

## 共同研究報告書

研究課題: 可視化を応用する AI 学習用データ作成の高速化の研究

研究担当者 北海道大学大学院情報科学研究科

特任教授 山本 強

あらまし IoT やビッグデータが産業や生活を変えるとと言われて久しい。コンセプトとしての第4次産業革命、超スマート社会（ソサエティ 5.0）の「絵」は描かれたが、その社会実装はまだ見えていない。その一方、IoT を実現する要素技術は急速に低価格化しており、サービスが企画できれば実現する手段は簡単に手に入るようになっている。自分が必要な IoT サービスは作ろうと思えばすぐ作れる時代になっている。しかし、エッジコンピューティングとしての IoT システムを作っても、それを活用する認識ロジックや識別器まで含むサービスシステムにするには生のデータでは不十分であり、ラベル付けされた学習用データに変換する必要がある。今のところ、ラベル付けされた学習用データの生成は人手作業に大きく依存しており、その作成時間短縮や精度向上には改善の余地が多い。本研究では、IoT によって実現される情報収集システムを想定して、それを可視化するシステムの提案と実装を行う。可視化システムを活用することで、ビッグデータに表れる特異現象やイベントとデータの関連などが容易に発見できるようになり、深層学習で必要になる学習用データベースの作成が容易になることを示す。

キーワード IoT, 可視化、数値データベース、ブラウザ、データマイニング、AI

## 1. まえがき

第五期科学技術基本計画ではキーワードとして IoT, ビッグデータ, AI が頻出し、それらを活用して実現される「超スマート社会」を今後の国の技術開発の方向性として定義してからすでに3年が経過した。関連した分野の研究開発も盛んであり、サービスとして実用化されているものも見えてきている。

一方でこの分野の研究アクティビティを見るに、日本の存在感は決して高いとは言えない。この分野に限らず、アカデミックコミュニティにおける日本の貢献が低下しているというのは以前から言われていることだが、今後の社会インフラとして重要になるのが確実な情報通信やそれを応用する IoT, ビッグデータ, AI などの分野で存在感が薄いというのは深刻な問題である。この研究分野を活性化させるためには次世代の研究者・エンジニアとなる学生さんがそういった分野の本質を理解し、関心を持って貰うことが重要なのは明らかなのだが、一方で理系離れや研究分野の細分化、セグメント化が進んでおり、概念は理解できるが実際に何をすればいいのか、IoT やビッグデータでどんな問題が解決できるかを具体的なケースとして知らない人が増えているのが現実である現実の問題、特に自分に直接かわる問題に対して、IoT を応用できということ、自分自身が技術を身に着けていることの証明ともいえる。自分のために使えないようでは、社会が求める応用に対応できないとは思えないのである。ということで、ここではあえて自分のためと話を矮小化して、IoT やビッグデータを取り扱う具体的な例を紹介することにする。

## 2. キーワード、コンセプトから実践へ

ビッグデータがあれば何でもわかるみたいな解説がある反面、具体例はと言われると教科書的なものしか出てこないという現実もある。Society 5.0 というキーワードは冒頭で紹介した第5期科学技術基本計画の文書で使われている用語であるが、こういった行政文書も見方によってはビッグデータとなる。第五期があるということは、第四期があるわけで、両

## 共同研究報告書

研究課題: 可視化を応用する AI 学習用データ作成の高速化の研究

研究担当者 北海道大学大学院情報科学研究科

特任教授 山本 強

文書は内閣府のホームページ上に PDF 形式で公開されている。ちなみに第五期は 522KB(63 ページ)、第四期は 397KB(54 ページ)である。行政文書をビッグデータ的に読む試みとして、単語の出現頻度を比較してみると面白いことがわかる。ここ数年の間にメディアで頻繁に取り上げられた単語の出現数を第四期と第五期でカウントし、その比率を算出した結果が表 1 である。

キーワード	4 期出現頻度	5 期出現頻度	5期/4期
超スマート社会	0	28	$\infty$
IoT	0	11	$\infty$
AI	0	10	$\infty$
ビッグデータ	0	5	$\infty$
サイバー	1	26	25.97
ビジネス	1	23	22.98
ベンチャー	7	60	8.57
セキュリティ	3	18	6
オープン	7	33	4.71
グローバル	10	34	3.4
持続的	14	27	1.93
地球	18	32	1.78
小子	4	7	1.75
イノベーション	127	214	1.69
安全保障	9	15	1.67
国家	15	16	1.07
高齢	15	16	1.07
国民	70	45	0.64
エネルギー	58	26	0.45
バイオ	6	1	0.17
震災	52	8	0.15
グリーン	13	0	0

