Artificial Intelligence for Human-Robot Interaction
AAAI Technical Report FS-17-01

A Theoretical Model of Indirect Request Comprehension

Sean Trott, Benjamin Bergen
University of California, San Diego, Cognitive Science
sttrott@ucsd.edu, bkbergen@ucsd.edu

Abstract

Natural human dialogue often contains ambiguous or indirect speech. This poses a unique challenge to language understanding systems because comprehension requires going beyond what is said to what is implied. In this paper, we survey related work on the particularly challenging case of understanding non-conventional indirect speech acts, then propose a more generalizable rule rooted in building a mental model of the speaker. Finally, we discuss experimental evidence pointing to the cognitive plausibility of this rule.

Introduction

People often speak indirectly in natural conversation—their intended message is different from the literal interpretation of their utterance (Searle, 1975; Gibbs, 1981). Indirect speech ranges from highly conventional phrases, like "Can you pass the salt", to less conventional, like "This kitchen is a pigsty" (as a request to clean the dishes). By some estimates, as much as 85% percent of requests that humans produce are indirect in one or the other of these ways (Gibbs, 1981).

Indirect speech creates challenges for human comprehenders and no less so for machines charged with determining a speaker's intentions. Wilske and Kruijff (2006) distinguish between using *idiomatic* and *inferential* strategies for the machine interpretation of indirect speech acts, roughly corresponding to the processes needed to understand conventional and non-conventional utterances like those above. Conventional indirect requests follow a highly predictable form, e.g. "Can you *X*", meaning a generalizable rule can be written for a language understanding system (Trott, Eppe, and Feldman, 2016). But non-conventional indirect requests, such as "This kitchen is a pigsty", are often context-dependent, and require inference to interpret.

In this paper, we first describe an existing architecture for understanding indirect requests. Then we propose a solution for interpreting non-conventional indirect requests, and discuss experimental evidence in support of it.

Previous Work

The state-of-the-art model for understanding indirect requests is the DIARC system. This is an end-to-end system for language understanding and production, including the interpretation of indirect requests (Briggs and Scheutz, 2013; Williams et al, 2014; Williams et al, 2015; Briggs and Scheutz, 2016). The system described in Williams et al (2014) and Williams et al (2015) uses a Dempster-Shafer probabilistic framework for reasoning about pragmatic rules, and ultimately inferring a speaker's intentions, **I**, given some utterance, **U**, and a representation of the context, **C**, e.g. $P(I \mid U, C)$.

Hand-coded pragmatic rules relate different contexts to different interpretations of the same utterance. Consider the utterance:

(1) *Commander Z needs a med-kit*.

This utterance could be interpreted as either a request to bring the speaker a med-kit, or as a request to know where the closest med-kit could be found. In both cases, the speaker is making a request, but the conditions for fulfilling the request depend on which action the speaker desires from the listener.

Here, the different interpretations are modulated by "the robot's assumptions regarding its interlocutor's beliefs about who is subordinate to whom" (Williams et al, 2015). If the robot believes that the speaker believes the robot is subordinate to the speaker, **U** is interpreted as a request to bring the med-kit; if the robot believes that the speaker believes the speaker is subordinate to the robot, **U** is interpreted as a request to know where a med-kit could be found. These beliefs can be represented as probability intervals using the Dempster-Shafer framework.

This system works very well for a known class of utterances and contexts, as long as there are pragmatic rules mapping **U** and **C** to **I**. In particular, it can distinguish successfully between the different interpretations of conventional indirect requests, in which the space of solutions is constrained by the form of the utterance, and the different disambiguating contexts are easily identified and represented.

Remaining Challenges

The problem becomes much more difficult for non-conventional indirect requests – cases like "This kitchen is a pigsty" – in which the space of solutions is not constrained by the utterance's form. There are at least two general challenges here:

The first challenge is **enumerating the mappings between utterances and intended interpretations**. The intended interpretation of non-conventional indirect requests cannot be reliably inferred from their form. This makes it difficult to both: a) enumerate all the ways in which someone might express some intention, **I**; and b) enumerate all the possible intentions of (or solutions to) some utterance, **U**. For example, a request to close the window could take many forms: "It's getting cold in here", "The breeze is getting in", or "Is there any chance you could close that window?" Conversely, the actionable responses to the utterance "It's getting cold in here" could include closing the window, turning on the heater, grabbing a blanket, sympathizing with the speaker's plight, and more.

