Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media

The Wisdom of Bookies? Sentiment Analysis Versus the NFL Point Spread

Yancheng Hong* **and Steven Skiena**†
Department of Computer Science, Stony Brook University
Stony Brook, NY 11794-4400, USA

Abstract

The American Football betting market provides a particularly attractive domain to study the nexus between public sentiment and the wisdom of crowds. In this paper, we present the first substantial study of the relationship between the NFL betting line and public opinion expressed in blogs and microblogs (Twitter).

We perform a large-scale study of four distinct text streams: LiveJournal blogs, RSS blog feeds captured by Spinn3r, Twitter, and traditional news media.

Our results show interesting disparities between the first and second halves of each season. We present evidence showing usefulness of sentiment on NFL betting. We demonstrate that a strategy betting roughly 30 games per year identified winner roughly 60% of the time from 2006 to 2009, well beyond what is needed to overcome the bookie's typical commission (53%).

Introduction

The *wisdom of crowds* (Surowiecki 2004) is notion that the collective opinion of a large, diverse group of individuals can produce more accurate information than the judgement of a particular expert.

We believe that the American Football betting market makes a particularly attractive domain to study the nexus between public sentiment and the wisdom of crowds, for several reasons:

- *Market Size* – The National Football League (NFL) represents the largest sports gambling market in the U.S.. The American Gaming Association states that $2.58 billion was legally wagered in Nevada's sports books, while the National Gambling Impact Study (NGISC) estimates that illegal football wagering totals as much as $380 billion annually (AmericanGamingAssociation 2009).
- *Degree of Popular Interest* – Professional football is the most popular spectator sport in the U.S., as measured by television ratings. This enormous popularity implies that a large community of self-proclaimed experts express themselves in blogs, forums, and microblogging environments. The weekly frequency of NFL games means this each game is subject to intense anticipation and analysis by this crowd.
- *Time Scale* – Each NFL team plays only one game per week, almost exclusively on Sunday. This provides a well defined news sampling period (Tuesday to Saturday) to capture commentary about each prospective game . The NFL point spread is quite stable compared with the price of more traditional financial securities, permitting meaningful analysis to be performed on a daily scale.

In this paper, we present the first substantial study of the relationship between the NFL betting line and public opinion expressed in blogs and microblogs (Twitter).

Major results from our work include:

- *The Deterministic Nature of the Line* – Relatively straightforward models prove sufficient to generate betting lines very close to what are seen in practice.
- *Early Season Effects on Sentiment and the Line* – The beginning of a football season is particularly challenging to both bookies and the betting public, because very little reliable statistical evidence of team performance exists.
- *The Interrelationship Between Social and Professional Media* – We perform a large-scale study of four distinct text streams: LiveJournal blogs, Spinn3r blog feeds , Twitter, and traditional news media. We present the degree to which team volume and sentiment coverage agrees across these separated sources, concluding that mainstream news serves as a good proxy for all of them.
- *Geographical Bias in Team Sentiment* – One may anticipate that each team's reported sentiment differs substantially between the local and national media. Local media sentiment appears less reliable than national media in terms of correlation with the betting line.
- *A Sentiment-Based Betting Strategy* – We present promising results on a sentiment-based betting strategy where we bet on games where our model line differs most substantially from the published line. We demonstrate that a strategy betting roughly 30 games per year identified winner roughly 60% of the time from 2006 to 2008, well beyond what is needed to overcome the bookie's typical commission (53%).

*yancheng@ust.hk
†skiena@cs.sunysb.edu

Background

NFL Point Spread Betting Market

The NFL sport betting market is organized in the form of point spread betting. The bettor bets on either the “Favorite” team or the “Underdog” team against a point spread line established by the bookies.

This mechanism is chosen in order to make an active market for both sides of the wager. The bookies need to compensate the weaker team so that the betting capital for both sides are roughly the same.

Sport betting has an “11 for 10” rule, which means that you need to bet 11 dollars to win 10 dollars. The extra one dollar is for bookies’ commission. The desired winning rate (P) for a bettor to break even is:

$$10 * P = 11 * (1 - P), \text{ where } P = 52.4\% \quad (1)$$

It is realistic to assumed that the bookies try to maintain a balanced book so that they can earn their commission without taking any risk. According to American Gaming Association, gross revenue for Nevada’s sports books was $136.4 million while more than $2.5 billion was wagered in 2008. The fact that more than 94 percent of all bets placed were returned to the winning bettors suggests that the above assumption is probably true. However, this assumption implies the point spread is the best predictor for bettor’s behavior instead of the game outcome.

