Pontifícia Universidade Católica
do Rio de Janeiro

Sand Luz Corrêa

Detecção estatística de anomalias de desempenho em sistemas baseados em middleware

Tese de Doutorado

Tese apresentada ao Programa de Pós–graduação em Informática do Departamento de Informática da PUC–Rio como requisito parcial para obtenção do título de Doutor em Informática

Orientador: Prof. Renato Fontoura de Gusmão Cerqueira

Rio de Janeiro
abril de 2011

Pontifícia Universidade Católica
do Rio de Janeiro

Sand Luz Corrêa

Detecção estatística de anomalias de desempenho em sistemas baseados em middleware

Tese apresentada ao Programa de Pós–graduação em Informática do Departamento de Informática do Centro Técnico Científico da PUC–Rio como requisito parcial para obtenção do título de Doutor em Informática. Aprovada pela Comissão Examinadora abaixo assinada.

Prof. Renato Fontoura de Gusmão Cerqueira
Orientador
Departamento de Informática — PUC–Rio

Profa. Noemi de la Rocque Rodriguez
Departamento de Informática — PUC-Rio

Prof. Ruy Luiz Milidiú
Departamento de Informática — PUC-Rio

Prof. Raimundo José de Araújo Macêdo
Departamento de Ciência da Computação — UFBA

Prof. Rodrigo Fernandes de Mello
Instituto de Ciências Matemáticas e de Computação — USP

Prof. José Eugenio Leal
Coordenador Setorial do Centro Técnico Científico — PUC–Rio

Rio de Janeiro, 1 de abril de 2011

Sand Luz Corrêa

Mestre em Ciência da Computação pela Universidade Estadual de Campinas. Bacharel em Ciência da Computação pela Universidade Federal de Goiás.

Ficha Catalográfica

Corrêa, Sand

Detecção estatística de anomalias de desempenho em sistemas baseados em middleware / Sand Luz Corrêa; orientador: Renato Fontoura de Gusmão Cerqueira. — Rio de Janeiro : PUC–Rio, Departamento de Informática, 2011.

v., 146 f: il. ; 29,7 cm

1. Tese (doutorado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática.

Inclui referências bibliográficas.

1. Informática – Tese. 2. Computação Autonômica. 3. Previsão de Problemas de Desempenho. 4. Diagnóstico de Problemas de Desempenho. 5. Técnicas de Aprendizado Estatístico. 6. Testes Estatísticos. 7. Sistemas baseados em *Middleware*. I. Cerqueira, Renato. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. III. Título.

CDD: 004

Agradecimentos

Ao meu pai e minhas irmãs, pelo amor e apoio durante toda a vida.

Ao meu companheiro, Kleber, pelo amor, carinho e pelas inúmeras discussões e críticas construtivas sobre este trabalho.

Ao Prof. Renato Cerqueira, por sua orientação e amizade.

Ao CNPq e ao Tecgraf, pelo suporte financeiro, sem o qual este trabalho não poderia ter sido realizado.

Resumo

Corrêa, Sand; Cerqueira, Renato. **Detecção estatística de anomalias de desempenho em sistemas baseados em middleware**. Rio de Janeiro, 2011. 146p. Tese de Doutorado — Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Tecnologias de *middleware* têm sido amplamente adotadas pela indústria de software para reduzir o custo do desenvolvimento de sistemas computacionais. No entanto, é difícil estimar o desempenho de aplicações baseadas em *middleware* devido a fatores como a especificidade de implementação das plataformas de *middleware* e a multiplicidade de serviços e configurações providas para diferentes cenários de implantação. Assim, o gerenciamento do desempenho de aplicações distribuídas baseadas em *middeware* pode ser uma tarefa não trivial. Computação autonômica é um novo paradigma para construir sistemas autogerenciáveis, isto é, que procuram operar com o mínimo de intervenção humana. Este trabalho investiga o uso de abordagens estatísticas para construir mecanismos autonômicos de controle do desempenho de aplicações baseadas em *middleware*. Particularmente, o tema é investigado sob três perspectivas. A primeira é pertinente à previsão de problemas de desempenho. O uso de técnicas de classificação é proposto para derivar modelos de desempenho que auxiliem o gerenciamento autonômico das aplicações. Nesse sentido, diferentes classes de modelos em aprendizado estatístico são avaliadas, tanto em cenários de aprendizado *offline* quanto *online*. A segunda perspectiva refere-se à redução da emissão de alarmes falsos, visando a construção de mecanismos robustos a falhas transientes dos classificadores. Este trabalho propõe um algoritmo que aumenta o poder de predição das técnicas de aprendizado estatístico, combinando-as com testes estatísticos para a detecção de tendência. Por fim, a terceira perspectiva é pertinente ao diagnóstico das causas de um problema de desempenho. Para isso, propõe-se também o uso de testes estatísticos. Os resultados apresentados nesta tese demonstram que abordagens estatísticas podem contribuir para a construção de ferramentas eficazes e eficientes para a caracterização do desempenho de aplicações baseadas em *middleware*. Portanto, essas abordagens podem contribuir de forma decisiva para diferentes perspectivas do problema.

Palavras–chave

Computação Autonômica; Previsão de Problemas de Desempenho; Diagnóstico de Problemas de Desempenho; Técnicas de Aprendizado Estatístico; Testes Estatísticos; Sistemas baseados em *Middleware*;

Abstract

Corrêa, Sand; Cerqueira, Renato (Advisor). **Statistical detection of performance anomalies in middleware-based systems**. Rio de Janeiro, 2011. 146p. DsC Thesis — Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Middleware technologies have been widely adopted by the software industry to reduce the cost of developing computer systems. Nonetheless, predicting the performance of middleware-based applications is difficult due to specific implementation details of a middleware platform and a multitude of settings and services provided by middleware for different deployment scenarios. Thus, the performance management of middleware-based applications can be a non-trivial task. Autonomic computing is a new paradigm for building self-managed systems, *i.e.*, systems that seek to operate with minimal human intervention. This work investigates the use of statistical approaches to building autonomic management solutions to control the performance of middleware-based applications. Particularly, we investigate this issue from three perspectives. The first is related to the prediction of performance problems. We propose the use of classification techniques to derive performance models to assist the autonomic management of distributed applications. In this sense, different classes of models in statistical learning are assessed in both offline and online learning scenarios. The second perspective refers to the reduction of false alarms, seeking the development of reliable mechanisms that are resilient to transient failures of the classifiers. This work proposes an algorithm to augment the predictive power of statistical learning techniques by combining them with statistical tests for trend detection. Finally, the third perspective is related to diagnosing the root causes of a performance problem. For this context, we also propose the use of statistical tests. The results presented in this thesis show that statistical approaches can contribute to the development of tools that are both effective, as well as efficient in characterizing the performance of middleware-based applications. Therefore, these approaches can contribute decisively to different perspectives of the problem.

Keywords

Autonomic Computing; Performance Problem Forecasting; Performance Problem Diagnosis; Statistical Learning Techniques; Statistical Testing; Middleware-based Systems;

Sumário

Lista de figuras

Lista de tabelas

Testemunhar com ares de quem possui a verdade é um rotundo desacerto... Pensar certo supõe a disponibilidade à revisão dos achados, reconhece não apenas a possibilidade de mudar de opção, de apreciação, mas o direito de fazê-lo. Mas como não há pensar certo à margem de princípios éticos, se mudar é uma possibilidade e um direito, cabe a quem muda - exige o pensar certo - que assuma a mudança operada.

Paulo Freire