Principal Curves to the nonlinear algorithms such as kernel methods and manifold methods, from the global algorithms such as Isometric Feature Mapping (ISOMAP) to local algorithms such as Locality Preserving Projections (LPP) and Neighborhood Preserving Embedding (NPE), from the unsupervised algorithms such as PCA and LPP to supervised algorithms such as Linear Discriminant Analysis (LDA) and Maximum Margin Criterion (MMC), and from the single-modal algorithms to the multi-view algorithms such as Canonical Correlation Analysis (CCA).

In recent years, many new dimensionality reduction algorithms were designed for particular applications whose data have special characteristics. For example, semi-supervised dimensionality reduction methods [4–7] were proposed to handle the situation where unlabeled examples are readily available but labeled ones are fairly expensive to obtain, which is the common case in tasks such as multimedia retrieval, image annotation, and medical data analysis. Multiple-view dimensionality reduction methods [8,9] were developed to deal with the applications where an instance may have multiple representations from different feature spaces or graph spaces.

In information retrieval tasks, relevance degree information is ubiquitous in ranking techniques, which always plays a key role and have received great attentions from
industry and academic communities. A ranking model can directly take the queries and documents as inputs and compute a matching score using some heuristic formulas, and can also extract some features for each query–document pair and combine these features to produce the matching score [10]. Popular information retrieval applications include visual search reranking [11–13], question answering [14], Video Annotation [15–17], multimedia retrieval and summarization [18–20], collaborative filtering [21], and so on. In most cases, each document will be labeled manually or automatically with a relevance degree to the query in the form of binary judgment or multiple ordered ones (Perfect, Excellent, Good, Fair, or Bad). This kind of special labels is different from the traditional ones, and has led to the emerging of a new research area named learning to rank [10]. However, learning to rank technique only focuses on the leveraging machine learning methods in the ranking process to build effective ranking models, which does not take into account the feature dimensionality reduction problem. Unfortunately, as far as we know, there has been little previous work employing the relevance degree information in the dimensionality reduction techniques.

Generally, multimedia data have multiple modalities, and each modality is usually represented with high dimensional features and has particular statistical properties. CCA copes with the mutual relationships between two modalities to extract the representation of the semantics. It is one of the valuable multi-data processing methods, and has been successfully utilized in many research areas such as multimedia content analysis and retrieval [22], action classification [23], facial expression recognition [24], fMRI data analysis [5], and so on.

Inspired by the pairwise constraints method [4] and semi-supervised canonical correlation analysis (Semi-CCA) method [25], we propose a method to incorporate relevance constraints with the dimensionality reduction technique. In [4], the authors proposed a semi-supervised dimensionality reduction method, which exploited the abundant unlabeled instances together with pairwise constraints to preserve the intrinsic training data structure in the projected low-dimensional space. Generally, pairwise constraints are relatively easy to obtain, which include must-link (instance pairs belong to the same class) and cannot-link (instance pairs belong to different classes) constraints. Peng and Zhang [25] further applied the idea of pairwise constraints to CCA, and presented a semi-supervised method called Semi-CCA. They proved its effectiveness via extensive experiments on UCI handwritten digit dataset, Yale and AR facial datasets.

Based on the framework of CCA, we extend the pairwise constraints from class labels to relevance degrees, and develop a novel dimensionality reduction algorithm named Rank-CCA for information ranking tasks. Compared to previous work, this paper makes the following contributions:

(1) To the best of our knowledge, our approach is the first to consider relevance degree information in dimensionality reduction technique.

(2) A novel Rank-CCA algorithm is proposed, which incorporates the relevance constraints to standard CCA method and can employ the knowledge of both unlabeled and labeled data.

(3) A visual search reranking method is investigated with the proposed Rank-CCA algorithm, which is confirmed to be superior to standard CCA and Semi-CCA algorithm, and is comparable to some state-of-the-art methods while preserving the superiority of low dimensional features.

