The 34th Annual Conference of the Japanese Society for Artificial Intelligence, 2020

STBM: Stochastic Trading Behavior Model for Financial Markets Based on Long Short-Term Memory

Masanori HIRANO Hiroyasu MATSUSHIMA Kiyoshi IZUMI Hiroki SAKAJI

School of Engineering, The University of Tokyo

In this study, we propose a stochastic model for predicting the behavior of financial market traders. First, using real ordering data that includes traders' information, we cluster the traders and select a recognizable cluster that appears to employ a high-frequency traders' market-making (HFT-MM) strategy. Then, we use an LSTM-based stochastic prediction model to predict the traders' behavior. This model takes the market order book state and a trader's ordering state as input and probabilistically predicts the trader's actions over the next one minute. The results show that our model can outperform both a model that randomly takes action and a conventional deterministic model. Herein, we only analyze limited trader type but, if our model is implemented to all trader types, this will increase the accuracy of predictions for the entire market.

1. Introduction

Today, there are increasing systemic risks in the financial market. It is because the complexity of the financial market has been increasing. The financial crisis of 2007–2008 was one of the most famous examples of systemic risks. To avoid future crashes, or more accurately predict future markets, we need to discover the underlying mechanism and take this into account. Many researchers are attempting to predict market movements using machine learning approaches using only macroscopic data, such as market prices and macroeconomic indicators. However, such approaches completely ignore financial markets' micro–macro mechanisms, the most important of which is that movements of the whole market are attributed to the accumulated interactions between individual traders' actions.

To deal with this problem, we need to build a trader model to predict each trader's actions for simulation. Recently, ordering data with masked trader information from the Tokyo Stock Exchange (TSE) was made available to certain people for particular usages, making it possible to filter the ordering data by traders. Moreover, our proposed clustering method, which extends the approach of [Uno 18], enables us to identify traders who show specific trading strategies. In this study, we propose a stochastic model for predicting traders' ordering behavior based on a type of neural network known as the long short-term memory (LSTM). We call this model the Stochastic Trading Behavior Model (STBM). Herein, we test the STBM on data from traders employing the well-known high-frequency trading market-making (HFT-MM) strategy. We focus on this because, as a first step, it is too challenging to handle all possible trader types at once. The HFT-MM strategy involves making frequent limit orders to obtain the margin price between the best ask and the best bid. However, the price is very likely to dramatically change beyond the spread; to hedge against this risk, HFT-MM traders monitor the market and make rapid series of orders, repeatedly making new orders and then canceling them. This strategy is distinctive and thus, relatively easy to identify in financial market data, so we decided to target HFT-MM traders when testing our STBM.

In our study, the STBM showed promising performance, outperforming both a by-chance model and the theoretical limit on deterministic models. Although these results only focus on HFT-MM traders, if our model can be applied to other types of traders, then it is possible to predict each trader's behavior and make trustable bottom-up simulations or financial market predictions based on real data.

Contact: Masanori HIRANO, hirano@g.ecc.u-tokyo.ac.jp; https://mhirano.jp/

2. Related Works

Over the years, there have been many attempts to predict future markets. Traditionally, technical or fundamental analyses of financial markets were conducted, but owing to technological advances, other prediction approaches are gaining in popularity. First, several researchers have used support vector machines to predict prices or price indices [Kim 03, Kercheval 15]. Some studies have also used deep learning to predict market prices, movements, or turning points [Sirignano 19, Tsantekidis 17, Dixon 19, Zhang 17]. Moreover, [Wang 19] proposed the CLVSA, a model for predicting price trends in financial markets based on the LSTM and Seq2seq deep learning models. [Tashiro 17] proposed an LSTM-based short-term price prediction model based on ordering data at millisecond timescales. They later extended this method to use a convolutional neural network instead of an LSTM [Tashiro 19]. However, the above work focused only on macroscopic data or mechanisms to predict macro indices such as market trends, price movements, or price indices, ignoring potential micro-mechanisms.

Alternatively, some researchers have taken other approaches, focusing on the traders' actions. [Chiarella 02] proposed a very basic trader model for use in simulations that determines the trader's strategy based on fundamental factors, trends, and noise. Other researchers later built and demonstrated a model of HFT-MM traders [Avellaneda 08] based on equations that consider the market ordering strength, current fundamental prices, and trader inventories. They used this to perform multi-agent financial market simulations.

