

Supplementary Materials

Learning Knowledge Bases with Parameters for Task-Oriented Dialogue Systems

Andrea Madotto, Samuel Cahyawijaya, Genta Indra Winata, Yan Xu,
Zihan Liu, Zhaojiang Lin, Pascale Fung

Center for Artificial Intelligence Research (CAiRE)

Department of Electronic and Computer Engineering

The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong

{amadotto, scahyawijaya, giwinata, yxucb, zliucr, zlinao}@connect.ust.hk,

pascale@ece.ust.hk

A Knowledge Embedded

We provide intuitive samples of our Knowledge Embedded approach in different datasets. Table A1 and Table A2 shows the user goal query in form of SQL syntax for tabular-formatted KB and how the KE-DELEX generate TEMPLATES. Similarly Table A3 shows the user goal query in CYPHER syntax for graph-formatted KB and how the KE-DELEX generates TEMPLATES. We further discuss the detail of the KE-DELEX for OpenDialKG in the following section.

A.1 OpenDialKG Knowledge Embedded

In OpenDialKG, we divide the KE-DELEX process into three steps: string matching, spanning tree, and dialogue generation. We perform string matching using cased letters, and we only select the entities with a minimum length of five characters to reduce the detection of false entities. To handle overlapping sequences, such as “The Dark” and “The Dark Knight” in “I enjoy watching The Dark Knight”, we perform a further filtering in each turn and we take the longest string when there is an overlapping between two or more entities.

String Matching Process We extract a set of entities that from in the dialogue based on the nodes in the graph. This set of entities are defined as the R of a user goal. To complete the user goal, we need to find the constraint C . This can be done by generating a spanning tree from the Knowledge Graph between all entities in R .

Spanning Tree We get all the relations and intermediary nodes between each pair of nodes in R . The collected relations are what we defined as constraint C of the user goal. With the given R and C , we can build a CYPHER query in form of `MATCH C RETURN R` as mentioned in the Methodology.

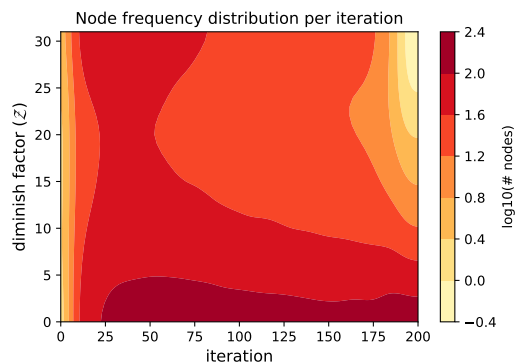


Figure A1: Distribution of #nodes over Z and iteration.

Dialogue Generation We use the CYPHER query to retrieve the equivalent nodes for the dialogue using neo4j, a graph database which supports diverse functionality for graph retrieval and manipulation. An example of our query generation is shown in Table A3. To ensure diversity of the dialogue generation, we set up a diminishing factor Z on each node, to restrict the access to the same node over time. We initialize Z with the number of edges on each node, and we decremented Z each time the node is used for the generation. In order to constraint the query with the limiting factor Z , we expand the CYPHER query into `MATCH C WHERE $Z_n > 0 \forall n \in \{C, R\}$ RETURN R`. We iteratively generate dialogues by sampling TEMPLATES. For each iteration, we randomly sampled 200 TEMPLATES and use KE-RELEX to generate the dialogues. To check the diversity of the entity in the generated dialogues, we measure the number of nodes per Z per iteration. As shown in Figure A1, the nodes with high Z is reduced over iteration and on each iteration, more and more nodes reach $Z = 0$, which ensure that the entity selected for the generation of the same TEMPLATE would include a different set of entities.

SMD

U: Where is the closest gas station?
S: Valero is 3 miles away.
U: What is the address?
S: Valero is located at 200 Alester Avenue.
SELECT type, poi, dist, addr
FROM navigation
GROUP BY type
HAVING dist = MIN(dist)

Table A1: Example of user goal query from SMD (Eric et al., 2017a). The requested information R are the attribute $type$, poi , $dist$, and $addr$, with the constrain C using as OP: GROUP BY, HAVING and MIN. The dialogues is filled with an instance from the query results, with the entities marked with the corresponding colour.

bAbI5

U: May I have a table for two people in a cheap price range?
S: Any preference for a type of cuisine?
U: With British cuisine.
S: Where should it be?
U: In Rome, thanks.
S: Let me look for some options
U: Instead could it be with Italian food?
S: Sure, is there anything else?
U: No, thanks.
S: How about Pizza Hut?
SELECT cuisine, loc, ppl, price, name
FROM restaurant

Table A2: Example of user goal query in SQL formation. The user goal specify a cuisine (i.e., Italian), but in the dialogue the user mention multiple once. To resolve this tie, we select the last mentioned cuisine entities in the dialogue.

