Natural Language Understanding,Generation, and Machine Translation

Lecture 21: Semantic Parsing

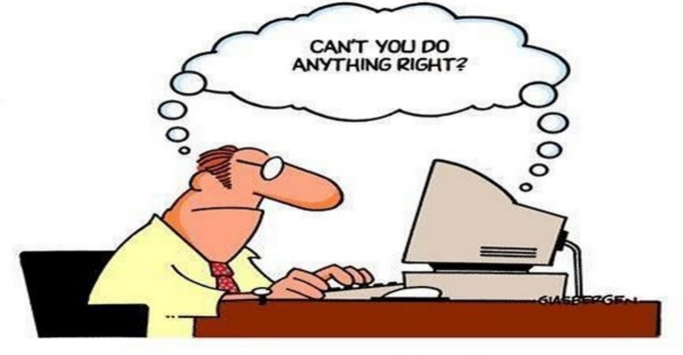
Laura Perez-Beltrachini

(Slides by Mirella Lapata)

School of Informatics
University of Edinburgh
mlap@inf.ed.ac.uk

Natural Language Interfaces

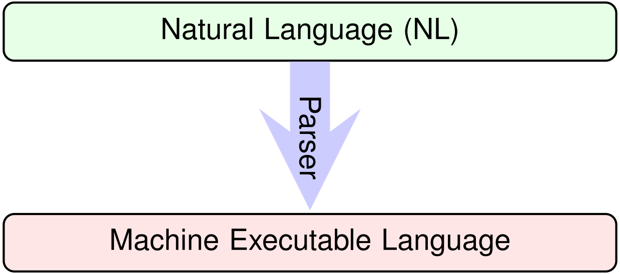
Semantic parsing: users interact with computers in human language!



CONVERSATION MACHINE (Green et al., 1959); BASEBALL (Green et al., 1961).

Natural Language Interfaces

Semantic parsing: users interact with computers in human language!



Semantic Parsing: Querying a Database

NL What are the capitals of states bordering Texas?

↓ Parser

DB $\lambda x.\ \text{capital}(y, x) \wedge \text{state}(y) \wedge \text{next\_to}(y, \text{Texas})$

Semantic Parsing: Instructing a Robot

[1]

NL: at the chair, move forward three steps past the sofa

Parser

LF: $\lambda a.pre(a, x.chair(x)) \wedge move(a) \wedge len(a, 3) \wedge dir(a, forward) \wedge past(a, y.sofa(y))$

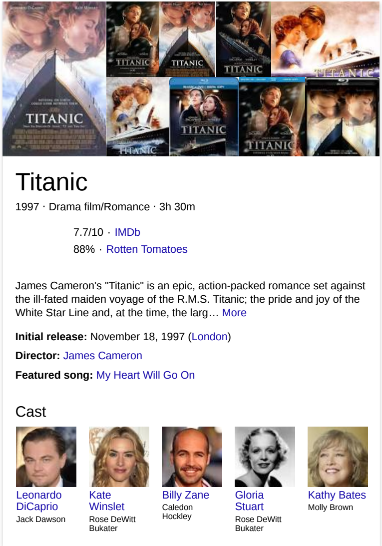
[1] Courtesy: Artzi et al., 2013

Semantic Parsing: Answering Questions with Freebase

NL Who are the male actors in Titanic?

Parser

KB $\lambda x.\, \exists y.\, \mathrm{gender}(\textsc{Male}, x) \wedge \mathrm{cast}(\textsc{Titanic}, x, y)$



Semantic Parsing: Digital Assistants

Bill Gates lists smooth-talking AI assistants as one of 2019 inventions that will change the world for the better (MIT Technology Review).

Semantic Parsing: Digital Assistants

Learning Setting

We will induce semantic parsers parsers from sentence-logical form pairs.

Question

Who are the male actors in Titanic?

