Initiative-Aware Self-Supervised Learning for Knowledge-Grounded Conversations

Chuan Meng, Pengjie Ren, Zhumin Chen,
Zhaochun Ren, Tengxiao Xi, Maarten de Rijke

SIGIR 2021

Introduction

In Knowledge-Grounded Conversations (KGC) tasks, external knowledge is used to generate more informative responses.

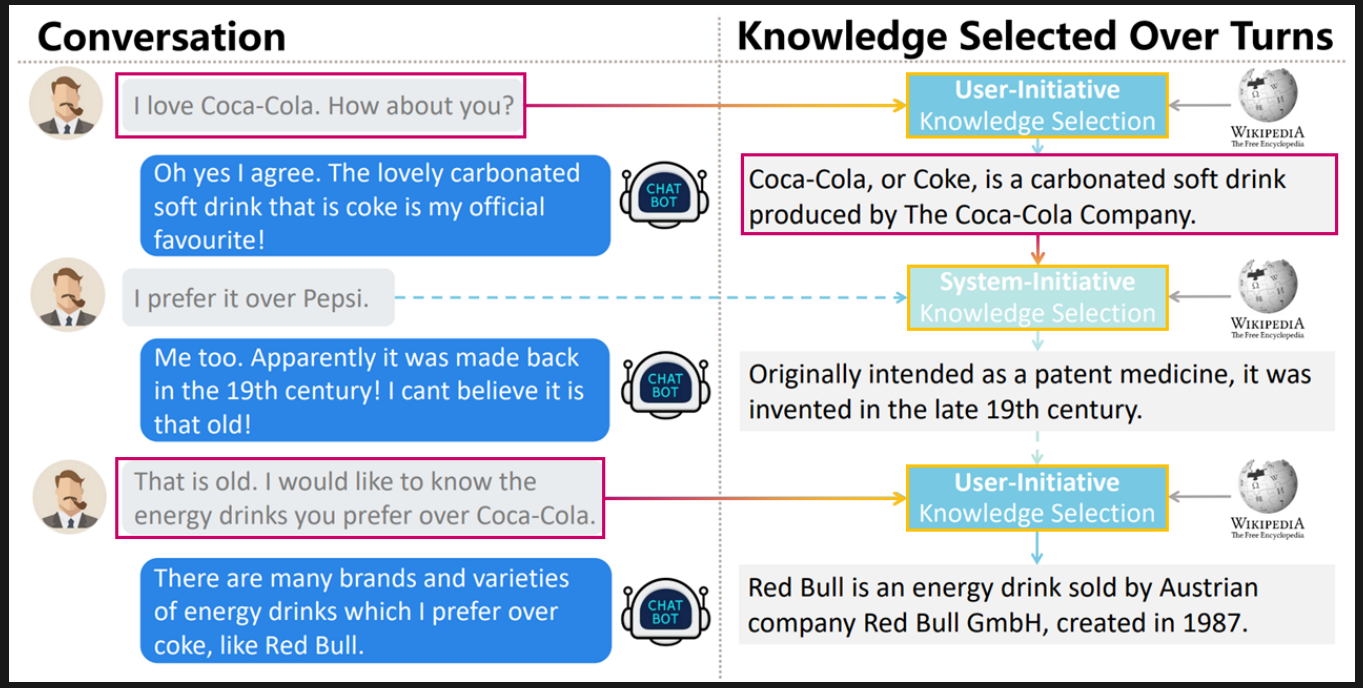
The method of selecting the required knowledge is an important part of KGC.

The conversation may be driven by the user or the system.

Introduction(cont.)

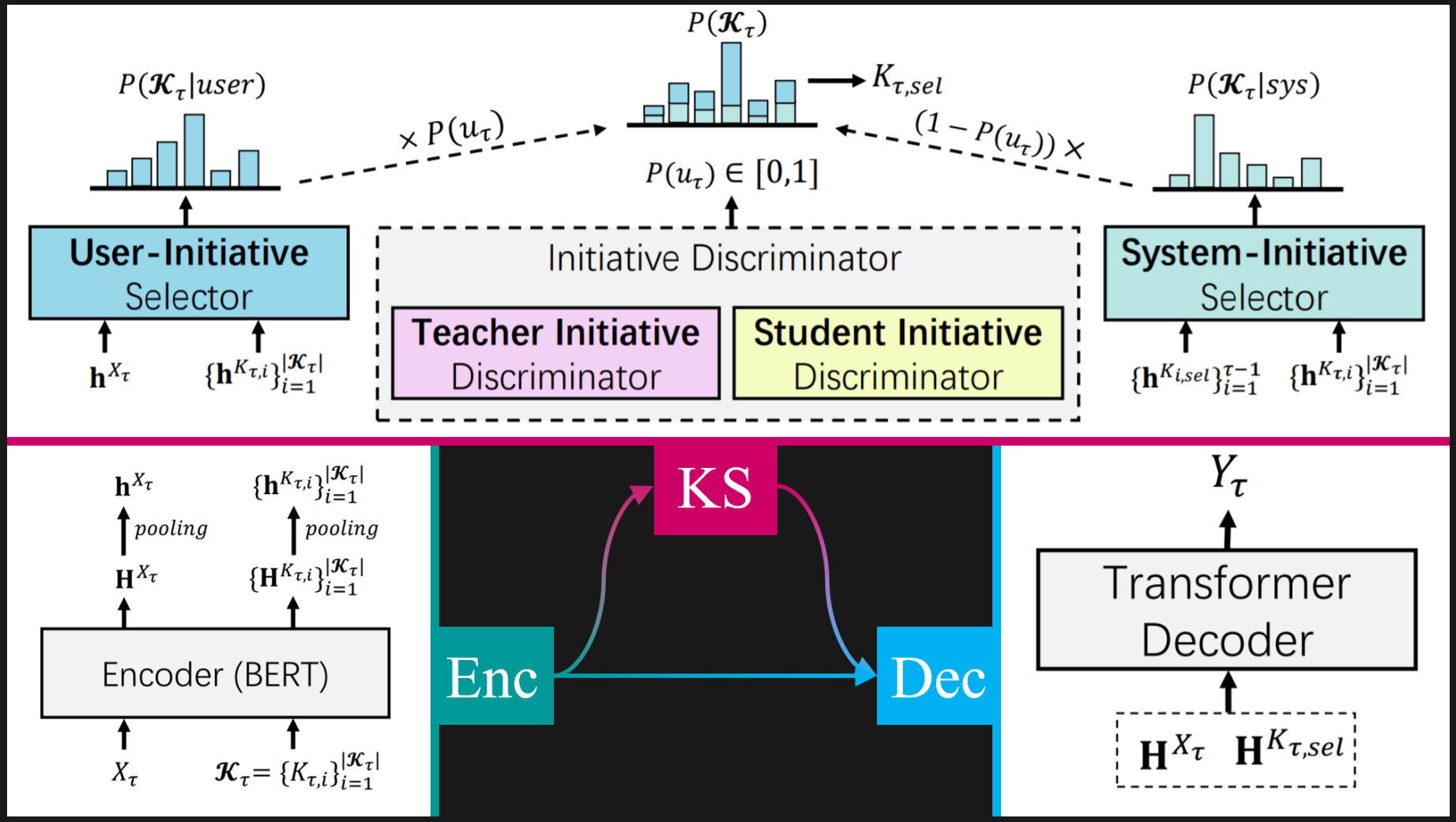
- User's Initiative
  Based on **New Topics** or **Questions** posed by the user
- System's Initiative
  Based on **The Previously Selected Knowledge**