Literature Reviews

Lacey (Lacey 1990) tested over 15 trading strategies over 1984-1986 NFL season and most of them could not gain abnormal returns.

Golec and Tamarkin (Golec and Tamarkin 1991) present some evidence showing that the NFL betting market are systematically biased estimator for outcomes via statistical tests.

Gray and Gray (Gray and Gray 1997) used probit model instead of regression model and found that widely documented inefficiencies in the NFL betting market dissipate over time.

Vergin (Vergin and Sosik 1999) suggested that there is a home field advantage based on 59.2% winning rate generated from betting on home team for Monday night and playoff games. They also gave possible explanations for this phenomenon, which are learning, travel and crowd factors.

Avery (Avery and Chevalier 1999) showed that the bettors do have the hypothesized betting proclivities on different sentiment source, like the advice of experts, prestige teams, and past winners. However, these sentiment variables serves as a significant predictor of point spread movements but not of the actual outcomes.

Vergin (Vergin 2001) found that the bettors overweigh outstanding positive performance over different number of recent games but do not overweigh the negative recent performance.

Dare (Dare and Holland 2004) found that there are fairly weak evidence for inefficiency in NFL betting market from bias for home underdog. But they conclude that it is too small to be exploited.

Lydia

The *Lydia* text analytics system uses news analysis on news, blog, and other web sources to build a relational model of people, places, and things.

Here we provide a brief overview of how sentiment analysis works in *Lydia*. We refer the reader to our original papers (Bautin, Vijayarenu, and Skiena 2008; Godbole, Srinivasaiah, and Skiena 2007) for details.

The *Lydia* sentiment analysis system is based on lexicons of positive and negative words, and associating entities with sentiment of co-occurring words from these lexicons. The *Lydia* sentiment lexicons were constructed by starting from small sets of seed words of incontrovertible polarity, targeted to each of six specific domains: *business*, *crime*, *health*, *politics*, *sports*, and *media*. The synonyms and antonyms of an electronic dictionary (Wordnet, (Miller 1995)) enable us to expand each seed set into a full sentiment lexicon, using a graph-theoretic approach described in (Godbole, Srinivasaiah, and Skiena 2007). Our *general* sentiment lexicons represent the union of these underlying sub-lexicons.

Sentiment Statistics

The public sentiment for NFL teams is generated from the Lydia. The original sentiment series are daily positive and negative raw counts for each NFL teams.Based on the raw sentiment counts series, we developed a measure of relative favorableness for the team A over team B as follows:

$$Favorable(A) = \frac{\frac{(Pos_A + Neg_B) - (Neg_A + PosB)}{Pos_A + Pos_B + Neg_A + Neg_B} + 1}{2} \quad (2)$$

The favorable score for team A lies between [0,1] and can be viewed as the winning possibilities of team A for this game suggested by their sentiments. We do our training and analysis based on the 683 NFL games from 2006 to 2008 and use 2009 data to evaluate our findings.

Data Sources

- *Dailies*, which includes over 500 newspaper from both United States and international sources. The time range is from 2004 to present.
- *Live Journal blogs*, which includes all the blogs provided by live journal. The time range is from 2006 to 2008.
- *Spinn3r*, which is a collection of worldwide blogs graciously provided by www.spinn3r.com since 2007.
- *Twitter*, is a free social networking and micro-blogging service that enables its users to send and read tweets, which is short messages within 140 characters. It covers the first 15 weeks of the regular NFL season in 2009.

Additional data sources concerning NFL statistics are from *http://www.footballlocks.com/nfl_point_spreads.shtml*, *http://espn.go.com/nfl/powerrankings*, and *http://www.jt-sw.com/football/pro/results.nsf*.

Line Model

Correlation Analysis

Overall Correlation The correlation analysis has been done between adjusted point spread and each of the following performance measurement.

- *Power Ranking score*, which is calculated from weekly subjective ranking given by experts at ESPN.
- *Point For/Against score*, which is calculated from accumulated points the two teams get/loss for current season.
- *Winning Percentage score*, which is calculated from accumulated winning percentage of the two teams in current season.
- *General Sentiment score*, which is calculated from weekly aggregated General sentiment counts for both teams.
- *Sport Sentiment score*, which is calculated from weekly aggregated Sport sentiment counts for both teams.

