

# データ分析手順理解のための操作ログ解析方法の提案

大澤奈々穂<sup>†1</sup> 鍵本麻美<sup>†1</sup> 山本南美<sup>†1</sup> ジョージ フェリックス<sup>†1</sup>  
大崎雅代<sup>†1</sup> 上野洋平<sup>†1</sup>

**概要**：公共プラント等の異常要因解析等のデータ分析作業において、膨大な信号の中から信号を見ていく順序等の分析手順は個々のデータ分析者のノウハウとして暗黙知となっていた。そこで我々は、人が分析手順を理解することを目的に、データ分析ツールの操作ログを取得・可視化・解析し、データ分析手順やノウハウを理解・比較する方法を提案する。操作ログの解析方法として、【案1】表示信号（信号IDと期間）の解析、【案2】分析手順の解析、【案3】信号関連解析の3案を提案し、実際に配水プラントシステムのデータ分析を実施した操作ログ4名分で検証する。

## 1. はじめに

電力や上下水道などの公共プラント、製造工場などで稼働するシステムでは、センサから得られる大量の信号データを監視・制御することにより日々の連続的な運転を維持している。これらのシステムでは、機器故障、異常気象等の様々な要因による非定常な異常が発生した際、その異常要因を迅速に特定し、対応策を講じることが求められる。その際、膨大な数の分析対象信号の時系列データを可視化（グラフ化）し、異常に関連する信号を特定して要因解析を行うデータ分析作業を行う必要があり、それを支援する分析ツール・分析支援ツールが多数開発されている[1][2]。

しかしこれらのデータ分析作業において、どの信号をどの順にグラフ化し関連を調べていくのかという分析手順は特にルール化されておらず、個々のデータ分析者の分析ノウハウは暗黙知となっていた。そこで我々はデータ分析ツールの操作ログを取得・可視化し、データ分析の手順やノウハウを理解・比較する方法を提案する。

2章で操作ログを活用したデータ分析手順理解の概要について述べ、3章で操作ログの解析方法を提案する。4章では、実際に操作ログを取得して3章で提案した解析を行った結果について述べ、5章でまとめる。

## 2. 操作ログを活用したデータ分析手順の理解

図1に操作ログを活用したデータ分析手順理解の概要



図1 操作ログを活用したデータ分析手順の理解

を示す。

### (1) 操作ログの取得

データ分析手順を理解するためには、データ分析作業を実施したときの分析手順の履歴を取得する必要がある。このとき、分析ツールや分析支援ツールを利用するのであれば、その操作ログを取得することで分析手順履歴を取得できる。ただし、分析者が異なる分析ツールや複数の分析ツールを利用した際にも手順理解や比較を行うためには、分析ツールに依存しない操作ログの取得が求められる。

分析ツールに依存しない操作ログの取得方法として、「表示信号（信号IDと期間）」と「表示形式」に分けて操作ログを取得することを提案する。「ボタンA押下」のような分析ツール依存の操作ログではなく、「表示信号=信号1:10月1日10時~12時」「表示形式=折れ線グラフ」のように、一般的な操作ログを記録する必要がある。

### (2) 操作ログの解析

(1)で取得した操作ログを解析し、データ分析手順を理解、ノウハウを抽出する。解析の方法としては、次の4種類が考えられる。

(2-1) 操作ログの可視化：データ分析手順を人が目視で理解しやすくするために、操作ログをグラフや色等で可視化する。

(2-2) 操作ログのフェーズ分解：どこからどこまでが同じ目的の一連の操作・手順で、どこで切り替わったのかをわかりやすく整理するために、操作ログをフェーズに分解する。

