NBT-BiLA

Vardhman Mehta, Divye Jain & Kevin Zhao

CSE 481N: NLP Capstone | University of Washington

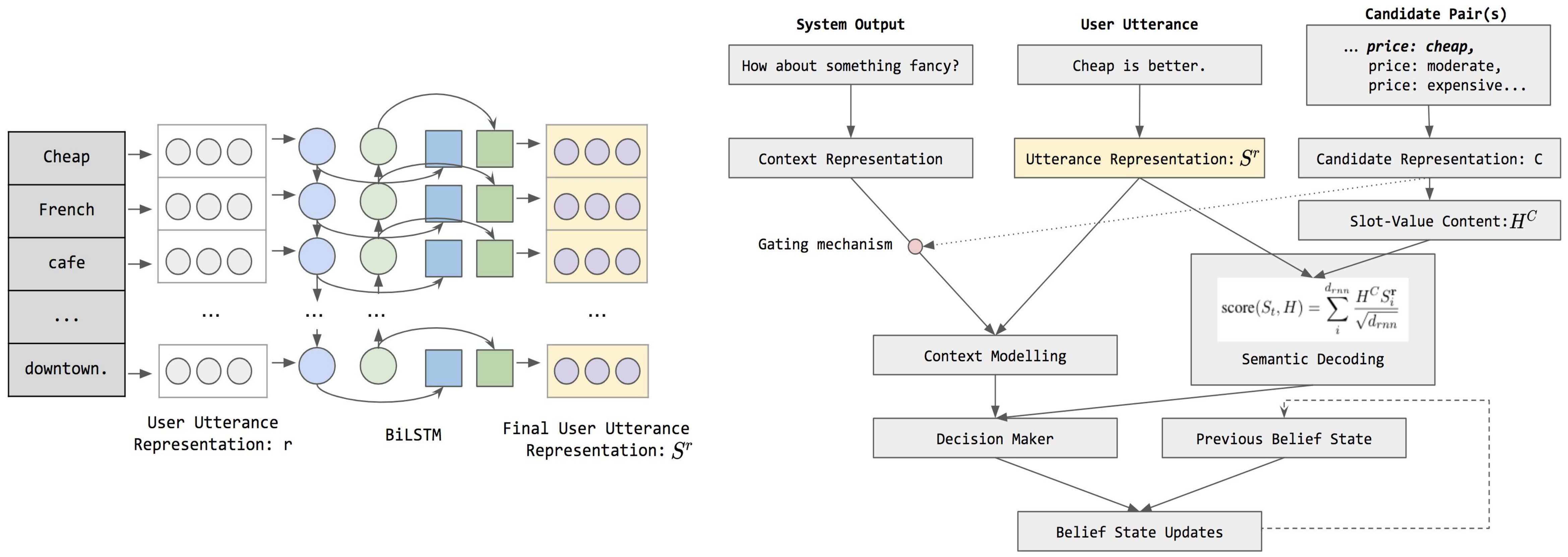
Abstract

Task-oriented dialogue systems have seen a recent boom, especially due to the advent and popularity of virtual assistants like Alexa and the Google Assistant. One of the core components of these modern dialogue systems is the **belief tracker**, which approximates the user's goal at each step in the conversation. There has been a significant amount of work done to design these state trackers, like an *End-to-End Trainable Task-Oriented Dialogue System* or a more recent model like the *Neural Belief Tracker*. In this capstone project we propose an improvement using **Bi**directional **L**STM and **A**ttention over the existing architecture to create ***NBT-BiLA***.

