

データセンタ消費電力予測向け ML モデル管理技術 Study of ML Model Management Technology for Data Center Power Consumption Prediction

河野 泰隆[†]
Yasutaka Kono

1. はじめに

脱炭素社会の実現による地球温暖化の抑止は近年、特に注目を集めている社会課題の 1 つである。一方、情報通信分野では全世界で生成・消費されるデータ量の爆発的な増加を背景として消費電力が増加している。特にデータセンタ(DC)の年間消費電力はグローバルで 190 [TWh] (2018)から 500,000 [TWh] (2050)に増加するとの予測がある[1]。このように大量の電力を消費する DC の消費電力効率改善は、脱炭素社会の実現に向けて重要な課題である。

DC の消費電力の大部分は IT 機器の消費電力と空調設備の消費電力が占める。このうち DC 事業者の観点では、自社が管理権限を有する空調設備が電力効率改善の主対象である。空調設備の消費電力は主に外気温とサーバールームの室温に影響され、室温は主に IT 機器からの排熱に影響される。外気温が空調消費電力に与える影響の程度は DC の立地条件やサーバールームの位置 (方角やフロア)によって異なる。また IT 機器からの排熱が室温を介して空調消費電力に与える影響の程度は、サーバールーム内のラック密度やラック内のサーバ密度によって異なる。したがって DC の電力効率改善に向けては、DC やサーバールームごとに異なる特徴に合わせた適切な空調設定が必要である。

これに対し現状の DC では、DC 運用者による経験知に基づく改善や、熱流体解析を用いた効率が行われている。前者では日々の運用作業工数が多いことや効率改善の属人化の問題がある。後者では効率改善対象の全 DC、全サーバールームの 3 次元空間モデル作成が必要であり、コストとリードタイムが問題である。いずれの方法も多拠点への横展開の容易性に課題がある。

そこで我々は、データ駆動でサーバールームごとに異なる特徴を捉えて消費電力を分析し、非効率を検知して原因を推定する技術を確認することで DC 電力効率改善を支援することをめざしている。具体的には DC のデータを用いて機械学習(ML: Machine Learning)により将来の空調消費電力を予測し、予測に反する実測値を非効率として検知する[2]。

DC 消費電力の傾向は時間経過に伴い変化するため、予測精度を維持する仕組みが不可欠である。機械学習を用いた DC 消費電力予測に関する先行研究には、例えば[3][4]がある。しかし、これらの先行研究では初期のモデル構築フェーズにおいて精度の高い予測モデルを実現することに焦点を当てており、運用フェーズにおけるモデルの精度維持の課題は扱っていない。

モデルの精度維持の課題に取り組むに当たっては、実際の DC のデータを用いて DC 消費電力の変動や、その傾向変化が予測モデルの精度に与える影響を明らかにする必要がある。そこで我々は自社 DC の実データを用いた予備実験を通じて具体的な影響を調査した。そして当該影響下において予測モデルの精度を維持するために、モデルの再学習を適切に制御する ML モデル管理技術を研究した。

2. DC 消費電力予測の精度維持における課題

我々は自社 DC の 2 年分の実データを用いて空調消費電力予測の予備実験を行った。実験では空調消費電力、IT 消費電力、外気温のデータを用いて、以下の通り ML モデルの学習(再学習)と予測、精度評価を繰り返し実行した。

- (1) 実験における現在日時を 2020/1/1 00:00:00 と設定
- (2) 直近 30 日分のデータを用いて空調消費電力の予測モデルを作成 (Linear Regression)
- (3) 当該モデルを用いて翌日の空調消費電力を予測
- (4) 実験における現在日時を 1 日後に更新
- (5) 手順(3)で算出した予測値と、それに対応する日時の実測値を用いて誤差率を評価
- (6) 実験における現在日時が 2021/12/31 23:00:00 に至るまで、手順(2)に戻り処理を継続

