



INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Team SequeL

Sequential Learning

Futurs

THEME COG

Activity
R *report*

2006

Table of contents

1. Team	1
2. Overall Objectives	1
2.1. Overall Objectives	1
3. Scientific Foundations	2
3.1. Statistical learning	2
3.1.1. Kernel methods for non parametric function approximation	3
3.1.2. Monte-Carlo methods	3
3.1.3. Non parametric Bayesian models	3
4. Application Domains	4
4.1. Applications	4
4.1.1. Optimal control of a depollution system	4
4.1.2. Audio signal analysis and processing	4
4.1.2.1. Sequential harmonic models	4
4.1.2.2. Kernel methods	4
4.1.3. Customer behavior modeling	4
4.1.4. Sequential learning of sensors localization	5
5. Software	5
5.1. Software	5
6. New Results	5
6.1. Reinforcement learning	5
6.1.1. Theoretical foundations of reinforcement learning	6
6.1.2. Non parametric function approximation for RL: the Equi-Gradient TD algorithm	6
6.1.3. Exploration vs. exploitation: the game of Go/Monte-Carlo methods	6
6.1.3.1. In the game of Go	6
6.1.3.2. Foundations	7
6.2. Non parametric Bayesian modeling	7
6.2.1. Functional clustering	7
6.2.2. Data fusion	7
6.2.3. Joint segmentation of piecewise constant autoregressive processes	7
6.2.4. Sequential selection of stochastic models	7
7. Contracts and Grants with Industry	8
7.1. Contracts and Grants with Industry	8
8. Other Grants and Activities	8
8.1. National activities	8
8.1.1. ANR Kernsig	8
8.1.2. Projet ANR ADAPT ^{MC}	8
8.1.3. Projet ACI “Complex Systems in Social and Human Sciences”	8
8.2. International activities	8
8.2.1. KRL-2006 workshop	9
9. Dissemination	9
9.1. Scientific Community animation	9
9.2. Teaching	9
10. Bibliography	9

1. Team

SEQUEL is a joint project with the LIFL (UMR 8022 of CNRS and University of Lille 1 and University of Lille 3) and the LAGIS (UMR 8021 of the École Centrale de Lille 1 and the University of Lille 1).

Head of the team

Philippe Preux [Professor, Université de Lille, HdR]

Vice-Head of the team

Rémi Munos [Research Director (DR), Inria, HdR]

Project assistant

Véronique Couvreur [Secretary (SAR) Inria, shared by 3 projects]

Staff member

Manuel Davy [Researcher (CR) CNRS, HdR]

Emmanuel Duflos [Professor, École Centrale de Lille, HdR]

Philippe Vanheeghe [Professor, École Centrale de Lille, HdR]

Rémi Coulom [Assistant professor, Université de Lille 3]

Jérémie Mary [Assistant professor, Université de Lille 3]

Ph.D student

Pierre-Arnaud Coquelin [École Polytechnique]

Robin Jaulmes [DGA]

Manuel Loth [INRIA-Région Nord-pas-de-calais Grant]

Post-Doctoral fellow

Thomas Bréhard [INRIA]

Stéphane Rossignol [CNRS & NTI]

2. Overall Objectives

2.1. Overall Objectives

SEQUEL means “Sequential Learning”. As such, SEQUEL focuses on the task of learning in artificial systems (either hardware, or software) that gathers information along time. Such systems are named (*learning*) *agents* in the following¹. These data may be used to estimate some parameters of a model, which in turn, may be used for selecting actions in order to perform some long-term optimization task.

For the purpose of model building, the agent needs to gather information collected so far in some compact representation and combine it to newly available data.

The acquired data may result from an observation process of an agent in interaction with its environment (the data thus represent a perception). This is the case when the agent makes decisions (in order to fulfill a certain goal) that impact the environment thus the observation process itself.

Hence, in SEQUEL, the term **sequential** refers to two aspects:

- The **sequential acquisition of data**, from which a model is learned (supervised and non supervised learning),
- the **sequential decision making task**, based on the learned model (reinforcement learning).

¹we might also have called them “learning machines”, since that’s what these agents are here.

We exemplify these various problems:

Supervised learning tasks deal with the prediction of some response given a certain set of observations of input variables and responses. New sample points keep on being observed.

Unsupervised learning tasks deal with clustering objects, these latter making a flow of objects. The (unknown) number of clusters typically evolves during time, as new objects are observed.

Reinforcement learning tasks deal with the control (a policy) of some system which has to be optimized (see [63]). We do not assume the availability of a model of the system to be controlled.

In all these cases, we assume that the process can be considered stationary for at least a certain amount of time, and slowly evolving.

We wish to have any-time algorithms, that is, at any moment, a prediction may be required/an action may be selected making full use, and hopefully, the best use, of the experience already gathered by the learning agent.

The perception of the environment by the learning agent (using its sensors) is generally not the best one to make a prediction, nor to take a decision (we deal with Partially Observable Markov Decision Problem). So, the perception has to be mapped in some way to a better, and relevant, state (or input) space.

Finally, an important issue of prediction regards its evaluation: how wrong may we be when we perform a prediction? For real systems to be controlled, this issue can not be simply left unanswered.

