CONVERSATIONAL AGENT IN ARGUMENTATION
Updating of Information States

Mare Koit
Institute of Computer Science, University of Tartu, J. Liivi 2, Tartu, Estonia

Keywords: Dialogue model, Reasoning model, Conversational agent, Information state, Update rules.

Abstract: The paper describes a computational model that we are implementing in an experimental dialogue system. Conversation process is modelled where one participant is trying to influence his/her partner to agree to do an action. In the paper we concentrate on the representation of information states of the conversational agent and update rules which allow moving from one information state into another. Information state includes a partner model which consists of evaluations of different aspects of the action under consideration. The partner model is changing, based on the arguments and counter-arguments presented during the interaction. As a practical realization of the model we have in view a computer program which we call communication trainer.

1 INTRODUCTION

Modelling of conversational agents and development of dialogue systems is aimed to make interaction of human users with the computer more convenient. Conversational agents communicate with users in natural language in order to make travel arrangements, answer questions about weather or sports, route telephone calls, act as a general telephone assistant, or perform even more sophisticated tasks (Jurafsky and Martin, 2008).

Four kinds of dialogue management architectures are most common. The earliest and also one of the most sophisticated models of conversational agent behavior is based on the use of planning techniques (Allen, 1994). Plan-based dialogue models take into account communicative goals of dialogue participants and ways of their achieving, and offer flexibility of interaction with the computer but their creation and implementation on the computer is hard.

The two simplest and most commercially developed architectures are finite-state and frame-based (Wilks et al., 2005). The existing dialogue systems that interact with a user in natural language are mostly implemented as simple finite state automata which use regular expressions. In this way, it is possible to achieve robustness as needed in practical implementations because user's options and vocabulary are limited in every dialogue state. Still, these systems lack the flexibility and functionality which are important characteristics of human-human communication.

The most powerful are information-state dialogue managers (Traum and Larsson, 2003). Information state represents cumulative additions from previous actions in the dialogue, motivating future actions. The functions of the dialogue manager can be formalised in terms of information state update. The information state may include aspects of dialogue state and also beliefs, desires, intentions, etc. of dialogue participants.

We are dealing with interactions where the goal of one of the participants (*A*) is to get the partner (*B*) to carry out a certain action *D* (cf. Koit and Õim, 2004, Koit et al., 2009). *A* as initiator of the communication makes a proposal to the partner *B* to do an action *D*. If *B* refuses then *A* must influence him/her in the process of communication trying to see on which step of the reasoning the partner reached the negative decision.

In this paper, we will develop the model considered in (Koit et al., 2009). The paper has the following structure. In section 2 we give an overview of modelling the communication process between two participants. A model of conversational agent which involves a reasoning model will be presented. Section 3 considers interaction with the conversational agent as updating of information states. Section 4 discusses some aspects of implementation of the model and section 5 makes conclusions.

Koit M..
CONVERSATIONAL AGENT IN ARGUMENTATION - Updating of Information States.
DOI: 10.5220/0003636503750378
In *Proceedings of the International Conference on Knowledge Engineering and Ontology Development* (KEOD-2011), pages 375-378
ISBN: 978-989-8425-80-5

2 MODELLING THE COMMUNICATION PROCESS

Let us consider communication between a conversational agent *A* and its partner *B* (another conversational agent or human user). The process is defined if the following is given (Koit et al., 2009):

1) set *G* of communicative goals where both participants choose their own initial goals (G^A and G^B, respectively). In our case , G^A = "*B* makes a decision to do *D*"

2) set *S* of communicative strategies of the participants. A communicative strategy is an algorithm which a participant uses for achieving his/her communicative goal. This algorithm determines the activity of a participant at each communicative step

3) set *T* of communicative tactics, i.e. methods of influencing the partner. For example, *A* can entice, persuade, or threaten *B* in order to achieve its goal G^A

4) set *R* of reasoning models which is used by participants when reasoning about an action *D*. A reasoning model is an algorithm the result of which is a positive or negative decision about the object of reasoning (in our case, an action *D*)

5) set *P* of participant models, i.e. a participant's depiction of himself/herself and his/her partner:

$P = \{P^A(A), P^A(B), P^B(A), P^B(B)\}$

6) set of world knowledge

7) set of linguistic knowledge.

2.1 Reasoning Model

The reasoning process of a subject who should make a decision, to perform an action *D* or not (in our case, *B*), consists of a sequence of steps where the resources, positive and negative aspects of *D* will be weighed. Partner (*A*) cannot take part in this reasoning process explicitly. (S)he can direct the reasoning of *B* only by giving information about certain aspects of *D*, by stressing the positive aspects of *D* and downgrading the negative aspects. Positive aspects are pleasantness and usefulness of doing *D* for *B* but also punishment for not doing *D* if *D* is obligatory. Negative aspects are unpleasantness and harmfulness of doing *D* and punishment for doing *D* if *D* is prohibited.

The reasoning model consists of two parts: 1) a model of human motivational sphere; 2) reasoning schemes. We represent the model of motivational sphere of a subject by the following vector of weights assigned by him/her to different aspects of an action:

w = (w(resources), w(pleasant), w(unpleasant), w(useful), w(harmful), w(obligatory), w(prohibited), w(punishment-for-doing-a-prohibited-action), w(punishment-for-not-doing-an-obligatory-action)).

