Filtering before Iteratively Referring for Knowledge-Grounded Response Selection in Retrieval-Based Chatbots

Jia-Chen Gu[1], Zhen-Hua Ling[1], Quan Liu[1,2], Zhigang Chen[2], Xiaodan Zhu[3]

[1]National Engineering Laboratory for Speech and Language Information Processing, University of Science and Technology of China

[2]State Key Laboratory of Cognitive Intelligence, iFLYTEK Research

[3]ECE & Ingenuity Labs, Queen's University

Outline

- **Introduction**
- Filtering before Iteratively Referring (FIRE)
- Experiments
- Conclusion

Introduction

- Knowledge-Grounded Response Selection

  The task is to select the best-matched response from a set of candidates given the context of a conversation and the background knowledge.

Background Knowledge	
Name	The inception
Year	2009
Director	Christopher Nolan
Genre	Scientific
Cast	Leonardo DiCaprio as Dom Cobb, a professional thief who specializes in conning secrets from his victims by infiltrating their dreams. Tom Hardy as Eames, a sharp-tongued associate of Cobb. ...
Critical Response	Response DiCaprio, who has never been better as the tortured hero, draws you in with a love story that will appeal even to non-scifi fans. The movie is a metaphor for the power of delusional hype for itself. ...
Introd--uction	Dominick Cobb and Arthur are extractors, who perform corporate espionage using an experimental military technology to infiltrate the subconscious of their targets and extract valuable information through a shared dream world. Their latest target, Japanese businessman Saito, reveals that he arranged the mission himself to test Cobb for a seemingly impossible job: planting an idea in a person's subconscious, or inception.
Rating	Rotten Tomatoes: 86% and average: 8.1/10; IMDB: 8.8/10
Conversation	
User 2:	Hi how are you today?
User 1:	I am good. How are you?
User 2:	Pretty good. Have you seen the inception?
User 1:	No, I have not but have heard of it. What is it about?
User 2:	It's about extractors that perform experiments using military technology on people to retrieve info about their targets.
User 1:	Sounds interesting. Do you know which actors are in it?
User 2:	I haven't watched it either or seen a preview. But it's scifi so it might be good. Ugh Leonardo DiCaprio is the main character.
User 2:	He plays as Don Cobb.
User 1:	I'm not a big scifi fan but there are a few movies I still enjoy in that genre.
User 1:	Is it a long movie?
User 2:	Doesn't say how long it is.
User 2:	The Rotten Tomatoes score is 86%.

Figure 1: An example from CMU_DoG dataset (Zhou et al., 2018a). Words in the same color are related.

Outline

- Introduction
- **Filtering before Iteratively Referring (FIRE)**
- Experiments
- Conclusion

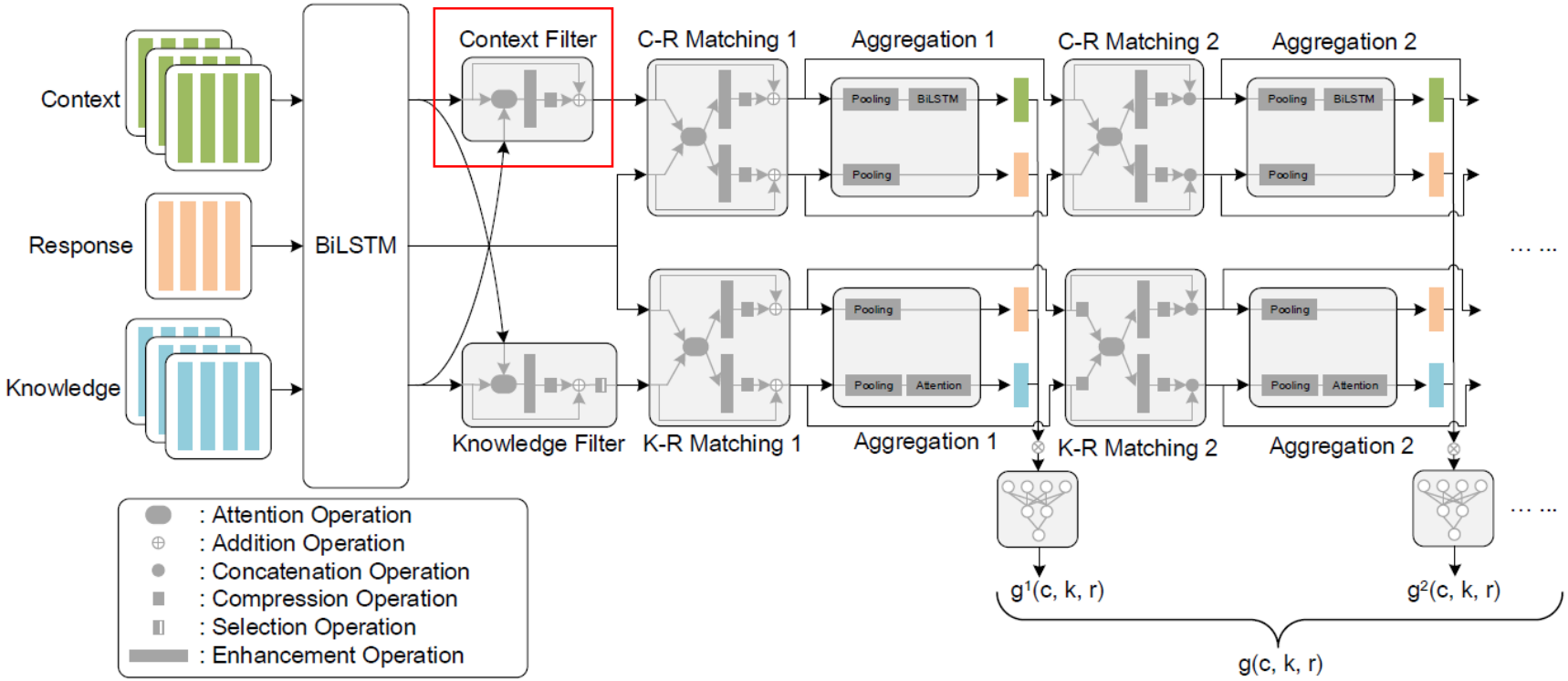
Filtering before Iteratively Referring (FIRE)