The remainder of this paper is organized as follows. In Section 2, we provide literature short reviews on the extensions of CCA and visual search reranking methods. Section 3 gives a brief review of CCA. Section 4 formulates and defines Rank-CCA in detail. Experimental results in image search reranking are given in Section 5. At last, we conclude this paper in Section 6.

2. Related works

2.1. Extensions of canonical correlation analysis

Canonical Correlation Analysis (CCA) is a technique for finding basis vectors such that the correlations between the projections of the paired variables onto the corresponding vectors are mutually maximized [26]. It is a fundamental technique in statistics and dimensionality reduction, and is typically used for multi-view data samples.

In recent years, many research studies have been carried out to improve CCA for different situations in different ways. For example, kernel CCA [24] and locality preserving CCA (LPCCA) [27] were developed to reveal the nonlinear relationships hidden behind original data. Kernel CCA first projects the data into a higher-dimensional feature space through well-known “kernel trick”, and then performs CCA in the new feature space. LPCCA decomposes the global nonlinear problem into many local linear ones, which can be processed by linear CCA. And then, the final global problem can be solved by optimizing the combination of these local sub-problems. Recently, Hardoon and Shawe-Taylor proposed a sparse CCA [28] algorithm to focus on the scenario when one is interested in a primal representation for the first view while having a dual representation for the second view. Sparse CCA minimizes the number of features used in both the primal and dual projections while maximizing the correlation between the two views.

Some research works [29,30] aimed at analyzing the correlations between more (than two) datasets and the corresponding methods are called multi-set CCA (MCCA). For example, in [30], the authors proposed an MCCA algorithm for color images, which can extract three color components and provide the analytical solution. It means the method can directly acquire the typically correlative features of three input datasets. In the meantime, a surge of efforts have been made in algorithms for semi-supervised extensions of CCA (Semi-CCA) [5,25,31], which are designed for the commonly encountered situation that the number of labeled samples are limited, while
abundant unlabeled samples are available. For example, to utilize both labeled and unlabeled data, Peng and Zhang [25] took advantage of pairwise constraints on CCA, and Kimura et al. [31] introduced principal component analysis (PCA) into CCA by bridging CCA with paired samples and PCA with paired and unpaired samples via a trade-off parameter.

2.2. Visual search reranking

Visual search reranking is a new paradigm followed by content-based image/video retrieval and concept detection in multimedia content analysis and retrieval domain. It is defined as reordering visual documents (images or video clips) based on the initial text search results by incorporating visual information. The surge of this paradigm is due to the fact that most popular image search engines still perform text-based search techniques by using the metadata associated with media contents, such as the title, comments, and so on. However, this kind of search approach often returns some noisy results on the top of the ranking list since the text cannot entirely reflect the image visual contents. Visual search reranking holds the simple search mechanism preferred by typical users, and exploits the visual information and multimedia analysis methods in another way. Therefore, it integrates characteristics of real-time and accuracy, and has great importance to establish practical image search system.

Over the past decade, visual search reranking has been receiving great attention, and the current approaches can be roughly divided into two categories: unsupervised reranking [12,13,32–34] and supervised reranking [11,35–39]. The former is generally based on the smoothness assumption, which considers that the relevance scores of visually similar documents should be closer. Therefore, this kind of method aimed at discovering and mining the visual patterns that have high visual similarity. Obviously, one of the effective ways is using the clustering algorithm since it can detect the recurrent patterns. For example, Hsu et al. [32] and Wei et al. [33] used information bottleneck and NCuts clustering algorithm, respectively, to refine the initial performance. Recently, graph-based methods have shown promising results, in which a graph is constructed with the visual documents as the nodes and the edges between them being weighted by their visual similarity. In [12,13,34], reranking was formulated as random walk over the graph and the ranking scores were propagated through the edges. Reranking results were finally obtained through the stationary probability distribution. Supervised reranking methods first train a classifier using the train data directly from the initial search results, and then reorder all the documents by the relevance scores predicted from the classifier. The key technical difficulties depend on the selection of train data. Currently, there are mainly three solutions. (1) Choosing some train data manually [35], or with active learning method [36], (2) Adopting the idea of pseudo-relevance feedback, in which a significant fraction of top-ranked documents are assumed to be relevant and are then used to build a model for reranking the search result set [11,37]. (3) Utilization of other information resources [38,39], such as the concept detectors and Internet-related resources.

