INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team tao

Thème apprentissage et optimisation

Futurs

THEME COG

Activity Report

2005

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]

Michèle Sebag [DR CNRS]

Permanent staff

Nicolas Bredèche [MDC Université Paris Sud]

Philippe Caillou [MDC Université Paris Sud – since October 2005]

Antoine Cornuéjols [MDC CNAM]

Cécile Germain [MDC Université Paris Sud]

Olivier Teytaud [CR INRIA]

Administrative assistant

Marie-Carol Lopes [Tao/Select/Proval]

Ph.D. Students

Nicolas BASKIOTIS [started September 2003]

Alexandre DEVERT [started September 2005]

Mary FELKIN [started September 2003]

Sylvain GELLY [started September 2004]

Nicolas GODZIK [defended in September 2005]

Cédric HARTLAND [started September 2005]

Thomas HEITZ [started September 2004]

Celso Yoshikazu ISHIDA [started December 2003, co-tutelle Universidade Federal do Parana - Brésil)]

Mohamed JEBALIA [started September 2004, co-tutelle Université Tunis - Tunisie)]

Claire LE BARON [CIFRE Renault, started September 2005]

Jérémie MARY [defended in December 2005]

Jérôme MALOBERTI [co-tutelle Univ. Yokohama, defended in June 2005]

Matthieu PIERRES [CIFRE Airbus, started February 2004]

Vijay PRATAP SINGH [CIFRE IFP, started November 2003]

Raymond ROS [started September 2005]

Post-doctoral fellows

Jérôme AZÉ [until September 2005]

Christian GAGNÉ [since May 2005]

Samuel LANDAU [until September 2005]

Mathieu ROCHE [until September 2005]

Research Engineer

Bertrand CHARDON [since November 2005]

Yann SEMET

Damien TESSIER [since October 2005]

Collaborators

Yves KODRATOFF [DR CNRS]

2. Overall Objectives
2.1. Overall Objectives

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 dynamical system parameters, and finding short cuts in their trajectories.

3. Scientific Foundations

3.1. Scientific Foundations

Abstract:

One of the goals of Machine Learning and Data Mining is to extract optimal hypotheses from (massive amounts of) data. What "optimal" means varies with the problem. The goal might be to induce useful knowledge, allowing new cases to be classified with optimal confidence (predictive data mining), or to synthesize the data into a set of understandable statements (descriptive data mining).

On the other hand, Evolutionary Computation and stochastic optimization are adapted to ill-posed optimization problems, such as involved in machine learning, data mining, identification, optimal policies, and inverse problems. However, optimization algorithms must adapt themselves to the search landscape; in other words, they need learning capabilities.

3.1.1. *Machine learning, Data Mining, Inductive Logic Programming*

Learning is concerned with i) choosing the form of knowledge to be extracted (rules, Horn clauses, distributions, patterns, equations, ..), referred to as hypothesis space or language ; ii) exploring this (HUGE) search space, to find the best hypotheses in it.

Learning thus is an optimization problem; however, the "real" optimization criterion is unknown. Learning is like a game, with incomplete information: i) in the statistical learning case, the player (learning algorithm) only knows some cards of the game (the available examples, in the data set); ii) in the data mining case, the player (algorithm) does not know the preferences of the expert (whom the algorithm tries to please).

New learning criteria (and the corresponding algorithms) are investigated, concerned with the Area under the ROC curve (Receiver Operating Characteristics), particularly for medical applications, and concerned with stable spatio-temporal data, with applications in Neurosciences.

3.1.2. *Evolutionary Computation, Stochastic Optimisation*

Considering the lack of a universal optimisation algorithm, the power of an optimisation algorithm is measured by its ability in acquiring and exploiting problem-specific information. The use of such a priori knowledge has long been heuristic. It leads for example to the development of operators specific to pattern optimisation, to constrained identification, etc. One of our objectives is to have operators able to adapt themselves by automatically exploiting regularities in the search space. Another objective is to investigate how domain knowledge can be introduced at all levels of evolutionary algorithms, starting with the representation itself, and in the corresponding variation operators.

3.1.3. *Robotics*

Autonomous robotics is a fascinating challenge to optimisation and machine learning. The TAO approach is mainly inspired by evolutionary robotics and cognitive science. It is based on defining a control problem (optimisation of behavioral traits leading to the desired behavior), on subordinating the controller to the real world (by a module which predicts the effects of its actions, by simplifying sensory-motor cognition, etc) and on intensive use of the informations acquired by the robot (log mining).

A new framework has arrived in the team with the OpenDP project, that combines methods from operational research (discretization of the Hamilton-Jacobi-Bellman equation), and learning of the Bellman-function. Robotics is not the usual application of such methods ; thanks to modern methods of learning, working in spaces of large dimension is possible, enabling the application of stochastic dynamic programming for robotics.

3.1.4. *Crosswise Axis : Phase Transition*

Coming from work done on constraint satisfactions, the phase transition concept is appropriate to describe the average performance of algorithms faced with NP-complete problems. This concept allows us to study how machine learning and relational data mining scale up, and to define the space of difficult problems for functional algorithms in this domain. The study of this perspective for propositional learning is in progress. The point of such an empirical and statistical study of algorithms taken as black boxes is first to pinpoint their weaknesses, then to understand these weaknesses, and then, hopefully, to fix them.

4. Application Domains

4.1. Application Domains

Applications are described all along the text, and referenced in the contract section.

The main application domains are Robotics (see section 6.3), Medical Data Mining (see sections 6.1.1, 6.1.2) and Inverse problems for Numerical Engineering (section 6.4 and

5. Software

5.1. Evolving Objects

Keywords: *Evolutionary Computation*, *Object-oriented*, *Standard Template Library*.

Participant: Marc Schoenauer [correspondent].

Abstract: EO is a templates-based, ANSI-C++ compliant evolutionary computation library. It contains classes for almost any kind of evolutionary computation you might come up to - at least for the ones we could think of. It is component-based, so that if you don't find the class you need in it, it is very easy to subclass existing

abstract or concrete class. EO works with main compilers, including GNU $g++$ (versions 2.95 and above) and Microsoft *Visual C++* (versions 6.00 and above).

In 2005, EO has been kept up-to-date with the most recent compilers (e.g., g++ 4.0), and fundamental improvements include the CMA-ES algorithm, the up-to-date method in parametric optimization and a very efficient GP algorithm for real function identification.

See main page at http://eodev.sourceforge.net//.

5.2. GUIDE: A graphical interface for EO

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

Participants: Marc Schoenauer [correspondent], Damien Tessier.

Abstract: GUIDE is a graphical user interface for the Open Soucre library EO (see above). 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).

The improvements that are now brought to GUIDE concern the link with WEKA (addition of a new type of genotype, compliant with Weka .aff files) and a better separation of GUIDE and EO, allowing to consider using GUIDE for another lirbary (e.g. BEAGLE, the library developed by Christian Gagné, post-doc ERCIM at TAO from May 2005 to February 2006).

5.3. World in a bottle

Keywords: *OpenGL*, *Robot simulator*.

Participant: Jeremie Mary [correspondent].

Abstract: "World in a Bottle" is a robot's simulator written in C++ that takes advantage of the OpenGL library. Kheperas robots are currently privileged but it is possible to easily implement other types of robots into the simulator.

- Real time 3D display of the simulation, including sensors view (display can be disable to achieve better simulation speed).
- The environment can be easily designed (walls, cylinders...), as well and saved or loaded to/from a file.
- It is possible to run as many robots as need in the simulator.
- It is possible to control the robot during the simulation through the keyboard controls.
- Simulation of IR proximity sensors both in active and passive modes.
- Simulation of 1D and 2D Cameras.
- Simulation of moving obstacles.
- It is possible to easily write a controller in C++ (C++ tutorials are available).
- It is also possible to write a controller in any language thanks to the simulator's server mode (i.e. one just have to write a client that connect to the simulator through a socket).
- The simulator can switch directly from simulation mode to real-world mode. This enables your controller to directly take control of a real-world robot. This is currently limited to khepera robots but may be extended.
- The simulator can be easily interfaced with EO (the Evolutionary library), enabling the evolution of robot controllers
- The simulator currently works under Linux, Windows and Mac OS X.

For further information and download, please refer to : http://www.lri.fr/~mary/WoB/.

5.4. Django

Keywords: *Fast theta-subsumption.*

Participants: Jérôme Maloberti [correspondent], Michèle Sebag.

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.
   See http://www.lri.fr/~malobert/Django/.

5.5. OpenDP

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

Participants: Olivier Teytaud [correspondent], Sylvain Gelly, Jérémie Mary.

Abstract: OpenDP is a young open source code for stochastic dynamic programming, based upon the use of (i) time-decomposition as in standard dynamic programming (ii) learning (iii) derivative-free optimization. It is designed in a very modular manner, including many 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 Weka algorithms and some others for learning. It also includes various benchmarks.

