

ARGUING ABOUT PLANNING ALTERNATIVES

ALEX QUILICI

Department of Electrical Engineering
2540 Dole Street, Holmes Hall 455
University of Hawaii at Manoa
Honolulu, HI, 96822

Abstract

In discourse processing, two major problems are understanding the underlying connections between successive dialog utterances and deciding on the content of a coherent dialog response. This paper presents a computational model of these tasks for a restricted class of argumentative dialogs. In these dialogs, each response presents a belief that justifies or contradicts another belief presented or inferred earlier in the dialog. Understanding a response involves relating a stated belief to these earlier beliefs, and producing a response involves selecting a belief to justify and deciding upon the set of beliefs to provide as its justification. Our approach is knowledge based, using general, common-sense justification rules to recognize how a belief is being justified and to form new justifications for beliefs. This approach provides the ability to recognize and respond to never before seen belief justifications, a necessary capability for any system that participates in dialogs involving disagreements.

1 Introduction

In discourse processing, two major problems are understanding the underlying connections between successive dialog responses and deciding on the content of a coherent dialog response. This paper presents an initial model that accomplishes these tasks for one class of *argumentative* dialogs. In this class, each dialog response presents a belief that justifies or contradicts a belief provided earlier in the dialog.

The following dialog fragment is an example:

- (1) TIDY: The members of the AI lab should clean it themselves.
- (2) SCRUFFY: But that interferes with doing research.
- (3) TIDY: There's no other way to keep it clean.
- (4) SCRUFFY: We can pay a janitor to keep it clean.
- (5) TIDY: We need money to pay a janitor.
- (6) SCRUFFY: We can transfer the money from the salary fund.
- (7) TIDY: But doing that interferes with paying the lab members.

- (8) SCRUFFY: It's more desirable to have a clean lab than to pay the lab members.

Each response states one or more plan-oriented beliefs, usually as part of a short chain of reasoning justifying or contradicting a belief provided earlier in the dialog.

In (1), TIDY begins by stating a belief: the lab members should execute the plan of cleaning the lab.

In (2), SCRUFFY responds with a belief that the lab members executing this plan interferes with their doing research. This belief justifies SCRUFFY's unstated belief that the lab members should not execute the plan of cleaning the lab, which contradicts TIDY's stated belief in (1). SCRUFFY's underlying reasoning is that the lab members shouldn't clean the lab because it interferes with their executing the more desirable plan of doing research.

In (3), TIDY presents a belief that there's no alternative plan for keeping the lab clean. This belief justifies TIDY's belief in (1). TIDY's underlying reasoning is that the lab members should clean the lab because it's the best plan for the goal of keeping the lab clean, and it's the best plan because it's the only plan that achieves the goal.

Finally, in (4), SCRUFFY states a belief that paying a janitor achieves the goal of keeping the lab clean. This contradicts TIDY's stated belief in (3). It also justifies a belief that the lab members cleaning the lab isn't the best plan for keeping the lab clean, which contradicts one of the beliefs inferred from (3). SCRUFFY's reasoning is that paying a janitor is a more desirable plan that achieves this goal.

The remaining responses follow the same pattern. Understanding responses like these involves relating a stated belief to beliefs appearing earlier in the dialog. That requires inferring the participant's underlying reasoning chain and the beliefs it justifies. Producing these responses involves selecting a belief to justify and deciding upon the set of beliefs to provide as its justification. That requires constructing an appropriate reasoning chain that justifies holding any unshared beliefs.

Our focus in this paper is on an initial method for representing, recognizing, and producing the belief justifications underlying dialog responses that provide coherent defenses of why beliefs are held.

The behavior modeled is limited in several significant ways. First, we do not try to recognize when an arguer's response contradicts one of his earlier responses, such as the contradiction between (2) and (8), nor do we try to avoid producing such responses. Second, we do not try to recognize or make use of high-level arguing strategies, such as *reductio ad absurdum*. Third, we restrict ourselves to a small class of beliefs involving planning. Finally, we start with representations of beliefs and ignore the linguistic issues involved in turning responses into beliefs. Clearly, all these limitations must eventually be addressed in order to produce a more realistic model of debate. Our belief, however, is that an initial model of the process of recognizing and producing belief justifications is a useful and necessary first step.

