

クラウド環境でのアイドルVM 識別のための機械学習適用事例

○住田 宏己† 吉本 安男†

富士通(株)†

1. はじめに

我々は仮想化されたクラウド環境の運用現場において資源の無駄遣いを削減する方法を研究している。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。クラウド環境を運用し始めた当初は、多くのVMが数日間、誰にも使われないまま実メモリを占有していた。

そこでメモリ資源の無駄な占有を減らすことを目的に、一定期間使われていないVM(アイドルVMと呼ぶ)を識別して資源を有効に活用する『選択的VM片寄せ方式』[1]を導入した。ただしアイドルVMを識別するためには熟練者の技術が必要であり、そのことが『選択的VM片寄せ方式』を多くの環境に導入する際の足かせとなっていた。本稿では熟練者が居なくてもアイドルVMを識別できるようにするために機械学習の手法を適用した事例を報告する。

2. 選択的VM片寄せ方式の概要と課題

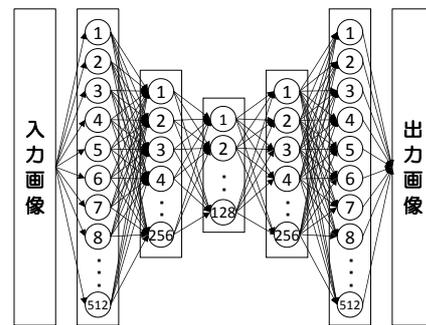
対象のクラウドシステムではソフトウェア開発チームの事情に応じてVMを自由に配備でき運用時間も自由に設定できる。例えば月曜日の早朝に起動して金曜日の深夜に停止させるような運用を可能にしている。利用者には便利な反面、週の初めに起動された後、数日の間、誰にも使われないようなVMもかなりの割合で見つかる。『選択的VM片寄せ方式』はアイドル状態が長期間続いているVMを業務用サーバから追い出すことで実メモリを有効に活用することを狙ったものである。

従来はVMの資源使用量を閾値と比較することでアイドルVMを識別していた。閾値はシステム環境の経年変化に伴い適宜見直す必要がある。しかし閾値の見直しの際には熟練者が稼働実績データを分析しなければならず、限られた熟練者に作業負荷が集中することが『選択的VM片寄せ方式』を導入する際の足かせとなっていた。

3. 機械学習によるアイドルVM識別

熟練者によるデータ分析作業を不要にするた

めに、アイドルVM識別のための閾値を設計する方式に替えてアイドルVMの特徴を学習する識別器の構築に取り組んだ[2]。学習モデルにはDeep Autoencoderを採用した(図1)。各VMの1日の資源使用量(CPU使用率とデータ転送量)の推移を表したグラフ画像を学習させることとした。



・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1. アイドルVM識別器の構成

アイドル状態のVMのデータだけを教師データとして利用し、アイドルVMの場合に入力画像と出力画像が一致するように学習させた。数個のVMを敢えてアイドル状態で放置しておくことで教師データを自動的に収集できる。収集したデータは全てアイドルVMのデータであるからラベル付け作業は不要である。ディープラーニングでは教師データを準備するコストが膨大になるが本稿のやり方であれば少ないコストで済む。アイドルVMを識別するには、各VMの資源使用量の推移を表すグラフ画像を学習済みモデルに入力し、学習済みモデルが出力する画像と似ているか否かをcos類似度で評価する方式とした。

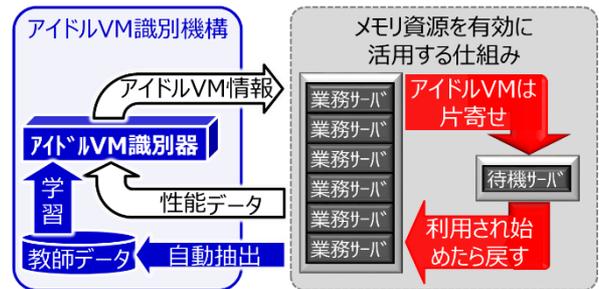


図2. 機械学習を適用した選択的VM片寄せ方式

Example of Using Machine Learning to Detect Idle VMs in a Cloud Environment

†Hiroki Sumida, Yasuo Yoshimoto
 Fujitsu Ltd.

