How is Natural Language Processing Utilized to Predict and Comprehend Various Metrics of Equities

Alek Seibel
Department of Computer Science, Lafayette College

INTRODUCTION

What is NLP for Financial Market Applications:
1. The utilization of textual data from sources such as news, social media, internet, financial filings, etc. to extract buy/sell signals, predict price movements, and/or understand volume figures, etc. [1]
2. Numerous hedge funds and proprietary trading firms use NLP, which is embedded in their low-latency trading systems

NLP Considerations for Financial Market Application:
1. Semantic Modeling → How is textual data represented so that the most accurate understanding of the text can be gained? [1]
    1. Bag-of-words (No word order, early approach) vs. Word embeddings (Considers word is context, enables topic modeling) [1].
2. Sentiment Analysis → What is the sentiment polarity of the textual data? [1]
    1. Requires doing aspect extraction, subjectivity detection, named entity recognition, sarcasm detection, personality recognition, user profiling, and multimodal fusion. [1]
    2. A KB of keywords is also used to help with understanding the sentiment. In many cases there are industry publications (i.e. Henry Word List) or moods, Profile of Mood States (POMS). [1]
3. Event Extraction → What are the narrative impacts of the textual data on the stock. [1]

Financial Market Perspectives for Prediction:
1. Connectionist Perspective → Stocks in similar sectors behave similarly [1]
2. Portfolio Management Perspective → Optimization problem (increase returns, decrease risk) [1]
3. Energy System Perspective → Stock movement is due to real world operations of a company [1]
4. Social Network Perspective → Stock movements are really due to mass sentiment [1]

Research Objectives:
1. What are the various methodologies, utilizing NLP, that effectively predict and/or comprehend a stock's price and other key attributes?
2. What are some of the challenges when representing textual data? What are the most effective techniques?

Event-Driven Prediction and Comprehension

What does Event-Driven Mean:
1. Capturing structured relationships with the textual data (i.e., viewing a statement from a holistic perspective)
2. Example [2]:
    1. Microsoft sues Barnes & Noble.
    2. Unstructured → Look at the words individually; What is the polarity of each word.
    3. Structured → Actor = Microsoft, Action = sues, Object = Barnes & Noble. What is the polarity of the statement.

Figure 1: The two plots highlight how certain events affect the pricing volatility of Google's stock. | Source: [2]

