Valuing On-the-Ball Actions in Soccer: A Critical Comparison of xT and VAEP

Maaike Van Roy, Pieter Robberechts, Tom Decroos, Jesse Davis

KU Leuven, Department of Computer Science

{firstname.lastname}@kuleuven.be

Abstract

Objectively quantifying a soccer player's contributions within a match is a challenging and crucial task in soccer analytics. Many of the currently available metrics focus on measuring the quality of shots and assists only, although these represent less than 1% of all on-the-ball actions. Most recently, several approaches were proposed to bridge this gap. By valuing how actions increase or decrease the likelihood of yielding a goal, these models are effective tools for quantifying the performances of players for all sorts of actions. However, we lack an understanding of their differences, both conceptually and in practice. Therefore, this paper critically compares two such models: expected threat (xT) and valuing actions by estimating probabilities (VAEP). Both approaches exhibit variety in their design choices, that leads to different top player rankings and major differences in how they value specific actions.

Introduction

A fundamental task in soccer analytics is to objectively quantify a player's performance during a match. Typically, the goal is to summarize a player's contribution to the team's performance using one, or a handful of numbers. This can help inform a variety of different decisions that a club must make in areas such as team selection, opponent scouting, and player acquisition. Furthermore, these approaches facilitate fan engagement as they provide fodder for debating the relative merits of different players (e.g., McHale, Scarf, and Folker 2012; Decroos et al. 2019) or they can help tell the story of a match (e.g. Decroos et al. 2017a).

In recent years, soccer analytics researchers and enthusiasts have proposed several of these performance metrics for assessing individual players. Although, the majority of these metrics focuses on measuring the quality of specific action types in a variety of specific game situations, such as shooting opportunities (Green, 2012), off-ball positioning (Spearman, 2018), passing (Bransen and Van Haaren, 2018) and set pieces (McKinley, 2018). The latest research has attempted to join these models together in a unifying framework that can value a wide range of action types in varying game scenarios (Decroos et al., 2019; Yam, 2019; Singh, 2019; Fernández, Bornn, and Cervone, 2019).

Besides soccer, such models were developed in many other sports, including basketball (Cervone et al., 2014), American football (Romer, 2006), ice hockey (Routley and Schulte, 2015; Liu and Schulte, 2018) and rugby (Kempton, Kennedy, and Coutts, 2016). However, the low-scoring nature of soccer and the small number of on-the-ball actions makes quantifying a player's contributions within a soccer match particularly challenging. Therefore, those models became popular in soccer analytics only recently, fueled by the availability of more extensive data.

Two primary data sources about soccer matches exist that can be used to value actions: event stream data and optical tracking data. Event stream data annotates the times and locations of specific events (e.g., passes, shots, and cards) that occur in a game. Optical tracking data records the locations of the players and the ball multiple times per second. While some work exists on valuing actions using tracking data (Fernández, Bornn, and Cervone, 2019; Link, Lang, and Seidenschwarz, 2016; Spearman, 2018), the vast majority of work focuses on event stream data as it is more widely available, both in terms of leagues covered and availability to clubs.[1]

Broadly speaking, there are three styles of approaches for valuing actions in soccer using event stream data:

Count-based approaches. These techniques (McHale and Scarf, 2007; McHale, Scarf, and Folker, 2012; Pappalardo et al., 2019) rate players by (1) assigning a weight to each action type, and (2) calculating a weighting sum of the number of times a player performs each action type (e.g., pass, dribble, cross, tackle) during a match. The weights are typically learned by training a model that correlates these counts with either the match outcome or the number of goals scored (or conceded).

Expected possession value (EPV) approaches. These techniques (Rudd, 2011; Mackay, 2017; Decroos et al., 2017b; Yam, 2019; Singh, 2019) divide a match into possessions or phases, which are sequences of consecutive on-the-ball actions where the same team possesses the ball. Hence, these models value each action that progresses the ball, typically by seeing how much the action changed the team's chances of producing a goal

---

[1]Often, tracking data is not shared across leagues, which makes event stream data valuable for player recruitment purposes.

scoring attempt. Conceptually, the vast majority of these approaches can be seen as modeling a possession using a Markov decision process.

Action-based approaches. VAEP (Decroos et al., 2019) is a recent approach that goes beyond the possession-based ones by trying to value a broader set of actions and by taking the action and game context into account. Decroos et al. frame the problem as a binary classification task and rate each action by estimating its effect on the short-term probabilities that a team will both score or concede.

Despite the prevalence and importance of these types of models, the various approaches are rarely, if ever, directly compared either conceptually or empirically. In this work, we will focus on providing such a comparison between the elegant and popular expected threat (xT) model (Singh, 2019) and the VAEP model (Decroos et al., 2019). We select these two models as they are canonical exemplars for the last two styles of approaches. Because the EPV and action-based approaches both focus on rating each individual action based on properties of the action, they are more closely related in spirit to each other than to the count-based approaches which look at aggregated actions without accounting for any aspect of an action's context. We highlight the key differences in design choices made by xT and VAEP, which yields different strengths and weaknesses. Qualitatively, we show several illustrative actions where this leads to each formalism producing different valuations for particular actions. Quantitatively, we show that this leads to different rankings of players with xT being slightly more correlated with playmaking whereas VAEP tends to favor shooting. Importantly, both rankings deviate from traditional metrics like goals or assists per 90 minutes, which shows they give novel insights into player performance.

