

素粒子物理学実験への機械学習の適用研究

岩崎 昌子（大阪公立大学）

概要

本研究課題は、大型加速器を用いた素粒子・原子核物理実験に機械学習を適用し、実験効率・実験性能を向上させ、測定精度を高めることを目的としている。研究テーマとして、1) 機械学習を用いた加速器制御、2) 機械学習を用いたデータ圧縮、および、データ較正・再構成を行った。

テーマ1では、環境駆動型機械学習（強化学習）による加速器制御のために必要となる、加速器シミュレーターの開発を行った。敵対的生成型ニューラルネットワーク (Generative Adversarial Networks, GAN) を用いた加速器シミュレーターを開発し、モード崩壊による問題を解決するには、変分オートエンコーダ (Variational Autoencoder, VAE) を用いた次元削減の導入が有効であることを示した。テーマ2では、スパースセンシング技術を基盤とした、データ処理（データ削減）技術の開発を行った。具体的なアプリケーションとして、ILC 実験用電磁カロリメータの測定器データを削減し、エネルギー較正性能評価を行った。機械学習を用いてセンシングパターンを最適化することで、データが持つ本質的な情報を保持し、エネルギー較正性能の劣化を低減させることが可能であることを示した。

以上、本研究により、重要な知見を得ることができた。

プログラム開発：加藤，井上，度会

1 共同研究に関する情報

1.1 共同研究を実施した拠点名

- 大阪大学 サイバーメディアセンター

1.2 課題分野

- データ科学・データ利活用課題分野

1.3 共同研究分野 (HPCI 資源利用課題のみ)

- 超大規模数値計算系応用分野
- 超大規模データ処理系応用分野

1.4 参加研究者の役割分担

研究統括：岩崎

機械学習：長原，中島，武村

素粒子実験：末原，山田

原子核実験：中野

2 研究の目的と意義

本研究の目的は、大型加速器を用いた素粒子・原子核物理実験に機械学習を適用し、実験効率・実験性能を向上させ、測定精度を高めることである。機械学習では、明確なモデルや事象の関連性が特定できなくても、学習によって入出力の関係がモデル化できる。したがって、非線形な応答を高速に、高精度で得ることができる。大型加速器実験の大量データ処理プロセスに情報分野における最先端機械学習を導入することで、実験データ処理性能の飛躍的な向上が期待できる。

大型加速器を用いた素粒子物理学実験では、

1) 加速器によりビーム衝突実験を行い、実験で生成された大量の粒子を測定器で測定する、2) 測定された膨大量の実験データを蓄積する、3) 蓄積された測定データを、粒子のエネルギー情報、位置情報へ変換するための較正・再構成を行う、4) 較正・再構成された実験データから、データ解析により極微な信号事象を抽出、を行い、抽出された信号事象を用いて物理パラメータの測定を行う。

素粒子実験では、これまでに、項目4)で述べられた、データ解析における信号事象抽出への機械学習適用開発が行われてきた。本研究では、従来の素粒子実験ではあまり例がみられない1)、2)、3)における機械学習の適応研究を行う。具体的には、以下の開発を行う：

1. 機械学習を用いた加速器制御
2. 機械学習を用いたデータ圧縮、および、データ較正・再構成

テーマ1は環境駆動型機械学習（強化学習）および、敵対生成型ニューラルネットワーク (Generative Adversarial Networks, GAN) を、テーマ2はスパースセンシング技術を基盤とした、研究開発である。

これらの機械学習適用研究を、素粒子・原子核実験研究者と情報科学研究者による学際共同研究として遂行する。

最先端機械学習技術を基盤とした先駆的な研究を行うことで、加速器実験におけるデータ処理技術を大きく進展させ、加速器実験と関連分野への多大な波及効果を与えることを目指す。

3 当拠点公募型研究として実施した意義

本研究課題は、素粒子物理学分野における新原理・新物理探索のための重要な実験アプローチ、大型加速器による超精密実験のための基盤

技術研究である。

加速器実験では、膨大量実験データを蓄積し、超精密測定を行うことで、標準理論を超える、より根源的な新原理を探求する。実験で捉えた信号事象数で物理感度が決まるので、1. 加速器の強度を高めて高頻度に信号事象を生成すること、2. 生成された信号事象を精度良く測定し、膨大な実験データを効率良く収集すること、3. 実験データの較正・再構成、物理解析を高精度で行うことが重要である。

