REINFORCEMENT LEARNING APPROACHES TO OPTIMAL MARKETING MAKING

03/07/22

Haoye Lu,
David R. Cheriton School of Computer Science

Gašperov B, Begušić S, Posedel Šimović P, Kostanjčar Z. Reinforcement Learning Approaches to Optimal Market Making. *Mathematics*. 2021; 9(21):2689.

UNIVERSITY OF WATERLOO | FACULTY OF MATHEMATICS

Contents

Formulation of market maker problems

A financial market: people sell/purchase products by putting orders on a book.

(For simplicity, we assume there is only one product.)

Scenario A

1. A seller puts an order (bid) on the book containing:
   - Price to sell
   - Quantities of the products
2. If a buyer accepts the price (and the quantities), the transaction will be executed.

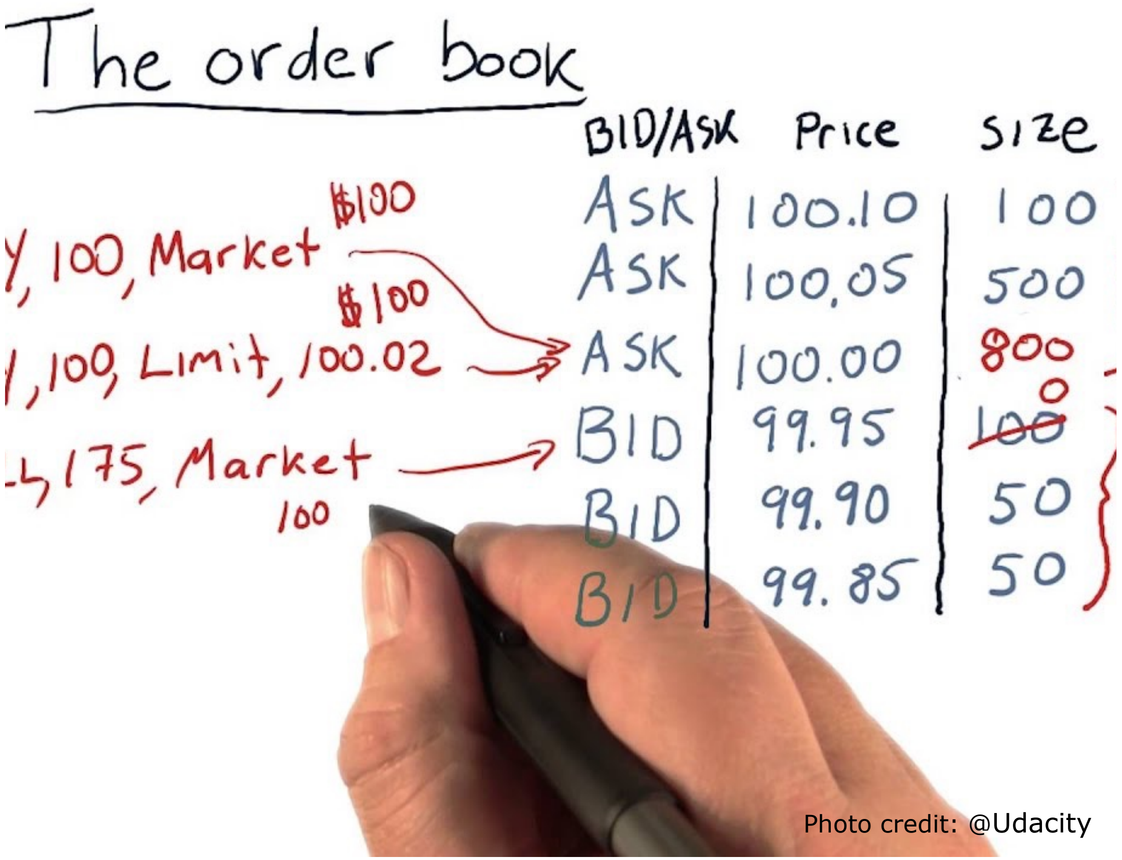


Photo credit: @Udacity

Formulation of market maker problems

Scenario B

1. A buyer puts an order (ask) on the book containing:
   - The highest price it can afford
   - Quantities of the products
2. If a seller accepts the price (and the quantities), the transaction will be executed.

