INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team DREAM

Diagnosis, REcommending Actions and Modeling

Rennes - Bretagne-Atlantique

Theme : Knowledge and Data Representation and Management

Activity Report

2009

Table of contents

1. Team

Research Scientist

Yves Moinard [CR]

René Quiniou [CR]

Faculty Member

Marie-Odile Cordier [Team Leader, Professor, University of Rennes 1, HdR]

Véronique Masson [Associate professor, University of Rennes 1 (50 % time)]

Sophie Robin [Associate professor, University of Rennes 1]

Laurence Rozé [Associate professor, Insa, Rennes]

Xavier Le Guillou [Ater,University of Rennes1]

External Collaborator

Philippe Besnard [Research scientist, DR CNRS, IRIT, Toulouse]

Thomas Guyet [Associate Professor, Agrocampus-Ouest, Rennes]

Christine Largouët [Associate Professor, Agrocampus-Ouest, Rennes]

Post-Doctoral Fellow

Alice Marascu [since the 15th of September 2009]

Administrative Assistant

Marie-Noëlle Georgeault [AA Inria]

2. Overall Objectives

2.1. Introduction

The research objectives of the Dream team are about monitoring complex systems. Our aim is to enhance monitors in such a way that they can achieve their task at best, especially when the monitored systems are subject to failures, to degraded functionality or more generally when the (internal or external) context is evolving. We consider mainly distributed systems composed of a set of components, or a set of agents (software agents, robots) which cooperate to achieve a common goal. They can be physical systems as telecommunication networks, software systems as web services, or even a set of collaborative robots. In all cases, they are supposed to achieve a goal (or contract), possibly expressed by a set of QoS (Quality of Service) constraints. These systems can be either supervised by a human operator or autonomous systems. In the first case, the operator is in charge of taking the decisions as it is often the case for telecommunication or power distribution networks. The second case includes embedded systems, such as web services, robotic or automotive systems, where the repair/reconfiguration actions are triggered by the system itself. The idea which has oriented our research activities for many years is that the repair/recovery actions must be based not only on symptoms (as exception handlers do for instance), but on diagnosing the deep causes of failures. In this view, the diagnosis task, which has the burden of analyzing the symptoms in order to locate and identify the causes, becomes the cornerstone on which relies the decision task. Since observed systems may be quite large, most of the time some part of the diagnosis must be computed locally, in a distributed way. But, to have a global view of the situation, a synchronization process is necessary. Thus, in our research, we assume that an agent (called user when it is a human agent and broker when it is a software agent), is in charge of exploiting the diagnosis to trigger the adequate action. The decision can consequently be globally consistent, even if actions are locally executed in a distributed way. These two credos impact our research directions both from an architectural point of view (not purely distributed) and from a diagnostic point of view (decision-oriented diagnosis): the ultimate goal of the system is not merely to diagnose itself, but to react, or suggest reaction, accurately, based on a diagnosis. We are especially interested in *self-healing* systems, i.e. systems able to repair themselves after the occurrence of a fault: they should be able to always provide a sufficient diagnosis to trigger an adequate repair process. We have chosen to develop a model-based approach. The diagnosis task assumes the existence of a model of the system, describing the expected behavior as well as potential faulty behaviors. Moreover, we give the preference to qualitative models, which give an abstract view of the system and which is often easier to understand. Qualitative model-based approaches are advocated for at least two main reasons:

- they are "glass-box" approaches which means that diagnoses and recommended actions can be explained to the user in an explicit and adequate language,
- they are flexible enough and are then adapted to quickly evolving systems such as technological systems (for instance telecommunication components).

Dealing with dynamical systems in an on-line context, we give a central role to temporal information, when modeling systems as well as when diagnosing situations, or deciding actions, and hence also when acquiring knowledge on the systems. The models are expressed using event-based formalisms such as discrete-event systems (mainly described by automata), or sets of chronicles (a chronicle is a temporally constrained set of events).

To sum up, the challenge we have in mind is to design smart (or lively) systems, both adaptable and dependable, to answer the demand for self-healing embedded systems. The approach we propose is to develop formal methods and efficient algorithms dedicated to the on-line monitoring and repair-oriented diagnosis of complex distributed systems using qualitative temporal models. In this context, the research questions we are investigating are the following. Even though they are clearly highly related, we have chosen for the sake of clarity to present them in two distinct paragraphs. The first one is devoted to on-line monitoring issues and the second one to design and model acquisition issues. Finally, we provide some application domains.

2.2. On-line monitoring issues

Classical model-based diagnosis methodologies have been shown to be inadequate for complex systems due to the intractable size of the model and the computational complexity of the process. It is especially true when on-line diagnosis is considered and when the system is composed of a number of interacting components (or agents). This is why we focus on a decentralized approach which relies on computing local diagnoses from local models and synchronizing them to get a global view of the current state of the system. The problems we are investigating are: Which strategy to select for synchronizing in an optimal way the local diagnoses in order to preserve the efficiency and the completeness of the process? Which kind of communication protocols to use? How to improve the efficiency of the computation by using adequate symbolic representations such as BDD and partial order reduction techniques? How to ensure an efficient incremental process, in an on-line diagnosis context where observations are incrementally collected? How to deal with reconfigurable systems, the topology of which can change at runtime?

The diagnosis task is usually a steady task which does not take into account the current context and the possible repair actions. However, designing an adaptive system requires to define monitoring and diagnosis tasks which adapt to the context. A first possible improvement is to have a more active diagnosis, capable of tuning the observability of the system, deciding for instance to activate other sensors to acquire more information. A second idea is to adapt the diagnosis granularity, i.e. the details of the information that is needed, to the available repair procedures in the system.

2.3. Design and model acquisition issues

When designing a dependable and adaptive system, a main point is to formally characterize the intended properties of the system such as the diagnosability (i.e. whether, given the system specifications, it is possible to detect and explain an error in due time), or the repairability (i.e. whether it is possible to get the system back to correctness, in due time). Moreover, these two properties must be combined to get the best compromise, so as to get actual self-healing systems. Some of these concepts have been defined, but in a centralized context. We are interested in extending the solutions proposed so far for discrete-event systems in the decentralized context.

It is well recognized that model-based approaches suffer from the difficulty of model acquisition. The first issue we have studied is the automatic acquisition of models from data with symbolic learning methods and data mining methods. The problems which are investigated are listed here. How to improve relational learning methods to cope efficiently with data coming from signals (as electrocardiograms in the medical domain) or alarm logs (in the telecommunication domain)? How to integrate signal processing algorithms to the learning or diagnosis tasks when these latter ones rely on a qualitative description of signals? How to adapt the learning process to deal with multiple sources of information (multi-sensor learning)? How to apply learning techniques to spatio-temporal data? How to combine data mining and visualization to help experts build their models?

Concerning evolving context management and adaptive systems, an emerging issue is to detect when a model is becoming obsolete and to update it by taking advantage of the current data. This difficult and new issue is related to data streams processing and is highly challenging in the monitoring research area where the model is used as a reference by the diagnosis task.

The last point we consider is the decision part itself, mainly the ability of proposing repair policies in order to restore the functionalities of the system or the expected quality of service. A first direction is to interleave diagnosis and repair and to design some decision-theoretic procedure to dynamically choose the best action to undertake. Another direction is to be able to automatically build the recommending actions from simulation or recorded data.

2.4. Application domains

Our application domains are related to funds and contracts we have got thanks to long-term relations with academic and industrial partners. These application domains serve us both as providers of real challenging problems and as test-beds for our research development. They should not be seen as distinct research areas but as distinct experimentation fields, enabling the confrontation of similar methodologies and techniques to various application contexts. The following application domains are investigated:

- large component-based system monitoring applications, the two focused ones being telecommunication networks, and software systems as those found in embedded systems or web services.
- surveillance systems for monitoring interaction data in telecommunication networks (internet) or electrical consumption data.
- health-care applications for monitoring physiological signals, e.g. from patients in hospital or from cows in big dairy herds.
- environmental protection, and more precisely the development of decision support systems to help the management of agricultural plots with the objective of preserving water quality threatened by pollution.

2.5. Highlights of the year

This year can be seen as a consolidation year. It has been dedicated to the submission to scientific journals of the research results obtained during the last years. Four journal articles and one book chapter were published; four other ones are submitted. They cover all our research activities, diagnosis, data stream mining, qualitative modelling and decision-aid in the environmental domain. Moreover, we participated in two books that will hopefully appear next year. The final report for the MOZAE project has been completed. This year has also seen the start of new projects, that has not yet given place to publications, as the cooperation with Orange Labs on data streams analysis, the cooperation with NICTA on interleaving diagnosis and repair, the launching of two new projects concerning the environmental domain (ACCASYA and CLIMASTER).

3. Scientific Foundations

3.1. Computer assisted monitoring and diagnosis of physical systems

Our work on monitoring and diagnosis relies on model-based approaches developed by the Artificial Intelligence community since the founding studies by R. Reiter and J. de Kleer [81], [94]. Our project investigates the on-line monitoring and diagnosis of systems, which are modeled as discrete events systems, focusing more precisely on monitoring by alarms management [58]. Computational efficiency is a crucial issue for real size problems. We are developing two approaches. The first one relies on diagnoser techniques [86], for which we have proposed a decentralized and generic approach. The second one uses chronicle recognition techniques, focusing on learning chronicles.

Early work on model-based diagnosis dates back to the 70-80's by R. Reiter, the reference papers on the logical theory of diagnosis being [81], [94]. In the same years was constituted the community known as DX, named after the *workshop on the principles of diagnosis*. Research in these areas is still very active and the workshop gathers about fifty people in the field every year. As opposed to the expert system approach, which has been the leading approach for diagnosis (medical diagnosis for instance) before 1990, the model-based approach lies on a deep model representing the expected correct behavior of the system to be supervised or on a fault model. Instead of acquiring and representing an expertise from experts, the model-based approach uses the design models of industrial systems. The approach has been initially developed for electronic circuits repair [95], focusing on off-line diagnosis of so-called static systems. Two main approaches have been proposed then: (i) the consistency-based approach, relying on a model of the expected correct behavior, which aims at detecting the components responsible for a discrepancy between the expected observations and the ones actually observed ; (ii) the abductive approach which relies on a model of the failures that might affect the system, and which identifies the failures or the faulty behavior explaining the anomalous observations. See the references [24], [26] for a detailed exposition of these investigations.

