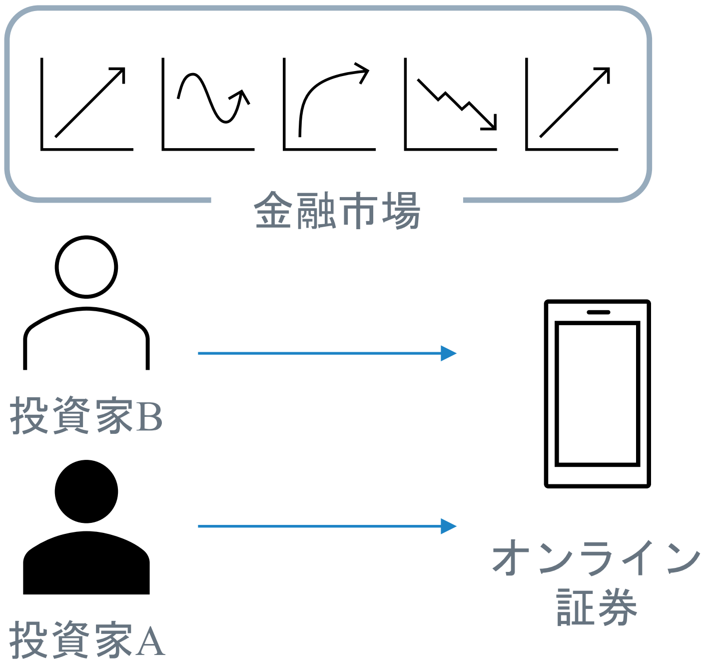
GPIF Finance Awards for Students

情報技術を活用し個人の選好を考慮した投資支援

高柳剛弘　工学系研究科　和泉研究室　博士後期課程

研究背景：貯蓄から投資へ



- 個人投資家の金融市場への参加が容易に
  - さまざまなオンライン証券の台頭
  - 手数料の低下
- 家計金融資産のリスク性資産への投資は限定的[*1]
  - 国内で投資信託あるいは株式を保有する投資人口は約20%
  - 50%以上の家計の金融資産は現金で保有
- リスク性資産保有への障壁
  - 複雑な金融市場
  - 時間制約
  - リスク定量化の困難さ

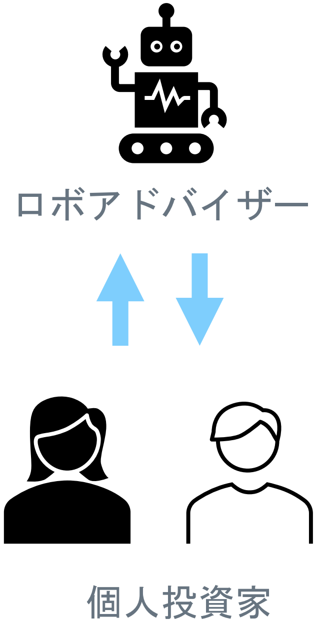
[*1] Flow of funds statistics (https://www.boj.or.jp/en/statistics/sj/index.htm), Bank of Japan

研究背景：推薦システム

- ▷ 情報科学の分野で推薦システムの研究が盛んに
  - ○ **膨大な情報(Information overload)**により必要な情報が埋もれてしまうという問題
    - ■ 欲しい情報が埋もれている
    - ■ 必要な除法を具体化できない
  - ○ **どれに価値があるかを特定**するのを助けるシステム
    - ■ ユーザーの好み，行動，**目的**に合わせて情報やアイテムを提案
- ▷ 多様なドメインで発展
    - ■ E-Commerce，音楽，動画，映画
- ▷ 情報学の一大分野
  - ○ ユーザー行動に関して新たな知見
  - ○ 予測モデルの発達

Ex) Amazon
特定の顧客の嗜好・目的に合わせて商品を提示

モチベーション：推薦システムの知見を活用した金融市場における投資家支援システムの構築



- ▷ 金融市場においても，**膨大な量の情報が継続的に生成され**（information overload ），それを適切に理解・解釈し，投資判断を行うことは**高度な能力が必要**
- ▷ 一般の個人投資家にとっては依然として**株式等のリスク性資産保有への参入障壁が非常に高い**という課題

=>金融分野においても**推薦システムの知見を活用**することで，**個人投資家への支援システム**が構築できるのではないか？

研究の方針

- 推薦システム研究の推移
  1. <u>行動予測精度（accuracy）向上</u>
     - ユーザーの行動予測のためにさまざまな手法が提案
     - ユーザーの選好に対する知見が多く蓄積
     - Filter bubbleや公平性などの問題が出現
  2. <u>Beyond accuracyの推薦システム構築の研究</u>
     - 予測精度（Accuracy）にのみならず，多様性，セレンディピティ，公平性などを考慮した推薦システムの構築
     - 各ドメインにおける重要指標の提案
       - 動画推薦におけるセレンディピティ
       - ニュース推薦における新鮮さ

- 本研究の方針
  1. <u>行動予測精度(Accuracy)の向上</u>
     - 投資家行動の予測手法の提案
     - 投資家行動に対する知見を蓄積
     - 予測精度を追求することから出現する問題・課題の整理
  2. <u>金融分野に適したBeyond Accuracyの推薦システムの構築</u>
     - 金融分野におけるBeyond Accuracyの指標を確立
       - 投資家のリスク許容度とリスクリターンとの合致，ポートフォリオのProfitability，投資目的
     - Beyond Accuracyの推薦システムの提案

研究の方針

現在の研究成果

- 推薦システム研究の推移
  1. **行動予測精度（accuracy）向上**
     - ユーザーの行動予測のためにさまざまな手法が提案
     - ユーザーの選考に対する知見が多く蓄積
     - Filter bubbleや公平性などの問題が出現
  2. **Beyond accuracyの推薦システム構築の研究**
     - 予測精度（Accuracy）にのみならず，多様性，セレンディピティ，公平性などを考慮した推薦システムの構築

- 本研究の方針
  1. 行動予測精度(Accuracy)の向上
     - 投資家行動の予測手法の提案
     - 投資家行動に対する知見を蓄積
     - 予測精度を追求することから出現する問題・課題の整理
  2. 金融分野に適したBeyond Accuracyの推薦分野の構築
     - 金融分野におけるBeyond Accuracyの指標を確立
       - 投資家のリスク許容度とリスクリターンとの合致，ポートフォリオのProfitability，投資目的
     - Beyond Accuracyの推薦システムの提案

投資家行動予測の精度向上

- 過去の取引データが多くある個人投資家に対する行動予測モデルの提案
  - Takehiro Takayanagi, Kiyoshi Izumi, Atsuo Kato, Naoyuki Tsunedomi, Yukina Abe, "Personalized Stock Recommendation with Investors' Attention and Contextual Information," In Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval.

- 過去の取引データがない個人投資家に対する行動予測モデルの提案
  - Takehiro Takayanagi, Kiyoshi Izumi, "Harnessing Behavioral Traits to Enhance Financial Stock Recommender Systems: Tackling the User Cold Start Problem," 2023 IEEE International Conference on Big Data (Big Data'23).

Personalized Stock Recommendation with Investors’ Attention and Contextual Information

