Predicting the Potential of Professional Soccer Players

Ruben Vroonen[1], Tom Decroos[1], Jan Van Haaren[2], and Jesse Davis[1]

[1] KU Leuven, Department of Computer Science, Celestijnenlaan 200A, 3001 Leuven, Belgium
[2] SciSports, Hengelosestraat 500, 7251 AN Enschede, The Netherlands

Abstract. Projecting how a player's skill level will evolve in the future is a crucial problem faced by sports teams. Traditionally, player projections have been evaluated by human scouts, who are subjective and may suffer from biases. More recently, there has been interest in automated projection systems such as the PECOTA system for baseball and the CARMELO system for basketball. In this paper, we present a projection system for soccer players called APROPOS which is inspired by the CARMELO and PECOTA systems. APROPOS predicts the potential of a soccer player by searching a historical database to identify similar players of the same age. It then bases its prediction for the target player's progression on how the similar previous players actually evolved. We evaluate APROPOS on players from the five biggest European soccer leagues and show that it clearly outperforms a more naive baseline.

1 Introduction

With more than 250 million players, soccer is the most popular sport in the world. Due to technological advances, new soccer data sources such as event streams and optical tracking from matches are rapidly becoming available. This has lead to an explosion of interest in the area of soccer analytics. Most research tends to focus on analyzing soccer gameplay (e.g., [11, 7, 8, 9, 10]). This has ranged from formation identification [7], to evaluating the quality of shots [9, 10] to detecting commonly employed offensive strategies [8].

Another relevant problem in soccer analytics is projecting how a player's skill level will change over time. This is particularly important for clubs, as it can influence a club's player acquisition and retention strategies. In other sports, projection systems have been developed that predict a player's future performance. Two well-known examples of such systems are PECOTA [5] for MLB baseball and CARMELO [2] for NBA basketball. In a similar spirit, this paper proposes APROPOS, a system that can predict the future potential of professional soccer players. Like past approaches, we project a target player's potential by searching a historical database to identify other players with a similar profile to the target player when they were the target player's age. Then, the target player's evolution is predicted based on the observed evolutions of the identified similar players. However, one challenge in soccer is the relative paucity of events, particularly those that can be related to a match outcome. Thus, we use a set of expert ratings for a number of skills that are available on the `SoFIFA.com` website to compare the similarity between two players. This contrasts with past systems (e.g., PECOTA and CARMELO) that measure similarity based on past statistics and personal descriptive characteristics such as height and weight.

2 Related work

Multiple projection systems exist that try to predict a player's future level of performance. We focus on two specific systems: PECOTA and CARMELO. Both follow the same high-level outline that our system uses. First, given a target player, they compare the target player's profile to previous players' profiles when they were at the same stage of development as the target player. Second, they project the target player's future performance based on how the similar previous players evolved.

2.1 PECOTA

The PECOTA system (Player Empirical Comparison and Optimization Test Algorithm), named after former professional baseball player Bill Pecota, is a projection system used within Major League Baseball (MLB). It predicts the career path of a baseball player by fitting previous statistics with similar players using Bill James's similarity scores [6]. Originally developed by Nate Silver in 2002, it is currently managed by Baseball Prospectus [1]. Each year, they release the seasonal predictions for every MLB player.

2.2 CARMELO

The CARMELO system is a simplified version of the PECOTA projection model and is adapted to NBA (National Basketball Association) players. The system, named after basketball player Carmelo Anthony, gathers player statistics, characteristics and vital attributes. Every player starts with a similarity score of 100 and points are subtracted for each difference in 19 weighted statistics. The final prediction of the level of a player is made by taking the weighted average of the Wins Above Replacement (WAR) of the players with a score above 0, where the similarity scores are used as weights. The system is maintained by the site `fivethirtyeight.com` [3].

