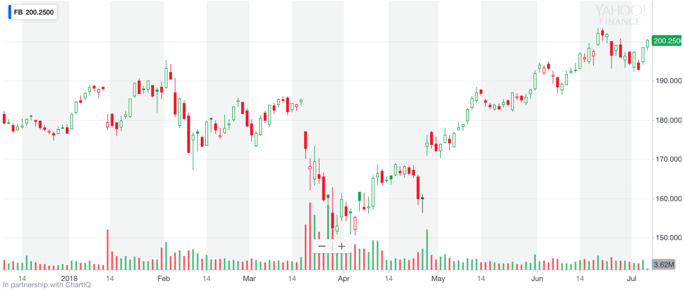
Stock Movement Prediction from Tweets and Historical Prices

Yumo Xu and Shay B. Cohen
Institute for Language, Cognition, and Computation
School of Informatics, University of Edinburgh

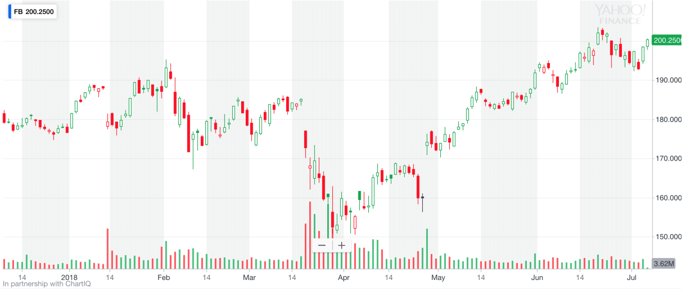
ACL, 2018.

https://yumoxu.github.io/, yumo.xu@ed.ac.uk

Who cares about stock movements?

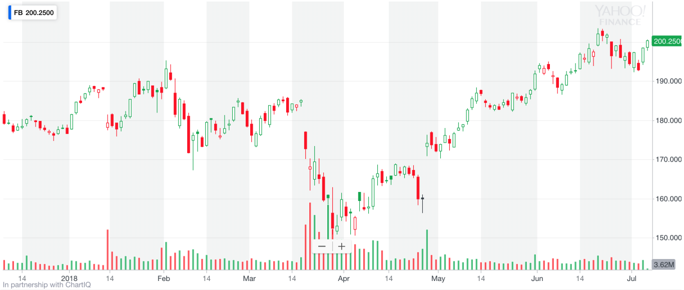


Who cares about stock movements?



No one would be unhappy if they could predict stock movements

Who cares about stock movements?



No one would be unhappy if they could predict stock movements

Investor

Government

Researcher

Background

- Two mainstreams in finance: technical and fundamental analysis

Background

- Two mainstreams in finance: technical and fundamental analysis
- Two main content resources in NLP: public news and social media

Background

- Two mainstreams in finance: technical and fundamental analysis
- Two main content resources in NLP: public news and social media
- History of NLP models

Feature engineering (before 2010)

↓

Topic models (2013-2015)

↓

Event-driven neural nets (2014-2015)

↓

Hierarchical attention nets (2018)

Background

- Two mainstreams in finance: technical and fundamental analysis
- Two main content resources in NLP: public news and social media
- History of NLP models

Feature engineering (before 2010)

↓

Topic models (2013-2015) | Generative

↓

Event-driven neural nets (2014-2015)

↓

Hierarchical attention nets (2018)

Background

- Two mainstreams in finance: technical and fundamental analysis
- Two main content resources in NLP: public news and social media
- History of NLP models

Feature engineering (before 2010)

↓

Topic models (2013-2015) Generative

↓

Event-driven neural nets (2014-2015)

↓

Hierarchical attention nets (2018)

However, it has never been easy...

Complexities

The market is highly stochastic, and we make temporally-dependent predictions from chaotic data.

