GOAALLL!: Using Sentiment in the World Cup to Explore Theories of Emotion

Jonathan Gratch, [1] Gale Lucas, [1] Nikolaos Malandrakis, [2] Evan Szablowski, [3] Eli Fessler[4] and Jeffrey Nichols[5]

[1]University of Southern California Institute for Creative Technologies, Playa Vista, CA
[2]University of Southern California Signal Analysis and Interpretation Laboratory, Los Angeles, CA
[3]Oxford University, Oxford UK
[4]Pomona College, Claremont, CA
[4]Google Inc., San Jose, CA
gratch@ict.usc.edu

Abstract—Sporting events evoke strong emotions amongst fans and thus act as natural laboratories to explore emotions and how they unfold in the wild. Computational tools, such as sentiment analysis, provide new ways to examine such dynamic emotional processes. In this article we use sentiment analysis to examine tweets posted during 2014 World Cup. Such analysis gives insight into how people respond to highly emotional events, and how these emotions are shaped by contextual factors, such as prior expectations, and how these emotions change as events unfold over time. Here we report on some preliminary analysis of a World Cup twitter corpus using sentiment analysis techniques. We show these tools can give new insights into existing theories of what makes a sporting match exciting. This analysis seems to suggest that, contrary to assumptions in sports economics, excitement relates to expressions of negative emotion. We also discuss some challenges that such data present for existing sentiment analysis techniques and discuss future analysis.

Keywords—sentiment, sport, emotion models

I. Introduction

When Germany defeated Argentina in the final of the 2014 World Cup, the city of Berlin exploded in fireworks, chants and cheers. Halfway across the world, the streets of Buenos Aires were filled with tearful dejected crowds, but also wracked with angry riots and dozens of arrests. Worldwide, the game was watched with a mixture of emotions by about $1/6^{th}$ of the planet.

Clearly, sporting events command our interest and evoke strong emotions, yet do so in a structured and repeatable way that makes them of particular interest to social science research. Athletics has long served as a natural laboratory for investigating psychological and economic theory. This research, in turn, shapes the way sports are played as, for example, sports economists use findings to make games more exciting and addictive to their fans [1]. Sport involves high stakes for the participants and stimulate strong emotions and considerable economic investment on the part of spectators. Here, we illustrate ways that sentiment analysis techniques can expand our ability to exploit this natural laboratory.

This article briefly reviews the importance of sporting events as a natural laboratory to study human behavior, especially emotion. We next describe a corpus we collected from the 680 million tweets generated during the 2014 World Cup. Our work on analyzing this data is in its preliminary stages but we illustrate one example of its potential. We use sentiment analysis techniques over this dataset to examine a theory from sports economics concerning what makes a sporting event exciting. Specifically, the Uncertainty of Outcome Hypothesis (UOH) argues that close games are more exciting and therefore draw a larger audience [2]. The role of emotion is assumed by this theory but never tested. We show that sentiment analysis of game tweets can give interesting insights into what emotional processes are involved in people's decisions to watch an athletic match.

II. Background

A. Sport as an Emotion Laboratory

Sporting events have long served as a natural laboratory to study cognitive processes and emotion in particular. Because they follow uniform rules, are repeated many times, and elicit a variety of measurable (and faithfully recorded) events, sport provides a rich source of data against which to test psychological and economic theories.

One early example is the work of Tversky and Gilovich on the concept of momentum. It is generally accepted amongst players, coaches and fans that momentum is a major factor in determining outcomes in sporting events. Early success by one side within a game is presumed to create confidence, positive emotions and continuing success, whereas early failure creates negative emotions and failure. By analyzing basketball statistics, they were able to show this widespread belief is not borne out by the objective data [3], an effect now known as the hot hand fallacy." (See [4] for a review that emphasizes this fallacy holds across a wide range of sports).

