

素粒子物理学実験への 機械学習の適用研究

拠点課題ID: EX21701

2021/7/9

大阪市立大学^A, 大阪市立大学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, 大阪大学RCNPF^F

岩崎 昌子^{ABEF}, 久野 彰浩^A, 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^E, 中野 貴志^F

JHPCNシンポジウム

研究目的

素粒子実験分野研究者と情報科学研究者が
部局、大学、機関の枠を超えてグループを形成し、

大型加速器を用いた素粒子実験へ

深層学習を適用し

物理実験の実験効率と実験性能を向上させる

素粒子
原子核物理



情報科学

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

大量データを処理を
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

フランス-スイス

日本

SuperKEKB

エネルギー ~11GeV
周長 = 3km

• Belle II 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

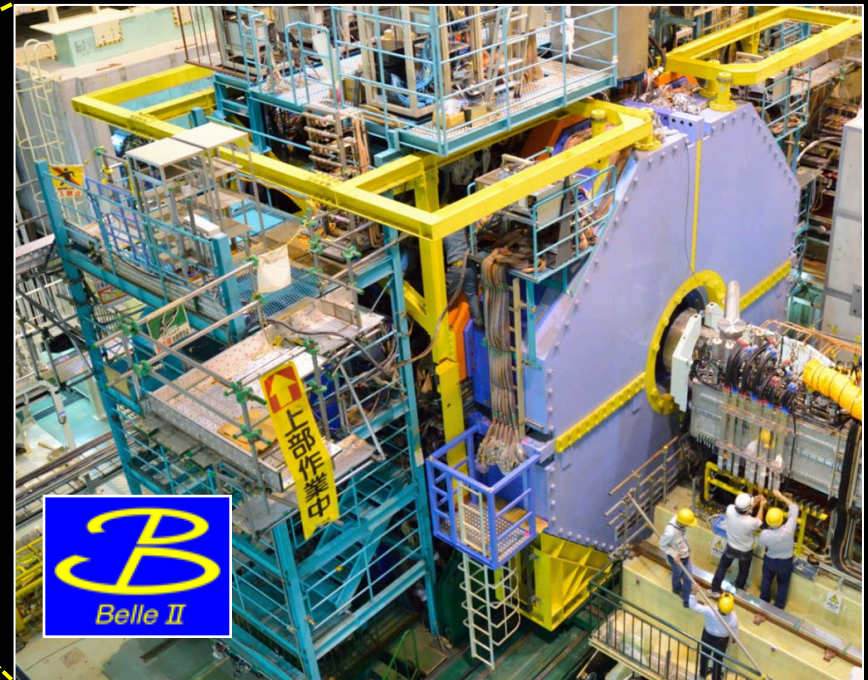
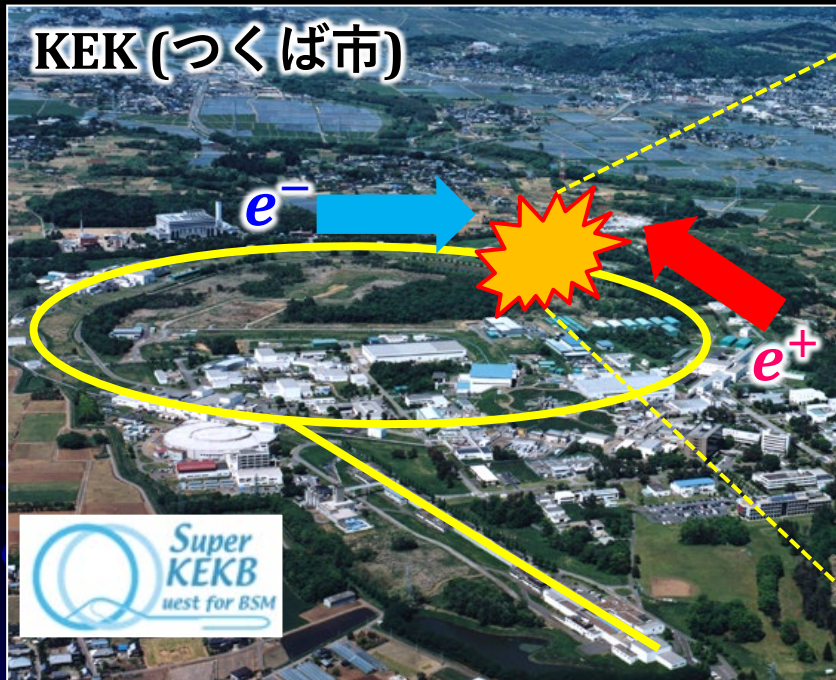


