MODELING PROCESSES FROM TIMED OBSERVATIONS

Marc Le Goc[1], Emilie Masse[1,2]
[1]*LSIS, Laboratory for Systems and Information Sciences, UMR CNRS 6168*
Aix-Marseille University, Marseilles, France

Corinne Curt[2]
[2]*Cemagref, Unité Ouvrages Hydrauliques et Hydrologie, Aix en Provence, France*

Keywords: Multi modeling, diagnosis reasoning, dynamic system.

Abstract: This paper presents a modelling approach of dynamic process for diagnosis that is compatible with the Stochastic Approach framework for discovering temporal knowledge from the timed observations contained in a database. The motivation is to define a multi-model formalism that is able to represent both the knowledge of these two sources. The aim is to model the process at the same level of abstraction that an expert uses to diagnose the process. The underlying idea is that at this level of abstraction, the model is simple enough to allow an efficient diagnosis. The proposed formalism represents the knowledge in four models: a structural model defining the components and the connection relations of the process, a behavioural model defining the states and the transitions states of the process, a functional model containing the logical relations between the values of the process's variables, which are defined in the perception model. The models are linked with the process's variables. This point facilitates the analysis of the consistency of the four models and is the basis of the corresponding knowledge modelling methodology. The formalism and the methodology is illustrated with the model of a hydraulic dam of Cublize (France).

1 INTRODUCTION

The design of knowledge based systems to supervise, diagnose and control industrial processes pose the difficult problem of the acquisition and the representation of temporal knowledge which is the core of the solving problem method of dynamic system diagnosis (Basseville and Cordier, 1996).

The aim of the MultiModel Based Diagnosis (MMBD) is to solve this problem. One of the major difficulties is the definition of the right level of abstraction at which the models have to be constructed to have an efficient diagnosis. This investigation being still an open problem, most of the proposed modeling approaches use the abstraction level of the available models, typically the design model(s). But the abstraction level required to the design of a process requires the definition of a lot of components. Using the design model(s) leads then to diagnosis models containing a large number of components. Such models entails strong computational difficulties for the usual diagnosis algorithms.

Our works consist then in developing a modeling method, called *TOM4D* (Timed Observations Modeling for Diagnosis) with the aim of using the same level of abstraction that the Experts use to diagnose a process: defined at this level of abstraction, the models lead up to a minimal set of components allowing an efficient and sure diagnosis reasoning. This notion of minimality is intuitive: only the components that are concerned with a diagnosis have to be identified in the models. Consequently, when using such models, the computational difficulties of the usual diagnosis algorithms will decrease. The *TOM4D* methodology is based on the idea that the Experts use implicit models to formulate their knowledge about the process and the way to diagnose about it. The methodology aims then to explicit these models when combining a CommonKads-like conceptual approach (Schreiber et al., 2000) with a multimodeling approach (Chittaro et al., 1993) and the *Stochastic Approach* framework for discovering temporal knowledge from timed observations (Le Goc et al., 2005), (Le Goc, 2006). The modeling methodology is directed by the timed observations provided either with a set of scenarios describing a concrete deterioration of a process or with the *a priori* knowledge about the main events that plays a role in the different diagnosis. The next section of

Le Goc M., Masse E. and Curt C. (2008).
MODELING PROCESSES FROM TIMED OBSERVATIONS.
In *Proceedings of the Third International Conference on Software and Data Technologies - PL/DPS/KE*, pages 249-256
DOI: 10.5220/0001884502490256

the paper (Section 2) provides a brief overview of the main modeling approaches. Section 3 presents the formalisms of the four models of the *TOM4D* modeling methodology, and section 4 describes the corresponding modeling process through a real world application to the Cublize dam (France). Finally, section 5 states our conclusions and proposes some perspectives to our works.

2 MODELING APPROACHES FOR DIAGNOSIS

The limitations and the problems of the heuristic approach (Clancey, 1987) have motivated the definition of the Model Based Diagnosis (MBD) approach (Reiter, 1987). Reiter's algorithm of diagnosis uses a logical model of the system. This model represents both the structure of the system (components and interconnections) and the correct behaviors of the components through a set of relations between input and output values of the components. The MBD, either in a component (De Kleer and Brown, 1984), in a process (Forbus et al., 1984) or in a constraint (Lee and Kuipers, 1988) based approach, applies the well known "no function in structure" principle. This provides a clear and general framework for diagnosis.

Nevertheless, these approaches present two major drawbacks. First, the number of potential diagnosis is exponential with the number of components. This leads to several difficulties with real world systems where the number of components is large. Second, even when the system contains few components, the consistency based theory of diagnosis provides no means to eliminate diagnosis which are logically acceptable but physically meaningless (Struss and Dressler, 1989). A lot of extensions have then been proposed to avoid these difficulties (Dague, 2001): using different levels of aggregation (Davis, 1984) corresponding to different approximations, using ontologies that represent the variable values at multiple resolutions level (Hayes, 1989), using different behavioral models either qualitative or quantitative models (Murthy, 1988), and using multiple types of knowledge (Abu-Hanna et al., 1991). Nevertheless, there is not a general solution to these problems. Moreover, none of these approaches really explains how to export partial results obtained with a model into other models during the problem solving task.

