大阪公立大学
Osaka Metropolitan University

機械学習を用いた トップクオークの識別手法の開発

© Osaka Metropolitan University All Rights Reserved.

吉岡優[1)], 岩崎昌子[1)2)], 田中純一[3)] ,齊藤真彦[3)] 末原大幹 [3)]

1)大阪公立大学 2)大阪大学 RCNP 3)東京大学 ICEPP

イントロダクション：本研究の概要

大阪公立大学
Osaka Metropolitan University

本研究ではLHC実験「Top Quark Tagging Reference Dataset」に公開されているシミュレーションデータを用いて機械学習を用いたTopクォークの識別手法を開発した

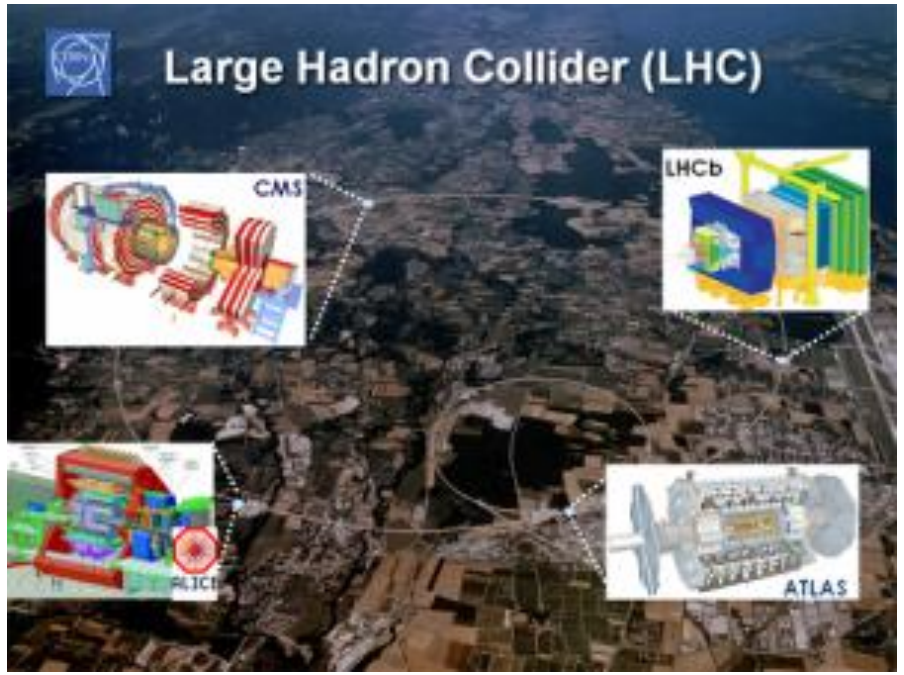
- 機械学習手法としてNN(ニューラルネットワーク）を用いてTopクォークの識別を行った
- 同様の先行研究はたくさん行われているが、本研究では比較的単純なNN・GNN(グラフニューラルネットワーク)により、良い識別性能を得ることを目指す
- 本研究ではNN ・GNNの識別性能を向上させるためにそれぞれ最適化されたネットワーク構造を使用している

© Osaka Metropolitan University All Rights Reserved.

イントロダクション：LHCについて

大阪公立大学
Osaka Metropolitan University

- Large Hadron Colliderの略称
- CERN(欧州合同原子核研究機構）に建設された世界最高エネルギーの陽子・陽子衝突加速器。
- 円周の長さは27km。
- 重心系エネルギー13TeV



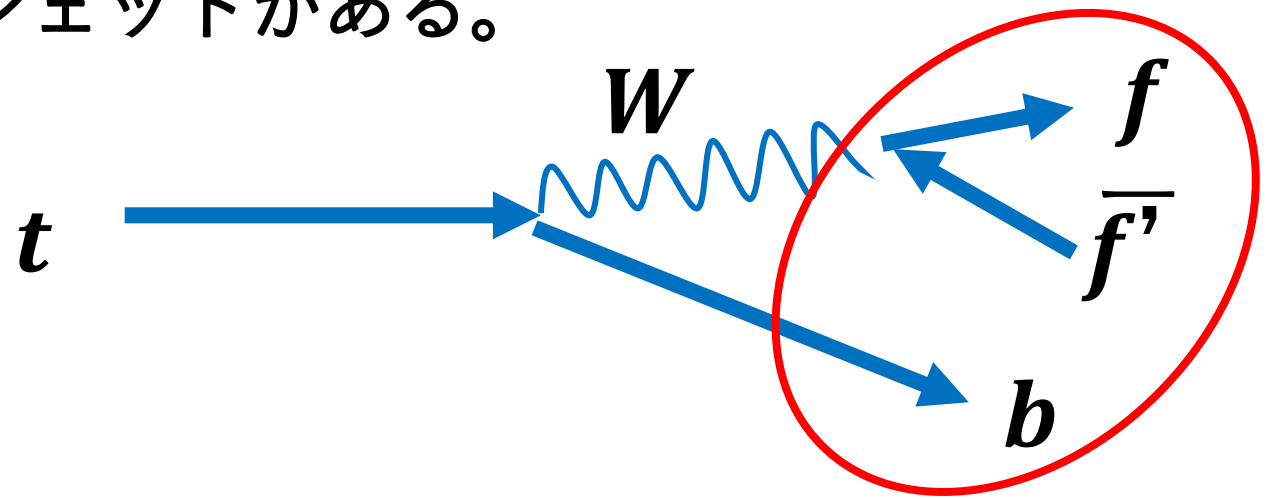
引用：http://alice-j.org/lhc%E5%8A%A0%E9%80%9F%E5%99%A8/

© Osaka Metropolitan University All Rights Reserved.

イントロダクション

LHC実験でのトップクオーク生成の識別を行った。

- Topクオーク崩壊にはbクオークに由来するジェットとW崩壊に由来する2ジェットがある。



この3つが
まとまっている

今回使用したではデータではすべて一つのジェットの中に集約されている。

- 実験ではTopクォークの信号以外にも膨大な量のバックグラウンド事象があり、これをQCDと呼んでいる。

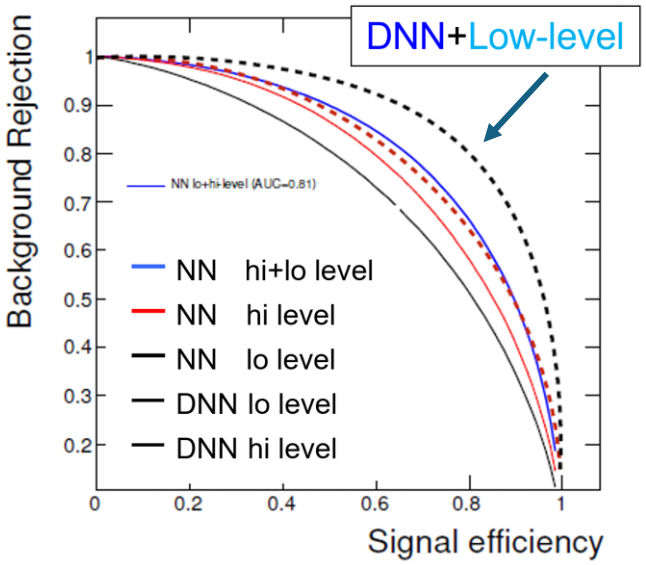
© Osaka Metropolitan University All Rights Reserved.

機械学習を用いた識別手法の開発

大阪公立大学
Osaka Metropolitan University

- 20~30年ほど前から高エネルギー実験では、機械学習を用いた信号識別が行われてきた
- それらの機械学習では、特徴量（特徴を示す物理量, high-level-data）を入力とした比較的単純な機械学習が行われてきた

LHCにおけるb-jet識別



D. Guest, K. Cranmer, D. Whiteson, Annu. Rev. Nucl. Part. Sci. **68** 1-22 (2018)

生データ（Low-level-data）を用いた機械学習を行うと、よりよい識別性能を示すことが分かってきた。

© Osaka Metropolitan University All Rights Reserved.

イントロダクション：先行研究について

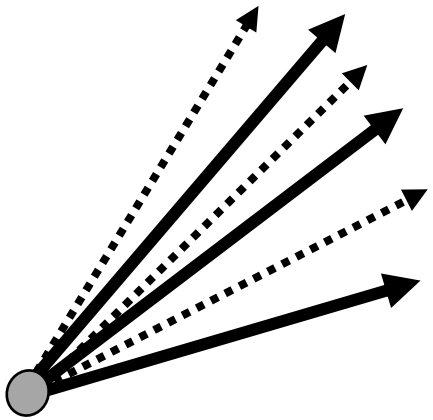
大阪公立大学
Osaka Metropolitan University

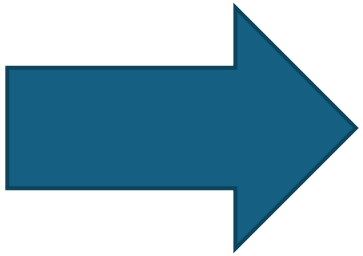
データサイエンスの分野では、画像や自然言語などの解析のために、
DNNなどの最新の機械学習技術が発展してきた。
→ 画像処理や自然言語解析のためのツールが数多く存在する。

CNNなどの最新のDNNツールを利用するため高エネルギー実験のデータを
「画像データ」に変換すると

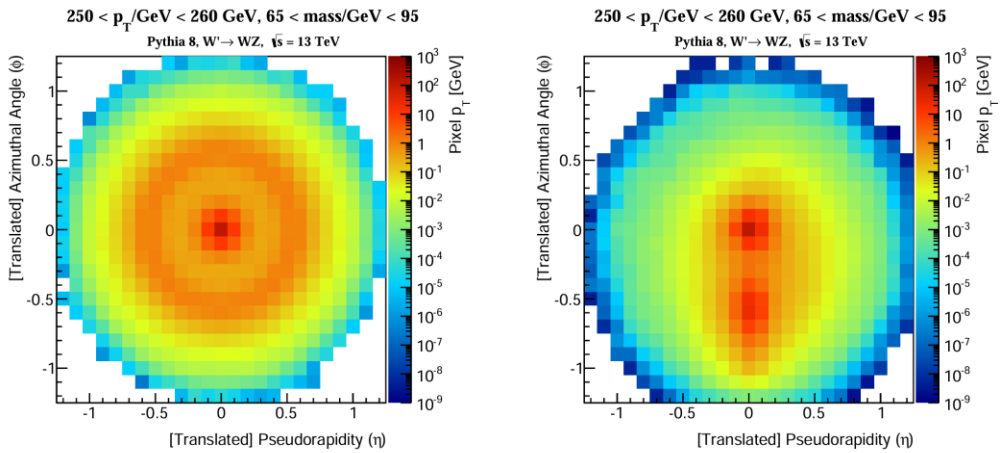
- 位置分解能が低下する
- 大量の学習データが必要になる

粒子ジェット





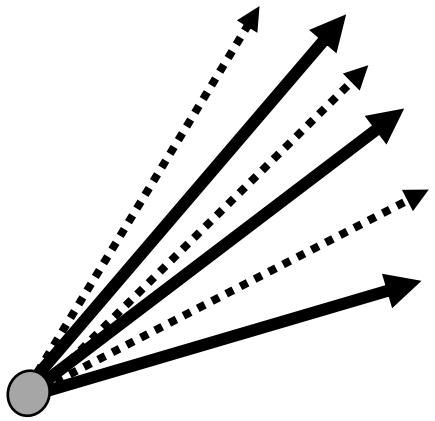
画像化されたジェット

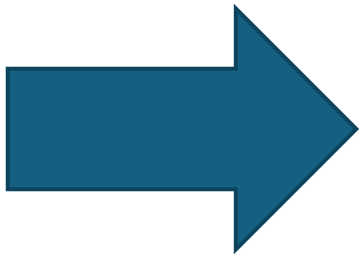


使用したデータの説明

大阪公立大学
Osaka Metropolitan University

- 本研究における生データ(Low-level-data)





