12th International Symposium "Computer Science for Environmental Protection"
(UI-98), Bremen, 1998

Diagnosis and Therapy Recognition for Ecosystems - Usage of Model-based Diagnosis Techniques

Ulrich Heller and Peter Struss[1]

Abstract

Environmental decision support involves different tasks like assessing the current situation from observations, diagnosing the causes of undesirable developments and planning therapy actions in order to improve the situation. While these task are knowledge intensive, we find that techniques of consistency-based diagnosis that have been successfully applied for component structures of technical devices can profitably be generalized to cover the mentioned tasks for a wider range of systems. We present a generalized theory of model-based diagnosis handling a variety of compositional modeling techniques (like the Qualitative Process Theory). This also provides formal logical foundations and the possibility to utilize existing computational tools.

1 Introduction

In complex ecosystems it is often a knowledge intensive task to infer the causes of undesirable developments. One has to know possible phenomena and their preconditions. Assessing the situation of an ecosystem from a limited number of observations can be defined as a challenging diagnosis task. Additionally, one wants to reason about possible cures or symptom treatments in order to influence the ecosystem in a direction that is in accordance with some specified goals.

There exist formal foundations, successful methodologies, and powerful tools for the task of diagnosis of technical systems being modeled as consisting of components and connections between them. We present an extension of these techniques in order to cover reasoning tasks aimed at ecosystems, which we prefer to model in a process-oriented compositional paradigm (Heller/Struss 1996a), similar to the Qualitative Process Theory (Forbus 1984). This provides the foundation for

[1] Model-based Reasoning and Qualitative Reasoning Group, Dept. of Computer Science, Technical University of Munich, Orleansstr. 34, 81667 München, Germany
email: {heller, struss}@in.tum.de

identifying additional objects not accounted for in the initial system model and for a broader variety of tasks including therapy recognition.

First, we give a short introduction to consistency-based diagnosis. Then, we will present a motivating example from hydro-ecology (section 3) and show how the presented theory of diagnosis has to be extended (section 4). This includes distinguishing three tasks that are identical from the component-oriented point of view (section 5), a description of how model formation and prediction are performed (section 6), and how to search for model revisions, with a special focus on revising the closed-world assumption, i. e. on how to look for additional entities (section 7). Finally, we will discuss benefits, limitations and open issues.

2 The Consistency-based Approach to Diagnosis

The development and improvement of consistency-based diagnosis techniques is an active research field, which benefits from the strict logical foundations provided by (Reiter 1987) and others (Hamscher et al. 1992). Our short introduction to the standard methodology follows (Dressler/Struss 1996).

Starting from a description of the (correctly working) system, SD (system description), the diagnostic system can make predictions (in terms of values of variables given certain situations). If any of these predictions contradict actual observations, OBS, the system is known to have some fault.

Additionally, all assumptions about correctly working components being used when making prediction are recorded. For each component model, C_i, out of a set of components, COMPS, contributing to the prediction, an assumption of the form $ok(C_i)$ is added. An assumption-based truth maintenance system (ATMS, (de Kleer 1986)) can be employed to compute the minimal set of assumptions necessary to assert a certain value to a variable. If a contradiction occurs, it is also supported by one or more minimal sets of such assumptions. We call these sets conflicts. Formally,

$$SD \cup OBS \cup \{ok(C_i) : C_i \in COMPS\} \vdash \bot$$

Clearly, not all of the components, C_i, included in a conflict can be working correctly. Therefore, a set of diagnostic candidates is generated from the obtained set of conflicts by identifying sets of components covering all these conflicts. The resulting "hitting sets" are candidates for representing the actual system fault. Notice, that while the simplest case is that a single component occurs in all conflicts, the algorithm works equally well for multiple faults.

