A Discrete Hard EM Approach for Weakly Supervised Question Answering

EMNLP 2019

Sewon Min[1], Danqi Chen[2,3], Hannaneh Hajishirzi[1,4], Luke Zettlemoyer[1,3]

[1]University of Washington, [2]Princeton University, [3]Facebook AI Research, [4]Allen Institute of AI

Overview

Reading comprehension

Discrete reasoning task

Semantic parsing

Task formulation

A unified weak supervision scenario with
a small set of possible solutions

Overview

Reading comprehension

Discrete reasoning task

Semantic parsing

Task formulation

A unified weak supervision scenario with
a small set of possible solutions

Learning method

A **hard EM learning scheme**

Problem

Multi-mention Reading Comprehension

Q: Which composer did pianist Clara Wieck marry in 1840?
A: Robert Schumann

Robert Schumann was a German composer and influential music critics of the Romantic era. (…) Robert Schumann himself refers to it as “an affliction of the whole hand” (…) Robert Schumann is mentioned in a 1991 episode of Seinfeld “The Jacket” (…) Clara Schumann was a German musician and composer. Her husband was the composer Robert Schumann. (…) Brahms met Joachim in Hanover, made a very favorable impression on him, and got from him a letter of introduction to Robert Schumann.

From TriviaQA (Joshi et al. 2017)

Problem

Multi-mention Reading Comprehension

Q: Which composer did pianist Clara Wieck marry in 1840?
A: Robert Schumann

Robert Schumann was a German composer and influential music critics of the Romantic era. (…) Robert Schumann himself refers to it as "an affliction of the whole hand" (…) Robert Schumann is mentioned in a 1991 episode of Seinfeld "The Jacket" (…) Clara Schumann was a German musician and composer. Her husband was the composer Robert Schumann. (…) Brahms met Joachim in Hanover, made a very favorable impression on him, and got from him a letter of introduction to Robert Schumann.

From TriviaQA (Joshi et al. 2017)

Problem

Multi-mention Reading Comprehension

Q: Which composer did pianist Clara Wieck marry in 1840?
A: Robert Schumann

Given

Robert Schumann was a German composer and influential music critics of the Romantic era. (…) Robert Schumann himself refers to it as "an affliction of the whole hand" (…) Robert Schumann is mentioned in a 1991 episode of Seinfeld "The Jacket" (…) Clara Schumann was a German musician and composer. Her husband was the composer Robert Schumann. (…) Joachim in Hanover, made a very favorable impression on him, and got from him a letter of introduction to Robert Schumann.

Not Given

From TriviaQA (Joshi et al. 2017)

Problem

Multi-mention Reading Comprehension

Q: Which composer did pianist Clara Wieck marry in 1840?
A: Robert Schumann

Given

z1 Robert Schumann was a German composer and influential music critics of the Romantic er z2 .) Robert Schumann himself refers to it as “an affliction of the whole hand z3) Robert Schumann is mentioned in a 1991 episode of Seinfeld “The Jacket” (…) Clara Schumann was a German musician and composer. Her husband was the comp z4 Robert Schumann. (…) Brahms met Joachim in Hanover, made a very favorable impression on him, and got from him a letter of introductio z5 Robert Schumann.

From TriviaQA (Joshi et al. 2017)

Problem

Multi-mention Reading Comprehension

Q: Which composer did pianist Clara Wieck marry in 1840?
A: Robert Schumann

Given

z1 Robert Schumann was a German composer and influential music critics of the Romantic er z2 .) Robert Schumann himself refers to it as "an affliction of the whole hand z3) Robert Schumann is mentioned in a 1991 episode of Seinfeld "The Jacket" (…) Clara Schumann was a German musician and composer. Her husband was the comp z4 Robert Schumann. (…) Brahms met Joachim in Hanover, made a very favorable impression on him, and got from him a letter of introductio z5 Robert Schumann.

Input: Q, Document

Solution z (span in this case)

Output: A (text)

From TriviaQA (Joshi et al. 2017)

Problem

Multi-mention Reading Comprehension

Q: Which composer did pianist Clara Wieck marry in 1840?
A: Robert Schumann

Given

z1 Robert Schumann was a German composer and influential music critics of the Romantic er z2 .) Robert Schumann himself refers to it as "an affliction of the whole hand z3) Robert Schumann is mentioned in a 1991 episode of Seinfeld "The Jacket" (…) Clara Schumann was a German musician and composer. Her husband was the comp z4 Robert Schumann. (…) Brahms met Joachim in Hanover, made a very favorable impression on him, and got from him a letter of introductio z5 Robert Schumann.

Input: Q, Document

Solution z (span in this case)

Output: A (text)

From TriviaQA (Joshi et al. 2017)

Problem

Multi-mention Reading Comprehension

Q: Which composer did pianist Clara Wieck marry in 1840?
A: Robert Schumann **Given**

z1 Robert Schumann was a German composer and influential music critics of the Romantic er z2 .) Robert Schumann himself refers to it as "an affliction of the whole hand z3) Robert Schumann is mentioned in a 1991 episode of Seinfeld "The Jacket" (…) Clara Schumann was a German musician and composer. Her husband was the comp z4 Robert Schumann. (…) Brahms met Joachim in Hanover, made a very favorable impression on him, and got from him a letter of introductio z5 Robert Schumann.

Input: Q, Document | **Solution z** (span in this case) | **Output: A** (text)

We can find a solution set **Z**

Z = {z1, z2, z3, z4, z5}

From TriviaQA (Joshi et al. 2017)

Problem

Discrete Reasoning Task

Q: How many yards longer was Rob Bironas' longest field goal compared to John Carney's only field goal?
A: 4

Titans responded with Kicker Rob Bironas managing to get a 37 yard field goal. … In the third quarter Tennessee would draw close as Bironas kicked a 37 yard field goal. … John Carney getting a 36 yard field goal. … Young and Williams hooking up with each other on a 41 yard td pass. … Bironas nailing a 40 yard and a 25 yard field goal.

$$40 - 36 = 4$$

From DROP (Dua et al 2019)

Problem

Discrete Reasoning Task

Q: How many yards longer was Rob Bironas' longest field goal compared to John Carney's only field goal?
A: 4

Titans responded with Kicker Rob Bironas managing to get a 37 yard field goal. … In the third quarter Tennessee would draw close as Bironas kicked a 37 yard field goal. … John Carney getting a 36 yard field goal. … Young and Williams hooking up with each other on a 41 yard td pass. … Bironas nailing a 40 yard and a 25 yard field goal.

$$40 - 36 = 4$$

From DROP (Dua et al 2019)

Problem

Discrete Reasoning Task

Q: How many yards longer was Rob Bironas' longest field goal compared to John Carney's only field goal?
A: 4

Titans responded with Kicker Rob Bironas managing to get a 37 yard field goal. … In the third quarter Tennessee would draw close as Bironas kicked a 37 yard field goal. … John Carney getting a 36 yard field goal. … Young and Williams hooking up with each other on a 41 yard td pass. … Bironas nailing a 40 yard and a 25 yard field goal.

$$40 - 36 = 4$$

From DROP (Dua et al 2019)

Problem

Discrete Reasoning Task

Q: How many yards longer was Rob Bironas' longest field goal compared to John Carney's only field goal?
A: 4

Titans responded with Kicker Rob Bironas managing to get a 37 yard field goal. … In the third quarter Tennessee would draw close as Bironas kicked a 37 yard field goal. … John Carney getting a 36 yard field goal. … Young and Williams hooking up with each other on a 41 yard td pass. … Bironas nailing a 40 yard and a 25 yard field goal.

$$40 - 36 = 4$$

From DROP (Dua et al 2019)

Problem

Discrete Reasoning Task

Q: How many yards longer was Rob Bironas' longest field goal compared to John Carney's only field goal?

A: 4

Given

Titans responded with Kicker Rob Bironas managing to get a 37 yard field goal. … In the third quarter Tennessee would draw close as Bironas kicked a 37 yard field goal. … John Carney getting a 36 yard field goal. … Young and Williams hooking up with each other on a 41 yard td pass. … Bironas nailing a 40 yard and a 25 yard field goal.

40 - 36 = 4

Not Given

From DROP (Dua et al 2019)

Problem

Discrete Reasoning Task

Q: How many yards longer was Rob Bironas' longest field goal compared to John Carney's only field goal?

A: 4 **Given**

Titans responded with Kicker Rob Bironas managing to get a 37 yard field goal. … In the third quarter Tennessee would draw close as Bironas kicked a 37 yard field goal. … John Carney getting a 36 yard field goal. … Young and Williams hooking up with each other on a 41 yard td pass. … Bironas nailing a 40 yard and a 25 yard field goal.

Input: Q, Document → **Solution z** (equation in this case) → **Output: A** (text)

From DROP (Dua et al 2019)

Problem

Discrete Reasoning Task

Q: How many yards longer was Rob Bironas' longest field goal compared to John Carney's only field goal?

