Understanding Twitter Activity During Live Sporting Events

Stephen Armstrong
sjarmstr@ucalgary.ca

ABSTRACT

In this paper, we investigate the relationships between Twitter, a microblogging and social networking service, and sporting events. Sporting events have always been a social occasion and now with Twitter people are able to socialize remotely. The problem proposed is, does momentum, as reflected on social media, correspond with in game events? The approach we have used is to collect tweets posted during several Canadian Football League (CFL) games and apply several varieties of analyses, including sentiment and frequency analyses, to the dataset. From this analysis we have made inferences about what is happening within the game. In terms of football, we say the team with more time in possession of the ball has more "momentum" and thus control over the game. We have attempted to quantify momentum using human generated summaries of the game's progression. Our collected data was then compared to these summaries. Unlike other papers exploring similar topics, this paper has shown that by using multiple low level analyses we can draw conclusions about what happened in the sporting event such as how many penalties, touchdowns and fumbles occurred, as well as the times that they occurred.

KEYWORDS

Twitter, Sports, Momentum

1. INTRODUCTION

Twitter is a social networking and microblogging service which enables its users to post and read 140 character, text-based messages on a public forum. These posts, called tweets, contain a wealth of information about what the user is thinking, doing and feeling. During sporting events, Twitter users regularly post status updates about the game, as well as feelings and opinions about the game's progress [1, 2].

The goal of this research is to develop a tool which will allow a user to draw conclusions about what happened in the game based purely upon the twitter data collected during the game. In this project we aim to show that twitter can be used for much more than just social networking. The ideas applied here extend beyond just football. In general the data can be applied to marketing data collection. Collecting twitter data provides nearly instantaneous feedback to various marketing strategies. For instance, one could capture the tweets posted during commercials during major events such as the super bowl and then perform sentiment analysis. This will give advertisers vital information about how the ads have performed and the information will be available very quickly and without investing large amounts of manpower. Similarly, sentiment analysis of tweets could be used to determine public support for politics. For example, during a presidential debate, tweets could be analyzed to determine the political point of view, as well as what they like and dislike about each presidential candidate. These are just a few of the potential applications of the tool being developed here.

2. BASIC DENITIONS AND EXPLINATION OF THE PROBLEM

The problem proposed is whether or not a user can draw conclusions about what happened during a sporting event based purely upon the Twitter data posted during the game. More specifically, is it possible to develop a piece of software which will perform a variety of analyses on the data collected and show this data in a way that the user can draw conclusions about the game. One of the key ways the user will be able to do this is by creating a visual representation of the momentum of the game.

Now we will explore some of the basic definitions we will be using throughout this project. The first of which is twitter itself, specifically the aspects of a tweet. Tweets, as described before are short text based messages posted by a user. These messages often include what are referred to as hashtags, denoted by "#" followed by a word. These hashtags are used to mark keywords or topics on the message [3]. In this project we also discuss twitter's retweet system. This is when a user reposts a tweet by another user using twitter's retweet system. Retweets are denoted by "RT" at the beginning of the message [4]. Tweets can also contain references to other users. This is referred to as a mention and is denoted by "@" followed by the username. A mention can occur anywhere in a tweet. In addition to the mention is a reply. This is when a user makes a reply to another tweet. Replys are denoted by "in reply to @username" at the bottom of the tweet [5].

For the purposes of this study we will be evaluating Canadian Football League (CFL) football games. Football is a sport in which two teams of at least 12 players attempt to gain points by getting the football into the opposing team's end zone. Each CFL team can be split into two parts, the offence and the defense. The offence is charged with scoring and the defense's task is to prevent the opposition from scoring. In CFL football the offense gets three plays to move the ball at least 10 yards. If they move the ball 10 or more yards down field they are awarded a "first down" in which the number of expired plays is reset giving them 3 more plays to gain 10 more yards. If the offense does not move the ball 10 yards then a turnover occurs and the other team gains possession of the ball.

Momentum can be very difficult to define and varies between sports. However, commonalities can be identified to provide a simplistic definition of momentum. For instance, nearly every sport involves some sort of object to be used as a scoring device. It could be argued that the team which has possession of this object, usually a ball, is in a more advantageous position than the other team at that particular time. Thus, we can say that the team with more time in possession of the ball on offence has more momentum than the other team. To determine the momentum of each team we will divide the number of first downs by the number of possessions a team had. This will give us the number of first downs per possession. Since in football a first down extends the team's time with the ball, we can also assume it will increase the team's momentum. As such we will use this average first down value as our momentum value.

Another important way that the user will be able to draw conclusions about the game is by examining the frequency of tweets posted during a specific period of time. It stands to reason that if something exciting happens in the game, a goal for instance, then there will be an influx of the number of tweets shortly thereafter [1]. Based on the timestamps of each tweet collected we can determine the frequency in which the tweets are posted.

