Hierarchical Methods for a Unified Approach to Discourse, Domain, and Style in Neural Conversational Models

João Sedoc
Department of Computer and Information Science
University of Pennsylvania
Philadelphia, PA
joao@cis.upenn.edu

Introduction

With the advent of personal assistants such as Siri and Alexa, there has been a renewed focus on dialog systems, specifically open domain conversational agents. Dialog is a challenging problem since it spans multiple conversational turns. To further complicate the problem, there are many contextual cues and valid possible utterances. Dialog is fundamentally a multiscale process given that context is carried from previous utterances in the conversation; however, current neural methods lack the ability to carry human-like conversation. Neural dialog models, which are based on recurrent neural network Encoder-Decoder sequence-to-sequence (Seq2Seq) models (Sutskever, Vinyals, and Le 2014; Bahdanau, Cho, and Bengio 2015), lack the ability to create temporal and stylistic coherence in conversations.

Hierarchical systems are natural forms of abstraction that can be used to capture natural language hierarchy. There are words → sentences → paragraphs → documents, and more specifically in dialog words → utterances (conversational turns) → conversations → interactions. Hierarchical models for such systems are well studied dating back to the work of El Hihi and Bengio (1995). These hierarchical models have also been incorporated into dialog (Serban et al. 2017) for carrying context across multiple turns. Other common approaches for maintaining coherence incorporate topic models or sentence clusters (Xing et al. 2017; Cheng, Fang, and Ostendorf 2017) as well as persona (Li et al. 2016).

This thesis incorporates dialog acts (such as *Statement-non-opinion* [”Me, I’m in the legal department.”], *Acknowledge* [“Uh-huh.”]) and discourse connectives (e.g. “because”, “then”), utterance clustering and domain prediction (for instance conversations about “politics”), and style shifting (i.e. depressed vs happy tone) using hierarchical methods. Our approach is unique as it focuses on the utterance-level discourse of neural dialog models, and thus only requires a statistic that captures word level representations. As shown in Figure 1, while focusing on utterance level ($h^{(2)}$), these methods directly alter representations both at conversational level and word level responses.

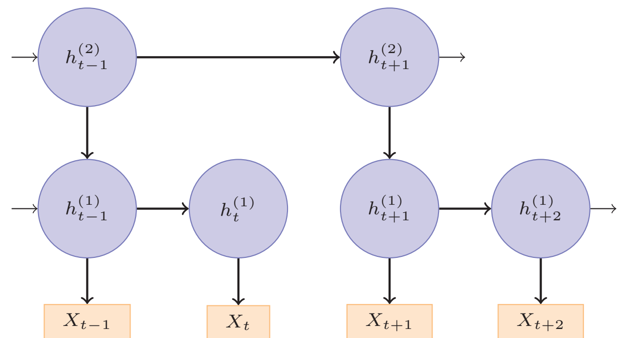

Figure 1: An example two state processes where x_i is a word vector, and $h_i^{(2)}$ is the higher level representation.

Methods

While our approach is intended to unify utterance domain (or cluster type), utterance style, and dialog acts / discourse connectives , we divide our overall method into three components: 1) factored responses, 2) conversational tree and 3) style shifting.

1. Factored Response Prediction While domain is often ill-defined, our intent is for domain to capture the topic or theme of the conversation utterances, which often change during a conversation. Factored response models generate multiple intelligent responses in a conversation, as shown in Figure 2.

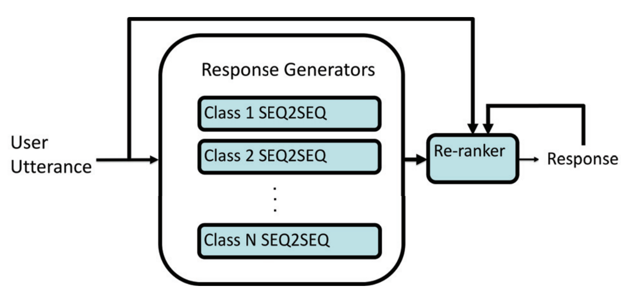

Figure 2: DOM-Seq2Seq model which selects the response from a particular class of concurrent Seq2Seq models.

To achieve this, we presented *DOM-Seq2Seq*, a domain aware neural network model based on the novel technique of using domain-targeted Seq2Seq models and a domain classifier. In DOM-Seq2Seq, the higher level of the hierarchy, the re-ranker, selects the Seq2Seq lower level response. The model captured features from current utterance and domains of the previous utterances to facilitate the formation of relevant responses.

In our previous work, we used labeled topics from Reddit (Choudhary et al. 2017); however, we will also use factored models similar to Cheng, Fang, and Ostendorf (2017) where utterances are clustered by the similarity between their representations.

