

NewsdAI backend / backtesting

TL;DR: Newsdai model of finding and measuring rate of propagation of financial memes in news can reverse returns from negative (reversal) to positive (momentum) and can possibly lead to trade-able strategies.

Overview

This is a description of the NewsdAI model and backtesting results outlying the key performance points.

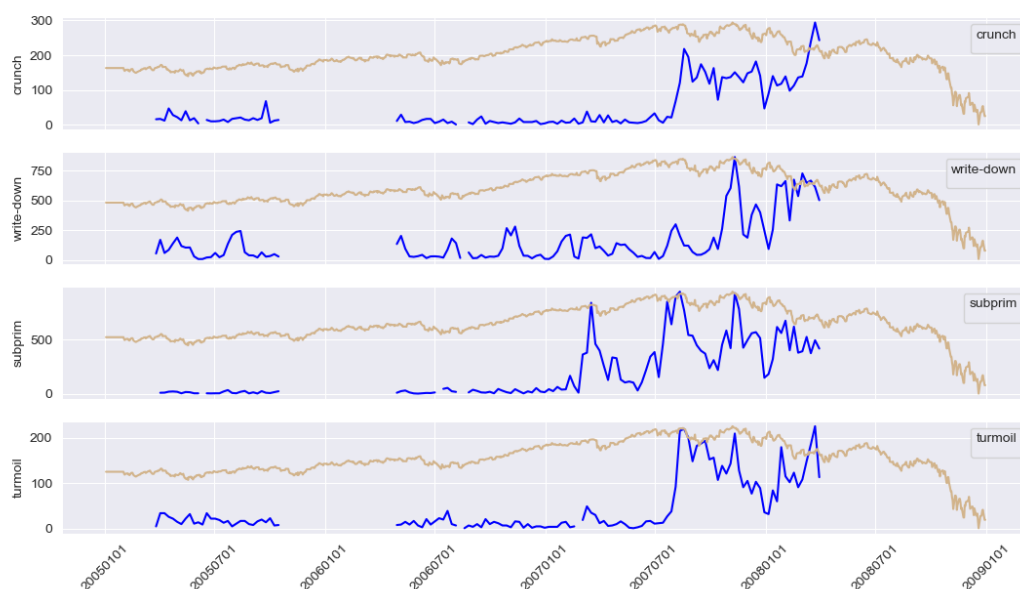
The NewsdAI is a news processing toolkit and a backend for search like UI where traders can combine string and numeric queries based on performance of instruments and news about them.

Here is a practical example, during and in the immediate aftermath of sudden market moves, we need to distinguish between the following scenarios resulting from breaking news coming about the company or instrument:

- price move is atomic, leaving no time to react
- price move becomes a trade, as story keeps developing



NewsdAI technologies breaks stories into 2 groups. Most of the stories are post-factum or trivial statement of facts that nonetheless leads to a singular "atomic" price move that creates a jump in return value. This category leaves no time to form an actionable strategy or trade. If the event



happen abruptly and became widely published, it may be too late to take action on it, or the signal could be too weak to bother.

A lot more interesting is second category where the story keeps developing. The timing could be just right for a profitable trade. NewsdAI backend technology identifies those story threads and extracts keywords belonging to an event category from the thread.

The escalation of the event is determined by estimating the rate of propagation of established keywords in the media and its effects. In essence *NewsdAI measures rate of propagation of financial memes with potential to create significant market impact in the financial news media.*

As an example let's look at housing crisis and find what were the market impacting keywords that were trending 6-12 months before the crisis hit.

