



Title	素粒子物理学実験への機械学習の適用研究
Author(s)	岩崎, 昌子
Citation	サイバーメディアHPCジャーナル. 2022, 12, p. 17-20
Version Type	VoR
URL	https://doi.org/10.18910/89340
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

素粒子物理学実験への機械学習の適用研究

岩崎 昌子¹

大阪公立大学 大学院理学研究科／大阪大学 核物理研究センター

1. はじめに

本研究の目的は、大型加速器を用いた素粒子物理学実験に機械学習を適用し、実験における様々なデータ処理性能を高めることである。機械学習は、車の自動走行や医療画像の自動診断などで注目されている AI の基盤技術である。大型加速器実験の大量データ処理へ適用することで、従来のデータ処理手法を上回る、より高性能なツールとして期待できる。ここで、大型加速器を用いた素粒子物理学実験では、実験の巨大化や高度化に伴い、ビッグデータの収集・処理と解析技術が重要な研究基盤である。また、実験遂行費用が高額であるため、高精度で実験装置を制御し、実験の効率化を図ることが必須である。

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処理・再構成処理を行う、
- 4) 較正・再構成された実験データから、物理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、これまでも機械学習の適応研究が行われていたが、主に 4) における開発、つまり、データ解析において効率よく信号事象を抽出するための開発が行われてきた。しかし、近年の機械学習手法の飛躍的な進展により、様々な最先端機械学習技術の適用が可能になった。

本研究では、従来の素粒子実験ではあまり例がみ

られなかった 1)、2)、3) における機械学習適応を目指す。具体的には、以下の開発を行う。

- 機械学習を用いた加速器制御技術の開発
- スパースセンシングを導入したデータ処理技術の開発

上記の研究開発は、どちらも最先端の機械学習手法をベースとした研究開発である。加速器実験による膨大量データに対する大量計算が必要である。

2. 機械学習を用いた加速器制御の開発

高エネルギー加速器研究機構(KEK)で稼働中の、電子・陽電子入射器、Linac 加速器[1]の入射効率向上を目標として、機械学習を用いた運転調整システムの開発(加速管の RF 位相調整や、ビーム位置補正のためのステアリング電磁石の調整)を行った。

Linac 加速器の運転調整には以下の問題がある。

- 運転調整に関連するパラメータの数が多く($\sim O(1000)$)、相関が複雑であるため、パラメータ調整が困難である。そのため、パラメータ調整の最適化の速さや精度が、加速器運転員の経験・技能によって異なる。
- 加速器構成機器の温度変化や振動、潮汐力による影響等、環境変化に応じて常時運転調整を行う必要がある。

以上の問題点を解決するために、複数パラメータ間の相関関係の記述に強みを持つ機械学習の適用が有効であると考えられる。

加速器調整時間の高速化や調整性能の向上、安定化を行うために、我々は、以下で述べられるようなディープニューラルネットワーク(DNN)を用いた次元削減(特徴抽出)による運転状況可視化手法の

¹ プロジェクトメンバー：久野 彰浩、加藤 睦代（大阪市立大学）、末原 大幹（九州大学）、佐藤 政則、佐武いつか、山田 悟（KEK）、中島 悠太、武村 紀子、長原 一（大阪大学 IDS）、中野 貴志（大阪大学 RCNP）

開発と、環境駆動型の機械学習（強化学習）の実装を行うための準備研究を行った。これらの開発は2018年から2021年に蓄積された、Linac 加速器運転の実データを用いて行った。

2.1 次元削減による運転状況可視化手法の開発

加速器データを可視化するために、変分オートエンコーダー(Auto-Encoding Variational Bayes, VAE) [2]を用いた次元削減を行うことで、1232 パラメータの加速器環境データを、2 次元の出力データに次元削減を行った。図 1 に、出力結果を示す[3]。2 次元の潜在変数 $Z(0)$, $Z(1)$ で表される加速器環境データのふるまいが、短期的には連続的に分布しているが、長期的には非連続であり、時期によって異なる分布であることが明らかになった。したがって、加速器運転調整を行うためには、周囲の環境変化に応じた運転調整手法の開発が重要であることが示された。