B Human Evaluation

In this section, we show the annotators instructions used the for the human evaluation.

B.1 Instructions for Humanness Evaluation

Overview In this task, you will be given a dialogue and a response, and you have to provide a rating of the response from 1 to 4 to indicate how human-like is the response. For instance, 4 means that the response is a very natural human response, and 1 indicates the response is obviously not a human-generated response.

Steps The steps of the humanness evaluation are as following:

OpenDialog

U: Any movies similar to Gangs of New York that you can recommend?
S: Sure, Quantum of Solace has the same actor Tim Pigott-Smith.
U: Is that the one with Daniel Craig?
S: Yes, it is a thriller also starred by Daniel Craig.
U: I really love thrillers. Any suggestion?
S: Daniel Craig also starred in The Girl with the Dragon Tattoo
U: Thanks for the suggestion

MATCH

n1-[ActorsIn]→ n2,
n1-[ActorsIn]→ n3,
n4-[ActorsIn]→ n3,
n4-[ActorsIn]→ n6,
n3-[HasGenre]→ n5,
n6-[HasGenre]→ n5
RETURN n1, n2, n3,
n4, n5, n6

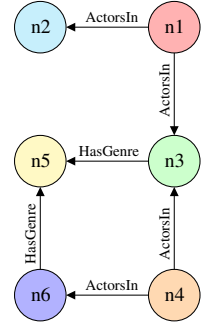


Table A3: Example of user goal query from OpenDialog (Moon et al., 2019) with CYPHER syntax (Webber, 2012), where the nodes are the requested information in R , and the labeled edges the constrains in C .

- There is a pre-filled columns with the dialogue history and a second column filled with the response text.
- There is 1 blank humanness column where you can put rating from 1 to 4, indicating how human-like is the response: 4 indicates the response is a very natural human response and 1 indicates the response is obviously not a human-generated response.
- 1. Read the dialogue from the first column.
- 2. Read the response from the second column.
- 3. Rate how human-like is the response and fill the humanness rating on the third column.

	Pre-Processing			Training/Testing					Model
	Goal	Span	KB	DST	S-ACT	KB	API	LEX-R	
<i>E2E+Pipelined</i>	✗	✓	✗	✓	✓/✗	✓	✗	✓	Sequicity (Lei et al., 2018), DAMD (Zhang et al., 2019), Structured Fusion (Mehri et al., 2019), HDSA (Chen et al., 2019), UniConv (Le et al., 2020), Soloist (Peng et al., 2020), SimpleTOD (Hosseini-Asl et al., 2020), MultiWOZ Benchmark (Budzianowski et al., 2018)
<i>E2E+API+KB</i>	✓	✗	✓	✓	✗	✓	✓	✗	MemoryNet (Bordes and Weston, 2017), Copy-Augmented Seq2Seq (Eric and Manning, 2017), Mem2Seq (Madotto et al., 2018), MLMN (Reddy et al., 2019), GLMP (Wu et al., 2019), BoSsNet (Raghu et al., 2019), KB-Trs (Haihong et al., 2019)
<i>E2E+GOLD KB</i>	✓	✗	✓	✗	✗	✓	✗	✗	KVRet (Eric et al., 2017a), Mem2Seq (Madotto et al., 2018), KBRet (Qin et al., 2019), Neural Assistant (Neelakantan et al., 2019), GLMP (Wu et al., 2019), DFF (Qin et al., 2020), GCN (Banerjee and Khapra, 2019),
<i>E2E+KB</i>	✗	✗	✗	✗	✗	✓	✗	✗	Neural Assistant (Neelakantan et al., 2019)
<i>OURS</i>	✓	✓	✓	✗	✗	✗	✗	✗	KE-Dialogue

Table A4: Comparison between different task-oriented methodologies in terms of annotation and mechanism used during pre-processing, training, and inference. *Goal* denotes user goal, *Span* denotes dialogue span, *KB* denotes knowledge base, *DST* denotes dialogue state tracking, *S-ACT* denotes speech act, *API* denotes API call, and *LEX-R* denotes lexicalization for the responses.

