Naturalistic Conversational Communication of Structured Data through Cognitive Belief Modeling

Eric Saund
Palo Alto Research Center*
Palo Alto, CA 94304, USA
saund@alum.mit.edu

Abstract
Even when communicating highly structured information, human conversants dynamically adjust their utterances based on perception of their partner's reactions and responses. How can machines borrow from human strategies to achieve more naturalistic communication abilities? One key insight is that human conversants maintain a mental model of what their partner has heard and understood. This model is continuously updated, and is used in real time to decide what to say next, and how to say it. This abstract outlines principles and a computational model for a conversational agent that implements this capability.

Author Keywords
conversational agents; conversation analysis; cognitive belief modeling; naturalistic conversational user interfaces.

ACM Classification Keywords
I.2.1 [Artificial Intelligence]: Natural Language Processing–Discourse

Introduction
Currently available conversational assistants are idiot savants. They can access vast databases and web resources at a large scale to answer one-shot natural language queries. Some agents are able to muster a small degree of multi-turn context for slot-filling, goal stacking, and pronoun res-

olution. But no automated system approaches the fluidity and naturalness of human conversation. In contrast to the broad and flat capabilities of current technology, we aim for narrow yet deep, to devise agents that can conduct richly interactive conversations about limited, highly structured topics.

Partial Listing of Logical Form Types

RequestTopicData
what is the area code?

InformTopicData
it's six five zero

RequestDialogManagement
was that the area code?

InformDialogManagement
that's what I said

CheckTopicData
six five zero
(echo in context)

ConfirmTopicData
that's correct

CorrectionTopicData
no not six five zero

To constrain the problem, we focus on a simple yet practical task. The computer agent's job is to convey a ten-digit telephone number to a human user. Obviously, the agent could simply recite the ten digits in open-loop fashion. Informal observations of people conveying telephone numbers to one another in real life indicate that this practice does sometimes occur, but more often, the communication is accomplished through subtle give-and-take that evoke strategies identified by Conversation Analysts. These strategies include linguistic and non-linguistic vocalizations; pauses; adjustment of pacing and chunk sizes; repeats of content items; solicitation and offerings of confirmations; detection of errors; conversational repair; references to just-spoken utterances; and variations in intonation and emphasis.

An example is provided in Table 1.

For a computer agent to communicate in this fashion, a dialog management system must support an open-ended path through a dialog space of very high dimension. In contrast to rote recitation of a number sequence, a sophisticated decision must be made at every step to choose an appropriate utterance that will advance the goal of transmitting all ten digits of topic information, while ensuring both that the user has received the information correctly, and that the user remains engaged and satisfied with the interaction.

Our research on this task addresses the following questions:

- What internal meta-parameters should be maintained by the computational agent to support naturalistic dialogue?
- Can an adequate dialog manager be built around the Entity/Intent paradigm that is currently ascendant in conversational agent technologies?
- How can theories of conversation be applied to harness the combinatorially explosive state space of possible dialogue pathways?

In greatly abbreviated form, we outline some answers to these questions.

Per contemporary theory, conversational acts occur at three Dialog Levels [1]:

Propositional (e.g. "The area code is six five zero.")
Dialog Management (e.g. "Could you say that again?")
Interpersonal (e.g. "You are not listening very well.")

Propositional acts refer to data values themselves e.g. "six", and also indices to data structures ("area code") and indexicals ("the next digit"). To establish goals for organized transmission of propositional information, we employ a *Cognitive Belief Model* under which a computer agent maintains data structures that hold its belief about the human recipient's knowledge of the target telephone number as a probability distribution [3]. At the outset, the recipient is assumed to be ignorant of the Topic Data values. During the course of conversation, belief probabilities that the human has acquired the correct number, or not, are adjusted continuously by evaluation of dialog content.

U: what is the telephone number?	[initial query]
A: *area code six five zero*	[inform topic data including explicit segment index]
U: six five zero	[check echo (correct)]
A: *six three seven*	[new topic data, implicit segment index]
U: six three eight	[check echo (incorrect)]
A: *i'll repeat that. the exchange is six three seven*	[disconfirmation, repeat of topic data]
U: six	[check echo]
A: *three*	[repeat of topic data, adjustment of chunk size to match user]
U: three	[check echo]
A: *seven*	[repeat of topic data]
U: did you say nine?	[confirmation request, reference to A's utterance]
A: *no. it's seven*	[correction, repeat of topic data]
U: did you say seven?	[confirmation request, reference to A's utterance]
A: *yes. seven*	[confirmation, repeat of topic data]
U: okay	[confirmation, turn handoff]
A: *the line number is one two one two*	[new topic data with explicit segment index, readjustment of chunk size to match structured data]
U: what does line number mean?	[request for topic data]
A: *the line number is the last four digits of the telephone number*	[new topic data in response to request]
U: okay	[confirmation, turn handoff]
A: *the line number is one two one two*	[repeat topic data with explicit segment index]
U: okay, got it	[confirmation, turn handoff]
A: *we're all done*	[closing]

