

PLATO-KAG: Unsupervised Knowledge-Grounded Conversation via Joint Modeling

Xinxian Huang, Huang He, Siqi Bao, Fan Wang, Hua Wu, Haifeng Wang

Baidu Inc., China

{huangxinxian01, hehuang, baosiqi, wang.fan, wu_hua, wanghaifeng}@baidu.com

Abstract

Large-scale conversation models are turning to leveraging external knowledge to improve the factual accuracy in response generation. Considering the infeasibility to annotate the external knowledge for large-scale dialogue corpora, it is desirable to learn the knowledge selection and response generation in an unsupervised manner. In this paper, we propose PLATO-KAG (**K**nowledge-**A**ugmented **G**eneration), an unsupervised learning approach for end-to-end knowledge-grounded conversation modeling. For each dialogue context, the top-k relevant knowledge elements are selected and then employed in knowledge-grounded response generation. The two components of knowledge selection and response generation are optimized jointly and effectively under a balanced objective. Experimental results on two publicly available datasets validate the superiority of PLATO-KAG.

1 Introduction

Recently, the capability of large-scale pre-trained models has been verified in open-domain dialogue generation, including Meena (Adiwardana et al., 2020), Blender (Roller et al., 2021), and PLATO-2 (Bao et al., 2020). Without introducing explicit knowledge in learning process, substantive knowledge is implicitly embedded into parameters from the training corpus. However, these models are found to suffer from knowledge hallucinations (Roller et al., 2021; Marcus, 2020), producing plausible statements with factual errors. To boost the generation accuracy, there is a trend to leverage external knowledge in addition to the parameters of large-scale pre-trained models (Guu et al., 2020; Lewis et al., 2020).

In knowledge-grounded conversation, several datasets have been collected through crowdsourcing (Dinan et al., 2019; Gopalakrishnan et al., 2019; Komeili et al., 2021). Given that manual annotation is expensive and time-consuming, it is not

feasible to annotate the corresponding knowledge for each response on a large scale. Therefore, it is desirable to develop knowledge-grounded dialogue generation models without reliance on explicit knowledge labels.

Some attempts have been made to learn the unsupervised retrieval of external knowledge based on semantic similarity (Ghazvininejad et al., 2018; Dinan et al., 2019). Whereas, there exists the one-to-many phenomenon in knowledge-grounded conversation (Kim et al., 2019), where multiple knowledge elements can be appropriate to reply a given context. The prior top-1 knowledge selection employed by these approaches (Ghazvininejad et al., 2018; Dinan et al., 2019) has difficulties to hit the knowledge contained in the target response, deteriorating the learning of knowledge utilization. As an improvement, PostKS (Lian et al., 2019) and KnowledGPT (Zhao et al., 2020) rely on the target response to identify the grounded knowledge. However, involving the posterior knowledge selection will inevitably cause discrepancy between the training and inference stages (Zhao et al., 2019).

In this paper, we propose an unsupervised approach for end-to-end knowledge-grounded conversation modeling, namely PLATO-KAG (**K**nowledge-**A**ugmented **G**eneration). As shown in Figure 1, given each dialogue context, the top-k relevant knowledge elements are selected for the subsequent response generation. Then, the model learns to generate the target response grounded on each of the selected knowledge. The generation probability can in turn provide back-propagating signal for the precedent knowledge selection. These two components of knowledge selection and response generation are optimized jointly.

Two essential ingredients contribute to the performance of PLATO-KAG: top-k knowledge selection and balanced joint training. Firstly, in comparison to the conventional top-1 selection, top-k

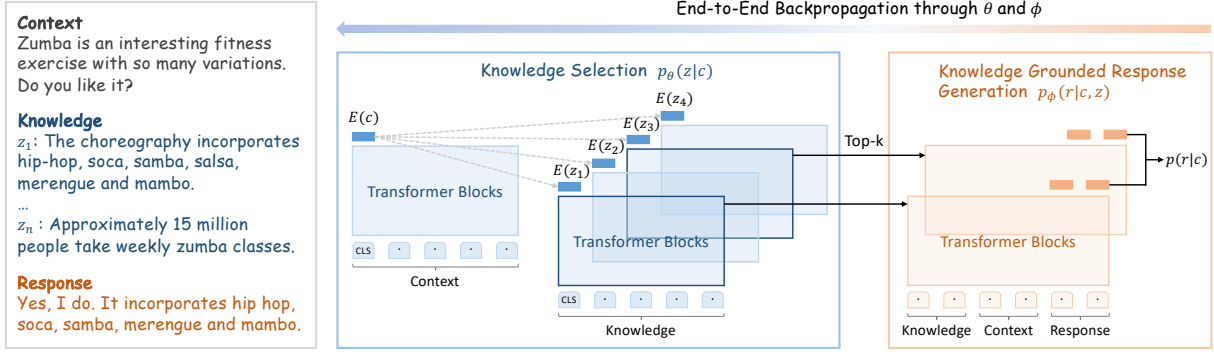


Figure 1: An overview of joint training in PLATO-KAG. For each dialogue context, top-k relevant knowledge elements are selected and employed in response generation. The generation probability can reflect the quality of the precedent knowledge selection. These two components of knowledge selection θ and response generation ϕ are optimized jointly in an unsupervised manner.

selection remarkably increases the chance to hit the grounded knowledge and improves the effectiveness of prior knowledge selection. Without the interlude of posterior knowledge selection, we manage to avoid the discrepancy between training and inference stages. Secondly, considering the difference of knowledge selection and response generation, balanced training is further designed for their effective joint optimization. To evaluate the performance of the proposed method, comprehensive experiments have been carried out on two publicly available datasets. Experimental results demonstrate that our method achieves better performance as compared with other state-of-the-art unsupervised approaches.¹

2 Methodology

There are two main components in PLATO-KAG: knowledge selection and knowledge-grounded response generation.