3 Data

To develop the projection system, we need data to measure the level of players. For this, we used the expert ratings on the `SoFIFA.com` website [4], which provides the player ratings that are included in the realistic FIFA video games

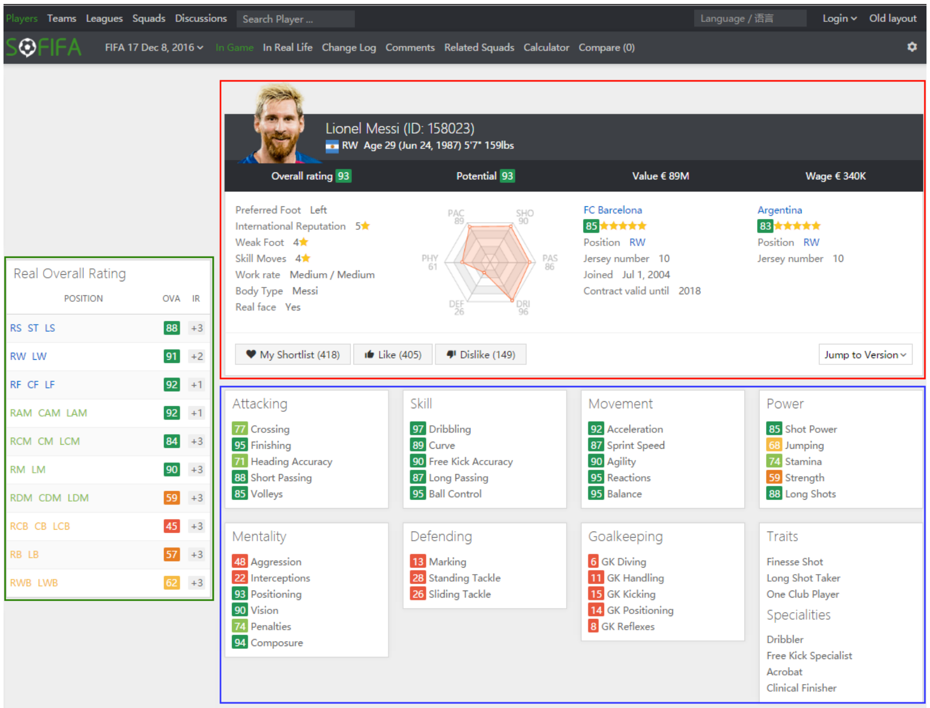

Fig. 1. Player card for Lionel Messi

published by EA Sports.[3] Each player is rated on 24 different skills and each skill is rated on a 0 to 100 scale. On the `SoFIFA.com` website, each player has a card which displays his ratings. Figure 1 shows Lionel Messi's card. The `SoFIFA.com` website has been publishing FIFA player ratings since 2007. Initially, the player cards were updated semi-annually. Since 2014, the ratings are released weekly.

We use data for players from the English, French, German, Italian and Spanish competitions. These competitions are the most popular and have the most accurate and complete information. Our database contains 57 860 player cards for 10 247 players. On average, there are 5.65 years of data for each player.

4 The APROPOS projection system

The algorithm we designed is called APROPOS (Algorithm for PRediction Of the Potential Of Soccer players). Like the PECOTA and CARMELO projection systems, it uses a nearest neighbors approach to predict how a soccer player's skill will evolve over time. Formally, the task can be defined as follows:

Given: A player p, a set of skill ratings $V_{p_{a_1}}$ for p at his current age a_1, and a future age a_2;

[3] `https://www.easports.com/fifa`

Predict: $V_{p_{a_2}}$ which is p's set of skill ratings at age a_2.

To tackle this problem, the APROPOS projection system also requires as input a similarity metric sim, a similarity threshold t, and a database of players D. Then given a player p and a future age a_2, it works as follows.

1. Add all players p' in D to set S if (1) the data for the age a_2 season of p' is available in D, and (2) when p''s age is a_1, $sim(p, p') \geq t$. If S contains less than ten players, then S consists of the ten most similar players.
2. Predict $V_{p_{a_2}}$ by combining the ratings of all players in S.

Next, we describe in detail how we perform each step.

