Vol.2022-NL-253 No.13
2022/9/30

T5 による特定キャラクター風発話への変換とその言語モデルの構築

岸野　望叶[1,a)]　古宮 嘉那子[2,b)]　新納　浩幸[3,c)]

概要：現在、Siri などの対話エージェントが盛んに利用されていたり、RPG などのゲームで大量のセリフが必要になったりする。それらの発話はキャラクターらしさを含んでいることが求められる。しかし、特定のキャラクターに特化した言語モデルの構築を行うには学習データが限られており精度の向上は困難である。そのため本論文では対象の発話者と同作品に出てくる別人物の発話を T5 を用いて、対象発話者の発話風に変換し、学習データを増補する。その学習データを「ドメイン」の学習データ、対象の発話者の発話を「タスク」の学習データとし、TAPT-DAPT の手法でベースの言語モデルとなる GPT2 に Fine-tuning を行った。その結果、GPT2 に対象の発話者の発話のみで学習を行った場合のパープレキシティが 46.23 であったのに対し、この手法で行った場合のパープレキシティは 43.93 となり、精度を向上させることができた。

Conversion to Specific Character-like Speech Using T5 and Construction of Language Model of Utterances of Fictional Characters

1. はじめに

本論文では特定の発話者に対する言語モデルの構築手法を提案する。

小説、アニメあるいはゲームなどの登場人物は、仮想的にではあるが、ある特徴を有しており、その人物による自然な発話にはその特徴を有したものにする必要がある。高度化された対話システムにおいても、ディスプレイに表示される仮想的な発話者あるいは実際のロボットの発話では、その見た目から想起される発話者らしい発話が自然である。このような背景から発話者の特徴を反映した発話文を生成する研究がいくつか行われているが、それらはみな規則ベースの手法である（[1][2] など）。規則ベースの手法は規則の構築が手作業となるためその構築コストが高く、また対象を別の発話者に変更することが困難という問題もある。

本論文では発話者の特徴を有した発話文を生成するために、その発話者の言語モデルを構築することを試みる。発話者の言語モデルはその発話者の発話文を大量に収集することで自動的に構築できる。ただしその発話者は、通常、仮想的であるため収集できる発話文の量は限られる。そのため本論文では GPT-2 [3] をベースの言語モデルに設定し、対象発話者の少量の発話文によって、そのベースの言語モデルを Fine-tuning することで目的の言語モデルを構築する。更に Fine-tuning により構築される言語モデルの性能を上げる（パープレキシティを下げる）ために T5 [4] を利用して別人物の発話文を、対象発話者の発話文に変換するモデルを学習し、対象発話者の発話文を増補する。このとき増補された発話文を対象発話者の発話文を含む「ドメイン」の発話文、対象発話者の発話文を「タスク」の発話文と捉え、TAPT-DAPT [5] の手法を利用することで効果的な Fine-tuning を目指す。

実験では株式会社スクウェア・エニックスのゲームであるドラゴンクエスト Ⅳ 内のセリフをコーパスとし、コー

[1] 茨城大学大学院理工学研究科情報工学専攻
Major in Computer and Information Sciences, Graduate School of Science and Engineering, Ibaraki University
[2] 東京農工大学大学院工学研究院先端情報科学部門
Institute of Engineering, Tokyo University of Agriculture and Technology
[3] 茨城大学大学院理工学研究科情報科学領域
Graduate School of Science and Engineering, Department of Computer and Information Sciences, Ibaraki University
[a)] 21nm722y@vc.ibaraki.ac.jp
[b)] kkomiya@go.tuat.ac.jp
[c)] hiroyuki.shinnou.0828@vc.ibaraki.ac.jp

パス内で現れるマリベルを対象発話者に設定した。このコーパスから取り出した対象発話者の発話文を訓練データとテストデータに分け、テストデータに対する言語モデルのパープレキシティから構築した言語モデルの評価を行った。訓練データを用いて GPT-2 を Fine-tuning した言語モデルのパープレキシティは 46.23 であったが、本論文の提案手法を用いることで 43.93 まで改善できた。

