DukeNet: A Dual Knowledge Interaction Network for Knowledge-Grounded Conversation

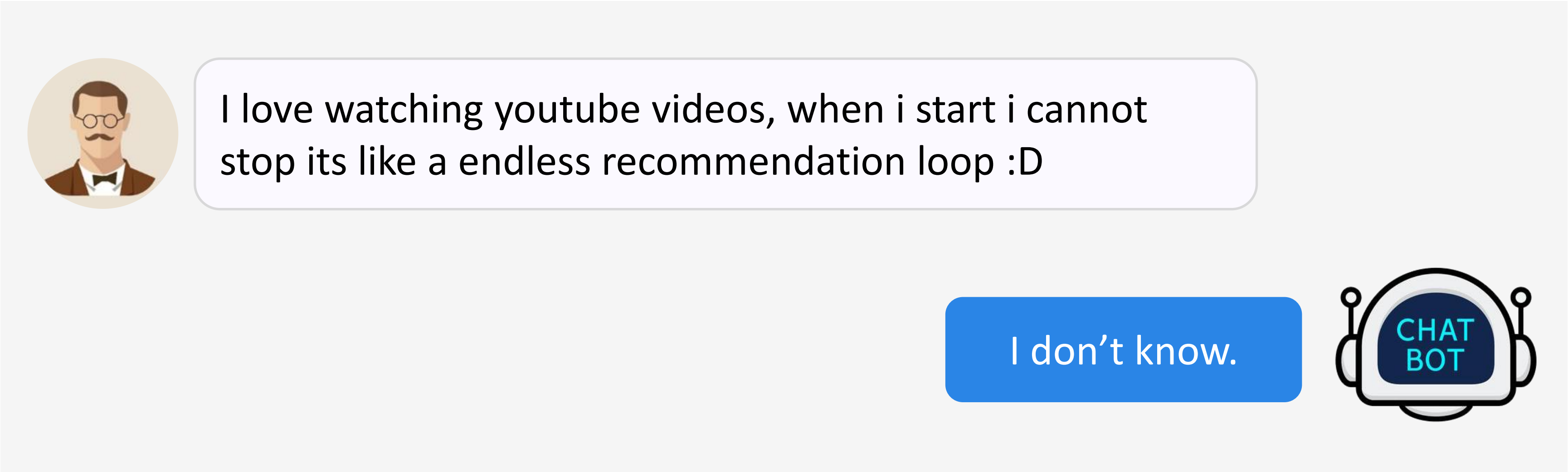
Chuan Meng[1], Pengjie Ren[2], Zhumin Chen[1], Weiwei Sun[1], Zhaochun Ren[1], Zhaopeng Tu[4], Maarten de Rijke[2,3]

[1]Shandong University
[2]University of Amsterdam
[3]Ahold Delhaize
[4]Tencent AI Lab

Outline

Non-Task-Oriented Conversational Agents (Chatbot)

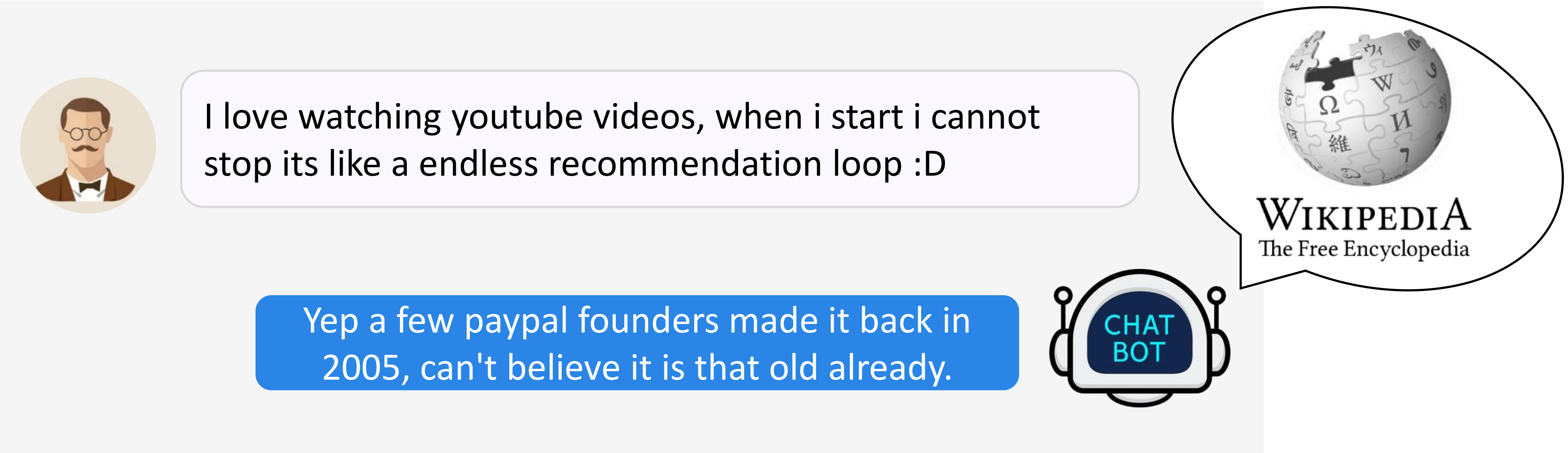
- Such agents using sequence-to-sequence learning[1] tend to produce non-informative responses[2].



