Focusing at High Frequency

An Attention-based Neural Network for Limit Order Books

Andrea Barbon (USI Lugano, SFI)

USI Brown Bag Seminar, Lugano, October 2019

Introduction

- Rise of Artificial Intelligence (AI) and Machine Learning (ML)
    - Unprecedented availability of data
    - Increase in computational power (GPUs, Cloud computing)
    - Surge of model-architecture research in computer science
- Successful applications to virtually any field
    - Image and speech recognition
    - Medical diagnosis
    - Sentiment analysis
    - Recommendation systems
    - ...

Machine Learning in Finance

- In finance, ML is very useful for practitioners:
  - Time-series forecasting
  - Robo advisory
  - Credit scoring
  - Fraud detection
  - Algorithmic trading

- Not so good for researchers:
  - Results are difficult to interpret
  - Impossible to make inference
  - Risk of over-fitting because of limited data

This Paper

- Proposes an application of ML to academic finance
- Novel neural network model with attention mechanism
  - Improve interpretability of the model
  - Allows to make inference
  - Investigate information transmission in equity markets
- Train on high-frequency data from limit order books (LOB)
  - Limit over-fitting problems
  - Reproduce tools used by high-frequency traders (HFTs)
  - Analyze behavior when facing institutional block orders

Related Literature

- Machine Learning and Limit Order Books
  - Sirignano and Cont (2018)
  - le Calvez and Cliff (2018)
  - Zhang, Zohren and Roberts (2019)

- Attention Mechanism
  - Bahdanau, Cho and Bengio (2014)
  - Zhou et al. (2016)
  - Vaswani et al. (2017)

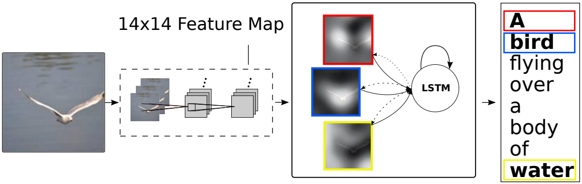
- High-Frequency Trading and Liquidity
  - Hendershott and Menkveld (JF, 2011)
  - Foucault, Kadan and Kandel (JF, 2013)
  - Chaboud, Chiquoine, Hjalmarsson and Vega (JF, 2014)
  - van Kervel and Menkveld (JF, 2019)

LOBster High-Frequency Data

- NASDAQ limit order book (LOB) data from LOBster
- Order book events
  - Limit orders submission / update / cancellation
  - Trade executions (market orders)
  - Ask / Bid price and share volume
- High-frequency observations (millisecond precision)
- Most liquid stocks traded in the NASDAQ exchange
- Long sample period (2009 to 2019)

Visual Attention

Y. Bengio et al. (2016) – Show, Attend and Tell: Neural Image Caption Generation with Visual Attention



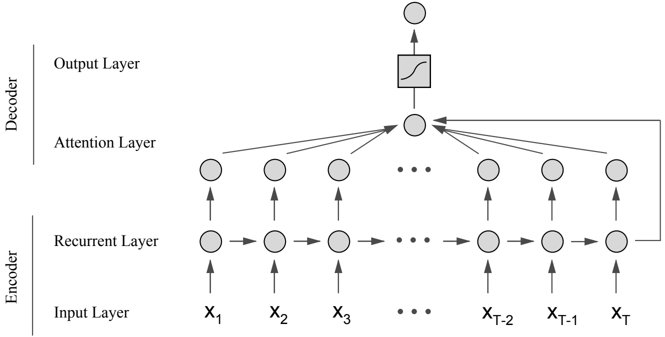
1. Input Image
2. Convolutional Feature Extraction
3. RNN with attention over the image
4. Word by word generation

A stop sign is on a road with a mountain in the background.

A large white bird standing in a forest.

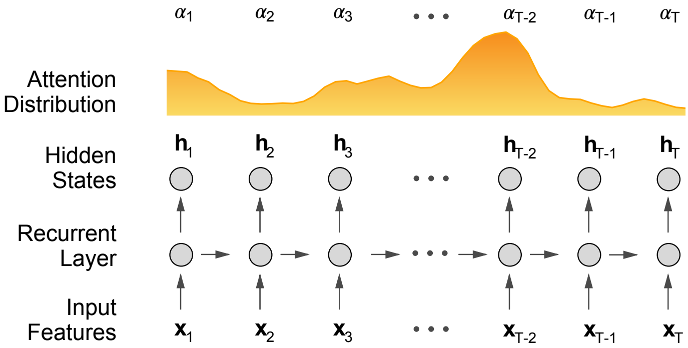
Model Architecture: RNN + Temporal Attention

- Deep neural network with temporal attention mechanism
- Implemented in python using the TensorFlow framework



- Recurrent layer: LSTM – Hochreiter and Schmidhuber 1997
- Attention layer: Bahdanau et al. 2014 (machine translation)