ジェットのデータ（粒子数,M,E,p_x, p_y, p_z）
粒子データ1（ E,p_x, p_y, p_z）
粒子データ2（ E,p_x, p_y, p_z）
粒子データ3（ E,p_x, p_y, p_z）
…

本研究では、ジェット内部の粒子の4元運動量をlow-level-data としてNNに直接入力する。

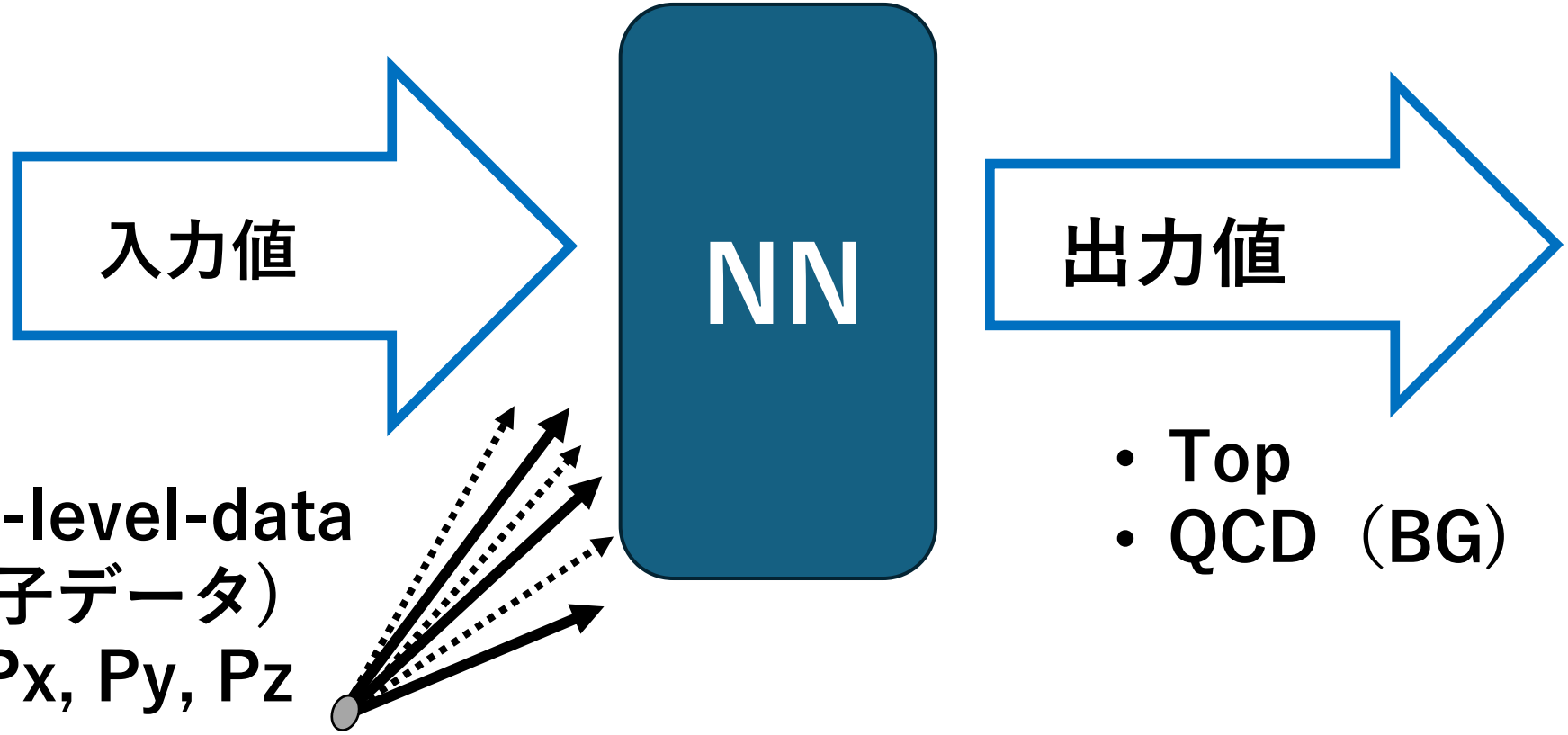
岸田　「$B^0 \rightarrow \gamma\gamma$ 崩壊過程の探索～機械学習を用いた解析手法の開発～」(2020年)
加藤　「高エネルギー実験のための機械学習を用いたスパースサンプリングによるデータ処理技術開発」(2023年)

© Osaka Metropolitan University All Rights Reserved.

NNを用いたLHCデータ識別

大阪公立大学
Osaka Metropolitan University

本研究のNNのイメージ



- Low-level-data
  （粒子データ）
  E, Px, Py, Pz

- Top
- QCD（BG）

本研究ではTop(学習用：約50万イベント, 評価用約20万イベント), QCD(約50万イベント, 評価用約20万イベント)を使用

© Osaka Metropolitan University All Rights Reserved.

識別性能を上げるための工夫①

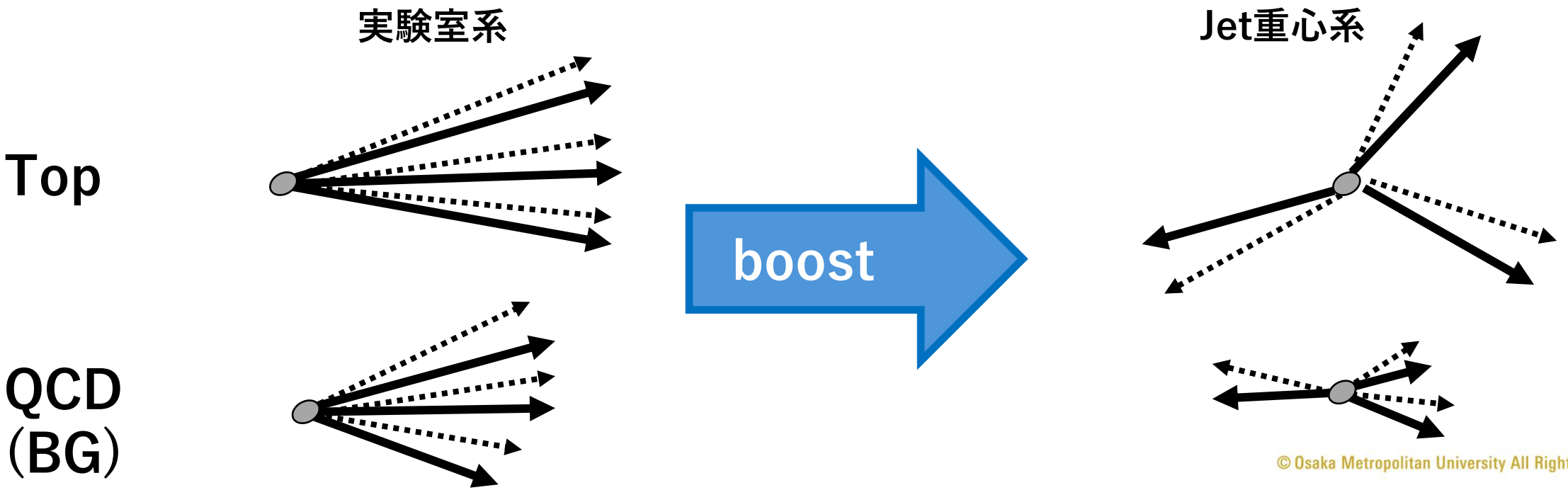
大阪公立大学
Osaka Metropolitan University

Jet重心系へのboost

Topクォークは静止質量が大きい

→Jet重心系ではjetのエネルギーと静止質量が等しくなる

$$M^2 = E_{lab}^2 - P_{lab}^2 = E_{CM}^2$$

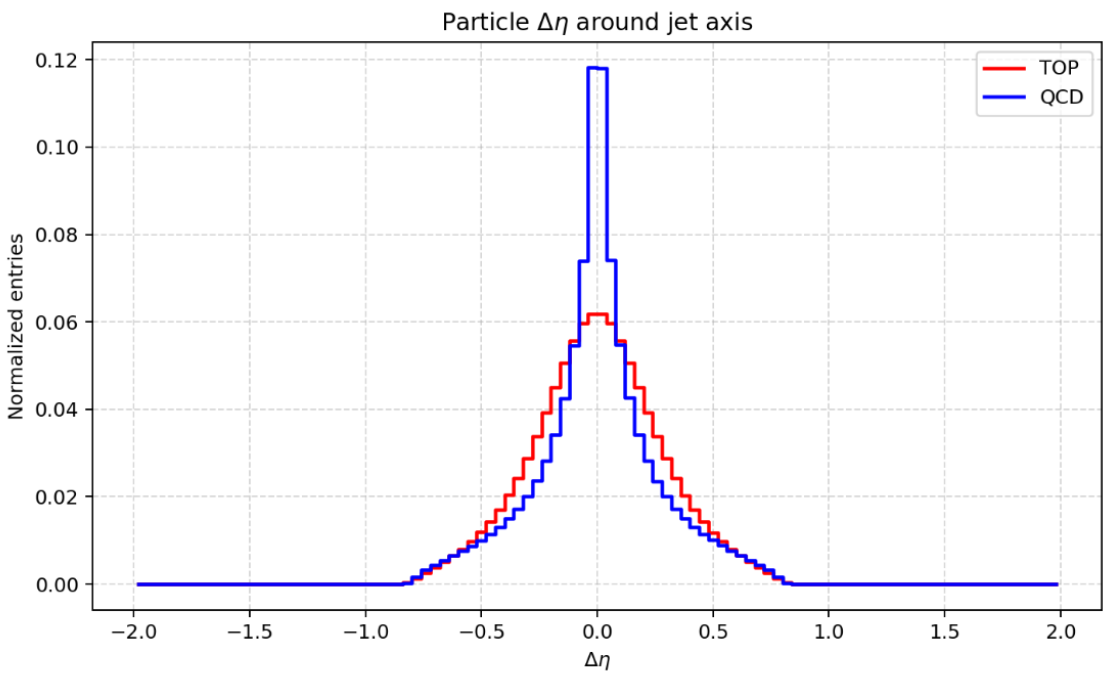


© Osaka Metropolitan University All Rights Reserved.

