DIALKI: Knowledge Identification in Conversational Systems through Dialogue-Document Contextualization

Zeqiu Wu ♠* Bo-Ru Lu ♠* Hannaneh Hajishirzi ♠♦ Mari Ostendorf ♠
♠University of Washington ♦Allen Institute for AI
{zeqiuwu1,roylu,hannaneh,ostendor}@washington.edu

Code : https: //github.com/ellenmellon/DIALKI

——EMNLP 2021

Reported by Yabo Yin

Artificial

1.Introduction

2.Method

3.Experiments

gesis
Leibniz-Institut
für Sozialwissenschaften

Introduction

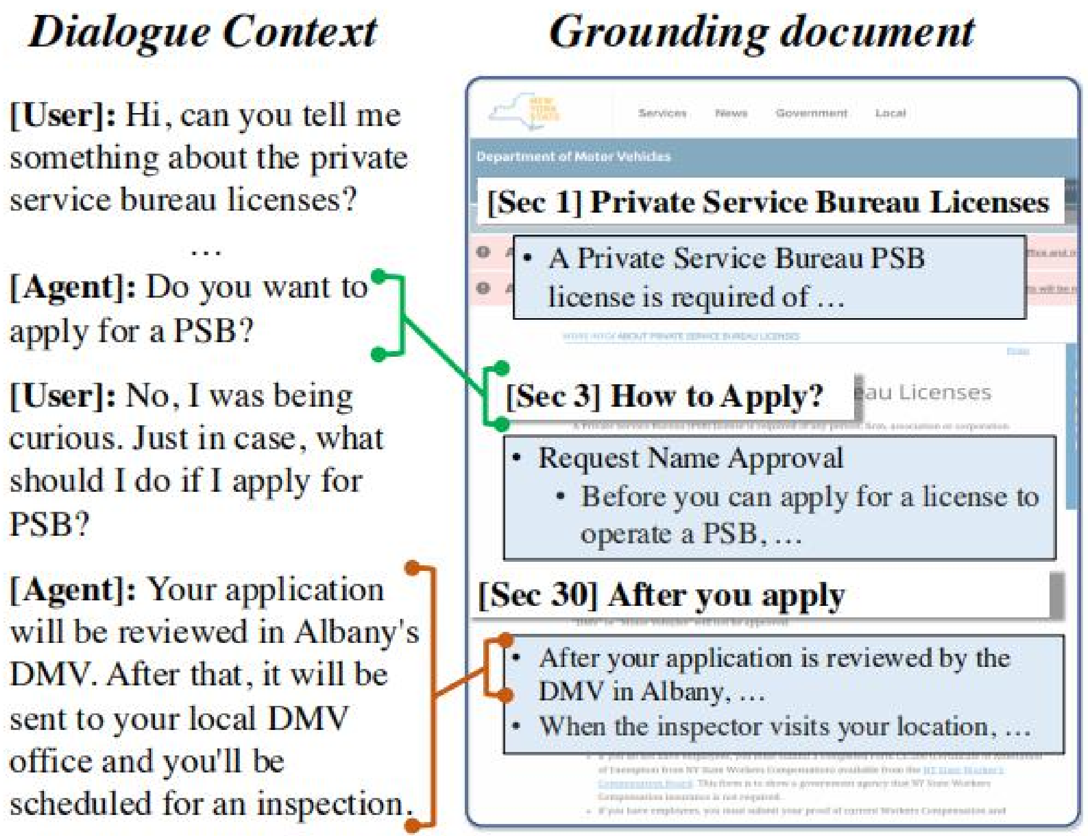


Figure 1: In a document-grounded conversation, *knowledge identification* targets to locate a knowledge string within a long document to assist the agent in addressing the current user query.