B.2 Instructions for Correctness Evaluation

Overview In this task, you will be given a KB, a dialogue history, and a response, and you have to provide a number of entity appearing in the KB and present in the response. You then need to check whether each of the entity is correct given the dialogue history, and the provided KB.

Steps The steps of the correctness evaluation are as following:

- There are 3 pre-filled columns, the first column is the ID to the KB, if the KB is dynamic else -1, the second column contains the dialogue history of the conversation, and the third column contains the response.
- There is 2 blank column, the first column (num_entity) is where you can put the number of entities existing in the response text and second column (correct_entity) is where you can put the number of correct entities based on the dialogue history and the KB.
- Another file for the KB is also provided in separate file named KB.txt
- 1. Read the dialogue history and the response from the second and third column.
- 2. Count how many entities on the response text that appears in the KB.
- 3. Find all the possible entities in the KB from the given the response on dialogue history and response and fill the num_entity column.

Statistics		CamRest	SMD	MWoZ
Humanness	#annotation	3	3	3
	#utterance	150	450	495
	avg. deviation	0.88	0.74	0.85
Correctness	#annotation	1	1	1
	#utterances	147	255	339

Table B5: Human evaluation statistics.

- 4. Decide whether the entities in the response are in one of the possible entities in the KB.
- 5. Check whether the entities in the response text answer the given dialogue history or not (you need to make sure that the relation between each entity’s attribute are also correct)
- 6. Count the number of correct entities attributes in the given text and fill the correct_entity column

B.3 Human Evaluation Results

In Humanness collected 3 annotations for each sample, while for correctness we used 1 annotation for each sample made by an expert. We take the mean of the annotation score to get the inter-rater agreement score. Our human evaluation reaches statistical significance with 95% confidence interval. We report the human evaluation statistics for each dataset in Table B5. The result of humanness and correctness human evaluation are shown in Figure B2 and Figure B3 respectively.

C System Comparison

To make a clear distinction of our work to existing task-oriented dialogue systems, we categorize

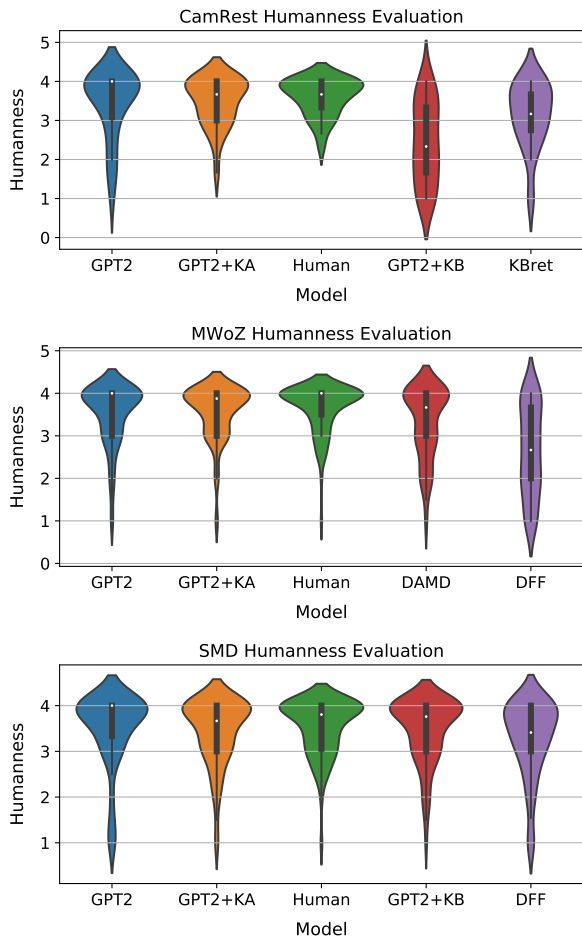


Figure B2: Humanness evaluation in CamRest, MWOZ, and SMD dataset.