Introduction(cont.)



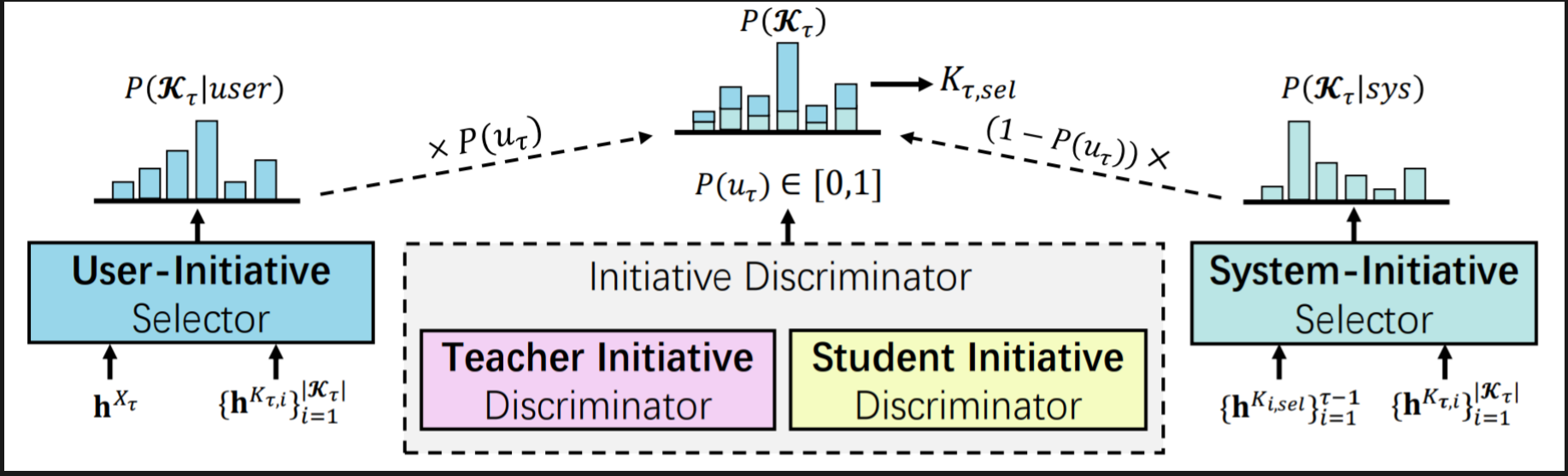
Explicitly distinguishes between user-initiative and system-initiative knowledge selection.

