Press Release

千葉工業大学 CHIBA INSTITUTE OF TECHNOLOGY

IRCN International Research Center for Neurointelligence wpi

研究成果 プロジェクト

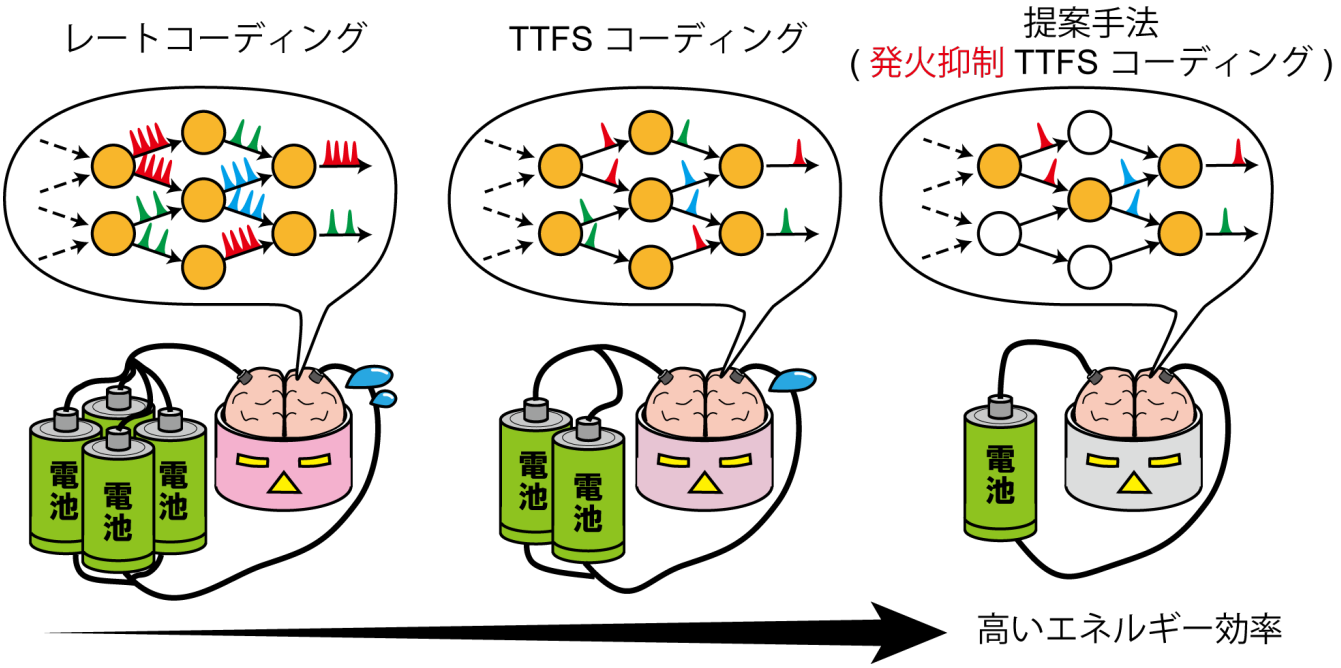
2024 年 1 月 17 日

報道機関　各位

千葉工業大学 数理工学研究センター

時間符号スパイキングニューラルネットワークの発火頻度を低減した学習に成功

キーワード：スパイキングニューラルネットワーク、テンポラルコーディング、エッジ AI



［ 発表者 ］

・酒見 悠介（千葉工業大学 数理工学研究センター 上席研究員／東京大学国際高等研究所ニューロインテリジェンス国際研究機構（WPI-IRCN） 連携研究者）

・山本 かけい（マサチューセッツ工科大学）

・細見 岳生（NEC）

・合原 一幸（東京大学 特別教授・名誉教授／東京大学国際高等研究所ニューロインテリジェンス国際研究機構（WPI-IRCN） 副機構長／千葉工業大学 数理工学研究センター 主席研究員）

[概要]

酒見悠介（千葉工大)、山本かけい（MIT)、細見岳生（NEC）、合原一幸（東大)は、脳の情報処理機構を模倣するスパイキングニューラルネットワーク（SNN)※1において、予測精度を保ちながらニューロンの発火頻度を低減する手法を開発した。特に、テンポラルコーディングと呼ばれる発火頻度が極めて少ない情報処理機構において、発火頻度を更に半減させることを実証しました。発火頻度の低減は消費エネルギーの低減をもたらすため、本技術は低電力性が求められるエッジ AI※2 において今後重要なものになると考えられます。この成果は、2023 年 12 月 21 日に査読付き国際学術雑誌「Scientific Reports」で公開されました。

