Proposal for a common research project INRIA-LRI as part of UR Futurs

TAO - Thèmes Apprentissage et Optimisation
Machine Learning, Data Mining and Evolutionary Optimization

Marc Schoenauer[1] **et Michèle Sebag**[2]

[1]INRIA, DR2 [2]CNRS, DR2

1 Preamble

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[1].

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.

2 The Team

- Marc Schoenauer, Directeur de Recherche INRIA,
- Michèle Sebag, Directeur de Recherche CNRS,

[1]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.

- Antoine Cornuejols, Maître de Conférence à l'Université d'Évry,
- Nicolas Bredèche, Maître de Conférence à l'Université de Paris-Sud Orsay
- Anne Auger, PhD student
- Nicolas Baskiotis, PhD student
- Nicolas Godzik, PhD student
- Jérémy Marie, PhD student
- Mary Felkin, PhD student
- Mathieu Pieres, PhD student, CIFRE Airbus
- Vijay Pratap Sing, PhD student, CIFRE IFP

The TAO team merges part of the *Inférence et Apprentissage* team of the Laboratoire de Recherche en Informatique, CNRS UMR 8623, Université Paris-Sud, and part of the former *Fractales* project at INRIA Rocquencourt.

The *Inférence et Apprentissage* team of LRI has an international reputation in Machine Learning (European projects Machine Learning Toolbox ; Inductive Logic Programming, I et II ; member of the Networks of Excellence MLNet, ILPNet, and KDNet). Michèle Sebag, head of the *I & A* team since 2001, is internationally known in that field: associate editor of Machine Learning Journal (Kluwer), co-president of the Inductive Logic Programming conference 2001, vice-chair of the IEEE conference on Data Mining (ICDM03), member of the Program Committees of the main conferences of the field since 1997, member of the *PASCAL* project, she has also been active at the national level, as founder of the National conference CAP (since 1999) and of the *Collège Apprentissage, Fouille et Extraction (CAFE)*, and co-leader of the CNRS "action spécifique" *Fouille de données et Extraction (2000-2002).* But she is also known in the area of Evolutionary Computation, being associate editor of the journals IEEE Transactions on Evolutionary Computation, and Genetic Programming and Evolvable Hardware (Kluwer).
Antoine Cornuejols is co-author of the latest French book on Machine Learning (2002), currently considered and cited as *the* French reference.

Marc Schoenauer is one of the leading persons world-wide in the area of Evolutionary Computation: he is Editor in Chief of *Evolutionary Computation* – MIT Press, associate editor of IEEE Transactions on Evolutionary Computation, Genetic Programming and Evolvable Hardware (Kluwer), Applied Soft Computing (Elsevier), Natural Computing (Springer Verlag). He has been involved in the organization of many international conference, as well as founder and organizer of the series of French conferences *Evolution Artificielle* and the French series of workshops JET (*Journées Évolutionnaires Trimestrielles*). He is member of the Executive Committee of *Evonet*, the Network of Excellence on Evolutionary Computing (FP5), and of the Executive Board of the *International Society of Genetic and Evolutionary Computation – ISGEC.*

The joint efforts of those researchers within their respective domain – Machine Learning, Data Mining and Evolutionary Computation – should permit significant advances in all three fields, tackling some common bottlenecks from several perspectives. The resulting INRIA/LRI project aims at becoming a world-wide center of excellence at the interface between Machine Learning and Optimization.

Several researchers have joined the TAO team for long stays in 2003-2004: Shengwu Xiong, associate professor at Wuhan University, China, and Aurora Trinidad Ramirez Pozo, associate professor at the Universidad Federale do Parana, Brésil (together with her PhD student Celso Ishida) are here for one year. Hatem Hamda, assistant professor at Tunis University (with his graduate student Mohamed Jebalia) is here for 4 months. Nico di Mauro and Mara Basile, PhDs of University of Bari, Italy, are here for a 3-months stage.

3 State(s) of the art

This section briefly describes the current bottlenecks and hot topics in ML, DM and EC.

3.1 Machine Learning and Data Mining

The field of Machine Learning [83] has been deeply modified in the last decade, due to two major scientific breakthroughs on one hand, and the applicative challenges of Data Mining [49, 50] on the other hand.

3.1.1 Recent advances in Machine Learning

A first breakthrough is that of the Statistical Learning Theory (SLT) [134, 135, 26]. In this framework, the stress is put on the generalization error, measuring how well an estimate extracted from a sample of data points, referred to as *hypothesis*, would fit all data points. The estimate is viewed as a random variable, depending on the data sample; the study then focuses on the deviations of the generalization error wrt the empirical error measured from the available sample.

Along those lines, SLT provides a rigorous framework for bounding with sufficiently high probability the deviation between the generalization error and the empirical error. The learning goal can thus be formulated as i) minimize the empirical error; ii) minimize the deviation bound.

This reformulation, together with a change of representation (the "kernel trick") leads to quadratic optimization problems: this is the basis of Support Vector Machine (SVMs) algorithms [108, 58, 109], both elegant and efficient.

As could have been expected [137], SVM algorithms are limited in some respects. Firstly, the choice of the kernel is problem dependent, and a significant part of the SVM research is devoted to crafting new kernels for specific problem domains [55, 39]. Secondly, SVMs turn a multi-objective problem (minimize both the empirical error and the deviation bound, linked to the richness of the search space and its Vapnik-Cervonenkis dimension) into a single optimization problem, defined as the weighted sum of the two objectives, with an expensive optimal weight setting [136]. Lastly, quadratic optimization problems in high dimensional spaces happen to suffer from numerical instabilities [113] and the quadratic complexity is not affordable when one comes to really huge data (an acceptable complexity would rather be $n\,log(n)$).

A second breakthrough in ML and DM is that of Ensemble Learning [107], originally rooted in the PAC Learning theory [133, 61]. Initially, the question was to compare strong learnability (being able to learn any concept in a search space, with an arbitrarily low generalization error) and weak learnability (the error is no greater than $1/2 - \epsilon$). Using ensemble learning, Schapire could prove that strong and weak learnability are equivalent [106], i.e. the vote of (a polynomial number of) weak learners can behave as a strong learner. More generally, the combination of diversified hypotheses [16, 77] improves on the best single hypothesis. An explanation for this good behavior is based on the decomposition of the error into error bias (usually reduced as ensemble learning explores the convex hull of the hypothesis search space) and error variance (reduced because hypotheses are independent in bagging [16], or anti-correlated in boosting [38]).

Again, ensemble learning suffer from some limitations. On one hand, the theoretical analysis of boosting in terms of margin [124] does not explain convincingly its good empirical behavior (i.e. directly optimizing the margin shows less efficient [16]). Empirical studies demonstrate that the criteria used in ensemble learning are not optimal [77] and offer room for improvement [29]. Last, though boosting is often empirically more efficient than bagging, it is hindered by the data noise [105].

3.1.2 Data Mining is not Machine Learning

Another main cause for the rapid evolution of ML is the challenges offered by Data Mining, concerned with the exploitation of massive real-world data for finding "valid, new and useful" regularities [33, 49, 50].

While significant advances have been made for specific DM tasks, e.g. identifying frequent item sets [56, 139, 17, 128], the DM process suffers from three bottlenecks.

Firstly, it is widely acknowledged that 80% of the DM effort is spent in the pre-processing step (cleaning, selecting, merging, etc, the datasets). This tedious phase involves a number of elementary redescription operations, the choice and setting of which are crucial to the quality of the process output[2]. More generally, Data Mining fully faces the problem of data representation, which can be viewed as an inverse problem: a good representation is one permitting a successful learning/mining. Actually, the feature selection task, which is but one part of the representation issue, is a NP-hard combinatorial problem [46].

