Using Financial Analysis Techniques on Twitter Sentiment to Improve NFL Predictions

Robert P. Schumaker[1], Leonard L. Brown, III[1], Chester S. Labedz Jr.[2], and A. Tomasz Jarmoszko[2]

[1]Computer Science Dept., University of Texas at Tyler, Tyler, Texas 75799, USA
[2]Central Connecticut State University, New Britain, Connecticut 06050, USA

rob.schumaker@gmail.com, lbrown@uttyler.edu, clabedz@gmail.com and jarmoszko@ccsu.edu

Abstract
We examine the application of financial analysis techniques to sentiment gathered from social media to predict NFL game outcomes. From our analysis we found a $14.84 average return per sentiment-based wager versus an odds-only approach of $12.21 average loss on the entire 256 games of the 2015-2016 NFL regular season. We further noted that wagers on underdogs (i.e., the less favored teams) that exhibited a more positive pattern in sentiment, netted a $35.28 return per wager on 117 wagers. These results show promise of cross-domain research, and we believe that applying stock market techniques to sports wagering may open an entire new research area.

Keywords: Business Intelligence, Decision Support, Sentiment Analysis, Sports Analytics

1. Introduction

Worldwide sports gaming market is worth an estimated $1 trillion dollars and is growing (Statista.com, 2016). The primary challenge of sports gaming is to predict the outcome of future events, so tools and techniques used successfully in other prediction-based applications, such as the stock market and election-forecasting, may potentially be useful in solving this challenge. However, these techniques need to be modified in order to suit the unique characteristics of sports gaming.

To illustrate the above issue, consider the field of sentiment-analysis which has been applied successfully to social media for forecasting elections (Tumasjan, Sprenger et al., 2010) and to news articles for forecasting stock market changes (Schumaker, Zhang et al., 2012). Applying a similar technique to forecasting NFL games means analyzing a fan base's expectations of their team's performance from their collective writings (in our case, tweets). The results are aggregated to form a crowdsourced signal that can be used to forecast the winning team.

The problem is that fan-bases of different teams may have very different ways of expressing sentiment. Some may tend to be eternally optimistic while others tend to become extremely pessimistic at the first sign of trouble. This type of regional fan behavior makes the comparison of sentiment between geographically different fan bases a non-trivial problem. Thus, making comparisons between teams in an absolute sense, without consideration of their fan base differences, may lead to less successful predictions.

The purpose of this paper is to demonstrate the effectiveness of utilizing techniques from stock price analysis to sports gaming. In particular, we analyze sentiment polarity as a time-series signal and examine the position and magnitude of signal change between two temporal windows. By analyzing the aggregated differences of signal between two moving averages and comparing it between teams, one may avoid problems associated with forecasting accurately despite the noise of NFL fan base differences.

The rest of this paper is organized as follows. Section 2 reviews the application of sentiment analysis, wagering and crowdsourcing, and technical charting to sports gaming. Section 3 contains our research questions. Section 4 introduces the CentralSport system, a fusion of sentiment analysis and technical stock market charting. Section 5 presents our experimental design. Section 6 contains the experimental results and a discussion of their implications. Section 7 concludes with a summary of findings and future research directions.

2. Literature Review

In this section, we focus on three key areas related to our proposed approach. The related work relies on using sentiment analysis for sports, using crowdsourcing for wagering, and using technical charting for forecasting.

2.1 Sentiment Analysis

Sentiment analysis is the science of identifying opinions and emotions in text (Abbasi, Chen et al., 2008), which can include author tone, whether a text is objective or subjective, and author polarity, whether the text is positive or negative (Wilson, Hoffmann et al., 2005). These features are embedded within a text and are a function of author word choice, which can often provide insight into an author's emotional state.

Tracking sentiment through social media has been shown useful to predict elections (Tumasjan, Sprenger et al., 2010), stock markets (Schumaker, Zhang et al., 2012) and sports. The forecasting of sport has an avid following, ranging from fantasy team owners to professional gamblers and even academics. In a study of NFL game prediction, Sinha et. al. used Twitter features, such as changes in tweet volume, to build a model capable of 55% accuracy, exceeding the breakeven point of a bookmaker's commission (Sinha, Dyer et al., 2013). In a study of sentiment on the prediction of English Premier League soccer results, Schumaker et. al. examined aggregated tweet sentiment in the 96 hours prior to match start between the two clubs. They found that using sentiment netted a higher payout, $2,704.63, compared with the return loss ($1,887.88) of an odds-only approach (Schumaker, Jarmoszko et al., 2016). They also observed that contrasting a club's weekly tweet difference with its seasonal average netted an even greater payout, $3,011.20. This examination of volume change versus club average was a step towards minimizing fanbase differences and shows the promise of further research in this direction.

