Diverse and Informative Dialogue Generation with Context-Specific Commonsense Knowledge Awareness

Sixing Wu, Ying Li, Dawei Zhang, Yang Zhou and Zhonghai Wu

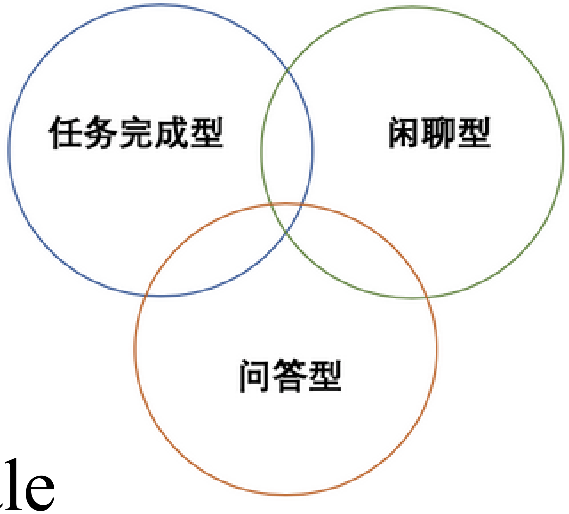
Speaker: ZhangLiang
2021.02.21

Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 5811–5820
July 5 - 10, 2020. c ?2020 Association for Computational Linguistics

Content

Introduction

- Open-domain dialogue response generation systems still suffer from generating generic and boring responses, such as ”I don’t know.”



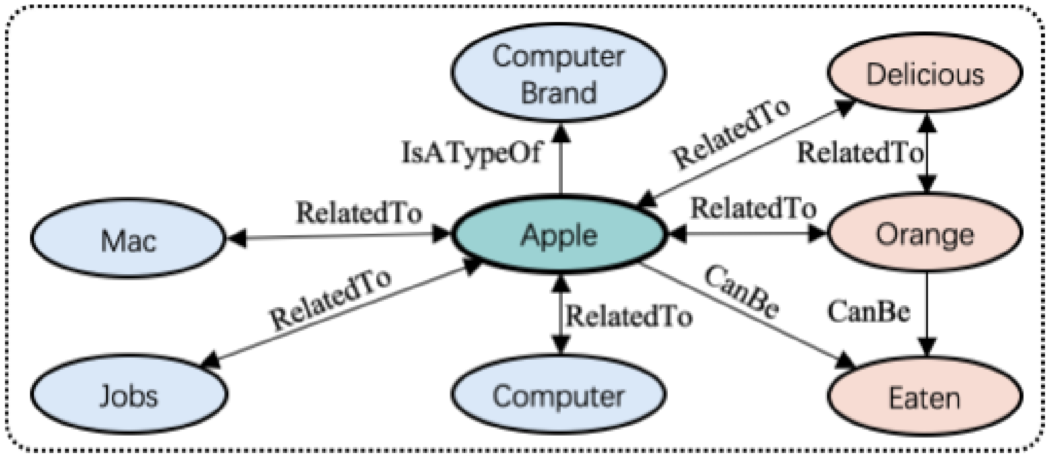
- Researchers have begun to introduce large-scale knowledge graphs for enhancing the dialogue generation

Introduction

Some challenges in knowledge-aware dialogue generation still keep unsolved.

- 1) An entity word usually can refer to different concepts, i.e., an entity has multiple meanings, but only one specific concept is involved in a particular context.
- 2) Even if we only consider a particular entity meaning, the related knowledge facts may cover various target topics. However, some of those topics do not contribute to the dialogue generation.
- 3) The integration of the knowledge and the dialogue generation in previous approaches is insufficient, including the way of integration, as well as the types of knowledge.

Introduction



Message: Apple's new product is awesome!

#1: Yes, a beautiful new Mac.

#2: I love it, as delicious as the orange.

Figure 1: An illustrative example. #1 shows the response generated with a highly relevant fact, #2 shows the response generated with irrelevant facts.

This paper's work:

- First, it design a Felicitous Fact mechanism to help the model highlight the knowledge facts that are highly relevant to the context, that is, "Felicitous Facts".
- Next, Context-Knowledge Fusion is proposed to lift the role of knowledge facts in the dialogue generation, by fusing the context and the felicitous knowledge before the decoding.
- Last, ConKADI can generate three types of words owing to the Flexible Mode Fusion module, which aims at simultaneously fusing multiple types of knowledge.

Related Work

- Seq2Seq (Sutskever et al., 2014; Vinyals and Le, 2015) has been widely used in the open-domain dialogue generation.

- ConceptNet (Speer 5813 et al., 2017) is a multilingual open-domain commonsense knowledge graph, which is designed to represent the general knowledge and to improve understanding of the meanings behind the words people use.

- the current state-of-the-art CCM (Zhou et al., 2018)

Approach

Given a training data D of triplets

a query message $X = (x_1, \ldots, x_n)$,

a response $Y = (y_1, \ldots, y_m)$,

a set of commonsense knowledge facts $F = \{f_1, \ldots, f_l\}$.

The training goal： maximize the probability $\sum_{(X,Y,F)\in\mathcal{D}} \frac{1}{|\mathcal{D}|} p(Y|X,F)$;

the inference goal： find $Y^* = \arg\max_Y p(Y|X,F)$.

Knowledge facts F are retrieved from the knowledge graph G ; each fact is organized as a triplet (h, r, t).

