

Discern: Discourse-Aware Entailment Reasoning Network for Conversational Machine Reading

Yifan Gao¹ Chien-Sheng Wu² Jingjing Li¹ Shafiq Joty^{2,3}
Steven C.H. Hoi² Caiming Xiong² Irwin King¹ Michael R. Lyu¹

1. The Chinese University of Hong Kong

2. Salesforce Research

3. Nanyang Technological University

Conversational Machine Reading

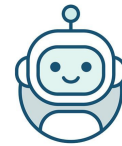
Introduction



Scenario: I am a 34-year-old man from the United States who owns their own business. We are an American small business.



Question: Is the 7(a) Loan Program for me?



Rule Text

7(a) loans are the most basic and most used type loan of the Small Business Administration's (SBA) business loan programs. Its name comes from section 7(a) of the Small Business Act, which authorizes the agency to provide business loans to American small businesses. The loan program is designed to assist for-profit businesses that are not able to get other financing from other resources.

Conversational Machine Reading

Introduction



Scenario: I am a 34-year-old man from the United States who owns their own business. We are an American small business.

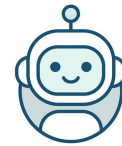


Question: Is the 7(a) Loan Program for me?



Yes

Are you a for-profit business?

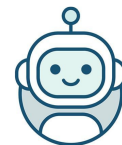


No

Are you able to get financing from other resources?



Yes. (You can apply the loan.)



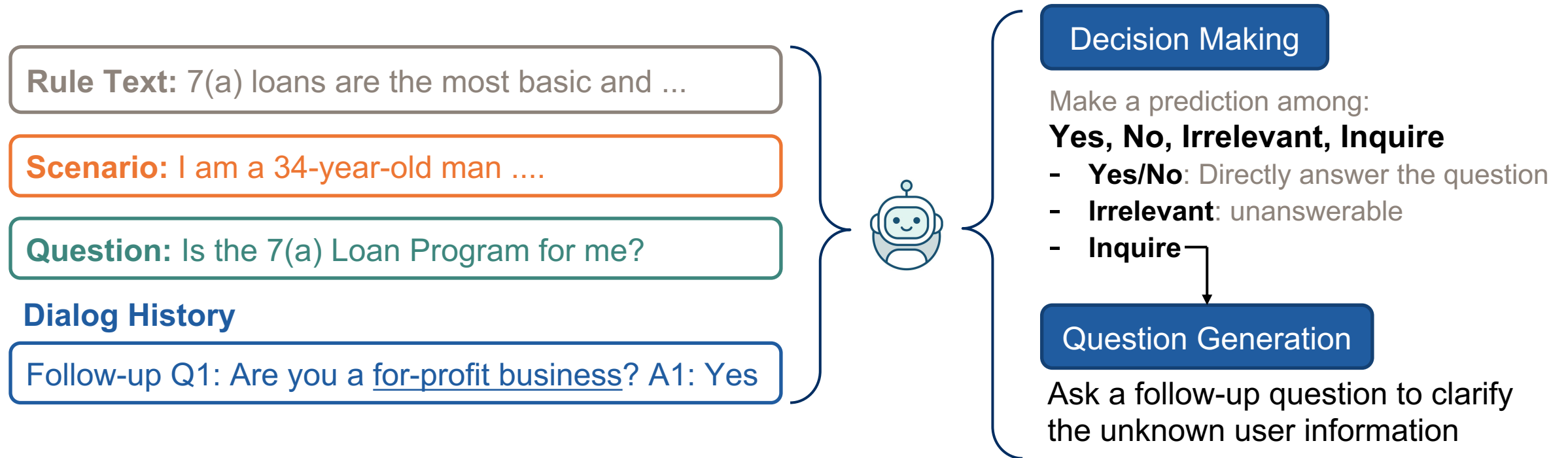
Rule Text

7(a) loans are the most basic and most used type loan of the Small Business Administration's (SBA) business loan programs. Its name comes from section 7(a) of the Small Business Act, which authorizes the agency to provide business loans to ^①American small businesses. The loan program is designed to assist ^②for-profit businesses that are ^③not able to get other financing from other resources.

Task Definition

ShARC: **Sh**aping **A**nswers with **R**ules through **C**onversation

ShARC
(Saeidi et al., 2018)



Challenges

1. Document Interpretation

- Identification of Conditions
- Determination of Logical Structures

7(a) loans are the most basic and most used type loan of the Small Business Administration's (SBA) business loan programs. Its name comes from section 7(a) of the Small Business Act, which authorizes the agency to provide business loans to American small businesses. The loan program is designed to assist for-profit businesses that are not able to get other financing from other resources.

“Eligible for 7(a) loans” = (①==True) and (② ==True) and (③ ==True)

Challenges

2. Dialog Understanding

- Track the user's fulfillment over the conditions
- Jointly consider the fulfillment states and the logical structure of rules

7(a) loans are the ... provide business loans to ^①American small businesses. The loan program is designed to assist ^②for-profit businesses that are ^③not able to get other financing from other resources.

“Eligible for 7(a) loans” = (①==True) **and** (② ==True) **and** (③ ==True) \Rightarrow **AND**: examine the fulfillments of all conditions

Scenario: I am a 34-year-old man from the United States who owns their own business. We are an American small business.

\Rightarrow ① == True

Fulfillment State: (①==True) **and** (② == ?) **and** (③ == ?) \Rightarrow **Decision: Inquire**

Follow-up Q1: Are you a for-profit business?

Challenges

2. Dialog Understanding

- Track the user's fulfillment over the conditions
- Jointly consider the fulfillment states and the logical structure of rules

7(a) loans are the ... provide business loans to ^①American small businesses. The loan program is designed to assist ^②for-profit businesses that are ^③not able to get other financing from other resources.

