TIGS: An Inference Algorithm for Text Infilling with Gradient Search

Dayiheng Liu[1], Jie Fu[2], Pengfei Liu[3], Jiancheng Lv[1]
College of Computer Science, Sichuan University
Mila, IVADO, Polytechnique Montreal
School of Computer Science, Fudan University

Introduction

Text Infilling:

- ✓ Sequence missing value reconstruction (e.g., for damaged or historical documents)
- ✓ Lexically constrained sentence generation
- ✓ Fill in the blanks (e.g., Lyrics and poetry generation)

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 .

Challenge: given a well-trained sequential generative model, generating missing symbols conditioned on the context is challenging for existing greedy approximate inference algorithms.

Main Contribution:

- We propose an *iterative inference algorithm* based on *gradient search*, which could be the first inference algorithm that can be broadly applied to any neural sequence generative models for text infilling tasks.
- Extensive experimental comparisons show the effectiveness and efficiency of the proposed method on *three different text infilling tasks*, compared with *five* state-of-the-art methods.

Experiments

Three Datasets:

- ❑ DailyDialog
- ❑ Chinese Poetry
- ❑ Amazon product reviews

Six settings:

- ❖ Mask ratio: 25% 50% 75%
- ❖ Mask strategy: random middle

Input: What is the weather like today ?

Ground Truth
It stops snowing , but there's a bit wind .

Mask strategy: Random Mask ratio: 75%
__ __ snowing __ __ __ __ __ wind .

Mask strategy: Random Mask ratio: 50%
It __ snowing __ but __ __ bit __ .

Mask strategy: Middle Mask ratio: 25%
It stops snowing , __ __ a bit wind .

Seven baselines:

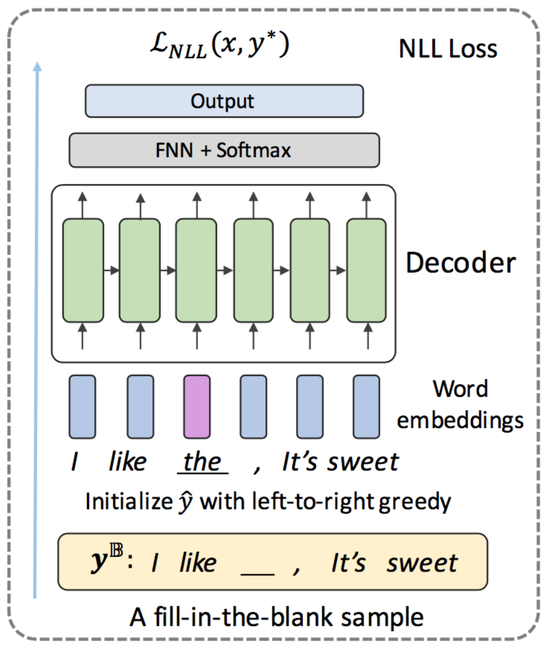
- ❖ UniRNN + left-to-right beam search
- ❖ UniRNN + right-to-left beam search
- ❖ UniRNN + f + b
- ❖ BiRNN-BiBS
- ❖ BiRNN-GSN
- ❖ Mask-Seq2Seq
- ❖ Mask-Transformer

Our code and data are available at
https://github.com/dayihengliu/Text-Infilling-Gradient-Search

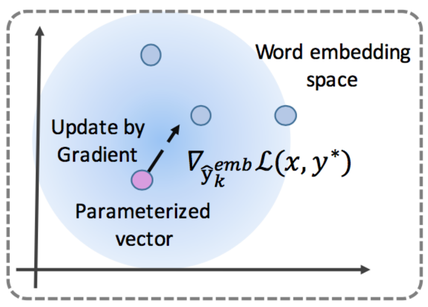
Methodology

Algorithm 1 TIGS algorithm

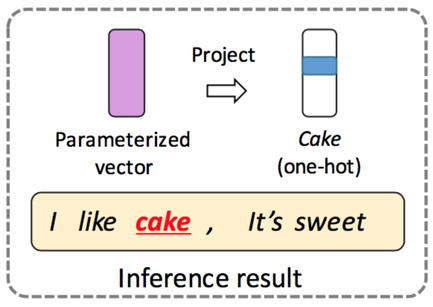
Input: a trained seq2seq model, a pair of text infilling data $(\boldsymbol{x}, \boldsymbol{y}^{\mathbb{B}})$, output length m.
Output: a complete output sentence $\boldsymbol{y}^*$.
Initialize the infilled word set $\hat{\boldsymbol{y}}$ and initialize $\boldsymbol{y}^*$ by infilling $\boldsymbol{y}^{\mathbb{B}}$ with $\hat{\boldsymbol{y}}$.
Initialize $\hat{\boldsymbol{y}}^{emb}$ by looking up the word embedding matrix $\mathbb{W}^{emb}$.
for $t = 1, 2, \ldots, T$ **do**
  for $j = 1, 2, \ldots, |\mathbb{B}|$ **do**
    O-step:
    Update $\hat{\boldsymbol{y}}_j^{emb}$ with gradient $\nabla_{\hat{\boldsymbol{y}}_j^{emb}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{y}^*)$
    P-step:
    Set $S = \underset{y_k \in \mathcal{V}}{\text{nearest-K}}\, dist(\hat{\boldsymbol{y}}_j^{emb}, \boldsymbol{y}_k^{emb})$
    Set $\hat{y}_j = \underset{\hat{y}_j \in S}{\arg\min}\, \mathcal{L}_{NLL}(\boldsymbol{x}, \boldsymbol{y}^*)$
  end for
  Update $\boldsymbol{y}^*$ with $\hat{y}_j$
  if convergence **then**
    break
  end if
end for
return y^*



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



(b) O-step



(c) P-step

Partial Results

Datasets	Metrics	Methods	r=25%		r=50%		r=75%	
			Random	Middle	Random	Middle	Random	Middle
Dialog	NLL	Seq2Seq-f	3.573	3.453	3.653	3.316	3.328	2.975
		Seq2Seq-b	3.657	3.558	3.911	3.542	3.713	3.421
		Seq2Seq-f+b	3.397	3.321	3.491	3.213	3.233	2.932
		BiRNN-BiBS	3.248	3.279	3.268	3.294	3.245	3.217
		BiRNN-GSN	3.239	3.270	3.219	3.199	3.086	2.938
		Mask-Seq2Seq	3.406	3.368	3.434	3.347	3.279	3.177
		Mask-Self-attn	3.567	3.524	3.694	3.466	3.509	3.205
		TIGS (ours)	**3.143**	**3.164**	**3.050**	**3.030**	**2.920**	**2.764**
	BLEU	Template	0.780	0.823	0.621	0.700	0.552	0.601
		Seq2Seq-f	0.834	0.861	0.670	0.737	0.584	0.640
		Seq2Seq-b	0.837	0.862	0.675	0.739	0.584	0.627
		Seq2Seq-f+b	0.860	0.881	0.692	0.751	0.594	0.643
		BiRNN-BiBS	0.828	0.852	0.661	0.725	0.575	0.626
		BiRNN-GSN	0.894	0.892	**0.726**	0.752	0.600	0.643
		Mask-Seq2Seq	0.867	0.887	0.719	0.769	0.614	**0.662**
		Mask-Self-attn	0.858	0.864	0.719	0.743	**0.623**	0.643
		TIGS (ours)	**0.895**	**0.894**	0.724	**0.754**	0.596	0.644