Twocker

David Hall

March 13, 2012

1 Introduction

The rise of microblogging services like Twitter means that millions of people are sharing aspects of their lives that they've never shared before at a much finer-grained resolution than what was possible. With so much information, is it possible that the records of their attitudes and economic activity can be harnessed?

Our task is to predict movements of financial market based on tweets. In particular, we hypothesize that the movement of *individual* stocks and bonds can be effectively predicted by tweets in the past, even relative to the stock market action the day? That is, is the market leaving tells in the form of tweets?

2 Prior Work

Perhaps unsurprisingly, the sentiment of tweets can be used to predict traditional opinion polls and other macro economic indicators. O'Connor et al. [2010] demonstrated that tweets can be predictive of presidential approval polls, election forecasts, and even consumer confidence. The method they used is surprisingly simple. They gathered tweets with topic-specific keywords. For example, for predicting presidential approval they looked for tweets containing "obama." Then, they counted the number of words with negative and positive valence in these tweets, and normalized. After some smoothing, their metric was positively correlated with approval. Moreover, it was positively correlated with *future* opinions.

In a similar vein, Zhang et al. [2009] showed that expressions of hope or fear (and indeed, any emotion at all) were negatively correlated with stock market indices. In fact, these same measures were positively correlated with measures of volatility. In a similar vein, Bollen et al. [2011] found that tweets reflecting "calm" moods were highly predictive of the Dow-Jones Industrial average, even 6 days in advance.

3 Approach

Our approach will build on this prior work by trying to link sentiment to particular stocks. Broadly, our methodology will be similar to O'Connor et al. [2010], though we suspect that other indicators (such as volume of tweets or deltas in other more basic measures) will be useful.

However, we are not interested in merely seeking correlates, but in prediction. Thus, we will also include standard predictors, such as volatility, volume, and previous returns into our model. Indeed, these same standard predictors can be used as a kind of baseline which our system can seek to beat.

Of course, it is likely the case that some stocks are more predictable than others. In particular, consumer-facing stocks are more likely to be talked about—and more likely to make the news—than non-consumer facing stocks. Therefore, the same statistics or methodology may not work for the same stocks.

However, while it might be nice from a sociological or economic perspective to be able to find a single set of indicators that can predict any given stock, from a financial or machine-learning perspective this is not strictly necessary. In particular, if this were a practical project, we would have the luxury of choosing which financial transactions to undertake.

Thus, this problem can be interpreted as an instance of the well-studied "bandit" problem, wherein at each time step t, we select an action (here, a financial transaction), and receive some reward r. Our goal is to maximize our reward over the long term. Thus, we will apply standard techniques from this area for the prediction task. Hopefully, the predictors may yield some insight into what behaviors are associated with twitter.

References

Johan Bollen, Huina Mao, and Xiao-Jun Zeng. Twitter mood predicts the stock market. *J. Comput. Science*, 2(1):1–8, 2011.

Brendan O'Connor, Ramnath Balasubramanyan, Bryan R. Routledge, and Noah A. Smith. From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series. In *Proceedings of the International AAAI Conference on Weblogs and Social Media*, 2010.

Xue Zhang, Hauke Fuehres, and Peter A Gloor. Predicting stock market indicators through Twitter: "I hope it is not as bad as I fear". *Anxiety*, 26:1–8, 2009.