A Step Toward a Cybernetic Whale:

An Interim Report on "A Study on the Use of Artificial Intelligence for Learning Characteristics of Funds' Behavior"

Takao Tajiri, Takahiro Sasaki, and Hiroaki Kitano

Sony Computer Science Laboratories, Inc.

December 2019

This report should be cited as: Tajiri, T. , Sasaki, T. and Kitano, H. (2019). A Step Toward a Cybernetic Whale: An Interim Report on "A Study on the Use of Artificial Intelligence for Learning Characteristics of Funds' Behavior". Tokyo, Japan: Government Pension Investment Fund.

ExecutiveSummary

Highlights of this intermediate progress report

This document reports an intermediate progress of the commissioned research by Government Pension Investment Fund (GPIF) to Sony Computer Science Laboratories, Inc. (Sony CSL) on "A Study on the Use of Artificial Intelligence for Learning Characteristics of Fund's Behavior".

Major progresses are:

1. Development of a new style detection mechanisms termed "Resembler" that captures an inherent style of each fund manager that may not be captured by mere combination of simplified styles.

2. Resembler detected potential style drifts that were not detected by Style Detector Array (SDA) used in the previous report. The use of both Resembler and SDA significantly improved analysis capability of fund manager's style and multi-dimensional style drift detection capability. Mutual-resemblance uncovers similarities among funds and their temporal changes assisting GPIF to maintain diversity of funds from their actual investment practices.

3. An explorative study was performed to identify characteristics of fund manager comparing against benchmark to identify where excess return actually come from and visualized using Self-Organized Map (SOM). This may enable us to analyze if excessive returns are due to luck or skill. An intermediate result is promising. It implies that GPIF may be able to uncover skill vs. luck of investment results data-driven manner and much more precisely than fund managers.

4. Continuation of such research and development shall enable GPIF to analyze, visualize, and explain explicit and implicit investment behaviors of each fund manager, manage overall risk profile, and optimize GPIF portfolio. Most analysis and visualization can be made semi-automated and dramatically improve the efficiency, accuracy, and accessibility of profiling on manager structure.

5. The system is now implemented on ABCI —a 1088 nodes GPGPU cluster at the National Institute of Advanced Industrial Science and Technology (AIST).

As a result of these improvement, the system is now capable of detecting and analyzing behaviors of fund managers and their relationship in the context of overall GPIF investment far more precise than previously possible. This shall dramatically improve GPIF's capability to understand behaviors of fund managers and engaging in-depth dialogue with them.

These outcomes have major implications on the direction of GPIF operation beyond improvement of the current practices. Two future directions for GPIF has been suggested that are:

1. GPIF's augmented capability to access long-tail side of the players, and

2. GPIF-metrics for evaluating fund managers that is based on AI and data analysis characteristics.

These directions may represent evolution of GPIF's fund management structure and more fundamentally transformation of the organization.

AI and data-driven approach as described in this report direct the future where major asset owners can dissect and analyze behaviors of fund managers from multiple dimensions and uncover their characteristics much accurately and precisely than fund managers themselves. It may result in creation of GPIF-metrics that reflects policies and value of GPIF. Furthermore, the overall analysis process can be semi-automated, including report generation used for due processes. This opens up a new opportunity for both GPIF and smaller funds, especially this gives GPIF access to the "long tail" of fund scale in the investment management industry. Due to statistical feature of complex systems including a global stock market, how to access and benefits from long-tail part of the players and assets is critically important. The use of AI that enable GPIF and other major asset owners to access long-tail part of the players could be a major breakthrough and may trigger qualitative transformation in the industry.

An implication is clear. Whale will turn into Cybernetic Whale, that is empowered by AI. Cybernetic Whale will be able to see what human may not be able to see without the assistance of AI.

1 The context of the study and the interim report

This is a follow-on to a research study conducted in FY2017, under which a proof-of-concept prototype system called Style Detector Array (hereinafter called "SDA") based on deep learning was developed [1]. The previous study drew interest from asset owners in Japan and globally, leading to awards from overseas media. Building on those results, we developed a system called "Resembler" as an extension and application of SDA in the current research program. Resembler provides quantitative metrics for a type of information about fund management firms that previously was only qualitative: their distinct characteristics or uniqueness. This system will help GPIF address two concerns about its qualitative evaluation of fund management firms: (1) difficulties in avoiding arbitrariness and subjectiveness, and (2) the dependence on a small pool of staff with specialized expertise.

	FY2017	October 2018 - March 2020
Universe size and asset class	Small universe (100 Japanese equities)	Large universe (1,000 Japanese or foreign equities)
Indices to be detected	Investment style based on widely-used traditional factors	Resemblance among fund managers' characteristics
Target of the development	Proof-of-concept prototype to validate the possibility of detecting investment style	New model and various system upgrades to support experimental implementations

This report covers interim results as of October 2019 of a research program that began in October 2018; since research is still ongoing, many of the findings contained herein are provisional. We plan to issue a final report covering the entire program of research after its completion in 2020.

2 Methodology: Investigating funds' characteristics based on trading data

SDA and Resembler are systems built on a deep learning based neural networks to classify and detect fund characteristics. These two systems have the same architecture and function. SDA is trained on virtual trading data generated by simulations of virtual fund managers. On the contrary, Resembler uses actual trading data from actually existing funds as a training data. This difference results in SDA detecting investment style, an objective indicator defined in terms of factor exposure, while Resembler supplies a relative indicator of a fund's "uniqueness", a distinct trading style of each fund.

In FY2019, we made significant advances in our research resulted in building Resembler as an extension of SDA and develop its applications. Resembler is a system that evaluates self-resemblance, by comparing a fund to its own past characteristics and uniqueness, and mutual-resemblance, which is the degree of inter-similarity among multiple funds. Development has proceeded through cooperation with GPIF on experimental implementations on their front-line.

SDA, developed during the previous research contract, was refined by re-evaluating what factors it should targets for detection, feeding more data by scaling up the universe of equities, introduction of round-robin sampling for equity selection, and refining the trading logic of the VFMs that generate training data for SDA.

Since beginning of FY2019, we have deployed our system on AI Bridging Cloud Infrastructure (ABCI) [2, 3], which is provided by the National Institute of Advanced Industrial Science and Technology (AIST). ABCI is a GPGPU cluster computing platform, which provides an optimal environment for the vast calculations that is required in training deep learning neural networks like SDA and Resembler. ABCI ' s exceptional cost-performance enabled our research to progress far more efficiently than in FY2017. We anticipate making even greater use of it as we move forward.

In addition to SDA and Resembler, we have been applying self-organizing maps (SOMs) as a means of visualizing fund characteristics to provide a means of deepening analytical insights. We are exploring usage of this technique to track trends in the equity holdings of a fund over time, and to evaluate the skill with which the fund is managed.

3 Results of experimental implementation in GPIF front-line operations

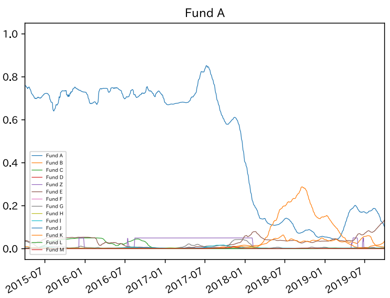
An initial evaluation was carried out using SDA and Resembler on all the actively managed Japanese and foreign equity funds in GPIF's manager structure. In this interim report, we highlight four case studies of special interest. (For this summary, abbreviated findings of two of those case studies follow). Here we would mention two particular cases that illustrate the value of the approach taken in this study.

Case Study 1: Resembler detected change while SDA unchanged

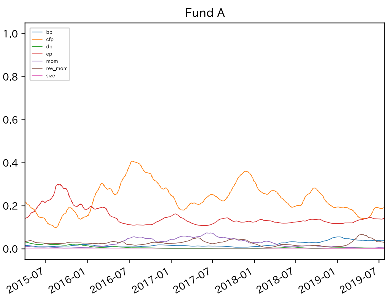
In Case Study 1, during the period of observation the fund's self-resemblance measured by Resembler plunged at one point and remained at a low level for some time, while mutual-resemblance to other funds edged upwards; later the self-resemblance indicator slightly trended up again. No corresponding changes were detected by SDA. The suggested cause is multifactor, involving equity holding shifts and changes in buying and selling behavior such as turnover; the timing of these can largely be seen as reflecting efforts to improve the fund performance. GPIF received advance notification of model changes and changes in the concentration of portfolio holdings. In reality, however, it is difficult to flag sequences of events as being a connected wave of subtle changes, since communication between GPIF and funds is wide-ranging, with many points of discussion, and mostly each anomaly has to be discussed as an isolated event. Thus, even if GPIF dives into the details with the fund manager, it would be difficult to uncover the whole picture without support of advanced analysis system as we discussed here. Resembler makes it possible to accurately flag large waves in order to establish new hypotheses, which support discussions about how to improve a fund's performance by narrowing down the points at issue. In addition, Resembler provides a quantitative basis for logging changes in fund behavior that up to now depended on the intuition of GPIF personnel that something is "off". This could yield an operationalizable basis for improvement.

Case Study 4: Fluctuating Resembler index

In Case Study 4, large decline in self-resemblance was observed concomitant with a certain degree of rises and falls of resemblance to various other funds Resembler's output. This implies the fund in question has a policy of opportunistically switching between different investment strategies based on market conditions, and this can be considered an investment strategy unto itself. In this case the lack of a firm self-resemblance is what is unique about the fund. On the other hand, funds of this nature can lose flexibility if they become locked into one strategy over a long period, and even if they retain their flexibility, their switching could start to lag behind changes in market conditions, scenarios which merit more detailed examination.



(a) Output of Resembler for fund A

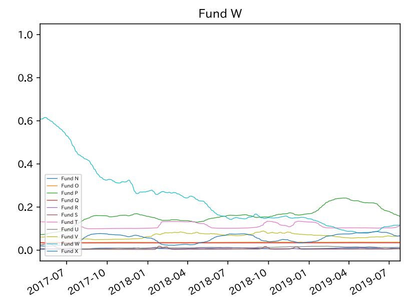


(b) Output of SDA for fund A

Case study 1

Interim summary of experimental implementations

Concerning the deployment of AI in front-line operations at GPIF, we figured out a three-step process which seems promising: 1) Using AI system (Resembler/SDA) for initial detection of changes; 2) Performing causal analysis on detected changes; 3) Delivering output in a form that is immediately actionable in the operational environment. Step 2) in particular would require a systematic framework for a seamlessly integrated processing flow. For example, a follow-on investigation is now initiated on whether approaches including topic analysis and VAE would be suitable. On step 3), in the course of ongoing experimental implementations, we are testing ways of integration of our AI system within current operations of IT systems that output human-readable text descriptions of the key features of changes, which makes communicating with fund management firms as smooth and efficient as possible.



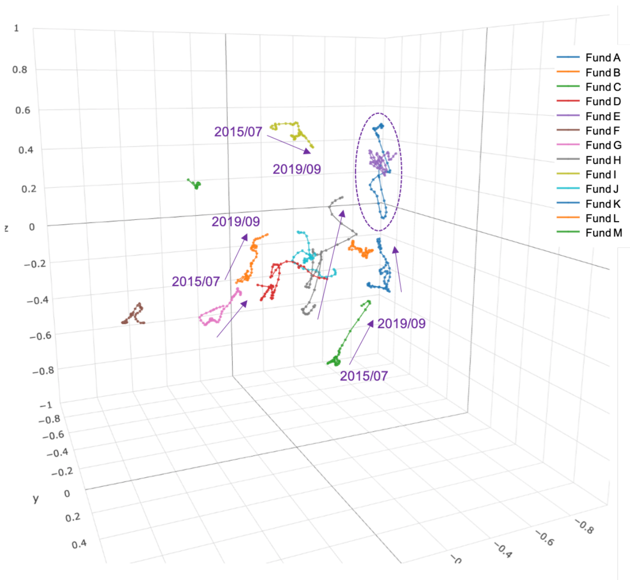
Case study 4: Output of Resembler for fund W

4 Exploratory Studies on Advanced Analysis

In parallel with the development of Resembler and SDA, we are making progress on a number of experimental implementations including: (1) a resembler-based investment strategy visualization and (2) a return quality analysis using Self-Organized Map.

The first one is an application of Resembler for visualization based on mutual-resemblance, which is the degree of similarity among the characteristic behavior patterns of multiple funds. It plots mutual-resemblance as time-series in phase space. In an example analysis case, although the characteristics of each fund, represented as coordinates on the resemblance map, are initially fairly well-scattered, they gradually converge toward two funds positioned in the upper right of the graph over time. It is considered that market condition in 2019, which performance was positive with low risk diversified investment, induced an overall shift toward similar strategy. Unlike relying on conventional diversification assessment by factor exposure, this is higher level diversification or convergence assessment with more expansive characteristic of investment management. GPIF that allocates among a set of many different fund managers is the entity which can overview this kind of trend and which is able to apply data for the purpose.

Secondly, experimental implementation was carried out to investigate a method of evaluating the “edge” of an active manager: the degree to which its performance (alpha) is attributable to skill rather than luck. When fund managers report to GPIF on their investment activities, there is a psychological incentive for them to attribute performance to their skill if performance has been good, but to blame bad luck due to unfavorable market conditions or other economic factors outside their control when performance has been poor. In that context, if there was a systematic methodology to classify what component of a fund’s positive (or negative) returns is attributable to skill and what component is attributable to luck, it would provide a common ground on which GPIF could evaluate fund performance without being swayed by the spin of fund managers. Based on this idea, we used the pattern of active weights in the fund ’ s portfolio to divide between domains of intentional active weight and non-intentional active weight, visualized these by SOMs, and attempted to compare their respective active returns. We found that for clusters of equities with a high degree of intentionality behind the active weight, some had correspondingly good returns, but some exhibited negative returns, meaning that the fund manager’s made a bad call. There were also cases where equities delivered positive returns even though the fund was not holding them with the intention of achieving active returns, but only out of necessity for risk management purposes.



Change in mutual-resemblance for Japanese funds over time

We consider firther investigation shall enable us to identify true skill of fund manager.

5 How AI could enable GPIF to unlock value

Transformation of operations at GPIF

Needs and expectation of GPIF for the use of AI in daily operation has been carefully identified and agreed between GPIF and Sony CSL team. During experimental implementations and trials, our team had close, deep, and frequent discussions on the requirements for business process at GPIF. Through such discussions, consensus has been developed on what GPIF expects to AI and how to apply it. There are feedbacks from persons in charge at GPIF regarding specific benefit and expectation once AI applied, these include: "Monitoring broader range of information", "Efficient managers evaluation", "Prediction based monitoring" and "Automation of selecting new manager".

The ultimate goal of this research is to find solutions for and more rigorously define issues raised at GPIF, referring the potential for accusations of arbitrariness and subjectivity with qualitative evaluation and the process being is heavily dependent on a small pool of staff with specialized expertise. These include detection of any changes in manager reports, confirming how much improvement is conducted by managers which have been put on notice to improve performance and detection of unreported changes. These would serve as the basis for improvement of more constructive and efficient discussion with managers and for standardization and commoditization of skills by exploiting big data at GPIF.

How "Whale" attracts the industry

If AI is fully deployed at front-line operations in the future, then there will no longer be a distinction or information gap in evaluating pre-contract candidates versus existing managers. In addition, GPIF's onerous reporting requirements have blocked smaller fund management firms which do not have sufficient resources, however the barrier would be lowered for them to pitch themselves to GPIF, giving GPIF access to the "long tail" of the fund scale in the investment management industry, if the output of AI systems was able to take over and eliminate much of this reporting obligation and workload. Asset management industry as a whole could benefit from being able to redirect resources from reporting to core investment performance. Since asset owners, including GPIF, are in a position to accumulate bigger data than asset managers, this could allow them to develop the most insightful analytical results, giving asset owners the information advantage over asset managers. This paradigm shift echoes the "Money Ball" strategy that enabled successful teams to be built in Major League Baseball by employing SABRmetrix. This new approach we might call "GPIF-metrix" would lead to the formation of "manager teams" which do not depend on traditional and established value and attract young generation and investment professionals.

GPIF's capability to efficiently evaluate and monitor smaller funds and potential GPIF-metrics will transform GPIF into AI and Data-driven organization that may set a new standard in the industry.

6 Future work

While the initial stage of our research started in October 2018 was basic and exploratory investigation, current one has been shifted to an application phase as pilot test. The application phase has cleared numerous suggestions and issues to be resolved from actual business process, and this phase is moving forward in parallel with the following agenda: "expansion of experimental implementation", "deepening, improving, and operationalizing based upon trial result" and "streamlining and integration of the technology platform".

