ログデータ分析による通信ネットワーク機器の異常検出技術

～ログ分析業務の効率化を目指して～

Anomaly Detection Technology for IP Network Devices by Data Analysis

~To make data analysis work efficient~

（エネルギー応用研究所　ネットワークG　通信T）

昨今、センサ情報や設備の巡視点検記録・障害記録など、ビッグデータの有効活用が注目を集めている。

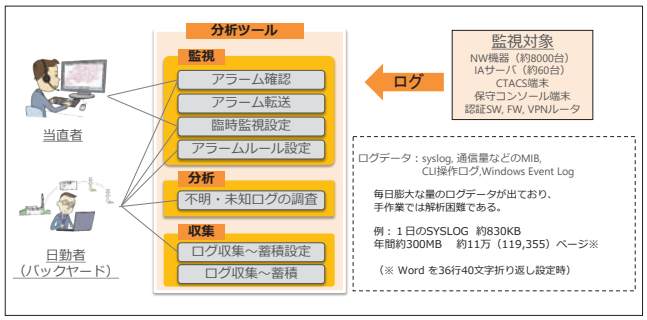
本取り組みでは、IPネットワーク機器から発生する大量のログデータに対して統計解析やAI（Artificial intelligence）を用いることにより、異常の検出を試行し、有効性を評価した。

(Telecommunication Team, Network Group, Energy and Applications Research & Development Center)

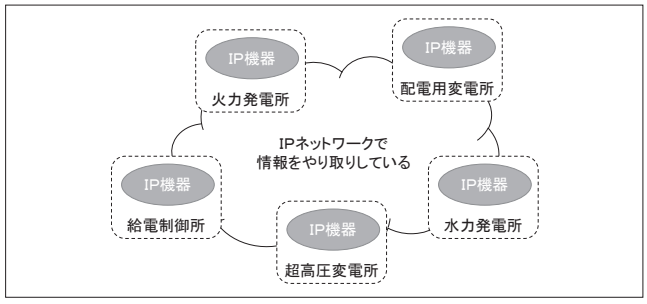
Recently, to use big data (sensor data, patrol data, inspection data, and failure data) of utility facilities effectively is attracting attention.
Therefore, we tried to detect anomalies by using statistical analysis and AI for a large amount of log data generated from IP network devices, and evaluated the effectiveness.

1 はじめに

電子通信部門では、経験豊かな技術者（スペシャリスト）が、異常を発見するため、膨大なログデータを手作業で個別に分析している（第1図にログ分析業務イメージ、第2図に当社IPネットワークイメージ）。今回、ログ分析ツールの機械学習（AI）機能を活用して、異常を自動的・効率的に検出することに取り組んだ。なお、分析ツールはIT機器のログデータ分析に定評のあるsplunkを使用した。



第1図　ログ分析業務イメージ



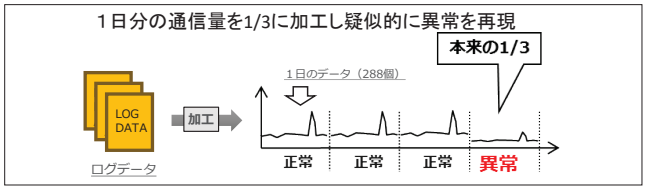
第2図　給電用IPネットワーク（イメージ図）

2 通信量の変化による異常兆候検出

ネットワーク上を通過する通信量の変化に着目しAIによる異常兆候の検出を試みた。また、AIによる解析が、従来方式よりも優れていることを確認するため、統計的手法でも異常兆候の検出を試みた。

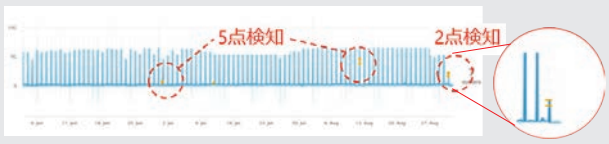
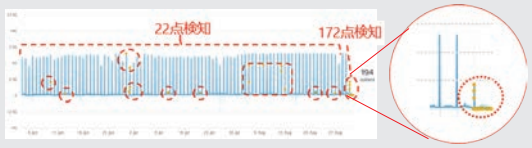
(1) AIによる異常兆候検出

系統運用情報や電気所TCなどを伝送している当社の給電用IPネットワークは、通信量に周期性がある事が特徴であるが、実際の通信量の変化と異常を関連付けたデータになってない。従って、検出アルゴリズムとしては、周期性がある通信量の予測に適し、教師データが不要な異常値検出機能（Karlman Filterを活用した機能）を使用した。検証用のデータは、3カ月間の通信量をベースに、1日の通信量を1/3に加工した疑似的な異常を付加して使用した。（第3図）



