A Hybrid Framework of Anomaly Detection for Mutual Fund Parent Companies

David Xuejun Wang[1][0000−0003−0724−7615] and Yong Zheng[2(✉)][0000−0003−4990−4580]

[1] Morningstar, Inc., Chicago, 60602, USA
david.wang@morningstar.com
[2] Center for Decision Making and Optimization,
Illinois Institute of Technology, Chicago, 60616, USA
yzheng66@iit.edu

Abstract. In this paper, we introduce a unique task, *the assessment of the mutual fund parent companies*, in our financial company, where the anomaly events associated with parent companies need to be identified and sent to financial experts to access the impact on related mutual funds. We propose a hybrid framework of anomaly detection to combine data-driven detection and experts-engaged tuning to enhance the identification process. Our experiments have demonstrated its effectiveness through the feedback from financial experts, utilizing a tracking record spanning from May 2022 to June 2023.

Keywords: Mutual Fund · Anomaly Detection · Hybrid Framework

1 Introduction

A *mutual fund* represents a collective investment scheme, aggregating capital from numerous investors to allocate in a diversified portfolio comprising equities, fixed-income instruments, and other securities. These funds are particularly favored by long-term investors for their accessibility, cost-efficiency, and potential for sustained growth. Financial analysts specializing in mutual funds undertake a comprehensive analysis of these funds, scrutinizing various dimensions such as performance metrics, associated risks, and alignment with investors' financial objectives. The establishment and operational framework of a mutual fund are structured by its sponsoring entity, known as *parent company*, which shoulders critical responsibilities encompassing the selection and supervision of fund managers, the implementation of robust risk management, adherence to regulatory standards, and the execution of strategic research and development initiatives. The spectrum of funds managed by a parent company is markedly variable. Prominent institutions, like Fidelity[3] and Invesco[4], manage an extensive array of over a thousand funds, whereas smaller entities may oversee a solitary fund.

[3] https://www.fidelity.com/
[4] https://www.invesco.com/

Periodic assessment of the parent companies is crucial, since the efficacy of a mutual fund is significantly contingent upon the attributes of its parent company. Such evaluative assessment is indispensable for elucidating the uniformity of fund performance amidst varying market scenarios and the parent company's adeptness in navigating fiscal and operational adversities. Moreover, an assessment of the parent company's fiscal solvency and governance practices offers indications regarding the prospective endurance, sources for risk management and expansion of the funds. This is a unique financial task identified in our company, Morningstar, Inc., which is a leading financial services firm renowned for its independent investment research and data analytics. Within our company, the evaluation of parent companies has primarily included manually monitoring company attributes and activities, and identifying unusual events on a biennial basis since 2002. This method is labor-intensive and prone to oversights, leading to occasional missed anomalies and off-schedule review (e.g., some incidents may occur within a shorter period, rather than 2 years).

Utilizing anomaly detection is a viable solution for streamlining and automating this process. Anomaly detection [4, 9] is the process of identifying data points that substantially deviate from the normative behaviors. It has been successfully applied in multiple areas or domains, such as financial markets [2, 1], healthcare [3], recommender systems [11, 10], etc. However, there are two significant challenges in our scenarios. On one hand, we do not have labels for anomaly detection, which results in the needs of unsupervised learning at the current stage. On the other hand, the anomalies flagged by these algorithms, although grounded in data patterns, may not necessarily correspond to genuine anomalous events that necessitate an unscheduled review of the parent company. Ensuring the quality and effectiveness of identified anomalies requires the involvement of feedback or perspectives from financial experts in the process, rather than using the data-driven anomaly detection algorithms only.

In response to these challenges, we propose a hybrid framework tailored for the detection of anomaly events pertinent to fund parent companies. This innovative solution comprises a curated selection of anomaly detection algorithms, the combination of a voting mechanism and a thresholding strategy, as well as experts-engaged tuning. The efficacy of the proposed framework was demonstrated through the feedback from financial experts in our experiments, utilizing a tracking record spanning from May 2022 to June 2023.

2 The Hybrid Framework of Anomaly Detection

The workflow in our proposed hybrid framework of anomaly detection can be depicted by Figure 1. This section presents a step-by-step discussion of various processing stages and the hybrid framework.

2.1 Data Preparation

The stage of data preparation is composed of two important parts – *feature collection* and *feature grouping*.