We can view the retraction of the assumptions encoding that all components are working correctly as a special revision of our initial model of the device,

$$SD \cup \{ok(C_i) : C_i \in COMPS\}$$

The revised model contains new assumptions about the "mode" an individual component is in. This could be a weaker assumption ($\neg ok(C_i)$, enabling no prediction at all) or an assignment of some potential failure mode ($mode_i(C_i)$), for which there exists a "fault model", which is to be consistent with the observations:

$$SD \cup OBS \cup \{mode_i(C_i) : C_i \in COMPS\} \nvdash \bot$$

Using a common example of a circuit containing multipliers (M_i) and adders (A_j) (Fig. 1), this reasoning mechanism can be demonstrated:

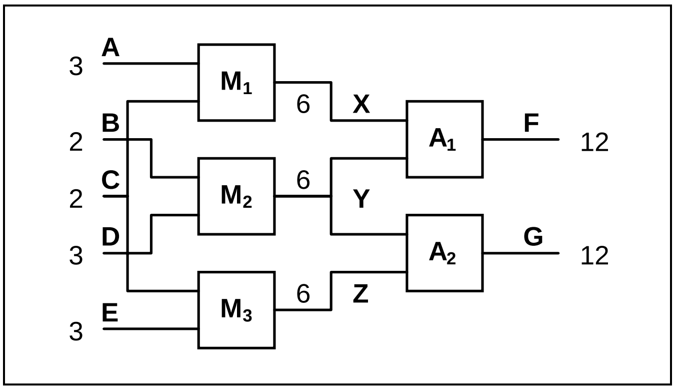

Figure 1: Example circuit (multipliers M_i and adders A_j)

Supplied with the input $A=3$, $B=2$, $C=2$, $D=3$ and $E=3$, constraint propagation can be used to predict outputs $F=12$ and $G=12$. The ATMS computes, that, for instance, $F=12$ depends on

$$\{ok(A_1), ok(M_1), ok(M_2)\}.$$

If we now measure $F=10$ and $G=12$ and enter these values as facts, then

$$\{ok(A_1), ok(M_1), ok(M_2)\}$$

is a conflict, and so is

$$\{ok(A_1), ok(A_2), ok(M_1), ok(M_3)\}.$$

This set of assumptions is involved in computing $G=10$ from $Z=6$ and $Y = F\text{-}X = 10 - 6 = 4$.

Thus, the minimal candidates are (each one explains both conflicts):

$$\{ok(A_1)\}, \{ok(M_1)\}, \{ok(A_2), ok(M_2)\}, \{ok(M_2), ok(M_3)\}.$$

As can be easily seen from the set COMPS and the design of the approach, this is explicitly tailored for diagnosing components. Our research group has successfully developed and applied model-based diagnosis to a number of technical domains.

However, we will see, that the theory can be generalized in order to handle a richer set of models. We are particularly interested in the process oriented ontology, which we apply in ecological reasoning. We will present a motivating example for diagnostic reasoning in the ecological domain in the next section and then discuss the necessary extensions of consistency-based diagnosis techniques in sections 4 and 5.

3 An Example from Hydro-Ecology

In many hydro-ecological systems, the variation of the methane concentration near the surface is monitored. If a significant rise in the methane concentration is measured, a human expert would use his domain knowledge to find a plausible cause of this.

The most probable source of methane is the metabolism of anaerobic bacteria, but they prosper only in anaerobic conditions (with very low oxygen concentrations). Thus, the methane couldn't possibly be generated near the surface, except in the unlikely case of anaerobic conditions throughout the water body. But methane ascends in the water column and can be produced near the ground, where anaerobic bacteria can be hypothesized and the necessary oxygen depletion is quite possible.

This involves a kind of diagnostic reasoning, which makes use of knowledge about the physical and biological phenomena that can occur in a water body. We will give an example of how to do this with a simple model of the system based on the Qualitative Process Theory (QPT, see (Forbus 1984)). In QPT, a model consists of three parts: the taxonomy (specifying the object types of the domain and their possible relations), the scenario (describing the system under consideration in terms of object instances and their relations) and a process library, consisting of generic process descriptions. The generic process descriptions contain instantiation and activation conditions (we use the abbreviations AC and IC) and a specification of what happens, when the process is instantiated (for a matching set of objects) and activated (in dependency of values of quantities associated with objects). For an extensive treatment of QPT, please refer to the original paper (Forbus 1984).

