Generating Sentences from Disentangled Syntactic and Semantic Spaces

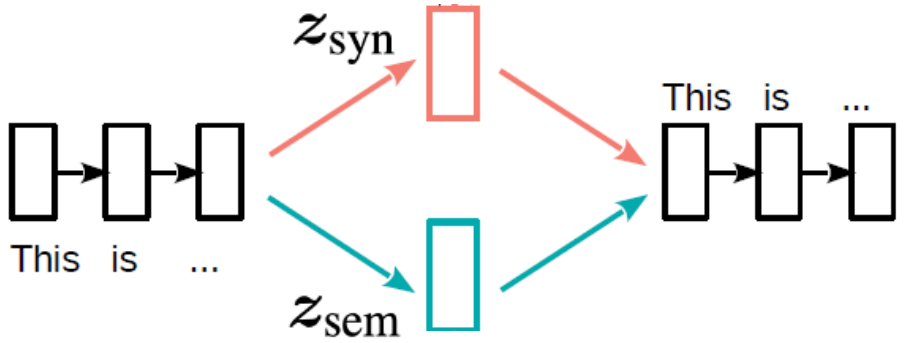
Yu Bao[1*] **Hao Zhou**[2*] **Shujian Huang**[1†] **Lei Li**[2] **Lili Mou**[3]
Olga Vechtomova[3] **Xinyu Dai**[1] **Jiajun Chen**[1]
[1]National Key Laboratory for Novel Software Technology, Nanjing University, China
[2]ByteDance AI Lab, Beijing, China
[3]University of Waterloo, Canada
{baoy,huangsj,dxy,chenjj}@nlp.nju.edu.cn
{zhouhao.nlp,lileilab}@bytedance.com
doublepower.mou@gmail.com, ovechtomova@uwaterloo.ca

ACL2019

紹介者: 三輪誠 (豊田工業大学)

スライド中の図表は論文のものを利用しています

概要



- VAEの潜在表現を構文的な表現と意味的な表現に分ける disentangled syntactic and semantic spaces of VAE （DSS-VAE）を提案
  - 文生成・言い換え生成で高い性能
  - スタイル変換 (syntax-transfer) も可能
    - Ref_{sem}の意味を保ちつつ，Ref_{syn}の構文に合わせる

Semantic and Syntactic Providers		Syntax-Transfer Output	
Ref_{syn}:	There is an apple on the table.	**VAE:**	The man is in the kitchen.
Ref_{sem}:	The airplane is in the sky.	**DSS-VAE:**	There is a airplane in the sky.
Ref_{syn}:	The shellfish was cooked in a wok.	**VAE:**	The man was filled with people.
Ref_{sem}:	The stadium was packed with people.	**DSS-VAE:**	The stadium was packed with people.
Ref_{syn}:	The child is playing in the garden.	**VAE:**	There is a person in the garden.
Ref_{sem}:	There is a dog behind the door.	**DSS-VAE:**	A dog is walking behind the door.

選んだ理由

- disentanglement representation learningの研究が増えているのとVAEや生成をあまり知らないので読んでみた
  - タイトルにdisentangleがある今年の論文はNAACL6件, ACL7件, ICML6件, ICLR2件, CVPR8件, NeurIPS11件

DSS-VAE (disentangled syntactic and semantic spaces of VAE) のモデル

- 文$\boldsymbol{x}$は構文的な潜在変数z_{syn}と意味的な潜在変数z_{sem}から生成

$$p(\boldsymbol{x}) = \int p(\boldsymbol{z}_{\text{sem}}, \boldsymbol{z}_{\text{syn}}) p(\boldsymbol{x} | \boldsymbol{z}_{\text{sem}}, \boldsymbol{z}_{\text{syn}}) \, \mathrm{d}\boldsymbol{z}_{\text{sem}} \, \mathrm{d}\boldsymbol{z}_{\text{syn}}$$