“Eligible for 7(a) loans” = (①==True) **and** (② ==True) **and** (③ ==True) \Rightarrow **AND**: examine the fulfillments of all conditions

Scenario: I am a 34-year-old man from the United States who owns their own business. We are an American small business.

\Rightarrow ① == True

Follow-up Q1: Are you a for-profit business? A1: Yes

\Rightarrow ② == True

Fulfillment State: (①==True) **and** (② == True) **and** (③ == ?) \Rightarrow **Decision: Inquire**

Follow-up Q2: Are you able to get financing from other resources?

Challenges

2. Dialog Understanding

- Track the user's fulfillment over the conditions
- Jointly consider the fulfillment states and the logical structure of rules

7(a) loans are the ... provide business loans to ^①American small businesses. The loan program is designed to assist ^②for-profit businesses that are ^③not able to get other financing from other resources.

“Eligible for 7(a) loans” = (①==True) **and** (② ==True) **and** (③ ==True) \Rightarrow **AND**: examine the fulfillments of all conditions

Scenario: I am a 34-year-old man from the United States who owns their own business. We are an American small business.

\Rightarrow ① == True

Follow-up Q1: Are you a for-profit business? A1: Yes

\Rightarrow ② == True

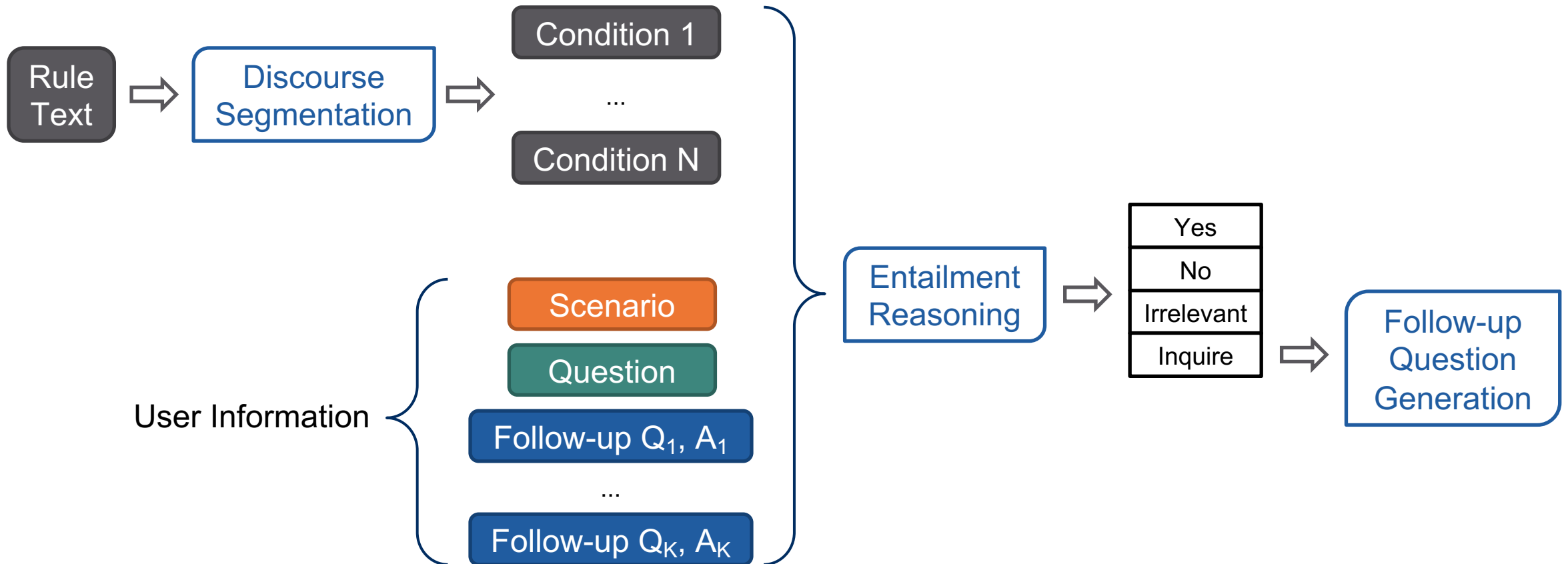
Follow-up Q2: Are you able to get financing from other resources? A2: No

\Rightarrow ③ == True

Fulfillment State: (①==True) **and** (② == True) **and** (③ == True) \Rightarrow **Decision: Yes** (You can apply the loan.)

