Investigation of Overall Rating Fairness in FIFA Games via Machine Learning Methods

Hanze Yao
Department of Civil and Environmental Engineering
University of California Berkeley
Berkeley, United States
yaohanze2016@berkeley.edu

Chao Wang
Department of Civil and Environmental Engineering
University of California Berkeley
Berkeley, United States
chao_wang@berkeley.edu

***Abstract*—FIFA is a football simulation game published by Electronic Arts. There is a long-existing claim in the FIFA community that English Premier League (EPL) and England players have an advantage on overall rating over other players with similar in-game stats. Because England has one of the largest FIFA gaming communities, and EPL has the highest market value among all professional football leagues globally. In this report, classification and neural network method are used to investigate the overall rating fairness from FIFA 15 to FIFA 20.**

Keywords—classification, neural network, fairness

I. Introduction

Machine Learning (ML) has become an emerging technology in many engineering and social science fields due to its advantages in fast decision making with high accuracy [1]. Despite these promising features, one of the main challenges for applying machine learning in social science is fairness [2-4]. Over the last several years, many definitions of fairness have been proposed [5-8].

FIFA is an industry-leading football simulation game series developed by EA Vancouver and EA Romania. Each in-game player has six categories of stats: pace, dribbling, shooting, defending, passing, and physical. Goalkeepers have a different rating system with six stats categories: diving, handling, kicking, reflexes, speed, and positioning. Each stats category contains several sub stats. For example, pace includes acceleration and sprint speed. The overall rating model based on in-game stats is never open to the public. As a result, FIFA has been criticized for favoring English Premier League and England players on the overall rating because of the huge market value of England FIFA community. An example of this potential overall rating unfairness is as follows. In the latest FIFA game, Ferland Mendy, a French left-back playing for Real Madrid in La Liga (top football league in Spain, one of the five major European football leagues), only has 83 overall ratings with 2168 total in-game stats. While Jamie Vardy, an English striker playing for Leicester City in English Premier League, has 86 overall rating with only 2143 total in-game stats. Three overall ratings is a big difference for top tier football players in FIFA. In this project, classification is used to investigate the overall rating fairness for EPL players. A neural network is used to investigate the overall rating fairness for English players. The methodology and results are discussed in detail in the following sections.

II. Methodology

A. Fairness between players from EPL and other leagues

Unlike the debate over COMPAS on the fairness between black and white defendants, there is not an attribute that can measure the overall rating like the COMPAS score to the risk of recidivism. Therefore, it is not possible to use the same method as setting a threshold for COMPAS score and investigating the false positive rate for different races. So, a new method is used in this project. In order to formulate this as a classification problem, a threshold is required for the overall rating. To select a reasonable threshold, only players from five major European Leagues (English Premier Leagues, Spanish La Liga, Germany Bundesliga, Italian Seria A, and French Ligue 1) are considered. Because choosing a single overall rating threshold can guarantee a similar percentage of players with overall above and below the threshold between different leagues. The mean and median overall for the five leagues are shown in Table 1. A threshold of 73 is selected for classification.

TABLE I. Mean and Median Overall for Different Leagues

Game	English Premier League	Other Four European Leagues
FIFA 15	70.5 / 72.0	69.9 / 71.0
FIFA 16	72.9 / 75.0	71.9 / 73.0
FIFA 17	72.2 / 75.0	72.6 / 74.0
FIFA 18	72.5 / 75.0	72.5 / 73.0
FIFA 19	72.6 / 75.0	72.8 / 74.0
FIFA 20	73.4 / 75.0	72.4 / 73.0

Now, this is a binary classification problem for players with 73 overall or above and players with below 73 overall based on eight features: pace, dribbling, shooting, defending, passing, physical, skill move, and weak foot. Goalkeepers are not considered because of the small data size. SGDClassifier from sklearn.linear_model [9] with l2 regulation and logistic loss is used. After training, calculate the False Negative Rate (FNR) and False Positive Rate (FPR) of the test set. Now apply the EPL classification model to the test set of four other major European Leagues. If there is an overall rating boost for EPL players, the FNR should decrease, and the FPR should increase compared to the original FNR and FPR of the four other major European Leagues' test set.

