Stock Price Time Series Forecasting Using Dynamic Graph Neural Networks and Attention Mechanism in Recurrent Neural Networks

Marco Gregnanin[1,2], Johannes De Smedt[2],
Giorgio Gnecco[1], Maurizio Parton[3]

1 - IMT School for Advanced Studies Lucca, Italy

2 - KU Leuven, Belgium

3 - University of Chieti-Pescara, Italy

22 September 2023

Outline

Overview

- Forecasting future stock prices is a challenging problem in financial markets.
- Classical statistical models encounter difficulties due to the presence of noise and non-linearities.
- Deep learning models show ability to capture and exploit non-linearities within the data.

Why do we require geometric patterns?

- Transforming the data in a non-Euclidean space can aid in capturing and analyzing inherent complexity.
- Identifying fractal series: self-similarity and repetition at different scales.

Objective

- Find a function that can predict the future q values $s_{t+1:t+q}$ of a univariate target time series:

$$s_{t+1:t+q} = f(G_t, X_t), \tag{1}$$

where G_t denotes the graph representation of the univariate target time series, and X_t is the feature matrix.

Visibility Graph Algorithm

- Idea: Connect data points if there is a clear line of sight between them.

- Visibility criterion: Given a time series $S = \{s_1, s_2, \ldots, s_T\}$ and the set of associated discrete time stamps $\mathcal{T} = \{t_1, t_2, \ldots, t_T\}$, check the following condition:

$$s_k < s_j + (s_j - s_i)\frac{t_j - t_k}{t_j - t_i}\,. \tag{2}$$

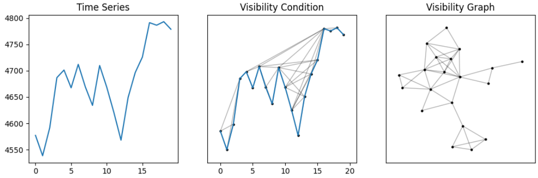
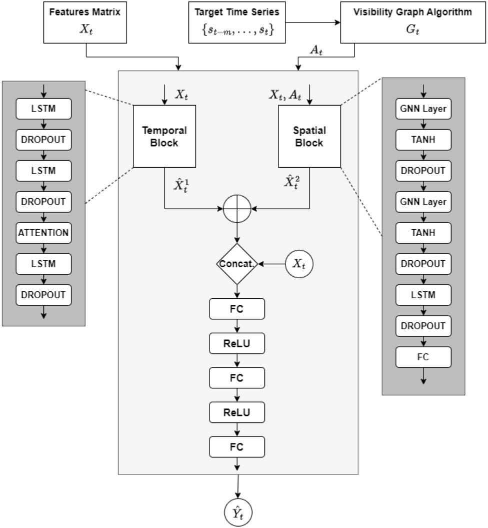


Figure: Visibility graph method on the Standard and Poor's 500 closing price from 02/12/2021 to 31/12/2021.

Temporal Attention - Dynamic Graph Neural Network (TA-DGNN) Architecture



Temporal Block

Input: X_t.

- Two LSTM layers: capture temporal and long-term patterns;
- Attention layer: extraction of important features;
- LSTM layer: processing of the patterns identified by the attention layer.

Output: $\hat{X}_t^1$.

Spatial Block

Input: X_t and A_t.

- GNN layers: capture information from neighboring nodes that are 1-hop away;
- GNN layers: capture information from nodes that are 2-hops away from the original nodes;
- LSTM layer: processing of the temporality of spatial patterns.

Output: $\hat{X}_t^2$.

Fully Connected Layers

Input: $\hat{X}_t^1$, $\hat{X}_t^2$, and X_t.

- FC layer: 128 cells
- FC layer: 64 cells
- FC layer: q cells

Output: $\hat{Y}_t$

Dataset

- Standard and Poor's 500 (S&P500).

- Collect "Close, Open, High, Low" daily prices from Monday 4th January, 2010 to Sunday 4th June, 2023, resulting in $3,377$ observations.

- We can reframe the regression problem $(s_{t+1:t+q} = f(G_t, X_t))$, by setting q equal to 1 and m equal to 20:

$$s_{t+1} = f(G_t, X_t)\,.$$

- We normalize the data based on a sliding window of 20 days:

$$\tilde{x}_t = \frac{x_t - \mu_t}{\sigma_t}\,.$$

- We split the data in Train-Validation-Test sets: $70 - 10 - 20\%$.