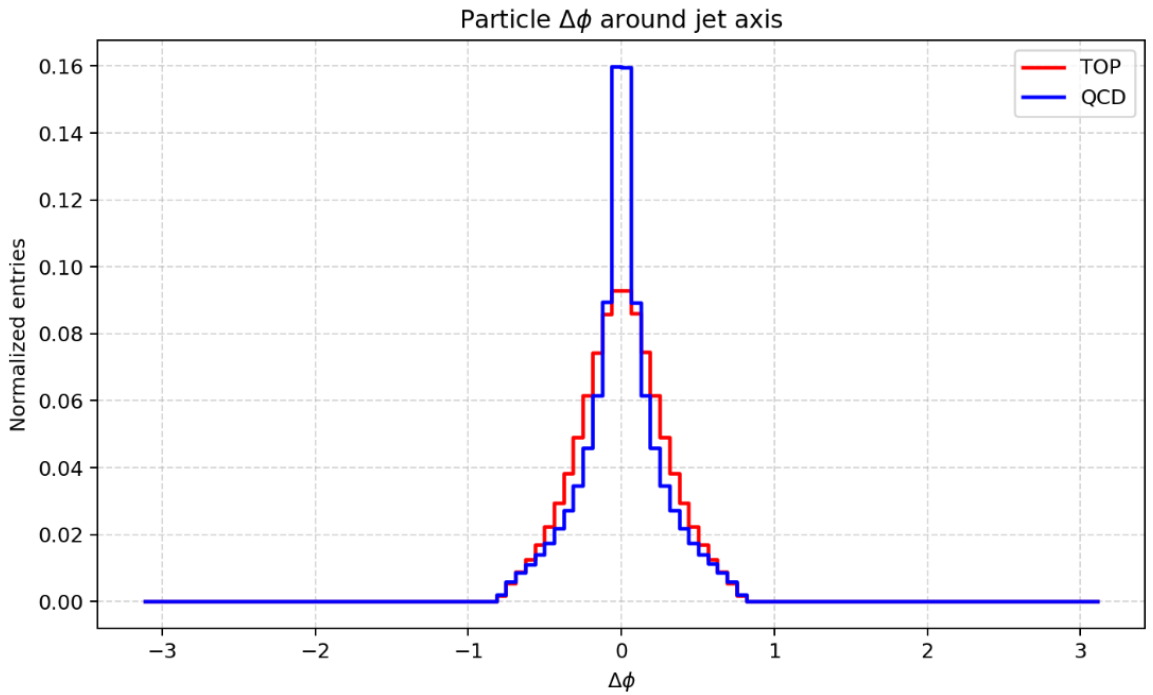
識別性能を上げるための工夫②

大阪公立大学
Osaka Metropolitan University

Δ η, Δ Φを入力に加える



Δ η



Δ Φ

トップジェットとQCDジェットでは粒子のΔ η, Δ Φ分布が異なり、ジェット内部構造の識別に役立つ。

Top
QCD

© Osaka Metropolitan University All Rights Reserved.

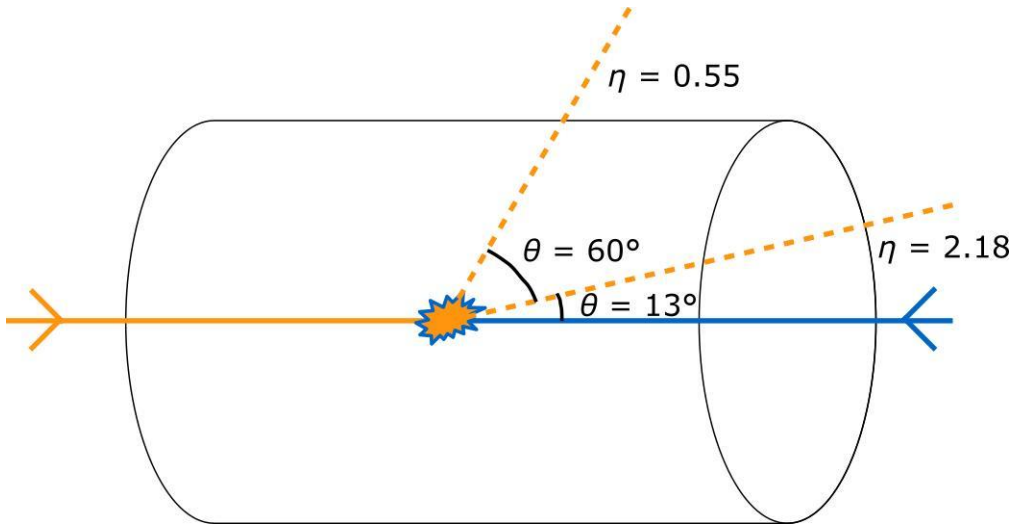
識別性能を上げるための工夫②

大阪公立大学
Osaka Metropolitan University

シュードラピディティ η の定義

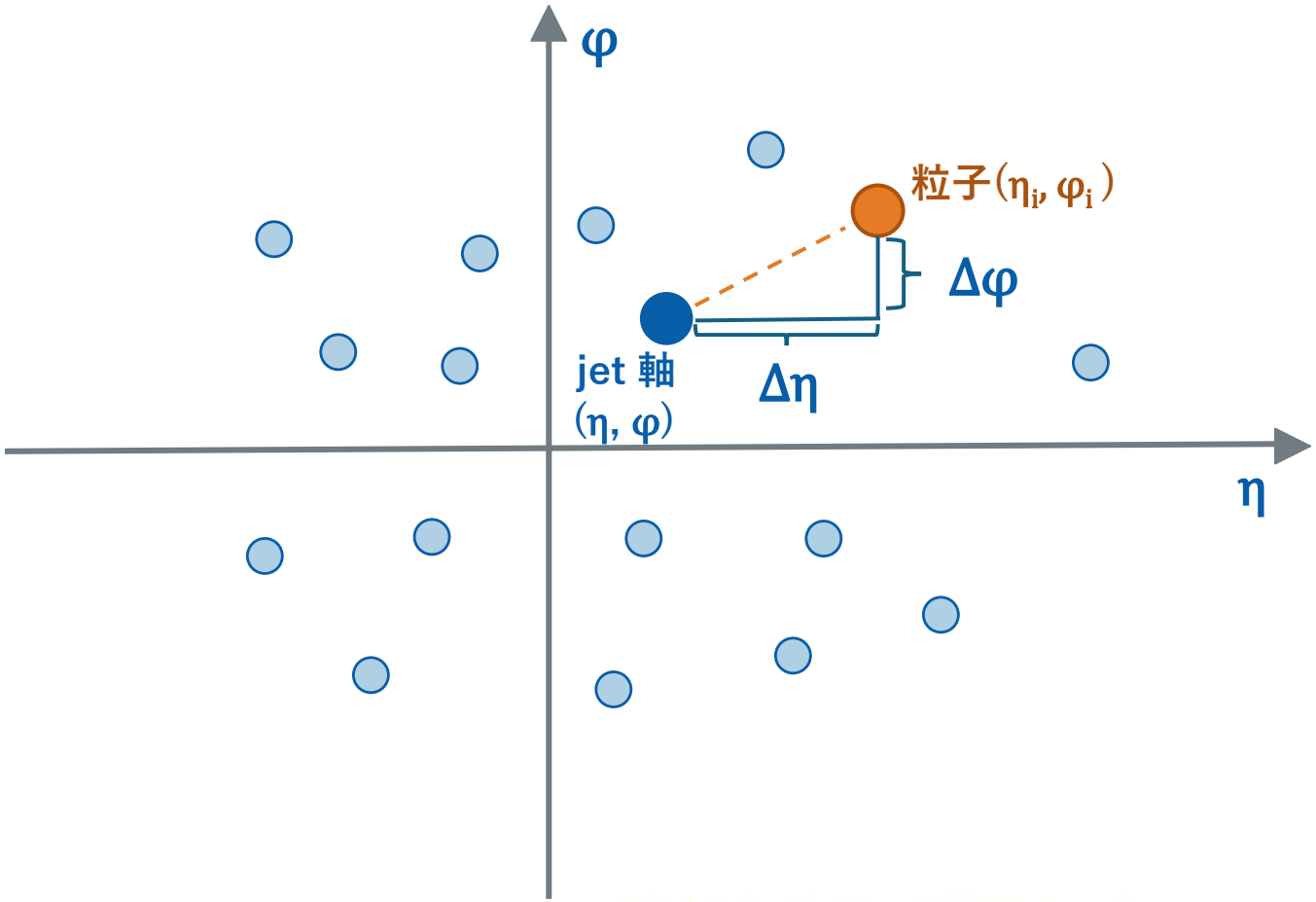
$$\eta = -\ln \tan(\theta / 2)$$

θ はビーム衝突軸（z軸）のなす角



出典：CMS Open Data Education

η, ϕ 平面



© Osaka Metropolitan University All Rights Reserved.

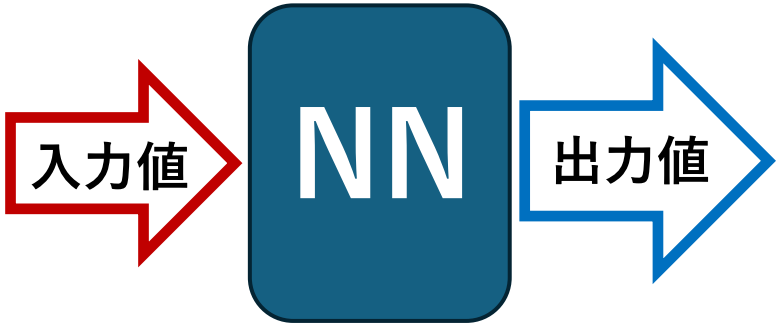
NNを用いたLHCデータ識別

大阪公立大学
Osaka Metropolitan University

NNへの入力パラメータ

低特徴データ(Low-level-data)(粒子データ)

- 静止系　：4元運動量(E, **P**), P_T, |**P**|
- 実験室系：4元運動量(E, **P**), P_T, |**P**|
- 相対角　：$\Delta\eta$, $\Delta\Phi$



特徴量(High-level-data)(ジェットデータ)