Since 1990, the researchers in the field have studied dynamic system monitoring and diagnosis, in a similar way as researchers in control theory do. What characterizes the AI approach is the use of qualitative models instead of quantitative ones and the importance given to the search for the actual source/causes of the faulty behavior. Model-based diagnosis approaches rely on qualitative simulation or on causal graphs in order to look for the causes of the observed deviations. The links between the two communities have been enforced, in particular for what concerns the work about discrete events systems and hybrid systems. Used formalisms are often similar (automata, Petri nets ,...) [33], [58].

Our team focuses on monitoring and on-line diagnosis of discrete events systems and in particular on monitoring by alarm management. In this context, a human operator is generally in charge of the system monitoring and receives time-stamped events (the alarms) which are emitted by the components themselves, in reaction to external events. These observations on the system are discrete pieces of information, corresponding to an instantaneous event or to a property associated to a time interval. The main difficulties for analyzing this flow of alarms are the following:

- the huge number of received alarms: the supervisor may receive up to several hundreds of messages per second, many of which being insignificant,
- the alarm overlapping: the order in which alarms are received may be different from the order in which alarms were emitted. Moreover, various sequences of alarms resulting from concurrent failures may overlap. The propagating delays, and sometimes the ways the alarms are transmitted, must be taken into account, not only for event reordering, but also to decide at what time all the useful messages can be considered as being received.
- the redundancy of received alarms: some alarms are only routine consequence of other alarms. This can provoke a phenomenon known as cascading alarms.
- the alarm loss or alarm masking: some alarms can be lost or masked to the supervisor when an intermediate component in charge of the transmission is faulty. The absence of an alarm must be taken into account, since it can give a useful information about the state of the system.

There are two cases focusing on very different issues. In the first one, the alarms must be dealt with *on-line* by the operator. In this case, alarm analysis must be done in real time. The operator must react in a very short period of time to keep the system working at best in spite of the inputs variability and the natural evolution of the processes. Consequently, the natural system damages (components wear, slow modification of the components properties, etc.) are not directly taken into account but are corrected by tuning some parameters.

This *reactive* treatment withstands the treatment of alarms maintenance. In this second case, a deeper *off line* analysis of the system is performed, by foreseeing the possible difficulties, by planning the maintenance operations in order to minimize significantly the failures and interruptions of the system.

The major part of our work focuses on on-line monitoring aid and it is assumed that the correct behavior model or the fault models of the supervised systems are available. However, an on-line use of the models is rarely possible because of its complexity with respect to real time constraints. This is especially true when temporal models are under concern. A way to tackle this problem is to make an off-line transformation (or compilation) of the models and to extract, in an adapted way, the useful elements for diagnosis.

We study two different methods:

- In the first method, the automaton used as a model is transformed off-line into an automaton adapted to diagnosis. This automaton is called a *diagnoser*. The transitions of the automaton are only triggered by observable events and the states contain only information on the failures that occurred in the system. Diagnosing the system consists in going through all the different states of the diagnoser as observable events become available. This method has been proposed by M. Sampath and colleagues [86]. We have extended this method to the communicating automata formalism [84] (see also [82]). We have also developed a more generic method which takes advantage of the symmetries in the architecture of the system [83].

  The main drawback of centralized approaches is that they require to explicitly build the global model of the system which is unrealistic for large and complex systems as telecommunication networks. It is why our more recent work deals with a decentralized approach [77]. This approach can be compared with R. Debouk and colleagues [49] and also to P. Baroni and colleagues [32]. Our method, unlike R. Debouk et al., relies on local models. We do not need to construct a global model. Indeed, the size of the global model would have been too large in our applications. Even if the methods are very close, P. Baroni et al. are concerned with an *a posteriori* diagnosis (off-line) whereas we propose an on-line diagnosis. Each time an alarm comes, it is analyzed and the diagnosis hypotheses are incrementally computed and given to the operator. Our main theme of study is close to E. Fabre and colleagues [56], [28]. The main difference is that they propose a multi-agent approach where the diagnoses are computed locally at the component level using message exchanges, whereas we construct a global diagnosis which is given to the operator at the supervisor level.

- In the second method, the idea is to associate each failure that we want to detect with a *chronicle* (or a scenario), i.e. a set of observable events interlinked by time constraints. The chronicle recognition approach consists in monitoring and diagnosing dynamic systems by recognizing those chronicles on-line [53], [80], [51].

  One of our research focus is to extend the chronicle recognition methods to a distributed context. Local chronicle bases and local recognizers are used to detect and diagnose each component. However, it is important to take into account the interaction model (messages exchanged by the components). Computing a global diagnosis requires then to check the synchronisation constraints between local diagnoses.

  Another issue is the chronicle base acquisition. A chronicle base must contain all the chronicles characterizing the behaviors to monitor. Moreover, the base must be updated each time the supervised system evolves physically or structurally. An expert is often needed to create the chronicle base, and that makes the creation and the maintenance of the base very expensive. That is why we are working on an automatic method to acquire the base.

  Applications generally deal with system monitoring (telecommunication network) and video-surveillance (underground, bank, etc...).

Developing diagnosis methodologies is not enough, especially when on-line monitoring is required. Two related concerns must be tackled, and are the topics of current research in the team:

- The ultimate goal is usually not merely to diagnose, but to put back the system in some acceptable state after the occurrence of a fault. That calls for considering the repair capabilities of a system and designing the diagnoser in such a way that the diagnoses are sufficiently discriminating to be able to trigger a valid repair procedure.
- When designing a system and equipping it with diagnosis capabilities, it may be crucial to be able to check off-line that the system will behave correctly, i.e. that the system is actually 'diagnosable'. Diagnosability is checked when two distinct faults (or one fault and the correct behavior) can never produce the same set of observations. A lot of techniques have been developed in the past (see Lafortune and colleagues [85]), essentially in automata models. Extending them to deal with temporal patterns, permanent faults, multiple faults, fault sequences and some problems of intermittent faults, or trying to relate such techniques with diagnosability of continuous systems, has been the main focus of our studies up to now. We intend now to study diagnosability and repairability capabilities together, in order to build self-healing systems.

3.2. Machine learning and data mining

The machine learning and data mining techniques investigated in the group aim at acquiring and improving models automatically. They belong to the field of machine or artificial learning [47]. In this domain, the goal is the induction or the discovery of hidden objects characterizations from their descriptions by a set of features or attributes. Our work has been grounded on Inductive Logic Programming (ILP) for several years but we are now also investigating the use of data-mining techniques.

We are especially interested in structural learning which aims at making explicit dependencies among data where such links are not known. The relational (temporal or spatial) dimension is of particular importance in applications we are dealing with, such as process monitoring in health-care, environment or telecommunications. Being strongly related to the dynamics of the observed processes, attributes related to temporal or spatial information must be treated in a special manner. Additionally, we consider that the legibility of the learned results is of crucial importance as domain experts must be able to evaluate and assess these results.

The discovery of spatial patterns or temporal relations in sequences of events involve two main steps: the choice of a data representation and the choice of a learning technique.

3.2.1. Temporal and spatial data representation

Temporal data comes mainly in two forms: time series and temporal sequences. Time series are sequences of real values which are often the result of sampling a signal at a regular rate. Temporal sequences are series of symbolic events being either ordered lists of simple events denoting the simple precedence temporal relation or list of timestamped events which provide more precise temporal information.

Temporal sequences come often from the abstraction of time series data. This is especially true for monitoring data recorded by sensors located on the observed system. For example, an ECG can be viewed as a time series of real values, i.e. the raw sampled signal, or as a series of time stamped events describing the waves that appears on the raw signal, e.g. P-waves, QRS-complexes, T-waves, etc. The main issue of data abstraction is to select the relevant features that will represent the raw data and the granularity of abstraction that are the most convenient for the particular task. On the one hand, data should be sufficiently abstracted to enable an efficient (learning) computation, and on the other hand, they should be precise enough in order to avoid omitting essential information that could affect the accuracy of learned models.

Since they are often noisy and some features are difficult to detect, we have mainly investigated signal processing techniques for abstracting health-care data. In collaboration with the LTSI University of Rennes 1, we have studied, among other techniques, wavelet transforms and neural networks for wave detection and classification [2]. For telecommunication data, where the semantics were not so obvious, we have enhanced discretization methods transforming a time series in a sequence of linear segments which are next associated to specific symbols [78].

Generally, simple sequences of symbolic locations are not sufficient to model spatial information. This is why graphs are often used to represent spatial data. In this model, nodes represent a geographic area and edges spatial relations between these areas. Sometimes the data are such that trees, which are specific graphs, can be used. This is the case for hydrological modeling, for instance. The considered area is split into sub-areas whose granularity can be adapted to particular tasks. In our case, the relationships between sub-areas model the runoff transfer from an area to a lower one. This kind of representation is used to compute simulations of water and pesticides transfer, in tight collaboration with the SAS-INRA research group.

3.2.2. Learning from structured data

We distinguish supervised and unsupervised learning methods. A learning method is supervised if samples of objects to be classified are available and labeled by the class they belong to. Such samples are often called *learning examples*. If the examples cannot be classified a priori, the learning method is unsupervised. Kohonen maps, association rule extraction in data mining or reinforcement learning are typical unsupervised learning methods. From another point of view, learning methods can be symbolic, such as inductive rule or decision tree learning, or numerical, such as artificial neural networks. We are mainly interested in symbolic supervised and unsupervised methods. Furthermore, we are investigating methods that can cope with temporal relationships in data. In the sequel, we will give some details about relational learning, relational data-mining and data streams mining.

Relational learning

Relational learning, called inductive logic programming (ILP) in the past, is a research topic at the intersection of machine learning, logic programming and automated deduction. The main goal of relational learning is the induction of classification or prediction rules from examples and from domain knowledge. As relational learning relies on first order logic, it provides a very expressive and powerful language for representing learning hypotheses especially those learnt from temporal data. Furthermore, domain knowledge represented in the same language, can also be used. This is a very interesting feature which enables taking into account already available knowledge and avoids starting learning from scratch.