2.1 Knowledge Selection

As shown in Figure 1, a dual encoder with shared parameters (Siamese network) (Bromley et al., 1993) is employed in knowledge selection, where the semantic representations of the dialogue context and knowledge are extracted independently. Then the relevance between the dialogue context c and each piece of knowledge z is estimated by:

$$f(c, z) = (W_c E(c))^T (W_z E(z)) \quad (1)$$

where $E(\cdot)$ is the encoder's output on the [CLS] token, corresponding to the input's pooled repre-

sentation. W_c and W_z denotes the linear projection matrix for the dialogue context and knowledge, respectively. The relevance function f calculates the inner product of these two projected embeddings.

For the subsequent response generation, the top-k knowledge elements with highest relevance scores are selected. The prior selection probability is further normalized as:

$$p_\theta(z|c) = \frac{\exp(f(c, z))}{\sum_{z'} \exp(f(c, z'))} \quad (2)$$

where z' is one element from the top-k relevant knowledge. The benefits brought by the top-k knowledge selection are two-fold. First, top-k selection significantly increases the robustness of prior knowledge selection, as compared with the widely adopted top-1 knowledge selection (Dinan et al., 2019). As mentioned before, there exists the one-to-many problem in knowledge-grounded conversation (Kim et al., 2019). The top-k selection remarkably increases the chance to hit the knowledge and facilitates the training of generation model grounded on appropriate knowledge. Second, for the generation of one response, it is computational intractable to marginalize over the whole knowledge set. The top-k selection is an effective approximation, as most knowledge elements are not relevant with the current dialogue context.

2.2 Knowledge-Grounded Response Generation

The overall probability of generating the target response is estimated as follows:

$$p(r|c) = \sum_z p_\theta(z|c) p_\phi(r|c, z) \quad (3)$$

¹Our training code and models will be released at <https://github.com/PaddlePaddle/Knover/tree/develop/projects/PLATO-KAG>.

where the summation is running over the top-k selected knowledge elements. The second part of knowledge-grounded response generation can be further decomposed into the following form, if conditioned on one piece of knowledge:

$$p_\phi(r|c, z) = \prod_t^T p_\phi(r_t|c, z, r_{<t}) \quad (4)$$

where $r_{<t} = r_1, \dots, r_{t-1}$. In fact, the above generation probability is dependent on the quality of knowledge selection. If the selected knowledge is coherent to the context and relevant to the target response, it is able to benefit the prediction of the target response and lead to a higher generative probability. Otherwise, it leads to a lower probability. As such, the generative probability given by Equation (4) can in turn provide learning signal for the precedent knowledge selection.

2.3 Balanced Joint Training

In PLATO-KAG, the knowledge selection and response generation are optimized jointly. Depending on the marginalization strategy over knowledge (Lewis et al., 2020), the objective in Equation (3) can be expanded in the following two ways:

$$p_{\text{seq}}(r|c) = \sum_z p_\theta(z|c) \prod_t^T p_\phi(r_t|c, z, r_{<t}) \quad (5a)$$

$$p_{\text{tok}}(r|c) = \prod_t^T \sum_z p_\theta(z|c) p_\phi(r_t|c, z, r_{<t}) \quad (5b)$$

In the sequence form of Equation (5a), it relies on one knowledge element to predict the whole sequence of the target response. In the token form of Equation (5b), the generative process can rely on different knowledge elements independently for each token.

With the sequence form, the selection of knowledge just weight like the generation of one response token. Given the long responses in knowledge-grounded conversation², the module of knowledge selection is at a distinct disadvantage during joint optimization. With the token form, the weight of knowledge selection becomes identical as that of response generation. However, in the preliminary experiments, some of its generated responses exhibit some degree of knowledge misuse, where knowledge fragments are mixed inappropriately.

²For example, the dialogue response has 18.431 words on average in the Wizard of Wikipedia dataset.

The proposed method combines the merits of these two forms and introduces the following joint training objective for knowledge-grounded dialogue generation:

$$p(r|c) \propto \sum_z p_\theta(z|c) \left(\prod_t^T p_\phi(r_t|c, z, r_{<t}) \right)^\alpha \quad (6)$$

where $\alpha > 0$ is a variable controlling the weight of knowledge selection and response generation. The sequence form is preserved for the sake of generation accuracy. It is worth noting that these two components are complementary to each other. A too small or too large value of α can lead to biased and ineffective optimization. When α is close to 0, the optimization focuses on knowledge selection, neglecting the signals from response generation. When α approaches positive infinity, the optimization focuses on response generation, ignoring the effects of knowledge selection. Therefore, it is crucial to keep the balance during the joint optimization. In PLATO-KAG, α is set to $1/T$, where T is the length of target response. Through the adaptive normalization on the second term, our method successfully maintains the balance between knowledge selection and knowledge-grounded response generation. More analyses on the component weight are included in the experiments.