B. *Fairness between players from England and other countries*

In this subsection, the fairness between English football players and football players from other countries is studied. A three-layer fully connected neural network [10] is trained to predict the overall rating of a player based on his stats attributes. The number of nodes for the two hidden layers of this neural network are 25 and 36, respectively. The activation functions are chosen as the rectified linear unit (ReLU).

From FIFA 15 to FIFA 20, in each year, two different models are then built with the same architecture given above. The difference is the training data sets. One model is trained with English players. The other model is trained with all non-English players. Finally, the model trained with English players is used to predict the overall rating of the non-English players. If for non-English players, the overall rating predicted by the model trained with English players is higher than the value predicted by the model trained with non-English players, we then can conclude the bias in the FIFA rating model.

III. RESULTS

In this section, Results from both parts of this project are shown.

A. *Results on league fairness*

The accuracy, false positive rate and false negative rate on the training set and test set of EPL players and players from four other major European Leagues are shown in Table II.

TABLE II. ACCURACY FPR AND FNR FOR DIFFERENT LEAGUES

Game			**English Premier League**	**Other four European Leagues**
FIFA 15	Training set	Accuracy	0.783	0.77
		FPR	0.267	0.279
		FNR	0.164	0.158
	Test set	Accuracy	0.782	0.753
		FPR	0.286	0.267
		FNR	0.143	0.219
FIFA 16	Training set	Accuracy	0.855	0.768
		FPR	0.37	0.244
		FNR	0.027	0.221
	Test set	Accuracy	0.795	0.77
		FPR	0.433	0.22
		FNR	0.014	0.238
FIFA 17	Training set	Accuracy	0.807	0.791
		FPR	0.222	0.234
		FNR	0.173	0.19
	Test set	Accuracy	0.822	0.818
		FPR	0.197	0.208
		FNR	0.165	0.163
FIFA 18	Training set	Accuracy	0.826	0.8
		FPR	0.306	0.258
		FNR	0.09	0.152
	Test set	Accuracy	0.849	0.776
		FPR	0.232	0.238
		FNR	0.1	0.215
FIFA 19	Training set	Accuracy	0.856	0.803
		FPR	0.242	0.323
		FNR	0.086	0.103
	Test set	Accuracy	0.926	0.823
		FPR	0.323	0.28
		FNR	0.103	0.105
FIFA 20	Training set	Accuracy	0.857	0.826
		FPR	0.265	0.197
		FNR	0.082	0.154
	Test set	Accuracy	0.838	0.806
		FPR	0.255	0.214
		FNR	0.119	0.179

After applying the classification model of EPL players to four other major European Leagues players (test set only this time), the Accuracy, FPR and FNR comparison are shown in Table III.

TABLE III. FPR AND FNR COMPARION

Game		**With other leagues classifier**	**With EPL classifier**
FIFA 15	Accuracy	0.782	0.748
	FPR	0.267	0.307
	FNR	0.219	0.178
FIFA 16	Accuracy	0.795	0.756
	FPR	0.22	0.498
	FNR	0.238	0.027
FIFA 17	Accuracy	0.822	0.701
	FPR	0.208	0.361
	FNR	0.163	0.253
FIFA 18	Accuracy	0.849	0.789
	FPR	0.238	0.333
	FNR	0.215	0.124
FIFA 19	Accuracy	0.926	0.811
	FPR	0.28	0.285
	FNR	0.105	0.122
FIFA 20	Accuracy	0.838	0.785
	FPR	0.214	0.41
	FNR	0.179	0.066

The FPR and FNR change by applying different classifiers to four other European leagues test set over the years are shown in Fig. 1.

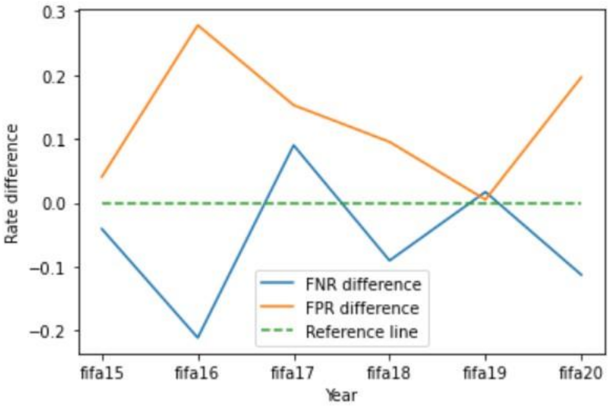

Fig. 1. FNR and FPR difference after applying EPL classifier to four other European League players

B. Results on nation fairness

After training the neural net for goalkeeper and non-goalkeeper from England and other countries, the Mean Squared Error (MSE) on the test set is shown in Table IV.

