1C4-3

制御システムのためのイベントの予測に基づく異常検知

Anomaly Detection based on Event Prediction for Control System

但馬 慶行[*1] Yoshiyuki Tajima　山形 知行[*2] Tomoyuki Yamagata　山本 秀典[*1] Hidenori Yamamoto　志村 明俊[*1] Akitoshi Shimura

[*1] 日立製作所 横浜研究所 Hitachi, Ltd. Yokohama Research Laboratory

[*2] 日立製作所 インフラシステム社 Hitachi, Ltd. Infrastructure Systems Company

It is important for control system represented by a social infrastructure to decrease the loss of system troubles. In this research, we propose a new method for anomaly detection by using system log. Firstly, we construct a prediction model of events that is observed by the log in the method. Then, we evaluate deviation from event probabilities that is calculated by the prediction model. A part of effectiveness in a typical situation has been evaluated by our experiments.

1. はじめに

昨今，社会インフラ等を支える制御システムは年々複雑化してきている．また，国内を含む先進国では，老朽化したシステムの維持保守費用の削減が求められている．一方，新興国では，先進国のように高度なスキルを持った保守員を確保することが難しい．このため，一般的な保守員でも障害対応できるように，[加藤, 13]など異常検知を自動化する取り組みが盛んとなっている．一方，異常検知技術は業務効率低下や機会損失となる事象の発見など業務改善に対する応用も期待されている．そこで本研究では，障害対応迅速化，業務改善のために把握すべき事象（これらをインシデントと呼ぶ）の発見を自動化するログを活用した異常検知手法を提案する．

2. 現状と課題

制御システムの障害対応の現状と課題を述べる．

2.1 障害対応の現状

現在の制御システムは非同期で動く様々なコントローラや計算機等サブシステムの集合体となっており，障害対応は年々複雑化している．大規模な制御システムでは，一般的にサブシステム毎に開発されるため，システム全体を把握できる人材が育ちにくい．この結果，障害やその原因となった事象の発見などの分析業務に関する人材が不足し，対応が長期化している．

2.2 インシデント発見の課題

制御システムは高信頼なつくりとなっており，インシデントの発生回数は少ない．また，システムごとに細部が異なるため，まったく同じインシデントが発生することも少ない．このため，インシデントのパターンを予め網羅的に得ておくことは困難である．さらに，制御システムは，センサから得られる数値データだけでなく，オペレーティングシステムやミドルウェア等が出力するログが大量に含まれる．ログはしばしば人が見てわかるようにテキスト形式となっている．そこで本研究では，インシデントを迅速に発見するために，（課題 1）事前に障害発生時のデータを必要とせず，（課題 2）テキスト形式のログに対応した，異常検知手法の確立を目指す．

3. 提案手法

本章では提案手法について基本原理と詳細を説明する．

3.1 異常検知の基本原理

制御システムは，周期的に同じような振舞いを繰り返す．そのため，正常時には見られなかった振舞いが生じた場合，何らかのインシデントが起きていることが多い．そこで，課題 1 を解決する方策として，正常時の振舞いから逸脱した振舞いが見られたとき，それを異常と捉え検知する．

この異常検知のフレームワークを簡単に説明する．制御システムが出力するログには，データ送受信，プロセスの開始や終了，部分交換や消耗品切れなどのアラートなど，制御システムのある側面の振舞いを表すイベントが時系列で記載されている．そこで，異常検知処理に先立って，試験運転時や異常が起きていない運用時等のログを用い，イベントの特徴量の生起確率を予測する統計モデルを構築しておく．運用時は，まず，観測した直近のログから将来起こりうる各イベントの特徴量の生起確率を予測する．そして，イベントの特徴量の生起確率と，その後実際に観測されたイベントの特徴量との間の逸脱具合を評価し，逸脱が大きい場合，異常であると判定する．

3.2 予測モデルとその学習

予測モデルは，観測したイベントから将来起こりうる各イベントの特徴量の生起確率を予測する．この予測モデルを条件付確率としてモデル化する．すなわち，ある時刻$\mathrm{t}-\tau_2$から時刻$\mathrm{t}-\tau_1$までのログL^xから得られるイベントの特徴量 X，時刻$\mathrm{t}-\tau_1$から tまでのログL^yから得られるイベントの特徴量 Y とする．ここで，$\tau_2 > \tau_1 > 0$である．このとき予測モデルは P(Y|X)となる．P(Y|X)には様々な統計モデルを用いることができる．本研究では，後述する特徴量との相性から，式(1)で示すような対数線形モデルを用いるものとする．

$$\mathrm{P}(\mathrm{Y}=\mathrm{Y_i}|X)=\frac{\exp(W\cdot\Phi(L^x,L_i^y))}{\sum_j\exp(W\cdot\Phi(L^x,L_j^y))} \quad (1)$$

ここで，W は重みベクトル，$\Phi(\cdot)$は，L^xとL^yから特徴量を算出する特徴関数（素性関数）である．

連絡先：但馬慶行，(株)日立製作所 横浜研究所，〒244-0817 神奈川県横浜市戸塚区吉田町 292 番地，045-860-3093，yoshiyuki.tajima.hh@hitachi.com

3.3 イベントの特徴量の算出方法

課題 2 で述べたとおり，テキストに対応する必要がある．しかしながら，テキストは特徴量の定義が自明ではない．そこで，特徴関数の一つ一つの要素を文字列検索に対応付け，検索に合致する文字列があれば 1，なければ 0 となるようにする．つまり，特徴関数$\Phi(L^x, L^y) = \{\phi_1(L^x, L^y), \phi_2(L^x, L^y), \ldots\}$とするとき，各$\phi_i$が，例えば，$L^x$に”ID100 ACTION1”を含み，かつ，$L^y$に”ID200 ACTION2”を含むかどうかといった文字列検索に対応させる．なお，各ϕ_iは検索パターンの論理演算等によって拡張することもできる．