Configuration of the Models

- Baseline models: vanilla GCN augmented with an LSTM layer, BiLSTM model, AT-LSTM model, and ARIMA model.
- Loss function: Mean Square Error (MSE).
- Hyperpameter setting: batch size 25, number of epochs 500, learning rate 0.0001, dropout rate 0.1.
- Evaluation metrics: Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Mean Absolute Scaled Error (MASE):

$$\mathrm{MASE} = \frac{\frac{1}{l}\sum_{i=1}^{l}|y_i - \hat{y}_i|}{\frac{1}{l-1}\sum_{i=2}^{l}|y_i - y_{i-1}|}.$$

First Analysis

Table: Model performance in terms of MSE, RMSE, MAE, MAPE and MASE for the test set. The best value for each metric is in bold.

Models	MSE	RMSE	MAE	MAPE	MASE
ARIMA(3, 0, 2)	2269.7678	47.64208	36.0789	0.8886 %	1.0137
GNN+LSTM	4742.4291	68.8653	55.5719	1.3396 %	1.5624
BiLSTM	9361.4256	96.7545	78.4831	1.9301 %	2.2066
AT-LSTM	1254.2723	35.4157	22.3470	0.5547 %	0.6276
TA-DGNN (GCN)	**123.4954**	**11.1128**	**7.6060**	**0.1881** %	**0.2136**
TA-DGNN (ChebConv)	216.5798	14.7166	9.5463	0.2371 %	0.2681
TA-DGNN (SAGE)	127.2517	11.2806	7.7775	0.1925 %	0.2184

Plot of the results

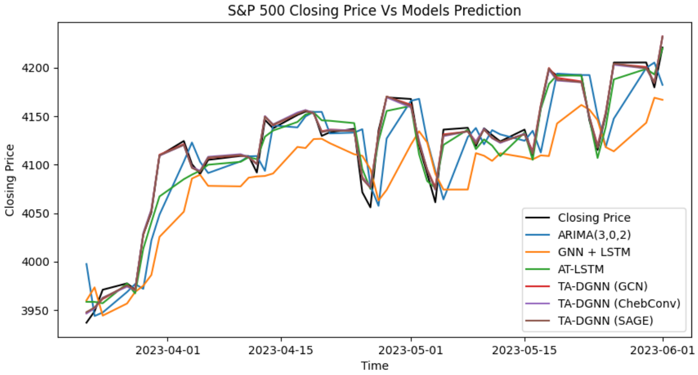


Figure: Comparisons among the models' predictions and the S&P 500 closing price from 22/03/2023 to 01/06/2023.

Second Analysis: Out-of-Sample Test

- We collected daily observations of the S&P500's Close, Open, High, and Low from Monday 1st October, 2007 to Tuesday 1st December, 2009, resulting in 547 observations.

Table: Model performance in terms of MSE, RMSE, MAE, MAPE and MASE for the out-of-sample analysis. The best value for each metric is in bold.

Models	MSE	RMSE	MAE	MAPE	MASE
AT-LSTM	28.3803	5.3273	4.2145	0.4168 %	0.4682
TA-DGNN (GCN)	**5.5232**	**2.3501**	1.7421	0.1715 %	0.1935
TA-DGNN (ChebConv)	7.8564	2.8029	1.9653	0.1949 %	0.2183
TA-DGNN (SAGE)	5.7383	2.3955	**1.7336**	**0.1710** %	**0.1926**

Plot of the results

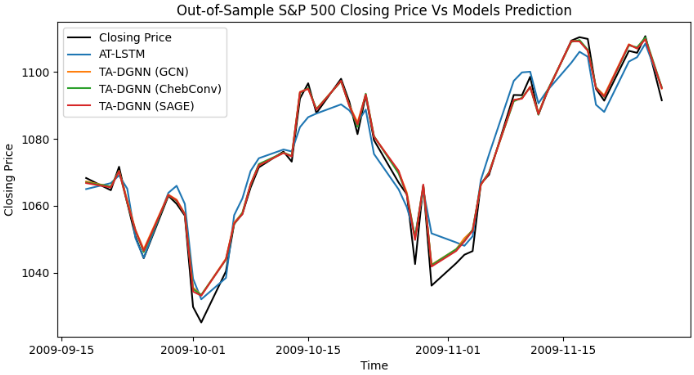


Figure: Comparisons among the models' predictions and the S&P 500 closing price from 19/09/2009 to 27/11/2009.

Conclusion

- We propose a model that combines DGNNs with LSTM networks enhanced with an attention mechanism.
- Evaluate the performance in predicting one-step ahead closing price of the S&P500.
- We propose to consider also the MASE metric.
- We conduct two analysis to assess the robustness and stability of our proposed model.

Future Research

- Exploring additional features or modify the existing features.
- Adding a mechanism to track and update the hidden layers.
- Studying the performance of the TA-DGNN model in a trading context.

THANKS FOR YOUR ATTENTION