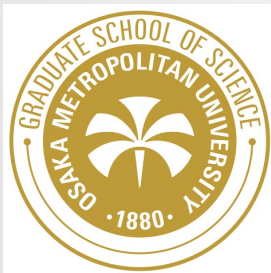


機械学習

2022/11/8

岩崎昌子 (阪公大/NITEP, 阪大RCNP/IDS)



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 ~1100人



BelleII実験

大型国際共同実験
26カ国約~1100人の研究者で構成

茨城県つくば市
高エネルギー加速器研究機構(KEK)

SuperKEKB加速器

周長3kmの
巨大加速器

1km

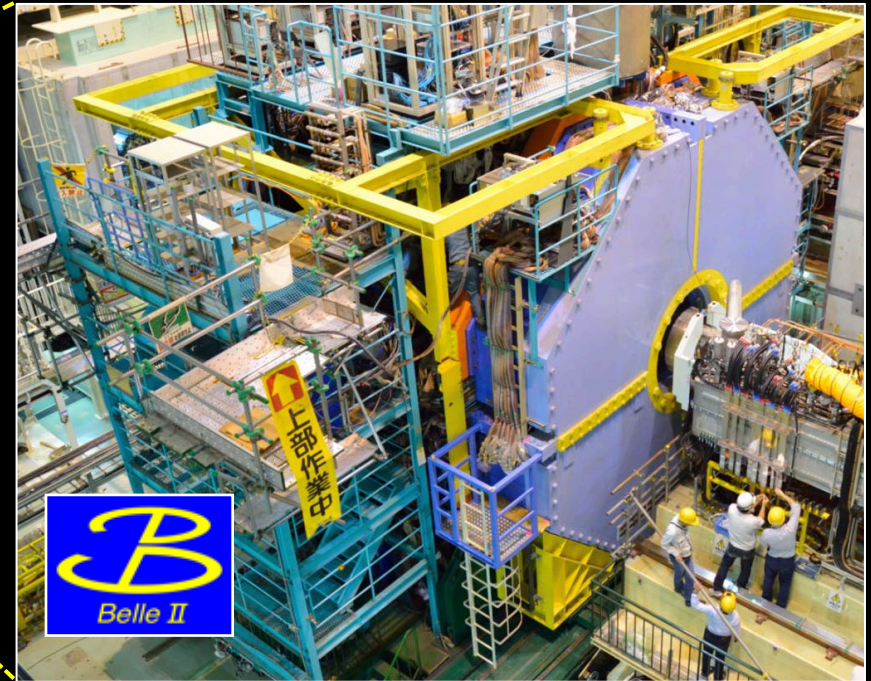
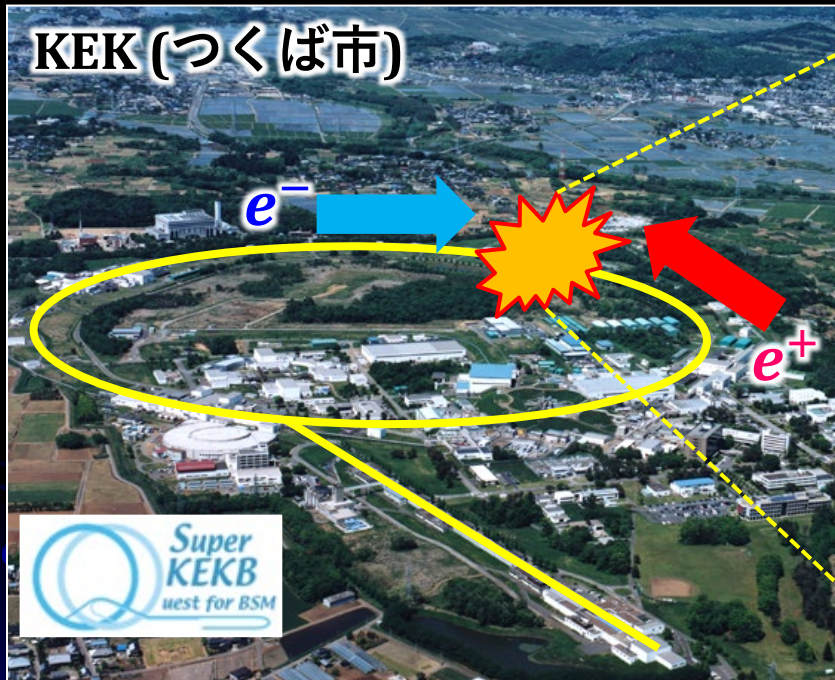
入射線形加速器

KEK 敷地

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)

● 大量の素粒子反応を生成

大型加速器を用いた素粒子実験

5

- ▶ 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

機械学習の適用が有効

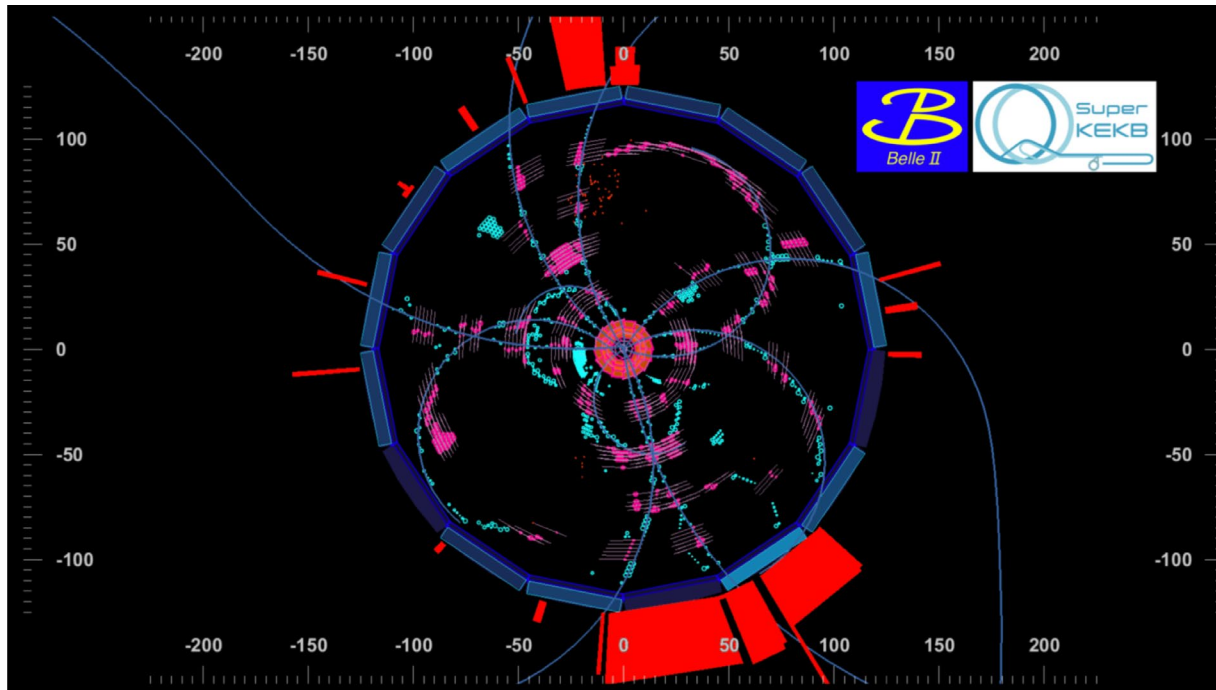
従来より**高性能**、**高速**なデータ処理が期待

強力なデータ処理ツール



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

機械学習 (Machine Learning)

人工知能(AI, Artificial Intelligence)

Any technique that enables computers to mimic the intelligence or behavioral pattern of humans

機械学習(ML, Machine Learning)

A subset of AI. Techniques that enable computers to learn the best rules to solve tasks.

