

文法項目を指定した第二言語学習用応答生成

岡野 裕紀¹ 船越 孝太郎¹ 永田 亮^{2,3} 奥村 学^{1,3}

¹ 東京工業大学 ² 甲南大学 ³ 理化学研究所

{okano, funakoshi, oku}@lr.pi.titech.ac.jp

nagata-nlp2023@ml.hyogo-u.ac.jp

概要

語学学習やライティング支援など教育分野において言語処理技術の活用が注目されている。本稿では、対話形式の文法学習に向けて、対話の応答文に特定の文法項目を含める言語生成に取り組む。生成手法として、強化学習によって生成モデルを訓練する手法と、大規模汎用言語モデルに少数サンプルの入力例を与えて生成する手法を検討する。そして、強化学習の報酬関数として、文法項目の含有に加え、多様性と流暢性の観点の導入を検討する。実験の結果、どちらの手法でも生成が可能であること、報酬関数としては多様性よりも流暢性の考慮が重要であることが確認された。

1 はじめに

言語処理技術を教育に活用する一つの方向性として、語学学習に対話システムを活用することが注目されており、多くの研究で語学学習者の対話相手として導入され、その効果が検証されてきた。先行研究 [1, 2, 3] からは、対話システムを語学教育に導入する利点となる特徴として、時間を問わず利用できる、語学学習をより手軽で身近なものにできること、多くの人々が慣れ親しんだチャットベースのアプリケーションなどに容易に組み込み可能であり、使用の敷居が低いこと、個人の学習状況に合わせて最適化できることなどが挙げられる。これらの特徴はいずれも、対話システムが語学学習のパートナーとして有用であり、より良い選択肢の一つとなりうることを示唆していると考えられる。

また、Nagata ら [4] により、単純な対話システムである ELIZA [5] を用いた場合でも学習者の文産出量が増加することが示された。さらに、彼らの実験では、チャットボットの文に出現する単語を学習者が採用した結果が報告されており、学習者が知らない言葉や自分では使わない傾向にある言葉を学習する

の対話システムが有効であることが示唆された。

本稿では、対話モデルの応答に学習者が習得したい文法項目を含ませることにより、語学学習における対話モデルの可用性の向上に取り組む。ここで、文法項目とは中学・高校の英語学習で扱われる、現在完了形、仮定法、関係詞節、不定詞などの項目を指す。対話システムとの対話を通して英文産出量を増やすだけでなく、学習したい文法項目に触れる機会を増やすことで優れた学習効果が期待できる。

所望の表現を含む応答文の生成を行う研究として、条件を満たすテキストを生成するように、事前学習済み言語モデルを誘導する補助モジュールを構築する手法 [6] や、大規模言語モデルを訓練する際に、制御コードと一緒に学習させ、それらのコードを用いて生成を制御する手法 [7] などがある。これらの手法は教師あり学習に基づいているため、教師付きデータセットが必要である。しかし、言語処理技術の教育応用については多くの資源が整備されているとは言えず、特に、文法項目に関する大規模なラベル付き対話データセットが充実していないという課題がある。したがって、教師あり学習を用いるこれらの手法においては、訓練に用いるデータセットの不足が懸念される。

そこで本稿では、大規模なラベル付きデータセットを必要とせず所望の文法項目を含む応答を生成する手法として、強化学習を用いる手法と、大規模汎用言語モデルに、プロンプトと呼ばれるタスクの指示と少数のサンプルを含んだ入力を与える手法を検討する。強化学習の報酬関数には、文法項目が含まれるかどうかを判定する分類器を用いる。また、学習が文法項目のみに偏って応答文の質が落ちないように、応答文の多様性、流暢性に関する報酬関数を設計し、それらの組み合わせの効果を検討する。