The inclusion of tools from various areas of science is under work, such as time-pca, robotic-mapping, derandomization of random processes. If many of these tools are not new, their use in the framework of dynamic programming is new. The software is already parallel and has provided many results, among which a comparison of function-values approximators (no so large comparison existed in the literature, many published papers only considering one learning method, not necessarily in the same conditions than other published results) and derivative-free optimization algorithms in the case of a very restricted number of iterates. The use of benchmarks that are not in the traditional goals of stochastic dynamic programming is also new.
As a side effect of this software development:

- a first publication has been done ([27]) about the problem of random processes modelization ;

- we started a collaboration with the ETH institute (Zurich), using the optimization library of OpenDP on the CEC benchmark for real-parameter optimization ;

- the software will be the basis for a challenge that has been accepted by the Pascal Network of Excellence. The goal is that any research team can include easily its optimization software or itslearning method. The challenges fills a hole, as no general comparison of various optimization methods / learning methods has been done in the framework of dynamic programming.

Contacts have been developped with industrial and applied users of dynamic programming, such as Artelys, Cemagref and EdF.
   See main page at http://opendp.sourceforge.net//.

6. New Results

6.1. Fundamentals of Machine Learning, Knowledge Extraction and Data Mining

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: Jérôme Azé, Nicolas Baskiotis, Nicolas Bredèche, Antoine Cornuéjols, Mary Felkin, Jérôme Maloberti, Jérémie Mary, Michèle Sebag, Olivier Teytaud.

Abstract: This theme focuses on machine learning, knowledge discovery and data mining (ML/KDD/DM) considered as optimisation problems, and particularly on the key issues of the search space/hypothesis language, and the learning criteria.

6.1.1. EC-based Learning and Mining

TAO participates to the PASCAL[1] (Pattern Analysis, Statistical Modelling and Computational Learning, 2003-2007) Network of Excellence, which includes the major European Research Centers in Machine Learning; M. Sebag is responsible for the UPS site, including the TAO and SELECT projects, and member of the Pascal Steering Committee as manager of the Challenge Programme.

While the mainstream of Machine Learning considers quadratic learning criteria and well posed optimisation problems, e.g., the structural risk minimization underlying kernel methods, our expertise in evolutionary computation allows us to consider non convex optimisation criteria such as the Wilcoxon statistics, a.k.a. area under the Receiver Operating Characteristics (ROC) curve. The ROC curve, describing the trade-off between the two types of error of a hypothesis, is well suited to imbalanced example distributions and cost-sensitive learning.

Our experiments with the evolutionary optimisation of the area under the ROC curve (AUC) have shown very good learning performances compared to prominent approaches such as Support Vector Machines. The ROGER algorithm, *ROC-based GEnetic learneR*, has been successfully applied to bio-informatics, more particularly to ranking candidate solutions for a protein docking problem [6], [16], and to text mining, to ranking candidate terms during the terminology extraction step [13], [14].

Although the optimization of the AUC criterion has been formalized as a constrained quadratic optimization (Joachims 2005), it must be emphasized that this approach is actually quadratic in the number N of examples; in opposition, ROGER complexity is in $N \log N$.

Along the same line, the search for stable patterns in spatio-temporal data mining was formalized as a multi-objective multi-modal optimization problem, applied to functional brain imaging and presented at IJCAI 2005 [37] (ACI NIM NeuroDyne contract, in coll. with Hôpital La Pitié Salpétrière, LENA).

6.1.2. Feature Selection (FS)

Most available databases were not constructed with data mining in mind, and they usually involve a number of features that are irrelevant to the learning task at hand. The irrelevant features not only result in a significant increase of the computational and memory resources needed; they might also mislead the search for a good hypothesis, ultimately resulting in a poor predictive accuracy. Feature Selection (FS) is thus recognized as a central task for ML/KDD/DM applications, and particularly in bio-informatics. In collaboration with the bio-informatics team of LRI and INSERM, a novel algorithm inferring the relevance of attributes from the structure and parameters of hypotheses was proposed; this work resulted in the discovery of a biological process likely to occur as a response to weak radiation exposures[2]

---

[1] http://www.pascal-network.org
[2] G. Mercier, N. Berthault, J. Mary, A. Antoniadis, J-P. Comet, A. Cornuéjols, Ch. Froidevaux, and M. Dutreix. Biological detection of low radiation by combining results of two analysis methods. *Nucleic Acids Research*, **32**:1, pp 1–8, 2004.

In response to the uncertainty that plague feature ranking in the case of data with very low signal to noise ratio, like in genomics, a new method was developed. It is based on a new method to measure the correlation between ranking techniques. Relying on a maximum likelihood method, it allows one to combine the results of two or more ranking methods to get a high precision estimation of the number of relevant features and their identity. The method has been applied to microarray data [20].

Theoretical studies about feature selection have been pursued, firstly in relation with the ACI MIST-R contract (Nouvelles Interfaces des Maths), and secondly in the PASCAL NoE framework.

The MIST-R contract, in collaboration with J.-M. Loubès, Lab. Math. UPS, is concerned with the modelling of road traffic; in this context, almost sure convergence proofs of feature selection under a VC-penalized scheme were obtained by Merve Amil and Olivier Teytaud [45], as well as non-convergence proofs for some non-VC penalized schemes.

Further, a theoretical Pascal challenge was launched (http://www.lri.fr/~teytaud/risq/ regarding the type I and type II error rates when extracting sets of solutions; in the FS framework, such errors respectively correspond to the selection of irrelevant and the pruning of relevant features.

6.1.3. Bayesian networks

Bayesian networks, a now classical framework for probability density estimation, involve the identification of both a structure (the dependency graph) and its parameters. Whereas many works have been devoted to the optimization of the structure, only a few papers deal with parameter learning, classically handled in a frequentist approach. A new approach related to parameter learning has been proposed in [29] – one of the 10 papers suggested for publication in the RIA journal. The results are :

- a new complexity measure that shows that the number of parameters is not the only important term in the complexity of bayesian networks ; the structure of the network, for a fixed number of parameters, has a strong influence that is theoretically predicted by a result in the paper above (confirmed by practical results in the journal version) ;

- a new criterion for parameter fitting, computationally harder than the usual one (which reduces in general to the frequentist approach), that leads to substantially better results, even in the limit of large samples, in the same way than loss minimization is preferable to likelihood maximization when the model is unperfect.

Some other results, coming from classical learning theory coupled with results above, include proofs of convergence to a minimal sufficient structure.

6.1.4. Relational Learning

Relational Learning, a.k.a. Inductive Logic Programming (ILP) is about learning from structured examples, such as chemical molecules (graphs), XML data (trees), and/or learning structured hypotheses such as toxicological patterns (graphs, sequences) or dimensional differential equations (mechanical models).

The TAO team developed an internationally acknowledged competence in ILP and Relational Learning; Céline Rouveirol and Michèle Sebag co-chaired the 11th International Conference in Inductive Logic Programming in 2001. Compared to propositional learning and data mining, relational learning faces two additional difficulties. On one hand, the covering test, checking whether a given hypothesis covers an example, is equivalent to a constraint satisfaction problem (CSP), with exponential complexity; on the other hand, the search space is doubly exponential.

A principled and computationally effective approach, based on the junction of ILP and CSP, was proposed in Jerome Maloberti's PhD toward the covering test. The resulting algorithm, termed Django,shows an improvement of several orders of magnitude on artificial and real-world problems over the existing algorithms; this achievement has been cited among the major UPS ones over 2005. Django, available under GPL (see section 5.4), has been widely used and cited in the literature (coll. with the Yokohama University, Japan, U. of Tufts in Arizona, U. of Bari, Italy).

Recent developments, done during Alexandre Termier's post-doc visiting Pr. Motoda's Lab at Osaka University after his PhD co-supervised by M.-C. Rousset and M. Sebag, were concerned with XML and more generally tree mining [43], [42].

Other developments have been related to the ACI IMPBIO ACCAMBA, concerned with cribbling molecules, and using to this aim the relational learner STILL, based on the stochastic sampling of the substitution space, initially designed by M. Sebag.

6.1.5. *Phase transition*

The abovementioned relationship between relational learning and CSPs led to extend the phase transition paradigm (Hogg et al., 96) into the ML framework. This study, initialized in collaboration with L. Saitta and A. Giordana from the University of Alessandria, Italy, was exceptionally fruitful: negative results were obtained concerning the scalability of ILP learners [3] and the failure region of the prominent FOIL algorithm was identified in 2003. It was shown in 2004 that phase transition could be observed also in propositional learning [4], and the failure region of the widely known C4.5 algorithm was identified.

In 2005, we similarly investigated the grammatical inference framework, the complexity of which is intermediate between relational and propositional learning. Unexpected results regarding the behavior of the prominent RedBlue and RPNI grammatical inference algorithms, and showing the shortcomings of the stopping criterion, were found and presented at IJCAI 2005 [34]; this paper was proposed for submission in the International Journal of Intelligent Information Systems.