表 1. 第四期、第五期の科学技術基本計画におけるキーワード出現頻度算出結果

この表は PDF リーダーの検索機能と EXCEL を使って手作業で作ったものだが、政策文書であってもその内容を数値的に評価できることがわかる。例えば、超スマート社会、IoT、AI、ビッグデータという用語は第四期では一度も出現しておらず、第五期で突然使われたキーワードであることがわかる。逆に、国民、エネルギー、バイオ、震災という用語は第五期が総文字数が多いにも関わらず出現頻度は第四期よりも減少している。グリーンに至っては第五期では一度も使われていないことがわかる。

この実験はテキスト、音声、画像といった非数値データであっても、統計的処理を経ることで数値データ（ベクトルデータ）として取り扱うことができることの例になっている。数値化されたデータは可視化により、全体構造が理解しやすくすることができる。

## 2. IoT 情報可視化の基本方針

IoT やセンサーネットワークというキーワードで検索すると、事例やサービスがたくさん出てくる。雑誌媒体でも Raspberry Pi や Arduino などのプラットフォームを応用した実装例が出てくるし、技術情報は沢山あるのだが、事例として出てくるのは LED 点滅と温度と明度のセンシングといった単純化された「サンプル」にとどまっている。「できることをする」のと、「必要なことをする」の間に大きな隔りがある。

IoT に関してよくある誤解は、IoT を称するデバイスは超小型で軽いシステムであることから、IoT システム全体が小さくてシンプルなものだと思われることである。実際にシステム構築するとわかるのだが、IoT のエッジ部分は確かにシンプルで小さいのだが、システム全体は多層構造で大規模なものになることが多い。つまり、システムを一から作るのは多岐に渡る技術分野を理解することを求められるのである。