To sum-up, in SEQUEL, the main issues regard:

- the learning of a model: we focus on models than map some input space \mathbb{R}^P to \mathbb{R} ,
- the observation to state mapping,
- the choice of the action to perform (in the case of sequential decision problem),
- the bounding of the performance,
- the implementation of usable algorithms,

all that being understood in a *sequential* framework.

3. Scientific Foundations

3.1. Statistical learning

Keywords: *Bayesian formalism, Monte-Carlo methods, kernel methods.*

machine learning Machine learning refers to a system capable of the autonomous acquisition and integration of knowledge. This capacity to learn from experience, analytical observation, and other means, results in a system that can continuously self-improve and thereby offer increased efficiency and effectiveness. (source: <http://www.aaai.org/AITopics/html/machine.html>)

statistical learning An approach to machine intelligence which is based on statistical modeling of data. With a statistical model in hand, one applies probability theory and decision theory to get an algorithm. This is opposed to using training data merely to select among different algorithms or using heuristics/"common sense" to design an algorithm. (source: <http://www.cs.wisc.edu/~hzhang/glossary.html>)

kernel method Generally speaking, a kernel function is a function that maps a couple of points to a real value. Typically, this value is a measure of dissimilarity between the two points. Assuming a few properties on it, the kernel function implicitly defines a dot product in some function space. This very nice formal property as well as a bunch of others have ensured a strong appeal for these methods in the last 10 years in the field of function approximation. Many classical algorithms have been "kernelized", that is, restated in a much more general way than their original formulation. Kernels also implicitly induce the representation of data in a certain "suitable" space where the problem to solve (classification, regression, ...) is expected to be simpler (non-linearity turns to linearity).

The fundamental tools used in SEQUEL come from the field of statistical learning [61]. We briefly present the most important for us to date, namely, kernel-based non parametric function approximation, sequential Monte-Carlo methods, and non parametric Bayesian models.

3.1.1. Kernel methods for non parametric function approximation

In SEQUEL, the model to be learned is a real-valued function defined in a multi-dimension space.

Many methods have been proposed for this purpose. We are looking for suitable ones to cope with the problems we wish to solve. In reinforcement learning, the value function may have areas where the gradient is large; these are areas where the approximation is difficult, while these are also the areas where the accuracy of the approximation should be maximal to obtain a good policy (and where, otherwise, a bad choice of action may imply catastrophic consequences).

For the moment, we consider non parametric methods since they do not make any assumptions about the function to learn. Locally weighted regression have yielded efficient methods to learn a policy in reinforcement learning, as well as good performance in regression settings. The kernelized version gives us a wide ability to handle sample points and combine them to obtain the approximation. To keep computation times of practical interest, a sparse representation is sought.

We currently devote a lot of efforts to LARS-like approximators [58], that we have fitted into the reinforcement learning framework [54], [49], [50].

3.1.2. Monte-Carlo methods

Sequential Monte-Carlo (or particle filtering, see [56]) methods are currently used for various purposes in SEQUEL:

- the estimation of the state of the agent given its current observation as well as its history;
- the estimation of parameters of a model.

3.1.3. Non parametric Bayesian models

Numerous problems in signal processing may be solved efficiently by way of a Bayesian approach. The use of Monte-Carlo methods let us handle non linear, as well as non Gaussian problems. In their standard form, they require the formulation of densities of probability in their parametric form. For instance, it is a common usage to use Gaussian likelihood, because it is handy.

However, in some applications such as Bayesian filtering, or blind deconvolution, the choice of a parametric form of the density of the noise is often arbitrary. If this choice is wrong, it may also have dramatic consequences on the estimation.

To overcome this shortcoming, non parametric methods provide an other approach to this problem. In particular, mixtures of Dirichlet processes [59] provide a very powerful formalism.

Mixtures of Dirichlet Processes are an extension of finite mixture models. Given a mixture density $f(\mathbf{x}|\theta)$, and $G(d\theta) = \sum_{k=1}^{\infty} \omega_k \delta_{U_k}(d\theta)$, a Dirichlet process ². Then, we define a mixture of Dirichlet processes as:

$$F(\mathbf{x}) = \int_{\Theta} f(\mathbf{x}|\theta)G(d\theta) = \sum_{k=1}^{\infty} \omega_k f(\mathbf{x}|U_k)$$

A mixture of Dirichlet processes is fully parameterized by the mixture density, as well as the parameters of G , that is G_0 and α .

The class of densities that may be written as a mixture of Dirichlet processes is very wide, so that these are really fit to very large amount of applications.

²A Dirichlet process is a random distribution almost surely discrete, where the centroids U_k are distributed along a *base distribution* $G_0(\cdot)$, and where weights follow a certain *stick breaking* law with parameter α [62].

Given a set of observations, the estimation of the parameters of a mixture of Dirichlet processes is performed by way of a *Monte Carlo Markov Chain (MCMC)* algorithm.

4. Application Domains

4.1. Applications

Keywords: *automatic transcription of music, automatic transcription of speech, civil engineering, customer behavior modeling, environment, multimedia, sensor localization, transportation systems.*

SEQUEL aims at solving problems of prediction, as well as problems of optimal control. As such, the application domains are very numerous. Furthermore, we consider that tackling real applications is necessary for us as a feedback for our fundamental researches, as well as a guide towards technological locks, and their resolution. The applications are being studied with industrial partners (such as function prediction), as well as with academic partners (such as in the control of depollution systems).