■ 背景

スパイキングニューラルネットワーク（SNN）※1はスパイク信号による情報処理が可能な、脳を模倣した人工知能モデルです。SNN は専用ハードウェア化を行うことで既存の深層学習モデルよりも高いエネルギー効率を達成することが可能であるため、エッジ AI※2 への応用が期待されています。

SNN の専用ハードウェアにおいては、主としてスパイクの生成（発火)によって電力が消費されるために、少ないスパイク数（発火頻度）によって情報処理を行うことが重要です。これを実現するものとして、スパイクの時間情報を活用したテンポラルコーディングによる学習アルゴリズムが開発されてきました。その中でも、Time-to-First-Spike Coding (TTFS)と呼ばれるコーディングは、1 ニューロンあたり、最大 1 回までしか発火しないという制約を課すことで、高い学習性能と高いエネルギー効率を両立できることが知られています。しかし、この制約を越えて発火が抑制可能かどうかは、ほとんど調べられていませんでした。

■ 内容

本研究では、TTFS コーディングされた SNN (TTFS-SNN)を更に発火抑制するための学習手法を開発しました。とくに、発火イベントに着目した発火抑制手法であることから ***SSR (Spike-Timing-Based Sparse-Firing Regularization)***と名付けました。一般的な教師あり学習で用いられるコスト関数に、SSR 正則化項を加えることで、発火を抑えつつ、データセットを学習させることができます。本研究では、異なる観点で二つの SSR 手法、M-SSR と F-SSR を導出しました。図 1 にそれぞれの導出方法の概要を示しています。重要なのは、どちらの正則化関数も、発火時刻およびそれに関連した重みの情報のみを用いる点です。この特徴により、発火現象を直接的に抑制させることが可能になっています。

図 2 に、M-SSR を導入した時の学習結果の様子を示しています。正則化強度が強いほど、中間層の発火は低減されていますが、出力層の発火時刻に大きな変化はないことがわかります。これは、発火を低減しても、予測が可能なことを示しています。このことをより詳細に調べる

ために、SSR 正則化強度を変化させた場合の、発火率と予測精度のトレードオフについて調べました。図 3 に示すように、SSR 正則化を導入しない場合には、ネットワーク全体での発火は 1 ニューロン当たり平均 0.5 回でしたが、正則化を導入することで、1 ニューロン当たり平均 0.2 回程度まで予測精度を著しく劣化させずに低減させることがわかります。

なお、SNN の発火を低減させる研究にはいくつか先行研究が存在しています。しかし、それらの研究は、スパイクの発生頻度に情報が込められるレートコーディングを基にしており、また離散時刻系の SNN に限定されていました。これらの手法は、膜電位を低減しているとみなすことが出来るので、本研究における M-SSR と同様の発想です。しかし、M-SSR では、極限操作により、膜電位を直接的に扱わない時刻型の正則化関数の導出に成功しました。これは、本研究が連続時間系の SNN をもとに学習アルゴリズムを構築したことによって初めて可能になっております。さらに、膜電位を低減する従来手法に比べて、M-SSR はより優れた発火率―予測精度トレードオフ特性を示すことができました(図 4)。

■　まとめと展望

エッジ AI のように高いエネルギー効率が求められる場合には、エネルギー効率を最適化した人工知能モデルが必要になります。一般的な深層学習で用いられている人工ニューラルネットワーク(ANN)において、そのような技術はネットワークの軽量化技術として発展しており、重み量子化、枝刈り、蒸留などの技術が知られています。今回開発した、発火頻度を抑制する手法は、SNN に特有のネットワーク軽量化技術と考えることができます。また重要なことは、この軽量化技術は SNN の専用ハードウェアにおいてのみ有効になる点です。私たちは、アルゴリズム研究だけでなく、ハードウェア研究にも取り組んでいるため、このような新しい軽量化技術にいち早く着手することができました。GPU などのデジタルハードウェア上で効率的に動作させる人工知能モデルの研究は成熟しつつありますが、アナログハードウェア上で効率的に動作させる人工知能モデルの研究は発展途上にあります。今後は、SNN のアルゴリズム・モデルの研究に加え、アナログハードウェア自体の開発にも取り組み、エッジ AI の実装を目指していきます。

