Literature Review

Our research assumes minute-level price fluctuation pattern is independent of corporate fundamentals and macro economy. Thus, unlike the studies of Chiang et al. (2016), Chourmouziadis and Chatzoglou (2016), and Zhong and Enke (2017) in which daily price data are used as input, we seek to develop a predictive model based on minute-level input price data. The prediction of future stock price had also been understood as both classification and regression problems in previous studies. T.-l. Chen and Chen (2016) and Zhong and Enke (2017) provided prediction of market direction as either up or down. In more complicated cases, Chourmouziadis and Chatzoglou (2016) specified cash and stock within the optimal portfolio composition. Our study intends to give a prediction of the stock return in the next minute compared to the current time point. Due to certain limitation and just for preliminary testing of our strategy, we are currently using 10 days of minute-level price data of 50 stocks, but our aim is to obtain data of 10 years for model training. The input will be 500 dimensional lagged stock returns, which are the returns of 50 stocks in the previous 10 minutes. We may adjust the number of lagged periods for better performance later.

There have been linear and nonlinear models to predict stock price movement with varying degrees of success. Chong, Han, and Park (2017) noted a multilayer artificial neural network might be particularly suitable with such time-series data, due to its higher computational power and sophistication of algorithm. Such model selects features based on raw input price data automatically and does not require understanding or providing data from the side of fundamentals or macro economy, which fits our assumption about minute-level price fluctuation pattern. Our model will be composed of two parts, with the first part being unsupervised learning with traditional ML techniques like RBM, PCA, etc. The second part takes advantage of recurrent neural network (RNN) model, especially its variant LSTM. For performance measurement, previous studies have used trade simulation or various MSE methods (Chiang et al., 2016; Chourmouziadis and Chatzoglou, 2016; Zhong and Enke, 2017; Chong, Han, and Park, 2017). We plan to use MSE in the assessment.