EvoApplications 2016 programme
(revised draft at 23 march 2016)

Wednesday 30 March

1110 - 1300 EvoAPP 1 : Real world applications

chair : Giovanni Squillero

Talks are 20 mins including setup and questions

Evolving classification models for prediction of patient recruitment in multicentre clinical trials using grammatical evolution
Gilyana Borlikova, Michael Phillips, Louis Smith, Michael O'Neill
Successful and timely completion of prospective clinical trials depends on patient recruitment as patients are critical to delivery of the prospective trial data. There exists a pressing need to develop better tools/techniques to optimise patient recruitment in multicentre clinical trials. In this study Grammatical Evolution (GE) is used to evolve classification models to predict future patient enrolment performance of investigators/site to be selected for the conduct of the trial. Prediction accuracy of the evolved models is compared with results of a range of machine learning algorithms widely used for classification. The results suggest that GE is able to successfully induce classification models and analysis of these models can help in our understanding of the factors providing advanced indication of a trial sites' future performance.

Evolving Coverage Optimisation Functions for Heterogeneous Networks using Grammatical Genetic Programming
Michael Fenton, David Lynch, Stepan Kucera, Holger Claussen, Michael O'Neill
Heterogeneous Cellular Networks are multi-tiered cellular networks comprised of Macro Cells and Small Cells in which all cells occupy the same bandwidth. User Equipments greedily attach to whichever cell provides the best signal strength. While Macro Cells are invariant, the power and selection bias for each small cell can be increased or decreased (subject to pre-defined limits) such that more or fewer UEs attach to that cell. Setting optimal power and selection bias levels for small cells is key for good network performance. The application of Genetic Programming techniques has been proven to produce good results in the control of Heterogenous Networks. Expanding on previous works, this paper uses grammatical GP to evolve distributed control functions for Small Cells in order to vary their power and bias settings. The objective of these control functions is to evolve control functions that maximise a proportional fair utility of UE throughputs.

A Variable Local Search based Memetic Algorithm for the Load Balancing Problem in Cloud Computing
Nasser Sabar, Andy Song, Mengjie Zhang
Load balancing (LB) is an important and challenging optimisation problem in cloud computing. LB involves assigning a set of services into a set of machines for which the goal is to optimise machine usages. This study presents a memetic algorithm (MA) for the LB problem. MA is a hybrid method that combines the strength of population based evolutionary algorithms with local search. However the effectiveness of MA mainly depends on the local

search method chosen for MA. This is because local search methods perform differently for different instances and under different stages of search. In addition, invoking local search at every generation can be computationally expensive and compromise the exploration capacity of search. To address these issues, this study proposes a variable local search based MA in the context of LB problem. The proposed MA uses multiple local search mechanisms. Each one navigates a different area in search space using a different search mechanism which can leads to a different search path with distinct local optima. This will not only help the search to avoid being trap in a local optima point, but can also effectively deal with various landscape search charateristics and dynamic changes of the problem. In addition, a diversity indicator is adoped to control the local search processes to encourage solution diversity. Our MA method is evaluated on instances of the Google machine reassignment problem proposed for the ROADEF/EURO 2012 challenge. Compared with the state of the art methods, our method achieved the best performance on most of instances, showing the effectiveness of variable local search based MA for the Load Balancing problem.

Challenging Anti-virus through Evolutionary Malware Obfuscation
Marco Gaudesi, Andrea Marcelli, Ernesto Sanchez, Giovanni Squillero, Alberto Tonda
The use of anti-virus software has become something of an act of faith. A recent study showed that more than 80% of all personal computers have anti-virus software installed. However, the protection mechanisms in place are far less effective than users would expect. Malware analysis is a classical example of cat-and-mouse game: as new anti-virus techniques are developed, malware authors respond with new ones to thwart analysis. Every day, anti-virus companies analyze thousands of malware that has been collected through honeypots, hence they restrict the research to only already existing viruses. This article describes a novel method for malware obfuscation based an evolutionary opcode generator and a special ad-hoc packer. The results can be used by the security industry to test the ability of their system to react to malware mutations.

Simheuristics for the Multiobjective Nondeterministic Firefighter Problem in a Time-Constrained Setting
Krzysztof Michalak, Joshua D. Knowles
The firefighter problem (FFP) is a combinatorial problem requiring the allocation of `firefighters' to nodes in a graph in order to protect the nodes from fire (or other threat) spreading along the edges. In the original formulation the problem is deterministic: fire spreads from burning nodes to adjacent, unprotected nodes with certainty. In this paper a nondeterministic version of the FFP is introduced where fire spreads to unprotected nodes with a probability P_{sp} (lower than~1) per time step. To account for the stochastic nature of the problem the simheuristic approach is used in which a metaheuristic algorithm uses simulation to evaluate candidate solutions. Also, it is assumed that the optimization has to be performed in a limited amount of time available for computations in each time step. In this paper online and offline optimization using a multipopulation evolutionary algorithm is performed and the results are compared to various heuristics that determine how to place firefighters. Given the time-constrained nature of the problem we also investigate for how long to simulate the spread of fire when evaluating solutions produced by an evolutionary algorithm. Results generally indicate that the evolutionary algorithm proposed is effective for $P_{sp} \geq 0.7$, whereas for lower probabilities the heuristics are competitive suggesting that more work on hybrids is warranted.

1400 - 1550 EvoAPP2 : Image Analysis and Signal Processing

chairs : Stefano Cagnoni & Mengjie Zhang

Talks are 20 mins including setup and questions

Binary Tomography Reconstruction by Particle Aggregation
Mohammad Majid al-Rifaie, Tim Blackwell
This paper presents a novel reconstruction algorithm for binary tomography based on the movement of particles. Particle Aggregate Reconstruction Technique (PART) supposes that pixel values are particles, and that the particles can diffuse through the image, sticking together in regions of uniform pixel value known as aggregates. The algorithm is tested on four phantoms of varying sizes and numbers of forward projections and compared to a random search algorithm and to SART, a standard algebraic reconstruction method. PART, in this small study, is shown to be capable of zero error reconstruction and compares favourably with SART and random search.

Population Based Ant Colony Optimization for Reconstructing ECG Signals
Yih-Chun Cheng, Tom Hartmann, Pei-Yun Tsai, Martin Middendorf
A population based ant optimization algorithm (PACO) for reconstructing electrocardiogram (ECG) signals is proposed in this paper. In particular, the PACO algorithm is used to find a subset of nonzero positions of a sparse wavelet domain ECG signal vector which is used for the reconstruction of a signal. The proposed PACO algorithm uses a time window for xing certain decisions of the ants during the run of the algorithm. The optimization behaviour of the PACO is compared with two random search heuristics and several algorithms from the literature for ECG signal reconstruction. Experimental results are presented for ECG signals from the MIT-BIT Arrhythmia database. The results show that the proposed PACO reconstructs ECG signals very successfully.

Speaker Verification on Unbalanced Data with Genetic Programming
Roisin Loughran, Alexandros Agapitos, Ahmed Kattan, Anthony Brabazon, Michael O'Neill *(best paper candidate)*
Automatic Speaker Verification (ASV) is a highly unbalanced binary classification problem, in which any given speaker must be verified against everyone else. We apply Genetic programming (GP) to this problem with the aim of both prediction and inference. We examine the generalisation of evolved programs using a variety of fitness functions and data sampling techniques found in the literature. A significant difference between train and test performance, which can indicate overfitting, is found in the evolutionary runs of all to-be-verified speakers. Nevertheless, in all speakers, the best test performance attained is always superior than just merely predicting the majority class. We examine which features are used in good-generalising individuals. The findings can inform future applications of GP or other machine learning techniques to ASV about the suitability of feature-extraction techniques.

Bare-Bone Particle Swarm Optimisation for Simultaneously Discretising and Selecting Features For High-Dimensional Classification
Binh Tran, Mengjie Zhang, Bing Xue
Feature selection and discretisation have shown their effectiveness for data preprocessing especially for high-dimensional data with many irrelevant features. While feature selection selects only relevant features, feature discretisation finds a discrete representation of data

that contains enough information but ignoring some minor fluctuation. These techniques are usually applied in two stages, discretisation and then selection since many feature selection methods work only on discrete features. Most commonly used discretisation methods are univariate in which each feature is discretised independently; therefore, the feature selection stage may not work efficiently since information showing feature interaction may be destroyed in the discretisation process. In this study, we propose a new method called PSO-DFS using bare-bone particle swarm optimisation (BBPSO) for discretisation and feature selection in a single stage. The results on ten high-dimensional datasets show that PSO-DFS obtains a substantial dimensionality reduction for all datasets. The classification performance is significantly improved or at least maintained on nine out of ten datasets by using the transformed ``small'' data by PSO-DFS. Compared to applying the two-stage approach which uses PSO for feature selection on the discretised data, PSO-DFS achieves better performance on six datasets, and similar performance on three datasets with a much smaller number of features selected.

