Attention-based Stock Price Movement Predictions using 8-K filings

Mohamed Masoud
masoud@stanford.edu

Problem:

- Stock forecasting is important part of financial applications
- Stock prices are impacted by newly revealed information
- 8-K filings report updates on companies' **31** major events: Financial Disclosure, Mergers, Bankruptcies, Change of Management. etc.
- Presenting an attention based architecture for short-term stock dynamics forecasting (UP, DOWN, STAY).
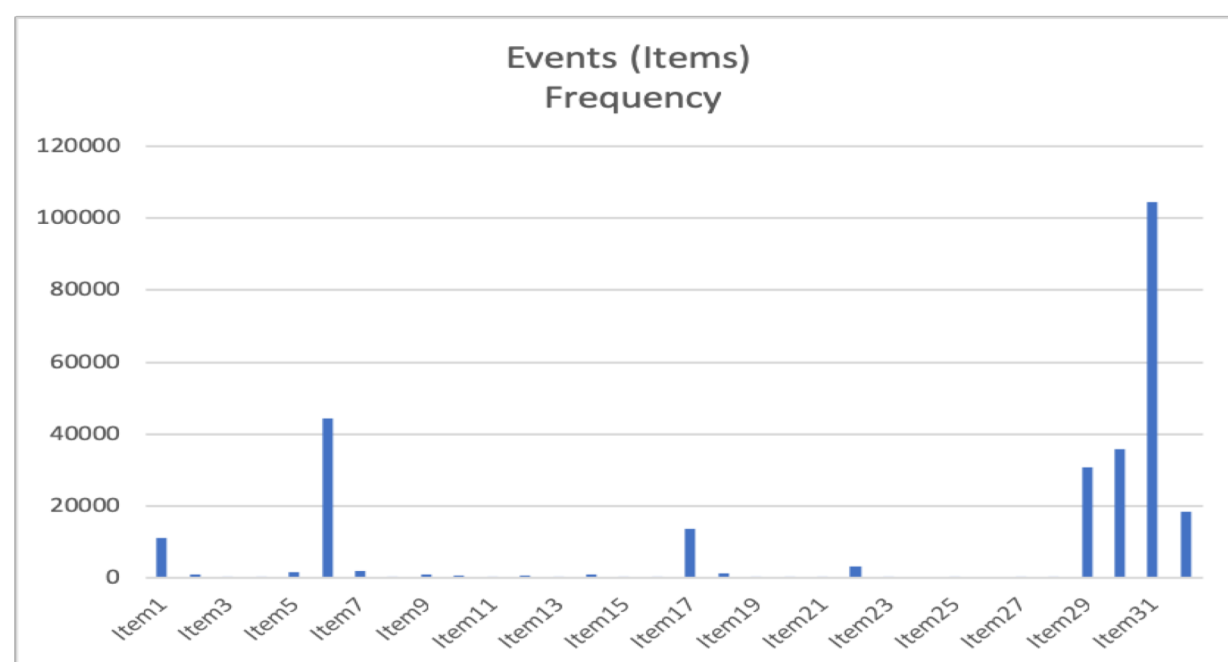
Data/Preprocessing:

- Lee et al., 2014 Stanford's NLP stock financial events dataset.
- Contains variable length 8-K reports for all S&P 500 companies between 2002 and 2012.
- **Preprocessing - Labeling**: to isolate the impact of 8-K report release: Normalized Rates

$$NR = \frac{(P_{after} - P_{before})}{P_{before}} - \frac{(GSPC_{after} - GSPC_{before})}{GSPC_{before}}$$

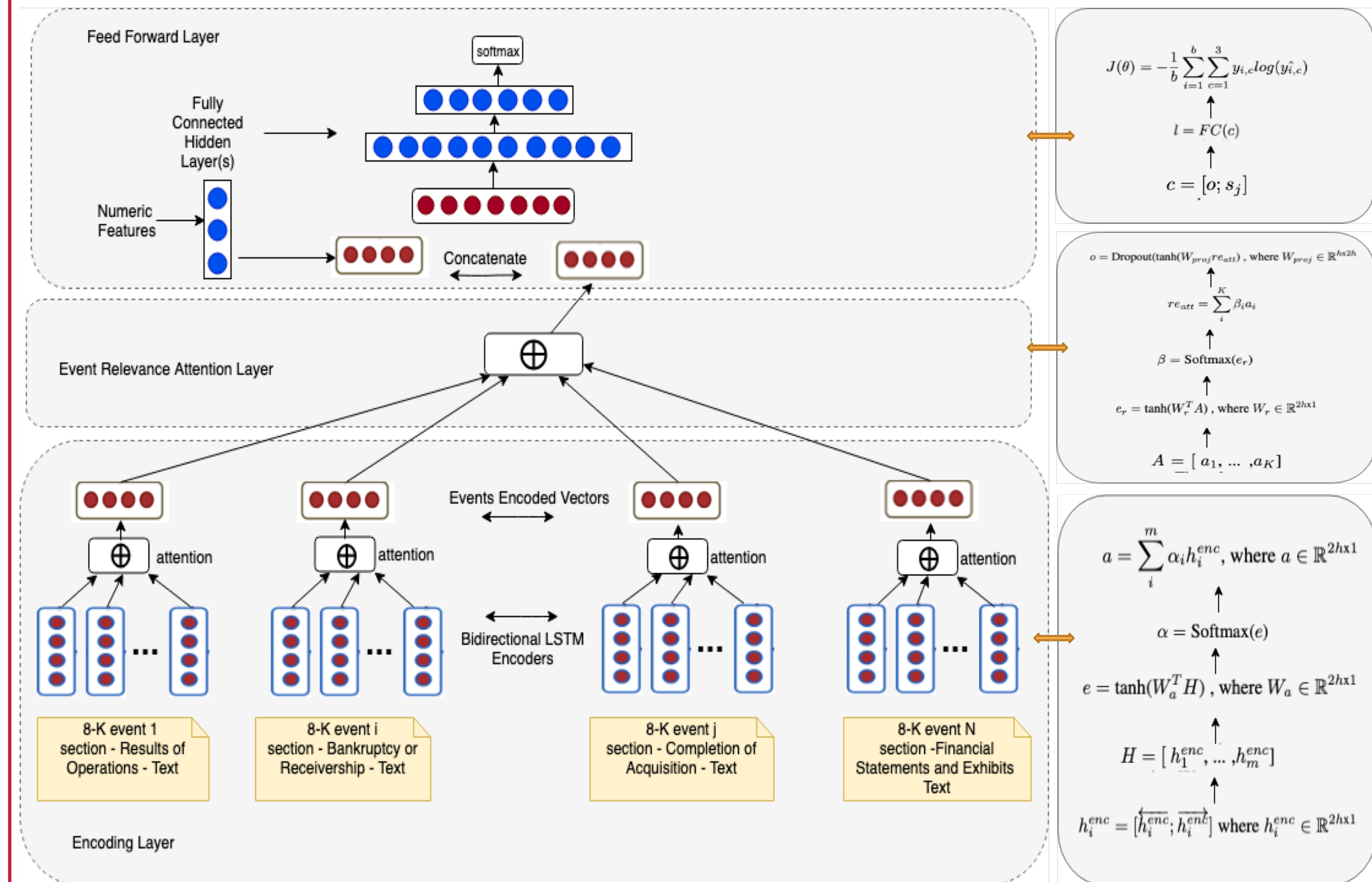- |NR| < 1% (STAY), NR>1% (UP), NR<-1% (DOWN)
- *138294* reports with < 3000 words.
- **Non-Linguistic Features:** Earnings (EPS) surprise = Reported EPS – consensus EPS Parsed EPS HTMLs → *16185* examples
- The reports parsed into 31 possible events' text