Two issues that FIRE is designed to address:

- Ground a conversation on the background knowledge from a global view and select relevant knowledge entries.
- Match response candidates with both context and knowledge simultaneously and deeply.

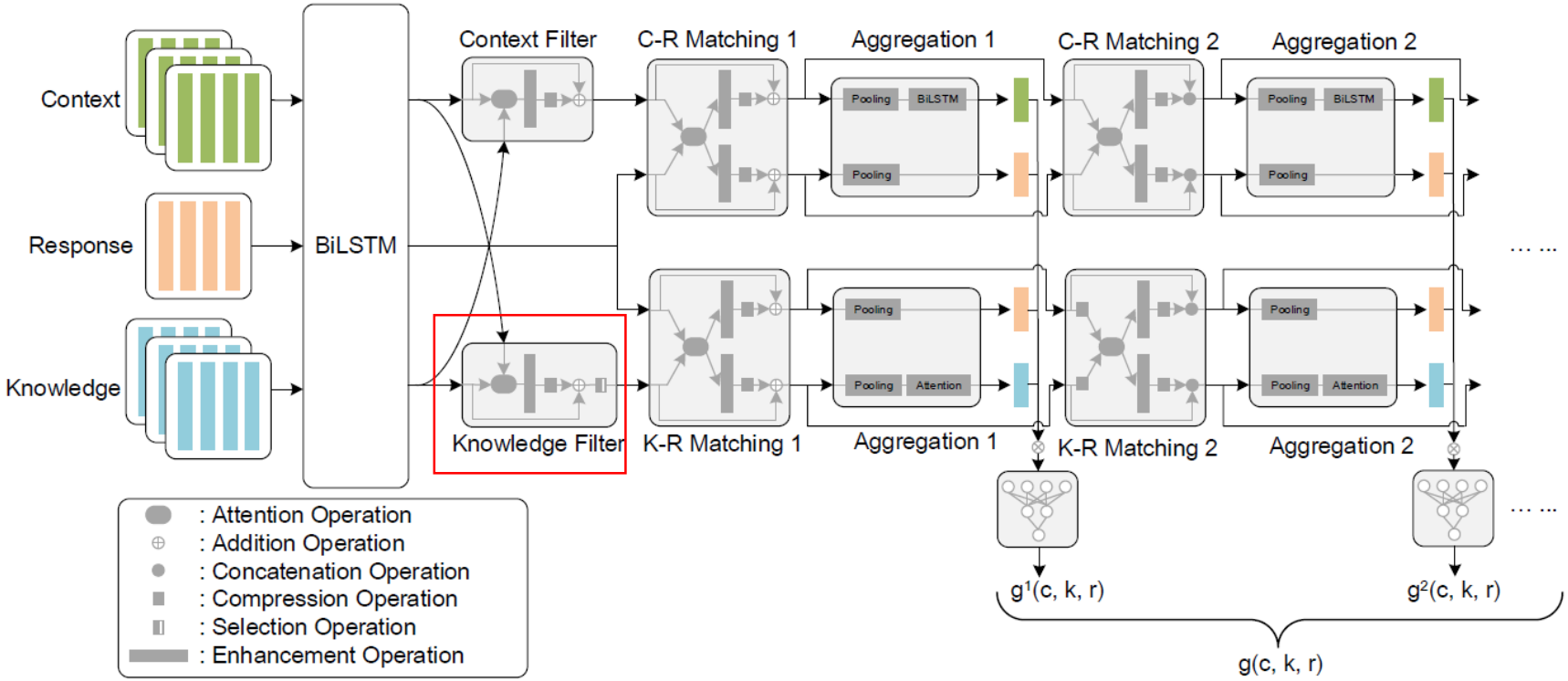
Filtering before Iteratively Referring (FIRE)

