

SuperKEKB 加速器の高ルミノシティ化に向けた因果探索手法の探求

EXPLORATION OF CAUSAL DISCOVERY METHODS FOR ACHIEVING HIGH LUMINOSITY AT THE SuperKEKB ACCELERATOR

有馬諒太^{#, A)}, 松岡広大^{B, C, D)}, 三塚岳^{B, C)}, 小関忠^{A, B)}
Ryota Arima^{#, A)}, Kodai Matsuoka^{B, C, D)}, Gaku Mitsuka^{B, C)}, Tadashi Koseki^{A, B)}

A) The University of Tokyo

B) KEK

C) SOKENDAI University

D) Nagoya University

Abstract

At SuperKEKB, commissioning is ongoing toward the early achievement of the target luminosity. The factors affecting luminosity are numerous and complex, and their mutual causal relationships remain unclear, making it difficult to identify them exhaustively through manual analysis. In this study, we aim to quantitatively uncover these causal relationships by applying machine learning techniques to identify the key parameters contributing to luminosity. The central challenge of this study lies in evaluating the applicability of causal discovery methods to actual accelerator operations. In this presentation, we introduce the theoretical basis of the LiNGAM model used for causal discovery, present the results of its application to SuperKEKB operational data, and discuss current limitations and future prospects.

1. はじめに

SuperKEKB 加速器は、高エネルギー電子・陽電子衝突実験において世界最高のルミノシティを実現することを目的とした円形加速器である。2025 年 7 月現在、SuperKEKB 加速器は世界最高記録となるルミノシティ $5.1 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ を達成しており、引き続き目標ルミノシティの早期達成に向けた調整が進められている。その性能向上のアプローチとして、ルミノシティに影響を与える各種パラメータの特定、各パラメータの影響度を可視化する方法がある。

しかしながら、加速器の運転状態には多数の変数が関与しており、各変数間の相互作用や因果関係は極めて複雑である。従来は、専門家の知見や定性的な分析に基づく経験則により調整が行われてきた部分があるが、この方法ではすべての要因を網羅的に把握することは困難であり、効率的な運転最適化の妨げとなっている。

そこで、ルミノシティに寄与する様々な要因と因果関係の定量的理解を目的として、近年発展する機械学習と因果探索の手法を用いた解析を試みた。

2. 機械学習の導入

因果探索モデルの選定、動作確認を行い、徐々に変数を増やしながら結果の妥当性を検証していく。機械学習モデルとして、比較的シンプルな線形モデルであり、既存の Python ライブラリにより安定した実装が可能であるため、LiNGAM[1]を選定した。

2.1 LiNGAM

本研究で使用した LiNGAM (Linear Non-Gaussian Acyclic Model) は、変数間の因果順序および因果関係を推定する手法である。

[#] ry0aryma@g.ecc.u-tokyo.ac.jp

まず各変数 x_i が他の変数 x_j による回帰残差 e_i とどれだけ独立であるかをカーネル独立性検定により評価する。最も独立性が高い変数は他の変数に影響を与えておらず、因果構造上で最も“上流”にあると推定される。その変数の効果を他の変数から回帰によって除去する。この操作を繰り返すことで、すべての変数に因果順序を割り当てる。

次に各変数を各々の上流の変数で回帰することで、因果構造と変数間の影響度合いを推定している。

2.2 LiNGAM の仮定

LiNGAM が用いる仮定は以下のとおりである。

- 因果関係は「非巡回的」
- データの回帰残差が非正規分布かつ互いに独立
- モデルは「線形」
- 隠れた潜在変数がない

SuperKEKB の実データには、因果構造に影響を与えるような隠れた潜在変数の有無や数が事前には不明であるという課題がある。本研究では、このような状況でも LiNGAM が因果構造を正しく推定できるかを、実験的に評価することを目的とした。

3. 使用パラメータ、データの前処理

3.1 使用パラメータ

解析には、SuperKEKB 加速器の実運転データのうち、パラメータとしたのは、LER および HER のビーム電流、衝突点近傍の LER の Y 方向エミッタンス、ルミノシティである。これらは因果関係が自明であるとされるパラメータである。