3.4 異常検知の手順

異常は，起こると思っていたのに起きなかった（偽陽性，α 過誤），起きないと思っていたのに起きたといった（偽陰性，β 過誤），といったこと判定することで検知する．より具体的には，図 1 に示すような手順で判定する．なお，予測モデルを構築した際に，予測モデルの予測精度を算出し，予測が難しいイベントを評価対象から除外することで，誤検知を低減することができる．

1. 観測したログからイベントの特徴量 X を算出
2. 発生しうるイベントの特徴量 Y について以下を実行:
   2. 1 P(Y|X)を算出
   2. 2 以下を実行:
      a)Y が実際には発生していない場合（偽陽性判別）:
         P(Y|X) > α ならば異常と判別
      b)Y が実際に発生している場合（偽陰性判別）:
         P(Y|X) < β ならば異常と判別

図 1 異常検知の手順

4. 評価実験

トラックをキャリアとする搬送システムのシミュレータにより原理検証を行った．これについて説明する．

4.1 実験内容とシミュレータの概略

多くの搬送システムでは稼働率が重要視される．そのため，致命的な障害でなくとも，稼働率を低下させるインシデントは排除もしくは早期対策できることが不可欠となる．これを踏まえ，本実験では搬送システムでの稼働率低下の要因となるインシデントを提案手法で検知できるか確かめた．以下に詳細を述べる．

搬送システムは，キャリアとなるトラックと，トラックに経路の指示を与える指示機能からなる．シミュレータは，これらに加え，図 2 に示すように荷物を受け取る搬入エリア，荷物を下ろす搬出エリア，その間をつなぐ道路といった環境を含む．シミュレータは，既存研究[Tajima, 12]のキューモデルに基づいて構築した．

本実験では，シミュレーションの途中でトラック 1 台の動作が他のトラックに対して遅くなるように設定することでインシデントを模擬した．そして，このインシデントを提案手法により正しく検知できるかどうか評価した．

シミュレーションの詳細な設定は次のようにした．トラックの台数は 20 台とし，各トラックはそれぞれ搬入エリアのどれかで荷物を受け取り，搬出エリアのどれかで荷物を降ろすといった動作を約 15 分で繰り返すものとした．また，トラックが荷物を搬入もしくは搬出した際，トラックの所在を伝えるイベントがログに記録されるものとした．このイベントには，図 3 に示すような時刻，トラック ID，搬入もしくは搬出エリアを表すエリア ID が含まれる．トラックの経路選択に関しては，最も混雑していないエリアをグリーディに選ぶようにした．

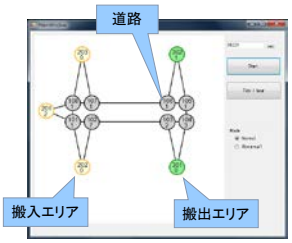

図 2 シミュレータの実行画面

経過時間(秒)	トラックID	エリアID
4203.1,	9,	203
4233.2,	17,	301
4288.3,	11,	201
4293.3,	6,	301
4353.4,	10,	301
4358.5,	13,	202

図 3 シミュレータのログ

4.2 本手法の適用方法と実験結果

本手法適用にあたり，トラック ID とエリア ID の組合せを一つの要素として，この要素の論理演算（積・和）および順序関係に基づく約 1 万次元の特徴量を生成するものとした．例えば，時刻tの特徴量は，L^xを時刻t − 7分～t − 2分，L^yを時刻t − 2分～時刻tまでのログとして，L^xの中にトラック ID=2，エリア ID=201 となるレコードがあり，L^yの中にトラック ID=19，エリア ID=301 となるレコードがあれば 1，なければ 0 を返すといったものとした．予測モデルの学習は全てのトラックが正常に動作した場合のログを使って行った．特徴量はスパースとなるため，学習には L1 正則化付きの確率的勾配法である FOBOS アルゴリズム[Duchi, 09]を用いた．また，異常検知では偽陽性（α =0．95）のみを評価した．評価実験の結果を図 4 に示す．正常な期間で一部誤検知される場合があるものの，異常な期間で多くの異常を検知することができていることが確かめられた．

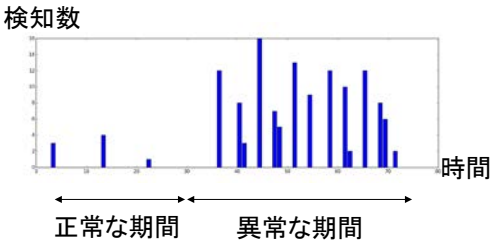

図 4 実験結果（検知数の時間推移）

5. おわりに

本稿では制御システムを対象にログに現れるイベントの予測に基づいた異常検知手法を提案した．実験の結果，時系列の異常を検知できることを確認した．今後は，特徴量の自動生成と縮約，および，より高精度な予測モデルの構築を行えるよう手法を拡張していく．

参考文献

[加藤, 13] : 加藤清志, 矢吹謙太郎: WebSAM の分析技術と応用例～インバリアント分析の特長と適用領域～, NEC 技報 Vol. 65 No. 2, 2013

[Tajima , 12] Yoshiyuki Tajima , Takashi Noguchi , Takashi Fukumoto: Predictive Transportation Control of a Complex Dynamical System for High TP and Short TAT, IEEE 7th International Conference on System of Systems Engineering, 2012

[Duchi , 09] John Duchi , Yoram Singer: Efficient Online and Batch Learning Using Forward Backward Splitting, Journal of Machin Learning 10, 2009