A context filter is designed to collect the global matching information between context and knowledge to derive the knowledge-aware context representations.



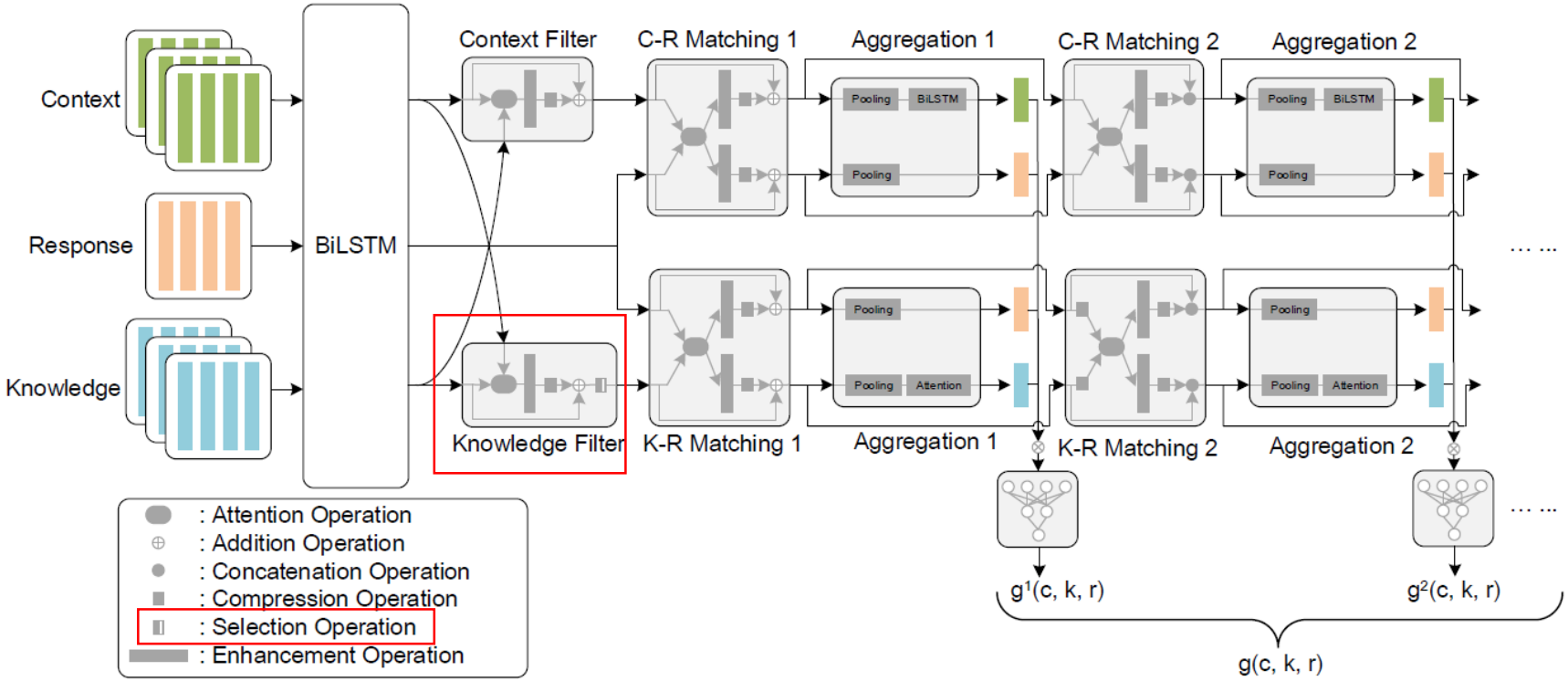
Filtering before Iteratively Referring (FIRE)

A knowledge filter is designed to collect the global matching information between context and knowledge to derive the context-aware knowledge representations.



