

システム性能評価と システム評価

木下俊之 東京工科大学

システム評価とは

当研究会は、「システム評価研究会」であり、「システム性能評価研究会」ではありません。ITの分野で「評価」というと性能評価を連想しますが、当研究会は性能だけでなくいろいろなファクタの評価を扱っています。表-1に、当研究会が全国大会で使っているキーワードを示しました。性能に関するもののほかに信頼性評価、運用性評価などがあり、またそれらとは異なる軸として評価技法(シミュレーション、実測、解析的手法など)があります。

しかしそうは言っても、情報処理学会の研究会ということで、性能評価に関する発表が多いです。性能評価は、その名の通り性能を評価するもので、開発段階、テスト段階、実用段階のいずれの局面でも必要になります。開発段階での性能評価は、開発中の製品等が所要の性能を持つかを確認することです。もし性能が不足と分かれば、性能向上策を追加して製品の完成前に性能不足を補うことができます。ここでは、ハードウェアとソフトウェア、およびその両者の相互作用による性能を確認します。

テスト段階での性能評価は、いわゆる性能バグをたたくことにあります。性能バグとは設計段階ではこうなるはずと思って作り込んだものが誤って性能低下を引き起こしているケースで、テスト段階での性能評価によりこの誤った作り込みを見つけ出すためのものです。

製品やシステムが実稼働を始めると、今度はシステムが実際のフィールドで所要の性能を出しているかを確認することになります。これが実稼働段階での性能評価です。このように性能評価は、設計開発、実稼働のさまざまな局面で必要になることが分かります。

信頼性に関しても、同様に開発段階、テスト段階、実稼働段階のそれぞれで評価が必要です。性能と違い信頼

- | |
|---------------|
| 5. システム評価 |
| 01 性能評価 |
| 02 信頼性評価 |
| 03 運用評価 |
| 04 予測評価 |
| 05 モデリング |
| 06 評価技法 |
| 07 評価シミュレーション |
| 08 実測評価 |
| 09 待ち行列による評価 |
| 99 その他 |

表-1 システム評価研究会のキーワード

性は予測評価することが難しいため、開発段階やテスト段階といった信頼性評価は、過去の実績に基づいた経験的な評価が多いようです。一方、実稼働段階の信頼性評価は、運用評価と密接に関連します。運用評価は、実稼働しているシステムを分析し、その性能や信頼性の特性を見つけてそれをシステムの運用に活かすことです。実稼働しているシステムの特性とは、たとえば性能でいえばオンラインシステムでは昼休み前やシステム終了間際、休前日の稼働率が高いといった時間的な特性や、どのアプリケーションの負荷が高いかといった業務処理特性などです。また信頼性上の特性とは、ホットスタンバイの頻度や切り替え時間、バックアップの取得方法などが挙げられます。この運用管理は、システムの実際の稼働にかかわる生の振る舞いが報告されるので、興味深い発表になることが多いです。

このように性能評価、信頼性評価、運用評価は、開発、テスト、運用の実際に基づいた評価に関する発表が行われます。

評価技法というのは、性能や信頼性などの評価ファクタについて、さまざまな種類のシステム評価を横断的かつ汎用的に応用可能な評価手法に関するものです。主な評価手法は、キーワード表にもあるシミュレーション、実測、机上・理論解析などですが、これらについて分野横断的に利用可能なものを紹介したり、その手法を使った評価の結果などが報告されています。特に実測による評価はフィールドでの生のデータが出てくるので、とても興味深い内容になります。

このようにシステム評価の研究は、主に実システムでの評価結果と評価法についての発表ですが、一部で評価技法について発表されています。

研究会の活動

システム評価研究会は、2001年に発足した比較的若い研究会です。研究会になる以前は、研究グループとして約4年間活動してきました。研究会のテーマはその名の通りシステムの評価全般で、単に性能だけでなくシ