Gilovich also used sporting events to lend support for prospect theory [5], one of the most important theories in behavioral economics. Prospect theory predicts that people's feelings towards an outcome depends on their reference point

Tweets during the 6/17 Russia vs Korea 2014 FIFA World Cup match

■ #WorldCup ■ #RUS ■ #KOR

Tweets during the 7/9 Netherlands vs Argentina 2014 FIFA World Cup match (logarithmic scale)
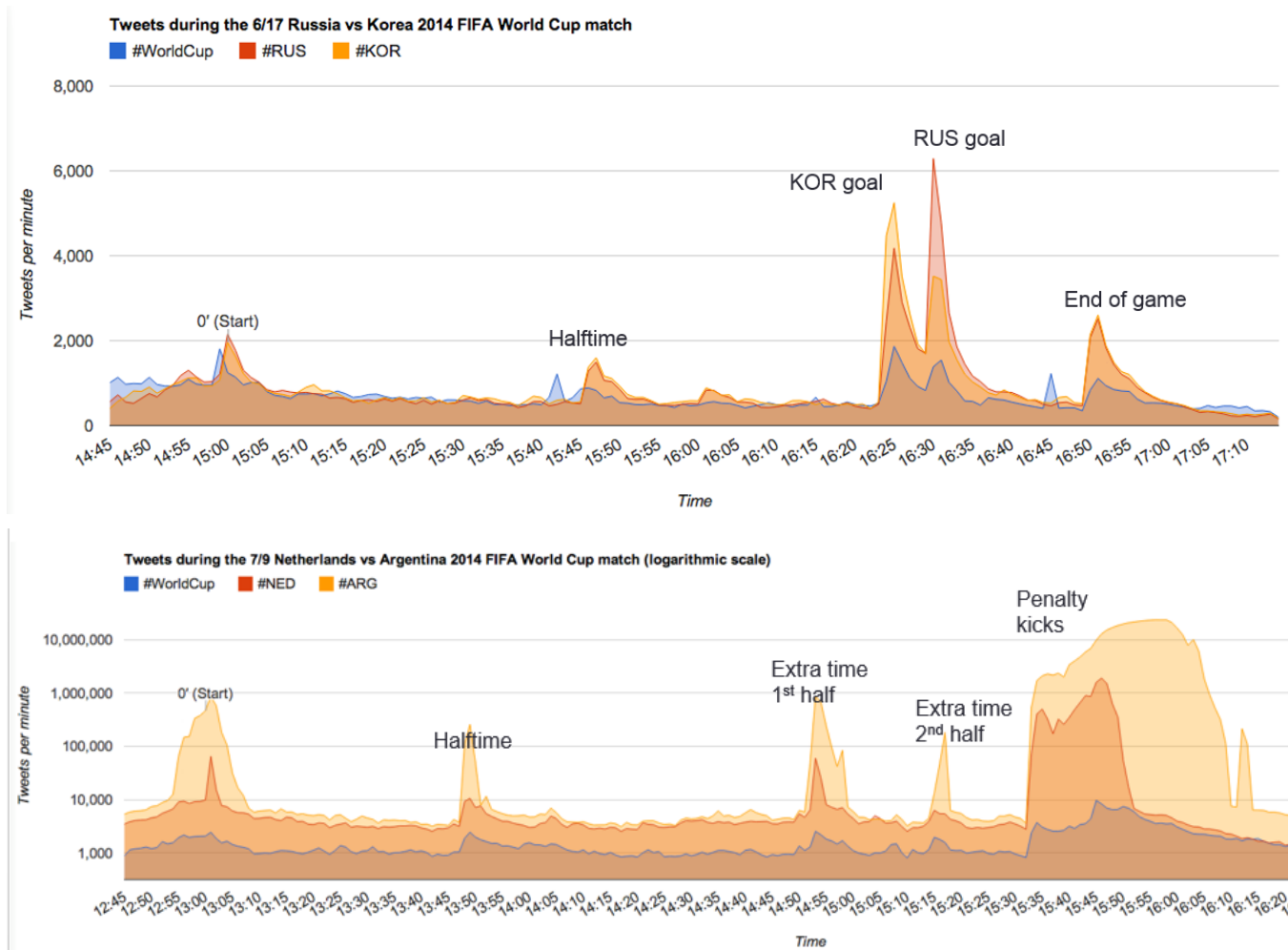
■ #WorldCup ■ #NED ■ #ARG

Figure 1: Graph summarizing tweets per minute during two games in the World Cup (annotated with game events).