第3図　検証用データ

第1表　異常検出の検証結果

	統計解析　時分毎に標準偏差（±5σ）外を異常として検出		AI　カルマンフィルタ
			
適合率（Precision）	検出できた真の異常数 / 検出した異常数	0.2857（2 ／ 7）	0.8865（172 ／ 194）
再現率（Recall）	検出できた真の異常数 / 検出すべき真の異常	0.0069（2 ／ 288）	0.5972（172 ／ 288）
評価		×	○

AIによる検証は、収集した3カ月間の実機通信量の周期性をAIに分析させた上で、異常検出精度を評価した。一方、統計解析では、検証用データの1日周期における測定時分毎の通信量の標準偏差（±5σ＝99.9994%）を閾値とし、そこから外れるものを検出させた。

(2) 結果

AIでは、疑似した異常288個に対して172個を検出した。AIと統計解析が正しく異常を検出できた割合を比較すると、AIの方が優位であった。(第1表)

(3) 課題

AIは統計解析よりも高い精度で異常兆候検出ができたが、実際の運用に当たっては様々な異常データを用意して検証を重ねる必要がある。

3 検証用NWでの異常兆候検出

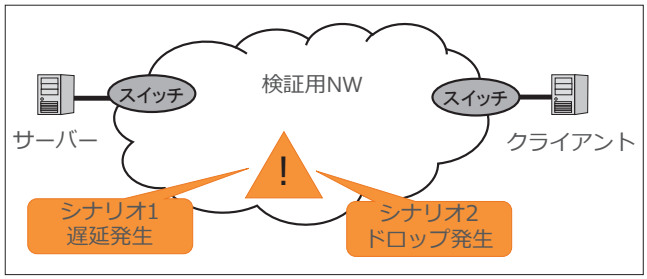
様々なデータで異常兆候検出を検証するため、検証用ネットワークにより多種多様のデータを作り出した。

＜具体的方法＞

検証用NW（第4図）で、通信遅延（200m秒）とパケットドロップ（50%）を模擬障害として発生させ、詳細なMIB情報（装置の状態や処理状況を収めたデータ）を収集した（第2表）。

＜収集データから観測できたこと＞

・遅延発生→通信量減少
・遅延発生→Round-Trip Time（往復時間）増加
・ドロップ発生→パケットロス



第4図　検証用NWイメージと収集情報

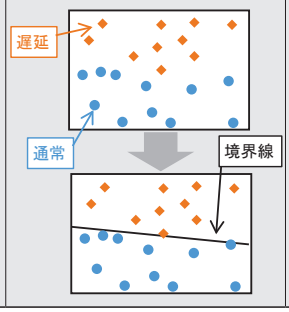
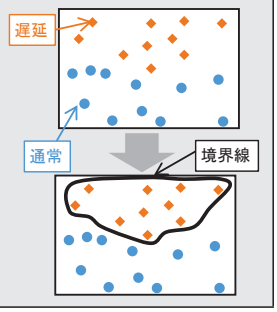
第2表　収集したMIB情報

情報種別	収集状況
① CPU使用率 ② Memory使用率 ③ Traffic情報 ④ Ping情報 ⑤ Round-Trip Time ⑥ パケットロス	250分間 15秒毎 （実機3.5日に相当）

(1) AIによる異常兆候検出

splunkが実装する教師あり学習機能のうち、時系列データに適した二種類のAI機能（ロジスティック回帰とサポートベクターマシン）を用いて分析を試みた。各機械学習の特徴などについて、第3表に示す。今回は、正解データに基づいて学習させる方法である、いわゆる「教師あり学習」による分析とした。前述の検証用NWのデータから収集したデータの約90%について通常／遅延／ドロップのラベル付けをして学習データとし、残りを検証データとした。(第4表)

第3表　各機械学習の特徴と学習イメージ

名称	ロジスティック回帰	サポートベクターマシン
特徴	2値予測に利用できるアルゴリズム。何かを買った、買わない、のような確率が「1－0」の場合に利用できる	2つのデータ群をグループに分けするアルゴリズム。非線形データでも分類が可能。
イメージ		

第4表　AI（教師あり学習）用データ

データ種別	学習用データ［個］	検証用データ［個］	備考
通常	120	0	
遅延（模擬故障）	120	40	200m秒
欠落（模擬故障）	120	40	50%
合計	360	80	

(2) 結果

両AI機能とも、遅延およびドロップを100%の精度で検知できた。これは、今回の検証用データは、作業や故障などで発生するログデータが入っておらず、学習用データと検証用データの変化の仕方が同じであったからと推定した。

また、異常発生前にその兆候を検知することはできなかった。これは、収集した詳細データを分析したが、兆候検知につながるような特徴的な現象をとらえる事ができなかったためであると推定した。

(3) 課題

異常兆候検出に向けては、更に多くの条件での確認が必要であるが、障害発生時の実データは教師あり学習として整っておらず、十分な検証が出来なかった。今後は教師データをいかに準備するかが課題である。

4 まとめ

今回の試行によりAIによる異常兆候検出は有効であると評価できた。今後は実運用に向けて様々な障害発生時のデータを分析し評価をしていく。

執筆者／田中弘生