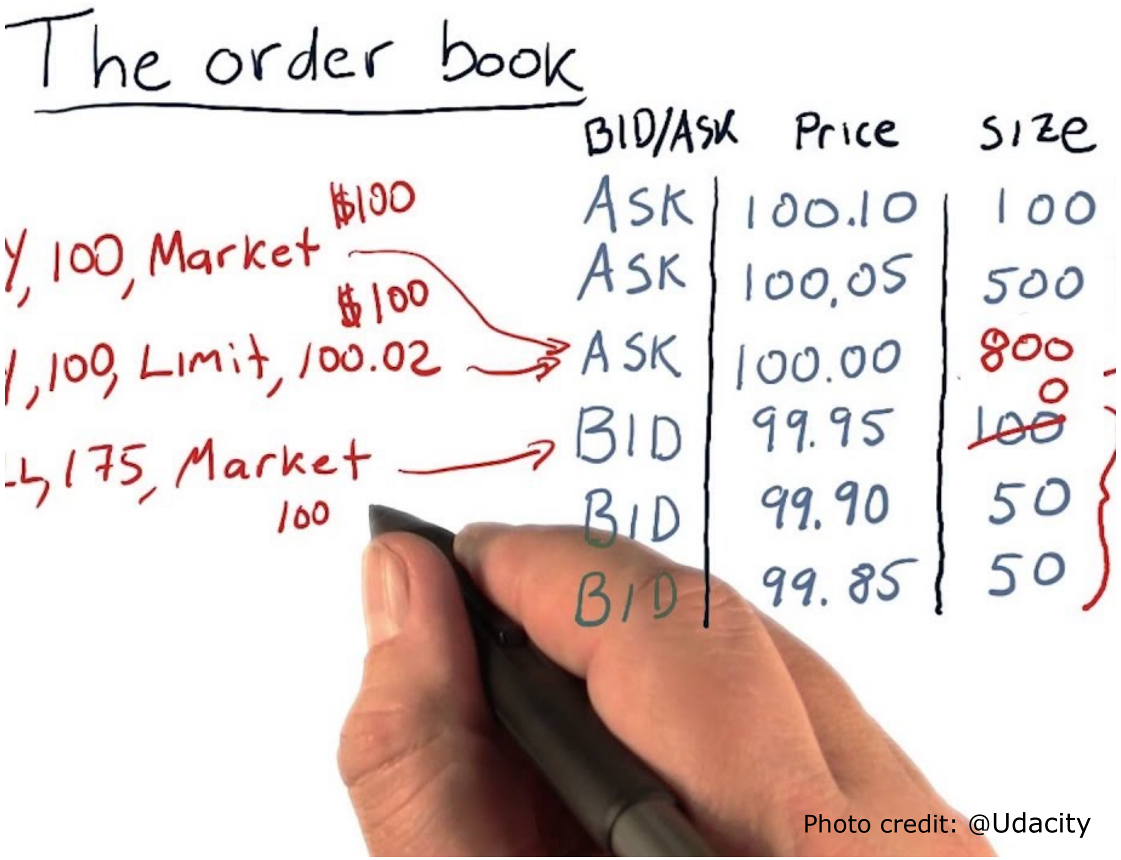


Photo credit: @Udacity

UNIVERSITY OF WATERLOO | FACULTY OF MATHEMATICS

Formulation of market maker problems

Market maker (MM)

- Post bids and asks at the same time.
- For example,
  - MM puts two orders:
    - Ask (buy): Price $99 of size 1.
    - Bid (sell): Price $101 of size 1.
  - If both orders are executed, the MM will earn
    - $101 - $99 = $2.
- Note: in principle, the price of bids is always higher than the one of asks. Otherwise, MM will lose money.

