Machine Learning Applications in Asset Management

Tugce Karatas, Columbia University

Satyan Malhotra, Flexstone Partners

Columbia University Data Science Institute

Columbia Engineering
The Fu Foundation School of Engineering and Applied Science

Bloomberg

Machine Learning Applications in Asset Management

Tugce Karatas (PhD candidate, Columbia University)
Satyan Malhotra (Managing Partner, Flexstone Partners*)

Advisor: Dr. Ali Hirsa
IEOR Department
Columbia University

**This presentation reflects the views and opinions of the individual authors at this date and in no way the official position or advices of any kind of Flexstone Partners, LLC (the "Firm") and thus does not engage the responsibility of the Firm nor of any of its officers or employees.*

Outline

- ➢ Overview
- ➢ Liquid products - Mutual Funds
- ➢ Illiquid products - Private Equity
- ➢ Conclusion

Machine Learning applications in Asset Management are here

Economic Regimes Identification Using Machine Learning Techniques

Machine Learning for Stock Selection

Deep Learning in Asset Pricing

Adaptive Portfolio Asset Allocation Optimization with Deep Learning

Classifying Mutual Funds based on Relative Performance using Artificial Neural Networks

Using Deep Learning to Detect Price Change Indications in Financial Markets

A Backtesting Protocol in The Era of Machine Learning

Machine Learning Algorithms for Financial Asset Price Forecasting

Deep Learning for Finance: Deep Portfolios

ARTIFICIAL INTELLIGENT BASED ASSET MANAGEMENT

Machine Learning's base pillars

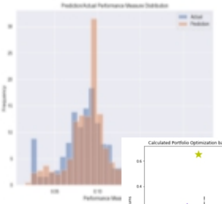
DATA

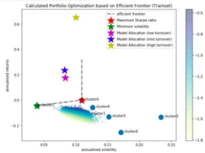
ANALYTICAL TOOLS

VISUALIZATION

- Filings
- Alternative Data
- Holdings Data
- News
- Price
- Information
- People
- Company

Source: http://batlab.web.unc.edu





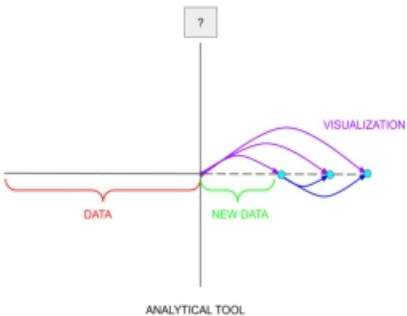
Data

Data is there but, needs to be carefully processed to be useful

- Accessing datasets
- Merging disparate datasets
- Imputing missing data
- Interpolating data
- Over-sampling or down-sampling to obtain balanced classes
- Feature engineering and dimension reduction
- Categorical data
- Managing non-stationary data

Analytical Tools

Q: What is the asset management application?



Signals
- How does the price / volume move?
- Which sector / asset is going to outperform?
- Which manager is better?
- What is the product's success rate?

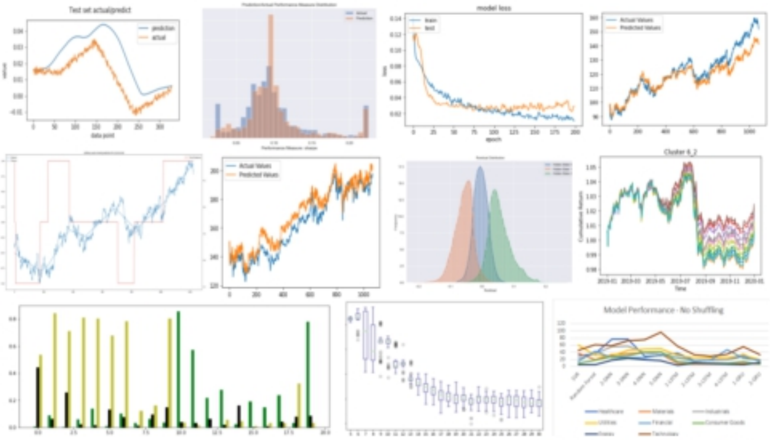
Sentiment
- Is there an up or down trend?
- What is the closing probability?

Scrapping
- What is the real time impact of news?

