

インバリアント分析技術を活用した 加速器システム等の異常に対する予兆検知・診断

PREDICTIVE DETECTION & DIAGNOSIS OF ACCELERATOR SYSTEM USING SYSTEM INVARIANT ANALYSIS TECHNOLOGY (SIAT)

相馬知也^{#, A)}, 石井恒次^{B)}, 吉岡正和^{C)}, 高城真弓^{A)}

Tomoya Soma^{#, A)}, Koji Ishii^{B)}, Masakazu Yoshioka^{B)}, Mayumi Takagi^{A)}

^{A)} NEC Corporation, ^{B)} J-PARC/High Energy Accelerator Research Organization, ^{C)} Tohoku University

Abstract

Detecting anomalies in the accelerator system is very important in carrying out accelerator experiments. In recent years, techniques for AI (artificial intelligence) and big data analysis have rapidly developed. As its application, an anomaly detection / sign diagnosis technique for infrastructure such as plant is being studied and developed. By applying this technology to the accelerator, we examine the possibility of detecting abnormality detection and failure signs by accurately and precisely grasping the state of the accelerator system. In this research, using System Invariant Analysis Technology (SIAT) developed by NEC, System Invariant Analysis Technology of infrastructure information accumulated in J - PARC is aimed at real - time anomaly detection and diagnosis of signs of failure Perform analysis. Moreover, if it becomes possible to capture the unknown relationship by adding the beam operation information, it can be expected to be useful not only for stabilizing the accelerator operation but also for improving the performance of the beam. In the future, it is highly likely that it is an indispensable technology for large-scale projects such as ILC accelerator, and its possibility is also considered.

1. はじめに

近年のネットワークやコンピュータなど ICT (Information and Communication Technology)の進歩は目覚ましいものがある。たとえば、ハードディスクや SSD をはじめとするストレージの高速化と低価格化によりデータの大量蓄積が容易になってきた。データベースでは、従来の RDB(リレーショナルデータベース)では扱えなかった、加速器などの大型設備から取得された時系列データの蓄積も、TSS-DB(時系列データベース)の出現により大量のデータを容易に蓄積し扱うことができるようになってきている。[1]

一方でデータ分析技術の進歩も目覚ましく、AI(人工知能)やビッグデータ解析と呼ばれる新しい技術の出現により、従来までの統計的手法では扱うことのできなかつた大量のデータを容易に扱えるようになってきた。加速器においても、設備の運転データはその実験の特性から短時間のサンプリングレートで取得されている。このためデータ量は膨大になり、蓄積されたデータの分析を統計的な手法で行うことは非常に時間のかかるものであった。

加速器設備の運用においては、スムーズな実験の遂行および安定した実験データの取得の観点から、設備自体の健全性をリアルタイムで認識しておく必要がある。また、故障の状態を早期に検知し、計画外停止を未然に防ぐことは、研究スケジュールを遂行するうえでも重要であり、研究経費の削減にもつながる。

本稿では、NECが開発したデータ分析技術であるインバリアント分析技術 (System Invariant Analysis

Technology : SIAT)を活用した加速器等の故障予兆検知と効率的な設備保全への適用について述べる。適用検討に当たって、J-PARC メインリングの運転データおよびユーティリティの運転データサンプルをもとに簡易的な分析を行った。これらの分析結果をもとに、適用の可能性について述べる。

2. インバリアント分析の概要

2.1 手法と特徴

インバリアント分析技術は、システムの性能情報やプラントのセンサ情報として得られる数値の時系列から、それらの関係性を網羅的にかつ高速に抽出する機械学習技術である。学習したセンサ情報間の関係性を対象システムの稼働モデルとし、その関係性が変化した時刻と場所をリアルタイムに監視することで、異常兆候を早期に発見することができる。

大規模設備の運転監視に適用した場合、従来の閾値監視やベースライン監視では見つけづらかつたサイレント障害 (監視制御システムで確認できない、閾値以下で発生している異常など)を検知できるため、障害影響が顕在化する前の早期対処により、計画外停止による損失を防止し、運用コストを低減することが可能となる。[2]

t-soma@vx.jp.nec.com



Figure 1: Invariant analysis technology overview.

インバリエント分析技術は基本的に数値の時系列であれば分析の対象とできるため、例えば、ポンプやモーターなどの振動計測データや、設備の性能データなどの情報をまとめて分析(クロスドメイン分析)し、管理体制に応じた障害の一次切り分けを迅速に行うことができる。更に、設備建屋内の温度や、外気温、冷却水温などのさまざまな状態や、設備運転データとの関係性を発見することで、より広範囲な知見の見える化やシミュレーションへの応用も可能となる。

2.2 分析技術の詳細

機械学習には、入力データとその入力データから得られるべき結果を学習データとして与え予測モデルを構築する「教師あり学習」と、入力データのみから予測モデルを構築する「教師なし学習」がある。