Approach

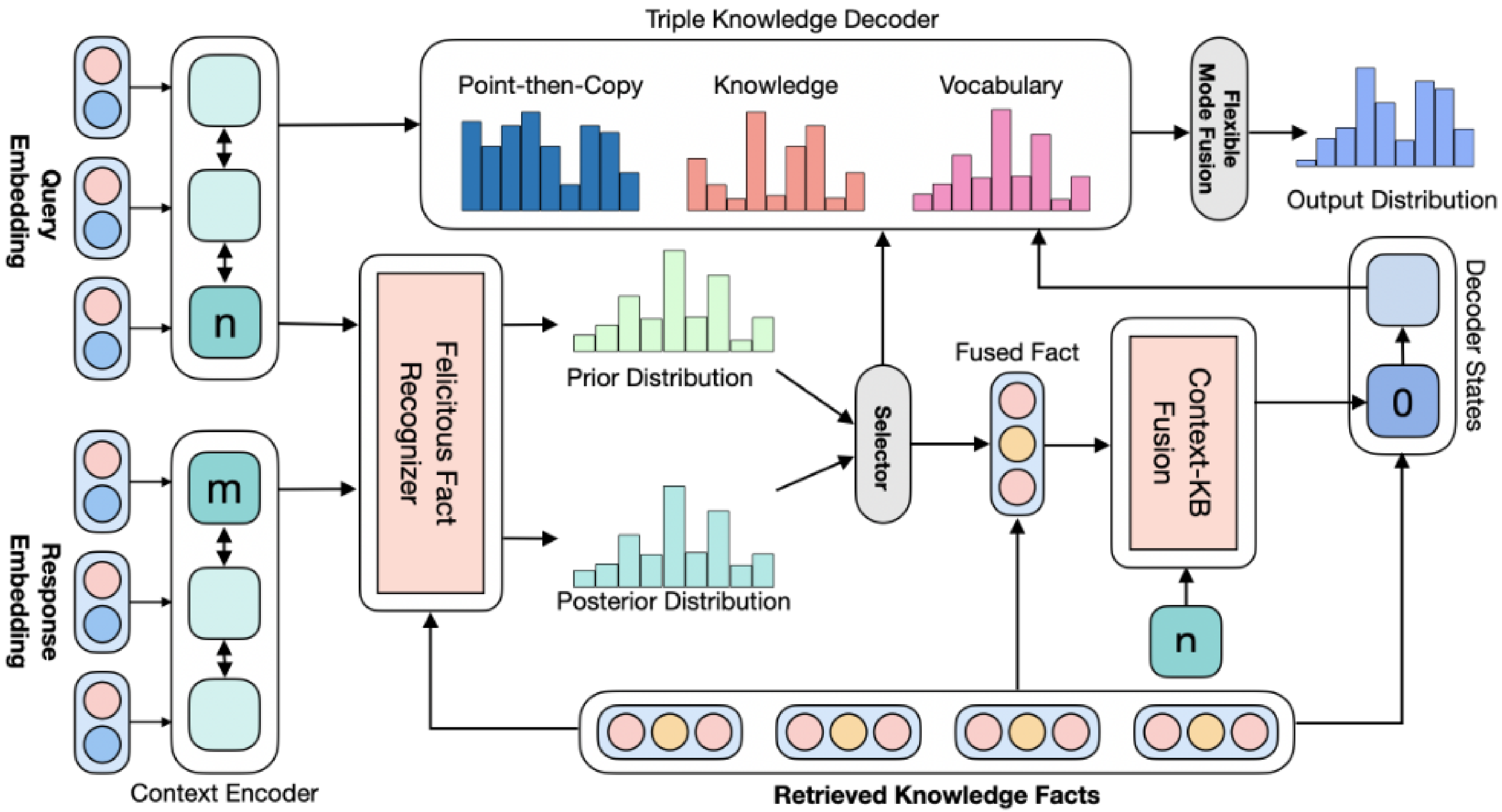


Figure 2: An overview of the proposed approach ConKADI.

Approach

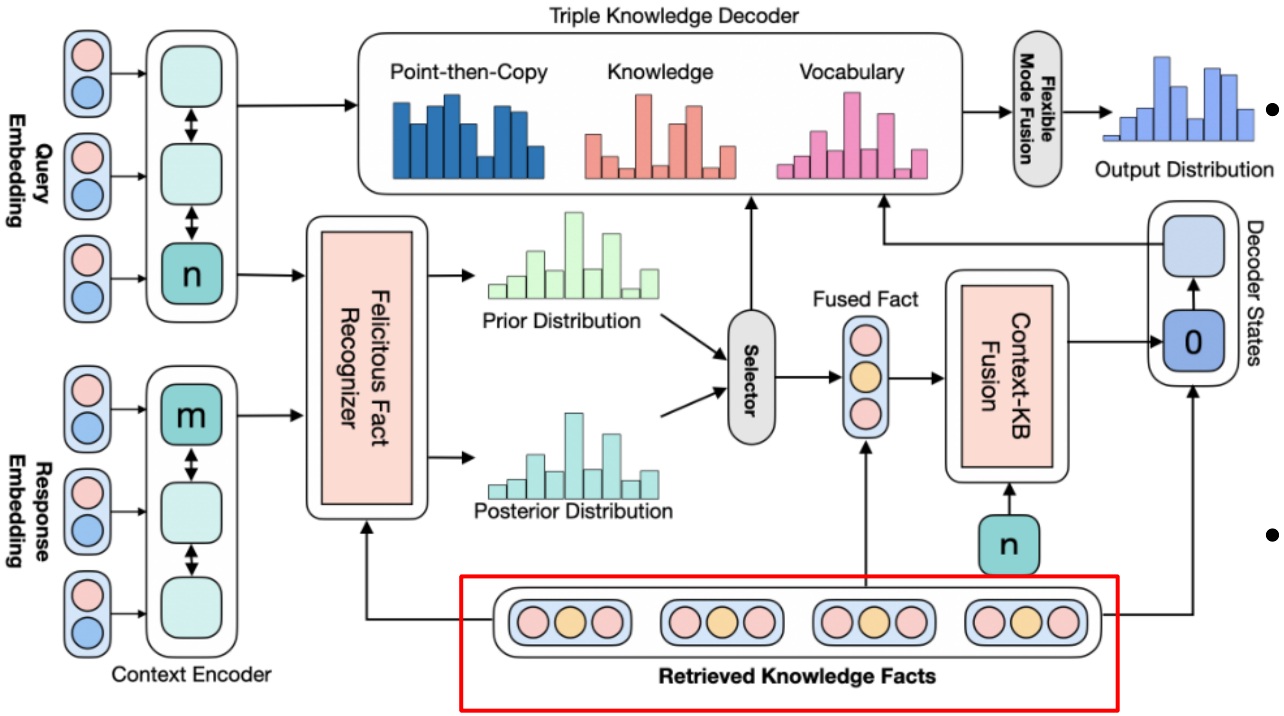


Figure 2: An overview of the proposed approach ConKADI.

- If $x_i \in X$ is recognized as an entity word and can be matched to a vertex e_{src} in the knowledge graph G, then each neighbour $e_{tgt} \in Neighbour(e_{src})$ and the corresponding relation r is retrieved as a candidate fact f.