Furthermore, there are also some new research works beyond the two categories mentioned above. For instance, Zhang et al. [40] presented a novel method by investigating visual search reranking from the view of global optimization. They introduced particle swarm optimization (PSO) mechanism and viewed the reranking as a mapping process from the initial text search list to the objective ranked list. Results diversity is also taken into account in visual search reranking. Wang et al. [41] and Ji et al. [42] studied the joint optimization of search relevance and diversity. Specifically, Wang et al. [41] proved that relevance reranking can be regarded as the process of optimizing the mathematical expectation of the conventional Average Precision (AP) measure, while diversity reranking can be viewed as an optimization process to a new Average Diversity Precision (ADP) measure.

3. Canonical correlation analysis

Consider a set of paired samples \( S_i = \{(x_i, y_i)\} \in \mathbb{R}^p \times \mathbb{R}^q, i = 1, \ldots, n \), where \( \{x_{i1}\} \) and \( \{y_{i1}\} \) are obtained from different information channels. Without loss of generality, we assume that \( X = [x_1, \ldots, x_n] \in \mathbb{R}^p \times n \) and \( Y = [y_1, \ldots, y_n] \in \mathbb{R}^q \times n \) are both centered, which can always be achieved by subtracting the sample means from each sample. The aim of CCA is to seek a pair of linear transforms, \( w_x \in \mathbb{R}^p \) and \( w_y \in \mathbb{R}^q \) for \( X \) and \( Y \), such that correlations between \( X' = w_x X \) and \( Y' = w_y Y \) are maximized. The objective function is formulated as follows:

\[
\rho = \max_{w_x, w_y} \frac{w_x^T C_{xy} w_y}{\sqrt{w_x^T C_{xx} w_x \cdot w_y^T C_{yy} w_y}}
\]

where \( C_{xx} = XX^T \in \mathbb{R}^{p \times p} \), \( C_{yy} = YY^T \in \mathbb{R}^{q \times q} \) are the within-set covariance matrices and \( C_{xy} = XY^T \in \mathbb{R}^{p \times q} \) is the between-sets covariance matrix and \( C_{xy} = \mathbb{C}_{xy} \). More details about derivation and solution of CCA can be found in [22].

4. Rank canonical correlation analysis

Relevance degree information is different from the class label in nature. The former measures the degree of closeness for a document to the query in information retrieval applications, while the latter refers to a common attribute for one category in machine learning and pattern recognition tasks. For example, in conventional pattern classification tasks, instances from the same category always share similar characteristics, and those instances from different categories generally have different characteristics, so semi-supervised dimensionality reduction methods can be adopted with must-link and cannot-link constraints [4,25]. However, in ranking applications, examples in different relevance degree may still possess similar characteristics since they are related to the same query. Therefore, pairwise constraints of must-link and cannot-link constraints cannot be employed directly.

Without loss of generality, we consider only the three-level relevance degree label method, which is the label of
“very relevant”, “relevant” and “irrelevant”. And their data sets are represented with A, B, C, respectively. Fig. 1 illustrates some example images at different relevance degree for the same query. We can see that images from set A and set B have some kind of similarity for the same query, since they reflect the same topic. We can also discover that images in set C are dissimilar even for the same query. Therefore, based on the framework of CCA, we obtain the following relevance constraints.

(1) Except for set C, different modality features between different instances from the same group have the maximum correlation. We regard these as relevant constraints, i.e. A-A and B-B.