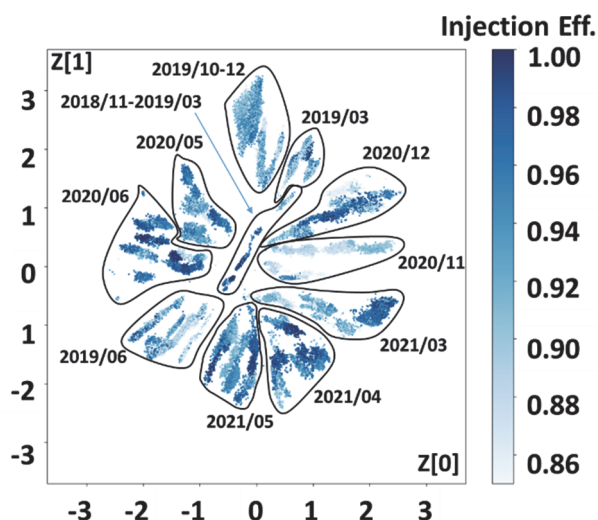


図 1: VAE からの出力 (2 次元潜在変数)。2018 年から 2021 年の Linac 運転データを使用した。色の濃淡は、入射効率の大きさを表す。

2.2 強化学習による加速器制御の準備研究

周囲の環境変化に応じた運転調整を行うためには、強化学習（環境適応型機械学習）の適用が有効であると考えられる。強化学習は、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が、報酬の指標と考えられる。

強化学習を導入するための準備研究として i) DNN を用いた加速器の運転状況と入射効率の相関関係を記述する手法の開発、ii) 機械学習を用いた加速器運転シミュレータの開発を行った。

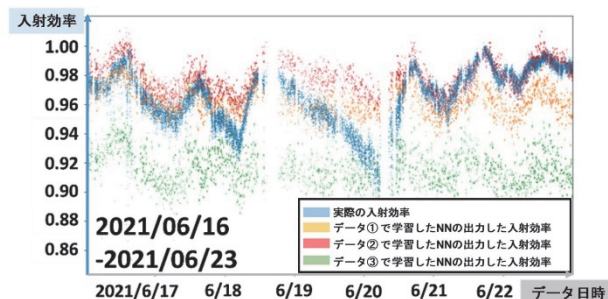


図 2: DNN による入射効率予測結果。2021 年 6 月について予測を行った。使用学習データは、①2018 年 11 月-2021 年 5 月、②2021 年 5 月、③2018 年 11 月-2021 年 4 月に取得された加速器データである。

図 2 は、DNN による入射効率の予測結果を示す[3]。2021 年 6 月の入射効率の予測を行った。DNN の学習に使用したデータは ①2018 年 11 月から 2021 年 5 月、②2021 年 5 月、③2018 年 11 月から 2021 年 4 月に取得された加速器データである。2021 年 6 月の入射効率を予測するためには、2021 年 5 月のデータ、つまり直近の実データによる学習が必要であることが示された。したがって、強化学習を行うためには、周囲の環境変化にあわせた学習用データを用意して、予測モデルを随時更新する必要がある。

また、加速器制御に強化学習を導入するために、加速器運転シミュレータの開発を行った。実機を用いたパラメータ最適化作業は危険を伴うため、シミュレータを用いたパラメータ決定 (pre-training) が必要である。そのため我々は、敵対的生成ネットワーク (Generative Adversarial Networks. 以下、GAN) を用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して、疑似データを生成する機械学習である。

我々は、加速器実データを GAN に入力して、疑似データ生成を試みた。結果を図 3 に示す[3]。図 3 は、実データおよび GAN が生成した疑似データによる、入射効率とステアリング磁石パラメータの相関を示したものである。疑似データは、実データの