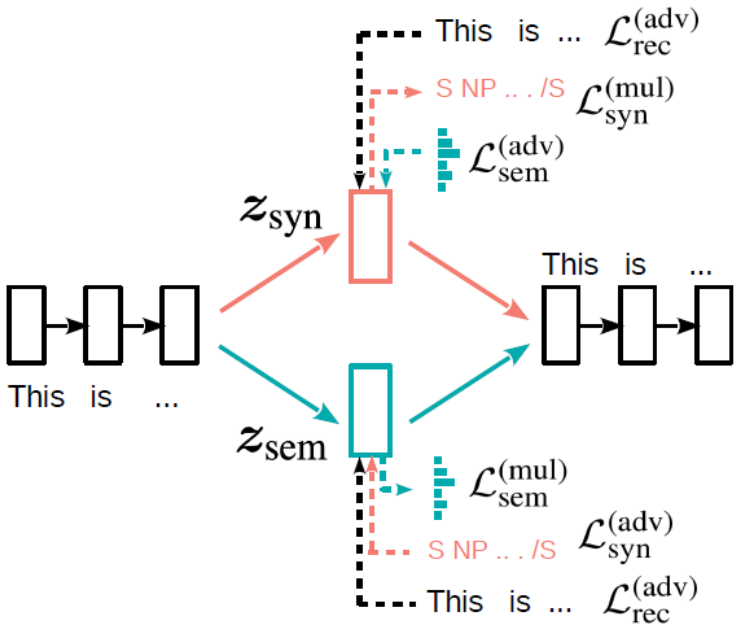
- 事前分布$p(z_{syn})$と$p(z_{sem})$は独立した正規分布 $\mathcal{N}(\mathbf{0}, \mathrm{I})$を仮定
  - 実験では100次元

- c.f. VAE

$$p(\boldsymbol{x}) = \int p(\boldsymbol{z}) p(\boldsymbol{x} | \boldsymbol{z}) \, \mathrm{d}\boldsymbol{z}$$

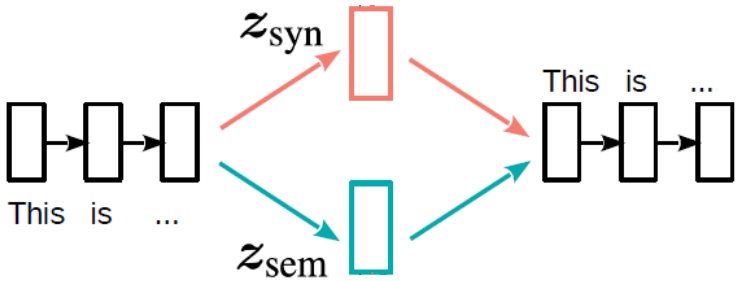
DSS-VAEの目的関数

- 以下の4つから目的関数を設定
  1. 変分下限 (ELBO; Evidence Lower BOund)
     - VAEの目的関数
  2. マルチタスク学習
     - 構文，意味それぞれのための補助タスク
  3. 敵対的損失
     - 構文と意味の表現を分離するための補助タスク
  4. 敵対的復元損失
     - 構文・意味単独では文を復元できないようにするための補助タスク



変分下限 (ELBO) の最大化

- VAE (文の復元) の目的関数



$$\log p(\boldsymbol{x}) \geq \mathrm{ELBO}$$
$$= \mathop{\mathbb{E}}_{q(\boldsymbol{z}_{\mathrm{sem}}|\boldsymbol{x})q(\boldsymbol{z}_{\mathrm{syn}}|\boldsymbol{x})} \left[\log p(\boldsymbol{x}|\boldsymbol{z}_{\mathrm{sem}}, \boldsymbol{z}_{\mathrm{syn}})\right] - \mathrm{KL}\left(q(\boldsymbol{z}_{\mathrm{sem}}|\boldsymbol{x}) \,\middle\|\, p(\boldsymbol{z}_{\mathrm{sem}})\right) - \mathrm{KL}\left(q(\boldsymbol{z}_{\mathrm{syn}}|\boldsymbol{x}) \,\middle\|\, p(\boldsymbol{z}_{\mathrm{syn}})\right)$$

