INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team tao

Thème apprentissage et optimisation

Futurs

THEME COG

Activity Report

2007

Table of contents

1. Team

TAO (Thème Apprentissage et Optimisation) is a joint project inside PCRI, including researchers from INRIA and the LRI team I & A – Inférence et Apprentissage (CNRS and University of Paris Sud), located in Orsay.

Co-Heads of project-team
Marc Schoenauer [DR INRIA, HdR]
Michèle Sebag [DR CNRS, HdR]

Permanent staff
Anne Auger [CR INRIA]
Nicolas Bredèche [MDC Université Paris Sud]
Philippe Caillou [MDC Université Paris Sud]
Cyril FURTLEHNER [since October 2007]
Cécile Germain [Professeur Université Paris Sud, HdR]
Olivier Teytaud [CR INRIA]

Administrative assistant
Marie-Carol Lopes [Tao/In Situ]

Ph.D. Students
Nicolas BASKIOTIS [started September 2003]
Jacques BIBAI [CIFRE Thalès, started September 2007]
Alexandre DEVERT [started September 2005]
Lou FEDON [started September 2006]
Mary FELKIN [started September 2003]
Romaric GAUDEL [started September 2006]
Sylvain GELLY [defended September 2007]
Cédric HARTLAND [started September 2005]
Mohamed JEBALIA [started September 2004, co-tutelle Université Tunis - Tunisie)]
Fei JIANG [started November 2006 TAO-ALCHEMY]
Claire LE BARON [CIFRE Renault, started September 2005]
Julien PEREZ [started October 2006]
Arpad RIMMEL [started September 2006]
Raymond ROS [started September 2005]
Xiangliang ZHANG [started September 2006]

Post-doctoral fellows
Ettore CAVALLARO [until October 2007]
Luis DACOSTA [since May 2007]

Research Engineer
Bertrand CHARDON [until November 2007]
Nikolaus HANSEN [since April 2006]
Miguel NICOLAU
Damien TESSIER [until September 2007]
Jean-Baptiste HOOCK [since October 2007]

Collaborators
Antoine Cornuéjols [professor INA-PG, Paris, HdR]
Balasz KEGL [CR CNRS, LAL]
Yves KODRATOFF [DR CNRS, HdR]

2. Overall Objectives

2.1. Introduction

Data Mining (DM) has been identified as one of the ten main challenges of the 21st century (MIT Technological Review, fev. 2001). The goal is to exploit the massive amounts of data produced in scientific labs, industrial plants, banks, hospitals or supermarkets, in order to extract valid, new and useful regularities. In other words, DM resumes the Machine Learning (ML) goal, finding (partial) models for the complex system underlying the data.

DM and ML problems can be set as optimization problems, thus leading to two possible approaches. Note that this alternative has been characterized by H. Simon (1982) as follows. *In complex real-world situations, optimization becomes approximate optimization since the description of the real-world is radically simplified until reduced to a degree of complication that the decision maker can handle. Satisficing seeks simplification in a somewhat different direction, retaining more of the detail of the real-world situation, but settling for a satisfactory, rather than approximate-best, decision.*

The first approach is to simplify the learning problem to make it tractable by standard statistical or optimization methods. The alternative approach is to preserve as much as possible the genuine complexity of the goals (yielding "interesting" models, accounting for prior knowledge): more flexible optimization approaches are therefore required, such as those offered by Evolutionary Computation.

Symmetrically, optimization techniques are increasingly used in all scientific and technological fields, from optimum design to risk assessment. Evolutionary Computation (EC) techniques, mimicking the Darwinian paradigm of natural evolution, are stochastic population-based dynamical systems that are now widely known for their robustness and flexibility, handling complex search spaces (e.g. mixed, structured, constrained representations) and non-standard optimization goals (e.g. multi-modal, multi-objective, context-sensitive), beyond the reach of standard optimization methods.

The price to pay for such properties of robustness and flexibility is twofold. On one hand, EC is tuned, mostly by trials and errors, using quite a few parameters. On the other hand, EC generates massive amounts of intermediate solutions. It is suggested that the principled exploitation of preliminary runs and intermediate solutions, through Machine Learning and Data Mining techniques, can offer sound ways of adjusting the parameters and finding short cuts in the trajectories in the search space of the dynamical system.

2.2. Context and overall goal of the project

The overall goals of the project are to model, to predict, to understand, and to control physical or artificial systems. The central claim is that Learning and Optimisation approaches must be used, adapted and integrated in a seamless framework, in order to bridge the gap between the system under study on the one hand, and the expert's goal as to the ideal state/functionality of the system.

Specifically, our research context involves the following assumptions:

1. The systems under study range from large-scale engineering systems to physical or chemical phenomenons, including games. Such systems, sometimes referred to as *complex systems*, can hardly be modelled based on first principles due to their size, their heterogeneity and the incomplete information aspects involved in their behaviour.
2. Such systems can be observed; indeed selecting the relevant observations and providing a reasonably appropriate description thereof is part of the problem to be solved. A further assumption is that these observations are sufficient to build a reasonably accurate model of the system under study.
3. The available expertise is sufficient to assess the system state, and any modification thereof, with respect to the desired states/functionalities. The assessment function is usually not a well-behaved function (differentiable, convex, defined on a continuous domain, etc), barring the use of standard optimisation approaches and making Evolutionary Computation a better suited alternative.

In this context, the objectives of TAO are threefold:

1. using Evolutionary Computation (EC) and more generally Stochastic Optimisation to support Machine Learning (ML);
2. using Statistical Machine Learning to support Evolutionary Computation;
3. investigating integrated ML/EC approaches on diversified and real-world applications.

Due to the unavoidable shift of the scientific environement and people interest after 4 years of activity, the detailed implementation of those objectives have been slightly revised since the initial project proposal 4 years ago, and updated lines of research will be described in next section 3.1.

2.3. Highlights

The MoGo program (see sections 6.2.3 and 5.1) is one of the most prominent successes of TAO in the period; MoGo repeatedly won international computer-go contests in 2006 and 2007; it was the first computer-go program to ever reach 2,000 ELO, and the first one to win over a (3e dan) human player (on 9x9 goban); it still is ranked best of the computer go programs. Its port onto a heavily parallel platform (collaboration with Bull), should increase this domination and open new horizons in computer-go.

3. Scientific Foundations

3.1. Introduction

This section is the prospective part of the report written for the Evaluation Seminar that took place in November 2007 for the INRIA theme COG A, to which TAO belongs. Because this was a unique opportunity to re-think the objectives of the project after 4 years of activity, it naturally becomes the Scientific Foundations of TAO activities for the next 4 years.
Four main objectives have been identified, related to scientific bottlenecks of Machine Learning and Stochastic Optimisation. The first one, illustrated by the results discussed in sections 6.1 and 6.3, is devoted to the search of representations with desirable properties. The second one, exploiting the prominent successes obtained with Multi-Armed Bandit algorithms and MoGo (section 6.2.3), considers the various challenges raised by extending MAB algorithms to dynamic contexts and hostile environments. The third one will investigate the conceptual and algorithmic shift required to deal with modern computing architectures. The fourth one is concerned with *Crossing the Chasm*[1], and making algorithmic advances available to the whole (digital) society.

These objectives are relevant to the main research projects in our agenda. For the sake of completeness the projects starting end 2007 or in 2008 are listed below as they do not appear in section 7.1:

- *Adaptive Combinatorial Search*, joint project with Microsoft Research (M. Schoenauer and Y. Hamadi, Microsoft Research Cambridge, coordinators), aims at the automatic tuning of search and optimization algorithms, from heuristics to meta-heuristics (more precisely Evolutionary Computation). Both off-line and on-line tunings will be investigated.
- SYMBRION, an FP7 IP (Integrated Project), coordinated by Sergey Kornienko (Univ. Stuttgart), involving 10 partners from Robotics (Electronics and Mechanics), Evolutionary Biology, and Computer Science (working on bio-inspired complex systems). Integrating hardware and software design, Symbrion IP aims at designing autonomous swarm robots. The software will involve both time-scales of evolutionary learning and on-line learning, in direct connection with TAO research themes.
- Observatory of the EGEE Grid (WP in EGEE-III, C. Germain coordinator). Resuming earlier studies in EGEE-II, the Grid Observatory will integrate the collection and publication of traces of the EGEE grid and users with the development of models and ontology of the domain knowledge.
- DigiBrain, a Digiteo project, coordinated by J.-D. Muller (CEA-LIST) and M. Sebag, involves CEA, INRIA, LRI and Neurospin (S. Dehaene and J.-B. Poline). DigiBrain will investigate EEG-based interfaces for neuroscience studies and Human Computer Interaction.

[1] Geoffrey A. Moore, *Crossing the Chasm: Marketing and Selling High-Tech Products to Mainstream Customer*, 1991

3.2. Deep Representations

The last years have seen some breakthroughs related to the ML search space, including Deep Networks (DN[2]), Echo State Networks (ESN[3]) or Liquid State Machines (LSM[4]). These frameworks offer a compact or sparse coding of complex functions (e.g. Deep Networks require a logarithmic vs exponential number of units to encode the canonical parity problem). While they have long been discarded as their full training is a hopelessly ill-behaved optimization problem, new heuristics have been shown to achieve efficient learning although their theoretical properties are not yet fully understood. For instance DNs consider a sequence of tractable optimization problems (unsupervised learning), iteratively growing the network and reaching a region where the supervised learning problem becomes “reasonably” tractable[5]. In ESNs, a compact representation (sparse graph + set of weights) is mapped onto a complex function space (dynamic systems and limit cycles).

Interestingly, the distinction between the search space (referred to as genotypic space) and the solution space (referred to as phenotypic space) has long been identified as a main source of effectiveness for Evolutionary Computation[6]. The merits of the distinction between genotypic and phenotypic spaces can be illustrated by the so-called developmental representations, a prototype of which is the Cellular Encoding [7], more recently extended toward embryogenic representations based on various biological models, from plant growth to Genetic Regulatory Networks. Developmental representations map a compact search space (parametric or tree-structured, e.g. programs) onto a complex, usually non parametric, solution space (e.g. analog circuits).

Taking advantage of the statistical learning and evolutionary cultures in TAO, our first research objective will be to analyze and study the diverse frameworks enabling a **compact description of complex solutions through procedural heuristics**, referred to as Deep Representations (DRs). The theoretical study will focus on the following two aspects:

- Expressivity/tractability tradeoff.
  DRs allow a huge solution space to be searched through exploring a comparatively restricted and well identified search space, which either offers some performance guarantees, or was found the only feasible way to obtain any result at all[8]. Tools from statistical ML, e.g. covering numbers, will be used to analyze the genotypic/phenotypic mappings in the EC literature.
- Stability/versatility tradeoff.
  The feasibility of learning/optimization requires some stability of the search landscape, meant as most genotypic changes result in little phenotypic differences. In the meanwhile, the search space should offer “sufficiently” many shortcuts toward various regions of the solution space. This property, referred to as versatility, implies that additional information enables efficient jumps in the phenotypic space, making the most of efficient active learning or exploration strategies. The stability/versatility tradeoff will be studied in the spirit of active learning[9].

