Estimating Soccer Team Strength Using a Markov Random Field

John Zech 1/17/2012

Who will win? It's not obvious…

espn soccernet ESPN.CO.UK ESPN.COM ESPNDEPORTES.COM Sign In or Register

Barclays Premier League

Search

NEWS & FEATURES | VIDEO | FIXTURES & RESULTS | PREMIER LEAGUE | ENGLAND | MLS | UCL | EUROPE | GLOBAL | EURO 2012 | FANTASY

Premier League Home | News | Features | Clubs | Fixtures | Results | Table | Stats | Transfers | Blogs | Galleries | History | Showdown

2011/12 Barclays Premier League Table / Standings

Jump To League

Season: 2011-12

Tables | Live Table

2011-2012 BARCLAYS PREMIER LEAGUE TABLE

		OVERALL						HOME					AWAY						
POS	TEAM	P	W	D	L	F	A	W	D	L	F	A	W	D	L	F	A	GD	Pts
1	Manchester City	20	15	3	2	56	16	10	0	0	31	4	5	3	2	25	12	40	48
2	Manchester United	21	15	3	3	52	20	8	1	2	33	14	7	2	1	19	6	32	48
3	Tottenham Hotspur	21	14	4	3	39	21	8	2	1	21	9	6	2	2	18	12	18	46
4	Chelsea	21	12	4	5	40	25	7	1	3	24	16	5	3	2	16	9	15	40
5	Arsenal	21	11	3	7	38	31	7	2	1	16	6	4	1	6	22	25	7	36
6	Newcastle United	21	10	6	5	30	25	6	3	2	16	11	4	3	3	14	14	5	36
7	Liverpool	21	9	8	4	24	18	4	7	0	14	8	5	1	4	10	10	6	35
8	Stoke City	21	8	6	7	22	31	4	4	2	14	11	4	2	5	8	20	-9	30
9	Norwich City	21	7	7	7	32	36	4	3	3	17	15	3	4	4	15	21	-4	28
10	Swansea City	21	6	8	7	23	25	5	5	1	15	6	1	3	6	8	19	-2	26
11	Everton	21	7	4	10	21	25	3	2	5	10	12	4	2	5	11	13	-4	25
12	Sunderland	21	6	6	9	27	24	3	4	3	14	11	3	2	6	13	13	3	24
13	Aston Villa	21	5	9	7	23	27	3	3	5	12	14	2	6	2	11	13	-4	24
14	Fulham	21	5	8	8	23	29	4	3	3	16	15	1	5	5	7	14	-6	23
15	West Bromwich Albion	21	6	4	11	20	30	2	2	7	8	14	4	2	4	12	16	-10	22
16	Wolverhampton Wanderers	21	4	6	11	23	37	3	2	5	14	17	1	4	6	9	20	-14	18
17	Blackburn Rovers	21	4	5	12	32	44	3	0	8	16	22	1	5	4	16	22	-12	17
18	Queens Park Rangers	21	4	5	12	19	36	1	4	5	9	17	3	1	7	10	19	-17	17
19	Bolton Wanderers	21	5	1	15	25	46	1	1	8	11	24	4	0	7	14	22	-21	16
20	Wigan Athletic	20	3	6	11	18	41	1	4	5	10	20	2	2	6	8	21	-23	15

VS ?

Let’s try to figure it out:

