L-005

機械学習によりサーバ異常を検知する管理システムの試作

A Management System to Detect Server Anomalies with Machine Learning.

堀 壮吾† 田島 孝治‡ 山田 博文‡
Hori Sougo Tajima Koji Yamada Hirobumi

1. はじめに

近年，ほとんどの企業がホームページを運用している．ホームページの運用にはサーバ管理が必要である．自社に IT 部門を持たない企業や官公庁ではそのサーバ管理を外部委託する場合がある．外部委託を受けた企業は IP アドレスやホスト名などのサーバ管理のための情報と顧客の情報を管理するとともに，状態把握が必要である．このような企業ではサーバに異常が発生した場合にメールを管理者に送信するシステムを使用しサーバの状態把握を行っている．このようなシステムではメモリなどのリソースが少なくなった場合にメールが送られてくる．これらのリソース使用率は，サーバの混雑度を示すものでありサーバの状態異常を示すものではない．サーバのリソース使用率にはサーバごとに特色がある．人間ならば特色を考慮して，正常，異常の状態判別ができるが，すべてのサーバについて人が常に監視するのは困難である．本研究ではこの問題について機械学習を用いて解決する手法を提案し，サーバの異常検知と管理が行えるシステムの試作を行う．

2. 先行研究

サーバの異常検知の学習手法の調査として小泉氏らの研究を挙げる[1]．この研究では大規模なデータセンターにおいてリソース使用量の異常検知を機械学習によって行っている．特に着目しているのは，リソース使用量の相関である．従来手法である主成分分析，確率的主成分分析では単一のリソース使用量にのみ着目するため，相関を考慮することができない．この研究ではスパース構造学習という学習手法を用いて，相関を考慮した異常検知を行った．リソース使用量の相関を考慮した結果，検知精度の向上が見られ，いくつかのサーバデータを学習させることで，異常検知精度が向上することが確認された．

次に AWS AI Labs が発表した Neural Contextual Anomaly Detection(以下 NCAD)を挙げる[2]．NCAD とは教師なし，教師あり学習をシームレスに取り扱うことが可能な時系列異常検知の機械学習手法である．この学習手法の特徴として，教師なし学習でありながら少量の異常データを追加データとして取り込み異常検知精度を向上させることができる．また，論文内でサーバマシンデータセットに対する実験を行っており，38 時系列のサーバデータに対して 80%の異常精度が確認された．

3. 予備実験

予備実験として One Class SVM という手法を使用し，岐阜高専サーバの異常検知を行った． SVM とは，分類や回帰などの問題に適用できる機械学習モデルであり，データを 2 つのクラスに分類する超平面のうち，各データから最も離れている超平面を決定する手法である．One Class SVM(以下 OCSVM)は，SVM を教師なし学習に改良したものであり，サーバの状態のような極端に異常状態が観測されないデータの学習に多く使用されている[3]．

実験を行うために,以下の 3 種類のデータを用意した．

① CPU 使用率データ 5 日分
② ①の 18:00-20:00 のデータを複製したもの
③ ①とは別日の CPU 使用率データ 5 日分

それぞれのデータは 1 分毎に CPU 使用率を計測しており 1 日あたり 1000 分のデータがある.これらのデータを重複ありの 30 分間隔で分割したデータを 1 データとして,学習および判定を行った.そのため 1 日あたりのデータ数は 970 である.データ①，②を OCSVM で学習させ,データ③の異常検知を行う．

表 1 予備実験結果

使用データ	データ①	データ②
異常検知数(分)	49	492

実験結果を表 1,図 1 に示す．この結果からデータ①を学習させた場合には,異常検知数が少量であるためそのサーバの特色を学習できているとわかる．

データ②を学習させた場合には,異常検知数が多量であるため,サーバの特色を全く学習できていない.これらの結果から,OCSVM を使用して特色を考慮した機械学習を行うことは可能であると推測される．

4. 課題

本システムの実現には(1)情報の収集方法，(2)学習項目の選定が課題となる．

情報の収集方法として，顧客サーバからの直接取得を想定している．予備実験で利用したサービスでは，情報が 1 ヶ月しか保存されないことに加え，取得可能なデータにも限りがある．そこで，各サーバに Sysstat を導入する．Sysstat は Linux 用サーバの負荷を分析するための監視ツールである[4]．Sysstat で収集できるリソース情報の加工を行い，状態判別システムに送信する．

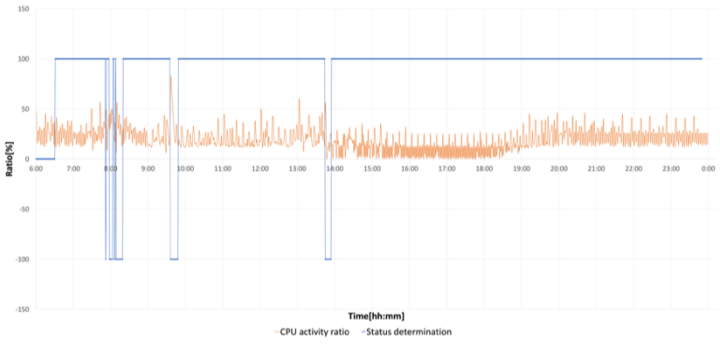

図 1 予備実験結果：赤線が CPU 使用率，青線が正常異常の判別を示している

† 岐阜工業高等専門学校 先端融合開発専攻
National Institute of Technology, Gifu College Advanced Course
‡ 岐阜工業高等専門学校 電気情報工学科
National Institute of Technology, Gifu College Dept. of Electrical and Computer Engineering

Copyright © 2023 by
The Institute of Electronics, Information and Communication Engineers and
Information Processing Society of Japan All rights reserved.

