Explaining soccer match outcomes with goal scoring opportunities predictive analytics

Harm Eggels[1], Ruud van Elk[2], and Mykola Pechenizkiy[1]

[1] Eindhoven University of Technology,
P.O. Box 513, NL-5600 MB, Eindhoven, The Netherlands
h.p.h.eggels@student.tue.nl, m.pechenizkiy@tue.nl
[2] PSV, Eindhoven, The Netherlands

Abstract. In elite soccer, decisions are often based on recent results and emotions. In this paper, we propose a method to determine the expected winner of a match in elite soccer. The expected result of a soccer match is determined by estimating the probability of scoring for the individual goal scoring opportunities. The outcome of a match is then obtained by integrating these probabilities. In our experimental study, we show that the probabilities of goal scoring opportunities accurately match reality.

Keywords: soccer analytics, scoring opportunity, predictive modeling

1 Introduction

The use of advanced big data analytic in soccer starts showing its potential, however, sports analytics as a research area is still only emerging. Some of the problems are well-defined, e.g. many studies have attempted to predict the result of soccer matches before the match actually started. Various perspectives have been used to tackle this problem. A common perspective to look at this problem is the prediction of soccer matches from a betting perspective [1, 2]. Both machine learning approaches, e.g. an ensemble of k-nn predictors [3], and statistical approaches, e.g. considering goals scored by a team by Poisson processes [4, 5]. However, based on the reported results, the practical applicability of the obtained models is still rather limited.

Other problem formulations are less straightforward to formalize, e.g. providing insights into how well each of the teams or individual players did in the match. Here the approaches range from plotting histograms to heatmaps aligned with the playing field, and from counting successful actions to computing complex features based on domain knowledge.

In our work, we take a complementary perspective. We consider the use of predictive modeling to explain the outcome of a match based on the available data from the match (rather than trying to predict the outcome of the game before the game starts).

By explaining the match outcome we mean accumulating evidence of which team should have won the match based on the created goal scoring opportunities and accounting for both the quantity and quality of such opportunities. The demand for such an approach comes from soccer clubs themselves. These soccer clubs often base their decisions on recent results, even if they do not completely understand where these results come from.

In this paper, we provide an empirical illustration that inducing a predictor from the past soccer matches and applying it on the current match data provides us with accurate probabilistic estimates of scoring opportunities to result in goals.

2 Methods

An important aspect of this paper is to lay the sound foundation for reasoning about scoring opportunities. We shall be able to get insights into two kinds of questions: "how can we quantify the value of a shot given the scoring opportunity?" and "how can we quantify the value of a goal scoring opportunity created by a team (disregarding whether it was realized or not)?" If we can provide good estimates, then someone can see how many opportunities (and their quality) each of the teams produced during a match and how many of them each team realized.

In particular, if we get probability estimates for each scoring opportunity, we can simply sum up this estimates and get an expected number of goals as follows from the Poisson binomial distribution. Thus, if we denote p_i be the probability that we scored a goal in the scoring opportunity i, and model each i as a Bernoulli random variable $y_i \sim Ber(p_i)$ then the expected number of goals in the match of n scoring opportunities is equal to: $E[\#goals] = E(\sum_{i=1}^{n} y_i) = \sum_{i=1}^{n} E(y_i) = \sum_{i=1}^{n} p_i$.

The idea of applying predictive modeling for quantifying the quality of scoring opportunities is not new. E.g. logistic regression was used in [6] to determine the quality of individual goal scoring opportunities. In our approach, however, we make sure that scores we obtain can be treated as probabilities of scoring a goal in considered match situation. For this purpose, we ensured that our predictive model learned from the data has good generalization performance and has low variance.

Formally, we learned a classifier that is functional mapping $y = f(X)$, where for each scoring opportunity X_i that is a feature vector describing it and any additional contextual information, the classifier should accurately predict $y_i \in \{\text{goal}, \text{no goal}\}$. The classifier must generalize well to previously unseen scoring opportunities and avoid overfitting to the training data. We expect that a classification technique that is not only accurate (i.e. has low error, bias, and variance) but also can provide good confidence estimates for the predicted output would work best for our purpose.

One could argue that training the classifiers on only the best players of the world would lead to a more accurate and insightful model of *desirable* player performance. Such a model would, however, have only limited business value since it would not be directly applicable to poorer players. Adding the player quality to the model allows the model to learn the relations between player quality and the value of a goal scoring opportunity. The scores could be corrected accordingly. Furthermore, the quality of a goal scoring opportunity is influenced by the opposing team. Since the location of the defenders is already defined, the influence of the defenders is likely to be limited. The goalkeeper, however, could have significant influence on the quality of a goal scoring opportunity.