This has led researchers to give up the "no function in structure" principle with the aim of using some teleological and functional knowledge. This provides information for driving the diagnosis reasoning about the structure and the behavior of the system. For example, the compositional modeling of (Falkeihainer and Forbus, 1988), (Falkeihainer and Forbus, 1991), uses the ideas of the Qualitative Physics theory. An explicit abstract representation of the process is then given where a model is made of fragments from a general-purpose domain theory. Similarly, the Function Behavior Structure approach (FBS) of (Franke, 1991) and the multimodeling approach of (Chittaro et al., 1993) consider the reasoning task as a cooperative activity between diverse models. Each model represents a specific type of knowledge and uses a specific representation formalism. Both approaches separate on one hand the structural and the behavioral knowledge about the physical system and on the other hand the knowledge about the functions and the purposes of the system. But only the multimodeling approach proposes a physical basis for the relations between the different models.

The multimodeling approach of (Chittaro et al., 1993) separates the available knowledge about a dynamic system in three main categories: the fundamental, the interpretative and the empirical knowledge. The fundamental knowledge is the basic knowledge (structural and behavioral models) used to reason about a system using the objective and neutral language of natural sciences (Chittaro et al., 1993). The interpretative knowledge is provided by a subjective interpretation of the fundamental knowledge. This interpretation is made in terms of functions (functional model) and goals (teleological model) of the system components assigned by the designer(s). The empirical knowledge is an explicit statement of the properties of the system and may refer to the two other categories. These three sets of knowledge are based on an ontology which defines the type of entities the system is made with. The functional model links the behavioral model and the teleological model. The role of a function is to describe how the behavior of the individual components contributes to the achievement of a goal assigned to the system. The links between the models allow a good continuity for modeling and reasoning.

The behavioral model is the key point of the multimodeling approach, notably because the diagnosis reasoning process is based on this model. Behavioral models are build according to physical equations of the system studied. This set of equations is difficult to determine because there is an important number of physical laws to respect. So the bond-graph concept has been used to facilitate the determination of these equations and the modeling of the behavior of components and the generation of functional roles (Zouaoui et al., 1997). These functional roles are identified by the interpretation of the behavior of a system or more precisely of its physical equations. The bond graph approach (Rosenberg and Karnopp, 1983) expresses system dynamics in terms of energy transfer between constituent elements and more precisely based on power echanges. This approach is based on the tetrahedron of state which represents the relations between effort and power variables. It allows the construction of dynamic system models. Nevertheless, the behavioral model and the functional model are closed to the set of components (Thetiot, 1999) because in most cases the “a priori” knowledge is mainly provided by the design model of the system. Thereby, the diagnosis process is still concerned with computational problems. There is then a crucial need to define the right level of abstraction for modeling with a level of aggregation that allows an efficient diagnosis process.

So, in this paper, we propose to use the same level of abstraction the Experts use to diagnose a dynamic system. This level of abstraction corresponds to a level of aggregation that minimizes the set of components and thereby improves the computational problem. This corresponds to associate the abstraction level with the diagnostic task and not with the design task.

3 TOM4D MODELING METHOD

TOM4D (”Timed Observations Multimodeling for Diagnosis”) is a timed observations centered modeling method for dynamic systems or processes. The aim of this method is to produce a model from a set of sequences of timed observations and the *a priori* expert’s knowledge, the so produced model being used to diagnose a dynamic process. The main constraints to define the *TOM4D* method is that the modeling formalisms must be compatible with (i) the Stochastic Approach Framework for discovering temporal knowledge from timed data of (Le Goc, 2006), (Le Goc and Bénayadi, 2008), (ii) the conceptual multimodeling framework of (Zanni et al., 2005) and (iii) the diagnosis algorithm of (Reiter, 1987).

3.1 The Stochastic Approach Framework

The *Stochastic Approach Framework* of (Le Goc, 2006), (Le Goc and Bénayadi, 2008) for discovering temporal knowledge from timed data provides a general framework for modeling dynamic processes. This framework considers that the timed messages of a series are written in a database by a program, called a monitoring cognitive agent *MCA*, that monitors, diagnoses or controls a production process *Pr*. A timed message is represented with an occurrence $o(t_k) \equiv (\delta_i, k)$ of a discrete event class $C^i = \{(x_i, \delta_i)\}$ that is an arbitrary set of discrete event $e_i \equiv (x_i, \delta_i)$, where δ_i is one of the discrete value of the variable x_i. A discrete event class is often a singleton because in that case, two discrete event classes $C^i = \{(x_i, \delta_i)\}$ and $C^j = \{(x_j, \delta_j)\}$ are only linked with the variables x_i and x_j when the constants δ_i and δ_j are independent (Le Goc, 2006). This condition is only concerned with the programs the *MCA* is made with. This means that a relation between the occurrences of two classes C^i and C^j subsumes a relation between the corresponding two functions $x_i(t)$ and $x_j(t)$. A sequence of discrete event class occurrences is then considered as the observable manifestation of a series of state transitions in a timed stochastic automaton representing the couple (Pr, MCA). The *BJT4G* algorithm represents a set of sequences of discrete event class occurrences with a Markov chain and apply an abductive reasoning on this representation to identify the set $M = \{R_{ij}(C^i, C^j, [\tau^-_{ij}, \tau^+_{ij}])\}$ of the most probable timed sequential binary relations $R_{ij}(C^i, C^j, [\tau^-_{ij}, \tau^+_{ij}])$ between discrete event classes leading to a given class. A timed sequential binary relation $R_{ij}(C^i, C^j, [\tau^-_{ij}, \tau^+_{ij}])$ is an oriented relation between two discrete event classes C^i and C^j that is timed constrained with the interval $[\tau^-_{ij}, \tau^+_{ij}]$. $[\tau^-_{i,j}, \tau^+_{i,j}]$ is the time interval for observing an occurrence of the C^j class after an occurrence of the C^i class. The set *M* of timed sequential binary relation is an abstract chronicle model that is graphically represented with the *ELP* language (Event Language for Process) where the nodes are discrete event classes and the links are timed sequential binary relations.

