EstRate: A novel approach to predict performance of soccer players based on previous performance

Turash Mosharraf
Department of Computing Science
Simon Fraser University
Burnaby, BC, Canada
tmosharr@sfu.ca

Maria Babaeva
Department of Computing Science
Simon Fraser University
Burnaby, BC, Canada
mbabaeva@sfu.ca

Mohammad Muhaimin
Department of Computing Science
Simon Fraser University
Burnaby, BC, Canada
mmuhaimi@sfu.ca

Volodymyr Kozyr
Department of Computing Science
Simon Fraser University
Burnaby, BC, Canada
vkozyr@sfu.ca

Abstract

The task of selecting suitable players has been a major problem in premier leagues since last decades. Researchers have proposed several techniques to predict future ratings from a set of known features. However, in a realistic scenario, many features are unknown and missing. In this project, we try to evaluate future rating of players in English Premier League (EPL) from a blend of both known and unknown features. In our first step, we use different machine learning techniques to estimate the values of unknown features. In our second step, we use the combination of known and estimated features to predict the overall ratings. We train our model using English Premier League data of 2016 and 2017. After that, we test the accuracy of our model by data of 2018. The experiment shows that our model outperforms the existing baseline approach in certain situations.

1 Introduction

With the growing popularity of soccer in modern era, the clubs are facing a new problem of selecting suitable player in transfer window. Unfortunately, the performance of a player depends on many physical, psychological and conditional features. Researchers have proposed several machine learning algorithms to learn the ratings based on the features. However, the managers may not be able to use these methods during a transfer window because all of the features required by the models are not available at the start of transfer window. For example, traditional approaches rely on shot accuracy of the current season, However, shot accuracy can only be correctly calculated at the end of the season.

In this project, we developed a model to address this particular issue of predicting overall performance from a blend of known and unknown features. Our model works in two steps. In first step, we identify the unknown features (we refer to them as key features from throughout the rest of this paper) and estimate their values from previous years performance. In the second step, we plug in the key features along with other known features (fixed features like height, weight, etc) to estimate the overall rating. Because of the initial error in calculating key features, our model are sometimes outperformed by the existing approaches. However, it can be easily applied to real life scenarios.

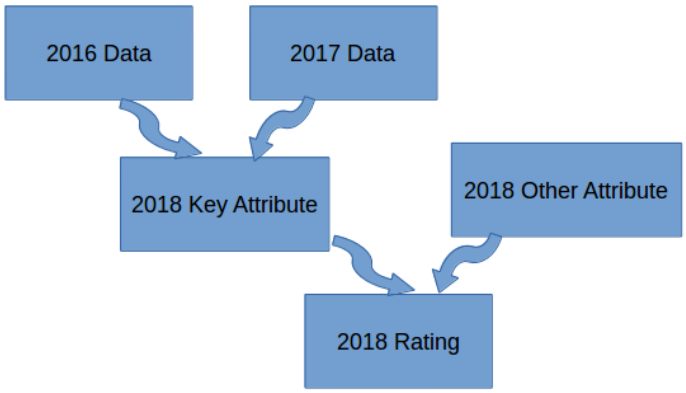

Figure 1: Graphical structure of our model

1.1 Related Works

Yuan He [1] developed a model to predict the market value of a player based on different predictors. The predictors are classified into two categories: personal information and performance data. Personal informations of a player does not vary (except age) over seasons. However, performance data shows significant variations. He used 8 personal and 5 performance predictors as features in a dataset consisting of records of 5 years and demonstrated the effectiveness of machine learning techniques on predicting the price of players. However, whereas the price of a player depends on last years performance, the rating depends totally on current years performance. Therefore, He's method are not effective for predicting future ratings.