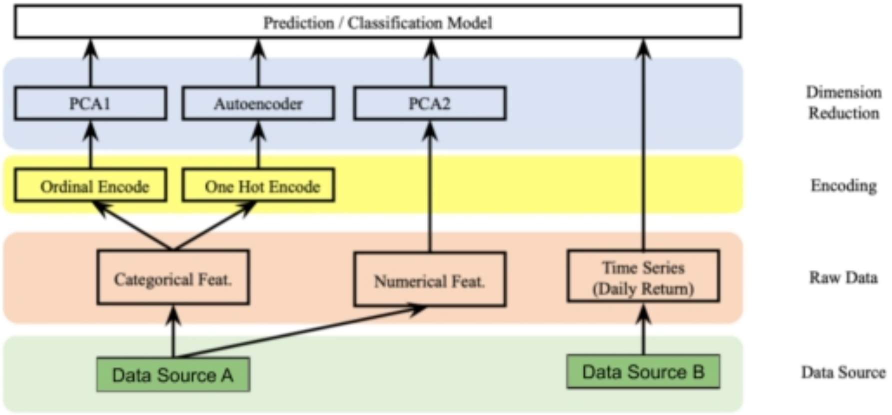
Clustering
- What should the allocation be?
- What are the cohorts and replications?
- What are the key questions?

Visualization

Q: What is the best visual representation for the application?



Base Framework



Outline

- ➢ Overview
- ➢ Liquid products - Mutual Funds
- ➢ Illiquid products - Private Equity
- ➢ Conclusion

Mutual Funds – Framework

- Manager Selection
  - Features: Alternative data, holding data
  - Optimization: LSTM
- Portfolio Construction / Asset Allocation
  - Traditional methods
  - Optimization, DL, Clustering, Reinforcement Learning
- Stress Testing / Real Time Impacts
- Visualization
- Market Indicators
  - Macroeconomic
  - Jump Prediction
  - Sentiment data

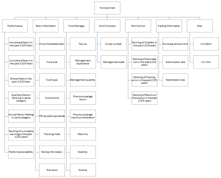
Tugce Karatas

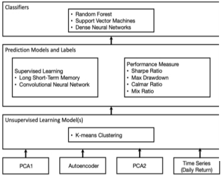
Mutual Funds – Manager Selection

Motivation:

Which managers “may” outperform given objective functions?

Data | Processing | Analytical Tools/Framework | Objective Function Visualization Example

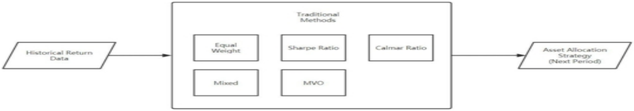




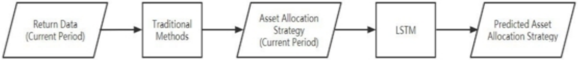
Mutual Funds – Portfolio Construction / Asset Allocation

Motivation:
How to allocate assets given objective functions?

Traditional methods – asset allocation is primarily done based on historical returns data – given risk/return appetite.



Machine Learning - In LSTM, the input is w = [wi ,t], for all funds given a time horizon.



Mutual funds – Traditional / EF

Introduced by Harry Markowitz in 1952.

Motivation

Maximize expected return of a portfolio for a given level of risk

$$\begin{aligned} \min_{w} \quad & \frac{1}{\lambda} w^T \Sigma w - w^T \mu \\ \text{s.t.} \quad & w^T e = 1 \\ & w_i \geq 0 \quad \text{for} \quad i = 1, \cdots, N \end{aligned}$$

where $\Sigma = [\sigma_{i,j}]_{i,j=1}^{N}$ is correlation matrix, $\mu = [\mu_i]_{i=1}^{N}$ is return vector, and λ is risk-tolerance factor.

Mutual funds – Traditional

Equal Weight Allocation

$w_i = \frac{1}{N}$, for $i = 1, \cdots, N$

w_i is the weight of mutual fund i , and N is the number of mutual funds in the portfolio.

Sharpe Ratio Based Allocation

Sharpe Ratio measures the performance of an asset compared to risk-free asset and adjusted for its risk.

$SR_i = \frac{(r_i - \bar{r})}{\sigma_i}$, and $w_i = \frac{SR_i}{\sum_{j=1}^{N} SR_j}$, for $i = 1, \cdots, N$

Mutual funds – Traditional

Calmar Ratio Based Allocation

Calmar Ratio uses max drawdown to measure the risk and combines it with profit and loss.