The second challenge is **enumerating the disambiguating contexts**. Given the sheer variety of "contexts" that can occur, it is difficult to identify all of the situations in which some utterance, **U**, would be a request (and if it is a request, what *kind* of request it would be), and all of the situations in which it would not. For example, "It's so cold here" could be interpreted as a request or a complaint depending on: where the speaker and listener are situated (a room vs. an icy tundra), the identities of the speaker and listener (e.g. whether one is subordinate to the other), how those identities relate to situational context and their respective entitlements to fulfill a request (e.g. whether they are in the speaker's room or the listener's room), whether there even exists a solution to the request interpretation at all, whether the speaker and listener know about this solution, and many more. Even worse, many of these contextual variables can be combined in various ways, with each combination favoring different interpretations.

Both of these challenges can be described more formally:

1. Given some utterance, **U**, generate the set of possible intended interpretations, $\{\mathbf{I_1}, \mathbf{I_2}\ldots\mathbf{I_n}\}$. This set could also be referred to as the set of "desired actions".
2. Given a set of possible intended interpretations, $\{\mathbf{I_1}, \mathbf{I_2}\ldots\mathbf{I_n}\}$, generate a probability distribution weighing the likelihood of each element.

Hand-coding rule-based solutions to either of these problems is not scalable for non-conventional indirect requests. Plan-based solutions, such as the one discussed in Perrault and Allen (1980), are better-suited, though still require generalizable rules to use for planning.

Below, we propose a generalizable solution to (2) grounded in experimental work from psycholinguistics. We briefly discuss the trade-offs and challenges of implementing such a solution and compare it to other proposals.

We do not propose a solution to (1), though we do address the challenges associated with (1) in both our proposed implementation and the Future Work section. We believe it is still worthwhile to address (2) in the absence of a solution for (1), as it takes us one step closer to a comprehensive solution for understanding non-conventional indirect requests.

Current Work

Given that some utterance *could* be a request, what are the contexts delineating whether or not it is? How do we weigh the likelihood of possible interpretations? The ideal solution is a *generalizable inference rule* about how to interpret pragmatically ambiguous utterances.

Proposal

One theory for how people make pragmatic inferences during conversation is the *mutual knowledge hypothesis* (Gibbs, 1987): interlocutors build up shared mental representations (e.g. linguistic common ground, shared experiences, etc.), which they use to perform inference.

We can extend this hypothesis into a general rule for understanding indirect requests: *An utterance that is a possible indirect request has an increased likelihood of being a request if the listener, **L**, has reason to the believe that the speaker, **S**, believes that **L** has the ability and/or obligation to fulfill the request, and that the request is not already being fulfilled.*

The example in (1) above fits into this rule. If **L** believes that **S** believes that **L** is subordinate to **S**, it's a request for the med-kit; otherwise, it's a request to know where a med-kit can be found. The rule also generalizes across many other contexts. Consider the sentences:

(2) *It's cold in here*.

(3) *This kitchen is a pigsty*.

The utterance in (2) could be interpreted as a request to turn on the heater, if **L** believes that **S** believes that **L** has the ability and/or obligation to turn it on (e.g. the heater is not broken). The utterance in (3) could be interpreted as a request to clean the kitchen, particularly if **L** believes that **S** believes that **L** is responsible for the mess.

Crucially, this rule requires that **L** maintain a mental model of **S**'s beliefs. That is, whether or not **L** actually is able or obligated to fulfill the request, **L** should be more likely to interpret the utterance as a request if **S** believes **L** is able.

Proposed Implementation

Implementing our proposed solution in an actual language understanding system also requires an existing solution to (1). Thus, in the proposed implementation below, we assume that the system in question has a way of generating $\{\mathbf{I_1}, \mathbf{I_2} \ldots \mathbf{I_n}\}$ from some utterance **U**.

Given the set $\{\mathbf{I_1}, \mathbf{I_2} \ldots \mathbf{I_n}\}$, the robot could weigh the likelihood of different interpretations using a model of what the speaker knows. One approach to mental modeling in HRI is the use of separate ontologies for different interlocutors, which allows the robot to represent divergent beliefs (Lemaignan et al, 2010; Lemaignan et al, 2017). A robot endowed with such a mechanism could simulate the logical inferences that result from different knowledge states, and hence disambiguate between possible interpretations of an utterance.