教師あり学習

教師なし学習

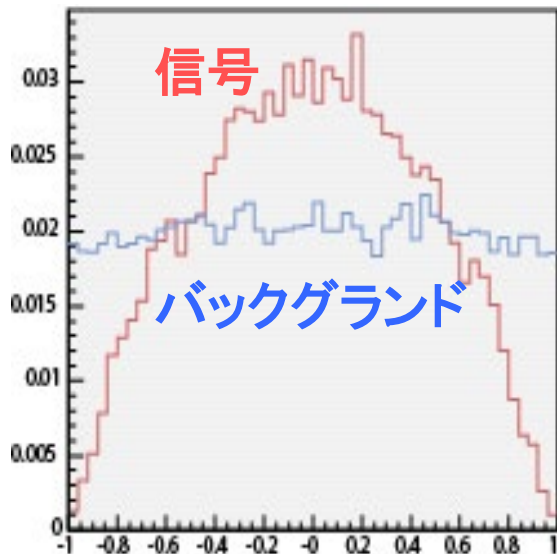
強化学習

機械学習 (Machine Learning)

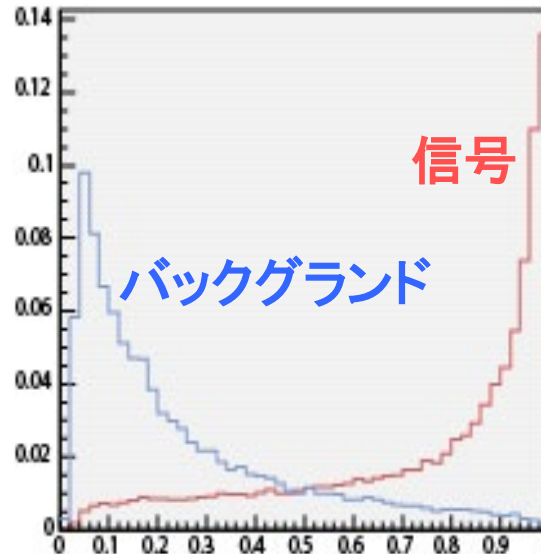
- 人工知能 (AI) の手法の1つ
機械 (コンピュータ) に **学習** させて **認識・判断** を行わせる
- 「**学習処理**」と「**判定処理**」の二つの手順で行われる
新人に、仕事の内容を教える (学習処理)、仕事をさせる (判定処理)

例: 加速器実験でのデータ解析 “信号事象選別処理”

パラメータA



パラメータB



学習

信号は、
Aの絶対値が小さい
Bの値が大きい

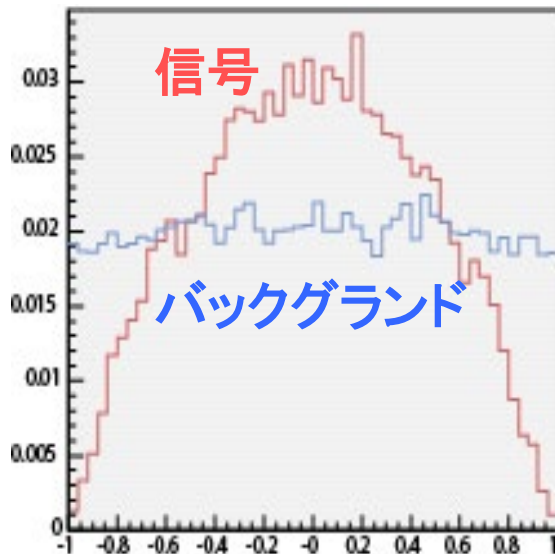
判定

$|A| < 0.6$
 $B > 0.5$
だったら、信号だと判断する
(信号事象選別)

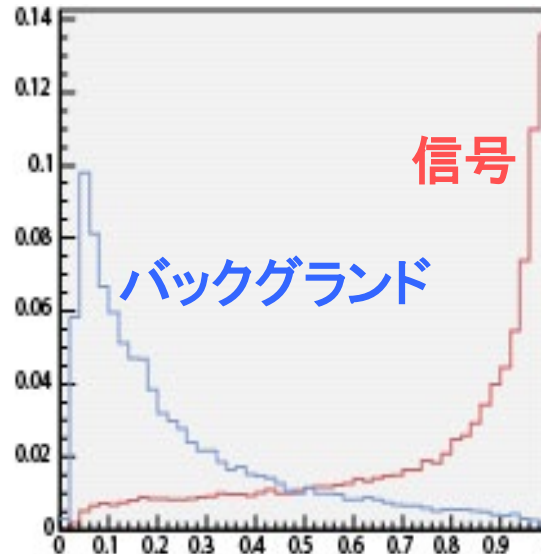
学習

- データを統計処理することで「モデル」を構築
- データの統計量が多いほどモデルの精度が向上
- データの統計量と信頼性が大切

パラメータA



パラメータB



学習

信号は、
Aの絶対値が小さい
Bの値が大きい

判定

$|A| < 0.6$
 $B > 0.5$
だったら、信号だと判断する
(信号事象選別)

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

機械学習

Supervised Learning

教師あり学習

Task driven

Classification

分類・識別

Regression

回帰

加速器実験で行ってきた機械学習
特徴量(High-level data)を用いた
教師あり学習による分類・識別

Dimensionality
Reduction
次元削減

Clustering

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

機械学習

計算機技術・計算手法の進展により
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

Classification

分類・識別

Regression

回帰

Dimensionality

Reduction

次元削減

Clustering

Real-time decisions

Game AI

Leaning Tasks

Robot Navigator

....