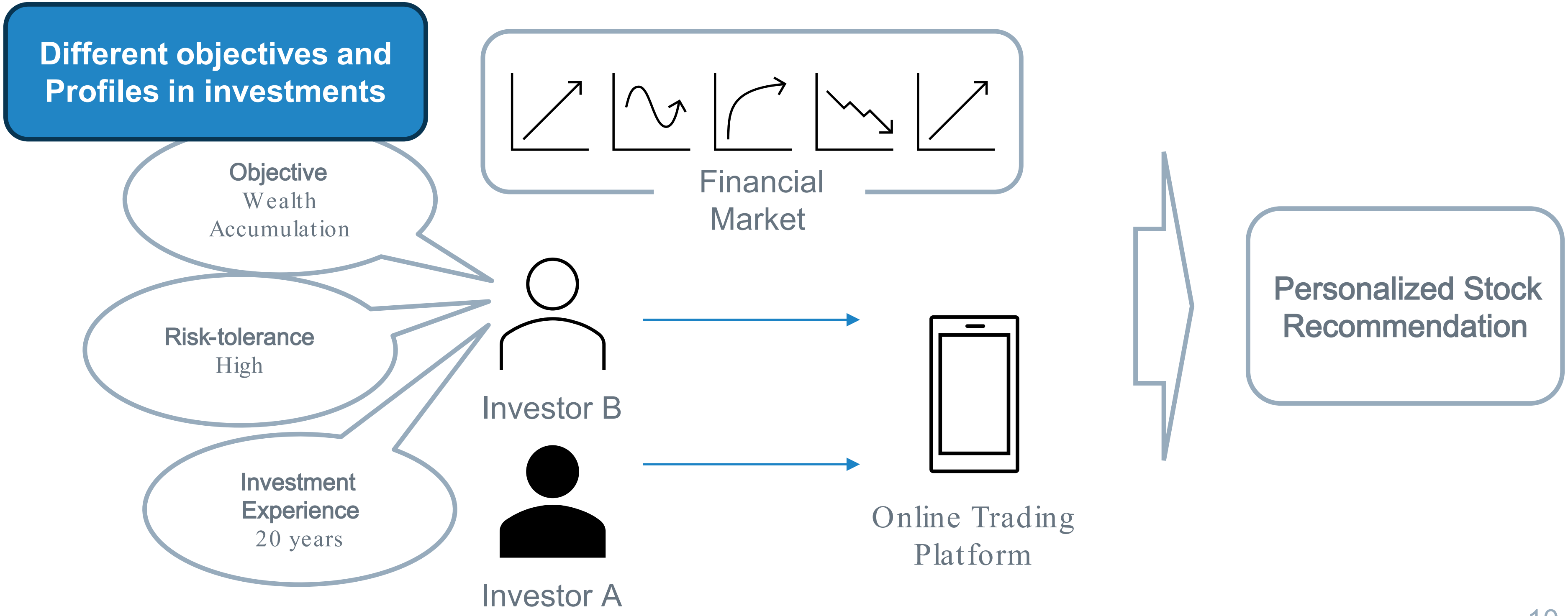
The International ACM SIGIR conference, Taipei Taiwan 2023

研究概要

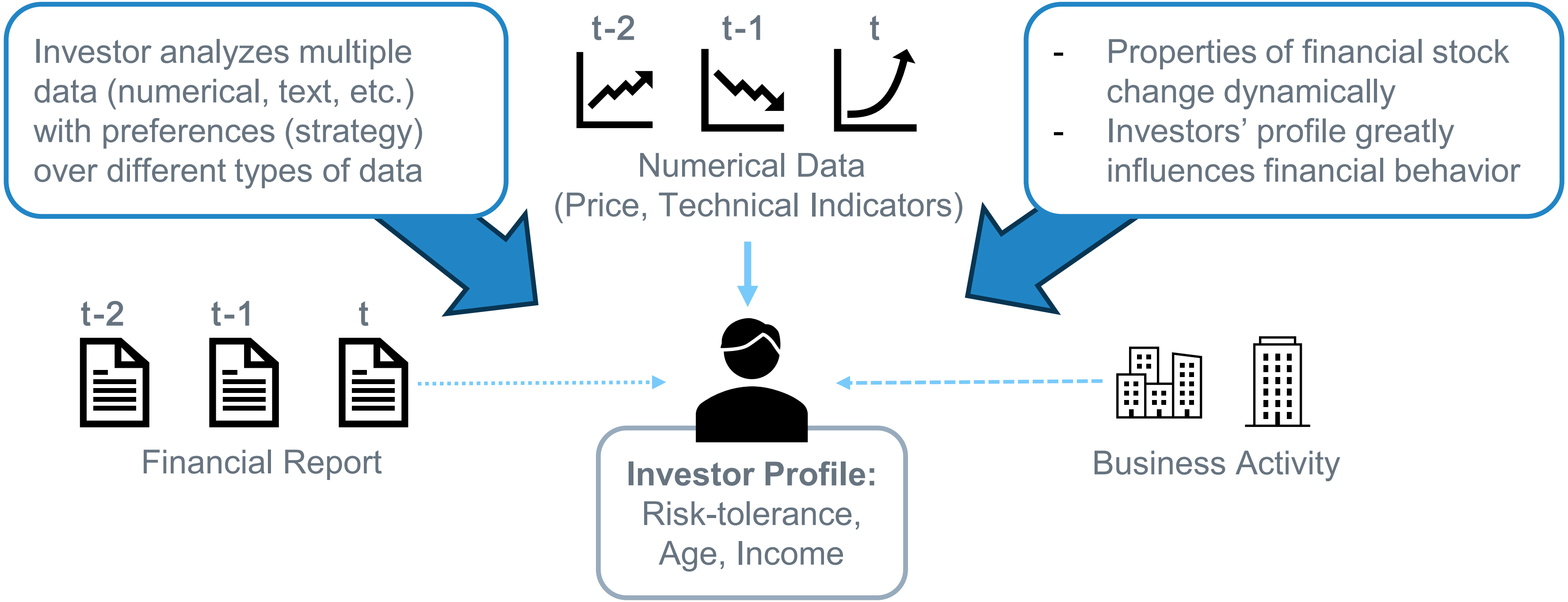
- ▷ 過去の取引データが多くある個人投資家に対する行動予測モデルの提案
- ▷ 個人投資家は意思決定の際に数値情報やテキスト情報などさまざまな情報を参照し，投資判断を行う
- ▷ 個人投資家はそれぞれの情報に対して各自の重みをもつ
  - 株価チャートの動きを重視する投資家やニュース情報を重視する投資家
- ▷ 個人投資家の情報選択の重みを考慮した協調フィルタリングに対する投資家行動予測モデルを提案
- ▷ 個人投資家の取引データを用いた比較実験でベースラインを超える予測精度を達成し，情報選択の重みを考慮する重要性を確認

本手法は新規参入など取引データのない個人投資家には適応できない（コールドスタート問題）＝＞取引データのない投資家に対する行動予測モデルの必要性

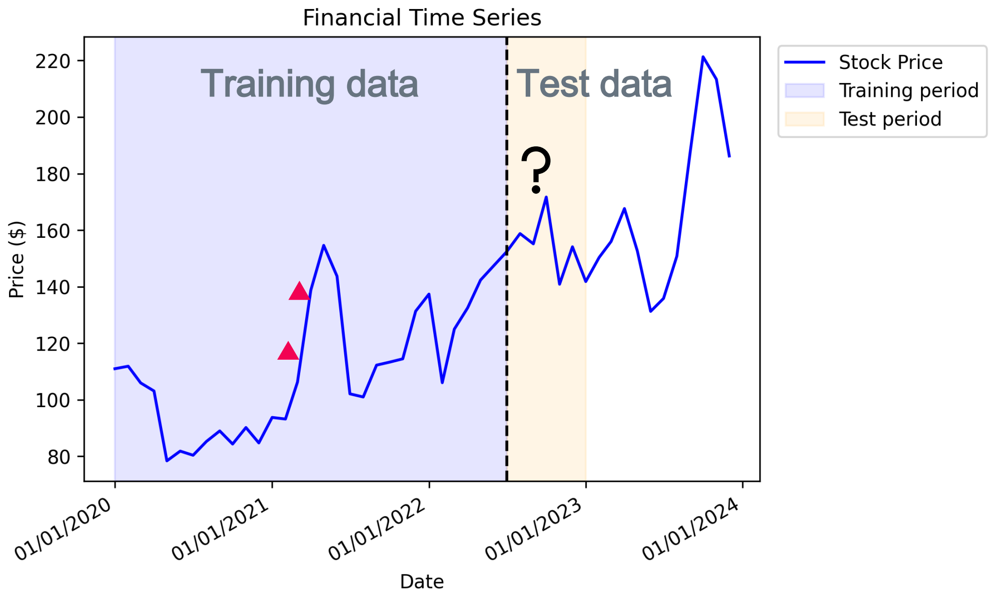
Personalized Stock Recommendation



Motivations



Financial Recommendation (FinRec)



- Overview of FinRecprocess
  - Financial asset types (individual stock, mutual fund, bonds,..)
  - Data split: training / test
    - Training: from 1st Jan 2018 to 30th June 2022 30th June
    - Test: from 1st Jul 2022 to 31st Dec 2023
  - Train model and evaluate
- Methods
  - Collaborative filtering approach [McCreadie et al., 2022; Takayanagi et al.,2023]

Dataset: Transaction Data

	Interaction#	Unique Assets#	Investor#
Original	518,847	3.065	967
Processed	91,317	1,627	956

