Soccer Talk: Exploring Differences in r/soccer Threads about Men’s and Women’s World Cup Match Threads

Kaitlin Swinnerton
August 3, 2019

Abstract

This paper examines whether soccer fans use different language patterns in online discussions of games played by male versus female players at their respective FIFA World Cup tournaments. I used text mining techniques to scrape comments from match threads on reddit.com/r/soccer, an online forum where soccer fans engage in casual discussion and analysis of games. I first trained a Multinomial Naive Bayes model to predict whether a given comment is from a thread about a men’s World Cup game or a women’s World Cup game. A baseline model achieved an accuracy of 67%. I then examined the words that are most predictive of gender. Next, to try to achieve improved performance, I trained recurrent neural networks to classify comments. Those models failed to improve accuracy, achieving an accuracy of about 65%. While all of these models perform greater than chance, suggesting that there may be differences in language used discussing men’s and women’s games, further investigation is needed to draw strong conclusions about those differences.

Introduction

Despite increased growth and participation in women’s sports in recent years, there are still significant differences in respect, attention, and earning potential for female athletes. In *Forbe’s* most recent list of the 100 highest paid athletes, Serena Williams is the sole woman. The back to back World Cup winning US Women’s National Soccer team is currently suing their federation, U.S. Soccer, for gender discrimination in the work place, arguing that they have been paid less and have experienced inferior working conditions compared to their male counterparts on the Men's national team. In the United States, women athletes are assumed to be inferior athletes and experience bias and migro-aggressions towards them (Kaskan et al, 2016). Media coverage of female athletes overwhelmingly focuses on their appearance. The combination of these experiences can cause negative biological, cognitive, and behavior effects on women experiencing them (Kaskan et al, 2016).

Online forums are another place where sexism can persist. A 2016 analysis by Alice Wu of comments on the Economics Job Market Rumors Forum found evidence of sexism directed to female economists discussed on the site. Among academics studying economics, women are underrepresented and face difficulty rising up the ranks in their careers. Since female athletes share similar challenges, I decided to look for evidence of sexism in online forums discussing sports. With the Women's World Cup ending recently, I decided to focus my analysis on soccer, comparing comments made on match threads posted on reddit.com/r/soccer.

Models

Data scraping, cleaning, and preparation

I used the reddit API to scrape comments from 20 match discussion threads on reddit.com/r/soccer from each of the 2019 FIFA Women's World Cup, and the 2018 FIFA World Cup, resulting in 7,613 comments about the men's World Cup and 7,717 comments about the women's World Cup. I also scraped metadata about each comment, including the number of upvotes it received, the number of replies, whether it was gilded, and whether it was marked as controversial. I used TextBlob sentiment analysis to create polarity and subjectivity scores for each comment. I also used the VADER model, designed to rate sentiment for social media posts, to create negative, neutral, positive, and compound scores for each post (Hutto et al, 2014). I cleaned the text data by character accents (which are common on the names of players participating in the World Cup), and removing punctuation, consistent with the methods used by Kim et al. (2014). I next took steps to reduce the impact of gender pronouns and tournament specific names on model the model. I replaced all gender pronouns with gender neutral pronouns. Next, I scraped Wikipedia to obtain lists of the names of all of the players and countries who participated in both 2018 and 2019 World Cups. Exploratory analysis did not show evidence of differences in sentiment between comments about the men's and women's tournaments, so I did not pursue further investigation into that topic.

Baseline Model

I first built a Multinomial Naive Bayes model to classify comments as either belonging to discussions of the men's or women's world cups. For these analyses, I used an 80%, 10%, 10% train, development, test set split. I began by using the CountVectorizer function from sklearn to transform the comments into a matrix of token counts.The vocabulary size of the training data was 10,818. I first trained a Multinomial Naive Bayes model with default parameters and achieved a development classification accuracy of 66.86%. A grid search to identify the optimal alpha parameter, identified the default value of 1.0 as optimal, so results were could not be improved by tuning hyper parameters. Next, I used the CountVectorizer function to create bigram and trigram vectors. MNB models trained on the bigram and trigram vectors did not yield improved accuracy. Next, I identified which words had the strongest predictive power for gender by calculating the difference of feature log probabilities predicted for each class for each word. Many of the words that were predictive of gender=male were words, such as 'Russian,', 'Iranians,' and "Mexicans,' that described people from countries that participated in the men's World Cup, but not the women's World Cup. This shows that attempts to remove country identifiers were insufficient, and need to be more thorough in order to reveal language differences. While there were still a few instances of country specific words among the words most predictive of gender=female, the top words list also included words such as 'illegal' and 'racism,' as well as both 'onside' and 'offside.' When expanding this analysis to the top 50 words, 'racist,' 'sloppy,' and 'attendance,' joined the women's list, and 'doping,' joined the men's list. While these analyses do not show evidence of sexism, they do reveal some points of interest, and more thorough analyses might yield increasingly interesting results.

