

# 発話スタイルに非依存な文章表現を用いたスタイルロバストな Seq2seq モデルの提案

古舞 千暁 滝口 哲也 有木 康雄

神戸大学 システム情報学研究科

kazuaki.furumai@stu.kobe-u.ac.jp, {takigu, ariki}@kobe-u.ac.jp

## 1 研究背景

近年、自然言語処理分野において翻訳や対話を行うシステムが盛んに研究されており、特に文章データから直接目的とする文章データの生成を行う Seq2seq モデルの登場により、その性能が飛躍的に向上している。しかし、そのようなシステムをユーザに使用してもらうにあたり、「標準語と方言」「話し言葉と書き言葉」「敬体と常体」などといった発話スタイルのギャップが学習に用いたデータと実際のデータとの間に存在する場合、Seq2seq モデルが適切な符号化が行えなくなることで性能が低下する問題がある。

パラレルデータコーパスなど、高品質な言語資源が存在する場合は、特異なスタイルを標準的なスタイルに変換することで対応が可能となるが、特に日本語ではそのようなコーパスは入手が困難である。方言データやノイズデータを対象とした研究も存在するが、データごとに人手によるラベル付けや変換辞書を必要としておりコストが高い。そこで本研究では、比較的に入手が容易なノンパラレルコーパスを活用して、発話スタイルにロバストな Seq2seq モデルを作成する手法を提案する。

文章を発話スタイルに非依存な文章表現（意味ベクトル）と発話スタイル情報を表すベクトル（スタイルベクトル）に分解するために敵対的生成ネットワークを利用し、Seq2seq モデルには Transformer Encoder-Decoder を拡張し利用した。

## 2 文章情報の分解

Romanov らによって提案された ADNet[1] を参考に、文章情報の分解を行なった。図 1 に ADNet の概略図を示す。

ADNet では、スタイルの異なる 2 つのコーパスを用いて、GRU エンコーダの隠れ状態から意味ベクトル

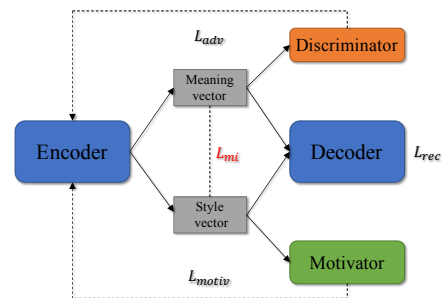


図 1: 相互情報量誤差を導入した ADNet

とスタイルベクトルを生成する。Discriminator は、分離の不完全性から意味ベクトルに含まれるスタイル情報を元に、元の文章がどちらのコーパスのものかの識別を試みる。Motivator は、スタイルベクトルから元の文章がどちらのコーパスのものかを識別する。この際、Discriminator の誤差関数には負の係数をかけて損失関数とする。この理由は、Discriminator が正しい識別を行えないようにすることで、スタイル情報を含まない意味ベクトルを生成するためである。また、Motivator の識別誤差については、識別が可能になるようなスタイルベクトルを生成する必要があるため、正の係数をかけて誤差関数とする。

しかし、Romanov らも述べている通り、これだけではスタイルベクトルに意味情報が含まれてしまうことを防ぐことができない。そこで我々は、意味ベクトルとスタイルベクトルが情報を共有しないことを促すために、生成される意味ベクトルとスタイルベクトルをそれぞれ確率変数と見なし、それらの間で計算される相互情報量を誤差とする損失関数を新たに加えた。

異なるスタイル  $a, b$  を持つ文章群  $X^a, X^b$  から生成される意味ベクトル、スタイルベクトル群をそれぞれ  $\mathbf{m}^a, \mathbf{m}^b, \mathbf{s}^a, \mathbf{s}^b$  とする。 $\mathbf{m}^a, \mathbf{m}^b, \mathbf{s}^a, \mathbf{s}^b$  を確率変数とみなすと、相互情報量誤差  $L_{mi}$  は以下のように表される。

$$L_{mi}(\theta_E) = I(\mathbf{m}^a; \mathbf{s}^a) + I(\mathbf{m}^b; \mathbf{s}^b) \\ = \sum_{i=a,b} (H(\mathbf{m}^i) + H(\mathbf{s}^i) - H(\mathbf{m}^i, \mathbf{s}^i))$$

ただし、 $I(\mathbf{m}^a; \mathbf{s}^a)$  は  $\mathbf{m}^a$  と  $\mathbf{s}^a$  の相互情報量、 $H(\mathbf{m}^i)$  は  $\mathbf{m}^i$  のエントロピー、 $H(\mathbf{m}^i, \mathbf{s}^i)$  は  $\mathbf{m}^i$  と  $\mathbf{s}^i$  の結合エントロピーである。ここで、 $\mathbf{m}$  と  $\mathbf{s}$  の真の確率分布関数は未知であるため、直接エントロピーの計算を行うことはできない。そこで本研究では Kraskov らが提案した推定手法 [2] を用いる。