Architecture

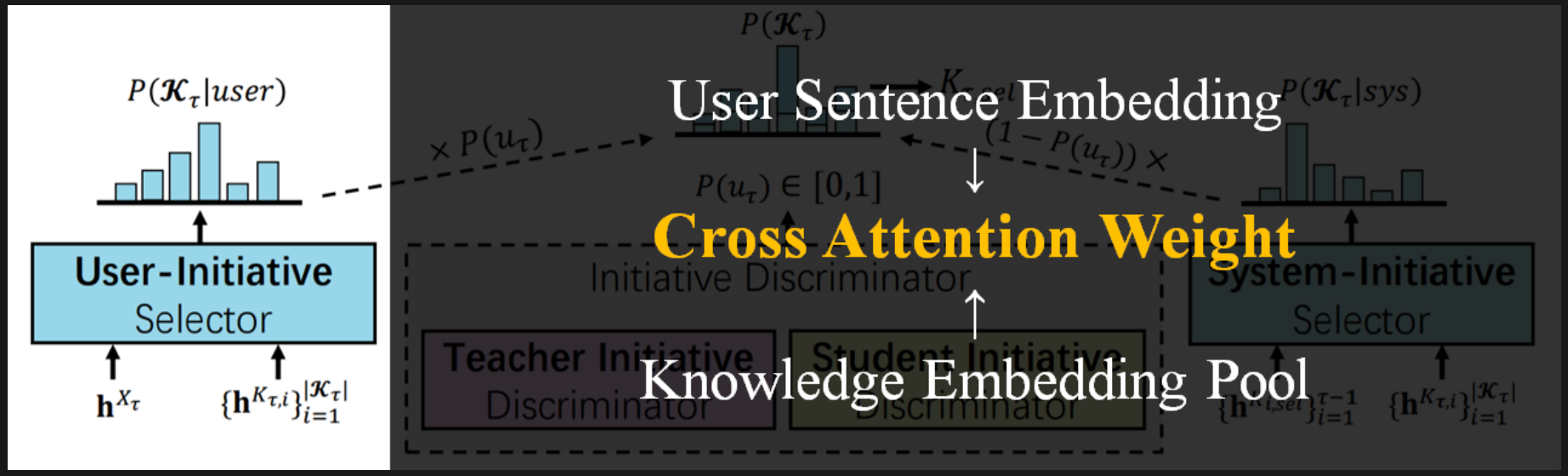


MIKe

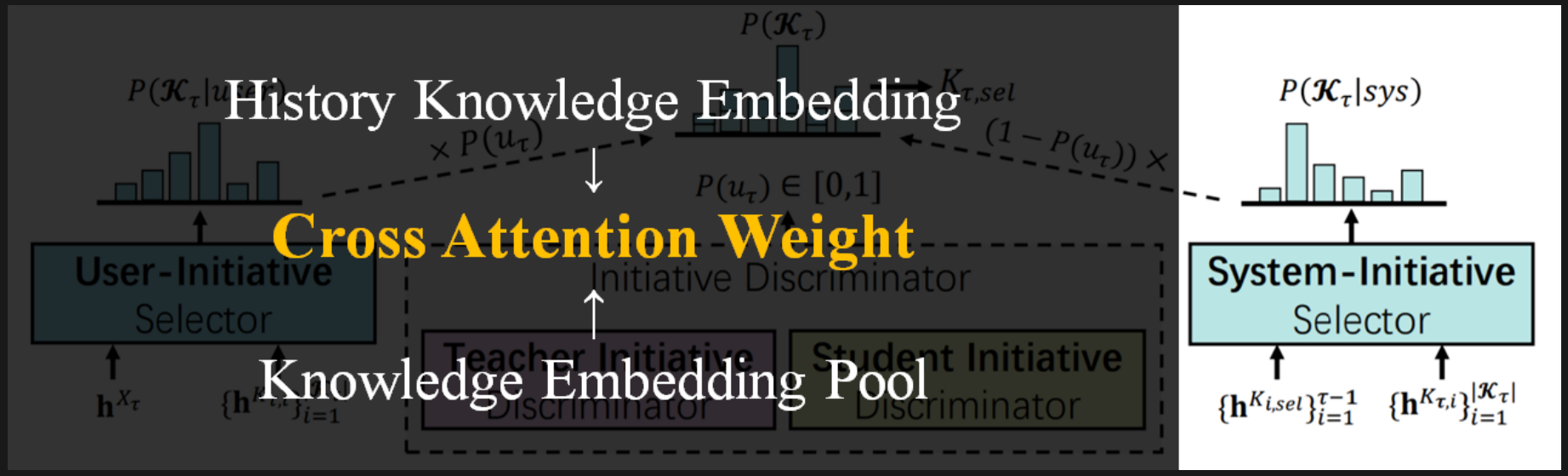
***M**ixed-**I**nitiative **K**nowledge Selection Method*



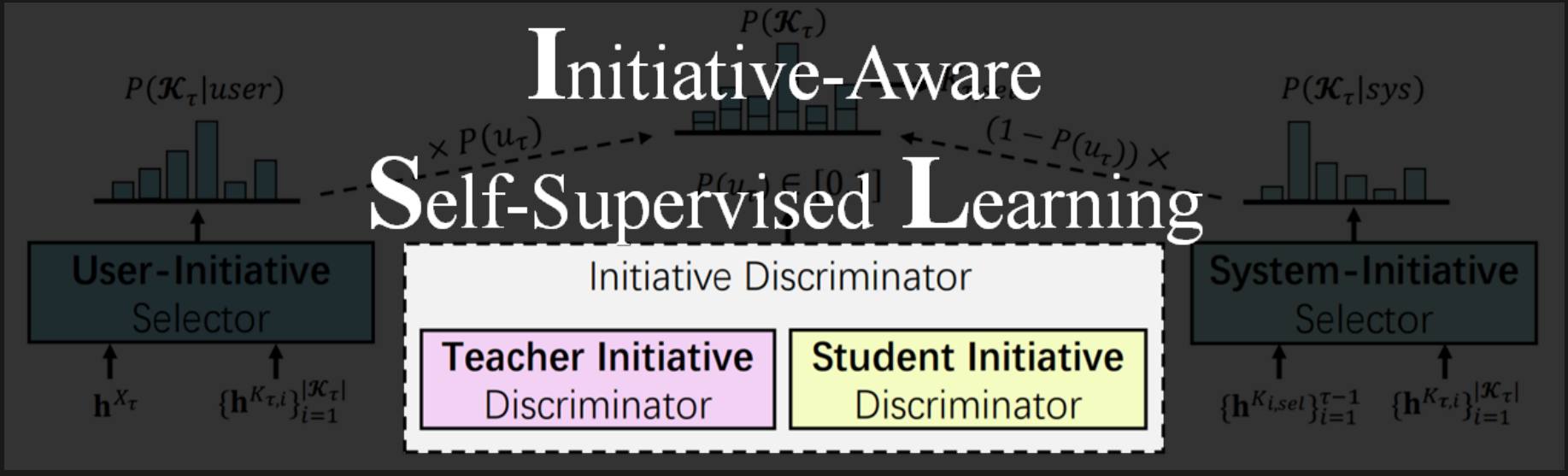
MIKe(cont.)