Moreover, some studies have focused on HFT or HFT-MM strategies. [Hosaka 14] performed an HFT analysis based on the same data used in this study, finding that many of the HFTs in the TSE market were executing market-making orders. [Uno 18] proposed a method of clustering traders based on financial market data (which we extend in this study) and demonstrated that it could identify the orders of traders employing distinctive strategies, such as HFT-MM.

Other researchers have also studied financial markets using real data. For example, [Miyazaki 14] detected illegal orders and trades by applying a Gaussian mixture model to raw ordering data, while [Nanex 10] reported several distinctive trader behaviors based on mining ordering data.

3. Method and Experiments

3.1 Data & Extracting HFT-MM Orders

In our study, we used two kinds of data on the Tokyo Stock Exchange (TSE). One is the FLEX Standard, which contains the order book information. The other is order book reproduction data. This data contains all complete orders and masked trader information and enables us to trace traders' behaviors. We used ordering data of HFT-MM traders, which is extracted from order book reproduction data. The way of the extraction is explained in the following. The data from January 2015 to July 2015 was used for building our new model. On the other hand, the ordering data from August 2015 was used for evaluation and comparison.

In the following, we explain the extraction of HFT-MM trader's ordering data from the order book reproduction data. Although order book reproduction data has masked traders' information, it cannot help us to identify the type or strategy of traders. So, to extract ordering data of HFT-MM only, we used a clustering analysis base on [Uno 18]. In our paper, we extended the method [Uno 18] in terms of the usage of the executed and market order ratio.

At first, for each trader, we calculated

$$(\text{ActionsPerTicker}) = \frac{(\text{newOrders}) + (\text{changeOrders}) + (\text{cancelOrders})}{(\text{numTickers})}.$$

Here, we calculate the index for each bussiness days and we employ the median of the indecies of bussiness days in the data. Then to extract HFT only, we filter traders by $(\text{ActionsPerTicker}) > 100$. This gave us a total of 181 traders. Then, we also calculated the following indices for each trader on each day.

$$(\text{InventoryRatio}) = \operatorname*{Median}_{\text{ticker}}\left(\left|\frac{(\text{soldVol.})_{\text{ticker}} - (\text{boughtVol.})_{\text{ticker}}}{(\text{soldVol.})_{\text{ticker}} + (\text{boughtVol.})_{\text{ticker}}}\right|\right)$$

$$\begin{aligned}(\text{ExecutedRatio}) &= (\text{executedOrders})/(\text{newOrders})\\ (\text{CancelRatio}) &= (\text{cancelOrders})/(\text{newOrders})\\ (\text{MarketOrderRatio}) &= (\text{marketOrders})/(\text{newOrders})\\ (\text{TickerPerVSLOG}) &= \ln\{(\text{numTikcer})/(\text{numVS})\}\\ (\text{ActionsPerTickerLOG}) &= \ln(\text{ActionsPerTicker})\end{aligned}$$

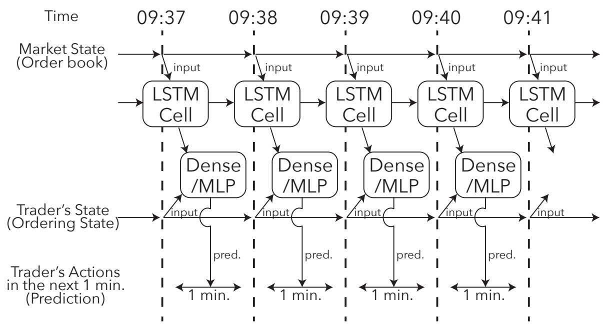


Figure 1: Overview of our neural network model.

After calculating these indices for each day, we took the median values. These indices are corresponding to the features of HFT-MM: low inventory ratios (InventoryRatio), low executed ratios (ExecutedRatio), high cancel ratios (CancelRatio), and very low market order ratios (MarketOrderRatio). Then, we normalized all indices over traders before clustering analysis. The clustering analysis we employed was hierarchical clustering with Ward's method, Euclidean distances, and a limit of ten clusters.

As a result, we found one cluster which satisfies all criteria for HFT-MM. In the cluster, there are seven traders. To ensure that these seven traders employ the HFT-MM strategy, we confirmed it by plotting their ordering histories and stock prices. Thus, finally, we got the ordering data of HFT-MM.

