INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team DREAM

Diagnosis, REcommending Actions and Modeling

Rennes - Bretagne Atlantique

THEME COG

Activity Report

2007

Table of contents

1. Team

Head of project
Marie-Odile Cordier [Professor, University of Rennes 1]

Administrative assistant
Marie-Noëlle Georgeault [AA Inria]

Inria staff members
Yves Moinard [Research Associate (CR) Inria]
René Quiniou [Research Associate (CR) Inria]
Thierry Vidal [Associate Professor on leave from ENIT Tarbes]

Faculty members
Véronique Masson [Associate professor, University of Rennes 1 (50 % time)]
Sophie Robin [Associate professor, University of Rennes 1]
Laurence Rozé [Associate professor, Insa, Rennes]

Ph. D. candidates
Ronan Trépos [Inra Ph. D. candidate, since October 2004]
Alexandre Vautier [MENRT Ph. D. candidate, since October 2004]
Xavier Le Guillou [University Ph. D. candidate, since October 2005]

Research scientists (partners)
Philippe Besnard [Research scientist, DR CNRS, IRIT, Toulouse]
Christine Largouët [Associate Professor, Ensar, Rennes]

Technical staff
Lucie Callens [Associate Engineer]
Tristan Moreau [Associate Engineer]

2. Overall Objectives

2.1. Introduction

Keywords: *diagnosis*, *machine learning*, *supervision*.

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, 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 suppose 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 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 fundings 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.
- medical applications and especially cardiac monitoring, i.e the on-line analysis of cardiac signals to detect arrhythmias and the development of “intelligent” cardiac devices (pacemakers and defibrillators) having signal analysis and diagnosis capabilities. A similar application concerns cattle healthcare, consisting in the monitoring of big dairy herds from temperature and cardiac data recorded through internal sensors.
- 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

Three points can be highlighted for 2007:

- We have been much involved in the WS-DIAMOND European project, adapting the model-based diagnosis techniques we developed before to this new application field. It gave us the opportunity to propose a novel repair-oriented view of diagnosis. Especially, we have shown the potential benefit of taking repair capabilities into account for designing a self-healable system.
- Ronan Trepos has completed his PhD thesis (to be defended in January 2008). It coincides with the end of the SACADEAU project, carried out in cooperation with our partners from SAS/INRA and dedicated to developing a decision tool to help catchment managers to preserve stream-water quality. We have started a new project, named APPEAU, funded by ANR/ADD, which aims at studying which politics, for which agronomic systems, are best adapted to improve water management. Our role is to demonstrate how machine learning techniques can help a decision maker to take a decision profit from simulation results.
- Alexandre Vautier has completed his PhD thesis (to be defended in February 2008). It coincides with the end of a CRE contract, funded by FRANCE-TELECOM R&D, concerning the use of data mining techniques in the context of intrusion detection. This research action is continuing, in a similar context, with a focus on data streams analysis, in the framework of the SESUR ARC project.

3. Scientific Foundations

3.1. Computer assisted monitoring and diagnosis of physical systems

Keywords: *chronicle acquisition*, *chronicle recognition*, *deep model*, *diagnosis*, *fault model*, *monitoring*, *simulation*, *temporal causal graph*.

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 [77], [88]. 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 [56]. Computational efficiency is a crucial issue for real size problems. We are developing two approaches. The first one relies on diagnoser techniques [82], 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 [77], [88]. 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 [89], 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 [31], [33] 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 ,...) [40], [56].

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 [82]. We have extended this method to the communicating automata formalism [80] (see also [78]). We have also developed a more generic method which takes advantage of the symmetries in the architecture of the system [79].

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 [70]. This approach can be compared with R. Debouk and colleagues [49] and also to P. Baroni and colleagues [39]. 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 [55], [35]. 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. One way to supervise dynamic systems is to recognize those chronicles on-line. The principle is to maintain and incrementally update the set of all the possible chronicles corresponding to a set of received failure messages until finding one or several instances that satisfy all the constraints. To perform this task, we have to create a chronicle base that contains all the possible chronicles. This base must be updated each time the supervised system evolves physically or structurally. An expert is needed to create the chronicle base. Moreover, this makes the maintenance of the base very expensive. That is why we prefer to use an automatic method to learn the base. Most of the studies on chronicle recognition are french [53], [76] and are based on C. Dousson's thesis [51]. Applications generally deal with system monitoring (telecommunication network) and video-surveillance (underground, bank, etc...). Our research studies do not focus directly on the development of chronicle recognition systems but on the automatic acquisition of the chronicle base. This idea is developed in the next section.

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 [81]), 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

Keywords: *Inductive Logic Programming (ILP)*, *Machine learning*, *data-streams*, *temporal abstraction*, *temporal data mining*.

Machine learning 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. They can be 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 problem 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 an essential attribute 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 [1]. 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 [73].

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 pesticids transfer, in tight collaboration with the SAS-INRA research group.

3.2.2. Learning from relational 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, induction of association rules 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 by 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, 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 hypotheses. Consequently, it is a good candidate for learning 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.

Formally, relational learning can be described as follows: given a set of positive examples P and a set of negative examples N of some concept to be learned, a logical theory B called the background knowledge and a language L_H specifying which clauses are syntactically and semantically acceptable, the goal is to discover a hypothesis H in the form of a logic program belonging to L_H such that $\forall p \in P\ B \cup H \models p$ and $\forall n \in N\ B \cup H \not\models n$. This definition can be extended to multi-class learning. From a computational point of view, the learning process consists in searching the hypothesis space, either *top-down* by refining clauses that are too general (that cover negative examples) by adding literals to clause body or *bottom-up* by generalizing clauses that are too specific (that do not cover enough positive examples) by deleting literals or transforming constants into variables in literals. An interesting property is that the clause space has a lattice structure which enables an efficient search.