[1] Sutskever et al. Sequence to sequence learning with neural networks. In NeurIPS 2014.
[2] Li et al. A diversity-promoting objective function for neural conversation models. In NAACL 2016.

Knowledge-Grounded Conversation (KGC)

- Knowledge-Grounded Conversation (KGC) can address this issue via referring to relevant knowledge[1-3]



[1] Meng et al. A Reference-aware Network for Background Based Conversation. In AAAI 2020.
[2] Ren et al. Thinking Globally, Acting Locally: Distantly Supervised Global-to-Local Knowledge Selection for Background Based Conversation. In AAAI 2020.
[3] Zhang et al. Improving Background Based Conversation with Context-aware Knowledge Pre-selection. In SCAI 2019

Knowledge-Grounded Conversation (KGC)

- KGC can be divided into two sequential subtasks[1,2]
  - Knowledge Selection (KS): select the appropriate knowledge fragment from a knowledge pool
  - Response Generation (RG): generate a response based on the selected knowledge fragment and context

Knowledge Pool

K1 YouTube was created by three former PayPal employees—Chad Hurley, Steve Chen, and Jawed Karim—in February 2005. **Selected knowledge**

K2 YouTube is an American video-sharing website headquartered in San Bruno, California.

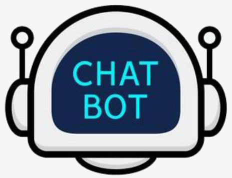
K3 Google bought the site in November 2006 for US$1.65 billion; YouTube now operates as one of Google's subsidiaries.

…

Conversation

I love watching youtube videos, when i start i cannot stop its like a endless recommendation loop :D

Yep a few paypal founders made it back in 2005, can't believe it is that old already.



[1] Dinan et al. Wizard of Wikipedia: Knowledge-Powered Conversational agents. In ICLR 2019.
[2] Kim et al. Sequential Latent Knowledge Selection for Knowledge-Grounded Dialogue. In ICLR 2020.

Knowledge-Grounded Conversation (KGC)

- Knowledge Selection (KS) is crucial[1,2]
  - Decides the topic to be talked in the future response
  - Selecting inappropriate knowledge will directly result in an inappropriate response

Knowledge Pool

K1 YouTube was created by three former PayPal employees—Chad Hurley, Steve Chen, and Jawed Karim—in February 2005.

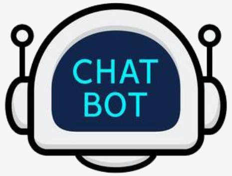
K2 YouTube is an American video-sharing website headquartered in San Bruno, California.

K3 Google bought the site in November 2006 for US$1.65 billion; YouTube now operates as one of Google's subsidiaries.

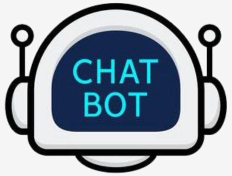
…

Correspoding Responses

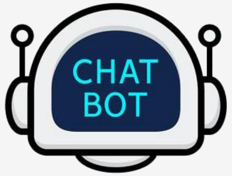
Yep a few paypal founders made it back in 2005, can't believe it is that old already.



Yep Youtube is a well-known American video-sharing website.



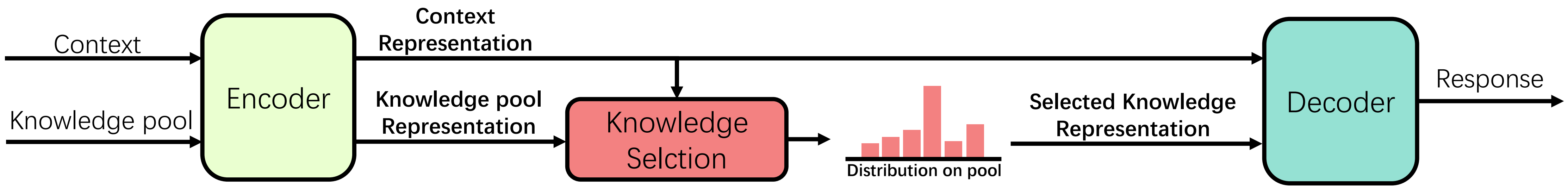
Yea to think google bought it only one year later for over a billion dollars, pretty insane but a good purchase.



[1] Zhang et al. Improving Background Based Conversation with Context-aware Knowledge Pre-selection. In SCAI 2019.
[2] Kim et al. Sequential Latent Knowledge Selection for Knowledge-Grounded Dialogue. In ICLR 2020.

Related work about **Modeling** on KS

- Most works predict the next knowledge fragment based on the context[1,2]



[1] Ghazvininejad et al. A knowledge grounded neural conversation model. In AAAI 2018.
[2] Dinan et al. Wizard of Wikipedia: Knowledge-Powered Conversational agents. In ICLR 2019.