いろいろなタイプの機械学習があります

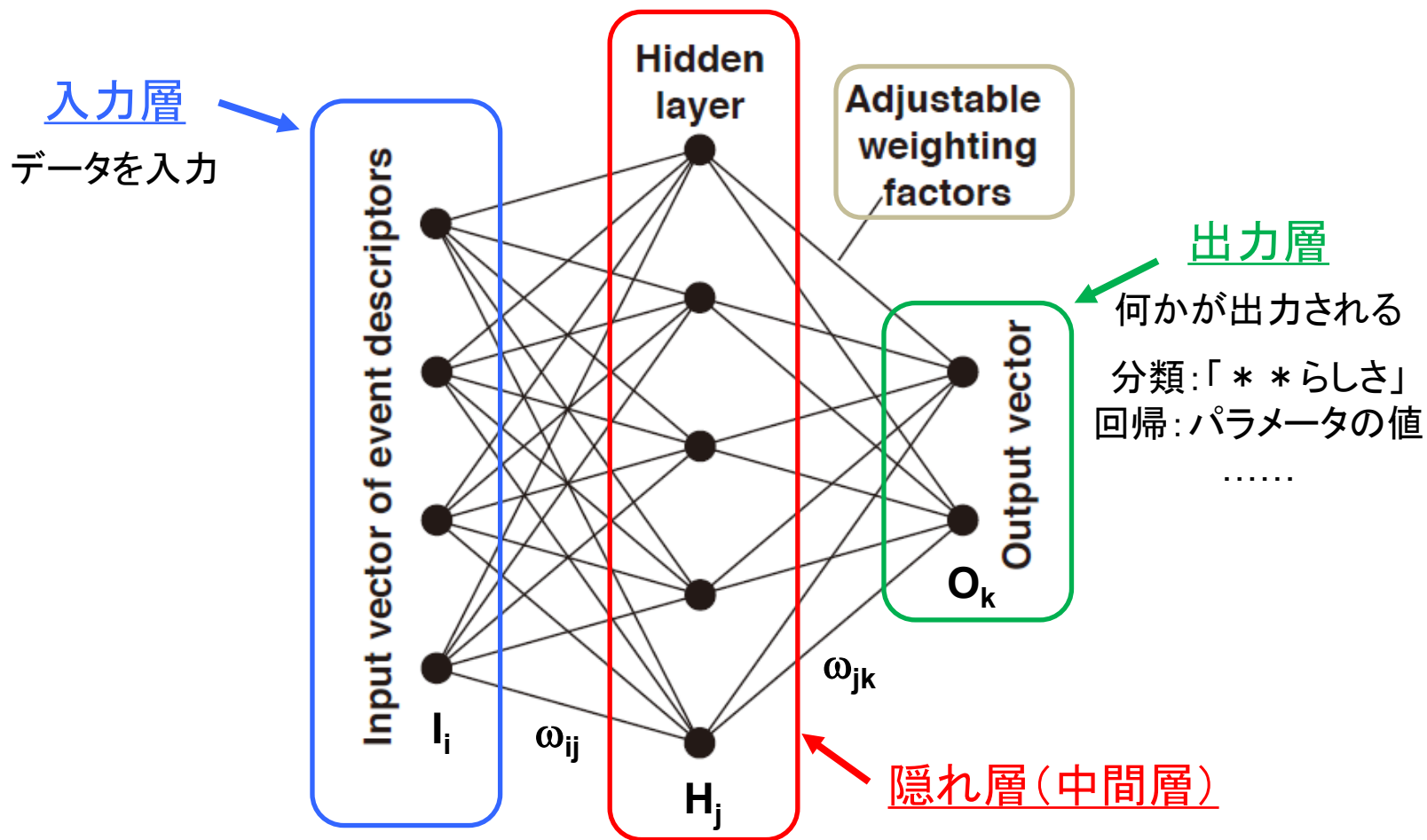
ニューラルネットワーク を用いた機械学習

ニューラルネットワーク(NN)に**学習**させて**認識・判断**を行わせる



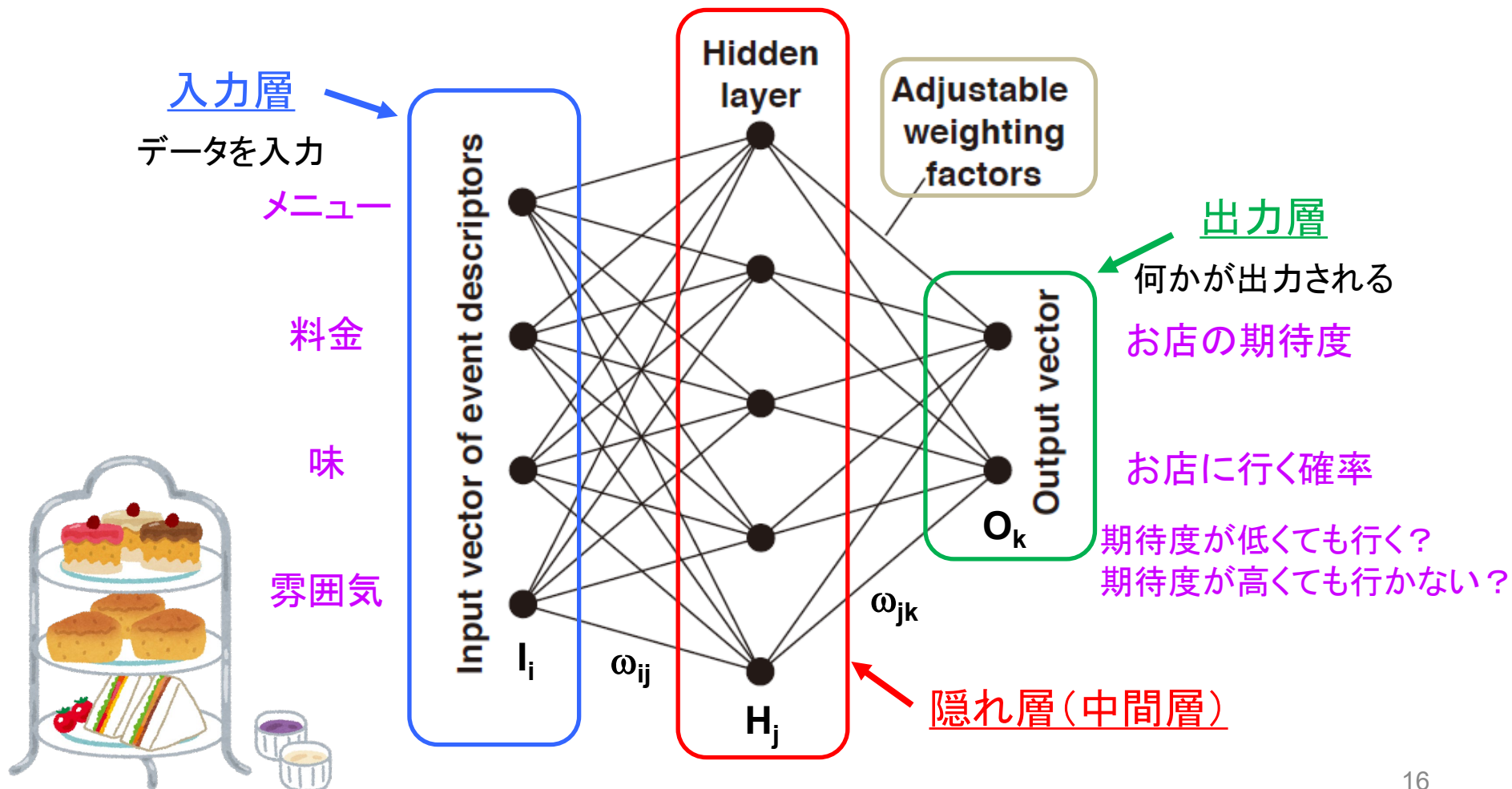
ニューラルネットワーク

Neural Network (NN)とは、脳の情報処理の最小単位である神経細胞ニューロンとそのつながりを模した数式的なモデル



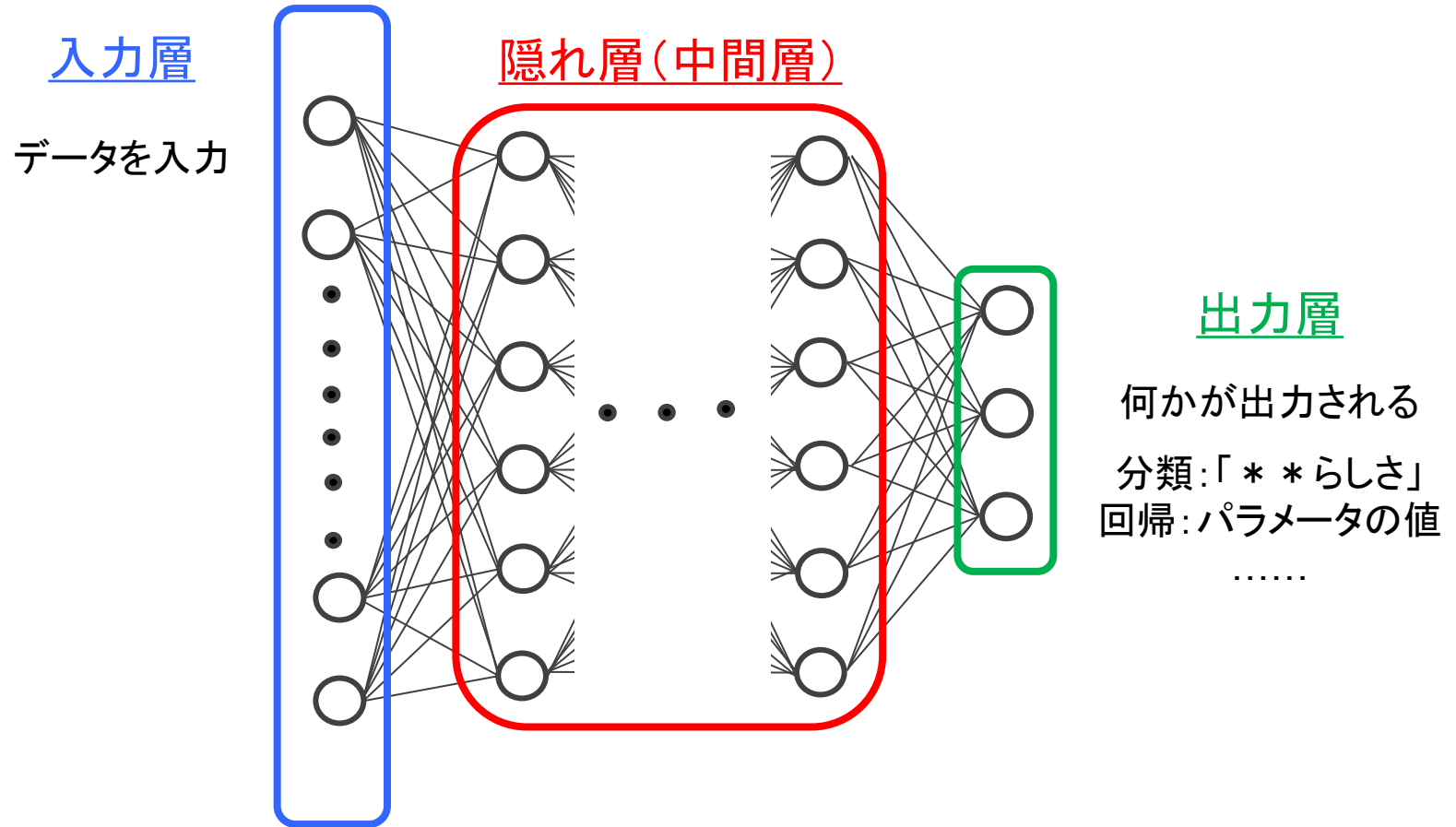
ニューラルネットワーク

Neural Network (NN)とは、脳の情報処理の最小単位である神経細胞ニューロンとそのつながりを模した数式的なモデル



深層学習

NNの隠れ層 (Hidden Layers) が多層化して深くなると深層学習になる



非線形で、より複雑なモデルを表すことが可能になる

最先端機械学習技術：深層学習

Low-level dataによる機械学習を行うためには
深層学習(DNN)の適用が必要