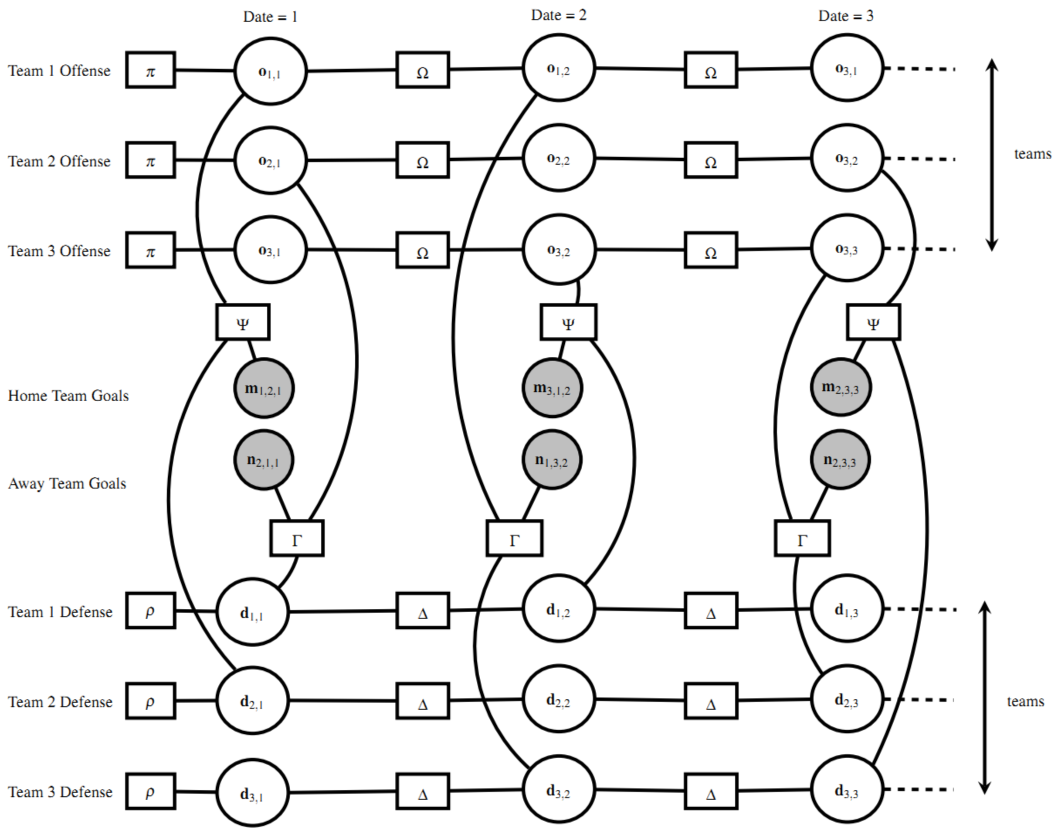
- We want to estimate team strength in soccer, and then we want to use that to predict outcomes. But:

- We only observe teams together → how can we figure any of them out individually?

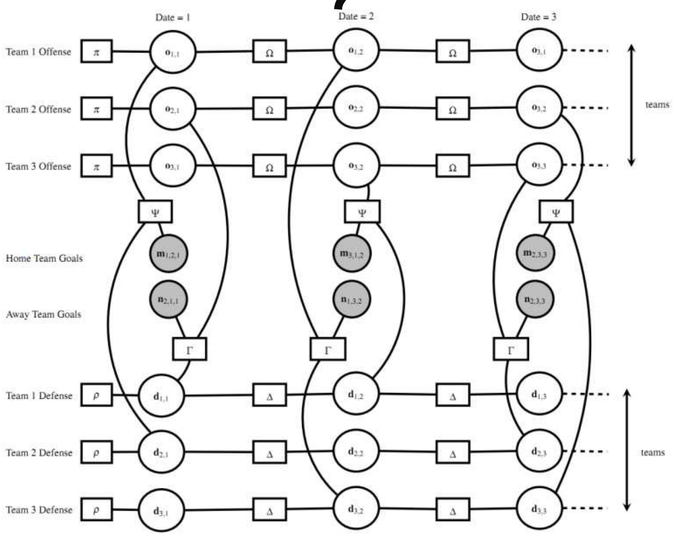
- We don’t know how to map strength to goals.
  We don’t know how team strength evolves over time.
  → can we learn these and strengths simultaneously?

Graphical Model

- This is a big graphical model
- Also, loopy



Graphical Model = Joint Distribution



$$P(\mathcal{X}, \mathcal{Z}, \theta|\Lambda) = P(\mathcal{Z}|\theta)P(\mathcal{X}|\mathcal{Z}, \theta)P(\theta|\Lambda)$$

where

$$P(\mathcal{Z}|\theta) = \prod_{t\in\mathcal{T}}\{\prod_{i\in\mathcal{S}}(\pi_i^{O_{t,1,i}})(\rho_i^{D_{t,1,i}})(\prod_{d\in\mathcal{D}\setminus 1}\prod_{k\in\mathcal{S}}[(\Omega_{i,k}^{(O_{t,d-1,i})(O_{t,d,k})})(\Delta_{i,k}^{(D_{t,d-1,i})(D_{t,d,k})})]\}$$

