Activity Report 2014

Project-Team DREAM

Diagnosing, Recommending Actions and Modelling

Table of contents

<div align="center">**Project-Team DREAM**</div>

Keywords: Artificial Intelligence, Knowledge Acquisition, Data Mining, Decision Aid, Environment

Creation of the Project-Team: 2004 October 04.

1. Members

Research Scientists

> Yves Moinard [Inria, Researcher]
> René Quiniou [Inria, Researcher]

Faculty Members

> Alexandre Termier [Team Leader, Univ. Rennes I, Professor, from Aug 2014, HdR]
> Marie-Odile Cordier [Univ. Rennes I, Professor, Emerite, HdR]
> Véronique Masson [Univ. Rennes I, Associate Professor]
> Laurence Rozé [INSA Rennes, Associate Professor]
> Thomas Guyet [Agrocampus-Ouest Rennes, Associate Professor]
> Christine Largouët [Agrocampus-Ouest Rennes, Associate Professor]
> Tassadit Bouadi [Univ. Rennes I, ATER until August 2014]

Engineer

> Louis Bonneau de Beaufort [Agrocampus-Ouest Rennes, Engineer]

PhD Students

> Yulong Zhao [Univ. Rennes I, until Jan 2014]
> Serge Vladimir Emteu Tchagou [STMicroelectronics, CIFRE, From December 2014 to December 2015]

Post-Doctoral Fellow

> Sid Ahmed Benabderrahmane [Inria, until Mar 2014]

Visiting Scientist

> Sophie Robin [Collaborator, ex-Associate Professor]

Administrative Assistants

> Marie-Noëlle Georgeault [Inria, from Dec 2014]
> Isobelle Kelly [Univ. Rennes I, from Feb 2014 until Nov 2014]

Others

> Philippe Besnard [CNRS, Research Director, HdR]
> Anne-Isabelle Graux [INRA, Visiting Scientist, from Apr 2014]
> Torsten Schaub [Univ. Potsdam, Professor, Advanced Research position, from Sep 2014]

2. Overall Objectives

2.1. Introduction

The objective of the Dream projet-team is to design smart surveillance and decision-support systems, using knowledge acquisition from massive and heterogeneous data. The team has a special interest on spatial and/or temporal data.

To achieve this ambitious objective, the Dream project-team concentrates its efforts on the following sub-problems:

- Facilitating queries in massive and heterogeneous data
- Mining complex patterns in massive data
- Developing novel decision-support systems that integrate the user in the analysis loop

This research is conduced in collaboration with our academic and industrial partners. Many decision-support work are motivated by environmental applications, which provide real use cases and collaboration with domain experts to validate the interest of the proposed methods.

3. Research Program

3.1. Introduction

The research agenda of the Dream project-team revolves around the following 4 main topics:

- Simulator-based decision support systems
- Incremental learning
- Mining complex patterns
- Answer Set Programming

3.2. Simulator-based decision support systems

A common way to investigate and understand complex phenomena, such as such as those related to ecosystems, consists in designing a computational model and implementing a simulator to test the system behavior under various parameters. These simulators enable a fine grained understanding of the system studied, however they produce huge quantities of data. To be able to exploit these simulators in decision support scenarios, it is thus critical to provide methods to simplify the interactions with the simulator and handle the large quantity of data produced.

- One approach is to store all the simulation data in a datawarehouse and provide scientists and experts with tools to analyze efficiently the simulation data. Providing users with means to dig through large amount of multidimensional data, from more or less abstract viewpoints, and express preferences on the returned results is an important research topic in databases and data mining. To this end, *Skyline queries* constitute a relevant approach as they retrieve the most interesting objects with respect to multi-dimensional criteria with the possibility of making compromises on conflicting dimensions. The challenge is to define and implement skyline queries in a datawarehouse context. In this field, we are investigating efficient interactive tools for answering dynamic [36] and hierarchical [10] skyline queries.
- Another approach is to simplify the simulation model. For some applications, the system is too complex for a traditional numerical simulation to give relevant results in a short amount of time. It is especially the case when data and knowledge are not available to supply numerical models. Qualitative models offers a good alternative to model complex systems in such context. This abstracted representation offers an efficient computation on model exploration and gives relevant results when querying the system behavior. In the Dream project-team we focused on qualitative models of dynamical systems described as Discrete Event Systems (DES). Recent studies have emphasized the great interest of coupling model-checking techniques with qualitative models. We propose to use the timed automata formalism that allow the explicit representation of time [29]. In this context, the research issues we investigate are the following.
  - The size of a global model constructed from an abstracted description of the system and domain knowledge is potentially huge. A challenging problem is to reduce the size of this model using artificial intelligence tools [37].

- It is necessary to propose a high-level language to explore and predict future changes of the system. Using this language, a stakeholder should express easily any requirements he wants to ask on the system behavior. We investigate the formalization of query patterns relying on recent temporal logics that can be exploited using model-checking techniques [52].

  - Another challenge is the computation of the optimal strategy for a reachability problem (*"what is the best sequence of actions to reach a specific state at a specific time ?"*). In this case we propose to use extended timed automata, such as timed game automata or priced time automata, with controller synthesis methods [30].

- When modelling becomes increasingly complex because of ever-increasing numbers of combined processes, making model-based decision aids are essential. Our approach uses symbolic learning techniques on simulated data to synthetise complex processses and help in decision making. Thus rule induction has attracted a great deal of attention in Machine Learning and Data Mining. However generating rules is not an end in itself because their applicability is not straightforward, especially when their number is high.

  Our goal is to lighten the burden of analyzing a large set of classification rules when the user is confronted to an "unsatisfactory situation" and needs help to decide about the appropriate action to remedy to this situation. The method consists in comparing the situation to a set of classification rules. For this purpose, we have proposed a framework for learning action recommendations dealing with complex notions of feasibility and quality of actions [63].

3.3. Incremental learning

The first learning algorithms were batch learning. They examine all examples and produce a concept description, that is generally not further modified. This is not adapted to dynamic settings where data are delivered continuously. For such settings, incremental algorithms have been proposed. These algorithms examine the training example one at a time (or set by set), maintaining a "best-so-far" description which may be modified each time a new example (or set of examples) arrives. In order to strengthen the learning process, some specific old examples are often kept: this is called partial memory systems. A more specific classification of incremental learning can be found in [58].

Current issues in incremental learning are

- for partial instance memory: how to select examples, [56]

- the problem of hidden context: the target concept may depend of unknown variables, which are not given as explicit attributes [66]

- the problem of concept drift: the target changes with time [65], [39]

- the problem of masked example: the data distribution may change and some examples may not be anymore visible.

As a human expert has to give his opinion on the learned description model, we focus our research on incremental learning of rules ([39]).

3.4. Mining complex patterns

Pattern mining, a subdomain of data mining, is an unsupervised learning method which aims at discovering interesting knowledge from data. Association rule extraction is one of the most popular approach and has received a lot of interest in the last 20 years. For instance, many enhancements have been proposed to the well-known Apriori algorithm [27]. It is based on a level-wise generation of candidate patterns and on efficient candidate pruning having a sufficient relevance, usually related to the frequency of the candidate pattern in the data-set (i.e., the support): the most frequent patterns should be the most interesting. Later, Agrawal and Srikant proposed a framework for "mining sequential patterns" [28], which extends Apriori by coping with the order of elements in patterns. Such approach initiated research on *temporal pattern mining*,

which is of particular interest for the DREAM team. The simplest temporal patterns are sequential patterns that constraints the order of the events in one of its occurrence. More advanced approaches also exploit quantitative information in order to provide significant patterns about both ordering and duration of events as well as inter-event delay. A challenge is that the classical anti-monotony property, used to prune the search space, is difficult to define in this case.

Many work in pattern mining have attempted to improve the runtime efficiency of algorithms, on the one hand, by proposing more efficient representation and execution schemes such as pattern-growth methods [48], or, on the other hand, by focusing on condensed representations such as closed patterns [60], [64]. Other research directions have been investigated to enhance the syntax of patterns e.g. temporal and periodic patterns, mutidimensional and hierarchical patterns, constrained patterns, contextual patterns, etc. Despite these improvements, the size of the results may still be too high. Thus, post-mining or visualization methods have been introduced to let the user focus on results that correspond to his own preferences.

Another challenge of pattern mining is that for each pattern mining task (such as mining itemsets, sequences or graphs) there are many specialized algorithms, each exploiting some ad-hoc optimizations. It is very hard for a practitioner to find an algorithm suited for his problem, and such an algorithm may not exist. There is a need to propose novel *generic* pattern mining algorithms, that exploit the main algorithmic advances proposed in the last 20 years, and that only require a description of their pattern mining problem from practioners. Recently, we have proposed ParaMiner [59], a generic pattern mining algorithm using state of the art optimizations and exploiting the parallelism of multicore processors. The practitioner only has to enter a pattern interest criteria and check that it verifies a *strong accessibility* property coming from set theory. As of now, ParaMiner is the fastest generic pattern mining algorithm, being competitive with specialized algorithm on several pattern mining tasks.

