SIGEVOlution

newsletter of the ACM Special Interest Group on Genetic and Evolutionary Computation

Summer 2008
Volume 3 Issue 2

in this issue

Editorial

Happy new year and welcome to the new issue of SIGEVOlution. You are probably mumbling "...the Summer issue arrives during Winter ...again?" Indeed, this is almost becoming a tradition. As it happened last year, we had some minor issues which delayed the editing process. In the meanwhile however, we have been working on a little surprise for you that is already sitting in my mailbox just waiting to reach you in the near future, so stay tuned for more.

Third volume, second issue, with a rich menu to choose from.

Bioinformatics has become a major research area which poses many challenges and, most important, provides many research opportunities. If you want to know more, the first paper by Casey Greene and Jason Moore is a very good point to start from.

Pherographia is instead a brand new drawing procedure inspired by artificial ants that has strong connections to the camera obscura used in the early days of photography. Carlos M. Fernandes tells you more about it in the second paper.

While if you wish to have fun with Grammatical Evolution, then GEVA, the tool presented in the third paper by Michael O'Neill, Erik Hemberg, Conor Gilligan, Eliott Bartley, James McDermott, and Anthony Brabazon, is probably the best choice.

At the end, we have the columns providing information about the latest edition of the International Conference on Evolvable Systems (ICES-2008), the new books, and the forthcoming events.

As always, this issue would not be here without the many people who helped me cooking it, Casey S. Greene, Jason H. Moore, Carlos M. Fernandes, Michael O'Neill, Erik Hemberg, Conor Gilligan, Eliott Bartley, James McDermott, Anthony Brabazon, Andrew Greensted, Jaume Bacardit, Gregory Hornby, Cristiana Bolchini, Stewart Wilson, Christian Gagné, and board members Dave Davis and Martin Pelikan.

Last but not least, I wish to thank Franz Rothlauf for the deadline extension, what a relief... Just kidding!

The cover is a photo of Montréal by Frank Slack, the original image is available here.

Pier Luca
January 19th, 2009

SIGEVOlution
Summer 2008, Volume 3, Issue 2

Newsletter of the ACM Special Interest Group on Genetic and Evolutionary Computation.

Contributors to this Issue

Casey S. Greene
Jason H. Moore
Carlos M. Fernandes
Michael O'Neill
Erik Hemberg
Conor Gilligan
Eliott Bartley
James McDermott
Anthony Brabazon
Andrew Greensted

Contents

Solving Complex Problems in Human Genetics Using GP

Challenges and Opportunities

Casey S. Greene, Dartmouth College, Lebanon, NH 03756 USA, Casey.S.Greene@dartmouth.edu
Jason H. Moore, Dartmouth College, Lebanon, NH 03756 USA, Jason.H.Moore@dartmouth.edu

The development of rapid data-collection technologies is changing the biomedical and biological sciences. In human genetics chip-based methods facilitate the measurement of thousands of DNA sequence variations from across the human genome. The collection of genetic data is no longer a major rate limiting step. Instead the new challenges are the analysis and interpretation of these high dimensional and frequently noisy datasets. The specific challenge we are interested in is the identification of combinations of interacting DNA sequence variations predictive of common human diseases. Specifically, we wish to detect epistasis or gene-gene interactions. Here we focus solely on the situation where there is an epistatic effect but no detectable main effect. The challenge for applying search algorithms to this problem is that the accuracy of a model is not indicative of the quality of the attributes within the model. Instead we use pre-processing of the dataset to provide building blocks which enable our evolutionary computation strategy to discover an optimal model.

We focus on the analysis of single nucleotide polymorphisms or SNPs. A SNP is a single nucleotide or point in the DNA sequence that differs among people. At least one SNP is expected every 100 nucleotides across the 3×10^9 nucleotide human genome. An important goal in human genetics is determining which of these SNPs are useful for predicting an individual's risk of disease. This "genome-wide" approach is expected to revolutionize the genetic analysis of common human diseases. Despite the promise of this broad unbiased approach, finding successful predictors will be difficult due to nonlinearity in the genotype to disease relationship that is due, in part, to epistasis.

The implication of epistasis from a data mining point of view is that SNPs need to be considered jointly in learning algorithms rather than individually. Moore argues that epistasis, far from being a corner case, is likely to be ubiquitous in common human diseases [8]. The challenge of modeling attribute interactions in data mining has been previously described [2]. Due to the combinatorial magnitude of this problem, analysis strategies which use outside knowledge are crucial.

Combining the challenge of modeling nonlinear interactions with the difficulty of attribute selection yields what Goldberg calls a needle-in-a-haystack problem [3]. There may be a particular combination of SNPs that, together with the right nonlinear functions, are a significant predictor of disease susceptibility, but considered individually they may not look any different than thousands of irrelevant SNPs. Under these conditions the learning algorithm is looking for a genetic needle in a genomic haystack. A recent report from the International HapMap Consortium suggests that approximately 300,000 carefully selected SNPs may be necessary to capture all relevant variation across the Caucasian human genome [23]. This number is probably a lower bound, but assuming it is true we would need to scan 4.5×10^{10} pairwise combinations of SNPs to find a single genetic needle. The number of higher order combinations is astronomical. We find that without expert knowledge these approaches perform no better than a random search, but methods that integrate expert knowledge into the search strategy succeed in this domain [25, 16]. Each of these successful approaches exploit knowledge gained by pre-processing the data.

Here we review genetic programming (GP) as a strategy for this problem. Genetic programming is an attractive approach because many of the difficulties found in biology are likely to also be issues here. Biological organisms evolve in a noisy environment with a rugged fitness landscape. Many of the interesting problems in human genetics also likely involve a rugged fitness landscape where models that contain some but not all of the relevant attributes may not have an accuracy greater than that of the surrounding noise. In addition these data are frequently noisy because two individuals with the same values at relevant attributes may have different disease states. In this context it is no surprise that we look to natural systems for inspiration when designing algorithms. Wagner discusses the role of robustness and evolvability in living systems [24]. We must design and use algorithms that, like living systems, are both robust to the noise in the data and evolvable despite the rugged fitness landscape.

Building Blocks

Genetic Programming and other evolutionary algorithms succeed when components that comprise successful solutions are, on their own, usually able to improve the fitness of a solution. These successful small components are the building blocks these algorithms exploit to find good solutions. Unfortunately with the epistasis problem in human genetics we lack these building blocks because interacting SNPs, though they may be predictive of disease risk when considered with other SNPs, are not individually predictive. Therefore for a GP strategy to be successful we require a source of outside knowledge which can provide the building blocks our problem lacks. We use an algorithm based on Relief, a method developed by Kira and Rendell [6], which is capable of pre-processing the input data and weighting attributes (SNPs) on how well they, in the context of other SNPs, are able to differentiate individuals with disease from those without. Kononenko improved upon Relief by choosing n nearest neighbors instead of just one [7]. This new algorithm, ReliefF, has been shown to be more robust to noisy attributes and missing data and is widely used in data mining applications [20]. Unfortunately the power of ReliefF is reduced in the presence of a large number of noisy attributes. This drove the development of Tuned ReliefF (TuRF), which systematically removes attributes that have low quality estimates so that the weights of the remaining attributes can be re-estimated. TuRF is significantly better than ReliefF in the domain of human genetics [15]. The GP strategies we discuss here use TuRF as their source of expert knowledge.

Multifactor Dimensionality Reduction (MDR)

Our GP approach focuses solely on attribute selection. We therefore need a method to identify whether or not the attributes selected are relevant. For this we use multifactor dimensionality reduction (MDR) which has been developed as a nonparametric and genetic model-free data mining strategy for identifying combinations of SNPs predictive of a discrete clinical endpoint [19, 9, 12, 10]. The MDR method has been successfully applied to the detection of gene-gene interactions in a variety of common human diseases. At the core of the MDR approach is an attribute construction algorithm that creates a new attribute by pooling genotypes from multiple SNPs. MDR constructs a new one-dimensional attribute with high and low risk levels. It is this new single variable that is returned by the MDR function in the GP root node (Section). The MDR method is described in more detail by Moore et al. [12]. Open-source MDR software is freely available from www.epistasis.org.

Genetic Programming with Expert Knowledge Building Blocks

We have developed GP-MDR which is MDR wrapped in a GP framework for attribute selection. Figure 1A illustrates an example GP tree for this problem. In this work the solution representation is kept simple with one function in the root node and two leaves to evaluate the best GP parametrization for attribute selection. More complex trees (e.g. Figure 1B) can be explored once the principles of how the GP performs attribute selection in simpler trees are understood. By first focusing on attribute selection, we separate the task of finding good attributes from the task of generating good models. To these ends the MDR approach is used at the root node because it is able to capture interaction information.

In this representation each tree has two leaves or terminals consisting of attributes. Given that the challenge has been described as a needle in a haystack, we were not surprised to discover that without building blocks from expert knowledge the GP approach performed no better than a random search [16].

Goldberg describes a class of genetic algorithms called "Competent Genetic Algorithms" [3]. These are genetic algorithms that solve specific problems quickly, reliably, and accurately and which exploit knowledge about the task or dataset to the fullest extent possible. In our successful approaches we use a multiobjective fitness function [16], an initialization operator [5], a recombination operator [14], and a mutation operator [4].

Multiobjective Fitness Functions

First, we integrate expert knowledge in a multiobjective fitness function consisting of two pieces in a simple linear combination of the form α*A+β*Q [16]. Here, A is the measure of accuracy obtained from the analysis of the single constructed attribute from the GP tree using a naïve Bayes classifier. The parameter α weights the accuracy measures. Q in this function represents the attribute quality estimate obtained from pre-processing the dataset using the TuRF algorithm. The parameter β weights the quality measures. The GP can identify useful building blocks because information from TuRF is used.

We found that a GP with accuracy (A) as the fitness function does no better than random search (R) across all genetic models and all genetic effect sizes. There is a clear difference between the GP that uses attribute quality (Q) in the fitness function versus the GP that only uses accuracy (A). This difference was statistically significant across most models and most heritabilities. It is interesting to note that increasing the weight of the attribute quality to twice that of accuracy (α=1 and β=2) performed no better than equal weighting, thus it is clearly important to provide expert knowledge but a higher weighting of expert knowledge is not necessarily better.

Sensible Operators

While a fitness function is one way to exploit expert knowledge in an evolutionary search, it is also possible to utilize expert knowledge via operators tuned to exploring areas of the search space thought to be beneficial. Work has been done to develop such operators. This knowledge is provided through TuRF scores gained from pre-processing. These operators have been titled "sensible" operators as they make sense in our context and heavily exploit dataset specific information. Here we examine three of these operators which could be useful for building "Competent Genetic Algorithms" for this domain, and then discuss results obtained using these operators.

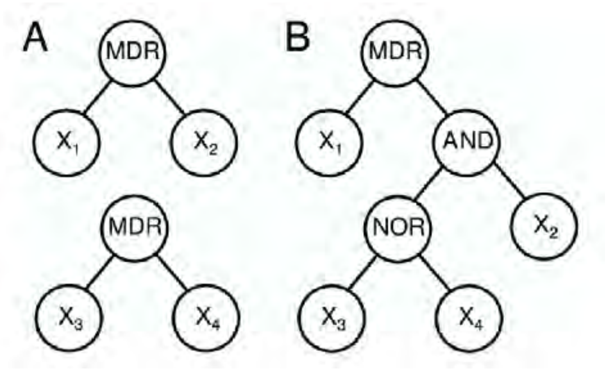


Fig. 1: Example GP trees for solutions (A). Example of a more complex tree that we hope to address in future studies (B).

Sensible Initialization. With our sensible initializer we focus on using expert knowledge for the initialization of terminals in this GP tree structure. O'Neill and Ryan discuss the importance of initialization and the negative impact of a lack of diversity on final solutions [17]. We apply their principles of sensible initialization in two different initialization operators [5]. The first is an exhaustive initializer focused on diversity and the second is an expert knowledge based initializer focused on exploiting pre-processing information for population initialization. The exhaustive initializer insures maximal diversity by selecting attributes to be leaves without replacement until all attributes have been used, at which point it replenishes the pool from which attributes are selected. This insures that all attributes are used once before any are used a second time. The expert knowledge aware probabilistic initializer selects attributes for terminals via a weighted roulette wheel. The same attribute is not allowed to be used twice within the same tree, but it may be used any number of times within the generated population.

The diversity focused initializer did not outperform a random initialization operator. The probabilistic initializer was successful however. This initializer was not as successful in its own right as either the sensible recombination or sensible mutation operators, but using the expert knowledge driven probabilistic initialization operator did improve the power of a GP also using the "competent fitness function."