TABLE IV. MSE ON TEST SET FOR TRAINED NEURAL NET

Game	Position	English players	non-English players
FIFA 15	Goalkeeper	0.353	0.307
	non-Goalkeeper	3.298	3.236
FIFA 16	Goalkeeper	0.82	0.542
	non-Goalkeeper	4.085	3.73
FIFA 17	Goalkeeper	0.851	1.153
	non-Goalkeeper	2.567	3.206
FIFA 18	Goalkeeper	0.689	0.68
	non-Goalkeeper	3.531	3.661
FIFA 19	Goalkeeper	0.574	0.855
	non-Goalkeeper	4.298	3.97
FIFA 20	Goalkeeper	0.758	0.836
	non-Goalkeeper	6.177	5.23

For goalkeepers, the overall rating prediction difference between using neural net trained with English goalkeepers and neural net trained with non-English goalkeepers on non-English goalkeeper test set has nearly the same trend as FIFA 15 except for FIFA 19 as shown in Fig. 2 and Fig. 3.

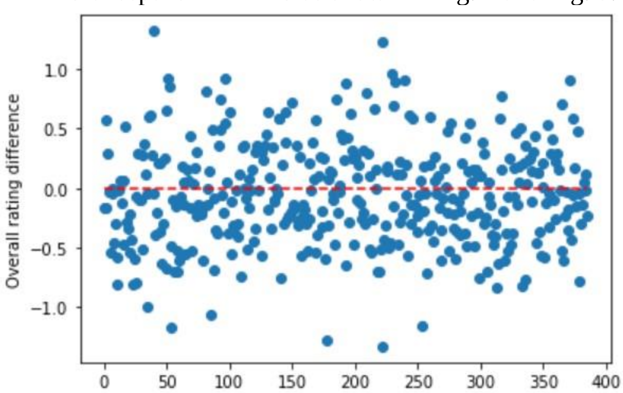

Fig. 2. FIFA 15 Overall rating prediction difference between two neural nets for goalkeeper

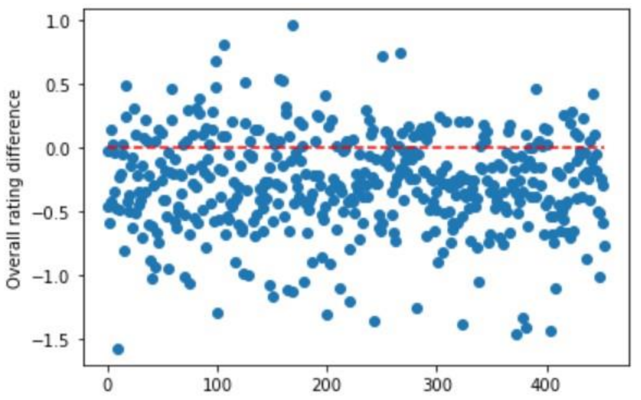

Fig. 3. FIFA 17 Overall rating prediction difference between two neural nets for goalkeeper

As for non-goalkeepers, the overall rating prediction difference on the non-English non-goalkeeper test set between the two neural nets generally has the same trend as FIFA 15, as shown in Fig. 4.

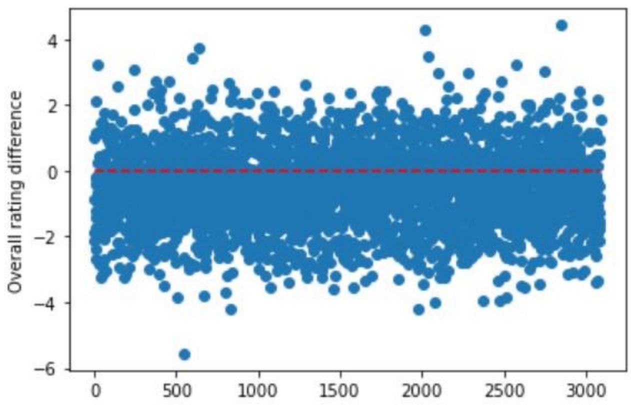

Fig. 4. FIFA 15 Overall rating prediction difference between two neural nets for non-goalkeeper

IV. DISCUSSION

What those results really show is discussed in this section.