Method

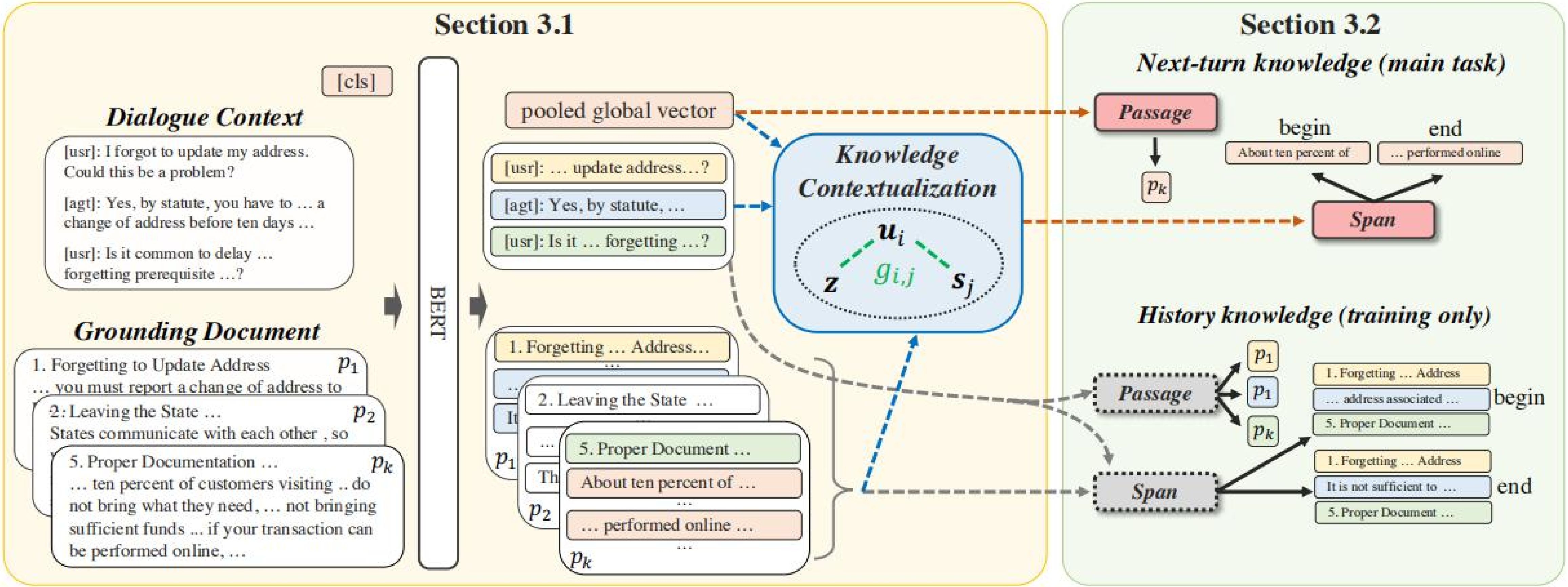


Figure 2: The overview of DIALKI. Each document is divided into passages. We apply BERT and a knowledge contextualization mechanism to obtain dialogue context and knowledge representations (left), for performing both next (main) and history (auxiliary) turn knowledge identification tasks (right). For each turn, DIALKI identifies knowledge by selecting the relevant passage as well as the begin/end spans in the passage.

Method

Problem Definition

dialogue context $(u_1, u_2, \ldots, u_n)$

grounding document $\mathcal{D} = \{p_1, p_2, \ldots, p_{|\mathcal{D}|}\}$

Each passage p consists of a sequence of semantic units $p = (s_1, s_2, \ldots, s_l)$