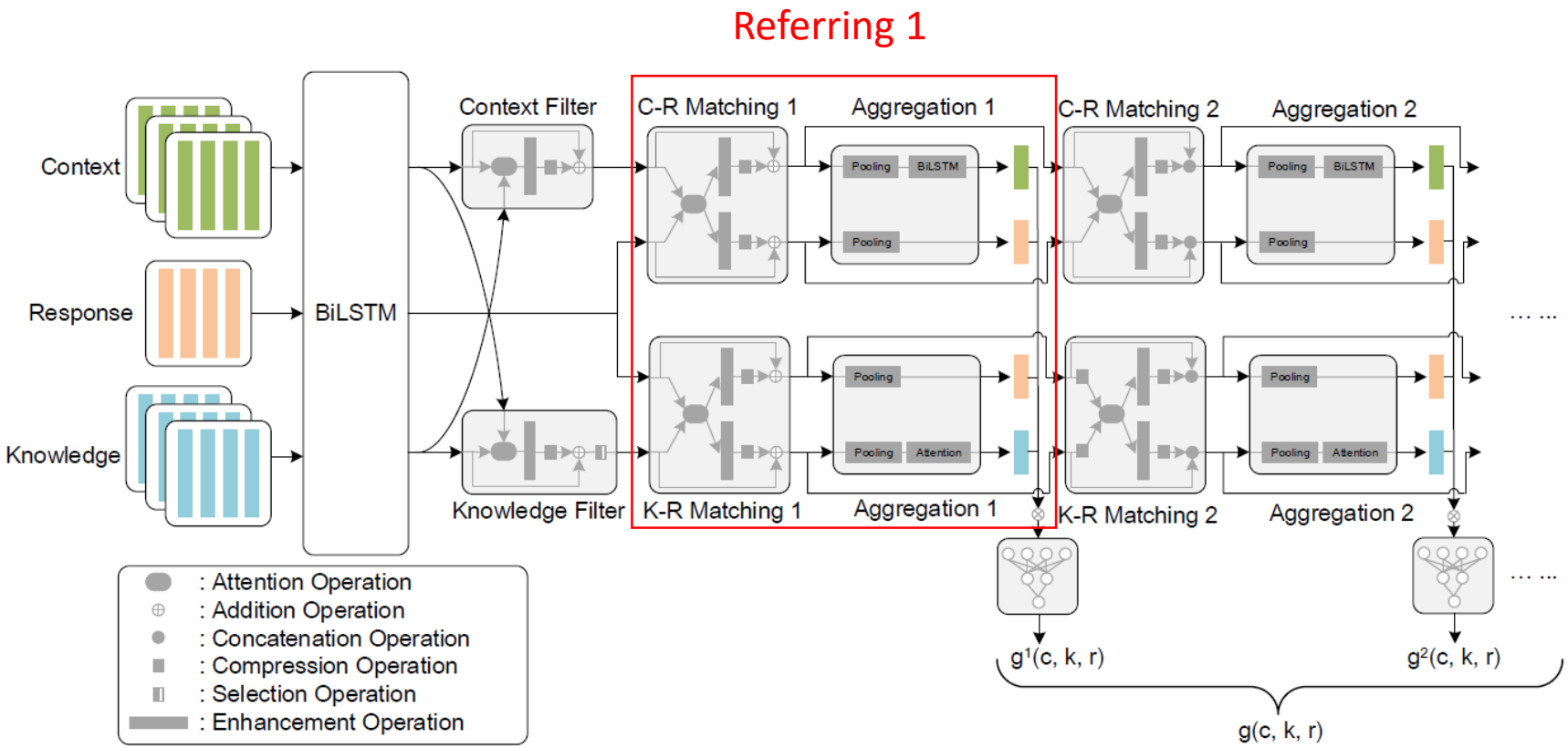
Filtering before Iteratively Referring (FIRE)

Additionally, the knowledge filter discards irrelevant entries, by calculating the similarity between each entry and the whole context, considering the knowledge entries are independent of each other.



