

TWITTER SENTIMENT ANALYSIS: FAN ENGAGEMENT IN ESPORTS MATCHES

Sarah Anne Yan *
Wellesley College
syana@wellesley.edu

Peter Mawhorter
Wellesley College
pmawhort@wellesley.edu

ABSTRACT

Sentiment analysis is used to gain an understanding of the opinions, emotions, and subjectivity of text. Twitter is a social networking service where millions of users post and interact with messages. For multiplayer video game competitions, known as esports, many fans use Twitter as a platform to react to the match progress and results. In this study, we utilize the Twitter API and analyze tweets relating to Overwatch League matches using TextBlob to compute the sentiment polarity based on POS tagging. Each tweet will be assigned a score in the range of -1 to +1. We hypothesized that tweets mentioning winning teams would include more positive sentiments on average than tweets mentioning losing teams, but our preliminary analysis (with limited data) indicates that the relationship between sentiment and match outcomes may be more complicated.

KEYWORDS

Esports, Twitter, Sentiment analysis, Polarity analysis, Flashbulb memory, Python.

1. INTRODUCTION

Multiplayer online battle arena games, such as Dota 2, real-time strategy games, such as Clash Royale, and first-person shooter games, such as Overwatch, all have active esports scenes: virtual leagues where professional players compete in a manner similar to physical sports, often sponsored by the developer of the game. Esports have seen a steady growth in both viewership and participation in the past few years. In 2017, 33% of U.S. esports fans are younger millennial consumers (ages 18~24) (The Nielsen Company, 2018). With fan data from channels like Twitter, Twitch, and YouTube, we have an ability to look at their aggregated expectations and opinions of the team's performance. In the 2019 season of the Overwatch League, known as OWL (Blizzard 2020), the stage 2 finals were the best performing esports broadcast across ESPN and ABC (Activision 2019). In this paper, we will focus on Overwatch League fan data gathered from Twitter.

Public opinion about teams changes based on their match performance. In a study that looked at fans of winning and losing teams, Talarico and Moore found that the two groups of fans remembered the same game differently (Talarico and Moore 2012). Winning teams' fans found the game to be more emotional and remembered the factual details more accurately. The other fans' memories had decreased vividness. These memories can be classified as flashbulb memories, which is a subset of autobiographical memory (Brown and Kulik 1977). They are recalled with distinct vividness, sense of reliving the memory, and confidence. According to several studies, when relating memories of an event, participants who viewed the event as positive were more vocal than those who viewed it as negative (Berntsen 2009). If fans of the winning team were generally more vocal, we would expect the overall balance of sentiment to be more positive for winning teams. We investigate this hypothesis by looking at the public statements about two opposing teams on Twitter in a given Overwatch match using sentiment analysis.

Sentiment analysis is a text analysis method used to understand the opinions, emotions, and subjectivity of text. This can be done at document, phrase, and sentence levels, and often assigns a simple polarity score to text, measuring the degree to which the text conveys positive or negative sentiment. The Natural Language ToolKit, NLTK, is a platform that was developed by Loper and Bird for natural language processing education (Loper and Bird 2002), which includes sentiment analysis tools. Previous works have algorithmically analyzed categories of fan engagement from tweets based on Part Of Speech (POS) tagging and per-word sentiment scores (Bhangle and Krishnan 2017). In order to understand the effectiveness of this technique in our target corpus, our initial hypothesis is quite simple: in a given Overwatch League match, we expect tweets mentioning the winning team to have more positive average polarity than tweets mentioning the losing team.

2. METHODOLOGY

There are four major steps in our study, starting with data collection. We used the Twitter Standard Search API (via the Tweepy Python library) to collect the data. Twitter's standard search API deletes data after 5~7 days, so after each Overwatch League (OWL) weekend of matches was completed, we downloaded the data immediately. For each OWL team, there are official and unofficial Twitter hashtags that are used as well as official team Twitter accounts. For the Twitter API call, we used these hashtags and official account IDs to search for a team in a specific match. For ease of filtering, we only included tweets in English. The example below shows how we utilized the API's query parameter to look at tweets about the Boston Uprising OWL team during the match between Boston Uprising and LA Valiant.

```
query = "BostonUprising OR #BostonUp -filter:retweets -ShanghaiDragons AND -#Breakthrough AND -Charge AND -Hunt AND -spark AND -TorontoDefiant AND -#RiseTogether AND -ParisEternal AND -Spitfire AND -NYXL AND -#EverUpward AND -Reign AND -Mayhem AND -Justice AND -SFSHock AND -#Winnable AND -LAGladiators AND -SeoulDynasty AND -TigerNation AND -DallasFuel AND -#BurnBlue AND -#PlayWithFire"
```