3.2 TOM4D Model Formalisms

The conceptual multimodeling framework of (Zanni et al., 2005) organizes the available knowledge about a process according to three models: a *Structural Model* describing the relations between the components of the process, a *Functional Model* providing the relations between the values of the process variables (i.e. a set of mathematical functions) and a *Behavioral Model* defining the states of the process and the discrete events firing the state transitions.

The *TOM4D* method extends this framework to add a complementary model, called the *Perception Model* (*PM*) of the process. This model specifies the process $P(t) = R(x_1(t), \ldots, x_i(t), \ldots, x_n(t))$ as a relation between n variables x_i of a set $X = \{x_i\}$ and a set $\{Q_i(x_i, \delta_i)\}$ of constraints about the values of the process variables. The Structural, Functional and Behavioral models must be deduced from Given PM. Consequently, with the *TOM4D* method, a model $M(P(t))$ of a process is a quadruplet $M(P(t)) = \langle PM, SM(P(t)), FM(P(t)), BM(P(t))\rangle$.

The *Perception Model* $PM = \langle \Psi, X, R^q\rangle$ defines the process $P(t)$ and its operating modes:

- $\Psi = \{\psi_i\}$ is a set of constants ψ_i typically corresponding to thresholds;
- $X(t) = \{x_i(t)\} \subseteq V(t)$ is a sub set of all the measured variables $V(t)$ of the process;
- $R^q = \{\prod(Q_i(x_i(t), \psi_i)\}$ is a set of binary predicate conjunctions $Q(x_i, \psi_i)$ linking a variable $x_i \in X$ and a constant $\psi_i \in \Psi$. These predicates formulate some constraints about the values of the process variables x_i. R^p is partitioned in three parts $C^g \cup C^n \cup C^a$ (i.e. $C^g \cap C^n \cap C^a = \{\phi\}$):
  - C^g contains the sub set of R^p describing the process operating goals;
  - C^n contains the sub set of R^p describing the normal operating modes of the process;
  - C^a contains the sub set of R^p describing the abnormal operating modes of the process;

The *Structural Model* $SM(P(t)) = \langle COMPS, R^i, R^x\rangle$ of a process $P(t)$ describes relations between components of the system and the relations between the components and the variable $X(t)$ of $P(t)$:

- $COMPS = \{c_i\}$ is a set of components $c_i \equiv \{\{e_i(c_i)\}, \{s_i(c_i)\}\}$ where $\{e_i(c_i)\}$ (resp. $\{s_i(c_i)\}$) is the set of input (reps. output) ports of the component c_i;
- $R^i = \{= (s_i(c_i), e_j(c_j))\}$ is a set of assignment of the binary predicate *'Equal'* linking an output of a component c_i with an input of the component c_j (i.e. the interconnection relations);
- $R^x = \{= (x_i(t), s_i(c_i))\}$ is a set of assignment of the binary predicate *'Equal'* linking each variable $x_i(t) \in X(t)$ with an output s_i of a component $c_i \in COMPS$.

The *Functional Model* $FM(P(t)) = \langle \Delta, F, R^f\rangle$ of a process $P(t)$ describes the relations between the values δ_i of the variables $x_i \in X$ with mathematical functions:

- $\Delta = \{\bigcup \Delta^{x_i}\} \bigcup \{\phi\}$ is the union of the sets $\Delta^{x_i} = \{\delta_i\}$ of constants associated with each variable x_i of X. The sets $\Delta^{x_i} = \{\delta_i\}$ is deduced from the set $\Psi = \{\psi_i\}$ of the *Perception Model PM* when applying the "*Spatial Discretization Principle*" of the Stochastic Approach Framework (Le Goc et al., 2005). The constant ϕ is added to assign an unknown value to a variable.
- $F = \{f_i(x_1(t), x_2(t), \ldots, x_m(t))\}$ is a set of function $f_i(x_1(t), x_2(t), \ldots, x_m(t))$ linking each value of the variables $x_1(t), x_2(t), \ldots, x_m(t)$ with the output value of the function $f_i(x_1(t), x_2(t), \ldots, x_m(t))$. Typically, these functions are defined with tables of values.
- $R^f = \{= (x_i(t), f_i(x_1(t), x_2(t), \ldots, x_m(t)))\}$ is a set of assignments of the binary predicate "Equal" linking the value of a variable x_i with the output values of a function $f_i(x_1(t), x_2(t), \ldots, x_m(t)))$.