そのため、大型加速器実験では、実験を高度化、巨大化し、それに伴ってビッグデータの収集・処理と解析が、重要な研究基盤となった。例えば、高エネルギー加速器研究機構 (KEK) における B ファクトリー実験では、年間数十ペタバイトの膨大量データを生成する。

このような超巨大データを扱うためには、情報分野との協業が必須である。さらに、大型加速器実験のための基盤研究を遂行するためには、大量高速計算が必要不可欠である。特に本研究では、加速器実験による生データ (low-level data) を直接入力し、深層学習を行っているので、膨大な入力パラメータ（数千パラメータ）を取り扱う必要がある。

大学研究室レベルの計算資源では、以上のよう大量高速計算を用いた開発を進めることが非常に困難である。拠点公募型共同研究として実施することで、膨大量計算が可能になり、本研究を遂行することが出来る。

4 前年度までに得られた研究成果の概要

5 今年度の研究成果の詳細

以下に、研究成果を示す。これらの研究成果は、国内学会、研究会、国際研究会で発表し、さらに修士論文としてまとめた。

【機械学習を用いた加速器制御】

本研究では、KEK 入射加速器 (Linac) における、機械学習を用いた加速器制御を開発する。機械学習を導入した加速器制御によって、人では到達できない高精度・高効率で安定な加速器運転の実現が期待される。

大型加速器は、構成する機器の点数が多く、複雑な運転システムの制御が必要である。KEK Linac 加速器の場合、約 1,000 点の運転パラメータを制御している。さらに、周囲の温度変化や振動、潮汐力等の影響を受けるため、常に高精度な運転調整が必要である。

複雑な運転システムをリアルタイムに制御するため、我々は、1. 強化学習による加速器コントローラーと、2. 敵対的生成型ニューラルネットワーク (Generative Adversarial Networks, GAN) による加速器シミュレータを用いた、新しい加速器制御手法を考案した。図 1 に、概念図を示す。

強化学習は、車の自動運転等、周囲の環境変化に応じて、最大の報酬が得られるように行動を決定する機械学習である。加速器制御の場合、入射効率やビーム輝度を、報酬の指標とする。強化学習では、実際の行動を決定する前に、高精度なシミュレータを用いた学習 (pre-training) が重要である。

強化学習におけるパラメータ決定 (pre-training) を行うために、我々は、GAN を用

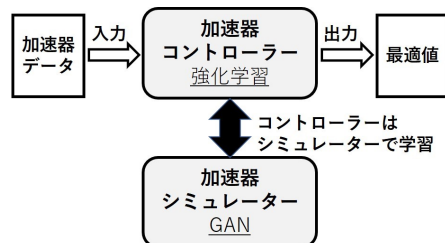


図 1 機械学習による加速器制御の概念図。

いた加速器運転シミュレータの開発を行った。

GAN は与えられたデータを模倣する機械学習である。生成されたデータと学習データの区別がつかないように学習を行うことで、実際の学習データの分布に近いデータを生成させる。約 1,000 パラメータで構成される加速器シミュレータの構築は、非常に挑戦的である。

我々の先行研究では、KEK Linac 加速器実データをもとに、GAN で疑似データ生成を行った場合、疑似データが、実データの一部区間しか再現できない、モード崩壊 (Mode collapse) と呼ばれる問題が起こった。

本研究では、モード崩壊を抑制するために、我々は 1) 変分オートエンコーダー (Auto-Encoding Variational Bayes, VAE) でのエンコーダを用いて加速器データを次元削減し、2)

