

3H-03 重回帰分析によるシステム監視項目の選定手法の評価

権藤 夏男, 小山 貴夫, 吉谷 文徳, 松田 栄之
(株)NTT データ 技術開発本部 情報科学研究所

1. はじめに

近年, Web システムを中心とした新しいビジネス形態が急速に広まって来ている. また, これらのシステムを集約的に運用するために, インターネットデータセンタ(IDC)のような大規模システムに対する注目が高まっている. 大規模システムにおいては, 様々な構成かつ大量の機器・ソフトウェアを統合的に運用管理し, 安定したサービスを提供することが従来以上に求められている. しかしながら, 従来のシステムの運用管理では, いわゆるキーマンの経験と試行錯誤によりシステムの監視項目が決定されてきた. このため, 管理可能なシステム規模が限定され, かつ個人のスキルの差による品質差が生じてしまう. この問題に対処するため, 運用管理の省力化・自動化[1]が今後のコンピュータシステムの大きな問題の一つとなる.

本稿では, サービス品質を維持する運用管理システムを一意的かつ短期に実現するため, 管理対象システムの実監視データを統計手法によって解析し, 着目するパラメータと相関の高い監視項目を定量的・自動的に設定する手法を提案する.

2. 監視項目の選定手法

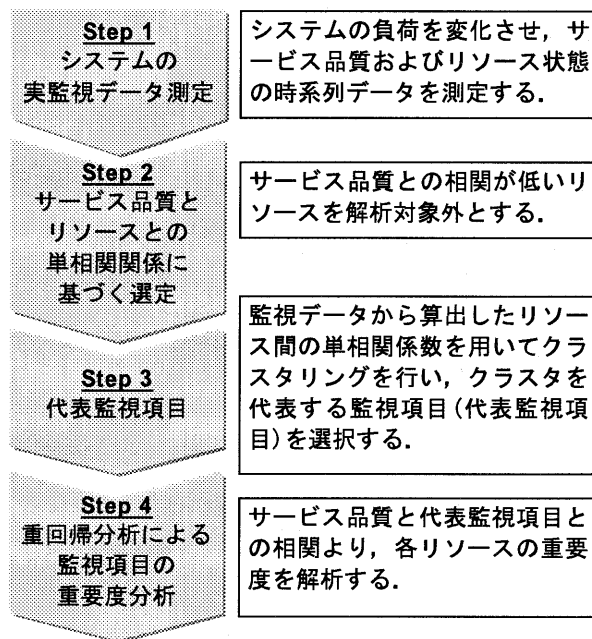
一般にシステムの運用監視対象項目として, CPU, メモリ, ディスクといったプリミティブな情報からサーバ, ネットワーク, ストレージなどのプリミティブな構成要素を集約したリソースなどがある. これら複数のリソースの状態は, 応答時間や処理件数といったサービス品質と相関があり, 相関度合いはリソースごとに異なる.

本手法では, サービス品質とリソース状態との関係を解析し, システムに適した監視項目を選定するために, 多変量解析の一手法である重回帰分析を用いる.

解析対象となるそれぞれのリソースには, 相互に相関の高いものと低いものとがある. 重回帰分析では, 説明変数同士に相関の高いものを排除する必要があるので, 説明変数であるリソース状態

間で相関の高いものをクラスタリングし, 各クラスタを代表する監視項目を用いて, 重回帰分析を適用する.

監視項目抽出の処理手順を以下に示す.



また, 本手法の Step3 における代表監視項目の決定方法の詳細を以下に示す.

- (1) 監視項目間の相関算出
全監視項目間の単相関係数を算出する.
- (2) 監視項目のクラスタリング
監視項目間の単相関係数に閾値を設定し, 相互に相関の高い監視項目のクラスタリングを行う. (今回は, 一般的に相関が高いと言われている 0.8 を閾値とした)
- (3) クラスタの代表監視項目の選択
代表監視項目の抽出基準を, クラスタの中心からの距離が最短であると定義する. 具体的には, 同一クラスタに属する他の監視項目との相関係数の総和が最大である監視項目を代表監視項目とする.