2 関連研究

強化学習による文生成 強化学習は、行動系列に

対して報酬関数によって与えられる非即時的な評価に基づき、状態に応じた最適な行動方策を獲得する機械学習の枠組みである。出力したトークンを行動と捉えることで、言語生成を強化学習の問題として扱うことができる。ニューラルネットで生成モデルを構成する場合、REINFORCE [8] などの方策勾配法に基づく手法によって、出力に対する報酬の期待値を最大化するようにパラメータを最適化できる。報酬関数でベースライン関数との差分を考慮することで報酬と確率勾配の分散を減らし学習を安定させる、ベースライン付き強化学習がよく用いられる。

報酬関数の設計はタスクによって様々だが、従来の教師あり学習における誤差関数とは異なり、微分不可能な関数を利用できることから、BLEU, ROUGE などの評価尺度を用いるテキスト生成タスクでよく利用されている [9, 10, 11]。深層学習に基づく言語生成では一般に誤差関数に交差エントロピーが用いられ、目的関数と評価尺度が異なることになるが、報酬関数に評価尺度を取り入れることで、そのギャップを緩和できる。

大規模汎用言語モデルによる生成 近年、タスクの説明とその例を含むプロンプトを事前学習済みの生成型言語モデルに入力として与えることで言語モデルの出力を制御する手法が盛んに研究されている [12, 13, 14]。特に、GPT-3 [15] は、タスクを説明する自然言語の指示のみが与えられる zero-shot、指示に加えてタスクの例が1つ与えられる one-shot、指示に加えてタスクの例が複数与えられる few-shot、いずれの場合にも学習データに含まれない例に対して、勾配の更新や fine-tune を行わずとも有望な結果を示した。特に、few-shot の実験設定では、CoQA や TriviaQA において他の fine-tune されたモデルと同等、もしくはそれ以上の値を達成している。

3 文生成手法

3.1 強化学習を用いた文生成

3.1.1 Self-critic 強化学習

まず、生成モデルの学習に用いる Self-critic 強化学習 [16] について説明する。文生成における強化学習では、生成モデルによってサンプリングされた文 \hat{x} に対して、報酬関数 R による報酬を係数とした尤度を最適化することで、高い報酬を得られる文が生成されやすくなるようにする。Self-critic 強化学習は、ベースライン関数として、生成モデル自身の貪

欲法による生成に対する報酬を設定する手法であり、以下の式で定義される。

$$\nabla_{\theta} L_{\text{SCST}} \propto -(R(\hat{x}) - R(x^*)) \nabla_{\theta} \log p_{\theta}(\hat{x}) \quad (1)$$

ここで、 x^* は生成モデルから貪欲法で生成された出力を表す。貪欲法により得られた文 x^* と比較して、サンプリングにより得られた \hat{x} の報酬の値が大きい場合、 \hat{x} の生成がより促進され、報酬の値が小さい場合、生成が抑制される。

本稿では3つの異なる報酬関数を検討する。一つ目は、主目的である文法項目を含めた生成のための報酬関数である。他に、文法項目のみに偏った学習を緩和するための関数として生成文の多様性と流暢性に関する報酬関数をそれぞれ用意する。また、batch 単位で処理を行うため、 M 文の入力文集合 \mathbf{C} に対する応答文集合 \mathbf{S} に対して報酬を計算する。

3.1.2 文法項目含有のための報酬関数

与えられた文が文法項目を含んでいるかどうかを判定する分類器を $F_g()$ として、報酬関数 $R_g()$ を以下のように定義する。

$$R_g(\mathbf{S}) = \frac{1}{|\mathbf{S}|} \sum_{s \in \mathbf{S}} F_g(s) \quad (2)$$

この値が大きいほど、 \mathbf{S} 中の文が所望の文法項目を用いている割合が高いことを示す。

$F_g()$ は、BERT [17] を用いて、入力文が文法項目を含むかどうかの判定を行う。入力文に [CLS] トークンを付与して BERT の入力とし、次式のように [CLS] トークンの埋め込み表現 $\mathbf{h}_{[\text{CLS}]}$ に対し文法項目を含有するスコアを予測する。

$$F_g(s) = \sigma(\mathbf{w}^{\top} \mathbf{h}_{[\text{CLS}]} + b) \quad (3)$$

ここで、 σ はシグモイド関数、 \mathbf{w}, b は学習可能なパラメータである。 $F_g(s)$ が閾値を上回る場合に文法項目を含有すると判定し、 s を項目含有文とよぶ。