Relational learning from temporal data is of major interest because many application domains, such as telecommunications, health-care or environment, provide huge amounts of such data. This is why we have chosen to rely upon work by C. Rouveirol and M. Sebag [83] who have shown the value of associating relational learning to CLP (Constraint Logic Programming) in order to infer numerical values efficiently. Other interesting extensions deal with learning dynamic models: one such extension uses a representation coming from the qualitative simulator QSIM [61], another enables the discovery of differential equations from examples describing the behavior of the dynamic system [54].

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 [68], the target application domains (such as signal processing in health-care) can benefit from the adaptation of relational learning to the particular features of the application data. Thus, 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 [75] in order to deal efficiently with the associated computations (such as the covering tests, for example). Concerning environmental data, we investigate tree structures where nodes are described by a set of attributes. Our goal is to find patterns expressed as sub-trees [44] with attribute selectors associated to nodes.

Data mining is an unsupervised learning method which aims at discovering interesting knowledge from data. Association rule extraction is one of the most popular pattern 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 [36]. 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" [37], an extension of Apriori where the order of elements in patterns are considered.

In [67], 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. Further, 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 [32] 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 [62], [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 [65]. We have investigated an extension of inductive database where patterns are very close to chronicles [87].

During the last years, a new challenge has appeared in the data mining community: mining from data streams [34]. These data come, for example, from monitoring system observing patients or from telecommunication systems. In this context, input data arrive in such a huge volume that they cannot be stored in totality for further processing: the key feature is that "you get only one look at the data" [57]. 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 or summaries have been proposed, as well as representation models taking advantage of the most recent data. 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. In this field, we are beginning a cooperative work (cf. section 7.5) on the extraction of sequential and temporal patterns from data streams with a particular focus on the adaptation of the extracted knowledge.

4. Application Domains

4.1. Telecommunication network monitoring

Keywords: *monitoring*, *security*, *supervision*, *telecommunications*.

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.

Given a monitoring system continuously receiving observations (alarms) sent by the system components, our purpose is to help operators identifying failures. In this context, we developed a decentralized component-oriented approach, able to incrementally compute on-line diagnoses [69]. We have extended the method to deal with reconfigurable systems, i.e. systems the topology of which is changing along time, due for instance to reconfiguration actions decided to remedy upload problems [59]. One of our current work is to deal with an uncertain flow of observations, and to ensure an efficient incremental computation of diagnosis [10].

An important issue in telecommunication is the security of networks. We have a cooperation with France-Telecom R&D funded by a CRE (external research contract - cf. section 7.3). We are working on two kinds of data: VPN data from which we have to extract patterns for intrusion detection ; data recorded on Internet routers from which we want to extract patterns related to DDoS attacks. Data mining techniques as well as visualization techniques are designed and experimented [86].

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. 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. Techniques mixing data-mining and symbolic learning are investigated for acquiring this kind of predictive knowledge. This is the main theme of the coordinated project SéSur (cf. section 7.5).

4.2. Software components monitoring

Keywords: *distributed diagnosis*, *software components*, *web services*.

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 are working on this problem within the WS-DIAMOND project (cf. section 7.1), a large European funded project involving eight partners in Italy, France, Austria and Netherlands http://wsdiamond.di.unito.it/. Our own work consists in two distinct contributions.

We first wish to extend the decentralized component-oriented approach, initially developed for monitoring telecommunication networks [3] 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.

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.

4.3. Decision aiding in medicine and health-care

Keywords: *data-streams*, *health-care*, *medicine*, *sensor networks*.

Since the development of expert systems in the 70's, decision aiding tools have been widely studied and used in medicine and health-care. The ultimate goal is to help a physician to establish his diagnosis or prognosis from observations delivered by sensors and the individual patient's data. This involves at least three tasks: patient monitoring, diagnostic and prognostic reasoning, modeling.

Time is a major feature of medical data, thus temporal reasoning plays an important role in data processing. This is true for data abstraction as well as for designing diagnostic and prognostic models as the course and outcome of a disease process is highly dynamic.

Another important aspect, also related to time, is adaptation. The context in which data are recorded may change with time: the noise level affecting data measurement as well as the state of the observed patient may evolve. In this case the monitoring system has to adapt his behavior to cope efficiently with this evolution. We have proposed solutions coming from the field of adaptive systems to solve this problem. Our solution is to use a pilot that activates dynamically, on the one hand, the best suited algorithm for signal processing according to signal properties, on the other hand, the models for disease recognition that have the right knowledge abstraction level according to the abstraction level of data and the patient's state.

The three points described above have been studied for cardiac patient monitoring in projects involving industrial (ELA medical), medical (University Hospital of Rennes) and academic (LTSI - University of Rennes) partners, especially in the field of cardiology. A framework prototype of Calicot achieving the above functionalities is being implemented (c.f. 5.2).

A similar approach is pursued in the agricultural domain for managing and monitoring the health of big dairy herds (cf. MOZAE project 7.4). We are collaborating with the Medria company which has designed a sensor that is placed in a cow's rumen and that enables individual as well as global monitoring. Such sensors send periodically temperature and cardiac rhythm data by radio to a monitoring station where they are analyzed. From a research point of view, our objective is, as for human cardiac monitoring, to devise techniques for extracting discriminant patterns from data that will be used to recognize interesting phenomena: estrus, disease or the proximity of birthing. From an industrial point of view the goal is to design an intelligent monitoring system that will assist the farmer, particularly that will generate alarms when a sanitary event requires some intervention.