- (1) コンピュータシステム評価……………11件
(メインフレーム、大規模サーバ、アプリケーションシステム)
(a) 仮想計算機、論理分割環境での資源管理
(b) Workload Manager 応用による資源管理
(c) 大規模サーバの負荷予測、キャパシティ管理、資源最適化
(d) ブレードサーバのスケラブル制御
(e) メモリ制御ベンチマーク
(f) バッチ処理システムのジョブ管理
- (2) ネットワークシステム評価……………8件
(ネットワーク制御、グリッドコンピューティング、P2P)
(a) ネットワーク上のトラフィック特性評価、輻輳制御
(b) ノードのルーティングプロトコル
(c) グリッドコンピューティング環境でのプロセス連携、実行時間短縮
(d) P2P 環境でのトラフィック徳性評価、ダウンロード性能特性、高速検索手法
- (3) Web システム、Java プログラム… 5件
(a) Java プログラムのシステムライブラリ、特性解析
(b) Web システムの性能限界予測、応答時間最適化
(c) シンクライアント環境での運用評価
(d) ユビキタス環境での負荷分散
- (4) その他のシステム評価……………4件
(低電力化、運用性評価、性能評価手法、災害対策)
(a) 低消費電力化手法
(b) シミュレーション手法
(c) 待ち行列による解析手法
(d) 災害対策システム

表-2 分野別の発表テーマ
(2006年6月から2007年8月までの6回, 31件)

システムの信頼性、運用性、あるいは価格性能比といった経済性やIT分野に限らない一般のシステム評価(その場合の評価指標にはいろいろなものが出てくる)も含めています。このようなシステム評価の技術は、対象とするシステムを横断的にわたる横串の技術です。したがって、1つの評価技術を他の評価対象に適用するとどうなるかといった、対象システムに共通な評価の議論があります。その一方で、各対象システムに固有の評価方法について報告されることもあります。特にIT分野でない一般の社会システムの評価の発表では、参加者が主にIT関連の専門家が多いので、そういう人たちにとっては分野違いの発表になり、とても新鮮で良い刺激を受けます。

最近1年間のシステム評価研究会の発表分野は、表-2のようになっています。この表に示すように、発表分野は多岐にわたっています。これはシステム評価がさまざまな分野で活発に行われていることを意味しています。表-3に発表者の所属を示しましたが、企業と大学・研究機関でちょうど半々になっています。企業からの発表は実稼働データに基づいた実際的な発表が多く、大学・研究機関からは多少理論的な発表になっています。両者の件数が同数ということは、理論と実際がバランスよく

(1)企業	15件
(2)大学	15件
(3)企業+大学 共著	1件

表-3 発表者の所属の内訳

- (1) 評価対象
(a) コンピュータ・ネットワークを中心としたITシステム
(b) 社会システム(交通、教育、医療、防災など)
(c) 自然環境(気象、生態、省資源など)
- (2) 評価項目
(a) 性能、信頼性などの評価項目
(b) ヒューマンファクタ、マンマシン連携
(c) 社会科学の指標
- (3) 評価方法
(a) 実測
(b) シミュレーション
(c) 統計分析
(d) ヒアリング、アンケート、世論調査

表-4 システム評価のスコープ

発表されているといえます。

研究会は1年に4回、毎年6月、8月、11月、3月に行っています。6月は首都圏の会場で、そのほかの3回は首都圏以外の地方で行っています。8月はSWoPPの1研究会として参加しています。シンポジウムや国際カンファレンスなどには参加していません。

システム評価の今後

冒頭にも書きましたが、システム評価はIT分野に限らず社会一般に行われています。当研究会では、これらの必ずしもITに限らないシステムの評価について、今後取り上げていきたいと思えます。考えられるシステム評価のスコープを表-4にまとめてみました。情報処理学会の研究会なので今後もIT関連の性能、信頼性評価が中心になると思いますが、これにできれば社会システム、自然環境の評価を加えて、コンピュータ・ネットワークシステムがこれらにどのような影響を与えるかについて考える場になればと思っています。コンピュータ・ネットワークシステムは急速に進化し、今や人間社会、自然環境に大きな影響を与えるまでになっています。今後このコンピュータ・ネットワークシステムが社会にどのようなかわりを持っていくか、システム評価を通じて明らかにしていければと思っています。

(平成20年2月22日受付)

木下俊之(正会員)

kinoshi@cs.teu.ac.jp

1977年東京大学大学院数学専門修士課程修了。同年(株)日立製作所システム開発研究所入所。2005年より東京工科大学コンピュータサイエンス学部教授。博士(理学)。ネットワークセキュリティ、オペレーティングシステム、性能評価の教育・研究に従事。