3.2 Trader Behavior Prediction Model

Finally, we built a neural network model to predict the behavior of the traders identified above. This model, based on an LSTM [Hochreiter 97], predicts each trader's behavior over the next one minute. We adopted an LSTM to evaluate the market state because it changes over time and the traders' actions depend on historical information, such as trends.

Figure 1 gives an overview of our model. This model is a kind of deep learning model to predict the future behavior of the traders. In this model, we employed an LSTM [Hochreiter 97] for processing time-series market states.

As a market state, we input data from FLEX Standard. The orders are divided into 18 classes: (1) best bid, (2–8) 1–7 tick(s) below the best bid, (9) 8 or more ticks below the best bid, (10) best ask, (11–17) 1–7 tick(s) above the best ask, and (18) 8 or more ticks above the best ask. Inputs data are turned into percentages of all volumes on the order book. In addition to these 18 classes, we include two features describing the best bid and ask prices. Thus, LSTM receives a total of 20 input features.

As another input representing the traders' ordering state, we input the detailed order data of HFT-MM. This data contains the orders' placement state at each minute. The

21 classes exists: (1) above best bid, (2) best bid, (3–9) 1–7 tick(s) below the best bid, (10) 8 or more ticks below the best bid, (11) below best ask, (12) best ask, (13–19) 1–7 tick(s) above the best ask, (20) 8 or more ticks above the best ask, and (21) other (set to 1.0 if all the other features are 0.0). These values are also the percentage of all volume of the current trader's orders.

After the processing by LSTM and Dense/MLP layer in Figure 1, our model predicts the probabilities of the actions in the next minute. The 21 possible ordering actions are (1) above best bid, (2) best bid, (3–9) 1–7 tick(s) below the best bid, (10) 8 or more ticks below the best bid, (11) below best ask, (12) best ask, (13–19) 1–7 tick(s) above the best ask, (20) 8 or more ticks above the best ask, and (21) other (do nothing). For the prediction, we used the softmax layer. And, for the model training, we used Adam optimizer and loss function of mean squared error (MSE).

The model details are tuned in hyperparameter tuning. Here, we divided the data into 90% training and 10% validation sets. The data was from January 2015 to July 2015 (143 business days). As a result, LSTM with a hidden layer size of 523 with an embedded size of 505, a dense layer with an embedded size of 310, and a ReLU activation function are employed.

After tuning the hyperparameters, we tested our model. As mentioned above, we used the August data, which was not used for training or validation, for testing. Here, we measured the MSE between the predicted actions and ground truth trader actions for the following minute. Particularly, we define the probabilities of taking each of these 21 classes at time t as p_1^t to p_{21}^t, and define the corresponding ground truth values as $\widehat{p_1^t}$ to $\widehat{p_{21}^t}$. The MSE at time t is defined as $\mathrm{MSE}_t = \sum_{i=1}^{21}\left(\widehat{p_i^t} - p_i^t\right)^2$. The final MSE is defined as $\mathrm{fMSE} = \frac{1}{150}\sum_{t=1}^{150}\mathrm{MSE}_t$, and is calculated for each data set. (TSE market have two 150 minutes sessions in one day.) Finally, we employed the mean fMSE value over the entire test data set as the performance index.

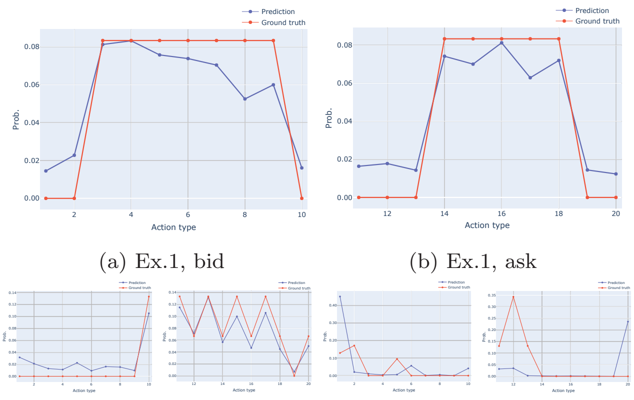
For comparison, we considered two other models. The first was a by-chance model; this zero-intelligence model was selected from the 21 classes at random with probability 1.0/21. The second model represented the theoretical limit on deterministic models. We assumed that the model would always take the best possible option, defined as

$$p_i = \begin{cases} 0.5 & (\text{if } i \in \{j,k\}) \\ 0.0 & (\text{other}) \end{cases},$$

$j = \arg\max_{i\in\{1,\cdots,10\}}\{\widehat{p_i}\}$, $k = \arg\max_{i\in\{11,\cdots,20\}}\{\widehat{p_i}\}$.

