

文書分類のための要約に基づくデータ拡張

王 悦綸

東京大学大学院 情報理工学系研究科
etsurin@tkl.iis.u-tokyo.ac.jp

吉永 直樹

東京大学 生産技術研究所
ynaga@iis.u-tokyo.ac.jp

概要

文書などの長文を対象とした自然言語理解タスクは、事前学習済みモデルを用いたとしてもデータの過疎性の問題から解くことが依然難しい。そこで、本稿では人間が文書読解能力を身につける過程に着想を得て、文書分類のための要約に基づくデータ拡張手法 SUMMAug¹⁾を提案する。提案手法では、元の学習データの入力を要約し、必要に応じ要約された入力に合わせるように元のラベルを縮退することで、学習容易な事例を生成し、中間学習に用いる。2つの文書分類データセットでの実験結果により、提案手法が分類精度及び学習の安定性の観点で既存手法より優れていることを確認した。

1 はじめに

事前学習済みモデル [1, 2, 3] の導入により、さまざまな自然言語理解タスクの精度が改善したものの、文書のような長いテキストを入力したり [4] や粒度の細かい出力ラベル [5] を扱う複雑なタスクでは、その効果は限定的である。例えば、文書分類のようなタスクでは、多様な入力と出力の関係性を捉えることが必要であり、データの過疎性の問題が深刻となる。

データの過疎性の問題に対処するには、データ拡張が有効であり、文書分類タスクにも過去に適用されている。これらは基本的に、入力に摂動を加えることで、元の学習事例から疑似学習事例を生成するものである。具体的な摂動としては、逆翻訳 [6] に加えて、テキスト [7, 8] および埋め込み空間 [9, 10, 11] 内での編集などが試みられている。これらの摂動で生成された疑似学習事例は、元の学習事例と本質的に同じ意味的内容しか含んでおらず、モデルが長い入力テキストを理解する能力を学習するのに役立つとは言い難い。

1) 実装: <https://github.com/etsurin/summaug>

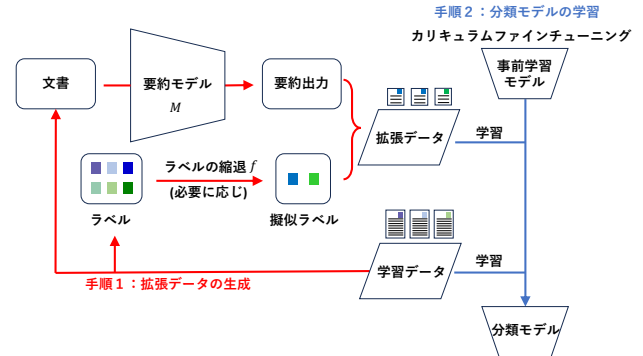


図1 提案手法 SUMMAug を用いた学習の流れ。

本研究では、文書分類タスクを対象として、長い入力テキストを理解する能力をモデルが獲得しやすくするため、単純で効果的な自動要約に基づく学習データ拡張手法 SUMMAug を提案する。提案手法では、まず、元の学習事例の入力を自動要約で要約する。このとき、必要に応じてラベルを縮退し、要約された入力に適合するようにする。このようにして得られた疑似訓練事例を用いて、人間が短いテキストから徐々に文書の読解能力を身につけることを考慮して、カリキュラム学習 [12] (微調整) を行う。

実験では、ラベル粒度の異なる2つの文書分類データセット [13, 14] において、ベースラインのデータ拡張手法 [8] と提案手法 SUMMAug を比較した。その結果、SUMMAug に基づくカリキュラム微調整により、精度と頑健性の両方でベースライン手法を上回るモデルが得られることを確認した。

2 関連研究

本節では、まず文書分類のための既存の深層学習モデルを説明し、次にテキスト分類のための既存のデータ拡張手法を紹介する。その後、テキスト分類で自動要約を活用した研究について触れる。

文書分類 文書分類において、複数の文から構成される長いテキストの内容を理解するために、より優れたモデル構造を探求する試みが広く行われて

表 1 IMDb データセットにおける元のテキストと生成された要約の例である。1 行目は元のテキストで、2 行目は生成された要約である。赤のテキストは元のテキストにおける要約の対応部分である。