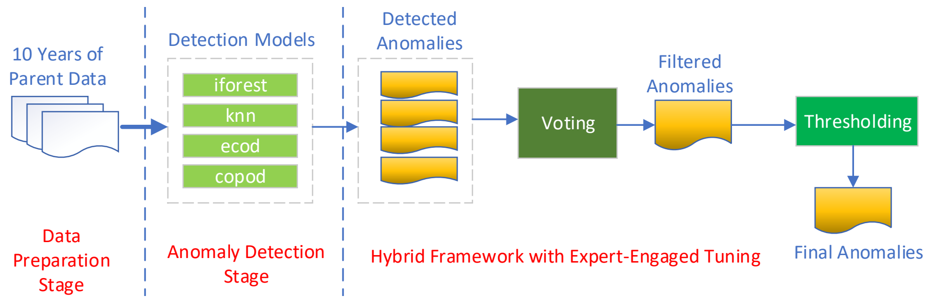


Fig. 1: The Workflow in the Hybrid Framework

Feature Collection. At the beginning, we asked fund analysts in our company to recommend the most relevant features for the purpose of anomaly detection among the mutual fund parent companies. We finally collected 39 features for anomaly detection. Below are some examples of these relevant features:

- Number of beginning funds: the number of funds under each parent company at the beginning of a month;
- Number of new funds: the number of new funds added under each parent company in a month;
- New fund percent: the new fund percent that refers to the proportion of new funds in relation to the total number of funds at the start of a month.;
- Number of fund managers: the number of fund managers in the parent company at the beginning of a month;
- Number of fund manager leaves: the number of fund managers left the parent company during a specific month;
- Manager leave percent: the ratio of manager leaves in relation to the total number of managers at the beginning of a month;

Feature Grouping. In practice, fund analysts consider anomaly events as abnormal behaviors over a group of related features. For example, a big jump in the number of manager leaves does not necessarily warrant an off-schedule review of the parent company, unless a significant negative cashflow appears at the same time. A substantial negative cash flow indicates a loss of investor confidence in the fund's operation, leading to a significant withdrawal of funds as a result of the manager leaves.

Therefore, detecting these anomalies in our scenarios requires examining sets of related features rather than analyzing each feature independently or utilizing the whole set of features. These sets of features, named as *events*, were derived from domain knowledge by the financial experts in our company. One example is the "Manager Leave $\rightarrow$ Fund Cashflow" event[5] which is composed by two features – the "Manager Leave Percent" that refers to the percentage of managers left the parent company in the current month, and "Cashflow Percent" that

[5] 'A $\rightarrow$ B' denotes the impacts on B by event A

denotes the percentage of total net cashflow over the total net asset of the parent company at the beginning of the month.

Based on feedback from analysts, we created 13 events that are relevant and useful to help detect different types of anomalies. These events or feature groups can be shown in Table 1.

Table 1: Number of Features in Events (Note: 'A $\rightarrow$ B' denotes the impacts on B by event A)

Event	**Number of Features**
Manager Leave → Fund Cashflow	2
Key Manager Leave → Fund Cashflow	2
Manager Leave → Fund Liquidation	2
Both Regular and Key Manager Leave	2
Manager Leave → Fund Return Rank	3
Manager Leave → Fund Relative Return	5
Fund Flow → Closed Fund and Assets	3
Fund Flow → Newly Open Fund and Assets	2
Fund Flow → Underperformed Funds and Assets	3
Fund Flow → Outperform Funds and Assets	3
Newly Created Funds → Merge and Liquidation Percent	2
Underperformed Funds and Assets → Rating Change & Performance Rank	5
Outperformed Funds and Assets → Rating Change & Performance Rank	5

Data Summary. We collected the monthly reports from 280 parent companies in the previous ten years. Therefore, we have 120 data points for each parent company at the beginning. Afterwards, we apply anomaly detection algorithms to assess each parent company based on each event or feature group defined above in the upcoming months. It's important to note that in the anomaly detection process, we consistently employ data from the preceding 10 years. For instance, we used data from the previous 10 years to predict or identify anomalies for May 2022. Subsequently, when making predictions for June 2022, the time window was shifted forward by one month to incorporate data from May 2022.

