ノンパラレルデータを用いたパラレルデータの作成に基づくテキストセンチメント変換

Yingxuan YUAN†, Sheng XU†, Jiyi LI††, Fumiyo FUKUMOTO††, and Kentaro GO††

† 山梨大学工学部コンピュータ理工学科　〒 400–8511 山梨県甲府市武田四丁目 3-11
†† 山梨大学大学院総合研究部工学域　〒 400–8510 山梨県甲府市武田四丁目 3-37
E-mail: †t19cs012@gmail.com, ††{g22dts03,jyli,fukumoto,go}@yamanashi.ac.jp

あらまし　自然言語処理技術の発展によって，実用性が向上した機械翻訳や質問応答システムなどは，人々の日常生活で欠かせない存在となりつつある．例えば，サポートチャットシステムを導入することで，顧客からの問い合わせに自動で対応でき，業務効率が大幅に改善された．しかし，これらの自然言語処理技術によって生成された文書は，ユーザーや文書の内容を考慮せず，特徴のないもしくは相応しくない文書が生成してしまう．これらの改善方法としては，ネガティブな文とポジティブな文との間で意味内容を保ちながら変換するタスク，テキストセンチメント変換が挙げられる．近年の研究では，自然言語分野において事前学習モデルの活用は進んでいるが，テキストセンチメント変換分野では教師データとして使われるパラレルデータがないのため事前学習モデルの活用が進めていない．そこで本研究は，事前学習モデルの活用を目的として，ノンパラレルデータを用いたパラレルデータの作成法とパラレルデータ拡張法によるテキストセンチメント変換手法を提案する．実験評価では，本手法の有効性を検討し，考察を行った．

キーワード　テキストセンチメント変換，テキストスタイル変換，自然言語生成, 事前学習，データ拡張

1　はじめに

テキストスタイル変換 (Text Style Transfer) とは，意味内容を保ちながら，文書のスタイルを変換することである．テキストスタイル変換には様々なサブタスクがある．例えば，難しい文書と専門外の方でも理解できる文書の間で変換するエクスパティーズ変換 [4](Expertise Style Transfer)，フォーマルな文と非フォーマルな文の間で変換するフォーマリティー変換 [5](Formality Style Transfer)．シェイクスピア風の英語と現代標準英語の間で変換するオーサーシップ変換 [6](Authorship Style Transfer) などがある．テキストセンチメント変換 [3] は，テキストスタイル変換分野の研究において最も注目されているタスクである．テキストセンチメント変換とは，ネガティブな文とポジティブな文との間で意味内容を保ちながら変換するタスクである．

自然言語処理技術の発展によって，実用性が向上した機械翻訳や質問応答システムなどは，人々の日常生活で欠かせない存在となりつつある．例えば，外国語による文章を理解するとき，辞書ではなく Google 翻訳ツールに頼る人が増えている．また，ネット通販などの EC サイトのサポートは，サポートチャットシステムを導入することで，顧客からの問い合わせに自動で対応でき，業務効率が大幅に改善された．しかし，これらの自然言語処理技術によって生成された文書は，ユーザーや文書の内容を考慮せず，特徴のないもしくは相応しくない文書が生成してしまい，これらの改善が期待されている．例えば，ビジネス文書を外国語に翻訳するとき，よりフォーマルな表現を使う．チャットボットと話すとき，よりポジティブや分かりやすい表現を使うことで，相手を配慮した会話を生成する．