3 Experiments

3.1 Settings

3.1.1 Datasets

We conducted experiments on two knowledge-grounded conversation datasets: Wizard of Wikipedia (WoW) (Dinan et al., 2019) and Holl-E (Moghe et al., 2018).

In **Wizard of Wikipedia**, two participants conduct in-depth discussion on a chosen beginning topic. One of the participants has access to relevant knowledge and plays the role of an expert (*wizard*). The other one acts as a curious learner (*apprentice*). There are 18,430/1,948/1,933 dialogues in the training/validation/test set. Validation and test sets are further split into seen and unseen parts, where the latter one is about new topics outside the training set.

In **Holl-E**, a single document about a specific movie is given as external knowledge for two participants to discuss in the conversation. There are 7,228/930/913 dialogues in the training/validation/test set. To facilitate the evaluation,

the test set includes multiple reference responses for each dialogue context. We use the scripts provided by Kim et al. (2019) to process this dataset.³

As these two datasets have annotated the ground truth knowledge used by participants to ground their conversation responses, both components of knowledge selection and knowledge-grounded response generation can be evaluated thoroughly in the experiments.

3.1.2 Baselines

We compared the proposed method with the following approaches.

Transformer Memory Network (TMN) is a classical knowledge-grounded dialogue generation method (Dinan et al., 2019). Its training can be carried out in a supervised or unsupervised way, depending on whether the ground truth knowledge label is involved or not. In our experiments, we also included the supervised TMN as the performance upper bound of unsupervised models for reference. **PostKS** is an unsupervised approach, which employs the target response to estimate the posterior distribution over knowledge (Lian et al., 2019). During training, the KL divergence is employed to reduce the gap between prior and posterior distributions. During inference, it will rely on the prior distribution to select knowledge for response generation.

KnowledGPT employs a cross encoder for knowledge selection (Zhao et al., 2020). It constructs pseudo knowledge labels based on word overlaps and uses them as weak supervision signals to warm up the models. The knowledge selection is then optimized using reinforcement learning with the rewards from generated responses. The response generation is learned gradually conditioned on knowledge selected from pseudo label to the prior distribution. They are optimized iteratively under their corresponding training objectives.

3.1.3 Implementation Details

We initialized the model parameters of knowledge selection and response generation with pre-trained dialogue generation models (Bao et al., 2020). There are 24 transformer blocks and 16 attention heads, with the embedding dimension of 1024. The maximum sequence length of context, knowledge

³<https://github.com/bckim92/sequential-knowledge-transformer/blob/master/data/holle.py>

and response is set to 256, 128 and 128, respectively. We used Adam optimizer (Kingma and Ba, 2015) with a learning rate of $2e - 5$ and a batch size of 64. The number of relevant knowledge elements (top-k) was set to 8 during training. Detailed explorations of top-k settings on the validation sets are included in the Appendix. The training process was carried out on 8 Nvidia Tesla V100 32G GPU cards. Following the convention in knowledge-grounded conversation, only the most relevant knowledge was selected for response generation during inference.

Since the original TMN and PostKS are developed on shallow networks, for the sake of fair comparison, we re-implemented them and initialized the model parameters in the way as the proposed method. For KnowledGPT, we used its open-sourced checkpoint⁴ in our experiments.

3.1.4 Evaluation Metrics

In the automatic evaluation, Perplexity (PPL) and Unigram F1 of ground truth responses (Dinan et al., 2019) are adopted to assess the response quality. Recall@1 (top-1 knowledge accuracy) is used to evaluate the performance of knowledge selection. We used the evaluation scripts provided by Dinan et al. (2019).⁵

In the human evaluation, we randomly sampled 100 examples from WoW seen and unseen test set, respectively. Each sample was distributed to three annotators and evaluated on the four aspects:

- **Coherence** evaluates whether the response is consistent and relevant with the context.
- **Informativeness** assesses whether the response contains appropriate information.
- **Engagingness** measures the annotator’s willingness to discuss with the speaker for a long conversation.
- **Hallucination** estimates the factual correctness in the response.

Coherence, informativeness and engagingness are scored on a range of [0, 1, 2], with the higher value, the better. Hallucination is evaluated on a range of [0, 1], where 0 means the response is factually correct and 1 means the response contains factual errors. The scoring criteria are provided in the Appendix. The final score of each sample was determined through majority voting.