Belle II 実験 970人

Belle II 実験の装置

SuperKEKB 加速器

Belle II 測定器



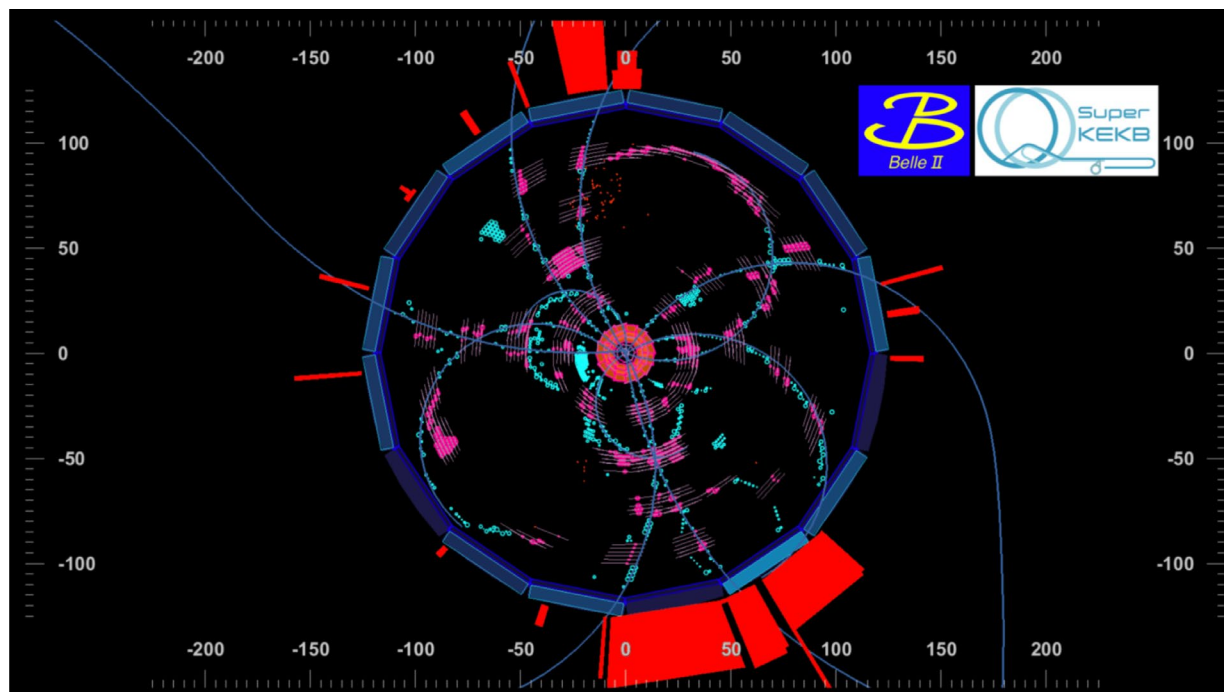
周長 3km

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

周長3kmの
巨大加速器

大量の素粒子反応を生成

大型加速器実験 × 最先端情報学



測定器の電気信号

機械学習

粒子の軌跡と種類

大型加速器実験
大量データ 高額な運転経費

BelleII実験
生成データ量 年間~30ペタバイト
運転経費 年間数十億円

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

大型加速器実験での データ処理

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

機械学習

Supervised Learning
教師あり学習

Task driven

Classification
分類・識別

Regression
回帰

Unsupervised Learning
教師なし学習

Data driven

Dimensionality Reduction
次元削減

Clustering

Reinforcement Learning
強化学習

Environment driven

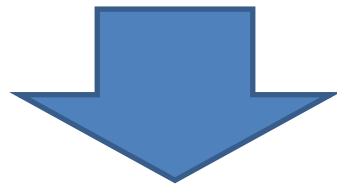
Algorithm learns to react to the environment

Real-time decisions
Game AI
Learning Tasks
Robot Navigator
....

いろいろなタイプの機械学習があります

加速器実験での機械学習

これまで加速器実験で行ってきた機械学習は
おもに特徴量(High-level data)を用いた
教師あり学習による分類・識別だった



計算機技術・機械学習技術の進展により
様々なタイプの機械学習が可能になった
低特徴データ(生データ、low-level data)を用いた
機械学習も可能になった

最先端機械学習技術：深層学習

深層学習 (Deep Neural Network, DNN) では
low-level dataを用いた機械学習が可能

深層学習

入力パラメータ数が多い

and / or

NNの中間層が多層

NNの構造が複雑

NNの表現力が向上

2000年代は計算リソースの制限のためDNNの適用が非現実的だった
→ 価格が現実的になった・処理速度が速くなった DNNが可能に!