3.1.3 多様性のための報酬関数

テキストの多様性を測る指標として n-gram ベースの指標である Distinct-N [18] を用いる。値が大きいほど多様性が高く、以下のように定義される。

$$R_d(\mathbf{S}) = \frac{\text{Count}(\text{unique}(\text{n-gram}(\mathbf{S})))}{\text{Count}(\text{n-gram}(\mathbf{S}))} \quad (4)$$

n-gram(\mathbf{S}) は文集合 \mathbf{S} 中のすべての n-gram を求める関数であり、unique はその集合から重複を除去する関数である。

3.1.4 流暢性のための報酬関数

対話の文脈における応答文の流暢性を考えるため、 C に対する S の尤度を、パラメータ θ を持つ事前学習済みの対話モデルにより計算し、流暢性のスコアとして定義する。

$$R_f(C, S) = \frac{1}{M} \sum_{i=1}^M P_{\theta}(s_i | c_i) \quad (5)$$

ここで、 s_i は c_i に対する応答文である。したがって、応答文 s_i の流暢性が高いほど、報酬の値が高くなる。

3.2 Few-shot プロンプトを用いた文生成

Few-shot プロンプトを用いた文生成では、大規模言語モデルに対して、図 1 のように、タスクの説明とそのタスクのいくつかの例、最後にタスクの入力(図 1 中の $\langle c_i \rangle$) を与え、モデルがその続きを補完することでタスクに沿った出力を得る。

4 実験

実験では、SCoRE データセット [19] に含まれる文法項目のうち、現在完了形、関係詞節、仮定法の 3 種類を用いた。実験では、項目含有文生成における強化学習と few-shot プロンプトの有効性を確認する。また、報酬関数の組み合わせ方の優劣と、その優劣の文法項目との関係を検証する。

4.1 データセット

対話コーパスである Daily dialog¹⁾ [20] に収録されている対話から各対話の最初の発話を直前の発話として抽出し、それに対する応答の生成を行った。これは、応答生成に以前の文脈情報が関わることを避け、タスクを単純化するためである。

公開されている学習データを分割し、10,618 件の学習データ、500 件の開発データ、1,000 件のテストデータとして用いた。

4.2 ハイパーパラメータ

Self-critic 強化学習における生成モデルの初期値と、流暢性のための報酬関数の θ には DialoGPT [21] を使用し、サンプリングには top- k サンプリング [22] を用いた。early stopping は開発データによって 10 batch ごとに評価し、patience を 3 として学習を停止した。学習が進むと項目含有文の生成数は増加す

A and B are speaking. Create B's response using the present perfect.

===

A: Good morning, how are you doing today?

B: I have been feeling pretty good, Dr. Smith.

===

A: What's your plan for your future?

B: I'd like to work in a law firm to enrich my experience and put what I've learned into practice.

===

A: I'm going to Japan this year on vacation.

B: Have you ever been to America?

===

A: $\langle c_i \rangle$

B:

図 1 現在完了形用のプロンプトのテンプレート

るが、同じような文が多く生成され、多様性が失われる結果が得られる。そのため、項目含有率と多様性はトレードオフの関係にあることを根拠として、それらの積をとったものを early stopping の指標として採用した。

また、大規模汎用言語モデルとして GPT-3 davinci²⁾ [15] を用い、temperature=1 とした。現在完了形の生成に用いたプロンプトを図 1 に示す。その他の設定については付録 A に詳細を記載する。

4.3 文法項目の含有判定に用いた分類器

3.1.2 節の報酬関数 R_g が用いる分類器には、SCoRE データセットによって学習されたモデルを使用した。現在完了形、関係詞節、仮定法のそれぞれについて分類器を学習し、含有判定の閾値は 0.5 とした。SCoRE データセットおよび分類器の学習、性能に関しては付録 B に詳細を記載する。