一般的に既知の障害事象を精度よく検出するためには、教師あり学習が適すと考えられる。しかし、設備の障害検知に利用しようと考えた場合、大きな課題に当たってしまう。それは学習データの量である。教師あり学習の多くは、予測精度と学習データ量が比例関係にある。質の低いデータを学習した場合、予測結果の質も低くなり誤検知の原因となる。すなわち、大量の質の高いデータが必要となるのである。通常の設定においては、同じ障害、特に大規模障害に至っては二度と起こさないように対策が取られる。このため、質の高いデータがあったとしても、その数は非常に少ない。設備の異常予兆検知にディープラーニングが限定的にしか使われていない原因はここにある。このため、ディープラーニングやニューラルネットワークなどの技術は、ある一点のセンサーデータのトレンド予測など、既存の監視の延長線上で限定的に使われているのみであり、設備全体を対象とすることが困難である。

対照的に、インバリエント分析技術は「教師なし」学習である。

インバリエント分析技術は、時系列データをもとに、計測項目(センサ等)間の不変的な関係性を網羅的に自動で抽出しその関係を関数として表現する。その関数を使用し関係性の崩れを「いつもと違う状態」として検知することで、プラント設備の異常やプロセスの状態変化を発見する技術である。

インバリエント分析技術の特徴の1点目は「自動化による網羅性」である。一般的に統計解析においては、分析者が仮説を立て検証を行う。このため、分析者の意図や「こうなってほしい」という恣意的な部分が必ずしも排除で

きないと考えられる。また、プラントデータの解析においては、その知見の多さから「設備設計上こんなことはない」や「このデータはこう動くはず」といった、人の意思が働くことがある。このような選択的アプローチを行うと、機械学習を適用したとしても今まで以上の発見はできない。

インバリエント分析技術においては、データを選択することなくすべてのデータを利用してモデル化を行う。つまり、 N 個のセンサがあれば組み合わせ数 C は最大で式(1)で示される。

$$C = \frac{N(N-1)}{2} \quad (1)$$

100個のセンサであれば、4,950個の組み合わせである。これらのすべての組み合わせについて不変的な関係性の有無を調べる。

自動制御はセンサ間の関係性をあらかじめ決めることで行われているが、設備には運転中に特性が変わっていくことによる、熟練者しかわからない関係性も存在する。これらの関係性は経験に基づくものであり、継承が難しい技術となっている。

このように、自動抽出を行うことで、従来の「教師あり学習」とは異なり、自動制御で行われている関係性はもとより、熟練者しか知らないような関係性、さらには「未知の関係性」を見つけることが可能となっている。

2点目の特徴は「関係性の維持」である。関係性の有無を調べる際、インバリエント分析技術では、学習データとして与えられた時系列データの期間において、関係性が維持されているか(不変的な関係性か)を調べる。これがインバリエント(英: Invariant) = 不変性の意味である。

また、設備においては、応答速度の差により同一時刻で比較を行っても関係が発見されない場合が多い。このため、単に瞬間的な関係性のあり/なしではなく、「学習期間で関係性が維持されているか」を調べ、かつ2点間の関係性の時間の遅れを考慮した関係式を作成することで、より高精度なモデルの作成を可能としている。

3点目の特徴は異常検知の仕組みにある。従来までの監視では、閾値(しきい値)超過を基本としている。また、設備設計や運用経験から得られた既知の関係性を監視システムに設定することで異常を検知していた。閾値監視では運転状態による閾値の細かな設定や、設備改修後の再設定などで大きな手間がかかるなど課題もあった。また、設定値は危険域に設定されていることが多く、異常事象が進まないと発見できないという課題もあった。

インバリエント分析技術では、センサーデータの動きから関係性を自動で発見しモデル化を行うため、設備の状態変更時にも容易にモデルの再作成が可能である。また異常検知においてはセンサ間の関係性の崩れを検知することで異常を発見する。このため、閾値を超過する前に異常を検知することが可能である。

設備においてはある一か所で異常が発生すると、そのセンサを中心に関係性が崩れる。インバリエント分析技術は「点」ではなく関係性という「線」で監視を行う。したがってモデル図の上に壊れたインバリエントを図示することで、視覚的に主原因センサの絞り込みが可能となる。[3]

3. J-PARC MainRing への適用検討

J-PARC 加速器はすでに 10 年程度の運転実績を有しており、システムのセンサーデータの蓄積は十分にある。インフラ情報をインバリエント分析技術により解析することで、リアルタイム異常検知や故障の予兆診断に活用できるものと考えられる。

今回の検証にあたっては、EPICS (Experimental Physics and Industrial Control System)によって収集/蓄積されているデータのうち 2017 年 3 月の 1 日間、トラブル発生のないデータを使用しモデル作成の簡易分析を行った。

Table 1: Data Specifications

データ期間	2017/3/1 0:00 ~ 2017/3/2 0:00
サンプリング間隔	最短で 1 秒
センサ数 (カラム数)	9,804 (コンピュータ関連を除く)
モデル化対象	BeamTransport/MainRing/Utility

