

Tue Hoang Vo

Math and Computer Science Department

Summer 2014

This summer, I conducted research on data mining at Sewanee's Computer Science department; Dr. Linda Lankewicz was my faculty advisor. My task was to test the effectiveness of different data mining algorithms and other parameters in a trading simulation based on historical data. More precisely, I pulled the historical data from the internet, ran feature selections techniques, tested different parameters for the artificial neural network, ran the trading simulation code and recorded the results. The research's finding is not conclusive (the best recorded performance is at approximately 2.31% per year, which outperforms hypothetical risk-premium rate at 2% but has significantly higher risk), but the result is optimistic towards a pure quantitative approach towards the challenging task of stock market prediction.

I was required to read a variety of R and data mining textbooks as preparation for the research. As such I developed new skills and greater understandings of the field of data mining. Specifically, first, I learn to use R programming language to perform statistical tasks, ranging from basic calculation, graphing to complex functions such as regression analysis. Second, I learned about the theoretical construction of different data mining algorithms (neural network, random forest classifier, information gain). Third, I learned to import data mining algorithms packages and performed different testings to optimize the parameters for specified task.

I also learned about research procedures. First, Dr. Lankewicz instructed me about the interchanged process of testing and forming hypothesis, as we continually ran test, looked at the data, formed new hypothesis and retested the data. Second, I learned the importance of relying heavily on collected data to provide assumption. Thirdly, Dr. Lankewicz constantly challenged me to explain my rationales clearly in both verbal and written communication. This helped me to communicate more clearly and concisely. Finally, during the process, as I communicated my ideas to Dr. Lankewicz, she helped me to either make my ideas clearer or give a different suggestions for me to consider. From that experience I learned to be receptive to new suggestion in a collaborative environment.

To conclude, the research was a great experience. Following the internship, I figured out that I enjoyed learning about the design of computer algorithms and about different techniques to extract information from a large pool of data. I am starting to consider a technical career in the future.