Abstract Temporal Diagnosis in Medical Domains

Johann Gamper*†and Wolfgang Nejdl

University of Hannnover
Institut für Rechnergestützte Wissensverarbeitung
Lange Laube 3, 30159 Hannover, Germany
Phone: 0049 / 511 / 762-9711
http://www.kbs.uni-hannover.de/
{gamper,nejdl}@kbs.uni-hannover.de

June 30, 2003

Abstract

Most current model-based diagnosis formalisms and algorithms are defined only for static systems, which is often inadequate for medical reasoning. In this paper we describe a model-based framework plus algorithms for diagnosing time-dependent systems where we can define qualitative temporal scenarios. Complex temporal behavior is described within a logical framework extended by qualitative temporal constraints. Abstract observations aggregate from observations at time points to assumptions over time intervals. These concepts provide a very natural representation and make diagnosis independent of the number of actual observations and the temporal resolution. The concept of abstract temporal diagnosis captures in a natural way the kind of indefinite temporal knowledge which is frequently available in medical diagnoses. We use viral Hepatitis B (including a set of real Hepatitis B data) to illustrate and evaluate our framework. The comparison of our results with the results of HEPAXPERT-I is promising. The diagnosis computed in our system is often more precise than the diagnosis in HEPAXPERT-I and we detect inconsistent data sequences which cannot be detected in the latter system.

Keywords: Diagnosis, Model-Based Reasoning, Temporal Reasoning

---

*Corresponding author. e-mail: jgamper@eurac.edu; Tel. (+39-471) 306114; Fax (+39-471) 306199.

†The authors present address is: European Academy Bozen, Weggensteinstraße 12/A, 39100 Bozen, Italy.

1 Introduction

Due to its enormous complexity, medicine has always been a challenging testbed for automatic diagnostic systems (e.g. [8, 31]). While most of these systems are based on shallow reasoning [8, 31], recently several researchers investigated the use of model-based techniques in medical diagnosis [5, 11, 14, 21]. These techniques allow the representation of deep knowledge, which is important for the development of more sophisticated problem solving systems [31].

Time constitutes an integral and important aspect of medical knowledge. Many diseases are characterized by complex temporal symptom patterns. Examples are infectious disease and diseases present over long time periods. Sophisticated medical diagnosis systems need formalisms, which allow for modeling complex dynamic behavior. Recently, several researcher have investigated temporal reasoning for medical problem-solving. Keravnou [22] describes temporal diagnostic reasoning based on time-objects which are defined in a multi-dimensional time space. Shahar and Musen [28] propose a knowledge-based system for temporal abstraction in clinical domains. Haimowitz and Kohane [19] are developing a system for the diagnosis of medical trends. Long [23] describes the addition of temporal reasoning facilities to the Heart Disease Program (HDP).

The model-based diagnosis approach [13, 26], which has been well studied for static systems, is based on a model of the structure and function (behavior) of the system to be diagnosed. The discrepancy and matching between the predicted behavior and the actual observations form the basis of the calculation of diagnoses. Different logical characterizations of diagnosis with different forms of explanation have been worked out: consistency-based [13, 16, 26, 30] and abductive [9, 12, 24, 25] diagnosis. Since many systems in the real world are dynamic, explicit or implicit representation of time had to be included in recent approaches. Friedrich and Lackinger [15] propose a very general extension of the traditional consistency-based approach to deal with temporal misbehavior, where the dynamic behavior is described by a set of logical sentences indexed by time. The approaches in [10, 14] approximate a dynamic system by a sequence of static states, each of which can be modeled using the traditional static framework. The temporal diagnosis framework in [11] is based on a causal network, where time intervals are associated with both arcs (representing delays) and nodes (representing temporal extents).

In this paper we present an alternative framework for the diagnosis of dynamic systems, which extends previous work in model-based diagnosis in several directions. Our main focus is the qualitative representation of complex temporal behavior and the abstraction of observations from single time points to time intervals. This yields an expressive and efficient framework for the diagnosis of time-dependent systems. In section 2 we briefly describe the diagnosis of hepatitis B, which motivated our work. We use it as an example throughout this paper. Section 3 introduces an interval-based temporal framework which provides a powerful language to represent qualitative temporal information. In section 4 we develop a new framework for temporal diagnosis.

Based on a process-oriented ontology, the temporal language allows a concise description of dynamic system behavior. We define the concept of an abstract observation as a temporal abstraction from observations at time points into "value assumptions over time intervals". Finally, we give a declarative definition of an explanation and show how to compute such explanations. Section 5 includes an evaluation of our system based on the diagnosis of real hepatitis B data samples. Section 6 discusses related work.

2   Example and Motivation

2.1   Diagnosis of Hepatitis B

Hepatitis B is characterized by the following serological findings: hepatitis B surface antigen (*hbsag*) and antibodies (*antihbs*), hepatitis B envelope antigen (*hbeag*) and antibodies (*antihbe*), and antibodies to hepatitis B core antigen (*antihbc* and *igmantihbc*). Each of these antigens and antibodies can take a *positive* (detectable in the patient's serum) or a *negative* value. The natural course of hepatitis B is characterized by a sequence of these findings. We distinguish four acute and four persisting courses. Figure 1 shows the typical sequence of findings in the four acute courses of hepatitis B. The bars indicate the time periods over which the findings are positive, while outside the findings are negative. Four different stages can be distinguished during the course of an hepatitis B: *incubation*, *acute*, *convalescence*, and *immunity*.

The first stage after an infection with an hepatitis B virus is the incubation stage which usually lasts from two to four months, although it may be very short (ten days) or extremely long (nine months) in some cases. Towards the end of this period HBs-antigen and HBe-antigen is detectable in the patient's serum. The appearance of *antihbc* and *igmantihbc* antibodies indicates the onset of the acute stage. Usually, after about ten weeks seroconversion of HBe-antigen occurs and *antihbe* antibodies appear. The next stage is the convalescence stage, in which seroconversion of HBs-antigen to *antihbs* occurs. In the first three courses we have a stage between the disappearance of HBs-antigen and the appearance of *antihbs*, which can last from several weeks to several months. The final stage in the course of an acute hepatitis B is the immunity stage. It is characterized by positive *antihbs* and *antihbc* antibodies. In some courses *antihbe* and *igmantihbc* antibodies can also be positive. The *antihbc* antibodies remain detectable in the patient's serum for the rest of his/her life.

2.2   Motivation

Adlassnig and Horak [1, 2] describe an expert system for the automatic interpretation of hepatitis A and B serology test results. Their system, HEPAXPERT-I, is able to handle typical courses of viral hepatitis, in addition to the situations of active and passive immunization against hepatitis A and B. The system does not consider any
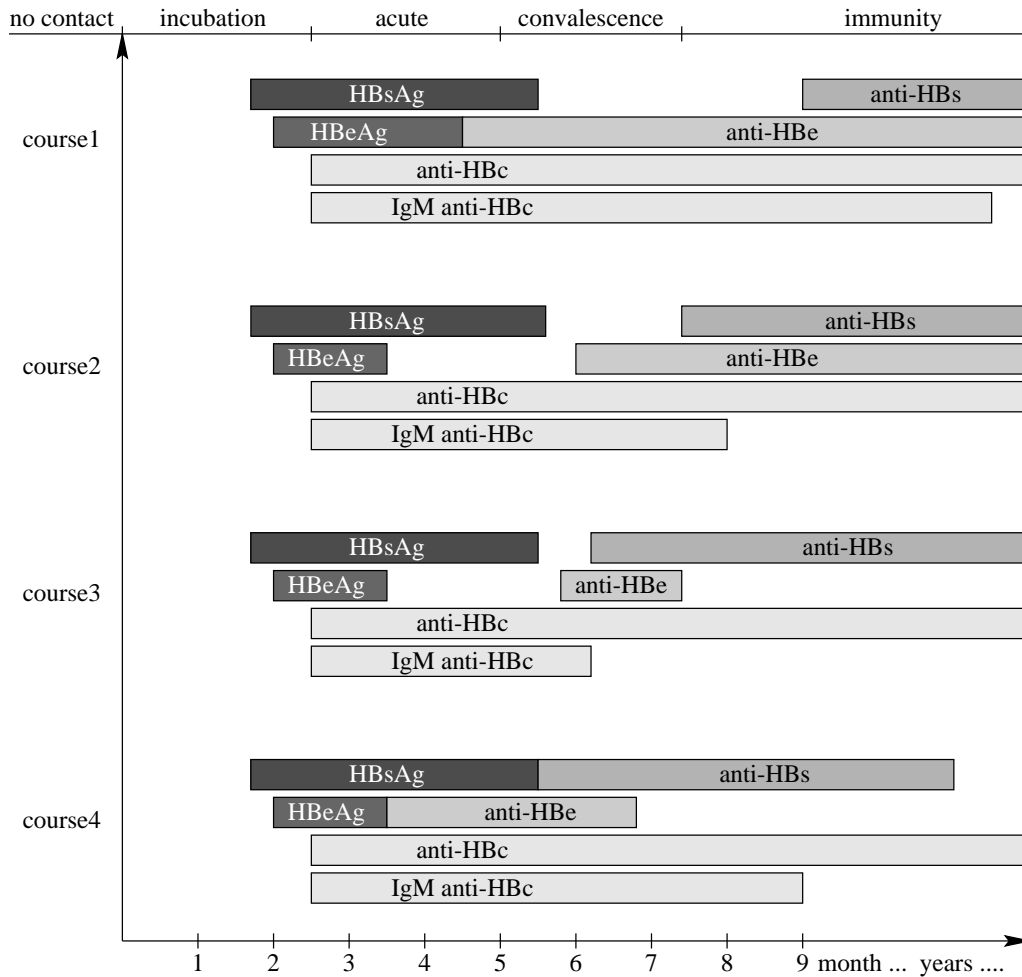
Figure 1: Typical sequence of findings in the four acute courses of hepatitis B.