Other approaches propose a completely declarative way to specify the pattern mining problem. In this case, the most used framework is Constraint Programming [44]. We are investigating another approach based on *Answer Set Programming*.

3.5. Answer Set Programing (ASP)

The DREAM team is investigating declarative approaches to solve complex problems such as causal reasoning, landscape simultation and pattern mining. One such approach is ASP.

ASP (Answer set programming) [43], [31] is an approach to declarative problem solving, combining a rich yet simple modelling language with high-performance solving capacities, tailored to Knowledge Representation and Reasoning. "Declarative problem solving" means that the program is close to the way a problem is enunciated, and not to the way the problem is solved. This facilitates writing and revising programs. ASP is an outgrowth of research on the use of non monotonic reasoning in knowledge representation. ASP programs[23] consist in rules that look like Prolog rules, but the computational mechanism is different [54].

ASP allows to solve search problems in NP (and theoretically in NP^{NP}) in a uniform way (being more compact than boolean approaches like SAT solvers). ASP is good when dealing with knowledge representation, particularly when logical rules or graphs are involved. The versatility of ASP is reflected by the ASP solver clasp, winning first places at ASP, SAT and other competitions.

ASP solvers deal with propositional rules, however in practice predicates are allowed. A *grounder* replaces each free variable of the program provided by users with any eligible constant symbol. The output of the grounder is thus a propositional program, which is piped into a *solver* which then computes *answer sets*. These answer sets are the models for the ASP theory, and they constitute the result of an ASP program. The user may ask for all the models, or only one, or any number n of models. The most powerful version (clingo, which combines the grounder gringo and the solver clasp) is from Torsten Schaub's team (see http://potassco. sourceforge.net for the last version of clingo, including a *guide*). These versions can be easily interfaced with python programs, which extends further the practical applicability of ASP [42].

The main interests of using ASP are: 1) the ease to write and to update programs, and 2) the efficiency of the ASP solvers (improved in the recent versions).

Our main challenge is to propose ASP modeling that scales up to solving real problems. We are especially working on the modeling of sequential pattern mining with ASP in order to mine real datasets in a flexible and efficient way.

Our second challenge is to model a wide range of expert knowledge to include reasoning into the solving processes, in order to output more meaningful results.

4. Application Domains

4.1. Introduction

The Dream project-team research applications have been oriented towards surveillance, monitoring and decision support. Our domains of application are:

- Agriculture and environment
- Health
- Exploitation of execution traces in an industrial setting

4.2. Environmental decision making

environment, decision methods

The need for decision support systems in the environmental domain is now well-recognized. It is especially true in the domain of water quality. The challenge is to preserve the water quality from pollutants as nitrates and herbicides, when these pollutants are massively used by farmers to weed their agricultural plots and improve the quality and increase the quantity of their crops. The difficulty is then to find solutions which satisfy contradictory interests and to get a better knowledge on pollutant transfer.

In this context, we are cooperating with INRA (Institut National de Recherche Agronomique) and developing decision support systems to help regional managers in preserving the river water quality. This work began in ANR projects like APPEAU and ACASSYA or the PSDR GO CLIMASTER project (Changement climatique, systèmes agricoles, ressources naturelles et développement territorial).

The approach we advocate is to rely on a qualitative modeling, in order to model biophysical processes in an explicative and understandable way. The SACADEAU model associates a qualitative biophysical model, able to simulate the biophysical process, and a management model, able to simulate farmers' decisions. One of our main contributions is the use of qualitative spatial modeling, based on runoff trees, to simulate the pollutant transfer through agricultural catchments.

The second issue is the use of learning/data mining techniques to discover, from model simulation results, the discriminant variables and automatically acquire rules relating these variables. One of the main challenges is that we are faced with spatiotemporal data. The learned rules are then analyzed in order to recommend actions to improve a current "unsatisfactory" situation.

Our main partners are the SAS INRA research group, located in Rennes and the BIA INRA and AGIR INRA research groups in Toulouse.

Ecosystem Management.
The objective of ecosystem management is to ensure sustainable ecosystems even when submitted to various stressors such as natural disturbances or human pressures. Several studies have already shown the interest of qualitative modelling for ecosystems [38]. In our case, we propose to couple a qualitative modelling with model-checking tools to explore marine ecosystems (as explained section 3.2). We applied our approach on a small-scale subsistence fishery in a coral reef lagoon (Uvea, New Caledonia). A well described foodweb model provides us with useful input data for steady-state biomass data and estimates of production and consumption. A timed automata model was developed using EcoMata to investigate the direct and indirect effects of various fishing strategies on a subset of the trophic network.

This work has been realized in collaboration with ecologists: Yves-Marie Bozec (today in position in Marine Spatial Ecology, University of Queensland, Australia) and Guy Fontenelle (Professeur at Agrocampus Ouest).

A second application has been studied in the dairy management area. Over an hybrid modelling on the grazing activities, four methods to generate the best grazing management activity has been proposed. The expert partners are researchers from the SAS INRA research group, located in Rennes.

4.3. Health

health-care, patient monitoring, medicament usage, pharmaco-immunology, health-care pathways, wireless sensors

Clinical monitoring, electronic patient records and computer supported disease management produce more and larger volumes of clinical data. This data is a strategic resource for healthcare institutions. Data mining brings the facility to discover patterns and correlation hidden within the data repository and assists professionals to uncover these patterns and to exploit them to improve medical care.

We are working on two aspects of health-care:

- exploitation of data from the french care insurance (Assurance Maladie) that contains records of medicament reimbursements for pharmaco-immunology purposes. Our goal is to reconstruct and mine patients' healthcare pathways in order to detect regularities and anomalies in the way patients take medicaments and alert medical authorities in case some problem is detected, such as non expected negative consequences of medicament intake. We are working in the framework of a project funded by the National Medicament Security Agency (ANSM - Agence Nationale de la Sécurité du Médicament) for building a platfom enabling focused studies on specific medicaments as well as discovering potential problems with medicament usage. This means selecting from billions of patients records, patients sharing similar medical contexts and showing different consequences of medicament intake,

- veterinary monitoring of feedlot cattle in big farms from sensors recording behavioral and physio-logical data. As farms are becoming bigger and bigger, detecting ill animals by visual appraisal is becoming more and more difficult. With the advent of cheap wireless sensors, animals (i.e. cows or steers) may be monitored in quasi real time for detecting relevant changes in their behavior that could be related to specific diseases. We are exploring diverse methods for detecting changes on multivari-ate data, such as cusum charts, specific sequential patterns or distribution of frequent patterns. We are specifically working with veterinaries from the university of Calgary (Canada) for monitoring feedlot cattle in farms growing up to 50.000 animals.

4.4. Exploitation of execution traces

log analysis, data mining, embedded systems.

We have an ongoing collaborations with STMicroelectronics, which is one of the world top-5 electronic chip makers. Nowadays, set-top boxes, smartphones or onboard car computers are powered by highly integrated chips called *System-on-Chip (SoC)*. Such chips contain on a single die processing units, memories, IO units and specialized accelerators (such as audio and video encoding/decoding). Programming SoC is a hard task due to their inherent parallelism, leading to subtle bugs when several components do not deliver their results within a given time frame. Existing debuggers and profilers are ill-adapted in this case because of their high intrusivity that modifies the timings. Hence the most used technique is to capture a *trace* of the execution and analyze it post-mortem. While Alexandre Termier was in Grenoble he initiated several works for analyzing such traces with data mining techniques [55], [50], which he is now pursuing with his colleagues of the Dream project-team [14].

4.5. Software components monitoring

software components, web services, distributed diagnosis

Web-services cover nowadays more and more application areas, from travel booking to goods supplying in supermarkets or the management of an e-learning platform. Such applications need to process requests from users and other services on line, and respond accurately in real time. Errors may occur, which need to be addressed in order to still be able to provide the correct response with a satisfactory quality of service (QoS): on-line monitoring, especially diagnosis and repair capabilities, becomes then a crucial concern.

We have been working on this problem within the WS-DIAMOND project [61], a large European funded project involving eight partners in Italy, France, Austria and Netherlands http://wsdiamond.di.unito.it/.

We do not work anymore on the diagnosis of web services, now we aim at coupling diagnosing and repair, in order to implement *adaptive web services*. We started this study by proposing an architecture inspired from the one developed during the WS-DIAMOND project and dedicated to the adaptive processing when faults occur and propagate through the orchestration.

5. New Software and Platforms

5.1. Introduction

The pieces of software described in this section are prototypes implemented by members of the project. Any interested person should contact relevant members of the project.

5.2. Platforms

The Dream project-team, in collaboration with their applicative partners, has proposed and maintains several important software platforms for its main research topics.