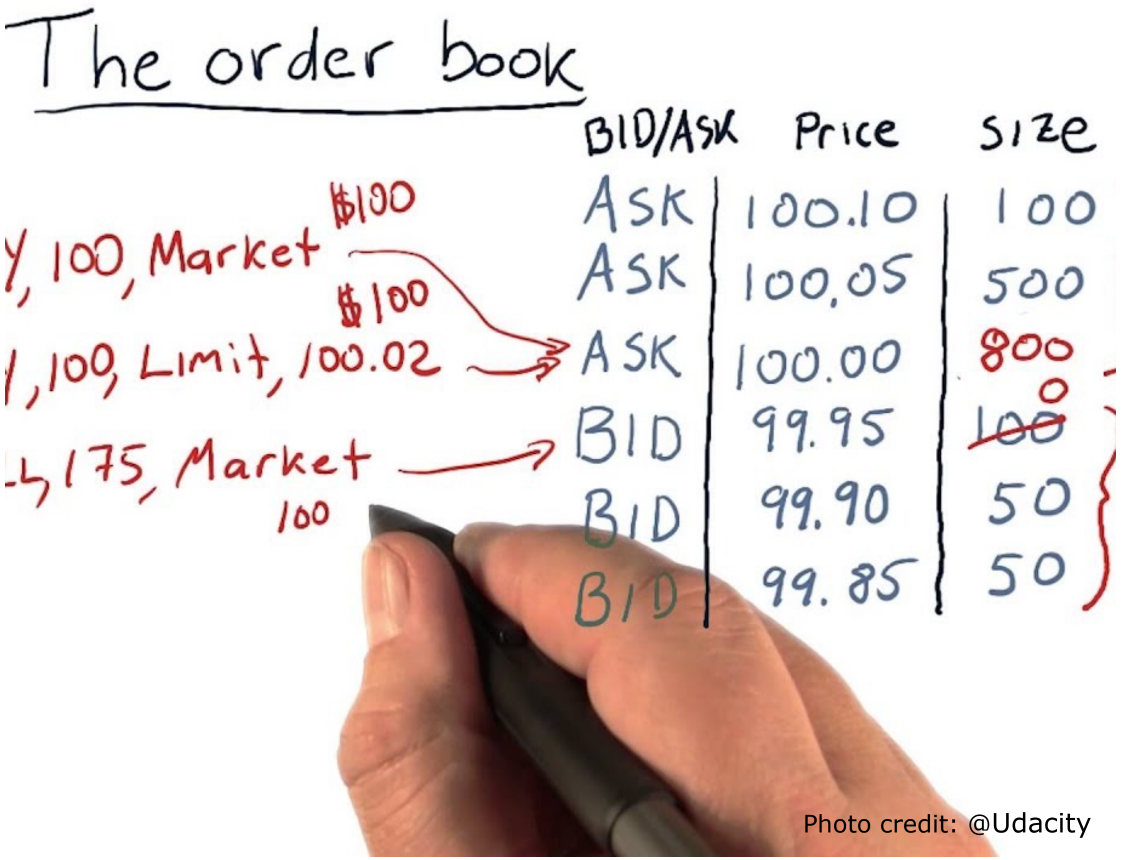


Photo credit: @Udacity

Examples

Foreign Exchange MM

- The price they sell to you is always higher than the one you buy from them.
- Assume you have 100 USD
  - You can exchange it for 124.29 CAD through the MM
  - Suppose you regret immediately and want to exchange the money back to USD, then you will only get 96.31 USD.
  - The difference 3.69 USD is earned by the MM.

CURRENCY	WE SELL TO YOU	WE BUY FROM YOU
Euro Member Countries (EUR)	1.4467290000	1.3648410000
Mexico (MXN)	0.0642800000	0.0575140000
United Kingdom (GBP)	1.7392880000	1.6403590000
United States (USD)	1.2905570000	1.2428500000
Argentina (ARS)	0.0134880000	0.0029320000
Aruba (AWG)	0.7874750000	0.6292770000

Fetched from: https://currencyconvertersinc.com/exchange-rates-calculator/, on Mar 22, 2022

Do MMs always make profits?

- Money markers earn money only when both orders (bid and ask) are executed.
- What if only the asks are processed?
  - Then the MM does not sell the products at the same time and has to put them in the warehouse.
  - As a result, the MM needs to pay the associated fees.
  - Such risk is called **inventory risk.**

Do MMs always make profits?

- **In general, the inventory risk stems from:**
  - **Fluctuations in the product market**
    - **Economic crises**
    - **Wars**
    - **Sanctions**
  - **Management Mistakes**

Examples

- Uncontrolled inventory risk could cause serious damage on MM's profit and the entire financial market.

The Crude Oil Treasure Incident of the Bank of China (BOC)

- BOC played a role like a MM in the transaction of crude oil future product.
- The oil future product works like a tradable contract which allows people to purchase oil that is not yet produced in advance.
- The contract becomes untradable after a deadline (roughly a week before it will be fulfilled).

Analysis from The Bank of China on The Crude Oil Treasure Incident

Authors

Bohan Sun[1, +], Cindy Ma[2, +], Wenzhi Sun[3, +], Yi zhai[4, +, *]