NSGA-II based Auto-Calibration of Automatic Number Plate Recognition Camera for Vehicle Speed Measurement

Patryk Filipiak, Bartlomiej Golenko, Cezary Dolega

This paper introduces an auto-calibration mechanism for an Automatic Number Plate Recognition camera dedicated to a vehicle speed measurement. A calibration task is formulated as a multi-objective optimization problem and solved with Non-dominated Sorting Genetic Algorithm. For simplicity a uniform motion profile of a majority of vehicles is assumed. The proposed speed estimation method is based on tracing licence plates quadrangles recognized on video frames. The results are compared with concurrent measurements performed with piezoelectric sensors.

1700 - 1850   EvoAPPS 3 : Evolutionary algorithms and meta-heuristics in stochastic and dynamic environments

chairs : Trung Thanh Nguyen & Michalis Mavrovouniotis

Talks are 20 mins including setup and questions

Direct Memory Schemes for Population-based Incremental Learning in Cyclically Changing Environments *(best paper candidate)*

Michalis Mavrovouniotis, Shengxiang Yang

The population-based incremental learning (PBIL) algorithm is a combination of evolutionary optimization and competitive learning. The integration of PBIL with associative memory schemes has been successfully applied to solve dynamic optimization problems (DOPs). The best sample together with its probability vector are stored and reused to generate the samples when an environmental change occurs. It is straight forward that these methods are suitable for dynamic environments that are guaranteed to reappear, known as cyclic DOPs. In this paper, direct memory schemes are integrated to the PBIL where only the sample is stored and reused directly to the current samples. Based on a series of cyclic dynamic test problems, experiments are conducted to compare PBILs with the two types of memory schemes. The experimental results show that one specific direct memory scheme, where memory-based immigrants are generated, always improves the performance of PBIL. Finally, the memory-based immigrant PBIL is compared with other peer algorithms and shows promising performance.

Benchmarking dynamic three-dimensional bin packing problems using discrete-event simulation

Ran Wang, Trung Thanh Nguyen, Shayan Kavakeb, Zaili Yang, Changhe Li

In this paper a framework is developed to generate benchmark problems for dynamic three-dimensional (3D) bin packing problems (BPPs). This framework is able to generate benchmark problems for different variants of BPPs by taking into account potential uncertainty in real-world BPPs, which are uncertainties in dimensions, costs, weights of upcoming items. This paper has three main contributions. First, a benchmark generator framework is developed for the first time using an open source discrete-event simulation platform. This framework generates benchmark problems for BPPs by reproducing uncertainty in real-world BPPs. Second, this framework can be integrated with any dynamic BPP algorithm so that the optimisation algorithm can be run alongside the simulation to solve dynamic BPPs. Third, various performance measures from the literature are included in the framework to evaluate the optimisation algorithms from different perspectives. Thanks to the 3D visualisation feature of this framework, the optimisation results can also be observed visually. Finally, empirical experiments on a real-world BPP are conducted to verify these contributions.

Genetic Programming Algorithms for Dynamic Environments

Joao Macedo, Ernesto Costa, Lino Marques

Evolutionary algorithms are a family of stochastic search heuristics that include Genetic Algorithms (GA) and Genetic Programming (GP). Both GAs and GPs have been successful in many applications, mainly with static scenarios. However, many real world applications involve dynamic environments (DE). Many work has been made to adapt GAs to DEs, but only a few efforts in adapting GPs for this kind of environments. In this paper we present novel GP algorithms for dynamic environments and study their performance using three dynamic benchmark problems, from the areas of Symbolic Regression, Classification and Path Planning. Furthermore, we apply the best algorithm we found in the navigation of an Erratic Robot through a dynamic Santa Fe Ant Trail and compare its performance to the standard GP algorithm. The results, statistically validated, are very promising.

A Memory-Based NSGA-II Algorithm for Dynamic Multi-Objective Optimization Problems

Shaaban Sahmoud, Haluk Topcuoglu

Dynamic multi-objective optimization problems (DMOPs) have been rapidly attracting the interest of the research community. Although static multi-objective evolutionary algorithms have been adapted for solving the DMOPs in the literature, some of those extensions may have high running time and may be inefficient for the given set of test cases. In this paper, we present a new hybrid strategy by integrating the memory concept with the NSGA-II algorithm, called the MNSGA-II algorithm. The proposed algorithm utilizes an explicit memory to store a number of non-dominated solutions using a new memory updating technique. The stored solutions are reused in later stages to reinitialize part of the population when an environment change occurs. The performance of the MNSGA-II algorithm is validated using three test functions from a framework proposed in a recent study. The results show that performance of the MNSGA-II algorithm is competitive with the other state-of-the-art algorithms in terms of tracking the true Pareto front and maintaining the diversity.

Hybrid Dynamic Resampling Algorithms for Evolutionary Multi-objective Optimization of Invariant-Noise Problems

Florian Siegmund, Amos H.C. Ng, Kalyanmoy Deb

In Simulation-based Evolutionary Multi-objective Optimization (EMO) the available time for optimization usually is limited. Since many real-world optimization problems are stochastic models, the optimization algorithm has to employ a noise compensation technique for the objective values. This article analyzes Dynamic Resampling algorithms for handling the objective noise. Dynamic Resampling improves the objective value accuracy by spending more time to evaluate the solutions multiple times, which tightens the optimization time limit even more. This circumstance can be used to design Dynamic Resampling algorithms with a better sampling allocation strategy that uses the time limit. In our previous work, we investigated Time-based Hybrid Resampling algorithms for Preference-based EMO. In this article, we extend our studies to general EMO which aims to find a converged and diverse set of alternative solutions along the whole Pareto-front of the problem. We focus on problems with an invariant noise level, i.e. a flat noise landscape.

Thursday 31 March

0930 - 1110 EvoAPP4 : Bio-inspired algorithms applied to networks

chairs : Ivanoe De Falco & Antonio Della Cioppa

Talks are 20 mins including setup and questions

Joint Topology Optimization, Power Control and Spectrum Allocation for Intra-Vehicular Multi-hop Sensor Networks using Dandelion-encoded Heuristics

Javier Del Ser, Miren Nekane Bilbao, Cristina Perfecto, Gonzalez-Pardo, Sergio Campos-Cordobes

In the last years the interest in multi-hop communications has gained momentum within the research community due to the challenging characteristics of the intra-vehicular radio environment and the stringent robustness imposed on critical sensors within the vehicle. As opposed to point-to-point network topologies, multi-hop networking allows for an enhanced communication reliability at the cost of an additional processing overhead. In this context this manuscript poses a novel bi-objective optimization problem aimed at jointly minimizing 1) the average Bit Error Rate (BER) of sensing nodes under a majority fusion rule at the central data collection unit; and 2) the mean delay experienced by packets forwarded by such nodes due to multi-hop networking, frequency channel switching time multiplexing at intermediate nodes. The formulated paradigm is shown to be computationally tractable via a combination of evolutionary meta-heuristic algorithms and Dandelion codes, the latter capable of representing tree-like structures like those modeling the multi-hop routing approach. Simulations are carried out for realistic values of intra-vehicular radio channels and co-channel interference due to nearby IEEE 802.11 signals. The obtained results are promising and pave the way towards assessing the practical performance of the proposed scheme in real setups.

A Hybrid Discrete Artificial Bee Colony Algorithm for the Multicast Routing Problem

Yannis Marinakis, Magdalene Marinaki, Athanasios Migdalas

In this paper, a new algorithm is proposed for the solution of the Multicast Routing Problem. The algorithm is based on the Artificial Bee Colony approach hybridized with Variable

Neighborhood Search. The quality of the algorithm is evaluated with experiments conducted on suitably modified benchmark instances of the Euclidean Traveling Salesman Problem from the TSP library. The results of the algorithm are compared to results obtained by several versions of the Particle Swarm Optimization algorithm. The comparisons indicated the effectiveness of the new approach.

An (MI)LP-based Primal Heuristic for 3-Architecture Connected Facility Location in Urban Access Network Design *(best paper candidate)*
Fabio D'Andreagiovanni, Fabian Mett, Jonad Pulaj
We investigate the 3-architecture Connected Facility Location Problem arising in the design of urban telecommunication access networks integrating wired and wireless technologies. We propose an original optimization model for the problem that includes additional variables and constraints to take into account wireless signal coverage represented through signal-to-interference ratios. Since the problem can prove very challenging even for modern state-of-the art optimization solvers, we propose to solve it by an original primal heuristic that combines a probabilistic fixing procedure, guided by peculiar Linear Programming relaxations, with an exact MIP heuristic, based on a very large neighborhood search. Computational experiments on a set of realistic instances show that our heuristic can find solutions associated with much lower optimality gaps than a state-of-the-art solver.