4.4. Environmental decision making

Keywords: *environment*.

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.

Our main partners are the SAS INRA research group, located in Rennes and the BIA INRA research group 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

Keywords: *chronicle recognition*, *diagnosis*, *monitoring*, *signal processing*, *temporal abstraction*.

Participants: Lucie Callens, 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.

[1] CRS (Chronicle Recognition System) from France Telecom R & D: http://crs.elibel.tm.fr/.

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: a chronicle-based distributed diagnosis platform

Keywords: *chronicle recognition*, *distributed diagnosis*, *monitoring*.

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

CARDECRS (Chronicles Applied to error Recognition in Distributed Environments, through CRS) is a generic diagnosis platform which is currently being developed within the DREAM project.

This platform aims at monitoring complex systems, thanks to specifically designed distributed chronicles. Its 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);
- an RMI/multithreaded based environment (for the low level part).

The platform is currently under development and should be completed in June 2008. It will be experimented in the framework of the WS-Diamond European project for monitoring web services.

6. New Results

6.1. Diagnosis of large scale discrete event systems

Participants: Marie-Odile Cordier, Xavier Le Guillou, Sophie Robin, Laurence Rozé, Thierry Vidal.

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 automata. 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 [3] 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 [3]), 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 systems, 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 [60].

More recently, we defined two independence properties (transition and state independence) and we showed their relevance to get a tractable representation of diagnosis [10]. The diagnosis slices are economically represented by a set of transition-independent diagnoses and its associated set of abstract descriptions, from which the set of final states and the trajectories of the global diagnosis can be easily retrieved.

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 physical systems [53]. Since then, it has been widely used in monitoring industrial or medical systems and in particular in the CALICOT system we present in section 5.2. But chronicle-based approaches have mostly been developed in a centralized way. Our main intention is to extend the chronicle-based diagnosis engine in order to deal with distributed systems.

First, we extend the formalism proposed by C. Dousson by defining a distributed chronicle model[12]. The standard chronicle concept is enriched with synchronization events that enable the expression of synchronization constraints between the different components of the system. Thus, a distributed chronicle associated to some component can be related to distributed chronicles in other components.

Then, we have proposed a decentralized architecture for a monitoring system where each component is equipped with a local diagnoser [14]. A global diagnoser (also called *broker*) is in charge of merging the local diagnoses, using the synchronization constraints that appear in the local chronicles being recognized by local diagnosers.

Each local diagnoser relies on a chronicle recognition system engine based on the CRS developed by C. Dousson [50], and a chronicle base representing the local scenarios which are of interest. When a chronicle is recognized, the local diagnoser may send a message to the broker or not, according to its diagnosis policy. This policy depends both on the *color* (i.e. the degree of importance) of the recognized chronicle and on the *filter* of this local diagnoser. The filter is in charge of telling which color of chronicle shall trigger the broker. The broker updates its knowledge base when it receives a message from a local diagnoser. Then, if necessary, it sends a request to other local diagnosers in order to receive complementary information and refine its global diagnosis.

This work finds natural application within the context of the WS-Diamond European project (http://wsdiamond.di.unito.it/ - cf. section 7.1). The overall goal is 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, via communication links, causing its primary effects later, in another service which is not directly responsible for the failure. The distributed chronicles we have designed describe the normal and abnormal behaviors of each web service and the communication with the other services.

The next step, currently under achievement, is the development of a test bed. This platform (cf. section 5.3) has been implemented with generic distributed diagnosis design principles, and should be re-used for other projects in the future.

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

After having addressed the problem of diagnosability of complex supervision patterns by [63], 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 in late 2006 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 [11]: 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.

Our current goal is 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. Self-adaptive systems

Monitoring systems, like CALICOT, are often faced with signals of varying quality. Taking advantage of context knowledge may generally facilitate signal processing and improve the results which are next exploited by diagnosis. We have proposed a system architecture to adapt such monitoring systems to the current context [71]. This means adapting signal processing algorithms to, e.g. the noise type and level of the raw signal or the shape of interesting waves, as well as to current failure or disease hypotheses inferred by diagnosis. Furthermore, diagnosis can also be adapted to the knowledge level allowed by signal processing or current possible diagnosis hypotheses. Our adaptive system has a pilot module that operates at three stages. It chooses and tunes the signal processing algorithms that analyze the input signal by using decision rules that are learned [17] automatically and then used to choose the most relevant algorithm according to the current context. It activates or deactivates processing tasks devoted to the extraction of specific events from the input signal. It adapts the diagnosis to the current resolution of the signal analysis.

In our approach, diagnosis is achieved by chronicle recognition. Chronicles are organized in a hierarchy of chronicle bases: more abstracted chronicles use less objects (event descriptions) and fewer attributes; more specific chronicles are more precise and more detailed. The pilot chooses the current chronicle base according to the level of abstraction imposed by the recognition task or the representation of events. For example, a more abstracted chronicle base needs less low-level computation than a more refined one but the abstraction level may be to high to discriminate between diseases. Using this kind of chronicle hierarchy leads to a smarter use of computational resources.

6.1.5. Optimal activity planning and scheduling for complex systems

Apart from the main DREAM scientific topics, research topics that were previously conducted at Laboratoire de Génie de Production, ENIT, Tarbes have also been extended in 2007. In one hand, a generic framework for scheduling the activities of complex systems under uncertainty have been completed and presented both to the international AI community [8] and to the specific planning and learning French community [9]. An extended version has also been submitted to a special track of the Journal of Scheduling. On the other hand, a co-supervised PhD addressing a complex vehicle routing problem for waste collecting in a district area, with multiple optimization criteria, has produced some first results [6], [22].

