<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0047678>

@article{heuer2012does,

title={How does the past of a soccer match influence its future? Concepts and statistical analysis},

author={Heuer, Andreas and Rubner, Oliver},

journal={PloS one},

volume={7},

number={11},

pages={e47678},

year={2012},

publisher={Public Library of Science}

}

**Abstract:** Scoring goals in a soccer match can be interpreted as a stochastic process. In the most simple description of a soccer match one assumes that scoring goals follows from independent rate processes of both teams. This would imply simple Poissonian and Markovian behavior. Deviations from this behavior would imply that the previous course of the match has an impact on the present match behavior. Here a general framework for the identification of deviations from this behavior is presented. For this endeavor it is essential to formulate an a priori estimate of the expected number of goals per team in a specific match. This can be done based on our previous work on the estimation of team strengths. Furthermore, the well-known general increase of the number of the goals in the course of a soccer match has to be removed by appropriate normalization. In general, three different types of deviations from a simple rate process can exist. First, the goal rate may depend on the exact time of the previous goals. Second, it may be influenced by the time passed since the previous goal and, third, it may reflect the present score. We show that the Poissonian scenario is fulfilled quite well for the German Bundesliga. However, a detailed analysis reveals significant deviations for the second and third aspect. Dramatic effects are observed if the away team leads by one or two goals in the final part of the match. This analysis allows one to identify generic features about soccer matches and to learn about the hidden complexities behind scoring goals. Among others the reason for the fact that the number of draws is larger than statistically expected can be identified.

* Apenas exploratório;
* Probabilidade de gol aumenta com o tempo da partida;
* Três desvios de um processo de taxa simples podem existir, a taxa de gols pode depender:

1. Do tempo exato dos últimos gols;
2. Do tempo passado desde o último gol;
3. Do placar atual.

* Data set: partidas de Bundesliga das temporadas 1968/69 até 2010/11 com exceção da temporada 1992/93.

<https://arxiv.org/abs/1906.05029>

@article{robberechts2019will,

title={Who will win it? An in-game win probability model for football},

author={Robberechts, Pieter and Van Haaren, Jan and Davis, Jesse},

journal={arXiv preprint arXiv:1906.05029},

year={2019}

}

**Abstract:** In-game win probability is a statistical metric that provides a sports team's likelihood of winning at any given point in a game, based on the performance of historical teams in the same situation. In-game win-probability models have been extensively studied in baseball, basketball and American football. These models serve as a tool to enhance the fan experience, evaluate in game-decision making and measure the risk-reward balance for coaching decisions. In contrast, they have received less attention in association football, because its low-scoring nature makes it far more challenging to analyze. In this paper, we build an in-game win probability model for football. Specifically, we first show that porting existing approaches, both in terms of the predictive models employed and the features considered, does not yield good in-game win-probability estimates for football. Second, we introduce our own Bayesian statistical model that utilizes a set of eight variables to predict the running win, tie and loss probabilities for the home team. We train our model using event data from the last four seasons of the major European football competitions. Our results indicate that our model provides well-calibrated probabilities. Finally, we elaborate on two use cases for our win probability metric: enhancing the fan experience and evaluating performance in crucial situations.

* Busca calcular a probabilidade de vitória/empate/derrota;
* Abordagens de outros esportes não funcionam bem no futebol;
* Propõe um modelo Bayesiano treinado com Auto-Differentiation Variation Inference;
* Divide o jogo em T=100 time frames em que T=50 é o intervalo e acréscimos podem ser estimados com random forest;
* Variáveis:
  1. **Base features:** Game Time e Score Differential;
  2. **Team strength features:** Elo Rating Differential;
  3. **Contextual features:** Team Goals, Yellows, Reds, Attacking Passes (média do número de passes certos no terceiro terço de campo nas últimas 10 time frames) e Duel Strength (média do percentual de duelos vencidos nas últimas 10 time frames).
* Data set: obtidos do Wyscout event stream data das primeiras divisões das ligas da Inglaterra, Espanha, Alemanha, Itália, França, Holanda e Bélgica das temporadas 2014/15, 2015/16, 2016/17 e 2017/18. As três primeiras temporadas foram utilizadas para treinar os modelos e a última foi utilizada para teste. Treino de 5967 partidas e teste de 2227 partidas.
* Added Goal Value: métrica que leva em conta o valor de cada gol para a probabilidade de vitória do time e permite avaliar os jogadores mais “clutch”.