A Heuristic Crossover Enhanced Evolutionary Algorithm for Clustering Wireless Sensor Network
Muyiwa Olakanmi Oladimeji, Mikdam Turkey, Sandra Dudley
In this paper, a Heuristic-Crossover Enhanced Evolutionary Algorithm for Cluster Head Selection is proposed. The algorithm uses a novel heuristic crossover operator to combine two different solutions in order to achieve a high quality solution that distributes the energy load evenly among the sensor nodes and enhances the distribution of cluster head nodes in a network. Additionally, we propose the Stochastic Selection of Inactive Nodes, a mechanism inspired by the Boltzmann Selection process in genetic algorithms. This mechanism stochastically considers coverage effect in the selection of nodes that are required to go into sleep mode in order to conserve energy of sensor nodes. The proposed selection of inactive node mechanisms and cluster head selections protocol are performed sequentially at every round and are part of the main algorithm proposed, namely the Heuristic Algorithm for Clustering Hierarchy (HACH). The main goal of HACH is to extend network lifetime of wireless sensor networks by reducing and balancing the energy consumption among sensor nodes during communication processes. Our protocol shows improved performance compared with state-of-the-art protocols like LEACH, TCAC and SEECH in terms of improved network lifetime for wireless sensor networks deployments.

Reducing Efficiency of Connectivity-Splitting Attack on Newscast via Limited Gossip
Jakub Muszynski, Sebastien Varrette, Pascal Bouvry
Newscast is a Peer-to-Peer, nature-inspired gossip-based data exchange protocol used for information dissemination and membership management in large-scale, agent-based distributed systems. The model follows a probabilistic scheme able to keep a self-organised, small-world equilibrium featuring a complex, spatially structured and dynamically changing environment. Newscast gained popularity since the early 2000s thanks to its inherent resilience to node volatility as the protocol exhibits strong self-healing properties. However, the original design proved to be surprisingly fragile in a byzantine environment subjected to cheating faults. Indeed, a set of recent studies emphasized the hard-wired vulnerabilities of

the protocol, leading to an efficient implementation of a malicious client, where a few naive cheaters are able to break the network connectivity in a very short time. Extending these previous works, we propose in this paper a modification of the seminal protocol with embedded counter-measures, improving the resilience of the scheme against malicious acts without significantly affecting the original Newscast's properties nor its inherent performance. Concrete experiments were performed to support these claims, using a framework implementing all the solutions discussed in this work.

0930 - 1110 EvoAPP5 : Evolutionary algorithms in games

chair : Antonio Mora

Talks are 20 mins including setup and questions

Orthogonally Evolved AI to Improve Difficulty Adjustment in Video Games
Arend Hintze, Randal Olson, Joel Lehman   *(best paper candidate)*
Computer games are most engaging when their difficulty is well matched to the player's ability, thereby providing an experience in which the player is neither overwhelmed nor bored. In games where the player interacts with computer-controlled opponents, the difficulty of the game can be adjusted not only by changing the distribution of opponents or game resources, but also through modifying the skill of the opponents. Applying evolutionary algorithms to evolve the artificial intelligence that controls opponent agents is one established method for adjusting opponent difficulty. Less-evolved agents (i.e.\ agents subject to fewer generations of evolution) make for easier opponents, while highly-evolved agents are more challenging to overcome. In this publication we test a new approach for difficulty adjustment in games: orthogonally evolved AI, where the player receives support from collaborating agents that are co-evolved with opponent agents (where collaborators and opponents have orthogonal incentives). The advantage is that game difficulty can be adjusted more granularly by manipulating two independent axes: by having more or less adept collaborators, and by having more or less adept opponents. Furthermore, human interaction can modulate (and be informed by) the performance and behavior of collaborating agents. In this way, orthogonally evolved AI both facilitates smoother difficulty adjustment and enables new game experiences.

There can be only one: Evolving RTS Bots via joust selection
Antonio Fernandez Ares, Pablo Garcia-Sanchez, Antonio Miguel Mora Garcia, Pedro A. Castillo, Juan J. Merelo
This paper proposes an evolutionary algorithm for evolving game bots that eschews an explicit fitness function using instead a match between individuals called {\em joust} and implemented as a selection mechanism where only the winner survives. This algorithm has been designed as an optimization approach to generate the behavioural engine of bots for the RTS game Planet Wars using Genetic Programming and has two objectives: first, to deal with the noisy nature of the fitness function and second, to obtain more general bots than those evolved using a specific opponent. In addition, avoiding the explicit evaluation step reduce the number of combats to perform during the evolution and thus, the algorithm time consumption is decreased. Results show that the approach performs converges, is less sensitive to noise than other methods and it yields very competitive bots in the comparison against other bots available in the literature.

8

Evolving Chess-like Games Using Relative Algorithm Performance Profiles

Jakub Kowalski, Marek Szykula

We deal with the problem of automatic generation of complete rules of an arbitrary game. This requires a generic and accurate evaluating function that is used to score games. Recently, the idea that game quality can be measured using differences in performance of various game-playing algorithms of different strengths has been proposed; this is called Relative Algorithm Performance Profiles. We formalize this method into a generally application algorithm estimating game quality, according to some set of model games with properties that we want to reproduce. We applied our method to evolve chess-like boardgames. The results show that we can obtain playable and balanced games of high quality.

Online Evolution for Multi-Action Adversarial Games

Niels Justesen, Tobias Mahlmann, Julian Togelius

We present Online Evolution, a novel method for playing turn-based multi-action adversarial games. Such games, which include most strategy games, have extremely high branching factors due to each turn having multiple actions. In Online Evolution, an evolutionary algorithm is used to evolve the combination of atomic actions that make up a single move, with a state evaluation function used for fitness. We implement Online Evolution for the turn-based multi-action game Hero Academy and compare it with a standard Monte Carlo Tree Search implementation as well as two types of greedy algorithms. Online Evolution is shown to outperform these methods by a large margin. This shows that evolutionary planning on the level of a single move can be very effective for this sort of problems.

The story of their lives: Massive procedural generation of heroes' journeys using evolved agent-based models and logical reasoning

Ruben H. Garcia-Ortega, Pablo Garcia-Sanchez, Juan J. Merelo, Aranzazu San-Gines, Angel Fernandez-Cabezas

The procedural generation of massive subplots and backstories in secondary characters that inhabit Open World videogames usually lead to stereotyped characters that act as a mere backdrop for the virtual world; however, many game designers claim that the stories can be very relevant for the player's experience. For this reason we are looking for a methodology that improves the variability of the characters' personality while enhancing the quality of their backstories following artistic or literary guidelines. In previous works, we used multi agent systems in order to obtain stochastic, but regulated, inter-relations that became backstories; later, we have used genetic algorithms to promote the appearance of high level behaviors inside them. Our current work continues the previous research line and propose a three layered system (Evolutionary computation - Agent-Based Model - Logical Reasoner) that is applied to the promotion of the monomyth, commonly known as the hero's journey, a social pattern that constantly appears in literature, films, and videogames. As far as we know, there is no previous attempt to model the monomyth as a logical theory, and no attempt to use the sub-solutions for narrating purposes. Moreover, this paper shows for the first time this multi-paradigm three-layered methodology to generate massive backstories. Different metrics have been tested in the experimental phase, from those that sum all the monomyth-related tropes to those that promote distribution of archetypes in the characters. Results confirm that the system can make the monomyth emerge and that the metric has to take into account facilitator predicates in order to guide the evolutionary process.

0930 - 1110 EvoAPP6 : Parallel and multi-agents systems

chairs : Ignacio Hidalgo & Francisco Fernandez de Vega

Talks are 20 mins including setup and questions

Implementing Parallel Differential Evolution on Spark *(best paper candidate)*

Diego Teijeiro, Xoan C. Pardo, Patricia Gonzalez, Julio R. Banga, Ramon Doallo
Metaheuristics are gaining increased attention as an efficient way of solving hard global optimization problems. Differential Evolution (DE) is one of the most popular algorithms in that class. However, its application to realistic problems results in excessive computation times. Therefore, several parallel DE schemes have been proposed, most of them focused on traditional parallel programming interfaces and infrastructures. However, with the emergence of Cloud Computing, new programming models, like Spark, have appeared to suit with large-scale data processing on clouds. In this paper we investigate the applicability of Spark to develop parallel DE schemes to be executed in a distributed environment. Both the master-slave and the island-based DE schemes usually found in the literature have been implemented using Spark. The speedup and efficiency of all the implementations were evaluated on the Amazon Web Services (AWS) public cloud, concluding that the island-based solution is the best suited to the distributed nature of Spark. It achieves a good speedup versus the serial implementation, and shows a decent scalability when the number of nodes grows.

ECJ+HADOOP: An easy way to deploy massive runs of evolutionary algorithm

Francisco Chavez, Francisco Fernandez, Cesar Benavides-Alvarez, Daniel Lanza, Juan Villegas, Leonardo Trujillo, Gustavo Olague, Graciela Roman
This paper describes initial steps towards allowing Evolutionary Algorithms (EAs) researchers to easily deploy computing intensive runs of EAs on Big Data infrastructures. Although many proposals have already been described in the literature, and a number of new software tools have been implemented embodying parallel versions of EAs, we present here a different approach. Given traditional resistance to change when adopting new software, we try instead to endow the well known ECJ tool with the MapReduce model. By using the Hadoop framework, we introduce changes in ECJ that allow researchers to launch any EA problem on a big data infrastructure similarly as when a single computer is used to run the algorithm. By means of a new parameter, researchers can choose where the run will be launched, whether in a Hadoop based infrastructure or in a desktop computer. This paper shows the tests performed, how the whole system has been tuned to optimize the running time for ECJ experiments, and finally a realworld problem is shown to describe how the MapReduce model can automatically deploy the tasks generated by ECJ without additional intervention.

