Input/Output Hidden Markov Models for Modeling Stock Order Flows

Adlar Jeewook Kim (with T. Poggio, CBCL)

The Problem: The project aims to develop a trainable system, which generates a sequence of orders. For the same market conditions as the training data, the generated orders will form a similar trading process (price and volume). The market conditions will be determined by the orders that are already generated, and by the market maker's actions.

Motivation: In the past few years, CBCL has developed an adaptive learning model for market-making by using reinforcement learning (Chan and Shelton, 2001[4]). The model was tested under a simply simulated market environment. However, testing under such simplified environment is inadequate to argue how well the market- making model will perform under the real market. In this project, we focus on designing a sophisticated and dynamic market environment using Input / Output Hidden Markov Models[2] .

Previous Work: Previously, the research on the information content of the trading process has been carried out (Easley, Kiefer, and O'Hara, 1997[6]). In this work, they fit a model of the trade process, which allowed to explain degrees of information content of the trading process of a particular stock. The model was fitted by maximum likelihood using transactions data on six stocks over 60 days. They showed that the trade process provided wealth of information as the following: The large and small trades have different information content, but this varies across stocks. The uninformed trades are history dependent. The large buys and large sells are equally informative.

Although this is not directly related to the order flow generation process, it helps defining the structure of our IOHMMs (particularly, input variables).

Approach: In this work, we propose to construct an IOHMM to generate a sequence of orders given the market conditions, which will form a trading process. Each state will emit an order based on their estimated conditional Gauissian distributions. The inputs will affect the transition probabilities distribution of the next following states, as well as the emission probabilities of the outputs (orders). The inputs will consist of variables, which can describe the market conditions.

For each new order, the arrival time and size are sampled from the estimated gamma distributions. For the side and price of the order, more work has to be done. First, the current input values are calculated by taking current values of the market condition (a five dimensional vector) and measuring the Euclidean distances to the means of the clusters of market condition vectors in training data. The nearest cluster is used as the input for the time step of the IOHMM. The output (order) of IOHMM is sampled from the output distribution Gaussian conditioned on the input and the state. If the price is less than the current best selling price (for buy orders) or greater than the current best buying price (for ask orders), the order is assumed to be a market order. Otherwise, it is treated as a limit order. Finally, the next state is generated from the transition probabilities given the current state and the input.

We compared the market model's reaction when market maker's quotes are significantly different. The Figure 1 shows the traded price over time when market maker was forced to quote bid and ask prices with spread of \$1 and \$10 respectively. We trained the market model based on the data from the stock symbol IBM for November 1, 1990. The result shows that when the market maker is forced to quote prices with larger spread, the market's volatility is significantly increased. The average bid / ask spread of above cases are \$0.0906 and \$7.7768. However, since the same market model were used for both cases, the volume weighted average prices were similar (\$227.51 and \$237.52).

Impact: Previously, modeling a financial market environment involved a number of unrealistic assumptions such as the existence of a true price process, or differentiating the informed and uninformed traders. And these assumptions over-simplified the model. The new approach will empirically model the aggregated behavior of the trading crowd. This will make the adaptive learning model for market-making possible to be tested under a more realistic market environment.

Future Work: Although the IOHMM domain model is interesting, the results of the IOHMM model still needs to be extensively studied by applying various scenarios of market conditions. Testing under

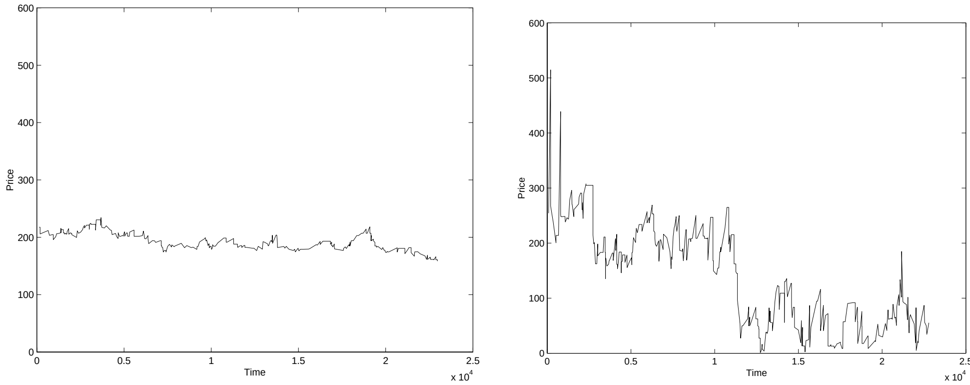

Figure 1: The price changes over time when the market maker was forced to quote with bid / ask spread of $1 and $10 respectively.

extreme conditions such as market-maker's bid-ask spreads of $1 and $10 are not sufficient enough to discuss the quality of the model.

Research Support: Research at CBCL is sponsored by grants from: Office of Naval Research (DARPA) under contract No. N00014-00-1-0907, National Science Foundation (ITR) under contract No. IIS-0085836, National Science Foundation (KDI) under contract No. DMS-9872936, and National Science Foundation under contract No. IIS-9800032

Additional support was provided by: Central Research Institute of Electric Power Industry, Center for e-Business (MIT), Eastman Kodak Company, DaimlerChrysler AG, Compaq, Honda R&D Co., Ltd., Komatsu Ltd., Merrill-Lynch, NEC Fund, Nippon Telegraph & Telephone, Siemens Corporate Research, Inc., Toyota Motor Corporation and The Whitaker Foundation.

References:

[1] Y. Bengio. Markovian models for sequential data. *Neural Computing Surveys*, 2:129–162, 1999.

[2] Y. Bengio and P. Frasconi. Input / output hmms for sequence processing. *IEEE Transactions on Neural Networks*, 7(5):1231–1249, 1996.

[3] Y. Bengio, V. Lauzon, and R. Ducharme. Experiments on the application of iohmms to model financial return series. *IEEE Transactions on Neural Networks*, 12(1):113–123, 2001.

[4] N. T. Chan and C. R. Shelton. An Electronic Market Maker. Technical Report 2001-005, MIT Artificial Intelligence Laboratory, 2001.

[5] T. Chan. *Artificial Markets and Intelligent Agents*. PhD thesis, Massachusetts Institute of Technology, 2001.

[6] D. Easley, N. Kiefer, and M. O'Hara. The information content of the trading process. *Journal of Empirical Finance*, 4:159–186, 1997.

[7] D. Easley, N. Kiefer, and M. O'Hara. One day in the life of a very common stock. *Review of Financial Studies*, 10(3):805–835, 1997.

[8] A. J. Kim and C. R. Shelton. Modeling Stock Order Flows and Learning Market-Making from Data. Technical Report 2002-009, MIT Artificial Intelligence Laboratory, 2002.

[9] L. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. In *Proceedings of IEEE, Vol. 77, No. 2*, pages 257–286, 1989.