



Some Recent Advance on Edit Based Generation Models

Jiachen Du^{1,2}

1. Harbin Institute of Technology, Shenzhen
2. Tencent AI Lab, NLP Center

2019.05.28



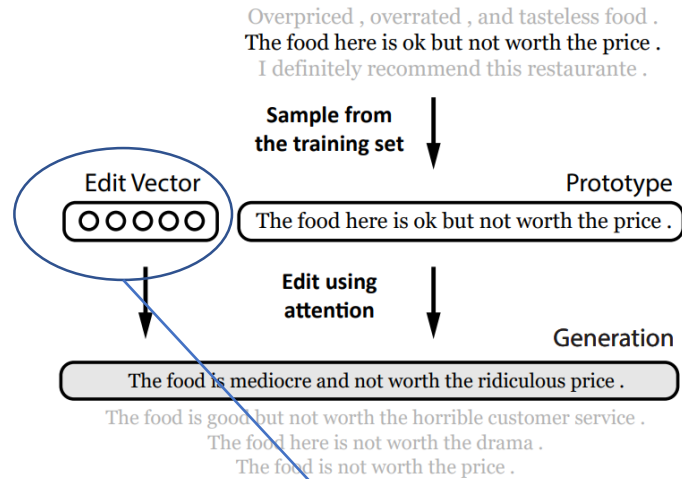
Overview

- Review of Edit Based Generation Models
- LEARNING TO REPRESENT EDITS (ICLR 2019)
- Text Infilling (arXiv)
- TIGS: An Inference Algorithm for Text Infilling with Gradient Search (arXiv)

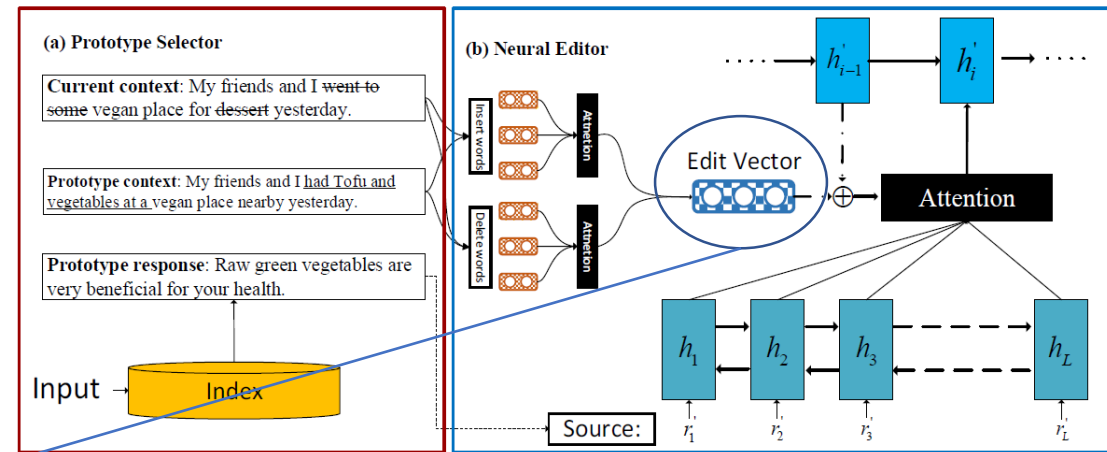


Review of Edit Based Generation Models

Generating Sentences by Editing Prototypes
Guu et al. 2018 (TACL)



Response Generation by Context-aware Prototype Editing
Wu et al. 2019 (AAAI)



$$e = \sum_{w \in I} \beta_w \Psi(w) \oplus \sum_{w' \in D} \gamma_{w'} \Psi(w')$$

Insertion Words Deletion Words

The core idea is to use distributed vector to represent the edit, and leverage the edit vector to