Dataset Statistics

stocks per investor

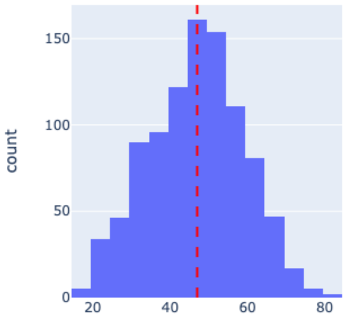
#transactions per investor

- Transaction data from 969 active investors from our platform
  - Active investors: Those who conducted more than 50 transactions within a year
  - Sampling period: Jul 2020 to Sep 2022
- Data processing
  - Focus on transactions involving individual stocks
  - Included only “buy” transactions

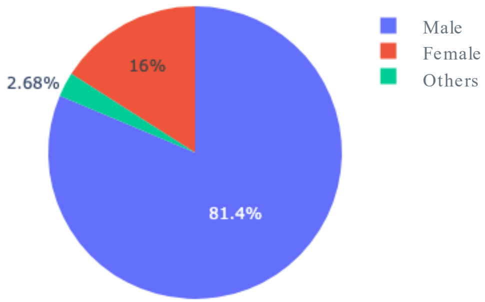
Dataset: Context Data

- Stock Data
  - Technical Indicators
    - Bollinger band, MACD, RSI, etc.
  - Fundamental Factors
    - Earning per share, book value per share, etc.
  - Business Activities
    - Stock embedding (SETN)[Takayanagi et al., 2022]
    - Graph Convolutional Network
    - RoBERTa

- Investor Profile
  - Age
  - Annual income
  - Investment experience
  - Risk tolerance



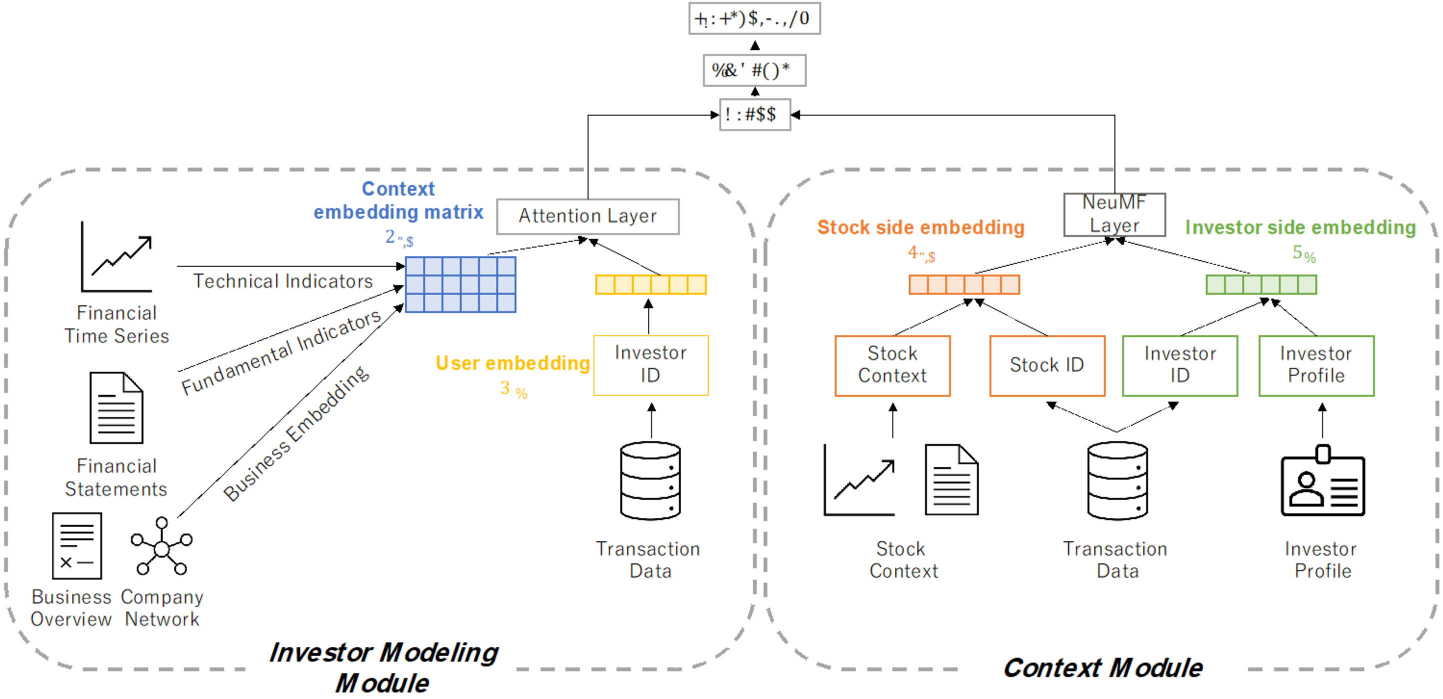
Age Distribution



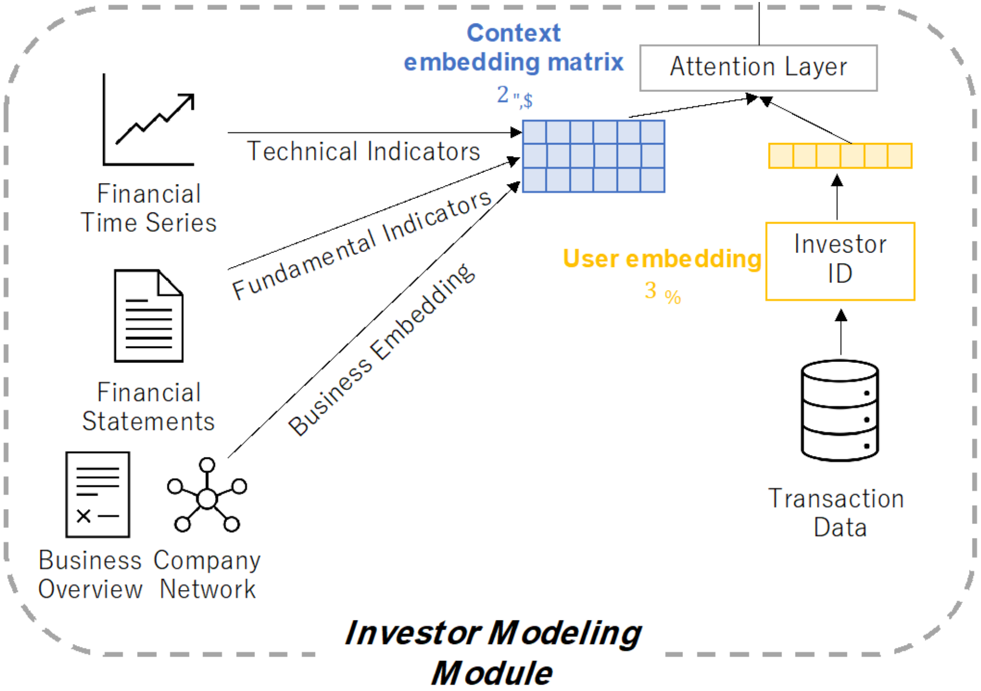
Gender Proportion

PSRIC model

Personalized Stock Recommendations with Investors' Attention and Contextual Information