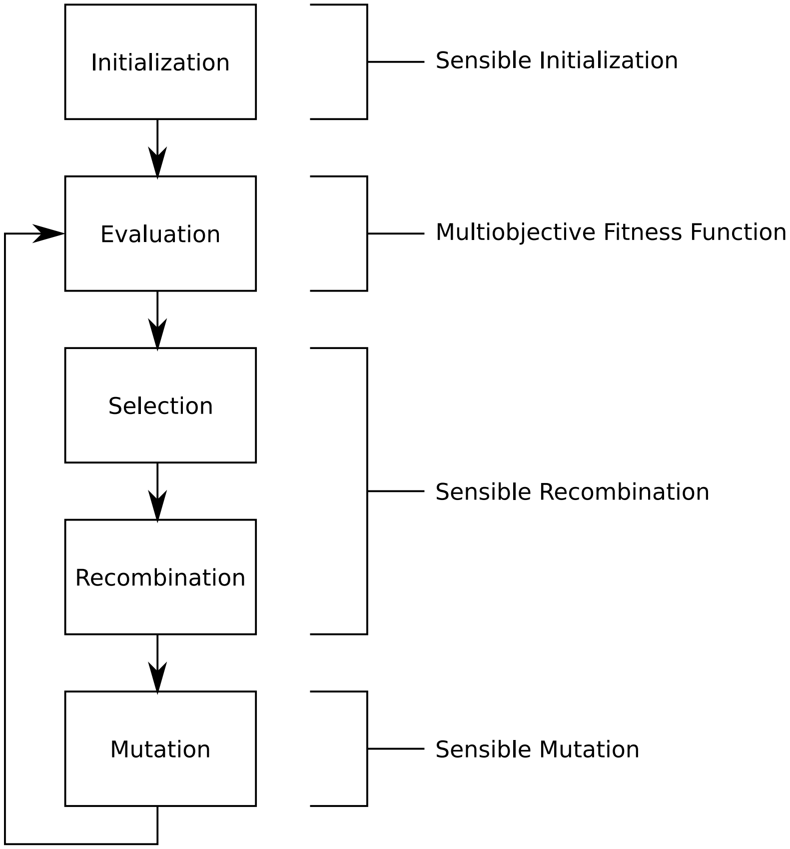


Fig. 2: A schematic showing the operators in GP and the population flow between them. The notation on the right shows where expert knowledge has been applied to the system.

Sensible Recombination. Our sensible recombination operator heavily exploits expert knowledge building blocks [14]. We modified this operator by first selecting the top 1% or 5% of trees according their maximum TuRF score. Each possible pair of these selected trees is recombined and the children are included in the next generation. For example, with a population size of 500 either 5 or 25 trees are selected. There are 5 choose 2 (10) and 25 choose 2 (300) possible recombination events. Thus, the new population consists of either 10 or 300 recombined trees generated from those with the maximum TuRF scores. The remaining 490 or 200 individuals in the population are generated using the standard binary tournament operator. This new selection operator ensures that the best trees as measured by pre-processed expert knowledge are selected and recombined. Generating the remaining trees by binary tournament ensures that functional attributes not assigned a high quality by TuRF still have a chance of being evaluated. This sensible recombination approach succeeds in this domain. Despite exploring only 1% of the available search space, this method is able to quickly and reliably discover interacting SNPs in simulated datasets when standard GP and random search cannot.

Sensible Mutation. The sensible mutation operator is based on the principles of the sensible recombination operator [4]. It aggressively applies information from pre-processing to generate solutions. One potential weakness of the recombination operator is that it can only use attributes still in the population. Sensible mutation, on the other hand, can exploit attributes not currently in the population. This operator mutates individuals with the greatest difference in TuRF scores between the leaves, changes the leaf with the lowest TuRF score, evaluates trees where the changed leaf is replaced with high TuRF scoring attributes, and retains the best individual for the next generation. This sensible mutator ensures that poor building blocks are replaced by good building blocks throughout evolution. This method is designed to change leaves unlikely to lead to success and provide attributes likely to be involved in successful models. For comparison with a random mutator, the percentage of the population chosen for mutation is adjusted to account for the iterative replacement attempts. This mutation driven approach also succeeds in this domain. Like the sensible mutation operator, this approach was limited to approximately 1% of the available search space. Despite these limitations, this method also quickly and reliably finds the simulated relevant SNPs.

Discussion

Both sensible recombination and sensible mutation were extremely successful. Despite exploring less search space (1% vs 10%) relative to the multiobjective fitness function, the power of each method was similar. One area that remains to be explored is how well these approaches work cooperatively. Are there any interactions between the operators that can increase the power of GP for attribute selection? One potential issue with the sensible operator approach is that the operators are rigid and predefined. They work for this problem, but will they generalize to other problems?

It is clear that work in this area will require the use of expert knowledge if GP approaches are going to successfully find predictors of common human diseases. Now the charge is to develop simple and efficient methods for attribute selection, implement and evaluate low parameter approaches for both attribute selection and epistasis modeling, and improve and develop new and better sources of expert knowledge. Developing systems with fewer parameters but high power will make the application of these systems to real biological problems more routine. Better expert knowledge will speed the search in these algorithms making them more efficient. Progress in these areas will help put this class of methods in the modern geneticist's toolbox.

We can improve our expert knowledge component by improving our preprocessing methods or by using information derived from biological data. Improvements to the power of ReliefF or TuRF will play a crucial role in our ability to find answers using nature inspired algorithms. Using biological expert knowledge from protein-protein interaction databases, the Gene Ontology, or known biochemical pathways we should be able to identify SNPs likely to be predictive of disease risk. For example, Pattin et al. argue that once we successfully develop the methods to extract expert knowledge from protein-protein interaction databases, we will improve our ability to identify important epistatic interactions in genome-wide studies [18]. By using biological knowledge to drive this search, the potential exists to enhance our comprehension of common human diseases. Eventually this could lead to an improvement in the prevention, diagnosis, and treatment of these diseases.

To develop parameter free or parameter robust approaches, we should explore methods that minimize or avoid user-defined parameters. While parameter sweeps are one way to approach the problem they are time consuming and cumbersome. In addition, instead of providing predefined "sensible" operators to enhance the GP, is there a way to allow the method to discover good operators on its own? It may be possible to evolve complex operators which tune themselves to the specific problem, outside knowledge, or dataset. Evolvable operators have already been applied to GP through Meta-GP [1] and PushGP [21, 22]. We have developed a prototype computational evolution system (CES) which can evolve operators that can exploit expert knowledge [11]. Furthermore we have shown that allowing the system to use expert knowledge improves the quality of the solutions [13].

Using the CES with an island model and entirely evolvable parameters could allow for a situation where the only two parameters are how many CPUs should be used and how long the run should continue. Our goal is to develop techniques which solve these hard genetics problems with minimal tweaking. If these nature-inspired approaches are to become widespread in human genetics, methods that powerfully analyze genome-wide datasets with minimal parameter sweeping will be critical.

Bibliography

[1] B. Edmonds. Meta-genetic programming: Co-evolving the operators of variation. CPM Report 98-32, Centre for Policy Modelling, Manchester Metropolitan University, UK, Aytoun St., Manchester, M1 3GH. UK, 1998.

[2] A. A. Freitas. Understanding the crucial role of attribute interaction in data mining. *Artif. Intell. Rev.*, 16(3):177–199, 2001.

[3] D. E. Goldberg. *The Design of Innovation: Lessons from and for Competent Genetic Algorithms*. Kluwer Academic Publishers, Norwell, MA, USA, 2002.

[4] C. S. Greene, B. C. White, and J. H. Moore. An expert knowledge-guided mutation operator for genome-wide genetic analysis using genetic programming. *Lecture Notes in Bioinformatics*, 4774:30–40, 2007.

[5] C. S. Greene, B. C. White, and J. H. Moore. Using expert knowledge in initialization for genome-wide analysis of epistasis using genetic programming. *Genetic and Evolutionary Computation Conference 2008*, In Press-.

[6] K. Kira and L. A. Rendell. A practical approach to feature selection. *In: Machine Learning: Proceedings of the AAAI'92*, 1992.

[7] I. Kononenko. Estimating attributes: Analysis and extension of relief. *Machine Learning: ECML-94*, pages 171–182, 1994.

[8] J. H. Moore. The ubiquitous nature of epistasis in determining susceptibility to common human diseases. *Human Heredity*, 56:73–82, 2003.

[9] J. H. Moore. Computational analysis of gene-gene interactions using multifactor dimensionality reduction. *Expert Review of Molecular Diagnostics*, 4(6):795–803, 2004.

[10] J. H. Moore. Genome-wide analysis of epistasis using multifactor dimensionality reduction: feature selection and construction in the domain of human genetics. In *Knowledge Discovery and Data Mining: Challenges and Realities with Real World Data*. IGI, 2007.

[11] J. H. Moore, P. C. Andrews, N. Barney, and B. C. White. Development and evaluation of an open-ended computational evolution system for the genetic analysis of susceptibility to common human diseases. *Lecture Notes in Computer Science*, 4973:129–140, 2008.

[12] J. H. Moore, J. C. Gilbert, C. T. Tsai, F. T. Chiang, T. Holden, N. Barney, and B. C. White. A flexible computational framework for detecting, characterizing, and interpreting statistical patterns of epistasis in genetic studies of human disease susceptibility. *Journal of Theoretical Biology*, 241(2):252–261, Jul 2006.

[13] J. H. Moore, C. S. Greene, P. C. Andrews, and B. C. White. Does complexity matter? artificial evolution, computational evolution and the genetic analysis of epistasis in common human diseases. In *Genetic Programming Theory and Practice VI*. Springer, In Press.

[14] J. H. Moore and B. C. White. Exploiting expert knowledge in genetic programming for genome-wide genetic analysis. *Lecture Notes in Computer Science*, 4193:969–977, 2006.

[15] J. H. Moore and B. C. White. Tuning ReliefF for genome-wide genetic analysis. *Lecture Notes in Computer Science*, 4447:166–175, 2007.

[16] J. H. Moore and B. C. White. Genome-wide genetic analysis using genetic programming: The critical need for expert knowledge. In R. Riolo, T. Soule, and B. Worzel, editors, *Genetic Programming Theory and Practice IV*. Springer, 2007.

[17] M. O'Neill and C. Ryan. *Grammatical Evolution: Evolutionary Automatic Programming in an Arbitrary Language*. Kluwer Academic Publishers, Norwell, MA, USA, 2003.

[18] K. A. Pattin and J. H. Moore. Exploiting the proteome to improve the genome-wide genetic analysis of epistasis in common human diseases. *Human Genetics*, 124(1):19–29, 2008.

[19] M. D. Ritchie, L. W. Hahn, N. Roodi, L. R. Bailey, W. D. Dupont, F. F. Parl, and J. H. Moore. Multifactor dimensionality reduction reveals high-order interactions among estrogen metabolism genes in sporadic breast cancer. *American Journal of Human Genetics*, 69:138–147, 2001.

[20] M. Robnik-Sikonja and I. Kononenko. Theoretical and empirical analysis of relieff and rrelieff. *Mach. Learn.*, 53(1-2):23–69, 2003.

[21] L. Spector. Autoconstructive evolution: Push, pushGP, and pushpop. In L. Spector, E. D. Goodman, A. Wu, W. B. Langdon, H.-M. Voigt, M. Gen, S. Sen, M. Dorigo, S. Pezeshk, M. H. Garzon, and E. Burke, editors, *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2001)*, pages 137–146, San Francisco, California, USA, 2001. Morgan Kaufmann.

[22] L. Spector and A. Robinson. Genetic programming and autoconstructive evolution with the push programming language. *Genetic Programming and Evolvable Machines*, 3(1):7–40, 2002.

[23] The International HapMap Consortium. A haplotype map of the human genome. *Nature*, 437(7063):1299–1320, Oct 2005. 10.1038/nature04226.

[24] A. Wagner. *Robustness and Evolvability in Living Systems (Princeton Studies in Complexity)*. Princeton University Press, Princeton, NJ, August 2005.

[25] B. C. White, J. C. Gilbert, D. M. Reif, and J. H. Moore. A statistical comparison of grammatical evolution strategies in the domain of human genetics. *Proceedings of the IEEE Congress on Evolutionary Computing*, pages 676–682, 2005.

About the authors

Casey S. Green received his B.S. in Chemistry from Berry College in 2005. He is currently a Ph.D. candidate in the Department of Genetics at Dartmouth College. Casey's work focuses on developing machine learning and evolutionary computing methods to detect and characterize genetic and environmental predictors of common human disease in the absence of single-factor effects. His research is supported by the Superfund Basic Research Program from the National Institute of Environmental Health Sciences.

Email: Casey.S.Greene@dartmouth.edu

Jason H. Moore received his B.S. in Biological Sciences from Florida State University. He then received an M.S. in Human Genetics, an M.A. in Statistics and a Ph.D. in Human Genetics from the University of Michigan. Dr. Moore is currently the Frank Lane Research Scholar in Computational Genetics, Professor of Genetics, Professor of Community and Family Medicine, and Director of Bioinformatics at Dartmouth Medical School and Dartmouth College.

Homepage: http://www.epistasis.org/
Email: Jason.H.Moore@dartmouth.edu

A Camera Obscura for Ants

Carlos M. Fernandes, Evolutionary and Biomedical Engineering Laboratory
Technical University of Lisbon, cfernandes@laseeb.org

Pherographia is a drawing procedure that arises from an artificial life model and addresses the hypothetical relationship between photography and the so-called cognitive maps created by the model. The theme is discussed within the emergent research field called *Artificial Art* and recent theoretical advances that link Swarm Intelligence and cognitive sciences are also addressed. The system and its equations are given and their outcome is shown in its different forms. Finally, the drawings and other potentialities of the system as a creative tool are discussed within an image-processing framework.

Introduction

The artificial worlds investigated by the new sciences of complexity and artificial life, together with the burst of technological development experienced in the last three decades, are serious candidates for producing analysis tools that will scrutinize, in the future, the blurred borders between art and science. In addition, these disciplines are creating "objects" that very often are classified as art, thus inducing the debate on the possibility of a kind of artificial creativity, completely independent from the creator of the machine. Man is regarded as the agent that sparks the process, and nothing more. This discussion around the artificial art hypothesis, born of Artificial Intelligence and Artificial Life research fields, is growing and becoming quite interesting. Shapes created by Genetic Algorithms [1] (usually guided by humans) and "abstract" drawings that emerge from artificial insects' colonies are usually regarded as artificial art.

