A new integer programming formulation and refined social choice property for expediting the solution to the consensus ranking problem

Yeawon Yoo, Adolfo R. Escobedo
School of Computing, Informatics, and Decision Systems Engineering, Arizona State University

IRA A. FULTON SCHOOLS OF engineering
ARIZONA STATE UNIVERSITY

Introduction

- Introduction
  - Group decision-making has been studied extensively since the shaping of democratic society
  - Many people devoted their efforts to develop a fair and consistent system that aggregates the opinion of each individual to make better social decisions
  - Example of group decision-making
    : University rankings, Proposal funding decisions, Online product review

Consensus ranking

- Limitation of ratings
  - The rating scales of two individuals are, in general, not comparable
- Consensus ranking problem (i.e. ranking aggregation)
  - The consensus ranking problem is at the center of many group decision-making processes
  - It entails finding an ordinal vector or ranking of a set of competing objects that minimizes disagreement with a profile of preferences (represented as ranking vectors)

Two frameworks of the consensus ranking problem

- Given a set of m rankings $\boldsymbol{a}^1, \ldots, \boldsymbol{a}^m \in \boldsymbol{\Omega}$, the median ranking is the optimal solution to the following problems:
- The **distance-based** ranking aggregation problem

$$\min_{\boldsymbol{r} \in \boldsymbol{\Omega}_C} \sum_{i=1}^{m} d(\boldsymbol{a}^i, \boldsymbol{r})$$

  - Objective is to minimize the cumulative distance or disagreement
- The **correlation coefficient-based** ranking aggregation problem

$$\max_{\boldsymbol{r} \in \boldsymbol{\Omega}_C} \sum_{i=1}^{m} \tau(\boldsymbol{a}^i, \boldsymbol{r})$$

  - Objective is to maximize the cumulative correlation or agreement

Computational difficulties

- Obtaining even just one consensus or median ranking via correlation-based methods (or the equivalent axiomatic distance-based methods) is an NP-hard problem. (Bartholdi et al. 1989)

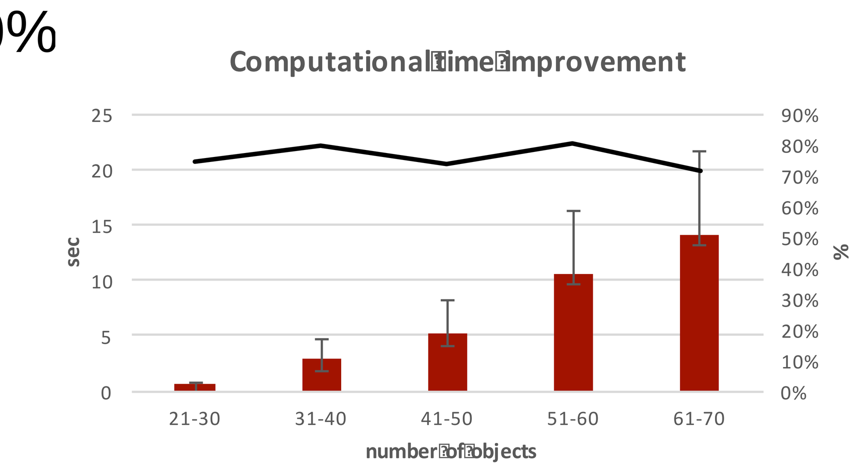
	number of solutions	
$n=2$	$3 \simeq \frac{1.442^3 2!}{2}$	
$n=3$	$\boxed{13 \simeq \frac{1.442^4 3!}{2}}$	$\boldsymbol{a} = (1,1),(1,2),(2,1)$
$n=4$	$75 \simeq \frac{1.442^5 4!}{2}$	
$n=5$	$541 \simeq \frac{1.442^6 5!}{2}$	
$n=6$	$4683 \simeq \frac{1.442^7 6!}{2}$	
$n=7$	$47293 \simeq \frac{1.442^8 7!}{2}$	
$\vdots$	$\vdots$	
$n=10$	$102247563 = \boxed{1.0222 \text{ billions}}$	
$n=20$	$2677687796244384203115 = 2677687796 \text{ trillions}$	

Approach 1 – Integer programming formulation (IP)

$$\begin{aligned}
\underset{y}{\text{maximize}} \quad & \sum_i \sum_j c_{ij}(2y_{ij}-1) & \\
\text{subject to} \quad & y_{ij} - y_{kj} - y_{ik} \geq -1 & i,j,k = 1,\ldots,n; \quad i \neq j \neq k \neq i \\
& y_{ij} + y_{ji} \geq 1 & i,j = 1,\ldots,n; \quad i \neq j \\
& 0 \leq y_{ij} \leq 1 & i,j = 1,\ldots,n; \\
& y_{ij} \in \mathbb{Z} & i,j = 1,\ldots,n;
\end{aligned}$$

Approach 1 – Computational experiments

- $\uparrow (|V|) \Rightarrow \uparrow$ (Computational time improvement)
- For all ranges of the number of objects, the computational time improves by 70%