4.4 比較手法

強化学習による生成とプロンプトによる生成を比較し、強化学習においては報酬関数に関して、以下の 5 つの組み合わせを比較した： R_g , $R_g + R_d$, $R_g \times R_d$, $R_g + R_f$, $R_g \times R_f$ 。

4.5 評価尺度

入力文に対して、強化学習による生成においては、beam 幅を 10 とした beam search により 10 通りの応答文を生成し、プロンプトによる生成においては、4.2 節の設定のもとで 10 回生成を行った。そのうち、分類器を用いて項目含有文を判定し、その中で最も尤度が高い文を出力文とする。そして、テストデータに対する出力文の尤度の平均値を評価

1) https://huggingface.co/datasets/daily_dialog

2) <https://beta.openai.com/docs/models/gpt-3>

表1 実験結果: 現在完了

	含有割合	Dist.-1	Dist.-2	GOAL
DialoGPT	0.145	0.223	0.490	0.0964
w/ R_g	0.789	0.111	0.279	0.0888
w/ $R_g + R_d$	0.781	0.052	0.127	0.1206
w/ $R_g \times R_d$	0.789	0.112	0.278	0.0931
w/ $R_g + R_f$	0.792	0.133	0.313	0.1078
w/ $R_g \times R_f$	0.603	0.136	0.280	0.1479
GPT-3 w/ few-shot	0.753	0.180	0.594	0.0332

値とする。これを GOAL³⁾と呼ぶ。尤度は DialoGPT により算出する。なお平均値の算出に際し、出力文に項目含有文が含まれない場合、その事例は除外する。また、含有割合は、項目含有文が出力文に含まれる割合である。

5 実験結果

表 1, 2, 3 に文法項目ごとの実験結果を示す。現在完了形、仮定法については $R_g \times R_f$ が最も高い GOAL 値を示し、関係詞節については $R_g + R_f$ が最も高い値を示した。 $R_g + R_f$ はいずれの実験設定においても高い GOAL を示した。生成例を付録 C に示す。強化学習の各設定を強化学習前の DialoGPT と比較すると、いずれも含有割合は向上しており、強化学習を用いて項目含有文を生成できることが確認された。また、いずれの場合においても、学習前と比べて Distinct-1, 2 の値は低下した。

現在完了と関係詞節では、多様性のための報酬関数 R_d を用いず、流暢性のための報酬関数 R_f を用いるほうが高い Distinct 値が得られており、 R_d の効果は限定的だと考えられる。理由として以下のことが考えられる。まず、多様性スコアが高いと判断された文が生成されやすくなったとしても、batch 単位という狭い範囲における多様性を計算しているため、モデルの多様性を正しく反映するとは限らない。また、batch に含まれる入力文が似ている場合、出力における多様性スコアは低下するのが自然だが、現状の報酬関数ではそれを十分考慮していない。また、流暢性を考慮することで、固定的なパターンの乱用（固定的なパターンにより R_g のスコアは上がる）が抑制される。これらのことから、 R_d の値が高い文を生成するように学習させることは必ずしも多様性を改善するとは限らず、多様性のための報酬関数にはさらなる改善が必要と考えられる。すべての文法項目において、流暢性のための報酬関数 R_f を適用した場合に GOAL 値の改善が見られた。従って、

3) GOAL: Grammar Oriented Average Likelihood

表2 実験結果: 関係詞節

	含有割合	Dist.-1	Dist.-2	GOAL
DialoGPT	0.822	0.186	0.431	0.1038
w/ R_g	0.911	0.173	0.393	0.1247
w/ $R_g + R_d$	0.888	0.156	0.355	0.1192
w/ $R_g \times R_d$	0.854	0.175	0.415	0.0960
w/ $R_g + R_f$	0.896	0.177	0.394	0.1398
w/ $R_g \times R_f$	0.833	0.180	0.423	0.1106
GPT-3 w/ few-shot	0.997	0.162	0.571	0.0363