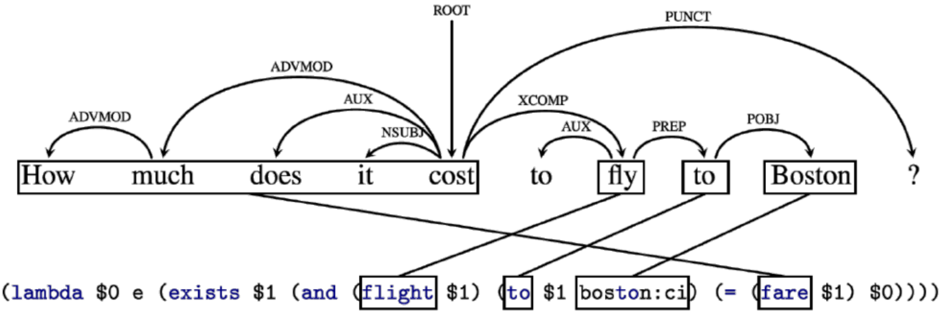
Logical Form

$$\lambda x.\, \exists y.\, \text{gender}(\textsc{Male}, x) \wedge \text{cast}(\textsc{Titanic}, x, y)$$

Miller et al., (1996); Zelle and Mooney, (1996); Tang and Mooney, (2000) ; Thomspon and Mooney, (2003); Kate et al., (2005); Ge and Mooney, (2005); Kate and Mooney, (2006); Wong and Mooney, (2006); Wong and Mooney, (2007); Zettlemoyer and Collins, (2005); 2007); Lu et al., (2008); Kwiatkowski et al., 2010); (2011); Andreas et al., (2013); Zhao and Huang, (2015); . . .

Challenge 1: Matching Natural and Artificial Language

Machine language is different from NL string and its syntactic representation.



Challenge 2: Well-formedness of Machine Language

NL Who are the male actors in Titanic?

↓ Parser

✗ **KB** $\text{gender}(\textsc{Male}, x) \wedge \text{cast}(\textsc{Titanic}, x, y)$

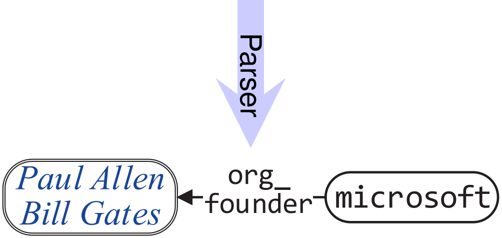
✗ **KB** $\lambda x.\ \text{gender}(\textsc{Male}, x \wedge \text{cast}(\textsc{Titanic}, x, y)$

✓ **KB** $\lambda x.\ \exists y.\ \text{gender}(\textsc{Male}, x) \wedge \text{cast}(\textsc{Titanic}, x, y)$

Challenge 3: Linguistic Coverage

The same meaning can be expressed in many different ways!

Who created Microsoft?
Microsoft was created by?
... ...
Who founded Microsoft?
Who is the founder of Microsoft?



Taking on the Challenges

1. **Structural Mismatches:** neural encoder-decoder architecture for mapping natural language expressions to their logical forms.
2. **Well-formedness Constraints**: coarse-to-fine decoding algorithm generates well-formed meaning representations.
3. **Linguistic Coverage**: query paraphrasing framework handles variation of natural language input.

Many more challenges I will **not** talk about: Where does the training data come from? What happens if the queries are out-of-domain or co-referring? Why doesn't Alexa understand me?

A Propos Speech Recognition

Microsoft's newest milestone? World's lowest error rate in speech recognition

Microsoft has leapfrogged IBM to claim a significant test result in the quest for machines to understand speech better than humans.

September 14, 2016

Amazon Alexa scientists reduce speech recognition errors by 20% with semi-supervised learning