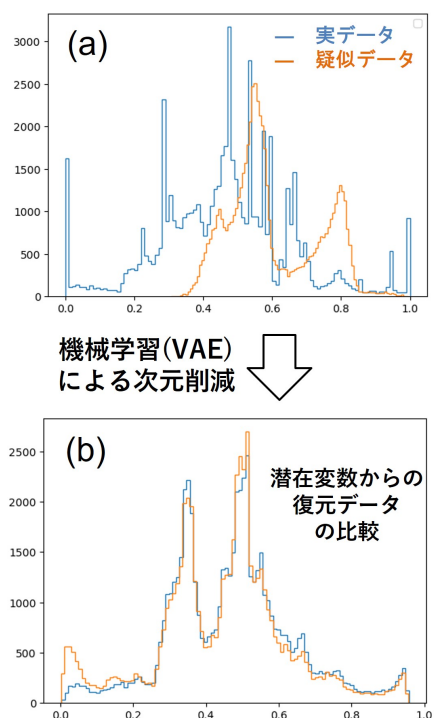


図 2 実データ (青), および GAN が生成した疑似データ (橙) による、ステアリング磁石パラメータを示す。VAE による次元削減を行うことで、モード崩壊が抑制されている。

GAN を用いて、次元削減された加速器データ (潜在変数) の疑似データを生成し、3) デコーダを用いて疑似データから加速器データを復元する、という手法を開発した。

図 2 は、実データ (青)、および GAN が生成した疑似データ (橙) による、ステアリング磁石パラメータを示したものである。図 2(a) では、モード崩壊により、疑似データが実データを再現できていないが、VAE による次元削減を行うことで、モード崩壊を抑制することができた (図 2(b))。ただし、デコーダを用いた加速器データの復元性能が低いため、今後は、加速器データの復元性能向上を目指す。

【機械学習を用いたデータ処理技術の開発】

大型加速器を用いた超精密測定実験では、捉えた信号事象数で物理感度が決まるため、加速器の衝突頻度を上げ (時間方向の高度化)、測定器のセンサー数を増やして信号検出能力を上げ (空間方向の高度化)、膨大量データを高頻度で取得することが必須になった。しかし、現行の実験データ生成量は、すでに既存のテクノロジーで対処可能な範囲を超え、トリガーやフィルタリングで取得事象数を抑えている。つまり、時間方向で削減している。

従来の「時間方向の削減+空間方向の高度化」を基軸としたデータ収集機構に対して、本研究では、機械学習によるスパースセンシングを導入した新しいデータ処理機構を提案する。

スパースセンシングでは、時間、空間についてスパースに (疎に) 信号を計測し、計測した



図 3 加速器実験データのスパース復元。サンプリングされたデータから復元する。

少数データから、信号のスパース性 (時間や空間についての近傍連続性) に注目して、元信号を復元する (スパース復元)。図 3 に、我々が考案した、加速器実験での生成粒子信号のスパース復元の概念図を示す。データの本質的な情報を保ちながらデータサイズをリアルタイムに低減し、高頻度なデータ収集が可能になる。そのため、「空間方向の削減+時間方向の高度化」へのパラダイムシフトが期待される。素粒子実験では例がない、画期的な手法である。

図 4 に、データ削減手法の評価の概念図を示す。スパースセンシングは、通常のデータのサンプリングのパターン (間隔) を変え、データ量を低減する。スパースにサンプリングされたデータを用いて較正や再構成を行い、通常のデータでの較正や再構成の性能と比較することで、定量的に性能を評価する。データ削減手法として手動ランダムサンプリングと、機械学習で最適化されたパターンに従う削減について評価を行った。

加速器実験のデータ処理にスパースセンシングを導入し、データの本質的な情報を保持したまま、データ量を削減する具体的なアプリケーションとして、本研究では、国際リニアコライダー将来実験計画 (International Linear

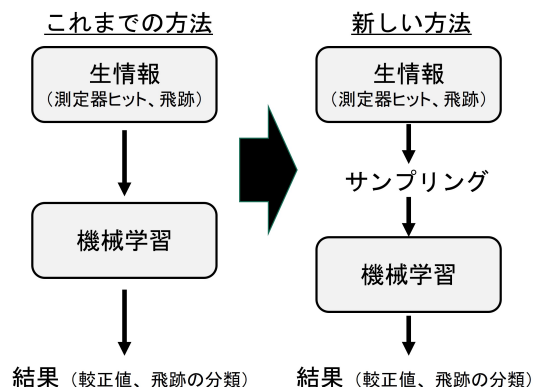


図 4 データ削減手法の評価の概念図。

Collider, ILC) での SiD 測定器用電磁カロリメーター (ECAL) のエネルギー較正において、スパースセンシングを導入し、その定量的な評価を行った。