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 ...

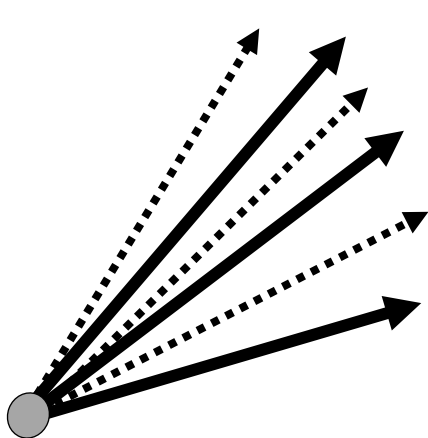
生データを用いた機械学習1

信号識別・フレーバー識別

Osaka-City U., IDS, RCNP

low-level data(粒子の4元運動量、位置情報) + **DNN**を用いた
識別方法を開発

Particle Jet



Jet as low-level data

Part1 (E, P_x , P_y , P_z , x, y, z)

Part2 (E, P_x , P_y , P_z , x, y, z)

Part3 (E, P_x , P_y , P_z , x, y, z)

.....

粒子の4元運動量と位置情報を
NNに入力

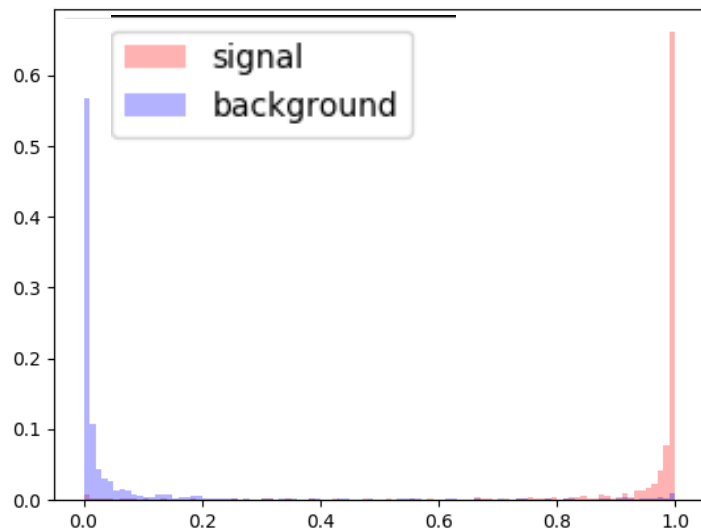
生データを用いた機械学習1

信号識別・フレーバー識別

信号識別性能向上!!

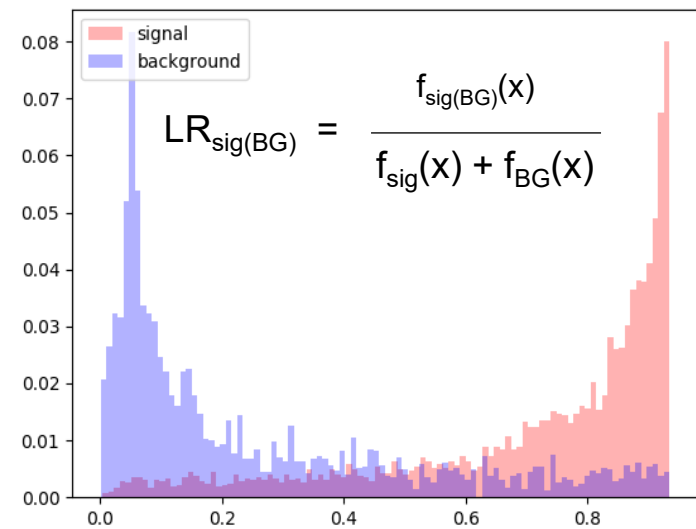
N.Kishida (Osaka City U.)