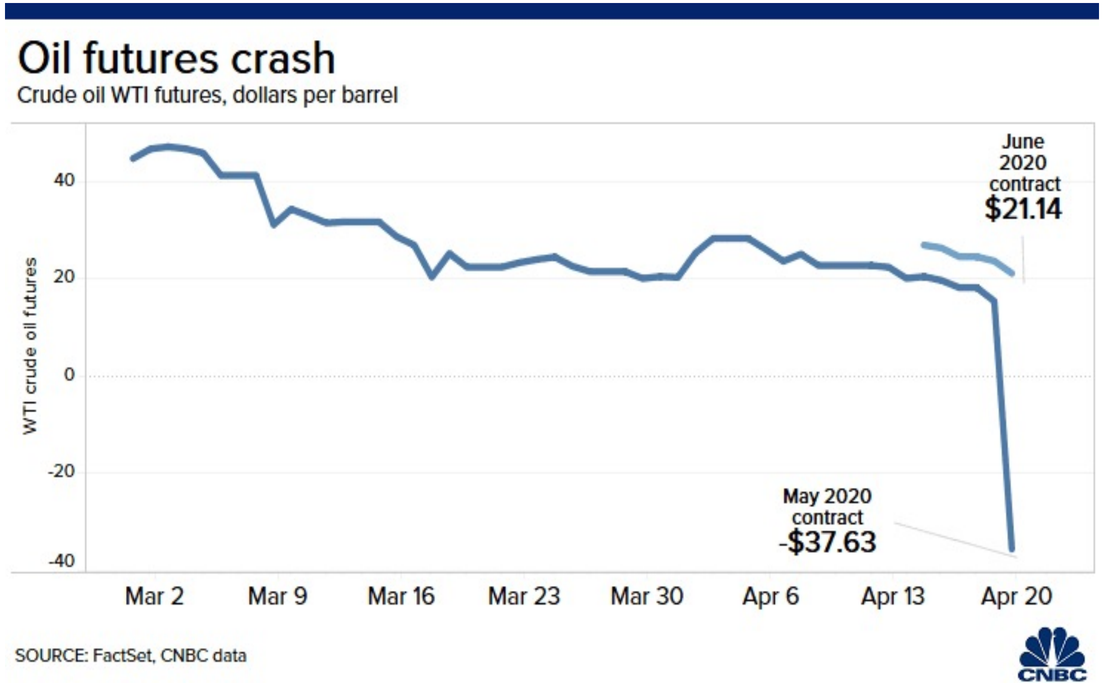
1 School of economic and business, Monash University, Melbourne, VIC3145, Australia
2 Ridge High School, Basking Ridge New Jersey, 07920, United States
3 Penn State University park, State College, Pennsylvania, 16801, United States
4 School of Politics and Public Administration, Soochow University, Suzhou, Jiangsu, 215123, China
+ These authors contributed equally to this work and should be considered co-first authors.
* Corresponding author. Email: eve.zhai@hotmail.com

Sun, B., Ma, C., Sun, W. and Zhai, Y. Analysis from The Bank of China on The Crude Oil Treasure Incident. Proceedings Of The 2021 3rd International Conference On Economic Management And Cultural Industry (ICEMCI 2021). pp. 3182-3190 (2021)

Examples

The Crude Oil Treasure Incident of the Bank of China (BOC)

- On April 20, 2020, BOC noticed that they had huge amount of oil futures unsold a few hours before the transaction deadline.
- BOC did not have any physical facilities to store oil, which means they would have nearly infinite inventory cost.
- As a result, they sold all the futures at all cost in a very short time, which caused the future price dropped to roughly negative $38 per barrel. (i.e. you do not need to pay for oil and will get $38 every barrel you purchase!)



Examples

The Crude Oil Treasure Incident of the Bank of China (BOC)

- The incident results in a huge loss of BOC's investors.
- It also caused a big chaos on the financial market.

THE WALL STREET JOURNAL.

SUBSCRIBE | SIGN IN

MARKETS | COMMODITIES

'Crude Oil Treasure' Debacle Leads to Fine for Chinese Bank

Investment product generated big losses for individual investors when U.S. oil prices briefly went negative

Other risks

Apart from the inventory risks, there are also many other risks that the economists may consider:

- Adverse selection risk
- Latency risk
- Model uncertainty risk

Formulation of MM problems

The market marker problem is to develop a policy to make orders (bid and ask) that maximize the profit with the controlled risk.

Classical approaches in Economics

A. Markov Chain-based model.

- Empirical study in Economy shows that the inventory risk is one of the most relevant factor in the MM problem.
- This prompts economists only focus on the inventory risk so that the corresponding models are not too complicated to analyze.
- Due to the need for inventory control, MM is naturally modelled as a Markov decision process (MDP), where the inventory levels correspond to the state values.

Classical approaches in Economics

B. Analytical approaches to MM problem

- Economists first construct a math model to describe all the relations between the buyers, sellers and MMs.
- Typically, the models are constructed using stochastic differential equations (SDEs). (like ODEs except that the function values could be random variables)
- Economists then solve the SDE systems and derive the optimal policy to ask and bid.

Drawbacks of the classical approaches

- To make the models computationally tractable and not too complicated to analyze, many strong assumptions must be made, which may not be true in practice.

- Many practitioners have reported that the developed methods tend to be ill-suited for real-world MM modeling.