- If a word can't match any vertex, a special fact f_{NAF} will be used.

Approach

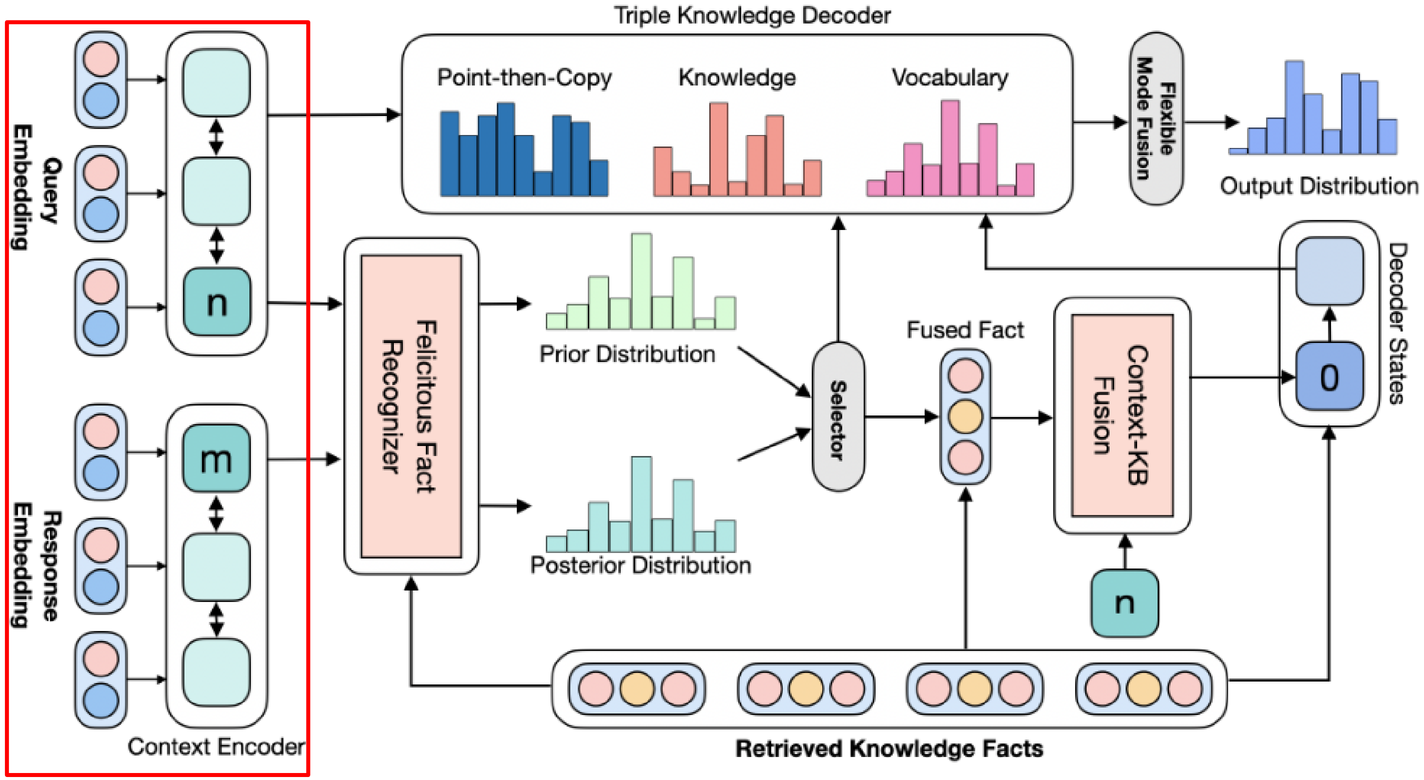


Figure 2: An overview of the proposed approach ConKADI.

$$\mathbf{h}_t^{x} = [\mathbf{h}_t^{fw}; \mathbf{h}_t^{bw}] \in \mathbb{R}^{2d_h \times 1}$$

$$\begin{aligned} \mathbf{h}_t^{fw} &= GRU^{fw}(\mathbf{h}_{t-1}^{fw}, \mathbf{x}_t, \mathbf{e}_{\mathbf{x}_t}) \\ \mathbf{h}_t^{bw} &= GRU^{bw}(\mathbf{h}_{t-1}^{bw}, \mathbf{x}_{n-t+1}, \mathbf{e}_{\mathbf{x}_{n-t+1}}) \end{aligned} \quad (1)$$