加速器実験への 最先端機械学習の適用

【キーワード】

- 低特徴データ (Low-level data, 生データ)
- 様々なタイプの機械学習
回帰、次元削減、強化学習、
再帰型ネットワーク、畳み込み型ネットワーク ...

本研究での開発項目

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開発を行う

1. 機械学習を用いたデータ圧縮
2. 機械学習を用いた加速器制御

1. 機械学習を用いたデータ圧縮

- 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



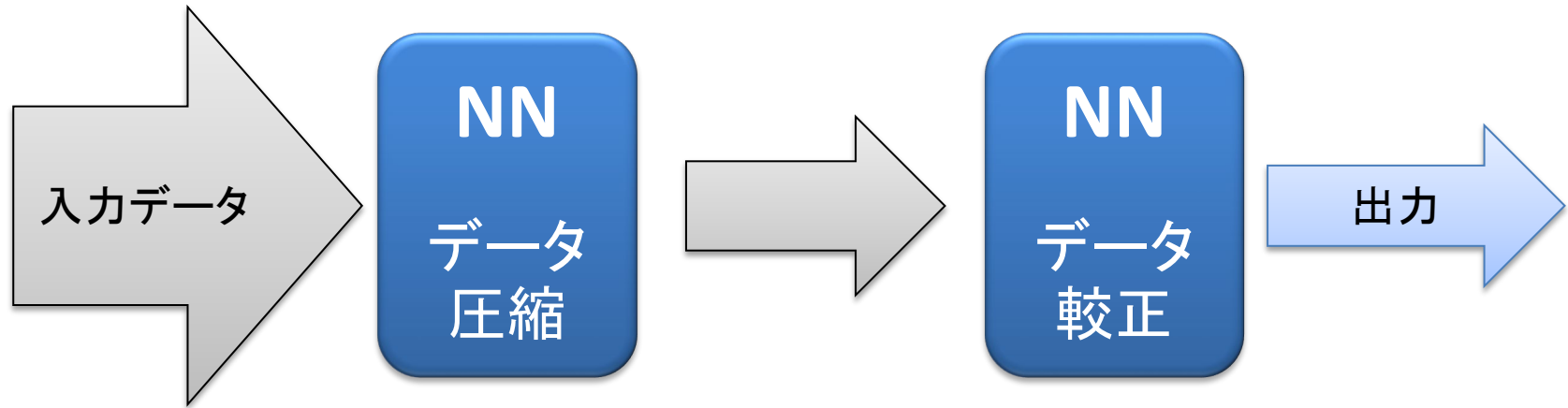
トリガー条件をみたさないデータは捨てられている

もし機械学習を用いて

リアルタイムにデータサイズが圧縮できれば
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

開発の指針



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校正
- 1-2: 機械学習を用いたデータ圧縮

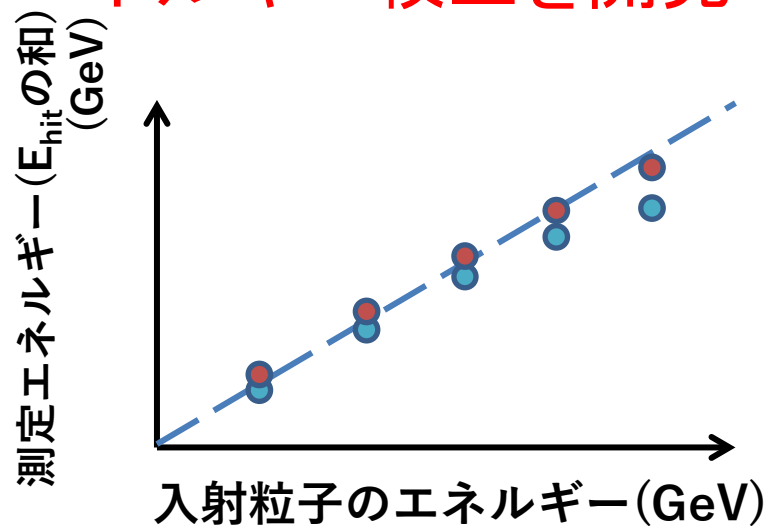
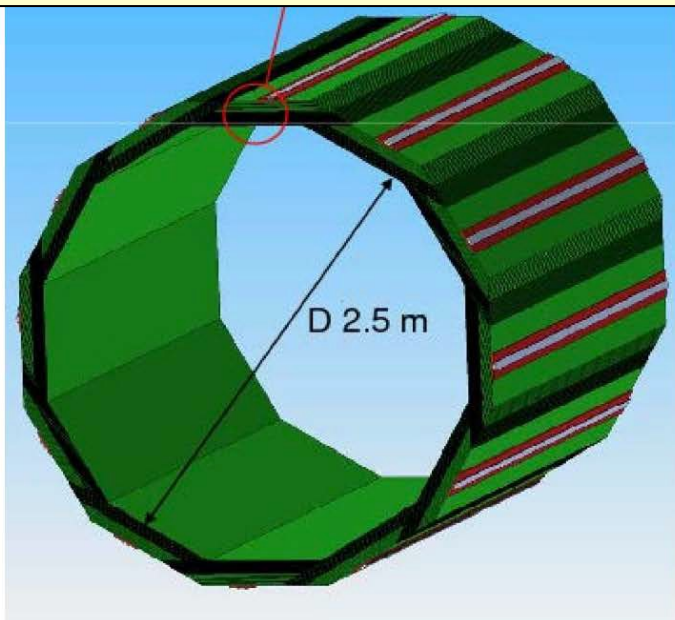
現在までに、1-1 の開発を行った
本研究では、続けて 1-2 を行う

