Enhancing Portfolio Optimization with Data Fusion and Machine Learning in Quantitative Finance

温州肯恩大学
WENZHOU-KEAN UNIVERSITY

Advisor:
Dr. Baha Ihnaini

Designed by:
Yixuan Mi
Zhentong Ye
Jiashu Li
Xianghao Zheng
Anxin Chen

CST
Wenzhou-Kean University
COLLEGE OF SCIENCE AND TECHNOLOGY

Introduction:

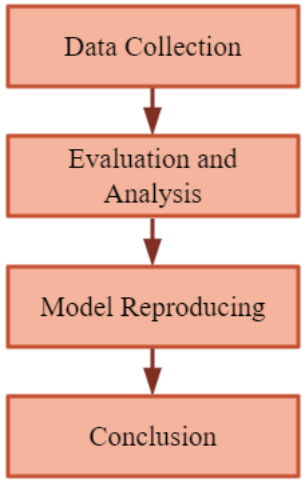
This project aims to revolutionize portfolio optimization in quantitative finance by integrating advanced machine learning techniques and data fusion methodologies. Our team strives to improve the problem that traditional methods will face: limitations in handling market complexities and integrating diverse data effectively.

Data Collection:

CSMAR (China Stock Market & Accounting Research) is a comprehensive database jointly established by Chinese regulatory authorities and the Shanghai Stock Exchange, providing data on Chinese A-share market companies for research and analysis.

Based on the CSMAR platform, we use CSMAR's downloadable CSV files to collect historical financial asset data and market sentiment data such as financial statements, investor sentiment indices, and sentiment consistency.

Research Flowchart:



Platforms Support:

PyCharm JetBrains IDE

CSMAR

Qlib

SSH Connection

MobaXterm

Model Reproducing:

Once we collected all the data, we used "Qlib" to choose the high-quality model (like LSTM) to fit our data; then, we used remote development tools (MobaXterm, Pycharm, etc.) to reproduce and improve the model.

- Signal-based evaluation: IC, ICIR, Rank IC, Rank ICIR
- $\text{corr}(\mathbf{x}, \mathbf{y}) = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}$
- $\text{IC}^{(t)} = \text{corr}(\hat{\mathbf{y}}^{(t)}, \mathbf{ret}^{(t)})$
- $\text{ICIR} = \frac{\text{mean}(\mathbf{IC})}{\text{std}(\mathbf{IC})}$
- $\text{Rank IC}^{(t)} = \text{corr}(\text{rank}(\hat{\mathbf{y}}^{(t)}), \text{rank}(\mathbf{ret}^{(t)}))$
- $\text{Rank ICIR} = \frac{\text{mean}(\mathbf{Rank\ IC})}{\text{std}(\mathbf{RankIC})}$

Results and Conclusion:

Model Name	Dataset	IC	ICIR	Rank IC	Rank ICIR
LSTM	Alpha360	0.0478±0.01	0.3620±0.05	0.0585±0.00	0.4578±0.04
ADD	Alpha360	0.0419±0.00	0.3066±0.04	0.0550±0.00	0.4205±0.03
ADARNN	Alpha360	0.0468±0.01	0.3706±0.08	0.0544±0.01	0.4416±0.07

Through trial and error, our team has reproduced three models, and each model was successfully run 50 times. It proves that our plan is working, paving the way for further advancements in our field.

Reference:

Song, C. (2023). Portfolio Optimization Based on Machine Learning. Advances in Economics, Management and Political Sciences. https://doi.org/10.54254/2754-1169/25/20230500.

Wang, Y. (2023). Review: Application of Machine Learning to Investment Portfolios. BCP Business & Management. https://doi.org/10.54691/bcpbm.v38i.4351.

Our GitHub Repository:

https://github.com/EthanYixuanMi/Machine-Learning-in-Quantitative-Finance