Other figures and forms that appear from other artificial models, not directly inspired by nature, but instead the result of sets of simple rules, from which emerges complex behavior, are also finding their way through exhibition halls and museums. Cellular automata [2] are one example of the achievements of the science of complexity in devising/finding models comprised of simple entities that give rise to complex and unpredictable behavior.

Those of Class IV [3], in particular, generate open-ended novelty by working between chaotic and ordered regimes, and were proven by Cook [4] to be capable of functioning as universal computers. The patterns created by these cellular automata are quite impressive in their diversity and constant innovation. Fractals are another example. The term – coined by Benoît Mandelbrot [5] in 1975 – refers to geometric shapes that exhibit self-similarity, that is, forms composed of parts that are similar to the whole. The property appears at all levels of magnification (in nature, some shapes approximate fractals to a certain degree), exhibiting a remarkable symmetry and complexity that instantly seduces the human eye. Other systems use robotics, combining hardware and software to take advantage of the complex interaction of the machine with the environment – much more complex than the interaction that occurs in simulated models – and generating “artistic” objects that are usually classified within the traditional clusters (painting and sculpture, for instance).

Fig. 1: Carlos M. Fernandes, A Tribute to Fernando Pessoa, 2008. The image shows the state of the model (in this case, the pheromone map) at different stages of the run.

But despite all this evidence of innovation and open-ended evolution found in some systems discovered or designed during the last five decades of scientific research on complex systems, we still need to devise a clear frontier between this so-called new Science (as Stephen Wolfram puts it, in the title of his controversial book: *A New Kind of Science* [3]) and the "old" Science — mathematics, for instance. Fernando Pessoa (1888-1935), Figure 1, wrote, under the pseudonym Álvaro de Campos [6]: *Newton's Binomial is as beautiful as Venus de Milo. But few people notice it* .

A fractal, or a landscape of artificial pheromone translated into a tridimensional object, is probably nothing more than the Binomial of the new Science. Few people would classify Newton's Binomial as an art object. Beauty is a labeling that certainly arises amongst the opinions of many viewers, at least those aware of the language (that is, Mathematics) behind Newton's formula, but Artificial Art, probably not. Therefore, what may be left as a (consensual) contribution of this new kind of Science to the dialogue between Art and Science? Maybe one thing that Technique has always given us: tools (analytical and stricto sensu). We shall proceed to the description of one of those tools, an object that is inseparable from the History of both Photography and Painting: the *camera obscura*.

The Camera Obscura

The *camera obscura* was the "photographic" camera before the invention of Photography. Its principles are understood from Aristotle's times, and, in the 16th century — when the narrow (pin)hole in front of it was replaced by a lens —, the *apparatus* became a medium for the permanent record of images. When Johann H. Schulze (1687-1744) was still making photograms with letters that he cut and glued to bottles full of silver chloride and nitric acid, and when the daguerreotype was only a dream of a young Niépce (1765-1833), there was already a camera obscura prepared to accept the new invention. But before the arrival of Photography (*drawing with light*), the camera obscura was used to draw over projected images. It was the canvas for the artist's pencil before being the magic box for *The Pencil of Nature* [7]. (This is the title of the celebrated book – see Figure — by William Henry Fox Talbot (1800-1877), the creator of the positive-negative photographic process). Johannes Vermeer (1632-1675), Canaletto (1697-1798) and Francesco Guardi (1712-1793) were some amongst many painters that probably1 used the device to create some of their masterpieces. Museum walls all over the world are filled with works of art that were carried out with the aid of the *camera*. The scenario was projected, through the lens, on a panel, and then the artist just had to “follow the lines“ in order to attain a draft. Only later, the camera was used to project images over a light-sensitive medium. The light, of course, was the same, but silver replaced the artist's hand.

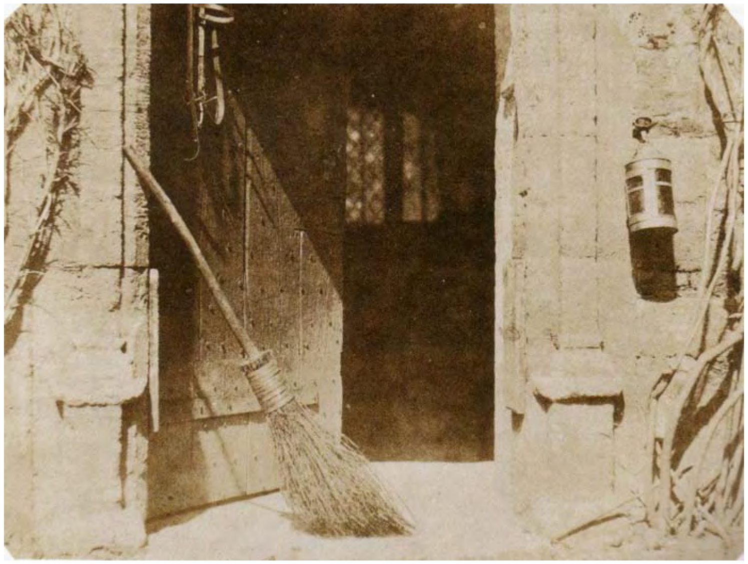


Fig. 2: The broom, by William Henry Fox Talbot. This photo was published in Talbot's celebrated book *The Pencil of Nature* [7].

Baudelaire (1821-1867) expressed his concern about the popularity of photography in a press article published [8, 9] in 1859: "(. . .) *During this lamentable period, a new industry arose which contributed not a little to confirm stupidity in its faith and to ruin whatever might remain of the divine in the French mind* (...)" [9]. But he had no reasons to be so exasperated; the bridges between Photography and Painting were already being built by the painters that carried the camera obscura when leaving the studio in search of a subject.

After 1839, Photography became another tool in the hands of mankind. For one hundred and fifty years, the camera was used as the first step in the magical act of capturing light on silver salts: light darkens silver, development speeds up the process, and sodium thiosulfate removes the unexposed salts, fixing the image for posterity. Keeping in mind the fundamental principles of the *camera obscura*, we now proceed to the description of an artifact, born of the simulation of an emergent phenomenon that is a kind of digital camera where artificial ants draw on top a ”projected” image using (artificial) pheromones.

The Ant System

The concept discussed in this paper is the result of a simple Artificial Life model (for details, the interested reader may refer to [10], [11] and [12]). The drawings arise from the interaction between entities (artificial ants) that follow simple rules, and also between the ants and the environment. A gray-scale digital image is used as a lattice, where in each cell lies the pixel's gray-level value (ranging from 0 to 255). We call it *habitat*. An ant possesses only local vision, that is, it is only aware of the gray-level value of the cell in which it lies on and of the eight neighboring cells (Moore neighborhood). The ants have no global perception of the image (habitat). That perception emerges from the local interactions (an ant isolated from a swarm is not able to create cognitive maps, made up of pheromone; it is necessary a minimum amount of ants – critical mass – for the system to establish communication networks and order emerge). Besides that sensorial capacity, the ants walk through the lattice and deposit pheromone on the cells. The amount of pheromone depends on the contrast observed in the position (and surroundings) of the ant: if an ant perceives a region of high contrast, then it deposits a larger amount (T) of pheromone than it would if lying on a low contrast area — see equation 1:

$$T = \eta + p\Delta_{gl}/255 \tag{1}$$

where η is a constant amount, p is a parameter of the system and Δ_{gl} measures the contrast.

After depositing pheromone, the ant decides which cell to move next by a probabilistic process that takes into account the direction from which the ant came and the intensity of pheromone in the neighboring cells.

$$P_{ik} = \frac{W(\sigma_i)w(\Delta_\theta)}{\sum_{j/k} W(\sigma_j)w(\Delta_\theta)} \tag{2}$$

$$W(\sigma) = \left(1 + \frac{\sigma}{1+\delta\sigma}\right)^{\beta} \tag{3}$$

In Equation 2, P_{ik} is the probability of moving from cell k to i, W(σ_i) is given by Equation 3 (where the parameters β, σ and δ control the behavior of the system) and $w(\Delta\phi)$ are the weights that assure, for instance, that the probability of a U-turn is low [10].

The colony is composed of N ants and the process is iterative. In each iteration, all ants must try to move to a neighboring cell, following pheromone, reinforcing it, and thus engaging in what is called a stigmergic process.[1] Another fundamental step in the stigmergic process, besides positive feedback (here represented by pheromone reinforcement and sensing), is evaporation: in each time step, the pheromone in all cells is decreased by a constant amount. This mechanism promotes the emergence of new paths and re-adaptation to changing environments. Without it the system would freeze in local optima. Evaporation is the way the system "forgets", and when dealing with dynamic environments, the rule played by this property is even more important. By using the capacity of the system to (re)adapt to changing environments (images), "moving" images may be obtained, either by recording the consecutive pheromone fields — like the ones in Figure 1, for instance — or the position of the ants in each iteration; videos and "living drawings" are just some of the additional potentialities of the artificial swarm as creative tool. In addition, by relating evaporation to a kind of "forgetting" process, we engage in the analogy between the model and some neurological phenomena, which is addressed in later paragraphs.

A simulated scheme of natural selection was added to the system [12], increasing its abilities to deal with image processing problems, and enhancing the speed of reaction to changing environments. To each is assigned a certain amount of energy that is decreased each iteration, by a value that depends on its position in the environment – see equation 4. When the energy reaches zero, the ant "dies".

$$e(t) = e(t-1) - \alpha + \alpha(\Delta_{gl} / \max \Delta_{gl}) \tag{4}$$

On the other hand, new ants may be created when two ants are in neighboring cells. The probability of generating a new ant depends on the number of ants in that region (uninhabited areas and highly populated surroundings diminish the reproduction probability) and on the environment (when two ants meet in a region with contrast, the probability of generating a child is higher).

$$P_r = W(n)(\mu + \frac{(1-\mu)\Delta_{gl}}{\max \Delta_{gl}}) \tag{5}$$

Fig. 3: A black-and-white photo (Carlos M. Fernandes, The Broom, Budapest, 2003) and the pheromone map that emerges from the interaction of the swarm with the image (100 iterations; 960x960 pixels; parameters as in [11].)

where $W(n)$ is: $W(0) = W(8)$ =0; $W(4)$ = 1; $W(5)$ = $W(3)$ = 0.75; $W(6)$ = $W(2)$ = 0.5; $W(7)$ = $W(1)$ = 0.25. As stated above, these values tend to favor crossover between ants standing in moderately populated regions of the lattice. Please note that the second term in the product is very similar to the pheromone deposition rate T. The constant μ assures that even ants moving in homogeneous regions may have some chance to reproduce (if μ is set to 0, ants moving between regions with Δ_{gl} = 0 do not have a chance to reproduce). In terms of pheromone fields/images, the reproduction process, by dynamically varying the swarm's population size, also eliminated wandering ants that added some noise to the pheromone maps created by the previous version of the model.

All the figures shown in this paper were obtained with the final version of the swarm, with ants dying and reproducing depending on their location in the field. This mechanism proved to be essential for the highly adaptive and self-regulated behavior of the artificial ant colony.

---

[1] The term stigmergia was introduced by Pierre-Paul Grassé in 1959 [13] and describes indirect communication trough the environment.

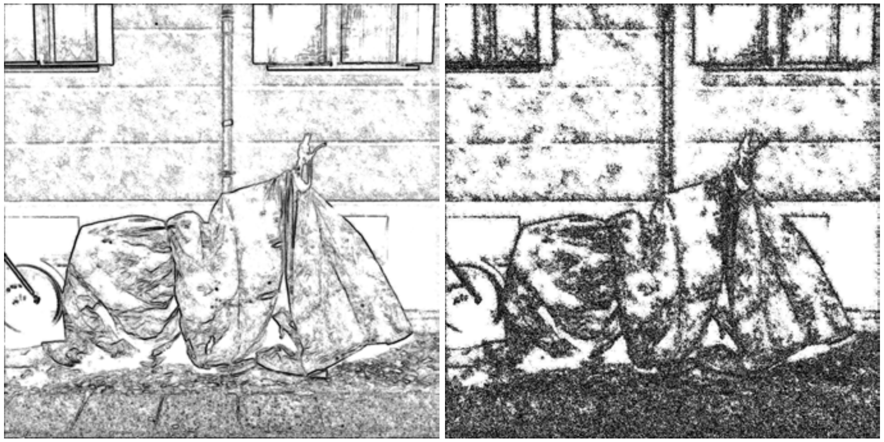


Fig. 4: A pheromone map (left) and a “snapshot” of the ants’ position (right) on the environment at iteration 100 (960x960 pixels; parameters as in [16])

This system exhibits what it is now called swarm intelligence [14, 15]. Swarm intelligence is the property of a system whereby the collective behaviors of simple entities interacting locally with their environment cause coherent global patterns to emerge; or, putting it another way, is the property of a system where self-organization arises from local interactions at the microscopic level and from the interaction between the system and its environment. In this case, ants interact locally, self-organize, and the swarm ends up perceiving the environment as whole.