That line of works is tightly connected to the approaches the DREAM team is now developing within the WS-DIAMOND project. It considers both diagnosis and planning of repair actions (see 6.1.3). It found an obvious extension within a collaborative work started in mid-2007 with the GREYC laboratory of the University of Caen, through a PhD thesis co-supervision, addressing the overall problem of configuration (i.e. service selection and activity planning), diagnosis and re-configuration/repair of distributed web-services.

6.2. Machine learning for model acquisition

Participants: Lucie Callens, Marie-Odile Cordier, Véronique Masson, Tristan Moreau, René Quiniou, Ronan Téepos, Alexandre Vautier.

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

6.2.1. Learning and mining from temporal data

6.2.1.1. Mining with poor a priori knowledge about the structure of data

Most knowledge discovery processes are biased since some part of the knowledge structure must be given before extraction. In many cases, such as in the analysis of large network logs, the user does not have sufficient domain knowledge to conduct an efficient Knowledge Discovery process. For this purpose, we propose a framework [21] that avoids this bias by supporting all major model structures, e.g. clustering, sequences, etc., as well as specifications of data and DM (Data Mining) algorithms, in the same language. The key concept of our work is the notion of schema which is related to the category theory and more precisely to the sketch theory enhanced for data mining requirements. A unification operation is provided to match automatically the data to the relevant DM algorithms in order to extract models and their related structure. In other words, the specification enables to automatically reuse specified algorithms for their adaption to the user's data. Finally, extracted models are evaluated and ranked using the MDL principle: given a data sample and an effective means to enumerate the relevant alternative theories that can explain the data, the best model is the one that minimizes length of the model description plus the length of the data description in this model.

The proposed framework is currently experimented with data randomly generated by models to validate the approaches and also with real network alarms. Precisely, in the framework of a CRE contract with France Telecom R&D (cf. section 7.3), we are working on two types of alarm logs: logs from a Virtual Private Network server that generates many alarms due to normal and abnormal connections and alarm logs generated by France-Telecom routers when they detect suspicious flows [86].

6.2.1.2. Chronicle discovery by mining data streams

We are interested in diagnosing and monitoring methods which aim at recognizing, on signal streams, temporal patterns that can be related to specific interesting events [1]. We are investigating learning from examples or mining from data streams where data arrive continuously at a very high rate. The volume of data is such that it is impossible to store them for offline processing. In fact, data can only be seen once and consequently they have to be processed on line. This is a challenge for monitoring systems using model-based techniques for diagnosis, prediction or prognosis. When observing a dynamic system, the underlying concept of the data may change and so the model must be updated. We are particularly interested in models composed of temporal patterns in the form of chronicles. In this case adaptation means to add or to retract patterns from the model or to add or to retract events from chronicles as well as to enlarge or to restrict temporal constraints.

Learning from data streams and model adaptation is achieved in the framework of the Calicot platform (cf. section 5.2 and within the SéSur project [24] (see also section 7.5). For the first action, our aim is to take into account the patient evolution. For the latter project, our aim is to take into account new attack methods which should be retrieved when some data evolution is noticed. A third application field is also investigated: zootechnic data recorded on dairy cattle (cf. section 7.4). After applying signal processing algorithms in order to improve the data quality, a methodology close to the one we have used on cardiac data for chronicle learning is being used. The results are currently evaluated and discussed by veterinary experts.

6.2.1.3. Learning decision rules for an adaptive system

The kind of adaptive systems we are investigating (cf. 6.1.4) is controlled by a pilot that uses decision rules to adapt his behavior to the current execution context. In a previous experiment [72], expert rules were devised from an analysis of algorithm performance in different signal contexts. This year we have worked on the automatic acquisition of decision rules for the pilot [17]. An example base has been built from the performance results of algorithms on signals variously affected by noise. Next, decision trees were learned from these examples and decision rules were extracted from the decision trees. A series of experiments showed that the learned decision rules performed nearly as good as expert decision rules obtained from a previous study whith a principal component analysis for extracting decision knowledge.

This work is performed in tight collaboration with Franois Portet who has a post-doc position in the University of Aberdeen.

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 [2]. In collaboration with Inra researchers, three actions have been conducted in parallel [4].

- 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. Given data on the climate over the year, on the catchment topology and on the farmer strategy, the model outputs the pesticide concentration in the stream along the year. The originality of our 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 is now implemented and used for simulation. An in-depth analysis of many simulation results leads us to refine the model. A paper on the model has been submitted to the "Computers and Geosciences" journal.

- The second action consists 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. Firstly, we decided to use Inductive Logic Programing (ILP) techniques on a simplified catchment model. The choice of ILP has been motivated by the aim to get easy-to-read and explicative rules. This first learning step (using ICL as software) showed the important impact of climate characteristics on streamwater pollution by pesticides. This work was then extended to deal with the newly implemented model and gave interesting preliminary results presented in [45].

  In order to deal with the complex spatial relations existing between the catchment plots, we decided to experiment two new approaches. The first one consists in extending Inductive Learning Programming to tree structured patterns. This was done using the Aleph software which proved to be more efficient than ICL. The second approach consists in propositionalizing the learning examples and using a propositional learning process, precisely CN2. These two approaches are described and compared in [20].

