Eras of the NBA

Nikhil J. Nathwani

September 7, 2014

1 Goal

The goal is to delineate the distinct eras of each NBA franchise. Taking a player-centric approach, this is equivalent to determining which players define the history of each franchise. I've chosen to quantify a player's worth based on his win share* percentage† (WS%), because a player with a large percentage of his team's win shares in a given season is likely to be crucial to that era of the franchise. Other metrics could be used to gain different perspectives on a team's history.

The method used to select these "franchise-definers" should keep the number of chosen players to a minimum, in order to weed out those who were not true flagships of any particular era. But of course, the obvious superstar players should not be excluded either.

2 Method

Consider an arbitrary team X. We would like to determine the set of franchise-defining players for X. Let:

P_j = the set of players on team X in season j.

$$w_j(p) = \begin{cases} \text{WS\% of player } p \text{ in season } j & p \in P_j \\ 0 & \text{otherwise} \end{cases}$$

As a preprocessing step, I use the following algorithm to exclude a number of players that clearly did not have a high enough impact:

*A measure that tries to quantify how many regular-season wins an individual accounted for. Complete explanation here.

†Simply calculated by dividing an individual's win shares by the sum of the win shares of all members of his team.

Algorithm 1: `Preprocess`(T)

```
for each season j do
    P* := ∅;
    totalWS := 0;
    while totalWS < T do
        p := argmax_{p∈P_j} {w_j(p)};
        totalWS ← totalWS + w_j(p);
        P_j ← P_j \ {p};
        P* ← P* ∪ {p};
    end
    P_j ← P*;
end
```

In words, this retains the win share leaders of team X each season, but only keeps as many as are needed to cover $T\%$ of the total win shares.

To make the problem more tractable and interesting, I impose the following constraint:

$$\forall_j \sum_{p \in P} w_j(p) \geq t \tag{1}$$

where t is a tweakable parameter. Ensuring that $t\%$ of the total win shares are accounted for each season allows every season to be well-represented in team X's history.

What we're now left with is a covering problem: we want to satisfy the constraints specifed by inequality (1) while minimizing the total number of players used. Since were dealing with decision variables (i.e. 0,1-valued variables that specify whether or not a player is used), we can formulate this problem as an integer program (IP). But before diving into the IP formulation, I want to define the decision variables in a way that is more sensible for this real-world scenario:

If we select player p in season j, and player p is also available to be selected in season $j + 1$, then it is befitting to select player p in season $j + 1$. This is because we know player p was valuable to the team in both of those years (because of the `Preprocess` algorithm), and being valuable for multiple consecutive seasons certainly warrants inclusion in team X's historical record.

So, let's define a *reign* to be a maximal-length sequence of consecutive seasons in which a specific player is in the preprocessed seasonal player sets. E.g. if player p is in P_0, P_1, P_2, P_4, and P_5, then $\{0, 1, 2\}$ and $\{4, 5\}$ are reigns. Now instead of selecting players, we'll select reigns, thereby selecting the player corresponding to that reign for each season of that reign's duration.

3 IP-Formulation

I'll index the reigns numerically[‡] so that our decision variables are:

$$r_i = \begin{cases} 1 & \text{reign } i \text{ is selected} \\ 0 & \text{otherwise} \end{cases}$$

Let W be a matrix with entries defined as follows:

[‡]In the implementation there is a mapping from reigns to the players they correspond to, but this is irrelevant when discussing the mathematical reasoning.

$$W_{ij} = \begin{cases} \text{WS\% of player } p \text{ corresp. to reign } i \text{ during season } j & p \in P_j \\ 0 & \text{otherwise} \end{cases}$$

Our problem has been reduced to the following binary integer program, where $\vec{r}$ is the vector of decision variables r_i and $\vec{t}$ a column vector (with length equal to the number of seasons) where each entry is t from inequality (1):

$$\begin{gathered} \text{min. } \|\vec{r}\|_1 \\ \text{s.t. } W\vec{r} \geq \vec{t} \\ r_i \in \{0, 1\} \ \forall_i \end{gathered}$$

This IP can be solved using branch-and-cut methods, for which there is a handy library called GLPK. The solution to this IP is a binary vector denoting which reigns have been selected. Mapping the selected reigns back to the players the correspond to gives us our desired set of players to represent team X's history.