Data

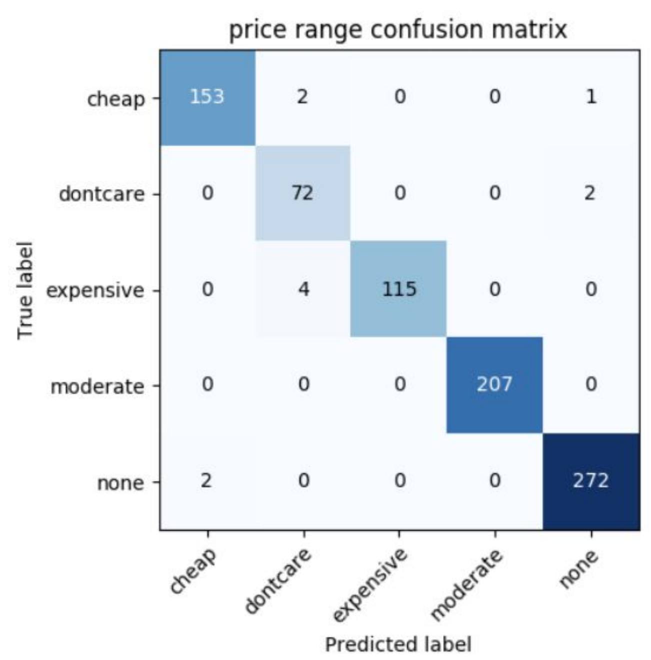
- Wizard-of-OZ v2.0 (***WoZ2.0***) dataset (avg: 4.0 turns)
  - Training (600 dialogues); Validation (200 dialogues); Test (400 dialogues).
- Second Dialogue State Tracking Challenge (***DSTC2***) dataset (avg: 5.5 turns)
  - Training (2207 dialogues); Test (1117 dialogues)

Network Architecture

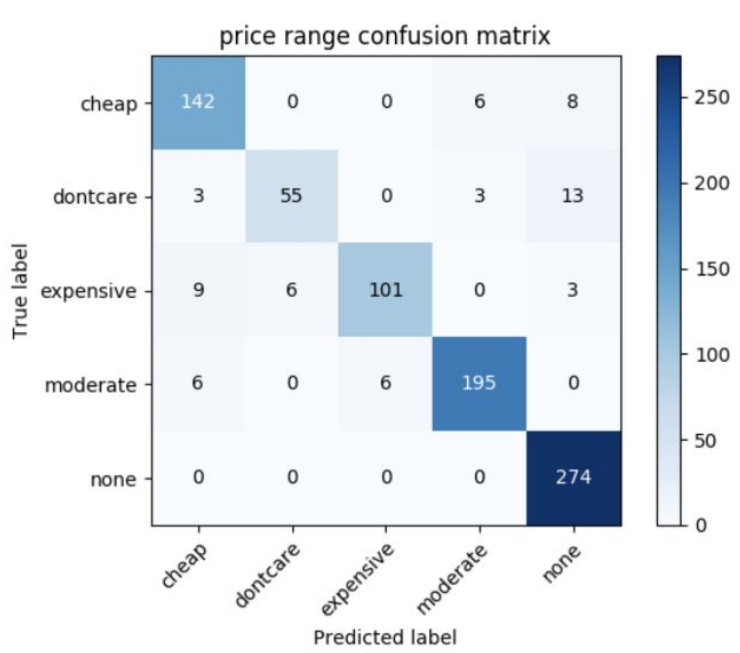


NBT-BiLA network architecture: A bidirectional LSTM layer is added on top of the *user utterances representation*. The model concatenates the representation from the forward and backward layers. This representation (S^r) is then coalesced with the context representation (H^c) of the slot value pairs by adding an Attention layer. The output is then added to the *semantic decoding layer* of the original NBT model and the data continues down the downstream architecture.

