Optimized Informer-based Stock Price Prediction with Feature Connection Mechanism and Multi-level Context Aggregation

Group 7: 余忠旻、宋沛潔、黃靖雅

Introduction

In modern financial markets, accurate prediction of stock price movements is a great importance. Therefore, we proposed a novel forecasting method, which based on the paper “Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting.” With our proposed method of Feature Connection Mechanism (FCM) and multi-level context aggregation, the MAPE and MSPE can be reduced by 0.03 and 0.007, respectively.

Dataset

The dataset for our analysis is sourced from Yahoo Finance, and it focuses on four companies that represent different sectors

- JPMorgan (JPM) for Finance
- Drive Shack (DS) for Entertainment
- ExxonMobil (XOM) for Energy
- Pfizer (PFE) for Pharmaceuticals

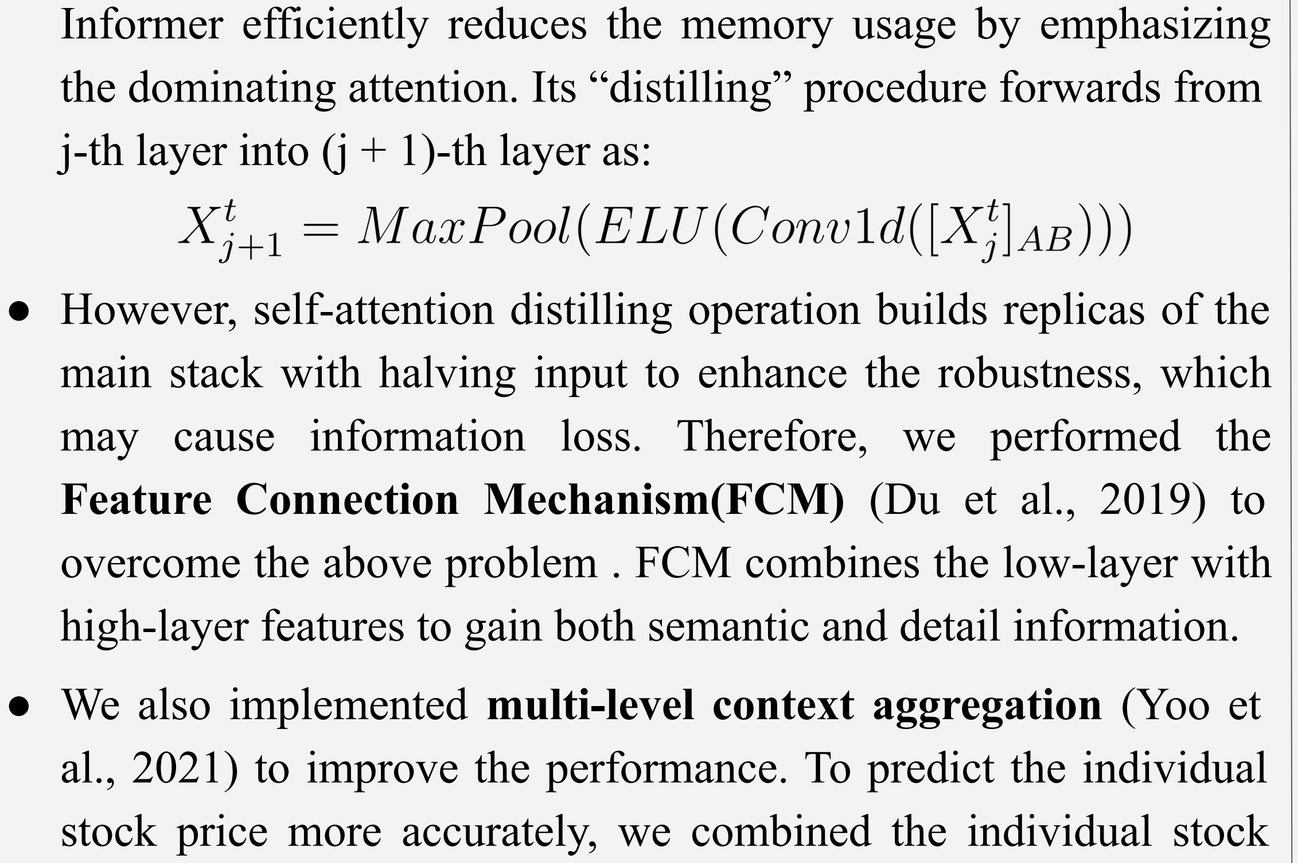
Additionally, the NASDAQ Composite, a broad market index, was selected to provide a macro perspective on the overall stock market performance.