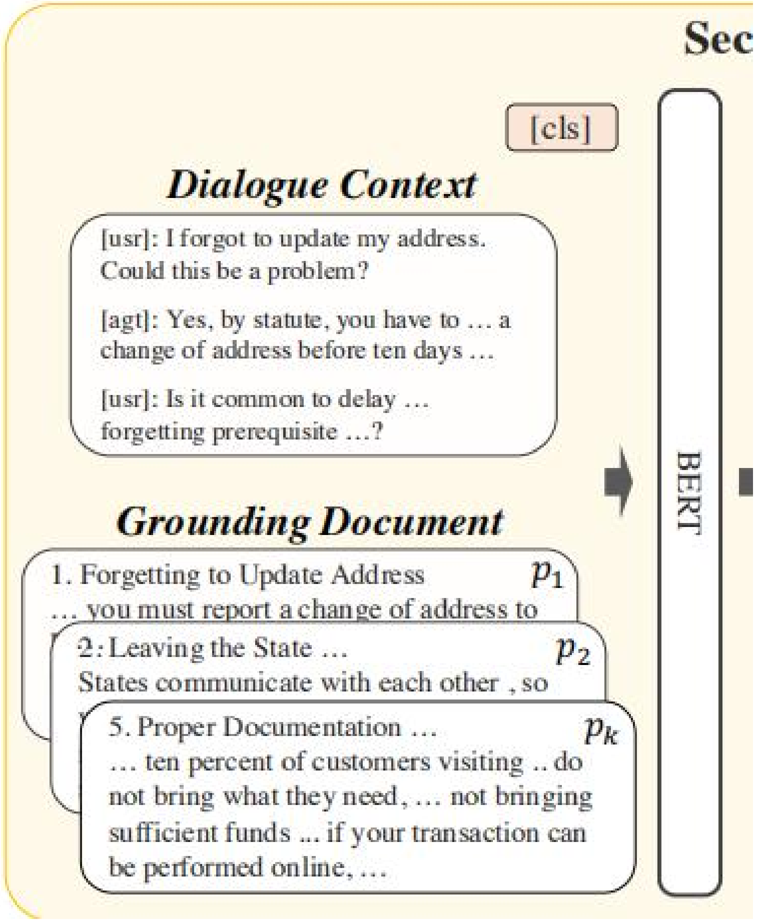
Multi-Passage Encoding

$$\begin{aligned}\mathbf{X} \quad = \quad & [\text{cls}][\text{usr}]\, u_1 [\text{agt}] u_2 \cdots [\text{usr}]\, u_n \\ & [\text{sep}]\, t\, [\text{cls}]\, s_1\, [\text{cls}]\, s_2 \cdots [\text{cls}]\, s_l\, [\text{sep}]\end{aligned}$$

$$\mathbf{H} = G(\text{BERT}(\mathbf{X}))$$

where $G(.)$ gathers vectors of all '[cls]', '[usr]' and '[agt]' tokens.

$\mathbf{H}$ as $[\mathbf{z}, \mathbf{u}_1, \ldots, \mathbf{u}_n, \mathbf{s}_1, \ldots, \mathbf{s}_l]$