Approach 1 – Convolutional Neural Network (CNN):
1. With a CNN, the long and short term affects of an event can be modeled. Additionally, event sparsity can also be accounted for. The CNN performs semantic composition over the event embeddings and delivers a prediction of the stock. [2]
2. This approach, can result in 6% increase in accuracy at predicting the S&P 500 and stocks' time series when compared to other methodologies. [2]
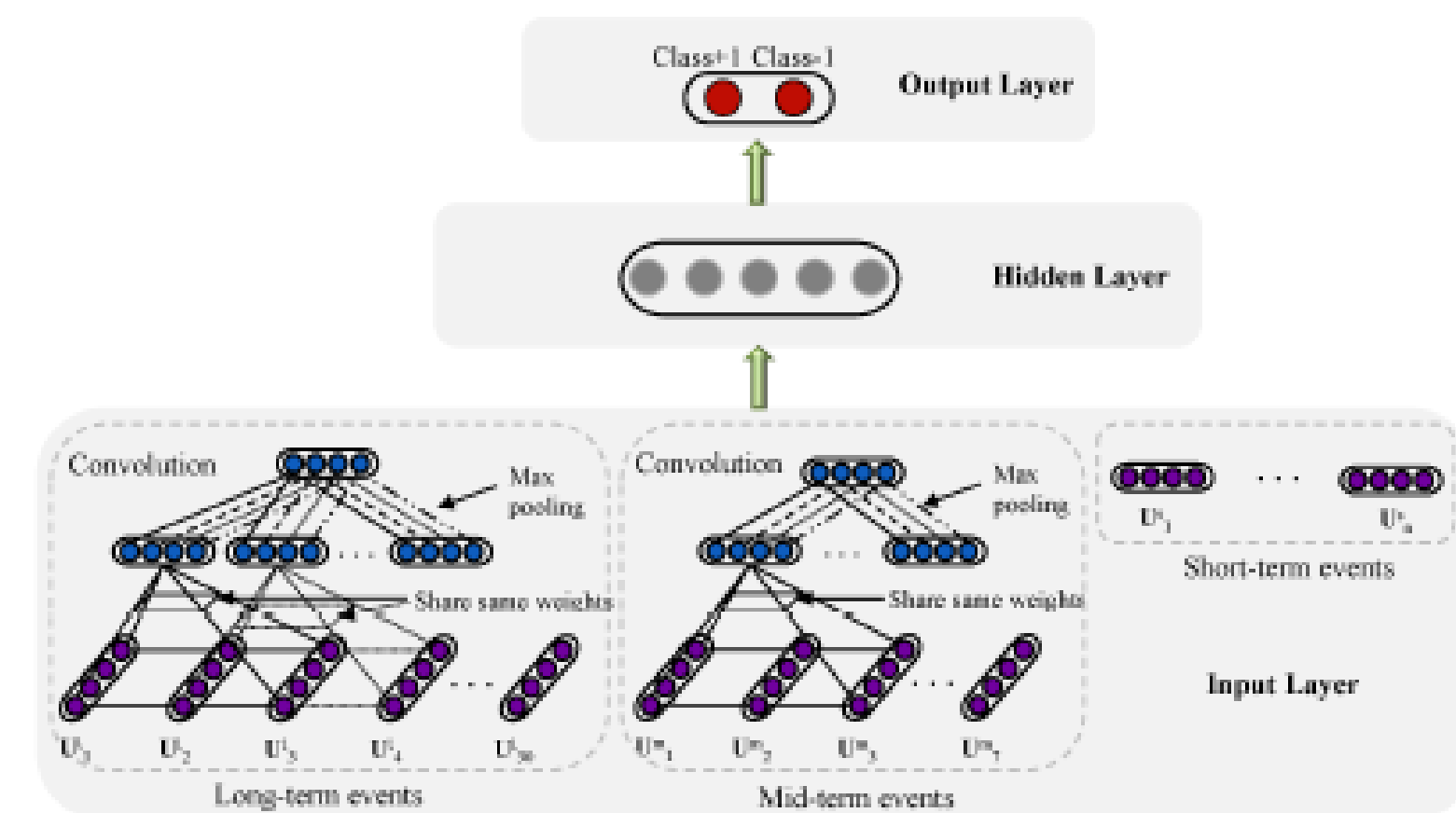

Figure 2: The diagram presents the CNN for Approach 1. The CNN is able to model various events based on their timeframes. | Source: [2]

