Event Detection and Live Summarization through Sports Tweet Analysis

Megan Williams
University of California, Los Angeles

Abstract— **Since the advent of Twitter in 2009, social media analysts have observed a phenomenon known as live-tweeting, wherein users provide short, real-time commentary of a sports game or other event. In aggregate, these tweets can provide insight and information about a game. In this project, we determine what information can be gleaned from sports tweets, and use this information to automatically generate new tweets summarizing major events. Using a recurrent neural network to produce new tweets, our program has successfully live-tweeted several games from the 2019 NFL season and post-season.**

I. Motivation and Background

A. Background

During the 2015 Superbowl, viewers wrote over 28 million tweets about the game. Each tweet is a short (140 characters or less) statement that represents a small snapshot of the game. While each tweet may not contain much information, the combination of tweets can paint a complete picture of the game. When a significant game event occurs, users tend to tweet about it. Therefore, we believe that by looking at the volume and content of live tweets we can extract specific information about what occurred during the game, including when teams score or turn over the ball, as well as other interesting information.

B. Previous Work

Event detection in sports games is not a new concept [3]. However, most existing papers focus on detecting a narrow set of game events including touchdowns, field goals, fumbles and interceptions [6]. While these events are no doubt important moments of the game, they are not a comprehensive list. For instance, a critical moment of the 2019 Cowboys-Bears game was a sack of the Cowboys quarterback in the third down that forced the Cowboys to give up the ball, a turning point in the game which effectively ended their chance at winning. Of all of the papers that we read, none of them detected sacks. In our paper we attempt to develop a method which can summarize any interesting event without restricting ourselves to a predetermined set of possible events.

C. Goal

This project has two goals. The first is to analyze tweets from past NFL games, and to determine what information we can obtain from these tweets. Specifically, we would like to infer when significant game events occurred, and if we can assign these events to a team. To accomplish this, we analyze tweets from the 2015 Superbowl. For the second part of this project, we wish to create a model which can detect events in real time, as well as generate a new tweet characterizing the event, with the ultimate goal of live-tweeting the game. This model will be tested on select games during the 2019 NFL season and post-season.

II. Investigation of Sports Tweets

A. Volume of Tweets

We can recover a surprisingly large amount of information from the volume of tweets generated, even without looking at the content of tweets. We analyzed tweet volume from previous Superbowls, and found that the number of tweets tends to spike when interesting events occur. Figure 1 shows tweet volume during the 2015 Superbowl. Several of the spikes appear to correspond to critical game moments, such as an interception that occurred at minute 22 and a touchdown at minute 42. Therefore we conclude that we can use tweet volume as well as the topography of the tweet peaks as features in determining whether an interesting event has occurred.

