Live It Up: Analyzing Emotions and Language Use in Tweets during the Soccer World Cup Finals

Marisa Vasconcelos
IBM Research
marisaav@br.ibm.com

Jussara Almeida
UFMG
jussara@dcc.ufmg.br

Paulo Cavalin
IBM Research
pcavalin@br.ibm.com

Claudio Pinhanez
IBM Research
csantosp@br.ibm.com

ABSTRACT

In this paper, we present a study to identify similarities and differences in how users express themselves on Twitter during two editions of the most watched sports events in the world, the finals of the FIFA Soccer World Cup of 2014 and 2018. Our findings suggest in 2014 users tended to post more negative content than in 2018, while less hateful and offensive messages were posted in 2018 than in 2014. This study also showcases the challenges of performing analysis of emotional reactions on sports-related posts due to the specificity of the colorful language employed by the fans.

1 INTRODUCTION

Sports are known by their capacity of generating emotional responses which influence both players and spectators. However, it is difficult to measure emotion dynamics of fans during a match, especially during a major worldwide event like the FIFA's Soccer World Cup. With the popularization of mobile phones, sports fans increasingly interact among themselves using social media during the match, displaying their emotions in social posts [5].

We chose to analyze the online behavior of soccer spectators because they are known by their strong emotional reactions which can vary depending on the spectator perspective [2]. Also, the World Cup is the most watched sports competition in the world: an estimated 700 million people followed the final match of the 2014 World Cup.

Our main contribution is a quantitative exploration of the use of emotional language during the 2014 and 2018 final matches of the FIFA's Soccer World Cup. Our findings suggest that the emotional language used in each edition differs in many ways, for instance, tweets posted during 2104 final present a more negative tone on average, compared to the 2018 edition.

In both datasets, we found a low fraction of tweets classified as hate speech or offensive language and we discuss the false positive issues with the hate speech detection approach. Soccer is particularly a game where unfortunately it is common for fans to use offensive vocabulary (for instance, racism, sexism, and homophobia) what makes hate speech and offensive classification important. We discuss some of the challenges of detecting hateful and offensive messages shared during those matches, pointing out limitations on the existing hate-detection tools.

WebSci '19, June 30-July 3, 2019, Boston, MA, USA

ACM ISBN 978-1-4503-6202-3/19/06…$15.00
https://doi.org/10.1145/3292522.3326051

2 THE WORLD CUP TWEET DATASETS

The datasets used in this work are comprised of Twitter posts in English language collected during the final matches of the 2014 and 2018 World Cups: Argentina vs. Germany (July 13th of 2014) and Croatia vs. France (July 14th of 2018). Those sets have been downloaded from different sources. The posts from 2014, to which we refer here as *Firehose 2014* dataset were gathered by means of the use of the Twitter *Firehose*, and contains over 4 million tweets. On the other hand, the posts from 2018, referred to here as the *Streaming 2018* dataset, were collected with the Twitter *Streaming API* (around 771 thousand tweets).

We are aware that a direct comparison of those two datasets has to be done with caution due to the fact that the sampling methodology of the Streaming 2018 dataset is not completely known. However, we argue that, for the goals of this paper, comparing the data in the two sets is fair, following the discussion and arguments of previous work [4]. Wang *et al.* [4] showed that Twitter data samples are able to preserve enough information for analyses focused on the general tweets (e.g., user activity pattern characterization, event detection) or content statistics (e.g., sentiment analysis).

Analyzing the Polarity of the Tweets. We start analyzing the textual content of the tweets by extracting the polarity of the fan's opinions. For this analysis, we utilized *SentiStrength*[1] a state-of-the-art sentiment analysis tool which gives a score that varies from -4 to +4 (negative to positive polarity). Before applying the tool, we filtered out from the tweets URLs, symbols, emojis, and mentions, and we perform lemmatization to the final tweet texts.

Figure 1 shows the mean polarity of the tweets over time in the two datasets. Note that tweets posted in 2014 stays most of the time below the neutral line while the tweets posted in 2018 seem to be more positive than negative. In both datasets, we found that most of the tweets were neutral (45% in 2014 and 53% in 2018). However, 21% and 34% of the tweets were classified as positive and negatives, respectively, in 2014; while in 2018 27% and 20% were classified as positive and negative, respectively.

Analyzing each selected event, we see that users were on average more positive right before the full time completion in both years. In 2014, the weird collision of Argentinian player Higuain with German goalkeeper Neuer had on average a negative view from users while in 2018, the second Croatia goal was also seen as negative, probably because of the failure of the French goalkeeper.

Hate Speech and Offensive Language. We also studied the presence of hate and offensive language. We follow the same hate speech definition proposed by Davidson *et al.* [1]: a "*language that is used to expresses hatred towards a targeted group or is intended to be derogatory, to humiliate, or to insult the members of the group*".

[1] http://sentistrength.wlv.ac.uk/

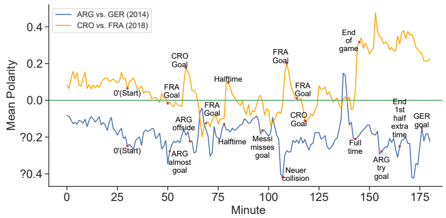

Figure 1: Mean polarity over time.

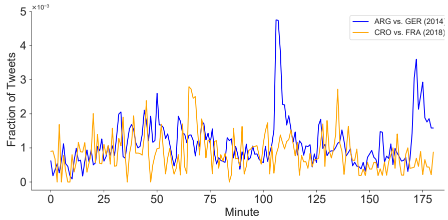

Figure 2: Hate speech tweets over time.

