

資源効率の良いプライベートクラウド運用を実現するための機械学習の活用

○住田 宏己[†] 吉本 安男[†]
富士通(株)[†]

1. はじめに

多くの企業がプライベートクラウドシステムを導入し、仮想化による IT 機器の資源効率の改善と設備費の削減を進めている。我々はプライベートクラウドの資源効率をさらに高めるため、閾値をもとに仮想マシン (VM: virtual machine) の稼働状態を識別して資源の無駄遣いを削減する『選択的 VM 片寄せ方式』[1]を開発し実践してきた。ただし、稼働状態を識別するための閾値の設計には性能評価技術が必要なため、熟練者がいない場合には適用できなかった。そこで、自律的な機械学習で分類器を構築することにより、熟練者に頼ることなく VM の稼働状態を識別する方法を探っている。本稿では分類器の試作結果について報告する。

2. 選択的 VM 片寄せ方式の概要と課題

対象は、社内の技術者にソフトウェア開発環境を提供するプライベートクラウドである。このシステムでは開発チームごとに必要な VM を必要な時点で利用者自身が操作して配備できる。開発環境を手軽に展開できるので重宝されているが、一方で資源の無駄遣いが生じている。

配備する VM は、開発チームの事情に応じて運用時間を自由に設定できる。例えば、月曜日の早朝に起動し、金曜日の深夜に停止させるような運用を可能にしている。利用者にとっては日々の VM の起動・停止に煩わされることなく開発業務に専念できるというメリットがある。ところが実際の運用では、週の初めに起動された後、数日の間、誰にも使われないままになっている VM が、かなりの割合で見つかる。このような VM をアイドル VM と呼び、アイドル VM 以外を利用中 VM と呼ぶことにする。

仮想環境で VM を起動し過ぎると実メモリ不足に起因する性能低下のリスクを伴う。リスクを避けるためには VM の配備総数を制限すればよいが、それだけではアイドル VM によるメモリ資源の無駄な占有は無くならない。

『選択的 VM 片寄せ方式』(図1)は、アイドル状態が長期間続いている VM を業務用サーバから追い出すことで、実メモリを有効に活用することを狙ったものである。業務用サーバ上のアイドル VM は特別に用意した片寄せ専用サーバに追い出し、利用中 VM に転じた時点で業務用サーバに戻す。アイドル VM しか存在しなければ実メモリの奪い合いにより利用者が迷惑するという問題は発生しにくい。片寄せ専用サーバには実メモリだけで賄える上限を超えてアイドル VM を詰め込むことができる。

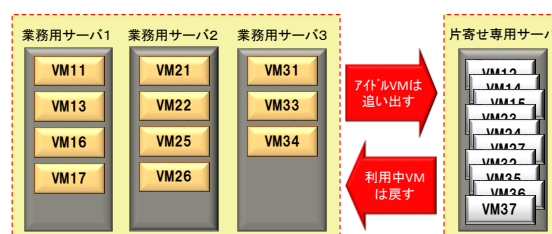


図1. 選択的 VM 片寄せ方式

我々は、VM ごとの資源使用量の推移を記録し、閾値を基準にアイドル VM を識別する方式を採用してきた。先行研究[1]で示したように、システム環境の経年変化に伴い、適宜、閾値を見直す必要がある。しかし、閾値を設計できる熟練者は限られていることから、『選択的 VM 片寄せ方式』を容易に導入することができなかった。そこで、閾値を前提としない、機械学習による分類器の構築に取り組んだ。

3. アイドル VM を識別するための分類器

3.1 学習モデル

教師データを準備するコスト、特に人手によるラベル付け作業を軽減することを狙い、学習モデルには Deep Autoencoder を採用した(図2)。

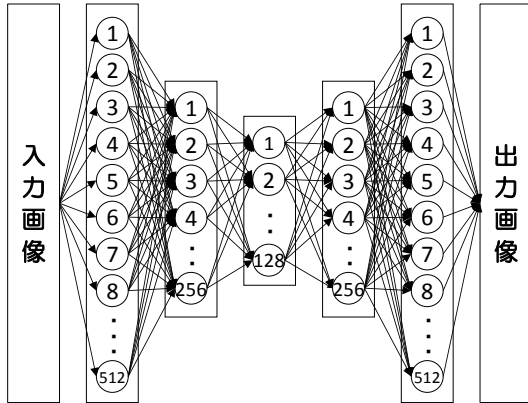
3.2 教師データ

熟練者は、VM ごとの資源使用量の時間推移をもとに経験に基づいてアイドル VM を識別している。熟練者のやり方を踏襲し、資源使用量(特に CPU 使用率に着目)の時間的な推移をグラフ表示した画像を教師データとした(表1)。

Use of machine learning to achieve private cloud operation with efficient use of resources

[†]Hiroki Sumida, Yasuo Yoshimoto
Fujitsu Ltd.