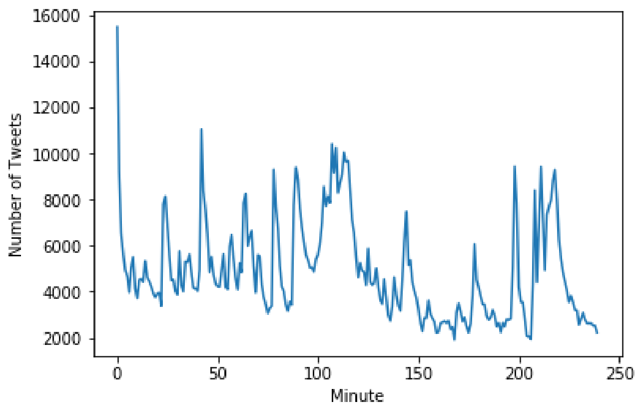

Fig. 1. Total number of tweets during each minute

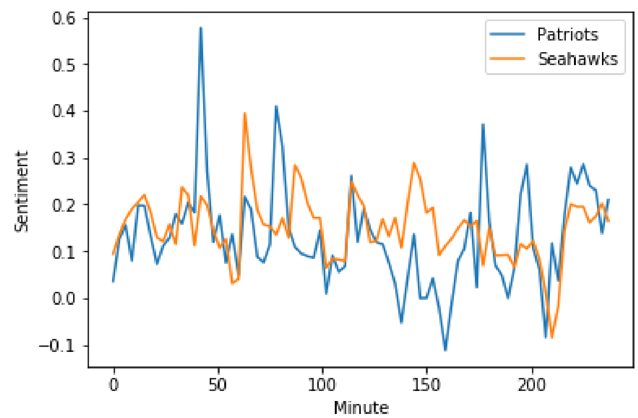

Fig. 2. Average fan sentiment during each minute

B. Sentiment Analysis

While a spike in tweets can be a good indicator of game events, relying on the number of tweets alone can lead to many false positives. This is because other events, such as the halftime show and viral commercials can cause spikes as well. Figure 1 shows a major spike that occurred around the 100 minute mark, during the halftime show. Therefore, we need to use other features to predict game events. One such feature that we can use is fan sentiment. We begin by identifying users that are fans of each team. We define Patriots fans as those who have used the hashtag #gopatriots at least once, and Seahawks fans as those who have used #gohawks. We found 427 users who used both hashtags, and we ignore these users. Next, we identified roughly 20 positive words such as 'good', 'yes', 'win', and 20 negative words such as 'bad', 'lose', 'angry', which are found in sentimental tweets [5]. For each tweet we count the number of positive words and negative words in the tweet. A tweet can then be classified as positive, negative, or neutral depending on if it has more, less, or the same number of positive and negative words. During each time interval, we calculate the average sentiment for fans of both teams. This is shown in Figure 2. Significant game events show a large discrepancy in sentiment between teams, whereas during other events, such as the halftime show, teams tend to express similar sentiments. Therefore, we use the difference in sentiment between the teams as a feature to predict events. We define this as the sentiment score.

C. Event Prediction

Using tweet volume and sentiment analysis, we attempt to predict interesting events during the 2015 Superbowl. In practice we find that there is often a delay of a few minutes between an event and a tweet being posted. For this reason, we consider intervals of three minutes and we wish to predict whether or not a major event occurred in the previous three minutes. The game can be broken down into 73 three-minute intervals. Of these intervals, eleven contained significant events, including seven touchdowns, one field goal, and three interceptions. We assign a time period a label of 1 if an event occurred during that time period, and a label of 0 otherwise.

The model that we choose to predict events is a Random Forest Classifier. For each time period we use the number of tweets and fan sentiment to predict if an event occurred. To evaluate our model, we use leave-one-out cross-validation (LOOCV) over the 73 time intervals. In LOOCV, one data point is withheld for validation and the model is trained on the rest of the samples. The average accuracy, precision, and recall from our 73 predictions are shown in Table I. We also present the confusion matrix in Figure 3 to demonstrate the performance of our classifier. Ultimately, we were able to correctly identify nine out of the eleven game events. The two events that we failed to predict were both interceptions. One of the interceptions was overshadowed by a touchdown, which made it difficult to predict.

In addition to predicting events, we also wish

	Accuracy	Precision	Recall
Random Forest Classifier	0.93	0.75	0.82

TABLE I
LEAVE-ONE-OUT CROSS-VALIDATION RESULTS FOR EVENT PREDICTION

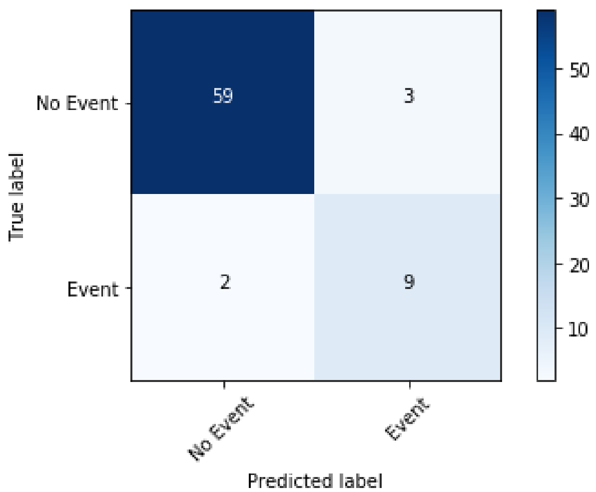

Fig. 3. Confusion Matrix for event prediction with the Random Forest Classifier

to assign the event to the team that benefited from it. We use the sentiment score to predict which team scored or made an interception, and we were able to correctly characterize 89% of our predicted events. The only event that we were not able to characterize was a field goal made by the Seahawks in the second half. This is likely due to the large amount of negative sentiments expressed by Seahawks fans at being forced to kick a field goal instead of scoring a touchdown. This particular issue makes characterizing field goals difficult.

Overall, we conclude that tweets can be used reasonably well to predict important events as well as characterize the event that occurred.

