I Don't Know What To Title This Poster

Justin Dieter, Tian Wang, Angel Chang, Gabor Angeli, Arun Chaganty

{jdieter, tian, angel, gabor, arun}@eloquent.ai

Eloquent Labs

Motivation

- Traditionally, dialog agents respond to unexpected input with a potentially frustrating generic "I don't understand" utterance.
- We propose a model for generating contextually aware responses to unexpected input, in order to acknowledge the user's request and gracefully inform them that it cannot be accomplished; for example:

Example User Input	Desired Output
What is the status of my order?	I do not know the status of your order.
I am having trouble with my order and would like to get connected to a support line.	I am not able to help you get connected to a support line.
I would like to know when BayLearn will take place cause I'm hyped for it!	I do not know when BayLearn will take place.

By rewriting the user's request, we can reliably generate responses that acknowledge the user's request while informing them that our dialog agent is unable to accomplish it.

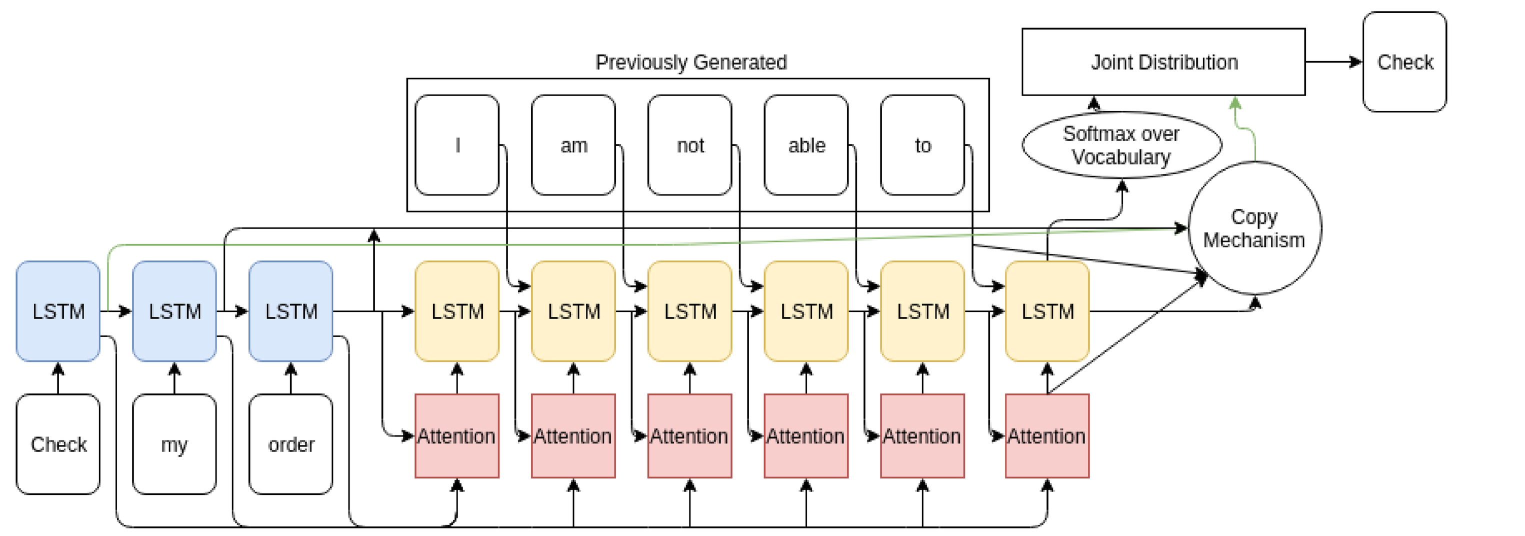
Challenges

- **Semantic understanding of the user's intent; is it a question or a request?**
  - For example, responding to the statement "*I am curious about the hours of operation*" requires understanding that it is actually a question, and to respond with "*I do not know what the hours of operation are*" instead of "*I am not able to be curious about the hours of operation*".
- **Identifying which contexts to include and which to ignore**
  - For example, responding to "*Do I need to submit my application to the French consulate if I am American?*" should capture the conditional clause, "*I do not know if you need to submit your visa application to the French consulate if you are American*".
  - On the other hand, responding to "*I am going on a holiday tour for 6 days in France, can I submit my visa application to the French consulate?*" can be more concise "*I do not know if you can submit your visa application to the French consulate*".
- **Changing frames of reference**
  - For example, responding to "*Do you need my receipt to process a refund?*" should flip the persons in the response, "*I do not know if we need your receipt to process a refund*".
  - On the other hand, responding to "*Do you need to register to pay online?*" needs us to recognize that "*you*" is in the third person and shouldn't be flipped, "*I do not know if you need to register to pay online.*".

