

# Reinforcement Learning for Improved Low Resource Dialogue Generation

**Ana V. González-Garduño**

University of Copenhagen, Department of Computer Science  
ana@di.ku.dk

## Introduction

Dialogue systems have become a very popular topic in recent years with the rapid improvement of personal assistants like Cortana, Siri, Alexa, among many others. Open ended conversational systems tend to work better for languages like English or Spanish, where a great amount of text and speech data is available. As the problem of data sparsity is one of the many barriers towards creating robust and accurate dialogue systems, conversational agents in low resource languages are not nearly as developed. In addition, goal oriented dialogue systems pose a different kind of challenge. These systems require tracking of states during the course of a conversation in order to make the next action. Generally, these two types of conversational systems have been approached separately which has led to improvements in both directions, however recent work has attempted to create systems that try to bridge the gap between the two (Weston, Dinan, and Miller 2018). In this thesis, I focus on language independent methods of improving utterance understanding and response generation and attempt to tackle some of the issues surrounding current systems. The aim is to create a unified approach to dialogue generation inspired by developments in both goal oriented and open ended dialogue systems. The main contributions in this thesis are: 1) Introducing hybrid approaches to dialogue generation using retrieval and encoder-decoder architectures to produce fluent but precise utterances in dialogues, 2) Proposing supervised, semi-supervised and Reinforcement Learning methods for domain adaptation in goal oriented dialogue and 3) Introducing models that can adapt cross lingually.

## Related Work

**Hierarchical Encoder-Decoder Models** Neural generation models such as Machine translation (MT) approaches (Cho et al. 2014) to dialogue generation and question-answering systems became of interest because of their ability to encode an utterance or question into a fixed sized vector representation, feed it into a decoder and decode it into a variable length sequence. These models are trained to maximize the conditional probability of a target sequence given a source sequence. However, MT based approaches like this

lack the ability to encode the context in which the current utterance is occurring in. This can lead to repetitive and meaningless responses, therefore it is important to encode the long term context of a conversation in order to make a more accurate prediction.

This has led to work that extends simple encoder-decoder models to deal with generation of larger structured texts such as paragraphs and documents (Li, Luong, and Jurafsky 2015; Serban et al. 2016; 2017). Many of these models work by encoding information at multiple levels i.e. a context encoder, an utterance encoder and a decoder that predicts the next utterance in a conversation.

**Dialogue State Tracking** Dialogue state tracking (DST), which consists of tracking the user's goal at each turn of the conversation, has seen rapid developments since the Dialog State Tracking Challenge (Henderson, Thomson, and Williams 2014). In DST, the goal of the user is typically represented using slot-value pairs, which are typically fixed. This limits the way that many models can be used in a practical setting. Some newer models have looked at how to dynamically add slots into DST models however, most works have focused on tracking states for single domains (Ren et al. 2018; Zhong, Xiong, and Socher 2018).

## Current status of the project

**Hybrid Hierarchical Dialogue Generation model** One problem with end-to-end open domain dialogue systems is the fact that these tend to generate fluent but meaningless responses. On the other hand, retrieval based methods to question-answering tend to return snippets with precise information that do not really follow a flow of conversation.

Early in the project, I experimented with simple baselines for relevancy ranking that are language independent. The model presented in Gonzalez, Augenstein, and Søgaard (2018) is a very simple multi-layer perceptron model that has a low computational cost compared to other more expressive recurrent models (Bowman et al. ) due to the lower number of parameters. The first attempt at investigating the grey area between open ended chitchat and goal oriented dialog has been by combining the benefits of a hierarchical encoder-decoder model (Serban et al. 2017) with exemplar-based retrieval. More specifically, we condition our decoder on the answer of the most similar dialogue

appearing on the training set in order to mimic the scenario of solving a problem based on past experience. Given the current utterance, we use the retrieval model in order to obtain the most related example answer, encode it and feed into the decoder, along with the hidden states of the context and utterance encoders. The ranking and dialogue models are implemented and currently we have results that suggest this method yields responses that are more accurate however, more analysis and human evaluation is needed and is expected to be done by early December.