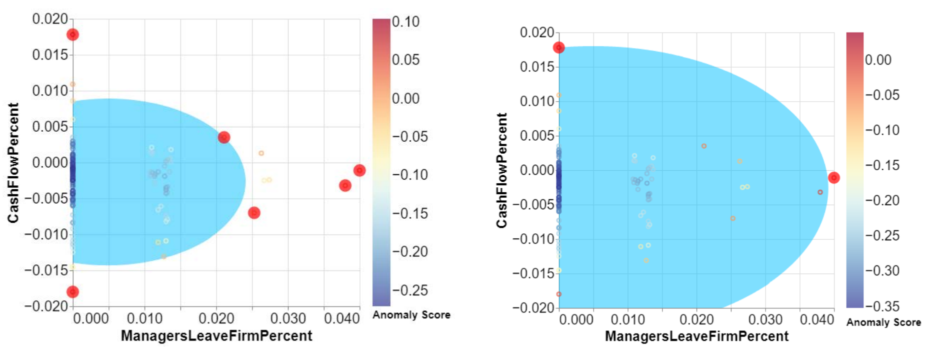
2.2 Anomaly Detection

The techniques of anomaly detection have been extensively studied. We employ the unsupervised anomaly detection algorithms in our experiments. More specifically, we utilized the following four anomaly detection techniques from the PyOD library [9] which is an open-source toolbox for scalable outlier detection with implementations of the state-of-art outlier detection approaches.

– K-Nearest Neighbor (KNN) [8] is a distance-based anomaly detection by measuring how close of the data point to the rest in its neighborhood.

- The ECOD [6] and COPOD [5] methods, where both of them are based on estimated joint probability distribution to identify anomalies by calculating tail probability, while ECOD approximates the joint distribution directly from empirical distribution functions of each variable, COPOD calculates the estimated joint distribution using a copula function.
- The isolation forest (IFOREST) [7] is an ensembling method which creates an ensemble of isolation trees to efficiently isolate outliers.

Parameter Tuning. There are two sets of hyper-parameters to be tuned in these anomaly detection algorithms. The first one is the contamination which is used to control the threshold of the decision function that output the anomalies. Contamination ranges from 0 to 0.5, and as its magnitude increases, the model generates a greater number of anomalies. Financial experts in our company helped us determine the optimal value for contamination. More specifically, the experts investigated the identified outliers from a visualization tool as shown by Figure 2. A larger contamination value may introduce more false positives (i.e., misclassified anomalies), where a small value may result in limited anomalies identified. From Figure 2, we can observe that using 0.05 as the value for contamination results in more identified outliers. However, financial experts also observed more false positives in these identified outliers. Through the feedback from experts, we finally decided to use 0.01 as the optimal value.



(a) contamination = 0.05 (b) contamination = 0.01

Fig. 2: Example: Visualizations for Decisions on Contamination

Another set is the algorithm-specific hyper-parameters, such as the number of neighbors and the distance measures in the KNN approach. We also tune up them carefully to seek the best results.

2.3 The Hybrid Framework of Anomaly Detection

At the beginning, we run each anomaly detection algorithm individually and reported the results to financial experts for evaluations. Nevertheless, the subsequent phase of anomaly validation, wherein the fund analysts were tasked with reviewing the anomalies flagged by these models, revealed a majority of false positives. Our fund analysts define an anomalous event as one meriting an unscheduled review of a parent company, a determination that hinges on a nuanced understanding of such events, while the anomaly detection algorithms operate purely on numerical data, devoid of this context-specific insight.

To address this challenge, we proposed and developed a hybrid framework. The methodology is hybrid due to two characteristics – on one hand, we utilize four anomaly detection together in the process of anomaly detection to be integrated with the *voting mechanism*; on the other hand, we additionally adopted a *thresholding strategy* where the analysts' knowledge and inputs were applied as additional filters to produce more accurate anomalies. It is worth noting that our framework is a combination of data-driven anomaly detection and expert inputs or expert-engaged tuning which will be highlighted in this section later.

The Voting Mechanism. Recall that the four anomaly detection algorithms can produce four sets of identified anomalies. For each unique identified anomaly, we define a metric, called *model popularity*, to determine if it is a reliable anomaly. More specifically, the model popularity (p) is defined as the fraction of models that identified a specific data point as anomaly. For example, if 3 out of 4 models detected a same parent company as an anomaly in the event of "Manager Leave $\rightarrow$ Fund Cashflow", then the model popularity p = 3/4 = 0.75. A minimum threshold, p_m, is used to filter out the detected anomalies with low model popularity. If the model popularity of a detected anomaly is greater than p_m, the anomaly is selected as a candidate of valid anomaly to enter the next step (i.e., the thresholding strategy). p_m is one of the hyper parameters. We experimented with values ranging from 0.5 to 1.0, incrementing by 0.1 at each step, in order to determine the optimal choice for p_m. Finally, the optimal value (i.e., it is 1.0 in our experiments) was determined based on the feedback from our fund analysts.