学習項目は，予備実験で行った CPU 使用率だけでなくディスク使用率やメモリスワップなどを考慮した学習を行う必要がある．特に注目している項目は平均ジョブ待ち数である．平均ジョブ待ち数を学習することで，そのサーバに適したリソース設定がされているかの確認が可能となると考えられる．加えて，昨年は 1 つの項目に注目し学習を行ったため，複数の学習項目を使用した場合の学習結果についても調査する必要がある．

5. 提案方式

5.1 状態判別システム

状態判別システムには(1) 機械学習による判別，(2) 管理者への通知の 2 つの機能を実装する．機械学習による判別は先行研究で使用されていた NCAD で実装する．学習データは各サーバから送られてくるデータをデータベースに保存し，それらのデータから学習を行う．管理者への通知では，異常発生時に管理者へメールを送信する．このメールにはどのリソースでどのような異常が発生したのかを詳細に記載し，管理者が容易に異常を把握できるよう実装する．

5.2 情報管理システム

情報管理システムには(1)サーバ・顧客情報の管理，(2)現在状態の表示の 2 つの機能を実装する．サーバ・顧客情報の管理では，登録，編集，削除，検索の機能を実装するとともに，アイコン，ボタンで視認性の向上を行う．現在状態の表示では，状態判別システムから送られてくる正常，異常の情報を表示する．またサーバの詳細画面に遷移することによって，サーバの現在のリソースを確認可能にし，リソースのグラフ表示についても実装する．

6. 進捗状況

状態判別システムでは，NCAD の調査，NCAD を使用した機械学習の試作を行った．

情報管理システムでは，サーバ・顧客情報の管理機能の開発，サーバの現在状態の表示の実装を行った．開発した画面を図 2 に示す．

図 2 情報管理システム表示画面

また，クラウドサーバを使用しサーバのリソースについて調査を行っている．このサーバに Sysstat とゲームサーバ用のシステムを導入しリソースの変化について調査を行っている．実際に CPU 情報を取得した結果を図 3 に示す．この結果はシステムレベルでの CPU 使用率を表示している．

7. まとめと今後の予定

本稿では，サーバ異常を検知する管理システムの提案・開発について述べた．現在までに，機械学習手法である NCAD の調査を行い，サーバ・顧客情報管理機能の開発を行った．加えて，実際にクラウドサーバ上に Sysstat を導入し，サーバ利用時に注目するべきリソースについて調査を行っている．

今後は 2 つのシステムの開発，着目すべきリソース情報の調査を続けていく．状態判別システムに関する開発として，各サーバからリソース情報の取得方法，NCAD を用いた機械学習システムの開発を行う．特にリソース情報の取得方法では，Sysstat が使用している sa ファイルを加工し各サーバから取得する方式を実装する．情報管理システムに関する開発として，システムの利便性の向上，状態判別システムからの結果表示，リソースのグラフ表示を行う．着目すべきリソース情報の調査では，アクセスによる負荷，データ I/O による負荷などサーバにかける負荷の種類を変更しながら，それぞれの負荷によるリソースの遷移についても調査していく．

23:06:01	CPU	%user	%nice	%system	%iowait	%steal	%idle
23:07:01	all	1.60	0.00	0.55	0.00	0.00	97.85
23:08:01	all	1.58	0.00	0.57	0.00	0.00	97.85
23:09:01	all	1.78	0.00	0.54	0.01	0.01	97.66
23:10:01	all	6.77	0.00	0.84	0.65	0.03	91.70
23:11:01	all	5.91	0.00	0.54	0.02	0.09	93.43
23:12:01	all	7.50	0.00	0.80	0.04	0.00	91.65
23:13:01	all	7.79	0.00	0.83	0.10	0.00	91.27
23:14:01	all	9.28	0.00	0.84	0.20	0.00	89.68
23:15:01	all	8.58	0.00	0.68	0.87	0.03	89.83
23:16:01	all	5.51	0.00	0.60	0.11	0.05	93.73
23:17:02	all	10.02	0.00	0.82	0.16	0.00	89.00
23:18:01	all	14.31	0.00	1.14	0.15	0.00	84.40
23:19:01	all	23.45	0.00	1.02	0.26	0.01	75.26
23:20:01	all	16.93	0.00	0.97	0.78	0.00	81.31
23:21:01	all	9.00	0.00	1.01	0.41	0.33	89.25
23:22:01	all	8.81	0.00	0.98	0.19	0.01	90.02
23:23:01	all	5.62	0.00	0.73	0.08	0.00	93.56
23:24:01	all	4.04	0.00	0.69	0.09	0.00	95.18
23:25:01	all	1.51	0.00	0.40	0.23	0.00	97.86
23:26:01	all	1.57	0.00	0.42	0.01	0.10	97.89
23:27:01	all	1.58	0.00	0.47	0.00	0.00	97.95
23:28:01	all	1.55	0.00	0.44	0.00	0.00	98.00

図 3 Sysstat CPU 情報

参考文献

[1] 小泉成司,鮫島正樹,菅野裕介,松下康. “スペース構造学習によるサーバの異常検知”, 情報処理学会,No.11, (2017).
[2] Chris U. Carmona, François-Xavier Aubet, Valentin Flunkert, Jan Gasthaus. “Neural Contextual Anomaly Detection for Time Series.”, Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, (2021).
[3] 五井雅登,片山昇,盛田克彦,大川浩,小杉明史,今井庸二.機械学習を用いた太陽光発電の異常検知. Proceedings of the International Council on Electrical Engineering Conference, 2019, pp. 1-6.
[4] SYSSTAT.”Documentation”,2022/4/6,http://sebastien.godard.pagesperso-orange.fr/documentation.html .

Copyright © 2023 by
The Institute of Electronics, Information and Communication Engineers and
Information Processing Society of Japan All rights reserved.