

学位論文題名

Binary-Real Coded Genetic Algorithm for Resource Allocation Problem in Cloud Platforms

(二進実数複合型進化計算によるクラウド資源割当問題の解決)

学位論文内容の要旨

Cloud computing is a promising emerging paradigm that enables enterprises to host multiple independent applications on a shared resource pool with the ability to lease computational power in the form of virtual machines on a per-demand basis. Resource allocation in cloud platforms is a good example of a complex real-world problem. It is expected that an efficient resource management system to ensure high Quality of Service (QoS) under challenging workload variability trends. If the end-user data and applications are moved to the cloud, cloud providers should guarantee the availability of the services against unpredictable system faults. It is also important to enhance the efficiency of the management systems by optimizing power management, minimizing management system overhead, reducing operation costs and limiting the escalating ecological footprint of cloud systems. Finally, virtualization is an enabling technology that adds great flexibility to the resource managers in addressing such challenges. However, utilizing virtualization techniques efficiently is a new challenge of management complexity added to the above mentioned challenges of the cloud platforms.

Many optimization problems in science and engineering involve constraints. Constraint optimization problems have a variety of great challenges resulting from the various limits on the decision variables, the constraints involved, the interference among constraints, and the interrelationship between the constraints and the objective functions. Genetic algorithms (GAs), on the other hand, are efficient search metaheuristics that mimic the operations of natural evolution. GAs play an increasingly important role in a variety of fields and applications. However, constraint optimization problems pose challenges to the performance of GAs, since GAs in their original versions (i.e. the ones that use traditional crossover and mutation operators) are blind to constraints.

Thus, the main objective of this Ph.D dissertation is to design an intelligent resource allocation manager that can flexibly adapt computational resources to satisfy the above mentioned challenges and objectives. To address this objective, I have proposed, firstly, a novel multi-level architecture that relies on a hybrid virtualization framework to address the problems of what it is the optimal number of VMs to be allocated for the running applications? What it is the optimal utilization levels to be assigned for the running VMs? What it the optimal distribution to place the running VMs on? This framework can flexibly and efficiently address a wide spectrum of objectives. Furthermore, I have developed a hybrid GA, Binary-Real coded Genetic Algorithm (BRGA), to act as a core for the management system intelligence. BRGA is a new hybrid approach that has the ability to handle both floating point and binary encoding schemes. It relies on a parameterized hybrid scheme to share the computational power and coordinate the cooperation between Binary coded GA (BGA) and Real coded GA (RGA). BRGA has been applied successfully to a wide spectrum of global and constraint optimization problems from the

known benchmark suites. I have used CEC'2005 benchmark of 25 problems to verify the performance of the algorithm against global optimization problems. To handle constraint optimization effectively, I have introduced a modified dynamic constraint handling technique within the architecture of BRGA. I have also used the CEC'2010 benchmark suite of 18 functions to analyze quality, time and scalability performance of BRGA. BRGA has been applied to the multi-level autonomic architecture. I have evaluated the feasibility, effectiveness and scalability of the proposed approach through simulation experiments.

The dissertation organized as follows: Chapter 1 gives a general introduction of the research. It explains the main motivation and major objectives behind the research. In addition, it reviews the related literature of BGAs and RGAs highlighting the main points of characteristics and disadvantages or critiques. Chapter 1 also discusses the anatomy of hybrid algorithms and clarifies the background of resource allocation problem in cloud platforms. Chapter 2 is mainly about the BRGA. It describes the algorithmic issues in details and discusses the conceptual issues underline the operation of BRGA and the key concepts behind the success of BRGA.

Chapter 3 analyzes the performance of BRGA against global optimization problems through extensive experimentation using the CEC'2005 benchmark suite of 25 functions. The outcomes of the parameter tuning procedure are reported and analyzed. Quality and time performance of BRGA against original binary and real coded GA components are reported and discussed. Finally, to validate the performance of BRGA, the performance of BRGA is compared with some other state-of-the-art evolutionary algorithms from the literature. Chapter 4 analyzes the performance of BRGA against real parameter constraint optimization problems through extensive experimentations using the CEC'2010 benchmark suite of 18 functions. To extend BRGA for constraint optimization problems, a modified dynamic constraint handling technique is introduced within the architecture of the BRGA. The modified penalty method is discussed in details here. To evaluate the affects of modification on BRGA performance, the outcomes of comparison experiments are reported and discussed. To validate the performance of BRGA, the performance is compared with some other state-of-the-art evolutionary algorithms for constraint optimization problems. Finally, the outcomes of the parameter tuning procedure are reported and analyzed.