Method

- Informer has shown it great ability of long sequence time series forecasting. The proposed **self-attention distilling** operation in Informer efficiently reduces the memory usage by emphasizing the dominating attention. Its “distilling” procedure forwards from j-th layer into (j + 1)-th layer as:

$$X_{j+1}^{t} = MaxPool(ELU(Conv1d([X_j^t]_{AB})))$$

- However, self-attention distilling operation builds replicas of the main stack with halving input to enhance the robustness, which may cause information loss. Therefore, we performed the **Feature Connection Mechanism(FCM)** (Du et al., 2019) to overcome the above problem . FCM combines the low-layer with high-layer features to gain both semantic and detail information.
- We also implemented **multi-level context aggregation** (Yoo et al., 2021) to improve the performance. To predict the individual stock price more accurately, we combined the individual stock context with market stock like NASDAQ or another individual stock context as background knowledge of stock correlations.



Results

1. With four methods to analyze stock trends (DS stock):
Sequence length : 64, token length: 40, forecast the next 5 trading days.

	Informer no distilling	Informer no distilling + FCM	Informer	Informerstack	Informer + FCM
Time (sec)	158	199	120	221	134
Memory	67%	77%	51%	67%	54%
MAPE / MSPE	0.08/0.01	0.07/0.01	0.13/0.02	0.08/0.01	0.07/0.01

★ **Table1**

2. With different lengths and methods (DS stock):

Seq_len/token_len/predict_len		64/40/5	64/60/15	128/60/15	128/80/25
Informer no distilling	Time (sec)	158	170	-	-
	Memory	67%	72%	-	-
	MAPE / MSPE	0.08/0.01	0.18/0.04	-	-
Informer no distilling + FCM	Time (sec)	199	212	-	-
	Memory	77%	79%	-	-
	MAPE / MSPE	0.07/0.01	0.16/0.03	-	-
Informer	Time (sec)	120	131	174	187
	Memory	51%	55%	72%	78%
	MAPE / MSPE	0.13/0.02	0.23/0.08	0.19/0.05	0.18/0.05
Informerstack	Time (sec)	221	233	287	298
	Memory	67%	70%	87%	94%
	MAPE / MSPE	0.08/0.01	0.19/0.05	0.15/0.03	0.15/0.04
Informer + FCM	Time (sec)	134	145	192	205
	Memory	54%	56%	80%	82%
	MAPE / MSPE	0.07/0.01	0.15/0.03	0.14/0.03	0.14/0.03

★ **Table2**

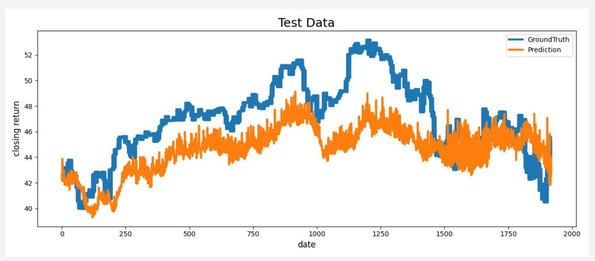
From Tables 1 and 2, we found that adding FCM to the Informer reduced MAPE and MSPE compared to using only the Informer without distilling, but at the cost of higher time and memory usage. Using Informer with distilling reduces time and memory cost but lowers accuracy. Hence, the Informer + FCM is optimal, improving accuracy and reducing memory usage.

3. We made a comparative analysis between the stocks and the overall market index.
Sequence length : 64, token length: 40, forecast the next 5 trading days.

4. We predicted JP Morgan's (JPM) price using Procter & Gamble (P&G) data and a negative beta (-0.1197), reflecting the inverse relationship between stable and high-return stocks.
Sequence length : 64, token length: 40, forecast the next 5 trading days.

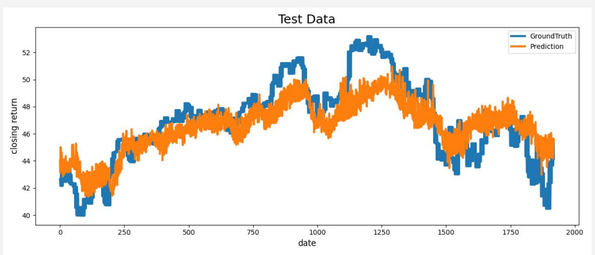
★ Predict XOM

★ Predict JPM



★ Predict XOM with NASDAQ

★ Predict JPM with P&G



Conclusion

- We implemented a proposed informer-based model with FCM to gain more fine-grained information to predict stock prices.
- Also, our proposed multi-level context aggregation mechanism, which incorporates market stock as background knowledge, leads to better forecasting performance of our model.