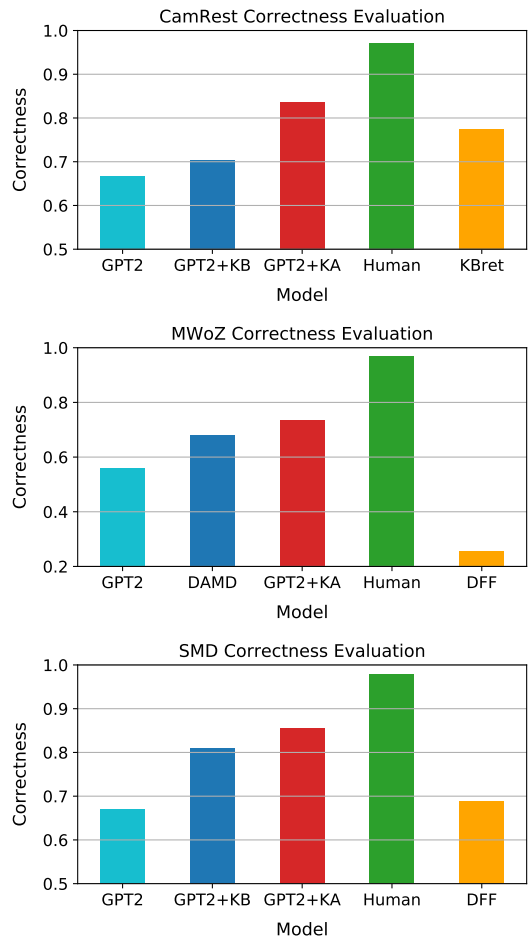


Figure B3: Humanness evaluation in CamRest, MWOZ, and SMD dataset.

them based on the annotated information and external dependencies used in the pre-processing phase and training-inference phase, such as knowledge base (*KB*), API call for retrieving information (*API*), user goal (*Goal*), dialogue span (*Span*), dialogue state tracking (*DST*), speech act (*S-ACT*), and lexicalization response (*LEX-R*). As shown in Table A4, we classify the existing work into four different categories *E2E+Pipelined*, *E2E+API+KB*, *E2E+GOLD KB*, and *E2E+KB*.

Our work is very distinct to all existing works because our approach does not incorporate any annotated information and external dependencies during training and inference time. Our approach utilizes some annotated information only on the pre-processing phase and it trains the model end-to-end with the knowledge-embedded dataset. Our approach is not only removing the dependencies to external dependencies but also eliminate most of the complexity of the whole training-inference process.

	GPT2	+KE25	+KE50	+KE75	+KE100
batch size	8	8	8	8	8
grad accu	4	4	4	4	4
lr	6.25e-5	6.25e-5	6.25e-5	6.25e-5	6.25e-5
epoch	30	30	30	30	30
fp16	-	-	-	-	-
max length	150	150	150	150	150
max history	50	50	50	50	50
num layer	12	12	12	12	12
num head	12	12	12	12	12
num emb	768	768	768	768	768
vocab size	50k	50k	50k	50k	50k
params	117M	117M	117M	117M	117M
topk	1	1	1	1	1

Table D6: Hyper-parameters on SMD, CAMREST, and OpenDialog. The experiments were run on several Nvidia 1080Ti.

D Experimental Settings

We report our hyper-parameters to train our model in Table D6 for SMD, CAMREST, and OpenDialog and Table D7 for MultiWOZ 2.1.

	GPT2	+KE25	+KE50	+KE100
batch size	6	6	6	6
grad accu	3	3	3	3
lr	6.25e-5	6.25e-5	6.25e-5	6.25e-5
epoch	10	10	10	5
fp16	O2	O2	O2	O2
max length	150	150	150	150
max history	50	50	50	50
num layer	12	12	12	12
num head	12	12	12	12
num emb	768	768	768	768
vocab size	50k	50k	50k	50k
params	117M	117M	117M	117M
topk	1	1	1	1

Table D7: Hyper-parameters on MultiWOZ. The experiments were run on a single Nvidia V100.

E Datasets Information

Table E8 shows the data splits (train/valid/test) and the link to download each dataset.

Dataset	Split			Source
	Train	Valid	Test	
bAbI	1,000	1,000	1,000	Website
CAMREST	406	135	135	Github repository
SMD (KVR)	2,425	302	304	Website
MultiWOZ	2,447	204	226	Github repository
attraction single	127	11	12	
hotel single	513	56	67	
restaurant single	1,199	50	62	
taxi single	326	57	52	
train single	282	30	33	
OpenDialKG	11,041	1,380	1,380	Facebook Github repository

Table E8: Dataset Statistics and Source.

F Detailed Experiment Results

We report more detailed results for bAbI-5, SMD, CamRest and MwoZ. Figure F9 shows all detailed results in bAbI dataset. Figure F11 shows all detailed results in SMD dataset. Figure F10 shows all detailed results on CamRest676 dataset. Figure F12 shows all detailed results on MWOZ 2.1 dataset.

G How many TEMPLATES are enough?