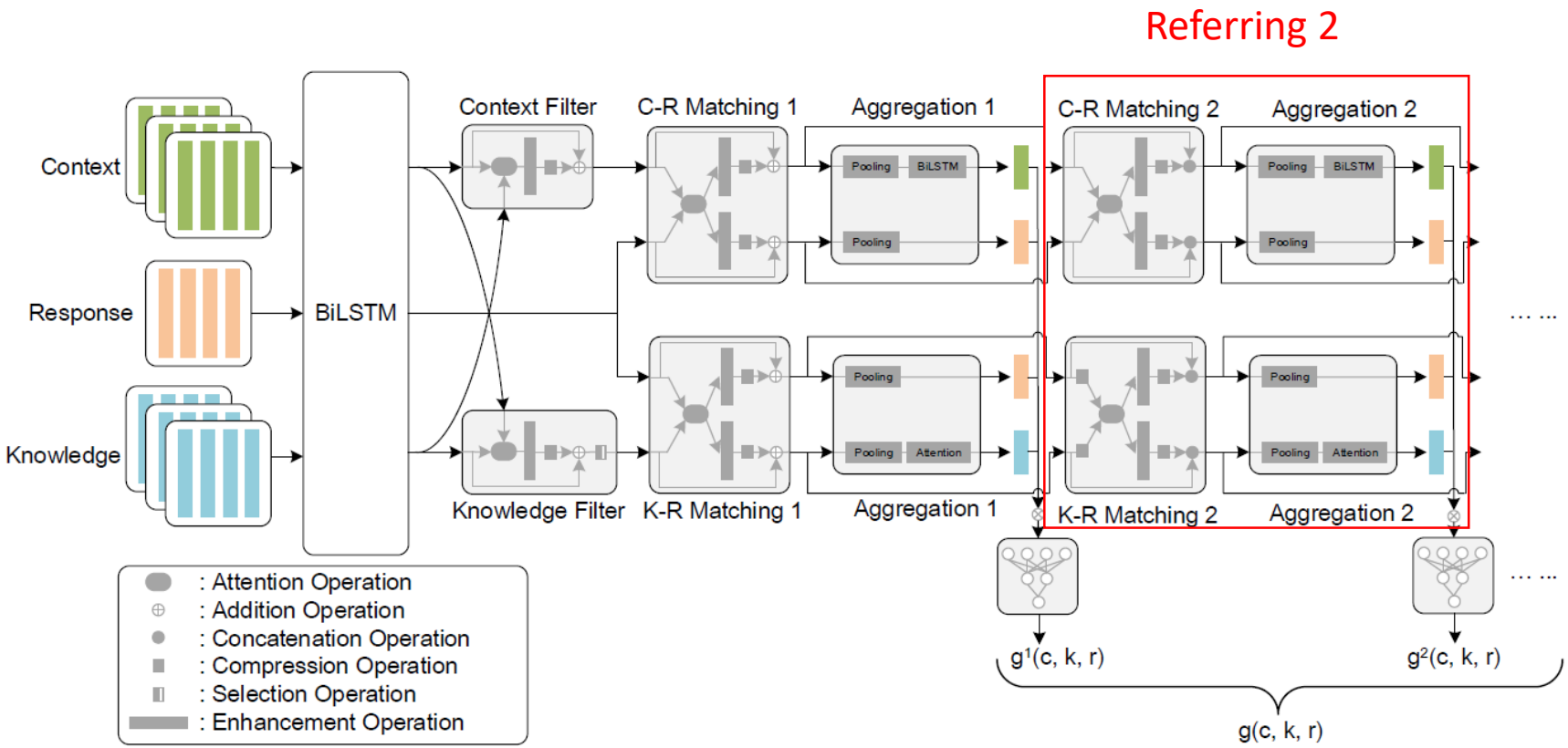
Filtering before Iteratively Referring (FIRE)

An iteratively referring network is designed to capture the deep matching information.