$$H^{x/y} = (h_1^{x/y}, \ldots, h_{n/m}^{x/y})$$

Approach

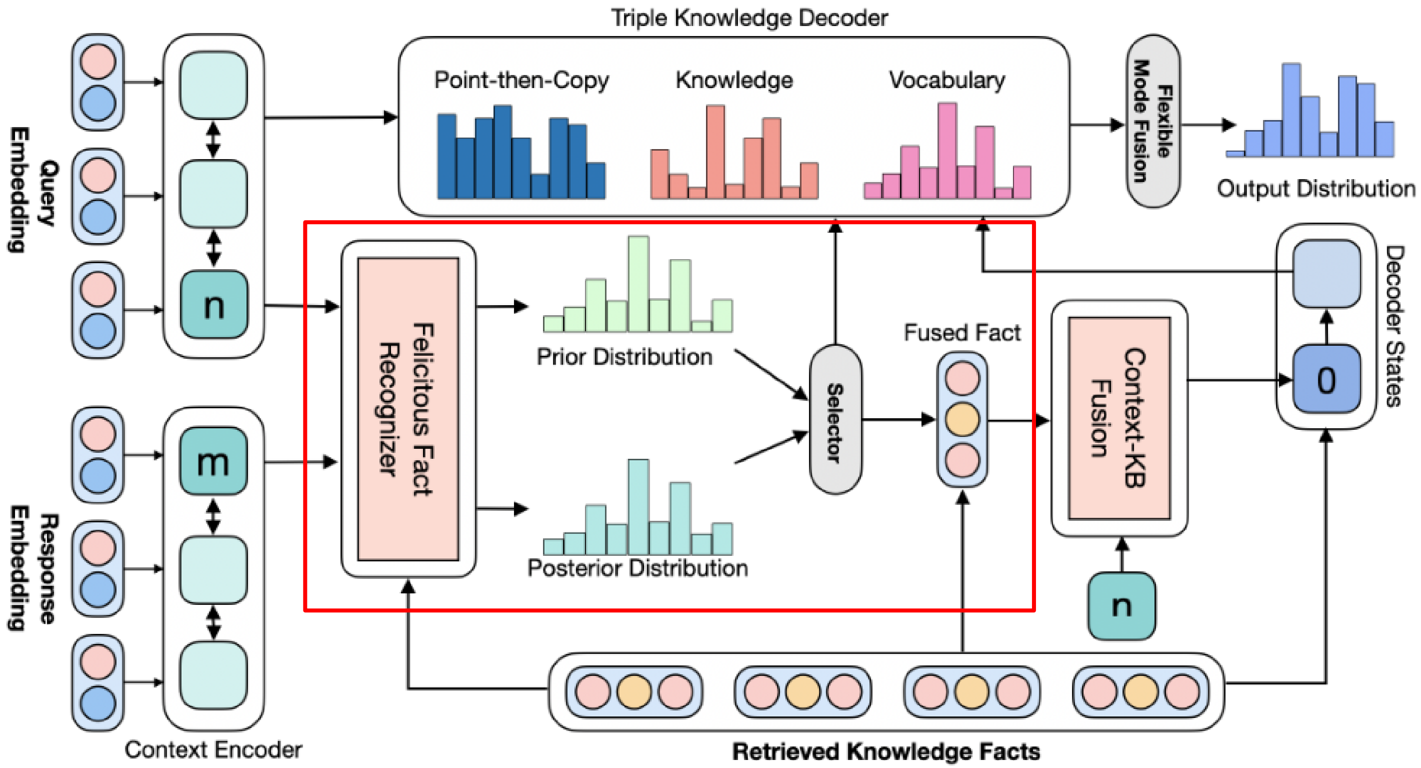


Figure 2: An overview of the proposed approach ConKADI.

$$\mathbf{z}_{\mathbf{post}} = \eta(\varphi(\mathbf{F} \cdot \mathbf{W}_{\mathbf{ft}}) \cdot \varphi([\mathbf{h}_{\mathbf{n}}^{\mathbf{x}\top}; \mathbf{h}_{\mathbf{m}}^{\mathbf{y}\top}] \cdot \mathbf{W}_{\mathbf{post}}))^{\top}$$

$$\mathbf{z}_{\mathbf{prior}} = \eta(\varphi(\mathbf{F} \cdot \mathbf{W}_{\mathbf{ft}}) \cdot \varphi(\mathbf{h}_{\mathbf{n}}^{\mathbf{x}\top} \cdot \mathbf{W}_{\mathbf{prior}}))^{\top} \quad (2)$$

where $\mathbf{F} \in \mathbb{R}^{l \times (d_e + d_r + d_e)}$ is the embedding matrix of the retrieved facts F, $\mathbf{W}_{\mathbf{ft}}$, $\mathbf{W}_{\mathbf{post}}$ and $\mathbf{W}_{\mathbf{prior}}$ are trainable parameters, η is $softmax$ activation , φ is $tanh$ activation.

$$\mathcal{L}_k = \mathbf{KLD}(\mathbf{z}_{\mathbf{post}}, \mathbf{z}_{\mathbf{prior}}) \quad (3)$$

Approach

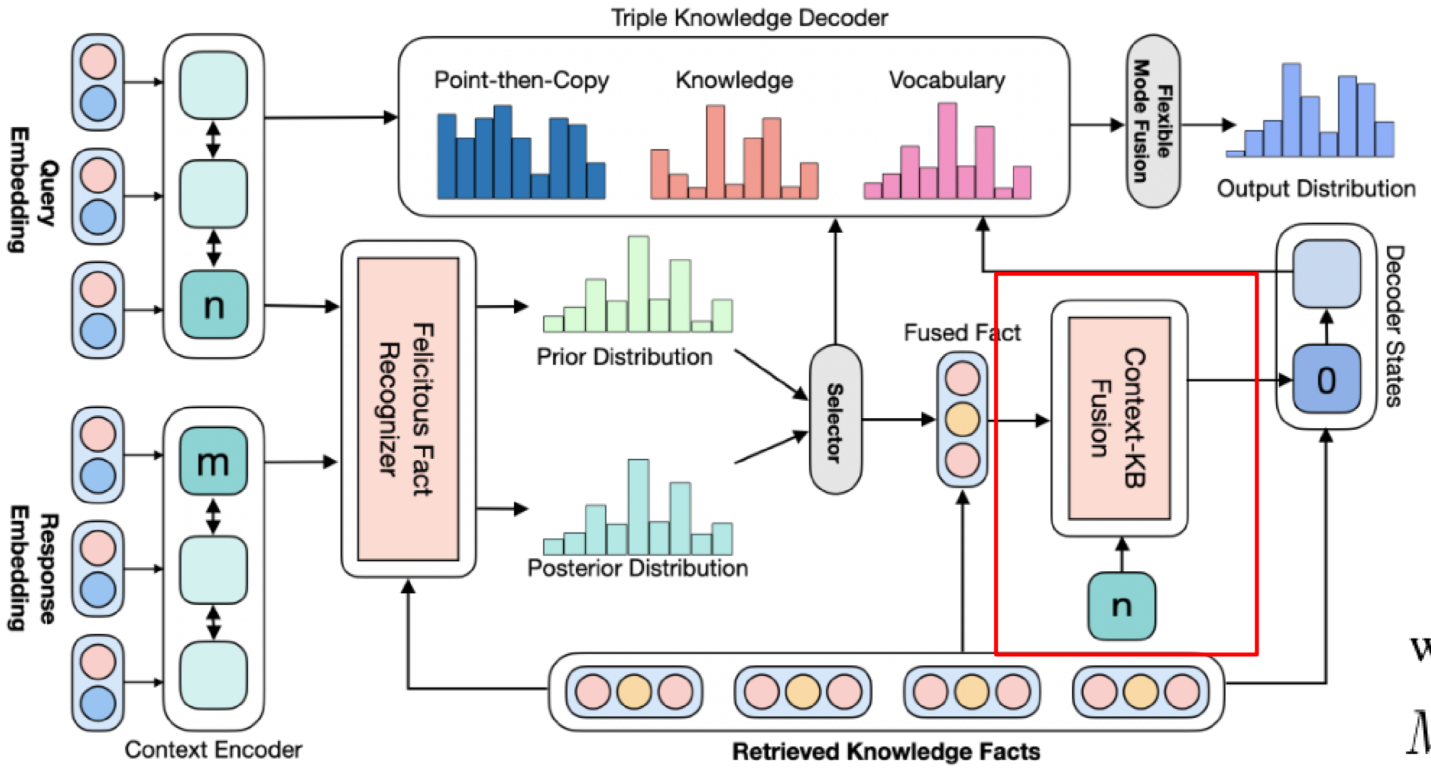


Figure 2: An overview of the proposed approach ConKADI.

