

物理リザバー・コンピューティングによる機械学習デバイスとエッジ・コンピューティングへの応用

近年、再帰的ニューラル・ネットワークの新たなアーキテクチャーとしてリザバー・コンピューティングが注目されています。リザバー・コンピューティングは、通常のニューラル・ネットワークと比較してモデルの構造が簡略化されているため、学習時や動作時（推論時）のコストが著しく低いという特徴があります。さらに、リザバー・コンピューティングは物理系のダイナミクスによっても実現できるため、ソフトウェア実装に比べて高速性や低消費電力性を持つ実装が可能です。本稿ではリザバー・コンピューティングが持つこれらの特徴を生かした機械学習デバイス、およびIoTエッジ・コンピューティングなどの次世代ITシステムへの応用について解説します。

▶▶ 1. はじめに

IT社会の進展に伴い、膨大な音声・画像・テキストといった構造化されていない、人間の認知に関わるビッグデータが指数的に増加しつつあります。特に、エッジ側のIoTデバイスが非構造化ビッグデータの主要な発生源になると予想されています。そこで、エッジの領域で、より一層の通信負荷の低減、認知データ処理能力の向上、エネルギー利用効率の向上を図り、認識、分類、予測などの高度な認知処理を高効率に行うエッジ・コグニティブ・コンピューティング技術の開発が必須となります。

一方、従来のクラウド側でのコグニティブ・コンピューティングにおいては、ノイマン型汎用プロセッサ上でのソフトウェア処理およびハードウェア・アクセラレーションが主流です。例えば、ディープ・ラーニングに代表される大規模な機械学習モデルをGPUによって動作させるアプローチが近年盛んに研究されていますが、このような実装では一般的に動作時の消費電力が大きくなることから、利用範囲がクラウド側に限定されます。さらにディープ・ラーニングはオフラインでの学習時にも極めて多数のパラメーター値の最適化を行うため、膨大な学習時間、消費電力を必要とし、エッジ側での運用には適していません。

そこで、IBM東京基礎研究所のチームではエッジ・コグニティブ・コンピューティングへの応用へ向けて、デ

バイス実装に非常に適した機械学習アルゴリズムであるリザバー・コンピューティングに基づき、超高効率動作とオンライン学習が可能なデバイスの開発を行っています。本稿ではこの研究を概観し、IoTエッジ・コンピューティングへの応用について解説します。

▶▶ 2. リザバー・コンピューティング

現在のところ、非構造化認知データの処理を計算機上で行うにあたって最も有力視されている手法が人工ニューラル・ネットワークです。ニューラル・ネットワークは、生物の脳の認知機能と構造に基礎を置いている機械学習モデルであり、その名のとおり神経細胞を数理的にモデル化した人工ニューロン素子がネットワークを構成しています。ニューラル・ネットワークをネットワークの構造で大別すると、常に後続の層に向かって一方向性の結合しか持たないフィード・フォワード・ニューラル・ネットワークと、自分自身へ再帰するような双方向性の結合パスを持つ再帰的ニューラル・ネットワークに分かれます。再帰的ニューラル・ネットワークはこの再帰的な構造により、時間的に相関を持つ時系列データを取り扱うことができます。

リザバー・コンピューティングとは2000年代初頭に提案された再帰的ニューラル・ネットワークのアーキテクチャーの一つで、**図1**に示すように、入力層、リザバー層、リードアウト層（出力層）の3つの部分から構成され

ます[1]。最大の特徴は、リザーバー層がランダムに初期化・固定化された結合重みをもつ高次元の再帰型ニューラル・ネットワークとして設計されるという点です。また、入力層もランダムに初期化し固定します。学習はリードアウト層のみで行われ、通常は比較的処理負荷の軽い線形適応フィルターによる信号処理技術が用いられます。リザーバー・コンピューティングは再帰的なニューラル・ネットワークの一種であるため、音声認識や時系列処理など時間的なダイナミクスを持つデータの処理が可能であり、いくつかの代表的なベンチマーク・タスクでは、従来手法とほぼ同程度の精度を示します。