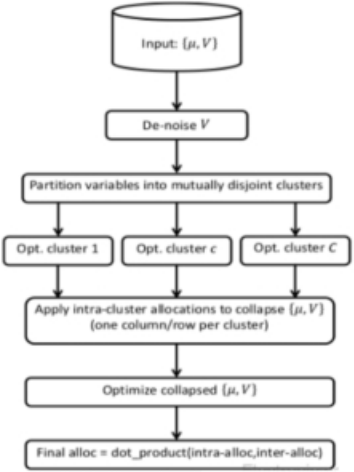
$CR_i = \frac{PL_i}{MDD_i}$, and $w_i = \frac{CR_i}{\sum_{j=1}^{N} CR_j}$, for $i = 1, \cdots, N$

Mixed Ratio Based Allocation

Mixed Ratio combines Sharpe Ratio and Calmar Ratio.

$MR_i = \frac{PL_i}{\sigma_i \cdot MDD_i}$, and $w_i = \frac{MR_i}{\sum_{j=1}^{N} MR_j}$, for $i = 1, \cdots, N$

Mutual Funds – Sample Nested Clustered Optimization



Lopez de Prado, Marcos. ”A Robust Estimator of the Efficient Frontier.” Available at SSRN 3469961 (2019).

Mutual Funds – Reinforcement Learning

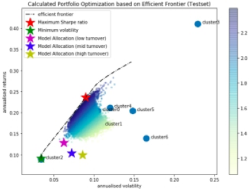
- After building LSTM models for weights, we need to come up with the final weights of the portfolio.
- β(i , t) is combination weights.
- w (i , t) is predicted portfolio weights from each LSTM model.
- **Reinforcement Learning Formula:**

$$w_t = \sum_{i=1}^{5} \beta(i,t) \cdot w(i,t)$$

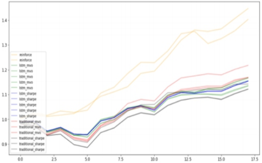
$$\text{where } \beta(i,t) = \frac{\frac{1}{\epsilon(i,t)}}{\sum_{j=1}^{5} \frac{1}{\epsilon(j,t)}}, \epsilon(i,t) = |w_{true}(i,t) - w_{LSTM}(i,t)|$$

Mutual Funds – Asset Allocation Visualization

Simulating future performance of the portfolios



Backtesting the asset allocation strategy by historically simulating its performance



Mutual Funds – Stress Testing / Real Time Impacts

Motivation

Testing the market conditions and portfolios against historic scenarios and news to assess impacts

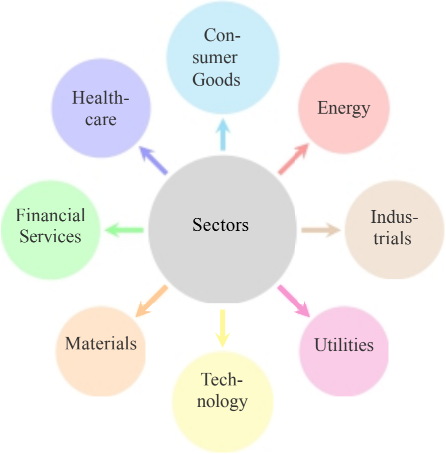
- ➢ We can do stress testing in order to see the resilience of the models under possible crisis like 2008 economic crisis, Covid-19, etc. As an example, a stable ranking system for Market Indicators can test the models under current market situation to see the effect of Covid-19.
- ➢ NLP and sentiment can be leveraged to assess the potential impact on sectors, portfolios, other

Market Indicators

Motivation

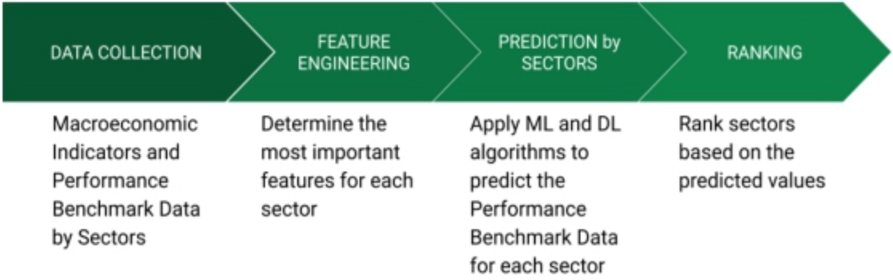
Identifying the relative potential and attractiveness of sectors, sub-sectors, other given the market and economic conditions