(i.e., the same event may be seen either as a loss or a gain depending on one's point of comparison). Examining the emotional reactions of players during the 1992 Olympics, Gilovich finds support for this theory by showing that silver medalists felt worse (because they felt the loss of not winning) than bronze medalists (who felt relieved for making it to the podium) [6].

Ortony and Clore used the emotional reactions of basketball fans to provide support for their appraisal theory of emotion. Throughout an entire college basketball season they asked 106 fans to record their appraisals and emotions before, during, and after all 20 games in one year's season. The data provided clear support for their structural theory (sometimes referred to as the "OCC model" after the names of the authors) as well as their model of emotional intensity [7].

As a final example, Fernández-Dols examined facial expressions of gold medal winners in the Olympic games [8]. He was trying to distinguish theories that argue that emotional expressions reflect true feelings from competing theories that they are communicative acts. Based on a manual coding of facial displays he argues that winning athletes don't smile when they learn that they have won, but only upon turning

towards the audience. From this, they concluded that expressions serve as a social signal, rather than a marker of emotion.

Collectively, these studies emphasize the wide range of psychological questions that sporting events might address. Here we aim to extend this body of research with automatic sensing techniques.

B. Why Soccer?

In this article we focus on emotions expressed through Twitter by fans of World Cup soccer. We argue that soccer is especially relevant as a laboratory to examine emotional reactions. Soccer is the most popular sport in the world with an estimated 3.5 billion fans (sporteology.com). The World Cup was also the most important event in the history of social media [9]. Twitter reported 672 million tweets during the tournament, with 618,000 tweets per minute during the finale. Facebook recorded a record 280 million posts, comments and likes as Germany beat Argentina.

Soccer also has some unique advantages over other supports with regard to relating these posts to actual game

Positive
• Let's go #COL with such great talent they will give #CIV a challenge. This game is going to be amazing!!
• Best celebration so far. Go on Colombia! https://t.co/LfsqZTHqnr #COL #worldcup
• Goooooooaaaaalllll!!!!!!!! I can't tell you how happy I am atm #COL
• Fantastic counter attack from Columbia, unselfish by @TeoG29 #COL #IVC #WorldCup2014
• Excellent game by both teams..#COLvsCIV Classic match..

Neutral
• why isn't Victor Moses and Samuel Eto'o playing for Ivory Coast? #ivo #col
• Since when was James pronounced Ham-Es" #Col"
• LIVE: #COL begin the second half in Brazil's capital #joinin #worldcup #COLCIV
• #COL #COL #COL #COL #COL #COL #COL #COL #COL #COL #COL #COL #COL #COL #COL #COL
• GOAAAAAAAAAAAAAAAAAALLLLLLLLLLLLL QUINTEROOOO!!!!!!!!!!!!! COLOMBIA 2-0 IVORY COAST! #COL

Negative
• Ahhhh Gutierez! Bad connection on the ball #COL
• How have you missed that! #COL
• YES! HE F***D THAT UP #COL
• My mom says I'm a traitor for supporting Colombia :-/ #COL #HON
• hat was thee shot! Perfectly placed, and they threw it away. Smh! ? #col

Table 1: Examples of sentiment classifications

events. Unlike many sports, soccer does not allow commercial breaks during play. This makes it especially easy to relate specific posts to moment-to-moment game events. Indeed, work by Jeffrey Nichols has shown that machines can automatically recover and summarize the sequence of events in a soccer game by analyzing simply the volume of tweets per minute [10]. This is illustrated vividly in Figure 1, which shows tweets per minute during two matches, one between Russia and South Korea, and the other between the Netherlands and Argentina. Both are annotated with the time of key game events. As can be seen in the first game, the only two goals in the game generated considerable excitement, as measured by the number of tweets per minute.

