

# A New Hierarchical Ranking Aggregation Method

Jiankun Ding, Deqiang Han  
Institute of Integrated Automation  
School of Electronic and Information Engineering  
Xi'an Jiaotong University  
Xi'an, Shaanxi, China 710049  
Email: d4574b@163.com; deqhan@gmail.com

Jean Dezert  
ONERA  
The French Aerospace Lab  
Chemin de la Hunière,  
F-91761 Palaiseau, France  
Email: jean.dezert@onera.fr

Yi Yang  
SKLSVMS  
School of Aerospace  
Xi'an Jiaotong University  
Xi'an, Shaanxi, China 710049  
Email: jiafeiyy@mail.xjtu.edu.cn

**Abstract**—The ranking fusion (or aggregation), which is an important branch in multiple attribute decision making, combines multiple rankings to a single one for decision making. Many traditional ranking fusion methods are implemented through heuristic ways to reduce the computational cost. They all have their own pros and cons. In this paper, a new hierarchical ranking aggregation method is proposed. All the items are first divided into multiple ranking levels (i.e., ordered items subsets) based on the information provided by different rankings to be fused. The items in high ranking levels are ranked higher than all the items in low ranking levels in the fused ranking, therefore those items in different levels never interact each other's ranking. Then, the items in the same ranking level are further divided into multiple sub-levels if possible. In the final, the items in a sub-level which cannot be further divided are further compared and ranked in particular. Simulation results show that our new hierarchical method performs well in ranking fusion for decision making.

**Index Terms**—Ranking fusion, hierarchical, pairwise comparison, multiple attribute decision making (MADM).

## I. INTRODUCTION

Different attributes or information sources reflect different aspects of the items concerned in decision making. Information fusion based decision making approaches such as multiple attribute decision making (MADM) [1], which comprehensively uses multiple attributes, are expected to obtain better results than those using only single attribute. The ranking fusion (or ranking aggregation) is an important branch in MADM where each information source provides a order list (ranking) of the corresponding items and these input rankings are fused to generate a consensus ranking. The ranking fusion has been a research topic for more than two centuries and has been increasingly used in a wide range of successful applications such as the information retrieval [2], bio-informatics [3], estimator ranking [4] and enterprise management [5], [6].

In the traditional MADM framework, the goal of ranking fusion is to find a ranking that can best represent all the input rankings, i.e., to find a ranking which has the smallest average distance to all the available rankings to be fused [2], [7]. Finding this fused ranking can be formulated as an optimization problem, which is usually NP-hard [2]. To reduce the computational cost, a series of heuristic ranking fusion methods have been proposed. According to their implementation processes, these methods can be divided into two branches, i.e., *batch mode* [8]–[12] (or *global mode*)

methods and *instant-runoff mode* [15] (or *Luce mode* [16], [17]) methods.

Batch mode methods generate the ranking positions of all the items simultaneously, where the Borda count [7], [8] is perhaps the most representative one. It works based on the items' positions in input rankings directly. Condorcet vote [9] is another popular traditional method, which works based on the pairwise comparisons of items. Dwork [11] used the Locally Kemeny optimal ranking as the fused ranking. It is implemented by finding Hamiltonian path. In 2002, Mark [10] designed a Condorcet-fuse method and it is also implemented by finding the Hamiltonian path. In 2012, Negahban [12] introduced a kind of iterative ranking from pairwise comparisons, and later Yin [4] proposed the Ranking Eigenvector (REV) method. Although these two methods analyze the ranking problem in different ways, they both use the eigenvector of the pairwise comparison matrix for ranking. Yin also introduced a Self-weighted Score Addition (SSA<sup>2</sup>) method. Moreover, some other methods apply stochastic optimization algorithms [13], [14] such as the genetic algorithm and cross-entropy Monte Carlo algorithm for searching the optimal ranking.

The instant-runoff mode methods decompose the process of generating a fused ranking of  $n$  items into  $n$  sequential rounds. In each round, only the one item is assigned a fused ranking position and eliminated immediately from the set of unranked items. The following rounds only concern about the unranked items. The most representative method is Instant-Runoff Voting (IRV) [15]. However, in each round of IRV, the eliminated items is selected just according to the first place votes, which can not represent the input rankings' opinions comprehensively. In 2010, Qin [17] proposed a probabilistic model, i.e., coset-permutation distance based stagewise (CPS) model, for the ranking aggregation. CPS works better than IRV, however, it has large computational cost for selecting the best item in each round. In 2015, we [18] proposed an instant-runoff ranking fusion method (IRRF) using the result of traditional batch mode ranking fusion methods and a top-2 comparison based instant-runoff ranking fusion method (T2-IRRF), which is an improved IRRF by introducing more local comparison information into the selection of the best item in each round.

Traditional ranking fusion methods all have their pros and cons (e.g., the problem of rank reversal [19]), and there is

no widely accepted one. In this paper, a hierarchical ranking aggregation (HRA) method is proposed to obtain better performances of ranking fusion. HRA divides the items into different ranking levels, i.e., multiple ordered items subsets, according to their comparisons with all the other items based on all the rankings to be fused. Then the items in the same ranking level are further divided into different sub-levels according to the comparisons among themselves if they could be further divided. The ranking fusion is implemented hierarchically. In the final, the items in a sub-level, which cannot be further divided, are compared and ranked in particular to complete the fused ranking. According to HRA, the items in high ranking levels are ranked higher than all the items in low ranking levels in the fused ranking, therefore, those items in different levels never interact each other's ranking. Simulation results show that the new ranking fusion method can effectively improve the performance of ranking fusion, which is desired for decision making.

