

粒子同定の手法へのベイズ統計の適用

木村 彰徳[†] 柴田 章博^{††} 清水 直美^{†††} 佐々木 節^{††}大阪工業大学工学部[†]高エネルギー加速器研究機構計算科学センター^{††}島根大学総合理工学部^{†††}

概 要

近年の高エネルギー物理学実験は複雑かつ大規模になり、データを効率良く蓄積・保守を行うために、データベースを利用するようになってきている。その蓄積された巨大なデータの中から大量のバックグラウンド事象に埋もれている未発見の粒子の探索や物理過程の高精度測定には、知識発見の技術が不可欠である。エネルギー測定に用いる電磁カロリメーターで起こる物理過程をモデリングし、ベイズ統計による予測分布を用いて粒子同定を行った事例を報告する。モンテカルロシミュレーションで生成した擬似的な事象に対して粒子同定を行った結果からベイズ統計の適用の可能性について述べる。

1. はじめに

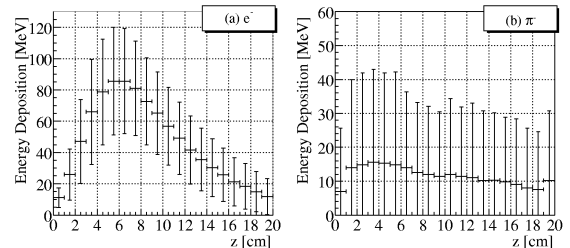
高エネルギー物理学実験では、粒子加速器により加速され高いエネルギーを持った粒子同士の正面衝突で生成した粒子を様々な検出器を用いて捕らえ、事象データを蓄積し素粒子反応の解析を行う。近年の高エネルギー物理学実験は、実験システムの複雑化かつ大規模化に伴い、巨大なデータの効率良い蓄積と保守および解析が必要となってきている。蓄積されたデータの中から解析を行う興味のある事象を何段階かに分けて抽出し、各段階においてデータモデルを適用し物理現象を理解するためのデータ解析を行う。このデータ解析の過程に、データベースの利用や情報科学における蓄積データからの知識発見の手法の成果を重要視し、応用研究がなされている。

本報告では、電磁カロリメーターによる粒子同定の問題を取り上げ、データマイニングの手法とその手法の有用性について述べる。これまでも同種の問題に対してニューラルネットワークを用いた様々な研究¹⁾がなされてきているが、オーバーラッピングなどのニューラルネットワークの抱える様々な問題と無縁ではなく課題も多い状態である。ここでは、粒子同定の問題を電磁シャワーの形状の抽出の問題としてモデリングし、ベイズ統計²⁾に基づく定式化を行う。検出器に入射した粒子の種類やエネルギーによって電磁シャワーの形状が異なるため、その特徴を抽出することで入射粒子の同定を行う。

2. パターン抽出と粒子同定

2.1 電磁カロリメーター

電磁カロリメーターは、エネルギー検出器の1つであり、電磁相互作用をする電子 (e^-)、陽電子 (e^+)、光子などの粒子

図1 e^- と π^- の入射方向の電磁シャワーのエネルギー分布。

のエネルギーを測定できる。検出器内で電磁シャワー³⁾を起させ、短い距離ですべてのエネルギーを多量の粒子に変換し、エネルギーを測定する。

他の運動量などを測定する検出器からの情報に依存せずに、この電磁カロリメーターのみで粒子同定が可能であれば、測定器を製作する費用や手間を減らすことが期待できる。また、既存の測定器においても、電磁カロリメーターによる粒子同定の情報を追加することで、実験解析の精度の向上が期待できる。

2.2 エネルギー分布のプロファイル

電磁カロリメーター内で発生する電磁シャワーは、入射粒子の種類やエネルギーによって、エネルギー分布の形状(シャワープロファイル)が異なることが知られている³⁾。そのため、異なるシャワープロファイルからパターン抽出を行うことで、その粒子を同定できると考える。ただし、本報告では電磁カロリメーターの手前に他の測定器を配置することを前提に、光子などの中性粒子と e^- とを判別できるとものとして、 e^- と間接的に電磁シャワーを起し、バックグラウンド事象となり得る荷電パイ中間子 (π^-) を同定することを試みる。

3. ベイズ統計に基づく判定法

3.1 シャワープロファイルのモデリング

エネルギー E_0 の入射粒子 ϕ の事象によって発生する電磁シャワーのプロファイルを粒子の種類とエネルギーの組ごとに分類するモデルを考える。粒子の入射方向に N_z 層からなる電磁カロリメーターの各層のエネルギー分布 $\varepsilon = (\varepsilon_1, \dots, \varepsilon_{N_z})^t$ は (図1)、その分布を与える母数 $\theta = (\theta_1, \dots, \theta_{N_z})^t$ を中心とした分散 Ω の正規分布に従うとすると、シャワープロファイルは $E_0, \phi, \theta, \Omega$ を与えた時の条件付確率分布で与えられる。

$$m(\varepsilon|\theta, \Omega, \phi, E_0) = \prod_{k=1}^{N_z} \frac{\exp(-(\varepsilon_k - \theta_k)^2 / (2\omega_k))}{\sqrt{2\pi\omega_k}} \quad (1)$$