4.1.1. *Optimal control of a depollution system*

The research concerns the control of waste-water treatment by an anaerobic digestion process. Organic waste-water is treated in a biological reactor by means of an anaerobic digestion process, which produces bio-gas (methane) that can be used to power electricity generators. The digestion of the organic material is done by an appropriate mixing of (mainly) two different species of bacteria, each doing a different job at a different step in the digestion process. Maintaining suitable conditions in the digester is essential for the viability of the bacterial population. The problem of designing a good controller is crucial. A mandatory condition is to stabilize the system and maintain the bacterial populations alive. Under these constraints, an additional objective is to optimize the production of bio-gas.

From the scientific standpoint, this problem is formalized as a partially observable Markov decision problem (POMDP), where the state dynamics is not perfectly known: there are several available models for the dynamics, but most of the parameters of these models are unknown. Besides, the observation process is very poor. We aim at designing an adaptive policy that is both secure and which optimizes the methane production.

This work is being held in collaboration with the Naskeo Environment spin-off (Paris), the INRA laboratory in Narbonne (France), and the COMORE team with INRIA-Sophia-Antipolis. The work is currently only beginning.

4.1.2. *Audio signal analysis and processing*

4.1.2.1. *Sequential harmonic models*

This work concerns the automatic transcription of music, as well as the detection of sonor effects in an urban environment.

To estimate audio signals, we have used sequential sinusoidal plus noise models. A particular filtering algorithm has been developed to estimate the parameters. An emphasis has been put on the reduction of the computation times (see [3]).

4.1.2.2. *Kernel methods*

We have worked on the speaker segmentation problem, on speech signals. We have used a kernel method originally developed by F. Desobry [57]. We obtained almost state-of-the-art results with this method (see [44], [45]). Further experimental assessments are on their way.

In collaboration with France Telecom, we have also worked on the automatic transcription of speech by using kernel methods (Stéphane Rossignol's post-doc fellowship).

4.1.3. *Customer behavior modeling*

Basically speaking, machine learning aims at predicting a function (rather than a single value). A function formalizes the behavior of a "system", for instance, the behavior of customers in a shop (either a material shop, or a virtual shop on the web).

Auchan is a major international group which operates more than 150 Hypermarkets worldwide. One of its crucial issues is to be able to predict many variables, such as the number of customers reaching the cashiers at a given time of the day, the number of breads to cook, etc. For each day, this is a functional prediction problem, which may be seen as non-stationary, because the customer habits evolve.

SEQUEL members and Auchan are currently investigating this problem by extensively using past events. State-of-the-art inference methods have been implemented with excellent performance (not described here for confidentiality reasons). To date, we have improved the prediction accuracy of about 25% percent. Due to these very promising results, a new research contract is currently under discussion between Auchan and SEQUEL, to extend the use of sequential learning approaches in commercial industries.

4.1.4. Sequential learning of sensors localization

We have worked in collaboration with Prof Carl Haas of the University of Waterloo (Canada) on a problem appearing in civil engineering: how can we automatically localize the building materials on a construction site. This is a real problem because a lot of time (hence of money) is lost to look for these materials that have often been moved away. The proposed solution is to equip each piece with a RFID and each person working on the construction site with a RFID receiver, a GPS for the localization and a transmitter. We then learn sequentially the position of the pieces using the incoming detection information sent automatically by the transmitter to a central processor when the workforces walk near these pieces and detect them. RFID systems and localization systems as GPS allow to treat such a problem in the more general context of randomly distributed communication nodes localization. When the nodes are moving the problem is still more complicated. Our work shows how the Transferable Belief Model can be used to learn the position of the communication nodes and to detect potential movements. This study also shows how to deal with the computation.

This work has also been applied for land vehicle localization. The vehicle is equipped with three sensors, including a GPS sensor.

This work has been carried out in collaboration with François Caron. See [15], [37].

5. Software

5.1. Software

For the moment, we do not have developed yet a software from which general users could benefit. We plan to start the development of a software platform in the coming year. In its first stage, the goal would be to integrate various programs we have designed separately.

The interested reader should check out our website for any new information regarding software development <http://www.grappa.univ-lille3.fr/SequeL>.

6. New Results

6.1. Reinforcement learning

Keywords: *LARS, L_p -norm, Monte-Carlo estimation, dynamic programming, exploration-exploitation trade-off, non parametric function approximation, performance bound, policy search, value function approximation, variance reduction, viability problem.*

Participants: Rémi Coulom, Manuel Davy, Manuel Loth, Philippe Preux, Pierre-Arnaud Coquelin, Rémi Munos, Robin Jaulmes.