一部区間しか、再現できなかった(モード崩壊、Mode collapse)。今後は、モード崩壊の抑制方法の開発が重要である。

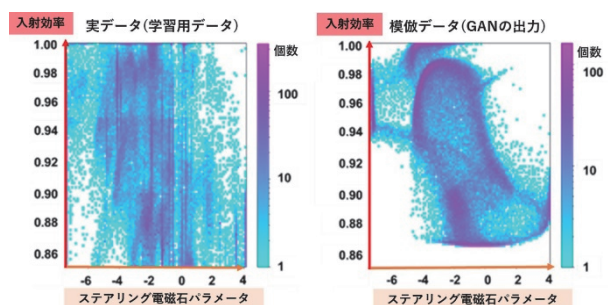


図3：実データ、およびGANが生成した疑似データによる、入射効率とステアリング磁石パラメータの相関を示したもの。疑似データは、実データの一部区間しか、再現できていない。

3. スパースセンシングを導入したデータ処理技術の開発

大型加速器を用いた素粒子・原子核実験では、測定感度を高めるために実験で収集する事象数を増やすことと、データの再構成、物理解析の精度及び効率を良くすることの二点が重要である。事象数を増やすためにデータ取得頻度を高める必要があり、測定器測定精度を高めるためには、センサー数を増加させる必要がある。それに伴ってデータサイズが年々増加し、膨大量のデータ処理の高精度化及び高効率化が課題となった。そこで、機械学習を用いたスパースセンシングによるデータ処理技術を導入することにより、加速器から高頻度で供給される衝突実験データを、測定器のデータ収集システム能力の限界まで収集する方法を開発する。

スパースセンシングは、信号のスパース性(時間や空間についての近傍連続性)に注目することで、時間や空間についてダウンサンプリングされた少数データから、元信号を復元する技術である(スパース復元)。加速器実験へ適用するには、図4に示すように、データをダウンサンプリングし、機械学習で再現する手法が考えられる。本研究では具体的なアプリケーションとして、ILC SiD 電磁カロリメータ[4]におけるエネルギー較正へ、スパースセンシングを適用させる。

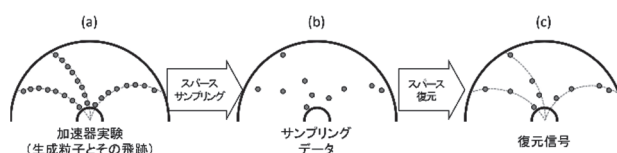


図4：加速器実験データのスパース復元手法の概念図。ランダムサンプリングで得られたデータから復元を行う。

スパースセンシングの導入により、データ解析のために重要で本質的な情報を十分に保ったままデータサイズをリアルタイムに圧縮し、膨大量のデータ処理の高精度化及び高効率化を目指す。

3.1 MNIST データを用いた準備研究

SiD 電磁カロリメータ較正への適用のための準備研究として、手書き文字データ(MNIST データセット[5])を用いた評価を行った。ここでは、データの削減方法として、ランダムサンプリング、および機械学習(VAE)による次元削減を行った。