ここで、それぞれのエネルギー分布は相関がない独立同時分布 $\Omega = \text{diag}(\omega_1^{-1}, \dots, \omega_{N_z}^{-1})^t$ と近似する。

このモデル分布の母数を $\theta = \theta(E_0, \phi)$, $\Omega = \Omega(E_0, \phi)$ とし、 (E_0, ϕ) それぞれについてシャワープロファイルの分布を決定する。いま n 個の事象 $\{\varepsilon^{(k)}; k = 1, \dots, n\}$ を用いた時の母数の分布 $P(\theta, \Omega|\{\varepsilon^{(k)}\})$ は、尤度 $L(\{\varepsilon^{(k)}\}) = \prod_{l=1}^n m(\varepsilon^{(l)}|\theta, \Omega)$

A Method of Particle Identification with Bayesian Statistics Method

[†] Akinori Kimura, Faculty of Engineering, Osaka Institute of Technology^{††} Akihiro Shibata and Takashi Sasaki, Computing Research Center, High Energy Accelerator Research Organization^{†††} Naomi Takashimizu, Interdisciplinary Faculty of Science and Engineering, Shimane University

を用いて

$$P(\theta, \Omega | \{\varepsilon^{(k)}\}) \propto \prod_{l=1}^n m(\varepsilon^{(l)} | \theta, \Omega) P(\theta, \Omega)$$

と書ける. このときの事前分布を局所一様であると仮定すると最大尤度を与える母数 $\theta = \bar{\varepsilon} = \frac{1}{n} \sum_l \varepsilon^{(l)}$ および $\omega_k = S_{kk} = \sum_l (\varepsilon^{(l)} - \bar{\varepsilon})_k (\varepsilon^{(l)} - \bar{\varepsilon})_k^t$ は, 正規分布の平均と分散を求めることに対応している.

3.2 オッズによる粒子判定

電磁シャワー ε^{id} を起こした粒子種別の推定は, 粒子, エネルギーを与えたときの母数 (θ, Ω) の分布から導かれる次の予測分布を用いて行う.

$$P(\varepsilon^{id} | \{\varepsilon^{(ex)}\}, \phi, E_0) = \left[\frac{\Gamma(\frac{n}{2})}{\sqrt{\frac{n}{2}} \Gamma(\frac{n-1}{2})} \right]^{N_z} \times \prod_{k=1}^{N_z} \sqrt{\frac{1}{2\pi S_{kk}/n} \frac{n}{n+1}} \left(\frac{S'_{kk}(\varepsilon^{id})}{S_{kk}} \right)^{-\frac{n}{2}} \quad (2)$$

ここで, S_{kk} は学習事象 $\varepsilon^{(ex)}$ によって計算された分散であり, $S'_{kk}(\varepsilon^{id})$ は, 次式で与えられる.

$$S'_{ij}(\varepsilon^{id}) = S_{ij} + \frac{n(\varepsilon^{id} - \bar{\varepsilon})_i (\varepsilon^{id} - \bar{\varepsilon})_j}{n+1}$$

予測分布は, 粒子とエネルギーの組 (ϕ, E_0) を指定したときのシャワープロファイルが実現される確率を与えるものである. 推定を行う粒子のシャワー分布 ε^{id} を用いればそれぞれの組におけるシャワー ε^{id} の実現確率を推定することができる. ここでは, 予測分布からシャワーの観測データが得られたとき, 入射粒子の状態 (ϕ, E_0) を推定するオッズを導入する. ベイズの定理より, $P(\phi, E_0 | \varepsilon^{id}, \{\varepsilon^{(ex)}\}) \propto P(\varepsilon^{id} | \{\varepsilon^{(ex)}\}, \phi, E_0) P(\phi, E_0)$ であるので, $P(\phi, E_0)$ が無情報事前分布であると仮定すると, (ϕ, E_0) に関するオッズは次式で与えられる.

$$O(\phi, E_0 | \varepsilon^{id}) = \frac{P(\varepsilon^{id} | \{\varepsilon^{(ex)}\}, \phi, E_0)}{\sum_{\phi' \in \Phi \setminus \phi} P(\varepsilon^{id} | \{\varepsilon^{(ex)}\}, \phi', E_0)}$$

4. 実験

4.1 電磁シャワーのシミュレーション

本報告で用いる電磁カロリメーターは, $1 \times 1 \times 1 \text{ cm}^3$ の立方体の鉛ガラスのブロックを x, y, z 軸に対して $20 \times 20 \times 20$ の 8,000 個を並べた $20 \times 20 \times 20 \text{ cm}^3$ の立方体となる. 鉛ガラスには, Pb (66.0%), O (19.9%), Si (12.7%), K (0.8%), Na (0.4%), As (0.2%) の割合で密度を 5.2 g/cm^3 で混合したものをを用いる. 同定を行う入射粒子はこの電磁カロリメーターの $x-y$ 面の中心に $+z$ 方向に入射するものとする.

