Feature definition for NBA result prediction through Deep Learning

Definizione delle features per la predizione dei risultati nella NBA tramite il Deep Learning

Manlio Migliorati, Eugenio Brentari

Abstract This contribution is focused on features' definition for the outcome prediction of matches of NBA basketball championship. It is shown how models based on one a single feature (Elo rating or the relative victory frequency) can have a quality of fit better than models using box-score predictors (e.g. the Four Factors). Features have been ex ante calculated for a dataset containing data of 16 NBA regular seasons, paying particular attention to home court factor. Models have been produced via Deep Learning, using cross validation.
Abstract *Questo contributo è focalizzato sulla costruzione di predittori per la predizione dei risultati degli incontri del campionato di basket NBA. In particolare si mostra come modelli basati su un unico predittore (Elo rating o la frequenza relative delle vittorie) possono avere una qualità di fit superiore a quella dei modelli basati sui box-scores (ad esempio i Four Factors). I predittori sono stati calcolati ex-ante su un dataset che comprende i dati di 16 regular seasons del campionato NBA, facendo particolare attenzione al fattore campo. I modelli sono stati prodotti tramite Deep Learning, applicando la cross-validation.*

Key words: basketball outcome prediction, features definition, court factor

1 Introduction

This contribution is focused on features selection for the problem of predicting the winner in NBA matches. It is shown how, for outcome prediction classification problem, a careful definition of single features used in model definition, can produce predictions with a quality better than quality of models built on the top of box-score statistics.

University of Brescia, Department of Economics and Management, Contrada S. Chiara 50
e-mail: manlio.migliorati@unibs.it
e-mail: eugenio.brentari@unibs.it

To this purpose, two features directly quantifying strength of teams involved in a match have been selected:

1. The Elo (from the name of its creator) rating system [2], originally defined for rating chess players and today widely used in several domains.
2. The difference of the relative frequency of victories for the two teams.

and used as covariates to build models to be compared, in terms of quality of fit, to models built using Oliver's Four Factors [5, 4] (few indexes synthesizing several *box-score* statistics) as regressors.
The models built in this work have been developed in a particular Deep Learning echosystem in R based on `Keras` package [1].

2 Features' definition

2.0.1 The Elo rating

The Elo rating system [2] has been originally defined for calculating the strength of players in zero-sum games (i.e. games where a player gains exactly what its opponent loses) as chess, the sport for which this system was created by Arpad Elo. More formally: if before a match Player1 has a rating $R1$ and Player 2 has a rating $R2$, and if S is the result of the match (where 1 means Player1 victory and 0 Player2 victory), after the match, the ratings of the 2 players will be updated as follows:

$$R1^{'} = R1 + K * (S - P(p1w)) \tag{1}$$

$$R2^{'} = R2 + K * (S - P(p2w)) \tag{2}$$

where K is a parameter addressing how strongly a result will affect ratings' update and $P(p1w)$ and $P(p2w)$ are the probabilities of victory (modelled as logistic curves) attributed to the two players before the match. The difference in Elo ratings between the two teams fighting in a match will be the first feature used in the present study: for the initial ratings we will follow [6] and 1300 will be used.

2.0.2 The difference in relative victory frequencies

A second feature directly quantifying the strength of opposing teams is the difference of their relative victory frequencies, named *diff* in the following. It can be formally defined as follows:

$$diff = \frac{won\_matches_{ht}}{played\_matches_{ht}} - \frac{won\_matches_{at}}{played\_matches_{at}} \tag{3}$$

Where the subscript *ht* means home team, and the subscript *at* mean away team. *Diff* statistics ranges from -1 to 1, where value 1 means that the home team is absolutely the strongest between the two teams. So, *diff* is a clear and concise way for showing the difference in class between the two teams, providing an analytical definition for a classic rule of thumb often used in naive fan predictions (the favorite is the team that won more in the (recent) past).

2.0.3 Four Factors

The Four Factors [5, 4] is a set of indexes built on top of classic *box-score* statistics. Four Factors are considered fundamental for winning a match, and summarize the attitude of a team with respect to shooting, turnovers, rebounds and free throws.

3 The dataset

The dataset includes data about 16 NBA regular seasons (from 2004-2005 to 2019-2020), counting more than 18.000 observations (one for each match). Elo, $diff$ and Four Factors have been calculated ex ante, i.e. considering only information from prior matches, to make them suitable for outcome predictions, taking into account:

- the periodicity, considering both the historical (considering all prior games) and the dynamic perspective (averaging on a subset of prior matches). Moreover, the mechanism of regression to mean [3] has been implemented for historical features, seeming particularly suitable for NBA [6], where at the end of each season there is an attempt to rebalance teams strength.
- the court where matches have been played: besides features usually calculated considering all matches, two new statistics based considering only either home or away data (called *the court issue* in the following) will be calculated, too.

4 Methods and Models: Deep Learning

4.1 Building Deep Learning models

All the models described in this work share the same sequential structure:

- one first input layer, with a number of input units corresponding to the number of features to be considered in building the model (1 for Elo and $diff$, 8 for Four Factors (4 for each team)
- one final output layer, with 1 output unit corresponding to the two possible results of a NBA match (basketball outcome prediction is a typical classification problem)

- a stack of several intermediate hidden sequential layers, connecting the input and output layers. Each hidden layer contains several elaboration units, to work on data received from the prior layer before sending them to the following layer.