---

[2] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle. Greedy Layer-Wise Training of Deep Networks. NIPS 2006, pp 153–160, MIT Press, 2007.

[3] H. Jaeger. The "echo state" approach to analysing and training recurrent neural networks. German National Research Center for Information Technology, technical report GMD Report 148, 2001.

[4] NIPS 2006 Workshop on Echo State Networks and Liquid State Machines. H. Jaeger and H. Haas and J. C. Principe Eds. 2006.

[5] Notably, TAO has done some early works along this line of research. Cascade mechanisms embedded in logical learning enables the construction of disjunctive hypotheses while handling conjunctive ones, in a transparent and cost-effective way (M. Sebag and M. Schoenauer, Incremental Learning of Rules and Meta-Rules. In B. Porter and R. Mooney, Eds, Proc. ICML90, Morgan Kaufmann, 1990); handling a sequence of objective functions within a population-based optimization scheme, thanks to the Behavioral Memory mechanism, was shown effective to tackle optimization problems which could not be addressed up front (M. Schoenauer and S. Xanthakis, Constrained GA Optimization. In S. Forrest, Ed., Proc ICGA'93, pp 573-580, Morgan Kaufmann, 1993)

[6] R. C. Lewontin, *The Genetic Basis of Evolutionary Change*, Columbia University Press, 1974.

[7] F. Gruau. Neural Network Synthesis using Cellular encoding and the Genetic Algorithm. PhD thesis, Ecole Normale Superieure de Lyon, 1994

[8] J. Koza, *Genetic Programming III: Automatic Synthesis of Analog Circuits*. MIT Press, 1999

[9] S. Dasgupta. Coarse Sample Complexity Bounds for Active Learning, NIPS'05, MIT Press, 2005).

In an application perspective, the search for deep representations is relevant to the on-going Gennetec project (investigating Gene Regulatory Networks in a Genetic Programming perspective), and to Symbrion IP (as the target representation should allow learning at different time scales, e.g. involving both evolutionary optimization and on-line learning).

3.3. Generalized Bandits

The blossoming use of Multi-Armed Bandit (MAB) algorithms to revisit reinforcement learning[10], tree search[11] including games [3], [6], [27] (see section 6.2.3), optimization[12] is explained from two factors. On the one hand, MABs aim at minimizing the regret, i.e. the cumulative loss over the oracle strategy; this elegant criterion is amenable to theoretical bounds; furthermore, it is relevant in an any-time perspective, whereas many theoretical studies mostly focus on asymptotic performances (see also section 3.4). On the other hand, the decision making process achieved by MABs enforce an Exploitation vs Exploration (EvE) tradeoff; a wide variety of Exploration-related criteria has been considered in the literature, with and (most often) without theoretical justifications.

Several extensions of MAB algorithms and analysis have been identified as theoretical and applicative priorities on our research agenda:

- A first extension is required to deal with **dynamic environments**, relaxing the assumption of iid rewards for each option (bandit arm). Let us consider for instance the EvE tradeoff at the core of evolutionary computation, of game strategies [3], of news recommendation[13]; the rewards associated to a given option (respectively, variation operator, game move or news item) are not stationary; they evolve while the search or the game goes on, or as the user's needs and mood change. Some algorithmic advances have been made in Tao (development of MoGo, winning participation to the Pascal Challenge on Online Trading of EvE), extending the standard Upper Confidence Bound algorithms to handle non stationary environments. Further work is required to extend MAB algorithms to Monte-Carlo-based planning (as an alternative to dynamic programming) and to provide theoretical guarantees on the global solution quality.
- A second extension regards **multi-variate bandits**. In quite a few application domains, some side information is available (e.g. the user profile in a news recommendation context) and can be used to handle the EvE tradeoff more efficiently. In the MoGo system, the so-called RAVE (*Rapid Action Value Estimate*) heuristics provides additional estimates of the move values; significant improvement of MoGo has been obtained by exploiting this additional, most often strongly biased, side information. Notably, multi-variate bandit algorithms have been acknowledged a prioritary research direction in the PASCAL-2 roadmap[14]. Further study is required to both design more efficient multi-variate bandit algorithms, and provide theoretical guarantees thereof.
- Thirdly, the extension of MAB algorithms to the **bounded rationality** framework, e.g., increasing the number of options and considering a short-term time horizon, is a both theoretical and applicative challenge. Quite a few application domains involve many options (e.g. circa 400 arms in computer-go, and infinitely many in continuous frameworks); further more, in games or planning, the stress is put on the short term performances (as opposed to, the asymptotic ones). We have developed efficient anytime algorithms, extending Berry et al.[15] in the so-called easy setting where the reward

[10] P. Auer and R. Ortner. Logarithmic Online Regret Bounds for Undiscounted Reinforcement Learning. NIPS'06 pp 49–56, MIT Press, 2007

[11] A typical example is UCT: L. Kocsis and C. Szepesvári. Bandits-based Monte-Carlo planning. In J. Fürnkranz et al., eds., ECML'06, pp 282-293, LNAI 4212, Springer Verlag, 2006.)

[12] P.-A. Coquelin and R. Munos, Bandit Algorithms for Tree Search, Technical report INRIA 6141, 2007.

[13] Cédric Hartland, Sylvain Gelly, Nicolas Baskiotis, Olivier Teytaud, and Michele Sebag. Multi-armed bandit, dynamic environments and meta-bandits, *Online Trading of Exploration and Exploitation Workshop, NIPS*, 2006.

[14] The PASCAL Network of Excellence (2003-2008) will be continued in the FP7 framework (2008-2013).

[15] D. A. Berry, R. W. Chen, A. Zame, D. C. Heath, and L. A. Shepp. Bandit problems with infinitely many arms, *Ann. Stat.* **25**(5):2103-2116, 1997.

distribution is favorable (many arms have a reward probability close to 1); we also proposed original heuristics in the so-called difficult setting (the reward probabilities are in $[0, \epsilon]$, with $\epsilon < 1$). While the empirical efficiency of these algorithms has been found satisficing, their theoretical fundations remain to be established.

- A fourth research perspective related to MAB is concerned with **multi-objective settings**. For instance in autonomous robot control, every option can be assessed along several criteria, such as its value (to which extent the option is instrumental to reach the robot goal) and its risks (to which extent the option can dammage the robot integrity). Although multi-objective optimization can always be cast as a mono-objective optimization problem (e.g. classically considering the weighted sum of the objective functions as single objective), it is believed that multi-objective bandits correspond to a relevant and daring extension of MAB algorithms. On the one hand, this extension aims at finding optimal, e.g. controlled risk-taking, decision strategies; on the other hand it requires to extend the regret definition (e.g. cumulative distance to the Pareto front).

These research themes are directly relevant to *Microsoft-TAO project*, SYMBRION IP and *Autonomic Computing*.

Typically, the online learning of hyperparameters tackled by *Microsoft-TAO project* can be formalized as a MAB problem; bounded rationality is similarly relevant at least in the calibration stage of the current application. The multi-objective aspect is also relevant as the algorithmic performance can usually be assessed along quite a few independent criteria (e.g. time-to-solution, memory usage, solution accuracy).

Independently, optimal decision making under bounded resource constraints is relevant to SYMBRION IP. Likewise, the mid-term goal of Autonomic Computing is to deliver satisficing job schedulers. In both cases, as experiments are done *in situ* the learning algorithm must find a way to preserve the system integrity, and limit the risks incurred by any system unit.

3.4. Shifting to modern computing

Notable breakthroughs have revolutionized computer science in the last decade, along several perspectives: hardware, communication/networks, and usage. With the evolution of means and ends (multi-core/multi-thread machines; grid systems; distributed/ubiquitous computing; resource-aware/anytime/scalable algorithms; ambient/pervasive intelligence), the software world gradually becomes more aware that new algorithmic paradigms are required to make the most of new architectures, to handle new demands, and to sail towards New Intelligence Frontiers[16].

While TAO is already interested in modern computing as a source of applications (e.g. through *Autonomic Computing* and our collaboration with Alchemy [31]), this third direction of research aims at developing a different approach of algorithms, along the *Computational Thinking*[17] perspective. Formally, Computational Thinking is meant to revisit the art of problem setting and algorithm design, through a broader view of computing systems.

A first theme in this research direction aims at the global optimization of the information processing chain[18]. Precisely, learning and optimization algorithms most often involve elementary processes in charge of i) information selection (e.g., active learning, feature construction, or population-based sampling); ii) model construction (learning, building surrogate models, covariance matrix adaptation); iii) optimization (assessing the current result and going further; tuning or adapting the hyper-parameters involved in the other two tasks).

[16] While this new world struggles with the same issues as the old one (scalability, reliability, communication efficiency), the paradigm shift can be best understood by analogy with Herbert Simon's bounded rationality. Bounded rationality indeed faces the same problems as good old rationality, except for one additional, distinctive, feature: the reasoning cost now is an element of the reasoning process. This simple modification entails the loss of equilibrium results; the overwhelming importance of initial conditions; the existence of information asymmetry; in brief, this modest modification leads to a significantly different Economics Theory.

[17] Term coined by Pr. Jeanette Wing, CMU. "Computing in the 21st Century", Microsoft Research Asia Conferences, Nov. 2005. Cited in Towards 2020 Science, Microsoft White Book.

[18] Reminding that the optimal chain is not usually obtained by combining elements that have been optimized independently.

While these processes have generally been considered independently, a fundamental issue is to handle their interactions, and for instance to be able to optimally (dynamically) allocate the computational efforts among them. Preliminary studies along this line, considering the adaptation of MoGo to highly parallel architectures, have started in collaboration with Bull. A longer-term research, concerned with automatic adjustment, is the goal of next section.

Secondly, Computational Thinking advocates the ability of dealing with large complex (algorithmic) systems without understanding every detail thereof, which shall be referred to as Smart Black Box Design. This second theme of research is relevant to the search of Deep Representations and the Gennetec project, already mentioned in section 3.2. It also includes the studies related to complex systems (section 6.4), where tools from statistical physics are used to provide a manageable model at the macro-scale of a system described at the micro-scale. Last, but not least, the DigiBrain project aims at creating the means for an effective interaction with the user without spelling the interaction rules.

3.5. Crossing the Chasm

Countless studies have underlined the fact that many a good scientific and algorithmic advances fail to make it outside research labs. Such failures are often blamed on the Knowledge Barrier, the entry ticket people have to pay in order to master new techniques because of their many options and parameters, because of their flexibility and versatility. Parameter adjustment is at the core of most if not all TAO applicative studies in Numerical Engineering [5].

Considerable efforts are deployed to anticipate this usage barrier when designing algorithms or devices. One approach aims at building "plug-and-play" variants, allowing naive users to benefit from 90% of the algorithm potentialities. It is generally deemed however that some form of adaptation to the problem at hand is required to deliver robust results; in most application domains, the search for "the" killer algorithm has been acknowledged to be hopeless. Alternatively, one might want to build a meta-layer topping a set of algorithms, selecting the best algorithm to use depending on the problem at hand[19] the bottleneck along this line has been identified as the description of the problems at the meta-level.

The fourth research direction of TAO will investigate two approaches for Crossing the Chasm, respectively based on online and offline hyperparameter tuning/learning.