This study was resumed during Raymond Ros's Master [52]. Contrary to all expectations, the results obtained from extensive experimentations show striking discontinuities in the exploration of the research space in the case of Deterministic Finite state Automata and sharp variations in the case of Non deterministic Finite state Automata [35], [34], [21].

6.1.6. *Competence Maps*

The proposed phase transition paradigm considers computational complexity, coverage rate or learning success as random variables conditioned by some order parameters (e.g. tightness and hardness of the constraints, size of clauses and examples in relational learning, alphabet size and number of states in finite state automata). The average behavior of the learning operators is observed through extensive experimentations, using a problem sampling mechanism relying on the order parameters. Accordingly, such studies allow for drawing the *Competence Map* of the learners in the landscape defined from the order parameters.

Such competence maps partially address the critical Meta-Learning problem, widely acknowledged as the main bottleneck for Machine Learning, and concerned with determining *a priori* the best algorithm for a given dataset. Indeed, competence maps estimate the average error rate of the algorithm given the characteristics of the dataset. Relatedly, the phase transition approach offers a methodology for the principled validation and verification of learning algorithms (invited talk at the NIPS 2004 Wshop of Verification and Validation of Learning Algorithms, M. Sebag). As an example of the "competence map" principle, a computationally expensive and parallel test is under progress for the competence maps of learning algorithms as function value estimators, in the OpenDP framework (see section 5.5).

It must be emphasized that this approach significantly differs from an analytical algorithmic study; instead, it postulates that many heuristics are packed into really efficient algorithms, the interaction of which is hardly amenable to analytical modeling. Therefore, an empirical framework originating from natural and physical sciences is a useful tool to determine the regions in the problem space where an algorithm generally fails or succeeds.

---

[3]M. Botta and A. Giordana and L. Saitta and M. Sebag. Relational Learning as Search in a Critical Region. *Journal of Machine Learning Research*, Vol. 4, pp 431–463, 2003.

[4]Nicholas Baskiotis and Michele Sebag. C4.5 Competence Map: a Phase Transition-inspired Approach. Proc. ICML 2004.

6.1.7. DataMining on the Grid

Cécile Germain, former member of the Parall group at LRI, joined the TAO project in 2005. Her strong expertise in grid computing, as member of the EGEE (Enabling Grid for E-Science in Europe) Network of Excellence, opens new and strategic research avenues for Grid-Aware Mining Algorithms.

Currently, Cécile Germain chairs the ACI MD AGIR[5] contract (starting sept. 2004), concerned with medical data mining and more precisely medical imaging through grid computing [7], [30]. Our specific interests in AGIR regard the interactive exploration and retrieval of relevant images and medical exams; these tasks require specific grid services and intelligent prefetch mechanisms, supporting the physician's search. A multi-disciplinary project, AGIR gathers researchers in computer science, physics and medicine from CNRS, University, INRIA, INSERM and hospitals.

A general architecture for interactive grid access has been designed, and is in the process of integration in the EGEE software[6]. The application to volume reconstruction [8], which grid-enables the PTM3D software developed at LIMSI, has been part of the first EGEE review and of the HealthGrid demonstrations at SC'05 .

An on-going project termed DEMAIN (*Des DonnéEs MAssives Aux InterpretatioNs*), centered on e-Science[7], is a PPF proposal to Paris-Sud University (approbation pending). DEMAIN, gathering computer scientists, mathematicians (Lab. de Mathématiques), physicists (Lab. Accélérateur Linéaire), and biologists (IBBMC), aims to develop principled and efficient algorithms to deal with huge amounts of data, describing complex natural or artificial systems, in the absence of ground truth. Among the first considered applications is the control of the EGEE system itself (result-checking algorithm based on Wald's sequential test for a typical class of grid computations (Monte-Carlo simulations) [31]), and the analysis of its failures from the grid logs (Samuel Deberles's master thesis [47]); another application concerns the analysis of the Auger experiment, in collaboration with A. Cordier (LAL). This junction between the LAL and the Tao group was instrumental in the recruitment of Balazs Kegl, CNRS CR1, researcher in machine learning at LAL.

6.2. Fundamentals of Evolutionary Computation

Keywords: *Asymptotic convergence rate*, *Convergence of evolutionary algorithms*, *Estimation of Distribution*, *Evolution Strategies*, *Self-adaptivity*.

Participants: Nicolas Bredeche, Sylvain Gelly, Mohamed Jebalia, Marc Schoenauer, Michèle Sebag, Olivier Teytaud.

Abstract: Evolutionary Computation (EC) is a unifying framework for population-based optimization algorithms. It relies on a crude imitation of the Darwinian evolution paradigm: adapted species emerge thanks to natural selection combined with blind variations. Historical approaches differ by the search space they work on: genetic algorithms work on bit-strings, evolution strategies on real-valued parameters, and genetic programming on structured programs.

EC is now widely acknowledged as a powerful optimization framework dedicated to ill-posed optimization problems. The main reason for its efficiency comes from the possibility for EC to incorporate background knowledge about the application domain into the representation and the variation operators.

6.2.1. Convergence results for the (1,λ)-Self-Adaptive Evolution Strategies

Evolution Strategies are evolutionary algorithms recommended by the state of the art in practical parametric optimisation. Since their invention in the mid-sixties, theoretical studies mainly concentrate on establishing local properties of this algorithm on the well known sphere function ($f(x) = \sum_{i=1}^{dim} x_i^2$). A recent paper [5] investigates the global convergence of the (1,λ)-SA-ES on this function and proves sufficient conditions ensuring the linear convergence (or divergence) of the algorithm. The proofs call upon the Theory of Markov chains on a continuous state space and make use of the so-called drift conditions to establish practical

---

[5] http://www.aci-agir.org
[6] http://egee-na4.ct.infn.it/wiki/index.php/ShortJobs
[7] http://www.lri.fr/~cecile/DEMAIN/DemainSc.htm

properties of the Markov chains investigated, and are the continuation of Anne Auger's PhD thesis, defended in 2004.

New theoretical works have been performed, in continuing collaboration with Anne Auger, now at ETHZ:

- A 3/2-order convergence proof for a Newton-Based surrogate model, as far as we know the best ever-published convergence rate for derivative-free methods [12]. This result is based on an ad hoc choice of the mutation step-size and a derandomized sampling ; it can indeed be seen as a result about the choice of the step-size in finite-difference Newton-methods. This work is therefore between evolution strategies and classical optimization.

- A linear convergence proof for a derandomized evolution strategy, under very mild hypothesis [44]. This result uses an ad hoc sampling for a linear convergence rate in a very difficult framework where objective functions have no continuity, even in the neighborhood of the optima.

6.2.2. *Genetic Programming*

Methods from Statistical Learning Theory have been used in the framework of Genetic Programming for function identification, leading to original results in the theoretical study of GP. It has been published in [28], and in CAP'05 conference, where it was suggested for publication in the journal RIA. This study (i) provides proofs of bloat in various standard frameworks for genetic programming (ii) provides proofs of no-bloat in other frameworks. This work has been recommended for publication in GPEM Journal and a joint work with C. Gagné (post-doctoral fellow in the Tao team) has been performed, including numerical experiments that show that the approach can be used in practice.

Another joint work where GP is used for Machine Learning, in the continuation of C. Gagné's PhD thesis, together with M. Schoenauer and in collaboration with M. Tomassini (U. Lausanne) and M. Parizeaux (U. Laval à Québec) studies the influence on the bloat of several heuristics for the choice of the best GP classifier after a multi-objective GP run [23].

6.2.3. *Multi-objective optimization*

The population-based dynamics of EAs allows one to tackle multi-objective optimization (sampling the Pareto front, i.e. the set of the best compromises between the contradictory objectives)[8].

This part of TAO's activities buildson on Olga Rudenko'PhD, defended in 2004, where, apart from applications in the automobile area, she designed a sound stopping criterion (the only available criterion was ...a given number of generations) and proposed a new crossover operator based on the dominance property.

Those works have been completed by a theoretical study [17], including the work of Y. Bonnemay, in training-course with St Gobain. This theoretical study shows, in the case of a simplified multi-objective evolutionary algorithm, that :

- the sample complexity for a given precision is roughly linear in the number of objective functions ;

- a good stopping/precision criterion is the ratio between the undominated points that have been already discovered and the total number of visited points.

Also note that C. Gagné's post-doctoral work on GP for Machine Learning is also concerned with evolutionary multi-objective optimzation.

---

[8] K. Deb, *Multi-Objective Optimization Using Evolutionary Algorithms*, John Wiley, 2000.

6.2.4. Surrogate models in Evolution Strategies

The work about surrogate models, that was started in the Tao team by K. Abboud in his Ph.D. thesis, defended in 2004, has been continued by the work of Y. Bonnemay (in training between Saint-Gobain and Tao) and O. Teytaud. This work concern the theoretical properties of surrogate models, and an empirical study on (i) standard test difficult test functions (ii) industrial problems.