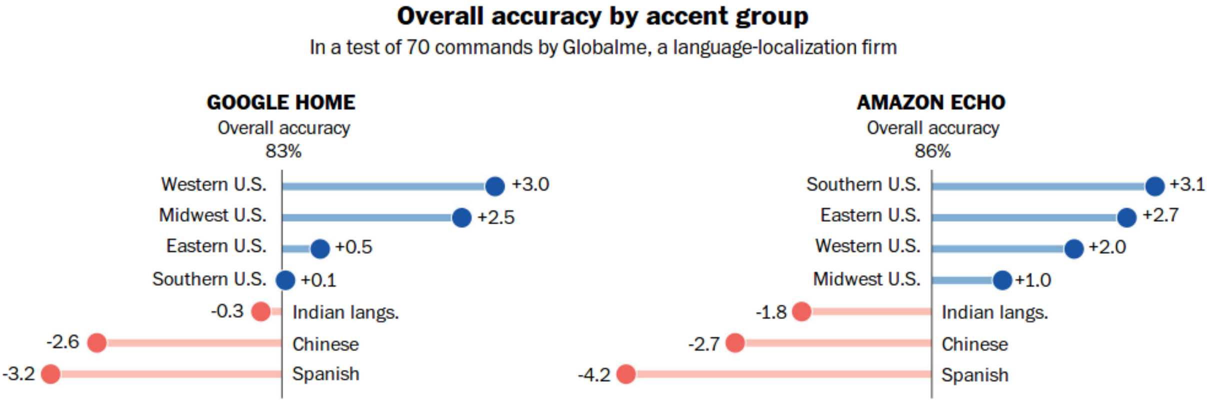
March 20, 2019

Google May Have Finally Made a Truly Usable Voice Assistant

May 8, 2019

Google says it will include a new version of its speech-recognition software in some new phones, potentially transforming how people accomplish tasks.

A Propos Speech Recognition



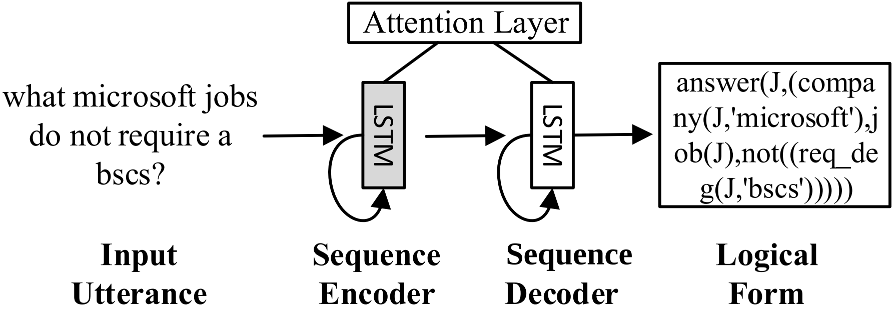
“These systems are going to work best for white, highly educated, upper-middle-class Americans, probably from the West Coast, because that’s the group that’s had access to the technology from the very beginning.”, The Washington Post, July 19, 2018.

A Propos Speech Recognition

Taking on the Challenges

1. **Structural Mismatches:** neural encoder-decoder architecture for mapping natural language expressions to their logical forms.

SEQ2SEQ Model



Dong and Lapata, (2016); Vinyals et al., (2015a,b), Kalchbrenner and Blunsom (2013), Cho et al., (2014), Sutskever et al., (2014), Karpathy and Fei-Fei, (2015)

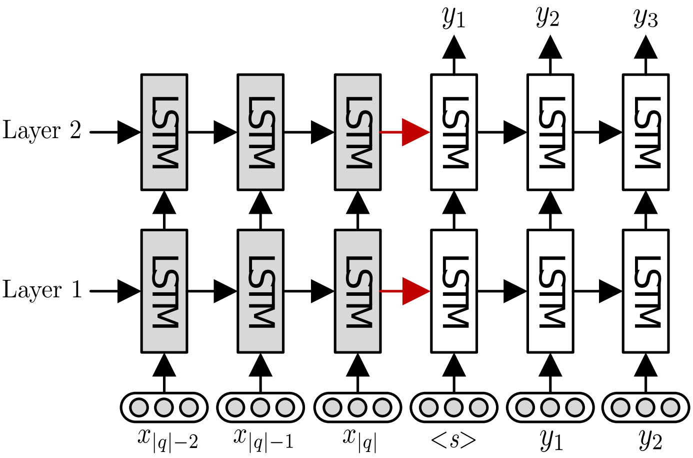
SEQ2SEQ Model

Model maps **natural language input** $q = x_1 \cdots x_{|q|}$ to a **logical form representation** of its meaning $a = y_1 \cdots y_{|a|}$.

$$p(a|q) = \prod_{t=1}^{|a|} p(y_t|y_{<t}, q) \qquad \text{where } y_{<t} = y_1 \cdots y_{t-1}$$

- **Encoder** encodes natural language input q into a vector representation
- **Decoder** generates $y_1, \cdots, y_{|a|}$ conditioned on the encoding vector.

