Making Hybrid Plans More Clear to Human Users

A Formal Approach for Generating Sound Explanations

Bastian Seegebarth, Felix Müller, Bernd Schattenberg, and Susanne Biundo

Institute of Artificial Intelligence

June 28th, 2012
Motivation – Transparent Decision Support

Applications of planning technology:
- Emergency planning
- Assistance of cognitively impaired and elderly people
- Support in daily activities

Planning systems provide recommendations when and how to act to subjects who are competent themselves
Humans might scrutinize the systems’ suggestions

To prevent this a planning system must be able to:
- Give reasons for decisions
- Present them in a comprehensible manner
- Build trust in the system’s competence
Our Contribution

- Provide a formal framework for the generation of (raw) plan explanations
- Raw Explanations can be used as input to a text generation system to produce the actual explanation
- Implement the explanation for the existence of actions and orderings on actions in a plan in the framework, i.e., the system can answer the questions:
  - Why does the user have to execute a given action from the plan?
  - Why does the user have to execute two actions from the plan in the order given in the plan?
- Implement a prototype explanation system
Hybrid Planning – Problem Formalization

Hybrid planning framework combines classical and hierarchical planning

\[ \pi = \langle T, M, P_{init} \rangle \] is a hybrid planning problem with

- \( T \) is a set of task schemata of the form \( \langle t(\bar{v}), pre, eff \rangle \), \( \bar{v} \) is the parameter list of \( t \)
- \( M \) is a set of decomposition methods of the form \( \langle t(\bar{v}), P \rangle \), \( P \) is a partial plan
- \( P_{init} \) is the initial partial plan that needs to be decomposed
  - plan step \( init \) has the initial state as effects
  - plan step \( goal \) has the goals as preconditions
Hybrid Planning – Partial Plans and Solutions

\[ P = \langle PS, \prec, V, C \rangle \] is a partial plan with

- \( PS \) is a set of plan steps of the form \( s:t(\bar{v}) \), \( s \) is a unique label and \( t(\bar{v}) \) is a (partially) instantiated task
- \( \prec \) is a partial order on \( PS \)
- \( V \) is a set of constraints on the variables appearing in \( PS \)
- \( C \) is a set of causal links of the form \( \langle s \rightarrow_p s' \rangle \)

A partial plan is a solution to a planning problem if

- Every precondition is established by a causal link
- No causal links are threatened
- \( P \) contains only primitive tasks and can be obtained from \( P_{init} \) by decomposition of abstract tasks and the insertion of plan steps, causal links, ordering, and variable constraints
Running Example

- **enterAlbum**
  - InAlbumMode

- **selectPic**
  - PicSelected

- **pressSendByEMail**
  - PicAttached

- **chooseRecipient**
  - RecipientSet

- **pressSend**
  - EMailSent

- **displayPic**
  - dPViaAlbum

- **sendPic**
  - sendPicByEMail

- **top**
  - m_{top}

- **goal**
Basic Framework

- Formalize information about the plan, its construction process, and basic arguments as a first-order logic axiomatic system
- Construct an explanation by finding a proof for a formula that represents the requested aspect
- Elements of axiomatic system are based on the problem specification, the construction process that led to the plan, and the underlying planning formalism
- Explanations are provably correct w.r.t. the underlying planning system
Construct axiomatic system $\Sigma$ from plan $P = (PS, \prec, V, C)$.

Causal Structure:
- Add $CR(s, p, s')$ to $\Sigma$ for every causal link $\langle s \rightarrow_p s' \rangle \in C$

Decomposition Structure:
- Add $DR(s, m, s')$ to $\Sigma$ if $s$ was introduced by the decomposition of $s'$ via method $m$
Examples:

- CR(enterAlbum, InAlbumMode, selectPic)
- CR(selectPic, PicSelected, pressSendByEMail)
- DR(pressSendByEMail, sendPicByEMail, sendPic)
- DR(sendPic, mTop, top)
Basic Explanations

\(\text{Nec}(s)\) denotes that \(s\) is necessary for the plan to be a solution (this does not mean that there cannot be a plan without \(s\))

A plan step is necessary if it establishes a goal:
- \(\forall s.\left(\exists g. \text{CR}(s, g, \text{goal})\right) \Rightarrow \text{Nec}(s)\)

...or if it establishes a precondition of a plan step that is necessary:
- \(\forall s.\left(\exists s', p. \text{CR}(s, p, s') \land \text{Nec}(s')\right) \Rightarrow \text{Nec}(s)\)

A plan step is necessary if it is a step of the initial partial plan:
- \(\forall s. \left[\text{DR}(s, m_{\text{top}}, \text{top}) \Rightarrow \text{Nec}(s)\right]\)