The Functional Model $FM(P(t))$ specifies the relations between the variables x_i of the process $P(t)$ but any mention about the normal and the abnormal values is contained in $FM(P(t))$. These properties are deduced from the Behavioral Model of $P(t)$.

The *Behavioral Model* $BM(P(t)) = \langle S, C, R^c\rangle$ of a process $P(t)$ describes its operating modes (according to (Chittaro et al., 1993)) with a set of states and discrete event classes firing the state transitions:

- $S = \{s_i\}$ is a set of distinguishable states. Each state s_i is identified with a unique value $X_i \equiv \{(x_i = \delta_i)\}$ of $X(t)$ so that:
  - $X(t) = X_i \Leftrightarrow \forall x_i(t) \in X(t), \exists \delta_i \in \Delta, x_i(t) = \delta_i$.
  - $\forall s_i, s_j \in S, s_i = s_j \Leftrightarrow X_i = X_j$.
- $C = \{C^i\} \cup C^\phi$ is a set of classes $C^i = \{e_i\}$ of discrete event $e_i = (x_i, \delta_i)$. $C^\phi = \{e_\phi\}$ is a technical class containing the event $e_\phi = (\phi, \phi)$ which matches with the observation of a date. In this modeling context, each class $C^i = \{e_i\}$ is a singleton (i.e. $C^i = \{e_i\} = \{(x_i, \delta_i)\}$).
- $R^c = \{= (s_i, \gamma_i(C^i, s_j)\}$ is a set of assignment of the binary predicate "Equal" linking a state s_i to a state s_j conditioned with the observation of an occurrence of the class C^i.

Given a sequence $\omega = \{o(k)\}$ of discrete event class occurrences $o(k) \equiv (\delta_i, t_k)$, a transition from the state s_i to the state s_j is fired when:

- there is an occurrence $o(k) \equiv (\delta_i, t_k)$ of the class $C^i = \{e_i\}$, $e_i = (x_i, \delta_i)$, in ω;
- the current state $s(t)$ of the finite state machine implementing $BM(P(t))$ is in the state s_i (i.e. $s(t) = s_i$);
- there exist an assignment $= (s_i, \gamma_i(C^i, s_j))$ in R^c.

In that case, the new value of the current state $s(t)$ is provided with the σ function defined on $S \times \omega \times R^c \to S$ of a finite state machine such as $s(t_k) = \sigma(s(t), o(k), R^c)$:

- $\forall t_k \geq t, \forall o(k) \in \omega, \forall s_i \in S, \forall = (s_i, \gamma_i(C^i, s_j)) \in R^c$
  $s(t) = s_i \wedge o(k) \equiv (\delta_i, t_k) \wedge C^i = \{(x_i, \delta_i)\}$,
  $\Rightarrow \sigma(s(t), o(k), R^c) = \gamma_i(C^i, s_j)$

According to the set R^q of constraints in the Perception Model PM of the process $P(t)$, the set S of states $BM(P(t))$ can be partitioned in three categories $S^g \cup S^n \cup S^a$ (i.e. $S^g \cap S^n \cap S^a = \phi$):

- A set S^g containing the states satisfying the process operating goals of $P(t)$:
  $\forall s_i \in S^g, \exists \prod(Q(x_i, \delta_i)) \in C^g$,
  $s(t) = s_i \Rightarrow X(t) = X_i$
  $X(t) = X_i \Rightarrow \prod(Q(x_i, \delta_i))$
- A set $S^n \subseteq S$ containing the states corresponding to the normal operating modes of $P(t)$:
  $\forall s_i \in S^n, \exists \prod(Q(x_i, \delta_i)) \in C^n$,
  $s(t) = s_i \Rightarrow X(t) = X_i$
  $X(t) = X_i \Rightarrow \prod(Q(x_i, \delta_i))$
- A set $S^a \subseteq S$ containing the states corresponding to the abnormal operating modes of $P(t)$:
  $\forall s_i \in S^a, \exists \prod(Q(x_i, \delta_i)) \in C^a$,
  $s(t) = s_i \Rightarrow X(t) = X_i$
  $X(t) = X_i \Rightarrow \prod(Q(x_i, \delta_i))$

Consequently, the S^a set of states defines the abnormal values $X(t) = X_i$ of the variables x_i of $P(t)$. So the qualification ”normal value” or ”abnormal value” depends on the states:

- $\forall t \in \Re$, $\forall s_i \in S^a$,
  $s(t) = s_i \Rightarrow X(t) = X_i \wedge AB(s(t))$
- $\forall t \in \Re$, $\forall s_i \in S^g \cup S^n$,
  $s(t) = s_i \Rightarrow X(t) = X_i \wedge \neg AB(s(t))$

The *Perception Model* $PM = \langle \Psi, X, R^q \rangle$, the *Structural Model* $SM(P(t)) = \langle COMPS, R^x, R^s \rangle$, the *Functional Model* $FM(P(t)) = \langle \Delta, F, R^f \rangle$ and the *Behavioral Model* $BM(P(t)) = \langle S, C, R^c \rangle$ degining a process $P(t)$ are linked together with the variable set $X = \{x_i\}$:

- each component $c_i \in COMPS$ is linked with at least one variable $x_i \in X$ through a relation $= (x_i(t), s_i(c_i))$ in R^x;
- each function $f_i(x_1(t), x_2(t), ..., x_m(t)) \in F$ is linked with at least one variable $x_i \in X$ through a relation $= (x_i(t), f_i(x_1(t), x_2(t), ..., x_m(t)))$ in R^f;
- each state transition $= (s_i, \gamma_i(C^i, s_j) \in R^c$ is linked with at least one variable $x_i \in X$ through a discrete event class $C^i = \{(x_i, \delta_i)\}$ in C.

4 TOM4D MODELING PROCESS

The *TOM4D* method is made with three steps (cf. Figure 1) aiming at producing a model $M(P(t)) = \langle PM, SM(P(t)), FM(P(t)), BM(P(t)) \rangle$ from the available knowledge and data. The *Knowledge Interpretation* step uses a CommonKADS-like template

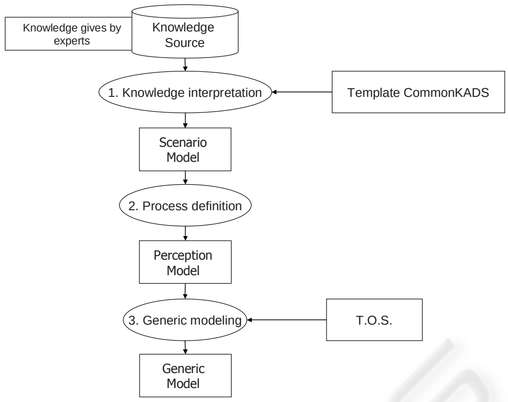


Figure 1: TOM4D Modeling Process.

(Schreiber et al., 2000) to interpret and organize the available knowledge about a process (i.e. the knowledge conceptual model of Sachem (Le Goc, 2004) in the example of this paper). This knowledge is provided by a knowledge source (an expert, a set of documents, etc) and at least one scenario $\omega = \{x_j(t_0) = \delta_j, \ldots, x_i(t_k) = \delta_i, \ldots\}$ describing a typical evolution of the process with a series of timed measures $x_i(t_k) = \delta_i$. This first step produces then the scenario model $M(\omega) = \langle SM(\omega), FM(\omega), BM(\omega) \rangle$ of the process that is coherent with the scenario $\omega = \{o_i(t_k) \equiv (\delta_i, k)\}$ when formulated with a series of occurrences $o_i(t_k)$ of discrete event classes $C^i = \{(x_i, \delta_i)\}$. The *Process Definition* step aims at defining the Perception Model (PM) of the process from the available knowledge and the scenario model $M(\omega)$. Finally, the *Generic Modeling* step uses the tetrahedron of states (TOS, (Rosenberg and Karnopp, 1983)) to provide a ”physical” dimension to each variable x_i of the process $P(t)$ and an interpretation of the relations linking the variables between them. The analysis of the properties of the perception model PM with this interpretation produces some knowledge about the process that is distributed in the different parts of the model $M(P(t))$. This model is called the ”*Generic Model*” of the process because it is independent of the concrete instrumentation. This section explains the steps 2 and 3 of the TOM4D modeling process trough its application with a real world process, the Cublize dam (France) that has been diagnosed wy the Expert's of the Cemagref, the French governmental organization that assumes the security of French hydraulic civil engineering structures. The step 1, the *Knowledge Interpretation* step), produces the scenario model $M(\omega)$ that is presented in (Masse and Le Goc, 2007).

This scenario model leads to define the process $P(t)$ as a relation $P(t) = R(V(t), Qs(t), Qf(t))$ between three variables: a volume $V(t)$ and two outflows, a ”normal” outflow $Qs(t)$ and a water leak outflow $Qf(t)$. The following constraints have been deduced from $M(\omega)$:

- The goal of the process operations is to avoid any water leak outflow:
  - $\forall t, Qf(t) = Qfmin$
- Two normal process operations have been identified, one before the first dam's filling and one after:
  - $\exists t_i, \forall t < t_i, V(t) = Vmin \wedge Qs(t) = Qsmin \wedge Qf(t) = Qfmin$
  - $\exists t_i, \forall t \geq t_i, Vmax > V(t) > Vmin \wedge Qsmax > Qs(t) > Qsmin \wedge Qf(t) = Qfmin$
- Three abnormal process operations have been identified:
  - $\exists t_j > t_i, \forall t \geq t_j, Qs(t) \leq Qsmin$
  - $\exists t_j > t_i, \forall t \geq t_j, Qs(t) \geq Qsmax$
  - $\exists t > t_i, Qf(t) > Qfmin$

Consequently, the set X of the process variable is then:

- $X(t) = \{V(t), Qs(t), Qf(t)\}$;

the set of threshold constants is:

- $\Psi = \{Vmax, Vmin, Qfmax, Qfmin, Qsmax, Qsmin\}$;

and the set of constraints is

- $R^q = \{ \ (Qf(t) = Qfmin), \ (Vmax > V(t) > Vmin) \wedge (Qsmax > Qs(t) > Qsmin) \wedge (Qf(t) = Qfmin), \ (Qs(t) \leq Qsmin), \ (Qs(t) \geq Qsmax), \ (Qf(t) > Qfmin)\}$.