2. 関連研究

Suchin Gururangan ら [5] は事前学習されたモデルに対して対象のドメインやタスクのデータセットで追加の事前学習を行った。その結果、適切なドメインやタスクで事前学習を行うことでモデルの精度は向上、特に適切なタスクでの学習においては少量のデータセットで精度が向上することを示した。また勝又ら [6] は話し言葉 BERT の作成に TAPT-DAPT の手法を利用し、TAPT-DAPT の 2 種類の事前学習を組み合わせることで様々なタスクに有効なモデルを作成した。また同時に話し言葉においてもこの手法は有効であることを示した。

また、キャラクター性についての研究も盛んにおこなわれている。Mairesse と Walker[7] によって開発された PERSONAGE (personality generator) は最初の高度なパラメータ化された会話生成器である。彼らは認識できる言葉のバリエーションや個性を作成した。Walker ら [8] は映画の対話コーパスを収集し言語構造やキャラクターの原型のアノテーションを行った。さらに彼らはジャンル、性別、ディレクター、映画の年代のようなグループによって言語のスタイルを分類するためにキャラクターの言語モデルの実験を行った。宮崎ら [9] は日本語の会話を特徴付ける言語表現を基礎的に分析し、部分的に言い換えることで会話を特徴付ける技術を開発した。また、宮崎ら [10] は、日本のフィクションの登場人物の言語的特徴のカテゴリーを報告した。さらに会話の特徴づけやバリエーションを豊かにする方法として、各文の機能部分を対象人物に適した確率で言い換えて、読者が対話者の特徴づけを理解できるかどうかを確認する実験を行った。[11] また、日本の架空の人物の発話を特徴付けるために、日本語の音変化表現に着目し、これらの表現を収集し、分類した研究もある [1]。さらに、奥井・中辻 [12] は、日本語の対話システムにおいて、ポインタ生成機構を用いて、複数の異なるキャラクターの応答を参照し、様々な応答を生成した。彼らは、少ないデータ量で応答の特徴付けを学習した。また岸野ら [2] の研究では、SentencePiece を用いて形態素解析を行い、TFIDF を用いてキャラクターの特徴と考えられる単語を抽出した。

3. 提案手法

3.1 TAPT+DAPT

本研究では GPT2 の small サイズのモデル[*1]を使用した。このモデルはウェブサイトから集められた 800 万文書のデータセットを利用し学習されたものである。

DAPT とは Domain-Adaptive Pretraining の略であり、ターゲットが属するドメインのデータで学習を行うものである。TAPT とは Task-Adaptive Pretraining の略であり、対象のタスクのデータで学習を行うものである。TAPT+DAPT の学習方法とは、既に一般的なコーパスで事前学習された言語モデルに対し、対象のドメインのデータで追加の事前学習を行い、さらに対象のタスクのデータで追加の事前学習を行う方法である。Suchin Gururangan ら [5] は、DAPT のみ、TAPT のみ、DAPT+TAPT の場合をそれぞれ比較し、DAPT+TAPT が最もよい精度であることを示している。また、対象のタスクのドメイン以外のデータを使って DAPT を行った場合、精度が悪くなることも示した。

本研究の提案手法を図 1 に示す。本研究での対象のタスクのデータは対象の発話者の発話であり、その対象が属するドメインのデータは同一作品の別の登場人物の発話と考えた。また同一作品の別の登場人物の発話を対象の発話者風の発話に変換し、それをドメインの学習データとすることで精度が上がると考えた。ここで対象の発話者風の発話への変換には T5 モデル[*2]を利用する。本研究では、以下の手順を用いて対象発話者の発話を増補する（図 1）。

(1) T5 を用いて同一作品の別の登場人物の発話を一般的な発話に変換する
(2) (1) で作成した一般的な発話を T5 を用いて対象者風の発話に変換する
(3) (2) で作成した対象者風の発話を DAPT の学習データとして扱い、GPT2 の Fine-tuning を行う
(4) （変換で得たのではない）もともとの対象の発話を TAPT の学習データとして扱い、さらに GPT2 の Fine-tuning を行う

3.2 T5 による発話文変換モデルの構築

学習データを増幅するため、T5 で発話文の変換を行った。

まず対象の発話者の発話から TFIDF を用いてより対象キャラクターの特徴を捉えていると考えられる発話を 100 文抽出した。詳しい抽出方法を 3.3 節で述べる。抽出した 100 文を、そのキャラクターの特徴をなくした一般的な発話に手作業で修正した。ここでの一般的な発話とは、性別、年代、人物像を想像させないような発話のことである。例