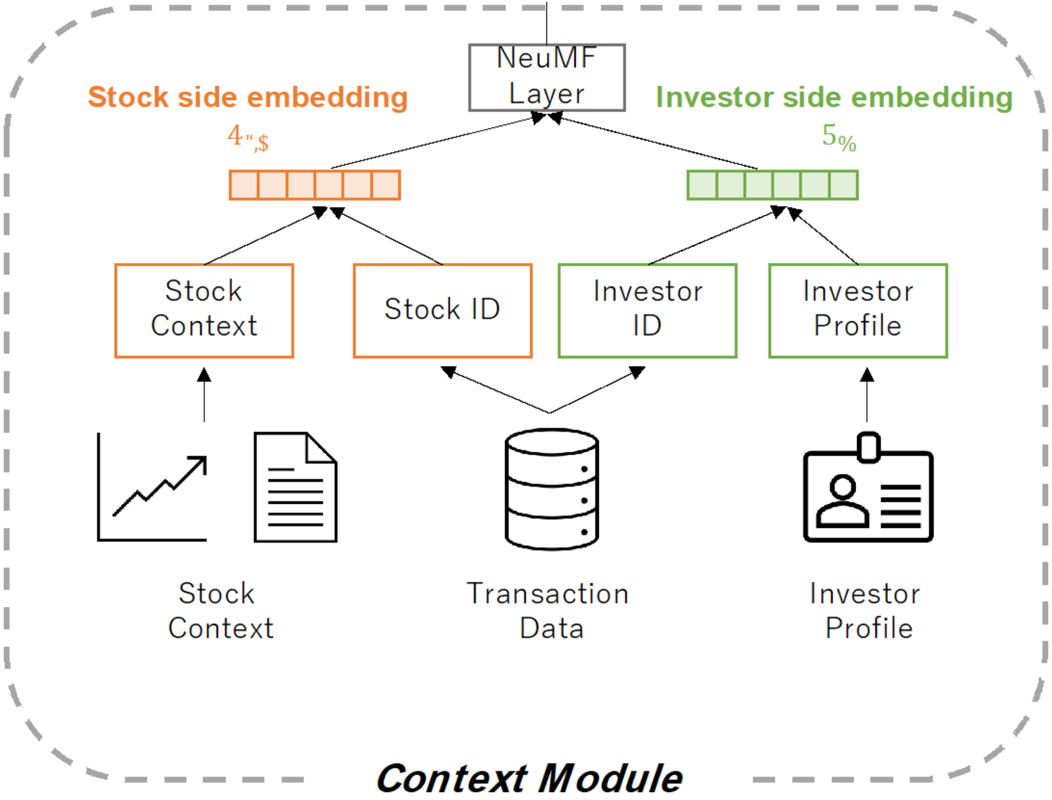
PSRIC model: Investor Modeling Module



Investor Modeling Module

- Captures the financial decision-making process of investors
- Utilizes a self-attention mechanism to model attention towards various stock contexts

PSRIC model: Context Module



Context Module

- Captures stock dynamics and investors' profile
- Incorporates side information through stock and investor side embedding

Experimental Settings

- Evaluation approaches
  - Leave-one-out approach
- Evaluation metrics
  - HR@K
  - nDCG@K

- Baselines
  - Pop Model
  - BPR [Rendle et al,. 2009]
  - itemKNN [Aiolli 2013]
  - NeuMF [He et al,. 2017]
  - NGCF [Wang et al,. 2019]
  - LightGCN [He et al,. 2020]
  - MultiVAE [Liang et al,. 2018]
  - RecVAE [Shenbin et al,. 2020]

Result & Discussion

	H@5	H@10	H@50	N@5	N@10	N@50
Pop	0.1726	0.2696	0.8089	0.0949	0.1258	0.2360
BPR	0.3821	0.5362	0.8906	0.2702	0.3199	0.3991
itemKNN	0.1726	0.2696	0.8089	0.0949	0.1258	0.2360
NeuMF	0.4222	0.5855	0.9168	0.2951	0.3476	0.4224
MultiVAE	0.4191	0.5686	0.9045	0.2924	0.3401	0.4149
NGCF	0.4160	0.5824	0.9153	0.2887	0.3428	0.4172
LightGCN	0.4191	0.5794	**0.9214**	0.2934	0.3458	0.4236
RecVAE	0.4083	0.5639	0.8937	0.2834	0.3338	0.4073
PSRIC	**0.4807**	**0.6441**	0.9106	**0.3772**	**0.4292**	**0.4874**

Performance comparison

Model	H@5	H@10	N@5	N@10
PSRIC	0.4807	0.6441	0.3772	0.4292
- Investor Modeling	-2%	-9%	-7%	-10%
- Context	-2%	-7%	-4%	-7%

Ablation Study

- Performance observations
  - PSRIC outperforms all baseline models across most evaluation metrics
  - Especially, PSRIC performs better in shorter recommendation lists
- Ablation studies
  - Both modules contribute to the overall performance, with investor modeling module being most effective
  - Eliminating each module leads to a more decrease in nDCG performance

Conclusion and Future Work

- Conclusion
  - We proposed a novel stock recommendation model
  - We validated the effectiveness of the model through experiments
- Future Work
  - We aim to proceed with application development and commence operation as a service
    - Need for online experiments to confirm the effectiveness
    - Investigate a model that aligns investors' interests with portfolio profitability

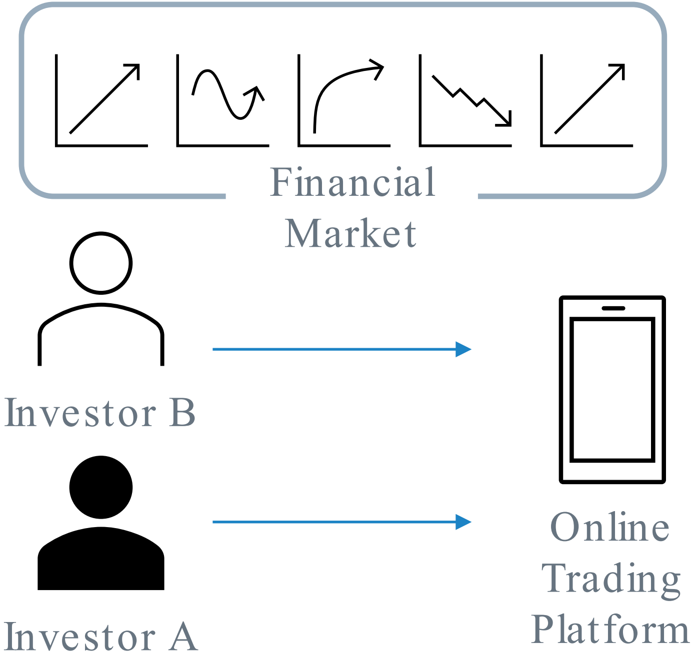
Harnessing Behavioral Traits to Enhance Financial Stock Recommender Systems: Tackling the User Cold Start Problem

IEEE International Conference of Big Data, Sorrento 2023

研究概要

- 過去の取引データがない個人投資家に対する行動予測モデルの提案
  - 行動予測モデルは協調フィルタリングを用いることが多い
  - 協調フィルタリングはUserとItemの取引データから学習するので、取引データのない新規ユーザーに対して精度が低下するコールドスタート問題が知られている
- 金融分野の推薦システムではコールドスタート問題は特に大きな課題
  - 投資経験の少ないnovice investorsが対象に
- 他ドメインの推薦ではユーザー行動と相関するユーザー情報（年齢，ジェンダーなどのデモグラや性格などの心理特性）を用いて取引データ不足を補う
- 金融分野では、行動ファイナンスにより投資家特性と投資家行動の関係性を分析した豊富な研究がある
- 行動ファイナンスの知見を用いることで、取引データの少ない個人投資家に対しても高い精度の投資家行動予測を達成

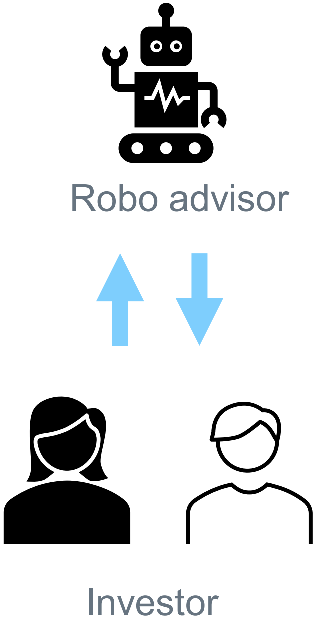
Challenges to Investment



- Ease of investment
  - Growing number of online trading platform
  - Lower commission fee
- Limited public participation in investments
  - Despite ease of access, a small percentage of the financial assets is invested
  - In Japan, over 50% of household financial assets are held in cash [*1]
- Barriers to investing in financial assets
  - The complexity of financial market
  - Time constraints
  - The difficulty in quantifying risks

[*1] Flow of funds statistics (https://www.boj.or.jp/en/statistics/sj/index.htm), Bank of Japan

