運用ログを活用したアノマリ事象の検知と活用に関する研究
－障害予兆検知への活用－

アブストラクト

1. 背景

近年、AI やビックデータ解析技術の発展に伴い、アノマリ（いつもと違う状態）による障害予兆検知が可能な環境は整いつつある。セキュリティ分野においては、不正侵入検知・防御を行う IDS（Intrusion Detection System）や IPS（Intrusion Prevention System）の予兆検知システムの実用化が進んでいる。

一方、システム運用の分野では、運用ログを用いた障害予兆検知への活用は進んでいない。システムの重要性が増すにつれ、障害予兆検知へのニーズが高まっているにも関わらず、システム運用の現場でアノマリ検知が進まない要因を、普及学の観点から以下の 3 点が不足しているためだと考えた。

(1)「わかりやすさ」
　アノマリ検知を行う上で、統計学などの専門的な知識が必要とされている点。
(2)「可視性」
　事例が少なく、効果が不明確な点。
(3)「試用可能性」
　商用サービス利用も含めアノマリ検知を行うためには専門知識を持った人材のサポートが必要となり、試用のハードルが高い点。

2. 目的

普及を妨げている要因に対する解決策を導き出し、システム運用の現場で運用ログを活用したアノマリ検知の導入ハードルを下げることが必要であると考えた。具体的には、以下の 3 点を研究対象とした。

(1) 統計学などの専門知識を必要としない手順の確立。
(2) アノマリ検知による障害予兆検知を実証することで効果を明確化。
(3) 低コストで試用可能な条件の整備。

上記を解決し、**「明日から使える（始められる）障害予兆検知」**の成果を得ることを目的とした。

3. アプローチ

障害予兆検知の成功事例として生産設備の故障予兆検知を、データマイニングの標準プロセスである CRISP-DM にあてはめ分析を行い、主に以下の 2 点を課題として設定し対策を行った。

(1) データの理解
　運用ログは、分析対象のログを選択することが重要なのは勿論のこと、そのデータの特徴を把握することが重要かつ難しい課題である。
　対象のシステムを熟知した管理者しか把握出来ない状態では、普及するはずがない。本分科会では、データの特徴の中で特に周期性に着目し、運用ログから機械的に検出する方法を独自に考案した。

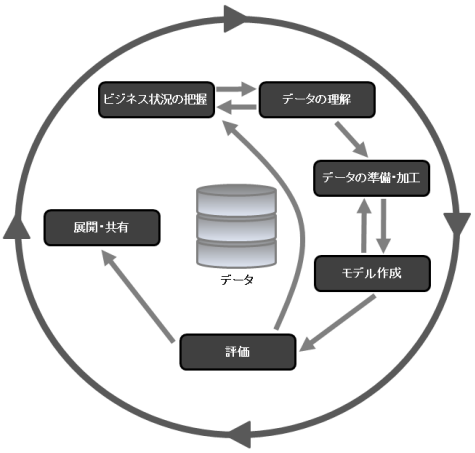

図 1 CRISP-DM のモデル

(2) データの加工・準備
　単純にデータを表形式や数値化するだけでは、満足な結果は得られない。同じ数値であっても、時間帯によって異常とみなされる場合や、みなされない場合があるからである。これを解決するために、本分科会では、周期性を考慮しつつデータを正規化する方法を考案した。

4. 実証結果

実際に発生した以下の障害に対して、本分科会で考案したアプローチを施す前後で検証した。

- 障害内容　　：　Web アクセス遅延事象
- 障害発生時間　：　2016 年 6 月 7 日（下図の四角囲み）
- 対象データ　　：　netstat ログ内の接続済みセッション数（1 か月分）

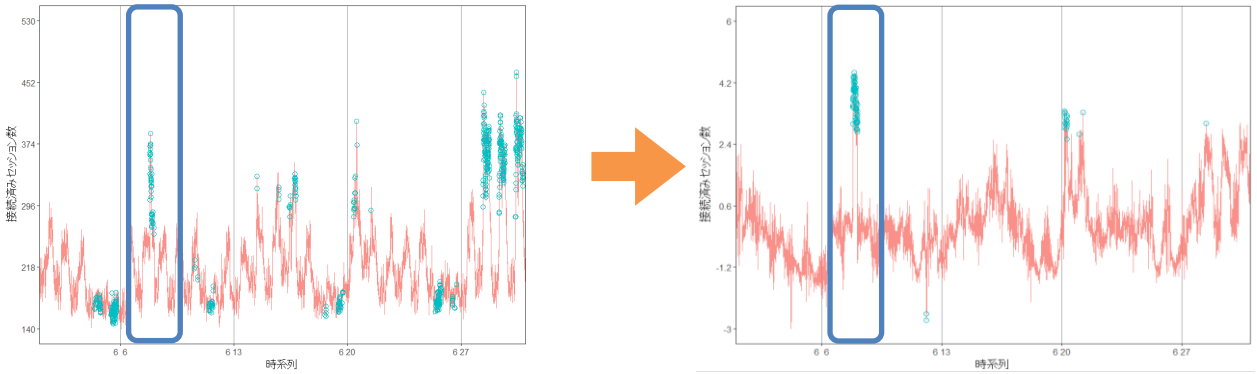

図 2 アプローチ前後での実証結果

アノマリ事象検知には、無償で公開されている Twitter, Inc 製 AnomalyDetection（統計用言語 R で動作）を利用し、評価基準は検知時刻と偽陽性率（過剰検知の少なさ）とした。検知は図の丸印で表記。

(1) 検知時刻：実際の検知時刻よりも加工前で 36 分前、加工後で 31 分前に検知（5 分遅延）。

(2) 偽陽性率：加工前 6.9%から加工後 0.7%（6.2 ポイント改善）。

わずかながら検知時刻は遅れたものの、大幅に誤った検知が削減され、加工による有用性を立証した。

5. 成果

本分科会の成果として、アノマリ検知技術による障害予兆検知の導入ハードルを下げるために、下記の解決案を示した。

(1) 既存ツールの問題点を改善する手法を考案し、利用方法と合わせて整理・体系化した。

(2) 体系化した方法論を活用ツール（プログラム及びマニュアル）として整理した。

上記活用ツールは、既存ツールにおいて普及の妨げとなっていた以下の点で優れている。

(1)「わかりやすさ」

具体的な考え方と手順を含む内容となっているため、障害予兆を検知するにあたり、専門的な知識を必要としない。

(2)「可視性」

本分科会にて実際の運用ログを用いた実証を行い、導入効果の実例として示している。具体的には、実運用に耐えうる精度の確保、及び従来の検知時刻よりも早期の障害検知に成功した。

(3)「試用可能性」

全て無償のソフトウェアにて構成しており、誰でも試用することが可能である。

以上より、システム運用の現場で**「明日から使える（始められる）障害予兆検知」**を実現するための環境を整備出来たと考える。

6. 総括

本分科会では、システム運用の現場へ導入されやすい環境を整備することを念頭に活動してきた。

そして、アノマリ検知を行う上で重要となる「データの理解」と「データの準備・加工」の手法を体系化し、汎用的な活用ツールとしてまとめた。これにより、導入に専門知識が不要になったと考える。

しかし、汎用的であるがゆえに、より精度の高い成果を得るにはカスタマイズが必要となる。今後は、各社がアノマリ検知による障害予兆検知を導入し、更なる事例・手法を公開することでアノマリ検知が普及していくことが強く望まれる。

アノマリ検知が普及することで、システム運用における障害防止、安定運用が図れるようになることを期待する。