- 事後分布 $q(\boldsymbol{z}_{\mathrm{sem}}|\boldsymbol{x})$ は正規分布 $\mathcal{N}(\boldsymbol{\mu}_{\mathrm{sem}}, \boldsymbol{\sigma}^2_{\mathrm{sem}})$ を仮定．分布パラメタは入力をGRUでエンコードした $\boldsymbol{r}_x$ を等分割したものから計算

$$\begin{bmatrix} \boldsymbol{\mu}_{\mathrm{sem}} \\ \boldsymbol{\sigma}_{\mathrm{sem}} \end{bmatrix} = \begin{bmatrix} W^{\mu}_{\mathrm{sem}} \\ W^{\sigma}_{\mathrm{sem}} \end{bmatrix} \mathrm{ReLU}(W_{\mathrm{sem}} \boldsymbol{r}^{\mathrm{sem}}_x + \boldsymbol{b}_{\mathrm{sem}}) \quad \boldsymbol{r}_x = [\boldsymbol{r}^{\mathrm{sem}}_x; \boldsymbol{r}^{\mathrm{syn}}_x]$$

- $q(\boldsymbol{z}_{\mathrm{syn}}|\boldsymbol{x})$ も同様

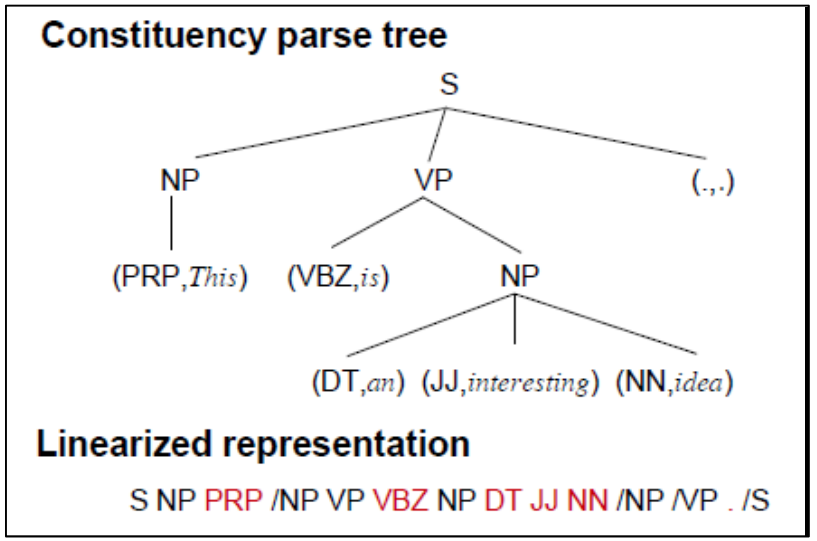
マルチタスク学習

- 構文・意味をそれぞれの表現から予測する
- Bag-of-words分布の予測

$$\mathcal{L}_{\text{sem}}^{(\text{mul})} = -\sum_{w \in \mathcal{V}} t_w \log p(w|\boldsymbol{z}_{\text{sem}})$$

- （線形化された）構文情報の予測

$$\mathcal{L}_{\text{syn}}^{(\text{mul})} = -\sum_{i=1}^{n} \log p(s_i|s_1 \cdots s_{i-1}, \boldsymbol{z}_{\text{syn}})$$



敵対的損失

- 構文から意味，意味から構文を生成できないようにする
- z_{syn} からBOWを，z_{sem} から構文情報を予測する2つのモデル p_{adv} を騙すように学習（←z_{syn} と z_{sem} を分離したい）

$$\mathcal{L}_{\text{sem}}^{(\text{adv})} = \sum_{w \in \mathcal{V}} t_w \log p_{\text{adv}}(w | \boldsymbol{z}_{\text{syn}})$$

$$\mathcal{L}_{\text{syn}}^{(\text{adv})} = \sum_{i=1}^{n} \log p_{\text{adv}}(s_i | s_1 \cdots s_{i-1}, \boldsymbol{z}_{\text{sem}})$$

- p_{adv} はVAEのパラメタを固定して更新
  - p_{adv} の学習とVAEの学習を交互に行う（はず）

敵対的復元損失

- 片方の表現だけでは文を再現できないようにする

$$\mathcal{L}_{\text{rec}}^{(\text{adv})}(\boldsymbol{z}_s) = \sum_{i=1}^{M} \log p_{\text{rec}}(x_i | x_{<i}, \boldsymbol{z}_s) \qquad (\boldsymbol{z}_s = \boldsymbol{z}_{\text{syn}} \text{ or } \boldsymbol{z}_{\text{sem}})$$