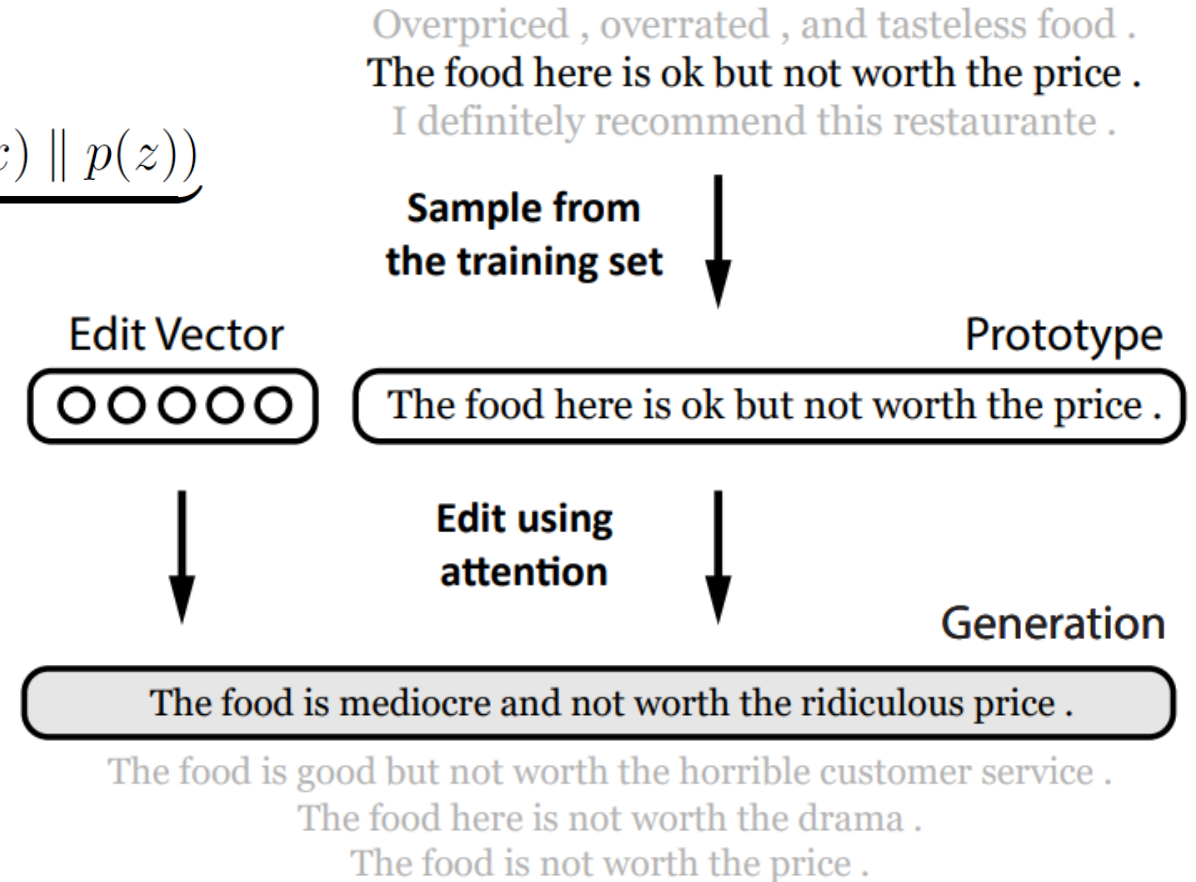


Review of Edit Based Generation Models

$$\log p(x|x') \geq \underbrace{\mathbb{E}_{z \sim q(z|x',x)} [\log p_{\text{edit}}(x | x', z)]}_{\mathcal{L}_{\text{gen}}} - \underbrace{\text{KL}(q(z | x', x) \| p(z))}_{\mathcal{L}_{\text{KL}}}$$

LSTM decoder concatenating z to the input of the decoder at each time step

a sample from q is obtained by adding von-Mises Fisher (vMF) noise, and we perturb the magnitude of e by adding uniform noise.

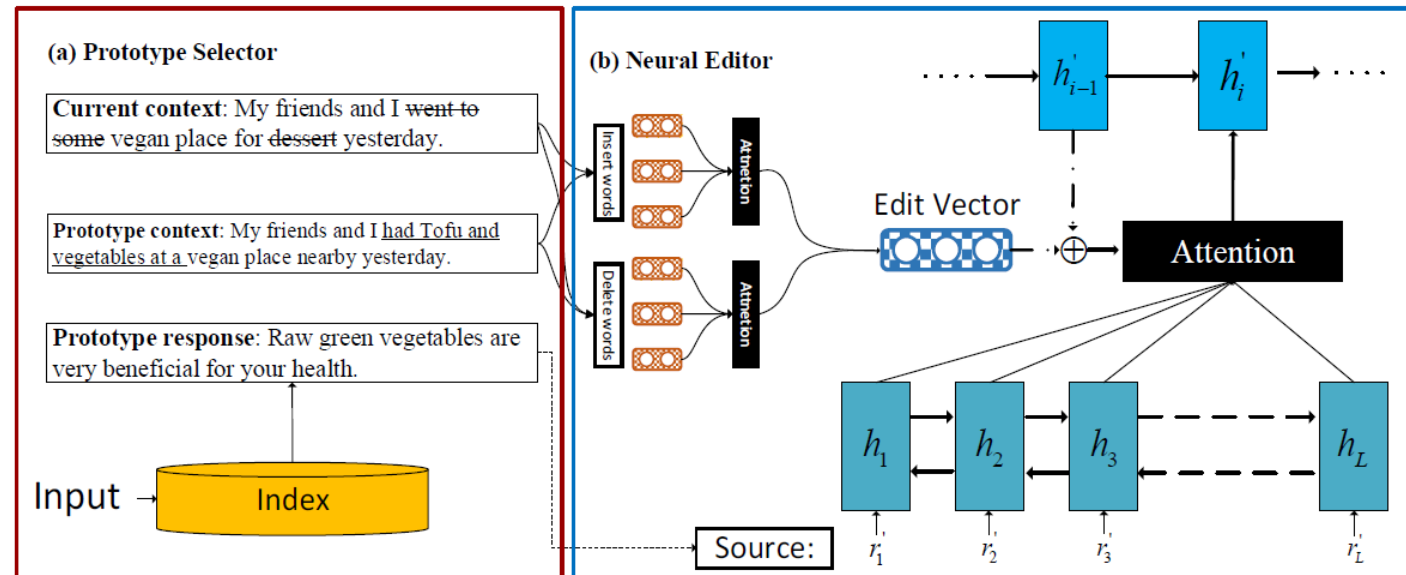




Review of Edit Based Generation Models

Response Generation by Context-aware Prototype Editing
Wu et al. 2019 (AAAI)

The edit vector is generated by summing the insertion and deletion words in context.



Some Thinking about Edit Based Generation

- Most works apply the edit changes in vector space by concatenating the edit vector to decoder's input. It is not clear how the model apply the edit change in token level.
- The edit vector only takes consideration of words deleted or inserted, but ignore the order of these changing words.
- Current edit based generation model focus on sequential editing. Applying editing to the structural objects (Tree, Graph) may be a promising idea ?