Concerning temporal data, our work is more concerned with the application of relational learning rather than developing or improving the techniques. Nevertheless, as noticed by Page and Srinivasan [76], the target application domains (such as signal processing in health-care) can benefit from the adaptation of relational learning scheme to the particular features of the application data. Therefore, constraint programming has been associated to ILP for relational learning in order to infer numerical values efficiently [87]. Extensions, such as QSIM [64], have also been used for learning a model of the behavior of a dynamic system [55]. Precisely, we investigate how to associate temporal abstraction methods to learning and to chronicle recognition. We are also interested in constraint clause induction, particularly for managing temporal aspects. In this setting, some variables are devoted to the representation of temporal phenomena and are managed by a constraint system [79] in order to deal efficiently with the associated computations (such as the covering tests, for example).

Concerning environmental data, we have investigated tree structures where nodes are described by a set of attributes. Our goal is to find patterns expressed as sub-trees [43] with attribute selectors associated to nodes.

Data mining

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 deserved a lot of interest in the last 10 years. For instance, many enhancements have been proposed to the well-known Apriori algorithm [29]. It is based on a levelwise generation of candidate patterns and on efficient candidate pruning based on a notion of minimal relevance, usually related to the frequency of the candidate pattern in the dataset (i.e. the support): the most frequent patterns should be the most interesting. Later, Agrawal and Srikant proposed a framework for "mining sequential patterns" [30], an extension of Apriori where the order of elements in patterns are considered.

In [73], Mannila and Toivonen extended the work of Aggrawal et al. by introducing an algorithm for mining patterns involving temporal episodes with a distinction between parallel and sequential event patterns. Later, in [52], Dousson and Vu Duong introduced an algorithm for mining chronicles. Chronicles are sets of events associated with temporal constraints on their occurrences. They generalize the temporal patterns of Mannila and Toivonen. The candidate generation is achieved by an Apriori-like algorithm. The chronicle recognizer CRS [50] is used to compute the support of patterns. Then, the temporal constraints are computed as an interval whose bounds are the minimal and the maximal temporal extent of the delay separating the occurrences of two given events in the dataset. Chronicles are very interesting because they can model a system behavior with sufficient precision to compute fine diagnoses, they can be extracted reasonably efficiently from a dataset and they can be efficiently recognized on an input data stream.

Relational data-mining [25] can be seen as a generalization of these works to first order patterns. Interesting propositions have been made in this field, for instance the work of Dehaspe for extracting first-order association rules which have strong links with chronicles. Another interesting research concerns inductive databases which aim at giving a theoretical and logical framework to data-mining [65], [48]. In this view, the mining process is considered as querying a database containing raw data as well as patterns that are implicitly coded in the data. The answer to a query is computed, either directly if the solution patterns are already present in the database, or computed by a mining algorithm, e.g. Apriori. The original work is concerned with sequential patterns only [71]. We have investigated an extension of inductive database where patterns are very close to chronicles [92].

Mining data streams

During the last years, a new challenge has appeared in the data mining community: mining from data streams [27]. Data coming for example from monitoring systems observing patients or from telecommunication systems arrive in such huge volumes that they cannot be stored in totality for further processing: the key feature is that "you get only one look at the data" [60]. Many investigations have been made to adapt existing mining algorithms to this particular context or to propose new solutions: for example, methods for building synopses of past data in the form of or summaries have been proposed, as well as representation models taking advantage of the most recent data. Sequential pattern stream mining is still an issue [75]. At present, research topics such as, sampling, summarizing, clustering and mining data streams are actively investigated.

A major issue in data streams is to take into account the fact that the process generating data is dynamic, i.e. that the underlying model is evolving, and so the extracted patterns have to be adapted constantly. This feature, known as *concept drift* [93], [67], occurs within an evolving system when the state of some hidden system variables changes. This is the source of important challenges for data stream mining [59] because it is impossible to store all the data for off-line processing or learning. Thus, changes must be detected on-line and the current mined models must be updated on line as well.

4. Application Domains

4.1. Software components monitoring

Web-services, i.e. services that are provided, controlled and managed through Internet, 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. Anyway, 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, become then a crucial concern.

We have been working on this problem within the WS-DIAMOND project [22],a large European funded project involving eight partners in Italy, France, Austria and Netherlands http://wsdiamond.di.unito.it/. Our own work consisted in two distinct contributions.

The first issue has been to extend the decentralized component-oriented approach, initially developed for monitoring telecommunication networks [4] to this new domain. To this end we have proposed the concept of distributed chronicles, with synchronization events, and the design of an architecture consisting of distributed CRSs (Chronicle Recognition Systems) communicating their local diagnoses to a broker agent which is in charge of merging them to compute a global diagnosis. During his thesis, X. Le Guillou developped two approaches before solving this problem [68], [69], [70].

We are also involved in formally characterizing intended properties of the system such as the diagnosability (i.e. the capability to detect and explain an error in due time) and repairability (i.e. the capability to get the system back to correctness, in due time), managing to relate them in order to define a so-called 'self-healability' property which will ensure the system is self-healing, i.e. it can always match a set of observations to an adequate repair procedure, which will resolve any of the faults which are consistent with the observations [45].

4.2. Telecommunication monitoring

Telecommunication networks are a good example of large-scale complex systems. Moreover, monitoring telecommunication networks is an important task to ensure a good quality of service.

We are focusing on another important issue, how to detect intrusion attempts in local networks or web servers. Two main techniques have been used so far: signature-based and anomaly-based detection. The first one makes use of a signature database which contains specific patterns that can be related to attacks. The second one makes use of a model representing the normal behavior of the observed system (local network, web server, etc.). Within the first method, any deviation from the behavior inferred by the model can be associated with a faulty state. Data mining is widely used to extract signatures from data. Concerning the second method data stream mining, e.g. dynamic clustering [16], can be used for detecting discrepancies between the system actual behavior and a normal one.

Another important issue for telecommunication networks monitoring is to predict the subjective quality of monitoring and diagnosis from collected technical data. If many false alarms are generated or many faults are missed during monitoring, new diagnostic knowledge must be acquired i.e. the models should be updated. Furthermore, in the context of data streams new data arrive continuously and cannot be stored in totality. Thus, the model used for diagnosis must be continuously adapted and knowledge acquisition must be performed *on the fly* concurrently with diagnosis [13].

4.3. Environmental decision making

The need of decision support systems in the environmental domain is now well-recognized. It is especially true in the domain of water quality. For instance the program, named Bretagne Eau Pure, was launched a few years ago in order to help regional managers to protect this important resource in Brittany. 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. 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 the farmer decisions. One of our main contribution 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 spatio-temporal data. The learned rules are then analyzed in order to recommend actions to improve a current situation.

This work is currently done in the framework of the APPEAU project, funded by ANR and of the ACCASYA project, funded by ANR, having started at the beginning of 2009. We are also involved in the PSDR GO CLIMASTER project, that started in september 2008 and will end in 2011.. CLIMASTER stands for "Changement climatique, systèmes agricoles, ressources naturelles et développement territorial" and is dedicated to the impact of climate changes on the agronomical behaviors in west of France (Grand Ouest). PSDR GO stands for "Programme Pour et Sur le Développpement Régional Grand Ouest".

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

5. Software

5.1. Introduction

The pieces of software described in this section are prototypes implemented by members of the project. They are not available through the APP. Any interested person should contact relevant members of the project.

5.2. Calicot: intelligent cardiac monitoring

Participant: René Quiniou [Correspondant].

CALICOT (Cardiac Arrhythmias Learning for Intelligent Classification of On-line Tracks) is a software that takes as input several signals coming from sensors and that delivers as output fault states or diseases that were diagnosed by recognizing characteristic temporal patterns on the monitored signals. CALICOT is devoted to monitoring cardiac patients and diagnosing cardiac arrhythmias.The software is mainly implemented in Java with a few modules in Prolog. The main features of CALICOT are:

- a base of signal processing algorithms for abstracting signals into time-stamped symbolic events,
- a base of chronicles that are used on line by the chronicle recognizer[1] Chronicles are discriminant temporal patterns related to arrhythmias. They are learned automatically off line from signal examples by using ILP techniques,
- a pilot that adapts the behavior of the system to the monitoring context: noise on signals, patient's state, relevant arrhythmias, etc,
- a graphical interface that displays the recognized patterns on the signal curves and shows the related diagnoses.

The software prototype has been completely redesigned this year and offers new functionalities. The following website gives many details about the goals, the conception and the implementation of Calicot: http://www.irisa.fr/dream/Calicot/.

5.3. CarDeCRS and Matrac: two chronicle-based distributed diagnosis platforms

Participants: Xavier Le Guillou, Laurence Rozé [Correspondant].

CARDECRS [68] and MATRAC are two generic diagnosis platforms which have been developed within the DREAM project.

[1] CRS (Chronicle Recognition System) from France Telecom R & D.

Those platforms aim at monitoring complex systems, thanks to specifically designed distributed chronicles. Their goal is to establish an end-to-end system, acting from the acquisition of the chronicles to the final diagnosis, with no need for extra software.

The implementation mainly relies on a chronicle recognition system called CRS, developed by C. Dousson and his team, from France Telecom Recherche & Developpement Lannion (for the diagnosis part), and an RMI/multithreaded based environment (for the low level part).

The main difference between those two approaches reflects the potentially dynamic aspect of the system to monitor:

- MATRAC uses an interaction model that lists all the exchanges between the participants of a topologically static distributed system;
- CARDECRS is interaction model-free and is adapted to topologically dynamic distributed systems.

The platforms were experimented in the framework of the WS-Diamond European project for monitoring web services between 2005 and 2008. New perspectives for those platforms consist in:

- enriching them with repair capabilities, which is currently the main research interest of the participants;
- making the software more user-friendly, by developing a new GUI capable of generating skeletons of web services and chronicles automatically.

5.4. LogAnalyzer: a workbench for analyzing log streams

LOGANALYZER is an experimental workbench for validating our scientific results on adaptive intrusion detection [19]. This software offers two possible uses: off-line and interactive computation of intrusion diagnosis, and on-line intrusion detection dedicated to the processing of massive data streams.

The software is organized in three layers:

- the adaptive intrusion detection layers: it constitues the core API which implements the multi-diagnoser adaptive diagnosis approach (see 6.2.1).
- the applicative layer: it applies the core API to the structure of the log data. In our software, we developed this layer to detect intrusions in Apache HTTP logs,
- the graphic user interface layer for off-line and interactive analysis.

The adaptive diagnosis layer is not dedicated to Apache HTTP logs processing. This layer is stand alone and can be easily instantiated to other kinds of streams of structured data. A multi-threaded framework organizes multi-diagnosers and meta-diagnoser processing and adaptations triggering. The concrete implementations of diagnosers are components of an application layer dedicated to a specific kind of data stream.

