

銘柄特徴と投資家特性を考慮した株式銘柄推薦の個別化

高柳剛弘¹ 坂地泰紀¹ 和泉潔¹

¹ 東京大学 大学院工学系研究科

m2021ttakayanagi@socsim.org

概要

本研究では銘柄特徴と投資家特性を考慮した銘柄推薦モデルを提案する。提案手法では銘柄のテクニカル指標、ファンダメンタル指標、事業活動の情報を抽出し銘柄ベクトルを作成し、投資家の心理特性、行動特性、属性情報を活用して投資家ベクトルを作成する。最後に協調フィルタリングにより銘柄の推薦を行う。個人投資家の取引データを用いて行った実験により、提案手法の精度が既存手法の精度を上回ることを確認し、銘柄推薦に対して銘柄特徴と投資家特性を考慮することの有用性を示した。

1 はじめに

資産運用を行う個人投資家の増加に伴い銘柄推薦の必要性が増加している。個人投資家に銘柄を推薦する銘柄推薦において、個人投資家がポートフォリオを構築する目的は個人投資家ごとに異なるため、銘柄推薦をパーソナライズすることが重要である。例えば、投資家の中には高いリターンのみを求める投資家に加えて、より慣れ親しんだ (familiar) ポートフォリオを求める投資家も存在する。

Personalized された銘柄推薦の研究では個人投資家のリスク許容度と銘柄のリスクリターンに基づいて銘柄推薦をパーソナライズする手法がよく用いられる [1, 2, 3]。例えば、Robin はユーザーのポートフォリオから投資家のリスク許容度を推定し、株式のリスクリターンと投資家のリスク許容度との関連性に基づいて、協調フィルタリングを用いて銘柄推薦のパーソナライズを行っている [1]。これらを背景とし、本研究では personalized された銘柄推薦に取り組む。

Personalized された銘柄推薦には二つの課題が存在する。一つ目の課題は銘柄の特性の考慮が十分でないという点である。先行研究では銘柄の特徴の中でもリスクリターンのみ焦点を当てることが多い。しかしながら、個人投資家は銘柄を選択する際

に銘柄のさまざまな側面を考慮して投資の意思決定を行っている。特に、個人投資家は主に3つの要因を分析して意思決定を行うと考えられる。

テクニカル指標

個人投資家は銘柄に対するテクニカル分析を行い過去のトレンド、パターンを分析する。

ファンダメンタル指標

個人投資家は経済指標や財務指標などを分析することで企業の財務状況や経営状況を評価する。

事業活動

個人投資家は決算短信やニュースデータなどのテキストデータから、企業の事業活動を分析する [4, 5, 6]。

したがって、より良いパーソナライズを行うためには銘柄のテクニカル指標、ファンダメンタル指標、事業活動を考慮することが重要である。テクニカル指標やファンダメンタル指標は量的な情報である一方で、事業活動はテキストデータやネットワークデータなどの非構造化データにあらわれる。このことから、銘柄推薦で事業活動を考慮するには非構造化データからの特徴抽出を行う必要がある。

二つ目の課題は個人投資家の特性に対するパーソナライズが不十分な点である。先行研究では個人投資家のリスク許容度に基づきパーソナライズをしているが、投資家モデリングの実証研究において性格特性や行動バイアスなどの心理特性と投資家行動の関連性が示されている [7, 8, 9, 10]。この中で、個人投資家の複雑な性質をリスク許容度という一つの変数により表すことが適切かどうかは明らかでない。したがって、より良いパーソナライズを行うためにはこれらの投資家特性を考慮する必要がある。