SEQ2SEQ Model



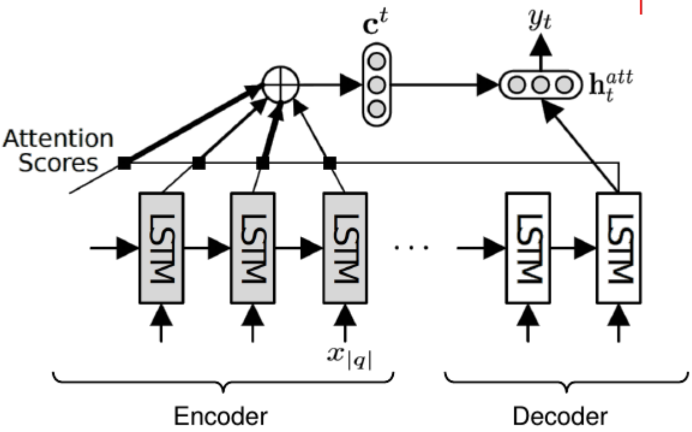
$$\mathbf{h}_t^l = \mathrm{LSTM}\left(\mathbf{h}_{t-1}^l, \mathbf{h}_t^{l-1}\right)$$

$$\mathbf{h}_t^0 = \mathbf{W}_q \mathbf{e}(x_t)$$

$$\mathbf{h}_t^0 = \mathbf{W}_a \mathbf{e}(y_{t-1})$$

$$p\left(y_t | y_{<t}, q\right) = \mathrm{softmax}\left(\mathbf{W}_o \mathbf{h}_t^L\right)^{\mathsf{T}} \mathbf{e}\left(y_t\right)$$

Attention Mechanism



$$r_{t,k} \propto \exp\{\mathbf{h}_t^L \cdot \mathbf{h}_k^L\}$$

$$\mathbf{h}_t^{att} = \tanh\left(\mathbf{W}_1 \mathbf{h}_t^L + \mathbf{W}_2 \mathbf{c}_t\right)$$

$$p\left(y_t | y_{<t}, q\right) = \mathrm{softmax}_{a_t}\left(\mathbf{W}_o \mathbf{h}_t^{att}\right)$$

Bahdanau et al., (2015), Luong et al., (2015b), Xu et al., (2015)

Training and Inference

Training maximizes likelihood of logical forms given natural language input:

$$\max \sum_{(q,a)\in\mathcal{D}} \log p\left(a|q\right)$$

where $\mathcal{D}$ is the set of all natural language-logical form training pairs

At test time we predict the logical form for an input utterance q by:

$$\hat{a} = \arg\max_{a'} p\left(a'|q\right)$$

- Iterating over all possible a's to obtain the optimal prediction is impractical.
- Probability $p\left(a|q\right)$ decomposed so that we can use greedy/beam search.

Taking on the Challenges

1. **Structural Mismatches:** neural encoder-decoder architecture for mapping natural language expressions to their logical forms.
2. **Well-formedness Constraints**: coarse-to-fine decoding algorithm generates well-formed meaning representations.

Coarse-to-Fine Decoding

NL: all flights from dallas before 10am

↓

Meaning Sketch: (lambda#2 (and flight@1 from@2 (< departure_time@1 ?)))

↓

Low-level Details: e.g., arguments and variable names

↓

LF: (lambda $0 e (and (flight $0) (from $0 dallas:ci) (< (departure_time $0) 1000:ti)))

Meaning Sketches

- **Disentangle** high-level from low-level semantics; different levels of granularity.
- More **compact** meaning representation (length: 21.1 $\rightarrow$ 9.2 on ATIS).
- Explicit **sharing** of coarse structure which is the same for examples with same basic meaning.
- Provide **global** context to fine meaning decoding.

Coarse-to-Fine Model

Model maps **natural language input** $x = x_1 \cdots x_{|q|}$ to a **logical form representation** of its meaning $y = y_1 \cdots y_{|y|}$.