Events (Items) Frequency

The Attention-Model Architecture:

$$J(\theta) = -\frac{1}{b}\sum_{i=1}^{b}\sum_{c=1}^{3} y_{i,c} \log(y_{i,c})$$

$$l = FC(c)$$

$$c = [o; s_j]$$

$$o = \text{Dropout}(\tanh(W_{proj} re_{att})), \text{ where } W_{proj} \in \mathbb{R}^{h \times 2h}$$

$$re_{att} = \sum_{i}^{K} \beta_i a_i$$

$$\beta = \text{Softmax}(e_r)$$

$$e_r = \tanh(W_r^T A), \text{ where } W_r \in \mathbb{R}^{2h \times 1}$$

$$A = [a_1, \dots, a_K]$$

$$a = \sum_{i}^{m} \alpha_i h_i^{enc}, \text{ where } a \in \mathbb{R}^{2h \times 1}$$

$$\alpha = \text{Softmax}(e)$$

$$e = \tanh(W_a^T H), \text{ where } W_a \in \mathbb{R}^{2h \times 1}$$

$$H = [h_1^{enc}, \dots, h_m^{enc}]$$

$$h_i^{enc} = [\overrightarrow{h_i^{enc}}; \overleftarrow{h_i^{enc}}] \text{ where } h_i^{enc} \in \mathbb{R}^{2h \times 1}$$

8-K examples:

Apple (AAPL) - April 19, 2006., report snippet: EPS surprise = 9.5%

... We've generated over $10 billion in revenue and almost $1 billion in earnings in the first half of fiscal 2006, said Steve Jobs, Apple's CEO ... We're very pleased to report the second highest quarterly sales in Apple's history, resulting in year-over-year revenue growth of 34 percent and earnings growth of 41 percent ...

Visa Inc (V) on October 29, 2008., report snippet: EPS surprise = 3.5%

....On a GAAP basis, the Company reported a net loss of $356 million ...We remain intensely focused on helping our financial institution and retail clients through this difficult period . . . providing them with products and services that build deeper cardholder relationships and boost their own bottom line ... fiscal fourth quarter transactions processed by VisaNet, were 9.6 billion, an 11% increase over the prior year

References:

[1] Lee et al.,(2014) On the importance of text analysis for stock price prediction
[2] Liu et al.,(2017) Attention-Based Event Relevance Model for Stock Price Movement Prediction

Results and Discussion:

- The Single Event models 1, 3 outperformed, despite lengthy reports. The event attention-based encoder was able to extract some signals from the text pertinent to the classification
- Model-2 (Single Event Text with EPS), Underperformed – overfitting - High generalization error – Train Avg. accuracy ~ 59%) - 16185 examples
- The Multi-Event model slightly underperformed model-1 and model-3. multi-event dataset is **sparse**. **Imbalanced dataset**. The encoders' ability to extract relevant linguistic signals are impacted for the less represented events.
- AAPL example: Correctly classified (UP) - Positive Sentiment – Multi Model: STAY
- Visa Inc (V) (DOWN) example, models incorrectly classified (STAY). Bias in the language in reporting the positive and negative events

Model	Accuracy
Random Guess[*]	33.3
NMF 200[*]	55.3
Ensemble[*]	55.5
Single-Event-Text-Only	**57.02**
Single-Event-Text-EPS	51.92
Single-Event-Text-EventsTypes	56.71
Mutli-Event-Text-EventsTypes	56.07

[*] Lee et al., [1] Models.

Conclusion:

- The attention based models were able to extract the relevant linguistic signals from the report that help with the stock dynamic predictions.
- Need to combine with non-linguistic features to enhance the performance of the model.