AI Forecasting Fama-French Risk Factors: vs Old School

Xinxuan Wang
shirleywxx@outlook.com
Supervisor: Dr. David Newton

MSc Finance
JMSB Concordia University

WHAT

- **AI Approach**: LSTM (Long Short-Term Memory) network, a type of neural network that can complete a very deep learning task and has outstanding predictive power on time series data because of its "memory" algorithm
- **Old School Approach**: ARIMA (Autoregressive Integrated Moving Average) model, a benchmark model for time series data analysis
- **Fama-French Risk Factors**: the risk factors are from the Fama-French five-factor model, a very classic model designed to describe stock returns. The five factors are: SMB, HML, RMW, CMA, and Rm-Rf

WHY

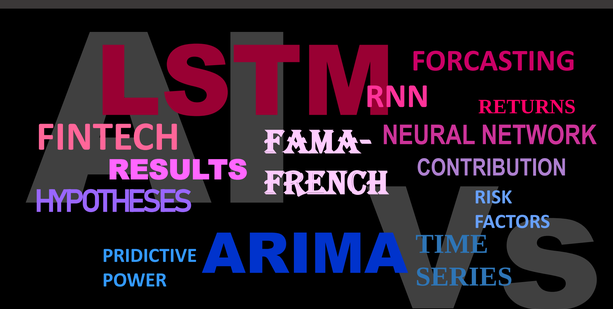
- **Background**: Deep learning architectures have been applied to a wide range of fields where they have brought many astonishing achievements and improvements. FinTech is a trend of future studies and activities in the finance field
- **Research Gap**: Financial data is a typical class of time series data; however, most studies focused on the traditional forecasting models which rely on a limited number of economic variables.

HOW

- We focus on the AI approach, building a well-designed LSTM network to predict the five risk factors
- The AI and classical finance approaches will be compared: LSTM results vs ARIMA results

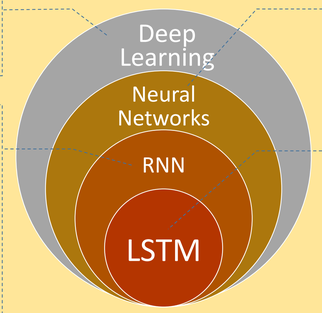
GOAL

- To demonstrate the predictive power of LSTM networks when applied to financial time series data
- Shed light on the applications of AI approaches in finance researches
- To further investigate the explanation power of Fama-French risk factors on the expected rate of return



LSTM

A class of machine learning algorithms that teach computers to learn and compute like human brain



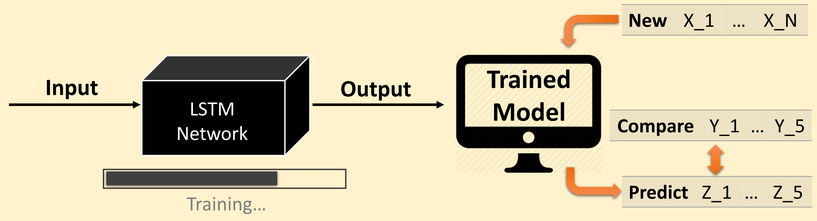
A computing system that processes information by mimicking the way that human neurons function

Recurrent Neural Network: A type of neural network that has "memory". This allows it to better process time sequences, exhibiting the connections among data from different time nodes
This is very meaningful to time-series datasets, such as financial data, whose current information reveals what happened in previous periods

A type of RNN with "gates", which regulate the amount of "memory" inflows and outflows. This benefits the whole process by effectively keeping the useful historical information
Compared with the normal RNN, LSTM has no limitation of lags' length between events, making the model very suitable for processing and predicting time series data

"Neural networks are commonly considered as black-box methods."