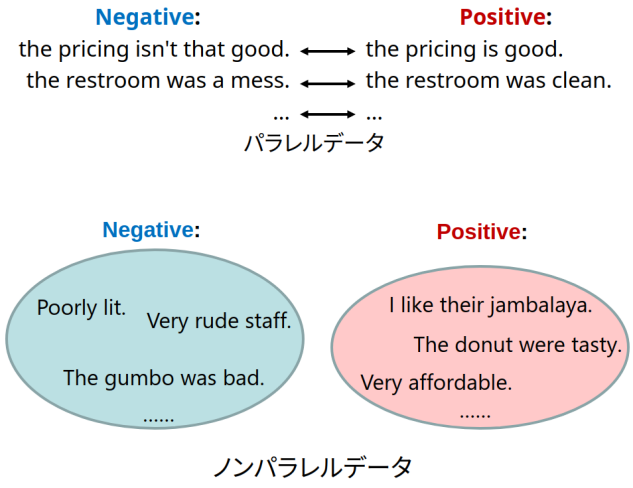

図 1　パラレルデータとノンパラレルデータの例

近年，深層学習により自然言語タスクを解く際に，ラベル付きデータやパラレルデータを用いた事前学習モデルの活用は優れた性能を達成している．しかし，テキストセンチメント変換タスクにおいて，互いに言い換えであるネガティブな文とポジティブな文のペアで構成されたパラレルデータがなく，同じ意味を持ったネガティブな文とポジティブな文が存在しても，それらが対応付けられていないノンパラレルデータしか利用できないため，事前学習モデルの活用は困難である．パラレルデータとノンパラレルデータの例を図 1 に示す．近年の研究では，主にノンパラレルデータでの教師なし学習を用いた手法が盛ん

に行われている．例として Li ら [1] の手法は，ネガティブな文からポジティブな文への変換を行うとき，検索と統計手法を用いて，ネガティブな文におけるネガティブな単語を適切なポジティブな単語に書き換えし，深層学習モデルで書き換えた文の流暢性を改善する手法である．しかしこの手法には，テキストの変換ごとに書き換え用のポジティブな単語の検索を行うため，大規模なデータセットでは検索に要する時間が長くなるという欠点があり，実用性に限界があると言える．また，敵対的生成ネットワークや強化学習によるセンチメント変換について多くの研究が行われていたが，複数のロスを最適化することは難しいことや，生成文の流暢性が低いことなどが問題点として挙げられる．

本研究では，ネガティブな文からポジティブな文への変換タスクを注目し，事前学習モデルの活用を目的として，ノンパラレルデータを用いたパラレルデータの作成法とパラレルデータ拡張法によるテキストセンチメント変換手法を提案する．本研究では，以下の貢献を果たした．(1) ノンパラレルデータを用いたパラレルデータの作成法を提案し，データセットを作成した．(2) パラレルデータを用いたパラレルデータの拡張法を提案し，データセットを作成した．(3) パラレルデータの作成法によるデータセットを用いたモデルを訓練し，3 つの観点から評価実験を行い，手法の有効性を示した．また，パラレルデータの作成法によるデータセットとパラレルデータの拡張法によるデータセットを用いたモデルを訓練し，評価実験と考察を行った．

2 関 連 研 究

2.1 事前学習モデル

事前学習とは，特定のタスクにモデル適応する前に，大規模なテキストデータで汎用的な知識を学習させる手法です．特定のタスクにモデル適応することは，微調整またはファインチューニングと呼ばれている．事前学習モデルを使用することで，少数の教師データでも高精度な性能を達成することができ，テキストを生成するときより自然な文書が作成できる．近年の自然言語処理分野における事前学習モデルは，Transformer [7] をベースにしたモデルである．Transformer は，Attention を用いたエンコーダ・デコーダネットワークアーキテクチャである．エンコーダ・デコーダネットワークのエンコーダは，入力系列の意味を捉えた特徴量の表現を算出し，デコーダは，エンコーダによって算出された特徴を利用し，タスクに応じてテキストを生成する．Attention とは，文脈を考慮して単語を解釈する技術である．本節では，本手法で使用するテキスト分類用事前学習モデル BERT とテキスト生成用事前学習モデル BART について説明する．