Consider two situations in which a player attempts to score with the only difference being the opposing goalkeeper. If one of these goalkeepers would be the best of the league and the other goalkeeper would be an average goalkeeper. Intuitively, these situations would not have the same probability of resulting in a goal. Therefore, the quality of the opposing goalkeeper taking into account.

Since goals are so rare in soccer, more non-goals than goals exist in the data set. Therefore, a combination of over-sampling (SMOTE) [13] and cluster based under-sampling [14] are used before applying classification algorithms to deal with class imbalance.

We also apply calibration to make the scores obtained with the classifier to be interpreted as a probability of scoring a goal given the scoring opportunity situation. The classification algorithms provide class membership probabilities, i.e. the confidence a sample belongs to a certain class. These class membership probabilities can not be interpreted as the probability that a goal attempt results in a goal. Calibrating the classifier ensures that its output can be interpreted as a probability that a goal attempt results in a goal. Two main calibration techniques exist: Platt's scaling [8] and Isotonic regression [9, 10]. Niculescu-Mizil and Caruana show that Platt scaling outperforms Isotonic regression when the data set is relatively small. When the size of the data set, however, increases (1000 samples or more) Isotonic regression outperforms Platt scaling [11]. Therefore, we use Isotonic regression.

With the use of classification algorithms and calibration techniques, point estimates can be determined for the goal scoring opportunities. In order to avoid misleading interpretation of the quality scores, we also estimate prediction intervals. To determine the standard deviation similar goal scoring opportunities, the samples are firstly clustered. The standard deviation of the samples in the cluster is then used to determine the prediction intervals. Gaussian Mixture clustering is used since this technique is often used in kernel density estimation. Intuitively, if the variation of the point estimate is too large, no valid statements about individual point estimates can be made. However, aggregating the data, however, still valid statements can be made due to the law of the large numbers.

3 Experimental Study

3.1 Data

We had access to three different data sources are available: 1) data about the main events during a match tracked by (employees of) ORTEC; 2) data about the quality of players [12]; the data from the soccer game FIFA is extracted from the web; and 3) spatiotemporal data about players tracked by Inmotio during matches with the help of cameras.

In total, we have data from seven different leagues over three seasons. This leads to a total of 5017 matches in which 128667 goal attempts were performed. Of these goal attempts, only 14109 resulted in a goal.

It is worth noting that each data source has its own data quality problems that can affect classifiers and the conclusions derived from their outputs. The data quality issues of the ORTEC data are related to the tracking of the events by ORTEC employees who have to select the location of the event at the right time and at the right location that is hard to do, especially in a near real-time setting. The main data quality issue with the FIFA data comes from the determination of the stats that is somewhat vague and could be incorrect for some of the players. Finally, the data quality issues of the Inmotio data come from the cases in which the cameras lose the correct player or accidentally selects the wrong player. In this case, the location of the player is incorrect and it is difficult get the correct location. We had too little data from the cameras and hence did not use it for inducing classifiers.

With the use of the considered data sources, various features can be extracted. A list of the extracted features for each data source is provided in Table 1.

Table 1: Features for the data sources

ORTEC	FIFA	Inmotio
Context	Player quality	Number of attackers in line
Part of body	Goal keeper quality	Number of defenders in line
Dist to goal		Distance nearest defender in line
Angle to goal		Distance goal keeper
Originates from		

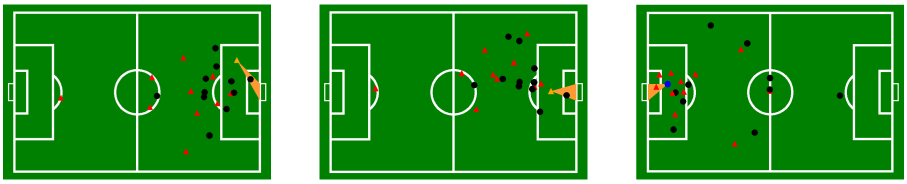
4 Results

We experimented with four different classification techniques algorithms (as implemented in scikit-learn [7]): Logistic Regression, Decision tree, Random Forrest, and a decision tree boosted with Ada-boost. Inner 10-Fold cross validation is used for parameter selection and generalization performance estimation. In the inner fold, the best parameters are selected. The generalization performance is then computed in the outer fold.