Action Valuing Frameworks

When considering event stream data, a soccer match can be viewed as a sequence of n consecutive actions $a_1, a_2, \ldots a_n$. Each action a_i is described by a number of properties such as its start location, its end location, its start time, and what type of action it was. The effect of an action a_i is to move the game from state $S_{i-1} = \{a_1, \ldots, a_{i-1}\}$ to state $S_i = \{a_1, \ldots, a_{i-1}, a_i\}$. Consequently, at a high-level EPV and action-based approaches all value actions according to the following equation:

$$V(a_i) = Q(S_i) - Q(S_{i-1}) \qquad (1)$$

where Q captures the value or quality of a particular game state. The differences among the various approaches arise in how they represent and assign values to the various game states. Next, we will describe xT's (Singh, 2019) and VAEP's (Decroos et al., 2019) approaches for doing so.[2]

Expected Threat

The expected threat or xT model (Singh, 2019) is a possession-based Markov model. This modelling approach

---

[2]Our implementations of these frameworks are available at https://github.com/ML-KULeuven/socceraction.

---

implies that soccer games are divided into possessions, which are periods of the game where the same team has control of the ball. Subsequently, each possession can be discretized in a consecutive sequence of ball-progressing actions. The key insight underlying xT and similar models is that players perform these actions with the intention to move the game into a state in which they are more likely to score. These game states directly correspond to the transient states of a Markov decision process: players transition the game from one state to another by passing or dribbling[3] until absorption (i.e., a goal or possession turnover).

Although the game states can be made arbitrarily complex (Rudd, 2011; Yam, 2019), xT represents each game state S_i by only considering the location of the ball. Therefore, xT overlays a $M \times N$ grid on the pitch in order to divide it into $M \cdot N$ zones. Each zone z is then assigned a value $\text{xT}(z)$ that reflects how threatening teams are at that location, in terms of scoring (Figure 1). The value $Q(S_i)$ of game state $S_i = \{a_1, \ldots, a_i\}$ is then simply the value of that zone corresponding to a_i's end location. The Markov decision process view allows deriving these xT values from historical data by iteratively solving the following equation:

$$\text{xT}(z) = s_z \cdot \text{xG}(z) + m_z \cdot \sum_{z'=1}^{M \times N} T_{z \to z'} \cdot \text{xT}(z'),$$

where s_z is the probability that a player will shoot when in zone z, $\text{xG}(z)$ is the probability of a shot from zone z being converted into a goal, m_z is the probability that a player will move the ball when in zone z, and T is a transition matrix that defines the probability that the player moves the ball to each of the other zones when in zone z. Intuitively, solving the equation boils down to looking another action ahead with each added iteration. In the first iteration, all $\text{xT}(z)$ values are initialized to zero. After iteration i, $\text{xT}(z)$ then represents the probability of scoring within the next i actions.

Subsequently, the model values a successful action a_i that moves the ball from zone z to zone z' by computing the difference between the threat value before and after that action:

$$V_{xT}(a_i) = \text{xT}(z') - \text{xT}(z). \qquad (2)$$

VAEP

VAEP uses a much more complex game state representation than xT. It considers the three last actions that happened during the game: $S_i = \{a_{i-2}, a_{i-1}, a_i\}$. Then each game state is represented using three types of features. The first category of features includes characteristics of the action itself such as its location and type as well as more complex relationships such as the distance and angle to the goal. The second category of features captures the context of the action, such as the current tempo of the game, by comparing

---

[3]Note that event stream data contains various definitions of what constitutes a pass. We regard all open play actions that intentionally move the ball between two players of the same team as passes. This includes regular passes, head passes, long range kicks, crosses, etc. Similarly, the term dribble is quite ambiguous as it can denote trying to move past a player or just carrying the ball from one location to another. We regard both types as a dribble.
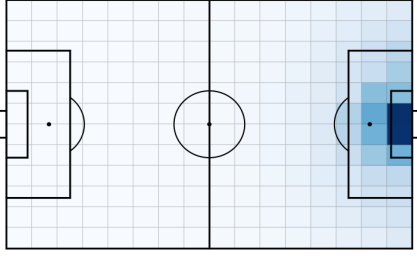
Figure 1: A heatmap of the xT values obtained after training the xT model on the training data set until convergence. Darker colours denote higher xT values.

the properties of consecutive actions. Examples of this type of feature include the distance covered and time elapsed between consecutive actions. The third category of features captures the current game context by looking at things such as the time remaining in the match and the current score differential.