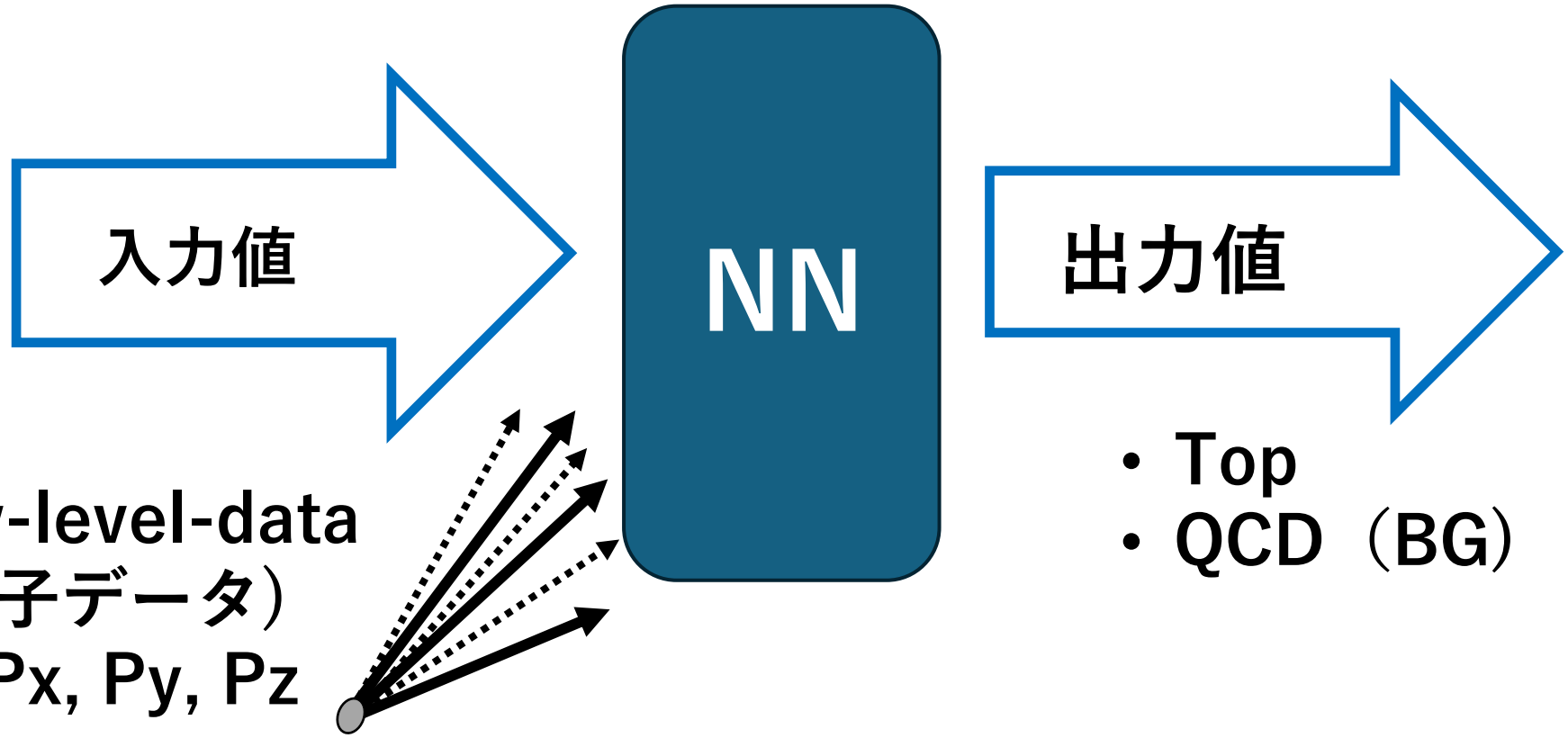
- 実験室系：4元運動量(E, **P**), P_T, |**P**|
- 静止質量M, 粒子数, η, Φ

© Osaka Metropolitan University All Rights Reserved.

NNを用いたLHCデータ識別

大阪公立大学
Osaka Metropolitan University

本研究のNNのイメージ



- Low-level-data
  （粒子データ）
  E, Px, Py, Pz

- Top
- QCD（BG）

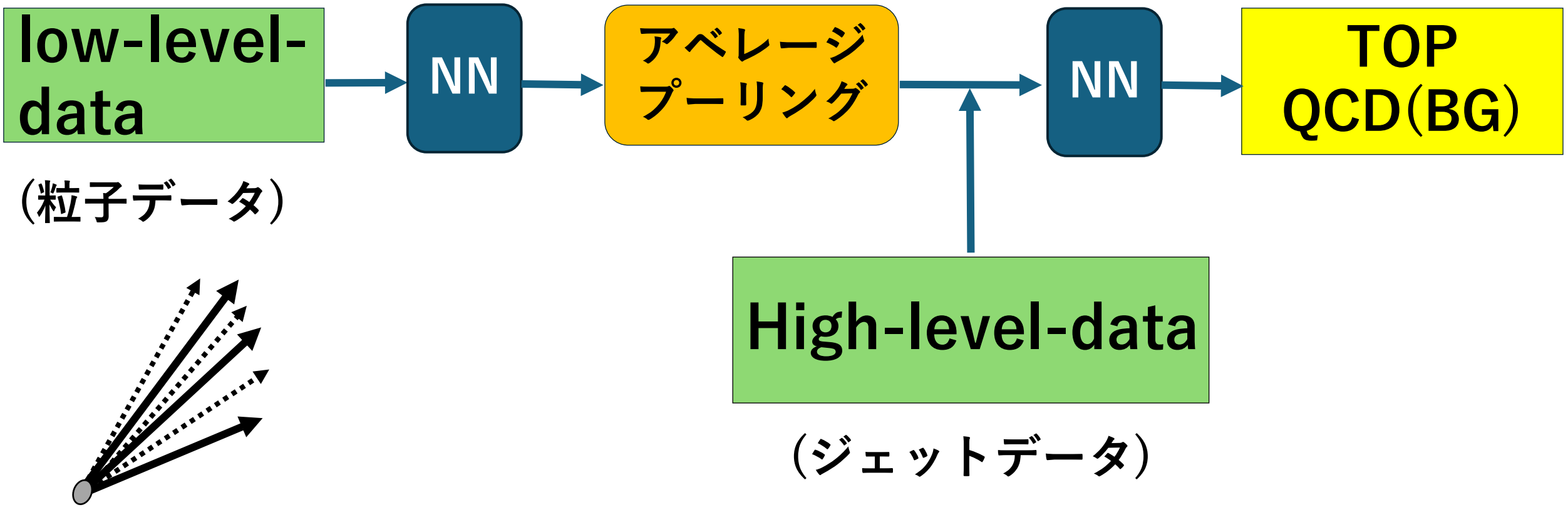
本研究ではTop(学習用：約50万イベント, 評価用約20万イベント), QCD(約50万イベント, 評価用約20万イベント)を使用

© Osaka Metropolitan University All Rights Reserved.

NNを用いたLHCデータ識別

大阪公立大学
Osaka Metropolitan University

本研究のNNのイメージ

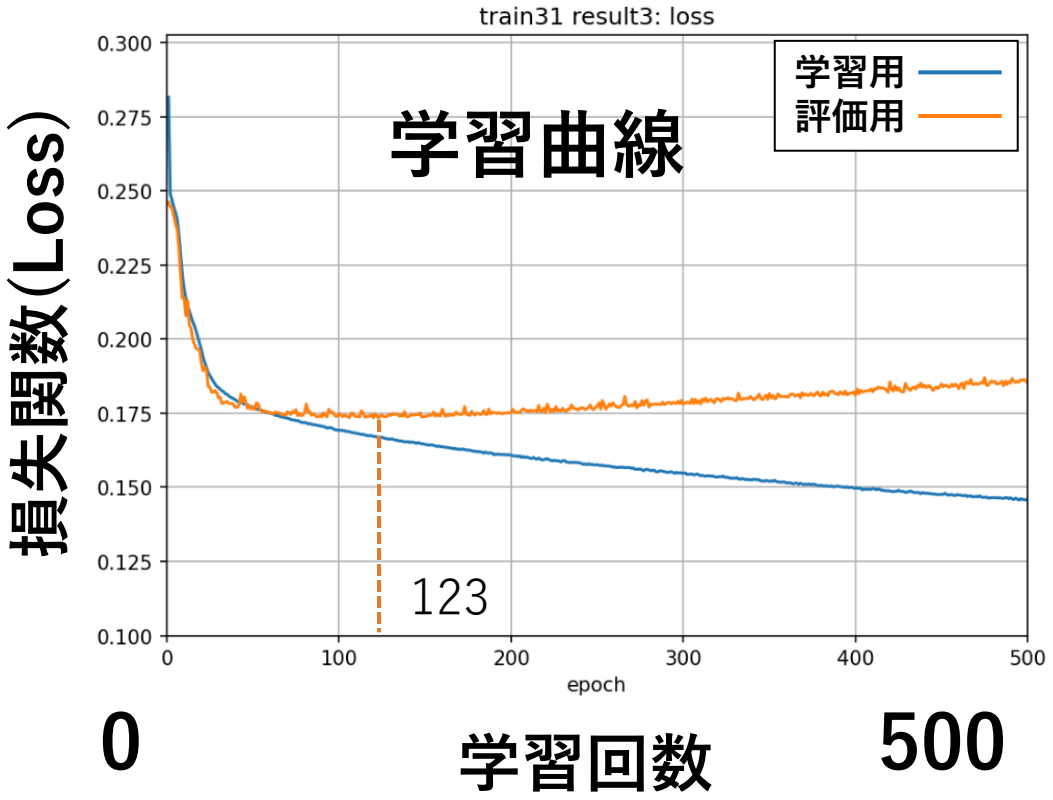


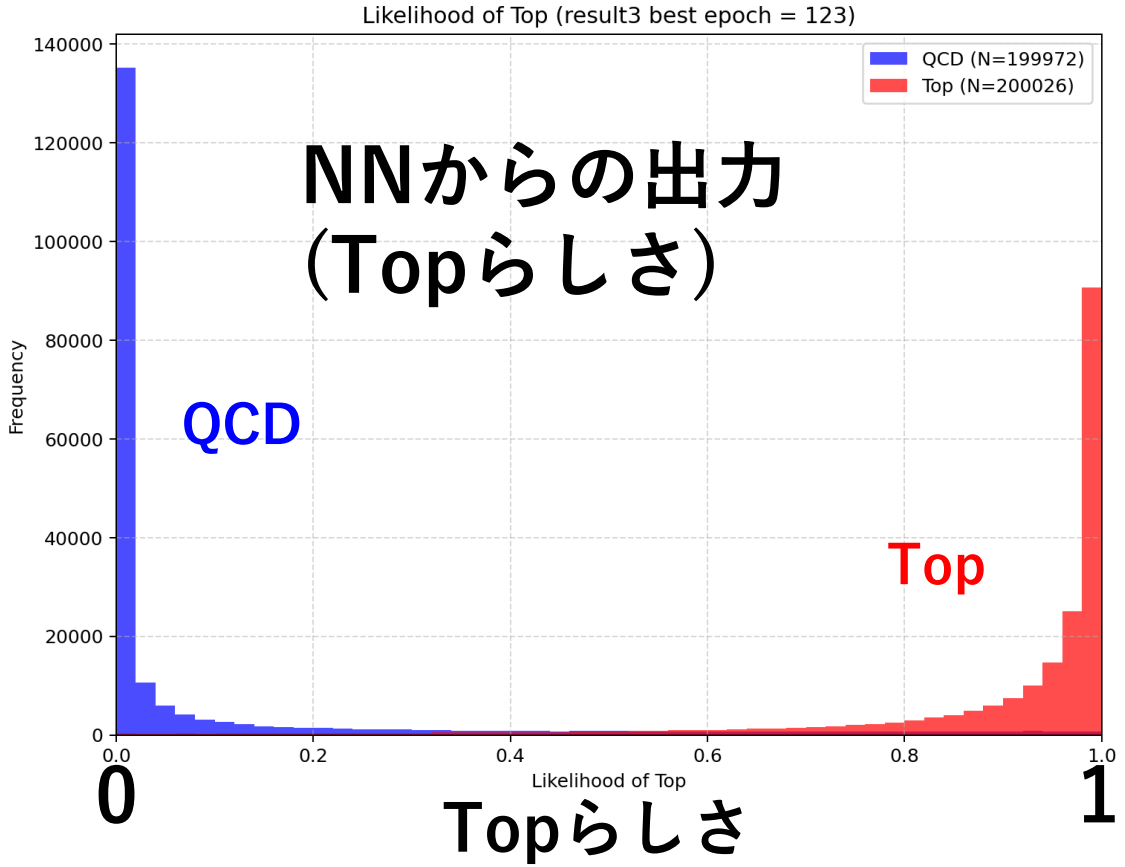
© Osaka Metropolitan University All Rights Reserved.