Temporal Attention



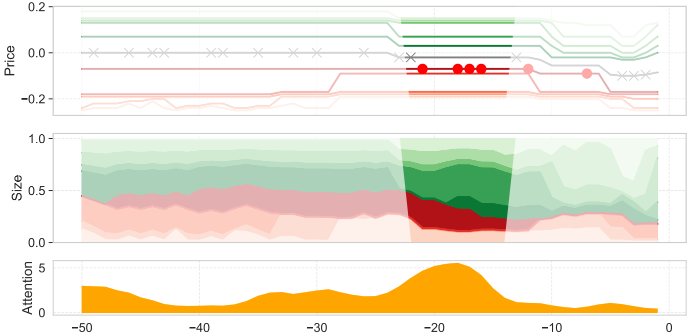
$$\mathbf{e} = \mathbf{v}^\top \tanh(\mathbf{W}^a \mathbf{h}_T + \mathbf{U}^a \mathbf{h}_{T,T-1} + \mathbf{b}^a)$$

$$\boldsymbol{\alpha}_t = \frac{\exp(\mathbf{e}_t)}{\sum_{s=1}^{T} \exp(\mathbf{e}_s)}, \quad t = 1, \ldots, T$$

vectors $\mathbf{v}$, $\mathbf{b}^a$ and matrices $\mathbf{W}^a$, $\mathbf{U}^a$ are parameters to be learned

Attention Distribution

- Visualize the attention levels for a specific example



- Attention levels reveal where the model is focusing
- Useful to interpret and supervise the model's decisions

Features and Target Variable

- **Universe:** 10 among the most liquid NASDAQ stocks
- **Sample period:** 2 weeks, from 2/22/2019 to 3/7/2019
- **Input vectors:** previous 50 LOB events
- **Features:** first 5 levels of the LOB:
  - Bid and ask prices and sizes
  - Elapsed time since last event
  - Percentage mid-price change since last event
- **Target variable:** expected profits for a market maker (MM) providing liquidity to a sell market order
  - The MM buys at the bid price $b(t)$, then she tries to unload during $\pi = [t+1, \dots, t+100]$
  - If there is at least one buy trade in π $\Longrightarrow$ the MM sells at the **minimum ask price** $\min_{s \in \pi} a(s)$
  - If there is no buy trade in π $\Longrightarrow$ the MM sells at the **last bid price** $b(t+100)$

Training and Cross-Validation

- **Loss Function:** Mean Squared Error (MSE)

$$MSE = \frac{1}{N}\sum_{i=i}^{N}(y_i - \hat{y}_i)^2$$

- **Minimizer:** Stochastic Gradient Descent + Adam Optimizer
  - Each epoch with 50 batches of 1024 samples
  - Same random seed for all benchmark models
- **Cross-Validation:**
  - Training set: the first 4 days
  - Validation set: the 5th day
  - Out-of-sample test set: the last 5 days
- **Early-stopping rule:** limits over-fitting (patience $= 5$ epochs)
- **Machine:** Google Cloud server with a NVIDIA Tesla P4 GPU

LOB Events – Summary Statistics

Stock	Ticker	Submissions per Minute	Deletions per Minute	Executions per Minute
American Airlines	AAL	438.2	449.7	16.1
Apple	AAPL	1074.9	1053.8	62.1
Adobe	ADBE	264.2	245.2	24.0
Comcast	CMCSA	583.4	621.0	14.2
eBay	EBAY	499.6	525.4	13.7
FaceBook	FB	643.8	616.5	65.4
Microsoft	MSFT	1472.8	1532.8	53.4
Nvidia	NVDA	378.5	362.7	49.4
Pesico	PEP	304.0	306.0	20.2
Tesla	TSLA	256.5	214.4	75.8

- More than 47 million observations
- More than 500 events per minute, on average

Experimental Results

Root Mean Squared Error: $RMSE = \left(\frac{1}{N}\sum_{i=i}^{N}(y_i - \hat{y}_i)^2\right)^{1/2}$

	(1) Attention	(2) No Attention	(3) DeepLOB*	(4) Neural Net
AAL	**3.334**	3.350	3.346	3.464
AAPL	**0.986**	0.988	0.989	1.063
ADBE	**3.044**	3.123	3.095	3.111
CMCSA	1.661	**1.658**	1.668	1.767
EBAY	**2.007**	2.055	2.018	2.151
FB	**1.599**	1.607	1.604	1.621
MSFT	**0.970**	0.976	0.979	1.045
NVDA	**3.009**	3.038	3.068	3.078
PEP	**1.456**	1.460	1.466	1.509
TSLA	**4.454**	4.590	4.594	4.623
Average	**2.252**	2.284	2.283	2.343

* The DeepLOB model is replicated from Zhang, Zohren & Roberts (2019)

Market Orders VS Limit Orders

- Which LOB events are most informative for price discovery?
  - **Market Orders:** certainty of execution, but price risk
  - **Limit Orders:** fixed price, but execution risk
- In microstructure models, only market orders are informative:
  - Used by traders with private information to capitalize on their informational advantage
- Recent empirical papers question this assumption
  e.g. Brogaard, Hendershott, and Riordan 2019, JF
  - Variance decomposition exercise using VAR model
    $\Longrightarrow$ limit orders are more informative
  - Potentially due to high-frequncy activity of HFTs