$$P(\mathcal{X}|\mathcal{Z}, \theta) = \prod_{d\in\mathcal{D}}\prod_{h\in\mathcal{T}}\prod_{a\in\mathcal{T}}\prod_{i\in\mathcal{S}}\prod_{j\in\mathcal{S}}\prod_{g\in\mathcal{G}}\{\Psi_{i,j,g}^{(M_{h,a,d,g})(O_{h,d,i})(D_{a,d,j})}\Gamma_{i,j,g}^{(N_{a,h,d,g})(O_{a,d,i})(D_{h,d,j})}\}$$

$$P(\theta|\Lambda) \propto \prod_{k\in\mathcal{S}}\prod_{j\in\mathcal{S}}\Omega_{k,j}^{\alpha_{j,k}-1}\Delta_{k,j}^{\alpha_{j,k}-1}\prod_{k\in\mathcal{S}}\prod_{j\in\mathcal{S}}\prod_{g\in\mathcal{G}}\Psi_{k,j,g}^{\beta_g-1}\Gamma_{k,j,g}^{\phi_g-1}$$

- An objective function involving both θ and Z

Tool used #1: Belief Propagation

Naïve marginalization is costly $O(N^K)$:

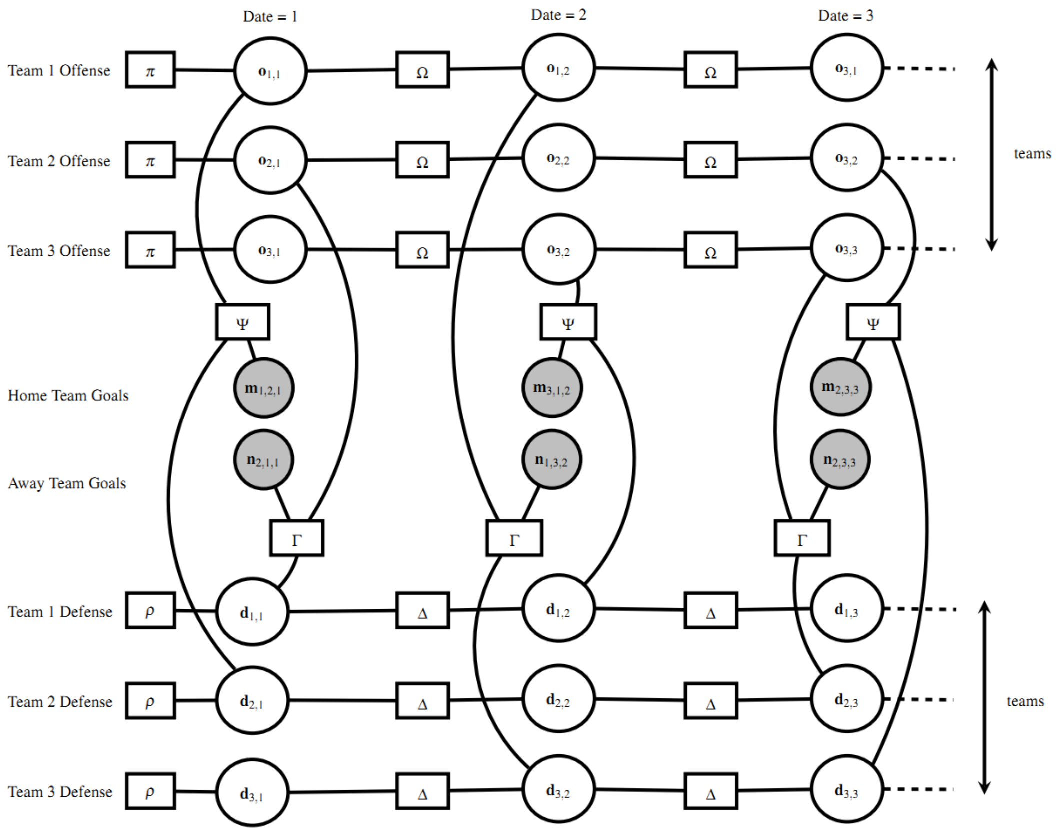
$$P(X_i) = \sum_{X_1}\sum_{X_2}\cdots\sum_{X_{i-1}}\sum_{X_{i+1}}\cdots\sum_{X_n} P(X_1)P(X_2|X_1)P(X_3|X_2)\ldots P(X_i|X_{i-1})P(X_{i+1}|X_i)\ldots P(X_N|X_{N-1})$$

