INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team tao

Thème apprentissage et optimisation

Futurs

THEME COG

Activity Report

2004

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]
Antoine Cornuéjols [MDC CNAM]
Olivier Teytaud [CR INRIA – since October 2004]

Administrative assistant

Marie-Carole Lopes [Tao/Select/Proval]

Ph.D. Students

Kamal ABBOUD [defended October 2004]
Anne AUGER [defended December 2004]
Nicolas BASKIOTIS [started September 2003]
Mary FELKIN [started September 2003]
Sylvain GELLY [started September 2004]
Nicolas GODZIK [to be defended early 2005]
Thomas HEITZ [started September 2004]
Mohamed JEBALIA [started September 2004, co-tutelle Université Tunis)]
Jérémie MARY [started September 2002]
Jérôme MALOBERTI [co-tutelle Univ. Yokohama]
Matthieu PIERRES [CIFRE Airbus, started February 2004]
Vijay PRATAP SINGH [CIFRE IFP, started November 2003]
Mathieu ROCHE [defended December 2004]
Olga ROUDENKO [defended March 2004]

Post-doctoral fellows

Jérôme AZÉ
Samuel LANDAU [since October 2004]
Shengwu Xiong [until November 2004]

Research Engineer

Yann SEMET

Collaborators

Yves KODRATOFF [DR CNRS]
Oriane MATTE TAILLIEZ [post-doc]

2. 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

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 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).

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

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 6.1.1,6.1.2,6.1.3) and Inverse problems for Numerical Engineering (see sections 6.4).

5. Software

5.1. Evolving Objects

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

Participant: Marc Schoenauer [correspondent].

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).

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], James Manley.

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. 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).

5.3. World in a bottle

Keywords: *OpenGL, Robot simulator*.

Participant: Jeremie Mary [correspondent].

"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.

Django is an algorithm of theta-subsumption of Datalog clauses. It is written in C by Jerome Maloberti and is freely available under the GNU Public License. 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 in order to find efficiently one or all the solutions.

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

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.

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. AUC-based Learning

While the mainstream of Statistical Learning focuses on well posed, quadratic, optimisation criteria, e.g., kernel methods, our expertise in evolutionary computation allows us to consider non convex optimisation criteria such as the Wilcoxon statistics, or area under the Receiver Operating Characteristics (ROC) curve. The ROC curve describes the trade-off between the two types of error of a hypothesis: the false positive and the false negative error rates. Among its main advantages, the ROC approach is adapted to imbalanced example distributions (e.g. one positive for 100 negative training examples), and to cost-sensitive learning (e.g. the cost of mistaking an ill person for a healthy one is much higher than vice versa).

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, as illustrated by applications in medical data mining and text mining [12][30][23].

6.1.2. Feature Selection

Most available databases were not constructed with data mining in mind, and they usually involve a number of features 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 so 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.

Along the same lines, the variability of solutions provided by stochastic optimisation can be exploited for sensitivity studies. In collaboration with the Vrije Universiteit of Amsterdam, we showed that AUC-based evolutionary optimization provides the means for Ensemble Feature Ranking [20]. This theoretical result extends the state of the art (see Special Issue on Feature and Variable Selection, Journal of Machine Learning Research, 2003) to non-linear target concepts, while only linearly increasing the size of the search space.

6.1.3. Human Computer Interaction and Visual Data Mining

Two main requirements are put on knowledge discovery and data mining systems. On one hand, the hypotheses that are produced must be intelligible to the experts, and, on the other hand, they must be accurate. To address these antagonistic requirements, a new graphical approach for comparing sub-populations, evaluating the expert's conjectures, etc., was proposed. The expert can interactively exploit precise and accurate hypotheses, asking for more details, and/or checking his/her own conjectures; the answers are expressed in terms of pictures, and human eyes are incomparably efficient at decoding the rich information conveyed by these pictures.

An estimation of the risk for cardio-vascular diseases (CVD) (Challenge Principles and Practice of Knowledge Discovery from Databases, 2001) is used to visually compare the impact of tobacco and alcohol ingestion (see figures 1 and 2).

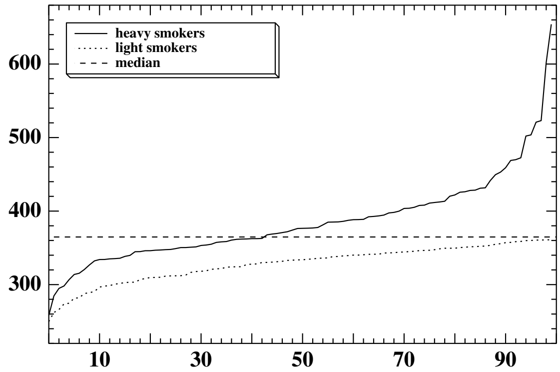

Figure 1. The lower curve shows the risk for the 100 non-smokers in the database. It is always lower than the median risk (horizontal line). The upper curve depicts the risk associated with the 100 heaviest smokers in the database. Although 40% of them have a CVD risk lower than the median risk, 10% of them present a huge risk.

