Reinforcement Learning for Optimized Trade Execution

Yuriy Nevmyvaka, Lehman Brothers (& CMU)
Yi Feng, Penn
Michael Kearns, Penn (& Lehman Brothers)

ICML 2006
Background on Market Microstructure

• Consider a typical exchange for some specific stock
• Limit order: specify price (away from the market)
• (Non-executable) Orders are placed in the buy or sell book
  – sorted by price; top prices are the bid and ask
• (Partially) Executable orders are filled immediately
  – prices determined by standing orders in the book
  – one order may execute at multiple prices
  – the "mechanical" component of market impact
• Market order: limit order with an extreme price
• Full order books now visible in real time
• What are they good for?

<table>
<thead>
<tr>
<th>MSFT</th>
<th>Stock</th>
<th>Symbol Search</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>LAST MATCH</th>
<th>TODAY'S ACTIVITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price</td>
<td>24.0700</td>
</tr>
<tr>
<td>Time</td>
<td>14:57:07.72</td>
</tr>
<tr>
<td>Volume</td>
<td>10,243,212</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>BUY ORDERS</th>
<th>SELL ORDERS</th>
</tr>
</thead>
<tbody>
<tr>
<td>SHARES</td>
<td>PRICE</td>
</tr>
<tr>
<td>500</td>
<td>24.0620</td>
</tr>
<tr>
<td>6,000</td>
<td>24.0610</td>
</tr>
<tr>
<td>5,000</td>
<td>24.0600</td>
</tr>
<tr>
<td>1,100</td>
<td>24.0650</td>
</tr>
<tr>
<td>100</td>
<td>24.0500</td>
</tr>
<tr>
<td>5,000</td>
<td>24.0500</td>
</tr>
<tr>
<td>200</td>
<td>24.0500</td>
</tr>
<tr>
<td>3,294</td>
<td>24.0500</td>
</tr>
<tr>
<td>1,000</td>
<td>24.0500</td>
</tr>
<tr>
<td>3,000</td>
<td>24.0430</td>
</tr>
<tr>
<td>100</td>
<td>24.0400</td>
</tr>
<tr>
<td>5,603</td>
<td>24.0400</td>
</tr>
<tr>
<td>2,100</td>
<td>24.0300</td>
</tr>
<tr>
<td>2,800</td>
<td>24.0300</td>
</tr>
<tr>
<td>(412 more)</td>
<td>(694 more)</td>
</tr>
</tbody>
</table>

As of 14:57:07.178
Optimized Trade Execution

• Canonical execution problem: sell V shares in T time steps
  – must place market order for any unexecuted shares at time T
  – trade-off between price, time… and liquidity
  – problem is ubiquitous

• Canonical goal: Volume Weighted Average Price (VWAP)
  • attempt to attain per-share average price of executions
  • widely used on Wall Street; reduces risk sources to execution
**RL for Optimized Execution**

- Basic idea: execution as **state-based stochastic optimal control**
  - **state:** time and shares remaining… what else?
  - **actions:** position(s) of orders within the book
  - **rewards:** prices received for executions
  - **stochastic:** because same state may evolve differently in time

- **This work:** large-scale application of RL to microstructure

- **Related work:**
  - Bertsimas and Lo
  - Coggins, Blazejewski, Aitken
"No Impact" State Factorization

Full OB State:

OB State Features:

OB execution simulation → reward (share prices)

OB(T) → OB(T-1) → OB(T-2)

Policy:

a(T) → a(T-1) → a(T-2)

Private State:

(T,V) → (T-1,V') → (T-2,V'')

Massive saving of data and computation...

Will it work?

Action: limit price for remaining volume

Training only, do full OB sim on test data
Experimental Details

- Stocks: AMZN, NVDA, QCOM (varying liquidities)
- $V = 5K$ and $10K$ shares
  - divided into 1, 4 or 8 levels of observed discretization
- $T = 2$ and 8 mins
  - divided into 4 or 8 decision points
- Explored a variety of OB state features
- Learned optimal strategy on 1 year of INET training data
- Tested strategy on subsequent 6 months of test data
- Evaluation:
  - compare to optimized submit and leave strategies
    - best single limit order price at start of trading interval
    - simplest form of learning
  - performance criterion: implementation shortfall
    - basis points compared to all shares at initial spread midpoint
    - an unattainable ideal (infinite liquidity assumption)
Results
Private State Variables Only: Time and Inventory Remaining

Average Improvement Over Optimized Submit-and-Leave

| T=4 I=1  | 27.16% | T=8 I=1  | 31.15% |
| T=4 I=4  | 30.99% | T=8 I=4  | 34.90% |
| T=4 I=8  | 31.59% | T=8 I=8  | 35.50% |
Strategy Visualization (10K, 2min)

General shape is intuitive, but (stock-specific) numerical optimization matters!
## Improvement From Order Book Features

<table>
<thead>
<tr>
<th>Feature</th>
<th>Improvement</th>
<th>Feature</th>
<th>Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bid Volume</td>
<td>-0.06%</td>
<td>Ask Volume</td>
<td>-0.28%</td>
</tr>
<tr>
<td>Bid-Ask Volume Misbalance</td>
<td>0.13%</td>
<td>Bid-Ask Spread</td>
<td>7.97%</td>
</tr>
<tr>
<td>Price Level</td>
<td>0.26%</td>
<td>Immediate Market Order Cost</td>
<td>4.26%</td>
</tr>
<tr>
<td>Signed Transaction Volume</td>
<td>2.81%</td>
<td>Price Volatility</td>
<td>-0.55%</td>
</tr>
<tr>
<td>Spread Volatility</td>
<td>1.89%</td>
<td>Signed Incoming Volume</td>
<td>0.59%</td>
</tr>
<tr>
<td>Spread + Immediate Cost</td>
<td>8.69%</td>
<td>Spread+ImmCost+Signed Vol</td>
<td>12.85%</td>
</tr>
</tbody>
</table>
Strategy Visualization II

![Graphs showing spread size and immediate cost for AMZN, NVDA, and QCOM](image)

- **AMZN**
  - Optimal Action vs. Spread Size and Immediate Cost

- **NVDA**
  - Optimal Action vs. Spread Size and Immediate Cost

- **QCOM**
  - Optimal Action vs. Spread Size and Immediate Cost
Q-Values: Trading Costs vs. Actions

predictive and actionable

predictive but not actionable
Future Work

• “Fancier” RL
  – function approximation
  – may permit richer feature set, but…

• RL for other stylized trading problems
  – market-making strategies

• Theory: low-impact RL?