III.   2014 WORLD CUP TWITTER CORPUS

The 2014 World Cup was played from June 12th through July 13th 2014 and consisted of 64 games (48 first-stage "group play" games and 16 second-stage "knockout" games). The character of games differed by stage. Ties are allowed in the first stage but second round games must result in a winner. Thus second-round games were potentially longer and optionally included two overtime periods and possibly penalty kicks. We attempted to record data for all 64 games but for technical reasons only 60 are included in the final corpus. To capture tweets, we utilized a java-based recorder employing Twitter Streaming API and Twitter4J Java library. Out of these games, our Twitter recorder failed to record properly for four games, reducing our final sample of games to 60. For these 60 games, we recorded tweets starting 15 minutes prior and continuing during the entire game, but only utilize tweets during game play for analyses reported below (and therefore exclude tweets recorded during halftime, for example).

During each game, we collected tweets that contained at least one of the official hashtags (released by FIFA) for the two teams currently playing. We had two recorders running concurrently for each game, one capturing tweets for each team. We had access to Twitter's "gardenhose" (public stream of data), which allows for collecting up to 1% of all world-wide tweets at any moment. For most games, this allowed us to capture all tweets for a given team. However, during more popular games, this restriction limited the number of tweets we could capture. For sentiment analysis described below, we were limited to this restricted (albeit still substantial) dataset because the content of tweets was only available for those that we collected using our recorder. However, for analyses of game's average tweet-per-minute (as a proxy for attendance by the Twitter audience), because twitter provided the rate at which it limited the data, we were able to calculate the number of tweets that would have been collected without limitation. To more accurately represent the games, we use this adjusted number for average tweet-per-minute of games in which rate-limiting occurred.

In addition to Twitter data on each match, the corpus contains other important information concerning pre-game expectations and how the game unfolded. For each game, we recorded the Vegas betting odds (as reported at OddsPortal.com) as a measure of which team was expected to win. As a measure within-game events, we accessed game statistics through ESPN.com's GameCast which provided a timeline of all shots, goals, penalties and substitutions.

IV.   SENTIMENT ANALYSIS

To extract sentiment from recorded tweets we used an "off the shelf" sentiment analysis technique without adapting it to the idiosyncrasies of soccer tweets.

A. Sentiment anaysis method

The sentiment analysis method utilized is a variation of the one presented in [11] as a submission to the SemEval 2014 twitter sentiment analysis challenge [12]. It is a supervised method, mainly based on lexicon derived features that describe the distribution of word emotional content through a multitude of statistics.

Sentiment is only extracted for English tweets. Language classification was performed on each tweet using the implementation of [13] included in NLTK [14]. Tweets identified as English were then part-of-speech (POS) tagged using the twitter specific ARK NLP tagger [15]. Lexicon based features were finally extracted by replacing words by the corresponding word-level emotional ratings, applying POS filters, and calculating statistics. The word-level ratings were provided by multiple lexica, some containing ratings for more than one affective dimension, combining for a total of up to 17 ratings per word, with each dimension used separately. Some novel lexica were generated according to the method described in [16]. Words not included in the lexica were ignored. The statistics used include the mean, minimum, maximum and variance and were calculated after applying some POS tag filtering, keeping only words with specific tags. Some examples of resulting features are "average of SentiWordNet positive over singular nouns" and "maximum of EmotiWord arousal over past participles". Beyond lexicon based features, the method used simple frequencies of capitalization, punctuation, emoticons and elongation (character repetition). Feature selection [17] was used to reduce the very large set of candidate features and the resulting set was used in training a Naive Bayes tree, a tree with Naive Bayes classifiers on each leaf. Both feature selection and training were performed using the annotated data of the SemEval 2014 challenge. The output is a tertiary rating for each tweet: negative, positive or neutral.

