

Algo Depth

Candlestick Pattern Recognition

July, 2017

By: Shiyuan Zhang

Alpha Signal

An equity trading strategy using nearest neighbor
algorithms to predict future return



Data Mining For Alpha Signals

In this research we mine past data for similar N day patterns, and then use that information to predict return for next several days. It's a strategy based on K-Nearest Neighbor algorithm. There are several parameters we choose, such as the size of the lookback window and how many days ahead to predict return.

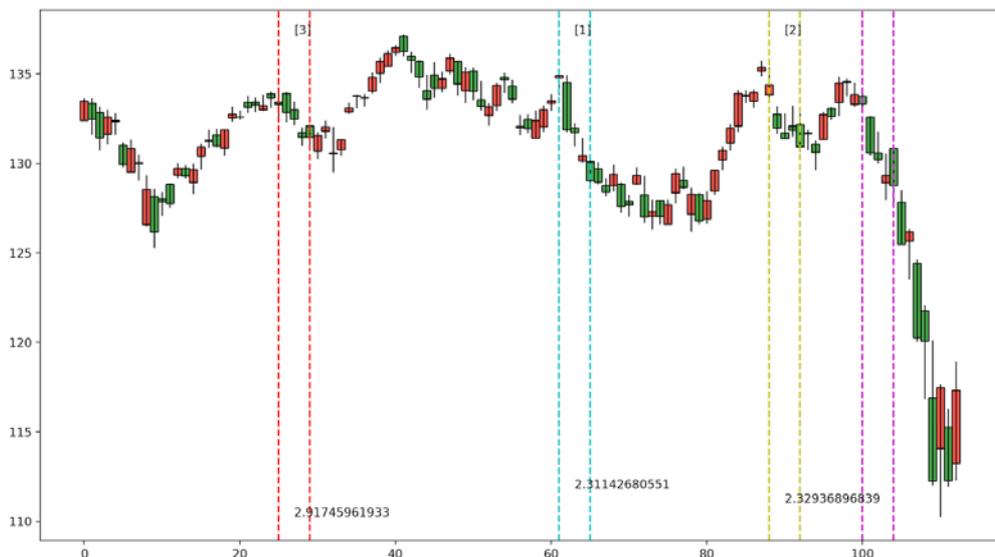
Pattern Recognition

How to measure the similarity between the patterns is the most important part of this research, the main idea behind this strategy is that if the two candlesticks pattern are very similar to each other, then we assume that the trend afterwards will also be very similar. To measure similarity, for example, if we choose the length of the pattern is 3, we need 11 numbers. Assume one pattern start from n, the other one start from m.

Then, the L2 distance between two patterns can be defined as:

$$l2 \text{ norm distance} = \left(\frac{\text{open}(n)}{\text{close}(n)} - \frac{\text{open}(m)}{\text{close}(m)}\right)^2 + \left(\frac{\text{open}(n+1)}{\text{close}(n+1)} - \frac{\text{open}(m+1)}{\text{close}(m+1)}\right)^2 + \left(\frac{\text{open}(n+2)}{\text{close}(n+2)} - \frac{\text{open}(m+2)}{\text{close}(m+2)}\right)^2 + \left(\frac{\text{high}(n)}{\text{close}(n)} - \frac{\text{high}(m)}{\text{close}(m)}\right)^2 + \left(\frac{\text{high}(n+1)}{\text{close}(n+1)} - \frac{\text{high}(m+1)}{\text{close}(m+1)}\right)^2 + \left(\frac{\text{high}(n+2)}{\text{close}(n+2)} - \frac{\text{high}(m+2)}{\text{close}(m+2)}\right)^2 + \left(\frac{\text{low}(n)}{\text{close}(n)} - \frac{\text{low}(m)}{\text{close}(m)}\right)^2 + \left(\frac{\text{low}(n+1)}{\text{close}(n+1)} - \frac{\text{low}(m+1)}{\text{close}(m+1)}\right)^2 + \left(\frac{\text{low}(n+2)}{\text{close}(n+2)} - \frac{\text{low}(m+2)}{\text{close}(m+2)}\right)^2 + \left(\frac{\text{close}(n+1)}{\text{close}(n)} - \frac{\text{close}(m+1)}{\text{close}(m)}\right)^2 + \left(\frac{\text{close}(n+2)}{\text{close}(n+1)} - \frac{\text{close}(m+2)}{\text{close}(m+1)}\right)^2$$

We utilize the distance above and go back through the entire history to find top three patterns that most similar to current reference pattern. As shown below, the purple line indicating current reference pattern(length=5). The other three are the most similar patterns in the past 100 days (for illustration purpose). The number below each pattern is its distance from the reference pattern.