Secondly, an essential part of the DM process is to allow the user to inspect the results, and to interact with the system in order to focus the search and find "useful" results. The literature is most often concerned with focusing the DM search through syntactical and logical constraints [74, 95], or defining interest measures [90]. Still, finding constraints with an adequate granularity (i.e. leading to readable and yet non-trivial solutions), or finding a measure of interest which would precisely rank first those regularities the user would find most interesting, still is an open problem to our best knowledge.

Last, the size of the data sets imposes severe requirements on the algorithms, making it necessary to reconsider most algorithms from ML, statistics and data analysis, or to reshape the learning task, e.g. incorporating progressive sampling [138] and active learning [22].

3.2 Evolutionary Computation

Evolutionary Algorithms (EAs) are stochastic optimization methods loosely inspired from the Darwinian evolution of biological populations [81, 9]. EAs historically include Genetic Algorithms [44], Evolution Strategies [114], Evolutionary Programming [37] and Genetic Programming [65, 8]; recent developments, referred to as Estimation of Distribution Algorithms [69, 43], extend the EC paradigm to the infinite population case.

Since the 90's an ever growing number of industrial applications have demonstrated significant improvements brought by EC over previous approaches from Numerical Optimization, Operational Research, Simulated Annealing [63] or Tabu Search [41].

Although some progress has been made regarding the theory of EC [18, 9, 13], the practice of EAs is far ahead of their theoretical understanding: EAs success stories emphasize their flexibility in handling the specificities of the problem at hand (e.g. dealing with non parametric search spaces such as program spaces [65, 112]) and taking advantage of prior knowledge (modeled through specific variation operators [60], and/or hybridization with domain heuristics, e.g. for TSP [80]).

Furthermore, the population-based dynamics of EAs allows one to tackle multi-modal optimization (sampling the multiple optima), constrained optimization (sampling the admissible optima), and multi-objective optimization (sampling the Pareto front, i.e. the set of the best compromises between the contradictory objectives) [28].

The main drawback of EAs is their cost; computational cost, as each optimization requires several hundred thousands evaluations of the objective function(s); but also manual cost, as the many parameters of EAs must be manually tuned almost anew for every problem.

The coupling of ML and EC techniques has been explored in the literature along several perspectives.

[2]In the particular case of Text Mining, concerned with mining documents and Web pages, the design of such elementary operators (stemming, tagging, etc) constitute learning problems *per se* [27].

First of all, EC techniques have been widely applied to ML goals (see [53, 131, 40, 120, 67] among many others).
More rarely, ML techniques have been used for adapting the EC dynamics, as pioneered in [121, 101, 19]; the major difficulty is that the computational savings (e.g. the reduction of the number of fitness evaluations needed to reach a given level of performance) are often canceled out by the ML induction cost (as the fitness evaluation cost is negligible for academic benchmark problems).

More recently, the coupling of EC with ML has been re-considered (GECCO Workshop 2002, 2003 [1, 100], Marc Schoenauer was among the organizers of the 2003 edition).

Finally, the development of Estimation Distribution Algorithms extensively relies on statistical learning techniques, estimating the (naive, chain-, tree- or graph-structured) Bayesian network fitting the best individuals in the population in order to generate the next population [7, 119, 123, 69, 43].

4 Proposed Research Themes

4.1 Using EC within Machine Learning and Data Mining

Among the many optimization problems occurring in ML and DM, this Research Project will consider the following ones: i) the domain representation (feature construction); ii) the search for a single hypothesis (learning criterion); iii) the search for a collection of hypotheses (diversity criterion); iv) the user's interest (implicit preferences). The choice of these four problems is motivated as one wants to consider raw data (problem i), find affordable hypotheses (problem ii), produce accurate classifiers (problems ii and iii) and finally, address the expert's implicit goals and produce satisfactory results (problem iv).

4.1.1 Change of representation and feature construction

It has been emphasized that the success of ML applications extensively depends on the problem representation. In application domains with abundant and low-level descriptors (e.g. image pixels, document words), the key learning task concerns the representation change, through selecting relevant features and/or constructing new compound features.

The literature involves two main approaches for representation changes: univariate approaches (see e.g. [126]) and wrapping approaches [64, 67, 14]. However, univariate approaches suffer from their myopia, e.g. the contribution of a single feature to a complex target concept appears poorly significant. Conversely, wrapping approaches are computationally intensive, as they use a learning component to evaluate the contribution of a feature subset to the target concept.

A third approach, based on an EC framework known as Parisian Evolution [23] will be investigated. In this approach, a population of features is evolved; each compound feature is independently evaluated using a new criterion (see section 4.1.2), enforcing a local and affordable evaluation; the global assessment of the features is achieved through the selection step, ensuring that each training example is accounted for by at least one compound feature in the population. This way, the local *vs* global aspects of representation changes are handled with a controllable computational cost.

Preliminary investigations have shown the potentialities of such an approach, extracting a sparse representation for images [59]. This approach, termed FIS-ICA, achieves the independent extraction of frequent patterns, sets of pairs (pixel-value), and loosely optimizes the independence of these patterns, inspired from Independent Component Analysis [54].

4.1.2 New criteria for hypothesis learning

Besides the generalization error (see section 3.1), a new criterion for comparing learning algorithms on problem instances has been proposed in [15, 91]. This criterion is based on the Receiver Operating Characteristics curve, originated from signal analysis and widely used in Medical Data Analysis.

Theoretical studies of the ROC curve (see e.g. [71, 35]) shows that the area under the ROC curve is a good measure of the learner quality; specifically, it addresses two main practical ML concerns, namely how to deal with imbalanced data distributions (when the target anomaly class involves 1% or 0,1% of the data) and asymmetric error costs (e.g. a false negative is usually more costly than a false positive) [30].

The area under the ROC curve (AUC) defines a combinatorial NP-complete optimization problem. Early approaches to this maximization problem [36, 84, 34], have considered GA-based or greedy optimization techniques.

Preliminary investigations [118, 117] have demonstrated the good performances of an ES-based algorithm named ROGER (ROC-based Genetic Learner) comparatively to one state-of-art SVM, SVMTorch [24]. Similar predictive accuracy is obtained for a fraction for the total SVMTorch cost, as the AUC criterion is computed with complexity $n\,log(n)$.

Further research will consider the extension of ROGER to deal with i) the learning of disjunctive concepts; ii) the sensitivity analysis of the features.

4.1.3 Ensemble learning

The principle underlying ensemble learning, (a complex target concept is better characterized by a set of "complementary" hypotheses than by a single best hypothesis) appears quite similar to the underlying principle of EC (a population-based approach is less prone to be trapped in local minima than a single individual-based approach).

Along these lines, we propose to hybridize ensemble learning with some EC heuristics, concerned with maintaining a pool of diversified and accurate solutions. Such heuristics might enforce both the robustness with respect to data noise, and the diversity of the hypotheses. Specifically, the research plan concern:

- Hybridization with multi-modal optimization heuristics, simultaneously evolving sub-populations covering different global optima (or quasi-optima) of the problem (e.g. sharing and crowding [73], segregational selection [70], clearing [92]). This line of research was first pioneered by the GA-based relational learner REGAL [40], dynamically distributing the computational effort with respect to "difficult" training examples.
  Additionally, intermediate results can be exploited to construct a (semi)-distance on the training examples [122, 115], based on their classification by the sequence of hypotheses. It then becomes possible to evaluate the typicity of an example, from the class distribution in its local neighborhood, and to detect outliers.

