

産業機器のアラート出現傾向に基づく事故予測手法の検討 Study on Failure Prediction Method Based on Alert Tendency for Industrial Equipment

太田 菜月[†] 桑島 昂平[†] 柴田 秀哉[†]
Natsuki Ota Kohei Kuwajima Hideya Shibata

1 はじめに

産業機器を対象とした保全方法の1つとして、予知保全の重要性が指摘されている。予知保全とは、何らかの方法で発見した事故予兆に基づき、事故発生前に機器の不具合を取り除く保全方法を指す。

電力やFA等の分野で扱う多くの産業機器は、耐用年数が長く、さらに事故発生以前に余裕を持って交換される傾向にある。したがって、事故発生時の異常データが少ないという特性がある。このような分野の予知保全では、一般に、異常データの学習に頼らない外れ値検出や変化点検出等の手法が活用されてきた。外れ値検出等を応用した異常検知手法では、期待する正常データとの差異を検出可能である。一方で、事故種別や事故発生場所などの詳細までは予測できなかった。

他方で、機器にセンサを取り付けて、センサデータを用いた外れ値検知等の手法が実システムに適用されるようになると、検知結果としてのアラートデータと、実際の事故の情報を紐づけて取得することが可能となる。

本研究では、これらの情報を予知保全に活用できないかという点に注目し、事故の詳細情報まで予測可能な事故予測手法の確立を目指す。送配電の分野を例にとると、「アラート1、アラート2、アラート3がこの順序で一定時間内に発生した場合、その後、系統A電柱Bという位置で種別Fの事故が発生する可能性が高い」等の予測実現を目標とする(図1)。

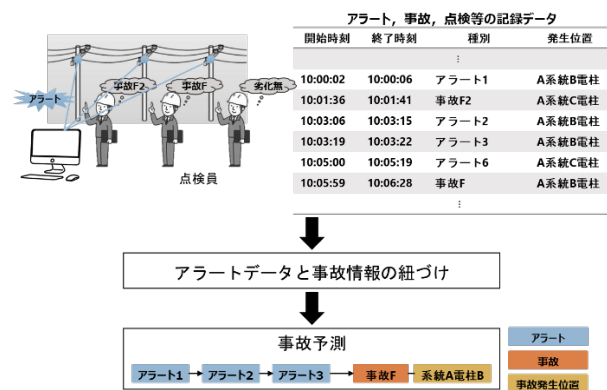


図1. アラートの出現傾向に基づく事故予測の概念図

本稿では、事故の記録と、当該事故が発生する直前に記録されているアラートデータを学習することで、ある種別の事故の直前に頻出するアラート系列を抽出する手法を検討する。なお、1つの種別の事故の前に頻出するアラート系列は複数存在する可能性があるものとし、その全てを抽出する方法を検討する。

[†] 三菱電機株式会社 Mitsubishi Electric Corporation

2 先行研究とその課題

先行研究[1][2]では、計算機に記録されたログに基づき、軽度のエラー情報から重度のエラーを予測する手法が提案されている。ここで、軽度のエラーがアラート、重度のエラーが事故に相当する。本章では、文献[1][2]で提案されている手法を紹介し、その課題について述べる。

2.1 先行研究の手法

先行研究では、ある種別の事故において、その事故が発生する前の一定時間(観測ウィンドウ)内に、その予兆として特定のアラート群が特定の順序で発生する可能性が高いと仮定する。この仮定に基づき、同一種別の事故の前に共通的に発生しているアラート系列を抽出する。

この手法は、注目している事故種別に関係の深いアラート系列の候補を抽出する候補抽出処理と、各候補の当該事故種別との関連度を判定する判定処理から成る。以下、それぞれについて説明する。

2.1.1 候補抽出処理

k 個の異なる種別のアラートの順列を「 k アラート系列」と呼ぶ。本処理では、アラート系列の長さ k を2から始めて、 k を1つずつ増加させることで、より長い系列の候補を抽出する。抽出した k アラート系列の候補集合を $C(k)$ と置く。 $C(k)$ は次項で述べる判定処理の対象となる。

以下の処理は、観測ウィンドウ毎に実施されるものとするが、出力である集合 $C(k)$ は共有とする。

$k=2$ のときは、観測ウィンドウ内に発生した全てのアラートのうち異なる種別のアラートの順列を $C(2)$ に追加する。

$k \geq 3$ のとき、次項で求める「 $C(k)$ の候補」である $(k-1)$ アラート系列同士の結合によって $C(k)$ を生成する。2つの $(k-1)$ アラート系列に対し、一方の $(k-1)$ アラート系列の最後の $(k-2)$ 個のアラートの種別が、もう一方の系列の最初の $(k-2)$ 個のアラートの種別と等しい場合、その2つの系列を結合する。

2.1.2 判定処理

k アラート系列の頻出判定には、サポートカウントと確信度(後述)を用いる。用語定義として、対象とする k アラート系列において、1番後ろのアラートを後行アラート、後行アラートより前に現れる $(k-1)$ アラート系列を先行アラート系列と定義する。

サポートカウントを定義する。観測ウィンドウ内の全てのアラートのうち、後行アラートと同じ種別のアラートを基準として、それ以前に発生しているアラート系列の中に、部分系列として含まれる先行アラート系列の数をカウントし、これを観測ウィンドウ毎に足し合わせたものがサポート要はなく、アラート間に関係のない種別のアラートが挟まっても良い。

次に後行カウントを定義する。観測ウィンドウ内の全てのアラートのうち、先行アラート系列を部分系列として含む位置を基準とし、それ以降に発生している後行アラートと同じ種別のアラート数を、観測ウィンドウ毎に足し合わせたものを後行カウントと呼ぶ。