Table 1: Typical dialogue pattern for communicating a telephone number between an Agent A: (Topic Information Sender) and a User U: (Topic Information Receiver).

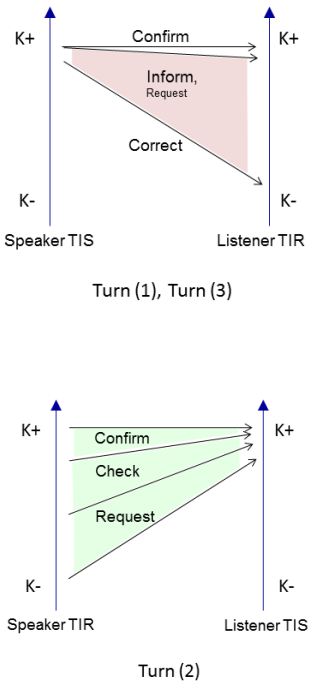

Figure 1: Conversants' appropriate dialog acts depend on the epistemic gradient between them, i.e. their beliefs about their relative knowledge.

The Dialog Management level is where CA theory applies most directly, including determination of readiness, transfer of turn, control of topic, pacing, chunk size, and handshake. These factors also can be modeled in terms of internal parameters, some categorical, some continuous-valued, and some probabilistic. For example, most utterances provide cues to turn ownership at the Dialog Management level, but there can be ambiguity and misalignment; a probabilistic variable is thus most apt.

The transmitter of information such as a telephone number is referred to as the *Topic Information Sender (TIS)*, and the recipient as the *Topic Information Receiver (TIR)*. In the theory of Heritage [2], this relationship creates an *epistemic gradient* that reflects the relative epistemic status of the speakers. Then, in production of conversation, the participants' subjective beliefs about their epistemic statuses give rise to *epistemic stance*. The notation, K+ indicates confidence in participant's knowledge, and K- indicates belief that the participant lacks knowledge. For the structured data transmittal task, we have found that this principle organizes five types of dialog acts that suffice to characterize a preponderance of Propositional and Dialog Management utterances:

Request K- stance party requests information from K+ partner
Inform K+ party conveys information to partner
Check party tests belief with K+ stance partner
Confirm K+ party ack./agrees with K+ stance partner
Correct K+ party corrects statement by K- partner

Figure 1 illustrates the role each of these Dialog Act types plays under the epistemic gradient framework.

When composed with the three Dialog Levels, we arrive at a categorization vocabulary of fifteen Dialog Intents (sidebar). Following standard practice, these exist in a formal vocabulary of Logical Forms that can be translated to and from natural language utterances by parsers and generators.

Note however, that isolated utterances can be ambiguous with respect to Logical Form Intent, and disambiguation requires appeal to conversational context. For example, the utterance "six five zero" is an InformTopicData LF when uttered by TIS, but it becomes a CheckTopicData LF when uttered by TIR, as an echoing of topic data issued by TIS.

Using these principles, we can build a conversational agent TIS that interacts with a human TIR in a naturalistic fashion. The dialog sample in Table 1 is not just a representation of prototypical human exchange, it is a literal transcription of our agent interacting with a user. We intend to present an audio clip, and preferably a live demo of the system at the workshop.

References

[1] Justine Cassell. 2001. Embodied Conversational Agents: Representation and Intelligence in User Interfaces. *AI Magazine* (2001).

[2] John Heritage. 2012. Epistemics in Action: Action Formation and Territories of Knowledge. *Research on Language and Social Interaction* 45, 1 (2012), 1–29. http://www.sscnet.ucla.edu/soc/faculty/heritage/Site/Publications_files/EPISTEMICS_IN_ACTION.pdf

[3] Antonio Roque and David Traum. 2008. Degrees of Grounding Based on Evidence of Understanding. *The 9th SIGdial Workshop on Discourse and Dialogue (SIGdial 2008)* (2008). https://aclweb.org/anthology/W/W08/W08-0107.pdf