実験 1

あるサーバールームの空調消費電力および外気温のデータを図 1 に示す。

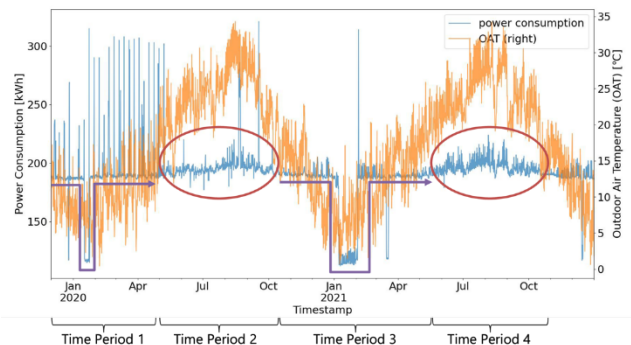


図 1: DC 消費電力の傾向変化の例

図 1 に示した例は、あるサーバールームにおいて空調設備の消費電力の傾向が時間変化している様子を示している。期間 1, 3 では空調消費電力は矩形的な変化をしており、外気温の動きとの連動性は弱い。これに対し期間 2, 4 では空調消費電力に連続的な変化が見られ、外気温の動きとの連動性が強い。このような変化は、サーバールーム内に異種の空調設備が混在していることにより発生していると考えられる。期間 2, 4 に見られるような外気温と連動性の強い空調消費電力の動きは、インバータ制御の空調設備において見られる特徴である。一方、期間 1, 3 に見られるような矩形的な空調消費電力の動きは、インバータ制御でない空調設備に見られる特徴である。このように異種の空調設備が混在するサーバールームでは、外気温の季節変化に伴ってどの種類の空調設備の制御の特徴が強まるかによって、図 1 に示したような傾向変化が生じることがわかった。

当該サーバールームのデータを用いた実験結果を図 2, 図 3 に示す。

[†]日立製作所 Hitachi, Ltd.

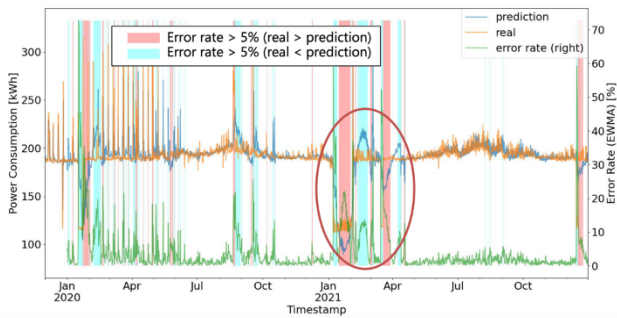


図 2: 実験結果 1 (1)

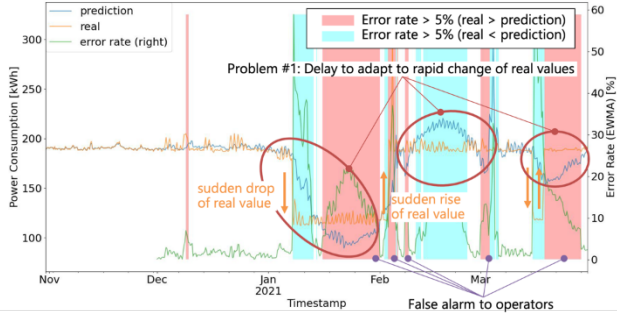


図 3: 実験結果 1 (2)

図 2, 図 3 では空調消費電力の予測値と実測値(左軸), および誤差率の EWMA (Exponentially Weighted Moving Average) (右軸)を示している。また, 誤差率の EWMA が閾値(5%)を超えている期間を示している。ここでは, 誤差率の EWMA が閾値を超過し, かつ実測値が予測値よりも大きい場合は消費電力の非効率の可能性があるとして DC 運用者に通知することを想定している。

図 2 に示した結果からは, 空調消費電力の急変化が起こった際に, ML モデルの再学習によって予測結果が実測値の変化に追従できるまでに時間差が生じている様子がわかる。図 3 はその期間を拡大表示したものである。図 3 に示した結果から, 消費電力の変動傾向が矩形的に変化した際に, 実測値に対する予測値の追従遅れと, その結果として生じる誤報が問題となることがわかった(問題点 1)。

実験 2

別のサーバールームの空調消費電力および外気温のデータを図 4 に示す。

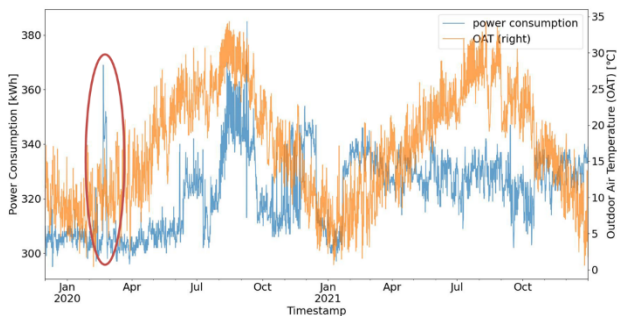


図 4: DC 消費電力の一時的変化の例

図 4 では, 一時的に空調消費電力が増加し, 約一週間後に元の程度に戻る様子を示している。このようなアノマリは, 日々の DC 運用において何らかの原因で生じた空調消