NNを用いたLHCデータ識別

大阪公立大学
Osaka Metropolitan University

学習の結果

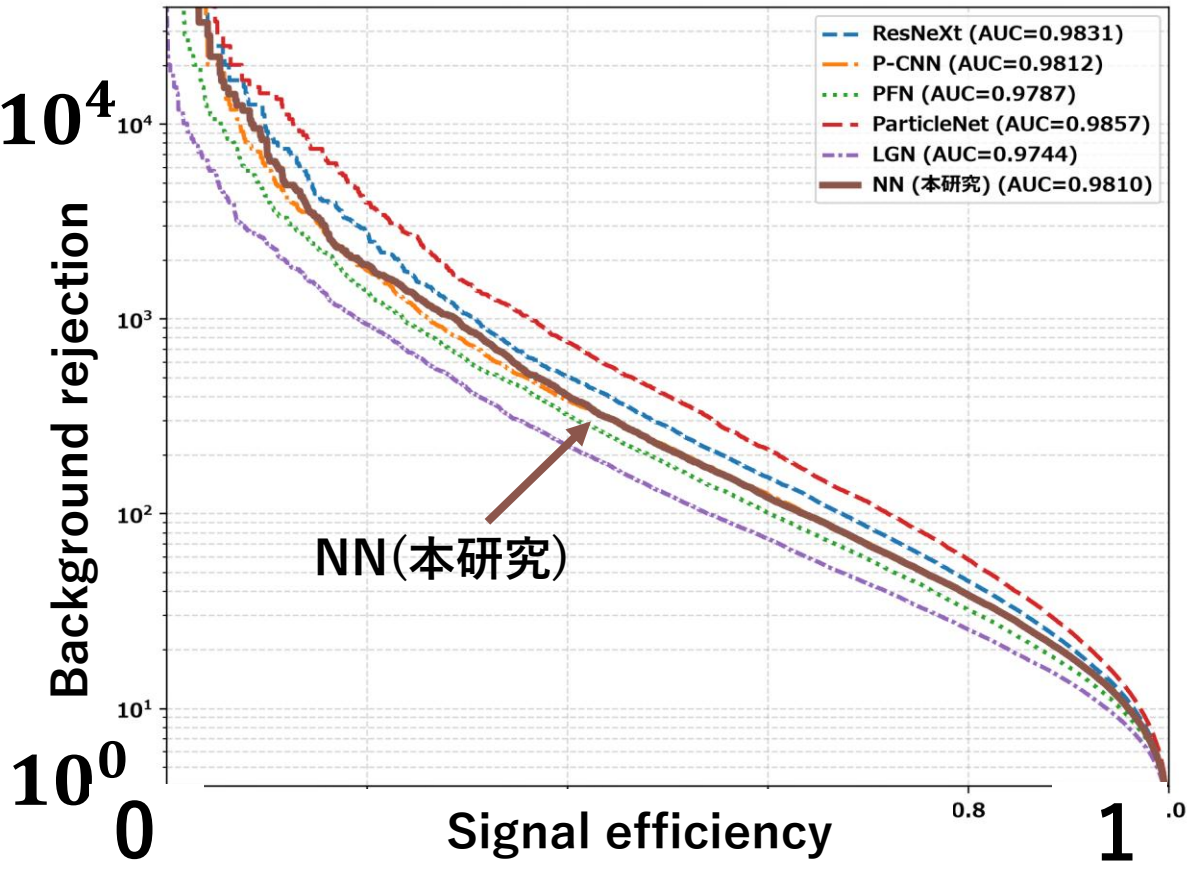




- 学習曲線が安定
- このNNでTopクォークとQCDを識別できている

© Osaka Metropolitan University All Rights Reserved.

先行研究との比較



Model	$\varepsilon\mathcal{S} = 0.3$	$\varepsilon\mathcal{S} = 0.5$
LorentzNet	2195	498
ParticleNet	1530	402
ResNeXt	1052	277
NN(本研究)	**885**	**215**
P-CNN	745	220
PFN	614	179
LGN	438	126

GNN

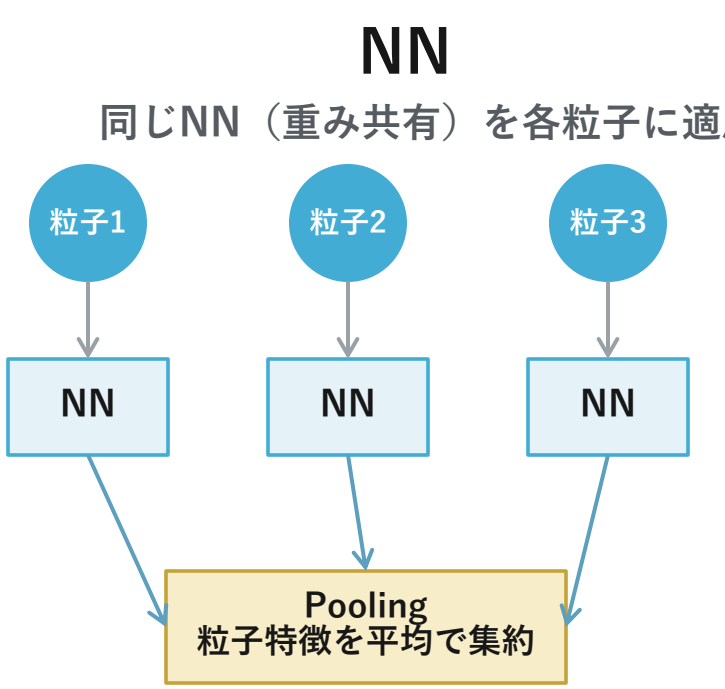
本研究で構築したNN(MLPベース)で、他のNN・CNNを用いた先行研究とほぼ同等の識別性能を示した。

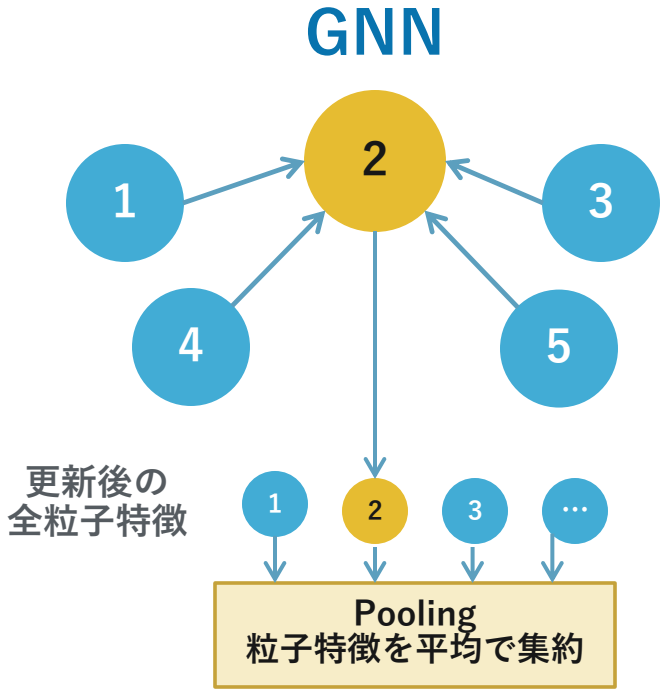
© Osaka Metropolitan University All Rights Reserved.

GNNを導入

大阪公立大学
Osaka Metropolitan University

NNは粒子ごとの特徴を扱うが、粒子間の近傍関係は明示的に扱えない。





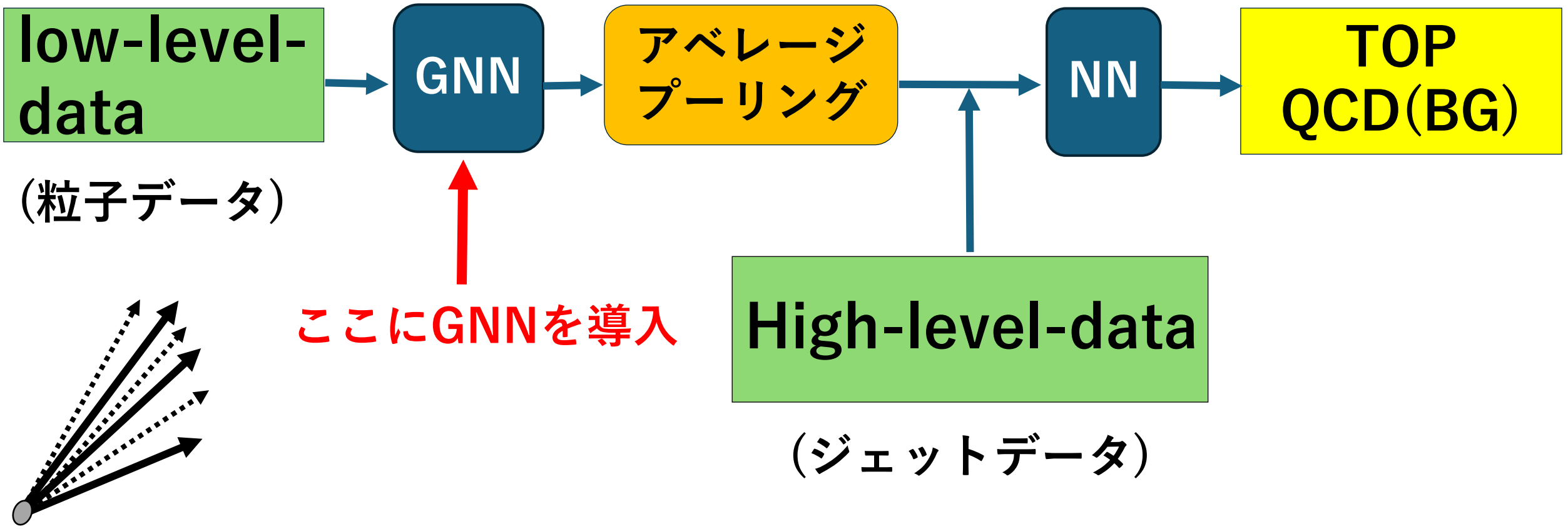
GNNでは、粒子間の近傍情報も識別に利用できる。

© Osaka Metropolitan University All Rights Reserved.