(2) Since images in set A and set B have visual similarity to some extent, we also regard A-B as relevant constraints. However, we add a coefficient $\alpha \in [0,1]$ to control its effect.

(3) The constraints A-C, B-C and C-C are set to be irrelevant constraints, which mean that different modality features between different instances from them have the minimum correlation. This is due to the fact that set C is actually a clutter group, its images have no visual similarity with the others either in intra-set or in inter-set.

Accordingly, in order to improve the performance of CCA in ranking applications, we incorporate the constraints of relevance degree information into CCA, which is referred to as Rank-CCA. The standard CCA optimization problem is modified so that the cross-covariance matrix $C_{xy}$ in Eq. (1) is replaced by a term $\hat{C}_{xy}$ that takes relevance constraints into account. Note that both $x$ and $y$ have been centralized. The objective function is formalized as

$$\rho = \max_{w_x, w_y} w_x^T \hat{C}_{xy} w_y / \sqrt{w_x^T C_{xx} w_x \cdot w_y^T C_{yy} w_y}$$ (2)

where

$$\hat{C}_{xy} = XY ^ {\top} + \gamma \left[ \sum_{(x,y) \in \alpha(AA)} (x_iy_j)^T + x_iy_j^T \right] + \sum_{(x,y) \in \beta(AB)} (x_iy_j)^T + x_iy_j^T \right]$$

$$+ \sum_{(x,y) \in \gamma(AC)} (x_iy_j)^T + x_iy_j^T \right] - (1-\gamma) \left[ \sum_{(x,y) \in \beta(BC)} (x_iy_j)^T + x_iy_j^T \right]$$

$$= XEY^T + \gamma (XC_{AA}Y^T + XC_{BB}Y^T + 2XC_{AB}Y^T)$$

$$- (1-\gamma) (XC_{CC}Y^T + XC_{AC}Y^T + XC_{BC}Y^T)$$

$$= X(E + \gamma C_{AA} + \gamma C_{BB} + \gamma 2C_{AB})$$

$$- (1-\gamma) C_{CC} - (1-\gamma) C_{AC} - (1-\gamma) C_{BC} Y^T = XSY^T$$ (3)

In Eq. (3), $E$ is an identity matrix, $C_{AA}$, $C_{BB}$, $C_{CC}$, $C_{AB}$, $C_{AC}$ and $C_{BC}$ are the constraint matrices, and $\alpha, \beta, \gamma \in [0,1]$ are the weighting parameters.

Similar to CCA, the solution of Rank-CCA can also be obtained by computing a generalized eigenvalue decomposition problem, and the solution of Rank-CCA is equivalent to the optimal problem as follows:

$$\max_{w_x, w_y} w_x^T \hat{C}_{xy} w_y$$

s.t. $w_x^T C_{xx} w_x = 1$, $w_y^T C_{yy} w_y = 1$ (4)

The corresponding Lagrangian is

$$L(\lambda, \lambda_x, \lambda_y, w_x, w_y) = w_x^T \hat{C}_{xy} w_y - \frac{\lambda_x}{2} \left[ w_x^T C_{xx} w_x - 1 \right] - \frac{\lambda_y}{2} \left[ w_y^T C_{yy} w_y - 1 \right]$$ (5)

Taking derivatives with respect to $w_x$ and $w_y$, we obtain

$$\frac{\partial L}{\partial w_x} = \hat{C}_{xy} w_y - \frac{\lambda_x}{2} \left[ C_{xx} w_x + C_{xx} w_x \right] = \hat{C}_{xy} w_y - \lambda_x C_{xx} w_x = 0$$ (6)

$$\frac{\partial L}{\partial w_y} = (w_x^T \hat{C}_{xy} w_y - \frac{\lambda_y}{2} \left[ C_{yy} w_y + C_{yy} w_y \right]) - \hat{C}_{xy} w_x - \lambda_y C_{yy} w_y = 0$$ (7)

