Objective: Improve Yumo’s AAAI code and beat Yumo’s deep generative model and get result close to his ACL paper

Tips

* tweet data is quite noisy.
  + advanced attention mechanism for learning to focus on higher-quality tweets could be a good way to go.
* Social media and historical prices are heterogeneous, depicting market features from different perspectives. Using a better alignment / combination between tweets and prices could also benefit the prediction.
  + Weight based-approach (Let’s say, an IB banker tweets compared to others)
    - Automatic scoring algorithm?
      * train a Twitter users accuracy rate with historical prices to obtain the weight
      * Legal though?
  + Tweet delay? Tweets could be mentioning @stock and refer to very old past events? “the new facebook” of X?
* Visualisation parts of the sentiment analysis?
* Sparse data on certain time scale (seconds, minutes, hours, days)
* Binary stock movement prediction, what about taking into account the magnitude? Hence we can identify the main types of events that’s closely associated with a sharp increase/decrease

Ideas

* A real time application (web/android):
  + Getting latest tweets are possible (open API)
  + Getting latest price from Yahoo
  + Visualisation dashboard
* Self learning (if have time)
* User-specific weights (twitter users with higher accuracy, eg, IBD banker or just celebrities)
* Bernoulli classification (too simplified) -> Maybe quantitative?
* Identify big news which correlates to industry?
* Segment sentiment analysis by industry? (initially focusing on certain stocks?)
* Do experiments with context windows (Yumo’s paper is set to 30 days)
* Expert opinions (buy,sell, neutral)
* Technical analysis (candle turn, etc… check investing.com)

Questions

* How does a sample contain multiple tweets? How are they related to each other?
* What is sequence representation learner?
* Frequency threshold to 2 (discarding all words below this threshold)
* What is ReLU?
* RECURRENTNOVICEANALYST = people or just the variation of StockNet? (confused)
* LSTM cell?

Povilas = Upstreaming/downstreaming