表3 実験結果: 仮定法

	含有割合	Dist.-1	Dist.-2	GOAL
DialoGPT	0.037	0.228	0.490	0.0840
w/ R_g	0.860	0.091	0.222	0.1143
w/ $R_g + R_d$	0.566	0.059	0.172	0.1012
w/ $R_g \times R_d$	0.794	0.085	0.221	0.0914
w/ $R_g + R_f$	0.941	0.064	0.119	0.2145
w/ $R_g \times R_f$	0.949	0.030	0.049	0.2414
GPT-3 /w few-shot	0.535	0.174	0.579	0.0317

R_f は文法項目含有の制御を果たした上で、生成文の質の低下を抑止する効果があったと考えられる。

強化学習とプロンプトを比較すると、プロンプトによる生成では高い Distinct 値が得られていた。DialoGPT が基準となる GOAL 値では強化学習の方が高いものの、目視で確認した範囲では GPT-3 の出力も遜色ない。仮定法以外では、含有割合も同程度得られている。本稿が想定する利用形態では、複数候補を生成して適当な項目含有文を選択すればよいので、0.53 程度の含有割合でも問題ない。したがって、本稿の単純な設定下では、学習も不要なプロンプトによる生成手法に分があるといえる。しかし、より複雑な生成の制御を行いたい場合には、強化学習に優位性が生まれる可能性もある。

6 おわりに

強化学習とプロンプトによる応答生成手法を比較し、どちらの手法でも期待した生成結果を得られることを確認した。多様性の高さと学習が不要であることを考慮すると、本稿の設定下ではプロンプトによる手法に優位性が認められる。強化学習で用いる報酬関数については、多様性のための報酬関数の効果が限定的である一方で、流暢性のための報酬関数の有効性が示唆された。今後の方向性として、対象とする文法項目の拡充や、より複雑な制御が可能な生成手法の検討が挙げられる。本稿では学習者が特定文法項目を含む発話に触れる機会の増加を目指したが、学習者が次発話で文法項目を使用することを促す先行発話の生成も検討したい。

