Front. Comput. Sci., 2024, 0(0): 1–3
https://doi.org/10.1007/sxxxxx-yyy-zzzz-1

LETTERS

# Visual Analysis Approach for Mutual Fund Selection

**Fan YAN[1], Yong WANG[2], Xuanwu YUE[3], Kam-Kwai WONG[4], Ketian MAO[1], Rong ZHANG[4], Huamin Qu[4], Haiyang ZHU[1,5], Minfeng ZHU(✉)[1], Wei CHEN[1]**

1 State Key Lab of CAD&CG, Zhejiang University, Hangzhou 310058, China
2 Computing and Data Science, Nanyang Technological University, 639798, Singapore
3 Baiont Quant, Nanjing 210038, China
4 Computer Science and Engineering, Hong Kong University of Science and Technology, Hong Kong, 999077, China
5 Wuchan Zhongda Digital Technology Co., Ltd., Hangzhou 310020, China

© Higher Education Press 2024

## 1 Introduction

Selecting mutual funds is a major challenge for ordinary investors due to the vast number of available options and diverse investment preferences. Existing methods to support fund selection heavily rely on extensive user investment behavior data [1]. Furthermore, current online fund selection platforms primarily offer basic sorting based on *Rate of Return*, neglecting the individual preferences of typical investors.

To address this, we propose a mixed-initiative approach designed to help ordinary investors efficiently choose mutual funds. This approach integrates a user preference-based fund classifier and interactive user input, allowing it to quantify user preferences, automatically rank mutual fund candidates, and incorporate human expertise seamlessly for efficient and reliable fund selection. We evaluated our approach with eight ordinary investors.

Received month dd, 2024; accepted month dd, 2024

E-mail: minfeng_zhu@zju.edu.cn

The major contributions of this work include:

- We propose a mixed-initiative approach, informed by domain experts and publicly generated investor content, to achieve efficient mutual fund selection.
- We evaluated our approach with a study and experiments involving eight investors.

## 2 Methodology

We introduce a mixed-initiative approach for fund selection comprising three modules. The data processing module calculates key technical indicators as the foundation for subsequent steps. The user preference-based fund classifier automatically ranks funds based on investor preferences, refined through feedback. The user interaction module presents refined investment preferences and ranked fund details, helping investors annotate misordered funds.

Our approach emphasizes the mixed-initiative iterative loops involving the user preference-based fund classifier and user interaction. Initially, the fund classifier generates a ranked list of funds using the current preferences and fund indicators as inputs. This ranked list is passed to the user interaction module. Then investors provide feedback on specific funds, such as indicating whether a fund is over- or under-ranked based on their preferences. The feedback, in the form of annotated user choices, is then sent back to the fund classifier. The classifier refines the investor's investment preferences based on the feedback and generates an updated ranked list of funds for the next iteration. This continuous engagement allows investors to refine the model progressively and gain a deeper understanding of their investment preferences, leading to more informed and insightful investment choices.

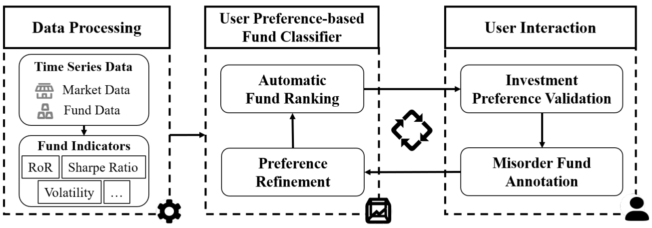


**Fig. 1** The approach for mutual fund selection.

## 3 Data Processing

The data processing module calculates seven indicators from fund Net Asset Values and Bond Market Index , categorized into three groups: profitability, risk resistance, and price-performance ratio.

Profitability: *Rate of Return(RoR)*is calculated as $RoR = (\frac{p_{end}}{p_{start}} - 1) \times 100\%$, with $p_{start}$ and $p_{end}$ marking the beginning and end prices. The *Capture Ratio* is defined as $CaptureRatio = \frac{n_{pos}}{n}$, where $n_{pos}$ is the number of days that the fund's *RoR* is better than the *RoR* of the overall market and $n$ is the day count. Additionally, *Winning Streak* is continuously positive during the investment period.

Risk Resistance: the *Maximum Drawdown (MDD)* is defined as $MDD = Max(\frac{p_i - p_{i+1}}{p_i})$, where $p_i$ is the price of the $i$-th time point. The *Volatility* is calculated as $Volatility = \sqrt{\frac{\sum_{i=1}^{n}(p_i - \bar{p}_i)^2}{n-1}} * \sqrt{n}$, where $p_i$ is the daily return of the $i$-th time point and $\bar{p}$ is the average of $p$ of $n$ days.

Price-Performance: the *Sharpe Ratio* [2] is defined as $SharpeRatio = \frac{r_p - r_f}{\sigma_p}$, where $r_p$ is the fund's *RoR*, $r_f$ is the rate of risk-free return which we use as the bond market index's *RoR*, and $\sigma_p$ is the volatility during the investment period. The *Calmar Ratio* [3] is calculated as $CalmarRatio = \frac{r_p - r_f}{m}$, where $m$ is the *MDD* during the investment period.