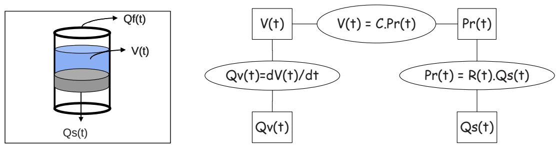
This mean that the *Perception Model* $PM = \langle \Psi, X, R^q \rangle$ is only concerned with a filled dam.

The hydraulic TOS of Figure 2 (b) provides a physical dimension to the variables of $X(t)$ leading to define the structural model $SM(P(t))$ of $P(t)$ as a pipe (a glass, Figure 2a) with a constant capacity C that contains a variable volume $V(t)$ of water and that is closed with a permeable stopper (a porous cork). The resistivity to water of the stopper evolve as a not measured variable denoted $R(t)$. The relation between the volume $V(t)$ and the outflow $Qs(t)$ is made through the pressure $Pr(t)$ on the top of the stopper, which is not measured. So the complete set of variable is:

- $\{V(t), Qs(t), Qf(t), Pr(t), R(t)\}$,

but the generic process is instrumented with only three abstract sensors measuring the outflow $Qs(t)$ through the stopper, the outflow over the pipe $Qf(t)$ and the volume $V(t)$ of the column of water in the pipe.

The Structural Model $SM(P(t)) = \langle COMPS, R^i, R^x \rangle$ of $P(t)$ is then made with three components:



(a) (b)

Figure 2: (a) Structural Model and (b) TOS Relations.

- $COMPS = \{ \ pipe, stopper, waterColumn\}$, $pipe \equiv \{e_1(pipe), s_1(pipe), s_2(pipe)\}$, $stopper \equiv \{e_1(stopper), s_1(stopper), s_2(stopper)\}$, $waterColumn \equiv \{s_1(waterColumn)\}\}$;
- $R^i = \{ \ s_1(waterColumn) = e_1(pipe), \ s_2(pipe) = e_1(stopper)\}$ (i.e. there is two interconnection relations);
- $R^x = \{ \ V(t) = s_1(waterColumn), \ Qf(t) = s_1(pipe), \ Pr(t) = s_2(pipe), \ Qs(t) = s_1(stopper), \ R(t) = s_2(stopper)\}$.

The set Ψ of PM allows to define a set of 8 ranges for each variables of $X(t)$ and consequently (i) the set Δ of 8 constants δ_i of the Functional Model $FM(P(t))$ (column "range values" in the Figure 3) and (ii) the corresponding set C of 8 discrete event classes $C^i = \{e_i\}$ (column "Events" in the Figure 3). These constants define also the set X_i of the 17 possible values for $X(t)$. The Structural Model $SM(P(t))$ and the TOS allows to eliminates 7 of these values that are note physically possible for $X(t)$: for example, a value X_i so that $V(t) = 0$ and $Qf = 1$ is physically impossible. The resulting set X_i of the 10 possible values of $X(t)$ defines then the 10 distinguishable states $S = \{s_i\}$ of the *Behavioral Model* $BM(P(t))$.

Variables	Constants	Range	Range Values	Events
V(t)	Vmax, Vmin	V(t)<Vmin	0	e1:V(t)<Vmin
		Vmin<V(t)<Vmax	1	e2:V(t)>Vmin
				e3:V(t)<Vmax
		V(t)>Vmax	2	e4:V(t)>Vmax
Qs(t)	Qsmax, Qsmin	Qs(t)<Qsmin	0	e5:Qs(t)<Qsmin
		Qsmin<Qs(t)<Qsmax	1	e6:Qs(t)>Qsmin
				e7:Qs(t)<Qsmax
		Qs(t)>Qsmax	2	e8:Qs(t)>Qsmax
Qf(t)	Qfmin	Qf(t)<Qfmin	0	e9:Qf(t)<Qfmin
		Qf(t)>Qfmin	1	e10:Qf(t)>Qfmin

Figure 3: Constraint table.

Given the sets $\{X_i\}$ of 10 values and C, three matrix are filled in to identify the physically possible temporal successions:

- an event-to-event matrix $E = [e_{ij}]$ defined on $E \times E$ at value on $\{0, 1\}$ so that $e_{ij} = 1$ when the an occurrence $o(k)$ of the C^i class can be followed by an occurrence $o(k+1)$ of the C^j class (i.e. $(o(k), o(k+1))$ is physically observable on $P(t)$).
- a value-to-value matrix $X = [x_{ij}]$ defined on $X \times X$ at value on $\{0, 1\}$ so that $x_{ij} = 1$ when a $X(t)$

can physically move from the value X_i at time t_k (i.e. $X(t_k) = X_i$) to the value X_j at the next time t_{k+1} in one occurrence of a discrete event class (i.e. $X(t_{k+1} = X_j)$.