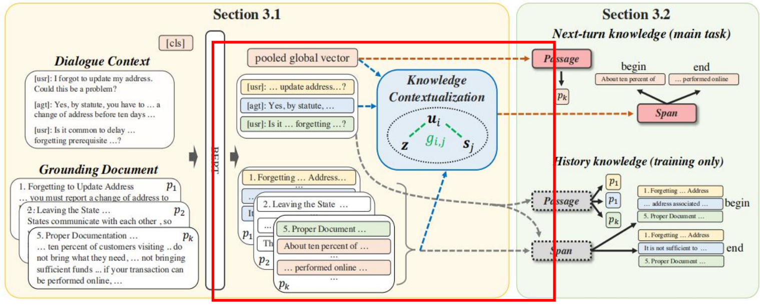


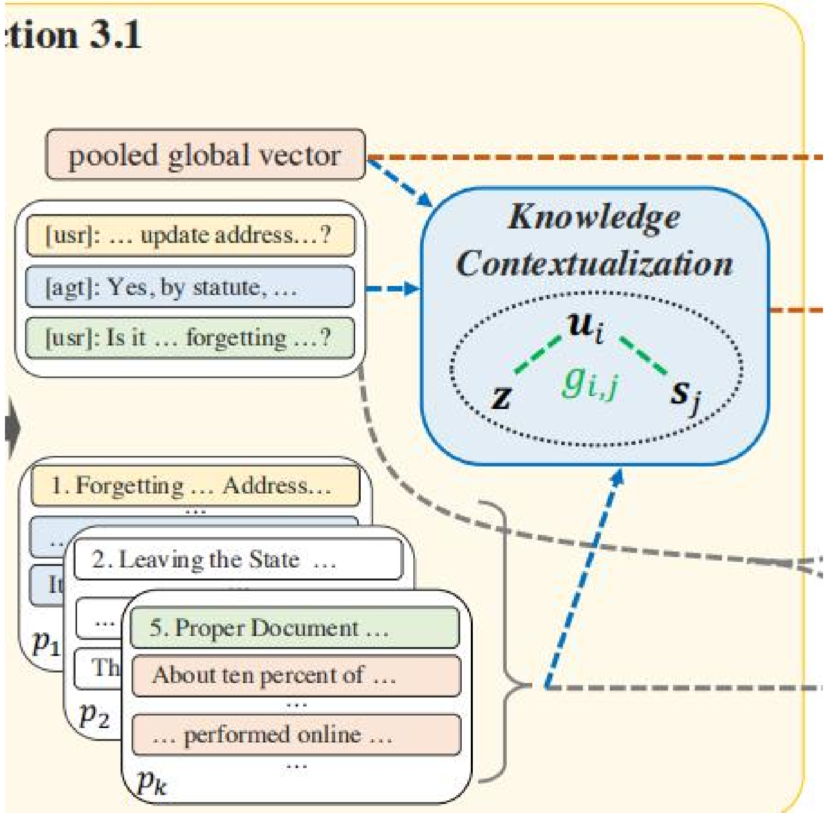
Method

$\mathbf{z}$, pooled global

$\mathbf{u}_i$ dialogue utterance u_i

$\mathbf{s}_j$ span s_j





$$\mathbf{a}_{i,j} = \mathbf{W}_s \mathbf{s}_j + \mathbf{W}_z \mathbf{z} + \mathbf{W}_u \mathbf{u}_i \,, \; i \in \mathbf{C}_u$$

$$g_{i,j} = \sigma\left(\mathbf{u}_i{}^{\top} \mathbf{z} + \mathbf{u}_i{}^{\top} \mathbf{s}_j\right),$$

$$\widehat{\mathbf{s}}_j = \upsilon\Big(\sum_{i \in \mathbf{C}_u} \left[\phi(\mathbf{a}_{i,j}) \odot g_{i,j}\right] + \mathbf{s}_j \Big)$$

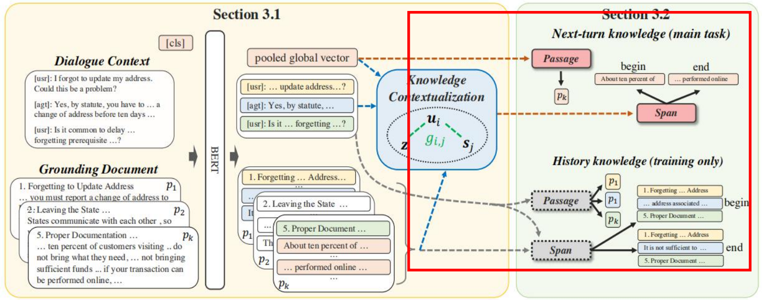
where $\mathbf{W}_s, \mathbf{W}_z, \mathbf{W}_u \in \mathbb{R}^{d \times d}$