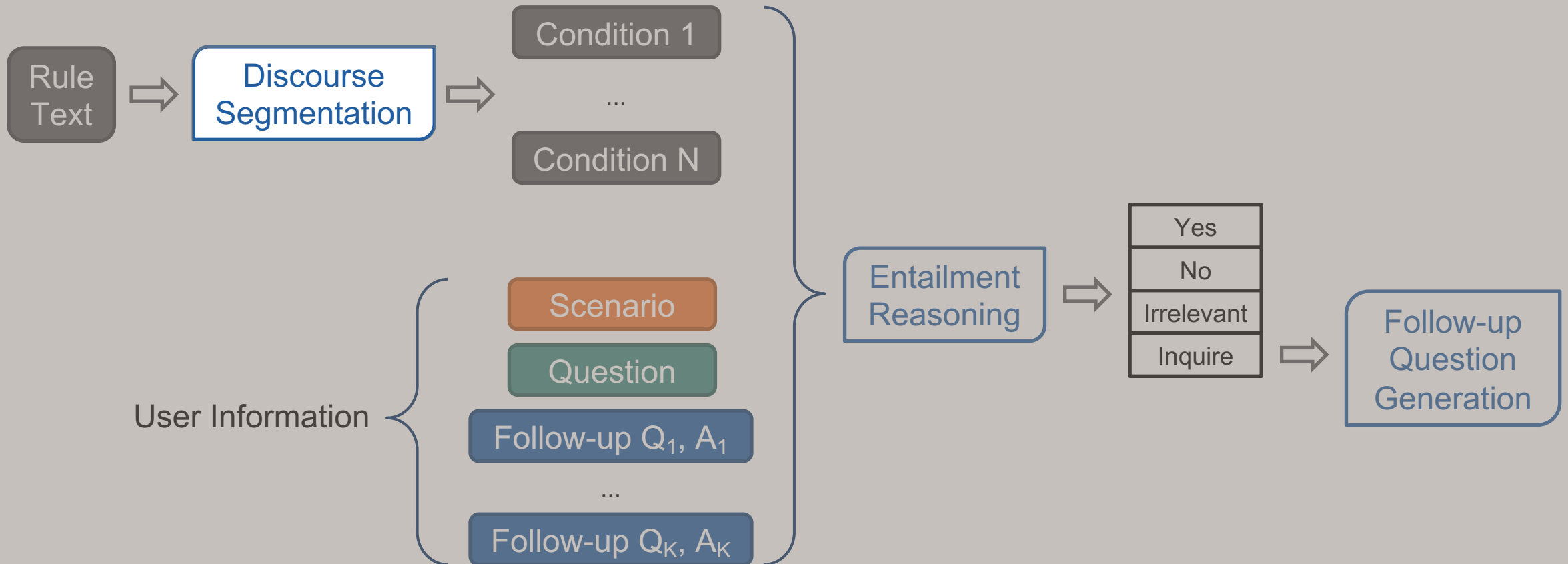
Discern: Discourse-Aware Entailment Reasoning Network for Conversational Machine Reading

Overview



Discern: Discourse-Aware Entailment Reasoning Network for Conversational Machine Reading

Overview



Discern: Discourse-Aware Entailment Reasoning Network for Conversational Machine Reading

Rule Segmentation

- Goal
 - Understand the logical structure of the rule text
 - Parse the rule into individual conditions for entailment reasoning
- Challenges
 - Sentence splitting is not enough: one rule sentence may contain several in-line conditions
- Solution: Discourse Segmentation
 - In the Rhetorical Structure Theory (RST) of discourse parsing (Mann and Thompson, 1988), texts are split into clause-like units called elementary discourse units (EDUs)

Rule Text: If a worker has taken more leave than they're entitled to, their employer must not take money from their final pay unless it's been agreed beforehand in writing.



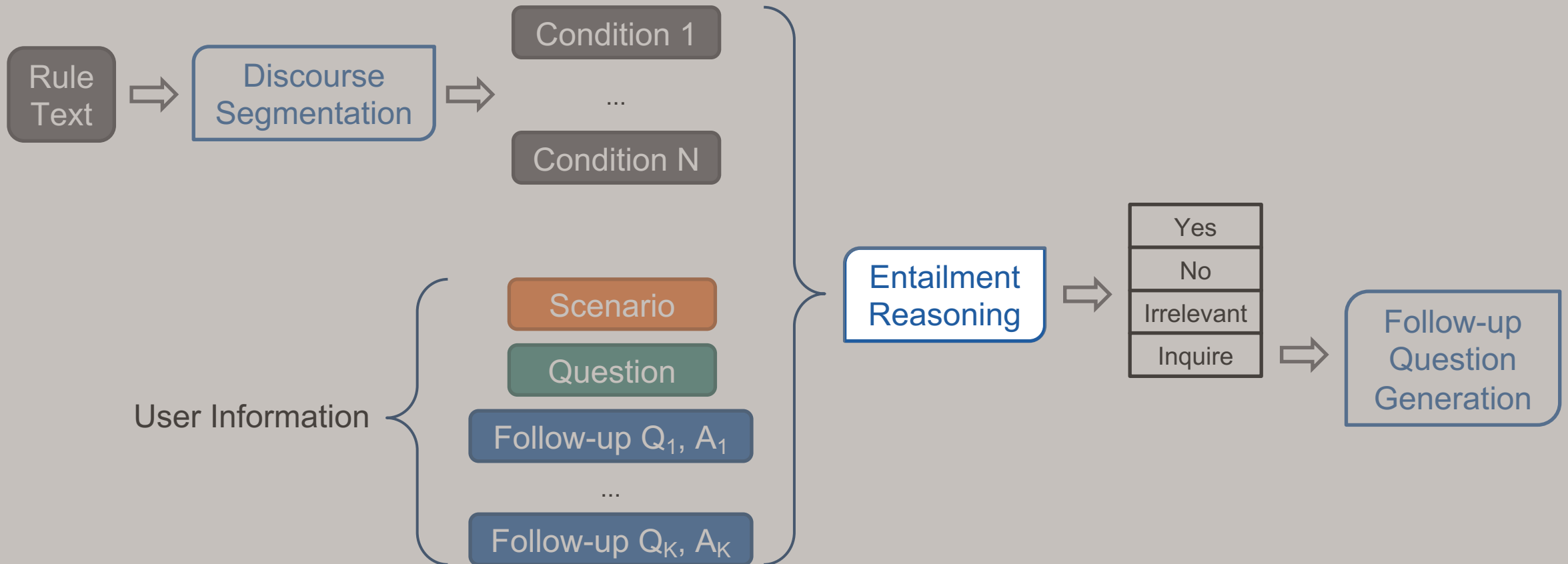
Discourse Segmentation



[If a worker has taken more leave than they're entitled to,]_{EDU1} [their employer must not take money from their final pay]_{EDU2} [unless it's been agreed beforehand in writing.]_{EDU3}