## II. RANKING AGGREGATION IN MULTIPLE ATTRIBUTE DECISION MAKING

### A. Conceptions and formulations in ranking aggregation

Given a set of items  $X = \{x_1, x_2, \dots, x_n\}$ , a ranking is the order list of these items according to a certain criterion (or attribute). The ranking in general can be categorized into three types [7], [20], [21]: the total ranking, the partial ranking, and the top- $k$  ranking. Here, we only focus on the total ranking, where all the items are assigned ranking positions. When there are more than one available rankings originated from different information sources (or attributes), the ranking fusion (or aggregation) can output a consensus ranking of these available rankings. Suppose there are  $m$  available rankings with respect to  $n$  items as shown in (1).

$$E = \begin{matrix} & x_1 & x_2 & \cdots & x_n \\ A_1 & \tau_{11} & \tau_{12} & \cdots & \tau_{1n} \\ A_2 & \tau_{21} & \tau_{22} & \cdots & \tau_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ A_m & \tau_{m1} & \tau_{m2} & \cdots & \tau_{mn} \end{matrix} \quad (1)$$

The  $i$ th row of  $E$  represents the ranking  $\tau^i = \{\tau_{i1}, \tau_{i2}, \dots, \tau_{in}\}$  given by attribute  $A_i$  ( $i = 1, 2, \dots, m$ ) and  $\tau_{ij}$  denotes the ranking position of  $x_j$  in  $\tau^i$ . In general, we assume that if the value of  $\tau_{ij}$  is smaller, the corresponding item  $x_j$  is more preferred according to  $\tau^i$ .

Each ranking in  $E$  can also be described using pairwise representation [22], i.e., the preference between each pair of items is represented using a matrix  $R = (r_{ij})_{n \times n}$ , where

$$r_{ij} = \begin{cases} 1 & \text{if } x_i \succ x_j, \\ 0 & \text{if } x_i \prec x_j, \\ 0.5 & \text{if } x_i = x_j \end{cases} \quad (2)$$

Here,  $x_i \succ x_j$  denotes that  $x_i$  is better than  $x_j$ , and  $x_i \prec x_j$  denotes the opposite situation. The ranking, e.g.,  $x_4 \succ x_1 \succ x_2 \succ x_3$  can be represented with  $R$ :

$$R = \begin{matrix} & x_1 & x_2 & x_3 & x_4 \\ x_1 & 0 & 1 & 1 & 0 \\ x_2 & 0 & 0 & 1 & 0 \\ x_3 & 0 & 0 & 0 & 0 \\ x_4 & 1 & 1 & 1 & 0 \end{matrix}$$

To make a decision, the multiple rankings should be fused or aggregated to generate a consensus ranking that best represents all the rankings:

$$\tau^* = H(\tau^1, \tau^2, \dots, \tau^m) \quad (3)$$

where  $\tau^*$  is the fused ranking,  $H(\cdot)$  denotes a kind of ranking fusion method, and  $\tau^1, \tau^2, \dots, \tau^m$  are the rankings provided by different attributes or information sources, e.g., the rankings in  $E$ . Traditional ranking fusion methods can mainly be divided into two branches, the batch mode method and the instant-runoff mode method. Batch mode (i.e. global mode) methods obtain the fused ranking in one run, e.g., Borda Count and the Ranking Eigenvector method (REV). The instant-runoff mode methods obtain the fused ranking using a series of sequential rounds, e.g., Instant-Runoff Voting (IRV).

### B. Batch mode methods

1) *Borda Count*: It works based on the ranking positions of items directly. It assigns each item  $x_j$  ( $j = 1, 2, \dots, n$ ) a score  $B^i(j)$  based on the position provided by  $\tau^i$  ( $i = 1, 2, \dots, m$ ).

$$B^i(j) = 1 - \frac{\tau_{ij} - 1}{n} \quad (4)$$

According to the scores of  $x_j$  generated from all the available rankings  $\tau^i$  ( $i = 1, 2, \dots, m$ ), Borda Count generates a final score for  $x_j$  using a certain aggregate function  $B_j = f(B^1(j), B^2(j), \dots, B^m(j))$ . For example, the mean aggregate function is

$$f(B^1(j), B^2(j), \dots, B^m(j)) = \frac{1}{m} \sum_{l=1}^m B^l(j) \quad (5)$$

There are also many other ways to construct aggregate function  $f$ , such as the  $p$ -norm mapping, median mapping and geometric mean mapping, etc. [7]. The fused ranking is obtained by sorting the aggregated scores of items.

2) *Ranking Eigenvector (REV)*: It works based on the pairwise comparison matrix (PCM). PCM's elements express the average preference between each items pair according to their comparisons in all the available rankings. In REV, all the available rankings  $\tau^i$  ( $i = 1, 2, \dots, m$ ) are reformulated using pairwise representation  $R^i = (r_{jk}^i)_{n \times n}$ . The PCM matrix is generated as:

$$\mathcal{M} \triangleq \begin{bmatrix} M(1,1) & M(1,2) & \cdots & M(1,n) \\ M(2,1) & M(2,2) & \cdots & M(2,n) \\ \vdots & \vdots & \ddots & \vdots \\ M(n,1) & M(n,2) & \cdots & M(n,n) \end{bmatrix} \quad (6)$$

where

$$M(j, k) = \frac{1}{m} \sum_{i=1}^m r_{jk}^i \quad (7)$$

$M(j, k)$  denotes the average preference of  $x_j$  over  $x_k$ . If  $M(j, k) > M(k, j)$ , then  $x_j \succ x_k$ . There exists

$$\begin{cases} M(j, k) + M(k, j) = 1 & \forall j \neq k \\ M(j, k) = 0 & \forall j = k \end{cases} \quad (8)$$