NNを用いたLHCデータ識別

大阪公立大学
Osaka Metropolitan University

本研究のNNのイメージ

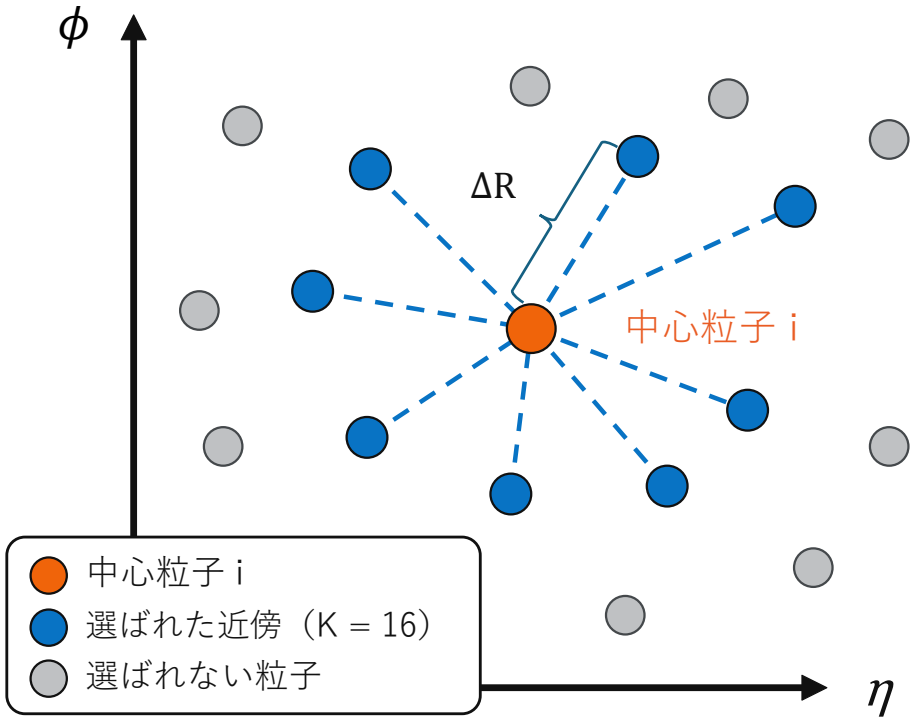


© Osaka Metropolitan University All Rights Reserved.

GNNを用いたLHCデータ識別

大阪公立大学
Osaka Metropolitan University

kNN



近いK個を選ぶ（K = 16）

ΔR に基づいて近傍を決定

$$\Delta R^2 = (\Delta\eta')^2 + (\Delta\phi')^2$$

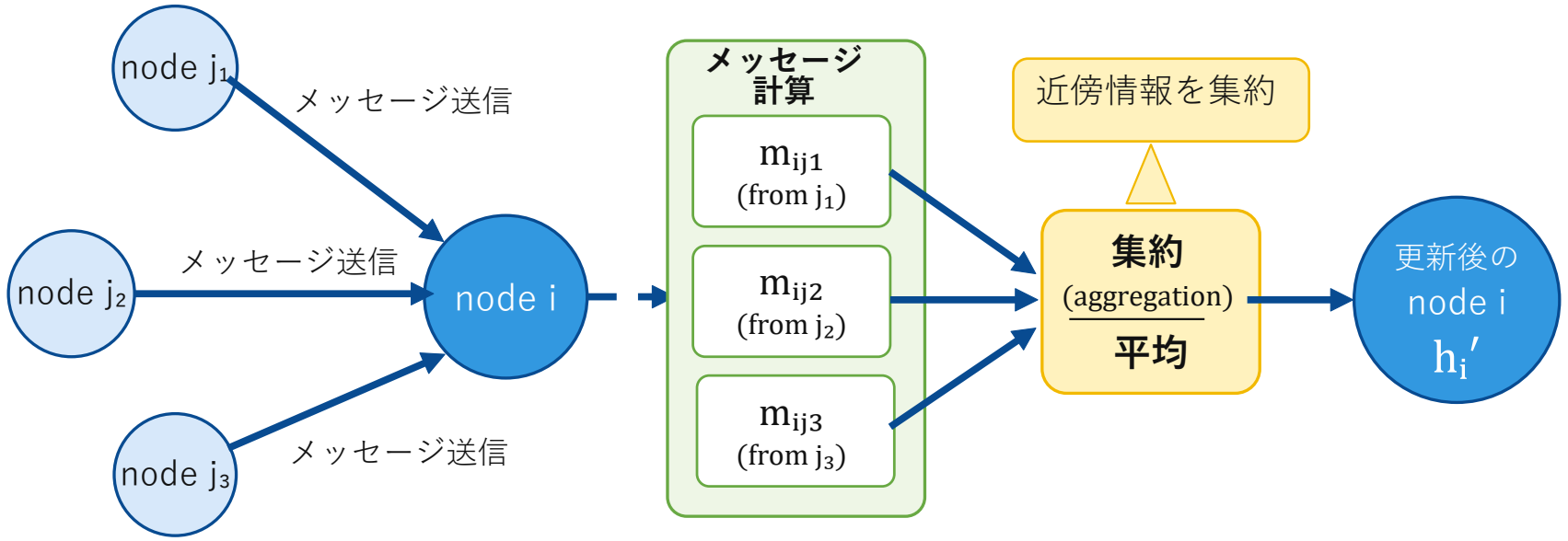
各粒子について、角度空間で近い粒子を近傍として選ぶ

© Osaka Metropolitan University All Rights Reserved.

GNNを用いたLHCデータ識別

大阪公立大学
Osaka Metropolitan University

Message Passing



m_{ij} → 平均 → h_i'

メッセージ計算　aggregation　更新後の表現

近傍粒子の情報を集約して、各nodeの表現を更新する

© Osaka Metropolitan University All Rights Reserved.

NNを用いたLHCデータ識別

大阪公立大学
Osaka Metropolitan University

GNNへの入力パラメータ

低特徴データ(Low-level-data)(粒子データ)

- 静止系　：4元運動量(E, **P**), P_T, |**P**|
- 実験室系：4元運動量(E, **P**), P_T, |**P**|
- 相対角　：$\Delta\eta$, $\Delta\Phi$



特徴量(High-level-data)(ジェットデータ)

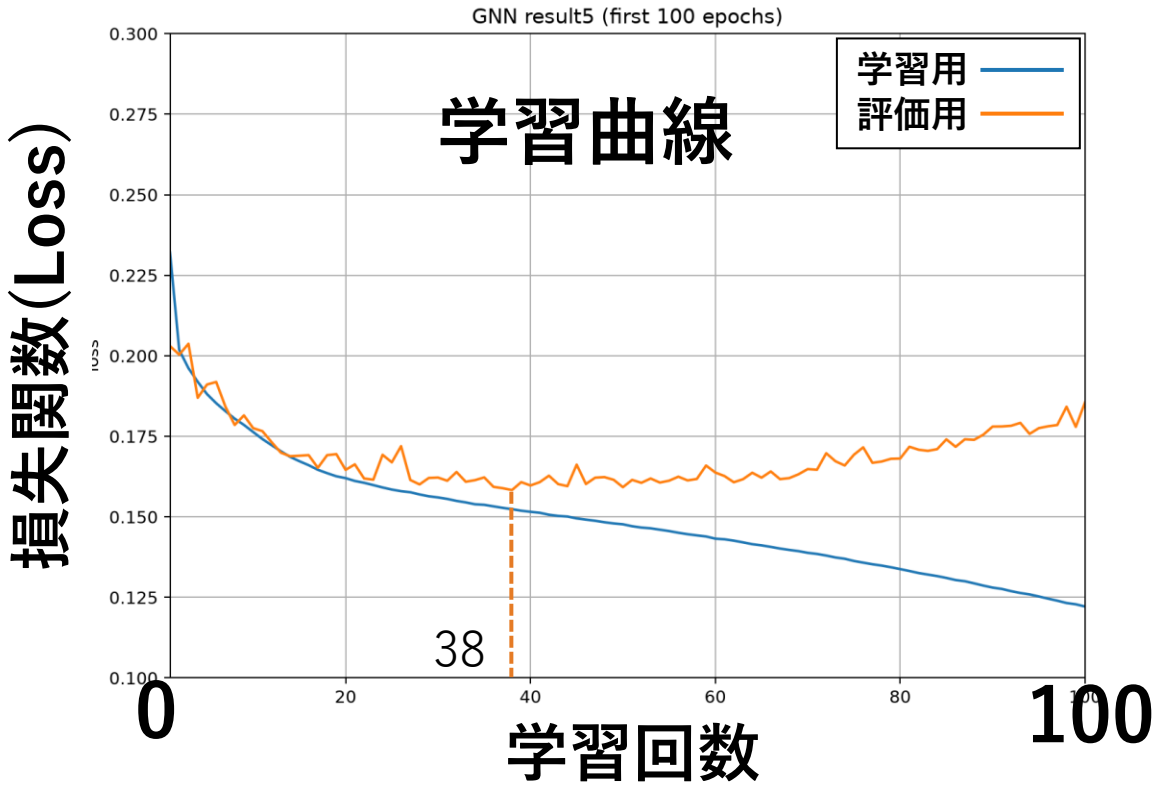
- 実験室系：E, P_z
- 静止質量M, 粒子数

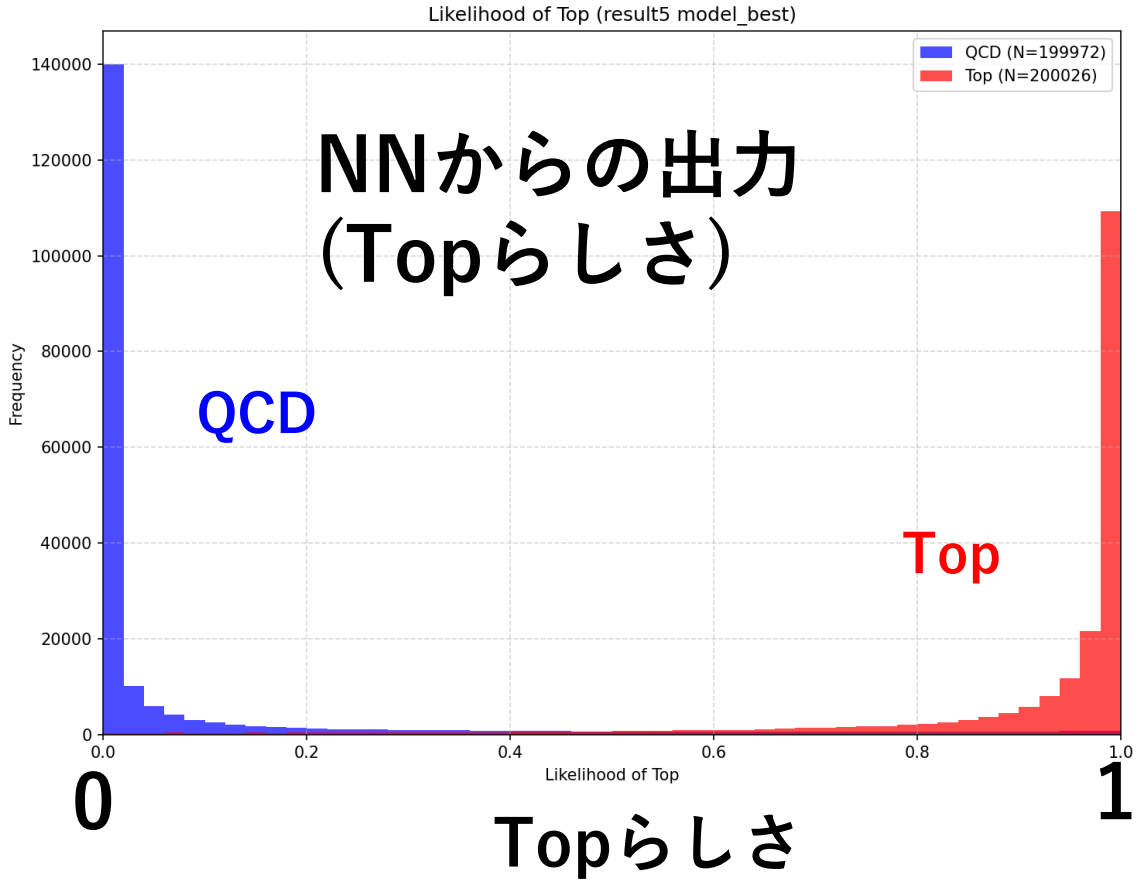
© Osaka Metropolitan University All Rights Reserved.