[*1] https://github.com/tanreinama/gpt2-japanese
[*2] https://huggingface.co/sonoisa/t5-base-japanese/discussions

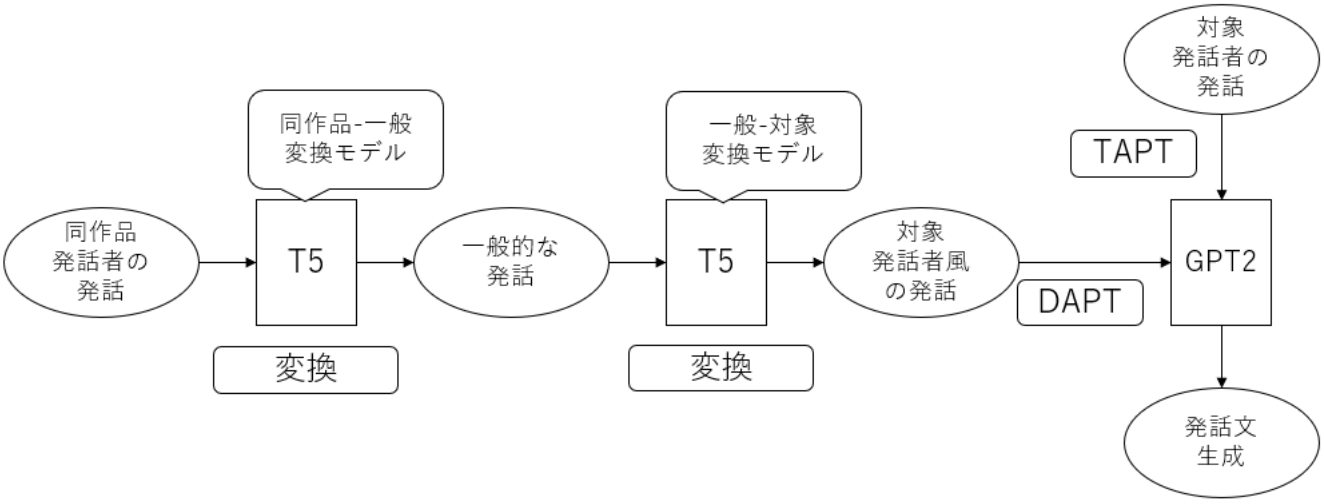

図 1 提案手法の流れ
Fig. 1 Flow of the proposed method

えば、対象発話者であるキャラクターの一人称である「あたし」から「私」、発話の末尾につけられる口癖である「だわ」や「よね」から「です」に修正した。これは「私」や「です」の方が使用者として考えられる人物や使われるシチュエーションが広いと考えられるからである。ここでの修正作業はルールに基づいたものではなく筆者の主観で行われた。また修正に際して、砕けた雰囲気からフォーマルな雰囲気になる、元の発話は語気の強いものだが、それがなくなるなど、若干のニュアンスが変わることは容認した。これは発話の持つ雰囲気が発話者の雰囲気でもあると考えたためである。同様にして、同一作品の別の登場人物の発話からも抽出、修正を行った。

続いて、T5 の学習を行い以下の 2 種類の変換モデルの作成を行った。

(1) 一般-対象変換モデル：一般的な発話を対象の発話者風の発話に変換する。対象の発話者の発話を一般の発話に修正したものを入力、修正前の対象発話者の発話を出力として学習した。
(2) 同作品-一般変換モデル：同一作品の別の登場人物の発話を一般的な発話に変換する。同一作品の別の登場人物の発話を入力、一般の発話に修正したものを出力として学習した。

