Researchers dive into big data to predict NBA winners

May 2 2016, by Lorena Anderson

UC Merced professor and applied mathematician Harish Bhat. Credit: UC Merced

Before you place any bets on the next NBA playoff games, take a look at UC Merced professor Harish Bhat's Twitter account.

He's tweeting daily predictions based on an innovative new [algorithm](link), and so far this postseason, he's right about 80 percent of the time.

In about 10 minutes a day, Bhat and two of his students run simulations for each team in each day's matchup between 10,000 and 100,000 times before tweeting their predictions. The calculations take about 10 minutes.

"We could simulate the entire season 10,000 times in about an hour," said Bhat, an applied mathematician with the School of Natural Sciences.

The algorithm they've designed takes a microscopic look at big data surrounding each team in the NBA playoffs—including analyzing the transcripts of each play-by-play file from all 1,230 regular season games—to predict who will win each day.

Model accounts for lineup changes

"We're just in the early stages, but it's fun to see how accurate we can make our statistic models," Bhat said. "Other analyses are taking a macro view, ranking the teams and calculating how well each one performs in a certain set of categories."

Bhat's dynamic, stochastic algorithm is different from traditional predictive models, which do not account for substitutions and focus mainly on offense. Bhat's model—which is available publicly to anyone who wants to use it—looks at fine details.

It has learned, for example, when and how to make player substitutions during game simulations, including the various lineups the coaches tend to use throughout games. It models for the average plus/minus rate of each unit, and by putting both elements together, simulates separate

game trajectories for each home and visiting team. Whichever has the higher final score wins.

The substitutions are important because "if you know who's playing at any given moment, you should be able to predict how many points they will contribute to the score, either through offense or defense," Bhat said.

The model can answer "what if" questions, as well, such as: What if Chris Paul and Blake Griffin cannot play for the Los Angeles Clippers? How does that impact the Clippers' probability of beating the Portland Trailblazers?

"With our algorithm, it's easy to put in a couple of 'stop signs,' so to speak, to ensure that certain players never start a game or are never substituted into the game," Bhat said.

Driven by student collaboration

Bhat and his students are using the same techniques they would use in complex scientific calculations, and are constantly improving the modeling. Each model has some elements of randomness figured in to account for variations in substitution patterns, and the researchers plan to add more variables, including fouls, which should hone the predictions even further.

Bhat has already published one paper on this method of big-data analysis, entitled "Learning Stochastic Models for Basketball Substitutions from Play-by-Play Data," and anticipates more in the future as the algorithm is improved. He said he and his students—Sebastian Rodriguez, an undergraduate from Costa Rica, and Li-Hsuan Huang, a graduate student from Southern California—have had to be quite creative in how they've developed the algorithm and

gathered data.

"The students have put in tons of research into how to get the data off various public sites and put it into a form that can be used with this cutting-edge modeling," Bhat said. "You can't just use some off-the-shelf algorithm and do what we're doing."

Provided by University of California - Merced

Citation: Researchers dive into big data to predict NBA winners (2016, May 2) retrieved 3 May 2024 from https://phys.org/news/2016-05-big-nba-winners.html