With belief propagation, marginalization is much cheaper $O(NK^2)$:

$$P(X_i) = [\sum_{X_{i-1}} P(X_i|X_{i-1})\ldots[\sum_{X_2} P(X_3|X_2)[\sum_{X_1} P(X_2|X_1)P(X_1)]\ldots]$$
$$[\sum_{X_{i+1}} P(X_{i+1}|X_i)\ldots[\sum_{X_N} P(X_N|X_{N-1})]\ldots]$$

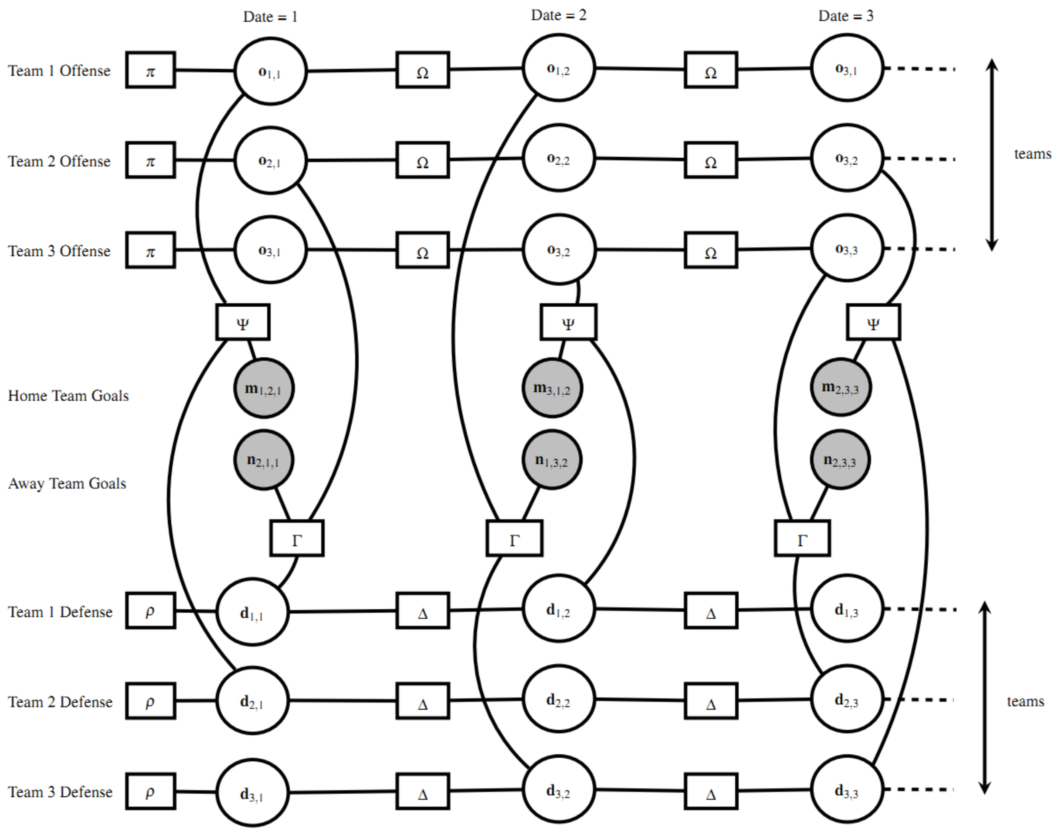
Tool used #1: Belief Propagation

- We can use loopy belief propagation
- Reduces marginalization computational cost from $O(K^N)$ to $O(NK^2)$.



Tool used #2: Expectation Maximization

- We need Z to estimate θ
- We need θ to estimate Z
- Solution: EM Algorithm – hold one constant to estimate the other
- Derive update parameters by maximizing objective w.r.t θ



Estimation procedure

- Run 25 cycles of message passing to find distribution over strengths

- Estimate state parameters (initializations / transitions) using updates derived from log-likelihood

- Estimate goal emission parameters using interior point methods (optimization constrained to enforce that higher strengths correspond to higher skill)

- →cycle until convergence of joint log-likelihood

Implementation

- C++ with MATLAB interfacing for Optimization Toolbox
- Run on Columbia's High-Performance Cluster (HPC) to find regularization parameters

