Maximum Likelihood Prediction of Team Ability

Christopher Tsai

June 12, 2008

1 Background: A Comparison of Team Abilities

What are we trying to model? A set of n teams compete in a tournament or round-robin league. Due to a number of debatable but deterministic factors - player skills, practice time, team chemistry, natural ability, coaching prowess - the team has some inherent ability relative to other teams. Placed on a numerical scale, some teams are naturally stronger at the game than other teams. We will call this strength 'team ability' $\mathbf{a}$. However, the tournament or regular season record do not often manifest this relative superiority/inferiority. Instead, the results of head-to-head matches contain significant noise: fluke plays, unforeseen weather conditions, player injuries, aberrant individual performances, emotional and motivational fluctuations, or simply game-based pure luck. Because this noise is difficult to characterize, we cannot decompose it into independent factors or predict outcomes with certainty. Instead, we group all of these potential factors into random fluctuations in 'team ability,' which we will treat as a random vector $\mathbf{a}$.

To which sports does this model apply? This model can apply to any sport that assigns results to one-on-one head-to-head matchups between teams. Some sports work better than others simply because teams encounter one another more frequently (such as in basketball, baseball, hockey, or tennis), thereby providing more data to analyze, but any sport that supports head-to-head results favoring some reasonable set of underlying skills can receive this modeling.

2 Motivation: Power Rankings and Match Prediction

Why do we care about team ability? In attempting to quantify 'team ability' $\mathbf{a}$, we seek to juxtapose all teams in the league on the same simple but sophisticated scale. From a statistician's perspective, this quantization allows us to rank the teams on a ladder, answering questions such as "Which teams seem to be the strongest in the league?" and "Which team has performed best in the league until now?" based on previously compiled results such as adaptively updated matches and last year's records. From a gambler's perspective, this quantization known as 'team ability' can aid us in prediction of future matches. For example, based on what we have seen from two given teams, we can compare their relative abilities and set a point spread (or at least declare a favorite and underdog) for the next match. The layman might simply extrapolate the two teams' most recent matchup results, but we can do better by incorporating matchups with other teams in the league and accounting for noise. Finally, we can evaluate a team's performance by using the past as a basis for our expectations of the upcoming season.

3 Defining Team Ability: Problem Statement

How do we quantify team ability? We will assign each team in the tournament a single scalar decimal a_j between 0 and 1: $0 < a_j < 1$, where j indexes one of our n teams: $j = 1, 2, 3, ..., n$. On this scale, the impeccably strong team boasts an ability of $a_j = 1$, where the hopelessly weak team has an ability of $a_j = 0$. However, in a balanced league with no clear goat or juggernaut, it is possible that no single team has an ability of 0 or 1.

How do we model game-specific noise? Because noise is difficult to decompose, we will treat game-to-game randomness as random fluctuation in team ability. Because this noise comprises the sum of numerous factors - from player health and biorhythm to the game's emotional significance or even temperature - we apply a Gaussian model to the fluctuation $v \sim \mathcal{N}(0, \sigma^2)$. We can hypothesize various values of variance depending on the specific game and league under study, but we will use the empirically viable variance of $\sigma^2 = 0.25$.

What do these team abilities entail? When two teams x and y compete in a game, the probability that the first team x emerges victorious is equal to $\mathbf{prob}(a_x - a_y + v > 0)$, where a_x and a_y represent the team abilities of teams x and y, respectively, and v represents random noise: $v \sim \mathcal{N}(0, 0.25)$.

The optimization variable: $\hat{\mathbf{a}} \in \mathbb{R}^n$, which is a column vector of maximum likelihood team abilities: a_j, one for each team. On our scale, an arbitrary choice, we mandate that $\mathbf{0} \preceq \hat{\mathbf{a}} \preceq \mathbf{1}$.

4 Use of Past Data: Constructing the Game Incidence Matrix

How do we incorporate the outcomes of m past games? We organize a game incidence matrix, given by A, with dimensions $(m \times n)$. Each row of this matrix represents one single past game (out of m past games) that we are considering, while each column of this matrix represents one of the teams (out of n teams) in the league. Typically, for best results, $m = (n-1)!$, so we can incorporate one head-to-head matchup between each pair of teams in the league. Each past outcome contains the numbers of two contending teams and a binary result:

$$(x^{(i)}, y^{(i)}, z^{(i)})$$

For each past game that we consider, we organize a single result. A total of m past games yields a total of m such results, indexed by $i = 1, 2, 3, ..., m$.
$x^{(i)}$ is the number (between 1 and n) of the first team in the i^{th} past contest.
$y^{(i)}$ is the number (between 1 and n) of the second team in the i^{th} past contest.
$z^{(i)}$ is the result of the i^{th} past contest, where $z^{(i)} = +1$ if team $x^{(i)}$ prevailed, or, contrarily, $z^{(i)} = -1$ if team $y^{(i)}$ prevailed.