For this string, we included any tweets that contain Boston Uprising's Twitter account ID or their official hashtag, while excluding any tweets that were retweets or which contained any Twitter account IDs or official hashtags from teams not in the match. We used searches to collect two different datasets per team per match. The first search was for pre-match fan engagement, starting 2 hours prior to and ending 10 minutes after the start of a match. The second search aimed to measure engagement during and after the match. Given that the same team played twice during a weekend (Saturday and Sunday), tweets about both matches would overlap if the search extended past the 12-hour mark. We shall refer to pre-match fan engagement as the Prior-Match Period and the 12-hour search period as the Match Period. To avoid potential biases, only the first set of matches in any given OWL weekend was collected.

After gathering the raw data, the next step is post-filtering. Post-filtering is necessary since some OWL teams utilize account IDs and hashtags that are also used in other contexts. Common non-related words are "military", "army", "virus", "college", and "law". These words were generated by adding non-related topics from 10 random tweets in each category (positive, negative, and neutral) to the filter array. We continued to add to the filter array until there were no unrelated tweets for any of the 3 categories of tweets. For each tweet, if they contained any of the filtered words, they would be excluded from the final tweet collection. From a single match between Team A and B, for example, the same post-filtering was applied to all 4 datasets: Team A's Prior-Match Period, Team B's Prior-Match Period, Team A's Match Period, and Team B's Match Period. While this post-filtering may have excluded some tweets that are about OWL, our initial observations confirmed that the tweets filtered out were overwhelmingly unrelated.

The third step in our procedure is analyzing the tweets using sentiment analysis. We utilized TextBlob, which is a Python library for processing textual data built on top of NLTK. TextBlob provides an easy-to-use API for diving into common natural language processing (NLP) tasks (Loria 2020). For sentiment analysis, TextBlob uses a Movies Reviews dataset, which has already been hand-labeled as positive or negative. We passed the post-filtered datasets through TextBlob and measured the average polarity of the tweets in each dataset, ignoring neutral tweets (tweets with 0 polarity).

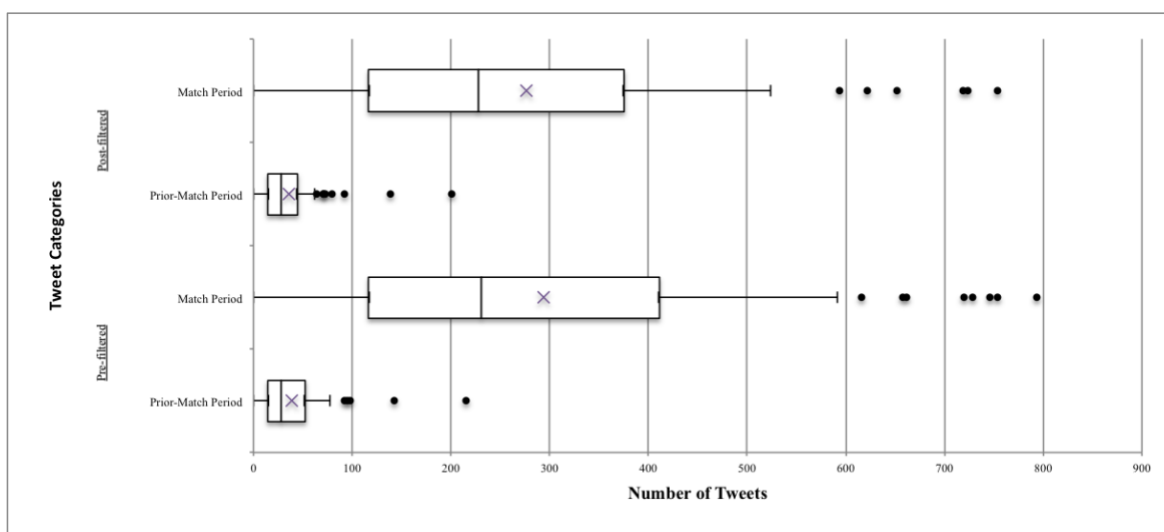
The final step in our analysis is determining statistical significance. To test our hypothesis, we compared the average polarities for tweets associated with different teams in both winning and losing matches. To compare the two groups of datasets (Match Period datasets from winning vs. losing matches) we ran the Student's t-test by using the t-test function in Google Sheets.