Related work about **Modeling** on KS

- **Problem**: knowledge tracking and knowledge shifting are not explicitly considered
  - Knowledge tracking and shifting capture the interaction between the knowledge at adjacent turns

Knowledge Pool

K1 YouTube was created by three former PayPal employees—Chad Hurley, Steve Chen, and Jawed Karim—in February 2005. Tracked knowledge

K2 YouTube is an American video-sharing website headquartered in San Bruno, California.

Interaction

K3 Google bought the site in November 2006 for US$1.65 billion; YouTube now operates as one of Google's subsidiaries. Shifted knowledge

…

Conversation

Context

I love watching youtube videos, when i start i cannot stop its like a endless recommendation loop :D

Yep a few paypal founders made it back in 2005, can't believe it is that old already.

CHAT BOT

Yes, its the most viewed multimedia place, if you can call it that way, in the world!

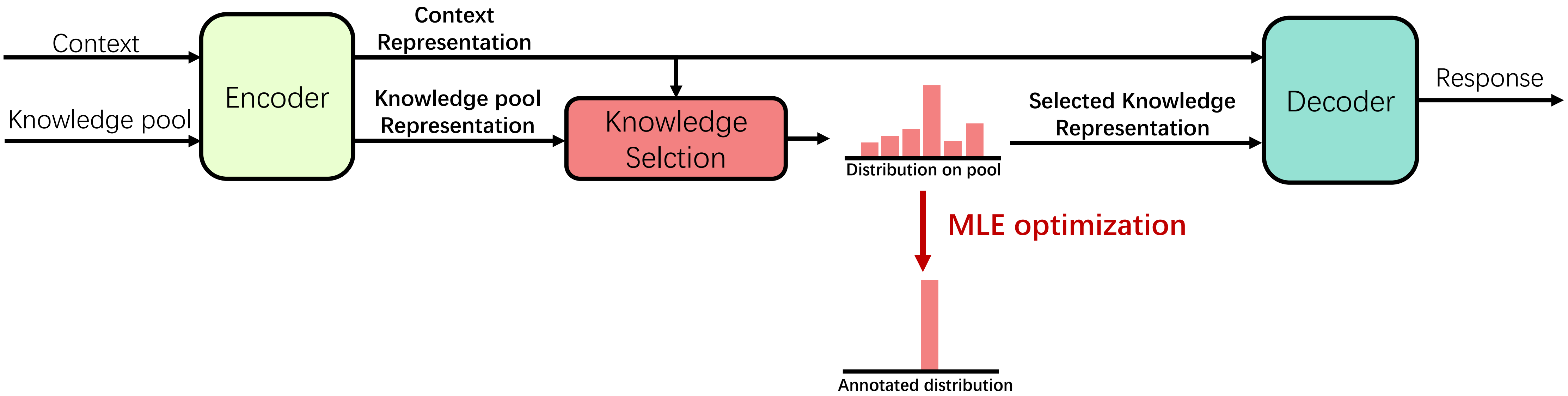
Response

Yea to think google bought it only one year later for over a billion dollars, pretty insane but a good purchase.

CHAT BOT

Related work about **Learning** on KS

- Most works optimize their models via **Maximum Likelihood Estimation** (MLE) based on annotation information in the training set[1-5]



[1] Meng et al. A Reference-aware Network for Background Based Conversation. In AAAI 2020.
[2] Ghazvininejad et al. A knowledge grounded neural conversation model. In AAAI 2018.
[3] Dinan et al. Wizard of Wikipedia: Knowledge-Powered Conversational agents. In ICLR 2019.
[4] Lian at al. Learning to Select Knowledge for Response Generation in Dialog Systems. In IJCAI 2019.
[5] Kim et al. Sequential Latent Knowledge Selection for Knowledge-Grounded Dialogue. In ICLR 2020.

Related work about **Learning** on KS

- **Problem**: many-to-many mapping phenomenon between context and knowledge is not addressed effectively
  - Unlike Question Answering (QA) tasks: almost each query has only one unique answer

Knowledge base

K1 Barack Obama

K2 Donald John Trump
Selected knowledge

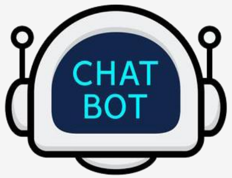
K3 Joseph Robinette Biden

…

Question answering

Who is the current President of the US?

Donald John Trump



Related work about **Learning** on KS

- **Problem**: many-to-many mapping phenomenon between context and knowledge is not addressed effectively
  - Unlike Question Answering (QA) tasks: almost each query has only one unique answer
  - It is really hard to annotate all suitable knowledge fragments for a context when constructing a dataset
  - MLE only optimize the annotated knowledge and do not try any other reasonable knowledge

Knowledge Pool

K1 YouTube was created by three former PayPal employees—Chad Hurley, Steve Chen, and Jawed Karim—in February 2005.

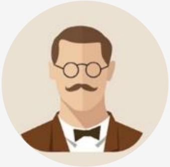
K2 YouTube is an American video-sharing website headquartered in San Bruno, California.

