A Split-and-Recombine Approach for Follow-up Query Analysis

Speaker：Qian Liu (刘乾)

Date : 2019.12.22

北京航空航天大學 BEIHANG UNIVERSITY
Microsoft Research 微软亚洲研究院

Outline

1. Introduction
2. Methods
3. Our Research
4. Take-home-message
5. Reference

1. Introduction

Database

Brand	Sales	Profit	Year
BMW	31020	5000	2009
Ford	25220	3000	2009
Benz	47060	6000	2009

Text-to-SQL

show the sales of BMW in 2009

SELECT Sales WHERE Brand = BMW and Year = 2009

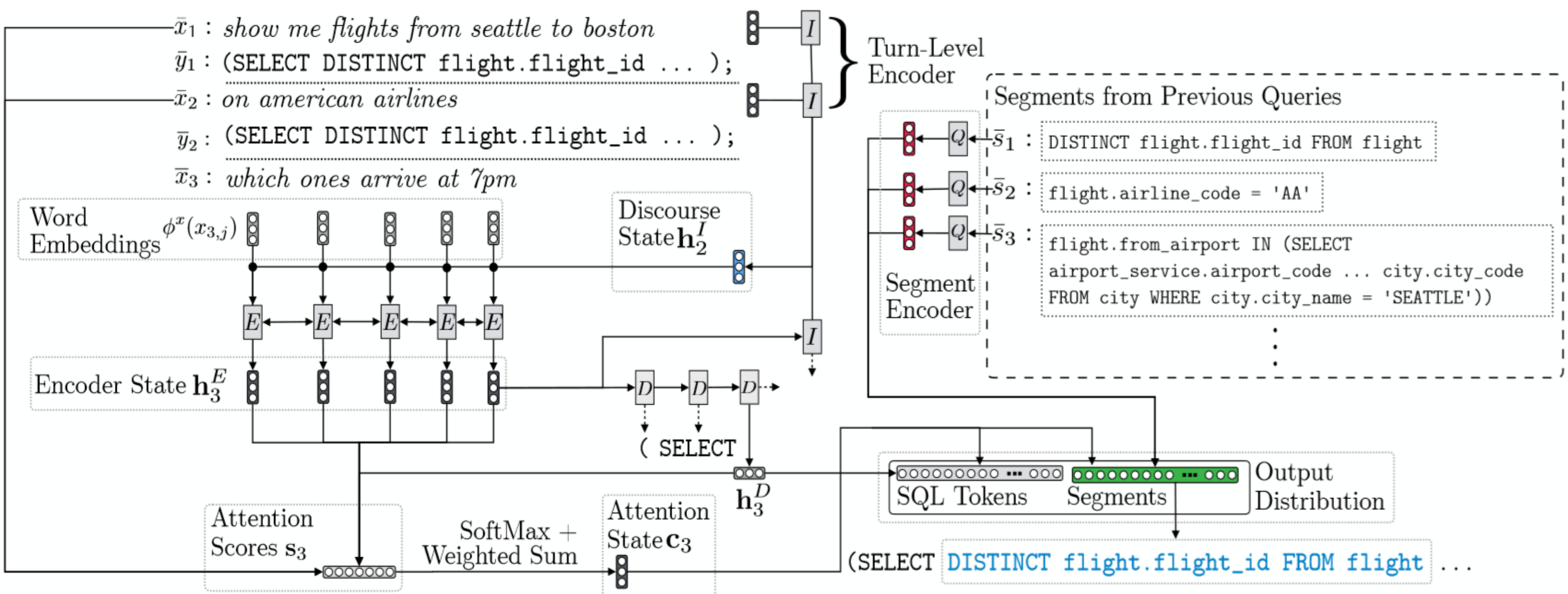
Conversation

what about profit ?

SELECT Profit WHERE Brand = BMW and Year = 2009

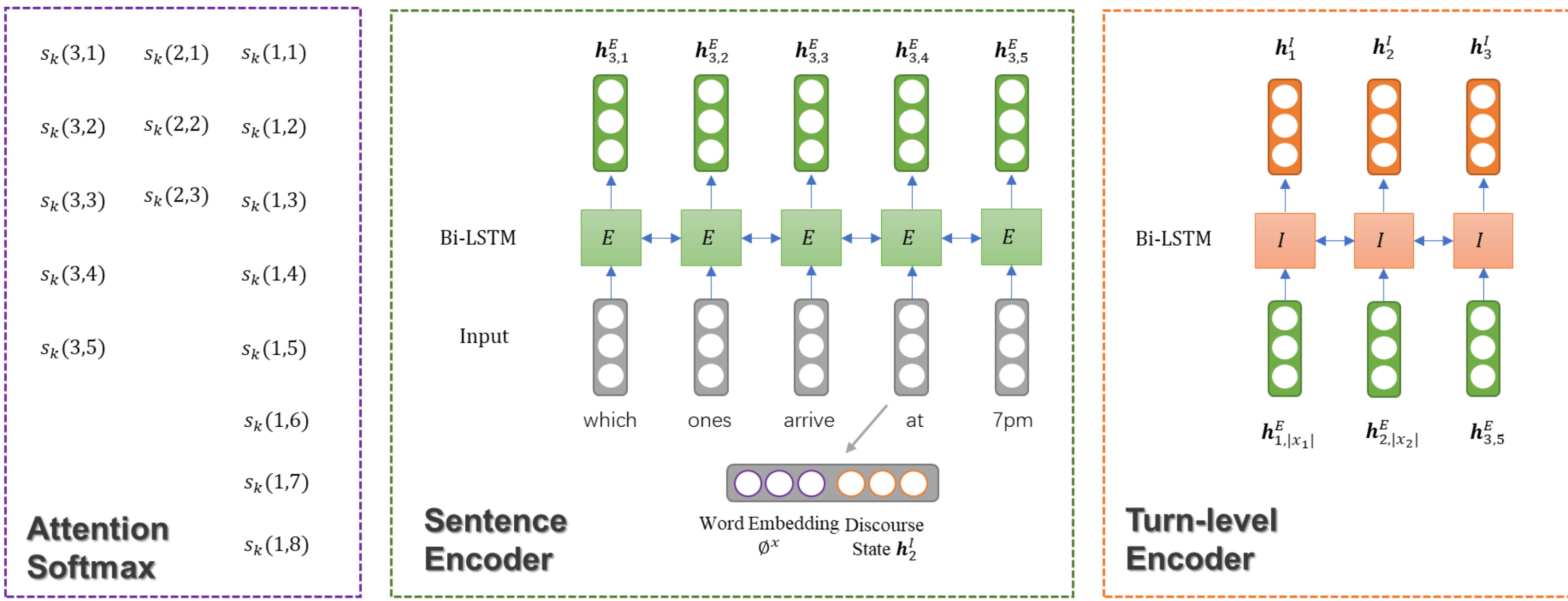
2. Methods

Traditional methods always model the problem as an end-to-end manner[1].



