Ranking by Stealing Human Cycles

Tingfan Wu
Department of Computer Science
National Taiwan University

Joint work with Tzu-Kuo Huang, Chih-Jen Lin and Ruby C. Weng

The Problem

- Given a large(>1M) number of photos.
- How to obtain the global ranking effectively and efficiently?
- Hot or Not's solution:
  Steal human cycles from Internet visitors.

Hot or Not: Score 1 to 10

Drawbacks of Scoring Method

Vulnerable to voter variation (if low #vote/photo)

- An example

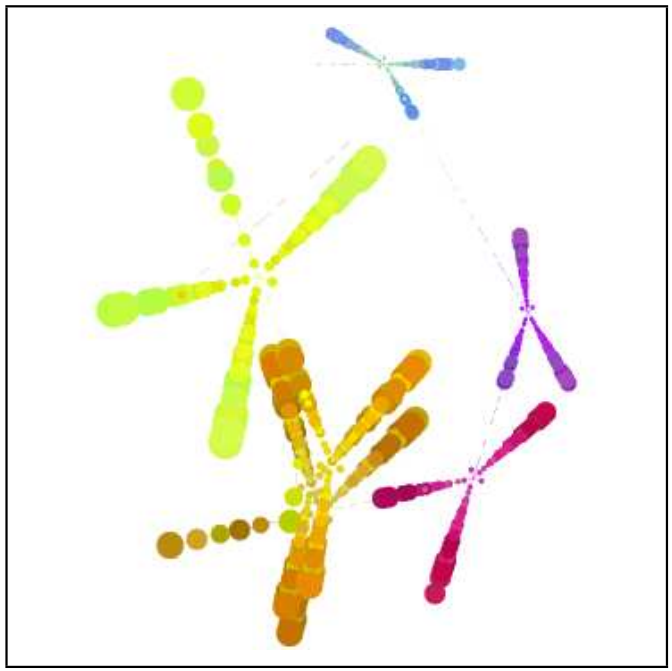


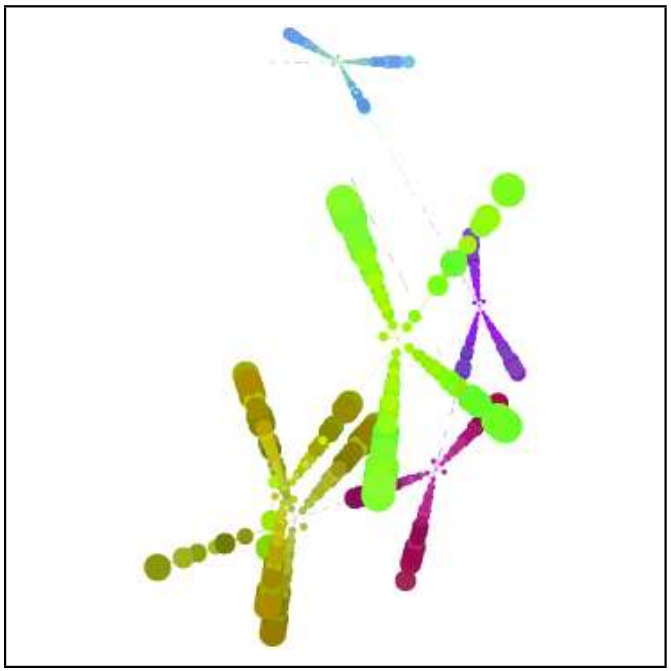
Real score	6	5	4
Nice voter		7	6
Tough voter	4	3	
Avg. Score	4	5	6

- Human compare the picture with previously seen ones. Dependent on one's previous experience.

Drawbacks of Scoring Method

- Some objects are hard to give score.





- Comparison is easier and more objective.

New Challenges

- Number of objects very large
  Hot or Not : 24.3M photos
- Impossible to compare all $\binom{k}{2}$ pairs: 300T pairs
- Selecting a small subset of pairs

Conditions:

1. Connectivity : for any $i \neq j$, $\exists$ comparisons
   $$i \text{ vs } i_1,\ i_1 \text{ vs } i_2,\ \ldots,\ i_s \text{ vs } j.$$
2. Fairness : each picture compared to equal number of opponents

Design of The System

- Cyclic design (#objects = 7)

{1}:	(0,1)	(1,2)	(2,3)	(3,4)	(4,5)	(5,6)	(6,0)
{3}:	(0,3)	(1,4)	(2,5)	(3,6)	(4,0)	(5,1)	(6,2)