The GUI the software provides several applicative features:

- anonymization: logs may be anonymized to cope with (client and server) privacy issues,
- server activity reporting: a module draws the curves of the evolution of the web server activity (number of requests per time unit, number of bot requests, ...),
- access logs browsing and annotating: three log views (sequential view, transactional view and server image view) are provided to aid the user browse huge amount of data. In particular, this tool can help her/him to identify intrusions manually and to label them.

The software has been developed in C++/Qt4 and is still under improvement. The following website is devoted to the presentation of the LOGANALYZER: http://www.irisa.fr/dream/LogAnalyzer/.

5.5. QTempIntMiner: quantitative temporal sequence mining

QTEMPINTMINER (Quantitative Temporal Interval Miner) is a software that implements the QTEMPINTMINER algorithm presented in section 1.

The software is mainly implemented in Matlab. It uses the Mixmod toolbox [39] to compute multi-dimensional Gaussian distributions. The main features of QTEMPINTMINER are:

- a generator of synthetic noisy sequences of temporal events,
- an implementation of the QTEMPINTMINER algorithm,
- a graphical interface that enables the user to generate or import data set and to define the parameters of the algorithm. The interfaces although displays the extracted temporal patterns.

This year, the software has been updated to mine long sequences of temporal events such as cardiac monitoring annotations. Long sequences are transformed into short temporal sequences that are used as instances of the QTEMPINTMINER algorithm. The software is currently applied to the characterisation of cardiac arrythmias.

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

5.6. Sacadeau: qualitative modeling and decision-aid to preserve the water quality from pollutants as herbicides

SACADEAU is a software that implements the SACADEAU transfer model presented in section 7.1. 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 is mainly implemented in Java.

6. New Results

6.1. Diagnosis of large scale discrete event systems

Participants: Marie-Odile Cordier, Christine Largouet, Xavier Le Guillou, Sophie Robin, Laurence Rozé.

The problem we deal with is monitoring complex and large discrete-event systems (DES) such as telecommunication networks or web services. Two approaches are used in our research group. The first one consists in representing the system model as a discrete-event system by an automaton. The diagnostic task consists in determining the trajectories (a sequence of states and events) compatible with the sequence of observations. From these trajectories, it is then easy to determine (identify and localize) the possible faults. In the second approach, the model consists in a set of predefined characteristic patterns. We use temporal patterns, called chronicles, represented by a set of temporally constrained events. The diagnostic task consists in recognizing these patterns by analyzing the flow of observed events.

6.1.1. Distributed and incremental diagnosis of discrete-event systems

One of the main difficulties of discrete event system modeling is the intractable size of the model and the huge number of states and trajectories to be explored. To cope with this problem, we proposed to use a decentralized approach [4] which allows us to compute on-line diagnosis without requiring the computation of the global model. Given a decentralized model of the system and a flow of observations, the program computes the diagnosis by combining local diagnoses built from local models (or local diagnosers).

In real systems, generally the observed events do not exactly correspond to the emitted events. Thus, instead of only considering partially ordered observations (as in [4]), we proposed to represent the uncertainty on emitted observations by an automaton and extended the decentralized approach to cope with this new representation. In order to deal with on-line diagnosis, we then proposed to slice the observation flow into temporal windows, introduced the concept of automata chains to represent the successive observation slices, and proposed an algorithm to compute the diagnosis in an incremental way from these diagnosis slices [62]. A revised version of a paper, written on this issue in collaboration with A. Grastien, currently researcher at ANU (Australian National University), is currently submitted to AIJ (Artificial Intelligence Journal).

6.1.2. Distributed monitoring with chronicles

The formalism of chronicles has been proposed a few years ago by C. Dousson in order to monitor dynamic systems in real time [53]. A chronicle, characterized by a set of temporally constrained events, describes the situations to monitor. Efficient algorithms for on-the-fly chronicle recognition exist, but only in a centralized way up to now. Our main contribution is an extension of the chronicle-based approach to deal with distributed systems.

First, we have extended the formalism proposed by C. Dousson by defining a distributed chronicle model [70]. The standard chronicle description language is enriched with synchronization constraints.

Then, we have proposed a decentralized monitoring architecture in which each component is equipped with a local diagnoser. A global diagnoser (also called *broker*) is in charge of merging the local diagnoses, by checking the synchronization constraints of the local diagnoses.

This work began in the context of the WS-DIAMOND European project where the overall goal of the project was to monitor and diagnose the processing of a request sent to a web service, which in turn usually requests other services to complete the task: it is common that a fault, occurring in a service, propagates to other services. A new formalism for describing distributed chronicles has been designed. It extends the classical chronicle formalism and is well-suited to describe both the normal or abnormal behaviors of each web service and the communication with the other services.

Each local diagnoser relies on a chronicle recognition system engine based on the CRS application developed by C. Dousson [50], and a chronicle base representing the local scenarios which are of interest wrt the diagnostic task. Concerning the global diagnoser, two approaches have been developed depending on the availability of a model of interactions between services. In the first one [9], the global diagnoser itself relies on a chronicle recognition system, its chronicle base being built off-line from the interaction model. In the second approach, when no model of interactions is available, for instance in case of a dynamic choreography,the global diagnoser itself has to ensure on-line the correspondence between the synchronization variables from different services and to build the diagnosis tree [69].

Two platforms corresponding to the two approaches have been developed (cf. section 5.3).

A paper comparing the two approaches has been submitted to a special issue of the french RIA journal and an english version of this paper is nearly ready to be submitted.

We are currently working in two directions. The first one consists in tackling the difficult problem of chronicle acquisition by addressing firstly the automatic construction of chronicle skeletons from the workflow of the different web services.

The second one consists in connecting the diagnosis and the repair tasks by integrating distributed repair processes within our decentralized diagnosis architecture. Interleaving diagnosis and repair in the framework of intelligent networks is one of the issue we investigate in cooperation with the planning and diagnosis group of NICTA's Canberra Research Laboratory and the team GEMO of Inria Saclay. A proposal for an Inria Associate Team has been submitted. Extending the distributed chronicle recognition approach for diagnosis to incluse repair actions is the first step we are currently working on to reach this objective.

6.1.3. Diagnosability and self-healability of discrete-event systems

After having addressed the problem of diagnosability of complex supervision patterns by [66], and proposed a common theoretical paradigm for diagnosability of both discrete and continuous systems, based on the common concept of signature [46], we have started a new line of collaborative work with the Disco/Laas research group, within our common participation in the WS-DIAMOND project. Based on the signature-based definition of diagnosability alone, and introducing the concept of repairability (i.e. the existence of at least one applicable repair procedure for each fault that may occur in the system), our goal was now to consider jointly diagnosis and repair capabilities in complex, discrete-event or continuous, systems. That led us to define formally the "self-healability" property of such systems [44], [45]: a system is said to be self-healable if and only if there exists a set of 'macro-faults' (i.e. identified situations in which several candidate faults may still not be discriminated) which can be matched to at least one repair procedure. We extended this definition to cope with conditional repairs and temporal constraints; this work has not been yet published. We intend to extend this first work which focuses on a centralized system, to deal with distributed systems, and with temporally related events (expressed as chronicles).

6.1.4. Scenario patterns for exploring qualitative ecosystems

This work aims at giving means of exploring complex systems (in our case, ecosystems). We propose to transform environmental questions about future evolution of ecosystems into queries that could be submitted to a simulation model. When dealing with environmental problems, scenarios are widely used tools for evaluating future evolution of ecosystems given policy options, potential climatic changes or impacts of catastrophic events. If the scenarios are generally expressed in natural language, when working with a model describing the ecosystem, it is necessary to transform them into formalised queries that can be given as input to the model.

We propose to model the system in a distributed qualitative way and we define a high-level language to query the model. The system behavior is represented as a discrete-event system, described by a set of interacting timed automata. The ecosystem is represented as a set of interacting subsystems and the global model obtained by composition on shared events. This technique is particularly suited to representing large-scale systems such as ecosystems. To explore the system, we define generic patterns, associated to the most usual types of scenarios, and translate them into temporal logic formula. The answer is computed thanks to model-checking techniques, that are efficient for analysing large-scale systems. Five generic patterns have been defined using TCTL (Timed Computation Tree Logic) : *WhichStates*, *WhichDate*, *WhichStates-Si*, *Stability*, *Safety*, three of these patterns have been implemented using the model-checker UPPAAL [14]. We experimented our approach on a marine ecosystem under fishing pressure. The (simplified) model describes the tropho-dynamic interactions between fish trophic groups as well as interactions with the fishery activities and with an environmental context. The results concern the impact of a fishery growing policy on the ecosystem [10].

6.2. Machine learning for model acquisition

Participants: Marie-Odile Cordier, Thomas Guyet, Christine Largouët, Alice Marascu, Véronique Masson, René Quiniou.

Model acquisition is an important issue for model-based diagnosis, especially as modeling dynamic systems. We investigate machine learning methods for temporal data recorded by sensors or spatial data resulting from simulation processes.

6.2.1. Learning and mining from sequential and temporal data

Our main interest is extracting knowledge, especially sequential and temporal patterns or prediction rules, from static or dynamic data (data streams). We are particularly interested in detecting and dealing with concept change in data stream mining in order to be able to adapt on-line the models used for diagnosis.

Mining temporal patterns with numerical information

This work aims at exploring a new solution to learn the temporal extents of events in chronicle models. Mining sequential patterns as well as extracting frequent temporal episodes has been widely explored. However, in a monitoring context, numerical information associated to event duration or to the delay between event occurrences is very meaningful for discriminating between faults or diseases. Due to the added complexity of such patterns, numerical temporal pattern mining methods have been less explored than methods for mining sequential patterns using the simpler notion of precedence.

We focus on mining temporal interval patterns from time-series databases or from sequence databases where a temporal sequence is a set of events with time-stamped begin and end. Our first proposition relied on an algorithm that processes temporal sequences represented by hyper-cubes [63]. The candidate generation uses the classical *Apriori* schema [29] enriched by constraints on temporal intervals. Frequent temporal pattern selection estimates the interval distribution to extract the statistically significant temporal intervals. The algorithm is implemented in Matlab (see section 5.5).