A: 4 **Given**

Titans responded with Kicker Rob Bironas managing to get a *37* yard field goal. … In the third quarter Tennessee would draw close as Bironas kicked a *37* yard field goal. … John Carney getting a *36* yard field goal. … Young and Williams hooking up with each other on a *41* yard td pass. … Bironas nailing a *40* yard and a *25* yard field goal.

Input: Q, Document

Solution z (equation in this case)

Output: A (text)

We can find a solution set **Z**

Z = { "41-37", "41-37", "40-36" }

From DROP (Dua et al 2019)

Related Work 1
– Reading Comprehension

1. Heuristics --- first span, random span
- Very competitive baseline (partially because of dataset bias)

2. Maximum Marginal Likelihood (MML)
- A latent variable learning method which maximizes $\sum_{z \in Z} P(z|Q, D)$

Example papers with first only: Joshi et al 2017, Tay et al 2018, Talmor & Berant 2019
Example papers with MML: Kadlec et al 2016, Swayamdipta et al 2018, Clark & Gardner 2018, Lee et al 2019

Related Work 1
– Reading Comprehension

1. Heuristics --- first span, random span
- Very competitive baseline (partially because of dataset bias)

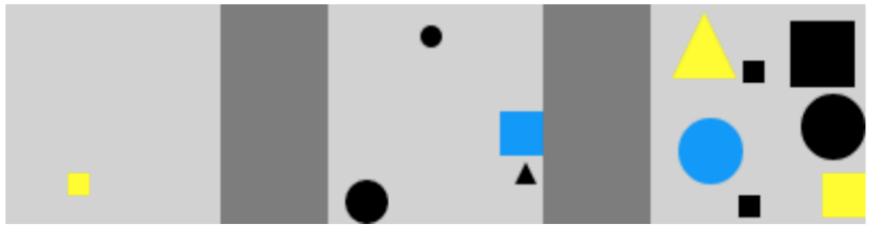
2. Maximum Marginal Likelihood (MML)
- A latent variable learning method which maximizes $\sum_{z \in Z} P(z|Q, D)$

We will show: First only & MML are similar; Our hard EM method outperforms them significantly

Example papers with first only: Joshi et al 2017, Tay et al 2018, Talmor & Berant 2019
Example papers with MML: Kadlec et al 2016, Swayamdipta et al 2018, Clark & Gardner 2018, Lee et al 2019

Related Work 2
– Semantic Parsing

(Zettlemoyer & Collins 2005, Liang et al 2013, Berant et al 2013, Artzi & Zettlemoyer 2013, …)



x : *There is a small yellow item not touching any wall*

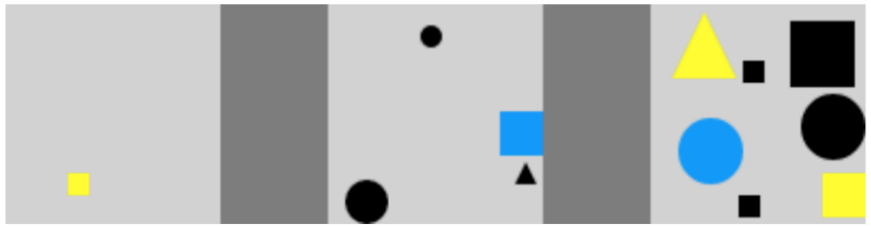
y : `True`

z : `Exist(Filter(ALL_ITEMS, λx.And(And(IsYellow(x), IsSmall(x)), Not(IsTouchingWall(x, Side.Any))))))`

CNLVR dataset (Suhr et al 2017); Figure from Goldman et al 2018

Related Work 2
– Semantic Parsing

(Zettlemoyer & Collins 2005, Liang et al 2013, Berant et al 2013, Artzi & Zettlemoyer 2013, …)



x : *There is a small yellow item not touching any wall*

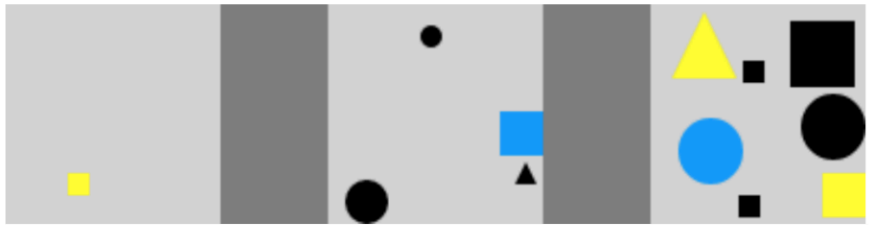
y : `True`

z :
```
Exist(Filter(ALL_ITEMS, λx.And(And(IsYellow(x),
  IsSmall(x)), Not(IsTouchingWall(x, Side.Any))))))
```

CNLVR dataset (Suhr et al 2017); Figure from Goldman et al 2018

Related Work 2
– Semantic Parsing

(Zettlemoyer & Collins 2005, Liang et al 2013, Berant et al 2013, Artzi & Zettlemoyer 2013, …)



x : *There is a small yellow item not touching any wall*

y : `True`

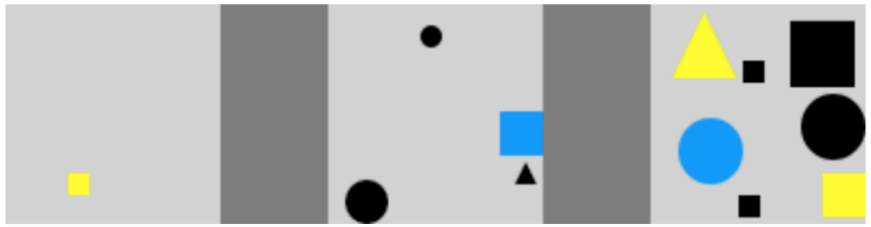
z : `Exist(Filter(ALL_ITEMS, λx.And(And(IsYellow(x), IsSmall(x)), Not(IsTouchingWall(x, Side.Any))))))`

Very large or infinite search space -> reward-based methods are used with no precomputed set of logical forms

CNLVR dataset (Suhr et al 2017); Figure from Goldman et al 2018

Related Work 2
– Semantic Parsing

(Zettlemoyer & Collins 2005, Liang et al 2013, Berant et al 2013, Artzi & Zettlemoyer 2013, …)



x : *There is a small yellow item not touching any wall*

y : `True`

z : `Exist(Filter(ALL_ITEMS, λx.And(And(IsYellow(x), IsSmall(x)), Not(IsTouchingWall(x, Side.Any))))))`

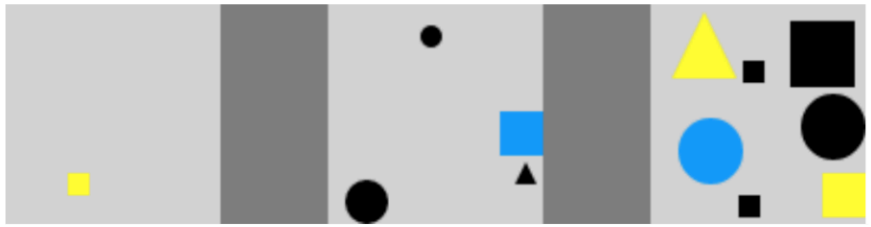
Very large or infinite search space -> reward-based methods are used with no precomputed set of logical forms

This paper: only focus on problems where a solution set can be precomputed

CNLVR dataset (Suhr et al 2017); Figure from Goldman et al 2018

Related Work 2
– Semantic Parsing

(Zettlemoyer & Collins 2005, Liang et al 2013, Berant et al 2013, Artzi & Zettlemoyer 2013, …)



x : *There is a small yellow item not touching any wall*

y : `True`

z : `Exist(Filter(ALL_ITEMS, λx.And(And(IsYellow(x), IsSmall(x)), Not(IsTouchingWall(x, Side.Any))))))`

Very large or infinite search space -> reward-based methods are used with no precomputed set of logical forms

This paper: only focus on problems where a solution set can be precomputed
→ Precomputing a solution set and using hard EM update is better than reward-based methods.

CNLVR dataset (Suhr et al 2017); Figure from Goldman et al 2018

Task Formulation

Reading comprehension

Discrete reasoning task

Semantic parsing

Task formulation

A unified weak supervision scenario with
a small set of possible solutions

Task Formulation

Reading comprehension | Discrete reasoning task | Semantic parsing

Q: Which composer did pianist Clara Wieck marry in 1840?
A: Robert Schumann

Robert Schumann was a German composer and influential music critics of the Romantic era. (…) Robert Schumann himself refers to it as "an affliction of the whole hand" (…) Robert Schumann is mentioned in a 1991 episode of Seinfeld "The Jacket" (…) Clara Schumann was a German musician and composer. Her husband was the composer Robert Schumann. (…) Brahms met Joachim in Hanover, made a very favorable impression on him, and got from him a letter of introduction to Robert Schumann.