$P(\mathcal{K}_\tau|user)$
User-Initiative Selector
$\mathbf{h}^{X_\tau}$
$\{\mathbf{h}^{K_{\tau,i}}\}_{i=1}^{|\mathcal{K}_\tau|}$
$P(\mathcal{K}_\tau)$
$\times P(u_\tau)$
$P(u_\tau) \in [0,1]$
$(1-P(u_\tau)) \times$
$P(\mathcal{K}_\tau|sys)$
User Sentence Embedding
Cross Attention Weight
Knowledge Embedding Pool
Initiative Discriminator
Teacher Initiative Discriminator
Student Initiative Discriminator
System-Initiative Selector
$\{\mathbf{h}^{K_{\tau,i}}\}_{i=1}^{|\mathcal{K}_\tau|}$

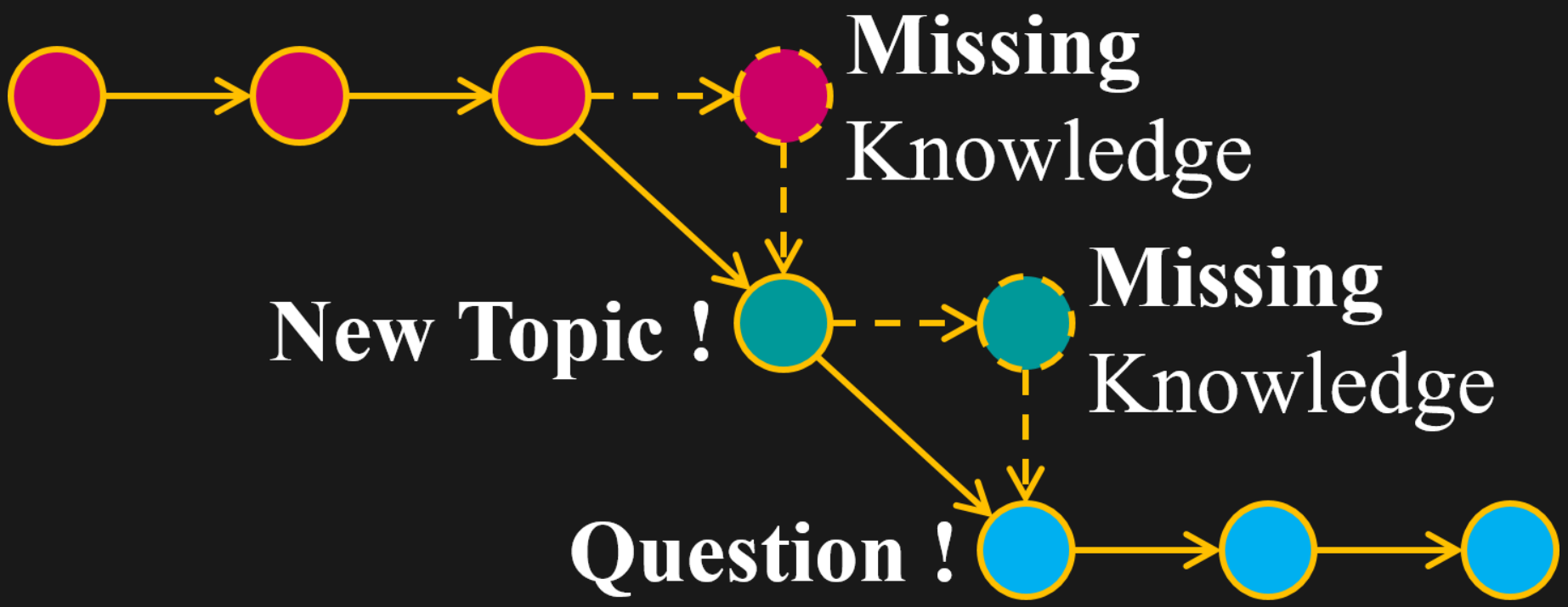
MIKe(cont.)


History Knowledge Embedding
Cross Attention Weight
Knowledge Embedding Pool
$P(\mathcal{K}_\tau)$
$P(\mathcal{K}_\tau|sys)$
$K_{\tau,sel}$
$\times P(u_\tau)$
$(1-P(u_\tau))\times$
$P(u_\tau) \in [0,1]$
User-Initiative Selector
System-Initiative Selector
Initiative Discriminator
Teacher Initiative Discriminator
Student Initiative Discriminator
$\mathbf{h}^{X_\tau}$
$\{\mathbf{h}^{K_{i,sel}}\}_{i=1}^{\tau-1}$
$\{\mathbf{h}^{K_{\tau,i}}\}_{i=1}^{|\mathcal{K}_\tau|}$