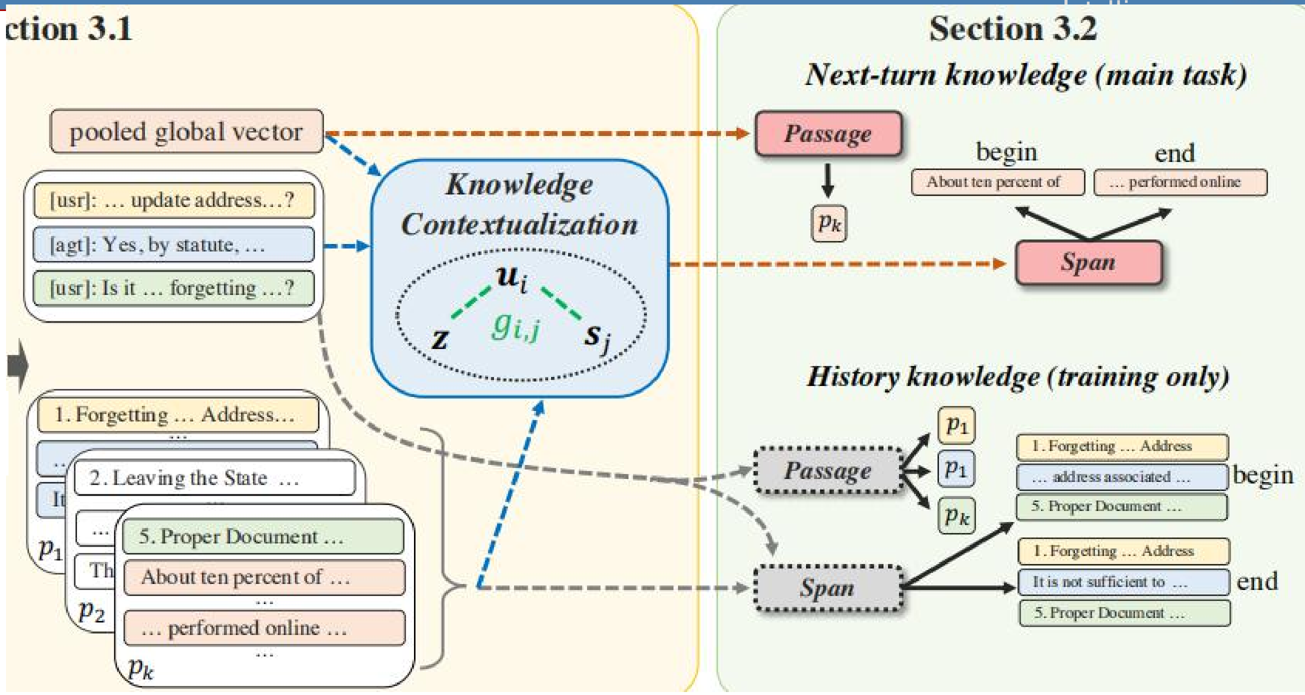
$\mathbf{C}_u$ indexes the most recent user turns.

$\widetilde{\mathbf{s}}_j$ with previous *agent* turns.

$$\dot{\mathbf{s}}_j = [\mathbf{s}_j, \widehat{\mathbf{s}}_j, \widetilde{\mathbf{s}}_j]$$

Method





$\mathbf{Z}$ matrix containing the pooled global vectors for all

$\mathbf{U}_i$ utterance representations for u_i in all passages

$\dot{\mathbf{S}}$ all span representations

$$\mathcal{L}_{\text{psg}} = -\log q(\mathbf{W}_p \mathbf{Z})_{\hat{k}} \quad (1)$$

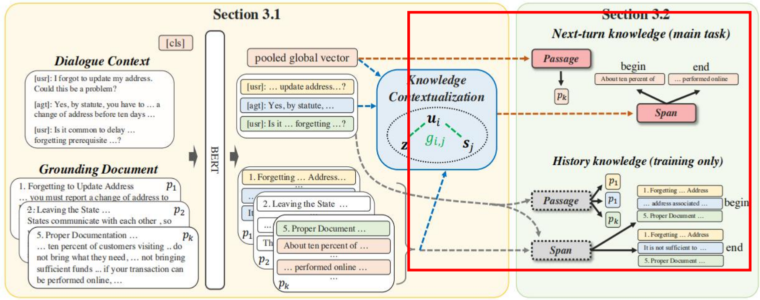
$$\mathcal{L}_{\text{begin}} = -\log q(\mathbf{W}_b \dot{\mathbf{S}})_{\hat{b}} \quad (2)$$