The approach by virtual examples, proposed to solve the inverse problem of chromatography (section 6.4.3) also makes use of surrogate models and builds on K. Abboud's PhD.

6.2.5. Estimation of Distribution Algorithms

Estimation of Distribution Algorithms (EDAs) proceed by alternatively sampling and updating a distribution on the search space. The sampled individuals are evaluated, i.e. their fitness is computed, and the distribution is updated and biased toward the best individuals in the current sample. Extensions of this framework to continuous optimization, initialized by Ducoulombier & Sebag (1998)[9], showed failures in specific cases where the solutions lie on the edge of the search space. Regularization heuristics, which calibrate the eigenvalues of this distribution, were shown to successfully overcome such failures.

On-going work (Ph.D. Celso Ishida, co-advised with A. Pozo, Universidad Federale do Parana, Brazil) is concerned with using mixtures of distributions, borrowing to the MIXMOD EM-like approaches developed in the SELECT project at INRIA, to extend EDAs to multi-modal optimization.

In the same way EDAs evolve a stochastic model mainly applied to the search in the space of continuous values, one can evolve a stochastic grammar model, with a grammar describing the building of structures, mainly applied to the search in the space of programs. This is called EDP (Evolution of Distribution Programming). In previous works[10], we already investigated a EDP model based on the coupling of context-free grammars with genetic programming (CFG-GP), with sound results. The Master stage of Jean-Philippe Braud aimed at further exploring this matter, by refining the program distribution model and endowing it the ability to represent probabilities for path-dependent derivations of the grammar model. This stage was supervised by Samuel Landau, during his postdoc at TAO (report not yet available).

6.3. Robotics

Participants: Nicolas Bredèche, Antoine Cornuéjols, Alexandre Devert, Mary Felkin, Sylvain Gelly, Nicolas Godzik, Cédric Hartland, Jérémie Mary, Raymond Ros, Marc Schoenauer, Michèle Sebag.

Abstract: Autonomous Robotics is a challenging application for both machine learning and evolutionary computation. From a machine learning perspective, robotics asks for extending machine learning algorithms beyond the classical assumptions of independent and identically distributed examples (indeed the robot exploration results in non independent and identically distributed data). Furthermore, an autonomous robot is immersed in a noisy and dynamic environment where its actions are bound to modify its very perception of the environment (the perception-action loop). Last, robotics perfectly illustrates the learnability issues: robotic sensors with high definition result in a very high dimensional instance space, hindering the learning task; but if the sensor definition is low, this results in "perceptual aliasing": the robot mistakes different places, e.g. corners in a maze, as being the same.

6.3.1. Controller Optimization and Modelling

In robotics, control architecture are either explicitly modelized (e.g. subsomption architecture) or learned (e.g. evolutionary robotics).

From human to robot behavior

In collaboration with Lutins (U.Paris 8 - Cognitive Psychology) within a joint project, Vincent Besson's M2 stage focused on the modeling of a control architecture from observed behaviors of humans involved in

---

[9]Extending Population-Based Incremental Learning to Continuous Search Spaces, PPSN'98

[10]A. Rattle and M. Sebag. Avoiding the Bloat with Stochastic Grammar-based Genetic Programming. In P. Collet et al., eds, *Evolution Artificielle*, pp 254–266, LNCS 2310, Springer Verlag, 2002.

an exploration task with reduced perceptual capabilities (i.e. no vision). The control architecture was then implement on simulated robot so as to reproduce these experiments in silico. The resulting architecture is based on the traditional subsumption architecture which is extended to deal with several contexts. Simulated experiments showed comparable behaviors in silico compared to in vivo experiments with human subject, thus providing insightful results for psychologists.

Symbolic controllers

In the field of automatic design of controllers, Nicolas Godzik's PhD [2] (defended in September 2005) focused on improving neural-network-based controllers. As a matter of fact, as the rewards for the robot actions are delayed, sparse and noisy, e.g., due to perceptual aliasing, and the state space is huge, an alternative to reinforcement learning is offered by Evolutionary Robotics, pioneered by Nolfi and Floreano (2000). Along this line, the robot controller is sought for by evolutionary computation, as the goal at hand, e.g., exploring the arena, getting to a specific location in a maze, combining exploration and search for a source of energy, is formalized as optimizing a manually designed fitness function. A central issue for Evolutionary Robotics is to decide whether the search space is discrete (as for classifier systems) or continuous (using, e.g., neural nets as the search space). A hybrid approach called "Symbolic Controllers" was proposed[11]. It involves continuous inputs and high-level outputs (discrete outputs that are transformed in continuous command for the actuators using either hand-made or previously-learned behaviors). Compared to the now standard Action Selection framework, this approach allows to gradually construct and exploit libraries of hierarchical actions, enforcing the scalability of the approach toward more complex target behaviours.

In a similar framework, Agustin Martinez, graduate student from University of Buenos Aires (Argentina) spent 3 months in TAO and studied the influence of the milieu on the results of evolution: even for the simple problem of obstacle avoidance, punctuated equilibrium-like phenomena can take place provided the milieu is complex enough [51].

Anticipation

Another issue is the generality of the robotic controller and the adaptation to environments that are different from those encountered during the (limited) training period. In collaboration with LIMSI (Robea project), an architecture combining three functionalities has been proposed[12]: action selection, anticipation (predicting the next robot sensations based on the selected action) and adaptation (reacting to the difference between the predicted sensations and those actually encountered by the robot). This architecture, inspired from cognitive science and sensory-motor contingency, contains a model of the world (the anticipation module) which can be costlessly confronted with the world and provides hints into the needed adaptation. Interestingly, this approach was shown to be much more robust in the long run than the competitive auto-teaching approach proposed by Nolfi and Parisi [2].

Cedric Hartland built on this architecture during his Master thesis [50]: he used anticipation to automatically calibrate a controller learned in a simulated world when it is implemented in a real robot. He is now continuing this work during his PhD, started in September 2005.

Fuzzy controllers and a priori rules

An alternative to neural networks as controllers for autonomous robots is the use of fuzzy controllers. Starting from the function approximation using Voronoi diagrams (see section 6.4.1, Carlos Kavka, a student at University of San Luis (Argentina) working under the supervision of Marc Schoenauer developed an evolutionary system to design fuzzy systems in inverse problems. This approach is original in the fuzzy system domain, because the algorithm here evolves rules whose domain is not restricted to hyper-rectangles, as in most

---

[11]N. Godzik and M. Schoenauer and M. Sebag, Evolving Symbolic Controllers, in G. Raidl et al., eds, *Applications of Evolutionary Computing*, pp 638-650, LNCS 2611, Springer Verlag, 2003

[12]N. Godzik and M. Schoenauer and M. Sebag, Robustness in the long run: Auto-teaching vs Anticipation in Evolutionary Robotics. In X. Yao et al., Eds, *Proc. PPSN VIII*, pp 932-941, LNCS 3242, Springer Verlag, 2004

previous work. Moreover, it is possible in this approach to define a priori rules (e.g., for an obstacle avoidance problem for an autonomous robot, "Go forward if there is no obstacle ahead") without defining strictly its domain of application: the system can restrict or expand this rule depending on the other rules it will find useful. The application to evolutionary robotic[13] demonstrated the efficiency of the proposed approach for fuzzy system design. Rcently, this work was extended to recurrent fuzzy systems [32], where some memory is required to perform the target task (e.g. a light on the right hand side of the corridor indicates that the robot must turn right in the next corridor).

6.3.2. *Cartography for autonomous robots*

Localization and Mapping is a very important thematic in autonomous robotics and has long been studied in the litterature. A basic definition of mapping is to incrementaly build a map of the environment so as to be able to navigate and relocalize in this very environment. In this context, the goal of the PhD thesis of Sylvain Gelly is to endow such a mapping algorithm with generalization capability so as to be able to build more compact map as well as to provide an efficient way to detect similarities (e.g. to speed up exploration). A probabilistic representation is studied that relies on extensions of traditional Hidden Markov Models and Bayesian networks. Several (national and international) papers have been published on this topic about the representation used as well as theoretical properties of the underlying concepts.

6.3.3. *Automatic design of robots and objects*

Aside from control-related problematic, Research also focus on the automatic design of robotic morphology. The first goal is to identify the relevant building blocks using simple lego-like structure to achieve construction such as Lego robot, or, as a first step, bridges and tables. A second goal is then to perform a relevant optimization in such a space to discover the fitted structure (current research focus on the use of Genetic Programming). Alexandre Devert started a PhD thesis this september and already worked on this subject during his Master thesis [48]. However, to make things simple, he started to evolve static structures, removing the controller issue (see section 6.4.1). But one long-term goal of his PhD is to actually come back to robotics and design both the morphology and the controller of a robot.