- Connectivity: $0 \rightarrow 1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 5 \rightarrow 6 \rightarrow 0$
- Fairness:

0:	1 3 4 6
1:	0 2 4 5

- Related to experimental design: efficiency

Paired Comparison Method

- Bradley-Terry model: paired comparisons $\longrightarrow$ global ranking
- Object j's hotness : $p_j \geq 0, j = 1, 2, \ldots, k$

$$P(\text{object } i \text{ beats } j) = \frac{p_i}{p_i + p_j}.$$

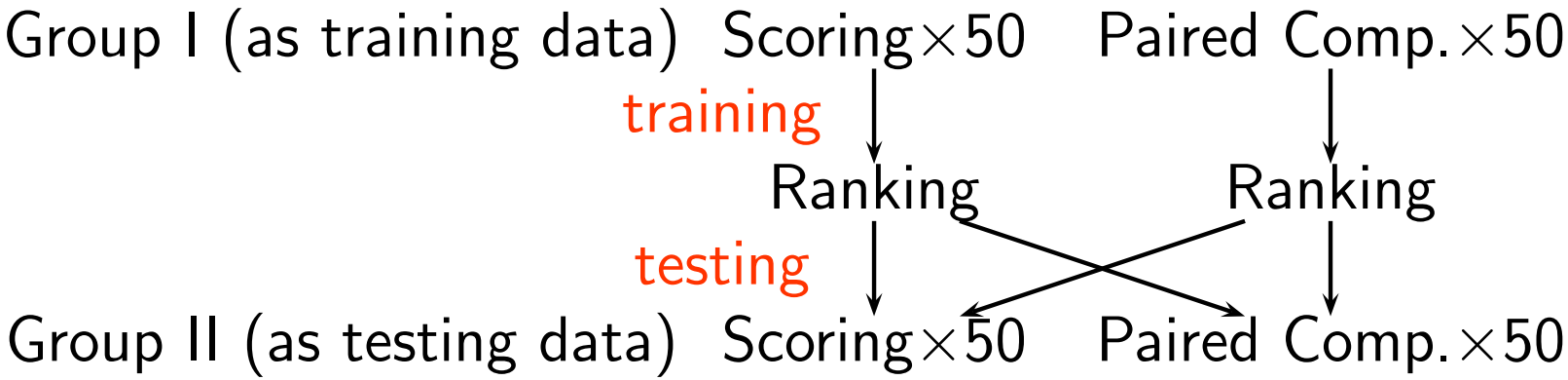
- Object i beats j r_{ij} times.
- Maximum log-likelihood

$$\max_{\mathbf{p}} \quad \sum_{i,j:i\neq j} r_{ij} \log \frac{p_i}{p_i + p_j}, \quad \text{s.t.} \sum_{i=1}^{k} p_i = 1,\ p_i \geq 0,\ i = 1, \ldots, k.$$

- Unique global maximum exists if *connectivity condition* holds.