リザーバー・コンピューティングは、神経科学の分野で大脳皮質の微小カラム構造のモデルとして提案されたLiquid state machine、および信号処理・工学の分野で時系列データ解析のモデルとして提案されたEcho state networkを端緒とします[1]。ニューラル・ネットワークとしてはLiquid state machineが連続時間、スパイクング・ニューロンモデルであるのに対して、Echo state networkは離散時間、レート・コーディング・ニューロンモデルに基づくという違いがありました。しかし、両者とも中間層の再帰的ネットワークの部分はランダムに固定され、出力層のみが学習されるという共通点があることが認識され、リザーバー・コンピューティングとして統一されたという経緯があります。

ここで、リザーバー・コンピューティングという名前の由来について補足しておきます。もともとリザーバー(reservoir)とは「貯水池、溜池」を意味します。比喩的に説明すると、「池に石を投げたとき水面にどのような波紋ができたかを見れば、投げられた石がどのような石

だったかが分かる」ということです。この比喩が示唆するように、リザーバー層の機能の本質は識別の難しい2つのパターンの差異を拡大し、識別しやすいパターンへと変換することです。これは機械学習の言葉で言い換えると、入力データを非線形変換によって高次元の特徴量空間(さらにはカーネル関数によって無限次元空間)へ写像し、線形分離の可能性を高める操作である、と言えます。機械学習においては個々の問題に応じてさまざまな特徴量が構成されますが、高次元性が本質であるならば、ある種の極論としてランダムかつ高次元に特徴量を構成するというアプローチも考えられます。そして、リザーバー層はまさにこの機能を担っているのです。

ただし、モデルの一部にランダムな構造を用いるという考え方自体は、実はそれほど新しいものではありません。実際、現在の人工ニューラル・ネットワークの源流となったRosenblattのパーセプトロンはS層(感覚層、入力層)、A層(連合層、中間層)、R層(反応層、出力層)の3つの層からなるモデルですが、S層とA層の間はランダムに接続されており、A層とR層の間の学習のみを考えます[2]。しかしながら、単純なパーセプトロンはモデルの表現能力に限界があり、線形分離不能であるXOR演算を実現できないことが人工知能研究者のMinsky氏とPapert氏によって指摘されると、パーセプトロンに対する関心は急速に失われました。その後のニューラル・ネットワークの研究は、モデルの表現力を高めるため多層のニューラル・ネットワークを考え、誤差逆伝播法などによりすべての結合重みを学習する方向へと進みました。さらに最近になって、事前学習やドロップアウトなどの多数の学習テクニックが導入されたため、多層化の傾向が急激