Addressing high dimensional multi-objective optimization problems by coevolutionary islands with overlapping search spaces

Pablo Garcia-Sanchez, Julio Ortega, Jesús Gonzalez, Pedro A. Castillo, Juan J. Merelo
Large-scale multi-objective optimization problems with many decision variables have recently attracted the attention of researchers as many data mining applications involving high dimensional patterns can be leveraged using them. Current parallel and distributed computer architectures can provide the required computing capabilities to cope with these problems once efficient procedures are available. In this paper we propose a cooperative

coevolutionary island-model procedure based on the parallel execution of sub-populations, whose individuals explore different domains of the decision variables space. More specifically, the individuals belonging to the same sub-population (island) explore the same subset of decision variables. Two alternatives to distribute the decision variables among the different sub-populations have been considered and compared here. In the first approach, individuals in different sub-population explore disjoint subsets of decision variables (i.e. the chromosomes are divided into disjoints subsets). Otherwise, in the second alternative there are some overlapping among the variables explored by individuals in different sub-populations. The analysis of the obtained experimental results, by using different metrics, shows that coevolutionary approaches provide statistically significant improvements with respect to the base algorithm, being the relation of the number of islands (subpopulations) to the length of the chromosome (number of decision variables) a relevant factor to determine the most efficient alternative to distribute the decision variables.

Leveraging Online Racing and Population Cloning in Evolutionary Multirobot Systems

Fernando Silva, Luis Correia, Anders Lyhne Christensen

Online evolution of controllers on real robots typically requires a prohibitively long evolution time to synthesise effective solutions. In this paper, we introduce two novel approaches to accelerate online evolution in multirobot systems. We introduce a racing technique to cut short the evaluation of poor controllers based on the task performance of past controllers, and a population cloning technique that enables individual robots to transmit an internal set of high-performing controllers to robots nearby. We implement our approaches over odNEAT, which evolves artificial neural network controllers. We assess the performance of our approaches in three tasks involving groups of e-puck-like robots, and we show that they facilitate: (i) controllers with higher performance, (ii) faster evolution in terms of wall-clock time, (iii) more consistent group-level performance, and (iv) more robust, well-adapted controllers.

Multi-Agent Behavior-Based Policy Transfer

Sabre Didi, Geoff Nitschke,

A key objective of transfer learning is to improve and speed-up learning on a target task after training on a different, but related, source task. This study presents a neuro-evolution method that transfers evolved policies within multi-agent tasks of varying degrees of complexity. The method incorporates behavioral diversity (novelty) search as a means to boost the task performance of transferred policies (multi-agent behaviors). Results indicate that transferred evolved multi-agent behaviors are significantly improved in more complex tasks when adapted using behavioral diversity. Comparatively, behaviors that do not use behavioral diversity to further adapt transferred behaviors, perform relatively poorly in terms of adaptation times and quality of solutions in target tasks. Also, in support of previous work, both policy transfer methods (with and without behavioral diversity adaptation), out-perform behaviors evolved in target tasks without transfer learning.

1130 - 1310 EvoAPP7   : Bioinspired algorithms and complex systems

chair : Carlos Cotta

Towards intelligent biological control: Controlling Boolean networks with Boolean networks

Nadia S. Taou, David W. Corne, Michael A. Lones

Gene regulatory networks (GRNs) are the complex dynamical systems that orchestrate the activities of biological cells. In order to design effective therapeutic interventions for diseases such as cancer, there is a need to control GRNs in more sophisticated ways. Computational control methods offer the potential for discovering such interventions, but the difficulty of the control problem means that current methods can only be applied to GRNs that are either very small or that are topologically restricted. In this paper, we consider an alternative approach that uses evolutionary algorithms to design GRNs that can control other GRNs. This is motivated by previous work showing that computational models of GRNs can express complex control behaviours in a relatively compact fashion. As a first step towards this goal, we consider abstract Boolean network models of GRNs, demonstrating that Boolean networks can be evolved to control trajectories within other Boolean networks. The Boolean approach also has the advantage of a relatively easy mapping to synthetic biology implementations, offering a potential path to in vivo realisation of evolved controllers.

The Emergence of Cooperation in Public Goods Games on Randomly Growing Dynamic Networks

Steve Miller, Joshua Knowles

According to evolutionary game theory, cooperation in public goods games is eliminated by free-riders, yet in nature, cooperation is ubiquitous. Artificial models resolve this contradiction via the mechanism of network reciprocity. However, existing research only addresses pre-existing networks and does not specifically consider their origins. Further, much work has focused on scale-free networks and so pre-supposes attachment mechanisms which may not exist in nature. We present a coevolutionary model of public goods games in networks, growing by random attachment, from small founding populations of simple agents. The model demonstrates the emergence of cooperation in moderately heterogeneous networks, regardless of original founders' behaviour, and absent higher cognitive abilities such as recognition or memory. It may thus illustrate a more general mechanism for the evolution of cooperation, from early origins, in minimally cognitive organisms. It is the first example of a model explaining cooperation in public goods games on growing networks.

Influence Maximization in Social Networks with Genetic Algorithms

Doina Bucur, Giovanni Iacca

We live in a world of social networks. Our everyday choices are often influenced by social interactions. Word of mouth, meme diffusion on the Internet, and viral marketing are all examples of how social networks can affect our behaviour. In many practical applications, it is of great interest to determine which nodes have the highest influence over the network, i.e., which set of nodes will, indirectly, reach the largest audience when propagating information. These nodes might be, for instance, the target for early adopters of a product, the most influential endorsers in political elections, or the most important investors in financial operations, just to name a few examples. Here, we tackle the NP-hard problem of influence maximization on social networks by means of a Genetic Algorithm. We show that, by using simple genetic operators, it is possible to find in feasible runtime solutions of high-influence that are comparable, and occasionally better, than the solutions found by a number of known heuristics (one of which was previously proven to have the best possible approximation guarantee, in polynomial time, of the optimal solution). The advantages of Genetic Algorithms

show, however, in them not requiring any assumptions about the graph underlying the network, and in them obtaining more diverse sets of feasible solutions than current heuristics.

Measuring Diversity of Socio-cognitively Inspired ACO Search
Ewelina Swiderska, Jakub Lasisz, Aleksander Byrski,Tom Lenaerts, Dana Samson, Bipin Indurkhya, Ann Nowe, Marek Kisiel-Dorohinicki
In our recent research, we implemented an enhancement of Ant Colony Optimization incorporating the socio-cognitive dimension of perspective taking. Our initial results suggested that increasing the diversity of ant population --- introducing different pheromones, different species and dedicated inter-species relations --- yielded better results. In this paper, we explore the diversity issue by introducing novel diversity measurement strategies for ACO. Based on these strategies we compare both classic ACO and its socio-cognitive variation.

Multiwinner Voting in Genetic Algorithms for Solving Ill-Posed Global Optimization Problems
Piotr Faliszewski, Jakub Sawicki, Robert Schaefer, Maciej Smolka
Genetic algorithms are a group of powerful tools for solving ill-posed global optimization problems in continuous domains. In case in which the insensitivity of the fitness function is the main obstacle, the most desired feature of a genetic algorithm is its ability to explore plateaus of the fitness function, surrounding its minimizers. In this paper we suggest a way of maintaining diversity of the population in the plateau regions, based on a new approach for the selection based on the theory of multiwinner elections among autonomous agents. The paper delivers a detailed description of the new selection algorithm, computational experiments that guide the choice of the proper multiwinner rule to use, and a preliminary experiment showing the proposed algorithm's effectiveness in exploring a fitness function's plateau.

1130 - 1310 EvoAPP8 : Bioinspired Algorithms in Energy Applications

chairs : Neil Urquhart & Kevin Sim

Talks are 20 mins including setup and questions

Stigmergy-Based Scheduling of Flexible Loads
Fredy Rios, Lukas Konig, Hartmut Schmeck
In this paper, we address the rescheduling of shiftable loads in a sub-section of the power grid (micro-grid) to maximize the utilization of renewable energy sources (RES). The objective is to achieve a schedule for all customers in the micro-grid such that the RES output utilization is maximized. Customers correspond to residential households provided with intelligent appliances with the ability to recalculate their operation times. We propose an approach based on stigmergy to efficiently find a close-to-optimal solution to the general problem. An empirical analysis of the internal functioning of the algorithm is performed. Furthermore, the performance of the algorithm is compared to a price-based approach.