additional biochemical or clinical data. In HEPAXPERT-I a sample of findings from a single time point is compared with those serological constellations that may occur during the course of viral hepatitis. The result of this comparison is an interpretation text. Since previous test results are not considered, the interpretation is not always unique.

The temporal evolution of the findings is crucial to the diagnosis of a viral hepatitis. Considering this temporal information might substantially improve the interpretation of the course of the disease. The following observations motivated our work:

- The typical courses of antigens and antibodies provide a good model for the evolution of a viral hepatitis. Additional diagnostic experience from the physician is not required to establish a diagnosis.

- The quantitative temporal information in the courses of viral hepatitis represents average values that usually differ from case to case. The qualitative temporal

4

relations among findings are much more reliable.

- All courses of hepatitis B look similar involving basically the same findings. What distinguishes these courses is the order in which antigens and antibodies occur.

- The findings are constant over long time periods. We can expect consecutive serology test results to be frequently identical.

The first observation supports the use of the model-based diagnosis approach. The last three observations stress the need to consider temporal information such as an explicit representation of complex temporal behavior or the provision of temporal abstraction mechanisms to reduce the computational complexity.

3   An Interval-Based Temporal Framework

The temporal language used is a subset of the interval-based temporal logic described by Allen and Hayes in [4]. The only primitive temporal objects consist of a nonempty set of *time intervals* which have a positive duration. We assume the existence of a maximal time interval i_{max} which includes all other time intervals. This convention is useful to represent propositions which are always true by stating that they are true throughout i_{max}.

There are thirteen basic, mutually exclusive, *qualitative temporal relations* between two time intervals (table 1). Indefinite knowledge is expressed as disjunction of basic relations. For example, the relation "time interval *I* begins before time interval *J* begins and *I* ends before *J* ends" is represented as $(I \ before \ J) \vee (I \ meets \ J) \vee (I \ overlaps \ J)$. To facilitate reading, we will use a set notation and write this disjunction as $I\{before, meets, overlaps\}J$. The empty relation $\{\}$ represents inconsistent temporal knowledge. The relation *noinfo* is defined as the set of all basic relations and indicates that we have no temporal knowledge at all. The relation *disjoint* is defined as the set $\{before, meets, after, met\}$ and means that two intervals have no common subinterval.

We do not need the full power of first-order predicate logic and restrict ourselves to conjunctions of binary relations between two time intervals. In the following we refer to these relations as *qualitative algebra relations*, for short QA relations. This restriction allows us to represent temporal information as binary constraint networks and to use efficient constraint satisfaction algorithms. A set *R* of QA relations can graphically be represented by a *QA network*, denoted as $\mathcal{R}$. Nodes represent time intervals. The directed arcs are labeled with the QA relation between the two time intervals represented by the connected nodes. We neither show arcs from a node to itself, nor reverse arcs from node *X* to node *Y* with inverse temporal relations. A missing arc between two nodes is implicitly labeled with the *noinfo* relation. The logical equivalent of a QA network is the conjunction of all relations in the network.

Basic relation	Inverse	Meaning
I before J	*J after I*	
I meets J	*J met I*	
I overlaps J	*J overlapped I*	
I starts J	*J started I*	
I during J	*J contains I*	
I finishes J	*J finished I*	
I equal J	*J equal I*	

Table 1: The basic mutually exclusive qualitative temporal relations that hold between two time intervals.

Given a QA network, the search for a consistent scenario and all feasible relations are basic temporal reasoning tasks. A *consistent scenario* is a labeling of the network, where every label consists of a single basic relation and it is possible to map the nodes to the time line such that these relations hold. There are only a finite number of consistent scenarios. A basic relation *B* is *feasible* iff there exists a consistent scenario containing *B*. Computing the feasible relations amounts to removing all inconsistent basic relations. A QA network is *consistent* iff a consistent scenario exists. In our temporal framework all of these reasoning tasks are NP-complete [33]. For the purpose of this paper, we are using standard approximation algorithms from the literature, e.g. [3, 32].

We assume a linear and dense time structure, called *time line*, which is unbound in both time directions, past and future. Time line intervals are denoted by $[B, E]$, where B is the start point and E is the end point of the interval. For convenience, we represent the time line by real numbers and implicitly assume the corresponding ordering axioms. These axioms result in QA relations between time line intervals, e.g. between the two intervals $[1, 2]$ and $[2, 5]$ we get a *meets* relation.

Given the primitive temporal objects and the time structure, we want to build propositions about the world by referring to these temporal objects. Time is introduced as an additional argument to the predicates in our language. For example, *fever*(*high*, *i*) represents a high fever throughout time interval *i*.

4   ATD — A Temporal Diagnosis Framework

In this section we present a new temporal diagnosis framework. We will use first-order predicate logic language. Variable names start with an upper case letter, constant and predicate names with a lower case letter. QA relations are represented in set notation, as discussed in the previous section.

4.1 Process Description

We begin with the description of the basic formalism. While a component-oriented ontology assuming a predefined structure is appropriate for modeling static systems, a process-oriented ontology is more suitable for representing dynamic systems. We call the model of the system the *process description PD*.

A system consists of a set *PROCS* of processes. One of these processes corresponds to the normal system behavior, while the other processes represent different faulty behaviors. Processes evolve over time, and it is useful to talk about consecutive states a process passes through. With each process we associate a set of different *states*. $s(p,i)$ represents that process p assumes state s throughout time interval i. By writing a process state assumption as $s(p,i) \wedge \tau_i$, the amount of temporal incompleteness is expressed by a conjunction of QA relations τ_i constraining the time interval i. For example, $s(p,i) \wedge i\{starts, during, finishes, equal\}[1,10]$ represents that process p assumes state s some time between 1 and 10, possibly but not necessarily over the whole time interval $[1,10]$. Processes cannot be observed directly, but evoke the appearance of observable parameters. $m(v,i)$ represents that parameter m assumes value v throughout time interval i and we use the terms manifestation, symptom, and finding interchangeably to refer to these effects. Processes are the only causes for changes in the system behavior. This strong relationship between processes and manifestations is explicitly captured by the concept of a state description model.

Definition 1 (State Description Model) A *state description model SDM* of a process $p \in PROCS$ is defined as a logical formula

$$\alpha \wedge \tau_\alpha \quad \supset \quad \beta \wedge \tau_\beta$$

where α and β are conjunctions of process state assumptions and/or manifestations, and τ_α and τ_β are conjunctions of QA relations. $\qquad \square$

A state description model specifies the causal and temporal relationships between two or more processes (process states) and between processes (process states) and the evoked manifestations. A state description model is read as follows: if a process p assumes the states as specified in α such that the temporal relations in τ_α are satisfied, then the manifestations in β and the temporal relations in τ_β are predicted. The process states in α are possible causes of the manifestations in β. The use of qualitative temporal relations in a state description model allows a concise representation of complex temporal relationships among causes and effects.