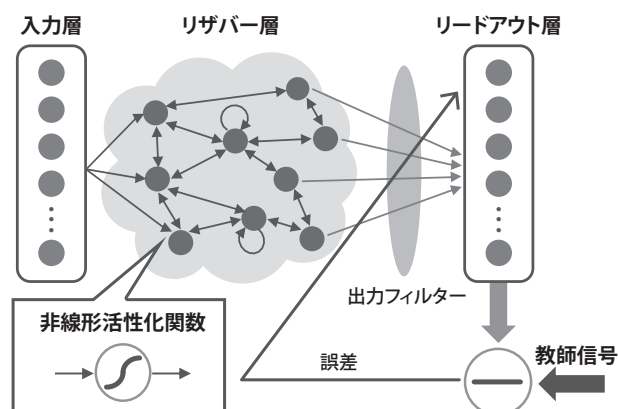


図1. リザーバー・コンピューティングの概念図

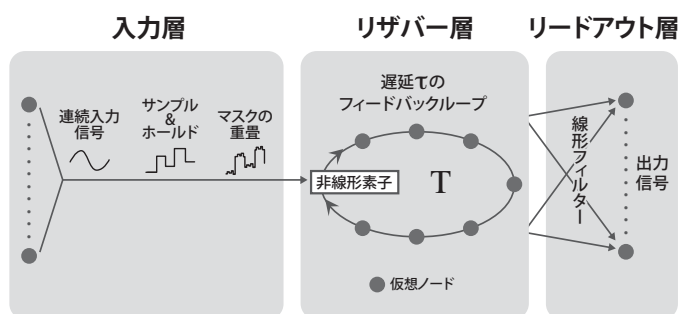


図2. 遅延フィードバック系による物理リザーバー・コンピューティング

に進みました。同時に大規模なモデルの学習に必要なビッグデータが容易に手に入る時代になり、現在のディープ・ニューラル・ネットワークの隆盛へとつながっています。しかし、多層ニューラル・ネットワークや全結合をもつ再帰的ニューラル・ネットワークは一般に非線形性をもつ複雑なモデルであるため、学習は大規模かつ非線形の最適化問題となり、局所最適解などの困難を抱えることが多くあります。そこで、複雑な最適化問題を避けるためにモデルの一部にランダムな構造を取り入れ、学習を単純化するタイプのニューラル・ネットワークが近年積極的に見直されつつあります[3]。リザーバー・コンピューティングも再帰的ニューラル・ネットワークの部分ランダムなネットワークに置き換えて簡略化したものであり、こうした流れの中にあると言えます。

近年のディープ・ラーニングに代表される極めて多数のパラメーターを最適化する方法に比べて、モデルを簡略化した分、プログラマビリティは失われることとなりますが、代わりに学習コストの低減や運用の容易性などの実用上のメリットが得られます。例えば、学習はリードアウト層における単純な(線形)識別器あるいは適応フィルタでのみ行われるため、システムの基本構成を変更することなく、広範なタスクへ展開することができます。ユースケースによっては、認識性能よりもこのようなメリットの方が重要である場合も多々あると考えられます。4章ではそのようなケースとして、エッジ・コンピューティングを取り上げます。

▶▶ 3. 物理系によるリザーバー・コンピューティング

従来のノイマン型プロセッサ上でニューラル・ネットワークをソフトウェア実装によって実現すると、十分な処理効率が得られない場合があります。この問題に対し、ニューロモルフィック・デバイスと呼ばれるニューラル・ネットワークの実行に特化したデバイスがいくつか提案されています。ニューロモルフィック・デバイスは脳の構造を模倣した非ノイマン型のアーキテクチャーが特徴であり、「IBM TrueNorth」が有名です。同様に、リザーバー・コンピューティングも従来のニューロモルフィック・デバイスやFPGA(Field Programmable Gate Array)などのハードウェアによって実装すること

ができます。ハードウェア実装の観点からリザーバー・コンピューティングを見た場合の特徴を以下に記載します。

- シンプルな構成とシンプルな学習アルゴリズムであるため低消費電力での動作が可能。
- オンライン学習や時間制約のあるリアルタイム処理が可能。
- 結合重み行列がスパースかつ固定されているため、実装の観点からは配線が少なく済み、低消費電力性につながる。

さらに近年、リザーバー・コンピューティングはニューラル・ネットワークとは直接関係のない、高次元の非線形物理系のダイナミクスによっても実現可能であることが分かってきました。例えば、遅延フィードバックを受けた非線形素子(具体的にはMackey-Glass振動子を実装したアナログ電気回路)をリザーバー層として用いる物理リザーバーが提案されています[4]。図2に示すとおり、この方式では入力信号にランダムなマスク信号を重畳し、非線形素子に入力します。さらにリードアウト層においては遅延フィードバック・ループ上に等間隔に配置された観測点を仮想的なニューロンと見なし、それらの観測値に対し線形フィルタを適用しています。一般に非線形の物理系は動作パラメーターの値によってその振る舞いをさまざまにコントロールできますが、ある程度の高次元性と複雑な挙動を実現できればリザーバーとして機能する可能性があります。このように物理系を利用してリザーバー・コンピューティングを実現したデバイスを物理リザーバー・デバイスと呼びます。一般的に機械学習において、高次元特徴空間への非線形マッピングやニューロンの非線形演算は計算負荷が大きい処理ですが、リザーバー層を物理ダイナミクスとして実現すれば、これらの負荷の大きい処理を物理系のダイナミクスそのものに委ねることができます。従って、アルゴリズムの観点だけでなく、実装の観点からも高速性や低消費電力性を得ることができます。