Brooks [2] et el. implemented a model to measure the defensive, midfield and attacking score of a player solely bessed on pass completion rate. This value is derived based on the relationship of pass locations in a possession and shot opportunities generated. This relationship is learned by ap- plying a supervised machine learning model to pass locations in event data from the 2012-2013 La Liga season. However, In reality, pass completion rate of current year is not known at transfer time and last year rate may not be a good indicator for current year. The Brooks model address the second problem of our model.

Vroonen [3] et el. developed APROPOS projection system. Given a players ratings of a particular skill and corresponding ages, APROPOS can effectively predict his future ratings for that skill at any other ages using k-nearest neighbors approach. APROPOS depends on the ratings of similar players to predict how a players skill can evolve in future. The first problem of our model is a specific case of the problem addressed by APROPOS in which:

$$age_{target} = max_{a \in age_{training}}(a) + 1. \tag{1}$$

APROPOS uses kNN as the prediction algorithm. Therefore it may suffer from the curse of dimensionality for large feature space. We try to solve this issue by using neural networks and ensemble learning based methods.

1.2 Style

Papers to be submitted to NIPS 2015 must be prepared according to the instructions presented here. Papers may be only up to eight pages long, including figures. Since 2009 an additional ninth page *containing only cited references* is allowed. Papers that exceed nine pages will not be reviewed, or in any other way considered for presentation at the conference.

Please note that this year we have introduced automatic line number generation into the style file (for LaTeX 2_ε and Word versions). This is to help reviewers refer to specific lines of the paper when they make their comments. Please do NOT refer to these line numbers in your paper as they will be removed from the style file for the final version of accepted papers.

The margins in 2015 are the same as since 2007, which allow for $\approx 15\%$ more words in the paper compared to earlier years. We are also again using double-blind reviewing. Both of these require the use of new style files.

Authors are required to use the NIPS LaTeX style files obtainable at the NIPS website as indicated below. Please make sure you use the current files and not previous versions. Tweaking the style files may be grounds for rejection.

1.3 Retrieval of style files

The style files for NIPS and other conference information are available on the World Wide Web at

`http://www.nips.cc/`

The file `nips2015.pdf` contains these instructions and illustrates the various formatting requirements your NIPS paper must satisfy. LaTeX users can choose between two style files: `nips15submit_09.sty` (to be used with LaTeX version 2.09) and `nips15submit_e.sty` (to be used with LaTeX2e). The file `nips2015.tex` may be used as a "shell" for writing your paper. All you have to do is replace the author, title, abstract, and text of the paper with your own. The file `nips2015.rtf` is provided as a shell for MS Word users.

The formatting instructions contained in these style files are summarized in sections 2, 3, and 4 below.

2 General formatting instructions

The text must be confined within a rectangle 5.5 inches (33 picas) wide and 9 inches (54 picas) long. The left margin is 1.5 inch (9 picas). Use 10 point type with a vertical spacing of 11 points. Times New Roman is the preferred typeface throughout. Paragraphs are separated by 1/2 line space, with no indentation.

Paper title is 17 point, initial caps/lower case, bold, centered between 2 horizontal rules. Top rule is 4 points thick and bottom rule is 1 point thick. Allow 1/4 inch space above and below title to rules. All pages should start at 1 inch (6 picas) from the top of the page.

For the final version, authors' names are set in boldface, and each name is centered above the corresponding address. The lead author's name is to be listed first (left-most), and the co-authors' names (if different address) are set to follow. If there is only one co-author, list both author and co-author side by side.

Please pay special attention to the instructions in section 4 regarding figures, tables, acknowledgments, and references.

3 Headings: first level

First level headings are lower case (except for first word and proper nouns), flush left, bold and in point size 12. One line space before the first level heading and 1/2 line space after the first level heading.

3.1 Headings: second level

Second level headings are lower case (except for first word and proper nouns), flush left, bold and in point size 10. One line space before the second level heading and 1/2 line space after the second level heading.

3.1.1 Headings: third level

Third level headings are lower case (except for first word and proper nouns), flush left, bold and in point size 10. One line space before the third level heading and 1/2 line space after the third level heading.