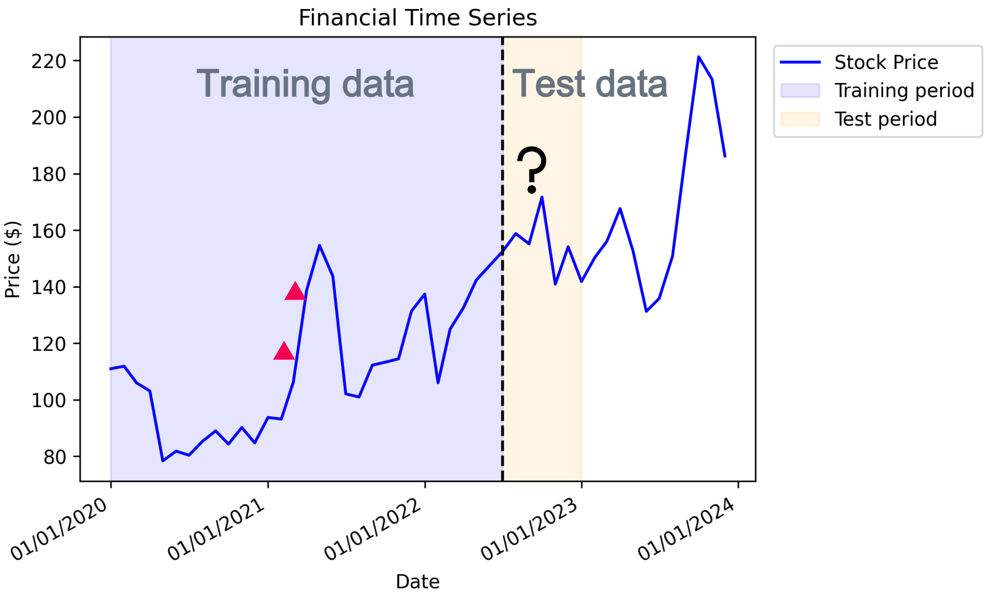
Financial recommendation (FinRec)



- It is difficult for an average individual to become a savvy investor on their own due to the challenges
- They need a **financial advisor** to analyze their position and recommend assets to invest in personalized to them
  - Manage investment risk to the customer by identifying profitable assets that meet their risk profile
- But consulting expert financial advisors can be costly

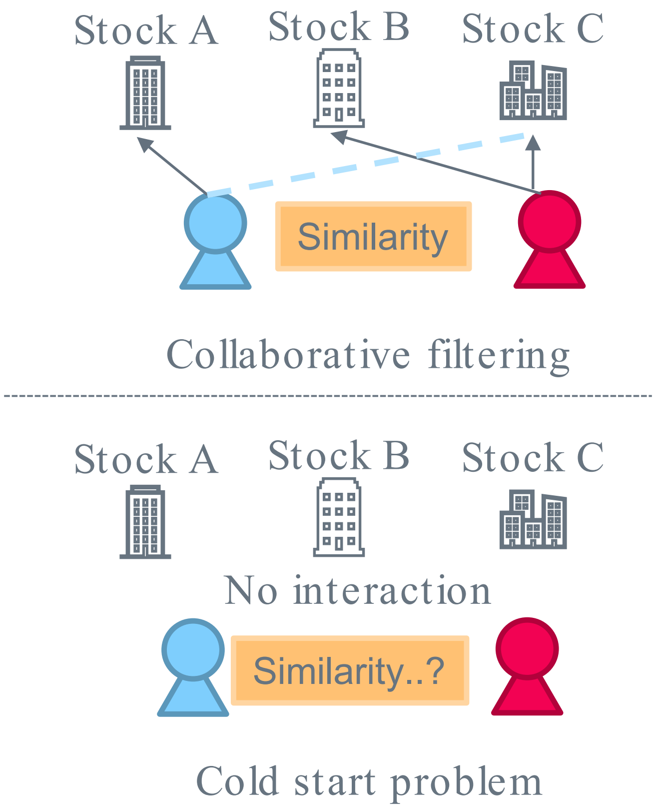
=> **Financial recommender system (FinRec)** can fill this gap

Financial Recommendation (FinRec)



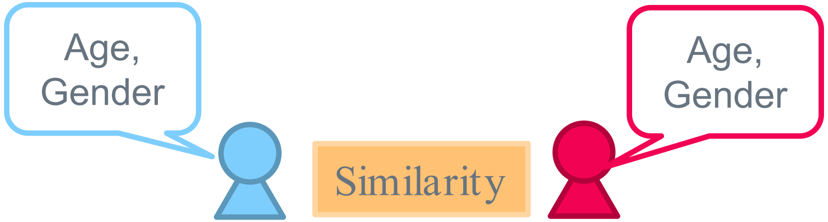
- Overview of FinRecprocess
  - Financial asset types (individual stock, mutual fund, bonds,..)
  - Data split: training / test
    - Training: from 1st Jan 2018 to 30th June 2022 30th June
    - Test: from 1st Jul 2022 to 31st Dec 2023
  - Train model and evaluate
- Methods
  - Collaborative filtering approach [McCreadie et al., 2022; Takayanagi et al.,2023]

The challenge of collaborative filtering: Cold Start



- Cold start problem
  - Recommendation performance deteriorates for new users with scarce item interactions [Zhang2019; Gope 2017; Briand 2021]
- FinRec also suffers from user cold start problem as it often employs collaborative filtering approach
- The user cold start problem is especially severe in FinRec
  - Novice investors typically lack experience and sound investment strategy

Tackling Cold Start Problem: General Domain



General domain

- Existing research to tackle the cold start problem employs meta-data, often times demographic data [Yanxiang 2013; Lika 2014]
- Ongoing efforts to find user traits that correlate with user preferences
- Personality in music recommendation [Ferwerda et al., 2017]
  => Personality-aware recommendations [Dhelim 2020,2022; Lex 2022]

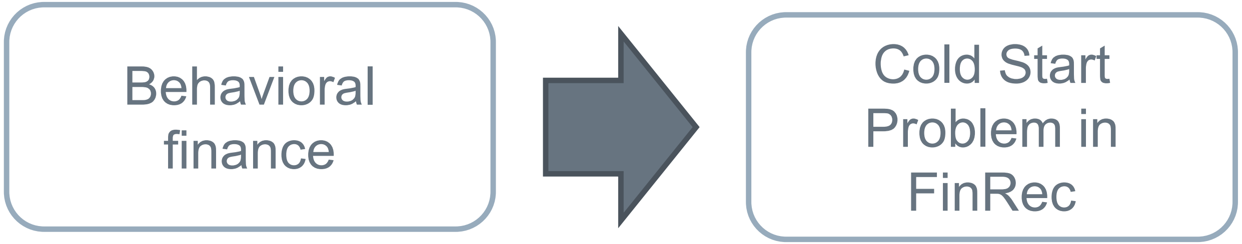
Tackling Cold Start Problem: FinRec

- Few research tackles the cold start problem in FinRec, while many research points out the significance of the problem [McCreadie et al., 2022; Takayanagiet al.,2023]
- In financial scenarios, **behavioral finance** studies the relationship of investors' traits and investment behavior [Barber & Odean 2013]
- Key factors include
  - Gender [Barber & Odean 2001]
  - Personality [Jiang et al., 2023; Tauni et al., 2015]
  - Cognitive ability [Grinbalt et al., 2011]
  - Investment objectives [Shefrin & Statman 2011]

→ The wealth of knowledge in behavioral finance provides a unique opportunity to leverage behavioral traits for the cold start problem in FinRec !

Key ideas

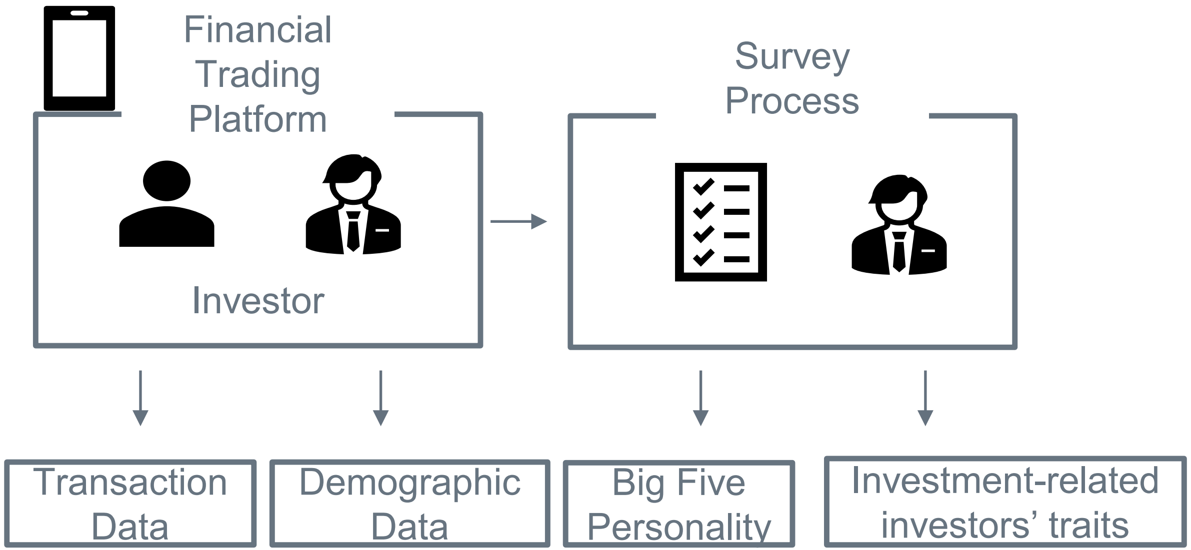
- Can we utilize the insights from behavioral finance to tackle the cold start problems in FinRec?



