Group Ranking with Application to Image Retrieval

Ou Wu, Weiming Hu, Bing Li
NLPR, Institute of Automation, Chinese Academy of Sciences, Beijing, China
{wuou, wmhu, bli }@nlpr.ia.ac.cn

ABSTRACT
Many existing ranking-related information processing applications can be summarized into one theoretical problem called group ranking (GR). A simple average-ranking approach is usually applied to GR. Although the approach seems reasonable, no theoretical analysis about its intrinsic mechanism has been presented, increasing the difficulty of evaluating the ranking results. This study provides a formal analysis for GR. We first construct an objective function for the GR problem, and discover that each GR problem can be transformed into a rank aggregation problem whose objective function is proved to be equal to the objective function of GR. As a consequence, the average-ranking approach can be explained by two well-known rank aggregation techniques. We incorporate two other effective rank aggregation methods into the GR problem and obtain two new GR algorithms. We apply the GR algorithms into image retrieval to diversify the image search results returned by search engines. Experimental results show the effectiveness of the proposed GR algorithms.

Categories and Subject Descriptors
D.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval – *Retrieval models*.

General Terms
Algorithms, Experimentation, Theory.

Keywords
Group Ranking, Rank Aggregation, Cluster Raking, Visual Diversity.

1. INTRODUCTION
In many real information processing applications, placing in order a set of instance groups such as people, document, product, and image groups is necessary. For example, school administrators usually require ranking classes, each of which contains dozens of students. Search engines rank web pages, each of which consists of several page blocks. We summarized such problems into one theoretical problem called group ranking (GR), which aims to rank a set of instance groups instead of a set of instances. A simple yet popular approach in addressing GR is to utilize the average-ranking algorithm, which calculates the average score/rank of each group and then order the average scores to achieve group order. For example, school administrators calculate the average scores of each class (i.e., student group) average scores and then arrange classes by their corresponding average scores. Although this heuristic strategy seems quite reasonable, it lacks a theoretical analysis: How can we explain the average-ranking approach in theory? Are the results of this approach reasonable? Are there any other alternative solutions?

CIKM'10, October 25–29, 2010, Toronto, Ontario, Canada.

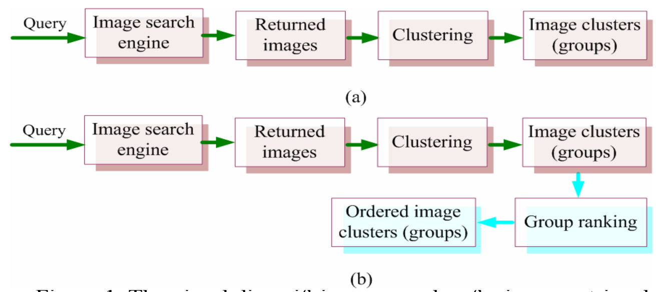

Figure 1. The visual diversifying approaches for image retrieval: (a) the conventional approach, (b) our approach.

To answer the questions listed above, we make a formal study on the GR problem. We first formalize the GR problem into an optimization problem geared toward a defined objective function. We are able to prove that the objective function is equal to a Kemeny optimal rank aggregation function. Rank aggregation techniques can be adapted to solve the GR objective function. Average-ranking strategy can be explained by two classical rank aggregation techniques, namely, linear combination method (LCM) [3] and Borda's count (BC)[7]. This conclusion conversely shows the rationality of the proposed objective function as well as the problem transformation. We further propose several new methods in the light of rank aggregation studies.

We explore the utilization of GR studies into the visual diversification of image search results. In visual diversification [1], the search images returned by search engines are clustered into a set of image groups to make the returned image as diverse as possible. This approach is shown in Fig. 1(a). Upon obtaining the image clusters (groups), these clusters are arranged from large to small according to the contained image counts in each cluster. Note that the relevance of clusters is also important to users. Existing studies ignore the ranking of image clusters and lack deep investigation on the ranking of the returned image clusters. In essence, the problem in ranking of image clusters is actually the problem of ranking a set of groups. We apply our proposed GR algorithms to rank the returned image clusters. Our approach is shown in Fig. 1(b).

The remainder of this paper is organized as follows: Section 2 briefly reviews related issues. Section 3 gives a formal description of the GR problem. Section 4 transforms GR into a classical rank aggregation problem. Section 5 introduces the solutions for GR. Section 6 reports experimental results on image retrieval. Conclusions are given in Section 7.