2 Our Approach

Our approach to these tasks rests on a simple assumption: *Dialog participants justify beliefs with instantiations of general, common-sense justification rules.* For plan-oriented beliefs, a justification rule corresponds to a planning heuristic that's based solely on structural features of plans in general, not on characteristics of specific plans themselves.

The first few responses in this dialog illustrate several justification rules. In (2), SCRUFFY uses the rule: *One reason why a plan shouldn't be executed is that it conflicts with executing a more desirable plan.* Similarly, in (3), TIDY chains together a pair of these rules: *One reason why a plan should be executed is that it's the best plan for achieving a goal, and One reason why a plan is the best plan for a goal is that it's the only plan that achieves the goal.*

Given our assumption, understanding a response is equivalent to recognizing which justification rules were chained together and instantiated to form it, determining which belief to address in a response is equivalent to determining which beliefs in a chain of instantiated justification rules are not shared, and producing a justification is equivalent to selecting and instantiating justification rules with beliefs from memory.

We make this assumption for two reasons. First, *dialog participants should be able to understand and respond to never before seen belief justifications.* That suggests applying general knowledge, such as our justification rules, to analyze and produce specific justifications, as that knowledge is likely to be shared by different participants, even if they hold different beliefs about specific courses of action. And second, *dialog participants should also be able to use the same knowledge for different tasks.* That suggests that arguments about planning should use the same knowledge as planning itself. The justifica-

tion rules for plan-oriented beliefs describe knowledge that a planner would also find useful in selecting or constructing new plans.

Our approach differs in two ways from previous models of participating in dialogs. First, these models emphasized *plan recognition*: the task of recognizing and inferring the underlying plans and goals of a dialog participant [4, 10, 17, 18, 2]. They view utterances as providing steps in plans (typically by describing goals or actions) and tie them together by inferring an underlying plan. But in an argument not only must the participant's plans and goals be inferred, but also their underlying *beliefs* about those plans and goals. Our approach suggests a model that infers these beliefs as a natural consequence of trying to understand connections between successive dialog utterances. In contrast, existing approaches to inferring participant beliefs take a stated belief and try to reason about possible justifications for it [12, 9]. Previous models have also tended to view providing a dialog response solely as a part of the question answering process. In contrast, our approach suggests that responses arise as a natural consequence of trying to integrate newly-encountered beliefs with current beliefs in memory, and trying to understand any contradictions that result.

3 Justification Rules

The argumentative dialogs we've examined have two types of plan-oriented beliefs: *factual* and *evaluative* [1]. Factual beliefs are objective judgements about planning relationships, such as whether a plan has a particular effect or enablement. They represent the planning knowledge held by most previous plan-understanding and plan-constructing systems. Evaluative beliefs, on the other hand, are subjective judgements about planning relationships, such as whether or not a plan should be executed. Although these beliefs have generally been ignored by previous systems, they are crucial to participating in arguments involving plan-oriented beliefs.

Our assumption is there exists a small set of justification rules for each planning relationship. Each rule is represented as an abstract configuration of planning relationships that, when instantiated, provides a reason for holding a particular belief. For example, the rule that a plan shouldn't be executed if it conflicts with a preferred plan is represented as:

```
IF   interferes(occur(P), occur(P')) AND
      favors(occur(P'), occur(P))
THEN ought(not(occur(P)))
```

That is, a plan shouldn't be executed if (1) it interferes with another plan, and (2) that plan is preferred to it. Figure 1 lists our current justification rules for

Reasons why executing plan X is desirable:

X is the best plan for a goal.

Executing X is an enablement for a goal.

Reasons why executing plan X is undesirable:

X conflicts with a more desirable plan.

X has an undesirable effect.

X has an undesirable enablement.

Reasons why plan X is the best plan for a goal:

X is the only plan that achieves the goal.