⁴<https://github.com/zhaoxlpku/KnowledGPT>

⁵<https://github.com/facebookresearch/ParLAI>

WoW Seen Test Model	Knowledge Label	Automatic Evaluation			Human Evaluation			
		PPL↓	Recall@1	Unigram F1	Coherence	Informativeness	Engagingness	Hallucination↓
TMN	N	10.136	0.041	0.168	1.27	1.10	1.13	0.34
PostKS	N	11.577	0.224	0.187	1.33	1.28	1.30	0.21
KnowledGPT	N	19.600*	0.262 ⁺	0.183	1.16	1.16	1.12	0.28
PLATO-KAG	N	9.767	0.253	0.188	1.54	1.44	1.40	0.17
TMN	Y	9.633	0.265	0.188	1.51	1.39	1.38	0.17

WoW Unseen Test Model	Knowledge Label	Automatic Evaluation			Human Evaluation			
		PPL↓	Recall@1	Unigram F1	Coherence	Informativeness	Engagingness	Hallucination↓
TMN	N	12.910	0.042	0.156	1.33	1.07	1.12	0.40
PostKS	N	13.668	0.199	0.176	1.33	1.29	1.28	0.28
KnowledGPT	N	22.849*	0.238 ⁺	0.173	1.16	1.11	1.03	0.31
PLATO-KAG	N	11.458	0.253	0.181	1.50	1.34	1.41	0.24
TMN	Y	11.362	0.260	0.180	1.50	1.40	1.42	0.19

Table 1: Evaluation results on the WoW seen and unseen test sets. The second column indicates whether the model is trained using knowledge labels (supervised) or not (unsupervised). * Not comparable to the rest models due to the different vocabulary. ⁺ Not comparable to the rest models due to the cross encoder in knowledge selection.

3.2 Experimental Results

The evaluation results on the WoW test sets are summarized in Table 1. Besides the unsupervised models, the supervised TMN with reliance on knowledge labels during training was also included in the experiments for reference. The automatic and human evaluation results demonstrate that PLATO-KAG achieves better performance as compared with other state-of-the-art unsupervised approaches, even on par with the supervised approach. Based on appropriate knowledge selection, PLATO-KAG produces high-quality responses that are coherent, informative and engaging. Moreover, it alleviates the problem of knowledge hallucinations and generates more factual accurate responses.

As shown in the Table 1, unsupervised TMN generates less informative responses and suffers from a higher degree of hallucination. As for PostKS, based on inferior prior knowledge selection, it generates less coherent responses. Since KnowledGPT employs a cross encoder in the knowledge selection, it achieves a higher value of Recall@1. While cross encoder is hardly feasible for practical deployment given its expensive computation cost. Another factor that attributes to the weak performance of KnowledGPT might be the pre-training models used for initialization⁶. The average Fleiss’s kappa (Fleiss, 1971) in human evaluation is 0.502,

⁶The released checkpoint of KnowledGPT is developed on the general language model GPT-2, while the rest models are developed on dialogue pre-training models.

indicating that annotators have reached moderate agreement.

The evaluation results on the Holl-E test set are summarized in Table 2. In the evaluation on the multiple reference test set, we took the best score over multiple reference responses for each dialogue context. The results demonstrate that PLATO-KAG also achieves competitive results in Holl-E. PostKS obtains a slightly higher value on Unigram F1 than PLATO-KAG and supervised TMN. While the values on Distinct-1/2 (Li et al., 2016) indicate the PLATO-KAG and supervised TMN might have better capacity on lexical diversity.

3.3 Discussions

3.3.1 Case Analysis

For further qualitative analysis, two examples of generated responses from the WoW test set are provided in Table 3. It can be observed that unsupervised TMN suffers from low-quality response generation, such as generic replies with little information or statements with factual errors. In comparison, PostKS and KnowledGPT are able to generate much more informative responses, depicting contents from the selected knowledge. However, the responses fail to be coherent with the dialogue context due to the inferior knowledge selection. Among these unsupervised approaches, PLATO-KAG achieves better performance, producing coherent and informative responses.

The above analysis is also validated by the re-

Holl-E Test Model	Knowledge Label	Distinct-1/2	Single Reference			Multiple References		
			PPL↓	Recall@1	Unigram F1	PPL↓	Recall@1	Unigram F1
TMN	N	0.091 / 0.292	16.693	0.035	0.169	14.100	0.044	0.223
PostKS	N	0.120 / 0.379	16.808	0.238	0.308	11.790	0.343	0.376
PLATO-KAG	N	0.128 / 0.416	10.222	0.271	0.300	7.724	0.371	0.368
TMN	Y	0.126 / 0.400	10.081	0.296	0.300	7.964	0.394	0.363

Table 2: Automatic evaluation results on the Holl-E test set.

