



Making Trading Great Again: Trump-based Stock Predictions via doc2vec Embeddings

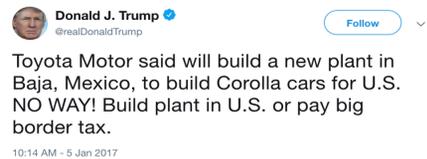
Rifath Rashid

Stanford University, Department of Computer Science

Stanford
Computer Science

Background

- One of the first seminal papers using Twitter to predict stock movements focused on relatively coarse categories of sentiment [1]
- More recent advancements in NLP and NLU have given rise to distributed representations (*word2vec*, *doc2vec*) [2, 3]
- On January 20th 2017, Donald Trump is inaugurated as President
- President Trump uses Twitter as his singular platform. His tweets and policies have been known to shake financial markets.



Problem

- Can contextually-rich document representations of Trump tweets successfully predict stock change?

Data

Trump Twitter Statistics (2010-2019)

Total Posts	38,471
Max posts on one day	160
Total business days with at least one post	2,139

Data

Frequent Words

'great'	2,891
'Great'	1,204
'America'	765
'big'	771
'Hillary'	657
'GREAT'	537
'Fake'	372

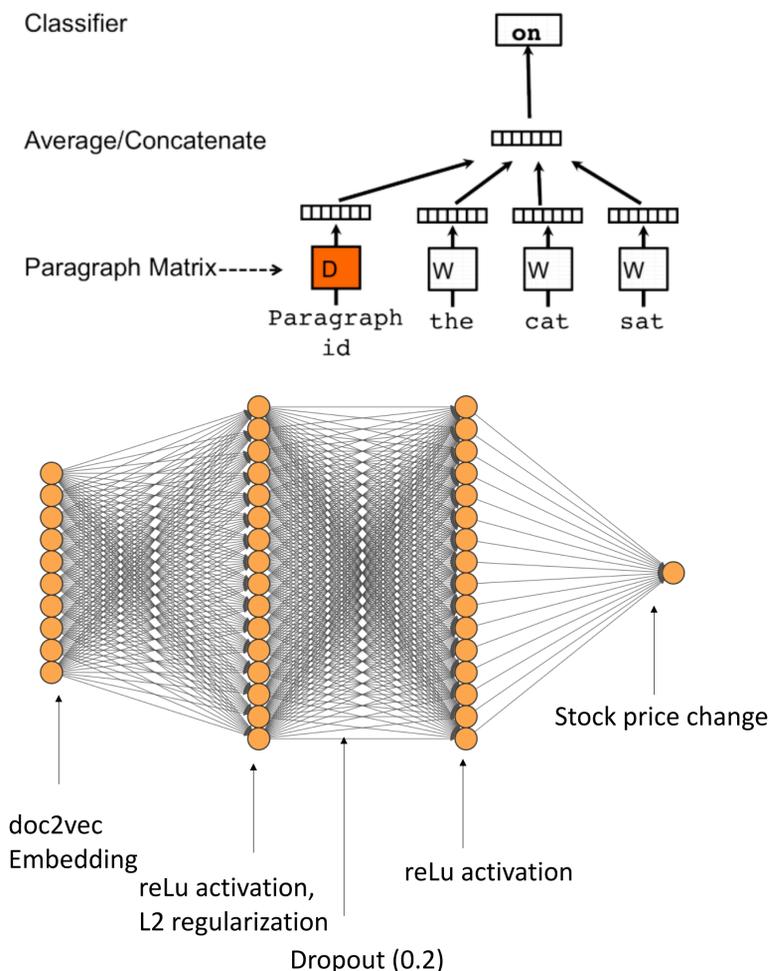
Stock Price Data

- End of day prices obtained from Quandl
- Parameter tuning performed on RUSSELL 3000 index – 3000 largest American companies