Divide and Treat

1. Chaotic market information
   - Noisy and heterogeneous
2. High market stochasticity
   - Random-walk theory (Malkiel, 1999)
3. Temporally-dependent prediction
   - When a company suffers from a major scandal on a trading day, its stock price will have a downtrend in the coming trading days
   - Public information needs time to be absorbed into movements over time (Luss and d'Aspremont, 2015), and thus is largely shared across temporally-close predictions

Divide and treat

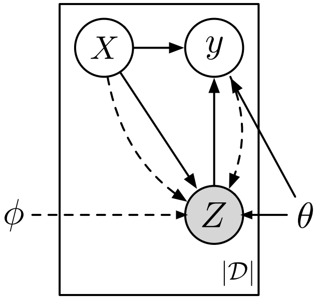
1. Chaotic market information — Market Information Encoder
   - Noisy and heterogeneous
2. High market stochasticity — Variational Movement Decoder
   - Random walk theory (Malkiel, 1999)
3. Temporally-dependent prediction — Attentive Temporal Auxiliary
   - When a company suffers from a major scandal on a trading day, its stock price will have a downtrend in the coming trading days
   - Public information needs time to be absorbed into movements over time (Luss and d'Aspremont, 2015), and thus is largely shared across temporally-close predictions

Problem Formulation

Stock Movement Prediction

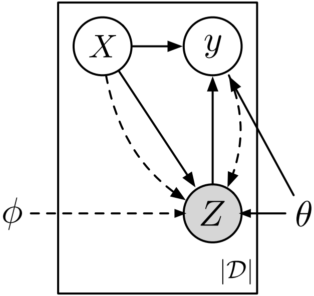
- We estimate the binary movement where 1 denotes rise and 0 denotes fall
- Target trading day: d
- We use the market information comprising relevant tweets, and historical prices, in the lag $[d - \Delta d, \ d - 1]$ where Δd is a *fixed* lag size

Generative Process



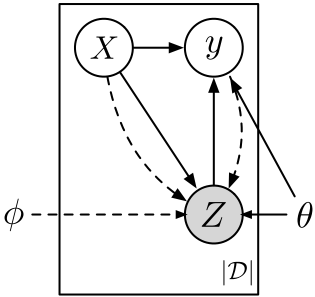
- T eligible trading days in the Δd lag
- Encode observed market information as a random variable $X = [x_1; \dots; x_T]$

Generative Process



- T eligible trading days in the Δd lag
- Encode observed market information as a random variable $X = [x_1; \ldots; x_T]$
- Generate the latent driven factor $Z = [z_1; \ldots; z_T]$

Generative Process



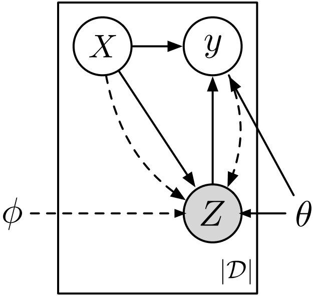
- T eligible trading days in the Δd lag
- Encode observed market information as a random variable $X = [x_1; \ldots; x_T]$
- Generate the latent driven factor $Z = [z_1; \ldots; z_T]$
- Generate stock movements $y = [y_1, \ldots, y_T]$ from X, Z

Factorization

- For multi-task learning, we model $p_\theta\left(y|X\right) = \int_Z p_\theta\left(y, Z|X\right)$ instead of $p_\theta(y_T|X)$
  - Main target: y_T
  - Temporal auxiliary target: $y^* = [y_1, \ldots, y_{T-1}]$
- Factorization

