Thesis update

2025/2/24

Agenda

1. Data
2. Experiment
3. Measurement*
4. Derive equation

FAR-Trans: An Investment Dataset for Financial Asset Recommendation

Abstract

Financial asset recommendation (FAR) is a subdomain of recommender systems which identifies useful financial securities for investors, with the expectation that they will invest capital on the recommended assets. FAR solutions analyse and learn from multiple data sources, including time series pricing data, customer profile information and expectations, as well as past investments. However, most models have been developed over proprietary datasets, making a comparison over a common benchmark impossible. In this paper, we aim to solve this problem by introducing FAR-Trans, the first public dataset for FAR, containing pricing information and retail investor transactions acquired from a large European financial institution. We also provide a bench-marking comparison between eleven FAR algorithms over the data for use as future baselines. The dataset can be downloaded from https://doi.org/10.5525/gla.researchdata.1658.

Item Type:	Conference or Workshop Item
Additional Information:	The work introduced in this paper was in part carried out within the Infinitech project which is supported by the European Union's Horizon 2020 Research and Innovation programme under grant agreement no. 856632. Subsequent development was also financially supported via Engineering and Physical Sciences Research Council (EPSRC) Impact Accelerator, part of UK Research and Innovation (UKRI) with grant ref. number EP/X525716/1.
Status:	Published
Refereed:	Yes
Glasgow Author(s) Enlighten ID:	Mccreadie, Dr Richard and Sanz-Cruzado Puig, Dr Javier
Authors:	Sanz-Cruzado, J., Droukas, N., and Mccreadie, R.
College/School:	College of Science and Engineering > School of Computing Science
Copyright Holders:	Copyright © 2024 IJCAI
Publisher Policy:	Reproduced with the permission of the publisher
Related URLs:	• Organisation
Data DOI:	10.5525/gla.researchdata.1658

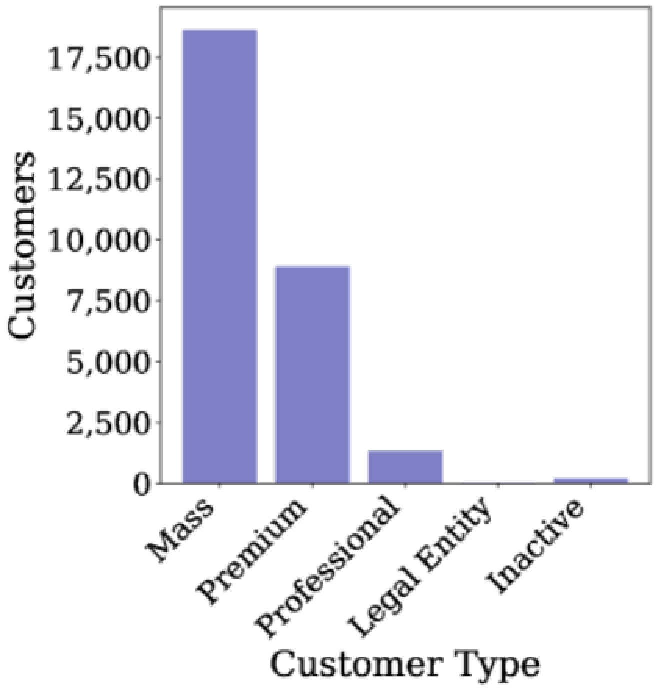
FAR-Trans: An Investment Dataset for Financial Asset Recommendation - Enlighten Publications