4.1 Similarity scores

We have developed two different scores to measure the similarity between two players. Both scores compare the similarity for every single skill rating reported on `SoFIFA.com` and combine them into a final similarity score. The final score is a real number between 0 and 1, where 0 means not similar at all and 1 means completely similar (i.e., the two players are identical in all their skill ratings).

Absolute similarity score The absolute similarity score first calculates a similarity score for each skill V_r as

$$absolute_{V_r}(p, p', a_1, y) = 1 - \frac{\sqrt{\sum_{a=a_1-y+1}^{a_1} (v_{r,p_a} - v_{r,p'_a})^2}}{\sqrt{\sum_{a=a_1-y+1}^{a_1} max(v_{r,p_a}, 100 - v_{r,p_a})^2}}, \quad (1)$$

where v_{r,p_a} represents player p's observed rating for skill V_r at age a, and y represents over how many years of data the similarity metric should consider. The denominator normalizes the score relative to the maximum Euclidean distance possible for player p to reflect the percentage of similarity.

The final absolute similarity score is computed as the average over all skills:

$$sim_{abs}(p, p', a_1, y) = \frac{\sum_{V_r \in V} absolute_{V_r}(p, p', a_1, y)}{|V|} \quad (2)$$

Evolutionary similarity score The skill level of a player is also partially dependent on his team and the competition in which he plays. Some competitions and teams are stronger than others. This may introduce a bias in the ratings as a player's skill may be over (under) estimated because his skill is rated relative to his less (more) talented teammates or opponents. To attempt to control for this, instead of comparing the absolute value of the skill rating, we look at changes in a player's skill rating between two consecutive years. The evolution similarity score first calculates a similarity score for each skill V_r as:

$$evolution_{V_r}(p, p', a_1, y) = \sqrt{\sum_{a=a_1-y+2}^{a_1} ((v_{r,p_a} - v_{r,p_{a-1}}) - (v_{r,p'_a} - v_{r,p'_{a-1}}))^2} \quad (3)$$

where v_{r,p_a} represents player p's observed rating for skill V_r at age a, and y represents how many years of past data the similarity metric should consider. Because the metric considers the change in skill between two consecutive years, the measure only looks at $y-1$ values when comparing each skill.

The total evolutionary score is computed by summing over all skill values:

$$evo_{tot}(p, p', a_1, y) = \sum_{V_r \in V} evolution\_score_{V_r}(p, p', a_1, y) \tag{4}$$

Then, the final score is computed by normalizing the total score relative to the range of similarity scores for all players in set S:

$$sim_{evo}(p, p', a_1, y) = 1 - \frac{evo_{tot}(p, p', a_1, y) - \min_{p'' \in S}(evo_{tot}(p, p'', a_1, y))}{\max_{p'' \in S}(evo_{tot}(p, p'', a_1, y)) - \min_{p'' \in S}(evo_{tot}(p, p'', a_1, y))} \tag{5}$$

where S is the set of similar players for p. This normalization maps the least similar player's score to 0 and the highest to 1. This similarity score no longer reflects the percentage of similarity. Instead, it can only be used to rank players according to their similarity. If player p has a higher evolutionary similarity to player p' than to player p'', we can conclude that player p is more similar to player p' than to player p''.

4.2 Prediction methods

We consider two ways to predict a player's future rating for a given skill: the absolute prediction and the evolutionary prediction.

Absolute prediction method The absolute prediction simply computes p's expected rating for a skill at age a_2 as a weighted average of the observed ratings for each similar player found in the data. Specifically, the predicted rating for skill V_r for player p at age a_2 is:

$$\hat{v}^{abs}_{r,p_{a_2}} = \frac{\sum_{p' \in S} sim(p, p', a_1, y) \times v_{r,p'_{a_2}}}{\sum_{p' \in S} sim(p, p', a_1, y)} \tag{6}$$

where sim is the chosen similarity metric, S is the set of similar players, and y represents how many years of data the similarity metric should consider.