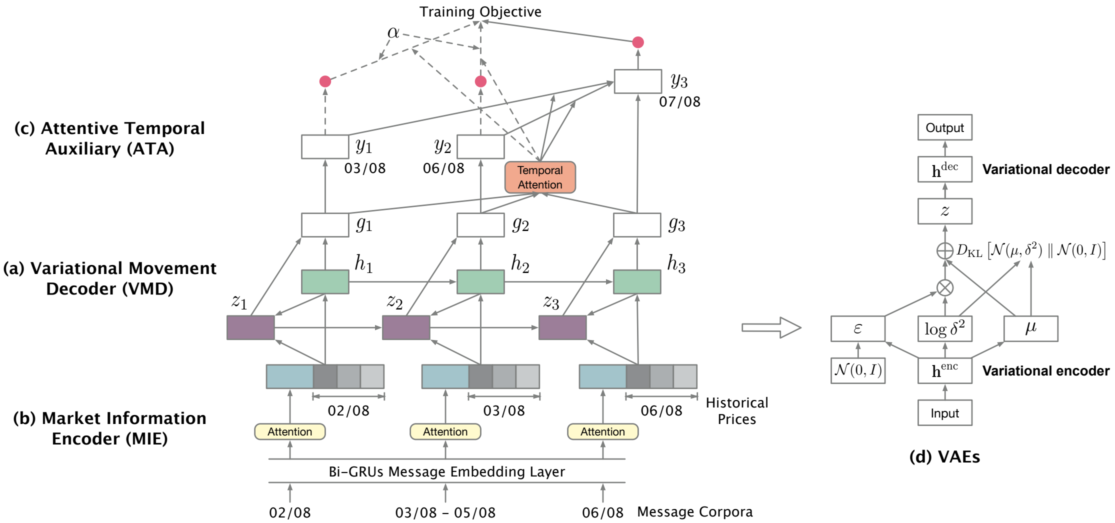
$$p_\theta\left(y, Z|X\right) = p_\theta\left(y_T|X, Z\right) p_\theta(z_T|z_{<T}, X)$$
$$\prod_{t=1}^{T-1} p_\theta\left(y_t|x_{\leq t}, z_t\right) p_\theta\left(z_t|z_{<t}, x_{\leq t}, y_t\right)$$

Primary components



1. Market Information Encoder (MIE)
   - Encodes X
2. Variational Movement Decoder (VMD)
   - Infers Z with X, y and decodes stock movements y from X, Z
3. Attentive Temporal Auxiliary (ATA)
   - Integrates temporal loss for training

StockNet architecture



Variational Movement Decoder

- Goal: recurrently infer Z from X, y and decode y from X, Z
- Challenge: posterior inference is intractable in our factorized model

Variational Movement Decoder

- Goal: recurrently infer Z from X, y and decode y from X, Z
- Challenge: posterior inference is intractable in our factorized model

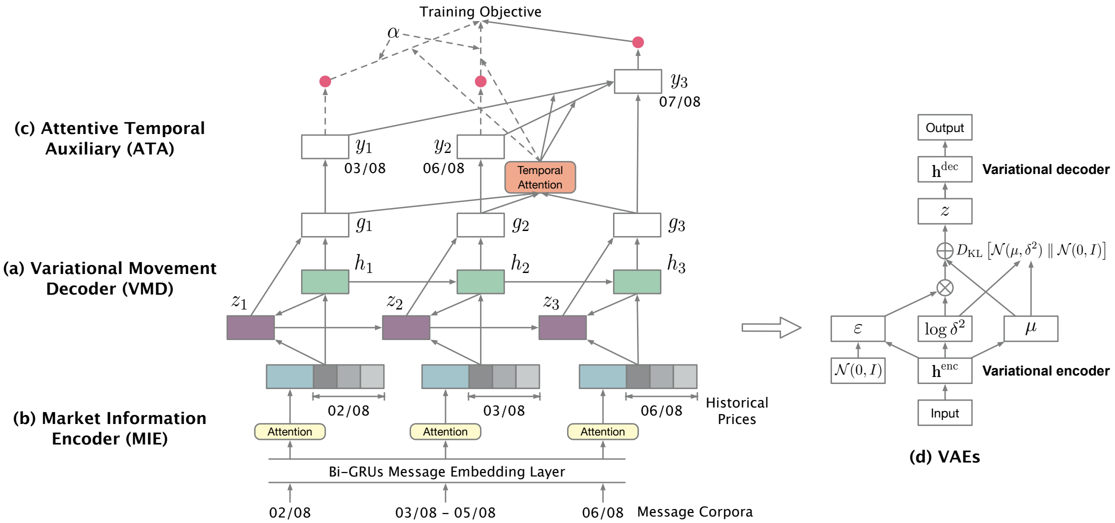
VAE solutions

- Neural approximation and reparameterization
- Recurrent ELBO
- Adopt a posterior approximator