(2-3) 複数人の手順比較：人毎や事象毎の分析手順の違い・類似性を比較する。特に、2-2でフェーズ分解しておくことにより、分析手順の比較がしやすくなる。

(2-4) 操作ログの集計・統計：データ分析手順を体系化するために、複数人・複数回のデータ分析手順の集計・統計量を、操作ログから計算する。

### (3) 解析結果の活用

(2) で解析したデータ分析手順やノウハウを活用する。活用方法としては下記 3 種類が考えられる。

(3-1) 人による分析手順・ノウハウの理解・共有・伝承:

(2) の解析結果を分析者同士が目視で読み解き理解することで、分析手順を学びスキルアップにつながる。例えば、(2-1) (2-2) で可視化・フェーズ分解されたベテラン分析者の分析手順を初心者が知るにより、ベテラン分析者の分析ノウハウを共有化できる。また、(2-3) により人毎の分析手順の類似性・差異が比較できると、手順の類似な分析者が集まって情報共有する、手順の異なる分析者が意見交換しお互いの方法を学び合うといったことも可能になる。

(3-2) リコメンド・ガイダンス: (2) の解析結果を分析ツールに組み込み、分析作業時に次の手順等をガイダンスとして表示する等、分析ツールの支援機能として利用できる。(2-2) で操作ログがフェーズ分解され、(2-4) で統計量が計算されると、例えば、「手順 A の後には手順 B を行う分析者が多い」という計算結果が得られれば、手順 A を実施中の分析者に「次は手順 B を実施してはどうでしょうか」等のリコメンドを行うことができる。他にも、「信号 1 と信号 2 は同時に表示し関連を分析する分析者が多い」という計算結果が得られれば、信号 1 を表示中の分析者に対して「信号 1 と信号 2 はよく一緒に表示されています」といった情報提示が可能となる。

(3-3) 分析者の状態検知: (3-2) のリコメンドやガイダンスは常に多数出ると、分析を阻害する要因にもなり得る。そこで、操作ログから分析者の状態を検知し、分析者の状態に応じたリコメンド・ガイダンスを出すより適切である。例えば、「一定時間操作がない」「何度も同じ操作を繰り返す」といった時には、分析者は困っていると判断し、このような時のみリコメンドを行うといったことが考えられる。

### 3. 操作ログ解析方法の提案

2 章 (3) で述べた操作ログ解析結果の活用のうち、(3-1) 「人による分析手順ノウハウの理解・共有・伝承」を目的

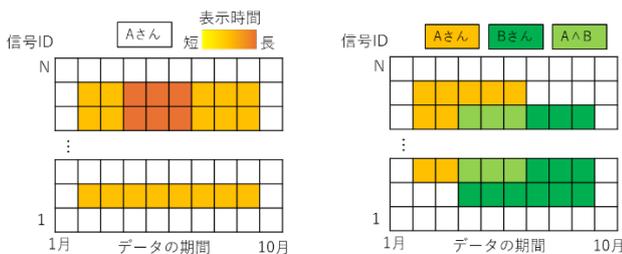


図 2 【案 1】表示信号 (信号 ID と期間) の可視化 (左) 表示時間を用いた色付け, (右) 分析者毎の違い

として、2 章 (2) 操作ログ解析の方法について 3 案提案する。

#### 【案 1】表示信号 (信号 ID と期間) の解析

分析者が分析システムで表示した時系列データの信号と期間を表にして各セルを色付けして可視化し (2-1: 可視化)、信号 ID と期間について解析する (図 2)。この可視化により、分析者がどの信号と期間を重要視して分析したのかを知ることができる。例えば、図 2 (左) のように表示時間により色の濃淡を変更する。また、図 2 (右) 分析者毎に色を変更し、一定時間以上表示していた信号 ID と期間の組み合わせに色付けして重ねることにより、同じ信号に着目していたのか否か等、手順を比較することができる (2-3: 複数人の手順比較)。膨大な信号・期間から特定の信号・期間を探し出すようなケースの場合には、分析者毎に分析する信号や期間の範囲を決めて分担することも考えられる。このようなときには、分析者毎に色付けすることにより誰がどの範囲を分析のしたのかがわかり、役割分担がしやすくなる。