DSS-VAEの目的関数のまとめ

KLのハイパーパラメタは0から sigmoid関数に従って，増やす (KL annealing)

β-VAEはβ ($=\lambda^{\mathrm{KL}}$)>1

$$\mathcal{L} = \mathcal{L}_{\text{vae}} + \mathcal{L}_{\text{aux}}$$

VAE

$$= \Big[- \mathop{\mathbb{E}}_{q(\boldsymbol{z}_{\text{sem}}|\boldsymbol{x})q(\boldsymbol{z}_{\text{syn}}|\boldsymbol{x})} \log \left[p(\boldsymbol{x}|\boldsymbol{z}_{\text{sem}}, \boldsymbol{z}_{\text{syn}}) \right] + \lambda^{\text{KL}}_{\text{sem}} \mathrm{KL}\left(q(\boldsymbol{z}_{\text{sem}}|\boldsymbol{x}) \,\|\, p(\boldsymbol{z}_{\text{sem}}) \right) + \lambda^{\text{KL}}_{\text{syn}} \mathrm{KL}\left(q(\boldsymbol{z}_{\text{syn}}|\boldsymbol{x}) \,\|\, p(\boldsymbol{z}_{\text{syn}}) \right) \Big]$$

$$+ \lambda^{\text{mul}}_{\text{sem}} \mathcal{L}^{(\text{mul})}_{\text{sem}} + \lambda^{\text{adv}}_{\text{sem}} \mathcal{L}^{(\text{adv})}_{\text{sem}} + \lambda^{\text{rec}}_{\text{sem}} \mathcal{L}^{(\text{adv})}_{\text{rec}}(\boldsymbol{z}_{\text{sem}})$$

$$+ \lambda^{\text{mul}}_{\text{syn}} \mathcal{L}^{(\text{mul})}_{\text{syn}} + \lambda^{\text{adv}}_{\text{syn}} \mathcal{L}^{(\text{adv})}_{\text{syn}} + \lambda^{\text{rec}}_{\text{syn}} \mathcal{L}^{(\text{adv})}_{\text{rec}}(\boldsymbol{z}_{\text{syn}})$$