A hybrid genetic algorithm for the interaction of electricity retailers with demand response
Maria Joao Alves, Carlos Henggeler Antunes, Pedro Carrasqueira
In this paper a bilevel programming model is proposed for modeling the interaction between

electricity retailers and consumers endowed with energy management systems capable of providing demand response to variable prices. The model intends to determine the pricing scheme established by the retailer (upper level decision maker) to the consumer (lower level decision maker) and the optimal load schedule adopted by the consumer under this price setting. The lower level optimization problem is formulated as a mixed-integer linear programming (MILP) problem. A hybrid approach consisting of a genetic algorithm and an exact MILP solver is proposed. The individuals of the population represent the retailer's choices (electricity prices). For each price setting, the exact optimal solution to the consumer's problem is obtained in a very efficient way using the MILP solver. An illustrative case is analyzed and discussed.

Comparison of Multi-objective Evolutionary Optimization in Smart Building Scenarios *(best paper candidate)*

Marlon Braun, Thomas Dengiz, Ingo Mauser, Hartmut Schmeck
The optimization of operating times and operation modes of devices and systems that consume or generate electricity in buildings by building energy management systems promises to alleviate problems arising in today's electricity grids. Conflicting objectives may have to be pursued in this context, giving rise to a multi-objective optimization problem. This paper presents the optimization of appliances as well as heating and air-conditioning devices in two distinct settings of smart buildings, a residential and a commercial building, with respect to the minimization of energy costs, CO_2 emissions, discomfort, and technical wearout. We propose new encodings for appliances that are based on a combined categorization of devices respecting both, the optimization of operating times as well as operation modes, e.g., of hybrid devices. To identify an evolutionary algorithm that promises to lead to good optimization results of the devices in our real-world lab environments, we compare four state-of-the-art algorithms in realistic simulations: ESPEA, NSGA-II, NSGA-III, and SPEA2. The results show that ESPEA and NSGA-II significantly outperform the other two algorithms in our scenario.

Electrical Load Pattern Shape Clustering using Ant Colony Optimization

Fernando Lezama, Ansel Y. Rodriguez, Enrique Munoz de Cote
Electrical Load Pattern Shape (LPS) clustering of customers is an important part of the tariff formulation process. Nevertheless, the patterns describing the energy consumption of a customer have some characteristics (e.g., a high number of features corresponding to time series reflecting the measurements of a typical day) that make their analysis different from other pattern recognition applications. In this paper, we propose a clustering algorithm based on ant colony optimization (ACO) to solve the LPS clustering problem. We use four well-known clustering metrics (i.e., CDI, SI, DEV and CONN), showing that the selection of a clustering quality metric plays an important role in the LPS clustering problem. Also, we compare our LPS-ACO algorithm with traditional algorithms, such as k-means and single-linkage, and a state-of-the-art Electrical Pattern Ant Colony Clustering (EPACC) algorithm designed for this task. Our results show that LPS-ACO performs remarkably well using any of the metrics presented here.

A Decentralized PSO with Decoder for Scheduling Distributed Electricity Generation

Jorg Bremer, Sebastian Lehnhoff
A steadily increasing pervasion of the distribution grid with rather small renewable energy resources imposes fluctuating and hardly predictable feed-in and thus calls for new predictive load planning strategies. On the other hand, combined with controllable, shiftable loads and

electrical storages, these energy units set up a flexibility potential for fine-grained control. To tap the full potential, distributed control strategies are discussed for scheduling due to the expected large number of controlled entities. Decoder strategies for unit independent algorithm implementation and feasibility assurance had recently been applied to some first optimization approaches for scheduling in smart grid. We extended a distributed particle swarm to harnesses such decoder approach for model independent constraint-handling and achieved a higher accuracy compared with other approaches. A multi swarm is integrated after the island model into a decentralized agent-based solution and compared with an established decentralized approach for predictive scheduling within virtual power plants. We demonstrate the superiority of the particle swarm in terms of achieved solution accuracy and the competitiveness in terms of sent messages.

1415 - 1555 EvoAPP9 : Natural computing methods in finance

chairs : Anthony Brabazon & Michael Kampouridis

Talks are 20 mins including setup and questions

Portfolio Optimization, a Decision-Support Methodology for Small Budgets
Igor Deplano, Giovanni Squillero, Alberto Tonda
Several machine learning paradigms have been applied to financial forecasting, attempting to predict the market's behavior, with the final objective of profiting from trading shares. While anticipating the performance of such a complex system is far from trivial, this issue becomes even harder when the investors do not have large amounts of money available. In this paper, we present an evolutionary portfolio optimizer for the management of small budgets. The expected returns are modeled resorting to Multi-layer Perceptrons, trained on past market data, and the portfolio composition is chosen by approximating the solution to a multi-objective constrained problem. An investment simulator is then used to measure the portfolio performance. The proposed approach is tested on real-world data from Milan stock exchange, exploiting information from January 2000 to June 2010 to train the framework, and data from July 2010 to August 2011 to validate it. The presented tool is finally proven able to obtain a more than satisfying profit for the considered time frame.

Improving Fitness Functions in Genetic Programming for Classification on Unbalanced Credit Card Data
Van Loi Cao, Nhien-An Le-Khac, Michael O'Neill, Miguel Nicolau, James McDermott
Credit card classification based on machine learning has attracted considerable interest from the research community. One of the most important tasks in this area is the ability of classifiers to handle the imbalance in credit card data. In this scenario, classifiers tend to yield poor accuracy on the minority class despite realizing high overall accuracy. This is due to the influence of the majority class on traditional training criteria. In this paper, we aim to apply genetic programming to address this issue by adapting existing fitness functions. We examine two fitness functions from previous studies and develop two new fitness functions to evolve GP classifiers with superior accuracy on the minority class and overall. Two UCI credit card datasets are used to evaluate the effectiveness of the proposed fitness functions. The results demonstrate that the proposed fitness functions augment GP classifiers, encouraging fitter solutions on both the minority and the majority classes.

Enhanced Multi-objective Population-Based Incremental Learning with Applications in Risk Treaty Optimization

Omar Andres Carmona Cortes, Andrew Rau-Chaplin

The purpose of this paper is to revisit the Multiobjective Population-Based Incremental Learning method and show how its performance can be improved in the context of a real-world financial optimization problem. The proposed enhancements lead to both better performance and improvements in the quality of solutions. Its performance was assessed in terms of runtime and speedup when parallelized. Also, metrics such as the average number of solutions, the average hypervolume, and coverage have been used in order to compare the Pareto frontiers obtained by both the original and enhanced methods. Results indicated that the proposed method is 22.1\% faster, present more solutions in the average (better defining the Pareto frontier) and often generates solutions having larger hypervolumes. The enhanced method achieves a speedup of 15.7 on 16 cores of a dual socket Intel multi-core machine when solving a Reinsurance Contract Optimization problem involving 15 Layers or sub-contracts.

Evolutionary Multi-objective Optimization for Portfolios in Emerging Markets: Contrasting Higher Moments and Median Models

Mai Ibrahim, Mohammed El-Beltagy, Motaz Khorshid

Multi-objective Evolutionary algorithms are well suited to Portfolio Optimization and hence have been applied in complex situations were traditional mathematical programming falls short. Often they were used in portfolios scenario of classical Mean-Variance which are not applicable to the Emerging Markets. Emerging Markets are characterized by return distributions that have shown to exhibit significance departure from normality and are characterized by skewness and fat tails. Therefore higher moments models and median models have been suggested in the literature for asset allocation in this case. Three higher moment models namely the Mean-Variance-Skewness, Mean-Variance-Skewness-Kurtosis, Mean-Variance-Skewness-Kurtosis for return and liquidity and three median models namely the Median-Value at Risk, Median-Conditional Value at Risk and Median-Mean Absolute Deviation are formulated as a multi-objective problem and solved using a multi-objective evolutionary algorithm namely the non-dominated sorting genetic algorithm II .The six models are compared and tested on real financial data of the Egyptian Index EGX. The median models were found in general to outperform the higher moments models. The performance of the median models was found to be better as the out-sample time increases.

Genetic Programming with Memory for Financial Trading

Alexandros Agapitos, Anthony Brabazon, Michael O'Neill

A memory-enabled program representation in strongly-typed Genetic Programming (GP) is compared against the standard representation in a number of financial time-series modelling tasks. The paper first presents a survey of GP systems that utilise memory. Thereafter, a number of simulations show that memory-enabled programs generalise better than their standard counterparts in most datasets of this problem domain.