III. Game Summarization

A. Objective

We have concluded that tweets can indeed provide useful information about game events, and now we wish to utilize these tweets to generate a game summary. The second goal for this project was to create a program which can live-tweet a game. We wish to do this without giving our

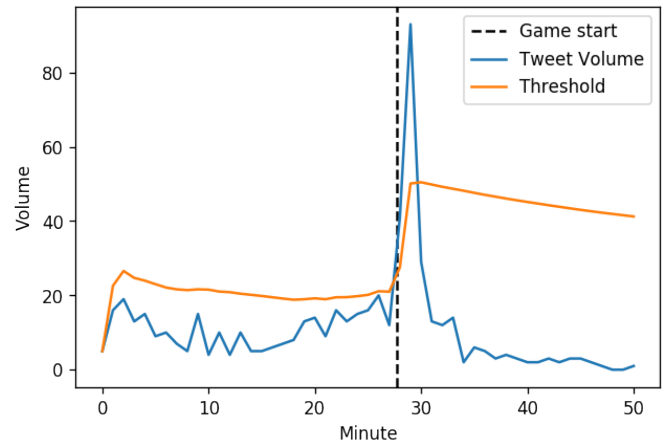

Fig. 4. Moving threshold for game detection

program any input about the game being played. The tweets generated from our program can be found at https://twitter.com/MeganWi14586600.

B. Game Detection

The first step is to identify when a game is starting. We do this by monitoring mentions of “NFL”. Specifically, our program looks for tweets containing words related to the beginning of games, such as “start” and “kickoff” [4]. We look for a spike in these tweets to identify the game start and use a moving threshold technique for the cutoff. We found that not only can our program identify when games are starting within five minutes of the start time, it is usually more accurate than the listed kickoff time. This is because games often start a couple of minutes late.

One of the biggest challenges in game detection is determining the threshold for the tweet spike that determines if a game is starting. Games like the Superbowl can get thousands of pre-game tweets, however, a regular season early Saturday morning football game between two less popular teams may only get hundreds. Therefore, we implement a moving threshold technique [1]. We look at the average number of tweets occurring each minute, and also the standard deviation in tweet volume. We determine a spike to be significant when the tweet volume exceeds the average by more than two times the standard deviation. Let $v_1, v_2, ..., v_k$ be the number of tweets binned into one minute intervals for minutes 1 through k. Then

$$m_i = \frac{1}{i}\sum_{k=1}^{i} v_k$$

and

$$s_i = \sqrt{\frac{1}{i}\sum_{k=1}^{i}(v_k - m_i)^2}$$

Then the threshold at minutes k, t_k, is given by

$$t_k = m_k + 2 * s_k$$

Figure 4 shows the volume of tweets related to the start of a game roughly 30 minutes before the 2019 Cowboys-Patriots game along with the calculated threshold. The game was supposed to start at 5:20 p.m., but the actual kickoff occurred at 5:27 p.m. A large spike in tweets occurs around 5:27 p.m., which we correctly detected as the start of the game.

C. Hashtag Identification

In general, each NFL game has one or more hashtags that people use to talk about the game. On Twitter, hashtags are words or short phrases preceded by the symbol '#', which are used to identify trending topics. For example, during the 2019 Cowboys-Bears game, the main hashtag that people used was '#DalvsChi'. In order to best determine what people are saying about a game, we wish to first identify the main hashtags for the game. This is particularly tricky, because at any given moment on Twitter, multiple hashtags can be trending for a number of reasons. In addition to the game hashtag, fans of a certain team also use team-specific hashtags such as '#GoBears' or '#CowboysNation'. We wish to avoid looking strictly at a team hashtag, because those tweets are often biased.

Once we have detected a game we aggregate all recent tweets mentioning 'NFL'. We then look for mentions of team names for the 32 teams in the NFL. We find the top two names. Then we find all hashtags co-occurring with mentions of the NFL. For each hashtag, we try to determine whether it favors one team or is neutral by looking at team name occurrence and performing sentiment analysis. If more than 70% of the favorable tweets for a hashtag mention one team and not the other, we consider that hashtag to be a team hashtag. Otherwise, the hashtag is considered neutral. Sometimes random, non-game related hashtags are detected, mostly due to spam bots. Therefore, we only pick the top hashtags to look at. Table II shows the hashtags we detected during the Cowboys-Bears game, including some seemingly off-topic ones, sorted by number of occurrences. The bolded game hashtags account for over 80% of tweets mentioning 'NFL', and the bolded team hashtags account for 90% of team name mentions.