- Use of co-evolutionary optimization heuristics, simultaneously evolving/evaluating two species in a competitive or cooperative interaction [52, 87]. Such a co-evolution framework would neatly formalize the motto underlying ML : *good hypotheses are learned from good examples*. The typicity of examples evolves with the sequence of hypotheses ; inversely, the quality of a hypothesis depends on its reliability (the average typicity of all the examples it covers), and its added value (the average typicity of the examples it covers and that are not already covered by other hypotheses).

- Hybridization with multi-objective optimization heuristics, sampling the Pareto front of the problem, i.e. the best compromises between contradictory objectives [28]. As already mentioned, many learners (e.g. based on MDL-like [94, 85], bias-variance-like [108, 26], or regularized [45] learning criteria) turn an intrinsically multi-objective optimization problem, into a single-objective optimization problem by considering the weighted sum of the objectives. In practice, the weight put on the regularization term, or penalization factor, is determined empirically along an expensive cross-validation procedure.

Along these lines, instead of rejecting all hypotheses but the one learned with the optimal penalization, we propose to keep them all.

On one hand, this might be a costless way of constructing a collection of hypotheses (since all hypotheses are constructed anyway) that are diversified (since they optimize distinct criteria). Low penalization factors correspond to specific and complex hypotheses, covering examples that are relatively isolated or close to the frontier of the target concept. Moreover, the vote of such hypotheses might provide interesting clues regarding the reliability of the examples[3].

4.1.4 The expert in the loop

As already noted, the goal of data mining (discovering *new, useful and valid* knowledge about the problem at hand [33]) is clearly subjective. Specifically, filtering out trivial regularities (e.g. husbands are men) and retaining meaningful ones (e.g. two entities are correlated in the data due to some unexpected reason), resort to an AI-hard problem (common sense reasoning).

Actually, adjusting a good filter for DM (based on frequency threshold [2] and/or syntactical constraints [95] or interestingness measures [90]) is a decision making problem: the user usually proceeds by trials and errors, along an expensive search.

The proposed alternative aims at *Minded Mining Assistants*, combining approaches of Interactive Evolution and On-Line Learning. Interactive Evolution is a subfield of EC, concerned with the optimization of non computable fitness functions, e.g. related to taste [51], or aesthetics criteria [102, 12]. Interactive evolution (IE) relies on the expert in the EC loop, in charge of the evaluation/selection step (two workshops on this topic have been organized in the recent GECCO conferences).

The main drawback of IE is the number of evaluations that the expert will willingly perform. To address this drawback, Minded Mining Assistants will combine IE with on-line learning of the expert's preferences, in the line of User Profiling [79]. On-line learning (see for instance [20]), is a principled framework for incremental learning from iid samples. The extension to non iid distributions, more realistic in a DM application, will be required.

Minded Mining Assistants could hit two birds with one stone: First, the expert input could directly guide the search for interesting hypotheses among the valid ones that the system can generate at will. Second, the assistant learns to characterize the expert's interest function. Such an interest function can be used to reduce the amount of "clicks" asked to the expert. But it would also be a result on itself, that could be later used for expertise exchange and archiving, as well as for educative purposes.

4.2 Using Machine Learning within Evolutionary Computation

Symmetrically, among the ML-relevant problems occurring in Evolutionary Computation, this Research Project will focus on: i) the domain representation; ii) the cost reduction of fitness evaluation over one run; iii) the overall cost reduction of EC over an entire application. A fourth research direction is that of Estimation of Distribution Algorithms, where evolution simultaneously learns and exploits a distribution over the problem domain [7, 123, 119, 69, 88, 43].

[3] Typically, if all "sufficiently complex" hypotheses classify an example as positive, and all simpler hypotheses classify this example as negative, the example reliability is much higher than for an example which is indifferently classified positive and negative by hypotheses of any complexity.

4.2.1 Adding domain knowledge to the representation

Constraints on the search space

As already mentioned, Evolutionary Computation and particularly Genetic Programming [66, 8] can handle non-parametric search spaces. Many problems can hence be revisited, as they were initially modeled with the idea of parametric optimization in mind (Marc Schoenauer's work in Structural Design of Mechanical Structures [47, 48], or in underground identification from seismic data [76, 75] are examples of such situations).

However, the larger and the more complex the search space, the more prior knowledge is needed to guide the search [111]. Prior knowledge can be explicit, e.g. expressed as dimensionality constraints on physical or mechanical, models. Earlier works coupling GP with context free grammars have demonstrated the benefit of reducing the search space of non-parametric model identification in Mechanics, through dimensionality constraints [97, 99, 98] (see also [62]).

It might also be desirable to capture implicit prior knowledge concerning the model plausibility. In some engineering problems, for instance underground model identification [75], the fitness function only provides partial indications in the sense that part of the optimal models (wrt the fitness function) are meaningless (from a geological perspective) [76]. In such cases, a learner trained from both plausible and meaningless models can be used to capture such an implicit prior knowledge.

The benefit of using such prior knowledge is twofold. On one hand, relevant restrictions of the search space most often improve the performance; on the other hand, complying with the expert's principles significantly increases his/her satisfaction, and facilitates the cooperation.

The use of prior knowledge in EC will be investigated along two directions. Firstly, the ML expertise concerning the representation and exploitation of explicit knowledge (e.g. through grammars, Horn clauses) will be exploited. Secondly, in many cases, no explicit prior knowledge is available. ML algorithms can thus be trained to distinguish plausible from non-sense solutions, and the learned classifier will be further used as a constraint on the EC search space.

Extensional representations

Many numerical engineering problems are concerned with the identification of models, such as the behavioral law for mechanical problems [112], the command in control problems, the equilibrium equations for physico-chemical systems [132], ... Such models are exploited through computationally demanding simulators, to characterize the underlying physical, chemical, etc, phenomenon.

The construction of such models, known as *system identification*, is not an ML problem as very few experimental data, if any, are available (this is the case for instance for the *isotherm* function in chromatography [32], where each sample point requires several months of experimentation). In such situations, it is important to use as much as possible the available sample points. Presently, this information is only accounted for in the optimality criterion, penalizing the solutions which do not model properly the experiments.

System identification problems are often mathematically ill-posed (no uniqueness, or even sometimes no existence result can be theoretically derived), making the application of standard numerical techniques (use gradient algorithms, and get the gradient by solving the adjoint problem) impossible without very severe restrictions. Standard Evolutionary Algorithms have been applied in such cases, and found to give satisfactory results [32, 112, 82].

Another way of enforcing the use of experimental samples within EC-based system identification, based on the genotypic/phenotypic heuristics, will be investigated. In this approach, the individual genotype is made of a set of candidate sample points. These data, plus the true experiment samples, are used to learn a function, e.g. using regression SVMs[4] [1]; this function is taken as the individual phenotype, and evaluated as candidate model.

[4]It must be noted that the parameters of the learning algorithm, which are not always easy to tune, can be incorporated into the genotype and thus adjusted along evolution, as done for self-adaptive mutation [114].

This way, the inclusion of an ML component in the genotype-to-phenotype fully enables to exploit the rare available data. The challenge comes from the size of the search space (the candidate sample points); a possibility is to use ensemble learning-based heuristics (section 4.1.3) to estimate the quality of candidate samples, and speed up the search.

4.2.2 Approximating the fitness function

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. Such approximations, referred to as surrogate fitness functions, have been combined with classical optimization methods for years in *Response Surface Method*, and they are increasingly used in EC [57]. Their use rise many issues: i) the type of surrogate function (built using classical approximation, neural nets, krieging [96], SVMs [1]); ii) the learning schedule, when is the surrogate function updated [3]; iii) the learning data: should the training sample focus on the good regions.

ML-based heuristics can be used for a better interaction between evolutionary optimization and surrogate estimation. Specifically, ensemble learning would allow to capture the multi-facets of the fitness function, and provide more accurate surrogates. Second, the sequence of surrogate functions acquired along evolution can be exploited to better select the promising solutions, the exact fitness of which must be computed.