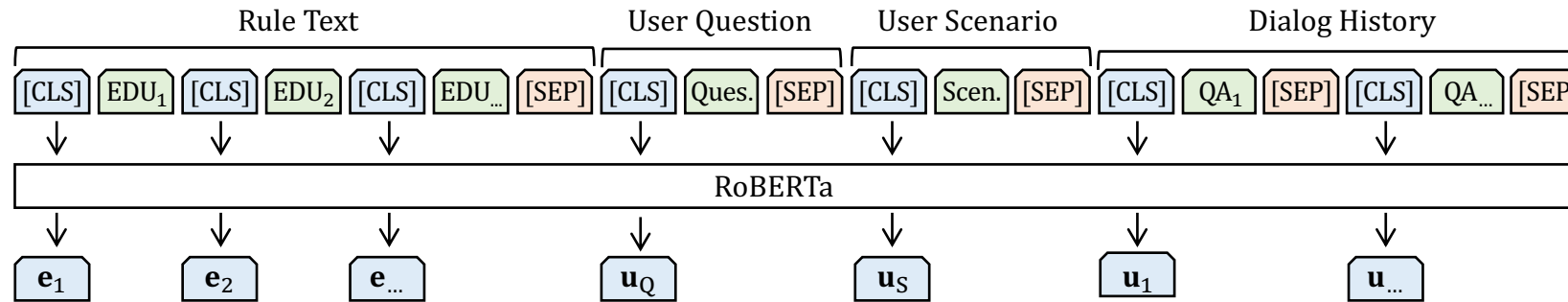
Discern: Discourse-Aware Entailment Reasoning Network for Conversational Machine Reading

Overview



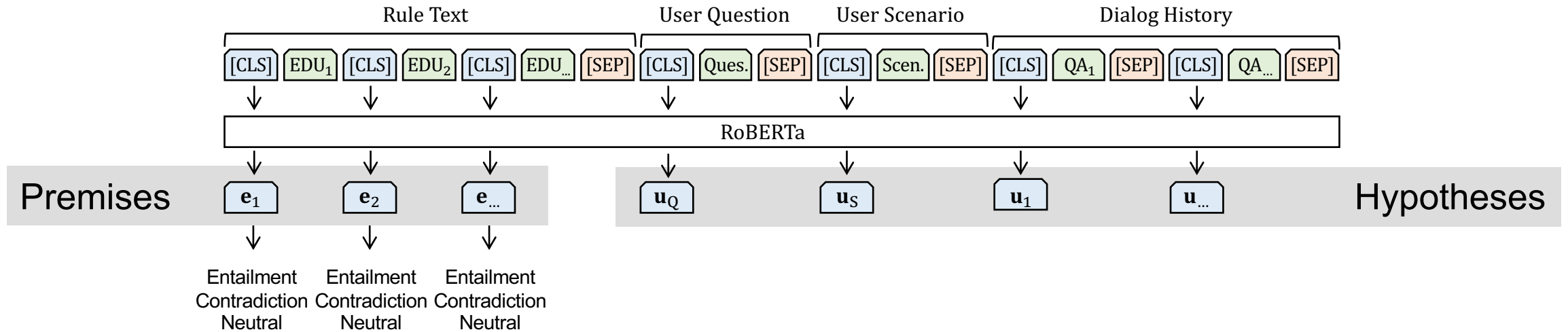
Discern: Decision Making via Entailment Reasoning