Reinforcement learning-based approaches

The reported approaches can be mainly categorized into two types:

1. Information-Based Approaches
   - Markov Chain-based model

2. Approaches Stemming from Analytical Models
   - Using neural networks to approximate the solution of the SDEs.

Information-Based Approaches

[Chan and Shelton 2001] Chan, N.T.; Shelton, C. An Electronic Market-Maker; Massachusetts Institute of Technology: Cambridge, MA, USA, 2001.

- Tabular Q-learning
- States (discretized):
  - Inventory level
  - Order imbalance (difference between quantities of bids and asks)
  - Current market quality measure
- Actions (discretized):
  - Changes of prices of asks and bids. (e.g. increase price of asks by $2, decrease price of bids by $0.5)

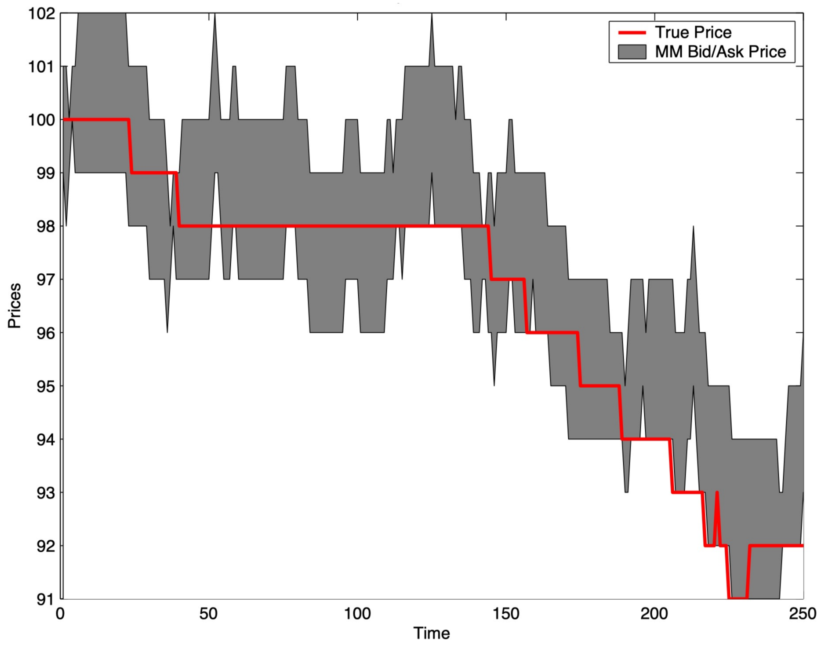
Information-Based Approaches

[Chan and Shelton 2001] Chan, N.T.; Shelton, C. An Electronic Market-Maker; Massachusetts Institute of Technology: Cambridge, MA, USA, 2001.

- Rewards:
  - The difference between the end-of-the-day and the beginning-of-the-day profit.
  - The environment only gives reward to the last step of the entire episode. (like the setting of AlphaGo)
- Training
  - Update the Q values using temporal difference (incremental updates).

Information-Based Approaches

- Simulation results:



UNIVERSITY OF WATERLOO | FACULTY OF MATHEMATICS

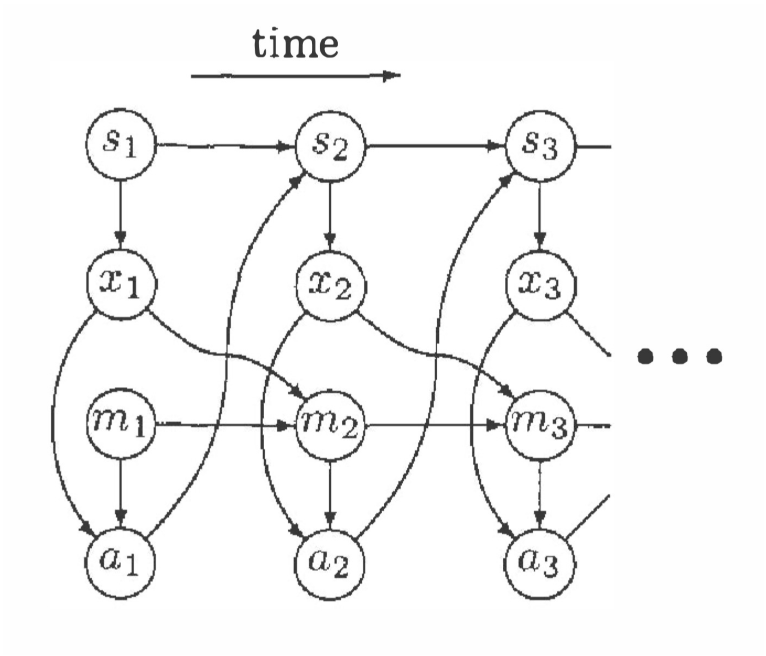
Information-Based Approaches

[Shelton 2001] Shelton, C.R. Policy Improvement for POMDPs using Normalized Importance Sampling; Conference on Uncertainty in Artificial Intelligence (UAI), 2001.