We further analyze our result to see how many TEMPLATES are enough to achieve good performance in the corresponding dataset. In CamRest dataset, as shown in Figure G5, we can see that there is a steep increase from without KE-dialogue to 10 TEMPLATES in term of F1 and a steep improvement from 10 TEMPLATES to 50 TEMPLATES in term of BLEU. This fact suggests that 50 TEMPLATES on CamRest dataset is

Model	Test	Test OOV
QRN ¹	99.60 (-)	67.80 (-)
Mem2Seq ²	97.90 (69.60)	84.50 (2.30)
BoSsNet ³	97.30 (65.60)	91.70 (18.50)
GLMP ⁴	99.20 (88.50)	92.00 (21.70)
GPT2	90.74 (31.00)	70.14 (0.00)
GPT2+KE 1	93.31 (46.10)	74.75 (2.00)
GPT2+KE 10	99.84 (98.10)	96.84 (77.20)
GPT2+KE 50	99.78 (97.10)	99.60 (95.70)
GPT2+KE 100	99.99 (99.90)	99.01 (94.90)

Table F9: Results on the bAbI dataset.¹ (Seo et al., 2017), ² (Madotto et al., 2018), ³ (Raghu et al., 2019), ³ (Wu et al., 2019).

Model	Success	BLEU	F1	Human	Correct
Human	86.08	-	-	3.60	96.97
KB-Tris ¹	-	14.80	45.30	-	-
MLMN ²	-	13.61	54.85	-	-
BoSsNet ³	-	15.20	43.10	-	-
KBRet ⁴	62.03	18.64	55.76	3.13	77.33
GPT2	30.38	13.58	34.69	3.42	66.67
GPT2+KB	62.03	13.59	50.45	2.42	70.37
GPT2+KE10	62.03	16.55	52.15	-	-
GPT2+KE50	70.89	17.85	55.81	-	-
GPT2+KE100	72.15	17.78	54.04	-	-
GPT2+KE161	74.68	18.00	54.85	3.48	83.50

Table F10: Detailed results on CAMREST dataset.¹ (Haihong et al., 2019). ² (Reddy et al., 2019). ³ (Raghu et al., 2019). ⁴ (Qin et al., 2019). We re-evaluate ⁴ using our script that includes postcode as entity and removes the API-call from the F1-count.

enough to represent the whole dataset. In MWOZ dataset, as shown in Figure G4, with 100 templates the inform and success scores are still increasing while the BLEU score remains stable over TEMPLATES. This suggests that we need more than 100 TEMPLATES to get the optimum benefit from our approach.

In SMD dataset, as shown in G6, in Schedule domain the F1-scores keep increasing steadily until 50 TEMPLATES and slowing down in 75 and 100 TEMPLATES. In Navigation domain there is a steep increase of F1-score from the one without KE-dialogue to the one with 10 TEMPLATES. In weather domain, the F1-score increases steadily from 10 to 100 TEMPLATES. This results suggest on Schedule domain, around 100 TEMPLATE is needed to get the optimal score, while on navigation domain, only a around 10 to 25 TEMPLATES is required, and Weather domain more than 100 TEMPLATES is required in order to achieve the optimal score.

Model	BLEU	Ent.	Nav.	Wea.	Sch.	Hum.	Cor.
KVRet ¹	13.20	48.00	44.50	53.30	62.90	-	-
MLMN ²	17.10	55.10	41.30	47.00	68.30	-	-
BoSsNet ³	8.3	35.9	-	-	-	-	-
Mem2Seq ⁴	12.20	33.40	20.00	49.30	32.80	-	-
KBRet ⁵	13.90	53.70	54.50	52.20	55.60	-	-
KB-Trs ⁶	13.90	37.10	23.30	48.20	51.20	-	-
GLMP ⁷	13.90	60.70	54.60	56.50	72.50	-	-
DFP ⁸	14.40	62.70	57.90	57.60	73.10	3.28	68.90
GPT2	15.60	39.11	23.41	53.74	52.26	3.49	67.05
GPT2+KB	17.03	58.60	48.37	62.87	72.22	3.47	81.03
GPT2+KE 10	14.18	52.88	50.26	51.64	58.62	-	-
GPT2+KE 25	14.22	55.00	50.46	52.91	64.87	-	-
GPT2+KE 50	14.90	56.43	50.04	54.25	69.60	-	-
GPT2+KE 75	16.31	58.79	52.56	56.39	71.89	-	-
GPT2+KE 100	17.35	59.78	53.53	57.73	72.58	3.44	85.56
Human ¹	13.50	60.70	55.20	61.60	64.30	3.54	97.92

