Electronic Transactions on Artificial Intelligence
Vol. 1 (1997): *73-104*
http://www.ep.liu.se/ej/etai/1997/004/

Refereed Article

Michael Thielscher:

A Theory of Dynamic Diagnosis

Abstract: *Diagnosis is, in general, more than a mere passive reasoning task. It often requires to actively produce observations by performing a test series on a faulty system. We present a theory of diagnosis which captures this dynamic aspect by appealing to Action Theory. The reactions of a system under healthy condition are modeled as indirect effects, so-called* ramifications, *of actions performed by the diagnostician. Under abnormal circumstances - i.e., if certain aspects or components of the system are faulty-one or more of these ramifications fail to materialize. Ramifications admitting exceptions is shown to giving rise to a hitherto unnoticed challenge - a challenge much like the one raised by the famous Yale Shooting counter-example in the context of the Frame Problem. Meeting this challenge is inevitable when searching for "good" diagnoses. As a solution, we adapt from a recent causality-based solution to the Qualification Problem the key principle of* initial minimization. *In this way, when suggesting a diagnosis our theory of dynamic diagnosis exploits causal information, in addition to possibly available, qualitative knowledge of the* a priori *likelihood of components to fail.*

Remark: Some of the results in this paper have been preliminarily reported in (Thielscher, 1997a).

Publication and review history

1. First version *published* by Linköping University Electronic Press on 7.10.1997 and permanently available at
   http://www.ep.liu.se/ea/cis/1997/011/
2. *Received* by the research area "Reasoning about Actions and Change" of the Electronic Transactions on Artificial Intelligence, posted on its web site, and advertised in its Newsletter on 7.10.1997; the summary on 21.10.1997.
3. Publicly *discussed* in the ETAI area of Reasoning about Actions and Change during November, 1997 - January, 1998. The discussion protocol is available and will be retained in the Article Interaction Page at
   http://www.ida.liu.se/ext/etai/received/rac/002/aip.html
4. Revised version, with minor corrections compared to the first version, *published* by Linköping University Electronic Press on 2.3.1998 and permanently available at
   http://www.ep.liu.se/ea/cis/1997/011/
   under the label "Revised publication 1998-03-02".
5. Revised version *accepted* after due refereeing and according to scientific journal standards by the Electronic Transactions on Artificial Intelligence (ETAI) on 3.5.1998. Comments by the referees are included in the Article Interaction Page.
6. Revised version appears in Electronic Transactions on Artificial Intellligence, Volume 1 (1997), Issue 4, pages 73–104. The issue and the annual volume are made permanently available at
   http://www.ep.liu.se/ej/etai/1997/
7. Paper editions of the ETAI issue and the ETAI volume containing this artice *republished* by the Royal Swedish Academy of Sciences. For purchase of this edition, please refer to the URL mentioned in the previous item.

The review policy and the quality requirements for acceptance are documented at
http://www.ida.liu.se/ext/etai/info/

Article maintenance

The Electronic Transactions on Artificial Intelligence maintains an Article Interaction Page for this article at
http://www.ida.liu.se/ext/etai/received/rac/002/aip.html
Besides the publication and review history, it contains links for contacting the author(s), as well as to amendments and other related information for the article. It is intended to keep these links up-to-date in the future.

Summary

ETAI authors are recommended that each article be accompanied by a summary. Longer and more specific than a conventional abstract, it should specify in concrete terms what are the new results in the article. If present, the summary also plays a role in the refereeing process: referees are asked to judge whether the results as specified in the summary *are of importance to the field, and whether the body of the article* gives sufficient evidence *for the claims made in the summary. – The Editor.*

Diagnosis in general requires more than just passively observing the behavior of a faulty system. Often it is necessary to actively produce observations by performing actions. Diagnosing then amounts to reasoning about more than a single state of the system to be examined. We propose to capture this dynamic aspect by appealing to Action Theory. A formal system description consists of a *static* and a *dynamic* part. The former introduces the system components and their static relations in form of so-called state constraints, like, for instance,

$$\texttt{active}(\texttt{relay}_1) \equiv \texttt{closed}(\texttt{switch}_1)$$

stating that a particular relay is active if and only if a corresponding switch is closed. The dynamic part of a system description specifies the actions which can be used to manipulate the system's state. These definitions are accompanied by so-called action laws, which focus on the direct effects. State constraints like the above then give rise to additional, indirect effects of actions, which we accommodate according to the theory of causal relationships [Thielscher, 1997b]. E.g., this causal relationship is a consequence of our example state constraint:

$$\texttt{closed}(\texttt{switch}_1)\ \underline{\text{causes}}\ \texttt{active}(\texttt{relay}_1)$$

Informally speaking, it means that whenever $\texttt{closed}(\texttt{switch}_1)$ occurs as direct or indirect effect of an action, then this has the additional, indirect effect that $\texttt{active}(\texttt{relay}_1)$. Generally, causal relationships are successively applied subsequent to the generation of the direct effects of an action until a satisfactory successor state obtains.

In this way, the reactions of a system under healthy condition are modeled as indirect effects, so-called *ramifications*, of actions. Under abnormal circumstances—i.e., if certain aspects or components of the system are faulty—one or more of these ramifications fail to materialize. We introduce an abnormality fluent `ab` by which we account for such exceptions to both state constraints and the ramifications they trigger. Thus our example constraint from above, for instance, may read weaker—e.g., to the effect that

$$\neg\texttt{ab}(\texttt{resistor}_1) \land \neg\texttt{ab}(\texttt{relay}_1) \supset [\texttt{active}(\texttt{relay}_1) \equiv \texttt{closed}(\texttt{switch}_1)]$$

where $\texttt{ab}(\texttt{resistor}_1)$ and $\texttt{ab}(\texttt{relay}_1)$ represent an abnormal failure of a corresponding resistor and the relay itself, respectively. This weakening transfers to our expectations regarding indirect effects: The aforementioned causal relationship becomes

$$\texttt{closed}(\texttt{switch}_1)\ \underline{\text{causes}}\ \texttt{active}(\texttt{relay}_1)\ \underline{\text{if}}\ \neg\texttt{ab}(\texttt{resistor}_1) \land \neg\texttt{ab}(\texttt{relay}_1)$$

An important contribution of this paper, now, is a proof that due to the phenomenon of causality straightforward global minimization of abnormality—which is suitable for static diagnosis—is problematic in case of dynamic diagnosis. This raises a challenge much like the one raised by the famous Yale Shooting counter-example in the context of the Frame Problem. Meeting this challenge is inevitable when searching for 'good' diagnoses.

As a solution, we adapt from a recent causality-based solution to the Qualification Problem the key principle of *initial minimization*. All instances of the abnormality fluent are assumed false initially but may be indirectly affected by the execution of actions. In this way, our theory of dynamic diagnosis suitably exploits causal information when generating diagnoses. Our theory moreover respects available knowledge of the *a priori* likelihood of component failures. Since it is often difficult if not impossible to provide precise numerical knowledge of probabilities, we deal with qualitative rather than quantitative information, and we do not rely on complete knowledge. Such possibly incomplete information as to different degrees of abnormality is formally represented by a partial ordering among the instances of the abnormality fluent.

For the entire theory there exists a provably correct axiomatization based on the Fluent Calculus paradigm and which uses Default Logic to accommodate the nonmonotonic aspect of the diagnostic problem.