- The last action consists in *automatically analyzing propositional rules learned in the second step to help the experts in decision making*. The 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. In 2005, we proposed the algorithm DAKAR (Discovery of Actionable Knowledge And Recommendations) [85] which works as follows: starting from an unsatisfactory situation and relying on a set of classification rules, DAKAR discovers a set of action recommendations and proposes them to the domain expert to improve the situation. The actions are built by selecting attributes among those describing the situation and proposing modifications on these attribute values. A heuristic evaluation function is used to evaluate the quality of the actions and rank them. We describe this work and compare it with other action recommendation systems in a paper submitted to the *Journal of Machine Learning Research*.

  This year, visualization techniques have been developed to help the experts in dealing with a large set of rules. It is now possible to locate, on a catchment map, the examples explained by the rules and thus to compare them. It is possible, for instance, to select a zone on the catchment map, and visualize the rules explaining their pollution degree.

6.3. Diagnostic and causal reasoning

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

This work stems on [41], [42] and is related to diagnosis where observed symptoms have to be explained by faults. The previously existing proposals are ad hoc or, as in [43], [58], they are too close to standard logic in order to make a satisfactory diagnosis. Our proposal starts from propositional logic and introduces new causal formulas, built on causal atoms such as *(α causes β)* intended to mean: "α causes β". There are also "ontological atoms" such as (α IS_A β) (roughly α is of kind β). We have provided a set of inference rules [7] stating 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 $\{\alpha\}$)*. This is the simplest set of "conditions" in the formalism. It is worth noting that the rules state also that if *(α causes β)* then $\alpha \rightarrow \beta$, thus in fact 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 limits the number of connexions between these two operators, while keeping the formalism simple and natural enough. One of the rules states that if *(α causes β)* and *(β_i IS_A β)* then *(α explains β_i if_possible $\{\alpha, \beta_i\}$)*. This is the way to deal with alternative effects. The use of an elementary ontology makes the formalism easy to use, making a clear separation between three kinds of data: ordinary, causal and ontological.

In order to test the theory with toy-like examples we have translated our causal formalism into DLV [66], an Answer Set Programming implementation. Answer Set Programming [38] is a recent evolution of logic programming, where a "program" is a set of rules close to logical formulas, and the "result" is a set of some kinds of models of these formulas. The present implementations of this declarative formalism are rather efficient. This translation was relatively easy, even if some little improvements in the existing implementations could help the task of the programmer [15].

We intend to test more examples, and to adapt the theory if some unwanted behavior still remains. The translation into ASP is of great help, since the rules in ASP reflect directly the rules of the formalism, so it is easy to test any modification of the theory.

Notice that the problems encountered by workers on diagnosis have been one of the main motivations for introducing default reasoning (reasoning in the "normal course of events", thus allowing exceptions) and that an important part of the presently active work on causation is an illustration of this long lasting close relationship.

We have continued our work on one of the leading default formalism: circumscription. The idea of circumscription is to deal with the problem of default reasoning as follows. A "rule by default" (R) "if α then β" will be transformed into the rule "if α and $\neg exceptionR$ then β". Circumscription will keep the exceptions to the minimum possible. Recently, a new method of computation has been proposed [64], where the old logical notion of forgetting propositional symbols (or reducing the logical vocabulary, already studied by Boole) has been generalized to the notion of forgetting literals, and this was shown to allow tractable computation of some kinds of circumscription. We have extended this notion by allowing some propositional symbols to vary while forgetting literals, in order to deal with the really useful kinds of circumscription. Our new results [5] consist in a formal study of the connexions of this notion with already known formalisms, and in an extensive description of a constructive method for computing this notion of literal forgetting. This is important since it improves significantly previously known results ([74], [84], [64]) applying this kind of method to the computation of circumscription or other default formalisms.

7. Contracts and Grants with Industry

7.1. WS-Diamond – Web Services: DIAgnosability, MONitoring and Diagnosis

Participants: Marie-Odile Cordier, Xavier Le Guillou, Sophie Robin, Laurence Rozé, Thierry Vidal.

WS-DIAMOND is a European project (Specific Targeted Research Project or STREP) dedicated to developing a framework for self-healing web services. It started in September 2005 and will last 36 months. It will produce

- an operational framework for self-healing service execution of conversationally complex web services. Monitoring, detection and diagnosis of anomalous situations, due to functional or non-functional errors (as quality of service), is carried on. Repair/reconfiguration is performed, guaranteeing reliability and availability of web services;
- a methodology and tools for service design that guarantee effective and efficient diagnosability/repairability during execution. The results will be demonstrated on real applications.

Our team is mainly involved in the fourth and fith work packages (WP4 and WP5). WP4 is concerned with model-based diagnosis and repair of cooperative web services. The challenge of this work package is to apply recent results and techniques developed for monitoring and reconfiguring complex physical systems, as telecommunication networks, to web services networks. We are currently developing a distributed chronicle recognition approach adapted to this new application area.

WP5 is dedicated to developing a design methodology to ensure the dependability of web services. Our work in this work package concerns the definition of self-healability properties in order to ensure that the system can be both diagnosed and repaired.

A general description of the project can be found in [25]. The deliverables and the annual report [29], [30], [28], [27] describe the work done in collaboration with our partners in this project.

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

Participants: Marie-Odile Cordier, Véronique Masson, Ronan Trépos.

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.

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.

7.3. CURAR: using chronicles for network security

Participants: Marie-Odile Cordier, René Quiniou, Alexandre Vautier.

This CRE no 171978 (External Research Contract) is a focused collaboration between the DREAM research group and France Telecom R & D on the problem of detecting specific network attacks. This study began in November 2004 and is planned to last three years. The first objective is to evaluate the use of chronicles, patterns of temporally constrained events, for representing and detecting attack scenarios on telecommunication networks. The second objective is to learn or discover automatically such attack scenarios from network logs, either generated by a simulation process or really observed on active networks.