Approach 2 – Tree Structured Long-Short Term Memory (LSTM):
1. A tree-structured LSTM is used to find a target specific representation of the textual data. In other words, the information content of the text is being measured. With this polarity deciphered, the Cumulative Abnormal Return (CAR) is assigned to the stock to see the economic value that the event has on the stock. [3]
2. This approach results in a more effective CAR calculation and this method can be used for equity screening. [3]
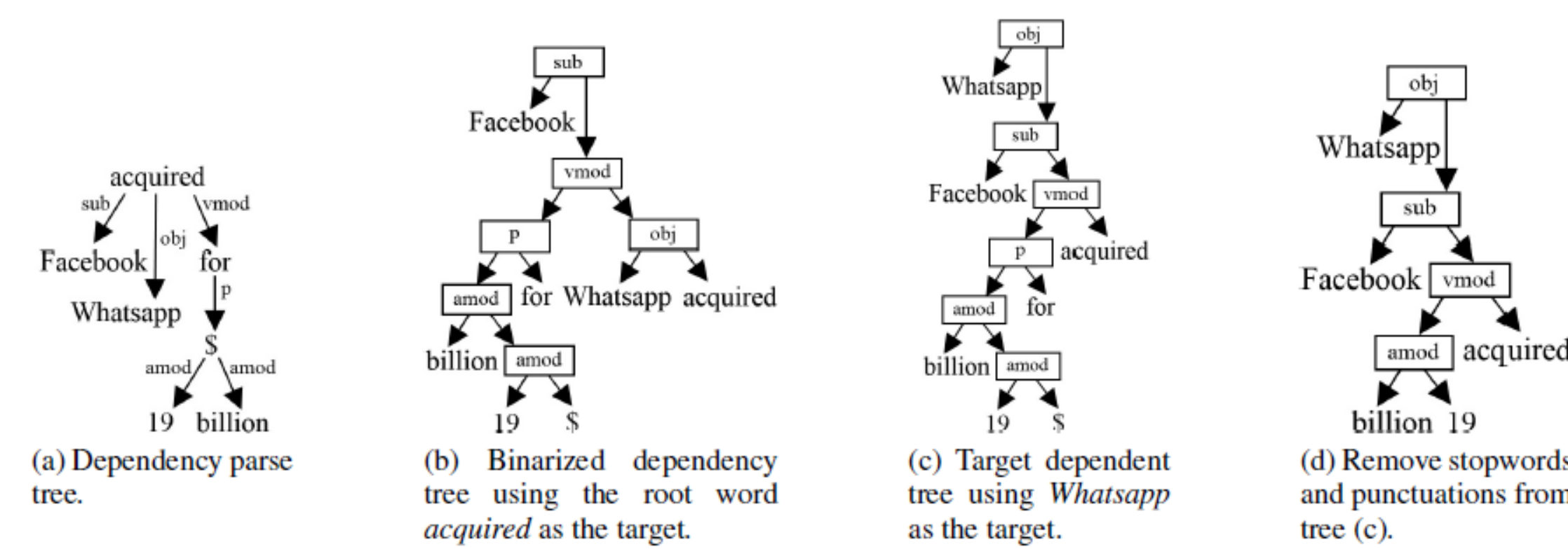

Figure 3: The various trees highlight the transformation steps required to create a binary target specific LSTM tree. | Source: [3]

LAYERED MODELS

Approach 1 – The Financial News and Tweet based Time Aware Network for Stock Trading (FAST) Model:
1. Utilizing just one LSTM, can be ineffective. Therefore, the FAST methodology proposes a layered approach where intra-day and temporal considerations are taken into account, along with adding a timing consideration to the model. [4]
2. FAST performs 8% better in terms of cumulative returns and Sharpe ratio at ranking various equities when compared with current widely used models. [4]
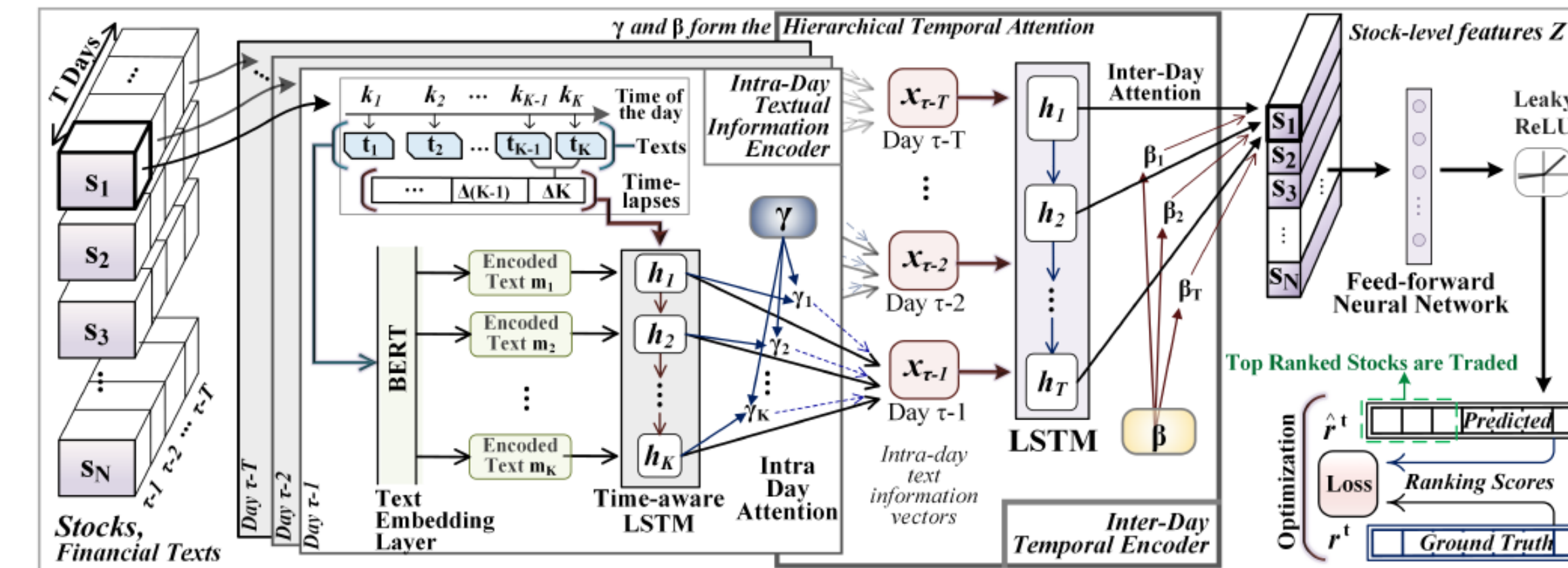