Example 1 (Hepatitis B) The various possible courses of hepatitis B are described in terms of nine processes: $PROCS = \{no\_contact, course1, \ldots, course8\}$. The first process *no_contact* (no contact with the virus) represents the normal situation. In this case, all possible findings are negative. Since only pathophysiological conditions need to be explained, the *no_contact* process is disregarded in the following. The other processes,

7

*course*1 to *course*8, stand for the various Hepatitis-B-related antibody and antigen titre time courses. The various stages in hepatitis B are represented as process states: *incubation*, *acute*, *convalescence* and *immunity*. Each of the courses is characterized by a specific pattern of positive and negative findings. We denote the six hepatitis B findings as *hbsag*, *antihbs*, *hbeag*, *antihbe*, *antihbc*, and *igmantihbc* and their possible values as *pos* (positive) and *neg* (negative). A portion of the state description model for the first course of hepatitis B is given below:

$$\forall I_{in}\forall I_{ac}\forall I_{co}\forall I_{im}\exists I_{hbsag}\ldots\exists I_{igmantihbc}$$
$$incubation(course1,I_{in})\wedge\ldots\wedge immunity(course1,I_{im})\wedge$$
$$I_{in}\{meets\}I_{ac}\wedge I_{ac}\{meets\}I_{co}\wedge I_{co}\{meets\}I_{im}$$
$$\supset\quad hbsag(pos,I_{hbsag})\wedge\ldots\wedge igmantihbc(pos,I_{igmantihbc})\wedge$$
$$I_{hbsag}\{contains\}I_{hbeag}\wedge\ldots\wedge I_{igmantihbc}\{starts\}I_{antihbc}\wedge$$
$$I_{in}\{overlaps\}I_{hbsag}\wedge\ldots\wedge I_{im}\{after\}I_{hbsag}\wedge\ldots$$

The antecedent specifies the causes: the four stages of Hepatitis B and the QA relations between these stages. The consequent specifies the effects: the positive findings, the QA relations among these findings, and the QA relations among the states in the antecedent and the findings in the conclusion. □

We refer to the qualitative temporal relations $\tau_\alpha\cup\tau_\beta$ as *temporal behavior* of a state description model. Graphically, we represent the temporal behavior as a QA network. The nodes represent the temporal extent of process states and manifestations. The arcs are labeled with the relations in $\tau_\alpha\cup\tau_\beta$. We denote the nodes by the atemporal proposition rather than by the time interval and show only a part of the temporal relations. The QA network representing the temporal behavior of *course*1 is depicted in figure 2. It differs from the temporal behavior of the other courses only by the temporal relations among the findings.

The state description model is the only concept to represent causal relationships among processes and evoked manifestations. Our main focus is on modeling faulty behavior and on generating explanations for abnormal observations. In the hepatitis example negative findings are normal, while positive findings represent abnormal values. Therefore, we use only the positive findings to describe *course*1.

In many applications additional knowledge is available and/or required. In the hepatitis example we relate a whole course to its four specific stages. To represent such knowledge we introduce definitional axioms.

Definition 2 (Definitional Axiom) A *definitional axiom* is defined as a logical formula

$$\gamma\supset\delta$$

where γ and δ are conjunctions of various domain concepts including QA relations. □

Figure 2: Temporal behavior of *course*1 using positive findings only (*convalescence* is abbreviated by *conval*, *course*1 is abbreviated by *c*1).

A definitional axiom allows us to relate various domain concepts to each other or to define new concepts from other concepts. This leads to a simpler knowledge representation and provides different views of the domain. Unlike state description models, definitional axioms do not represent causal knowledge and will not be used to explain observations.

While a manifestation assumes different values over different time intervals, only one value at a time is allowed. Similarly, only one course of a viral hepatitis actually occurs. This leads to the definition of exclusiveness axioms.

Definition 3 (Exclusiveness Axioms) The *exclusiveness axioms* for properties / processes are defined as follows:

- a property *m* can assume only one value at a time:

$$\forall V_1 \forall V_2 \forall I_1 \forall I_2 \ m(V_1, I_1) \wedge m(V_2, I_2) \wedge V_1 \neq V_2 \quad \supset \quad I_1\{disjoint\}I_2$$

- two processes p_1 and p_2 are mutually exclusive:

$$\forall I_1 \forall I_2 \ active(p_1, I_1) \wedge active(p_2, I_2) \quad \supset \quad I_1\{disjoint\}I_2$$

where $active(p, I)$ means that process *p* occurs over time interval *I*. $\qquad\Box$

Example 2 (Hepatitis B continued) We complete the modeling of hepatitis B by adding definitional and exclusiveness axioms to the process description. For each of the eight courses one definitional axiom linking the stages to the course is added. The definitional axiom for *course*1 is given as

$\forall I_{in} \forall I_{ac} \forall I_{co} \forall I_{im} \exists I_{course1} \ incubation(course1, I_{in}) \wedge acute(course1, I_{ac}) \wedge$

$convalescence(course1, I_{co}) \wedge immunity(course1, I_{im}) \wedge$

$I_{in}\{meets\}I_{ac} \wedge I_{ac}\{meets\}I_{co} \wedge I_{co}\{meets\}I_{im}$

$\quad \supset \quad active(course1, I_{course1}) \wedge I_{in}\{starts\} \ I_{course1} \wedge I_{im}\{finishes\} \ I_{course1}$

For each of the six findings we include an exclusiveness of properties axiom. For *hbsag* this axiom is given as

$$\forall V_1 \forall V_2 \forall I_1 \forall I_2 \ hbsag(V_1, I_1) \wedge hbsag(V_2, I_2) \wedge V_1 \neq V_2 \quad \supset \quad I_1 \{disjoint\} I_2$$

Only one course of hepatitis B actually occurs. This knowledge is captured by:

$$\forall P_1 \forall P_2 \forall I_1 \forall I_2 \ active(P_1, I_1) \wedge active(P_2, I_2) \wedge P_1 \neq P_2 \quad \supset \quad I_1 \{disjoint\} I_2$$

If a process P_1 occurs over time interval I_1 and a process P_2 which is different from P_1 occurs over time interval I_2, then I_1 and I_2 have no common subinterval. □

Sometimes a stronger form of exclusiveness is necessary which not only forces two processes to be disjoint, but additionally requires that only one of them can occur at all. Such knowledge can be represented by using the empty relation {} instead of the *disjoint* relation in the corresponding exclusiveness of processes axiom.

4.2 Observations and Abstract Observations

Observations tell us how the system actually behaves. An *observation* is defined as a measurement of an observable parameter at a particular time point. In the hepatitis example, observations are the results of hepatitis B serology tests. For example, $hbsag(pos, 1)$ represents a positive test result for *hbsag* at time 1.

For hepatitis, but also in other domains it is reasonable to assume continuous persistence of parameters. This allows a more intuitive and natural interpretation of observations over time intervals. We define the following temporal abstraction strategy:

1. For each observation construct a small time interval i around the observation point, such that the parameter m assumes value v throughout i, i.e. we use only intervals in our representation and algorithms.

2. Join consecutive time intervals representing the same parameter assuming the same value and construct larger intervals. Consecutive time intervals for the same parameter assuming different values are constrained by the *meets* relation.

The first abstraction is justified by the continuous persistence of parameter values in dynamic systems and the dense time model we are using. The second highly depends on the dynamics of the system as well as on the measurement frequency. Including such information is beyond the scope of this paper and not necessary in the hepatitis domain. More sophisticated methods on measurement abstraction from time points into time intervals are described in the literature, e.g. [27, 28, 29].

The result of performing this temporal abstraction for a set of observations is a set of parameter assumptions over time intervals plus a set of QA relations constraining these intervals. This leads to the concept of an abstract observation.

Definition 4 (Abstract Observation) An *abstract observation* for a parameter *m* assuming value *v* during time interval *i* is defined as a logical formula

$$m(v,i) \wedge \tau$$

where τ is a conjunction of QA relations between the temporal extent *i* and the real time line and/or other abstract observations. □

Example 3 (Hepatitis B continued) Suppose that a patient's blood is examined each month for hepatitis B antigens and antibodies. The results of these examinations are shown in figure 3a. The letters "p" and "n" denote positive and negative test results. From these observations we construct fourteen abstract observations indicated by thick lines for positive and by thin lines for negative values. The symbol "?" indicates an uncertain time of parameter value change. A portion of the abstract observations is given below:

$$hbsag(neg, i_1) \wedge hbsag(pos, i_2) \wedge hbsag(neg, i_3) \wedge antihbs(neg, i_4) \wedge \ldots \wedge$$
$$i_1\{meets\}i_2 \wedge i_2\{meets\}i_3 \wedge i_2\{during\}i_4 \wedge \ldots \wedge$$
$$i_1\{overlaps\}[1,2] \wedge i_2\{overlapped\}[1,2] \wedge i_2\{overlaps\}[5,6] \wedge \ldots$$

For example, *hbsag* turns positive between time 1 and 2 and turns again negative between times 5 and 6; the time interval of positive *hbsag* is during the time interval of negative *antihbs*. □

Two different types of qualitative temporal information can be distinguished: temporal location and order. For an abstract observation $m(v,i) \wedge \tau$, the *temporal location* consists of those relations in τ which constrain the temporal extent *i* relative to the time line. In the above example, the temporal location of the positive *hbsag* is $i_2\{overlapped\}[1,2] \wedge i_2\{contains\}[5,6]$. The *temporal order* of an abstract observation $m(v,i) \wedge \tau$ consists of all relations in τ between the temporal extent *i* and the temporal extent of other abstract observations. A portion of the temporal order of the positive *hbsag* in the above example is $i_1\{meets\}i_2 \wedge i_2\{meets\}i_3 \wedge i_2\{during\}i_4$.