Task Formulation

Reading comprehension | Discrete reasoning task | Semantic parsing

Q: Which composer did pianist Clara Wieck marry in 1840?
A: Robert Schumann

Average: 1.8 - 6.7
Median: 1 - 5

Robert Schumann was a German composer and influential music critics of the Romantic era. (…) Robert Schumann himself refers to it as "an affliction of the whole hand" (…) Robert Schumann is mentioned in a 1991 episode of Seinfeld "The Jacket" (…) Clara Schumann was a German musician and composer. Her husband was the composer Robert Schumann. (…) Brahms met Joachim in Hanover, made a very favorable impression on him, and got from him a letter of introduction to Robert Schumann.

Task Formulation

Reading comprehension | Discrete reasoning task | Semantic parsing

Q: How many yards longer was Rob Bironas' longest field goal compared to John Carney's only field goal? **A:** 4

Titans responded with Kicker Rob Bironas managing to get a 37 yard field goal. … In the third quarter Tennessee would draw close as Bironas kicked a 37 yard field goal. … John Carney getting a 36 yard field goal. … Young and Williams hooking up with each other on a 41 yard td pass. … Bironas nailing a 40 yard and a 25 yard field goal.

Task Formulation

Reading comprehension | Discrete reasoning task | Semantic parsing

Q: How many yards longer was Rob Bironas' longest field goal compared to John Carney's only field goal? **A:** 4

Titans responded with Kicker Rob Bironas managing to get a 37 yard field goal. … In the third quarter Tennessee would draw close as Bironas kicked a 37 yard field goal. … John Carney getting a 36 yard field goal. … Young and Williams hooking up with each other on a 41 yard td pass. … Bironas nailing a 40 yard and a 25 yard field goal.

37+37 37-37 37+36 37-36 37+41 37-41 36+41 36-41 …

Task Formulation

Reading comprehension | **Discrete reasoning task** | Semantic parsing

Q: How many yards longer was Rob Bironas' longest field goal compared to John Carney's only field goal? **A:** 4

Titans responded with Kicker Rob Bironas managing to get a 37 yard field goal. … In the third quarter Tennessee would draw close as Bironas kicked a 37 yard field goal. … John Carney getting a 36 yard field goal. … Young and Williams hooking up with each other on a 41 yard td pass. … Bironas nailing a 40 yard and a 25 yard field goal.

37+37 37-37 37+36 37-36 37+41 37-41 36+41 36-41 …

Execute

41-37 41-37 40-36 10-6 …

Task Formulation

Reading comprehension | Discrete reasoning task | Semantic parsing

Q: How many yards longer was Rob Bironas' longest field goal [illegible] Carney's only field goal? **A:** 4

Average: 8.2
Median: 3

Titans responded with Kicker Rob Bironas managing to get a 37 yard field goal. … In the third quarter Tennessee would draw close as Bironas kicked a 37 yard field goal. … John Carney getting a 36 yard field goal. … Young and Williams hooking up with each other on a 41 yard td pass. … Bironas nailing a 40 yard and a 25 yard field goal.

37+37 37-37 37+36 37-36 37+41 37-41 36+41 36-41 …

Execute

41-37 41-37 40-36 10-6 …

Task Formulation

Reading comprehension | Discrete reasoning task | Semantic parsing

Q: Which player played guard for Toronto in 1996-1997? **A:** John Long

Player	No.	Position	Year in Toronto
Kyle Lowry	3	Guard	2012-Present
John Long	25	Guard	1996-1997
Popeye Jones	54	Forward	1996-1998

All non-compositional SQL queries with up to 3 conditions

Task Formulation

Reading comprehension | Discrete reasoning task | Semantic parsing

Q: Which player played guard for Toronto in 1996-1997? **A:** John Long

Player	No.	Position	Year in Toronto
Kyle Lowry	3	Guard	2012-Present
John Long	25	Guard	1996-1997
Popeye Jones	54	Forward	1996-1998

Select player where No.="1996"
Select max(player) where No.="1996" …
Select min(player) where No.="1996"

Task Formulation

Reading comprehension | Discrete reasoning task | Semantic parsing

Q: Which player played guard for Toronto in 1996-1997? **A:** John Long

Player	No.	Position	Year in Toronto
Kyle Lowry	3	Guard	2012-Present
John Long	25	Guard	1996-1997
Popeye Jones	54	Forward	1996-1998

Select player where No.="1996"
Select max(player) where No.="1996"
Select min(player) where No.="1996"

Select player where No.="1997"
Select max(player) where No.="1997"
Select min(player) where No.="1997"

...

Task Formulation

Reading comprehension | Discrete reasoning task | Semantic parsing

Q: Which player played guard for Toronto in 1996-1997? **A:** John Long

Player	No.	Position	Year in Toronto
Kyle Lowry	3	Guard	2012-Present
John Long	25	Guard	1996-1997
Popeye Jones	54	Forward	1996-1998

Select player where No.="1996"
Select max(player) where No.="1996"
Select min(player) where No.="1996"

Select player where No.="1997"
Select max(player) where No.="1997"
Select min(player) where No.="1997"

…

Execute

Select player where position="guard" and Year in Toronto="1996-1997"
Select max(player) where position="guard" and Year in Toronto="1996-1997"
Select min(player) where position="guard" and Year in Toronto="1996-1997"
Select min(player) where position="guard"
Select min(player) where Year in Toronto="1996-1997"

Task Formulation

Reading comprehension | Discrete reasoning task | Semantic parsing

Q: Which player played guard for Toronto in 1996-1997?

Average: 346.1
Median: 5

Player	No.	Position	Year
Kyle Lowry	3	Guard	2012-Present
John Long	25	Guard	1996-1997
Popeye Jones	54	Forward	1996-1998

Select player where No.="1996"
Select max(player) where No.="1996"
Select min(player) where No.="1996"

Select player where No.="1997"
Select max(player) where No.="1997"
Select min(player) where No.="1997"

…

Execute

Select player where position="guard" and Year in Toronto="1996-1997"
Select max(player) where position="guard" and Year in Toronto="1996-1997"
Select min(player) where position="guard" and Year in Toronto="1996-1997"
Select min(player) where position="guard"
Select min(player) where Year in Toronto="1996-1997"