Like xT (equation 2), VAEP values an action based on how it alters the game state,:

$$V_{\text{VAEP}}(a_i) = Q(S_i) - Q(S_{i-1}).$$

However, it differs in how it values each game state:

$$Q(S_i) = P_{\text{score}}^k(S_i, t) - P_{\text{concede}}^k(S_i, t).$$

where $P_{score}^k(S_i, t)$ and $P_{concede}^k(S_i, t)$ are the probabilities that team t which possesses the ball in state S_i will respectively score or concede in the next k actions. This valuation arises from the insight that players tend to perform actions not only to increase their team's chance of scoring a goal, but also to decrease their team's chance of conceding a goal in the near future.

Hence, an alternative way to view this is that VAEP estimates the risk-reward trade-off of an action:

$$V_{\text{VAEP}}(a_i) = \Delta P_{\text{score}}(a_i, t) - \Delta P_{\text{concede}}(a_i, t) \quad (3)$$

where

$$
\begin{aligned}
\Delta P_{\text{score}}(a_i, t) &= P_{\text{score}}^k(S_i, t) - P_{\text{score}}^k(S_{i-1}, t) \\
\Delta P_{\text{concede}}(a_i, t) &= P_{\text{concede}}^k(S_i, t) - P_{\text{concede}}^k(S_{i-1}, t).
\end{aligned}
$$

Thus, ΔP_{score} is referred to as the "offensive value" of an action, while $\Delta P_{concede}$ can be thought of as the "defensive value" of an action.

VAEP models the scoring and conceding probabilities separately as these effects may be asymmetric in nature and context-dependent. Hence, it trains one gradient boosted tree model to predict each one based on the current game state. For estimating $P_{score}^k(S_i, t)$, each game state is given a positive label (= 1) if the team that possesses the ball after action a_i scores a goal in the subsequent k actions. Otherwise, a negative label (= 0) is given to the game state. Analogously, for estimating $P_{concede}^k(S_i, t)$, each game state is given a positive label (= 1) if the team that possesses the ball after action a_i concedes a goal in the subsequent k actions. If not, a negative label (= 0) is given to the game state.

Comparing xT and VAEP

xT and VAEP are similar approaches in the sense that they both value individual on-the-ball actions of soccer players by evaluating how actions increase or decrease the likeliness of yielding a goal. To do so, both approaches rely on the insight that on-the-ball actions are distinct events that modify the game state. The goal of both frameworks is to measure how valuable an action's resulting change of game state is by computing the differences between the game state values before and after the action.

However, xT and VAEP approach two aspects differently. First, xT uses a very limited game state representation that is purely location-based, while VAEP employs a detailed feature-based representation that captures the action and game context. Second, xT is possession-based, meaning that the xT framework splits up the game in possessions and estimates the likelihood of any goal occurring within the same possession. VAEP splits up the game in action sequences of a fixed length and looks beyond turnovers. Below, we outline how these design choices impact the generated action values.

Location-based vs feature-based

VAEP models the game state as an extensive feature-based description of the previous three actions and the game context. xT's game state representation, on the other hand, is purely location-based. It discretizes the pitch into $M \cdot N$ zones by overlaying a grid and encodes the game state as the zone in which the ball is. This enables using simple and elegant dynamic programming approaches designed for Markov decision processes to compute each game state's xT value. However, it seriously limits the game dynamics that its values can capture in terms of the action types that can be valued, and the action and game context that is captured.

xT can only value ball-progressing actions. Since a game state is fully captured by a zone on the pitch, xT can only value actions that move the ball from one zone to another (i.e., passes, dribbles and crosses). Hence, it ignores defensive actions like tackles and interceptions, as well as valuable offensive actions such as take-ons within the same zone of the pitch.[4] Therefore, xT and related models are often referred to as "ball-progression models" (Yam, 2019).

VAEP captures the action context. xT captures only a very limited portion of the context in which actions are performed. Most of the time, the location of an event in space is not sufficient to fully evaluate its potential impact. The probability of scoring in a game state can depend on the type, accuracy and speed during the previous actions leading up to the current state. For instance, a subsequent shot might be easier when a player is positioned in front of goal via a through ball, compared to when he had to dribble past a couple of defenders first. Therefore, a critical aspect to properly evaluate soccer situations is to have a clear understanding of the ongoing context. VAEP has a much more accurate representation of this context.

---

[4]By definition, xT assigns all such actions a value of zero.

VAEP captures the game context. VAEP also includes features in its model to capture the game context such as the number of goals scored by each team, the time remaining in the match, and the score difference. This could be valuable as it is known that the chances of scoring vary slightly according to goal difference (Robberechts, Van Haaren, and Davis, 2019; Decroos and Davis, 2019a). Again, xT does not consider these factors.

xT values are interpretable. The detailed game state representation used in the VAEP framework has a cost in terms of interpretability. Where each game state in the xT framework is assigned one out of $M \cdot N$ possible values which only depends on the location of the ball, a function approximator (e.g., a gradient boosted tree ensemble) is needed to value game states in the VAEP framework. As such, game state values are derived from complex interactions between a large set of features. Explaining why a particular value is assigned to a specific game state is no longer straightforward in this framework.