Based on the PCM matrix, REV method generates scores for different items using an certain mapping  $F(\cdot)$ . For example, the simple linear mapping is as follows.

$$\lambda_{\max} v_{REV} = F(v_{REV}) = \mathcal{M} v_{REV} \quad (9)$$

where  $\mathcal{M}$  is the PCM matrix,  $\lambda_{\max}$  is the largest eigenvalue of  $\mathcal{M}$ , and  $v_{REV}$  is the ranking eigenvector (REV), which is the positive eigenvector<sup>1</sup> corresponding to the largest eigenvalue  $\lambda_{\max}$ .  $v_{REV}$  is defined as

$$v_{REV} = [v_1 \quad v_2 \quad \dots \quad v_n] \quad (10)$$

where  $v_i$  ( $i = 1, 2, \dots, n$ ) is positive and stands for the fusion score of item  $x_i$ . The fused ranking is generated by sorting the scores in  $v_{REV}$ .

### C. Instant-runoff mode methods

Both two methods introduced above determine the positions of all items in the fused ranking simultaneously, while instant-runoff mode methods determine the items' ranking positions one by one. Instant-Runoff Voting is the most representative instant-runoff mode method. In each round, IRV only determines the ranking position of the item with the fewest (or the most) votes. The item whose ranking is determined in current round will be eliminated. The following rounds are only for determining the rankings of remaining items. The process of IRV is as follows:

i) By counting first place votes according to the rankings with respect to the remaining items we get

$$V(x_j) = \sum_{i=1}^m \delta_i(x_j) \quad (11)$$

where

$$\delta_i(x_j) = \begin{cases} 1 & x_j \succ x_k, \forall k \neq j \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

$\delta_i(x_j)$  indicates the first place vote according to  $\tau^i$ .  $V(x_j)$  is the count of votes for item  $x_j$ . The item with the fewest (or the most) votes will be assigned the lowest (or the highest) ranking position.

ii) The item who has been ranked in step i) is put into the fused ranking and eliminated from all the input rankings. These updated input rankings will be used as input rankings in the next round.

iii) Repeat the above two steps until we have obtain the whole fused ranking.

<sup>1</sup>The way of taking this eigenvector as ranking vector is also adopted in many other ranking fusion methods, i.e., AHP [6].

### D. Average ranking distance criterion

As aforementioned, the purpose of the ranking aggregation is to find a ranking that can most represent all of the rankings available for fusion. Kemeny optimal aggregation (KOA) [2] aims to find a ranking with the smallest average distance between the input rankings and the fused one, which is defined as

$$\tau^* = \arg \min_{\tau} \frac{1}{m} \sum_{i=1}^m K(\tau, \tau^i) \quad (13)$$

where  $K(\cdot, \cdot)$  is the Kendall distance [23], which is a measurement for describing the consensus degree between rankings. The Kendall distance compares rankings by counting the pair-wise disagreements between two rankings, which is defined as follows:

$$K(\tau_1, \tau_2) = \frac{1}{C_n^2} \sum_{(i,j), i \neq j} K_{ij}^*(\tau_1, \tau_2) \quad (14)$$

where

$$K_{ij}^*(\tau_1, \tau_2) = \begin{cases} 0 & \text{if } x_i, x_j \text{ are in the} \\ & \text{same order in } \tau_1, \tau_2 \\ 1 & \text{if } x_i, x_j \text{ are in the} \\ & \text{inverse order in } \tau_1, \tau_2 \end{cases} \quad (15)$$

Kendall distance is calculated in time  $O(n^2)$ .

Finding the Kemeny optimal aggregation is usually NP-hard, which is very difficult especially when the number of items is very large. The most direct way to find this optimal fused ranking is traversing all of the possible rankings. If the number of the items to be ranked is  $n$ , there will be  $n!$  (the factorial of  $n$ ) rankings to be traversed. Traditional ranking fusion methods find the feasible fused rankings, among which the one with a smaller average ranking distance is preferred. Therefore, the average ranking distance criterion (ARDC) is often used to evaluate the fused ranking, which is defined as

$$\varepsilon(\tau) = \frac{1}{m C_n^2} \sum_{j=1}^m K(\tau, \tau^j) \quad (16)$$

If the average ranking distance  $\varepsilon$  is small, the corresponding fused ranking is good, i.e., the fused ranking has a higher consensus degree to the input rankings. In this paper, ARDC is adopted to evaluate the fused ranking.

Traditional ranking fusion methods introduced above have their own pros and cons. There is no well accepted ranking fusion approach till now.

## III. NOVEL HIERARCHICAL RANKING AGGREGATION

Here, a novel hierarchical ranking aggregation (HRA) approach is proposed. The hierarchical fusion process of HRA is illustrated using a simple example in Figure 1. HRA first divides the items into multiple ranking levels, e.g., Level<sub>2</sub>, Level<sub>1</sub> and Level<sub>0</sub><sup>2</sup>, according to the comparisons between each item with all the other items. All the items in high ranking levels are ranked higher than the items in low ranking levels.

<sup>2</sup>The ranking level with larger subscript number is a higher ranking level.