6.1.4. Methodological aspects, Meta-learning and Competence Maps

The Meta-Learning problem (a priori finding the algorithm most suited to a given dataset and learning task) is a bottleneck issue for machine learning algorithms, hampering their widespread use outside research labs, in industry and services. For this purpose, a methodology inspired by the Constraint Satisfaction and Phase Transition paradigm (see section 6.1.5 below) was developed, based on the definition of order parameters problem. In this paradigm, assessing an algorithm amounts to drawing its "Competence Map", picturing its average performance in the landscape defined by the order parameters.

The Competence Map of the best-known (propositional) learning algorithm, C4.5 (decision tree inducer) has been drawn using an extensive experimentation protocol. This led to characterize the regions of the problem space where this well-known algorithm is inefficient on average. Further, this map helped to provide a unified and precise perspective on long known empirical observations, e.g., regarding the small disjunct problem in learning [14][13].

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

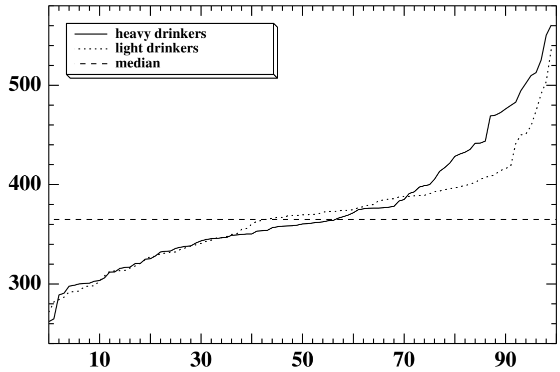

Figure 2. The lower curve shows the risk for the 100 heaviest and lightest drinkers in the database. These curves, unexpectedly showing little or no impact of alcohol ingestion (see figure 1), were later explained by the fact that the studied population was found to involve no light drinkers...

sciences is a useful tool to determine the regions in the problem space where an algorithm generally fails or succeeds.

6.1.5. Inductive Logic Programming, Constraint Satisfaction and Phase Transition

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. The major difficulties of ILP can be illustrated on an application such as the Predictive Toxicology Evaluation (IJCAI Challenge, looking for patterns discriminating carcinogenic molecules in nitro-aromatic compounds). The cornerstone of Machine Learning is the covering test, checking whether a given hypothesis (a chemical pattern) appears in the training examples (molecules). In first-order logic, this problem is NP-complete.

ILP thus faces an additional difficulty compared to propositional machine learning, namely the fact that the hypothesis assessment is intractable in terms of worst-case complexity analysis. Therefore, relational data mining is seriously challenged regarding its scalability. Our research on phase transition in Inductive Logic Programming aims at a better understanding of the obstacles to large-scale relational learning and at devising possible solutions.

Bounded Relational Reasoning

We devised an approach addressing the relational covering test, or theta-subsumption, based on stochastic reasoning, sampling the subsumption lattice. This work pioneered the field of any-time algorithms for relational induction and deduction and was successfully applied to predictive toxicology evaluation. Furthermore, it provided hints into a principled junction between first order logic and probabilistic frameworks, needed to achieve efficient knowledge discovery in highly relational and noisy domains, e.g., Web mining or ontology construction.

Phase Transitions

The formalization of the subsumption lattice led us to investigate more deeply the connections between ILP and Constraint Satisfaction Problems (CSP). Independently, L. Saitta and A. Giordana (Università del Piemonte) formalized this connection, and imported into ILP the phase transition complexity paradigm developed for CSP. In this paradigm, the performance of the algorithm (predictive accuracy, running time) is viewed as random variables depending on the order parameters of the problem instances, e.g., the density and tightness of constraints.

The practical implications on ILP of the phase transition paradigm were investigated in collaboration, showing the existence of a "failure region" where the up-to-date algorithms behave no better than random guessing. This negative result is acknowledged as significant in the literature, and the explanations offered for this failure region shed unexpected light on the actual limitations of ILP.

Junction with Constraint Satisfaction

A new theta-subsumption algorithm was devised by Jérôme Maloberti, based on an ad hoc rewriting of theta-subsumption problems into CSPs, and calling upon efficient CSP procedures. This algorithm was shown to outperform the state of the art by several orders of magnitude. Finally, a new relational learning mechanism was proposed to address the failure region mentioned earlier. Formally, this bottom-up approach exploits the constraints that originated from training examples, confronted to that of the order parameters, e.g., total number of variables, of the sought hypothesis (to be guessed by the expert). A stochastic complexity analysis, confirmed by empirical validation, shows that the approach actually addresses part of the failure region observed for greedy (top-down) algorithms.