EPICS で取得したデータは最短で 1 秒間隔でサンプリングされているものの、抜けや間隔が一定ではないデータが多く存在した。このため分析にあたってはデータの前処理を行って入力データとした。サンプリング間隔のずれは 1 秒未満の範囲内で起きていたため、1 秒間隔に成型した、また、抜けていたデータは、収集システムの仕様上、変化のないものについては記録しない設定のため、前値補間を行った。これらの成型を行ったデータを使用してモデル作成を行った。作成したモデルの一部をイメージとして Figure 2 に示す。

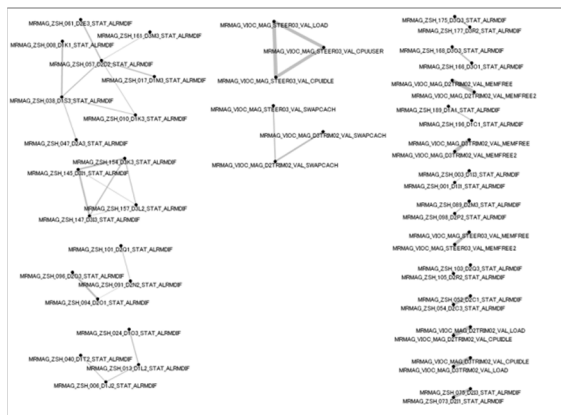


Figure 2: Invariant model Image (MR).

Figure 2 は MR の一部を図示しているものであるが、人が持っている設備知見との比較においても、十分に関係性を捉えられている。また、比較的強い関係を持つインバリエントが多いことも特徴的であった。Table 2 に作成されたモデルの概要を示す。

Table 2: Model Specifications

ThB	ThE	# of Invariant
0.3	0.5	740,542
0.5	0.7	488,248
0.7	0.9	43,660

ここで ThB は両方向の関係性(Threshold Both)、ThE は片方向の関係性(Threshold Either)を示す。インバリエントは2点間の関係性のため 2 点のうちどちらを関数の x とし、どちらを y とするか方向性がある。インバリエント分析におけるモデル作成は、自動で網羅的に関係性を抽出するため非常に弱い関係性もモデルとして作成される。関係性の強さ(Fitness Score:Fs)は、最大を 1.0 として算出される。弱い関係のインバリエントは異常検知時には誤検知の原因となることがあるため、モデル作成においては Fs を利用してフィルタリングを行う。

今回作成したモデルの概要を見ると、ThB/ThE=0.5/0.7 以上のインバリエントでも約 49 万インバリエント、ThB/ThE=0.3/0.5 以上になると約 74 万インバリエントと非常に多くのインバリエントが抽出され、MainRing においては各センサ間が強い関係の下で動作していることが確認できた。これは、我々が過去に経験した設備と比べても、強い関係性が多い。

今回のサンプルは安定している期間の 1 日間データを使用している。このデータで強い相関関係が得られたことは、リアルタイム異常検知や故障の予兆診断に大きな役割を果たすものと期待できる。今後、不安定期のデータや故障時のデータを解析し、異常検知の可能性・確実性等を検証したいと考えている。

また今回のサンプルはセンサ数の一部しか使用していない。それでも高い数値のインバリエントを示しているため、現状では埋もれてしまっており、未知の関連性を捉えることはできていない。解析データを精査するに加え、少し前処理等を行うことで未知の関係性が抽出できてこないかを模索したいと考えている。

4. まとめと今後

今回の簡易分析では、MainRing 及び周辺ユーティリティのデータを中心にモデル作成までを行い初期の有効性確認を行った。ここで有効性に期待が持てたことから、今後は MainRing やユーティリティなどの個別部分の検討に加え、それら全体を含めた関係性のモデルを作成し、相互の状態が及ぼす影響についても分析/検討を行っていく。

また設備の監視データにビーム運転情報を加味し、未知の関係性を捉えることが可能になれば、加速器運転の安定化のみならず、ビームの性能向上にも役立つものと期待できる。

将来に渡っては、ILC 加速器のような大規模プロジェクトでは必須の技術となっている可能性が高く、J-PARC 加速器で十分に経験を積むことで、次期加速器 (ILC 等) 計画では初期段階からの導入を目指す。本技術の適用

PASJ2017 WEOL01

により、運転開始直後のシステム調整・ビーム調整においても大いに役立つ技術と確信している。

参考文献

- [1] T.Soma *et al.*, “Big Data Analytics in the Cloud - System Invariant Analysis Technology Pierces the Anomaly –”, NEC Technical Journal Vol.9 No.2 2014;
<http://www.nec.com/en/global/techrep/journal/g14/n02/pdf/140218.pdf>
<http://jpn.nec.com/techrep/journal/g14/n02/pdf/140219.pdf>
- [2] 相馬知也, “運転中のリアルタイム検査を実現するビッグデータ解析技術の活用”, 機関誌「非破壊検査」2016年10月度.
- [3] 高木敏行、相馬知也 他, “非破壊検査領域におけるインバリアント分析技術の適用研究”, 2017/4.