An example of a pheromone map that appears after 100 iterations of the swarm on top of a black-and-white image with 960×960 pixels is shown in Figure 3. The edges of the image are detected — by the swarm — by pheromone (darker dots/lines refer to pixels with more pheromone). Figure shows the pheromone map and also the ants’ position on the environment, after the swarm had evolved for 100 iterations on another black-and-white photograph.

Please note the distinct characteristics of the two drawings (as a matter of fact, the right-hand side of Figure is a “drawing of the drawers”). The way the pheromone maps emerge from a plain pheromone field can be visualized in Figure 1. Gradually, the pheromone around high contrast regions becomes more intense. The role of evaporation is also clear in this example: pheromone in homogeneous areas gradually disappears, after being deposited during the initial exploratory stage. Those pictures illustrate the emergence of global perception from the local interaction of simple entities.

This camera obscura for ants was inspired by a model developed by the scientists Dante Chialvo and Mark Millonas in its simplest form [10], and later adapted by Vitorino Ramos and Filipe Almeida in order to evolve over gray-scale images [11]. The swarm was projected to deal with image processing problems, like segmentation and edge detection. Later, its abilities to evolve over dynamical landscapes shaped by mathematical functions were investigated [16]. Recently, the original system was again adapted in order to solve classification and clustering problems [17]. The model belongs to the class of Ant Algorithms [15], which is a computational paradigm inspired by the behavior of ants in nature. These simple entities that give rise to complex collective behavior (for a comprehensive study on insect societies and their complex behavior, see [18]) inspired not only new algorithms, but may also have induced progresses in Complexity studies and opened new horizons for Neurosciences. Chialvo and Milonas argued that an analogy (other than metaphorical) between the behavior of social insects and the self-organization of neurons might exist. Douglas Hofstadter, in his book Gödel, Escher e Bach, an Eternal Golden Braid [19], had already referred such idea:

> *(. . .) There is some degree of communication among the ants, just enough to keep them from wandering off completely at random. By this minimal communication they can remind each other that they are not alone but are cooperating with teammates. It takes a large number of ants, all reinforcing each other this way, to sustain any activity – such as trail building – for any length of time. Now my very hazy understanding of the operation of brain leads me to believe that something similar pertains to the firing of neurons (. . .)*

Following these ideas, we also designate the pheromone maps as cognitive maps. From now on, the possibilities are endless. We shall stick to the aesthetical outcome of the swarm being applied to gray-scale images and call it Pherographia: drawing by pheromones. Ants reinforce the "lines" by depositing more pheromone – like the chemical developer enhances the exposed silver –, while evaporation eliminates that pheromone that is no longer useful in the process of self-organization – like the fixer removes unexposed silver. Grain, in a film, appears as the result of the aggregation of silver salts when developing time is increased; the lines in this camera obscura for ants are enhanced by the constant reinforcement of pheromone over desired regions – as grain emerges from "reinforcement" of silver clusters, created by a longer developing time.

Pherographia is a rather naïve approach to drawing. There are no shadows or highlights, only lines delimiting the main areas of the image (although some detail emerges in some regions). The ants' drawings sometimes resemble other edge detection methods, but we still feel, when looking at the images, to be facing a children's sketch or some neo-Palaelolithic kind of representation of reality. In that sense, Pherographia departs from Photographia. Solarization, a photographic process popularized by such artists as Man Ray (1890-1976) and László Moholy-Nagy (1895-1946), comes to mind when looking at pherographic images. Due to the discontinuities imposed by pheromone trails, pherographic representations of images that hold rich tonal gradations may also resemble cloisonnism (if one mentally fills the blank regions with colors). Perhaps the most notable artist that engaged in such style was the post-impressionist painter Paul Gauguin (1848-1903), who was influenced by Japanese Ukiyo-e prints. As stated by Roy R. Behrens [22]:

> *(. . .) There is a persuasive resemblance between gestalt principles and the Japanese-inspired aesthetics (. . .)*

Gestalt principles also show some resemblances with Swarm Intelligence studies. Both aim at understanding how local perceptions become organized into wholes, and this it is precisely what happens in the ant system discussed in this paper: the restricted perception of individual ants gives rises to a global perception of the environment. A braid appears to arise that embraces all these concepts.

Pherographia and Image Processing

It is unquestionable that these pherographs are somewhat close to the results provided by some image processing software tools, but a careful look reveals characteristics of its own. For instance, in Figure 5 we can compare the resulting pheromone map of the well-known Lena1 image with the outcome of the "find edges" Photoshop tool, and the Canny [20] and Sobel [21] algorithms for edge detection (created in a Matlab toolbox with the suggested settings). Sobel and Canny attain well defined and minimalist graphic styled drawings. Photoshop's "find edges" creates a more intricate drawing, with some shadows over the sharp lines that cover high contrast regions. Pherographic version of Lena is rougher, with some blurred lines and with what seems to be noise but it is actually the detection of some minor details of the image.

But even if the swarm evolves drawings that somehow resemble the results attained by deterministic methods, its own nature — self-organized, evolutionary and stigmergic — is important enough to place it not only in the debate on the Artificial Life hypothesis, but also in some contemporary art trends (those that focus mainly on the means, the creator, and their relationship with the public, rather than the aesthetical outcome of the creative process). In addition, the swarm model holds some traits that are not present in deterministic methods. The intermediary steps, as stated above and shown in some pictures, allow us to create other compositions besides the static pherographs made of artificial pheromone. These "snapshots" of the system may be also gathered in videos. Finally, the ants' position, when recorded in a video, result in a kind of living drawing: even after the population converges to the desired regions, a high activity remains around the edges, and sometimes we fell as if we are observing pictures that are "alive". Some experiments with videos are on the way, but for now, the ant system is being used as a drawing tool that also explores some concepts that arise in contemporary art, photography and culture. Vernacular photography and recycling are some those issues that we are working with. The site www.pherographia.com was designed to relate our ongoing experiments with Pherographia.

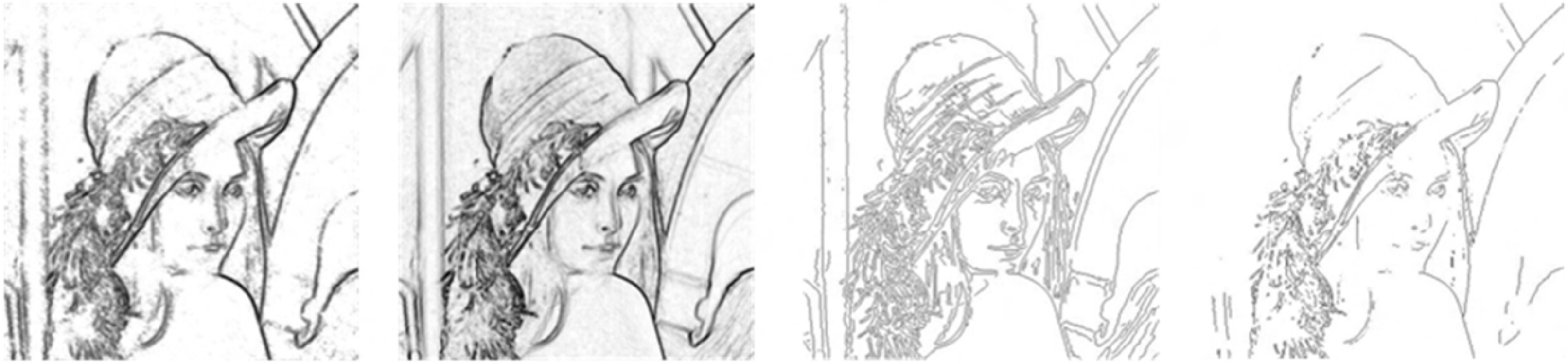


Fig. 5: Edge detection techniques: 1) Swarm (pherograph); 2) Photoshop's find edges tool; 3) Canny; 4) Sobel.

Conclusions

In this paper, we analyzed an image processing tool, based on an artificial life model, that may be regarded (metaphorically or not) as a modern version of the old camera obscura, the apparatus that was first used as canvas for the artist's pencil, and later, with the invention of Photography, became the camera that records the lines drawn by the *pencil of nature*. The *camera obscura* for ants evolves monochromatic drawings out of artificial pheromone fields and ants' position in the environment. In order to name the results attained by representing the pheromone fields, we coined the expression *Pherographia*: drawing by pheromones. The system is also able to react to changing environments and self-adapt to new images, and this ability gives us the opportunity to generate videos that illustrate not only the emergence of global perception but also the swarm's capacity to "forget" previous images and create new cognitive maps. Presently, a system that generates coloured pherographs is being developed.

Bibliography

[1] J. Holland, Adaptation in Natural and Artificial System, Ann Arbor, MI: The University of Michigan Press, 1975.

[2] J. von Neumann, A. W. Burks, Theory of self-reproducing automata (Urbana, IL: The University of Illinois Press, 1966).

[3] S. Wolfram, A New Kind of Science (Champaign, IL: Wolfram Media, Inc., 2002)

[4] M. Cook, Universality in Elementary Cellular Automata, Complex Systems Vol. 15, No.1, pp. 1-40 (2004).

[5] B. Mandelbrot, Les Objects Fractals (Paris: Flammarion, 1975)

[6] F. Pessoa, Poesias de Álvaro de Campos, a Obra Poética, Rio de Janeiro: Editora Nova Aguilar, 1995.

[7] W. H. F. Talbot, The Pencil of Nature, n.p., 1844-46.

[8] C. Baudelaire, Salon de 1859, Révue Française, Paris, June 10-July 20, 1859.

[9] J. Mayne, Charles Baudelaire, The Mirror of Art, London: Phaidon Press Limited, 1955.

[10] D. Chialvo, M. Milonas, How Swarms Build Cognitive Maps, Luc Steels (Ed.), The Biology and Technology of Intelligent Autonomous Agents, No. 144, NATO ASI Series, 1995, pp. 439-450.

[11] V. Ramos, F. Almeida, Artificial Ant Colonies in Digital Image Habitats A Mass Behaviour Effect Study on Pattern Recognition", Marco Dorigo, Martin Middendoff and Thomas Suetzle (Eds.), Proceedings 2nd International Workshop on Ant Algorithms, 2000, pp. 113-116.

[12] C. M. Fernandes, V. Ramos, A. C. Rosa, Self-Regulated Artificial Ant Colonies on Digital Image Habitats, International Journal of Lateral Computing Vol. 2, No. 1, 2005, pp. 1-8.

[13] P.-P. Grassé, La reconstrucion du nid et les coordinations interindividuelles chez bellicositermes et cubitermes sp. La théorie de la stigmergie: Essai d'interpretation du comportement des termites constructeurs, Insectes Societies, No. 6, 1959, pp. 41-80.

[14] J. Kennedy, R. C. Eberhart, Swarm Intelligence, San Francisco, CA: Morgan Kaufmann, 2001.

[15] E. Bonabeau, M. Dorigo, G. Theraulaz, Swarm Intelligence: From Natural to Artificial Systems, New York, NY: Oxford University Press, 1999.

[16] C. M. Fernandes, V. Ramos, A. C. Rosa, Varying the Population Size of Artificial Foraging Swarms on Time Varying Landscapes, Artificial Neural Networks: Biological Inspirations, Proceedings of 15th International Conference on Artificial Neural Networks, LNCS series, Vol. 3696, Part I, Springer-Verlag, 2005, pp. 311-316.

[17] A. Mora, C. M. Fernandes, J.-J. Merelo, V. Ramos, J. L. J. Laredo, KohonAnts. A Self-Organizing Ant Algorithm for Clustering and Pattern Classification, Proceedings of the XI Artificial Life Conference, 2008, pp. 428-435.

[18] E. O. Wilson, Insect Societies, Cambridge: Belknam Press, 1971.

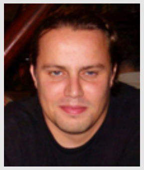
[19] D. Hofstadter, Gödel, Escher e Bach, an Eternal Golden Braid, New York, NY: Basic Books, 1979.

[20] J. Canny, A Computational Approach to Edge Detection, IEEE Transactions on Pattern Analysis and Machine Intelligence, No. 8, 1986, pp. 679-714.

[21] I. Sobel, G. Feldman, A 3x3 Isotropic Gradient Operator for Image Processing, presented at a talk at the Stanford Artificial Project, unpublished but often cited, 1968.

[22] R. R. Behrens, Art, Design and Gestalt Theory, Leonardo Vol. 39, No. 4, 1998, pp. 299-303.

About the authors



Carlos M. Fernandes was born in Luanda (Angola) in 1973 and lives between Lisbon (Portugal) and Granada (Spain). He graduated (Technical University of Lisbon, 1998) in Electrotechnics Engineering and owns a master degree in the same field since 2002 (Technical University of Lisbon). He is currently pursuing a Ph.D. on Bio-inspired Computing applied to Dynamic Optimization. He works between the Evolutionary and Biomedical Engineering Laboratory of the Technical University of Lisbon and the Department of Computers Architecture of the University of Granada. He is also a photographer and since 1996 he has been involved in several group and solo exhibitions. Recently, he exhibited the work Timor Mortis Conturbat Me at a gallery in Lisbon, making use of the creative potentialities of the Pherographia and finally blending his main interests: Science and Photography.

Email: cfernandes@laseeb.org

GEVA: Grammatical Evolution in Java

Michael O'Neill, Erik Hemberg, Conor Gilligan, Eliott Bartley, James McDermott, & Anthony Brabazon
Natural Computing Research & Applications Group, University College Dublin, Ireland

We are delighted to announce the release of **GEVA** [1], an open source software implementation of Grammatical Evolution (GE) in Java. Grammatical Evolution in Java (GEVA) was developed at UCD's Natural Computing Research & Applications group (http://ncra.ucd.ie).