4 Citations, figures, tables, references

These instructions apply to everyone, regardless of the formatter being used.

4.1 Citations within the text

Citations within the text should be numbered consecutively. The corresponding number is to appear enclosed in square brackets, such as [1] or [2]-[5]. The corresponding references are to be listed in the same order at the end of the paper, in the **References** section. (Note: the standard BIBTEX style `unsrt` produces this.) As to the format of the references themselves, any style is acceptable as long as it is used consistently.

As submission is double blind, refer to your own published work in the third person. That is, use "In the previous work of Jones et al. [4]", not "In our previous work [4]". If you cite your other papers that are not widely available (e.g. a journal paper under review), use anonymous author names in the citation, e.g. an author of the form "A. Anonymous".

4.2 Footnotes

Indicate footnotes with a number[1] in the text. Place the footnotes at the bottom of the page on which they appear. Precede the footnote with a horizontal rule of 2 inches (12 picas).[2]

4.3 Figures

All artwork must be neat, clean, and legible. Lines should be dark enough for purposes of reproduction; art work should not be hand-drawn. The figure number and caption always appear after the figure. Place one line space before the figure caption, and one line space after the figure. The figure caption is lower case (except for first word and proper nouns); figures are numbered consecutively.

Make sure the figure caption does not get separated from the figure. Leave sufficient space to avoid splitting the figure and figure caption.

You may use color figures. However, it is best for the figure captions and the paper body to make sense if the paper is printed either in black/white or in color.

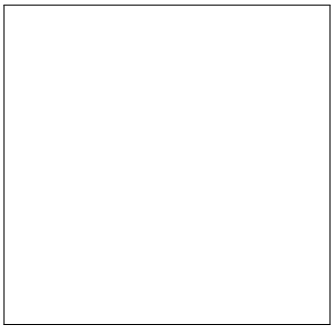

Figure 2: Sample figure caption.

4.4 Tables

All tables must be centered, neat, clean and legible. Do not use hand-drawn tables. The table number and title always appear before the table. See Table 1.

[1] Sample of the first footnote

[2] Sample of the second footnote

Table 1: Sample table title

PART	DESCRIPTION
Dendrite	Input terminal
Axon	Output terminal
Soma	Cell body (contains cell nucleus)

Place one line space before the table title, one line space after the table title, and one line space after the table. The table title must be lower case (except for first word and proper nouns); tables are numbered consecutively.

5 Final instructions

Do not change any aspects of the formatting parameters in the style files. In particular, do not modify the width or length of the rectangle the text should fit into, and do not change font sizes (except perhaps in the **References** section; see below). Please note that pages should be numbered.

6 Preparing PostScript or PDF files

Please prepare PostScript or PDF files with paper size "US Letter", and not, for example, "A4". The -t letter option on dvips will produce US Letter files.

Fonts were the main cause of problems in the past years. Your PDF file must only contain Type 1 or Embedded TrueType fonts. Here are a few instructions to achieve this.

- You can check which fonts a PDF files uses. In Acrobat Reader, select the menu Files>Document Properties>Fonts and select Show All Fonts. You can also use the program `pdffonts` which comes with `xpdf` and is available out-of-the-box on most Linux machines.
- The IEEE has recommendations for generating PDF files whose fonts are also acceptable for NIPS. Please see `http://www.emfield.org/icuwb2010/downloads/IEEE-PDF-SpecV32.pdf`
- LaTeX users:
  - Consider directly generating PDF files using `pdflatex` (especially if you are a MiKTeX user). PDF figures must be substituted for EPS figures, however.
  - Otherwise, please generate your PostScript and PDF files with the following commands:

    ```
    dvips mypaper.dvi -t letter -Ppdf -G0 -o mypaper.ps
    ps2pdf mypaper.ps mypaper.pdf
    ```