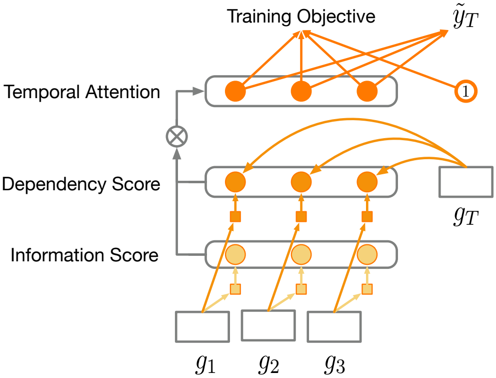
$$q_\phi\left(z_t|z_{<t}, x_{\leq t}, y_t\right) \sim \mathcal{N}(\mu, \delta^2 I)$$

where $\phi = \{\mu, \delta\}$

StockNet architecture



Interface between VMD and ATA



- Integrate the deterministic feature h_t and the latent variable z_t

$$g_t = \tanh(W_g[x_t, h_t^s, z_t] + b_g)$$

- Decode movement hypothesis: first auxiliary targets, then main target
- Temporal attention: v^*

Attentive Temporal Auxiliary

- Break down the approximated $\mathcal{L}$ to temporal objectives $f \in \mathbb{R}^{T \times 1}$

$$\begin{aligned} f_t = & \log p_\theta\left(y_t | x_{\leq t}, z_{\leq t}\right) \\ & - \lambda D_{\mathrm{KL}}\left[q_\phi\left(z_t | z_{<t}, x_{\leq t}, y_t\right) \| p_\theta\left(z_t | z_{<t}, x_{\leq t}\right)\right] \end{aligned}$$

- Reuse v^* to build the final temporal weight vector $v \in \mathbb{R}^{1 \times T}$

$$v = [\alpha v^*, 1]$$

  where $\alpha \in [0, 1]$ controls the overall auxiliary effects
- Recompose $\mathcal{F}$

$$\mathcal{F}(\theta, \phi; X, y) = \frac{1}{N} \sum_n^N v^{(n)} f^{(n)}$$

Experimental setup

- Dataset
  - Two-year daily price movements of 88 stocks
  - Two components: a Twitter dataset and a historical price dataset
  - Training: 20 months, 20,339 movements
  - Development: 2 months, 2,555 movements
  - Test: 2 months, 3,720 movements
- Lag window: 5
- Metrics: accuracy and Matthews Correlation Coefficient (MCC)
- Comparative study: five baselines from different genres and five StockNet variations

Baselines and variants

Baselines

- RAND: a naive predictor making random guess
- ARIMA: Autoregressive Integrated Moving Average
- RANDFOREST (Pagolu et al., 2016)
- TSLDA (Nguyen and Shirai, 2015)
- HAN (Hu et al., 2018)

StockNet variants

- HEDGEFUNDANALYST: fully-equipped
- TECHNICALANALYST: from only prices
- FUNDAMENTALANALYST: from only tweets
- INDEPENDENTANALYST: optimizing only the main target
- DISCRIMINATIVEANALYST: a discriminative variant

Results

Baseline models	Acc.	MCC
RAND	50.89	-0.002266
ARIMA	51.39	-0.020588
RANDFOREST	53.08	0.012929
TSLDA	54.07	**0.065382**
HAN	**57.64**	0.051800