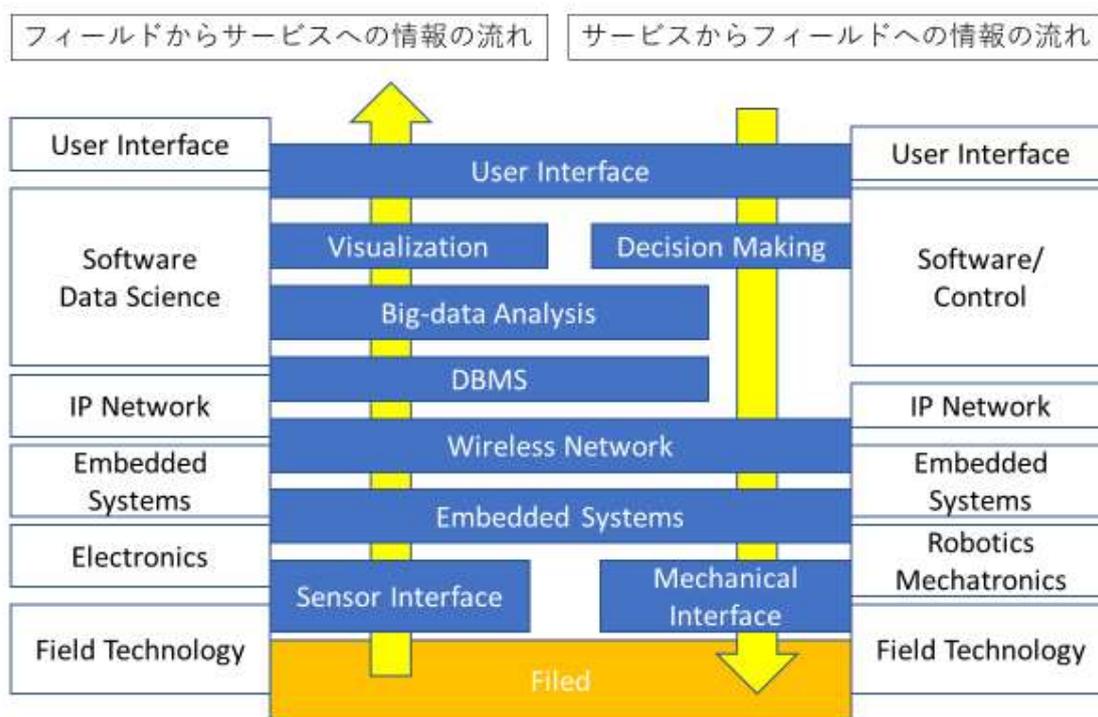


図 1. IoT サービスを構成する要素技術の階層

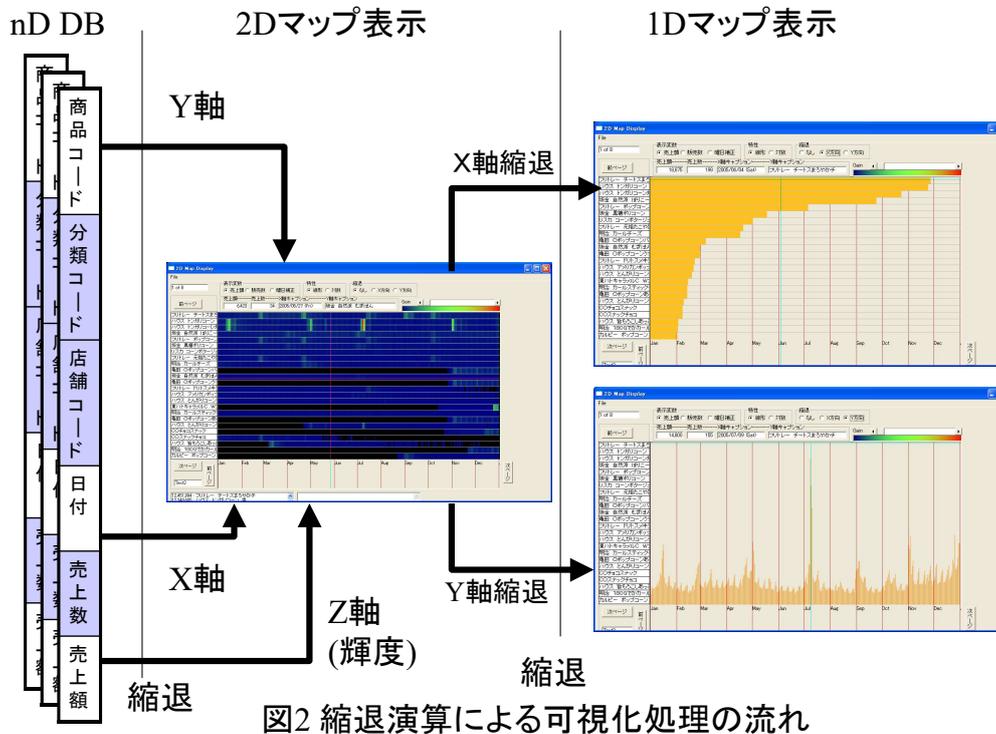
## 共同研究報告書

研究課題: 可視化を応用する AI 学習用データ作成の高速化の研究

研究担当者 北海道大学大学院情報科学研究科

特任教授 山本 強

数値データの可視化、グラフ化は商用システムで既に行われているが、それはその分野での標準的な集計手法や定番のグラフを作成するものがほとんどで、一般化された可視化手法ではない。本研究では、数値データの分析と可視化のワークフローを分析し、それを汎用かつ高速にこなすためのアーキテクチャを提案する。



### 3. スタック指向の仮想マシンアーキテクチャ

本提案ではデータベースの縮退と画像化処理の組み合わせにより、可視化処理を行う。従来の代表的アーキテクチャは可視化処理に必要な機能をブロック分割し、データフローモデルで結合するものが多い。データフローモデルは3次元画像生成のグラフィックスパイプラインモデルのアナロジーがあり、ハードウェア集合体として可視化システムを抽象化している。これはコンピュータが発明された当時に ENIAC が演算ユニットのパッチボード接続により計算を定義していたのと同じモデルである。その後、コンピュータアーキテクチャが、いわゆるノイマン型のスタックマシンに基本構造を変化させてきたことを考えると、可視化システムの基本構造をよりフレキシブルな仮想マシンアーキテクチャで構築するほうが汎用性があり、自由度が高くなる可能性がある。

本研究で提案するアーキテクチャは縮退演算の結果を2値1変数の正規化されたチャートオブジェクトとし、それを一時的に格納するスタックを介在させることにより、コンパクトな可視化アーキテクチャを実現する。図3に提案するアーキテクチャの基本構造を示す。