Game Hashtags	**#DALvsCHI**, **#tnf**, #nfl100, #ps4live, #sportsbetting
Team 1 Hashtags	**#Cowboys**, **#DallasCowboys**, **#CowboysNation**, #nba, #cbb
Team 2 Hashtags	**#Bears**, **#GoBears**, #Beardown, #teddybeartoss, #shopmycloset

TABLE II
GAME HASHTAGS

D. Event Detection

Once we have identified relevant game hashtags, we measure the volume of tweets using these hashtags to identify significant game events. Earlier, we trained a Random Forest Classifier to identify events during the 2015 Superbowl, but we want to develop a model that can be generalized to any game and that works in real time. We decided to, again, use the moving threshold technique to identify important events. We chose to use the threshold cutoff as 1.5 times the standard deviation:

$$t_k = m_k + 1.5 * s_k$$

During the 2019 Patriots-Texans game, our program was able to successfully identify the start of the game, the teams playing, and the relevant game hashtag #NEvsHOU. We monitored the tweet volume using the game hashtag, and plotted the results in Figure 5. This figure also shows the moving threshold we applied to determine significant game events. Out of eight scoring plays, we were able to identify seven of them. Our model missed a late touchdown by the Patriots, likely because the Patriots were losing by a significant margin at this point in the game, and many fans had stopped watching or live-tweeting the game. In addition, we identified three other interesting

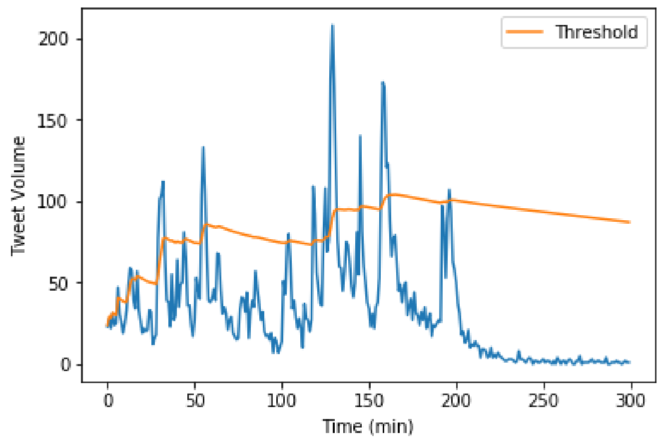

Fig. 5. Moving threshold for event detection

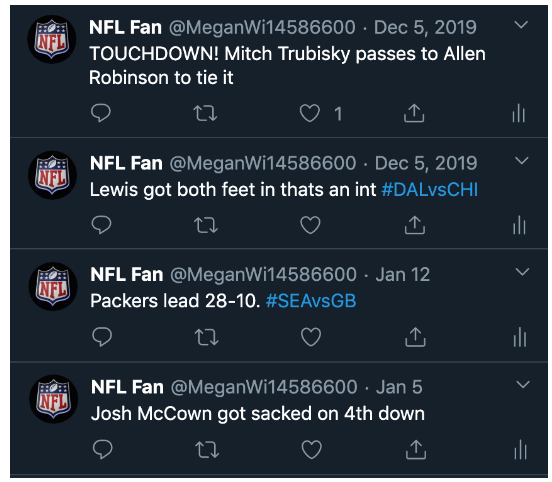

Fig. 6. Tweets generated from our model during detected events

plays, including an interception, a touchdown play that was reversed, and a pivotal sack of Tom Brady.

However, our method of detecting interesting events is not flawless. Our model is particularly susceptible to missing field goals. While there generally is a spike in tweets when a field goal is scored, the overall volume of tweets is generally much smaller than when a touchdown occurs in the same game. In addition, because our tweet volumes are binned in one minute intervals, we sometimes miss events that occur close together. For example, in one instance a touchdown was missed because it was too close to an interception that occurred several plays before. Lastly, we achieve the best results when the game is competitive. If one team starts losing by a large margin, those fans generally stop tweeting about the game, making it difficult to detect when that team scores.

Despite these drawbacks, moving threshold detection works well for our model, as it allows for flexibility in its ability to detect different kinds of events. Our model is even able to identify non-football related events. In one of the most surprising moments in the 2019 season, a black cat ran onto the field during the Giants-Cowboys game. Officials had to remove the cat before play could resume, and the game was delayed by several minutes. The cat instantly achieved internet stardom, and was later listed on the Giant's team roster. Our model was successfully able to detect the appearance of the cat as an interesting event, whereas other models looking for a fixed set of plays or events would not.