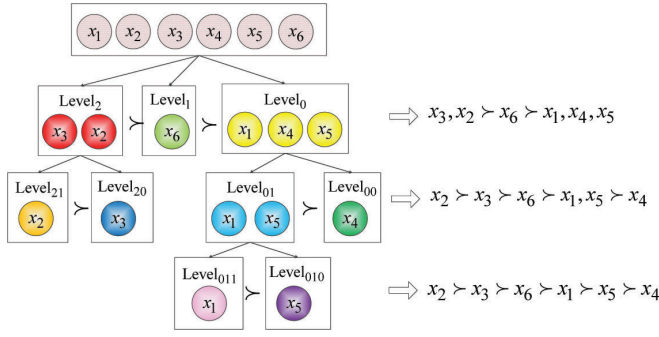


Fig. 1. Hierarchical fusion process of HRA.

The items in the same ranking level are further compared and divided into multiple ranking sub-levels, e.g.,  $x_2, x_3$  in  $\text{Level}_2$  are further compared and divided into  $\text{Level}_{21}$  and  $\text{Level}_{20}$ . HRA refines the fused ranking hierarchically. In the following, we introduce the way for dividing the items into different ranking levels in detail.

Assume that  $x_1, x_2, \dots, x_n$  are the items to be ranked and there are  $m$  rankings with respect to these  $n$  items available for fusion. HRA assigns each item a certain ranking level according to its comparisons with all the other items. All the comparisons of each item with all the other items can construct a matrix, i.e., the pairwise comparison matrix  $\mathcal{M}$  which has already been introduced in (6) as.

$$\mathcal{M} \triangleq \begin{bmatrix} M(1,1) & M(1,2) & \cdots & M(1,n) \\ M(2,1) & M(2,2) & \cdots & M(2,n) \\ \vdots & \vdots & \ddots & \vdots \\ M(n,1) & M(n,2) & \cdots & M(n,n) \end{bmatrix}$$

As aforementioned,  $M(j, k)$  denotes the average preference of  $x_j$  over  $x_k$  on the opinions of all input rankings (attributes).  $x_j$  is preferred than  $x_k$  if and only if  $M(j, k) > M(k, j)$ .

Then HRA generates the preference order between each items pair according to the PCM matrix and obtains the pairwise comparison decision (PCD) matrix, as

$$\mathcal{D} \triangleq \begin{bmatrix} D(1,1) & D(1,2) & \cdots & D(1,n) \\ D(2,1) & D(2,2) & \cdots & D(2,n) \\ \vdots & \vdots & \ddots & \vdots \\ D(n,1) & D(n,2) & \cdots & D(n,n) \end{bmatrix} \quad (17)$$

where

$$D(i, j) = \begin{cases} 1 & M(i, j) > M(j, i) \\ 0.5 & M(i, j) = M(j, i), \text{ and } i \neq j \\ 0 & M(i, j) < M(j, i), \text{ or } i = j \end{cases} \quad (18)$$

The  $i$ th row of  $\mathcal{D}$  denotes the comparisons results between  $x_i$  and all the other items.

According to  $i$ th row of  $\mathcal{D}$ ,  $x_i$  is assigned the ranking level by counting the times of  $x_i$  winning the pairwise comparisons.

$$L(i) = \sum_{j=1}^n D(i, j) \quad (19)$$

where  $L(i)$  is defined as the ranking level of  $x_i$ .  $x_i$  and  $x_j$  are in the same ranking level if  $L(i) = L(j)$ , and  $x_i$  locates in a higher ranking level than  $x_j$  if  $L(i) > L(j)$ . In the fused ranking, all the high ranking level items are ranked higher than the low ranking level items.

For the items in the same ranking level, HRA further compares their preference order and divides them into multiple ranking sub-levels. The ranking level (or sub-level), which contains more than one items but can not be divided any more, is defined as stationary level. The same ranking level items will stop being divided when this level only contains one item or it is a stationary level. For the items in a stationary level, HRA uses some traditional ranking fusion method, e.g., Borda count as an auxiliary method to generate the fused ranking of those items. The pseudo code of HRA is as Algorithm 1.

---



---

**Algorithm 1:** Hierarchical Ranking Aggregation

---



---

**Input:** Rankings to be fused,  $\{\tau^j\}_{j=1}^m$ ;

fusedRanking = **HierarchicalRank** ( $\{\tau^j\}_{j=1}^m$ )

{

$n$  = number of items in  $\{\tau^j\}_{j=1}^m$

if  $n=1$

return

endif

$\mathcal{M}$  = calculate PCM matrix  $\{\tau^j\}_{j=1}^m$

$\mathcal{D}$  = calculate PCD matrix ( $\mathcal{M}$ )

$L(i)$ , ( $i = 1, 2, \dots, n$ ) = calculate each item's ranking level ( $\mathcal{D}$ )

if  $L(1) = L(2) = \dots = L(n)$

fusedRanking = traditional ranking fused method ( $\{\tau^j\}_{j=1}^m$ )

return fusedRanking

endif

for  $k = 1$ : number of different levels

ranking of  $k$ th level = **HierarchicalRank** (input rankings with respect to the items in  $k$ th level )

endfor

fusedRanking = Rank the items according to their levels

return fusedRanking

}

---



---

The computational cost of HRA mainly relates to the comparison times between the same level items. The best case is that all the items are divided to different ranking levels in one time, where HRA will compute in time  $O(n^2)$  comparing each pair of the  $n$  items. The worst case is that, in each ranking level (or sub-level) division, the items in the same ranking level (or sub-level) are divided into two sub-levels, one of which only contains one item. In this case, HRA will perform  $n - 1$  times ranking sub-level divisions for generating the complete fused ranking. Therefore, HRA will compute in time  $O(n^3)$  in the worst case.