AOBS denotes a set of abstract observations. The qualitative temporal relations of a set *AOBS* are graphically represented by a QA network $\mathcal{AOBS}$. The nodes in the upper level represent the temporal order. The lower level represents the real time line. The arcs between the two levels represent the temporal location. Figure 3b shows part of the QA network for the abstract observations in the above example. For clarity, not all relations are shown.

The concept of abstract observations is an important shift from a discrete view based on time points to a view driven by changes of observations. This can considerably improve the efficiency of the diagnostic process. In the worst case the value of a parameter changes at every observation. Then each abstract observation covers exactly one observation. Our example shows that in many cases parameters are stable over long time periods.
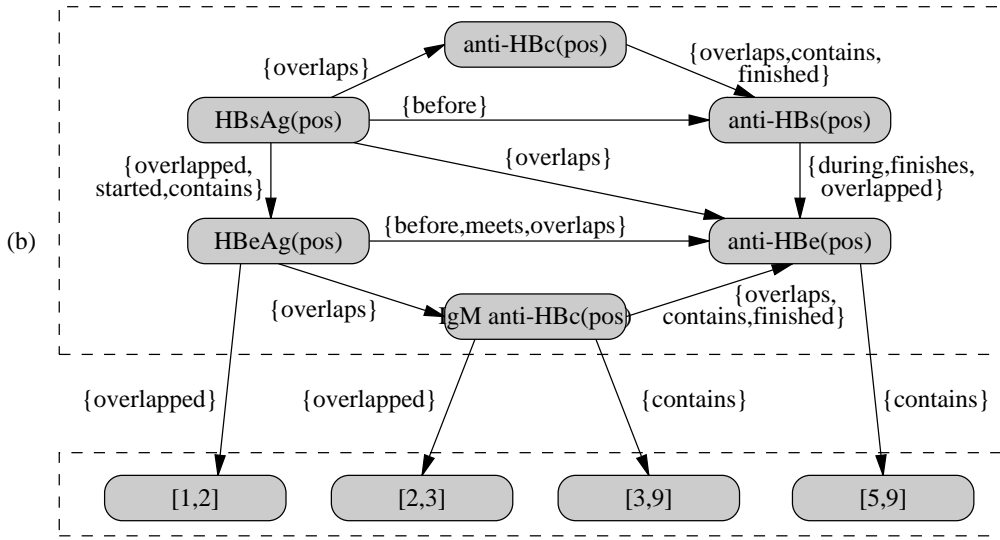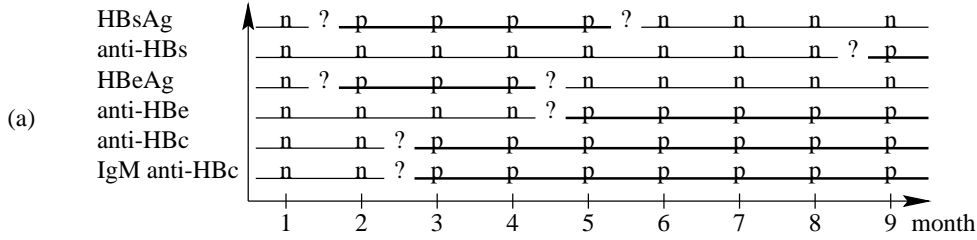
(a)

HBsAg	n	?	p	p	p	p	?	n	n	n	n
anti-HBs	n	n	n	n	n	n	n	n	?	p
HBeAg	n	?	p	p	p	?	n	n	n	n
anti-HBe	n	n	n	n	?	p	p	p	p	p
anti-HBc	n	n	?	p	p	p	p	p	p
IgM anti-HBc	n	n	?	p	p	p	p	p	p

1 2 3 4 5 6 7 8 9 month

(b)

anti-HBc(pos)

{overlaps}

{overlaps,contains, finished}

HBsAg(pos)   {before}   anti-HBs(pos)

{overlapped, started,contains}

{overlaps}

{during,finishes, overlapped}

HBeAg(pos)   {before,meets,overlaps}   anti-HBe(pos)

{overlaps}   IgM anti-HBc(pos)   {overlaps, contains,finished}

{overlapped}   {overlapped}   {contains}   {contains}

[1,2]   [2,3]   [3,9]   [5,9]

Figure 3: (a) Observations (b) QA network representing the temporal relations of the positive abstract observations.

The concept of abstract observations can also express indefinite knowledge per se. Patients frequently do not remember the exact time course of their symptoms, but are able to temporally relate the symptoms to each other. By using the *equal* relation, an abstract observation can also represent exact temporal knowledge.

4.3 Abstract Temporal Diagnoses

The aim of diagnostic reasoning is to find an explanation for the observed abnormal system behavior. Using abstract observations as described above, we will provide a declarative definition of the concept of explanation in our framework, and develop an algorithm for computing diagnostic explanations.

4.3.1 Declarative Definition

First, we have to determine those literals which we accept as an explanation. Following the terminology of abductive reasoning we refer to these literals as abducibles. An

abducible is a literal for which no further explanation can be generated, and hence should be accepted unconditionally. This is the case for all literals occurring only in the antecedent of state description models, for example process state assumptions ($acute(course1, i)$ etc.) Usually, it is not possible to determine the exact temporal extent of such process state assumptions, although they can be constrained using temporal relations. As a consequence, an explanation may include temporal relations, which are also abducible literals. Using such abducibles, we are able to generate explanations like $acute(course1, i) \wedge i\{overlapped\}[2, 3] \wedge i\{overlaps\}[4, 6]$, stating that the acute stage of *course*1 starts between time 2 and 3 and ends between time 4 and 6.

Two extremes of diagnostic reasoning using different notions of explanation have widely been used: consistency-based diagnosis [13, 16, 26, 30] where an explanation has to be consistent with the observations, and abduction-based diagnosis [9, 12, 24, 25] where an explanation logically entails the observations. Following [12] we develop an abduction-based framework with additional consistency constraints. Abnormal observations are explained abductively, and it is required that an explanation is consistent with all observations, both normal and abnormal observations. Recall that abstract observations are denoted by *AOBS*. $AOBS^+ \subseteq AOBS$ denotes the abnormal abstract observations. In the hepatitis example positive findings are abnormal, while negative findings are normal.

Definition 5 (def:atd Abstract Temporal Diagnosis Given are a process description *PD* and abstract observations $\langle AOBS^+, AOBS \rangle$, where $AOBS^+ \subseteq AOBS$ are the abnormal abstract observations. A conjunction Δ of abducibles is an *abstract temporal diagnosis* for $\langle AOBS^+, AOBS \rangle$ iff

1. $PD \cup \Delta \models AOBS^+$

2. $PD \cup \Delta \cup AOBS$ is consistent.

$\square$

Condition (1) captures the abductive part: abnormal abstract observations $AOBS^+$ are logically entailed by the process description and Δ. Condition (2) captures the consistency-based part: an explanation is consistent with the process description and all abstract observations. By including the temporal order and/or temporal location into $AOBS^+$, we can force stronger temporal conditions for specific temporal relations, establishing various forms of abstract temporal diagnoses. Which temporal relations should be included in $AOBS^+$ depends mainly on the available domain knowledge. A detailed discussion on this issue is given in [17].

4.3.2 Computation

The computation of diagnoses is best understood as a non-monotonic process, where assumptions are successively revised until some termination criteria are satisfied.

These assumptions consist of abducibles and represent hypotheses about the functioning of the system to be diagnosed. The strategy to compute an abstract temporal diagnosis can roughly be described as follows: we start with the empty assumption set and consistently backchain from the abnormal abstract observations to the abducibles by applying the cause-effect relationships. Two sequential tasks can be distinguished: abduction and prediction. The idea of *abduction* is to find a set of abducibles which, together with the model, logically entail the data to be explained. In the context of our framework, abduction looks for a process description model which predicts a nonempty set of abstract observations in $AOBS^+$. The literals in the antecedent of such a model are introduced as new assumptions, which can either be abducibles or need to be explained by further abduction steps. The *prediction* task is a deductive process and explicitly infers all new facts which logically follow from a knowledge base. In the context of our framework, after the introduction of new assumptions in the abductive step, the prediction task infers all new facts which follow from the process description and the actual assumptions.

In the algorithms for computing abstract temporal diagnoses we make the following assumptions which are all satisfied in the hepatitis example:

- The process description *PD* is acyclic, i.e. we cannot build chains of state description models such that a parameter appears more than once. Put more formally, for each causal chain $\alpha_1 \supset \alpha_2$, $\alpha_2 \supset \alpha_3$, ..., $\alpha_{n-1} \supset \alpha_n$, α_i ground, a parameter appears at most in one of the α_i's. This assumption is reasonable in any causal theory, since a parameter cannot be explained by itself.