※1）スパイキングニューラルネットワーク（SNN）
スパイキングニューロンによって構築されるネットワークであり、脳により近い特性を持つ、スパイクとよばれる二値の短パルス信号による情報処理を行うことができます。スパイキングニューロンは、入力スパイクに依存して膜電位が時間変化し、膜電位が発火閾値を超えるとスパイクを発生し、同時に膜電位がリセットされます。発生したスパイクは、接続された他のニューロンへと伝達されます。

※2）エッジ AI
車載システムやロボット上での AI 動作は低レイテンシ性が重要であるため、センサーから得られたデータをその場で処理することが求められ、それを実現する技術をエッジ AI と呼びます。エッジ AI の技術課題の一つとして高い電力効率が挙げられ、人工知能モデルの軽量化や、専用ハードウェア化が取り組まれています。

■ 論文情報
論文誌: Scientific Reports
論文題目: Sparse-Firing Regularization Methods for Spiking Neural Networks with Time-to-First-Spike Coding
著者: Yusuke Sakemi、Kakei Yamamoto、Takeo Hosomi、and Kazuyuki Aihara
URL: https://www.nature.com/articles/s41598-023-50201-5
DOI: 10.1038/s41598-023-50201-5
出版日: 2023 年 12 月 21 日

■ 謝辞
本研究の一部は、JST さきがけ JPMJPR22C5、セコム科学技術振興財団、日本電気株式会社、JST Moonshot R&D Grant Number JPMJMS2021、AMED under Grant Number JP21dm0307009、the International Research Center for Neurointelligence (WPI-IRCN) at The University of Tokyo Institutes for Advanced Study (UTIAS)、JSPS KAKENHI Grant Number JP20H05921 から助成を受けて行われました。

＜お問い合わせ先＞

【本研究内容に関する問い合わせ先】
千葉工業大学　数理工学研究センター　上席研究員
酒見　悠介
TEL: 047-478-0345
E-mail: yusuke.sakemi@p.chibakoudai.jp
【取材・大学広報関連に関する問い合わせ先】
千葉工業大学　入試広報部
大橋　慶子
TEL: 047-478-0222
E-mail: ohhashi.keiko@it-chiba.ac.jp

■ 添付資料

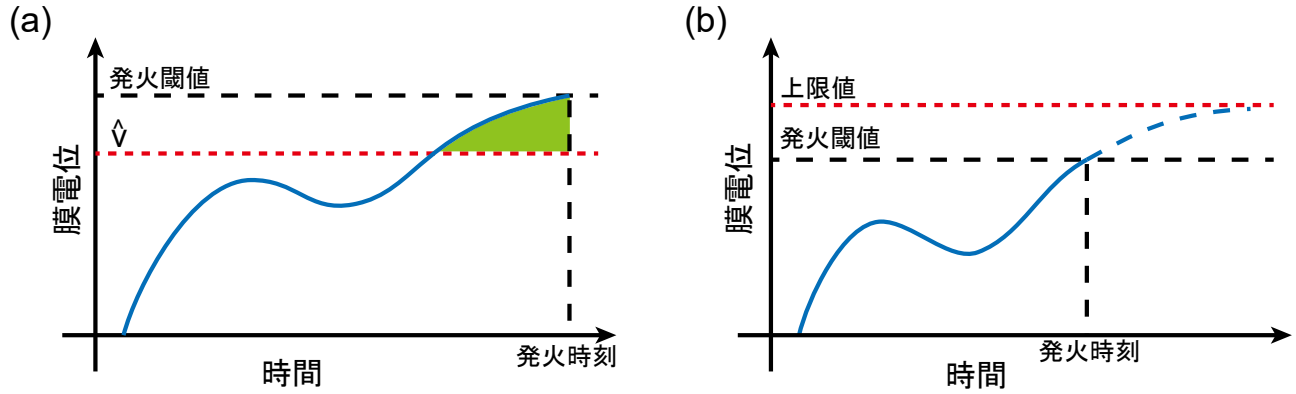


図 1: SSR 正則化の導出

(a) M-SSR 正則化は、膜電位がある電圧 $\hat{v}$ よりも大きい場合に損失を発生させる（緑色の領域)。そして、電圧 $\hat{v}$を発火閾値(firing threshold voltage)と一致させることで、発火時刻(firing time)のみに依存した正則化関数が得られる。 (b) F-SSR は、膜電位が発火後もリセットされずに時間発展した場合の上限値(upper limit)を正則化関数とする。