The standard covering test in ILP (theta-subsumption) shows a phase transition, that is the percentage of clauses covering an example abruptly drops from almost 1 to almost 0 as the specificity of the clause (number n of variables, number m of predicates) and of the example (number L of constants, number N of literals built on each predicate symbol) increases. The figure below depicts the percentage of successful theta-subsumption test for n=10 and NN=100, in plane (m, L), averaged over 100 pairs (clause,example) uniformly drawn for each (m, L) point.

The complexity of the former best known theta-subsumption algorithm (Scheffer, 1994) is depicted in figure 3-left. The complexity of the Django algorithm [22][8] shows an improvement of several orders of magnitude. Note in particular in figure 3-right that the complexity is negligible in the region to the right, contrasting with Scheffer (1994).

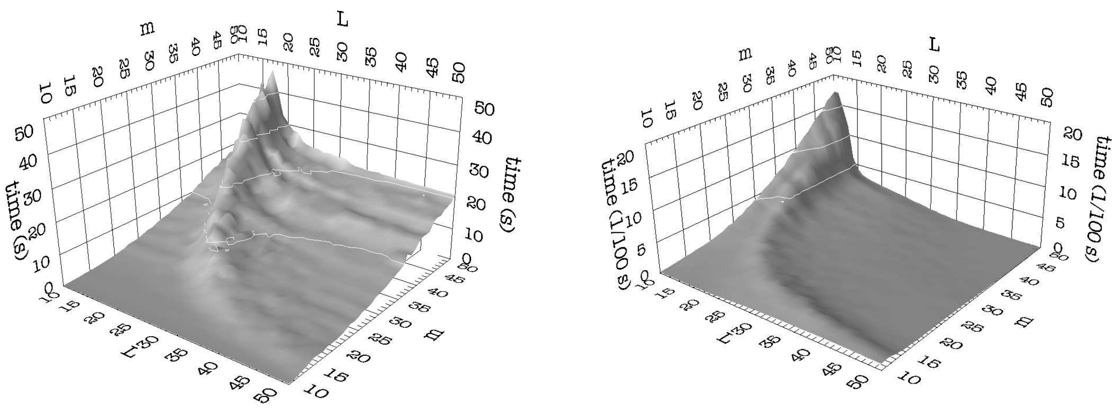

Figure 3. Comparative complexities of (Scheffer, 1994) (left) and the Django algorithm (right).

Software: The Django algorithm is available; it is used at University of Bari, Tufts University and Yokohama University.

6.2. Fundamentals of Evolutionary Computation

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

Participants: Kamal Abboud, Anne Auger, Olga Rudenko, Marc Schoenauer, Michèle Sebag, Olivier Teytaud.

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-stings, 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. Self-adaptive operators in continuous search spaces

A first research direction considers the self-adaptive operators embedded in evolution strategies, which are the most effective EC algorithms for optimization in R^n.

In continuous search spaces, genetic individuals are mutated by adding a Gaussian perturbation. It is clear that the most critical issue is the tuning of the Gaussian law, most generally its covariance matrix. A flexible mechanism was proposed, using a quadratic approximation of the target function and constructing the covariance matrix from the Hessian of this estimate in "sufficiently" regular regions, and switching to the standard Covariance Matrix Adaptation otherwise [11].

6.2.2. 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 [7] 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 properties of the Markov chains investigated.

Note that both the above subsections, together with a paper published in GECCO'2003, are the fundamental part of Anne Auger's PhD thesis [4].

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)[1].

In her PhD thesis, Olga Rudenko [6] studied several aspects of evolutionary multi-objective optimization, some of them presented in conferences. First, she designed a sound stopping criterion (the only available criterion was ...a given number of generations) [28]. Second, she proposed a new crossover operator based on teh dominance property. She demonstrated its efficiency in terms of the speed of convergence toward the Pareto front – the quality of the solution being the same than with the standard crossover [29].

Last, she studied two problems of automobile industry in collaboration with FIAT industry in Milan, in the framework of the European Project *Ingenet*. She obtained far better results than the previous best design for a car-front-end part, and she successfully tackled the very difficult problem of engine running points optimization, with the heavy constraints of European regulations on gaz emissions[2].

[1] K. Deb, *Multi-Objective Optimization Using Evolutionary Algorithms*, John Wiley, 2000.
[2] O. Rudenko, T. Bosio, R. Fontana, and M. Schoenauer, Optimization of car front crash members, in selected papers from Evolution Artificielle'01, pp 202-214, Springer Verlag, LNCS 2310, 2002.

The on-going collaboration with EADS-LV company was continued. It deals with the design of Launchers using evolutionary algorithms. Whereas previous results had shown the feasibility of evolutionary optimization for the global design of a launcher considered as a single-objective problem [16], the 6 months project by a student from *Ecole Centrale* investigated this year the multi-objective problem. A single launcher was optimized with two operational objectives, corresponding to two missions with two different orbits for the satellites that the launcher is carrying.