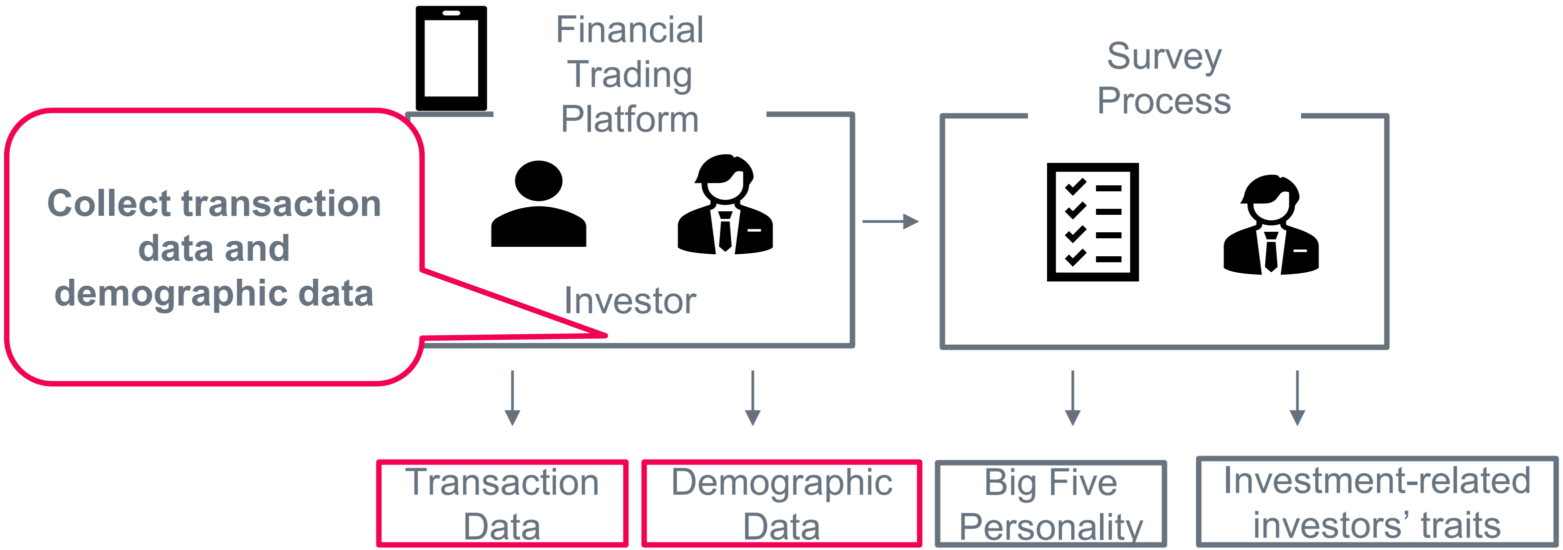
Our approach

- To tackle the research questions
  - Collection of a **comprehensive dataset** encompassing investors' actual transaction and investors' traits
  - Introduction of a **model to tackle the cold start problem** in FinRec