For our example, we assume a stratified water body, i. e. the vertical mixing is largely impeded by a thin separating layer, called metalimnion, with a steep temperature gradient (but ascending gases can cross the boundary). The upper layer (above the metalimnion) is referred to as the epilimnion, while the lower part of the water body is called hypolimnion. Neglecting the extent of the metalimnion, we model this as a scenario with two adjacent compartments, each of them associated with a number of quantities describing the physical and chemical properties or state, including the methane concentration in a given compartment at a given time. There might be a number of processes from the library that can be instantiated for this simple scenario, but nothing will produce or otherwise influence the methane concentration. In particular, we don't expect anaerobic bacteria in either compartment.

In our process library, we include process descriptions of the splitting of organic matter by anaerobic bacteria and ascending of methane. The following diagrams use the formalism of Qualitative Influence Diagrams (QIDs) for specifying the effects of the processes on object variables. For details about the QID notation please refer to (Heller/Struss 1996b).

The process of splitting of organic matter by anaerobic bacteria (see figure 2) can be located in any compartment with anaerobic bacteria (instantiation condition) and anaerobic conditions (activity condition). The (absolute) splitting rate of organic matter is modeled as the product of (linear functions of) the concentration of organic matter in the layer and the biomass of the anaerobic bacteria. The splitting rate (amount of organic matter split per time unit) in turn has a negative influence on the derivative of the organic matter (consumption), but a positive influence on the derivatives of the bacteria biomass (growth) and the by-products methane and organic acids (given as concentrations).

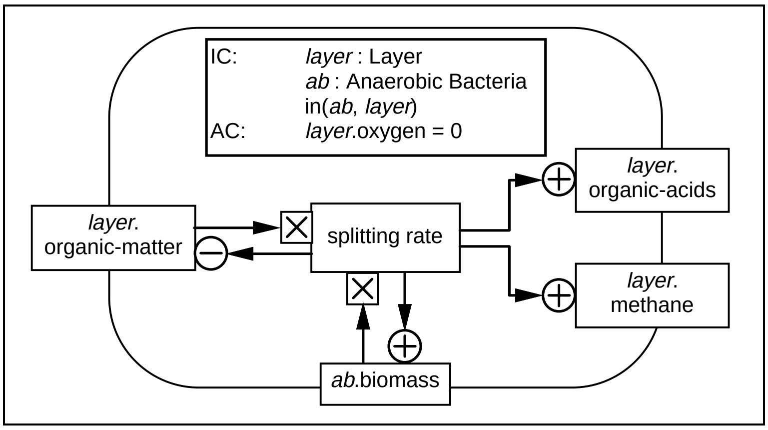

Fig. 2: Process description: Splitting of organic matter by anaerobic bacteria

The second process, namely ascending of methane from one layer to the another one directly above it, is shown in figure 3. The (absolute) rising rate of methane is modeled to be proportional to the amount of methane present in the lower layer.:

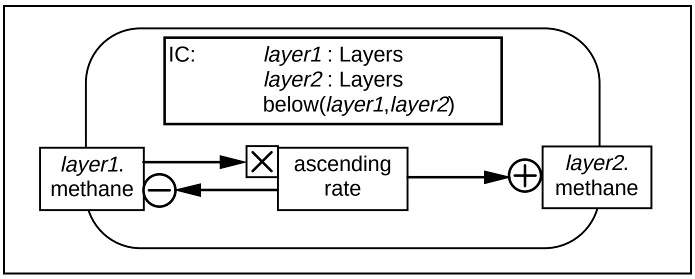

Fig. 3: Process description: Ascending of methane

A third process (not shown in a diagram) is the evaporation of methane from a layer, if that layer is the topmost, i. e. has air contact (IC). It is active, whenever there is methane in the respective layer (AC).

Now, we take the observation of a significant rise in the concentration of methane in the upper layer (d/dt epilimnion.methane > 0). This is a discrepancy to the prediction of a decreasing methane concentration from the fact that there are no other influences on it in our initial model except the evaporation, predicting a decreasing concentration.

The fact that the evaporation of methane is always (instantiated and) active, i. e. not dependent on any retractable assumptions, we are lead to giving up the closed-world assumption and searching the process library for possible instances acting on the methane concentration in the upper layer. We find two possibilities:

- An instance of the splitting process in the upper layer: methane is generated there.
- An instance of the ascending process, i. e. methane is transported into the compartment from the hypolimnion.