6.1.1. Theoretical foundations of reinforcement learning

We have worked on several aspects of reinforcement learning and optimal control, including the use of function approximation to represent the value function or the policy. We have worked in collaboration mainly with Csaba Szepesvári (University of Alberta, Canada), András Antos (Hungarian Academy of Sciences), Jean-Yves Audibert (CERTIS, Ecole des Ponts et Chaussées), Guillaume Deffuant and Sophie Martin (Cemagref, Clermont-Ferrand), Hasnaa Zidani (ENSTA), Olivier Bokanowski (Paris VII), Olivier Teytaud (LRI, Orsay). This work can be summarized as follows:

- **Establishing links between statistical learning and reinforcement learning.** Performance bounds on the policies deduced by approximate dynamic programming methods (such as approximate value iteration, approximate policy iteration) when using sampling devices are established in terms of the capacity (using VC dimension, covering numbers) of the function space considered in the approximations. See [8], [29], [33].
- **Analysis of dynamic programming using L_p -norms.** This work extends usual analysis in L_∞ -norm to L_p -norms, which opens promising new directions towards an analysis of dynamic programming and reinforcement learning combined with function approximation. See [28], [27].
- **Adaptive variance reduction method for Monte-Carlo estimation in Markov chains.** This has been applied to fast estimation of value functions in Markov decision processes and their gradient (with respect to control parameters). See [5].
- **Policy gradient estimation in continuous time.** This method allows to search directly for a locally optimal controller in a class of parameterized policies, in the case of continuous-time state-dynamics. See [6]. An application of this method to a control problem in finance is described in [11].
- **Analysis of the exploration-exploitation tradeoff using variance estimate.** We investigate the multi-armed bandit framework using new deviation inequalities that takes into account the variance estimate. This results in a great sharpening of the regret bounds. See [10].
- **Numerical approximation of viability problems.** We use several ideas from the ultra-bee schemes used for transport equation with discontinuous solutions and the dynamic programming approach combined with function approximation to approximate the viability kernel of a viability problem. See [12], [39].

6.1.2. Non parametric function approximation for RL: the Equi-Gradient TD algorithm

As well as in many machine learning situations, a key component of a reinforcement learning algorithm lies in the approximation of a real function, generally in a high dimension space. Even though various approaches have been studied for the last 15 years in the reinforcement learning framework, a scalable, workable, ... solution is still awaited. We have considered the use of non parametric function approximator, and in particular, the LARS algorithm introduced in statistics in [58] and later kernelized [60]. We have adapted the algorithm to the RL problem setting, and actually completely rethought and re-interpreted it, to approximate the value function, providing the so-called “equi-gradient descent” algorithm introduced in [54], [49]. This work is currently going on [50]. We have recently proposed a unified view of many algorithms (TD, residual TD, iLSTD, LSTD, LSPE and our equi-gradient TD algorithm), showing that they all fit into a single, and simple, formalism [55]. Experiments demonstrate the interest of this new approach.

6.1.3. Exploration vs. exploitation: the game of Go/Monte-Carlo methods

6.1.3.1. In the game of Go

The research on the game of Go conducted in 2006 followed the Master thesis of Guillaume Chaslot. Guillaume introduced us to Monte-Carlo techniques applied to move selection. When he left after defending his master in 2005, a new Go-playing program (Crazy Stone) was developed by Rémi Coulom. This led to the design of original and efficient methods for combining Monte-Carlo evaluation with tree search. Crazy Stone won the 9x9 Go tournament at the 11th Computer Olympiad, and a paper describing its algorithm was presented in the “5th International Conference on Computer and Games” [40], [41].

6.1.3.2. Foundations

We have also investigated the exploration vs. exploitation trade-off more theoretically by way of multi-arm bandit models. This has already been discussed in the theoretical foundations section. See [10].

6.2. Non parametric Bayesian modeling

Keywords: *clustering, land vehicle localization, mixture of Dirichlet processes, model selection, rupture detection, state estimation.*

Participants: Manuel Davy, Emmanuel Duflos, Philippe Vanheeghe.

6.2.1. Functional clustering

With E. Jackson (PhD student) and A. Doucet (Professor at the U. of British Columbia), we have investigated Bayesian functional clustering. Given observations obtained by sampling different functions at random locations (and different from one function to an other), we have developed an algorithm that clusters these functions into coherent groups. For instance, each observation may be a signal (in one or more dimensions), such that the sampling instants are different from one signal to an other. We have then modeled the underlying signals by using Gaussian processes and the clustering itself is performed by using a Dirichlet mixture process. This study is currently under submission [46]. The target applications concerns the clustering of expression data of messenger-RNA, as well as sampling data originating from geostatistics.

6.2.2. Data fusion

We consider the problem of sequentially estimating the state \mathbf{x}_t of a system. In the Bayesian framework, one writes down a transition equation $\mathbf{x}_t \sim p(\mathbf{x}_t|x_{t-1})$ and an observation equation $\mathbf{z}_t \sim p(\mathbf{z}_t|\mathbf{x}_t)$. This may be written as:

$$\begin{aligned}\mathbf{x}_t &= g(\mathbf{x}_{t-1}) + \mathbf{v}_t \\ \mathbf{z}_t &= h(\mathbf{x}_t) + \mathbf{w}_t\end{aligned}$$

where \mathbf{v}_t and \mathbf{w}_t are noise of given distributions. However, it may be very difficult to forethink a parametric representation for $p(\mathbf{v}_t)$ and $p(\mathbf{w}_t)$. We have considered the case where the two densities are given by a mixture of Dirichlet processes. This let us estimate them in the same time as we estimate the state \mathbf{x}_t . With regards to the state of the art, our contribution concerns the modeling, as well as the Monte-Carlo approach for estimation that as been used [34].