The online tuning approach includes both self-adaptation heuristics (dating back to the 90's[20]), and the original use of multi-armed bandit algorithms which has been discussed in section 3.3. While self-adaptation and specifically Covariance-Matrix Adaptation ES is acknowledged a major breakthrough in continuous evolutionary optimization[21], TAO has a world-wide theoretical and applicative expertise on CMA-ES. Several extensions of self-adaptive evolution will be investigated, aiming at bridging the gap between convex optimisation and continuous evolutionary computation: the convergence of the CMA-ES algorithm, along the lines of that of the SA-ES algorithm[22]; a separable version of CMA-ES (can separability be detected – and taken advantage of?); the handling of different types of constraint (at the moment, only bound constraints can be set on the variables); an efficient version of multi-objective CMA-ES (as the existing version does not really use multi-objective information, and poorly samples the Pareto Front). Surrogate versions of CMA-ES (and other stochastic optimization algorithms) will also be considered, in particular within the Digiteo PhD grant "Simplified models in Numerical Engineering", that has been allocated to be co-supervised by M. Sebag and J.-M. Martinez, SF2MS, CEA (2007-2010).

[19] Meta- Learning project, ESPRIT project 26.357 (2002-2004).

[20] See [5] and other chapters of *Parameter Setting in Evolutionary Algorithms*, Springer Verlag 2007.

[21] Hansen, N. and A. Ostermeier (2001). Completely Derandomized Self-Adaptation in Evolution Strategies. Evolutionary Computation, 9(2), pp. 159-195

[22] Anne Auger. Convergence results for (1,λ)-SA-ES using the theory of φ-irreducible markov chains. *Theoretical Computer Science*, 334(1-3):35–69, 2005.

The second approach, concerned with offline learning, resumes the Phase-Transition studies pioneered by TAO in collaboration with L. Saitta and A. Giordana (Univ. del Piemonte Orientale), (see section 6.1). Based on the appropriate order parameters, the phase transition framework allows for building the "competence map" describing the average algorithm behaviour. Clearly, such competence maps make it easy to achieve meta-learning[23] and decide for each problem which algorithm/setting will be the most appropriate on average. Indeed similar approaches have been developed for CSP solvers[24], exploiting a range of indicators developed over the years as order parameters.

The fundamental bottleneck, i.e. designing relevant order parameters, has been tackled empirically so far. Further studies will use the statistical physics tools[25] to construct phase diagrams (physics equivalent of competence maps). Recent results at the frontier of statistical physics and computer science [26] have bridged together optimization (free energy optima, K-SAT solutions,...) and learning with message passing algorithms (belief propagation, affinity propagation[27]). In addition[28] an algorithmic hierarchy (warning propagation, survey propagation) appears to be in correspondence to the hierarchy provided by the order parameter of the mean-field theory. These results are synthesized in a phase diagram; the purpose of new adaptive heuristics will be to infer the position of the instance problem in the phase diagram.

This direction of research subsumes several on-going projects, specifically Evotest and *Microsoft-TAO project* where the goal is to produce off-the-shelf algorithms and to enforce technology transfer. In contrast, many previous [8] or current projects (e.g. GENNETEC) have enforced technology transfer ... manually.

4. Application Domains

4.1. Application Domains

The mainstream applications of TAO are, since its creation, autonomous robot control, medical applications, and engineering applications. Software robotics is still an active application domain (see section 6.4.2, and several applications involving numerical engineering are on-going (section 6.4.1).

However, most medical applications have been hindered as the dialogue with experts could not take place to the required extent. In the neighbour domain of neurosciences (in the context of Neurodyne ACI), the collaboration was effective[29], and led to start the DigiBrain project, concerned with Brain Computer Interface, within the Digiteo structure, involving LRI, INRIA, CEA List and Neurospin, and starting in 2008.

Moreover, two new fields of applications have appeared, due to the arrival of new staff on the project:

- Autonomic Computing applications have started to become an important part of TAO applications, since Cecile Germain-Renaud, now Professor at Université Paris-Sud, joined TAO in 2005, and Balázs Kégl was hired at LAL in Orsay and became a full associate member of the team (see section 6.4.3).
- the study of Social Systems, more precisely economical multi-agent models and road traffic models, have started at TAO, following the arrival of respectively Philippe Caillou and Cyril Furtlehner.

[23] **J. Maloberti and M. Sebag**. Fast theta-subsumption with constraint satisfaction algorithms. *Machine Learning Journal*, 55:137–174, 2004.

[24] F. Hutter, Y. Hamadi, H. Hoos, and K. L. Brown. Performance prediction and automated tuning of randomized and parametric algorithms. In CP'06, pp 213–228, 2006.

[25] M.Mézard, G.Parisi, and M.A. Virasoro, *Spin glass theory and beyond*, World Scientific, 1987.

[26] J. S. Yedidia, W. T. Freeman, and Y. Weiss, *Generalized belief propagation*, Advances in Neural Information Processing Systems (NIPS 2001).

[27] B.J. Frey and D. Dueck. Clustering by Passing Messages Between Data Points Science (2007) Vol. 315. pp. 972 - 976.

[28] M. Mézard and R. Zecchina, The Random K-satisfiability problem : from an analytic solution to an efficient algorithm, 2002 *Phys. Rev. Letters E* **66**

[29] Michele Sebag, Nicolas Tarrisson, Olivier Teytaud, Julien Lefevre, and Sylvain Baillet. A multi-objective multi-modal optimisation approach for mining stable spatio-temporal patterns. In *20th Int. Joint Conf. on Artificial Intelligence (IJCAI-07)*, pages 859-864, 2007.

All those applications are described in detail along the text, most of them in section 6.4.

5. Software

5.1. MoGo

Keywords: *Go*, *Multi-armed bandit*.

Participants: Olivier Teytaud [correspondent], Sylvain Gelly, Jean-Baptiste Hoock, Arpad Rimmel.

MoGo (55000 lines of code in C++) is currently one of the best Computer-Go programs worldwide (including advanced options as multithreading and now message-passing parallel version) (see Section 2.3. Only the binary code is released, with hundreds of downloads[30]. Many computer-go programmers discuss their experiments with MoGo on the computer-go mailing-list.

5.2. GUIDE: A graphical interface for Evolutionary Algorithms

Keywords: *Evolutionary Computation*, *GUI*, *Java*, *Object-oriented*.

Participants: Marc Schoenauer [correspondent], Luis DaCosta.

Abstract: GUIDE is a graphical user interface for easy Evolutionary Algorithm design and coding. It allows the user to describe its genome (the structure that will evolve) graphically, represented as a tree, using containers and elementary types (booleans, integers, real numbers and permutations). All representation-dependent operators (initialization, crossover and mutation) can then be defined either using default values, built bottom-up from the elementary types, or user-defined operators. Developing a prototype for a new search space involving complex structures has now become a matter of minutes.

GUIDE was programmed in JAVA by James Manley during the 6 months of his DESS stage in 2004. It is a follow-up of a previous tool developed in collaboration with Pierre Collet in the DREAM project (http://www.dcs.napier.ac.uk/~benp/dream/dream.htm).
GUIDE has been chosen as the evolutionary basis for the EvoTest project: testing a given program means feeding it with data of a specific structure. In this context, the goal of EvoTest is to automatically evolve test data, relying on an automatic code generator that only requires a description of the structure of the data to evolve – and this is precisely what GUIDE is doing.

The main changes in GUIDE in 2007 have been a complete cleaning of the code (preserving, but not increasing, the functionalities). Consequently, it is now much easier to generate code for different libraries, and the next version will handle EO as well as ECJ (Evolutionary COmputation in Java). Moreover, because of the arrival of Luis DaCosta as developper, GUIDE is now available on GForge as an Open Source software [18]. Also, to fit in the complete loop of Automated testing, GUIDE has been interfaced with the partners' systems through a TPTP interface.

5.3. Simbad Autonomous and Evolutionary Robotics Simulator

Keywords: *Java*, *evolutionary robotics*, *robot simulation*.

Participant: Nicolas Bredèche [correspondent].

[30] http://www.lri.fr/~gelly/MoGo_Download.htm

Abstract: Simbad is an open source Java 3D robot simulator for scientific and educational purposes (Authors: Louis Hugues and Nicolas Bredèche). Simbad embeds two stand-alone additional packages: (1) a Neural Network library (PicoNode) and (2) an Artificial Evolution Engine (PicoEvo). The Simbad package is targeted towards Autonomous Robotics and Evolutionary Robotics for research and education. The packages may be combined or used alone. In the scope of Research in Evolutionary Robotics, the Simbad package package helps quick development of new approaches and algorithms thanks to the complete and easy-to-extend libraries. Real-world interface can be easily written to transfer simbad controllers to real robots (the Khepera interface is available). The open source nature of the project combined with easy-to-understand code makes it also a good choice for teaching Autonomous and Evolutionary Robotics. Simbad is used in several AI and robotics courses: IFIPS engineering school (4th and 5th year) ; Master 1 at Université Paris-Sud ; Modex at Ecole Polytechnique.
Please refer to : http://simbad.sourceforge.net/.

5.4. PuppetMaster - Generic 3D Robotics Simulator

Keywords: *physics-based engine*, *robot simulation*.

Participant: Alexandre Devert [correspondant].

Abstract:

PuppetMaster is an open source C++ 3d robotic simulation framework for scientific and educational purposes. It allows to describe simulation scenarios, robot morphologies and behaviors as a C++ plugin. A visualizer makes it possible to see a plugin in action. The simulation is based on realistic physical simulations, so the range of the representable robots and simulations scenarii covers all the practical cases. It allows rapid prototyping of both control algorithm and robot morphology. Combined with a numerical optimization framework, it allows fully automatic design of robots, with simulations scenarios as fitness measure. PuppetMaster was used to design a robot control algorithm independent from the morphology, allowing tests on snake-like and multi-legged robots.

5.5. Django

Keywords: *Fast theta-subsumption*.

Participant: Michèle Sebag [correspondent].

Abstract: Django is an algorithm of theta-subsumption of Datalog clauses, written in C by Jerome Maloberti and freely available under the GNU Public License. This algorithm is an exact one, with a gain of two or three orders of magnitude in computational cost over other theta-subsumption algorithms. Django uses Constraint Satisfaction techniques such as Arc-Consistency, Forward-Checking and M.A.C. (Maintaining Arc-Consistency) and heuristics based on the First Fail Principle.

Django has been widely used and cited in the literature (coll. with the Yokohama University, Japan, U. of Tufts in Arizona, USA, U. of Bari, Italy).

http://tao.lri.fr/TikiWiki/tiki-index.php?page=Django/.

5.6. OpenDP

Keywords: *Learning*, *Object-oriented*, *Stochastic Dynamic Programming*.

Participants: Olivier Teytaud [correspondent], Sylvain Gelly.

Abstract: OpenDP is an open source code for stochastic dynamic programming[31], based upon the use of (i) time-decomposition as in standard dynamic programming (ii) learning (iii) derivative-free optimization. Its modular design was meant to easily integrate existing source codes: OpenBeagle (with the help of Christian Gagné), EO (with the help of Damien Tessier), CoinDFO, Opt++, and many others, for optimization; the Torch library and the Weka library and some others for learning. It also includes various derandomized algorithms (for robust optimization and sampling); other algorithms (e.g. time-pca and robotic-mapping) are underway. OpenDP has been experimented on a large set of benchmark problems (included in the environment), allowing for an extensive comparison of function-values approximators and derivative-free optimization algorithms with a tiny number of iterates.