For now, we assume that we can exclude horizontal transport (by either knowledge about the system boundary or by contradictions with other observations) and there are no other potential sources of methane in our library.

Pursuing the first possibility, we have to postulate the existence of anaerobic bacteria and (simultaneously) anaerobic conditions in the epilimnion as a candidate. The second possibility leaves us with the same problem as before: somehow methane is generated or transported into the hypolimnion. The latter can be excluded, since there is no third layer below the hypolimnion in our system model. Thus, (an instance of) anaerobic splitting of organic matter in the lower layer is another way to eliminate the discrepancy with observations, and we hypothesize the existence of anaerobic bacteria and anaerobic conditions there (hypolimnion.oxygen = 0). Further measurement of (or specification of additional knowledge about) the oxygen level in the upper compartment will easily discriminate between these two candidates and leave the second one as the only plausible one.

The oxygen concentration is treated as exogenous in this model. With a larger process library we could search for causes for oxygen deprivation in the hypolimnion. This might give clues of how to take action, if we want to change (or prevent) the situation.

3 Using Model-based Diagnosis Techniques

The example discussed in the previous section is certainly a diagnosis problem. And we are able to model the phenomena relevant to the problem. But it refuses a straightforward application of standard model-based diagnosis theories and techniques that work nicely for component-oriented models of devices.

A naive attempt to apply model-based diagnosis might treat the constituents of the model, objects and/or processes, as the system components and associate correct or faulty behavior models to them. However:

- We could not call (the existence of) anaerobic bacteria a fault, and the ascending process is not abnormal in itself, nor does it have a "fault mode". It is simply inappropriate to ask which phenomena or processes involved in the description of the system fail or behave abnormally. They don't. Natural processes do not break or fail like components. This means: diagnosis by assigning modes is inappropriate.
- It is not the case that one of the constituents of our original system description can be blamed for the observed deviation. The cause is some **additional** constituent, namely anaerobic bacteria we were not aware of. This means: diagnosis cannot be confined to a set of given constituents of a system description ("COMPS").
- There are no "failures of nature". Anaerobic conditions in a part of the water body or even severe oxygen depletion causing fish die-off are not faults, even though we would like to avoid it. The model of a lake with oxygen depletion may be perfectly consistent with the observations, while inconsistencies arise only **with our goals** and intentions, with what we consider to be a well-behaving hydro-ecological system from the perspective of health, economy, etc. This means: diagnosis based on inconsistencies between the model and the observations does not work.

Without the concepts of mode assignments, of the set COMPS, and of inconsistencies between the model and observations, only the core of the "theory of diagnosis from first principles" as outlined in a previous section remains: The idea of formalizing and implementing diagnostic problem solving as a search for a model that is consistent with (or entailing) something and, perhaps, that this search can be appropriately organized by revising some initial system description in a minimal way. We have constructed a richer and more powerful theory on this foundation, described in detail in (Struss/Heller 1998). The two main issues are:

- First, we clarify the **nature of the consistency check**, explicitly introduce goals and use this opportunity to distinguish different tasks related to diagnosis.
- In order to really create a diagnosis based on first principles, we have to be specific about **what constitutes the system description**, because this is where the first principles reside. This system description has to be expressive enough to accommodate a more general class of systems, models, and diagnosis problems. In particular, this provides the hooks for potential revisions of the system description. The search for diagnosis candidates is then based on these revisions.