Encoding



Discern: Decision Making via Entailment Reasoning

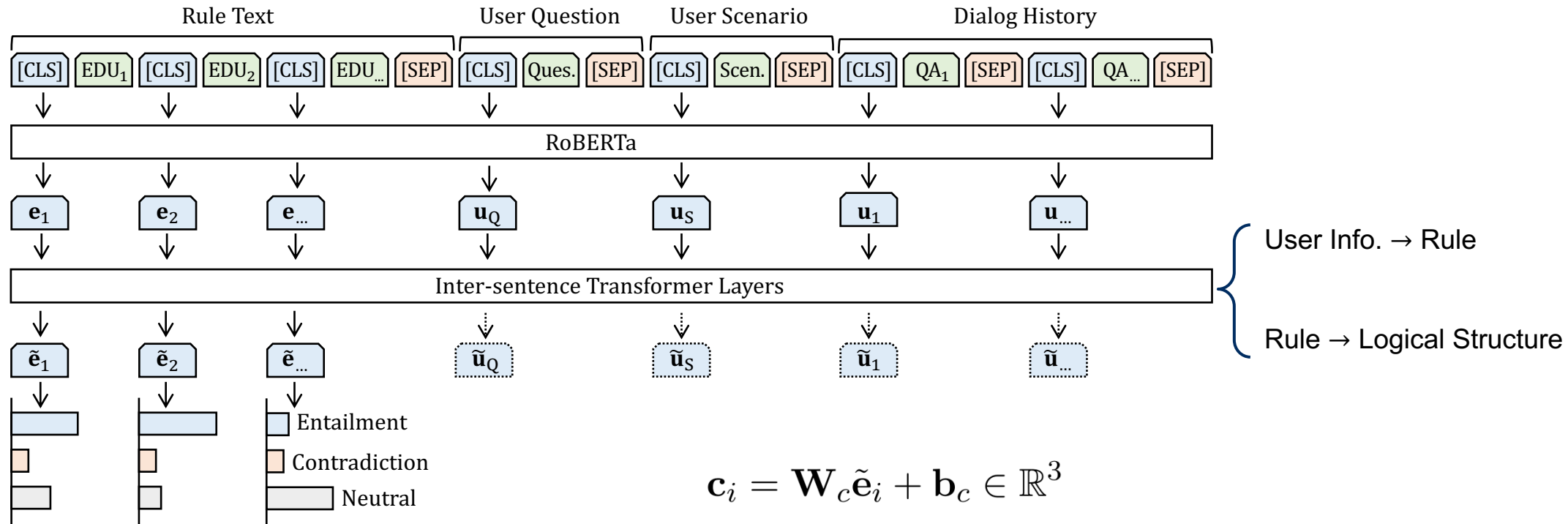
Entailment Prediction



Multi-Sentence Entailment Prediction

Discern: Decision Making via Entailment Reasoning

Entailment Prediction

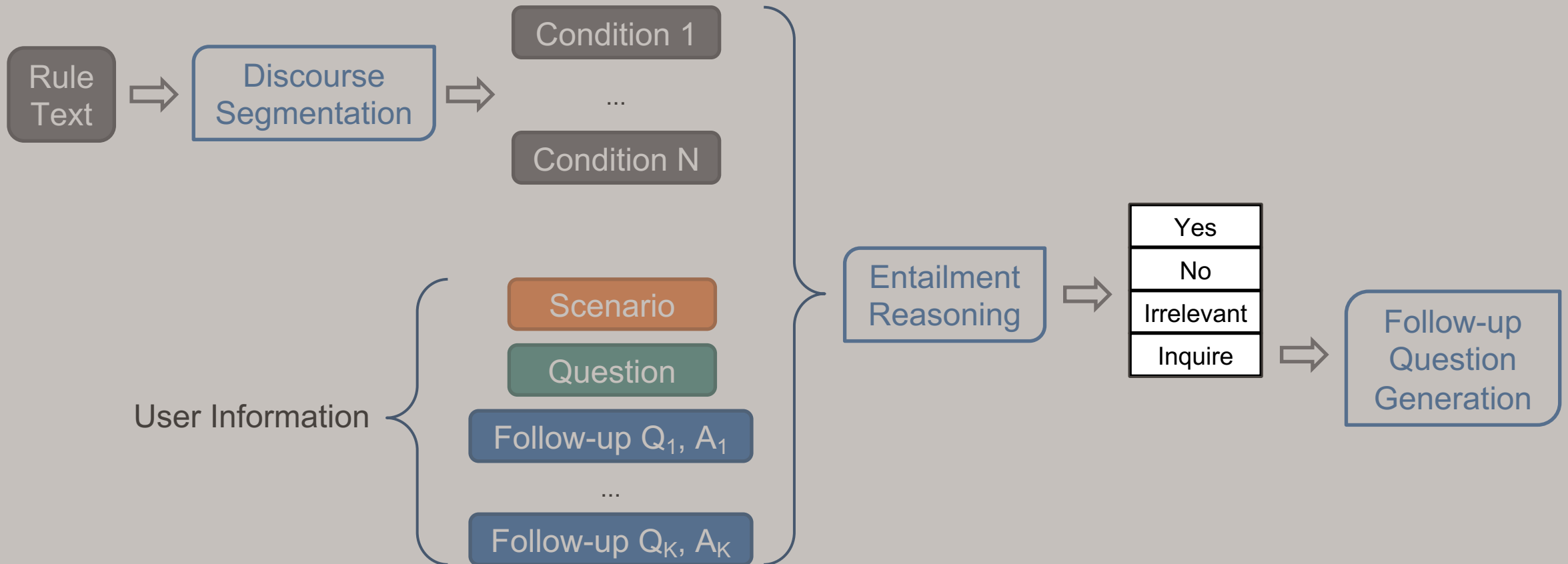


$$\mathbf{c}_i = \mathbf{W}_c \tilde{\mathbf{e}}_i + \mathbf{b}_c \in \mathbb{R}^3$$

$$\mathcal{L}_{\text{entail}} = -\frac{1}{K} \sum_{i=1}^K \log \text{softmax}(\mathbf{c}_i)_r$$