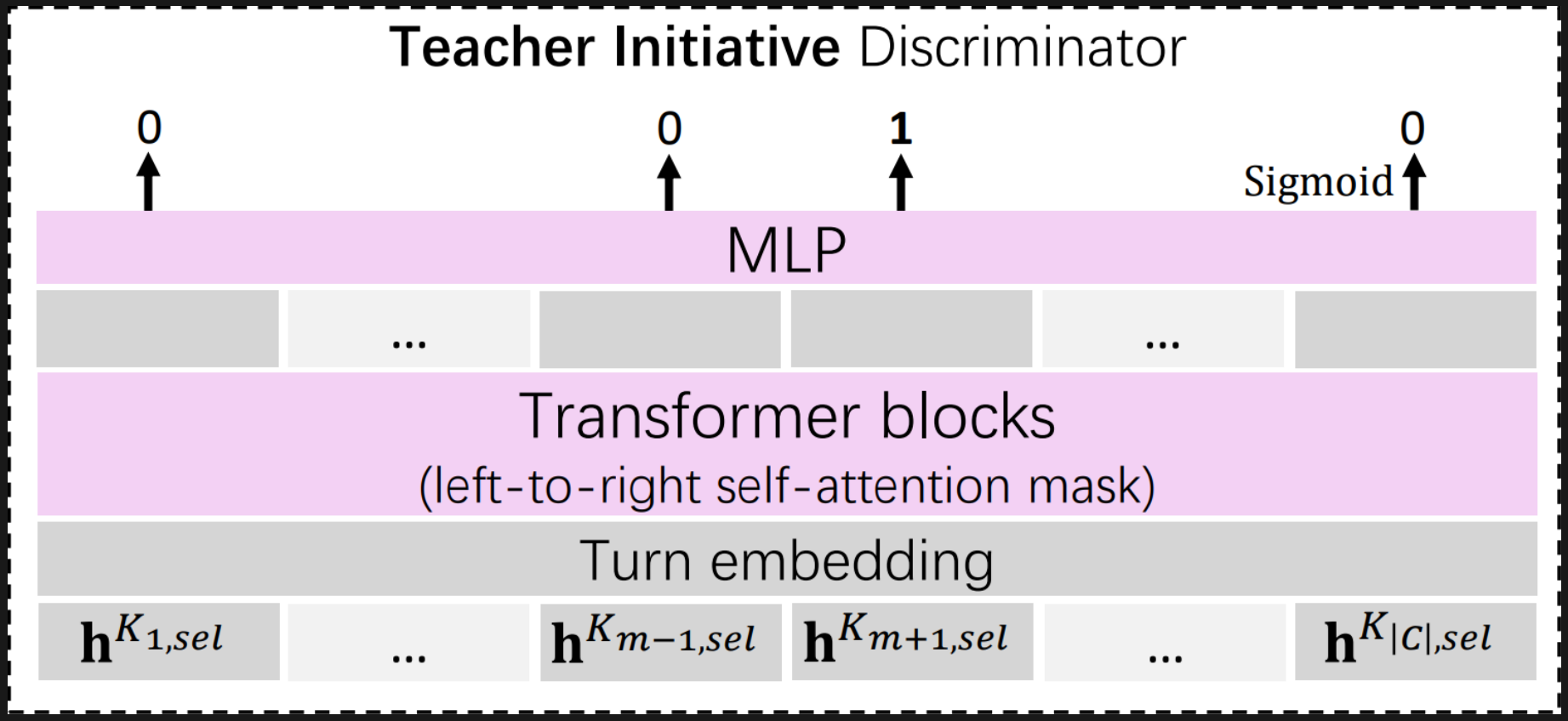
ISLe


$P(\mathcal{K}_\tau)$
Initiative-Aware
$P(\mathcal{K}_\tau|user)$
$\times P(u_\tau)$
$(1-P(u_\tau))\times$
$P(\mathcal{K}_\tau|sys)$
Self-Supervised Learning
User-Initiative
Selector
Initiative Discriminator
Teacher Initiative
Discriminator
Student Initiative
Discriminator
System-Initiative
Selector
$\mathbf{h}^{X_\tau}$
$\{\mathbf{h}^{K_{\tau,i}}\}_{i=1}^{|\mathcal{K}_\tau|}$
$\{\mathbf{h}^{K_{i,sel}}\}_{i=1}^{\tau-1}$
$\{\mathbf{h}^{K_{\tau,i}}\}_{i=1}^{|\mathcal{K}_\tau|}$

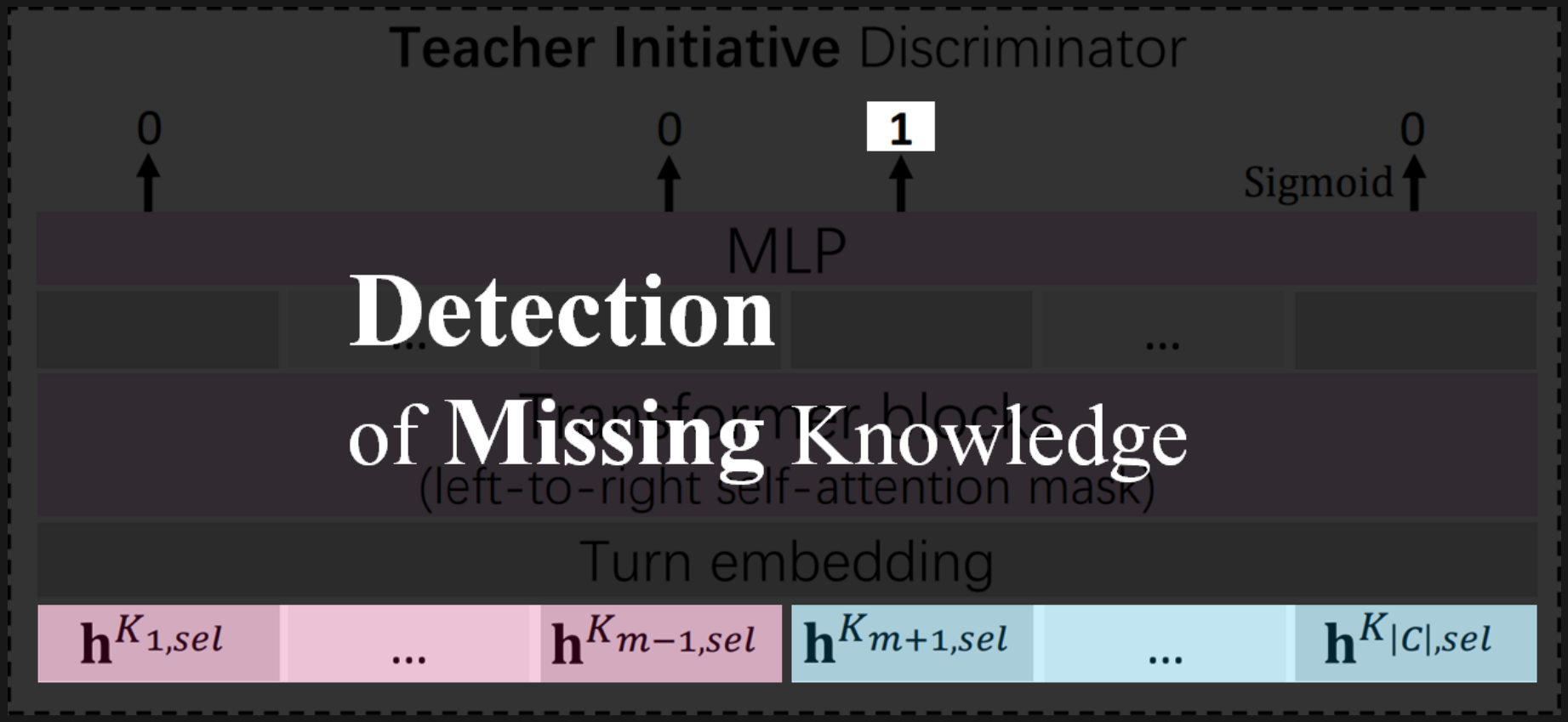
ISLe(cont.)