2.2 BERT

BERT(Bidirectional Encoder Representations from Transformers) は 2018 年に Google の Devlin ら [8] により提案された事前学習モデルである．BERT の入力は，1 つもしくは 2 つの文書であり，文書はトークン化したものである．入力系列の先頭に分類用の特別トークン <cls> を付加し，2 つの文章を間で特別トークン <sep> を使用し，入力系列の末尾にも特別トークン <sep> を使用する．BERT は，Transformers のエンコーダをベースにしたモデルであり，2 つの手法を用いて事前学習を行う．1 つ目の事前学習手法 Masked Language Model(MLM) は，文のトークンをマスクトークン <mask> もしくは別のトークンに置き換えた文に対し，そのマスクトークンの元単語を予測するタスクである．2 つ目の事前学習手法 Next Sentence Prediction(NSP) は，2 つの文章を入力し，それは前後関係であるかどうかを予測するタスクである．BERT の出力は，各トークンの文脈を考慮したベクトル表現である．分類タスクにおける BERT の微調整は，出力である特別トークン <cls> のベクトル表現を全結合層に入力し，Softmax 関数を介して各クラスの確率を出力する．本研究における BERT モデルは，bert-base-uncased[1]を使用する．

2.3 BART

BART(Bidirectional Auto-Regressive Transformer) は 2019 年に Facebook の Lewis ら [9] により提案された事前学習モデルである．BART は Transformer をベースにしたモデルであり，エンコーダは BERT となり，デコーダは自己回帰モデルであるため，文書生成タスクに適している．自己回帰モデルとは，系列の各時点におけるトークンが，過去の時点のトークンに基づいて予測されるモデルである．最初に入力されるトークンは特別トークン <s> である．BART には，以下の 5 つのノイズを加える手法が用意されている．

- Token Masking: ランダムに単語をマスクトークン <mask> に置き換える．
- Token Deletion: ランダムに単語をマスクトークン <mask> に置き換える．
- Text Infilling: 複数の単語の並びを一つの <mask> で置き換える．
- Sentence Permutation: 複数の文からなる文書について，文章の順番をシャッフルする．
- Document Rotation: 文章から一つの単語を選び，その単語が一番初めになるように，文章を回転させる．

BART は，これらの手法によりノイズを加えた文書をノイズを加える前の元の文書に復元する事前学習タスクを行う．本研究における BART モデルは，facebook/bart-base[2]を使用する．

3 提 案 手 法

提案手法の流れを図 2 に示す．まず，ノンパラレルデータからパラレルデータを作成する．次に，作成したパラレルデータをデータ拡張する．最後に，作成されたパラレルデータを用いて，生成用の事前学習モデルを微調整し，入力文がポジティブ

1：https://huggingface.co/bert-base-uncased
2：https://huggingface.co/facebook/bart-base