StockNet variations	Acc.	MCC
TECHNICALANALYST	54.96	0.016456
FUNDAMENTALANALYST	58.23	0.071704
INDEPENDENTANALYST	57.54	0.036610
DISCRIMINATIVEANALYST	56.15	0.056493
HEDGEFUNDANALYST	**58.23**	**0.080796**

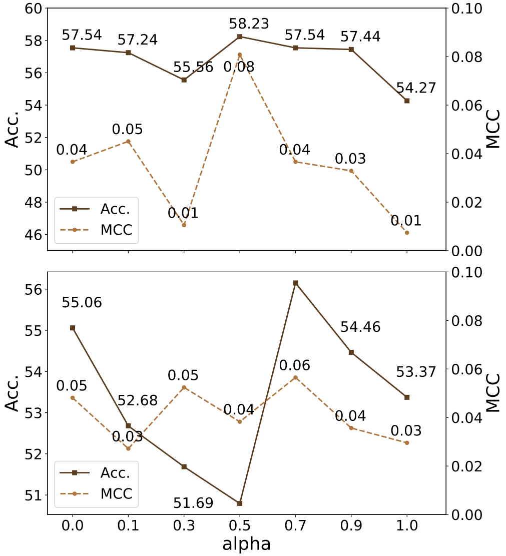
Baseline comparison

- The accuracy of 56% is generally reported as a satisfying result (Nguyen and Shirai, 2015)
- ARIMA: does not yield satisfying results
- Two best baselines: TSLDA and HAN

Variant comparison

- Two information sources are integrated effectively
- Generative framework incorporates randomness properly

Effects of temporal auxiliary



- The auxiliary weight $\alpha \in [0, 1]$ controls overall auxiliary effects

$$v = [\alpha v^*, 1]$$

- Our models do not linearly benefit from temporal auxiliary
- Tweaking α acts as a trade-off between focusing on the main target and generalizing by denoising

Summary

- We demonstrated the effectiveness of deep generative approaches for stock movement prediction from social media
- Outlook
  - Better way to integrate fundamental information and technical indicators
  - Other market signals, e.g. financial disclosures, periodic analyst reports and company profiles
  - Investment simulation with modern portfolio theory
- Dataset is available at `https://github.com/yumoxu/stocknet-dataset`

Ziniu Hu, Weiqing Liu, Jiang Bian, Xuanzhe Liu, and Tie-Yan Liu. 2018. Listening to chaotic whispers: A deep learning framework for news-oriented stock trend prediction. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*. ACM, Los Angeles, California, USA, pages 261–269.

Ronny Luss and Alexandre d'Aspremont. 2015. Predicting abnormal returns from news using text classification. *Quantitative Finance* 15(6):999–1012.

Burton Gordon Malkiel. 1999. *A random walk down Wall Street: including a life-cycle guide to personal investing*. WW Norton & Company.

Thien Hai Nguyen and Kiyoaki Shirai. 2015. Topic modeling based sentiment analysis on social media for stock market prediction. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*. Beijing, China, volume 1, pages 1354–1364.

Venkata Sasank Pagolu, Kamal Nayan Reddy, Ganapati Panda, and Babita Majhi. 2016. Sentiment analysis of twitter data for predicting stock market movements. In *Proceedings of 2016 International Conference on Signal Processing, Communication, Power and Embedded System*. IEEE, Rajaseetapuram, India, pages 1345–1350.