Data Collection Process

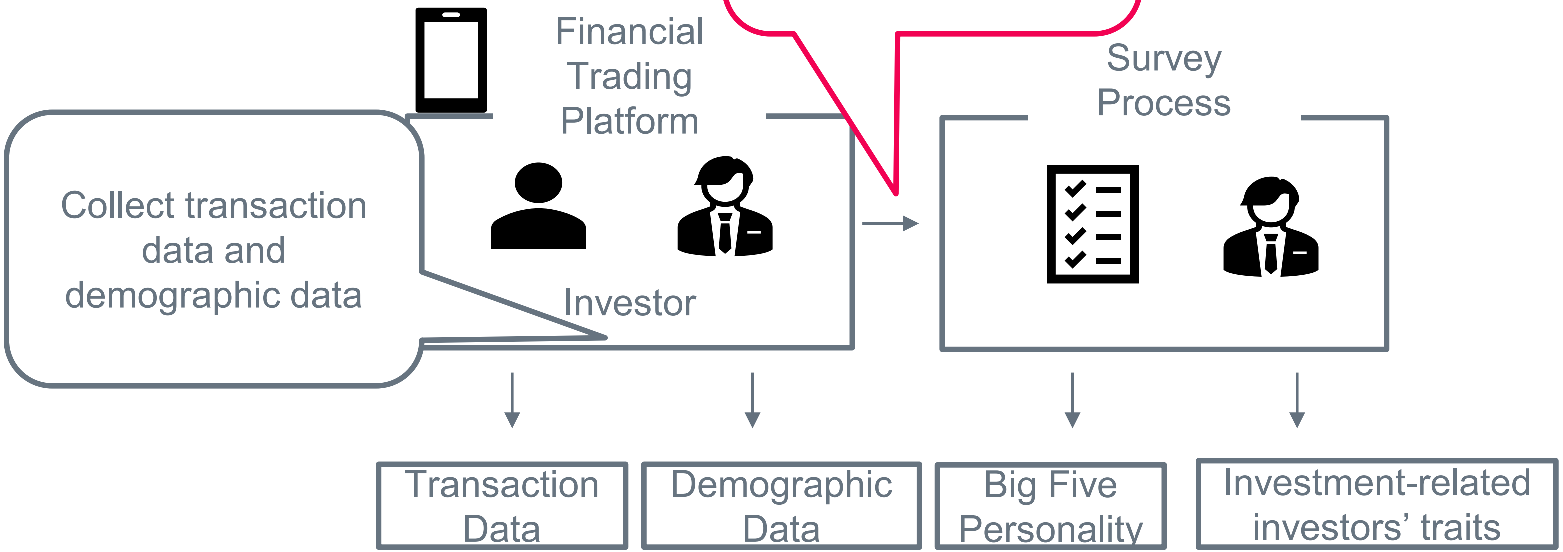


Data Collection Process

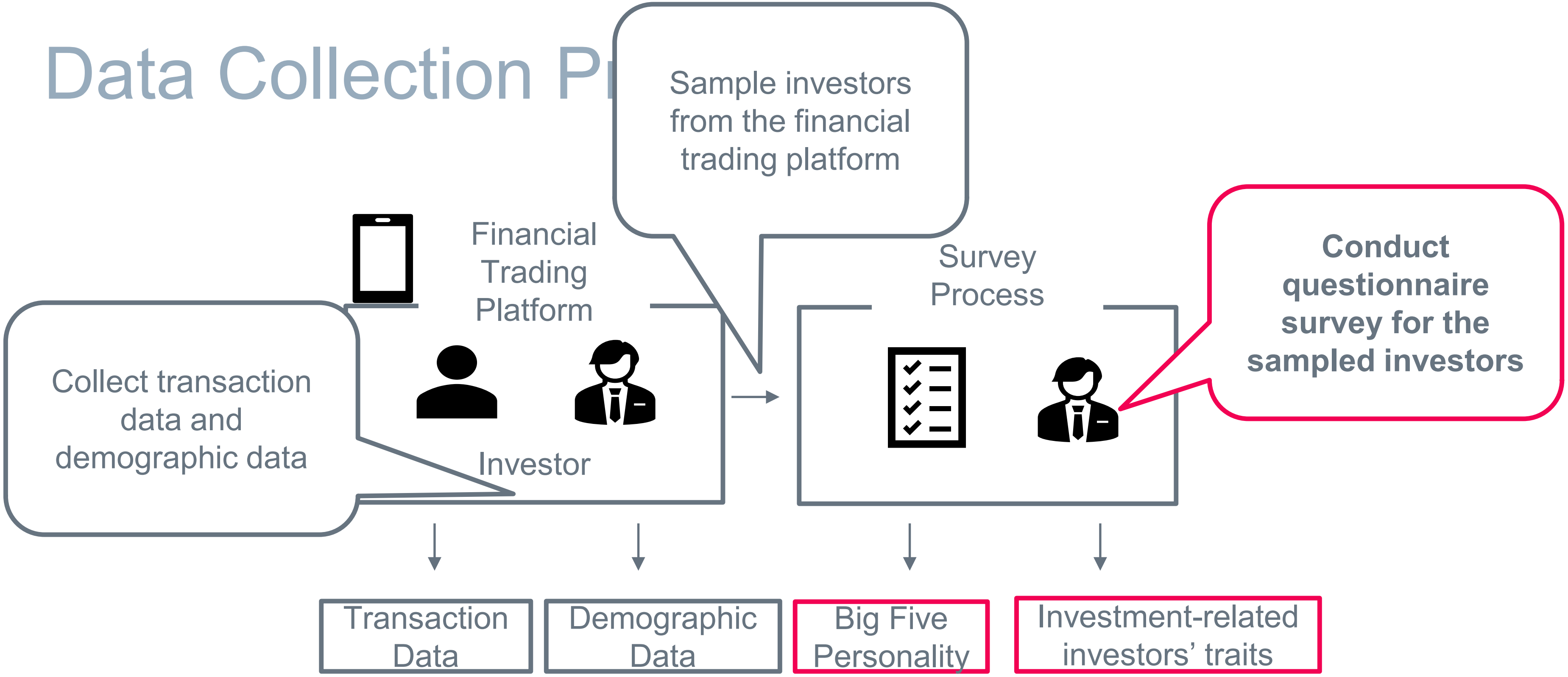


Data Collection Pr

Sample investors from the financial trading platform



Data Collection Pr



Data Overview

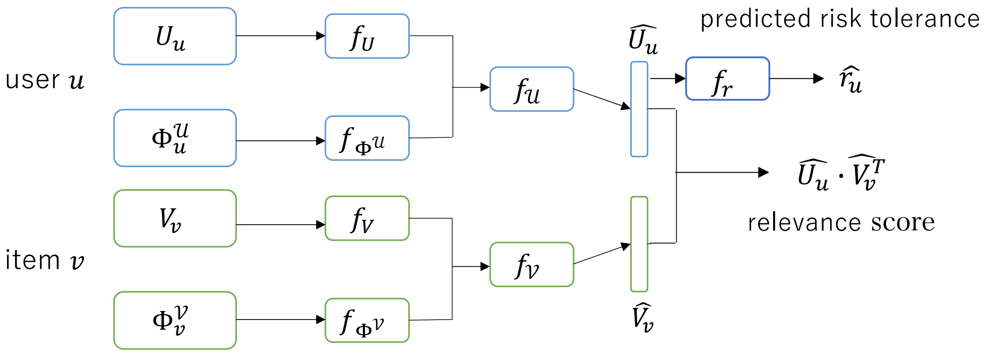
TABLE I
THE STATISTICS OF OUR TRANSACTION DATASET.

Interaction#	Unique Assets#	Investor#	Average Age	Sampling Period
32,678	2,962	964	47	Jul 2020 to Jul 2023

Financial trading platform	Questionnaire survey
Transaction history	**Big five personality traits**[*1] [Gosling et al., 2003, Oshio et al., 2013]
Demographic (age, gender, income financial asset amount, investment experience)	**Investment-related Behavioral Traits** [*2] Cognitive ability [Frederick, 2005] Subjective financial literacy [Collins 2012] Objective financial literacy [Lusardi & Mitchell 2008] Investment purpose Risk preference Time discounting

*1 Ten item personality inventory (TIPI)
*2 Japan Household Panel Survey (JHPS)

IRAD: Investor Risk-tolerance Aware DropoutNet



Algorithm 1 Learning Algorithm

Input: R, U, V, Φ^U, Φ^V

Initialize: user model f_U, item model f_V, risk tolerance model f_r

repeat ▷ DNN optimization

  Sample mini-batch $B = \{(u_1, v_1), \ldots, (u_k, v_k)\}$

  for each $(u, v) \in B$ **do**

    if dropout is applied **then**

      $[U_u, \Phi_u^U] \rightarrow [0, \Phi_u^U]$ ▷ User dropout

    else

      $[U_u, \Phi_u^U] \rightarrow [U_u, \Phi_u^U]$ ▷ Leave as is

    end if

  end for

  Update f_V, f_U, f_r using B based on combined loss $\mathcal{L}$ from Equation 11

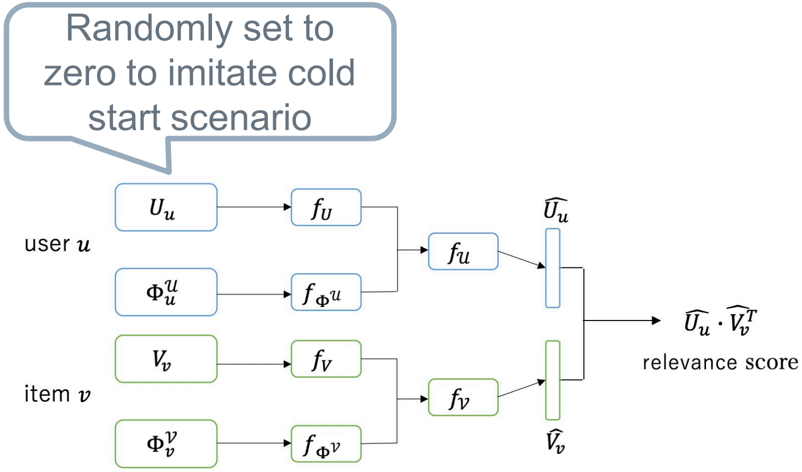
until convergence

Output: f_V, f_U, f_r

Model component

- Base component: Utilizing DropoutNet (Volkovs et al., 2017)
- Multitask learning: Incorporates a loss function for risk tolerance

Base component (Volkovs et al., 2017)



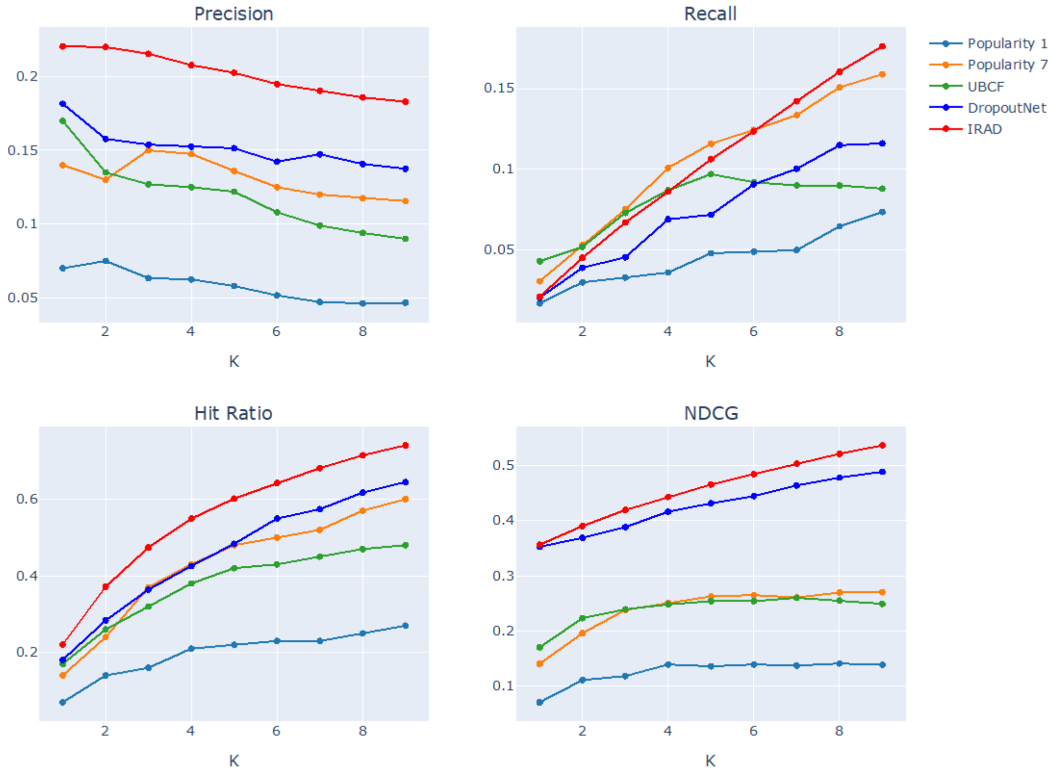
- Input: latent representations U_u (user) and V_v (item)[*1], and meta data Φ_u^U (user) and Φ_v^V(item)
- Model: deep neural networks f_{Φ^U}, f_{Φ^U}, f_U, and f_V for processing user and item inputs
- Training: the model is trained to generate reconstructions. *want to minimize* $U_u \cdot V_v^T - \widehat{U_u} \cdot \widehat{V_v^T}$
- The input are randomly set to zero, using Dropout [Srivastava et al., 2014]
  => enables the model to learn from content-only scenarios and both content and preference scenarios.