7 Implications

- GPIF, the world's largest public pension fund, AIST that operates ABCI, a GPGPU cluster attaining the world-class computing performance, and Sony CSL with deep expertise and a proven track record in AI, as a trio of Japanese based organizations, have combined forces to break new ground in experimental implementation of AI in investment management industry. This demonstrates managing major funds may require cross-boundary highly competent team and resources at the highest-level.

- As these research results have implications for all asset owners collectively, we propose a "Global Data Consortium" to solve essential problems at investment management industry, such as "Whales breach wildly because an earthquake happened, or whales breaching wildly causes the earthquake to be happened". We expect a series of discoveries can be made that should uncover reality of investment practices for large-scale asset owners, and scientific understanding should be possible. Such efforts shall open a new era of science of investment and asset management.

- There are broader subject critical to asset management where basic scientific and engineering insights can be highly applicable. Many of them are implicated in the development of AI applications.

- As AI systems move toward operational deployment, the time has come to make clear plans for redirection of human resources. This affects the organizational composition of asset owners as well as fund managers.

Contents

1 The context of the study and the interim report

This paper is an interim report on "A Study on the Use of Artificial Intelligence for Learning Characteristics of Funds' Behavior" by Sony Computer Science Laboratories, Inc. (hereinafter called "CSL") which was commissioned in October 2018, by the Government Pension Investment Fund of Japan (hereinafter called "GPIF").

The following sections describe background of research and technical details of the research commissioned in October 2018. Section 5 discusses implications of outcome obtained to date.

1.1 Revisiting FY2017 research

In 2017, GPIF commissioned Sony CSL to conduct a research titled "A Study on the Use of Artificial Intelligence within Government Pension Investment Fund's Investment Management Practices." The research was oriented toward harnessing the power of AI for one of GPIF's core functions, which is maintaining the manager structure. This function means to oversee the allocation of GPIF's pension assets among a diversified set of fund management firms (hereinafter called "funds" or "managers"). The aim of the research was to provide more sensitive tools for detecting whether the investment behaviors of funds are kept consistent with their prospectus, and if not so evaluate whether this deviation is appropriate or need to be watched carefully.

For this research program, we developed a proof-of-concept prototype of deep learning AI system which we named as Style Detector Array (SDA). SDA takes fund manager's trading behavior data as input, and outputs the fund's character which is expressed as a combination of factor-based investment styles such as "value", "market cap", "momentum", etc. For the ease of explanation toward audiences in the financial industry, we chose these well-known investment styles as "reference indices" for our first SDA. The system's architecture, however, can take any other reference indices, not limited only to the factor-based styles mentioned above. The initial goal of the development was to validate whether the fundamental principle of our idea works properly or not. Therefore, instead of considering the full universe, we conduct experiments on small universe that consists of 100 representative Japanese equities, selected mainly according to market cap. By directly analyzing the investment behavior of fund managers, we confirmed that style analysis could be performed in real time with an evidence-based approach, enabling GPIF to evaluate and select its fund managers more accurately.

The Summary Report published in 2018 [1] earned an "Academic Research Paper Of The Year - Machine Learning & Big Data" award for GPIF and the Sony CSL project members at The Volatility & Risk Premia Awards 2019 presented by EQ Derivatives, which is a special interest magazine serving the financial industry that caters to asset owners and hedge funds; this was the first Japanese organization to receive the award, and Sony CSL was the first non-financial institution to receive the award. Our research outputs contributed also to GPIF's win of 2018 Institutional Excellence Awards (its fourth consecutive win) from Asian Investor magazine, a special interest magazine for institutional investors in the Asia-Pacific region.

The paper received various feedbacks from asset owners, fund management firms, portfolio advisors, legal practitioners, and so on. The feedbacks included their expectation for extending our method to other asset classes than Japanese equities, questions about the difference between what SDA can detect and what existing tools do, questions about SDA's output and funds' performance relationship, awareness and debate about the "black box" property of AI, and more.

1.2 Research agenda and activities in FY2018-2019

With the success of on the preliminary outcomes obtained in FY2017, an extensional research program was initiated titled as; “A Study on the Use of Artificial Intelligence for Learning Characteristics of Funds’ Behavior”, runs from October 2018 to March 2020. In order to investigate the applicability of our system within the actual business scenes of GPIF, we have set our research agenda and experimental conditions as the following table, based on the discussion described in last year’s report and from extensive communications within GPIF.

	FY2017	October 2018 - March 2020
Universe size and asset class	Small universe (100 Japanese equities)	Large universe (1,000 Japanese or foreign equities)
Indices to be detected	Investment style based on widely-used traditional factors	Resemblance among fund managers’ characteristics
Target of the development	Proof-of-concept prototype to validate the possibility of detecting investment style	New model and various system upgrades to support experimental implementations

Table 1: Comparison between FY2017 and current research activities

As described in the section titled “Process for selecting fund management firms” within GPIF’s annual report 2018 [4], GPIF emphasized the importance of qualitative evaluation of funds’ investment policies, operating procedures, and strength of organization/personnel. On the other hand, according to the minutes of 2017 Board of Governors, there are some concerns that qualitative evaluation of fund management firms could invite criticism for arbitrariness and subjectivity. Meanwhile, interviews with GPIF personnel revealed a process that is heavily dependent on a small pool of staff with specialized expertise. Meaning that if GPIF had access to quantitative metrics for a fund’s characteristics encompassing qualitative aspects such as investment policies, operating procedures, and strength of organization/personnel, changes in those aspects could be detect without any ambiguity. Thus, it could be the solution to GPIF’s concerns mentioned above. Selection and evaluation of funds could be fairer and more accurate. It could also provide a quantitative backing for the “expert’s intuition” or “sixth sense” of specialized GPIF personnel who noticed something is “off” about any fund’s behavior.

To solve these problems, we focused out research to implement extensions on SDA so that it could detect the fund managers’ “characteristics” embodying qualitative information, such as investment policies and operating procedures, that is difficult to quantify with existing analytical tools. IT was built on the basis of the core technology of SDA, which we initially implemented in FY2017 to output easy-to-understand investment style (such as “value”, “market cap”, “momentum”, etc.), but redesigned to meet these emerging demands.

This report covers interim results as of October 2019 of a research program that began in October 2018; since research is still ongoing, many of the findings are provisional. We plan to issue a final report covering the entire program of research after its completion in 2020.

2 Methodology: Investigating funds' characteristics based on trading data

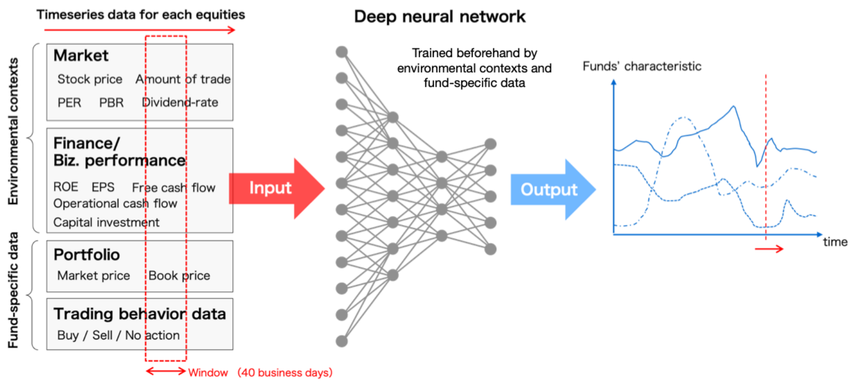


Figure 2: Overview of SDA and Resembler

Deep learning, which mimics the neural networks found in living organisms, is a technology currently at the forefront of the AI field. Without algorithms explicitly preprogrammed by humans, deep learning systems automatically learn rules for identifying or classifying tremendous amount of data from training datasets.Thanks to the abundancy of computing powers and the availability of huge data, it has proven to outperform human capability in some tasks especially in gaming, voice and image recognition and natural language processing. This triggered major excitements on deep learning as a major contributor to the current AI boom [5, 6]. Utilizing this technology, we have been developing Style Detector Array and its extension as Resembler that aim to derive funds' characteristics from trading behavior data.

In addition to deep leatning, we introduced self-organizing map, which is another technique in the field of AI, as a method of creating intuitive visualization of fund characteristics and changes that help humans to easily draw decision making from data.

In this chapter, we describe our methodology, focusing on the technical aspects.

2.1 Deep learning systems to detect and classify the characteristics of fund managers

Style Detector Array (hereinafter referred as "SDA") and Resembler are systems that detect and classify the characteristics of funds using deep learning neural network technology. Both systems need to be trained in advance before they can function as detectors/classifiers; in fact, they both have the same underlying system architecture, differing only in what data is used for initial training. They take as input data about the market environment and a fund's trading behavior, and output a n-dimensional vector whose attribute values represent the intensity of one of more (n) characteristics of interest (Figure 2)[1]. Since the input consists of time-series data, the n-dimensional attribute vector will track changes over time. Because the input includes not only fund's activity, but also the exogenous context of market conditions, the pattern of how the fund's behavior responds to particular changes in the market can be interpreted as that fund's characteristics.

[1] The name of Style Detector "Array" comes from its form of architecture that outputs n-dimensional vector.

2.1.1 Style Detector Array(SDA)

SDA was updated to handle larger asset universe with higher precision. Given a fund manager's trading behavior data as input, SDA is designed to output the fund's character which is expressed in n-dimensional vector, whose elements represent strength of (or similarity to) pre-defined reference indices. The references are selected from factor-based investment styles such as "value", "market cap", "momentum", etc. Although prototype development was completed as of our FY2017 research program [1], considerations of real world application were not a priority because the primary purpose at that time was to validate whether the fundamental principle of our idea works properly or not. Therefore, one objective of the current research program has been to upgrade SDA to be practically useful.

The key system component in enabling SDA to accurately classify investment behavior patterns by similarity to pre-defined reference styles is the deep learning neural network, which must be trained in advance. We took an approach to train SDA by using virtual trading data generated by Virtual Fund Managers (hereinafter called "VFM"). As such, analysis of a fund using SDA consists of the following three steps (Figure 3):

1. Generate virtual trading data through simulation of VFMs.
2. Train SDA using virtual trading data.
3. Input actual trading data from real funds into the trained SDA for analysis.

The reason for using virtual trading data in training the SDA is that there are no real-world funds which maintain a "pure" pattern of trading as defined in the reference styles over a sufficiently long time span to provide the quantity of data that would be needed for training.

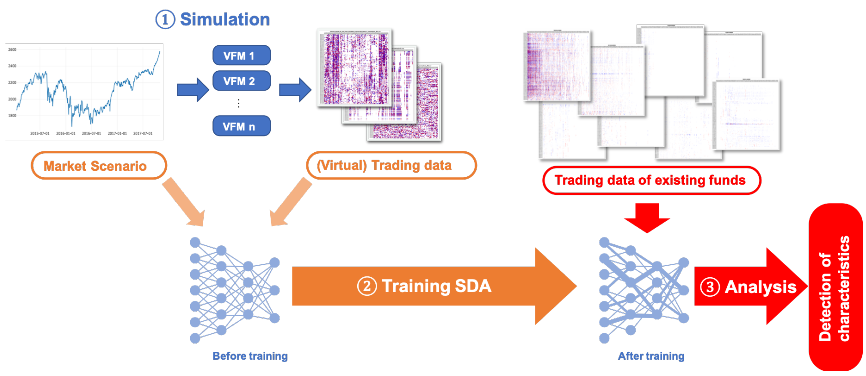


Figure 3: Flow of analysis made by SDA

Upgrading SDA by improving the Virtual Fund Managers

A VFM is an agent program that formalizes the logic of a reference investment style and it is used to carry out trading simulations under exogenous market conditions provided by historical data. Since SDA receives the virtual trading data generated by VFMs as training datasets, the effectiveness of SDA heavily depends on the quality of the VFMs implementation. Therefore, upgrading SDA must be accomplished by improving the VFMs. Below, we describe how our current generation of VFM has been improved over the previous prototype.

- Scaling up the universe of equities
  In developing the prototype system, we had to limit the execution time required for training SDA within affordable length, in order to carry out iterations of trials and errors as many as possible within a limited period of time. Therefore, we assumed a small universe consisting of the 100 largest Japanese equities by market cap to perform the VFM simulations and SDA training. In contrast, our current VFMs and SDA are now able to run on a much larger universe of 1,000 equities, so that the amount of data fed into the system has been increased. It is technically feasible to go beyond 1,000 equities and work off a full-sized universe of data. However, with progress by stages in validating SDA, we came to a judgment that by also applying round-robin sampling (discussed below), we can obtain virtual fund data that is sufficient for our current stage of analytical purposes. Accordingly, we have established a universe of 1,000 equities for our current testing environment.

- Round-robin sampling for defining the universe
  With the small universe used by the SDA prototype, limited to the top equities ranked by market cap, it was not possible to analyze funds holding a substantial number of equities outside that universe. Namely, it was impossible to analyze funds focused on small-cap stocks. Therefore, besides expanding the size of universe in our current SDA, we define it by picking equities by turns (i.e., round-robin) from each fund which is entrusted with investing for GPIF, in order of the fund's quantity of shares in that equity (Figure 4). Using this method, even when the size of the universe is capped at 1,000 equities, it provides 98% coverage (in the case of Japanese domestic equity funds) on a market value basis of equities held by all active funds in GPIF's portfolio. (Figure 5). This means SDA can now analyze a wider range of funds while making efficient use of the available data.

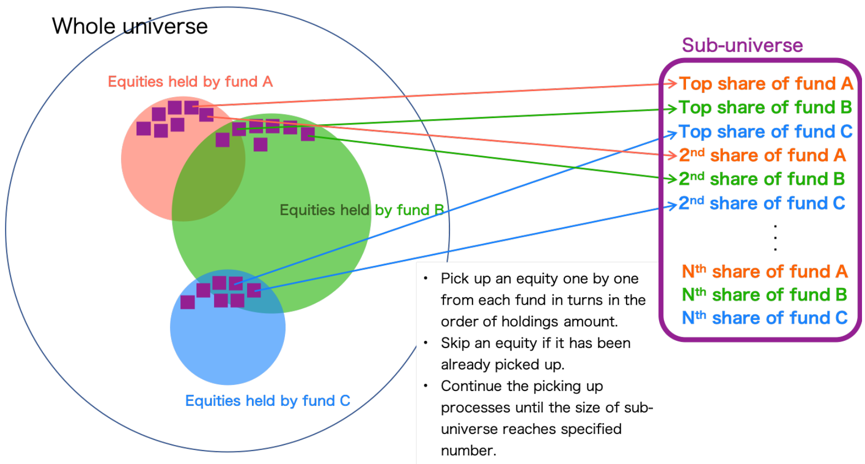


Figure 4: Round-robin sampling

- Investment strategies adopted as reference styles
  While a group of extremely simple and naive implementations of investment strategies were

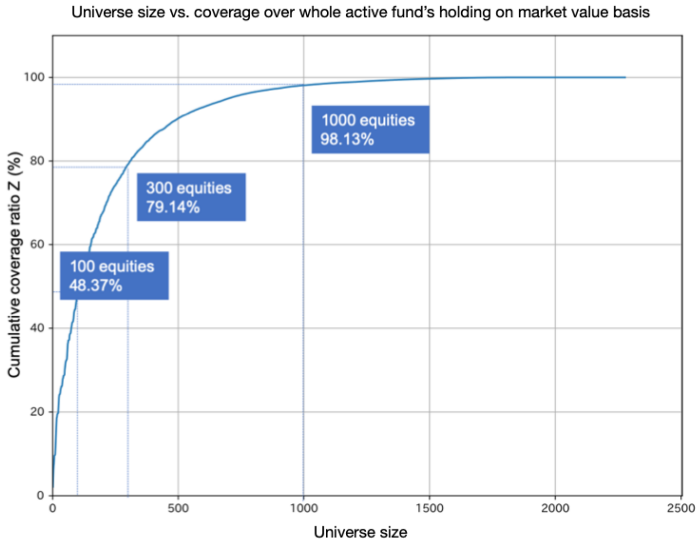


Figure 5: Coverage changes according to the size of sampled sub-universe

set as reference styles for the prototype systems, our current VFMs have been redesigned with real-world application in mind. Specifically, we referred to theoretical frameworks whose effectiveness has been demonstrated by academic research, including the Fama-French three factor model [7] and the Carhart four factor model [8], in order to capture orthodox risk factors that all bottom-up fund managers are conscious of. We defined VFMs that select equities based on seven indicators: BP (net assets), CFP (cashflow), DP (dividend yield), EP (earnings), MOM (long (12-month) positive momentum), Rev-MOM (short (1-month) negative momentum), and SIZE (market cap) [2].