Possession-based vs window-based

Both approaches require historical observations of action sequences to estimate the value of a game state. They differ in how they split up the game in those sequences. xT is possession based. The framework splits up the game in possessions (i.e., sequences of consecutive on-the-ball actions where the same team possesses the ball) and estimates the likelihood of any goal occurring within the same possession. In contrast, VAEP values an action by looking at the probability of a goal being scored within a finite number of actions. This leads to the following two consequences:

VAEP captures the risk involved in an action. The xT model only values an action's offensive contribution, that is, how it changes the team's chance of scoring. In contrast, because VAEP considers what happens after turnovers, it can estimate how an action alters a team's chance of conceding in addition to the action's offensive contribution. Hence, it may better capture the risks associated with taking certain actions. For example, a square pass in the middle of the field enables the other team to quickly launch a counter attack if they would intercept the ball. While VAEP may capture the risks associated with such a pass, xT ignores it.

VAEP can value 'failed' actions accurately. Not all losses of possession are equal in soccer. For example, consider a scenario in which a player has the opportunity to clear the ball having no opportunity to reach another teammate. In such a scenario, he can simply kick the ball forward giving the other team the opportunity to recover the ball easily and quickly build up a new attack. Alternatively, he can kick a long ball out of bounds, giving his team the opportunity to escape the pressing and try to recover the ball by aggressively pressing the subsequent throw-in. The second option is clearly the better one, but cannot be valued by the xT framework.

Data

The data used for the experiments in this paper is the StatsBomb data from the English Premier League for the 2017/2018 and 2018/2019 seasons. Both models are trained on the data of the first season. Our match event stream data is encoded in the SPADL format (Decroos et al., 2019), which is a language designed for analysis that unifies the representations used by different vendors. This language facilitates analysis by ignoring the optional information (e.g., about weather changes) in the data and by representing all on-the-ball actions using the same fixed set of attributes. Moreover, a publicly available converter is available that translates the data from various providers into this format.[5] The XGBoost algorithm was used as the prediction method for the VAEP model. A 16×12 grid was used for the xT model and convergence was reached after 6 iterations. Afterwards, both models are used to rate actions in the 2018/2019 season.

Experimental comparison of action values

In this section, we compare and contrast how xT and VAEP assess different actions in different contexts. Concretely, we will explore four actions: a (risky) backward pass in the own half, recovering the ball to set up a counter attack, a forward dribble into the opponent's penalty box, and a through ball near the opponent's penalty box.

Backward passes into a team's own penalty box

Backward passes have an interesting risk-reward trade-off as they usually open up space (reward), but also move the ball closer to the team's own goal (risk). In particular, a backward pass into your team's penalty box is especially risky as losing the ball in this position may lead to a big scoring chance for the opposing team. These passes happen roughly 19 times per game.
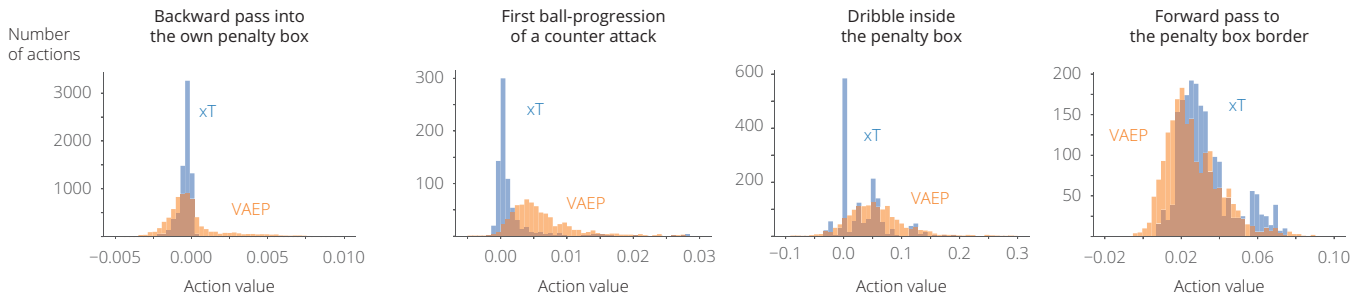
xT neither captures the risk nor reward of these passes, as all zones near a team's own goal are valued close to zero (Figure 1). Figure 2a shows how VAEP assigns more diverse values to these passes, both positive and negative.

First ball progression of counter attacks

The first ball progression of a counter attack is an action that interrupts the opponent's possession sequence. This action must occur in your own half and be the first action in a sequence of actions that leads to a shot in the next 20 seconds. Such counter attacks occur roughly 2 to 3 times per game.

This ball progression can be a valuable action, both from an offensive and a defensive viewpoint. On the one hand, the action interrupts the opponent's attack, thus reducing the odds of conceding a goal. On the other hand, recovering the ball while the opponent is still in an offensive position gives the player's team the opportunity of building a fast counter attack that exploits the opponents' unorganized defensive positioning and thus increasing the odds of scoring a goal. The fact that the ball was only very recently recovered (and how this differs from normal open play) is a contextual clue that can only be leveraged by VAEP's more powerful

---

[5]https://github.com/ML-KULeuven/socceraction

(a) xT does not capture the risk nor reward of backward passes into a team's own penalty box.