IBM東京基礎研究所と東京大学社会連携講座のグループでは、物理ダイナミクスのうち特に波動現象を用いた物理リザーバー・デバイスとして、VCSELレーザー系や、ガーネット磁性薄膜上でのスピン波のダイナミクスに基づいた物理リザーバー・デバイス構造を提案しました[5][6]。図3にこれらの波動物理リザーバー・デバイスの構造と物理リ

ザバー・コンピューティングの全体図を示します。波動現象は物理現象として極めて一般的であるため、既存の技術を用いて実現できます。さらに、波動現象は周期的運動からカオスに至るまで多様な挙動を示すことが知られており、物理リザーバ・デバイスの実現に適しています。また、波同士の相互作用がニューラル・ネットワークとしての非線形演算を実現しているため、内部に物理的な配線が不要な点も魅力です。このように波動現象を用いた物理リザーバ・デバイスは前章で説明した「リザーバ＝溜池」の比喩を物理的に実現していると言えるでしょう。

また最近、これらの例を含んださまざまな物理系を用いたリザーバ・コンピューティングの包括的なレビューが出版されました[7][8]。

物理リザーバ・コンピューティングと従来の大規模機械学習手法との比較を表1にまとめました。

4. IoTエッジ・コンピューティングへの応用

1章で述べたとおり、今後は人間の認知に関わるデータがビッグデータの大きな割合を占めるため、その効率的な処理が今後のIT技術の大きな課題となります。特に2020年にはビッグデータの40%がセンサーで発生すると予想されています[9]。そのため、従来のクラウド型の「端末機器で取得したデータを物理的に遠く離れたデータセンターに転送して、処理した情報を端末機器へとフィードバックする」というアプローチは、遅延時間、ネットワーク負荷およびサーバー側での電力消費の観点から困難になります(図4左)。

そこで、従来クラウド側で行われていた認知処理の一部をデータ発生源に近いエッジ側において高速に低消費電力で処理するエッジ・コンピューティングが必須となります(図4真ん中)。このような背景のもと現在エッジ・コンピューティングの研究が盛んに行われていますが、主なアプローチは以下のようにまとめられます。

- エッジ側(ルーター、アクセス・ポイント、ゲートウェイの周辺)にエッジ・サーバーを設置する。
- センサーやエッジ・サーバーにおいて、ニューラル・ネットワークなどの機械学習アルゴリズムをソフトウェアによって実装する。

しかし、エッジ側においては厳しい電力制約下での動作や高速処理などの高効率性、リアルタイム性などクラウド側とは大きく異なる性能要求があります。従って、従来の汎用プロセッサ・ベースでのソフトウェア処理ではデータ処理能力や効率性が十分でなく、適用範囲が限定されてしまうという問題があります。

これに対して本提案の物理リザーバ・コンピューティングはIoTセンサーと一体化した実装が可能であり、物理ダイナミクスを計算過程として利用することから、従来のエッジ・コンピューティングと比較してさらなる効率化が達成できます。その結果、センサー・デバイスが生み出すビッグデータをIoTデバイス側で十分に縮約した後にクラウド側へ送信することが可能になり、ネットワークやサーバーへの負荷の低減と、IoTプラットフォーム全体の高速性と低消費電力性とを両立させたエッジ・コグニティ

