高频金融交易中的价格变动预测建模及实现

开题报告

李若昊 2022年3月18日

研究内容

目的

- 预测某一只股票未来数个周期内价格的变化方向
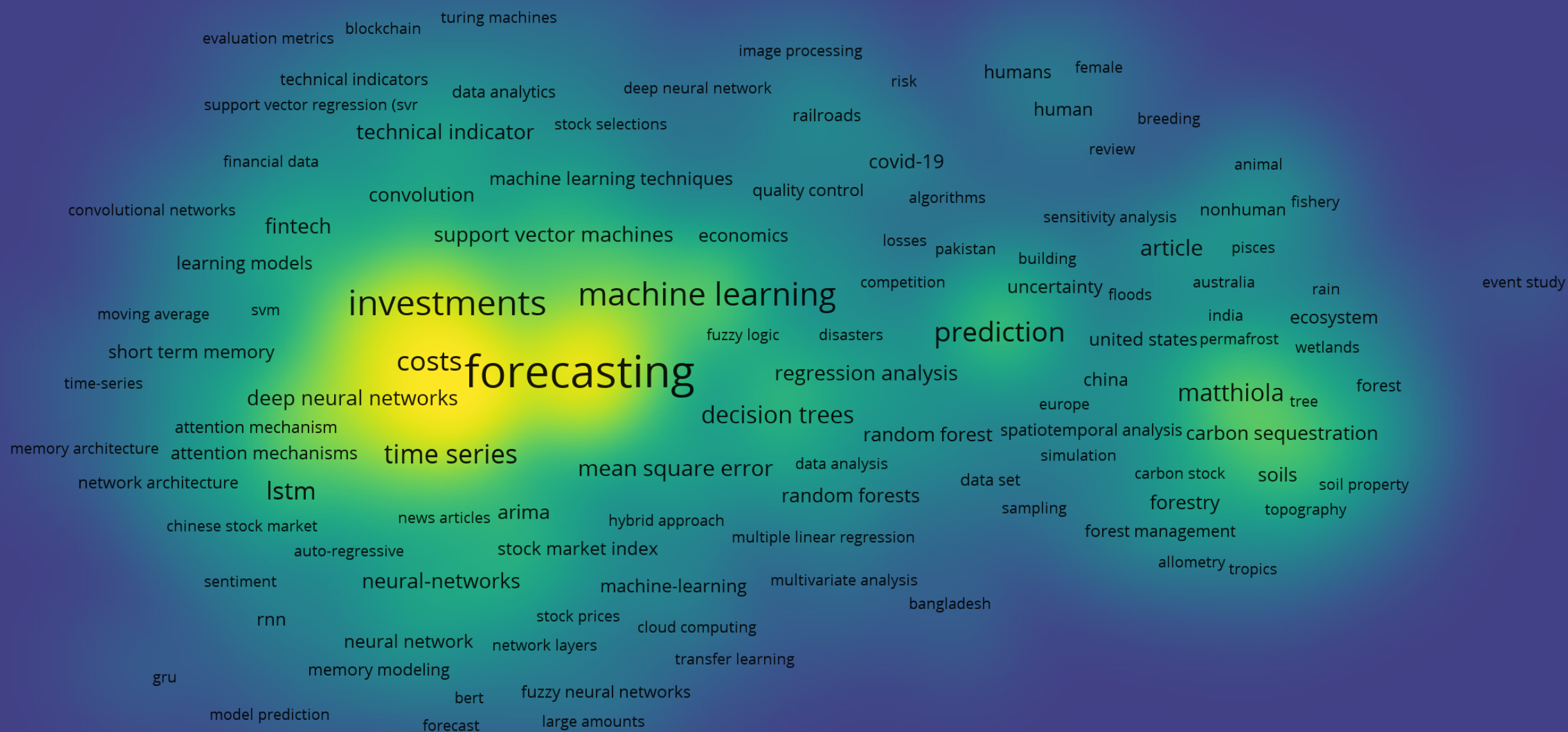
工具

- 循环神经网络、卷积神经网络

数据

- 公开的高频限价订单簿基准数据集 FI-2010

主要工作

- 建模、调参、训练

有大量相关文献支撑

时间	作者	题目
2018-02-13	Xingyu Zhou	Stock Market Prediction on High-Frequency Data Using Generative Adversarial Nets
2018-04-20	Bruno Miranda Henrique	Stock price prediction using support vector regression on daily and up to the minute prices
2018-12-20	Faisal Qureshi	Investigating Limit Order Book Characteristics for Short Term Price Prediction: a Machine Learning Approach
2019-03-25	Zihao Zhang	DeepLOB: Deep Convolutional Neural Networks for Limit Order Books
2020-03-23	AH Bukhari	Fractional Neuro-Sequential ARFIMA-LSTM for Financial Market Forecasting
2020-05-19	Adamantios Ntakaris	Mid-price prediction based on machine learning methods with technical and quantitative indicators
2020-08-06	ZinebLanbouri	Stock Market prediction on High frequency data using Long-Short Term Memory
2020-08-09	YuChen Tu	Predicting High-Frequency Stock Market by Neural Networks
2020-10-19	Yuechun Gu	Price Forecast with High-Frequency Finance Data
2021-05-18	Konark Yadav	Stock values predictions using deep learning based hybrid models
2021-06-16	Muye Wang	Essays on the Applications of Machine Learning in Financial Markets
2021-06-21	Yue Yang	Stock Price Prediction Based on XGBoost and LightGBM
2021-07-26	Liang Zeng	Trade When Opportunity Comes Price Movement Forecasting via Locality-Aware Attention and Adaptive Refined Labeling
2021-08-07	Xuerui Lv	Residual Gated Recurrent Unit-Based Stacked Network for Stock Trend Prediction from Limit Order Book

研究工具

为什么？

- 数据量大且精度要求高：传统机器学习方法有性能瓶颈

适合吗？

- RNN—擅长处理时序数据

- CNN—独特的特性令我想要结合

依托数据

- 已标注数据——适合监督学习

- 大规模——500ms一笔 5支股票合计约40万条的快照行情

- 高质量——10档买卖数据信息

工作内容

不能提出新模型算法、目前没有能力研发框架…那?

注重模型是怎么训练出来的，如

- 超参数的调整

- 试解决过拟合的问题

- 为什么选择某个优化器?

- 学习速率为何设置为这个值?

- 通过阅读有关文献添加手工特征

- 对结果的分析

自身能力

- 单变量微积分 94
  多变量微积分 82
  线性代数 94
  机器学习 87.6
  数据挖掘 88

- MIT 18.01 18.02 18.06
  Coursera Deep Learning Specialization

- IELTS Reading 8.5

- 知识储备可以应付

评测方法

指标

• 准确度与F-score

目标

• 比数据集源论文中给出的基线准确率高

时间安排

表1 - 时间安排

事项	开始周	结束周	5	6	7	8	9	10	11	12	13	14	15	16	17
设计/调试模型	5	9	■	■	■	■	■								
中期检查	10	10						■							
完成后期实验	11	12							■	■					
撰写/修改初稿	13	14									■	■			
接受查重	15	15											■		
接受评阅	16	16												■	
毕业答辩	17	17													■