5.2.1. *Platform: Environmental decision-support systems*

Participants: Marie-Odile Cordier, Christine Largouët, Véronique Masson, Yulong Zhao.

SACADEAU: the SACADEAU system is an environmental decision software (cf. 4.2) that implements the SACADEAU transfer model. The SACADEAU simulation model couples two qualitative models, a transfer model describing the pesticide transfer through the catchment and a management model describing the farmer decisions. Giving as inputs a climate file, a topological description of a catchment, and a cadastral repartition of the plots, the SACADEAU model simulates the application of herbicides by the farmers on the maize plots, and the transfer of these pollutants through the catchment until the river. The two main simulated processes are the runoff and the leaching. The output of the model simulation is the quantity of herbicides arriving daily to the stream and its concentration at the outlets. The originality of the model is the representation of water and pesticide runoffs with tree structures where leaves and roots are respectively up-streams and down-streams of the catchment.

The software allows the user to see the relationships between these tree structures and the rules learnt from simulations. A more elaborated version allows to launch simulations, to learn rules on-line and to access to two recommendation action algorithms. This year, we have developed a new visualization tool designed to compare two sets of rules learnt from simulations. The user can choose one (or more) rule(s) to compare from one set of rules, and a distance to apply from several multidimensional distances. The most similar rules in the second set of rules are found and the comparison can be easily visualized. The examples covered by "similar" rules can also be presented to the user by highlighting shared positive and negative covered examples. The software is mainly in Java.

The following website is devoted to the presentation of the SACADEAU: http://www.irisa.fr/dream/SACADEAU/. See also [57] for a presentation.

ECOMATA: The ECOMATA tool-box provides means for qualitative modeling and exploring ecosystems and for aiding to design environmental guidelines.We have proposed a new qualitative approach for ecosystem modeling (cf. 4.2) based on timed automata (TA) formalism combined to a high-level query language for exploring scenarios.

To date, ECOMATA is dedicated to ecosystems that can be modeled as a collection of species (prey-predator systems) under various human pressures and submitted to environmental disturbances. It has two main parts: the Network Editor and the Query Launcher. The Network Editor let a stakeholder describe the trophic food web in a graphical way (the species icons and interactions between them). Only few ecological parameters are required and the user can save species in a library. The number of qualitative biomass levels is set as desired. An efficient algorithm generates automatically the network of timed automata. EcoMata provides also a dedicated window to help the user define different fishing pressures, a nice way being by using chronograms. In the Query Launcher, the user selects the kind of query and the needed parameters (for example the species biomass levels to define a situation). Results are provided in a control panel or in files that can be exploited later. Several additional features are proposed in EcoMata: building a species library, import/export of ecosystem model, batch processing for long queries, etc. EcoMata is developed in Java (Swing for the GUI) and the model-checker called for the timed properties verification is UPPAAL.

The following website is devoted to the presentation of ECOMATA: http://oban.agrocampus-ouest.fr:8080/ecomata.

PATURMATA: The Paturmata tool-box provides means for qualitative modeling and exploring agrosystems, specifically management of herd based on pasture [5]. The system is modelled using a hierarchical hybrid model described in timed automata formalism.

In PaturMata software, users can create a pasture system description by entering herds and plots information. For each herd, the only parameter is the number of animals. For each plot, users should enter the surface, the density, the herb height, the distance to the milking shed, a herb growth profile and an accessibility degree.

Users then specify pasturing and fertilization strategies. Finally, users can launch a pasture execution. PaturMata displays the results and a detailed trace of pasture. Users can launch a batch of different strategies and compare the results in order to find the best pasture strategy.

PaturMata is developed in Java (Swing for the GUI) and the model-checker that is called for the timed properties verification is UPPAAL.

Another feature which will be soon added to PaturMata is strategy synthesis. Users choose a pasture configuration or a type of pasture configuration and PaturMata proposes the best pasture and fertilization strategy in order to minimize the pasture procedure cost and use of nitrogen fertilizer.

5.2.2. *Platform: Pattern Mining*

Participants: Thomas Guyet, René Quiniou.

QTempIntMiner: the QTEMPINTMINER (Quantitative Temporal Interval Miner) data mining software implements several algorithms presented in [46] and [3] (QTIAPRIORI and QTIPREFIXSPAN). The software is mainly implemented in Matlab. It uses the Mixmod toolbox [33] to compute multi-dimensional Gaussian distributions. The main features of QTEMPINTMINER are:

- a tool for generating synthetic noisy sequences of temporal events,

- an implementation of the QTEMPINTMINER, QTIAPRIORI and QTIPREFIXSPAN algorithms,

- a graphical interface that enables the user to generate or import data set and to define the parameters of the algorithm and that displays the extracted temporal patterns.

- a sequence transformer to process long sequences of temporal events. Long sequences are transformed into a database of short temporal sequences that are used as input instances for the available algorithms.

The software includes one new algorithm based on the separation of the set of interval to extract more efficiently but less accurately the time interval in temporal patterns. This new algorithm version is still under evaluation on simulated and real datasets (care pathways).

The following website gives many details about the algorithms and provides the latest stable implementation of QTEMPINTMINER: http://www.irisa.fr/dream/QTempIntMiner/.

5.2.3. *Platform: Diagnostic and Monitoring Systems*

Participants: Marie-Odile Cordier, René Quiniou, Sophie Robin.

Odisseptale: the Odisseptale software implements disease detectors using monitoring of data provided by sensors placed on calves or cows. Sensors record streams of data such as body temperature, physical activity, feeding behavior, etc. These data are transmitted regularly to a monitoring software that aims to detect if a noticeable change has occurred on the data streams. Several detectors can be simultaneously active and each contribute to the final decision (detection of a disease). Two kinds of detectors have been implemented: a generic detector based on adaptive CUSUM and a symbolic pattern-based detector. Odisseptale provides also facilities for parameter setting and performance evaluation.

ManageYourself: the ManageYourself software comes from a collaborative project between Dream and the Telelogos company aiming at monitoring smartphones from a stream of observations made on the smartphone state.

Today's smartphones are able to perform calls, as well as to realize much more complex activities. They are small computers. But as in computers, the set of applications embedded on the smartphone can lead to problems. The aim of the project ManageYourself is to monitor smartphones in order to avoid problems or to detect problems and to repair them.

The ManageYourself application includes three parts :

- A monitoring part which triggers preventive rules at regular time to insure that the system is working correctly, e.g. *if the memory is full then delete the tmp directory*. This part is always running on the smartphone.
- A reporting part which records regularly the state of the smartphone (the memory state - free vs allocated -, the connection state, which applications are running, etc.). This part also is always running on the smartphone. The current state is stored in a report at regular period and is labeled *normal*. When an application or the system bugs, the current buggy state is stored in a report and is labeled *abnormal*. At regular timestamps, all the reports are sent to a server where the learning process is executed.
- A learning part which learns new bug rules from the report dataset. This part is executed offline on the server. Once the bug rules are learnt, human experts translates them into preventive rules which are downloaded and integrated in the monitoring part of the smartphones.

5.3. TraceSquiz: reduction of captured trace volume

Participants: Serge Vladimir Emteu Tchagou, Alexandre Termier.

TraceSquiz is a software developed in collaboration with STMicroelectronics. Its goal is to reduce the volume of execution trace captured during endurance tests of multimedia applications. It uses anomaly detection techniques to "learn" regular parts of the trace and only capture the irregular ones. The software is written in C++.

6. New Results

6.1. Simulator-based decision support

Participants: Louis Bonneau de Beaufort, Tassadit Bouadi, Marie-Odile Cordier, Thomas Guyet, Christine Largouët, Véronique Masson, René Quiniou, Sophie Robin, Laurence Rozé, Yulong Zhao.

6.1.1. *Model-checking an ecosystem for decision-aid*

In previous work we have proposed to use qualitative modelling to model ecosystems and we defined a set of high level query patterns to explore th system [53]. This approach has been applied on real-case ecosystems (coral-reef ecosystem in New-Caledonia, fisheries ecosystem in the English channel) and implemented in a tool called EcoMata.

In recent studies we have focussed on the formalization of the qualitative model automatically built from an abstracted ecosystem description. Ecosystems share some common features with concurrent systems represented in the model-checking field: the system complexity is due to interacting components and the system evolution is event-driven and submitted to temporal constraints. However if model-checking techniques are dedicated to finite state systems, ecosystem models are usually represented by analytical models as a set of differential equations. Some studies present how to quantize continuous-time systems in order to diagnose them as discrete-event systems. We proposed a method to build automatically a network of timed automata from various information on the system: description of interactions between components, human knowledge, simple models of population dynamics. The key point is to quantize the continuous-time sub-systems and to get a qualitative model described as network of timed automata. To reduce the size of this network, important after the automatic generation, a learning machine algorithm has been applied in order to reduce the number of "similar" locations. This work has been published in [37].