Context	[Topic] zumba [Wizard] hello dear, have you hard of zumba? is an exercise fitness program created by colombian dancer [Apprentice] i have! my best friend and i occasionally take classes! [Wizard] do you know that zumba is a trademark owned by zumba fitness, . [Apprentice] i didn't know that. it's interesting because there seem to be so many variations. do you like it?
Unsupervised TMN	yes i have tried it and i like it a lot.
PostKS	approximately 15 million people take weekly zumba classes in over 200,000 locations across 180 countries.
KnowledGPT	its an exercise fitness program created by colombian dancer and cyclist/choreographer alberto "beto" "power pedal"
PLATO-KAG	yes. it incorporates hip hop, soca, samba, merengue and mambo.
Supervised TMN	yes i do. it includes hip-hop and samba, social, salsa, choreography, mambo.
Ground Truth	yes, zumba fitness ambassador is abrazilian pop singer claudia leitte
Context	[Topic] elvis presley [Wizard] do you like elvis presley? he was a famous american singer [Apprentice] he is pretty good, i am from nashville, so i've even been to graceland. [Wizard] presley was born in tupelo, mississippi, and then to memphis [Apprentice] what was elvis's most popular songs?
Unsupervised TMN	his first big hit was "favorite song"
PostKS	he had a series of successful netflix appearances and chart-topping records
KnowledGPT	he became the leading figure of the new popular rock group, the new sound of rock and roll. he was very popular.
PLATO-KAG	his first rca single was "heartbreak hotel" and became a number one hit in the united states
Supervised TMN	his first rca single, "heartbreak hotel" was released in 1956 and became a number one hit in the us
Ground Truth	heartbreak hotel was a number one hit in the usa

Table 3: Examples of the generated responses on the WoW seen (upper) and unseen (bottom) test sets.

	WoW Seen Test	WoW Unseen Test
Unsupervised TMN	0.107	0.106
PostKS	0.443	0.430
KnowledGPT	0.385	0.375
PLATO-KAG	0.347	0.340
Supervised TMN	0.314	0.306
Ground Truth	0.334	0.335

Table 4: Knowledge F1 on the WoW test set.

sults on knowledge dependency. Knowledge F1 (Lian et al., 2019), which is defined as the unigram F1 between the generated response and the selected knowledge, can measure the degree of knowledge dependency. A too low Knowledge F1 value means the models hardly reference the knowledge when generating responses. A too high value indicates the models "copy" the knowledge too much, which might cause unnatural responses. The values of

	Hallucination		
	PLATO-KAG Win	Tie	PLATO-KAG w/o EK Win
WoW Seen	13	78	9
WoW Unseen	26	67	7
	Informativeness		
	PLATO-KAG Win	Tie	PLATO-KAG w/o EK Win
WoW Seen	23	61	16
WoW Unseen	35	46	19

Table 5: Comparison of hallucination and informativeness between PLATO-KAG and PLATO-KAG w/o EK on the WoW test sets.

Knowledge F1 are reported in Table 4. The results indicate that unsupervised TMN suffers from poor knowledge utilization. As for PostKS and KnowledGPT, since posterior knowledge selection

Context	[Topic] harry potter [Apprentice] everyone has seen harry potter except me. have you ever read the books or watched the movies? [Wizard] yes i am a big fan of it actually. i love reading about chronicles of the life of a young wizard [Apprentice] was he born a wizard? [Wizard] as far as the main story arc concerns, he was and he went to wizard school to improve himself. [Apprentice] who wrote these stories?
PLATO-KAG w/o EK	it was written by j.k. rowling. and it was published <i>in 1977 by hermione granger.</i>
PLATO-KAG	harry potter is a series of fantasy novels written by british author j. k. rowling.
Context	[Topic] bowling [Apprentice] i used to be really into bowling. ... [Wizard] you aim to knock over the pins at the end [Apprentice] yup, though it can be hard at times haha.
PLATO-KAG w/o EK	the most common way to bowl is to use <i>a rectangular ball</i>
PLATO-KAG	the max score is 300 which is achieved by getting 12 strikes in a row

Table 6: Examples of PLATO-KAG and PLATO-KAG w/o EK on the WoW seen and unseen test sets. Italic blocks are contents with factual errors.

is involved in their training process, their generation models learn to rely heavily on the provided knowledge, resulting in very high Knowledge F1 values. During inference with their inferior prior knowledge selection, this kind of strong dependency will lead to unrelated and unnatural response generation. Our method gets exempt from this discrepancy with end-to-end modeling and optimization. The close values of PLATO-KAG and the ground truth (0.347/0.334 on seen and 0.340/0.335 on unseen) indicates our method achieves a natural degree of knowledge utilization.

3.3.2 External Knowledge Effects on Response Quality

As discussed in the introduction, conversation models are turning to leveraging external knowledge explicitly to boost generation accuracy. To quantitatively analyze the performance, one dialogue generation model was trained on the WoW dataset without grounding on external knowledge, denoted as PLATO-KAG w/o EK. We asked annotators to compare the hallucination and informativeness between our method and PLATO-KAG w/o EK, with results summarized in Table 5. It is notable that the tie score of hallucination from PLATO-KAG w/o EK is a little inflated. This is because the model generates less informative responses, which helps keep the factual correctness (less talk, less mistake). With access to external knowledge, our method achieves better performance consistently. Moreover, the performance gaps on both metrics get enlarged from the seen to unseen test set. PLATO-KAG w/o EK produces plausible statements with

factual errors more easily under unseen topics.

Two examples of generated responses by these two models are shown in Table 6, where the contents with factual errors are displayed in italic blocks. It reveals that PLATO-KAG w/o EK has difficulties to memorize and describe the knowledge details precisely. In fact, the initial publication of Harry Potter is in 1997 and Hermione Granger is one representative character in the book instead of a publisher. Sometimes, PLATO-KAG w/o EK produces statements that are obviously problematic and against the common sense, like "a rectangular ball". By leveraging external knowledge, PLATO-KAG can generate more accurate and informative responses.