Task Formulation

Reading comprehension

Discrete reasoning task

Semantic parsing

Task formulation

A unified weak supervision scenario with
a small set of possible solutions

Task Formulation

Reading comprehension | Discrete reasoning task | Semantic parsing

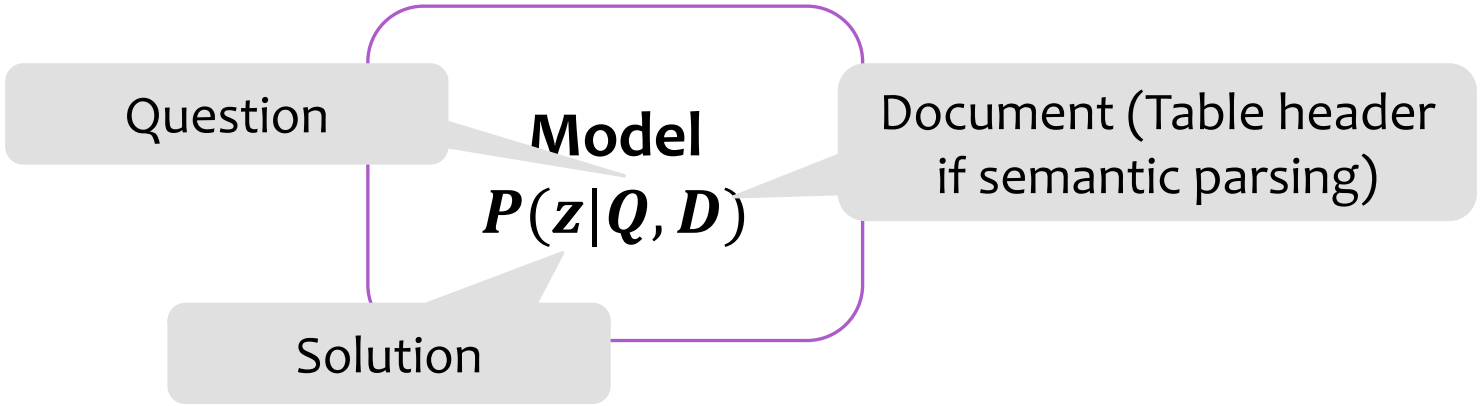
Task formulation

A unified weak supervision scenario with
a small set of possible solutions

Learning method

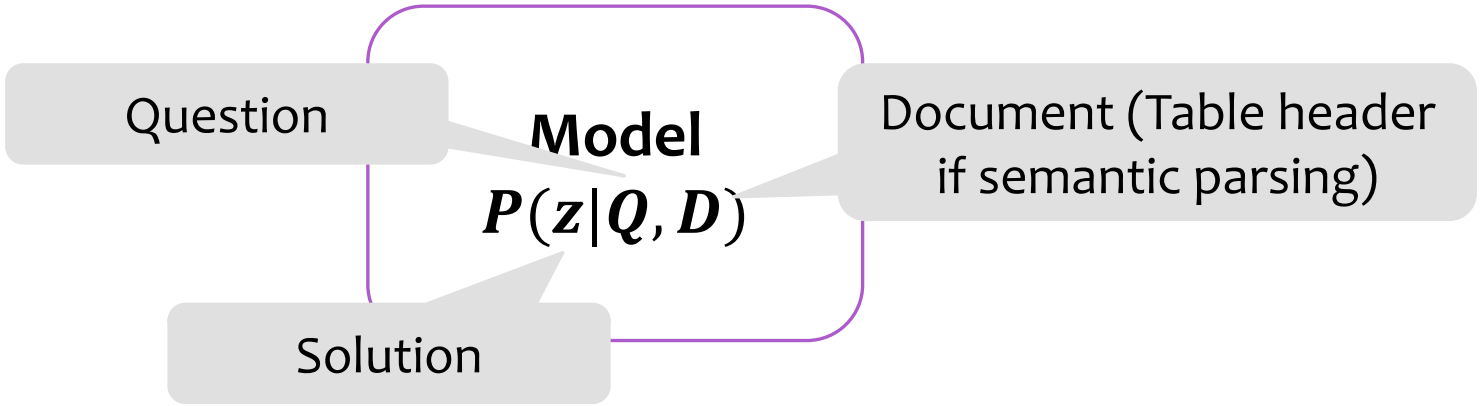
Learning

Goal: train $P(z|Q,D)$



Learning

Goal: train $P(z|Q,D)$
Groundtruth solution: $\bar{z}$

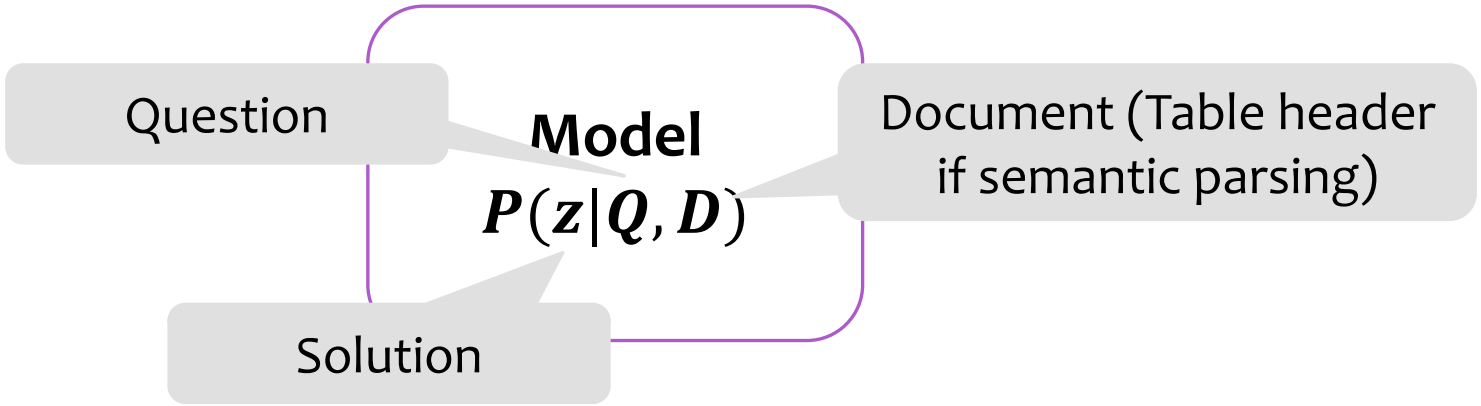


Supervised model (given $\bar{z}$) $\quad P(\bar{z}|Q,D)$

Learning

Goal: train $P(z|Q,D)$

~~Groundtruth solution: $\bar{z}$~~ $\{z_1, z_2, \ldots, z_n\}$ is a solution set executing the correct answer



Supervised model (given $\bar{z}$) $\mathrm{P}(\bar{z}|Q,D)$

Learning

Goal: train $P(z|Q,D)$

~~Groundtruth solution: $\bar{z}$~~ $\{z_1, z_2, \ldots, z_n\}$ is a solution set executing the correct answer

At each parameter update

Model
$\boldsymbol{P(z|Q,D)}$

z_1 0.32

z_2 0.58

z_n 0.01

Supervised model (given $\bar{z}$) $\mathrm{P}(\bar{z}|Q,D)$

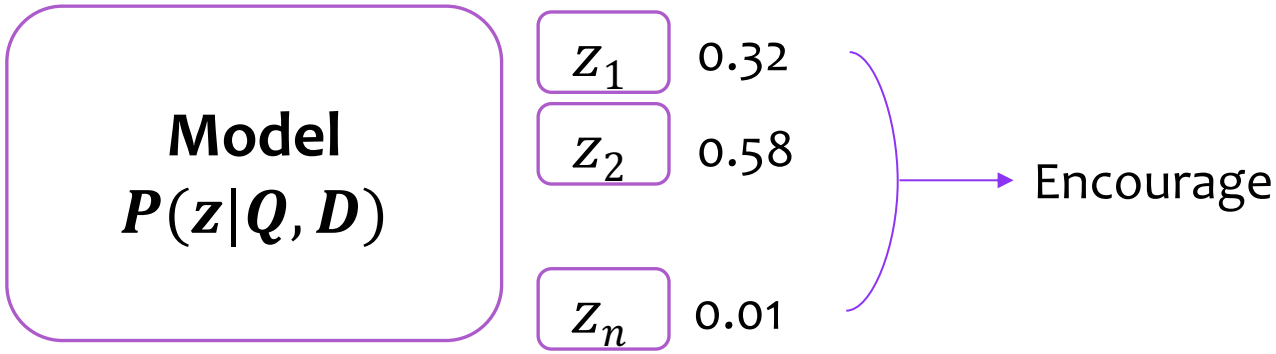
Learning - MML

Goal: train $P(z|Q,D)$

~~Groundtruth solution: $\bar{z}$~~ $\{z_1, z_2, \ldots, z_n\}$ is a solution set executing the correct answer

MML: Marginalize over z1,…, zn

At each parameter update



Supervised model (given $\bar{z}$) $\mathrm{P}(\bar{z}|Q,D)$

MML $\sum_{z \in Z} P(z|Q,D)$

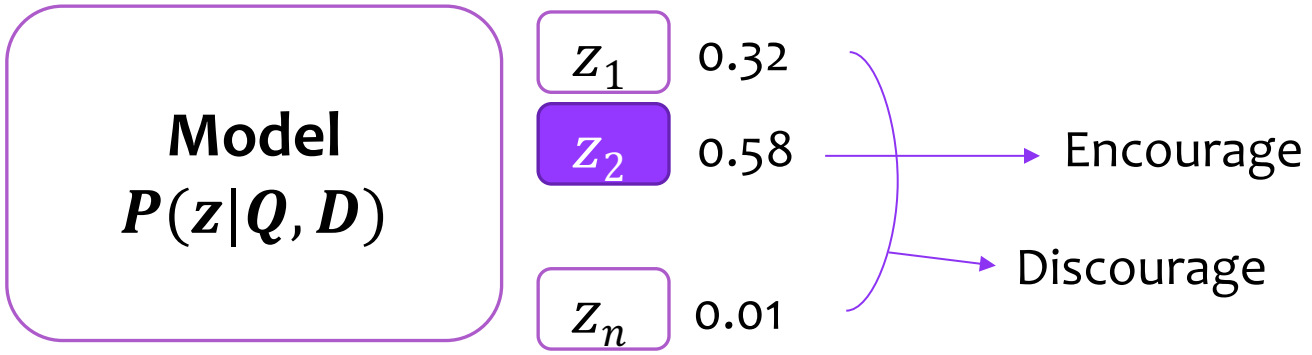
Learning - Ours

Goal: train $P(z|Q,D)$

~~Groundtruth solution: $\bar{z}$~~ $\{z_1, z_2, \ldots, z_n\}$ is a solution set executing the correct answer