2. RELATED RESEARCH

In this section, we will introduce a related work, namely, rank aggregation. This study explores the connections between rank aggregation and our proposed GR problem. Rank aggregation is a fundamental and classical optimization problem addressed in various areas including theoretical computer science, economics, statistics, and information retrieval [2]. It combines different rank orderings on the same set of instances to obtain a "better" ordering. Let τ represent an ordering list and $\tau(\cdot)$ represent the position of a given element in a rank list. Formally, rank aggregation aims to find an optimal rank list τ^* with the following objective function:

$$\tau^* = \min_{\tau} \sum_i d(\tau,\ \tau_i)$$

where $d(\cdot,\cdot)$ is the distance function between the lists τ and τ_i. A widely accepted distance function for two fully ordered lists is the *Kendall* tau distance [2]:

$$K(\tau_i, \tau_j) = |\{(x, y) \mid \tau_i(x) < \tau_i(y),\ \text{but}\ \tau_j(x) > \tau_j(y)\}|$$

There are numerous rank aggregation techniques. One simple and well-known method is Borda's count (BC)[7]. This method assigns a weight to each element. Elements can be ranked according to their weights in increasing order. Some other methods calculate the weights by combining the scores of elements in a linear way. One linear combination method (LCM) proposed in [3] calculates the weights according to the elements' scores. Both the BC and LCM methods can be seen as direct aggregation approaches. There are some other methods utilizing optimization approach. Dwork [2] proposed the Markov chains method to find the *Kemeny* optimal aggregation. Markov chains method is a series of algorithms differed in the strategy of constructing the state transmission matrix. We will bring two Markov chains algorithms namely MC1 and MC4 proposed in [2] into our study.

3. PROBLEM DESCRIPTION

In this section, we attempt to give a formal description for GR. We first define some symbols used in the paper. Let $T = \{B_1, B_2, \ldots, B_{|T|}\}$ be a set of groups, where each element denotes a group of instances. We use x to represent an arbitrary instance. A given group is then represented as $B_i = \{x_{i1}, x_{i1}, \cdots, x_{i|B_i|}\}$. All the instances compose the instance set denoted by X. Let *Scs* denote the score list and $Scs(x_{ij})$ be the score of the instance x_{ij}. When referring to a general rank list, we use τ; if referring to a rank list of groups over T, we use τT; and when referring to a rank list of instances over X, we use τX. The smaller of the value of $\tau(\cdot)$, the higher order the element ranks. As the score list *Scs* is already known for a given GR problem, we are able to derive an ordering list over X according to the scores. We denote this derived rank list as Γ (in some cases, lower score means better rank; thus, the above rule should be modified accordingly). With the above symbols, the GR problem can be formally described as follows:

Group Ranking (GR) Problem: *Given a set of groups T and the score list (Scs) as well as the instance rank list Γ, how can we derive a reasonable ordering of all the groups?*

As mentioned in the beginning of the paper, the widely used approach for GR is based on the ranking of average scores/ranks of the groups. The average scores (a_s_i) and ranks (a_r_i) are:

$$a\_s_i = \sum\nolimits_{j=1}^{|B_i|} Scs(x_{ij}) / |B_i| \ \text{ and } \ a\_r_i = \sum\nolimits_{j=1}^{|B_i|} \Gamma(x_{ij}) / |B_i|$$

These two formulas seem quite simple and reasonable. Nevertheless, they lack theoretical basis.

In other ranking studies such as rank learning and rank aggregation, *pairwise comparison* is the base stone of the whole task [2, 4]. Motivated by these studies, we first attempt to establish a rule to compare any two groups by considering the following case:

Round-robin comparison: Consider a given pair of groups named B_i and B_j. Each instance in B_i takes turns to compare with each instance in B_j. If the instance from B_i ranks better, the score of B_i increases by 1; vise versa, the score of B_j increases by 1. There are $|B_i|\cdot|B_j|$ times of comparison. The comparing results (scores) are represented as follows:

$$\begin{aligned} \Theta_{ij} &= \Theta(B_i \succ B_j) = |\{(x_{il}, x_{jk}) \mid \Gamma(x_{il}) < \Gamma(x_{jk}), x_{il} \in B_i, x_{jk} \in B_j\}| \\ \Theta_{ji} &= \Theta(B_j \succ B_i) = |\{(x_{jl}, x_{ik}) \mid \Gamma(x_{jl}) < \Gamma(x_{ik}), x_{ik} \in B_i, x_{jl} \in B_j\}| \end{aligned} \quad (1)$$