Ablation: ***Representation Learning***



Paragram Embeddings



Elmo Embeddings

Embedding Variant	Validation Set Accuracy			
	Food	Joint	Price Range	Area
Paragram + BiLSTM	**97.1**	**95.3***	98.7	**98.3**
ELMo + BiLSTM	96.3	94.1	**98.8**	98.1

Results and Conclusion

This project shows that by improving:

- Representation learning
- Semantic decoding

With the help of

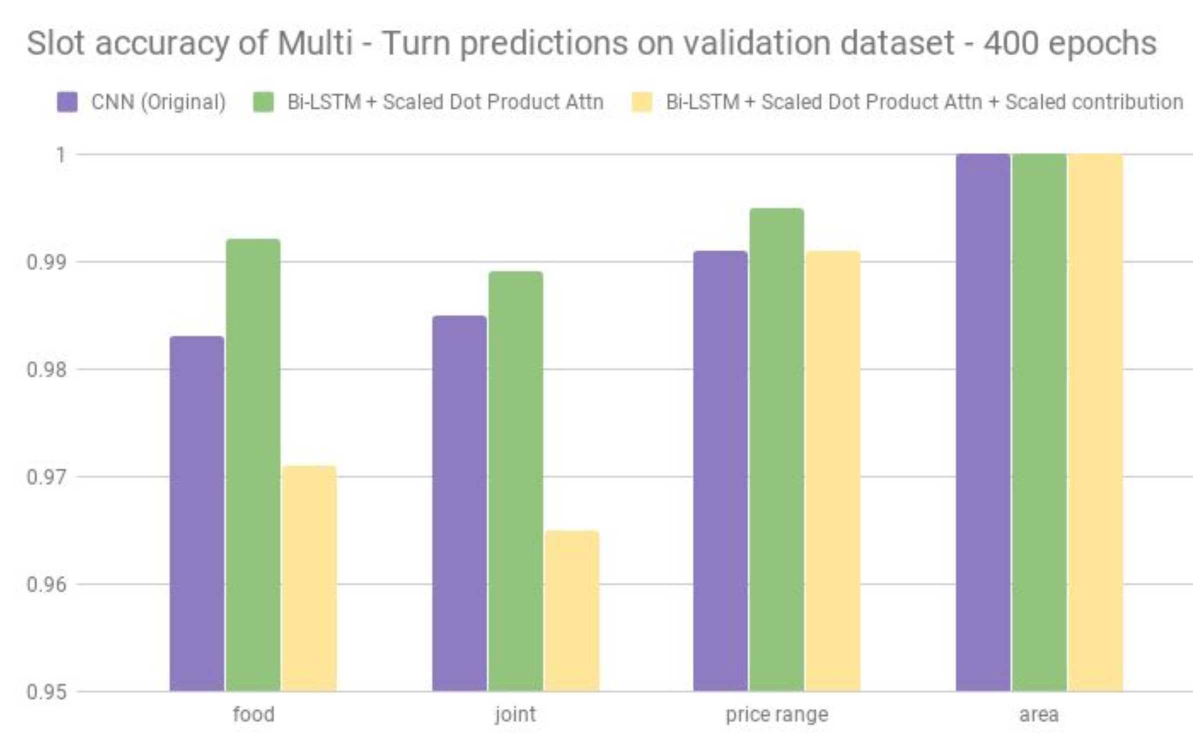
- Recurrent architecture (*BiLSTM*)
- Scaled Dot-product Attention

Reasonable gains on the joint accuracy

- WoZ2.0 dataset
- DSTC2 dataset

DataSet	Model	Test Set Accuracy			
		Food	Joint	Price Range	Area
WOZ	NBT-CNN	-	84.8	-	-
WOZ	NBT-BiLA	90.4	**86.3***	95.1	96.3
DSTC2	NBT-CNN	80.2	68.7	90.9	88.2
DSTC2	NBT-BiLA	**86.1**	**73.2***	82.1	**92.6**

Ablation: ***Semantic Decoding***



Semantic Decoding	Validation Set Accuracy			
	Food	Joint	Price Range	Area
General Attention	0.0	0.0	10.8	7.4
Dot Product Attention	7.9	31.9	98.7	99.3
Scalar Dot Product Attention	**99.2**	**98.9***	**99.5**	**100**