Figure 4: The figure shows a diagram of the FAST model. The model has three key elements. The first being the input of textual data with timestamps. The second being the time aware LSTM. The third being the hierarchal temporal attention LSTM for ranking the stock. | Source: [4]

Approach 2 – The Multi-Layer Perception (MLP) Model:
1. The MLP model combines the outputs of four different models. A CNN is used to understand the regression characteristics of textual data. A LSTM is used to understand the long-term dependences of the textual data. A Gated Recurrent Unit (GRU) is used similar to the LSTM; however, it is much more computationally efficient. Finally, a Support Vector Regression (SVR) is used for predicting the sentiment score of the textual data. [5]
2. This model is successful at outperforming other social sentiment score models, particularly when the textual data came from microblogs and news headlines and when it comes to bearish and bullish signals. [5]
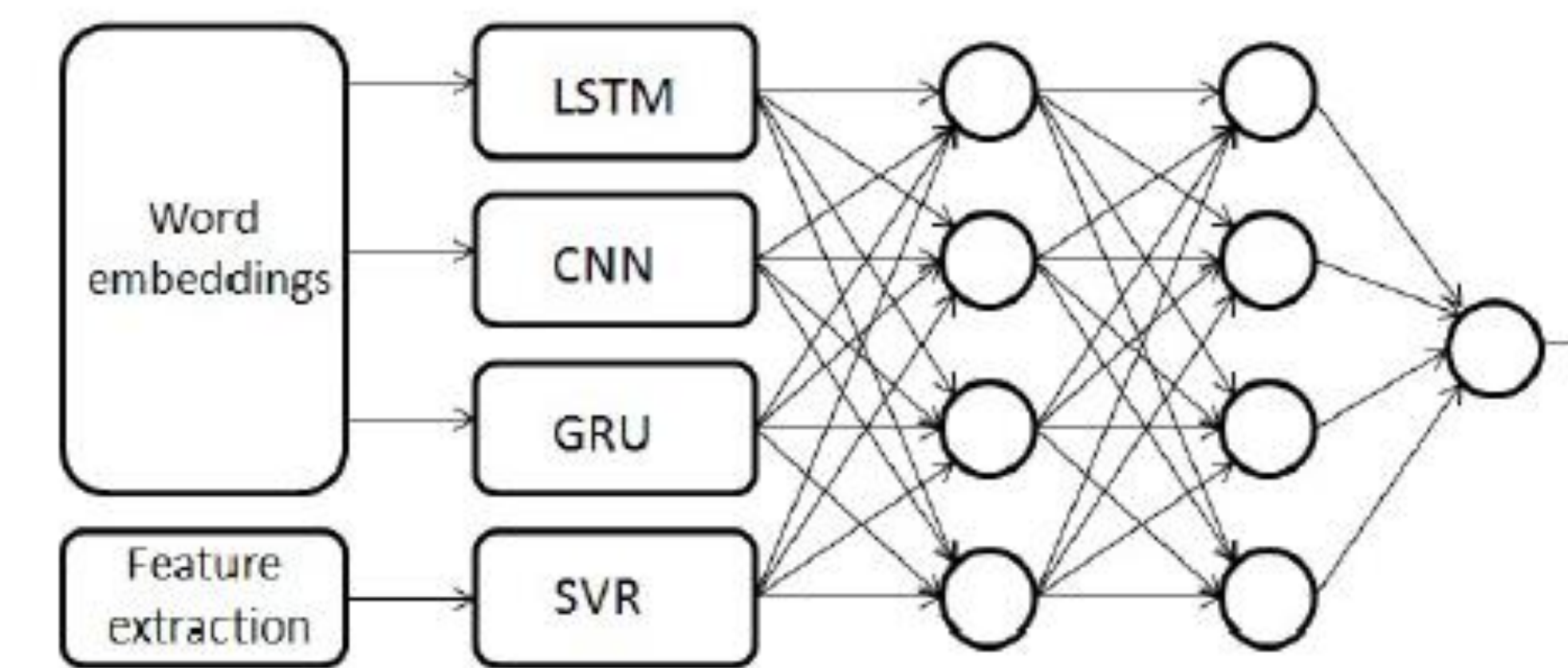