...or if it is a sub step of a step from the initial partial plan:
- \(\forall s.\left[\exists s', m. \text{DR}(s, m, s') \land \text{Nec}(s')\right] \Rightarrow \text{Nec}(s)\)
A First Explanation

- \( \forall s. [\exists g. CR(s, g, \text{goal})] \Rightarrow \text{Nec}(s) ] \)
- \( \forall s. [\exists s', p. [CR(s, p, s') \land \text{Nec}(s')]] \Rightarrow \text{Nec}(s) ] \)
- \( \forall s. \left[ DR(s, m_{\text{top}}, \text{top}) \Rightarrow \text{Nec}(s) \right] \)
- \( \forall s. [\exists s', m. [DR(s, m, s') \land \text{Nec}(s')]] \Rightarrow \text{Nec}(s) ] \)

To explain why executing \textit{enterAlbum} is necessary:

1. \text{Nec}(\text{enterAlbum})
2. \text{CR}(\text{enterAlbum, \text{InAlbumMode, selectPic}})
3. \text{Nec}(\text{selectPic})
4. \text{CR}(\text{selectPic, PicSelected, pressSendByEMail})
5. \text{Nec}(\text{pressSendByEMail})
6. \text{DR}(\text{pressSendByEMail, sendPicByEMail, sendPic})
7. \text{Nec}(\text{sendPic})
8. \text{DR}(\text{sendPic, } m_{\text{top}}, \text{top})
Translation to Natural Language

1. \textit{Nec}(\textit{enterAlbum})
2. \textit{CR}(\textit{enterAlbum}, \textit{InAlbumMode}, \textit{selectPic})
3. \textit{Nec}(\textit{selectPic})
4. \textit{CR}(\textit{selectPic}, \textit{PicSelected}, \textit{pressSendByEMail})
5. \textit{Nec}(\textit{pressSendByEMail})
6. \textit{DR}(\textit{pressSendByEMail}, \textit{sendPicByEMail}, \textit{sendPic})
7. \textit{Nec}(\textit{sendPic})
8. \textit{DR}(\textit{sendPic}, m_{\text{top}}, \textit{top})

In natural language (Future Work):
"Entering the album is necessary to select the picture. You must select the picture in order to use the \textit{send by EMail}...-function. That is a necessary sub step of sending the picture which is part of your initial problem specification."
Levels of Abstraction for Explanations

- Explaining only on primitive level leads to overly long and detailed explanations
- Level of detail should be variable for different parts of an explanation
  - Explain on primitive levels for parts of the plan that the user is not familiar with
  - Skip over other parts by explaining on high level of abstraction
- Through decomposition relations the explanation can be moved to higher level of abstraction

Problem: the Causal Structure of a plan is usually given only in terms of primitive plan steps

Therefore, abstract plan steps can only be explained via decomposition relations
Causal Relations for Abstract Plan Steps

What is the causality produced and consumed by an abstract plan step?

Let plan steps inherit causal relations from their sub steps:

- $\forall s, p, s'.\exists m, s''.[DR(s'', m, s) \land CR(s'', p, s')]) \Rightarrow CR(s, p, s')$
- $\forall s, p, s'.\exists m, s''.[DR(s'', m, s') \land CR(s, p, s'')]) \Rightarrow CR(s, p, s')$

Example:

```
selectPic  pressSendByEMail
    PicSelected
          displayPic
                   sendPic
```
Causal Relations over Abstract State Features

Sometimes inherited causal relations do not seem reasonable:

- \( CR(\text{sendPic}, \text{EMailSent}, \text{goal}) \)
- \( CR(\text{sendPic}, \text{PicAttached}, \text{goal}) \)

Introduce set of decomposition axioms:

- \( \text{PicTransferred} \Leftrightarrow [\text{PicAttached} \land \text{EMailSent}] \lor [\text{PicPrinted} \land \text{FaxSent}] \)

Derive causal relations over abstract state features:

- \( \forall s, s', p, p'.[[CR(s, p, s') \land CR(s, p', s')] \Rightarrow CR(s, p \land p', s')] \)
- \( \forall s, s', a.[[\exists da, d. [\text{AbsL}(a, da) \land \text{FDec}(d, da) \land CR(s, d, s')]]) \Rightarrow CR(s, a, s')] \)

Example:

- From the above causal relations and axioms we can derive: \( CR(\text{sendPic}, \text{PicTransferred}, \text{goal}) \)
We have implemented a prototype system to generate explanations as specified by the formal framework.

Thousands of explanations can be found within a few seconds.

How to select among the abundance of possible explanations?

- Size of explanation
- Type of arguments
- Means for presentation (text, graphics, speech)
- Existing user knowledge
Communication of plans is crucial for the acceptance of planning technology

We presented a general framework for the generation of explanations

Instantiation of the framework for the explanation of plan steps and the ordering of plans

Future work has to deal with the presentation and selection of raw explanations