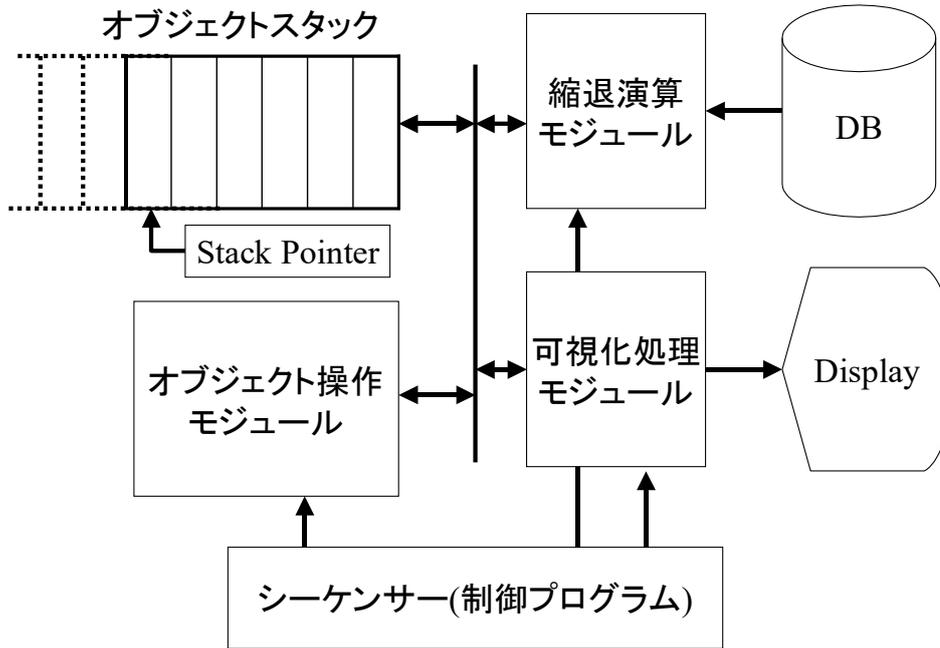


図3 スタックを併用する可視化処理用仮想マシンアーキテクチャ

スタックに対する基本操作として表2の操作命令を定義する。

移動操作	<b>PUSH</b> (表示中オブジェクトをスタックに移動)
	<b>POP</b> (スタックを1段POP)
演算操作	<b>加減乗除</b> (スタックの先頭の2オブジェクトに対する2項演算)
	<b>縮退</b> (スタック全体を加算した結果のみをスタックに残す)
	<b>正規化</b> (スタックの先頭オブジェクトをスタックの一番底にあるオブジェクトで除算する)
展開操作	<b>展開</b> (変数を2まで縮退した後、1変数についてスタックに展開する)
初期化操作	<b>クリア</b> (スタックを初期化する)
定数生成操作	<b>定数</b> (値が定数、固定関数であるオブジェクトを生成し、スタックにPUSH)

表2 可視化支援バーチャルマシンの基本オペレーション (仮想命令セット)

シーケンサはデータベース縮退とスタック操作命令を組み合わせ、順序実行することで高度な可視化処理を実現することができる。可視化処理にスタックを取り入れることの利点はソフトウェアの構造的な整理だけでなく、マニュアル操作による可視化操作にもある。これは、かつて Hewlett Packard 社, Texas Instrument 社のプログラム電卓の基本操作 (逆ポーランド記法) と似ている。シーケンサの制御プログラムをマニュアル操作で置き換える

## 共同研究報告書

研究課題: 可視化を応用する AI 学習用データ作成の高速化の研究

研究担当者 北海道大学大学院情報科学研究科  
特任教授 山本 強

ことにより、高レベルな電卓操作と似た感覚で可視化処理を行うことが可能になる。マニュアルでの操作をジャーナル記録することで、その操作を新しい可視化プロセスとすることもできることになる。

従来、ビッグデータ分析はデータベースに対してあらかじめ決められた集計処理手順を行い、結果を表示するという形式のものが多かったが、ここで提案するモデルでは、集計手順をモジュール化しその中間処理結果をスタックに複数保存してそれらを演算処理して様々な表示形態を作成することができ、分析者主体で表現を短時間で作成することが可能になる。

### 4. IoT センサーネットワークと可視化システムの融合

提案したバーチャルマシンを元に、IoT センサー端末が収集したビッグデータの可視化およびブラウジングのためのクラソフトウェアツールを開発した。本研究では、ビッグデータを A.I.の学習用データにするプロセスを高速化することがその目的であり、突発的に発生した災害や事象に適用できる即応性を持つことを確認することに意味がある。研究実施中の平成 30 年 9 月 7 日に発生した平成 30 年北海道胆振東部地震が引き起こした北海道全域の停電災害、いわゆるブラックアウト現象の分析に研究成果を活用した。