Adding more and more indicators would be one possible approach to give SDA a more rich and elaborated ability of detection. On the other hand, our basic stance has been to avoid thoughtlessly expanding the parameters, since the nonlinearity of the deep learning algorithm within SDA will capture some of the nuanced spread of investment styles. However, since we have found in the outputs of SDA's testing to date that there are certain discrepancies with the understanding at GPIF, we plan to revise the mix of indicators that we use in order to correct this.

- VFM as aggregated clusters of sub-VFMs
  Because our prototype VFMs were implemented as monolithic agents and each executing trading on the basis of a "pure" and simple investment strategy logic, there was a tendency for them to exhibit extreme turnover of equity holdings ("all buy" or "all sell"). We enabled our current VFMs retain the simple logic while expressing a more natural pattern of trading activity, by having each VFM be an aggregate of multiple sub-VFMs with slight offsets to their trade timing. This eliminates discontinuities of equity turnover, reduces noise, and allows us to model the totality of market participants adopting an arbitrary spread of strategies.

[2]Factors which end in a "P" indicate the named parameter is "to market price", i.e., on a per market price basis

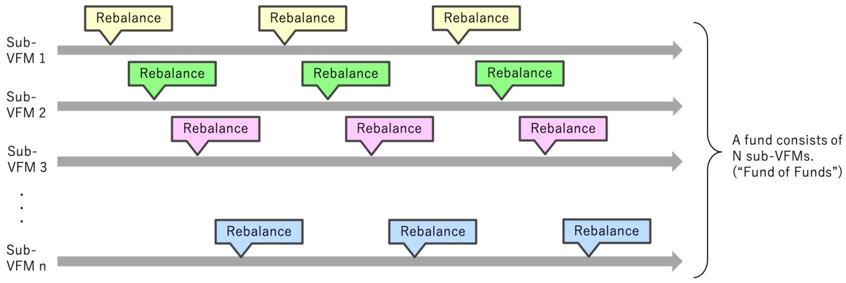


Figure 6: VFM as aggregated clusters of sub-VFMs

- Smoothing by category during equity selection
  Our prototype VFMs did not distinguish among categories such as industries, domicile (territory), etc., treating the universe of equities as monolithic with respect to these, and making decisions about whether to hold or not hold each equity by comparing all equities in the universe solely on the basis of the indicators targeted by the investment strategy style. However, since these indicators have mean baseline values that systematically differ among categories of industry, domicile, etc., there is a concern that this simplistic selection method just described would cause the portfolio to be skewed toward certain categories. Therefore, we introduced into the current VFMs' equity selection logic a weighting adjustment by category to ensure that the portfolio is unskewed.

Applying the latest version of SDA, upgraded in line with the concepts above, we analyze the characteristics of existing domestic and foreign active funds which GPIF has mandated asset management. Those results are the basis of case studies on several funds discussed in Chapter 3. For a list of all the results, including those not featured in case studies, see the Appendix. Although these results are closer to how GPIF's personnel understand funds' behavior than our previous prototype achieved, SDA is still not yet fine-tuned to a fully satisfactory capability. We are currently planning improvements by reconsidering the mix of factors given to train SDA, and by modifying the VFM logic.

2.1.2 Resembler

In parallel with upgrading SDA, we undertook the development of Resembler as an extensional application of SDA.

Given investment behavior pattern, SDA judges the degree of similarity to preset reference investment styles such as "value", "market cap", "quality", "momentum", and so on. As discussed in Chapter 1, we chose these well-known investment styles as "reference indices", for the ease of explanation of our technology and application toward audiences in the finance and investing sphere.

One of the key features of our system is that it is built with an array of neural networks that can be trained by any given data. Therefore, by choosing training data that reflects reference styles of analytical interest, users of the system can freely customize SDA for their purposes.

Resembler is one of such extensions and applications of that principle. Resembler takes as its training data the actual trading data of actually existing funds (instead of virtual trading data generated by VFMs as with SDA). Thus, Resembler's output is not a vector whose components show its degree of similarity to objective investment styles, but is instead a time-series of vectors that track the resemblance of a fund to each of the actual funds used as references: i.e., its resemblance to Fund A, resemblance to Fund B, etc. The system architecture and operating principle of Resembler are exactly the same as SDA; only the data used to train it are different. Nevertheless, Resembler is able to analyze existing funds in terms of their resemblance to other funds (or to themselves), which illuminates characteristics of fund behavior that SDA alone cannot adequately capture (Table 2).

	SDA	Resembler
Training data	Virtual trading data generated by VFMs	Actual trading data from existing funds
Target of detection	Objective indicators of "investment style" defined according to widely-used factor exposure	Relative indicators such as "resemblance to Fund A"

Table 2: Comparison between SDA and Resembler

The following are two examples of Resembler applications we have been carrying out so far. We, however, continue to explore other possible applications of Resembler, as well.

- Evaluating self-resemblance: comparing a fund to its own past characteristics.
  By including the past investment behavior patterns of the fund itself in the training data set, we can assess the degree to which a fund is maintaining consistency with its own characteristics and uniqueness; alternatively, to analyze self-resemblance as an indicator of whether the fund has been changing its style. To obtain such self-resemblance, firstly the data of all active funds targeted for analysis are fed as training dataset into the detector array, which creates the Resembler system, and then give to that system as input the data for each individual fund. During periods where a fund is strongly maintaining its particular characteristic, the self-resemblance score will remain high and stable. But when the fund deviates from its own particular characteristic, its self-resemblance score will also drop. Furthermore, in the event that the self-resemblance score is constantly fluctuating[3], we can conclude that the fund lacks a particular characteristic of its own, or that it is unclassifiable.

  Here it is important to note that changes in SDA and changes in Resembler are independent: in some cases they both happen at the same time, but in some cases one changes while the other does not change. This indicates the fact that SDA and Resembler address different aspects of the characteristics of funds. While SDA assesses fund behavior from the third person perspective in terms of commonly-used investment styles like value, momentum, etc. It is possible that actual investment style change without affecting composition of simplified style index as captured in SDA. Resembler, from the internal party perspective, can be thought of as quantifying characteristics such as the uniqueness and consistency of how the fund in question is managed, which conventional analysis could only describe qualitatively. By combining the analyses of SDA and Resembler, we expect that a more multifaceted understanding of fund behavior characteristics could be delivered. In Chapter 3, we use this framework to analyze several domestic and foreign equity funds as case studies. A table of the full results for self-resemblance of all funds is in the Appendix.

[3] In many cases, relatively high values of resemblances to other funds appear at the same time.

- Evaluating mutual-resemblance: comparing a fund to other funds.
  Firstly, Resembler is trained with investment activity data of a certain number of, say, n funds selected as "anchors". Next, given the data of another fund, the target of analysis, as an input, the already-trained Resembler outputs n-dimensional vector, whose attribute values express the degree of resemblance to the corresponding anchor funds. By measuring the distance between these vectors for multiple funds, we can evaluate the mutual-resemblance among these funds; i.e., how similar they are to one another. With regard to the anchor funds, we need to select a combination suitable to the purpose of the analysis. In principle, it is important to select anchors whose characteristics except for the control conditions vary from one another. In the case where we want to evaluate the resemblance only among quant funds, for example, then we should choose anchors from quant funds whose characteristics are different as much as possible from one another.

  By tracing the changes in these resemblance scores over time, it may support asset owners to judge whether their manager structures are maintaining the desired diversification. Our currently underway experiments to explore that concept are covered in detail in Chapter 4.

2.1.3 Using ABCI to improve the efficiency of development for SDA and Resembler

Traning deep learning system requires large data set and extensive computing time. Thus, we made a decision to deploy our system on a large-scale GPGPU cluster machine owned by the Japanese Government.

Training SDA and Resembler requires massive computing resources. There are a large number of adjustable external parameters for execution of SDA and Resembler. In order to obtain useful results, it is necessary to seek the optimal combination of parameters. This tuning process requires performing a large number of trial iterations while slightly tweaking the conditions of each run.

In addition, even after the combination of parameters has been set, when training a deep learning neural network, especially for complex problems, it is known that the output is sensitive to the values of parameters that are initially set at random at the beginning of training [9, 10, 11]. For that reason, we use the same data set to perform multiple training runs, and the ensemble average of the independent outputs is taken as the final output of the SDA or Resembler.

All of the procedures described above require enormous computation. In developing the prototype in 2017, a major reason that we opted for a small universe and simplified the implementation was the limited computing resources available.

Since beginning of FY2019, we have employed AI Bridging Cloud Infrastructure (hereinafter called "ABCI") [2], which is provided by the National Institute of Advanced Industrial Science and Technology (AIST). ABCI is a computing cluster constructed from 1,088 nodes with GPUs, and provides an optimal infrastructure for parallel process execution of deep learning algorithms [3]. While it depends on the number of nodes available simultaneously, with the typical configuration that we use, iteration cycles which took a week or more to run previously can now be completed in about a day and a half. As a result, our research progress has become far more efficient than in FY2017. We project that a vast number of iterations will be needed moving forward, so ABCI with its exceptional cost-performance will be a crucial key to the success of our research.

2.2 Where the alpha comes from? – Self-Organizing Maps for visualization of fund characteristics to explain what SDA and Resembler detected

Quality and sustainability of the fund's performance can be evaluated by identifying where the alpha comes from. Combination of Self-Organizing Map (SOM) and a series of deep learning based identifiers such as SDA and Resembler enable us to visualize characteristics of investment behaviors of funds in the context of the source of excessive return. Deep learning neural networks, which are the foundational technology for SDA and Resembler, can automatically and instantaneously identify rules within vast data sets that would not be evident from an ordinary perusal. They make it possible to tackle applications in areas such as pattern recognition and pattern classification without explicitly predefining those rules in a human-written formal computer program. However, this advantage comes with the disadvantage of having to accept the operation of the neural network itself as a black box. Since GPIF and other institutions which are fiduciaries of the public interest have an obligation to be transparent about the basis of their decision-making, this black box nature of neural networks is undesirable. In practice, even if a system like SDA or Resembler detects some kind of changes, but the reasons for those change cannot be explained, GPIF would, in many cases, not be able to proceed to decision making or taking action on that basis alone.

Therefore, we are seeking various ideas and use cases for existing methods in data science that would be suited to probing the causes of changes detected by SDA and Resembler. One promising method that we have tried is the self-organizing map (hereinafter called "SOM") as a technique for visualizing fund characteristics.

A SOM is a type of neural network that maps a large number of individual items, each of which has a certain number of attributes, onto a 2D or 3D spaces according to the resemblance among those attributes. In essence, SOM performs unsupervised clustering on the attribute vectors of complex high-dimensional data, and creates visual representations enabling humans to grasp trends and correlations that would be difficult or impossible to spot in the raw data. The basic flow of SOM is as follows:

1. An item is selected randomly from the dataset on which clustering is to be performed.
2. Compare the attribute vector of that item with the reference vector of each unit on the map and select the best matching unit.
3. Place the item in the best matching unit and adjust the reference vectors of the surrounding units to be closer to the item's attribute data vector.
4. Return to step 1.

When this series of operations is performed over a sufficient number of iterations, the resulting map will ultimately consist of clusters of items whose attribute vectors are most similar to each other[4] .

For our application, the items are individual equities held by a certain fund, and their corresponding attribute vectors consist of each equity's attributes, some of which (e.g., market cap, PER, PBR) are values independent of the fund, and other attributes whose values are specific to that fund (e.g., active weight, active return). Through this procedure, the equities held by a fund are plotted on a 2D map with a color gradation corresponding to the values of the attribute of interest, which makes it possible to readily discern fund characteristics and their changes over time.

[4]It is called "self"-organizing map, since items are automatically classified in an unsupervised way without any prior knowledge.

Although strictly speaking, we use SOM only for visualization, its usefulness is tremendous. For example, if we place the items (equities) on the 2D space simply in the order of securities tickers, we will just get random pattern of colors without any meaningful underlying structure, because the tickers are arbitrarily assigned to each equity without relationship to any of its attributes.

For example, in Figure 7, the equities held by a certain fund are clustered using a SOM, which has been colored according to the active weight in the equities corresponding to each unit. Red zones indicate equities which the fund overweight and blue zones indicate underweighted equity holdings. Zooming in on the units, it is possible to see which specific equities are in the cluster and investigate the attributes corresponding to the units. By generating a time series of SOMs depicting active weight, we can visually track changes in the fund's portfolio. In Chapter 3, we discuss the use of SOMs in this fashion to perform causal analysis of changes detected by Resembler.

Moreover, by fixing the positions of equities on maps colored with different attributes, we can conduct multifaceted complex analysis among various attributes of equities. For example, given two SOMs, one of which is colored according to active weight and the other by active return, we can visualize in a readily discernable form whether clusters of high-active-weight equities are actually generating high active returns. In Chapter 4, we discuss the case of what we named "Distiller", an experimental effort to differentiate the contributions of luck from those of skill in generating active return.

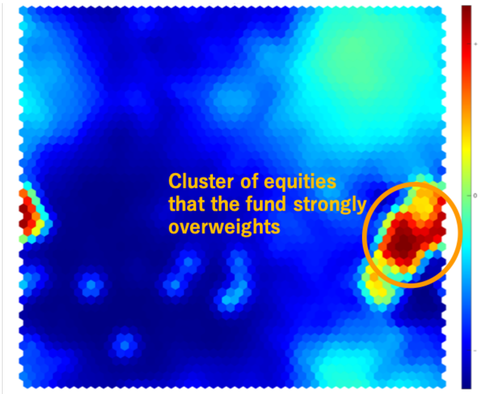


Figure 7: Visualization of holding patterns through SOM

3 Results of experimental implementation in GPIF front-line operations

A series of technologies described above has been implemented in our experimental system and tested using data on all the actively managed Japanese and foreign equity funds to which GPIF has allocated assets.

In this section, we will report four cases in depth as we judged them to be of particular value for gaining insight. The results for all funds analyzed can be found in the Appendix.

The workflow in all cases began with firstly examining the output of Resembler and SDA, and then proceeding to causal analysis using other techniques where features of the case warranted closer study.

3.1 Case studies on domestic (Japanese) equity funds

Case Study 1: Assessment using data from all periods

Case Study 1 considers Fund A. In this case, using data from all periods, we assessed whether changes took place in the past or not, and the causes of those changes; we then considered how Resembler could be deployed most effectively into GPIF's operations.

< Detection of changes >

The output of Resembler shown in Figure 8a exhibits the following three trends.

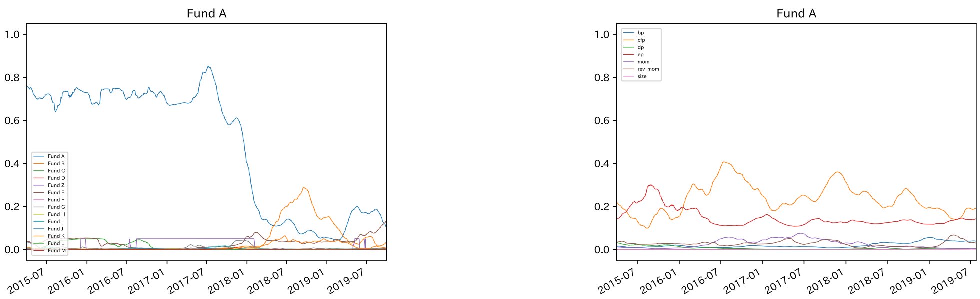
- Between July 2017 and January 2018, there was a slight uptrend in self-resemblance, followed by a large drop.
- Between January 2018 and March 2019, self-resemblance was stable at a low level, while another reference (resemblance to another fund) trended up.
- From March 2019 onward, self-resemblance began to slightly trend up again.

On the other hand, during the same time frames, the output of SDA (which detects investment style) did not exhibit any notable changes (Figure 8b). This indicates that while no changes were noticeable from the viewpoint of conventional investment style factors, Resembler did detect some kind of changes from the standpoint of the fund's uniqueness.

< Causual analysis of changes >

We attempted to analyze the cause behind changes in Resembler's output from several perspectives, as follows:

- Active weight:
  In Figure 9, each equity has been mapped into a SOM with a color gradation according to each equities' active weight, which visually highlights patterns in the evolution of portfolio composition.
  - Comparing the SOM for Jan. 2017 (Figure 9a) to the one for January 2019 (Figure 9b), the red units inside solid circles shrank, while the red units inside dotted circles expanded. This indicates that portions of the equities in which Fund A strongly overweight were



(a) Resembler

(b) SDA

Figure 8: Case study 1: Analysis of characteristics for fund A

rebalanced. At the same time, the total area of dark red on the SOM shrank, indicating that strong overweighting and increasingly concentrated in fewer stocks than before.