Remark: Several other *Mater theses* dealing with autonomous robotics have been made by students in TAO, but only the ones that are being continued as a regular activity have been listed here.

See also section 5.3 for the description of the robot simulator that has been developed by Jérémie Mary within this theme.

6.3.4. *Dynamic programming*

OpenDP, a framework for stochastic dynamic programming optimization has been developed by TAO (see section 5.5). Stochastic dynamic programming is a classical tool for discrete-time control problems, but its use for robotics, thanks to the combination with learning techniques with a good generalization ability, is a prospective issue. Some scalable benchmarks of robotics are already included and some others are under development for a systematic study of the possibilities.

6.4. Inverse problems

Participants: Nicolas Bredèche, Antoine Cornuéjols, Alexandre Devert, Mohamed Jebalia, Matthieu Pierres, Marc Schoenauer, Michèle Sebag, Vijay Pratap Singh.

Abstract:

Inverse Problems (IP) aim to determine unknown causes based on the observation of their effects. In contrast, direct problems are concerned with computing the effects of (exhaustively described) causes. Inverse problems are often mathematically ill-posed in the sense that the existence, uniqueness and stability of solutions cannot be assured.

---

[13]C. Kavka and M. Schoenauer. Evolution of Voronoi-based Fuzzy Controllers. In Xin Yao et al., eds, PPSN'04, LNCS 3242, Springer Verlag, 2004.

IPs are present in many areas of science and engineering, such as mechanical engineering, meteorology, heat transfer, electromagnetism, material science, etc. The TAO project has focused on the problems of system identification, modeling physical (mechanical, chemical, biological, etc.) phenomena from available observations and current theories.

The key issue in Inverse Problem is the choice of the search space, i.e., in Evolutionary Computation terminology, of the representation.

6.4.1. *Topological Optimum Design*

Several results have been obtained by TAO team members (Marc Schoenauer and his former PhD student Hatem Hamda – PhD defended in 2003) regarding the Topological Optimum Design of Mechanical Structures using the so-called Voronoi representations for structures, that overcome many limitations of the more widely-used bitarray representation[14]. In collaboration with *EZCT* architecture, we have worked on the elementary benchmark of architectural schools, the design of a chair. Results of this work have been exposed in the *Innovative Design Techniques* section of the *ArchiLab* exhibition, an architectural exhibition in Orléans in Fall 2004, and published in an architectural journal [9]. This cooperation is now being officialized through a research contract with *ECZT*, and Alexandre Devert, who just started his PhD under the supervision of Marc Schoenauer and Nicolas Bredèche, started to work on the automatic building of construction plans during his Master 2 [48]: compared to previous work, this original approach ensures the contructibility of the resulting structure. The continuation of this work has been acceptetd for publication in EuroGP'06 [22], and the work will now focus on technologically vaible designs, with as mid-term objective the small-scale industrialisation through EZCT collaboration.

On the other hand, Mohamed Jebalia continues his PhD work under Marc Schoenauer's supervision, to investigate other representations based on Genetic Programming for the same problem, trying to grasp the holy grail of modularity in the representation. A good example is such modularity is given by crane structures, where the same elements are duplicated to build the complete crane - whereas an evolutionary algorithm must "discover" that structure again and again.

6.4.2. *Geologically sound representation for seismic inversion*

Vijay Pratap Singh's PhD thesis in geophysics, funded by IFP : *Représentations géophysiquement fondées pour la reconstruction du profil de vitesses du sous-sol par algorithmes évolutionnaires*, is a very good example of why domain knowledge must be included in the representation itself - and how it can be. This inverse problem of geophysics aims at identifying some underground characteristics from recorded seismic data. A previous PhD (F. Mansanné, Université de Pau, 2000) had used the same Voronoi representation proposed for Structural Design problem, obtaining some satisfactory results ...but many other results, though actually optimizing the target Least Square objective function, were geophysically absurd (as any 7-years-old geophysicist would have immediately noticed). The proposed representation now evolves an initial state of layered underground as well as the geological conditions across the geological ages and tries to fit the resulting underground to the seismic data. Such representation at least ensures the geophysical relevance of the identified underground. First results, on the purely geological problem, demonstrated the power of this approach [41], [40]. On-going work is concerned with the complete geophysical problem.

6.4.3. *Representations for isotherm law in chromatography*

In the framework of the ACI NIM *Nouvelles Interfaces des Mathématiques*, Marc Schoenauer is part of the *Chromalgema* project, whose aim is the identification of the isotherm function in analytical chromatography. This is an inverse problem for which the direct problem is solved by standard numerical approaches (e.g. Godunov scheme for Non-Linear Hyperbolic Systems).

When the unknown isotherm function is sought as a rational fraction of the concentrations (e.g. in the so-called "Langmuir" model), the inverse problem amounts to parametric optimization. It can then be solved by

---

[14]H. Hamda, F. Jouve, E. Lutton, M. Schoenauer and M. Sebag. Compact Unstructured Representations in Evolutionary Topological Optimum Design. *Applied Intelligence*, 16, pp 139–155, 2002.

Evolution Strategies, coupled with standard deterministic gradient-based methods. A recent improvement has been brought by replacing the now-standard "self-adaptive" ES by the recent "CMA-ES" (see section 6.2.1).

But a more complex situation is when no known model is available for the isotherm function. Two innovative representations have been proposed in that case:

- Evolving a set of virtual examples: From those examples, the isotherm is computed by a Support Vector Machine algorithm. This representation allows one to take into account actual measurements made by the chemist engineers, including them as fixed data in the set of example[15]. Moreover, using SVM should scale up for a large number of components.

- Using Genetic Programming: Mohamed Jebalia is now working on using GP to represent the unknown isotherm function, trying to identify both the degree of the model and its parameters.

6.4.4. *Variable selection in chemical engineering*

Anne Auger and Marc Schoenauer are also member of another project in the ACI NIM framework, Contrôle quantique, headed by Claude LeBris (Cermics, ENPC). Anne Auger addressed the optimization (using Evolution Strategies) of the laser characteristics to better align the molecules in order to control a chemical reaction. But one of the key issues in quantum control is that only a few among the many variables are actually useful to control. On-going work is concerned with the identification of those variables – and hence makes the bridge with what is called in Machine Learning field *Feature Selection*. The particular context of optimization should allow to use mixed techniques pertaining to both Evolutionary Computation (e.g. monitor the standard deviations associated with each variable to detect the "good" ones) and Machine Leaning (e.g. "learn" to discriminate between the good and the bad points, and use Data Mining Feature Selection to eliminate non-discriminant variables).

The first part of this work program has now been tackled by Mikhail Zaslavskiy during his *Stage d'Option de l'École Polytechnique*, who proposed several ways to identify useful variables, both ex abrupto and within the evolutionary optimization process (CMS-ES in this case) [53].

6.4.5. *Automatic design of mesh topologies*

Another closely related Inverse Problem is known as Feature Construction, mapping the problem at hand onto a space more amenable to the resolution of the problem. The Feature Construction problem, subsuming the Feature Selection one, is central to Artificial Intelligence in general, and Machine Learning and Data Mining in particular.

The problem of Feature Construction is studied on an application, concerned with characterizing good meshes for numerical engineering, particularly the design of 3D meshes in the aerospace industry (Ph.D. Mathieu Pierres, Airbus CIFRE, co-advised by Marc Schoenauer and Michèle Sebag). This challenging real-world application involves relational issues (a mesh is but a set of finite elements and their relations) and probabilistic issues (as usual for real-world applications, the solution is to be sought as a trade-off between conflicting logical rules).

In order to characterize "what is a good mesh", a representation had to be designed, that allows one to represent efficiently for the subsequent learning task very different meshes – in particular, meshes with very different number of nodes, edges, surfaces and elements. This has been achieved by going at the level of a block (or element). Blocks are all described by a fixed set of features, which enables to use any learning algorithm to discriminate good blocks from bad blocks ...provided you have examples of bad blocks. Hence a first task was to build "plausible" bad blocks. Then standard learning algorithms can be applied to characterize good blocks. A good mesh is then a mesh which contains only good blocks - of whose blocks are as good as possible. Next step will be to start generating topologies automatically. Fist, a parametrized script will be used, and only its parameters will be optimized. Later, the complete topology will be designed by the Evolutionary

---

[15]Another representation allowing the user to specify known points for the target law has been implemented in collaboration with Carlos Kavka, in the framework of fuzzy controllers for robots – see Evolutionary Robotics section.

Algorithm ,using a variable length representation – the series of "cuts" that are to be made to generate a complete topology.

6.5. Other applications

Participants: Jérôme Azé, Thomas Heitz, Yves Kodratoff, Samuel Landau, Claire Le Baron, Mathieu Roche, Marc Schoenauer, Michèle Sebag, Yann Semet.

Abstract: The TAO group is also historically involved in other applications of either Machine Learning or Evolutionary Computation that are not directly linked to its main streams of research. They are surveyed below.