3.2 データの処理

- 欠損値補完

実運転データでは測定値の欠損が発生することがあり、そのまま解析に使用すると因果推定の精度に悪影響を

及ぶ可能性がある。そのため、局所的な連続性を維持できる手法として、直前の有効なデータを用いることで欠損値を埋めた。

・時系列の整列処理

各パラメータは個別の計測システムから収集されているため、時間軸の粒度や記録タイミングが一致していない。これにより、同一時刻での値を比較することが困難になる。因果探索では時系列の整合性が重要であるため、すべての変数を共通の時間軸にリサンプリングして整列させた。

・LER または HER の電流が 300 mA 未満のデータ行の削除

ビーム電流が著しく低い状態では、加速器の安定運転が達成されていない場合が多く、パラメータの挙動も通常運転時とは大きく異なる可能性がある。したがって、異常系として解析結果を乱す要因となるこれらのデータは除外した。

4. 検証と検定

4.1 研究手順

解析手順としては以下を実施した：

1. どれくらいのデータ統計量で正しい結果が得られるか検証
2. 同じデータ統計量だが、異なる時間帯データでの結果比較
3. 得られる結果の妥当性を2つの検定で評価。

以下の出力結果ではブートストラップ法により統計的有意性が高いと判断されたパスのみを表示している。

4.2 検証

4.2.1 データ統計量ごとの結果

どれくらいのデータ統計量で正しい因果関係が得られるかを検証するため、1 分間、1 時間、1 日間でデータ統計量を変えながら検証した。結果が Fig. 1 である。赤×は出力結果で因果関係がないと判断されたパスを、赤矢印は因果関係の向きが誤っているパスを意味する。

出力結果は毎回異なり、データ統計量が増えても正しい因果関係が出るとは限らなかった。

4.2.2 同じデータ統計量だが、異なる時間帯データでの結果比較

次に同じデータ統計量で、出力ごとに同じ結果が得られるのかを検証した。1 時間での複数回の出力結果を比較した。結果が Fig. 2 である。

データ統計量が同程度でも用いるデータごとに出力結果は毎回異なった。

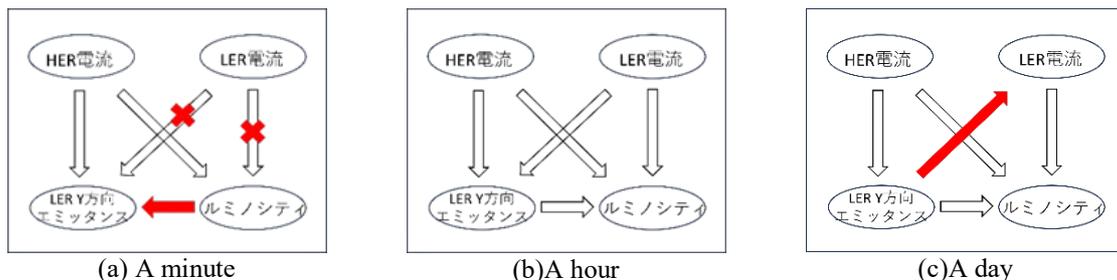


Figure 1: Variation in LiNGAM Outputs with Respect to Data Statistics.

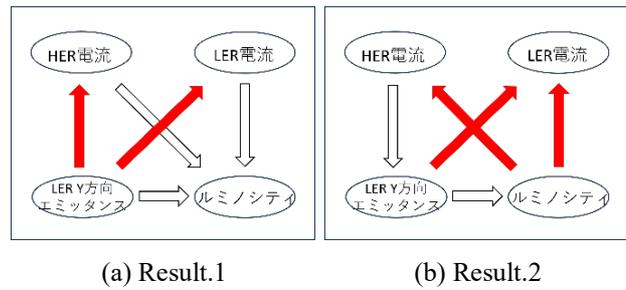


Figure 2: Examples Comparison of LiNGAM Results for the Same Data Statistics across Different Time Periods.