Marginalization Strategy	Component Weight	WoW Seen	WoW Unseen	Holl-E
Sequence Form	$\alpha = T$	11.455	12.926	12.043
	$\alpha = 1$	10.965	11.962	11.089
	$\alpha = 1/T$ (PLATO-KAG)	9.863	11.387	10.148
	$\alpha = 1/T^2$	10.399	11.787	12.551
Token Form	-	11.841	13.099	12.679

Table 7: Perplexity under different marginalization strategies and component weights on the WoW and Holl-E validation sets.

3.3.3 Impacts of Marginalization Strategies and Component Weight

As discussed in Section 2.3, the quality of joint optimization is effected by the marginalization strategies and component weight. Explorations on these

settings have been carried out on the validation sets, with the perplexity results summarized in Table 7. For the marginalization strategy, the token form (Equation (5b)), which depends on various knowledge elements to predict one response token, obtains relatively poor results. Under this training paradigm, the model tends to mix information from various knowledge fragments and is prone to generate low-quality responses. Two more examples are included in the Appendix to illustrate this phenomenon.

As comparison, with the marginalization strategy in sequence form (Equation (6)), the models achieve relatively better performance on perplexity. For the sequence form, one crucial factor affecting the performance is the component weight α between knowledge selection and knowledge-grounded response generation. Under the straightforward setting ($\alpha = 1$), knowledge selection weighs like one single response token. In PLATO-KAG ($\alpha = 1/T$, where T is the length of target response), the weight of knowledge selection becomes identical to that of the whole response generation. The results indicate PLATO-KAG achieves better performance with the help of balanced training. A too large or too small weight value (such as $\alpha = T$ or $\alpha = 1/T^2$) will lead to ineffective optimization and performance degradation.

4 Related Work

Knowledge-grounded conversation is becoming a more important and popular topic, with several datasets (Zhang et al., 2018; Moghe et al., 2018; Zhou et al., 2018; Dinan et al., 2019; Gopalakrishnan et al., 2019; Komeili et al., 2021) collected to study it. Besides interactive dialogues, some of these datasets have annotated the corresponding knowledge for each response, aiming to ease the learning difficulty of knowledge-grounded conversation. However, given that manual annotation is expensive and time-consuming, it is not feasible to carry out the knowledge labelling on a large scale.

Unsupervised approaches have been introduced to model knowledge-grounded conversation. Some of these such as Li et al. (2019); Yavuz et al. (2019); Lin et al. (2020) perform implicit soft fusion over provided knowledge elements and do not select knowledge explicitly. Some attempts have been made to learn the unsupervised selection of external knowledge based on semantic similarity (Ghazvininejad et al., 2018; Dinan et al., 2019).

Due to the one-to-many problem in knowledge-grounded conversation (Kim et al., 2019), the prior top-1 knowledge selection employed by these approaches has difficulties to hit the knowledge contained in the target response, and deteriorates the learning of knowledge utilization. Our top-k selection improves the robustness of prior knowledge selection. Some other works (Lian et al., 2019; Zhao et al., 2020; Ren et al., 2020) employ the target response to identify the grounded knowledge. Since the posterior knowledge selection is involved, it will inevitably cause discrepancy between the training and inference stages (Zhao et al., 2019). With end-to-end modeling and optimization, PLATO-KAG gets exempt from this discrepancy. KIF (Fan et al., 2021) explicitly selects external knowledge through a retrieval module, and fuses into one integrated representation to assist dialogue generation. While some knowledge details might be obscured with this fusion. As comparison, the knowledge keeps its independence and integrity in our response generation, which helps reduce the hallucination.

More recently, Shuster et al. (2021) attempts to utilize the pre-trained retriever DPR (Karpukhin et al., 2020). DPR has been trained on Wikipedia which includes the knowledge sets of WoW and Holl-E. Due to the concern of potential data contamination, we chose to initialize our knowledge selection module with a general dialogue model which is pre-trained on Reddit. Thus, we facilitated an unbiased setting for our experiments and the analysis of framework generalization.

5 Conclusion

In this paper, an unsupervised approach is proposed for end-to-end knowledge grounded conversation modeling. There are two main components in our method: knowledge selection and response generation. Given a dialogue context, top-k relevant knowledge elements are selected and utilized for response generation. The generation probability can in turn provide training signal for the precedent knowledge selection. Joint balanced training is further introduced for the effective optimization of these two components. Comprehensive experiments have been carried out on WoW and Holl-E, verifying the effectiveness and superiority of the proposed method.

Acknowledgements

We would like to thank the anonymous reviewers for their constructive suggestions and Yunyi Yang for the helpful discussions. This work was supported by the Natural Key Research and Development Project of China (No. 2018AAA0101900).