費電力の非効率である可能性がある。当該サーバールームのデータを用いた実験結果を図 5, 図 6 に示す。

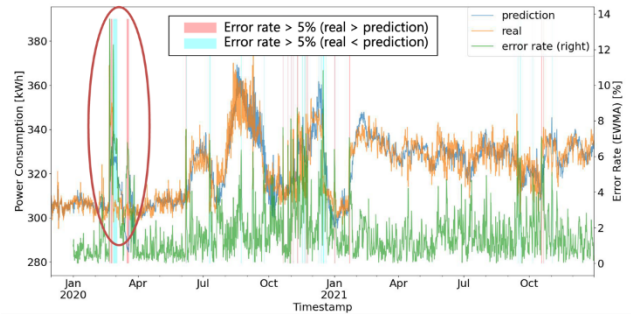


図 5: 実験結果 2 (1)

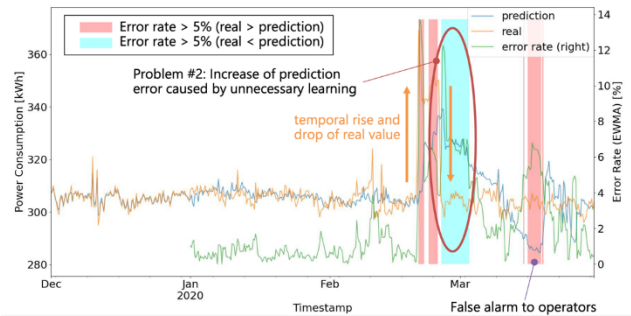


図 6: 実験結果 2 (2)

図 5 に示した結果では, 一部の期間において一時的な空調消費電力の急増・急減(アノマリ)が発生しており, その影響で予実差が生じていることがわかる。図 6 は該当期間を拡大表示したものである。図 6 に示した結果から, アノマリを学習したことによる予測誤差の増加と, その結果として生じる誤報が問題となることがわかった(問題点 2)。

以上から我々は, 消費電力の傾向変化に追従すること, 非効率な消費電力変動の学習を回避することを DC 消費電力予測の精度維持における課題と定めた。

3. 提案方式

本章では, 2 章で述べた課題を解決するための ML モデル管理方式の提案を行う。提案方式の概要を図 7 に示す。

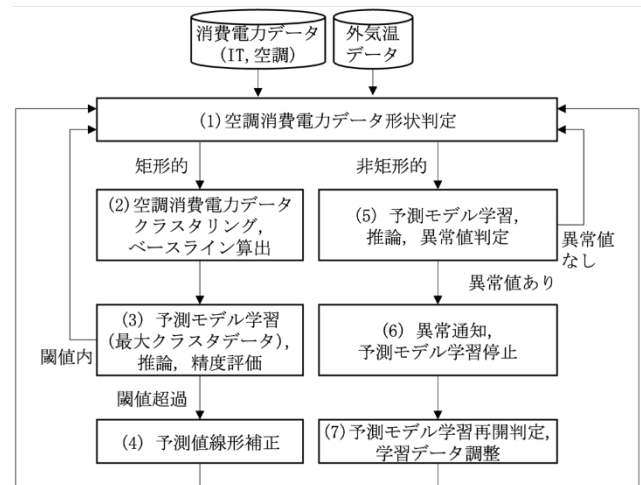


図 7: 提案方式の概要

提案方式では、まずサーバールームごとに空調消費電力の変動傾向を分析し、矩形的な急変化があるか否かを判定する(図 7 (1))。本判定処理はデータ分布の尖度を閾値と比較することで行う。これは、矩形的な変化のあるデータはそうでないデータに比べて取りうる値の範囲が狭い(分布の尖度が大きい)という考えに基づく。但し、我々が扱った自社 DC の実データの中には、矩形の段数が多いために取りうる値の範囲が広いデータが存在しており、単純にデータ分布の尖度による判定を行うと誤判定を起こすことがわかった。そこで我々は、空調消費電力のデータを 1 つ前のデータ点との差分を算出した上で、その分布の尖度を判定に用いる方式を採用した。図 8 に異なる 3 パターンの空調消費電力データと、それぞれの差分データの例を示す。