6.2.4. Surrogate models in Evolution Strategies

It is well-known that Evolutionary Algorithms require a large number of evaluation of the fitness function – the price to pay for their success on problems that are intractable for other methods. But on the other hand, those many individuals evaluated during the course of evolution are as many examples of the fitness function, and can be used to learn an (inexpensive) approximation of this (often computationally heavy) fitness function (note that another use of such information is the computation of an approximation of the Hessian matrix as described section 6.2.1).

In his PhD thesis [3], Kamal Abboud used Support Vector Machines to compute a local approximation of the fitness around the best-so-far individual. This approximation is then deterministically minimized (using the classical BFGS method), and the minimum, if better than the previous minimum, is introduced in the population. Experiments demonstrate the efficiency of the *Surrogate Deterministic Mutation*.

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)[3], 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.

Along the same lines, a project in the DEA IARFA (University Paris 6) investigated the efficiency of the *LEM* method proposed by Michalski[4], and its connections to Estimations of Distribution Algorithms. While Michlaski used his own AQ15 program (NOT open-source, but the binary executable program is freely distributed), the *Version Space* approach[5] was used. First, it was impossible to reproduce Michalski's published results using his own code ; second, the performances of Michalski's LEM were outrageously degraded when the fitness function was translated so the minimum is not 0 ; last, the Version Space based approach, though providing better results than the original LEM, was far too costly compared to the state-of-the-art Evolution Strategies. See the stage report [35] for further details.

6.3. Robotics

Participants: Nicolas Bredèche, Antoine Cornuéjols, Mary Felkin, Sylvain Gelly, Nicolas Godzik, Jérémie Mary, Marc Schoenauer, Michèle Sebag.

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

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

[4] Learnable Evolution Model : Evolutionary Processes guided by Machine Learning, *Machine Learning* 38, pp 9–40, 2000.

[5] M. Sebag, Constructive Induction: A Version Space based Approach, in IJCAI'99, pp 708–713, Morgan Kaufmann

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. Symbolic controllers

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[6]. It involves continuous inputs and discrete outputs. 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.

6.3.2. 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), we proposed an architecture combining three functionalities [18] (see figure 4): 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.

6.3.3. 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 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 [21] demonstrated the efficiency of the proposed approach for fuzzy system design.

6.3.4. Cartography for autonomous robots

Cartography is another very active domain of research in autonomous robotics – sometimes considered as one of the most important for actual autonomy. Many issues are yet unsolved, even though there exist today a few robust approaches to automatically draw maps of static structures small environments.

Building on his *Stage de DEA* [34], the goal of Sylvain Gelly during his PhD (that started in September 2004) is to design a system with the following properties – that is viewed as crucial for actual autonomous robotics: automatically draw a map of its usual environment that allows it to make accurate predictions when moving; to adapt such an existing map to unknown (but more or less similar) environments; to take advantage of possible existing a priori knowledge (as given by some human expert for instance).

[6] 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

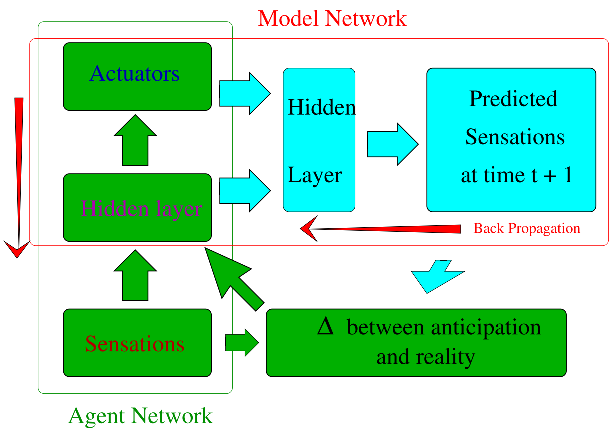

Figure 4. The Action-Anticipation-Adaptation architecture (A3) is composed of three interacting modules, currently implemented as Elman neural nets. The decision making module ("agent network" on the left) computes the values of the actuators from the sensors values. The anticipation module (on the upper right) predicts the sensations at time t+1 from the sensations and actions at time t. The differences between the actual and the anticipated sensations provides the robot with a (very basic) self-awareness, and allows it to modify its weights using the back-propagation mechanism.

Remark: Several other *Stages de DEA* 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.

6.4. Inverse problems

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

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.

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 in the past regarding the Topological Optimum Design of Mechanical Structures using the so-called Voronoi representation for structures, that overcomes many limitations of the more widely-used bitarray representation. Some results of this work have been exposed in hte *Innovative Design Techniques* section of the *ArchiLab* exhibition, an architectural exhibition in Orléans.

Mohamed Jebalia has just started his PhD under the supervision of Marc Schoenauer to investigate other representations for the same problem, trying to implement some sort of modularity in the representation. A good example is that of a crane structure, where the same elements are duplicated to build the complete crane - whereas an evolutionary algorithm must "discover" that structure again and again. During his *stage de DEA* he has used Genetic Programming for a simple inverse problem (find a given target shape) with, as one usually says, promising results.