6.5.1. *Text Mining*

Text Mining (TM) is concerned with exploiting/transforming documents to achieve particular tasks. The difficulty lies in the delicate balance to keep between texts, transformations and tasks. Problem resolution implies the existence of cognitive entities, called concepts of specialty, necessary to the resolution of the current tasks.

A key data preparation step in Text Mining, Term Extraction selects the terms, or collocation of words, attached to specific concepts. The task of extracting relevant collocations can be achieve through a supervised learning algorithm, exploiting a few collocations manually labelled as relevant/irrelevant. In our work, an evolutionary learning algorithm (ROGER, see section 6.1.1), based on the optimization of the Area under the ROC curve criterion, extracts an order on the candidate terms. The robustness of the approach is demonstrated on two real-world domain applications, considering different domains(biology and human resources) and different languages (English and French) [13], [14].

The team has organized the first french text-mining challenge (DEFT : DÉfi Fouille de Textes), which consisted of removing the non relevant sentences from french corpora of political speeches. It took place in a workshop of TALN'05 conference with thirty participants belonging to eleven french teams. The results by the participating teams are presented in [11].

6.5.2. *Train scheduling*

Scheduling problems are a known success area for Evolutionary Algorithms (EAs). The French Railways (SNCF) were interested to find out whether they could benefit from EAs to tackle the problem of rescheduling the trains after an incident has perturbed one (or more) train(s). At the moment, they are using commercial software (*CPlex* from Ilog), and they experience serious difficulties when several incidents occur simultaneously on large networks.

The first results in 2004 on a simplified instance of the problem on a small network have shown that in fact an iterated hill-climber can solve the problem better than any other algorithm (including CPlex and complex EAs). Significant results on full size problems have later been obtained [38], showing that EAs can indeed give better and faster results that *CPlex* thanks to a very specialized "scheduler", result of Yann Semet's one year work. This motivated a renewal of the contract by SNCF for an additional year. Moreover, two publications in *Rail et Recherche*, the SNCF internal mazazine, have mentionned this work.

6.5.3. *Time-dependent planning with bounded resources*

A one-year contract between TAO and Thales - Land & Vision division was concerned with temporal planning with limited resources. Two approaches have been tried: coupling Evolutionary Algorithms on the global scale with Constraint Programming to solve local (hopefully small) problems on the one hand; and using Petri Networks, representing partial plans, inside some Parisian-like Evolutionary Algorithm to derive a global optimal plan. The *ATNoSFERES* program (Samuel Landau) is being used for the latter approach.

Only the first approach succeeded, and feasibility results for the *TGV* approach have been obtained during the contact (that ended in July 2005). Collaboration will continue with Pierre Savéant (Thales) and Vincent Vidal (Université de Lens) as this approach allowed to obtain breakthrough results : it is the first Pareto approach for multi-objective temporal planning problems [36].

6.5.4. Multi-disciplinary multi-objective otpimization

TAO team is involved in the RNTL project on Multi-disciplinary optimization coordinated by Rodolphe Leriche (Ecole des Mines de St-Etienne), for its expertise in multi-objective optimization (see section 6.2.3). Claire Le Baron started a PhD in September 2005 co-funded by the automobile company Renault, whose goal is to optimize the complete motor of a car, thus involving strutural mechanics, vibration and acoustics, combustion and thermics. Evolutionary Algorithms are good candidate for such ill-posed problems – but the thesis will focus on comparing a wide range of approaches invooving some trade-off between solving a faithfull but complex optimization problem (using Evolutionary Algorithms), and solving an easy but maybe not really significant optimization problem using standard numerical methods.

7. Contracts and Grants with Industry

7.1. Contracts and Grants with Industry

Keywords: *Airbus*, *Chromalgena*, *IFP*, *Quantum chemistry*, *RNTL*, *Renault*, *SNCF*, *Thalès*.

Contracts managed by Inria

- **Chimie Quantique**, CNRS Program ACI-NIM (New Interfaces of mathematics) – 2004-2007 (8 kEur), coordinator Claude Le Bris(Cermics) ;
  participants : Marc SCHOENAUER, Anne AUGER (section 6.4.4).

- **Chromalgena**, CNRS Program ACI NIM (New Interfaces of mathematics) – 2003-2006 (14 kEur), coordinator F.James (Université d'Orléans) ;
  participant : Marc SCHOENAUER (section 6.4.3).

- **AIRBUS** – 2004-2007 (45 kEur), side-contract to Mathieu PIERRES's CIFRE Ph.D. (section 6.4.5).

- **IFP** – 2003-2006 (18 kEur), side-contract to Vijay Pratap SINGH's CIFRE Ph.D. (section 6.4.2).

- **SNCF** – 2004 (80 kEur) : research contract, Yann SEMET, expert engineer (section 6.5.2). A follow-up contract (80 kEur) has been signed in June 2005 for one additional year.

- **Thalès** – 2004-2005 (50 kEur) : research contract (section 6.5.3). This contract ended in July 2005, but the collaboration continues (submitted publication).

- **ONCE-CS** – 2005-2008 (147 kEur) European *Coordinated Action* from FP6. Coordinator Jeff Johnson, Open University, UK.

- **OMD-RNTL** – 2005-2008 (72 kEur) Coordinator Rodolphe Leriche, Ecole des Mines des St Etienne.

- **Renault** – 2005-2008 (45 kEur) side-contract to Claire LEBARON's CIFRE Ph.D. (section 6.5.4). Not yet notified.

- **EZCT** – 2005-2006 (10 kEur) side-contract to Alexandre Devert's PhD (section 6.4.1). Not yet notified.

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 a WP.

- **Neurodyne** ACI-NIM (New Interfaces of mathematics) – 2003-2006 (6 kEur). Coordinator Sylvain Baillet, LENA, Hôpital La Pitié-Salpètriere, CNRS UPR 640. Participants: M. Sebag, O. Teytaud, A. Cornuéjols.

- **Traffic** ACI-NIM (New Interfaces of mathematics) – 2004-2007 (17,5 kEur). Coordinator Jean-Michel Loubès, Project SELECT. Participants: M. Sebag, O. Teytaud.
- **Télémédecine** ACI technologie de la médecine – 2002-2005 (15,245 kEur). Coordinator Marie-Christine Jaulent, Hôpital Broussais. Participant: M. Sebag.
- **Indana**, ACI NIM – 2002-2005 (15 kEur). Coordinator Marie-Christine Jaulent, Hôpital Broussais. Participant: A. Cornuéjols.
- **Egide**, Slovenia, coll. Nada Lavrac, Institut Jozef Stefan, Ljubljana; Croatia, coll. Dragan Gamberger, Zagreb, U. Mazaryk.
- **AGIR** ACI Masses de Données (section 6.1.7) - 2004-2007 (260 kEur) Coordinator Cécile Germain-Renaud. Participants are from CRAN, CREATIS, INRIA-Sophia, I3S, LPC, LIMSI, LORIA, PCRI, Centre Antoine Lacassagne, CHRU Clermont-Ferrand , Tenon Hospital, Fédération de la Mutualité Parisienne.
- **Herobot**, TCAN CNRS – 2004-2006 (28 kEur). Participant and coordinator : Nicolas Bredeche.

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.

8.1.2. Collaborations with joint publications

- Universidad de San Luis, Argentina [32].
- Université Laval à Québec [23].

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.

8.2.2. Working groups

- EGEE, Enabling Grids for E-SciencE : Cécile Germain-Renaud is a member, 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, member of Executive Committee since 2005.

8.2.3. Collaborations with joint publications

- Colab, ETH Zurich [12], [44].
- 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.

- 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.

- FERA, Fédération des Equipes de Recherche en Apprentissage: Michèle Sebag co-organized the first venue in May 2005 (Apprentissage Pro-Actif) which led to the constitution of FERA, and co-organized the first FERA-Day, October 21th, 2004[16].

- Rencontres Représentation des Données et des Connaissances (RDC) was co-organized by G. Hébrail and M. Sebag, March 21st, 2005, at ENST.

- Apprenteo (June 24th, 2005) was organized by Michèle Sebag, as the first meeting devoted to Machine Learning and gathering members of the Digiteo Labs (PCRI, CEA, SupElec, LIMSI, CMAP).

8.3.2. Management positions in scientific organizations

- National research program on knowledge management, machine learning and new technologies ACI TCAN Traitement des Connaissances, Apprentissage, Nouvelles Technologies: Michèle Sebag, member of the steering committee, 2002-2004; Antoine Cornuéjols, member of the steering committee since 2004.

- CNRS Network "Discovering and Summarizing", Réseau Thématique Pluridisciplinaire Découvrir et résumer, RTP 12: Michèle Sebag, member of the steering committee since 2002.

- Member of the Steering Committee PASCAL NoE, M. Sebag.

