

AI That Picks Stocks Better Than the Pros

Blog Invited Guests

A computer science professor uses textual analysis of articles to beat the market.

By Christopher Mims

The ability to predict the stock market is, as any Wall Street quantitative trader (or quant) will tell you, a license to print money. So it should be of no small interest to anyone who likes money that a new system that works in a radically different way than previous automated trading schemes appears to be able to beat Wall Street's best quantitative mutual funds at their own game.

It's called the Arizona Financial Text system, or AZFinText, and it works by ingesting large quantities of financial news stories (in initial tests, from Yahoo Finance) along with minute-by-minute stock price data, and then using the former to figure out how to predict the latter. Then it buys, or shorts, every stock it believes will move more than 1% of its current price in the next 20 minutes - and it never holds a stock for longer.



The system was developed by Robert P. Schumaker of Iona College in New Rochelle and and Hsinchun Chen of the University of Arizona, and was first described in a paper published early this year. Both researchers continue to experiment with and enhance the system - more on that below.

Using data from five non-consecutive weeks in 2005, a period chosen for its lack of unusual stock market activity, here's how AZFinText performed versus funds that traded in the same securities (which were all chosen from the S&P 500):

	Return
AZFinText	8.50%
Vanguard Growth & Income (VONPX)	6.44%
BlackRock Investment Trust Portfolio Inv A (CEIAX)	5.48%
RiverSource Disciplined Equity Fund (ALEIX)	4.69%

Table 2. Simulated Trading results of S&P 500 quants

And here's how it performed compared to the top 10 quantitative mutual funds in the world, all of which draw from a much larger basket of securities, except of course for the included S&P 500 itself:

	Return
ProFunds Ultra Japan Inv (UJPIX)	24.73%
ProFunds Ultra Japan Svc (UJPSX)	24.59%
American Century Global Gold Adv (ACGGX)	12.96%
American Century Global Gold Inv (BGEIX)	12.93%
AZFinText	8.50%
Quantitative Advisors Emerging Markets Instl (QEMAX)	8.16%
Quantitative Advisors Emerging Markets Shs (OFFOX)	8.15%
S&P 500 Index	5.62%
Lord Abbett Small-Cap Value Y (LRSYX)	5.22%
Lord Abbett Small-Cap Value A (LRSCX)	5.19%
Quantitative Advisors Foreign Value Instl (QFVIX)	4.99%
Quantitative Advisors Foreign Value Shs (QFVOX)	4.95%

Table 1. Simulated Trading results of the Top 10 Quants

Software that analyzes textual financial information - quarterly reports, press releases, news articles - is nothing new. Researchers have been publishing on the subject since at least the mid-1990's.

However, previous approaches to this technique were hampered by either poor performance (averaging little better than chance) and / or requirements for unreasonable amounts of computational horsepower. Schumaker and Chen get around these issues by first radically shrinking the amount of text their system has to parse by boiling down all the financial articles the system ingests into words falling into specific categories of information.

Interestingly, these techniques and categories derive from classification schemes described at the 7th Message Understanding Conference, held in 1997, which was a Defense Advanced Research Projects Agency project to create new and better ways to extract information and meaning from texts. (At the time, they were concentrating on terrorist activities in Latin America, airplane crashes, rocket and missile launches and other things relevant to national security.)

Schumaker and Chen's system concentrates on Proper Nouns - people and companies - and combines information about their frequency with stock prices at the moment a news article is released. Using a machine learning algorithm on historical data, they look for correlations that can be used to predict future stock prices.

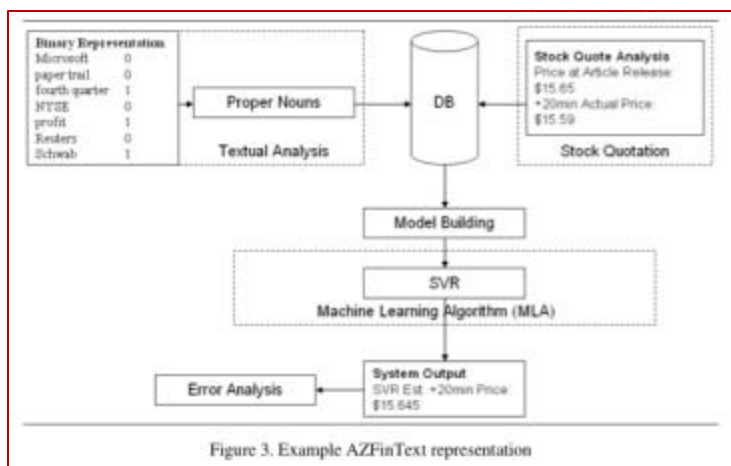


Figure 3. Example AZFinText representation

Further work with the AZFinText system has revealed oddities that may or may not remain relevant as researchers continue to apply it to other bodies of historical stock market and financial news data. For example, in a paper described on June 6 at the Computational Linguistics in a World of Social Media workshop, Schumaker went fishing for the Verbs most likely to cause a stock to move up or down in the next 20 minutes, and came up with a list of 211 terms that had some power to move stock prices. (In his work, 'verb' is a technical term, and does not exactly correspond with the conventional definition of the word.)

According to Schumaker:

The five verbs with highest negative impact on stock price are *hereto*, *comparable*, *charge*, *summit* and *green*. If the verb *hereto* were to appear in a financial article, AZFinText would discount the price by \$0.0029. While this movement may not appear to be much, the continued usage of negative verbs is additive.

The five verbs with the highest positive impact on stock prices are *planted*, *announcing*, *front*, *smaller* and *crude*.

Schumaker did not attempt to determine why these particular terms move stock prices, but it's interesting to note that the stock market does not appear to like the marketing buzzword "green," but is quite happy to hear any news at all about the term "crude," as in oil. ~END ~