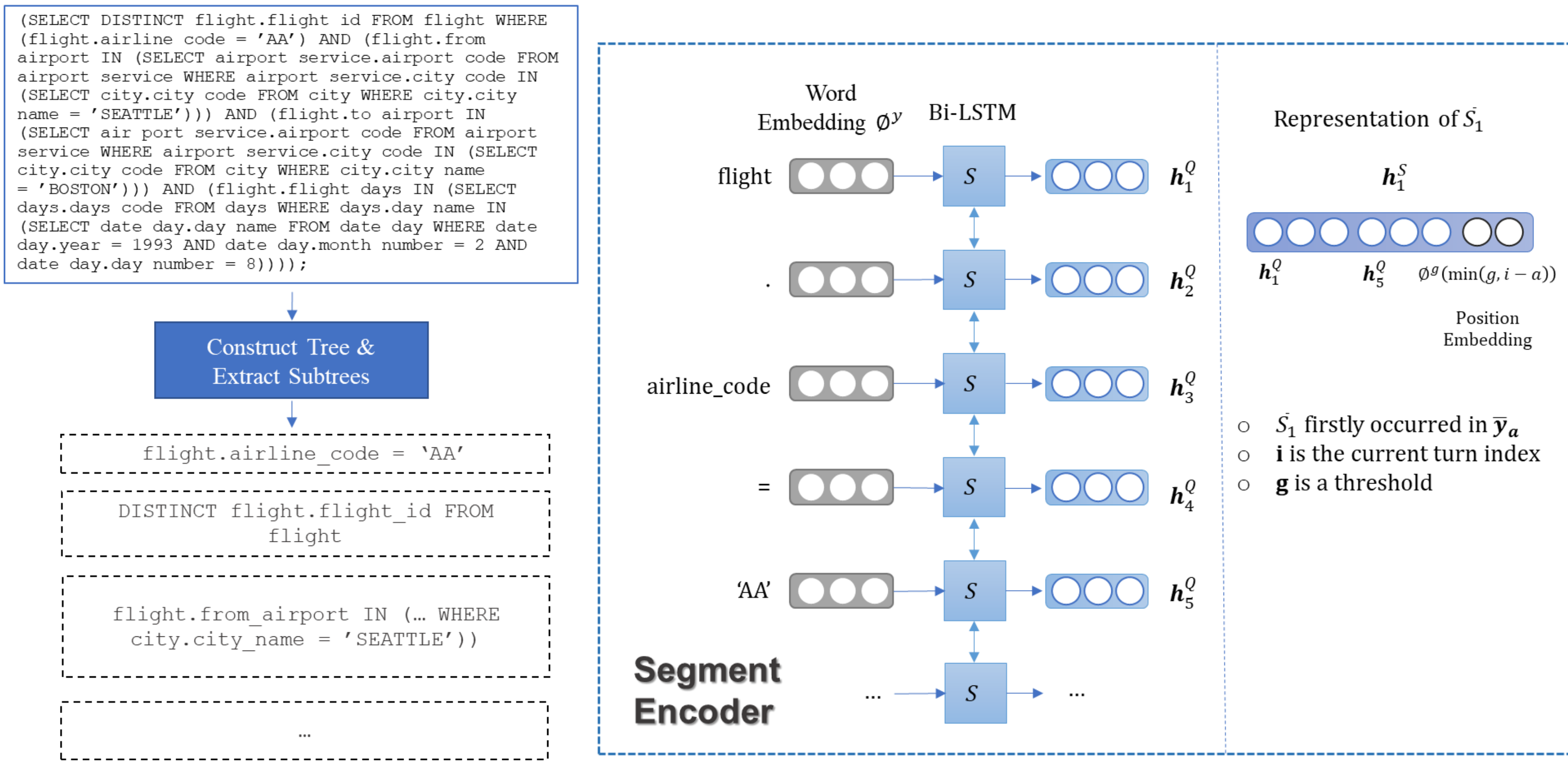
2. Methods

- **Turn-level encoder**: hierarchically represent the dialogue history



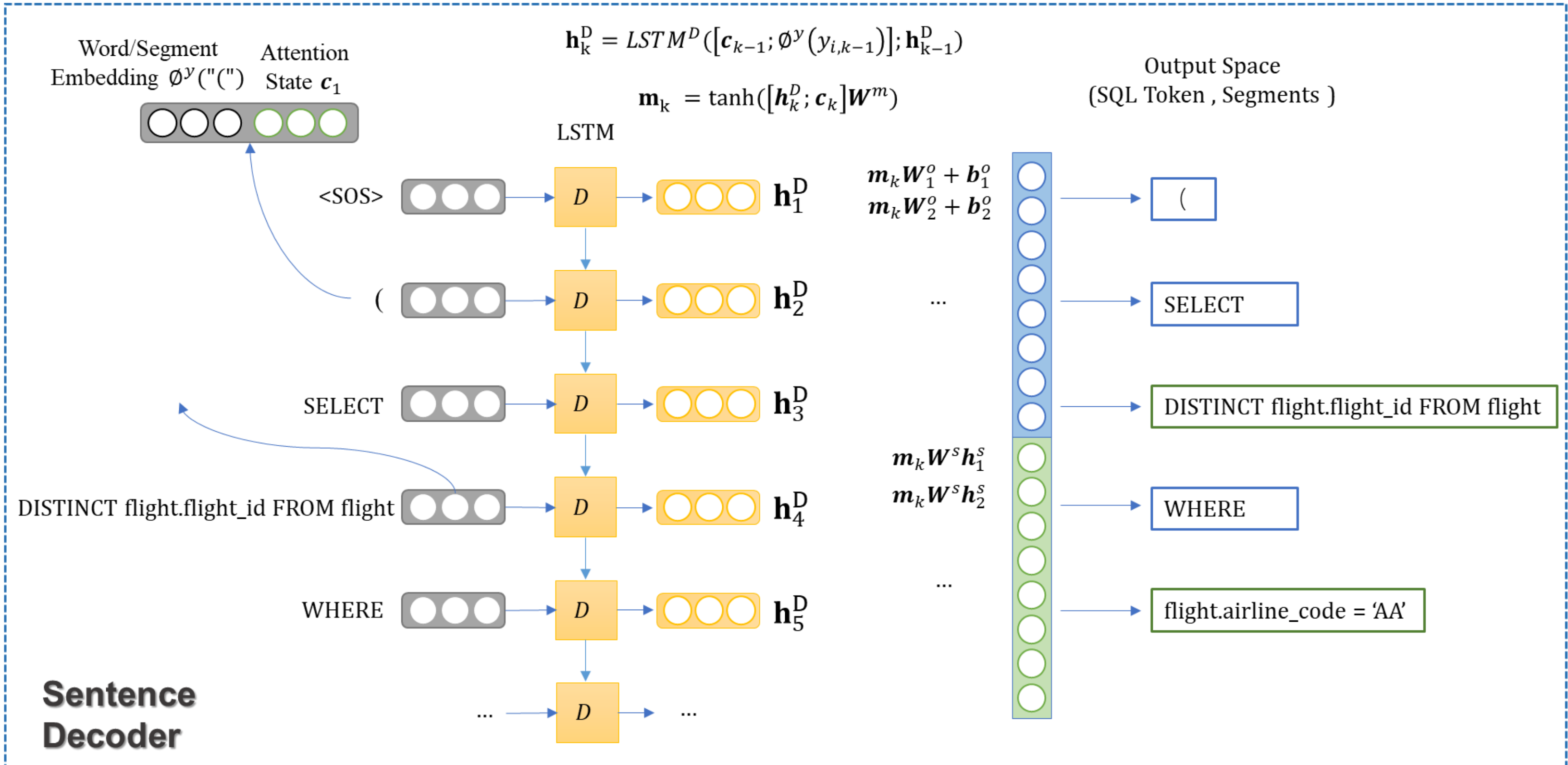
2. Methods

- **SQL Segment Encoder**: reduce decoding steps by a large margin



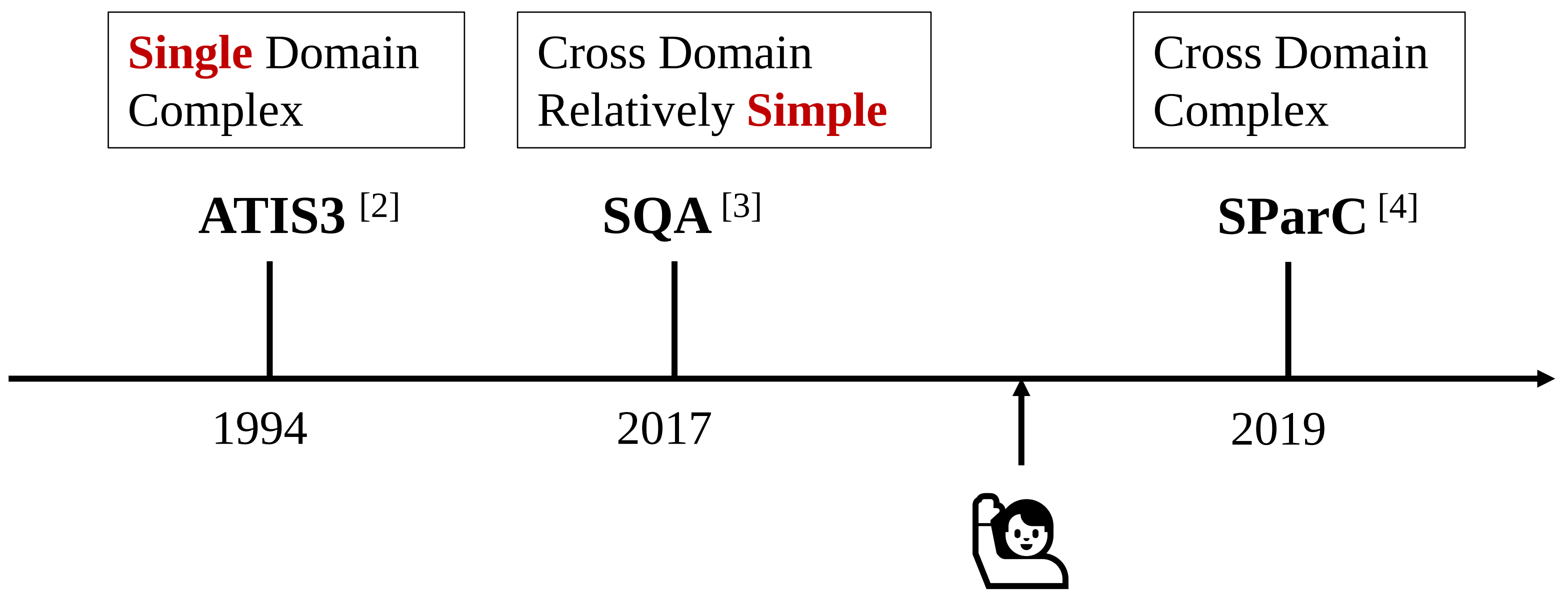
2. Methods

- **Joint Decoding:** decode SQL Token & Segment jointly



2. Methods

Previous works focus on end-to-end context-dependent parsing (e.g. Conversational Text-to-SQL). However, related datasets are very rare .



2. Methods

However, semantic parsing techniques obtained a large attention.
Therefore, we have a question:

Could we use trained semantic parsers to solve
the context-dependent semantic parsing problem?

2. Methods

However, semantic parsing techniques obtained a large attention.
Therefore, we have a question:

Could we use trained semantic parsers to solve
the context-dependent semantic parsing problem?

Yes, we could make it when imposing an auxiliary task

Follow-up Query Analysis (FQA)

2. Methods

Brand	Sales	Profit	Year
BMW	31020	5000	2009
Ford	25220	3000	2009
Benz	47060	6000	2009

Precedent Query ← show the sales of BMW in 2009

SELECT Sales WHERE Brand = BMW and Year = 2009

Follow-up Query ← what about profit

Restate Query ← show the profit of BMW in 2009.

SELECT Profit WHERE Brand = BMW and Year = 2009

2. Methods

- Previous works aims to achieve the task end-to-end.
- We aim to use a pretrained semantic parser to do it.
- The annotation of FQA is cheaper than conversational semantic parsing.

<u>Previous Work</u>

Context-dependent Parsing

<u>Our Method</u>

Follow-up Query Analysis → Parsing

2. Methods

Scenario	Example
Analytics	Precedent : In 1995, is there any network named CBC? Follow-up : Any TSN? Fused : In 1995, is there any network named TSN?