We are currently working on improving the algorithm complexity, precisely on how to estimate the distribution of temporal intervals more efficiently. The method is evaluated on cardiac monitoring data for extracting typical patterns that can be associated to cardiac arrhythmias. We also plan to mine electrical consumption data in the context of a collaboration with EDF (Électricité de France).

Mining sequential patterns from data streams

During her thesis [74], Alice Marascu has proposed methods for data stream processing and analysis. Precisely, she has devised CLUSO, a method that computes clusters from a sequence stream and extracts frequent sequential patterns from the clusters and maintain their history in order to summarize the stream. During her post-doctoral stay, extensions to the CLUSO method will be investigated in order to deal with more complex patterns, in particular temporal patterns with numerical information, to detect and characterize changes in the data, such as trends and deviations, to take into account the quality of the data (missing values, noisy data, etc.).

Dealing with concept change in data stream mining

We are investigating a multi-diagnoser approach for detecting changes from data stream and for adaptating the diagnosers of the surveillance systems.

In this framework, several diagnosers process the input data stream to construct their own diagnosis. The global diagnosis is constructed by making the fusion of the diagnoses. We use the Demspter-Shafer evidence theory [88] to model the diagnosis and to operate their fusion. The diagnosers are themselves continuously (meta-)diagnosed to assess whether a change has occurred and to decide whether one or several diagnosers should be adapted. The meta-diagnosis is computed by checking predefined relations between the diagnosers to detect faulty diagnoses. These relations takes the form of integrity constraints that express information redundancy between diagnosers. If a set of diagnoses does not satisfy an integrity constraint, it means that two or more diagnosers disagree given some input observations. In such a situation, the faulty diagnosers are detected by comparing their own diagnosis to the global diagnosis decision. The involved diagnosers have the ability to adapt themselves to concept change with respect to the recommended diagnosis decision.

The method has been evaluated on an adaptive intrusion detection system monitoring the queries addressed to a web server [13], [23], [19]. Several diagnosers have been implemented by taking different views of the input data: distribution of character in single queries or sessions (set of queries associated to the same user in a logical temporal window), distribution of tokens in queries or sessions, distribution of bi- grams (sequence of two characters), etc. The system is able to detect changes due to the arrival of new kinds of attacks not previously encountered by the system and to adapt the diagnoser model (the distributions of objects they rely upon) accordingly. The adaptive system shows a better performance (higher detection rate and lower false positive rate) than a non adaptive version of the same system.

We have also proposed a novel intrusion detection method that detects anomalies in an on line and adaptive fashion through dynamical clustering of unlabelled audit data streams [16], [12], [11]. The framework shows self-managing capabilities: self-labeling, self-updating and self-adapting. The method is based on a recently developed clustering algorithm, Affinity Propagation (AP) [57]. Given an audit data stream, our method identifies outliers (suspicious accesses or requests) with AP. If an outlier is identified, it is then marked as suspicious and put into a reservoir. Otherwise, the detection model is updated until a change is detected, triggering a new model rebuilding through clustering. The suspicious examples are considered as really anomalous if they are marked as suspicious once again after model rebuilding. Thus, our detection model does not need labeled data and can be used on line. The method has been evaluated on a very large size of real HTTP Log data collected at INRIA as well as on a subset of the KDD 1999 benchmark data. The experimental results show that the method obtains better results than three static methods in terms of effectiveness as well as efficiency.

6.2.2. Learning decision-oriented rules from simulation data

In the framework of the SACADEAU project, our aim is to build decision support systems to help catchment managers to preserve stream-water quality [3]. In collaboration with INRA researchers, three actions have been conducted in parallel [20].

- The first one consisted in *building a qualitative model* to simulate the pesticide transfer through the catchment from the time of its application by the farmers to the arrival at the stream. 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. Though INRA is the main contributor, we have participated actively to its realization. This model has been implemented and used for simulation. An in-depth analysis of many simulation results leads us to refine the model. A paper on the SACADEAU model appeared in [5].
- The second action consisted in *identifying some of the input variables* as main pollution factors and in learning rules relating these pollution factors to the stream pesticide concentration. During the learning process, we focus on actionable factors, in order to get helpful rules for decision-makers. Moreover, we take a particular interest in spatial relations between the cultivated plots and in the characteristics of crop management practices [90]. To deal with the complex spatial relations existing between the catchment plots, we experimented two learning approaches. The first one consisted in extending Inductive Learning Programming to tree structured patterns. The choice of ILP has been motivated by the aim to get easy-to-read and explicative rules. This was done using the Aleph software (after a first experiment with ICL). The second approach consisted in propositionalizing the learning examples and using a propositional learning process, namely CN2. A comparison of these two approaches and an analysis of the results can be found in R. Trepos's PhD thesis [91]. A paper has been submitted on this issue and a revised version will be submitted soon to the Environmental Modelling and Software journal.
- The final aim is to go beyond the simple use of classification rules for prediction, by assisting the user in the post-analysis and in the exploitation of a large set of rules. The goal is then to find advices in order to reduce pollution whereas the learned rules are classification rules predicting if a given farmer strategy or climate leads to a polluted or not polluted situation. The propositional rules learned in the second step are automatically analysed by the algorithm DAKAR (Discovery of Actionable Knowledge And Recommendations) [89] to propose actions well-suited to improve a given situation. Another way to help the experts in dealing with a large set of rules is using visualization techniques which have been developed. The two approaches to learn rules, the algorithm DAKAR and visualization techniques are fully described in [91].

In the APPEAU context, the idea is to study how the SACADEAU-style model can be used in a more generic way and to compare, and possibly unify, our work with what is done by our partners from SAS/INRA concerning nitrate transfer. The main difference between the two contexts, pesticide in one hand, nitrate in the other hand, is that the spatial issues (water paths, connected plots) are the most important one in the first case, when the temporal issues are the important ones in the second case. Two actions are planned:

- The first action consisted in using the TNT2 model, an hydrological model build by our colleagues from SAS/INRA and dedicated to topography-based simulation of nitrogen transfer and transformation, in order to run scenarios and get interesting simulation tesults to be exploited in a next phase.
- The second action just started and consists in analysing the simulation results got in the previous step in order to get information on the main influencial variables. The idea is to use adapted learning and data mining tools.

In the ACCASYA context, the challenge is to transform these simulation models into decision-aid tools, able to answer queries about future evolution of ecosystems. Two issues are studied :

- When dealing with environmental problems, scenarios are widely used tools for evaluating future evolution of ecosystems given policy options, potential climatic changes or catastrophic event impacts. If the scenarios are generally expressed in natural language (especially in the field of environmental sciences), when using ecological model simulators, it is necessary to transform them into formalised queries that can be given as input to the model. A first step in this direction is described in 6.1.4 (see also [10]). It concerns halieutic ecosystems but we intend to use a similar approach in the context of catchment management.
- Another issue consists in answering user queries by using user-oriented incremental learning. A first approach dedicated to incremental learning has been undertaken. The idea is to learn more interesting rules by selecting well-suited learning examples. The idea is to iteratively run the model : each cycle consists in (i) setting parameters after analysing earlier cycle results; (ii) running the model and getting new data; (iii) learning rules from these data; (iv) selecting the fittest rules according to given criterias; and adding them to the global theory. One of the key issues is the choice of the parameters from the previous cycles results. Currently, the choice is done (either automatically or guided by an expert) in order to produce rules improving the quality of the global theory. In our case, the weed control strategies appeared to be the most relevant parameters in the sense that they heavily impact the final results. We proposed to group the strategies in six clusters according to their characteristics, and to assign choice probabilities to each cluster according to distances. this protocol. This preliminary work has been presented at the SIDE/Inforsid workshop “Systèmes d’Information et de Décision pour l’Environnement” [17]. The next step is then to design and use more sophisticated criteria for choosing rules (step (iv) above).

  A thesis funded thanks to the ANR/ADD Accassya project will start at the beginning of 2010.

6.3. Diagnostic and causal reasoning

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

This work stems on [34], [35], [36] and is related to diagnosis where observed symptoms have to be explained by faults. The previously existing proposals are ad-hoc or, as in [40], [61], they are too close to standard logic in order to make a satisfactory diagnosis. Our proposal starts from propositional logic and introduces *causal formulas*, built on causal atoms such as (α *causes* β) intended to mean: “α causes β”. There are also *taxonomy atoms* such as (α IS_A β) (roughly α is of kind β). The user describes the system thanks to classical formulas, causal formulas, and taxonomy atoms. Then, a set of inference rules computes where β can be explained by some α. These explanations are formalized by *explanation atoms* of the kind (α *explains* β *if_possible* $\{\gamma_1, \cdots, \gamma_n\}$): when all the γ_i’s are possible together in the context of the given data, then *(α explains β)*. The simplest generating rule states that if (α *causes* β) then (α *explains* β if_possible $\{\alpha\}$). One rule states that $\alpha \rightarrow \beta$ if (α *causes* β), thus here “if_possible$\{\alpha\}$” is equivalent to “if_possible$\{\alpha, \beta\}$”. Despite this rule, the connexions between *causes* and $\rightarrow$ are rather loose. These connexions are a known difficulty in causal formalisms. Our solution avoids too much connexions between these two operators, while keeping the formalism simple and natural enough. Our system [37], [38] deals with alternative effects without too much computational effort.

This kind of formalism is well fitted for an implementation in logic programming, precisely in the Answer Set Programming (ASP) [31] version. Indeed, not only ASP is closed to classical logic, thus allowing an immediate translation of the rules into an ASP language, but also ASP is known to be good in finding paths in graphs. ASP is a fast evolving area and the systems (e.g. DLV [72] and claspD [54]) are becoming more and more efficient and user friendly. This last aspect is important since, even if an ASP program is written in a language allowing to express directly rules of the kind of our formalism, writing an ASP program generally happens to be more bothering than expected due to some limitations of the systems: reduced kind of objects (till very recently: no lists or sets, reduced use of functions,...) and difficulty in writing real "sub programs" (generally, a specific part of program should be rewritten differently for each occurrence). These two limitations are not crucial from a theoretical point of view. They are due to the novelty of this branch of logic programming, and are now in the way of being overcome. In particular, DLV-Complex [41] allows lists and function symbols, plus and easy way to built "pre definite predicates" in another language when necessary. Also, DLV Templates [42] allows "templates" which make "routines" easy to write. These are crucial improvements that simplify the task of a user. Then, gaining efficiency can be done by modifying the original rules, while keeping them close to their logical expression, and again this task is easier when using some "routines". Then, ASP is really easy to use and, since the core formalisms '(such as DLV or claspD) are also evolving rapidly, we can deal with more complex situations. Due to the natural way ASP rules can now be written, a user can directly write a program adapted to his needs without the need to learn a dedicated tool. Part of these results should appear in the "Logic programming and ASP" chapter in a book called "Panorama actuel de l'intelligence artificielle, ses bases méthodologiques, ses développements" , supervided by Pierre Marquis, Odile Papini and Henri Prade.