Figure 1 illustrates two examples of probabilities of the individual goal scoring opportunities obtained with our approach.

The features for these examples and the probabilities are shown in Table 2.

Since we want our classifiers to provide higher scores for better goal scoring opportunities, we report AUC performance, but also provide the precision, recall,



(a) Example 1 (b) Example 2 (c) Example 3

Fig. 1: Three examples of goal scoring opportunities

Table 2: Features and probabilities for the three events

Feature	Example 1	Example 2	Example 3
Context	Open-play	Open-play	Open-play
Part of body	Foot	Foot	Foot
Dist to goal	10.488	11.652	8.408
Angle to goal	0.351	0.648	0.724
Originates from	Pass	Pass	Pass
Current scoreline	In front	In front	Behind
Player quality	72	35	60
GK quality	77	77	77
Goal	Yes	Yes	No
Probability	0.125	0.228	0.276
Prediction Interval	[0, 0.191]	[0.011, 0.445]	[0.124, 0.428]

and F-score value for reference. Table 3 summarizes the results. Also the standard deviation of the AUC over the different cross validation phases is provided. We can see from the table that Random Forest performs reasonably well and outperforms other classifiers.

Table 3: Performance of the classification algorithms

Classifier	Precision	Recall	F-score	AUC	Std. dev. AUC
Random Forrest	0.785	0.822	0.800	0.814	0.053
Decision Tree	0.698	0.678	0.676	0.677	0.142
Logistic Regression	0.715	0.650	0.673	0.697	0.082
Ada-boost	0.624	0.773	0.688	0.670	0.069

Next, we perform the calibration step to make the scores more accurate. We used the reliability graph introduced in [11] that show how close the predicted values are to the actual ratios of goals scored. The obtained reliability graph is shown in Figure 2.

Figure 2 shows that the predicted values are indeed close to the actual ratios of goal scoring opportunities resulting in goals. We also show the confidence intervals of the predicted values in the bins. Since these confidence intervals

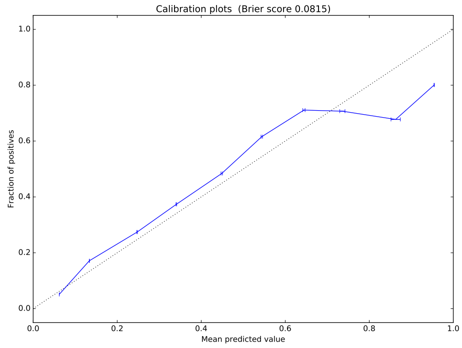


Fig. 2: Calibration plot

are narrow, it is safe to consider the scores as the probability of a goal scoring opportunity to result in a goal.

We also report the Brier score [15] that determines the accuracy of probabilistic predictions of a set of mutually exclusive discrete outcomes. The Brier score of the calibrated scores is pretty low as well, which supports conclusions drawn from the reliability graph.

To determine whether the expected goals can be used to explain match results, the predicted match outcomes are compared to the actual match outcomes. Table 4 provides the number of matches that are correctly predicted with the use of the expected goals model. Furthermore, the number of games where the number of predicted goals was at most one goal off is shown, the number of matches in which the result was correct (win team 1, draw, win team 2). Finally, the Mean Squared Error (MSE) for the number of goals per match is provided.

What stands out from Table 4 is that in only 1366 of the 5020 matches, the exact score of the match was predicted based on the expected goals. If, however, one goal difference is accepted, 3443 of the 5020 matches have correctly predicted scores. Therefore, it seems that the expected goals model is, in most cases, almost correct. The MSE Match strengthens this statement. The MSE match shows that the average MSE of the result of a match is 2.366. Therefore, the average number of goals predicted difference goals of both teams differs $\sqrt{2.366} \approx 1.538$ from the actual difference in goals.

