**MPM Capstone Proposal**

**Machine Learning Application on Chinese Stock Market: Internship Report of High-frequency Trading**

Haoyang Han

Master of Project Management Program

Department of Civil and Environmental Engineering

McCormick School of Engineering and Applied Science

Northwestern University

Evanston, Illinois

October 05, 2018

**Title:** Machine Learning Application on Chinese Stock Market: Internship Report for High-frequency Trading

Project Context and Purpose:

High frequency trading (HFT), which is also called algorithmic trading or ‘flash’ trading, is getting increasingly popular in Chinese stock trading market. Generally, it’s divided by stock trading and derivatives trading (Lattemann el al. 2012; Menkveld 2013). By developing algorithm for trading and metrics for evaluation, high frequency trading could get higher alpha factor which lead to higher revenue for hedge fund (Tortoriello, R. 2009; Leote el al. 2012). Quantitative analysis is using mathematical, statistical and economical knowledge to transfer real-world problem into mathematical problem then solving them. In high-frequency trading, the detailed quantitative analysis should be abstracting statistical patterns hidden in trading data into predictable model and applying those models into practice.

By designing and analyzing algorithms which enable computer to do regression, classification and clustering, machine learning has been growing rapidly in all fields of study, including data mining, natural language processing, computer vision, recommendation system, vision recognition and voice recognition (Goldberg et al., 1988; Collobert et al., 2008;). In high frequency trading, people use stocking trading data to feed machine learning algorithms for future stock price prediction. Deep learning method is the subset of machine learning. By combining multiple levels of regressor or classifier together in a single model, deep learning algorithms are trying to make more accurate classification or regression. In practice, researchers and engineering would develop deep neural network (DNN), convolutional neural network (CNN), recurrent neural network (RNN), generative adversarial nets (GAN), autoencoder (AE) and restricted Boltzmann machine (RBM) to solve real-world problems (Hinton et al., 2012; Girshick, 2015; Krizhevsky et al., 2012; Mikolov et al., 2010).

Since the dataset for training machine learning models are getting increasingly huge, many machine learning and deep learning algorithms would take long time to get trained from datasets. Meanwhile, hyperparameters, such as learning rate, iteration time and cost which should be considered before training. To improve efficiency and accuracy of model, this experiment used grid search to search for optimized hyperparameter to achieve the model’s best time and accuracy performance.

Since it’s meaningful for individuals and institutes to discover the law under stock market, and doing internship in a quantitative field is extremely beneficial for career development, I chose the machine learning quantitative researcher position for my summer internship. The purpose of this paper is to provide an overall evaluation towards the summer internship, to summarize an general and introductory module for how to apply machine learning techniques into quantitative analysis practice, and to provide a new possibility of what a MPM student could do after graduation. I believe this capstone should be beneficial.

The experiments this capstone would discuss about are following:

1. What’s the shape, size and statistical characters of experiment data;

2. Different method we will use to standardize the data;

3. Different machine learning models and their performance in high frequency trading;

4. Some basic architecture of deep learning in high frequency trading;

5. Weakness of the experiments and what else could be done in future experiment design.

**Proposed table of contents**

Abstract

Executive summary

1.0 Introduction

1.1 Report background

1.3 Report purpose and scope

2.0 Background

2.1 History of high frequency trading

2.2 Quantitative analysis application in high frequency trading

2.3 Machine learning application in high frequency trading

3.0 Experiment method and data resources

3.1 Data resources

3.2 Data pre-processing pipeline

3.3 Data standardization method

3.4 Model selection criteria

3.5 Grid search for hyper-parameters

3.6 Optimization method for performance improvement

3.7 Strategy evaluation criteria

4.0 Experiment Result

4.1 Comparison between different standardization method

4.2 Model comparison result

4.3 Case study: ensemble method application for performance improvement

5.0 Conclusions

6.0 Discussions

7.0 References

**Timeline**

June 16, 2018: Internship Start, design experiment and read enough essay in high frequency trading, machine learning and deep learning.

September 11, 2018: Complete internship and save all data and files about internship project.

October 15, 2018: Submit first half of report of introduction, background, experiment method and data recourses. Receive feedbacks and make adjustments.

October 30, 2018: Submit second half of report for remaining session.

November 10, 2018: Submit final draft and prepare for final presentation.

November 15, 2018: Give presentation.

**Possible technical readers**

Not Applicable.

**References**

Collobert, R., & Weston, J. (2008, July). A unified architecture for natural language processing: Deep neural networks with multitask learning. In Proceedings of the 25th international conference on Machine learning (pp. 160-167). ACM.

Girshick, R. (2015). Fast r-cnn. In Proceedings of the IEEE international conference on computer vision (pp. 1440-1448).

Goldberg, D. E., & Holland, J. H. (1988). Genetic algorithms and machine learning. Machine learning, 3(2), 95-99.

Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A. R., Jaitly, N., ... & Kingsbury, B. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. IEEE Signal processing magazine, 29(6), 82-97.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).

Lattemann, C., Loos, P., Gomolka, J., Burghof, H. P., Breuer, A., Gomber, P., ... & Zajonz, R. (2012). High Frequency Trading. Business & Information Systems Engineering, 4(2), 93-108.

Leote, R. A. U. L., Lu, X., & Moulin, P. (2012). Demystifying equity risk-based strategies: A simple alpha plus beta description. Journal of Portfolio Management, 38, 56-70.

Menkveld, A. J. (2013). High frequency trading and the new market makers. Journal of Financial Markets, 16(4), 712-740.

Mikolov, T., Karafiát, M., Burget, L., Černocký, J., & Khudanpur, S. (2010). Recurrent neural network based language model. In Eleventh Annual Conference of the International Speech Communication Association.

Tortoriello, R. (2009). Quantitative strategies for achieving alpha. McGraw Hill.

**Additional sources to read**

Pazzani, M. J., & Billsus, D. (2007). Content-based recommendation systems. In The adaptive web (pp. 325-341). Springer, Berlin, Heidelberg.

Prix J, Loistl O, Huetl M (2007) Algorithmic trading patterns in Xetra orders. European Journal of Finance 13(8):717–739

Watt, Borhani, Katsaggelos, Borhani, Reza, & Katsaggelos, Aggelos Konstantinos. (2016). Machine learning refined : Foundations, algorithms, and applications. Cambridge, United Kingdom: Cambridge University Press.