Let A_{ij} denote the real scalar element in the i^{th} row and j^{th} column of matrix A . Let α_j denote the j^{th} column of the game incidence matrix. Each vector α_j has m elements $(m \times 1)$.

$$A = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{pmatrix} = \begin{pmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_n \end{pmatrix}$$

where

$$A_{ij} = \begin{cases} +z^{(i)} & \text{if } j = x^{(i)}; \\ -z^{(i)} & \text{if } j = y^{(i)}; \\ 0 & \text{otherwise.} \end{cases}$$

This says that, if the first team in $(x^{(i)}, y^{(i)}, z^{(i)})$ won, we use the result $z^{(i)} = +1$ as our matrix entry, and, if the second team won, we use $-z^{(i)}$ to make the entry $+1$. Therefore, every row ("game") of A contains two nonzero entries, one for each of the two teams in previous game i.

5 Reformulation of Maximum Likelihood Team Abilities

We want to find the set of team abilities $\hat{a}$ that is *most likely* given the results of previous matches as reflected in matrix A. We begin by expressing the probability of victory in terms of the standard normal (Z) Gaussian cumulative distribution function $\Phi(z)$:

$$\textbf{prob}\,(\text{team } x \text{ beats team } y) = \textbf{prob}\,(a_x - a_y + v > 0) \tag{1a}$$

$$\textbf{prob}\,(\text{team } x \text{ beats team } y) = \textbf{prob}\,(v > -a_x + a_y) \tag{1b}$$

$$\textbf{prob}\,(\text{team } x \text{ beats team } y) = \textbf{prob}\left(\frac{v-0}{\sigma} > \frac{-a_x + a_y}{\sigma}\right) \tag{1c}$$

$$\textbf{prob}\,(\text{team } x \text{ beats team } y) = \textbf{prob}\left(Z > \frac{-a_x + a_y}{\sigma}\right) \tag{1d}$$

$$\textbf{prob}\,(\text{team } x \text{ beats team } y) = \textbf{prob}\left(Z < \frac{a_x - a_y}{\sigma}\right) \tag{1e}$$

$$\textbf{prob}\,(\text{team } x \text{ beats team } y) = \Phi\left(\frac{a_x - a_y}{\sigma}\right) \tag{1f}$$

$$\textbf{prob}\,(\text{team } x \text{ beats team } y) = \Phi\left(\frac{\alpha_i^T \mathbf{a}}{\sigma}\right) \tag{1g}$$

$$\textbf{prob}\left(z^{(i)} = 1\right) = \Phi\left(\frac{\alpha_i^T \mathbf{a}}{\sigma}\right) \tag{1h}$$

Likewise, because the standard normal distribution function has even symmetry and unit area,

$$\textbf{prob}(\text{team } y \text{ beats team } x) = \Phi\left(\frac{a_y - a_x}{\sigma}\right) \tag{2a}$$

$$\textbf{prob}\left(z^{(i)} = -1\right) = 1 - \Phi\left(\frac{\alpha_i^T \mathbf{a}}{\sigma}\right) \tag{2b}$$

Assuming that past results are independent, we define the likelihood of the outcomes $z^{(i)}$ to be

$$\textbf{prob}(z|a) = \prod_{i=1}^{n} \textbf{prob}\left(z^{(i)}\right) \tag{3a}$$

$$\textbf{prob}(z|a) = \prod_{i=1}^{n} \Phi\left(\frac{\alpha_i^T \mathbf{a}}{\sigma}\right) \tag{3b}$$

Instead of maximizing the *likelihood* over all possible team abilities $\mathbf{a}$, we can equivalently maximize the *log*-likelihood, which attains its optimum at the same location as the pure likelihood $\textbf{prob}(z|a)$ because the logarithm is monotonically increasing in its argument. Therefore, in order to convert our product into a more tractable summation, we maximize the log-likelihood function:

$$\log \textbf{prob}\,(z|a) = \log \prod_{i=1}^{n} \textbf{prob}\left(z^{(i)}\right) \tag{4a}$$

$$\log \textbf{prob}\,(z|a) = \sum_{i=1}^{n} \log \Phi\left(\frac{\alpha_i^T \mathbf{a}}{\sigma}\right) \tag{4b}$$

