Conversational Semantic Parsing

Submitted by Armen Aghajanyan et al (18, Sep 2020)

YIFEI Zhang(.5387)

Abstract

1. Proposes a semantic representation for task-oriented conversational systems that can represent concepts such as **co-reference and context carryove**r
2. Proposes a new family of Seq2Seq models for the **session-based** parsing.
3. **Achieve better or comparable performance t**o the current state-of-the-art on ATIS, SNIPS, TOP and DSTC2.
4. Releases a new session-based, compositional task-oriented **parsing dataset** of 20k sessions consisting of 60k utterances.

Semantic Representation

Task oriented dialog system typically first parse user utterances to semantic frames comprised of Intents and slots

Intents: an abstract meaning which always refers to a sentence or sub-sentence.
Slots: It is attribute or key, which should have a value.

Example (from ATIS):

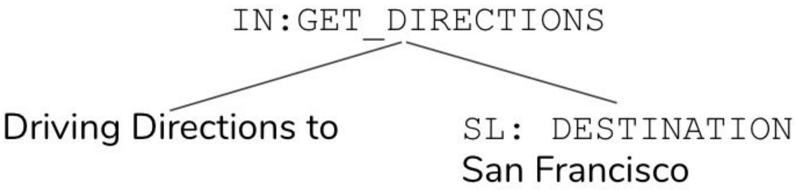
```
Query: What flights are available from pittsburgh to baltimore on thursday
Intent: flight info
Slots:
    - from_city: pittsburgh
    - to_city: baltimore
    - depart_date: thursday
    - depart_time: morning
```

Source: http://nlpprogress.com/

Semantic Representation

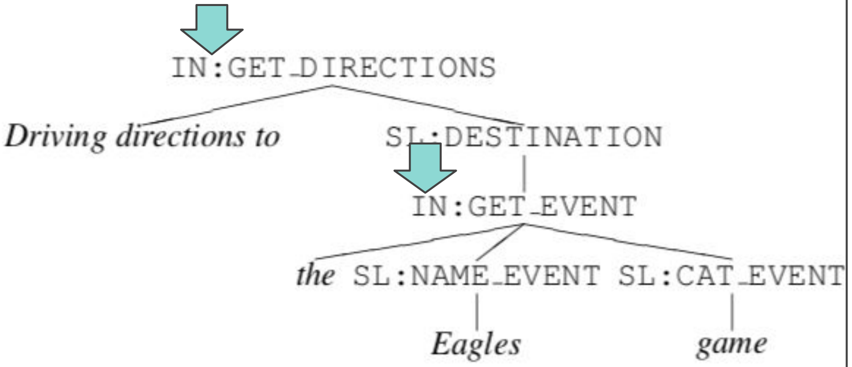
Task : Driving Directions to San Francisco

- Classify the intent of a query (e.g. GET DIRECTIONS)
- Tag the necessary slots (e.g. San Francisco)



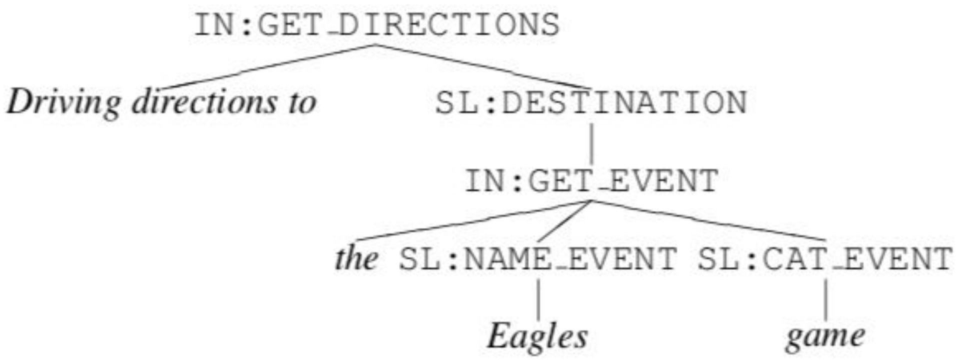
Flat Representation

Task : Driving Directions to <u>the Eagle Game</u>



Compositional Representation (Gupta et al 2018)

Compositional Representation (Gupta et al 2018)