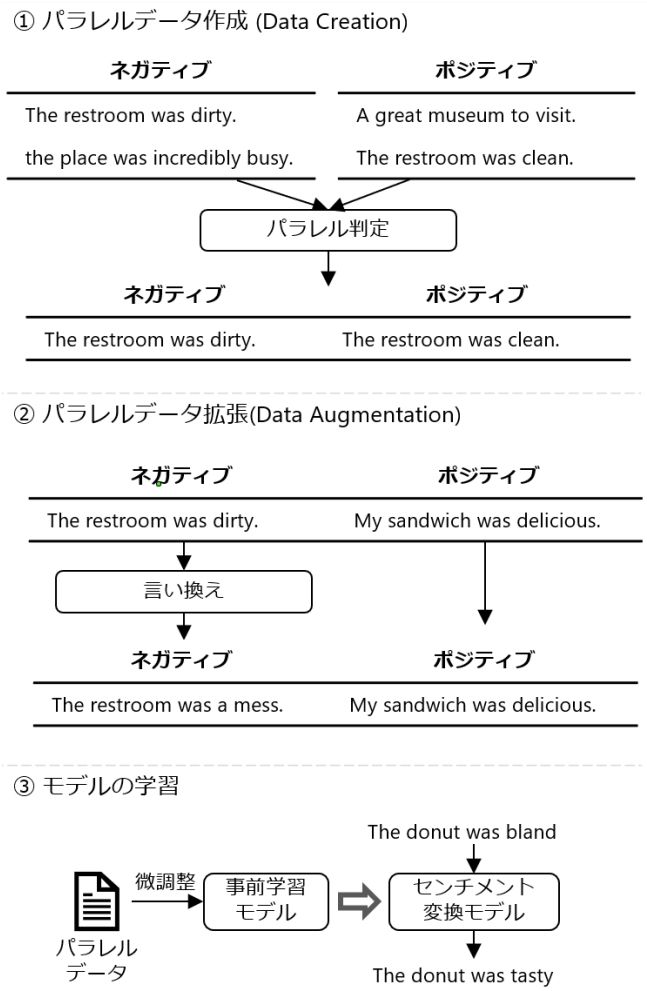

図 2　手法の流れ

な文に変換されるように学習する．

3.1　パラレルデータの作成

本研究におけるノンパラレルデータは，Li ら [1] が作成したデータセットのテストサブセットを使用する．このノンパラレルデータには 177,218 文のネガティブな文と 266,041 文のポジティブな文で構成されている．まず，ノンパラレルデータをデータクリーニングをする．ノンパラレルデータでは同じテキストが複数存在することがあるため，重複するデータを削除する．また，ネガティブデータセットではポジティブな文が混在し，ポジティブデータセットでもネガティブな文が混在することがあり，このようなノンパラレルデータからパラレルデータを作成すると，ポジティブな文とポジティブな文のようなパラレルデータになり，最終的にモデルの学習に影響を与えるため，混在するデータを削除する．削除する方法は，既存のセンチメント分類器 cardiffnlp/twitter-roberta-base-sentiment-latest[3] を使用し，ネガティブデータセットにおけるポジティブな文と識別された文を削除し，ポジティブデータセットにおけるポジティブな文と識別された文を削除する．最後に，認識できない・トークン化できない記号を削除する．データクリーニング後のノンパラレルデータは 142,559 文のネガティブな文と 210,571 文のポジティブな文に削減した．

削減後のノンパラレルデータを用いて，パラレルデータの作成を行う．あるポジティブな文とあるネガティブな文が互いに言い換え表現である場合，そのポジティブな文とネガティブな文の文対をパラレルデータとして抽出する．２文が互いに言い換え表現であるかどうかは言い換え表現分類器を用いる．言い換え表現分類器は，４つの言い換え表現分類用データセット qqp [10], mrpc [11], paws [12], stsb [13] を用いて，事前学習分類モデル bert-base-uncased を微調整したモデルである．言い換え表現分類器の出力層では，softmax を使い，言い換え表現である確率を出力する．本研究では，確率が 0.5 以上であれば入力された２文が互い言い換え表現と見なされる．しかし，全てのノンパラレルデータについて言い換え表現である確率を算出すると，14 万 × 21 万 回の計算を行う必要となり，膨大な時間が要する．そのため，ネガティブな文とポジティブな文が互いに同じ名詞や動詞を１つ以上持つ場合のみ言い換え表現である確率を計算する．また，ある 1 つネガティブな文に対し，同じ名詞や動詞を持つポジティブな文が 64 文を超える場合には，そのうちの 64 文をランダムに選出し確率を計算し，確率が一番高いジティブな文を選出する．文書における名詞と動詞の推定は，既存の品詞推定モデル vblagoje/bert-english-uncased-finetuned-pos[4]を使用する．最終的に作成したパラレルデータは 8,891 文対である．

3.2　パラレルデータの拡張

データ拡張とは，既存のデータセットを用いて，データをさらに増やすことである．データ拡張は，データの多様性を増やすことで，過学習を減らし，モデルの精度や汎化能力を向上させる．本節は，2.1 節で作成されたパラレルデータを対象とし，ネガティブな文の言い換え文を生成することで，新しいパラレルデータを作成する．言い換え文の生成は，パラフレーズ生成器を用いる．パラフレーズ生成器は，言い換え表現生成用データセット ParaBank2 [14] と MSCOCO [15] のキャプションデータを用いて，事前学習生成モデル facebook/bart-base を微調整したモデルである．また，生成した言い換え文が原文と一致した場合はその言い換え文による拡張パラレルデータは作成しない．最後に，生成したネガティブな言い換え文と生成文の原文に対応するポジティブな文と対応付け，拡張パラレルデータを作成する．最終的に作成したパラレルデータは 8,417 文対である．