上述した電磁カロリメーターに粒子が入射したときに起こる電磁シャワーのシミュレーションのために, GEANT4⁴⁾ を用いる.

粒子同定のパラメータには, 入射粒子 ϕ の (e^-, π^-), そのエネルギー E_0 の (0.5, 1.0, 2.0, 3.0, 4.0) GeV を組み合わせ, 学習のためにそれぞれ 1,000 事象を生成した. 判定する事象として, (a) それぞれの粒子を (1.0, 2.0, 3.0) GeV のエネルギーで入射した場合と, (b) $0.5 \leq E_0 \leq 4.0$ GeV の範囲で同様乱数で発生させたエネルギーの粒子の場合について, それぞれ 100,000 事象を生成した.

4.2 粒子同定の結果

まず簡単のためにモデルの学習に入射粒子 ϕ の (e^-, π^-) とエネルギー E_0 の (1.0, 2.0, 3.0) GeV の組み合わせを用いる. このとき, 入射粒子の e^- およびバックグラウンドとなる π^-

表 1 粒子を正しく判定できた正答率.

Energy set for learning model [GeV]	Ratio of correct inferences [%]	
	e^-	π^-
(a) (1.0, 2.0, 3.0)	96.8 ± 0.3	94.6 ± 0.2
(b) (1.0, 2.0, 3.0)	81.4 ± 0.4	95.2 ± 0.4
(0.5, 1.0, 2.0, 3.0, 4.0)	88.6 ± 0.4	93.2 ± 0.4

の $96.8 \pm 0.3\%$ および $94.6 \pm 0.2\%$ を正しく判定できている (表 1(a)). ただし, 誤差は統計誤差である.

次に, より実際の実験に近い条件にするために, E_0 が (1.0, 2.0, 3.0) GeV または (0.5, 1.0, 2.0, 3.0, 4.0) GeV の組み合わせの事象で学習したモデルに対して, 入射粒子のエネルギーを同様乱数で $0.5 \leq E_0 \leq 4.0$ GeV の範囲で振った事象で判定を行う. このときの判定結果を表 1(b) に示す. E_0 が (1.0, 2.0, 3.0) GeV の学習により, 入射粒子の e^- および π^- をそれぞれ $81.4 \pm 0.4\%$ および $95.2 \pm 0.4\%$ の割合で正しく判定できている. さらに, E_0 が (0.5, 1.0, 2.0, 3.0, 4.0) GeV と学習に用いるエネルギーの範囲を広げた場合には, 入射粒子の e^- および π^- の正答率は, それぞれ $88.6 \pm 0.4\%$ および $93.2 \pm 0.4\%$ となり, e^- は改善したが π^- が低下するという結果となった. これは, π^- が起こすシャワーのエネルギー分布のモデルが正規分布の近似からずれているため, 高エネルギーの π^- と低エネルギーの e^- の作る類似するシャワープロファイルが判別できないことなどが原因と考えられる. シャワープロファイルのモデリングをエネルギー分布の相関を含めて再検討することが必要である.

5. まとめと討論

ベイズ統計に基づくモデルを用いて, 理想化した電磁カロリメーター内で発生する電磁シャワーのみから, e^- とバックグラウンドの π^- の同定を行った. ベイズ統計のモデルを設定し, モデルパラメータを決定する方法は, 少ない学習データからでも良い粒子同定の成績を得ることができた.

今後は, モデリングと統計処理を系統的に扱うことができるベイズ統計に基づく本手法の利点を生かして, モデルの改良や実際の実験環境に近づけた解析を進めていくことが課題として挙げられる.

参考文献

- 1) M. Hoppner et al., Nucl. Instrum. Meth. **A389**: 154-156, 1997. and Halina Abramowicz et al., Nucl. Instrum. Meth. **A365**: 508-517, 1995.
- 2) J.M. Bernardo and A.F.M. Smith, Bayesian Theory, John Wiley and Sons, 2000.
- 3) D. H. Perkins, Introduction to High Energy Physics 4th edition, Cambridge University Press, 2000.
- 4) S. Giani et al., Geant4: An Object-Oriented toolkit for Simulation in HEP, CERN/LHCC 98-44, 1998. (URL: <http://wwwinfo.cern.ch/asd/geant4/geant4.html>)
- 5) 木村彰徳 et al., ベイズ統計モデルによる粒子推定の試み, KEK Preprint 2002-122, December 2002.