UNIFIED SEMANTIC PARSING WITH SEQ2SEQ MODELS

Subendhu Rongali, Luca Soldaini, Emilio Monti, Wael Hamza
Summer Internship 2019

THE TASK

- Find the intent and slots in a user utterance - core task in digital voice assistants like Alexa, Siri, Cortana etc.
- Utterances can be **simple** with flat slots, or **complex** with hierarchical slots

Artist Name

Play the song don't stop believin by Journey

Song Name

Play Song Intent

TRADITIONAL APPROACHES

- **Slot Filling mechanisms** for simple utterances – CRFs, Bi-LSTMs with CRFs, Capsule networks
- **Recurrent Neural Network Grammars** for complex utterances – Shift-reduce parsers
- All these models only work on **specific grammars**
- These models cannot process slots that are **overlapping or non-consecutive**
- Our model is **unified** and can process slots in **any kind of tagging format**

OUR APPROACH

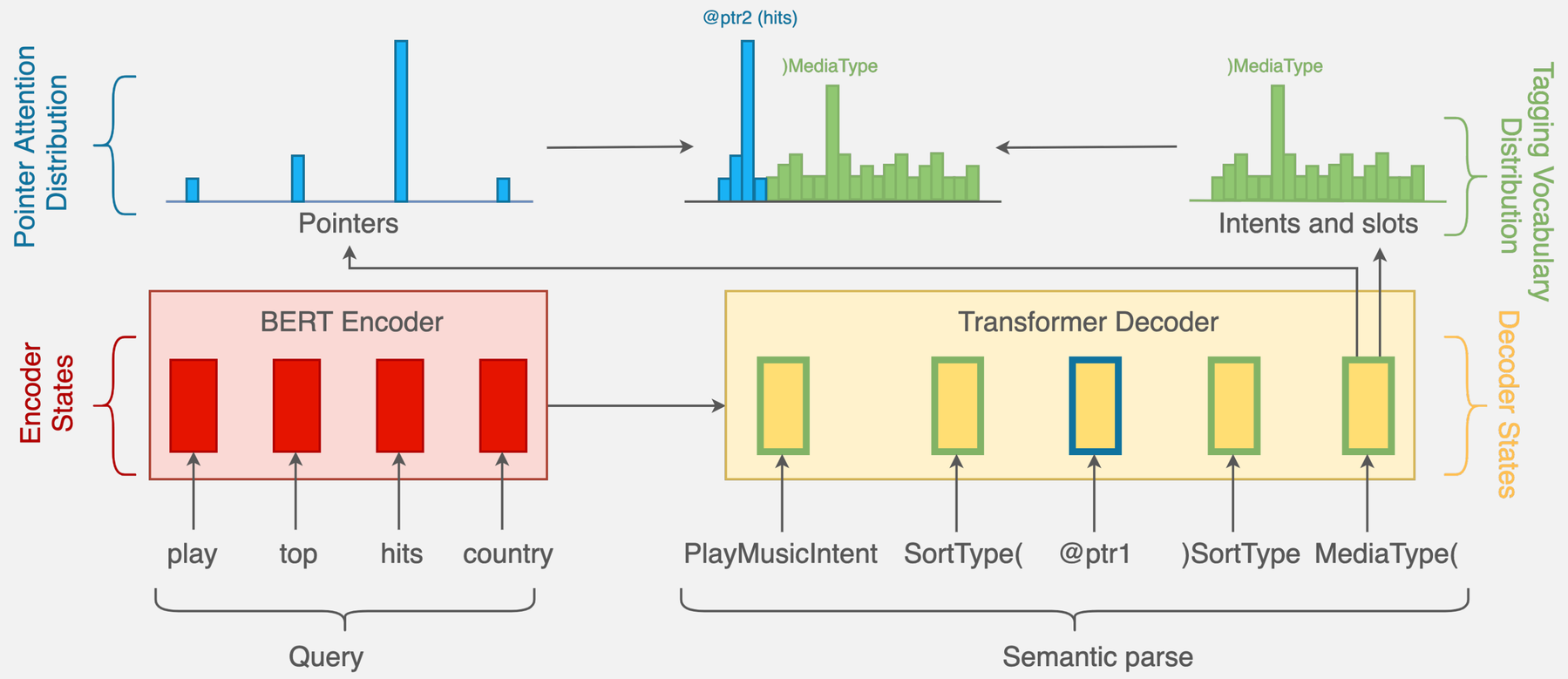
- Use sequence to sequence models to solve this task
- Use a Pointer Generator Network for source tokens
- Easy to see how complex utterances can be framed