– In June 2019 (Figure 9c) the overweight areas circled in solid line expanded, indicating that Fund A was reverting slightly toward its previous level of self-resemblance. However, since strongly overweight clusters circled in dotted line remained, and the recovery of overweighting of whole map was limited, this suggests that the fund did not fully revert to its previous self-resemblance.

– Because the time frame of all these SOMs changes align with the changes in Resembler's result, Resembler's output may reflect the evolution of portfolio composition revealed by the SOMs.

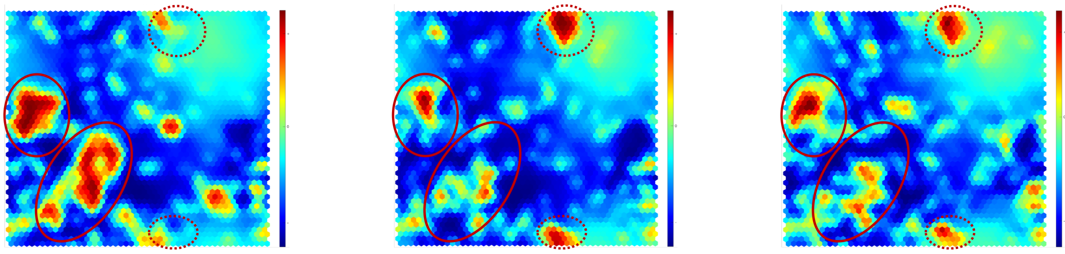
- Number of equities held
  From July 2017 to January 2018, January 2018 to March 2019, and from March 2019 onward, we can observe prominent fluctuations in the number of equities held by Fund A, which are generally synchronous with changes in Resembler's output, indicating the possibility that these changes in the number of equities held are also being noticed by Resembler (Figure 10).

- Portfolio turnover
  During the timeframes when Resembler noticed changes, turnover showed less frequent changes than in the past. Although the relationship is not as prominent as those with active weight and number of equities, the possibility that portfolio turnover is one cause of change in Resembler's output cannot be ruled out (Figure 11).

<Considerations for deployment at GPIF>

Fund A, on multiple occasions in 2016, gave GPIF advance notification of model revisions, and Resembler re-validated that changes took place from July 2017 to January 2018. In addition, Resembler detected a slight reversion to greater self-resemblance from March 2019 onward, which we assume to be reflection of the changes in the pattern of number of equities held and portfolio turnover during this period.

Concerning the latter changes, although GPIF consulted with the fund manager to understand them as isolated moves, it is possible that these changes should be scrutinized as part of a large-scale



(a) January 2017 (b) January 2019 (c) June 2019

Figure 9: Case study 1: Evolution of the pattern of portfolio composition (active weight)

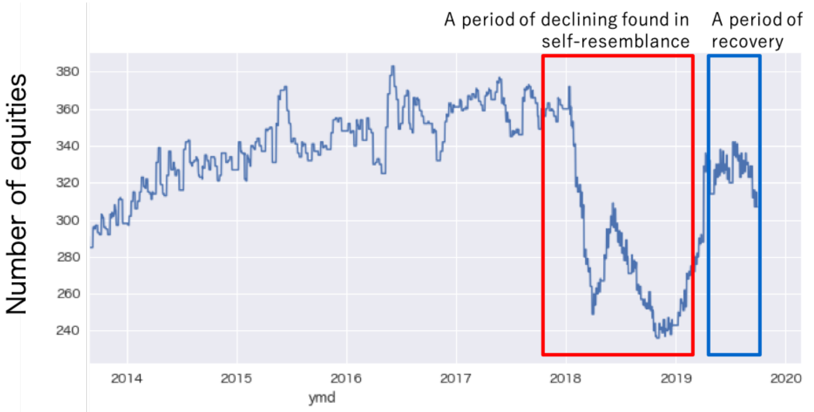


Figure 10: Case study 1: Evolution of number of equities held

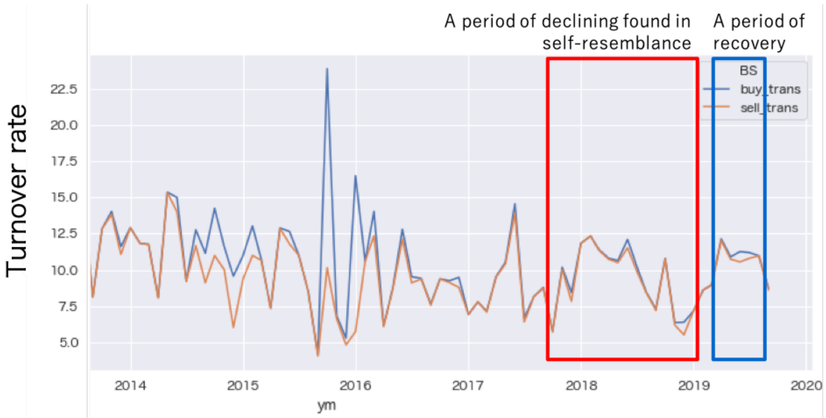


Figure 11: Case study 1: Evolution of turnover

qualitative shift in investment style in a bid to boost performance, rather than as a series of isolated moves in light of the performance issues of the fund in question,

It is possible that GPIF is not able to capture slight changes in investment style that may leads to stagnating performance because signals are subtle and often embedded in a wave of larger shift in economic conditions and other issues. Furthermore, daily communication between GPIF and funds is wide-ranging, with many points of discussion. Thus, even if GPIF dives into the details with the fund manager on each occasion and probe each issue's nature in isolation, it would be difficult to constantly spot such signal in advance without support of such a system. Resembler makes it possible to detect specific instances as part of sequences of large-scale shifts in order to establish new hypotheses. This capability could support discussions about how to improve a fund's performance by narrowing down the points at issue to the assessment of these hypotheses.

Furthermore, Resembler could contribute to operational improvements at GPIF through enabling quantitative logging of strategy changes by fund managers. When changes in fund behavior are detected solely through the "expert's intuition" or "sixth sense" of GPIF oversight personnel that tells them something is "off", no matter how strong that intuitive conviction might be, it nevertheless makes a weak justification for logging officially a change in fund behavior. However, if Resembler registers a change, this can be used as a quantitative basis for logging a change in fund behavior, and triggering the start of comprehensive monitoring. In other words, while declines in fund performance, key person risk, etc. have been used up to now to justify logging fund changes, Resembler allows new considerations to be added. By increasing the frequency and robustness of monitoring, this could deliver better fund selection and monitoring over the long term.

Case Study 2: Detecting change that is currently underway

Our next case study concerns Fund F. Resembler's output did not exhibit any change in over a long period of time in the past, but recently a change was observed. In this case study, we consider how the start of the change would be reflected in Resembler's output, assuming that new updated data are continuously provided on a monthly basis.

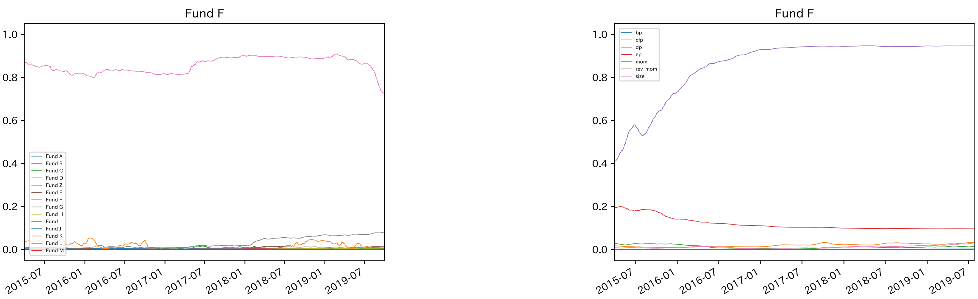
<Detection of changes>

During June-July 2019, the fund in question exhibited a change in Resembler that had never been observed previously (Figure 12a). In contrast, there was no comparable change detected by SDA over the same period. (Figure 12b). As well as the previously mentioned case study, although no style change was evident from the standpoint of conventional factors, Resembler did detect some kind of change from the standpoint of the fund's uniqueness.

<Causal analysis of the changes>

Next, we probed for the cause of this change from several angles.

- Active weight
  - In December 2018 (Figure 13b), in comparison to June 2018 (Figure 13a), the regions that are strongly overweight in the middle of the left and right sides (shown in dark red units inside solid circles) have contracted slightly, and a few white dots surrounded by dotted circle have appeared in the center. At this point, a certain amount of change is already being recognized.

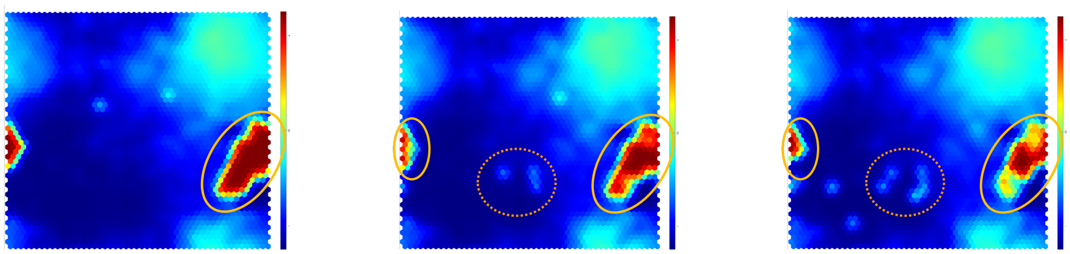


(a) Resembler (b) SDA

Figure 12: Case study 2: Analysis of characteristics for fund F

– On the other hand, in June 2019 (Figure 13c), the strongly overweight region circled with a solid line has noticeably shrunk. And we can clearly see the scattering of white dots in the center now, indicating that these equities in which the fund was strongly underweighting (dark blue) are shifting to neutral weighting (white). These tell us that there is a shift from the pattern of investment being pursued previously, and at the same time that the fund is slightly cutting back its active risk.

Since these changes visualized in SOMs all correspond to changes in Resembler, it is possible that Resembler was flagging these changes.



(a) June 2018 (b) December 2018 (c) June 2019

Figure 13: Case study 2: Evolution of the pattern of portfolio composition (active weight)

<Considerations for deployment at GPIF>

During the recent period when changes were detected by Resembler, Fund F was showing strong performance. Conventional wisdom would suggest a hypothesis that a fund which is delivering good performance would not have much motivation to adjust its investment policies or models. Yet Resembler might detect the emergence of an equity selection pattern not seen before, and a change in the degree of concentration in their investment policy.

GPIF has confirmed these changes by "separately" receiving notifications as "individual instances" occurring in "different timeframes independent from one another". In contrast, Resembler might detect the "onset of a sequence of systematic shift taking place from now on." For seamless integration of AI detection into GPIF's actual operating environment with respect to continuity, there are two issues to deal with: first, "Why pay attention to changes at a fund which is performing well?"; second, "How can the system configure output as actionable material in the GPIF operating environment?"

- "Why pay attention to changes at a fund which is performing well?"
  Because GPIF's operations are broad and multifaceted, and much of that work is handled by small pools of personnel with specialized expertise, we assume that the inevitable reality is that relatively worse-performing funds receive the most attention. On the other hand, a fund that is performing well, as in Case Study 2, will not necessarily continue to perform well indefinitely. If there are precursor signals of the fund's results taking a turn for the worse, and change responsible for that deterioration happens gradually, being able to detect such patterns in advance would unquestionably be valuable.

  The change that was detected in this case does not imply such changes are always bad. It merely means that there are changes in the style. Resembler and SDA, however, can be systematically vigilantly in assessing changes currently underway. If, at the point in time at which a certain threshold is crossed, a fund manager exhibiting precursor signals of change could be given priority monitoring by GPIF staff. Thus, it would allow the organization to expand their range of scrutiny in their operations without increasing their workload.

  Moreover, funds that have good performance are constantly making efforts to maintain those results; the changes detected by Resembler in this case may represent the fund's proactive moves in anticipation of possible future market conditions. By combining indicators from Resembler, fund performance, changes in investment behavior and market conditions, GPIF could build up an archive of quantitative insights into the moves made by the best-performing funds, raising the knowledge base throughout the organization, and enabling sharing of quantitative insights among relevant personnel. The better your knowledge, the better the discoveries you can make, and we accordingly recommend the development of applications that will further elevate GPIF's fund selection capabilities.

- "How can the system configure output as actionable material in the GPIF operating environment?"
  Using other methods to analyze the cause of the changes detected by Resembler yielded results suggesting the possibility of "the emergence of an equity selection pattern not previously exhibited" and "a point of departure from the initial prospectus with respect to the degree of concentration in their investment policy." However, a request was made by GPIF that there should be "a human readable text materials that breaks down to an even simpler level how to explain the nature of the change (assuming the change is not routine factor rotation)".

  The backdrop to this feedback is that although both the GPIF and the fund manager sides are in close contact on equity selection and portfolio rebalancing plans on a daily basis, they may not be able to detect subtle changes in the context of dynamically changing environment. Resembler may systematically detect such changes and signal them to GPIF. We consider this is a great example of complementation between the understanding of GPIF personnel and the AI. In order to reinforce such a complementation, the system needs to configure its output as actionable information as described in the feedback from GPIF. We will investigate how AI should be integrated with the appropriate processes or modules of IT systems currently used, in order to obtain materials based on which GPIF personnel can make actual decisions.

Regarding SOMs, we consider it can be used to visualize the cause of change in fund behavior more intuitively understandable. Substantial training costs, however, would be incurred to enable staff without preexisting expertise to make full use of this evaluation method and interpretation in operations. Improvements of the user interface is desired to make the output of causal analysis more visually and intuitively accessible.

3.2 Case studies for foreign equities funds

With foreign equities funds, we selected two case studies that reflect a different perspective than the Japanese equities case studies; "assessing the validity of the contents of reports received from fund managers" and "identifying a lack of self-resemblance as uniqueness".

Case Study 3: Assessing the validity of the contents of reports received from fund managers

Case Study 3 is about Fund N. In this case, we validated that the content of the fund manager's past reports was valid.

Fund N experienced a fairly important corporate event in its organization in 2018, and GPIF received a report giving advance notification of this change. In addition, GPIF received a report to the effect that the fund would be continuing its previous investment policies unchanged, and GPIF had used existing tools to confirm through factor analysis, etc., whether any change actually took place. We also re-validated this in Resembler, and confirm that its output showed almost no changes, meaning that policies remained basically the same as before (Figure 14).

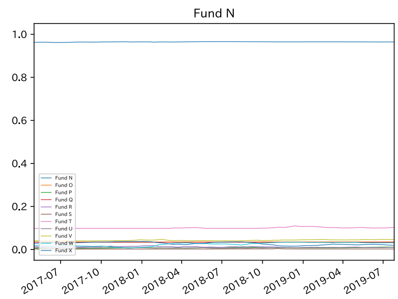


Figure 14: Case study 3: Analysis of characteristics for fund N (Output of Resembler)

We also performed follow-up assessment using several supplementary analytical methods, which confirmed that Fund N's investment behavior had not changed to any extent worthy of special mention.

- Number of equities
  Although a slight change in the number of equities held occurred in 2018, when the event happened, this was not flagged as a serious change because scale of that change was same in the past (Figure 15).

