Annual Reports for the Novartis Project 2006-2009

Karl Lieberherr

June 3, 2009

1 Introduction

While my sabbatical at Novartis started with the goal to study crosscuting concerns in biology with the goal to better model biological systems for disease understanding, the interest of my working group headed by Carolyn Cho shifted towards techniques to reason about biological systems under uncertainty. Given a set of conflicting, weighted constraints expressing biological facts, we seek an assignment to the variables that maximizes the fraction of satisfied weighted constraints. In other words, we are finding the most likely state of the modeled biological world given the current incomplete and conflicting knowledge. We studied algorithms for the Maximum Constraint Satisfaction Problem (Max CSP) and their applications in biology.

This eventually lead to a very interesting artificial world in which autonomous agents compete in solving Max CSP problems. We generalized this world to any kind of optimization problem. Success in this world requires to be good at solving instances of the optimization problem.

The world construction has several interesting applications from testing and evaluating optimization algorithms to teaching computer science where students are challenged to grow a simple baby agent into an intelligent agent that succeeds in the world inhabited by agents developed by their peers. The most recent use of the artificial world in a Software Development Course is here:

`http://www.ccs.neu.edu/home/lieber/courses/csu670/sp09/csu670-sp09.html`

14 agents written by 14 student teams took part in the weekly competitions.

2 November 2006 - November 2007

The artificial world construction started with the Evergreen game, only a two person game and continued the following year with the multi-player Specker Derivative Game (SDG). The formulation of the Evergreen game for life scientists is here:

`http://www.ccs.neu.edu/home/lieber/evergreen/game-life-science.html`

The paper explaining the Evergreen game and its solution is here:

`http://www.ccs.neu.edu/research/demeter/biblio/evergreen.html`

We published in 2007 a workshop paper at a workshop on Local Search at the Principles and Practice of Constraint Programming - CP 2007 Conference at Brown University:

`http://www.ccs.neu.edu/research/demeter/biblio/local-search-eg.html`

I gave presentations at Microsoft Research in Redmond and the University of British Columbia in Vancouver.

3 November 2007 - November 2008

This year we created an artificial world out of Max CSP. The artificial world is inhabited by autonomous agents that are driven to produce better algorithms for Max CSP. The autonomous agents were implemented by competing student teams in my software development course. The incoming students were given a simple baby agent that they had to develop into a successful agent that competed with the agents of their class mates to solve Maximum CSP. The rules of artificial world were enforced by an administrator. The artificial world is described in the following two documents:

Requirements:

`http://www.ccs.neu.edu/home/lieber/courses/csu670/f08/requirements/1sep15/sdg.pdf`

Specker Derivative Game Home Page: `http://www.ccs.neu.edu/home/lieber/evergreen/specker/sdg-home.html`

An important development this year was the generalization from classic SDG to secret SDG: `http://www.ccs.neu.edu/home/lieber/evergreen/specker/secret-hiding.html`

I gave a presentation at ETH Zurich in July 2008 on SDG. Also in 2008, I was invited keynote speaker for a European National Science Foundation Workshop on Correlations where I also talked about SDG.

4 November 2008 - November 2009 (partial report)

We developed a much better baby agent to let the agent developers focus on improving the intelligence of their agent and not the basic mechanisms like "walking" and "talking". The administrator was also significantly improved to make sure that the rules of the world are reliably enforced.

We want now to bring SDG to the web. This lead to a wider dissemination of this technology:

`http://www.ccs.neu.edu/home/lieber/evergreen/specker/web/WebSDG`

We started the study of the broader applications of SDG. It seems to have applications to the study of cancer cell behavior (the "perfect" autonomous agent acts like a cancer cell), evaluation of software products in a software ecosystem and testing of software. Our current focus is on optimization problems of which Max CSP is a special case.

A derivative consists of a predicate and a price. Once a derivative is delivered, an optimization problem instance, satisfying the predicate, will be delivered with a secret quality ($secret\_quality$) and the instance solved by the buyer with $achieved\_quality$. The payoff is a function

$$payoff(secret\_quality, achieved\_quality, price).$$

For example,

$$if\ achieved\_quality >= secret\_quality * price\ then\ 3 * price\ else\ 0.$$

For the classic version we would ignore the $secret\_quality$:

$$if\ achieved\_quality >= price\ then\ 3 * price\ else\ 0.$$

A simpler version:

$$achieved\_quality - price.$$

We could also ignore the price in the condition:

$$if\ achieved\_quality >= secret\_quality\ then\ 3 * price\ else\ 0.$$

Many other payoff functions are possible and each one leads to a different world.

What is important about the payoff function is that a good algorithm for solving the optimization problem will be rewarded by a good payoff. To succeed in the game, it is also necessary to analyze the predicates and find the hardest instance satisfying a predicate. But those skills are likely to lead to better algorithms for solving the optimization problem.

5 Conclusions and Plan

We believe that SDG is a useful technology to test and evaluate optimization processes of relevance to Novartis. Biological systems perform numerous optimization processes that are relevant for disease understanding. SDG can help to find more efficient simulation algorithms for simulating the biological processes at a higher level than differential equations.

The Novartis project has come a long way from reasoning with uncertainty using Max CSP to an artificial world inhabited by autonomous agents with the purpose to evaluate and test algorithms for optimization problems.

For the final year of the project, we will polish and evaluate our web implementation of SDG and make it available to a broader community:

- Instructors who want to teach students reliable software development within the context of an optimization process. We did this at Northeastern University for four courses with Maximum CSP as the optimization process. We plan it again for the fall 2009:

  `http://www.ccs.neu.edu/home/lieber/courses/csu670/f09/requirements/two-player-sdg.pdf`

- Researchers who want to objectively measure the quality of their optimization algorithms in a dynamic context. SDG complements the static benchmark technology.

- Software Developers who want to thoroughly test their optimization process implementations.

- Software Ecosystems developing software for optimization processes.

We will write two papers about SDG. One will be for the game community that is interested in serious games for Computer Science education. The second one is for the software development community showing how SDG is useful as a sophisticated testing and evaluation tool for optimization processes that is effective at producing robust and efficient implementations of optimization processes. We believe that the techniques are applicable to computational processes in general.