**What I’d like to analyze?**

I would like to analyze the social media to collect **public opinion**, official channels to collect **professional** **comments**, general news for **technological** **and** **political** **changes**, in descending order of priority. After collecting such inputs, I wish to see its impact on **volatility** and **prices** taking into account **calendar effect**, **days of the week**, etc.

I also wish to add, as an additional function, the ability of the system to detect unusual price movements, especially those that defies the price relationship between the underlying and derivatives.

**What data to use?**

For public opinion, I intend to ask Seeking Alpha for their archived articles. following previous related research. I also wish to use Facebook, Twitter, forums such as HKGolden if possible.

For professional comments, I wish to use whichever is willing to provide archived news articles among Dow Jones News Services, CNN Money and other services.

For financial data, I hope that I could obtain historical data from either HKEX, NYSE or CME, depending on the most obtainable type of news. If most discussions were focusing on stocks in the US, for example, using financial data from NYSE would be most suitable.

**How to use it?**

For the preprocessing stage, I am thinking of two methods. The first one is a dictionary approach. This includes a large dictionary that links up all key words with relevant stocks / derivatives (Eg. iPhone with Apple Inc.). A similar dictionary would store the words with negative / positive emotions, in order to gauge whether a comment or article indicates primarily positive or negative prospect and to what extent. It will be conducted with the note that negative keywords are more suitable according to the research “Wisdom of Crowds”, however the reason behind at that time was that positive keywords are often used with a negation. The possibility of bypassing such limit would also be investigated in this research.

The second method involves using Natural Language Processing to do the gauging of prospect among comments and articles. I believe the technology now is mature enough to attempt such an action, and we can always use cloud services such as Google Cloud NLP for testing purposes.

**How to evaluate the result?**

Quite naturally, like all machine learning tasks, the accuracy of our prediction would be our utmost concern. First by historical data, followed by real time data. Response time would also be important after we passed the training stage.

**What do I have now?**

Wisdom of Crowds: The value of stock opinions transmitted through social media (2013) which claims to be both statistically significant and economically meaningful.

September 2018 CS-HSI Call Option Prices