Expectation Propagation for Rating Players in Sports Competitions

Adriana Birlutiu and Tom Heskes

Institute for Computing and Information Sciences
Radboud University Nijmegen

Outline

Purpose: develop and evaluate methods for the analysis of paired comparison data.

Illustrate the methods by rating players in sports, in particular in tennis.

- Bayesian probabilistic model for rating players
- Expectation propagation algorithm and variants of it
- Experimental results on a large tennis dataset
- Challenges

Probabilistic Bayesian Framework

- Input: match outcomes
- Goal: infer the players' strengths

- Approach: consider the players' strengths as a probabilistic variable in a *Bayesian* setting
  - Prior: information available about the players
  - Likelihood: the Bradley-Terry model
  - Bayes' rule to compute the posterior over the players' strengths

$$p(\boldsymbol{\theta}|R) = \frac{1}{d} p(R|\boldsymbol{\theta}) p(\boldsymbol{\theta}) = \frac{1}{d} p(\boldsymbol{\theta}) \prod_{i \neq j} p(r_{ij}|\theta_i, \theta_j)$$

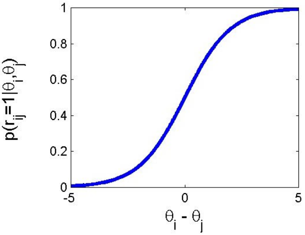
(d = normalization constant)

The Bradley-Terry Model

Expresses the probability of one player winning against another one, as a function of the difference of their strengths:

$$p(r_{ij}|\theta_i, \theta_j) = \frac{1}{1 + \exp[-r_{ij} \cdot (\theta_i - \theta_j)]}$$

where $r_{ij} = 1$ if player i wins against player j, and $r_{ij} = -1$ otherwise.

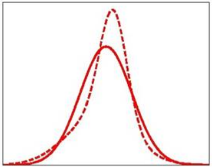


Assumed Density Filtering

$$p(\boldsymbol{\theta}|R) = \frac{1}{d} p(\boldsymbol{\theta}) \prod_{i \neq j} p(r_{ij}|\theta_i, \theta_j)$$

Let q (Gaussian distribution) be the assumed density.

- Include the prior $q(\boldsymbol{\theta}) = p(\boldsymbol{\theta})$
- Include and project one likelihood term at a time



$$\tilde{p}(\boldsymbol{\theta}) = \Psi_{ij}(\theta_i, \theta_j) q(\boldsymbol{\theta}) \,, \quad \Psi_{ij}(\theta_i, \theta_j) = p(r_{ij}|\theta_i, \theta_j)$$

$$q^{\text{new}}(\boldsymbol{\theta}) = \text{Project}\{\tilde{p}(\boldsymbol{\theta})\}$$

Using the KL-divergence projection becomes moment matching.

Term approximation the quotient between the new and old Gaussian approximation.

Expectation Propagation

EP: *backward-forward iterations* to refine the term approximations.

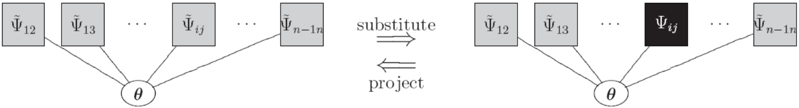
- Initialize the term approximations $\tilde{\Psi}_{ij}(\theta_i, \theta_j)$, e.g., by performing ADF; and compute the initial approximation

$$q(\boldsymbol{\theta}) = p(\boldsymbol{\theta}) \prod_{i \neq j} \tilde{\Psi}_{ij}(\theta_i, \theta_j)$$

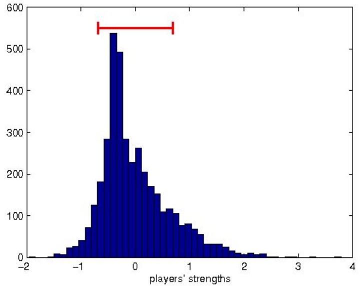
- Repeat until all $\tilde{\Psi}_{ij}$ converge:

$$q^{\setminus ij}(\boldsymbol{\theta}) = \frac{q(\boldsymbol{\theta})}{\tilde{\Psi}_{ij}(\theta_i, \theta_j)}, \quad \tilde{p}(\boldsymbol{\theta}) = \Psi_{ij}(\theta_i, \theta_j) q^{\setminus ij}(\boldsymbol{\theta})$$

$$q^{\text{new}}(\boldsymbol{\theta}) = \underset{q \in Q}{\text{argmin}}\, KL[\tilde{p}||q], \quad \tilde{\Psi}_{ij}^{\text{new}}(\theta_i, \theta_j) = \frac{q^{\text{new}}(\boldsymbol{\theta})}{q^{\setminus ij}(\boldsymbol{\theta})}$$



ATP tennis dataset: 38538 tennis matches, 1139 players, 1995-2006.



A histogram of the players' strengths (means of the posterior distribution) for all years. The bar indicates the average width of the posterior distribution for each of the individual players.

	EP (means)	ATP (points)
Federer, Roger	1 (3.78)	1 (6725)
Nadal, Rafael	2 (2.92)	2 (4765)
Hewitt, Lleyton	3 (2.47)	4 (2490)
Roddick, Andy	4 (2.27)	3 (3085)
Agassi, Andre	5 (2.22)	7 (2275)
Gasquet, Richard	6 (1.92)	16 (1506)
Ljubicic, Ivan	7 (1.85)	9 (2180)
Gaudio, Gaston	8 (1.75)	10 (2050)
Gonzalez, Fernando	9 (1.65)	11 (1790)
Nalbandian, David	10 (1.62)	6 (2370)

Comparison between the ranking obtained using the Bayesian probabilistic framework and the ATP ranking for the year 2005.

EP-Correlated, EP-Independent

We distinguish between two variants of EP:

- EP-Correlated: full distribution coupling all the components of $\boldsymbol{\theta}$

$$q(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\theta}; \boldsymbol{m}, \boldsymbol{\Sigma})$$

- EP-Independent: distribution which factorizes over the elements of $\boldsymbol{\theta}$

$$q(\boldsymbol{\theta}) = \prod_i \mathcal{N}(\theta_i; m_i, \sigma_i)$$

  - Gaussian: diagonal covariance matrix

Accuracy of predicting future matches

Compute ratings for the players at the end of a year, based on the matches from that year.

Perform predictions for matches in the next year:

	ADF		EP-Independent	
	correct	incorrect	correct	incorrect
EP-Correlated				
correct	16636 (54.48%)	2395 (7.81%)	17857 (58.46%)	1174 (3.83%)
incorrect	1902 (6.21%)	9620 (31.50%)	945 (3.09%)	10577 (34.62%)

Comparison between EP-Correlated, ADF and EP-Independent based on the number of matches correctly/incorrectly predicted.

Conclusions

Binomial test ⇒

- EP-Correlated performs significantly better than ADF (iterative improvement helps)
- EP-Correlated performs slightly, but significantly better than EP-Independent (correlations do matter)

Challenges

We considered the most basic probabilistic rating model; this model performs as well as the ATP rating system.

We expect that the more complex models would outperform ATP.

- Generalize to more complex models, e.g., including dynamics over time and team effects
- Specifically for tennis, incorporate the effect of the surface the games are played on
- Apply the comparison to other types of data