$$\mathbf{f_z}^\top = \mathbf{z}^\top \cdot \mathbf{F}$$

$$\mathbf{h_0^y}^\top = tanh([\mathbf{h_n^x}^\top; \mathbf{f_z}^\top] \cdot \mathbf{W_{init}}) \tag{4}$$

$$-\frac{\sum_{y_b \in B} \log p_b(y_b|\mathbf{h_n^x}, \mathbf{f_z}))}{|B|} - \frac{\mathbf{I_f}^\top \cdot \log(\mathbf{z_{post}})}{|\mathbf{I_f}|} \tag{5}$$

where B is the word bag of Y, p_b is a 2-layer MLP_{bow} activated by $softmax$

Approach

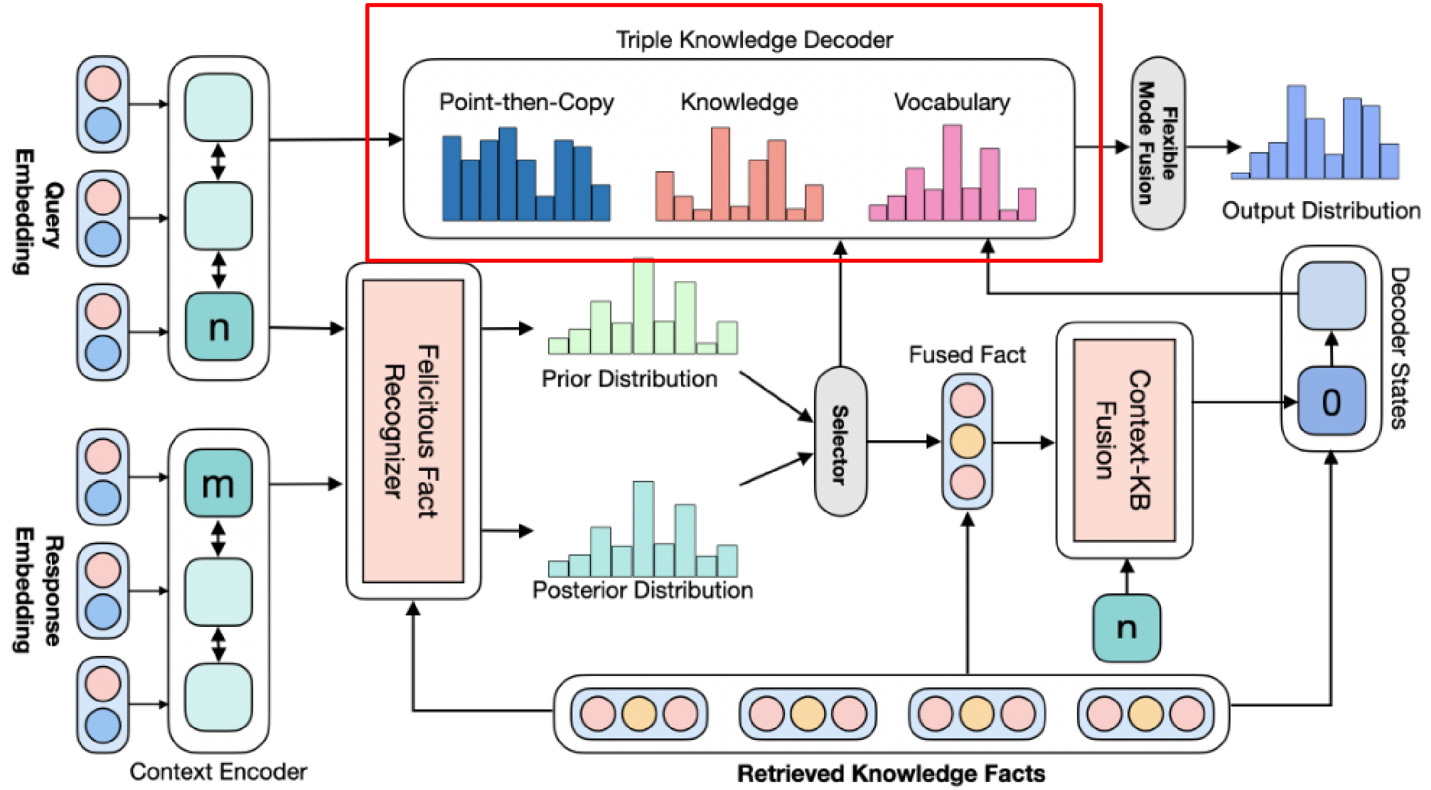


Figure 2: An overview of the proposed approach ConKADI.

$$\mathbf{h}_{\mathbf{t}}^{\mathbf{y}} = g(\mathbf{h}_{\mathbf{t-1}}^{\mathbf{y}}, \mathbf{u}_{\mathbf{t-1}}, \mathbf{c}_{\mathbf{t-1}}) \tag{6}$$

where $\mathbf{u}_{\mathbf{t-1}}{}^{\top} = [\mathbf{y}_{\mathbf{t-1}}^{\top}; \mathbf{e}_{\mathbf{y}_{\mathbf{t-1}}}^{\top}; \mathbf{h}_{\mathbf{y}_{\mathbf{t-1}}}^{\mathbf{x}}{}^{\top}]$, and $\mathbf{y}_{\mathbf{t-1}}$, $\mathbf{e}_{\mathbf{y}_{\mathbf{t-1}}}$, $\mathbf{h}_{\mathbf{y}_{\mathbf{t-1}}}^{\mathbf{x}}$ are the word embedding, the entity embedding and the pointed-then-copied source state of the last predicted token y_{t-1}, respectively; and $\mathbf{c}_{\mathbf{t-1}}$ is the Attention [2].

