Student Academic Conference

Computer Models and Predictions in Baseball

Noah Vetrone
noah.vetrone@go.minnesota.edu

Follow this and additional works at: https://red.mnstate.edu/sac

Researchers wishing to request an accessible version of this PDF may complete this form.

Noah Vetrone

CSIS 492

Final Paper

5/2/2022

Computer Models and Predictions in Baseball

Computer modeling and projections are becoming a part of every industry, and it is no different in professional sports, especially baseball. Teams are relying on these systems now more than ever and are involved in an arms race of trying to acquire the best technology to win more games. The road to get here hasn't been easy. This report goes over a brief history of computer modeling's development in professional sports, before going over the technical aspects of how these models work, looking and them from a Computer Science background. Analysis of different types of projection models that include Statcast and PITCHf/x. These models extend to other aspects of computing such as A.I. and machine learning. Will also look at the current development of computer projections and how it could change the future of baseball.

The history of professional team using computer projection models goes as far back as the beginning of the personal computer revolution, and in some cases even further. In the NFL new general manager of the Dallas Cowboys; Hank Schram employed a system for drafting players using a computer model that was created with the help of IBM engineers, with this system The Cowboys became a dynasty of the 1970's and won five NFC Titles and two Super

Bowls. The system didn't spread to other teams however as computers at the time were expensive and took up a whole room. As far as baseball was concerned, we look again to IBM. John Bergeson a computer programmer working for IBM created a rudimentary baseball simulation. It wasn't so much a projection model or a tool, but it laid much of the groundwork for what would be to come. It wasn't until the late 90's and early 2000's however that that the ubiquitous use of these systems in front offices of baseball teams became a thing. One such use is with the Oakland Athletics and their General Manager Bill Beane who needed a better way to build his team [1]. He hired an analytics-oriented staff and revolutionized the way teams sign, scout, and draft players. Today almost every MLB team employs a computer scouting department of some kind.

Baseball projection models are computer models that predict various aspects of baseball, whether it be a team's record or a player's success all these things are encompassed in baseball projection systems. So how do these projection systems work, there are two aspects. The first part is ascribing value or figuring out what leads to wins or success on the baseball diamond. Using baseballs ever increasing database of statistics dating as far back as 1871 we can see what statistics and aspects lead to not only successful players but successful teams as well [5]. With this research we have determined what leads to the most success. An example of this is a new stat has been created called WAR (Wins Above Replacement), This stat directly states how much value an individual player is worth to their team. The second aspect is trying to project performance to future results, this is the most difficult part to try and do. It is almost impossible to do this without the speed and computing power of computers. How do we predict a player's individual players future statistics, the answer is like how we found value

above? By using things like a player's previous seasons while also considering their age we can come up with a basic projection of a players upcoming season. Actual projections are much more complicated, but it is similar. You add up all the players projections into a team's model and you can produce a projection of a team's season [3]. That is just a basic understanding of value and modeling. There are hundreds of different models and projections created by various companies and organizations, some teams even created their own model. Each model is different, one model might change how things are valued or how they project a player's future seasons. As a result of this models very regarding accuracy. We must understand however that 100% accuracy is not something that these models are trying to do, instead these models are supposed to give people an understanding of how talented a team is supposed to be or a general understanding of how a player's season is going to be. If a model projects the twins to win 86 games this year but instead, they only win 82 that doesn't mean the model was wrong. It could mean the opposite. Depending on how the other 29 teams did as well. There are thousands of factors that go into the creation of projection systems and things are always changing.

This technology has completely changed the way the front offices build baseball teams, players practice, and fans watch the game, whether you know it or not the game you are watching out on the field has been influenced by the use of computer-based projection systems. In terms of the MLB front office the change couldn't be greater. Almost every Front office uses these systems in some way, and for many it is the most important tool that they have. It gives general managers a better understanding of the strengths and weaknesses of their team, for instance if the model says they have a weak bullpen it would make sense to try

and sign a relief pitcher, and because the model also projects players if can give them a better idea of who the best relief pitchers are on the market. Front Offices most likely already know the weaknesses of their teams and who the best free agents are, but these tools help them to discover hidden value or data that they otherwise might have not seen, as a result it also helps players whose skills aren't as obvious as others. The most tools you have in your arsenal the better equipped you will be to handle any task. Projection models are just another tool scouts, players, and General Managers have at their disposal. That doesn't mean we completely discount traditional ways of scouting and building teams. There is a lot of talk about people who see this as taking the human element out of baseball and having a computer building teams and drafting players but that couldn't be further from the case, people are the ones how created the models and are the ones who have the final decision. It makes zero sense to ignore something that helps your favorite teams win games, the goal is to win as many games as possible and this is something that greatly increases your chances, especially if some teams are slow to embrace new technology.