3.3 T5 学習データの抽出方法

T5 の学習に用いたデータの抽出を以下の方法で行った。

まず、対象の発話をそれぞれ T5 で用いられているトークナイザーを用いて単語へ分割を行った。そして単語ごとに以下の式で TFIDF 値を求めた。

$$tf(t,d) = \frac{n_{(t,d)}}{\sum_{s \in d} n_{(s,d)}} \tag{1}$$

ここで、$tf(t,d)$ は文書 d のサブワード t における tf 値、$n(t,d)$ は文書 d におけるサブワード t の出現回数、$\sum_{s \in d} n(s,d)$ は文書 d における全サブワードの総出現回数を示す。

$$idf(t) = \log \frac{N}{df(t)} \tag{2}$$

ここで、$idf(t)$ は文書 d のサブワード t における idf 値、N は総文書数、$df(t)$ はサブワード t が出現した文書数を示す。

$$TFIDF = tf(t,d) * idf(t) \tag{3}$$

TFIDF 値は tf 値と idf 値を掛け合わせることで求められる。ここで TFIDF 値が高い単語はそのキャラクターの特徴を表していると考える。

続いて、これらの単語の TFIDF 値を用いて発話ごとの平均 TFIDF 値を求める。発話内に出てくる TFIDF 値の合計を出現単語数で割る。平均 TFIDF 値が高い発話ほどキャラクターの特徴を含んだ発話と考える。そのため、平均 TFIDF 値が高い発話を 100 個抽出し T5 の学習に利用した。

4. 実験

4.1 TAPT,DAPT 用のデータの作成

本研究では以下の 5 種類のデータを利用した。

- 対象の発話者の発話
- 同一作品の発話者の発話：同一作品の別の登場人物の発話
- CSJ：日本語話し言葉コーパス
- 対象風同作品の発話：「同一作品の発話者の発話」を「対象の発話者の発話」風に T5 を用いて変換したもの
- 対象風 CSJ：「CSJ」を「対象の発話者の発話」風に T5 を用いて変換したもの

本研究では「対象の発話者の発話」を TAPT に使用する対象タスクのデータとする。また、「同一作品の発話者の発話」、「対象風同作品の発話」を DAPT に利用する対象ドメインの学習データとする。さらに、比較対象として「CSJ」、「対象風 CSJ」を DAPT に用いて対象ドメインのデータとして利用した。本研究では対象の発話者としてゲーム、ドラゴンクエスト Ⅳ のキャラクターであるマリベルを利用し、同一作品の別の登場人物にはドラゴンクエスト Ⅳ のキャラクターであるキーファ、ガボ、メルビン、アイラ、リーサ姫を利用した。マリベルを対象としたのは、セリフを収集したキャラクターの中でもセリフ量が多かったためと、性別や人物像を表す話し方をしているためである。

「対象風同作品の発話」は、まず「同一作品の発話者の発話」を同作品-一般変換モデルを使い一般の発話に変換する。続いて一般-対象変換モデルを使い先ほど変換した一般の発話から「対象風同作品の発話」に変換する。以上の 2 段階の変換を行い「対象風同作品の発話」を作成した。「対象風 CSJ」は一般-対象変換モデルを使い CSJ を対象発話者風の発話に変換させて作成した。

4.2 実験の設定

4.2.1 TAPT,DAPT を利用した GPT2 の学習方法

GPT2 をそれぞれ以下の方法で Fine-tuning を行い、モデルごとに平均パープレキシティを求めた。

(a) TAPT のみ
(b) DAPT のみ
(c) TAPT+DAPT

(a)TAPT のみ　で「対象の発話者の発話」のみで学習したものをベースラインとする。(b)DAPT のみ　では、3.1 節で対象ドメインのデータとして説明した「同一作品の発話者の発話」、「対象風同作品の発話」を学習データとして Fine-tuning を行い、パープレキシティを求めた。また、比較対象として「CSJ」、「対象風 CSJ」を DAPT 用の対象のドメインデータとして利用する手法についても実験した。(c)TAPT+DAPT　では、(b)DAPT のみ　で作成したモデルにさらに「対象の発話者の発話」で Fine-tuning を行った。

GPT2 の学習は 20 エポックずつ 100 エポックまで行い、検証データを用いてパープレキシティを算出し、最も精度がよかった学習回数でテストを行った。この時、最も精度がよかった学習回数が 80 エポック、もしくは 100 エポックだった場合、追加で 200 エポックまで学習、検証を行った。

「対象の発話者の発話」の 9/10 を学習データとし、GPT2、T5 の学習に用いた。残りの 1/10 をパープレキシティの算出に利用した。そのうち 1/5 を検証データとし、4/5 をテストデータとした。