In the description, w(pleasant), etc. means weight of pleasant, etc. aspects of *D*. Such a vector (w^{AB}) is used by *A* as the partner model $P^A(B)$. The weights of the aspects of *D* are *A*'s beliefs about *B*. When interacting, *A* is making changes in the partner model if needed.

The second part of the reasoning model consists of reasoning schemes that supposedly regulate human action-oriented reasoning. A reasoning scheme represents steps that the agent goes through in its reasoning process; these consist in computing and comparing the weights of different aspects of *D*; and the result is the decision to do or not to do *D* (cf. Koit and Õim, 2004). In the motivational sphere three basic factors that regulate reasoning of a subject concerning *D* are differentiated. First, subject may *wish* to do *D*, if pleasant aspects of *D* for him/her overweigh unpleasant ones; second, subject may find reasonable to do *D*, if *D* is *needed* to reach some higher goal, and useful aspects of D overweigh harmful ones; and third, subject can be in a situation where (s)he *must* (is obliged) to do D – if not doing D will lead to some kind of punishment. We call these factors *wish-, needed-* and *must-*factors, respectively. They trigger the reasoning procedures *wish, needed* and *must*, respectively.

It is supposed here that the dimensions pleasant/unpleasant, useful/harmful, etc. have numerical values and that in the process of reasoning (weighing the pro- and counter-arguments) these values can be summed up.

In general this reasoning model follows the ideas of the Belief-Desire-Intention model (Allen, 1994).

2.2 Reasoning in Interaction

In the goal base of one participant (the conversational agent *A*) a goal G^A gets activated. *A* checks the partner model – supposed weights of the aspects of *D*. Then *A* chooses tactics of influencing of *B* (e.g. to persuade *B*, i.e. to stress the usefulness of *D*). Therefore, the agent sets up a sub-goal – to trigger in *B* a certain reasoning process (in case of persuading, by the *needed*-factor). *A* plans the dialogue acts and determines their verbal form as the first turn tr_1. This turn triggers a reasoning process in *B* where two types of procedures should be distinguished: the interpretation of *A*'s turn tr_1 and the generation of *B*'s response tr_2. The turn tr_2 triggers in *A* the reasoning cycle, *A* builds a new turn tr_3. Dialogue comes to an end, when *A* has reached

or abandoned its goal.

3 INTERACTION AS UPDATING OF INFORMATION STATES

3.1 Representation of Information States

The key of an information state is the partner model which is changing during the interaction.

There are two parts of an information state of a conversational agent – private (information accessible only for the agent) and shared (accessible for both participants). The private part consists of the following information slots:

- Current partner model (vector w^{AB} of weights – *A*'s picture about *B*)
- A tactic t_i^A which *A* has chosen for influencing *B*
- Reasoning procedure r_j which *A* is trying to trigger in *B* and bring to a positive decision (is determined by the chosen tactic, e.g. when persuading, *A* triggers the reasoning procedure *needed* in *B*)
- Stack of (sub-)goals under consideration. In the beginning, *A* puts its initial goal into the stack ("*B* decides to do *D*"). In every information state, the stack contains an aspect of *D* under consideration (e.g. when *A* is persuading *B* then usefulness is on the top)
- Set of dialogue acts $DA=\{d_1^A, d_2^A, ..., d_n^A\}$. There are the following DA-s for *A*: proposal, assessments for increasing or decreasing weights of different aspects of *D* for *B*, etc.
- (Finite) set of utterances as verbal forms of DA-s, incl. utterances for increasing or decreasing the weights ("arguments for/against") $U=\{u_{i1}^A, u_{i2}^A, ..., u_{iki}^A\}$. Every utterance has its own weight/numerical value: $V=\{v_{i1}^A, v_{i2}^A, ..., v_{iki}^A\}$ where v_{i1}^A, etc. is the value of u_{i1}^A, etc., respectively. Every argument can be chosen by *A* only once.

The shared part of an information state contains

- Set of reasoning models $R=\{r_1,...,r_k\}$
- Set of tactics $T=\{t_1, t_2, ..., t_p\}$
- Dialogue history – the utterances together with participants' signs and dialogue acts $p_1:u_1[d_1]$, $p_2:u_2[d_2],..., P_i:u_i[d_i]$ where $p_1=A$, p_2, etc. is *A* or *B*.

3.2 Update Rules

There are different categories of update rules which will be used for moving from the current information state into the next one:

I. Rules used by *A* in order to generate its turns:

1) For the case if the "title" aspect of the used tactic is located on top of the goal stack (e.g. if the tactic is persuasion then the "title" aspect is usefulness)
2) For the case if another aspect is located on the "title" aspect of the used tactic (e.g. if *A* is trying to increase the usefulness of *D* for *B* but *B* argues for unpleasantness, then the unpleasantness lies over the usefulness)
3) For the case if there are no more utterances for continuing the current tactic (and a new tactic should be chosen if possible)
4) For the case if *A* has to abandon its goal
5) For the case if *B* has made the positive decision and therefore, *A* has reached the goal.