No plan more desirable than X achieves the goal.

Reasons why plan X is not the best plan for a goal:

X has an unachievable enablement.

X's execution is undesirable.

Some more desirable plan achieves the goal.

Reasons why plan X is more desirable than plan Y:

X has a desirable effect that Y doesn't have.

X doesn't have an undesirable effect that Y has.

X doesn't have an undesirable enablement that Y has.

Y conflicts with a more desirable plan and X doesn't.

X is an enablement of a more desirable plan than Y.

X has an effect more desirable than Y.

Reasons why achieving goal G is undesirable:

The only plan for achieving G is undesirable.

Achieving G has an undesirable effect S.

Reasons why achieving goal G is desirable:

Achieving G is an enablement for another goal.

Not achieving G has an undesirable effect S.

Figure 1: Justification rules.

evaluative beliefs (see [13] for representational details and criteria for deciding what is a reasonable justification rule). These rules were abstracted from examining a variety of different plan-oriented argumentative dialogs.

The power of these justification rules comes from their generality: A single rule can be instantiated in different ways to provide justifications for different beliefs. In (2), SCRUFFY uses the above rule to justify a belief that the lab members shouldn't clean the lab themselves. In (7), TIDY uses the same rule to justify a belief that the lab members shouldn't transfer money from the salary fund. Here, TIDY's justification is that transferring the money interferes with the more desirable plan of paying researchers.

4 Recognizing Justifications

The process of understanding a dialog response is modeled as a forward-chaining search for a chain of instantiated justification rules that (1) contains the user's stated belief, and (2) justifies an earlier dialog belief or its negation.

We briefly illustrate this process by showing how SCRUFFY understands TIDY's response in (3). The

input belief is that the lab members cleaning the lab is the only plan that achieves the goal of keeping the lab clean. This belief matches an antecedent in a pair of justification rules, so the process begins by instantiating these rules, resulting in pair of possible justification chains that contain TIDY's stated belief: (1) *the lab members cleaning is the best plan for keeping the lab clean because it's the only plan for keeping the lab clean*, and (2) *the lab shouldn't be kept clean because the only plan for that goal is the undesirable plan of having the lab members cleaning it*.

Neither justification directly relates to the dialog, so the next step is to determine which one to pursue further, and whether either can be eliminated from further consideration. Here, the second justification contains a belief that the lab members cleaning the lab is undesirable, which contradicts TIDY's stated belief in (1). Applying the heuristic "*Discard any potential justification containing beliefs that contradict the speaker's earlier beliefs*" leaves only the first justification to pursue further. Its consequent is the antecedent of a single justification rule, and instantiating this rule leads to this justification chain: *the lab members should clean the lab because their cleaning the lab is the best plan for the goal of keeping the lab clean because it's the only plan for keeping the lab clean*. The justified belief is TIDY's belief in (1), so the process stops.

In general, the understanding process is more complex, since justification rules may not be completely instantiated by a single antecedent, and may therefore need to be further instantiated from beliefs in the dialog context and memory. There also may be many possible chains to pursue even after heuristically discarding some of them, requiring the use of other heuristics to determine which path to follow, such as "*Pursue the reasoning chain which contains the most beliefs found in the dialog context*."

5 Selecting A Belief To Justify

After recognizing a participant's reasoning chain, it's necessary to select a belief to justify as a response. This task involves determining which beliefs are not shared, and selecting the negation of one of those beliefs to justify.

An intuitive notion of agreement is that a belief is shared if it it's found in memory or can be justified, and it's not shared if its negation is found in memory or can be justified. But this notion is computationally expensive, since it could conceivably involve trying to justify all the beliefs in the participant's reasoning chain, as well as their negations. As an alternative, our model determines whether a belief is shared by searching memory for the belief and its negation and, if that fails, applying a small

set of agreement heuristics. One such heuristic is "Assume a belief is shared if a justifying generalization is found in memory". So, for example, if the belief "keep everything clean" is found in memory, the belief "keep the AI lab clean" is considered to be shared. If no agreement heuristic applies, the belief is simply marked as "unknown".