6.2.3. Joint segmentation of piecewise constant autoregressive processes

Certain electrical devices include a set of electric cables which, under certain circumstances, may give rise to electrical arcs (a typical example is the command circuitry of an airplane). This is basically a problem of detecting ruptures in a multichannel environment.

We have designed a Bayesian algorithm to detect these ruptures which models the signals on each cable individually, by an autoregressive model and that assumes a correlation between the instants of rupture in the different cables. See [43], [21].

6.2.4. Sequential selection of stochastic models

Our main contribution takes place in the frame of the PhD work of François Caron who has defended his PhD on the 10th of November. In this work we have been interested in the addition of uncertainty in hidden Markov models. In this work a multisensor system is considered where each sensor may switch between several states of work. A new jump model has been developped based on an original modeling of the transition probabilities of the sensors reliability probabilities using Dirichlet Processes. An algorithm based on a particle filter with efficient importance law has also been developped. This algorithm has been applied with success to the positioning of a land vehicle equipped with three sensors, one of the sensors being a GPS. It is shown that the proposed algorithm allows the detection and the rejection of GPS data corrupted with multipathes, improving therefore the estimation process.

In such systems and application, it is generally assumed that the state noise and the measurement noise is gaussian and stationary so as to be able to use Kalman filtering. In real applications this isn't often the case. A possible solution developed in this work consists in estimating the density probability function (pdf) using Dirichlet processes mixture modelling. First, the case of linear models has been addressed and MCMC and particle filter algorithms have been developed. Then, the case of the estimation of pdf in non linear models has been addressed. For that purpose, time-varying Dirichlet processes have been defined for the on line estimation of time-varying pdf.

This work has been carried out in collaboration with Manuel Davy. It has been published in [34], [16], [14], [36], [38], [35][2].

7. Contracts and Grants with Industry

7.1. Contracts and Grants with Industry

As an INRIA team, SEQUEL has not yet signed any contract by way of the INRIA. However, various works in 2006 have been done under contract, such as France Telecom, Auchan, and Neutrik Test Instruments (NTI, Liechtenstein).

As an INRIA team, we have had serious contacts with a few industrial partners. These contacts are currently under development. To date, these contacts concern EADS Space-Transportation, ATOS, Naskeo-Environment, and Auchan.

8. Other Grants and Activities

8.1. National activities

8.1.1. ANR Kernsig

Participants: Manuel Davy, Thomas Bréhard.

This project is headed by Prof. S. Canu with the INSA-Rouen. It deals with the study of kernel methods for signal processing. Thomas Bréhard has been recruited on a post-doc fellowship on this contract.

8.1.2. *Projet ANR ADAPT'MC*

Participant: Rémi Munos.

This ANR deals with adaptive Monte Carlo Methods, in collaboration with the teams from the École des Ponts (B. Lapeyre, B. Jourdain), the Université de Dauphine (C. Robert, A. Guillin), the École Nationale Supérieure des Télécommunications (E. Moulines, O. Cappé) and the Centre de Mathématiques Appliquées (R. Douc). This is a 3 years project 2006-2008.

8.1.3. *Projet ACI "Complex Systems in Social and Human Sciences"*

Participant: Preux Philippe.

This is the last year of this action (2004-2006). The project, named DYNAPP for "Dynamics of Learning", aims at modeling the learning of behavior of an animal, at a functional level, using concepts from system dynamics. We use reinforcement learning as a model and as a way to simulate this learning. This work is being done with a team of experimental psychologists (U. Lille 3).

8.2. International activities

Rémi Munos is co-supervisor of Yuval Tassa's Phd at the ICNC, Hebrew University of Jerusalem, Israel.

8.2.1. KRL-2006 workshop

Participants: Manuel Davy, Rémi Munos, Philippe Preux.

Along with Csaba Szepesvári (U. Alberta), we have organized a workshop on Kernel Machines and Reinforcement Learning at the International Conference on machine Learning, in Pittsburgh, in June 2006 (<http://www.grappa.univ-lille3.fr/ppreux/krl/>). The workshop has gathered 30 people during the day.

9. Dissemination

9.1. Scientific Community animation

- Manuel Davy is associate editor for the IEEE Transactions on Signal Processing review. He has also reviewed papers for IEEE Trans. on Signal Processing, IEEE Signal Processing Letter, Speech communications, Signal Processing, IEEE Trans. on Circuits and Systems I.
- Along with A. Klapuri (Tampere University of Technology, Finland), Manuel Davy has coordinated a book on “signal processing methods for music transcription” [1]
- Rémi Munos is member of Journal of Machine Learning Research, Artificial Intelligence Journal, Revue d’Intelligence Artificielle. He has been a member of the PC of the 2006 conferences Neural Information Processing Systems, International Conference on Machine Learning, Conférence Francophone sur l’Apprentissage Automatique.
- Along with Csaba Szepesvári (U. Alberta), Philippe Preux, Rémi Munos and Manuel Davy co-organized the **workshop on Kernel Machines and Reinforcement Learning**, International Conference on Machine Learning, 2006. See: <http://www.icml2006.org/icml2006/workshops.html>
- Rémi Munos is Co-chair of the **IEEE International Symposium on Approximate Dynamic Programming and Reinforcement Learning**, 2007. See : <http://liu.ece.uic.edu/dliu/ADPRL07/>
- Philippe Preux is a member of the program committee of the **IEEE International Symposium on Approximate Dynamic Programming and Reinforcement Learning**, 2007.
- Rémi Coulom has been a reviewer for the **Computer & Games** 2006 conference.
- Philippe Preux, Rémi Coulom, together with Samuel Delepoulle, have prepared a special issue of the “Revue d’Intelligence Artificielle” on Markov Decision Processes (to appear in 2007).
- Philippe Preux has served as workshop co-chairman of the EGC 2006 (french-speaking yearly conference on data mining).