本研究では銘柄特徴と投資家特性を考慮した銘柄推薦モデルを提案する。

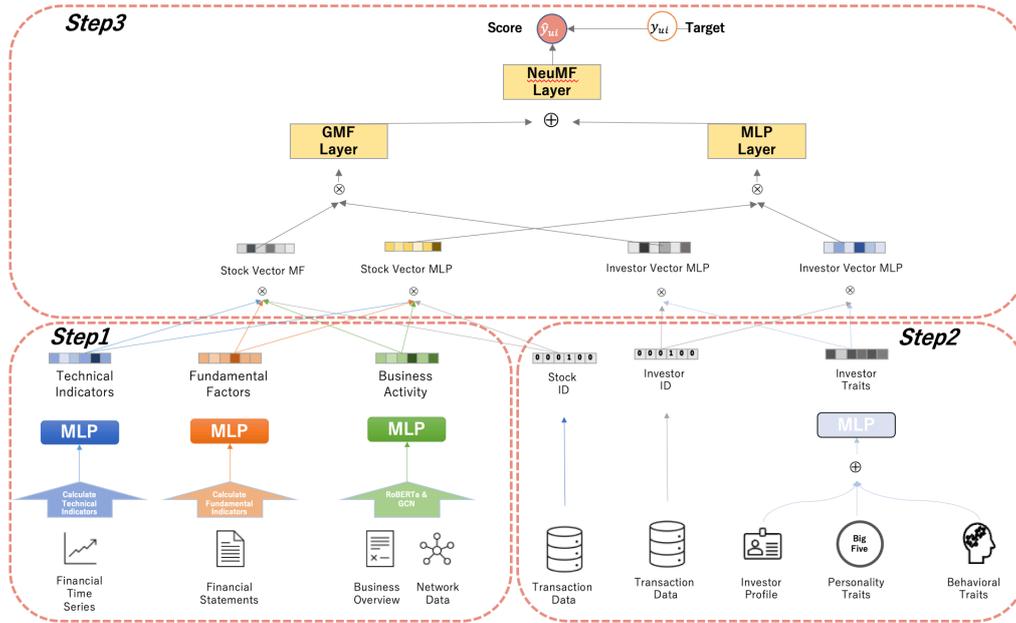


図 1 提案手法の概要

2 提案手法

図 1 に示すよう提案手法は以下の三つのステップで構成される。Step1 では銘柄のテクニカル指標、ファンダメンタル指標、事業活動など銘柄の特徴を抽出し、銘柄のベクトル表現を獲得する。Step2 では投資家の性格特性、行動特性、投資家の属性情報から投資家のベクトル表現を獲得する。Step3 では銘柄と投資家のベクトル表現を利用し Neural Collaborative Filtering Model を構築し、銘柄推薦を行う。

2.1 銘柄特徴の獲得

時刻 t における銘柄 i のテクニカル指標とファンダメンタル指標は次のように計算される。

$$\text{technical}_{t,i} = TC(v_{t,i} \dots v_{t+n,i}) \quad (1)$$

$$\text{fundamental}_{t,i} = FC(w_{t,i}) \quad (2)$$

ここで $v_{t,i}$ は時刻 t における銘柄 i の終値、 n はテクニカル指標の計算に必要な期間、 $w_{t,i}$ は時刻 t における銘柄 i の利益や売上高などの財務情報、 TC はボリンジャーバンドなどのテクニカル指標を株価時系列から計算する関数、 FC は Earnings Price Share (EPS) などのファンダメンタル指標を財務情報から計算する関数、そして $\text{technical}_{t,i} \in \mathbb{R}^{k_t}$ 、と $\text{fundamental}_{t,i} \in \mathbb{R}^{k_f}$ は時刻 t における銘柄 i のテ

クニカル指標、ファンダメンタル指標を表すベクトルである。非構造データから事業活動を抽出する特徴抽出では、Takayanagi *et al.* で提案されたネットワークデータとテキストデータから事業活動情報を抽出する銘柄埋め込み (Stock Embedding) 手法を用いる [11]。Stock embedding は以下のように計算される

$$h_{\text{BERT},i} = \text{BERT}(x_i) \quad (3)$$

$$h_{\text{GCN},i} = \text{ReLU} \left(\sum_{j \in N_i} \hat{D}^{-\frac{1}{2}} \hat{A} \hat{D}^{-\frac{1}{2}} h_j W + b \right) \quad (4)$$

$$h_{\text{BERT},i} = \text{BERT}(x_i)$$

$$h_{\text{GCN},i} = \text{GCN}(h_{\text{BERT},i}) \quad (5)$$

$$\text{emb}_i = h_{\text{BERT},i} + h_{\text{GCN},i}$$

ここで x_i は target stock i のテキストディスクリプションで、 $h_{\text{BERT},i} \in \mathbb{R}$ はプールされた BERT の出力、 N_i は銘柄 i の隣接銘柄で、 $\hat{A} = A + I$ は因果チェーン [12] により構築される企業ネットワークの隣接行列で、 $\hat{D}_{i,i} = \sum_{j=0} \hat{A}_{ij}$ は diagonal degree matrix であり、 h_i は Graph Convolutional Network (GCN) への入力、 W は重み行列、 b はバイアス、そして $S_{\text{emb},i}$ は銘柄 i の事業活動のベクトル表現である。