Approach

Vocabulary Words: The probability distribution $p_{w,t} \in \mathbb{R}^{|V| \times 1}$ over the V is given by:

$$p_{w,t}{}^{\top} = \eta(elu([\mathbf{h_t^y}{}^{\top}; \mathbf{u_{t-1}}{}^{\top}; \mathbf{c_t}{}^{\top}] \cdot \mathbf{W_{v1}}) \cdot \mathbf{W_{v2}}) \tag{7}$$

Copied Words: The Decoder can further point out a word x from X, and then copies the x . The corresponding probability distribution $p_{c,t} \in \mathbb{R}^{n \times 1}$ over the query message X is calculated as:

$$\begin{aligned} p_{c,t} &= \eta(\varphi(\mathbf{H^x} \cdot \mathbf{W_{cs}}) \cdot \varphi(\mathbf{u_t^c}{}^{\top} \cdot \mathbf{W_{ct}})^{\top}) \\ \mathbf{u_t^c}{}^{\top} &= [\mathbf{h_t^y}{}^{\top}; \mathbf{u_{t-1}}{}^{\top}; \mathbf{c_t}{}^{\top}] \end{aligned} \tag{9}$$

Knowledgeable Entity Words: An entity word can be generated by extracting the target entity of the best-matched fact f at each time step. The corresponding probability distribution $p_{k,t} \in \mathbb{R}^{l \times 1}$ over the F is calculated as:

$$\begin{aligned} \mathbf{z_{d,t}} &= \eta(\varphi(\mathbf{F} \cdot \mathbf{W_{fd}}) \cdot \varphi([\mathbf{h_t^y}{}^{\top}; \mathbf{u_{t-1}}{}^{\top}] \cdot \mathbf{W_d})^{\top}) \\ \gamma_t &= sigmoid([\mathbf{h_t^y}{}^{\top}; \mathbf{u_t}{}^{\top}; \mathbf{c_t}{}^{\top}] \cdot \mathbf{W_{gate}}) \in \mathbb{R}^1 \\ p_{k,t} &= \gamma_t \times \mathbf{z} + (1.0 - \gamma_t) \times \mathbf{z_d} \end{aligned} \tag{8}$$

Approach

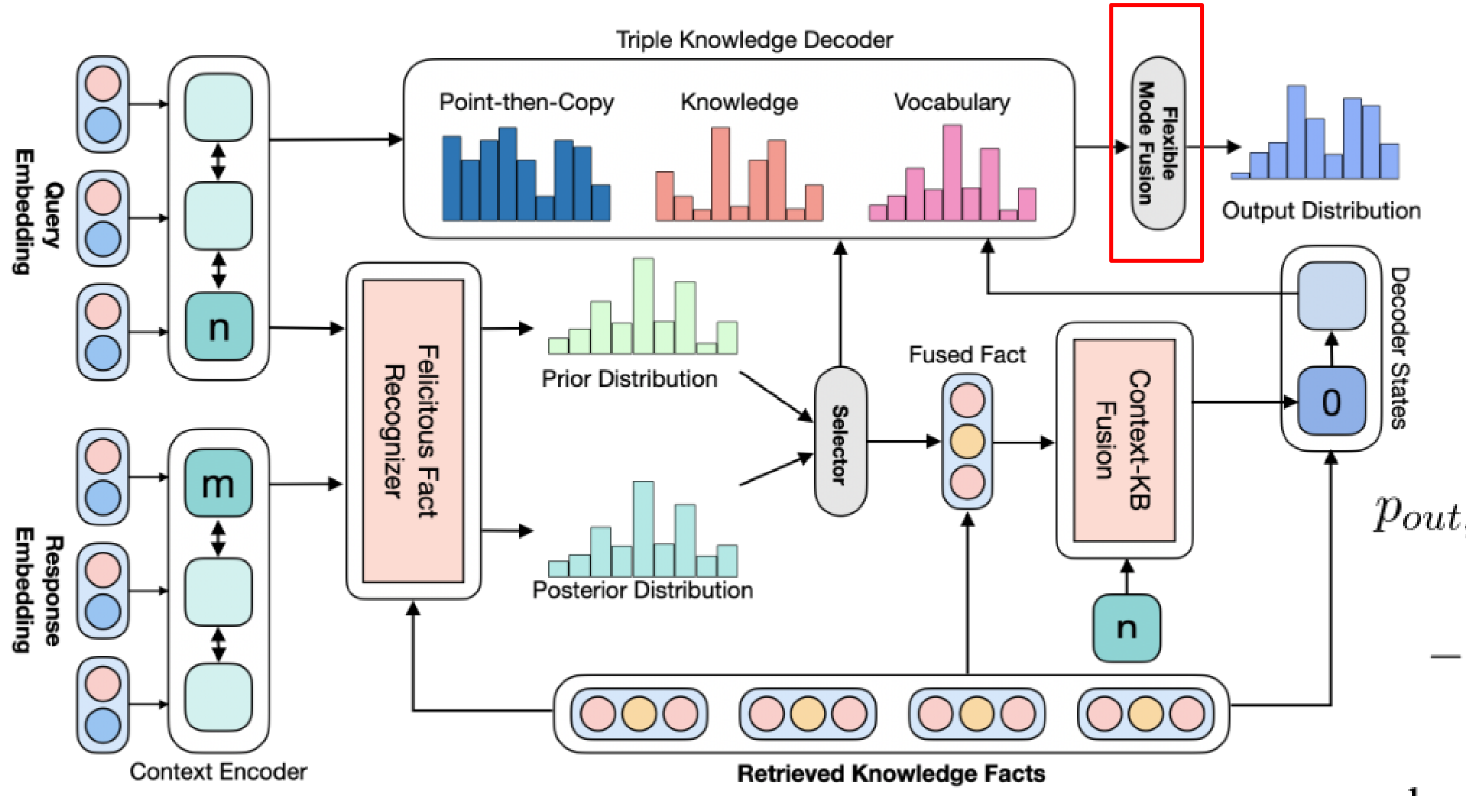


Figure 2: An overview of the proposed approach ConKADI.

$$MF(\mathbf{h}_t^y, \mathbf{u}_{t-1}, \mathbf{c}_t),$$

$$(\gamma_{w,t}, \gamma_{k,t}, \gamma_{c,t})$$

$$p_{out,t} = \gamma_{w,t} \times p_{w,t} + \gamma_{k,t} \times p_{k,t} + \gamma_{c,t} \times p_{c,t} \quad (10)$$

$$-\sum_t \lambda_t \log p_{out,t}(y_t | y_{t-1:1}, X, F) + \frac{\mathcal{L}_m}{2} \quad (11)$$

where λ_t is a normalization term to penalize the out-of-vocabulary words, $\lambda_t = \frac{1}{\#(unk \in Y)}$ [3] if y_t is an unk, otherwise $\lambda_t = 1$.