I am Anthony Park, Glenn Park is my father. First off I want to say that the story behind this movie and the creation of the Amber Alert system is a good one. However **the movie itself was poorly made and the acting was terrible**. The major problem I had with the movie involved the second half with Nichole Timmons and father Glenn Park. **The events surrounding that part of the story were not entirely correct**. My father was suffering from psychological disorders at the time and picked up Nichole without any intent to harm her at all. He loved her like a daughter and was under the mindset that he was rescuing her from some sort of harm or neglect that he likely believed was coming from her mother who paid little attention to her over the 3 plus years that my father took care of her and summarily raised her so her mother could frolic about. The movie depicted my father in a manner that he was going to harm her in some way shape or form. The funny thing is that Nichole had spent many nights sometimes consecutively at my fathers place while Sharon would be working or doing whatever she was doing. The reason that my father was originally thought to be violent was because he had items that could be conceived to be weapons on his truck. My father was a landscaper. The items they deemed to be weapons were landscaping tools that he kept in his truck all the time for work. **My recommendation is take this movie with a grain of salt, it is a good story and based on true events** however the details of the movie (at least the Nichole Timmons - Glenn Park portion) are largely inaccurate and depict the failure of the director to discover the truth in telling the story. The funny thing is, that if the director would have interviewed any of Sharon's friends who knew the situation they would have stated exactly what I have posted here.

The movie itself was poorly made and the acting was terrible. The events surrounding that part of the story were not entirely correct. My recommendation is take this movie with a grain of salt, it is a good story and based on true events.

いる。具体的には、グラフニューラルネットワーク [15, 16] や畳み込みアテンションネットワーク [5] などがある。しかしながら、近年、Transformer [17] ベースのモデルの有効性が再検証され [18]、これらの文書分類に特化したモデル構造を上回る性能が報告されている。本研究で提案するデータ拡張は、特定のモデル構造に依存しないが、この現状を踏まえて、Transformer に基づく事前学習済みモデル RoBERTa [2] を使用する。

テキスト分類のためのデータ拡張 テキスト分類タスクにおいて、データの過疎性の問題に対処するため、元の学習事例から擬似学習事例を生成するデータ拡張が広く試みられている。Shleifer [6] は、機械翻訳を用いて学習事例を別の言語に翻訳し、その出力を逆翻訳して元の言語に戻すことで、入力と同じ意味で表現が異なる学習事例を生成している。Wei ら [7] は、単語を削除、挿入、または同義語で置き換えることによって入力に摂動を加えている。Karimi ら [8] は、ランダムに句読点を挿入するという、より単純ながら効果的な摂動を提案している。これら以外にも、学習事例のテキストに直接摂動を加える代わりに、その埋め込み空間にノイズを加えることで擬似学習事例を生成する研究 [9, 10, 11] もある。しかしながら、これらの手法で得られる擬似学習事例は元の学習事例の言い換えに過ぎず、長いテキストの理解に直接役立つとは言い難い。

テキスト分類における自動要約の活用 Li ら [19] と Hartl ら [20] は、フェイクニュース検出のため、要約を介して対象記事から事実を抽出している。これらの手法が推論時に要約を行うのに対し、提案手

法はモデルの学習のために要約を用いる。

3 提案手法

文書分類では、数十以上の文が含まれる長いテキストを理解する必要がある。人間は、成長するにつれ、長いテキストでも正しく理解できるようになるが、そのような読解能力をどのように身につけたのだろうか。学校での学習を振り返ると、短くて簡潔なテキストから読み始めて、徐々に長いテキストを読むことで読解能力を磨いたことに気づくだろう。

本研究では、この点に着想を得て、文書分類のための要約に基づくデータ拡張手法 SUMMAug を提案する。提案手法では、自動要約で抽象的な訓練例を生成し、拡張データ、元のデータの順でカリキュラム学習（微調整）を行う。

3.1 要約に基づくデータ拡張

SUMMAug ではまず、後述する要約モデル M を用いて、文書分類のための擬似学習事例を生成する。具体的に、各学習事例 $\{x, y\}$ の入力 x に要約モデル M を適用し、 x の要約 $\hat{x} = M(x)$ を得る。