K3 Google bought the site in November 2006 for US$1.65 billion; YouTube now operates as one of Google's subsidiaries.

...

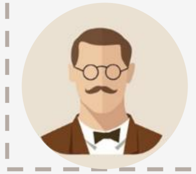
The Cases of contexts

One-to-many mapping: K1, K2 and K3 are all suitable in this context



I love watching youtube videos, when i start i cannot stop its like a endless recommendation loop :D

Many-to-one mapping: K1 is also reasonable in this context



I would like to know who created YouTube?

Concerns

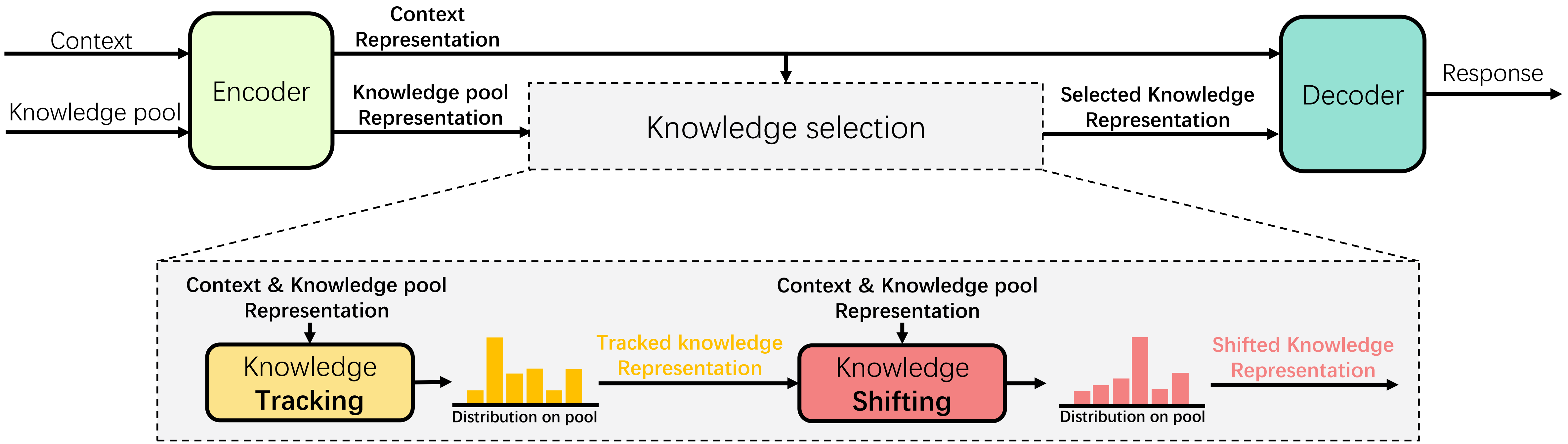
Can we explicitly model **knowledge tracking and shifting** while handling the **many-to-many mapping phenomenon**?