Subtracting $w_y^T$ times the first equation from $w_x^T$ times the second, we have

$$w_x^T \hat{C}_{xy} w_y - \lambda_x w_y^T C_{xx} w_x = 0$$ (8)

$$w_y^T \hat{C}_{xy} w_x - \lambda_x w_x^T C_{yy} w_x = 0$$ (9)

$$0 = w_x^T \hat{C}_{xy} w_x - \lambda_x w_y^T C_{yy} w_y - w_y^T \hat{C}_{xy} w_x + \lambda_x w_x^T C_{xx} w_x$$

$$= -\lambda_x w_x^T C_{yy} w_x + \lambda_x w_x^T C_{xx} w_x = -\lambda_x + \lambda_x = 0$$ (10)

Let $\lambda_x = \lambda_y = \lambda$, assuming $C_{xx}$ is invertible, we have

$$w_x = \frac{1}{\lambda} C_{xx}^{-1} \hat{C}_{xy} w_y$$ (11)
and so substituting in Eq. (9) gives
\[ \mathcal{C}_{yy}^{-1/2} \mathcal{C}_{xy} W_x = \lambda^2 \mathcal{C}_{yy} W_y \]

Similarly, we can have
\[ \mathcal{C}_{yy}^{-1/2} \mathcal{C}_{yx} W_x = \lambda^2 \mathcal{C}_{xx} W_x \]  

Then we employ singular value decomposition (SVD) to solve Rank-CCA equation following Ref. [43]. Let \( Z = \mathcal{C}_{xx}^{-1/2} \mathcal{C}_{xy}^{-1/2}, \) \( u = \mathcal{C}_{xx}^{-1/2} W_x, \) \( v = \mathcal{C}_{yy}^{-1/2} W_y, \) Eqs. (12) and (13) can be rewritten as

\[
\begin{align*}
Z^T u &= \lambda^2 u \\
Z^T Z v &= \lambda^2 v
\end{align*}
\]

Let \( Z = UDV^T = \sum_{i=1}^{d} u_i v_i^T \) be the SVD decomposition of matrix \( Z, \) where the \( i \)th diagonal element of diagonal matrix \( D \) is just \( \lambda_i u_i \) and \( v_i \) is the \( i \)th row of matrices \( U \) and \( V, \) respectively, corresponding to singular value \( \lambda_i, \) we have
\[
\begin{align*}
W_d &= \mathcal{C}_{xx}^{1/2} u_i \\
W_y &= \mathcal{C}_{yy}^{1/2} v_i
\end{align*}
\]

From Eq. (15), we obtain the \( i \)th \( (i = 1, \ldots, d, \) \( d \leq \min(p,q) \)) pairs of basis vector of Rank-CCA.

After obtaining eigenvectors \( W_x = [w_{x1}, \ldots, w_{xd}] \in \mathbb{R}^{d \times p} \) and \( W_y = [w_{y1}, \ldots, w_{yd}] \in \mathbb{R}^{d \times q}, \) we get new transform eigenvectors \( W_x^* = [w_{x1}, \ldots, w_{xd}] \in \mathbb{R}^{d \times d}, \) \( W_y^* = [w_{y1}, \ldots, w_{yd}] \in \mathbb{R}^{d \times d} \) by taking the first \( d \) \((d \leq \min(p, q)) \) eigenvectors from \( W_x \) and \( W_y. \) Then for any sample \((x, y), \) we can extract features as \([W_x^* x, W_y^* y]^T.\) The pseudo-code of Rank-CCA algorithm is summarized in Table 1.

5. Experiments on visual search reranking with Rank-CCA

In this section, we evaluate the performance of the proposed Rank-CCA algorithm in image visual search reranking task. We first introduce the dataset and methodologies, and then demonstrate the effectiveness of Rank-CCA algorithm from two aspects. On one hand, we label the relevance degree to some images manually. With these true labeled data and unlabeled data, we demonstrate that our proposed Rank-CCA algorithm is superior to standard CCA [26] and Semi-CCA [25] algorithms in ranking applications. On the other hand, we adopt a pseudo-relevance feedback method to label the images automatically, which aims to prove that Rank-CCA based visual search reranking algorithm is comparable to some state-of-the-art methods with pseudo-true label data and unlabeled data. Note that for all the statistical experiments, we repeat them for three times and report the average results.