1415 - 1555 EvoAPPS10 : Evolutionary algorithms in industrial and simulated environments

chairs : Kevin Sim & Neil Urquhart

Constrained Level Generation through Grammar-Based Evolutionary Algorithms

Jose M. Font, Roberto Izquierdo, Daniel Manrique, Julian Togelius

This paper introduces an evolutionary method for generating levels for adventure games, combining speed, guaranteed solvability of levels and authorial control. For this purpose, a new graph-based two-phase level encoding scheme is developed. This method encodes the structure of the level as well as its contents into two abstraction layers: the higher level defines an abstract representation of the game level and the distribution of its content among different inter-connected game zones. The lower level describes the content of each game zone as a set of graphs containing rooms, doors, monsters, keys and treasure chests. Using this representation, game worlds are encoded as individuals in an evolutionary algorithm and evolved according to an evaluation function meant to approximate the entertainment provided by the game level. The algorithm is implemented into a design tool that can be used by game designers to specify several constraints of the worlds to be generated. This tool could be used to facilitate the design of game levels, for example to make professional-level content production possible for non-experts.

Can Evolutionary Algorithms Beat Dynamic Programming for Hybrid Car Control?

Tobias Rodemann, Ken Nishikawa

Finding the best possible sequence of control actions for a hybrid car in order to minimize fuel consumption is a well-studied problem. A standard method is Dynamic Programming (DP) that is generally considered to provide solutions close to the global optimum in relatively short time. To our knowledge Evolutionary Algorithms (EAs) have so far not been used for this setting, due to the success of DP. In this work we compare DP and EA for a well-studied example and find that for the basic scenario EA is indeed clearly outperformed by DP in terms of calculation time and quality of solutions. But, we also find that when going beyond the standard scenario towards more realistic (and complex) scenarios, EAs can actually deliver a performance en par or in some cases even exceeding DP, making them useful in a number of relevant application scenarios.

Environment-Model Based Testing with Differential Evolution in an Industrial Setting

Annamaria Szenkovits, Noemi Gasko, Erwan Jahier

Reactive systems interact continuously with their environments. In order to test such systems, one needs to design executable environment models. Such models are intrinsically stochastic, because environment may vary a lot, and also because they are not perfectly known. We propose an environment-model based testing framework optimized for reactive systems, where Differential Evolution (DE) is used to fine-tune the environment model and to optimize test input generation. In order to evaluate the proposed method, we present a case study involving a real-world SCADE system from the domain of railway automation. The problem specification was proposed by our industrial partner, Siemens. Our experimental data shows that DE can be used efficiently to increase the structural coverage of the System Under Test.

Workforce Scheduling in Inbound Customer Call Centres With a Case Study

Goran Molnar, Domagoj Jakobovic, Matija Pavelic

Call centres are an important tool that businesses use to interact with their clients. Their efficiency is especially significant since long queuing times can reduce customer satisfaction. Assembling the call centre work schedule is a complex task that needs to take various and often mutually conflicting goals into account. In this paper, we present a workforce scheduling system suited for small to medium call centres and adjusted to the needs of two real-world client institutions. The scheduling problem is to minimise the difference between allocated and forecasted number of staff members while also caring for numerous legal and organisational constraints as well as staff preferences. A flexible constraint handling framework is devised to enable rapid prototyping methodology used during the development. Based on it, two metaheuristics are devised for schedule construction: GRASP and iterated local search. Performance analysis and comparisons for these two methods are provided, on a real-world problem example. The devised system is successfully implemented in a real world setting of call centres at PBZCard, Croatian largest credit card vendor and PBZ (Intesa Sanpaolo), one of the largest Croatian banks.

Optimization of Operation and Control Strategies for Battery Energy Storage Systems by Evolutionary Algorithms
Jan Muller, Matthias Marz, Ingo Mauser, Hartmut Schmeck
To support the utilization of renewable energies, an optimized operation of energy systems is important. Often, the use of battery energy storage systems is stated as one of the most important measures to support the integration of intermittent renewable energy sources into the energy system. Additionally, the complexity of the energy system with its many interdependent entities as well as the economic efficiency call for an elaborate dimensioning and control of these storage systems. In this paper, we present an approach that combines the forward-looking nature of optimization and prediction with the feedback control of closed-loop controllers. An evolutionary algorithm is used to determine the parameters for a closed-loop controller that controls the charging and discharging control strategy of a battery in a smart building. The simulation and evaluation of a smart residential building scenario demonstrates the ability to improve the operation and control of a battery energy storage system. The optimization of the control strategy allows for the optimization with respect to variable tariffs while being conducive for the integration of renewable energy sources into the energy system.

1615 - 1745   EvoAPP11 : Biological applications

chair : Federico Divina

Talks are 20 mins including setup and questions

On Combinatorial Optimisation in Analysis of Protein-Protein Interaction and Protein Folding Networks
David Chalupa
Protein-protein interaction networks and protein folding networks represent prominent research topics at the intersection of bioinformatics and network science. In this paper, we present a study of these networks from combinatorial optimisation point of view. Using a combination of classical heuristics and stochastic optimisation techniques, we were able to identify several interesting combinatorial properties of biological networks of the COSIN project. We obtained optimal or near-optimal solutions to maximum clique and chromatic

number problems for these networks. We also explore patterns of both non-overlapping and overlapping cliques in these networks. Optimal or near-optimal solutions to partitioning of these networks into non-overlapping cliques and to maximum independent set problem were discovered. Maximal cliques are explored by enumerative techniques. Domination in these networks is briefly studied, too. Applications and extensions of our findings are discussed.

Automating biomedical data science through tree-based pipeline optimization

Randal Olson, Ryan Urbanowicz, Peter Andrews, Nicole Lavender, La Creis Kidd, Jason Moore   *(best paper candidate)*

Over the past decade, data science and machine learning has grown from a mysterious art form to a staple tool across a variety of fields in academia, business, and government. In this paper, we introduce the concept of tree-based pipeline optimization for automating one of the most tedious parts of machine learning---pipeline design. We implement a Tree-based Pipeline Optimization Tool (TPOT) and demonstrate its effectiveness on a series of simulated and real-world genetic data sets. In particular, we show that TPOT can build machine learning pipelines that achieve competitive classification accuracy and discover novel pipeline operators---such as synthetic feature constructors---that significantly improve classification accuracy on these data sets. We also highlight the current challenges to pipeline optimization, such as the tendency to produce pipelines that overfit the data, and suggest future research paths to overcome these challenges. As such, this work represents an early step toward fully automating machine learning pipeline design.

Bicliques in Graphs with Correlated Edges: From Artificial to Biological Networks

Aaron Kershenbaum, Alicia Cutillo, Christian Darabos, Murray Keitha, Schiaffino Robert, Jason H. Moore

Networks representing complex biological interactions are often very intricate and rely on algorithmic tools for thorough quantitative analysis. In bi-layered graphs, identifying subgraphs of potential biological meaning relies on identifying bicliques between two sets of associated nodes, or variables -- for example, diseases and genetic variants. Researchers have developed multiple approaches for forming bicliques and it is important to understand the features of these models and their applicability to real-life problems. We introduce a novel algorithm specifically designed for finding maximal bicliques in large datasets. In this study, we applied this algorithm to a variety of networks, including artificially generated networks as well as biological networks based on phenotype-genotype and phenotype-pathway interactions. We analyzed performance with respect to network features including density, node degree distribution, and correlation between nodes, with density being the major contributor to computational complexity. We also examined sample bicliques and postulate that these bicliques could be useful in elucidating the genetic and biological underpinnings of shared disease etiologies and in guiding hypothesis generation. Moving forward, we propose additional features, such as weighted edges between nodes, that could enhance our study of biological networks.

A Multi-objective Genetic Programming Biomarker Detection Approach in Mass Spectrometry Data

Soha Ahmed, Mengjie Zhang, Lifeng Peng, Bing Xue

Mass spectrometry is currently the most commonly used technology in biochemical research for proteomic analysis. The main goal of proteomic profiling using mass spectrometry is the classification of samples from different clinical states. This requires the identification of proteins or peptides (biomarkers) that are expressed differentially between different clinical

states. However, due to the high dimensionality of the data and the small number of samples, classification of mass spectrometry data is a challenging task. Therefore, an effective feature manipulation algorithm either through feature selection or construction is needed to enhance the classification performance and at the same time minimise the number of features. Most of the feature manipulation methods for mass spectrometry data treat this problem as a single objective task which focuses on improving the classification performance. This paper presents two new methods for biomarker detection through multi-objective feature selection and feature construction. The results show that the proposed multi-objective feature selection method can obtain better subsets of features than the single-objective algorithm and two traditional multi-objective approaches for feature selection. Moreover, the multi-objective feature construction algorithm further improves the perfomance over the multi-objective feature selection algorithm. This paper is the first multi-objective genetic programming approach for biomarker detection in mass spectrometry data.

1615 - 1745   EvoAPP12   :   Interactive Presentations I

chair : Antonio Mora

Short talks are 6 mins

Dangerousness Metric for Gene Regulated Car Driving
Sylvain Cussat-Blanc, Jean Disset, Stephane Sanchez
In this paper, we show how a dangerousness metric can be used to modify the input of a gene regulatory network when plugged to a virtual car. In the context of the 2015 Simulated Car Racing Championship organized during GECCO 2015, we have developed a new cartography methodology able to inform the controller of the car about the incoming complexity of the track: turns (slipperiness, angle, etc.) and bumps. We show how this dangerousness metric improves the results of our controller and outperforms other approaches on the tracks used in the competition.

