

意味的類似性を報酬とした強化学習による文生成

安井 豪[†] 鶴岡 慶雅[†] 永田 昌明[‡]

[†] 東京大学 大学院情報理工学系研究科

[‡] NTT コミュニケーション科学基礎研究所

[†]{gyasui,tsuruoka}@logos.t.u-tokyo.ac.jp

[‡]nagata.masaaki@lab.ntt.co.jp

1 はじめに

近年の情報技術の発展に伴い、様々な自然言語処理の課題を高い精度で実現できるようになってきた。中でも機械翻訳や会話文生成などに代表される文生成タスクは盛んに研究されており、人による文と遜色ない文を出すこともできるようになってきている。一方で、従来の教師データを用いた文生成には問題点も存在する。そのひとつとして挙げられるのが、学習の仕組みが生成される単語の揺れや文構造の変化に寛容でないことである。

多くの文生成モデルは入力から情報を抽出する符号化器 (Encoder) と、抽出された情報からトークン列を出力する復号器 (Decoder) をつなぎ合わせた、Encoder-Decoder モデル [8] の形で実現される。教師あり学習で Encoder-Decoder モデルを学習する場合、損失関数として Cross-entropy 損失を用いることが多い。翻訳の正解例となる単語列を $y = \{y_1, y_2, \dots, y_T\}$ 、とすれば、この出力列に対する Cross-entropy 損失は

$$L_{CE} = - \sum_{t=1}^T \log \pi_{\theta}(y = y_t | y_{t-1}, s_t) \quad (1)$$

と定義される。ただし、 s_t は時間ステップ t における Decoder モデル π_{θ} の内部状態である。式 (1) からわかるように、Cross-entropy 損失では Decoder が出力と正解データをトークン単位で比較し、全体で和をとって損失を求める。その性質上 Cross-entropy 損失は文内での単語の位置のずれに対しても非寛容である。例えば、副詞が挟まるなどしてトークンが一つずれているだけで、以降のトークン列すべてが間違いとみなされる他、フレーズの順番を入れ替えるなどした文も、間違い扱いになってしまう。

本研究では Cross-entropy のように単語毎の一致ではなく、文全体の意味でモデル出力と正解データを比較することを目的とし、文同士の意味的類似度を

評価するタスクである Semantic Textual Similarity (STS) [1] に注目した。さらに自然言語処理のクラス分類問題を高精度で実現できる汎用事前学習モデル BERT [2] を STS に転移学習させ、翻訳学習時の生成文評価に用いることを試みた。

2 関連研究

2.1 Semantic Textual Similarity

STS [1] は文間の意味的類似度を推定するタスクである。関連タスクとされる含意推論や言い換え検出と違い、STS では似ているか否かの離散的な判別をするのではなく、0 (意味的に全く異なる) から 5 (意味的に同一) までの連続的な数値として推定しなければならない。したがって、STS では、文同士の段階的な意味的類似度を把握する必要があり、より高度な文の意味的理解を必要とするタスクと解釈することもできる。

2.2 BERT

Delvin らが考案した Bidirectional Encoder Representations from Transformers (BERT) [2] 以前の事前学習モデルでは、与えられた単語列から次の単語を予測するタスクを事前学習に用いることが多かった。例として ELMo [6] では Encoder に再帰的構造を持つ Long Short-Term Memory (LSTM) [3] を用いて単語予測をしている。再帰的なモデルの欠点は入力系列を前後のどちらか一方からしか入力することができない点にある。汎用性の高い文の特徴抽出をするには、一方向の関係性だけでなく、双方向の関係性を同時に学習できる方が望ましい。

BERT では機械翻訳で成果を上げた Transformer [9] を Encoder に採用した上で、事前学習のタスクも次の

単語の予測から、単語列の中でマスクにより隠された単語の予測に変更しており、事前学習タスクの方向依存性が解消されている。さらに、単語予測の他にも与えられた2文が隣接文であるかを予測するタスクも事前学習しており、単語単位だけでなく、文単位の特徴量も捉えることも試みている。