また、複数人の手順を比較する際、表示していた信号 ID と期間の組み合わせが重複する面積で類似度を計算することも可能である (2-3: 複数人の手順比較) (2-4: 集計・統計)。例えば、2 人の分析者 A, B の類似度を下記式で表す。

$$\text{類似度} = (\text{A} \cap \text{B}) \text{ の面積} / (\text{A} \cup \text{B}) \text{ の面積}$$

#### 【案 2】分析手順の解析

分析した時間に対して、表示した信号をガントチャートのように並べて可視化し (2-1: 可視化)、表示した信号順を元に分析手順の解析を行う (図 3)。このような可視化により、いつどの信号をどの順序で表示したかが目視で確認でき、人が分析手順を理解できるようになる。

また、表示した信号 ID の組み合わせが大きく切り替わるところで手順の区切りがあったと判断し、フェーズ分解を行う (2-2: フェーズ分解)。図 3 (左) の例では、信号 ID1~3 を主に表示している T0~T1 をフェーズ 1、信号 ID4~6 を表示している T1~T2 をフェーズ 2 としている。フェーズ分解は表示した信号 ID の組み合わせを元に一般的なクラスタリング手法を用いて行う。各分析者でこのようなフェーズ分解を行い、さらに複数分析者間でフェーズに含まれる表示信号からどのフェーズが類似なのかの計算を行うと、分析者の手順をフェーズで比較することが可能になる (2-3: 複数人の手順比較)。図 3 (右) の例では、分析者

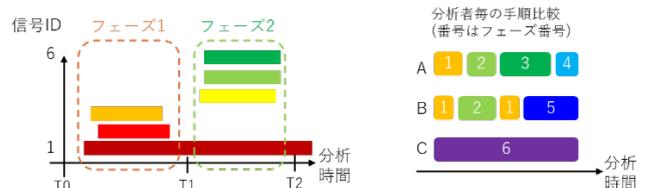


図 3 【案 2】分析手順の可視化 (左) 表示信号とフェーズ分解, (右) 分析者毎の違い

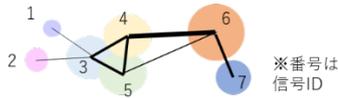


図 4 【案 3】信号関連の可視化

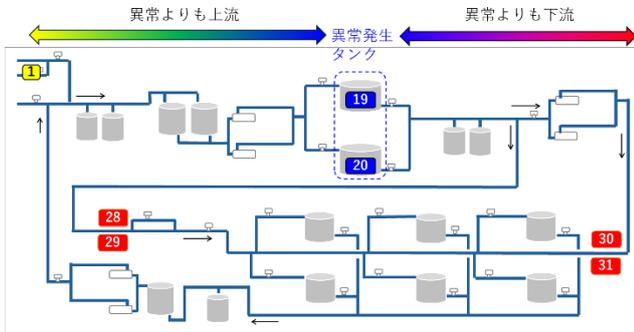


図 5 分析対象の配水プラント

A と分析者 B は最初の手順は似ていることがわかり、分析者 C は全く異なる手順で分析していることがわかる。

### 【案 3】信号関連解析

信号の時系列データを分析画面上に表示した時間の長さ、分析画面上に同時に表示した信号をグラフネットワークで表現し可視化することで、信号の重要度と関連性を解析する (2-1:可視化)。例えば、図 4 に示すように、各信号をノードとし、ノードの大きさを各信号の表示時間、信号間のエッジに同時に表示していた時間を設定し、時間の長さによってエッジの太さを変更することで信号の重要度と関連性を表現する。複数分析者の信号関連ネットワークを並べることで、分析者による信号の見方を比較することができる (2-3:複数人の手順比較)。また、複数の分析者の信号関連ネットワークに共通する部分を抽出することで、元々の信号が保有している関連情報をドメイン知識として抽出することが可能となる (2-4:集計・統計)。特に初心者には、ベテラン分析者の保有しているドメイン知識 (どの信号とどの信号が関連しているのか) が足りない場合が多いため、このような信号間同士の関連が知識として抽出され図示されると分析の際に役立つと考えられる。