Also, it might be useful to construct more abstract hypotheses on the search space, e.g. characterizing the "interesting regions", with above-average fitness. Such hypotheses can be used to guide (or replace [19]) the variation operators.

4.2.3 Learn across runs

On each problem domain, many EC runs are routinely performed, generating massive amounts of data. Such data are poorly exploited at the moment: preliminary runs are used to adjust the EC parameters; other runs are essentially used to compute the best, average best and standard deviation of the solutions.

However, the exploitation of an EC archive has been shown to contribute to multi-objective optimization (sampling the Pareto front [28]) or co-evolution (e.g. using a Hall of Fame to avoid cycles [103]).

More generally, an EC archive, gathered along and across runs could be exploited in a more principled and systematic way:

- in order to recover the genetic diversity when evolution stagnates. Instead of partially reinitializing the population (random migrants), a better alternative might be to incorporate individuals from the archive which are both sufficiently different (at the genotypic level), and similarly fit; in this way, random migrants would be replaced by diversified and competent migrants, preserving the population diversity throughout the selection effects.

- in order to detect and stop earlier the unpromising runs. The stop criterion most widely used depends on the number of consecutive fitness evaluations bringing no improvement. However, the EC dynamics greatly varies from one run to another, typically depending on the first basins of attraction encountered by the population. By observing the EC trajectories, one might estimate the "expected progress rate" associated to a given population distribution, measuring its probability $p(\Delta(fitness), cost)$ of improving the best fitness by $\Delta(fitness)$ for a given computational *cost*. The challenge is to find an adequate description of population distributions, ensuring a good trade-off between its complexity (construction and exploitation cost) and its precision (accuracy of the progress rate).

- in order to reconsider the domain representation. Basically, the relative impact of the domain attributes can be studied from evolution-specific variables, such as the self adaptive mutation rates. On-going research is concerned with determining the key design variables for quantum control [6, 5] along these lines.

4.2.4 Estimation of Distribution Algorithms

A new type of Evolutionary Algorithms has been proposed in the last decade, called *Estimation of Distribution Algorithms* (EDAs) [7, 69, 119, 88, 43]. EDAs extend EC to the infinite population case: instead of considering a set of individuals, EDAs consider a distribution over the problem domain, which is alternatively used to generate a (finite) population, and updated from the best individuals in the population.

Recent investigations have shown the equivalence of EDAs and standard EC, modeling the variation + selection operators as the convolution of the distribution with some ad hoc kernel [127].

The central step of EDAs is the construction of a distribution, after the best individuals in the current population. This step will be reconsidered using Machine Learning techniques. Firstly, more flexible EDAs will be devised, replacing Bernoulli, Gaussian or Bayesian nets-based distributions with mixtures of distributions; this research direction meets some of the research topics of the *SELECT* project.

Secondly, the search for a distribution on the problem domain generalizes many of the research topics mentioned in this section (e.g. constructing fitness surrogates, or exploiting an EC archive to determine the expected progress rate associated to a given population distribution). A formal framework will be defined to study the required properties of such distributions (e.g. inducing a partial order on the problem domain [21]; ensuring a generative process; addressing the exploration *vs* exploitation dilemma).

5 Applications

This Research Project will focus on three main application domains: Robotics, Medical Data Mining, and Numerical Engineering. The first domain illustrates the coupling of evolution-based optimal control with data mining, systematically exploiting the wealth of information gathered along the robot explorations. The second domain focuses on the interaction with the expert, navigating along intelligible and yet precise representations of critical (noisy, disjunctive, with imbalanced distributions, asymmetrical classification cost) data. The third domain is concerned with incorporating knowledge domain into computationally heavy inverse problems issued from industrial needs.

5.1 Robotics

Marc Schoenauer, Michele Sebag and Antoine Cornuejols collaborate since 2001 on the Robea project *Action, Anticipation and Adaptation* (A3) [42] (coll. LIMSI, Philippe Tarroux and Jean-Sylvain Liénard).

This project is concerned with evolutionary robotics [86] and the design of autonomous robot controllers. Specifically, the considered controller is based on the interaction of three modules: an Action module, an Anticipation module and an Adaptation module (PhD Nicolas Godzik, Mary Felkin, Jérémie Mary).

The Action module is classically concerned with the action selection, based on high-level features (perceptions) constructed from the low-level robot sensors (sensations).

The Anticipation module is concerned with constructing a Robot "Cognitive Map", modeling the environment. Such a model is commonly recognized mandatory for achieving complex tasks (e.g. localizing a target locus in an environment with perceptual aliasing [68], or reaching an ambiguous goal [130]). The approach investigated, inspired from sensori-motor contingencies [89], encodes the cognitive map into an anticipation module, predicting the next sensations of the robot depending on its current sensations and the selected action. In this way, the cognitive map sidesteps the AI-hard problems of building a declarative description of the world; instead, a procedural description, using the additional information of the robot action, is built.

Interestingly, the anticipation module provides the robot with a meta-information: either everything goes as predicted (the anticipation is accurate), or there is something wrong (the effective sensations differ from that predicted). In the latter case, the robot is informed that either its action does not produce the correct effects (e.g. due to some motor problems), or the action selection is not appropriate (e.g. due to ambiguous goals) or the perceptions are ambiguous (due to perceptual aliasing).

In such cases, the Adaptation module will enforce the modification of the other modules.

The A3 scheme relies on the tight coupling of evolution and learning: the Action and Anticipation modules are optimized by EC, while the Anticipation module is trained by ML, exploiting the examples gathered along all robot runs. In particular, the Anticipation module can be trained independently and constitute a common resource for the task at hand.

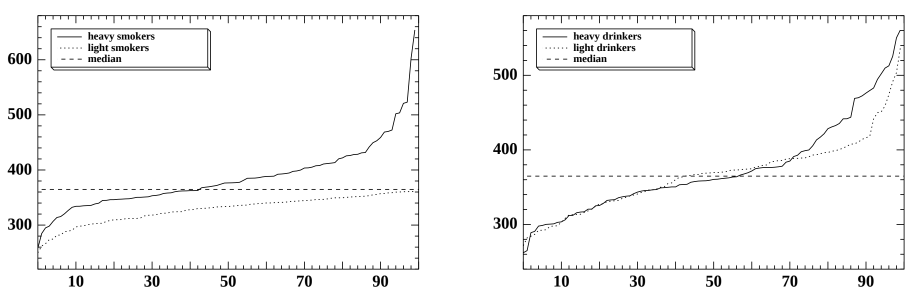
5.2 Medical Data Mining

A strong expertise in medical risk assessment has been developed in the I & A team: first and second contributions to the Predictive Toxicology Evaluation Challenge, IJCAI 97 [125, 120]; identification of cardio-vascular risk factors, Indana project 2000-2004 (coll. Hopital Broussais, Equipe Grappa-Lille3, LIP6); Atherosclerosis risk factor identification, Challenge PKDD 2002, 2003 [117]; readmission modeling, coll. HUG Genève, Master thesis Gilles Cohen 2001.

Medical data present specific characteristics: intrinsic variability of the target concept; skewed data distributions (aimed at comparatively rare diseases); asymmetric error cost (a false positive being in general much better than a false negative).

These characteristics have inspired the ROC-based learner ROGER (section 4.1.2); ROC curves, used for decades in medical data analysis [25], permit a robust assessment of hypotheses wrt noise, rare phenomenons, and misclassification costs.