Transaction with ISIN code and Customer ID

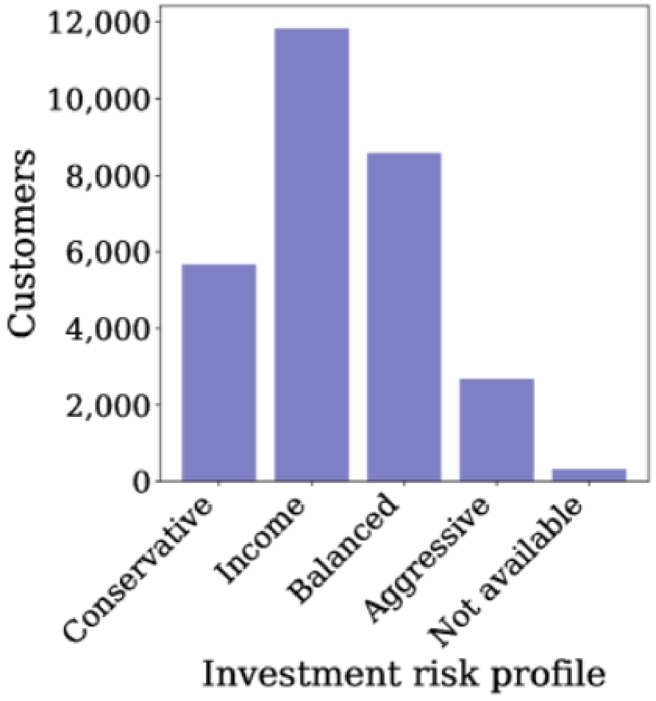
customerID varchar	ISIN varchar	transactionID int64	transactionType varchar	timestamp varchar	totalValue double	units double	channel varchar	marketID varchar
00017496858921195E5A	GRS434003000	7590224	Buy	2020-03-27	11000.0	5000.0	Internet Banking	XATH
00017496858921195E5A	GRS434003000	7607029	Sell	2020-04-06	12080.0	5000.0	Internet Banking	XATH
00017496858921195E5A	GRS434003000	7634872	Buy	2020-04-24	13400.0	5000.0	Internet Banking	XATH
00017496858921195E5A	GRS434003000	7652627	Sell	2020-05-07	12700.0	5000.0	Internet Banking	XATH
00017496858921195E5A	GRS434003000	7664807	Buy	2020-05-15	12150.0	5000.0	Internet Banking	XATH
00017496858921195E5A	GRS434003000	7674608	Sell	2020-05-22	14470.0	5000.0	Internet Banking	XATH
00017496858921195E5A	GRS434003000	7709457	Buy	2020-06-10	17080.0	5000.0	Internet Banking	XATH
00017496858921195E5A	GRS434003000	7754022	Sell	2020-07-01	17320.0	5000.0	Internet Banking	XATH
00017496858921195E5A	GRS434003000	7910112	Buy	2020-10-26	20000.0	4000.0	Internet Banking	XATH
00017496858921195E5A	GRS434003000	8263985	Sell	2021-01-05	29760.0	4000.0	Internet Banking	XATH

no_row int64	no_cust int64	no_stock int64	no_txn int64	min_date varchar	max_date varchar
388048	29090	320	359128	2018-01-02	2022-11-30

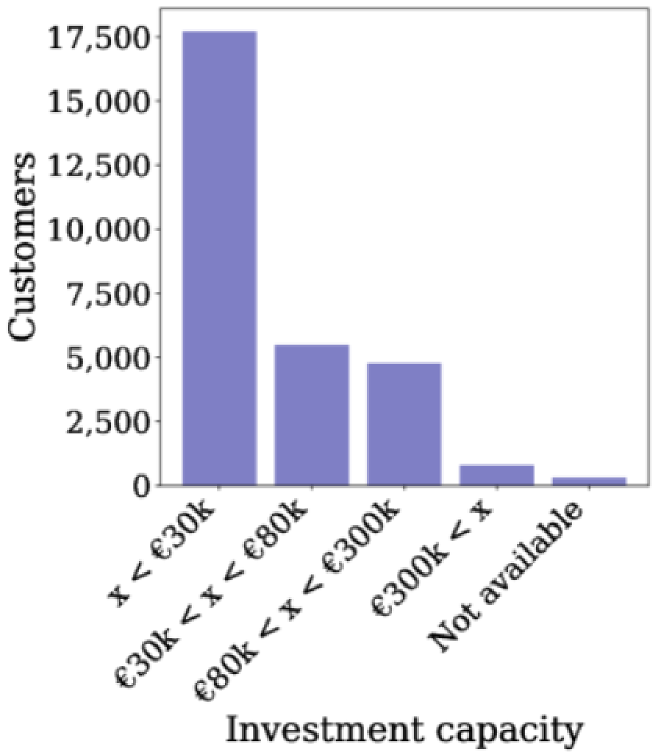
Customer Risk Profile: Parameter for Behavioral portfolio optimization