## 4. 操作ログ取得・解析実験

実際に、配水プラントのデータ分析を実施し、その操作ログを取得して提案の解析方法の有効性を確認した。

### 4.1 実験状況

#### (1) データ分析対象の配水プラント

配水プラントのオープンデータである WADI テストベッド[3][4]を用い、ある期間に発生した配水異常に関連する信号の特定を実施した。

対象となる配水プラントは複数のタンクやバルブ、ポンプ等が設置され、それぞれのタンクの水位レベルや水質、流入量、流出量、バルブの開度等が信号値として計測され

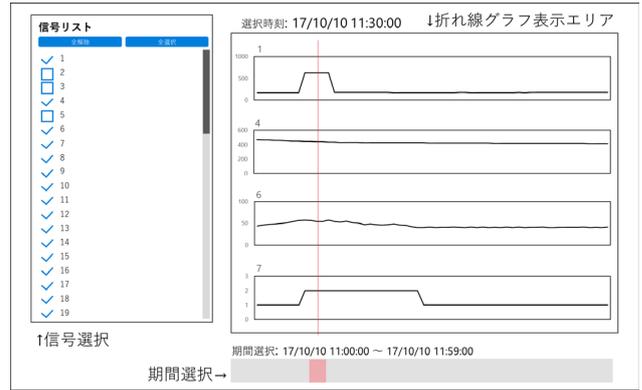


図 6 データ分析システムの画面

表 1 操作ログの例

時刻	操作	値
2024/04/17 11:06:00	折れ線	19,20
2024/04/17 11:06:10	期間	[2017/10/10 04:08:14, 2017/10/10 18:15:17]
2024/04/17 11:06:30	折れ線	
2024/04/17 11:06:34	折れ線	1,2,3,4,5,6
...		

ている (図 5)。今回、配水プラントの信号のうち分析対象の信号を 31 信号とし、おおよそ上流から下流の順になるように信号 ID を 1 から並べた。異常が発生したタンクの水位の信号 ID は 19 または 20 であり、1~18 が異常設備よりも上流、21~31 が異常設備よりも下流の信号である。分析には異常期間を含む 3 日間のデータ (1 分毎の値) を用いた。

#### (2) データ分析システム

(1) の配水プラントのデータを分析する画面を図 6 に示す。データ分析システムは「信号選択」「折れ線グラフ表示エリア」「期間選択」の機能を持つ。「信号選択」では、

(1) の配水プラントの信号 ID1~31 を選択することができる。「信号選択」で選択した信号の折れ線グラフが「折れ線グラフ表示エリア」に表示される。このとき並べられる折れ線グラフは最大 6 個までだが、6 個を超える信号が選択された場合はスクロールバーでグラフ表示エリアをスクロールさせることが可能である。「期間選択」では表示させる期間範囲を自由に変更することができ、期間範囲を選択すると「折れ線グラフ表示エリア」に表示される折れ線グラフの期間 (X 軸) が全て共通に変更される。

#### (3) データ分析システムから取得した操作ログ

(2) のデータ分析システムの操作履歴を操作ログとして取得する。表 1 に取得する操作ログの例を示す。操作ログとして、折れ線グラフ表示エリアに表示される信号 ID と期間選択を行った際に選択された期間の範囲を取得した。折れ線グラフ表示エリアに表示される信号 ID は、「信号選択」で信号の選択有無を変更したとき、既に「信号選択」で選択され表示済みの折れ線グラフがスクロールバーをス