2018年10月より実際のクラウド環境とアイドルVM識別器を連結して運用している(図2)。

4. 実環境への適用結果

本システムを継続して安定稼働させるにはアイドルVM識別精度を高いレベルに保つ必要があるが、システム環境の経年変化に伴い識別精度の低下が懸念される。そこで2018年10月初旬の稼働実績データに対するアイドルVM識別精度を二つのモデルで比較した。一つは2017年9月の一ヶ月間のデータで学習したモデル、もう一つは2018年9月の一ヶ月間のデータで学習したモデルである(図3.a、図3.b)。

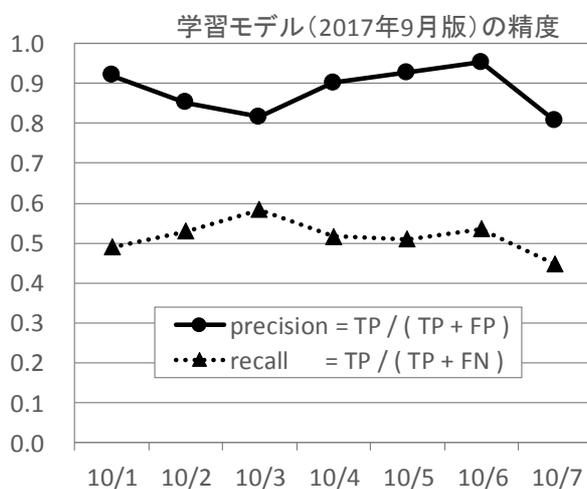


図3.a 学習済みモデル(2017年9月版)による2018年10月のアイドルVMの識別精度

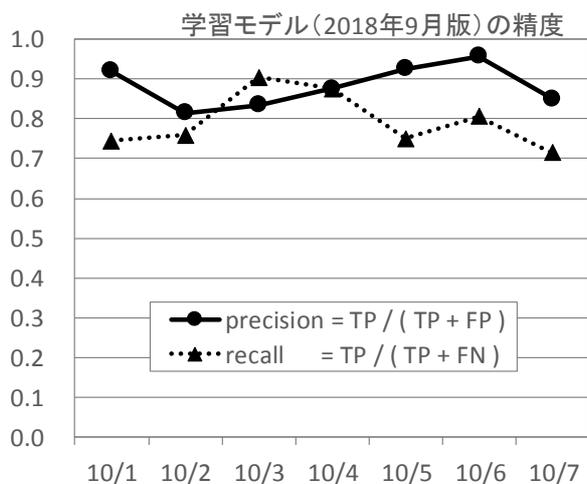


図3.b 学習済みモデル(2018年9月版)による2018年10月のアイドルVMの識別精度

アイドルVMの予測精度(precision)は図3.aと図3.bとで大きな違いは無いが、アイドルVMの再現率(recall)は図3.aの方が明らかに低い。2017年9月のデータで学習した識別器は1年後には精度が低下したものと考えている。

アイドルVMの予測精度が低下した理由を確認するため、2017年7月の一ヶ月間の教師データで学習したモデルを使って、8月以降の数ヶ月分についてアイドルVMのデータを評価した。意図的にアイドル状態で放置してある数個のVMについて、学習済みモデルへの入力画像と出力画像のcos類似度の各月ごとの平均値を表1に示す。cos類似度の平均値は月を追うごとに低下しており、数ヶ月程度の時間経過であってもアイドル状態のVMの振舞が変わり資源使用量の推移を表すグラフ画像の特徴が変化しているものと推測される。

アイドル状態のVMの振舞が変化したのであれば、最新の教師データを使って改めて学習し直すことで当初の精度に戻せる可能性がある。

表1. アイドルVMのcos類似度

	2017/8	2017/9	2017/10	2017/11
cos類似度(平均値)	0.94	0.93	0.90	0.91

5. 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のメモリ資源の有効活用を図る目的で、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した事例を紹介した。加えて、クラウド環境の経年変化に起因してアイドルVM識別精度が低下することに備えて自律的に学習し直すことで、環境変化に耐性のあるアイドルVM識別システムを構築できる可能性を示した。

引き続き、自律的に学習し直しを行うシステムを構築し効果を検証していく予定である。

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
- [2]住田宏己, 吉本安男, “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”, 情報処理学会第80回全国大会講演論文集, 2A-03