LEARNING TO REPRESENT EDITS

Pengcheng Yin*, Graham Neubig

Language Technology Institute

Carnegie Mellon University

Pittsburgh, PA 15213, USA

{pcyin, gneubig}@cs.cmu.edu

Miltiadis Allamanis, Marc Brockschmidt, Alexander L. Gaunt

Microsoft Research

Cambridge, CB1 2FB, United Kingdom

{miallama, mabrocks, algaunt}@microsoft.com

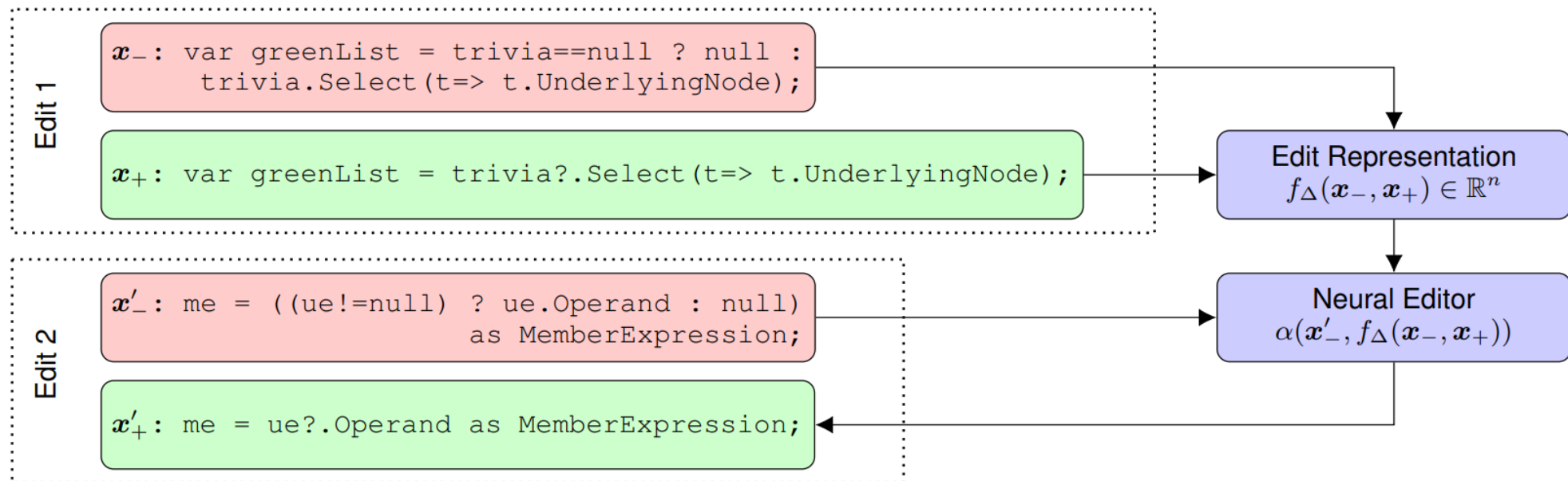


Introduction

Edit Representation: $f_{\Delta}(x_{-}, x_{+})$ maps the editing difference between original text x_{-} and edited version x_{+} .

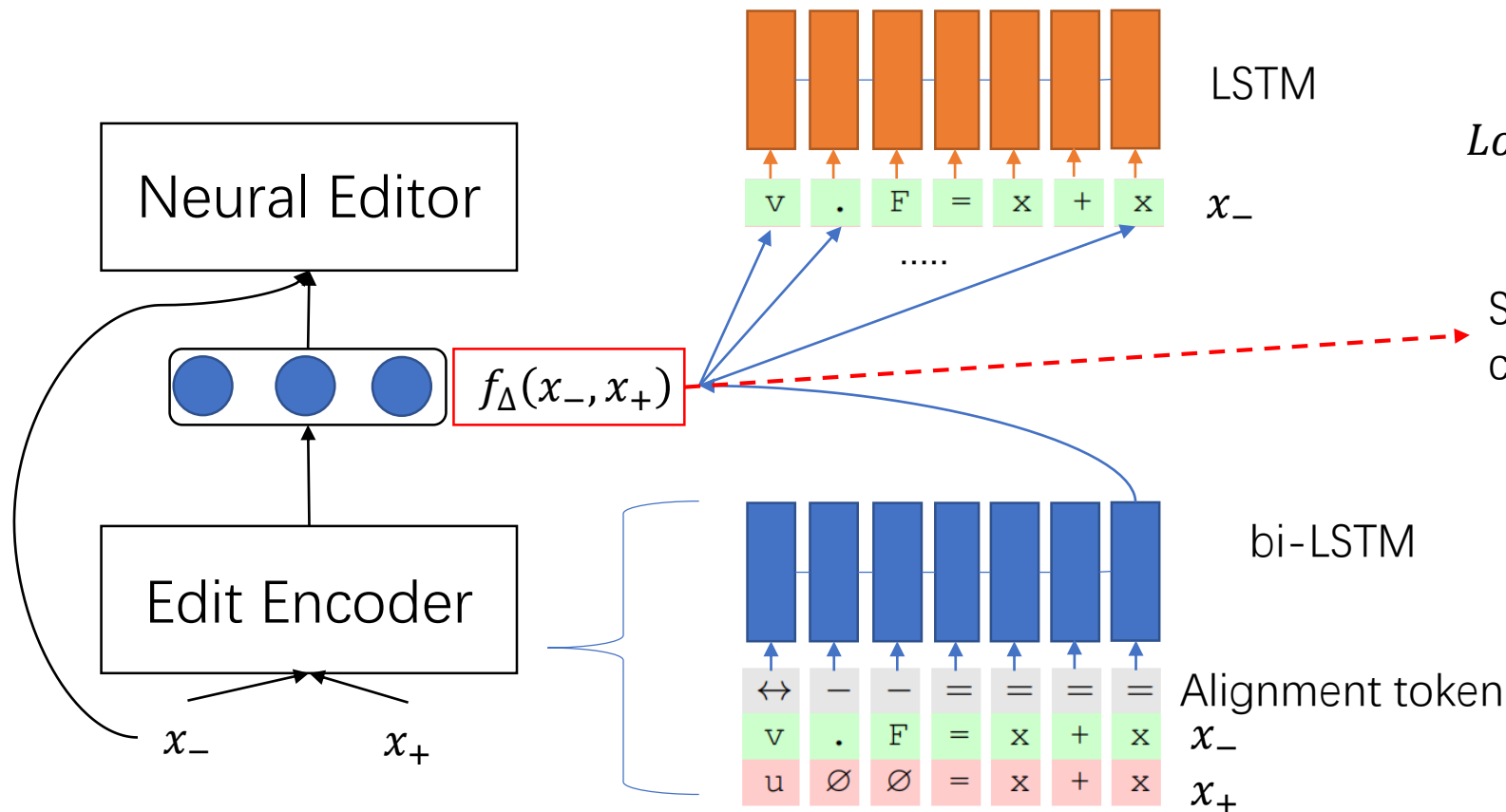
Neural Editor: $\alpha(x'_{-}, f_{\Delta}(x_{-}, x_{+}))$ reconstructs edited text from x_{-} and edit representation $f_{\Delta}(x_{-}, x_{+})$.