### **Reinforcement Learning for Domain Adaptation in DST**

As mentioned earlier, much of the work on DST has focused on single domains. One of the recent works that have attempted to learn information across domains is Ramadan, Budzianowski, and Gasic (2018), in which they jointly identify domain and track the states for that domain, as well as using semantic similarity between the slots in order to allow the model to share parameters across domains. Currently, we are using this model to train on N number of domains and adapt this model to new domains for which we don't have turn level supervision using Reinforcement Learning. More specifically, we incorporate a dialogue level reward with policy gradients (using the joint goal accuracy) in order to adapt to the new domain. This work is expected to be done by mid December.

### **Future Work**

#### **Ellipsis Resolution, Adversarial, Transfer Learning**

Another problem in current dialogue systems -both open ended chat and goal oriented dialogue- is the fact that many models fail to recognize whether an utterance in the dialogue refers to a previous one and the interactions between entities across turns. In the next year, I will be attempting to tackle this problem by informing end to end models about ellipsis resolution. Introducing this bias will also allow me to create models that are generally more robust and could potentially adapt to other languages. In addition, starting in February 2019, the plan is to incorporate Adversarial Learning and Reinforcement learning into the dialogue generation models. The idea is that when training a dialogue generation model on human conversation logs, a discriminative model will be trained to distinguish between the human and machine generated. In addition, deriving from the findings of Li et al. (2017) who use reinforcement learning in order to improve dialogue generation, I will experiment with ways of incorporating experiments using customer feedback signals in a reinforcement learning setup in order to improve the performance of goal oriented models in languages such as Danish, Norwegian, Swedish and Icelandic. For the last 1.5 years of my project, I aim to focus on introducing new techniques for transfer learning and multi-task learning with the goal of improving dialogue generation in low resource languages as well as low resource domains.

### **References**

Bowman, S. R.; Angeli, G.; Potts, C.; and Manning, C. D. A large annotated corpus for learning natural language inference.

Cho, K.; van Merriënboer, B.; Bahdanau, D.; and Bengio, Y. 2014. On the properties of neural machine translation: Encoder-decoder approaches. *Syntax, Semantics and Structure in Statistical Translation* 103.

Gonzalez, A.; Augenstein, I.; and Søgaard, A. 2018. A strong baseline for question relevancy ranking. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 4810–4815.

Henderson, M.; Thomson, B.; and Williams, J. D. 2014. The second dialog state tracking challenge. In *Proceedings of the 15th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*, 263–272.

Li, J.; Monroe, W.; Shi, T.; Jean, S.; Ritter, A.; and Jurafsky, D. 2017. Adversarial learning for neural dialogue generation. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2157–2169.

Li, J.; Luong, T.; and Jurafsky, D. 2015. A hierarchical neural autoencoder for paragraphs and documents. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, volume 1, 1106–1115.

Ramadan, O.; Budzianowski, P.; and Gasic, M. 2018. Large-scale multi-domain belief tracking with knowledge sharing. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, volume 2, 432–437.

Ren, L.; Xie, K.; Chen, L.; and Yu, K. 2018. Towards universal dialogue state tracking. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 2780–2786.

Serban, I. V.; Sordoni, A.; Bengio, Y.; Courville, A. C.; and Pineau, J. 2016. Building end-to-end dialogue systems using generative hierarchical neural network models. In *AAAI*, volume 16, 3776–3784.

Serban, I. V.; Sordoni, A.; Lowe, R.; Charlin, L.; Pineau, J.; Courville, A. C.; and Bengio, Y. 2017. A hierarchical latent variable encoder-decoder model for generating dialogues. In *AAAI*, 3295–3301.

Weston, J.; Dinan, E.; and Miller, A. 2018. Retrieve and refine: Improved sequence generation models for dialogue. In *Proceedings of the 2018 EMNLP Workshop SCAI: The 2nd International Workshop on Search-Oriented Conversational AI*, 87–92.

Zhong, V.; Xiong, C.; and Socher, R. 2018. Global-locally self-attentive encoder for dialogue state tracking. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, 1458–1467.