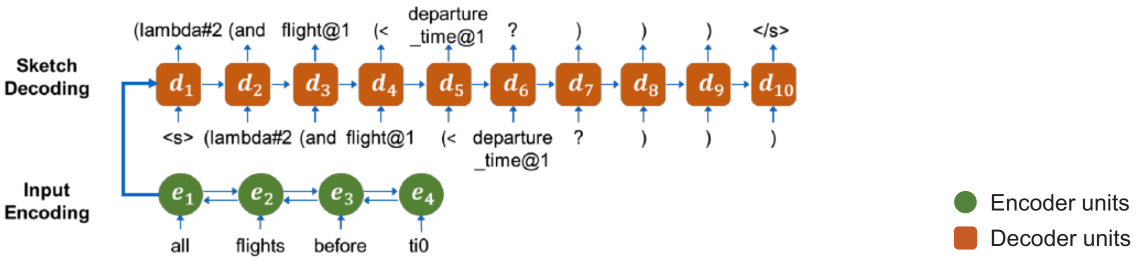
$$p(y|x) = p(y|x, a)p(a|x)$$

where $a = a_1 \ldots a_{|a|}$ is an abstract sketch representing the meaning of y.

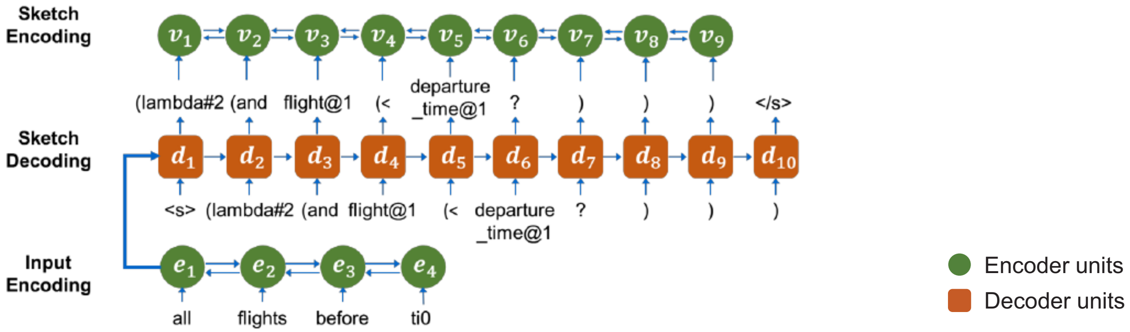
$$p(y|x, a) = \prod_{t=1}^{|y|} p(y_t|y_{<t}, x, a) \qquad p(a|x) = \prod_{t=1}^{|a|} p(a_t|a_{<t}, x)$$

where $a_{<t} = a_1 \cdots a_{t-1}$, and $y_{<t} = y_1 \cdots y_{t-1}$.