Because our individual Gaussian cumulative distributions monotonically increase in their arguments, the log-likelihood function attains its maximum value at the same team abilities **a** as the pure likelihood function. Furthermore, because the Gaussian cumulative distribution is log-concave, the log-likelihood function is a concave function. For proof of the Gaussian CDF's log-concavity, please read Stephen Boyd and Lieven Vandenberghe's *Convex Optimization*, paying particular attention to Example 3.42 on page 107. You can view the book at **http://www.stanford.edu/∼boyd/cvxbook/bv_cvxbook.pdf**.

To predict team abilities **a** with *maximum* likelihood, we *maximize* the likelihood function, which is equivalent to maximizing the log-likelihood function:

$$\boxed{\begin{array}{ll} \textbf{maximize} & \displaystyle\sum_{i=1}^{n} \log \Phi\left(\frac{\alpha_i^T \mathbf{a}}{\sigma}\right) \\ \textbf{subject to} & 0 \preceq \mathbf{a} \preceq 1 \end{array}}$$

Since the objective function of this problem (the log-likelihood function) is a sum of concave functions, its maximization subject to linear constraints is a convex optimization problem. We can solve this problem with a semidefinite program (SDP) solver, or a sparse solver, such as `cvx`. The solution to this convex optimization problem is our maximum likelihood estimate of team abilities, $\hat{\mathbf{a}}$.

6 Matlab Implementation

```
% NOTE: train contains results in form (x, y, x>y, xScore, yScore)
% Truncating the scores (using win/loss only):
train = train(:, 1:3);
m = sum(1:(nTeams-1));
n = nTeams;

% Forming the game incidence matrix A:
A1 = sparse(1:m, train(:,1), train(:,3), m, n);
A2 = sparse(1:m, train(:,2), -train(:,3), m, n);
A = A1 + A2;

% Predicting team ability from the problem:
cvx_begin;
variables a(n)
maximize (sum(log_normcdf(A * a / sigma), 2));
subject to
a >= 0;
a <= 1;
cvx_end;

teamAbility = a
```

7 Experimental Results from Real Sports Leagues

7.1 NBA Western Conference Team Abilities 2006-2007:

'Dallas Mavericks HOME'	[0.73895189860095]
'Phoenix Suns HOME'	[0.67189479138584]
'San Antonio Spurs HOME'	[0.67022007665423]
'San Antonio Spurs AWAY'	[0.61544162317452]
'Minnesota Timberwolves HOME'	[0.61292934589290]
'Utah Jazz HOME'	[0.58857760837800]
'Phoenix Suns AWAY'	[0.58590545144983]
'Houston Rockets HOME'	[0.56230413989584]
'Los Angeles Lakers HOME'	[0.56225980269953]
'Utah Jazz AWAY'	[0.56225980236008]
'Golden State Warriors HOME'	[0.53713756836931]
'Dallas Mavericks AWAY'	[0.53673317729809]
'Memphis Grizzlies HOME'	[0.48883239855274]
'Seattle Supersonics HOME'	[0.43981898145400]
'New Orleans Hornets HOME'	[0.43981898143543]
'Houston Rockets AWAY'	[0.43981898143400]
'Los Angeles Clippers HOME'	[0.43981898142633]
'Portland Trail Blazers HOME'	[0.43981898138172]
'Denver Nuggets HOME'	[0.43981898137830]
'Golden State Warriors AWAY'	[0.41832916427066]
'New Orleans Hornets AWAY'	[0.41504863563522]
'Denver Nuggets AWAY'	[0.41504863562618]
'Los Angeles Lakers AWAY'	[0.39078467616673]
'Sacramento Kings AWAY'	[0.38941735883707]
'Sacramento Kings HOME'	[0.38941735880315]
'Portland Trail Blazers AWAY'	[0.33686158861764]
'Memphis Grizzlies AWAY'	[0.31132722874134]
'Los Angeles Clippers AWAY'	[0.30836838141511]
'Seattle Supersonics AWAY'	[0.28054710795776]
'Minnesota Timberwolves AWAY'	[0.27782776765596]

COMMENTS:

- Expanding the field to 60 teams significantly retards convex optimization.
- Some results are nonsensical (ex: Certain AWAY teams being better than their corresponding HOME team).
- Rankings can apply to playoff matchups in isolated circumstances, but generally do not.
- Predicting results for the next team encounters works especially well (with accuracy above 75%), since the teams meet again within the same season.
- When we cross seasons, draft and free agency will initially derail rankings (Boston 2008).
- We might imagine translating abilities into next season's win/loss record.