Table F11: Results on the SMD (KVR) dataset. ¹Eric et al. (2017b) ²(Reddy et al., 2019) ³(Raghu et al., 2019) ⁴(Madotto et al., 2018) ⁵(Qin et al., 2019) ⁶(Haihong et al., 2019) ⁷(Wu et al., 2019) ⁸(Qin et al., 2020)

Model	Inform	Success	BLEU	F1	Train	Attraction	Hotel	Rest	Taxi	Human	Correct
Human	-	-	-	-	-	-	-	-	-	3.66	96.85
Mem2Seq ²	-	-	6.60	21.62	-	22.00	21.00	22.40	-	-	-
DSR ³	-	-	9.10	-	30.00	28.00	27.00	33.40	-	-	-
GLMP ⁴	-	-	6.90	-	32.40	24.40	28.10	38.40	-	-	-
DFP ⁵	-	-	9.40	-	35.10	28.10	30.60	40.90	-	2.65	25.53
GPT2	64.60	51.77	14.33	30.38	23.30	15.11	23.56	25.62	89.76	3.51	55.91
GPT2+KE-25	70.80	57.52	14.24	36.96	22.27	43.30	29.74	35.71	87.62	-	-
GPT2+KE-50	72.12	58.41	13.44	37.20	21.95	44.72	30.03	36.10	87.38	-	-
GPT2+KE-100	72.57	64.16	15.05	39.58	23.79	43.32	33.44	37.10	92.38	3.56	73.38
DAMD ¹	85.40	70.40	13.50	-	-	-	-	-	-	-	-
DAMD*	72.12	61.06	11.48	22.58	16.96	31.05	15.50	22.23	55.95	3.31	67.97

Table F12: Detailed results on MultiWOZ dataset. ¹(Zhang et al., 2019). ²(Madotto et al., 2018). ³(Wen et al., 2018). ⁴(Wu et al., 2019). ⁵(Qin et al., 2020). *We evaluate DAMD with our scorer.

H Example of Template Generation

Examples illustrating the KE-DELEX and KE-RELEX process for MultiWOZ, CamRest, and SMD datasets are, respectively, shown in Table H13, Table H14, and Table H15. For CamRest dataset, we remove all of the API calls in the dialogue. For MultiWOZ we keep booking API call and keep the booking reference number as is.

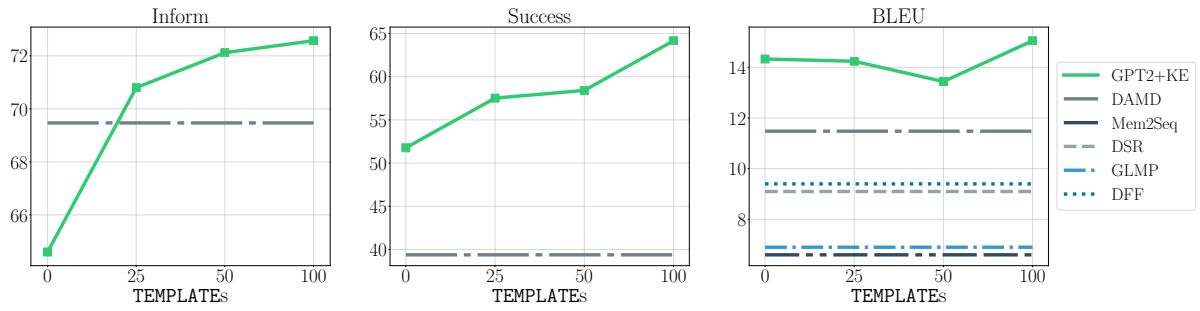


Figure G4: Inform, Success, BLEU score versus number of TEMPLATES in the MultiWOZ dataset.

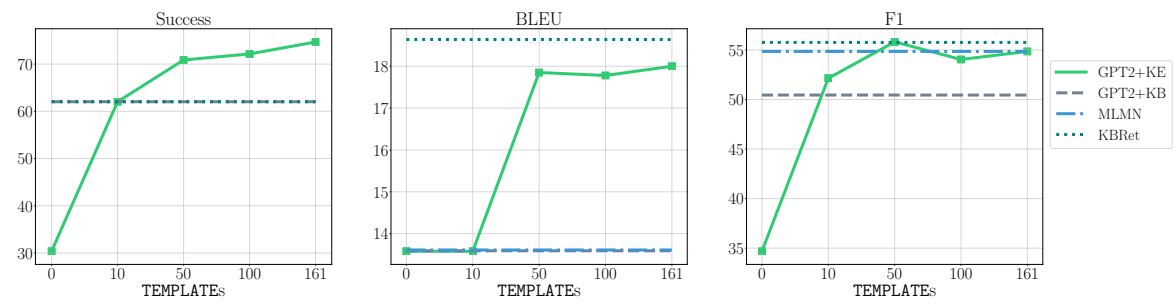


Figure G5: BLEU and F1-Score versus number of TEMPLATES by domain in the CamRest dataset.