3.3　モデルの学習

モデルの学習では，3.1 や 3.2 で作成したパラレルデータを用いて，テキスト生成用の事前学習モデルを微調整し，入力文がポジティブな文に変換されるように学習する．

3：https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment-latest 4：https://huggingface.co/vblagoje/bert-english-uncased-finetuned-pos

表 1 実験結果

	Att	Con						Flu
	Acc ↑	ROU ↑			BS ↑			ppl ↓
		inp	ref	geo	inp	ref	geo	
human	82.0%		55.48			92.72		373.21
D&R	**91.6%**	57.55	38.94	47.34	91.47	89.53	90.49	274.52
SST	72.2%	**74.89**	**46.68**	**59.13**	**93.26**	89.96	91.60	1156.56
Cross	79.6%	50.08	33.44	40.92	89.74	88.04	88.89	657.21
DC	82.0%	55.95	43.01	49.06	92.99	**91.47**	**92.23**	**192.01**
DC+DA	86.4%	49.43	39.13	43.98	92.90	91.05	91.62	192.41

4 実　　験

実験は，本手法のモデルを作成し，テスト用データ，本手法の出力，比較手法の出力について評価し，比較を行う．

4.1 データセット

本手法のモデルは，3.1 節のパラレルデータによるセンチメント変換モデル (以下「DC」という) と，3.1 節のパラレルデータと 3.2 節のパラレルデータによるセンチメント変換モデル (以下「DC+DA」という) を作成する．本研究はネガティブな文からポジティブな文への変換のため，教師データの原文がパラレルデータのネガティブな文とし，ラベル文がパラレルデータのポジティブな文である．また教師データは，学習用と検証用に 8:2 の割合で分割する．テスト用のデータは，Li ら [1] が作成したテスト用データ (原文がネガティブな文，ラベルが原文に対応する人手によるポジティブな文) を使用する．実験に使用するデータセットを表 2 に示す．

表 2　データセットの概要

	DC	DC+DA
訓練	7112	13846
検証	1779	3462
テスト (Li ら)	500	

4.2 実験設定

モデル DC と DC+DA は Huggingface と Pytorch を用いて実装し，モデルは事前学習モデル facebook/bart-base を使用する．Adam アルゴリズムを用いて，バッチサイズ 64 でモデルを学習する．初期学習率を 0.00007 に設定し，2 エポックの学習を行った．

4.3 評　　価

関連研究に従い，自動評価はポジティブな文への変換精度 (Att)，意味内容の保存度 (Con)，生成文の流暢性 (Flu) という 3 つの観点から評価する．Acc は既存のセンチメント分類器 cardiffnlp/twitter-roberta-base-sentiment-latest で評価する．Con は ROUGE-L の F1 スコア (ROU) と BERTScore(BS) という 2 つの指標で評価する．ROUGE-L [16] は最長一致単語列の長さをもとに単語の一致度を評価する指標であり，BERTScore [17] は 2 文の分散表現ベクトル集合による類似度を評価する指標である．また，Con には inp, ref, geo という 3 種類がある．inp は原文とモデルによる変換文の Con,ref はモデルによる変換文と人手による変換文 (ポジティブな文) の Con，geo は inp と ref の幾何平均である．Flu はパープレキシティ (ppl) で評価する．ppl とは，言語モデル (gpt-2 [18]) による予測分布と評価対象であるデータ分布との不一致を示す指標である．

比較手法は，Li ら [1] の DeleteAndRetrieve(D&R)，Lee [2] の SST，Shen ら [3] の Cross-align(Cross) である．