6.1.2. Controler synthesis for optimal strategy search

Similarly to previous work, this approach relies on a qualitative model of a dynamical system. The problem consists in finding a strategy in order to help the user achieveing a specific goal. The model is now considered as a timed game automata expressing controllable and uncontrollable actions. The strategy represents the sequence of actions that can be performed by a user to reach a particular state (in case of a reachability problem for instance). A first approach based on a "generate and test" method has been developped for the marine ecosystem example [69].

More recently, two new methods for finding the optimal strategies have been proposed. The first one uses controller synthesis on timed automata and exploits the efficency of well-recognized tools. The second one deals with a set of similar models and extracts a more general strategy, closer to what is expected by the stakeholders. These methods have been applied in the context of herd management on a catchment. Yulong Zhao defended his Phd this year on this research subject [5].

6.1.3. A datawarehouse for simulation data

In previous work we have proposed a datawarehouse architecture to store the huge data produced by deep agricultural simulation models [35]. This year, we have worked on hierarchical skyline queries to introduce skyline queries in a datawarehouse framework. Conventional skyline queries retrieve the skyline points in a context of dimensions with a single hierarchical level. However, in some applications with multidimensional and hierarchical data structure (e.g. data warehouses), skyline points may be associated with dimensions having multiple hierarchical levels. Thus, we have proposed an efficient approach reproducing the effect of the OLAP operators "drill-down" and "roll-up" on the computation of skyline queries [10], [25]. It provides the user with navigation operators along the dimensions hierarchies (i.e. specialize / generalize) while ensuring an online calculation of the associated skyline.

6.1.4. Post-mining classification rules

We consider sets of classification rules with quantitative attributes inferred by supervised machine learning, as in the framework of the SACADEAU project. Our aim is to improve human understanding of such sets of rules. Often, output quantitative rules contains too many intervals that are difficult to intepret. It is thus important to merge some of these intervals in order to get more understandable rules. However, blindly merging rules may decrease rule quality. To counter that, we proposed two algorithms for merging intervals via clustering techniques that take into account the final rule quality. The approach automatically detects the most adapted number of clusters required to merge intervals while maintaining rule quality.

6.2. Data Mining

Participants: Sid Ahmed Benabderrahmane, Marie-Odile Cordier, Serge Vladimir Emteu Tchagou, Thomas Guyet, Yves Moinard, René Quiniou, Alexandre Termier.

6.2.1. Application of sequential pattern mining with intervals

Our theoretical work on sequential pattern mining with intervals [47] has been applied to two real issues: the customer relationship management and analysis of care pathways.

Customer Relationship Management (CRM) comprises a set of tools for managing the interactions between a company and its customers. The main objective of the data analysts is to propose the correct service to a customer at the correct moment by applying decision rules. If rules or sequential patterns can predict the interaction that can follow a sequence of actions or events, they can not predict at what time such actions have the highest probability to occur. The objective of temporal pattern mining is to refine the prediction by extracting patterns with information about the duration and delay between the events. This year we have experimented two algorithms on a CRM databases, QTIPrefixSpan [47] and TGSP [68], to extract sequential patterns with quantitative temporal information. We have integrated the TGSP algorithm into an interface to visualize and to browse the extracted patterns. A paper describing this contribution have been recently accepted in a workshop [41].

The QTIPrefixSpan algorithm has also been applied to the analysis of care pathways. The pharmaco-epidemiology platform of the Rennes hospital was interested in characterizing the care pathways preceeding the epileptic seizures of stable epilepic patients. A care pathway consist of the sequence of drug exposures (temporal intervals). The objective is to study the ability of QTIPrefixSpan to identify drug switches between original and generic anti-epileptic drugs. This work is still in progress and will be extended in the PEPS project (see section 8.1.1).

6.2.2. Multiscale segmentation of satellite image time series

Satellite images allow the acquisition of large-scale ground vegetation. Images are available along several years with a high acquisition frequency (1 image every two weeks). Such data are called satellite image time series (SITS). In [45], we presented a method to segment an image through the characterization of the evolution of a vegetation index (NDVI) on two scales: annual and multi-year. This work is now under submission to the journal on Remote Sensing in Environment. The main issue of this approach was the required computation resources (time and memory). Last year, we applied 1D-SAX to reduce data dimensionality [21]. This approach on the supervised classification of large SITS of Senegal and we showed that 1D-SAX approaches the classification results of time series while significantly reducing the required memory storage of the images.

This year, we first continued to explore the supervised classification of SITS using classification trees for time-series [40] by implementing a parallelized version of this algorithm. Secondly, we explored the adaption of the object-oriented segmentation to SITS. The object-oriented segmentation [34] is able to segment images based on the segment uniformity. We proposed a measure for time-series uniformity to adapt the segmentation algorithm and applied it on large multivariate SITS of Senegal. This work have been presented to the conference on spatial analysis and geography [16]. A collaboration with A. Fall (Université Paris-13) have been initiated to compare our results on the Senegal with ground observations. Moreover, we planned to apply our algorithm to analyse the land use in Peru (collaboration with A. Marshall, Université Paris 13/PRODIG).

6.2.3. Analysis and simulation of landscape based on spatial patterns

Researchers in agro-environment need a great variety of landscapes to test their scientific hypotheses using agro-ecological models. Real landscapes are difficult to acquire and do not enable the agronomist to test all their hypothesis. Working with simulated landscapes is then an alternative to get a sufficient variety of experimental data. Our objective is to develop an original scheme to generate landscapes that reproduce realistic interface properties between parcels. This approach is made of the extraction of spatial patterns from a real geographic area and the use of these patterns to generate new "realistic" landscapes. It is based on a spatial representation of landscapes by a graph expressing the spatial relationships between the agricultural parcels (as well as the roads, the rivers, the buildings, etc.), of a specific geographic area.

In past years, we worked on the exploration of graph mining techniques, such as gSPAN [67], to discover the relevant spatial patterns present in a spatial-graph. We assume that the set of the frequent graph patterns are the characterisation of the landscape. Our remaining challenge was to simulate new realistic landscapes that will reproduce the same patterns.
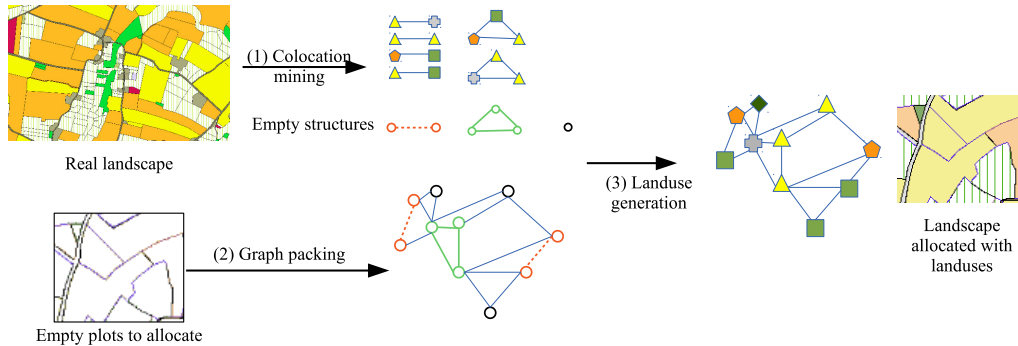

Figure 1. Simulation process in three steps : 1) characteristic graph-patterns mining, 2) graph packing of the cadastral landscape and 3) land use assignement.

This year, we formalized the simulation process by a formal problem of graph packing [51]. The process is illustrated by Figure 1. Solving instances of the general graph packing problem has a high combinatorics and there does not exists any efficient algorithm to solve it. We proposed an ASP program to tackle the combinatorics of the graph packing and to assign the land use considering some expert knowledge. Our approach combines the efficiency of ASP to solve the packing issue and the simplicity of the declarative programming to take into account the expert contraints on the land use. Contraints about the minimum surface of crops or about the impossibility of some crops colocation can be easily defined. This work have been presented at the conference RFIA [19] and we have been invited to provide an extended version to the Revue d'Intelligence Artificielle (RIA). The application results have been presented to the national colloquium on landscape modelling (http://www.reseau-payote.fr/?q=colloque2014).

In addition to the landscape simulation, the challenging tasks of solving the general graph packing with ASP raises interests in more general problem (such as graph compression). We initiated a collaboration with J. Nicolas (Inria/Dyliss) to improve the efficiency of our first programs.

6.2.4. *Mining with ASP*

In pattern mining, a pattern is considered interesting if it occurs frequently in the data, i.e. the number of its occurrences is greater than a fixed given threshold. As non informed mining methods tend to generate massive results, there is more and more interest in pattern mining algorithms able to mine data considering some expert knowledge. Though a generic pattern mining tool that could be tailored to the specific task of a data-scientist is still a holy grail for pattern mining software designers, some recent attempts have proposed generic pattern mining tools [44] for itemset mining tasks. In collaboration with Torsten Schaub, we explore the ability of a declarative language, such as Answer Set Programming (ASP), to solve pattern mining tasks efficiently. A first attempt have been proposed by Jarvisälo for simple settings [49].