3. ANALYSIS AND RESULTS

We collected and analyzed tweets during the Overwatch League 2020 Season. For the 25 matches that we collected, there were 16,649 tweets and after the post-filter was applied, there were 15,653 remaining tweets, or an average of 626.12 tweets per match.

3.1 Data Collection

Figure 1. Prior-match period and match period tweets collected



These 25 matches were from weeks 11, 13, 14, and 15. Out of 25 matches, 16 required post filtering. Figure 1 shows the number of tweets we gathered across all of our Prior-Match Period and Match Period datasets. There were an average of 38.7 tweets in the Prior-Match Period (minimum 0, maximum 201, with an average of 36.3 after filtering) and an average of 294.28 tweets in the Match Period (minimum 26, maximum 754, with an average of 276.16 after filtering).

3.2 Sentiment Analysis

Table 1. Examples of collected tweets with polarities

Tweet	Polarity
@BostonUprising it was hard fought but I know you will keep improving and surprising everyone.	0.204166667
@Closer @SFShock Nah. Y'all did good https://t.co/YDeoVeGHAp	0.7
@Crimzo @SFShock You guys popped off	0
This game is kind of sad. Numbani sucks to attack on anyway, rip Outlaws.	-0.15
@Outlaws Hey outlaws, a tip of advice. Fire that sorry sack of crap you call a coach. Sincerely, a former fan	-0.433333333
I BELIEVED. I BELIEVED THEY CAN WIN WAHOOOOOOOO https://t.co/aGChgT4Kex	0.8

After we applied TextBlob to our data, we calculated the average polarity for tweets with non-zero polarity for every team in every match by using just the Match Period tweets. Table 1 lists some examples of the collected tweets plus TextBlob’s calculated polarity. The third tweet is an example of misclassification: it is labeled as 0 for polarity, but in the context of gaming, saying that a player “popped off” indicates they played well. Out of 25 matches, 17 had higher average polarity for the winning team, as shown in Figure 2. We ran a paired two-sample two-tailed Student’s t-test to determine whether the mean difference in polarity between the winning and losing side of each match was different from zero (our hypothesis was that the mean difference would be positive, indicating that the winner of each match was mentioned in overall more positive tweets than the loser). Although our results did not reach statistical significance, Table 2 shows that as we have been able to collect more and more data, the p-value for our t-test has steadily decreased.

Figure 2. Comparing winning team and losing team average polarities.

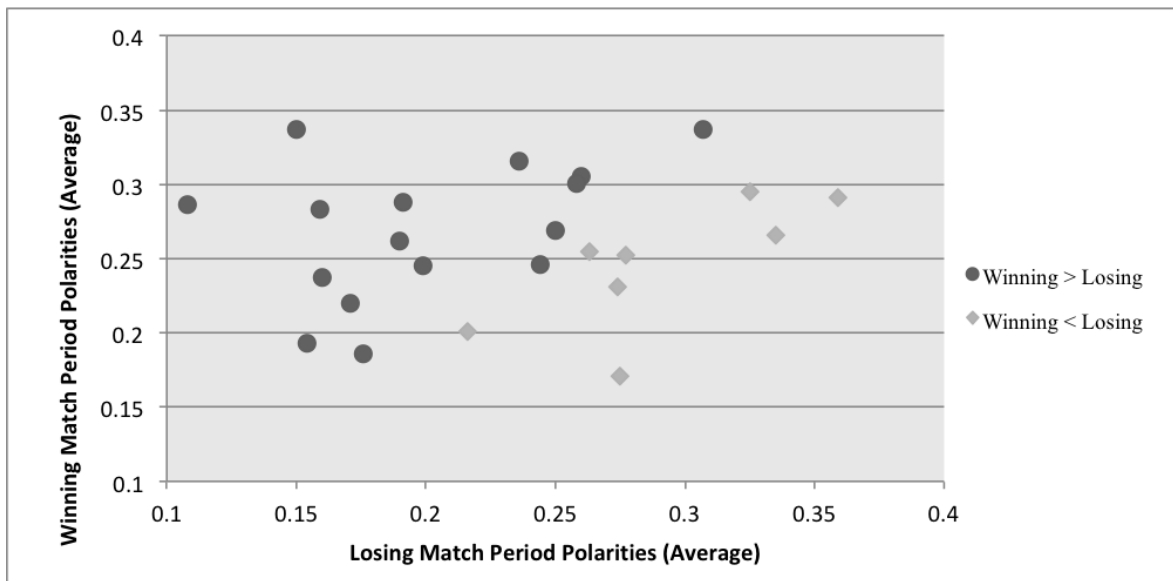


Table 2. T-test p-values for the mean difference in average polarities for winning and losing sides of each match as more data points are added.