機械学習を用いたデータ較正 カロリメータのエネルギー較正(回帰)

Osaka-City U., IDS, U. Oregon, SLAC

ILC SiD測定器電磁カロリメータのエネルギー較正を開発

- 30層Si + W サンプル型
- 物質質量(total) $\sim 26X_0$
- $\Delta E \sim (17/\sqrt{E} \oplus 1)\%$ (デザイン値)

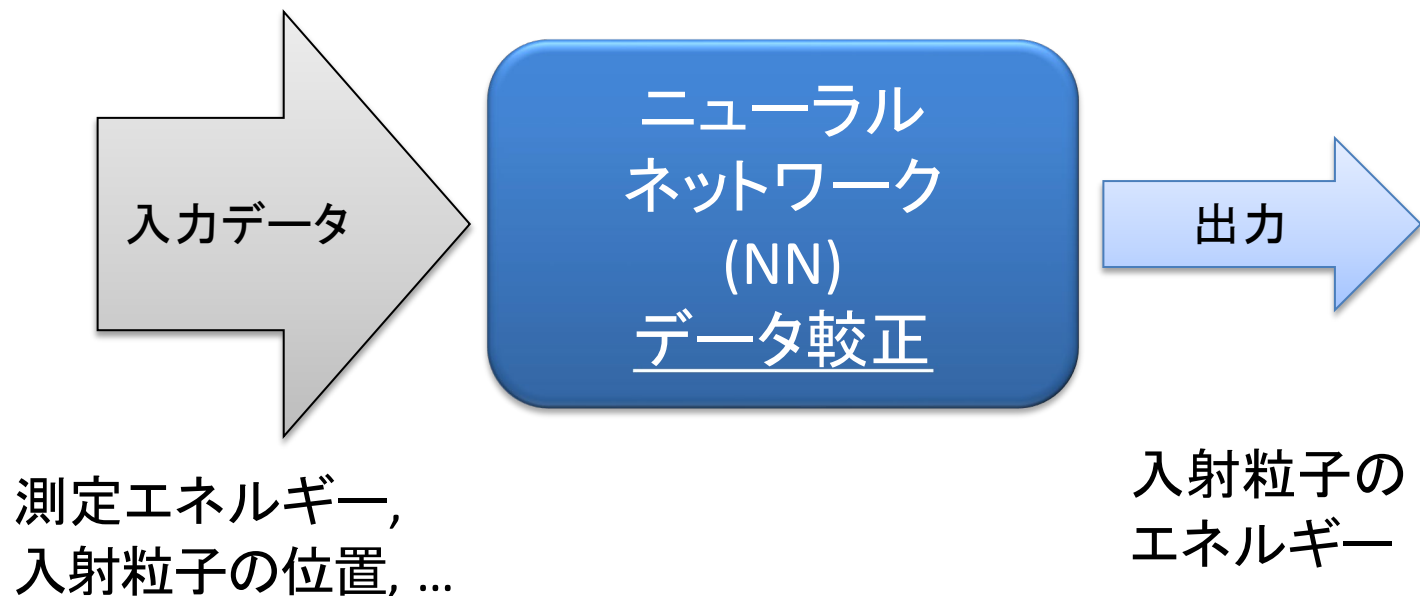


問題点

1. 入射粒子のエネルギーと測定エネルギーが**非線形**
2. 電子と光子で、測定器の応答が異なる(**粒子依存性**)

機械学習を用いたデータ較正 カロリメータのエネルギー較正(回帰)