Approach

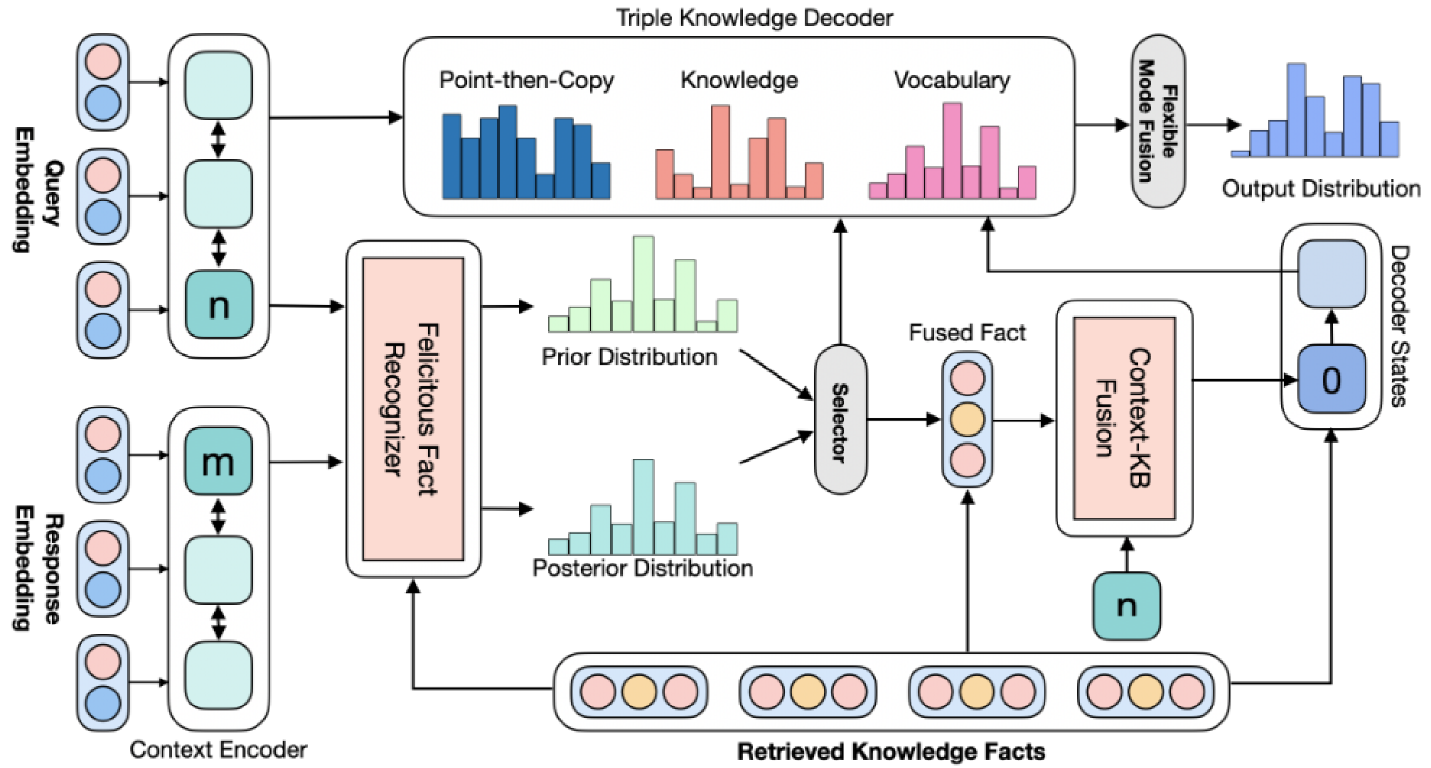


Figure 2: An overview of the proposed approach ConKADI.

Training Objective: Finally, the ConKADI can be trained by minimizing the following objective:

$$\mathcal{L} = \mathcal{L}_n + \mathcal{L}_k + \mathcal{L}_f \tag{12}$$

Experiments

	Reddit	Weibo
#Train	1,352,961	1,019,908
#Dev/#Test	40,000	56,661
#Vocab	30,000	50,000
Batch Size	100	50
#Entity/#Relation	21,471/44	27,189/26
#Fact	149,803	696,466

Table 1: The statistics of two datasets.