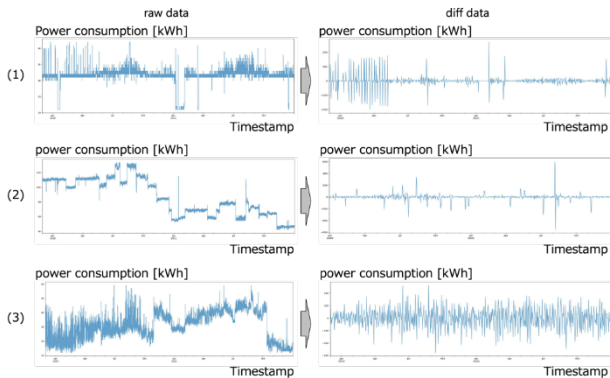


図 8: 空調消費電力データと差分データの例

図 8 の(1)は 2 段の矩形的変化を伴うデータ(図 1 に示したデータと同一)、(2)は 3 段以上の矩形的変化を伴うデータ、(3)は矩形的変化を伴わないデータである。これらを含む 18 のサーバールームのデータについて、元データの分布の尖度と、差分データの分布の尖度を図 9 に示す。

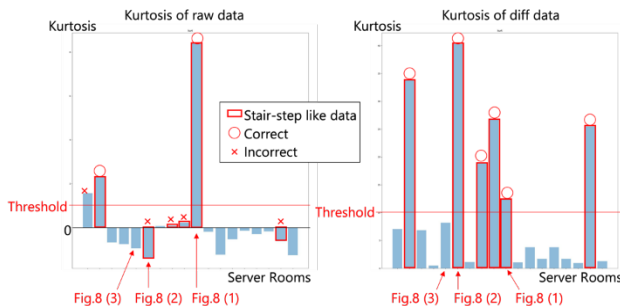


図 9: データ分布の尖度 (元データ, 差分データ)

図 9 左のグラフは元データの尖度である。図 8 の(1)に示したデータの尖度は閾値を超えており、矩形的であると判定できる。また、図 8 の(3)に示したデータの尖度は閾値を下回っており、矩形的でないとして判定できる。しかし図 8 の(2)に示したデータは尖度が閾値を下回っており、矩形的であると正しく判定することができない。また他に 4 つのサーバールームのデータで正しい判定が行えない結果となった。図 9 右のグラフは差分データの尖度である。図 8 の(1)(2)を含め、矩形的変化を伴うデータは全て尖度が閾値を超えており、矩形的であると判定できる。また図 8 の(3)を含め、矩形的変化を伴わないデータは全て尖度が閾値を下回っており、矩形的でないとして判定できる。このように、差分データの分布の尖度による判定が有効であることがわかった。

空調消費電力データが矩形的であると判定した場合、提案方式では次に空調消費電力データをクラスタリングし、全てのクラスタについてベースラインを算出する(図 7 (2))。クラスタリングには k-means 法を用いる。k-mean 法ではクラスタ数 k を与える必要があるが、 k は矩形の段数に相当するため事前に与えることはできない。そこでエルボー法を用いて適切な k の値を自動的に判定する。具体的には、 k を 1 ずつ増加させながらクラスタ内誤差平方和(SSE: Sum of Squared Errors)を算出し、SSE の変化が収束した時点の k の値を採用する。各クラスタのベースラインは、クラスタに属する空調消費電力データの平均値とする。

次に、クラスタに属する空調消費電力データ数が最も多いクラスタ(最大クラスタ)のデータを用いて空調消費電力予測モデルを学習する(図 7 (3))。そして当該モデルを用いて推論(空調消費電力の予測)を行い、精度評価を行う。予測誤差が閾値を下回っている場合は処理(1)に戻る。

予測誤差が閾値を超過した場合、予測値を線形補正する(図 7 (4))。図 10 に本処理の概要を示す。本処理では、最新の実測値に最も近いクラスタ(最近傍クラスタ)を特定する(図 10 (a))。そして最大クラスタのベースライン(b1)と最近某クラスタのベースライン(b2)との差を予測値に加算することで予測値を線形補正する(図 10 (b))。実測値が最大クラスタの近傍に戻った場合は、最大クラスタと最近傍クラスタが同一となるためベースラインの差がゼロとなり、自動的に線形補正が行われなくなる(図 10 (c))。