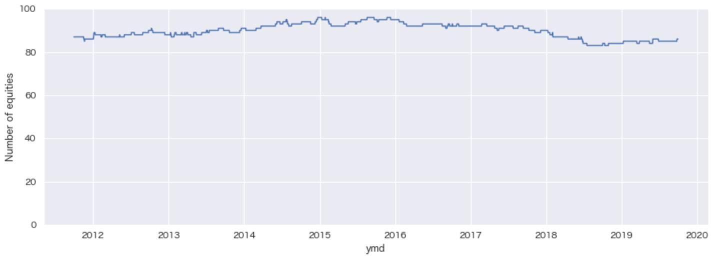


Figure 15: Case study 3: Changes in number of equities

- In and Out of equities
  We confirm that the numbers of INs and OUTs of equites changed slightly, but not substantively, around 2018, when the important corporate event occurred. (Figure 16)

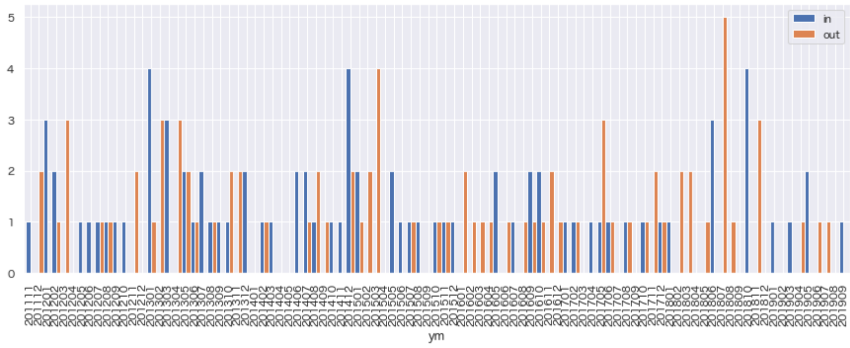


Figure 16: Case study 3: In and Out of equities

- Change in equities making up the portfolio
  Between the portfolio of any given month and the portfolio of six months previous to that, we calculated the correlation coefficient of the holding ratio and the concordance rate of the equities held (Figure 17). Correlation of holding ratios and concordance ratio maintained a steady high value indicating that investment style was continued after the event.

Fund management firms have an obligation to make reports of material changes to GPIF, and GPIF validates the contents of those reports to the extent possible with existing tools. Resembler is able to detect changes that existing tools cannot fully characterize, at a more high-level, holistic level than traditional indicators such as investment styles and risk factors. It might allow a more sensitive reassessment from a different standpoint of whether fund managers have changed their behaviors.

Case study 4: Uniqueness defined by absence of self-resemblance

Case Study 4 is about Fund W, which does not stick to a certain single investment strategy, but continuously switches among various strategies. Resembler detects large decline in self-resemblance,

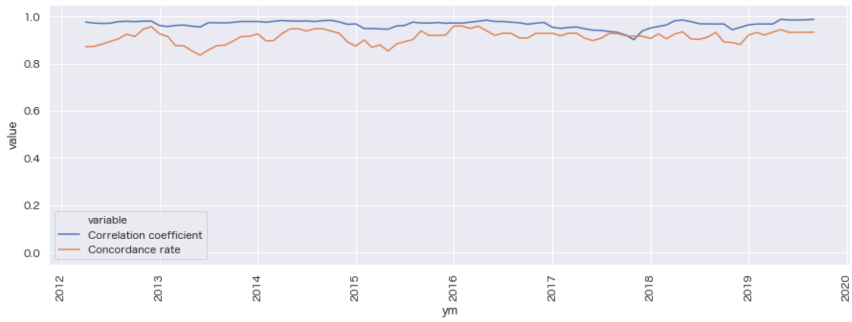


Figure 17: Case study 3: Changing in equities making up the portfolio

as well as a certain degree of rises and falls of resemblance to various other funds during the time. This lack of self-resemblance is, in a sense, what constitutes the uniqueness of this fund, and is being consistently detected by Resembler (Figure 18).

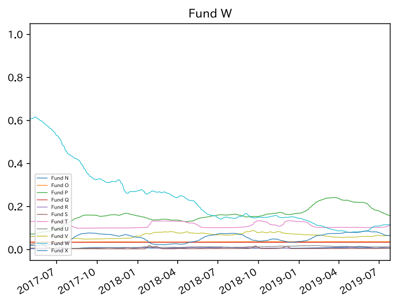


Figure 18: Case study 4: Analysis of characteristics for fund W (Output of Resembler)

Regarding those like fund W which from the outset take an approach of opportunistically switching between different investment strategies, if those changes are consciously and proactively made, can be considered an investment philosophy unto itself. On the other hand, if such a fund become locked into one style over a long period, or if their strategy switching starts to lag rather than lead changes in market conditions, it could be considered a precursor signal that an in-depth consultation should be held.

3.3 Interim summary of experimental implementations

Through the experimental implementation of AI in actual operations at GPIF, we became aware that the following three-step process is needed:

1. Using AI system (Resembler/SDA) for initial detection of changes;
   The first step for AI (Resembler/SDA) is to detect whether any change has taken place or not. The type of change that can detected is not limited to those which relate to factor

exposure and style, but also qualitative aspects such as "uniqueness" of the fund behavior that is difficult for existing tools to capture.

2. Performing causal analysis on detected changes;
The next step is to analyze the cause of the change. In case studies described above, other data science methods including SOM were used, but these require labor-intensive ad-hoc work. What is needed is a systematic framework for a seamlessly integrated processing flow and system that takes over after change detection occurs. a follow-on investigation into ideas like topic analysis, various clustering methods, Variable auto-encoders (VAE), and other methods are currently underway.

3. Delivering output in a form that is immediately actionable in the operational environment;
As mentioned in Case Study 2, we received feedbacks that GPIF would want output when change is detected that clearly presents the salient points of the change ("what did they do and why is it concerning") as human-readable text. This will help GPIF's personnel to make conversation more fruitful with the fund manager especially who has not been aware of the changes. Moreover, if the outputs are given in easily accessible format based on which concrete decisions can be made, this would reduce workload and enable the situations to be handled in a way that is more likely to have results than with current systems. Through ongoing experimental implementations, we will work on perfecting the integration of our AI system for data sharing with existing operational and IT systems.

4 Exploratory Studies on Advanced Analysis

4.1 Assessment of portfolio diversification using Resembler

GPIF allocates its overall manager structure among a diversified set of managers to optimize the robustness to market changes. An important aspect of GPIF's operations is to monitor diversification, detect when the behavior of fund managers exhibits convergence in response to the economy, and evaluate whether this convergence can be mitigated to maintain diversification.

Generally, diversification of the manager structure has been assessed on the basis of fund performance and factor-based analysis. We introduce an experimental application within the framework of Resembler, whose output is the uniqueness of fund managers, to assess diversification of manager structure at GPIF. Mutual-resemblance among funds is calculated as discussed in section 2.1.2, then funds are visualized together within the same phase space by dimensionality reduction method called multi-dimensional scaling (MDS) [12].

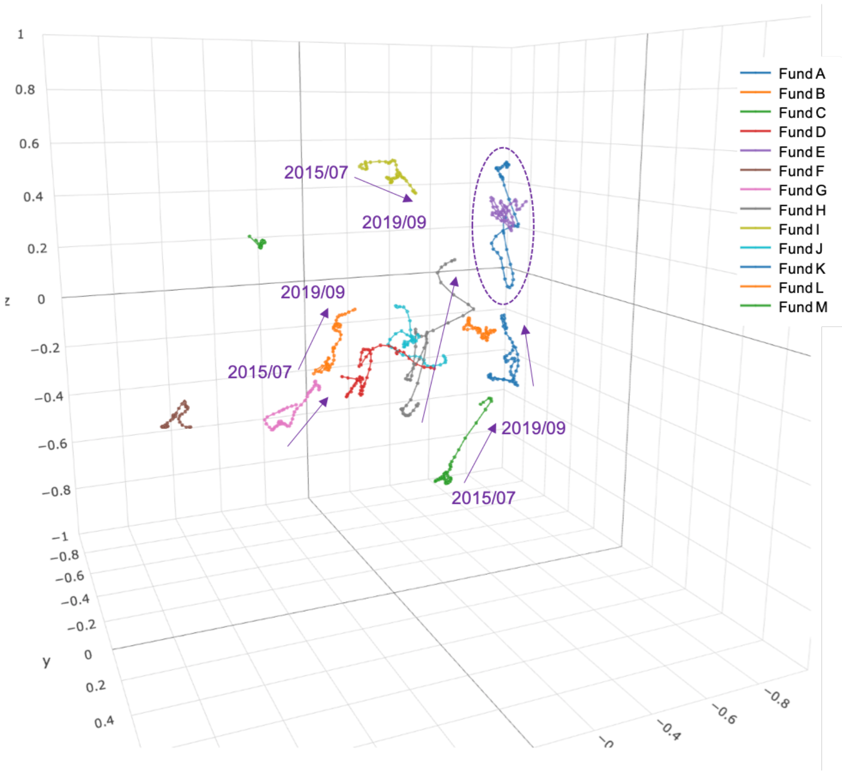


Figure 19: Changes in mutual-resemblance for Japanese funds over time

It is required to select several "anchor" funds to serve as references to obtain a map of the mutual-resemblance. We chose four funds as anchors which are markedly different from one another in terms of investment policy, performance, etc. to assess the resemblance among all funds comprehensively. Figure 19 shows the resemblance among domestic active funds over the period from July 2015 to September 2019. Trajectories over time from start to end (in the direction indicated by purple arrows on the graph) indicate that at the outset they are dispersed through the phase space, representing the uniqueness of each fund manager, over time they gradually converge toward Fund A and Fund E at the upper right of the graph. There are several conceivable causes for this trend toward convergence; as one example, Figure 20 graphs estimated tracking error, a measure of the portfolio risk that fund managers have taken on that is calculated from the variance-covariance matrix of daily returns for each equity on past twelve month.

The estimated tracking error for most of the funds squeezes into the same narrow range in early 2019. This indicates that the active risk of each fund's portfolio has similarity. It is considered that market condition in 2019, which performance was positive with low risk diversified investment, induced an overall shift toward similar strategy. Unlike relying on assessment of specific factor exposure, this is higher level diversification assessment with more expansive characteristic of investment management, i.e. its uniqueness requires further assessment what kind of risk can be occurred in case of the convergence. GPIF that allocates among a set of many different fund managers is the entity which can overview this kind of trend, otherwise each fund manager does not. It follows that, even if each manager behaves to maximize their own performance, their investment behavior could increase the risk exposure of the entire structure of GPIF with respect to some particular factor. Even though it is not necessarily to hedge the convergence of mutual resemblance, and it can be allowed or even encouraged in some scenarios, Resembler has potentiality to apply to capture high level and entire information to optimize portfolio at GPIF.

As for another application, Resembler-based diversification assessment can be applied to selection process of a new fund manager; to evaluate similarity to the currently outsourced managers and to avoid of convergence at manager structure. It could also contribute to resolve concern "arbitrariness of qualitative evaluation" with the quantitative evaluation of qualitative information by resemblance and the causal analysis at the process of similarity assessment to the currently outsourced managers. It would become possible that GPIF with limited resources fairly and efficiently evaluates a broad range of qualitative information from numerous candidates at the selection process by a system which streamlines from AI to causal analysis to use the data provided from candidates to evaluate qualitative information.

Further studies are needed on how new candidate provides their data under constraint due to the importance and confidentiality and whether AI could output properly with monthly or quarterly based data instead of daily based one. the limited number of data such as monthly or quarterly. Accordingly, it is also to be experimented to achieve an acceptable quality by data science methodology even with such coarse data.

4.2 Evaluating "edge" in active fund management

GPIF allocates approximately 20% of its total assets to active funds. However, over the three year period from 2014 to 2016, only a few of these active funds reached their targets for excess returns. GPIF has identified the primary reason for this as a problem with fund selection capability at GPIF, along with other possibilities such as fund managers not setting their target for excess returns appropriately, or being more focused on increasing the amount of funds under management than on capacity management. GPIF has taken various actions to address this, including revising its previous policies of fixed management fees or moderate performance fee in favor of performance base fee [13].

In this context, we set out to experimentally evaluate the "edge" of active fund managers: whether an active fund's performance comes from skill as opposed to luck. This is expressed as a "batting average" comparing how much of the return that they are achieving on equity investment is due to the manager's intentional active weight (skill), to how much is derived from unintentional

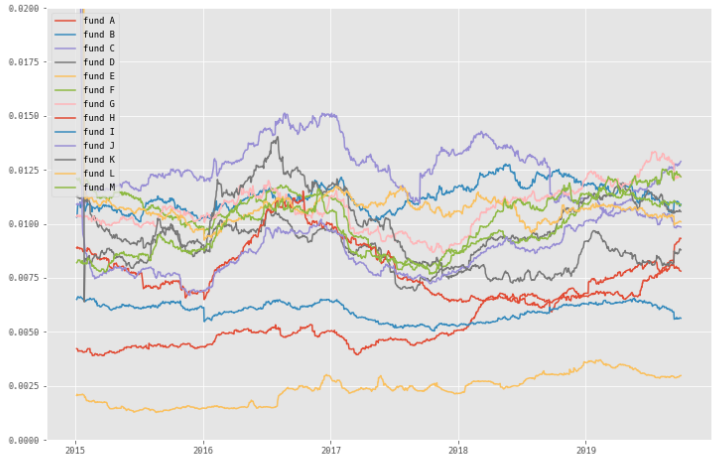


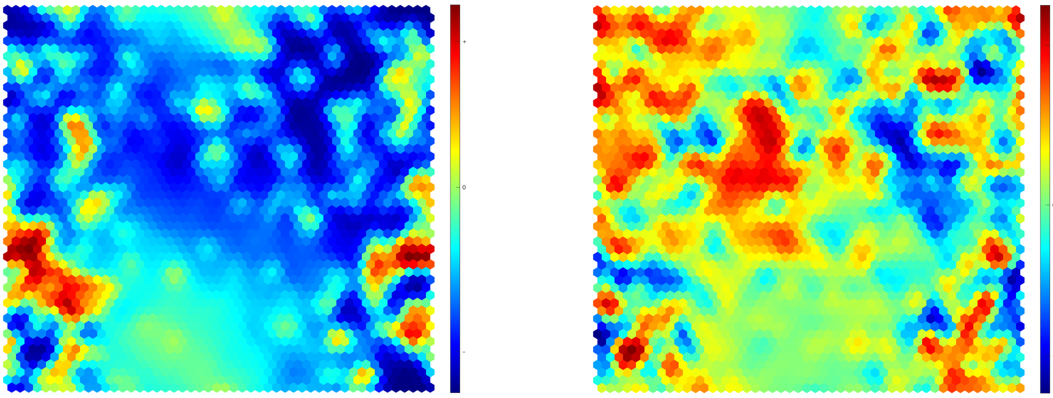
Figure 20: Changes in tracking errors

active weight (luck). Figure 21a shows SOM of the active weight by actively manager of domestic equity fund. As discussed in section 2.2 above, this method takes all the equities composing TOPIX and clusters them on a 2D map according to a rule that results in equities with the most resemblance to each other being closest together, without meanings both on vertical and horizontal axes.

In this case equities in which the fund in question is overweight are colored in red, those in which it is underweight are colored in blue, and the saturation of the color indicates the degree of over- or under-weightiness. In Figure 21b the active returns for equities held by the same fund are shown; as in Figure 21a, positive returns are shown in red and negative returns in blue. In other words, we can juxtapose the two SOMs in order to assess whether equities with high active weight also have high active return. For this fund, Figure 21a shows that high-active weight equities are clustered midway up the right and left edges. When we juxtapose Figure 21b, in those same areas we can see both positive and negative returns. On the other hand, while the central area of Figure 21a shows that the fund is underweight in those equities, in Figure 21b we see that returns on those equities are fairly good. Note that while this pair of SOMs only show the results for a predefined time period, a time series of SOMs can be generated to analyze trends over time.

As for next step, we defined categories of active weight as described below to classify the attributability of the respective returns (Figure 22):

- Intentional active weight:
  Within the distribution of active weight, this domain – defined by having the large absolute values of active weight – is where manager applied the strategy to make an investment decision that these are promising stocks to hold. The intentional active weight domain is divided into two subdomains: where the values are positive and relatively large number is intentional overweight and where the values are negative and relatively large number is intentional underweight.



(a) Active weight (b) Active return

Figure 21: Visualization of active weight and active return with SOM

- Non-intentional active weight:
  As opposed to the intentional active weight noted above, non-intentional active weight domain is divided into two subdomains: where the values are positive and relatively small number is non-intentional overweight and where the values are negative and relatively small number is non-intentional underweight. This type of position is required by portfolio risk management, thus does not always reflect managers' market perspective. Hence performance yielded by non-intentional active weight has no difference between overweight and underweight.

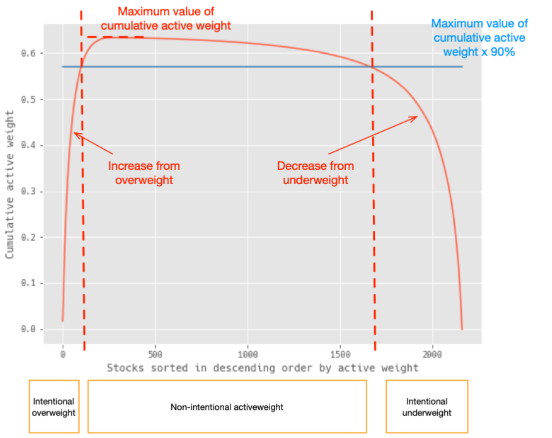


Figure 22: Separation of intentional and non-intensional positions

The procedure we used to calculate the degree of attributability to the intentionality of the fund manager is as follows:

- Calculate the active weights of all the equities, and arrange them in order from largest to smallest.

