Pearson Rank: A Head-Weighted Gap-Sensitive Score-Based Correlation Coefficient

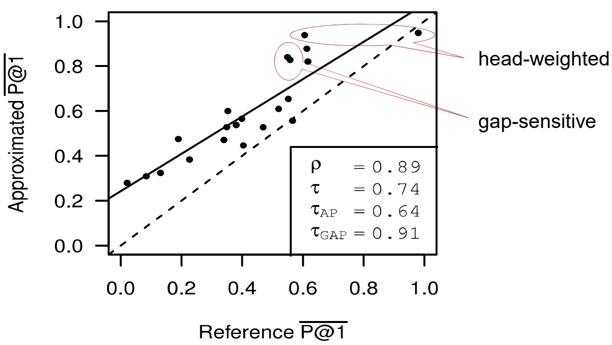
Ning Gao,[1,2] **Mossaab Bagdouri,**[1] and **Douglas W. Oard**[1,2] {ninggao,mossaab,oard}@umd.edu

[1] University of Maryland, College Park

[2] Human Language Technology Center of Excellence, The Johns Hopkins University

Motivation

- Participating systems are ranked in TREC, CLEF, NTCIR, FIRE evaluations
  - ➢ Goal: reliable system comparisons despite incomplete judgments
- Key idea: compare system rankings using complete or incomplete judgments
  - ➢ Which differences matter most? Large gaps? Those between good systems?
- *Pearson Rank*: **head-weighted** correlation coefficient for an **interval** scale.



Pearson Rank Definition

number of ranked items

any item j above i in the ground truth ranking

ground-truth ranking

approximated ranking

$$\rho_r(Y \mid X) = \frac{1}{\sum_{i=2}^{m} x_i} \cdot \sum_{i=2}^{m} \left[x_i \cdot \frac{\sum_{j=1}^{i-1}(x_j - x_i)\cdot(y_j - y_i)}{\sqrt{\sum_{j=1}^{i-1}(x_j - x_i)^2 \sum_{j=1}^{i-1}(y_j - y_i)^2}} \right]$$

score difference in ground-truth ranking

score difference in approximated ranking

Example

x_1 y_1

x_2 y_3

x_3 y_2

$$\rho_r = \frac{\rho_{r,2} + \rho_{r,3}}{x_2 + x_3}$$

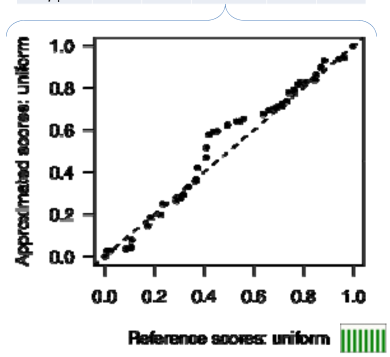
$$\rho_{r,2} = x_2 \cdot \frac{(x_1 - x_2)(y_1 - y_2)}{\sqrt{(x_1 - x_2)^2 (y_1 - y_2)^2}}$$

$$\rho_{r,3} = x_3 \cdot \frac{(x_1 - x_3)(y_1 - y_3) + (x_2 - x_3)(y_2 - y_3)}{\sqrt{(x_1 - x_3)^2 + (x_2 - x_2)^2}\sqrt{(y_1 - y_3)^2 + (y_2 - y_3)^2}}$$

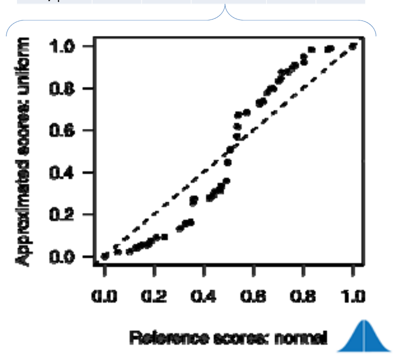
Simulations

- Perturb scores in ways that maintain ranking
  - $\tau = \tau_{AP} = \tau_{GAP} = 1$ (by construction)

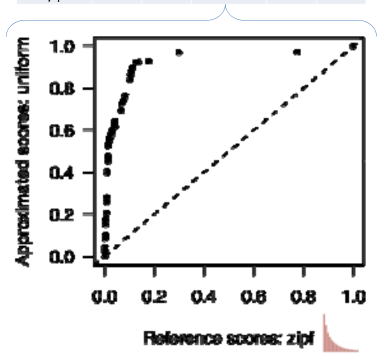
Quartile	0th	1st	median	3rd	4th
ρ_r	0.89	0.97	**0.98**	0.99	1.00



Quartile	0th	1st	median	3rd	4th
ρ_r	0.80	0.95	**0.97**	0.98	1.00



Quartile	0th	1st	median	3rd	4th
ρ_r	0.51	0.80	**0.87**	0.91	1.00



Properties

- $-1 \le \rho_r \le +1$
- Early swaps yield greater reduction in ρ_r
- Swaps with larger gaps yield greater reduction in ρ_r
- Gap errors without swaps yield reduction in ρ_r

	Ordinal	Interval	Head-Weighted	Symmetric
Pearson ρ	✓	✓	✗	✓
Kendall's τ	✓	✗	✗	✓
Yilmaz's τ_{AP}	✓	✗	✓	✗
Gao's τ_{GAP}	✓	✓	✓	✗
Pearson Rank ρ_r	✓	✓	✓	✗