So far, just the exact results are examined. Maybe even more interesting, is how often the expected goals model predicted the correct winner. This is given by the number of correct results in Table 4. Obviously, the number of correctly predicted matches is higher than the correctly predicted scores. What stands out, however, that the number of correctly predicted matches is not close to the number of scores predicted correctly where one goal difference was allowed. This shows that games where the model is one goal off in the match, this one goal also influences the result of the match. To evaluate in which cases the one goal difference most often influences the result, the problem of predicting the winner of a match is defined as a three class problem where either Team 1 wins, Team 2

Table 4: Evaluation of match outcomes according to the expected goals model

League	Season	#Matches	#Correct	#1 Goal	#Result	MSE Match
1.Bundesliga	2015-2016	2	1	1	1	1.655
Premier League	2015-2016	1	0	0	0	2.289
Champions League	2015-2016	133	37	96	74	1.956
Eredivisie	2013-2014	322	86	221	175	2.556
	2014-2015	322	81	222	177	2.395
	2015-2016	322	84	222	178	2.284
Jupiler League	2013-2014	380	95	239	202	2.902
	2014-2015	379	94	255	215	2.517
	2015-2016	342	88	208	185	2.762
KNVB Beker	2015-2016	26	6	17	16	2.601
Ligue 1	2013-2014	380	105	257	183	2.106
	2014-2015	380	106	289	189	2.172
	2015-2016	380	101	272	198	2.153
Primeira Liga	2014-2015	280	92	202	165	1.948
	2015-2016	231	58	150	123	2.426
Primera Division	2013-2014	380	118	271	220	2.471
	2014-2015	380	112	254	207	2.153
	2015-2016	380	102	267	201	2.385
Total		5020	1366	3443	2709	2.366

wins or the game ends in a draw. The confusion matrix of the tree-class problem is provided in Table 5.

Table 5: Confusion matrix for the match winner prediction

		Actual			
		Win 1	Draw	Win 2	Total
	Win 1	1079	329	210	1618
Predicted	Draw	599	524	591	1714
	Win 2	220	362	1106	1688
Total		1898	1215	1907	5020

Table 5 shows that most of the incorrect classified match outcomes originate from predicted draws. Predicted draws are, most likely, games which were very tight. Table 4 already showed that in many cases, the model was only one goal off. In tight games, one goal off means that the result of the match is predicted incorrectly. This was most likely the case of the predicted draws.

5 Potential Applications for Soccer Clubs

We consider three immediate applications of the quality scores by soccer clubs: 1) performance evaluation over a given period of time; 2) analysis of matches and team performance; and 3) assessment of players and individual training sessions management.

5.1 Performance Management

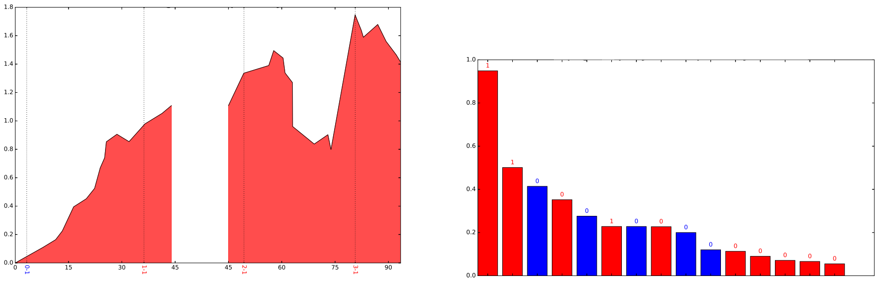
Soccer clubs tend to base decisions on results of a short period of times and emotions. Since many factors influence the results of soccer matches, these results could not closely match the reality. These decisions could, therefore, be based on misperceptions. A more objective metric of the quality of goal scoring opportunities would provide a more objective decision-making strategy. Before firing staff, for example, an expected league table could be created to determine whether the team is actually performing badly.

Furthermore, the results of matches could be plotted together with injuries, suspensions, fired coaches and many more factors to find out the relation of the events that happened during a season on the performance of the team.

5.2 Match Analysis

Goals are very rare in soccer that leads to high influence of a single goal on the result of a match. By analyzing goal scoring opportunities instead of actual goals scored, a more objective way of analyzing the result is obtained. Adding the quality of the goal scoring opportunities makes this analysis even better.

Figure 3 provides an illustrative example of simple visualization of match scoring opportunities and whether they were realized. On the left side of this figure, the progress of a match in terms of expected goals is provided. Here, one could see that the upper team (represented by red), should have been in front from the beginning of the match and had the upper hand during the match. The right side of the figure shows from which players the expected goals (the bars) and the actual goals (the numbers) were coming. Since this data is classified, the names of the players on the x-axis are removed.



(a) Progress of a match over time

(b) Expected goals per player

Fig. 3: Match goal scoring opportunities

Furthermore, similar to the analysis for periods of time, the expected result of a match could be plotted over time. By adding important events such as goals scored, cards, substitutions and many more factors, the influence of these factors could be researched in more detail.