Market Indicators – Data

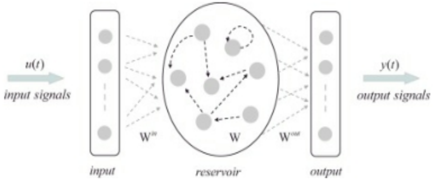


- Classify sectors, sub-sectors, other
- Compile macro and cohort specific data
- Train models for common macro-economic and sector-based indicators over various time horizons

Market Indicators – Analytical Tools



Market Indicators – Echo State Networks Application



Source: https://doi.org/10.1016/j.ins.2016.08.081

$$x(n+1) = f(Wx(n) + W^{in}u(n+1))$$

$$y(n+1) = f^{out}(W^{out}x(n+1))$$

Jaeger, Herbert. "The "echo state" approach to analysing and training recurrent neural networks-with an erratum note." Bonn, Germany: German National Research Center for Information Technology GMD Technical Report 148.34 (2001): 13.

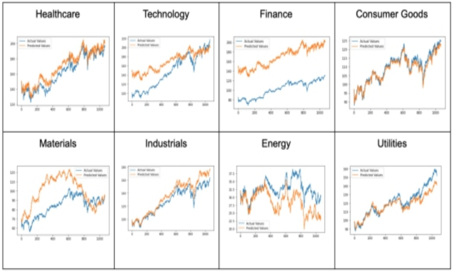
- The reservoir with fixed weights solves the vanishing gradient problem in traditional RNNs.
- The output is a linear combination from the reservoir. Therefore, linear regression algorithms can be used to predict the output weights.
- It works faster than traditional RNNs.

Market Indicators – Extensions

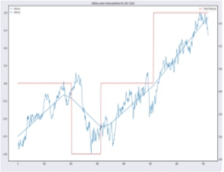
- Jump Prediction
  - Objective: Detecting the effect of unexpected conditions on the models and its effect on the ranking
    Methodology:
    - Filtering the data with H-P filtering
    - Signal prediction on the filtered data
- Sentiment Analysis
  - Objective: Detecting the effect of news on the models and its effect on the ranking
    Methodology:
    - News data collection from different sources
    - Labeling the data Loughran and McDonald Dict
    - BERT model vs TFIDF model

Market Indicators – Visualization Examples

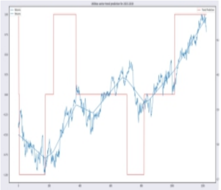
Sector trend simulation



Utilities 3-year trend simulation



Utilities 5-year trend simulation



Outline

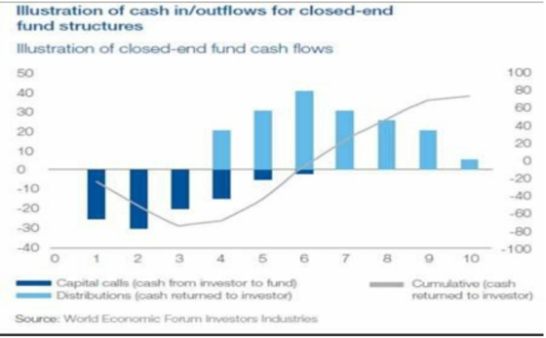
- ➢ Big Picture
- ➢ Liquid products - Mutual Funds
- ➢ Illiquid products - Private Equity
- ➢ Conclusion

Illiquid products – Private Equity

- ➢ Manager Selection*
- ➢ Investment Selection*
- ➢ Market Indicators*
- ➢ Cash Flow Forecasting

* Unique dataset but, extension of the Analytical Tools and Visualization from the Liquid Products

Illiquid products – Cash Flow Forecasting



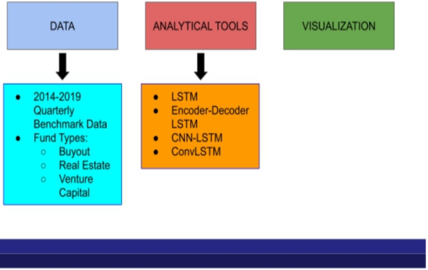
- ➢ Cash Flow questions
  - ➢ What are the contribution and distributions profiles?
  - ➢ What impact do unplanned events have on the profile?
  - ➢ How close is the profile to the generic sectors or sub-sectors profiles?
  - ➢ What is the tracking error and reinforcement method?
- ➢ Data challenges
  - ➢ Sparse and difficult to access
  - ➢ Not standard
  - ➢ Infrequent updates

Illiquid Products – Illustrative Traditional Models

- Takahashi, Dean, and Seth Alexander. "Illiquid alternative asset fund modeling." The Journal of Portfolio Management 28.2 (2002): 90-100.
    - The model is discrete-time and deterministic.
    - There are certain assumptions and input parameters to be estimated
    - Contributions and distributions are dependent.
    - It is difficult to update to the recent data.