事前学習済み BERT の Encoder を目標とするタスクにさせると、様々な言語処理タスクで高精度のモデルが実現できることがわかっている。実際 BERT は9種類の言語処理タスクの総合成績を競う GLUE [10] において、発表当時最高精度を達成していた。

2.3 強化学習による文生成

強化学習は主にゲーム AI やロボット制御の分野で広く使われる学習手法であり、エージェントとそれに報酬を返す環境の間のフィードバックループを通して、累積報酬値を最大化することを目的としてエージェントを学習させるものである。このときエージェントの行動を評価する報酬関数が微分可能でなくても損失関数を計算できる。したがって強化学習を用いれば、ニューラルネットワークのパラメータに対して微分不可能な評価尺度を学習時にも利用することができる。文生成においても、強化学習を用いる事例は増えてきており、例えば Ranzato[7] らは翻訳評価尺度 BLEU[5] 等を報酬関数として用いることで、学習時に単語単位ではなく文単位での比較を行った。

比較的シンプルな強化学習の手法として REINFORCE [11] が知られている。文生成に REINFORCE を用いる場合、その目的はモデルの出力列 $\hat{y} = \{\hat{y}_1, \hat{y}_2, \dots, \hat{y}_T\}$ 全体に対する報酬の和の期待値 $\mathbb{E}_{\hat{y}_1, \dots, \hat{y}_T \sim \pi_\theta(\hat{y}_1, \dots, \hat{y}_T)} [r(\hat{y}_1, \dots, \hat{y}_T)]$ を最大化することである。ここで、 r は単語列を入力とする報酬関数であり、出力モデル π_θ のパラメータ θ に依存しない。REINFORCE の損失関数は出力モデル π_θ から文をトークン分布に従ってサンプリングすることによって式 (2) のように近似される。

$$L_{RL} = - \sum_{t=1}^N (r - r_b) \log(y = y_t | \hat{y}_{t-1}, s_t) \quad (2)$$

ただし、 r_b はサンプリングすることで生じる損失の分散を緩和するための期待報酬の推定量で、ベースライン報酬と呼ばれる。

3 提案手法

3.1 STS による生成文評価

本研究では STS によって翻訳モデルの生成文と正解文の意味的類似度を評価する。そのためには STS を高精度を評価できる評価器が必要となるが、これには事前学習済みの BERT モデルを STS タスクにファインチューンさせたものを利用することを提案する。

BERT モデルは文章の意味表現を捉える Encoder に当たる部分のモデルであり、意味表現から出力を生成する部分はタスクによって適切なモデルを選ぶ必要がある。STS は 0 から 5 までの連続値で文の意味的類似度を評価するタスクであるが、BERT はクラス分類問題に威力を発揮するモデルであり、連続値を出力することを想定されていない。そこで、STS を 0 から 5 までの離散的な 6 クラスへの分類タスクに変更する。具体的な生成モデルは線形層 1 層で BERT の状態数を出力数の 6 まで減らすだけの単純なもので、入力された 2 文に対する類似度評価値が正解値と同じになるよう Cross-entropy 損失で学習させる。STS の正解値も本来であれば連続値だが、Cross-entropy 損失では連続値の正解データをうまく利用できないので、小数点以下を四捨五入をすることで 0 から 5 の整数値になるようにする。

3.2 STS を報酬とする強化学習

翻訳モデルには Attention 機構を持った LSTM ベースの Encoder-Decoder モデル [4] を用いる。学習時のモデルの出力評価は STS に転移学習した後パラメータを固定した BERT によって行う。得られる STS 評価値は翻訳モデルのパラメータについて微分できないので、REINFORCE でその値を報酬として利用する。

前述の通り BERT は連続値を出力することには向かないが、そのまま最も尤度が高い STS の評価値を使ってしまうと出力文と正解文の比較の基準が離散的になってしまう。そこで、BERT の出力を評価値クラスの確率分布に見立て、STS 値の期待値を計算することにした。これにより、連続値で意味的類似度の目安値を得ることができるようになった。