Here, we give an example to illustrate our new ranking aggregation approach.

**Example 1:** Assume that  $\tau^1, \tau^2, \tau^3, \tau^4, \tau^5$  are five rankings originated from five attributes  $A_1, A_2, A_3, A_4, A_5$  with respect to nine items  $x_1, x_2, \dots, x_9$  as

$$E = \begin{bmatrix} 4 & 1 & 9 & 8 & 3 & 6 & 5 & 7 & 2 \\ 3 & 1 & 2 & 8 & 9 & 6 & 7 & 5 & 4 \\ 9 & 3 & 5 & 6 & 1 & 2 & 8 & 7 & 4 \\ 5 & 4 & 2 & 7 & 9 & 3 & 1 & 8 & 6 \\ 2 & 6 & 7 & 1 & 9 & 4 & 8 & 5 & 3 \end{bmatrix}$$

where the  $i$ th row of  $E$  represents ranking  $\tau^i$  obtained based on  $A_i$  and its column represents the ranking position of corresponding item in each ranking. Here we show the fusion process of HRA.

HRA first generates the pairwise comparison matrix  $\mathcal{M}_1$ , according to  $E$ . For example, considering the comparison between  $x_1$  and  $x_2$ ,  $x_1$  loses the comparison according to  $\tau^1, \tau^2, \tau^3, \tau^4$  and wins the comparison according to  $\tau^5$ . According to (2) and (7), we have  $M(1,2) = 0.2$ .

$$\mathcal{M}_1 = \begin{bmatrix} 0.0 & 0.2 & 0.4 & 0.6 & 0.6 & 0.6 & 0.6 & 0.8 & 0.6 \\ 0.8 & 0.0 & 0.8 & 0.8 & 0.8 & 0.4 & 0.8 & 0.8 & 0.8 \\ 0.6 & 0.2 & 0.0 & 0.6 & 0.6 & 0.4 & 0.6 & 0.6 & 0.4 \\ 0.4 & 0.2 & 0.4 & 0.0 & 0.6 & 0.2 & 0.4 & 0.6 & 0.2 \\ 0.4 & 0.2 & 0.4 & 0.4 & 0.0 & 0.4 & 0.4 & 0.4 & 0.2 \\ 0.4 & 0.6 & 0.6 & 0.8 & 0.6 & 0.0 & 0.6 & 0.8 & 0.4 \\ 0.4 & 0.2 & 0.4 & 0.6 & 0.6 & 0.4 & 0.0 & 0.4 & 0.2 \\ 0.2 & 0.2 & 0.4 & 0.4 & 0.6 & 0.2 & 0.6 & 0.0 & 0.0 \\ 0.4 & 0.2 & 0.6 & 0.8 & 0.8 & 0.6 & 0.8 & 1.0 & 0.0 \end{bmatrix}$$

Then HRA generates the preference order between each pair of items according to  $\mathcal{M}_1$  and obtains the pairwise comparison decision matrix  $\mathcal{D}_1$ .

$$\mathcal{D}_1 = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \end{bmatrix}$$

Then different items are assigned ranking levels by counting their winning times according to  $\mathcal{D}_1$ .

$$L_1 = [ 7 \ 8 \ 6 \ 3 \ 1 \ 7 \ 3 \ 3 \ 7 ]$$

where  $L_1(i)$  denotes the ranking level of  $x_i$ . Since  $L_1(2) = 8$  is the biggest, then  $x_2$  is in the highest ranking level. Similarly,  $x_5$  is in the lowest ranking level.  $x_1, x_6, x_9$  are in the same ranking level, and  $x_4, x_7, x_8$  are also in the same ranking level. Then we have

$$x_2 \succ x_1, x_6, x_9 \succ x_3 \succ x_4, x_7, x_8 \succ x_5$$

where the preference order among  $x_1, x_6, x_9$  and the preference order among  $x_4, x_7, x_8$  has not been determined, as they are in the same ranking level, i.e., they are close to each other. They will be further compared and divided into multiple ranking levels.

For  $x_1, x_6, x_9$ , the pairwise comparison matrix with respect to these three items is  $\mathcal{M}_2$ .

$$\mathcal{M}_2 = \begin{bmatrix} 0.0 & 0.6 & 0.6 \\ 0.4 & 0.0 & 0.4 \\ 0.4 & 0.6 & 0.0 \end{bmatrix}$$

According to  $\mathcal{M}_2$  the preference order between each pair of items is generated and the pairwise comparison decision matrix  $\mathcal{D}_2$  is as

$$\mathcal{D}_2 = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

So we have

$$L_2 = [ 2 \ 0 \ 1 ]$$

According to  $L_2$  we have

$$x_1 \succ x_9 \succ x_6$$

Then this branch will stop and return the ranking with respect to  $x_1, x_6, x_9$ , as each ranking level only has one item according to  $L_2$ .