3. RELATED WORK

Similar work has been conducted in several papers. One example of such is a paper on "Summarizing Sporting Events Using Twitter" [1]. The problem presented was to create an algorithm which can create a summary of events using only Twitter status updates from the Twitter streaming Application Programming Interface (API). The authors wanted to see if it was possible to create a summary from the Twitter data which was as good as summaries published by sports writers. The authors collected tweets posted during soccer games and identified important events which occurred during the match. Summaries were then generated by copying the text from tweets which correlated to the events. These summaries were then compared to the ones written by sports writers for the same game. The conclusion reached is that the algorithm does produce a summary that could be interpreted by someone who was not watching the game [1]. Our paper will utilize similar means of filtering the data collected as was presented in "Summarizing Sporting Events Using Twitter".

Another paper, presenting a similar problem is "Personalized and automatic social summarization of events in video" [2]. The problem presented was to create a video highlight reel from time stamped Twitter posts. As in the previous paper, Twitter posts were collected using the Twitter streaming API based on key terms and hash tags relating to world cup soccer. Summaries were generated by either using frequency based data or content based data. The frequency based method would select the top number of documents with the highest number of tweets and assign a time slice. These time slices were concatenated together to produce the highlight reel. For the content based method, users based terms were provided to limit the content of the tweets available. Then the highlight reel was built in the same way using the largest number of tweets. The results found that the summaries were satisfying, but had some evident flaws. One example of such a flaw is that if the user submitted poor queries for the content based approach the results would be poor [2]. Like both of the previously discussed, papers this paper will be utilizing the Twitter streaming API to collect tweets for analysis.

Similar work can also be seen in the paper "TwitInfo: Aggregating and Visualizing Micoblogs for Event Exploration" [6]. Their system, TwitInfo, automatically identifies peaks in Twitter data and marks them as 'events'. Users can then view the events as well as a summary of what happened during the event. TwitInfo also makes use of sentiment analysis, in that tweets are classified as positive, negative or neutral and then a sentiment analysis is shown along with the search results. Users however, found the sentiment to be misleading as when they searched for topics such as "earthquake" they would often see a positive sentiment, not the negative one expected. Upon further investigation, users found this was mainly caused by tweets which offer well wishes to the people affected by the earthquake [6]. In our paper we will also be using sentiment analysis; however, we will be using the SentiWordNet [7] database to determine the polarity of the tweets.

SentiWordNet, as described in "SENTIWORDNET 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining" is an automatic annotation of "WORDNET" according to the notions of "positivity", "negativity", and "neutrality" [7]. The idea is to assign the definition of each word with three values: a positive score, a negative score and an objective score. Each score ranges from 0.0 to 1.0 and the sum of these scores must always equal 1.0. The user can then evaluate these scores and categorize the words based on sentiment [7]. This is crucially important to our work as it will be the basis of our sentiment analysis of the tweets. By using this database we hope to be able to identify which team has the favor of the crowd at any given time.

4. DEVELOPED SOLUTION

Our solution to the problem was to develop a tool which will collect the twitter data posted during the sports game, analyze this data and then visualize the data in a way which would aid the user in drawing conclusions about what happened during the game. Three parts were required to complete the tool. The first is a collector to gather the Twitter data. The second is an analyzer which will perform a variety of functions to interpret the data. The third and final part will be a visualizer to graph the data. In addition, game statistics were manually collected for use as a baseline to compare the tweets to. To build our dataset we collected tweets posted during several regular season CFL games.

4.1 Collector

The collector uses the Twitter public streaming API to collect tweets posted during CFL football games. The public streaming API will take search terms and return any tweets which include the specified search terms. The collector requests the search terms from the user which can include anything that can show up in the tweet message. In order to ensure that the tweets collected actually pertained to the sporting event in question we used "hashtags" and "mentions" involving the team names for our search parameters. Three streams were opened in total, one to collect tweets for each team and a third to collect tweets under the general #CFL hashtag. For example if there was a game between the BC Lions and the Calgary Stampeders we would start three streams, one would have the search parameters "#CFL", the second "#Stampeders, @Stampeders" and the third "#BCLions, @BCLions". The collector for each stream saved the data collected into separate comma separated value files (.csv).

Alongside the automated tweet collection we manually recorded a list of game events with timestamps based off of the live television feed. This list included every a team gained possession of the ball, gained a first down, touchdown, fieldgoal, had a penalty or interception, or injury. Any other interesting events were also recorded, for instance during one game a streaker ran across the field. The game events were saved into two .csv files, one for each team in the game.

4.2 Analyzer

The analyzer is the backend of the visualizer. It performs all the analyses of the collected data so that the visualizer can graph it. While the analyzer is a large portion of the software, the workings of each function will be described along with the visualizer. A key feature of the analyzer is that it reads in the twitter data from the .csv files in order to build a result set. The data files are organized in a specific way. Each line denotes a separate tweet. The first item on the line is the user name, second the date and time at which the tweet was posted, third is the tweet id number and lastly the tweet body containing the message. The analyzer reads each line and splits the strings into a list of tweets. The manually created event files are organized in a specific way, similar to the tweet data files. Each line denotes a type of event and each item in a line is an occurrence of an event.