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

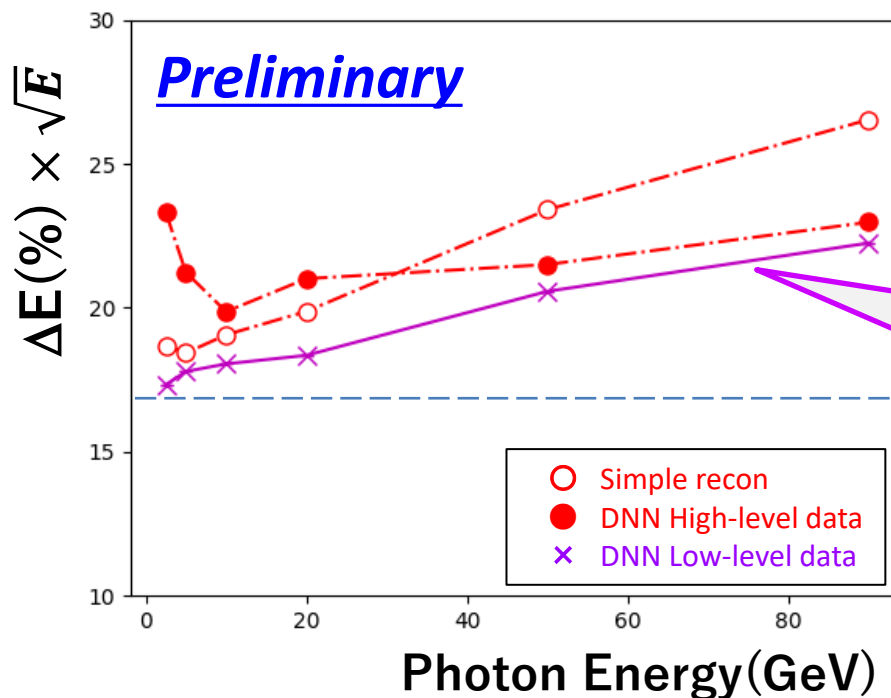


機械学習を用いたデータ較正 カロリメータのエネルギー較正(回帰)

エネルギー較正性能向上!!

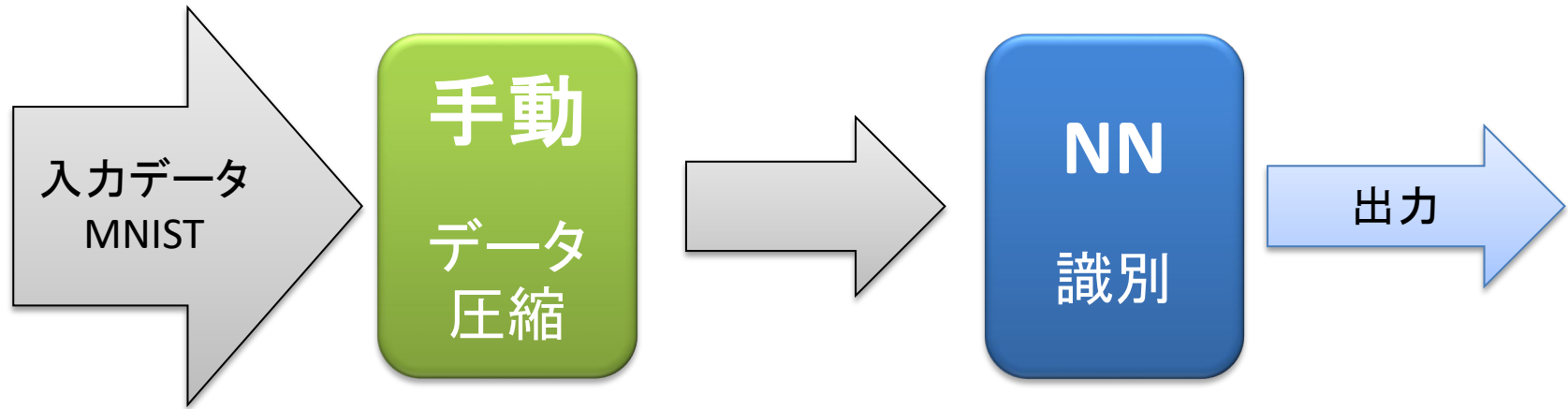
Photon Energy Resolution $\Delta E(\%) \times \sqrt{E}$

Y.Naka (Osaka-City U.)