## 4 User Preference-based Fund Classifier

The user preference-based fund classifier uses the pairwise algorithm [4], transforming personalized recommendations into a binary classification problem, allowing investors to select the superior option from fund pairs based on their preferences.

### 4.1 Automatic Fund Ranking

Automatic fund ranking uses a fund classifier to score funds based on investor preferences. In this process, each fund is paired with every other fund in the market. The classifier evaluates these pairs by comparing the differences in each pair of funds' indicators and determines the superior fund by outputting a value between $-1$ and 1. For Fund $A$, its overall preference is calculated by summing all pairwise comparisons with other funds $F_i$:

$$C_A^k = \sum_{F_i \in S, F_i \neq A} \omega^k (A^k - F_i^k), \quad (1)$$

where $C_A^k$ represents the $k$-th indicator's contribution to Fund A's Sum Score. $A^k$ and $F_i^k$ represent the $k$-th indicator for Funds $A$ and $F_i$, respectively. $\omega^k$ represents the $k$-th indicator's weight, which can be either positive or negative. As a result, $C_A^k$ can be positive or negative.

### 4.2 Preference Refinement

To refine investment preferences, we prompt investors to identify misordered pairs from the fund ranking. For example, if Fund $A$ is deemed superior to Fund $B$, the indicator difference is labeled 1; otherwise, it's $-1$. By training the fund classifier through multiple loops of investor interaction, it can refine investors' investment preferences for fund indicators and predict which fund is superior between pairs of funds.

## 5 Evaluation

### 5.1 User Study

We invited eight ordinary investors from online financial service platforms, such as Alipay and Eastmoney, to participate in our study. These participants, representing individual investors with basic but not extensive investment knowledge, engaged in a one-and-a-half-hour semi-structured study.

We introduced our fund selection approach, demonstrated a simplified prototype, and then asked them to use the system for selecting funds from 2015-2016 and 2016-2017. To mitigate time biases, we randomly assigned periods for using and not using our approach. Participants completed a five-point Likert scale questionnaire (Fig. 2) assessing the approach's usefulness and usability.

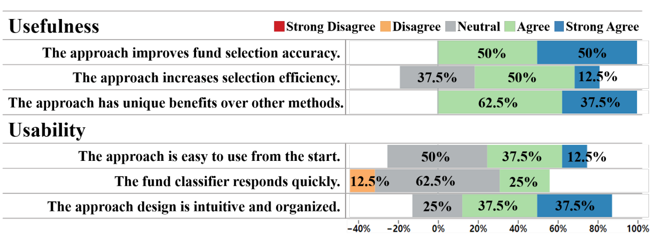


**Fig. 2** The user feedback on our approach.

### 5.2 Quantitative Experiments

To evaluate our fund classifier, we assumed investors have consistent preferences and defined rankings for funds. If our classifier can accurately match these predefined rankings, it demonstrates its effectiveness in mimicking investor preferences. Our data spanned 2019-2020 Chinese funds, with expert-defined target rankings guiding iterative refinements made by eight participants. These refinements were driven by discrepancies identified between the classifier's predictions and the expert rankings.

Ranking accuracy was measured for each of the eight participants using Mean Average Precision (MAP) and Normalized Discounted Cumulative Gain (NDCG) as shown in Fig. 3. After four rounds, all participants achieved a MAP above 0.875 and NDCG@20 over 0.9 with the fund classifier. Most top 20 target funds matched the recommended ones. However, there were minor discrepancies, primarily in the order of some funds at the list's end. Despite these variations, the classifier's accuracy is enough to support fund selection aligned with user preferences.

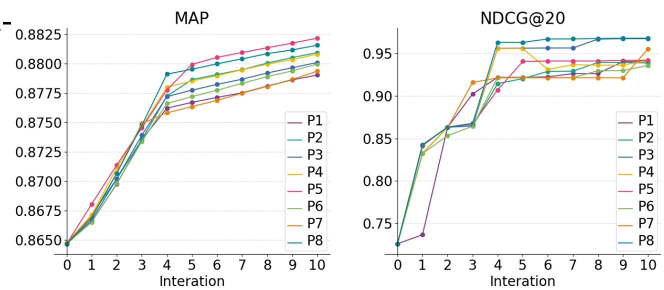


**Fig. 3** There are the MAP and NDCG@20 of the approach's ranking.

**Acknowledgements** This work was supported by the National Natural Science Foundation of China (No. 62132017), the Zhejiang Provincial Natural Science Foundation of China (No. LD24F020011), the Key R&D "Pioneer" Tackling Plan Program in Zhejiang Province (No. 2023C01119), and the "Ten Thousand Talents Plan" Science and Technology Innovation Leading Talent Program of Zhejiang Province, China (No. 2022R52044).

## References

1. Pan X, Wu L, Long F, Ma A. Exploiting user behavior learning for personalized trajectory recommendations. Frontiers of Computer Science, 2022, 16: 1–12
2. Sharpe W F. The sharpe ratio. Journal of Portfolio Management, 1994, 21(1): 49–58
3. Young T W. Calmar ratio: A smoother tool. Futures, 1991, 20(1): 40
4. Liu T Y. Learning to rank for information retrieval. Information Retrieval, 2009, 3(3): 225–331