The merit of the OpenDP platform is twofold. On one hand, while many of the above algorithms are well-known, their use in a dynamic programming framework is new. On the other hand, such a systematic comparison of these algorithms on general benchmarks did not exist in the literature of stochastic dynamic programming, where many papers only consider one learning method, not necessarily in the same conditions than other published results. These thorough experimentations inspired some theoretical work in progress about the criteria for learning in dynamic environments, noting that cross-validation is neither satisfactory (for example the σ^2 parameter in Gaussian SVM chosen by cross-validation is usually too small in the context of dynamic programming) nor fast enough in that framework.
See main page at http://opendp.sourceforge.net/.

6. New Results

6.1. Stochastic Optimisation for ML and DM

Keywords: *AUC-based Learning*, *Bounded Relational Reasoning*, *Constraint Satisfaction and Phase Transition*, *Feature Selection*, *Human Computer Interaction and Visual Data Mining*, *Inductive Logic Programming*, *Meta-learning and Competence Maps*, *Methodological aspects*, *Phase Transitions*.

Participants: Nicolas Baskiotis, Antoine Cornuéjols, Cécile Germain-Renaud, Marc Schoenauer, Michèle Sebag, Olivier Teytaud, Xiangliang Zhang.

6.1.1. Representation, Feature Selection, and Learning Criteria

At the core of Machine Learning is the representation of the problem domain. Building an appropriate representation, aka Feature Extraction[32] or Constructive Induction, includes Feature Selection and Feature Construction.

Some results previous to TAO regarding Feature Selection exploits the stochasticity of EC-based learning: The ROGER algorithm (*ROC-based GEnetic learneR*)[33] is based on the evolutionary optimisation[34] of the Area Under the ROC Curve (AUC) criterion, equivalent to the Wilcoxon-Mann-Whitney statistic. The ensemble of the various hypotheses provided by independent ROGER runs is used for sensitivity analysis and to achieve feature selection[35].

Another kind of ensemble-based feature selection has been recently devised and applied for DNA microarrays analysis [24], focusing on the notion of Type I and Type II errors (distinguishing relevant and irrelevant features using feature rankings based on independent criteria)[36]

[31] Sylvain Gelly and Olivier Teytaud. Opendp: a free reinforcement learning toolbox for discrete time control problems. In *NIPS Workshop on Machine Learning Open Source Software*, 2006.

[32] I. Guyon, S. Gunn, M. Nikravesh, and L. Zadeh, eds. Feature Extraction, Foundations and Applications. Physica-Verlag, Springer, 2006.

[33] M. Sebag, J. Azé, and N. Lucas. Roc-based evolutionary learning: Application to medical data mining. In Xin Yao et al., eds, Proc. PPSN 2004, LNCS 3242, pages 384–396. Springer Verlag, 2004.

[34] While it has been shown by Thorsten Joachims that Support Vector Machines were amenable to AUC optimisation (Int. Conf. on Machine Learning, 2005, best paper award), in practice only greedy optimization is performed for the sake of tractability.

[35] K. Jong, E. Marchiori, and M. Sebag. Ensemble Learning with Evolutionary Computation: Application to Feature Ranking. In Xin Yao et al., eds, Proc. PPSN 2004, LNCS 3242, pages 1133–1142. Springer Verlag, 2004.

A new criterion for graphical model learning, stressing the graph structure complexity, has been proposed in Sylvain Gelly's PhD [3]; the advantage of this criterion in terms of learning consistency has been demonstrated in the specific but applicatively relevant case of a small learning sample when the graph structure is not the true one.

In the context of unsupervised learning, a new latent-clustering based criterion has been proposed in the SELECT project-team, and tackled by TAO using evolutionary approaches [35].

Finally, AUC-like learning criteria are being considered in Arpad Rimmel's PhD, aimed at handling imbalanced classification problems with many more features than examples, motivated by chemometry applications. The dialogue with the experts in the applicative context (Accamba ANR) did not permit the assessment of the approach up to now.

6.1.2. Hypothesis search space

The great applicative successes of Support Vector Machines [37] are partly explained as prior knowledge about the problem domain can be rigorously encapsulated in the (manually designed) kernel. Previous works related to the use of prior knowledge in TAO, such as Carlos Kavka's PhD[38], aimed at merging the best of two worlds: the expert provides his knowledge (specific fuzzy rules), the scope of which is automatically determined and optimized using EC.
In a more theoretical perspective, the feasibility of learning in higher-order logic spaces has been investigated in an average-case perspective [14]; a new framework has been developed to study the expected undecidability (the probability of meeting undecidable clauses) and the convergence thereof along learning.

Another way of exploring the hypothesis space is based on ensemble learning, in the hope that the whole can perform better than the sum of its parts. Along this line, a multi-objective evolutionary ensemble learning approach has been proposed, leading to some insights into how to ensure the diversity of the hypotheses along evolution or in the final population, and how to select the best ensemble [23].

Independently, motivated by the search for active neural cell assemblies or relevant patterns (in the context of the ACI NeuroDyne), the spatio-temporal data mining of Magneto-Encephalographic datasets has been formalised as a multi-objective optimisation problem (finding large spatio-temporal areas with high signal correlation). An extension to multi-objective multi-modal optimisation was required to capture the several neural cell assemblies in interplay [4]. Interestingly, the search for discriminant patterns among the relevant patterns turns out to be significantly easier than directly searching for discriminant patterns[39].

6.1.3. Phase Transition and Competence Maps

ML can simultaneously be viewed as an optimisation problem and a constraint satisfaction problem (CSP). Inspired from the Phase Transition paradigm developed in the CSP community since the early 90s, Lorenza Saitta and Attilio Giordana have been studying relational learning after some order parameters; some implications on the limitations of existing relational learners have been demonstrated in an early collaboration with TAO members (1999). The order parameters define a landscape, enabling to depict the average behaviour of any related algorithm through a *Competence Map*. The comparison of the competence maps related to various θ-subsumption algorithms was shown instrumental to building a meta-layer, automatically selecting the best (on average) algorithm depending on the problem instance at hand[40]. The approach has been applied to propositional decision tree learning in Nicolas Baskiotis' PhD[41] and to grammatical inference [4]. In the latter case,

[36] A Pascal workshop related to Type I and Type II errors, *Multiple Simultaneous Hypotheses Testing*, has been organized by O. Teytaud et al. in May 2007.

[37] V. N. Vapnik, *Statistical Learning Theory*, J. Wiley, 1998.

[38] Carlos Kavka. *Evolutionary Design of Geometric-Based Fuzzy Systems*. PhD thesis, Université Paris-Sud, July 2006.

[39] Vojtech Krmicek and Michèle Sebag. Functional brain imaging with multi-objective multi-modal evolutionary optimization. In Th. Runarsson et al., eds, Proc. PPSN 2006, LNCS 4193, pages 382–391. Springer Verlag, 2006.

[40] Fast Theta-Subumption with Constraint Satisfaction Algorithms, J. Maloberti and M. Sebag, Machine Learning Journal, 2004, pp 137-174.

[41] C4.5 Competence Map: a Phase Transition-inspired Approach, N. Baskiotis and M. Sebag. In Proc. Int. Conf. on Machine Learning, ICML 2004, Morgan Kaufman pp 73–80.

some unexpected biases of prominent learners, e.g. RPNI, have been discovered and tentative explanations have been provided. The link between CSP and linear programming and Support Vector Machines has been further investigated in Romaric Gaudel's PhD, leading to new bounds on the generalization error. The merit of the approach is twofold: it provides lower bounds which are applicable for small sample size [25].

6.2. Machine Learning for Stochastic Optimisation

Participants: Anne Auger, Sylvain Gelly, Nikolaus Hansen, Mohamed Jebalia, Marc Schoenauer, Michèle Sebag, Olivier Teytaud.

This research direction investigates how the theoretical and algorithmic body of knowledge developed in Machine Learning can advance the fundamental study of stochastic optimisation, extend its scope and support more effective algorithms. Three types of contributions have been made; the first one is related to the theoretical study of Estimation of Distribution Algorithms; the second one considers surrogate optimisation, extending EC to deal with computationally expensive objective functions; the third one is related to Optimal Decision Making, revisiting dynamic programming and investigating multi-armed bandit algorithms.

6.2.1. Fundamental studies of evolutionary algorithms.

Estimation of Distribution Algorithms (EDAs) evolve a probability distribution on the search space by repeating a series of sampling/selection/learning steps. Using Statistical Learning theory, EDAs have been studied in the context of expensive optimisation problems allowing only a small numbers of iterates [3], and some optimal (w.r.t. robustness) comparison-based EDAs have been proposed [7]. Note that this research is closely related to that on surrogate models decribed in next section 6.2.2.
Genetic Programming (GP) extends the Evolutionary Computation paradigm to tree-structured search spaces, essentially the space of programs. In this field where theory still lags far behind practice, several classical results of Statistical Learning theory have allowed to delineate the applicability of this technique, and to derive sufficient conditions on the penalty term used in practice to limit the uncontrolled growth of the solution (aka *bloat*)[42]; more recently, necessary conditions have been established, barring the use of some heuristics as inconsistent (with no guarantee of asymptotic convergence to the optimal solution) [37].

More generally, several techniques borrowed from Machine Learning and Complexity Theory have been applied to theoretical investigations of Evolutionary Algorithms. This includes results on the consistency of halting criteria and sufficient conditions for convergence in non-convex settings [36]; the non-validity of the No-Free-Lunch Theorem in continuous optimisation [11]; some limitations of multi-objective optimisation without feedback from the user [10]; an analysis of the parametrisation of the computational effort in stochastic optimisation [29], [38]; and studies of different ways to use quasi-random points in Evolutionary Algorithms [39].

6.2.2. Approximations of the fitness function.

Evolutionary Algorithms are known to be computationally expensive. They are hence particularly concerned with what Mechanical Engineers have called Response Surface Methods, revisited in the last few years as Surrogate Models methods. The idea is to build an approximation of the objective function, and to run the optimisation algorithm (whatever the algorithm) on the approximation rather than on the original function. A first crucial issue is the choice of the approximation model. And, because the approximation has to be updated as the search proceeds, another important issue is how often this update has to be done.

Within the ANR/RNTL OMD project, TAO is in charge of the technology transfer related to surrogate methods, motivated by the expensive benchmark problems of the industrial OMD partners (Dassault, Renault and EADS), using in particular the surrogate-based version of CMA-ES[43] as in [32]. In particular, TAO contribution to OMD includes the port to Scilab of both the original and the surrogate versions of CMA-ES.

[42] S. Gelly, O. Teytaud, N. Bredèche, and M. Schoenauer. Universal consistency and bloat in GP.*Revue d'Intelligence Artificielle*, 20(6):805–827, 2006.

[43] Kern, S., N. Hansen, P. Koumoutsakos. Local Meta-Models for Optimisation Using Evolution Strategies. In Th. Runarsson, ed., *Proc. PPSN IX*, pp.939-948, LNCS 4193, Springer Verlag, 2006.