[*1] In our experiment, we utilize weighted matrix factorization

Evaluation method

- Evaluation metrics
  - Precision@K
  - Recall@K
  - Hit Ratio
  - NDCG (Normalized discounted cumulative gain)
- Cold start settings
  - Evaluation of recommendation relevance against actual stock preferences of new users in their first 30 days
- Baseline models
  - Popularity 1
  - Popularity 7
  - UBCF (User based collaborative filtering) [Dhelim 2020, 2022]
  - DropoutNet [Volkovs et al., 2017]

Result



- IRAD consistently outperforms the baseline models for the majority of K values across multiple evaluation metrics
- Particularly, displaying higher precision

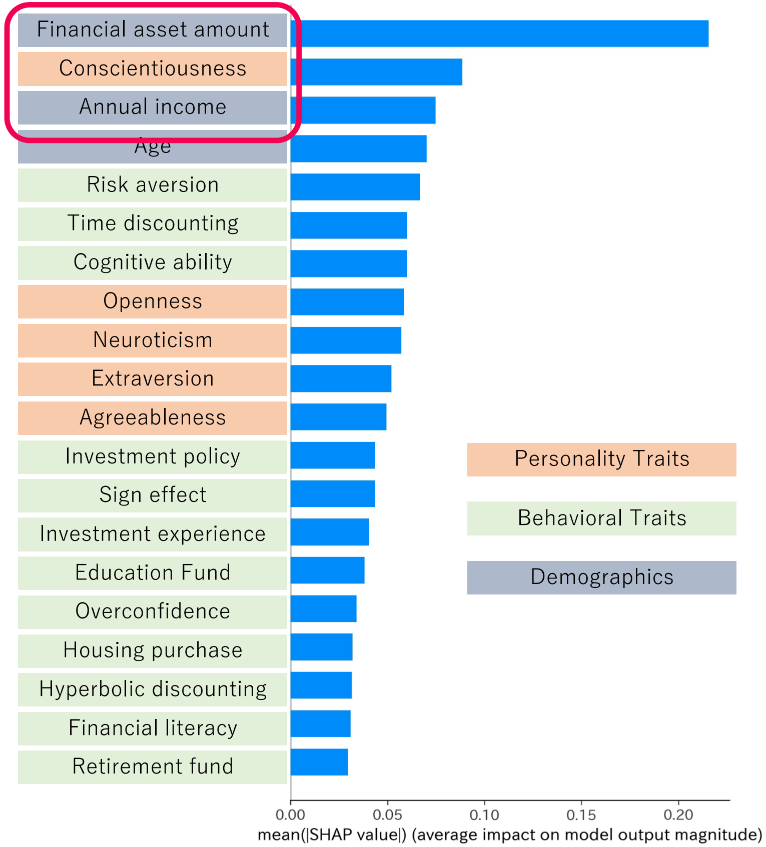
Result: Ablation studies

TABLE II
EXPERIMENTAL RESULTS AND ABLATION STUDIES.

	P@5	R@5	HR@5	NDCG@5
Pupularity 1	0.058	0.048	0.220	0.236
Popularity 7	0.136	**0.116**	0.480	0.263
UBCF [32], [33]	0.122	0.097	0.420	0.254
DropoutNet [18]	0.151	0.072	0.484	0.432
IRAD	**0.203**	0.106	**0.601**	**0.466**
-RTL	-0.051	-0.034	-0.118	-0.034
-RTL-IBT	-0.053	-0.031	-0.101	-0.036

- Impact of risk tolerance loss (RTL)
  - RTL significantly contributes to the IRAD's performance
- Effect of removing investment-related behavioral traits (IBT)
  - Negligible performance loss observed upon removal of IBT

Interpretations with investors' traits



- We can also utilize the obtained investors' traits for interpretating the model output
- The fig presents a SHAP [Lundberg and Lee 2017] score derived from 1000 samples
- The result indicates the financial asset amount, conscientiousness, and annual income contribute the most to the results.

Conclusions & Future work

Conclusion

- ▷ This study presents a novel framework that addresses the user cold start problem in financial recommendations by incorporating insights from behavioral finance
- ▷ We conduct survey study from investors in online financial trading platform to build dataset
- ▷ We propose IRAD, a model for tackling cold start problem in FinRec, outperforming baseline models
- ▷ We observe that the obtained investment-related traits influence the recommendation performance marginally

Future work

- ▷ We plan to extend our research by exploring more effective methods for utilizing investor behavioral traits
- ▷ We plant to enhance recommendation interpretability by utilizing investor traits

研究の方針

今後の方針

- 推薦システム研究の推移
  1. 行動予測精度（accuracy）向上
     - ユーザーの行動予測のためにさまざまな手法が提案
     - ユーザーの選考に対する知見が多く蓄積
     - Filter bubbleや公平性などの問題が出現
  2. Beyond accuracyの推薦システム構築の研究
     - 予測精度（Accuracy）にのみならず，多様性，セレンディピティ，公平性などを考慮した推薦システムの構築

- 本研究の方針
  1. 行動予測精度(Accuracy)の向上
     - 投資家行動の予測手法の提案
     - 投資家行動に対する知見を蓄積
     - 予測精度を追求することから出現する問題・課題の整理
  2. 金融分野に適したBeyond Accuracyの推薦分野の構築
     - 金融分野におけるBeyond Accuracyの指標を確立
       - 投資家のリスク許容度とリスクリターンとの合致，ポートフォリオのProfitability，投資目的
     - Beyond Accuracyの推薦システムの提案

今後の研究方針

- 行動予測精度を追求することから出現する問題・課題の整理
  - 行動予測精度とProfitabilityの相関など
- 金融分野のBeyond Accuracyの指標を確立
  - 投資家のリスク許容度とポートフォリオのリスクリターン
  - ポートフォリオのProfitability
- Beyond Accuracyを考慮した推薦システムの構築
- Public Dataを作成・公開し，研究コミュニティーに寄与
  - Twitterを用いた個人投資家の興味推移を表すデータセットの公開
  - SECのForm13Fから作成した機関投資家の行動データセットの公開

年金運用研究へのつながり

- 本研究で作成された投資家の行動モデルを活用した投資家エージェントによる人工市場の作成
- 人工市場シミュレーションにより強化学習を用いてマーケットインパクトを考慮した最適執行戦略の研究

研究成果一覧

- 国際会議4件
  - Takehiro Takayanagi, Chung-Chi Chen, Kiyoshi Izumi, “Personalized Dynamic Recommender System for Investors,” In Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR‘23). (Acceptance Rate: 25.12%, 154/613)<-情報系のトップカンファレンス
  - Takehiro Takayanagi, Kiyoshi Izumi, Atsuo Kato, Naoyuki Tsunedomi, Yukina Abe, "Personalized Stock Recommendation with Investors’ Attention and Contextual Information," In Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'23) <-情報系のトップカンファレンス.
  - Takehiro Takayanagi, Kiyoshi Izumi, "Harnessing Behavioral Traits to Enhance Financial Stock Recommender Systems: Tackling the User Cold Start Problem," 2023 IEEE International Conference on Big Data (Big Data'23).
  - Takayanagi Takayanagi, Hiroki Sakaji and Kiyoshi Izumi, "SETN: Stock Embedding Enhanced with Textual and Network Information," 2022 IEEE International Conference on Big Data (Big Data'22).
- ジャーナル2件
  - Takehiro Takayanagi, and Kiyoshi Izumi, ”Incorporating Domain-Specific Traits into Personality-Aware Recommendations for Financial Applications, ” New Gener. Comput.
  - Takehiro Takayanagi, and Kiyoshi Izumi, “Context-Aware Stock Recommendations with Stock’s Characteristics and Investors,’ Traits, ” 2023 IEICE TRANSACTIONS on Information and Systems.
- 国内会議5件
- 受賞2件
  - 人工知能学会全国大会優秀賞（ https://www.t.u-tokyo.ac.jp/topics/tp2023-11-28-001 ）
  - Amazon Science Fellow
- プレスリリース2件
  - https://www.t.u-tokyo.ac.jp/press/pr2023-07-20-001
  - https://www.t.u-tokyo.ac.jp/press/pr2022-07-28-001