### 3 意味ベクトルを用いた Seq2seq モデルの拡張

Seq2seq モデルとして、近年翻訳や対話タスクでも用いられている Transformer Encoder Decoder [3] を用いた。図 2 のように、Decoder の Multi-head attention からの出力において入力文章の意味ベクトルを結合し、その後線型結合層を経て Softmax 関数によりトークンの予測を行う。ここで意味ベクトルは入力単語系列から ADNet によって生成されたものであり、L2 正規化を行なったものを用いている。

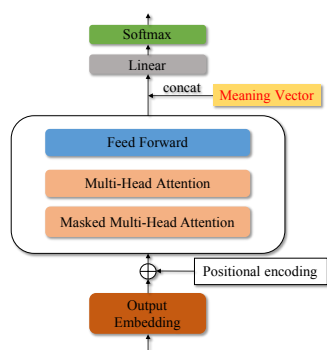


図 2: 意味ベクトルを用いた Transformer Encoder-Decoder

## 4 実験

相互情報量誤差の効果と、意味ベクトルによる拡張の効果それぞれ実験によって確認した。本節では、各実験に用いた評価指標とデータセットについて述べる。

### 4.1 評価指標

ADNet について、相互情報量誤差を用いた場合と用いなかった場合の比較のため、Shen らが提案した “Transfer Strength”, “Content Preservation” [4] という 2 つの客観評価指標を用いた。これは言い換えタスクにおける評価のために提案された指標であり、文章の変換強度と内容の保持具合を評価するものである。

まず、各文章のスタイルベクトルを変換先のコーパスからランダムに 10 文サンプリングし、それらのスタイルベクトルを平均したものと交換し、Decoder で再構成されることによって言い換えられた文章群を作成する。“Transfer Strength”については、ADNet の学習に用いたそれぞれのスタイルのコーパスを教師データとして LSTM による識別器を学習し、それを用いて先ほど作成した言い換え文章群の識別を行い、変換先のスタイルだと識別できた Accuracy をスコアとする。

“Content Preservation”については、事前学習された単語の分散表現を用いて文章ベクトルを作成して計算する。文章ベクトルは、文章に含まれる単語それぞれの分散表現の各次元の平均と最大と最小を結合したベクトルである。言い換え元、言い換え後の文章それぞれで文章ベクトルを作成し、それらの Cos 類似度をスコアとする。

また、拡張した Seq2seq モデルの評価のために翻訳実験を行ない、BLEU で評価を行なった。対話応答文生成も行い、その生成例を示す。

### 4.2 データセット

まず、言い換えタスクで用いられている Yelp Open Dataset<sup>1</sup> に対して、Positive, Negative の 2 つのスタイルを付与したデータセットを用いた。ここで、Yelp データセットは店舗に対する 1 点から 5 点の得点評価付きのレビューデータセットであり、positive データは評価が 3 点より高いもの、negative データは評価が 3 点より低いものを指す。また、方言データとして関西弁コーパス [5]、対話タスクのためにおーぷん 2 ちゃんねる対話コーパス [6]、標準語かつ翻訳データとして田中コーパス<sup>2</sup>を用いた。

<sup>1</sup><https://www.yelp.com/dataset>

<sup>2</sup>[http://www.edrdg.org/wiki/index.php/Tanaka\\_Corpus](http://www.edrdg.org/wiki/index.php/Tanaka_Corpus)

## 5 結果と考察

意味ベクトル, スタイルベクトルはいずれも 128 次元で生成し実験を行なった.

### 5.1 相互情報量誤差の評価

表 1: positive と negative の言い換え実験

	TS(%)	CP
Original	0.786	0.868
Proposed	<b>0.844</b>	<b>0.893</b>

表 2: 関西弁と標準語<sup>1</sup>の言い換え実験

	TS(%)	CP
Original	0.923	<b>0.771</b>
Proposed	<b>0.987</b>	0.768

表 3: 「なんでも実況 (ジュピター)」と「ニュー速 VIP」の言い換え実験 (おーぶん 2 ちゃんねる対話コーパス)

	TS(%)	CP
Original	0.513	0.450
Proposed	<b>0.895</b>	<b>0.829</b>

各データセットの組み合わせによる “Transfer Strength” (TS), “Content Preservation” (CP) の結果を表 1 から表 3 に示す. “Original” が相互情報量を用いなかった場合, “Proposed” が相互情報量を用いた場合を表す. “Transfer Strength” と “Content Preservation” はトレードオフの関係となりやすいが, どちらの指標でも上回っている場合が確認され, 表 2 に関しても “Content Preservation” に関して差は少なかった. また, 表 3 に関しては, 相互情報量を用いなかった場合は ADNet の学習がうまくいかなかった. “Positive”, “Negative” について, 生成した意味ベクトルとスタイルベクトルを t-SNE で 2 次元に圧縮し可視化したところ (図 3), スタイルベクトルがより分離されている様子が確認できた.