References

- Daniel Adiwardana, Minh-Thang Luong, David R So, Jamie Hall, Noah Fiedel, Romal Thoppilan, Zi Yang, Apoorv Kulshreshtha, Gaurav Nemade, Yifeng Lu, and Quoc V. Le. 2020. Towards a human-like open-domain chatbot. *arXiv preprint arXiv:2001.09977*.
- Siqi Bao, Huang He, Fan Wang, Hua Wu, Haifeng Wang, Wenquan Wu, Zhen Guo, Zhibin Liu, and Xinchao Xu. 2020. Plato-2: Towards building an open-domain chatbot via curriculum learning. *arXiv preprint arXiv:2006.16779*.
- Jane Bromley, Isabelle Guyon, Yann LeCun, Eduard Säckinger, and Roopak Shah. 1993. Signature verification using a "siamese" time delay neural network. *Advances in neural information processing systems*, 6:737–744.
- Emily Dinan, Stephen Roller, Kurt Shuster, Angela Fan, Michael Auli, and Jason Weston. 2019. Wizard of wikipedia: Knowledge-powered conversational agents. *International Conference on Learning Representations*.
- Angela Fan, Claire Gardent, Chloé Braud, and Antoine Bordes. 2021. Augmenting transformers with knn-based composite memory for dialog. *Transactions of the Association for Computational Linguistics*, 9:82–99.
- Joseph L Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological bulletin*, 76(5):378.
- Marjan Ghazvininejad, Chris Brockett, Ming-Wei Chang, Bill Dolan, Jianfeng Gao, Wen-tau Yih, and Michel Galley. 2018. A knowledge-grounded neural conversation model. In *Proceedings of the AAAI Conference on Artificial Intelligence*.
- Karthik Gopalakrishnan, Behnam Hedayatnia, Qinglang Chen, Anna Gottardi, Sanjeev Kwatra, Anu Venkatesh, Raefer Gabriel, Dilek Hakkani-Tür, and Amazon Alexa AI. 2019. Topical-chat: Towards knowledge-grounded open-domain conversations. In *Interspeech*, pages 1891–1895.
- Kelvin Guu, Kenton Lee, Zora Tung, Panupong Pasupat, and Ming-Wei Chang. 2020. Realm: Retrieval-augmented language model pre-training. In *Proceedings of the 37th Annual International Conference on Machine Learning*, pages 3929–3938.
- Vladimir Karpukhin, Barlas Oguz, Sewon Min, Patrick Lewis, Ledell Wu, Sergey Edunov, Danqi Chen, and Wen-tau Yih. 2020. Dense passage retrieval for open-domain question answering. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6769–6781.
- Byeongchang Kim, Jaewoo Ahn, and Gunhee Kim. 2019. Sequential latent knowledge selection for knowledge-grounded dialogue. In *International Conference on Learning Representations*.
- Diederik P Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*.
- Mojtaba Komeili, Kurt Shuster, and Jason Weston. 2021. Internet-augmented dialogue generation. *arXiv preprint arXiv:2107.07566*.
- Patrick Lewis, Ethan Perez, Aleksandara Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. In *Advances in Neural Information Processing Systems*, pages 9459–9474.
- Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao, and William B Dolan. 2016. A diversity-promoting objective function for neural conversation models. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 110–119.
- Zekang Li, Cheng Niu, Fandong Meng, Yang Feng, Qian Li, and Jie Zhou. 2019. Incremental transformer with deliberation decoder for document grounded conversations. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 12–21.
- Rongzhong Lian, Min Xie, Fan Wang, Jinhua Peng, and Hua Wu. 2019. Learning to select knowledge for response generation in dialog systems. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, pages 5081–5087.
- Xiexiong Lin, Weiyu Jian, Jianshan He, Taifeng Wang, and Wei Chu. 2020. Generating informative conversational response using recurrent knowledge-interaction and knowledge-copy. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 41–52.
- Gary Marcus. 2020. The next decade in ai: four steps towards robust artificial intelligence. *arXiv preprint arXiv:2002.06177*.
- Nikita Moghe, Siddhartha Arora, Suman Banerjee, and Mitesh M Khapra. 2018. Towards exploiting background knowledge for building conversation systems. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2322–2332.

- Pengjie Ren, Zhumin Chen, Christof Monz, Jun Ma, and Maarten de Rijke. 2020. Thinking globally, acting locally: Distantly supervised global-to-local knowledge selection for background based conversation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 8697–8704.
- Stephen Roller, Emily Dinan, Naman Goyal, Da Ju, Mary Williamson, Yinhan Liu, Jing Xu, Myle Ott, Kurt Shuster, Eric M Smith, et al. 2021. Recipes for building an open-domain chatbot. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics*, page 300–325.
- Kurt Shuster, Spencer Poff, Moya Chen, Douwe Kiela, and Jason Weston. 2021. Retrieval augmentation reduces hallucination in conversation. *arXiv preprint arXiv:2104.07567*.
- Semih Yavuz, Abhinav Rastogi, Guan-Lin Chao, and Dilek Hakkani-Tur. 2019. Deepcopy: Grounded response generation with hierarchical pointer networks. In *Proceedings of the 20th Annual SIGdial Meeting on Discourse and Dialogue*, pages 122–132.
- Saizheng Zhang, Emily Dinan, Jack Urbanek, Arthur Szlam, Douwe Kiela, and Jason Weston. 2018. Personalizing dialogue agents: I have a dog, do you have pets too? In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, pages 2204–2213.
- Tiancheng Zhao, Kaige Xie, and Maxine Eskenazi. 2019. Rethinking action spaces for reinforcement learning in end-to-end dialog agents with latent variable models. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1208–1218.
- Xueliang Zhao, Wei Wu, Can Xu, Chongyang Tao, Dongyan Zhao, and Rui Yan. 2020. Knowledge-grounded dialogue generation with pre-trained language models. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, pages 3377–3390.
- Kangyan Zhou, Shrimai Prabhumoye, and Alan W Black. 2018. A dataset for document grounded conversations. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 708–713.