4.2.2 T5 の学習データセット

T5 の学習データセットには以下の 3 種類を使った。

A　TFIDF 値上位 100 発話
B　TFIDF 値上位 200 発話
C　ランダムな 100 発話

データセット A,B は、3.3 節で説明した方法で抽出した。またデータセット B,C は

- 「対象風同作品の発話」をドメインの学習データとし DAPT のみで学習
- 「対象風同作品の発話」をドメインの学習データ、「対象の発話者の発話」をタスクの学習データとし TAPT+DAPT で学習

の 2 パターンだけ検証を行った。

4.3 実験結果

TFIDF 値上位 100 発話を利用した場合の学習データによるパープレキシティの変化を以下の表 1 に示す。また、学習データに TFIDF 値上位 200 発話を利用した実験と、ランダムな 100 発話を利用した際の、T5 用の学習データの抽出方法によるパープレキシティの変化を表 2 に表す。ベースラインよりよい精度のものを太字で表す。

5. 考察

5.1 訓練データの選択

表 1 の結果から「CSJ」や「対象風 CSJ」を学習データに使ったモデルは精度がよくならなかった。これは「CSJ」には物語に出てこない学術的な単語が多く出てくるため、語尾や一人称が対象の発話者風である発話が生成されるモデルであったとしてもパープレキシティの精度は上がらなかったためだと考える。

また学習手法 (b)DAPT のみ　で「同一作品の発話者の発話」と「対象風同作品の発話」を比較した場合、T5 を使って対象の発話者風の話し方に変換したはずの「対象風同作品の発話」のほうが精度が悪かった。これは、T5 を使って 2 回変換を行っており、対象の発話者の特徴を掴んだ発話にはなっていても不自然な日本語になってしまったものが多くあったためだと考える。しかし、学習手法 (c)TAPT+DAPT　で比較すると「対象風同作品の発話」の方が良い結果になった。これは後で「対象の発話者の発話」で学習することで発話の不自然さから受ける影響が少なくなったためではないかと考える。

5.2 TF-IDF を用いた T5 訓練用発話文選択

表 2 の学習手法 (b)DAPT のみ　の結果を比較したとき、T5 の学習データとして TFIDF 値上位 200 文を抽出したときが最もよく、学習手法 (c)TAPT+DAPT　を比較した場合も同様によくなっている。学習手法 (b)DAPT のみ　で T5 の学習データとして TFIDF 値上位 100 文を抽出したときとランダムに 100 文を抽出したときはほぼ同程度であるが、このモデルに対象の発話者の発話を学習させた

表 1 学習データによるパープレキシティ（PPL）の変化

学習方法	T5 の学習データの抽出方法	ドメインの学習データ	タスクの学習データ	PPL
(a)TAPT のみ	TFIDF 値上位 100 発話		対象の発話者の発話	46.2295
(b)DAPT のみ	TFIDF 値上位 100 発話	同一作品の発話者の発話		59.4187
(b)DAPT のみ	TFIDF 値上位 100 発話	CSJ		136.7374
(b)DAPT のみ	TFIDF 値上位 100 発話	対象風同作品の発話		72.9026
(b)DAPT のみ	TFIDF 値上位 100 発話	対象風 CSJ		110.8460
(c) TAPT+DAPT	TFIDF 値上位 100 発話	同一作品の発話者の発話	対象の発話者の発話	**45.6060**
(c) TAPT+DAPT	TFIDF 値上位 100 発話	CSJ	対象の発話者の発話	48.2447
(c) TAPT+DAPT	TFIDF 値上位 100 発話	対象風同作品の発話	対象の発話者の発話	**44.1731**
(c) TAPT+DAPT	TFIDF 値上位 100 発話	対象風 CSJ	対象の発話者の発話	49.7503

表 2 T5 用の学習データの抽出方法によるパープレキシティ（PPL）の変化

学習手法	T5 の学習データの抽出方法	ドメインの学習データ	タスクの学習データ	PPL
(b)DAPT のみ	TFIDF 値上位 100 発話	対象風同作品の発話		72.9026
(b)DAPT のみ	TFIDF 値上位 200 発話	対象風同作品の発話		69.8814
(b)DAPT のみ	ランダムに選んだ 100 発話	対象風同作品の発話		72.6223
(c) TAPT+DAPT	TFIDF 値上位 100 発話	対象風同作品の発話	対象の発話者の発話	**44.1731**
(c) TAPT+DAPT	TFIDF 値上位 200 発話	対象風同作品の発話	対象の発話者の発話	**43.9297**
(c) TAPT+DAPT	ランダムに選んだ 100 発話	対象風同作品の発話	対象の発話者の発話	47.8923