Experiments

	Entity Score			Embedding		Overlap (%)		Diversity (%)		Informativeness	R-Score	
Metric	E_{match}	E_{use}	E_{recall}	Emb_{avg}	Emb_{ex}	BLEU-2	BLEU-3	Distinct-1	Distinct-2	Entropy	R_a	R_g
Chinese Weibo												
S2S	0.33	0.58	13%	0.770	0.500	2.24	0.80	0.21	1.04	6.09	0.78	0.75
ATS2S	0.33	0.59	12%	0.767	0.513	1.93	0.69	0.27	1.23	5.99	0.77	0.75
$ATS2S_{MMI}$	0.40	0.74	15%	0.773	0.528	4.01	**1.61**	0.75	3.91	7.49	1.24	1.21
$ATS2S_{DD_{1.5}}$	0.35	0.62	13%	0.780	0.542	2.14	0.86	1.03	4.86	7.62	1.16	1.10
Copy	0.33	0.68	13%	0.786	0.501	2.28	0.84	0.59	2.18	6.13	0.92	0.91
GenDS	0.75	0.84	26%	0.789	0.524	2.09	0.73	0.30	1.66	5.89	0.94	0.91
CCM	0.99	1.09	28%	0.786	0.544	3.26	1.20	0.48	2.59	6.16	1.18	1.15
AVG	0.49	0.74	17%	0.779	0.522	2.56	0.96	0.52	2.50	6.48	1.00	1.00
ConKADI	1.48	**2.08**	**38%**	**0.846**	**0.577**	**5.06**	1.59	**3.26**	**23.93**	**9.04**	**2.98**	**2.24**
$ConKADI_{-cp}$	**1.60**	1.89	**38%**	0.833	0.567	5.00	1.52	2.34	18.29	8.75	2.55	2.08
English Reddit												
S2S	0.41	0.52	4%	0.868	0.837	4.81	1.89	0.38	1.77	7.59	0.82	0.78
ATS2S	0.44	0.59	5%	0.863	0.831	4.50	1.81	0.82	3.44	7.62	0.92	0.91
$ATS2S_{MMI}$	0.45	0.65	6%	0.858	0.825	4.95	2.13	0.75	3.22	7.62	0.95	0.94
$ATS2S_{DD_{0.3}}$	0.31	0.43	4%	0.830	0.784	1.70	0.75	0.97	3.50	7.47	0.77	0.72
Copy	0.13	0.67	9%	0.868	0.841	**5.43**	**2.26**	1.73	8.33	7.87	1.19	1.09
GenDS	1.13	1.26	13%	**0.876**	0.851	4.68	1.79	0.74	3.97	7.73	1.14	1.10
CCM	1.08	1.33	11%	0.871	0.841	5.18	2.01	1.05	5.29	7.73	1.21	1.18
AVG	0.55	0.77	7%	0.860	0.829	4.40	1.79	0.94	4.32	7.69	1.00	1.00
ConKADI	1.24	**1.98**	**14%**	0.867	0.852	3.53	1.27	**2.77**	**18.78**	8.50	**1.76**	**1.46**
$ConKADI_{-cp}$	**1.41**	1.73	13%	0.865	**0.855**	3.09	1.07	2.29	16.70	**8.68**	1.63	1.37

Table 2: Objective Experimental Results. The ablation $ConKADI_{-cp}$ removes the ability to copy source words.

Experiments

ConKADI	Appropriateness			Informativeness		
vs.	Win	Tie	Lose	Win	Tie	Lose
ATS2S	71.3%	11.0%	17.7 %	87.3%	6.9%	5.8%
$ATS2S_{MMI}$	59.3%	9.2%	31.5%	82.5%	7.3%	10.2%
Copy	71.7%	8.8%	19.5%	89.7%	3.8%	6.5%
GenDS	87.2%	7.3%	5.5%	93.8%	2.3%	3.5%
CCM	83.8%	6.9%	9.3%	93.0%	3.5%	3.5%

Table 3: Human annotation results on the Chinese Weibo. ConKADI significantly (sign test, p-value < 0.005, ties are removed) outperforms other baselines in terms of both appropriateness and informativeness.

Experiments

#	Settings	E_{use}	Distinct-2	Entropy	R_g
#1	Copy+GlFact+CKF+$\mathcal{L}_f$	2.08	23.93	9.04	2.24
#2	Base+GlFact+CKF+$\mathcal{L}_f$	1.89	18.29	8.75	2.02
#3	Copy+GlFact+CKF	1.79	18.18	8.73	2.08
#4	Base+GlFact+CKF	1.92	17.38	8.87	2.01
#5	Base+CKF	1.87	15.72	8.66	1.96
#6	Base+GlFact	1.05	2.90	6.31	1.10
#7	Base	1.06	2.50	6.46	1.10

Table 5: Ablation study on the Chinese Weibo.

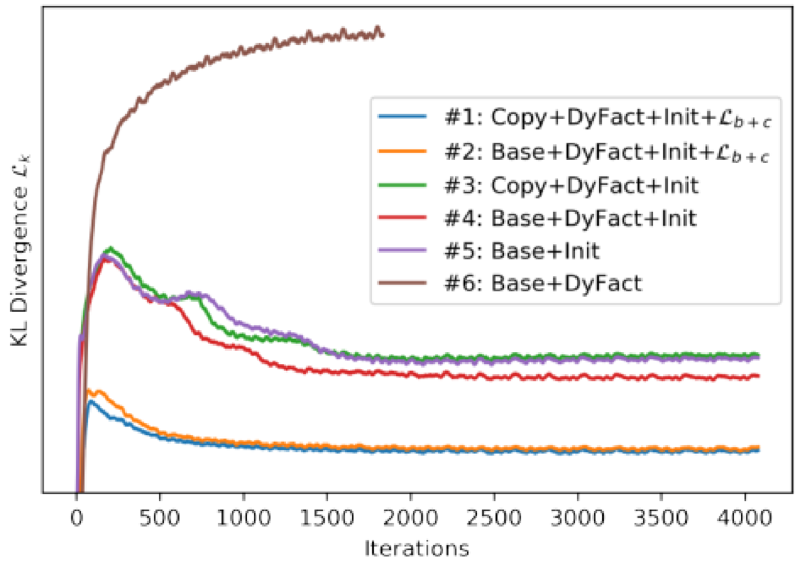


Figure 3: The Kullback–Leibler Divergence between the between the $\mathbf{z}_{\mathbf{prior}}$ and $\mathbf{z}_{\mathbf{post}}$ on Chinese Weibo against the training iteration number.

Experiments

Query	#1:My cat likes bananas and bread.	#2:Yeah , but what website?	#3:我会唱霉霉的歌。 I can sing the song of Taylor Swift
ATS2S	I'm a banana and I don't know what you're talking about.	I'm not sure. I'm just curious.	我也是，我唱的是 **unk**。 Me too. I'm singing **unk**。
ATS2S_{MMI}	Do you have a cat?	It's a site site.	你唱的是哪种歌? What kind of song are you singing?
Copy	I'm a cat.	I'm not sure what site you're talking about.	我也是，我也是，我也是，我也喜欢。 Me too, me too, me too, I like it。
GenDS	I'm a banana.	I'm not sure , but I'm not sure if it's a link to the original post. I'm not sure what the site is.	你可以听我唱的唱。 You can listen to my singing singing.
CCM	I'm a banana and I love my cat.	I'm not sure, I just got a link to the site.	我也是,我也喜欢,听着歌着歌听着歌听着歌 Me too. I like it, too. Listening to songs. Listening to songs. Listening to songs
ConKADI	And your cat is the best.	Looks like Youtube, the site is blocked.	我听了,他的音乐好听。 I heard it. His music is good.

Table 4: Case Study: #1 #2 are sampled from the English Reddit, #3 is sampled from the Chinese Weibo.

Thanks