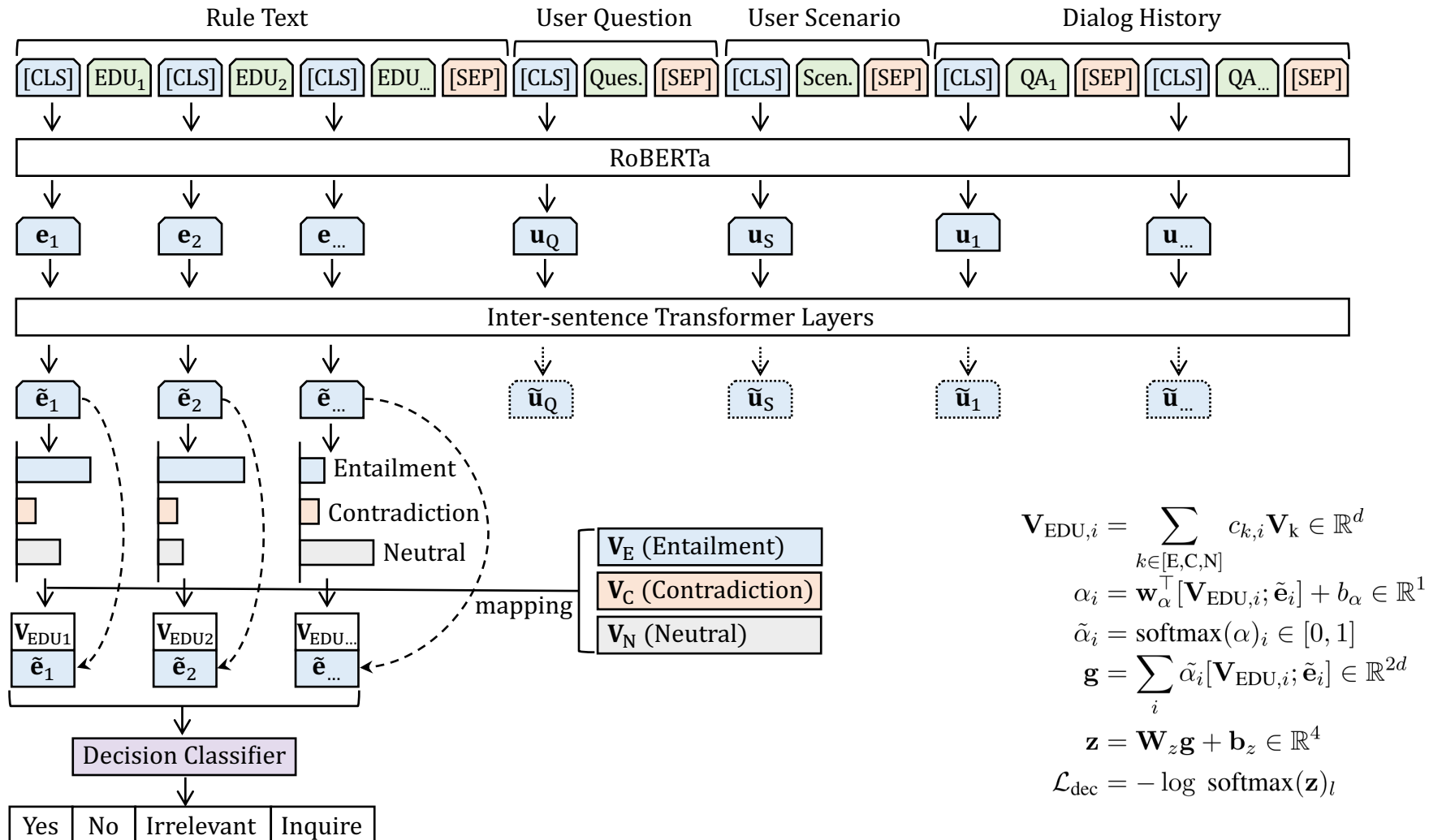
Discern: Discourse-Aware Entailment Reasoning Network for Conversational Machine Reading

Overview



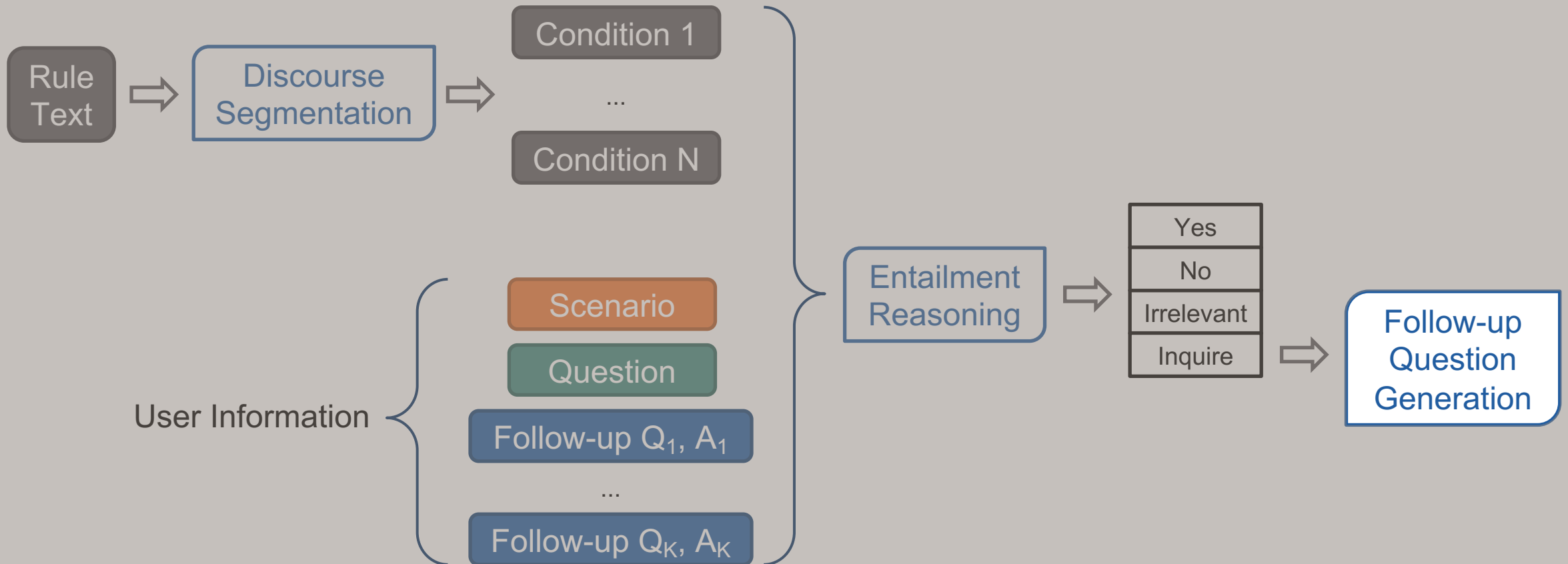
Discern: Discourse-Aware Entailment Reasoning Network for Conversational Machine Reading

