CS 230

Visualization and Prediction for a High Frequency Limit Order Book

Alan Dunetz

Poster Talk: https://youtu.be/GRJ9jJx5--Y

Abstract

Four Deep Learning architectures were applied to a mid-price change prediction problem for a high frequency Limit Order Book (LOB) dataset:

- A Wavenet-type architecture [3] employing dilated 1D CNNs with causal padding significantly outperformed other architectures.
- The addition of expert hand-crafted features improved performance.
- Performance for a single stock was improved by adding LOB data from other stocks.
- A variety of visualization techniques were employed to provide insight into what our networks learned.

Background

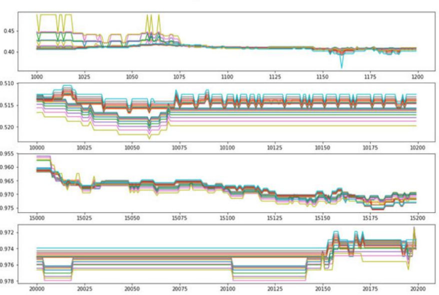
LOBs aggregate all orders on an exchange to buy and sell a security at different prices. The LOB therefore provides a snapshot of the market's cumulative demand to buy and sell securities. Imbalances between bid and offer order sizes, or sudden changes in order sizes or prices, for example, may be informative about future price direction.

Data Set

We used the FI2010 public dataset described in [1]:

- 10-days of limit order book data from June 2010 for five stocks that trade on the Helsinki exchange.
- Each record in the time series shows prices and aggregate order sizes for the first ten levels on each side (bid and offer) of the market
- The total number of messages (i.e. arriving buy/sell orders or cancels) reflected in the time series is approximately four million.
- The dataset includes an orderbook snapshot after every 10 messages resulting in approximately 400,000 records in total for the five stocks.
- On average, there is one snapshot for approximately every one-half second, though timing varies based on the level of market activity.

Sample LOB Data



Models

Four model architectures were compared (details in paper/github):

- LSTM: LSTM layer-> DO -> Softmax.
- CNN: 3 Conv layers-> MP-> Conv layer -> MP->FC-> Softmax.
- CNN-LSTM: Same as CNN model with LSTM layer replacing FC layer.
- Wavenet-type: 5 dilated CNN layers with causal padding -> FC->DO->Softmax

Key parameters tuned: learning rate, batch size, sequence length, regularization methods, layer size and number of layers.

RELU activations were used in most cases. 100 epochs of training were applied.

Methodologies

- Model inputs consisted of sequences (sliding windows) of LOB feature data.
- Target was smoothed 10 period prediction category from last observation in sequence. Category labels were 38.4% up; 24.7% unchanged; 36.9% down.
- Models were tuned using days 1-6 for training and days 7-8 for validation.
- Models were tested using days 1-8 for training and days 9-10 for testing.
- Metrics: Loss, Accuracy, F1 and Cohen's Kappa.
- Loss Function: Categorical cross entropy . Adam was used for optimization.

Prediction

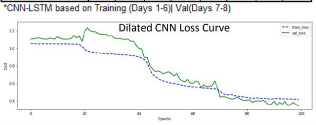
Notes

- CNNs required low learning rates (<=.001); small batch sizes (<=50); and would benefit from longer training.
- LSTM model tended to over fit despite regularization.
- Adding LSTM layer improved CNN model.
- Performance of Wavenet-type model warrants additional study/scrutiny.

Training(Days 1-8) | Test (Days 9-10)

Model	Loss	Acc	F1	Kappa
LSTM	0.71|0.90	0.71|0.64	0.67|0.64	0.50|0.46
CNN	0.88|0.88	0.60|0.62	0.62|0.62	0.41|0.42
CNN-LSTM*	0.72|0.79	0.70|0.65	0.69|0.65	0.51|0.47
Dilated CNN	0.42|0.34	0.85|0.89	0.89|0.89	0.83|0.84

*CNN-LSTM based on Training (Days 1-6)| Val(Days 7-8)



Hand-crafted Features Experiment: (using CNN model):

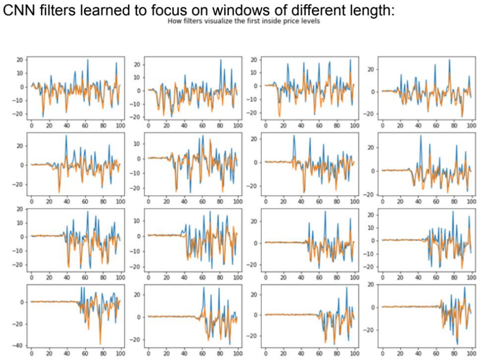
- Deep LOB features were replaced with expert-designed features [4] and results were compared with baseline model using all 10 LOB layers.
- Adding expert designed features improved test performance:
- F1 increased by 14.5%. Kappa increased by 21.4%.

Transfer Learning Experiment (using LSTM model):

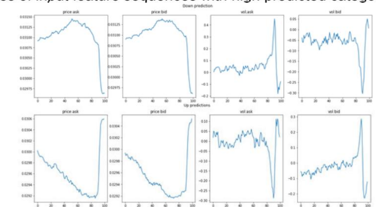
- Model trained on all 5 stocks (5->1 model) was compared to 5 models trained on each stock separately (1->1 models).
- Test data set prediction performance for each individual stock was averaged for the two classes of models.
- 5->1 model outperformed 1->1 models: F1: +7.65%; Kappa: +49%.

Visualization

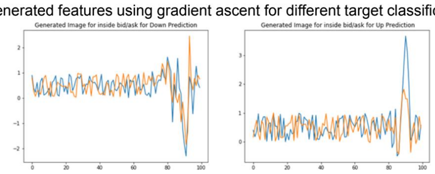
- CNN filters learned to focus on windows of different length:



- Averages of input feature sequences with high predicted category values:



- Generated features using gradient ascent for different target classifications:



References

1. Ntakaris, Benchmark Dataset for Mid-Price Forecasting of LOB.
2. Tsasntekidis, Forecasting Stock Prices from the LOB using CNN.
3. Borovykh, Conditional time series forecasting with CNN.
4. Kercheval, Modeling High-Frequency LOB Dynamics with SVM.

Future Work

- Outperformance of Wavenet-type model warrants additional study.
- Explore higher dimensional visualization, such as t-sne.
- Apply models to larger LOB data sets.