(a) Customer classification by segment



(b) Customer classification by risk profile



(c) Customer classification by capacity

5-Fold Forward Chaining CV (User-Item)

201801	...	201912	202006	202012	202106	202112	202206
Train			Val				
Train				Val			
Train					Val		
Train						Val	
Train							Val

Avg. MAP@K

Dataset Temporal Splitting: This dataset spans 39 months (just over 3 years). The effectiveness of different recommendation algorithms will naturally vary as market conditions change (as we will demonstrate later). Hence, it is important to examine how performance varies over time if we are to gauge more accurately when and where different recommendation strategies succeed and fail. To this end, we divide our dataset into 29 distinct variants, each representing a recommendation setting for a different point in time. Each variant defines a time point when recommendations are produced $t \in T$, with a pricing data and investment transactions recorded prior to t available for model training/validation, and the pricing data and investment transactions made after t being used

Measurement

- Transactional based
  - MAP@K
- Profit based
  - ROI, Sharpe ratio …

```
pred_top_k_mf = get_top_k(pred_array_mf, 5)
mapk(val_array, pred_top_k_mf, k=5)
[54] ✓ 0.0s
... 0.0034000000000000002

pred_top_k_mmf = get_top_k(pred_array_mmf, 5)
mapk(val_array, pred_top_k_mmf, k=5)
[55] ✓ 0.0s
... 0.0006666666666666666

pred_top_k_wmf = get_top_k(pred_array_wmf, 5)
mapk(val_array, pred_top_k_wmf, k=5)
[56] ✓ 0.0s
... 0.8055

pred_top_k_mwmf = get_top_k(pred_array_mwmf, 5)
mapk(val_array, pred_top_k_mwmf, k=5)
[57] ✓ 0.0s
... 0.8132
```

Equation (see cf_model.py)

1. SGD MF (Done)
2. SGD Mean MF (Done)
3. SGD WMF (Done)
4. SGD Mean WMF (Done)
5. SGD Mean-Variance MF (TBD)
6. SGD Mean-Variance-Prospect-Theory MF (TBD)
7. SGD Mean-Variance WMF (TBD)
8. SGD Mean-Variance-Prospect-Theory WMF (TBD)

Equation

$$\min_{p_u, q_i} c_{ui}(y_{ui} - p_u q_i)^2 + \lambda_{L2}(||p_u||^2 + ||\mathrm{q_i}||^2)$$

$$\min_{p_u, q_i} c_{ui}(y_{ui} - p_u q_i)^2 - \lambda_{MV}(p_u q_i \mu_i) + \lambda_{L2}(||p_u||^2 + ||\mathrm{q_i}||^2)$$

(Mean)

$$\min_{p_u, q_i} c_{ui}(y_{ui} - p_u q_i)^2 - \lambda_{MV}(p_u q_i \mu_i) + \lambda_{MV}\left((p_u q_i)^2 \sigma_i^2 + p_u q_i \Sigma_{j:j \neq i} p_u q_j \sigma_{uj}\right) + \lambda_{L2}(||p_u||^2 + ||\mathrm{q_i}||^2)$$

(Mean) *(Variance-Covariance)*

$$\min_{p_u, q_i} c_{ui}(y_{ui} - p_u q_i)^2 - \lambda_{MV}(p_u q_i \mu_i) + \lambda_{MV}\left((p_u q_i)^2 \sigma_i^2 + p_u q_i \Sigma_{j:j \neq i} p_u q_j \sigma_{uj}\right) - \lambda_{PT}(p_u q_i \pi_{ui} V_{ui}) + \lambda_{L2}(||p_u||^2 + ||\mathrm{q_i}||^2)$$

(Mean) *(Variance-Covariance)* *(Prospect)*
or any theory..