Outline

- Background and motivation
- **Model**
- Experiments
- Conclusion

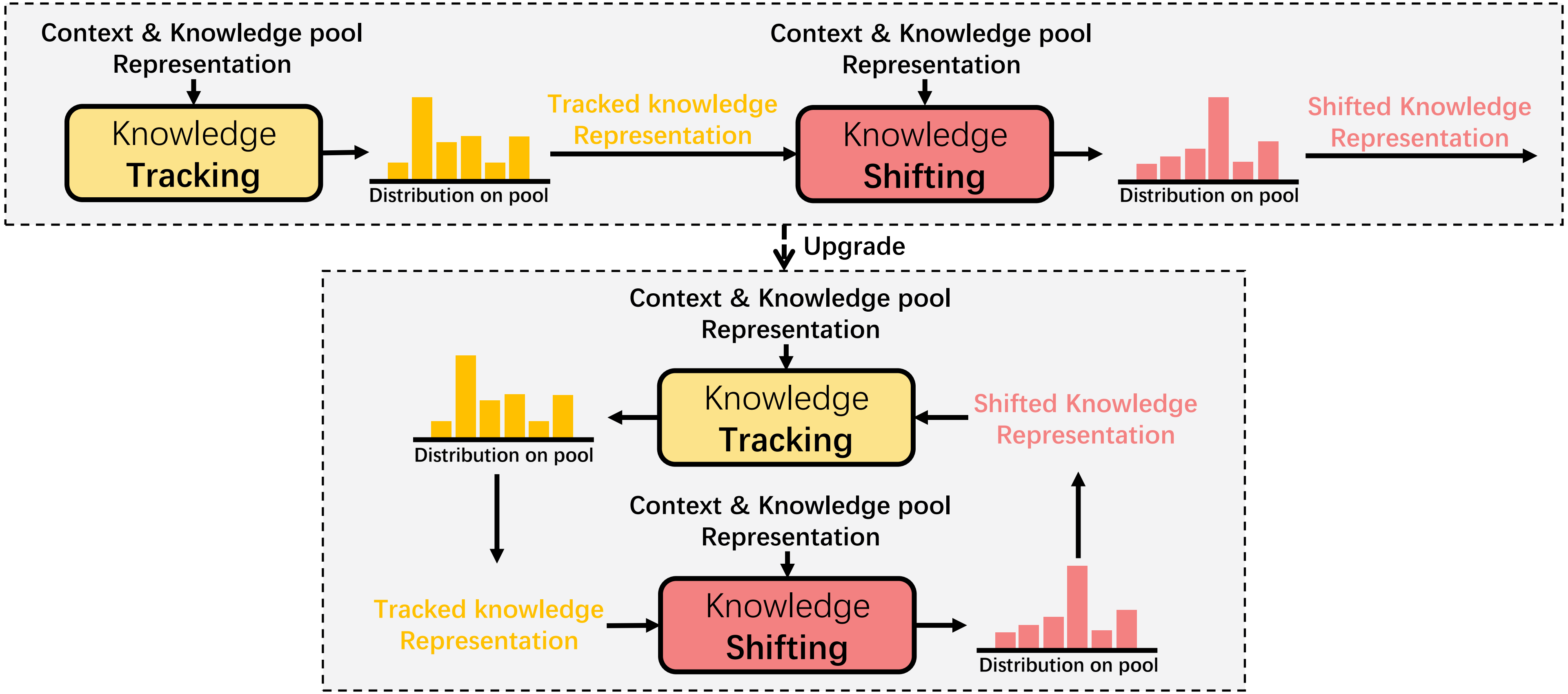
DukeNet

- **DukeNet (Dual Knowledge Interaction Network)**
  - Explicitly models knowledge tracking and knowledge shifting



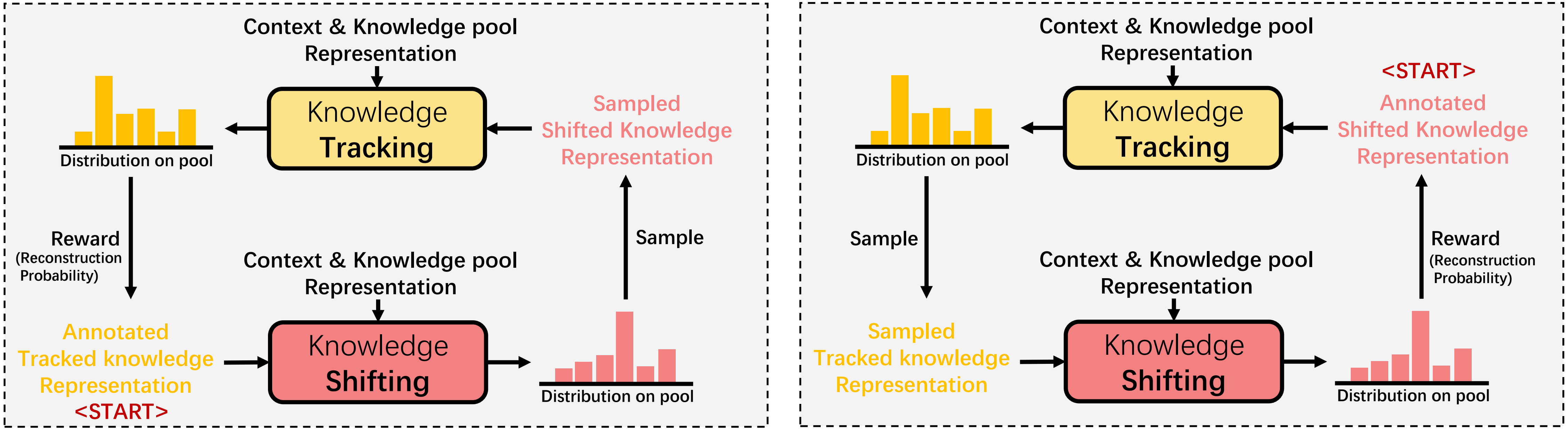
DukeNet

- **DukeNet (Dual Knowledge Interaction Network)**
  - Explicitly models knowledge tracking and knowledge shifting
  - Further regards them as **dual tasks**



DukeL

- **DukeL (Dual Knowledge Interaction Learning)**
  - Enable knowledge tracking and shifting to teach each other in an unsupervised learning way
  - Alternate **i** and **ii** until convergence
  - The processes of **sampling** and **rewarding** will introduce randomness to **explore extra suitable knowledge** that is not annotated in the training set, which handles the many-to-many mapping phenomenon