ブラックアウトは発電所の電力供給力と負荷のバランスが大きく崩れ、交流周波数が維持できなくなった時に周波数低下による二次被害を防止するために電力供給を遮断することが連鎖的に発生することで起こると言われている。私たちは日常的に 50/60Hz の商用交流電源に依存しているが、その周波数の変動など気にすることはない。しかし、今回のブラックアウトを経験したことにより交流周波数の変動がエネルギー供給と消費のバランス崩壊の前兆ということらしいのだが、その情報はリアルタイムでは公開されていない。

しかし瞬時電源周波数を継続して測定し、その変動をビッグデータ化し分析することで電力システムのバランス状態を消費者側でも観測できる。しかし、その観測機器や分析システムが市販されているわけではない。本研究では、ブラックアウト後に瞬時電源周波数測定用 IoT 端末を短期間開発し、9 月 12 日から観測を開始した。図 4 は開発した瞬時電源周波数測定モジュールの主要回路である。このモジュールは交流電源の周期を 1 サイクル毎に 50ppm 精度で計測することが可能である。

この測定モジュールを IoT エッジコンピュータ Raspberry Pi に装着し、インターネット接続することによりクラウド上に周波数データを蓄積することができる。写真 1 は実装した IoT センサーノードである。この IoT センサーノードは小型軽量であり、WiFi 環境があれば AC 電源(100-220V)に接続するだけで電源周波数変動を測定できる。

共同研究報告書

研究課題: 可視化を応用する AI 学習用データ作成の高速化の研究

研究担当者 北海道大学大学院情報科学研究科

特任教授 山本 強

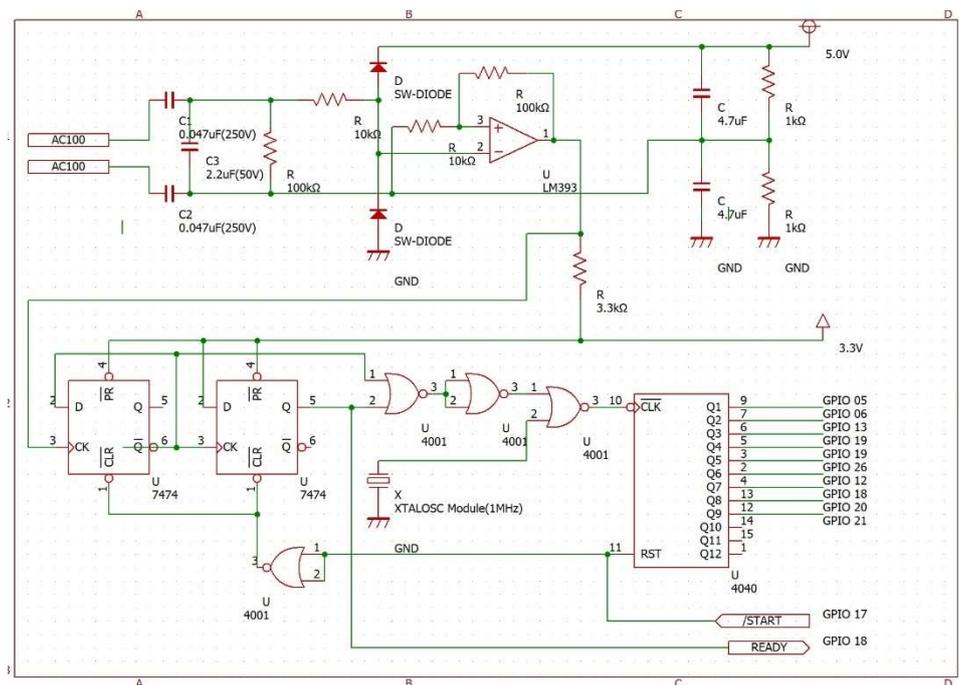


図 4. 瞬時電源周波数測定回路



写真 1. 電源周波数測定用 IoT センサーノードの実装

このセンサーノードを 9 月 12 日夜に立ち上げ、以後現在まで継続して測定を続けている。そのデータは 1 分単位での最大・最低・平均の周波数データとして数値データベース化されている。図 5 は測定開始翌日の 9 月 13 日終日の観測データの可視化の一例である。この可視化では上から 1 分間最大周波数 (桃色)、平均周波数 (白色)、最低周波数 (黄色) が表