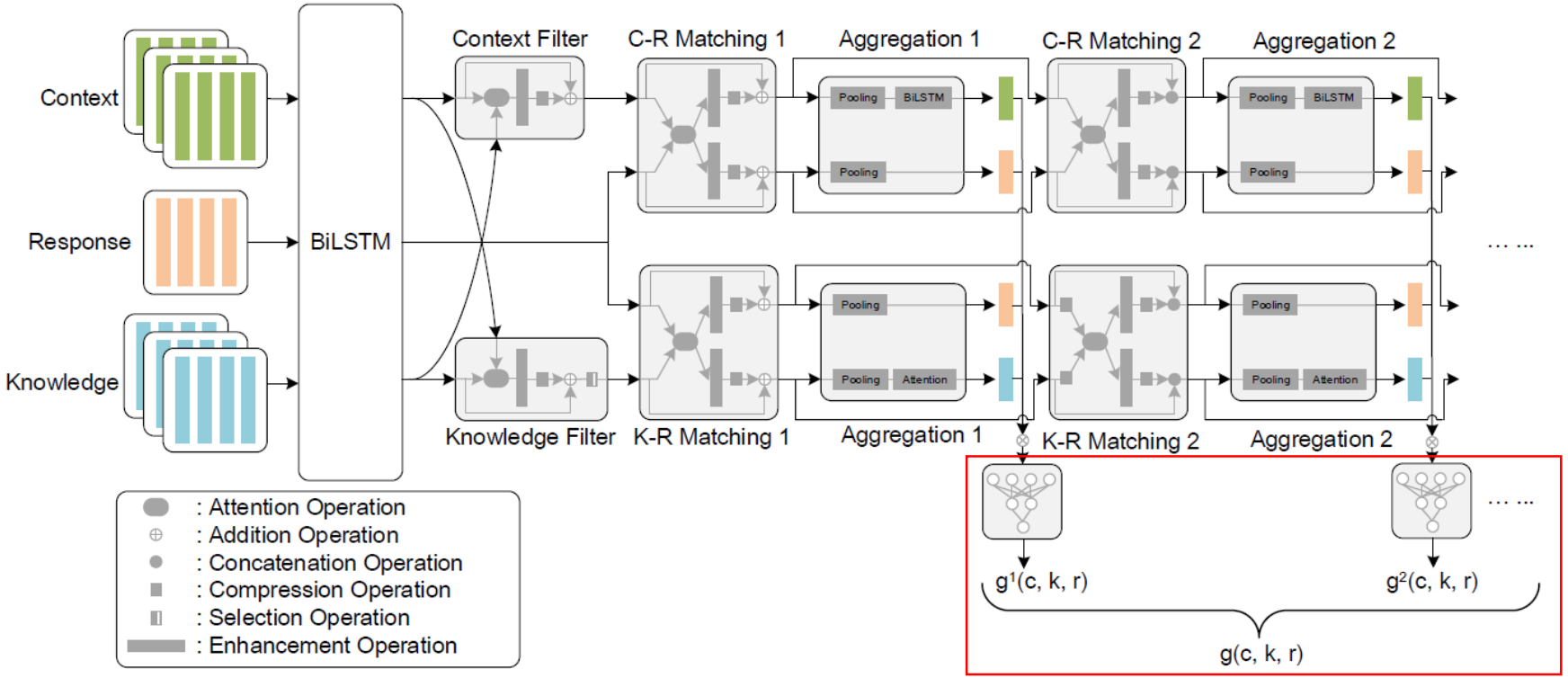
Filtering before Iteratively Referring (FIRE)

The outputs of the *l*-th iteration are the inputs of the (*l+1*)-th iteration.



Filtering before Iteratively Referring (FIRE)

Accumulating all these iterations can help to derive the deep and comprehensive matching features for response selection



Outline

- Introduction
- Filtering before Iteratively Referring (FIRE)
- **Experiments**
- Conclusion

Experiments

- Datasets

Datasets	# candidates	Train	Valid	Test
Persona-Chat	20	65719	7801	7512
CMU_DoG	20	36159	2425	6637

- Metrics

  The recall of true positive replies by selecting *k* best-matched response from available candidates for the given context and knowledge, denoted as R@k.

Experiments

- Overall Performance

Model	PERSONA-CHAT						CMU_DoG		
	Original			Revised					
	R@1	**R@2**	**R@5**	**R@1**	**R@2**	**R@5**	**R@1**	**R@2**	**R@5**
Starspace (Wu et al., 2018)	49.1	60.2	76.5	32.2	48.3	66.7	50.7	64.5	80.3
Profile Memory (Zhang et al., 2018a)	50.9	60.7	75.7	35.4	48.3	67.5	51.6	65.8	81.4
KV Profile Memory (Zhang et al., 2018a)	51.1	61.8	77.4	35.1	45.7	66.3	56.1	69.9	82.4
Transformer (Mazaré et al., 2018)	54.2	68.3	83.8	42.1	56.5	75.0	60.3	74.4	87.4
DGMN (Zhao et al., 2019)	67.6	80.2	92.9	58.8	62.5	87.7	65.6	78.3	91.2
DIM (Gu et al., 2019b)	78.8	89.5	97.0	70.7	84.2	95.0	78.7	89.0	97.1
FIRE (Ours)	**81.6**	**91.2**	**97.8**	**74.8**	**86.9**	**95.9**	**81.8**	**90.8**	**97.4**

FIRE achieves new state-of-the-art performances.

Experiments

- Ablation
  - Remove iteratively referring by setting the number of iterations L to one.
  - Remove the two filters.