Decision Making



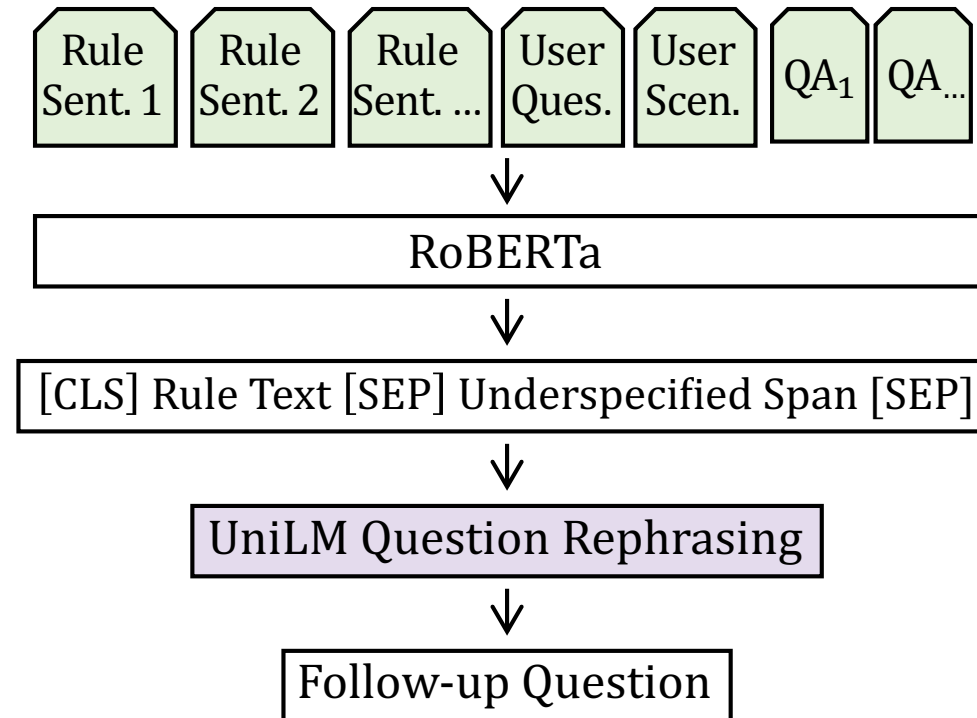
Discern: Discourse-Aware Entailment Reasoning Network for Conversational Machine Reading

Overview



Discern: Discourse-Aware Entailment Reasoning Network for Conversational Machine Reading

Follow-up Question Generation



Experiments

Experimental Setup

Dataset:

- ShARC CMR dataset (Saeidi et al. 2018)
- Train/Dev/Test dataset sizes are 21980/2270/8276
- Test set is not public
- <https://sharc-data.github.io/leaderboard.html>

Evaluation

- Decision Making: Macro- & Micro- Accuracy
- Question Generation: BLEU1 & BLEU4

ShARC: End-to-end Task							
#	Model / Reference	Affiliation	Date	Micro Accuracy[%]	Macro Accuracy[%]	BLEU-1	BLEU-4
1	Discern (single model)	The Chinese University of Hong Kong	May 2020	73.2	78.3	64.0	49.1
2	EMT	Salesforce Research & CUHK	Nov 2019	69.4	74.8	60.9	46.0
3	EMT + entailment	Salesforce Research & CUHK	Mar 2020	69.1	74.6	63.9	49.5
4	UrcaNet (ensemble)	IBM Research AI	Dec 2019	69.0	74.6	56.7	42.0
5	E3	University of Washington	Feb 2019	67.6	73.3	54.1	38.7

Experiments

Main Results

Models	End-to-End Task (Leaderboard Performance)			
	Micro Acc.	Macro Acc.	BLEU1	BLEU4
Seq2Seq (Saeidi et al., 2018)	44.8	42.8	34.0	7.8
Pipeline (Saeidi et al., 2018)	61.9	68.9	54.4	34.4
BERTQA (Zhong and Zettlemoyer, 2019)	63.6	70.8	46.2	36.3
UrcaNet (Sharma et al., 2019)	65.1	71.2	60.5	46.1
BiSon (Lawrence et al., 2019)	66.9	71.6	58.8	44.3
E ³ (Zhong and Zettlemoyer, 2019)	67.6	73.3	54.1	38.7
EMT (Gao et al., 2020)	69.4	74.8	60.9	46.0
EMT+entailment (Gao et al., 2020)	69.1	74.6	63.9	49.5
DISCERN (our single model)	73.2	78.3	64.0	49.1

Experiments

Ablation Study

Models	Micro Acc.	Macro Acc.	
DISCERN	74.97 ± 0.27	79.55 ± 0.35	
DISCERN (BERT)	73.07 ± 0.21	77.77 ± 0.24	➡ RoBERTa > BERT
DISCERN (w/o EDU)	73.34 ± 0.22	78.25 ± 0.57	➡ Discourse Segmentation > Sentence Splitting
DISCERN (w/o Trans)	74.25 ± 0.36	78.78 ± 0.57	➡ Inter-Sentence Transformer IS Necessary!
DISCERN (w/o $\tilde{\mathbf{e}}$)	73.55 ± 0.26	78.19 ± 0.30	} Both Condition Representations and Entailment Vectors Facilitate Decisions
DISCERN (w/o \mathbf{V}_{EDU})	72.95 ± 0.23	77.53 ± 0.19	