This year, we worked on several classical pattern mining tasks: episodes, sequences and closed/maximal itemsets. In [20], we explore the use of ASP to extract frequent episodes (without parallel events) in a unique long sequence of itemsets. We especially evaluate the incremental resolution to improve the efficiency of our program. We next worked on sequence mining to extract pattern from the sequence of TV programs (V. Claveau, CNRS/LinkMedia). This tasks was simpler, but the computation time was significantly higher than

dedicated algorithms. Nonetheless, our recent programs extracting closed or maximal patterns have better results.

6.2.5. *Monitoring cattle*

Following the lines of a previous work [62], we are working on a method for detecting Bovine Respiratory Diseases (BRD) from behavioral (walking, lying, feeding and drinking activity) and physiological (rumen temperature) data recorded on feedlot cattle being fattened up in big farms in Alberta (Canada). This year, we have especially worked on multivariate sensor analysis to devising multivariate decision rules for improving the specificity of detectors [15].

6.2.6. *Subdimensional clustering for fast similarity search over time series data. Application to Information retrieval tasks*

Information retrieval and similarity search tasks in time series databases remains a challenge that requires to discover relevant pattern-sequences that are recurrent over the overall time series sequences, and to find temporal associations among these frequently occurring patterns. However, proposed methods suffer from a lack of flexibility of the used similarity measures, a lack of scalability of the representation model, and a penalizing runtime to retrieve the information. Motivated by these observations, we have designed a framework tackling the query by content problem on time series data, ensuring (i) fast response time, (ii) multi-level information representation, and (iii) representing temporal associations between extracted patterns. This year we have compared several distance measures on time series with different criteria and proposed a hybrid retrieval method based on pattern extraction and clustering [8].

6.2.7. *Knowledge Extraction from Heterogeneous Data*

Recently, mining microarrays data has become a big challenge due to the growing sources of available data. We are using machine learning methods such as clustering, dimensionality reduction, association rules discovery on transcriptomic data, by combining a domain ontology as source of knowledge, in order to supervise the KDD process. Our objectives concern the identification of genes that could participate in the development of tumors. This year, we have introduced a new method for extracting enriched biological functions from transcriptomic databases using an integrative bi-classification approach based on formal concept analysis [7].

6.2.8. *Trace reduction*

One problem of execution trace of applications on embedded systems is that they can grow very large, typically several Gigabytes for 5 minutes of audio/video playback. Some endurance tests require continuous playback for 96 hours, which would lead to hundreds of Gigabytes of traces, that current techniques cannot analyze. We have proposed TraceSquiz, an online approach to monitor the trace output during endurance test, in order to record only suspicious portions of the trace and discard regular ones. This approach is based on anomaly detection techniques, and as been accepted in the DATE'15 conference [14].

6.3. Causal reasoning and argumentation

Participants: Philippe Besnard, Marie-Odile Cordier, Yves Moinard.

We have continued our work on reasoning (precisely search for explanations) from causal relations and ontology [9]. Mainly, we have enforced the use or argumentation [32] in order to help choosing the best explanations among the (rather big) set of explanations given by our previous formalism. Then, we hope to be able to use the last versions of clingo in order to get an efficient tool to deal with complex situations (our example is Xynthia storm, february 2012 in western France for which there exists a huge amount of data from various official reports) by using clingo. For now we have a preliminary program which provides (besides the applications already mentioned: mining and landscape simulation) another application of the recent versions of ASP. One interest is that the nature of the rules in ASP should allow to translate rather directly (and hopefully efficiently) our previous formalism together with the improved argumentation part.

7. Bilateral Contracts and Grants with Industry

7.1. Bilateral Contracts with Industry

7.1.1. SoCTrace

Participants: Serge Vladimir Emteu Tchagou, Alexandre Termier.

SoCTrace is a FUI project led by STMicroelectronics, with the companies ProbaYes and Magilem, Université Joseph Fourier and Inria Rhône-Alpes. Its goal is to provide an integrated environment for storing and analyzing execution traces. In this project, we are working on data mining techniques for analyzing the traces, and on the use of ontologies to enable querying traces with a higher level of abstraction.

7.1.2. ManageYourSelf: diagnosis and monitoring of embedded platforms

Participants: Marie-Odile Cordier, Sophie Robin, Laurence Rozé.

ManageYourSelf is a project that deals with the diagnosis and monitoring of embedded platforms, in the framework of a collaboration with Telelogos, a French company expert in mobile management and data synchronization. ManageYourSelf aims to perform diagnostic and repair on a fleet of mobile smartphones and PDAs. The idea is to embed on the mobile devices a rule-based expert system and its set of politics, for example "if memory full then delete (directory)". At regular intervals the recognition is performed, using the parameters of the phones as the fact base. Of course, it is impossible to foresee all the rules in advance. Upon detection of a non anticipated problem, a report containing all the system's information prior to the problem is sent to a server. The learning step is realised using rules: crash rules are learnt, they are transformed in preventive rules by an expert and embedded on the phone.

7.1.3. Particular contract of the strategic action EDF/Inria

Participants: Thomas Guyet, René Quiniou, Véronique Masson.

At the time of digitalization of multi-channel customer relations, the analysis of customer pathways has become a strategic issue for any business unit. The interaction traces left by clients when connecting to the customer services can be combined with data from other communication channels (phone, web form, e-mail, mail, fax, SMS, shop, etc.) and allow to analyse the customer pathways in details.

Pattern mining tools are able to extract the frequent customer behaviors in very large database of client pathways, but taking into account the duration and the delay between the customer actions remains a challenging issue for pattern mining. The objective of this one year particular contract was to design and to develop a frequent mining tool taking into account the time dimension for analysis of multichannel customer pathways.

8. Partnerships and Cooperations

8.1. National Initiatives

8.1.1. PEPS: Pharmaco-epidemiology for Health Products

Participants: Thomas Guyet, René Quiniou, Véronique Masson, Alexandre Termier.

The PEPS project (Pharmaco-epidemiology des Produits de Santé) is funded by ANSM (national agency for health security). The project leader is E. Oger from the clinical investigation center CIC-1414 INSERM/CHU Rennes. The other partners located in Rennes are the Institute of Research and Technology (IRT) B<>Com, EHESP and the LTSI. The project will start in january 2015 and is funded for 4 years (3.6M€).

The PEPS project has two parts: the clinical studies and a research program dedicated to the development of innovative tools for pharmaco-epidemiological studies with medico-administrative databases. The pharmaco-epidemiology is the study of the uses, the effectiveness and the effects of health products (especially drugs) for the patients in a real live context, on a large population. Using medico-administrative databases – that contains information about the reimbursement of the medication, the medical visits and the cares – is a recent approach to enable studies on large cohortes and to reduce the response time to a pharmaco-epidemiology question.

Our contribution to this project will be the proposal of pattern mining algorithms and reasoning techniques to analyze typical care pathways of specific groups of insured patients.

8.2. International Initiatives

8.2.1. Inria International Partners

8.2.1.1. Informal International Partners

8.2.1.1.1. Monitoring cattle in big herds with multiple sensors
Participant: René Quiniou.

The state of Alberta produces a significant part of the beef meat in Canada. Big farms feeds up around 40.000 bull calves in feedlots grouping 200-300 animals. Diseases such as Bovine Respiratory Diseases (BRD) are frequent and may propagate quickly in such conditions. So, it is important to detect as soon as possible when an animal is sick. We are collaborating with the Department of Production Animal Health, University of Calgary for designing monitoring systems able to generate early alarms when an animal is sick. Precisely, we are studying the properties of new sensors and their aptitude to provide relevant data for BRD detectors. This year, we had a contract with the university of Calgary to fund a grant for a master student.

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific events organisation

9.1.1.1. general chair, scientific chair

- Local chair of EGC 2014 in Rennes (R. Quiniou).
- Local chairs of PFIA 2015 in Rennes (T. Guyet, R. Quiniou).
- Organization chairs and program committee members of FST-CERGEO workshop at EGC 2014 (T. Guyet, R. Quiniou).
- Organization chairs and program committee members of GAST workshop at EGC 2015 (T. Guyet, R. Quiniou).

9.1.1.2. member of the organizing committee

- Organization committee member of EGC 2014 in Rennes (T. Guyet, T. Bouadi, S. Benabderrahmane).

9.1.2. Scientific events selection

9.1.2.1. responsable of the conference program committee

- Steering Committee of RFIA'2014 (T. Guyet).

9.1.2.2. member of the conference program committee

- Program committee member of DX'14 (Principles of Diagnosis)(M.-O. Cordier).
- Program committee member of JIAF'14 (Journés Intelligence Artificielle Fondamentale)(M.-O. Cordier).