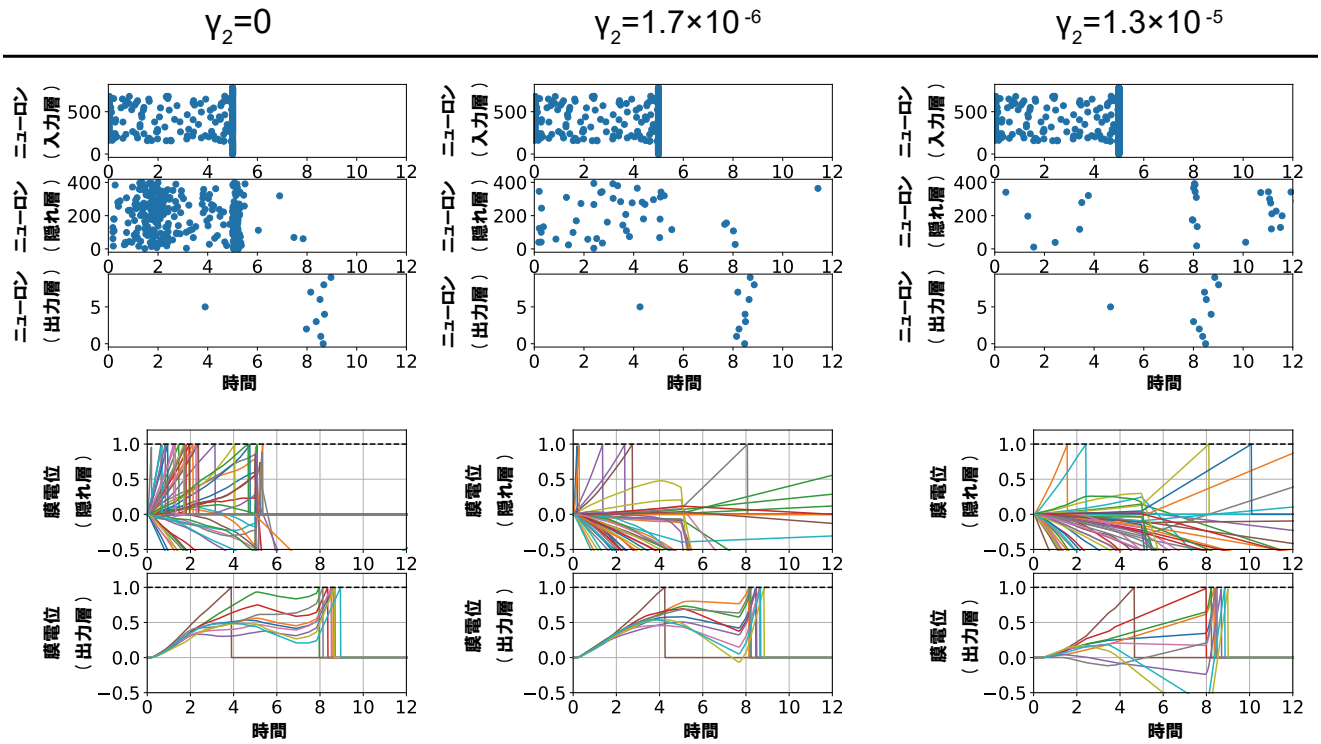


図 2: 発火抑制効果の典型例

隠れ層が 1 層のみの SNN(784-400-10)において M-SSR 正則化の強度 γ_2 を変化させた時の学習結果の変化を示している。上図は各層の発火時系列（ラスタープロット)を示しており、下図は各層の膜電位の時間発展を示している。正則化強度が強いと（$\gamma_2 = 1.3 \times 10^{-5}$)、中間層ニューロンは発火が強く抑制されるが、出力層ニューロンの発火分布に大きな変化がないことがわかる。