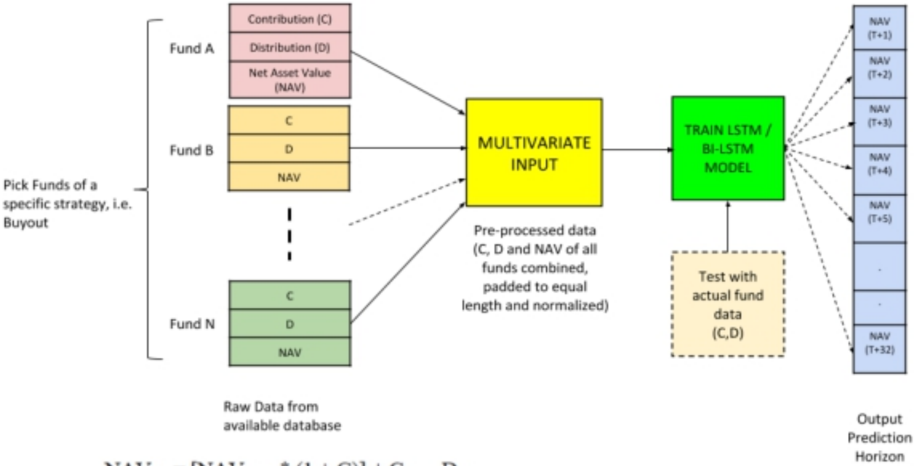
- Buchner, Axel, Christoph Kaserer, and Niklas Wagner. Stochastic modeling of private equity: an equilibrium based approach to fund valuation. No. 2006-02. CEFS working paper series, 2006.
    - It is continuous-time and stochastic
    - There are 2 independent stochastic process for Capital Contributions and Distributions.
    - It allows performing risk analysis
    - Rate of contribution is modelled with mean-reverting square-root process
    - Certain assumptions: Capital distributions follow lognormal distribution.

Illiquid Products – Machine Learning Applications



- Both models have used certain assumptions. ML allows us to build assumption-free models.
- Key Challenge: Inconsistent and insufficient data
- Each fund type has different characteristics. We need to have separate models for each fund type.
- Initially, historical benchmarks for each fund type are used for cash flow forecasting.
- Contributions and distributions are predicted independently.

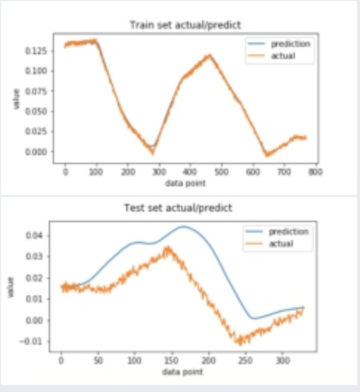
Illiquid Products – Machine Learning Applications

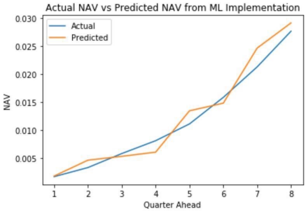


$$NAV_{(t)} = [NAV_{(t-1)} * (1 + G)] + C_{(t)} - D_{(t)}$$

Illiquid Products – Visualization Sample

Data: 2014-2019 quarterly benchmark data
Interpolation: Brownian Bridge
Model: CNN-LSTM Model





Outline

- ➢ Overview
- ➢ Liquid products - Mutual Funds
- ➢ Illiquid products - Private Equity
- ➢ Conclusion

Conclusion

- Machine Learning Applications in Asset Management is a huge space to explore!

- There is a broad range of projects* we are working on, but all the projects are based on the Machine Learning base pillars: Data, Analytical Tools, and Visualization.

- Research and applicability of the solutions highlight the edge market participants can leverage.

* Columbia research on all topics covered in the presentation is available.

THANK YOU