6.2.3. Approximate Dynamic Programming and Multi-Armed Bandit Problems

In some problems, the goal is to find the optimal policy in the sense of some (delayed) reward function; an intermediate step thus is to learn the value function, associating to each problem state the reward expectation. Dynamic programming, a sound and robust approach dating back to the 50's, suffers from the curse of dimensionality[44] and Monte-Carlo planning approaches are being investigated to address this limitation. TAO has been working in both areas.

OpenDP[45] is an open source platform for approximate dynamic programming[46], which has been thoroughly benchmarked to assess diverse sampling, learning, optimization and frugal non-linear programming algorithms. Experimental comparisons have been reported in Sylvain Gelly's PhD [3], together with theoretical results related to deterministic, random and quasi-random sampling [40], [41].

Monte-Carlo planning approaches have been investigated, with the domain of computer-go as motivating application. The MoGo program[3], [6], [27], embedding a Monte-Carlo evaluation function within the Multi-Armed Bandit framework, currently is the best computer-go program[47] See also sections 2.3 and 5.1.

The Multi-Armed Bandit setting has been intensively studied in TAO from a theoretical [42] and applicative [27] perspective. The MAB extension to dynamic settings has been considered in relation with the Pascal Challenge *Online Trading of Exploration vs Exploitation*[48]; TAO won the challenge in 2006, emphasising the successful use of change point detection techniques [3][49]. Directions for further research are to extend the Multi-Armed Bandit algorithms underlying MoGo, and to address multi-variate, multi-objective bandit problems.

6.3. Representations and Prior Knowledge for Evolutionary Computation

Participants: Nicolas Bredèche, Alexandre Devert, Cédric Hartland, Fei Jiang, Miguel Nicolau, Marc Schoenauer.

One key issue in Evolutionary Computation is to design a representation and the associated variation operators, crossover and mutation, best suited to the problem at hand. In order to do so, the use of prior knowledge is highly recommended in practice. Resuming early work devoted to Structural Optimisation, the research done in TAO focuses on the design of scalable and modular representations. A promising direction in this respect is that of developmental representations.

Specifically in the domain of combinatorial optimization, all reported successes of EAs are based on the use of prior knowledge and domain-specific heuristics. An application in temporal planning have witnessed this, and further demonstrated the need for some characterisation of problem instances in order to facilitate the choice of the hyper-parameters.

6.3.1. Scalable and modular representations

Incorporating domain knowledge in Evolutionary Algorithms is mandatory after the No Free Lunch Theorem in boolean settings (and is a good idea in any other setting); the choice of the representation (and the associated variation operators, e.g. crossover and mutation) is the very first place where this can be done.

[44] W.G. Powell: Approximate Dynamic Programming; Solving the curses of dimensionality, John Wiley and Sons, 2006.

[45] Sylvain Gelly and Olivier Teytaud. Opendp: a free reinforcement learning toolbox for discrete time control problems. In *NIPS Workshop on Machine Learning Open Source Software*, 2006.

[46] http://opendp.sourceforge.net

[47] http://www.lri.fr/~gelly/MoGo.htm

[48] http://www.pascal-network.org/Challenges/EEC/

[49] Cédric Hartland, Sylvain Gelly, Nicolas Baskiotis, Olivier Teytaud, and Michele Sebag. Multi-armed bandit, dynamic environments and meta-bandits, *Online Trading of Exploration and Exploitation Workshop, NIPS*, 2006.

Marc Schoenauer's early work on Evolutionary Design included an original representation based on Voronoi diagrams[50]. This work led to a collaboration with EZCT Design and Architecture Research. A first application was to design chairs, and the results were exposed in different contemporary architecture exhibitions[51], including the Beaubourg modern art museum [47].

However, this approach has many drawbacks: its lack of flexibility makes it almost impossible to address the constructibility issue. An original representation directly handling construction plans of Lego-like structures was proposed in Alexandre Devert's PhD in 2006[52]. Despite its efficiency and modularity, this representation nevertheless scales up poorly with the number of elements used to build the structure. An embryogenic approach has hence recently been proposed. The idea is to evolve local rules for some "cells" that will exchange "chemicals", and the steady-state of those chemicals will describe the target structure. Preliminary experiments on evolving 2D images have shown its robustness compared to previous approaches, explained as it automatically adjusts the stopping criterion for the developmental stage [19].

The core of this developmental approach is a Neural Network, duplicated in all the 'cells' of the underlying substrata. Whereas the first work above uses standard NN evolution of topology [53], the design of an original procedure to evolve Echo State Networks (ESNs) proved to greatly increase the speed of evolution [20]. Note that ESNs have also been used in TAO as models for robot controllers [28] – see section 6.4.2.

An on-going collaboration within the MIT-France program aims at further extending developmental systems in the context of architectural design. Ongoing work focus on applying this approach in the domain of truss structure design (phd of A. Devert) as well as evolving a developing multi-cellular system in a continuous substrate (work of N. Bredeche).

6.3.2. Toward developmental representations

Gradually shifting toward complex system modelling, we investigated, in collaboration with the Alchemy INRIA project-team, the influence of the topology of large Neural Networks on their computational ability, with the co-supervised PhD thesis of Fei Jiang. First studies investigated the influence of topology of SOM networks on their classification performances [31]. On-going work is concerned with Echo State Networks and their use in Engineering and Control problems.

In the meantime, on-going study on Gene Regulatory Networks (GRNs) provides yet another source of diversity for representation of (Neural) Networks: within the GENNETEC project, and starting from Banzhaf's model of GRN[54], Miguel Nicolau investigates the links of GRN models with Complex Neural Networks, related to both developmental representations and the abovementioned network topology studies.

6.3.3. Representations for combinatorial optimisation

The representation issue also arises for combinatorial optimisation problems, as witnessed by the new paradigm for Evolutionary Temporal Planning developped in TAO.

This original representation, based on a sequential splicing of the problem in the state space, was designed during a collaboration with Thalès [9]. The idea is to use a deterministic constaint programming solver (CPT[55] was chosen here) to solve the (hopefully) small problems between two states of the spliced sequence. This approach allowed us to solve yet unsolved instances of well-known benchmarks of the IPC suite in Jacques

[50] H. Hamda and M. Schoenauer. Topological Optimum Design with Evolutionary Algorithms. *Journal of Convex Analysis*, 9, pp 503–517, 2002

[51] P. Morel, H. Hamda, and M. Schoenauer. Computational chair design using genetic algorithms. *Concept*, 71(3):95–99, 2005.

[52] A. Devert, N. Bredèche, and M. Schoenauer. Blindbuilder : a new encoding to evolve lego-like structures. In Proc. EuroGP'06, pp 61–72, Springer Verlag LNCS 3905, 2006.

[53] K. O. Stanley and R. Miikkulainen. Evolving Neural Networks through augmenting topologies. Evolutionary Computation 10(2):99-127, 2002.

[54] W. Banzhaf, Artificial Regulatory Networks and Genetic Programming, Chapter 4 in Rick L. Riolo and Bill Worzel, Eds, *Genetic Programming Theory and Practice*, pp 43–62, Kluwer, 2003.

[55] V. Vidal, H. Geffner, Branching and Pruning: An Optimal Temporal POCL Planner based on Constraint Programming, Artificial Intelligence 170 (3), pp. 298-335, 2006.

Bibai's Master thesis, now continuing with a Cifre PhD: one research direction, here again, is to be able to assess the difficulty of a sub-problem a priori, i.e. without having to run the deterministic solver.

6.4. Applications

Participants: Anne Auger, Nicolas Bredèche, Philippe Caillou, Cyril Furtlehner, Cécile Germain-Renaud, Cédric Hartland, Nikolaus Hansen, Mohamed Jebalia, Claire Le Baron, Julien Perez, Marc Schoenauer, Michèle Sebag, Xiangliang Zhang.

6.4.1. Numerical engineering

Several studies pertaining to engineering optimisation (including the work on Structural Design mentioned in Section 6.3.1) are on-going, in collaboration with Renault, Dassault and EADS.

Isotherm identification in analytic chromatography, an important challenge for chemical engineers, has been considered within the ACI *Chromalgema*. A simulation/identification platform had been designed, using Self-Adaptive Evolution Strategies and gradient based methods in a first phase. In a second phase, the use of CMA-ES allowed to significantly improve the results ...and the usability of the platform for chemists [30].
Multi-Disciplinary Optimisation, a typical application domain for Evolutionary Algorithms since several objectives (usually non-regular) are involved, has been considered in TAO. In 2007, C. LeBaron's PhD (CIFRE Renault, in its 3rd year) is interested in the global optimisation of the propulsion engine, mixing static and dynamic structural optimization with acoustic design.

Along similar lines, TAO participates in the OMD RNTL project, in charge of the general surrogate approach Work Package, and is working on one test-case provided by EADS that concerns the global optimization of a complete launcher considering structural, combustion, and trajectory planification at the same time for several possible missions (satellite positionning around the earth).

6.4.2. Software Robotics

Evolutionary Robotics is another domain of application where Machine Learning and Evolutionary Computation can be usefully combined. In TAO, in the recent years, diverse controller representations have been investigated, ranging from classical multi-layer perceptrons[56] to evolved Voronoï-based classifiers[57].

During his on-going PhD, C. Hartland first extended the use of an anticipation module to address the reality gap problem[58]. In 2007, the novelty was to consider Echo State Networks (ESNs) [28]. In the scope of learning human demonstrated robot behavior, we showed that by using specific neural network known as Echo State Network, makes it possible to grasp relevant information, both punctual and temporal, regarding the demonstrated sequence. Experiments were conducted on a real world khepera robot and showed promising directions [59]. We also considered using ESN in the scope of evolutionary optimisation without oracle [20] (see section 6.3.1).

We are also concerned with optimizing robot locomotion and morphologies. Recent works this year focused on bridging the reality gap for an evolved locomotion controler for a tetrapodal walking robot using Central Pattern Generator as a locomotion basic block. On the practical aspects, we implemented a howebrew bios update for a real world robotic kit and then ran an evolved controler on the real tetrapodal robot (in the scope of the SCOUT STIC-asie projet). We also proposed a representation formalism to deal with representing robot morphologies. The goal is to use such a representation to evolve morphologies for a given objective (typicaly a locomotion gait) (Master's thesis M2R, 2007).

[56] Nicolas Godzik. *Une approche évolutionnaire de la robotique modulaire et anticipative*. PhD thesis, Université Paris-Sud, September 2005.

[57] Carlos Kavka. *Evolutionary Design of Geometric-Based Fuzzy Systems*. PhD thesis, Université Paris-Sud, July 2006.

[58] Cedric Hartland and Nicolas Bredèche. Evolutionary robotics, anticipation and the reality gap. In *IEEE Intl Conf. on Robotics and Biomimetics (ROBIO)*, pp 1640 - 1645, IEEE Press, 2006.

[59] C. Hartland, N. Bredeche. Using Echo State Networks for Robot Navigation Behavior Acquisition. Accepted at the 4rd IEEE International Conference on Robotics and Biomimetics (ROBIO 2007). Sanya, China, December 2007.

