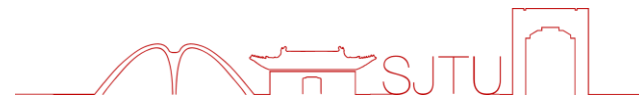




上海交通大学  
SHANGHAI JIAO TONG UNIVERSITY



# 基于机器学习的预测与量化因子开发

2022.12.23

饮水思源 · 爱国荣校



## ◆ 课题目标：

基于股票高频交易行为对未来收益率有一定的预测作用这一假设，希望通过一定的深度学习模型，挖掘股票高频交易数据与股票未来收益率之间的隐藏关系，建立预测模型，对未来一段时间的个股截面收益进行预测。



## ◆ 课题任务：

针对高频数据、文本数据等，开发能够对未来预期收益率有一定预测能力的信号。具体包括三方面：

- ◆ 高频截面因子的开发；
- ◆ 文本因子的开发；
- ◆ 其他因子的开发。

一方面期望通过**机器学习模型**能够提高因子研发的效率和预测准确性；另一方面，通过**文本因子**的研究，开拓另外数据因子，提高Alpha的预测的广度和深度。



## ◆ 高频截面因子的开发

L2行情数据的因子开发。通过研究L2行情的订单簿数据、逐笔成交、报单以及集合竞价数据的研究，通过Python开发，包括但不限于，通过机器学习模型进行Alpha预测模型的开发等工作。



## ◆ 文本因子的开发

舆情因子的开发。通过研究卖方报告，新闻文本，股票公告数据等与股票相关的新闻舆情文本类数据，对于个股的未来收益有一定的正面或负面影响，希望建立NLP模型，对相关文本数据进行分析，输出标准化打分，使得得分与股票未来收益之间具有较强的相关性。



## ◆ 其他因子的开发

通过机器学习模型获取新的因子，并进行建模，构建Alpha模型。



## ◆ 课题的研究思路及关键内容：

- ◆ CNN+LSTM;
- ◆ Transformer;
- ◆ 高频因子的学习;
- ◆ L2数据的学习;
- ◆ 文本内容分析;