In the round-robin comparison, if $\Theta_{ij} > \Theta_{ji}$, B_i wins; else if $\Theta_{ij} < \Theta_{ji}$, B_j wins; otherwise, they are equal. Note that numbers of instances in different groups may differ a lot. The comparing results are normalized as below:

$$\theta_{ij} = \Theta_{ij} / (|B_i| \cdot |B_j|), \quad \theta_{ji} = \Theta_{ji} / (|B_i| \cdot |B_j|) \quad (2)$$

If each instance of B_i ranks higher than all the instances in B_j, θ_{ij} equals 1; if each instance of B_i ranks lower than all the instances in B_j, θ_{ij} equals 0. These two cases are consistent with the real applications and our intuition. Based on the pairwise comparison, we are now able to measure a candidate rank list τT by the index defined below:

Definition 1 (***Inconsistency index***)

$$Q(\tau T) = \sum_{i,j} \theta_{ji} \varpi_{ij} \quad (3)$$

where

$$\varpi_{ij} = \begin{cases} 1 & \text{if } \tau T(B_i) < \tau T(B_j) \\ 0 & \text{otherwise} \end{cases}$$

And θ_{ij} is defined in Eq. (2). This index reflects the inconsistency between a candidate rank list τT and the pairwise round-robin comparing results derived from a given GR problem. The lower the $Q(\tau T)$, the better the list τT. As a result, the goal of GR is formalized into the searching of an optimal rank list σ such that $Q(\sigma)$ is minimized. The GR problem aims to solve the following objective function:

$$\sigma = \min_{\tau T} Q(\tau T) \quad (4)$$

The inconsistency index virtually reflects the intensity of the "opposed sound" towards a candidate rank list. To solve the function, we transform GR into the rank aggregation problem and explore the connections between them in the following section. We find that the GR problem can be addressed by using rank aggregation techniques introduced in Section 2.

4. PROBLEM TRANSFORMATION

In this section, we transform the GR problem into a typical rank aggregation problem. We prove that the objective function of GR is equivalent to its transformed rank aggregation.

We are able to generate a number of instance combinations from T. Each combination consists of $|T|$ instances coming from the groups in T separately (different instances come from different groups). The total number of such combinations we can generate is $|B_1| \times |B_2| \times \cdots \times |B_{|T|}|$. Given that the order of each instance in each combination is already known, each combination can be arranged into a rank list of the instances it contained. We name the set of combination rank lists as Ω.

If each instance is replaced with the group it locates, each combination rank list becomes a rank list of all the groups in T. This means each combination (instance) rank list can produce a group rank list. We denote the set of the produced group rank lists as Φ. Given that we have produced a number of full rank lists for all the groups in T, we obtain an optimal rank list by solving the following rank aggregation objective function:

$$\tau^* = \min_{\tau T} \; d(\tau T, \Phi) \qquad (5)$$

where τ^* denotes the optimal rank list and $d()$ is a distance function. Note that each GR problem corresponds to a rank aggregation problem. Consequently, we are interested to know whether or not there is connection between the two optimal objective functions, i.e., Eq. (4) and Eq. (5). We have the following Theorem.

Theorem 1: *Eq.(4) and Eq.(5) are equivalent when Kendall tau distance is used.*

Proof. Ommited due to lack of space (The interested readers can refer to the full version of this paper).

Theorem 1 reveals the GR problem is equivalent to the *Kemeny* optimal rank aggregation problem transformed from GR. As a result, many useful properties can be directly inherited from rank aggregation studies and applied to GR. In the following section, we attempt to solve Eq. (4) by solving Eq. (5). It is should be ***noted*** that, during the calculation steps it is ***unnecessary*** to produce the combination rank lists so we can obtain all the middle parameters without accessing the rank list sets Φ and Ω.

5. SOLUTIONS

This section will show that the heuristic average-score/rank approach is essentially identical to two rank aggregation techniques. To differ from the rank aggregation techniques, the methods proposed for GR are added by the prefix 'GR-'. There are two main classes of the rank aggregation techniques for *Kemeny* optimal aggregation (also for GR): direct approaches and optimization approaches.