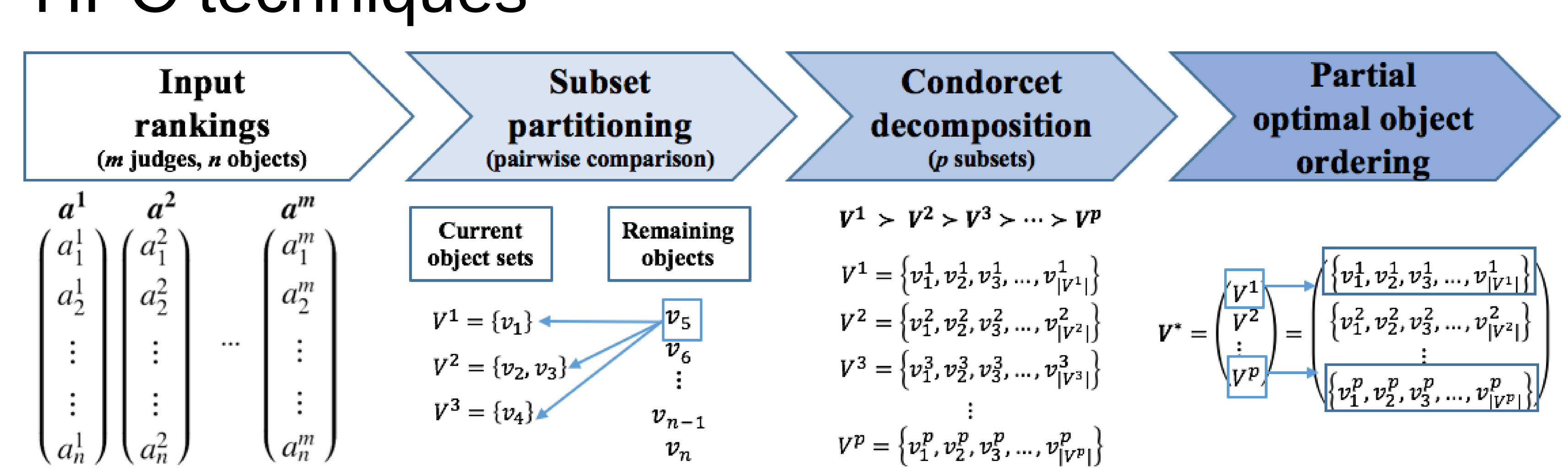


Approach 2 – Condorcet criterion

- Condorcet criterion (Condorcet 1785)
- Extended Condorcet criterion (Truchon 1998)
- Generalized Condorcet criterion
  - Definition: If **a decisive majority** ranks **every** objects $v_i \in V$ ahead of **every** objects $v_j \in V'$ (i.e. a less people believes $v_j \in V'$ ahead of $v_i \in V$), then every objects $v_i \in V$ must be ahead of $v_j \in V'$ in the consensus ranking

Approach 2 – HPC techniques applicability

- Each independent object set can be solved by using HPC techniques



Approach 2 – Computational experiments

- $\uparrow (|V|) \Rightarrow \uparrow$ (Computational time)
- $\uparrow (\phi) \Rightarrow \uparrow$ (Computational time)
- Even for ϕ <0.5, the difference in computational times increases

AVG net computational time **improvement**

	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
10	0.003	0.003	0.004	0.003	0.003	0.003	0.003	0.004	0.004
20	0.030	0.029	0.030	0.030	0.028	0.027	0.041	0.066	0.086
30	0.112	0.114	0.115	0.125	0.123	0.137	0.237	0.357	24.121
40	0.212	0.214	0.212	0.224	0.211	0.453	0.714	2.283	784.355
50	0.491	0.485	0.484	0.493	0.513	0.770	1.397	10.240	351.555
60	0.899	1.010	0.915	0.971	1.061	1.225	3.545	62.429	2979.755
70	1.413	1.418	1.446	1.419	1.530	2.113	17.307	175.678	6508.484
80	2.617	2.465	2.527	2.315	2.286	6.246	69.752	1246.854	7199.981
90	3.970	3.598	3.624	3.735	4.564	16.864	65.423	896.681	7199.973
100	5.088	5.055	5.082	5.126	8.022	28.183	837.353	1875.160	7199.965
200	62.384	62.384	62.611	62.576	85.368	1741.126	7199.961	7199.961	7199.935
300	291.330	291.804	291.504	290.909	651.276	7199.886	7199.890	7199.891	7199.685

Unit: sec

AVG computational time **speedups**

	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
10	3.11	3.62	3.79	3.41	3.14	2.87	3.20	4.04	3.19
20	11.50	10.73	10.15	11.83	10.04	9.89	14.85	24.15	28.07
30	27.87	27.00	29.31	29.69	29.60	32.75	53.67	84.20	1884.3
40	38.23	39.05	38.46	41.04	38.07	80.79	126.88	361.94	50734.6
50	65.06	72.04	72.65	79.90	79.03	123.36	222.38	1653.26	45071.4
60	104.06	113.81	105.98	112.73	121.06	138.49	401.00	7209.70	139751.0
70	168.18	159.28	157.57	139.82	170.81	244.28	1936.46	19208.84	469169.2
80	228.47	211.96	222.63	211.47	203.76	585.85	6538.07	112919.05	432937.7
90	265.23	228.89	255.37	291.58	427.69	1509.25	6180.38	85445.58	296263.2
100	295.89	407.12	412.22	403.69	657.20	2413.04	85180.78	161875.84	259982.8
200	1162.35	1162.35	1172.96	1178.59	1821.24	42891.95	183841.20	183298.09	120995.4
300	2151.66	2257.02	2337.72	2388.17	5534.73	63335.16	65255.11	66417.23	49515.0

* (Computational time w/o subset partitioning / w/ subset partitioning)

Conclusions and future work

- Summary
  - We derived a property that aligns better with Kemeny-Snell distance
  - We introduced an IP formulation to expedite solution process
- Future work
  - More computational experiments will be conducted on various types of preference data and ranking models
  - We can apply distributed computing to obtain optimal rankings separately and then combine sub-problem solutions