Further, it appears that the ROC framework eases the communication with the medical experts. Firstly, they are used to read ROC curves, which facilitates their assessment of the learned hypotheses. Secondly, the ROC hypotheses permit an impact study of the risk factors (e.g. tobacco, alcohol); again, the impact study is visually represented (Fig. 1), materializing the risk of e.g. smokers against non smokers, drinkers against non drinkers in a precise and yet readable way.



(a) Tobacco (b) Alcohol

Figure 1: Tobacco and Alcohol Impacts on Atherosclerosis Risks

Last, the collection of hypotheses provided through independent EC runs enables a sensitivity analysis of the medical factors, determining their average contribution to the target concept conditioned by the other factors.

5.3 Learning and Optimization for Numerical Engineering

The learning and mining approach originally developed in CMAP (Marc Schoenauer) and LMS (Michèle Sebag), Ecole Polytechnique, to address ill-posed inverse problems, will be continued on the following applications, tightly coupled with the fundamental research themes of section 4.2:
– Model identification in chromatography, coll. U-Tenessee USA, Guiochon, with very few and expensive examples (see section 4.2.1);
– Underground identification in geology (PhD Vijay Pratap Singh, CIFRE IFP), with implicit prior knowledge about plausible solutions (section 4.2.1);
– Fatigue modeling from DVD aging experimental data, coll. Laboratoire d'Acoustique Musicale, Paris VI;
– Interactive optimization of design parameters for musical instruments (section 4.1.4), coll. Laboratoire d'Acoustique Musicale, Paris VI;
– Optimization for quantum control, involving a feature selection problem (section 4.2.3), coll. Cermics;
– Characterization of good FE meshes for aeronautic design (PhD Mathieu Pieres, CIFRE Airbus), first involving the design of a representation for meshes (section 4.2.1, then the resolution of a Machine Learning problem (characterizing good meshes), and finally an inverse optimization problem (generating good meshes).

6 Collaborations

Most on-going national and international collaborations have been cited in this document. Other collaborations are with Università di Piemonte Orientale, L. Saitta and A. Giordana; Vrije Universiteit Amsterdam, A. Eiben; ETH Zurich, P. Kammoutsakos; Università di Bari, F. Esposito; Leuven University, H. Blockeel; U. Yokohama, E. Suzuki (co-advisor J. Maloberti); U. Sapporo, F. Furukawa.

The *TAO* project also collaborates with other INRIA projects: Micmac-Cermics (PhD A. Auger, co-supervised by Claude LeBris and Marc Schoenauer); Gemo (PhD A. Termier, co-supervised by Marie-Christine Rousset and Michèle Sebag [129, 128]); Mostrare (R. Gilleron, Indana project); and Select (G. Celeux, P. Massart, J.-M. Loubes).

Other collaborations concern LIMSI (Robea project), and LRI, Genomics.

References

[1] K. Abboud, X. Merlhiot, and M.Schoenauer. Deterministic function learning within evolutionary algorithms. In *GECCO 2002 Workshop Program*, pages 148–151, 2002.

[2] R. Agrawal and R. Srikant. Fast algorithms for mining association rules. In J. B. Bocca, M. Jarke, and C. Zaniolo, editors, *Proc. 20th Int. Conf. Very Large Data Bases, VLDB*, pages 487–499. Morgan Kaufmann, 1994.

[3] L. Albert and D. Goldberg. Efficient discretization scheduling in multiple dimensions. In W.B. Langdon & al., editor, *Proceedings of the Genetic and Evolutionary Conference 2002*, pages 271–278. Morgan Kaufmann, 2002.

[4] M.G. Arenas, P. Collet, A.E. Eiben, M. Jelasity, J.J. Merelo, B. Paechter, M. Preuß, and M. Schoenauer. Dream: A framework for distributed evolutionary algorithms. In J.J. Merelo et al., editor, *Proceedings of the* 7^{th} *Conference on Parallel Problems Solving from Nature*, pages 665–675. Springer-Verlag, LNCS 2439, 2002.

[5] A. Auger, C. Le Bris, and M. Schoenauer. Dimension-independent convergence rate for non-isotropic $(1, \lambda) - es$. In E. Cantu-Paz et al., editor, *Proceedings of the Genetic and*

Evolutionary Conference 2003, pages 512–524. LNCS 2723 and 2724, Springer Verlag, 2003.

[6] A. Auger, A. Ben Haj Yedder, E. Cancès, C. Le Bris, C. M. Dion, A. Keller, and O. Atabek. Optimal laser control of molecular systems: Methodology and results. *Mathematical Models and Methods in Applied Sciences*, 12(9):1281–1315, 2002.

[7] S. Baluja and R. Caruana. Removing the genetics from the standard genetic algorithms. In A. Prieditis and S. Russel, editors, *Proceedings of the 12^{th} International Conference on Machine Learning*, pages 38–46. Morgan Kaufmann, 1995.

[8] W. Banzhaf, P. Nordin, R.E. Keller, and F.D. Francone. *Genetic Programming — An Introduction On the Automatic Evolution of Computer Programs and Its Applications*. Morgan Kaufmann, 1998.

[9] T. Bäck. *Evolutionary Algorithms in theory and practice*. New-York:Oxford University Press, 1995.

[10] S. BenHamida and M. Schoenauer. An adaptive algorithm for constrained optimization problems. In M. Schoenauer et al., editor, *Proceedings of the 6^{th} Conference on Parallel Problems Solving from Nature*, LNCS, pages 529–539. Springer Verlag, 2000.

[11] S. BenHamida and M. Schoenauer. Aschea: New results using adaptive segregational constraint handling. In *Congress on Evolutionary Computation 2002*, pages 884–889. IEEE Press, 2002.

[12] P.J. Bentley and D. W. Corne, editors. *Creative Evolutionary Systems*. Morgan Kaufmann, 2001.

[13] H.-G. Beyer. *The Theory of Evolution Strategies*. Springer Verlag, 2001.

[14] M. Botta, A. Giordana, L. Saitta, and M. Sebag. Relational learning as search in a critical region. *Journal of Machine Learning Research*, 4:431–463, 2003.

[15] A.P. Bradley. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 1997.

[16] L. Breiman. Arcing classifiers. *Annals of Statistics*, 26(3):801–845, 1998.

[17] A. Casali, R. Cicchetti, and L. Lakhal. Extracting semantics from data cubes using cube transversals and closures. In *Proc. of Knowledge Discovery from Databases*, pages 69–78, 2003.

[18] R. Cerf. *Une théorie assymptotique des algorithmes génétiques*. PhD thesis, Université de Montpellier II, March 1994.

[19] G. Cervone, R.S. Michalski, K. Kaufman, and L.A. Panait. Combining Machine Learning with Evolutionary Computation: Recent Results on LEM. In *Multistrategy Learning 2000*, pages 41–58, 2000.

[20] N. Cesa-Bianchi, Y. Freund, D. Haussler, D. Helmbold, R. Schapire, and M. Warmuth. How to use expert advice. *Journal of the ACM*, 44(3):427–485, 1997.

[21] W. W. Cohen, R. E. Schapire, and Y. Singer. Learning to order things. In Michael I. Jordan, Michael J. Kearns, and Sara A. Solla, editors, *Advances in Neural Information Processing Systems*, volume 10. The MIT Press, 1998.

[22] D. A. Cohn, Z. Ghahramani, and M. I. Jordan. Active learning with statistical models. In G. Tesauro, D. Touretzky, and T. Leen, editors, *Advances in Neural Information Processing Systems*, volume 7, pages 705–712. The MIT Press, 1995.

[23] P. Collet, E. Lutton, F. Raynal, and M. Schoenauer. Polar ifs + parisian gp = efficient inverse ifs problem solving. *Genetic Programming and Evolvable Machines*, 1(4):339–361, 2000.