5.1 Direct approaches

A simple way of approximately solving the objective function given in Eq. (5) is using the linear combination aggregation method (LCM). The weight for a group is calculated as:

$$w(B_i) = |\Phi| \cdot \sum_{j=1}^{|B_i|} Scs(x_{ij}) / |B_i| \propto \sum_{j=1}^{|B_i|} Scs(x_{ij}) / |B_i|$$

This formula shows that the weight of a group is the average score of its contained instances. By ordering the weights of all the groups, we can obtain their order. We call this method GR-LCM. Similarly, we can obtain a new method which is called GR-BC based on Borda's count (BC) method. Obviously, the above two methods are identical to the average-ranking (average score/rank) approach introduced in the beginning of the paper. That is, the average score/rank approach can be viewed as the direct approached in aggregation approach and thus its properties can be inherited from studies for LCM and BC.

5.2 Approximate optimization approaches

Next, we modify the Markov chains method proposed in [2] to address the GR problem. The key step of this method is the transition probability calculation between any two states (here, we used groups). For an arbitrary group B_i, its multiset is named as $S(B_i)$. Let $\#B_{ij}$ be the number of B_j in $S(B_i)$. Assume the current state is B_i and the next state is B_j. Based on Lemma 1, we can obtain:

$$\#B_{ij} = |\{\tau T \mid \tau T(B_j) \le \tau T(B_i),\ \tau T \in \Phi\}| = |\Phi|(1-\theta_{ij})$$

When j equals i, we prescribe that $\#B_{ii} = |\Phi|$. For the MC1 matrix, if i does not equal to j,

$$M_{ij} = p(B_i \to B_j) = \frac{\#B_{ij}}{\sum_{k\ne i} \#B_{ik} + \#B_{ii}} = \frac{(1-\theta_{ij})}{\sum_{k\ne i}(1-\theta_{ik}) + 1} \qquad (6)$$

Otherwise,

$$M_{ii} = 1 - \sum_{k\ne i} M_{ik} = \frac{1}{\sum_{k\ne i}(1-\theta_{ik}) + 1} \qquad (7)$$

Based on the above formulas, we can easily modify MC1 to GR-MC1 as shown in Table 1. The modification of MC4 to GR-MC4 can also be obtained according to the similar manner.

Table 1. Steps of GR-MC1

Input: T, τX^*, ε	
Output: σ	
Steps:	
a)	Calculate θ_{ij} using Eq.(2) ($i, j = 1,\ldots, \lvert T\rvert$);
b)	Calculate M_{ij} using Eq.(6) and Eq.(7) ($i, j = 1,\ldots, \lvert T\rvert$);
c)	Perturb M into a new matrix M' (= M-εI+(ε/rank(M))1*1^T);
d)	Calculate the left principal eigenvector of M'.
e)	Order the groups according to their corresponding values in the left principal eigenvector.
f)	Output will be the rank of the groups (σ).

5.3 Remark on the solutions

GR-LCM and GR-BC theoretically explain the average-score/rank approach which conversely shows the rationality of the proposed inconsistency index and the objective function. There are other numerous rank aggregation methods such as median rank [5], which can also be used to address GR. Considering that this study is mainly concerned the formal description of GR, we do not attempt to list all the available methods but leave them for future research instead. The computational complexities of the direct methods are $O(|X|)$ in calculating the weights and $O(TlogT)$ in ordering the weights. For optimization methods, the complexity of calculating scores and creating the state transition matrix is $O(|X|^2)$, while the complexity of calculating the left principal eigenvector is, for example, $O(|T|^3)$ when QR decomposition is used. The whole complexity for MC-GR1 and MC-GR4 is $O(|X|^2) + O(|T|^3)$.

6. EXPERIMENTS

This section reports our experimental results in the visual diversification for image retrieval results. Four algorithms were compared: GR-MC1, GR-MC4, GR-BC, and the conventional algorithm that is based on the Instance Counts in each Cluster (referred to as the ICC algorithm). In GR-MC2 and GR-MC4, both their permutation parameters (ε) were set to 0.01.

We posted twenty image queries to the Yahoo! image search engine and downloaded top-50 and top-100 images of the returned results from the search engine for each query. We only report the results on top-50 image collections due to lack of space. To establish the ground truth, three users were invited to manually cluster the downloaded images and then order the image clusters. In our experiments, the image features (including color, texture, edges) were extracted according to the methods described in [1]. Three image clustering algorithms proposed in [1] were then applied: Folding, Maxmin, and Reciprocal election. Finally, the clustered results were ordered using our proposed GR algorithms as well as the instance counts based cluster ranking method (ICC). The *Kendall's* Tau-b [6] value was used to evaluate an ordered clustering with respect to three ordered clustering given by users.