6.4.3. Autonomic Grids

Autonomic Computing, acknowledged a Grand Challenge by IBM in 2001[60] aims at self-aware, self-healing and self-configuring complex computer systems.

Autonomic Grids were considered a promising field of ML applications as Cécile Germain-Renaud provided both her expertise and extensive datasets related to the EGEE[61] grid. A first contribution related to Feature Construction was motivated by the EGEE job modelling application. Exploiting prior knowledge regarding the data heterogeneity (e.g. the job lifecycle depends on both the current grid load, and the job user's expertise), the dataset was aggressively sub-sampled; the bunch of hypotheses extracted from every sub-sample was used along the dual attribute/example clustering principle[62] to derive relevant clusters and sidestep the lack of natural metric on the native representation [44], [45]. Self-protection has been explored through abrupt changepoint detection methods [44] with the perspective of new applications for MAB algorithms. In the area of grid scheduling, preliminary work [34] showed that grid job traffic shares some properties with Internet traffic. Julien Perez's PhD considers reinforcement learning applied to scheduling, with utility function including the QoS requirements of reactive (interactive) grids, in line with previous work on the middleware[63].

6.4.4. Software Testing

The application of EC and ML to software testing has been investigated along the EvoTest STREP (Tanja Voss, ITI, Valencia, Spain coordinator, 2006-2009), where TAO is in charge of the Evolutionary Engine at the core of the search process. The generation of test data is set as an optimisation problem. Depending on the context, the fitness can range from coverage measures (for structural testing), to CPU time or memory consumption (for functional testing of real-time embedded systems). The automatic generation of the Evolutionary Algorithm from the test objectives will be based on GUIDE [18] (see section 5.2), and one of the main challenges as far as Evolutionary Optimisation is concerned is to make the search engine fully autonomous, relieving the burden of adjusting evolutionary parameters by trials and errors.

Independently, in a joint work with the Software Engineering group in LRI (Marie-Claude Gaudel and Sandrine Gouraud), Nicolas Baskiotis' PhD has proposed a hybrid approach based on ML and stochastic heuristics to overcome the drawbacks of statistical structural software testing[64] [12].

6.4.5. Social Systems

Preliminary studies concerning the modeling of speculative bubbles on a financial market under different rationality frameworks [65] were extended as Philippe Caillou's and later Cyril Furtlehner's arrivals strengthened TAO's competence in complex systems (TAO also participates in the European Coordinated Action devoted to Complex Systems, ONCE-CS on the executive side). Further studies were devoted to the interaction of social and economic phenomenons, examining how the structure of the social network governs the macro-indicators in a rent-seeking game [15], [16], [17].

Transportation systems provide other examples of social systems where interesting collective phenomena may emerge from local interactions. We have investigated such complex systems from two perspectives. In the first one we proposes a new approach of traffic reconstruction and prediction based on floating car data, by application of distributed algorithms (belief propagation) and ideas inherited from statistical physics [22], [46]. In the second perspective, the mechanism of jam emergence due to variability of driver behaviours was

[60] J.O. Kephart, D.M. Chess, The Vision of Autonomic Computing, in Computer Magazine, reprinted by IEEE Press, pp 41-50, 2003.

[61] Enabling Grids for E-Science in Europe, infrastructure project (2001-2003), (2003-2007), (2008-2010).

[62] Noam Slonim, Naftali Tishby: Document clustering using word clusters via the information bottleneck method. SIGIR 2000: 208-215.

[63]

C. Germain-Renaud, C. Loomis, J. Moscicki, and R. Texier. Scheduling for responsive grids. *Journal of Grid Computing*, 2007. To appear.

C. Germain-Renaud, C. Loomis, R. Texier, and A. Osorio. Grid scheduling for interactive analysis. In *HealthGrid 2006*, volume 120 of *Studies in Health Technology and Informatics*, pp 25–33, 2006. IOS Press.

[64] Alain Denise, Marie-Claude Gaudel, Sandrine-Dominique Gouraud: A Generic Method for Statistical Testing. ISSRE 2004: 25-34.

[65] semet:CF04

analysed using probabilistic tools of queueing networks and statistical physics models (exclusion processes, condensation mechanisms and phase transitions) [21].

7. Contracts and Grants with Industry

7.1. Contracts and Grants with Industry

Contracts managed by INRIA

- **Chromalgema**, CNRS Program ACI NIM (New Interfaces of mathematics) – 2003-2007 (14 kEur), coordinator F.James (Université d'Orléans);
participant: Anne Auger, Mohamed Jebalia, Marc Schoenauer.
- **ONCE-CS** – 2005-2008 (147 kEur) European *Coordinated Action* from FP6. Coordinator Jeff Johnson, Open University, UK;
Participants: Bertrand Chardon, Marc Schoenauer.
- **OMD-RNTL** – 2005-2008 (72 kEur) Coordinator Rodolphe Leriche, Ecole des Mines de St Etienne; Participants: Anne Auger, Olivier Teytaud and Marc Schoenauer.
- **Renault** – 2005-2008 (45 kEur) side-contract to Claire LeBaron's CIFRE Ph.D.;
Participants: Claire LeBaron, Marc Schoenauer.
- **EvoTest** – 2006-2009 (231 kEur) European *Specific Targeted Research Project* from FP6. Coordinator Tanja E.J. Vos, Instituto Tecnológico de Informática, Spain;
Participants: Marc Schoenauer.
- **GENNECTEC** – 2006-2009 (379 kEur) European *Specific Targeted Research Project* from FP6. Coordinator François Képès, Génopôle and CNRS, France;
Participants: Miguel Nicolau, Marc Schoenauer.
- **SCOUT** – 2007-2008 from STIC Asia program, coordinated by MICA, Hanoi (Vietnam). TAO is the INRIA correspondant for this project (15kEur). Other partners are from Vietnam (IFI, CARGIS), China (LIAMA), Cambodia (ITC) and France (IRD, LRI-Paris Sud / TAO, MaIA-LORIA, IGN).

Contracts managed by CNRS or Paris Sud University

- **PASCAL**, Network of Excellence, 2003-2007 (34 kE in 2005). Coordinator John Shawe-Taylor, University of Southampton. M. Sebag is manager of the Challenge Programme.
- **KD-Ubiq**, Coordinated Action, 2005-2008 (19 kE). Coordinator Michael May, Fraunhofer Institute. M. Sebag is responsible of the Benchmarking WP.
- **Traffic** ACI-NIM (New Interfaces of mathematics) – 2004-2007 (17,5 kEur). Coordinator Jean-Michel Loubès, Project SELECT;
Participants: M. Sebag, O. Teytaud.
- **AGIR** ACI Masses de Données - 2004-2007 (30 kEur) Coordinator Cécile Germain-Renaud.
- **EGEE-II** FP6 IP - 2006-2008 (10 kEur) Participants: Cécile Germain, Michèle Sebag, Xiangliang Zhang, Julien Perez.
- **NeuroLog** RNTL 2007-2009 (10 kEur) Participants: Cécile Germain.
- **DEMAIN** PPF (Interdisciplinary program) of the Ministry of Research 2006-2009 (10KE) Participants: Cécile Germain (coordinator), Michèle Sebag, Xiangliang Zhang, Julien Perez.
- **Galileo**, Programme d'actions intégrés franco-italien – 2007 (4.2 kEur);
Participant and coordinator: Antoine Cornuéjols.

8. Other Grants and Activities

8.1. International actions

8.1.1. Management positions in scientific organizations

- Marc Schoenauer, Board Member of ISGEC (International Society on Genetic and Evolutionary Algorithms) since 2000. This Society became the ACM SIGEVO (Special Interest Group in Evolutionary Computation) in 2006, but the board remained unchanged.

8.1.2. Collaborations with joint publications

- The Games Lab (University of Alberta) invited S. Gelly in 2006 and a collaboration on tree-based Monte-Carlo planning started [27].

8.2. European actions

8.2.1. Management positions in scientific organizations

- Marc Schoenauer, Member of PPSN Steering Committee (Parallel Problem Solving from Nature) since 1998.
- Michèle Sebag, Member of PASCAL Steering Committee (Pattern Analysis, Statistical Modeling and Computational Learning, FP6 NoE) since 2004; Member of KD-Ubiq Steering Committee (Ubiquitous Knowledge Discovery, FP6 CA) since 2006.

8.2.2. Working groups

- EGEE, Enabling Grids for E-SciencE : Cécile Germain-Renaud is a member of the NA4 steering committee, and chair for the Working Group *Short Deadline Jobs*.
- ONCE-CS, Coordinated Action, 6th Framework Program: TAO (Marc Schoenauer) is one of the main contracting nodes, responsible of WP2 - Web Portal and Services. Bertrand Chardon is paid as engineer and works on this WP.
- PASCAL, Network of Excellence, 6th Framework Program: Michèle Sebag, corresponding member for Université Paris-Sud since 2003, Manager of the Challenge Programme since 2005.

8.2.3. Collaborations with joint publications

- Colab, ETH Zurich [32].
- Université Lausanne [23].

8.3. National actions

8.3.1. Organization of conferences and scientific events

- JET, Journées Évolutionnaires Trimestrielles: Marc Schoenauer organized the first editions since their creation in 1998 until 2004. Now member of the steering Committee. A 2-days Summer School was organised in this framework at Yravals in June 2007.
- Evolution Artificielle: the international conference on Evolutionary Computation, is organized in France every second year, and has acquired a world-wide reputation not only because of the good wine and food ...Marc Schoenauer is in the organizing committee since the first edition in 1994. The 8th edition took place in Tours in Ocrober 2007.
- Dagstuhl Seminar: Anne Auger co- organizer for the Seminar “Theory of Evolutionary Computation” 2008.

- Apprenteo, gathering the researchers of the Digiteo Lab (PCRI, CEA, SupElec, LIMSI, CMAP) now RTRA, had a second meeting on March 16th, organized by Michele Sebag.

8.3.2. Management positions in scientific organizations

8.3.3. Associations

- *Evolution Artificielle* : Marc Schoenauer, founding president (1995-2004), now member of the Steering Committee. Anne Auger and Nicolas Bredeche, members of the Administrative Committee since October 2007.
- *AFIA, Association Française d'Intelligence Artificielle* : Marc Schoenauer, member of Executive (since 1998, former president (2002-2004)) ; Michèle Sebag, member of Executive since 2000, treasurer in 2003-2004, president since 2004.
- *FERA, Fédération des Equipes de Recherche en Apprentissage* : Michèle Sebag, member of the Steering Committee with Stéphane Canu, Manuel Davy and Jean-Gabriel Ganascia.

8.3.4. Collaborations with joint publications

- Collaborations with Remi Munos (at Ecole Polytechnique before joining the team-project SEQUEL) grounded the development of MoGo [6].
- Vincent Vidal, Université de Lens, started to collaborate to the temporal planning application by providing the sources of his temporal planner CPT, and gradually became a full co-author of the project [9].
- The collaboration with Gilles Celeux, Project team Select, INRIA Futurs, resulted in some original results on Latent Class Clustering [35].
- The research on representations for the topology of large Neural Networks [31] started after several discussions with Alchemy INRIA, about modern computing architectures.
- Balàzs Kègl, from Laboratoire de l'Accélérateur Linéaire (LAL), Université Paris-Sud, is an associate member of TAO [34].
- Collaboration with the Programming and Software Engineering group, led by Marie-Claude Gaudel, LRI, Université Paris-Sud, is about applying Machine Learning to Software Testing [13].
- The collaboration with the young architecture group EZCT motivated the research on compact representations for Structural Design [47]; this collaboration also led to a joint submission to the Serousi House contestwhere we won the first price together with another candidate.

