SemDial 2011: Proceedings of the 15th Workshop on the Semantics and Pragmatics of Dialogue, pages 178–179.
Los Angeles, California, 21–23 September 2011.

Unveiling the Information State with a Bayesian Model of the Listener

Hendrik Buschmeier and **Stefan Kopp**
Sociable Agents Group, CITEC and Faculty of Technology, Bielefeld University
PO-Box 10 01 31, 33501 Bielefeld, Germany
{hbuschme, skopp}@techfak.uni-bielefeld.de

Abstract

Attentive speaker agents – artificial conversational agents that can attend to and adapt to listener feedback – need to attribute a mental 'listener state' to the user and keep track of the grounding status of their own utterances. We propose a joint model of listener state and information state, represented as a dynamic Bayesian network, that can capture the influences between dialogue context, user feedback, the mental listener state and the information state, providing an estimation of grounding.

1 Introduction

Listeners providing communicative feedback reveal – not always deliberately – their mental state of processing to speakers (Allwood et al., 1992). Producing a backchannel (e.g., a quick 'yeah' or a nod) at appropriate places in the dialogue signals that they attend to and perceive what the speaker is saying. Looking puzzled or producing a hesitant 'yeah', on the other hand, might show that they have difficulties understanding what the speaker wants to express. Speakers attend to these signals, use them as information for grounding (Clark and Schaefer, 1989), and take them into account when producing their ongoing and subsequent communicative actions.

To be able to do this, speakers need to interpret a listener's feedback signal in its context and infer what the listener indicates, displays, or signals. Using this information, speakers can refine the model they have of their interlocutor and conjecture about the grounding status of dialogue moves in the information state that caused the listener to produce this feedback signal.

In the context of enabling virtual conversational agents to attend to and adapt to user feedback, we proposed that such an 'attentive speaker agent' maintains an 'attributed listener state' (ALS) of its user (Buschmeier and Kopp, 2011). The ALS is the part of the agent's interlocutor model that is particularly relevant when processing communicative listener feedback since it represents the agent's knowledge about the user's current ability to perceive and understand the agent's actions.

Here we propose a more sophisticated approach to ALS that integrates with the agent's information state and is modelled as a (dynamic) Bayesian network, giving the agent degrees of belief in the user's mental state as well as the grounding status of the current dialogue move.

2 Model

Figure 1 shows a schema of the model. Each step in time (t, t+1) corresponds to one dialogue move of the agent. The attributed listener state contains three nodes C, P and U that model whether the user is in contact with the agent and perceives and understands the agent's utterance. Allwood et al. (1992) propose that these functions of feedback relate to each other: being in contact is, for instance, a prerequisite for perception, which in turn is a prerequisite for understanding. We can easily capture these relations in terms of influences in our Bayesian network model. Evidence that contact is established increases the degree of belief in the user being able to perceive what is said, which in turn increases the degree of belief in her understanding the utterance.

Variables influencing the nodes in the ALS are hidden in the boxes 'Context' and 'User FB'. Important

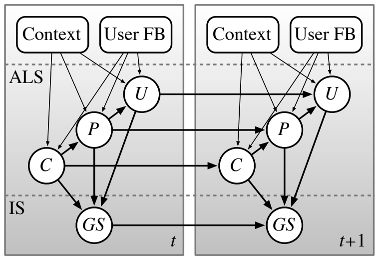

Figure 1: Attributed listener state (ALS) and information state (IS) modelled as one dynamic Bayesian network. User feedback and dialogue context influence the degrees of belief in contact, perception and understanding (C, P, U) in the ALS. These determine the grounding status (GS) of the current dialogue move kept in the information state.

contextual factors for perception might, for example, be whether noise is present in the environment, or the occurrence probability of the agent's utterance calculated by an *n*-gram language model. The type of the user's feedback function as well as certain features of the feedback signal obviously also influence the ALS nodes. Presence of feedback signalling 'understanding' increases the agent's degree of belief in U and should certainly influence P and C as well. Similarly, the influence of a prosodically flat 'yeah' on U should be smaller than an enthusiastic one.

Our model also enables the agent to relate different kinds of user feedback to the grounding status of the dialogue move it refers to. This is modelled with a node GS in the information state part of the model. If we have evidence from feedback signals that the user understood the agent's utterance, the degree of belief in the dialogue move being in the common ground should be high. If, in contrast, the agent only got feedback of the communicative function 'perception', the degree of belief in the dialogue move being grounded should be lower. Nevertheless, depending on the context (for example, the dialogue move is simple and there is no apparent reason for the user not to understand it) the degree of grounding can still be high enough to take it as being grounded.

Finally, our model is a *dynamic* Bayesian network since the previous dialogue move influences the current one. If the previous move has a high degree of being grounded this should increase belief in the current move being grounded as well. Similar assumptions can also be made about the values of C, P and U in the ALS.

3 Discussion and Conclusion

We presented first steps towards a joint model of attributed listener state and information state for artificial conversational agents. Modelled as a dynamic Bayesian network, it can easily capture the influences between dialogue context, user feedback, the mental listener state the agent attributes to the user and the grounding status of the agent's dialogue moves.

This is an improvement on our previous model of listener state (Buschmeier and Kopp, 2011), since dialogue context and features of feedback signals can be taken into account during state estimation. In contrast to state of the art models of (degrees of) grounding (Traum, 1994; Roque and Traum, 2008) the model presented here allows for continuous instead of discrete grounding values, based on the user's feedback signals and the dialogue context.

Several issues, however, have not yet been addressed. It is, for instance, still unclear how exactly the timing of feedback signals will be handled. Furthermore, although a simple hand-crafted prototype looks promising, the question how such a network can be learnt is open as well.

Acknowledgements This research is supported by the Deutsche Forschungsgemeinschaft (DFG) in the Center of Excellence in 'Cognitive Interaction Technology' (CITEC).

References

Jens Allwood, Joakim Nivre, and Elisabeth Ahlsén. 1992. On the semantics and pragmatics of linguistic feedback. *Journal of Semantics*, 9:1–26.

Hendrik Buschmeier and Stefan Kopp. 2011. Towards conversational agents that attend to and adapt to communicative user feedback. In *Proceedings of the 11th International Conference on Intelligent Virtual Agents*, pages 169–182, Reykjavik, Iceland.

Herbert H. Clark and Edward F. Schaefer. 1989. Contributing to discourse. *Cognitive Science*, 13(2):259–294.

Antonio Roque and David R. Traum. 2008. Degrees of grounding based on evidence of understanding. In *Proceedings of the 9th SIGdial Workshop on Discourse and Dialogue*, pages 54–63, Columbus, OH.

David R. Traum. 1994. *A Computational Theory of Grounding in Natural Language Conversation*. Ph.D. thesis, University of Rochester, Rochester, NY.