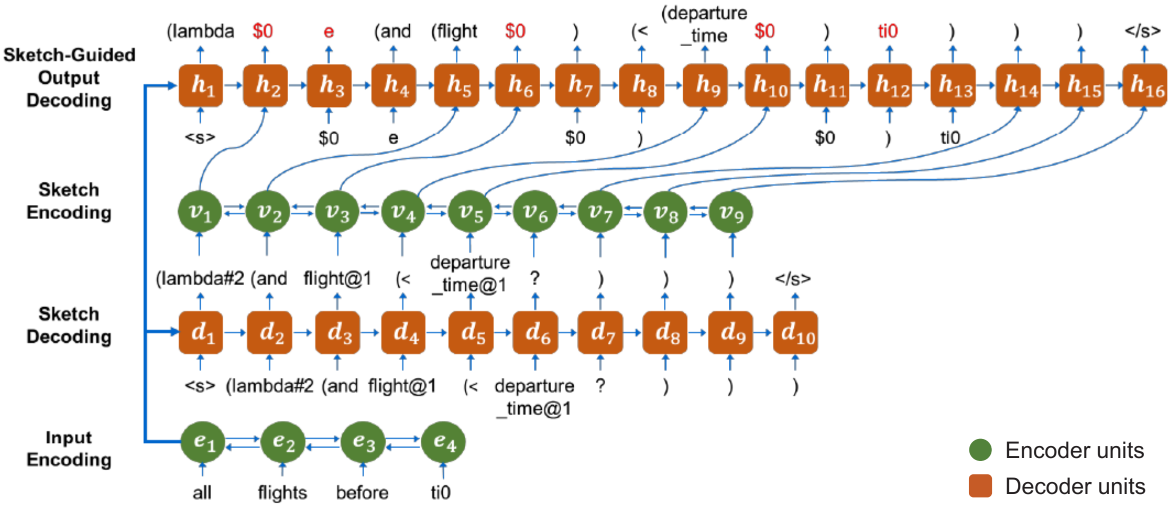
Modeling Framework



Modeling Framework



Modeling Framework



Training and Inference

Training maximizes the log likelihood of the generated meaning representations given natural language expressions:

$$\max \sum_{(x,a,y)\in\mathcal{D}} \log p\,(y|x,a) + \log p\,(a|x)$$

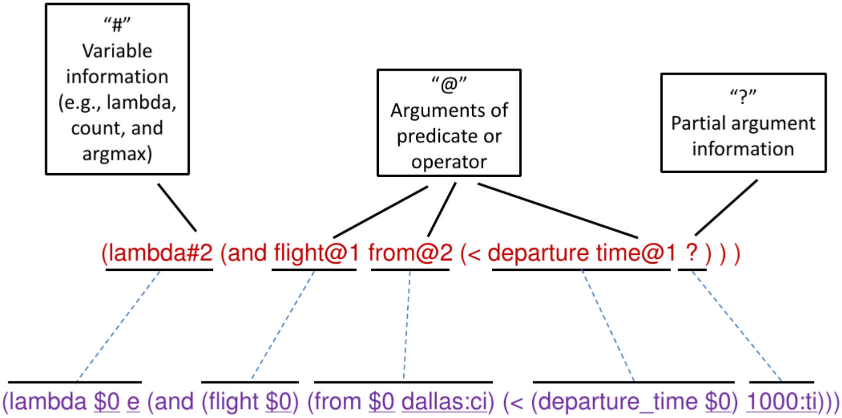
$\mathcal{D}$ are training pairs, x input, a sketch, and y meaning representation.

At test time we predict a and y via greedy search:

$$\hat{a} = \arg\max_{a'} p\,(a'|x) \qquad \hat{y} = \arg\max_{y'} p\,(y'|x,\hat{a})$$

a' and y' represent coarse- and fine-grained meaning candidates.

Tasks: Natural Language to Logical Form



GEOQUERY and ATIS (Zettlemoyer and Collins, 2005; Kwiatkowski et al., 2011).

Tasks: Natural Language to Source Code

if NAME [: NUMBER] . NAME ( ) == STRING :

if s [: 4] . lower ( ) == 'http':

Substitute tokens[2] with their token types, except:

- delimiters (e.g., "`[`", and "`:`")
- operators (e.g., "+", and "⋆")
- built-in keywords (e.g., "`True`", and "`while`")

DJANGO dataset (Oda et al., 2015).

[2] `https://docs.python.org/3/library/tokenize.html`

Tasks: Natural Language to SQL

```
SELECT agg_operator agg_column
WHERE (cond_column cond_operator cond_value)
AND ...
```

SELECT *Record Company*
WHERE (*Year of Recording > 1996*) AND (*Conductor = Mikhail Snitko*)

WHERE __ > AND __ =

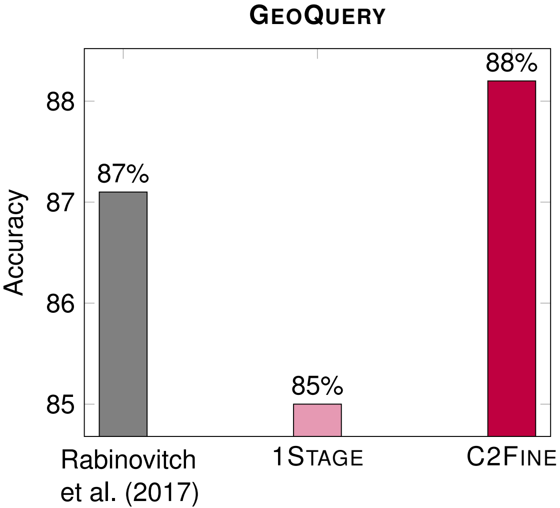
WIKISQL dataset (Zhong et al., 2017), 80,654 questions and queries distributed across 24,241 tables from Wikipedia).