The analyzer, using a user defined length of time, will create a dictionary of “time buckets”. Each time bucket has a list of events associated with it. This list includes all events, if any, which occurred during the bucket. The result is a list of all the events which occurred during the game and the general time in which they occurred. We use time buckets to condense the data into points so that we can effectively evaluate the tweet frequency and sentiment values. If we did not condense the data this way it would be much more difficult to identify peaks in the tweet frequency and sentiment.

4.3 Visualizer

The visualizer is the graphical interface of the software. As seen in figure 1, the visualizer can be divided into several sections. Section 1 allows the user to choose the input files which will be analyzed. Section 2 displays the tweet information. Section 3 allows the user to select from a variety of visualization options. Finally, section 4 is where the graph data appears.

Figure 1: Graphical User Interface

4.3.1 *Tweet Sentiment*

The visualizer can provide a large variety of information for the user. The tweet data in section two utilizes the result set built by the analyzer and puts it into a spreadsheet. It also calculates the sentiment of each individual tweet and displays it in the spreadsheet. The Sentiment analysis is performed using the SentiWordNet database. This is a database of words, and definitions. Each word and definition is associated with two values. The first value is the positive value which indicated a word has a positive meaning, the second is the negative value which indicates a word has a negative meaning [1]. The sum of these values will then range from -1 to 1. A value of -1 indicates a negative implication, 0 indicates neutral and 1 indicates a positive implication (Equation 1). For example, the word ‘living’ is given a positive score of 0.5 and a negative score of 0.125. So its score is 0.375.

$$\boldsymbol{Score_w = w_p - w_n} \qquad (1)$$

$$\boldsymbol{Score_t = \sum Score_w} \qquad (2)$$

$$\boldsymbol{Score_{at} = \frac{\sum Score_w}{\#\, words}} \qquad (3)$$

In this project we have calculated the sentiment in two different ways. In the first method the analyzer compares each word in a tweet to the database and sums the resulting values of each word (Equation 2). For example, a tweet of “go stamps go” has 2 unique words. “Stamps” the team name has a positive and negative score of 0.0 giving it a total score of 0.0. “Go” has a positive score of 0.5 and a negative score of 0.0 giving it a total score of 0.5. The score of the tweet is then the sum of all the words, so 0.5+0.0+0.5 = 1.0.

The second method divides the result of the first method by the number of words giving us the average sentiment of the tweet (Equation 3). For example, take the tweet “go stamps go” again the tweet score was 1.0 but there were 3 words in the tweet so its average score is 1.0 / 3 = 0.3.

In the case of the tweet information the sentiment is calculated using the first method. This is because we found this to be more representative of the tweet’s meaning. Since tweets are so short, tweets such a “go team go” score relatively low with the averaged sentiment method, however, they imply a very strong message.

4.3.2 *Visualizer Graphing Options*

The visualizer provides a large variety of graphing options. There are two graphs in the visualizer, the left graph shows all line graphs while the right graph shows bubble graphs and bar graphs. There are two types of line graphs available, a frequency graph and a sentiment graph. Each of these graphs has several options available which will be explained in detail in their section. The bubble and bar graphs depict a variety of information including, tweet sentiment, tweet frequency, and word frequencies.

4.3.2.1 *Frequency and Sentiment Graphs*

The graphs get their data from the input files selected in the data input section. A series is added to each graph for each input file included. For example, if 3 separate data files are included there will be 3 separate frequency graphs plotted as can be seen in figure 2.

For the frequency graph, the user defines the size of a time bucket, the length of the game is then split into buckets of this given size. The analyzer counts how many tweets were posted during each bucket and the visualizer graphs the resulting frequency over time relation in the left graph. An example of the frequency graph can be seen in the top graph of figure 2. The y axis represents the number of tweets posted and the x axis represents time. The frequency graph also includes an option to show an "event overlay" when enabled, this option will draw another line graph which has the number of events per bucket over time. When the user hovers over a non-zero bucket, a list of events which occurred during that particular bucket is displayed. Figure 6 is an example of the frequency graph with the event overlay enabled the y axis of the event overlay depicts the number of events that occurred during that time bucket and the x axis represents time. Like with the frequency graph itself, one event overlay will be plotted for each event file graphed.