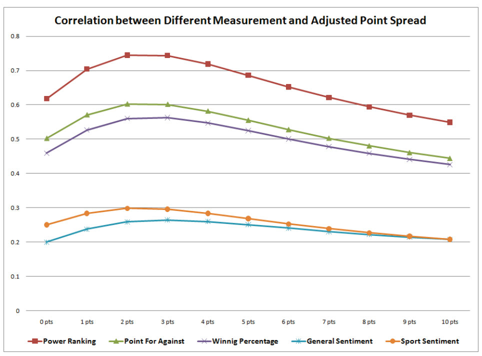

Figure 1: Correlation between different relative power measurement and homefield adjusted point spread

From Figure 1, we can notice the highest correlation coefficients always appear when we assume the home team get 2-points advantage. It is also consistent with what we got from linear regression model , which is about 1.8.

It is clear that the performance statistics scores have stronger correlation than sentiment score. However, sentiment score provide extra information as there are lots of literature suggesting people often overreact to information. For sentiment alone, the correlation line for sports sentiment are consistently higher than for general sentiment.

Sentiment Score Correlation Several factors must be considered before building models by sentiment variables:

- *Text data source*: Only Dailies and Live Journal covers the entire experimental period from 2006 to 2008. For Dailies, we can partition content into local newspapers, newspaper whose location is within 1.5 degree from each NFL team's stadium in latitude and longitude, for each team and non-local newspapers.
- *Entity name*: The experiments are done for sentiment time series based on both short name and full name for all the teams. The full name is the official name appears in NFL record. The short name is only the team name.

Sentiment Source	Name Match	Counting Method	Corre- -lation	Deg Free	P Value	T Stat.
National	Full	Cumulative	0.299	681	< 0.001	8.17
Local	Full	Cumulative	0.214	681	< 0.001	5.72
Non-lcl	Full	Cumulative	0.291	681	< 0.001	7.94
LiveJ	Full	Cumulative	0.183	681	< 0.001	4.85
National	Full	Weekly	0.261	681	< 0.001	7.05
Local	Full	Weekly	0.151	666	< 0.001	3.95
Non-lcl	Full	Weekly	0.253	681	< 0.001	6.83
LiveJ	Full	Weekly	0.016	527	0.722	0.36
National	Short	Cumulative	0.086	680	0.0254	2.24
Local	Short	Cumulative	0.109	680	0.0043	2.87
Non-lcl	Short	Cumulative	0.076	680	0.0467	1.99
LiveJ	Short	Cumulative	0.104	679	0.0064	2.73
National	Short	Weekly	0.072	680	0.0619	1.87
Local	Short	Weekly	0.066	662	0.087	1.71
Non-lcl	Short	Weekly	0.063	680	0.098	1.66
LiveJ	Short	Weekly	0.072	569	0.0838	1.73

Table 1: Statistics For all Sentiment Scores

- *Cumulative or week by week*: We calculate sentiment score with two different aggregated periods. One is individual-weekly aggregation, which aggregates the news from the Tuesday to Saturday before the game occurs on Sunday. The other is cumulative aggregation, which aggregates the all individual-weekly sentiments before the game for the current season.

The respective correlation coefficient and statistical significance of the coefficient for different sentiment scores are shown in Table 1. There are several observations:

- Most of the correlations are statistical significant at 95% significance level, except those calculated from the individual-weekly short name sentiment series.
- Cumulative sentiment score always has a higher correlation than individual-weekly sentiment score.
- The strongest correlation appears for sentiment scores derived from cumulative full name sentiment series. They are statistical significant at 0.001 significance level.
- The sentiment score calculated from the time series with the short name of the team have relatively significant correlation in the local area and live journal blogs.

We also investigate the correlation among sentiment score from different medias. From table 2 we can see that the sentiment from national professional media has a good correlation with sentiment generated by all other medias, including regional sentiment and social media sentiment. In other word, professional media information represents people's sentiment pretty well for both blogs or Twitters.

Model To the Line

There are three major linear regression models we have applied to fit the line. For the models incorporate sentiment scores, we can use different sentiment sources. The three linear regression models are:

- Objective-only model, in which only statistical performance score, including home field advantage, point for

	National	Local	Non-Lcl	LiveJ	Spinn3r	Twitter
National	1	0.469	0.986	0.432	0.449	0.509
Local		1	0.355	0.152	0.165	0.356
Non-Lcl			1	0.444	0.452	0.490
LiveJ				1	N/A	N/A
Spinn3r					1	0.083
Twitter						1

Table 2: Pairwise Sentiment Score Correlation

& against, winning percentage, and power ranking, are taken into the model.