モデルである学習手法 (c)TAPT+DAPT　では、TFIDF 値上位 100 文を選んだほうがが精度が上がっている。これらの結果から、TAPT+DAPT の手法で学習する場合は TFIDF を用いて T5 の訓練用発話文の選択を行う方法は有効であると考える。

6. おわりに

本研究では、特定の話者の言語モデルの構築方法として、T5 を用いて対象の発話者の発話を増補する方法を提案した。対象の発話者と同一作品の登場人物の発話をドメイン、対象の発話者の発話をタスクとし、TAPT-DAPT の手法を利用して Fine-tuning を行った。ベースの言語モデルには GPT2 を利用した。その結果、T5 を用いて対象の発話者と同作品のキャラクターの発話を対象の発話者風に変換したものをドメインのデータセットとして学習し、その後対象の発話者の発話をタスクのデータセットとして学習したものが最もよい精度となった。この結果から、提案手法が有効であることを示した。

また、T5 の学習データとして TFIDF を用いてキャラクターの特徴を表していると考えられる発話を選ぶことで、より精度を高めることができた。

謝辞 本研究は 2022 年度国立情報学研究所公募型共同研究（22FC04）また、JSPS 科研費 17KK0002 の助成を受けています。

参考文献

[1] 宮崎千明, 佐藤理史. 発話テキストへのキャラクタ性付与のための音変化表現の分類. 自然言語処理, Vol. 26, No. 2, pp. 407–440, 2019.

[2] Mika Kishino and Kanako Komiya. Extracting linguistic speech patterns of japanese fictional characters using subword units. *arXiv preprint arXiv:2203.02632*, 2022.

[3] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. Language models are unsupervised multitask learners. *OpenAI blog*, Vol. 1, No. 8, p. 9, 2019.

[4] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, Peter J Liu, et al. Exploring the limits of transfer learning with a unified text-to-text transformer. *J. Mach. Learn. Res.*, Vol. 21, No. 140, pp. 1–67, 2020.

[5] Suchin Gururangan, Ana Marasović, Swabha Swayamdipta, Kyle Lo, Iz Beltagy, Doug Downey, and Noah A Smith. Don't stop pretraining: adapt language models to domains and tasks. *arXiv preprint arXiv:2004.10964*, 2020.

[6] 勝又智, 坂田大直. CSJ を用いた日本語話し言葉 BERT の作成. 言語処理学会 第 27 回年次大会 発表論文集, pp. 805–810, 2021.

[7] François Mairesse and Marilyn Walker. Personage: Personality generation for dialogue, 2007.

[8] Marilyn A. Walker, Grace I. Lin, and Jennifer E. Sawyer. An annotated corpus of film dialogue for learning and characterizing character style. In *the Proceedings of LREC 2012*, pp. 1373—-1378, 2012.

[9] 宮崎千明, 平野徹, 東中竜一郎, 牧野俊朗, 松尾義博, 佐藤理史. 話者のキャラクタ性に寄与する言語表現の基礎的分析. 言語処理学会 第 20 回年次大会 発表論文集, pp. 232–235, 2014.

[10] Chiaki Miyazaki, Toru Hirano, Ryuichiro Higashinaka, and Yoshihiro Matsuo. Towards an entertaining natural language generation system: Linguistic peculiarities of japanese fictional characters. In *the Proceedings of SIGDIAL 2016*, pp. 319—-328, 2016.

[11] 宮崎千明, 平野徹, 東中竜一郎, 牧野俊朗, 松尾義博, 佐藤理史. 文節機能部の確率的書き換えによる言語表現のキャラクタ性変換. 人工知能学会論文誌, Vol. 31, No. 1, pp. DSF–515, 2016.

[12] 奥井颯平, 中辻真. ポインタ生成機構を用いたキャラク

ター応答生成の検証. 第 34 回人工知能学会全国大会論文集, pp. 1I4–GS–2–01, 2020.