Evolutionary prediction method Because players' skill levels can vary, an alternative idea is to consider p's current skill level as a baseline and predict how this will evolve over time. This can be done by adjusting p's current skill by a weighted average of the observed difference in rating for the skill at age a_2 and age a_1 for each similar player found in the data. Thus the predicted rating for skill V_r for a player p at age a_2 is:

$$\hat{v}^{evo}_{r,p_{a_2}} = v_{r,p_{a_1}} + \frac{\sum_{p' \in S} sim(p, p', a_1, y) \times (v_{r,p'_{a_2}} - v_{r,p'_{a_1}})}{\sum_{p' \in S} sim(p, p', a_1, y)} \tag{7}$$

5 Experiments

We now evaluate the predictive accuracy of the APROPOS projection system. Our goal is to evaluate the following four questions:

Q1 How well can APROPOS predict ratings one year in the future?
Q2 How does APROPOS' predictive accuracy vary with how far in the future it projects ratings for?
Q3 What is the effect of the number of years of data used to compute the similarity between two players on APROPOS' predictive performance?
Q4 How does the threshold used to identify similar players effect APROPOS' predictive performance?

5.1 Experimental setup

We compare four different systems:

Baseline: Given a prediction age a_2, the baseline finds all players of the same age and for each skill simply predicts the average rating over all players.
ABS-ABS This uses our absolute similarity metric and absolute prediction mechanism.
ABS-EVO This uses our absolute similarity metric and evolutionary prediction mechanism.
EVO-EVO This uses our evolutionary similarity metric and evolutionary prediction mechanism.

To run the experiments, we predict the potential ratings of 1000 players in the English and German competitions. For each player, we use data from 2012 and earlier to compute similarities and make predictions for year 2013 and onwards. We select this cutoff as it yields five years of data in both train and test sets, which allows us to vary how much data is used to identify similar players while also making predictions upto five years in the future. As an error metric, we report mean absolute error (MAE) which is an average over all players and all skills. Recall that each skill is scored from 0 to 100.

To evaluate the first question, we predict the ratings for 2013. We use three years to compute player similarities and the threshold for selecting the best players is set to 0.9. For the second question, we use an identical setup except we predict the results for each year in the period from 2013 to 2017 inclusive. For the third question, we predict the 2013 rating using a similarity threshold of 0.9 and vary the number of years used to compute player similarities from one to five. Finally, for the fourth question, we predict the 2013 rating using three years to compute player similarities and vary the threshold used to identify similar players from 0.7 to 0.9 in increments of 0.05.

5.2 Results

Figure 2 shows the mean absolute error for the predictions for 2013 to address question 1. Each model performs better than the baseline model. The baseline's MAE is 12.62 while the worst APROPOS model, ABS-ABS, has a MAE of 5.45. Using the evolutionary prediction method seems to result in more accurate predictions than the absolute prediction metric.

Figure 3 shows how the prediction period influences MAE. Unsurprisingly, the predictions that are farther in the future are worse than predictions that are less far away. Predictions five years in the future are about twice as bad as predictions one year in the future. However, the predictions are substantially more accurate than the baseline model.

Figure 4 shows how the number of years used to compute the similarity between players affects performance. Interestingly, this parameter only seems to have a limited effect on performance.

Finally, Figure 5 shows how the threshold used to identify similar players affects performance. On the ABS-ABS model this parameter has a strong effect, whereas on the other two models its effect is quite limited.

6 Conclusions & Future work

We presented a first approach for predicting the potential of professional soccer players. We developed and evaluated the APROPOS projection system which makes predictions for the potential using a k-nearest neighbours approach. We introduced multiple metrics to measure the similarities between players and multiple methods to predict player potentials leveraging the resulting similarities. Our best models predict the player potentials sufficiently accurate. The most influential parameter is the choice of the predictive method. The best model has a maximum mean absolute error of only 2.15 on 100.