DNN Output



確率

Likelihood Ratio (LR) using $\cos\theta$, KSFW



LR

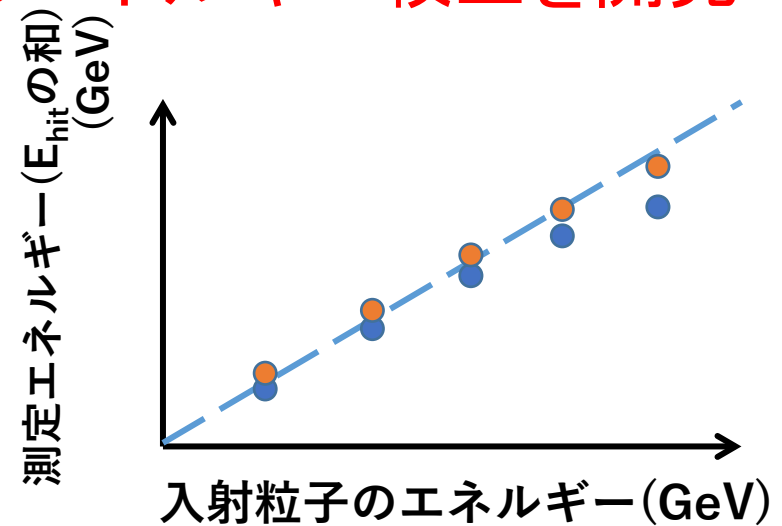
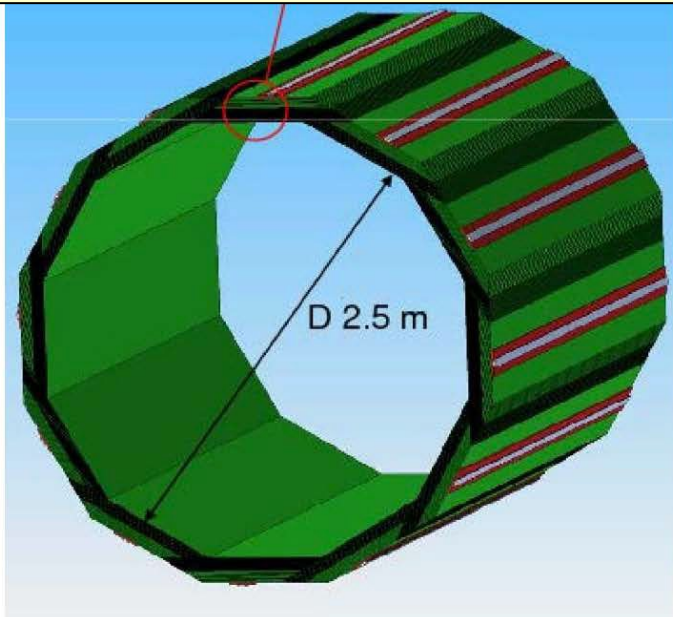
生データを用いた機械学習2

カロリメータのエネルギー較正

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 サンプルング型
- 物質量(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**)

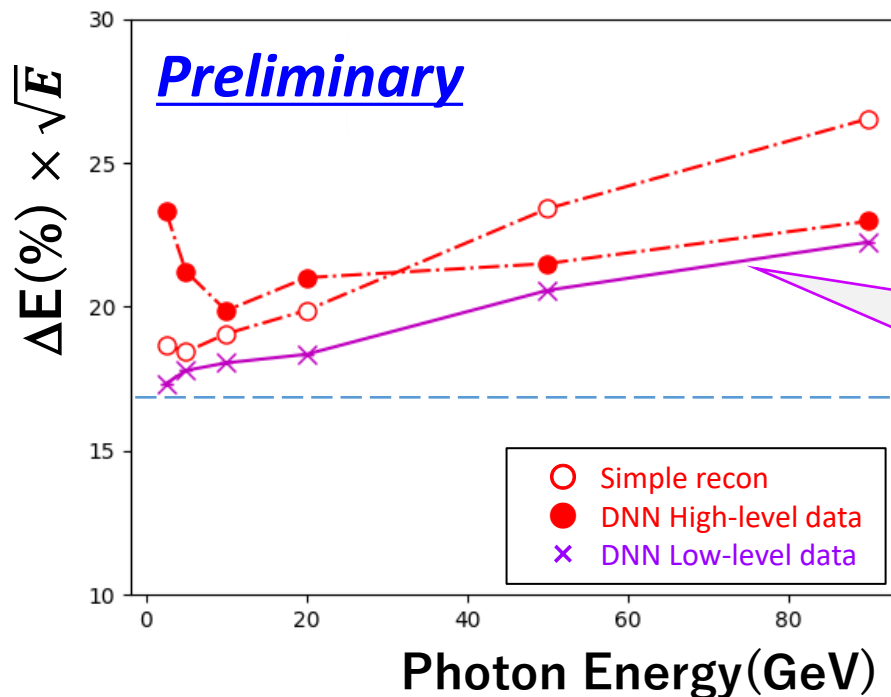
生データを用いた機械学習2

カロリメータのエネルギー較正

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

機械学習を用いた加速器調整

機械学習を用いたKEK Linac加速器調整の開発

KEK Linac、大阪市大、阪大IDS



- 加速器周囲の環境変化により
運転調整が常時必要
- 複雑な調整システム
→ 調整する人によって最適化
のスピードや精度が変わる

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

数百パラメータの調整が必要



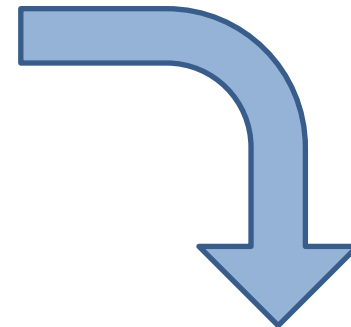
加速器制御の問題点

問題点① 制御システムが複雑

構成要素が多い $\sim O(1000)$

→ 参照するパラメータが多数で複雑

→ 調整速度が個人の技量に依存



機械学習を用いた加速器の可視化

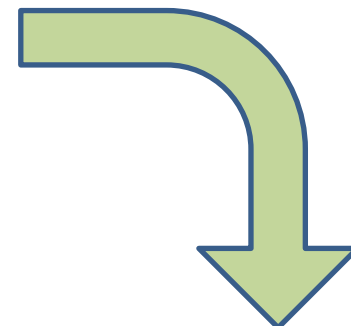
教師なし学習を用いた次元削減(特徴抽出)

問題点② 運転状況が変化する

温度変化や振動、潮汐力等

周囲の環境が変化する

→ 常時調整が必要



環境適応型機械学習(強化学習) を用いた加速器制御

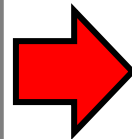
1. 機械学習による加速器状況の可視化

教師なし学習による次元削減（特徴抽出）で
加速器の可視化を行った

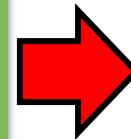
機械学習 (VAE, Variational Auto Encoder)を用いて行った

入力：加速器データ

運転パラメータ：500個
環境パラメータ：732個



機械学習
(次元削減)