## 共同研究報告書

研究課題: 可視化を応用する AI 学習用データ作成の高速化の研究

研究担当者 北海道大学大学院情報科学研究科

特任教授 山本 強

示されている。

この可視化結果を見ると、午前 4 時から 5 時 40 分過ぎまでの約 100 分に 0.025Hz の安定した周波数上昇と、午後 4 時-8 時の間に不規則な周波数変動が観測されていることがわかる。後日、北海道電力関係者に確認したところ、早朝の周波数上昇区間は電気時計の時刻補正のための周波数上昇であることが判明した。これは交流電源周波数を原クロックに使用する電気時計のために、波数カウントによる時刻計測の誤差が 3 秒累積した時に 0.025Hz (0.05%) の周波数上昇により補正するために意図的に挿入するものとのことである。

### Powerline Frequency of 2018-09-13

[Data View](#)

[Previous day](#)

[Next day](#)

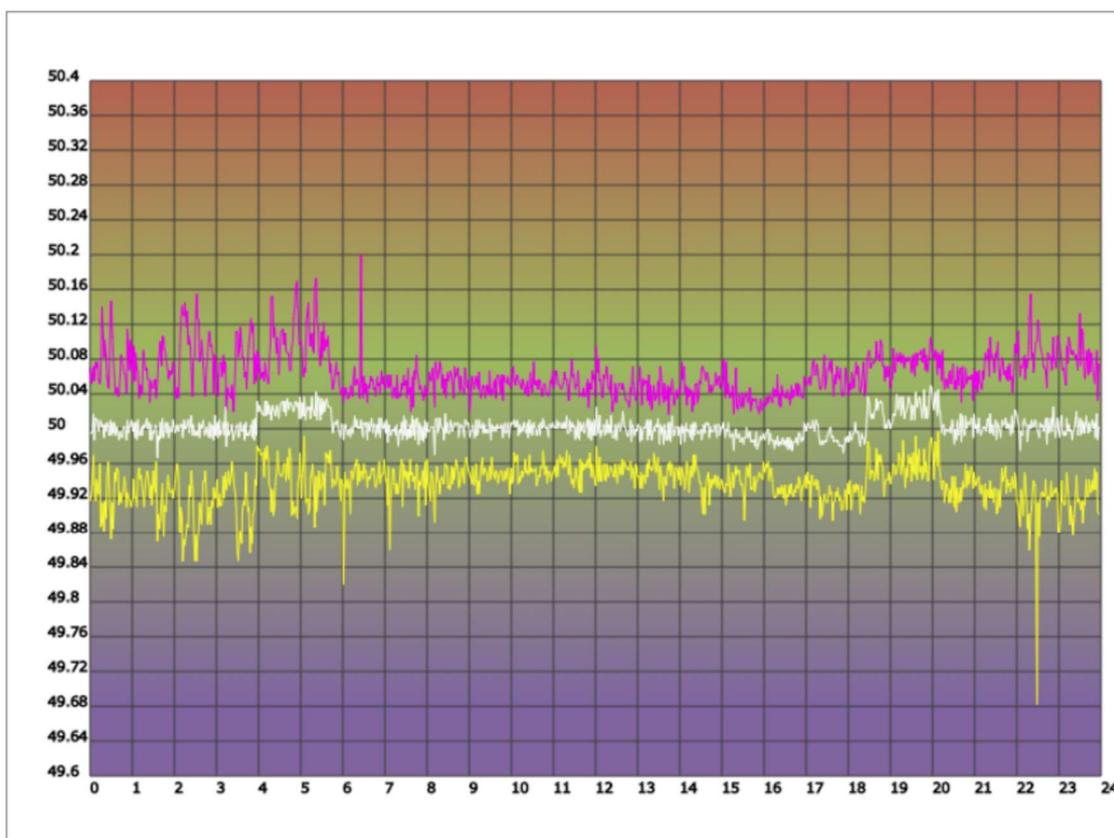


図 5. ブラックアウト直後 (9 月 13 日) の北海道電力周波数変動測定結果の可視化

A.I.用学習データの生成では、大量データから特徴あるデータを抽出し、分類する必要がある。特徴がある程度数値化されれば以降の分類作業は自動化可能であるが、初見のデータでは何が特徴であるかもわからない。新しいビッグデータが入手された場合、最初に行ななければならないのはその特徴の把握である。その段階で重要になるのは、データを俯瞰するためのブラウジング機能である。今回のブラックアウト後の周波数モニタリングでは、対話的なブラウザ環境を実装しており、それを用いて異常データの検出を簡単に行う