This voting or consensus mechanism operates by computing the proportion of models that concur on the identification of a specific anomaly, which results in more confidence in the pool of identified anomalies.

The Thresholding Strategy. We additionally introduce thresholds for the upper bounds or lower bounds associated with the scale of relevant features, instituted on the foundation of analyst expertise, serve to discern authentic anomalous events, thereby triggering the requisite off-schedule scrutiny of parent companies.

As previously articulated, the anomalies identified by algorithmic approaches may not align with the actual concerns of analysts, primarily due to the omission of expert knowledge. Illustrating this with the event "Manager Leave $\rightarrow$ Fund Cashflow", an anomaly may be marked by the anomaly detection algorithm for a scenario exhibiting at a 30% positive cashflow concurrent with a managerial departure. However, from an analyst's perspective, it is not necessary to be an anomaly, given the positive cashflow and the absence of managerial departures during the pertinent timeframe. An event would only be categorized as anomalous if it manifested a substantial negative cashflow coupled with manager leaves.

To address this, we established feature-specific thresholds within each event to authenticate the anomalies detected. These thresholds, which could represent either the minimum or maximum values (i.e., lower or upper bound) contingent upon the inherent characteristics of the data features, are meticulously defined within a configuration file. In other words, we invited financial experts to define the thresholds or limits for the upper bounds and/or lower bounds associated with relevant features, in order to filter out irrelevant anomalies. These threshold values were initialized based on the consultation with analysts, then they were finalized during the experts-engaged tuning phase. Namely, the fund analysts need to engage in the iterative tuning process, in order to determine the optimal thresholds for these features. After a series of tuning iterations (e.g., May to Aug in 2022), we finalized these thresholds and utilized them directly for the upcoming predictions (e.g., from Sep, 2022) without no more changes.

Experts-Engaged Tuning. Our proposed methodologies uniquely incorporate an expert-engaged tuning process. Crucial hyper-parameters, such as the threshold for model popularity in the voting mechanism and the feature thresholds in the thresholding strategy, are refined through an iterative, hands-on approach involving the feedback-loops from our fund analysts. This dynamic, experts-engaged refinement process ensures meticulous calibration of hyper-parameters, progressively enhancing the tool's performance and its congruence with the intricate demands of anomaly detection in the domain of fund parent companies. It is worth noting that these thresholds or hyper-parameters need to be initialized at the beginning of tuning iterations. Once they are finalized, the anomaly detection can be run seamlessly without anymore experts-engaged tuning.

3 Results and Findings

We consistently employ data from the preceding 10 years (i.e., using 10-years as the time window) to identify anomalies for the next month. We started anomaly detection and evaluations from May 2022 to June 2023. Fund analysts engaged in the evaluation process to examine the effectiveness of our frameworks. More specifically, we generate a list of detected anomalies for each month to be examined by the fund analysts manually. Analysts labeled each detected event with true or false. We updated the thresholds of features based on the feedback from analysts for the next prediction iteration. This experts-engaged tuning was executed from May to Aug, 2022 to finalize corresponding settings. For evaluation purposes, we compared methods in three scenarios:

- In the first scenario, we discard the thresholding strategy and employed the voting mechanism only. The optimal p_m was identified as 100%.
- In the second case, we used the hybrid framework with $p_m = 75\%$ in the voting mechanism and the thresholding strategy mentioned above.
- Last, we used the hybrid framework with $p_m = 100\%$ in the voting mechanism and the thresholding strategy mentioned above. This combination was demonstrated as the best choice in our experiments.

During the process of hyperparameter tuning and evaluations, we adopt precision as the metric, striving to ensure that a high percentage of the detected anomalies represent genuine irregularities. This precision-centric approach not only strengthens the tool's analytical prowess but also aligns closely with the analysts' expectations, thereby bolstering their reliance on and satisfaction with the tool's performance. We cannot assess the models using recall because it requires significant human effort to identify true positives over multiple feature sets within the entire dataset.