What is Grammatical Evolution?

Grammatical Evolution (GE) (e.g., [2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14]) is a grammar-based form of Genetic Programming [15]. It marries principles from molecular biology to the representational power of formal grammars. GE's rich modularity gives a unique flexibility, making it possible to use alternative search strategies, whether evolutionary, or other heuristic, be it stochastic or deterministic, and to radically change its behaviour by merely changing the grammar supplied. As a grammar is used to describe the structures that are generated by GE, it is trivial to modify the output structures by simply editing the plain text grammar. This is one of the main advantages that makes the GE approach so attractive. The genotype-phenotype mapping also means that instead of operating exclusively on solution trees, as in standard GP, GE allows search operators to be performed on the genotype (e.g., integer or binary chromosomes), in addition to partially derived phenotypes, and the fully formed phenotypic derivation trees themselves.

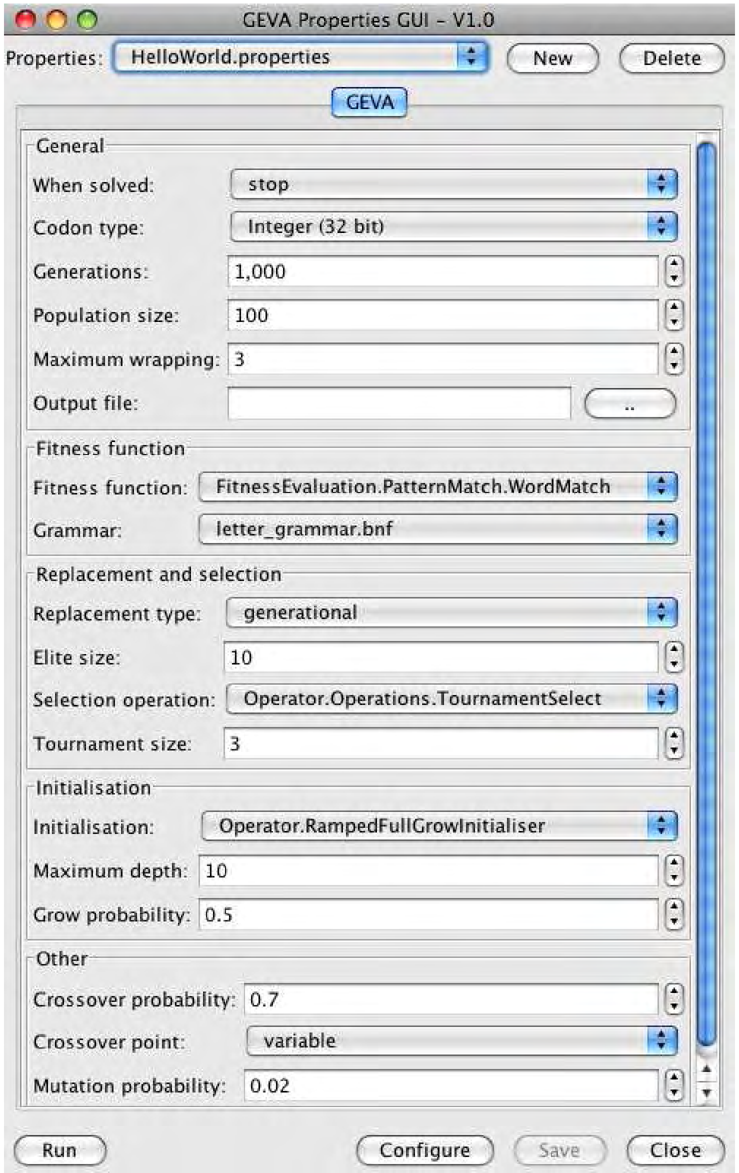
Grammatical Evolution in Java

GEVA has been released under GNU GPL version 3, and uses Java 1.5 and greater. As well as providing the characteristic genotype-phenotype mapper of GE, an evolutionary search engine and a GUI are also provided.

GEVA comes out-of-the-box with a number of demonstration problems that can be easily switched between from the GUI or command line. Sample problems include simple String Pattern Matching, an LSystem generator, the Paint problem, and a number of classic Genetic Programming problems such as an example of Symbolic Regression, the Santa Fe ant trail and Even Five Parity.

A screenshot of the default GUI screen showing settings for the pattern matching problem can be seen in Fig. 1. The goal of the problem is to rediscover the string "*geva*". Simple graphing support is also provided, which allows the user to observe various attributes of the population live during the course of a run. The resulting graphs can then be saved for later use. Attributes that can be plotted include the best fitness, the average fitness (with error bars), the number of invalid (incompletely mapped individuals), the average number of codons in each individual, and the average number of expressed codons in the population. For an example see Fig. 2.

A number of tutorials have been developed to help the novice user get up to speed: from running the software out-of-the-box, to using the command-line parameters, writing your own grammars and fitness functions, to developing your own search engine. These include a tutorial describing a bonus demo problem, Battleship, that the interested user can add to a GEVA installation.



GEVA Properties GUI - V1.0

Properties: HelloWorld.properties

General
When solved: stop
Codon type: Integer (32 bit)
Generations: 1,000
Population size: 100
Maximum wrapping: 3
Output file:

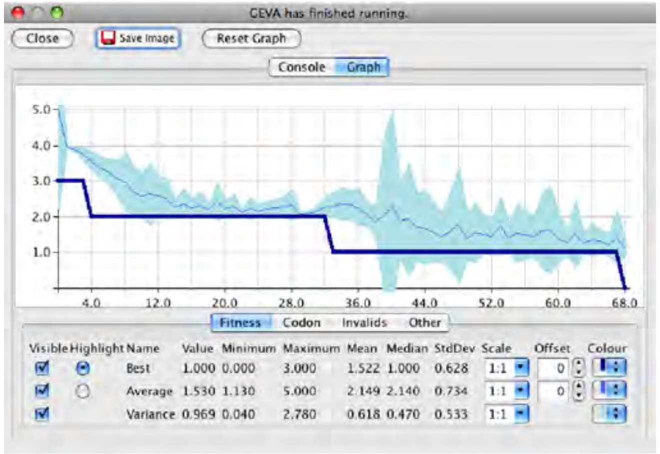
Fitness function
Fitness function: FitnessEvaluation.PatternMatch.WordMatch
Grammar: letter_grammar.bnf

Replacement and selection
Replacement type: generational
Elite size: 10
Selection operation: Operator.Operations.TournamentSelect
Tournament size: 3

Initialisation
Initialisation: Operator.RampedFullGrowInitialiser
Maximum depth: 10
Grow probability: 0.5

Other
Crossover probability: 0.7
Crossover point: variable
Mutation probability: 0.02

Fig. 1: A screenshot of the GEVA GUI. When opened for the first time it adopts the parameter settings for the pattern matching example problem where the goal is to rediscover the string "*geva*".



GEVA has finished running.

Visible	Highlight	Name	Value	Minimum	Maximum	Mean	Median	StdDev	Scale	Offset	Colour
☑	●	Best	1.000	0.000	3.000	1.522	1.000	0.628	1:1	0	
☑	○	Average	1.530	1.130	5.000	2.149	2.140	0.734	1:1	0	
☑		Variance	0.969	0.040	2.780	0.618	0.470	0.533	1:1		

Fig. 2: An example graph produced while running GEVA on the pattern matching example problem. Displayed are the best fitness and average fitness with error bars. It is also possible to observe the average codon length of individuals, the average expressed codon length and the number of invalid individuals in the population.

Design Overview

GEVA takes advantage of GE's modular structure as outlined in Fig. 3. This allows us to create a framework in which any search engine algorithm can be used to generate the genotypes (the `GEChromosome` class) that are used to direct the GE Mapper's use of the Grammar during the development of the *output* solution. In recent years this approach has included the adoption of a Particle Swarm algorithm and Differential Evolution as alternative search engines resulting in Grammatical Swarm [14] and Grammatical Differential Evolution [16] variants. GEVA facilitates the adoption of alternative search engines through the provision of an `Algorithm` interface. This will work correctly as long as a GE `Mapper` object is provided with a legal `GEChromosome` object, so any alternative algorithm must ensure to map its search engine's individual representation to a `GEChromosome` to generate an output solution. In this first release a standard Genetic Algorithm engine is provided with plans to add alternative engines in future releases. The current version uses individuals with (32-bit) integer codon values and adopts a corresponding integer mutation operator.

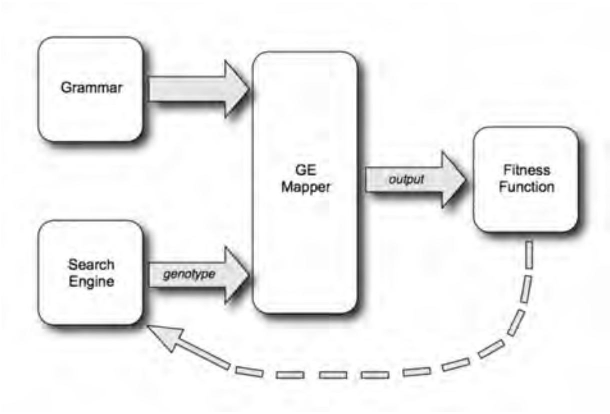


Fig. 3: Modular components of Grammatical Evolution.

Grammars are made available to GEVA through plain text files that adopt BNF notation. A simple parser is provided which handles standard BNF and can also recognise special symbols including `GECodonValue`, which returns the current codon's numerical value as a terminal symbol to the developing output phenotype sentence.

Simply by altering the contents of a BNF text file you can radically change the output generated by GEVA. A number of studies have illustrated this flexiblity: for example, grammars have been used to represent a diverse array of structures including binary strings, code in various programming languages (e.g., C, Scheme, Slang, Postscript), music, financial trading rules, 3D surfaces, and even grammars themselves (examples include [18, 4, 19, 20, 21, 22, 23, 24]). A number of demonstration grammars are provided in the example problems and are available through the GUI and from the command line.

How to find out more

GEVA is available for download from the UCD NCRA group website (http://ncra.ucd.ie/geva) or http://www.grammatical-evolution.org. Included in the release are instructions on how to run GEVA out-of-the-box, and more detailed tutorials for those who wish to modify the software for new purposes.

We also welcome feedback on the software as we plan to actively maintain the code, releasing new versions as features are added. A GEVA Google group has been set up to facilitate communication amongst the GEVA community [25]. We hope that GEVA will be a useful resource for the EC community and beyond.

Acknowledgments

We would like to thank past and present members of the UCD Natural Computing Research & Applications group, especially Tiberiu Simu, Jonathan Huggoson, and Jeff Wright for initial testing and some additions to the code. GEVA was beta-tested by the students of *COMP30290 Natural Computing*, an elective offered by UCD School of Computer Science & Informatics [27], from September to November 2007 and 2008, and by Patrick Middleburgh who developed the seed of the LSystem interface as part of his Final Year Computer Science project at UCD. We would also like to thank Miguel Nicolau for many interesting discussions that helped inform the design of GEVA, in particular the genotype-phenotype mapper which draws upon the design adopted in Miguel's libGE C++ library. This publication has emanated from research conducted with the financial support of Science Foundation Ireland under Grant No. 06/RFP/CMS042, and UCD Research Seed Funding.

Bibliography

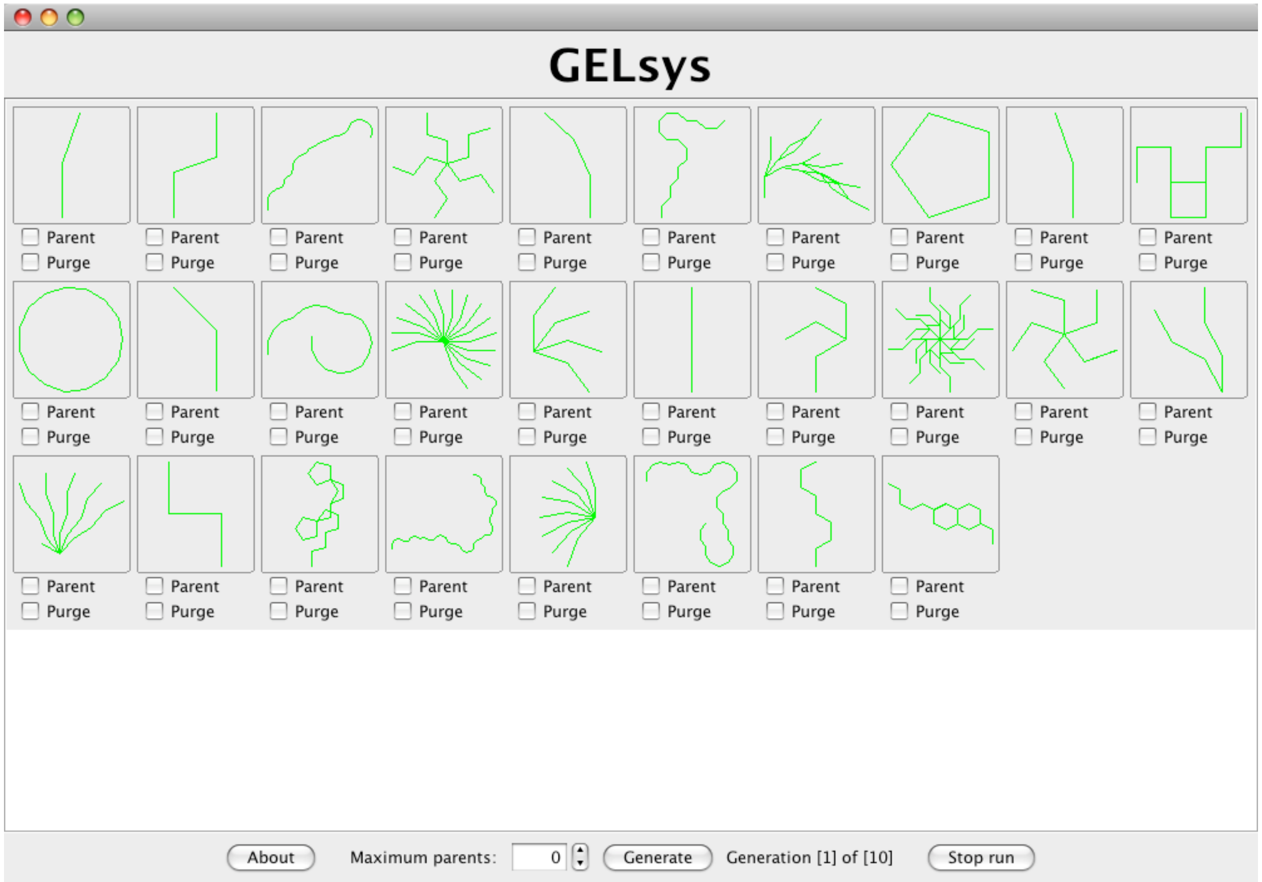
[1] GEVA - Grammatical Evolution in Java. (2008). http://ncra.ucd.ie/geva/

[2] O'Neill, M., Ryan, C. (2003). *Grammatical Evolution: Evolutionary Automatic Programming in an Arbitrary Language*. Kluwer Academic Publishers.