Ours: Encourage the most likely solution

At each parameter update



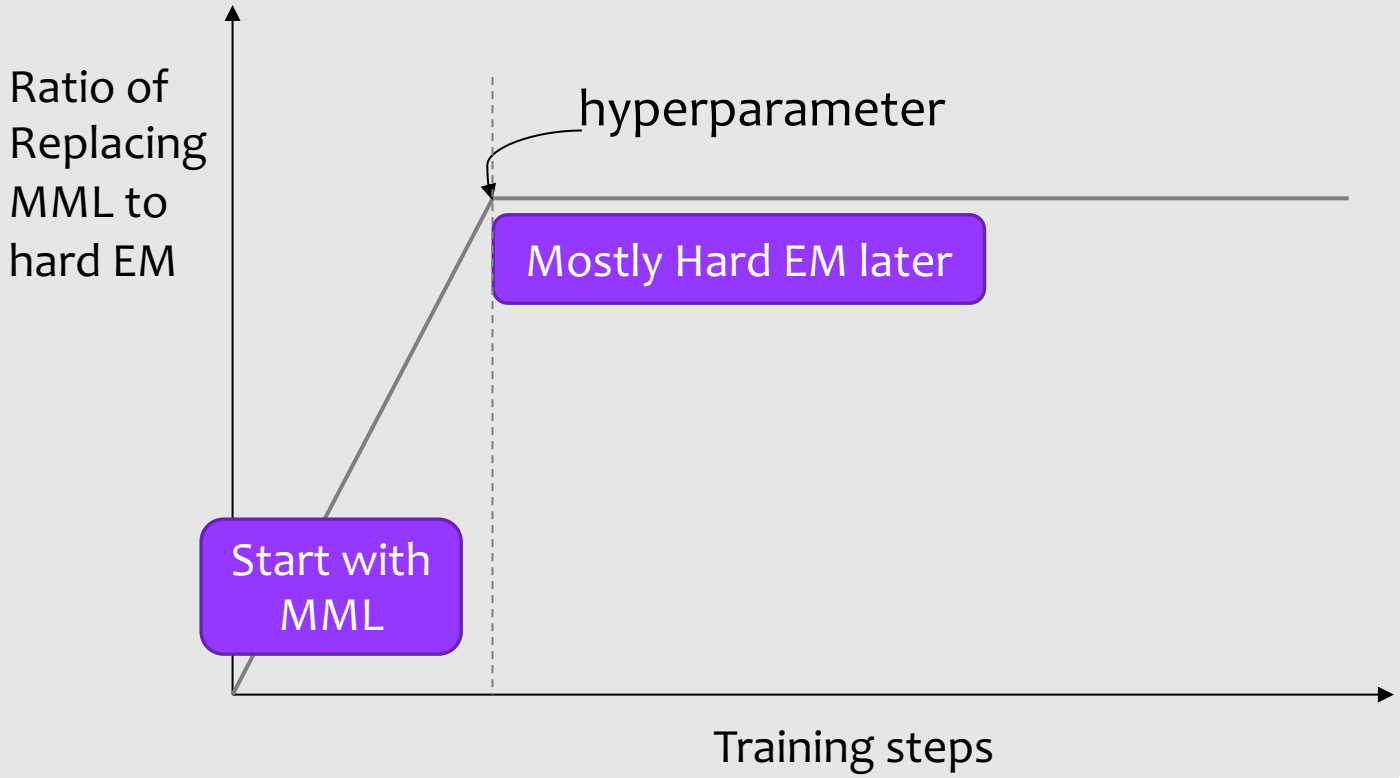
Supervised model (given $\bar{z}$)	$\mathrm{P}(\bar{z}\|Q,D)$
MML	$\sum_{z \in Z} P(z\|Q,D)$
Ours	$\max_{z \in Z} P(z\|Q,D)$

Learning - Ours

Goal: train $P(z|Q,D)$

~~Groundtruth solution: $\bar{z}$~~ $\{z_1, z_2, \ldots, z_n\}$ is a solution set executing the correct answer

In practice, we perform annealing:



Datasets

1. Multi-mention Reading Comprehension	
TriviaQA	Distantly-supervised RC
NarrativeQA	Generative RC
TriviaQA-open	Open-domain QA
Natural Questions-open	Open-domain QA
2. Discrete Reasoning Task	
DROP-num	Numeric reasoning
3. Semantic Parsing	
WikiSQL	Non-compositional SQL query generation

Datasets are from: Joshi et al 2017; Kocisky et al 2018; Joshi et al 2017; Kwiatkowski et al 2019; Dua et al 2019; Zhong et al 2017.
Note that TriviaQA-open & Natural Questions-open are open-domain versions of TriviaQA & Natural Questions, respectively.

Datasets

1. Multi-mention Reading Comprehension	
TriviaQA	Distantly-supervised RC
NarrativeQA	Generative RC
TriviaQA-open	Open-domain QA
Natural Questions-open	Open-domain QA
2. Discrete Reasoning Task	
DROP-num	Numeric reasoning
3. Semantic Parsing	
WikiSQL	Non-compositional SQL query generation

Base Model

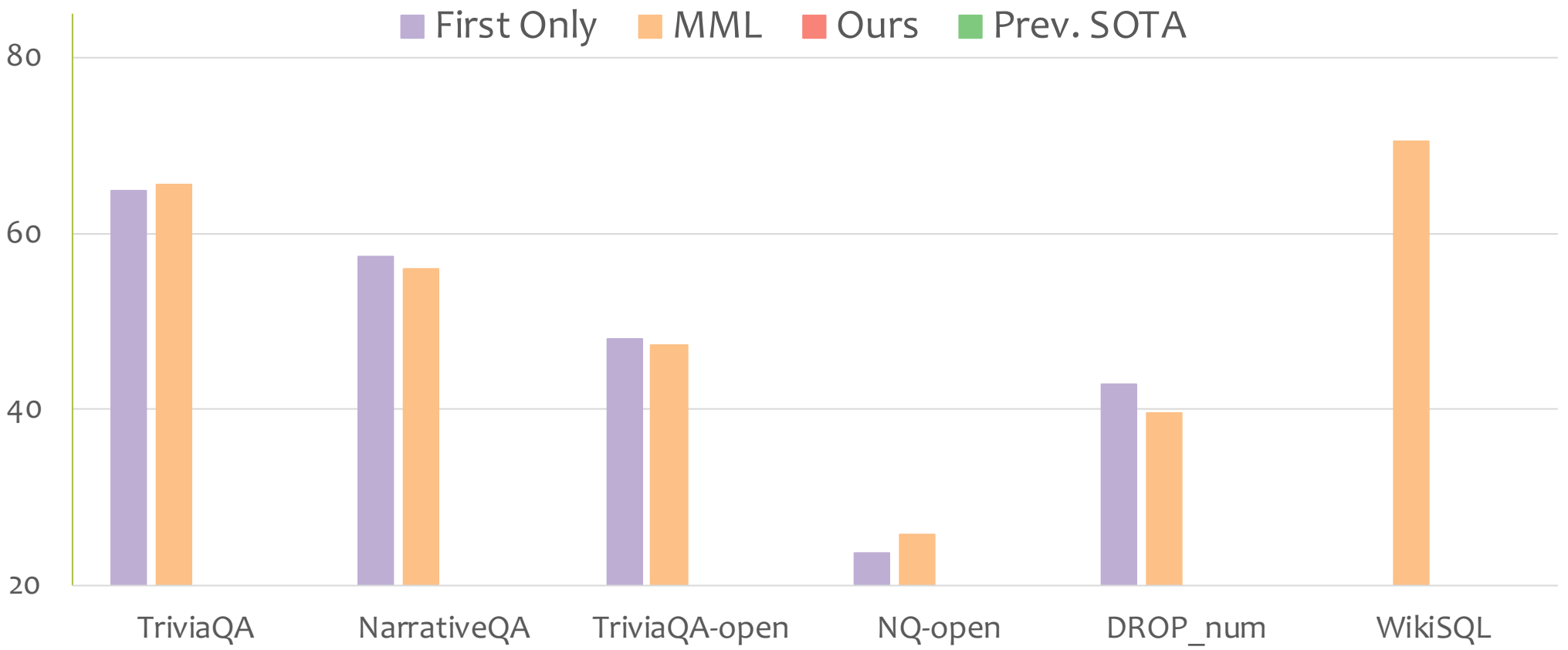
Multi-paragraph BERT-QA
(Devlin et al 2019 & others)

Augmented BERT
(Dua et al 2019)

SQLova
(Hwang et al 2019)

Datasets are from: Joshi et al 2017; Kocisky et al 2018; Joshi et al 2017; Kwiatkowski et al 2019; Dua et al 2019; Zhong et al 2017.
Note that TriviaQA-open & Natural Questions-open are open-domain versions of TriviaQA & Natural Questions, respectively.