- The abnormal abstract observations $AOBS^+$ contain no temporal relations. No temporal relations are abductively explained, but are only used for consistency checking. This restriction makes the computation of diagnoses more efficient. Moreover, since the process description models in the hepatitis example contain only definite temporal knowledge (single basic QA relations), using the temporal order only for consistency checking yields the same diagnoses as explaining the temporal order abductively.

- The temporal abstraction from the observations into the abstract observations is correct.

- Each abstract observation has only one cause, i.e. the entire temporal extent of an abstract observation can be explained by a single state description model.

In some applications the latter three assumptions are rather restrictive. It remains part of our future work to extend the algorithms for more general cases.

ATD

The recursive function ATD in figure 4 implements the main algorithm to compute abstract temporal diagnoses. ATD has four input parameters: the process description

PD; a set *C* of data which have to be covered by an abstract temporal diagnosis and at the first call is equal to *AOBS*$^+$; a QA network $\mathcal{N}$ which at the first call is equal to the QA relations in *AOBS*; a set *A* of abducibles which at the first call is equal to the empty set. ATD returns all abstract temporal diagnoses for *AOBS*$^+$ and *AOBS*, respectively. *A* and $\mathcal{N}$ together define a partial abstract temporal diagnosis, which initially is equal to the empty diagnosis and is extended during the computation process. The set *D* is used to collect the diagnoses computed so far and is initialized to the empty set. Since every parameter has a temporal extent, a consistency check of a diagnosis amounts to testing temporal consistency. The global network $\mathcal{N}$ is used to maintain temporal consistency.

If *C* is empty, all abnormal data are consistently covered. The function DIAGNOSIS constructs an abstract temporal diagnosis consisting of the abducibles *A* and those QA relations in $\mathcal{N}$, that exist among these abducibles and from these abducibles to the time line. ATD terminates and returns *D*.

Otherwise, there are still data to be covered. We begin with the abduction and consider each relevant state description model combined with each possible covering of data in *C* by that model. A state description model is relevant in a particular state of the diagnostic process, if it predicts some elements in *C*. Given a particular state description model *SDM* of the form $\alpha \wedge \tau_\alpha \supset \beta \wedge \tau_\beta$, we abductively infer the antecedent of *SDM* as an explanation for those elements in *C* which are covered by a manifestation in β. Pairs of a manifestation in β and a covered element in *C* are specified in the binding list *BL*. The binding list is a nonempty set of tuples $\langle m, c \rangle$, where *m* is a manifestation in β, *c* is an element in *C*, and *m* and *c* both represent the same parameter and the same value. Then the function ASSUME is invoked, which extends the global network $\mathcal{N}$ by the temporal behavior of *SDM* and returns the network $\mathcal{N}_{tmp}$.

If $\mathcal{N}_{tmp}$ is consistent, we use PREDICTION (simple forward-chaining). PREDICTION recursively infers all new facts triggered by assuming the antecedent of *SDM* in the abductive step. It applies the definitional axioms and the exclusiveness axioms, possibly extending the global network $\mathcal{N}$ by new nodes and new temporal constraints. This process is repeated until no additional rules can be applied or until an inconsistency is detected. The global network may be altered during prediction and is returned in $\mathcal{N}_{new}$.

If $\mathcal{N}_{new}$ is consistent, *SDM* consistently explains a nonempty subset of *C*. We add the abducibles in α to *A*, remove the covered data in *BL* from *C*, and add the non-abducibles in *SDM* to C^1. Now ATD is called recursively to compute all diagnoses for $AOBS_{new}^+$ by extending the actual partial diagnosis.

---

[1]The antecedent of *SDM* might contain both abducibles and non-abducibles. The non-abducibles introduced by assuming the antecedent of *SDM* must be explained by other models. In the hepatitis example the state description models contain only abducibles in the antecedent.

function ATD$(PD, C, \mathcal{N}, A)$ **returns** set of all abstract temporal diagnoses
inputs PD... process description
          C... data to be covered; at the first call, C is equal to the set $AOBS^+$
          $\mathcal{N}$... QA network; at the first call, $\mathcal{N}$ is equal to the QA relations in $AOBS$
          A... set of abducibles; at the first call, A is equal to the empty set
begin
  $D \leftarrow \emptyset$
  if $C = \emptyset$ **then** % *No more data have to be covered*
     $D \leftarrow \text{DIAGNOSIS}(A, \mathcal{N})$
  else
       % *Apply a state description model to explain a nonempty set of elements in C*
       for each Relevant *SDM* of the form $\alpha \wedge \tau_\alpha \supset \beta \wedge \tau_\beta$ in *PD* and
               binding list *BL* among manifestations in β and data in C **do**
       begin
          % *Add the temporal behavior of SDM to the global network* $\mathcal{N}$
          $\mathcal{N}_{tmp} \leftarrow \text{ASSUME}(SDM, BL, \mathcal{N})$
          if $\text{CONSISTENT}(\mathcal{N}_{tmp})$ **then**
          begin
             % *Infer all new facts triggered by assuming the antecedent* $\alpha \wedge \tau_\alpha$
             $\mathcal{N}_{new} \leftarrow \text{PREDICTION}(PD, SDM, \mathcal{N}_{tmp})$
             if $\text{CONSISTENT}(\mathcal{N}_{new})$ **then**
             begin
                % *SDM explains some elements in C; update A and C*
                $A_{new} \leftarrow A \cup \{\text{Abducibles in } \alpha\}$
                $C_{new} \leftarrow C \setminus \{\text{Covered data in } BL\} \cup \{\text{Non-abducibles in } \alpha\}$
                % *Compute all diagnoses for* C_{new} *by extending the*
                % *current partial diagnosis represented by* $\mathcal{N}_{new}$ *and* A_{new}
                $D \leftarrow D \cup \text{ATD}(PD, C_{new}, \mathcal{N}_{new}, A_{new})$
             end
          end
       end
    return D
end

Figure 4: Function ATD which computes all abstract temporal diagnoses.

Assume

The function ASSUME in figure 5 has three input parameters: a state description model *SDM*, a binding list *BL*, and the global network $\mathcal{N}$. The return value is $\mathcal{N}$ extended by the temporal behavior of *SDM* and containing only the feasible relations. The basic task of the function is to "assume" the abducibles in *SDM*. This is done by adding the temporal behavior of *SDM* to the global network $\mathcal{N}$ and then computing all feasible relations of the extended network.

For each process state assumption and manifestation in the state description model a new node in $\mathcal{N}$ is created. We then compute temporal constraints among these new nodes and the old nodes in $\mathcal{N}$. This task is performed within two loops ranging over the new and the old nodes, respectively. Three cases are distinguished. First, two nodes representing a pair in the binding list *BL* are constrained by the *equal* relation. Second, if two nodes represent the same parameter value assumption, but are not in the binding list, the first node is either within the second one or it is strictly before or after the second one. This is represented by the relation $\{before, after, starts, during, finishes, equal\}$. Finally, if none of the two previous cases is given, we assert the *noinfo* relation. The QA relations resulting from this analysis plus the temporal behavior of *SDM* are added to the global network $\mathcal{N}$, which then is passed to the function FEASIBLE to compute all feasible relations of the extended network.

Example 4 (Hepatitis B continued) In this example, the computation of abstract temporal diagnoses is illustrated in detail for hepatitis B. The process description *PD* consists of a state description model and a definitional axiom for each of the eight courses of hepatitis B, one exclusiveness of processes axiom, and one exclusiveness of properties axiom for each of the six findings. We compute all abstract temporal diagnoses for the abstract observations which result from the observations shown in figure 3a. The abnormal abstract observations which have to be covered by a diagnosis are the positive abstract observations without any temporal constraints, i.e.

$$
\begin{aligned}
AOBS^+ \;=\; & \{hbsag(pos, i_{hbsag}), antihbs(pos, i_{antihbs}), \\
& hbeag(pos, i_{hbeag}), antihbe(pos, i_{antihbe}), \\
& antihbc(pos, i_{antihbc}), igmantihbc(pos, i_{igmantihbc})\}
\end{aligned}
$$

Since *C* is not empty (it is equal to $AOBS^+$), we start to consider each relevant state description model. Suppose that we begin with the model SDM_{c1} for *course*1 and a binding list *BL* which contains six tuples like $\langle hbsag(pos, I_{hbsag}), hbsag(pos, i_{hbsag})\rangle$ for the positive *hbsag* (I_{hbsag} is an existentially quantified variable, i_{hbsag} is a constant). The function ASSUME adds the temporal behavior of SDM_{c1} to the global network $\mathcal{N}$ and adds an *equal* relation between each manifestation in SDM_{c1} and the corresponding abstract observation. The network $\mathcal{N}_{tmp}$ returned by ASSUME is consistent. We invoke the function PREDICTION. The exclusiveness axioms for properties yield