- Program committee members of RFIA'2014 (M.-O. Cordier, T. Guyet, A. Termier).
- Program committee members of EGC 2014 and 2015 (R. Quiniou, A. Termier).
- Program committee member of BDA'2014 (A. Termier).
- Program committee member of Data Mining on Networks Workshop of ICDM 2014 (A. Termier)
- Program committee member of International Conference on Data Science and Advanced Analytics (DSAA) 2014 (A. Termier)

9.1.2.3. reviewer

- Reviewer for EDBT'2014 (A. Termier).

9.1.3. Journal

9.1.3.1. member of the editorial board

- *AAI: Applied Artificial Intelligence* (M.-O. Cordier).
- *Interstices webzine* (M.-O. Cordier).
- *Revue d'Intelligence Artificielle* (T. Guyet).

9.1.3.2. reviewer

- *AAI: Applied Artificial Intelligence* (M.-O. Cordier).
- *Interstices webzine* (M.-O. Cordier).
- *Journal of Biomedical Informatics* (T. Guyet).
- *ACM Computing Surveys* (T. Guyet).
- *Data Mining and Knowledge Discovery* (A. Termier).
- *Transactions on Knowledge and Data Engineering* (A. Termier).
- *National Academy Science Letters* (A. Termier).

9.1.4. Scientific animation

9.1.4.1. Scientific and administrative boards

- ECCAI fellow + Honorific member of AFIA (Association Française d'Intelligence Artificielle): M. -O. Cordier
- Member of "Agrocampus-Ouest" scientific board: M.-O. Cordier.
- Member of "Conseil d'administration de l'ISTIC": M.-O. Cordier.
- Head of IRISA department "Data and Knowledge Management" and member of the IRISA scientific management committee: M.-O. Cordier
- Member of the "Prix de thèse AFIA 2014" award committee (selects the best French PhD thesis in the Artificial Intelligence domain): M.-O. Cordier.
- Chair of the INRA CSS-MBIA (Commission scientifique spécialisée "Mathématiques, Biologie et Intelligence Artificielle"): M.-O. Cordier.
- Member of the CoNRS (Comité national recherche scientifique (since october 2012, until september 2014) : M.-O. Cordier
- Chair of an AERES-HCERES evaluation committee and member of two AERES-HCERES evaluation committees : M.-O. Cordier
- Member of two recruitment committees for INRA (CR1 and DR2) : M.-O. Cordier
- Chair of a recruitment committee for a professor position, member of three recruitment committees for assistant-professor positions : M.-O. Cordier
- Member of the AFIA board (since october 2011): T. Guyet.
- Member of the COREGE (Research Committee-COmité de la REcherche du Grand Etablissement) of Agrocampus-Ouest: T. Guyet.

- Evaluator for the Mines-Telecom Foundation, "Futures et Ruptures" program: T. Guyet.
- Evaluator for the National Research Agency (ANR): T. Guyet.
- Member of the Payote-Network board: T. Guyet.

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

Many members of the EPI DREAM are also faculty members and are actively involved in computer science teaching programs in ISTIC, INSA and Agrocampus-Ouest. Besides these usual teachings DREAM is involved in the following programs:

Master: *Module DSS: Apprentissage sur des données séquentielles symboliques*, 10 h, M2, ISTIC University of Rennes (R. Quiniou).

Master: *C++ Programming*, M1, ENSAI, Rennes (T. Guyet),

Master: *Géoinformation*, M2, Agrocampus Ouest Rennes (L. Bonneau, T. Guyet, C. Largouët)

9.2.2. Supervision

PhDs: Yulong Zhao, "Modélisation d'agroécosystèmes dans un formalisme de type systèmes à événements discrets et simulation de scénarios utilisant des outils de model-checking. Application à l'étude des impacts des changements climatiques et des pratiques agricoles sur les flux de nutriments vers les eaux de surface.", january 13th 2014, supervisor Marie-Odile Cordier and Chantal Gascuel

PhD in progress: Philippe Rannou, "Modèle rationnel pour humanoïdes virtuels", october 1st 2010, co-supervisors Marie-Odile Cordier and Fabrice Lamarche

PhD in progress: Serge Vladimir Emteu Tchagou, "Stream mining techniques for online monitoring of MPSoC applications", february 1st 2012, co-supervisors Alexandre Termier, René Quiniou, Miguel Santana and Jean-François Méhaut

PhD in progress: Léon Constantin Fopa, "Mise en contexte de traces pour une analyse en niveaux d'abstraction", january 1st 2012, co-supervisors Fabrice Jouanot, Alexandre Termier and Jean-François Méhaut

PhD in progress: Behrooz Omidvar Tehrani, "Interactive Pattern Space Exploration", october 1st 2012, co-supervisors Sihem Amer-Yahia and Alexandre Termier

PhD in progress: Hamid Mirisaee, "Matrix decomposition for social network analysis and itemset mining", october 1st 2012, co-supervisors Eric Gaussier and Alexandre Termier

PhD in progress: Oleg Iegorov, "Data Mining Environment for Degugging Real Time Issues on MPSoCs", january 1st 2013, co-supervisors Alexandre Termier, Vincent Leroy, Miguel Santana and Jean-François Méhaut

PhD in progress: Rémy Dautriche, "Techniques d'interaction multi-échelles pour la visualisation interactive de traces d'exécution", november 1st 2013, co-supervisors Renaud Blanch, Miguel Santana and Alexandre Termier

9.2.3. Juries

- Committee member of Yulong Zhao's PhD defence (Université de Rennes 1): M.-O. Cordier , C. Largouët.
- Committee member of Sébastien Silva's PhD defence (Université de Lorraine): M.-O. Cordier
- Committee chair of Hervé Jegou's HDR defence (Université de Rennes 1): M.-O. Cordier
- Committee member of Alef Denguir's PhD defence (Université de Montpellier 2): M.-O. Cordier
- Committee member and reviewer of Jeremy Sanhes' Phd defence (Université de Nouvelle Calédonie): A. Termier

- Committee member of Christiane Kamdem-Kengne Phd defence (Université de Grenoble Alpes): A. Termier

9.3. Popularization

- M.-O. Cordier is editorial board member of Interstices webzine.

10. Bibliography

Major publications by the team in recent years

[1] P. BESNARD, M.-O. CORDIER, Y. MOINARD. *Ontology-based inference for causal explanation*, in "Integrated Computer-Aided Engineering", 2008, vol. 15, n⁰ 4, pp. 351-367, http://hal.inria.fr/inria-00476906/en/

[2] C. GASCUEL-ODOUX, P. AUROUSSEAU, M.-O. CORDIER, P. DURAND, F. GARCIA, V. MASSON, J. SALMON-MONVIOLA, F. TORTRAT, R. TRÉPOS. *A decision-oriented model to evaluate the effect of land use and agricultural management on herbicide contamination in stream water*, in "Environmental modelling & software", 2009, vol. 24, pp. 1433-1446

[3] T. GUYET, R. QUINIOU. *Extracting temporal patterns from interval-based sequences*, in "International Joint Conference on Artificial Intelligence (IJCAI)", Barcelone, Spain, July 2011, http://hal.inria.fr/inria-00618444

[4] Y. PENCOLÉ, M.-O. CORDIER. *A formal framework for the decentralised diagnosis of large scale discrete event systems and its application to telecommunication networks*, in "Artificial Intelligence Journal", 2005, vol. 164, n⁰ 1-2, pp. 121-170, http://hal.inria.fr/inria-00511104/en/

Publications of the year

Doctoral Dissertations and Habilitation Theses

[5] Y. ZHAO. *Qualitative modelling and strategy synthesis of grazing activities*, Université Rennes 1, January 2014, https://tel.archives-ouvertes.fr/tel-00933443

Articles in International Peer-Reviewed Journals

[6] W. WANG, T. GUYET, R. QUINIOU, M.-O. CORDIER, F. MASSEGLIA, X. ZHANG. *Autonomic Intrusion Detection: Adaptively Detecting Anomalies over Unlabeled Audit Data Streams in Computer Networks*, in "Knowledge-Based Systems", November 2014, https://hal.inria.fr/hal-01052810

International Conferences with Proceedings

[7] S. BENABDERRAHMANE. *Formal Concept Analysis and Knowledge Integration for Highlighting Statistically Enriched Functions from Microarrays Data*, in "International Work-Conference on Bioinformatics and Biomedical Engineering - IWBBIO", Granada, Spain, April 2014, 1 p. , https://hal.archives-ouvertes.fr/hal-00935378

[8] S. BENABDERRAHMANE, R. QUINIOU, T. GUYET. *Evaluating Distance Measures and Times Series Clustering for Temporal Patterns Retrieval*, in "IEEE IRI - 15th IEEE International Conference on Information Retrieval and Reuse", San Francisco, United States, S. RUBIN, S.-C. CHEN (editors), August 2014, https://hal.inria.fr/hal-01052805

[9] P. BESNARD, M.-O. CORDIER, Y. MOINARD. *Arguments using ontological and causal knowledge*, in "FoIKS", Bordeaux, France, C. BEIERLE, C. MEGHINI (editors), Springer, March 2014, vol. 8367, pp. 79–96, https://hal.inria.fr/hal-00931679