Linearized version of tree

[IN:GET_DIRETIONS Driving directions to [SL:DESTINATION [IN:GET_EVENT the [SL:NAME_EVENT Eagles][SL:CAT_EVENT game]]]]

Limitation: In-order traversal constraint disallow long-distance dependencies within semantic representation

eg:***On Monday**, set an alarm for **8am**.*

Want a single date-time slot: [SL DATETIME **8am on Monday**]

Decoupled Representation

Convert Compositional Form into a logical form containing two label type (SLOT and INTENT)
Removing all text in the compositional semantic parse that does not appear in a leaf slot

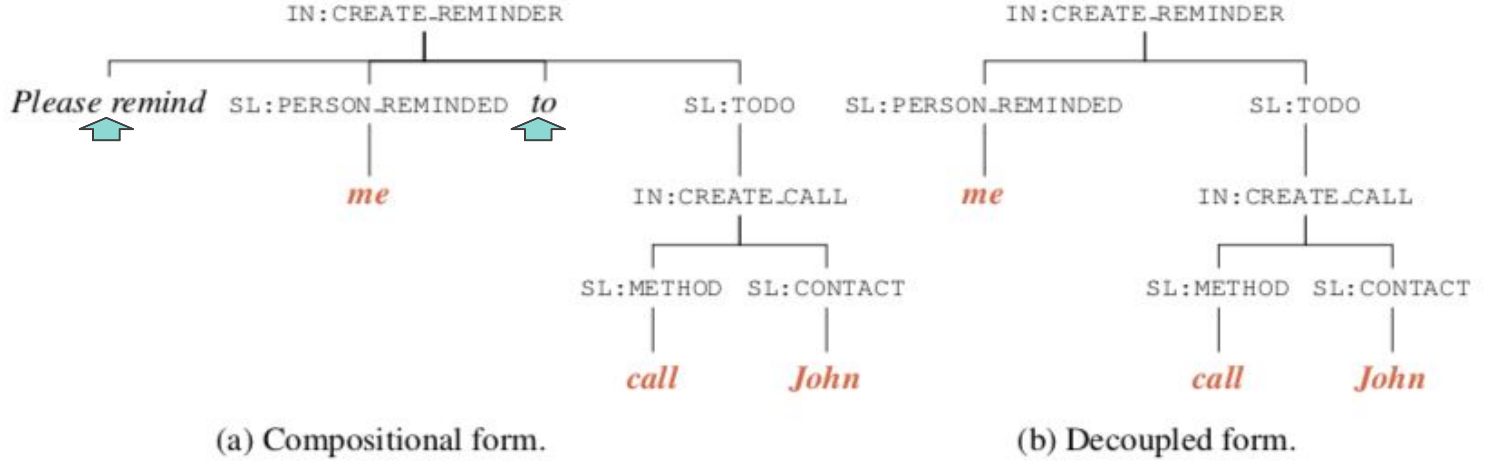
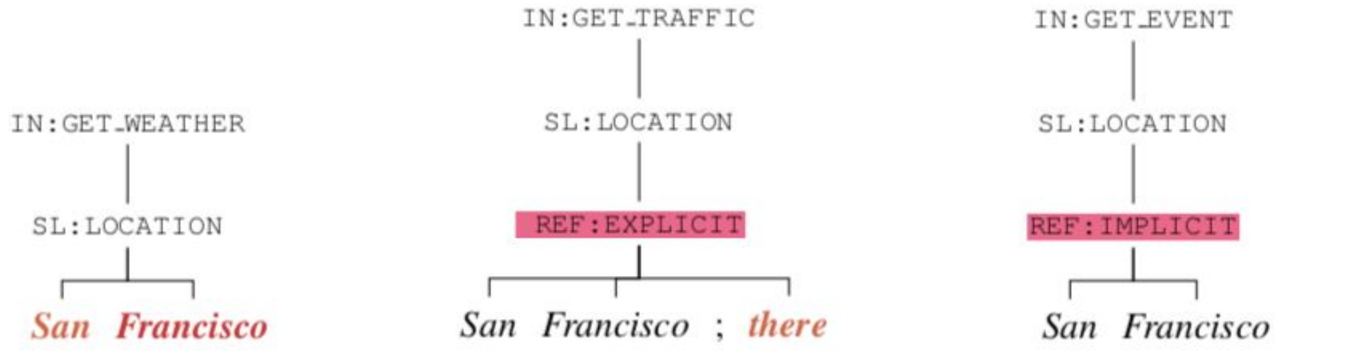


Figure 1: Compositional and decoupled semantic representations for the single utterance “Please remind me to call John”.