Figure 5: The figure shows the diagram from the MLP model. | Source: [5]

NETWORK-BASED REPRESENTATION

What is a Stock Network:
1. A graph structure is used to represent the relationships between various investment instruments, with nodes being instruments and edges being connections. The most common stock network is a Correlation Stock Network (CSN), which runs a Pearson correlation function over the time series of a stock and connects them based on this correlation value. [6]

Sentiment Stock Network (SSN):
1. Textual data can be used to associate various stocks with each other. This network is called a Semantic Stock Network (SSN). With this structure, sentiment from close neighbors can be evaluated to inform a time series prediction of the stock under evaluation. [6]
2. This methodology performs much better for time series prediction than CSNs as CSNs can sometimes consider unrealistic relationships. [6]
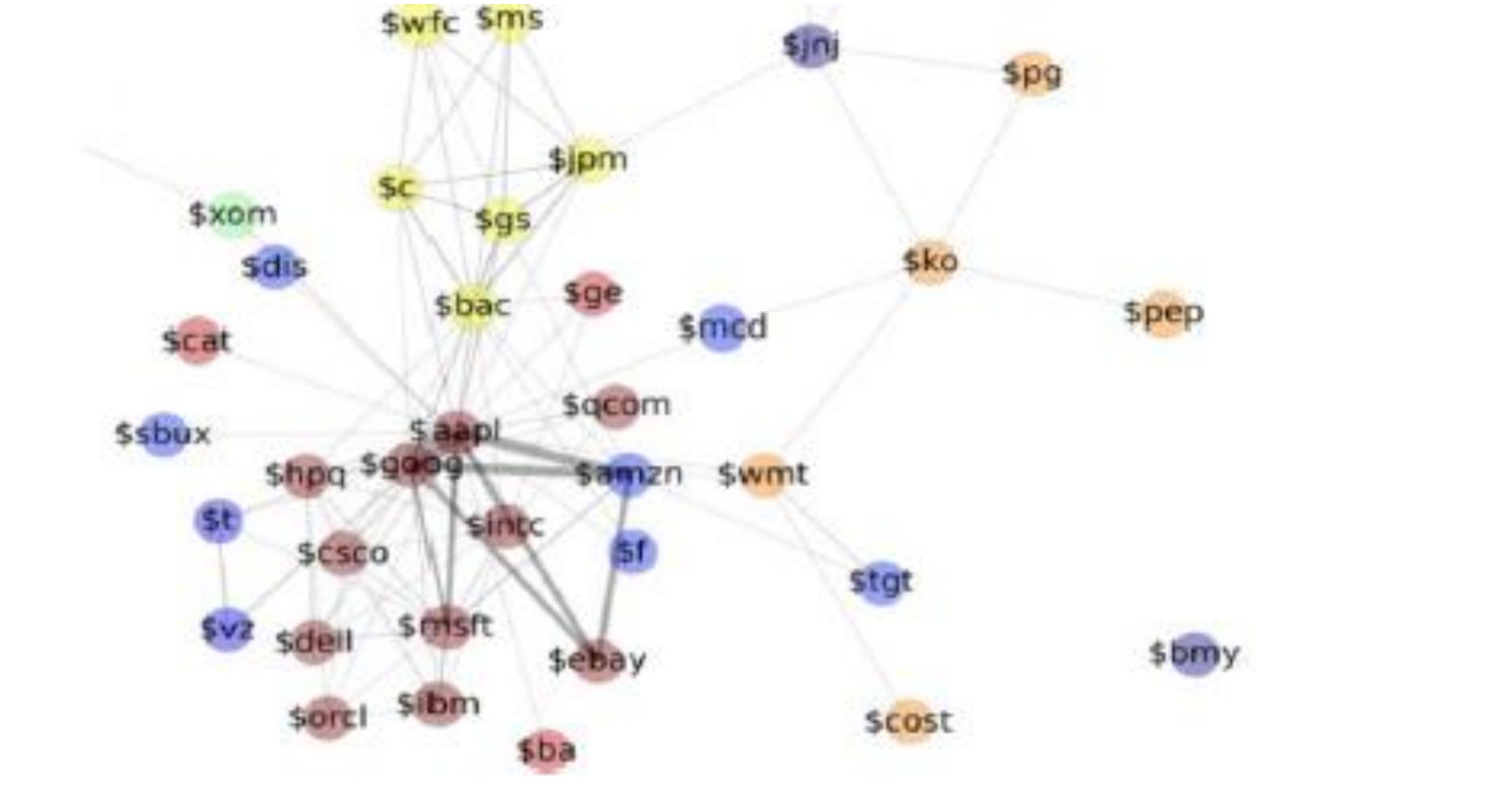