[10] T. BOUADI, M.-O. CORDIER, R. QUINIOU. *Computing Hierarchical Skyline Queries "On-the-fly" in a Data Warehouse*, in "DAWAK 2014 - 16th International Conference on Data Warehousing and Knowledge Discovery", Munich, Germany, LNCS, Springer Verlag, September 2014, https://hal.inria.fr/hal-01052864

[11] J. C. CODY, M.-O. CORDIER, A. GRASTIEN. *Critical Observations in a Diagnostic Problem*, in "IEEE Conference on Decision and Control (CDC-14)", Los Angeles, United States, December 2014, https://hal.inria.fr/hal-01100627

[12] J. C. CODY, M.-O. CORDIER, A. GRASTIEN. *Critical observations in a diagnostic problem*, in "25th International Workshop on Principles of Diagnosis (DX-14)", Graz, Austria, 2014, https://hal.inria.fr/hal-01100645

[13] M.-O. CORDIER, C. LARGOUËT, Y. ZHAO. *Model-Checking an Ecosystem Model for Decision-Aid*, in "IEEE 26th International Conference on Tools with Artificial Intelligence", Limassol, Cyprus, November 2014 [*DOI :* 10.1109/ICTAI.2014.87], https://hal.archives-ouvertes.fr/hal-01094742

[14] S. V. EMTEU TCHAGOU, A. TERMIER, J.-F. MÉHAUT, B. VIDEAU, M. SANTANA, R. QUINIOU. *Reducing trace size in multimedia applications endurance tests*, in "Design, Automation & Test in Europe Conference & Exhibition (DATE)", Grenoble, France, 2015, https://hal.archives-ouvertes.fr/hal-01093576

[15] A.-V. FREYA MANG, E. TIMSIT, R. QUINIOU. *Early detection of bovine respiratory disease (BRD) in feedlot using multiple health monitoring systems*, in "XIX Congress on bovine medicine", Oviedo, Spain, June 2014, https://hal.inria.fr/hal-01093905

[16] T. GUYET, S. MALINOWSKI, M.-C. BENYOUNÈS. *Extraction des zones cohérentes par l'analyse spatio-temporelle d'images de télédétection*, in "Proceedings of the Spatial Analysis and Geography conference (SAGEO)", Grenoble, France, November 2014 [*DOI :* 10.3166/HSP.2014.1-14], https://hal.inria.fr/hal-01088643

National Conferences with Proceedings

[17] K. GASHTEOVSKI, T. GUYET, R. QUINIOU, A. DA SILVA. *Sequential pattern mining for customer relationship management analysis*, in "Atelier GAST@EGC2015", Luxembourg, Luxembourg, January 2015, https://hal.inria.fr/hal-01097466

[18] T. GUYET, F. LE BER, S. DA SILVA, C. LAVIGNE. *Comparaison des chemins de Hilbert adaptatif et des graphes de voisinage pour la caractérisation d'un parcellaire agricole*, in "Conférence Extraction et Gestion de Connaissances", Rennes, France, January 2014, https://hal.inria.fr/hal-00916964

[19] T. GUYET, Y. MOINARD. *Programmation par ensembles réponses pour simuler l'assolement d'un paysage*, in "Reconnaissance de Formes et Intelligence Artificielle (RFIA) 2014", France, June 2014, https://hal.archives-ouvertes.fr/hal-00989199

[20] T. GUYET, Y. MOINARD, R. QUINIOU. *Using Answer Set Programming for pattern mining*, in "Intelligence Artificielle Fondamentale", Angers, France, June 2014, https://hal.inria.fr/hal-01069092

[21] S. MALINOWSKI, T. GUYET, R. QUINIOU, R. TAVENARD. *1d-SAX : une nouvelle représentation symbolique pour les séries temporelles*, in "Conférence Extraction et Gestion de Connaissances", Rennes, France, January 2014, https://hal.inria.fr/hal-00916970

Scientific Books (or Scientific Book chapters)

[22] L. AMGOUD, P. BESNARD, C. CAYROL, P. CHATALIC, M.-C. LAGASQUIE-SCHIEX. *Argumentation et raisonnement en présence de contradictions*, in "Panorama de l'intelligence artificielle Ses bases méthodologiques, ses développements", P. MARQUIS, O. PAPINI, H. PRADE (editors), Représentation des connaissances et frtomalisation des raisonnements, Cépaduès, May 2014, vol. 1, https://hal.inria.fr/hal-00770580

[23] A. LALLOUET, Y. MOINARD, P. NICOLAS, I. STÉPHAN. *Programmation logique*, in "Panorama de l'intelligence artificielle Ses bases méthodologiques, ses développements", P. MARQUIS, O. PAPINI, H. PRADE (editors), Algorithmes pour l'intelligence artificielle, Cépaduès, May 2014, vol. 2, chapitre 4, https://hal.inria.fr/hal-00758896

[24] C. REYNAUD, A. MARTIN, R. QUINIOU. *EGC - 14èmes Journées Francophones Extraction et Gestion des Connaissances*, Revue des Nouvelles Technologies de l'Information, Hermann, January 2014, vol. E-26, https://hal.archives-ouvertes.fr/hal-01108551

Scientific Popularization

[25] T. BOUADI, M.-O. CORDIER, R. QUINIOU. *Requêtes skyline hiérarchiques*, in "10e journées francophones sur les Entrepôts de Données et l'Analyse en ligne (EDA 2014)", Vichy, France, June 2014, https://hal.inria.fr/hal-01094661

Other Publications

[26] T. GUYET, S. DA SILVA, C. LAVIGNE, F. LE BER. *Caractérisation d'un parcellaire agricole : comparaison des sacs de noeuds obtenus par un chemin de Hilbert adaptatif et un graphe de voisinage*, January 2014, 13 p. , Atelier FST "fouille de données spatiales et temporelles" associé à la conférence EGC, Rennes, 28 janvier 2014, https://hal.inria.fr/hal-01100583

References in notes

[27] R. AGRAWAL, T. IMIELINSKI, A. N. SWAMI. *Mining Association Rules between Sets of Items in Large Databases*, in "Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data", Washington, D.C., P. BUNEMAN, S. JAJODIA (editors), 26–28 1993, pp. 207–216

[28] R. AGRAWAL, R. SRIKANT. *Mining sequential patterns*, in "Eleventh International Conference on Data Engineering", Taipei, Taiwan, P. S. YU, A. S. P. CHEN (editors), IEEE Computer Society Press, 1995, pp. 3–14

[29] R. ALUR, D. DILL. *A theory of timed automata*, in "Theoretical Computer Science", 1994, vol. 126, pp. 183-235

[30] E. ASARIN, O. MALER, A. PNUELI. *Symbolic Controller Synthesis for Discrete and Timed Systems*, in "Hybrid Systems II, LNCS 999", Springer, 1995, pp. 1–20

[31] C. BARAL. *Knowledge Representation, Reasoning and Declarative Problem Solving*, Cambridge University Press, 2003

[32] P. BESNARD, A. HUNTER. *Elements of Argumentation*, The MIT Press, june 2008, http://mitpress.mit.edu/books/elements-argumentation

[33] C. BIERNACKI, G. CELEUX, G. GOVAERT, F. LANGROGNET. *Model-Based Cluster and Discriminant Analysis with the MIXMOD Software*, in "Computational Statistics and Data Analysis", 2006, vol. 51, n° 2, pp. 587-600

[34] T. BLASCHKE. *Object based image analysis for remote sensing*, in "J. Photogramm.", 2010, vol. 65, n° 1, pp. 2–16

[35] T. BOUADI. *Analyse multidimensionnelle interactive de résultats de simulation. Aide à la décision dans le domaine de l'agroécologie*, Université Rennes 1, November 2013, https://tel.archives-ouvertes.fr/tel-00933375

[36] T. BOUADI, M.-O. CORDIER, R. QUINIOU. *Computing Skyline Incrementally in Response to Online Preference Modification*, in "Transactions on Large-Scale Data- and Knowledge- Centered Systems", 2013, vol. 10, pp. 34-59 [*DOI :* 10.1007/978-3-642-41221-9_2], http://hal.inria.fr/hal-00920548

[37] M. CORDIER, C. LARGOUËT, Y. ZHAO. *Model-Checking an Ecosystem Model for Decision-Aid*, in "26th IEEE International Conference on Tools with Artificial Intelligence, ICTAI 2014, Limassol, Cyprus, November 10-12, 2014", 2014, pp. 539–543

[38] J. DAMBACHER, W. HIRAM, P. ROSSIGNOL. *Qualitative predictions in model ecosystems*, in "Ecological Modelling", 2003, vol. 161, pp. 79-93

[39] M. DECKERT. *Incremental rule-based leraners for handling concept drift : an overview*, in "foundations of computing and decision sciences", 2013, vol. 38, pp. 35-65