そして、提案手法では Multi-layer perceptron (MLP) 層を用いる。

$$S_{\text{technical},t,i} = \text{ReLU} (W_{\text{tech}} \text{technical}_{t,i} + b_{\text{tech}}) \quad (6)$$

$$S_{\text{fundamental},t,i} = \text{ReLU} (W_{\text{fund}} \text{fundamental}_{t,i} + b_{\text{fund}}) \quad (7)$$

$$S_{\text{emb},i} = \text{ReLU}(W_{\text{emb}}\text{emb}_i + b_{\text{emb}}) \quad (8)$$

ここで $W_{\text{tech}} \in \mathbb{R}^{l \times k_t}$, $W_{\text{fund}} \in \mathbb{R}^{l \times k_f}$, $W_{\text{emb}} \in \mathbb{R}^{l \times k_e}$, $b_{\text{tech}}, b_{\text{fund}}, b_{\text{emb}} \in \mathbb{R}^l$ であり, l は銘柄のベクトル表現の次元数である. 銘柄のベクトル表現 q_i は以下のように計算される.

$$q_i^{\text{MF}} = S_{\text{technical},t,i} \odot S_{\text{fundamental},t,i} \odot S_{\text{emb},i} \odot p_i \quad (9)$$

$$q_i^{\text{MLP}} = [S_{\text{technical},t,i} \odot S_{\text{fundamental},t,i} \odot S_{\text{emb},i} \parallel p_i] \quad (10)$$

ここで p_i は銘柄 i のワンホットベクトルであり, \odot は Hadamard product を表し, \parallel は concatenation operation を表す. ここでは, Hadamard product を先に計算し次に concatenation operation を行う. q_i^{MLP} は MLP 層の銘柄ベクトルの入力で, q_i^{MF} は matrix factorization(MF) 層の銘柄ベクトルの入力である.

2.2 投資家特性の獲得

次に投資家の特性を分析する. u を投資家で, $I_{\text{Attributes},u}$, $I_{\text{Personality},u}$, $I_{\text{Behavioral},u}$ を投資家 u の属性, 性格特性, 行動特性とする. 投資家特性 I_u は以下のように示される.

$$I_u = [I_{\text{Attributes},u} \parallel I_{\text{Personality},u} \parallel I_{\text{Behavioral},u}] \quad (11)$$

投資家ベクトル r_u は以下のように計算される.

$$r_u^{\text{MF}} = I_u \odot p_u \quad (12)$$

$$r_u^{\text{MLP}} = [I_u \parallel p_u] \quad (13)$$

ここで p_u は投資家 u のワンホットベクトルである. r_u^{MLP} は MLP 層の投資家ベクトルの入力で, r_u^{MF} は MF 層の投資家ベクトルの入力である.

2.3 協調フィルタリング

最後に, 協調フィルタリングを用いて銘柄の推薦を行う. He *et al.* に従い predicted preference $y_{u,i}$ を以下のように計算する.

$$\phi^{\text{GF}} = q_i^{\text{MF}} \odot r_u^{\text{MF}} \quad (14)$$

$$\phi^{\text{MLP}} = a_L(a_{L-1}(\dots a_2(W_2[q_i^{\text{MLP}} \parallel r_u^{\text{MLP}}] + b_2)\dots) + b_L) \quad (15)$$

$$\hat{y}_{u,i} = \sigma(W_F[\phi^{\text{GF}} \parallel \phi^{\text{MLP}}] + b_F) \quad (16)$$

ここで σ はシグモイド関数, a_l は ReLU などの活性化関数, W_l と W_F は重み行列, b_l と b_F はバイアスを表している.

2.4 損失関数

損失関数は以下のように定義される.

$$p(\mathcal{Y}, \mathcal{Y}^- | \mathbf{Q}, \mathbf{R}, \Theta f) = \prod_{(u,i) \in \mathcal{Y}} y_{u,i} \prod_{(u,i) \in \mathcal{Y}^-} (1 - y_{u,i}) \quad (17)$$

$$\begin{aligned} L &= - \sum_{(u,i) \in \mathcal{Y}} \log y_{u,i} - \sum_{(u,i) \in \mathcal{Y}^-} \log(1 - y_{u,i}) \\ &= - \sum_{(u,i) \in \mathcal{Y} \cup \mathcal{Y}^-} y_{u,i} \log y_{u,i} + (1 - y_{u,i}) \log(1 - y_{u,i}) \end{aligned} \quad (18)$$

ここで \mathcal{Y} は正例, \mathcal{Y}^- は負例, $y_{u,i}$ は投資家 u の銘柄 i への implicit feedback であり, \mathbf{Q} と \mathbf{R} は銘柄と投資家の latent factor matrix である. 負例 \mathcal{Y}^- に関しては, それぞれの投資家が未だ保有していない銘柄の集合から uniform sampling を行った.