4 Distinguish Different Tasks

Diagnostic reasoning and systems may attempt to answer different questions:

- **What's going on?**
  We start off with some system description, SD_0, which we assume, but which turns out to contradict what we observe. The tasks is then to find a revision of SD_0, SD_1, that agrees with the observations

  $$SD_0 \cup OBS \vdash \perp \qquad \longrightarrow \qquad SD_1 \cup OBS \nvdash \perp ,$$

  or we even require that the new system description entails the observations:

  $$SD_1 \vdash OBS .$$

  If the revision concerns assumptions about the system structure and parameters, the task is *model identification*. *Situation assessment* has the goal of identifying the actual values of variables. Often, we want both.
- **What's going wrong?**
  What is "wrong", is determined by some goal, rather than physics itself. This means inconsistencies and the necessity for diagnosis is caused by criteria that are external to the physical system and its model. When building a diagnosis system in such a case, we cannot expect it to detect and be driven by such inconsistencies unless we make the intentions explicit. The initial system description, possibly SD_1 obtained in the previous step, violates certain goals or intentions we have. The task is to find out what causes this violation, because this may enable us to fix the problem. One way to tackle the problem is to ask how SD_1 has to be revised in order to no longer contradict the goals:

  $$SD_1 \cup GOALS \vdash \perp \;\longrightarrow\; SD_2 \cup GOALS \nvdash \perp.$$

  The "difference" between SD_1 and SD_2, i. e. the revised elements of SD_1 (including removal of added ones) which eliminate the contradiction can be considered to "cause" the trouble. This diagnosis may focus on ultimate or external causes such as values of exogenous variables (in our example, the concentration of oxygen) or simply identify system variables that deviate from a healthy value but can be influenced for treatment.
- **What can be done?**
  Here we try to identify actions that modify the system such that it becomes healthy again. This can be based on the previous step if we make sure that the logical revision of the system description, SD_1, corresponds to some real action. We can aim at a **real cure** if the diagnosis step identified the ultimate causes and there are actions to eliminate them, i. e. shift the system to SD_2. If, in our example, we had searched for reasons for the oxygen depletion, their removal would be the best way to cure the water body. If diagnosis focuses on variables that have abnormal values and can be influenced, a **treatment of symptoms**

could be achieved by introducing actions that add appropriate influences to the respective variables in order to remove the conflict with goals:

$$\mathrm{SD}_1 \cup \mathrm{ACTIONS} \cup \mathrm{GOALS} \not\vdash \perp,$$

or, stronger, to establish them:

$$\mathrm{SD}_1 \cup \mathrm{ACTIONS} \vdash \mathrm{GOALS}.$$

For example, artificial aeration (by using machines for increasing the turbulence of the water body) would be an option. Of course, in reality the goal may not be reached immediately, but only ultimately, so it is a useful subgoal to try to establish the right direction of change (derivative) for each variable (as long as there are no constraints on the trajectory the system has to follow, this gradient driven approach will provide useful results). But we will exclude more sophisticated temporal aspects for now.

Interestingly, these tasks have not been distinguished in the standard theory, for the simple reason that they fall into one, if we postulate that the correctly working device fulfills our goals, i. e.

$$\{\mathrm{ok}(C_i) : C_i \in \mathrm{COMPS}\} \Rightarrow \mathrm{GOALS}$$

and the set of possible actions consists of the replacement of each single component.

This formal description of the tasks reveals their common core: searching for a revision of the model. However, we are not able to devise effective search algorithms and to really perform diagnosis from first principles unless we become more specific about what is contained in SD. This is what we attempt in the following section. Although this basically reconstructs qualitative process theory (or a generalization of it), also "classical" component-oriented modeling can be regained as a (very) special instance of the formalism.

5 The System Description and Prediction

We follow the methodology of compositional modeling, or, more precisely, structure-to-behavior reasoning (a classical example of this is the Qualitative Process Theory (see above)). The basic idea is to compose the model of a system from a set of generic model fragments (we call behavior constituents, processes), which capture some clearly separable piece of domain knowledge in a context independent way. This forms the domain theory.

For each behavior constituent, preconditions for its validity are stated in terms of structural and quantity conditions. In turn, a valid model fragment establishes constraints of the quantities of the system (quantity effects) and might also create new objects in certain relations (structural effects).

The knowledge about the system at hand is expressed in terms of objects and relations between objects (structure description). Knowledge about the state of the system at any given time is described via assignment of values to quantities associated with the present objects.

Therefore, our system description, SD, consists of three parts:

- **The domain theory**
  This includes the ontology (object types and object relation types), i. e. a vocabulary for the system structure description, basic laws of the domain and the behavior models, i. e. a set of behavior constituents, associated instantiation and activity rules, and quantity associations.
- **The system structure description**
  In terms of the defined object types and object relation types.
- **Quantity specification**
  We distinguish between parameter specifications and the state description.