II. Rules used by *A* in order to interpret *B*'s turns.

Special rules of the category I exist for updating the initial information state.

4 DISCUSSION

When *A* tries to bring *B* to a decision, *A* uses several statements to increase the weights of the positive aspects and to decrease the weights of the negative aspects of the action *D* under consideration. If *B* indicates a certain aspect which actual weight (too low or too high) does not allow him/her to do *D* then *A* simply can choose a statement for attacking this aspect. If *B* does not indicate a certain reason of rejection then *A* only can stress the usefulness when persuading.

Let us consider a brief example where the action *D* is "to prepare a potato salad" (cf. Koit et al., 2009). *A* has such a partner model that the reasoning procedure *needed* would give a positive decision. *A* will implement the tactic of persuasion.

The initial information state of *A* is as follows.

Private part

- Initial partner model w^{AB} = (w^{AB}(resources)=1, w^{AB}(pleasant)=3, w^{AB}(unpleasant)=0, w^{AB}(useful)=7, w^{AB}(harmful)=0, w^{AB}(obligatory)=0, w^{AB}(prohibited)=0, w^{AB}(punishment-for-doing-a-prohibited-action)=0, w^{AB}(punishment-for-not-doing-an-obligatory-action)=0)
- The tactic chosen by *A* – persuasion
- *A* tries to trigger the reasoning procedure *needed* in *B*
- Stack of goals under consideration contains only *A*'s initial goal
- Set of dialogue acts at *A*'s disposal

- Set of utterances for expressing the dialogue acts, together with their values {*I will help you* – value 1, etc.}.

The shared part of the initial information state contains

- The reasoning procedures *wish, needed*, and *must*
- The tactics of enticement, persuasion, and threatening
- Dialogue history – empty set.

A (computer): Please prepare a potato salad. [Proposal]

B (user): I do not have enough time. [Refusal to do D + assertion for decreasing the weight of resources]

Therefore, the actual value of w^B(resources) is 0. The computer tries to increase the value:

A: I will help you. [Rejection of the argument + assertion for increasing the weight of resources]

B: It is very hot in the kitchen. [Refusal to do D + rejection of the argument + assertion for increasing the weight of harmfulness]

Therefore, the weight w^{AB}(harmful) has to be corrected in the user model:

A: My kitchen has good ventilation. [Rejection of the argument + assertion for decreasing the harmfulness],

etc.

An experimental dialogue system is implemented which in interaction with a user can play the role of *A*. At the moment, the computer operates with semantic representations of linguistic input/output only, the surface linguistic part of interaction is provided in the form of a list of ready-made and classified utterances both for the computer and user.

5 CONCLUSIONS

We are dealing with interactions where the goal of one participant is to get the partner to carry out a certain action. The paper describes a computational model that we are implementing in an experimental dialogue system. We concentrate on the representation of information states and update rules. Information state includes a partner model which consists of evaluations of different aspects of the action under consideration. The partner model is changing during the interaction, based on the arguments and counter-arguments presented. As a practical realization of the model we have in view a computer program which we call communication trainer.

We are continuing our work in the refining the model, considering different scenarios, e.g. *A* and *B* have opposite goals and one of them has to abandon his/her initial goal (as considered so far), or they collaborate in order to achieve a common goal; both of *A* and *B* are conversational agents with their own information states and update rules. Different communicative strategies/tactics used by participants will be evaluated taking into account their success in achieving the initial goal.

ACKNOWLEDGEMENTS

This work is supported by the European Regional Development Fund through the Estonian Centre of Excellence in Computer Science (EXCS), and the Estonian Science Foundation, grants 7503 and 8558.

REFERENCES

Allen, J., 1994. *Natural Language Understanding.* Second Edition. The Benjamins/Cummings Publ. Co. Redwood City etc., 678 p.

Jurafsky, D., Martin, J., 2008. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition.* Prentice Hall, 1028 p.

Koit, M., Õim, H., 2004. Argumentation in the Agreement Negotiation Process: A Model that Involves Natural Reasoning. In *Proc. of the Workshop W12 on Computational Models of Natural Argument. 16th European Conference on Artificial Intelligence.* Valencia, Spain, 53–56.

Koit, M., Roosmaa, T., Õim, H., 2009. Knowledge representation for human-machine interaction. In *Proceedings of the International Conference on Knowledge Engineering and Ontology Development: International Conference on Knowledge Engineering and Ontology Development, Madeira (Portugal), 6-8 October 2009.* Jan L.G. Dietz (Ed.), Portugal: INSTICC, 396–399.

Traum, D., Larsson, S., 2003. The Information State Approach to Dialogue Management. In *Current and New Directions in Discourse and Dialogue*, J. van Kuppevelt and R. Smith (Eds.), Kluwer, 325–353.

Wilks, Y., Webb, N., Setzer, A., Hepp, M., Catizone, R., 2005. Machine learning approaches to human dialogue modelling. In *Text, Speech and Language Technology. Advances in Natural Multimodal Dialogue Systems*, Vol. 30. Springer, 355–370.