Acknowledgements

Tom Decroos is supported by the KU Leuven Research Fund (C22/15/015) and FWO-Vlaanderen (G.0356.12). Jan Van Haaren was supported by the Agency for Innovation by Science and Technology in Flanders (IWT). Jesse Davis is partially supported by the KU Leuven Research Fund (C22/15/015) and FWO-Vlaanderen (G.0356.12, SBO-150033).

References

[1] Basketball Prospectus. URL: `http://www.basketballprospectus.com`, last checked on 2017-6-5
[2] CARMELO Projection System. URL: `https://fivethirtyeight.com/features/how-were-predicting-nba-player-career/`, last checked on 2017-6-5
[3] FiveThirtyEight (538). URL: `https://fivethirtyeight.com`, last checked on 2017-6-5

[4] SoFIFA. URL: https://sofifa.com, last checked on 2017-6-5
[5] Wikipedia: PECOTA. URL: https://en.wikipedia.org/wiki/PECOTA, last checked on 2017-6-5
[6] Wikipedia: Similarity score. URL: https://en.wikipedia.org/wiki/Similarity_score, last checked on 2017-6-5
[7] Alina Bialkowski, Patrick Lucey, Peter Carr, Yisong Yue, Sridha Sridharan en Iain Matthews: Large-Scale Analysis of Soccer Matches Using Spatiotemporal Tracking Data. Proceedings of the IEEE International International Conference on Data Mining (ICDM) (2014)
[8] Jan Van Haaren, Vladimir Dzyuba, Siebe Hannosset, and Jesse Davis: Automatically Discovering Offensive Patterns in Soccer Match Data. Advances in Intelligent Data Analysis XIV, pp.286-297 (2015)
[9] Martin Eastwood: Expected Goals and Support Vector Machines. URL: http://pena.lt/y/2015/07/13/expected-goals-svm/, last checked on 2017-6-5
[10] Patrick Lucey, Alina Bialkowski, Mathew Monfort, Peter Carr en Iain Matthews: Quality vs Quantity: Improved Shot Prediction in Soccer Using Strategic Features from Spatiotemporal Data. MIT Sloan Sports Analytics Conference (2015)
[11] Tom Decroos, Vladimir Dzyuba, Van Haaren en Jesse Davis: Predicting Soccer Highlights from Spatio-temporal Match Event Streams. Proceedings of the AAAI Conference on Artificial Intelligence (2017)

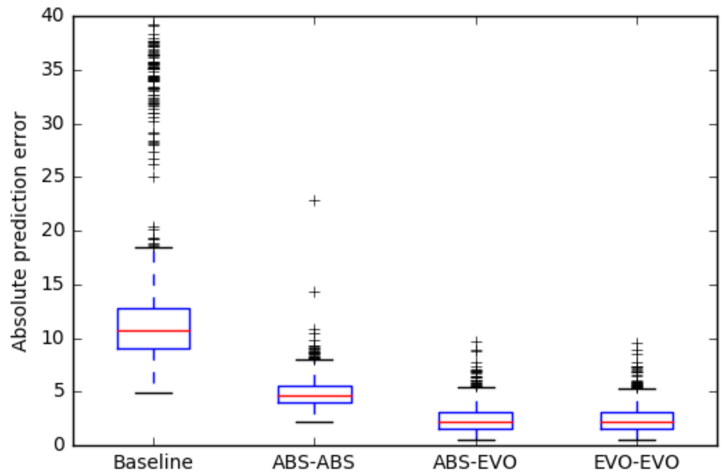

Fig. 2. The MAE for predicting skill ratings for 2013 using three years of data to compute similarity scores and 0.9 as threshold for identifying similar players.