Compare	Precedent : How much money has Smith earned? Follow-up : Compare it with Bill Collins. Fused : Compare money Smith earned with Bill Collins.
Calc & Stats	Precedent : List all universities founded before 1855. Follow-up : Show their number. Fused : Show the number of all universities founded before 1855.
Extremum	Precedent : Which stadium has the most capacity? Follow-up : Which get the highest attendance? Fused : Which stadium get the highest attendance?
Filter	Precedent : How many roles are from studio paramount? Follow-up : List all titles produced by that studio. Fused : List all titles produced by studio paramount.
Group	Precedent : Show the industry which has the most companies? Follow-up : Show in different countries. Fused : Show the industry which has the most companies in different countries.
Sort	Precedent : Show all chassis produced after the year 1990. Follow-up : Sort them by year. Fused : Show all chassis produced after the year 1990 and sort by year.
Search	Precedent : What position did Sid O'Neill play? Follow-up : Which players else are in the same position? Fused : Which players play in the position of Sid O'Neill excluding Sid O'Neill?

2. Methods

The challenge turns to be:

How to perform the task FQA ?

2. Methods

The restate query always has **a large overlap** with the input queries, hence we have following options:

- Sequence to sequence with attention
- Copy mechanism (CopyNet/Point Generator) [5]
- Concatenate

show the sales of BMW in 2009

what about profit

→ show the profit of BMW in 2009

2. Methods

The restate query contains **anaphora** which refers to the entities in the input queries, we consider:

- End to end coreference resolution [6]