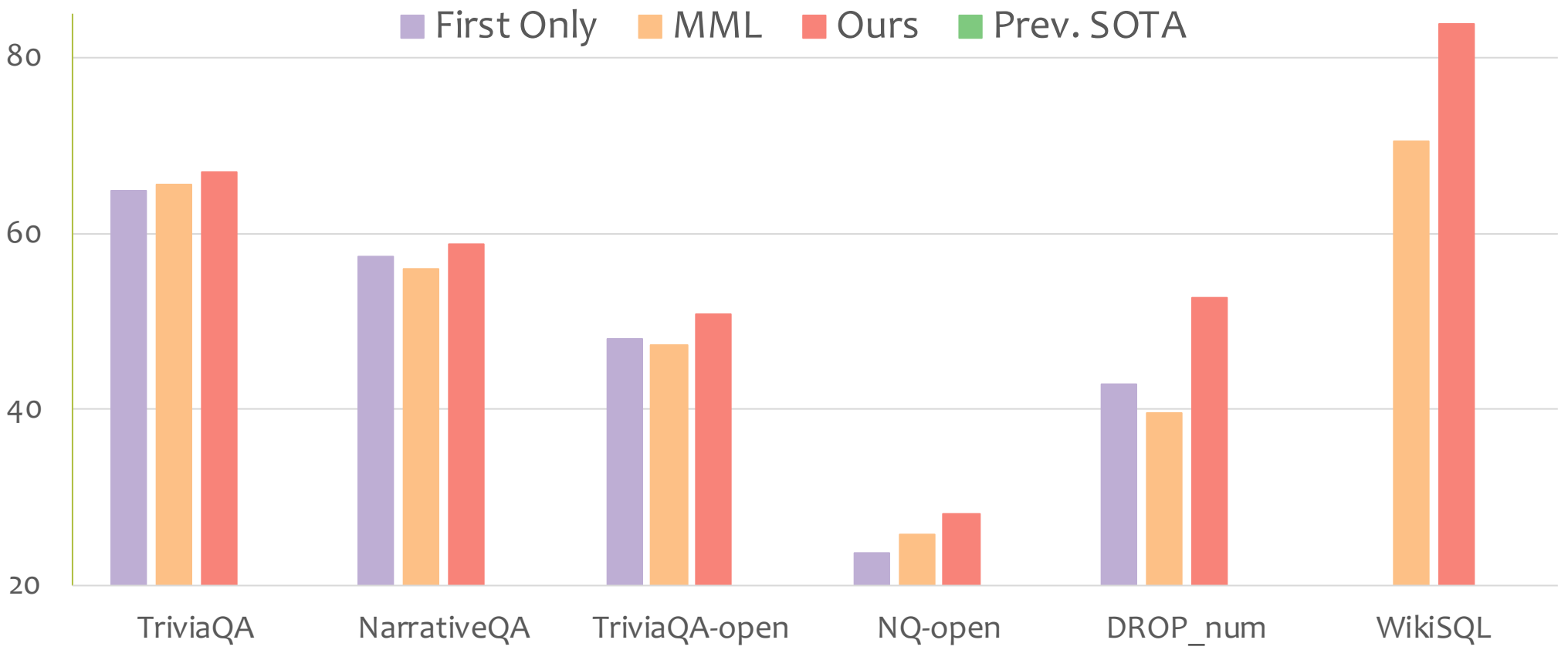
Result Summary



1) First-only and MML are similar.

SOTAs are from: Wang et al 2018; Nishida et al 2019; Lee et al 2019; Lee et al 2019 ; Dua et al 2019; Agarwal et al 2019

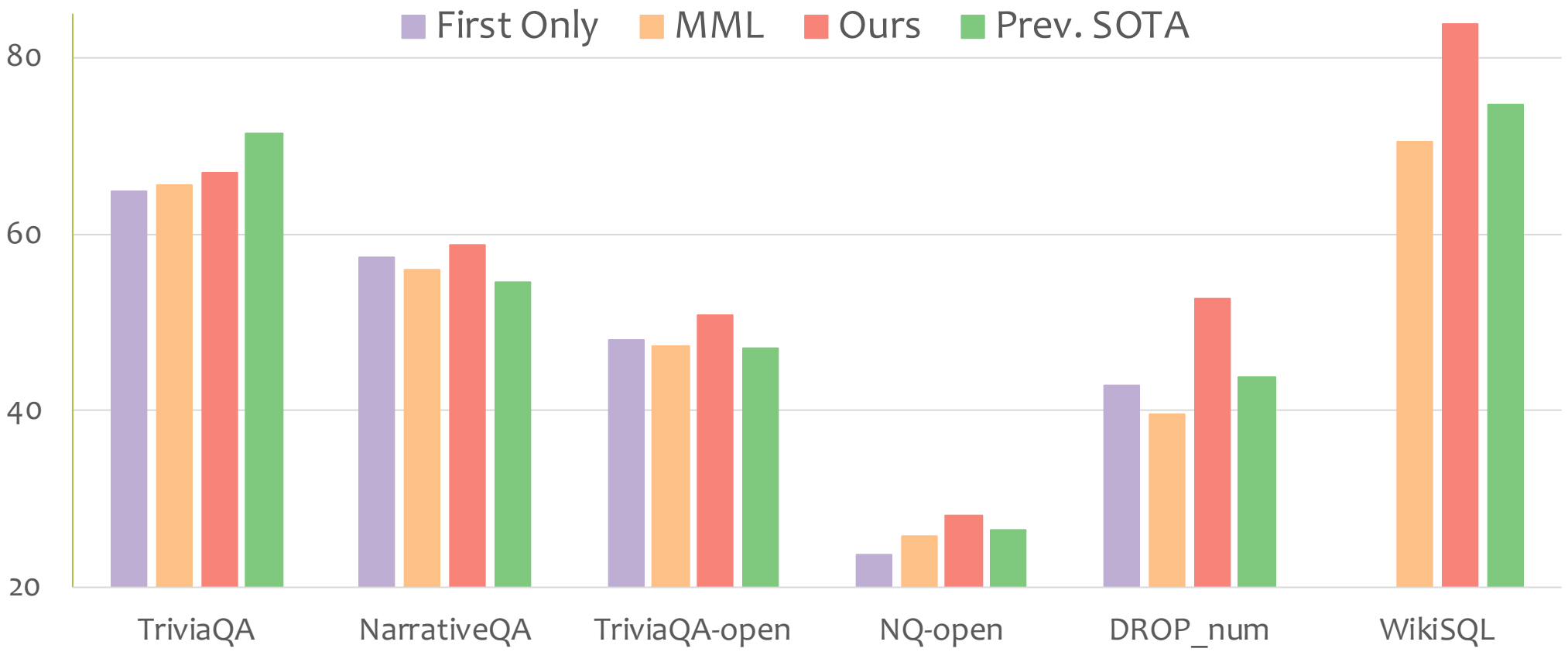
Result Summary



1) First-only and MML are similar.
2) Our Hard-EM method outperforms First-only & MML consistently.

SOTAs are from: Wang et al 2018; Nishida et al 2019; Lee et al 2019; Lee et al 2019 ; Dua et al 2019; Agarwal et al 2019

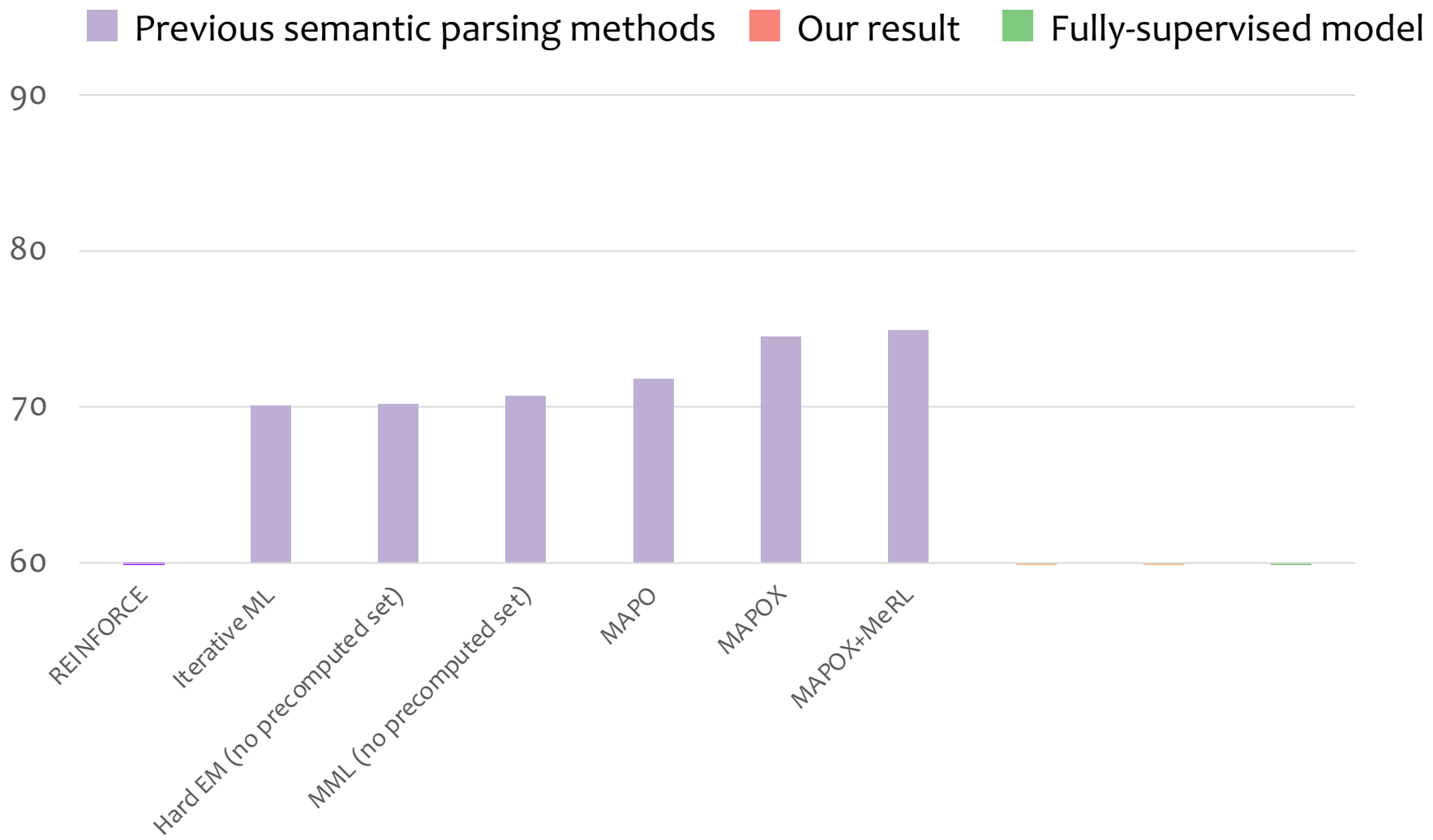
Result Summary



1) First-only and MML are similar.
2) Our Hard-EM method outperforms First-only & MML consistently.
3) SOTA on five datasets.