Equation (p_u)

$$\frac{dL}{dp_u} = -2c_{ui}(y_{ui} - p_u q_i)q_i - \lambda_{MV}(q_i \mu_i) + \lambda_{MV}(2p_u q_i^2 \sigma_i^2 + 2p_u q_i \sum q_j \sigma_{uj}) - \lambda_{PT}(q_i \pi_{ui} V_{ui}) + \lambda_{L2}\,(2p_u)$$

Take -1/2

$$\frac{dL}{dp_u} = c_{ui}(y_{ui} - p_u q_i)q_i + \lambda_{MV}\frac{1}{2}q_i \mu_i - \lambda_{MV}(p_u q_i^2 \sigma_i^2 - p_u q_i \sum q_j \sigma_{uj}) + \lambda_{PT}\frac{1}{2}q_i \pi_{ui} V_{ui} - \lambda_{L2} p_u$$

$$X = X - lr * \frac{d}{dX}f(X)$$

$$X = X + \; lr * c_{ui}(y_{ui} - p_u q_i)q_i + \lambda_{MV}\frac{1}{2}q_i \mu_i - \lambda_{MV}(p_u q_i^2 \sigma_i^2 - p_u q_i \sum q_j \sigma_{uj}) + \lambda_{PT}\frac{1}{2}q_i \pi_{ui} V_{ui} - \lambda_{L2} p_u$$

Equation (q_i)

$$\frac{dL}{dp_u} = -2c_{ui}(y_{ui} - p_u q_i)p_u - \lambda_{MV}(p_u \mu_i) + \lambda_{MV}(2p_u^2 q_i \sigma_i^2 + p_u^2 \sum q_j \sigma_{uj}) - \lambda_{PT}(p_u \pi_{ui} V_{ui}) + \lambda_{L2}\,(2q_i)$$

Take -1/2

$$\frac{dL}{dp_u} = c_{ui}(y_{ui} - p_u q_i)p_u + \lambda_{MV}\frac{1}{2}p_u \mu_i - \lambda_{MV}(p_u^2 q_i \sigma_i^2 - \frac{1}{2}p_u^2 \sum q_j \sigma_{uj}) + \lambda_{PT}\frac{1}{2}p_u \pi_{ui} V_{ui} - \lambda_{L2}\,q_i$$

$$X = X - lr * \frac{d}{dX}f(X)$$

$$X = X + \ lr * c_{ui}(y_{ui} - p_u q_i)p_u + \lambda_{MV}\frac{1}{2}p_u \mu_i - \lambda_{MV}(p_u^2 q_i \sigma_i^2 - \frac{1}{2}p_u^2 \sum q_j \sigma_{uj}) + \lambda_{PT}\frac{1}{2}p_u \pi_{ui} V_{ui} - \lambda_{L2}\,q_i$$

Final gradient descent equation

$$X_{pu} = X_{pu} + \ lr * c_{ui}(y_{ui} - p_u q_i)q_i - \lambda_{MV} p_u q_i^2 \sigma_i^2 - \lambda_{MV} p_u q_i \sum q_j \sigma_{uj} + \lambda_{MV} \frac{1}{2} q_i \mu_i + \lambda_{PT} \frac{1}{2} q_i \pi_{ui} V_{ui} - \lambda_{L2} p_u$$

$$X_{qi} = X_{qi} + \ lr * c_{ui}(y_{ui} - p_u q_i)p_u - \lambda_{MV} p_u^2 q_i \sigma_i^2 - \lambda_{MV} \frac{1}{2} p_u^2 \sum q_j \sigma_{uj} + \lambda_{MV} \frac{1}{2} p_u \mu_i + \lambda_{PT} \frac{1}{2} p_u \pi_{ui} V_{ui} - \lambda_{L2}\ q_i$$