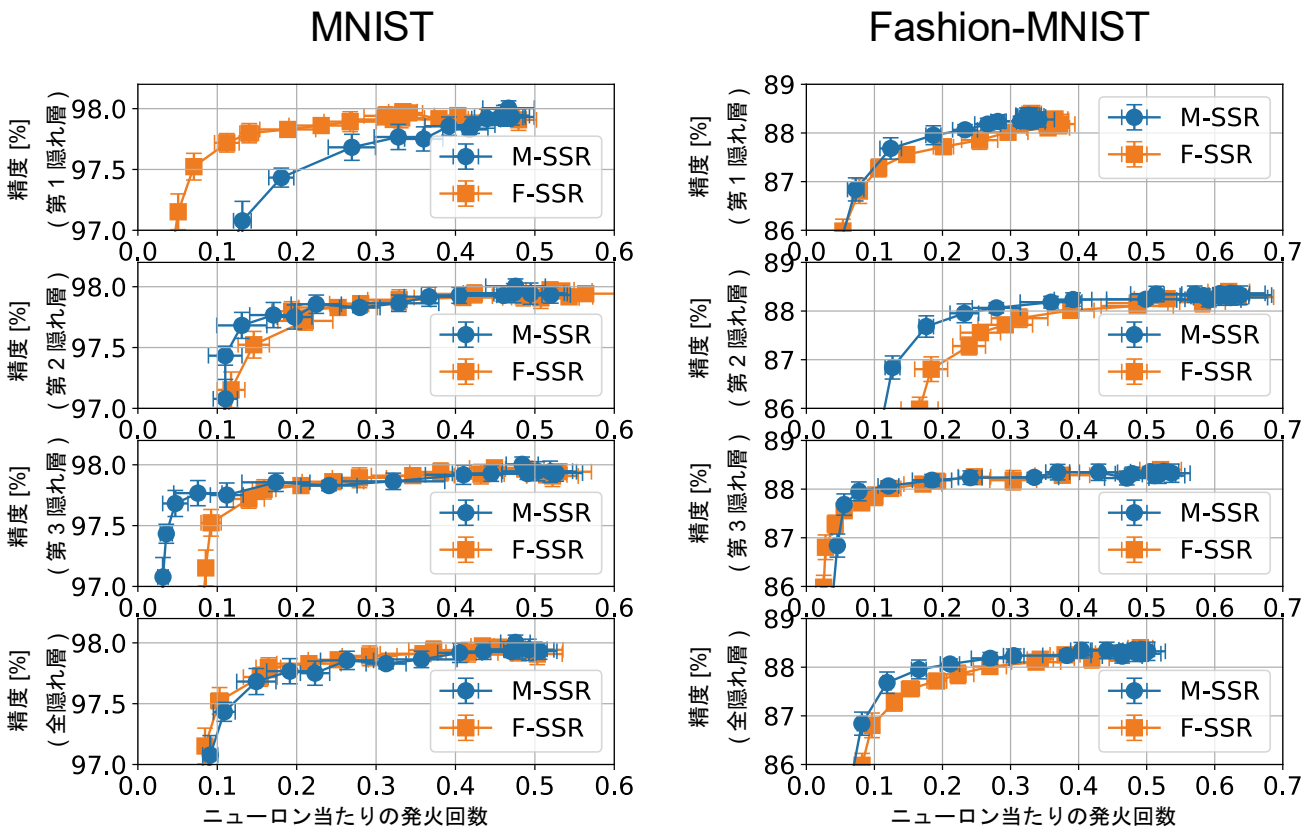


図 3: SSR 正則化による発火率と認識精度のトレードオフ

中間層を 3 層含む SNN(784-400-400-400-10)において SSR 正則化を導入した時の発火率（横軸）と認識精度（縦軸)のトレードオフの関係性を示しています。左図は MNIST データセットを用いたときであり、右図は Fashion-MNIST データセットを用いたときの結果です。各図において、上から、第一隠れ層、第二隠れ層、第三隠れ層の結果を示しており、一番下のパネルは、発火頻度を全隠れ層で平均した結果を示しています。SSR 正則化強度を大きくすることで、各パネルにおいて右上の点から、左下の点へと推移していくトレードオフ関係が得られました。M-SSR および F-SSR の場合も、正則化を導入することで発火頻度が 0.5 程度から、0.2 程度へと、認識精度をほとんど損なわずに抑制できていることがわかります。