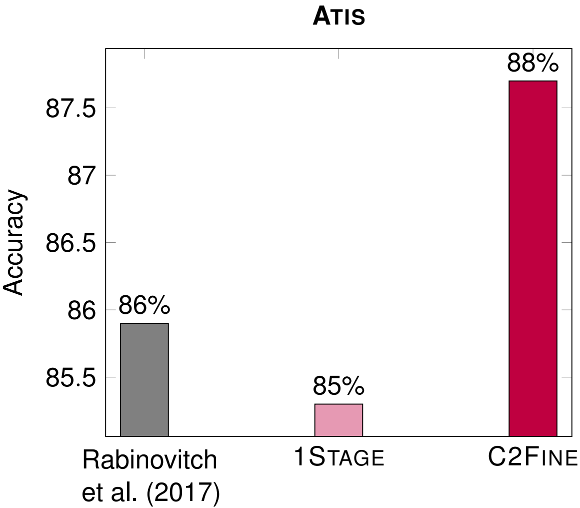
Evaluation

Exact Match Accuracy: proportion of input queries that are correctly parsed to their gold standard logical forms.

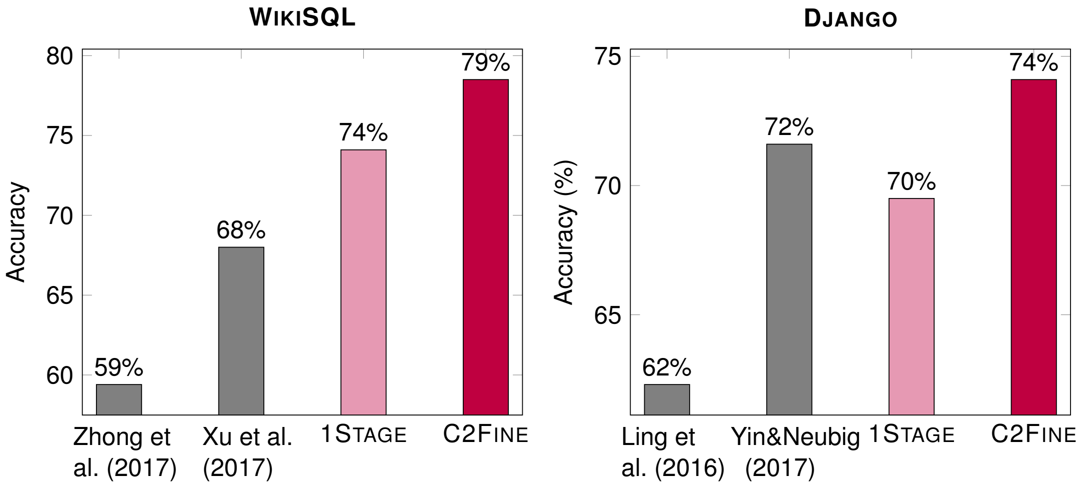
Denotation Match Accuracy: proportion of correct denotations (answers that the logical forms give when executed against the knowledge base).

Results


GeoQuery
Accuracy
88
87
86
85
87%
85%
88%
Rabinovitch et al. (2017)
1Stage
C2Fine


Atis
Accuracy
87.5
87
86.5
86
85.5
86%
85%
88%
Rabinovitch et al. (2017)
1Stage
C2Fine

Experimental Results



Taking on the Challenges

1. **Structural Mismatches:** neural encoder-decoder architecture for mapping natural language expressions to their logical forms.
2. **Well-formedness Constraints**: coarse-to-fine decoding algorithm generates well-formed meaning representations.
3. **Linguistic Coverage**: Next Lecture!

What Have we Learned?

- **Encoder-decoder** neural network model for mapping natural language to meaning representations (minimal engineering effort).
- **Constrained decoding** improves performance (Coarse-to-fine).
- **General** models could transfer to other semantic parsing tasks/architectures.
- Future work: learn meaning sketches, take multiple languages into account, learn model from database alone (without parallel data).

Data and code from: `https://github.com/donglixp/`