6.4.2. Geologically sound representation for seismic inversion

Vijay Pratap Singh's PhD thesis, 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 - nad 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 teh seismic data. Such representation at least ensures the geophysical relevance of the identified underground.

6.4.3. Representation by virtual examples

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. The inverse problem is solved by Evolution Strategies coupled with standard deterministic gradient-based methods. An innovative representation aims at evolving a set of virtual examples. The isotherm is then 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. Moreover, using SVM should scale up for a large number of components. Note that a *stage de DEA* was proposed on that topic, and chosen by a student – but unfortunately, that student resigned rapidly due to some family problems.

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.

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).

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).

6.5. Other applications

Participants: Jérôme Azé, Yves Kodratoff, Samuel Landau, Mathieu Roche, Michèle Sebag, Yann Semet.

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.
Our studies [26][25][27][24] focus on the induction of concepts from texts and examples provided by a domain expert. Our approach is rooted in systemic functional linguistics (Halliday 1978, 1994), and relies on two postulates: The importance of the field expert, and the idea that text mining is a chain of linguistic processes (e.g., lemmatisation, tagging, terminology, concept recognition, and information extraction) that must be optimized together.

We are presently working with five technical corpora: a "biology" corpus extracted from the National Library of Medicine (NLM) Medline database (6119 abstracts, 10Mb); a "Data Mining" corpus (369 Kb), the introductions of papers published at KDD conferences; a "Human Resources" corpus (3784 Kb) belonging to the company PerformanSe (in French). a "Resume" corpus (2470 Kb) that contains 1144 Resumes from the company VediorBis (in French). a "newspapers" corpus (40 Mb), part of the TREC competition.

Our basic hypothesis calls upon a theorization of the language that is minimal but sufficient to both characterize concepts of specialty and spot their occurrences in the documents. For example, in corpora related to Human Resources, the verb "to give up", followed by the noun "influence" evokes the presence of a concept that the psychologist recognizes as describing the nature of relations in the company and that he named "relational". In the same way, the verb "to give up" followed by the noun "task" is said to evoke a kind of behavior called "implication" (in this case, it is evidently the lack of implication). Along the same lines, the noun "action", followed by the adjective "together" evokes the concept "relational", while the adjective "adequate" followed by the noun "action", evokes the concept "activity".

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 on a simplified instance of the problem on a small network [36] have shown that in fact an iterated hill-climber can solve the problem better than any other algorithm (including CPlex and complex EAs). On-going work addresses the complete problem on large networks.

6.5.3. Time-dependent planning with bounded resources

On-going work in collaboration with Thales - Land & Vision division is concerned with temporal planning with limited resources. Two approaches will be compared: 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.

7. Contracts and Grants with Industry

Keywords: *Airbus*, *Chromalgena*, *EADS*, *IFP*, *Quantum chemistry*, *SNCF*.

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** – 2003-2005 (18 kEur), side-contract to Vijay Pratap SINGH's CIFRE Ph.D. (section 6.4.2).
- **EADS** – April-November 2004 (5 kEur) : side-contract to Stephane D'ESCRIVAIN internship (section 6.2.3).
- **SNCF** – 2004 (50 kEur) : research contract, Yann SEMET, expert engineer (section 6.5.2).
- **Thalès** – 2004-2005 (50 kEur) : research contract (section 6.5.3).
- **Robea** CNRS Program – 2002-2005 (56 kEur at Rocquencourt), coordinators Michèle SEBAG and Marc SCHOENAUER (section 6.3).

Contracts managed by Paris Sud University

- **Neurodyne** ACI-NIM (New Interfaces of mathematics) – 2003-2006 (6 kEur).
- **Traffic** ACI-NIM (New Interfaces of mathematics) – 2004-2007 (17.5 kEur). The partners are Laboratoire de Mathématiques (Paris-Sud), INRETS, Laboratoire Imagerie cérébrale et Neurosciences, LENA (UPR 640).
- **Télémédecine** ACI technologie de la médecine – 2002-2005 (15245 kEur). Coordinator Marie-Christine JAULENT, hopital Broussais.
- **Indana**, ACI NIM – 2002-2005 (15 000 kEur). Coordinator Marie-Christine JAULENT, hopital Broussais.

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.

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

- PASCAL, Network of Excellence, 6th Framework Program: Michèle Sebag, corresponding member for Université Paris-Sud since 2003.
- Evonet, European Network on Evolutionary Computation: Marc Schoenauer was member of the Executive Committee and Chair of the Electronic Communications Committee during the whole funding period (1997-2003), and is now responsible for the Web Site *EvoWeb* after the end of the funding period. Michèle Sebag was a member (1999-2003).

8.2.3. Collaborations with joint publications

- Università del Piemonte Orientale, Italie.
- Vrije Universiteit Amsterdam, The Netherlands ([19],[20]).
- University of Leuven, Belgium.

8.3. National actions

8.3.1. Organization of conferences and scientific events

- JET, Journées Évolutionnaires Trimestrielles: organization since1998.
- 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 very first edition in 1994.