[3] O'Neill, M. (2001). Automatic Programming in an Arbitrary Language: Evolving Programs in Grammatical Evolution. PhD thesis, University of Limerick.

[4] Brabazon, A., O'Neill, M. (2006). Biologically Inspired Algorithms for Financial Modelling. Springer.

[5] Dempsey I. (2007). Grammatical Evolution in Dynamic Environments, PhD Thesis, University College Dublin.



Maximum parents: 0

Generation [1] of [10]

Fig. 4: GEVA includes an interactive form of GE as an example problem that allows users to generate interesting LSystems. A similar approach was used to evolve the logo for the Natural Computing Research & Applications Group at UCD (see Fig. 5).

[6] Dempsey I., O'Neill M., Brabazon A. (2007). Constant Creation with Grammatical Evolution. International Journal of Innovative Computing and Applications, pp.23-38 Vol.1 No.1

[7] O'Neill, M., Ryan, C. (2001). Grammatical Evolution, *IEEE Trans. Evolutionary Computation*. 2001.

[8] Ryan, C., Collins, J.J., O'Neill, M. (1998). Grammatical Evolution: Evolving Programs for an Arbitrary Language. *Proc. of the First European Workshop on GP*, pp. 83-95, Springer-Verlag.

[9] O'Neill, M., Brabazon, A. (2005). Recent Adventures in Grammatical Evolution. In Proceedings of CMS 2005 Computer Methods and Systems. Vol.1, pp. 245-253, November 2005, Krakow, Poland.

[10] Hemberg E., Gilligan C., O'Neill M., Brabazon A. (2007). A Grammatical Genetic Programming Approach to Modularity in Genetic Algorithms. In Ebner M., O'Neill M., Ekart A., Vanneschi L., and Esparcia Alcazar A. (eds.) *EuroGP (Tenth European Conference on Genetic Programming)*, Valencia, Spain. Springer.

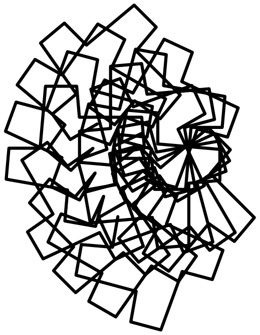


Fig. 5: The UCD NCRA group logo was evolved using Grammatical Evolution. See [17] for more details.

[11] Brabazon A., O'Neill M., Dempsey I. (2008). An Introduction to Evolutionary Computation in Finance. IEEE Computational Intelligence Magazine, November, pp. 42-55, IEEE Press.

[12] Cleary R. (2005). Extending Grammatical Evolution with Attribute Grammars: An Application to Knapsack Problems, Masters Thesis, University of Limerick. Available from http://ncra.ucd.ie/downloads/pub/thesisExtGEwithAGs-CRC.pdf

[13] Amarteifio S. (2005). Interpreting a Genotype-Phenotype Map with Rich Representations in XMLGE, Masters Thesis, University of Limerick. Available from http://ncra.ucd.ie/downloads/pub/SaoirseMScThesis.pdf

[14] O'Neill, M. and Brabazon, A. (2006). Grammatical Swarm: The Generation of Programs by Social Programming. Natural Computing, pp. 443-462 Vol.5 No.4

[15] Poli, R., Langdon, W.B., McPhee, N.F. (2008). A field guide to genetic programming. Published via `http://lulu.com` and freely available at `http://www.gp-field-guide.org.uk`.

[16] O'Neill, M. and Brabazon, A. (2006). Grammatical Differential Evolution. International Conference on Artificial Intelligence (ICAI'06), pp. 231-236 CSEA Press Las Vegas, Nevada.

[17] O'Neill, M., Brabazon, A. (2008). Evolving a Logo Design using Lindenmayer Systems, Postscript and Grammatical Evolution. In Proceedings of the IEEE World Congress on Evolutionary Computation, pp. 3788-3794. Hong Kong, 1-6 June 2008. IEEE Press.

[18] O'Reilly U-M., Hemberg M. (2007). Integrating generative growth and evolutionary computation for form exploration. Genetic Programming and Evolvable Machines, 8(2), pp.163-186.

[19] Moore J.M., Hahn L.W. (2003). Petri net modeling of high-order genetic systems using grammatical evolution. BioSystems, 72(1-2), pp.177-186.

[20] de la Puente A.O., Alfonso R.S., Moreno M.A. (2002). Automatic composition of music by means of grammatical evolution. Proceedings of the 2002 conference on APL, pp. 148-155, ACM Press.

[21] Karpuzcu U.R. (2005). Automatic Verilog Code Generation through Grammatical Evolution. In Proceedings of Genetic and Evolutionary Computation Conference (GECCO2005) workshop program, pp. 394-397, ACM Press.

[22] Tsoulos I.G., Lagaris I.E. (2006). Solving differential equations with genetic programming. Genetic Programming and Evolvable Machines, 7(1), pp. 33-54.

[23] Berarducci P., Jordan D., Martin D., Seitzer J. (2004). GEVOSH: Using Grammatical Evolution to Generate Hashing Functions. In Proceedings of Genetic and Evolutionary Computation Conference (GECCO2004) workshop program.

[24] Cebrian M., Alfonseca M., Ortega A. (2007). Automatic generation of benchmarks for plagiarism detection tools using grammatical evolution. GECCO '07: Proceedings of the 9th annual conference on Genetic and evolutionary computation, Vol. 2, pp. 2253-2253, ACM Press.

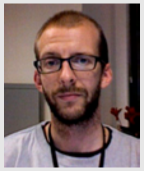
[25] GEVA user group. http://groups.google.com/group/geva
Email: geva@googlegroups.com

[26] http://www.grammatical-evolution.org

[27] O'Neill, M. (2007). COMP30290 Natural Computing.
http://ncra.ucd.ie/COMP30290/

About the authors

Michael O'Neill (m.oneill@ucd.ie) is with the School of Computer Science and Informatics at University College Dublin. He is co-founder and co-director of UCD's Natural Computing Research & Applications group, and has authored a number of books including Grammatical Evolution and Biologically Inspired Algorithms for Financial Modelling. (m.oneill@ucd.ie)



Erik Hemberg is a PhD student at UCD' s Natural Computing Research & Applications Group, in Dublin, Ireland.

Conor Gilligan is a PhD student at the UCD Natural Computing Research & Applications Group, Dublin, Ireland.

Eliott Bartley is an undergraduate studying Computer Science at UCD. He is a visiting researcher with the NCRA having recently undertaken a Summer Internship with the group.

James McDermott is a post-doctoral researcher in the NCRA. He obtained his PhD from the University of Limerick, with a thesis on interactive EC applied to sound synthesis. His research interests include representations for aesthetic EC and EC applications in music and design.

Anthony Brabazon is currently Head of Research in the School of Business at University College Dublin. He is also co-founder and co- director, of the Natural Computing Research and Applications Group at UCD. His primary research interests concern the development of natural computing theory and the application of natural computing algorithms in finance. He is a member of the IEEE Computational Finance and Economics Technical Committee.

Homepage: http://ncra.ucd.ie/geva/
Email: geva@googlegroups.com

Events Reports

International Conference on Evolvable Systems (ICES-2008)

Andrew Greensted, Department of Electronics, University of York

Overview

The field of biologically inspired hardware aims to provide solutions to a number of electronic system design challenges; increasing design complexity of electronic architectures, automation of architecture design; and system adaptability and reliability in dynamic environments. The International Conference on Evolvable Systems provides a forum in which these ideas can be presented and discussed.

ICES 2008, the 8th conference in the series so far, was organised by Gregory Hornby, Lukas Sekanina and Pauline Haddow, and held over the 21st-24th September in Prague, Czech Republic. This years conference demonstrated that the field is still developing and producing novel work. This year, 41 papers were accepted, 28 for oral presentation, 14 as posters. The conference attracted people from all over the world, with 70 registered delegates from 18 countries across 4 continents.

The conference program included tutorials, a discussion forum, paper presentations, a poster session and panel debate.

Presentations

As ICES covers a wide scope of bio-inspired electronics themes, there was a diverse range of sessions. The main sessions being Digital and Analogue Circuits, Neural Networks and Development. The addition of the last of these, is evidence of the growing interest in using developmental approaches as a solution to the scalability issues generally found in evolutionary design.

The presentations contained the normal collection of subtle variations and improvements on EA approaches, and the application of EAs to the design of well established circuits. There were also some examples of interesting new work.

ISCLEs: Importance Sampled Circuit Learning Ensembles for Trustworthy Analog Circuit Topology Synthesis, by Peng Gao, Trent McConaghy, Georges Gielen, is an interesting approach to overcoming industry's wariness of using evolved circuits. This system utilises proven analogue circuits as building blocks for evolving more complex circuits. The use of established sub-circuits endows the results with a greater level of trustworthiness.

Evolving Variability-tolerant CMOS Designs, by James Alfred Walker, James A. Hilder, Andy M. Tyrrell, highlighted that EA based circuit design is still a worthwhile research area when the results have real-world application. This work involves evolving relatively small circuits, so from the outset, this task is placed in a proven area of EA based circuit design. Despite the small circuit size, evolving nano-scale designs resistant to the intrinsic variability is not a simple task, yet this work has produced promising results.

The best paper award was given to an excellent demonstration of an evolutionary based solution to a real world problem. The winner was 'The Segmental-Transmission-Line: Its Design and Prototype Evaluation' by Moritoshi Yasunaga, Yoshiki Yamaguchi, Hiroshi Nakayama, Ikuo Yoshihara, Naoki Koizumi, and Jung H. Kim. This paper provides a solution to PCB signal integrity problems when operating in the GHz region. Using a GA to evolve the shape of a segmented transmission line, the signal waveform could be optimised at pre-set positions along a PCB track. The success of the approach was demonstrated with results measured from a fabricated evolved segmented transmission line.

Keynotes

The first keynote, given by Prof. Richard B. Fair (Duke University, Durham, NC, USA), presented recent progress in the field of lab-on-chip, in particular the use of digital microfluidics. Using the long known principle of the electrowetting, Prof. Fair explained how droplets of liquid can be moved and mixed.

The result is the automation of many different lab techniques at a small scale. This fascinating presentation, although not directly related to evolvable or bio-inspired hardware, provided great inspiration for possible new application areas.

Prof. Jordan Pollack's (Brandeis University, USA) keynote focused on the application of evolutionary techniques to the design of limbed robots. An interesting aspect of the keynote was an insight into the problems of evolutionary design. The void between simulation and realisation is not always easily crossed when evolving complex designs. Pollack suggested that for successful machine embryogenesis, developmental techniques that take into account the 'noise' of manufacture and environment are required; an approach that continuously optimises the assembly process in order to reach the goal of fully automated and unsupervised robot construction.

The final keynote, given by Prof. Jaroslav Flegr (Charles University in Prague, Czech Republic), gave a update of the current thinking on evolutionary theory. Highlighting possible flaws in Dawkins selfish-gene theory. It is interesting to consider that there is still more to learn about the theory of evolution, and that there remains, for the time being, an ongoing source of inspiration for future work.

Discussions

There were two opportunities for delegates to voice their feelings about the past, present and future of evolvable and bio-inspired electronics. A number of topics were discussed at the forum and panel debate, although there were a few main themes.

The perennial topic of what new application areas should be investigated by the community was discussed. Opinions included putting applications to one side and focusing on improving techniques and understanding; accepting that evolvable hardware is just another tool to be used when an application really requires it; start the search by looking for application areas with good quality simulators that can be embedded in a EA.