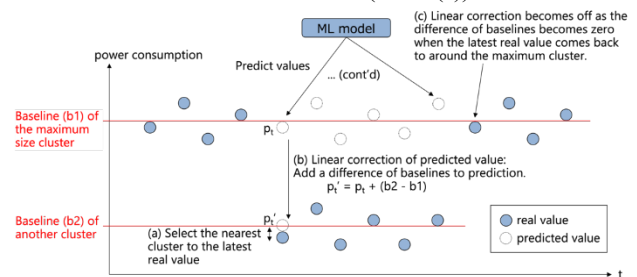


図 10: 予測値線形補正処理の概要

処理(1)にて、空調消費電力データが矩形的でないとして判定した場合、提案方式では空調消費電力、IT 消費電力、外気温のデータを用いて予測モデルを学習する(図 7 (5))。そして当該モデルを用いて推論を行い、異常値判定を行う。ここでは予測値と実測値とを比較し、一定以上の乖離がある場合に当該実測値を異常値であると判定する。

次に、空調消費電力の異常な変化が検知されたことを DC 運用者または DC 運用管理ソフトウェアに通知すると共に、予測モデルの学習を一時停止する(図 7 (6))。この間は最後に学習した予測モデルを用いて予測を継続する。

一定時間後、空調消費電力の異常な変化に対する是正対策が行われたか否かに基づき、予測モデルの学習再開の是非を判定する(図 7 (7))。是正対策が行われた場合は、アノマリを学習データから除外した上で予測モデルの学習を再開する。是正対策が行われなかった場合は、単に予測モデルの学習を再開する。これは、是正対策が行われなかったことから、当該変化は電力の傾向変化であってアノマリではないと見なすべきという考え方に基づいている。そのような傾向変化の例には、例えばサーバラックの増設による排熱増に対して空調の温度設定を見直した場合がある。

提案方式では以上に述べた処理を繰り返すことで予測モデルの学習と推論を制御する。なお、提案方式では空調消

費電力の形状判定処理(図 7 (1))を含めて繰り返し実行する。これは図 1 に示した例のように、同じサーバームのデータでも矩形的な変化をする期間とそうでない期間が変化することに对应するためである。

4. 評価

我々は自社 DC の実データを用いて、空調消費電力の矩形的な急変化への追従時間と、予測誤差の観点で提案方式の評価を行った。それぞれの評価結果を表 1, 表 2 に示す。

表 1: 消費電力の矩形的変化への追従時間

	Simple re-learning	Proposed Method
Server room #1	5.25	0.55
Server room #2	5.82	0.50
Server room #3	4.13	0.71
Server room #4	6.17	0.64
Server room #5	7.10	0.84
All rooms (average)	5.69	0.65

表 1 は、2 章で述べた実験方法と提案手法における空調消費電力の矩形的な急変化への追従時間を示している。表 1 に示した結果から、2 章で述べた実験方法では平均で約 6 日掛かるのに対し、提案手法では 1 日以内(0.65 日)で追従できたことがわかる。一例として、図 3 に示した結果に対応する提案手法の結果を図 11 に示す。この結果から、提案手法では空調消費電力の矩形的な急変化に対して予測値の線形補正により追従している様子が確認できる。これにより 2 章で述べた問題点 1 を解決した。

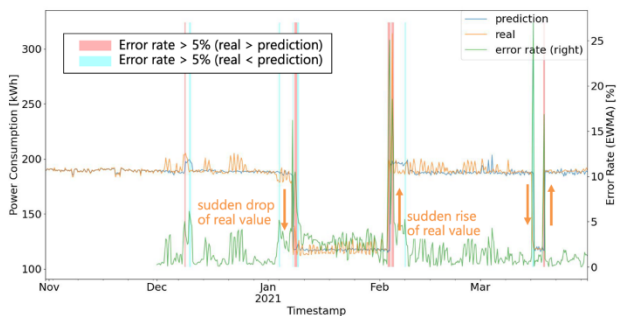


図 11: 矩形的変化への追従の例

また、矩形的な変化をするデータが矩形的な変化をしないデータに変化した場合の提案手法の結果の一例を図 12 に示す。この結果から、データの特徴が変化した場合でも提案手法により対応できている様子が確認できる。