- In [Chan and Shelton 2001], the Markov chain model has the state consisting of inventory level, order imbalance and current market quality.
- The model is obviously suboptimal as many other factors can also affect the optimal bids and asks.
- For example,
  - If the weather forecast says there is a winter storm in Waterloo, the local heating oil price will increase.
  - Then the price of the optimal bids and asks will also increase.

Information-Based Approaches

The issue motivates Shelton (2001) to model the problem using a hidden Markov chain (HMM).



s_t: hidden states at time t (not observable)

x_t: observations from the market at time t

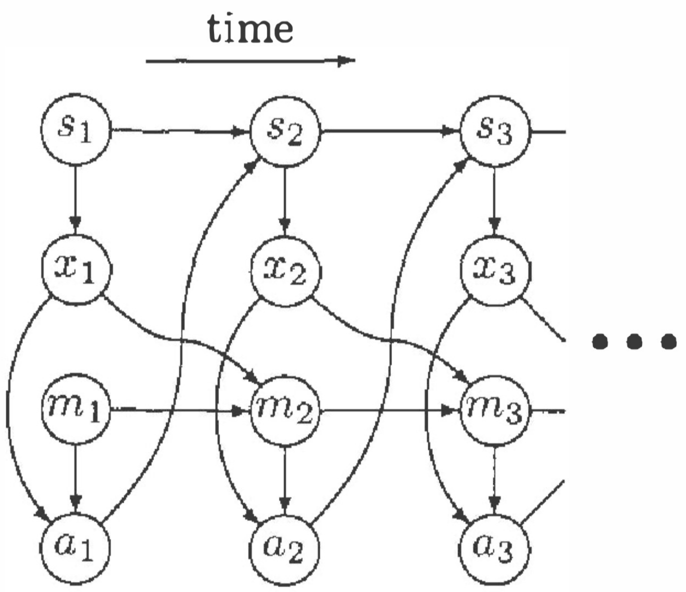
a_t: prices for bids and asks at time t

m_t: memory nodes (a hidden variable)

[Shelton 2001] Shelton, C.R. Policy Improvement for POMDPs using Normalized Importance Sampling; Conference on Uncertainty in Artificial Intelligence (UAI), 2001.

Information-Based Approaches

The problem motivates Shelton (2001) to model the problem using a hidden Markov chain (HMM).



The problem to solve:

Given $x_1, x_2, \dots, x_N$ and $a_1, a_2, \dots, a_N$, find the best a_{N+1} to maximize the profit.

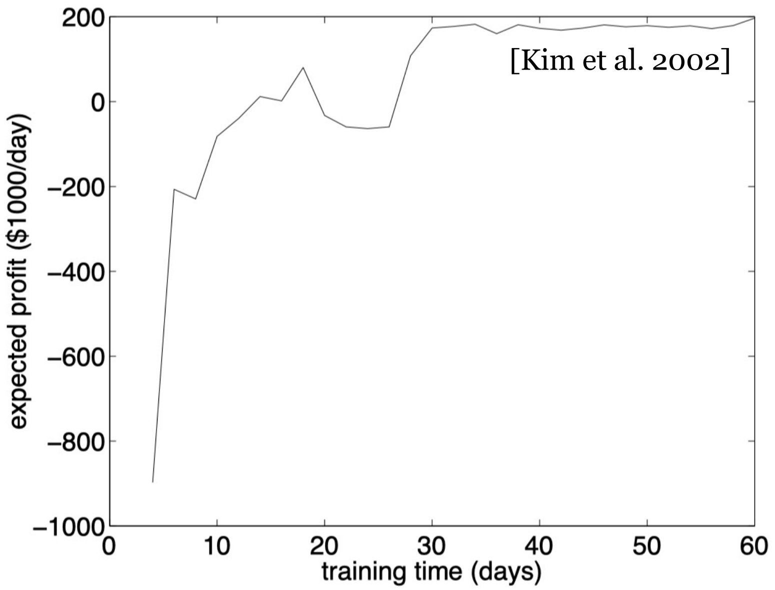
The author then used the statistical model to estimate the expected reward:

$$\Pr(\mathrm{R}|\mathrm{a}_{\mathrm{N+1}}; x_1, x_2, \dots, x_N, a_1, a_2, \dots, a_N)$$

and then adopted a hill-climbing algorithm to find the optimal a_{N+1}.