参考文献

- [1] Na-Young Kim. Effects of voice chat on efl learners' speaking ability according to proficiency levels. **Multimedia-Assisted Language Learning**, Vol. 19, No. 4, pp. 63–88, 2016.
- [2] Stergios Tegos, Stavros Demetriadis, and Thrasylvoulos Tsiatsos. A configurable conversational agent to trigger students' productive dialogue: a pilot study in the call domain. **International Journal of Artificial Intelligence in Education**, Vol. 24, No. 1, pp. 62–91, 2014.
- [3] Sherry Ruan, Angelica Willis, Qianyao Xu, Glenn M Davis, Liwei Jiang, Emma Brunskill, and James A Landay. Bookbuddy: Turning digital materials into interactive foreign language lessons through a voice chatbot. In **Proceedings of the Sixth (2019) ACM Conference on Learning@ Scale**, pp. 1–4, 2019.
- [4] Ryo Nagata, Tomoya Hashiguchi, and Driss Sadoun. Is the simplest chatbot effective in english writing learning assistance? In Le-Minh Nguyen, Xuan-Hieu Phan, Kôiti Hasida, and Satoshi Tojo, editors, **Computational Linguistics**, pp. 245–256, Singapore, 2020. Springer Singapore.
- [5] Joseph Weizenbaum. Eliza—a computer program for the study of natural language communication between man and machine. **Communications of the ACM**, Vol. 9, No. 1, pp. 36–45, 1966.
- [6] Zhaoliang Lin, Andrea Madotto, Yejin Bang, and Pascale Fung. The adapter-bot: All-in-one controllable conversational model. In **Proceedings of the AAAI Conference on Artificial Intelligence**, Vol. 35, pp. 16081–16083, 2021.
- [7] Nitish Shirish Keskar, Bryan McCann, Lav R Varshney, Caiming Xiong, and Richard Socher. Ctrl: A conditional transformer language model for controllable generation. **arXiv preprint arXiv:1909.05858**, 2019.
- [8] Ronald J Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. **Machine learning**, Vol. 8, No. 3, pp. 229–256, 1992.
- [9] Romain Paulus, Caiming Xiong, and Richard Socher. A deep reinforced model for abstractive summarization. In **International Conference on Learning Representations**, 2018.
- [10] Lijun Wu, Fei Tian, Tao Qin, Jianhuang Lai, and Tie-Yan Liu. A study of reinforcement learning for neural machine translation. In **Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing**, pp. 3612–3621, Brussels, Belgium, October-November 2018. Association for Computational Linguistics.
- [11] Karthik Narasimhan, Adam Yala, and Regina Barzilay. Improving information extraction by acquiring external evidence with reinforcement learning. In **Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing**, pp. 2355–2365, Austin, Texas, November 2016. Association for Computational Linguistics.
- [12] Jiaoda Li, Ryan Cotterell, and Mrinmaya Sachan. Probing via prompting. **arXiv preprint arXiv:2207.01736**, 2022.
- [13] Laria Reynolds and Kyle McDonell. Prompt programming for large language models: Beyond the few-shot paradigm. In **Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems**, pp. 1–7, 2021.
- [14] Yao Dou, Maxwell Forbes, Rik Koncel-Kedziorski, Noah A Smith, and Yejin Choi. Is gpt-3 text indistinguishable from human text? scarecrow: A framework for scrutinizing machine text. In **Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 7250–7274, 2022.
- [15] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners. **CoRR**, Vol. abs/2005.14165, 2020.
- [16] Steven J Rennie, Etienne Marcheret, Youssef Mroueh, Jerret Ross, and Vaibhava Goel. Self-critical sequence training for image captioning. In **Proceedings of the IEEE conference on computer vision and pattern recognition**, pp. 7008–7024, 2017.
- [17] 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, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.
- [18] Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao, and Bill Dolan. A diversity-promoting objective function for neural conversation models. In **Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies**, pp. 110–119, San Diego, California, June 2016. Association for Computational Linguistics.
- [19] Kiyomi Chujo, Kathryn Oghigian, and Shiro Akasegawa. A corpus and grammatical browsing system for remedial efl learners. **Multiple affordances of language corpora for data-driven learning**, pp. 109–130, 2015.
- [20] Yanran Li, Hui Su, Xiaoyu Shen, Wenjie Li, Ziqiang Cao, and Shuzi Niu. Dailydialog: A manually labelled multi-turn dialogue dataset. **arXiv preprint arXiv:1710.03957**, 2017.
- [21] Yizhe Zhang, Siqi Sun, Michel Galley, Yen-Chun Chen, Chris Brockett, Xiang Gao, Jianfeng Gao, Jingjing Liu, and Bill Dolan. DIALOGPT : Large-scale generative pre-training for conversational response generation. In **Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations**, pp. 270–278, Online, July 2020. Association for Computational Linguistics.
- [22] Angela Fan, Mike Lewis, and Yann Dauphin. Hierarchical neural story generation. In **Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 889–898, Melbourne, Australia, July 2018. Association for Computational Linguistics.
- [23] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In **International Conference on Learning Representations**, 2019.
- [24] 投野由紀夫 (編) . 学習者コーパスによる英語 C E F R レベル基準特性の特定と活用に関する総合的研究, 2016. <https://kaken.nii.ac.jp/ja/grant/KAKENHI-PROJECT-24242017>.

入力文	Look at the show on TV. I am watching a food show at a very famous seafood restaurant. I really want to eat at that restaurant. I am a seafood lover.
Daily dialog	Speaking of seafood, my mouth is watering. Let's go to the seafood restaurant in our neighborhood.
DialoGPT	I love seafood!
w/ $R_g \times R_f$	I've never been to a seafood restaurant.
GPT-3 w/ few-shot	I've never been to that restaurant for a lunch buffet.
入力文	Do you have a package tour to Hollywood?
Daily dialog	Yes, here is the table, which plans do you like best?
DialoGPT	I don't have a package tour.
w/ $R_g \times R_f$	I don't, but I've heard of it.
GPT-3 w/ few-shot	Definitely. In fact, I have been living in Hollywood for around twenty days.