2. Coherent Conversational Tree Our coherent conversational tree uses both a multilayer hidden Markov model as well as multiresolution Seq2Seq model. This model attempts to capture both dialog acts as well as discourse connectives, such as *by the way*. For such higher order language tasks, we use spectral HMM and deep learning to efficiently capture sentence compositionality using both the words in each sentence and the sentence context. We further extend this model by not only using the compressed representation from the word level, but also by capturing the change in the sentence compositionality using tensor factorization as shown in Figure 1.

Using dialog act prediction for generative neural conversation models (Tran, Haffari, and Zukerman 2017) has also only recently been examined. Our model is close to the recent work of Cheng, Fang, and Ostendorf (2017); however, their model is not generative. In Figure 3 we show how our model further extends to condition the response on the dialog act.

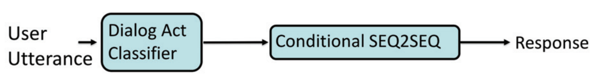

Figure 3: The conversational tree model conditions the Seq2Seq model based on the prediction of the next dialog act in the conversation.

3. Response Style Shifting Aside from domain or conversational-type coherence, personality and emotions also play a vital role in defining how humans interact with each other. Adding a specific persona is important to approach more natural conversation. A dialog system's linguistic style can enhance its perceived personality which can be adapted to different social situations.

In our previous work (Jena et al. 2017), we propose a design for a dialog agent that captures a "style" by incorporating references along with peculiar tones of fictional characters therein. We use two Seq2Seq models: one trained for specific dialog style and the other for general style. The model uses a modified implementation of the Word Graph algorithm (Filippova 2010; Banerjee, Biyani, and Tsioutsiouliklis 2016), which uses in-domain words to replace generated word sequence using matching between sentence graphs and generates domain-specific linguistic styles.

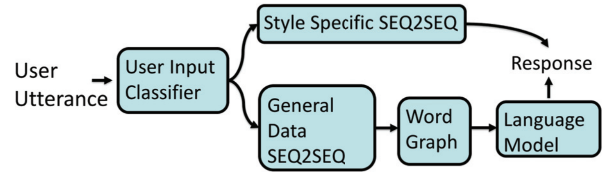

Figure 4: In the style shifting model, the Seq2Seq model chosen is conditioned on the similarity between the user utterance and training data.

Summary: Current and Future Work

Our research focuses on structured hierarchical models for compositional natural language understanding to create human-like dialog acts. In particular, this thesis focuses on domain, style, and discourse cues. Our current work shows that modeling at an utterance level is effective, but a missing component is dialog acts and discourse connectives. In the next few months, we plan to use discourse cues for generation.

References

Bahdanau, D.; Cho, K.; and Bengio, Y. 2015. Neural machine translation by jointly learning to align and translate. *ICLR*.

Banerjee, S.; Biyani, P.; and Tsioutsiouliklis, K. 2016. Transforming chatbot responses to mimic domain-specific linguistic styles. In *Second Workshop on Chatbots and Conversational Agent Technologies*.

Cheng, H.; Fang, H.; and Ostendorf, M. 2017. A factored neural network model for characterizing online discussions in vector space. In *EMNLP*, 2286–2296.

Choudhary, S.; Srivastava, P.; Ungar, L.; and Sedoc, J. 2017. Domain aware neural dialog system. *arXiv preprint arXiv:1708.00897*.

El Hihi, S., and Bengio, Y. 1995. Hierarchical recurrent neural networks for long-term dependencies. In *NIPS*, volume 400, 409. Citeseer.

Filippova, K. 2010. Multi-sentence compression: Finding shortest paths in word graphs. In *ACL*, 322–330.

Jena, G.; Vashisht, M.; Basu, A.; Ungar, L.; and Sedoc, J. 2017. Enterprise to computer: Star trek chatbot. *arXiv preprint arXiv:1708.00818*.

Li, J.; Galley, M.; Brockett, C.; Spithourakis, G. P.; Gao, J.; and Dolan, B. 2016. A persona-based neural conversation model. *arXiv preprint arXiv:1603.06155*.

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.

Sutskever, I.; Vinyals, O.; and Le, Q. V. 2014. Sequence to sequence learning with neural networks. In *NIPS*, 3104–3112.

Tran, Q. H.; Haffari, G.; and Zukerman, I. 2017. A generative attentional neural network model for dialogue act classification. In *ACL*, volume 2, 524–529.

Xing, C.; Wu, W.; Wu, Y.; Liu, J.; Huang, Y.; Zhou, M.; and Ma, W.-Y. 2017. Topic aware neural response generation. In *AAAI*, 3351–3357.