<https://ntnuopen.ntnu.no/ntnu-xmlui/bitstream/handle/11250/2625819/no.ntnu:inspera:2439616.pdf?sequence=1>

@mastersthesis{haugaard2019predicting,

title={Predicting live matches},

author={Haug{\aa}rd, Mii Erik Samyo and Vu, Kim Long},

year={2019},

school={NTNU}

}

**Abstract:** Football betting has increased in popularity over the past years. Several studies have attempted to predict the outcome of matches, both for fun and in an attempt to gain profit by using the predictions for betting. Most previous works have tried to predict the outcome of a match, prior to its start. This project will, on the other hand, focus on making predictions in an ongoing match. In this project, neural networks are used in order to predict the next goal-scoring team in a live football match. Different set of features are used to detect which have the best predictive capabilities. These predictions are combined with several betting strategies, to decide if a bet should be placed or not, and how much to potentially bet, in an effort to generate a profit. Through the project, we have achieved results that show it is possible to achieve high accuracy when predicting the next goal-scoring team in a football match. Further, the results show that using these predictions in order to beat the bookmakers can be achievable.

* Busca prever a próxima equipe a marcar um gol;
* Combina as previsões com diferenças estratégias para apostar;
* Data set: segunda metade das últimas três temporadas das ligas da Inglaterra, Espanha, Itália, França e Alemanha obtido na api do Sportradar, 1830 partidas no treino, 353 no teste e 20% do teste utilizado para validação;
* Utiliza RNN e feedforward network, a primeira obteve os melhores resultados;
* Variáveis utilizadas: xG (Expected Goals, métrica do Understat), EGRT (calculada com as forças de ataque e defesa das equipes) e eventos da partida;
* xG e EGRT foram as features mais importantes;
* Utilizar apenas partidas da Premier League no treino obteve melhores resultados para prever partidas desta liga do que treinar com todo o data set disponível.

<https://www.dropbox.com/s/k0es29ecdb8v9pe/Prosjektoppgave.pdf?dl=0>

**Abstract:** Football betting has increased in popularity over the past years. With the increased computation power we have today and the emergence of machine learning, analyzing sports-event data is more available and common than it has previously been. The analyzed data can be used by the athletes and teams to gain advantages over their opponents, but also by ordinary persons that try to beat the bookies. In order for the bookies to stay in front, their models need to be better than the bettors.

In this project, we explore what has been done in the field of predicting the outcome of football matches, in order to attempt building a better model. We consider what has been done with regards to methods that have been used, but also what kind of inputs been fed into the models and their importance. Previous work has shown good results with the use of neural networks when trying to predict the outcome of a match, prior to its start. We also explore the effect different match events has shown to have on the outcome of a football match.

We will try to build a model that is able to predict the outcome of an ongoing football match, instead of trying to predict pre-game. This includes finding out who will win the rest of the match (e.g. the last 30 minutes). The main focus is to predict matches in the English Premier League, where historic data is provided by Sportradar.

* Comenta os seguintes artigos que buscam fazer previsões em tempo real:

Boice (2018) que utiliza Poissons;

* 1. Petterson & Nyquist (2017) que utiliza RNN.

<https://fivethirtyeight.com/features/how-our-2018-world-cup-predictions-work>

* Busca fornecer previsões atualizadas em tempo real para a Copa do Mundo 2018;
* Utiliza SPI ratings: estimativas das forças gerais das equipes e podem ser divididas em ratings de ataque e de defesa;
* Previsão das partidas:
  1. Poissons independentes para cada equipe utilizando os SPI ratings;
  2. Matriz dos resultados possíveis;
  3. Chance de empate aumentada;
* Previsão ao vivo:
  1. Em qualquer momento da partida, calcula-se o número esperado de gols de cada equipe utilizando os SPI;
  2. Poissons independentes são geradas com estes valores e a matriz de resultados das Poissons é combinado com o placar atual do jogo;
  3. Taxa de gols aumenta ao decorrer da partida;
  4. Estima-se acréscimos com número de cartões e se a partida está acirrada;
  5. Considera cartões vermelhos durante a partida no cálculo da força das equipes;
  6. Times mais fortes tendem a marcar gols numa taxa maior quando estão atrás no placar;
  7. Utiliza xG para alterar os gols esperados.