7. Contracts and Grants with Industry

7.1. SACADEAU-APPEAU: Decision-aid to improve streamwater quality

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

The SACADEAU project (Système d'Acquisition de Connaissances pour l'Aide à la Décision pour la qualité de l'EAU - Knowledge Acquisition System for Decision-Aid to Improve Streamwater Quality) was funded by INRA (French institute for agronomy research) from October 2002 to October 2005. The main partners were from INRA (SAS from Rennes and BIA from Toulouse) and from IRISA. We have continued to develop the SACADEAU model with our partners until now and a PdD thesis has been funded by INRA (ASC).

We are now involved in a new project, named APPEAU and funded by ANR/ADD, which started in February 2007. The APPEAU project aims at studying which politics, for which agronomic systems, are best adapted to improve water management. It includes our previous partners as well as new ones, mainly from INRA.

In this framework, our work aims at building a decision-aid tool to help specialists in charge of the catchment management to preserve the streamwater quality. 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. The simulation results are analyzed, thanks to classification and symbolic learning techniques, in order to discover rules explaining the pesticide concentration in the stream by the climate, the farmer strategy, the catchment topology, etc. and, finally, in order to build recommendation actions for a given situation. In the APPEAU context, the idea is to study how this kind of model can be used to simulate scenarios in a more generic way and to compare, and possibly unify, our work with what is done by our partners from SAS concerning nitrate transfer.

7.2. ACCASYA: Supporting the agro ecological evolution of breeding systems in coastal watersheds

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

The ACCASYA project (ACcompagner l'évolution Agro-écologique deS SYstèmes d'élevage dans les bassins versants côtiers) is funded by ANR/ADD and started at the beginning of 2009. The main partners are our colleagues from INRA (SAS from Rennes. One of the objectives is to develop modelling tools supporting the management of ecosystems, and more precisely the agro ecological evolution of breeding systems in coastal watersheds. In this context, the challenge is to transform existing simulation tools (as SACADEAU or TNT2 into decision-aid tools, able to answer queries or scenarios about the future evolution of ecosystems.

We are also involved in a PSDR GO (programmes de recherche "Pour et Sur le Développement Régional Grand Ouest") project, named CLIMASTER that started in 2009. Our participation in this project is planned in 2011. For the moment, we participate to the general meetings and discuss the main scientific orientations to be ready to start.

7.3. ManageYourSelf: diagnosis and monitoring of embedded platforms

Participants: Marie-Odile Cordier, Xavier Le Guillou, 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 diagnostics and repairs on a fleet of mobile smartphones and PDAs. The idea is to integrate on the mobile devices a small expert system and a knowledge base containing a set of rules, 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 will be sent to a server. In the longer term aim is to develop on the server a module capable of learning from the reports of all the phones and able to generate new rules.

8. Dissemination

8.1. Services to the Scientific Community

8.1.1. Journal editorial board

- *AAI: Applied Artificial Intelligence* (M.-O. Cordier).
- *ARIMA: Revue Africaine de la Recherche en Informatique et en Mathématiques Appliquées* (M.-O. Cordier).
- *Revue I3* (M.-O. Cordier).
- *Revue JESA (Journal Européen des Systèmes automatisés* (M.-O. Cordier).

8.1.2. Conference program committees and organizations

- Program committee chair RFIA'2010; program committee member for IJCAI'09 and DX'09 (M.-O. Cordier).
- Program committee member for RFIA'2010 (R. Quiniou).
- Co-organizer and program committee member of the workshop "Temporal data mining" at EGC 2009 (R. Quiniou with G. Hébrail, ENST, Paris and P. Poncelet, LIRMM Montpellier).

8.1.3. Scientific and administrative boards

- ECCAI (European Coordinating Committed for Artificial Intelligence) board member, in charge of the new ECCAI fellows nomination: M.-O. Cordier

- ECCAI fellow + Honorific member of AFIA (Association Fran caise d'Intelligence Artificielle) : M.-O. Cordier
- Member of "Agrocampus Rennes (ENSAR)" scientific board: M.-O. Cordier.
- Member of "Conseil de l'IFSIC" and "Comité scientifique de l'IFSIC": M.-O. Cordier
- Member of the selection committee for the "Prix de thèse AFIA 2009" (selecting the best French PhD thesis in the Artificial Intelligence domain)
- Member of the INRA DR2 selection committee in 2009.

8.2. Academic teaching

Many members of the DREAM team are also faculty members and are actively involved in computer science teaching programs in Ifsic, INSA and ENSAR. Besides these usual teachings DREAM is involved in the following programs:

- Master2 in computer science (IFSIC): *Module DSS: Apprentissage sur des données séquentielles symboliques* (M.-O. Cordier (coordinator), R. Quiniou).
- Master2 in computer science (IFSIC): *Module CDD: Gestion de larges collections de documents décrits* (R. Quiniou).
- Master 2 in computer science (IFSIC). *Module OCI : Interfaces graphiques en C++/GTKMM* (T. Guyet)

8.3. PhD committees

- Marie-Odile Cordier was member of the examining committees (thesis defense) of M. Bayoud, Univ. Toulouse (rapporteur), A. Marascu, Univ. Nice (rapporteur) and A. Mechouche, Univ. Rennes1, présidente.
- Marie-Odile Cordier is member of the thesis committees Lucie Clavel (Inra Toulouse).
- René Quiniou is member of the thesis committee of Brivael Trelhu (University of Rennes 1).

9. 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", vol. 15, n° 4, 2008, p. 351-367.

[2] G. CARRAULT, MARIE-ODILE. CORDIER, R. QUINIOU, F. WANG. *Temporal abstraction and Inductive Logic Programming for arrhyhtmia recognition from electrocardiograms*, in "Artificial Intelligence in Medicine", vol. 28, 2003, p. 231-263.

[3] M.-O. CORDIER, F. GARCIA, C. GASCUEL-ODOUX, V. MASSON, J. SALMON-MONVIOLA, F. TORTRAT, R. TRÉPOS. *A machine learning approach for evaluating the impact of land use and management practices on streamwater pollution by pesticides*, in "MODSIM'05 (International Congress on Modelling and Simulation)", MODELLING AND SIMULATION SOCIETY OF AUSTRALIA AND NEW ZEALAND (editor), December 2005.

[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", vol. 164, n° 1-2, 2005, p. 121-170.

Year Publications

Articles in International Peer-Reviewed Journal

[5] P. AUROUSSEAU, H. SQUIVIDANT, R. TRÉPOS, F. TORTRAT, C. GASCUEL-ODOUX, M.-O. CORDIER. *A plot drainage network as a conceptual tool for the spatialisation of surface flow pathways for agricultural catchments*, in "Computer and Geosciences", vol. 35, 2009, p. 276-288, http://www.sciencedirect.com/science/article/B6V7D-4TXF7YP-2/2/7d57a4059309e57c3230cf52ef4928d4.

[6] E. FROMONT, R. QUINIOU, M.-O. CORDIER. *Learning Rules from Multisource Data for Cardiac Monitoring*, in "International Journal of Biomedical Engineering and Technology (IJBET)", vol. 2, n^{o} 3, 2009, to appear.

[7] 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 and Software", 2010, p. 1-14, to appear.

[8] Y. YAN, P. DAGUE, Y. PENCOLÉ, M.-O. CORDIER. *A Model-based Approach for Diagnosing Faults in Web Service Processes*, in "International Journal of Web Services Research (IJWSR)", vol. 6, n^{o} 1, 2009, p. 87–110.

International Peer-Reviewed Conference/Proceedings

[9] M.-O. CORDIER, X. LE GUILLOU, S. ROBIN, L. ROZÉ. *Monitoring WS-CDL-based choreographies of Web Services*, in "DX'09 (20th International Worshop On Principles of Diagnosis), Stockholm, Sweden", june 2009.

[10] C. LARGOUET, M.-O. CORDIER, G. FONTENELLE. *Scenario templates to analyse qualitative ecosystem models*, in "18th World IMACS Congress and MODSIM09 (International Congress on Modelling and Simulation. Modelling and Simulation)", R.S. ANDERSSEN, R.D. BRADDOCK, L.T.H. NEWHAM (editors), 2009, p. 2129-2135, http://mssanz.org.au/modsim09, ISBN: 978-0-9758400-7-8.

[11] W. WANG, T. GUYET, S. KNAPSKOG. *Autonomic Intrusion Detection System*, in "Symposium on recent advanced in intrusion detection (RAID)", 2009, p. 359–361, http://dx.doi.org/10.1007/978-3-642-04342-0_24.

[12] W. WANG, F. MASSEGLIA, T. GUYET, R. QUINIOU, M.-O. CORDIER. *A General Framework for Adaptive and Online Detection of Web attacks*, in "international world wide web conference (WWW 2009)", 2009, p. 1141-1142, http://doi.acm.org/10.1145/1526709.1526897.

National Peer-Reviewed Conference/Proceedings

[13] T. GUYET, R. QUINIOU, M.-O. CORDIER, W. WANG. *Diagnostic multi-sources adaptatif Application à la détection d'intrusion dans des serveurs Web*, in "EGC 2009", january 2009.

[14] C. LARGOUET, M.-O. CORDIER. *Patrons de scénarios pour l'exploration qualitative d'un écosystème*, in "RFIA 2010 (Reconnaissance des Formes et Intelligence Artificielle)", 2010, http://rfia2010.info.unicaen.fr/, to appear.

[15] X. LE GUILLOU, M.-O. CORDIER, S. ROBIN, L. ROZÉ. *Surveillance de choregraphies de Web Services basees sur WS-CDL*, in "RJCIA'09 (9e Rencontres des Jeunes Chercheurs en Intelligence Artificielle), Hammamet, Tunisie", mai 2009.

[16] W. WANG, T. GUYET, R. QUINIOU, M.-O. CORDIER, F. MASSEGLIA, B. TROUSSE. *Online and Adaptive anomaly Detection: detecting intrusions in unlabelled audit data streams*, in "EGC 2009", 2009.