Experiments

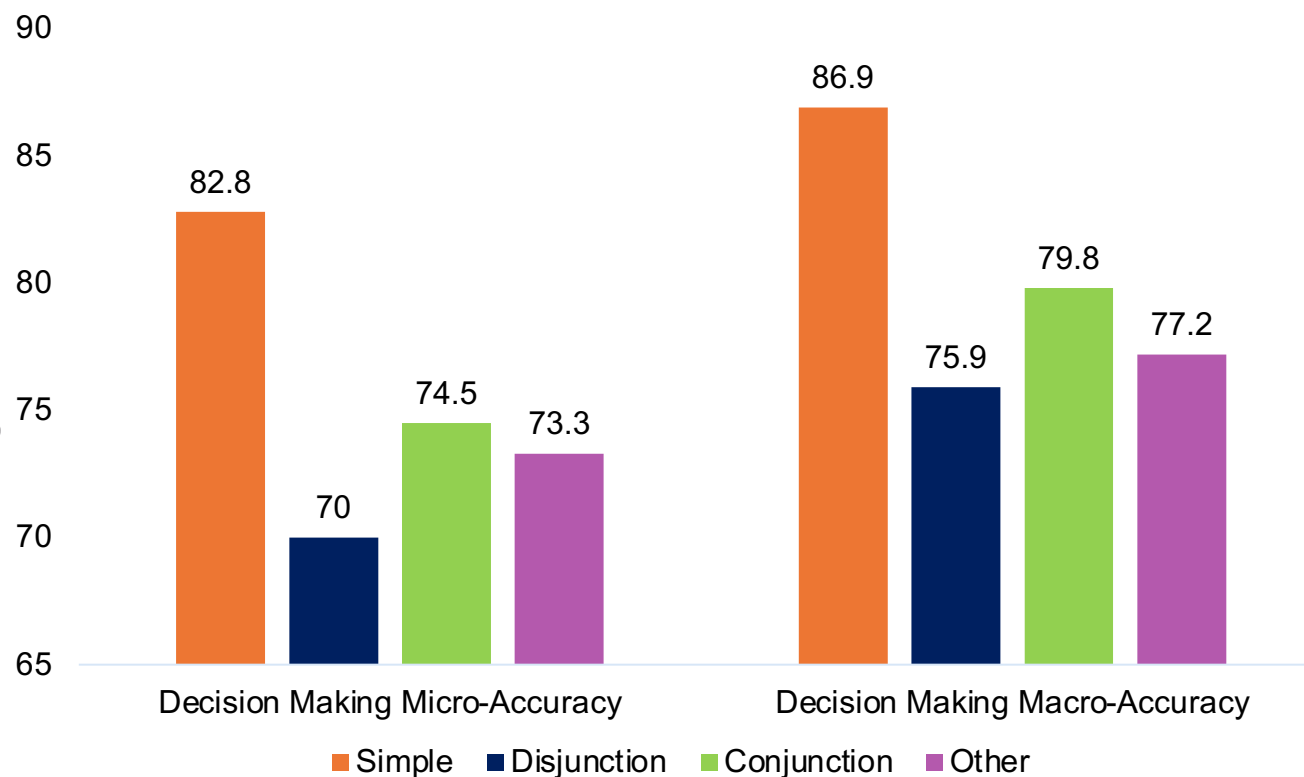
Analysis of Logical Structure of Rules

Simple: If A, then B

Disjunction: If A **OR** B **OR** C **OR** ..., then D

Conjunction: If A **AND** B **AND** C **AND** ..., then D

Other: Complicated AND/OR relationship

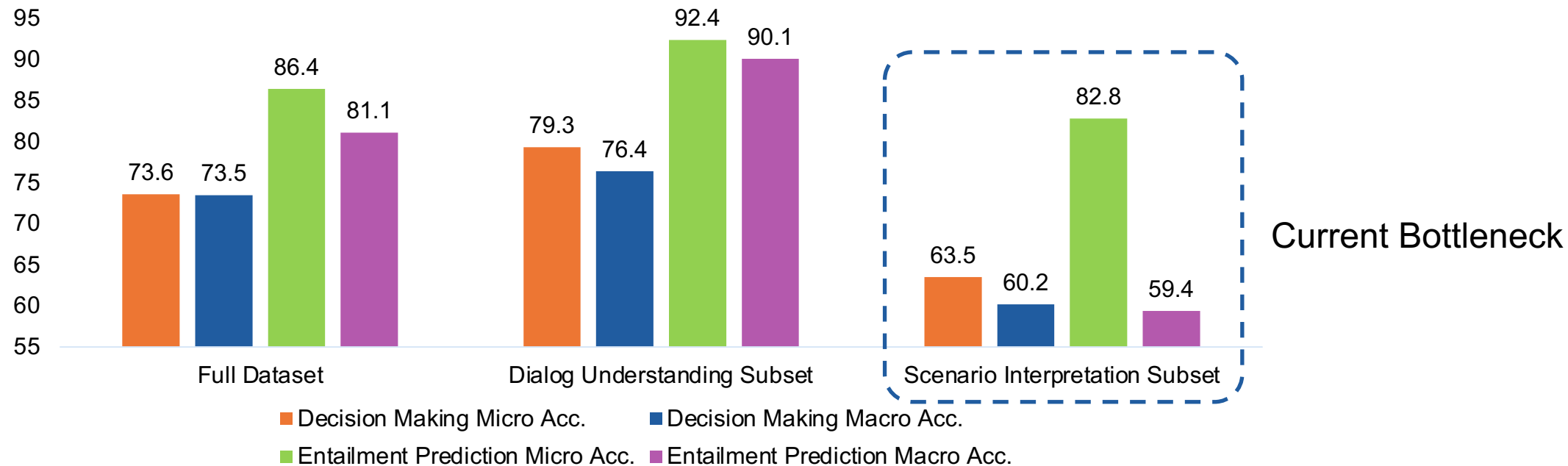


Experiments

How Far Has the Problem Been Solved?

Idea: Disentangle the challenge between scenario interpretation and dialog understanding

- **Full Dataset:** Full development set of ShARC
- **Dialog Understanding Subset:** User information only contains dialog history
- **Scenario Interpretation Subset:** User information only contains user scenario



Conclusion

- We present Discern, a system that does discourse-aware entailment reasoning for conversational machine reading.
- Results on the ShARC benchmark shows that Discern outperforms existing methods by a large margin.
- We also conduct comprehensive analyses to unveil the limitations of Discern and challenges for ShARC.

Thanks!

Code & Models: <https://github.com/Yifan-Gao/Discern>