ISLe(cont.)

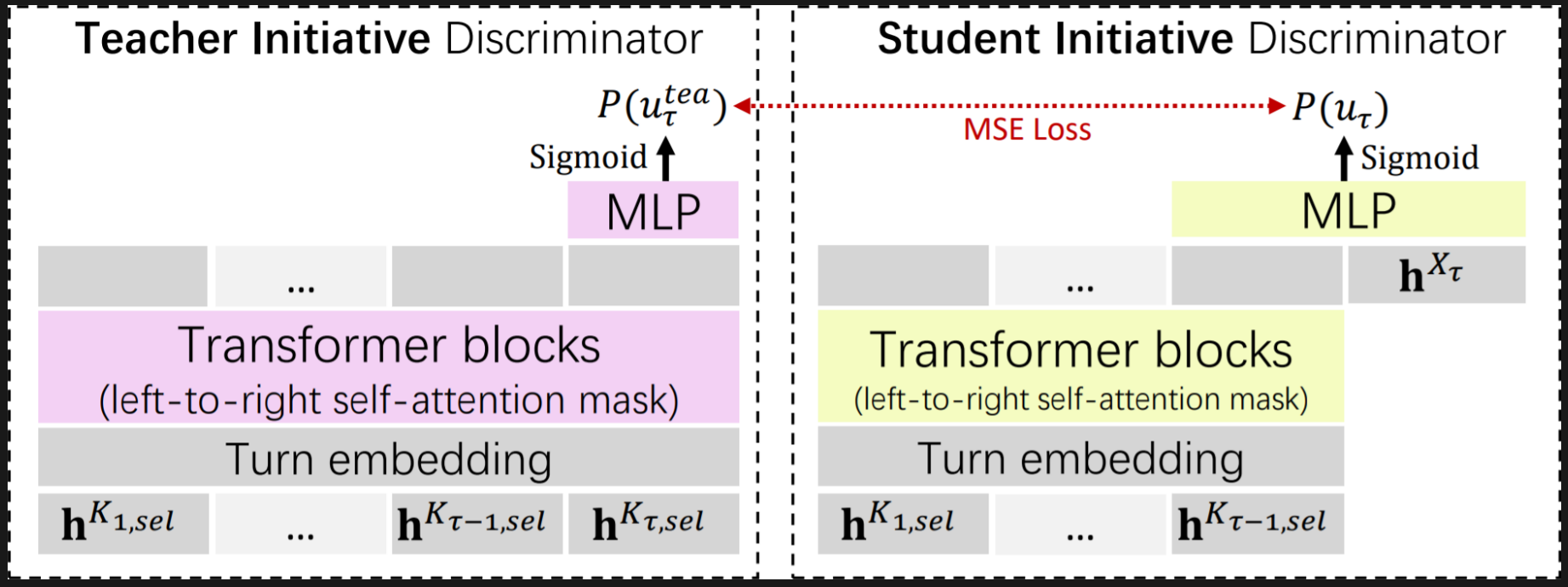

Teacher Initiative Discriminator
0
0
1
0
Sigmoid
MLP
...
...
Transformer blocks
(left-to-right self-attention mask)
Turn embedding
$\mathbf{h}^{K_{1,sel}}$
...
$\mathbf{h}^{K_{m-1,sel}}$
$\mathbf{h}^{K_{m+1,sel}}$
...
$\mathbf{h}^{K_{|C|,sel}}$

ISLe(cont.)