While this method is not identical to the one used in [11], it is very similar. That method performed very well in the SemEval challenge, reaching ranks of 7th and 2nd (out of 42 participants) in the two sub-tasks of tweet classification [12] and exhibiting a notable robustness with respect to sarcastic tweets. The overall model has proven versatile and robust, mainly due to the underlying lexicon expansion method. Prior versions of the method as well as individual components have been used in the past to achieve state-of-the-art results in a variety of tasks related to affective state: assessing the emotional content of words and news headlines [16] and evaluating psychological experiment transcripts [18] in English and diagnosing depression [19] in German. It is this robustness that makes the method particularly suitable to the task at hand.

Table 1 illustrates some examples of tweets classified by this method from a stage-one game between Columbia and Ivory Coast.

B. Measurement Check

Although we selected a method that performed well on the SemEval 2014 twitter sentiment analysis challenge, it is possible that the technique performs poorly on soccer tweets. We performed a preliminary evaluation to assess the effectiveness of this technique on our corpus. We randomly selected 154 tweets that were classified as English by the method and manually coded them as positive, neutral or negative. There was a strong and significant correlation between coders rating (-1, 0, 1) and sentiment analysis (-1, 0, 1): $r(152)=.53$, $p < .001$. A chi square test, treating these as categories, is also significant, $\chi^2 (4) = 62.92$, $p < .001$ (see Table 2). Separate chi squared test for negative vs neutral ($\chi^2 (4) = 12.72$, $p < .001$) and positive vs neutral ($\chi^2 (4) = 37.50$, $p < .001$) are also significant.

		Coder			
		negative	neutral	positive	Total
SA	negative	3	5	0	8
	neutral	4	96	28	128
	positive	0	1	17	18
Total		7	102	45	154

Table 2: Confusion matrix of automatic and manual codes

These results suggest that the automatic annotations can serve proxy for manual annotations. It should also be noted that there is a tendency for the technique to misclassify certain positive tweets as neutral, probably due to idiosyncratic terms used in soccer (as can be seen in Table 1). This should be kept in mind when interpreting the results.

V. A TEST OF EMOTION THEORY

To assess the potential of sentiment analysis to answer psychological questions, we present an example of how Twitter data about sporting events can be used as laboratories to explore emotion and how computational tools provide new ways to examine emotional processes "in the wild". Specifically, we use the corpus of twitter sentiment to provide a novel perspective on an important theory in sports economics.

A. A Novel Approach to the Uncertainty of Outcome Hypothesis

Across sports economics, emotional processes are assumed but most remain untested. For example, according to the well-studied uncertainty of outcome hypothesis (UOH), "close" games are more exciting and therefore better attended [2]. If one team were certain to win, it would take away a major source of excitement, reducing positive affect, and therefore decreasing attendance. The role of emotion here is assumed but has not been tested; furthermore, the

measures used (ticket sales, attendance, TV-viewership) do not allow for such a test because they are devoid of emotional content. To address this problem, we use tweets per minute (specifically, tweets posted during 2014 World Cup with official game hashtags). Sentiment analysis of these tweets can give interesting insights into what emotional processes are involved.

Another benefit of tweets is that they are dynamic, and novel results from dynamic analyses (of TV-viewership) suggest that the UOH effect can actually *reverse* as games unfold (people switch channels away from close games). Alavy, Gaskel, Leach, and Szymanski [20] used TV-viewership of football games to test the *"uncertainty of outcome" hypothesis*. While much of the previous research measuring attendance supports this hypothesis, because TV-viewership is a more dynamic measure of excitement, they could find that the effect reversed as the game wore on: people switched channels away from close games [20]. This, and other mixed evidence for the UOH [21], suggests that further testing is needed, especially using real-world data and novel approaches to explain these mixed findings.

We therefore also reconsider the UOH, specifically, extending it by both examining sentiment and dynamic changes during the game. Towards this end, we use the 2014 World Cup corpus to examine these questions.