---

function ASSUME(*SDM*, *BL*, $\mathcal{N}$) **returns** modified $\mathcal{N}$
inputs   *SDM*...state description model $\alpha \wedge \tau_\alpha \supset \beta \wedge \tau_\beta$
        BL...binding list
        $\mathcal{N}$...QA network
begin
  $Tmp \leftarrow \emptyset$
  $\mathcal{N} \leftarrow \mathcal{N} \cup$ NEW-NODES($\alpha \cup \beta$) `% Add new nodes for αUβ to` $\mathcal{N}$
  for each New node n_1 in $\mathcal{N}$ **do**
    for each Old node n_2 in $\mathcal{N}$ **do**
    begin
      if $\langle n_1, n_2 \rangle$ represents a pair in *BL* **then**
        `% Manifestation` n_1 `covers abstract observation` n_2
        $R \leftarrow n_1\{equal\}n_2$
      else if n_1 and n_2 represent the same parameter assumption
          (but are not a pair in the binding list) **then**
        `%` n_1 `is within` n_2 `or completely outside`
        $R \leftarrow n_1\{before, after, starts, during, finishes, equal\}n_2$
      else
        $R \leftarrow n_1\{noinfo\}n_2$
      $Tmp \leftarrow Tmp \cup \{R\}$
    end
  return FEASIBLE($\mathcal{N} \cup \tau_\alpha \cup \tau_\beta \cup Tmp$)
end

Figure 5: Function ASSUME which adds the temporal behavior of a state description model to the global network and computes all feasible relations.

a *disjoint* relation between each manifestation in SDM_{c1} and each corresponding negative abstract observation. Finally, the feasible relations of the extended network are computed and returned as $\mathcal{N}_{new}$ (see figure 6). $\mathcal{N}_{new}$ is consistent. We update the set of abducibles A and get

$$A_{new} \;=\; \{incubation(course1, I_{in}), acute(course1, I_{ac}),$$
$$convalescence(course1, I_{co}), immunity(course1, I_{im})\}$$

Since *course*1 covers all abstract observations in C and does not introduce new non-abducibles, the set C_{new} is empty.

ATD is called recursively in order to compute all diagnoses for C_{new}. In the next call, C is empty, and the function DIAGNOSIS composes the first final diagnosis Δ. Figure 7 shows a portion of the global network, where Δ is indicated by light nodes and arcs. Manifestations and the covered abstract observations are represented by a single node. The first recursive call of ATD successfully terminates and returns Δ which is collected in D. In turn, the remaining courses are considered; all of them
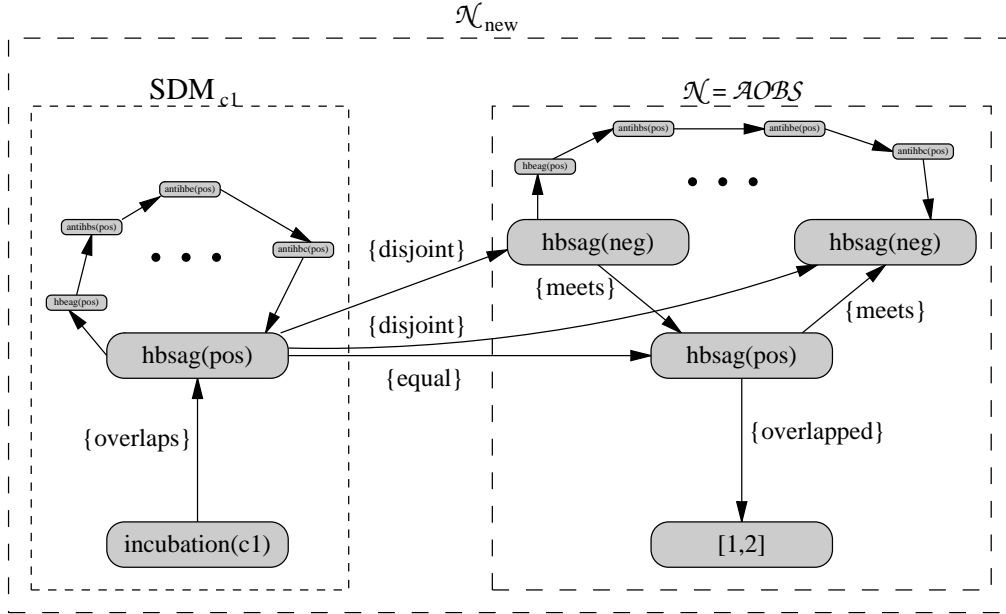
Figure 6: A portion of the network $\mathcal{N}_{new}$ after inserting the temporal behavior of *course*1 into the network $\mathcal{N}$ and applying the exclusiveness and definitional axioms.

are temporally inconsistent with the abstract observations. The original call of ATD terminates and returns the set *D* which contains the abstract temporal diagnosis

$$
\begin{aligned}
\Delta \quad = \quad & \exists I_{in} \exists I_{ac} \exists I_{co} \exists I_{im} \; incubation(course1, I_{in}) \wedge acute(course1, I_{ac}) \wedge \\
& convalescence(course1, I_{co}) \wedge immunity(course1, I_{im}) \wedge \\
& I_{in}\{overlaps\}[2,3] \wedge I_{ac}\{overlapped\}[2,3] \wedge I_{ac}\{overlaps\}[4,6] \wedge \\
& I_{co}\{overlapped\}[4,6] \wedge I_{co}\{overlaps\}[5,9] \wedge I_{im}\{overlapped\}[5,9] \wedge \\
& I_{in}\{meets\}I_{ac} \wedge I_{ac}\{meets\}I_{co} \wedge I_{co}\{meets\}I_{im}
\end{aligned}
$$

The incubation stage starts before time 2, in the interval $[2,3]$ the incubation stage finishes and the acute stage starts, in the interval $[4,6]$ the acute stage finishes and the convalescence stage starts, in the interval $[5,9]$ the convalescence stage finishes and the immunity stage starts, the latter one lasts beyond time 9. The *meets* relation constrains consecutive stages.

To verify that *course*2 is not a diagnosis, let us look at the relation between the positive *hbsag* and the positive *antihbe*: the state description model predicts a *before* relation, and in the abstract observations there is an *overlaps* relation. □

A few remarks are worthwhile. The use of abstract observations instead of observations in the diagnostic process leads to an event-driven [18] reasoning independent of the number of specific observations and the temporal resolution. A total of 54 observations at single time points have been reduced to 14 abstract observations to be
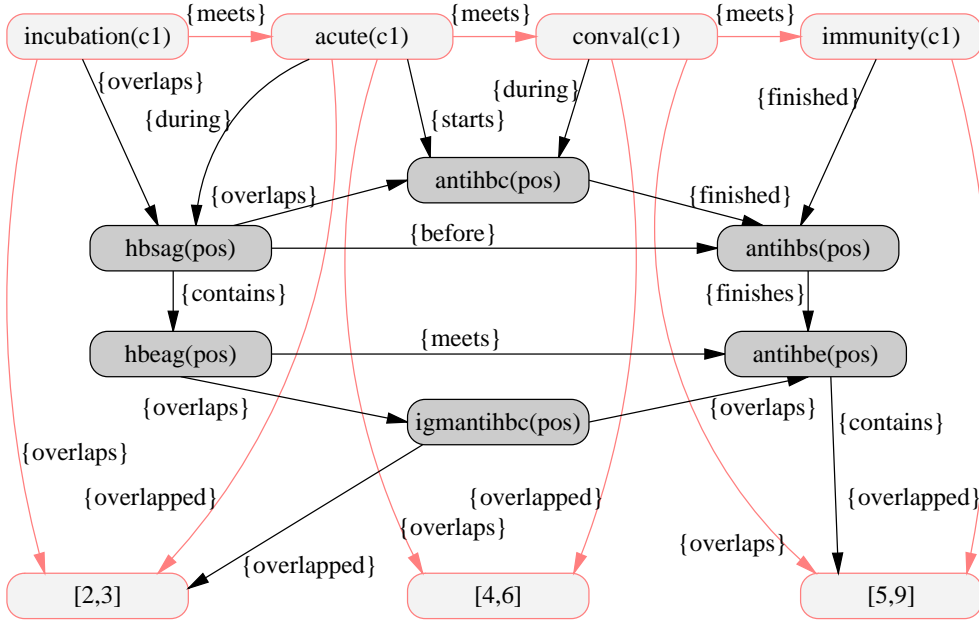
Figure 7: A portion of the global network $\mathcal{N}$ after the termination of the function ATD. Light nodes and arcs indicate the computed abstract temporal diagnosis.

considered by the diagnosis function. The diagnosis systems discussed in [10, 14], that approximate a dynamic system by a sequence of static systems, generate local diagnoses for each observation time and combine them to global diagnoses, possibly leading to a large number of global diagnoses. In the above example, these systems generate local diagnoses at nine time points. Let us look at the observations at time 6, 7 and 8 in figure 3a. At each of these time points both the convalescence and the immunity stage provide an explanation, yielding four possibilities for global diagnoses. Contrary to [10, 14], we get a single abstract temporal diagnosis which captures all these possible diagnoses. Obviously, the efficiency gain from our framework highly depends on the frequency and persistence of observations.