- ◆ **Distributed edit representation** explicitly models the discrete edit operations to have the property that semantically similar edits have nearby representations in continuous space.





Models – Sequential Editing

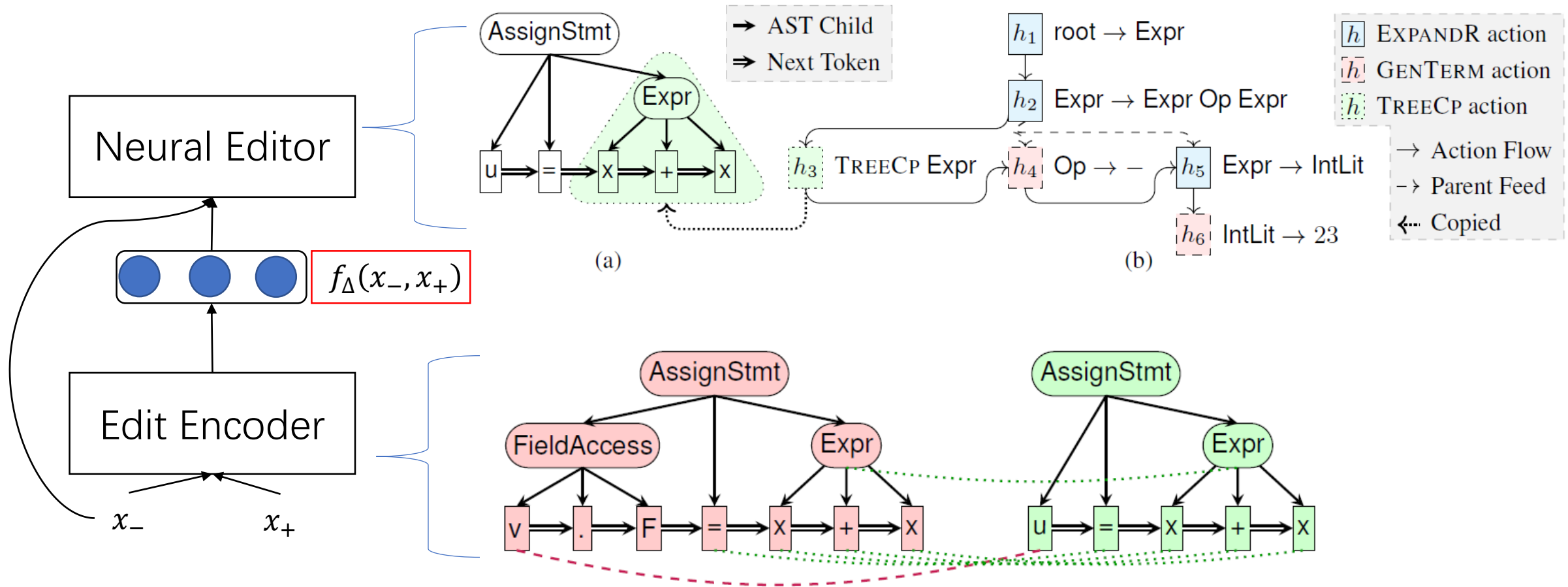


$$Loss = \frac{1}{N} \sum_i -\log P_{\alpha}(x_+ | x_-, f_{\Delta}(x_-, x_+))$$

Semantic similar changes are clustered in the edit vector space.