2.2 Wagering and Crowdsourcing

Crowdsourcing offers one way to generate accurate and reliable forecasts that can be used for wagering. This technique uses an average of individual forecasts to predict future events (Surowiecki, 2004). In a study of the 2006 FIFA World Cup soccer matches, crowds were found better able to predict winners than were comparative FIFA national team rankings or random chance (Luckner, Schroder et al., 2008). In a study of German Premier League soccer, the wisdom of crowds was found to be more accurate than bookies (Spann and Skiera, 2008). In a study of English Premier League soccer, social media sentiment (a type of crowdsourced data) was similarly accurate (Schumaker, Jarmoszko et al., 2016).

2.3 Technical Charting

Stock prices can be thought of as time-series data that is difficult to discern patterns from in the near-term, but easier to forecast in the long-term. This signal will adjust with the introduction of new information, regardless of whether that information is public or private. Technical analysis strategies examine price and/or volume signals and compare that data to historical values in an attempt to identify known activity patterns. These strategies, because of their relatively short duration, are a favorite of market daytraders.

One such strategy uses moving averages, which helps to lessen daily price volatility by providing a moving window of averaged historical prices which functions as a stabilizing weight when trying to discern a trend. Typically, two or more moving averages are computed and the interaction between them becomes of interest (Blair, 1996). When these time-series signals cross one another, it typifies a buy/sell recommendation. Several studies in the finance literature have used moving averages. In a study that used twenty-one years worth of data from the Singapore Stock Exchange, researchers concluded that investment decisions based on moving averages could generate a significant positive financial return (Wong, Manzur et al., 2003). Another study examined a hundred years of Dow Jones Industrial Average data and found support for technical analysis techniques (Brock, Lakonishok et al., 1992). A third study used technical analysis techniques coupled with portfolio building and found that a one-week formation period using a momentum strategy netted a 20.79% trading return over a five-week holding period (Schumaker and Chen, 2008).

2.4 Research Gaps

Our review of the literature identified several opportunities not previously pursued, notably very few sentiment studies on sports wagering. We seek to extend the existing research by investigating polarity measures from a technical charting standpoint and determine if the application of finance techniques can be successfully applied to improve betting returns.

Another gap was that prior studies treated the sentiment of each team equally, ignoring fan-base differences. Our work not only takes these differences into account but also allows an inspection of sudden, recent changes in sentiment. Our intent is to derive a profitable system based on crowdsourced information that is not typically used in a wagering market.

3. Research Questions

To address the research gaps and further explore this area, we consider these questions.

1. *Can the technical charting of fan sentiment predict game outcomes?*

From prior studies, counts and contents of tweets and blogs appear to function as adequate proxies for prediction. Those studies all rely on one set of time-series data which ignores fan-base differences and in turn obscure the data signal. By teasing out a baseline of fan sentiment and only tracking the changes, we suspect to find a more predictable model.

2. *Can fan sentiment be profitable?*

Prior studies in other sports domains have shown an inverse relationship between accuracy and profitability. Does the same hold true in the NFL and if so, can fan sentiment be used for profitable wagering?

4. System Design

Our approach is to test modified stock price charting techniques used by Wall Street to represent sentiment signals relative to regional affect. In technical charting two moving averages are typically utilized. The idea is that the longer moving average provides a type of baseline behavior and the shorter moving average represents recent activity but is long enough to smooth out any short-term volatility. Applied to our study of NFL Twitter sentiment, we chose 96 hour and 24 hour periods prior to game start.
Our system is illustrated in Figure 1 with an explanation of critical components to follow.