[24] R. Collobert and S. Bengio. Svmtorch: Support vector machines for large-scale regression problems. *Journal of Machine Learning Research*, 1:143–160, 2001.

[25] I. Colombet, M.-C. Jaulent, P. Degoulet, and G. Chatellier. Logistic regression model: An assessment of variability of predictions. In V. Patel et al., editor, *MEDINFO 2001*, pages 1314–1318. IOS Press, 2001.

[26] N. Cristianini and J. Shawe-Taylor. *An introduction to Support Vector Machines*. Cambridge University Press, 2000.

[27] J. Cussens. Part-of-speech tagging using progol. In S. Džeroski and N. Lavrač, editors, *Proceedings of the 7th International Workshop on Inductive Logic Programming*, volume 1297, pages 93–108. Springer-Verlag, 1997.

[28] K. Deb. *Multi-Objective Optimization Using Evolutionary Algorithms*. John Wiley, 2001.

[29] Philip Derbeko, Ran El-Yaniv, and Ron Meir. Variance optimized bagging. In T. Elomaa, H. Mannila, and H. Toivonen, editors, *Proc. ECML 2002*, pages 60–71. LNCS 2430, Springer Verlag, 2002.

[30] P. Domingos. Meta-cost: A general method for making classifiers cost sensitive. In *Knowledge Discovery from Databases*, pages 155–164. Morgan Kaufmann, 1999.

[31] A.E. Eiben and M. Schoenauer. Evolutionary computing. *Information Processing Letter*, 82(1):1–6, 2002.

[32] A. Fadda and M. Schoenauer. Evolutionary chromatographic law identification by recurrent neural nets. In J. R. McDonnell, R. G. Reynolds, and D. B. Fogel, editors, *Proceedings of the 4^{th} Annual Conference on Evolutionary Programming*, pages 219–235. MIT Press, March 1995.

[33] U.M. Fayyad, G. Piatetsky-Shapiro, and P. Smyth. From data mining to knowledge discovery: An overview. In *Advances in Knowledge Discovery and Data Mining*, pages 1–34. MIT Press, 1996.

[34] C. Ferri, P. A. Flach, and J. Hernández-Orallo. Learning decision trees using the area under the ROC curve. In Morgan Kaufmann, editor, *Proceedings of the 19^{th} International Conference on Machine Learning*, pages 179–186, 2002.

[35] P. Flach. The geometry of roc space: Understanding machine learning metrics through roc isometrics. In *Proceedings of the 20^{th} International Conference on Machine Learning*. Morgan Kaufmann, 2003.

[36] D.B. Fogel, E.C. Wasson, E.M. Boughton, V.W. Porto, and P.J. Angeline. Linear and neural models for classifying breast cancer. *IEEE Trans. Medical Imaging*, 17:3:485–488, 1998.

[37] L. J. Fogel, A. J. Owens, and M. J. Walsh. *Artificial Intelligence through Simulated Evolution*. New York: John Wiley, 1966.

[38] Y. Freund and R.E. Shapire. Experiments with a new boosting algorithm. In L. Saitta, editor, *Proceedings of the 13th International Conference on Machine Learning*, pages 148–156. Morgan Kaufmann, 1996.

[39] T. Gartner, K. Driessens, and J. Ramon. Graph kernels and gaussian processes for relational reinforcement learning. In T. Horvath and A. Yamamoto, editors, *Proceedings of ILP 2003*, pages 146–163. LNCS 2835, Springer Verlag, 2003.

[40] A. Giordana and F. Neri. Search intensive concept induction. *Evolutionary Computation*, 3(4):375–416, 1994.

[41] F. Glover. Genetic algorithms and scatter search: Unsuspected potentials. *Statistics, and Computing*, 4:131–140, 1994.

[42] N. Godzik, M. Schoenauer, and M. Sebag. Evolving symbolic controllers. In G. Raidl et al., editor, *Applications of Evolutionary Computing*, LNCS 2611, pages 638–650. Springer Verlag, 2003.

[43] D. Goldberg. *The Design of Innovation: Lessons from and for Genetic and Evolutionary Algorithms.* MIT Press, 2002.

[44] D. E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning.* Addison Wesley, 1989.

[45] Y. Grandvalet and S. Canu. Comments on "noise injection into inputs in back propagation learning" [78]. *IEEE transactions on Systems, Man and Cybernetics*, 1995.

[46] I. Guyon and eds A. Elisséeff, editors. *Special issue on Variable and feature selection.* Journal of Machine Learning Research, 2003.

[47] H. Hamda, F. Jouve, E. Lutton, M. Schoenauer, and M. Sebag. Compact unstructured representations in evolutionary topological optimum design. *Applied Intelligence*, 16:139–155, 2002.

[48] H. Hamda and M. Schoenauer. Toward hierarchical representations for evolutionary topological optimum design. In J. Périaux et al., editor, *Eurodays 2000, in memoriam of B. Mantel.* John Wiley, 2002.

[49] J. Han and M. Kamber. *Data Mining: Concepts and Techniques.* Morgan Kaufmann, 2000.

[50] D. J. Hand, H. Mannila, and P. Smyth. *Principles of Data Mining.* MIT Press, 2000.

[51] M. Herdy. Evolution strategies with subjective selection. In H.-M. Voigt, W. Ebeling, I. Rechenberg, and H.-P. Schwefel, editors, *Proceedings of the 4th Conference on Parallel Problems Solving from Nature*, volume 1141 of *LNCS*, pages 22–31. Springer Verlag, 1996.

[52] W. D. Hillis. Co-evolving parasites improves simulated evolution as an optimization procedure. In C. G. Langton, editor, *Proceedings Artificial Life II*, pages 313–324. Addison Wesley, 1992.

[53] J. H. Holland. *Adaptation in Natural and Artificial Systems.* University of Michigan Press, Ann Arbor, 1975.

[54] A. Hyvarinen, J. Karhunen, and E. Oja. *Independent Component Analysis.* Wiley New York, 2001.

[55] T. Jaakkola and D. Haussler. Exploiting generative models in discriminative classifiers. In M.J. Kearns, S.A. Solla, and D.A. Cohn, editors, *Proc. NIPS-11.* MIT Press, 1999.

[56] B. Jeudy and J.-F. Boulicaut. Using condensed representations for interactive association rule mining. In *Principles and Practive of Knowledge Discovery from Databases*, pages 225–236. Springer Verlag LNAI 2431, 2002.

[57] Y. Jin. A comprehensive survey of fitness approximation in evolutionary computation. *Soft Computing Journal*, 2003.

[58] T. Joachims. *Learning to Classify Text using Support Vector Machines.* Kluwer, 2002.

[59] S. Jouteau, A. Cornuéjols, M. Sebag, P. Tarroux, and J.-S. Liénart. Nouveaux résultats en classification à l'aide d'un codage par motifs fréquents. *Revue Extraction des Connaissances et Apprentissage (ECA), Actes EGC 2003*, 17:521–532, 2002.

[60] C. Kane. *Algorithmes génétiques et Optimisation topologique.* PhD thesis, Université de Paris VI, July 1996.

[61] M.J. Kearns and U.V. Vazirani. *An Introduction to Computational Learning Theory.* MIT Press, 1994.

[62] M. Keijzer and V. Babovic. Dimensionally aware genetic programming. In *Proceedings of the Genetic and Evolutionary Conference 1999*, pages 1069–1076. Morgan Kaufmann, 1999.

[63] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220:671–680, 1983.

[64] R. Kohavi and G.H. John. Automatic parameter selection by minimizing estimated error. In A. Prieditis and S. Russell, editors, *Proceedings of the 12th International Conference on Machine Learning*, pages 304–312. Morgan Kaufmann, 1995.