(b) Only VAEP captures initiating a counter attack.

(c) Short dribbles that do not move the ball to a different zone are not valued by xT.

(d) xT values the positional advantage gained by through balls higher than VAEP.

Figure 2: Histograms of the VAEP and xT values for a set of actions that are rated differently by both frameworks.

reasoning on game states. Figure 2b illustrates how the distribution of VAEP values of ball recoveries in your own half has a higher mean and variance, whereas the distribution of xT values of ball recoveries in your own half is extremely skewed towards zero.

Forward dribbles inside the penalty box

Next, we consider forward dribbles inside the penalty box that end up in the 23×13 meter rectangle in front of goal. This rectangle corresponds to the rectangle made up of 4×2 cells in front of the goal with the highest xT values. Although our definition of dribbles does not require that they pass a defender, they still considerably raise the odds of scoring a goal for two reasons: (1) the ball got moved a lot closer to the goal, and (2) the player kept control of the ball close to goal. Successful forward dribbles inside the penalty box occur roughly four times per game.

Figure 2c shows how xT assigns a value of zero to a major part of these forward dribbles inside the penalty box. Since xT discretizes the pitch into relatively large zones, many short dribbles do not move the ball into a different zone and therefore do not increase the xT value. Yet, these short dribbles may suffice to take on a defender and – when the ball is extremely close to the goal – small differences in location can considerably increase the odds of scoring.

Completed forward passes to the border of the penalty box

Here we explore successful forward passes from the third quarter of the field to the border of the penalty box. More precisely, we consider passes that end within 3 meters of either side of the penalty box border line that is parallel to the goal. Such completed forward passes occur roughly 5 times per game.

A successful forward pass to the border of the penalty box can considerably raise the odds of scoring a goal for two reasons: (1) it moves the ball closer to the goal, and (2) it often bypasses at least one player from the opposing team. Figure 2d shows how, on average, xT values the positional advantage gained by a through ball more than VAEP. Due to the detailed game state representation used by VAEP, it is hard

to explain why VAEP assigns lower values to these types of actions than xT. Using only the reasoning behind xT, one possible explanation for this is that xT is better than VAEP at capturing the positional advantage. However, it is possible that only using the positions of the actions will overestimate the action values and more information about the game state is needed to value these actions correctly. Unfortunately, determining the ground truth of these action values is very difficult, if not impossible. Thus, deciding which method is better suited for valuing these types of actions, or which method makes a better estimation of these action values is not straightforward.

Experimental comparison of player ratings

The most important application of an action valuing model is summing the action values of players to construct player ratings. Since spending more time on the pitch offers more opportunities to contribute, player ratings are normalized per 90 minutes of game time (Decroos et al., 2019). Given a time frame T and player p, a player's rating is computed as

$$\text{rating}(p) = \frac{90}{m} \sum_{a \in A_p^T} V(a),$$

where A_p^T is the set of actions the player p performed during time frame T, $V(a_i)$ is the value of an action according to either xT or VAEP, and m is the number of minutes the player played during T.

In this section, we compare and contrast how xT and VAEP rate players. We compare each method on (1) their top player rankings, (2) their correlation to traditional player performance metrics, and (3) their robustness.

Comparison of top-25 player rankings

To some extent, it is possible to subjectively evaluate both rating systems. Table 1 shows the top-25 players in the 2018/2019 season of the English Premier League according to xT and VAEP. Football fans would agree that the rankings produced by both xT and VAEP feature top players. Yet, there are some major differences between both rankings.

A notable emission from the top-25 players according to xT is Manchester City's Sergio Agüero as he is ranked 19th by VAEP. One possible explanation for this is that Agüero is a world class striker and thus probably scores more goals than expected from his shots due to his superior finishing skill. Overperforming on shots is a skill that will be rewarded heftily by the VAEP framework, as this directly influences the scoreline. In summary, the reason why VAEP picks up Agüero and xT does not is his average contribution from ball-progressing actions, but superior finishing.

A notable emission from the top-25 players according to VAEP is Manchester United's Alexis Sánchez as he is ranked 7th by xT. One possible reason for this is that it was Sánchez's first full season at Mourinho's defensively organised Manchester United, a playing style he was not used to after four seasons at Arsenal. During this season, his expected goals per 90 minutes (i.e. xG/90) more than halved. Yet, his number of key passes leading to a shot per 90 minutes (i.e. KP/90) stayed roughly the same. These stats indicate that he created roughly the same amount of threat as he did at Arsenal by positioning others in front of goal, but that he had fewer goal-scoring opportunities himself. As xT does not consider shots when ranking players, the fact that Sánchez still managed to complete key passes into high-value zones delivers him a higher xT value. On the other hand, the decrease in shots leads to VAEP valuing him less.