サポートカウントや後行カウントを求める際、一度カウントしたアラート又はアラート系列を重複してカウントしないものとする。

確信度とは、抽出された k アラート系列が、同一種別の複数の事故で共通的に発生しているかどうかを判断する指標であり、サポートカウントを後行カウントで割った数として定義される。サポートカウントを後行カウントで割ることで、後行アラートが先行アラート系列に真に起因するかの度合いを定量化している。

サポートカウントと確信度が、予め設定したそれぞれの閾値をどちらも超過する場合、その k アラート系列を「注目している事故種別に関係の深いアラート系列」として出力する。サポートカウントのみが閾値を超えていた場合、その k アラート系列は出力せず、「 $C(k+1)$ の候補」とする。 $C(k+1)$ の候補が1つでも存在する場合は、 $k = k+1$ として候補抽出処理へ移行する。 $C(k+1)$ の候補が1つもない場合は、処理を終了する。

2.2 課題

先行研究の課題として、アラートの継続時間に対応できないことが挙げられる。図 2 (a)に示すように、先行研究が想定するログでは、各アラートの発生時刻のみが記録されており、アラートは継続時間を持たない。一方で、図 2(b)に示すように、本研究では、産業機器から取得したセンサーデータに対する何らかの検知結果をアラートと見做しているため、各アラートは、開始時刻と終了時刻、すなわち継続時間を持つことが多い。このことから、アラートの発生順序だけでなく、アラートの継続時間も考慮するように、手法の改良が必要である。

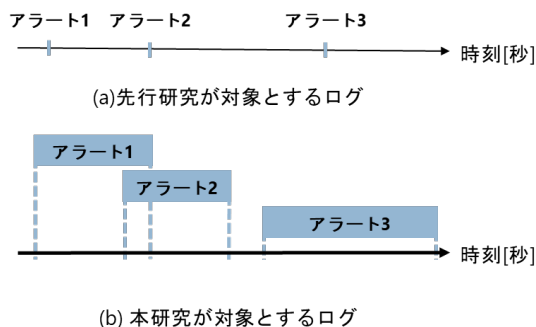


図 2. アラートの継続時間の有無

3 継続時間への対応に向けた検討

提案手法では、アラートの継続時間を考慮に入れ、2点改良を行う。1点目は、アラートの継続時間に基づいて、アラートの種別を分類することである。同一種別のアラートであっても、応用上は継続時間の長さによって、そのアラートの重大性や意味が変わることが多い。本稿では、継続時間を短、中、長の3段階に分類し、元のアラート種別との組合せによって種別を再定義する。

2点目は、アラート同士の時間的な重複へ対応することである。アラートが継続時間を持つと、図 2(b)のアラート1とアラート2のように、重複発生するアラートの扱いを考える必要がある。提案手法では、各アラートの開始時刻を基にアラートの順列を決定する。これは、提案手法を実運用する際に、アラートの発生順序を逐次的に確定でき、複雑な計算を要しないためである。これにより、事故の予兆となり得るアラート系列の候補を早期に検出でき、時間経過に伴う危険度の再計算も容易に実行可能となる。

4 評価対象データ

提案手法の課題抽出を目的とし、今回は、オープンデータである PRONTO heterogeneous benchmark[2]に手法を適用し、評価を行った。PRONTO heterogeneous benchmark は、混相流設備を対象としたアラート及び事故の記録データである。混相流とは、物質の複数の相が混ざり合って流動する現象であり、物質の状態によって気液二相流、固液二相流、液液二相流などがある。本データでは、事故を様々な運転状況下で意図的に誘発し、その際のデータを取得している。

今回の評価では「空気遮断」という事故を対象とした。空気遮断の事故は10回発生しており、今回はその過半数以上で、事故の前の一定時間内に共通的に発生していたアラート系列を抽出できるかどうかを確認した。また、今回のデータでは、4つの位置から継続時間による分類も含めて21種類の種別のアラートが上がる。

評価の結果、46個の頻出するアラート系列を抽出した。抽出した k アラート系列の長さは $2 \leq k \leq 4$ であり、21種類のうち、9つのアラート種別を含んでいた。今回は、アラート系列の発生頻度を確認したのみに留まっており、抽出したアラート系列が実際の事故予兆と見做せるかについて検証できていない。そのため、今後はデータの意味に基づく妥当性検証や、交差検定等による事故の予測精度の検証を行っていく必要がある。

5 おわりに

本稿では、事故発生時の異常データが少なく、発生し得る事故やその発生原因を事前に全て定義できないような分野に対する事故予測手法を検討した。先行研究では、アラートの継続時間を考慮していなかったため、この点について主に改良検討を行った。さらに、提案手法について、オープンデータを使用し定性的な評価を行った。

今後は、提案手法を利用して実際に事故予測が可能かを検証していく。

参考文献

- [1] Xiaoyu Fu, Rui Ren, Sally A. McKee, Jianfeng Zhan, Ninghui Sun, "Digging Deeper into Cluster System Logs for Failure Prediction and Root Cause Diagnosis", IEEE International Conference on Cluster Computing (CLUSTER), p103-112 (2014).
- [2] Xiaoyu Fu, Rui Ren, Jianfeng Zhan, Wei Zhou, Zhen Jia, Gang Lu, "LogMaster: Mining Event Correlations in Logs of Large-scale Cluster Systems", IEEE 31st Symposium on Reliable Distributed Systems, p71-80 (2012).
- [3] Anna, S., Ruomu, T., Yi, C., James R.O., Nina F.T., Jerzy B.: A heterogeneous benchmark dataset for data analytics: Multiphase flow facility case study, Journal of Process Control, Vol.79, pp41-55 (2019)