Model	PERSONA-CHAT		CMU_DoG
	Original	Revised	
	R@1	**R**@1	**R**@1
FIRE	82.3	75.2	83.4
- Ite. Ref.	81.3	73.8	81.6
- Filters	78.9	71.1	78.8
C-R	65.6	66.2	79.7
C-R → Fusion	67.0	66.4	80.9
Filter → C-R	78.8	70.2	81.4
K-R	51.6	34.3	57.8
K-R → Fusion	54.2	39.4	63.1
Filter → K-R	63.6	51.0	73.5

Table 2: The results of ablation tests on the validation sets. Here, C-R denotes context-response matching and K-R denotes knowledge-response matching. The symbol → indicates the order of operations.

Experiments

- Operation Order
  - For single context-response matching, fusion after matching and filtering before matching can both improve the performance of response selection after introducing knowledge.
  - Filtering before matching outperformed fusion after matching.
  - The same conclusion for knowledge-response matching after introducing context.

Model	PERSONA-CHAT		CMU_DoG
	Original	Revised	
	R@1	**R**@1	**R**@1
FIRE	82.3	75.2	83.4
- Ite. Ref.	81.3	73.8	81.6
- Filters	78.9	71.1	78.8
C-R	65.6	66.2	79.7
C-R → Fusion	67.0	66.4	80.9
Filter → C-R	78.8	70.2	81.4
K-R	51.6	34.3	57.8
K-R → Fusion	54.2	39.4	63.1
Filter → K-R	63.6	51.0	73.5

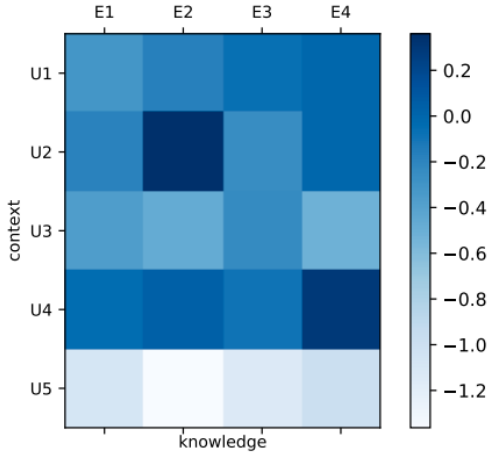
Table 2: The results of ablation tests on the validation sets. Here, C-R denotes context-response matching and K-R denotes knowledge-response matching. The symbol → indicates the order of operations.

Experiments

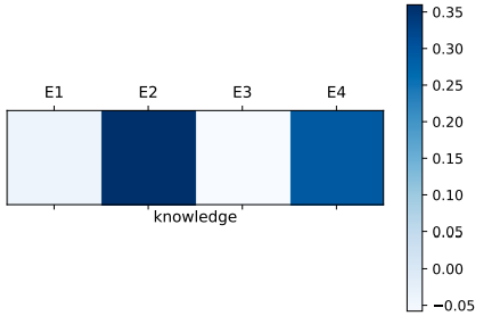
- Case Study

Context Utterances	
U1	hey , are you a student , i traveled a lot , i even studied abroad .
U2	no , i work full time at a nursing home . i am a nurses aide .
U3	nice , i just got a advertising job myself . do you like your job ?
U4	nice . yes i do . caring for people is the joy of my life .
U5	nice my best friend is a nurse , i knew him since kindergarten .
Knowledge Entries	
E1	i have two dogs and one cat .
E2	i work as a nurses aide in a nursing home .
E3	i love to ride my bike .
E4	i love caring for people .

Table 3: Context utterances and knowledge entries of a sample in the test set of the PERSONA-CHAT dataset.



Utterance-entry similarity. U2 and U4 obtained large attention weights with E2 and E4 respectively.



Aggregated conversation-entry similarity. Irrelevant entries E1 and E3 obtained small similarity scores with the conversation.

Experiments

- Knowledge Selection
  - When γ = 0, no knowledge entries were filtered out.
  - The performance was improved when increasing γ at the beginning, which indicates that filtering out irrelevant entries indeed helped.
  - The performance started to drop when γ was too large since some indeed relevant entries may be filtered out by mistake.