Week	end of week 11	end of week 13	end of week 14	end of week 15
P-value	0.658	0.471	0.274	0.158

3.3 Limitations

Fan engagement overall has recently decreased due to OWL moving from the Twitch streaming platform to YouTube. For this season, the average number of viewers at a time was around 44,000 whereas the 2019 season had an average of at least 100,000 viewers. Those games were available to watch in person at homestead arenas, but due to COVID-19 circumstances, matches were postponed during the middle of the current season and subsequent matches (including those we collected data for) were only viewable online.

Furthermore, some teams simply were not mentioned in many of the tweets collected by the search parameters, and most of these are teams that have predominantly Asian players. Given this, their supporters may also be predominantly Asian and would therefore be tweeting in their respective languages and not in English. Also, the classification of tweets is not perfect. The high percentage of neutral tweets might be explained by the classifier not picking up positive or negative language in the context of a video game.

4. CONCLUSION

4.1 Implications

With the data we were able to gather for this study, the p-value for our hypothesis is 0.158, which is not significant, indicating that the relationship between tweet polarity and match outcomes may be subtler than our initial hypothesis proposes. We are hopeful that our originally proposed relationship may be confirmed with more data, but the fact that it is not confirmed given the data we have collected so far suggests that the relationship is not a straightforward one. If we do achieve statistical significance, it would imply that the public opinion on Twitter is more favorable to the winning team than the losing team, but the effect size might be small. Our analysis so far seems to indicate that sentiment analysis can be productively applied to look for patterns in Twitter responses to OWL matches, but we may need to develop a more sophisticated theory about what kinds of relationships might exist.

4.2 Future Work

Aside from collecting more data to potentially confirm our hypothesis, future work could go in a few different directions. We could try to identify teams' fan accounts and more directly explore our hypothesis in the context of fan engagement by looking at tweets from those specific accounts. With enough data, we could also look at how the public opinion and fan base opinion change based on the closeness of a match, such as comparing the polarities between 3-2 matches and 3-0 matches. As for the sentiment analysis part of future research, changing how the tweets' polarities are calculated such that they account for gaming linguistics as pointed out in Section 3.2 could be productive. There are other Python libraries for sentiment analysis such as SpaCy, Stanford CoreNLP, and Gensim; one of these libraries might yield more accurate results for polarity.

REFERENCES

- Activision Blizzard CEO says 2019 season of Overwatch League breaking records. (2019, August 8). The Fly. https://link-gale-com.ezproxy.wellesley.edu/apps/doc/A595916939/ITOF?u=mlln_m_wellcol&sid=ITOF&xid=3a564133
- Berntsen, D. (2009). Flashbulb memory and social identity. In O. Luminet, & A. Curci (Eds.), *Flashbulb memories: New issues and new perspectives* (pp. 187–205). New York, NY: Psychology Press.
- Bhangle R.S. and Sornalakshmi K. (2018) Twitter Sentimental Analysis on Fan Engagement. In: Rajsingh E., Veerasamy J., Alavi A., Peter J. (eds) *Advances in Big Data and Cloud Computing. Advances in Intelligent Systems and Computing*, vol 645. Springer, Singapore
- Blizzard Entertainment, INC. (2020). *Overwatch League*. <https://overwatchleague.com/en-us>
- Brown, R., & Kulik, J. (1977). Flashbulb memories. *Cognition*, 5(1), 73–99. [https://doi.org/10.1016/0010-0277\(77\)90018-X](https://doi.org/10.1016/0010-0277(77)90018-X)
- Loper, E. and Bird, S. (2002). NLTK: the Natural Language Toolkit. In *Proceedings of the ACL-02 Workshop on Effective tools and methodologies for teaching natural language processing and computational linguistics - Volume 1 (ETMTNLP '02)*. Association for Computational Linguistics, USA, 63–70. <https://doi.org/10.3115/1118108.1118117>
- Loria, S. et al. (2020). *TextBlob: Simplified Text Processing*. TextBlob. <https://textblob.readthedocs.io/en/dev/>
- The Nielsen Company (2018). *Distribution of eSports fans in the United States in 2017, by age group*. In Statista: The Statistics Portal. <https://www.statista.com/statistics/532322/esports-viewers-by-age-usa/>
- Talarico, J.M. and Moore, K.M. (2012). Memories of 'The Rivalry': Differences in How Fans of the Winning and Losing Teams Remember the Same Game. *Appl. Cognit. Psychol.*, 26: 746-756. doi:10.1002/acp.2855