GNNを用いたLHCデータ識別

大阪公立大学
Osaka Metropolitan University

学習の結果



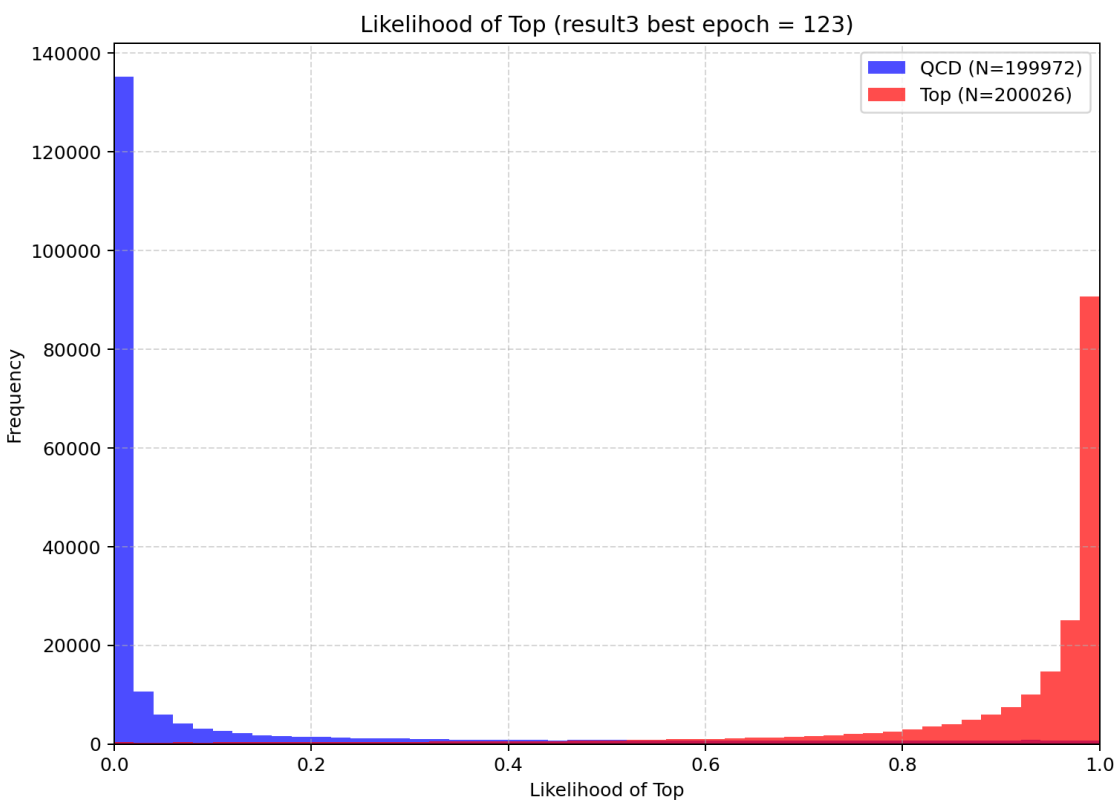


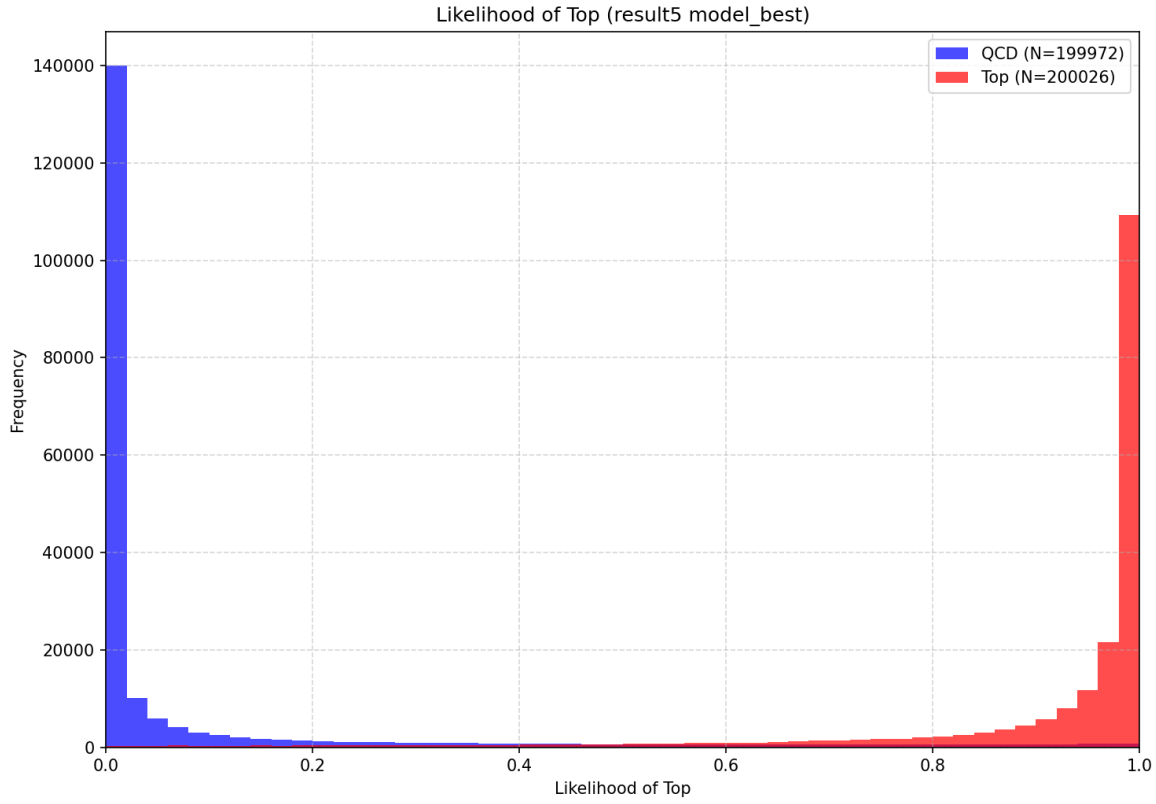
- 学習曲線安定
- このGNNでTopクォークとQCDを識別できている

© Osaka Metropolitan University All Rights Reserved.

NNとGNNの比較

大阪公立大学
Osaka Metropolitan University

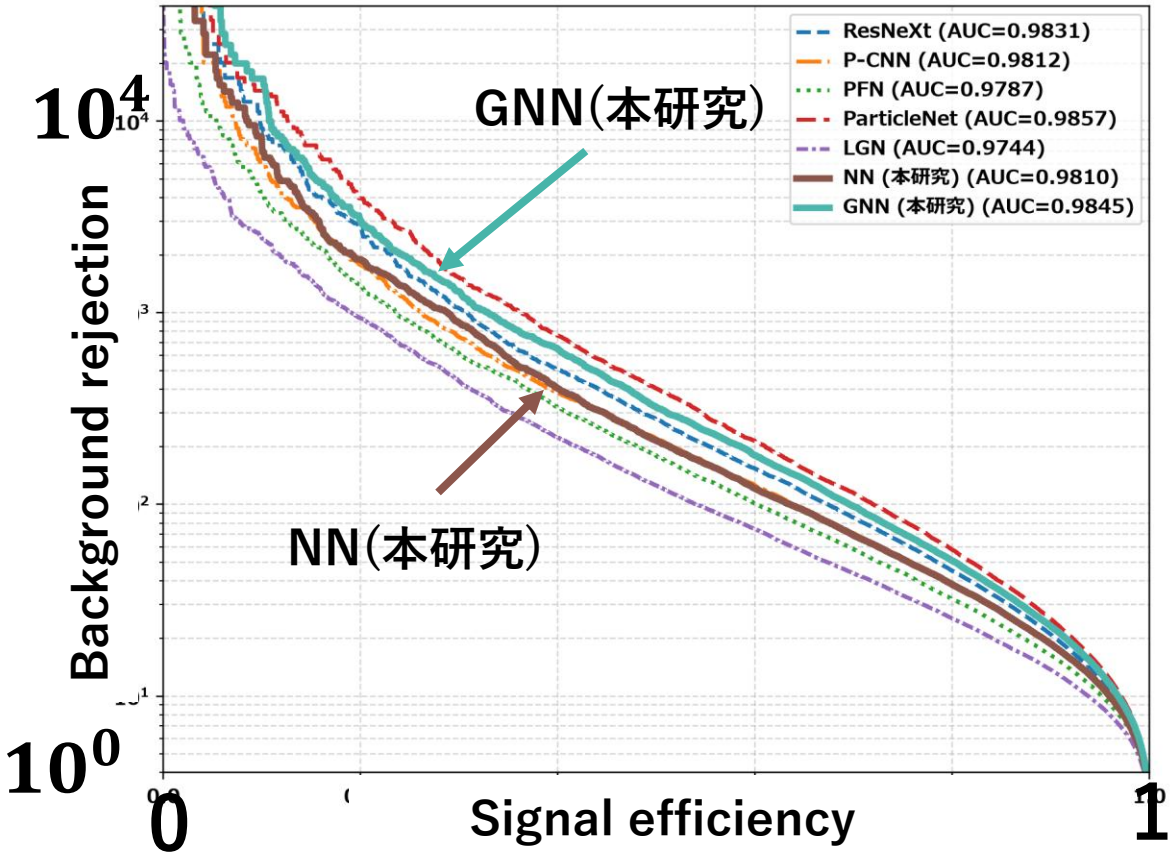




NN(MLP)		GNN
0.042	TopのTopらしさ1に対する乖離度	0.040
0.060	QCDのTopらしさ0に対する乖離度	0.052

sity All Rights Reserved.