- Cumulative active weight starts from zero and increases as the active weight of each equity is added. It reaches a peak value, where underweights start to be added and cumulative active weight starts declining, ultimately reaching zero.
- A level of 90% of the peak value of the cumulative active weight distribution is defined as the threshold for attributing an equity holding to intentionality. We anticipate that this threshold would be adjusted as necessary depending on the analysis scenario, taking into account, e.g., investment approach (highly active fund, quant, etc.) and the active risk level.

The result of this is Figure 23, where we see on the left side of the graph the "intentional overweight" domain as defined above, and on the bottom right of the graph the "intentional underweight" domain, while the rest of the graph (circled with a dotted blue line) is the "non-intentional active weight" domain.

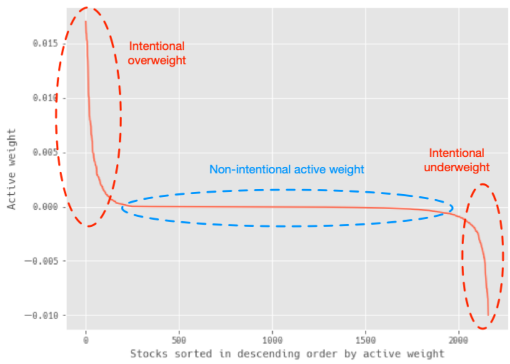


Figure 23: Order of equities by active weight

Based on this methodology, we calculated performance for each of the domains across the entire portfolio, to obtain the daily return against each day's active weight, to create a metric for skill shown in Figure 24.

Looking at the results of the above calculation for the "intentional active weight" and "non-intentional active weight" domains, intentional active weight is neutral, while non-intentional active weight exhibits slightly negative returns, with overall performance being neutral. In other words, this fund failed to achieve positive returns in equities where it aimed to generate them, and experienced slight negative returns in equities it held without having such aims, which can be interpreted as grading the fund's skill as neutral to slightly negative.

The goal of this experiment was as follows:

- When managers report to GPIF on their investment activities, a psychological incentive comes into play: if performance has been good they attribute it to their skill, but when performance has been poor, they blame bad luck due to unfavorable market conditions or other economic factors outside their control. While such explanations may seem to constitute a plausible story-telling from a short-term perspective, it is mandatory for GPIF to verify or falsify these claims scientifically and sustainable way from a long-term investment perspective. Thus, using

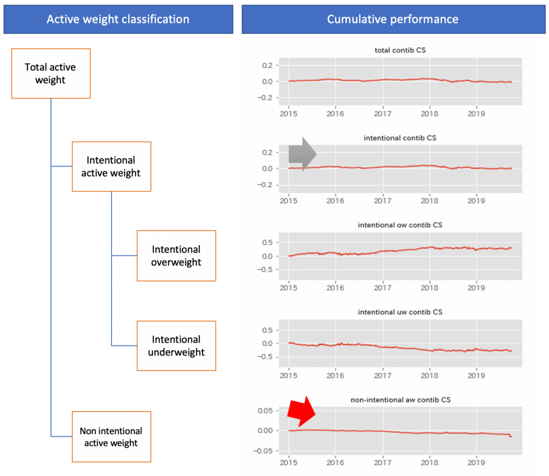


Figure 24: Decomposition of cumulative active return

a methodology that makes it possible to discriminate between returns to intentional/non-intentional active weight over the long-term based on consistent rules would provide a common ground on which GPIF can conduct reviews of fund performance of funds, and supply fund managers with reference materials for further improvement.

- Recognizing that there are cases where it is difficult to judge the validity of fund managers' own text-based reports about how the attribution of their active returns, SOM-based visualization and analysis may enable GPIF to share and discuss performance with a higher level of awareness of the causes behind it.

In conjunction with this, we received feedback in response to last year's research summary report that it would be "desirable research has correlation to performance". Evaluation and selection of managers based on a combination of a fund's "edge", assessing performance quality, with Resembler and SDA are promising basis that provide new and different angle than existing practice. With this research initiative, we intend to go deeper to develop a comprehensive methodology for GPIF and its fund managers to evaluate returns. Impacts of this initial outcome is far reaching because such approach shall uncover true skill of fund managers as opposed to luck, and it may dramatically change the way active funds are managed.

5 How AI could enable GPIF to unlock value

Insight obtained from the trial results and experimental implementation, applies to the operation level and the management level: how AI could transform GPIF's operations and how GPIF should deploy AI.

5.1 Transformation of operations at GPIF

Feedback from GPIF:

During experimental implementations and trial with close, deep and frequent discussion about the requirements for business process at GPIF, consensus has been developed what GPIF expects to AI and how to apply it. There are feedbacks from persons in charge of daily manager monitoring operation at GPIF regarding benefit and expectation once AI applied.

- Monitoring broader range of information
  Contribution to better monitoring is expected by detecting changes in organizational information that are not covered in reporting guidelines of GPIF.
- Efficient managers evaluation
  Resembler takes over a part of the evaluation report, which currently requires a large amount of manpower.
- Prediction-based monitoring
  Deploying Resembler to predict how managers will behave in response to market environment could prepare for communication with managers and be used for stress-testing portfolio at GPIF.
- Automation of selecting new manager
  If the same type and quantity of data could be obtained from funds that are prospective additions to GPIF's manager structure as from those that are already in it, then monitoring of prospective funds could begin in the pre-contract stage, which would eliminate the information gap in evaluating new versus existing funds. This would allow GPIF to gauge the uniqueness of funds which are being contemplated for addition, in order to avoid those which exhibit a high degree of resemblance with current funds whose performance has been unsatisfactory. Also, it could help identify funds with an edge that is not in current manager structure at GPIF.
- Manager Replication
  Once a large amount of data from Resembler has been accumulated, it may be possible to replicate idealized "cloned managers." These model funds could be used as benchmarks for grading existing fund managers and selecting new ones.

Boosting GPIF's front-line operational capabilities with AI

In our FY2017 research, target to deploy AI is to support GPIF's manager structure by providing more sensitive tools for detecting when the managers' behavior deviates from its prospectus and whether the change is appropriate or should trigger an alert, if it is changed from original policy. Our research from October 2018 onward has set out to find solutions for and more rigorously define the issues raised at GPIF. GPIF had up to that point emphasized qualitative evaluation of managers' policies, operating procedures, and organization/human resource capability, however in 2017 the Board of Governors at GPIF expressed concern that qualitative evaluation of managers could invite criticism for arbitrariness and subjectivity, while persons in charge also noted that the process is heavily dependent on a staff with specialized expertise. Based on these organizational concerns and the results of experimental implementations of AI, GPIF expects the following transformations of operations:

- Improvement of constructive and efficient discussion with managers.
  - Detection of any changes in investment behavior contained in reports from managers.
  - Confirming how much improvement is conducted by managers which have been put on notice to improve performance.
  - Detection of unreported changes (whether or not those changes should have been reported according to guidelines).
  - Direct capture of investment behavior of foreign managers through data, in addition to information from gatekeeper.
  - Elimination of reporting costs by funds and reduction call time.
- Standardization and commoditization of skills by exploiting big data at GPIF.
  - The output of AI running on GPIF's in-house data will enable novel perspectives and unique hypotheses to be built.
  - Integration spanning from AI to day-to-day operational systems will allow a reliable quality level that is not dependent on a small pool of personnel with years of experience or specific career backgrounds.

5.2 How "Whale" attracts the industry

Longer tail —Cybernetic Whale will have a longer tail

If this research is fully deployed at front-line operations in the future, then to the extent that it is possible to obtain the same level of data from pre-contract candidates as from existing fund managers, there will be no longer a distinction or information gap between them as far as GPIF is concerned. In that scenario, the long-term relationships between asset managers who are already part of the manager structure and GPIF may lose significance. Up to now asset managers, once under contract, have been able to maintain a privileged position by assigning personnel to manage the relationship with GPIF and supply various kinds information that creates a gap with competitors. Deploying AI would diminish the value of these efforts and potentially lead to a revolution in the business models of asset management firms. For new funds who want mandate from GPIF, the barrier would be lowered for them to pitch themselves to GPIF.

Because GPIF's reporting requirements are so onerous, smaller asset management firms which do not have sufficient resources have been blocked from doing business with GPIF. But if the output of AI system was able to take over and eliminate much of this reporting obligation and workload, these managers would be able to pitch themselves based on performance and practices. And whole investment management industry itself could benefit from being able to redirect resources from reporting to core investment performance.

This opens up a new opportunity for both GPIF and smaller funds, especially this gives GPIF access to the “long tail” of fund scale in the investment management industry. Due to statistical feature of complex systems including a global stock market, how to access and benefits from long-tail part of the players and assets is critically important. The use of AI that enable GPIF and other major asset owners to access long-tail part of the players could be a major breakthrough and may trigger qualitative transformation in the industry.

“Moneyball” for GPIF: Toward GPIF-Metrics

Asset owners, including GPIF, can have outside vendors and consultancies provide with variety of market and product information, benchmark analytics, and so on, in addition to the performance and trading data of the funds they invest into. This means that asset owners can be in a position of having bigger data than asset managers. Once AI becomes more heavily used in coming years, it should be realized that asset owners have the information advantage over asset managers, since they have access to the bigger data to develop the more insightful analytical results.

This represents a paradigm shift away from the situation up to now, where the professional and specialized expertise of asset managers gave them an information dominance over asset owners. In the future, the use of statistical metrics, including AI, will untangle the data that spans the investment world and form “manager teams” which do not depend on traditional and established value. It will be interesting to see if this quest to unravel the mysteries of GPIF-metrics may impact young generation and investment professionals.

6 Future work

While the initial stage of our research started in October 2018 was basic and exploratory investigation, current one has been shifted to an application phase as pilot test. The application phase has cleared numerous suggestions and issues to be resolved from actual business process, and this phase is moving forward in parallel with the following agenda: "enriching report generating function of the system", "deepening, improving, and operationalizing based upon trial result" and "streamline and integration of the technology platform".

Enriching report generating function of the system

As of October 2019, the system operates in the way Sony CSL generated reports based on output from Resembler on Japanese equity manager every month and SDA a few times with monthly detection of changes and causal analysis of these changes. The report was provided to GPIF for their internal discussion with managers. The next step is Sony CSL to provide monthly report using both outputs from Resembler and SDA on Japanese and global equity space.

Deepening, improving, and operationalizing the system

- Improving and deepening causal analysis output protocols with AI detection:
  In case Resembler and SDA have detected change, causal analysis of the change is required to be explainable. One of the analysis methods produces output through SOMs, clustering, and other means. However, a certain level of specialized expertise is required to evaluate and interpret this type of output, which would incur training costs to roll out to GPIF's operations. The user interface has to be improved to make the output of causal analyses more visually and intuitively accessible. SOM was applied to analyze changes of holding data so far. Sony CSL will conduct research exploring whether this method can be extended to identify background caused these changes (e.g., change in growth forecast).

- Operationalizing and streamlining final output for actionability:
  After AI detected change of managers' bahavior, Sony CSL conducted causal analysis of the changes and suggested how to communicate with the managers. GPIF provided feedback that a clearer presentation of the salient points is needed in order to succeed in bringing about "more constructive and efficient discussions with fund managers". For example, it is assumed that actionable context is at the level of specificity of, e.g., "we determined that investment of equity 'A' last month appears out of character with the investment behavior you have done ever," or "you have retained your position in equity 'B' even though it has been in quite negative performance, which is not consistent with your past pattern". Multiple different format of reporting can be assumed corresponding to each case. Priorities and categorization will be made to implement such automated report generation functions.

- Updating the system for better handling over dynamic change of universe:
  The current generation of SDA and Resembler are composed of dedicated input units in the neural network for each equity, which means that the system cannot take into account any newly listed equities. In the event of an IPO of a large market cap equity that will play an important part in the market, the output of SDA, especially, could veer far from the market[5].

[5]Since Resembler is based on comparisons among funds' characteristics with past of their own, the influence of

In order to solve this problem, the data structure of the input to thße neural network should be re-engineered.

- Further improvements of SDA:
  Although Resembler provides output on the characteristics of funds from a different viewpoint of conventional analytical tools, in considering the validity of the output, it is essential to check the validity of the functioning of SDA, which shares the same core system with Resembler. Accordingly, although the operating principle has already been confirmed, there is room for improvement in the virtual data for VFMs to train, we are continuously upgrading that aspect with reference to feedback from experimental implementation of the system. In addition, we will work to polish SDA into a tool that provides a novel analytical angle in its own right.

- Granularity of data for new fund selection:
  Resembler, which can measure self-resemblance and resemblance to other funds, is under consideration as one piece of information to use in fund selection. One issue is different granularity of data between currently contracted funds and new candidate funds. Current Resembler outputs base on daily data from contracted funds. However, data from a candidate fund could be less granular, e.g., only monthly or quarterly, due to the proprietary value and confidentiality. We will investigate whether Resembler produces appropriate output, even when the system is fed much less granular data, and whether the use of data science methods could enable the results to still achieve an acceptable level of quality.

Streamline and integration of the technology platform

Since FY2017, Sony CSL has carried out research and development of the system and their modules with numbers of training and verification processes. Each module of the system is not fully integrated. and some still at the stage of proof of concept. However, it must be integrated seamlessly in order to configure the final output to be operational in GPIF ' s daily use with proper output formats, as noted above in "deepening, improving, and operationalizing based upon trial result." For the full scale deployment of AI and causal analysis programs, data cleansing and database architecture are to be organized and integrated as a part of the daily operation process.

ignoring newly listed equities would be limited to some extent, until the number of such new comers is accumulated to a certain level.

7 Implications

Combined power of a major asset owner, AI expert, and massive computing resource

The process of R&D of such system requires large numbers of trials and verification on how well the idea work against large-scale real-world data. This is where AIST's AI Bridging Cloud Infrastructure (ABCI), one of the world's most powerful computing facilities, came along to make a huge contribution. With the trio of AIST with their ABCI, GPIF (the world's largest public pension fund) and Sony CSL (with deep expertise and a proven track record in AI), Japan-based organizations have combined forces to not just carry out basic research, but carry that forward to a ground-breaking experimental implementation of AI for detecting the investment behavior of active manager. This demonstrates managing major funds may require cross-boundary highly competent team and resources at the highest-level.

Formation of a "Global Data Consortium" of asset owners

This research was commissioned by GPIF, and the findings must first and foremost meet the needs of GPIF. However, evaluation of active funds is a need well shared by all asset owners. One possible collaborative project that maximizes the benefit of this study is to apply the method of assessing diversification/convergence of funds using Resembler, as covered in section 4.1, to manager structure of multiple asset owners.

Projections show that in 2020 global active investment will be USD74 trillion, reaching 87.6 trillion in 2025 [14]. If global asset owners with a certain scale joined together in a data consortium, Resembler could run on that database to determine convergence among the active managers, which would capture not only the manager structure risk confined to each asset owner, but a global asset owner-level view of the investment behavior of active funds at all market environments. Within the manager structure of a single asset owner, investment behavior may be appeared as diversified, yet there may be convergence when viewed through the global data consortium. This would detect any rise in localized risk among the world's active funds that could ultimately develop into a risk affecting all asset owners.

There are some practices utilizing inter-aggregated data at the exchange level. For example, NASDAQ has adopted deep learning systems to enhance its capability to detect trades that may be attempts at market manipulation. The system "works in tandem with human analysts" and "augments surveillance system that uses statistics and rules to flag any signs of market abuse." [15]

There is a question for extremely large asset owners that need to be answered: "Whales breach wildly because earthquake happened, or whales breaching wildly causes the earthquake to be happened?" Public institutions which share the mission of investing over the very long term have a keen interest in the mechanism behind this problem and what the precursor signals might be. And tackling a problem on this scale requires cooperation among asset owners. GPIF, AIST and Sony CSL, a trio of three highly unique Japan-based institutions, propose that asset owners with a shared interest in this issue form a global data consortium to further develop a robust solution to this market challenge. We expect a series of discoveries can be made that should uncover reality of investment practices for large-scale asset owners, and scientific understanding should be possible. Such efforts shall open a new era of science of investment and asset management.