Consider the example in (2) above. We are assuming that the robot already has a mapping from **U**, *It's cold in here*, to at least two possible interpretations and their corresponding actions:

(a) $\mathbf{I_1}$ --> *Complaint* --> *Offer sympathy*
(b) $\mathbf{I_2}$ --> *Request* --> *Turn on heater*

Of course, mapping **U** to $\{\mathbf{I_1}, \mathbf{I_2}\}$ above is nontrivially difficult, but here, we focus on how a robot could select the best interpretation – e.g. the one the speaker probably intended.

Even if the robot tracks its interlocutor's beliefs about "who is subordinate to whom", as discussed in Williams et al (2015), the possibility of misinterpretation can still arise if the speaker and the robot have world models that are misaligned in other respects, such as what they know to be true about their environment.

For example, the robot might know that the heater is broken and represent it accordingly: *<heater broken TRUE>*. If the robot assumes that the speaker knows this too, the correct interpretation would be $\mathbf{I_1}$ – after all, why would someone request that you turn on the heater if they know the heater is broken? It is possible, however, that the speaker believes the heater is functional: *<heater broken FALSE>*. In this case, the speaker might still be intending to make a request. If the robot represents the speaker's divergent belief, it can select the best interpretation according to the *speaker's* knowledge state: $\mathbf{I_2}$.

In essence, then, we are proposing that robots track and represent the belief states of their interlocutors. The robot can then use its model of the *speaker's* model of the world to infer the intended meaning.

Advantages and Disadvantages of This Approach

This solution is in a way more parsimonious than writing separate pragmatic rules for each context. Only a single rule is encoded, which can be applied to many situations. Instead of relying on hand-coded rules about how different contexts constrain the likelihood of different interpretations, **L** selects the best interpretation based on what they think **S** knows.

But simplifying pragmatic rules comes at the cost of requiring **L** to maintain extensive world knowledge, both for their own world model and for representing the world models of any possible interlocutor. Maintaining divergent world models could prove to be computationally expensive; equally difficult is the task of inferring what **S** knows in the first place, and which pieces of knowledge are relevant to track for computing pragmatic interpretations in the future.

However, we argue that these drawbacks are outweighed by the advantage of a more generalizable rule. Additionally, some of the mechanisms required to implement the rule, such as Theory of Mind, are already being researched and developed by the HRI community (Lemaignan et al, 2010; Talamadupula et al, 2014; Devin and Alami, 2016; Lemaignan et al, 2017). Mental models will also be necessary for tasks other than simply understanding indirect requests; as Lemaignan et al (2017) point out, mechanisms like Theory of Mind "are mandated" for the facilitation of naturalistic and successful human-robot interactions, and in particular, human-robot collaborations.

Experimental Evidence

While the cognitive mechanisms for interpreting indirect requests are still under investigation, there is evidence suggesting that a human L makes inferences about S's intentions using mental models of S's belief states.

Early work by Gibbs (1980; 1986; 1987) demonstrated the importance of *context* in producing and interpreting indirect requests, and laid the theoretical groundwork for the notion that interlocutors perform inference through coordinated mental representations. Around the same time, clinical research found a relationship between deficits in Theory of Mind, the ability to reason about other mental states (Wimmer and Perfner, 1983), with deficits in pragmatic reasoning (Hirst, LeDoux, and Stein, 1984).

More recently, fMRI studies by Ackeren et al (2012; 2016) found a correlation between understanding indirect requests and increase blood flow to brain regions associated with Theory of Mind. While this evidence does not prove a causal or mechanistic role for Theory of Mind in understanding indirect requests, it does suggest a relationship between mental modeling and understanding pragmatically ambiguous utterances.

Ultimately, however, more research is needed to determine the extent to which people use mental models of their interlocutor for pragmatic reasoning, and the actual role that these models serve for pragmatic inference.

Future Work

We are currently investigating whether people use mental models of their interlocutor for understanding indirect requests. This will shed additional light on the mechanisms underlying pragmatic inference in humans.

Future work in HRI and language understanding could attempt to integrate our proposed solution into existing architectures. As mentioned in the paper, we have only proposed an implementation for the problem of adjudicating between different interpretations of an ambiguous utterance. More work is needed to uncover the mechanisms whereby people generate this set of possible interpretations from a speech act. Hand-coding these interpretations is not scalable. It is possible that mental models could address this problem as well, but a robot would also need to have general world knowledge to construct novel solutions to a problem posed in language.