図2 比較手法および DialoGPT の出力の比較

A 実験の詳細設定

Self-critic 強化学習における top- k サンプリングでは, k を 50 とした. R_d における Distinct-N において, $N=2$ とした. 学習時にパラメータを AdamW [23] によって最適化し, 学習率は $2e^{-5}$, L2 正則化の係数を $1e^{-2}$ とした. Distinct-N を適切に計算するために, 最小出力長を 10 とした. batch size は 10, 最大 1100 イテレーションとした.

B 文法項目の判定に用いた分類器

強化学習の報酬関数として文法を含むかどうかの判定を行う分類器を用いた.

分類器の構造は, 本文の 3.1.2 節で述べたように, 判定を行う入力文に対して, BERT の [CLS] トークンの埋め込みから線形層とシグモイド関数により文法項目を含むかどうかの推定を行う.

今回作成した分類器には, 学習のためのデータセットが必要である. ただし, 必要なデータは対話形式である必要はなく, 言語モデルを教師学習する場合と比べ小規模でよい. また, データが得られない場合には正規表現による分類で代替可能である.

ここでは, 分類器の学習に用いたデータセットと学習時の設定について述べ, 分類器の性能を正規表現を用いたルールベースによる分類と比較する. 正規表現は CEFR-J の正規表現リスト [24] に基づいて作成し, 現在完了形において分類器と性能を比較した.

B.1 SCoRE データセット

文に対して文法項目が人手により付与されている SCoRE データセット⁴⁾ [19] を分類器の学習に用いた. したがって, 文法項目は SCoRE データセットに含まれる用例について扱うこととする. SCoRE データセットは大分類としておよそ 20 種類の文法項目が付与されており, 表 4 に今回対象とした文法項目に該当するデータの件数を示す. また, 項目内においても複数の用例が収録されており, 例として仮定法では, I wish, if I were, if 過去形, if had 過去分詞などが収録されている.

分類器の学習には, 対象の文法項目が付与された正例の他に, 対象の文法項目が付与されていない負例が必要となる. そこで, 負例には SCoRE データセット内から対象の文法項目ではない文法項目が付与されている文を用いる. ただし, 対象の文法項目が付与されていない全て

表4 SCoRE データセットの統計量

文法項目	例文数
現在完了	547
関係詞節	1,142
仮定法	783

表5 実験結果: 分類器の精度

文法項目	BERT	正規表現
現在完了	0.9902	0.9641
関係詞節	0.9879	-
仮定法	0.9919	-

の文を負例とすると, 適していないデータが含まれる可能性があり, また, データ中の割合が大きく偏るため, 人手によって負例に適しているかを確認して分類の学習に用いるデータセットを構築した. 得られた分類用のデータセットから 80% を学習データ, 20% をテストデータに分割した. 最終的に, 現在完了には, 学習データとテストデータをそれぞれ 1,222, および 306 件とした. 関係詞節と仮定法には, 学習データを 1,977 件, テストデータ 495 件とした.

B.2 ハイパーパラメータ

分類モデルには BERT (bert-large-uncased) を初期値として使用した. 学習時にパラメータを AdamW によって最適化し, 学習率は $2e^{-5}$, L2 正則化の係数を $1e^{-2}$ とした. Batch size は 10, エポック数は 10 とした. 本実験においては, 分類器はテストデータにおいて最も高い性能を示したモデルを用いた.

B.3 分類性能

表 5 にそれぞれの文法項目に対する分類器の分類性能を示す. 評価尺度に正解ラベルと予測ラベルの間の正解率を用いて評価した. 実験では, 現在完了において正規表現よりも BERT の方が分類性能が高かったため, 他の項目についても BERT を用いた分類器を報酬関数として採用した.

C 出力例

現在完了形の出力例を図 2 に示す. Daily dialog や DialoGPT と比較して, 学習後は現在完了を用いた応答文となっていることが確認できる. その上, 入力に対する応答として破綻のない応答となっている.

4) <https://www.score-corpus.org/>