表 2 被験者毎の分析時間

被験者	分析時間
A	18分06秒
B	36分25秒
C	30分59秒
D	1時間05分23秒

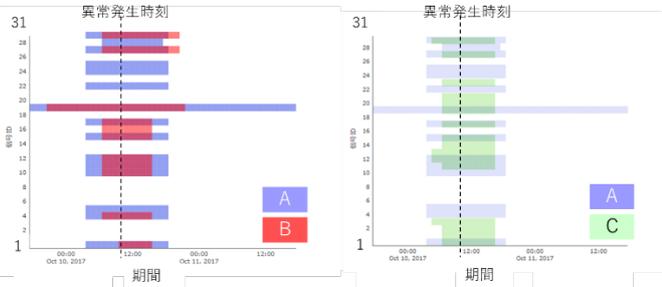


図 7 長時間表示した信号 ID と期間の組み合わせの可視化 (左) 被験者 A と B, (右) 被験者 B と C

クロールすることで変更された場合に取得される。

#### (4) データ分析のタスク

(2) のデータ分析システムを用いて (1) の配水プラントの信号を分析し、その際に分析時の操作ログ (3) を取得した。データ分析のタスクとして「配水プラントのタンク水位 (信号 ID19 または 20) に異常が発生した。異常発生のおおよその時刻と、異常に関連する信号を特定せよ」という課題を与えた。4 名の被験者に同じ課題でデータ分析を行ってもらい、その操作ログを取得した。

#### 4.2 実験結果

4 名の被験者の分析時間を表 2 に示す。被験者毎にばらつきはあるものの、おおよそ 20 分～1 時間程度で分析を終えている。

##### 【案 1】表示信号 (信号 ID と期間) の解析

全ての分析時間に対して長時間表示した信号 ID と期間の組み合わせを可視化した結果を図 7 に示す。「長時間」の定義について、今回は被験者毎に長時間表示した順に上位 11～16 信号を可視化した。被験者 A と B を比較すると 10 信号が同じで 6 信号が異なっていた。被験者 B と C を比較すると、8 信号が同じで、15 信号が異なっていた。3 章で述べた案 1 の類似度の値は、被験者 A と B は  $10/16=0.625$ 、被験者 B と C は 0.348 となり、被験者 A と B の方が同じ信号を見ていたことになる。

このような可視化を行うと、分析者毎にどの信号のどの期間に着目して分析を行ったのが比較できることがわかった。また、提案の類似度計算式を用いて、長時間見ていた信号を元にした被験者の分析信号類似度を計算することができた。

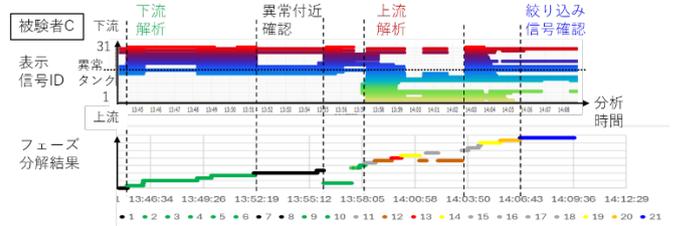
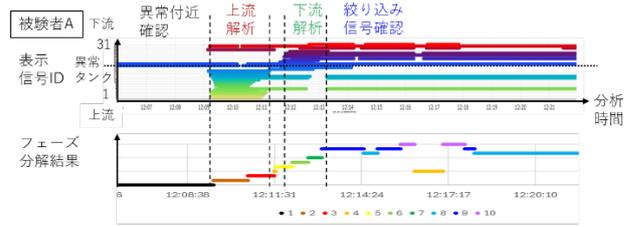