E. Generating Tweets

Finally, we take our detected events and generate tweets about them. To accomplish this, we use tweets from an event to train a recurrent neural network (RNN). The RNN that we use for this task is a basic model designed for text generation in the style of Shakespeare [2]. It consists of a single gated recurrent unit (GRU) layer with a sigmoid activation followed by a dense layer. We train our RNN on all of the tweets from an event time period. We then use the RNN to generate several new tweets. As the majority of users are tweeting about the event, the RNN will learn to generate tweets relevant to the event.

Ultimately, we found that we were fairly successful in generating tweets that accurately characterize an event. Several example tweets that we generated are shown in Figure 6. These tweets correctly identify a touchdown, an interception, and a sack that occurred. For example, the tweet "Lewis got both feet in thats an int" refers to the interception by Jourdan Lewis during the 2019 Cowboys-Bears game. In addition to identifying important events that have occurred, our generated tweets also recognized which players were involved. For example, one of the example tweets acknowledges a touchdown pass from quarterback Mitch Trubisky to wide receiver Allen Robinson. Our tweets can even identify the score. One tweet correctly reported the current score of a Seahawks-

Fig. 7. Questionable tweets generated from our model

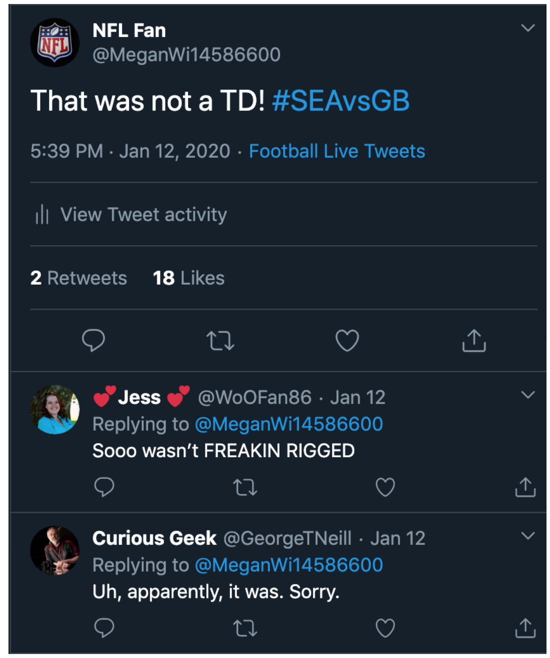

Fig. 8. A controversial live-tweet from our program

Packers game to be 28-10.

However, our method of generating tweets is not perfect. Several tweets were generated which make little sense or contain almost no useful information. Some of these tweets are shown in Figure 7.

To improve the quality of the tweets we generate dozens of tweets for a given event and select the best one among them. Each generated tweet is assigned a quality score, based on the frequency of each term in the corpus of training tweets. The idea is that we give a higher score to words most frequently used in the corpus as these words are likely related to the event that just occurred. We ignore stopwords like "the" and "and". We also give more weight to tweets that contain the game score such as "28-10".

In addition to providing summaries of scoring plays, our program has tweeted about other significant events during games, such as a game-ending sack of Eagles quarterback Josh McCown during the 2020 Seahawks-Eagles game. Another critical non-scoring play happened during the fourth quarter of the 2020 Seahawks-Packers game, when quarterback Aaron Rodgers completed a critical pass to seal the Packer's victory against the Seahawks. Our program tweeted "Aaron Rodgers with a clutch throw! #SEAvsGB".

Our model also occasionally expresses popular fan opinions and sentiment in addition to event summaries. During the 2019 Cowboys-Bears game the program tweeted "Cowboys shouldve kept Bailey, Maher keeps missing field goals", after a missed field goal attempt by Cowboy's Brett Maher. More generated tweets are listed in Table III.

Remarkably, our model learns grammar and syntax uniquely associated with Twitter. For example, our model picks up the usage of game and team hashtags, which commonly occur at the end of tweets. In one example, our model picked up the game hashtag "#SEAvsGB" during the Seahawks-Packers game, and used it in tweets such as "TD Seahawks 28-16 #SEAvsGB". Our model also learned the ability to tweet at other users, which is done by prefixing the "@" symbol to a username. Our program tweeted at the Chief's official Twitter account, congratulating them on a win after the 2020 AFC Championship with the tweet "Congrats @Chiefs! AFCChampionship TENvsKC".