[65] J. R. Koza. *Genetic Programming: On the Programming of Computers by means of Natural Evolution.* MIT Press, Massachusetts, 1992.

[66] J. R. Koza. *Genetic Programming II: Automatic Discovery of Reusable Programs.* MIT Press, Massachussetts, 1994.

[67] K. Krawiec. Genetic programming based construction of features for machine learning and knowledge discovery tasks. *Genetic Programming and Evolvable Machines*, 3(4):329–344, 2002.

[68] Pier Luca Lanzi. Adaptive Agents with Reinforcement Learning and Internal Memory. In *SAB2000*, pages 333–342. MIT Press, 2000.

[69] P. Larranaga and J. A. Lozano. *Estimation of Distribution Algorithms. A New Tool for Evolutionary Computation.* Kluwer Academic Publishers, 2001.

[70] R. G. Leriche, C. Knopf-Lenoir, and R. T. Haftka. A segregated genetic algorithm for constrained structural optimization. In L. J. Eshelman, editor, *Proceedings of the 6th International Conference on Genetic Algorithms*, pages 558–565, 1995.

[71] C.X. Ling, J. Hunag, and H. Zhang. Auc: a better measure than accuracy in comparing learning algorithms. In *Proc. of 16th Canadian Conference on AI 2003*, 2003.

[72] N. Lucas, J. Azé, and M. Sebag. Atherosclerosis risk identification and visual analysis. In *Discovery Challenge ECML-PKDD 2002*. http://lisp.vse.cz/challenge/ecmlpkdd2002/, 2002.

[73] S. W. Mahfoud. A comparison of parallel and sequential niching techniques. In L. J. Eshelman, editor, *Proceedings of the 6^{th} International Conference on Genetic Algorithms*, pages 136–143. Morgan Kaufmann, 1995.

[74] H. Mannila and H. Toivonen. Levelwise search and borders of theories in knowledge discovery. *Data Mining and Knowledge Discovery*, 1(3):241–258, 1997.

[75] F. Mansanné and M. Schoenauer. An automatic geophysical inversion procedure using a genetic algorithm. In P. Wong, F. Aminzadeh M., and Nikravesh, editors, *Soft Computing for Reservoir Characterization and Modeling*, volume 80 of *Studies in Fuzziness and Soft Computing*, pages 331–353. Physica-Verlag, 2002.

[76] F. Mansanne, F. Carrère, A . Ehinger, and M. Schoenauer. Evolutionary algorithms as fitness function debuggers. In Z. W. Ras and A. Skowron, editors, *Foundation of Intelligent Systems 12th International Symposium, ISMIS99*, volume 1609 of *LNCS*. Springer Verlag, 1999.

[77] Dragos Margineantu and Thomas G. Dietterich. Pruning adaptive boosting. In *Proceedings of the Fourteenth International Conference on Machine Learning*, pages 211–218. Morgan Kaufmann, 1997.

[78] K. Matsuoka. Noise injection into inputs in back propagation learning. *IEEE transactions on Systems, Man and Cybernetics*, 22(3):436–440, 1992.

[79] A.K. McCallum, K. Nigam, J. Rennie, and K. Seymore. Automating the construction of internet portals with machine learning. *Information Retrieval*, 3(4):127–163, 2000.

[80] P. Merz and B. Freisleben. Genetic local search for the TSP: New results. In T. Bäck, Z. Michalewicz, and X. Yao, editors, *Proceedings of the Fourth IEEE International Conference on Evolutionary Computation*, pages 159–164. IEEE Press, 1997.

[81] Z. Michalewicz. *Genetic Algorithms+Data Structures=Evolution Programs*. Springer Verlag, New-York, 1996. 3rd edition.

[82] Z. Michalewicz, D. Dasgupta, R. Leriche, and M. Schoenauer. Evolutionary algorithms for constrained engineering problems. *Computers & Industrial Engineering Journal*, 30(2), April 1996.

[83] T.M. Mitchell. *Machine Learning*. McGraw-Hill, 1997.

[84] M.C. Mozer, R. Dodier, M. C. Colagrosso, C. Guerra-Salcedo, and R. Wolniewicz. Prodding the ROC curve: Constrained optimization of classifier performance. In *Proc. NIPS-13*. MIT Press, 2001.

[85] S. Muggleton. Inverse entailment and PROGOL. *New Gen. Comput.*, 13:245–286, 1995.

[86] S. Nolfi and D. Floreano. *Evolutionary Robotics*. MIT Press, 2000.

[87] J. Paredis. Coevolving cellular automata: Be aware of the red queen! In Th. Bäck, editor, *Proceedings of the 7^{th} International Conference on Genetic Algorithms*, pages 393–400. Morgan Kaufmann, 1997.

[88] M. Pelikan and E. Goldberg, D.E.and Cantu-Paz. Boa: the bayesian optimization algorithm. In D.E. Goldberg & al., editor, *Proceedings of the Genetic and Evolutionary Conference 1999*, pages 525–532. Morgan Kaufmann, 1999.

[89] D. Philipona, J.K. O'Regan, and J.-P. Nadal. What is space? is there something out there?, 2002. Submitted.

[90] G. Piatetsky-Shapiro and B. Masand. Estimating campaign benets and modeling lift. In *Proc. Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 185–193. ACM Press, 1999.

[91] F. Provost, T. Fawcett, and R. Kohavi. The case against accuracy estimation for comparing classifiers. In *Proceedings of the 15th International Conference on Machine Learning*, pages 445–553. Morgan Kaufmann, 1998.

[92] A. Pétrowski. A new selection operator dedicated to speciation. In Th. Bäck, editor, *Proceedings of the 7th International Conference on Genetic Algorithms*, pages 144–151. Morgan Kauffman, 1997.

[93] D. Quadraglia, J. Périaux, C. Poloni, and G. Winter, editors. *Genetic Algorithms and Evolution Strategies in Engineering and Computer Sciences*. John Wiley, 1997.

[94] J.R. Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann, 1993.

[95] L. De Raedt and S. Kramer. The levelwise version space algorithm and its application to molecular fragment finding. In *Proceedings of Int. Joint Conf. on Artificial Intelligence, IJCAI'01*, pages 853–862. Morgan Kaufmann, 2001.

[96] A. Ratle. Accelerating the convergence of evolutionary algorithms by fitness landscape approximation. In Th. Bäck, G. Eiben, M. Schoenauer, and H.-P. Schwefel, editors, *Proceedings of the 5th Conference on Parallel Problems Solving from Nature*, pages 87–96. Springer-Verlag, LNCS 1498, 1998.

[97] A. Ratle and M. Sebag. Genetic programming and domain knowledge: Beyond the limitations of grammar-guided machine discovery. In M. Schoenauer et al., editor, *Proceedings of the 6th Conference on Parallel Problems Solving from Nature*, pages 211–220. Springer-Verlag, LNCS 1917, 2000.

[98] A. Ratle and M. Sebag. Grammar-guided genetic programming and dimensional consistency: Application to non-parametric identification in mechanics. *Applied Soft Computing*, 1(3):105–118, 2001.

[99] A. Ratle and M. Sebag. A novel approach to machine discovery: Genetic programming and stochastic grammars. In S. Matwin, editor, *Proceedings of Twelfth International Conference on Inductive Logic Programming*. Springer Verlag, 2003, to appear.

[100] Alain Ratle and Michèle Sebag. Stochastic grammatical programming to avoid the bloat. In *Proceedings of the GECCO-2002 Workshop Program*. Morgan Kaufmann, 2002.

[101] C. Ravisé and M. Sebag. An advanced evolution should not repeat its past errors. In L. Saitta, editor, *Proceedings of the 13th International Conference on Machine Learning*, pages 400–408. Morgan Kaufmann, 1996.