[Shelton 2001] Shelton, C.R. Policy Improvement for POMDPs using Normalized Importance Sampling; Conference on Uncertainty in Artificial Intelligence (UAI), 2001.

Information-Based Approaches



Shelton's group [Kim et al. 2002] tested the algorithm based on the data from the stock symbol GE for November 1, 1990.

- The results showed that the model can make profits after sufficiently long time of training.

[Shelton 2001] Shelton, C.R. Policy Improvement for POMDPs using Normalized Importance Sampling; Conference on Uncertainty in Artificial Intelligence (UAI), 2001.
[Kim et al. 2002] Kim, A.J.; Shelton, C.R.; Poggio, T. Modeling Stock Order Flows and Learning Market-Making from Data; Massachusetts Institute of Technology: Cambridge, MA, USA, 2002.

Approaches Stemming from Analytical Models

[Zhang and Chen, 2020] Zhang, G.; Chen, Y. Reinforcement Learning for Optimal Market Making with the Presence of Rebate. 2020.

- Recall that in analytical model-based approaches, economists use stochastic differential equation (SDE) systems to describe the dynamics among different entities in a market (like market makers, sellers, buyers, true prices, etc.).

- Even if multiple strong assumptions have been made, solving SDE systems are usually computationally intractable. More seriously, in most of the cases, the solutions do not have a closed-form expression.

Approaches Stemming from Analytical Models

It is well known that neural networks can be used to approximate functions without closed-form expressions as well as reduce computational complexity when exact solutions are not required.

Zhang and Chen (2020) developed a market marker model with the risk-adjusted rewards $v(t,q)$ satisfying:

$$v(t,q) = \max_{\pi} \mathrm{E}_{\mathrm{t}}\left[-\exp\left(-\gamma\left(\int_t^T v(s,q)\,ds\right)\right)\right] = \max_{\pi} \mathrm{E}_{\mathrm{t}}\left[v(t+\Delta, \mathrm{q}_{\mathrm{t}+\Delta})\exp\left(-\gamma\left(\int_t^{t+\Delta} v(s,q)\,ds\right)\right)\right],$$

- $v(t,q)$: the risk-adjusted rewards at time t with the inventory level q.
- γ: a coefficient related to the risk aversion.
- π: policy of adjusting the prices for bids and asks.
- Δ: a short period of time.

Note: the π^* that maximizes the expectation has a closed-form expression in terms of $v(t,q)$ and can be computed efficiently. So, we can easily derive the optimal prices for bids and asks given $v(t,q)$.

[Zhang and Chen, 2020] Zhang, G.; Chen, Y. Reinforcement Learning for Optimal Market Making with the Presence of Rebate. 2020.

Approaches Stemming from Analytical Models

$$v(t,q) = \max_{\pi} \mathrm{E_t}\left[-\exp\left(-\gamma\left(\int_t^T v(s,q)\,ds\right)\right)\right] = \max_{\pi} \mathrm{E_t}\left[v(t+\Delta, \mathrm{q_{t+\Delta}})\exp\left(-\gamma\left(\int_t^{t+\Delta} v(s,q)\,ds\right)\right)\right]$$

- The equation is very complicated and cannot be solved directly.
- However, it can be computed recursively as shown on the right-hand side of the equation.
- This motivates the author to use a neural network $v_\theta(t,q)$ to approximate it and use a method similar to temporal difference to optimize $v_\theta(t,q)$. In particular, let