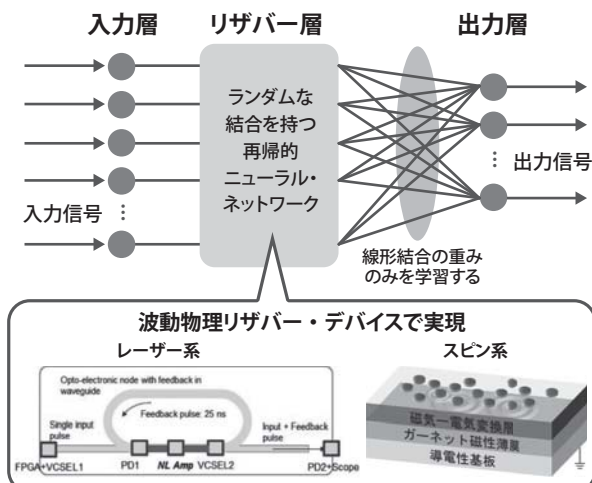


図3. 波動物理デバイスによる物理リザーバ・コンピューティングの概念図

表1. 従来の機械学習を実行するGPU、ニューロモルフィック・デバイスと本研究対象の物理リザーバ・コンピューティングとの比較表

項目	従来の機械学習 (Deep Learningなど) + GPU, ニューロモルフィック・デバイス	物理リザーバ・コンピューティング (リザーバ・コンピューティング + 物理リザーバ・デバイス)
ネットワークトポロジー	使用者が自由にソフトウェア上で設計	トポロジーはスパースで固定
ニューロン素子間配線	すべてのニューロン素子間が双方向でつながる必要があり、ニューロン数が増大すると配線爆発を起こす。	(波動)物理ダイナミクスによる相互作用を用いるため配線はないか、または非常に少ない。
学習にかかる時間	パラメーター数が膨大なために学習に時間がかかる。	出力層のパラメーターのみを学習すればよいため高速。
オンライン学習	オンライン学習できるデバイスはなく、オフラインでのソフトウェアによる学習に頼る。	学習アルゴリズムがシンプルかつリザーバ層と分離しているため、オンライン学習可能。
消費電力	学習はオフラインで、別途ソフトウェアにより行うため、パラメーター数が膨大になると消費電力が大きくなる。また、推論時の消費電力も高い。	物理ダイナミクスを直接計算過程として用いるため、消費電力を抑えられる。リードアウト層は単純な線形適応フィルターであるため、学習時・動作時ともに消費電力は低い。

ブ・コンピューティングを実現できます (図4右)。

5. まとめと将来展望

本稿では、再帰的ニューラル・ネットワークのアーキテクチャーの一つとして注目されているリザーバー・コンピューティングについて紹介しました。リザーバー・コンピューティングはランダムな再帰的ネットワークであるリザーバー層とシンプルな適応的信号処理を行うリードアウト層からなるため、従来の大規模機械学習システムに比べて低コストでの運用が可能です。さらに物理ダイナミクスを用いてリザーバー・コンピューティングを実現することで、さらなる高効率化を図ることができます。これらの特性を生かしIoTエッジ・コンピューティングの分野において、デバイスとの一体化によるセンサーの知能化、エッジ・コグニティブ・コンピューティングの実現が期待できます。

【参考文献】

- [1] Lukoševičius, M., Jaeger, H.: Reservoir computing approaches to recurrent neural network training. *Computer Science Review* 3(3), pp. 127-149 (2009)
- [2] F. Rosenblatt : The perceptron : a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386, 1958.
- [3] Scardapane, Simone & Wang, Dianhui : Randomness in Neural Networks: An Overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 7. 10.1002/widm.1200. 2017.
- [4] L. Appeltant et al, "Information processing using a single dynamical node as complex system", *Nature Comm.* 2011.
- [5] J. B. Héroux, H. Numata, Naoki Kanazawa, D. Nakano : Optoelectronic Reservoir Computing with VCSEL. *IJCNN*, 2018.
- [6] RYOSHO NAKANE , GOUHEI TANAKA, AND AKIRA HIROSE : Reservoir Computing With Spin Waves Excited in a Garnet Film, *IEEE Access*, 2018. <http://ieeexplore.ieee.org/document/8262616/>
- [7] G. Tanaka et. al : Recent Advances in Physical Reservoir Computing: A Review, *Neural Network Journal*, Elsevier (2019), <https://doi.org/10.1016/j.neunet.2019.03.005>
- [8] 中野大樹他編 : 小特集 リザーバー・コンピューティング, 電子情報通信学会誌, 2019年2月号
- [9] Cisco Global Cloud Index 2015-2020