IV. Conclusions

A fully autonomous version of this program is running and has live-tweeted several real NFL games. It tweets from a real Twitter account that can be found at https://twitter.com/MeganWi14586600, which the reader is encouraged to check out. The only prior information provided to the program is the names of the 32 NFL teams. All other information including which teams are playing,

Game	Actual Event	Generated Tweet
2019 Rams vs Seahawks	Rams score a touchdown.	TOUCHDOWN RAMS #SEAvsLAR
2019 Rams vs Seahawks	Quandre Diggs completes a pick six.	Diggs pick 6
2019 Rams vs Seahawks	End of regulation time. Rams win.	RAMS WIN IT #SEAvsLAR #LARams #SEAvsLAR #Seahawks
2020 Seahawks vs Eagles	Eagles block field goal attempt by Seahawks.	Blocked FG attempt #SEAvsPHI
2020 Seahawks vs Eagles	Quarterback Carson Wentz is injured, leaves the field.	Wentz going to the locker room #Eagles #SEAvsPHI
2020 Seahawks vs Packers	Seahawks score a touchdown, bringing the score to 28-16.	TD Seahawks 28-16 #SEAvsGB
2020 Seahawks vs Packers	Aaron Rodgers throws a 4th quarter completion that effectively ended the game.	Aaron Rodgers with a clutch throw! #SEAvsGB
2020 Superbowl	The 49ers throw their 2nd interception of the game.	Another #49ers interception #SuperBowl
2020 Superbowl	Superbowl halftime show.	Shakira is todays highlight #SuperBowl #HalftimeShow
2020 Superbowl	End of regulation time. Chiefs win the Superbowl.	Congrats to the Chiefs!! #SuperBowl

TABLE III

SELECT TWEETS GENERATED FOR DETECTED EVENTS

the game start time, and game hashtags are inferred by the program. The code can be found at https://github.com/mwilliams123/capstone.

The game summaries generated from our program are timely and accurate. After a significant event, our program generally tweets about it within three minutes. The summaries are accurate enough to be, for the most part, indistinguishable from real tweets. Over the course of several games, the tweets from our program have garnered nearly two dozen likes and a few retweets from other Twitter users.

One of our most popular tweets even managed to generate some controversy. The tweet came after a controversial touchdown call during the NFL divisional playoff between the Seahawks and Packers, expressing the opinion “That was not a TD! #SEAvsGB”. The tweet received several likes and re-tweets from other Twitter users, as well as several replies expressing agreement and disagreement.

Overall, while many tweets are not perfect, the program does an impressive job of identifying and characterizing interesting events.

V. FUTURE WORK

One area that we did not explore much was using different RNN architectures. The architecture we chose was purposefully minimal in order to reduce training time, allowing our tweets to be generated quickly. Perhaps a deeper architecture could provide better tweets. In addition, we would like to explore pre-training our RNN using a larger, general set of sports tweets. This would allow the network to pre-learn information about general sports vocabulary and Twitter syntax. Then, we would use event-specific tweets to fine-tune the network. In addition to improving tweet quality, we think this would allow us to use a deeper architecture without increasing the live training time.

Lastly, we would like to test our model on sports other than football, such as baseball or soccer. The fact that our model is flexible when it comes to detecting different types of events, and the fact that it uses a moving threshold for event detection rather than a fixed threshold, indicates that our model is highly adaptable and may be suited for other applications.

References

[1] Hsieh, L., Lee, C., Chiu, T., and Hsu, W. Live semantic sport highlight detection based on analyzing tweets of twitter. In *2012 IEEE International Conference on Multimedia and Expo* (July 2012), pp. 949–954.

[2] Karpathy, A. The Unreasonable Effectiveness of Recurrent Neural Networks, May 2015.

[3] Lanagan, J., and Smeaton, A. F. Using twitter to detect and tag important events in live sports.

[4] Oorschot, G., Erp, M., and Dijkshoorn, C. Automatic extraction of soccer game events from twitter. *CEUR Workshop Proceedings 902* (01 2012).

[5] Zhao, S., Zhong, L., Wickramasuriya, J., and Vasudevan, V. Analyzing twitter for social tv: Sentiment extraction for sports. In *in Proceedings of the 2nd International Workshop on Future of Television* (2011).

[6] Zhao, S., Zhong, L., Wickramasuriya, J., Vasudevan, V., LiKamWa, R., and Rahmati, A. Sportsense: Real-time detection of NFL game events from twitter. *CoRR abs/1205.3212* (2012).