Day 1	X_1	...	X_N	Y_1	...	Y_5
Day 2	X_1	...	X_N	Y_1	...	Y_5
...	X_1	...	X_N	Y_1	...	Y_5
Day N	X_1	...	X_N	Y_1	...	Y_5



Training Set

Data sources: WRDS daily stock return, Fama-French five factor model dataset
Period: 1990-2010
X: raw training data – the daily returns of all North America firms listed in the dataset
Y: raw target data – the daily F-F risk factors

The training procedure is a process to teach the computer to learn from data. The basic question it tries to figure out is **what Y will be outputted with certain X.** This process will be done by thousands of computing iterations to finalize the relationship.
The LSTM net will **capture the complicated relationship** between **X and Y**, as well as between **previous status and current status**, that is hidden from human sense and the traditional models.

Testing Set

Data sources: WRDS daily stock return, Fama-French five factor model dataset
Period: 2011-2018
X: raw testing data – the daily returns of all North America firms listed in the dataset
Z: Predicted data – the daily F-F risk factors
Y: Raw comparing data – daily F-F risk factors

Methodology & Hypotheses

Hypotheses: LSTM network exhibits better predictive power on financial data than the classical financial approaches regarding two measures – MSE(Mean Square Error) and MAD(Mean Absolute Deviation)

H1: MSE(LSTM) < MSE(ARIMA)
H2: MAD(LSTM) < MAD(ARIMA)

ARIMA

ARIMA is a very classic and important model for time-series data forecasting
Data source: Fama-French five factor model dataset
Period: 1990-2010
Pre-test: to dig the characteristics of Fama-French risk factors, then decide the parameters in the ARIMA models
Structure: here we take the factor "SMB" as an example

Date	SMB
19900102	-0.67
19900103	0.72
19900104	0.44
...	...
20101230	0.13
20101231	-0.6

ARIMA

Date	Prediction	Real
20110103	0.33	0.57
20110104	-0.27	-1.32
20110105	0.44	0.61
20110106	0.12	-0.13
20110107	0.13	-0.25
...	...	...

Measures

$$MSE = \frac{1}{n}\sum_{k}(Z_k - Y_k)^2$$

$$MAD = \frac{1}{n}\sum_{k}|Z_k - Y_k|$$

Z_k is the predicted value output from either LSTM or ARIMA
Y_k is the actual historical value recorded on the Fama-French website
Comparisons of MSE and MAD will be conducted on five factors

Robustness Test

- Using LSTM to forecast Returns directly
- Fama-Macbeth
- Monte Carlo Analysis

Projected Results

- The LSTM network outperforms the ARIMA model with more accurate forecasting on the Fama-French risk factors
- With a better prediction of the risk factors, we look forward to a more trustful return calculation, also validating the Fama-French model

Conclusions

Limitation

- The black-box process will not give details about how the results are calculated. The curiosity of WHAT and HOW the stock returns are being influenced still cannot be answered

Further Research

- LSTM networks are very diversified, modifications can be applied to many parameters and their structures

Contributions

- The paper differs from previous studies by introducing a well-designed LSTM neural network as a forecasting vehicle of financial data
- Instead of making predictions directly on return which is the target of all the other researches, we forecast the risk factors which are the essential components of the rate of return.
- This study would confirm that the AI method of investigating financial forecasting has a large potential to be explored.

Acknowledgement

I deeply appreciate all the support and help from my supervisor, Dr. David Newton. His understanding, intelligence and acceptance are the best inspiration for me. Life is tough sometimes, but everything will be beautiful again for whatever we go through.

References

Babu, C. N., & Reddy, B. E. (2014). A moving-average filter based hybrid ARIMA-ANN model for forecasting time series data. *Applied Soft Computing Journal*, *23*, 27–38.
Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, *270*(2), 654–669.
Guresen, E., Kayakutlu, G., & Daim, T. U. (2011). Using artificial neural network models in stock market index prediction. *Expert Systems with Applications*, *38*(8), 10389–10397.
Khashei, M., & Bijari, M. (2010). An artificial neural network (p, d, q) model for timeseries forecasting. *Expert Systems with Applications*, *37*(1), 479–489.