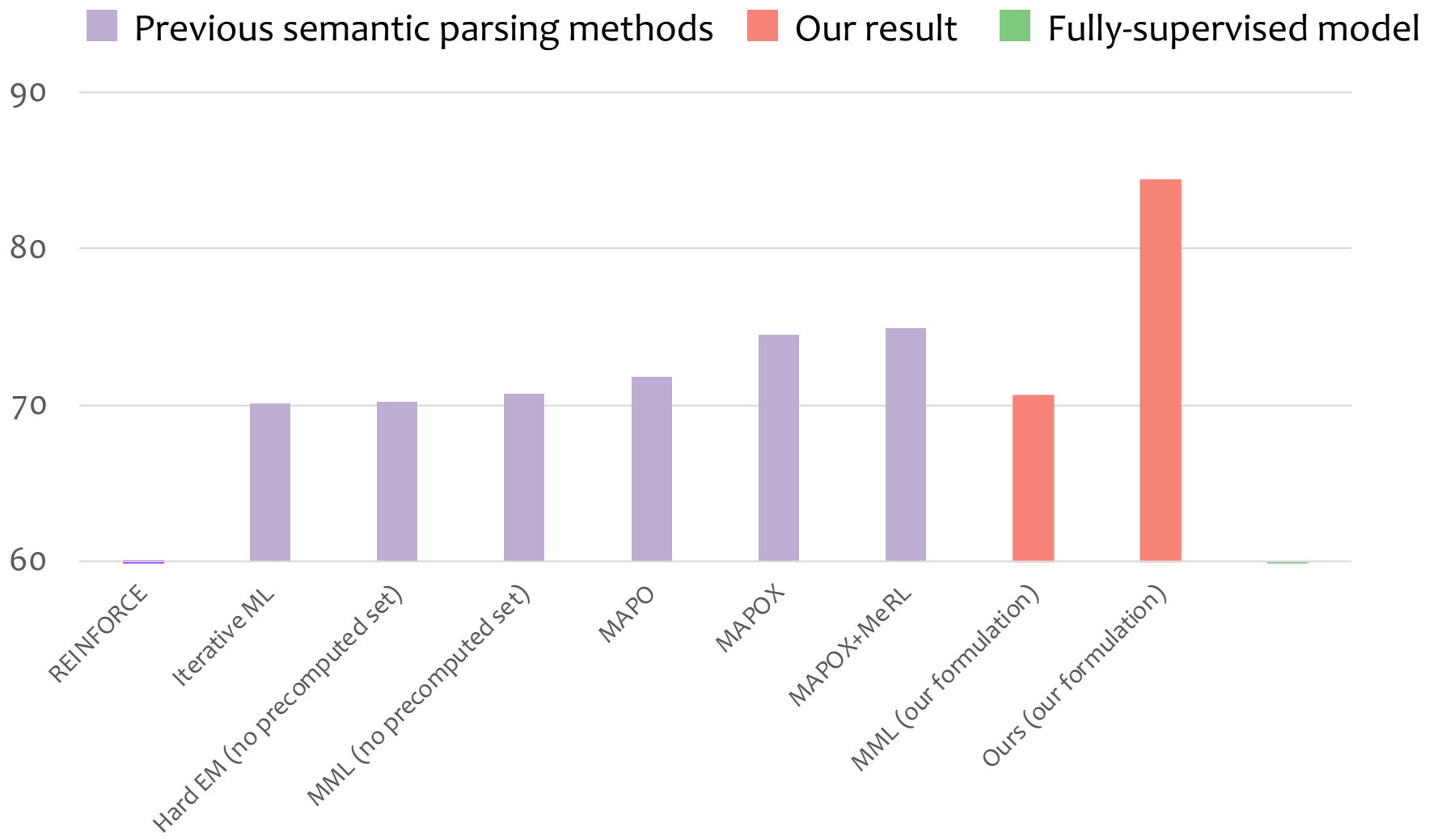
SOTAs are from: Wang et al 2018; Nishida et al 2019; Lee et al 2019; Lee et al 2019 ; Dua et al 2019; Agarwal et al 2019

Results - WikiSQL



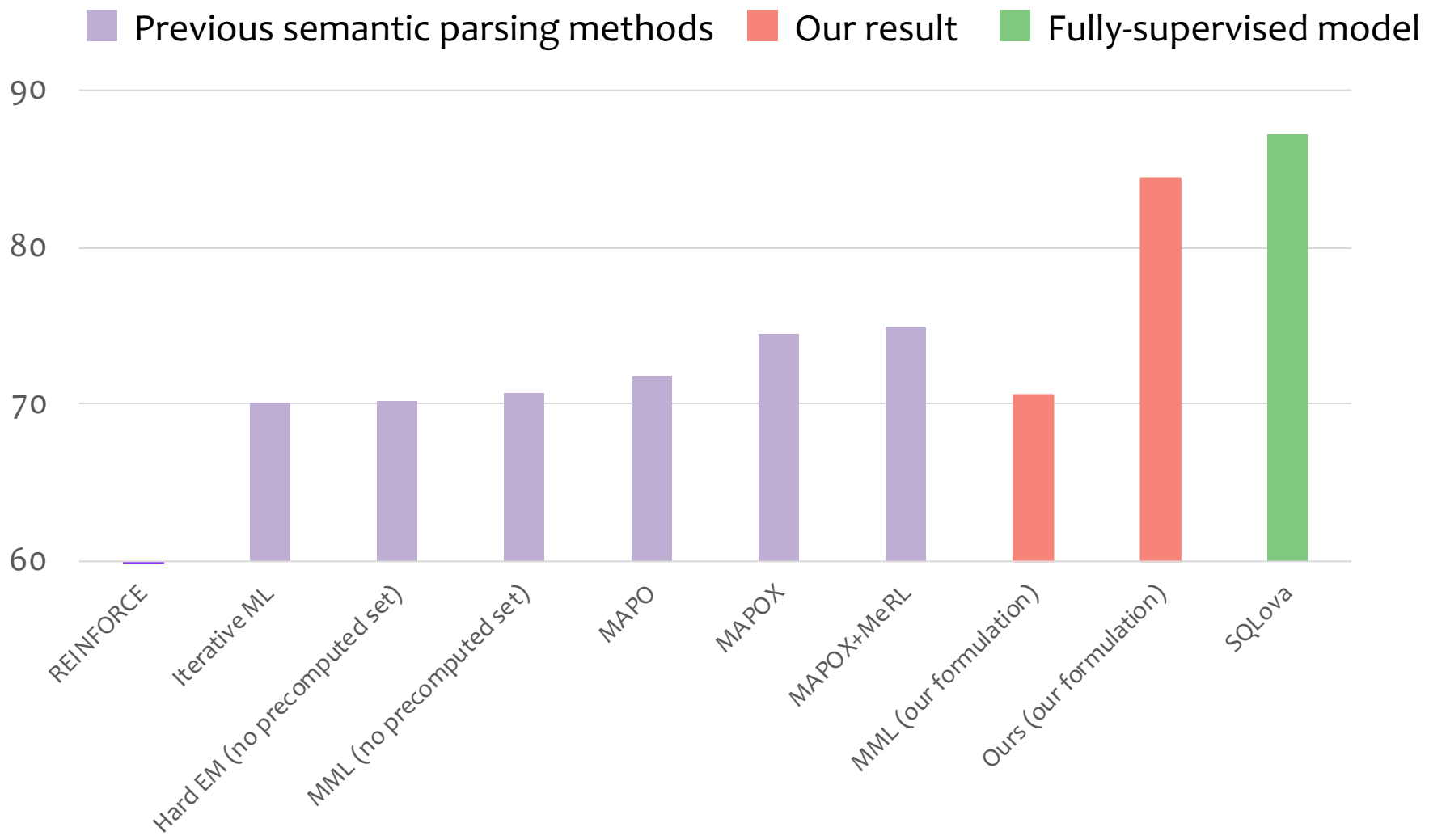
We outperform a wide range of previous semantic parsing methods.