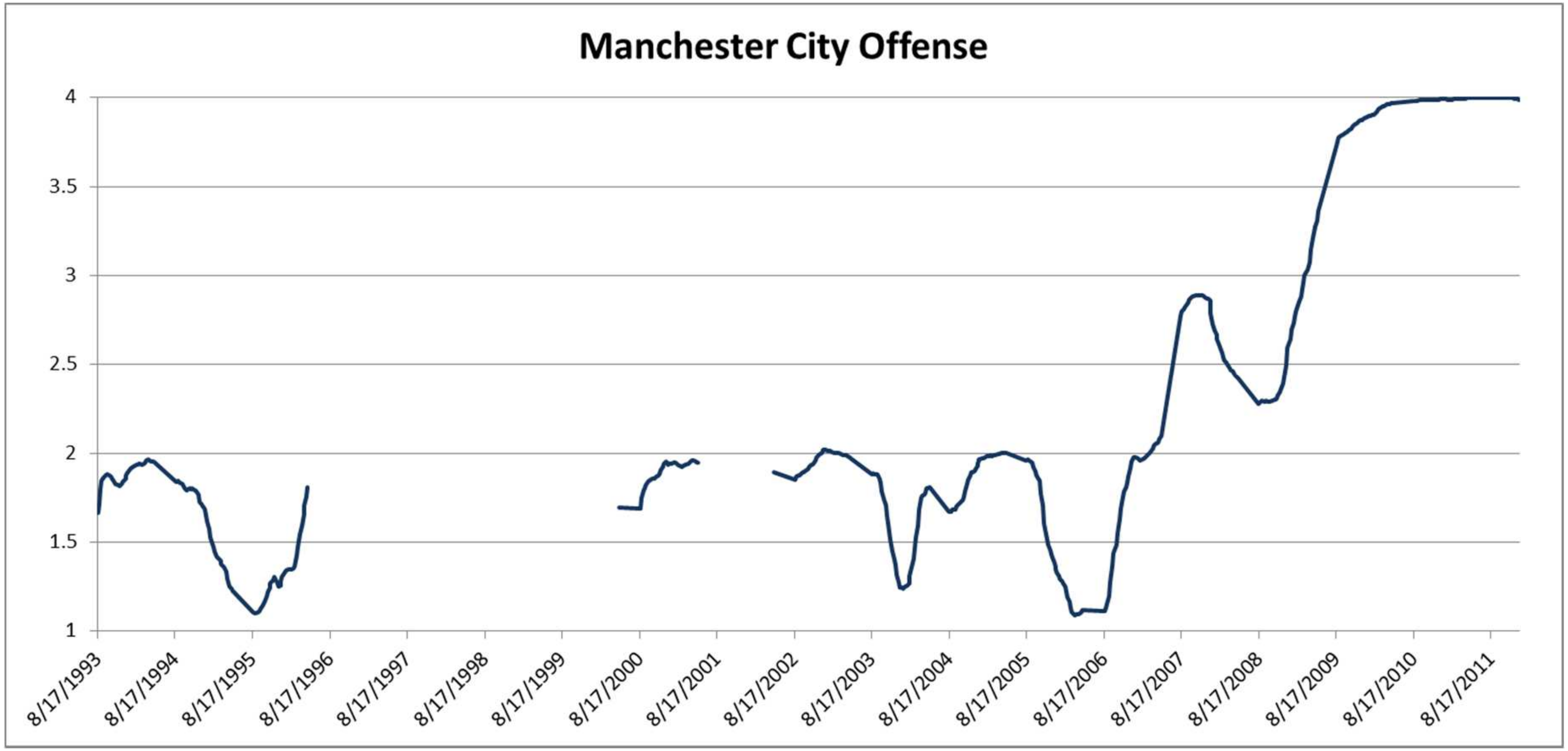
Competing method: Elo

$$h \sim N(\mu_h + \eta, \sigma^2) \qquad a \sim N(\mu_a, \sigma^2)$$

$$s = (h - a) \sim N(\eta + \mu_h - \mu_a, 2\sigma^2)$$

$$P(win) = P(s > \varepsilon) \qquad P(draw) = P(s <= |\varepsilon|) \qquad P(loss) = P(s < -\varepsilon)$$

Results (Diagnostic)



Results (Prediction)