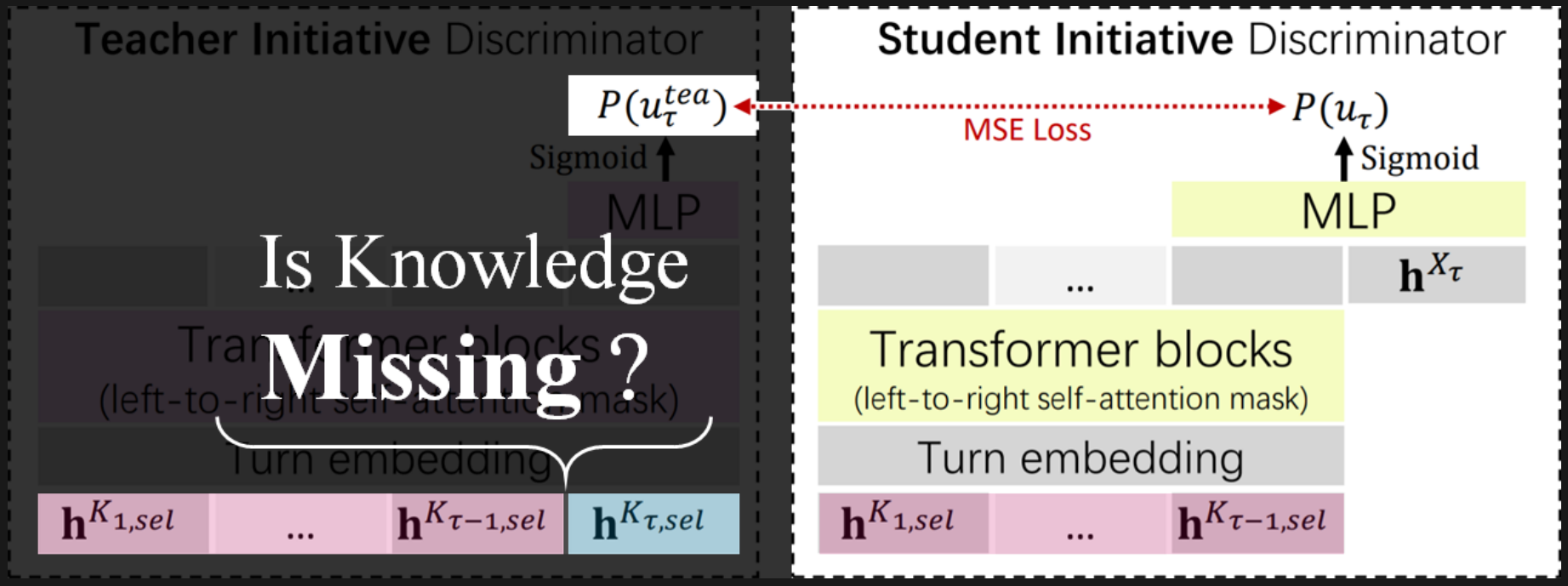


Detection
of **Missing** Knowledge

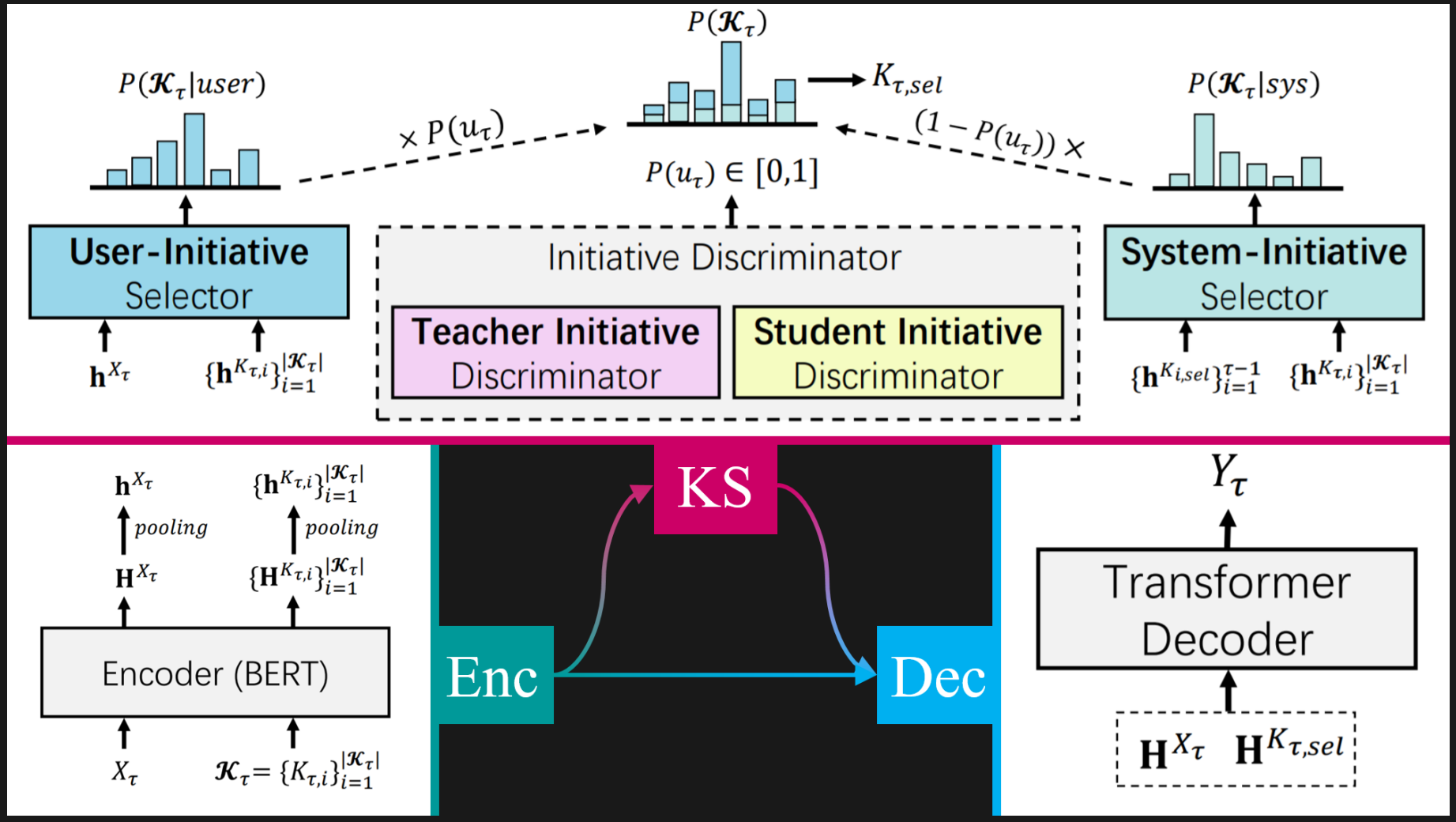
ISLe(cont.)


Teacher Initiative Discriminator
Student Initiative Discriminator
$P(u_\tau^{tea})$
MSE Loss
$P(u_\tau)$
Sigmoid
MLP
Sigmoid
MLP
...
$\mathbf{h}^{X_\tau}$
Transformer blocks
(left-to-right self-attention mask)
Turn embedding
$\mathbf{h}^{K_{1,sel}}$
...
$\mathbf{h}^{K_{\tau-1,sel}}$
$\mathbf{h}^{K_{\tau,sel}}$
Transformer blocks
(left-to-right self-attention mask)
Turn embedding
$\mathbf{h}^{K_{1,sel}}$
...
$\mathbf{h}^{K_{\tau-1,sel}}$

ISLe(cont.)


Teacher Initiative Discriminator
$P(u_\tau^{tea})$
Sigmoid
MLP
...
Transformer blocks
(left-to-right self-attention mask)
Turn embedding
$\mathbf{h}^{K_{1,sel}}$
...
$\mathbf{h}^{K_{\tau-1,sel}}$
$\mathbf{h}^{K_{\tau,sel}}$
Is Knowledge Missing ?
Student Initiative Discriminator
MSE Loss
$P(u_\tau)$
Sigmoid
MLP
...
$\mathbf{h}^{X_\tau}$
Transformer blocks
(left-to-right self-attention mask)
Turn embedding
$\mathbf{h}^{K_{1,sel}}$
...
$\mathbf{h}^{K_{\tau-1,sel}}$

Architecture(review)