Areas where academic insights could contribute to application

Outcome of this research uncovered interesting opportunities to scientifically understand behaviors of market, fund manager's behaviors, and impacts of large scale funds in the makret. For example, experimental implementation of the Resembler system to detect changes in the investment behavior of fund managers has been based on the idea that a nonlinear model could provide more encompassing and responsive detection than interpretation of conventional linear factor-based models. We believe there is a need for a deeper discussion of on detection of investment behavior in linear versus nonlinear models. Given the amount of data from the asset owners, computing resources available today, and computational approaches including deep learning that captures high dimensional non-linear space, we may be entering the new era of computational asset management.

Will AI steal jobs, or create jobs?

Although there are still many hurdles to clear before this research (now at the experimental implementation level) reaches full operational deployment, the roadmap to achieving full-scale deployment of AI systems becomes clearer. Once these systems are in place at GPIF in the future, it will transform the way GPIF maintains manager structure and monitor their overall portfolio. It will also results in a major reduction in workload as part of the process of selecting and monitoring active managers is handled by AI. It is also possible that this would also eliminate some of the reporting obligations that fund managers have towards GPIF. In that case, what will happen to all those man-hours that people no longer need to perform?

According to the provocative thesis that "AI will steal jobs", reduced workload will lead to staff cutbacks, because employers will take it as a way to cut labor costs. On the other hand, consider what happened in the banking industry in the past: once, a huge amount of staff and labor were needed to calculate manually, until the development and deployment of accounting system allowed that function to be executed with systematically greater speed and accuracy. The human labor saved by this was used to provide better customer service, develop new financial products, and manage the increasingly complex risks.

Humans and AI have each own excellence, and in cases where AI becomes able to take over work that had previously been done by people, GPIF and investment management firms will have to carefully study how to redirect human resources into areas of human strength, and what those areas are.

It is critically important to recognize that the use of AI shall have higher impact when it directed to trigger qualitative transformation of the organization, mode of operations, and quality of the work by assisting us with its capability to see what human cannot see. Reduction of workload takes place along with the organizational change, but mostly directly to enable the organization to refocus their effort to the issues that was not possible before. We consider there is unprecedented opportunity in the proper introduction of AI technologies to asset management industry for quality management and transparency which shall benefits the society as a whole.

Acknowledgement

We received a great deal of feedback in response to the report we issued about our FY2017 research. As this is an interim report and many findings are provisional, we hope to once again enjoy a healthy

amount of feedback. And we will take into account that feedback to help build out our research into pioneering, highly unique, standard-setting AI systems. At the same time, by continuing to deepen dialogue with all parties involved, we intend to elevate the standard of operations on all sides.

We extend our sincere thanks to all the people we worked with at GPIF who took time out of their extremely busy day-to-day responsibilities, to the part time researchers and assistant researchers at Sony CSL who constantly managed to complete difficult research assignments within tight timeframes, and to everyone else who supported the efforts that resulted in this paper.

A Appendix: Full results of all funds

A.1 Outputs of Resembler for Japanese equity funds

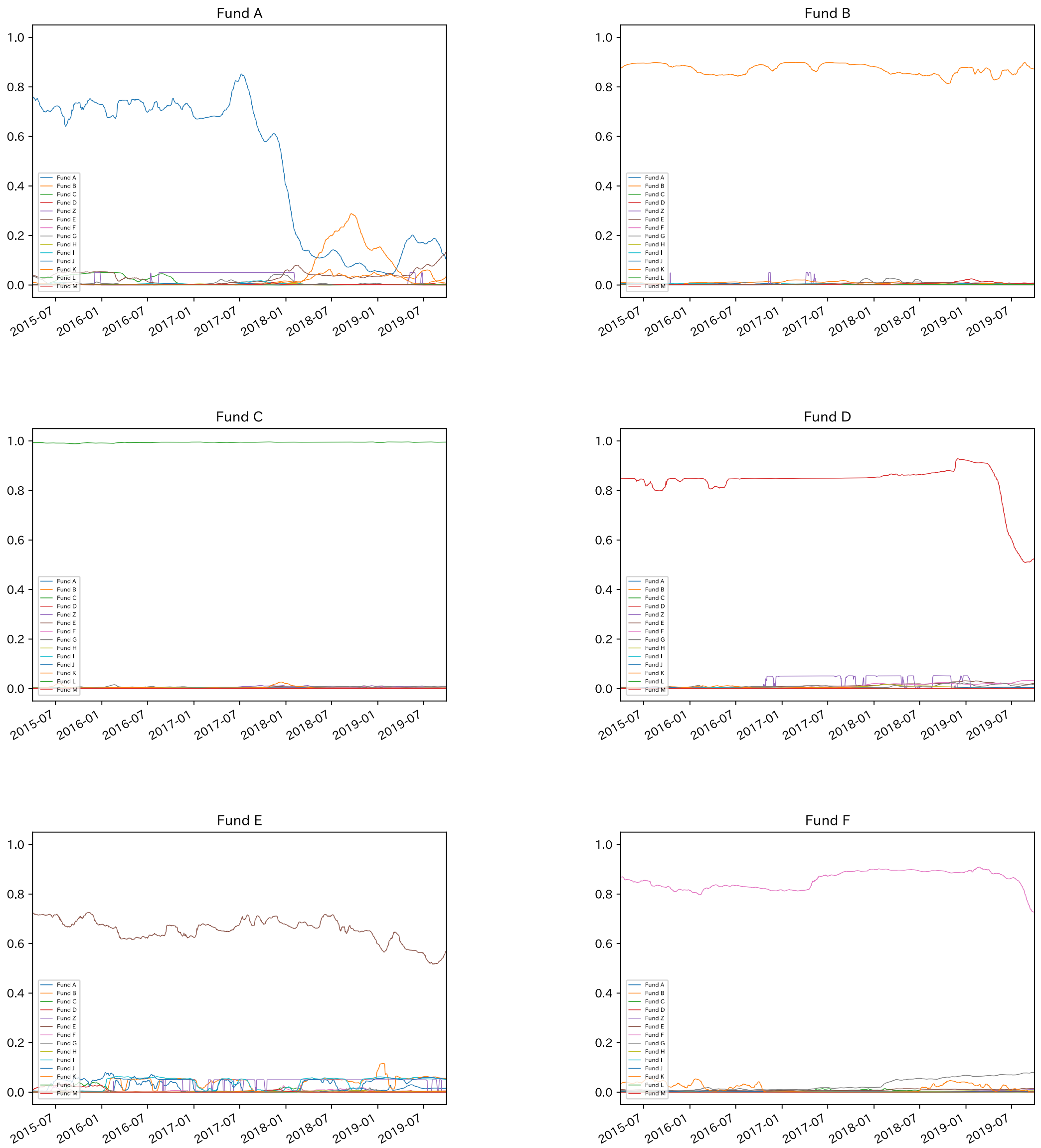


Figure 25: Outputs of Resembler — Japanese equity funds (1/3)

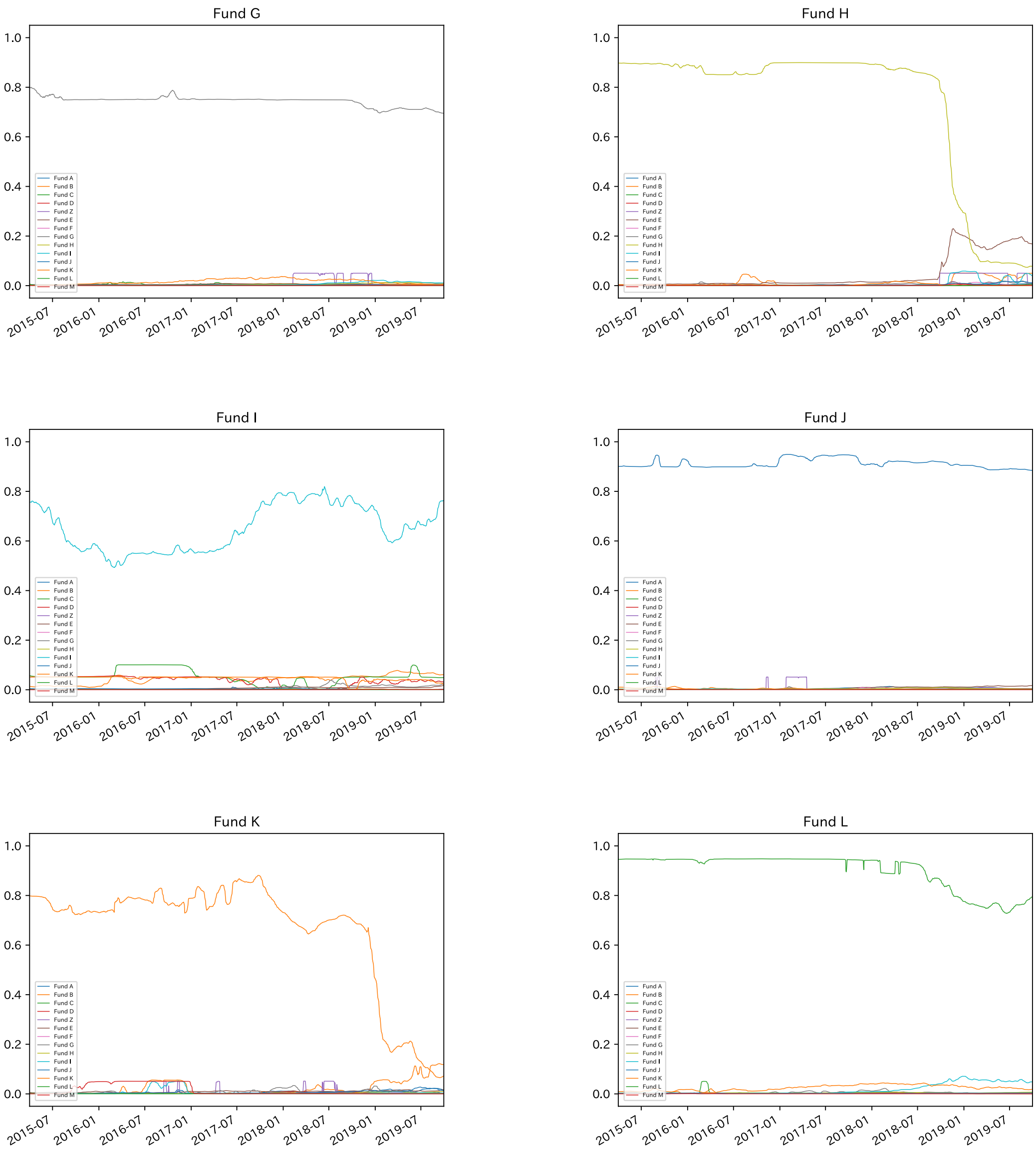


Figure 25: Outputs of Resembler — Japanese equity funds (2/3)

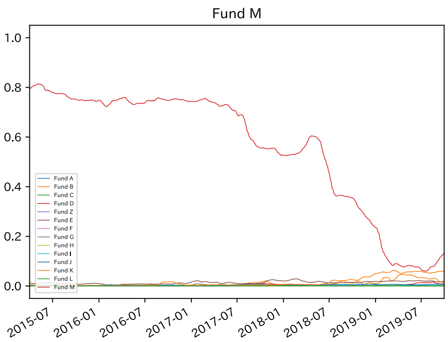


Figure 25: Outputs of Resembler — Japanese equity funds (3/3)

A.2 Outputs of SDA for Japanese equity funds

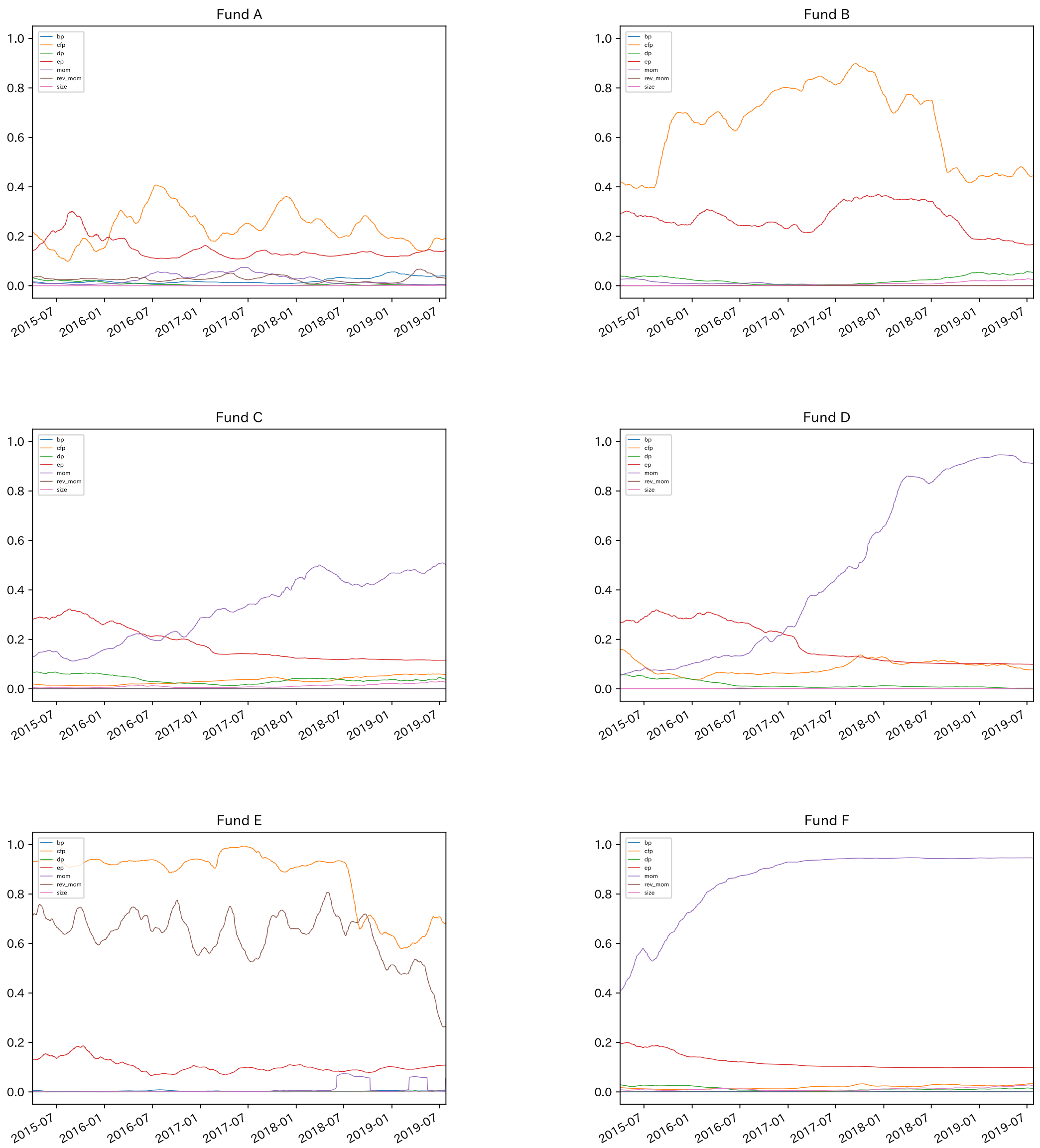


Figure 26: Outputs of SDA — Japanese equity funds (1/3)