For a more detailed discussion, please refer to (Struss/Heller 1998).

Now, in order to predict system behavior from SD, we compose and use the system model. In all of these steps, we collect the set of assumptions that provide clues of how to revise SD in case of conflicts. In the first place, we consider SD as consisting of a fixed and a revisable partition:

$$SD = SD_{fix} \cup SD_{rev}$$

We proceed in the following steps depicted in figure 4:

- Model composition
  From the structure description, we infer all the instances of the generic behavior constituents that can possibly be active. This is an important distinction we make for reasons of tractability (avoiding non-monotonicity): processes are **instantiated** if their structural conditions are satisfied and **active**, if their quantity (state-dependent) conditions are met. Instantiation of behavior constituents also creates new structural elements (structural effects). In this cycle and the resulting behavior constituents, we collect all assumptions from $STRUCT_{rev}$, the revisable part of the structure description.
- Constraint Generation
  From the set of behavior constituents, we infer constraints and influences (partially specified constraints) between variables. The latter will be turned into constraints ("influence resolution") under the assumption, that we have knowledge about all relevant entities in the system. This is commonly called the closed-world assumption (CWA).
  In order to focus search for revisions from conflicts, we localize the closed-world assumption, i. e. we create an individual assumption CWA_v for each variable v, stating that all influences on this particular variable are known. The set of these assumptions is called CWA_{infl}.
- Prediction
  Finally, from the specified quantities (part of which are revisable: $QUANT_{rev}$) and the generated constraints, new variable values can be determined. This is the prediction phase, which still propagates the sets of assumptions supporting each variable value.

For a more detailed description of how instantiation and activation are separated and the techniques of influence resolution for potentially active behavior constituents is carried out, refer to (Struss/Heller 1998).

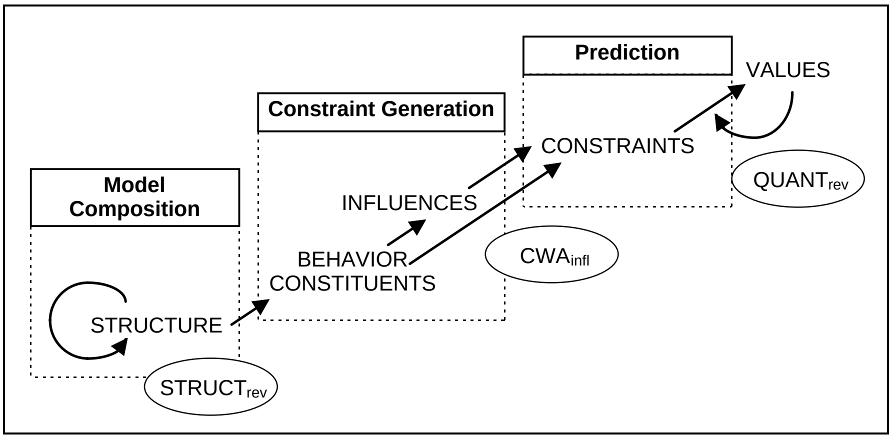

Fig. 4: Model formation and prediction

6 Searching for Revisions

If a conflict is detected, there are four kinds of assumptions that can be contained in the supporting set. Value assignments (from QUANT_{rev}), object existence and object relations (part of STRUCT_{rev}) are simply traced back to SD_{rev} and can be revised there (which we do for minimal candidates covering all of the conflicts found).

But to account for additional objects (or actions, in case of therapy recognition), the closed-world assumption for a particular variable can be a candidate, in which case a search for potential additional influences is carried out. Fortunately, the search space is restricted to the domain theory, i. e. we assume that everything of interest that can possibly happen is described there.

Compare figure 5 for the search tree. In our example from chapter 3, we look for additional influences on the concentration of methane. The domain theory provides only a limited number of behavior constituents for that particular quantity type associated with that object type, so we find ascending and splitting. The preconditions of one of these processes have to become true for the process to come into effect, so we look for the objects and relations missing in the initial system description (e. g. the anaerobic bacteria) . Of course, some minimality principle is to be employed, but there is a difficulty here: a non-minimal set of revisions might have been created by a process to be instantiated by a single object, so minimality heuristics are not easily applicable.