3. 検証

商用 WWW サイトのモデルである TPC-W を WWW(兼 AP)サーバ, DB サーバ各 1 台で構成し, 本手法の検証を行った. 一般に WWW サイトでは, クライアントが不特定多数であるため, ここ

A method for selecting metrics in system monitoring using multiple regression analysis

N. Gondo, T. Koyama, F. Yoshitani, S. Matsuda
Laboratory of Information Technology, R&D Headquarters,
NTT Data Corporation
Kayaba-cho Tower Bldg., 1-21-2, Shinkawa, Chuo-ku, Tokyo,
Japan

では監視対象をサーバ及び WWW サイト内ネットワークとした。

リソース状態として、性能管理ツールで監視可能な項目のうち、マシン単位の監視項目(サーバの CPU やメモリの使用率等)である 106 項目を、サービス品質として WWW サーバの応答時間を、表 1 に示す条件で測定した。なお、測定後、応答時間については 5 分平均を求めた。

表 1：監視データ測定条件

	サービス品質	リソース状態
測定項目	応答時間	サーバ、ネットワークリソース
測定方法	ダミー・トランザクションを実行	性能監視ツール
測定間隔	20 秒	5 分
負荷状態	WWW クライアントを 20 分毎に 1 つずつ、40 クライアントまで増加させる	

監視項目の時系列データ(168 サンプル)を用いて、サービス品質との相関が低いリソースを解析対象外とした(手順 Step 2)結果、67 項目が抽出された。次に、クラスタリング及び代表監視項目の選定(手順 Step 3)を行い、67 個の監視項目から 10 個の代表監視項目を選定した(表 2)。

表 2：監視項目数の推移

全測定監視項目	106
サービス品質とリソースとの単相関関係に基づく選定	67
クラスタリング及び代表監視項目の選定	10

重回帰分析による監視項目の重要度分析(手順 Step 4)を行った結果を表 3 に示す。

表 3：重回帰分析結果

代表監視項目	標準偏回帰係数
WWW メモリ使用率	0.668
DB システム CPU 使用率	0.173
DB 空きメモリ率	-0.127
DB Sleep プロセスの数	0.097
WWW 物理 I/O 完了待ち率	0.071
DB 仮想メモリライト回数/秒	0.047
DB 仮想メモリリード回数/秒	-0.045
DB アクティブなプロセス数	-0.023
WWW ディスクキャッシュヒット率	0.021
WWW ディスク競合プロセス数	0.020

標準偏回帰係数(応答時間への影響の大きさを示す係数)から、応答時間に最も影響を与えるリソースは、WWW サーバのメモリ使用率であることが分かる。これらの時系列データ(図 1)からも、測定期間内の WWW サーバのメモリ使用率は応答時間と高い相関を示していることが確認できる。また、その他の代表監視項目については、一般的な監視項目である CPU やメモリに加えて、キャッシュヒット率やプロセス数などシステム固有と考えられる監視項目を選定している。以上の結果より、本手法が有効であるとの見込みを得た。

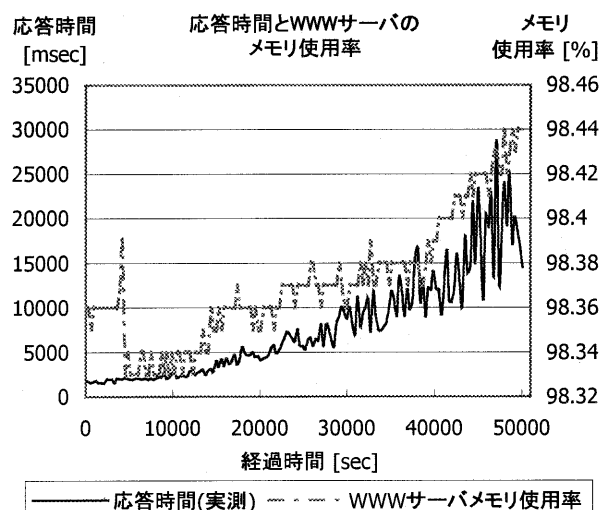


図 1：応答時間と WWW サーバのメモリ使用率

4. まとめ

人手に頼っているシステムの監視項目の設定について、重回帰分析を用いて定量的かつ自動的に設定する手法を提案した。本報告では、応答時間をサービス品質とした実データをもとに検証し、相関の高い監視項目選定が可能となる事を確認した。

今後は、複数の負荷パターンを用いて、多様なシステム状態に対応できることを検証していく予定である。また、他のサービス項目への適用可能性の確認や、システム障害時などに相関が高いパラメータの抽出方法に関しても取り組んでいきたいと考える。

参考文献

[1] “WWW サービスの品質劣化要因分析手法の検討”，堀之内剛史，野上慎也，2000 年電子情報通信学会総合大会 B-7-124, pp.217