Research process

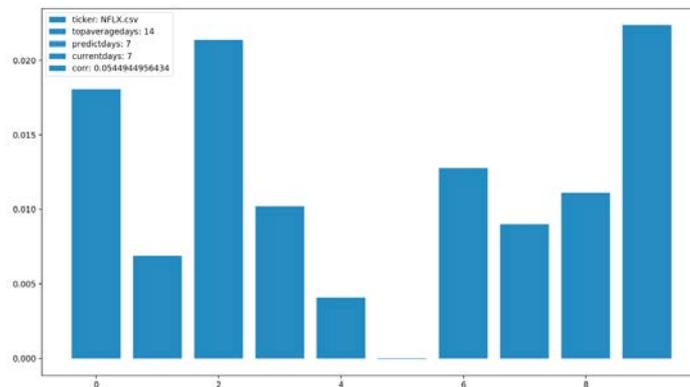
This simple trading idea is based on this pattern recognition, expecting the history will repeat to some extent. In this part I tried several distance measures like l1 norm, l2 norm, hamming distance etc. Then, observe them to see whether this recognition make sense or not. For example, after visualization, I found that because the volatility level is evolving all the time. Some two very similar patterns, because the body of one pattern are all much longer than the reference pattern, in this case their distance is actually very large. So, for each item in the distance, I choose to normalize them with their own bar size. In this way, the pattern looks more similar to the reference one.

Trading strategy

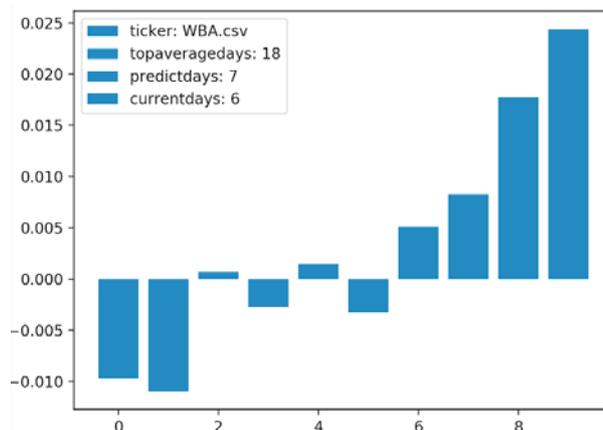
The strategy here is to select a number of the closest patterns we find, and simply average their next-day returns to arrive at an expected return (also tried weighted version using distance as weight). If it is a positive return we open long position and short position vice versa.

To test this trading idea, after finding proper distance measure, we do grid search using in-sample data of several tickers to see whether there is some positive signal of this strategy. Specifically, I use bar-plot to present result. The x-axis of bar-plot is predicted return and y-axis is real-return. If bar-plots of several set of parameter of a tickers is really good, we then switch strategy to out-of-sample data. As shown below, good bar-plot means predicted returns are highly correlated with real return.

