Ranking Individuals by Group Comparisons

Chih-Jen Lin
Department of Computer Science
National Taiwan University

Joint work with Tzu-Kuo Huang and Ruby C. Weng

Outline

- Problem

  Ranking individuals by group comparisons
- Existing approaches
- Our approaches
- Real applications:

  Ranking bridge partnerships
- Discussion and conclusions

The Problem

- Many sports are team comparisons
  How to rank individuals?
- Rank a basketball player by average points
  But ignore teammates'/opponents' abilities.
- In bridge
  Two partnerships vs. two partnerships
  Match record shows which two are better
  But how to rank partnerships?

- Multi-class classification by error-correcting codes [Dietterich and Bakiri, 1995, Allwein et al., 2001]

  Some classes vs. some others

  Finding the winning class (individual)

A Naive Approach: SUM

- Summing # of winning games

$$\frac{\sum_{i:s\in I_i^+} n_i^+ + \sum_{i:s\in I_i^-} n_i^-}{\sum_{i:s\in I_i} 1}.$$

- Not consider opponents' abilities

  Susceptible to individuals playing very few (or many) games
- Not consider teammates' abilities

  Strong and weak players: the same credits
- Ranking by SUM similar to that of teams.

Our 1st Approach [Huang et al., 2005]

- Games: $\{1,3\}$ vs. $\{2\}$; $\{1,4\}$ vs. $\{2,3,5\}$, etc.
- Individual j's ability: $p_j \geq 0$
- The ith setting: team I_i^+ vs. team I_i^-

$$P(I_i^+ \text{ beats } I_i^-) = \frac{\sum_{j:j\in I_i^+} p_j}{\sum_{j:j\in I_i} p_j} = \frac{\sum_{j:j\in I_i^+} p_j}{\sum_{j:j\in I_i^+} p_j + \sum_{j:j\in I_i^-} p_j}$$

Our 1st Approach [Huang et al., 2005]

- Games: {1,3} vs. {2}; {1,4} vs. {2,3,5}, etc.
- Individual j's ability: $p_j \geq 0$
- The ith setting: team I_i^+ vs. team I_i^-

$$P(I_i^+ \text{ beats } I_i^-) = \frac{\sum_{j:j \in I_i^+} p_j}{\sum_{j:j \in I_i} p_j} = \frac{\sum_{j:j \in I_i^+} p_j}{\sum_{j:j \in I_i^+} p_j + \sum_{j:j \in I_i^-} p_j}$$

- Extension of Bradley-Terry model [Bradley and Terry, 1952]

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

- Minimizing the negative log-likelihood

$$\min_{\mathbf{p}} \quad -\sum_{i=1}^{m} \left(n_i^+ \log \frac{\sum_{j:j\in I_i^+} p_j}{\sum_{j:j\in I_i} p_j} + n_i^- \log \frac{\sum_{j:j\in I_i^-} p_j}{\sum_{j:j\in I_i} p_j} \right)$$

$$\text{subject to} \quad \sum_{j=1}^{k} p_j = 1, 0 \leq p_j, j = 1, \ldots, k.$$

- n_i^+ and n_i^-: # wins by I_i^+ and I_i^-; $\quad n_i \equiv n_i^+ + n_i^-$
- May be non-convex

- Estimation by an iterative method

  stationary point
- Extensions

  weighted individual skill

  home-field advantage

  ties

  comparisons with more than two teams.

Our 2nd Approach: [Huang et al., 2006]

- An exponential model
- k individuals' abilities: a vector $\mathbf{v} \in R^k$, $-\infty < v_s < \infty$, $s = 1, \ldots, k$.
- Ability of team I_i^+ and I_i^-: sum of members'

$$T_i^+ \equiv \sum_{s:s \in I_i^+} v_s \quad \text{and} \quad T_i^- \equiv \sum_{s:s \in I_i^-} v_s.$$

- Teams' actual performances: random variables Y_i^+ and Y_i^-

$$P(I_i^+ \text{ beats } I_i^-) \equiv P(Y_i^+ - Y_i^- > 0).$$

- Distribution unknown, assume doubly-exponential extreme-value distribution

$$P(Y_i^+ \leq y) = \exp(-e^{-(y-T_i^+)}),$$

- Probability that I_i^+ wins

$$P(I_i^+ \text{ beats } I_i^-) = \frac{e^{T_i^+}}{e^{T_i^+} + e^{T_i^-}}.$$

- Can assume normal; but model more complex

Estimation: Regularized Least Square

- n_i^+ and n_i^-: # games teams I_i^+ and I_i^- win

$$\frac{e^{T_i^+}}{e^{T_i^+} + e^{T_i^-}} \approx \frac{n_i^+}{n_i^+ + n_i^-} \quad \Rightarrow \quad e^{T_i^+ - T_i^-} \approx \frac{n_i^+}{n_i^-}.$$