先行研究との比較



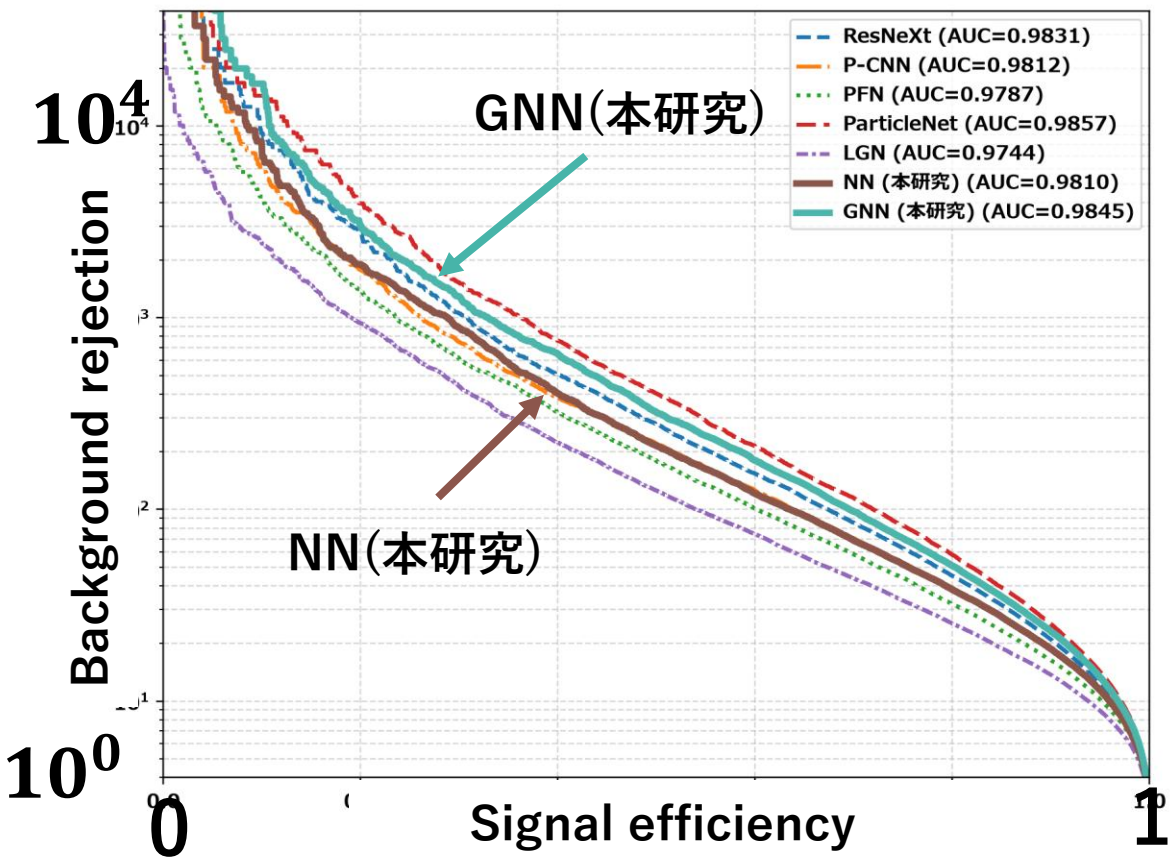
Model	$\varepsilon S = 0.3$	$\varepsilon S = 0.5$
LorentzNet	2195	498
ParticleNet	1530	402
GNN(本研究)	**1290**	**322**
ResNeXt	1052	277
NN(本研究)	**885**	**215**
P-CNN	745	220
PFN	614	179
LGN	438	126

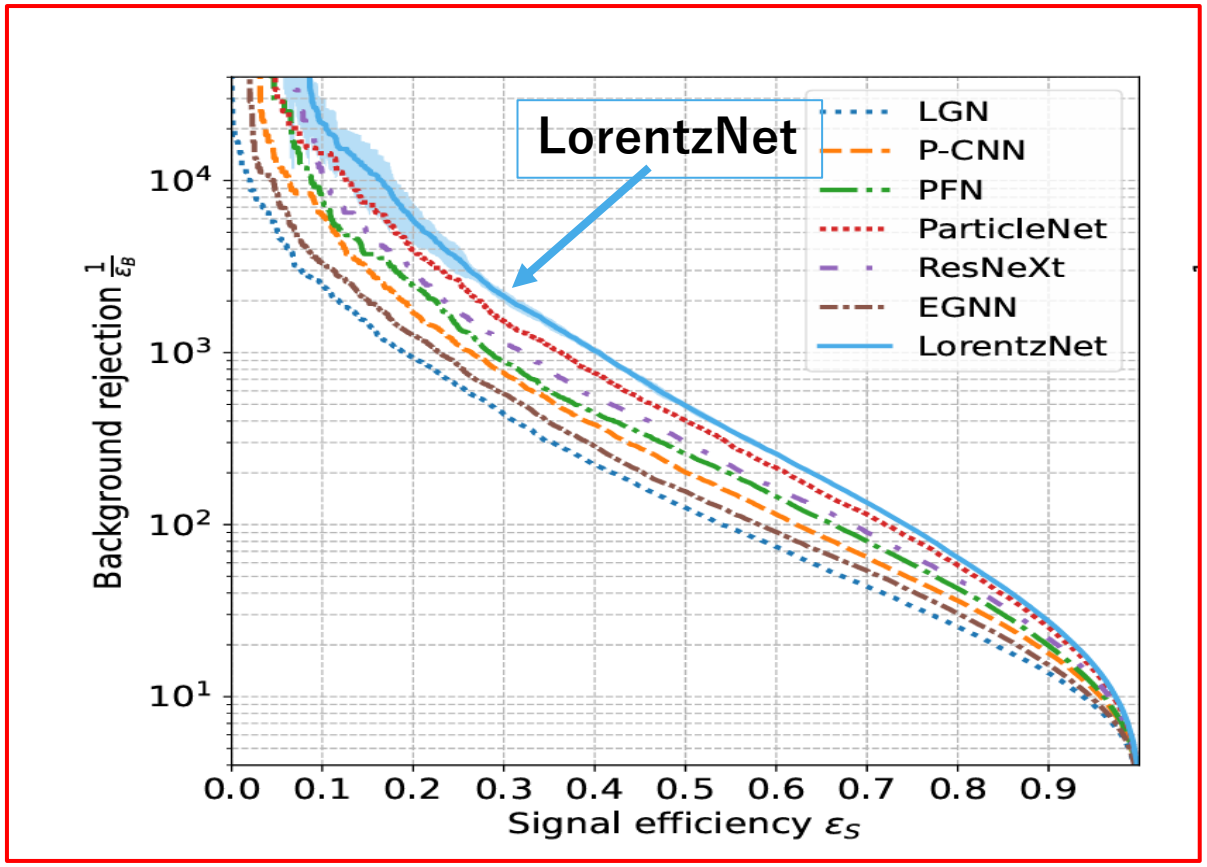
GNN

GNNの性能は、NN(MLPベース)を用いたモデルを上回る識別性能を示した。

© Osaka Metropolitan University All Rights Reserved.

先行研究との比較





本研究のGNNは、他のNNを用いたモデルを上回る識別性能を示した。

© Osaka Metropolitan University All Rights Reserved.

まとめ

大阪公立大学
Osaka Metropolitan University

本研究では機械学習を用いたトップクオークの識別手法の開発を行った。

- Low-level-data(生データ) によるNN・GNNを構築した。性能は以下のようになった。

Model	Signal efficiency=0.3	0.5
NN	Background rejection=885	322
GNN	1290	215

- 粒子単位のLow-level dataを用いた本研究は、特徴量を用いた先行研究との比較で、より高い識別性能を示した。
- GNNはNNを上回るBackground rejectionを示した。粒子間の関係を取り入れることが、性能向上に有効である可能性が示唆された。

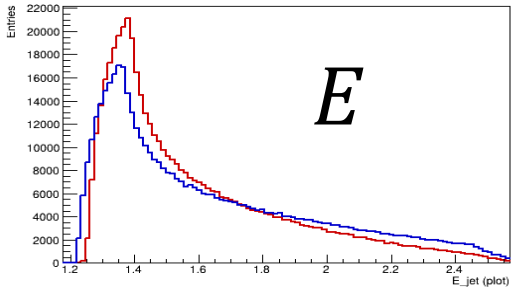
© Osaka Metropolitan University All Rights Reserved.

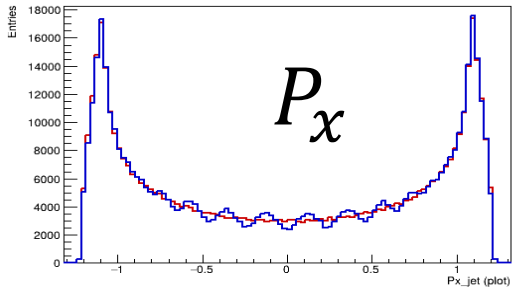
バックアップ

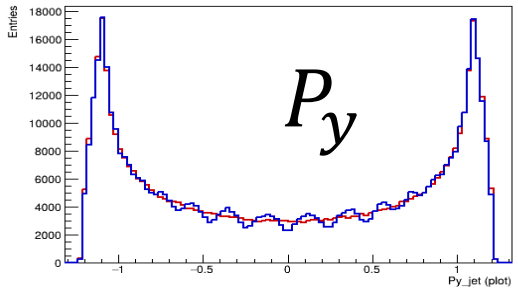
© Osaka Metropolitan University All Rights Reserved.

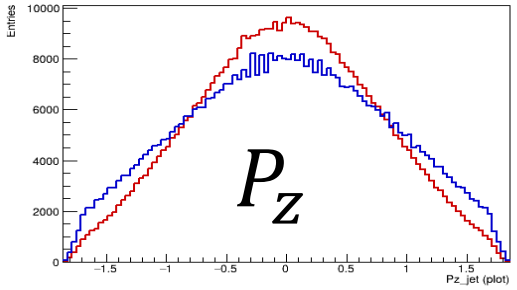
boost前のJetのヒストグラム

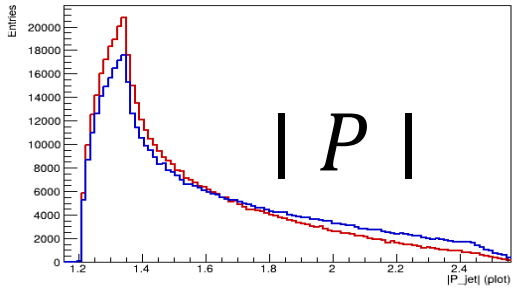
大阪公立大学
Osaka Metropolitan University

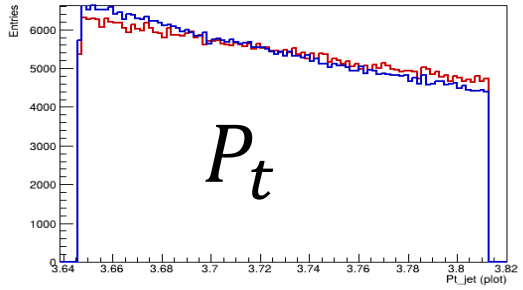


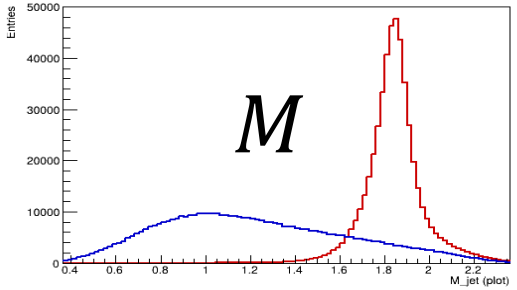


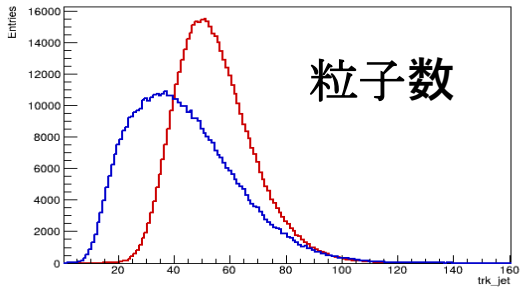












top
qcd

© Osaka Metropolitan University All Rights Reserved.

NN・GNNの最適化設定

大阪公立大学
Osaka Metropolitan University

項目	NN	GNN
粒子側隠れ次元	512	128
粒子側層数	2	2
イベント側隠れ次元	128	128
イベント側層数	2	3
Message Passing層数	—	2
近傍数 k	—	16
学習率	1.0×10^{-4}	1.0×10^{-4}
Batch size	512	128
Optimizer	Adam	Adam
活性化関数	ReLU	ReLU

© Osaka Metropolitan University All Rights Reserved.