Another issue concerns the natural representation of indefinite temporal knowledge. Abstract temporal diagnoses as process state assumptions over indefinite time intervals represent in a natural way the vague knowledge humans often have about the evolution of a system. We believe that a representation in terms of QA relations is more natural than representing different alternative diagnoses as in [10, 14].

5 Evaluation

We implemented a prototype of our temporal diagnosis system ATD and performed a practical evaluation. A set of real hepatitis B data samples was used. From a total of approximately 5000 patients with an overall of more than 24000 serological hepatitis B test results we randomly chose 100 patients as test cases: 32 of them consisted of only

a single test result and 68 consisted of more than one data sample from different time points, the most being 23 data samples in one test case. The function ATD was run on each of these 100 test cases, and we compared the computed abstract temporal diagnoses with the interpretation of the data samples (each at a time) by the HEPAXPERT-I system [1, 2].

Two differences between the two diagnosis systems are important to keep in mind during the analysis. First, while ATD only uses the eight typical courses of hepatitis B, HEPAXPERT-I uses additional knowledge such as the courses after an active or passive immunization. Hence, some data samples which are correctly interpreted in HEPAXPERT-I are inconsistent in ATD. Second, because of the different nature of diagnosis in the two systems, a comparison is not always straightforward. An abstract temporal diagnosis explains the sequence of all data samples in a test case as a whole by specifying the temporal extension of various stages. This "overall" diagnosis must be compared with the HEPAXPERT-I-interpretation text generated for each single data sample. Since the interpretation of a single data sample is not always unique, the generated text captures all possible interpretations.

Before we present a detailed analysis, let us briefly summarize the results of this practical evaluation (see table 2). In 65 cases ATD returned an abstract temporal diagnosis, whereas in 35 cases ATD was not able to generate a diagnosis at all. The HEPAXPERT-I system considered 96 test cases as consistent (all data samples are consistent), only four cases were inconsistent (at least one data sample is inconsistent). All 65 cases for which ATD computed a diagnosis were considered consistent in HEPAXPERT-I. Similarly, for all cases that HEPAXPERT-I found at least one inconsistent data sample, ATD computed no diagnosis. A more detailed analysis of our practical evaluation is summarized in table 3 and is discussed in the next section.

ATD		HEPAXPERT-I	
A diagnosis computed	65	All data samples are consistent	96
No diagnosis computed	35	At least one data sample is inconsistent	4

Table 2: Overall evaluation of 100 test cases: ATD versus HEPAXPERT-I.

5.1 Consistent Test Cases

Let us consider the 65 test cases for which ATD computed at least one abstract temporal diagnosis. In HEPAXPERT-I all data samples of these cases were interpreted as consistent data samples and this interpretation did not contradict the abstract temporal diagnosis. In fifteen cases the abstract temporal diagnosis is considerably more specific than the HEPAXPERT-I interpretation. Two aspects should be stressed: First, ATD explicitly distinguishes between different courses of hepatitis B, which is not the case in the interpretation texts of HEPAXPERT-I. Second, an abstract temporal diagnosis rules out several possible interpretations of a single isolated data sample, which

Result of ATD	Number of test cases
Consistent test cases	**65**
More specific than HEPAXPERT-I	15
Inconsistent test cases	**35**
Due to a single data sample	20
ATD and HEPAXPERT-I	4
ATD only	16
Due to a data sequence (ATD only)	15
Oscillating *antihbc*	8
antihbc(pos){before}antihbc(neg)	7
antihbc(pos){before}hbsag(pos)	2

Table 3: Detailed evaluation of 100 test cases: ATD versus HEPAXPERT-I.

have to be considered in the HEPAXPERT-I interpretation text. The accuracy of ATD is mainly a result of the ability to consider previous data and to interpret the whole sequence of data samples.

We illustrate the power of abstract temporal diagnoses on a test case which consists of data samples at three consecutive time points as shown in table 4. The interpretation of these three data samples in HEPAXPERT-I can be summarized as follows: the first data sample indicates the immunity stage, the acute stage, or the convalescence stage of an acute course of hepatitis B; the second data sample indicates either the transition from the acute to the convalescence stage of an acute course, the immunity stage of an acute course, or a persisting course of hepatitis B; the interpretation of the third data sample states that in general the serum is not infectious, but to be sure, additional tests are recommended. ATD identifies the fourth course of an acute hepatitis B with the following temporal constrains: incubation and acute stages finish before time 1; the convalescence stage starts before time 1 and finishes before time 2; the immunity stage starts before time 2 and persists beyond time 3. This diagnosis is much more specific than the HEPAXPERT-I interpretation: we rule out the acute stage mentioned in the interpretation text of the first data sample, since the acute stage finishes before time 1; we rule out some of the possibilities in the interpretation text of the second data sample; we identify the third data sample as the immunity stage and therefore do not recommend additional tests.

Findings	Test results		
hbsag	*neg*	*neg*	*neg*
antihbs	*pos*	*neg*	
antihbc	*pos*	*pos*	
Time points	1	2	3

Table 4: A test case with hepatitis B data samples at three consecutive time points.

5.2 Inconsistent Test Cases

In 35 cases ATD was not able to compute a diagnosis at all. Assuming that our knowledge about hepatitis B is complete we consider them as inconsistent cases. In 20 cases the inconsistency was caused by a single data sample. In the remaining 15 cases each single data sample was consistent, but evaluating the sequence of data samples yielded an inconsistency.

5.2.1 Inconsistent Data Samples

First, we consider those twenty test cases that are inconsistent due to a single data sample. In four cases, the inability of ATD to compute a diagnosis agrees with the interpretation of HEPAXPERT-I, i.e. each of these test cases contains at least one data sample which HEPAXPERT-I classified as inconsistent. HEPAXPERT-I identified no other inconsistencies (see also table 2).

The remaining sixteen inconsistencies were only found in ATD. Three different patterns of findings were responsible for these inconsistencies. Since none of these patterns occurs in any of the eight typical courses, they are beyond the knowledge base incorporated in ATD. Fourteen cases contained a data sample with a positive *antihbs* and a negative *antihbc*, which contradicts the fact that after a natural course of an hepatitis B *antihbc* will be detectable in the patient's serum for the rest of his/her life. HEPAXPERT-I interpreted this data sample as follows: "The patient is immune to the hepatitis B virus owing to vaccination or passive immunization. In one case an inconsistency arose due to a positive *hbsag* which coincides with a positive *antihbs*. HEPAXPERT-I interpreted this constellation as the late acute stage of an acute hepatitis B. Finally, in one case a positive *hbeag* and a positive *antihbs* coincided, which was interpreted as a very improbable constellation by HEPAXPERT-I. So HEPAXPERT-I uses additional knowledge beyond the eight typical courses, e.g. courses of findings after an immunization.

We can extend our system in order to capture such knowledge by including additional state description models into the process description or by relaxing some of the constraints in the eight typical courses. For example, the case that a positive *hbsag* coincides with a positive *antihbs* could be explained by *course*4 if we relax the *meets* relation between these two findings to the indefinite relation {*meets*, *overlaps*}. In fact, due to the variability of biological systems the *meets* relation is rather strong.

5.2.2 Inconsistent Sequences of Data Samples

In fifteen cases ATD detected an inconsistent sequence of data samples. These inconsistencies could of course not be detected by HEPAXPERT-I. In eight cases we had an oscillating *antihbc*. In two cases *antihbc* appeared before *hbsag*. In seven cases the test result of *antihbc* was first positive and later negative. All these inconsistencies were correctly detected with respect to the eight typical courses of hepatitis B.

5.3 Final Discussion

The practical evaluation illustrates the benefits of explicitly representing temporal information and diagnosing whole sequences of data samples. In fifteen cases the abstract temporal diagnosis was much more concise than the interpretation in HEPAXPERT-I. The four inconsistencies detected in HEPAXPERT-I were detected in ATD as well. The remaining 31 inconsistencies that were only detected in ATD are justified. Sixteen of them were caused by a single data sample due to the fact that ATD used only the eight typical courses. The remaining fifteen inconsistencies, caused by an inconsistent sequence of (consistent) data samples, were correctly detected by ATD.

Reasoning explicitly with temporal information is rather important to detect inconsistent data, in particular inconsistent data sequences due to measurement errors or faulty documentation of test results. Depending on the domain and the accuracy of measurements, the recognition of inconsistent data sequences is an important reasoning task. An isolated interpretation of the most recent data sample cannot detect such inconsistencies and might yield wrong diagnoses.

