Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16)

Architectural Mechanisms for Situated Natural Language Understanding in Uncertain and Open Worlds

Tom Williams
Human Robot Interaction Laboratory
Tufts University, Medford, MA, USA

Introduction

As natural language capable robots and other agents become more commonplace, the ability for these agents to understand truly natural human speech is becoming increasingly important. What is more, these agents must be able to understand truly natural human speech in realistic scenarios, in which an agent may not have full certainty in its knowledge of its environment, and in which an agent may not have full knowledge of the entities contained in its environment. As such, I am interested in developing architectural mechanisms which will allow robots to understand natural language in uncertain and open-worlds. My work towards this goal has primarily focused on two problems: (1) reference resolution, and (2) pragmatic reasoning. In the following sections, I will introduce and discuss the work of myself and others on these problems.

Reference Resolution

Reference resolution is the problem of identifying the entities referenced in a natural language utterance. For example in the utterance "The medkit is in the room at the end of the hall", an agent must determine what entities are being referred to by "the medkit", "the room", and "the hall". There are two main drawbacks common to most previous approaches to this problem. First, the majority of these approaches are *domain-dependent*: they are specifically targeted toward resolving one kind of reference, such as references to locations (i.e., *spatial reference resolution*). Second, the majority of these approaches operate in a *closed world*: they assume that all entities which could be referenced are already known. Matuszek et al. use Statistical Machine Translation techniques to translate natural language directions into routes through an open world. However, this approach is limited to resolving spatial references, and because utterances are translated into actions rather than world model modifications, their approach is unable to discuss or reason about new locations without first visiting them (Matuszek et al. 2012). Fasola and Mataric presented *Semantic Fields*, an approach that represents spatial relations as probability density functions over locations in a known metric map. This approach is restricted to *spatial* reference resolution and is limited to operation in fully known environments (Fasola and Mataric 2013). Chai et al. present a greedy algorithm which uses a subset of the Givenness Hierarchy to resolve a wide array of referential expressions, but this approach operates under a closed-world assumption (Chai, Prasov, and Qu 2006). Kollar, Tellex et al. present *Generalized Grounding Graphs*, which instantiate probabilistic graphical models based on the structure of incoming NL utterances, and use those models to resolve references (Tellex et al. 2011). However, this approach is limited to the domain it is trained on, and like Matuszek et al., does not allow an agent to discuss or reason about new entities without first experiencing them.

In (Williams et al. 2013), I presented an algorithm for resolving references to unknown locations by making world model modifications based on utterance semantics, and an algorithm for generating plans to unknown locations based on their properties. These algorithms, however, were domain-dependent, and did not handle uncertainty. To address these shortcomings, my Probabilistic, Open-World Entity Resolution algorithm POWER (Williams and Scheutz 2015) performs reference resolution in uncertain and open worlds, and is not tied to a particular domain, instead interfacing with a domain-specific knowledge base (KB) through a *consultant* capable of providing a list of resolution candidates, assessing the probabilities that certain properties hold for certain entities, and making KB modifications. Recently, I extended POWER in two significant ways. First, I adapted POWER to use *multiple* consultants associated with separate knowledge bases (Williams and Scheutz 2016). This allows POWER to perform resolution using multiple KBs that may be located on different machines, may use different knowledge representation schemes, and may have different methods for accessing and modifying data. Second, I devised what I believe to be the first *complete* implementation of the Givenness Hierarchy. My implementation substantially improves on the implementation presented by Chai et al. in several ways, most significantly by using POWER to handle uncertain and open worlds (Williams et al. 2015b).

Pragmatic Reasoning

Humans frequently use *indirect* utterances whose *literal* meanings do not match their *intended* meanings. "Do you know what time it is?" is likely not *literally* a question about whether or not you know the time, but more likely an *indirect request* for the time, using indirectness as a politeness strategy; "The commander is wounded and needs a medical kit!" is likely not a simple statement of fact, but is more likely an *indirect command* to bring the wounded commander a medical kit, or perhaps a request to know where to find a medical kit. The few language-capable robot architectures able to handle these types of non-literal utterances (known as *indirect speech acts* (ISAs)) (e.g., (Wilske and Kruijff 2006), and our own architecture, (Briggs and Scheutz 2013)) are not robust to uncertain contexts.