Models – Graph to Tree Editing

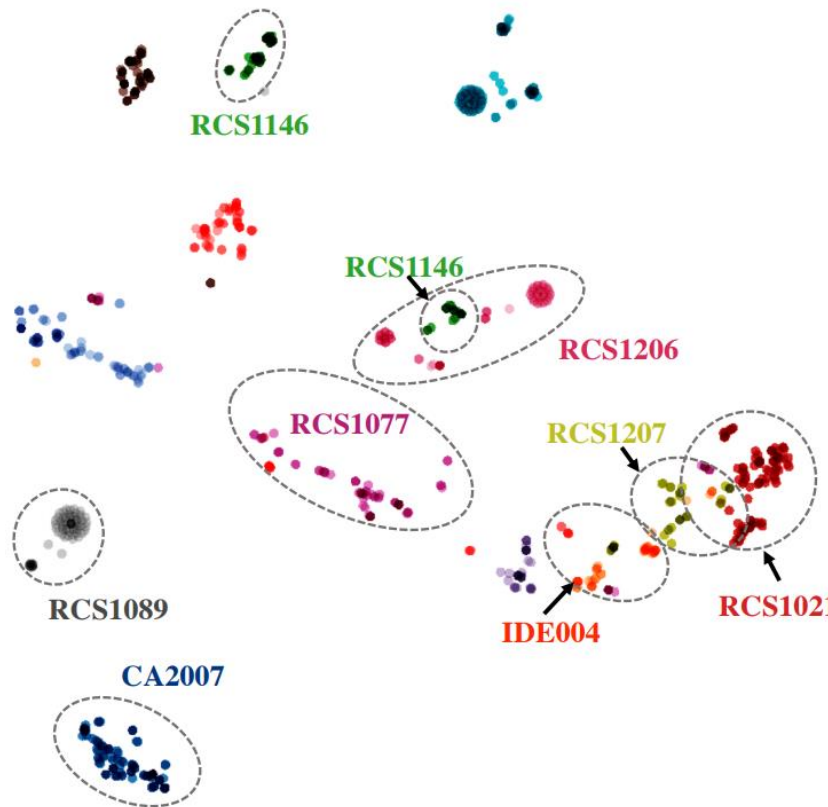


Experiments

- Natural Language Edits: WikiatomicEdit (Faruqui et al., 2018 EMNLP) : Contains Wikipedia dump document and corresponding editing history. They sampled 1040K edits from the English split the samples into 1000K/20K/20K train-valid-test sets.
- Source Code Edits: They clone a set of 54 C# projects on GitHub. They selected all changes in the projects that are no more than 3 lines long and whose surrounding 3 lines of code before. They split the dataset into 91,372 / 10,176 / 10,176 samples as train/valid/test sets.



Visualization of Edit Vector



Fixer ID	Description	Num. Edits	Example
CA2007	apply <code>.ConfigureAwait(false)</code> to await statements	1051	x_- : <code>await Console.WriteLineAsync()</code> x_+ : <code>await Console.WriteLineAsync().ConfigureAwait(false)</code>
IDE0004	Cast is redundant	53	x_- : <code>var x = 1; var b = (int)x;</code> x_+ : <code>var x = 1; var b = x;</code>
RCS1015	Use nameof operator	35	x_- : <code>Exception("parameter");</code> x_+ : <code>Exception(nameof(parameter));</code>
RCS1021	Simplify lambda expression	411	x_- : <code>var x = items.Select(f => { return f.ToString(); });</code> x_+ : <code>var x = items.Select(f => f.ToString());</code>
RCS1032	Remove redundant parentheses	24	x_- : <code>if ((x)) {}</code> x_+ : <code>if (x) {}</code>
RCS1058	Use compound assignment	43	x_- : <code>i = i + 2;</code> x_+ : <code>i += 2;</code>
RCS1077	Optimize LINQ method call	200	x_- : <code>items.Where(f => Foo(f)).Any();</code> x_+ : <code>items.Any(f => Foo(f));</code>



Nearest Neighbors of Edited Examples

Five nearest neighbors of 200 randomly sampled seed edits from our training set, using both our trained sequence-to-sequence editing model with sequential edit encoder, as well as a simple bag-of-words baseline based on TF-IDF scores.

Example 2 she , along with her follow artist carolyn mase studied with ► *impressionist landscape painter* ◀ john henry twachtman at the art students league of new york .

- | | |
|---|---|
| NN-1 his brother was draughtsman william daniell and his uncle was ► <i>landscape painter</i> ◀ thomas daniell . | the first painting was a portrait of a young girl , emerantia van beresteyn , the sister of ► <i>the landscape painter</i> ◀ nicolaes van beresteyn , the later founder of half of this hofje . |
| NN-2 william james linton (december 7 , 1812 - december 29 , 1897) was an english - born american wood engraver , ► <i>landscape painter</i> , ◀ political reformer and author of memoirs , novels , poetry and non-fiction . | he was the club ' s top scorer with 22 goals in all competitions , one more than ► <i>senegalese striker</i> ◀ lamine diarra , who left the club at the end of the season . |
| NN-3 early on , hopper modeled his style after chase and french ► <i>impressionist</i> ◀ masters douard manet and edgar degas . | caforio ” aggressively attacked ” his opponent , ► <i>republican incumbent</i> ◀ steve knight , for his delayed response to the leak . |
-



Nearest Neighbors of Edited Examples

each fixer category F of semantically similar edits, They randomly select a seed edit from one category of bug fixes, and use its edit representation $f_{\Delta}(x_{-}, x_{+})$ to predict the updated code for all examples

Model	Acc.(%)	Acc.* (%)	Recall@5(%)	Recall@5*(%)
Seq2Seq – Seq Edit Encoder	38.35	77.67	41.50	83.84
Graph2Tree – Seq Edit Encoder	49.21	77.30	51.93	81.77
<u>Baselines</u> (no edit encoding)				
Seq2Seq w/o Edit Encoder	7.07	—	14.29	—
Graph2Tree w/o Edit Encoder	8.81	—	11.90	—



Summary

- An **explainable** edit representation is proposed and some experiments are conducted to visualize the edit vector.
- Instead of representing the change by Bag-of-Word, the editing token sequence is used to model the editing procedure.
- Interesting downstream application ...