Session-Based Semantic Parsing

Introduce a new reference(REF) label type



(a) “Weather in San Francisco”. (b) “Traffic there”. (c) “Any events going on?”

Figure 4: Decoupled semantic representations for a three-utterance session.

Co-references - Explicit reference

Slot-carryover: Implicit reference

Model -Seq2Seq Architecture

Encoder :

$$e_1, ..., e_T = \text{Encoder}(w_1, ..., w_T)$$

Bidirectional LSTM / Transformer

Decoder :

Every decoding step, the model can either **generate an element from the ontology**, or **copy a token** from the source sequence

Ontology Generation Distribution

$$\begin{aligned} \boldsymbol{x}_t &= \text{Decoder}\left(\boldsymbol{e}_1, ..., \boldsymbol{e}_t; \boldsymbol{d}_{t-1}; \boldsymbol{s}_{t-1}\right), \\ \boldsymbol{p}_t^{\mathrm{g}} &= \text{softmax}\left(\text{Linear}_{\mathrm{g}}[\boldsymbol{x}_t]\right), \end{aligned}$$

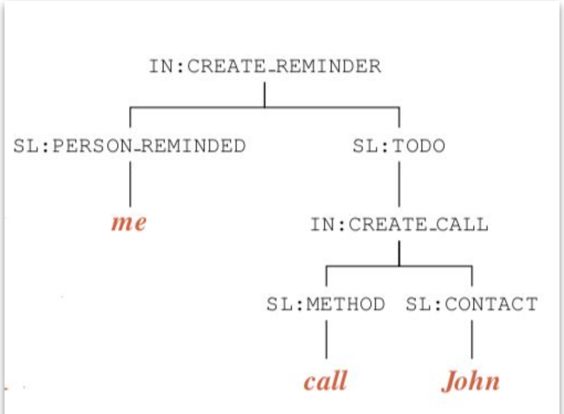
$$\mathbf{W}^\theta \boldsymbol{x} + \boldsymbol{b}^\theta$$

Utterance Copy Distribution (attention distribution)

$$\boldsymbol{p}_t^{\mathrm{c}}, \boldsymbol{\omega}_t = \text{MhAttention}\left(\boldsymbol{e}_1, ..., \boldsymbol{e}_t; \text{Linear}_{\mathrm{c}}[\boldsymbol{x}_t]\right),$$

$$\boldsymbol{p}_t = p_t^{\alpha} \cdot \boldsymbol{p}_t^{\mathrm{g}} + (1 - p_t^{\alpha}) \cdot \boldsymbol{p}_t^{\mathrm{c}}.$$

$$p_t^{\alpha} = \sigma\left(\text{Linear}_{\alpha}\left[\boldsymbol{x}_t \| \boldsymbol{\omega}_t\right]\right)$$



(b) Decoupled form.

```
[IN:CREATE REMINDER,
[SL:PERSON REMINDED,
me, ], [SL:TODO, ...,
]
```

Tokens are divided into 2 class:
Utterance token (the leaves of tree) and Ontology Symbols

Model -Encoder and Decoder

1. **RNN based**

   Encoder: 2 Stacked bidirectional LSTM + Decoder: UniDirectional LSTM

   Extentened Version (Incorporating Contextualized word vector ElMo Embeddings)

2. **Transformer**

   Encoder: RoBERTa +Decoder: 3 Layers Transformer

   Encoder: BART+ Decoder: BART

Experiment

DataSet Released(SB-TOP): 20K annotated session in 4 Domains: calling, weather, music and reminder.(Allow for mixture of domain within a session) http://dl.fbaipublicfiles. com/sbtop/SBTOP.zip

Collection: Crowdsourced Workers to write session, then ask 2nd group of annotators to annotate.