Another major theme was how to improve the widespread acceptance and utilisation of evolvable hardware. Related to this, why are evolved circuit design not more trusted by industry? Again, there were various views. More diversity in applications of evolvable hardware is required to demonstrate its usefulness. Solutions need to be evolved with real-world implementation in mind to avoid the requirement for non-standard fabrication techniques.

Socialising

The conference organisers put together a great social programme. Even though ICES 2008 was only a 4 day conference, the schedule included a reception dinner with local musicians, a visit to Troja palace, a boat trip on the Vltava river and the conference dinner at the Strahov monastery. As well as being great entertainment, these events gave everyone a good chance to meet up after the days presentations, discuss the days events and discuss ideas for future work.

Summary

In summary, ICES 2008 was a very informative, useful and fun conference. There was a good mixture of regular and new delegates, with representatives from both academia and industry. The papers and proceedings demonstrate that the field of evolutionary and bio-inspired electronics is still producing novel work and slowly maturing. Major conclusions from discussions were that there is a need to find and investigate new problem domains, perhaps by looking for reliable simulation tools and using them as a basis for an evolutionary algorithm. Also, more should be done to disseminate the field by more effectively pitching results and techniques to industry. Finally, many congratulations should be given to the conference organisers who put together and ran a great event.

ICES 2010

The next ICES conference is set for September 2010, located in York, England. The combination of rich history and a fantastic setting makes York a very attractive conference location. The 9th ICES promises to be a great event and will no doubt continue the excellent standard set at the Prague ICES conferences.

New Books

Natural Computing Series (www)

- Farooq, "Bee-Inspired Protocol Engineering - From Nature to Networks" (http://www.springer.com/978-3-540-85953-6)
- Hingston et al., "Design by Evolution - Advances in Evolutionary Design" (http://www.springer.com/978-3-540-74109-1)

Genetic and Evolutionary Computation Series (www)

- Riolo et al., "Genetic Programming Theory and Practice VI" (http://www.springer.com/978-0-387-87622-1)

Studies in Computational Intelligence (Engineering) (www)

- Kramer, "Self-Adaptive Heuristics for Evolutionary Computation" (http://www.springer.com/978-3-540-69280-5)
- Chen and Lim (Eds.), "Linkage in Evolutionary Computation" (http://www.springer.com/978-3-540-85067-0)
- Cotta and van Hemert (Eds.), "Recent Advances in Evolutionary Computation for Combinatorial Optimization" (http://www.springer.com/978-3-540-70806-3)
- Apolloni et al., "The Puzzle of Granular Computing" (http://www.springer.com/978-3-540-79863-7)
- Fulcher and Jain (Eds.), "Computational Intelligence: A Compendium" (http://www.springer.com/978-3-540-78292-6)
- Fink and Rothlauf (Eds.), "Advances in Computational Intelligence in Transport, Logistics, and Supply Chain Management" (http://www.springer.com/978-3-540-69024-5)

Lecture Notes in Computer Science (www)

- Branke et al., "Multiobjective Optimization - Interactive and Evolutionary Approaches", LNCS 5252 (http://www.springer.com/978-3-540-88907-6)
- Pezzulo et al., "The Challenge of Anticipation - A Unifying Framework for the Analysis and Design of Artificial Cognitive Systems", LNCS 5225 (http://www.springer.com/978-3-540-87701-1)
- Zurada et al. (Eds.), "Computational Intelligence: Research Frontiers", LNCS 5050 (http://www.springer.com/978-3-540-68858-7)
- Marinaro et al. (Eds.), "Dynamic Brain - from Neural Spikes to Behaviors", LNCS 5286 (http://www.springer.com/978-3-540-88852-9)
- Giacobini et al., "EvoWorkshops 2008", LNCS 4974 (http://www.springer.com/978-3-540-78760-0)
- O'Neill et al., "EuroGP 2008", LNCS 4971 (http://www.springer.com/978-3-540-78670-2)
- van Hemert and Cotta, "EvoCOP 2008", LNCS 4972 (http://www.springer.com/978-3-540-78603-0)
- Blesa Aguilera et al., "Hybrid Metaheuristics", HM 2008, LNCS 5296 (http://www.springer.com/978-3-540-88438-5)
- Marchiori and Moore, EvoBIO 2008, LNCS 4973 (http://www.springer.com/978-3-540-78756-3)
- Rudolph et al., PPSN 2008, LNCS 5199 (http://www.springer.com/978-3-540-87699-1)

Other Volumes

- Kordon, "Applying Computational Intelligence" (http://www.springer.com/978-3-540-69910-1)
- Floreano et al. (Eds.), "Flying Insects and Robots" (http://www.springer.com/978-3-540-89392-9)
- Rutkowski, "Computational Intelligence - Methods and Techniques" (http://www.springer.com/978-3-540-76287-4)
- Alba and Dorronsoro, "Cellular Genetic Algorithms" (http://www.springer.com/978-0-387-77609-5)
- Siciliano and Khatib (Eds.), "Springer Handbook of Robotics" (http://www.springer.com/978-3-540-23957-4)
- Gawiejnowicz, "Time-Dependent Scheduling" (http://www.springer.com/978-3-540-69445-8)
- Michiels et al., "Theoretical Aspects of Local Search" (http://www.springer.com/978-3-540-35853-4)
- Tautz, "The Buzz About Bees" (http://www.springer.com/978-3-540-78727-3)
- Würtz, "Organic Computing" (http://www.springer.com/978-3-540-77656-7)
- Barbieri, "The Codes of Life" (http://www.springer.com/978-1-4020-6339-8)
- Doerner et al., "Metaheuristics" (http://www.springer.com/978-0-387-71919-1)
- Marques de Sá, "Chance - The Life of Games & the Game of Life" (http://www.springer.com/978-3-540-74416-0)
- Williams, "Explorations in Quantum Computing", 2nd ed. (http://www.springer.com/978-1-84628-886-9)
- Mehlhorn and Sanders, "Algorithms and Data Structures" (http://www.springer.com/978-3-540-77977-3)

Calls and Calendar

March 2009

2009 IEEE Symposium Series on Computational Intelligence
March 30 - April 2, 2009, Nashville, TN, USA
Homepage: www.ieee-ssci.org

This international event promotes all aspects of the theory and applications of computational intelligence, by hosting 24 technical meetings in one location. Sponsored by the IEEE Computational Intelligence Society, this event will attract top researchers, professionals, and students from around the world. The registration, which will allow participants to attend any session in any technical meeting, will also include the complete set of the proceedings of all the meetings, coffee breaks, lunches, and the banquet. The event will be held in the magical town of Nashville, city of the country music.

Symposia And Workshops

- IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning (ADPRL 2009)
- IEEE Symposium on Computational Intelligence in Control and Automation (CICA 2009)
- IEEE Workshop on Robotic Intelligence in Informationally Structured Space (RiiSS 2009)
- IEEE Workshop on Evolving and Self-Developing Intelligent Systems (ESDIS 2009)
- IEEE Workshop on Evolvable and Adaptive Hardware (WEAH 2009)
- IEEE Swarm Intelligence Symposium (SIS 2009)
- IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB 2009)
- IEEE Symposium on Artificial Life (ALIFE 2009)
- IEEE Symposium on Multicriteria Decision-Making (MCDM 2009)
- IEEE Symposium on Computational Intelligence in Scheduling (CI-Sched 2009)
- IEEE Symposium on Computational Intelligence and Data Mining (CIDM 2009)
- IEEE Workshop on Hybrid Intelligent Systems (HIS 2009)
- IEEE Workshop on Organic Computing (OC 2009)
- IEEE Symposium on Intelligent Agents (IA 2009)
- IEEE Workshop on Memetic Algorithms (WOMA 2009)
- IEEE Symposium on Computational Intelligence in Cyber Security (CICS 2009)
- IEEE Symposium on Computational Intelligence in Vehicles and Vehicular Systems (CIVVS 2009)
- IEEE Workshop on Computational Intelligence in Aerospace Applications (CIAA 2009)
- IEEE Symposium on Computational Intelligence for Multimedia Signal and Vision Processing (CIMSVP 2009)
- IEEE Symposium on Computational Intelligence for Image Processing (CIIP 2009)
- IEEE Workshop on Computational Intelligence in Virtual Environments (CIVE 2009)
- IEEE Workshop on Computational Intelligence for Visual Intelligence (CIVI 2009)
- IEEE Workshop on Computational Intelligence in Biometrics: Theory, Algorithms, and Applications (CIB 2009)
- IEEE Symposium on Computational Intelligence for Financial Engineering (CIFEr 2009)

Please visit www.ieee-ssci.org for call for papers, guidelines, submission information, additional details, and up to the minute information.

Memetic Computing Journal:
Special Issue on Metaheuristics for Large Scale Data Mining

Submission deadline: March 31st, 2009

Guest Editors

- Jaume Bacardit
  School of Computer Science and School of Biosciences
  University of Nottingham
  jaume.bacardit@nottingham.ac.uk
- Xavier Llorà
  National Center for Supercomputing Applications
  University of Illinois at Urbana-Champaign
  xllora@illinois.edu

Aim and Scope

Data mining and knowledge discovery are crucial techniques across many scientific disciplines. Recent developments such as the Genome Project (and its successors) or the construction of the Large Hadron Collider have provided the scientific community with vast amounts of data. Metaheuristics and other evolutionary algorithms have been successfully applied to a large variety of data mining tasks. Competitive metaheuristic approaches are able to deal with rule, tree and prototype induction, neural networks synthesis, fuzzy logic learning, and kernel machines -to mention but a few. Moreover, the inherent parallel nature of some metaheuristics (e.g. evolutionary approaches, particle swarms, ant colonies, etc) makes them perfect candidates for approaching very large-scale data mining problems.

Although a number of recent techniques have applied these methods to complex data mining domains, we are still far from having a deep and principled understanding of how to scale them to datasets of terascale, petascale or even larger scale. In order to achieve and maintain a relevant role in large scale data mining, metaheuristics need, among other features, to have the capacity of processing vast amounts of data in a reasonable time frame, to use efficiently the unprecedented computer power available nowadays due to advances in high performance computing and to produce when possible- human understandable outputs.

Several research topics impinge on the applicability of metaheuristics for data mining techniques: (1) proper scalable learning paradigms and knowledge representations, (2) better understanding of the relationship between the learning paradigms/representations and the nature of the problems to be solved, (3) efficiency enhancement techniques, and (4) visualization tools that expose as much insight as possible to the domain experts based on the learned knowledge.

We would like to invite researchers to submit contributions on the area of large-scale data mining using metaheuristics. Potentially viable research themes are:

- Learning paradigms based on metaheuristics, evolutionary algorithms, learning classifier systems, particle swarm, ant colonies, tabu search, simulated annealing, etc
- Hybridization with other kinds of machine learning techniques including exact and approximation algorithms
- Knowledge representations for large-scale data mining
- Advanced techniques for enhanced prediction (classification, regression/function approximation, clustering, etc.) when dealing with large data sets
- Efficiency enhancement techniques
- Parallelization techniques
- Hardware acceleration techniques (vectorial instuctions, GPUs, etc.)
- Theoretical models of the scalability limits of the learning paradigms/representations
- Principled methodologies for experiment design (choosing methods, adjusting parameters, etc.)
- Explanatory power and visualization of generated solutions
- Data complexity analysis and measures
- Ensemble methods
- Online data mining and data streams
- Examples of real-world successful applications

Instructions for authors

Papers should have approximately 20 pages (but certainly not more than 24 pages). The papers must follow the format of the Memetic Computing journal. Papers should be submitted following the Memetic Computing journal guidelines and stating in the accompanying letter that the paper is being submitted to the special issue on Large Scale Data Mining.

Important dates

Manuscript submission:	March 31st, 2009
Notification of acceptance:	May 31st, 2009
Submission of camera-ready version:	July 31st, 2009

April 2009

2009 EVO*
www.evostar.org
tübingen::germany

Evostar 2009 - EuroGP, EvoCOP, EvoBIO and EvoWorkshops

April 15-17, 2009, Tübingen, Germany

Homepage: www.evostar.org

The EuroGP, EvoCOP and EvoBIO conferences and the workshops collectively entitled EvoWorkshops compose EVO*: Europe's premier co-located events in the field of Evolutionary Computing. Featuring the latest in theoretical and applied research, EVO* topics include recent genetic programming challenges, evolutionary and other meta-heuristic approaches for combinatorial optimisation, evolutionary algorithms, machine learning and data mining techniques in the biosciences, in numerical optimisation, in music and art domains, in image analysis and signal processing, in hardware optimisation and in a wide range of applications to scientific, industrial, financial and other real-world problems.

EuroGP

Twelfth European Conference on Genetic Programming: high quality papers are sought on topics strongly related to the evolution of computer programs, ranging from theoretical work to innovative applications.

EvoCOP

Ninth European Conference on Evolutionary Computation in Combinatorial Optimisation: practical and theoretical contributions are invited, related to evolutionary computation techniques and other meta-heuristics for solving combinatorial optimisation problems.

EvoBIO

Seventh European Conference on Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics: the emphasis is on evolutionary computation and other advanced techniques addressing important problems in molecular biology, proteomics, genomics and genetics, that have been implemented and tested in simulations and on real-life datasets.