Some Inspiration

- It is hard to directly apply this model to response generation, since the explicit editing pairs is missing. However, maybe it can be applied it to knowledge enhanced dialogue, i.e.

Question : <TOPIC> 100 metres <EOS> <MESSAGE> usain bolt is the best 100 m sprinter to ever live <EOS>

Answer : <RESPONSE> yeah , that 's true . he is the first person to hold both the 100 meter and 200 meters records . <EOS>

Generated: he is the first person to hold the the 100 meter 200

Checked: <CHECKED_SENT> he is the first person to hold both the 100 metres and 200 metres world records since fully automatic time became mandatory . <EOS>

Text Infilling

Wanrong Zhu¹, Zhiting Hu^{2,3}, Eric P. Xing^{2,3}
Peking University¹, Carnegie Mellon University², Petuum Inc.³



Introduction

Given a text template where portions of a body of text are deleted or redacted, we want to fill in the blanks properly to produce complete, semantically coherent and meaningful text.

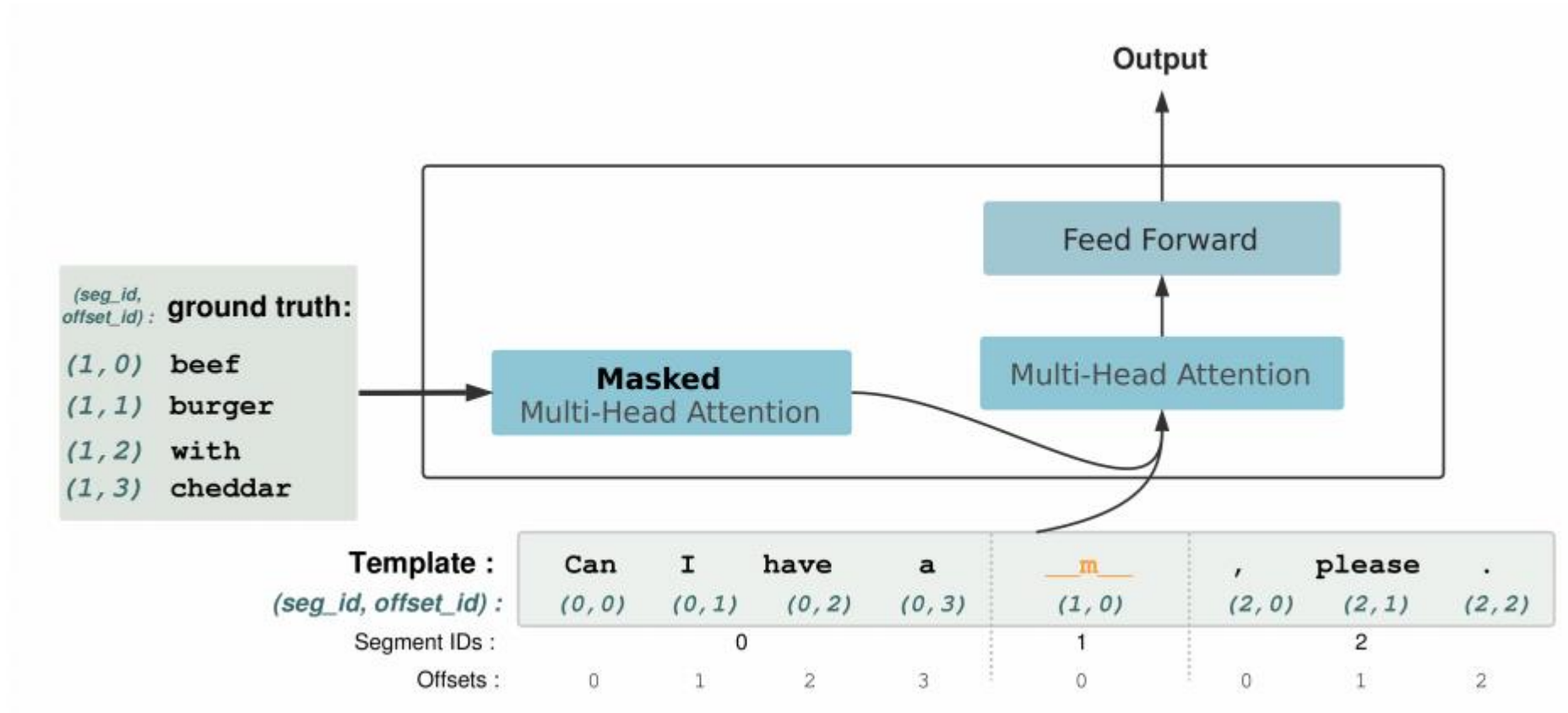
Different

Template : ^m have a ^m , please .

Filled Text : Can I have a beef burger with cheddar , please .