8.4. Honors

- MoGo, developed in the Team by S. Gelly, J.-B. Hoock, A. Rimmel, O. Teytaud and Y. Wang, won many Kgs-tournaments (http://www.weddslist.com/kgs/past/index.html) and has been for a long time first-ranked on both Cgos-servers (9x9 and 19x19, resp. http://cgos.boardspace.net/9x9/standings.html and http://www.lri.fr/~teytaud/standings/cgosStandings.html). The message-passing parallelization is now on progress.

8.4.1. Keynote addresses

- IEEE Congress on Evolutionary Computation, Sept. 25.-28 2007. Marc Schoenauer, invited plenary speaker.
- ISICA'07, International Symposium on Intelligence Computation and Applications, Wuhan, Sept. 21-23. Marc Schoenauer, invited plenary speaker.
- 3rd Franco-Japanese Workshop, Sapporo, June 25-29, Michele Sebag, invited speaker.
- First Entente Cordiale meeting, London, May 25, Michèle Sebag, invited speaker.

- CMFBD07, Nancy. Complexity in learning. May, 21st, 2007. Olivier Teytaud, invited speaker.
- NIPS Workshop on Machines Learning in Games, December 8, 2007. Monte-Carlo planning in the game of Go. Olivier Teytaud, Invited speaker.

9. Dissemination

9.1. Animation of the scientific community

9.1.1. Editorial boards

- Marc Schoenauer is Editor in Chief of MIT Press Evolutionary Computation Journal (since 2002)
- Marc Schoenauer is Associate editor of Kluwer Genetic Programming and Evolvable Machines (since its creation in 1999), of Elsevier Theoretical Computer Science - Theory of Natural Computing (TCS-C) since its creation in 2002, of Elsevier Applied Soft Computing since its creation in 2000, of Springer Memetic Computing Journal (first issue scheduled in March 2008), and has been Associate Editor of of IEEE Transactions on Evolutionary Computation (1996-2004) and of Kluwer Journal of Heuristics (1997-2003).
- Marc Schoenauer is on the Editorial Board of the book series *Natural Computing* by Springer Verlag, and *Mathématiques Appliquées* by SMAI (Springer-Verlag).
- Michèle Sebag is member of the Editorial Board of Knowledge and Information Systems (since 2003), of Machine Learning Journal (since 2001), of Genetic Programming and Evolvable Hardware (since 2000); she has been Associate Editor of of IEEE Transactions on Evolutionary Computation (1998-2004) and of Revue d'Intelligence Artificielle (2002-2005).

9.1.2. Chair in Organizing Committees

- Olivier Teytaud was chair of the Multiple Simultaneous Hypothesis Testing Wshop, (May, 2007, Paris).
- Antoine Cornuéjols was co-chair of CAP'07 (Conférence Francophone d'Apprentissage) (July, 2007, Grenoble).

9.1.3. Program Committee Member (international events)

- Anne Auger: Genetic and Evolutionary Computation Conference, IEEE Congress on Evolutionary Computation, Parallel Problem Solving from Nature.
- Nicolas Bredèche: European Conference on Genetic Programming, ICINCO Workshop on Multi-agent Robotic Systems.
- Marc Schoenauer: Genetic and Evolutionary Computation Conference, IEEE Congress on Evolutionary Computation, Parallel Problem Solving from Nature, European Conference on Genetic Programming, Evolutionary Computation for Combinatorial Optimization Problems, European Conference on Complex Systems, ...
- Michèle Sebag: PC of ICML 06, 23rd International Conference on Machine Learning, ECML-PKDD 06, 17th European Conference on Machine Learning, 10th Conference on Principle and Practice of Knowledge Discovery from Databases, IJCAI 07, 20th International Conference on Artificial Intelligence; ILP, Inductive Logic Programming, PPSN, Parallel Problem Solving from Nature, EuroGP, European Conference on Genetic Programming, GECCO, Genetic and Evolutionary Computation Conference, CEC, IEEE Congress on Evolutionary Computation, ICDM, IEEE Conf. on Data Mining...
- Olivier Teytaud: Approximate Dynamic Programming and Reinforcement Learning ADPRL'07

- Cécile Germain: IFIP Network and Parallel Computing (NPC) since 2007; EGEE User Forum since 2007; Europar Advisory Board.

9.1.4. Program Committee Member (national events)

- CAP, Conférence d'apprentissage: Michèle Sebag since 1999; Antoine Cornuéjols, since 1999; Olivier Teytaud since 2005.
- EA, Evolution Artificielle: Marc Schoenauer and Michèle Sebag since 1994, Anne Auger, Nicolas Bredèche and Olivier Teytaud since 2005.
- RFIA, Reconnaissance des Formes et IA: Michèle Sebag, member of the Editorial Committee.
- *Les 40 ans de l'INRIA*, Marc Schoenauer member of the Program Committee.

9.1.5. Evaluation committees and invited expertise

- Marc Schoenauer, reviewer for both ANR programs Young Researchers and Open Call ("appel blanc"); reviewer of European STREP Perplexus; Professorship evaluation for Profs K. Deb and Jon Rowe (Birmingham University, UK), and Jim Smith (UWE, Bristol, UK).
- Michèle Sebag, reviewer for FP7 Strep (June 11-15, Bruxels); for both ANR programs Young Researchers and Open Call ("appel blanc"); for RNTL; member of the CNRS evaluation committee for the LINA Lab, Nantes.

9.1.6. Other evaluation activities

- Reviewer for PhD dissertation: Marc Schoenauer (3) ; Michèle Sebag (2) ;
- Reviewer for Habilitation: Michèle Sebag (1)

9.1.7. Popularisation of research results

- The collection of Chairs, designed within a collaboration between TAO and with the architect consortium ECZT, have been bought in the permanent Design Collection of Beaubourg, the French National Modern Art Museum, and exhibited from April to October 2007 (see Section 6.3.1, and a more detailed desription at URL http://www.inria.fr/saclay/ressources-1/computer-culture/artlgorithm/view?set_language=en).

9.1.8. Summer schools, tutorials, invited seminars

- Marc Schoenauer
  - Lecture at the SEBASE Summer School in Birmingham in July 2007;
  - Two tutorials at the JET Summer School in Yravals in June 2007;
  - Séminaire "Complexité" organised by EuroBios in December 2007.
- Michèle Sebag, Lecture at the Dagstuhl Wshop on Probabilistic and Relational Logic, May 2007.
- Anne Auger: Tutorial at the JET Summer School in Yravals in Jume 2007.

9.2. Enseignement

9.2.1. Defended doctorates

- Sylvain Gelly, 25/9/07, Université Paris-Sud.

9.2.2. Graduate courses

- Master 2 Recherche (U. Paris-Sud), Data Mining and Machine Learning (24 h): Michèle Sebag, Antoine Cornuéjols.

- Master 2 Recherche (U.Paris-Sud), Artificial and Natural Perception : Nicolas Bredeche (3h).
- Master 2 Recherche (U.Paris-Sud), Multi-agent Systems : Nicolas Bredeche (3h).
- Master 2 Recherche (U.Paris-Sud), Artificial Evolution and Evolutionary Robotics, Anne Auger, Nicolas Bredèche and Marc Schoenauer.
- Master 1 Recherche (ENS Cachan), Introduction to Machine Learning, Michèle Sebag (3h).

9.2.3. Other research-related teaching activities

- *Ecole Polytechnique*, Projects in Evolutionary Robotics in the *Modex d'Electronique*: Marc Schoenauer, Cédric Hartland.
- *Ecole Polytechnique*, Majeure "SEISM" (Engineering Science): one lesson (+ hands-on experiments) on Evolutionary Topological Optimum Design;
- *Ecole Polytechnique*, *Stages d'option*: Michèle Sebag, Marc Schoenauer.

10. Bibliography

Year Publications

Books and Monographs

[1] *Change Point Detection and Meta-Bandits for Online Learning in Dynamic Environments*, Cepadues, 2007, p. 237-250, http://hal.inria.fr/inria-00164033/en/.

[2] O. TEYTAUD, C. ANTONINI, P. BORGNAT, A. CHATEAU, E. LEBEAU. *Mathématiques pour l'agrégation*, Dunod, 2007, http://hal.inria.fr/inria-00173250/en/.

Doctoral dissertations and Habilitation theses

[3] S. GELLY. *Une contribution à l'apprendissage par renforcement ; application au Computer GO*, Ph. D. Thesis, Université Paris-Sud, September 2007, http://tao.lri.fr/Papers/thesesTAO/SylvainGellyThesis.pdf.

Articles in refereed journals and book chapters

[4] A. CORNUÉJOLS, M. SEBAG. *A note on phase transition and computational pitfalls of learning from sequences*, in "Journal of Intelligent Information Systems", 2007, To appear.

[5] A. EIBEN, Z. MICHALEWICZ, M. SCHOENAUER, J. SMITH. *Parameter Control in Evolutionary Algorithms*, in "Parameter Setting in Evolutionary Algorithms", F. G. LOBO, C. F. LIMA, Z. MICHALEWICZ (editors), Studies in Computational Intelligence, http://www.springerlink.com/content/978-3-540-69431-1/, vol. 54, Springer Verlag, 2007, p. 19-46, http://hal.inria.fr/inria-00140549/en/.

[6] S. GELLY, R. MUNOS. *L'Ordinateur, champion de go ?*, in "Pour la sciences", vol. 354, 2007, p. 28-35, http://hal.inria.fr/inria-00177084/en/.

[7] S. GELLY, S. RUETTE, O. TEYTAUD. *Comparison-based Algorithms are Robust and Randomized Algorithms are Anytime*, in "Evolutionary Computation Journal", vol. 15, n^{o} 4, 2007, p. 411-434, http://hal.inria.fr/inria-00173221/en/.

[8] F. MAREUIL, C. SIZUN, J. PEREZ, M. SCHOENAUER, J.-Y. LALLEMAND, F. BONTEMS. *A simple genetic algorithm for the optimization of multidomain protein homology models driven by NMR residual dipolar coupling and small angle X-ray scattering data*, in "European Biophysics Journal", vol. 37, 2007, p. 95-104, http://hal.inria.fr/inria-00177279/en/.

[9] M. SCHOENAUER, P. SAVÉANT, V. VIDAL. *Divide-and-Evolve: a Sequential Hybridization Strategy using Evolutionary Algorithms*, in "Advances in Metaheuristics for Hard Optimization", Z. MICHALEWICZ, P. SIARRY (editors), Natural Computing Series, Springer Verlag, 2007, p. 179-198, http://hal.inria.fr/inria-00176967/en/.

[10] O. TEYTAUD. *On the hardness of offline multiobjective optimization*, in "Evolutionary Computation Journal", vol. 15, n^{o} 4, 2007, p. 475-491, http://hal.inria.fr/inria-00173239/en/.