Figure 6: An example of a Sentiment Stock Network (SSN). | Source: [6]

CONCLUSION

1. There are numerous ways to go about utilizing NLP for financial markets
    1. Event-Driven
    2. Layered Models
    3. Network-Based
2. Different methods succeed at different tasks (i.e., Layered Model → bullish and bearish signaling)
3. It should be noted, however, that just using NLP to trade and/or invest in stocks is a poor idea as there are many more items that needed to be taken into consideration for a strategy to be robust and ideally successful.
4. While much work has been done to see how NLP can be used for stocks, derivative products offer another opportunity for this technology to be used; albeit, in a much more complex environment. [1]

BIBLIOGRAPHY

[1] Xing, F.Z.; Cambria, E.; and Welsch, R.E. 2017. Natural language based financial forecasting: a survey. *Artificial Intelligence Review* 50: 49-73.

[2] Ding, X.; Zhang, Y.; Ting, L.; and Junwen, D. 2015. Deep Learning for Event-Driven Stock Prediction. In Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, 2327-2333. International Joint Conference on Artificial Intelligence.

[3] Chang, C.Y.; Zhang, Y.; Teng, Z.; Bozanic, Z.; and Ke, B. 2016. Measuring the Information Content of Financial News. In Proceedings of the 26th International Conference on Computational Linguistics, 3216-3225. International Conference on Computational Linguistics.

[4] Sawhney, R.; Wadhwa, A.; Agarwal, S.; and Shah, R.R. 2021. FAST: Financial News and Tweet Based Time Aware Network for Stock Trading. In Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics, 2164-2175. Conference of the European Chapter of the Association for Computational Linguistics.

[5] Akhtar, M.S.; Kumar, A.; Ghosal, D.; Ekbal, A.; and Bhattacharyya, P. 2017. A Multilayer Perception based Ensemble Technique for Fine-grained Financial Sentiment Analysis. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 540-546. Conference on Empirical Methods in Natural Language Processing.

[6] Si, J.; Mukherjee, A.; Liu, B.; Pan, S.J.; Li, Q.; and Li, H. 2014. Exploiting Social Relations and Sentiment for Stock Prediction. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, 1139-1145. Conference on Empirical Methods in Natural Language Processing.