Extralinguistic State Localization in Service of Turn Generation in Task-Oriented Dialogue

Petr Babkin
Cognitive Science Department / 110 8th St
Rensselaer Polytechnic Institute / Troy, NY 12180
babkip@rpi.edu

Abstract

The tremendous role of context in understanding natural language dialogue has been amply emphasized in the literature. Alas, in much research to date, context is defined simply as preceding linguistic material within some window. In real life, however, linguistic content amounts to only a fraction of contextual information that helps humans to act appropriately in a conversation. In fact, in some cases it is non-linguistic cues that are most informative e.g., in certain stereotypical situations such as the famous restaurant script. This study, explores the notion of context as latent extralinguistic state underlying task-oriented dialogue. This view is put to test of deriving a coherent task-relevant dialogue turns in the face of arbitrarily ablated input. The paper outlines the approach to be presented as a poster along with preliminary results.

1 Introduction

It is no secret that a better informed decision is bound to lead to a better outcome. When it comes to natural language processing, effective feature engineering is often attributed much success, in many cases, offsetting the merit of the actual algorithms that utilize them. With the growing complexity of NLP tasks, the heuristics, too, become more sophisticated, capturing the decision's increasing dependence on the context in which the input is observed. Alas, in much of NLP research, context is limited to features that are directly available from linguistic input. Such reliance on a single source of heuristics assumes error-free input, which is not always the case[1]. This study explores the capabilities of extra-textual heuristics by artificially encouraging the exploitation of pragmatic context over the observed input, through the use of ablation. The hypothesis is the more ablated the input[2] — the more the system has to rely on pragmatic reasoning to compensate for the deficiency. In other words, a good sense of the situation may enable one to come up with a correct answer without necessarily understanding the question.

2 Domain and representation

Task-oriented dialogue appears to be a promising domain, being a rich source of goal-based heuristics that could support pragmatic reasoning. Specifically, the Cards corpus, with its clear goal structure and highly goal-oriented linguistic content, appears well suited for modeling of this sort (Potts, 2012). Analysis of a sample of dialogues from the corpus revealed that each speech act is not simply conditioned on its preceding utterances *per se* but also depends upon a) a persistent extralinguistic state that is maintained via speech acts, and b) goal-directed implications of this state. Therefore, the problem of response generation is preconditioned on the following two subproblems:

- inference of the current extralinguistic state from the observed language input,
- selection of the desirable state and generation of the state-inducing linguistic output.

In order to capture the goal dynamics underlying language communication (herein hypothesized to be necessary for handling imperfect input), a variant of the state-space representation was used with two modifications. First, the standard definition of states as sets of domain-specific predicates and truth values was replaced with a more abstract notion of states of common ground (CG)

[1] Error propagation is one good analogy of a system's fragility because of the unrealistic expectations about the quality of upstream information (Caselli and Postma, 2015).

[2] To isolate the effects of ablation and bypass issues unrelated to it, semantic meaning representations were used as input rather than raw text.

Appears in *JerSem: The 20th Workshop on the Semantics and Pragmatics of Dialogue,* edited by Julie Hunter, Mandy Simons and Matthew Stone. New Brunswick, NJ USA, July 16-18, 2016, pages 108-109.

(Clark, 2006) — points in time when certain facts become shared knowledge (e.g., through verbalization by one of the agents). Second, domain-specific actions are assumed latent and state transitions are modeled solely as knowledge dependencies among the states. The choice of this proxy representation can be justified by a) the primary need for heuristics to aid language understanding rather than to help actually solve an associated planning problem and b) such a representation can be derived from the language material irrespective of the subject domain[3].

3 Model

For the reasons of space, much detail is omitted from the model description and the purpose of this section is limited to providing a high-level overview.

As noted in the previous section, the agent's choice of a response depends not directly on the input speech but on the unobservable state, which, in turn, is conditioned on both the observed input and the previous state:

$$o_{t+1} \leftarrow f(s_t|o_t, s_{t-1}) \quad (1)$$

This naturally brings us to hidden state models, such as HMM, where emitted states correspond to observed utterances and the hidden states are the underlying situations[4]. Such a model of course needs to be extended for there is not a one-to-one correspondence among observable and hidden states. For example, in Figure 1, the predicate $cur\text{-}cards(a,\ 5s)$[5] is added to the common ground as a result of an exchange between the conversants rather than a single utterance. While this compositionality could be modeled as a joint distribution, it appears reasonable to employ a feature-based grammar instead. A binary result returned by the corresponding recognizer effectively replaces the coefficient from the emission matrix in the state probability equation. In order to account for ablation, this value also needs to be weighed based on the likelihood of the ablation instance. The data

$$\{cur\text{-}cards(a,\ 5s)\} \left\langle \begin{array}{l} \frac{B\text{: } you\ have\ 5\ spades}{\mathbf{suggest}(b,\ cur{-}cards(a,\ 5s))} \\ \\ \frac{A\text{: } yep}{\mathbf{acknowledge}(a)} \end{array} \right.$$

Figure 1: A fact grounding adjacency pair.

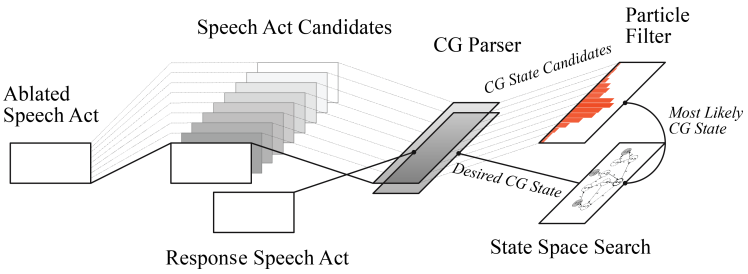

Figure 2: Model architecture.

flow in the model is summarized by the following stages.

1. For the observed speech act, alternatives increasingly dissimilar to the original are generated by enumerating feature values up until a set cutoff probability threshold.
2. The candidates are then mapped to their corresponding states (if any) by the CG parser/recognizer.
3. The resulting CG states along with their weights are passed to the particle filter, which outputs the belief distribution over CG states.
4. The most likely current CG state is used to compute the desirable CG state via breadth-first search.
5. The grammar is used again to induce a speech act that would expand the desired CG state.

[3] which in turn opens up an exciting possibility of generating models of novel domains automatically cf. (Kasch and Oates, 2010)

[4] It is important to note that hidden states in this case are not true values of observed inputs (emissions) common for noisy channel model, but extralinguistic states comprised of domain predicates.

[5] For the sake of brevity, the notation as in (Langley et al., 2014) is used for speech acts and domain predicates.

References

Tommaso Caselli and Marten Postma. 2015. When it's all piling up: investigating error propagation in an NLP pipeline. In *WNACP 2015*, Passau, Germany.

Herbert Clark. 2006. Context and Common Ground. *Concise Encyclopedia of Philosophy of Language*, pages 105–108.

Niels Kasch and Tim Oates. 2010. Mining script-like structures from the web. In *Proceedings of the First International Workshop on Formalisms and Methodology for Learning by Reading*, pages 34–42, Stroudsburg, PA.

Pat Langley, Ben Meadows, Alfredo Gabaldon, and Richard Heald. 2014. Abductive understanding of dialogues about joint activities. *Interaction Studies*, 15(3):426–454.

Christopher Potts. 2012. Goal-driven answers in the Cards dialogue corpus. In *Proceedings of the 30th West Coast Conference on Formal Linguistics*, Somerville, MA.