EvoWorkshops

The twelve workshops which make up this event are focused on the use of Evolutionary Computation in different application areas:

- EvoCOMNET: Telecommunication networks and other parallel and distributed systems
- EvoENVIRONMENT: Environmental issues
- EvoFIN: Finance and economics
- EvoGAMES: Games
- EvoHOT: Design automation
- EvoIASP: Image analysis and signal processing
- EvoINTERACTION: Interactive evolution and humanized computational intelligence
- EvoMUSART: Music, sound, art and design
- EvoNUM: Continuous parameter optimisation

- EvoPHD: Graduate student workshop on evolutionary computation
- EvoSTOC: Stochastic and dynamic environments
- EvoTRANSLOG: Transportation and logistics

In 2009, the event will take place in Tübingen, a traditional university town in Baden-Württemberg, Germany, situated on a ridge between the Neckar and Ammer rivers in the southwest of the country, about 30 kms southwest of Stuttgart. EVO* 2009 will be hosted at Eberhard Karls University in Tübingen, founded in 1477 and one of the oldest universities in Germany. The website www.evostar.org offers information relevant to all events, including calls for papers, deadlines, organising committees, submission requirements, local information and a thorough view on the previous editions.

May 2009

1st International Symposium on Search Based Software Engineering

May 13-15, 2009, Cumberland Lodge, Windsor, UK

Homepage: www.ssbse.org

We are pleased to announce SSBSE 2009, the inaugural meeting of an annual symposium dedicated to Search Based Software Engineering (SBSE) held in cooperation with the IEEE. The symposium's objective is to build on the recent flourishing of interest in SBSE by not only creating a welcoming forum for discussion and dissemination, but also by establishing a regular event that will strengthen the rapidly growing international community.

The venue for SSBSE 2009 is Cumberland Lodge, a beautiful and historic royal residence in heart of Windsor Great Park with world-class conference facilities. The Lodge is close to London, and only a short taxi ride from Heathrow Airport, yet is surrounded by some of the finest parkland in the country.

The symposium program includes three keynote speakers who are internationally renowned leaders in their research fields:

- Enrique Alba, University of Malaga, Spain
- Lionel C. Briand, Simula Research Lab & University of Oslo, Norway
- David E. Goldberg, Illinois Genetic Algorithms Laboratory (IlliGAL), University of Illinois at Urbana-Champaign, USA

Organizing Committee

- General Chair: Mark Harman, King's College London, UK
- Program Co-Chairs: Massimiliano Di Penta (University of Sannio, Italy) and Simon Poulding (University of York, UK)
- PhD Student Track Chair: Myra B. Cohen, University of Nebraska, Lincoln, USA
- Publicity Chair: Per Kristian Lehre, University of Birmingham, UK
- Website: Paul Emberson, University of York, UK
- Sponsors: Berner & Mattner, Germany; Engineering and Physical Science Research Council (EPSRC), UK

2009 IEEE Congress on Evolutionary Computation (CEC 2009)

May 18-21, 2009, Trondheim, NORWAY

Homepage: www.cec-2009.org

The 2009 IEEE Congress on Evolutionary Computation (CEC 2009) will be at the Nova Conference Centre and Cinema, Trondheim, Norway during May 18-21, 2009. Sponsored by the IEEE Computational Intelligence Society, co-sponsored by the Evolutionary Programming Society and the Institution of Engineering and Technology, CEC 2009 continues the successful sequence of World-class events going back to 1999.

CEC 2009 will feature a world-class conference that will bring together researchers and practitioners in the field of evolutionary computation and computational intelligence from all around the globe. Technical exchanges within the research community will encompass keynote speeches, special sessions, tutorials, panel discussions as well as poster presentations. On top of these, participants will be treated to a series of social functions, receptions and networking sessions, which will serve as a vital channel to establish new connections and foster everlasting friendship among fellow researchers. The annual IEEE Congress on Evolutionary Computation (CEC) is one of the leading events in the area of evolutionary computation.

CEC covers all topics in evolutionary computation, including, but not limited to: Ant colony optimization, Artificial immune systems, Artificial life, Autonomous mental & behaviour development, Bioinformatics & bioengineering, Coevolution & collective behaviour, Cognitive systems & applications, Combinatorial & numerical optimization, Computational finance & economics, Constraint & uncertainty handling, Estimation of distribution algorithms, Evolutionary data mining, Evolutionary design, Evolutionary games, Evolvable hardware & software, Evolutionary intelligent agents, Evolutionary learning systems, Evolving neural networks & fuzzy systems, Molecular & quantum computing, Particle swarm intelligence, Representation & operators, etc.

More information can be found at: www.cec-2009.org.

For general inquiries, please contact General Chair Andy Tyrrell at amt@ohm.york.ac.uk.

June 2009

2009 World Summit on Genetic and Evolutionary Computation

June 12-14, 2009, Shanghai, China

Homepage: http://www.sigevo.org/gec-summit-2009

The 2009 World Summit on Genetic and Evolutionary Computation (2009 GEC Summit) will be held June 12-14, 2009, in Shanghai, China. It is sponsored and organized by ACM/SIGEVO, the Special Interest Group for Genetic and Evolutionary Computation, sponsor of the annual GECCO conferences, and will feature the latest research and demonstrated successes in this dynamic area, including new approaches and breakthrough applications to problems in biology, medicine, engineering design, agriculture, logistics, traffic, security, scheduling, military affairs, and other fields.

Topics (including, but not limited to): Genetic Algorithms, Genetic Programming, Evolution Strategies, Evolutionary Programming Coevolution, Learning Classifier Systems, Ant Colony Optimization, Swarm Intelligence/Particle Swarm, DNA-Based Evolutionary Computation, Interactive Computational Models, Evolutionary Multiobjective Optimization, Evolutionary Combinatorial Optimization, Evolutionary Scheduling and Routing, Simulated Annealing, Estimation of Distribution Algorithms, Tabu Search, Biological Applications, Medical Applications, Industry Applications, Agricultural Applications, Military and Security Applications, Artificial Life, Search-Based Software Engineering, Evolutionary Robotics, Nature-Inspired Computation, Swarm Intelligence Optimization Applications, Intelligent Control, Intelligent Management, Intelligent Information Processing, Multi-Agent Theory, Pattern Recognition, Web Intelligence, Intelligent Transportation Systems, and others.

Organizers

Executive Chairs:	Qidi Wu Tongji Univ., China David Goldberg, University of Illinois, USA
Advisory Committee Chairs	Ruwei Dai, Inst. of Automation, Chinese Academy of Science, China John Koza, Stanford Univ., USA
Co-Chair	Dongyuan Yang, Tongji Univ., China
General Chairs	Lihong Xu, Tongji Univ., China Erik D. Goodman, Michigan State Univ., USA
Program Committee Chairs	Guoliang Chen, Univ. of Sci. & Tech. of China Darrell Whitley, Colorado State Univ., USA
Co-Chair	Yongsheng Ding Donghua Univ., China
Local Arrangements Chairs	Wanggen Wan Shanghai Univ., China Mark Kotanchek, Evolved Analytics, USA
Co-Chair	Xiaoguang Yang Tongji Univ., China
Treasurer	Jie Chen Tongji Univ., China

GEC Summit is sponsored by the Association for Computing Machinery Special Interest Group for Genetic and Evolutionary Computation.

July 2009

GECCO 2009 - Genetic and Evolutionary Computation Conference

July 8-12, 2009, Montréal, Canada

Homepage: http://www.sigevo.org/gecco-2009

Author notification: March 11, 2009

Camera-ready: April 22, 2009

The Genetic and Evolutionary Computation Conference (GECCO-2009) will present the latest high-quality results in the growing field of genetic and evolutionary computation.

Topics include: genetic algorithms, genetic programming, evolution strategies, evolutionary programming, real-world applications, learning classifier systems and other genetics-based machine learning, evolvable hardware, artificial life, adaptive behavior, ant colony optimization, swarm intelligence, biological applications, evolutionary robotics, coevolution, artificial immune systems, and more.

Organizers

General Chair:	Franz Rothlauf
Editor-in-Chief:	Günther Raidl
Business Committee:	Wolfgang Banzhaf Erik Goodman Una-May O'Reilly
Publicity Chair:	Martin Pelikan
Workshops Chair:	Anna I. Esparcia
Competitions Chairs:	Pier Luca Lanzi
Tutorials Chair:	Martin V. Butz
Late Breaking Papers Chair:	TBA
Local Chair:	Christian Gagné
EC in Practice Chairs:	David Davis Jörn Mehnen
Graduate Student Workshop Chair:	Steve Gustafson
Undergraduate Student Workshop Chair:	Frank Moore Clare Bates Congdon Larry Merkle

Important Dates

Paper Submission Deadline	January 14, 2009
Decision Notification	March 11, 2009
Camera-ready Submission	April 22, 2009

Venue

Delta Centre-Ville hotel is located in the heart of downtown, where Old Montreal and new Montreal blend seamlessly, and adjacent to vibrant nightlife, boutique shops and eclectic cuisine. For more information on Delta Centre-Ville, please visit:

www.deltahotels.com/hotels/hotels.php?hotelId=35

Visiting GECCO-2009 will be a great opportunity to visit the famous Montreal Jazz Festival (July 2-12, 2009):

www.montrealjazzfest.com/Fijm2008/festival_en.aspx

More Information

Visit www.sigevo.org/gecco-2009 for information about electronic submission procedures, formatting details, student travel grants, the latest list of tutorials and workshop, late-breaking papers, and more.

For technical matters, contact Conference Chair Franz Rothlauf at rothlauf@uni-mainz.de. For conference administration matters contact Primary Support Staff at gecco-admin@tigerscience.com.

GECCO is sponsored by the Association for Computing Machinery Special Interest Group for Genetic and Evolutionary Computation.

September 2009

IEEE Symposium on Computational Intelligence and Games (CIG-2009)
September 7-10, 2009, Milan, Italy
Homepage: http://www.ieee-cig.org
Submission deadline: May 22nd, 2009

Aim and Scope

Games are an ideal domain to study computational intelligence methods. They provide cheap, competitive, dynamic, reproducible environments suitable for testing new search algorithms, pattern based evaluation methods or learning concepts. At the same time they are interesting to observe, fun to play, and very attractive to students. This symposium, sponsored by the IEEE Computational Intelligence Society aims to bring together leading researchers and practitioners from both academia and industry to discuss recent advances and explore future directions in this field.

Topics of interest include, but are not limited to:

- Learning in games
- Evolutionary Computation for games
- Neural-based approaches for games
- Fuzzy-based approaches for games
- Console and video games
- Character Development and Narrative
- Opponent modeling in games
- CI/AI-based game design
- Multi-agent and multi-strategy learning
- Comparative studies
- Applications of game theory
- Board and card games
- Economic or mathematical games
- Imperfect information and non-deterministic games
- Evasion (predator/prey) games
- Realistic games for simulation or training purposes
- Player satisfaction in games
- Games for mobile or digital platforms
- Games involving control of physical objects
- Games involving physical simulation

Conference Committee

General Chair:	Pier Luca Lanzi
Program Chair:	Sung-Bae Cho
Proceedings Chair:	Luigi Barone
Publicity Chair:	Julian Togelius
Competition Chair:	Simon Lucas
Sponsorship Chair:	Georgios N. Yannakakis
Local Chairs:	Nicola Gatti and Daniele Loiacono

Important Dates (tentative schedule)

Tutorial proposals:	15th April 2009
Paper submission:	22nd May 2009
Decision Notification:	26th June 2009
Camera-ready:	24th July 2009
Symposium:	7-10 September 2009

For more information please visit: http://www.ieee-cig.org

About the Newsletter

SIGEVOlution is the newsletter of SIGEVO, the ACM Special Interest Group on Genetic and Evolutionary Computation.

To join SIGEVO, please follow this link [WWW]

Contributing to SIGEVOlution

We solicit contributions in the following categories:

Art: Are you working with Evolutionary Art? We are always looking for nice evolutionary art for the cover page of the newsletter.

Short surveys and position papers: We invite short surveys and position papers in EC and EC related areas. We are also interested in applications of EC technologies that have solved interesting and important problems.

Software: Are you are a developer of an EC software and you wish to tell us about it? Then, send us a short summary or a short tutorial of your software.

Lost Gems: Did you read an interesting EC paper that, in your opinion, did not receive enough attention or should be rediscovered? Then send us a page about it.

Dissertations: We invite short summaries, around a page, of theses in EC-related areas that have been recently discussed and are available online.

Meetings Reports: Did you participate to an interesting EC-related event? Would you be willing to tell us about it? Then, send us a short summary, around half a page, about the event.

Forthcoming Events: If you have an EC event you wish to announce, this is the place.

News and Announcements: Is there anything you wish to announce? This is the place.

Letters: If you want to ask or to say something to SIGEVO members, please write us a letter!

Suggestions: If you have a suggestion about how to improve the newsletter, please send us an email.

Contributions will be reviewed by members of the newsletter board.

We accept contributions in $\LaTeX$, MS Word, and plain text.

Enquiries about submissions and contributions can be emailed to editor@sigevolution.org.

All the issues of SIGEVOlution are also available online at www.sigevolution.org.

Notice to Contributing Authors to SIG Newsletters

By submitting your article for distribution in the Special Interest Group publication, you hereby grant to ACM the following non-exclusive, perpetual, worldwide rights:

- to publish in print on condition of acceptance by the editor
- to digitize and post your article in the electronic version of this publication
- to include the article in the ACM Digital Library
- to allow users to copy and distribute the article for noncommercial, educational or research purposes

However, as a contributing author, you retain copyright to your article and ACM will make every effort to refer requests for commercial use directly to you.