- a value-to-event matrix $T = [t_{ij}]$ defined on $X \times X$ at value on C so that $t_{ij} = \{C^i\}$ is a set of discrete event class that is not empty when there exist at leat an occurrence of a class C^i making the value of X(t) moving from X_i to X_j in $P(t)$.

This latter transition matrix T is deduced from the other two matrix E and X.

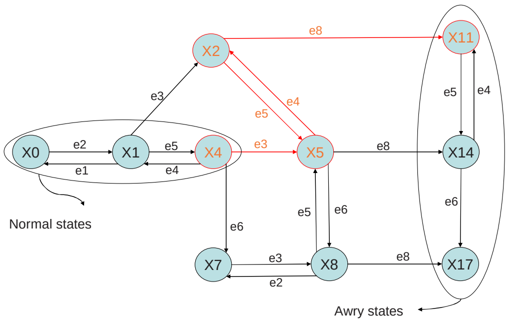


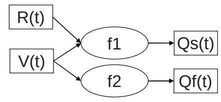
Figure 4: Behavioral model and Cublize dam story.

The state graph of the Figure 4 is a graphical representation of the T transition matrix and represents the *Behavioral Model* $BM(P(t))$ of $P(t)$. In this graph, a node is a state s_i and a relation $R(s_i, s_j, C^i)$ between two states is labeled with a discrete event class C^i when $C^i \in t_{ij}$. This state graph is then transformed in an discrete event graph where a node is a class C^i and a link is binary relation between two classes C^i and C^j if and only if $e_{ij} = 1$. Next, the classes C^i of the nodes are replaced with the associated assignation of the variables $x_i(t) = \delta_i$ according to the Stochastic Approach framework. The resulting graph is made a set of relations between two assignations $R_{ij}(x_i(t) = \delta_i, x_j(t) = \delta_j)$. Using the sets R^x and the R^i of the Structural Model $SM(P(t))$, the set $R = \{R_{ij}\}$ is made with the relations $R_{ij}(x_i(t) = \delta_i, x_j(t) = \delta_j)$ that are physically supported with an interconnection relation $= (s_i(c_i), e_j(c_j))$ of R^i. The set R is then transformed in the set F of functions $f_i(x_1(t), x_2(t), \ldots, x_m(t))$ of the *Functional Model* $FM(P(t))$ and their corresponding table of values (cf. Figure 5), and the set R^f is then constituted (i.e. $R^f = \{\, Qs(t) = f_1(V(t), R(t)),$ $Qf(t) = f_2(V(t))\}$. The $R(t)$ variable is an internal variable of the component "stopper" that explains the different values $Qs(t)$ can take when $V(t) = 1$. This is the role of the diagnosis to determine that $P(t)$ is in an abnormal state like X_{11} for example because of the value of $R(t)$.

The red paths of the *Behavioral Model* $BM(P(t))$ of Figure 4 illustrates the concrete story of Cublize's

V(t)	Qs(t)	V(t)	Qf(t)
0	0	0	0
1	0,1,2	1	0
2	0,1	2	1

(a)



(b)

Figure 5: Table of values and the functional model associated.

dam, that is to say the *Behavioral Model* $BM(\omega)$ of the scenario model $M(\omega)$ provided in (Masse and Le Goc, 2007). It can be see that $BM(\omega) \subset BM(P(t))$: $BM(P(t))$ can be used to generate a large set of scenarios that can be used to simulate the behavior and to learn to diagnose a dam. Similarly, the Figure 6) shows the relation between the Functional Model $FM(\omega)$ of the scenario model and the Functional Model $FM(P(t))$ of the process $P(t)$. The Generic Model $M(P(t))$ is then an abstract description of the relations between the concrete variables of Cublize dam. The experts of the Cemagref having validated this generic model, this shows that the TOM4D method can be used to represent a real world set of knowledge.

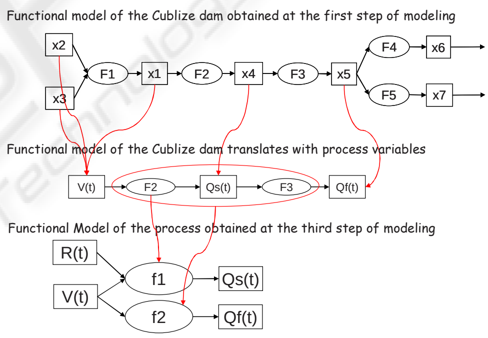


Figure 6: Comparison between the functional models.

5 CONCLUSIONS

This paper presents the principles of the *TOM4D* methodology used to represent a process at the level of abstraction that an Expert uses when diagnosing. At such a level of abstraction, the number of components is minimized so that the computational problem of the usual diagnosis algorithm is decreased. The *TOM4D* methodology represents the implicit model of an Expert with four models: a perception, a structural, a behavioral and a functional model. These models are linked together with the concept of variable, which allows to define a way to analyze the consistency between them. The methodology is directed with the timed observations to be adequate with the Stochastic Approach framework for discovering temporal knowledge from the timed observations contained in a database. One of the main advantage of the models of the *TOM4D* methodology is to be humanly understandable.