In both models, offensive actions have access to higher rewards. It is thus easier for offensive players to get higher ratings than for defensive players. Therefore, the top of the lists mainly contain attacking players, whereas defensive players are ranked lower. For example, top class defender Virgil van Dijk is ranked 81st by VAEP and 142nd by xT.

Finally, Figure 3 shows the evolution of the Jaccard similarity coefficient between top-k rankings created by xT and VAEP. This metric essentially measures the similarity between the sets of players in both rankings. The top-25 player rankings of both models have a relatively small similarity coefficient of 0.48, indicating again that xT and VAEP value different qualities. After about the top 25 players, most players get a similar, average rating such that small rating differences can have large effects in the rankings. Hence, the similarity coefficient drops to 0.35 before steadily increasing.

Comparison with traditional performance metrics

Currently, players' offensive contributions are usually quantified by counting goals and assists, as those events directly influence the scoreline. Although they largely fail to account for the circumstances under which the actions were performed, these statistics provide some insights into the performances of individual soccer players. Therefore, we correlate the ratings of both models to these two baseline metrics, normalized for game time.

For VAEP ($\rho_{g/90} = 0.41$) we obtain a stronger correlation with goals per 90 minutes than xT ($\rho_{g/90} = 0.26$), while assists per 90 minutes is correlated stronger with xT ($\rho_{a/90} = 0.53$) than with VAEP ($\rho_{a/90} = 0.33$). That is because VAEP generally assigns goals high action values, such that players can boost their VAEP ranking by scoring many goals. In xT, players do not get credit for scoring goals
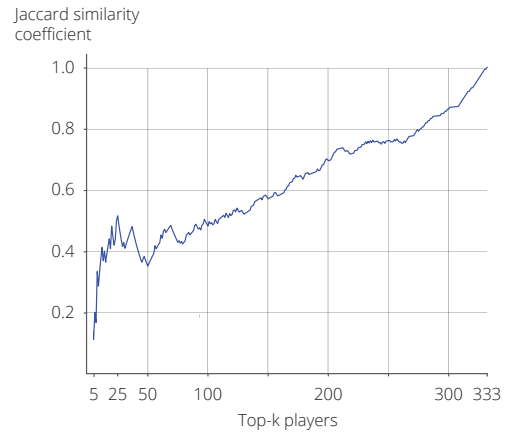

Figure 3: The Jaccard similarity coefficient between the top-k rankings created by the xT and VAEP model. The plots shows an increase in similarity for the top-25 players of both models and then a drop when adding the next 25 players before steadily increasing.

such that the xT-based rankings are biased towards creative players that complete many key passes and dribbles.

Robustness

A good rating system should capture the true quality of all players. Although some fluctuations in performances are possible across games, over the course of a season a few outstanding performances (possibly stemming from a big portion of luck) should not dramatically alter an assessment of a player. To measure which rating system produces the most robust player ratings over the course of a season, we split the data in two random disjoint subsets. Subsequently, we compute each players' average rating separately for both subsets and evaluate the Pearson correlation between both.

Figure 4 plots the relation between each players' average rating on the two subsets for xT (Figure 4a) and VAEP (Figure 4b). As can be observed visually from these figures, the xT model achieves a much stronger correlation ($\rho = 0.89$) than the VAEP model ($\rho = 0.25$). The correlation for the VAEP model improves when only the ball-progressing actions that are valued by the xT model are included (i.e., passes, dribbles and crosses) and when only the offensive value of actions is considered (Figure 4c). In this setting, VAEP achieves a correlation of $\rho = 0.59$.

This ultimately leads to the conclusion that the ratings produced by the xT model are more robust than the ones from the VAEP model, even when adjusting for the different actions and risk (i.e., defensive value). These differences can be attributed to two factors. First, because VAEP assigns high action values to goals, the aggregated ratings can vary significantly based on whether or not these goals are included. Especially for defensive players, a difference in only three goals can double or half their ratings. Second, since players are pretty consistent in what type of actions they perform at which locations (Decroos and Davis, 2019b), it makes sense that a metric purely based on zonal changes

Table 1: The top-25 players who played at least 900 minutes in the 2018/2019 English Premier League season with ratings according to both the xT and VAEP models. Goals per 90 minutes (g/90) and assists per 90 minutes (a/90) are also shown for each player.

(a) The top-25 players according to the xT model.