A Human Evaluation Scoring Criteria

The detailed criteria used in human evaluation are provided in Table 8. To evaluate the criteria of hallucination, the human annotators were provided with referenced knowledge and allowed to use search engine to check the factual correctness.

Score	Coherence
0	<ul style="list-style-type: none"> The response is irrelated with the context. The response simply repeats the context. The response has obvious conflicts with the context. There are serious logic conflicts within the response.
1	<ul style="list-style-type: none"> The response is less coherent with the context. There are minor logic conflicts within the response.
2	<ul style="list-style-type: none"> The response is consistent and relevant with the context.

Score	Informativeness
0	<ul style="list-style-type: none"> The response contains no information. The response simply repeats the context and contains no additional information. Since coherence score is 0, the information is invalid.
1	<ul style="list-style-type: none"> The response contains a little additional information.
2	<ul style="list-style-type: none"> The response has appropriate information.

Score	Engagingness
0	<ul style="list-style-type: none"> I am not willing to talk with the speaker.
1	<ul style="list-style-type: none"> It is a bit boring, but still ok to go on talking with this speaker.
2	<ul style="list-style-type: none"> I would like to talk with the speaker for a long conversation.

Score	Hallucination
0	<ul style="list-style-type: none"> The response is factually correct.
1	<ul style="list-style-type: none"> Some details in the response are factually incorrect.

Table 8: Scoring criteria of four metrics in human evaluation.

B Knowledge Hallucination with Token Form Marginalization Strategy

In our preliminary experiments, the model trained with token form marginalization strategy exhibits a certain degree of knowledge misuse. Some examples are provided in Table 9. In the upper case, the fragments from the reference knowledge "in his spare time" and "selling pharmaceuticals" are mixed to "selling his spare time to a pharmacist" in the generated response. The bottom case also exhibits a low level of fidelity to the selected knowledge. One hypothesis to this phenomenon is that under the objective of token form (Equation (5b)), the model learns to blend information across various knowledge fragments to predict one response token. Under this training paradigm, the learned

model is prone to hallucinate knowledge, mixing knowledge fragments inappropriately.

Context	<p>[Topic] nicholas sparks [Apprentice] i've heard nicholas sparks is a great writer but i'm not sure if i've ever read any of his books ... [Wizard] yes it was called the passing, though his first published work didn't come until 1990. [Apprentice] what was the passing about?</p>
Referenced Knowledge	<p>in 1993, sparks wrote his breakthrough novel the notebook in his spare time while selling pharmaceuticals in washington, d.c.. two years later, his novel was discovered by literary agent theresa park who offered to represent him.</p>
Generated Response	<p>i believe it was about him <i>selling his spare time to a pharmacist</i> who helped him sell his notebook in 1994.</p>

Context	<p>[Topic] john grisham [Apprentice] did you know john grisham is a popular author? [Wizard] yes, his first bestseller "the firm" sold more than 7 million copies! ... [Apprentice] do you know any other movies that were made into movies from his books?</p>
Referenced Knowledge	<p>eight of his other novels have also been adapted into films: "the chamber", "the client", "a painted house", "the pelican brief", "the rainmaker", "the runaway jury", "skipping christmas", and "a time to kill".</p>
Generated Response	<p>yes, "the runaway jury" has also been <i>adapted into other novels "the client of christmas", "children of men" and "a time to kill"</i></p>

Table 9: Examples of knowledge misuse under token form marginalization strategy on the WoW seen and unseen test sets. Referenced knowledge is the model selected to ground response generation. Italic blocks are contents with factually errors.

C Explorations of Top-k Settings in Knowledge Selection

Top-k	WoW Seen		WoW Unseen		Holl-E	
	PPL↓	Recall@1	PPL↓	Recall@1	PPL↓	Recall@1
1	10.583	0.064	12.842	0.061	16.720	0.031
2	9.897	0.250	11.344	0.228	10.634	0.251
4	9.865	0.258	11.339	0.228	10.359	0.262
8	9.863	0.257	11.325	0.231	10.246	0.266
16	9.871	0.256	11.321	0.231	10.309	0.263

Table 10: Perplexity and Recall@1 under different top-k settings on the WoW and Holl-E validation sets.

To decide the proper number of relevant knowledge elements (top-k) for the training process, we conducted 3 runs of experiments for each top-k setting (k = 1, 2, 4, 8, 16). The median results

on the validation sets are reported in Table 10. As discussed in the introduction, the prior top-1 knowledge selection hardly hits the grounded knowledge and suffers from relatively poor results. It also reveals a trend that models with larger k values can achieve better performance. It reaches stable states around $k = 8$. To balance the efficiency and performance, we set $k = 8$ in our experiments.