|  |  |
| --- | --- |
| **Namn:** | **Philip Arvidsson**  **Tobias Ånhed** |
| Titel: | **Sequence-to-sequence learning of financial time series in algorithmic trading** |
| Uppdragsgivare: | CSL@BS |
| Kontaktperson: | Patrick Gabrielsson, [patrick.gabrielsson@hb.se](mailto:patrick.gabrielsson@hb.se) |
| Bakgrund: | Time series forecasting is an important topic in multiple domains; such as predicting the future value of temperature, humidity and wind speed in weather forecasting applications, predicting the future value of power consumption in an electrical grid, or predicting the future value of volatility and returns in a financial market.  Many statistical tools and machine learning algorithms aren’t designed to handle sequential datasets and assume that data points are IID (Independent and Identically Distributed). Time series datasets include a temporal component, producing serially correlated data points (hence violating the IID assumption). Furthermore, most time series datasets are non-stationary.  In statistics, econometrics and signal processing, one traditional approach, used to model time series, is to transform the dataset (using i.e. differencing and logarithms) into a (weak) stationary distribution, induce a predictive ARIMA-based (AutoRegressive Integrating Moving Average) model, and re-transform the predicted output signal into its original form.  In machine learning, HMMs (Hidden Markov Models), DBNs (Dynamic Bayesian Networks), TDNNs (Tapped-Delay Neural Networks) and RNNs (Recurrent Neural Networks) are commonly used to handle sequential datasets. Neural network-based methods (connectionism) have become popular machine learning methods in both academia and industry, especially since the inception of the big data era and the deep learning revolution (2006).  RNNs (Recurrent Neural Networks) have historically been plagued with the vanishing (or exploding) gradient problem, but this problem was solved by Hochreiter and Schmidhuber when they introduced their LSTM-based (Long Short-Term Memory) RNNs. Recently, deep LSTM-based RNNs have successfully been applied to predict future time series sequences from historical time series sequences, such as predicting the weather for the next 24 hours, or predicting the next sequence of words in written and spoken language.  Financial time series are commonly used in algorithmic trading. In algorithmic trading, it is important to make predictions about numerous variables (such as volatility and returns) in order to time market entry and exit. Furthermore, in a trading scenario, it isn’t sufficient to predict when to place a single bid or ask order into the market, since a profit isn’t realized until a long or short position is closed-out (by placing a corresponding ask or bid into the market). Since deep LSTM-based RNNs have never been applied within the algorithmic trading domain before, and since they have shown great success in solving similar problems in other domains, it raises the question if the technique can be used to predict a future sequence of financial variables that can be used to time both entry and exit positions within a certain time horizon. |
| Kort beskrivning: | The thesis project will investigate the applicability of deep LSTM-based RNNs in sequence-to-sequence learning of financial time series, from which an algorithmic trading strategy can be devised. In essence, given an input sequence of vectors Xt, Xt-1, …, Xt-n at time *t*, the predictive model should produce a prediction consisting of a sequence of output vectors Yt+1, Yt+2, …, Yt+m, and the associated probability P(Yt+m, …, Yt+2, Yt+1 | Xt, Xt-1, …, Xt-n).  Furthermore, the project will include a comparative study between the chosen method and the state-of-the-art (e.g. ARIMA, SVR, DBN), using common evaluation metrics found in both the machine learning community and computational finance.  The project will include common steps used in an algorithmic trading system, such as the acquisition of data (freely available FOREX tick data), pre-processing (e.g. cleaning, data point interpolation, transformation, feature extraction), designing and training the predictive deep LSTM RNN architectures, trade strategy development, back testing, execution and evaluation (results analysis). |
| Vetenskapligt/innovativt bidrag: | The major contribution of the thesis project will be to assess the applicability of sequence-to-sequence learning using deep LSTM-based RNNs within the domain of financial time series modelling. Furthermore, the thesis work will produce a comparative study of the chosen method, benchmarked against the state-of-the-art (e.g. ARIMA, SVR, DBN). The results produced will also add to the knowledge base of sequence-to-sequence learning in time series, in general. |
| Krav på förkunskaper: | Basic knowledge of predictive machine learning methods and proficiency in any programming language. Although, proficiency with python and the scipy stack (including scikit-learn, tensor flow and keras), is ideal. |
| Handledare/Önskad handledarkompetens | Patrick Gabrielsson |