Model





Evaluation

#Blanks	Metric	Template	Seq2Seq	GAN	Self-attn
1	BLEU	63.916	69.097	68.470	71.104
	Perplexity	-	107.480	144.127	38.304
	Human Eval	-	1.950	1.775	2.275
2	BLEU	42.233	64.174	64.337	65.914
	Perplexity	-	43.044	36.704	21.028
	Human Eval	-	1.838	1.975	2.188
#Blanks	Metric	Template	Seq2Seq	GAN	Self-attn
1	BLEU	44.369	48.865	48.861	51.55
	Perplexity	-	244.862	287.415	43.688
	Human Eval	-	1.725	1.863	2.412
2	BLEU	32.498	42.613	42.535	44.418
	Perplexity	-	99.421	107.558	32.397
	Human Eval	-	1.875	1.913	2.238

Template	i live <u>__m__</u> and i was <u>__m__</u> chinese food .
Golden	i live <u>right down the street</u> and i was <u>craving some good</u> chinese food .
Seq2Seq	i live <u>at a ten times</u> and i was <u>at appreciated by</u> chinese food .
GAN	i live <u>right of the app</u> and i was <u>looking for</u> chinese food .
Self-attn	i live <u>in the neighborhood area</u> and i was <u>impressed with the</u> chinese food .
Template	<u>__m__</u> sound <u>__m__</u> be <u>__m__</u>
Golden	<u>if you bear it without letting a sound</u> <u>escape you , i shall be</u> <u>free</u>
Seq2Seq	<u>and</u> sound <u>the</u> be <u>and the little , and the little , and the</u>
GAN	<u>and</u> sound <u>the</u> be <u>and the , and and</u>
Self-attn	<u>the</u> sound <u>said , i will</u> be <u>the king</u>
Template	<u>__m__</u> Toronto_Raptors <u>__m__</u> 114 - 110 <u>__m__</u>
Golden	<u>The</u> Toronto_Raptors <u>defeated the Detroit_Pistons</u> 114 - 110 <u>on Sunday at ...</u>
Seq2Seq	<u>The</u> Toronto_Raptors <u>defeated the the</u> 114 - 110 <u>on Wednesday at the Center</u>
GAN	<u>The</u> Toronto_Raptors <u>defeated the visiting</u> 114 - 110 <u>on Friday .</u>
Self-attn	<u>The</u> Toronto_Raptors <u>defeated the Philadelphia_76ers</u> 114 - 110 <u>on Friday .</u>



Summary

The proposed task is lack of novelty however, the experiments validate that the self attention model is capable of generating continuous phrases for multiple slots.

It inspires me that, for knowledge-grounded dialogue, maybe we can firstly generate a response with slot by a ordinary Seq2seq model, then use the given knowledge to fill the slots.

TIGS: An Inference Algorithm for Text Infilling with Gradient Search

Dayiheng Liu[†], Jie Fu[‡], Pengfei Liu[§], Jiancheng Lv^{†*}

[†]College of Computer Science, Sichuan University

[‡]Mila, IVADO, Polytechnique Montreal

[§]School of Computer Science, Fudan University

losinuris@gmail.com

lvjiancheng@scu.edu.cn

Text Infilling in Dialogues

Different from unconstrained text infilling, Dialogue text infilling is a more practical problem.

Given a pair of text infilling data $(x, y^{\mathbb{B}})$ the method aims at finding an infilled word set $\hat{y} = \{\hat{y}_1, \hat{y}_2, \dots, \hat{y}_{|\mathbb{B}|}\}$.

The goal is to develop an inference algorithm to fill the slots when given a trained seq2seq model.

Input: Hey, how about going for a few beers after dinner ?

Ground Truth

You know that is tempting but is really not good for our fitness .

Seq2seq + Left-to-Right Beam Search

You know that I like it very much, let's for our fitness .

Seq2seq (backward) + Right-to-Left Beam Search

You know that not going, it is really bad for our fitness .

Text Infilling with Gradient Search

Algorithm 1 TIGS algorithm

Input: a trained seq2seq model, a pair of text infilling data $(\mathbf{x}, \mathbf{y}^{\mathbb{B}})$, output length m .

Output: a complete output sentence \mathbf{y}^* .

Initialize the infilled word set $\hat{\mathbf{y}}$ and initialize \mathbf{y}^* by infilling $\mathbf{y}^{\mathbb{B}}$ with $\hat{\mathbf{y}}$.

Initialize $\hat{\mathbf{y}}^{emb}$ by looking up the word embedding matrix \mathbb{W}^{emb} .

for $t = 1, 2, \dots, T$ **do**

for $j = 1, 2, \dots, |\mathbb{B}|$ **do**

O-step:

 Update $\hat{\mathbf{y}}_j^{emb}$ with gradient $\nabla_{\hat{\mathbf{y}}_j^{emb}} \mathcal{L}(\mathbf{x}, \mathbf{y}^*)$

P-step:

 Set $S = \text{nearest-K } \underset{y_k \in \mathcal{V}}{\text{dist}}(\hat{\mathbf{y}}_j^{emb}, \mathbf{y}_k^{emb})$