ECAL は、30 層のサンプリング型構造の電磁カロリメーターで、タングステン吸収体と半導体センサーによる読み出しで構成されている。ビーム衝突点からの入射粒子（電子または光子）のエネルギーを測定する。ビーム衝突点に近い最初の 20 層は、1 層あたり 2.5mm 厚のタングステン吸収体と 1.25mm の読み出しギャップ、最後の 10 層は、1 層あたり 2.5mm 厚のタングステン吸収体と 1.25mm の読み出しギャップを持つ（総物質質量 26 放射長）。読み出し用の半導体センサーは、4mm×4mm のピクセル型センサーで、ECAL ので、エネルギー測定性能デザイン値は $17/\sqrt{E} \oplus 1\%$ (E は入射粒子のエネルギー) である。本研究では、ピクセル型センサーからのデータを削減し、エネルギー較正性能を評価した。

図 5 は入射電子 (5GeV) に対する、エネルギー較正精度を示す。データを削減することで、データ削減なし (青線) の結果よりも、エ

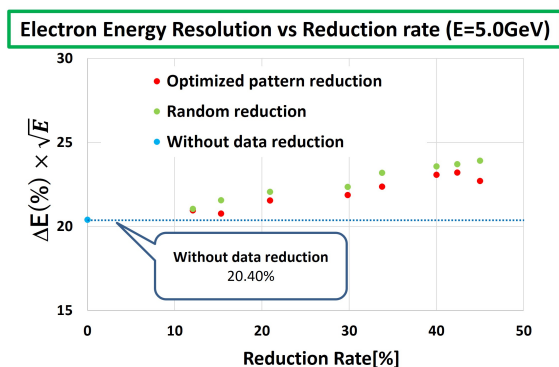


図 5 入射電子 (5GeV) に対する、エネルギー較正精度。データ削減なし (青線)、ランダム削減 (緑点)、および、機械学習で最適化されたパターンに従う削減 (赤点) を示す。

ネルギー較正精度が悪くなるが、ランダム削減 (緑点) に比べて、機械学習で最適化されたパターンに従う削減 (赤点) は、精度の劣化を抑制させることができる。

表 1 は、データ削減率 34% のときの、入射電子 (2.5, 5GeV) に対するエネルギー較正精度を示す。入射電子エネルギー 2.5GeV, 5GeV, どちらのエネルギー領域においても、機械学習で最適化されたパターンに従う削減は、ランダム削減に比べて精度の劣化を抑制させることができることが示された。

入射電子エネルギー [GeV]	2.5	5.0
エネルギー較正精度 [%]		
データ削減なし	17.59	20.40
最適化パターンに従う削減	19.69	22.36
ランダム削減	20.04	23.19

表 1 データ削減率 34% のときの、入射電子 (2.5, 5GeV) に対するエネルギー較正精度。

今後はさらに、最適化手法の改良、および、他のアプリケーションへの適用を検討する。

6 今年度の進捗状況と今後の展望

以下、申請時に設定した 2022 年度の計画と、自己評価を示す。

[申請時に設定した 2022 年度の計画]

● 研究テーマ 1, 2 共通項目

ニューラルネットワークの詳細なデザインと、GPU ノードを用いた大量高速計算を遂行し、原理の検証と、実現に向けた定量的な評価を行う。

● 研究テーマ 1

1) GAN による加速器シミュレータ開発 (約 1,000 パラメータ) を行う。我々の先行研究により、GAN による加速器シミュ

レータを構築すると、パラメータ間で再現性の違いが現れることが明らかになった(モード崩壊)。モード崩壊の問題を解決するため、入力パラメータ数の次元削減や Conditional GAN の適用を行う。2) コントローラーの開発を行う。GAN によるシミュレータを用いてコントローラーの pre-training を行い、定量的な評価を行う。