More generally, models for pragmatic inference could look to research on how extralinguistic information (e.g. prosody, gesture, gaze) is used for comprehension. For example, previous work (Kelly et al, 1999; Kelly, 2001) suggests that speakers indicate they are making a request through gestures and gaze patterns, such as looking at the window when they say *A breeze is getting in*. These gaze patterns could help with constructing the set of possible interpretations by triangulating the target of the speaker's gaze. Other research suggests that speakers sometimes gaze towards the *listener* as a form of "mobilizing a response" (Stivers and Rossano, 2010), which could help disambiguate between speakers making a request or simply complaining.

References

Briggs, G. M., & Scheutz, M. 2013. A Hybrid Architectural Approach to Understanding and Appropriately Generating Indirect Speech Acts. In *AAAI*.

Briggs, G., & Scheutz, M. 2016. The Pragmatic Social Robot: Toward Socially-Sensitive Utterance Generation in Human-Robot Interactions. In *AAAI Fall Symposium Series: Artificial Intelligence for Human-Robot Interaction* (pp. 12-15).

Devin, S., & Alami, R. 2016. An implemented theory of mind to improve human-robot shared plans execution. In *Human-Robot Interaction (HRI), 2016 11th ACM/IEEE International Conference on* (pp. 319-326). IEEE.

Gibbs, R. W. 1981. Your wish is my command: Convention and context in interpreting indirect requests. *Journal of Verbal Learning and Verbal Behavior*, *20*(4), 431-444.

Gibbs, R. W. 1986. What makes some indirect speech acts conventional?. *Journal of memory and language*, *25*(2), 181-196.

Gibbs, R. W. 1987. Mutual knowledge and the psychology of conversational inference. *Journal of Pragmatics*, *11*(5), 561-588.

Hirst, W., LeDoux, J., & Stein, S. 1984. Constraints on the processing of indirect speech acts: Evidence from aphasiology. *Brain and language*, *23*(1), 26-33.

Lemaignan, S., Warnier, M., Sisbot, E. A., Clodic, A., & Alami, R. 2017. Artificial cognition for social human–robot interaction: An implementation. *Artificial Intelligence*, *247*, 45-69.

Lemaignan, S., Ros, R., Mösenlechner, L., Alami, R., & Beetz, M. 2010. ORO, a knowledge management platform for cognitive architectures in robotics. In *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on* (pp. 3548-3553). IEEE.

Perrault, C. R., & Allen, J. F. 1980. A plan-based analysis of indirect speech acts. *Computational Linguistics*, *6*(3-4), 167-182.

Searle, J. R. 1975. *Indirect speech acts* (pp. 59-82)..

Stivers, T., & Rossano, F. (2010). Mobilizing response. *Research on Language and social interaction*, *43*(1), 3-31.

Trott, S., Eppe, M., & Feldman, J. 2016. Recognizing Intention from Natural Language : Clarification Dialog and Construction Grammar. *Workshop on Communicating Intentions in Human-Robot Interaction*.

Van Ackeren, M. J., Casasanto, D., Bekkering, H., Hagoort, P., & Rueschemeyer, S. A. 2012. Pragmatics in action: indirect requests engage theory of mind areas and the cortical motor network. *Journal of Cognitive Neuroscience*, *24*(11), 2237-2247.

Van Ackeren, M. J., Smaragdi, A., & Rueschemeyer, S. A. 2016. Neuronal interactions between mentalising and action systems during indirect request processing. *Social cognitive and affective neuroscience*, *11*(9), 1402-1410.

Williams, T., Briggs, G., Oosterveld, B., & Scheutz, M. 2015. Going Beyond Literal Command-Based Instructions: Extending Robotic Natural Language Interaction Capabilities. In *AAAI* (pp. 1387-1393).

Williams, T., Núñez, R. C., Briggs, G., Scheutz, M., Premaratne, K., & Murthi, M. N. 2014. A Dempster-Shafer theoretic approach to understanding indirect speech acts. In *Ibero-American Conference on Artificial Intelligence* (pp. 141-153). Springer, Cham.

Wilske, S., & Kruijff, G. J. 2006. Service robots dealing with indirect speech acts. In *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on* (pp. 4698-4703). IEEE.

Wimmer, H., & Perner, J. 1983. Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103-128.