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.

8.3.3. Associations

- *AFIA, Association Française d'Intelligence Artificielle* : Marc Schoenauer, member of Executive since 1998, president since 2002 ; Michèle Sebag, member of Executive since 2000, treasurer in 2004 ; Jérémie Mary, vice-treasurer since 2004.
- *Institut de la complexité* : Michèle Sebag, member of the Controversy Seminar since 2003.

8.3.4. Collaborations with joint publications

- IASI group, LRI ([33]).
- Situated Perception group (Perception Située), LIMSI-CNRS.

8.4. Honors

8.4.1. Prizes and awards

- Anne Auger, Marc Schoenauer, Best Student Paper Award [11], PPSN (Birmingham),2004.
- Marc Schoenauer, Senior Fellows of the International Society for Genetic and Evolutionary Computation, awarded at GECCO 2004.
- Yann Semet, best paper award of GECCO'04 for "An Interactive Artificial Ant Approach to Non-Photorealistic Rendering" (in collaboration with Una-May O'Reilly, Fredo Durand).

8.4.2. Keynote addresses

- Evolvable Hardware 2004, Marc Schoenauer, invited plenary speaker
- International Workshop on Nature Inspired Computation and Applications, Marc Schoenauer, invited plenary speaker
- Kick-off meeting of the dutch project *Evolutionary Algorithms for Physics*, AMOLF (FOM-Institute for Atomic and Molecular Physics).

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 IEEE Transactions on Evolutionary Computation (1996-2004), 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 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), and of Revue d'Intelligence Artificielle (since 2002).
- Michèle Sebag was Guest Editor of Revue des Nouvelles Technologies de l'Information in 2004.

9.1.2. Chair in Organizing Committees

- Michèle Sebag was Editor of RFIA, 14ème Congrès Reconnaissance des Formes et Intelligence Artificielle, 2004.

9.1.3. Program Committee Member (international events)

- Nicolas Bredèche: ICIIP, International Conference on Intelligent Information Processing).
- 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, Frontiers on Evolutionary Algorithms, ...
- Michèle Sebag: ECAI, 16th European Conference on Artificial Intelligence, ICDM, IEEE International Conference on Data Mining, ICML, International Conference on Machine Learning, ILP, Inductive Logic Programming, PPSN, Parallel Problem Solving from Nature, EuroGP, European Conference on Genetic Programming, GECCO, Genetic and Evolutionary Computation Conference, ECML-PKDD, European Conference on Machine Learning, Conference on Principle and Practice of Knowledge Discovery from Databases, CEC, IEEE Congress on Evolutionary Computation, ...

9.1.4. Program Committee Member (national events)

- RFIA, Reconnaissance des Formes et Intelligence Artificielle: Michèle Sebag, member since 1996.
- EGC, Extraction et Gestion des Connaissances: M. Sebag since 2002.
- CAP, Conférence d'apprentissage: Michèle Sebag since 1999; Antoine Cornuéjols, since 2001.
- EA, Evolution Artificielle: Michèle Sebag since 1994.

9.1.5. Evaluation committees and invited expertise

- Marc Schoenauer, reviewer of 3 European projects (STREPS) from FP5 framework in the Neuro-IT program: SIGNAL, HYDRA and POETIC.
- Michèle Sebag, reviewer of the National research program ACI Nouvelles Interfaces des Mathématiques.
- Michèle Sebag, Expert for teaching programs LMD (Licence-Master-Doctorat).

9.1.6. Other evaluation activities

- Reviewer for PhD dissertation: Marc Schoenauer (5) ; Michèle Sebag (4) ; Antoine Cornuéjols (1).
- Reviewer for Habilitation: Marc Schoenauer (1) ; Michèle Sebag (2)

9.1.7. Popularisation of research results

- *ArchiLab exhibition*, Orléans, 15 October – 31 December 2004: Marc Schoenauer in collaboration with the architect consortium ECZT (see section 6.4.1).

9.1.8. Summer schools, tutorials, invited seminars

- Invited seminars, Michèle Sebag: Université Lille 2004; Laboratoire de Physique Théorique et Modèles Statistiques, Université Paris-Sud, 2004.

9.2. Enseignement

9.2.1. Defended doctorates

- ROUDENKO Olga 05/03/04 Bourse Ecole Polytechnique
- ABBOUD Kamal 20/10/04 ATER, Université Evry
- ROCHE Mathieu 13/12/04 ATER, Université Paris Sud
- AUGER Anne 20/12/04 MENRT, Université Pierre et Marie Curie

9.2.2. Graduate courses

- DEA I3, Data mining and machine learning : Michèle Sebag, Antoine Cornuéjols.
- DEA ECD, Optimization for data mining: M. Sebag.
- DEA Analyse Numérique, Université Sant-Joseph, Beyrouth: M. Schoenauer.