Workshops without Proceedings

[17] M.-O. CORDIER, T. GUYET, C. LARGOUET, V. MASSON, H.-M. SUCHIER. *Apprentissage incrémental de règles de décision à partir de données d'un simulateur*, in "Atelier INFORSID : Systèmes d'Information et de Décision pour l'Environnement", 2009, http://www.irisa.fr/dream/Pages_Pros/Thomas.Guyet/publis/SIDE2009_Cordier.pdf.

[18] M.-O. CORDIER, Y. MOINARD. *Remarks about the use of Answer Set Programming for a causal formalism*, in "MICRAC project Workshop (Toulouse)", jun 2009, http://www.irit.fr/MICRAC/colloque/index_C.html.

[19] T. GUYET, R. QUINIOU, M.-O. CORDIER. *LogAnalyzer : monitoring adaptatif d'un flux de données*, in "RFIA 2010 (Congrès Reconnaissance des Formes et Intelligence Artificielle)", 2010, Demonstration, to appear.

Scientific Books (or Scientific Book chapters)

[20] M.-O. CORDIER, M. FALCHIER, F. GARCIA, C. GASCUEL-ODOUX, D. HEDDADJ, L. LEBOUILLE, V. MASSON, J. SALMON-MONVIOLA, F. TORTRAT, R. TRÉPOS. *Chapitre 9 : Modélisation d'un transfert d'herbicides dans un bassin versant dans le cadre d'un outil d'aide à la décision pour la maîtrise de la qualité des eaux*, in "Construire la décision : démarches en agriculture, agro-alimentaire et espace rural", Quae, 2009, p. 41-55.

[21] R. QUINIOU, L. CALLENS, G. CARRAULT, M.-O. CORDIER, E. FROMONT, P. MABO, F. PORTET. *Intelligent adaptive monitoring for cardiac surveillance*, in "Computational Intelligence in Healthcare", L. C. JAIN (editor), LNAI, Springer Verlag, 2009, to appear.

[22] THE WS-DIAMOND TEAM. *Chapter 2: WS-DIAMOND: Web Services DIAgnosability, MONitoring and Diagnosis*, in "At your service: An overview of results of projects in the field of service engineering of the IST programme", J. MYLOPOULOS, M. PAPAZOGLOU (editors), Series on Information Systems, MIT Press Series on Information Systems, 2009.

Research Reports

[23] T. GUYET, R. QUINIOU, W. WANG, M.-O. CORDIER. *Self-adaptive web intrusion detection system*, n^o 6989, INRIA, 2009, http://hal.inria.fr/inria-00406450/en/, Technical report.

References in notes

[24] B. DUBUISSON (editor). *Diagnostic, intelligence artificielle et reconnaissance des formes*, Traité IC2 : Information - Commande - Communication, Hermes, 2001.

[25] S. DZEROSKI, N. LAVRAČ (editors). *Relational Data Mining*, Springer, Berlin, 2001.

[26] W. HAMSCHER, L. CONSOLE, J. DE KLEER (editors). *Readings in Model-Based Diagnosis*, Morgan Kaufmann, San Meteo, CA, Etats-Unis, 1992.

[27] C. AGGARWAL. *Data Streams: Models and Algorithms*, Advances in Database Systems, Springer, 2007.

[28] A. AGHASARYAN, E. FABRE, A. BENVENISTE, R. BOUBOUR, C. JARD. *Fault detection and diagnosis in distributed systems : an approach by partially stochastic Petri nets*, in "Discrete Event Dynamic Systems", vol. 8, n^{o} 2, Juin 1998, p. 203-231, Special issue on Hybrid Systems.

[29] 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, p. 207–216.

[30] 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, p. 3–14.

[31] C. BARAL. *Knowledge representation, reasoning and declarative problem solving*, Cambridge University Press, 2003, 2003.

[32] P. BARONI, G. LAMPERTI, P. POGLIANO, M. ZANELLA. *Diagnosis of large active systems*, in "Artificial Intelligence", vol. 110, 1999, p. 135-183.

[33] M. BASSEVILLE, M.-O. CORDIER. *Surveillance et diagnostic de systèmes dynamiques : approches complémentaires du traitement de signal et de l'intelligence artificielle*, n^{o} 1004, Irisa, 1996, http://www.irisa.fr/centredoc/publis/PI/1996/irisapublication.2006-02-14.3475788484, Technical report.

[34] PHILIPPE. BESNARD, M.-O. CORDIER. *Inferring causal explanations*, in "Symbolic and Quantitative Approaches to Reasoning and Uncertainty (ECSQARU'99)", A. HUNTER, S. PARSONS (editors), Lecture Notes in Artificial Intelligence, vol. 1638, Springer-Verlag, 1999, p. 55-67.

[35] P. BESNARD, M.-O. CORDIER, Y. MOINARD. *Configurations for Inference between Causal Statements*, in "KSEM 2006 (First Int. Conf. on Knowledge Science, Engineering and Management)", J. LANG, F. LIN, J. WANG (editors), LNAI, n^{o} 4092, Springer, aug 2006, p. 292–304, http://www.irisa.fr/dream/dataFiles/moinard/causeksempub.pdf.

[36] P. BESNARD, M.-O. CORDIER, Y. MOINARD. *Ontology-based inference for causal explanation*, in "KSEM07 (Second International Conference on Knowledge Science, Engineering and Management)", Z. ZHANG, J. SIEKMANN (editors), LNAI, n^{o} 4798, Springer, nov 2007, p. 153-164, http://www.irisa.fr/dream/dataFiles/moinard/ksem07bcausesonto.pdf.

[37] P. BESNARD, M.-O. CORDIER, Y. MOINARD. *Deriving explanations from causal information*, in "ECAI 2008 (18th European Conference on Artificial Intelligence), Patras, Greece", M. GHALLAB, C. D. SPYTOPOULOS, N. FAKOTAKIS, N. AVOURIS (editors), IOS Press, jul 2008, p. 723–724.

[38] P. BESNARD, M.-O. CORDIER, Y. MOINARD. *Ontology-based inference for causal explanation*, in "Integrated Computer-Aided Engineering", vol. 15, n^{o} 4, 2008, p. 351-367.

[39] C. BIERNACKI, G. CELEUX, G. GOVAERT, F. LANGROGNET. *Model-Based Cluster and Discriminant Analysis with the MIXMOD Software*, in "Computational Statistics and Data Analysis", vol. 51, n^{o} 2, 2006, p. 587-600.

[40] A. BOCHMAN. *A Causal Theory of Abduction*, in "7th Int. Symposium on Logical Formalizations of Common Sense Reasoning", S. MCILRAITH, P. PEPPAS, M. THIELSCHER (editors), 2005, p. 33–38.

[41] F. CALIMERI, S. COZZA, G. IANN, N. LEONE. *An ASP System with Functions, Lists, and Sets*, in "10th International Conference on Logic Programming and Non- monotonic Reasoning (LPNMR'09)", E. ERDEM, F. LIN, T. SCHAUB (editors), LNCS, vol. 5753, Springer Verlag, 2009, p. 483-489.

[42] F. CALIMERI, G. IANNI. *Template Programs for Disjunctive Logic Programming: An Operational Semantics*, in "AI Communications", vol. 19, 2006, p. 193-206.

[43] Y. CHI, S. NIJSSEN, R. R. MUNTZ, J. N. KOK. *Frequent Subtree Mining–An Overview*, in "Fundamenta Informaticae", vol. 66, IOS Press, 2005, p. 161–198.

[44] M.-O. CORDIER, Y. PENCOLÉ, L. TRAVÉ-MASSUYÈS, T. VIDAL. *Self-healablity = diagnosability + repairability*, in "DX07 (18th International Workshop on Principles of Diagnosis), Nashville, TN, USA", May 2007, p. 251–258.

[45] M.-O. CORDIER, Y. PENCOLÉ, L. TRAVÉ-MASSUYÈS, T. VIDAL. *Characterizing and checking self-healibility*, in "ECAI 2008 (18th European Conference on Artificial Intelligence), Patras, Greece", M. GHALLAB, C. D. SPYTOPOULOS, N. FAKOTAKIS, N. AVOURIS (editors), IOS Press, 2008, p. 789–790.

[46] M.-O. CORDIER, L. TRAVÉ-MASSUYÈS, X. PUCEL. *Comparing diagnosability in continuous and discrete-events systems*, in "Safeprocess'2006, Beijing, China", 2006, http://www.irisa.fr/dream/dataFiles/cordier/safe06-cordetal.pdf.

[47] A. CORNUÉJOLS, L. MICLET. *Apprentissage artificiel : concepts et algorithmes*, Eyrolles, 2002.

[48] L. DE RAEDT. *A perspective on inductive databases*, in "SIGKDD Explor. Newsl.", vol. 4, n^{o} 2, 2002, p. 69–77, http://doi.acm.org/10.1145/772862.772871.

[49] R. DEBOUK, S. LAFORTUNE, D. TENEKETZIS. *Coordinated Decentralized Protocols for Failure Diagnosis of Discrete Event Systems*, in "Discrete Event Dynamic Systems", vol. 10, n^{o} 1-2, 2000, p. 33-86.

[50] C. DOUSSON. *Chronicle Recognition System*, 1994.

[51] C. DOUSSON. *Suivi d'évolutions et reconnaissance de chroniques*, Université Paul Sabatier de Toulouse, LAAS-CNRS, Toulouse, 1994, Ph. D. Thesis.

[52] C. DOUSSON, T. V. DUONG. *Discovering Chronicles with Numerical Time Constraints from Alarm Logs for Monitoring Dynamic Systems.*, in "IJCAI", T. DEAN (editor), Morgan Kaufmann, 1999, p. 620-626.

[53] C. DOUSSON, P. GABORIT, M. GHALLAB. *Situation recognition: representation and algorithms*, in "Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI), Chambéry, France", 1993, p. 166-172.

[54] C. DRESCHER, M. GEBSER, T. GROTE, B. KAUFMANN, A. KÖNIG, M. OSTROWSKI, T. SCHAUB. *Conflict-driven disjunctive answer set solving*, in "Eleventh International Conference on Principles of Knowledge Representation and Reasoning (KR'08)", E. ERDEM, F. LIN, T. SCHAUB (editors), AAAI Press, 2008, p. 422–432.