5.3 Player Evaluation, Training, and Acquisition

A major advantage of determining the quality of the individual goal scoring opportunities is that it generates more possibilities than only determining match results. Aggregating over players, instead of matches, leads to insights into player performance. These insights could be used to evaluate players, adjust training programs or perform player acquisition.

An example of interesting insights from the expected goals is when the expected goals are plotted for different locations on the field. This could, for example, show that a player is often shooting from one specific part of the field but never scores. If the probabilities of these goal scoring opportunities are high, the player is obviously doing something wrong in these cases and his actions could be analyzed in more detail. If, however, the probabilities are low for a player on the field, but that player shoots very often, someone could point out to him that shooting might not be the best decision at that part of the field.

Another example comes from the case where players, especially strikers, score many goals in one season (take for example Jamie Vardy of Leicester City during the English Barclays Premier League in 2015-2016). Those strikers are often bought by big clubs since they did score a lot. It could, however, be the case that such a player did score a lot but had a much lower amount of expected goals. This could suggest that the specific player was lucky during that season. Of course, more research has to be performed on that player's performance, but the expected goals indicator could be a useful tool in player acquisition.

6 Conclusion

In this paper, we presented a method with which the results of soccer matches can be described in a more objective manner by evaluating the quality of the goal scoring opportunities for both teams during that match. It is shown that the proposed method performs well in terms of classification performance as well as on calibration of probabilistic estimates. Further applications of the expected goals are given by evaluating seasons, matches, and individual players.

An important point to make when using the probability estimates of the goal scoring opportunities is that these estimates may have a high standard deviation. The scores for goal scoring opportunities, could, therefore vary quite a bit, even thought the goal scoring opportunities are similar. Users should, therefore, be very careful when making statements of individual goal scoring opportunities with too few point estimates.

References

1. A.C. Constantinou, N.E. Fenton, M. Neil, ”Profiting from an inefficient Association Football gambling market: Prediction, Risk and Uncertainty using Bayesian networks.” Knowledge-Based Systems 50 (2013): 60-86.
2. H. Langseth, ”Beating the bookie: A look at statistical models for prediction of football matches.” SCAI. 2013.
3. V. Hoekstra, P. Bison, G. Eiben, ”Predicting football results with an evolutionary ensemble classifier.” (2012).
4. D. Karlis, I. Ntzoufras, ”Analysis of sports data by using bivariate Poisson models.” Journal of the Royal Statistical Society: Series D (The Statistician) 52.3 (2003): 381-393.
5. A. Heuer, C. Mueller, O. Rubner, ”Soccer: Is scoring goals a predictable Poissonian process?.” EPL (Europhysics Letters) 89.3 (2010): 38007.
6. P. Lucey et al. ”quality vs quantity: Improved shot prediction in soccer using strategic features from spatiotemporal data.” Proc. 8th Annual MIT Sloan Sports Analytics Conference. 2014.
7. F. Pedregosa et al. ”Scikit-learn: Machine learning in Python.” Journal of Machine Learning Research 12.Oct (2011): 2825-2830.
8. J. Platt, ”Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods.” Advances in large margin classifiers 10.3 (1999): 61-74.
9. B. Zadrozny, C. Elkan, ”Obtaining calibrated probability estimates from decision trees and naive Bayesian classifiers.” ICML. Vol. 1. 2001.
10. B. Zadrozny, C. Elkan. ”Transforming classifier scores into accurate multiclass probability estimates.” Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2002.
11. A. Niculescu-Mizil, C. Rich, ”Predicting good probabilities with supervised learning.” Proceedings of the 22nd international conference on Machine learning. ACM, 2005.
12. K Stuart. Why clubs are using football manager as a real-life scouting tool. The Guardian, 2014.
13. N.V. Chawla et al. ”SMOTE: synthetic minority over-sampling technique.” Journal of artificial intelligence research 16 (2002): 321-357.
14. S.J. Yen, L. Yue-Shi, ”Cluster-based under-sampling approaches for imbalanced data distributions.” Expert Systems with Applications 36.3 (2009): 5718-5727.
15. G.W. Brier, ”Verification of forecasts expressed in terms of probability.” Monthly weather review 78.1 (1950): 1-3.
16. Anderson, Chris, and David Sally. The numbers game: Why everything you know about football is wrong. Penguin UK, 2013.