R_{xT}	R_{vaep}	Player	Rating	g/90	a/90
1	1 =	Eden Hazard	0.547	0.49	0.46
2	16 ▼	Adama Traoré	0.490	0.10	0.10
3	24 ▼	Kevin De Bruyne	0.456	0.19	0.19
4	25 ▼	Alex Iwobi	0.441	0.14	0.27
5	6 ▼	Anthony Martial	0.400	0.55	0.11
6	8 ▼	Felipe Anderson	0.399	0.26	0.12
7	106 ▼	Alexis Sánchez	0.395	0.10	0.31
8	2 ▲	Gerard Deulofeu	0.386	0.43	0.21
9	12 ▼	Wilfried Zaha	0.384	0.30	0.15
10	3 ▲	Riyad Mahrez	0.371	0.47	0.27
11	9 ▲	Raheem Sterling	0.340	0.55	0.32
12	67 ▼	Willian	0.338	0.13	0.25
13	30 ▼	Kieran Trippier	0.337	0.04	0.12
14	7 ▲	Mohamed Salah	0.327	0.60	0.22
15	20 ▼	James Milner	0.324	0.25	0.20
16	80 ▼	Nathan Redmond	0.321	0.16	0.11
17	15 ▲	Trent Alexander-Arnold	0.316	0.04	0.44
18	10 ▲	Jonjo Shelvey	0.313	0.10	0.10
19	51 ▼	Benjamin Mendy	0.310	0.00	0.50
20	14 ▲	Ryan Fraser	0.308	0.20	0.40
21	36 ▼	Oleksandr Zinchenko	0.302	0.00	0.23
22	53 ▼	Andrew Robertson	0.298	0.00	0.31
23	21 ▲	David Silva	0.296	0.22	0.30
24	4 ▲	Xherdan Shaqiri	0.289	0.52	0.26
25	110 ▼	Marc Albrighton	0.287	0.11	0.11

(b) The top-25 players according to the VAEP model.

R_{vaep}	R_{xT}	Player	Rating	g/90	a/90
1	1 =	Eden Hazard	0.558	0.49	0.46
2	8 ▼	Gerard Deulofeu	0.535	0.43	0.21
3	10 ▼	Riyad Mahrez	0.525	0.47	0.27
4	24 ▼	Xherdan Shaqiri	0.489	0.52	0.26
5	47 ▼	Son Heung-Min	0.479	0.52	0.26
6	5 ▲	Anthony Martial	0.469	0.55	0.11
7	14 ▼	Mohamed Salah	0.466	0.60	0.22
8	6 ▲	Felipe Anderson	0.466	0.26	0.12
9	11 ▼	Raheem Sterling	0.463	0.55	0.32
10	18 ▼	Jonjo Shelvey	0.442	0.10	0.10
11	210 ▼	Ruben Loftus-Cheek	0.425	0.56	0.19
12	9 ▲	Wilfried Zaha	0.411	0.30	0.15
13	33 ▼	Mesut Özil	0.404	0.26	0.10
14	20 ▼	Ryan Fraser	0.398	0.20	0.40
15	17 ▼	Trent Alexander-Arnold	0.382	0.04	0.44
16	2 ▲	Adama Traoré	0.381	0.10	0.10
17	35 ▼	Dwight McNeil	0.379	0.17	0.28
18	63 ▼	Sadio Mané	0.375	0.64	0.03
19	109 ▼	Sergio Agüero	0.368	0.75	0.29
20	15 ▲	James Milner	0.361	0.25	0.20
21	23 ▼	David Silva	0.351	0.22	0.30
22	55 ▼	Christian Eriksen	0.351	0.26	0.39
23	81 ▼	Pierre-Emerick Aubameyang	0.347	0.72	0.16
24	3 ▲	Kevin De Bruyne	0.345	0.19	0.19
25	4 ▲	Alex Iwobi	0.343	0.14	0.27

gives consistent results. The VAEP ratings add more context and therefore allow more variation.

Conclusion

xT and VAEP are two prominent approaches for the important task of valuing actions in a soccer match. We performed a critical comparison of these two approaches, conceptually, qualitatively and quantitatively. Key differences arise in how each approach represents the game state and what actions are valued. These lead to interesting differences such as VAEP better capturing the risk-reward tradeoff of actions and xT being more robust. Importantly, both metrics produce rankings that deviate from those produced by considering traditional metrics (goals or assists). Hence, they provide additional insights into player performance.

Acknowledgements

References

Bransen, L., and Van Haaren, J. 2018. Measuring football players' on-the-ball contributions from passes during games. In *International Workshop on Machine Learning and Data Mining for Sports Analytics*, 3–15. Springer.

Cervone, D.; D'Amour, A.; Bornn, L.; and Goldsberry, K. 2014. POINTWISE: Predicting Points and Valuing Decisions in Real Time with NBA Optical Tracking Data. In *MIT Sloan Sports Analytics Conference*.

Decroos, T., and Davis, J. 2019a. Interpretable prediction of goals in football. In *Innovation in Football Conference*. StatsBomb.

Decroos, T., and Davis, J. 2019b. Player vectors: Characterizing soccer players' playing style from match event streams. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*.

Decroos, T.; Dzyuba, V.; Van Haaren, J.; and Davis, J. 2017a. Predicting soccer highlights from spatio-temporal match event streams. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence*, 1302–1308.