7.4. MOZAE: Zootechnic monitoring of dairy herds

Participants: Marie-Odile Cordier, Tristan Moreau, René Quiniou.

The MOZAE (Monitoring Zootechnique des Animaux d'Élevage) project started in March 2006 and will last two years. It is granted by the ministry of Agriculture and the region of Brittany. The partners are the company Medria, the regional chamber of agriculture of Brittany, INRA (National Research Institute of Agronomy), Agrocampus-Rennes, ENVN (National Veterinary Scool of Nantes) and IRISA-INRIA.

The MOZAE project aims at designing and evaluating monitoring and assisting software tools for farmers managing big dairy herds (up to 500 cows). Furthermore, the project aims at discovering veterinary knowledge and at redefining protocols for cattle health-care. The temperature and cardiac rhythm of each individual cow is continuously monitored by a sensor introduced in the cow's stomach. These signals are regularly analyzed in order to detect physiological events that interest farmers, such as estrus, disease or proximity of birthing.

DREAM is in charge of processing signals and discovering the signatures of interesting events. Machine learning and data mining techniques are being used for discovering and refining temporal patterns with the aid of veterinary experts.

7.5. SeSur: SEcurity and SURveillance on data streams

Participants: Marie-Odile Cordier, René Quiniou.

This ARC (Action de Recherche Concertée-Research Concerted Action) is funded by INRIA. It gathers researchers from the INRIA AxIS group at Sophia Antipolis (coordinator of the project), the LGI2P from EMA in Nîmes, the IDC group from LIRMM in Montpellier and DREAM. It began in january 2007 and will last two years. It is mainly focused on new knowledge extraction techniques for data streams, especially temporal knowledge and knowledge maintenance related to changes observed on the data that reflect some evolution of the observed system. The proposed methods are evaluated on network security data, especially data recorded on an htpp server.

8. Other Grants and Activities

8.1. National projects

Members of the DREAM team are involved in the following national collaboration programs:

- IMALAIA (common working group of the GdR Automatique, GDR- PRC I3 and Afia group) which brings together researchers from automatics and artificial intelligence fields on the subject of dynamic system monitoring. M.-O. Cordier is co-chair with L. Travé-Massuyès.
- GdR I3 working group GT 3.4 (machine-learning, knowledge discovery in databases, data mining) - R. Quiniou, A. Vautier.

8.2. Visits and invitations

- Yannick Pencolé, LAAS, CNRS, Toulouse, april 2007.
- Yuhong Yan, researcher at the National Research Council in Fredericton, Canada, has been invited researcher (CNRS) in our team from 10th april to 10th may 2007. She gave seminars on her work on web services and participated to our annual seminary.
- Luca Console, University of Torino, Italy. november 2007

9. Dissemination

9.1. Services to the Scientific Community

9.1.1. 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).

9.1.2. Conference program committees and organizations

- Program committee member for DX'07 and senior program committee member for IJCAI'07; area chair for ECAI'08 and RFIA'08 (M.-O. Cordier).
- Evaluator of the CASYS laboratory from CEMAGREF (M.-O. Cordier).
- Co-organizer and program committee member of the workshop "Temporal data mining" at EGC 2008 (R. Quiniou with G. Hébrail, ENST, Paris and P. Poncelet, Ecole des Mines d'Alès).
- Program committee member for IJCAI'07, ICAPS'07, JFPDA'07, RFIA'08, ICAPS'08; Invited reviewer for journals EAAI (Engineering Applications of Artifical Intelligence), JOS (Journal of Scheduling) and JEDAI (Journal Electronique d'Intelligence Artificielle) (T. Vidal).

9.1.3. Scientific boards

- ECCAI (European Coordinating Committed for Artificial Intelligence) board member: M.-O. Cordier
- Member of "Agrocampus Rennes (ENSAR)" scientific board: M.-O. Cordier

9.1.4. Thesis supervision

- PhD principal supervisor for Frédérique Baniel, ENIT-INP Toulouse (co-supervisor: Marie-José Huguet, LAAS-CNRS, Toulouse); PhD co-supervisor for Mohamad El Falou, University of Caen (principal supervisor: Abdel-Illah Mouaddib, GREYC Caen, co-supervisor: Maroua Bouzid, GREYC Caen) (T.Vidal)

9.2. 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).
- In charge of the DRT (diplôme de recherche technologique) at IFSIC (M.-O. Cordier)

10. Bibliography

Major publications by the team in recent years

[1] G. CARRAULT, M.-O. 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.

[2] 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.

[3] 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^{o} 1-2, 2005, p. 121-170.

Year Publications

Articles in refereed journals and book chapters

[4] M.-O. CORDIER, M. FALCHIER, F. GARCIA, C. GASCUEL-ODOUX, D. HEDDADJ, L. LEBOUILLE, V. MASSON, J. SALMON-MONVIOLA, F. TORTRAT, R. TRÉPOS. *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, méthodes et instrumentations de l'aide à la décision pour l'agriculture, l'agro-alimentaire et l'espace rural", to appear, Aquae, 2007.

[5] Y. MOINARD. *Forgetting literals with varying propositional symbols*, in "Journal of Logic and Computation", vol. 17, n^{o} 5, October 2007, p. 955–982, http://logcom.oxfordjournals.org/.

Publications in Conferences and Workshops

[6] F. BANIEL, M.-J. HUGUET, T. VIDAL. *Evolution et réorganisation de la collecte des déchets ménagers et stabilité des tournées*, in "LT'07 (Workshop international Logistique & Transport), Sousse, Tunisie", November 2007.