[40] A. DOUZAL-CHOUAKRIA, C. AMBLARD. *Classification Trees for Time Series*, in "Pattern Recogn.", 2012, vol. 45, n° 3, pp. 1076–1091

[41] K. GASHTEOVSKI, T. GUYET, R. QUINIOU, A. GOMES DA SILVA, V. MASSON. *Sequential pattern mining for customer relationship management analysis*, in "Workshop GAST at EGC'2015", 2015

[42] M. GEBSER, R. KAMINSKI, B. KAUFMANN, T. SCHAUB. *Clingo = ASP + Control: Preliminary Report*, in "Proc. of Technical Communications of the Twenty-ninth International Conference on Logic Programming (ICLP'14)", 2014

[43] M. GELFOND, V. LIFSCHITZ. *The Stable Model Semantics for Logic Programming*, in "Proceedings of ICLP", 1988

[44] T. GUNS, A. DRIES, G. TACK, S. NIJSSEN, L. DE RAEDT. *MiningZinc: A modeling language for constraint-based mining*, in "Proceedings of the Twenty-Third international joint conference on Artificial Intelligence", AAAI Press, 2013, pp. 1365–1372

[45] T. GUYET, H. NICOLAS, A. DIOUCK. *Segmentation multi-échelle de séries temporelles d'images satellite : Application à l'étude d'une période de sécheresse au Sénégal*, in "Reconnaissance de Forme et Intelligence Artificielle (RFIA)", Lyon, France, January 2012, http://hal.inria.fr/hal-00646158

[46] T. GUYET, R. QUINIOU. *Mining temporal patterns with quantitative intervals*, in "4th International Workshop on Mining Complex Data at ICDM 2008", December 2008

[47] T. GUYET, R. QUINIOU. *Extracting temporal patterns from interval-based sequences*, in "International Joint Conference on Artificial Intelligence (IJCAI)", Barcelone, Spain, July 2011, http://hal.inria.fr/inria-00618444

[48] J. HAN, J. PEI, Y. YIN. *Mining Frequent Patterns Without Candidate Generation*, in "SIGMOD Rec.", May 2000, vol. 29, n° 2, pp. 1–12, http://doi.acm.org/10.1145/335191.335372

[49] M. JÄRVISALO. *Itemset mining as a challenge application for answer set enumeration*, in "Logic Programming and Nonmonotonic Reasoning", Springer, 2011, pp. 304–310

[50] C. KAMDEM KENGNE, L. C. FOPA, A. TERMIER, N. IBRAHIM, M.-C. ROUSSET, T. WASHIO, M. SANTANA. *Rewriting Large Multimedia Application Execution Traces with few Event Sequences*, in "KDD 2013 - International Conference on Knowledge Discovery and Data Mining", Chicago, IL, United States, INDERJIT S. DHILLON ET AL. (editor), ACM, August 2013, pp. 1348-1356, Session Industrial and government emerging (Security & Privacy) [*DOI :* 10.1145/2487575.2488211], https://hal.archives-ouvertes.fr/hal-00923538

[51] D. G. KIRKPATRICK, P. HELL. *On the Completeness of a Generalized Matching Problem*, in "Proceedings of the Tenth Annual ACM Symposium on Theory of Computing", New York, NY, USA, STOC '78, ACM, 1978, pp. 240–245, http://doi.acm.org/10.1145/800133.804353

[52] C. LARGOUËT, M.-O. CORDIER. *Patrons de scénarios pour l'exploration qualitative d'un écosystème*, in "RFIA 2010 (Reconnaissance des Formes et Intelligence Artificielle)", France Caen, 2010

[53] C. LARGOUËT, M.-O. CORDIER, Y.-M. BOZEC, Y. ZHAO, G. FONTENELLE. *Use Of Timed Automata And Model-Checking To Explore Scenarios On Ecosystem Models*, in "Environmental Modelling and Software", November 2011, n° 30, pp. 123-138, On-line publication : 26 November 2011 [*DOI :* 10.1016/J.ENVSOFT.2011.08.005], http://hal.inria.fr/hal-00649275

[54] V. LIFSCHITZ. *What Is Answer Set Programming?*, in "Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence, AAAI 2008, Chicago, Illinois, USA, July 13-17, 2008", 2008, pp. 1594–1597

[55] P. LÓPEZ-CUEVA, A. BERTAUX, A. TERMIER, J.-F. MÉHAUT, M. SANTANA. *Debugging Multimedia Application Traces through Periodic Pattern Mining*, in "EMSOFT 2012, part of ESWEEK - Embedded Systems Week", Tampere, Finland, ACM, October 2012, pp. 13-22, Session 1A: Testing and Characterization of Embedded Software [*DOI :* 10.1145/2380356.2380366], https://hal.archives-ouvertes.fr/hal-00750941

[56] M. A. MALOOF, R. S. MICHALSKI. *Incremental learning with partial instance memory*, in "Artificial Intelligence", 2004, vol. 154, pp. 95-261

[57] V. MASSON, F. PLOYETTE, M.-O. CORDIER, C. GASCUEL-ODOUX, R. TRÉPOS. *Sacadeau-Software, un logiciel d'aide à la décision pour améliorer la qualité de l'eau*, in "Revue d'Intelligence Artificielle", September 2013, vol. 27, n° 4-5, pp. 443-469 [*DOI :* 10.3166/RIA.27.443-469], http://hal.inria.fr/hal-00881658

[58] R. S. MICHASKI. *A theory and methodology of inductive leraning*, in "Machine learning : an Arificial Intelligence Approach", 1983, vol. I

[59] B. NEGREVERGNE, A. TERMIER, M.-C. ROUSSET, J.-F. MÉHAUT. *ParaMiner: a Generic Pattern Mining Algorithm for Multi-Core Architectures*, in "Data Mining and Knowledge Discovery", May 2014, vol. 28, n° 3, pp. 595-633 [*DOI : 10.1007/S10618-013-0313-2*], https://hal.archives-ouvertes.fr/hal-00923535

[60] N. PASQUIER, Y. BASTIDE, R. TAOUIL, L. LAKHAL. *Discovering Frequent Closed Itemsets for Association Rules*, in "Proceedings of the 7th International Conference on Database Theory", London, UK, ICDT '99, Springer-Verlag, 1999, pp. 398–416, http://dl.acm.org/citation.cfm?id=645503.656256

[61] THE WS-DIAMOND TEAM. *2*, in "WS-DIAMOND: Web Services DIAgnosability, MONitoring and Diagnosis", J. MYLOPOULOS, M. PAPAZOGLOU (editors), MIT Press Series on Information Systems, 2009

[62] E. TIMSIT, S. ASSIÉ, R. QUINIOU, H. SEEGERS, N. BAREILLE. *Early detection of bovine respiratory disease in young bulls using reticulo-rumen temperature boluses*, in "The Veterinary Journal", 2010 [*DOI : 10.1016/J.TVJL.2010.09.012*]

[63] R. TRÉPOS, A. SALLEB-AOUISSI, M.-O. CORDIER, V. MASSON, C. GASCUEL-ODOUX. *Building Actions From Classification Rules*, in "Knowledge and Information Systems (KAIS) journal", 2013, vol. 34, n° 2, pp. 267-298 [*DOI : 10.1007/S10115-011-0466-5*], http://hal.inria.fr/hal-00649388

[64] T. UNO, T. ASAI, Y. UCHIDA, H. ARIMURA. *An efficient algorithm for enumerating closed patterns in transaction databases*, in "Proc. DS'04, LNAI 3245", 2004, pp. 16–31

[65] G. WIDMER, M. KUBAT. *Learning flexible Concepts form Streams of Examples : FLORA2*, in "proceeding of the 10th European Conference On Artificial Intelligence", WILEY AND SONS (editor), 1992, pp. 463-467

[66] G. WIDMER, M. KUBAT. *Learning in the presence of concept drift and hidden contexts*, in "Machine Learning", 1996, vol. 23, n° 1, pp. 69-101

[67] X. YAN, J. HAN. *gSpan: Graph-Based Substructure Pattern Mining*, in "Proceedings of the 2002 IEEE International Conference on Data Mining", Washington, DC, USA, ICDM '02, IEEE Computer Society, 2002, pp. 721–

[68] S.-J. YEN, Y.-S. LEE. *Mining Non-redundant Time-gap Sequential Patterns*, in "Applied Intelligence", 2013, vol. 39, n° 4, pp. 727–738

[69] Y. ZHAO, M.-O. CORDIER, C. LARGOUËT. *Répondre aux questions "Que faire pour" par synthèse de contrôleur sur des automates temporisés - Application à la gestion de la pêche*, in "RFIA 2012 (Reconnaissance des Formes et Intelligence Artificielle)", Lyon, France, January 2012, pp. 978-2-9539515-2-3, Session "Posters", http://hal.inria.fr/hal-00656543