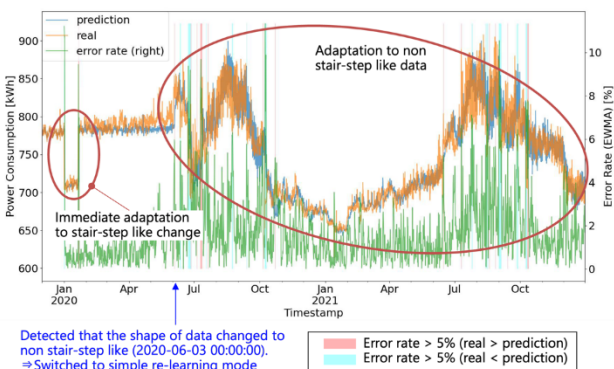


図 12: データ形状変化に対する追従の例

次に、予測誤差の評価結果を表 2 に示す。なお本評価では、図 7 の処理(6)(7)の実装を簡略化した。具体的には空調消費電力の異常変化を通知する処理と、異常変化に対する

是正対策が行われたか否かに基づいて予測モデルの学習再開の是非を判定する処理を省略した。代わりに空調消費電力の異常変化が起きている日時を予め選定し、当該期間を過ぎるとモデルの学習を再開するように処理(6)(7)を実装した(当該期間のデータは学習データから除外)。

表 2: 予測誤差

	Simple re-learning		Proposed Method	
	RMSPE (All periods)	RMSPE (Rapid change periods)	RMSPE (All periods)	RMSPE (Rapid change periods)
Server room #1	3.07	8.50	2.37	4.21
Server room #2	4.97	9.19	3.29	3.97
Server room #3	3.51	8.17	2.77	2.94
Server room #4	4.78	9.13	3.18	3.26
Server room #5	6.28	10.69	4.40	4.17
All rooms (average)	4.52	9.14	3.20	3.71

表 2 に示した結果から、矩形的な急変化が起きた期間における予測誤差率(RMSPE: Root Mean Square Percentage Error)は、2 章で述べた実験方法では平均 9.14%に対し、提案手法では平均 3.71%であった。実験期間全体(2年間)では前者が 4.52%、後者が 3.2%である。これは 2 年間の中で空調消費電力の矩形的な急変化のない期間が長いためである。

また、図 6 に示した結果に対し、提案手法でアノマリの学習回避を行った結果の一例を図 13 に示す。この結果から、空調消費電力のアノマリに対して提案手法では学習を回避した様子を確認することができる。これにより 2 章で述べた問題点 2 を解決した。

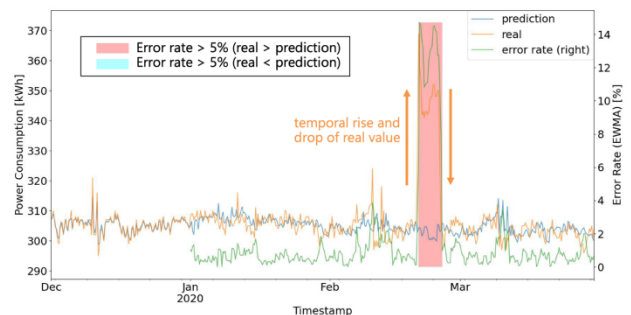


図 13: 一時的変化(アノマリ)の学習回避の例

5. おわりに

本研究では自社 DC の実データを用いた実験を通じて、消費電力の変動や、その傾向変化が予測モデルの精度に与える影響を明らかにした。そして消費電力の傾向変化に追従しつつ、非効率な電力変動の学習を回避して精度維持する ML モデル管理手法を提案した。実データを用いた評価により、空調消費電力の矩形的な急変化に対する短期間での追従と、予測誤差率の低減が可能であることを確認した。

参考文献

- [1] 国立研究開発法人科学技術振興機構 低炭素社会戦略センター, “情報化社会の進展がエネルギー消費に与える影響 (Vol.2)”, 2021.
- [2] Backhus Jana, Yasutaka Kono, “Cooling Power Consumption Dependency Simulation for Real-World Data Center Data”, International Conference on Software, Telecommunications and Computer Networks (SoftCOM), 2022.
- [3] A. Merizig, T. Bendahmane, S. Merzoug, and O. Kazar, “Machine learning approach for energy consumption prediction in datacenters”, International Conference on Mathematics and Information Technology (ICMIT), 2020.
- [4] G. Smpokos, M. A. Elshatshat, A. Lioumpas, and I. Iliopoulos, “On the energy consumption forecasting of data centers based on weather conditions: Remote sensing and machine learning approach”, International Symposium on Communication Systems, Networks & Digital Signal Processing (CSNDSP), 2018.