<https://odr.chalmers.se/bitstream/20.500.12380/250411/1/250411.pdf>

@mastersthesis{nyquist2017football,

title={Football match prediction using deep learning},

author={Nyquist, Robert and Pettersson, Daniel},

year={2017}

}

**Abstract:** In this thesis, the deep learning method Recurrent Neural Networks (RNNs) has been investigated for predicting the outcomes of football matches. The dataset consists of previous recorded matches from multiple seasons of leagues and tournaments from 63 different countries and 3 tournaments that include multiple countries. In the thesis work, we have studied several different ways of forming up input data sequences, as well as different LSTM architectures of RNNs that may lead to effective prediction, along with LSTM hyper-parameter tuning and testing. Extensive tests have been conducted through many case studies for the prediction and classification of football match winners. Using the proposed LSTM architectures, we show that the classification accuracy of the football outcome is 98.63% for many-to-one strategy, and 88.68% for manyto-many strategy. The prediction accuracy starts from 33.35% for many-to-one and 43.96% for many-to-many, and is increasing when more information about a match from longer time duration of data sequence is fed to the network. Using the full time data sequence, the RNN accuracy reached 98.63% for many-to-one, and 88.68% for many-to-many strategy. Our test results have shown that deep learning may be used for successfully predicting the outcomes of football matches. For further increasing the performance of the prediction, prior information about each team, player and match would be desirable.

* Utiliza LSTMs para prever o resultado de partidas;
* Considera line ups e técnicos das equipes além de gols, assistências, cartões, substituições e pênaltis;
* Testa diferentes tipos de inputs para as RNN mas poderia explicar melhor o input dos dados;
* 35234 partidas entre 2015 e 2017 de diversas ligas ao redor do planeta;
* Apresenta previsões de 15 em 15 minutos em que, obviamente, a acurácia cresce conforme o tempo restante de jogo diminui;
* Exibe alguns estudos de caso com diferentes configurações das RNN;
* Comenta acurácias de sites de apostas, jornais e aplicativos de celular com o objetivo de tentar mensurar a acurácia de humanos.

<http://wwwf.imperial.ac.uk/~ejm/M3S4/Problems/football.pdf>

@article{dixon1998birth,

title={A birth process model for association football matches},

author={Dixon, Mark and Robinson, Michael},

journal={Journal of the Royal Statistical Society: Series D (The Statistician)},

volume={47},

number={3},

pages={523--538},

year={1998},

publisher={Wiley Online Library}

}

**Abstract:** Data from over 4000 recent association football (soccer) matches from the main English competitions show clear evidence that the rate of scoring goals changes over the course of a match. This rate tends to increase over the game but is also influenced by the current score. We develop a model for a soccer match that incorporates parameters for both the attacking and the defensive strength of a team, home advantage, the current score and the time left to play. This model treats the number of goals scored by the two teams as interacting birth processes and shows a satisfactory fit to the data. We also investigate football cliches and find evidence that contradicts the cliche that a team is more vulnerable just after it has scored a goal. Our model has applications in the football spread betting market, where prices are updated during a match, and may be useful to both bookmakers and bettors.

* Utiliza processos de nascimento para modelar os gols em tempo real;
* Apenas usa o histórico dos placares das equipes e o tempo em que cada gol foi marcado;
* Descreve os modelos de Dixon & Coles (1997) e Maher (1982) e também utiliza a força de equipes e mando de campo;
* Apresenta alguns modelos, o melhor deles tem a intensidade dos gols crescendo ao longo da partida e variando de acordo com o placar atual;
* Exibe um exemplo de aplicação deste modelo no mercado de apostas.

<https://www.tandfonline.com/doi/full/10.1080/03610918.2017.1348518>

@article{zou2018discrete,

title={A discrete-time and finite-state Markov Chain model for association football matches},

author={Zou, Qingrong and Li, Qi and Guo, Hao and Shi, Jian},

journal={Communications in Statistics-Simulation and Computation},

volume={47},

number={8},

pages={2476--2485},

year={2018},

publisher={Taylor \& Francis}

}

**Abstract:** A birth process model proposed by Dixon and Robinson has been widely used in football spread betting market. However, multiple goals in a minute are permitted in the model, which does not conform to historical record. Moreover, it is difficult to calculate the outcome probability of the process accurately. The article presents a discrete-time and finite-state Markov chain model for real-time forecast of football matches and a recursive algorithm is derived to calculate the outcome probability accurately. The empirical study shows that the proposed model outperforms the models of Dixon and Robinson and Dixon and Coles.

* Apresenta um modelo de cadeias de Markov em tempo discreto e o compara com o modelo de Dixon e Robinson;
* Assume que até 3 gols podem ser marcados no minuto 90, 2 gols no minuto 45 e 1 gol nos outros minutos;
* Exibe um algoritmo para o cálculo das previsões que é bem mais rápido do que as simulações de Monte Carlo;
* O modelo sem os parâmetros do placar atual apresentou os melhores resultados;
* Dataset: temporadas 2009-10, 2010-11, 2012-13, 2013-14 da Premier League e da Bundesliga.