

選擇適當模型以建立交易策略與預測股價

宋相鎬
應數系計算科學所數據組,
中興大學
台中市, 台灣
g108053112@mail.nchu.edu.tw

Abstract—為挖掘時間序列資料的商業價值，希望得到合適的模型以建立有效的交易策略。文中的7篇論文將幫助我了解，預測較為準確的幾個模型，並釐清模型背後的理論，與使用到的機器學習方法。

I. INTRODUCTION

前兩篇論文，以M競賽的結果為主軸[1][2]，介紹一些近期用於預測時間性資料的模型，接著選定從上述得知，有較好預測結果的模型，再用以下幾篇論文做進一步的探討。神經網絡(NN)的時間序列模型[3]，在宏觀經濟數據中有良好表現，透過高斯過程的演算法判斷股票的交易策略[4]，其中有使用到逆向強化學習(IRL)[5]方法，另外用股市取得的資料，比較幾個機器學習方法所建構出的模型[6][7]。

II. PAPER OVERVIEW

以下將分為7個小節，簡略介紹論文內容。

A. An Empirical Comparison of Machine Learning Models for Time Series Forecasting

M競賽是由Makridakis組織創辦，由尼科西亞大學的Spyros Makridakis教授所發起。論文中描述了2008年的M3競賽，比賽中使用的時間序列資料，來自各種領域類型，涵蓋工業、金融、人口統計等等，共有3003個資料集，每個資料集的資料個數是63到108個之間。

最終比賽結果，使用多層感知器(multilayer perceptron)或高斯過程(Gaussian process)所建構的模型，在各領域中皆有較好的預測結果。同時研究中發現，神經網絡往往優於傳統的線性模型，文中將進一步探討與比較各種機器學習模型，如：貝式神經網絡。

B. The M4 Competition: Results, findings, conclusion and way forward

時間來至2018年的M4競賽，M4是前三場M系列競賽的延續，此次比賽將資料集的時間序列資料個數增加為100000，大幅提升了電腦的計算量，以避免一些需要過份運算的模型。

本文介紹M4競賽的初步結果，而其中有幾個主要發現：

1. 在較為準確的17個方法中，有12種幾乎是只使用統計模型的方法
2. 預測最準的模型是，同時使用統計和機器學習的組合方法，其平均準確度(sMAPE)比其他模型高出近10%。
3. 預測第二準的模型是，由七種統計方法和一種機器學習組合而成的模型，其中的權重藉機器學習訓練得出，有效地提升其預測正確率。

C. Linear models, smooth transition auto regressions, and neural network s f or forecasting macroeconomic time series: A re-examination

在此篇論文中，將比較線性自回歸、平滑過渡自回歸(STAR)和神經網絡的時間序列模型(AR-NN)，對G7經濟體47個月的總體經濟數據預測。最終發現，線性自回歸的預測準確度較差，而在較長遠的時間預測中，使用貝式理論生成的AR-NN有更準確的預測，論文將討論以上的結果。

D. Gaussian process-based algorithmic trading strategy identification

現今有許多金融市場交易員使用程式自動下單，然而有一個關鍵點是，演算法在多大程度上理解，市場多數交易員的交易行為。

在此篇論文中使用E-Mini期貨的資料，藉由觀察限價單的交易量與掛單量，再透過模型來判斷買賣策略。其中模型使用逆向強化學習(IRL)，以此建構特徵空間，並分類或識別交易者，接著比較線性IRL和高斯過程IRL的優劣，最後對以上的預測結果做出討論。

E. Inverse Reinforcement Learning with Gaussian Process

模仿學習是機器學習的一個子領域，其目的是要學習模仿人類行為，僅需要觀察對象所採取的行動即可。模仿學習通常分為兩大類：

1. 使用監督學習的克隆方法：直接從對個體的觀察中預測動作環境。
2. IRL方法：通過比對較好的模型與現今模型的差異，學習得到獎勵函數，再利用獎勵函數進行強化學習，以提高模型策略水平。

此篇論文將進一步探討高斯過程的逆向強化學習。論文作者認為，在有限空間中的IRL，使用貝式理論會有較好的表現；然而在大到近似無限空間下的IRL，會遇到瓶頸，因而提出一個新的高斯過程IRL模型。

F. Benchmark dataset for mid-price forecasting of limit order book data with machine learning methods

論文中使用納斯達克北歐股票市場上，五隻股票連續10天的時間的限價單價格，約有4,000,000個數據個數的資料集。論文中使用回歸方法與Radial Basis Function Network搭配k-means的演算法做比較，希望對高頻率交易的價格走勢做出預測。

G. An integrated framework of genetic network programming and multi-layer perceptron neural network for prediction of daily stock return: An application in Tehran stock exchange market

論文使用德黑蘭證券交易所的9支股票價格作為資料，透過genetic network programming(GNP)、強化學習和多層感知器(MLP)神經網絡對數據進行分類，搭配使用時間序列模型來預測股價。最終比較GNP-ARMA與ARMA-GARCH模型(在收益估算和風險評估的一般研究中，所使用的時間序列模型)，發現GNP-ARMA模型少了近16%的平均預測錯誤。

REFERENCES

- [1] Nesreen K. Ahmed , Amir F. Atiya , Neamat El Gayar ,and Hisham El-Shishiny, "An Empirical Comparison of Machine Learning Models for Time Series Forecasting," *Econometric Reviews, Taylor Francis Journals*, Vol.29, No.5-6, p.94-621, 2010.
- [2] Spyros Makridakis, Evangelos Spiliotis, Vassilios Assimakopoulos, "The M4 Competition: Results, findings, conclusion and way forward," *INTERNATIONAL JOURNAL OF FORECASTING*, Vol.34, No.4, p.802-808, 2018
- [3] Timo Terasvirtaa, Dick van Dijk, Marcelo C.Medeiros, "Linear models, smooth transition auto regressions, and neural network s f or forecasting macroeconomic time series: A re-examination," *INTERNATIONAL JOURNAL OF FORECASTING*, Vol.21, No.4, p.755-774, 2005
- [4] Steve Y. Yang, Qifeng Qiao, Peter A. Beling, William T. Scherer Andrei A. Kirilenko, "Gaussian process-based algorithmic trading strategy identification," *QUANTITATIVE FINANCE*, Vol.15, No.10, Pages: 1683-1703, 2015
- [5] Qifeng Qiao and Peter A. Beling, "Inverse Reinforcement Learning with Gaussian Process," *IEEE, Proceedings of the 2011 American Control Conference*, 2011
- [6] Adamantios Ntakaris, Martin Magris, Juho Kannianen, Moncef Gab-bouj, Alexandros Iosifidis, "Benchmark dataset for mid-price forecasting of limit order book data with machine learning methods," *JOURNAL OF FORECASTING*, Vol.37, No.8, p.852-866, 2018
- [7] Reza Ramezani, Arsalan Peymanfar, Seyed Babak Ebrahimi, "An integrated framework of genetic network programming and multi-layer perceptron neural network for prediction of daily stock return: An application in Tehran stock exchange market," *APPLIED SOFT COMPUTING*, Vol.82, Article Number: UNSP 105551, 2019