9.2. Teaching

We list the courses that are related to the research activities in SEQUEL that happened in 2006.

- Rémi Munos teaches a class in reinforcement learning in the M2 “Mathematics-Vision-Learning” (MVA) at the ENS-Cachan; he also teaches a cognitive science class in M1 at the EHESS (Paris).
- Philippe Preux teaches in the M2 of computer science at the University of Lille a class in reinforcement learning.

Otherwise, each of the 5 professors and assistant professors of the SEQUEL team teaches 192 hours per year, mostly at master level. Taught classes include machine learning, data mining, and signal processing classes.

10. Bibliography

Major publications by the team in recent years

- [1] A. Klapuri, M. Davy (editors). *Signal Processing Methods for Music Transcription*, Springer, New York, 2006.

- [2] F. CARON, M. DAVY, E. DUFLOS, P. VANHEEGHE. *Particle Filtering for Multisensor Data Fusion with Switching Observation Models. Application to Land Vehicle Positioning*, in "IEEE transactions on Signal Processing", 2006.
- [3] C. DUBOIS, M. DAVY. *Joint Detection and Tracking of Time-Varying Harmonic Components: a Flexible Bayesian Approach*, in "IEEE transactions on Speech, Audio and Language Processing", 2006.
- [4] E. GOBET, R. MUNOS. *Sensitivity analysis using Itô Malliavin calculus and martingales. Application to stochastic optimal control*, in "SIAM journal on Control and Optimization", vol. 43(5), 2005, p. 1676-1713.
- [5] R. MUNOS. *Geometric variance reduction in Markov chains. Application to value function and gradient estimation*, in "Journal of Machine Learning Research", vol. 7, 2006, p. 413-427.
- [6] R. MUNOS. *Policy gradient in continuous time*, in "Journal of Machine Learning Research", vol. 7, 2006, p. 771-791.

Year Publications

Books and Monographs

- [7] A. KLAURI, M. DAVY (editors). *Signal Processing Methods for Music Transcription*, Springer, New York, 2006.

Articles in refereed journals and book chapters

- [8] A. ANTOS, C. SZEPEŠVÁRI, R. MUNOS. *Learning near-optimal policies with Bellman-residual minimization based fitted policy iteration and a single sample path*, in "Conference on Learning Theory", 2006.
- [9] A. ANTOS, C. SZEPEŠVÁRI, R. MUNOS. *Learning near-optimal policies with Bellman-residual minimization based fitted policy iteration and a single sample path*, in "submitted to Machine Learning Journal", 2006.
- [10] J.-Y. AUDIBERT, R. MUNOS, C. SZEPEŠVÁRI. *Use of variance estimation in the multi-armed bandit problem*, in "NIPS workshop On-line Trading of Exploration and Exploitation", 2006.
- [11] C. BARRERA-ESTEVE, F. BERGERET, E. GOBET, A. MEZIOU, R. MUNOS, D. REBOUL-SALZE. *Numerical methods for the pricing of Swing options: a stochastic control approach*, in "Methodology and Computing in Applied Probability", to appear, 2007.
- [12] O. BOKANOWSKI, S. MARTIN, R. MUNOS, H. ZIDANI. *An anti-diffusive scheme for viability problems*, in "Applied Numerical Mathematics, special issue on Numerical methods for viscosity solutions and applications", vol. 45-9, 2006, p. 1147-1162.
- [13] F. CARON, M. DAVY, E. DUFLOS, P. VANHEEGHE. *Particle Filtering for Multisensor Data Fusion with Switching Observation Models. Application to Land Vehicle Positioning*, in "IEEE transactions on Signal Processing", 2006.
- [14] F. CARON, M. DAVY, E. DUFLOS, P. VANHEEGHE. *Particle Filtering for Multisensor Data Fusion with Switching Observation Models*, in "IEEE Transactions on Signal Processing", to appear, 2007.