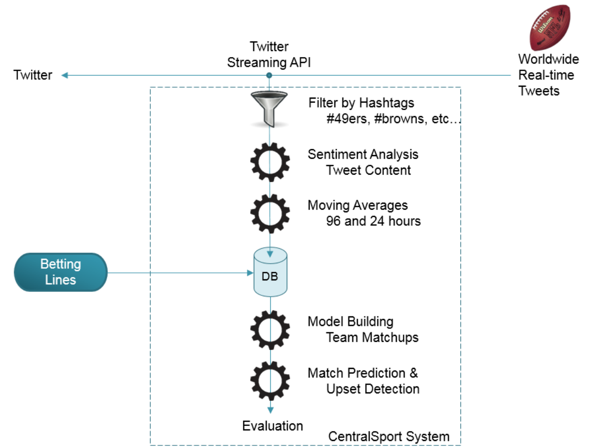

Figure 1. The CentralSport System

Tweets are gathered in real-time from the Twitter Streaming API and filtered by hashtag for each of the 32 NFL teams. To limit the effect any one tweeter could have on the results, we only analyzed one tweet per tweeter in each of the moving average windows; that is, if a tweet author created more than one message we only used the first one. While we recognize that this could be a study limitation, we sought to prevent a minority of prolific tweeters from skewing the results. The content of each tweet is then passed through OpinionFinder (OpinionFinder.com, 2016) to determine tweet polarity: positive, negative or neutral.
Four models are then built to test regional sentiment values. The first is a baseline odds model that wagers on favorites using data from OddsPortal.com, a compilation of several wagering firms that are averaged using the Moneyline approach. The second model uses the sentiment gathered 96 hours prior to kickoff. The team with the higher normalized positive to negative sentiment ratio is predicted to win. The third model is a variation of the second and uses the sentiment gathered 24 hours prior to kickoff. Both the 96 hour and 24 hour models ignore fan-base differences. The fourth model uses technical charting and takes into account the change in sentiment during these periods and calculates the sentiment surge or drop (we refer to it as swing) for each team prior to game start.
Swing values for each team are then compared, and the team with the higher swing value is predicted to win. We then evaluate each model's predictions against actual outcomes and calculate accuracy and

payout values. Payout uses a simulated trading engine that wagers $100 on each prediction and then calculates the theoretical winnings (or losses) using Moneyline odds.

An example of the system using the Steelers and the Patriots from Week 1 of the 2015-2016 NFL season is shown in Table 1. In Table 2, all four models predicted the Patriots to win. Baseline, using the odds-only approach from OddsPortal.com, chose the lower odds Patriots (-345 to +272). The three sentiment models also picked the Patriots because of the higher positive sentiment. Focusing on the swing model, the Steelers' swing value is 4.99% indicating that, after normalizing the data, the 24 hour sentiment was 4.99% more positive than the 96 hour sentiment. Fans were expressing more optimism in the hours leading up to kickoff. Additionally, the Patriots' swing value was 21.34% indicating greater twitter enthusiasm. All four models correctly picked the Patriots to win, which they did by a score of 28-21.

	96 hour Sentiment			24 hour Sentiment		
	Positive	Negative	Neutral	Positive	Negative	Neutral
Steelers	6,400	3,696	2,893	5,214	2,575	2,436
Patriots	31,510	7,126	2,178	21,088	1,265	2,094

Table 1. Sentiment values for Steelers vs Patriots 9/10/15

		Sentiment Models			
	Baseline	96hr	24hr	Swing	Score
Steelers	272	20.82%	25.81%	4.99%	21
Patriots	**-345**	**59.74%**	**81.08%**	**21.34%**	**28**

Table 2. Prediction Results of Steelers vs Patriots 9/10/15

5. Experimental Design

For the experiment we chose the 2015-2016 NFL regular season as our testbed. This dataset encompassed 17 weeks and a total of 256 games. During this period we collected tweets in real-time using one hashtag identified by a domain expert for each of the 32 teams using Twitter's streaming API. Following prior work we used the 96 hours prior to each game's kickoff for our study, acknowledging that some teams had Sunday followed by Thursday games which may cause a slightly more impure dataset than we would have liked. From this process, 4,509,260 tweets were used in the study. The New England Patriots had the most tweets of any team with (315,525) and the New York Giants had the fewest at 28,238 because the hashtag used, #nygiants, was selected to not interfere with the San Francisco Giants baseball team. The average team number of tweets was 140,914 with a standard deviation of 69,331.

6. Experimental Results and Discussion

In this section, we look at the results of our tests on each research question.

6.1 Technical Charting of Fan Sentiment and Accuracy

To answer our first research question of *can the technical charting of fan sentiment predict game outcomes,* we made accuracy predictions on all 256 NFL games during the 2015-2016 regular season, comparing the four models against the actual outcome. The results are presented in Table 3.