5.1. Dataset and methodologies

We conduct experiments on the publicly available MSRa-MM image dataset [44] which consists of 68 popular queries collected from the image search engines of Microsoft Bing Search. These queries cover a wide variety of categories, including objects, people, event, entertainments, and location. For each query, about top 1000 images along with the surrounding texts are collected. As a result, the dataset contains 65,443 images in total. The rank orders of these images are obtained as the initial ranked lists.

In the dataset, each image to the corresponding query was manually labeled with three levels: (0) “irrelevant,” (1) “relevant,” and (2) “very relevant.” To evaluate the ranking performance, Normalized Discounted Cumulative Gain (NDCG) is adopted, which is widely used in information retrieval tasks, especially when there are more than two relevance levels [45]. Given a query \( q, \) the NDCG score at the depth \( d \) in the ranked documents is defined by

\[
\text{NDCG@}d = \frac{1}{\log (1 + j)} \sum_{j=1}^{d} \frac{2^{r_j} - 1}{\log (1 + j)}
\]

where \( r_j \) is the rating of the \( j \)th document, \( Z_d \) is a normalization constant and is chosen so that a perfect ranking's NDCG@\( d \) value is 1.

To make our results reproducible and to enable comparisons with other approaches, we adopt the provided features. Specifically, we select 144D color correlogram and 128D wavelet texture as the image features since they come from different modalities. And they are \( x \) and \( y \) in Eq. (3). In addition, we used top 500 images in the initial search results for reranking in our experiments, since it is typical that there are very few relevant images after the top 500 search results [12].

The framework of the proposed Rank-CCA based image visual search reranking method is given in Fig. 2. Take the

Table 1

| Input: | Training data: \( S_i \{ i \in [1, n] \}, \) of which \( S_A, S_B, S_C \) are the labeled subset for sets \( A, B \) and \( C, \) respectively. Parameters: \( x \) and \( y \) |
| Output: | Projection vectors: \( W_x = [w_{x1}, \ldots, w_{xd}] \in \mathbb{R}^{d \times p} \) and \( W_y = [w_{y1}, \ldots, w_{yd}] \in \mathbb{R}^{d \times q}, \) Constraints: \( \mathcal{C}_{xx}, \mathcal{C}_{yy}, \mathcal{C}_{xy}, \mathcal{C}_{yx}, \mathcal{C}_{AB}, \mathcal{C}_{BC}, \mathcal{C}_{CA}, \mathcal{C}_{AC}, \mathcal{C}_{AC} \) and \( \mathcal{C}_{BC} \) to be \( n \times n \) zero matrix |
| Step 1: | Centering data: \( x_i^* = x_i - \bar{x}, y_i^* = y_i - \bar{y}, \) \((i = 1, \ldots, n)\) |
| Step 2: | Initialize constraint matrices \( \mathcal{C}_{xx}, \mathcal{C}_{yy}, \mathcal{C}_{xy}, \mathcal{C}_{yx}, \mathcal{C}_{AB}, \mathcal{C}_{BC}, \mathcal{C}_{CA}, \mathcal{C}_{AC}, \mathcal{C}_{AC} \) and \( \mathcal{C}_{BC} \) to be \( n \times n \) zero matrix |
| Step 3: | Compute constraint matrices |
| Step 4: | Compute covariance matrices \( \mathcal{C}_{XX}, \mathcal{C}_{YY}, \mathcal{C}_{XY}, \mathcal{C}_{YX}, \mathcal{C}_{AB}, \mathcal{C}_{BC}, \mathcal{C}_{CA}, \mathcal{C}_{AC}, \mathcal{C}_{AC} \) to be \( n \times n \) zero matrix |
| Step 5: | Compute matrix \( Z = \mathcal{C}_{xx}^{-1/2} \mathcal{C}_{xy} \) |
| Step 6: | Perform SVD decomposition \( Z = UDV^T \) |
| Step 7: | Choose \([u_1, \ldots, u_d] \) and \([v_1, \ldots, v_d] \), \( d < q \) |
| Step 8: | Obtain \( W_x = \mathcal{C}_{xx}^{-1/2} [u_1, \ldots, u_d], W_y = \mathcal{C}_{yy}^{-1/2} [v_1, \ldots, v_d] \)
query term “boy” as an example. When “boy” is submitted to the web image search engine, an initial text-based search result is returned to the user (only the top ten images are given for illustration). The result is unsatisfactory because some woman and cartoon images are retrieved as top results. To rerank these images, multimodal features are first extracted to represent their visual contents. And then, since there is generally no explicit training data, a manual labeling or pseudo-relevance feedback mechanism is adopted to label some data with relevance degrees. Both labeled and unlabeled data are employed in the proposed Rank-CCA algorithm to reduce the image features’ dimensionality. Next, Ranking SVM algorithm [46] is taken to use the labeled data as training data to build a ranking function. Finally, all the images are reranked with the reduced low dimensional features and ranking function.