-
- [15] F. CARON, S. RAZAVI, J. SONG, E. DUFLOS, P. VANHEEGHE, C. CALDAS, C. HAAS. *Particle Filtering for Multisensor Data Fusion with Switching Observation Models*, in "Locating sensor nodes on construction projects", to appear, 2007.
- [16] F. CARON, B. RISTIC, E. DUFLOS, P. VANHEEGHE. *Least Committed Basic Belief Density Induced by a Multivariate Gaussian: Formulation with Applications*, in "International Journal of Approximate Reasoning", to appear, 2007.
- [17] M. DAVY. *Signal Processing Methods for Music Transcription*, A. KLAPURI, M. DAVY (editors). , chap. An introduction to Statistical Signal Processing Methods, Springer, New York, 2006.
- [18] M. DAVY. *Signal Processing Methods for Music Transcription*, A. KLAPURI, M. DAVY (editors). , chap. Multiple F0 estimation methods based on generative models, Springer, New York, 2006.
- [19] M. DAVY, F. DESOBRY, A. GRETTON, C. DONCARLI. *An online Support Vector Machine for Abnormal Events Detection*, in "Signal Processing", vol. 86, n^o 8, August 2006, p. 2009–2025.
- [20] M. DAVY, S. GODSILL, J. IDIER. *Bayesian Analysis of Ployphonic Western Tonal Music*, in "Journal of the Acoustical Society of America", vol. 119, n^o 4, April 2006, p. 2498-2517.
- [21] N. DOBIGEON, J.-Y. TOURNERET, M. DAVY. *Joint segmentation of piecewise constant autoregressive processes by using a hierarchical model and a Bayesian sampling approach*, in "IEEE transactions on Signal Processing", to appear, 2007.
- [22] C. DUBOIS, M. DAVY. *Joint Detection and Tracking of Time-Varying Harmonic Components: a Flexible Bayesian Approach*, in "IEEE transactions on Speech, Audio and Language Processing", 2006.
- [23] E. GOBET, R. MUNOS. *Sensitivity analysis using Itô Malliavin calculus and martingales. Application to stochastic optimal control*, in "SIAM journal on Control and Optimization", vol. 43(5), 2005, p. 1676-1713.
- [24] P. HERRERA-BOYER, A. KLAPURI, M. DAVY. *Signal Processing Methods for Music Transcription*, A. KLAPURI, M. DAVY (editors). , chap. Automatic Classification of Pitched Musical Instrument Sounds, Springer, New York, 2006.
- [25] R. MUNOS. *Geometric variance reduction in Markov chains. Application to value function and gradient estimation*, in "Journal of Machine Learning Research", vol. 7, 2006, p. 413-427.
- [26] R. MUNOS. *Policy gradient in continuous time*, in "Journal of Machine Learning Research", vol. 7, 2006, p. 771-791.
- [27] R. MUNOS. *Analyse en norme L_p de l'algorithme d'itérations sur les valeurs avec approximations*, in "Revue d'Intelligence Artificielle", to appear, vol. 21, 2007, p. 55-76.
- [28] R. MUNOS. *Performance Bounds in L_p -Norm for Approximate Value Iteration*, in "SIAM Journal on Control and Optimization", to appear, 2007.
- [29] R. MUNOS, C. SZEPESVÁRI. *Finite time bounds for sampling based fitted value iteration*, in "Submitted to Journal of Machine Learning Research", 2006.

- [30] D. POTIN, E. DUFLOS, P. VANHEEGHE, M. DAVY. *An abrupt change detection algorithm for buried landmines localization*, in "IEEE Transactions on Geosciences and Remote Sensing", n^o 2, 2006, p. 220-272.
- [31] D. POTIN, E. DUFLOS, P. VANHEEGHE. *Landmines Ground Penetrating Radar Signal Enhancement by Digital Filtering*, in "IEEE Transactions on Geosciences and Remote Sensing", n^o 1, 2006, p. 2393-2406.
- [32] D. POTIN, P. VANHEEGHE, E. DUFLOS, M. DAVY. *An abrupt change detection algorithm for buried landmines localization*, in "IEEE trans. on Geophysics and Remote Sensing", vol. 44, n^o 2, February 2006, p. 260-272.

Publications in Conferences and Workshops

- [33] A. ANTOS, C. SZEPESVÁRI, R. MUNOS. *Sampling based fitted Q-iteration: learning with a single trajectory*, in "Proc. of the IEEE International Symposium on Approximate Dynamic Programming and Reinforcement Learning", to appear, 2007.
- [34] F. CARON, M. DAVY, A. DOUCET, E. DUFLOS, P. VANHEEGHE. *Bayesian Inference for Dynamic Models with Dirichlet Process Mixtures*, in "FUSION, Florence, Italy", September 2006.
- [35] F. CARON, M. DAVY, A. DOUCET, E. DUFLOS, P. VANHEEGHE. *Bayesian Inference for Dynamic Models with Dirichlet Process Mixtures*, in "International Conference on Information Fusion (FUSION'06), Florence, Italia, July 10-13, 2006", 2006.
- [36] F. CARON, M. DAVY, E. DUFLOS, P. VANHEEGHE. *Multipath Effects Reduction in Land Vehicle Positioning by Particle Filter*, in "4th CESA Multiconference on Computational Engineering in Systems Applications (CESA 2006) , Beijing, China, October 4-6, 2006", 2006.
- [37] F. CARON, S. RAZAVI, J. SONG, E. DUFLOS, P. VANHEEGHE, C. CALDAS, C. HAAS. *Models for locating RFID nodes*, in "Joint International Conference on Computing and Decision Making in Civil and Building Engineering (ICCCBE'06), Montreal, Canada, June 14-16, 2006", 2006.
- [38] F. CARON, B. RISTIC, E. DUFLOS, P. VANHEEGHE. *Least Committed basic belief density induced by a multivariate Gaussian pdf*, in "International Conference on Information Fusion (FUSION'06), Florence, Italia, July 10-13, 2006", 2006.
- [39] P.-A. COQUELIN, S. MARTIN, R. MUNOS. *A dynamic programming approach to viability problems*, in "Proc. of the IEEE International Symposium on Approximate Dynamic Programming and Reinforcement Learning", to appear, 2007.
- [40] R. COULOM. *Efficient Selectivity and Backup Operators in Monte-Carlo Tree Search*, in "Proceedings of the 5th International Conference on Computer and Games", P. CIANCARINI, H. J. VAN DEN HERIK (editors). , 2006.
- [41] R. COULOM. *Propagation de valeurs et sélectivité dans une recherche arborescente Monte-Carlo*, in "Actes des Journées Francophones Planification, Décision, Apprentissage", F. GARCIA, G. VERFAILLIE (editors). , 2006.