図 8 分析手順の可視化：(上) 被験者 A, (下) 被験者 C

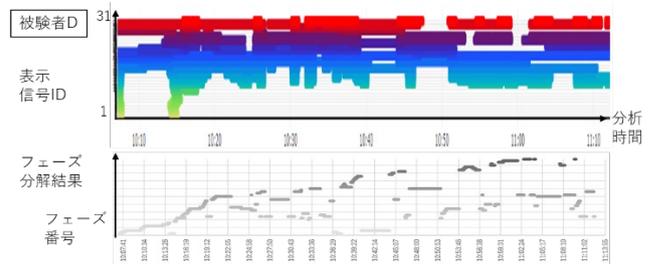


図 9 分析手順の可視化：被験者 D

##### 【案 2】分析手順の解析

分析画面に表示した信号をガントチャートのように可視化した結果を図 8 に示す。図 8 を見ると被験者 A はまず異常付近の信号である信号 19,20 (青) を確認した後、信号 19,20 よりも上流の信号 (黄色～青) を表示し、その後信号 19,20 よりも下流の信号 (青～赤) を表示し、最後にいくつかの信号に絞り込んでいることがわかる。被験者 C は、まず信号 19,20 よりも下流の信号 (青～赤) を表示し、次に異常付近の信号を確認し、信号 19,20 よりも上流の信号 (黄色～青) を表示した後、最後にいくつかの信号に絞り込んでいることがわかる。このように、分析画面に表示した信号をガントチャートのように可視化することで、分析手順を人が目視で確認し、意味を読み解き理解することが可能になる。

一般的なクラスタリング手法を用い、表示信号をクラスタリングし、フェーズ分解も実施した。被験者 A の操作ログは 10 個のフェーズに分解され、フェーズ 1 が異常付近の確認、フェーズ 2,3 が上流解析、フェーズ 6,7 が下流解析、フェーズ 8～10 が絞り込み信号の確認に当たる。目視での理解と分解粒度は多少異なるものの、おおよそ同等の分割ができています。被験者 C の操作ログも同様にフェーズ分解を行い、被験者 A と被験者 C で同種の信号を表示していたフェーズは同じ色で表示した。フェーズ間の距離計算は各

フェーズで表示していた信号割合のユークリッド距離を用いた。被験者 C では、被験者 A のフェーズ 1 (黒), フェーズ 2,3 (茶, 赤), フェーズ 7 (緑), フェーズ 9 (青) と類似のフェーズが存在することを確認できた。このようなフェーズ分解の結果を用いると、被験者 A と被験者 C では、同様の組み合わせで信号を表示しているものの順序が異なる (上流を先に見るか, 下流を先に見るか) という手順の類似性や差異を簡単に比較することができる。

図 9 は、被験者 D の表示信号とフェーズ分解結果である。被験者 D は被験者 A,C と比較して表示意図やフェーズの境界が目視では明確ではない。しかし、このようなケースでもフェーズ分解を行うと、何度も同じフェーズが繰り返されており、何度か同じ組み合わせの信号を表示しながら分析していることがわかる。また、このように操作ログを可視化することで、被験者 A,C と被験者 D では全く異なる分析スタイルを取っていることが明確になる。

### 【案 3】信号関連解析

信号の表示時間を円の大きさで表し、同時に表示した信号を線でつないで、分析対象の配水プラントの図面上に重畳表示した (図 10)。この可視化方法では、どの信号を長時