8.3.3. Associations

- *Evolution Artificielle* : Marc Schoenauer, founding president (1995-2004), now member of the Executive Committee.

- *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 ; Jérémie Mary, treasurer 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

- Université de Lens [36].
- IASI group, LRI ([42]).
- Bioinformatics, LRI [20], [43].
- Situated Perception group (Perception Située), LIMSI-CNRS.

---

[16]http://proml.lri.fr

8.4. Honors

8.4.1. Keynote addresses

- CERCIA Workshop on Computational Intelligence and Applications, 15-16 November 2005, Marc Schoenauer.
- IFP, journée *Algorithmes Génétiques*, 10 novembre 2005, Marc Schoenauer.

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, 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

- Marc Schoenauer was Member of the Organizing Committee of the *1st European Conference on Complex System* in Paris, 14-18 Nov. 2005, and co-edited the Proceedings [1].

9.1.3. Program Committee Member (international events)

- 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, ...
- Michèle Sebag: Area Chair of ICML 05, 22nd International Conference on Machine Learning, ECML-PKDD 05, 16th European Conference on Machine Learning, 9th Conference on Principle and Practice of Knowledge Discovery from Databases, ECAI 06, 17th European Conference on Artificial Intelligence;
  PC Member of IJCAI 05, 19th International Conference on Artificial Intelligence; ICDM05, IEEE International Conference on Data Mining, 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, ...

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: Michèle Sebag since 1994.
- EGC, Extraction et Gestion des Connaissances: M. Sebag since 2002.

9.1.5. Evaluation committees and invited expertise

- Antoine Cornuéjols, reviewer of the National research program ACI Masses de données.
- Cécile Germain-Renaud, reviewer of the National research program ACI Masses de données.
- Marc Schoenauer, reviewer of 3 European projects (STREPS) from FP5 framework in the Neuro-IT program: SIGNAL, HYDRA and POETIC. Final review took place in 2005. Professor evaluation for Profs I. Parmee (U. of Bath, England), B. Paechter (Napier U., Edinburgh, Scotland) and J. Rowe (Birmingham U., England). Tenure evaluation for Profs S. Luke (George Mason U., Fairfax, VA, USA) and M. Vose (U. of Tennessee at Knoxville, TE, USA).
- Michèle Sebag, reviewer for US National Science Foundation programme on Data Mining, June 27-28, 2005; reviewer of the European project New Ties (STREPS FP6-IST), November 10, 2005; reviewer for U. Leuven, professor evaluation; reviewer for U. Tufts, tenure report; member of the CNRS evaluation committee for the TIMC Lab, Grenoble.
- Michèle Sebag, reviewer of the National research program ANR Masse de Données, Projet Blanc.

9.1.6. Other evaluation activities

- Reviewer for PhD dissertation: Marc Schoenauer (4) ; Michèle Sebag (4) ; Antoine Cornuéjols (1).
- Reviewer for Habilitation: Michèle Sebag (1)

9.1.7. Brevets

- Marc Schoenauer et Vijay Pratap Singh, *Méthode pour construire des coupes géologiques équilibrées par algoritmes évolutionnaires*, coll. Michel Léger, IFP, numéro d'enregistrement national 05/09 023.

9.1.8. Summer schools, tutorials, invited seminars

- Invited seminars, Michèle Sebag: Università di Bari, march (6 hours); Institut Pasteur, mars; Ecole Nationale Supérieure des Télécommunications, april; 2nd Franco-Japonese Wshop, Lyon-II, june; ENS-Cachan, september.
- Invited seminars, Antoine Cornuéjols: LIMSI, Université Paris-Sud, Orsay, 2005.

9.2. Enseignement

9.2.1. Defended habilitations

- Cécile Germain, juin 2005, *Contributions à la modélisation et à l'optimisation des systèmes de calcul à grande échelle*.
- Antoine Cornuéjols, décembre 2005.

9.2.2. Defended doctorates

- Nicolas Godzik, 29/09/05, Bourse INRIA (Robea), Université Paris Sud
- Jérôme Maloberti, 28/06/05, Université Paris Sud
- Jérémie Mary, 12/12/05, Bourse MNRT, Université Paris Sud

9.2.3. Graduate courses

- DEA I3, Data mining and machine learning : Michèle Sebag, Antoine Cornuéjols.
- DEA ECD, Optimization for data mining: M. Sebag.
- 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)

9.2.4. Other research-related teaching activities

- *Ecole Polytechnique*, Projects in Evolutionary Robotics in the *Modex d'Electronique*: Marc Schoenauer, Cédric Hartland.
- *Ecole Polytechnique*, *Stages d'option*: Michèle Sebag, Marc Schoenauer.
- ENPC (*Ecole Nationale des Ponts et Chaussées*), in charge of the *Optimization* module in the *Applied Math* track: Marc Schoenauer.
- ENSTA (*Ecole Nationale Supérieure de Techniques Avancées*), in charge of the *Machine Learning*: Antoine Cornuéjols.

10. Bibliography

Books and Monographs

[1] P. BOURGINE, F. KÉPÈS, M. SCHOENAUER (editors). *Proceedings of the First European Conference on Complex Systems – ECCS'05*, 2005.

Doctoral dissertations and Habilitation theses

[2] N. GODZIK. *Une approche évolutionnaire de la robotique modulaire et anticipative*, Ph. D. Thesis, Université de Paris Sud – Orsay, 2005.

[3] J. MALOBERTI. *Improving Inductive Logic Programming with Constraint Satisfaction Techniques: Applications to Frequent Query Discovery*, Ph. D. Thesis, Université de Paris Sud – Orsay, 2005.

[4] J. MARY. *Étude de l'apprentissage actif. Application à la conduite d'expériences*, Ph. D. Thesis, Université de Paris Sud – Orsay, 2005.

Articles in refereed journals and book chapters

[5] A. AUGER. *Convergence results for $(1,\lambda)$-SA-ES using the theory of ϕ-irreducible Markov chains*, in "Theoretical Computer Science", n° 334(1-3), 2005, p. 35–69.

[6] J. BERNAUER, A. POUPON, J. AZÉ, J. JANIN. *A docking analysis of the statistical physics of protein-protein recognition*, in "Journal Physical Biology", 2005.

[7] C. GERMAIN, V. BRETON, P. CLARYSSE, Y. GAUDEAU, T. GLATARD, E. JEANNOT, Y. LEGRÉ, C. LOOMIS, J. MONTAGNAT, J.-M. MOUREAUX, A. OSORIO, X. PENNEC, R. TEXIER. *Grid-enabling medical image analysis*, in "Journal of Clinical Monitoring and Computing", Extended version of the BioGrid 2005 paper, vol. 19, nº 4-5, 2005, p. 339-349, http://dx.doi.org/10.1007/s10877-005-0679-9.

[8] C. GERMAIN, R. TEXIER, A. OSORIO. *Exploration of Medical Images on the Grid*, in "Methods of Information in Medecine", vol. 44, nº 2, 2005, p. 227-232.

[9] P. MOREL, H. HAMDA, M. SCHOENAUER. *Computational Chair Design using Genetic Algorithms*, vol. 71, nº 3, 2005, p. 95-99.

[10] A. SELIKHOV, C. GERMAIN. *A channel memory based environment for MPI applications*, in "Future Generation Computer Systems", vol. 21, nº 5, 2005, p. 709-715, http://dx.doi.org/10.1016/j.future.2004.05.011.

Publications in Conferences and Workshops

[11] E. ALPHONSE, A. AMRANI, J. AZÉ, A.-D. M. THOMAS HEITZ, M. ROCHE. *Préparation des données et analyse des résultats de DEFT'05*, in "DEFT'05", 2005.

[12] A. AUGER, M. SCHOENAUER, O. TEYTAUD. *Local and global oder 3/2 convergence of a surrogate evolutionary algorithm*, in "GECCO - genetic and evolutionary computation conference, Washington", 2005, p. 857-864, http://hal.inria.fr/inria-00000540/en/.

[13] J. AZÉ, M. ROCHE, Y. KODRATOFF, M. SEBAG. *Preference Learning in Terminology Extraction: A ROC-based approach*, in "Applied Stochastic Models and Data Analysis – ASMDA'05", 2005.

[14] J. AZÉ, M. ROCHE, M. SEBAG. *Bagging Evolutionary ROC-based Hypotheses Application to Terminology Extraction*, in "ROCML 05", 2005.

[15] N. BASKIOTIS, M. SEBAG, O. TEYTAUD. *Inductive-deductive systems : a learning theory point of view*, in "CAP'05", 2005, 2.

[16] J. BERNAUER, A. POUPON, J. AZÉ, J. JANIN. *A docking analysis of the statistical physics of protein-protein recognition*, in "JOBIM'05", 2005.

[17] Y. BONNEMAY, M. SEBAG, O. TEYTAUD. *Convergence proofs, convergence rates and stopping criterions for multi-modal or multi-objective evolutionary algorithms*, in "Evolution Artificielle, Lille", 2005, 12, http://hal.inria.fr/inria-00000545/en/.