Decroos, T.; Van Haaren, J.; Dzyuba, V.; and Davis, J. 2017b. STARSS: a spatio-temporal action rating system
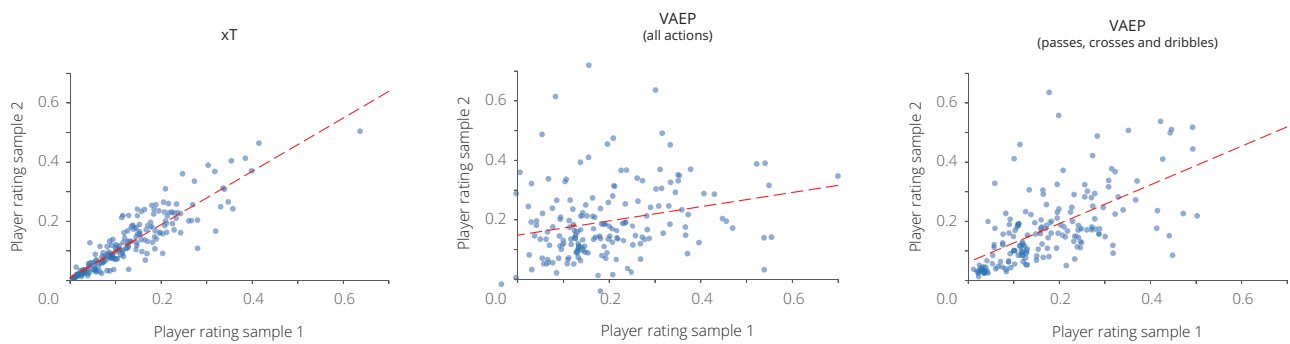
(a) xT ratings for players in both samples.

(b) VAEP ratings for players in both samples.

(c) VAEP ratings considering only offensive values for passes, dribbles and crosses in both samples.

Figure 4: Scatter plots of the player ratings produced on both samples. The linear least squares regression line is shown in red. The plot for xT displays a high correlation between both ratings. For VAEP, the plots are more random than the one for the xT model, but a correlation becomes visible when only looking at the offensive values for passes, dribbles and crosses of VAEP.

for soccer. In *Machine Learning and Data Mining for Sports Analytics ECML/PKDD 2017 workshop*, volume 1971, 11–20.

Decroos, T.; Bransen, L.; Van Haaren, J.; and Davis, J. 2019. Actions speak louder than goals: Valuing player actions in soccer. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 1851–1861.

Fernández, J.; Bornn, L.; and Cervone, D. 2019. Decomposing the immeasurable sport: A deep learning expected possession value framework for soccer. In *MIT Sloan Sports Analytics Conference*.

Green, S. 2012. Assessing the performance of Premier League goalscorers. http://www.optasportspro.com/about/optaproblog/posts/2012/blog-assessing-the-performance-of-premier-league-goalscorers. Accessed: 2019-11-14.

Kempton, T.; Kennedy, N.; and Coutts, A. J. 2016. The expected value of possession in professional rugby league match-play. *Journal of sports sciences* 34(7):645–650.

Link, D.; Lang, S.; and Seidenschwarz, P. 2016. Real time quantification of dangerousity in football using spatiotemporal tracking data. *PLOS ONE* 11(12).

Liu, G., and Schulte, O. 2018. Deep reinforcement learning in ice hockey for context-aware player evaluation. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, 3442–3448.

Mackay, N. 2017. Improving my 'xg added' model. https://mackayanalytics.nl/2017/07/28/improving-my-xg-added-model/. Accessed: 2019-06-21.

McHale, I., and Scarf, P. 2007. Modelling soccer matches using bivariate discrete distributions with general dependence structure. *Statistica Neerlandica* 61(4):432–445.

McHale, I.; Scarf, P.; and Folker, D. 2012. On the development of a soccer player performance rating system for the english premier league. *Interfaces* 42(4):339–351.

McKinley, E. 2018. A feast for throws. https://www.americansocceranalysis.com/home/2018/12/4/a-feast-for-throws. Accessed: 2019-11-14.

Pappalardo, L.; Cintia, P.; Ferragina, P.; Massucco, E.; Pedreschi, D.; and Giannotti, F. 2019. Playerank: Data-driven performance evaluation and player ranking in soccer via a machine learning approach. *ACM Trans. Intell. Syst. Technol.* 10(5):59:1–59:27.

Robberechts, P.; Van Haaren, J.; and Davis, J. 2019. Who will win it? An in-game win probability model for football. *arXiv preprint arXiv:1906.05029*.

Romer, D. 2006. Do firms maximize? Evidence from professional football. *Journal of Political Economy* 114(2):340–365.

Routley, K., and Schulte, O. 2015. A markov game model for valuing player actions in ice hockey. In *Proceedings of the 31st Conference on Uncertainty in Artificial Intelligence*, 782–791.

Rudd, S. 2011. A Framework for Tactical Analysis and Individual Offensive Production Assessment in Soccer Using Markov Chains. In *New England Symposium on Statistics in Sports*.

Singh, K. 2019. Introducing expected threat. https://karun.in/blog/expected-threat.html. Accessed: 2019-06-21.

Spearman, W. 2018. Beyond expected goals. In *MIT Sloan Sports Analytics Conference*.

Yam, D. 2019. Attacking contributions: Markov models for football. https://statsbomb.com/2019/02/attacking-contributions-markov-models-for-football/. Accessed: 2019-06-21.