Using machine learning to assess and compare athletes in team sports

Albrecht Zimmermann

Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC, 14000 Caen, France,
albrecht.zimmermann@unicaen.fr

Abstract. Using machine learning techniques (ML) to assess action (and derived from it player) quality is a recent and promising alternative to expert-based assessments. The big challenges for these approaches consist of data acquisition, data transformation and augmentation, which often require in-depth knowledge of the sport in question, as well as understanding which ML techniques are well-suited not only for final modeling but also for data preparation.

Keywords: sports analytics, player assessment, machine learning

1 The case for assessing players algorithmically

It is well-understood that team sports, in most countries and most leagues, is a business, often big business. As such, significant amounts of money are involved in player acquisition and player compensation, and teams employ a number of people whose only job it is to answer questions such as:

- Is a particular player worth a large contract?
- Can we find a player to replace one that is currently on a roster? Especially if they are a star player about to leave around whom much of the rest of the roster is constructed. And ideally at a lower salary than that player received.
- Can we find a hidden gem, i.e. a player who is underpaid relative to their performance?

The right answers to those questions free up money that can be invested elsewhere to improve the overall quality of the team. After player acquisition, player assessment does not stop and questions such as:

- What actions does a player perform better (worse) than the average player?
- Does a player make good decisions under pressure (and execute them well)?

can help with training and substitution management.

The question whether someone is a good player or not is easy to answer on a high level: a player who does things that help the team win is a good player, one who does things detrimental to winning is not. Identifying those "things", or rather actions, is already somewhat harder, however: scoring obviously helps, as does putting teammates in the position to score, turning the ball over does not.

But what about actions further in the past, such as the pass that occurred three seconds before the final assist? Or ones that prevent dangerous situations from even developing, such as the presence of a defender that discourages attackers from moving into spaces from which they could attempt so score?

Traditionally[1], the assessment of player actions is done by human experts but this approach comes with a number of pitfalls:

- Developing expertise requires time during which the game changes, meaning that experts might be biased w.r.t. the actions they consider valuable.
- Having experts score actions is time-consuming: a soccer match lasts 90 minutes, with several actions occurring per minute. For an expert to reliably assess them, they would probably have to go back and forth to understand context, making the scoring procedure longer than those 90 minutes.
- There are relatively few good experts around, making them expensive.

An recent alternative is learning to score actions automatically: starting from large amounts of data, ML algorithms learn models that allow to assign a numerical value to each action performed during a match. Once such a model is available, it can be used to score all actions performed by players during a match, a season, or their career, and compare their performance to that of the average player in their league, for a particular action type or aggregated over all of them, to put a number on how good (or not) that player is.

1.1 Data

A quick word about data: to learn ML models, learners need large amounts of data in which each action of interest is labeled with a ground truth label. This label can either be derived from expert assessments, who now do not need to score all actions but only a large enough amount, or by back-propagating positive or negative outcomes in time (such as a goal scored or conceded).

There are two main types of data used in the literature:

- Event stream data, which contain all events that involve the ball or puck, identified by human annotators based on video recordings of matches.
- Player tracking data, which contain the positions of all players, as well as the ball or puck, sampled at frequencies between 10Hz and 25Hz.

The former requires the use of human experts with all the constraints this entails, and zooms in on a small subset of actions, but carries the semantic information necessary to understand what happened. Acquisition of the latter can be largely automated, and off-ball data offers information that tracking data is missing, but semantic meaning is absent.

[1] A tradition, the systematic assessment of player actions, that is not that old to start with.

2 Some examples of action/player assessment

In this section, I will quickly review a handful of papers that perform ML-based action and player assessment. This overview is far from exhaustive and necessarily quickly out-of-date but should give an idea of the complexity of the steps involved.

2.1 "Not all passes are created equal"

The authors of [5] begin with the observation that passes are by far the most-often performed action in soccer matches and that all passes carry both an inherent risk – in terms of their success probability – and reward – in, terms of their leading to a shot attempt in the following ten seconds. The main work of the paper consists of data transformation and augmentation, such as deriving player speed, defender distance and angle, or using a variety of clustering techniques (e.g. [1]) to identify team formations on the pitch. Based on these augmented data, the authors learn logistic regression models for risk and reward and propose different metrics to identify the most valuable pass in a sequence, players that systematically successfully perform high-risk passes, or receivers that often corral high-quality passes.