To address this problem, my algorithms for understanding and generating ISAs (Williams et al. 2015a), use a Dempster-Shafer (DS) theoretic representation to encode the uncertainty of (1) incoming utterances, (2) the robot's knowledge, and (3) the pragmatic rules used by the robot to map an utterance and context to an intention and to to abduce the best utterance to use to communicate an intention. Using a DS-theoretic representation allows the robot to assess its own *ignorance*, and thus to acknowledge when it does not have enough information to understand an utterance.

Research Plan and Contribution

As an extension to my previous work on open-world reference resolution, I am designing mechanisms to allow a robot to perform inference using information contained in distributed, heterogeneous KBs. As an extension to my previous work on pragmatic reasoning, I am designing one-shot learning algorithms for learning new pragmatic rules. I expect to perform research on the following problems between AAAI 2016 and my expected graduation date of May 2017: As an extension of my work on open-world reference resolution, I will: (1) examine the effect of different variable-ordering strategies in the POWER algorithm, (2) examine the effects of different probability thresholds in the POWER algorithm, (3) convert the POWER algorithm to use a DS-theoretic knowledge representation scheme, in order to allow the robot to reason about its own ignorance when performing reference resolution, (4) integrate a consultant with semantic mapping capabilities into our architecture, (5) examine different ways of calculating cross-modality salience scores for use in my Givenness Hierarchy implementation, and (6) integrate common-sense and affordance-based reasoning capabilities into my Givenness Hierarchy implementation. As an extension of my work on pragmatic reasoning, I will develop mechanisms for altering the context and confidence of previously learned pragmatic rules. Finally, I will perform a comprehensive, extrinsic evaluation of my suite of architectural mechanisms on a physical robot, and examine human perceptions of the capabilities afforded by these mechanisms.

The contribution of this dissertation will thus be a set of architectural mechanisms for natural language understanding and generation in uncertain and open worlds, extending the state of the art on a variety of problems important for natural human-robot dialogue.

Acknowledgments

My work has been supported in part by ONR grants #N00014-11-1-0289, #N00014-11-1-0493, #N00014-10-1-0140 #N00014-14-1-0144, #N00014-14-1-0149, #N00014-14-1-0751*l*, and NSF grants #1111323, #1038257.

References

Briggs, G., and Scheutz, M. 2013. A hybrid architectural approach to understanding and appropriately generating indirect speech acts. In *Proceedings of the 27th AAAI Conf. on Artificial Intelligence*.

Chai, J. Y.; Prasov, Z.; and Qu, S. 2006. Cognitive principles in Robust multimodal interpretation. *Journal of Artificial Intelligence Research*.

Fasola, J., and Mataric, M. J. 2013. Using semantic fields to model dynamic spatial relations in a robot architecture for natural language instruction of service robots. In *IEEE/RSJ Int'l Conference on Intelligent Robots and Systems*.

Matuszek, C.; Herbst, E.; Zettlemoyer, L.; and Fox, D. 2012. Learning to parse natural language commands to a robot control system. In *Proceedings of the 13th International Symposium on Experimental Robotics*.

Tellex, S.; Kollar, T.; Dickerson, S.; Walter, M.; Banerjee, A.; Teller, S.; and Roy, N. 2011. Approaching the symbol grounding problem with probabilistic graphical models. *AI magazine*.

Williams, T., and Scheutz, M. 2015. POWER: A domain-independent algorithm for probabilistic, open-world entity resolution. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*.

Williams, T., and Scheutz, M. 2016. A framework for resolving open-world referential expressions in distributed heterogeneous knowledge bases. In *Proc. of the Thirtieth AAAI Conference on Artificial Intelligence*.

Williams, T.; Cantrell, R.; Briggs, G.; Schermerhorn, P.; and Scheutz, M. 2013. Grounding natural language references to unvisited and hypothetical locations. In *Proceedings of the 27th AAAI Conference on Artificial Intelligence*.

Williams, T.; Briggs, G.; Oosterveld, B.; and Scheutz, M. 2015a. Going beyond command-based instructions: Extending robotic natural language interaction capabilities. In *Proceedings of Twenty-Ninth AAAI Conference on Artificial Intelligence*.

Williams, T.; Schreitter, S.; Acharya, S.; and Scheutz, M. 2015b. Towards situated open world reference resolution. In *AAAI Fall Symposium on AI for HRI*.

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