    Check that the PDF files only contains Type 1 fonts.
  - xfig "patterned" shapes are implemented with bitmap fonts. Use "solid" shapes instead.
  - The `\bbold` package almost always uses bitmap fonts. You can try the equivalent AMS Fonts with command

    ```
    \usepackage[psamsfonts]{amssymb}
    ```

    or use the following workaround for reals, natural and complex:

    ```
    \newcommand{\RR}{I\!\!R} %real numbers
    \newcommand{\Nat}{I\!\!N} %natural numbers
    \newcommand{\CC}{I\!\!\!\!C} %complex numbers
    ```
  - Sometimes the problematic fonts are used in figures included in LaTeX files. The ghostscript program `eps2eps` is the simplest way to clean such figures. For black and white figures, slightly better results can be achieved with program `potrace`.

- MSWord and Windows users (via PDF file):
  - Install the Microsoft Save as PDF Office 2007 Add-in from `http://www.microsoft.com/downloads/details.aspx?displaylang=en&familyid=4d951911-3e7e-4ae6-b059-a2e79ed87041`
  - Select "Save or Publish to PDF" from the Office or File menu
- MSWord and Mac OS X users (via PDF file):
  - From the print menu, click the PDF drop-down box, and select "Save as PDF..."
- MSWord and Windows users (via PS file):
  - To create a new printer on your computer, install the AdobePS printer driver and the Adobe Distiller PPD file from `http://www.adobe.com/support/downloads/detail.jsp?ftpID=204` *Note:* You must reboot your PC after installing the AdobePS driver for it to take effect.
  - To produce the ps file, select "Print" from the MS app, choose the installed AdobePS printer, click on "Properties", click on "Advanced."
  - Set "TrueType Font" to be "Download as Softfont"
  - Open the "PostScript Options" folder
  - Select "PostScript Output Option" to be "Optimize for Portability"
  - Select "TrueType Font Download Option" to be "Outline"
  - Select "Send PostScript Error Handler" to be "No"
  - Click "OK" three times, print your file.
  - Now, use Adobe Acrobat Distiller or ps2pdf to create a PDF file from the PS file. In Acrobat, check the option "Embed all fonts" if applicable.

If your file contains Type 3 fonts or non embedded TrueType fonts, we will ask you to fix it.

6.1 Margins in LaTeX

Most of the margin problems come from figures positioned by hand using `\special` or other commands. We suggest using the command `\includegraphics` from the graphicx package. Always specify the figure width as a multiple of the line width as in the example below using .eps graphics

```
\usepackage[dvips]{graphicx} ...
\includegraphics[width=0.8\linewidth]{myfile.eps}
```

or

```
\usepackage[pdftex]{graphicx} ...
\includegraphics[width=0.8\linewidth]{myfile.pdf}
```

for .pdf graphics. See section 4.4 in the graphics bundle documentation (`http://www.ctan.org/tex-archive/macros/latex/required/graphics/grfguide.ps`)

A number of width problems arise when LaTeX cannot properly hyphenate a line. Please give LaTeX hyphenation hints using the `\-` command.

Acknowledgments

Use unnumbered third level headings for the acknowledgments. All acknowledgments go at the end of the paper. Do not include acknowledgments in the anonymized submission, only in the final paper.

References

References follow the acknowledgments. Use unnumbered third level heading for the references. Any choice of citation style is acceptable as long as you are consistent. It is permissible to reduce the font size to 'small' (9-point) when listing the references. **Remember that this year you can use a ninth page as long as it contains *only* cited references.**

References

[1] Y. He, "Predicting market value of soccer players using linear modeling techniques," *Unpublished manuscript*, 2013.

[2] J. Brooks, M. Kerr, and J. Guttag, "Developing a data-driven player ranking in soccer using predictive model weights," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2016, pp. 49–55.

[3] R. Vroonen, T. Decroos, J. Van Haaren, and J. Davis, "Predicting the potential of professional soccer players," in *Machine Learning and Data Mining for Sports Analytics ECML/PKDD 2017 workshop*, 2017.