Experiment Design : Evaluation

- Training & Testing



- Calculate accuracy

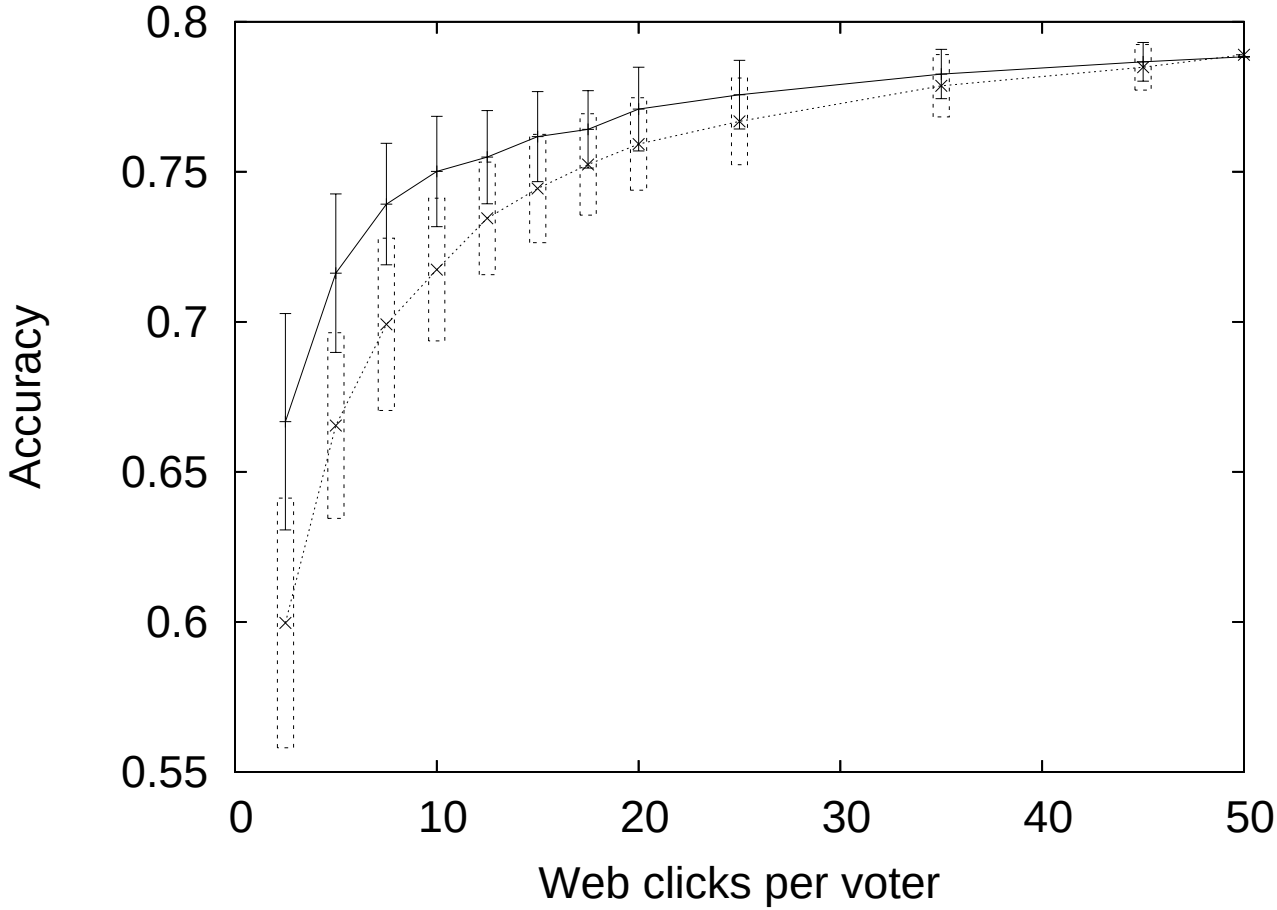
  ranking $5 > 3 > 4 > 2 > 1$

  testing data $(2 > 1)_{\circ}\ (3 < 4)_{\times}\ (5 > 2)_{\circ}\ (3 > 1)_{\circ}$

  3 correct + 1 error = 75%

Experiment Result

- Training data subsampled to reflect different #clicks/voter.



paired comparison ($+$, solid) scoring ($\times$, dotted)

Conclusion

Summary

- If low #clicks/voter, paired comparison outperforms scoring.
- Binary choices vs 10 choices
  each pair: 1.93s vs each score: 2.17s.

Ongoing and Future Work

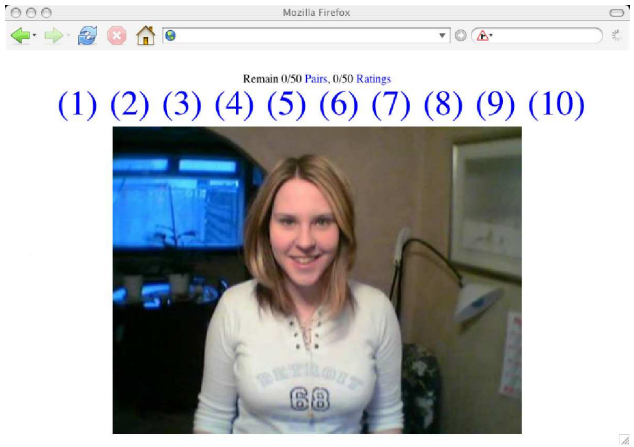
- Hotter Or Notter
  http://hotterornotter.csie.org/
  Please help to cast votes.
- Pair selection in incremental/decremental scenarios.

- Testing instaces from consequtive scores

ID	2	1	4	3	0 ...
score	9	7	6	2	5 ...
pairs	(**2**,1)	(~~**1**,4~~)	(**4**,3)	(3,**0**)	...

- Data collection (for each voter)

Scoring×50



Paired comparison×50