Publications in Conferences and Workshops

[11] A. AUGER, O. TEYTAUD. *Continuous lunches are free!*, in "GECCO, London Royaume-Uni", D. THIERENS, ET AL (editors), 2007, p. 916-921, http://hal.inria.fr/inria-00173209/en/.

[12] N. BASKIOTIS, M. SEBAG. *Structural Statistical Software Testing with Active Learning in a Graph*, in "17th Annual International Conference on Inductive Logic Programming, Oregon États-Unis d'Amérique", 2007, http://hal.inria.fr/inria-00171162/en/.

[13] N. BASKIOTIS, M. SEBAG, M.-C. GAUDEL, S.-D. GOURAUD. *A Machine Learning approach for Statistical Software Testing*, in "Twentieth International Joint Conference on Artificial Intelligence", 01 2007, p. 2274-2279, http://hal.inria.fr/inria-00112681/en/.

[14] N. BASKIOTIS, M. SEBAG, O. TEYTAUD. *Inductive-deductive systems: a mathematical logic and statistical learning perspective*, in "CAp'07, Grenoble France", Cepadues, 2007, http://hal.inria.fr/inria-00173259/en/.

[15] P. CAILLOU, F. DUBUT, M. SEBAG. *Economy-driven Shaping of Social Networks and Emerging Class Behaviors*, in "Artificial Economics 2007 Artificial Markets Modeling LNEMS, Palerme Italie", A. CONSIGLIO (editor), LNEMS, vol. 599, Springer Verlag, 2007, p. 195-208, http://hal.inria.fr/inria-00166089/en/.

[16] P. CAILLOU, F. DUBUT, M. SEBAG. *Prêts bilatéraux et réseaux sociaux*, in "Modelling and learning in Social and Human Sciences MASHS 07, Brest France", 2007, http://hal.inria.fr/inria-00167459/en/.

[17] P. CAILLOU, M. SEBAG, F. DUBUT. *Negotiation protocols and dynamic social networks*, in "Multi-Agents for modeling Complex Systems (MA4CS'07) Satellite Workshop of the European Conference on Complex Systems (ECCS'07), Dresde Allemagne", 2007, http://hal.inria.fr/inria-00167465/en/.

[18] L. DACOSTA, M. SCHOENAUER. *GUIDE, a Graphical User Interface for Evolutionary Algorithms Design*, in "GECCO Workshop on Open-Source Software for Applied Genetic and Evolutionary Computation (SoftGEC)", J. H. MOORE (editor), 2007.

[19] A. DEVERT, N. BREDÈCHE, M. SCHOENAUER. *Robust Multi-Cellular Developmental Design*, in "GECCO, London", D. THIERENS, ET AL (editors), ACM Press, ACM SIGEVO, 07 2007, p. 982-989, http://hal.inria.fr/inria-00145336/en/.

[20] A. DEVERT, N. BREDÈCHE, M. SCHOENAUER. *Unsupervised Learning of Echo State Networks: A case study in Artificial Embryogeny.*, in "Evolution Artificielle, Tours France", 2007, 12 pages, http://hal.inria.fr/inria-00174593/en/.

[21] C. FURTLEHNER, J.-M. LASGOUTTES. *A queueing theory approach for a multi-speed exclusion process.*, in "Traffic and Granular Flow '07, Orsay France", 10 pages, 6 figures, 10 2007, unknown yet, http://hal.archives-ouvertes.fr/hal-00175628/en/.

[22] C. FURTLEHNER, J.-M. LASGOUTTES, A. DE LA FORTELLE. *A Belief Propagation Approach to Traffic Prediction using Probe Vehicles*, in "10th International IEEE Conference on Intelligent Transportation Systems, Seattle États-Unis d'Amérique", 6 pages, 5 figures, 10 2007, p. pp. 1022-1027, http://hal.archives-ouvertes.fr/hal-00175627/en/.

[23] C. GAGNÉ, M. SEBAG, M. SCHOENAUER, M. TOMASSINI. *Ensemble Learning for Free with Evolutionary Algorithms ?*, in "GECCO, London Royaume-Uni", D. THIERENS, ET AL (editors), ACM, ACM SIGEVO, 2007, p. 1782-1789, http://hal.inria.fr/inria-00144010/en/.

[24] R. GAUDEL, A. CORNUÉJOLS. *Combining feature ranking methods for high dimensional data analysis*, in "5th Workshop on Statistical Methods for Post-Genomic Data, Paris France", 2007, http://hal.inria.fr/inria-00181429/en/.

[25] R. GAUDEL, M. SEBAG, A. CORNUÉJOLS. *A Phase Transition-Based Perspective on Multiple Instance Kernels*, in "ILP 2007, Corvallis États-Unis d'Amérique", Jude Shavlik, Hendrik Blockeel, Prasad Tadepalli, 2007, http://hal.inria.fr/inria-00175300/en/.

[26] R. GAUDEL, M. SEBAG, A. CORNUÉJOLS. *A Phase Transition-based Perspective on Multiple Instance Kernels*, in "Conférence francophone sur l'apprentissage automatique, Grenoble France", Jean-Daniel Zucker, Antoine Cornuéjols, 2007, p. 173–186, http://hal.inria.fr/inria-00171406/en/.

[27] S. GELLY, D. SILVER. *Combining Online and Offline Knowledge in UCT*, in "Proceedings of the 24th Annual International Conference on Machine Learning (ICML 2007)", Z. GHAHRAMANI (editor), Omnipress, 2007, p. 273–280, http://hal.inria.fr/inria-00164003/en/.

[28] C. HARTLAND, N. BREDÈCHE. *Using Echo State Networks for Robot Navigation Behavior Acquisition*, in "ROBIO 07, Sanya Chine", 2007, http://hal.inria.fr/inria-00176817/en/.

[29] M. JEBALIA, A. AUGER. *On the convergence of the* $(1+1)$*-ES in noisy spherical environments*, in "Evolution Artificielle, Tours France", 2007, http://hal.inria.fr/inria-00173483/en/.

[30] M. JEBALIA, A. AUGER, M. SCHOENAUER, F. JAMES, M. POSTEL. *Identification of the Isotherm Function in Chromatography Using CMA-ES*, in "IEEE Congress on Evolutionary Computation, Singapour", 2007, p. 1879-1886, http://hal.inria.fr/inria-00175852/en/.

[31] F. JIANG, H. BERRY, M. SCHOENAUER. *Optimising the topology of complex neural networks*, in "ECCS'07, Dresden Allemagne", J. JOST (editor), Complex Systesms Society, 2007, http://hal.inria.fr/inria-00175721/en/.

[32] S. KERN, N. HANSEN, P. KOUMOUTSAKOS. *Fast Quadratic Local Meta-Models for Evolutionary Optimization of Anguilliform Swimmers*, in "EUROGEN 2007, Helsinki Finlande", NEITTAANMAKI (editor), 2007, http://hal.inria.fr/inria-00173469/en/.

[33] N. LE ROUX, Y. BENGIO, P. LAMBLIN, M. JOLIVEAU, B. KÉGL. *Learning the 2-D Topology of Images*, in "Advances in Neural Information Processing Systems, Vancouver Canada", 2007, http://hal.inria.fr/inria-00176067/en/.

[34] J. PEREZ, C. GERMAIN-RENAUD, B. KÉGL. *Predicting Bounds on Queuing Delay in the EGEE grid*, in "2nd EGEE User Forum, Manchester Royaume-Uni", 2007, http://hal.inria.fr/inria-00174290/en/.

[35] D. TESSIER, M. SCHOENAUER, C. BIERNACKI, G. CELEUX, G. GOVAERT. *Parameter Setting for Evolutionary Latent Class Clustering*, in "Second International Symposium, ISICA 2007 LNCS, Wuhan Chine", L. KANG, Y. LIU, S. Y. ZENG (editors), LNCS, vol. 4683, Springer Verlag, 2007, p. 472-484, http://hal.inria.fr/inria-00179186/en/.

[36] O. TEYTAUD. *Conditionning, halting criteria and choosing lambda*, in "EA07, Tours France", 2007, http://hal.inria.fr/inria-00173237/en/.

[37] O. TEYTAUD. *Slightly beyond Turing-Computability for studying Genetic Programming*, in "MCU'07, Orléans France", 2007, http://hal.inria.fr/inria-00173241/en/.

[38] O. TEYTAUD, A. AUGER. *On the adaptation of the noise level for stochastic optimization*, in "IEEE Congress on Evolutionary Computation, Singapour", 2007, 2030,2037, http://hal.inria.fr/inria-00173224/en/.

[39] O. TEYTAUD, S. GELLY. *DCMA, yet another derandomization in covariance-matrix-adaptation*, in "GECCO, London Royaume-Uni", D. THIERENS, ET AL (editors), 2007, p. 955-922, http://hal.inria.fr/inria-00173207/en/.

[40] O. TEYTAUD, S. GELLY. *Non linear programming for stochastic dynamic programming*, in "Icinco 2007, Angers France", 2007, http://hal.inria.fr/inria-00173202/en/.

[41] O. TEYTAUD, S. GELLY, J. MARY. *Active learning in regression, with an application to stochastic dynamic programming*, in "ICINCO 2007, Angers France", 2007, http://hal.inria.fr/inria-00173204/en/.

[42] O. TEYTAUD, S. GELLY, M. SEBAG. *Anytime many-armed bandits*, in "CAp'07, Grenoble France", Cepadues, 2007, http://hal.inria.fr/inria-00173263/en/.

[43] C. TIJUS, N. BREDÈCHE, Y. KODRATOFF, M. FELKIN, C. HARTLAND, E. ZIBETTI, V. BESSON. *Human Heuristics for a Team of Mobile Robots*, in "IEEE International Conference on Research, Innovation and Vision for the Future, Hanoi Viet Nam", 2007, p. pp.122-129, http://hal.inria.fr/inria-00175523/en/.

[44] X. ZHANG, M. SEBAG, C. GERMAIN-RENAUD. *Fault Detection and Diagnosis from the Logging and Bookkeping Data*, in "2nd EGEE User Forum, Manchester Royaume-Uni", 2007, http://hal.inria.fr/inria-00174286/en/.

[45] X. ZHANG, M. SEBAG, C. GERMAIN-RENAUD. *Toward Behavioral Modeling of a Grid System: Mining the Logging and Bookkeeping files*, in "DSMM07, Omaha États-Unis d'Amérique", 2007, http://hal.inria.fr/inria-00174285/en/.

Internal Reports

[46] C. FURTLEHNER, J.-M. LASGOUTTES, A. DE LA FORTELLE. *Belief Propagation and Bethe approximation for Traffic Prediction*, 29 pages, 7 figures, Research Report, n^{o} RR-6144, INRIA, 2007, http://hal.inria.fr/inria-00136657/en/.

Miscellaneous

[47] H. HAMDA, M. SCHOENAUER, P. MOREL. *A(rt)lgorithm, design informed by mathematics*, April-October 2007, http://www.inria.fr/saclay/ressources-1/computer-culture/artlgorithm/view?set_language=en, Centre Pompidou, Collection permanente.