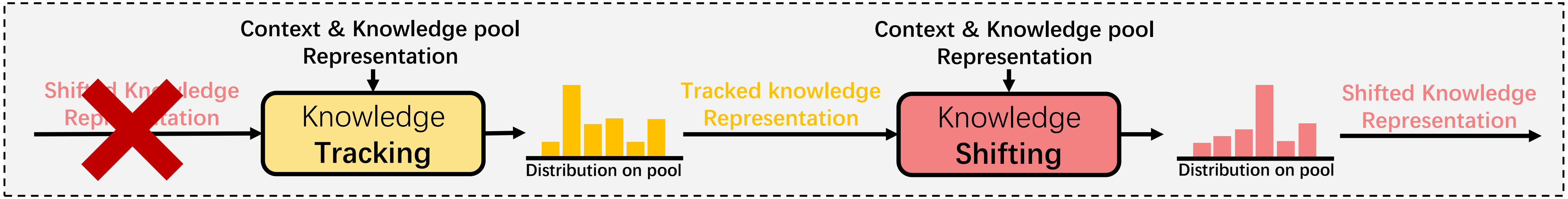


i. Optimize knowledge shifting

ii. Optimize knowledge tracking

Incompatible Dual Processes

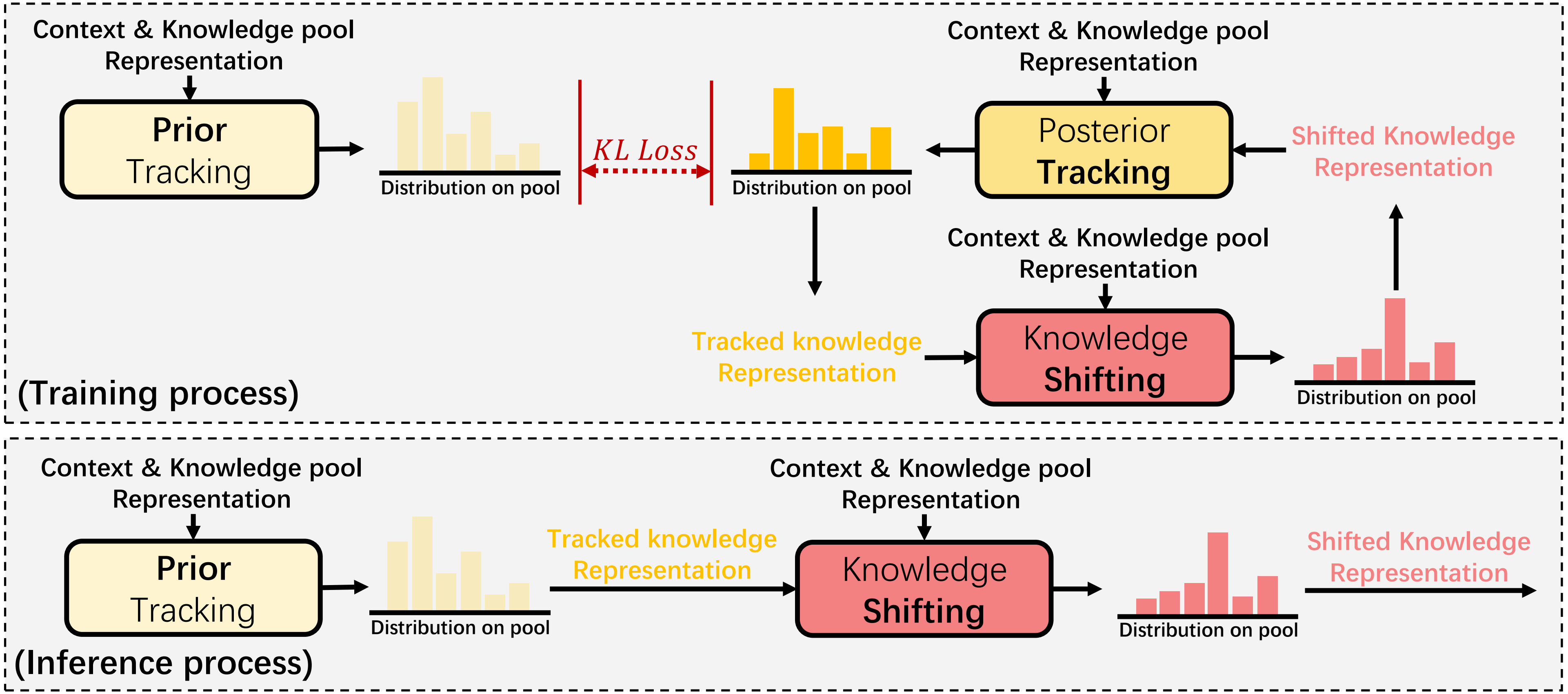
- **Problem: Incompatible dual processes between training and inference**
  - During inference knowledge tracking and shifting are executed **in order**
  - Knowledge tracking cannot get the shifted knowledge as input



Cannot get it during inference

Incompatible Dual Processes

- **Solution: Distinguish knowledge tracking as prior and posterior knowledge tracking**
  - During training, force the prior knowledge tracking to simulate the output of posterior knowledge tracking
  - During inference, only prior knowledge tracking and shifting are executed



Outline

- Background and motivation
- Model
- **Experiments**
- Conclusion

Experiments: Dataset

- Wizard of Wikipedia[1]
    - Contain the ground truth annotation for KS
    - Two versions of the test set: Test Seen and Test Unseen
- Holl-E[2]
    - We use the modified version released by [3]
    - Two versions of the test set: Single golden reference and Multiple golden references

[1] Dinan et al. Wizard of Wikipedia: Knowledge-Powered Conversational agents. In ICLR 2019.
[2] Moghe et al. Towards exploiting background knowledge for building conversation systems. In EMNLP 2018.
[3] Kim et al. Sequential Latent Knowledge Selection for Knowledge-Grounded Dialogue. In ICLR 2020.