- [42] M. DAVY, F. DESOBRY, S. CANU. *Estimation of Minimum Measure Sets in Reproducing Kernel Hilbert Spaces and Applications*, in "ICASSP, Toulouse, France", May 2006.
- [43] N. DOBIGEON, J.-Y. TOURNERET, M. DAVY. *Joint segmentation of piecewise constant autoregressive processes by using a hierarchical model and a Bayesian Sampling Approach*, in "ICASSP, Toulouse, France", May 2006.
- [44] B. FERGANI, M. DAVY, A. HOUACINE. *Application des machines à vecteurs support mono-classe à l'indexation en locuteurs de documents audio*, in "Journées d'Etude sur la Parole, Dinard, France", June 2006.
- [45] B. FERGANI, M. DAVY, A. HOUACINE. *Unsupervised Speaker Indexing Using One-Class Support Vector Machines*, in "EUSPICO, Florence, Italy", September 2006.
- [46] E. JACKSON, M. DAVY, A. DOUCET, W. FITZGERALD. *Bayesian Unsupervised Signal Classification by Dirichlet Process Mixtures of Gaussian Processes*, in "ICASSP 2007, Honolulu, USA", (submitted), April 2007.
- [47] A. JOHANSEN, A. DOUCET, M. DAVY. *Maximum Likelihood Parameter Estimation for Latent Variable Models using Sequential Monte Carlo*, in "ICASSP, Toulouse, France", May 2006.
- [48] M. LOTH, R. COULOM, M. DAVY, P. PREUX. *Least Angle Temporal Difference Learning: LATD(λ)*, in "Journées Francophones Planification, Décision, Apprentissage, Toulouse, France", May 2006.
- [49] M. LOTH, M. DAVY, R. COULOM, P. PREUX. *Equi-gradient Temporal Difference Learning*, in "ICML Workshop on Kernel machines and Reinforcement Learning, Pittsburgh, USA", June 2006.
- [50] M. LOTH, M. DAVY, P. PREUX. *Sparse temporal difference learning using LASSO*, in "IEEE International Symposium on Approximate Dynamic Programming and Reinforcement Learning", to appear, April 2007.
- [51] F. MONTAGNE, P. PREUX, S. DELEPOULLE. *Introducing interactive help for reinforcement learners*, in "Workshop on planning, learning and monitoring with uncertainty and dynamic worlds, ECAI Workshop", August 2006.

Internal Reports

- [52] E. JACKSON, M. DAVY, W. FITZGERALD. *Unsupervised Classification of Functions using Dirichlet Process Mixtures of Gaussian Processes*, Technical report, n^o CUED/F-INFENG/TR.562, Engineering Department, University of Cambridge, UK, August 2006.

Miscellaneous

- [53] M. DAVY. *Outils d'accès à l'information audio - Méthodes de Monte Carlo et Méthodes à Noyaux*, 2006, Habilitation à Diriger les Recherches, Université des Sciences et Technologies de Lille.
- [54] M. LOTH. *Apprentissage par renforcement et méthodes à noyaux: apprendre la fonction valeur par un réseau de fonctions évolutif*, Technical report, Computer Science Dpt, University of Lille, 2006.
- [55] M. LOTH, P. PREUX, M. DAVY. *A Unified View of TD Algorithms; Introducing Full-Gradient TD and Equi-Gradient Descent TD*, (submitted).

References in notes

- [56] A. DOUCET, N. DE FREITAS, N. GORDON (editors). *Sequential Monte Carlo Methods in Practice*, Springer, 2001.
- [57] F. DESOBRY, M. DAVY, C. DONCARLI. *An online kernel change detection algorithm*, in "IEEE trans. on Signal Processing", vol. 53, n^o 8, August 2005.
- [58] B. EFRON, T. HASTIE, I. JOHNSTONE, R. TIBSHIRANI. *Least Angle Regression*, in "Annals of Statistics", 2004.
- [59] T. FERGUSON. *A Bayesian Analysis of Some Nonparametric Problems*, in "The Annals of Statistics", vol. 1, n^o 2, 1973, p. 209–230.
- [60] V. GUIGUE. *Méthodes à noyaux pour la représentation et la discrimination de signaux non-stationnaires*, Ph. D. Thesis, INSA, Rouen, 2005.
- [61] T. HASTIE, R. TIBSHIRANI, J. FRIEDMAN. *The elements of statistical learning — Data Mining, Inference, and Prediction*, Springer, 2001.
- [62] J. SETHURAMAN. *A constructive definition of Dirichlet priors*, in "Statistica Sinica", vol. 4, 1994, p. 639-650.
- [63] R. SUTTON, A. BARTO. *Reinforcement learning: an introduction*, MIT Press, 1998.