Table 2: Effectiveness of Anomaly Detection

Month	# of True Anomalies	Voting (p_m = 100%)		Voting (p_m = 75%) + Thresholding		Voting (p_m = 100%) + Thresholding	
		# of Detected Anomalies	Precision	# of Detected Anomalies	Precision	# of Detected Anomalies	Precision
May-22	1	4	25%	4	25%	1	100%
Jun-22	9	45	20%	89	10%	45	20%
Jul-22	11	26	42%	24	46%	19	58%
Aug-22	4	16	25%	12	33%	6	67%
Sep-22	4	20	20%	19	21%	8	50%
Oct-22	4	19	21%	18	22%	5	80%
Nov-22	3	22	14%	15	20%	5	60%
Dec-22	2	12	17%	7	29%	5	40%
Jan-23	3	25	12%	10	30%	5	60%
Feb-23	6	16	38%	8	75%	8	75%
Mar-23	7	32	22%	19	37%	8	87.5%
Apr-23	6	20	30%	12	50%	7	86%
May-23	N/A	14	N/A	15	N/A	6	N/A
Jun-23	7	19	37%	22	32%	9	78%

The experiment results from the three approaches above, in terms of precision values, can be shown by Table 2, where the precision metric was calculated based on a matching process from the total number of identified anomaly parent companies and the number of true positive provided by our fund analysts.

For the rows with yellow background in Table 2, spanning from May 2022 to Aug 2022, fund analysts were engaged in the tuning process to finalize the thresholding strategy. From Sep 2022, we used the same thresholds in the following months. The rows with pink background indicate the results associated with Sep 2022, Dec 2022 and Jan 2023, where the precision values were lower due to identified issues in the data sources (e.g., data input mistakes, invalid or inconsistent values, etc.). These issues were fixed in the following month, and have no impacts on the anomaly detection in next iterations. Moreover, we missed human evaluations in May 2023, where we used "N/A" to fill Table 2.

We can observe that precision values were generally higher by the third hybrid method, in comparison with the other two approaches. By ignoring the pink rows (i.e., unreliable results due to data issues), the third approach (i.e., voting with

$p_m = 100\%$ and the thresholding strategy) can help achieve a stable precision around 80%. The financial analysts considered it as the most effective approach, leading to a significant reduction in their workloads. Our company started to adopt this approach in our real-world practice from July, 2023.

4 Conclusions and Future Work

In this paper, we proposed a hybrid framework of anomaly detection for mutual fund parent companies, where both data-driven detection methods and experts-engaged tuning were adopted to enhance the performance. In the future work, we first will focus on creating on an automated adaptation mechanism that can automatically adjust the voting mechanism and the thresholding strategy based on the real-time feedback from analysts. We will also try semi-supervised outlier detection with collection of analysts' labels, though our data is growing slowly.

References

1. Ahmed, M., Choudhury, N., Uddin, S.: Anomaly detection on big data in financial markets. In: Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017. pp. 998–1001 (2017)
2. Ahmed, M., Mahmood, A.N., Islam, M.R.: A survey of anomaly detection techniques in financial domain. Future Generation Computer Systems **55**, 278–288 (2016)
3. van Capelleveen, G., Poel, M., Mueller, R.M., Thornton, D., van Hillegersberg, J.: Outlier detection in healthcare fraud: A case study in the medicaid dental domain. International journal of accounting information systems **21**, 18–31 (2016)
4. Chandola, V., Banerjee, A., Kumar, V.: Anomaly detection: A survey. ACM computing surveys (CSUR) **41**(3), 15 (2009)
5. Li, Z., Zhao, Y., Botta, N., Ionescu, C., Hu, X.: Copod: copula-based outlier detection. In: 2020 IEEE international conference on data mining (ICDM). pp. 1118–1123. IEEE (2020)
6. Li, Z., Zhao, Y., Hu, X., Botta, N., Ionescu, C., Chen, G.: Ecod: Unsupervised outlier detection using empirical cumulative distribution functions. IEEE Transactions on Knowledge and Data Engineering (2022)
7. Liu, F.T., Ting, K.M., Zhou, Z.H.: Isolation forest. In: 2008 eighth ieee international conference on data mining. pp. 413–422. IEEE (2008)
8. Ramaswamy, S., Rastogi, R., Shim, K.: Efficient algorithms for mining outliers from large data sets. In: Proceedings of the 2000 ACM SIGMOD international conference on Management of data. pp. 427–438 (2000)
9. Zhao, Y., Nasrullah, Z., Li, Z.: Pyod: A python toolbox for scalable outlier detection. Journal of Machine Learning Research **20**(96), 1–7 (2019)
10. Zheng, Y.: An empirical comparison of outlier detection methods for identifying grey-sheep users in recommender systems. In: International conference on advanced information networking and applications. Springer (2024)
11. Zheng, Y., Agnani, M., Singh, M.: Identification of grey sheep users by histogram intersection in recommender systems. In: The 13th International Conference on Advanced Data Mining and Applications. pp. 148–161. Springer (2017)