出力：潜在変数
 $z(2)$

潜在変数は入力データの特徴を示す

教師無し学習

特徴量

2変数 $z[0], z[1]$

加速器運転・環境データ
 $\sim O(1000)$ パラメータ

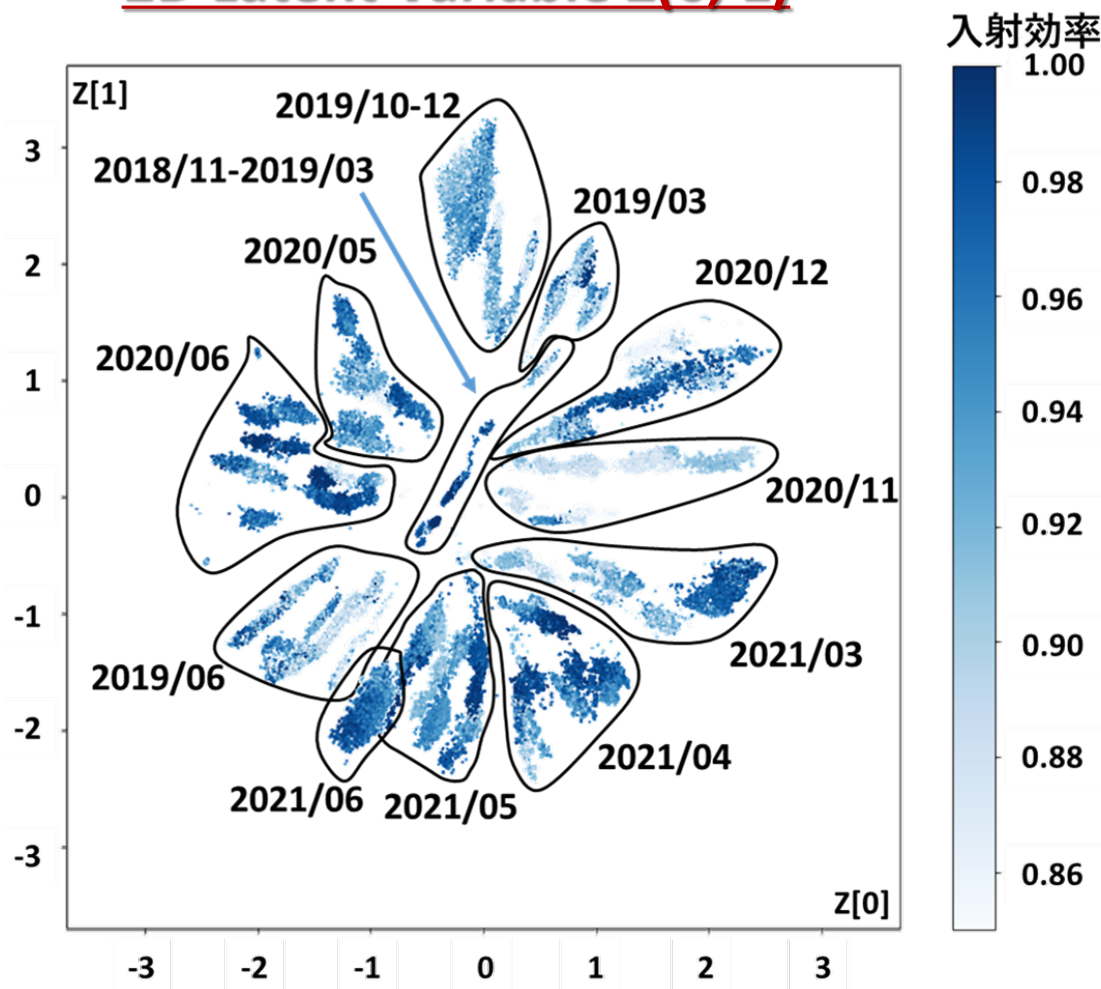
1. 機械学習による加速器状況の可視化

機械学習による次元削減（特徴抽出）で

1232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.)



「加速器の状態」の
可視化に成功

短期間では、加速器の状態
はあまり変化しない

長期間(数か月以上)では
加速器の状態が変化する

2. 強化学習による加速器制御

環境変化に適応した運転調整→ **強化学習**が有効だと思われる

強化学習：周囲の環境変化に応じ、最適な行動を選択する
(報酬を最大化する行動を選択)

→入射効率を報酬とした強化学習を導入したい

強化学習を導入するために、以下の準備開発を進めている

1. 機械学習を用いた **入射効率の予測**

加速器パラメータと入射効率の相関を学習可能か評価

2. 機械学習を用いた **加速器シミュレーターの作成**

Pre-trainingのために、加速器シミュレーターが必要

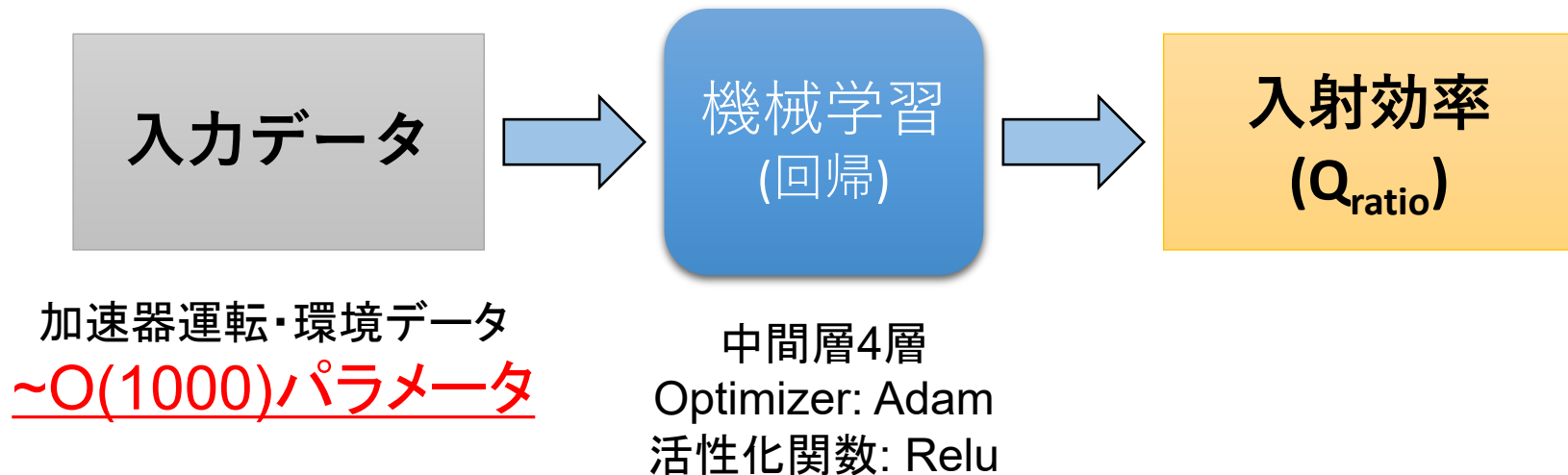


入射効率を最大化するために、実際に加速器で調整することは危険
(車の運転で、ハンドルを左右に切ってから進む方向を決めるのは危険)