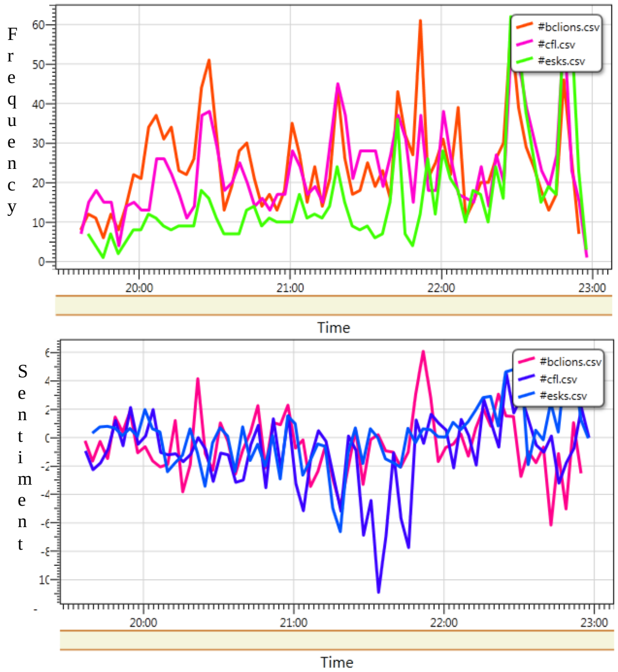

Figure 2: Frequency and Sentiment Graph

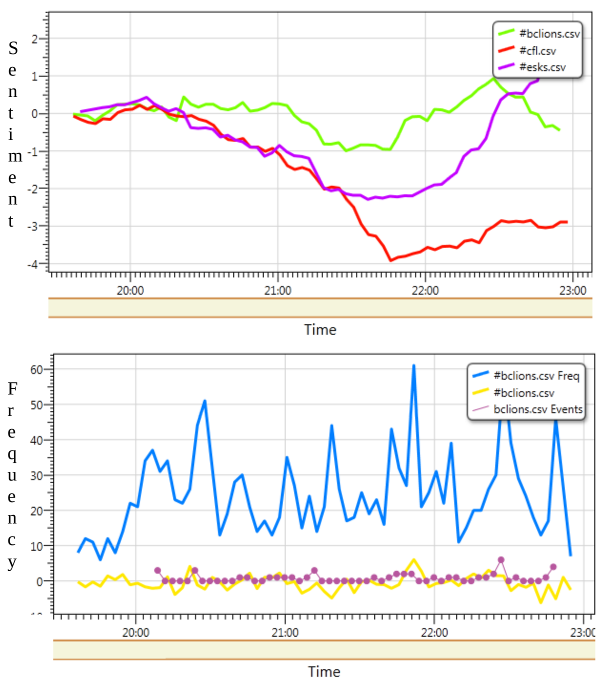

Figure 3: Sentiment Graphs with Options Enabled

The sentiment graph again utilizes a user defined bucket to graph the data. In addition, the user must also choose between an averaged graph, a summed graph or a graph containing both. The sentiment is calculated and summed within each bucket and graphed to the screen as can be seen in the bottom graph of figure 2. For this graph the y axis represents sentiment and the x axis represents time. Sentiment graphs also include options to enable both the frequency graph and event graph as overlays. The final option for the sentiment graph shows the cumulative sentiment of all previous buckets. These features can be seen in figure 3, the top graph shows the sentiment graph with the cumulative option turned on and the bottom graph has the event and frequency overlays enabled.

4.3.2.2 Bubble and Bar Graphs

The bubble and bar graphs have four options. The first option is a frequency bubble graph, which displays time in seconds on the x axis, the sentiment of the bucket on the y axis and the size of each bubble represents the number of tweets posted during that bucket.

The second option is a word frequency graph. This is a bar graph showing the 15 most common words posted for each input file. The number of words is restricted to 15 in order to improve the legibility of the graph. The words are listed on the y axis and the x axis represents the number of occurrences of the word.

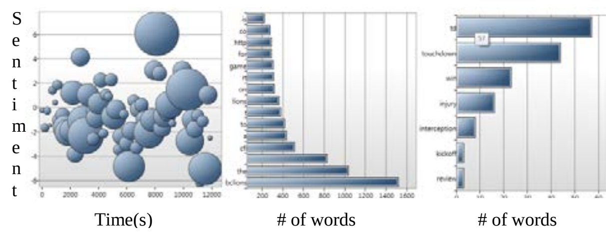

Figure 4: Bubble Frequency, Word Frequency, and Keyword Frequency Graphs

The third option is a keyword frequency graph. This graph displays the 15 most common words which also occur in a user defined "keyword" list in the form of a .txt file. The keyword graph allows the user to filter out words that are unimportant and focus on interesting words, unlike the word frequency graph which includes all words in the tweets. Like with the word frequency graph the keywords are listed on the y axis and the x axis represents the number of occurrences of the word. The frequency bubble graph, word frequency graph and keyword frequency graph are shown in figure 4.

The final graphing option is the significant event graph. This graph has a user defined "minimum tweets" variable which reflects the minimum number of tweets required to define an event. The analyzer takes the tweet frequency list and filters out any buckets which contain less tweets than the threshold. This resulting list is then represented as a bubble graph with x representing time in seconds, y the number of tweets and the size of the bubble also showing the number of tweets. The significant events function will also add a threshold line to the frequency graph and the sentiment graph if the frequency overlay is enabled. This will allow the user to better define the number of tweets for a significant event. Like with the line graphs, a new series is graphed for each data file included. An example of the significant event graph can be found in figure 5.