謝辞

本研究は東京大学 大学院 工学系研究科 国際工学教育推進機構 社会連携講座「省エネルギー情報処理 (Energy Efficient Information Processing)」のもとで行われました。また、この論文は国立研究開発法人新エネルギー・産業技術総合開発機構 (NEDO) の委託業務 (契約管理番号: 18102284-0) として執筆されました。



日本アイ・ピー・エム株式会社
東京基礎研究所
主任研究員

山根 敏志
Toshiyuki Yamane

2000年日本IBM入社。東京基礎研究所に入所後、光通信向け誤り訂正符号、パフォーマンス・エンジニアリング、ミリ波帯高速無線通信などの研究開発を経て、現在は次世代コグニティブ・コンピューティングや機械学習による新化学構造物の探求などの研究開発に従事。

【執筆協力】

日本アイ・ピー・エム株式会社
東京基礎研究所
沼田 英俊 Hidetoshi Numata

日本アイ・ピー・エム株式会社
東京基礎研究所
ヘロー ジャンベノ Jean Benoit Héroux

日本アイ・ピー・エム株式会社
東京基礎研究所
武田 征士 Seiji Takeda

日本アイ・ピー・エム株式会社
東京基礎研究所
金澤 直輝 Naoki Kanazawa

日本アイ・ピー・エム株式会社
東京基礎研究所
中野 大樹 Daiju Nakano

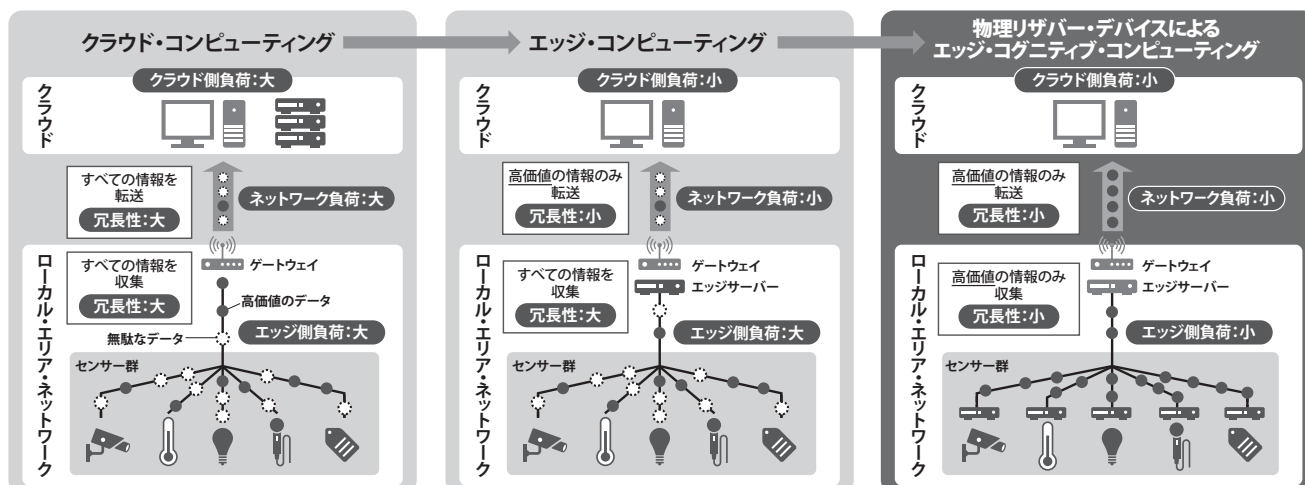


図4. クラウド・コンピューティングから、物理リザーバー・デバイスを用いたエッジ・コグニティブ・コンピューティングへ