Dataset	Characteristic	
TOP	31k training: Navigation, event, and navigation to event domain	
SNIPS-NLU	Playlist, Restaurant, Weather, Music, RateBook, etc.	
ATIS	Airline Travel Information Systems	
DSTC2	DSTC 2 (Restaurant Information Domain),Over 2k sessions	**Multiple Turns**
Internal Dataset	170k training, annotated with **flat** representation (from variety of domain: weather, communication, music, weather, device control)	Cover 140 distinct intent
Internal Dataset	67k training, annotated with **hierarchical** representation	Cover 60 intent in communication domain

Experiment

Table 1: Frame accuracy of the decoupled models on semantic parsing tasks. † indicates results from Hakkani-Tür et al. (2016); ‡, from Goo et al. (2018); ∗, from Zhang et al. (2018); ×, from Chen et al. (2019a).

(a) Accuracy on TOP.

Model	Acc.
RNNG	80.86
RNNG + Ensembling	83.84
RNNG + ELMo	83.93
Decoupled biLSTM	79.51
Decoupled transformer	64.50
Decoupled ELMo	84.85
Decoupled RoBERTa	84.52
Decoupled BART	**87.10**
Best Seq2SeqPtr	86.67

(b) Accuracy on ATIS and SNIPS.

Model	ATIS	SNIPS
Joint biRNN†	80.7	73.2
Slot gated‡	82.2	75.5
CapsuleNLU*	83.4	80.9
Joint BERT×	88.2	**92.8**
Joint BERT CRF×	88.6	92.6
Decoupled BART	**89.25**	91.00
Best Seq2SeqPtr	87.12	87.14

(c) Accuracy on internal datasets.

Model	Acc.
Multi-domain (170k)	
Decoupled ELMo	86.03
Decoupled RoBERTa	87.32
Decoupled BART	**88.29**
Single-domain (67k)	
Decoupled ELMo	90.52
Decoupled RoBERTa	91.51
Decoupled BART	**92.16**

New Release DataSet(**SB-TOP**)

Table 2: Decoupled model architecture results over the SB-TOP dataset. FA is exact match between canonicalized predicted and tree structures. Ref Only FA does not distinguish between implicit/explicit references. Intent accuracy is accuracy over top level intents while Inner Parse Accuracy is FA not considering top level intent.

Model	Oracle@Beam	FA	Ref-only FA	Intent Acc.	Inner Parse Acc.
Humans	1	55.04	57.4	84.32	60.12
Decoupled biLSTM	1	48.48	49.19	78.60	52.74
	5	60.24	69.88	93.71	72.01
Decoupled ELMo	1	51.22	52.03	80.93	55.07
	5	62.58	70.08	94.73	72.11
Decoupled BART	1	**53.45**	**54.18**	**82.46**	**56.84**
	5	**65.19**	**72.78**	**96.67**	**76.45**

Experiment (Slot Carryover)

Table 3: Performance of the decoupled models on a state tracking task (DSTC2).

Model	Accuracy	Slot distance			
		0	1	2	≥ 3
LSTM-based (Naik et al., 2018)	—	92.42	91.11	91.34	87.99
Pointer network decoder (Chen et al., 2019b)	—	92.70	92.04	92.90	91.39
Transformer decoder (Chen et al., 2019b)	—	93.00	92.69	92.80	89.49
GLAD (Zhong et al., 2018)	74.5	–	–	–	–
Decoupled biLSTM	88.3	93.34	94.73	95.28	95.73
Decoupled RoBERTa	89.8	91.98	92.94	93.58	94.28
Decoupled BART	**90.2**	94.21	95.47	95.90	97.05

Related Work

Traditional work on Semantic Parsing(question answering or task- oriented request understanding) has focused on mapping utterances to **logical form** representations (Zelle and Mooney, 1996; Zettle- moyer and Collins, 2005; Kwiatkowksi et al., 2010; Liang, 2016; van Noord et al., 2018).

Logical Form Lambda Calculus Encoding

a) What states border Texas

$$\lambda x.state(x) \wedge borders(x, texas)$$

Intent-slot representations have less expressive power , but have the major advantage of being simple enough to enable the creation of large-scale datasets. Gupta et al. (2018) introduce a hierarchical intent-slot representation ,more expressive.

Recent approaches: RNNGs (Gupta et al., 2018), RNNGs augmented with ensembling and re-ranking techniques or contextual embeddings (Einolghozati et al., 2018), sequence-to-sequence recurrent neural networks augmented with pointer mechanisms (Jia and Liang, 2016), capsule networks (Zhang et al., 2019), and Transformer-based architectures (Rongali et al., 2020).