Evaluation
on the WoW dataset

Methods	Test Seen (%)					
	BLEU-4	METEOR	ROUGE-1	ROUGE-2	ROUGE-L	R@1
PostKS + BERT	0.77	14.16	22.68	4.27	16.59	4.83
TMemNet + BERT	1.61	15.47	24.12	4.98	17.00	23.86
SKT	1.76	16.04	24.61	5.24	17.61	25.36
DiffKS + BERT	2.22	16.82	24.75	6.27	17.90	25.62
DukeNet	2.43	17.09	25.17	6.81	18.52	26.38
SKT+PIPM+KDBTS	2.47	17.14	25.19	7.01	18.47	27.40
MIKe (ours)	**2.78**	**17.76**	**25.40**	**7.11**	**18.78**	**28.41**
MIKe-ISLe	2.63	17.22	25.15	6.97	18.67	27.52
MIKe-ISLe-ID	2.48	17.28	24.90	6.64	18.24	26.58
MIKe-ISLe-ID-UIS	1.70	15.88	24.37	5.17	17.33	23.95
MIKe-ISLe-ID-SIS	1.68	15.76	24.33	5.08	17.21	23.88

Evaluation(cont.)
on the WoW dataset

Methods	Test Unseen (%)					
	BLEU-4	METEOR	ROUGE-1	ROUGE-2	ROUGE-L	R@1
PostKS + BERT	0.39	12.59	20.82	2.73	15.25	4.39
TMemNet + BERT	0.60	13.05	21.74	3.63	15.60	16.33
SKT	1.05	13.74	22.84	4.40	16.05	18.19
DiffKS + BERT	1.69	14.69	23.62	5.05	16.82	20.11
DukeNet	1.68	15.06	23.34	5.29	17.06	19.57
SKT+PIPM+KDBTS	1.71	14.83	23.56	5.46	17.14	20.20
MIKe (ours)	**2.00**	**15.64**	**23.78**	**5.61**	**17.41**	**21.47**
MIKe-ISLe	1.67	15.38	23.42	5.28	17.04	20.44
MIKe-ISLe-ID	1.46	14.70	22.87	5.16	16.36	19.35
MIKe-ISLe-ID-UIS	0.89	13.68	22.17	4.09	15.98	16.67
MIKe-ISLe-ID-SIS	0.87	13.44	22.01	3.88	15.79	15.99

Initiative Discrimination Evaluation

Methods	Test Seen (%)			Test Unseen (%)		
	M-F1	U-F1	S-F1	M-F1	U-F1	S-F1
MIKe	**62.87**	**61.79**	**63.95**	**61.79**	**61.10**	**62.48**
MIKe-ISLe	51.04	60.59	41.49	47.29	60.89	33.69
Heuristic	51.74	48.16	55.31	52.69	49.52	55.86

Initiative Discrimination Evaluation(cont.)

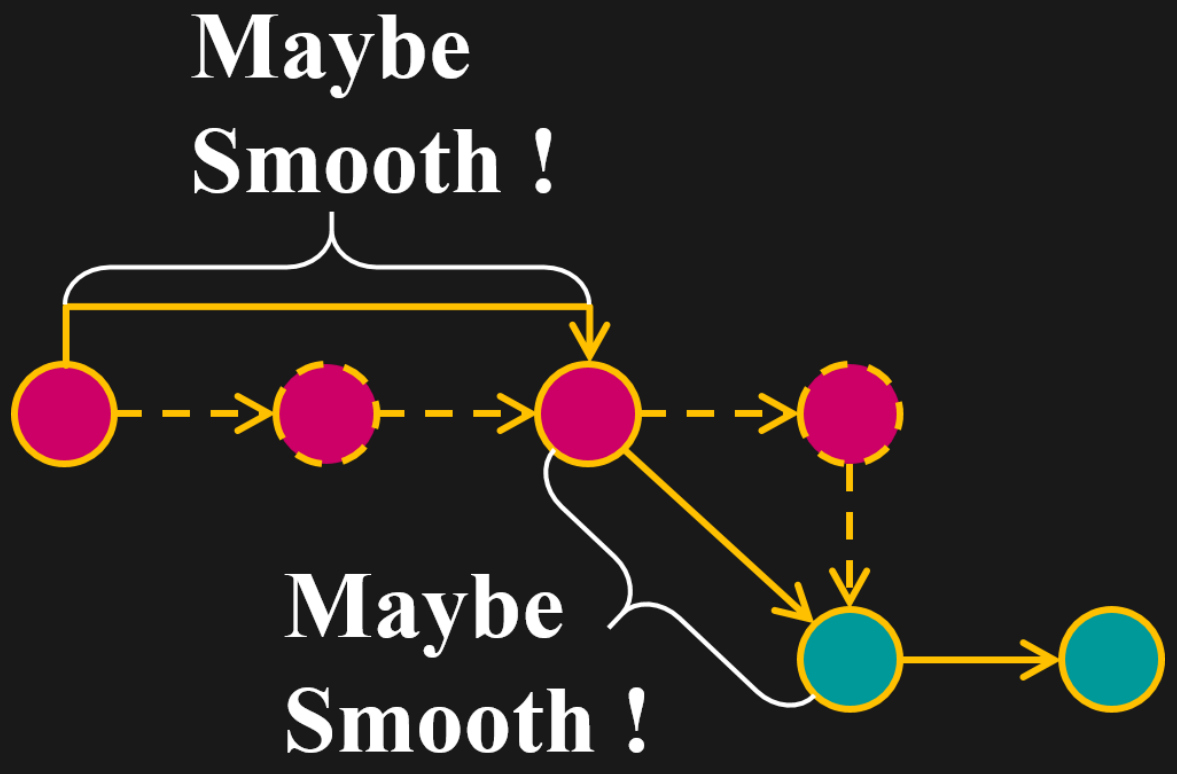
Heuristic Method

- **User Initiative**
  The current user utterance contains a question mark or begins with a question word.
- **System Initiative**
  Other

Conclusion

- With the help of ISLe, the accuracy of KS has been significantly improved.
- There is still a large room for improvement in distinguishing initiative types.

Conclusion(cont.)



Unable to determine switching point.

Thanks for your attention.