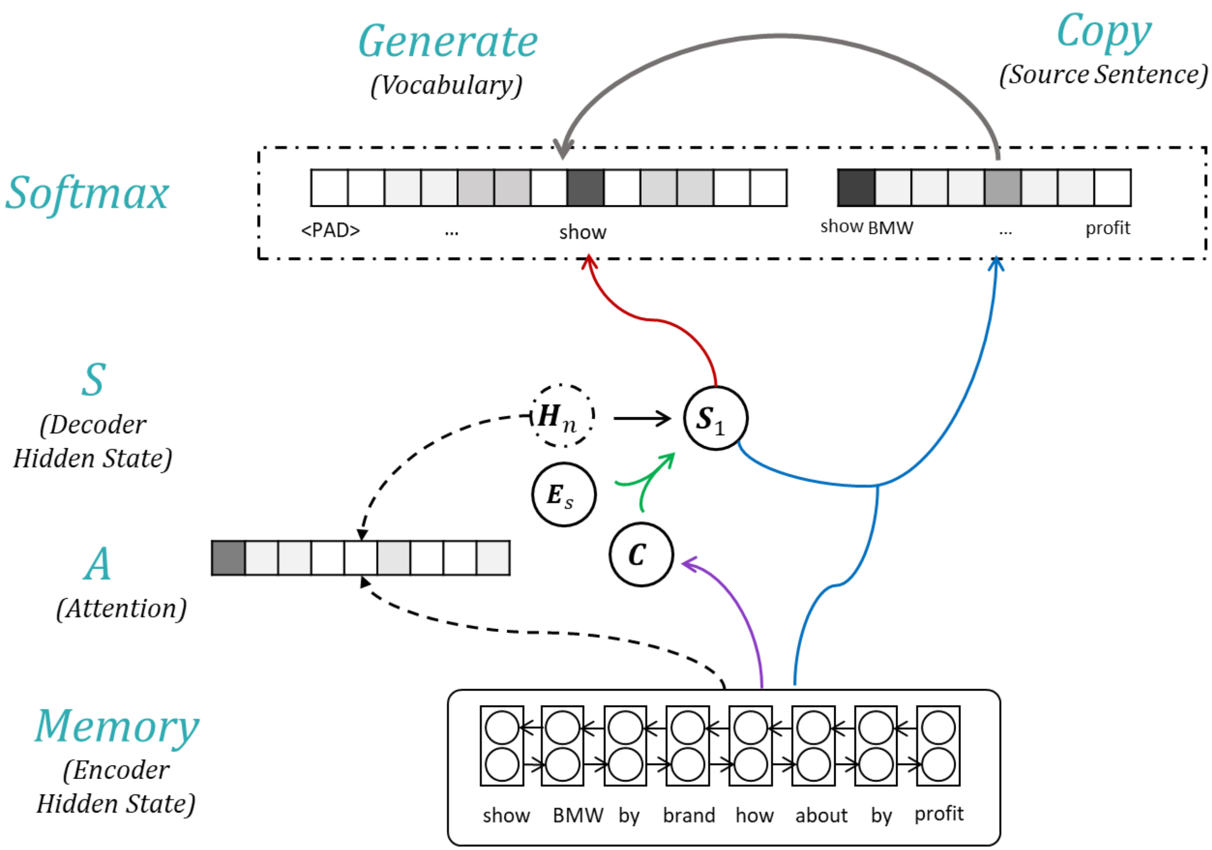
How much money has Bill earned

Compare it with Smith

→ Compare money Bill has earned with Smith

2. Methods

COPYNET



$$\alpha_i = V_a^T \tanh(\boldsymbol{W}[\boldsymbol{H}_n \boldsymbol{H}_i])$$

$$a_i = \frac{e^{\alpha_i}}{\sum e^{\alpha_k}}$$

$$\boldsymbol{C} = \sum a_i \cdot \boldsymbol{H}_i$$

$$\boldsymbol{S}_1 = RNN(\boldsymbol{H}_n, [\boldsymbol{E}_s, \boldsymbol{C}])$$

$$\boldsymbol{O}_G = \boldsymbol{W}_o \boldsymbol{S}_1$$

$$\beta_j = \tanh(\boldsymbol{W}_C \boldsymbol{H}_j) \boldsymbol{S}_1$$

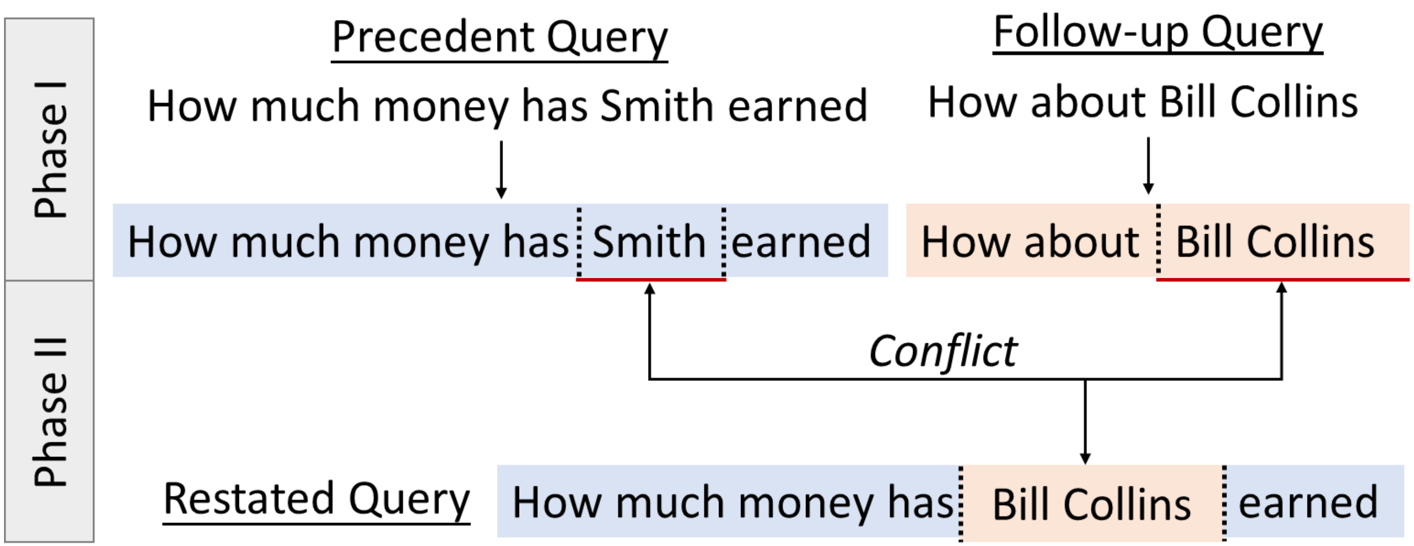
$$\boldsymbol{O}_C = [\beta_1, \ldots, \beta_s]$$

$$\boldsymbol{P} = Softmax([\boldsymbol{O}_G, \boldsymbol{O}_C])$$

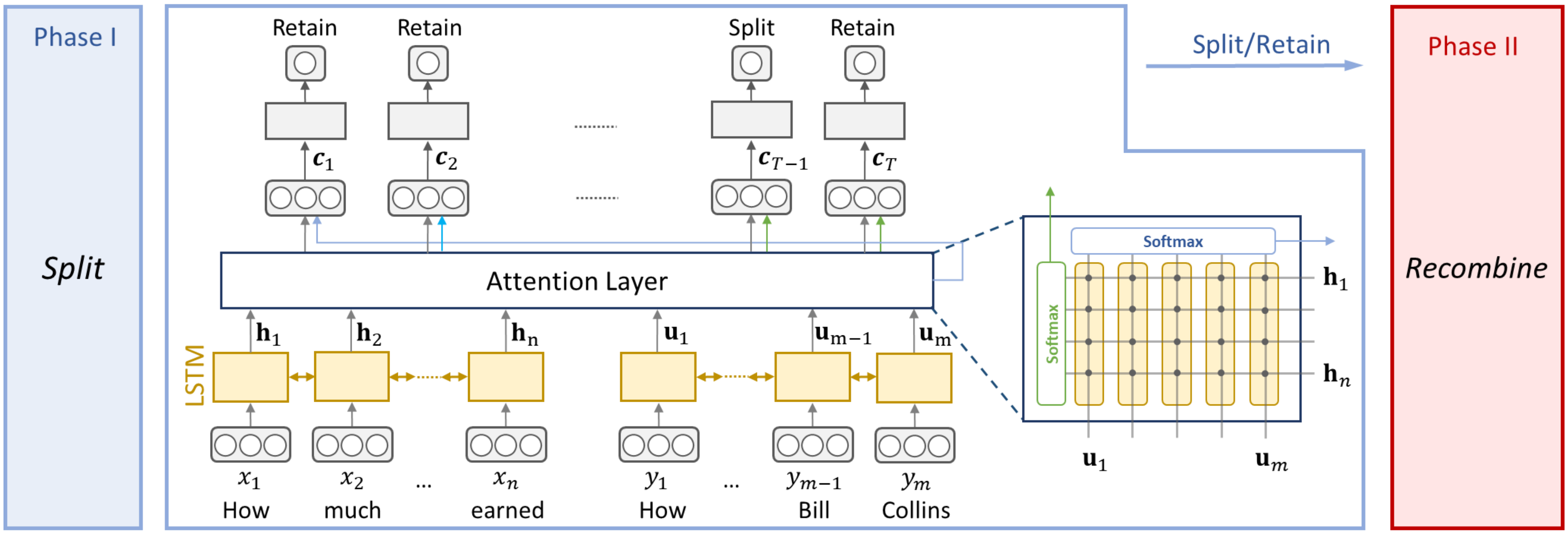
$$p(y_t) = p(y_t, g|\cdot) + p(y_t, c|\cdot)$$