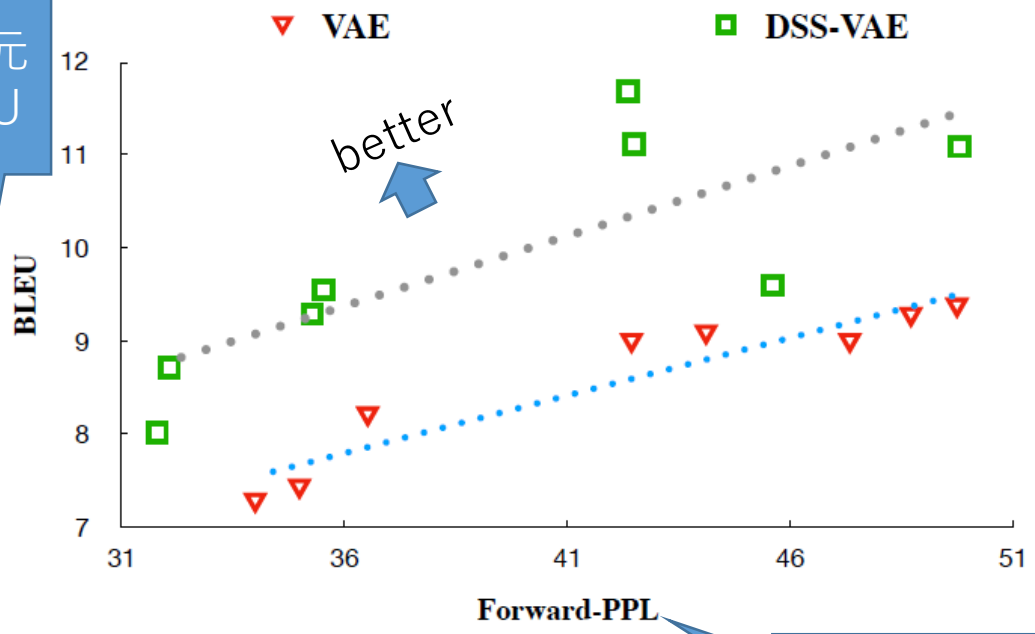
マルチタスク学習　　敵対的損失　　敵対的復元損失

Hyper-parameters	PTB	Quora
$\lambda^{\text{KL}}_{sem}$	1.0	1/3
$\lambda^{\text{KL}}_{syn}$	1.0	2/3
$\lambda^{\text{mul}}_{sem}$	0.5	5.0
$\lambda^{\text{mul}}_{syn}$	0.5	1.0
$\lambda^{\text{adv}}_{sem}$	0.5	0.5
$\lambda^{\text{adv}}_{syn}$	0.5	0.5
$\lambda^{\text{rec}}_{sem}$	0.5	1.0
$\lambda^{\text{rec}}_{syn}$	0.5	0.05
Batch size	32	50
GRU Dropout	0.1	0.3

言語生成

- PTBで実験
- KLのハイパーパラメタで決定的かどうかのトレードオフ

VAEで復元後のBLEU



決定的

生成文で学習した言語モデルをPTB test setsで評価

Model	Reverse PPL↓
Real data	70.76
LSTM-LM	132.46
PRPN-LM	116.67
VAE	125.86
DSS-VAE	**116.23**

生成文を事前学習した言語モデルで評価

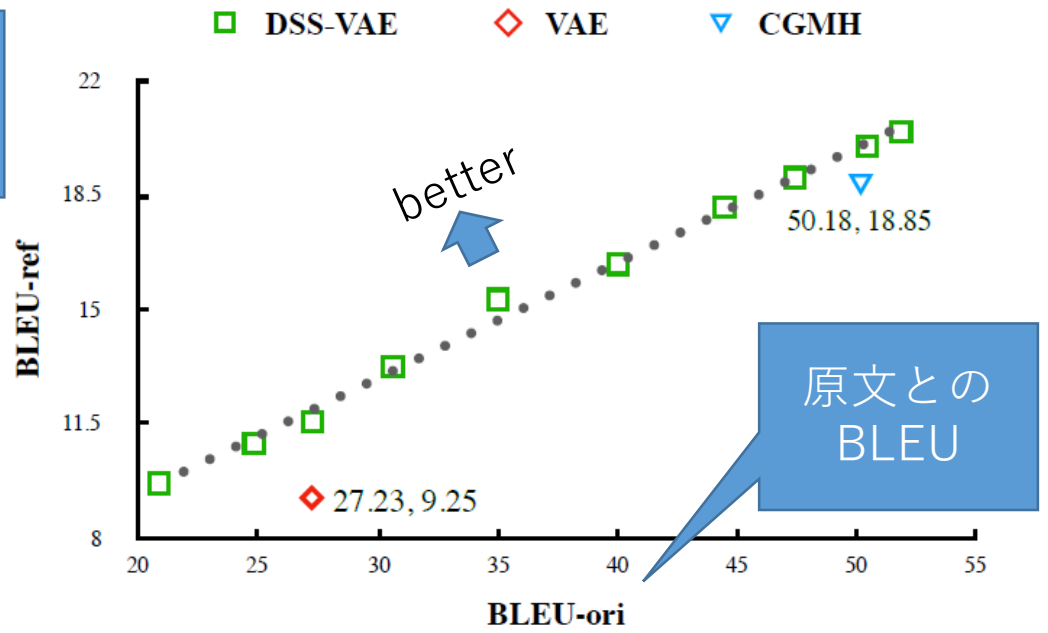
言い換え生成

$$z^*_{\text{sem}} = \text{argmax}_{z_{\text{sem}}}\ q(z_{\text{sem}}|x^*)$$
$$z^*_{\text{syn}} \sim q(z_{\text{syn}}|x^*)$$
$$z^* = [z^*_{\text{syn}}; z^*_{\text{sem}}]$$

構文表現のみサンプリング➔
意味が同じで構文の異なる
言い換えを生成

- Quoraデータセットを利用
  - 言い換えペア140k, 非言い換えペア260k (3kを開発，30kをテスト)