教師データを自動的に収集するために、アイドル状態で放置されていることが判っている VM のデータを利用することにした。このやり方であれば、業務で利用しない VM を予め配備しておくだけで、アイドル VM の教師データを自動的に収集できる。



- ・ 中間層は5層で全結合
- ・ 活性化関数はReLU
- 注) ノードやエッジは一部のみ記載

図2. 学習モデルの構成

表1. 教師データ

グラフデータ	アイドルVMのCPU使用率
グラフ構成	1日の中でのCPU使用率の推移
データ粒度	CPU使用率の5分間の平均値
グラフ種類	棒グラフ
画像形式	jpg形式、size=(198, 304)
色彩	白黒(背景は黒、グラフは白)
画像数	6798枚(2015年6月1日～12月20日)

3.3 分類方法

Autoencoder の入力画像と出力画像の cos 類似度をもとにアイドル VM か否かを分類することにした。2015年7月1日～7日の期間の VM のうち教師データ以外の画像について cos 類似度を測定した(図3)。

熟練者が利用中 VM だと識別した画像 56 件では

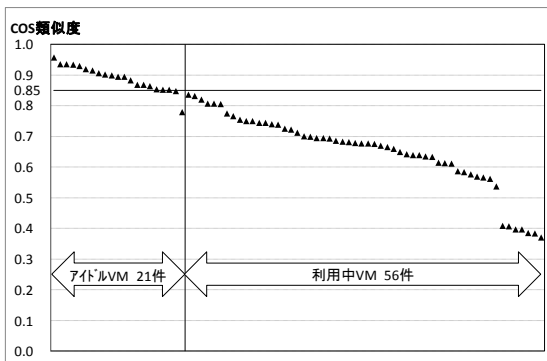


図3. 入出力画像の cos 類似度の分布 (2015年7月1日～7日のVMの画像)

cos 類似度の最大値が 0.843、アイドル VM だと識別した画像 21 件では cos 類似度の最小値が 0.790 であった。『選択的 VM 片寄せ方式』では、利用中 VM が誤ってアイドル VM に分類される「誤検出」は大きなリスクとなる。そのため少なくとも7月1日～7日のデータで「誤検出」が発生しないことを基準とし、cos 類似度が 0.85 以上のものをアイドル VM に分類することにした。一方、アイドル VM の recall 率は、機械学習モデルによる分類で 90%、従来閾値による分類で 81%であり、比較的高い比率でアイドル VM を検出できた。

3.4 環境変化が分類精度に与える影響

環境変化により、アイドル VM の「誤検出」がどの程度悪化するか、機械学習モデルによる分類と従来の閾値による分類を比較した(表2)。

2016年3月にセキュリティ対策ソフトの変更という大きな環境変化が発生している。そのため、2016年4月1日～7日の期間について、運用当初の閾値では「誤検出」の割合が極端に悪化し、利用中 VM の 51%がアイドル VM に分類されてしまった。

一方、本稿の機械学習モデルによる分類では、再学習も分類方法の変更もせずに「誤検出」を 6%に抑えることができる。つまり、機械学習モデルを適切に活用することで環境変化に対して一定の耐性を持たせられることを示せた。

表2. アイドル VM 誤検出数と誤検出率 (*)

対象期間	【環境変化前】	【環境変化後】
	2015年7月1日～7日 利用中VMの数は56	2016年4月1日～7日 利用中VMの数は440
機械学習モデルでの誤検出数と誤検出率	0件(0%)	26件(6%)
従来閾値での誤検出数と誤検出率	3件(5%)	223件(51%)

(*)利用中VMのうち、誤ってアイドルVMに分類した件数と割合(%)
件数は延VM数を表し、1つのVMが7日間稼働した場合は7件

4. おわりに

本稿ではソフトウェア開発用プライベートクラウド環境において、メモリ資源を無駄に占有しているアイドル VM を識別する際、機械学習モデルを活用することで、熟練者による閾値の見直しを不要とする、環境変化に耐性のある分類器を構築できる可能性を示した。

引き続き、機械学習モデルの精度向上に取り組み、実運用システムに適用する予定である。

参考文献

[1]住田宏己, 吉本安男, “ソフトウェア開発用プライベートクラウドにおける資源効率の改善事例”, 情報処理学会第79回全国大会講演論文集, 2A-02