[7] 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.

[8] J. BIDOT, T. VIDAL, P. LABORIE, J. C. BECK. *A general framework for scheduling in a stochastic environment*, in "IJCAI'07 (International joint conference on Artificial Intelligence), Hyderabad, India", January 2007, p. 56–61.

[9] J. BIDOT, T. VIDAL, P. LABORIE, J. C. BECK. *Une architecture générale pour ordonnancer dans un environnement stochastique*, in "JFPDA'07 (Journées francophones de planification, décision et apprentissage pour la conduite de systèmes), Grenoble, France", July 2007, p. 13–24.

[10] M.-O. CORDIER, A. GRASTIEN. *Exploiting Independence in a Decentralised and Incremental Approach of Diagnosis*, in "IJCAI'07 (20th International Joint Conference on Artificial Intelligence)", 2007, p. 292-297, http://www.ijcai.org/papers07/Papers/IJCAI07-045.pdf.

[11] 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.

[12] M.-O. CORDIER, X. LE GUILLOU, S. ROBIN, L. ROZÉ, T. VIDAL. *Distributed Chronicles for On-line Diagnosis of Web Services*, in "DX'07 (18th International Worshop On Principles of Diagnosis), Nashville, TN, USA", may 2007, p. 37–44.

[13] X. LE GUILLOU, M.-O. CORDIER, S. ROBIN, L. ROZÉ. *Diagnosing Web Service On-line with Distributed Chronicles*, in "PhdForum'07 (The Second Beijing-Hong Kong International Doctorial Forum 2007), Hong-Kong, China", july 2007.

[14] X. LE GUILLOU, M.-O. CORDIER, S. ROBIN, L. ROZÉ, T. VIDAL. *Chroniques distribuées pour le diagnostic en-ligne de Web Services*, in "RJCIA'07 (8èmes Rencontres des Jeunes Chercheurs en Intelligence Artificielle), Grenoble, France", juillet 2007, p. 115–132.

[15] Y. MOINARD. *An Experience of Using ASP for Toy Examples*, in "Fourth Workshop on Answer Set Programming (ASP 07)", S. COSTANTINI, R. WATSON (editors), Facultade de Ciencias, Universidade do Porto, sept 2007, p. 133–147, http://www.cs.ttu.edu/asp07/.

[16] F. PORTET, F. GAO, J. HUNTER, R. QUINIOU. *Reduction of Large Training Set by Guided Progressive Sampling: Application to Neonatal Intensive Care Data*, in "Intelligent Data Analysis in bioMedicine and Pharmacology (IDAMAP-2007)", C. COMBI, A. TUCKER (editors), 2007, p. 43-44.

[17] F. PORTET, R. QUINIOU, M.-O. CORDIER, G. CARRAULT. *Learning Decision Tree for Selecting QRS Detectors for Cardiac Monitoring*, in "11th Conference on Artificial Intelligence in Medicine (AIME 07)", R. BELLAZZI, A. ABU-HANNA, J. HUNTER (editors), LNCS, vol. 4594, Springer Verlag, 2007, p. 170-174.

[18] THE WS-DIAMOND TEAM. *WS-Diamond: An Approach to Web Services - DIAgnosability MONitoring and Diagnosis*, in "eChallenges 2007, The Hague, The Netherlands", october 2007.

[19] THE WS-DIAMOND TEAM. *WS-Diamond: Web Services - DIAgnosability, MONitoring and Diagnosis*, in "DX'07 (18th International Worshop On Principles of Diagnosis), Nashville, TN, USA", may 2007, p. 243–250.

[20] R. TRÉPOS, M.-O. CORDIER, V. MASSON, C. GASCUEL-ODOUX. *Apprentissage de motifs spatiaux et agronomiques jouant un rôle dans la contamination de l'eau par les pesticides sur un bassin versant*, in "RJCIA'07 (8èmes Rencontres Jeunes Chercheurs en Intelligence Artificielle), Grenoble, France", juillet 2007, p. 231–246.

[21] A. VAUTIER, M.-O. CORDIER, R. QUINIOU. *Towards data mining without information on knowledge structure*, in "PKDD'07 (Principles and Practice of Knowledge Discovery in Databases), Warsaw, Poland", LNCS, vol. 4702, sep 2007, p. 300-311.

Miscellaneous

[22] F. BANIEL, M.-J. HUGUET, T. VIDAL. *Aide à la décision multicritères en logistique des déchets ménagers*, in " Journées STIC & Environnement, Lyon, France", November 2007.

[23] M.-O. CORDIER, T. MOREAU, R. QUINIOU. *Détection et restauration de l'effet buvée sur des séries de température ruminale mesurées par un bolus*, Rapport d'avancement N°1 - contrat Mozae, 2007.

[24] SÉSUR PROJECT. *Présentation du projet SéSur*, 2007, Poster aux journées des ARC 2007.

[25] THE WS-DIAMOND TEAM. *WS-Diamond: An Approach to Web Services - Monitoring and Diagnosis*, ERCIM News (European Research Consortium for Informatics and Mathematics), vol. 70, july 2007.

[26] A. VAUTIER, M.-O. CORDIER, R. QUINIOU. *Vers une méthode de fouille de données sans information sur la structure de la connaissance*, Rapport d'avancement N°4 - contrat CRE/FT Curar R & D n°171938, 2007.

[27] WS-DIAMOND TEAM. *Annual activity and management report*, october 2007.

[28] WS-DIAMOND TEAM. *Characterization of diagnosability and repairability for self-healing Web services– Phase I. Deliverable 5.1*, 2007.