After determining whether each belief in the participant's reasoning chain is shared, the model first searches for an existing justification for an unshared belief's negation. If that fails, it then tries to create a new justification for an unshared belief's negation. And if that fails, it tries to create a new justification for the negation of one of the unknown beliefs. This way existing justifications are presented before an attempt is made to construct new ones. If none of these steps succeed, the assumption is that the reasoning chain is shared, and an attempt is made to form a new justification for the belief it contradicts.

Thus, the belief our model addresses in a response arises from trying to discover whether or not it agrees with another participant's reasoning.

6 Forming Justifications

To form a new justification for a belief, our model performs a backward chaining search for a chain of justification rules that justify the given belief and that can be instantiated with beliefs from memory.

We briefly illustrate this process by showing how SCRUFFY forms the response in (2). The belief to justify is that it's not desirable to have the lab members clean the lab. The first step is to instantiate the justification rules that have this belief as their consequent. That results in several possible justifications: (1) *there's an undesirable enablement of cleaning the lab*, (2) *there's an undesirable effect of cleaning the lab*, or (3) *the lab members cleaning the lab conflicts with a more desirable action*.

The next step is to try to fully instantiate one of these rules. Applying the heuristic "*Pursue the most instantiated justification rule*" suggests working on the last rule. Here, SCRUFFY instantiates it with a belief from memory that research is more desirable than cleaning. Once a rule is instantiated, it's necessary to verify that the beliefs it contains are shared. Here, that involves verifying that cleaning conflicts with research. It does, so the instantiated rule can be presented as the response.

In general, the process is more complex than outlined here, since not all of the belief in an instantiated justification rule may be shared, and there may be several ways to instantiate a particular rule. Those rules containing unknown beliefs require further justification, while those rules containing unshared beliefs can be discarded.

7 Background

The closest related system is ABDUL/ILANA [8], which debated the responsibility and cause for historical events. It focused on the complementary problem of recognizing and providing *episodic justifications*, rather than justifications based on the relationships between different plans.

There are several models for recognizing the relationship between argument propositions. Cohen's [5] takes each new belief and checks it for a justification relationship with a subset of the previously-stated beliefs determined through the use of dialog structure and clue words. That model assumes the existence of an *evidence oracle* capable of determining whether a justification relationship holds between any pair of beliefs. Our model implements this oracle for a particular class of plan-oriented belief justifications. OpEd [3] recognizes belief justifications in editorials about economic planning through the use of argument units, a knowledge structure that can be viewed as complex configurations of justification rules. The approaches are complementary, just as scripts [7] and plans [6, 18] are both useful methods for recognizing the connections between events in a narrative.

Several systems have concentrated on producing belief justifications. Our own earlier work [14, 15, 16] used a primitive form of justification rules for factual beliefs as a template for producing corrective responses for user misconceptions. Our current model extends this work to use these rules in both understanding and responding, and provides additional rules for evaluative beliefs.

ROMPER [11] provides justifications for beliefs about an object's class or attributes. But it provides these justifications purely by template matching, not by constructing more general reasoning chains.

8 Current Status

We've completely implemented the model discussed in this paper. The program is written in Quintus Prolog and runs on an HP/APOLLO workstation. Its input is a representation for a stated participant belief, and its output is a representation for an appropriate response. It currently includes 30 justification rules and over 400 beliefs about various plans. We've used the program to participate in short argumentative dialogs in two disparate domains: day-to-day planning in the AI lab, and removing and recovering files in UNIX. We're currently using it to experiment with different heuristics for controlling the search process involved in recognizing and constructing these reasoning chains.

Our model has several key limitations we've only

now starting to address. First, it views plans as atomic units and considers only a small set of "all or nothing" plan-oriented beliefs. This means it can't produce or understand justifications involving steps in a plan, conditional planning relationships, or beliefs not directly involving plans. Second, our model can understand only those responses that justify an earlier belief. It can't, for example, understand a response that contradicts an inferred justification for an earlier belief. These more complex relationships can be represented using justification rules, but our model must be extended to recognize them. Third, our model is reactive rather than initiatory: it produces responses only when there's a perceived disagreement. It needs to be extended to know why its in an argument, and to be aware of the underlying goals of the other argument participants.