Approach

- We implement a strong rule-based baseline using dependency tree rewrite patterns for syntactic reordering.
- In order to handle more complex inputs, we implement a sequence-to-sequence model.
- We augment our model with attention, and with a copy mechanism, allowing for higher reliability rephrasals of the input. The model has a relatively small vocabulary of words it can generate organically, and otherwise copies user input.

Model



An outline of the model. Here, the input phrase is *"Check my order"* with a correct output of *"I am not able to check your order"*. In this figure, we are generating the output word *"check"*.

- Attention is computed as a weighted sum of the hidden states with scores given below:

$$\text{score}(h_t, \bar{h}_s) = v_a^T \cdot \tanh(W_a[h_t; \bar{h}_s]),$$

  where v_a^T and W_a are learnable parameters of a fixed dimension.
- The scores (unnormalized log-probabilities) for copying are computed similarly by:

$$r(\text{copy}(x_s)|h_{t+1}) = v_c^T \cdot \tanh(W_c[h_{t+1}; c_t; \mathbf{E}(y_t); \bar{h}_s; \mathbf{E}(x_s)]),$$

  where $\mathbf{E}$ is the embedding layer. In our experiments, we use GloVE embeddings augmented with ELMo.
- The copy scores and the LSTM scores are additively combined through a softmax distribution to generate from the vocabulary.
- We train the model using a modified cross-entropy loss to account for copy tokens:

$$J = \sum_{i=1}^{m} \left[-\log p(y_i = \hat{y}_i | y_{<i}, x) + \sum_{\{j|x_j=\hat{y}_i\}} -\log p(y_i = copy(x_j)|y_{<i}, x) \right],$$

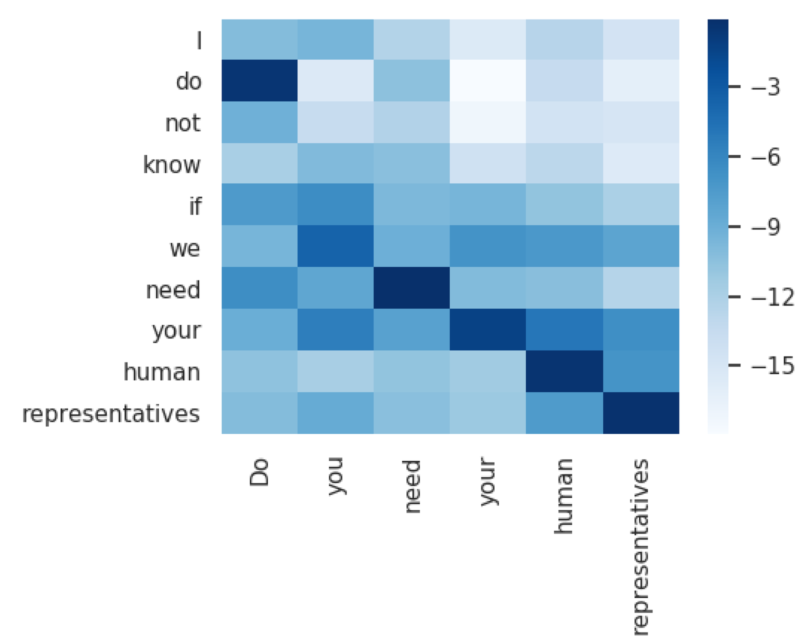
  where $x_1, \ldots, x_n$ are the input sequence of tokens, $\hat{y}_1, \ldots, \hat{y}_m$ are the correct sequence of output tokens, and y_i is the model's next proposed output token.

Evaluation

- We collected a dataset of 9182 examples of user input and response pairs through online crowdsourcing. The dataset is split into a 70:15:15% train:development:test sets.
- We compare the generated responses with the correct responses within the data using BLEU as a metric, and give an example rephrasing (of "If I win my case, what am I entitled to?").

Model	Dev	Test	Example
Rule-based	55.0	55.1	I do not know what you if you win your case, am entitled to.
`seq2seq`	56.4	56.1	I do not know what you are required to.
`+copy+ELMo`	75.8	76.0	I do not know what you are entitled to.

- The `rule-based` model fails on non-trivial syntactic constructs. The `seq2seq` model is more robust syntactically, but often mimics the training data at the expense of adapting to the user's question.
- The `seq2seq+copy+ELMo` model is the most robust model we have but still fails when either words need to be edited on a character level rather than just copied (i.e. add an 's' to be plural), or when it needs to rephrase pronouns like "your" to "our". For example, in the figure below, showing the copy scores for an example input:



Discussion

- Our proposed method enables a graceful back-off for dialogue agents that are unable to satisfy the user's request.
- More broadly, generating contextual back-off responses combines the well-defined goals of task-oriented dialog with the arbitrary input of open-ended chit-chat style dialog.