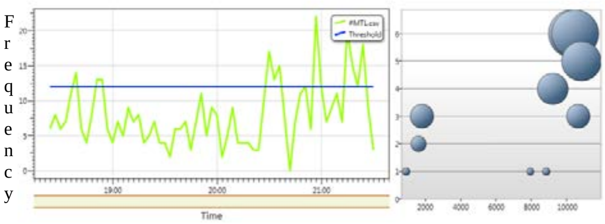

Figure 5: Significant Event Graph

5. EVALUATION

5.1 Identifying Events Based on Tweet Frequency

In total, data was collected for 4 separate regular season CFL football games. Upon inputting the data into the software some interesting patterns emerged. First, it was fairly easy to identify large peaks in the tweet frequency graph. These peaks, in most cases, occurred at the same time or just after some sort of event in the game. For example, in a game between Montreal and Winnipeg, Montreal intercepted the ball at 20:30:29 as seen in figure 6. The red line indicates the point in which the interception occurred. Each point on the graph represents a 3 minute time bucket. As such, during the bucket in which the interception occurred 17 tweets were posted. In the subsequent two buckets another 27 tweets were posted.

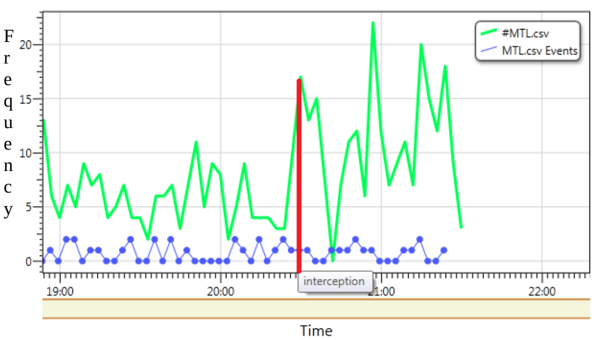

Figure 6: Montreal Interception Frequency

While not all of the tweets posted are related to the interception, the influx of tweets is consistent with the in game event. This pattern is repeated several other times as well. Figure 7 illustrates the peaks occurring concurrently with two touchdowns following the interception.

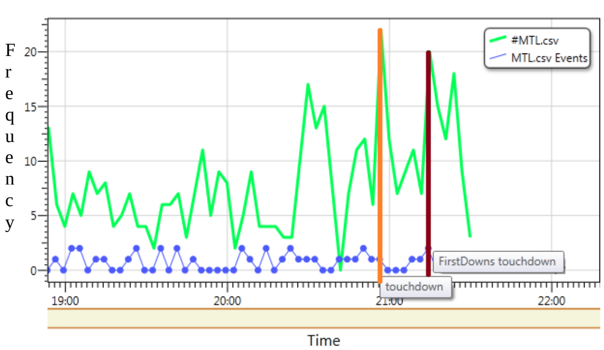

Figure 7: Montreal Touchdowns Frequency

Another example of the tweet frequency increasing alongside an in game event can be seen in figure 8. In a game between Saskatchewan and British Columbia three separate events can easily be identified. First, halftime starts; this event is depicted in figure 8 by a red line (leftmost line). In this particular case the BC Lions prevented the Saskatchewan Roughriders from scoring just before halftime leading to an increase of tweets. The second event is a fumble, depicted by an orange line (middle line), which had to go to review and was eventually overturned. The third event, shown in blue (rightmost line), is a touchdown for the BC Lions. In all three events there is a significant peak in the tweet frequency.

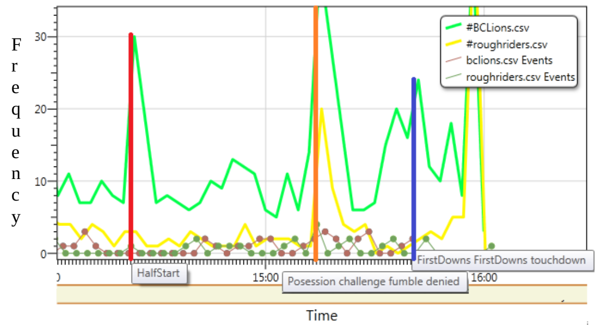

Figure 8: Saskatchewan vs. BC Frequency

5.2 Identifying Events Based on Tweet Sentiment

Like with the tweet frequency, peaks in the sentiment analysis are also representative of an event. In figure 8 we see the sentiment analysis of the game between BC and Saskatchewan. The time interval shown is the same as in figure 7. Once again halftime is represented by a red line (leftmost line), the fumble an orange line (middle line) and the touchdown a light blue line (rightmost line). It can be observed that during all three events there is a spike in the sentiment. The direction of the spike is dependent on whether or not the event is favorable for the team in question. For example, in the fumble event, BC lost the ball and so we see a net negative sentiment while Saskatchewan's sentiment is slightly positive.

This trend also continues in the cumulative sentiment. Figure 10 shows the cumulative sentiment of the game between Saskatchewan and BC. The event marked in red is the same fumble that we have observed in figures 8 and 9. As expected, we see a large decrease in the cumulative sentiment.

