Chapter 16
Schedulers Based on Ant Colony Optimization for Parameter Sweep Experiments in Distributed Environments

Elina Pacini
Institute for Information and Communication Technologies, Universidad Nacional de Cuyo, Argentina

Cristian Mateos
Instituto Superior de Ingenieria de Software, Consejo Nacional de Investigaciones Cientificas y Tecnicas, Argentina

Carlos García Garino
Institute for Information and Communication Technologies, Universidad Nacional de Cuyo, Argentina

ABSTRACT

Scientists and engineers are more and more faced to the need of computational power to satisfy the ever-increasing resource intensive nature of their experiments. An example of these experiments is Parameter Sweep Experiments (PSE). PSEs involve many independent jobs, since the experiments are executed under multiple initial configurations (input parameter values) several times. In recent years, technologies such as Grid Computing and Cloud Computing have been used for running such experiments. However, for PSEs to be executed efficiently, it is necessary to develop effective scheduling strategies to allocate jobs to machines and reduce the associated processing times. Broadly, the job scheduling problem is known to be NP-complete, and thus many variants based on approximation techniques have been developed. In this work, the authors conducted a survey of different scheduling algorithms based on Swarm Intelligence (SI), and more precisely Ant Colony Optimization (ACO), which is the most popular SI technique, to solve the problem of job scheduling with PSEs on different distributed computing environments.

INTRODUCTION

Parameter Sweep Experiments, or PSEs for short, is a very popular way of conducting simulation-based experiments among scientists and engineers through which the same application code is run several times with different input parameters resulting in different outputs (Youn and Kaiser, 2010). Representative examples of PSEs are sensitivity studies of results in terms of defined parameter changes like is the case of imperfections in the simulation of simple tension test, or the study of buckling of imperfect columns.

From a purely software perspective, most PSEs are cluster friendly in the sense that individual inputs of an experiment can be handled by independent jobs. Therefore, using a software platform such as Condor (Thain et al., 2005), which is able to exploit the distributed nature of a computer cluster, allows these jobs to be run in parallel. In this way, not only PSEs execute faster, but also experiments more computing intensive can be computed, and hence more complex simulations can be performed. The same idea has been systematically applied to execute PSEs on Grids (Foster and Kesselman, 2003), which are basically infrastructures that connect clusters via wide-area connections to increase computational power. To this end, software platforms designed to exploit Grids provide the illusion of the existence of a large supercomputer, which in turn virtualizes and combines the hardware capabilities of many much less powerful, geographically-dispersed machines to run resource intensive applications (Coveney et al., 2005).

On the downside, for users not proficient in distributed technologies, manually configuring PSEs is tedious, time-consuming and error-prone. As a consequence, users typically waste precious time that could be instead invested into analyzing results. The availability of elaborated GUIs -especially for Grids- that help in automating an experimentation process has in part mitigated this problem. However, the highly complex nature of today's experiments and thus their associated computational cost greatly surpasses the time savings that can be delivered by this automation.

A recent distributed computing paradigm that is rapidly gaining momentum is Cloud Computing (Buyya et al., 2009), which bases on the idea of providing an on demand computing infrastructure to end users. Typically, users exploit Clouds by requesting from them one or more machine images, which are virtual machines running a desired operating system on top of several physical machines (e.g. a datacenter). Interaction with a Cloud is performed by using *Cloud services*, which define the functional capabilities of a Cloud, i.e. machine image management, access to software/data, security, and so on.

Due to the fact that PSEs perform the processing of a lot of jobs, it is necessary to address how they will be executed in distributed computing environments, which is a complex endeavor. For jobs to be properly executed, it is necessary to allocate them to different resources reasonably and optimally. This problem is known as job scheduling and is an NP-complete problem, which aims to minimize the overall execution time of all jobs. Job scheduling is one of the bottlenecks in distributed computing.

In the last ten years or so, Swarm Intelligence has received increasing attention in the research community. Swarm Intelligence refers to the collective behavior that emerges from a swarm of social insects (Bonabeau et al., 1999). Social insect colonies solve complex problems collectively by intelligent methods. These problems are beyond the capabilities of each individual insect, and the cooperation among them is largely self-organized without any supervision. Through studying social insect colonies behaviors such as ant colonies, researchers have proposed some algorithms or theories for combinational optimal problems. Moreover, job scheduling in Grids or Clouds is also a combinational optimal problem.

Related Content

COVID-19 Vaccination Perceptions, Issues, and Challenges: An Analysis of Tweets Using Machine Learning Models
Sreekantha Desai Karanam, Krithin M.and R. V. Kulkarni (2023). *Convergence of Big Data Technologies and Computational Intelligent Techniques (pp. 113-146).*
www.irma-international.org/chapter/covid-19-vaccination-perceptions-issues-and-challenges/314339

Machine Learning Approach for Multi-Layered Detection of Chemical Named Entities in Text
Usha B. Biradar, Harsha Gurulingappa, Lokanath Khamariand Shashikala Giriyan (2016). *International Journal of Software Science and Computational Intelligence (pp. 1-15).*
www.irma-international.org/article/machine-learning-approach-for-multi-layered-detection-of-chemical-named-entities-in-text/161709

A Maximal Covering Location Model for the Organ Transplant Organization in Turkey
Bar Özkanand Eren Özceylan (2020). *Computational Intelligence and Soft Computing Applications in Healthcare Management Science (pp. 63-86).*
www.irma-international.org/chapter/a-maximal-covering-location-model-for-the-organ-transplant-organization-in-turkey/251968

Implementing Web and Mobile Applications From Linked Open Data
Leila Zemmouchi-Ghomari, Djamaleddine Belilet, Ikram Mekidecheand Abdessamed Réda Ghomari (2022). *International Journal of Software Science and Computational Intelligence (pp. 1-18).*
www.irma-international.org/article/implementing-web-and-mobile-applications-from-linked-open-data/301567

Relevance Feedback as New Tool for Computer-Aided Diagnosis in Image Databases
Issam El Naqa, Jung Hun Ohand Yongyi Yang (2012). *Machine Learning in Computer-Aided Diagnosis: Medical Imaging Intelligence and Analysis (pp. 86-106).*
www.irma-international.org/chapter/relevance-feedback-new-tool-computer/62225