生データを用いた
機械学習

機械学習を用いたデータ圧縮



入力: 784 パラメータ

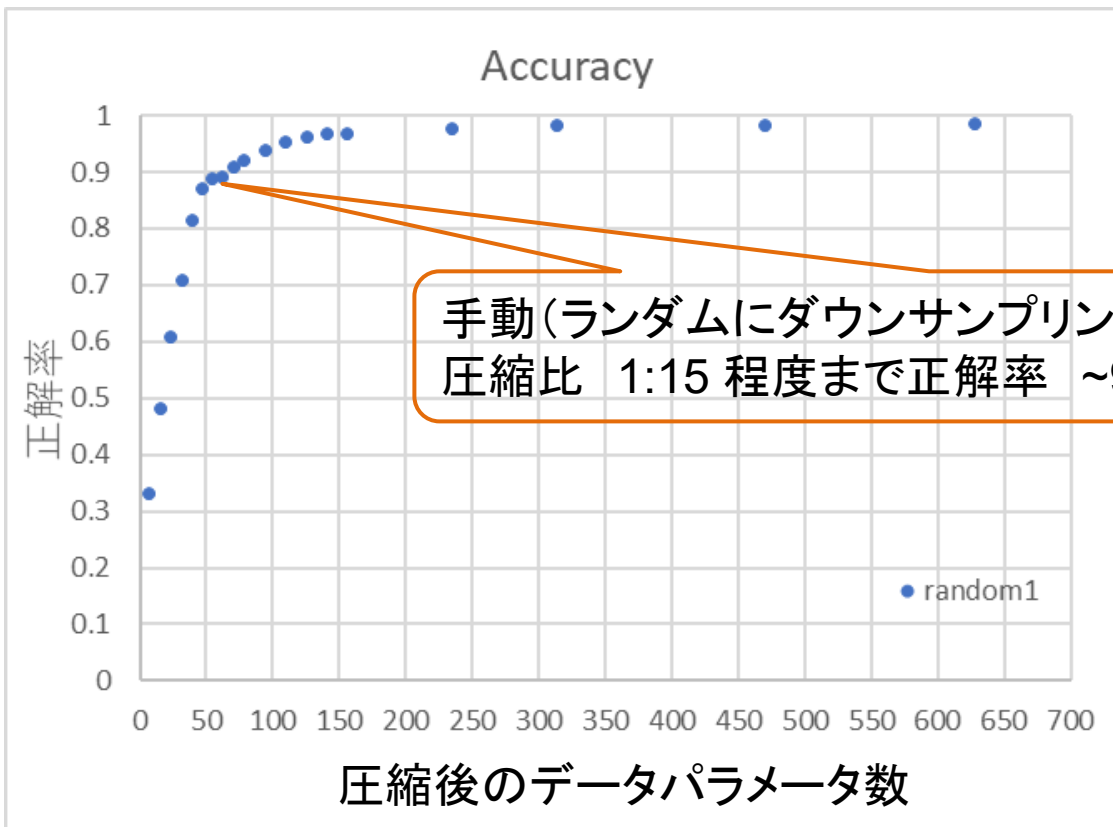
基礎研究として、文字データ(MNIST)を入力し、
手動(ランダムにダウンサンプリング)で
データ圧縮を行ったときのNN(識別)の性能を評価した

機械学習を用いたデータ圧縮



入力: 784 パラメータ

C.Kato (Osaka-City U.)



手動(ランダムにダウンサンプリング)で圧縮した場合
圧縮比 1:15 程度まで正解率 ~90% を保つことができた

今後の展望

機械学習で圧縮することで
圧縮性能の向上を目指す

2. 機械学習を用いた加速器制御

KEK Linac加速器調整の開発

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、**複雑なシステム**
→ 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