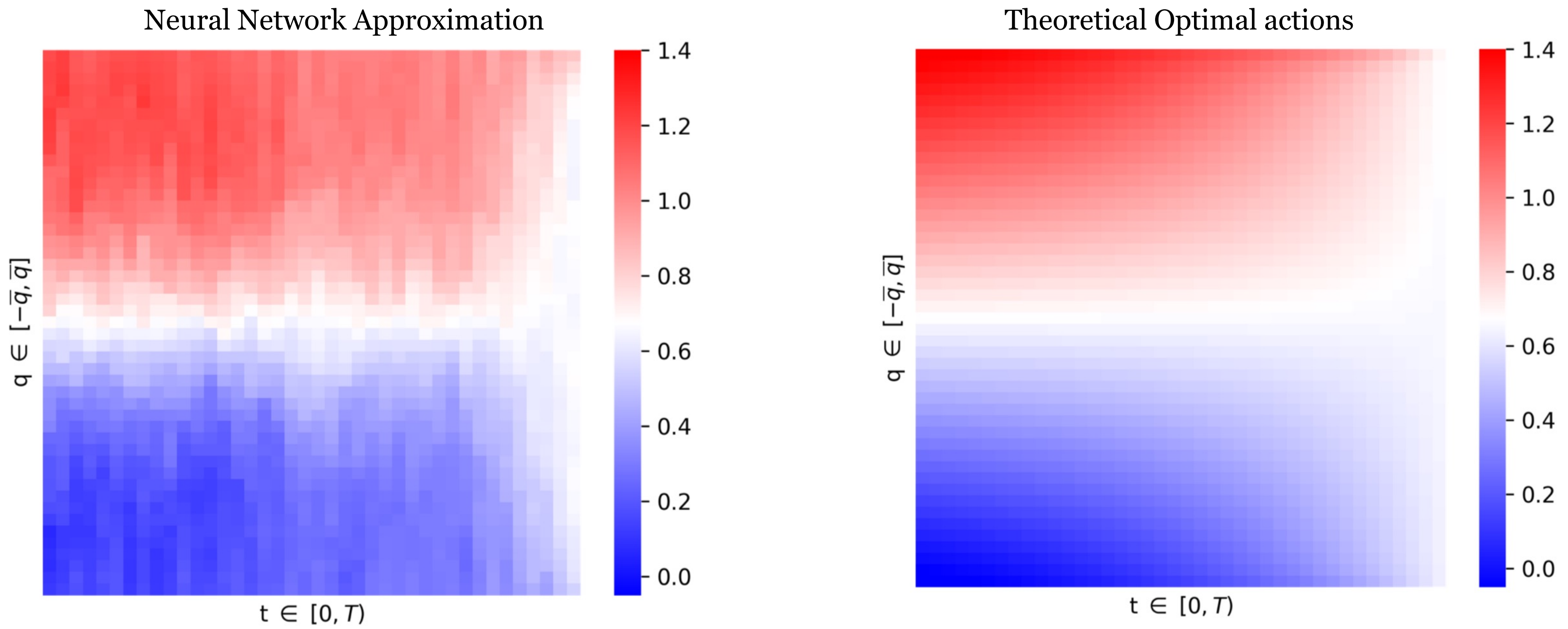
$$\tilde{v} = \max_{\pi} \mathrm{E_t}\left[v(t+\Delta, \mathrm{q_{t+\Delta}})\exp\left(-\gamma\left(\int_t^{t+\Delta} v_\theta(s,q)\,ds\right)\right)\right]$$

$$\theta \leftarrow \theta + \eta \cdot (\tilde{v} - v_\theta)\nabla_\theta v_\theta$$

where η denotes the learning rate.

[Zhang and Chen, 2020] Zhang, G.; Chen, Y. Reinforcement Learning for Optimal Market Making with the Presence of Rebate. 2020.

Approaches Stemming from Analytical Models



- The plots compare the adjustment of bid prices produced by the neural network with the theoretical optimal counterparts.
- We observe that the neural network-based implementation can largely learn the optimal actions, which shows it has sufficient capacity to capture and learn the dynamics of the entities in a complicated market setting.

[Zhang and Chen, 2020] Zhang, G.; Chen, Y. Reinforcement Learning for Optimal Market Making with the Presence of Rebate. 2020.

The conservative development of the automated MM system

- The development of the automated MM system is relatively slow in comparison to other deep learning and reinforcement areas.
- Most of the algorithms are only tested in simulators and their performances are questionable in practice.

The conservative development of the automated MM system

Why?

- In comparison to those classical RL problems like self-driving, playing games, etc., market maker problem is much more complicated as it requires us to have sufficient insights into the entire economy system and potentially the fundamental working mechanism of the human society.
- Current economic theories are still unable to give sufficient characterizations of the behavior of today's economy, which stops us from developing good RL-based models.
- More seriously, many important assumptions in Economy are not held in practice. For example, rational choice theory states that individuals rely on rational calculations to make rational choices that result in outcomes aligned with their own best interests. (However, we know this is not generally true in our daily life due to people's cognitive limitations and interpersonal relationships)

The conservative development of the automated MM system

Why?

- The lack of understanding of economy itself makes it difficult to develop a good simulator to train models, not to mention to test them in practice.
- Since MM plays a pivotal role in the financial system, an automated MM is required to have a provably safe and robust performance, which is very challenging to achieve considering the chaotic nature of markets.
- As a result, the development of RL-based MM system is very conservative, and the whole area is still in a nascent stage.

Summary

In this presentation, we

- Formulated market maker problems
- Introduced classical approaches in Economics
- Discussed typical Reinforcement Learning (RL)-based approaches
- Explained why the development of this area is very conservative.

UNIVERSITY OF WATERLOO

FACULTY OF MATHEMATICS

Thank you!