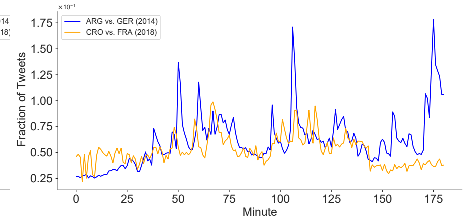

Figure 3: Offensive tweets over time.

For this analysis, we start using a classifier trained with generic tweets composed by posts labeled in three classes (hate speech, offensive language, and regular posts) released as described in [1]. Figures 2 and 3 plot the percentages of hate speech and use of offensive language over time during each match, respectively. Surprisingly, the fraction of tweets classified as hate speech is very low reaching up to only 0.005 in 2014 and 0.003 in 2018. Similarity, offensive language is also considered lower reaching up to 0.18 in 2014 and 0.1 in 2018.

We have also investigated some of the peaks shown in the plots (see Table 1). In 2014, the peak corresponds to the "Neuer collision", in which German goalkeeper Neuer clashed and knocked out Argentinian forward player Higuain. We observed many posts using the word "kill". This type of expression is common in soccer, a kind of hyperbolic language which is not necessarily a real case of hate speech. In 2018, the fourth goal of France team also reached a peak of hate speech due to the use of the word "kill". Offensive words seem to coincide with some major events in matches.

We also use another approach to detect hate speech proposed by Mondal *et al.* [3] which is based on the sentence structure of the text. Using that method, we found posts which match the template with "hate" expression accounted for the majority of matches in both datasets (57% for 2014 and 41% for 2018), while the word "n*igga" and "n*gger" accounted for 51% and 1.52% of the 2014 and 2018 posts. We did not find clear hate speech posts targeting specific nationalities, gender, or ethnicity using this tool.

As we notice, detecting hate speech in soccer (e.g., sports) related posts is not trivial. While analyzing the datasets, we observed many examples of false positives of both hate and offensive language. As mentioned, sports events evoke strong emotions from spectators who often use very expressive and colorful language [2] which tends to "mislead" automatic detection tools. Both state-of-the-art strategies of hate speech detection used here were not efficient to detect hate speech on soccer events [3]. We manually foraged for examples of false positives analyzing some of the posts which were not classified as hate speech and listed some of them in Table 2.

It is important to point out that the frequency of real hate speech may be lower in the Streaming 2018 dataset since Twitter has started restricting hate speech use on its platform in 2017. Their *hateful conduct policy* forbids users from directly attacking individuals on basis of characteristics such as race, sexual orientation or gender.

Example	Class	Event	Dataset
I think Neuer just tried to **kill** Higuain. How the f*ck did Higuain foul him.	hate offensive	Neuer collision	2014
Argentina is so g*y f*cking If you dare f*cking injure Messi	hate offensive	Peak before GER goal	2014
What a stupid decision! **F*ck** VAR **Bullsh*t** penalty. Joke final.	hate offensive	2nd FRA goal	2018
Mbappe that **n*gga** Holy **sh*t** Mbappe!!	hate offensive	4th FRA goal	2018

Table 1: Tweets found during hate speech peaks.

Examples	Dataset
You'd swear Germany watched Hitler's WWII speech in the locker room at halftime.	2014
Who say "Go Germany" actually mean "Heil Hitler"	2014
100 yrs from now". The last non-colored Frenchm...	2018
Kante is one of the most intelligent players I have ever seen. The pundits will never talk about a black player like that.	2018
Afrikans supporting France coz "they're represented by black players" shows how ignorant we are in this continent	2018

Table 2: Examples of no detected hate speech.

3 CONCLUSION AND FUTURE WORK

We present a quantitative analysis of the content posted on Twitter during the two FIFA Soccer World Cup final matches of 2014 and 2018. The analysis consisted of applying emotional language techniques, such as polarity analysis, and hate speech and offensive language detection to evaluate two large sets of tweets collected during those events. Our work demonstrates the need to improve emotional language analysis tools since we show that current methods seem to suffer from lack of context or biases towards some words which may not be offensive on many occasions.

ACKNOWLEDGMENTS

This work received financial support from CNPq and FAPEMIG.

REFERENCES

[1] T. Davidson, D. Warmsley, M. Macy, and I. Weber. 2017. Automated Hate Speech Detection and the Problem of Offensive Language. In *Proc. of ICWSM'17*.
[2] J. Gratch, G. Lucas, N. Malandrakis, E. Szablowski, E. Fessler, and J. Nichols. 2015. GOAALLL!: Using Sentiment in the World Cup to Explore Theories of Emotion. In *Proc. of the ACII*.
[3] M. Mondal, L. Silva, and F. Benevenuto. 2017. A Measurement Study of Hate Speech in Social Media. In *Proc. of the HT*.
[4] Y. Wang, J. Callan, and B. Zheng. 2015. Should We Use the Sample? Analyzing Datasets Sampled from Twitter's Stream API. *ACM Transactions on the Web* 9, 3, Article 13 (June 2015), 23 pages.
[5] Y. Yu and X. Wang. 2015. World Cup 2014 in the Twitter World: A big data analysis of sentiments in U.S. sports fans' tweets. *Computers in Human Behaviour* 48, C (jul 2015), 392–400.