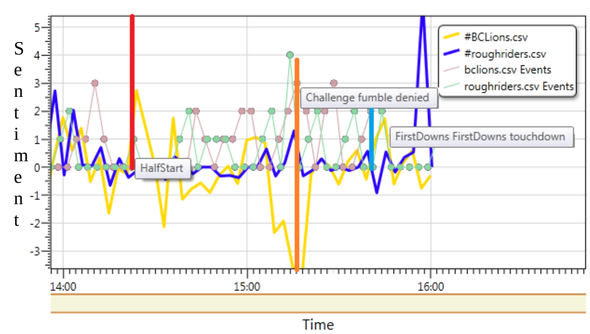

Figure 9: Saskatchewan vs. BC Sentiment

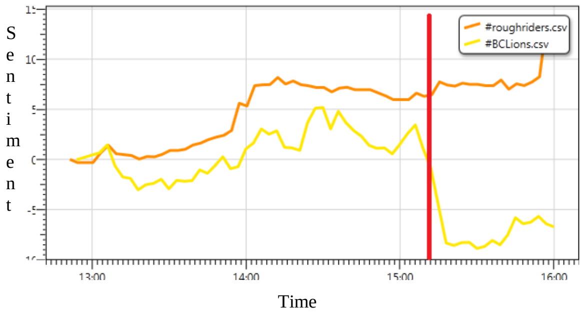

Figure 10: Saskatchewan vs. BC Cumulative Sentiment

We have also discovered that it is significantly easier to see negative events in the sentiment charts than it is to see them in the frequency charts. This of course makes sense, the frequency charts are objective and do not depict whether the event that occurred is good or bad but the sentiment charts do, thus, when looking at a specific team's sentiment graph you can determine whether the event was beneficial or not. Figure 11 shows a game between the Hamilton Tiger-cats BC Lions. The top graph shows the sentiment and the bottom graph the tweet frequency. The event marked with a blue line represents a touchdown scored by Hamilton on a kick return. The green line is a fumble by Hamilton which subsequently led to a touchdown by BC. We can clearly see in the frequency graph that something happens at both of these times but we cannot tell whether it was a good or bad thing. In the frequency graph you can clearly see a positive and negative peak at the respective events.

Figure 11: Tiger-Cats Sentiment vs. Frequency

5.3 Data Trends

5.3.1 Sentiment of Winning Team

Several other trends in the data have also been observed. In all data collected, the cumulative sentiment of the winning team, by the end of the game, is higher than that of the losing team. Figure 12 depicts this association in four sentiment-over-time graphs, one for each game. The winning teams are Montreal (top left in blue), BC (top right in green), Saskatchewan (bottom left in orange), and Edmonton (bottom right in blue).

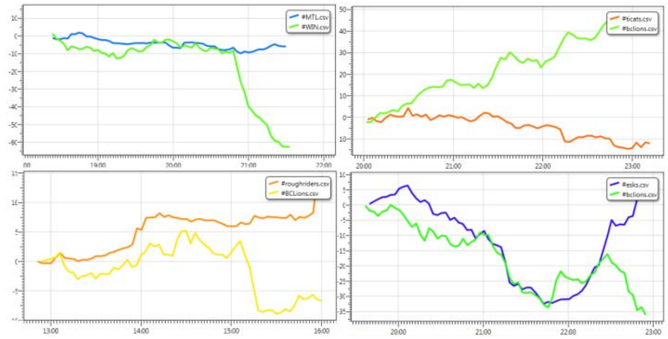

Figure 12: Cumulative Sentiment Across Games

5.3.2 Tweet Frequency of Home vs Away Team

Another trend which was observed is that in three of four games the home team has a higher frequency of tweets than the away team. In the game between the BC Lions and the Hamilton Tiger-Cats; the home team, BC, had 1559 tweets to Hamilton's 1097. In the game between the Montreal Alouettes and the Winnipeg BlueBombers, the home team, Winnipeg, had 1258 tweets to Montreal's 510. In the game between the BC Lions and the Edmonton Eskimos, the home team, BC, had 1646 tweets to Edmonton's 1050. The exception to the trend was the game between the BC Lions and the Saskatchewan Roughriders the home team, Saskatchewan, only had 257 tweets to BC's 750.

5.3.3 Momentum of Winning Team

One of the key trends which we intended to evaluate was whether or not sentiment of the tweets was representative of which team had the most momentum. As previously described, momentum in terms of this project, will be evaluated as the average of the number of first downs per possession. That is to say, we will be calculating the average number of first downs per possession and we will use this as our momentum value. Figure 13 illustrates the cumulative sentiment of a game between the BC Lions and the Edmonton Eskimos.

