素粒子物理学実験への機械学習の適用研究

拠点課題ID: jh220034

2022/7/7

大阪公立大学^A,大阪公立大学NITEP^B,九州大学^c, 高エネルギー加速器研究機構^D,大阪大学IDS^E,九州工業大学^F,大阪大学RCNP^G

> <u>岩崎 昌子</u>ABEG,加藤 睦代^A,末原 大幹^C,山田 悟^D 長原 一^E, 中島 悠太^E, 武村 紀子^{EF}, 中野 貴志^G

> > JHPCNシンポジウム



大型加速器を用いた素粒子・原子核物理実験で 機械学習の適用研究を行い

実験効率・実験性能を向上させ、測定精度を高める 素粒子・原子核実験と情報分野との共同研究を行い 加速器実験のアプローチに変革を生み出したい



大型加速器を用いた素粒子実験 ●大型実験施設(巨大加速器、測定器)を使用 →実験装置の制御点数は数千-数万点以上 ➡膨大量の実験データ →BelleII実験では、年間数十ペタバイトの予定 ■∕膨大な背景事象から極微の信号事象を抽出 大量データ処理を

いかに、効率よく・精度よく行うか

大型加速器を用いた素粒子実験

LHC エネルギー 14TeV 周長 = 27km

・ATLAS 実験 ・CMS 実験 ・LHCb 実験



ATLAS 実験 5000人



- ・素粒子実験用の主な大型 加速器は世界に2つ
- 素粒子や加速器の研究者 技術者・学生が多数集結

SuperKEKB エネルギー~11GeV 周長 = 3km

・Belle II 実験







周長 3km

7.7m(W) x 7.2m(D) x 7.9m(H)



<u>大量の素粒子反応を生成</u>



大型加速器実験	Bellell実験
<u>大量データ 高額な運転経費</u>	生成データ量 年間~30ペタバイト 運転経費 年間 <mark>数十億円</mark>

実験の効率化、大量データを効率的に学術的結果へ導くために、 最先端機械学習の適応が有効

例:10%の向上 → 年間数億円の費用効果



情報分野で開発された、最先端機械学習技術を 加速器実験でのデータ処理へ導入することで 従来よりも**高性能、高速**なデータ処理を期待

機械学習 → 加速器実験のビッグデータに対する 強力なデータ処理ツール



大型加速器を用いた素粒子実験への 機械学習の適用として 以下の研究テーマの開発を行う

 1. 機械学習を用いた加速器制御技術の開発
 2. 機械学習を用いたデータ圧縮, データ較正・ 再構成手法の開発

1. 機械学習を用いた加速器制御

KEK Linac加速器調整の開発



RFモニター:60台 ステアリング電磁石:200台 ビーム位置モニター(BPM):100台

KEK Linac、大阪市大、阪大IDS

加速器制御の問題点

- 構成要素が多く、<mark>複雑なシステム</mark> →調整速度が個人の技量に依存
- ・ 温度変化等、周囲の環境が変化する
 → 常時調整が必要

入射効率を高めるために 効率のよい加速器調整が重要

機械学習を用いた加速器制御を検討



機械学習を用いた加速器制御 開発の現状



→教師なし学習(変分オートエンコーダー VAE)を導入することで ~800パラメータの"加速器の状態"を可視化することに成功

周囲の環境変化に適合

→ 直前のデータ(~1日分)を用いた学習により最適なパラメータの 予測に成功

→予測精度をあげるために、学習用加速器シミュレータを作成

<u>本研究ではGANを用いたシミュレータ開発を行う</u>



機械学習を用いた加速器制御

加速器の設定値を実際に変化させながら調整することは危険 (車の運転で、ハンドルを左右に切ってから進む方向を決めるのは危険)

加速器運転調整に強化学習を導入するためには学習用加速器シミュレーターが必要

GAN: Generative Adversarial Network 敵対的生成ネットワーク



https://dzone.com/articles/working-principles-of-generative-adversarial-netwo



加速器データ(1233パラメータ)を入力し、_{A.Hisano} GANで疑似データを生成した (Osaka-City U.)



疑似データは実データ(学習用データ)の 一部しか再現していない



加速器データ(1233パラメータ)を入力し、 A.Hisano GANで疑似データを生成した (Osaka-City U.)





2パラメータのみの加速器データを入力して 疑似データを生成 A.Hisano (Osaka-City U.)



モード崩壊が起きている



モード崩壊が起きている



- 加速器実験では膨大量のデータが生成される
- ・ データ収集能力限界があるため、Filteringが必須
 →物理事象を選別するFilteringを設定している

トリガー条件をみたさないデータは捨てられている

もし機械学習を用いて <u>リアルタイムに</u>データサイズが<u>圧縮できれば</u> 全ての事象を取得することが可能になる?!

機械学習による「トリガーレスDAQ」を目指す





開発は以下の要素開発を組み合わせる

1-1: 機械学習を用いたデータ較正1-2: 機械学習を用いたデータ圧縮

現在までに、1-1の開発を行った 本研究では、続けて 1-2 を行う





電磁カロリメータからの生データ(エネルギー、位置) を用いた機械学習(回帰)でエネルギー較正を行った



測定エネルギー, 入射粒子の位置, …

エネルギー



LCWS2021 2021/3/17 Presented by M. Iwasaki



入力:784 パラメータ

基礎研究として、文字データ(MNIST)を入力し、 手動(ランダムにダウンサンプリング)で データ圧縮を行ったときのNN(識別)の性能を評価した



入力:784 パラメータ





2~5GeVの測定器データに、削減率5% - 20%でランダムサンプリン グを適用し、削減なしの場合のエネルギー較正性能と比較

	2.5GeV	5GeV
削減なし	14.93	16.15
5%削減	15.60	17.65
10%削減	16.48	19.00
20%削減	17.62	19.68

Preliminary

~10%のランダム削減であれば 元データを使用した場合の精度を誤差の範囲内で維持できた。

<u>機械学習を用いてサンプリングパターンを最適化することで、</u> <u>さらに削減を行っても精度を維持できる可能性がある</u> 23