	Baseline	96hr	24hr	Swing
# Correct	153	128	137	150
Average	59.8%	50.0%	53.5%	58.6%
p-value		0.013	0.077	0.394

Table 3. Accuracy Predictions of the Models

From this table, using the odds-only approach, Baseline, achieved an accuracy of 59.8% by correctly predicting 153 of 256 games. The model using the Twitter sentiment 96 hours prior to kickoff was accurate 128 times for 50.0% accuracy. The 24 hour model was slightly more accurate at 53.5%, correct on 137 games. Swing, the fusing of the two prior models into a technical charting technique, was accurate 150 times or 58.6%. From these values, Baseline and Swing were statistically equivalent, p-value = 0.394. Comparing Baseline accuracy to the 96 hour and 24 hour models independently, the sentiment models alone were found to perform statistically worse. This shows that using absolute sentiment by itself in either a 96hr or 24hr timeframe does not have better accuracy than odds, which is consistent with the accuracy results of absolute sentiment comparisons in the English Premier League study.

Diving deeper into the results and looking at predictions on favorites versus upsets, the Swing model selected favorites 54.3% of the time on 139 games, and upsets 45.7%, as shown in Table 4. It was interesting to note that all models were more correct versus incorrect when predicting favorites to win and more incorrect than correct in predicting upsets.

	Favorites		Upsets	
	Correct	Incorrect	Correct	Incorrect
Baseline	153	103		
96hr	80	55	48	54
24hr	83	49	54	70
Swing	93	46	57	60

Table 4. Model Accuracy on Favorites and Upsets

From this data, when Swing wagered on the odds favorite team, the accuracy was 66.9% (93 to 46) versus wagering on upsets with 48.7% accuracy. Comparing Baseline to Swing on Favorites Accuracy, 59.8% to 66.9% respectively, we noted a p-value of 0.05 indicating that Swing had a statistically significant improvement in accuracy. However, each absolute sentiment model (96hr and 24hr) for Favorites, yielded 59.3% and 62.9% accuracy respectively. Neither of which was found to be statistically different from Baseline accuracy.
It would appear that a wagering strategy of selecting those teams that are both odds favorites and have a more positive Swing value can yield better accuracy than odds alone. From the data, neither the 96 hour nor 24 hour sentiment models had statistically different results than the odds-only Baseline approach on favorites. This would indicate that the technical charting method of the Swing model is adding value. This finding can have implications on betting houses by exposing a technique for improving wagering odds, and bettors for improving their wagering accuracy.

6.2 Technical Charting of Fan Sentiment and Profitability

To answer our second research question of *can fan sentiment be profitable* we made hypothetical $100 wagers on matches and analyzed their returns for each model, as shown in Table 5.

	Baseline	96hr	24hr	Swing
Sum	($3,125.79)	($545.10)	$1,237.91	$3,798.15
Average	(12.21)	(2.13)	4.84	14.84
p-value		0.115	0.022	0.0009

Table 5. Payout Predictions of the Models

From this table, the odds-only Baseline had a return of -$3,125.79 for an average loss of $12.21 per wager. Given the 59.8% accuracy from earlier coupled with the lower odds for favorites, the accuracy was not sufficient to offset losses which amounted to $10,300 versus $7,174.21 in winnings. For the absolute sentiment models, the 96 hour model had a payout loss of $545.10 and the 24 hour model demonstrated a payout gain of $1,237.91. These two models coupled with their lower accuracies from earlier indicate that crowdsourced sentiment was identifying longshot opportunities, which is in line with prior studies. The technical charting model Swing had a payout gain of $3,798.15 or an average of $14.84 profit per wager. Swing was found statistically equivalent in terms of accuracy, however, the total payouts from longshot predictions is much improved (p-value = 0.0009).
Looking further into longshot predictions, we examine the payout differences between wagering favorites and upsets as shown in Table 6.