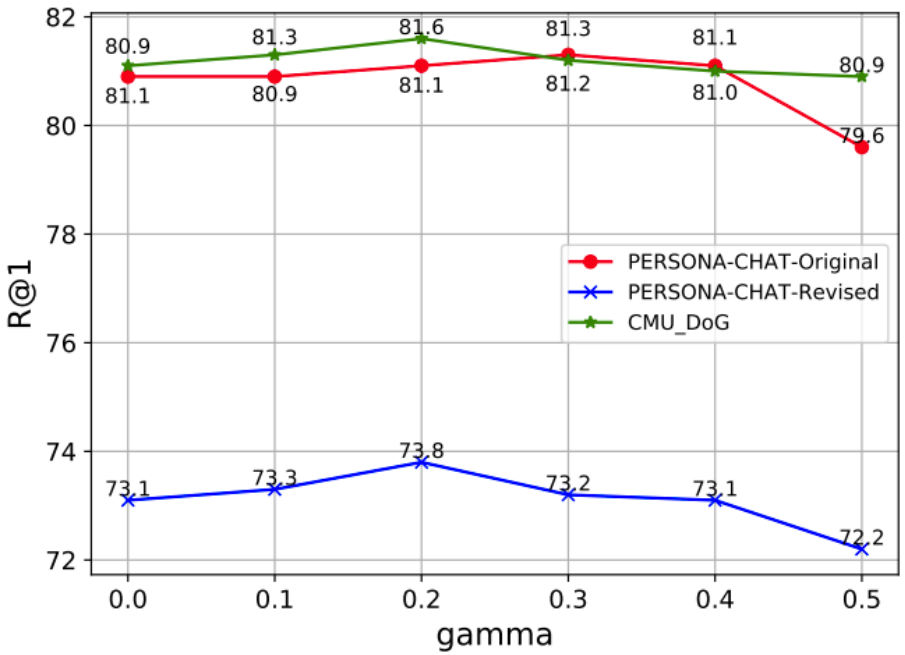


Figure 4: Validation set performance of FIRE with different threshold γ in the knowledge filter.

Experiments

- Iteratively Referring
  - Three iterations led to the best performance on datasets.

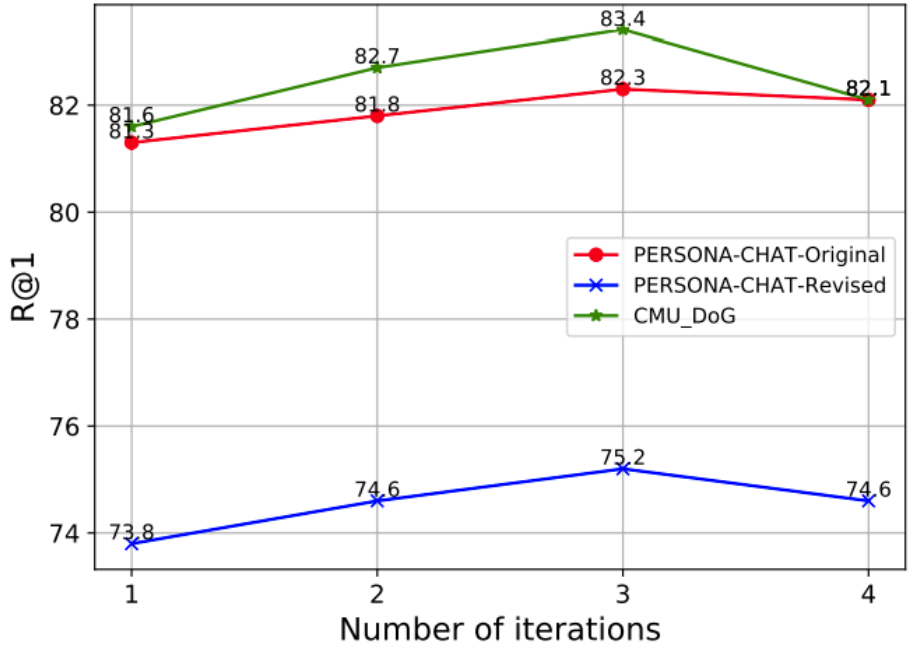


Figure 5: Validation set performance of FIRE with different number of iterations in iteratively referring.

Experiments

- Complexity
  - It takes FIRE 109.5s and DIM 160.4s to finish the inference over the validation set of PERSONA-CHAT using a GeForce GTX 1080 Ti GPU, which shows that FIRE is more time-efficient.
  - The reason is that we design a lighter aggregation method in FIRE.

Outline

- Introduction
- Filtering before Iteratively Referring (FIRE)
- Experiments
- **Conclusion**

Conclusion

- Grounding a conversation on its background knowledge and selecting relevant knowledge entries are important steps for knowledge-grounded dialogues.
- Designing deep matching models for response and given context and knowledge is useful for selecting an appropriate response.

Jia-Chen Gu

Zhen-Hua Ling

Quan Liu

Zhigang Chen

Xiaodan Zhu

Thanks!

Code: https://github.com/JasonForJoy/FIRE