言い換え文とのBLEU



原文とのBLEU

スタイル変換

$$z^*_{\text{sem}} = \text{argmax}_{z_{\text{sem}}}\ q(z_{\text{sem}}|x_2)$$
$$z^*_{\text{syn}} = \text{argmax}_{z_{\text{syn}}}\ q(z_{\text{syn}}|x_1)$$
$$z = [z^*_{\text{sem}}; z^*_{\text{syn}}]$$

x_2の意味表現とx_1の構文表現から文を作成

- SNLI (1,000非言い換えペア) を利用
- “compatible”な文間であれば有望な結果

Semantic and Syntactic Providers		Syntax-Transfer Output	
Ref$_{\text{syn}}$:	There is an apple on the table.	**VAE**:	The man is in the kitchen.
Ref$_{\text{sem}}$:	The airplane is in the sky.	**DSS-VAE**:	There is a airplane in the sky.
Ref$_{\text{syn}}$:	The shellfish was cooked in a wok.	**VAE**:	The man was filled with people.
Ref$_{\text{sem}}$:	The stadium was packed with people.	**DSS-VAE**:	The stadium was packed with people.
Ref$_{\text{syn}}$:	The child is playing in the garden.	**VAE**:	There is a person in the garden.
Ref$_{\text{sem}}$:	There is a dog behind the door.	**DSS-VAE**:	A dog is walking behind the door.

スタイル変換

$$z^*_{\text{sem}} = \text{argmax}_{z_{\text{sem}}}\ q(z_{\text{sem}}|x_2)$$
$$z^*_{\text{syn}} = \text{argmax}_{z_{\text{syn}}}\ q(z_{\text{syn}}|x_1)$$
$$z = [z^*_{\text{sem}}; z^*_{\text{syn}}]$$

x_2の意味表現とx_1の構文表現から文を作成

- 片方の表現だけ復元できないようにすると，反対の表現が強くなる
  - 両方を復元できないようにするとバランスの取れた表現

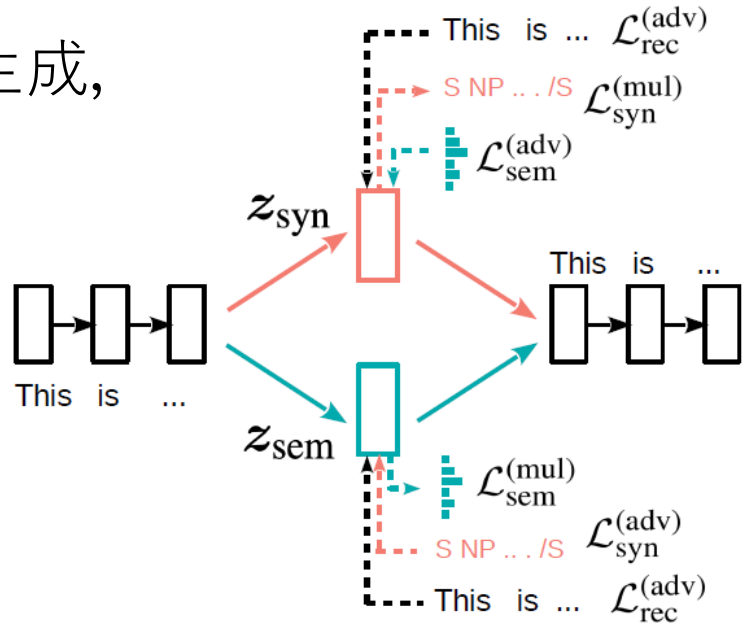
Tree edit distance

Model	word-BLEU (corpus)		Average TED (per sentence)		Geo Mean $\Delta^{\uparrow}$
	$\text{Ref}_{\text{sem}}{}^{\uparrow}$	$\text{Ref}_{\text{syn}}{}^{\downarrow}$	$\text{Ref}_{\text{sem}}{}^{\uparrow}$	$\text{Ref}_{\text{syn}}{}^{\downarrow}$	
VAE	6.81	6.68	149.22	148.59	0.29
$\mathcal{L}^{(\text{mul})}_{\text{sem}} + \mathcal{L}^{(\text{mul})}_{\text{syn}} + \mathcal{L}^{(\text{adv})}_{\text{sem}} + \mathcal{L}^{(\text{adv})}_{\text{syn}}$	12.14	6.22	159.51	134.80	12.09
$+\mathcal{L}^{(\text{adv})}_{\text{rec}}(z_{\text{sem}})$	11.83	6.60	**163.40**	**131.27**	12.96
$+\mathcal{L}^{(\text{adv})}_{\text{rec}}(z_{\text{syn}})$	**14.33**	**6.07**	159.20	134.22	14.36
$+\mathcal{L}^{(\text{adv})}_{\text{rec}}(z_{\text{syn}}) + \mathcal{L}^{(\text{adv})}_{\text{rec}}(z_{\text{sem}})$	13.74	6.15	161.94	131.09	**15.30**