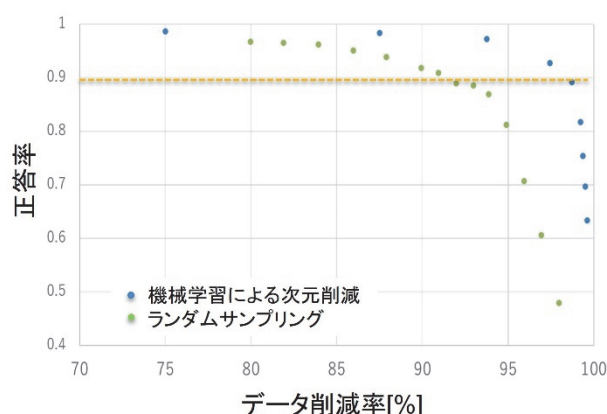


図5：機械学習による次元削減を用いた文字分類(MNISTを用いた準備研究)。

図5に結果を示す。データを1/10(ランダムサンプリング)、あるいは1/50(機械学習による次元削減)に圧縮しても、90%の識別性能を維持できることが示された。

3.2 ILC SiD 電磁カロリメータのエネルギー較正への適用

ILC SiD 電磁カロリメータのシミュレーションデータを用いて、本開発を行った。電磁カロリメータは、入射された粒子のうち、電子、または光子のエ

エネルギーを測定する測定器である。測定器からの出力データを較正して、入射粒子のエネルギー値を得る。一般的に、入射粒子のエネルギーの値は、測定器の出力データの値と比例、つまり線形性を仮定してエネルギー較正を行うが、我々は、先行研究を行い、深層学習（回帰問題）によるエネルギー較正手法を開発し、エネルギー測定精度の向上を示した[6]。

このアプリケーションにおいて、データ削減を行い、深層学習を用いた回帰および、スパース復元を行うことで、データ削減の定量的な評価を行った。結果を表 1 に示す。

電子のエネルギー	2GeV	5GeV
データ削減なし	20.2%	20.5%
10%削減（ランダムサンプリング）	19.2%	20.1%

表 1：SiD 電磁カロリメータで検出された電子のエネルギー測定精度 (Pre liminary)。

入射エネルギーが 2GeV または 5GeV の電子に対して、データ削減なしと、10%のランダム削減を行ったものを比較したところ、ほぼ同等のエネルギー測定精度が得られた[7]。ただし、データの統計量および、評価のための計算量が少ない結果であるため、今後計算を進め、より定量的な評価を行う。

4. おわりに

素粒子物理学実験への機械学習の適用研究として、機械学習を用いた加速器制御および、スパースセンシングを導入したデータ処理技術の開発を行った。

機械学習を用いた加速器制御開発では、強化学習の重要性が示された。強化学習のための、GAN を用いた加速器シミュレータ開発では、モード崩壊の抑制が重要であることが明らかになった。

スパースセンシングを導入したデータ処理技術の開発では、MNIST データを用いた予備実験を行い、定量的な評価を得た。また、ILC SiD 電磁カロリメータへの適用研究をスタートさせた。

今後さらに、素粒子物理学実験への機械学習の適用実証研究を進め、データ処理技術の向上を目指す。

謝辞

本研究は、大阪大学核物理研究センター(RCNP)研究プロジェクト、大阪大学データビリティフロンティア機構学際プロジェクトから、多大なご協力を頂きました。本研究の一部は、文部科学省「ダイバーシティ研究環境実現イニシアティブ（牽引型）」連携型共同研究助成を受けて、実施されました。また、大阪大学サイバーメディアセンターの公募型利用制度による支援を受けました。

参考文献

- (1) KEK electron-positron injector (LINAC), <https://www.kek.jp/en/Research/ACCL/LINAC/>
- (2) D. P. Kingma and M. Welling, “Auto-Encoding Variational Bayes”, arXiv:1312.6114 [stat.ML], (2013).
- (3) A. Hisano, M. Iwasaki, *et. al.*, “R&D of the KEK Linac Acc. Tuning Using ML”, Proc. ICALEPCS 2021, doi:10.18429/JACoW-ICALEPCS2021-WEPV010, (2021).
- (4) T. Behnke, *et. al.*, “The Int. Linear Collider Technical Design Report – Vol. 4: Detectors”, arXiv:1306.6329 [physics.ins-det], (2013).
- (5) Y. LeCunL, C. Cortes, and C.J.C. Burges, “The MNIST database of handwritten digits”, <http://yann.lecun.com/exdb/mnist/>.
- (6) M. Iwasaki, Y. Naka, *et. al.*, “R&D of the Energy Calibration for the SiD EM Calorimeter based on Machine Learning”, The 3rd edition of the Cal. for High Energy Frontier conf. (CHEF 2019), (2019).
- (7) C. Kato, M. Iwasaki, *et.al.*, 77th Annual Meeting of The Physical Society of Japan, (2022).