A. Results on league fairness

According to Table II, the accuracy of all classifiers on the test set is around 0.8, which is similar to the accuracy on the training set. Therefore, there is no overfitting problem, and the performance of the classifier is good. As shown in Fig. 1, for FIFA 15, FIFA 16, FIFA 18, and FIFA 20, after applying the classification model for EPL players to four other European League players test set, the False Positive Rate (FPR) increases, but the False Negative Rate (FNR) decreases. This means a higher percentage of four other European League players with an overall rating below 73 are predicted to be 73 overall or higher, and a lower percentage of four other European League players with 73 overall or above are predicted to be below 73 overall with EPL classifier. Therefore, for these four games, the overall rating advantage for EPL players over the other four major European Leagues players with similar in-game stats does exist. However, for FIFA 17, the result can only show that the classification model for EPL and the other four leagues are noticeably different. And the result from FIFA 19 shows that it did a great job at maintaining the overall rating fairness between different leagues.

B. Results on nation fairness

For goalkeeper, according to Fig. 2, the overall rating prediction difference between neural nets trained on English and non-English players spreads almost evenly around 0 with rather a small margin. Therefore, the overall rating models for English and non-English goalkeeper are almost the same. This is true except for FIFA 17. According to Fig. 3, FIFA 17 has a slightly stricter overall rating model for English goalkeeper.

For non-goalkeeper, the overall rating prediction difference between neural nets trained on English and non-English players shifts slightly to the negative side. This means all six investigated FIFA games tend to give slightly lower rating to English players than non-English players with similar in-game stats.

V. CONCLUSION

The following conclusions can be drawn from the previous discussion:

- The overall rating advantage for English Premier League (EPL) over other leagues does exist in four of the six FIFA games investigated in this project (except for FIFA 17 and FIFA 19).
- The overall rating model for goalkeeper does not have a noticeable difference between English players and non-English players. Except for FIFA 17, it has a slightly stricter model for English goalkeepers.
- For non-goalkeeper, FIFA game has a slightly stricter model for English players than players from other countries generally.

REFERENCES

[1] I Goodfellow, Y Bengio, A Courville, "Machine learning basics" - Deep learning, 2016 - deeplearningbook.org

[2] A Chouldechova, A Roth, "The frontiers of fairness in machine learning" - arXiv preprint arXiv:1810.08810, 2018 - arxiv.org

[3] S Corbett-Davies, S Goel, "The measure and mismeasure of fairness: A critical review of fair machine learning" - arXiv preprint arXiv:1808.00023, 2018 - arxiv.org

[4] N Mehrabi, F Morstatter, N Saxena, K Lerman, "A survey on bias and fairness in machine learning" - arXiv preprint arXiv, 2019 - arxiv.org

[5] Toon Calders and Sicco Verwer, "Three naive bayes approaches for discrimination-free classification." Data Mining and Knowledge Discovery, 21(2):277–292, 2010.

[6] Toshihiro Kamishima, Shotaro Akaho, and Jun Sakuma, "Fairness-aware learning through regularization approach." In Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on, pages 643–650. IEEE, 2011.

[7] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel, "Fairness through awareness." In Proceedings of the 3rd innovations in theoretical computer science conference, pages 214–226. ACM, 2012.

[8] Michael Feldman, Sorelle A Friedler, John Moeller, Carlos Scheidegger, and Suresh Venkatasubramanian. "Certifying and removing disparate impact." In KDD, 2015.

[9] Kingma, J Ba, "Adam: A method for stochastic optimizationDP" - arXiv preprint arXiv:1412.6980, 2014 - arxiv.org

[10] B Müller, J Reinhardt, MT Strickland, "Neural networks: an introduction" - 2012 - books.google.com