- **Source**: Play the song don't stop believin by Journey
- **Target**: PlaySongIntent SongName(@ptr3 @ptr4 @ptr5)SongName ArtistName(@ptr7)ArtistName

768
DIMENSIONS

12
LAYERS

12
HEADS



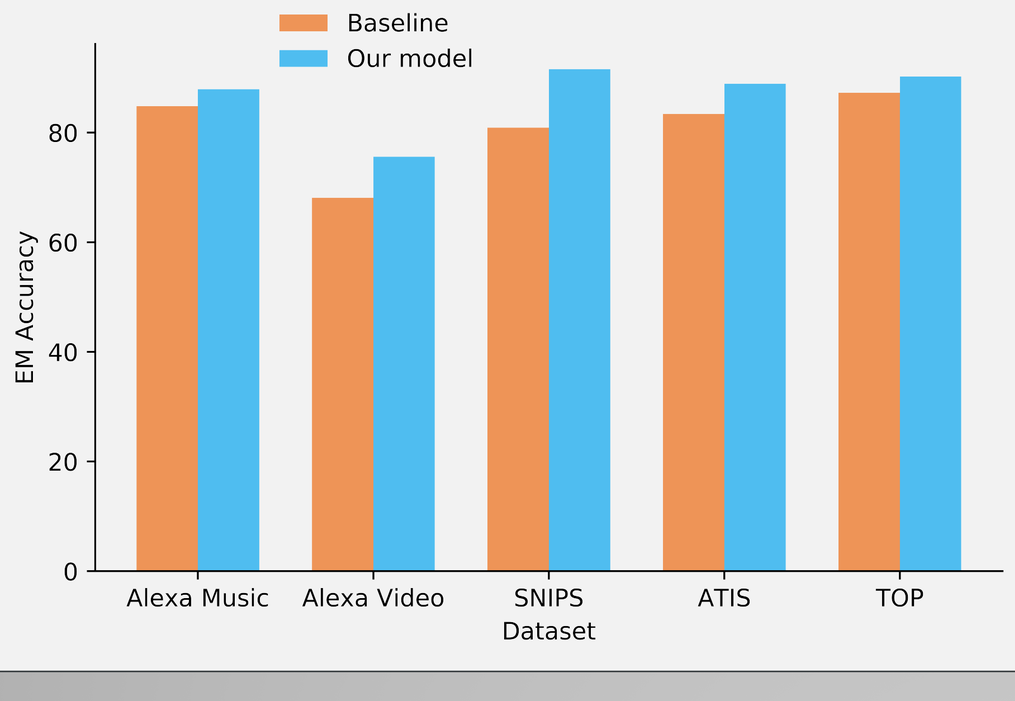
2000
WORD BATCH

CE
LABEL SMOOTHING

5
BEAM SEARCH

RESULTS

- Evaluate on 5 datasets – Alexa Music, Alexa Video, SNIPS, ATIS, and Facebook TOP
- Metric – Exact match accuracy. The entire semantic parse should match
- Best previous baseline models – **Bi-LSTM with CRF** for Alexa, **Capsule Network** for SNIPS & ATIS, **Shift-Reduce ensemble with ELMo and SVM reranking** for TOP



CONCLUSION

- We achieve improved performance on all five datasets
- We have SOTA performance for 3 public datasets – SNIPS, ATIS, and Facebook TOP
- Our model is unified for both simple and complex utterances without any grammar specifications
- We can also handle overlapping and non-consecutive slots

REFERENCES

- Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized BERT pretraining approach (2019).
- Oriol Vinyals, Meire Fortunato, and Navdeep Jaitly. 2015. Pointer networks. In Advances in Neural Information Processing Systems.