ここで問題となるのは、生成された簡潔な要約 \hat{x} に対するラベルの与え方である。要約は本質的な情報しか保持しないため、ラベルが細かいと、元のラベル y は不適切となる可能性がある。そこで、ラベルを粗いカテゴリに縮退する関数 f を定義し、擬似学習事例 $\{\hat{x}, f(y)\}$ を得る (図 1)。

表 1 は IMDb データセットに含まれる事例と後述の要約モデルによる要約である。元の入力は中立で少しネガティブの傾向があるレビューであるが、要

約ではこの傾向が強まっている。この例からも、擬似学習事例のラベルをより粗い粒度に縮退する必要性が示唆される。

要約モデル M について 手法の適用範囲を広げるためには、要約モデル M は多様なドメインの文書を要約できることが望ましい。そこで本稿では、CNN-Dailymail データセット [21] で微調整した BART [22] を要約モデル M として用いた。CNN-Dailymail はニュース記事に関する要約データセットであり²⁾、広範なトピックを含むため幅広いドメインの文書に適応可能である。また、CNN-Dailymail に含まれる要約の多くは抽出型要約であり、XSum [23] など抽象型要約データセットから学習したモデルで深刻な幻覚 [24] が生じる可能性も低いと期待できる。

3.2 分類器のカリキュラム微調整

データ拡張に関する既存研究では、擬似学習データと生成元の学習データのラベル集合は同一であり、混合して区別せず学習に用いている。しかしながら、提案手法で生成する擬似学習データ、すなわち、拡張データでは、要約に適合するようラベルを縮退する場合がある。そこで、学習戦略として以下の2つを考える。

混合微調整 元の学習データと擬似学習データを混合し、事前学習済みモデルを微調整する。この戦略では、ラベルを潰さず、 $f(y) = y$ とする。

カリキュラム微調整 まず、擬似学習データで事前学習済みモデルを微調整し、得られたモデルを、元の学習データで微調整する。この提案手法はカリキュラム学習 [25] に着想を得たものである。この戦略では、必要に応じてラベルを縮退するが、その場合は、元の学習データで微調整する前に、出力層のパラメータを破棄する。

以下の実験では、ラベルの粒度が異なる文書分類データセットに対して提案手法を適用し、2つの戦略を比較する。

4 実験

提案手法をラベル粒度の異なる2つの文書分類データセット [13, 14] に適用し、i) 提案手法が、ベースライン手法と比べて分類精度と学習の安定性の観点で優れていること、また ii) ラベル縮退を伴うカ

表2 IMDB データセット。 C , L , L_M はそれぞれ、ラベル数、入力 の平均長、要約の平均長である。

| データセット | 学習・検証・評価データ | C | L | L_M |
|---------|--------------------|-----|-------|-------|
| IMDb-2 | 22500/2500/25000 | 2 | 279.5 | 51.3 |
| IMDb-10 | 108670/13432/13567 | 10 | 394.2 | 50.2 |

リキュラム微調整が提案手法の有効性を発揮する上で重要な役割を果たすことを確認する。

4.1 データセット

提案手法の評価には、大規模映画レビューデータセット IMDb-2 [13] と IMDb-10 を用いた。IMDb-2 には肯定・否定の2値ラベルが付与されており、IMDb-10 には1から10までの10種の極性ラベルが付与されている。IMDb-2 データセットには開発データが含まれないため、学習データをランダムに9:1に分割し、学習データと検証データとした。IMDb-10 データセットについては、Adhikari ら [14] の実験と同じ分割を用いた。2つのデータセットの詳細を、表2に示す。

4.2 実験設定

評価には以下の3つのモデルを用いた。全てのモデルは、ベースラインの RoBERTa [2] を微調整したものである。

RoBERTa 事前学習済み RoBERTa³⁾ を元の学習データで微調整したベースライン。

RoBERTa + AEDA 元の学習事例の入力に既存のデータ拡張手法 AEDA⁴⁾ を適用し、得られた擬似学習データと元の学習データを混合して前述の RoBERTa を微調整したベースライン。

RoBERTa + SUMMaug CNN-Dailymail で微調整した BART を要約モデルに用いた提案手法。得られた擬似学習データと元の学習データを用いて前述の RoBERTa を微調整する。