Bad barplot



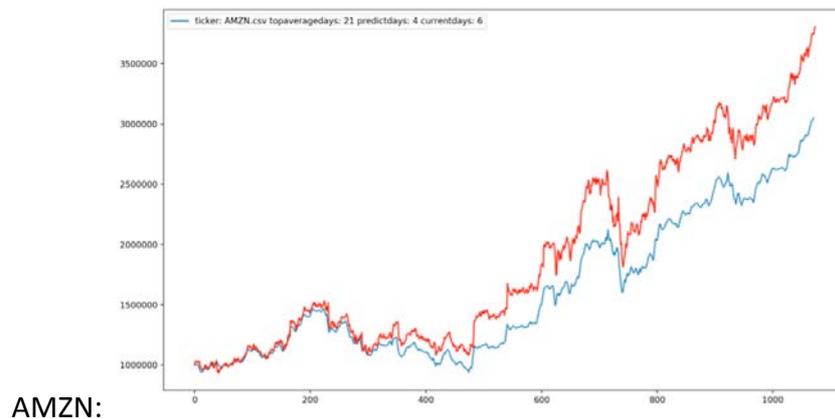
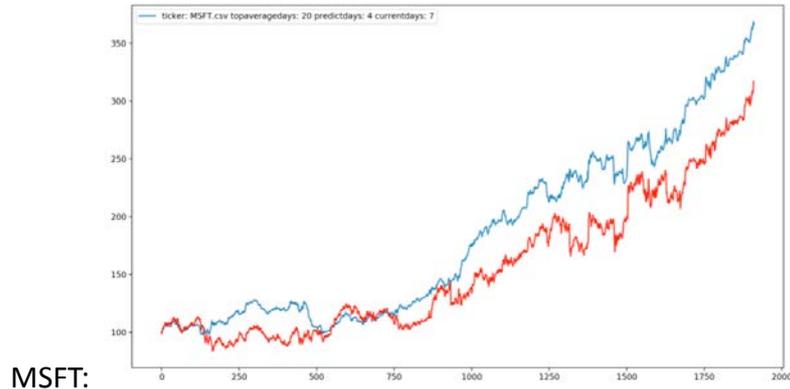
Good barplot



Result:

The strategy was tested on several ETF tickers as well as top 25 tickers in Nasdaq(daily data). The result is not robust, but there are some signals. This strategy can generate positive result on 4 tickers (INTC, PXI, AMZN, MSFT) according to our result (both in-sample and out-of-sample, transaction cost included). The PnL curve shown below (The red curve is stock close price series itself, the blue curve is the PnL curve on this ticker. X-axis represents number of trade, Y-axis represents profit).

PnL curve of AMZN and MSFT:



A potential problem here is that the PnL curve for MSFT and AMZN is very similar to its original price series, though this is a long-short strategy.

We attempt to include volume information into the strategy. As shown below, in the historical data set, we need to find the most similar patterns. Then we add this volume's distance information to our original strategy. After testing this idea, the result is not pleasant. More research is necessary to identify better trends in current candlesticks, using this information to predict future returns.

Important Disclaimer and Disclosure Information

Algo Depth makes no representations of any kind regarding this report. This includes, without limitation, warranties of title, merchantability, fitness for a particular purpose, non-infringement, absence of latent or other defects, accuracy, or the absence of errors, whether or not known or discoverable. In no event shall the author(s), Algo Depth or any of its officers, employees, or representatives, be liable to you on any legal theory (including, without limitation, negligence) or otherwise for any claims, losses, costs or damages of any kind, including direct, special, indirect, incidental, consequential, punitive, exemplary, or other losses, costs, expenses, or damages, arising out of the use of the report, including the information contained herein.

This report is prepared for informational and educational purposes only, and is not an offer to sell or the solicitation of an offer to buy any securities. The recipient is reminded that an investment in any security is subject to many risks, including the complete loss of capital, and other risks that this report does not contain. As always, past performance is no indication of future results. This report does not constitute any form of invitation or inducement by Algo Depth to engage in investment activity.

Algo Depth has not independently verified the information provided by the author(s) and provides no assurance to its accuracy, reliability, suitability, or completeness. Algo Depth may have opinions that materially differ from those discussed, and may have significant financial interest in the positions mentioned in the report.

This report may contain certain projections and statements regarding anticipated future performance of securities. These statements are subject to significant uncertainties that are not in our control and are subject to change.

Algo Depth makes no representations, express or implied, regarding the accuracy or completeness of this information, and the recipient accepts all risks in relying on this report for any purpose whatsoever. This report shall remain the property of Algo Depth and Algo Depth reserves the right to require the return of this report at any time.