In Rank-CCA, Semi-CCA and standard CCA algorithms, we set $d=30$, i.e. reduce the original feature dimension to 30D. We randomly select $k=10$ labeled images from each relevance degree group. The trade-off parameter $C$ is set to be 0.1 in the Ranking SVM model. In addition, $\alpha$ and $\gamma$ are set to 0.4 and 0.8, respectively, in Rank-CCA.

### 5.2. Visual search reranking with manually labeled data

In this section, we demonstrate the performance of Rank-CCA with manually labeled data in image visual search reranking task. Fig. 3 presents the NDCG results with different depths for algorithms of CCA, Semi-CCA and Rank-CCA, which use the same training data. The baseline refers to the performance of the initial text-based search result. From the figure we can observe that (1) The proposed Rank-CCA algorithm outperforms the others. (2) Both Rank-CCA and Semi-CCA have significant performance improvements to the baseline and standard CCA, which demonstrate pairwise constraints are effective to enhance the discriminative power of original features. (3) The superiority of Rank-CCA to Semi-CCA identifies the relevance constraints are more suitable than must-link and cannot-link constraints in ranking tasks. (4) The performance differences between Rank-CCA and the others decline with the increase of the depth. However, it becomes stable when the depth is greater than 40. This phenomenon demonstrates that Rank-CCA is a robust algorithm and can achieve steady performance gain.

We then evaluate the impact of weighting parameters $\alpha$ and $\gamma$ in Eq. (3). Fig. 4 depicts the performance of the Rank-CCA based reranking method with different $\alpha$ and $\gamma$ ranging from 0.0 to 1.0 in terms of NDCG@50. From the figure, we discover that the best weight for $\gamma$ is 0.8, which means that the relevant constraints are more important than irrelevant constraints. Even if we only use the relevant constrains ($\gamma=1.0$), the performance is relatively better than most of the other cases. Our conjecture is that relevant constraints are more helpful to get the “visual pattern” [12,32] which is a basic assumption in reranking applications. It can also be found that 0.4 is the best
choice for $\alpha$. It confirms the fact the relevant constraint $A-B$ has a smaller impact than $A-A$ and $B-B$.

Further, the influence of a different labeled number $k$ is investigated in Fig. 5. We can see that the performance steadily rises with the increase of $k$. It demonstrates that more constraints in number can bring more improvement in performance. Meanwhile, it also confirms that Rank-CCA is superior to Semi-CCA.