2. 強化学習による加速器制御: 準備研究

2-1: 機械学習を用いた入射効率の予測

加速器パラメータを入力すれば
入射効率を予測する機械学習（回帰）で、入射効率を予測した



2. 強化学習による加速器制御: 準備研究

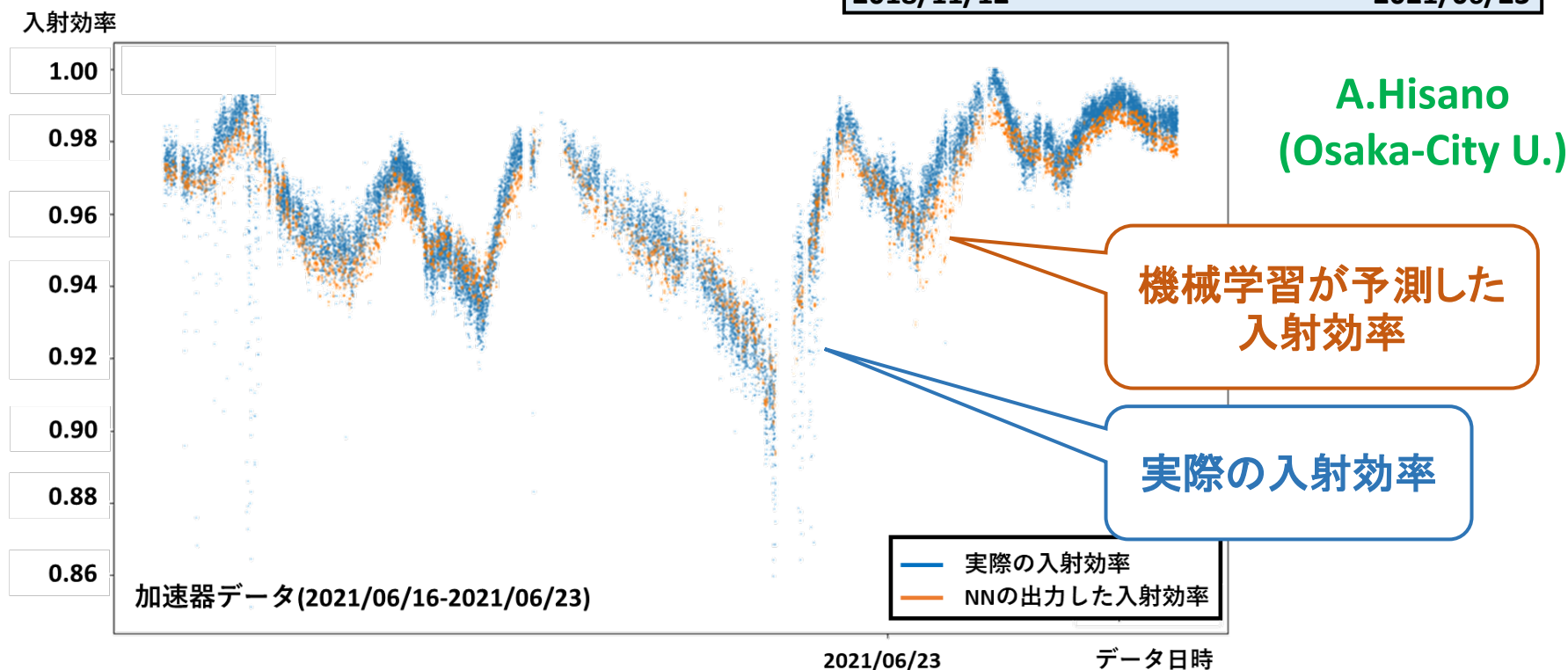
2-1: 機械学習を用いた入射効率の予測

入射効率予測結果 その1

学習・検証データ 同期間 (2018年11月～2021年6月)

2018/11/12	学習データ	2021/06/23
------------	-------	------------

2018/11/12	検証データ	2021/06/23
------------	-------	------------



非常に良く一致 → 加速器パラメータと入射効率の関係を学習可能

2-1: 機械学習を用いた入射効率の予測

入射効率予測結果 その2

検証データは、2021年6月

学習用データは、以下の3パターンを使用 (学習用データは、それぞれ15万イベント)

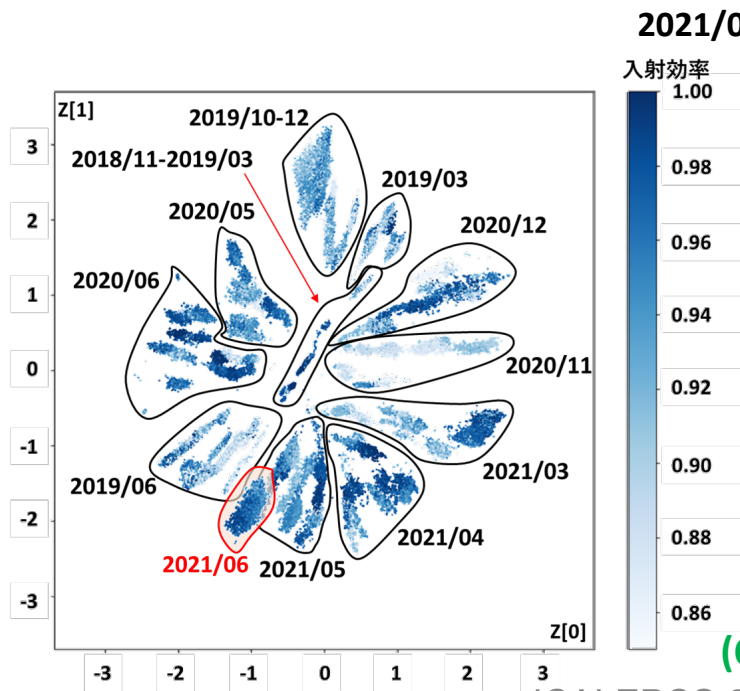
2018/11/12

2021/05/26

2021/06/16

2021/06/23

データ①	学習データ	検証データ
データ②	使用しない	学習データ
データ③	学習データ	使用しない



2021年6月の加速器パラメータの状態は、
2021年5月の加速器パラメータの状態と類似

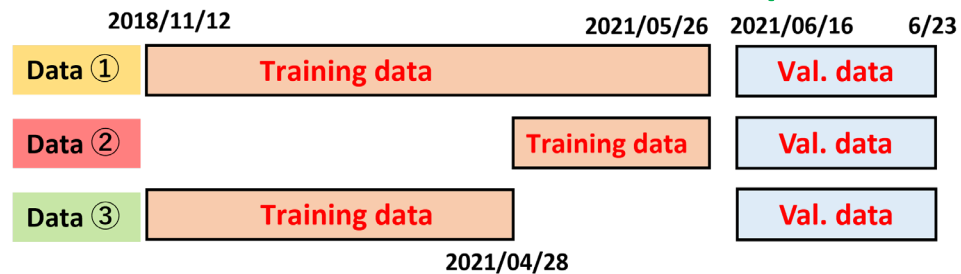
→ 2021年5月のデータで学習すると、
2021年6月の入射効率が予測可能？！

A.Hisano
(Osaka-City U.)

機械学習を用いた加速器運転調整：入射効率の予測

A.Hisano (Osaka-City U.)