### 5.2 意味ベクトルを用いた Seq2seq モデルの評価

意味ベクトルを用いた Seq2seq モデルの評価のために, 翻訳と対話タスクで実験を行なった. 田中コー

<sup>1</sup>標準語として田中コーパスの日本語を用いた

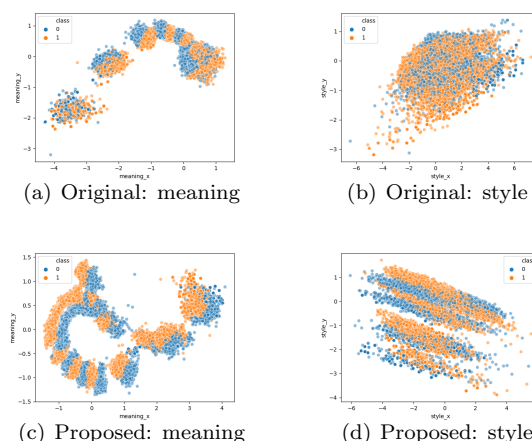


図 3: 意味ベクトル, スタイルベクトルの比較

パスを用いた翻訳実験に関する BLEU を表 4 に示す. Transformer Encoder-Decoder に関して, エンコーダ・デコーダは共に 4 層, モデルサイズは 128 次元, フィードフォワード層は 256 次元とした.

表 4: 田中コーパスを用いた翻訳実験

Model	1-gram	2-gram	3-gram	4-gram
Original	64.62	53.96	45.28	38.65
Proposed	<b>65.50</b>	<b>55.21</b>	<b>46.77</b>	<b>40.35</b>

表 4 の翻訳実験は, 方言などを入力したものでなく, 通常通り田中コーパスのみを用いたものであるが, それらの意味ベクトルを追加するだけでも性能が向上することが確認できた.

関西弁コーパス内の関西弁文章を入力した場合の翻訳例を表 5 に示す. ここで, A が意味ベクトルを用いた Seq2seq モデル, B が通常の Seq2seq モデルによる出力である. 田中コーパスの日本語と関西弁コーパスの間には, 方言というギャップに加え, 話し言葉と書き言葉というギャップも存在するため翻訳が難しくなっているが, それでも意味ベクトルを用いた Seq2seq モデルがより適切な翻訳を生成する傾向が確認できた.

次に, 「なんでも実況 (ジュピター)」と「ニュー速 VIP」コーパスで対話システムの学習を行なった. 「なんでも実況 (ジュピター)」では, 特有のスタイルが認められる発話がネットスラングとして流行している. 「ニュー速 VIP」への書き込みを学習データとし, 「なんでも実況 (ジュピター)」への書き込みを入力として生成した例を表 6 に示す. A が意味ベクトルを用いた Seq2seq モデル, B が通常の Seq2seq モデルによる

表 5: 関西弁の翻訳文生成例 (A: 意味ベクトル有, B: 意味ベクトル無)

入力	出力
外国で働いたほうが良いんちゃうん。	A) It's nice to work in foreign countries. B) Work is a good deal.
慣れたらそれでええのかな	A) I wonder if I could get used to B) You get used to your own way
試験難しかったんちゃん?	A) Did you enjoy your examination? B) You're in the examination.
どれぐらいしゃべれるん?	A) How long can you talk? B) How long is it?

表 6: 「なんでも実況」板の書き込みに対する応答文生成例 (A: 意味ベクトル有, B: 意味ベクトル無)

入力	出力
シュールストレミング試食会した人のブログ超おもしろい	A) ググってみるわ B) ありがとう!
まあ今ワイが考えた適当な嘘やけどな	A) 嘘つくな。嘘つきは無理 B) はいはい。
そマ? 数年後頑張るンゴ。ちな公務員志望	A) 公務員ってどんな仕事? B) <num> 年後半年くらい前から、
お疲れ様や、めげずに頑張るんやで	A) ありがとう! 頑張ります! B) ありがとう。ありがとう

出力である。特殊なスタイルが入力された場合でも、スタイルに非依存な単語を反映した応答をある程度生成できていると考えられる。

今後の課題として、これら生成例を用いた主観評価実験や相互情報量誤差の推定手法の比較を行う必要がある。また、ADNetにおいてDiscriminotrやMotivatorの出力を多次元とし、交差エントロピー誤差を用いることで複数のスタイルを一度に扱えるよう拡張することが考えられる。

## 6 謝辞

本研究の一部は、JSPS 科研費 JP17K00236 の助成を受けたものである。

## 参考文献

[1] Alexey Romanov et al., “Adversarial Decomposition of Text Representation,” Proc. of

NAACL, pp. 815–825, 2019.

- [2] Alexander Kraskov et al. “Estimating Mutual Information,” Phys. Rev. E69, 066138, 2004.
- [3] Ashish Vaswani et al., “Attention is all you need,” Proc. of NeurIPS, pp. 5998–6008, 2017.
- [4] Tianxiao Shen et al., “Style Transfer from Non-Parallel Text by Cross-Alignment,” Proc. of NeurIPS, 2017.
- [5] ヘファナン ケビン, “関西弁コーパスの紹介,” 総合政策研究 41 号, pp157-163, 2012
- [6] 稲葉 通将, “おーぷん 2 ちゃんねる対話コーパスを用いた用例ベース対話システム,” 第 10 回対話システムシンポジウム, 人工知能学会研究会資料 SIG-SLUD-B902-33, pp. 129–132, 2019.