学習データが少ない設定下での有効性を評価するために、学習データからそれぞれ200および1500事例をランダムに選択し、これらの部分データセットを用いた学習も行った。ただし、IMDb-10 データセットでは、学習データが200事例の場合は、全てのモデルの学習が発散したため、評価しなかった。

カリキュラム微調整の効果を明らかにするために、SUMMaug の擬似学習データだけでなく AEDA の擬似学習データを用いた学習でもカリキュラム微

2) <https://huggingface.co/facebook/bart-large-cnn>

3) <https://huggingface.co/roberta-large>

4) https://github.com/akkarimi/aeda_nlp

表3 IMDb-2での平均精度と標準偏差. 混合とカリキュラムはそれぞれ混合微調整とカリキュラム微調整を指す.

| モデル | 学習データの件数 | | |
|--------------------|------------------------------|------------------------------|------------------------------|
| | 200 | 1500 | all |
| RoBERTa | 92.19 _{1.21} | 94.21 _{0.62} | 94.63 _{0.56} |
| + AEDA (混合) | 90.91 _{1.44} | 94.43 _{0.49} | 94.75 _{0.66} |
| + AEDA (カリキュラム) | 93.59 _{1.16} | 94.26 _{0.74} | 95.56 _{0.12} |
| + SUMMaug (混合) | 92.94 _{0.99} | 94.61 _{0.64} | 94.85 _{0.62} |
| + SUMMaug (カリキュラム) | 93.36 _{0.97} | 94.77 _{0.28} | 95.45 _{0.17} |

表4 IMDb-10での平均精度と標準偏差. 表記は表3に従う.

| モデル | 学習データの件数 | |
|--------------------|------------------------------|------------------------------|
| | 1500 | all |
| RoBERTa | 39.99 _{8.46} | 56.58 _{0.34} |
| + AEDA (混合) | 36.58 _{10.64} | 51.23 _{14.39} |
| + AEDA (カリキュラム) | 41.77 _{3.01} | 56.63 _{1.65} |
| + SUMMaug (混合) | 40.65 _{2.71} | 55.81 _{2.00} |
| + SUMMaug (カリキュラム) | 42.14 _{1.48} | 57.55 _{0.29} |

調整を試した. なお, IMDb-10 データセットでは, §3.1 で述べたように, 拡張データのラベルを縮退した. 具体的には, 0-4 のラベルを 0 (否定的) に, 5-9 のラベルを 1 (肯定的) に写像し, 2 値分類問題とした. 異なるラベル縮退を行なった場合の結果についても, 本実験の結果の後で報告する.

4.3 学習の詳細

学習のハイパーパラメータは以下のように設定した. IMDb-2 での実験では, バッチサイズを 64, 学習率を $1e-5$ とした. IMDb-10 データセットでの実験では, Adhikari ら [14] に従い, バッチサイズを 16, 学習率を $2e-5$ にした. 学習エポック数は付録を参照されたい. 学習は NVIDIA Quadro P6000 GPU 4 枚を用いて行った.

評価用のモデルは, 開発データでの分類精度に基づき選択した. 5 回の試行での平均値を報告する.

5 実験結果

表3と表4に, ベースラインと我々の提案手法の結果を示す. 提案手法は, 全ての実験設定においてベースラインを上回った. さらに, 低資源設定でも, 提案手法が有効であることを確認した.

SUMMaug に基づくデータ拡張による学習安定性 SUMMaug は, データセット全体で高い精度を達成するだけでなく, 結果の分散の小さい安定した学習に寄与している. 一方で, 既存手法である AEDA (混合) では, IMDb-2 で学習データが 200 件の場合

表5 SUMMaug (カリキュラム) 手法で異なるラベル縮退関数を用いた場合の平均精度と標準偏差の変化. N は縮退後のラベル数, $f(x)$ は 0-9 のラベルに対する具体的な縮退関数である.

| N | $f(x)$ | 精度 |
|-----|-----------------------|-----------------------|
| 2 | [0,0,0,0,0,1,1,1,1,1] | 57.55 _{0.29} |
| 3 | [0,0,0,1,1,1,1,2,2,2] | 57.47 _{0.20} |
| 4 | [0,0,0,1,1,2,2,3,3,3] | 57.66 _{0.31} |
| 5 | [0,0,1,1,2,2,3,3,4,4] | 57.32 _{0.60} |
| 10 | [0,1,2,3,4,5,6,7,8,9] | 57.20 _{0.56} |

精度が下がり, IMDb-10 では不安定な結果をもたらしてしまう.