一般に REINFORCE による強化学習は、報酬値の分散が大きいため、特に学習初期において損失の減少が遅いことが知られている。それを踏まえて、初めに

ある程度 Cross-entropy 損失で教師あり学習をしてから REINFORCE を用いた学習に移行する。

強化学習をしている間も、多少は Cross-entropy を損失に用いないと学習が不安定になることがあるため、実際に使う損失関数は

$$L = \lambda L_{CE} + (1 - \lambda)L_{RL} \quad (3)$$

のように式 (1) と式 (2) の線形和となる。 $\lambda \in [0, 1]$ は Cross-entropy による損失と強化学習による損失の比率を決める値で、一般に非常に小さな値が選ばれる。

また、強化学習時の損失関数の分散を安定させるため、Decoder の各出力ステップでの隠れ状態から未来の報酬を予測するベースライン報酬モデル $r_b(s_t)$ を設定した。 r_b は線形層 1 層とシグモイド関数からなり、式 (4) のような損失関数を用いた。損失関数は式 (4) のようなを設定した。

$$L_{BSE} = \|r_b(s_t) - r(\hat{y})\|^2 \quad (4)$$

ただし、 r_b は生成モデル π_θ とは異なるパラメータを持ち、その学習時の誤差は生成モデルには伝搬しない。

4 実験

4.1 実験データ

事前学習済みの BERT モデルは BERT-base-uncased¹ を用いた。また、STS の学習には STS-Benchmark (STS-B)² を用いた。これは英語のニュース・画像のキャプション・チャット文からなるデータセットで、学習、検証、評価用にそれぞれ約 5,700、1,500、1,400 文対と人手による文対の STS 評価が含まれる。

翻訳の学習には、STS-B のドメインとの親和性を考え、WMT の複数メディア翻訳タスクで使われる画像と多言語キャプションのデータセット multi30k-dataset³ を選んだ。実際のタスクにはそこから task1 の英語とドイツ語のキャプション文のデータのみを利用した独英翻訳とした。このデータセットには学習用 29,000 文と検証用の約 1,000 文に加えて、1,000 文程度の評価用セットが複数含まれている。今回は評価用セットとして flickr2017 と mscoco2017 を用いた。

BERT のトークナイズは単語に分けた後、wordpiece によるサブトークン化を行った [2] が、翻訳モデルで

表 1: 独英翻訳の BLEU スコア

モデル	flickr2017	mscoco2017
ML	23.06	16.70
RL	24.90	18.77

は単語ごとに区切るのみした。また、翻訳モデルの出力トークン列を BERT モデルに入力する際には、一度トークン列を文に戻してから BERT 用のトークナイズを施した。翻訳出力と比較するための正解文はとくに前処理はせず、そのまま BERT 用のトークナイズを施している。翻訳モデルでは語彙数を英語とドイツ語それぞれで 10,000 語に制限した。

4.2 学習設定

BERT モデル BERT-base を用い、バッチ数 32、学習率 2×10^{-5} の Adam で、最大 10 エポック STS-B の学習データに対し転移学習させた。このとき検証用データで STS 評価の正解値との平均二乗誤差が最小だったものを以降の翻訳学習の評価器として利用した。

翻訳モデルは、Encoder 側は双方向 LSTM、Decoder 側は順方向 LSTM とし、各 LSTM の層数は 1、隠れ状態は 256 とした。パラメータ更新は momentum 値 0.75 の SGD で行い、学習率の減衰率を 0.5 に設定した。また、ドロップアウト率は一律 0.2 に設定した。

学習率については教師あり学習で 1.0、強化学習で 0.01、報酬予測モデルは 1.0×10^{-3} とした。教師あり学習は最大 30 エポック、強化学習は最大 20 エポック行い、検証データに対する平均 STS スコアが最良だったものを評価に用いた。式 (3) の λ は 0.005 とした。

4.3 結果

Cross-entropy 損失のみで学習させた翻訳モデル (ML) とそこから更に STS を用いた強化学習を行った翻訳モデル (RL) の BLEU スコアを表 1 に、検証用データに対する各モデルの出力の例を表 2 に示す。

5 考察

表 1 を見ると、双方のコーパスで ML と比べて RL でスコアが向上していることがわかる。その上昇値の大きさから STS 評価値と BLEU に正の相関があるこ

¹<https://github.com/google-research/bert>

²<http://ixa2.si.ehu.es/stswiki/index.php/STSBenchmark>

³<https://github.com/multi30k/dataset>

表 2: 出力文の例

モデル	出力文
ML	a man and a female <UNK> examining the very examine on the floor.
RL	a man and a female is examining a very large tiger around the floor.
正解文	a man and woman veterinarian checking on a very large tiger laying on the ground.