- Solve

$$\min_{\mathbf{v}} \quad \sum_{i=1}^{m} ((T_i^+ - T_i^-) - \log(n_i^+ / n_i^-))^2$$

- Unique solution:

Adding a regularized term $\mu \mathbf{v}^T \mathbf{v}$

Estimation: Maximum Likelihood (ML)

- Negative log-likelihood function:

$$\arg\min l(\mathbf{v})$$

where

$$l(\mathbf{v}) \equiv -\sum_{i=1}^{m}\left(n_i^+ \log \frac{e^{T_i^+}}{e^{T_i^+}+e^{T_i^-}} + n_i^- \log \frac{e^{T_i^-}}{e^{T_i^+}+e^{T_i^-}}\right)$$

- Convex

Maximum Likelihood: Iterative Scaling

- Standard optimization methods: Newton's etc.

Maximum Likelihood: Iterative Scaling

- Standard optimization methods: Newton's etc.
- Simple iterative scaling

  Update the sth component;
  $\boldsymbol{\delta} \equiv [0, \ldots, 0, \delta_s, 0, \ldots, 0]^T$

$$l(\mathbf{v} + \boldsymbol{\delta}) - l(\mathbf{v}) \leq -\left(\sum_{i:s \in I_i^+} n_i^+ + \sum_{i:s \in I_i^-} n_i^-\right)\delta_s +$$

$$\left(\sum_{i:s \in I_i^+} \frac{n_i e^{T_i^+}}{e^{T_i^+} + e^{T_i^-}} + \sum_{i:s \in I_i^-} \frac{n_i e^{T_i^-}}{e^{T_i^+} + e^{T_i^-}}\right)(e^{\delta_s} - 1)$$

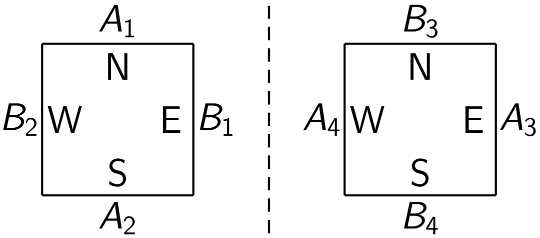
- Updating rule

$$v_s \leftarrow v_s + \log \frac{\sum_{i:s\in I_i^+} n_i^+ + \sum_{i:s\in I_i^-} n_i^-}{\sum_{i:s\in I_i^+} \frac{n_i e^{T_i^+}}{e^{T_i^+}+e^{T_i^-}} + \sum_{i:s\in I_i^-} \frac{n_i e^{T_i^-}}{e^{T_i^+}+e^{T_i^-}}}.$$

- Under a minor assumption,
  any limit point a global minimum

Ranking Bridge Partnerships

- A match setting:



Team A: two partnerships (A_1, A_2), (A_3, A_4)
Team B: two partnerships (B_1, B_2), (B_3, B_4)

- Same board for (B_1, B_2), (A_3, A_4)
  Avoid uneven hands

Bridge Scoring

- First three boards; India vs. Portugal in Bermuda Bowl 2005

Board	Table I		Table II		IMPs	
	NS	EW	NS	EW	IN	PT
1		1510		1510		
2	100		650			11
3		630		630		

- International Match Points (IMPs): difference in two teams' total scores
- IMPs of all rounds ⇒ Victory Points (VP)

  Overall results between two teams; used as n_i^+, n_i^-

Experimental Settings

- Bermuda Bowl 2005, the most prestigious bridge event
- 22 teams, round robin
  $\binom{22}{2} = 231$ matches
- Most teams: six players $\Rightarrow$ three fixed partnerships
  playing similar $\#$ matches
- Total 69 partnerships
  Other details not shown here

Results:
HNG: generalized Bradley-Terry model
RLS, ML: estimations of the exponential model

Team	Partnership rankings											
	RLS			ML			HNG			SUM		
IT	21	18	13	8	17	18	7	14	16	5	4	12
US2	63	67	1	52	66	1	43	66	1	47	29	2
US1	9	36	41	10	19	37	10	15	38	23	6	10
SE	2	55	37	2	25	53	2	12	64	1	19	39
IN	14	40	42	9	32	41	9	30	42	20	14	15
AR	33	26	32	26	21	29	25	23	34	16	17	28
EG	47	30	27	43	27	13	52	22	13	38	22	3
	46			57			50			8		
. . .												
AU	44	51	13	42	51	20	40	53	21	42	51	41
NZ	68	23	3	68	48	5	67	39	4	64	36	13
UK	10	24	59	12	36	64	17	33	63	45	18	54