3. Ours Research

Our solution: ***Split*** and ***Recombine***



3. Ours Research



- ***Split*** is modeled as a sequence labeling problem {0,1}.
- ***Recombine*** generates the restate based on semantic conflicting.

北京航空航天大学 BEIHANG UNIVERSITY

Microsoft Research 微软亚洲研究院

3. Ours Research

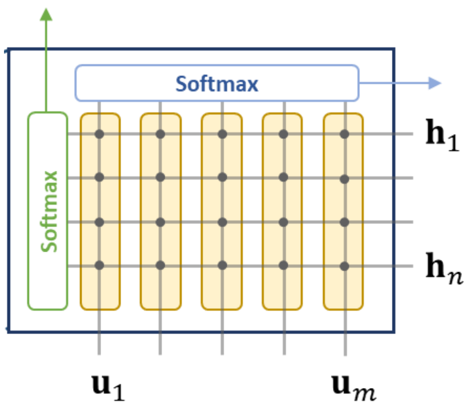
The architecture of our **SplitNet** is as follows:

- Embedding Layer: character, word and segment $[\phi_c; \phi_w; \phi_s]$
- Context Layer: Bi-LSTM $\overrightarrow{\mathbf{h}}_i = \overrightarrow{\mathbf{LSTM}}(\phi(x_i); \overrightarrow{\mathbf{h}}_{i-1})$

3. Ours Research

The architecture of our **SplitNet** is as follows:

- Embedding Layer: character, word and segment $[\phi_c; \phi_w; \phi_s]$
- Context Layer: Bi-LSTM $\overrightarrow{\mathbf{h}}_i = \overrightarrow{\mathbf{LSTM}}(\phi(x_i); \overrightarrow{\mathbf{h}}_{i-1})$
- Bi-Attention Layer: precedent-to-follow attention & follow-to-precedent attention



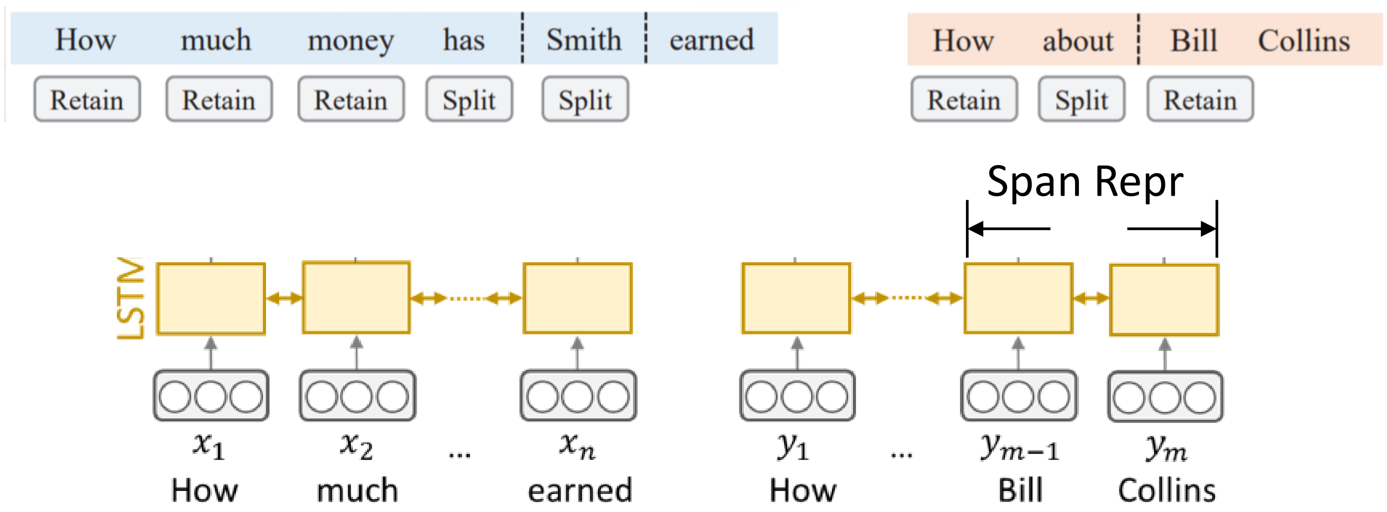
- Output: combine adjacent states to predict the action

$$\mathbf{c}^{x_i} = [\mathbf{h}_i;\ \mathbf{h}_i \circ \tilde{\mathbf{h}}_i;\ \mathbf{h}_{i+1} \circ \tilde{\mathbf{h}}_{i+1}] \qquad \sigma(\mathbf{W} * \mathbf{c}_t + b)$$

3. Ours Research

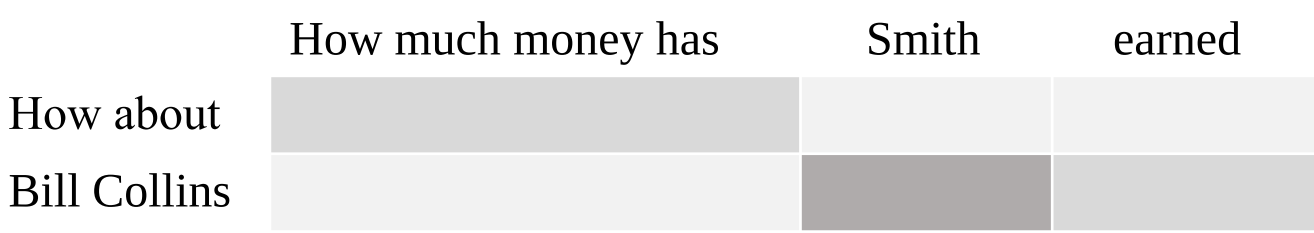
The recombination process is to generate the restate query:

1. Giving spans, we could obtain the representations of them.

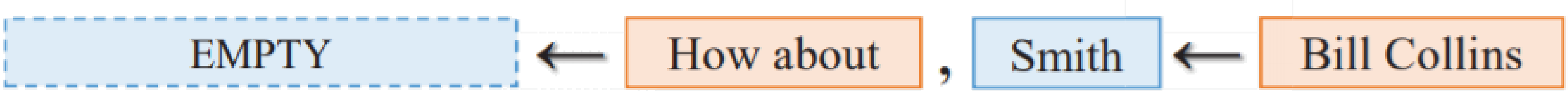


3. Ours Research

2. Using the representation, we could identify whether the spans semantically conflict.



3. Then use the 1-1 replacement principle, we could get the result.



3. Ours Research

An ideal solution, but there is no annotation!