For  $x_4, x_7, x_8$ , the pairwise comparison matrix with respect to these three items is  $\mathcal{M}_3$ .

$$\mathcal{M}_3 = \begin{bmatrix} 0.0 & 0.4 & 0.6 \\ 0.6 & 0.0 & 0.4 \\ 0.4 & 0.6 & 0.0 \end{bmatrix}$$

According to  $\mathcal{M}_3$ , the pairwise comparison decision matrix  $\mathcal{D}_3$  is generated, as

$$\mathcal{D}_3 = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

So we have

$$L_3 = [ 1 \ 1 \ 1 ]$$

According to  $L_3$ ,  $x_4, x_7, x_8$  can not be divided into different ranking levels any more, i.e., they are in an stationary level. Their preference order will be determined according to some traditional ranking fusion method. In this example, we use Borda count method. According to Borda count we have

$$x_7 \succ x_4 \succ x_8$$

Then this branch will stop and return the ranking with respect to  $x_4, x_7, x_8$ , as these three items are in an stationary level according to  $L_3$ .

Till now we have obtain the order list with respect to all the items, i.e.,

$$x_2 \succ x_1 \succ x_9 \succ x_6 \succ x_3 \succ x_7 \succ x_4 \succ x_8 \succ x_5$$

thus the final fused ranking is

$$\tau^f = [ 2 \ 1 \ 5 \ 7 \ 9 \ 4 \ 6 \ 8 \ 3 ]$$

The average ranking distance of  $\tau^f$  together with the results of traditional ranking fusion methods are listed in TABLE I, where HRA is our new method and KOA is the Kemeny optimal aggregation found by traversing all of the possible rankings. According to TABLE I, in this example, the result of HRA is better than traditional ranking fusion methods based on ARDC, and the fused ranking of HRA is the same with Kemeny optimal aggregation.

Note that the stationary level appears only when all the items can not be divided into multiple levels (or sub-levels)

TABLE I  
RESULTS OF DIFFERENT RANKING FUSION METHODS.

Method	Ranking	Average ranking distance
Borda	4, 1, 5, 7, 8, 3, 6, 9, 2	0.3500
REV	4, 1, 5, 7, 8, 3, 6, 9, 2	0.3500
SSA <sup>2</sup>	4, 1, 5, 7, 8, 3, 6, 9, 2	0.3500
IRV	3, 1, 2, 8, 5, 7, 6, 9, 4	0.3667
CPS	3, 2, 5, 8, 7, 1, 6, 9, 4	0.3500
HRA	2, 1, 5, 7, 9, 4, 6, 8, 3	0.3333
KOA	2, 1, 5, 7, 9, 4, 6, 8, 3	0.3333

any more, i.e., all the items still locate in the same ranking level according to the comparisons among themselves, such as  $x_4$ ,  $x_7$ ,  $x_8$ . In this case, every item wins half of the other items in this stationary level, i.e., these items are really close to each other. Although HRA uses Borda count to deal with the stationary level items as a auxiliary method, using HRA is better than only using Borda count to determine the whole fused ranking.

#### IV. SIMULATIONS

To verify the validity of the novel hierarchical ranking aggregation method, in this section, we use Monte-Carlo simulations with different settings. The rankings to be fused are randomly generated with different parameter settings in different simulations. Then different ranking fusion methods are used to fuse these rankings generated. The performance of the ranking fusion methods used are compared.

The parameters involved in the generations of simulation settings are given in TABLE II. In each run, a seed ranking

TABLE II  
PARAMETERS INVOLVED IN SIMULATION SETTINGS.

Parameters	Meanings
$m$	number of rankings to be fused
$I$	number of items
$T$	swapping times

$\tau_{seed}$  with respect to  $I$  items is randomly generated. Then, randomly select two items in  $\tau_{seed}$  and swap their positions. Such a random selection and swapping will be repeat  $T$  times to generate a ranking to be fused. The swapping times  $T$  can be different for different single rankings. The rankings generated with small  $T$  are similar to each other, as they are all similar to the seed ranking. The ranking generated with large  $T$  are quite different from all the other rankings generated. In each run,  $m$  rankings are generated.

In the simulations, Borda Count, REV, SSA<sup>2</sup>, IRV, CPS and HRA are used to fuse the input rankings. The fused rankings generated by different methods are evaluated according to average ranking distance criterion (ARDC). Totally 100 runs are executed, where in each run a seed ranking is randomly regenerated. The performance of different methods used are evaluated using the average of the 100 runs. When the number

of items is small, the Kemeny optimal aggregation is found by traversing all of the possible rankings.

**Simulation 1:** This simulation compares the performance of different methods in the cases with small number of items and input rankings. We design three simulation settings with different similarity degrees. The simulation settings are shown in TABLE III. Note that there are in total three rankings to

TABLE III  
SIMULATION SETTINGS IN SIMULATION 1.

settings	$m$ (Rankings NO.)	$I$ (Items NO.)	$T$ (Swapping times)
1	3	9	2
2	2	9	2
	1	9	6
3	3	9	6

be fused in setting 2, one is generated by swapping 6 times and the other two are generated by swapping 2 times. All the three rankings in setting 1 are similar to each other. In setting 2, two rankings are similar to each other but quite different from the other one. Setting 3 contains three quite different rankings. The average ranking distance of different methods are listed in TABLE IV.

TABLE IV  
SIMULATION RESULTS IN SETTINGS 1 ~ 3.

Methods	settings		
	1	2	3
Borda	0.2314	0.2456	0.2916
REV	0.2165	0.2479	0.2828
SSA <sup>2</sup>	0.2137	0.2469	0.2828
IRV	0.2356	0.2636	0.3045
CPS	0.2085	0.2397	0.2722
HRA	0.1966	0.2257	0.2622
KOA	0.1950	0.2233	0.2581

Because in this simulation the number of items is small, so we find the Kemeny optimal aggregation (KOA) by traversing all of the possible rankings. According to TABLE IV, the performance of HRA is better than traditional ranking fusion methods, i.e., Borda count, REV, SSA<sup>2</sup>, IRV and CPS, and its results are very close to the results of KOA.