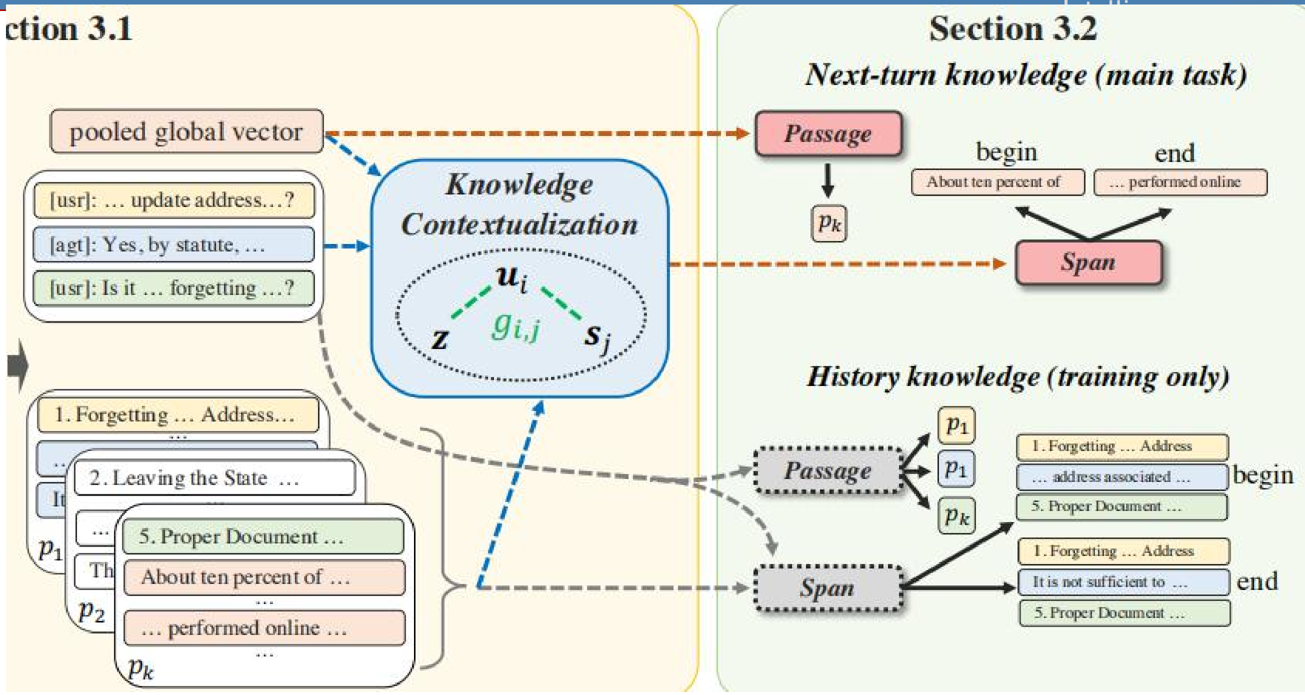
$$\mathcal{L}_{\text{end}} = -\log q(\mathbf{W}_e \dot{\mathbf{S}})_{\hat{e}} \quad (3)$$

where $\mathbf{W}_p, \mathbf{W}_b, \mathbf{W}_e \in \mathbb{R}^d$.

$$\mathcal{L}_{\text{next}} = \mathcal{L}_{\text{psg}} + \mathcal{L}_{\text{begin}} + \mathcal{L}_{\text{end}}.$$

Method





$\mathbf{U}_i$ utterance representations for u_i in all passages

$$\mathcal{L}^h_{\text{psg}} = \frac{1}{\|\mathbf{U}^*\|} \sum_{u_i \in \mathbf{U}^*} -\log q\left(\mathbf{W}^h_p\ \phi(\mathbf{W}^h\ \mathbf{U}_i)\right)_{k_i}$$

where $\mathbf{W}^h \in \mathbb{R}^{d\times d}, \mathbf{W}^h_p \in \mathbb{R}^d$

$\mathbf{U}^*$ is the set of history turns that can find their knowldge strings in the document $\mathcal{D}$. k_i is the gold passage index for turn u_i. ϕ is a non-linear