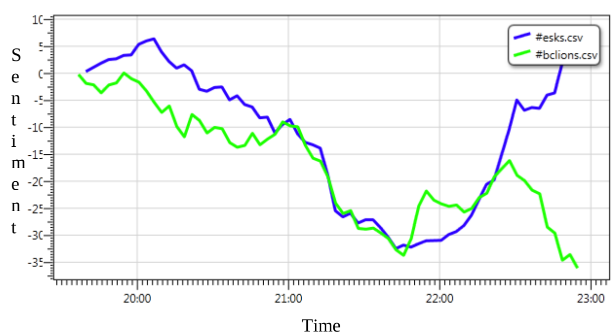

Figure 13: BC vs. Edmonton Cumulative Sentiment

If we look at the graph we can see that Edmonton had a slightly higher sentiment from the start of the game until about 21:00. Based on the data manually collected during the game Edmonton gained possession of the ball 4 times and had 4 first downs giving us a momentum of 1. BC had possession of the ball 4 times as well but had 5 first downs giving a momentum value of 1.25. Between 21:00 and 21:45 both teams had approximately the same sentiment. Edmonton had possession of the ball 3 times and got 4 first downs giving a momentum of 1.33. BC had possession 2 times and got 3 first downs giving a momentum of 1.5. Between 21:45 and 22:25 BC had a slightly higher sentiment. Edmonton had possession of the ball 2 times and got 5 first downs giving a momentum of 2.5. BC had possession 3 times and got 8 first downs giving a momentum of 2.66. Between 22:25 and the end of the game Edmonton had a significantly higher sentiment. Edmonton had possession of the ball 2 times and got 1 first down giving a momentum of 0.5. BC had possession 2 times and got 3 first downs giving a momentum of 1.5. The momentum numbers calculated suggest that in the first time frame BC had more momentum, in the second time frame BC was winning, in the third time frame Edmonton was winning and in the fourth time frame BC was winning. However, if we once again refer to figure 13 this contradicts the sentiment graph.

6. POTENTIAL IMPROVEMENTS

6.1 Momentum

There are many factors that influence the momentum of the game and not just how much time a team is in possession of the ball. Since points are only awarded for scoring in football it is possible for a team to have possession of the ball for a very long time and score no points or have possession of the ball for a very short period of time and score points. By improving our definition of momentum and analyzing more aspects of the game we may be able to effectively evaluate the momentum of each team.

6.2 Sentiment Analysis

If we evaluate the tweets at a per tweet level, we can identify two major issues with the sentiments which are returned. First, there are a significant number of "off topic" tweets included in the sentiment analysis. Second, in some cases the sentiment analysis for some tweets is not representative of the tweet's content.

6.2.1 Off Topic Tweets

Tweets that are completely unrelated to the football game could be filtered out before the sentiment analysis or even frequency analysis is performed. The issue of course is to find a way of filtering the tweets without removing tweets which are referring to the game. We already attempt to do this by utilizing hashtags and mentions in our search parameters but some erroneous tweets will always make their way through.

6.2.2 Inaccurate Sentiments

In some cases the sentiment calculated for tweets is not representative of the tweet's content. For example, the tweet "This BC Lions vs. Edmonton Eskimos game is hard hitting. #tsn #cfl", received a sentiment score of -0.58375. However, the tweet does not have a negative connotation. Or a more common case is where a tweet just receives a score of 0 when it should not. For example, the tweet "RT @RenoThreeZ: Khalf Mitchell is a scumbag. Payback is coming 2nd half. #esks #cfl", has a sentiment of 0 when it has a clearly negative connotation. This incorrect assignment of tweet sentiment is influenced by two factors.

6.2.2.1 SentiWordNet Database

The largest issue is the SentiWordNet database. While the database is quite large and contains a wide variety of terms and definitions found commonly in writing, the phrases and words associated with sports can have highly specialized meanings. As such, a word from the SentiWordNet database could have a different meaning from what was said in the tweet. A prime example is the phrase "hard hitting". In sports this phrase is almost always seen as positive, as it is what the fans want to see. However, the SentiWordNet database evaluates it as negative since in an everyday context no one really wants to be hit hard.

6.2.2.2 Sarcasm

Another example where the SentiWordNet database is insufficient for proper sentiment analysis is in regards to sarcasm. The database has no way of detecting sarcasm and will in most cases assign a sentiment value which is opposite of what it should be. The implementation of the sentiment analysis method also causes the sentiments values to be incorrect. Since the SentiWordNet database has multiple definitions for the same word the method has to decide what definitions to use. Unfortunately, implementation of a method which chooses the most appropriate definition to the context was beyond the scope of this project. Instead the sentiment analysis method averages the sentiment of all the different meanings leadings to an inaccurate sentiment, especially in cases where the word has many definitions.

6.3 Graph Legibility

Due to time constraints, there were several other modifications which could have been made but were not. These modifications would not only affect the quality of the data but the usability of the system. First, the bubble graph and bar graph functions, represented by figure 14, could be improved by adding separate colors for the different data sets. This would allow the user to differentiate between the datasets and compare the data. As it stands the program is unable to color the graphs separately for each set of data being analyzed and as such there is no way of telling what data is from which set.

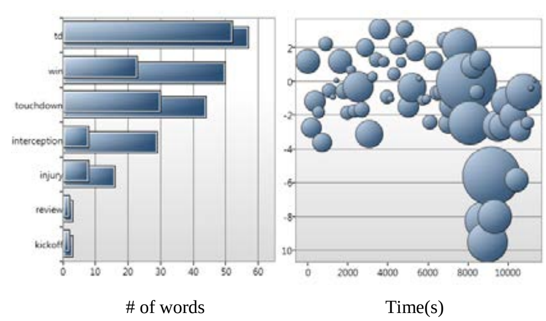

Figure 14: Bar and Bubble Graphs

6.4 Event Overlay

The next modification is for the event overlay option in both the frequency graph and sentiment graphs. In the current software this function draws a line graph with time on the x axis and the number of events that occurred in the time bucket on the y axis. The user can then hover over a point on the graph to see a list of the events which occur during this time period. This function could be improved by simply overlaying the list of events in text to the graph. This way the user does not have to hover over a point to see the events, also, this would also allow the user to see all the events which occurred at the same time. However, the library used to make this graph does not support the ability to draw words directly on the graph.