The methodology is applied to a real world problem: the hydraulic dam of Cublize (France). The resulting models have then been validated by the hydraulic dam Expert's of the Cemagref, the French governmental organization that assumes the security of French hydraulic civil engineering structures. Our current works are oriented towards the adaptation of Reiter's algorithm of diagnosis to timed observations and the generalization of this modeling approach to introduce a recursive principle of modeling.

REFERENCES

Abu-Hanna, A., Benjamins, R., and Jansweijer, W. (1991). Device understanding and modeling for diagnosis. *IEEE Expert*, 6(2):26–31.

Basseville, M. and Cordier, M.-O. (1996). *Surveillance et diagnostic de systèmes dynamiques : approche complémentaire du traitement de signal et de l'intelligence artificielle*. Rapport.

Chittaro, L., Guida, G., Tasso, C., and Toppano, E. (1993). Functional and teological knowledge in the multimodeling approach for reasoning about physical systems: A case study in diagnosis. *IEEE Transactions on systems*.

Clancey, W. (1987). Heuristic classification. *Artificial Intelligence*, 25(3):289–350.

Dague, P. (2001). Théorie logique du diagnostic à base de modèles. *Diagnostic, Intelligence Artificielle et Reconnaissance des Formes*, pages 17–105.

Davis, R. (1984). Diagnostic reasoning based on structure and behavior. *Artificial Intelligence*, 24:347–410.

De Kleer, J. and Brown, J. (1984). A Qualitative Physics Confluences. *Artificial Intelligence*, 24:7–83.

Falkeihainer, B. and Forbus, K. D. (1988). Seting up large scale qualitative models. In *Proceedings of the 7th National Conference of Artificial Intelligence*, pages 301–306, St. Paul, MN, USA.

Falkeihainer, B. and Forbus, K. D. (1991). Compositional modeling: Finding the right model for the job. *Artificial Intelligence*, 51:95–143.

Forbus, K. et al. (1984). Qualitative Process Theory. *Artificial Intelligence*, 24(1-3):85–168.

Franke, D. W. (1991). Deriving and using descriptions of purpose. *IEEE Expert*, 6(2):41–47.

Hayes, P. (1989). *Naive physics I: ontology for liquids*. Morgan Kaufmann Publishers Inc. San Francisco, CA, USA.

Le Goc, M. (2004). Sachem, a real time intelligent diagnosis system based on the discrete event paradigm. *Simulation, The Society for Modeling and Simulation International Ed.*, 80(11):591–617.

Le Goc, M. (2006). *Notion d'Observation pour le Diagnostic des Processus Dynamiques: Application à Sachem et à la découverte de Connaissances Temporelles*. Hdr, Faculté des Sciences et Techniques de Saint Jérôme, Marseille, France.

Le Goc, M. and Bénayadi, N. (2008). Dicovering expert's knowledge from sequences of discrete event class occurrences. In *To appear in the proceedings of the 10th International Conference on Enterprise Information Systems (ICEIS08)*, Barcelona, Spain, 12-16th of June 2008.

Le Goc, M., Bouché, P., and Giambiasi, N. (2005). Stochastic modeling of continuous time discrete event sequence for diagnosis. *16th International Workshop on Principles of Diagnosis (DX'05) Pacific Grove, California, USA*.

Lee, W. and Kuipers, B. (1988). Non-Intersection of Trajectories in Qualitative Phase Space: A Global Constraint for Qualitative Simulation. *Proceedings of the Seventh National Conference on Artificial Intelligence, AAAI-88*, pages 286–291.

Masse, E. and Le Goc, M. (2007). Modeling dynamic systems from their behavior for a multi model based diagnosis. In *Proceedings of the 18th International Workshop on the Principles of Diagnosis (DX'07)*, Nashville, TN, USA, May 29-31 2007.

Murthy, S. (1988). Qualitative reasoning at multiple resolutions. *Proceedings of the National Conference on Artificial Intelligence*, pages 296–300.

Reiter, R. (1987). A theory for diagnosis for first principles. *Artificial Intelligence*, 32:57–95.

Rosenberg, R. and Karnopp, D. (1983). *Introduction to Physical System Dynamics*. McGraw-Hill, Inc. New York, NY, USA.

Schreiber, A., Akkermans, J. M., Anjewierden, A. A., de Hogg, R., Shadbolt, N. R., Van de Velde, W., and Wielinga, B. J. (2000). *Knowledge Engineering and Management, The CommonKADS Methodology*. MIT Press.

Struss, P. and Dressler, O. (1989). Physical negation – integrating fault models into the general diagnostic engine. In *Proceedings of the 11th International Joint Conference on Artificial Intelligence (IJCAI-89)*, pages 1318–1323.

Thetiot, R. (1999). *Utilisation de l'Approche Multi-modèles pour l'Aide au Diagnostic d'Installations Industrielles*. Phd in sciences, Université d'Evry Val d'Essonne, France.

Zanni, C., Le Goc, M., and Frydman, C. (2005). A conceptual framework for the analysis, classification and choice of knowledge-based diagnosis systems.

Zouaoui, F., Thétiot, R., and Dumas, M. (1997). Multimodeling representation of pwr primary coolant loop. In *Proceedings Poster of International Joint Conference on Artificial Intelligence (IJCAI-97)*, page 116.