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

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

複雑なシステム

- 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

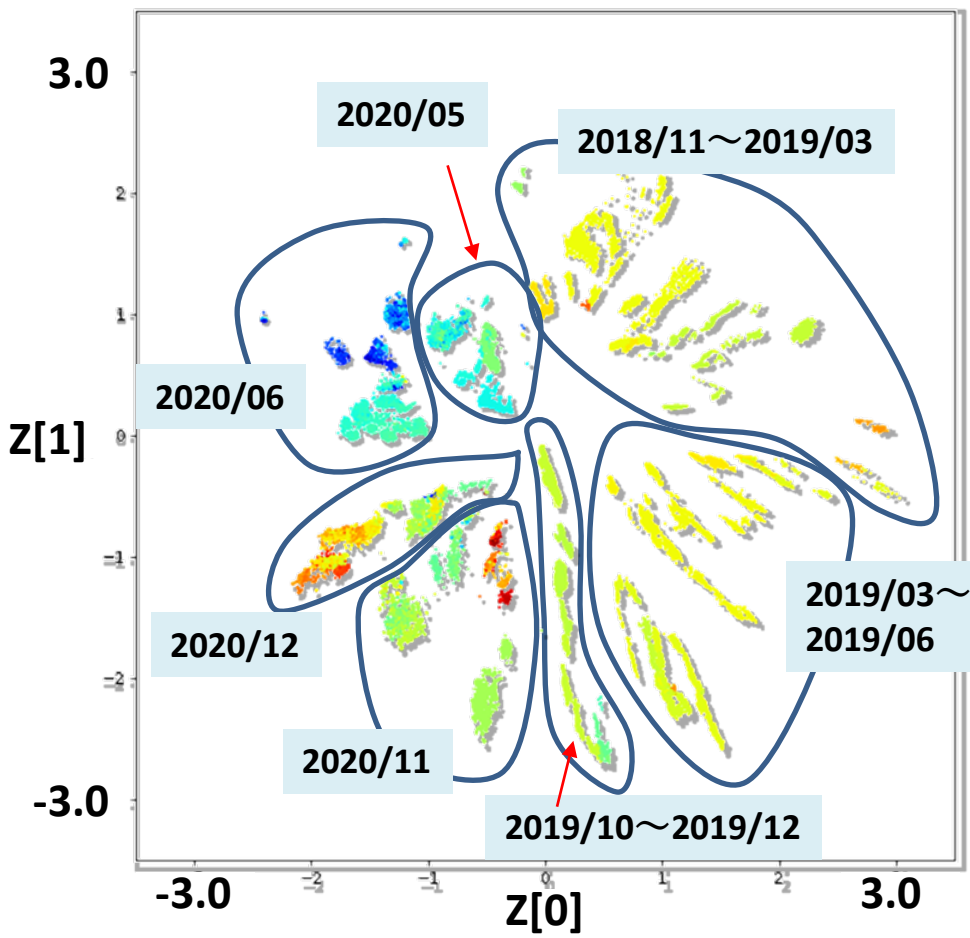
- 直前のデータ(~1日分)を用いた学習により最適なパラメータの
予測に成功
- 予測精度をあげるために、学習用加速器シミュレータを作成

本研究ではGANを用いたシミュレータ開発を行う

教師なし学習(VAE)による加速器状況の可視化

教師なし学習(VAE)による次元削減で
815パラメータの加速器データを2次元で表現 → 可視化!

2D Latent Variable Z(0, 1)



A.Hisano (Osaka-City U.)

Steering Magnet (3 sector 2nd)
Current value (normalized to 0-1)

Visualize the accelerator status with 2 parameters

In short term (~1month)
acc. status does not drastically change

In long term (> 6 month)
acc. parameters vary over a wide range

機械学習を用いた加速器調整

A.Hisano (Osaka-City U.)

2018/11

2020/06/08

2020/06/22



2019/12

2020/06/15

Steering Magnet
(3 sector 2nd)
Current value

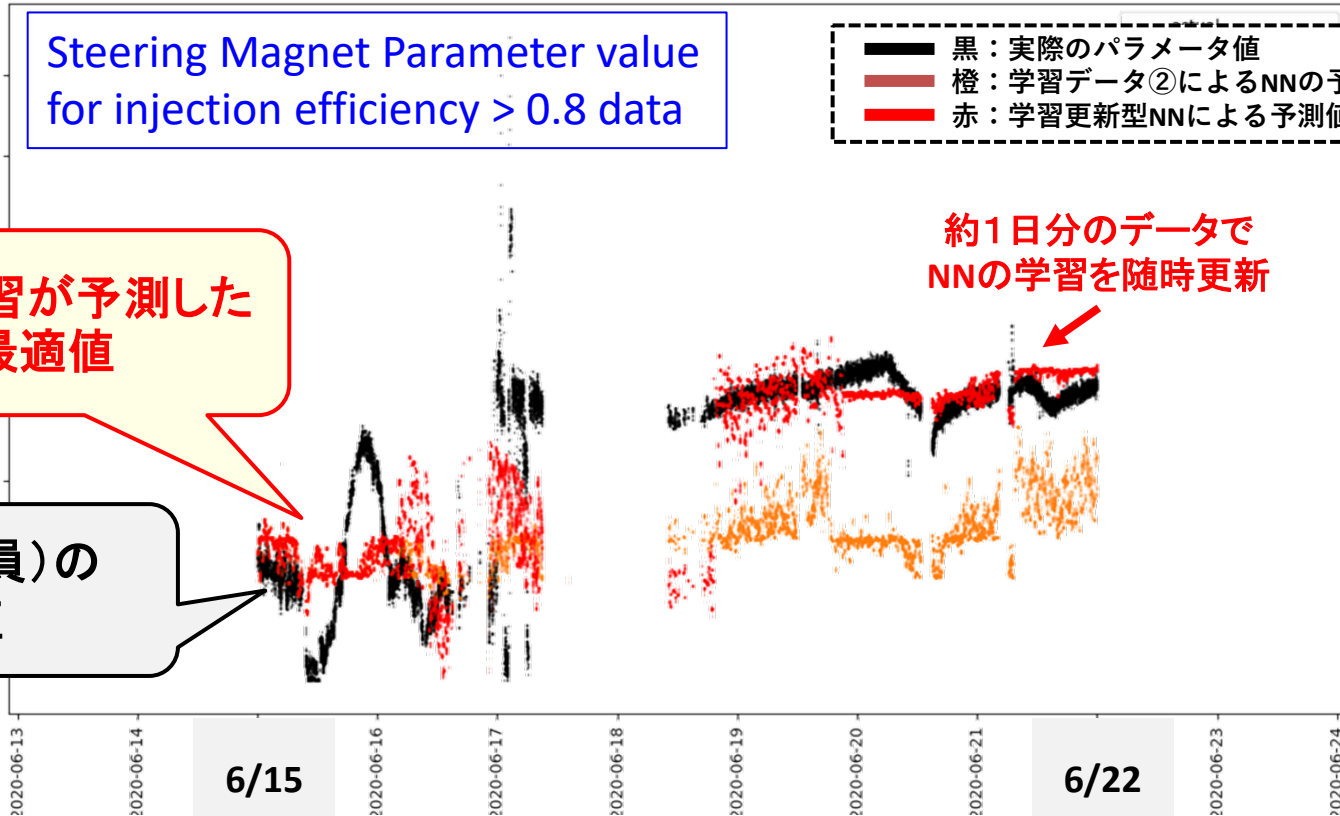
Steering Magnet Parameter value
for injection efficiency > 0.8 data