Geo Meanは両評価の$\text{Ref}_{\text{sem}} - \text{Ref}_{\text{syn}}$の平均

まとめ

- 構文表現と意味表現を分離するDSS-VAEを提案
  - 3つの補助タスクにより分離を促進
  - VAEと違い，構文情報の操作により言い換え生成，スタイル変換に利用可能
- 単語生成・言い換え生成・スタイル変換の3つのタスクで高い性能・有望な結果



A Multi-Task Approach for Disentangling Syntax and Semantics in Sentence Representations

NAACL2019

Mingda Chen　　Qingming Tang　　Sam Wiseman　　Kevin Gimpel
Toyota Technological Institute at Chicago, Chicago, IL, 60637, USA

- 読んだ論文とは異なる損失で似た効果（事前分布なども違う）
  - paraphraseデータから意味の表現だけを入れ替えた損失

$$\mathop{\mathbb{E}}_{\substack{y_2 \sim q_\phi(y|x_2) \\ z_1 \sim q_\phi(z|x_1)}} [-\log p_\theta(x_1|y_2, z_1)] + \mathop{\mathbb{E}}_{\substack{y_1 \sim q_\phi(y|x_1) \\ z_2 \sim q_\phi(z|x_2)}} [-\log p_\theta(x_2|y_1, z_2)]$$

y: syntax, z: semantics

  - 意味の表現だけからparaphraseを予測するタスクの損失
  - syntaxの表現とBOWから単語位置を予測するタスクの損失

Query Sentence	Semantically Similar	Syntactically Similar
i have much more colours at home .	even if there was food , would n't it be at least 300 years old ?	you have a beautiful view from here .
victor had never known darkness like it .	he had never experienced such darkness as this .	you seem like a really nice kid .
this is , uh , too serious .	but this is too serious .	it is , however , illegal discrimination .
you 're gon na save her life .	you will save her .	you 're gon na give a speech .
we 've got to get a move on .	come on , we got ta move .	you 'll have to get in there .
and that was usually the highlight of my day .	i really enjoyed it when i did it .	and yet that was not the strangest aspect of the painting .

Disentangled Representation Learning for Non-Parallel Text Style Transfer

Vineet John, Lili Mou, Hareesh Bahuleyan, Olga Vechtomova
University of Waterloo

同じ著者

ACL2019

Preliminary work rejected by EMNLP-18

- style (sentiment) とcontentを分ける

$$J_{\text{ovr}} = J_{\text{AE}}(\theta_{\text{E}}, \theta_{\text{D}})$$ VAE

$$+\lambda_{\text{mul(s)}} J_{\text{mul(s)}}(\theta_{\text{E}}, \theta_{\text{mul(s)}}) - \lambda_{\text{adv(s)}} J_{\text{adv(s)}}(\theta_{\text{E}})$$ Sentimentの予測と敵対的損失

$$+\lambda_{\text{mul(c)}} J_{\text{mul(c)}}(\theta_{\text{E}}, \theta_{\text{mul(c)}}) - \lambda_{\text{adv(c)}} J_{\text{adv(c)}}(\theta_{\text{E}})$$ BoW (感情語を除いた内容語)の予測と敵対的損失

Original (Positive)	DAE Transferred (Negative)	VAE Transferred (Negative)
the food is excellent and the service is exceptional	the food was a bit bad but the staff was exceptional	the food was bland and i am not thrilled with this
the waitresses are friendly and helpful	the guys are rude and helpful	the waitresses are rude and are lazy
Original (Negative)	**DAE Transferred (Positive)**	**VAE Transferred (Positive)**
the desserts were very bland	the desserts were very good	the desserts were very good
it was a bed of lettuce and spinach with some italian meats and cheeses	it was a beautiful setting and just had a large variety of german flavors	it was a huge assortment of flavors and italian food