$$\mathcal{L}_{\text{adv}} = \max_{\|\epsilon\| \le a} \sum_{f \in \{f_{psg}, f_{begin}, f_{end}\}} \text{Div}\left(f(x) \| f(x+\epsilon)\right)$$

$$\mathcal{L}_{\text{hist}} = \mathcal{L}^h_{\text{psg}} + \mathcal{L}^h_{\text{begin}} + \mathcal{L}^h_{\text{end}}.$$

$$\mathcal{L} = \mathcal{L}_{\text{next}} + \alpha \mathcal{L}_{\text{hist}} + \beta \mathcal{L}_{\text{adv}} \qquad (4)$$

Experiments

Method	Overall EM	Overall F1
BERTQA-Token	34.6	53.2
BERTQA-Token (our version)	35.8	52.6
DIALKI ($\mathcal{L}_{\text{next}}$ only)	51.2	64.7
DIALKI	59.5	71.0
DIALKI (BERT-large)	**61.8**	**73.1**

Table 1: Evaluation results on the Doc2Dial test set.

Method	Seen EM	Seen F1	Unseen EM	Unseen F1
Transformer MemNet	22.5	33.2	12.2	19.8
Transformer MemNet + Pretrain	24.5	36.4	23.7	35.8
DiffKS (RNN)	25.5	–	19.7	–
SLKS (RNN)	23.4	–	14.7	–
SLKS (BERT-base)	26.8	–	18.3	–
Multi-Sentence (BERT-base)	30.4	37.7	27.6	35.4
DIALKI (BERT-base)	**32.9**	**40.7**	**35.5**	**43.4**

Table 2: Evaluation results of WoW test sets.

Experiments

Method	Doc2Dial Overall EM	Doc2Dial Overall F1	Doc2Dial Seen EM	Doc2Dial Seen F1	Doc2Dial Unseen EM	Doc2Dial Unseen F1	WoW Overall EM	WoW Overall F1	WoW Seen EM	WoW Seen F1	WoW Unseen EM	WoW Unseen F1
BERTQA-Token	42.2	58.1	48.3	61.1	37.0	55.6	–	–	–	–	–	–
BERTQA-Span	46.3	59.3	54.4	63.5	39.4	55.6	–	–	–	–	–	–
Multi-Sentence	59.5	68.8	63.6	71.6	56.0	66.4	29.2	37.0	32.4	39.7	26.1	34.3
DIALKI ($\mathcal{L}_{\text{next}}$ only)	60.4	71.2	64.2	72.3	57.1	70.2	31.5	39.7	33.3	41.1	29.8	38.3
$+\mathcal{L}_{\text{hist}}$	63.0	72.6	66.5	73.9	59.9	71.9	33.6	41.6	35.1	42.7	32.2	40.5
$+\mathcal{L}_{\text{hist}}$, *know-ctx*	63.8	73.4	**67.7**	74.8	60.5	72.3	33.6	41.5	35.2	42.8	32.1	40.3
$+\mathcal{L}_{\text{adv}}$	64.4	73.8	66.2	73.9	62.8	73.7	32.9	40.8	34.6	42.2	31.1	39.5
$+\mathcal{L}_{\text{hist}}$, $\mathcal{L}_{\text{adv}}$, *know-ctx*	**65.9**	**74.8**	67.6	**74.9**	**64.4**	**74.7**	**34.2**	**42.1**	**35.9**	**43.5**	**32.6**	**40.7**

Table 3: Ablation results on Doc2Dial and WoW dev sets.