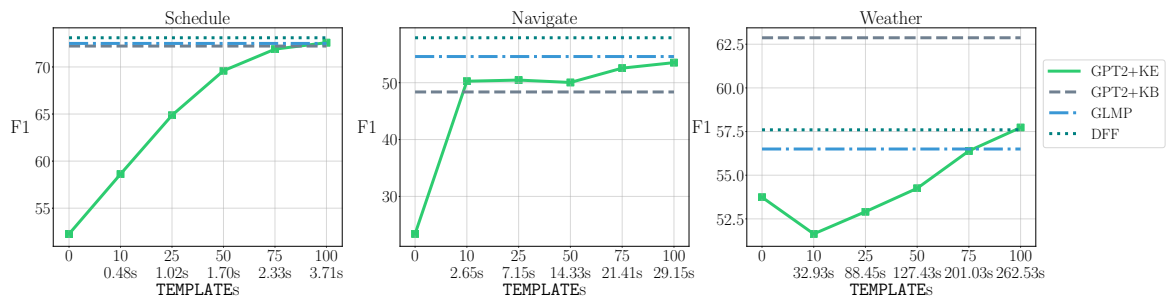


Figure G6: F1-Score versus number of TEMPLATES by domain in the SMD (Eric et al., 2017a) dataset. In the x-axis we report the average fine-tuning time per epoch.

TURN	ORIGINAL TEXT	TEMPLATE
USR	i am looking for a particular restaurant . it is called pizza hut city centre .	i am looking for a particular restaurant . it is called [name_0] .
SYS	it s a cheap italian place in the centre at regent street city centre cb21ab . the phone number is 01223323737 . would you like more info ?	it s a [price_0] [food_0] place in the [area_0] at [address_0] [post_0] . the phone number is [phone_0] . would you like more info ?
USR	that sounds like just the place ! can you book me a table for 2 at 1945 on thursday ?	that sounds like just the place ! can you book me a table for 2 at 1945 on thursday ?
SYS-API	restaurant name = 'pizza hut city centre' time = '1945' day = 'thursday' people = '2'	restaurant name = ' [name_0] ' time = '1945' day = 'thursday' people = '2'
API	f3k2pqzz	f3k2pqzz
SYS	you are all set . i have got a table for 2 at 1945 this thursday . your reference number is f3k2pqzz .	you are all set . i have got a table for 2 at 1945 this thursday . your reference number is f3k2pqzz .
USR	ok sounds good . thank you !	ok sounds good . thank you !
SYS	you are welcome . is there anything else i can assist you with today ?	you are welcome . is there anything else i can assist you with today ?
USR	that is all . thank you for your help !	that is all . thank you for your help !
SYS	you are welcome . have a great day .	you are welcome . have a great day .

USER GOAL QUERY

SELECT area, food, price, name, address, phone, postcode **FROM** Restaurant

Area	Food	Price	Name	Address	Phone	Postcode
centre	italian	cheap	pizza hut city centre	Regent Street City Centre	01223323737	cb21ab
centre	indian	cheap	kohinoor	74 Mill Road City Centre	01223323639	cb12as
centre	chinese	expensive	hk fusion	21 Burleigh Street City Centre	01223355909	cb11dg
south	modern european	moderate	restaurant alimentum	152 - 154 Hills Road	01223413000	cb28pb

TURN	RELEX 1	RELEX 2
USR	i am looking for a particular restaurant . it is called kohinoor .	i am looking for a particular restaurant . it is called restaurant alimentum .
SYS	it s a cheap indian place in the centre at 74 mill road city centre cb12as . the phone number is 01223323639 . would you like more info ?	it s a moderate modern european place in the south at 152 - 154 hills road cb28pb . the phone number is 01223413000 . would you like more info ?
USR	that sounds like just the place ! can you book me a table for 2 at 1945 on thursday ?	that sounds like just the place ! can you book me a table for 2 at 1945 on thursday ?
SYS-API	restaurant name = ' kohinoor ' time = '1945' day = 'thursday' people = '2'	restaurant name = ' restaurant alimentum ' time = '1945' day = 'thursday' people = '2'
API	f3k2pqzz	f3k2pqzz
SYS	you are all set . i have got a table for 2 at 1945 this thursday . your reference number is f3k2pqzz .	you are all set . i have got a table for 2 at 1945 this thursday . your reference number is f3k2pqzz .
USR	ok sounds good . thank you !	ok sounds good . thank you !
SYS	you are welcome . is there anything else i can assist you with today ?	you are welcome . is there anything else i can assist you with today ?
USR	that is all . thank you for your help !	that is all . thank you for your help !
SYS	you are welcome . have a great day .	you are welcome . have a great day .