2.2 "Actions speak louder than goals"

The authors of [3] consider not only passes but a larger set of actions, which move the match from "game state" to game state. Each state is labeled with whether a goal was conceded/scored in the next ten actions, and two Gradient Boosted Tree Models are learned to model the scoring/conceding probabilities for a given game state. The change to these probabilities when going from one state to the next is the score assigned to the action that caused the transition. The resulting VAEP (Valuing Actions by Estimating Probabilities) values, summed over a given time frame, and normalized for 90 minutes, allow to quantify the overall performance of players, zoom into particular action types, and compare players to each other.

2.3 "Choke or shine"

In [2], the research question is whether players make good decisions under pressure, and whether they execute the chosen actions well. To this end, the authors need to model game pressure, which they split into pre-game pressure, based on expert labels, season-level statistics, clustering, and monte carlo simulations, learned using a Gradient Boosted Tree model, and in-game pressure, learned using a Automatic Differentiation Variational Inference algorithm. They use the VAEP values from their preceding work to assess the quality of chosen actions, and combine them with probability models to find that, e.g., Neymar makes worse decisions under pressure even if his execution remains high-level.

2.4 "Deephoops"

The focus of [7] is on "micro-actions" in basketball, e.g. screens or cuts to the basket. After transforming their player tracking data into "moments" containing all players' positions expressed in polar coordinates, and splitting the match into sequences of 128 moments, the authors learn a neural network model based on Long Short-Term Memory (LSTM) cells to model the distribution of "terminal actions" – field goal attempt, shooting foul, non-shooting foul, turnover, no action. Assigning each such action an expected points value allows to translate these distributions into numerical values, and combining the model with a player embedding, which groups players who contribute to similar distributions closely together, leads to a final model that allows to decide whether a given player's action increased or decreased the expected points value.

2.5 "Bhostgusters"

The authors of [6] use five two-layer LSTM networks to model where defensive players can be expected to be given all other players' position, team tactics, and their own tendencies. Using a regression-based expected points model for game states, they use their position modeling to answer questions such as whether players were where they were supposed to be, whether their positioning best reduced expected points, but also counterfactuals, such as the changes that a different offensive action would have induced.

2.6 "Deep Reinforcement Learning in Ice Hockey for Context-Aware Player Evaluation"

Whether ice hockey players positively impact their team's chances to score is explored in [4], finally. Reinforcement learning is a technique originally employed to learn control strategies for robots in environments for which specifying expert rules is difficult but the arrival in a final state can be clearly identified. Combining it with deep learning allows to learn quality functions for large state-action spaces, which here are not used to make action recommendations but rather to retroactively assess player actions over the course of a match or a season.

3 Conclusion

As the preceding section shows, using the raw data is not enough and data transformation and augmentation are vital, a step during which ML techniques can already be employed and which requires a good understanding of the sport. The concrete goal, in turn, informs which ML techniques out of the large available tool kit to use to learn the final model, and how to employ it to assess players.

References

1. Bialkowski, A., Lucey, P., Carr, P., Yue, Y., Sridharan, S., Matthews, I.: Large-scale analysis of soccer matches using spatiotemporal tracking data. In: ICDM. pp. 725–730. IEEE (2014)
2. Bransen, L., Robberechts, P., Van Haaren, J., Davis, J.: Choke or shine? Quantifying soccer players' abilities to perform under mental pressure. In: MIT Sloan Sports Analytics Conference. pp. 1–25 (2019)
3. Decroos, T., Bransen, L., Van Haaren, J., Davis, J.: Actions speak louder than goals: Valuing player actions in soccer. In: KDD. pp. 1851–1861 (2019)
4. Liu, G., Schulte, O.: Deep reinforcement learning in ice hockey for context-aware player evaluation. In: IJCAI. pp. 3442–3448 (2018)
5. Power, P., Ruiz, H., Wei, X., Lucey, P.: Not all passes are created equal: Objectively measuring the risk and reward of passes in soccer from tracking data. In: KDD. pp. 1605–1613 (2017)
6. Seidl, T., Cherukumudi, A., Hartnett, A., Carr, P., Lucey, P.: Bhostgusters: Realtime interactive play sketching with synthesized NBA defenses. In: MIT Sloan Sports Analytics Conference (2018)
7. Sicilia, A., Pelechrinis, K., Goldsberry, K.: Deephoops: Evaluating micro-actions in basketball using deep feature representations of spatio-temporal data. In: KDD. pp. 2096–2104 (2019)