- **We do not know** how to split the utterance to spans.
- **We do not know** which span conflicts with which one.

3. Ours Research

An ideal solution, but there is no annotation!

- **We do not know** how to split the utterance to spans.
- **We do not know** which span conflicts with which one.

We propose to solve the problem using RL.

- **Pretrain** the model under noisy distant supervision.
- **Fine-tune** it with rewards using policy gradient.

3. Ours Research

Find the **maximum common string** to export the training data

How much money has Smith earned
How about Bill Collins

How much money has Bill Collins earned

3. Ours Research

A simple way to fine-tune the model is to sample both ***Split*** and ***Recombine***. The reward could be BLEU or other task-related metrics.

$$\mathcal{L}_{\mathrm{rl}} = \mathbb{E}[\sum_{\tilde{\mathbf{z}} \in \mathcal{Z}} \sum_{q \in \mathcal{Q}} P_{\mathrm{split}}(q|\mathbf{x}, \mathbf{y}) P_{\mathrm{rec}}(\tilde{\mathbf{z}}|q) r(\mathbf{z}, \tilde{\mathbf{z}})]$$

3. Ours Research

A simple way to fine-tune the model is to sample both ***Split*** and ***Recombine***. The reward could be BLEU or other task-related metrics.

$$\mathcal{L}_{\mathrm{rl}} = \mathbb{E}[\sum_{\tilde{\mathbf{z}} \in \mathcal{Z}} \sum_{q \in \mathcal{Q}} P_{\mathrm{split}}(q|\mathbf{x}, \mathbf{y}) P_{\mathrm{rec}}(\tilde{\mathbf{z}}|q) r(\mathbf{z}, \tilde{\mathbf{z}})]$$

But sadly 😭, it convergences **slow** and performs **bad**.

3. Ours Research

Why ?

- **The space Q**: 2 ^ |Sequence|
- **The space Z**: span-based permutation

The space Q x Z is vast, it is hard to train well via direct sampling.

3. Ours Research

Is there any way to

reduce the sampling space?

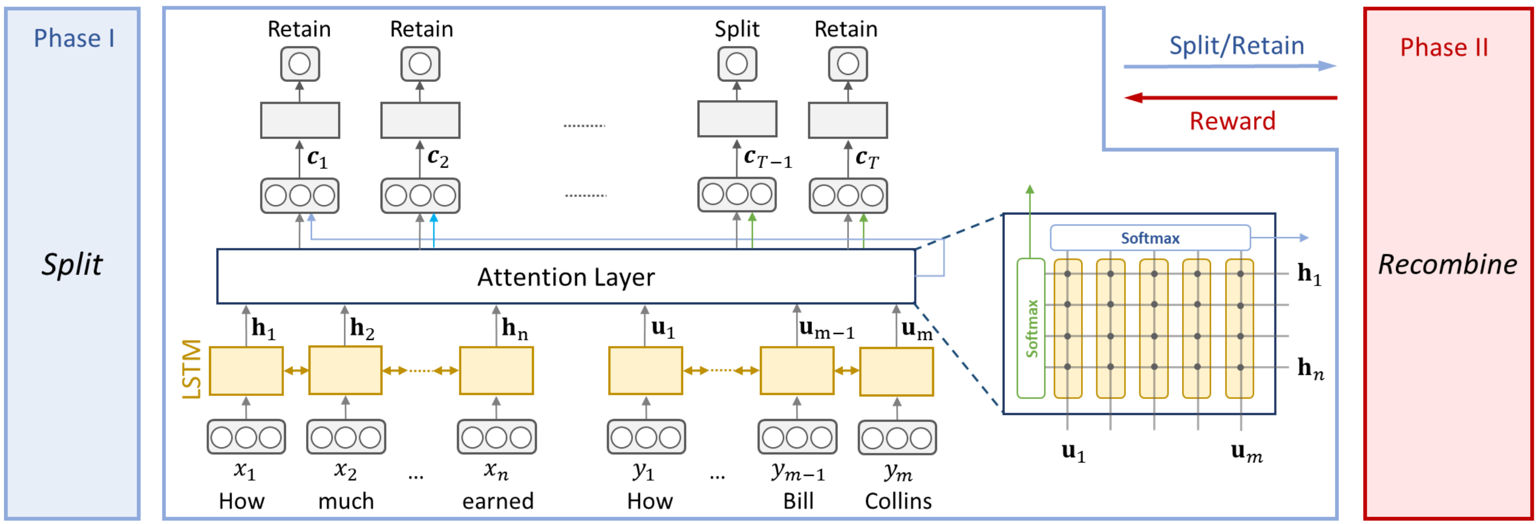
3. Ours Research

Yes, if we enumerate the space Z.

3. Ours Research

Yes, if we enumerate the space Z.

Reward shifting :Enumerating the space Z (a small effort) to provide rewards for action sampling in the space Q.



3. Ours Research

Recalling the equation of objective, the reward comes out:

$$\mathcal{L}_{\text{rl}} = \mathbb{E}[\sum_{\tilde{\mathbf{z}} \in \mathcal{Z}} \sum_{q \in \mathcal{Q}} P_{\text{split}}(q|\mathbf{x}, \mathbf{y}) P_{\text{rec}}(\tilde{\mathbf{z}}|q) r(\mathbf{z}, \tilde{\mathbf{z}})],$$

$$\downarrow$$

$$\mathcal{L}_{\text{rl}} = \mathbb{E}[\sum_{q \in \mathcal{Q}} P_{\text{split}}(q|\mathbf{x}, \mathbf{y}) \sum_{\tilde{\mathbf{z}} \in \mathcal{Z}} P_{\text{rec}}(\tilde{\mathbf{z}}|q) r(\mathbf{z}, \tilde{\mathbf{z}})],$$

3. Ours Research

Recalling the equation of objective, the reward comes out:

$$\mathcal{L}_{\mathrm{rl}} = \mathbb{E}[\sum_{\tilde{\mathbf{z}} \in \mathcal{Z}} \sum_{q \in \mathcal{Q}} P_{\mathrm{split}}(q|\mathbf{x}, \mathbf{y}) P_{\mathrm{rec}}(\tilde{\mathbf{z}}|q) r(\mathbf{z}, \tilde{\mathbf{z}})],$$

