

遺伝的アルゴリズムを利用した情報システム性能予測の 精度改善手法

西岡大祐[†]

(nisioka@crl.hitachi.co.jp)

情報システム設計において精度の高い性能見積りを実現し、SI 業務の生産性を向上させるシステム性能予測手法を開発した。本手法では、性能シミュレーション技術の活用を前提として、性能予測実施対象のシステムのシミュレーションモデルに関して、予測精度に影響するパラメータ調整を自動化する。従来は人手でこの調整作業を実施していたため、工数がかかり実用的でなかった。この問題を最適探索手法の一つである遺伝的アルゴリズムの採用により解決した。代表的な基幹アプリケーションの性能予測において本手法を適用して有効性を確認したところ、全パラメータ候補の0.3%の探索で、予測誤差を手法適用前の18%から2%へ縮小できた。

The Method of improving accuracy of IT system simulation with Genetic Algorithm

DAISUKE NISHIOKA[†]

We have developed a new parameter tuning method for simulation model in order to improve accuracy of simulation. Our method optimizes the parameter set which affect accuracy of simulation using Genetic Algorithm (GA). We have applied this method to famous Enterprise Resource Planning software benchmark simulation model. Our experiments show that the accuracy of simulation improved from 18% to 2%.

1. はじめに

近年の情報システムの構築や運用は、オープン化による技術変化への迅速な対応、低価格化の実現、システム構築時間短縮など、多くの課題を抱えている。これらの課題を解決する上で、SI (System Integration) 業務の役割が重要となる。SI 業務では、様々なプラットフォームを用いて、三層構成 Web 業務システムなどの複雑なシステムを構築する。また、同システムの性能、セキュリティ、可用性について顧客要求に見合った最適設計を行う必要がある。これらの要件が満たされない場合、様々な形で顧客業務に支障をきたし、顧客にビジネス上の大きな損失をもたらす可能性がある。

上記三つの要件のうち、性能については、システム構築前の段階でその値を見積る必要がある点に難しさがある。特に Web 業務システムではサーバ台数規模や負荷分散方法が全体性能に大きく影響するため、この性能を正確に予測できれば、顧客に対する適切なシステム提案によって信頼を獲得でき、また、システム構築後の性能問題を回避できるといったメリットがある。しかし、Web 業務システムの性能を、システム構成要素の性能単価の積み上げや、過去の事例からの類推といった経験則だけで予測することは困難である。そのため、解析的な手法や性能シミュレーションによる性能予測技術が期待されている。性能シミュレーションによる性能予測では、顧客に提供予定のシステムをシミュレ

[†] (株)日立製作所中央研究所
Central Research Laboratory, Hitachi, Ltd.

ーション実行環境向けにモデル化（シミュレーションモデル作成）してイベントシミュレーションを行い、結果として得られるシステム応答時間やハードウェア使用率などの統計量から、当該システムが処理可能な最大負荷と性能上のボトルネックを特定する。

上記シミュレーションモデルは、一度作成すると他の性能見積り案件でも再利用が可能である。ただし、同モデルを構成するハードウェア、ソフトウェアの各モデルについては、スペック変更やバージョンアップ等に対応して、各種パラメータの値を調整する必要がある。このパラメータ値の調整が適切でないと、性能見積り精度が悪化する。パラメータの組み合わせの数は数十から多いときは数万通りにもなるため、顧客への短期間での応答を必要とする SI 業務の現場にとって、網羅的なパラメータ調整を行うことは現実的ではない。本研究は、この課題を解決し、SI 業務の生産性を向上させることを目的とする。

2. シミュレーションによる性能見積り方法とその技術課題

シミュレーションを用いた性能予測では、スケリングや積み上げ計算では難しいシステム全体の性能が予測可能であり、実測環境を用意できないような大規模システムでも、システムの応答時間や、各サーバの利用率などについて検討が可能である。

本性能見積りは、次の三つのステップで実現される。

- (i)対象システムをシミュレーション実行環境向けにモデル化（シミュレーションモデル作成）
- (ii)作成したシミュレーションモデルを用いて、繰り返しシミュレーションを実施
- (iii)シミュレーション結果から対象システムの性能曲線を導出

図 1 に例示するとおり、性能曲線とは横軸に負荷、縦軸に応答時間をプロットしたグラフである。典型的な性能曲線では、システムに対する負荷が臨界点に達すると急激に応答時間が悪化するポイントが示される。システム構築者は、この悪化ポイントをもとにシステムの性能限界を予想し、顧客要求を満たすシステム構成の検討や製品選定を行う。そのため、この悪化ポイントを正確に予測することが重要となる。