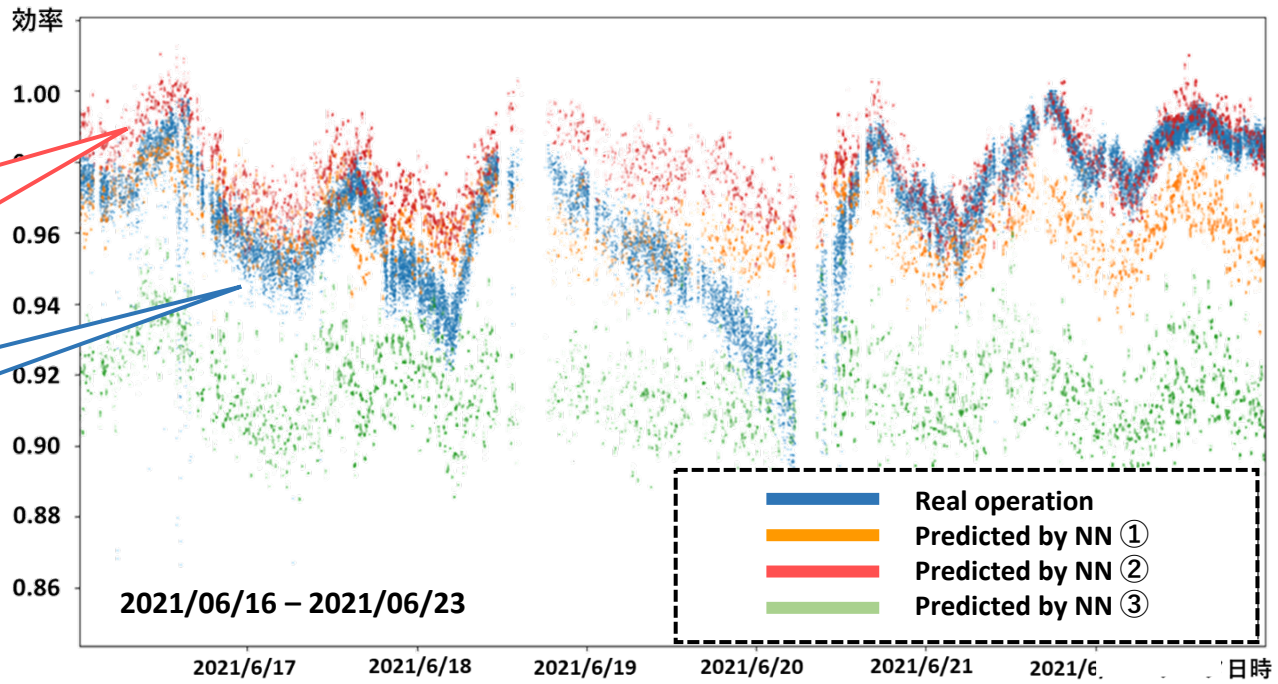
入射効率予測結果 その2



入射効率

機械学習が予測した
入射効率

実際の入射効率



2021年5月を含むデータ(①②)で学習すると
2021年6月の入射効率をほぼ予測可能

直近データによる学習が重要

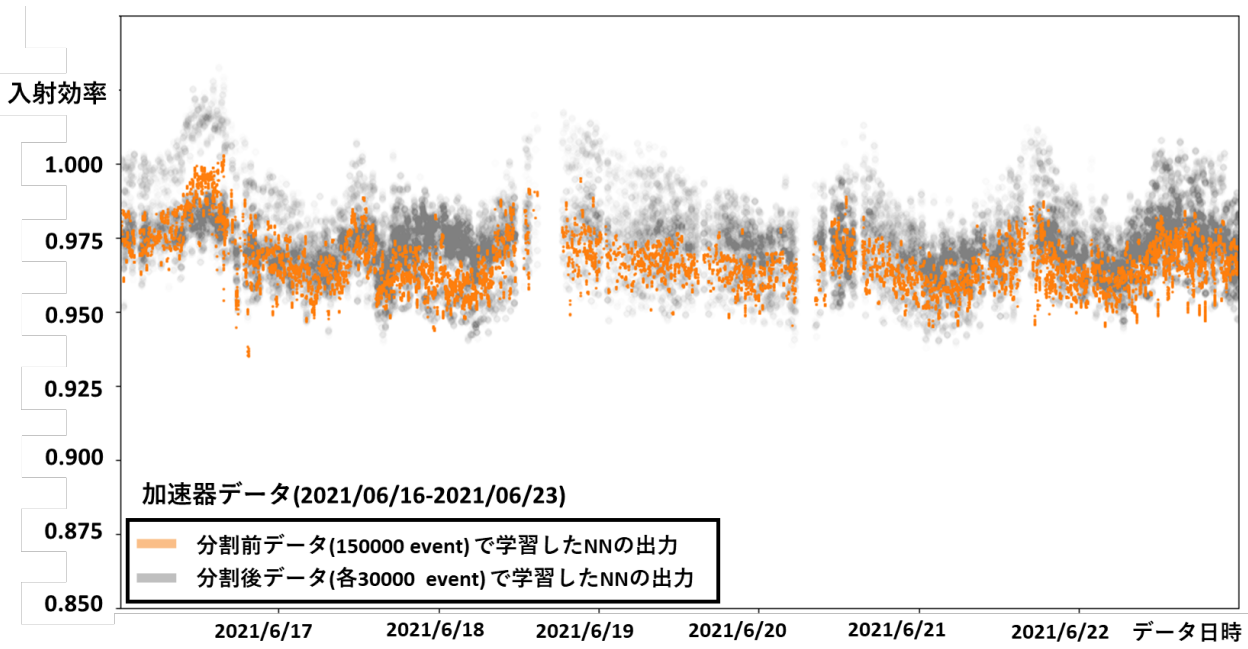
機械学習を用いた加速器運転調整：入射効率の予測

KEK Linac、大阪市大、阪大IDS

入射効率予測結果 その3

検証データは、2021年6月
学習用データは、学習データ①を5分割
(それぞれ3万イベント)

2018/11/12	2021/05/23	2021/06/16	2021/06/23
学習データ 150000 event	30000 event	検証データ	
	30000 event		
	30000 event		
	30000 event		
	30000 event		



A.Hisano
(Osaka-City U.)

DNN が予測した入射効率と実際の
入射効率の平均二乗誤差

分割前(150000 event)	Ave 0.00036
分割後(30000 event)	Ave 0.00043
	Ave 0.00036
	Ave 0.00045
	Ave 0.00067
	Ave 0.00038

学習用データの統計が高い方が入射効率の予測精度が高くなる

→ 加速器シミュレータで学習用データの統計量を確保することが有効?!

2. 強化学習による加速器制御: 準備研究

2-2: 機械学習を用いた加速器シミュレータの作成

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

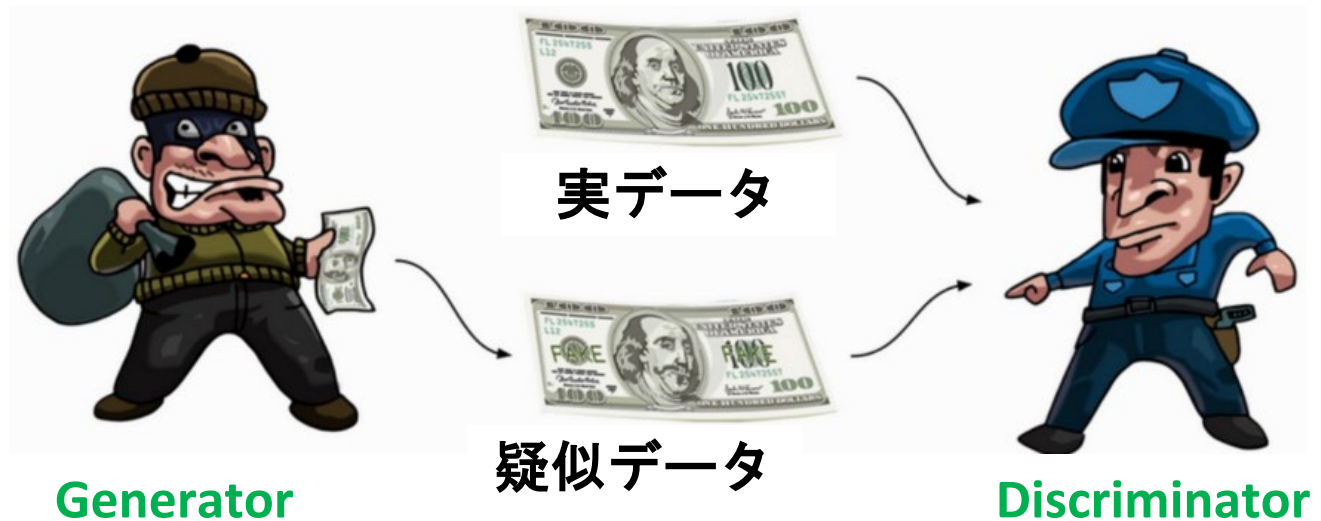


GANを用いた加速器シミュレーターの開発を進めている

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

入力データを模倣して
疑似データを生成

互いに競合する
2つのネットワーク
で構成



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

まとめ

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

最先端機械学習を適用することで
より高性能なデータ処理が可能



加速器実験の
新たな基盤技術への発展が期待