[29] WS-DIAMOND TEAM. *Specification of diagnosis algorithms for Web Services – Phase I. Deliverable 4.3*, 2007.

[30] WS-DIAMOND TEAM. *Specification of new repair/reconfiguration algorithms for Web Services – Phase I. Deliverable 4.4*, 2007.

References in notes

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

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

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

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

[35] 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", Special issue on Hybrid Systems, vol. 8, n^{o} 2, Juin 1998, p. 203-231.

[36] 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.

[37] 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.

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

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

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

[41] P. 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.

[42] 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.

[43] 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.

[44] 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.

[45] M.-O. CORDIER, F. GARCIA, C. GASCUEL-ODOUX, V. MASSON, A. SALLEB, R. TRÉPOS. *SACADEAU project: recommending actions from simulation results*, in "BESAI06 (ECAI'06 workshop on Binding Environmental Sciences and Artificial Intelligence)", poster, August 2006.

[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.

[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, http://crs.elibel.tm.fr/.

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

[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] 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.

[55] 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.

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

[57] 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.

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

[59] A. GRASTIEN, M.-O. CORDIER, C. LARGOUËT. *Extending decentralized discrete-event modelling to diagnose reconfigurable systems*, in "DX'04 (International workshop on principles of Diagnosis), Carcassonne", 2004, p. 75–80, http://www.irisa.fr/dream/dataFiles/agrastie/dx04.ps.

[60] 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.

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

[62] 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.

[63] T. JERON, 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", (extended version in Irisa Technical Report 1784), IEEE (ISBN 1-4244-0053-8), 2006, http://www.irisa.fr/vertecs/Publis/Ps/PI-1784.pdf.

[64] J. LANG, P. LIBERATORE, P. MARQUIS. *Propositional Independence - Formula-Variable Independence and Forgetting*, in "Journal of Artificial Intelligence Research", vol. 18, 2003, p. 391–443.

[65] S. D. LEE, L. DE RAEDT. *Constraint Based Mining of First Order Sequences in SeqLog*, in "Database Support for Data Mining Applications", P. L. LANZI, R. MEO (editors), LNCS, vol. 2682, Springer-Verlag, 2003.

[66] 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.

[67] 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.

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

[69] Y. PENCOLÉ, M.-O. CORDIER, L. ROZÉ. *A decentralized model-based diagnostic tool for complex systems*, in "ICTAI'01 (IEEE international conference on tools with artificial intelligence)", IEEE computer society, 2001, p. 95-102.

[70] 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.

[71] F. PORTET, G. CARRAULT, M.-O. CORDIER, R. QUINIOU. *Pilotage d'algorithmes pour un diagnostic médical robuste en cardiologie*, in "RFIA 2006 (Congrès Reconnaissance des Formes et Intelligence Artificielle), Tours, France", janvier 2006.

[72] F. PORTET, A. I. HERNANDEZ, G. CARRAULT. *Evaluation of real-time QRS detection algorithms in variable contexts*, in "Medical & Biological Engineering & Computing", vol. 43, n^{o} 3, 2005, p. 381-387.

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

[74] T. C. PRZYMUSINSKI. *An Algorithm to Compute Circumscription*, in "Artificial Intelligence Journal", vol. 38, n^{o} 1, 1989, p. 49–73.

[75] R. QUINIOU, M.-O. CORDIER, G. 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.

[76] 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.

[77] R. REITER. *A theory of diagnosis from first principles*, in "Artificial Intelligence", vol. 32, n^{o} 1, 1987 (repris dans Readings in Nonmonotonic Reasoning, M. L. Ginsberg (dir.), Morgan Kaufmann, 1987 ; aussi dans Readings in Model-Based Diagnosis, W. Hamscher, L. Console, J. de Kleer (dir.), Morgan Kaufmann,p. 29-48, 1992), p. 57-96.

[78] 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^{o} 1, 2002, p. 43-81, http://www.irisa.fr/dream/dataFiles/cordier/jeds.pdf.

[79] 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.

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

[81] 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.

[82] 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^o 7, 1998, p. 908-929.

[83] M. SEBAG, C. ROUVEIROL. *Constraint Inductive Logic Programming*, in "Advances in Inductive Logic Programming", L. DE RAEDT (editor), IOS Press, 1996, p. 277-294, http://citeseer.ifi.unizh.ch/sebag96constraint.html.

[84] M. A. SUCHENEK. *Evaluation of Querries under Closed-World Assumption*, in "J. of Automated reasoning", vol. 18, n^o 3, 1997, p. 357–398.

[85] R. TRÉPOS, A. SALLEB, M.-O. CORDIER, V. MASSON, C. GASCUEL. *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.

[86] A. VAUTIER, M.-O. CORDIER, M. DUCASSÉ, R. QUINIOU. *Agrégation d'alarmes faiblement structurées*, in "Atelier fouille de données temporelles à EGC 06", jan 2006, http://www.irisa.fr/dream/Pages_Pros/Alexandre.Vautier/publis/egc_06_vpn.pdf.

[87] 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), Poster, 2005, p. 1640-1641.

[88] J. DE KLEER, A. MACKWORTH, R. REITER. *Characterizing diagnoses and systems*, in "Artificial Intelligence", vol. 56, n^o 2-3, 1992 (repris dans Readings in Model-Based Diagnosis, W. Hamscher, L. Console, J. de Kleer (dir.), Morgan Kaufmann,p. 54-65, 1992), p. 197-222.

[89] 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 (also in Readings in Model-Based Diagnosis, W. Hamscher, L. Console, J. de Kleer (dir.), Morgan Kaufmann,p. 124-130, 1992), p. 1324-1330.