● 研究テーマ2

1) 我々が先行研究により進めてきた、機械学習を用いた電磁カロリメータ較正、崩壊点再構成において、データ圧縮を適用する。入力パラメータ数が約 3,000-10,000 のニューラルネットワークの前段に、スパースサンプリングを導入し、数 100 万事象の計算を行う。2) データ削減手法(ランダムサンプリング、VAE による次元削減)の事前評価を行う。ランダムサンプリングは最も性能が低いセンシングパターンである。VAE は、最も性能が高いが、リアルタイム処理ができない。機械学習によるスパースセンシングの性能は、これらの間になることが予想される。3) 機械学習でスパースセンシングのパターン最適化を行い、定量的な評価を行う。

[自己評価]

- 研究テーマ1, 2 共通項目: 100% 達成
研究テーマ1, 2, について、ニューラルネットワークの詳細なデザインを行い、GPU ノードを用いて、大量高速計算を遂行した。実現に向けた定量的な評価を行った。
- 研究テーマ1: 70% 達成
GAN による加速器シミュレータ開発(約 1,000 パラメータ)を行った。入力パラメータ数の次元削減を行うことで、モード

崩壊の問題を解決することができた。ただし、デコーダを用いた加速器データの復元性能が低いため、今後は、加速器データの復元性能向上を目指し、さらに、加速器コントローラーの開発を行う。

● 研究テーマ2: 80% 達成

機械学習を用いた電磁カロリメータ較正において、スパースサンプリングによるデータ圧縮を適用させた。ランダム削減に比べて、機械学習で最適化されたパターンに従う削減は、精度の劣化を抑制させることが明らかになった。今後さらに、最適化手法の改良、および、他のアプリケーションへの適用を検討する。

7 研究業績一覧(発表予定も含む)

学術論文(査読あり)

国際会議プロシーディングス(査読あり)

国際会議発表(査読なし)

- Masako Iwasaki, “Overview of AI application in Accelerator”, KEK IINAS 5th International School on Beam Dynamics and Accelerator Technology (ISBA22), Nov. 2022. 【招待講演】
- C. Kato, M. Iwasaki, H. Nagahara, M. Yoshida, T. Suehara, S. Yamada, Y. Nakashima, N. Takemura, T. Nakano, “R&D of the Data Processing with Sparse Sampling using Machine Learning for High Energy Experiments”, ML at HEP workshop, Feb. 2023.

国内会議発表(査読なし)

- 岩崎 昌子 “素粒子物理学実験への機械学習の適用研究”, JHPCN: 学際大規模情報基盤共同利用・共同研究拠点 第14回 シンポジウム, Jul. 2022.

- 加藤睦代, 岩崎昌子, 長原 一, 吉田道隆, 末原大幹, 山田 悟, 中島悠太, 武村紀子, 中野貴志, “機械学習を用いたスパースサンプリングによるデータ処理技術の基礎開発 (II)”, 日本物理学会 2022 年秋季大会, Spt. 2022.
- 岩崎 昌子, “加速器制御への機械学習の適用”, 第 19 回日本加速器学会年会, Oct. 2022. 【招待講演】
- 岩崎 昌子, “機械学習”, Flavor Physics Workshop 2022 (FPWS2022), Nov. 2022. 【招待講演】
- 加藤睦代, 岩崎昌子, 長原 一, 吉田道隆, 末原大幹, 山田 悟, 中島悠太, 武村紀子, 中野貴志, “機械学習を用いたスパースサンプリングによる データ処理技術の基礎開発”, Flavor Physics Workshop 2022 (FPWS2022) , Nov. 2022.
- 岩崎昌子, “機械学習を適用した KEK 電子陽電子入射器ビーム調整のための開発”, 加速器・ビーム物理の機械学習ワークショップ , Mar. 2023. 【招待講演】
- 度会龍, 岩崎昌子, 中島悠太, 武村紀子, 長原一, 中野貴志, 佐藤政則, 佐武いつか, “GAN を用いた加速器シミュレータの開発”, 日本物理学会 2023 年春季大会, Mar. 2023.

公開したライブラリ等

その他（特許，プレス発表，著書等）

- [修士論文] 加藤 睦代, “高エネルギー実験のための機械学習を用いたスパースサンプリングによるデータ処理技術開発”, 修士論文, 大阪市立大学, 大阪, 2023.