

大規模データセンターにおける 機械学習を用いた運用ノウハウ共有に関する研究

1180382 森尾勝弘 【 コミュニケーション&コラボレーション研究室 】

1 はじめに

近年、データセンターは物理環境と仮想環境の混在により大規模複雑化され運用管理は熟練管理者一人では困難となり、担当を分け複数の管理者で運用を行なっている。初級管理者は、障害発生時、欲しいノウハウ情報をキーワード検索を用いて取得し対応していたが、適切なキーワードで検索を行わなければ不適切なノウハウ情報が出力されるため、試行錯誤しなくてはならなかった。西野らの研究では、各ノウハウ情報に熟練管理者が類似度推定範囲を指定することにより、ノウハウ情報に対応するシステムと現在運用しているシステムの類似度が算出され、管理者に適切なノウハウ情報を提示することができた。[1] しかしながら、各ノウハウ情報に類似度推定範囲を指定することは非常に手間がかかることになる。

本研究では、過去に起こった障害に関するノウハウ情報を機械学習させ共有することにより、熟練管理者の手間を省くこと、さらに未知の障害が発生した際にも障害のきっかけとなる箇所を推定し、迅速に対応ができることを目的とする。

2 運用ノウハウ共有

森らの研究では、サーバを構成する部品をオブジェクトという単位で扱い、リンクに依存度という値を持つ有向グラフで示すことにより依存関係を表現した。[2] 本稿で考察するノウハウ情報は、この依存関係を用い、障害時復旧に当たった管理者により以下の3つの項目にまとめてそれぞれ設定したIDを用い蓄積する。

- 障害原因箇所
障害の原因として障害原因解析によって提示された箇所。ID:< 障害原因オブジェクト ID>
- 障害原因操作
障害復旧時に運用管理ツールの操作履歴を確認し、障害を発生させるきっかけであると判断された操作。ID:< ツール ID> < ツール別操作 ID>
- 障害理由箇所
実際に障害を引き起こしたきっかけとなった箇所。ID:< 障害理由オブジェクト ID>

3 機械学習を用いた運用ノウハウ共有

本稿は、ノウハウ情報を機械学習する手段としてベイズ推定を用いた。与えるデータの観測事象、推定したい事柄の原因事象を以下の通りに設定した。

- 観測事象 X

$X = (< \text{障害原因オブジェクト ID}>, < \text{ツール ID}> < \text{ツール別操作 ID}>)$

- 原因事象 Y

$Y = (< \text{障害理由オブジェクト ID}>)$

4 実験環境

本研究で想定するシステム環境について記述する。

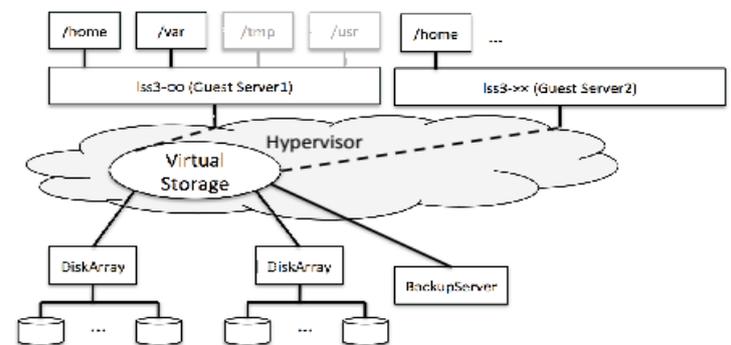


図1 システム環境 [1]

図1のシステムは実際のある大学の学内システムの運用仕様に基づいて作成した例である。

5 まとめ

本稿では、大規模データセンター運用ノウハウ情報を機械学習し共有することにより未知の障害に対しても障害の理由となる箇所を推定することができる。

将来、一層データセンターは大規模複雑化され、障害が発生し運用管理はより困難なものとなる。その際、本研究による機械学習を用いたノウハウ共有は有益なものであると考えられる。

参考文献

- [1] 西野博之, "大規模データセンターにおける運用ノウハウ共有による障害再発防止方式", 北陸先端化学技術大学院大学, 2014.
- [2] 森一, "サーバの依存関係を考慮したシステム構成管理の支援法", 情報処理学会論文誌, 46(4), 940-948(2005-04-15).