Experiments: Baselines

- Given only context
  - Seq2Seq[1]
  - Transformer[2]
- Given knowledge and context
  - MemNet[3]
  - TMemNet[4] (+BERT [7] enocder)
  - PostKS[5] (+BERT[7] enocder)
  - SKT [6]

[1] Sutskever et al. Sequence to sequence learning with neural networks. In NeurIPS 2014.
[2] Vaswani et al. Attention Is All You Need. In NeurIPS 2017.
[3] Ghazvininejad et al. A knowledge grounded neural conversation model. In AAAI 2018.
[4] Dinan et al. Wizard of Wikipedia: Knowledge-Powered Conversational agents. In ICLR 2019.
[5] Lian at al. Learning to Select Knowledge for Response Generation in Dialog Systems. In IJCAI 2019.
[6] Kim et al. Sequential Latent Knowledge Selection for Knowledge-Grounded Dialogue. In ICLR 2020.
[7] Devlin et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In NAACL 2018.

Experiments: Metrics

- Automatic evaluation
  - For **KS**: Hit@1 (top 1 accuracy)[4,5]
  - For **RG**: BLEU-4[1], METEOR[2], ROUGE-1[3], ROUGE-2[3], ROUGE-L[3]
- Human evaluation
  - For **KS**: Appropriateness
  - For **RG**: Informativeness[6], Engagingness[6]

[1] Papineni et al. BLEU: a Method for Automatic Evaluation of Machine Translation. In ACL 2002
[2] Denkowski et al. Meteor Universal: Language Specifc Translation Evaluation for Any Target Language. In WSMT 2014.
[3] Lin. Rouge: A package for automatic evaluation of summaries. In WTSBO 2004.
[4] Dinan et al. Wizard of Wikipedia: Knowledge-Powered Conversational agents. In ICLR 2019.
[5] Liu et al. Knowledge Aware Conversation Generation with Reasoning on Augmented Graph. In EMNLP 2019.
[6] Kim et al. Sequential Latent Knowledge Selection for Knowledge-Grounded Dialogue. In ICLR 2020.

Experiments: Automatic Evaluation

Methods	Test Seen (%)						Test Unseen (%)					
	BLEU-4	METEOR	ROUGE-1	ROUGE-2	ROUGE-L	Hit@1	BLEU-4	METEOR	ROUGE-1	ROUGE-2	ROUGE-L	Hit@1
Seq2Seq	0.46	12.22	20.32	3.03	14.46	–	0.34	10.21	19.01	2.16	13.55	–
Transformer	0.39	12.82	20.50	3.27	15.00	–	0.39	11.36	18.81	2.17	14.33	–
MemNet	0.41	12.35	21.51	3.17	15.33	4.27	0.32	11.75	20.06	2.51	14.68	4.13
PostKS	0.57	13.50	21.61	3.66	16.07	4.70	0.36	12.60	20.79	2.52	14.88	4.45
TMemNet	1.35	14.52	22.84	4.31	16.77	21.57	0.43	12.82	21.33	3.05	15.39	12.10
PostKS + BERT	0.77	14.16	22.68	4.27	16.59	4.83	0.39	12.59	20.82	2.73	15.25	4.39
TMemNet + BERT	1.61	15.47	24.12	4.98	17.00	23.86	0.60	13.05	21.74	3.63	15.60	16.33
SKT	1.76	16.04	24.61	5.24	17.61	25.36	1.05	13.74	22.84	4.40	16.05	18.19
DukeNet	**2.43***	**17.09***	**25.17**	**6.81***	**18.52***	**26.38**	**1.68***	**15.06***	**23.34**	**5.29***	**17.06***	**19.57**

i. Wizard of Wikipedia

Methods	Single golden reference (%)						Multiple golden references (%)					
	BLEU-4	METEOR	ROUGE-1	ROUGE-2	ROUGE-L	Hit@1	BLEU-4	METEOR	ROUGE-1	ROUGE-2	ROUGE-L	Hit@1
Seq2Seq	4.84	17.12	26.25	8.41	20.08	–	7.41	21.47	30.68	12.01	24.58	–
Transformer	5.09	16.39	25.96	8.62	19.64	–	7.58	21.01	30.43	12.25	24.60	–
MemNet	5.49	17.70	26.88	9.51	21.15	3.39	7.75	21.60	31.63	12.21	24.94	5.32
PostKS	5.85	18.53	27.52	9.21	21.23	3.60	8.01	22.23	31.57	12.55	25.15	5.95
TMemNet	6.77	20.67	28.25	9.97	22.37	24.15	8.98	25.29	32.46	13.05	26.37	33.95
PostKS + BERT	6.54	19.30	28.94	9.89	22.15	3.95	8.49	23.97	32.85	13.10	26.17	6.40
TMemNet + BERT	8.99	24.48	31.65	13.24	25.90	28.44	12.36	28.61	35.29	16.14	29.51	37.30
SKT	17.81	29.41	35.28	21.74	30.06	28.99	24.69	35.78	41.68	28.30	36.24	39.05
DukeNet	**19.15***	**30.93***	**36.53**	**23.02***	**31.46***	**30.03**	**26.83***	**37.73***	**43.18***	**30.13***	**38.03***	**40.33**

ii. Holl-E dataset

- DukeNet achieves the best results in terms of **KS** and **RG** on both datasets
- The improvement of DukeNet over the strongest baseline SKT on the **Test Unseen** is more obvious