7.2 NFL Regular Season Team Abilities 2003-2007:

'New England Patriots'	[0.74579624454565]
'Indianapolis Colts'	[0.68729107724874]
'San Diego Chargers'	[0.65983010576501]
'Pittsburgh Steelers'	[0.62670514985769]
'Baltimore Ravens'	[0.60591844284704]
'Jacksonville Jaguars'	[0.60481613939959]
'Philadelphia Eagles'	[0.58268768193079]
'Cincinnati Bengals'	[0.55897208193558]
'Denver Broncos'	[0.55865699832195]
'Seattle Seahawks'	[0.53615085752018]
'Carolina Panthers'	[0.51366900870979]
'St. Louis Rams'	[0.51366900854123]
'Chicago Bears'	[0.51366900839723]
'Kansas City Chiefs'	[0.46925261377860]
'Dallas Cowboys'	[0.46814891407345]
'Atlanta Falcons'	[0.46814891392106]
'Miami Dolphins'	[0.46814891385583]
'Buffalo Bills'	[0.46814891381156]
'Minnesota Vikings'	[0.44520532967604]
'New Orleans Saints'	[0.44520532966594]
'New York Giants'	[0.44520532966176]
'Tennessee Titans'	[0.42257148235246]
'Green Bay Packers'	[0.42194659916275]
'New York Jets'	[0.39827971782327]
'Washington Redskins'	[0.37940044695046]
'Tampa Bay Buccaneers'	[0.37430185034614]
'Cleveland Browns'	[0.37430185014708]
'Arizona Cardinals'	[0.34959930908343]
'Oakland Raiders'	[0.32622136734782]
'Detroit Lions'	[0.32528020000854]
'San Francisco 49ers'	[0.30123495899457]
'Houston Texans'	[0.29984569201820]

COMMENTS:

- If we separate HOME and AWAY, then we must delve farther back in time to obtain a full round of previous matchups.
- High turnover (free agency) renders older data increasingly misleading; the more we try to isolate factors like home field advantage, the more irrelevant data we must use.
- 2007-2008 season: If we simply use the result of the previous meeting to predict future results, then we are correct only 53.08% of the time.
- 2007-2008 season: If we bet on the team with higher ML team ability, then we are correct 58.77% of the time.
- Accuracy rapidly improves as we adaptively replace old data with ongoing (current) season data; these more recent results account for free agency and draft.

8 Conclusion

What can we see from the numbers themselves? It is hard to distinguish mediocrity. Several teams in the league share very similar team abilities, especially in the NBA, where the amount of game-to-game fluctuation is extremely high. Perhaps, for such leagues, a higher noise variance σ^2 would have been more appropriate. Nevertheless, even with an infinitely fine scale, some teams are hard to distinguish simply because they beat the best teams one night and lose to the worst teams the next night. One might think that more games yield a better picture, but, sometimes, more games - especially when taken from the same year - only obfuscate the truth even more, if only because upsets happen on both sides; the best teams still lose to the worst teams, sometimes for no obvious reason.

What happens if we utilize fewer results (a lower m)? If we do not account for every possible matchup – if we do not include one game for every possible pair of teams – then the results skew slightly. Missing one or two games does not noticeably affect the output, but it is hard to argue the benefit of omitting matchups. Nobody can say with certainty if the omitted match is the telling tale of a team's success, especially if the match pits two of the top teams in the league. Omitting upsets might arguably improve the results, but these omissions constitute selection bias, which we want to avoid in a mathematical method like maximum likelihood.

What happens if we utilize more results (a higher m)? If using more results means taking results from the more distant past, then the team abilities arising from the augmented match data will be a less accurate reflection of the *current* team abilities. If, however, we are using data from the same season, and even using more recent results (for duplicate intradivision matchups, for example), then we can be more confident in our team abilities. For consistency, the results prevented in the previous section use only the most recent matchups between each pair of teams, including playoff meetings.