	Favorites		Upsets	
	Correct	Incorrect	Correct	Incorrect
Baseline	$7,174.21	($10,300.00)		
96hr	$3,894.89	($5,500.00)	$8,360.01	($5,400.00)
24hr	$3,950.32	($4,900.00)	$9,187.59	($7,000.00)
Swing	$4,270.32	($4,600.00)	$10,127.83	($6,000.00)

Table 6. Model Payouts on Favorites and Upsets

From this data, none of the models wagering on favorites posted gains which was not surprising given the inherently lower odds. Despite accuracies of 59.8%, 59.3%, 62.9% and 66.9% for each of the four models respectively, none managed a positive payout. However, turning attention towards predicting upsets, the three sentiment models posted gains of $2,960.01, $2,187.59 and $4,127.83 respectively on 102, 124 and 117 wagers. This translates into an average return per wager of

$28.73, $17.64 and $35.28 per model respectively on upsets, despite accuracies of 47.1%, 43.5% and 48.7%. We acknowledge that these wagers are hypothetical and that real wagers might influence the betting line, albeit marginally, for slightly lower returns.

These results provide strong evidence that technical charting of social media sentiment can be used profitably in sports gaming. We believe that the longer 96 hour window provides a sort of baseline of sentiments that accounts for fanbase differences when measured against a 24 hour window. We believe that this method provides additional information especially in identifying upsets.

7. Conclusions and Future Directions

From the results we found strong evidence of technical charting's potential in sports wagering. In terms of accuracy, Swing was 58.6% accurate versus the odds-only approach of 59.8% which was found to be statistically equivalent. Swing did statistically outperform the absolute sentiment models in both the 96hr (50.0% accuracy) and 24hr (53.5% accuracy) timelines. If we were to examine predictions of just favorites, Swing's accuracy increased to 66.9%, a 7.1% higher accuracy value than the odds-only approach which is more than sufficient to offset a bookmaker's commission. In terms of payouts, Swing wagering netted a hypothetical return of $3,798.15 or $14.84 per wager versus a $12.21 loss per wager on odds favorites. For predictions of upsets, Swing again outperformed with an average return of $35.28.

References

Abbasi, A., H. Chen, et al., (2008). Sentiment Analysis in Multiple Languages: Feature Selection for Opinion Classification in Web Forums. *ACM Transactions on Information Systems* 26(3).

Blair, A., (1996). Guide to Charting. *Pitman, London*.

Brock, W., J. Lakonishok, et al., (1992). Simple technical trading rules and the stochastic properties of stock returns. *The Journal of Finance* 47(5):1731-1764.

Luckner, S., J. Schroder, et al., (2008). "On the Forecast Accuracy of Sports Prediction Markets" in *Negotiation, Auctions, and Market Engineering*. H. Gimpel, N. Jennings, G. Kersten, A. Ockenfels and C. Weinhardt. Berlin, Springer. 2**:** 227-234.

OpinionFinder.com, (2016). Subjectivity Analysis System. Retrieved on November 13, 2016 from http://mpqa.cs.pitt.edu/opinionfinder/ .

Schumaker, R. P. and H. Chen, (2008). Evaluating a News-Aware Quantitative Trader: The Effects of Momentum and Contrarian Stock Selection Strategies. *Journal of the American Society for Information Science and Technology* 59(1):1-9.

Schumaker, R. P., A. T. Jarmoszko, et al., (2016). Predicting Wins and Spread in the Premier League Using a Sentiment Analysis of Twitter. *Decision Support Systems* 88(8):76-84.

Schumaker, R. P., Y. Zhang, et al., (2012). Evaluating Sentiment in Financial News Articles. *Decision Support Systems* 53(3):458-464.

Sinha, S., C. Dyer, et al., (2013). Predicting the NFL using Twitter. *ECML/PKDD 2013 Workshop on Machine Learning and Data Mining for Sports Analytics*.

Spann, M. and B. Skiera, (2008). Sports Forecasting: A Comparison of the Forecast Accuracy of Prediction Markets, Betting Odds and Tipsters. *Journal of Forecasting* 28(1):55-72.

Statista.com, (2016). Statistics and Facts on Sports Betting. Retrieved on July 13, 2016 from http://www.statista.com/topics/1740/sports-betting/.

Surowiecki, J., (2004). "The Wisdom of Crowds". New York, Doubleday.

Tumasjan, A., T. O. Sprenger, et al., (2010). Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. *The Fourth Annual International AAAI Conference on Weblogs and Social Media* 10:178-185.

Wilson, T., P. Hoffmann, et al., (2005). OpinionFinder: A System for Subjectivity Analysis. *Human Language Technology Conference*, Vancouver, Canada.

Wong, W.-K., M. Manzur, et al., (2003). How rewarding is technical analysis? Evidence from Singapore stock market. *Applied Financial Economics* 13(7):543-551.