## 共同研究報告書

研究課題: 可視化を応用する AI 学習用データ作成の高速化の研究

研究担当者 北海道大学大学院情報科学研究科

特任教授 山本 強

ことができた。図 6 は 2019 年 4 月 18 日の電源周波数観測データであるが、早朝 7 時過ぎに急激な周波数上昇が観測され、その後意図的な周波数低下制御が行われていたことがわかる。この時の周波数上昇は 0.1Hz 以上の上昇が 5 分間以上継続しており、その波数誤差が大きかったため、直後に時刻補正が入ったことが読み取れる。

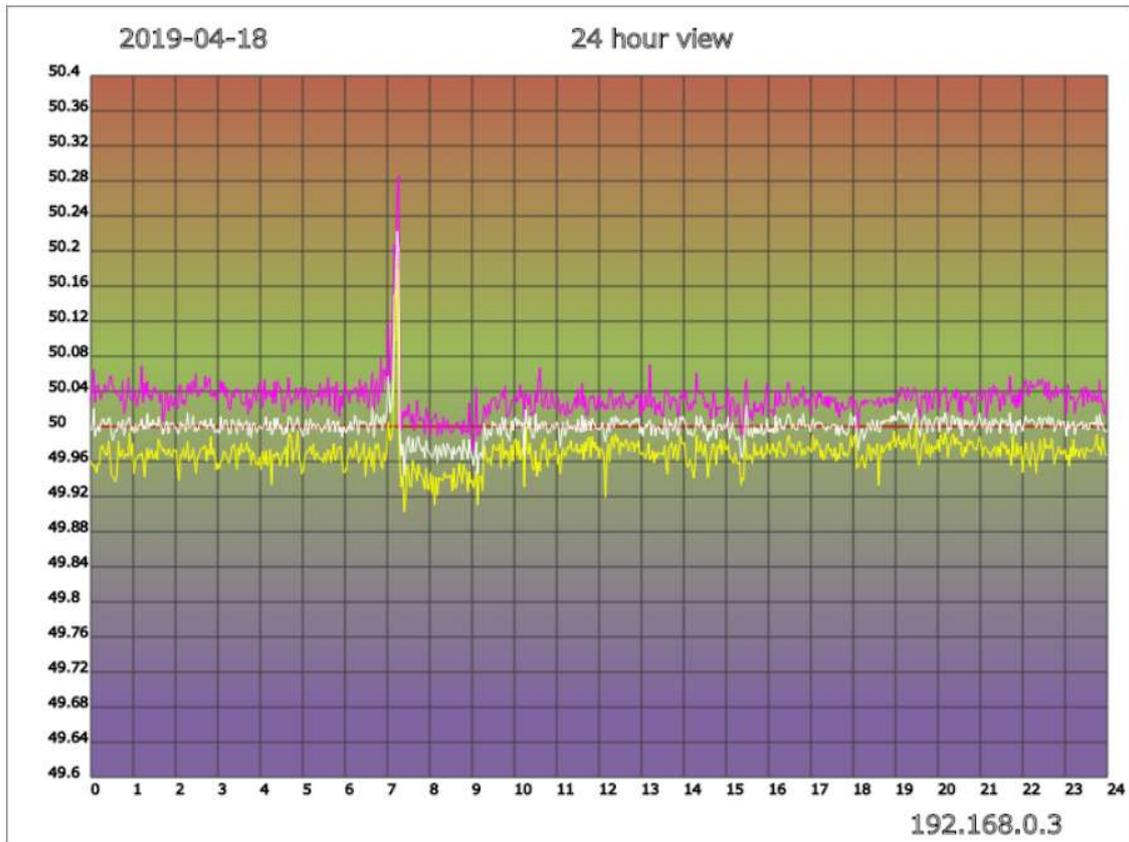


図 6. 2019 年 4 月 18 日に観測された、異常周波数変動パターン

本研究の一環でヘルシンキ（フィンランド）での A.I.研究動向に関する調査を実施した際に、フィンランドでの電源周波数変動の調査を実施した。図 7 は 2018 年 12 月 4 日、ヘルシンキ・バンター空港隣接のホテルでの測定した 1 日の周波数変動の観測結果である。明らかに札幌（北海道電力）との周波数変動傾向の差が見える。フィンランドでの商用交流電源の変動は日本のそれと比較して明らかに大きく、不安定である。しかし、それでも EU の電源周波数変動基準から外れているわけではない。