[102] A. Rosete, A. Ochoa, and M. Sebag. Automatic graph drawing and stochastic hill-climbing. In *Proc. of the Genetic and Evolutionary Computation Conf.*, pages 1699–1706. Morgan Kaufmann, 1999.

[103] C. D. Rosin and R. K. Belew. New methods for competitive coevolution. *Evolutionary Computation*, 5(1):1–29, 1997.

[104] O. Roudenko, T. Bosio, R. Fontana, and M. Schoenauer. Optmization of car front crash members. In *Artificial Evolution'01*, pages 202–214. Springer Verlag, LNCS 2310, 2002.

[105] Gunnar Rätsch, Bernhard Scholkopf, Alex Smola, Sebastian Mika, Takashi Onoda, and Klaus-Robert Müller. Robust ensemble learning for data analysis, 2000.

[106] R.E. Schapire. The strength of weak learnability. *Machine Learning*, 5:197, 1990.

[107] R.E. Schapire. Theoretical views of boosting. In *Proceedings of EuroCOLT-99, European Conference on Computational Learning Theory*, pages 1–10, 1999.

[108] B. Schölkopf, C. Burges, and A. Smola. *Advances in Kernel Methods: Support Vector Machines*. MIT Press, Cambridge, MA, 1998.

[109] T. Schneider and D. N. Mastronarde. Fast multiple alignment of ungapped DNA sequences using information theory and a relaxation method. *DAMATH: Discrete Applied Mathematics and Combinatorial Operations Research and Computer Science*, 71, 1996.

[110] M. Schoenauer and M. Sebag. Evolution artificielle et problèmes spatiaux. In H. Prade, R. Jeansoulin, and C. Garbay, editors, *Le temps, l'espace et l'évolutif en sciences du traitement de l'information*, volume 2, pages 377–396. Cepadues, 2000.

[111] M. Schoenauer and M. Sebag. Using Domain Knowledge in Evolutionary System Identification. In K. Giannakoglou et al., editor, *Evolutionary Algorithms in Engineering and Computer Science*. John Wiley, 2002.

[112] M. Schoenauer, M. Sebag, F. Jouve, B. Lamy, and H. Maitournam. Evolutionary identification of macro-mechanical models. In P. J. Angeline and Jr K. E. Kinnear, editors, *Advances in Genetic Programming II*, pages 467–488, Cambridge, MA, 1996. MIT Press.

[113] Greg Schohn and David Cohn. Less is more: Active learning with support vector machines. In Pat Langley, editor, *Proceedings of the 17^{th} International Conference on Machine Learning*, pages 839–846. Morgan Kaufmann, 2000.

[114] H.-P. Schwefel. *Numerical Optimization of Computer Models*. John Wiley & Sons, New-York, 1981. 1995 – 2^{nd} edition.

[115] M. Sebag. Distance induction in FOL. In *Proceedings of ILP-97*, pages 264–272. Springer Verlag, LNCS 1297, 1997.

[116] M. Sebag. Genetic Programming applied to model identification. In L. Györfi, editor, *Principles of non parametric learning*, pages 271–335. Springer Wien New York, 2002.

[117] M. Sebag, J. Azé, and N. Lucas. Impact Studies and Sensitivity Analysis in Medical Data Mining with ROC-based Genetic Learning. In *Proceedings of IEEE International Conference on Data Mining, ICDM03*, pages 637–640, 2003.

[118] M. Sebag, J. Azé, and N. Lucas. Roc-based evolutionary learning: Application to medical data mining. In *Artificial Evolution'03*, pages 384–396. LNCS 2936, Springer Verlag, 2004.

[119] M. Sebag and A. Ducoulombier. Extending population-based incremental learning to continuous search spaces. In Th. Bäck, G. Eiben, M. Schoenauer, and H.-P. Schwefel, editors, *Proceedings of the* 5^{th} *Conference on Parallel Problems Solving from Nature*, pages 418–427. Springer Verlag, 1998.

[120] M. Sebag and C. Rouveirol. Resource-bounded relational reasoning: Induction and deduction through stochastic matching. *Machine Learning*, 38:41–62, 2000.

[121] M. Sebag and M. Schoenauer. Controlling crossover through inductive learning. In Y. Davidor, H.-P. Schwefel, and R. Manner, editors, *Proceedings of the* 3^{rd} *Conference on Parallel Problems Solving from Nature*, pages 209–218. Springer-Verlag, LNCS 866, 1994.

[122] M. Sebag and M. Schoenauer. A rule-based similarity measure. In S. Wess, K.-D. Althoff, and M. M. Richter, editors, *Topics in Case-Based Reasoning*, volume 837 of *LNCS*, pages 119–130. Springer Verlag, 1994.

[123] M. Sebag, M. Schoenauer, and M. Peyral. Revisiting the memory of evolution. *Fundamenta Informaticae*, 38:1–39, 1998.

[124] R.E. Shapire, Y. Freund, P.Bartlett, and W.S. Lee. Boosting the margin: a new explanation for the effectiveness of voting methods. In *Proceedings of the 14^{th} International Conference on Machine Learning*, pages 322–330. Morgan Kaufmann, 1997.

[125] A. Srinivasan, R.D. King, and D.W. Bristol. An assessment of submissions made to the Predictive Toxicology Evaluation Challenge. In *Proceedings of Int. Joint Conf. on Artificial Intelligence, IJCAI'99*, pages 270–275. Morgan Kaufmann, 1999.

[126] H. Stoppiglia, G. Dreyfus, R. Dubois, and Y. Oussar. Ranking a random feature for variable and feature selection. *Journal of Machine Learning Research*, 3:1399–1414, 2003.

[127] Vivek Pandey (Anne Auger supervisor). Comparing Estimation of Distribution Algorithms and Evolution Strategies. Summer Internship report, INRIA, July 2002.

[128] A. Termier, M.-C. Rousset, and M. Sebag. Treefinder: a first step towards xml data mining. In *Proceedings of IEEE International Conference on Data Mining, ICDM02*, pages 450–457, 2002.

[129] A. Termier, M. Sebag, and M.-C. Rousset. Combining statistics and semantics for word and document clustering. In *Ontology Learning Workshop, IJCAI'01*, pages 49–54, 2001.

[130] E. Tuci, I. Harvey, and M. Quinn. Evolving integrated controllers for autonomous learning robots using dynamic neural networks. In B. Hallam et al., editor, *Proc. of SAB'02*. MIT Press, 2002.

[131] H. Vafaie and K.A. DeJong. Improving the performance of rule induction system using genetic algorithms. In *Proceedings of* 1^{st} *International Workshop on MultiStrategy Learning*, pages 305–315. Harpers Ferry, 1991.

[132] P. Valentin and G. Guiochon. Propagation of finite concentration in gas chromatography. *Separation Science*, 10:245–305, 1976.

[133] L.G. Valiant. A theory of the learnable. *Communication of the ACM*, 27:1134–1142, 1984.

[134] V. N. Vapnik. *The Nature of Statistical Learning.* Springer Verlag, 1995.

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

[136] J. Weston. Leave-one-out support vector machines. In *IJCAI99*, pages 727–733. Morgan Kaufmann, 1999.

[137] D.H. Wolpert and W.G. Macready. No Free Lunch Theorems for Search. Technical report, Santa Fe Institute, 1995.

[138] S. Wrobel. Scalability, search and sampling: From smart algorithms to active discovery. In L. de Raedt and P. Flach, editors, *Proceedings of the 12^{th} European Conference on Machine Learning*, volume LNAI 2167, page 615. Springer Verlag, 2001.

[139] M.J. Zaki. Efficiently mining frequent trees in a forest. In *Proc. of Knowledge Discovery from Databases*, pages 71–80, 2002.