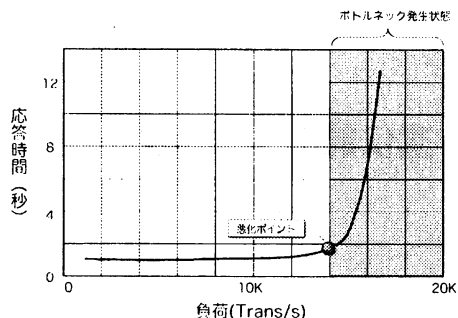


図 1 性能曲線の例

上記の悪化ポイントを正確に予測するためには、シミュレーションで用いるシミュレーションモデルの正確さが鍵となる。シミュレーションモデルには、性能予測対象のシステムを構成するハードウェア、ソフトウェアの特性を表現する各種パラメータが設定されている。シミュレーションモデルの正確さを向上させるためには、このパラメータの値を適切に決定する必要がある。以降、このパラメータ値を適切に決定するための作業をパラメータ調整と呼ぶ。

従来手法によるパラメータ調整は、次の手順により実現されていた。

- (1)各パラメータがとりうる値の列挙
- (2)各パラメータがとりうる値を個別に組み合わせたシミュレーションモデルの作成
- (3) (2)のシミュレーションモデルを用いたシミュレーションの実行
- (4) (3)の実行結果（性能曲線）と試験用パイロットシステムを用いた実測結果（性能曲線）の比較
- (5) (2)～(4)をすべてのパラメータ値の組み合わせに対して実施し、(4)の比較の結果、最も性能曲線が一致するモデルを性能見積り用シミュレーションモデルとして選定

なお、試験用パイロットシステムとは、パラメータ調整結果を検証するために、必要最低限のシステムリソースによって構成されたシステムであり、性能見積り対象システムのサブセットである。

上記(1)～(5)の手順に関わる工数は、パラメータ数とパラメータが取りうる値の組み合わせの数によって決まる。例えば、5 種類のパラメータがそれぞれ 5 通りの値を取りうる時、 $5^5=3125$ 通りのシミュレーションモデルの中から精度の高いモデルを探索する必要がある。こ

のため、組み合わせの数が多いときは、人手では最適な組み合わせを探索しきれないという問題がある。このような多数のパラメータの組み合わせの中から、現実的な時間の範囲で効率よく最適解を導き出す手法を開発するのが本研究の課題である。

3. システム性能予測の精度改善手法

3.1 パラメータ探索方法

多数のパラメータの組み合わせが持つ探索空間を探索する代表的な方法としては、全探索、ランダムサーチ、最急勾配法、遺伝的アルゴリズム(Genetic Algorithm)が挙げられる^{[1][2]}。

全探索はすべての組を検証する最も単純な方法である。確実に最適解が得られるが、パラメータの組み合わせの数により指数的に探索回数が増えてしまい、実用的でない。ランダムサーチは、解の候補を無作為に選択して条件に適合するかを評価する。この方法は、選択が完全にランダムなため、解を求める効率が悪い。最急勾配法は山登り法とも呼ばれ、はじめに無作為に選択した解の集合(初期集団)を用意し、それらのパラメータを評価値がよくなる方向へ変化させ、変化できなくなるまで繰り返す方法である。この方法では、全探索空間の中で部分的に評価値が高い所があると、局所安定と呼ばれる状態に陥り、最適解を発見せずに終了してしまう可能性がある。

遺伝的アルゴリズムは、ランダムサーチと最急勾配法の欠点を解消した方法であり、ランダムサーチに比べて効率がよく、最急勾配法のように局所安定に陥ることがないため、遺伝的アルゴリズムによるパラメータ探索を実施することにした。

3.2 遺伝的アルゴリズムのシステム性能予測への適用

遺伝的アルゴリズムを用いてシミュレーションモデルのパラメータ調整を自動的に行うプログラムを、Microsoft Excel 2000 上に作成した。本プログラムは、2章の手順(1)~(5)を遺伝的アルゴリズムに従って処理する。また、本プログラムは、手順(3)の自動化を目的として、報告者らが米国 HyPerformix^[3]社の性能シミュレータの上に開発した性能予測システム WISE (What-If Scenario Evaluator) ツール^[4]と連動する。