9.2.3. Other research-related teaching activities

- *Ecole Polytechnique*, Projects in Evolutionary Robotics in the *Modex d'Electronique*: Marc Schoenauer.
- *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.

10. Bibliography

Books and Monographs

[1] H. BRIAND, M. SEBAG, R. GRAS, F. GUILLET (editors). *Numéro spécial "Mesures de qualité pour la fouille de données"*, vol. 1, 2004.

[2] C. ROUVEIROL, M. SEBAG (editors). *Machine Learning Journal, Special Issue on Inductive Logic Programming and Relational Learning*, vol. 55, n^o 2, Kluwer, 2004.

Doctoral dissertations and Habilitation theses

[3] K. ABBOUD. *Algorithmes évolutionnaires assistés par méthodes d'approximation*, in French, Ph. D. Thesis, Ecole Polytechnique, 2004.

[4] A. AUGER. *Contributions théoriques et numériques à l'optimisation continue par algorithmes évolutionnaires*, in French, Ph. D. Thesis, Université Paris 6, December 2004.

[5] M. ROCHE. *Intégration de la construction de la terminologie de domaines spécialisés dans un processus global de fouille de textes*, in French, Ph. D. Thesis, Université Paris 11, 2004.

[6] O. RUDENKO. *Application des algorithmes évolutionnaires aux problèmes d'optimisation multi-objectif avec contraintes*, in French, Ph. D. Thesis, École Polytechnique, March 2004.

Articles in referred journals and book chapters

[7] A. AUGER. *Convergence results for (1,λ)-SA-ES using the theory of ϕ-irreducible Markov chains*, in "Theoretical Computer Science", in press, 2004.

[8] J. MALOBERTI, M. SEBAG. *Fast Theta-Subsumption with Constraint Satisfaction Algorithms*, in "Machine Learning Journal", vol. 55, 2004, p. 137-174.

[9] J. MARY, G. MERCIER, J.-P. COMET, A. CORNUÉJOLS, C. FROIDEVAUX, M. DUTREIX. *Utilisation d'une méthode d'estimation d'attributs pour l'analyse du transcriptome de cellules de levures exposées à de faibles doses de radiation*, in "Informatique pour l'analyse du transcriptome", J.-F. BOULICAUT, O. GANDRILLON (editors)., Hermès, 2004, p. 189-205.

[10] G. MERCIER, N. BERTHAULT, J. MARY, A. ANTONIADIS, J.-P. COMET, A. CORNUÉJOLS, C. FROIDEVAUX, M. DUTREIX. *Biological detection of low radiation by combining results of two analysis methods*, in "Nucleic Acids Research (NAR)", vol. 32, n^{o} 1, 2004, p. 1-8.

Publications in Conferences and Workshops

[11] A. AUGER, N. VANHAECKE, M. SCHOENAUER. *LS-CMA-ES: Better Covariance Matrix Adaptation using Least Square Approximation*, in "8th International Conference on Parallel Problems Solving from Nature (PPSN 2004)", LNCS, Springer Verlag, 2004, p. 182-191.

[12] J. AZÉ, N. LUCAS, M. SEBAG. *A Genetic ROC-based Classifier*, in "Physiological Data Modeling Contest, ICML 2004, Canada", July 2004.

[13] N. BASKIOTIS, M. SEBAG. *C4.5 Competence Map: a Phase Transition-inspired Approach*, in "21st International Conference on Machine Learning, ICML 2004", R. GREINER, D. SCHUURMANS (editors)., Morgan Kaufmann, 2004, p. 73-80.

[14] N. BASKIOTIS, M. SEBAG. *Carte de competence de C4.5 : une approche inspiree de la transition de phase*, in "Actes de la Conference d'Apprentissage, CAp04", 2004.

[15] A. CORNUÉJOLS, M. SEBAG, J. MARY. *Classification d'images à l'aide d'un codage par motifs fréquents*, in "Workshop sur la fouille d'images (RFIA-04), Toulouse (France)", 2004.

[16] N. DURANTÉ, A. DUFOUR, V. PAIN, G. BAUDRILLARD, M. SCHOENAUER. *Multidisciplinary Analysis and Optimisation Approach for the Design of Expendable Launchers*, in "Proc. of MDAO'04", AAAI, 2004.

[17] M. DUTREIX, J.-P. COMET, A. CORNUÉJOLS, C. FROIDEVAUX. *Determination of cellular drug targets: searching for functional information in the jungle of microarrays data*, in "Current Trends in Drug Discovery Research (CTDDR-04), India", 2004.

[18] N. GODZIK, M. SCHOENAUER, M. SEBAG. *Robustness in the long run: Auto-teaching vs Anticipation in Evolutionary Robotics*, in "8th International Conference on Parallel Problem Solving from Nature (PPSN 2004)", X. YAO ET AL. (editor)., Springer Verlag, 2004, p. 932-941.