4.4 結　　果

テスト用データ human，D&R, SST, Cross, 本研究の DC と DC+DA におけるテスト用データについての評価結果を表 1 に示す．表の結果から，DC の変換精度と意味内容の保存度は人手による変換文に近い上，流暢性の性能が向上し，事前学習モデルによる手法が有効であることを示された．しかし，DC よりデータ拡張を用いた DC+DA は，DC より変換精度が向上する一方，意味内容の保存度が低下した．その要因として，増加したデータによりモデルの変換精度を改善した一方，言い換え文により教師データ全体の意味内容の保存度が低くなったので，モデルの学習に影響を与えたと考えられる．

5 ま と め

本研究では，事前学習モデルの活用を目的として，ノンパラレルデータを用いたパラレルデータの作成法とパラレルデータ拡張法によるテキストセンチメント変換手法を提案した．評価実験により，事前学習モデルとパラレルデータの自動作成法による手法の有効性を示した．また，パラレルデータ拡張法によるモデルが内容の保存度が低い問題について考察を行った．今後の課題として，テキストスタイル変換タスクに対して本手法を適用すること，パラレルデータ作成時に使われていないデータを活用すること，よりよいパラレルデータ拡張法による精度

の向上などに取り組む予定である．

謝　　辞

本研究の一部は，科研費 20K11904 の支援を受けて実施したものである．

文　　献

[1] Juncen Li, Robin Jia, He He, and Percy Liang. Delete, Retrieve, Generate: a Simple Approach to Sentiment and Style Transfer. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), pp 1865–1874 (2018)

[2] Joosung Lee. Stable Style Transformer: Delete and Generate Approach with Encoder-Decoder for Text Style Transfer. In Proceedings of the 13th International Conference on Natural Language Generation, pp 195–204(2020).

[3] Shen, Tianxiao, Tao Lei, Regina Barzilay, and Tommi Jaakkola. Style transfer from non-parallel text by cross-alignment. In Advances in neural information processing systems, pp 6830–6841(2017).

[4] Yixin Cao, Ruihao Shui, Liangming Pan, Min-Yen Kan, Zhiyuan Liu, and Tat-Seng Chua. Expertise Style Transfer: A New Task Towards Better Communication between Experts and Laymen. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp 1061–1071(2020).

[5] Fadi Abu Sheikha and Diana Inkpen. Generation of Formal and Informal Sentences. In Proceedings of the 13th European Workshop on Natural Language Generation, pp 187–193(2011)

[6] Wei Xu, Alan Ritter, Bill Dolan, Ralph Grishman, and Colin Cherry. Paraphrasing for Style. In Proceedings of COLING 2012, pp 2899–2914(2012)

[7] Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA, pp 5998–6008(2017).

[8] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pp 4171–4186(2019).

[9] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp 7871–7880(2020).

[10] Quora Question Pairs, kaggle, 2017, `https://www.kaggle.com/c/quora-question-pairs`

[11] William B Dolan and Chris Brockett. Automatically constructing a corpus of sentential paraphrases. In Proceedings of the Third International Workshop on Paraphrasing(2005).

[12] Yuan Zhang, Jason Baldridge, and Luheng He. PAWS: Paraphrase Adversaries from Word Scrambling. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 1298–1308(2019).

[13] Daniel Cer, Mona Diab, Eneko Agirre, Iñigo Lopez-Gazpio, and Lucia Specia. SemEval-2017 Task 1: Semantic Textual Similarity Multilingual and Crosslingual Focused Evaluation. In Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017), pages 1–14(2017)

[14] J. Edward Hu, Abhinav Singh, Nils Holzenberger, Matt Post, and Benjamin Van Durme. Large-Scale, Diverse, Paraphrastic Bitexts via Sampling and Clustering. In Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL), pages 44–54(2019).

[15] Tsung-Yi Lin, Michael Maire, Serge J. Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft coco: Common objects in context. In ECCV.(2014)

[16] Abigail See, et al. Get To The Point: Summarization with Pointer-Generator Networks. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL), Volume 1: Long Papers, pages 1073–1083(2017)

[17] Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q Weinberger, and Yoav Artzi. Bertscore: Evaluating text generation with bert. In Proceedings of the International Conference on Learning Representations.(2020)

[18] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners. OpenAI Blog, 1(8)(2019).