To consider the UOH in a dynamic way (like [20]), we consider games that could have been close (high in uncertainty), but ended up being lower in uncertainty. We operationalize such *unexpected* certainty of outcome by taking the absolute value of the difference between two differences: the difference between the predicted scores for the two teams in a given game based on Vegas betting odds and the difference between *actual* final scores for the two teams. This resulted in a metric indexing as the extent to which games were predicted to be "close" (based on betting odds), but ended up with a bigger difference between the teams' scores than was expected. We use tweets per minute as an index of game excitement.

B. Results

First, replicating and extending previous research Alavy et al., [20], we tested whether games that could have been close (high in uncertainty), but ended up being lower in uncertainty were more or less attended to by the Twitter audience. Although the UOH would suggest that these games that became blowouts (as one team pulled farther ahead than expected) would become boring, and therefore lose attendance, Alavy et al. find that more dynamic analyses reveal different patterns.

To consider this possibility, we tested the correlation between game's average tweet-per-minute (as a measure of attendance by the Twitter audience) and the extent to which games was predicted to be "close" (based on betting odds), but ended up with a bigger difference between the teams' scores than was expected (its *unexpected* certainty of outcome). This analysis revealed that, contrary to the UOH,

games with a bigger difference in score between teams than expected had *higher* tweets per minute, $r(58) = .54$, $p < .001$.

As described above, we also performed sentiment analysis, categorizing each tweet as positive, negative or neutral. Specifically, we examined the proportion of tweets in a given game that were positive, negative or neutral (which summed to one hundred percent for each game). We considered the relationship between the proportion of positive, negative or neutral tweets and our other variables of interest: volume (tweets per minute) and *unexpected* certainty of outcome (the extent to which games are predicted to be "close" (based on betting odds), but ended up with a bigger difference between the teams' scores than was expected).

Analyses revealed that games with higher tweets per minute also have a higher percentage of negative tweets, $r(58) = .26$, $p < .05$. However, percentages of positive and neutral tweets were not significantly related to tweets per minute, rs $< -.12$, ps $> .36$. Furthermore, games that have a bigger difference than expected also have a higher percentage of negative tweets (compared to games closer to what is expected), $r(58) = .26$, $p < .05$. However, again, percentages of positive and neutral tweets were not significantly related, rs $< -.20$, ps $> .12$. These analyses suggest that, contrary to assumptions in sports economics, excitement relates to expressions of negative emotion (and not positive emotion).

VI. DISCUSSION

Sporting events can serve as natural laboratories for understanding how people emotionally respond to situations. Sport often involves high stakes and certainly evokes strong emotions in both participants and observers. Sport has the advantage over other natural situations in that it regularly repeats with variation, but where this variation is constrained by the rules of the game.

Here we have explored one aspect of sporting events – sentiment analysis of fan tweets – and shown the potential of this data for addressing important questions about human emotion, but this research has just scratched the surface of the potential of affective computing techniques to understand and model emotions in such situations. Future work will seek to improve the accuracy of the sentiment classification and consider if any biases in the method influence our interpretation. Many additional analyses could be considered. For example, how do emotions change as a result of the "trajectory" of a game? Is a game that starts as a blowout but ends close more exciting than one that stays close throughout? What can we learn from the social network or activity across multiple games? For example, one could identify fans of particular teams by looking if the same people tweeting during two games of the same team vs. different opponents. Using geographic information, one could compare responses to fans within or outside the stadium.

Of course, twitter is simply one view into the emotions of sports fans. Social signal processing techniques could be used to extract other information. For example, it is possible to measure excitement from the volume of crowd noise within a stadium or movement within a video sequence (see [22]).

Here we focused on using sentiment analysis of sport to test theories of human emotion but such data has other practical applications, such as automatically creating a summary of a game [10] or identifying specific events [23]. More broadly, the intersection of affective computing and sport analytics is a promising direction for future research.