We apply the three clustering algorithms (i.e., Folding, Maxmin, and Reciprocal election) to cluster the top-50 images for each query. The average numbers of clusters for each query obtained from the three clustering algorithms were 11, 13, and 19, respectively. We then calculated the average mutual information of the clusters achieved by the three algorithms. The average values were 0.9084, 1.3693, and 0.9118, respectively. These results show that the clusterings obtained using the Maxmin algorithm are better than those of the two other algorithms.

The four GR algorithms (GR-MC1, GR-MC2, GR-BC, and ICC) were then used to rank the image clusters obtained by the three clustering algorithms. The *Kendall*'s Tau-b values for each combination of the three clustering algorithms and four GR algorithms (totally 12 combinations) are listed in Table 2. When the GR algorithms were compared, results showed that GR-MC1 achieved the best overall performances among the four GR algorithms. The combination of GR-MC4 and Maxmin obtained the highest *Kendall*'s Tau-b value (0.3632). GR-BC achieved the worst results partially because many lower ranked images affected the order of the image group. ICC also achieved good results. The underlying reason is that in most queries, the top-1 image clusters returned by users had the largest numbers of images.

Given that current image search engines index images are based on surrounding texts, some irrelevant images are also returned and may have higher ranks. We observed that many irrelevant images usually are like outliers. They locate the clusters that contain only one or two images. Hence, if at least two clustering algorithms take one or two specific images as a cluster (2/50 < 5%), these images are filtered out before performing the four ranking algorithms. The experimental results are shown in Table 3. Results show that most combinations behave better after discarding irrelevant images. The best result was 0.3707, which was achieved by the combination of GR-MC4 and Maxmin.

Table 2. The *Kendall*'s Tau-b values of different combinations

	Folding	Maxmin	Reciprocal election
GR-MC1	0.2165	0.3491	0.2918
GR-MC4	0.2068	0.3632	0.2764
GR-BC	-0.1384	-0.2162	-0.0155
ICC	0.1922	0.2898	0.1755

Table 3. The *Kendall*'s Tau-b values of different combinations after filtering irrelevant images

	Folding	Maxmin	Reciprocal election
GR-MC1	0.2434	0.3337	0.3044
GR-MC4	0.1586	0.3707	0.2918
GR-BC	0.0152	0.0987	0.1127
ICC	0.2024	0.3179	0.1813

7. CONCLUSIONS

This paper proposed a new ranking problem, i.e., ranking groups of instances. We have set up a formal description; in addition, a criterion named round robin comparison and an index named inconsistency index are defined respectively. Using the criterion and the index, we evaluate any candidate rank list on a given set of groups. We have proved that the optimal function of the proposed problem equals to a rank aggregation problem. As a consequence, we extended several well known rank aggregation algorithms into four new solutions to our ranking problem. The conducted experiments on visual diversification in image retrieval suggest that the proposed solutions are useful in the real applications. Our future work will focus on: 1) collecting large image data sets to study the performance of the proposed approach and 2) employing supervised information such as [7] to achieve better results.

8. ACKNOWLEDGMENT

We would like to thank the anonymous reviewers for their useful comments and suggestions. This work is partly supported by NSFC (Grant No. 60825204, 60672040) and the Excellent SKL Project of NSFC (No. 60723005).

9. REFERENCES

[1] Leuken, R., Garcia, L. and Olivares, X., Visual diversification of image search results. In Proc. WWW, pages 341–350, 2009.

[2] Dwork, C., Kumar, R., Naor, M. and Sivakumar, D., Rank aggregation methods for the Web, In Proc. WWW, 2001, pp. 613-622.

[3] Lee, J. H., Analyses of multiple evidence combination. In Proc. SIGIR, pp. 267-276, 1997.

[4] Dekel, O., Manning, C. and Singer, Y., Log-Linear models for label ranking. In Proc. NIPS, 2003.

[5] Fagin, R., Kumar, R. and Sivakumar, D., Efficient similarity search and classification via rank aggregation, In Proc.SIGMOD, pp. 301-312, 2003.

[6] Pedro, J. S. and Siersdorfer, S., Ranking and classifying attractiveness of photos in folksonomies. In Proc. WWW, pp. 771–780, 2009.

[7] Chen, S., Wang, F., Song, Y., and Zhang, C., Semi-supervised ranking aggregation. In Proc. CIKM, pp. 1427-1428, 2008.

[8] http://en.wikipedia.org/wiki/Borda_count