5.3. Visual search reranking with pseudo-relevance feedback

In this section, we demonstrate the performance of Rank-CCA with training data automatically acquired with pseudo-relevance feedback (PRF). PRF is initially introduced in [47], and has been shown to be effective in improving initial text search results in both text and video retrieval [11,37]. We modify the typicality-based idea developed by Liu et al. [37] to carry out PRF. More specifically, the initial search results are first clustered into a set of clusters with affinity propagation [48]. And then, pseudo-relevance scores of each sample are obtained by combining the cluster typicality and local typicality measures [37], which are decided by visual similarities and initial order information. Finally, the samples are ordered by descending pseudo-relevance scores, from which we select the labeled data from the top, middle and bottom.

To demonstrate the effectiveness of the proposed Rank-CCA method in reranking task, we implement the performance comparison with the following three state-of-the-art reranking algorithms, as illustrated in Fig. 6. In the three approaches, the parameters are selected to achieve the best performance, and the results are obtained from the bar histogram presented in [12].

(a) Harvesting reranking (Harvesting-TV) [11]: A typical supervised reranking method used both the textual and visual features, and also utilized PRF approach to get training data.

(b) Co-reranking (Co-reranking-TV) [12]: A typical graph-based reranking method which reinforced the textual and visual information mutually via coupling two random walk graphs.

(c) Context reranking (Context-V) [13]: A classic graph-based reranking method which made use of random walk with visual features to perform reranking.

It should be mentioned that the visual features employed in the three methods were mainly based on bag-of-visual-words (BOW). From Fig. 6, we observe that the proposed Rank-CCA method has comparable performance with the other methods. Co-reranking is better than ours, because it employed a novel mutual reinforcement mechanism with both text and visual features. Only in terms of visual features, the BOW feature it used is up to 2000D. However, ours is only 60D from 272D low-level features. Moreover, the performance of Rank-CCA is similar to Harvesting reranking and Context reranking.
methods, which indicates the effectiveness of the relevance constraints in feature dimensionality reduction. 

Fig. 7 shows the top ten images of some queries: “Hawaii”, “Baby” and “Cat.” We can observe that the proposed Rank-CCA method gets satisfactory reranking results.

6. Conclusions

This paper has proposed an effective and practical algorithm for dimensionality reduction in ranking tasks. The proposed Rank-CCA algorithm introduces relevance constraints into CCA method, and achieves superior performance than standard CCA and Semi-CCA algorithms. In addition, only employing low-level and low-dimensional visual features, Rank-CCA performs similarly to several state-of-the-art methods in image visual search reranking application.

Further, we will incorporate the relevance constraints into some extensions of CCA, such as kernel CCA and sparse CCA to handle more complicated circumstances. Multiple-view dimensionality reduction methods, such as the method of Multiview Spectral Embedding (MSE) proposed in [9], are also very effective to deal with multiple modalities in multimedia data, and therefore will be one of our study directions. And moreover, we will employ these dimensionality reduction algorithms to perform some other ranking tasks, such as question answering and multimedia summarization.

Acknowledgments

The work was supported by the National Natural Science Foundation of China (Grant nos. 60975001, 61172121, 61170239), the Tianjin Research Program of Application Foundation and Advanced Technology (Grant nos. 10JCQYJC07700, 09JCQYJC00900), the Specialized Research Fund for the Doctoral Program of Higher Education (no. 20090032110028), the Innovation Foundation of Tianjin University (no. 60302019), and the Program for New Century Excellent Talents in University (no. NCET-10–0620). We would like to thank Linjun Yang in MSRA for providing the public MSRA-MM Dataset.

References

学霸图书馆
www.xuebalib.com

本文献由“学霸图书馆-文献云下载”收集自网络，仅供学习交流使用。

学霸图书馆（www.xuebalib.com）是一个“整合众多图书馆数据库资源，提供一站式文献检索和下载服务”的24小时在线不限IP图书馆。

图书馆致力于便利、促进学习与科研，提供最强文献下载服务。

图书馆导航：
图书馆首页 文献云下载 图书馆入口 外文数据库大全 疑难文献辅助工具