Furthermore, TABLE IV only lists the average ranking distance of different methods in different simulation settings. It should be better to provide further detailed comparisons among the different ranking fusion methods by considering KOA as the optimal one. Here we define the difference rate  $\gamma$ , which is calculated according to (20).

$$\gamma = \frac{\varepsilon^* - \varepsilon^{KOA}}{\varepsilon^{KOA}} \quad (20)$$

where  $\varepsilon^{KOA}$  is calculated using (16) with respect to the result of KOA and  $\varepsilon^*$  is calculated with respect to the results of different ranking fusion methods. As aforementioned, if  $\varepsilon$  is small, the corresponding fused ranking is good and KOA finds the optimal fused ranking by traversing all of the possible

rankings. Thus  $\varepsilon^* \geq \varepsilon^{KOA}$ , i.e.,  $\gamma \geq 0$ , and a smaller  $\gamma$  is better, i.e., the corresponding fused ranking is closer to the optimal fused ranking. The mean, maximum and minimum of  $\gamma$  with respect to different ranking fusion methods are calculated in 100 runs and shown in Figures 2~4.

In these three figures, the histogram indicates the mean of  $\gamma$ , and the top and the bottom of the error bar (blue vertical line segment) denote the maximum and minimum of  $\gamma$  in 100 runs, respectively. According to Figures 2~4, the performance of HRA is better than the other traditional ranking fusion methods. The mean  $\gamma$  of HRA is the smallest. Especially in setting 1, the mean  $\gamma$  of HRA nearly equals to 0, i.e., in each run, the result of HRA is very close to the optimal fused ranking. Furthermore, the top of HRA's error bar is the lowest one and the length of its error bar is the shortest one, i.e., the fused ranking obtained using HRA is more reliable than the results of the other traditional ranking fusion methods.

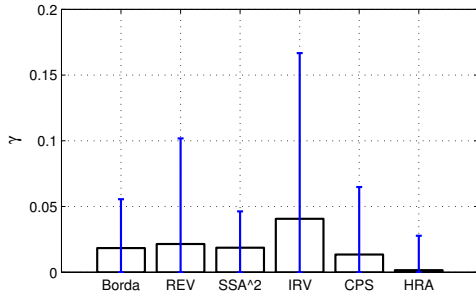


Fig. 2. Results in Setting 1.

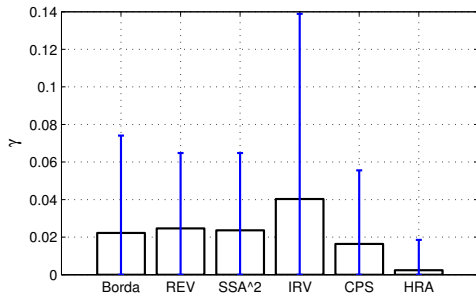


Fig. 3. Results in Setting 2.

**Simulation 2:** This simulation compares the performance of different methods in the cases with large number of items and more input rankings. We also design three simulation settings with different similarity degrees. The simulation settings are shown in TABLE V. Note that both setting 4 and setting 5 totally contain ten rankings to be fused. Most of the rankings in setting 4 are similar to each other and most of the rankings in setting 5 are quite different from the others. All the rankings in setting 6 are quite different from the others. The average ranking distance of different methods are listed in TABLE VI.

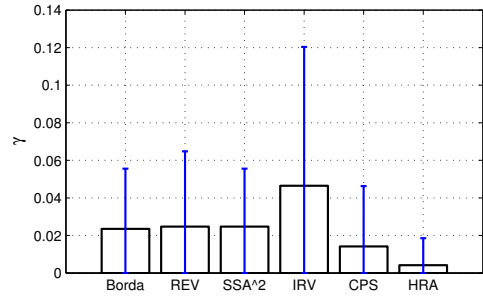


Fig. 4. Results in Setting 3.

TABLE V  
SIMULATION SETTINGS IN SIMULATION 2.

settings	$m$ (Rankings NO.)	$I$ (Items NO.)	$T$ (Swapping times)
4	7	50	5
	3	50	25
5	3	50	5
	7	50	25
6	10	50	25

According to TABLE VI, the performance of HRA is better than traditional ranking fusion methods, as the average ranking distance of HRA is smaller than traditional ranking fusion methods in all these three settings. Because in this simulation the number of items is too large to find the KOA by traversing all of the possible rankings. Here, we do not provide the result of KOA and the difference rates of different ranking fusion methods compared with KOA.

## V. CONCLUSION

A novel Hierarchical Ranking Aggregation approach is proposed in this paper. According to HRA, the ranking fusion is implemented by dividing all the items into multiple ranking levels hierarchically. The items in the same ranking level are further divided according to the pairwise comparisons between the items in this ranking level, where the preference order among the close items are further compared. In simulations no matter when the number of items and the number of input rankings are small or big, HRA performs better than traditional ranking fusion methods as its results have smaller average ranking distances. Moreover, the fused ranking of HRA has

TABLE VI  
SIMULATION RESULTS IN SETTING 4 ~ 6.