[19] K. JONG, E. MARCHIORI, M. SEBAG. *Ensemble Learning with Evolutionary Computation: Application to Feature Ranking*, in "8th International Conference on Parallel Problem Solving from Nature (PPSN 2004)", X. YAO ET AL. (editor)., Lecture Notes in Computer Science, Springer Verlag, 2004, p. 1133-1142.

[20] K. JONG, J. MARY, A. CORNUÉJOLS, E. MARCHIORI, M. SEBAG. *Ensemble feature ranking*, in "8th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD-04), Pisa, Italy", Springer Verlag, 2004, p. 267-278.

[21] C. KAVKA, M. SCHOENAUER. *Evolution of Voronoi-based Fuzzy Controllers*, in "8th International Conference on Parallel Problem Solving From Nature (PPSN 2004)", Springer Verlag, 2004, p. 541-550.

[22] J. MALOBERTI, E. SUZUKI. *An Efficient Algorithm for Reducing Clauses Based on Constraint Satisfaction Techniques*, in "14th International Conference on Inductive Logic Programming (ILP 2004)", Lecture Notes in Artificial Intelligence, Springer Verlag, 2004.

[23] M. ROCHE, J. AZÉ, Y. KODRATOFF, M. SEBAG. *Learning Interestingness Measures in Terminology Extraction. A ROC-based approach*, in "Workshop ROC Analysis in AI, 16th European Conference on Artificial Intelligence (ECAI 2004)", 21 august 2004 2004, p. 81-88.

[24] M. ROCHE, J. AZÉ, O. MATTE-TAILLIEZ, Y. KODRATOFF. *Mining texts by association rules discovery in a technical corpus*, in "Conference on Intelligent Information Processing and Web Mining (IIPWM 2004)", Advances in Soft Computing, Springer Verlag, 2004, p. 89-98.

[25] M. ROCHE, T. HEITZ, O. MATTE-TAILLIEZ, Y. KODRATOFF. EXIT*: Un système itératif pour l'extraction de la terminologie du domaine à partir de corpus spécialisés*, in "International Conference on Statistical Analysis of Textual Data (JADT 2004)", vol. 2, 2004, p. 946-956.

[26] M. ROCHE, O. MATTE-TAILLIEZ, Y. KODRATOFF. *Etude de Mesures de Qualité pour Classer les Termes Extraits de Corpus Spécialisés*, in "Actes de INFORSID'04", 2004, p. 371–386.

[27] M. ROCHE. *Utilisation de mesures pour extraire la terminologie de corpus spécialisés de domaines différents*, in "Actes de l'atelier "Fouille de Textes" à la conférence EGC'04", 2004.

[28] O. RUDENKO, M. SCHOENAUER. *A Steady Performance Stopping Criterion for Pareto-based Evolutionary Algorithms*, in "6th International Multi-Objective Programming and Goal Programming Conference", 2004.

[29] O. RUDENKO, M. SCHOENAUER. *Dominance Based Crossover Operator for Evolutionary Multi-objective Algorithms*, in "8th International Conference on Parallel Problem Solving from Nature (PPSN 2004)", Springer Verlag, 2004, p. 812-821.

[30] M. SEBAG, J. AZÉ, N. LUCAS. *ROC-based Evolutionary Learning: Application to Medical Data Mining*, in "Artificial Evolution'03", Lecture Notes in Computer Science, Springer Verlag, 2004, p. 384-396.

[31] Y. SEMET, S. GELLY, M. SCHOENAUER, M. SEBAG. *Artificial Agents and Speculative Bubbles*, in "Computational Finance and its Applications, Southampton, UK", WESSEX INSTITUTE OF TECHNOLOGY (editor)., WIT Press, 2004, p. 35–44.

[32] Y. SEMET, U.-M. O'REILLY, F. DURAND. *An Interactive Artificial Ant Approach to Non-PhotoRealistic Rendering*, in "Proceedings of the Genetic and Evolutionary Computation COnference (GECCO'04), Seattle, WA, USA", KALYANMOY DEB ET AL. (editor)., LNCS 3102, vol. 1, Springer Verlag, June 26-30 2004, p. 188–200.

[33] A. TERMIER, M.-C. ROUSSET, M. SEBAG. *Dryade: a new approach for discovering closed frequent trees in heterogeneous tree databases*, in "IEEE International Conference on Data Mining, ICDM04", 2004.

Internal Reports

[34] S. GELLY. *Rapport de stage de DEA : Cartographie en robotique autonome*, Technical report, Université Paris 6, July 2004.

[35] T. MAHMOOD. *Rapport de stage de DEA : Experiments with Learnable Evolution Models*, Technical report, DEA IARFA - U. Paris 6, 2004.

[36] Y. SEMET, M. SCHOENAUER. *Optimisation évolutionnaire pour la régulation des circulations*, Technical report, Collaboration SNCF/INRIA, Juin 2004.

Miscellaneous

[37] H. HAMDA, M. SCHOENAUER. *Les algorithmes évolutionnaires pour la conception originale de chaises*, Octobre-décembre 2004, Exposition ArchiLab, Orlèans.