Appendix - Market Information Encoder

Temporal input: $x_t = [c_t, p_t]$

Corpus embedding c_t

- Multiple tweets with varied quality
- Message embedding: Bi-GRU
- Corpus embedding: messages composition with salience

$$u_t = \text{softmax}(w_u^\intercal \tanh(W_{m,u} M_t))$$
$$c_t = M_t u_t^\intercal$$

Historical price vector p_t

- Price signals: the adjusted closing, highest and lowest

$$\tilde{p}_t = \left[\tilde{p}_t^c, \tilde{p}_t^h, \tilde{p}_t^l\right]$$

- Normalization

$$p_t = \tilde{p}_t / \tilde{p}_{t-1}^c - 1$$

Appendix - Variational Inference

Latent factorization

$$q_\phi\left(Z|X,y\right) = \prod_{t=1}^{T} q_\phi\left(z_t|z_{<t},x_{\le t},y_t\right)$$

Likelihood equation

$$\begin{aligned}
&\log p_\theta\left(y|X\right)\\
=&D_{\mathrm{KL}}\left[q_\phi\left(Z|X,y\right) \parallel p_\theta\left(Z|X,y\right)\right]\\
+&\mathbb{E}_{q_\phi(Z|X,y)}\left[\log p_\theta\left(y|X,Z\right)\right]\\
-&D_{\mathrm{KL}}\left[q_\phi\left(Z|X,y\right) \parallel p_\theta\left(Z|X\right)\right]
\end{aligned}$$

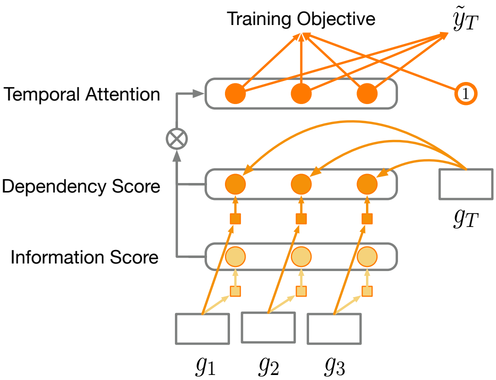
Recurrent ELBO

$$\begin{aligned}
&\mathcal{L}\left(\theta,\phi;X,y\right)\\
&= \sum_{t=1}^{T} \mathbb{E}_{q_\phi\left(z_t|z_{<t},x_{\le t},y_t\right)}\Big\{\log p_\theta\left(y_t|x_{\le t},z_{\le t}\right) -\\
&\quad D_{\mathrm{KL}}\left[q_\phi\left(z_t|z_{<t},x_{\le t},y_t\right) \parallel p_\theta\left(z_t|z_{<t},x_{\le t}\right)\right]\Big\}\\
&\le \log p_\theta\left(y|X\right)
\end{aligned}$$

where the likelihood term

$$p_\theta\left(y_t|x_{\le t},z_{\le t}\right) = \begin{cases} p_\theta\left(y_t|x_{\le t},z_t\right), & \text{if } t < T\\ p_\theta\left(y_T|X,Z\right), & \text{if } t = T. \end{cases}$$

Appendix - Attentive Temporal Auxiliary



- Information score

$$v_i' = w_i^{\mathsf{T}} \tanh(W_{g,i} G^*)$$

- Dependency score

$$v_d' = g_T^{\mathsf{T}} \tanh(W_{g,d} G^*)$$

- Integration

$$v^* = \zeta(v_i' \odot v_d')$$

Appendix - Trading-day Alignment

- We reorganize our inputs, including the tweet corpora and historical prices, by aligning them to the T trading days in a lag
- Specifically, on the tth trading day, we recognize market signals from the corpus $\mathcal{M}_t$ in $[d_{t-1}, d_t)$ and the historical prices p_t on d_{t-1}, for predicting the movement y_t on d_t

Appendix - Denoising Regularizer

- Objective-level auxiliary can be regarded as a denoising regularizer: for a sample with a specific movement as the main target, the market source in the lag can be heterogeneous

Example

- Affected by bad news, tweets on earlier days are negative but turn to positive due to timely crises management
- Without temporal auxiliary tasks, the model tries to identify positive signals on earlier days only for the main target of rise movement, which is likely to result in pure noise

- Temporal auxiliary tasks help to
  - Filter market sources in the lag as per their respective aligned auxiliary movements
  - Encode more useful information into the latent driven factor Z