A model of intentional communication: AIRBUS (Asymmetric Intention Recognition with Bayesian Updating of Signals)

J. P. de Ruiter and Chris Cummins
Bielefeld University

Abstract

The rapid and fluent nature of human communicative interactions strongly suggests the existence of an online mechanism for intention recognition. We motivate and outline a mathematical model that addresses these requirements. Our model provides a way of integrating knowledge about the relationship between linguistic expressions and communicative intentions, through a rapid process of Bayesian update. It enables us to frame predictions about the processes of intention recognition, utterance planning and other-repair mechanisms, and contributes towards a broader theory of communication.

Introduction

The ability to communicate effectively and flexibly with other humans is one of our species' most impressive cognitive capacities. However, there are very few comprehensive theories that aim to address this capacity, and those that do are often sketchy and fail to capture the essential and unique aspects of human communication.

Most notably among these, Shannon's (1948) mathematical theory of signal transmission is of limited use in modeling human-human communication. This model assumes that the encoder function that the sender uses to convert a message into a signal is the inverse of the decoder function that the receiver uses to reconstruct the message from the signal. This is not descriptively adequate for human communication, whose complex many-to-many mappings sometimes break down, resulting in miscommunication. The influential recent Interactive Alignment model (Pickering and Garrod 2004) implicitly assumes even similar encoding and decoding functions, namely the identity function.

From a more linguistic perspective, Grice's (1957) theory of meaning provides a very concise definition of what constitutes (intentional) communication, but is atheoretic as to how this process is accomplished. Research in this tradition encounters the daunting complexities, and potential infinite regress, associated with the recognition of mutual knowledge or *common ground* (Stalnaker, 1978; Clark and Marshall, 1981). Reductionist approaches to this problem are motivated by the intuition that full common ground processing is implausible given the speed and efficiency of typical dialogue. The immediacy of turn-taking (Stivers et al. 2009) and back-channel responses (Yngve 1970) speak to the need for rapid online heuristics that enable hearers to identify the general nature of the speaker's communicative intention or illocution.

The absence of models of human communication that address these competing concerns is keenly felt. Here we propose a mathematical model of communication that crucially relies upon the use of *shared conventions* to achieve efficiency, and that applies a form of Bayesian updating to address the many-to-many mapping problem. Rather than attempting to apply machine learning techniques such as POMDP to learn optimized mappings from utterances to appropriate responses in one fell swoop, we focus on the more tractable problem of recognizing the category of utterance involved. This enables us to consider the full range of different communicative contexts without succumbing to unsolvable complexity in the case of infinitely productive human language.

In the following we outline the technicalities of the model and discuss some of its implications.

Outline specification of the model

The AIRBUS model takes a signal as its input and calculates the corresponding intention. The model assumes a finite, predefined set of communicative intentions. It has access to three forms of information: a convention database C, which specifies the probability of communicative intentions given a certain signal; a likelihood database L, specifying the probability of signals given a certain communicative intention, and a set of prior probabilities E as to the communicative intention, conditioned by the social and discourse context.

The operation of the model consists of updating the prior probabilities E in the light of a new incoming signal, taking into account the information in C and L. We propose the following stages of update. Given a new signal s, the model examines whether there is an entry in C corresponding to the signal s. If so, the probabilities in this entry are averaged with the probabilities in E, creating a set of revised probabilities R. R is then treated as a prior and subjected to Bayesian update in the light of L. The resulting probability distribution over I is used to infer the speaker's intention. This process cycles as the signal continues and further convention-bearing units are transmitted.

Within this model, we can measure the success or the usefulness of a communicative act by considering the extent to which it reduces the hearer's uncertainty as to the speaker's intention. Following Shannon (1948), we can measure this by considering the entropy of the prior and posterior probability distributions over the possible intentions in I. We propose that the hearer commences the planning of a response when the entropy is low enough.

Separately, we propose that repair mechanisms are activated if there is too large a difference between prior and posterior distributions: that is, if the hearer's understanding of the speaker's intention is radically altered during the update process. A large difference would suggest disalignment between speaker and hearer, and the possible need for explicit repair negotiation. We can measure this difference using Kullback-Leibler divergence, a standard measure of relative entropy, and posit that sufficiently high K-L divergence triggers explicit repair.

Discussion

The model outlined above provides a rapid means to infer communicative intentions. It posits a powerful decoding process, using the hearer's knowledge about both directions of the relationship between signal and intention to draw pragmatic conclusions about the speaker's intended meaning. Moreover, by its use of probability distributions rather than categorical rules, the model is able to handle improbable events gracefully. Relative entropy allows us to predict when the model does break down to such an extent that explicit negotiation is required.

In this brief sketch we have necessarily left many issues open. We did not discuss how the likelihood and convention databases are to be populated. Another open question is which aspects of the utterance are listed in the convention database: that is, do the conventions relate to lexical items, syntactic categories (such as VP), or some other form of regular expression? Finally, a very general question concerns the nature of the possible intentions themselves, an issue that has been explored from many directions. However, although we concede that the correct set of intentions must be posited in order to precisely simulate human behaviour, we would argue that the use of any plausible proxy set should be adequate in principle to achieve a close approximation to this behaviour.

In future work we aim to explore the capabilities of this model through a range of qualitative and quantitative tests. The model gives rise to testable predictions as to a wide range of behaviours. These include the attribution of communicative intentions, the planning of conversational turns, and instances of repair. We feel that the model has considerable practical potential in providing enhanced artificial discourse capabilities, and that if this promise is borne out, it could also have substantial implications for the modelling of dialogic behaviour in natural language.

References

Herbert H. Clark and Catherine Marshall. 1981. Definite reference and mutual knowledge. In A. K. Joshi, B. L. Webber and I. A. Sag (eds.), Elements of Discourse Understanding. New York: Cambridge University Press. 10-63.

H. Paul Grice. 1957. Meaning. Philosophical Review, 67: 377-388.

Martin J. Pickering and Simon Garrod. 2004. Toward a mechanistic psychology of dialogue. Behavioral and Brain Sciences, 27: 169-226.

Claude E. Shannon. 1948. A mathematical theory of communication. Bell System Technical Journal, 27: 379-423.

Robert Stalnaker. 1978. Assertion. Syntax and Semantics, 9: 315-332.

Tanya Stivers et al.. 2009. Universals and cultural variation in turn-taking in conversation. Proceedings of the National Academy of Sciences of the United States of America, 106: 10587-10592.