Methods	settings		
	4	5	6
Borda	0.2151	0.2927	0.3390
REV	0.2183	0.2933	0.3393
SSA <sup>2</sup>	0.2174	0.2933	0.3393
IRV	0.2461	0.3448	0.4016
CPS	0.2154	0.2923	0.3385
HRA	0.1932	0.2837	0.3358

a smaller difference rate  $\gamma$ , i.e., it is more reliable than the results of other methods.

Furthermore, evaluation criteria are crucial for ranking fusion. In future work, we will try to design better evaluation criteria for analysing the performance of HRA and the robustness of HRA to the rank reversal.

#### ACKNOWLEDGMENT

This work was supported by the Grant for State Key Program for Basic Research of China (973) (No. 2013CB329405), National Natural Science Foundation (No. 61573275), Science and technology project of Shaanxi Province (No. 2013KJXX-46), Specialized Research Fund for the Doctoral Program of Higher Education (20120201120036), and Fundamental Research Funds for the Central Universities (No. xjj2014122, xjj2016066).

#### REFERENCES

- [1] J. Figueira, S. Greco, and M. Ehrgott, *Multiple Criteria Decision Analysis: State of The Art Surveys*, New York: Wiley, 2005.
- [2] C. Dwork, R. Kumar, M. Naor, and D. Sivakumar, "Rank aggregation methods for the web," *Proceedings of 10th International Conference on World Wide Web*, New York, USA, May 2001, pp. 613–622.
- [3] S. L. Lin and J. Ding, "Integration of ranked lists via cross entropy Monte Carlo with applications to mRNA and microRNA studies," *Biometrics*, vol. 65, no. 1, pp. 9–18, Mar. 2009.
- [4] H. L. Yin, J. Lan, and X. R. Li, "Measures for ranking estimation performance based on single or multiple performance metrics," *16th International Conference on Information Fusion*, Istanbul, Turkey, July 2013, pp. 453–460.
- [5] Y. L. Chen and C. C. Li, "A novel collaborative filtering approach for recommending ranked items," *Expert Systems with Applications*, vol. 34, no. 4, pp. 2396–2405, May 2008.
- [6] G. Mitra, H. J. Greenberg, F. A. Lootsma, M. J. Rijkaert, and H. J. Zimmermann, *Mathematical Models for Decision Support*, Berlin Heidelberg: Springer-Verlag, 1988, pp. 109–121.
- [7] A. Argentini, *Ranking Aggregation Based on Belief Function Theory*. PhD Dissertation, University of Trento, 2012.
- [8] J. C. de Borda, *Mémoire sur les élections au scrutin*, *Histoire de l'Académie Royale des Sciences*. Paris, France, 1781.
- [9] M. de Condorcet, *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. L'imprimerie royale, 1785.
- [10] M. Montague and J. A. Aslam, "Condorcet fusion for improved retrieval," *Proceedings of 11th International Conference on Information and Knowledge Management*, McLean, Virginia, USA, Nov. 2002, pp. 538–548.
- [11] C. Dwork, R. Kumar, M. Naor, and D. Sivakumar, "Rank Aggregation Revisited," *Proceedings of the 10th International Conference on World Wide Web*, HongKong, China, May 2001, pp. 613–622.
- [12] S. Negahban, S. Oh, and D. Shah, "Iterative ranking from pair-wise comparisons," *Advances in Neural Information Processing Systems*, vol. 9, no. 93, pp. 2474–2482, 2012.
- [13] V. Pihur, S. Datta, and S. Datta, "Weighted rank aggregation of cluster validation measures: a Monte Carlo cross-entropy approach," *Bioinformatics*, vol. 23, no. 13, pp. 1607–1615, May 2007.
- [14] V. Pihur, S. Datta, and S. Datta, "RankAggreg, an R package for weighted rank aggregation," *Bmc Bioinformatics*, vol. 10, no. 1, pp. 62–72, Feb. 2009.
- [15] C. M. Burnett and V. Kogan, "Ballot (and voter) "exhaustion" under instant runoff voting: An examination of four ranked-choice elections," *Electoral Studies*, vol. 37, pp. 41–49, Mar. 2015.
- [16] R. D. Luce, *Individual Choice Behavior: A Theoretical Analysis*, New York: John Wiley, 1959.
- [17] Q. Tao, X. B. Geng, and T. Y. Liu, "A new probabilistic model for rank aggregation," *Advances in Neural Information Processing Systems 23*, Curran Associates, Inc., 2010, pp. 1948–1956.
- [18] J. K. Ding, D. Q. Han, and Y. Yang, "Novel instant-runoff ranking fusion approaches," *IEEE 2015 International Conference on Multisensor Fusion and Integration for Intelligent Systems*, San Diego, CA, USA, Sep. 2015, pp. 226–232.
- [19] Y. B. Shin, S. Lee, S. G. Chun, and D. Chung, "A critical review of popular multi-criteria decision making methodologies," *Issues in Information Systems*, vol. 14, no. 1, pp. 358–365, 2013.
- [20] J. González-Pachón and C. Romero, "Aggregation of partial ordinal rankings: an interval goal programming approach," *Computers & Operations Research*, vol. 28, no. 8, pp. 827–834, July 2001.
- [21] R. Fagin, R. Kumar, and D. Sivakumar, "Comparing top k lists," *SIAM Journal on Discrete Mathematics*, vol. 17, no. 1, pp. 134–160, 2003.
- [22] W. D. Cook, "Distance-based and ad hoc consensus models in ordinal preference ranking," *European Journal of Operational Research*, vol. 172, no. 2, pp. 369–385, Aug. 2005.
- [23] M. G. Kendall, *Rank Correlation Methods*. London: Griffin, 1948.