[18] N. BREDECHE, L. HUGUES. *Evolutionary Robotics : incremental learning of sequential behavior*, in "Fourth IEEE International Conference on Development and Learning (ICDL 2005)", 2005.

[19] N. BREDECHE, L. HUGUES. *Speeding up Learning with Dynamic Environment Shaping in Evolutionary Robotics*, in "Fifth international Workshop of Epigenetic Robotics (Epirob2005)", 2005, p. 137-138.

[20] A. CORNUÉJOLS, C. FROIDEVAUX, J. MARY. *Comparing and combining feature estimation methods for the analysis of microarray data*, in "JOBIM-05 : Journées Ouvertes Biologie Informatique Mathématiques, Lyon, France", 2005.

[21] A. CORNUÉJOLS, M. SEBAG. *Phase transition and inductive learning*, in "Second Franco-Japanese Workshop on Information Search, Integration and Personalization (ISIP-05), Lyon, France", 2005.

[22] A. DEVERT, N. BREDECHE, M. SCHOENAUER. *Blindbuilder : A new encoding to evolve Lego-like structures*, in "EuroGP'06", M. EBNER (editor). , To appear, Springer Verlag, 2006.

[23] C. GAGNÉ, M. SCHOENAUER, M. TOMASSINI, M. PARIZEAU. *Genetic Programming, Validation Sets, and Parsimony Pressure*, in "EuroGP'06", M. EBNER (editor). , To appear, Springer Verlag, 2006.

[24] S. GELLY, N. BREDECHE, M. SEBAG. *From Factorial and Hierarchical HMM to Bayesian Network: A Representation Change Algorithm.*, in "SARA", 2005, p. 107-120.

[25] S. GELLY, N. BREDECHE, M. SEBAG. *HMM hiérarchiques et factorisés: mécanisme d'inférence et apprentissage à partir de peu de données.*, in "CAP", 2005, p. 143-144.

[26] S. GELLY, N. BREDECHE, M. SEBAG. *Inference dans les HMM hierarchiques et factorises : changement de representation vers le formalisme des Reseaux Bayesien*, in "EGC, Atelier sur les Modèles Graphiques Probabilistes", 2005, p. 35-44.

[27] S. GELLY, J. MARY, O. TEYTAUD. *Taylor-based pseudo-metrics for random process fitting in dynamic programming.*, in "PDMIA, Lille", publié Non, 2005, p. 21-36, http://hal.inria.fr/inria-00000217/en/.

[28] S. GELLY, O. TEYTAUD, N. BREDECHE, M. SCHOENAUER. *A statistical learning theory approach of Bloat*, in "GECCO - genetic and evolutionary computation conference, Washington", 2005, p. 1783-1784, http://hal.inria.fr/inria-00000546/en/.

[29] S. GELLY, O. TEYTAUD. *Bayesian networks : a better than frequentist approach for parametrization, and a more accurate structural complexity measure than the number of parameters*, in "CAP, Nice", 2005, p. 147-162, http://hal.inria.fr/inria-00000541/en/.

[30] C. GERMAIN, V. BRETON, P. CLARYSSE, Y. GAUDEAU, T. GLATARD, E. JEANNOT, Y. LEGRÉ, C. LOOMIS, J. MONTAGNAT, J.-M. MOUREAUX, A. OSORIO, X. PENNEC, R. TEXIER. *Grid-enabling medical image analysis*, in "3rd Bio-Grid Workshop at IEEE CCGRID", 2005.

[31] C. GERMAIN-RENAUD, D. MONNIER-RAGAIGNE. *Grid Result Checking*, in "Conf. Computing Frontiers", ACM, 2005, p. 87-96, http://doi.acm.org/10.1145/1062280.

[32] C. KAVKA, P. ROGGERO, M. SCHOENAUER. *Evolution of Voronoi based fuzzy recurrent controllers*, in "GECCO'2005", H.-G. B. ET AL. (editor). , ACM Press, ACM-SIGEVO, 2005, p. 1385-1392,

http://hal.inria.fr/inria-00000539/en/.

[33] S. LANDAU, O. SIGAUD, M. SCHOENAUER. *ATNoSFERES revisited*, in "Proc. GECCO-2005, Washington DC", H.-G. BEYER, ET AL. (editors). , ACM Press, june 25-29 2005, p. 1867-1874, http://hal.inria.fr/inria-00000158/en/.

[34] N. PERNOT, A. CORNUÉJOLS, M. SEBAG. *Phase transition in grammatical inference*, in "Int. Joint Conf. on Artificial Intelligence (IJCAI-05), Edinburgh, UK", 2005, p. 811-816.

[35] N. PERNOT, A. CORNUÉJOLS, M. SEBAG. *Phase transition in grammatical inference*, in "CAP-05, Nice", 2005, p. 49-60.

[36] M. SCHOENAUER, P. SAVÉANT, V. VIDAL. *Divide-and-Evolve: a New Memetic Scheme for Domain-Independent Temporal Planning*, in "EvoCOP'06", J. GOTTLIEB, G. RAIDL (editors). , To appear, Springer Verlag, 2006.

[37] M. SEBAG, N. TARRISSON, O. TEYTAUD, J. LEFEVRE, S. BAILLET. *A multi-objective multi-modal optimisation approach for mining stable spatio-temporal patterns*, in "IJCAI'05", 2005, 6p.

[38] Y. SEMET, M. SCHOENAUER. *An efficient memetic, permutation-based evolutionary algorithm for real-world train timetabling*, in "CEC'05", B. MCKAY, ET AL. (editors). , IEEE Press, 2005, p. 661-667, http://hal.inria.fr/inria-00000538/en/.

[39] V. P. SINGH, B. DUQUET, M. LÉGER, M. SCHOENAUER. *Velocity determination in foothills using evolutionary algorithms*, in "Thrust Belts and Foreland Basins", IFP, 2005.

[40] V. P. SINGH, M. LEGER, M. SCHOENAUER. *Foothill Simulation using multi-objective evolutionary Algorithms*, in "AAPG'05", American Association of Petroleum Geologists, 2005.

[41] V. SINGH, M. SCHOENAUER, M. LÉGER. *A geologically-sound representation for evolutionary multi-objective subsurface identification*, in "CEC'05", B. MCKAY, ET AL. (editors). , IEEE Press, 2005, p. 454-462, http://hal.inria.fr/inria-00000855/en/.

[42] A. TERMIER, M.-C. ROUSSET, M. SEBAG, K. OHARA, T. WASHIO, H. MOTODA. *Computation-time efficient and robust attribute tree mining with DryadeParent*, in "ECML/PKDD'05", 2005.

[43] A. TERMIER, M.-C. ROUSSET, M. SEBAG, K. OHARA, T. WASHIO, H. MOTODA. *Efficient Mining of High Branching Factor Attribute Trees*, in "ICDM 2005", IEEE Press, 2005.

[44] O. TEYTAUD, M. JEBALIA, A. AUGER. *XSE: Quasi-random mutations for evolution strategies*, in "Evolution Artificielle, Lille", 2005, p. 12-21, http://hal.inria.fr/inria-00000544/en/.

Miscellaneous

[45] M. AMIL. *Théorie VC pour la sélection de variables. Application au trafic routier*, Technical report, Université Paris-Sud, 2005.

[46] A. AMRANI, J. AZÉ, Y. KODRATOFF. *Logiciel d'aide à l'étiquetage morpho-syntaxique de textes de spécialité*, 2005, Session Démonstration de logiciels, EGC 2005.

[47] S. DEBERLES. *Un Observatoire de la Grille : schémas de fautes dans EGEE*, Technical report, Université Paris-Sud, 2005, http://www.lri.fr/~cecile/DEMAIN/DEA_Samuel_Deberles.pdf.

[48] A. DEVERT. *Evolution Artificielle de structures*, Technical report, Université Paris 6, 2005.

[49] C. GERMAIN-RENAUD. *Contributions à la modélisation et à l'optimisation des systèmes de calcul à grande échelle*, 2005, http://www.lri.fr/~cecile/RAPH/germainHdr.pdf, HDR report, Université Paris-Sud.

[50] C. HARTLAND. *Robotique Evolutionaire et Anticipation pour le controle*, Technical report, Université Paris-Sud, 2005.

[51] A. MARTINES. *Obstacle Avoidance: A Contribution to Evolutionary Robotics*, Technical report, Universidad de Buenos Aires and LRI, 2005.

[52] R. ROS. *Transition de phase en apprentissage artificiel. Pistes pour sa mise en évidence en robotique*, Technical report, Université de Paris Sud – Orsay, 2005.

[53] M. ZASLAVSKIY. *Méthodes de définition de sensibilité des variables à partir des résultats d'algorithmes évolutionnaires*, Technical report, École Polytechnique, 2005.