

Data Mining in the NBA

In any NBA game – particularly a playoff game -- the end goal of both teams is quite simple: to win. To do this, the coach of a team needs to recognize and exploit any advantages his team might have over their opponent. In 1995, IBM developed a tool to aid coaches in this endeavor. Advanced Scout (AS) is an application which analyzes collected data from previous basketball games played by a team, and compiles a list of “interesting events” that occurred during the game. The premise here is that once the coach is made aware of the events, he can then examine them more closely by looking at videos, and finally interpret the events, incorporating this additional information into future games.

The data collection involves recording a variety of statistical information for every possession that occurs throughout a basketball game. The sub-events recorded include everything from who brought the ball up the court, to who shot the ball, to whether the shot resulted in points for the team (and if not, whether possession was regained through an offensive rebound, foul call, etc).

Normally, it would be very taxing for a coach to have to sit on the sidelines, and simply jot down notes about what he noticed during a game. With AC, however, the coach can simply query the database to receive statistically-significant information. For example, AC would notice if one player suddenly scored many more points in a certain time-span, or this player's shot percentage suddenly went up – essentially, any event which caused a deviation from the team's regular performance, and that was not deemed to be only a chance occurrence.

Advanced Scout was made available to all the NBA teams that year, and was quickly taken advantage of by roughly half of them, one of them being the Orlando Magic, whose assistant coach was then Tom Sterner. By the end of that season, he was able to provide valuable insight and evidence regarding the performance of the application. Although the coaches were generally open to the idea of adding this extra element to their coaching procedures, some were (unsurprisingly) cautious in using it.

In the basketball world, it is very common to analyze player and team statistics, almost to a fault. Many people are critical of overly basing decisions based on statistics, simply because, from experience, they know that oftentimes there are other elements that need to be considered. For example, just because a player is averaging 20 points and 10 rebounds per game does not mean that adding him to a team's roster will improve that team's chances to do well in

the playoffs. Moreover, another player who is only averaging 10 points and 5 rebounds is not necessarily less beneficial to the team than the statistically better player.

There are several reasons behind this. For one thing, certain statistics can be inflated. If a weak team only has one tall center, and consistently plays him, he is likely to garner more rebounds than a better center on a strong team, playing alongside one or two other quality rebounders. Secondly, the common set of statistics usually provided is very incomplete. They usually do not, for example, take into account how a player affects the performance of his teammates. Some players can motivate their team to play at a higher level, whereas others are openly hostile against some of their teammates.

Nevertheless, many coaches were willing to give AS a try, and in a relatively quick period of time, they realized just how beneficial it could be. Sterner described the impact it had when Orlando played against Miami in the playoffs: "We were beaten so badly in games 1 and 2 at Miami - 99-64 in Game 1 and 104-87 in Game 2. After Game 2, we ran our analysis and watched the tape to determine what happened. Advanced Scout gave us some interesting pieces of data."

AS indicated that in the backcourt, playing Penny Hardaway and little-used reserve point guard Darrell Armstrong provided a significant boost to the team's performance, relative to Miami's. This had clearly gone unnoticed by the coaching staff, until it was pointed out by the application. Sterner recalls the next game: "with about six or seven minutes left to go in the half, we were down by 20 points again. We brought Darrell into the game, and by the end of the half we were tied, 42-42. By the end of the game, we had won, 88-75. Penny Hardaway had scored 42 points, and Armstrong scored 21. Most importantly, we avoided elimination." Orlando went on to win the playoff series, and AS was fully incorporated into the coaching routines.

The application clearly isn't a replacement for a coach. It is certainly an additional method of gaining information, but in the end, it is still the coach who must interpret the results and make the decisions. It still remains to be seen just how accurate AS, or any other program of this type, can be, relative to other methods coaches use. This will become clearer the more the program is refined, and the more past games are included in the database. Eventually, it would be interesting to observe how the suggestions of the application compare with a coach's "gut feeling," and whether or not the two strategies complement each other. As with other statistics, this type of software won't be able to predict anything with 100% certainty, but it is also likely that it will often be able to spot certain elements of a game that might go unnoticed by the coaching staff.

References:

Francett, Barbara. *The NBA gets a jump on Data Mining*. Software Magazine, 52 (09/1997).

Bhandari, Inderpal. *Advanced Scout: Data Mining and Knowledge Discovery in NBA Data*. Data Mining and Knowledge Discovery, 1, 121-125 (1997).