6 Comparison to Related Work

A few researchers have recently begun to develop frameworks for diagnosing time-dependent systems [6, 7, 10, 11, 14, 15]. The approach described by Brusoni et al. in [6, 7] is most closely related to our work. Brusoni et al. propose a logical characterization of temporal abductive diagnosis as well as an algorithm to compute such diagnoses. Their approach extends the traditional model-based approach for atemporal abductive diagnosis by a temporal constraint language and a temporal reasoning system. This temporal diagnosis framework is strongly related to our system, with a few but significant differences. In the state description models we distinguish temporal relations appearing in the antecedent from those appearing in the consequent, which is not the case in an explanatory formula representing cause-effect relationships in [6]. This distinction allows us to capture a larger class of logical definitions of temporal diagnoses. Depending on the specification of abnormal abstract observations ($AOBS^+$), we can explain various temporal relations in an abductive way. Brusoni et al. cannot explain any temporal relations abductively, rather temporal relations are only used for consistency checking. We capture their definition of a diagnosis if $AOBS^+$ contains no temporal relations. While Brusoni et al. use a more expressive temporal language (including qualitative and quantitative temporal constraints) and a more efficient temporal reasoning system than we do, they provide no temporal abstraction mechanisms like our concept of abstract observations.

7 Conclusion

We proposed a framework for model-based diagnosis of dynamic systems, which extends previous work in this field in several ways. The use of an interval-based temporal language considerably improves the expressiveness at the knowledge representation level. In the process description we represent a dynamic system in terms of processes and process states present over time intervals. Each of these process states represents a part of the process evolution and is characterized by a complex pattern of predicted parameters and their values, that are present over arbitrary time intervals related by qualitative temporal relations. We introduced the concept of an abstract observation which allows us to summarize consecutive observations for the same parameter taking the same value. Abstract observations make us independent from the actual number of observations and the temporal resolution, leading an efficient event-driven diagnosis process. Given a process description and abstract observations, we defined an abstract temporal diagnosis as a conjunction of process state assumptions over indefinite time intervals constrained by QA relations. This represents in a natural way indefinite temporal knowledge about the causes of the observations. We applied our framework to the diagnosis of hepatitis B and evaluated it by using real patient data, comparing the abstract temporal diagnoses with the interpretation of these data by the HEPAXPERT-I system. The results are promising and show the importance to consider sequences of findings in the diagnosis of hepatitis B.

Acknowledgments

References

[1] K.-P. Adlassnig and W. Horak, HEPAXPERT-I: Automatic interpretation of tests for hepatitis A and B, *M.D. Computing 8* (1991) 118–119.

[2] K.-P. Adlassnig and W. Horak, Development and retrospective evaluation of HEPAXPERT-I: a routinely-used expert system for interpretive analysis of hepatitis A and B serological findings, *Artificial Intelligence in Medicine 7* (1995) 1–24.

[3] J.F. Allen, Maintaining knowledge about temporal intervals, *Communications of the ACM 26* (1983) 832–843.

[4] J.F. Allen and P.J. Hayes, Moments and points in an interval-based temporal logic, *Computational Intelligence 5* (1989) 225–238.

[5] I. Bratko, I. Mozetič, and N. Lavrač, *KARDIO: A Study in Deep and Qualitative Knowledge for Expert Systems* (MIT Press, 1989).

[6] V. Brusoni, L. Console, P. Terenziani, and D. Theseider Dupré, Characterizing temporal abductive diagnosis, in *Working Papers of the 6th International Workshop on Principles of Diagnosis* (1995) 34–40.

[7] V. Brusoni, L. Console, P. Terenziani, and D. Theseider Dupré, An efficient algorithm for computing temporal abductive diagnoses, in *Working Papers of the 6th International Workshop on Principles of Diagnosis* (1995) 41–48.

[8] B.G. Buchanan and E.H. Shortliffe, editors, *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project* (Addison-Wesley Publishing Company, 1984).

[9] L. Console, D. Theseider Dupré, and P. Torasso, A theory of diagnosis for incomplete causal models, in *Proceedings of the 11th International Joint Conference on Artificial Intelligence* (Morgan Kaufmann Publishers, Inc., 1989) 1311–1317.

[10] L. Console, L. Portinale, D. Theseider Dupré, and P. Torasso, Diagnostic reasoning across different time points, in *Proceedings of the 10th European Conference on Artificial Intelligence* (John Wiley & Sons, 1992) 369–373.

[11] L. Console and P. Torasso, On the co-operation between abductive and temporal reasoning in medical diagnosis, *Artificial Intelligence in Medicine 3* (1991) 291–311.

[12] L. Console and P. Torasso, A spectrum of logical definitions of model-based diagnosis, *Computational Intelligence 7* (1991) 133–141, also in [20].

[13] J. de Kleer and B.C. Williams, Diagnosing multiple faults, *Artificial Intelligence 32* (1987) 97–130.

[14] K.L. Downing, Physiological applications of consistency-based diagnosis, *Artificial Intelligence in Medicine 5* (1993) 9–30.

[15] G. Friedrich and F. Lackinger, Diagnosing temporal misbehavior, in *Proceedings of the 12th International Joint Conference on Artificial Intelligence* (Morgan Kaufmann Publishers, Inc., 1991) 1116–1122.

[16] G. Friedrich and W. Nejdl, MOMO — Model-based diagnosis for everybody, in *Proceedings of the IEEE Conference on Artificial Intelligence Applications* (1990), a slightly revised and extended version appears in W. Hamscher, L. Console, J. de Kleer, eds, Readings in Model-Based Diagnosis, Morgan Kaufmann Publishers, Inc., 1992.

[17] J. Gamper, *A Temporal Reasoning and Abstraction Framework for Model-Based Diagnosis Systems* (”infix”-Verlag, St. Augustin, 1996).

[18] M.L. Ginsberg, The computational value of nonmonotonic reasoning, in *Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning* (Morgan Kaufmann Publishers, Inc., 1991) 262–268.

[19] I.J. Haimowitz and I.S. Kohane, Managing temporal worlds for medical trend diagnosis, *Artificial Intelligence in Medicine 8* (1996) 299–321.

[20] W. Hamscher, L. Console, and J. de Kleer, editors, *Readings in Model-Based Diagnosis* (Morgan Kaufmann Publishers, Inc., 1992).

[21] L. Ironi, M. Stefanelli, and G. Lanzola, Qualitative models in medical diagnosis, *Artificial Intelligence in Medicine 2* (1990) 85–101.

[22] E.T. Keravnou, Temporal diagnostic reasoning based on time-objects, *Artificial Intelligence in Medicine 8* (1996) 235–265.

[23] W. Long, Temporal reasoning for diagnosis in a causal probabilistic knowledge base, *Artificial Intelligence in Medicine 8* (1996) 193–215.

[24] D. Poole, Explanation and prediction: An architecture for default and abductive reasoning, Technical Report 89-4, Department of Computer Science, U.B.C., 1989.

[25] D. Poole, Normality and faults in logic-based diagnosis, in *Proceedings of the 11th International Joint Conference on Artificial Intelligence* (Morgan Kaufmann Publishers, Inc., 1989) 1304–1310.

[26] R. Reiter, A theory of diagnosis from first principles, *Artificial Intelligence 32* (1987) 57–95.

[27] T.A. Russ, Use of data abstraction methods to simplify monitoring, *Artificial Intelligence in Medicine 7* (1995) 497–514.

[28] Y. Shahar and M.A. Musen, Knowledge-based temporal abstraction in clinical domains, *Artificial Intelligence in Medicine 8* (1996) 267–298.

[29] Y. Shahar, S.W. Tu, and M.A. Musen, Knowledge acquisition for temporal-abstraction mechanisms, *Knowledge Acquisition 4* (1992) 217–236.

[30] P. Struss and O. Dressler, Physical negation — Integrating fault models into the general diagnostic engine, in *Proceedings of the 11th International Joint Conference on Artificial Intelligence* (Morgan Kaufmann Publishers, Inc., 1989) 1318–1323.

[31] S. Uckun, Model-based reasoning in biomedicine, *Critical Reviews in Biomedical Engineering 19* (1992) 261–292.

[32] P. van Beek, Reasoning about qualitative temporal information, *Artificial Intelligence 58* (1992) 297–326.

[33] M. Vilain and H. Kautz, Constraint propagation algorithms for temporal reasoning, in *Proceedings of the 5th National Conference on Artificial Intelligence (AAAI-86)* (1986) 377–382, a revised version appeared in [34].

[34] D. Weld and J. de Kleer, editors, *Readings in Qualitative Reasoning about Physical Systems* (Morgan Kaufmann Publishers, Inc., 1989).