9 Conclusions

Previous dialog models have focused primarily on recognising a participant's plans and goals. But to participate in an argument it's also necessary to recognise when participants are providing beliefs about their plans and goals and how they're justifying these beliefs. It's also necessary to be able to determine which beliefs require further justification and to formulate justifications for these beliefs. This paper suggests a knowledge-based approach for these tasks.

Our approach has several attractive features. First, it builds a model of many relevant but unstated participant beliefs as a side-effect of trying to relate their utterance to the dialog. Second, it decides which belief to address in a response as a natural consequence of trying to understand why it disagrees with another participant's belief. Third, it understands belief justifications using the same general, common-sense planning knowledge that it uses to formulate them. Finally, it suggests how never before seen belief justifications can be understood, so long as they were formed from general justification rules known to the participants. That ability is crucial for participating in dialogs whose participants hold differing beliefs.

References

- [1] R. Abelson. Differences Between Beliefs and Knowledge Systems. *Cognitive Science* 3, 1979.
- [2] J.F. Allen and C.R. Perrault, Analysing Intention in Utterances. *Artificial Intelligence* 15, 1980.
- [3] S.J. Alvarado. *Understanding Editorial Text: A Computer Model of Argument Comprehension*. Kluwer, Boston, MA, 1990.
- [4] S. Carberry. Modeling the User's Plans and Goals. *Computational Linguistics*, 14(3), 1988.
- [5] R. Cohen. Analysing the Structure of Argumentative Discourse. *Computational Linguistics*, 13(1), 1987.
- [6] M.G. Dyer. *In-depth Understanding: A Computer Model of Narrative Comprehension*. MIT Press, Cambridge, MA, 1983.
- [7] M.G. Dyer, R.E. Cullingford, and S.J. Alvarado. Scripts. In *Encyclopedia of Artificial Intelligence*, John Wiley, NY, NY, 1987.
- [8] M. Flowers, R. McGuire, and L. Birnbaum. Adversary Arguments and the Logic of Personal Attacks. In *Strategies for Natural Language Processing*. Lawrence Erlbaum, Hillsdale, NJ, 1982.
- [9] R. Kass. Building a User Model Implicitly from a Cooperative Advisory Dialog. *User Modeling and User-Adapted Interaction*, 1(3), 1991.
- [10] D.J. Litman and J.F. Allen. A Plan Recognition Model for Subdialogues in Conversations. *Cognitive Science*, 11, 1987.
- [11] K. McCoy. Reasoning on a Highlighted User Model to Respond to Misconceptions. *Computational Linguistics*, 14(3), 1988.
- [12] M. Pollack. *Inferring Domain Plans in Question-Answering*. PhD Thesis, University of Pennsylvania, Philadelphia, PA, 1986.
- [13] A. Quilici. *The Correction Machine: A Computer Model Of Recognizing and Producing Belief Justifications in Argumentative Dialogs*. PhD Thesis, University of California, LA, CA, 1991.
- [14] A. Quilici. Participating In Plan-Oriented Dialogs. In *Proceedings of the 12th Conference of the Cognitive Science Society*, Boston, MA, 1990.
- [15] A. Quilici. The Correction Machine: Formulating Explanations for User Misconceptions. In *Proceedings of the 1989 International Joint Conference on Artificial Intelligence*, Detroit, MI, 1989.
- [16] A. Quilici, M.G. Dyer, and M. Flowers. Recognizing and Responding to Plan-oriented Misconceptions. *Computational Linguistics*, 14(3), 1988.
- [17] C. L. Sidner. Plan Parsing For Intended Response Recognition in Discourse. *Computational Intelligence*, 1(1), 1985.
- [18] R. Wilensky. *Planning and Understanding*. Addison Wesley, Reading, MA, 1983.