Feature	Men's World Cup	Women's World Cup	coef_diff
thai	-11.429555	-8.276627	3.152928
hd	-11.429555	-8.367599	3.061956
norwegian	-11.429555	-8.367599	3.061956
band	-11.429555	-8.416389	3.013166
bein	-11.429555	-8.467682	2.961872
graham	-11.429555	-8.521749	2.907805
hara	-11.429555	-8.578908	2.850647
vd	-11.429555	-8.578908	2.850647
illegal	-11.429555	-8.578908	2.850647
onside	-11.429555	-8.639532	2.790022
supersport	-10.330942	-7.583480	2.747463
olympics	-11.429555	-8.704071	2.725484
x200b	-11.429555	-8.704071	2.725484
themself	-11.429555	-8.704071	2.725484
racism	-11.429555	-8.773064	2.656491
wwc	-11.429555	-8.773064	2.656491
chilean	-11.429555	-8.847172	2.582383
female	-10.736408	-8.193245	2.543162
offside	-9.126970	-6.616331	2.510639
orange	-11.429555	-8.927215	2.502340

Feature	Men's World Cup	Women's World Cup	coef_diff
russian	-7.270672	-10.718974	-3.448302
vuvuzelas	-8.028357	-11.412121	-3.383764
iranian	-8.133718	-11.412121	-3.278403
mexicans	-8.171458	-11.412121	-3.240663
peruvian	-8.251501	-11.412121	-3.160620
kdb	-8.294061	-11.412121	-3.118061
croatian	-8.338512	-11.412121	-3.073609
itv	-8.385032	-11.412121	-3.027089
2010	-8.385032	-11.412121	-3.027089
colombian	-8.433822	-11.412121	-2.978299
doot	-8.539183	-11.412121	-2.872938
belgian	-8.539183	-11.412121	-2.872938
russians	-8.596341	-11.412121	-2.815780
putin	-8.656966	-11.412121	-2.755155
subasic	-8.721505	-11.412121	-2.690617
honk	-8.721505	-11.412121	-2.690617
ba	-8.790497	-11.412121	-2.621624
bus	-8.133718	-10.718974	-2.585256
2014	-8.171458	-10.718974	-2.547516
columbia	-8.864605	-11.412121	-2.547516

Figure 1. Words with the strongest predictive power for gender - left table is words with the strongest predictive power for women, right is for men

Next, I investigated what types of mistakes were made by the classifier by printing out the comments that had the biggest R ratio, which represents divergence between predicted class probability and actual class. Notably, one of the most poorly classified comments was a discussion of a team's uniform. The model predicted that it was about the women's tournament, while it was actually about the men's tournament. Here is an instance where the model may be reflecting existing tendencies to comment on the appearance of female athletes. The model predicted that the comment discussing the appearance of players was about the women's tournament.

```
Top five most poorly classified messages

Message 1:
Actual Label: 0,  Predicted Label: [1]
R ratio: 4081.78

this is either name brilliant tackle or clear penalty depending on the poster , which is why soccer name var in socce
r is always going to be controversial name many calls are name to the refs name made in the context name what has hap
pened in the game already


Message 2:
Actual Label: 1,  Predicted Label: [0]
R ratio: 1716.19

we keep avoiding the germans , 2010 world cup 2014 world cup now 2019


Message 3:
Actual Label: 0,  Predicted Label: [1]
R ratio: 801.15

country name country should play 3 on 3 ot but first give them sticks to hit the ball , then put them on ice with ska
tes , then change ball to name puck


Message 4:
Actual Label: 0,  Predicted Label: [1]
R ratio: 409.72

motion to make yellow country 's primary kit colour it looks name much nicer yes , i know they 're called the red dev
ils but the yellow with the black shorts look like it would make name better first choice kit


Message 5:
Actual Label: 0,  Predicted Label: [1]
R ratio: 344.95

surely on the line doesn t mean outside the box for keepers \? i don t know how the referee could be sure about that
one
```

Figure 2: The five most poorly classified comments

Neural network models

Next, I built two different neural network models to see if they would perform better on classification of comments. I built a recurrent neural network with an embedding layer, recurrent layer, a fully connected layer, a dropout layer, and an output layer. Experiments were run to test the ideal number of neurons and batch size. The optimal model yielded a test accuracy of 65%, failing to outperform the Naive Bayes model.

Discussion

These experiments show some evidence that there are differences in how /r/soccer commentators discuss games from the 2019 FIFA Women's World Cup compared to games from the 2018 FIFA World Cup. Models were able to perform better than chance at classifying a comments. However, much more analysis is needed to draw conclusions about how the language differs and whether or not sexism against female athletes is exhibited in these threads. The scope of analysis could be increased immensely by examining comments about games outside of these two world cups. It is conceivable that comments from World Cup match threads may exhibit fewer tendencies towards sexism than other discussions. The World Cup is the pinnacle of the sport, and the match threads on r/soccer are for commenting on the specifics of each game. More informal comment threads or threads about games with less prestige may be more susceptible to demonstrating sexism and microaggressions. Additionally, in the US, the US Women's National Soccer team is one of the most popular and most well respected women's sports teams. Investigating language used to talk about female athletes in less respected sports or on less respected teams may yield different results.

Additionally, further analyses could be done to dive deeper into this data. Investigating the content of the comments more thoroughly, for example, looking to see if comments about women's games are more likely to involve discussions of the player's physical appearance, could yield interesting results.

References:

Hutto, C. J., & Gilbert, E. (2014, May). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In Eighth international AAAI conference on weblogs and social media.

Kaskan, E. R., & Ho, I. K. (2016). Microaggressions and female athletes. Sex Roles, 74(7-8), 275-287.

Kim, Y. (2014). Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882.

Wu, A. H. (2017). Gender stereotyping in academia: Evidence from economics job market rumors forum. Unpublished manuscript.