カリキュラム微調整の効果 SUMMaug で得られた擬似学習データを用いてカリキュラム微調整の代わりに混合微調整を行った場合, 提案手法による改善幅は小さくなり, IMDb-10 ではベースラインも下回る結果となった. このことから, SUMMaug で得られた擬似学習データを用いる場合はカリキュラム微調整を行うことが必須であることが分かる.

ラベルの縮退関数の影響 表5は, 異なるラベル縮退関数 f の下での SUMMaug (カリキュラム) の結果を示している. $N \leq 4$ の場合, 精度は同等であるが, $N = 5, 10$ では恐らくではラベルの縮退が不十分なため, 精度と学習の安定性に低下が見られる. しかしながら, ラベルの縮退を行わない場合でも, 元の学習データでの微調整に伴い拡張データに含まれるノイズの影響が弱まり, 表4で報告した混合微調整の結果よりは精度・学習安定性ともに良い結果となっている.

6 おわりに

本研究では, 文書分類のための単純で効果的な要約に基づくデータ拡張手法 SUMMaug を提案した. 提案手法では, 人間が文書読解能力を習得する過程を踏まえ, 短いテキストの分類問題から長いテキストの分類問題の順で学習する. 2つの文書分類データセットでの実験結果により, SUMMaug がベースラインのデータ拡張手法と比較して, 精度と学習の安定性の観点で性能を向上させることを確認した. また, 低資源設定でも提案手法が効果的であることも確認した.

今後の課題としては, ラベル縮退関数 f を自動で最適化する手法の検討や, 異なる要約モデルを用いた場合の効果を検証することが挙げられる.

