

Bayesian Skill Ranking

Leland Chen, Joseph Huang, Ryan Thompson

March 13, 2011

1 Introduction

Template

2 Related Work

Let's talk about [1] and [2] and then just follow the citations that those guys make to introduce the general challenges and state-of-the art.

Basketball brings two key challenges to the field of Bayesian Skill estimation.

- Results are not binary Win/Lose outcomes.
- Results are influenced by the skill differential of more than two players.

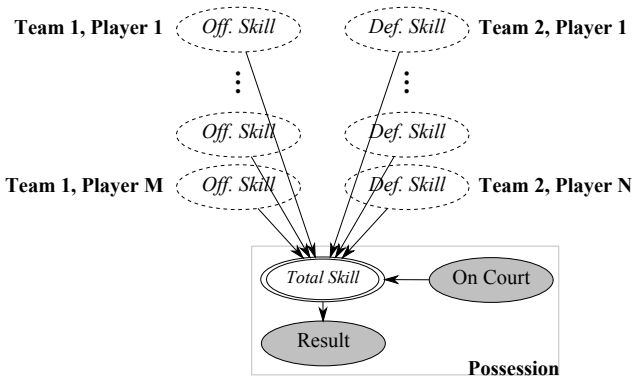
3 Method

3.1 Traditional Networks

We model the result of each possession as an independent and identically distributed random variable, *Result*. An NBA roster has 12 players, but only five of them are on the court for a given team during any single possession.

To reduce noise, we ignore “garbage time” possessions. To simplify the model we ignore possessions that end in a time-out.

We begin with a very basic model that follows traditional “skill ranking” in the sense that each player has an unobserved (or parameterized) skill that has some distribution, and using inference (or parameter estimation) we can determine the value of each player's skill. In this network, we decided each player would have a hidden offensive and defensive skill value, shared across possessions:



These skills would contribute deterministically in some way to an “effective” total skill differential between the two teams, and then the *Result* variable would be one of four outcomes:

- $R = 0$ Offensive Team Scores nothing, change of possession (e.g. turnover, defensive rebound, etc.)
- $R = 1$ Offensive Team Scores 1 point

- $R = 2$ Offensive Team Scores 2 points
- $R = 3$ Offensive Team Scores 3 points

The *On Court* random variable “multiplexes” between those players that are on the court and those that are not.

3.2 Issues

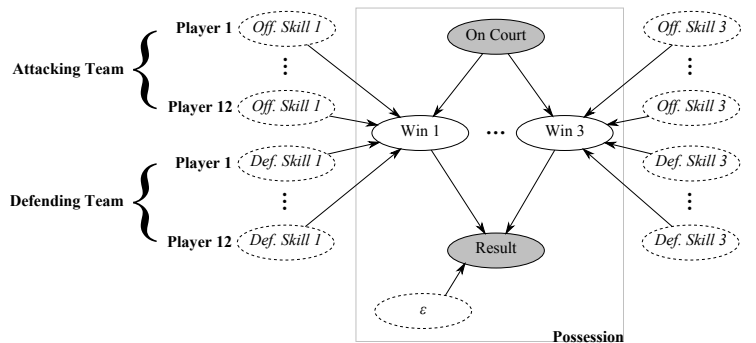
These traditional models [1] have difficulty capturing the proper causalities when Results can have multinomial-valued outcomes. For example, regardless of the parameterization of the *Results* CPD, both $\Pr\{R = 2\}$ and $\Pr\{R = 3\}$ would depend on the same skills of the same players. The relative distribution of outcomes $\Pr\{R = 2\}$ vs. $\Pr\{R = 3\}$ would be shared across all “units” (i.e. all combinations of *On Court* assignments).

In reality, a team that scores $R = 3$ half the time and $R = 1$ half the time is just as good as a team that scores $R = 2$ all the time. However, any traditional Win/Lose model will unfairly penalize the likelihood of one of these teams over the other and leads to under-fitting. This might suggest that we have a random variable that represents $E[\text{pointsscored}]$, but this doesn't pass the clarity test.

Secondly, there is a lot of value in being able to compare with the state-of-the-art in the Win/Lose based Bayesian Skill Ranking literature. Specifically, there is common debate over logistic vs. Gaussian skill/performance distributions and we wish to be sensitive to that conversation in this project. Having a multinomial outcome makes it difficult to directly compare logistic vs. Gaussian in the standard skill/performance framework because there is no consensus in the literature about how to extend these models just to support ties [3], let alone general multinomial outcomes.

3.3 Proposed Network

To address some of the shortcomings discussed in **Section 3.2**, we propose the following:



In this network, each player is represented by three skill parameters:

1. Skill 1: Ability to score (or defend against) one-point opportunities

2. Skill 2: Ability to score (or defend against) two-point opportunities
3. Skill 3: Ability to score (or defend against) three-point opportunities

On each possession, there are three binary hidden random variables:

1. *Win 1*: True if there was an opportunity to score one point during the possession.
2. *Win 2*: True if there was an opportunity to score two points during the possession.
3. *Win 3*: True if there was an opportunity to score three points during the possession.

Now, *Result* simply selects the highest-valued scoring opportunity available during the possession. If *Win 3* is true, *Result* is 3; If *Win 3* is false but *Win 2* is true, *Result* is 2; etc.

Result can be specified with either a tree-CPD or table-CPD, but it has the following probabilities.

	$R = 0$	$R = 1$	$R = 2$	$R = 3$
$w_3^1 w_2^1 w_1^1$	ε	ε	ε	$(1 - 3\varepsilon)$

The single parameter ε is inspired by the noisy-max and essentially indicates unmodelled errors (e.g. offensive or defensive mistakes). The better our model fits the actual flow of the game, the smaller ε should be.

4 Implementation

Summary

5 Results

Data

6 Analysis

Details

7 Conclusions

Done

8 References

References

- [1] R. Herbrich, T. Minka, and T. Graepel, “TrueSkillTM: A Bayesian skill rating system,” *Advances in Neural Information Processing Systems*, vol. 20, pp. 569–576, 2007.
- [2] R. Coulom, “Whole-history rating: A bayesian rating system for players of time-varying strength,” *Computers and Games*, pp. 113–124, 2008.
- [3] D. Hunter, “MM algorithms for generalized Bradley-Terry models,” *The Annals of Statistics*, vol. 32, no. 1, pp. 384–406, 2004.