Chapter 5 describes the proposed multi-level autonomic architecture in detail by illustrating a general overview of the system and defining key terminologies within the system. This Chapter discusses the feasibility of architecture implementation and highlights the main characteristics. The underlying mixed integer mathematical model is presented and explained. To validate the main hypothesis of the research, this Chapter reports the outcomes of an extensive experimental investigation using the randomly generated static datasets, which are snapshots that simulating challenging dynamic scenarios, were used to evaluate the feasibility, effectiveness and scalability of our approach through simulation experiments. Finally, Chapter 6 and 7 conclude the dissertation. BRGA is a new hybrid approach. It enjoys a competitive position when compared with the literature. Multi-level autonomic architecture is a recent approach for the management of virtualized resources within cloud platforms. It has the potential to impact the business environment of cloud computing.

学位論文審査の要旨

主 査 教 授 棟 朝 雅 晴
副 査 教 授 赤 間 清
副 査 教 授 鈴 木 恵 二
副 査 教 授 高 井 昌 彰

学 位 論 文 題 名

Binary-Real Coded Genetic Algorithm for Resource Allocation Problem in Cloud Platforms

(二進実数複合型進化計算によるクラウド資源割当問題の解決)

クラウドコンピューティングにおいて、その資源をサービス要求に応じて適切に配分することは重要な課題であり、利用者からのサービス要求を満足するだけでなく、電力などエネルギー効率の向上、システムのオーバーヘッドや管理コストの削減など、さまざまな基準をもとに最適化を行う多目的最適化問題として定式化することができる。また、その決定変数においては数多くの実数と整数が混在しており、さまざまな制約条件を有することから、その解決は困難な課題となる。

本論文では、クラウドコンピューティングの資源最適配分問題において、上記の困難な課題を解決し柔軟かつ適切な資源配分を実現することを目的としている。そのために、Horizontal Scaling により実行すべきバーチャルマシンの最適数をまず決定し、さらに Vertical Scaling によりそれぞれのバーチャルマシンに割当てられる必要資源量を調整した後、割当先となる物理資源を決定する、マルチレベルの管理アーキテクチャのフレームワークを提案した。このフレームワークにより、さまざまな条件に対して柔軟かつ効率的な資源割当が可能となった。

さらに、実数と整数が混在した困難な最適化問題を解決するため、二進実数複合型進化計算アルゴリズム BRGA (Binary-Real coded Genetic Algorithm) を提案した。BRGA は実数値により符号化された GA と、整数値により符号化された GA を組み合わせた手法であり、それら両方のメカニズムを切り替えて最適化を行うことで、実数と整数が混合した幅広い最適化問題に対応することを目的としている。進化計算の分野で代表的なベンチマーク問題である、CEC2005 ベンチマーク問題、および CEC2010 ベンチマーク問題に対して BRGA を適用した結果、制約付きの困難な最適化問題に対して、その解の質、最適化効率、スケーラビリティにおいて優れた性能を示すことが確かめられた。最終的に、BRGA をマルチレベルのアーキテクチャに基づくクラウドコンピューティングの資源割当問題に適用し、その有効性をシミュレーション実験を通して検証した。

本論文の構成は以下の通りである。第 1 章においては、本論文の背景とその目的につい

て述べており、従来型の GA についてその特徴や有効性について議論するとともに、クラウドコンピューティングにおける資源割当最適化への適用についてその背景について紹介した。第 2 章においては、BRGA についてその遺伝的操作等、アルゴリズムの詳細について紹介した後、アルゴリズム設計の背景にある考え方について議論した。第 3 章においては、BRGA の性能について CEC2005 ベンチマーク問題における 25 の関数最適化の結果について報告し、BRGA と従来型の高性能とされる GA との性能について比較し、その有効性を検証した。第 4 章においては、制約付きの実数値最適化問題のベンチマークである CEC2010 ベンチマーク問題における 18 の関数最適化に BRGA を適用し、制約条件を取り扱うための拡張を行うとともに、BRGA と従来型の高性能とされる GA との比較により、その有効性を検証した。第 5 章においては、マルチレベルのアーキテクチャを提案するとともに、その資源最適割当問題に対して BRGA を適用したシミュレーション実験を行い、提案手法の妥当性、有効性、スケーラビリティについて検証した。第 6 章、第 7 章においては、結論として、BRGA は実数値と整数値による符号化を融合した新たな手法であり、クラウドコンピューティングにおけるマルチレベルアーキテクチャの資源割当最適化問題に対する有効な解決策となることを確認することができた。結果としてクラウドコンピューティングにおいて有用な最適化のフレームワークおよびその解決に必要となる先進的な進化計算アルゴリズムを実現したものであり、情報科学の分野に貢献するところ大なるものがある。よって著者は博士 (情報科学) を授与される資格あるものと認める。