Using Isovists to Evolve Terrains with Gameplay Elements
Andrew William Pech, Chiou-Peng Lam, Philip Hingston, Martin Masek
The virtual terrain for a video game generally needs to exhibit a collection of gameplay elements, such as some areas suitable for hiding and others for large scale battles. A key problem in automating terrain design is the lack of a quantitative definition of terrain gameplay elements. In this paper, we ad-dress the problem by proposing a representation for gameplay elements based on a combination of space-based isovist measures from the field of architecture and graph-connectivity metrics. We then propose a genetic algorithm-based approach that evolves a set of modifications to an existing terrain so as to exhibit the gameplay element characteristics. The potential for this approach in the design of computer game environments is examined by generating terrain containing instances of the "hidden area" game element type. Results from four preliminary tests are described to show the potential of this research.

A spatially-structured PCG method for content diversity in a Physics-based simulation game
Raul Lara-Cabrera, Alejandro Gutierrez-Alcoba, Antonio Jose' Fernandez-Leiva
This paper presents a spatially-structured evolutionary algorithm (EA) to procedurally generate game maps of different levels of difficulty to be solved, in Gravityvolve!, a physics-

20

based simulation videogame that we have implemented and which is inspired by the n-body problem, a classical problem in the field of physics and mathematics. The proposal consists of a steady-state EA whose population is partitioned into three groups according to the difficulty of the generated content (hard, medium or easy) which can be easily adapted to handle the automatic creation of content of diverse nature in other games. In addition, we present three fitness functions, based on multiple criteria (i.e:, intersections, gravitational acceleration and simulations), that were used experimentally to conduct the search process for creating a database of maps with different difficulty in Gravityvolve!

Design and Evaluation of an Extended Learning Classifier-based StarCraft Micro AI

Stefan Rudolph, Sebastian von Mammen, Johannes Jungbluth, Jorg Hahner

Due to the manifold challenges that arise when developing an artificial intelligence that can compete with human players, the popular realtime-strategy game \SC\ (BW) has received attention from the computational intelligence research community. It is an ideal testbed for methods for \textit{self-adaption at runtime} designed to work in complex technical systems. In this work, we utilize the broadly-used Extended Classifier System (XCS) as a basis to develop different models of BW micro AIs: the Defender, the Attacker, the Explorer and the Strategist. We evaluate theses AIs with a focus on their adaptive and co-evolutionary behaviors. To this end, we stage and analyze the outcomes of a tournament among the proposed AIs and we also test them against a non-adaptive player to provide a proper baseline for comparison and learning evolution. Of the proposed AIs, we found the Explorer to be the best performing design, but, also that the Strategist shows an interesting behavioral evolution.

Benchmarking languages for evolutionary algorithms

JJ Merelo, Pedro Castillo, Israel Blancas, Gustavo Romero, Pablo Garcia-Sánchez, Antonio Fernandez-Ares, Víctor Rivas, Mario Garcia-Valdez

Although performance is important, several other issues should be taken into account when choosing a particular language for implementing an evolutionary algorithm, such as the fact that the speed of different languages when carrying out an operation will depend on several factors, including the size of the operands, the version of the language and underlying factors such as the operating system. However, it is usual to rely on compiled languages, namely Java or C/C++, for carrying out any implementation without considering other languages or rejecting them outright on the basis of performance. Since there are a myriad of languages nowadays, it is interesting however to measure their speed when performing operations that are usual in evolutionary algorithms. That is why in this paper we have chosen three evolutionary algorithm operations: bitflip mutation, crossover and the fitness function OneMax evaluation, and measured the speed for several popular, and some not so popular, languages. Our measures confirm that, in fact, Java, C and C++ not only are the fastest, but also have a behaviour that is independent of the size of the chromosome. However, we have found other compiled language such as Go or interpreted languages such as Python to be fast enough for most purposes. Besides, these experiments show which of these measures are, in fact, the best for choosing an implementation language based on its performance.

On the Closest Averaged Hausdorff Archive for a Circularly Convex Pareto Front

Gunter Rudolph, Oliver Schutze, Heike Trautmann

The averaged Hausdorff distance has been proposed as an indicator for assessing the quality of finitely sized approximations of the Pareto front of a multiobjective problem. Since

many set-based, iterative optimization algorithms store their currently best approximation in an internal archive these approximations are also termed archives. In case of two objectives and continuous variables it is known that the best approximations in terms of averaged Hausdorff distance are subsets of the Pareto front if it is concave. If it is linear or circularly concave the points of the best approximation are equally spaced. Here, it is proven that the optimal averaged Hausdorff approximation and the Pareto front have an empty intersection if the Pareto front is circularly convex. But the points of the best approximation are equally spaced and they rapidly approach the Pareto front for increasing size of the approximation.

Evolving Smoothing Kernels for Global Optimization
Paul Manns, Kay Hamacher
The Diffusion-Equation Method (DEM) - sometimes synonymously called the Continuation Method - is a well-known natural computation approach in optimization. The DEM continuously transforms the objective function by a (Gaussian) kernel technique to reduce barriers separating local and global minima. Now, the DEM can successfully solve problems of small sizes. Here, we present a generalization of the DEM to use convex combinations of smoothing kernels in Fourier space. We use a genetic algorithm to incrementally optimize the (meta-)combinatorial problem of finding better performing kernels for later optimization of an objective function. For two test applications we derive and show their transferability to larger problems. Most strikingly, the original DEM failed on a number of the test instances to find the global optimum while our transferable kernels - obtained via evolutionary computations - were able to find the global optimum.

On-line Evolution of Foraging Behaviour in a Population of Real Robots
Jacqueline Heinerman, Alessandro Zonta, Evert Haasdijk, A.E.Eiben
This paper describes a study in evolutionary robotics conducted completely in hardware without using simulations. The experiments employ on-line evolution, where robot controllers evolve on-the-fly in the robots' environment as the robots perform their tasks. The main issue we consider is the feasibility of tackling a non-trivial task in a realistic timeframe. In particular, we investigate whether a population of six robots can evolve foraging behaviour in one hour. The experiments demonstrate that this is possible and they also shed light on some of the important features of our evolutionary system. Further to the specific results we also advocate the system itself. It provides an example of a replicable and affordable experimental set-up for other researches to engage in research into on-line evolution in a population of real robots.

Friday 1 April

1130 - 1300 EvoAPP13 : Pattern Recognition & Numerical Optimisation

chair : Anna I Esparcia-Alcázar

Talks are 20 mins including setup and questions

Mutual Information Estimation for Filter Based Feature Selection Using Particle Swarm Optimization
Bach Hoai Nguyen, Bing Xue, Peter Andreae
Feature selection is a pre-processing step in classification, which selects a small set of

important features to improve the classifi- cation performance and efficiency. Mutual information is very popular in feature selection because it is able to detect non-linear re- lationship between features. However the existing mutual information approaches only consider two-way interaction between features. In ad- dition, in most methods, mutual information is calculated by counting approach, which might lead to an inaccurate results. This paper proposes a filter feature selection algorithm based on particle swarm optimization (PSO) named PSOMIE, which employs a novel fitness function using nearest neighbor mutual information estimation (NNE) to measure the quality of a feature set. PSOMIE is compared with using all features and two traditional feature selection approaches. The experiment results show that the mutual information estimation successfully guides PSO to search for a small number of features while maintaining or improving the classification performance over using all features and the traditional feature selection methods. In addition, PSOMIE provides a strong con- sistency between training and test results, which may be used to avoid overfitting problem.

A Wrapper Feature Selection Approach to Classification with Missing Data
Cao Truong Tran, Mengjie Zhang, Peter Andreae, Bing Xue
Many industrial and real-world datasets suffer from an unavoidable problem of missing values. The problem of missing data has been addressed extensively in the statistical analysis literature, and also, but to a lesser extent in the classification literature. The ability to deal with missing data is an essential requirement for classification because inadequate treatment of missing data may lead to large errors on classification. Feature selection has been successfully used to improve classification, but it has been applied mainly to complete data. This paper develops a wrapper feature selection approach to classification with missing data and investigates the impact of this approach. Empirical results on 10 datasets with missing values using C4.5 for an evaluation and particle swarm optimisation as a search technique in feature selection show that a wrapper feature selection for missing data not only can help to improve accuracy of the classifier, but also can help to reduce the complexity of the learned classification model.

Local Fitness Meta-Models with Nearest Neighbor Regression
Oliver Kramer
In blackbox function optimization, the results of fitness function evaluations can be used to train a regression model. This meta-model can be used to replace function evaluations and thus reduce the number of fitness function evaluations in evolution strategies (ES). In this paper, we show that a reduction of the number of fitness function evaluations of a (1+1)-ES is possible with a combination of a nearest neighbor regression model, a local archive of fitness function evaluations, and a comparatively simple meta-model management. We analyze the reduction of fitness function evaluations on set of benchmark functions.

