Stock2Vec: Optimizing Predictions of Company Characteristics with Stock Embedding
Zhelu Mai, Diya Deepak, Sugam Kafle, Madhav Thamaran, Ting Xiao, Mark V. Albert
Department of Computer Science and Engineering, University of North Texas
G. Brint Ryan College of Business, University of North Texas | Texas Academy of Mathematics and Science

Abstract
Price changes in stocks play a significant role in many ways. Not only can it capture the similarity between companies, but it can also dive deeper into hidden characteristics like weather conditions. For example, hurricanes in a particular region can also impact the stock price changes of companies in that region. Thus, it’s valuable to create an Embedding of company stocks, Stock2Vec, which can be easily added on to and optimize any Prediction Model that applies to companies with associated stock prices. In our work, based on the framework of the previous Stock2Vec paper, we built a fine-tuned Word2Vec and FastText model. Using three different regression models, we predicted a new target variable, the Market Capital of companies, with three different Embedding Models. The experiment results demonstrate that our models achieved at least a 10% improvement in performance.

Methodology
- Utilized the fine-tuned Word2Vec model
- Clusters were less widely distributed and overlapped than existing model
- Better clustered data points than the existing model

Results & Findings
- In the same prediction task (ESG ratings):
  - the fine-tuned Word2Vec model performed best in the task
  - Random Forest Regression got the highest R squared score in the task
- In predicting the Market Capital of SP 500 Companies:
  - the fine-tuned Word2Vec model did a better job in new task
  - Linear Regression reached the worst result and others regressors conducted barely the same

Future Work:
- New input and new embedding models are needed because the Word2Vec and FastText models cannot capture the ordering of the sentence
- More data is needed because only five years of data may not capture the transformation of companies and sectors in the long term
- Fine-tuning the model using a dense layer is welcome

Conclusion & Future Work
To Summarize:
- Compared to the PCA, better clustered data points increased performance
- Random Forest Regression outperformed Linear Regression in predicting ESG and Market Capital,
- Fine-tuned Word2Vec and FastText embeddings work better than the previous model in predicting ESG and Market Capital, meaning they can adapt general tasks.

Acknowledgements