性能予測対象としては、独 SAP AG 社の

ERP(Enterprise Resource Planning)ソフトウェアである SAP R/3 を採用し、シミュレーションモデルとしては、WISE ツール向けに用意されている標準的なパラメータ(初期パラメータ)を持つモデル(初期モデル)を利用した。

上記プログラムでは、シミュレーションモデルに設定されている各種パラメータのうち、性能に影響がある6種類のパラメータに対して遺伝的アルゴリズムを適用した。このパラメータを表1に示す。次に表2に示すとおり、各パラメータについて、初期パラメータを含めて各々が5通りの値を取るよう設定した。ここで、各パラメータに対して表中の任意の値を設定した組み合わせたシミュレーションモデルを生成する。

さらに、遺伝的アルゴリズムの各種条件を表3に示す形で設定した。また、シミュレーションモデルのパラメータ設定の評価関数としては、シミュレーションによる性能曲線と試験用パイロットシステムで実測した性能曲線の一致度を採用した。その計算式を表3に示す。なお、1個体に対するシミュレーションを用いた性能曲線の導出においては、シミュレーション条件として与える負荷の値をいくつか変化させ、各々シミュレーションを実施する。今回、この負荷の値については、5種類とした。

表 1 遺伝子として設定したパラメータ

パラメータ(遺伝子)	説明
WISEServerSpec	実測サーバのSPECint2000の値
WISECPUBackground	CPUのバックグラウンド使用率
WISEAPITSServer_SD	SAP R/3のITSServerプロセスのCPU使用量 (SPECint2000換算)
WISEAPMessageHandler_SD	SAP R/3のMessageHandlerプロセスのCPU使用量 (SPECint2000換算)
WISEAPDialogWorkers_SD	SAP R/3のDialogWorkersプロセスのCPU使用量 (SPECint2000換算)
WISEAPDBServer_SD	SAP R/3のDBServerプロセスのCPU使用量 (SPECint2000換算)

表 2 パラメータの値設定

パラメータ(遺伝子)	値1	値2	値3	値4	値5
WISEServerSpec	248	279	319	341	372
WISECPUBackground	0	5	10	15	20
WISEAPITSServer_SD	3.75	4.22	4.69	5.16	5.63
WISEAPMessageHandler_SD	0.16	0.18	0.2	0.22	0.24
WISEAPDialogWorkers_SD	3.15	3.54	3.94	4.33	4.73
WISEAPDBServer_SD	0.50	0.56	0.62	0.69	0.75

□ 初期パラメータ

表 3 遺伝的アルゴリズム条件設定

実行条件	値
1世代の個体数	16
交叉確率	0.5
突然変異確率	0.2
評価関数	$\sum \left(\alpha \left(\frac{U_i - T_i}{T_i} \right)^2 + (1-\alpha) \left(\frac{U_i - U_i}{U_i} \right)^2 \right)$ <small>T: シミュレーション平均の悪化ポイント U: 実測平均の悪化ポイント α: シミュレーション平均の悪化ポイントと実測平均の悪化ポイントの比率を決める係数 (α: 0.5)</small>
終了条件	同じ解が2回連続した段階で打ち切り

4. 提案手法適用効果の評価

4.1 システム性能予測の精度改善

図 2に、以下の各ケースでの性能曲線を示す。

- (a) 試験用パイロットシステムでの実測結果
- (b) 初期モデルによる性能予測結果
- (c) 提案手法を適用した性能予測結果

提案手法を用いた結果、性能予測の際に重視される悪化ポイントのずれが初期モデルの 31% から 9%へ改善された。

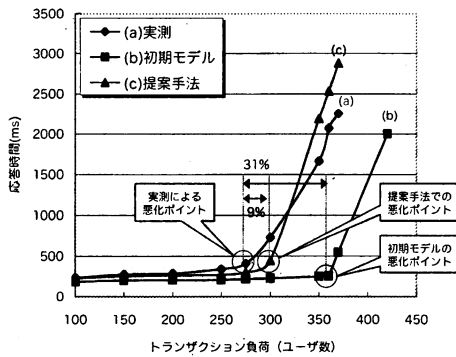


図 2 性能曲線の比較

ところで、SAP R/3では、性能見積りの指標として、標準ベンチマーク結果を基に SAPS 値という性能指標を算出する¹⁵⁾。これは同社の QuickSizer という見積りツールで用いられており、SAPS 値の定義は次の通りである。