[55] S. DZEROSKI, L. TODOROVSKI. *Discovering dynamics: from inductive logic programming to machine discovery*, in "Journal of Intelligent Information Systems", vol. 4, 1995, p. 89-108.

[56] E. FABRE, A. BENVENISTE, C. JARD, L. RICKER, M. SMITH. *Distributed State Reconstruction for Discrete Event Systems.*, in "Proc. of the 2000 IEEE Control and Decision Conference (CDC), Sydney, Australie", 2000.

[57] B. FREY, D. DUECK. *Clustering by passing messages between data points*, in "Science", vol. 315, 2007, p. 972–976.

[58] GROUPE ALARME. *Monitoring and alarm interpretation in industrial environments*, in "AI Communications", vol. 11, 3-4, 1998, p. 139-173, S. Cauvin, Marie-Odile Cordier, Christophe Dousson, P. Laborie, F. Lévy, J. Montmain, M. Porcheron, I. Servet, L. Travé-Massuyès.

[59] J. GAO, B. DING, W. FAN, J. HAN, P. S. YU. *Classifying Data Streams with Skewed Class Distributions and Concept Drifts*, in "IEEE Internet Computing", vol. 12, n^{o} 6, 2008, p. 37-49.

[60] M. GAROFALAKIS, J. GEHRKE, R. RASTOGI. *Querying and Mining Data Streams: You Only Get One Look. Tutorial notes*, in "ACM Int. Conf. on Management of Data", 2002.

[61] E. GIUNCHIGLIA, J. LEE, V. LIFSCHITZ, N. MCCAIN, H. TURNER. *Nonmonotonic causal theories*, in "Artificial Intelligence Journal", vol. 153, n^{o} 1–2, March 2004, p. 49–104.

[62] A. GRASTIEN, M.-O. CORDIER, C. LARGOUËT. *Automata Slicing for Diagnosing Discrete-Event Systems with Partially Ordered Observations*, in "AIIA'05 (Congress of the Italian Association for Artificial Intelligence)", Septembre 2005, Milan, Italy, http://hal.inria.fr/inria-00000531.

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

[64] D. T. HAU, E. W. COIERA. *Learning qualitative models of dynamic systems*, in "Machine Learning", vol. 26, 1997, p. 177-211.

[65] T. IMIELINSKI, H. MANNILA. *A database perspective on knowledge discovery*, in "Comm. of The ACM", vol. 39, 1996, p. 58–64, http://citeseer.ist.psu.edu/imielinski96database.html.

[66] T. JÉRON, H. MARCHAND, S. PINCHINAT, M.-O. CORDIER. *Supervision patterns in discrete event systems diagnosis*, in "WODES'06 (8th International Workshop on Discrete Event Systems), Ann Arbor, Michigan, USA", IEEE (ISBN 1-4244-0053-8), 2006, http://www.irisa.fr/vertecs/Publis/Ps/PI-1784.pdf, (extended version in Irisa Technical Report 1784).

[67] M. KUBAT, J. GAMA, P. E. UTGOFF. *Incremental learning and concept drift*, in "Intell. Data Anal.", vol. 8, n^{o} 3, 2004.

[68] X. LE GUILLOU, M.-O. CORDIER, S. ROBIN, L. ROZÉ. *Chronicles for On-line Diagnosis of Distributed Systems*, n° 1890, IRISA, Rennes, France, may 2008, Technical report.

[69] X. LE GUILLOU, M.-O. CORDIER, S. ROBIN, L. ROZÉ. *Chronicles for On-line Diagnosis of Distributed Systems*, in "ECAI'08 (18th European Conference on Artificial Intelligence), Patras, Greece", M. GHALLAB, C. D. SPYTOPOULOS, N. FAKOTAKIS, N. AVOURIS (editors), IOS Press, july 2008.

[70] X. LE GUILLOU. *Chronicles for On-line Diagnosis of Distributed Systems*, in "CAISE-dc'08 (Conference on Advanced Information Systems Engineering - doctoral consortium), Montpellier, France", june 2008.

[71] S. D. LEE, L. DE RAEDT. *Constraint Based Mining of First Order Sequences in SeqLog*, in "Database Support for Data Mining Applications", LNCS, vol. 2682, Springer-Verlag, 2003.

[72] N. LEONE, G. PFEIFER, W. FABER, T. EITER, G. GOTTLOB, S. PERRI, F. SCARCELLO. *The DLV System for Knowledge Representation and Reasoning*, in "ACM Transactions on Computational Logic (TOCL)", vol. 7, n° 3, 2006, p. 499–562.

[73] H. MANNILA, H. TOIVONEN, A. I. VERKAMO. *Discovery of Frequent Episodes in Event Sequences*, in "Data Mining and Knowledge Discovery", vol. 1, n° 3, 1997, p. 259–289.

[74] A. MARASCU. *Extraction de motifs séquentiels dans les flux de données*, University of Nice-Sophia Antipolis, september 2009, Ph. D. Thesis.

[75] A. MARASCU, F. MASSEGLIA. *Mining sequential patterns from data streams: a centroid approach*, in "J. Intell. Inf. Syst.", vol. 27, n° 3, 2006, p. 291–307, http://dx.doi.org/10.1007/s10844-006-9954-6.

[76] D. PAGE. *ILP: Just Do It*, in "Proceedings of ILP'2000", J. CUSSENS, A. FRISCH (editors), LNAI, vol. 1866, Springer, 2000, p. 3-18.

[77] Y. PENCOLÉ, M.-O. CORDIER, L. ROZÉ. *Incremental decentralized diagnosis approach for the supervision of a telecommunication network*, in "DX'01 (International Workshop on Principles of Diagnosis), Sansicario, Italy", 2001, http://www.irisa.fr/dream/dataFiles/ypencole/DX01.pdf.

[78] K. B. PRATT, E. FINK. *Search for Patterns in Compressed Time Series*, in "International Journal of Image and Graphics", vol. 2, n° 1, 2002, p. 89–106.

[79] RENÉ. QUINIOU, M.-O. CORDIER, GUY. CARRAULT, F. WANG. *Application of ILP to cardiac arrhythmia characterization for chronicle recognition*, in "ILP'2001", C. ROUVEIROL, M. SEBAG (editors), LNAI, vol. 2157, Springer-Verlag, 2001, p. 220-227, http://www.irisa.fr/dream/dataFiles/quiniou/ilp01.pdf.

[80] N. RAMAUX, M. DOJAT, D. FONTAINE. *Temporal scenario recognition for intelligent patient monitoring*, in "Proc. of the 6th Conference on Artificial Intelligence in Medecine Europe (AIME'97)", 1997.

[81] R. REITER. *A theory of diagnosis from first principles*, in "Artificial Intelligence", vol. 32, n° 1, 1987, p. 57-96.

[82] L. ROZÉ, M.-O. CORDIER. *Diagnosing discrete-event systems : extending the "diagnoser approach" to deal with telecommunication networks*, in "Journal on Discrete-Event Dynamic Systems : Theory and Applications (JDEDS)", vol. 12, n° 1, 2002, p. 43-81, http://www.irisa.fr/dream/dataFiles/cordier/jeds.pdf.

[83] L. ROZÉ, M.-O. CORDIER. *Diagnosing Discrete Event Sytems : An experiment in Telecommunication Networks*, in "WODES98, Fourth Workshop on Discrete Event Systems, Cagliari, Italy", 1998, p. 130-137.

[84] LAURENCE. ROZÉ. *Supervision de réseaux de télécommunication : une approche à base de modèles*, Université de Rennes 1, 1997, Ph. D. Thesis.

[85] M. SAMPATH, R. SENGUPTA, S. LAFORTUNE, K. SINNAMOHIDEEN, D. TENEKETZIS. *Diagnosability of discrete event systems*, in "Proceedings of the International Conference on Analysis and Optimization of Systems", vol. 40, 1995, p. 1555-1575.

[86] M. SAMPATH, R. SENGUPTA, S. LAFORTUNE, K. SINNAMOHIDEEN, D. TENEKETZIS. *Active Diagnosis of Discrete-Event Systems*, in "IEEE Transactions on Automatic Control", vol. 43, n° 7, 1998, p. 908-929.

[87] M. SEBAG, C. ROUVEIROL. *Constraint Inductive Logic Programming*, in "Advances in Inductive Logic Programming", L. DE RAEDT (editor), IOS Press, 1996, p. 277-294.

[88] G. SHAFER. *A Mathematical Theory of Evidence*, Princeton University Press, 1976.

[89] R. TRÉPOS, A. SALLEB, M.-O. CORDIER, V. MASSON, C. GASCUEL-ODOUX. *A Distance Based Approach for Action Recommendation*, in "ECML 05 (European Conference on Machine Learning), Porto, Portugal", Lecture Notes in Artificial Intelligence, Springer, october 2005.

[90] R. TRÉPOS, V. MASSON, M.-O. CORDIER, C. GASCUEL-ODOUX. *Induction de motifs spatiaux décrivant les chemins de ruissellement*, in "Atelier Représentation et Raisonnement sur le Temps et l'Espace (RTE'08), Montpellier, France", Juin 2008, p. 1-13.

[91] R. TRÉPOS. *Apprentissage symbolique à partir de données issues de simulation- Gestion d'un bassin versant pour une meilleure qualité de l'eau*, Université de Rennes 1, january 2008, Ph. D. Thesis.

[92] A. VAUTIER, M.-O. CORDIER, R. QUINIOU. *An Inductive Database for Mining Temporal Patterns in Event Sequences (short version)*, in "Proceedings of IJCAI-05 (International Joint Conference on Artificial Intelligence), Edinburgh", L. P. KAELBLING, A. SAFFIOTTI (editors), 2005, p. 1640-1641, Poster.

[93] G. WIDMER, M. KUBAT. *Learning in the Presence of Concept Drift and Hidden Contexts*, in "Machine Learning", vol. 23, n° 1, 1996, p. 69-101.

[94] J. DE KLEER, A. MACKWORTH, R. REITER. *Characterizing diagnoses and systems*, in "Artificial Intelligence", vol. 56, n° 2-3, 1992, p. 197-222.

[95] J. DE KLEER, B. C. WILLIAMS. *Diagnosis with behavioral modes*, in "Proceedings of the 11th International Joint Conference on Artificial Intelligence IJCAI'89, Detroit, MI, Etats-Unis", 1989, p. 1324-1330.