3 実験設定

3.1 データセット

3.1.1 銘柄特性

本研究においてテクニカル指標はボリンジャーバンド, 移動平均収束発散法, 相対力指数, 平均方向性指数, そしてストキャスティクスを用いた. ファundamental指標は一株当たり利益 (earnings per share), 一株当たり純資産 (book value per share), 株価収益率 (price-to-earnings ratio), そして自己資本利益率 (return on equity) を用いた.

3.1.2 個人投資家特性

本研究では個人投資家の特性として性格特性, 行動特性, 属性情報を用いた. 性格特性は 10 項目からならアンケート ten-item personality inventory (TIPI) [13, 14] を用いて収集した. 行動特性は危険選好 (risk preference), 時間割引率 (time discount), 自信過剰 (overconfidence), 双曲割引 (hyperbolic discounting), 符号効果 (sign effect), マグニチュード効果 (magnitude effect) を大阪大学の Japan Household Panel Survey (JHPS)¹⁾を参考にしてアンケートを作成して収集した. 属性情報は年齢, 収入, 金融資産の量, 投資経験を用いた.

1) https://www.iser.osaka-u.ac.jp/survey_data/survey_eng.html

3.1.3 取引データ

本研究では証券会社から提供された 969 名の投資家の 2020 年の 7 月から 2022 の 9 月までの取引データを用いて研究を行った。前処理として、取引データを m 投資家と n 銘柄の user-item matrix $U_f \in \mathbb{R}^{m \times n}$ を以下のように定義した。

$$(U_f)_{ij} = f(i, j) \quad (19)$$

ここで $q_{i,j,t}$ は投資家 i の時刻 t における銘柄 j のポートフォリオである。

$$q_{i,j,t} = \begin{cases} 1 & \text{if user } i \text{ holds stock } j \text{ in time } t \\ 0 & \text{otherwise} \end{cases} \quad (20)$$

本研究では implicit feedback collaborative filtering user-item matrix R を U_{fR} として Equation 21 に定義している。

$$f_R(i, j) = \begin{cases} 1 & \text{if there is } t \in T \text{ s. t. } q_{i,j,t} \neq 0 \\ 0 & \text{otherwise} \end{cases} \quad (21)$$

ここで T は全体の期間である。ここで R の各列は、それぞれの投資家 i が銘柄 j を保有したか否かを示している。

3.2 タスクと評価指標

本研究では投資家が次に保有する新規銘柄を予測するタスクである新規銘柄推薦タスクにより評価をおこなう。データセットから各投資家の最新のトランザクションをテストデータとして、それ以外のデータをトレインデータとして用いる。銘柄推薦の評価指標として HR@10 と NDCG@10 を用いる。評価時には過去に投資家が購買したことの無い 99 銘柄をランダムにサンプリングして、テストアイテムと合わせて合計 100 銘柄のランキングを取り、リストの上位 10 銘柄をもちいて評価を行う。

4 結果

表 1 実験結果

	HR@10	NDCG@10
Model		
GMF	0.476	0.307
MLP	0.551	0.390
NEUMF	0.551	0.388
Proposed Model	0.612	0.422

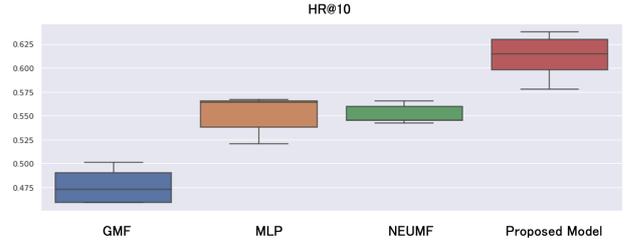


図 2 5つのシード値を用いた HR の箱ひげ図。ひげは最小値と最大値を示し、箱は第 1 四分位数、中央値、第 3 四分位数を表している。

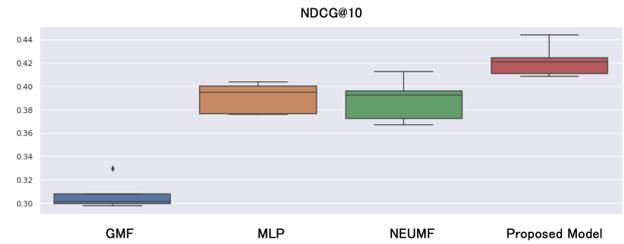


図 3 5つのシード値を用いた NDCG の箱ひげ図。ひげは最小値と最大値を示し、箱は第 1 四分位数、中央値、第 3 四分位数を表している。

表 1 は実験の結果を示す。提案手法、ベースラインモデルに対して五つの異なるシード値で評価し、スコアの平均を報告した。結果として提案手法が HR, NDCG 双方の評価指標でどちらもベースラインである Generalized Matrix Factorization(GMF), MLP, Neural Matrix Factorization(NeuMF) [15] をアウトパフォームしていることが確認できる。