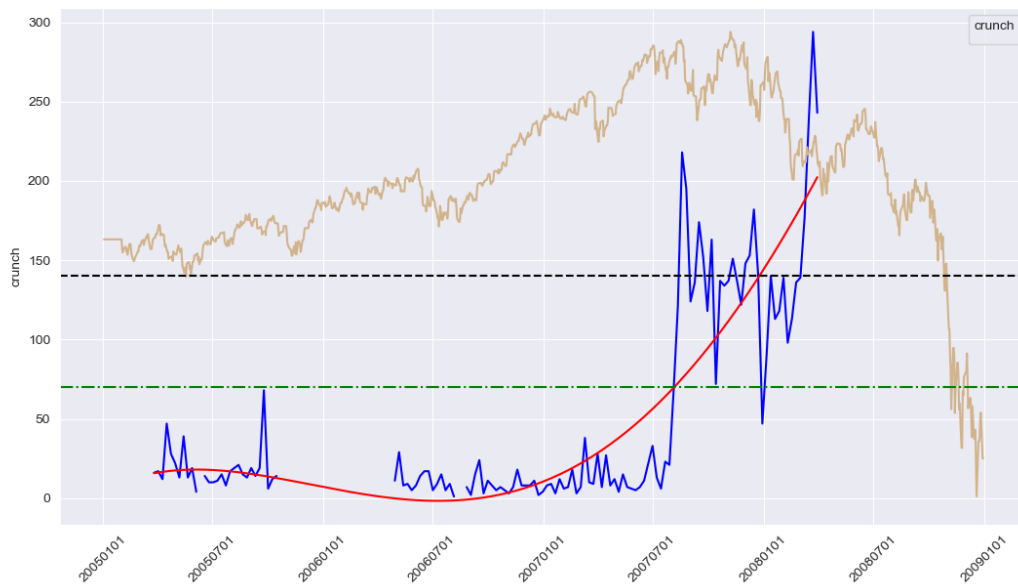
The plot below show most trending keywords with horizontal axis in dates and vertical axis is term frequency of keywords found summarily in published on the internet news.

For perspective, the term frequencies are plotted against scaled Standard and Poor index.

The keywords trending in months before 2008 crisis (Sept 18 Lehman bankruptcy happened) were:

- crunch - which expands to "credit crunch"
- write-down - has to do with banks writing down their investments due to bad financial decisions
- turmoil - expands to "market turmoil"
- subprime - expands to credit subprime

In the graph below I fit term frequency curve with a simple polynomial to check where fitted curve crosses thresholds set at multiples of standard deviation from the noise. When the curve is crossing the high water mark of the threshold that's an indication of a danger territory.



From the graph the threshold have been breached in Feb 2008 more than 6 months before the actual crisis.

Of course those examples are unrealistic in terms of time frame, almost no one forms a market view or enters a trade months before the event.

In order to test NewsdAI meme propagation rate technique I created a simple test.

NewsdAI trend escalation backtesting

In this backtest, I looked at available corpus of news stories linked reliably to the financial instruments, most of them are US company stock. Due to NDA I cannot disclose the source but my initial examination proved it's high quality.

To test my hypothesis that the NewsdAI model that identifies trending stories can also forecast trend in the stock price I developed a following "play" strategy. I was looking at the opposite of "No news is good news" event. I scanned a corpus for a news story about a company followed by significant market move, followed by one or more stories within a week after the market move happened. That would be an indication that the story keeps developing and the initial market move would continue in the same direction.

The "play" strategy steps were as follows:

- Identify all the significant market moves that exceeds 5% preceded by a large (>1000 words) story within previous 10 business days.
- If there is a follow up to the story (>1000) words, NEXT business day enter a position in the company in the same direction as proceeding market move. If the stock price went up, long the stock at the closing price NEXT business day after the news, if down, short it at the NEXT day closing price.
- Liquidate the position after 10 business days.

This "play" strategy could be further refined by filtering into subcategories by requiring certain keywords that would represent this category to be present in the body of the story or in the headline.

To test against the null hypothesis, I also considered the same strategy without any news input:

- Identify all the significant market moves that exceeds 5%
- If the stock price went up, long the stock at the closing price NEXT business day after the market move, if down, short it at the NEXT day closing price.
- Liquidate the position after 10 business days.

Below is the table with results for the following strategies:

- “trend” strategy that looks for a trending escalating story but without any keywords
- “refined_trend” strategy is same as above but requiring the following keywords to be present in the story: (“approv”;“licens”;“propos”;“clearanc”;“approv”)
- “null” strategy is the strategy that does not require

name	total return	sample size	sharpe
trend	1.6%	1680	0.05
refined_trend	6.5%	52	1.3
null	-1.2%	33512	-0.01

Notice that the third “null” strategy have negative total return, which indicates that general tendency of the market is to recover after large market moves in what is called reversal pattern.

In first two trend detection and refined trend strategy the return is positive and reversal becomes momentum based pattern. Still in terms of significance the total return still remains practically zero. Only after refining the strategy with matching for keywords:

“approv”;“licens”;“propos”;“clearanc”;“approv” that the return gets anywhere close to being practical.

The keywords being chosen as an example. They are usually indication of some sort of deal between company and a vendor or a regulatory body. Those keywords belong to the classification category the NewsdAI technology assigns dynamically.

Conclusion

NewsdAI model of detecting and measuring escalation of news trend reverses returns from negative (reversal) to positive (momentum) values. Narrowing the strategy with additional filters on keywords one can increase the signal and start getting noticeable returns.