 Set $\hat{y}_j = \arg \min_{\hat{y}_j \in S} \mathcal{L}_{NLL}(\mathbf{x}, \mathbf{y}^*)$

end for

 Update \mathbf{y}^* with \hat{y}_j

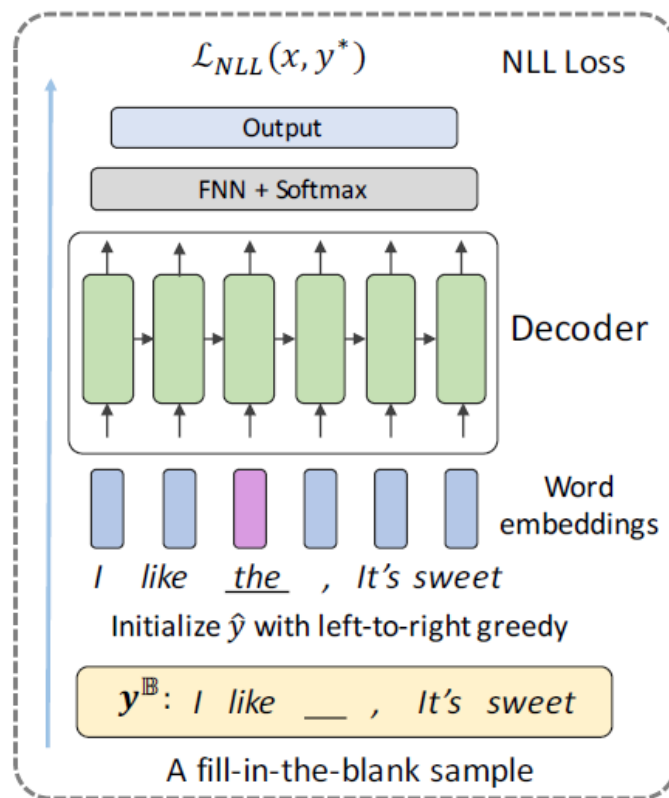
if convergence **then**

break

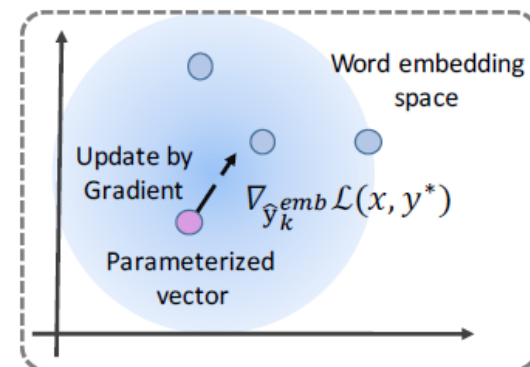
end if

end for

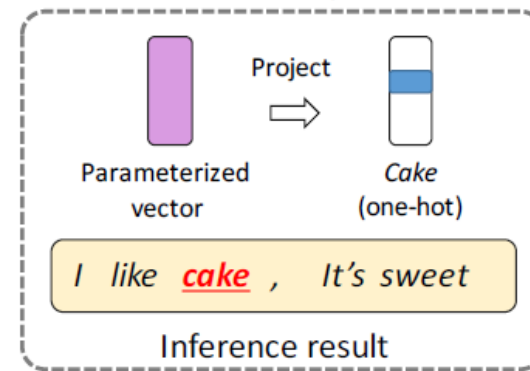
return \mathbf{y}^*



(a) Initialize $\hat{\mathbf{y}}$ and get NLL loss



(b) O-step



(c) P-step

Experiments

Input (Query)	can you study with the radio on ?
Template	__, __ listen __ music .
Ground Truth	<u>no</u> , <u>I</u> listen <u>to</u> <u>background</u> music .
Seq2seq-f	<u>i'd</u> , <u>I'm</u> listen <u>to</u> <u>the</u> music .
Seq2seq-b	<u>music</u> , <u>can</u> listen <u>to</u> <u>the</u> music .
Mask-Seq2Seq	<u>yes</u> , <u>they</u> listen <u>to</u> <u>the</u> music .
Mask-Self-attn	<u>yes</u> , <u>it's</u> <u>a</u> <u>lot</u> <u>of</u> music .
BiRNN-BiBS	<u>I</u> , <u>to</u> listen <u>to</u> <u>the</u> music .
BiRNN-GSN	<u>yes</u> , <u>I'll</u> listen <u>to</u> <u>the</u> music .
TIGS	<u>yes</u> , <u>I</u> listen <u>to</u> <u>classical</u> music .
Input (Query)	pretty good , thanks . i'm going to see my uncle .
Template	__ __ then __ and keep __ touch .
Ground Truth	<u>good</u> <u>bye</u> then <u>.</u> and keep <u>in</u> touch .
Seq2seq-f	<u>nice</u> <u>to</u> then <u>.</u> and keep <u>your</u> touch .
Seq2seq-b	<u>minutes</u> <u>.</u> then <u>go</u> and keep <u>in</u> touch .
Mask-Seq2Seq	<u>ok</u> <u>.</u> then <u>go</u> <u>then</u> keep <u>in</u> touch .
Mask-Self-attn	<u>then</u> <u>.</u> then <u>keep</u> and keep <u>in</u> touch .
BiRNN-BiBS	<u>you</u> <UNK> then <UNK> and keep <u>it</u> touch .
BiRNN-GSN	<u>ok</u> <u>.</u> then <u>go</u> and keep <u>in</u> touch .
TIGS	<u>alright</u> <u>.</u> then <u>.</u> and keep <u>in</u> touch .

Methods	Dialog	Poetry	APRC
BiRNN-BiBS	1.524	1.478	1.558
BiRNN-GSN	2.979	2.675	2.261
Mask-Self-attn	2.270	2.727	3.042
TIGS	3.226	3.120	3.137



Summary

How to generate the infilled text sequence with mask tags is still a huge challenge for applying this model to dialogue generation.

FIN