Experiments: Human Evaluation

Methods	Test Seen (%)									Test Unseen (%)								
	Appropriateness			Informativeness			Engagingness			Appropriateness			Informativeness			Engagingness		
	Win	Tie	Lose	Win	Tie	Lose	Win	Tie	Lose	Win	Tie	Lose	Win	Tie	Lose	Win	Tie	Lose
DukeNet vs PostKS + BERT	51	46	3	74	23	3	59	36	5	74	25	1	70	26	4	78	21	1
DukeNet vs TMemNet + BERT	37	55	8	19	79	2	48	42	10	45	50	5	23	74	3	38	57	5
DukeNet vs SKT	15	79	6	15	80	5	16	77	7	20	76	4	15	81	4	21	68	11

- Compare with the three most competitive baselines on Wizard of Wikipedia
  - DukeNet achieves the best performance in terms of all metrics
  - The performance gaps between DukeNet and baselines are more prominent on Test Unseen

Experiments: Analysis

Methods	Test Seen (%)						Test Unseen (%)					
	BLEU-4	METEOR	ROUGE-1	ROUGE-2	ROUGE-L	Hit@1	BLEU-4	METEOR	ROUGE-1	ROUGE-2	ROUGE-L	Hit@1
Full model	**2.43**	**17.09**	**25.17**	**6.81**	**18.52**	**26.38**	**1.68**	**15.06**	**23.34**	**5.29**	**17.06**	**19.57**
w/o DukeL	1.99	16.34	24.54	6.19	17.83	25.57	1.28	13.75	22.20	4.67	15.76	18.56
w/o Knowledge tracking	1.74	15.99	24.34	5.88	17.66	23.82	0.99	13.39	22.69	4.35	15.90	16.25

- Ablation study
  - Removing DukeL, the performance drops a lot in terms of all metrics, almost degenerating to SKT
  - Removing Knowledge tracking, the performance drops dramatically in terms of all metrics, almost degenerating to TMemNet + BERT

Methods	Test Seen (%)		Test Unseen (%)	
	Tracking-Hit@1	Hit@1	Tracking-Hit@1	Hit@1
w/o DukeL	81.66	25.57	76.04	18.56
Full model	82.77	26.38	76.82	19.57

Tracking-Hit@1 (top 1 accuracy for knowledge tracking)

- Further analyze DukeL
  - DukeL improves the performences of knowledge tracking and shifting simultaneously

Experiments: Case Study

	Example (Test seen)
Context	$X_{\tau-1}$: pizza delivery $Y_{\tau-1}$: for dinner i had pizza delievered to my house by a pizzeria X_{τ}: love a good pizzaria or restaurant that specializes in pissa
Knowledge pool	K_1: pizza delivery is a service in which a pizzeria or pizza chain delivers a pizza to a customer . K_2: pizzas may be delivered in pizza boxes or delivery bags , and deliveries are made with either an automobile . K_3: an order is typically made either by telephone or over the internet to the pizza chain, in which the customer can request pizza type , motorized scooter . ...
Tracking to shifting	PostKS + BERT: $\varnothing \rightarrow K_1$ ✗ TMemeNet + BERT: $\varnothing \rightarrow K_1$ ✗ SKT: K_2 ✗ $\rightarrow K_2$ ✗ DukeNet: K_1 ✓ $\rightarrow K_3$ ✓
Response	PostKS + BERT: i love pizza delivery . i love pizza delivery . TMemeNet + BERT: i can do it with my pizza delivery , but i can pay my pizza at home . SKT: i love frozen pizza ! i love the delivery bagels with either an automobile , or a motorized scooter . DukeNet: i love pizza delivery and i could pay online or online ordering .

- DukeNet captures the knowledge interaction from the defnition of “Pizza delivery” to “Pizza online ordering”
- Baselines all select less suitable knowledge and generate less attractive responses

Outline

- Background and motivation
- Model
- Experiments
- **Conclusion**

Conclusion and Future Work

- Conclusion
    - We propose DukeNet regarding knowledge tracking and knowledge shifting as dual tasks to improve KS
    - We devise DukeL to explore knowledge that is not annotated in the training set during training for KS
    - DukeNet enhanced by DukeL can select more appropriate knowledge and hence generate more informative and engaging responses.

- Future work
    - Consider dual knowledge interactions between **long-term turns** rather than **adjacent turns**

Thanks!
Q & A

Name: Chuan Meng

Email: mengchuan@sdu.edu.cn

Code: https://github.com/ChuanMeng/