Table H13: MultiWOZ 2.1 KE-DELEX and KE-RELEX Examples

TURN	ORIGINAL TEXT	TEMPLATE
USR	i want a moderately priced restaurant in the east part of town .	i want a [price.1] ly priced [type.1] in the [area.1] part of town .
SYS	would you prefer indian food or italian food ?	would you prefer [food.2] food or [food.3] food ?
USR	i do n't care . give me their phone number .	i do n't care . give me their phone number .
SYS	curry prince is a moderately priced restaurant in the east part of town that serves indian food . would you like the phone number ?	[name.2] is a [price.2] ly priced [type.2] in the [area.2] part of town that serves [food.2] food . would you like the phone number ?
USR	yes give me the phone number .	yes give me the phone number .
SYS	the phone number for Curry Prince is 01223566388 .	the phone number for [name.2] is [phone.2] .
USR	awesome , thank you ! goodbye !	awesome , thank you ! goodbye !
SYS	you 're welcome , goodbye .	you 're welcome , goodbye .

USER GOAL QUERY

SELECT area, food, price, name, phone **FROM** Restaurant

Area	Food	Price	Name	Address	Phone	Postcode	Location
east	indian	moderate	curry prince	451 newmarket road	01223566388	cb58jj	52.213072,0.149771
east	italian	moderate	pizza hut	cambridge retail park	01223323737	cb58wr	52.209742,0.146975
east	indian	moderate	rajmahal	7 barnwell road	01223244955	cb58rg	52.211144,0.16324

TURN	RELEX 1	RELEX 2
USR	i want a moderately priced restaurant in the east part of town .	i want a moderately priced restaurant in the east part of town .
SYS	would you prefer indian food or italian food ?	would you prefer italian food or indian food ?
USR	i do n't care . give me their phone number .	i do n't care . give me their phone number .
SYS	rajmahal is a moderately priced restaurant in the east part of town that serves indian food . would you like the phone number ?	pizza hut is a moderately priced restaurant in the east part of town that serves italian food . would you like the phone number ?
USR	yes give me the phone number .	yes give me the phone number .
SYS	the phone number for rajmahal is 01223244955 .	the phone number for pizza hut is 01223323737 .
USR	awesome , thank you ! goodbye !	awesome , thank you ! goodbye !
SYS	you 're welcome , goodbye .	you 're welcome , goodbye .

Table H14: CamRest Delex and Relex Examples

TURN	ORIGINAL TEXT	TEMPLATE
USR	where is the closest rest stop ?	where is the closest [poi_type_1] ?
SYS	the westin is 2 miles away	[poi_11] is [distance_11] away
USR	what is the address ?	what is the address ?
SYS	the westin is located at 329 el camino real	[poi_11] is located at [poi_address_11] .
USR	thanks set the navigation to there please	thanks set the navigation to there please
SYS	you are welcome , i set the navigation	you are welcome , i set the navigation

User Goal Query

SELECT type, poi, dist, addr **FROM** navigation **GROUP BY** type **HAVING** dist = **MIN**(dist)

POI	POI Type	Address	Distance	Traffic
whole foods	grocery store	819 alma st	2 miles	heavy traffic
hotel keen	rest stop	578 arbol dr	3 miles	no traffic
home	home	338 alester ave	3 miles	no traffic
round table	pizza restaurant	113 anton ct	4 miles	heavy traffic
mandarin roots	chinese restaurant	271 springer street	3 miles	heavy traffic

TURN	RELEX 1	RELEX 2
USR	where is the closest grocery store ?	where is the closest home ?
SYS	whole foods is 2 miles away	home is 3 miles away
USR	what is the address ?	what is the address ?
SYS	whole foods is located at 819 alma st	home is located at 338 alester ave
USR	thanks set the navigation to there please	thanks set the navigation to there please
SYS	you are welcome , i set the navigation	you are welcome , i set the navigation

Table H15: SMD Delex and Relex Example