4. Results

First, Figure 2 shows the prediction results for three representative examples, plotting the bid and ask action probabilities separately for each example. Here, the blue and red lines indicate the predictions and ground truth (actual) actions, respectively. Examples 1 ((a) and (b)) and 2 ((c) and (d)) show quite good results, with MSEs of 0.004972 and 0.007460, respectively. However, the results for example 3 are poor, with an MSE of 0.324278.



(a) Ex.1, bid (b) Ex.1, ask

(c) Ex.2, bid (d) Ex.2, ask (e) Ex.3, bid (f) Ex.3, ask

Figure 2: Three example predictions. Here, (a) & (b), (c) & (d), and (e) & (f) are the results for bid and ask actions, respectively. The vertical axes show the action probabilities, while the horizontal axes correspond to the ordering action types mentioned above. However, these plots skip (21). The blue and red lines indicate the predictions and ground truth (actual) actions, respectively.

Table 1: MSE results for three different models.

Model	MSE
STBM	0.170129 ± 0.274952
By-chance	0.396102 ± 0.362223
Ideal deterministic model	0.530091 ± 0.506592

Next, we consider the statistical results. Table 1 shows the MSE for each of the three models. Here, our STBM outperforms the other two models, but the significant variances indicate that the prediction accuracies substantially fluctuated. But each pair of models were significantly different. These results clearly show that our STBM outperformed both the by-chance model and theoretical limit on any deterministic model.

5. Discussion

First, we discuss our STBM's performance, particularly the fact that it outperformed the theoretical limit on any deterministic model. Although we employed the MSE metric where deterministic models have a disadvantage, in real markets, traders cannot determine their actions without resorting to probabilities because future financial markets are inevitably uncertain. Thus, we believe that stochastic action models represent more natural trading strategies. Moreover, our model substantially outperformed the by-chance model.

Some of the features in Figure 2 also suggest that our model worked quite well. In example 1 ((a) and (b)), it correctly predicted wherein actions were more probable, and its predictions were excellent in example 2 ((c) and (d)). In bid action predictions of example 2, our model correctly predicted that the most likely bid action was to make an bid order at 8 or more ticks below the best bid. Conventional models, such as that in [Avellaneda 08], would never have predicted this because this behavior is atypical of HFT-MM trading.

The advantage of our STBM is that its neural network can account for the market and ordering states. In addition, the LSTM can remember previous market movements when evaluating the market state, enabling it to make more accurate and detailed predictions than conventional models. However, such data-driven approaches have only recently become possible, owing to the huge volume of data required. Moreover, as previously mentioned, such trader-based predictions have only become possible owing to the availability of new data that includes masked trader information.

This work can also be extended to improve conventional financial simulations, which depend heavily on manually created trader models which may, therefore, be biased. Moreover, many simulation-based approaches are deterministic. With our STBM, financial simulations will become more trustable and realistic.

In future work, we plan to improve our model in three main areas. First, We also plan to test other neural network architectures. We could consider other possible architectures, such as a convolutional neural network, instead of a dense layer/MLP.

Second, we want to apply the STBM to other types of traders. Here, we focused on HFT-MM traders, selected via a clustering analysis. However, if we are to predict the movements of the whole market by combining the predicted actions of all traders, we will have to consider other trader types; this will imply overcoming several obstacles. For example, currently, the STBM only predicts the action probabilities for the next minute, but we also need to consider the number of orders made during that time.

Third, to combine our STBM with simulation-based approaches for financial markets, we should conduct tests to confirm that it also correctly works under simulated conditions, as real and simulated financial markets may be somewhat different. However, if we can combine an STBM based on real market data with hypothetical market simulations, we can potentially make predictions for scenarios that have never occurred.

6. Conclusion

In this study, we have proposed the STBM for financial markets. This study aims to predict a trader's actions during the next minute based on the previous market state and the trader's ordering state. It is a neural network model based on an LSTM and predicts the trader's actions probabilistically with the possibility of doing nothing.