とが推測できる。BLEU は単語単位でなく、連続する単語の集合の出現の有無で文を比較するため、Cross-entropy 損失と比べると、単語のズレをある程度許容する。それを踏まえると、より緩い文の評価尺度という点で STS と相関を持っていても不自然ではない。

学習中に利用した検証用データに対する出力を見ると、RL では ML よりも未知単語トークン (UNK) の出力が少なかった。これは、STS 評価値の特性が原因になっていると思われる。

STS の評価値は、文同士が酷似していても、単語の置換によって文全体の意味が異なるような場合に、大きく下がる傾向がある。例えば、STS-B の学習データの中に “A man is playing a guitar.” と “A girl is playing a guitar.” という文対が存在する。単語単位で比べると “man” と “girl” が異なるのみだが、STS のスコアは 2.80 と低めに設定されている。逆に、文の意味をさほど変えない単語の欠落等ではあまり STS 評価値は下がらない。例として、“A panda bear is eating some bamboo.” と “A panda is eating bamboo” では 2 単語抜けていても STS の評価値は 4.20 である。

表 2 の出力例のように、ML の出力に UNK トークンがあるのに、RL の出力からは UNK トークンが消えているときは、正解文に低頻出単語が含まれていることが多かった。正解文には直接 BERT 用のトークナイズを施したために UNK トークンが存在しない一方で、モデルの出力文では UNK トークンを含んだ文を BERT 用にトークナイズした。結果として、UNK トークンが文の意味が変わってしまうと判定されて取り除かれ、UNK がなくなった後の文は低頻出語がなくても正解文との類似度は変わらないと評価されたと思われる。この現象はモデル出力と正解文のトークナイズの差異によるところが大きいと考えらえるので、トークナイズを統一した場合にはまた違った結果が見られる可能性がある。

このように、STS が Cross-entropy 損失よりも抽象的に生成文と正解文の類似度を比較できている形跡は見つかった。しかし、それが出力文中の未知語と同義語の入れ替えや、意味を保持した文構造の変化など、有益に働いている例はあまり見られなかった。

6 おわりに

本研究では STS タスクに対してファインチューンした BERT モデルを評価器として用いることで、意味的類似性を報酬した強化学習による文生成の仕組みを提案した。強化学習による BLEU の向上が確認できたものの、出力文を見ても STS が有効に働いたと理由付けられる例はあまり見られなかった。今後の研究課題として、より大規模なコーパスでの実験や、BLEU で強化学習をしたモデルとの比較により、STS の文評価の有用性を詳しく検証する必要がある。

参考文献

- [1] Cer et al. SemEval-2017 Task 1: Semantic Textual Similarity Multilingual and Crosslingual Focused Evaluation. In *SemEval*, 2017.
- [2] Devlin et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv:1810.04805 [cs]*, 2018.
- [3] Hochreiter et al. Long Short-Term Memory. *Neural Computation*, 9(8), 1997.
- [4] Luong et al. Effective Approaches to Attention-based Neural Machine Translation. In *EMNLP*, 2015.
- [5] Papineni et al. BLEU: a method for automatic evaluation of machine translation. In *ACL*, 2002.
- [6] Peters et al. Deep Contextualized Word Representations. In *NACL*, 2018.
- [7] Ranzato et al. Sequence Level Training with Recurrent Neural Networks. *arXiv:1511.06732 [cs]*, 2015.
- [8] Sutskever et al. Sequence to Sequence Learning with Neural Networks. In *NIPS*. 2014.
- [9] Vaswani et al. Attention is all you need. In *NIPS*. 2017.
- [10] Wang et al. GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding. In *EMNLP Workshop BlackboxNLP*, 2018.
- [11] Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*, 8(3), 1992.