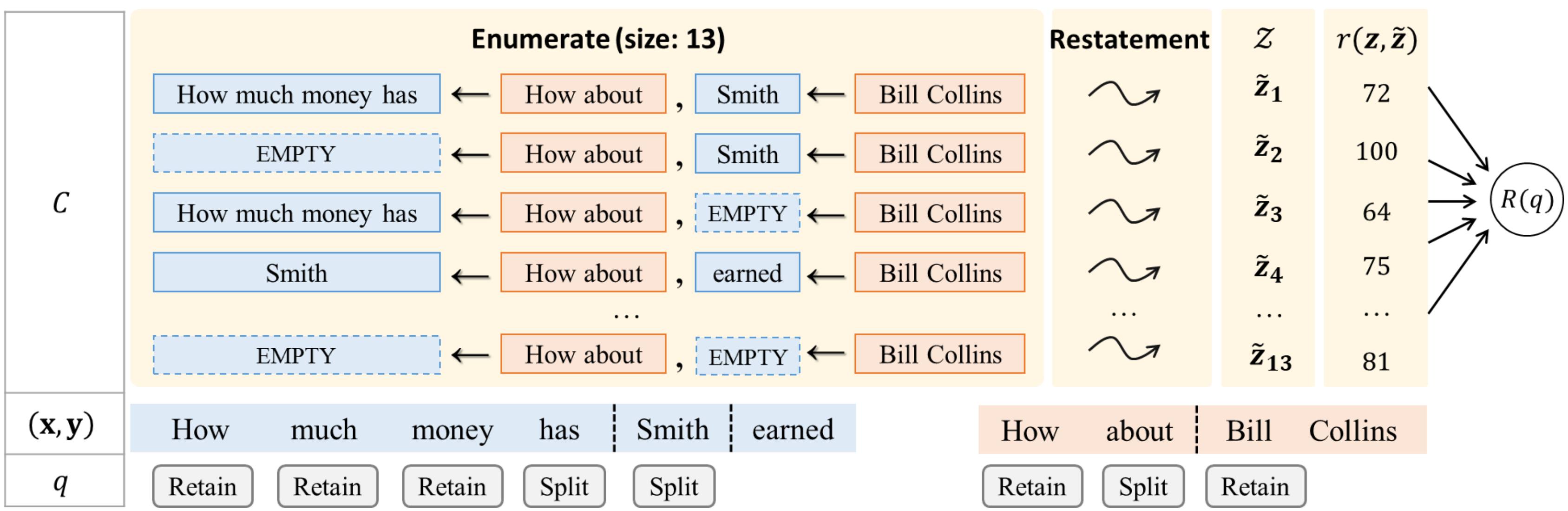
↓

$$\mathcal{L}_{\mathrm{rl}} = \mathbb{E}[\sum_{q \in \mathcal{Q}} P_{\mathrm{split}}(q|\mathbf{x}, \mathbf{y}) \boxed{\sum_{\tilde{\mathbf{z}} \in \mathcal{Z}} P_{\mathrm{rec}}(\tilde{\mathbf{z}}|q) r(\mathbf{z}, \tilde{\mathbf{z}})}],$$

Reward for **SplitNet** $R(q, \mathbf{z})$

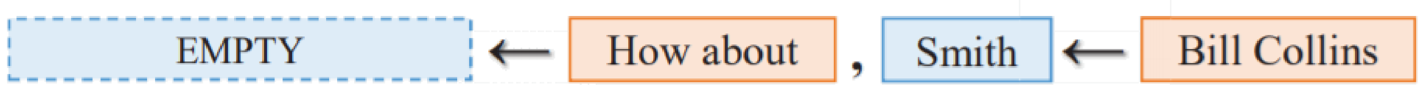
3. Ours Research

- Specifically, we systemically enumerate all valid **candidates**, and obtain the **R** .
- For unexpected large Z, we directly return the reward 0.



3. Ours Research

- The next question is, how to train the ***recombine***, a neural process where the span representations are learned.
- We take the candidate which gains the highest reward and regard it as the ground-truth.



3. Ours Research

- The next question is, how to train the ***recombine***, a neural process where the span representations are learned.
- We take the candidate which gains the highest reward and regard it as the ground-truth.

EMPTY ← How about , Smith ← Bill Collins

Objective Function

cos[Smith, Bill Collins] > cos[EMPTY, Bill Collins]

3. Ours Research

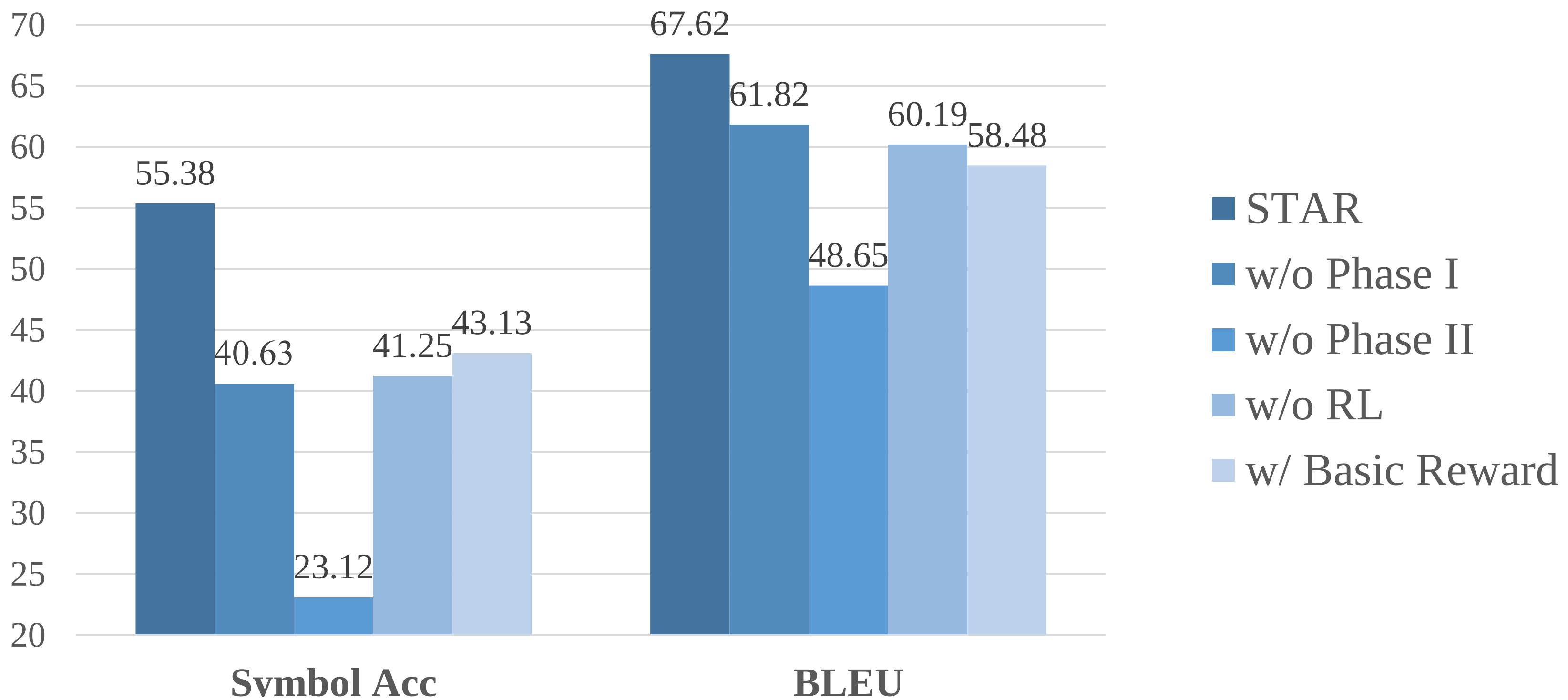
We mainly conducted experiments on **FollowUp** dataset [7]. Following metrics are employed to evaluate our method:

- ✓ **SymAcc**: detects whether all the SQL-related words are correctly involved in the predicted queries.
- ✓ **BLEU**: evaluates how similar the predicted queries are to golden ones.
- ✓ **AnsAcc**: check the answer accuracy of predicted queries manually.