図 2 と図 3 は異なるシードに対するそれぞれ HR, NDCG の箱ひげ図である。図からも提案手法が安定的にベースラインを上回ることが確認できる。

5 終わりに

本研究では銘柄特徴と投資家特性の Context を考慮した銘柄推薦モデルの提案を行った。本研究では銘柄特性として銘柄のテクニカル指標、ファンダメンタル指標、事業活動を、投資家特性として性格特性、行動特性、属性情報を考慮したモデルを提案した。実験結果により提案手法はベースラインを超える性能を示した。

謝辞

本研究は大和証券グループ本社とそのグループ会社である株式会社 CONNECT の支援を受けたものである。加えて、本研究は JSPS 科研費 JP21K12010 と JST 未来社会創造事業 JPMJMI20B1 の助成を受けたものである。

参考文献

- [1] Robin M. E. Swezey and Bruno Charron. Large-scale recommendation for portfolio optimization. In **Proceedings of the 12th ACM Conference on Recommender Systems**, pp. 382–386, 2018.
- [2] Yang Yujun, Li Jianping, and Yang Yimei. An efficient stock recommendation model based on big order net inflow. **Mathematical Problems in Engineering**, Vol. 2016, pp. 1–15, 2016.
- [3] Kohsuke Kubota, Hiroyuki Sato, Wataru Yamada, Keiichi Ochiai, and Hiroshi Kawakami. Content-based stock recommendation using smartphone data. **Journal of Information Processing**, Vol. 30, pp. 361–371, 2022.
- [4] JOSEPH E. ENGELBERG and CHRISTOPHER A. PARSONS. The causal impact of media in financial markets. **The Journal of Finance**, Vol. 66, No. 1, pp. 67–97, 2011.
- [5] Brad M. Barber and Terrance Odean. All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors. **The Review of Financial Studies**, Vol. 21, No. 2, pp. 785–818, 2007.
- [6] Stefano Dellavigna and Joshua M. Pollet. Investor Inattention and Friday Earnings Announcements. **Journal of Finance**, Vol. 64, No. 2, pp. 709–749, 2009.
- [7] Chris Brooks and Louis Williams. The impact of personality traits on attitude to financial risk. **Research in International Business and Finance**, Vol. 58, p. 101501, 2021.
- [8] Fatima Akhtar, K. Thyagaraj, and Niladri Das. Perceived investment performance of individual investors is related to the big-five and the general factor of personality (gpf). **Global Business Review**, Vol. 19, p. 097215091771352, 2017.
- [9] Cheng-Po Lai. Personality traits and stock investment of individuals. **Sustainability**, Vol. 11, No. 19, 2019.
- [10] Ummuhan Mutlu and Gökhan Özer. The effects of personality traits on financial behaviour. **Pressacademia**, Vol. 8, pp. 155–164, 2019.
- [11] Takayanagi Takehiro, Sakaji Hiroki, and Izumi Kiyoshi. Setn: Stock embedding enhanced with textual and network information. In **2022 IEEE International Conference on Big Data (Big Data)**, 2022.
- [12] Izumi Kiyoshi and Sakaji Hiroki. Economic causal-chain search using text mining technology. In **the First Workshop on Financial Technology and Natural Language Processing In conjunction with the 28th International Joint Conference on Artificial Intelligence**, pp. 61–65, 2019.
- [13] Samuel D Gosling, Peter J Rentfrow, and William B Swann. A very brief measure of the big-five personality domains. **Journal of Research in Personality**, Vol. 37, No. 6, pp. 504–528, 2003.
- [14] Atsushi Oshio, Atsushi Abe, Shingo Cutrone, and Pino Gosling Samuel. Big Five Content Representation of the Japanese Version of the Ten-Item Personality Inventory. **Psychology**, Vol. 4, pp. 924–929, 2013.
- [15] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. Neural collaborative filtering. **WWW '17**, p. 173–182. International World Wide Web Conferences Steering Committee, 2017.