中部電力

委託研究
シスコシステムズ
ユニアデックス

ビッグデータを分析して異常検出を行います

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

背景・目的

- 経験豊かな技術者（スペシャリスト）が、膨大なログデータから通常と異なる振る舞いに気づき、個別に分析を行い、異常を発見することがあります。
- 今回、ログ分析ツールの機械学習（AI）機能を活用して、自動的・効率的に異常を検出することに取り組みました。

特長

- ①運用中ﾈｯﾄﾜｰｸのﾛｸﾞﾃﾞｰﾀ（実ﾃﾞｰﾀ）を用いてAI（教師なし学習）による異常検出技術を検証。
  AIを用いない場合（統計解析）より高い精度で異常を検出できたが、一部誤検出もあった。
- ②検証用ﾈｯﾄﾜｰｸで生成した模擬ﾃﾞｰﾀを用いてAI（教師あり学習）による異常検出技術を検証。
  誤検出なく異常を検出できた。

用途

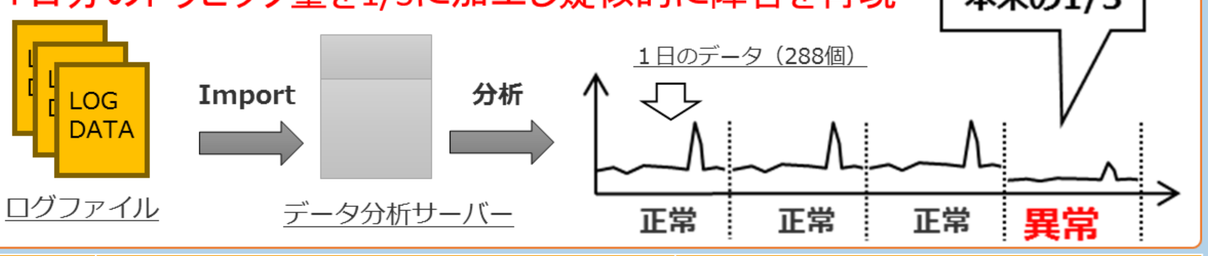
- ログ分析業務の自動化・効率化（特別なスキルやノウハウがなくても異常を検出できる）
- サイレント故障を含む異常兆候の検出（AIが未知の異常を検出する）

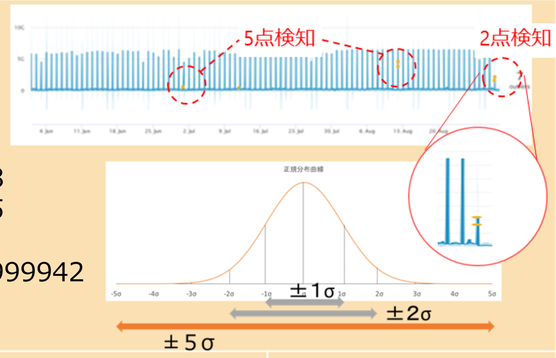
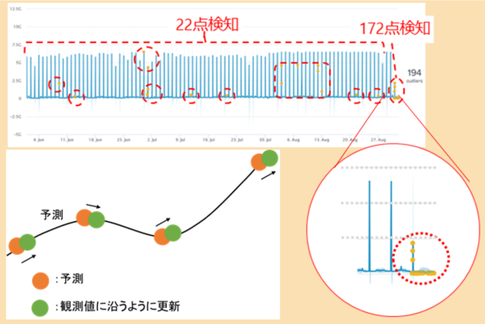
①教師なし学習による異常検出

AI（ｶﾙﾏﾝﾌｨﾙﾀ※）を用いて、ﾄﾗﾋｯｸ量の変化傾向を予測させ、異常（ﾄﾗﾋｯｸ量1/3低下）の検出に成功した（統計解析より高精度で検出）。

※時系列ﾃﾞｰﾀを次々に予測するｱﾙｺﾞﾘｽﾞﾑ

1日分のトラヒック量を1/3に加工し疑似的に障害を再現



	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
	 ±1σ＝0.68 ±2σ＝0.95 ： ±5σ＝0.9999942		
適合率 Precision	検出できた真の異常数 / 検出した異常数	0.2857 $(\frac{2}{7})$	0.8865 $(\frac{172}{194})$
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 $(\frac{2}{288})$	0.5972 $(\frac{172}{288})$
F値 予測精度の評価指標	2×適合率×再現率 / 適合率＋再現率	0.0137	**0.7146**

②教師あり学習による異常検出

2種類のAI(ﾛｼﾞｽﾃｨｯｸ回帰、ｻﾎﾟｰﾄﾍﾞｸﾀﾏｼﾝ※)を用いて、模擬故障（遅延・欠落）を学習させ、異常の検出に成功した（精度100%）。

※時系列ﾃﾞｰﾀに適した教師あり学習ｱﾙｺﾞﾘｽﾞﾑ

表　AI（教師あり学習）用データ

データ種別	学習用[個]	検証用[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	

ロジスティック回帰での検出

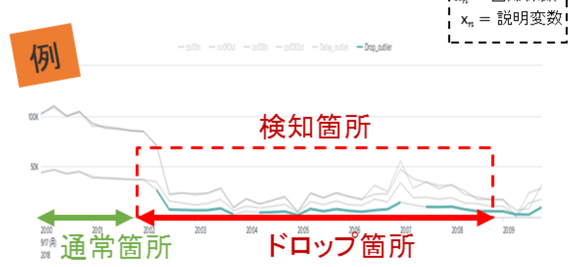
2値分類のアルゴリズム
買った、買わない、のような
「1－0」の場合に適用できる

今回は、
「通常－遅延」
「通常－欠落」
を分類した。

$$y = \frac{1}{1 - \exp(a_1x_1 + a_2x_2 + \cdots + anx_n + b)}$$

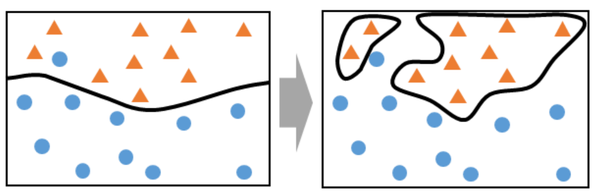
y ＝ 応答変数
a_n ＝ 回帰係数
x_n ＝ 説明変数

例



サポートベクターマシンでの検出

データ群を2つのグループに分類するために使用される手法
（分類対象はロジスティック回帰と同じ）



例

開発者のひとこと

機械学習により一定の精度で異常検出できました。
業務への適用に向けては更に多くの条件での確認が必要であり、
教師情報となる実際の障害発生時のデータをいかに確保するのかが課題です。