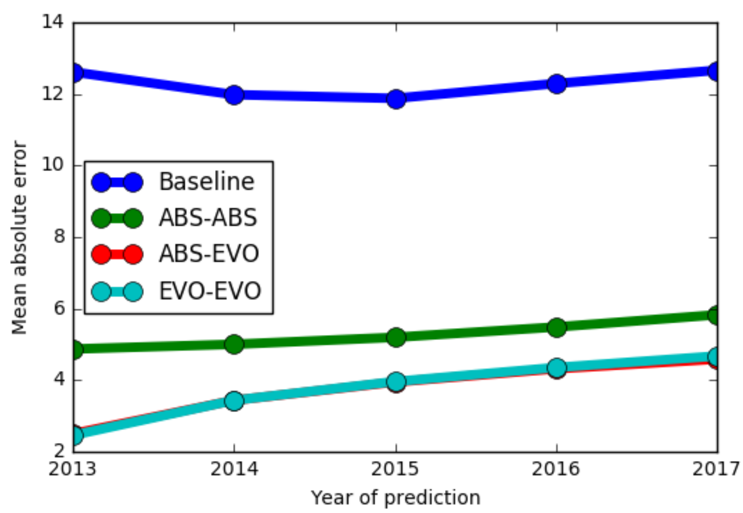

Fig. 3. The MAE for predicting skill ratings for 2013, 2014, 2015, 2016, and 2017 using three years of data to compute similarity scores and 0.9 as threshold for identifying similar players.

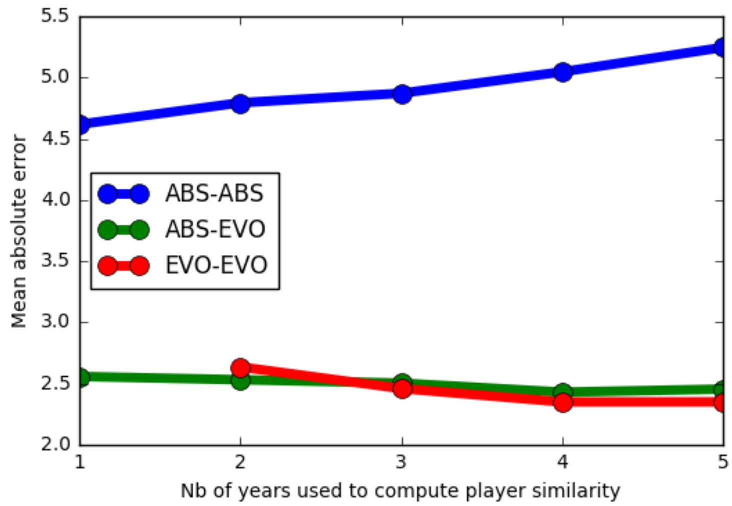

Fig. 4. The MAE for predicting skill ratings for 2013 using 0.9 as threshold for identifying similar players and varying the number of years of used to compute similarity scores from one to five.

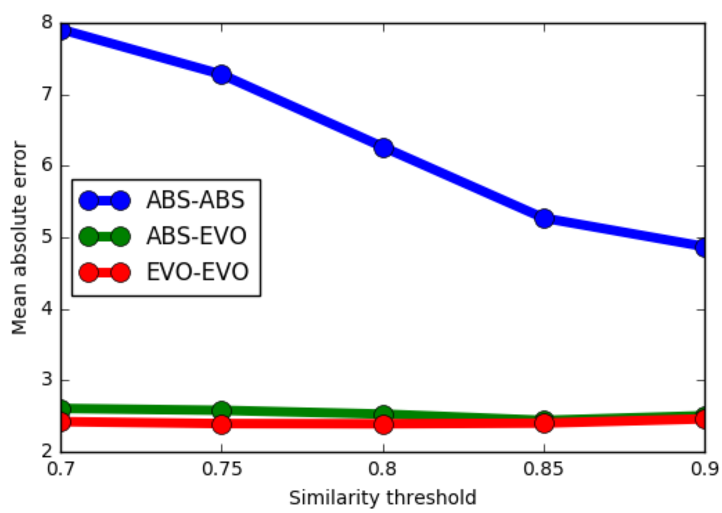

Fig. 5. The MAE for predicting skill ratings for 2013 using three years of data to compute similarity scores and varying the threshold used to identify similar players from 0.7 to 0.9