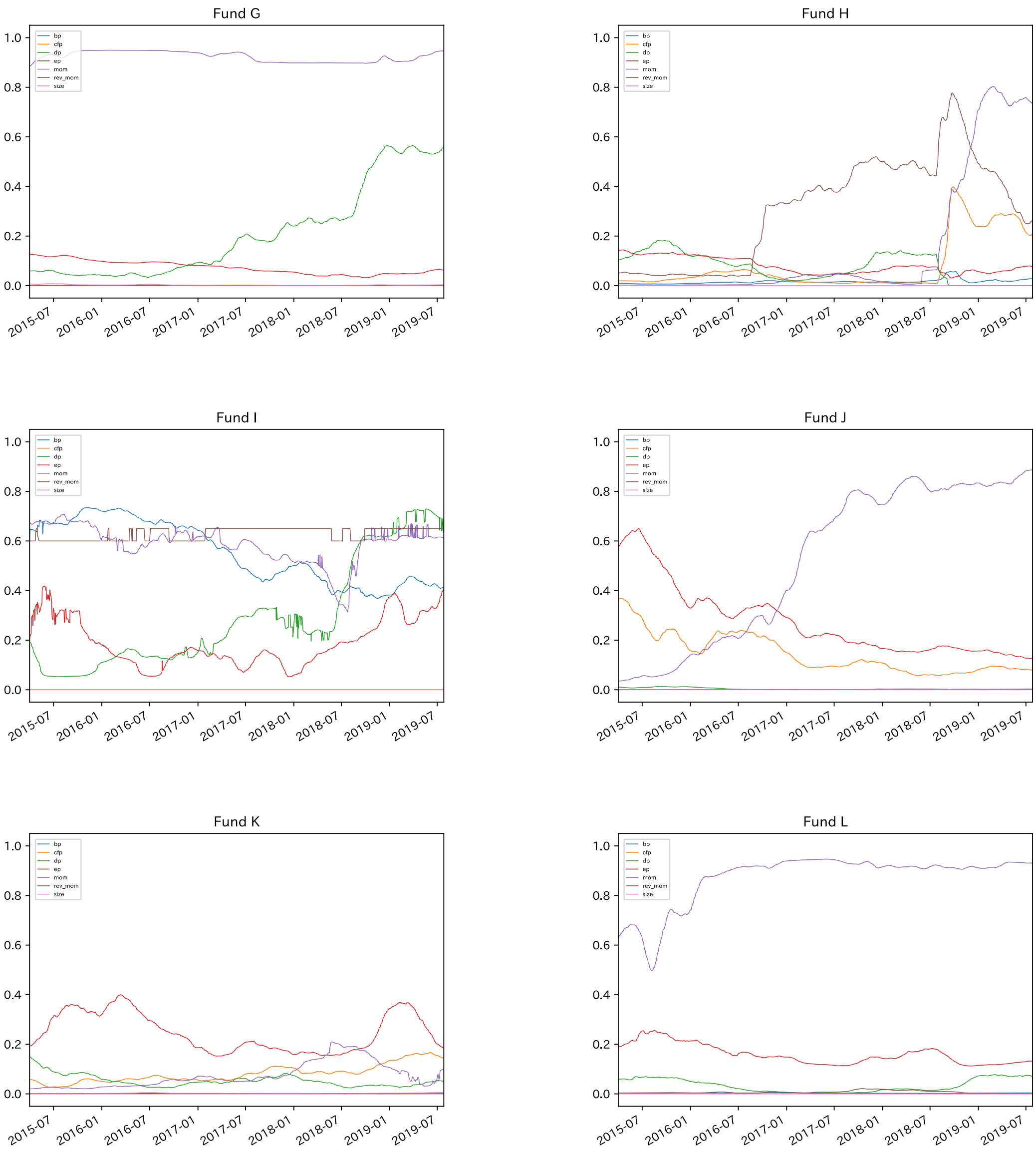


Figure 26: Outputs of SDA — Japanese equity funds (2/3)

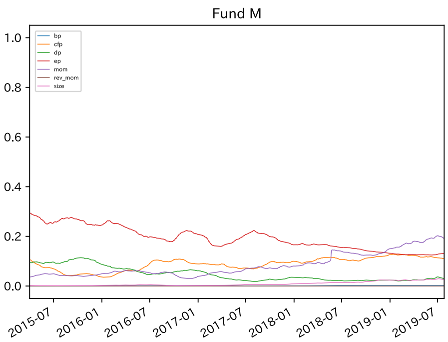


Figure 26: Outputs of SDA — Japanese equity funds (3/3)

A.3 Outputs of Resembler for foreign equity funds

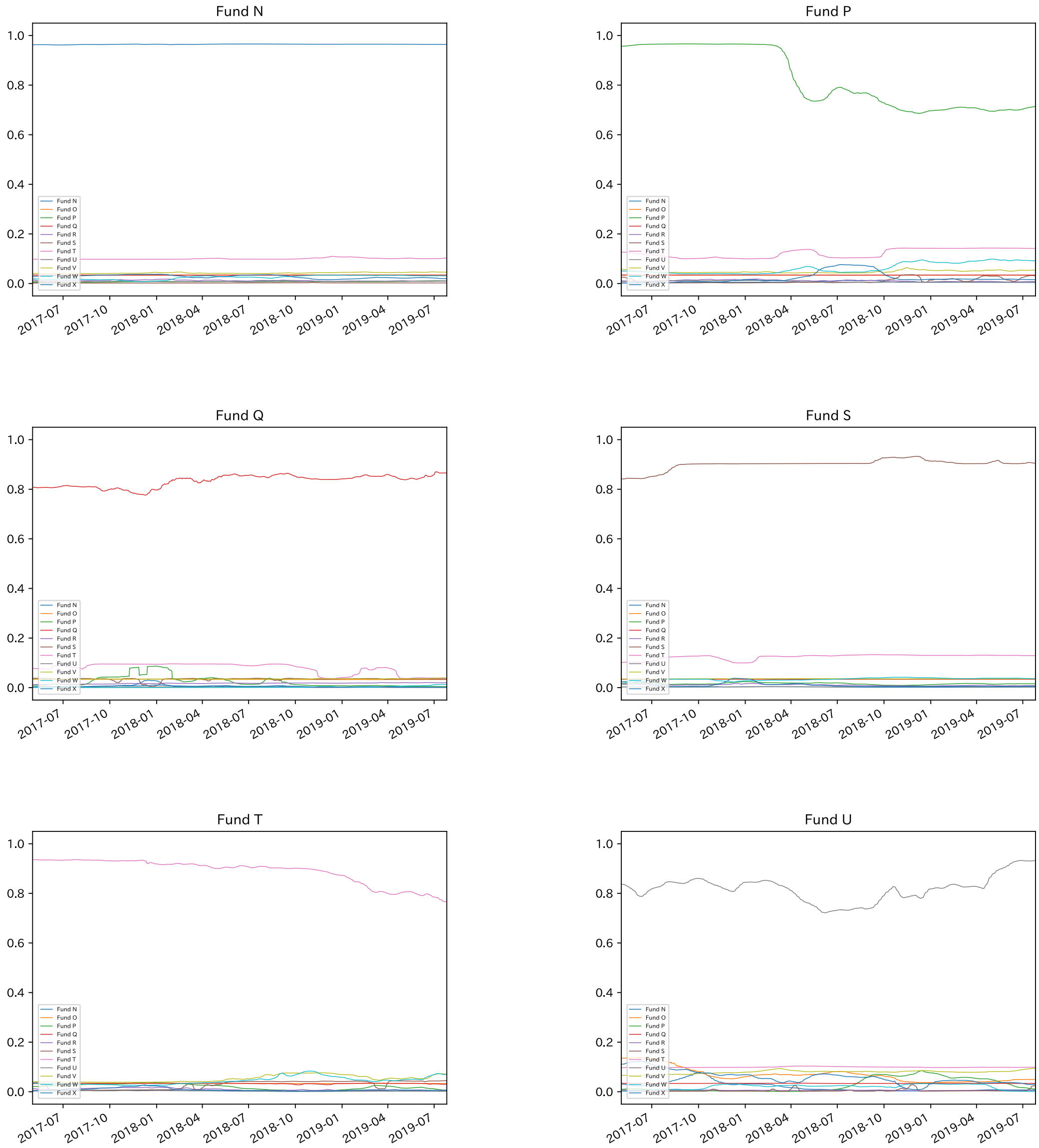


Figure 27: Outputs of Resembler — Foreign equity funds (1/2)

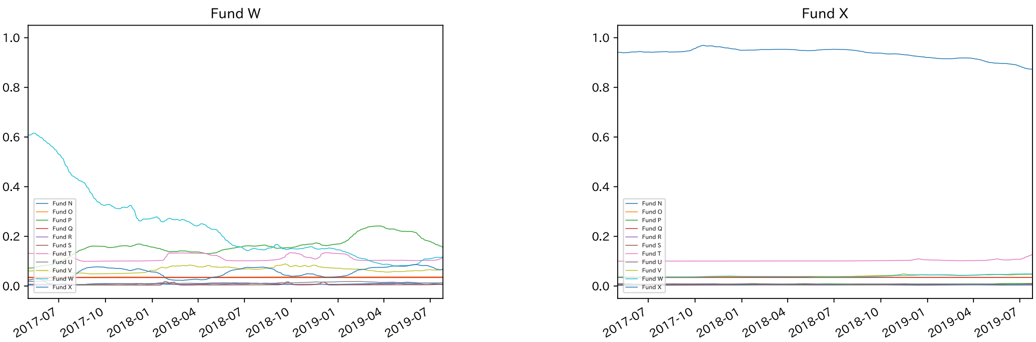


Figure 27: Outputs of Resembler — Foreign equity funds (2/2)

A.4 Outputs of SDA for foreign equity funds

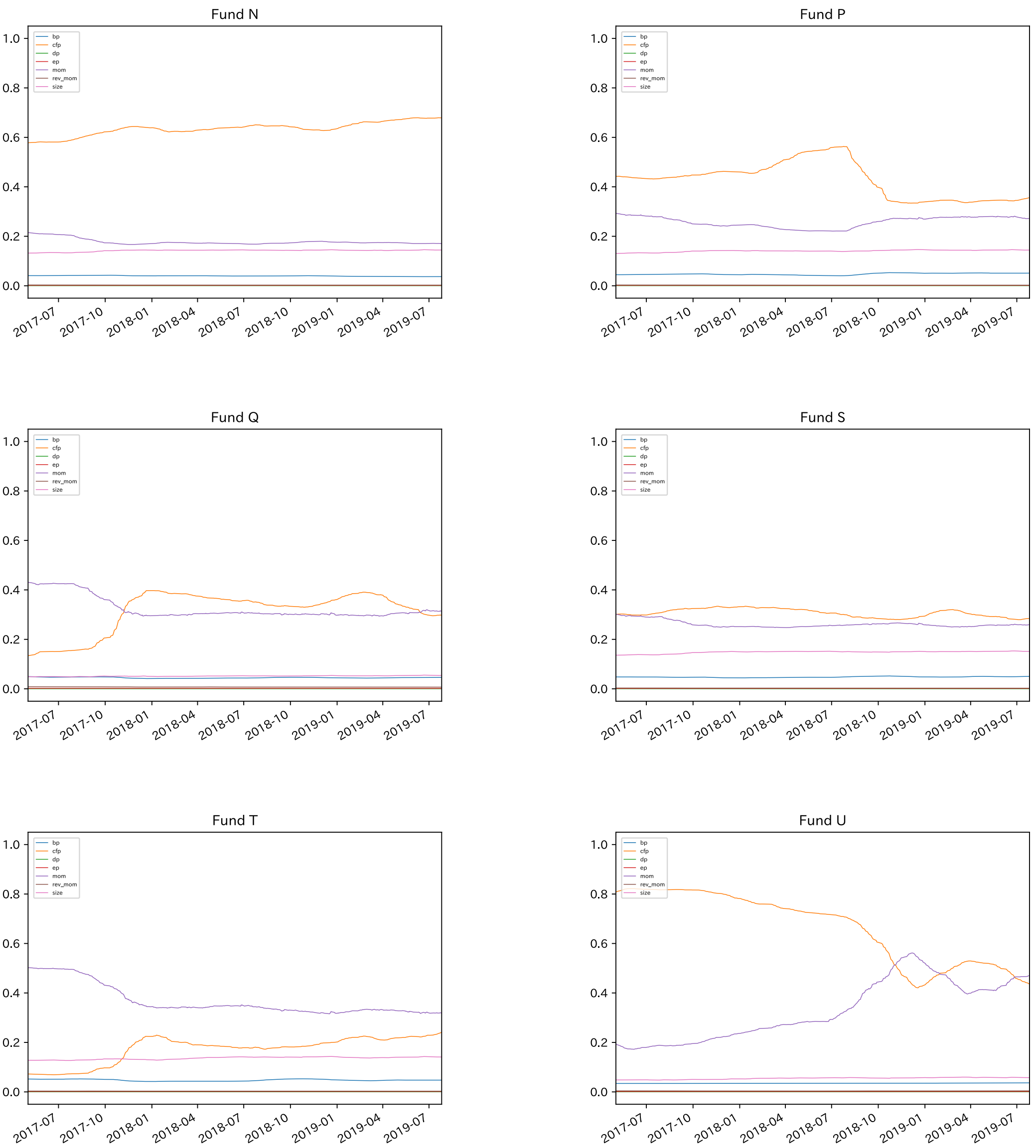


Figure 28: Outputs of SDA — Foreign equity funds (1/2)

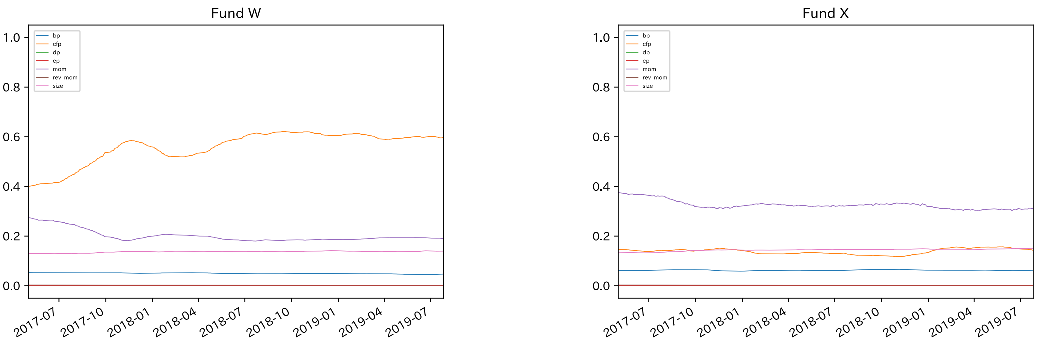


Figure 28: Outputs of SDA — Foreign equity funds (2/2)

References

[1] Takahiro Sasaki, Hiroo Koizumi, Takao Tajiri, and Hiroaki Kitano. A study on the use of artificial intelligence within government pension investment fund ' s investment management practices (summary report). Technical report, Tokyo, Japan: Government Pension Investment Fund, 2018.

[2] AI Bridging Cloud Infrastructure (ABCI). `https://abci.ai`.

[3] Fujitsu. Accelerating the societal implementation of AI with AI Bridging Cloud Infrastructure (ABCI) Realizing world-class performance and energy saving with 4,352 GPUs. `https://www.fujitsu.com/global/documents/solutions/business-technology/tc/fields/case-studies-e/abci/abci201909-e.pdf`, 2019.

[4] Government Pension Investment Fund (GPIF). Annual report fiscal year 2018, 2018.

[5] G E Hinton and R R Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, Vol. 313, No. 5786, pp. 504–507, July 2006.

[6] Quoc V. Le, Marc'Aurelio Ranzato, Rajat Monga, Matthieu Devin, Greg Corrado, Kai Chen, Jeffrey Dean, and Andrew Y. Ng. Building high-level features using large scale unsupervised learning. In *ICML*. icml.cc / Omnipress, 2012.

[7] Eugene F. Fama and Kenneth R. French. Common risk factors in the returns on stocks and bonds. *Financial Ecomomics*, Vol. 33, No. 1, pp. 3–56, February 1993.

[8] Mark M. Carhart. On persistence in mutual fund performance. *The Journal of Finance*, Vol. 52, No. 1, pp. 57–82, March 1997.

[9] Yi Shang and Benjamin W. Wah. Global optimization for neural network training. *IEEE Computer*, Vol. 29, No. 3, pp. 45–54, 1996.

[10] M. S. Iyer and R. R. Rhinehart. A method to determine the required number of neural-network training repetitions. *IEEE Trans. Neural Networks*, Vol. 10, No. 2, pp. 427–432, 1999.

[11] Akarachai Atakulreka and Daricha Sutivong. Avoiding local minima in feedforward neural networks by simultaneous learning. In Mehmet A. Orgun and John Thornton, editors, *Australian Conference on Artificial Intelligence*, Vol. 4830 of *Lecture Notes in Computer Science*, pp. 100–109. Springer, 2007.

[12] Warren S. Torgerson. Multidimensional scaling: I. theory and method. *Psychometrika*, Vol. 17, No. 4, pp. 401–419, December 1952.

[13] Takashi Jimba. GPIF Working paper: GPIF ' s New Performance-Based Fee Structure. Technical report, June 2018.

[14] PwC. Asset & wealth management revolution: Embracing exponential change, 2017.

[15] MIT Technology Review. AI will now watch for fraudsters on the world ' s largest stock exchange (November 7, 2019). `https://www.technologyreview.com/s/614691/nasdaq-ai-will-now-watch-for-fraudsters-on-the-worlds-largest-stock-exchange/`.

Reporter:
Takao Tajiri
Takahiro Sasaki
Hiroaki Kitano

Sony Computer Science Laboratories, Inc.
Takanawa Muse Bldg. 3F,
3-14-13, Higashigotanda, Shinagawa-ku,
Tokyo, Japan 141-0022 Tel: 03-5448-4380