Model

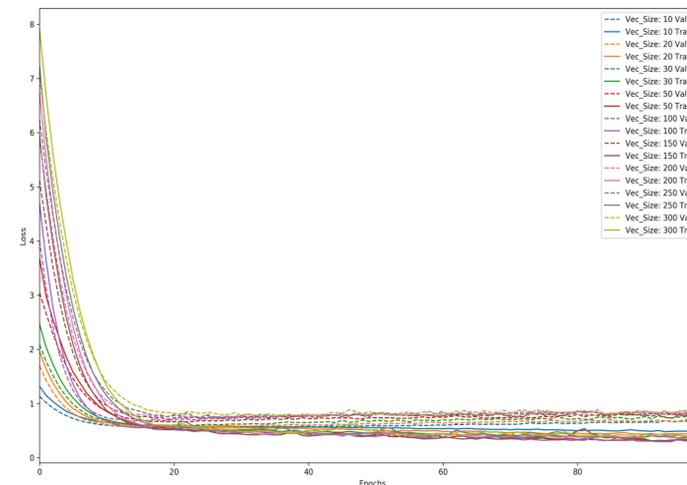
Original diagram from seminal doc2vec paper by Mikolov and Le

Classifier

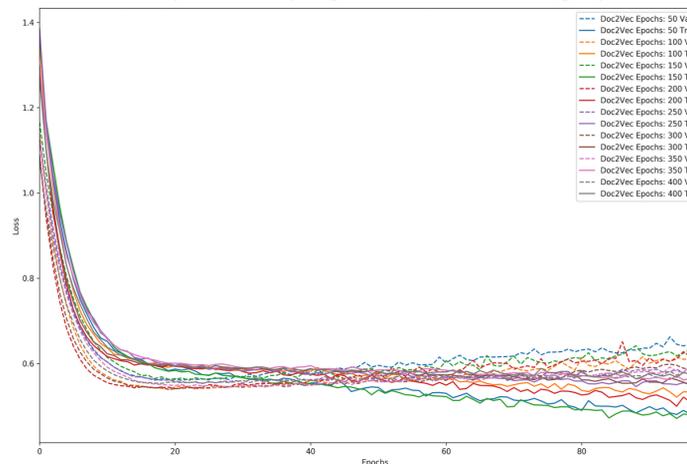


Results

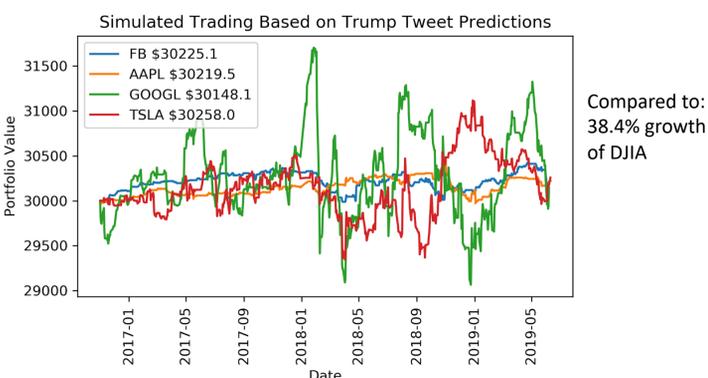
Loss vs Epoch varying Embedding Vector Size



Loss vs Epoch varying doc2vec Training Epochs



Trading Simulations



Analysis/Conclusion

- Smaller embedding vector sizes result in lower loss. Since we are considering only one person's lexicon, larger dimensions likely add more noise.
- Accuracy on up/down are generally a little over 50%. Testing data is evenly split, and predictions are also evenly split (i.e. Google up/down predictions have a 55%:45% distribution)
- Companies show varying responses to tweet events (i.e. 1/1/19—new year and heavy discussion about the wall)

Future Directions

- Members of the theory community have claimed that weighted word2vec embeddings might be a better baseline than doc2vec. The financial research community has shown empirical support for doc2vec
- Explore second or millisecond stock price data as twitter effects may be relevant for a short time window
- Explore effects of model on more complex trading strategies

References

[1] Bollen, J., Mao, H. & Zeng, X. (2010) Twitter mood predicts the stock market. *Journal of Computer Science*, pp. 1-8.

[2] Lutz, Bernhard & Pröllochs, Nicolas & Neumann, Dirk. (2019). Sentence-Level Sentiment Analysis of Financial News Using Distributed Text Representations and Multi-Instance Learning. 10.24251/HICSS.2019.137.

[3] Q. Le and T. Mikolov, "Distributed representations of sentences and documents," in *Proceedings of the 31st International Conference on Machine Learning (ICML)*, pp. 1188–1196, 2014.