4.3 検定

LiNGAM の適用結果が安定しなかった原因を調査するため、以下の 2 種類の前提条件に対する検定を行った。

4.3.1 独立性検定

LiNGAM は各変数の回帰残差同士の独立性を仮定している。独立性の評価には HSIC、有意水準 $p=0.05$ を用いた。検定結果が Table 1 である。多くのパラメータ間で回帰残差同士の従属関係が確認され、LiNGAM の仮定を満たさないことが分かった。

Table 1: Independence Between Regression Residuals of Each Parameter

変数間	P 値	独立の有無
HER 電流 - LER 電流	1.27×10^{-12}	なし
HER 電流 - LER Y 方向エミッタンス	7.03×10^{-6}	なし
HER 電流 - ルミノシティ	5.00×10^{-14}	なし
LER 電流 - LER Y 方向エミッタンス	5.94×10^{-33}	なし
LER 電流 - ルミノシティ	1.49×10^{-47}	なし
LER Y 方向エミッタンス - ルミノシティ	6.28×10^{-3}	なし

4.3.2 正規性検定

LiNGAM は回帰残差が非正規分布であることを前提としている。これに従い、各変数における回帰残差分布について視覚的な判断により、正規性を検定した。結果が Fig. 3 である。データセットごとに回帰残差分布が変わる。また、そのデータセットを用いても回帰残差が正規分

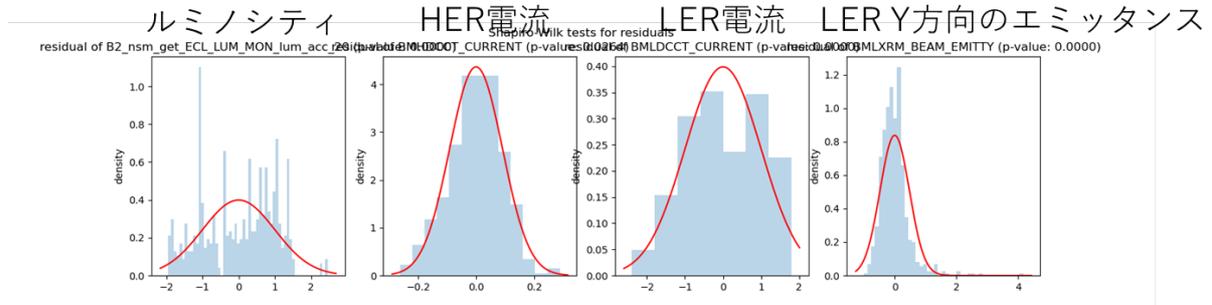


Figure 3: Distribution of Regression Residuals for Each Parameter.

布になるパラメータがあった。

5. まとめ

本研究では、SuperKEKB 加速器におけるパラメータ間の因果関係を定量的に明らかにすることを目的として、因果探索モデル LiNGAM を適用し、その有効性を検証した。特に SuperKEKB の実データには、因果構造に影響を与えるような隠れた潜在変数の有無や数が事前には不明であるという課題が存在しており、そのような前提違反の可能性を含んだ環境下でも LiNGAM が因果構造を適切に推定できるかどうかを、実験的に評価することを目的とした。

LiNGAM は、回帰残差の非ガウス性と独立性を仮定することで因果方向を識別可能とする理論的に有用な手法であるが、実データへの適用においては、時間帯や時間スケールの違いによって出力結果が大きく変動した。原因を探るため、LiNGAM の主要な前提条件である「回帰残差の非正規性」と「変数と回帰残差の独立性」の

妥当性を検証すべく、以下の 2 種類の統計的検定を実施した。

正規性検定では、どのデータセットを用いても、回帰残差が正規分布であるパラメータがあった。

独立性検定の結果、回帰残差間に従属関係が多数確認された。

これらの検定から、LiNGAM の回帰残差についての仮定が満たされておらず、ルミノシティに影響する 3 つのパラメータだけでは隠れた潜在変数の影響を受けて正しい結果が得られないことが分かった。

今後は、ルミノシティと強い相関を持つパラメータを含めることを検討している。また、回帰残差の仮定を含まないモデルへの変更も考え、PCMCi+を次の候補として有力視している。

参考文献

- [1] S.Shimizu *et al.*, “DirectLiNGAM: A direct method for learning a linear non-Gaussian structural equation model”, <https://arxiv.org/pdf/1101.2489>