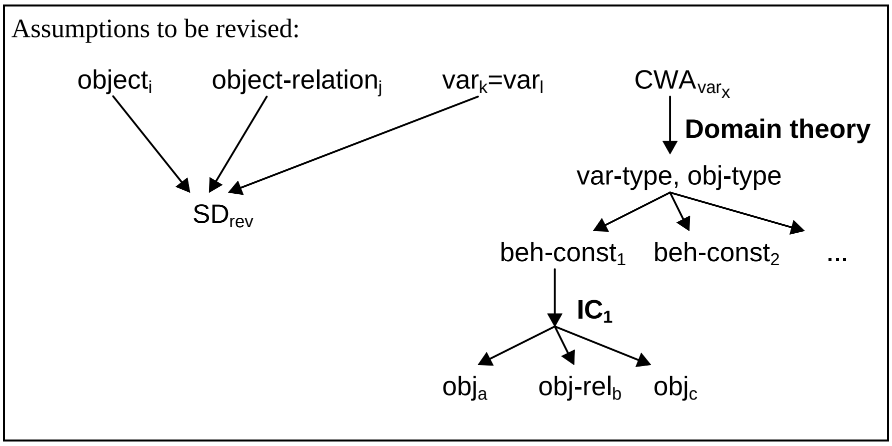

Fig. 5: How to search for revisions

7 Discussion

The presented approach provides a formal foundation for the tasks of situation assessment, diagnosis and some basic mechanisms of therapy recognition. Moreover, well-defined techniques and a number of powerful existing tools from the technical diagnosis community can be used.

Of course, there is an inherently high complexity in this extended approach, which is largely reduced in the component-oriented approach because of rigid structural conditions on the instantiation of behavior constituents (in this case behavior models for specific modes). Nevertheless, we expect a carry-over of results in focusing and controlling the diagnostic process from research in model-based diagnosis, e. g. optimized design of truth-maintenance systems.

A limitation of the approach is neglecting dynamic aspects. So far, we consider the system in a single time-slice, both for situation assessment and for therapy recognition. In particular this does not account for the inertia of the system, which might have to be "moved" to a desired state in accordance with the specified goals.

Development of search heuristics, e. g. in terms of characterization of what is to be considered an "ultimate cause" is needed.

Acknowledgments

We like to thank Robert Illner, who helped in modeling hydro-ecological systems during his stay at our Brazilian partners. And the municipal department of water and sewage in Porto Alegre, Brazil, for supporting our work with their knowledge about the domain.

References

de Kleer, J. (1986): An assumption-based TMS, in: Artificial Intelligence 28(2).

Dressler, O., Struss, P. (1996): The Consistency-based Approach to Automated Diagnosis of Devices. In: Brewka, G. (ed.), Principles of Knowledge Representation, CSLI Publications, Stanford.

Forbus, K. (1984): Qualitative Process Theory, in: Artificial Intelligence 24(1-3).

Hamscher, W., Console, L., de Kleer, J. (eds.) (1992): Readings in Model-based Diagnosis, Morgan Kaufmann Publishers, San Mateo.

Heller, U., Struss, P. (1996a): Qualitative Modeling for Environmental Decision Support, in: 10th International Symposium "Computer Science for Environmental Protection" (UI-96), Hannover (available from Metropolis Verlag, Marburg).

Heller, U., Struss, P. (1996b): Transformation of Qualitative Dynamic Models - Application in Hydro-Ecology, in: 10th International Workshop on Qualitative Reasoning (QR-96), Stanford (available from AAAI Press).

Reiter, R. (1987): A Theory of Diagnosis from First Principles, in: Artificial Intelligence 32 (1987) 57-95.

Struss, P., Heller, U. (1998): Process-oriented Modeling and Diagnosis - Revising and Extending the Theory of Diagnosis from First Principles, in: Working Notes of the 9th International Workshop on Principles of Diagnosis (DX-98), Sea Crest, Cape Cod.