The nets, calibrated to produce models with a good prediction quality, are built considering the two hyperparameters (i.e. the number of layers and the number of units for each layer) small in size, a natural consequence of the small number of features.

5 Results

The results reported in this section have been obtained using a v-fold cross-validation with v=4.

5.1 Using Elo features

Execution results for models based on Elo variants are reported in Table 1. The quality of predictions for models built using historical Elo without considering the court issue is the best one, with an AUC equal to 0.7117 and an accuracy equal to 0,6721 (using a threshold equal to 0.5047). These values have been obtained using a regression to mean percentage *P%* equal to 20.
Between the models built using dynamic Elo, the model not considering the court issue, obtained with a depth equal to two, is the best one: its AUC is equal to 0.7117 and its accuracy equal to 0.6736 (threshold equal to 0.5049), the best among the models we built in this work. Also predictions' quality for the model built using dynamic Elo considering the court issue, obtained with a depth equal to three, is good, with an AUC equal to 0.7103 and an accuracy equal to 0.6705 (threshold equal to 0.5148).

Table 1 Best quality of predictions for models based on Elo. For each variant, the best AUC measure, the corresponding threshold and the accuracy measure are reported, together with parameters' values used in Elo calculation

periodicity	court issue	AUC	threshold	accuracy	regression to mean P%
historical	not considered	0.7117	0.5047	0.6721	20
historical	considered	0.7001	0.5058	0.6650	60
periodicity	court issue	AUC	threshold	accuracy	depth
dynamic	not considered	0.7117	0.5049	0.6736	2
dynamic	considered	0.7103	0.5148	0.6705	3

5.2 Using $diff$ features

Results are reported in Table 2. The quality of predictions of the model built using $diff$ without considering the court issue is the best one, with an AUC equal to 0.6925 and an accuracy equal to 0.6626 (using a threshold equal to 0.5236). For the model built using dynamic $diff$, the quality of predictions not considering the court issue is the best one, with an AUC equal to 0.7020 and an accuracy equal to 0,663 (threshold equal to 0.5255).

Table 2 Best quality of predictions for models based on $diff$. For each variant, the best AUC measure, the corresponding threshold and the accuracy measure are reported, together with parameters' values used for calculation

periodicity	court issue	AUC	threshold	accuracy	regression to mean P%
historical	not considered	0.6925	0.5236	0.6626	90
historical	considered	0.6775	0.4788	0.6572	78
periodicity	court issue	AUC	threshold	accuracy	depth
dynamic	not considered	0.7020	0.5255	0.663	50
dynamicl	0.6944	0.5057	0.6586	27	

5.3 Using Four Factors

Table 3 reports some results: the model built on historical Four Factors without considering the court issue is the best one, with an AUC equal to 0.6655 and an accuracy equal to 0.6400 (threshold equal to 0.5334). Between dynamic features, the two models are equivalent in terms of quality of fit, slightly less than quality of historical model.

Table 3 Best quality of predictions for models based on Four Factors. For each variant, the best AUC, the corresponding threshold and the accuracy measure are reported, together with the parameter's value used for calculation

periodicity	court issue	AUC	threshold	accuracy	regression to mean P%
historical	not considered	0.6655	0.5334	0.6400	78
historical	considered	0.6527	0.4968	0.6347	74
court issue	AUC	threshold	accuracy	depth %	
dynamic	not considered	0.6495	0.4934	0.6371	42
dynamic	considered	0.6492	0.5091	0.6372	36

6 Conclusions

In this contribution we showed how appropriately defined statistics can profitably be used as single features in fitting models for outcome predictions on a basketball dataset including 16 NBA regular seasons from 2004-2005 to 2019-2020.
The models quality is better than quality of models fitted using Four Factors, a synthesis of *box-score* statistics.
The best prediction quality for a model considering the whole period has been produced using a single dynamic Elo feature (not considering the court issue), with an averaging depth equal to two (i.e. only Elo rating of prior two matches are considered in feature calculation). For this model, the AUC is equal to 0.7117 and the accuracy (using a threshold equal to 0.5049) is equal to 0.6736 (same AUC of the model built using historical Elo, but higher accuracy).
Results suggest that the court issue approach to features definition produces predictions comparable in the quality to models based on usual single feature, offering more interpretation details. Moreover, we verified how regression to mean can play a relevant role in prediction quality.
In general, quality of models built using $diff$ based features is close to quality of models built using Elo, and this is an expected result if we take into account how both these features express a direct measure of the strength of a team. Instead, the quality of models based on Four Factors is remarkably the lowest among the three approaches, suggesting how the approaches based on *box-score* statistics are close to their limit in outcome prediction quality.

Acknowledgements We would like to thank Prof. Marica Manisera for the help she gave us during this work.

References

1 Allaire, J., Chollet, F.: keras: R Interface to Keras (2021). URL https://CRAN.R-project.org/package=keras
2 Elo, A.E.: The Rating of Chess players, Past and Present. Ishi Press International (1978)
3 Galton, F.: Natural Inheritance. MacMillan (1889)
4 Kubatko, J., Oliver, D., Pelton, K., Rosenbaum, D.: A starting point for analysing basketball statistics. Journal of Quantitative Analysis in Sports **3**(3), 1–22 (2007)
5 Oliver, D.: Basketball on Paper: Rules and Tools for Performance Analysis. Potomac Books inc (2004)
6 Silver, N.: How We Calculate NBA Elo Ratings (2015). URL https://fivethirtyeight.com/features/how-we-calculate-nba-elo-ratings/