While projection models are a great way of analyzing and building baseball teams they are not without their flaws. The major disadvantage comes from the issue of trying to predict human behavior and achievement. It's not like predicting a coin flip or a roll of a dice. A projection system might project Player A to have a good season but if Player A's is injured and because of that he has a bad season or if Player B is taking Performance enhancing drugs with gives him a season that the projection model never predicted the problems start adding up. These models can only make projections with the information that is given to them that is quantifiable and trying to predict a human being behavior on the microlevel is hard to do. The

examples given were extreme cases of that but something even as small of Player A waking up with a headache and having a bad game is something that computer models cannot predict. To be far to the model it is not something it is supposed to, or we want it to predict, that would just make it even more accurate than it already is. The human element doesn't just come from the players side but also from the people creating the model as well. People creating these models must put a value on certain sets of data, that later through more research might turn out to have more value in baseball than previously thought overwise [6]. An example of that is s stat applied to catchers called Pitch Framing, 10 years ago this wasn't as valued as it should have been, as a result certain catchers were perceived to be less valuable, and because of the devaluing the projections were inaccurate.

While the previously discussed systems are only used by font offices and fans, the next emerging piece of technology is being used by coaching staffs and players and it is revolutionizing the way baseball is being taught. Full Motion tracking systems use technology to analyze a player's body movement. With this analysis we can track the speed of a pitchers Fastball, the break of a curveball, the efficiency of his pitching motion. This combination of Computer Science and Kinesiology can help improve a player's game. Various Tracking systems like Trackman, Rapsodo, and MLB's own system called Statcast popularity has exploded over the last 5 years with every team using some form of the technology in their training facilities or ballpark [4]. Related to this technology is Virtual Reality, which is being talked about as the next revolution in baseballs ever revolution arms race to have the best technology to win games. In the next few years players could be preparing against a starting pitcher by taking pitches against him in V.R.

Works Cited

[1] M. Lewis, *Moneyball*. New York: W.W. Norton, 2011.

[2] M. Maccambridge, in *America's game The epic story of how pro football captured a nation*, New York: Anchor, 2005. Chpt 5

[3] T. Sawchik, *Big Data Baseball: Math, miracles, and the end of a 20-year losing streak*. New York: Flatiron Books, 2016.

[4] J. Dyson, "The oldest-known computer baseball simulation," *The Strong National Museum of Play*, 26-Oct-2021. [Online]. Available: https://www.museumofplay.org/2014/06/20/the-oldest-known-computer-baseball-simulation/.

[5] J. Boice, "How Our MLB Predictions Work," *FiveThirtyEight*, 28-Mar-2018. [Online]. Available: https://fivethirtyeight.com/methodology/how-our-mlb-predictions-work/.

[6] G. Nourse, "A machine learning algorithm for predicting outcomes of MLB games," *Medium*, 08-Oct-2021. [Online]. Available: https://towardsdatascience.com/a-machine-learning-algorithm-for-predicting-outcomes-of-mlb-games-fa17710f3c04.

[7] D. Avevados," Simulation-Based Projections for Baseball Statistics" *CSPU,* 2018

[8] "MLB stats, scores, History, & Records," *Baseball*. [Online]. Available: https://www.baseball-reference.com/. [Accessed: 07-Feb-2022].

[9] G. Healey, "The New Moneyball: How Ballpark Sensors Are Changing Baseball," in *Proceedings of the IEEE*, vol. 105, no. 11, pp. 1999-2002, Nov. 2017, doi: 10.1109/JPROC.2017.2756740.

[10] A. Nathan. (Oct. 21, 2012). Determining Pitch Movement From PITCHf/x Data. [Online]. Available: http://baseball. physics.illinois.edu/Movement.pdf