Results of U.S.A.2

P1	P2	P3	Score		vs.
⋆	⋆		0	25	NL(10)
⋆	⋆		12	18	ZA(12)
⋆	⋆		11	19	UK(17)
⋆	⋆		14	16	TW(19)
⋆		⋆	25	5	IN(5)
⋆		⋆	22	8	NZ(16)
⋆		⋆	22	8	PL(20)
⋆		⋆	12	18	US1(3)
	⋆	⋆	22	8	SE(4)
	⋆	⋆	19	11	EG(7)
	⋆	⋆	19	11	BR(8)

P1	P2	P3	Score		vs.
	⋆	⋆	24	6	RU(13)
	⋆	⋆	20	10	PT(14)
	⋆	⋆	20	10	CA(18)
	⋆	⋆	25	3	JO(22)
	⋆	⋆	15	15	AU(15)
	⋆	⋆	12	18	IT(1)
	⋆	⋆	13	17	AR(6)
	⋆	⋆	14	16	JP(9)
	⋆	⋆	13	17	CN(11)
	⋆	⋆	14	16	GP(21)

- (P1,P3): significantly win 3, but lose 1
- (P2,P3): win 7 matches but lose 5
- P2 is not better than P1

Ranking Many Objects with Limited Resources

- New research we are doing

  See talk by Ting-Fan Wu in the workshop next Friday
- Current method: scoring

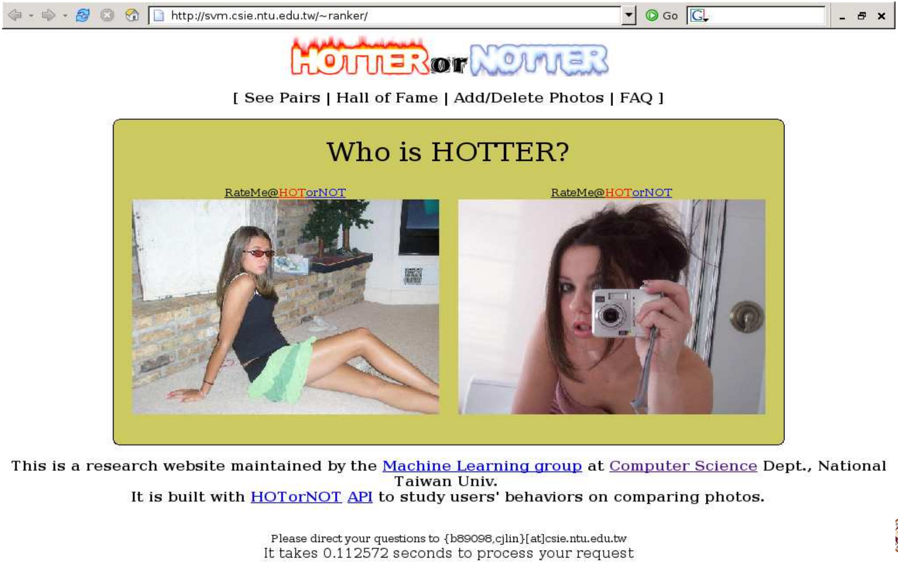
  Conference reviewing

  Hot or Not `http://www.hotornot.com`

Hot or Not

http://www.hotornot.com/
Go
HOT or NOT
Login
Submit picture
Please select a rating to see the next picture.
1 2 3 4 5 6 7 8 9 10
NOT
HOT
Show me women only of any age
Show her to a friend:
http://www.hotornot.com/r/?eid=EUEZOMH-LNT
Email picture
Please help keep this site FUN, CLEAN, and REAL.
Flag this picture as: inappropriate broken best of what is this?
Are you HOT or NOT ?
Submit your picture and find out!

Our New Website: Hotter or Notter



Please help to cask votes. We need more data
http://svm.csie.ntu.edu.tw/~ranker

Thank You

References I

Allwein, E. L., Schapire, R. E., and Singer, Y. (2001).
Reducing multiclass to binary: a unifying approach for margin classifiers.
JMLR, 1:113–141.

Bradley, R. A. and Terry, M. (1952).
The rank analysis of incomplete block designs: I. the method of paired comparisons.
Biometrika, 39:324–345.

Dietterich, T. G. and Bakiri, G. (1995).
Solving multiclass learning problems via error-correcting output codes.
J. Artificial Intelligence Res., 2:263–286.

Huang, T.-K., Lin, C.-J., and Weng, R. C. (2006).
Ranking individuals by group comparisons.
In *Proceedings of the Twenty Third International Conference on Machine Learning (ICML)*.

Huang, T.-K., Weng, R. C., and Lin, C.-J. (2005).
A generalized Bradley-Terry model: From group competition to individual skill.
In *Advances in Neural Information Processing Systems 17*. MIT Press, Cambridge, MA.