How does prediction from team ability compare to prediction from extrapolation of past results? Prediction from team ability performs unequivocally better, if only because league-wide results have been taken into account. When we use team abilities to predict the outcome of a future matchup **within the same season**, the proportion of correct predictions rises from $\approx 63\%$ to $\approx 76\%$ in the NBA. When we use team abilities to predict the outcome of a future matchup **in the following season**, the proportion of correct predictions rises from $\approx 58\%$ to $\approx 64\%$ in the NBA, and from $\approx 53\%$ to $\approx 59\%$ in the NFL. In other words, free agency defections and additions likely render cross-seasonal predictions more difficult and less accurate. For games like football, in which some teams meet only once every four years, both team ability and extrapolation of previous results are essentially crap shoots. Fortunately, we can employ adaptive processing to help assuage some of the early-season inaccuracies; instead of using results from past seasons, we can adaptively update the game incidence matrix after every new game played. For the NFL, this adaptive update increases the proportion of correct predictions from $\approx 53\%$ to $\approx 66\%$, meaning that the team abilities are much more accurate reflections of true ability when some of the incorporated past is very recent. In the age of free agency, team abilities are much more dynamic and transitory, constantly in flux due to player movement, so their accuracy is much more pronounced within season.

What happens if we use team abilities to estimate each team"s win/loss record in the following season? Unless we employ adaptive processing as detailed above, the results are typically off by an average of 5-10 wins in the NBA and 2-5 wins in the NFL. So much fluctuation occurs between seasons that teams often look entirely different; just consider the Boston Celtics of 2007-2008, one of the worst teams in the league, and juxtapose them with the Boston Celtics of 2008-2009, now the best team in the league with Kevin Garnett, Ray Allen, and Paul Pierce. Adaptive processing helps us by replacing the wrong (or misleading) results from the past season with a more recent showing that better reflects the current team's talents.

Why did we not split NFL team abilities into HOME and AWAY? Teams play very differently in front of their home crowds than they do in another team's stadium. In fact, some teams, such as the Sacramento Kings and Philadelphia Eagles, have traditionally enjoyed notoriously supportive home crowds. Unfortunately, in the NFL, teams do not meet frequently enough to justify this separation of HOME and AWAY; if we made this distinction, then we would need to delve more than 10 years back in time to find adequate results, and these antiquated results would prove only more misleading since teams often change quarterbacks and coaches several times within that period. The more feasible option would be ranking the two conferences (AFC and NFC) separately, since the styles of football differ immensely, and cross-conference matchups are the prime reasoning behind use of antiquated results; if we restricted rankings in the two conferences to separate scales with no consideration of the infrequent interconference matchups, then we would not need to travel more than two years back in time to fill the game incidence matrix satisfactorily. The results would likely better reflect the current team abilities within conference, but the drawback of this method would be the infeasibility of cross-conference comparisons, such as the one we might want to make between the New England Patriots and New York Giants.

Why did we split the NBA into two conferences when cross-conference results are plentiful? If we make our game incidence matrix too large (on the order of 60-70 teams), then convex optimization problem solvers consume gargantuan amounts of time and memory. For the scope of this project and the size of the machines used to perform the computations, these larger problems simply were not feasible.

How might we find a past result for a single team's HOME vs the same team's AWAY? For self matchups, the most logical comparison would be average score. If the arithmetic mean of a team's HOME scores is higher than the arithmetic mean of a team's AWAY scores, then the HOME version of the team is likely superior to the AWAY version. Of course, lower scores might simply reflect a differing style of play on road games (such as a more defensive, less risky approach), but, in general, this average score comparison is sound.

How might we find a past result for a HOME-vs-HOME or AWAY-vs-AWAY? This is a bit trickier. To compare how the Dallas Cowboys HOME might perform against the Washington Redskins HOME, for example, we have no previous account of such a matchup. Instead, we use the *scores* from the team's extant matchups (Dallas Cowboys HOME vs Washington Redskins AWAY, and Dallas Cowboys AWAY vs Washington Redskins HOME) to determine how the teams might perform in the corresponding venues. We then compare these scores to determine the probable victor in the game incidence matrix. For example, if Dallas HOME beat Washington AWAY 21-17, and Washington HOME defeated Dallas AWAY 14-7, then we say that Dallas HOME beat Washington HOME 21-14, and Washington AWAY beat Dallas AWAY 17-7.

9 Acknowledgments

Many thanks to Stephen Boyd's textbook *Convex Optimization* and Stanford graduate student Argyrios Zymnis for inspiring the use of convex optimization to solve the team ability problem. This formulation would not have been possible without their guidance.