## 共同研究報告書

研究課題: 可視化を応用する AI 学習用データ作成の高速化の研究

研究担当者 北海道大学大学院情報科学研究科

特任教授 山本 強

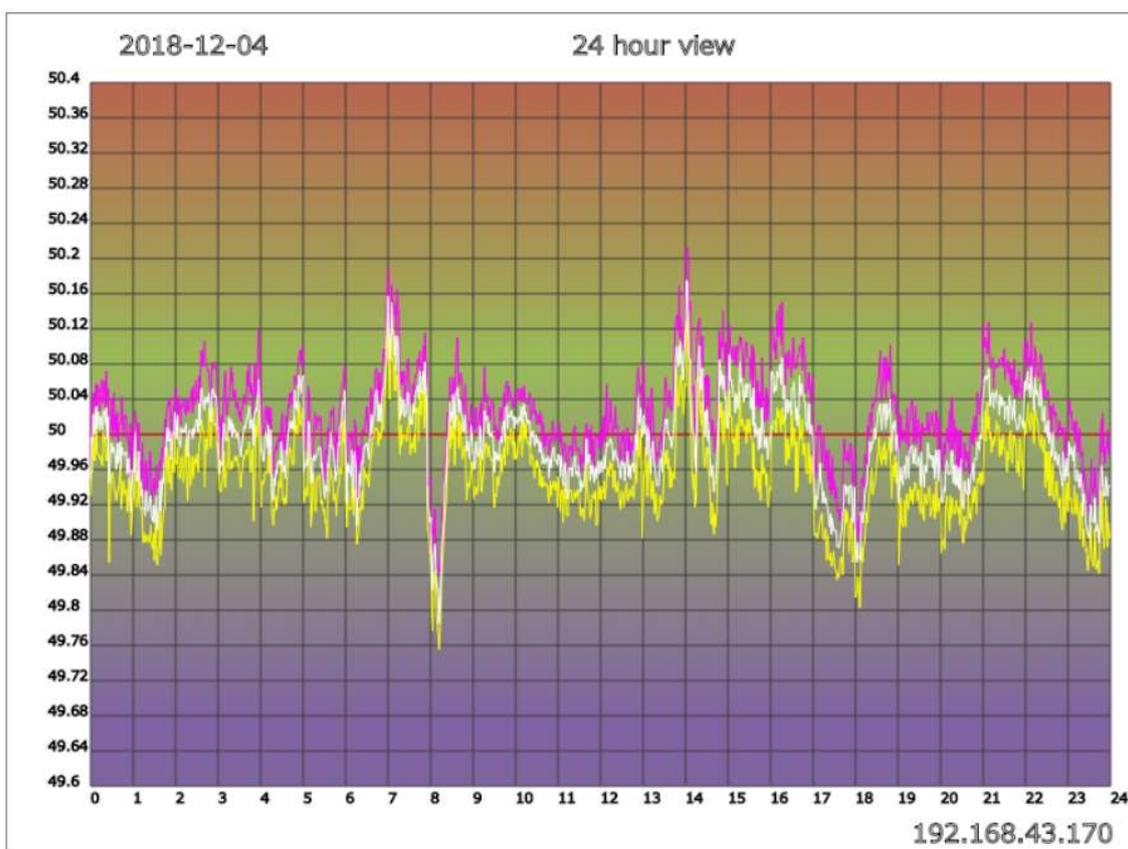


図 7. ヘルシンキ（フィンランド）での周波数変動測定結果

## 5. むすび

A.I.による社会変革が期待されているが、実際の応用に A.I.を組み込もうとするとビッグデータの収集、クレンジング、そして学習用データの作成が課題となることが多い。特に、短時間でデータの構造や特徴を理解し、実現可能な抽象的な話に陥りがちな IoT やビッグデータも自分の日常生活に必要な機能、実際に起こった事から出てくる機能を実装すると考えれば、できることは意外と多いものである。これが必要と考えたときに、それを実現する基本的な技術やノウハウを備えていることでできる IoT 応用分野は多い。あれが無い、これが無いと出来ない理由を探すのではなく、今手元にある技術で何ができるか、それに何かを加えると新しいことができる、それならばそれを体得しようという精神が、Society5.0 を実現するには肝心である。