Here, we focused on applying the STBM to traders employing high-frequency trading market-making (HFT-MM) strategies. First, using real ordering data that included masked trader information (order-book reproduction data), we performed a hierarchical clustering analysis only for high-frequency traders, selecting one cluster that we believed to consist of HFT-MM traders. Then, we used the data from these traders to train an STBM. The results show that it outperformed both a by-chance model and the theoretical limit on conventional deterministic models.

In future work, we plan to apply the STBM to other trader types. In addition, we could apply it to agent-based simulations.

Acknowledgement

We thank the Japan Exchange Group, Inc. for providing the data. This research was supported by MEXT via Exploratory Challenges on Post-K computer (study on multilayered multiscale space-time simulations for social and economic phenomena). This research also used the computational resources of the HPCI system provided by the Information Technology Center at The University of Tokyo, and the Joint Center for Advanced High Performance Computing (JCAHPC) through the HPCI System Research Project (Project ID: hp190150).

References

[Avellaneda 08] Avellaneda, M. and Stoikov, S.: High-frequency trading in a limit order book, *Quantitative Finance*, Vol. 8, No. 3, pp. 217–224 (2008)

[Chiarella 02] Chiarella, C. and Iori, G.: A simulation analysis of the microstructure of double auction markets, *Quantitative Finance*, Vol. 2, No. 5, pp. 346–353 (2002)

[Dixon 19] Dixon, M. F., Polson, N. G., and Sokolov, V. O.: Deep learning for spatio- temporal modeling: Dynamic traffic flows and high frequency trading, *Quantitative Finance*, Vol. 19, No. 4, pp. 549–570 (2019)

[Hochreiter 97] Hochreiter, S. and Schmidhuber, J.: Long Short-Term Memory, *Neural Computation*, Vol. 9, No. 8, pp. 1735–1780 (1997)

[Hosaka 14] Hosaka, G.: Analysis of High-frequency Trading at Tokyo Stock Exchange (2014)

[Kercheval 15] Kercheval, A. N. and Zhang, Y.: Modelling high-frequency limit order book dynamics with support vector machines, *Quantitative Finance*, Vol. 15, No. 8, pp. 1315–1329 (2015)

[Kim 03] Kim, K.: Financial time series forecasting using support vector machines, *Neurocomputing*, Vol. 55, pp. 307–319 (2003)

[Miyazaki 14] Miyazaki, B., Izumi, K., Toriumi, F., and Takahashi, R.: Change Detection of Orders in Stock Markets Using a Gaussian Mixture Model, *Intelligent Systems in Accounting, Finance and Management*, Vol. 21, No. 3, pp. 169–191 (2014)

[Nanex 10] Nanex, : Nanex - Market Crop Circle of The Day (2010)

[Sirignano 19] Sirignano, J. A.: Deep learning for limit order books, *Quantitative Finance*, Vol. 19, No. 4, pp. 549–570 (2019)

[Tashiro 17] Tashiro, D. and Izumi, K.: Estimating stock orders using deep learning and high frequency data [in Japanese], in *Proceedings of the 31nd Annual Conference of the Japanese Society for Artificial*, pp. 2D2–2 (2017)

[Tashiro 19] Tashiro, D., Matsushima, H., Izumi, K., and Sakaji, H.: Encoding of high-frequency order information and prediction of short-term stock price by deep learning, *Quantitative Finance*, Vol. 19, No. 9, pp. 1499–1506 (2019)

[Tsantekidis 17] Tsantekidis, A., Passalis, N., Tefas, A., Kanniainen, J., Gabbouj, M., and Iosifidis, A.: Using Deep Learning to Detect Price Change Indications in Financial Markets, in *Proceedings of the 25th European Signal Processing Conference*, pp. 2580–2584 (2017)

[Uno 18] Uno, J., Goshima, K., and Tobe, R.: Cluster Analysis of Trading Behavior: An Attempt to Extract HFT [in Japanese], in *The 12th Annual Conference of Japanese Association of Behavioral Economics and Finance* (2018)

[Wang 19] Wang, J., Sun, T., Liu, B., Cao, Y., and Zhu, H.: CLVSA: A Convolutional LSTM Based Variational Sequence-to-Sequence Model with Attention for Predicting Trends of Financial Markets, in *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI-19)*, pp. 3705–3711 (2019)

[Zhang 17] Zhang, L., Aggarwal, C., and Qi, G.-J.: Stock Price Prediction via Discovering Multi-Frequency Trading Patterns, in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 2141–2149 (2017)