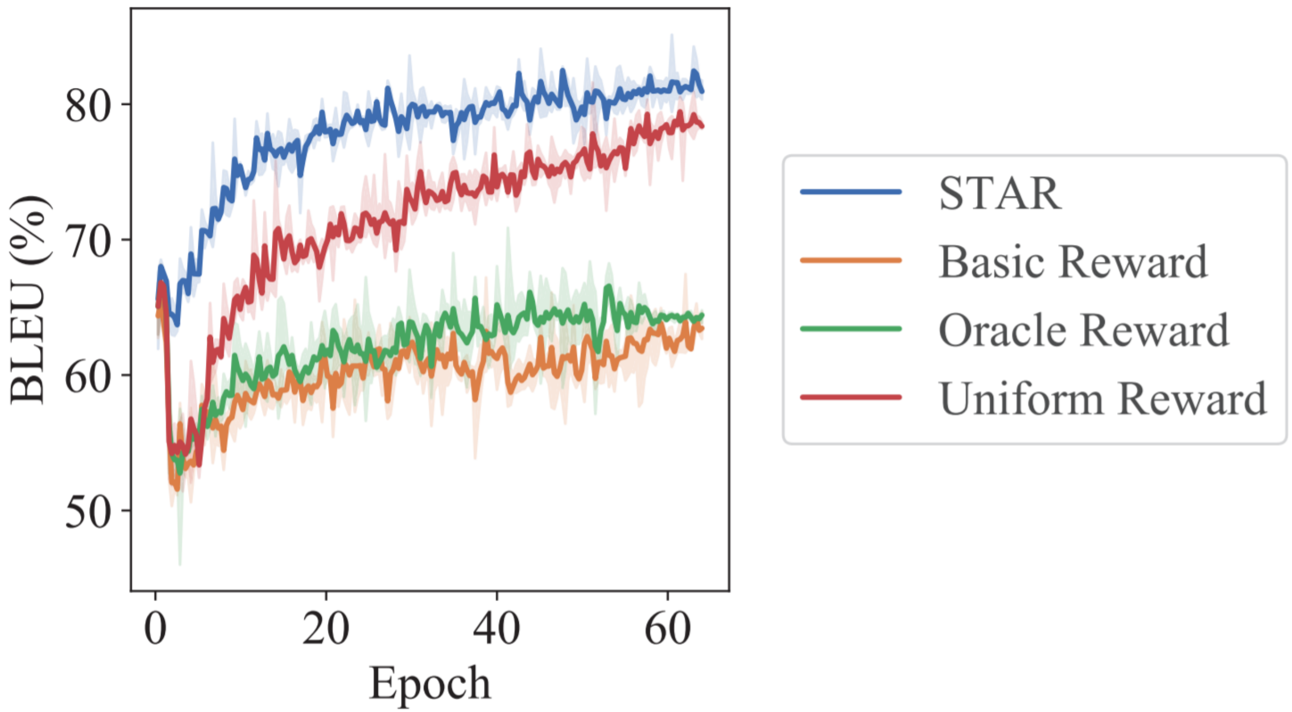
Model	**Dev**		**Test**		
	SymAcc (%)	BLEU (%)	SymAcc (%)	BLEU (%)	AnsAcc (%)
SEQ2SEQ† (Bahdanau et al., 2015)	0.63 ± 0.00	21.34 ± 1.14	0.50 ± 0.22	20.72 ± 1.31	–
COPYNET† (Gu et al., 2016)	17.50 ± 0.87	43.36 ± 0.54	19.30 ± 0.93	43.34 ± 0.45	–
COPY+BERT (Devlin et al., 2019)	18.63 ± 0.61	45.14 ± 0.68	22.00 ± 0.45	44.87 ± 0.52	–
CONCAT†	–	–	22.00 ± –	52.02 ± –	25.24
E2ECR† (Lee et al., 2017)	–	–	27.00 ± –	52.47 ± –	27.18
FANDA† (Liu et al., 2019)	49.00 ± 1.28	60.14 ± 0.98	47.80 ± 1.14	59.02 ± 0.54	60.19
STAR	**55.38** ± 1.21	**67.62** ± 0.65	**54.00** ± 1.09	**67.05** ± 1.05	**65.05**

3. Ours Research

We also do ablation study to identify the modules of our method.



3. Ours Research



3. Ours Research

Our model could be incorporated with any parser. Thus, we performed extensive experiments on the **SQA** dataset [3].

Model	Precedent	Follow-up
DynSP (Iyyer et al., 2017)	70.9	35.8
NP (Neelakantan et al., 2016)	58.9	35.9
NP + STAR	58.9	38.1
DynSP + STAR	70.9	39.5
DynSP* (Iyyer et al., 2017)	70.4	41.1

3. Ours Research

Here are three reals cases. Blue spans are those have conflicts, and histograms represent the conflict probabilities to all the spans in precedent queries.

No		Case Analysis
1	*Precedent*	: [compared to glebe park] [, does] [hampden park] [holds more attendances at capacity ?]
	Follow-up	: [how about] [compared to balmoor]
	STAR	: compared to balmoor , does hampden park holds more attendances at capacity ?
2	*Precedent*	: [Is there any book which belongs to] [Nancy miller]
	Follow-up	: [I mean] [the writer Nancy miller]
	STAR	: Is there any book which belongs to the writer Nancy miller
3	*Precedent*	: [show directors of] [greatest love and promised land]
	Follow-up	: [show air date of] [those two films]
	STAR	: show air date of greatest love and promised land

3. Ours Research

Our code and data are released in Github:

- https://github.com/microsoft/EMNLP2019-Split-And-Recombine
- https://github.com/SivilTaram/FollowUp

4. Take-home-message

- For conversational scenarios, we can combine pretrained context-independent models and the task of FQA to handle the scenarios.
- Imposing a latent intermediate structure encounters the problem of hard training. We may do reward shifting to avoid large sampling space with a small enumerating effort.

5. References

[1] Suhr A, Iyer S, Artzi Y. Learning to Map Context-Dependent Sentences to Executable Formal Queries

[2] Deborah A. Dahl, Madeleine Bates, Michael Brown, William M. Fisher, Kate Hunicke-Smith, David S. Pallett, Christine Pao, Alexander I. Rudnicky, and Elizabeth Shriberg. 1994. Expanding the scope of the ATIS task: The ATIS-3 corpus.

[3] Mohit Iyyer, Wen-tau Yih, and Ming-Wei Chang. 2017. Search-based neural structured learning for sequential question answering.

[4] Yu T, Zhang R, Yasunaga M, et al. SParC: Cross-Domain Semantic Parsing in Context.

[5] Gu J, Lu Z, Li H, et al. Incorporating copying mechanism in sequence-to-sequence learning.

5. References

[6] Kenton Lee, Luheng He, Mike Lewis, and Luke Zettlemoyer. 2017. End-to-end neural coreference resolution.

[7] Qian Liu, Bei Chen, Jian-Guang Lou, Ge Jin, and Dongmei Zhang. 2019. FANDA: A novel approach to perform follow-up query analysis.

Thanks

Speaker : Qian Liu