謝辞

本研究は東京大学生産技術研究所特別研究経費および JSPS 科研費 JP21H03494 の助成を受けています。

参考文献

- [1] 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.
- [2] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. RoBERTa: A robustly optimized bert pretraining approach. **arXiv preprint arXiv:1907.11692**, 2019.
- [3] Pengcheng He, Xiaodong Liu, Jianfeng Gao, and Weizhu Chen. DeBERTa: Decoding-enhanced BERT with disentangled attention. 2020.
- [4] Yu-Chen Lin, Si-An Chen, Jie-Jyun Liu, and Chih-Jen Lin. Linear classifier: An often-forgotten baseline for text classification. In **Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)**, pp. 1876–1888, Toronto, Canada, July 2023. Association for Computational Linguistics.
- [5] Yang Liu, Hua Cheng, Russell Klopfer, Matthew R. Gormley, and Thomas Schaaf. Effective convolutional attention network for multi-label clinical document classification. In **Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing**, pp. 5941–5953, Online and Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics.
- [6] Sam Shleifer. Low resource text classification with ULMFit and backtranslation. **arXiv preprint arXiv:1903.09244**, 2019.
- [7] Jason Wei and Kai Zou. EDA: Easy data augmentation techniques for boosting performance on text classification tasks. In **Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)**, pp. 6382–6388, Hong Kong, China, November 2019. Association for Computational Linguistics.
- [8] Akbar Karimi, Leonardo Rossi, and Andrea Prati. AEDA: An easier data augmentation technique for text classification. In **Findings of the Association for Computational Linguistics: EMNLP 2021**, pp. 2748–2754, Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics.
- [9] Jiaao Chen, Zichao Yang, and Diyi Yang. MixText: Linguistically-informed interpolation of hidden space for semi-supervised text classification. In **Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics**, pp. 2147–2157, Online, July 2020. Association for Computational Linguistics.
- [10] Hui Chen, Wei Han, Diyi Yang, and Soujanya Poria. DoubleMix: Simple interpolation-based data augmentation for text classification. In **Proceedings of the 29th International Conference on Computational Linguistics**, pp. 4622–4632, Gyeongju, Republic of Korea, October 2022. International Committee on Computational Linguistics.
- [11] Xing Wu, Chaochen Gao, Meng Lin, Liangjun Zang, and Songlin Hu. Text smoothing: Enhance various data augmentation methods on text classification tasks. In **Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)**, pp. 871–875, Dublin, Ireland, May 2022. Association for Computational Linguistics.
- [12] Petru Soviany, Radu Tudor Ionescu, Paolo Rota, and Nicu Sebe. Curriculum learning: A survey. **International Journal of Computer Vision**, Vol. 130, No. 6, pp. 1526–1565, 2022.
- [13] Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, and Christopher Potts. Learning word vectors for sentiment analysis. In **Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies**, pp. 142–150, Portland, Oregon, USA, June 2011. Association for Computational Linguistics.
- [14] Ashutosh Adhikari, Achyudh Ram, Raphael Tang, and Jimmy Lin. DocBERT: BERT for document classification. **arXiv preprint arXiv:1904.08398**, 2019.
- [15] Haopeng Zhang and Jiawei Zhang. Text graph transformer for document classification. In **Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)**, pp. 8322–8327, Online, November 2020. Association for Computational Linguistics.
- [16] Chong Zhang, He Zhu, Xingyu Peng, Junran Wu, and Ke Xu. Hierarchical information matters: Text classification via tree based graph neural network. In **Proceedings of the 29th International Conference on Computational Linguistics**, pp. 950–959, Gyeongju, Republic of Korea, October 2022. International Committee on Computational Linguistics.
- [17] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, **Advances in Neural Information Processing Systems**, Vol. 30. Curran Associates, Inc., 2017.
- [18] Xiang Dai, Ilias Chalkidis, Sune Darkner, and Desmond Elliott. Revisiting transformer-based models for long document classification. In **Findings of the Association for Computational Linguistics: EMNLP 2022**, pp. 7212–7230, Abu Dhabi, United Arab Emirates, December 2022. Association for Computational Linguistics.
- [19] Qifei Li and Wangchunshu Zhou. Connecting the dots between fact verification and fake news detection. In **Proceedings of the 28th International Conference on Computational Linguistics**, pp. 1820–1825, Barcelona, Spain (Online), December 2020. International Committee on Computational Linguistics.
- [20] Philipp Hartl and Udo Kruschwitz. Applying automatic text summarization for fake news detection. In **Proceedings of the Thirteenth Language Resources and Evaluation Conference**, pp. 2702–2713, Marseille, France, June 2022. European Language Resources Association.
- [21] Karl Moritz Hermann, Tomas Kocisky, Edward Grefenstette, Lasse Espeholt, Will Kay, Mustafa Suleyman, and Phil Blunsom. Teaching machines to read and comprehend. **Advances in neural information processing systems**, Vol. 28, , 2015.
- [22] 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, Online, July 2020. Association for Computational Linguistics.
- [23] Shashi Narayan, Shay B. Cohen, and Mirella Lapata. Don’t give me the details, just the summary! topic-aware convolutional neural networks for extreme summarization. In **Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing**, pp. 1797–1807, Brussels, Belgium, October–November 2018. Association for Computational Linguistics.
- [24] Joshua Maynez, Shashi Narayan, Bernd Bohnet, and Ryan McDonald. On faithfulness and factuality in abstractive summarization. In **Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics**, pp. 1906–1919, Online, July 2020. Association for Computational Linguistics.
- [25] Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. Curriculum learning. In **Proceedings of the 26th Annual International Conference on Machine Learning, ICML ’09**, p. 41–48, New York, NY, USA, 2009. Association for Computing Machinery.

A 付録

A.1 学習エポック数

| データセット (学習手法) | 学習データの件数 | | |
|-------------------|----------|-------|-----|
| | 200 | 1500 | all |
| IMDb-2 (データ拡張なし) | 70 | 18 | 2 |
| IMDb-2 (ミックス) | 70 | 18 | 2 |
| IMDb-2 (カリキュラム) | 70/70 | 18/18 | 2/2 |
| IMDb-10 (データ拡張なし) | - | 20 | 4 |
| IMDb-10 (ミックス) | - | 20 | 4 |
| IMDb-10 (カリキュラム) | - | 5/20 | 2/6 |

表 6 実験で学習エポック数の詳細. カリキュラムファインチューニングの手法では、 x/y は「拡張データで x エポック学習し、元の学習データで y エポック学習する」と意味する.

表 6 では、実験での学習エポック数をまとめた. エポック数は検証データでの精度により設定したものである.