REFERENCES

1. Szymanski, S., *The Economic Design of Sporting Contests*. Journal of Economic Literature, 2003. **41**(4): p. 1137-1187.
2. Rottenberg, S., *The Baseball Players' Labor Market*. Journal of Political Economy, 1956. **64**(3): p. 242-258.
3. Gilovich, T., R. Vallone, and A. Tversky, *The hot hand in basketball: On the misperception of random sequences*. Cognitive Psychology, 1985. **17**(3): p. 295-314.
4. Bar-Eli, M., S. Avugos, and M. Raab, *Twenty years of "hot hand" research: Review and critique*. Psychology of Sport and Exercise, 2006. **7**(6): p. 525-553.
5. Kahneman, D. and A. Tversky, *Prospect Theory: An Analysis of Decision under Risk*. Econometrica, 1979. **XLVII**: p. 263-291.
6. Medvec, V.H., S.F. Madey, and T. Gilovich, *When less is more: counterfactual thinking and satisfaction among Olympic medalists*. Journal of Personality and Social Psychology, 1995. **69**(4): p. 603-610.
7. Clore, G.L., A. Ortony, and S. Brand, *The joy of victory and the agnony of defeat: the emotions of sports fans*. 2007, University of Illinois.
8. Fernández-Dols, J.M. and M.-A. Ruiz-Belda, *Are Smiles a Sign of Happiness? Gold Medal Winners at the Olympic Games*. Journal of Personality and Social Psychology, 1995. **69**(6): p. 1113-1119.
9. Dredge, S. *World cup was biggest event yet for Twitter with 672m tweets*. the Guardian, 2014.
10. Nichols, J., J. Mahmud, and C. Drews, *Summarizing sporting events using twitter*, in *Proceedings of the 2012 ACM international conference on Intelligent User Interfaces*. 2012, ACM: Lisbon, Portugal. p. 189-198.
11. Malandrakis, N., et al., *SAIL: Sentiment Analysis using Semantic Similarity and Contrast Features*, in *proc. SemEval*. 2014. p. 512-516.
12. Rosenthal, S., et al., *SemEval-2014 Task 9: Sentiment Analysis in Twitter*, in *Proc. SemEval*. 2014.
13. Artemenko, O., et al., *Evaluation of a Language Identification System for Mono- and Multilingual Text Documents*, in *Proceedings of the 2006 ACM Symposium on Applied Computing*. 2006. p. 859-860.
14. Bird, S., E. Klein, and E. Loper, *Natural Language Processing with Python*. 2009: O'Reilly Media, Inc.
15. Owoputi, O., et al., *Improved Part-of-Speech Tagging for Online Conversational Text with Word Clusters*, in *proc. NAACL*. 2013. p. 380-390.
16. Malandrakis, N., et al., *Distributional Semantic Models for Affective Text Analysis*. Audio, Speech, and Language Processing, IEEE Transactions on, 2013. **21**(11): p. 2379-2392.
17. Hall, M.A., *Correlation-based feature selection for machine learning*. 1999, The University of Waikato.
18. Malandrakis, N., et al., *Affective language model adaptation via corpus selection*, in *proc. ICASSP*. 2014. p. 4838-4842.
19. Gupta, R., et al., *Multimodal Prediction of Affective Dimensions and Depression in Human-Computer Interactions*, in *Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge*. 2014. p. 33-40.
20. Alavy, K., et al., *On the Edge of Your Seat: Demand for Football on Television and the Uncertainty of Outcome Hypothesis*. International Journal of Sport Finance, 2010. **5**(2): p. 75-95.
21. Pawlowski, T., *Testing the Uncertainty of Outcome Hypothesis in European Professional Football: A Stated Preference Approach*. Journal of Sports Economics, 2013.
22. Hanjalic, A., *Adaptive extraction of highlights from a sport video based on excitement modeling*. Multimedia, IEEE Transactions on, 2005. **7**(6): p. 1114-1122.
23. Zhao, S., et al., *Human as real-time sensors of social and physical events: A case study of twitter and sports games*. arXiv preprint arXiv:1106.4300, 2011.