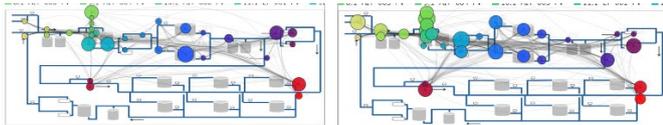


図 10 信号関連の可視化：  
(左) 被験者 A (右) 被験者 C

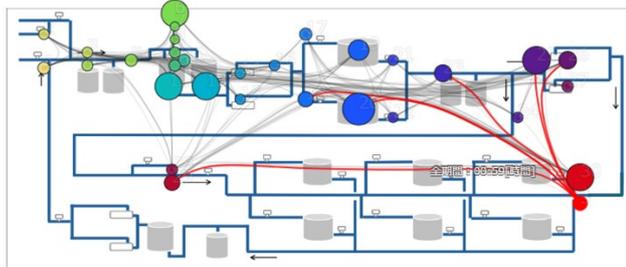


図 11 被験者 A, 信号 30 の関連信号

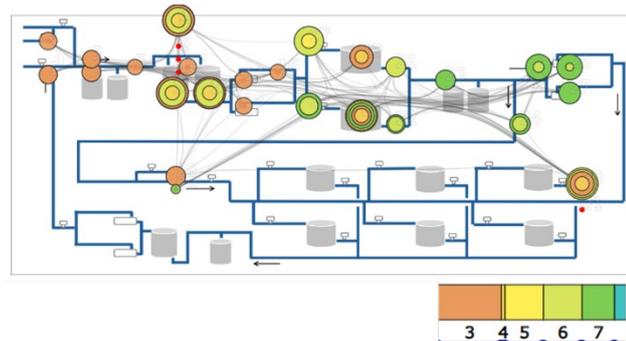


図 12 被験者 A, フェーズ 3~7 での表示信号

間見ているのかを被験者毎に比較することができる。

また、図 11 に示すように同時に見ている信号を線でつないで表示すると、全信号が接続されているわけではないことがわかる。このように同時に表示している信号を接続して可視化することで、信号の関連解析ができ、膨大な数のすべての信号の関連を調べる必要はなく関連するいくつかの信号を同時に表示して解析すればよいという知見が得られる。このような知見が蓄積されると、物理的に離れた場所にあっても関連の強い信号等もわかってくる。関連の強い信号がわかれば、分析システムにこの知見を組み込んで、同時に表示し分析した方がよい信号としてリコメンドすることも可能になる。

図 12 は【案 2】分析手順の解析のフェーズ分解と【案 3】の信号関連解析を組み合わせた可視化方法である。図 12 のように、フェーズ毎に色をつけて表示した信号を可視化すると、表示した信号が上流から下流に移動していったという手順と信号との関連をよりわかりやすく可視化することが可能になる。

## 5. おわりに

膨大な信号の異常要因解析等データ分析作業において、その分析手順やノウハウは個々の分析者の暗黙知となっていた。そこで我々は、人が分析手順を理解することを目的に、データ分析ツールの操作ログを取得・可視化・解析し、データ分析の手順やノウハウを理解・比較する方法を提案した。操作ログの解析方法としては、【案 1】表示信号 (信号 ID と期間) の解析, 【案 2】分析手順の解析, 【案 3】信号関連解析の 3 案を提案した。実際に配水プラントシステムのデータ分析を実施した操作ログを 4 名分取得し、提案の 3 案で操作ログの解析を行ったところ、3 案ともにデータ分析手順を理解することに役立ち被験者毎の手順比較も実施できることがわかった。今後は被験者の数を増やし、定量的な評価を行っていく予定である。

## 参考文献

- [1] 米田詠美, 大崎雅代, 上野洋平, 今井健. 異常兆候分析支援ツールの提案. 電気学会全国大会. 2021, pp.237-238.
- [2] 鍵本麻美, 大崎雅代, 上野洋平. 時系列信号データの可視化による異常要因分析支援システムの開発. 第 52 回可視化情報シンポジウム 講演論文集. 2024.
- [3] C. M. Ahmed, V. R. Palleti, and A. P. Mathur, "WADI: a water distribution testbed for research in the design of secure cyber physical systems," in Proc. of the 3rd International Workshop on Cyber Physical Systems for Smart Water Networks (2017) pp. 25-28.
- [4] iTrust, Centre for Research in Cyber Security, Singapore University of Technology and Design, "Datasets - iTrust," (2024) [https://itrust.sutd.edu.sg/itrust-labs\\_datasets/](https://itrust.sutd.edu.sg/itrust-labs_datasets/)