黒：実際のパラメータ値
橙：学習データ②によるNNの予測値
赤：学習更新型NNによる予測値

機械学習が予測した
最適値

人間(運転員)の
設定値

約1日分のデータで
NNの学習を随時更新



人間 vs 機械学習?!

GANを用いた加速器シミュレータ

GAN : **G**enerative **A**dversarial **N**etwork
敵対的生成ネットワーク

学習時に入力した画像やデータをもとに
新しい疑似データを生成するモデル
(教師無し学習)

互いに競合する二つのネットワークから構成される

Generator

入力されたノイズデータ(乱数)
から疑似データを出力する

Discriminator

Generatorが生成した疑似データ
と学習用データ(本物データ)を識別する

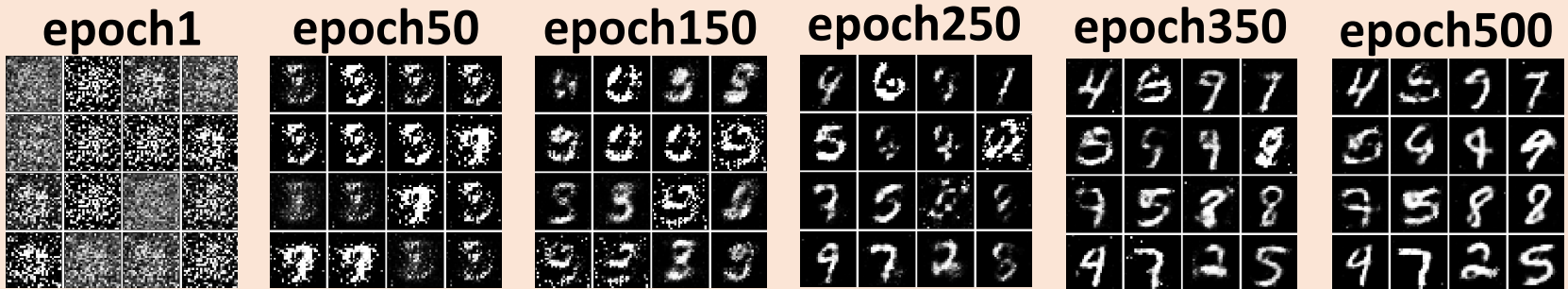
基礎研究として、文字データ(MNIST)の疑似データを生成した

GANを用いた加速器シミュレータ

基礎研究として、文字データ(MNIST)の疑似データを生成した

A.Hisano (Osaka-City U.)

MNISTデータをサンプルにGANの演習を実行



学習が進むにつれて、GAN(のGenerator)はMNISTデータらしい画像を生成できるようになった

今後の展望

実際の加速器データの疑似データを作成し
性能評価を行う

まとめ

機械学習→素粒子実験の
ビッグデータに対する強力なデータ処理ツール

これまでの要素技術開発にあわせて、
本研究でさらに、
最先端機械学習の適用研究を行う



より高性能なデータ処理を開発

加速器実験の新たな基盤技術に発展させたい