In Proceedings of the 12th International Conference on Intelligent Virtual Agents, IVA-2012, pp. 517–519, Santa Cruz, CA. The original publication is available at springerlink.com PREPRINT

Understanding How Well *You* Understood – Context-sensitive Interpretation of Multimodal User Feedback

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}@uni-bielefeld.de

1 Introduction

Human interlocutors continuously show behaviour indicative of their perception, understanding, acceptance and agreement of and with the other's utterances [1,4]. Such evidence can be provided in the form of verbal-vocal feedback signals, head gestures, facial expressions or gaze and often interacts with the current dialogue context. As feedback signals are able to express subtle differences in meaning, we hypothesise that they tend to reflect their producer's mental state quite accurately.

To be cooperative and human-like dialogue partners, virtual conversational agents should be able to interpret their user's evidence of understanding and to react appropriately to it by adapting to their needs [2]. We present a Bayesian network model for context-sensitive interpretation of listener feedback for such an 'attentive speaker agent', which takes the user's multimodal behaviour (verbal-vocal feedback, head-gestures, gaze) as well as its own utterance and knowledge of the dialogue domain into account to form a model of the user's mental state.

2 Bayesian Model of the Listener

In previous work [2], we adopted the concept of 'listener state' [5] for a model that an attentive speaker agent *attributes* to its user, i.e., a representation that emulates the user's listener state. Here we present an implementation of 'attributed listener state' (ALS) by modelling it probabilistically as a Bayesian network. This (1) allows us to manage the uncertainties inherent in the mapping between feedback signal and meaning; (2) gives us a natural and robust mechanism of interpreting feedback in its dialogue context; and (3) enables inference and learning within a well understood formalism.

The Bayesian network (Figure 1a) models the notions of contact, perception, understanding, acceptance and agreement with one random variable each, so that the values of C, P, U, AC and AG are to be interpreted in terms of degrees of belief. Assuming discrete variables for simplicity, strengths are modelled via their states: *low, medium* and *high*. The influences between ALS-variables are modelled after Allwood's hierarchy of feedback functions [1], e.g., if understanding is assumed, perception and contact can be assumed as well; a lack of perception, on the other hand, usually implies that understanding cannot be assumed either. Apart from influencing each other, the ALS-variables are influenced by the dialogue context and the user's multimodal feedback behaviour, which we model here, exemplarily, in the form of six influencing variables.

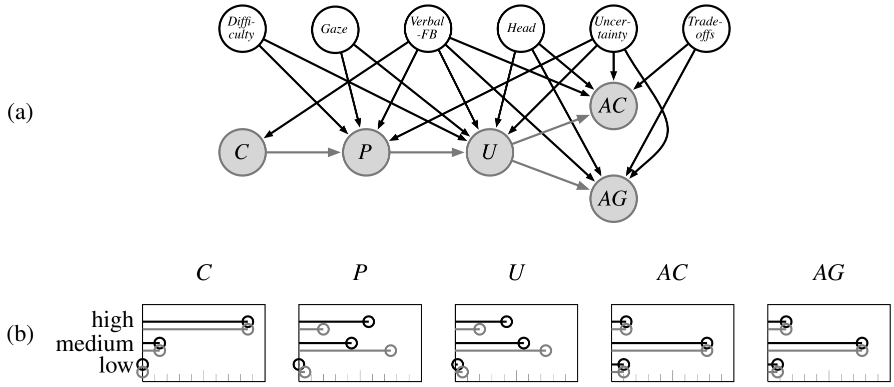

Figure 1. (a) Structure of the Bayesian model of the listener. The attributed listener state, drawn in shades of grey, consists of five random variables C, P, U, AC and AG. These are influenced by variables representing the dialogue context and the user's behaviour (drawn with black lines). (b) Plot of the example belief states. The x-axes show the probabilities of each variable's state. Black lines show the values for the first, grey lines values for the second variant of the example.

3 Worked Example

To demonstrate that the model behaves reasonably, we have tested its performance in an example situation in a calendar management domain. Figure 1b shows the belief state of the ALS-variables C, P, U, AC and AG under a certain assignment of (some of) the variables that represent the user's behaviour and the dialogue context. Model parameters (i.e., the conditional probability tables) were generated from structured representations [3]. In the example situation, the agent produces the utterance 'Your supervisor would like to meet you for lunch on Thursday at 12 PM', which is of *medium* difficulty as it might be a bit surprising but does not involve any complex structures or lexical items. The agent further knows that there will be no conflict with any other appointment of the user (*Trade-offs* = *low*). The user verbally signals understanding (*Verbal-FB* = *u*) but does not move her head (*Head* = *none*). In the first variant (black lines), she gazes at the correct target slot (*Gaze* = *on-target*), in the second variant (grey lines) at an incorrect target slot in the calendar (*Gaze* = *off-target*). As shown in Figure 1b, the belief state makes reasonable predictions of the mental state that the user might be in. The probability mass of the variables P and U is distributed mostly between *medium* and *high* in both variants. This makes sense, as the utterance was not too difficult and the user verbally signalled understanding. In the first variant, however, where the user looks at the correct target, the probability mass is shifted towards *high* for both variables P and U, whereas in the second variant, where the user is not looking at the correct target slot, the probability mass is shifted more towards *medium*. Since gazing at the correct target is not a strong indicator of acceptance or agreement, there is only a minimal difference between the two variants for the variables AC and AG.

More information on the model's causal structure and inner details along with its perfomance in further example situations is presented in [3].

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

References

1. Allwood, J., Nivre, J., Ahlsén, E.: On the semantics and pragmatics of linguistic feedback. Journal of Semantics 9, 1–26 (1992)
2. Buschmeier, H., Kopp, S.: Towards conversational agents that attend to and adapt to communicative user feedback. In: Proceedings of the 11th International Conference on Intelligent Virtual Agents. pp. 169–182. Reykjavik, Iceland (2011)
3. Buschmeier, H., Kopp, S.: Using a Bayesian model of the listener to unveil the dialogue information state. In: Proceedings of the 16th Workshop on the Semantics and Pragmatics of Dialogue. Paris, France (to appear)
4. Clark, H.H.: Using Language. Cambridge University Press, Cambridge, UK (1996)
5. Kopp, S., Allwood, J., Grammar, K., Ahlsén, E., Stocksmeier, T.: Modeling embodied feedback with virtual humans. In: Wachsmuth, I., Knoblich, G. (eds.) Modeling Communication with Robots and Virtual Humans, pp. 18–37. Springer-Verlag, Berlin, Germany (2008)