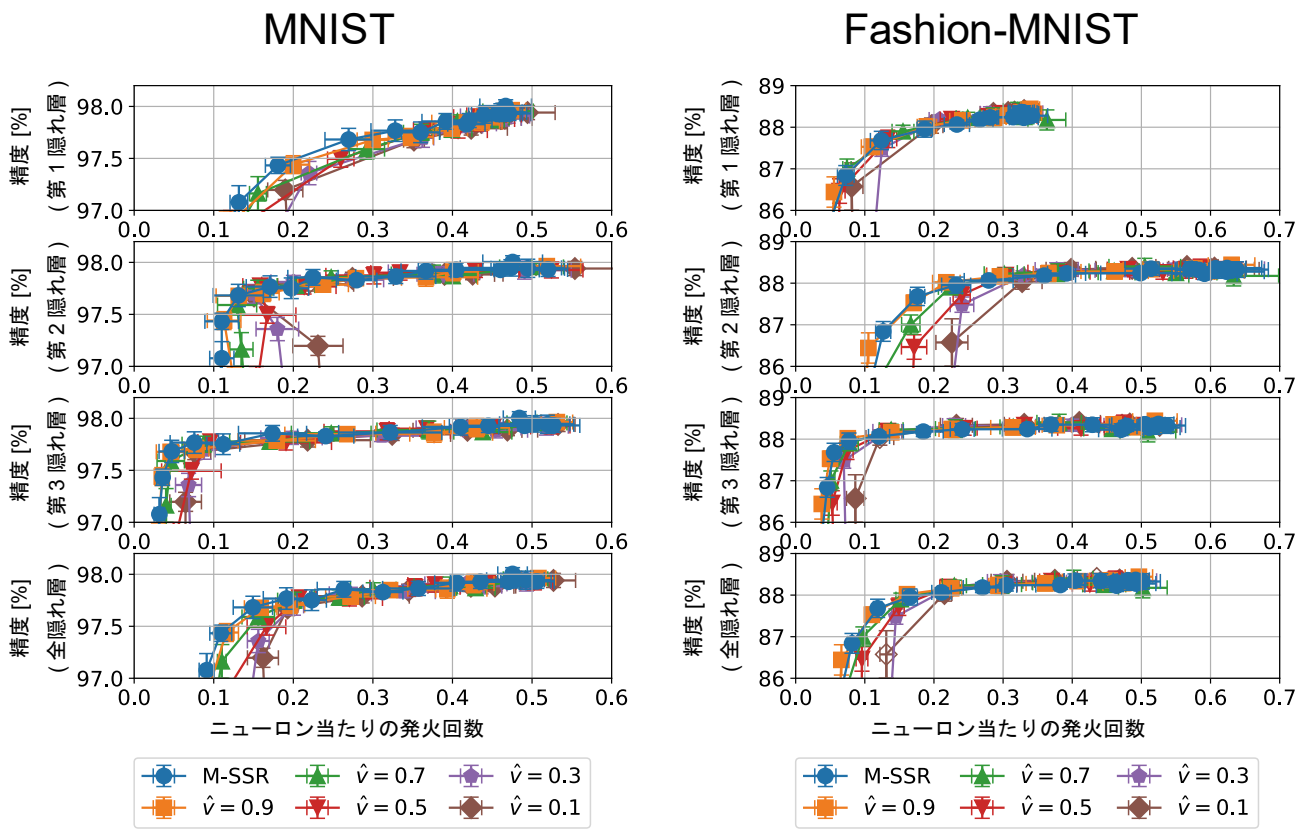


図 4: M-SSR 正則化と積分損失による正則化の比較

中間層を 3 層含む SNN(784-400-400-400-10)において膜電位に損失を与える正則化項と M-SSR の発火率―予測精度トレードオフ特性の比較。膜電位損失の正則化項は様々な参照電圧$\hat{v}$の場合の結果を示している。$\hat{v}$が大きいほどトレードオフ特性が良くなり M-SSR($\hat{v} = V_{\mathrm{th}} = 1$に相当、図 1 参照)が最もよいことがわかる。