6.5 Significant Events

The significant event feature could be improved in several ways. This feature was intended to identify all of the frequency peaks which exceed a certain threshold. The x axis represents time, and both the y axis and the size of the bubble represent the number of tweets posted. The function also determines the most common word within each event and assigns it as the "event name". However, this event name is not displayed by the graph. In addition there is a logic error in the graph and it graphs the incorrect number of events. The cause of this bug is unknown and was not investigated due to time constraints. The feature could be greatly improved by displaying the event name on the bubble as well as assigning the sentiment of the event to the y axis which would indicate if the event was in favor or against the team as opposed to repeating information already available.

6.6 Data Granularity

The frequency bubble graph, keyword frequency, word frequency graph and significant event graphs all share a common weakness. These graphs can only display data from the entire game. A major improvement would be to implement a variable selection feature. This would allow the user to select a portion of data from the sentiment or frequency graph and only graph the frequency bubbles, words, keywords or significant events which occur during the selected time frame. With this the user would have much more control over the depth of the data.

7. CONCLUSION

This software is capable of revealing some very interesting data. However, the data is still in a very raw form. With further revisions the software will be able to show relations between twitter data and the events of the sporting event. For this to occur there are several issues which would need to be addressed. First, the definition of momentum we have used is far too simplistic to be able to draw conclusions about how the game progressed. In half of the data collected, the losing team was evaluated to have more momentum than the winning team. Second, as was demonstrated earlier, it is relatively easy to identify events on the frequency and sentiment graphs, however, without the event overlay it is nearly impossible to tell what kind of event occurred. Proper implementation of a system which filters the data could help to mitigate this issue.

If these issues can be resolved then there are many avenues of advancement for this software, ranging from marketing, to psychology. With internet related social networking becoming such an integral part of people's everyday lives there is a wealth of knowledge which can be collected by observing the publicly available information. The software which has been developed here is merely a stepping stone for future projects. There are countless applications for the use of this data. In regards to CFL football there are numerous marketing applications; observing the tweet frequency during a game could give insight to the marketing department towards better placement of advertisements and commercials. CFL teams can collect valuable data about fan sentiment. They could potentially tell whether fans are content with a team's performance despite a loss or be disappointed with a team even if they won the game. Instances of such events are likely caused by a very strong team playing against a weak team and losing. The data collected and its visualization could also lead to several other interesting statistics. With the addition of weather tracking one could determine how much of an impact the weather has on a team's performance and on the fan's sentiment. For instance do certain teams only perform well in optimal conditions or are there teams which can excel in poor conditions? Also do fans get upset more easily when the weather is poor? Other interesting stats include which type of event the fans prefer, how much of an impact previous events make on a fan's sentiment, whether people tweet more about positive events or negative events and whether or not a fan will tweet at all if his team is performing poorly.

8. REFERENCES

[1] Jeffrey Nichols, Jalal Mahmud, and Clemens Drews. 2012. *Summarizing sporting events using twitter.* In Proceedings of the 2012 ACM international conference on Intelligent User Interfaces (IUI '12). ACM, New York, NY, USA, 189-198. DOI=10.1145/2166966.2166999 http://doi.acm.org/10.1145/2166966.2166999

[2] John Hannon, Kevin McCarthy, James Lynch, and Barry Smyth. 2011. *Personalized and automatic social summarization of events in video.* In Proceedings of the 16th international conference on intelligent user interfaces (IUI '11). ACM, New York, NY, USA, 335-338. DOI=10.1145/1943403.1943459 http://doi.acm.org/10.1145/1943403.1943459

[3] Twitter Help Center: What Are Hashtags ("#" Symbols)? https://support.twitter.com/groups/31-twitter-basics/topics/109-tweets-messages/articles/49309-what-are-hashtags-symbols#

[4] Twitter Help Center: How to Retweet a Tweet https://support.twitter.com/articles/20169873-how-to-retweet-a-tweet#

[5] Twitter Help Center: What are @Replies and Mentions? https://support.twitter.com/groups/31-twitter-basics/topics/109-tweets-messages/articles/14023-what-are-replies-and-mentions#

[6] Adam Marcus, Michael S. Bernstein, Osama Badar, David R. Karger, Samuel Madden, and Robert C. Miller. 2011. *Twitinfo: aggregating and visualizing microblogs for event exploration*. In Proceedings of the 2011 annual conference on Human factors in computing systems (CHI '11). ACM, New York, NY, USA, 227-236. DOI=10.1145/1978942.1978975 http://doi.acm.org/10.1145/1978942.1978975

[7] Stefano Baccianella, Andrea Esuli and Fabrizio Sebastiani. *SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining.* In Proceedings of LREC-10, 7th Conference on Language Resources and Evaluation, Valletta, MT, 2010, pages 2200-2204.