- Sentiment-only Model, in which only the sentiment score and home field advantage are taken into the model.
- Full combined model, in which the sentiment score, home field advantage, and statistical performance score are all taken into the model.

Sentiment Based Betting

Based on the models we discussed in previous section, we developed simple betting strategies for NFL games.

If the predicted line from our model is less than the real point spread, we bet on the underdog team. If the predicted line from our model is bigger than the real point spread, we bet on the favorite team.

We select the games to bet on according to the absolute difference between predicted point spread and actual point spread. As sentiment score are cumulative, such point for and against, winning percentage and news sentiment score, we expected the model will operate different on first and second halves of the season. Therefore, we have separated results for both halves of the season as shown in Figure 2.

Model Comparison

To our surprise, sentiment-based models perform much better than any other models in second half of the season. The reason could be that people are not good at correctly interpreting public sentiment. Sentiment-based models can achieve 60% of winning rate from 2006 to 2008.

Since the performance of a model with different sentiment sources are highly consistent, we can compare the average winning rate of each model. From figure 2, we can notice that (1) for the first half of the season, the winning rates of all the three models are quite close to each other, and (2) for the second half of the season, the winning rate of news model is consistently higher than others. These findings are consistent with experimental results for 2009. Combining with the result for 2006-2009, the significant level of news model becomes significant to less than 0.08.

Conclusion

In the paper we examined the NFL games happened during year 2006 to year 2009. We have seen that sentiment score has significant correlation with the betting, just like the objective performance statistics. We also showed that the social media, like blogs and twitter, are as informative as professional newspaper media.

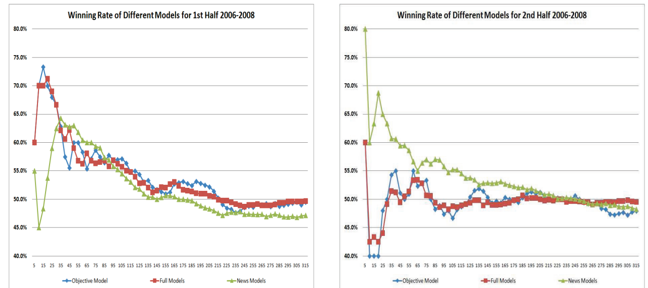

Figure 2: Models' performance comparison 2006-2008

Certain evidences for geographical sentiment difference on NFL teams are also presented. It is reasonable to say that local media is more likely to exaggerate the prospects of their own football team.

Several linear regression models were built based on the teams' performance statistics and/or sentiment of the teams. These models can generate predicted betting line relative close to real one. Simple betting strategies were implemented to take advantage of sentiment data and they seem to be profitable especially for the second half of the season.

References

AmericanGamingAssociation. 2009. Fact sheets about sports wagering.

Avery, C., and Chevalier, J. 1999. Identifying investor sentiment from price paths: the case of football betting. *Journal of Business*.

Bautin, M.; Vijayarenu, L.; and Skiena, S. 2008. International sentiment analysis for news and blogs. Second Int. Conf. on Weblogs and Social Media (ICWSM 2008).

Dare, W. H., and Holland, S. A. 2004. Efficiency in NFL betting market: modifying and consolidating research methods. *Applied Economics*.

Godbole, N.; Srinivasaiah, M.; and Skiena, S. 2007. Large-Scale Sentiment Analysis for News and Blogs. In *Proc. First Int. Conf. on Weblogs and Social Media*, 219–222.

Golec, J., and Tamarkin, M. 1991. The degree of inefficiency in the football betting market : statistical tests. *Journal of Financial Economics* 30(2):311–323.

Gray, P. K., and Gray, S. F. 1997. Testing market efficiency: evidence from the NFL sports betting market. *Journal of Finance*.

Lacey, N. J. 1990. An estimation of market efficiency in the NFL point spread betting market. *Applied Economics*.

Miller, G. A. 1995. WordNet: a lexical database for English. *Commun. ACM* 38(11):39–41.

Surowiecki, J. 2004. *The wisdom of crowds*. Doubleday.

Vergin, R. C., and Sosik, J. J. 1999. No plce like home: an examination of the home field advantage in gambling strategies in NFL football. *Journal of Economics and Business*.

Vergin, R. C. 2001. Overreaction in the NFL point spread market. *Applied Financial Economics*.