Validating the Grid Diversity Operator: an Infusion Technique for Diversity Maintenance in Population-based Optimisation Algorithms
Ahmed Salah, Emma Hart, Kevin Sim
We describe a novel diversity method named Grid Diversity Operator (GDO) that can be incorporated into population-based optimization algorithms that support the use of {\em infusion} techniques to inject new material into a population. By replacing the random infusion mechanism used in many optimisation algorithms, the GDO guides the containing algorithm towards creating new individuals in sparsely visited areas of the search space. Experimental tests were performed on a set of 39 multimodal benchmark problems from the literature using GDO in conjunction with a popular immune-inspired algorithm (opt-ainet) and

a sawtooth genetic algorithm. The results show that the GDO operator leads to better quality solutions in all of the benchmark problems as a result of maintaining higher diversity, and makes more efficient usage of the allowed number of objective function evaluations. Specifically, we show that the performance gain from using GDO increases as the dimensionality of the problem instances increases. An exploration of the parameter settings for the two main parameters of the new operator enabled the performance of the operator to be tuned empirically.

1130 - 1300 EvoAPP14 : Interactive Presentations II

chair: Evert Haasdijk

Short talks are 6 mins

Hybrid biclustering algorithms for data mining
Patryk Orzechowski, Krzysztof Boryczko
Hybrid methods are a branch of biclustering algorithms that emerge from combining selected aspects of pre-existing approaches. The syncretic nature of their construction enriches the existing methods providing them with new properties. In this paper the concept of hybrid biclustering algorithms is explained. A representative hybrid biclustering algorithm, inspired by neural networks and associative artificial intelligence, is introduced and the results of its application to microarray data are presented. Finally, the scope and application potential for hybrid biclustering algorithms is discussed.

Discovering potential clinical profiles of Multiple Sclerosis from clinical and pathological free text data with Constraint Non-negative Matrix Factorization
Jacopo Acquarelli, Elena Marchiori, Monica Bianchini
Constrained non-negative matrix factorization (CNMF) is an effective machine learning technique to cluster documents in the presence of class label constraints. In this work, we provide a novel application of this technique in research on neuro-degenerative diseases. Specifically, we consider a dataset of documents from the Netherlands Brain Bank containing free text describing clinical and pathological information about donors affected by Multiple Sclerosis. The goal is to use CNMF for identifying clinical profiles with pathological information as constraints. After pre-processing the documents by means of standard filtering techniques, a feature representation of the documents in terms of bi-grams is constructed. The high dimensional feature space is reduced by applying a trimming procedure. The resulting datasets of clinical and pathological bi-grams are then clustered using non-negative matrix factorization (NMF) and, next, clinical data are clustered using CNMF with constraints induced by the clustering of pathological data. Results indicate the presence of interesting clinical profiles, for instance related to vision or movement problems. In particular, the use of CNMF leads to the identification of a clinical profile related to diabetes mellitus. Pathological characteristics and duration of disease of the identified profiles are analysed. Although highly promising, results of this investigation should be interpreted with care due to the relatively small size of the considered datasets.

Application of Evolutionary Algorithms for the Optimization of Genetic Regulatory Networks
Elise Rosati, Morgan Madec, Abir Rezgui, Quentin Colman, Nicolas Toussaint, Christophe Lallement, Pierre Collet

Synthetic biology aims at reinvesting theoretical knowledge from various domains (biology, engineering, microelectronics) for the development of new biological functions. Concerning the design of such functions, the classical trial-error approach is expensive and time consuming. Computer-aided design is therefore of key interest in this field. As for other domains, such as microelectronics or robotics, evolutionary algorithms can be used to this end. This article is a first step in this direction: it describes the optimization of an existing artificial gene regulatory network (a band-pass detector) using evolutionary algorithms. Evolutionary algorithms successfully find a good set of parameters (the simulated response of the system which fits at 99% the expected response) in about 200 seconds (corresponding to 5000 generations) on a standard computer. This is the proof of concept of our approach. Moreover, results analysis allows the biologist not only to save time during the design process but also to study the specificity of a system.

A Distributed Intrusion Detection Framework based on Evolved Specialized Ensembles of Classifiers

Gianluigi Folino, Francesco Sergio Pisani, Pietro Sabatino

Modern intrusion detection systems must handle many complicated issues in real-time, as they have to cope with a real data stream; indeed, for the task of classification, typically the classes are unbalanced and, in addition, they have to cope with distributed attacks and they have to quickly react to changes in the data. Data mining techniques and, in particular, ensemble of classifiers permit to combine different classifiers that together provide complementary information and can be built in an incremental way. This paper introduces the architecture of a distributed intrusion detection framework and in particular, the detector module based on a meta-ensemble, which is used to cope with the problem of detecting intrusions, in which typically the number of attacks is minor than the number of normal connections. To this aim, we explore the usage of ensembles specialized to detect particular types of attack or normal connections, and Genetic Programming is adopted to generate a non-trainable function to combine each specialized ensemble. Non-trainable functions can be evolved without any extra phase of training and, therefore, they are particularly apt to handle concept drifts, also in the case of real-time constraints. Preliminary experiments, conducted on the well-known KDD dataset and on a more up-to-date dataset, ISCX IDS, show the effectiveness of the approach.

UAV Fleet Mobility Model with Multiple Pheromones for Tracking Moving Observation Targets

Christophe Atten, Loubna Channouf, Gregoire Danoy, Pascal Bouvry

The last years, UAVs have been developed to address a variety of applications ranging from searching and tracking to the surveillance of an area. However, using a single UAV limits the range of possible applications. Therefore, fleets of UAVs are nowadays considered to work together on a common goal which requires novel distributed mobility management models. This work proposes a novel nature-inspired mobility model for UAV fleets based on Ant Colony Optimisation approaches (ACO). It relies on two types of pheromones, a repulsive pheromone to cover the designated area in an efficient way, and an attractive pheromone to detect and to track the maximum number of targets. Furthermore, all decision takings are taken online by each UAV and are fully distributed. Experimental results demonstrate promising target tracking performances together with a small increase in the exhaustivity of the coverage.

Compilable phenotypes: Accelerating the evaluation of individuals in Grammatical Evolution

J. Manuel Colmenar, J. Ignacio Hidalgo, Juan Lanchares, Oscar Garnica, Jose-L. Risco-Martín, Ivan Contreras, Almudena Sanchez, J. Manuel Velasco
This paper presents a method for accelerating the evaluation of individuals in Grammatical Evolution. The method is applied for identification and modeling problems, where, in order to obtain the fitness value of one individual, we need to compute a mathematical expression for different time events. We propose to evaluate all necessary values of each individual using only one mathematical Java code. For this purpose we take profit of the flexibility of grammars, which allows us to generate Java compilable expressions. We test the methodology with a real problem: modeling glucose level on diabetic patients. Experiments confirms that our approach (compilable phenotypes) can get up to 300x reductions in execution time.

GPU Accelerated Molecular Docking Simulation with Genetic Algorithms

Serkan Altuntas, Zeki Bozkus, Basilio B. Fraguela
Receptor-Ligand Molecular Docking is a very computationally expensive process used to predict possible drug candidates for many diseases. A faster docking technique would help life scientists to discover better therapeutics with less effort and time. The requirement of long execution times may mean using a less accurate evaluation of drug candidates potentially increasing the number of false-positive solutions, which require expensive chemical and biological procedures to be discarded. Thus the development of fast and accurate enough docking algorithms greatly reduces wasted drug development resources, helping life scientists discover better therapeutics with less effort and time. In this article we present the GPU-based acceleration of our recently developed molecular docking code. We focus on offloading the most computationally intensive part of any docking simulation, which is the genetic algorithm, to accelerators, as it is very well suited to them. We show how the main functions of the genetic algorithm can be mapped to the GPU. The GPU-accelerated system achieves a speedup of around ~14x with respect to a single CPU core. This makes it very productive to use GPU for small molecule docking cases.

Hybrid Control for a Real Swarm Robotics System in an Intruder Detection Task

Miguel Duarte, Jorge Gomes, Vasco Costa, Sancho Moura Oliveira, Anders Lyhne Christensen
Control design is one of the prominent challenges in the field of swarm robotics. Evolutionary robotics is a promising approach to the synthesis of self-organized behaviors for robotic swarms but it has, so far, only been shown in relatively simple collective behaviors. In this paper, we explore the use of a hybrid control synthesis approach to produce control for a swarm of aquatic surface robots that must perform an intruder detection task. The robots have to go to a predefined area, monitor it, detect and follow intruders, and manage their energy levels by regularly recharging at a base station. The hybrid controllers used in our experiments rely on evolved behavior primitives that are combined through a manually programmed high-level behavior arbitrator. In simulation, we show how simple modifications to the behavior arbitrator can result in different swarm behaviors that use the same underlying behavior primitives, and we show that the composed behaviors are scalable with respect to the swarm size. Finally, we demonstrate the synthesized controller in a real swarm of robots, and show that the behavior successfully transfers from simulation to reality.