「応答時間 2 秒以内のときの最大ユーザ数を 5 倍にした値」

そこで、(a)、(b)、(c)の各グラフから、応答時間が 2 秒となるユーザ数を読みとり、SAPS 値に換算して比較した。比較結果を図 3に示す。SAPS 値換算で、誤差は約 18%から 2%へ改善された。

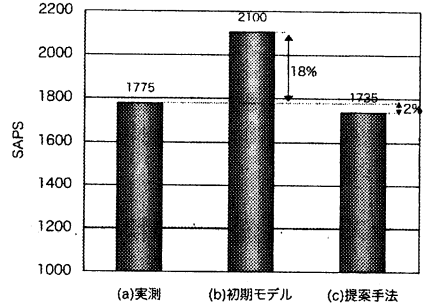


図 3 SAPS 値による精度の比較

4.2 システム性能予測の工数改善

提案手法を用いたシミュレーションモデルのパラメータ調整は、人手を介さずに自動探索を行う。4.1節で述べた評価のケースでは、遺伝的アルゴリズムによるパラメータ調整は、第 3 世代をもって終了条件を満たして終了した。この工数と全探索手法による工数を比較した結果を表 4に示す。遺伝的アルゴリズムにより、全探索手法の 0.3%の工数で、精度を向上させるパラメータ調整が可能となった。

表 4 パラメータ調整に要する工数

調整方法	設定	シミュレーション回数	所要時間*
遺伝的アルゴリズム	3世代、16個体 5回シミュレーション/個体	240回 (3x16x5)	約2日
全探索	パラメータ6種類×5値 5回シミュレーション/個体	78125回 (5 ⁶ ×5)	約1.5年

*1シミュレーション=10分で計算

5. 本研究の水準比較

本報告の研究と直接的に関係する性能予測手法としては、前述の QuickSizer がある。これは、SAPS 値を基にしたサーバ単体を対象とした積み上げ型の性能予測手法である。この方法の場合、SAP R/3 が稼働するサーバ単体についての知見は得られるが、システム全体の性能や応答時間などの検証ができない。それに対して本報告による性能見積りでは、システムの応答時間、各サーバの必要性能の見積り、悪化ポイントの特定など、多角的な分析が可能である。本報告の成果を活用することにより、QuickSizer よりも、詳細な事前見積りが可能となる。

6. おわりに

6.1 結論

情報システム設計において重要な高精度の性能見積りを行い、SI 業務の生産性を向上させるシステム性能予測手法を開発した。本手法では、ソフトウェアシミュレーション技術の活用を前提として、性能予測実施対象のシステムのシミュレーションモデルに関して、最適解探索手法の一つである遺伝的アルゴリズムを活用して予測精度に影響するパラメータ調整を自動化する。代表的な ERP ソフトウェアである SAP R/3 を題材として、本手法の有効性を検証した結果、以下の結論を得た。

- ・本手法により、性能予測精度を SAPS 値換算で 16 ポイント改善した。
- ・性能予測で重要な性能悪化ポイントの予測精度が 22 ポイント改善した。
- ・全探索空間の 0.3% の探索工数で予測精度が向上してきた。

6.2 今後の課題

本報告では、遺伝的アルゴリズムを SAP R/3 シミュレーションモデルのパラメータ調整に用いたが、今後、他のアプリケーションへも適用範囲を広げたい。また、遺伝的アルゴリズムに関しては、評価関数の設定や、交叉、突然変異する確率の設定、初期集団数の設定について方法論が確立していないため、現時点では試行錯誤しなければならない部分がある。今後、設定方法について検討する。

7. 参考文献

- [1] <http://kyu.pobox.ne.jp/softcomputing/ga/ga2.html> (探索法としての GA)
- [2] 平野：遺伝的アルゴリズムと遺伝的プログラミング—オブジェクト指向フレームによる構成と応用—，パーソナルメディア，2000 年，ISBN4-89362-173-4
- [3] <http://www.hyperformix.com>
- [4] 西岡ほか：システム性能シミュレーションのリモート実行環境の開発、情処学会コンピュータシステムシンポジウム論文集，Vol.2001，No.16，pp.129-136
- [5] http://www.sap.co.jp/partner/techno_partner/sizing/ (サイジングについて)

商標その他

HyPerformix は、米国 HyPerformix Inc. の登録商標です。

SAP, R/3, QuickSizer は、独 SAP AG 社の登録商標です。

そのほか一般に用いられている製品名については、各社の登録商標です。