Inference using Attention

- The data from LOBster features 3 types of events:
  - Submission/update of a limit order
  - Cancellation of a limit order
  - Execution of a market order

- Panel Regression (stock j, LOB event t):

$$\begin{aligned} \text{Attention Level}_{j,t} \;=\; \alpha \;&+\; \beta_0 \,\text{Event Type}_{j,t} \\ &+\; \beta_1 \,\text{Elapsed Time}_{j,t} \\ &+\; \beta_2 \,|\,\text{Price Change}_{j,t}\,| \;+\; \varepsilon_{j,t} \end{aligned}$$

- Control for Elapsed $\text{Time}_{j,t}$ and $|\,\text{Price Change}_{j,t}\,|$ to account for bid-ask bounce and lack of trading

Order Informativeness – Results

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var.	Attention Level	Attention Level	Attention Level	Attention Level	Attention Level	Attention Level
Execution	6.00***			5.93***		
	(5.63)			(5.59)		
Cancellation		-1.46***			-1.45***	
		(-7.96)			(-7.98)	
Submission			0.88***			0.88***
			(5.72)			(5.71)
Elapsed Time				1.81***	2.06***	2.08***
				(3.62)	(4.22)	(4.26)
|Price Change|				0.24***	0.56***	0.72***
				(3.35)	(4.17)	(4.82)
Constant	1.85***	2.71***	1.57***	1.78***	2.62***	1.47***
	(66.26)	(30.11)	(20.93)	(58.83)	(30.01)	(18.14)
Observations	2,560,000	2,560,000	2,560,000	2,560,000	2,560,000	2,560,000
R-squared	0.03	0.02	0.01	0.03	0.02	0.01
SEs Clustering	Day×Stock	Day×Stock	Day×Stock	Day×Stock	Day×Stock	Day×Stock

- Significantly higher attention for trade executions (+6%)
  $\Longrightarrow$ Market Orders are more informative than limit orders

Block Trades

- **Question:** does algorithmic trading harm market liquidity?
- **Idea:** study how the model behave during liquidity events
- Use transactions from institutional investors (AbelNoser/Ancerno proprietary data, period 2000-2014)
- Identify large block trades executed over multiple executions during a trading day (during year 2014)

Number of events	716
Number of institutions with events	51
Number of stocks with events	78
Number of days with events	234
Average block duration (minutes)	39.91
Average number of trades	22.89
Average dollar volume (million $)	6.61
Average volume ratio wrt CRSP	2.36%

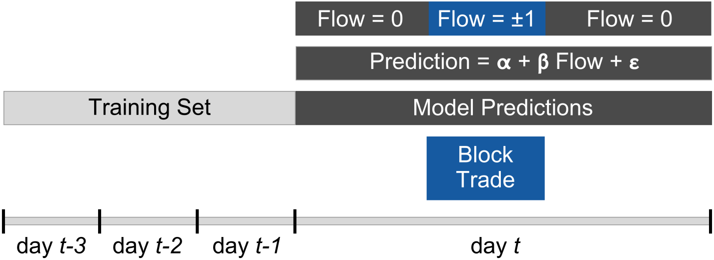
Model VS Block Trades

- How does the model behave when facing block orders?
- Block-order **Flow** $\in \{-1, 0, 1\} \cong \{\text{Sell}, \text{No trade}, \text{Buy}\}$
- Model's **Prediction** $\in [-1, 1]$
- Panel Regression (block order b, stock j, minute m)

$$\begin{aligned}\text{Prediction}_{j,m} &= \alpha + \beta_0 \text{ Flow}_{b,j,m} \\ &\quad + \beta_1 \text{ Stock FE}_j \\ &\quad + \beta_2 \text{ Day FE}_m + \varepsilon_{b,j,m}\end{aligned}$$

- Model Prediction is averaged over each trading minute

Model VS Block Trades – More Details



Model VS Block Trades – Results

	(1)	(2)	(3)	(4)
Dependent Variable	Model Prediction	Model Prediction	Model Prediction	Model Prediction
Flow	-0.08***	-0.08***	-0.06***	-0.05**
	(-3.39)	(-3.30)	(-2.74)	(-2.57)
Constant	-0.01			
	(-0.64)			
Observations	277,112	277,112	277,112	277,112
R-squared	0.01	0.01	0.13	0.14
Stock Fixed Effects		Yes		Yes
Day Fixed Effects			Yes	Yes
SEs Clustered By	Stock-Day	Stock-Day	Stock-Day	Stock-Day

- A market maker using the model would mostly provide liquidity to the block trades

Concluding Remarks

- Novel neural network model applied to LOB data
  - Attention mechanism delivers better prediction accuracy
- Attention allows to make inference
  - Market orders are the most informative
  - Other questions could be explored using this inference method
  - Cross-sectional attention?
- The model provides liquidity to large block orders
  - ML algorithms may improve overall market liquidity
  - Future research could use more granular data (millisecond time-stamped) with institutional trades
  - The behavior of alternative models should be tested