Experiments

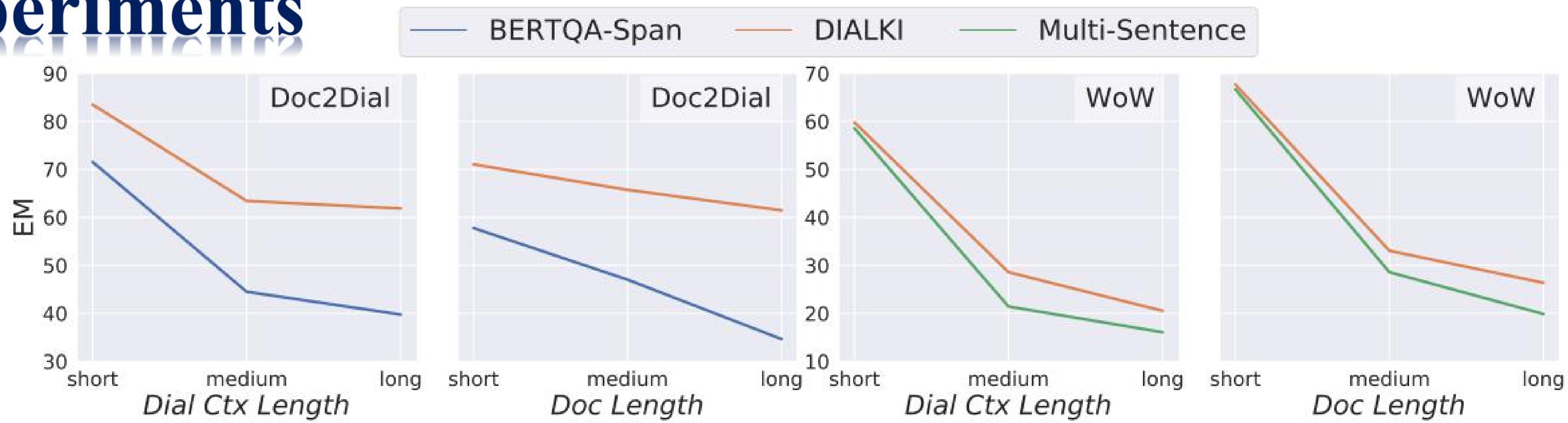


Figure 4: EM versus the length of dialogue context (# previous turns) or document (# tokens).

Figure 3: Similarity between $\mathbf{z}$ and the latest 4 history turn representations (i.e.,$\mathbf{u}_1 \ldots \mathbf{u}_4$ from left to right) on Doc2Dial (left) and WoW (right).

KI Model	**Knowledge Input**	**sacrebleu**
–	full doc	22.84
BERTQA-Token	pred span	21.42
DIALKI	pred span	25.16
DIALKI	pred span & passage	**25.84**

Table 4: Response generation results on Doc2Dial dev set. KI stands for Knowledge Identification.

Experiments

Dialogue Context	*User*: I want to trade in my license for a New York one. *Agent*: You have to exchange your-out-state driver license within 30 days of ... *User*: What if I need my license for when I go back to my other country? ... *Agent*: Are you studying in New York State? *User*: No	*User*: I heard something about co-op training program. Could you tell me about it, please? *Agent*: A co-op training program refers to ... *User*: Sounds awesome. What should I do to get that? *Agent*: Well, for that, lets do a little fact check, shall we? Are you using VA educational assistance? *User*: Yes, I'm using that.
Baseline	Do you need a New York State driver license?	Good. In that case, you may be able to get money for books, tuition and housing.
Ours	Are you a driver from another country?	Are you enrolled at an approved institution of Higher Learning?
Gold Response	Are you a licensed driver from another country?	Good. Now, are you enrolled at an approved institution of Higher Learning or IHL?

Table 5: Sample generated responses from BART with the full grounding document (baseline) or the predicted grounding span and passage by DIALKI (ours) as the additional input to the dialogue context.

Method	**Doc2Dial**		**WoW**	
	Seen	Unseen	Seen	Unseen
BERTQA-Span	76.9	72.7	–	–
Multi-Sentence	85.3	81.6	68.0	57.8
DIALKI ($\mathcal{L}_{\text{next}}$ only)	86.6	84.4	72.9	69.0
DIALKI	**88.5**	**87.5**	**73.4**	**69.7**

Table 6: Passage prediction accuracy on dev sets.

Thank you!