Results - WikiSQL



We outperform a wide range of previous semantic parsing methods.

Results - WikiSQL



We outperform a wide range of previous semantic parsing methods.

How does likelihood change over training?

Discrete Reasoning Task as example

Q: How many yards longer was Rob Bironas' longest field goal compared to John Carney's only field goal? (**Answer:** 4)
P: ... The Titans responded with Kicker Rob Bironas managing to get a 37 yard field goal. ...Tennessee would draw close as Bironas kicked a 37 yard field goal. The Chiefs answered with kicker John Carney getting a 36 yard field goal. The Titans would retake the lead with Young and Williams hooking up with each other again on a 41 yard td pass. ...Tennessee clinched the victory with Bironas nailing a 40 yard and a 25 yard field goal.

Desired equation:

"40-36"

How does likelihood change over training?

Discrete Reasoning Task as example

Q: How many yards longer was Rob Bironas' longest field goal compared to John Carney's only field goal? (**Answer:** 4)
P: ... The Titans responded with Kicker Rob Bironas managing to get a 37 yard field goal. ...Tennessee would draw close as Bironas kicked a 37 yard field goal. The Chiefs answered with kicker John Carney getting a 36 yard field goal. The Titans would retake the lead with Young and Williams hooking up with each other again on a 41 yard td pass. ...Tennessee clinched the victory with Bironas nailing a 40 yard and a 25 yard field goal.

Desired equation:

"40-36"

Solution set:

{"41-37", "41-37", "40-36", "10-6", ...}

How does likelihood change over training?

Discrete Reasoning Task as example

Q: How many yards longer was Rob Bironas' longest field goal compared to John Carney's only field goal? (**Answer:** 4)
P: ... The Titans responded with Kicker Rob Bironas managing to get a 37 yard field goal. ...Tennessee would draw close as Bironas kicked a 37 yard field goal. The Chiefs answered with kicker John Carney getting a 36 yard field goal. The Titans would retake the lead with Young and Williams hooking up with each other again on a 41 yard td pass. ...Tennessee clinched the victory with Bironas nailing a 40 yard and a 25 yard field goal.

Training step

Top 1 prediction

t	Pred	Z (ordered by $\mathbb{P}(z\|x;\theta_t)$)			
1k	10-9	10-6	41-37	40-36	41-37‡
2k	37-36	40-36	41-37	41-37‡	10-6
4k	40-36	40-36	41-37‡	41-37	10-6
8k	40-36	40-36	41-37‡	41-37	10-6
16k	37-36	40-36	41-37	41-37‡	10-6
32k	40-36	40-36	41-37	41-37‡	10-6

Solution set (ordered by likelihood)

How does likelihood change over training?

Discrete Reasoning Task as example

Q: How many yards longer was Rob Bironas' longest field goal compared to John Carney's only field goal? (**Answer:** 4)
P: ... The Titans responded with Kicker Rob Bironas managing to get a 37 yard field goal. ...Tennessee would draw close as Bironas kicked a 37 yard field goal. The Chiefs answered with kicker John Carney getting a 36 yard field goal. The Titans would retake the lead with Young and Williams hooking up with each other again on a 41 yard td pass. ...Tennessee clinched the victory with Bironas nailing a 40 yard and a 25 yard field goal.

Training step

Top 1 prediction

t	Pred	Z (ordered by $\mathbb{P}(z\|x;\theta_t)$)			
1k	10-9	10-6	41-37	40-36	41-37‡
2k	37-36	40-36	41-37	41-37‡	10-6
4k	40-36	40-36	41-37‡	41-37	10-6
8k	40-36	40-36	41-37‡	41-37	10-6
16k	37-36	40-36	41-37	41-37‡	10-6
32k	40-36	40-36	41-37	41-37‡	10-6

Solution set (ordered by likelihood)

Correct equation is ranked first since the early stage of training.

How does likelihood change over training?

Discrete Reasoning Task as example

Q: How many yards longer was Rob Bironas' longest field goal compared to John Carney's only field goal? (**Answer:** 4)
P: ... The Titans responded with Kicker Rob Bironas managing to get a 37 yard field goal. ...Tennessee would draw close as Bironas kicked a 37 yard field goal. The Chiefs answered with kicker John Carney getting a 36 yard field goal. The Titans would retake the lead with Young and Williams hooking up with each other again on a 41 yard td pass. ...Tennessee clinched the victory with Bironas nailing a 40 yard and a 25 yard field goal.

Training step

Top 1 prediction

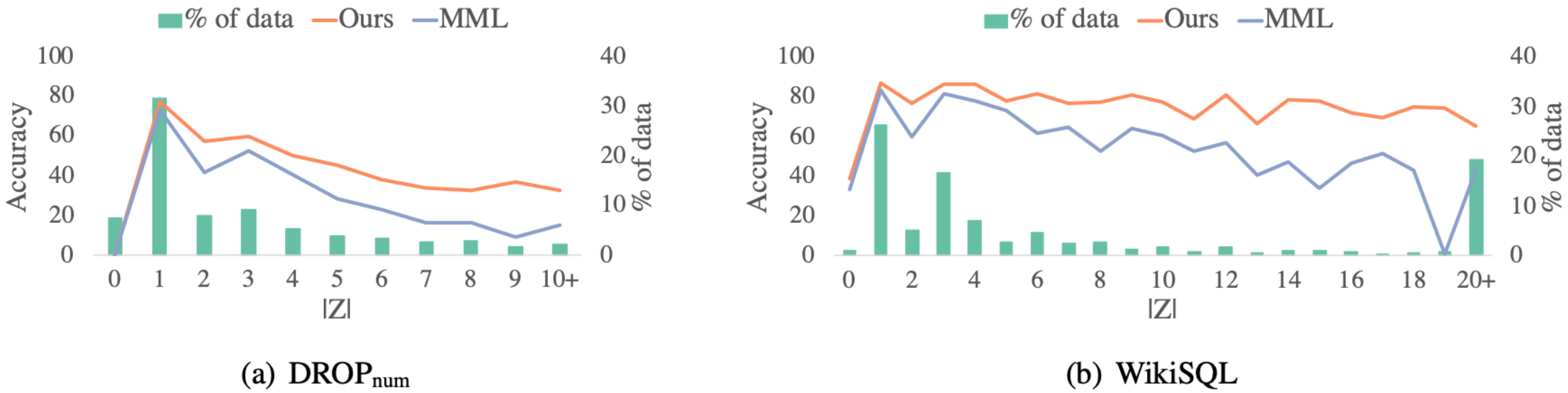
t	Pred	Z (ordered by $\mathbb{P}(z\|x;\theta_t)$)			
1k	10-9	10-6	41-37	40-36	41-37‡
2k	37-36	40-36	41-37	41-37‡	10-6
4k	40-36	40-36	41-37‡	41-37	10-6
8k	40-36	40-36	41-37‡	41-37	10-6
16k	37-36	40-36	41-37	41-37‡	10-6
32k	40-36	40-36	41-37	41-37‡	10-6

Solution set (ordered by likelihood)

Correct equation is ranked first since the early stage of training.

"Pushing hard towards the most likely solution is helpful"

Effect of solution set size



(a) DROP_{num} (b) WikiSQL

Figure 2: **Varying the size of solution set ($|Z|$) at test time.** We compare the model trained on MML objective (blue) and our training strategy (orange). Our approach consistently outperforms MML on DROP_{num} and WIKISQL, especially when $|Z|$ is large.

More performance gains when the size of solution set is large!

Effect of solution set size

More analyses in the paper!

Figure 2: **Varying [illegible] size of solution set ($|Z|$) at test time.** We compare the model trained on MML objective (blue) and our training strategy (orange). Our approach consistently outperforms MML on DROP_{num} and WIKISQL, especially when $|Z|$ is large.

More performance gains when the size of solution set is large!

Summary

We formulate various QA problems into a weak supervision problem where a **solution** is not given, but **a small set of potential solutions** can be precomputed.

We develop **a hard EM learning scheme** that computes gradients relative to the most likely solution at each parameter update.

Our method outperforms baselines significantly across 6 datasets, and set new SOTA on 5 datasets by only modifying the objective.

Use cases already!

A larger solution set with more extensive search, and further improved hard EM which encourages one or zero solution using thresholding.

	EM	F1
Hard EM with thresholding	**80.58**	**83.42**
Hard EM	73.72	77.46
Maximum Likelihood	63.96	67.98

Table 7: Results of different training algorithms on DROP development set.

Anonymous. "Neural Symbolic Reader: Scalable Integration of Distributed And Symbolic Representations for Reading Comprehension". **Submitted to ICLR 2020.**

(Disclaimer: we do not know who the authors are 😅)

Thank you for listening

Code https://github.com/shmsw25/qa-hard-em
Paper https://arxiv.org/abs/1909.04849
Contact sewon@cs.washington.edu