

抽象型要約における教師なしドメイン適応のための データ拡張

名原 光洋¹ 藤田 桂英^{†1}

概要: 自然言語処理 (NLP) タスクの 1 つである自動要約タスクでは、他の NLP タスクと同様に大規模なデータによる事前学習を行ったモデルを、小規模なターゲットドメインにドメイン適応させることができる。これにより様々なドメインの文書に対して、自動要約を行うことが可能である。自動要約タスクはその手法によって抽出型と抽象型の 2 つに大別することができ、その中でも抽象型要約は被要約文書の表現に捉われることなく柔軟かつ完成度の高い要約文を生成する。抽象型要約は深層学習に基づいており、その学習には正解データとなる要約が付与された大量のデータが必要になる。このような大規模な抽象要約に適したデータセットはドメインが限られているため、抽象要約が実現できるドメインも限られている。抽象要約で使われるデータセットの 1 つである CNN - Daily Mail コーパス (CNN-DM) はニュース記事とその要約になるいくつかの文からなる見出しがセットになっているデータセットである。本研究ではニュース記事とは性質を大きく異なるレビューサイトに寄せられたレビューの抽象要約に注目する。小規模な教師なしレビューデータセットの抽象要約を実現するために大規模データセットで事前学習したモデルのドメイン適応を前提としたデータ拡張手法を提案する。文脈を考慮した単語の置き換えによるデータ拡張によって、ROUGE において既存手法を上回る精度を確認した。

1. はじめに

自然言語処理タスクは多様かつ複雑な言語を相互変換する翻訳タスクや、様々な質問に適切な解答を返す質疑応答タスクを始め様々な形で実際の生活の中に応用されるようになった [1]。そして様々な自然言語処理タスク (NLP タスク) が研究の対象となり、それらを解くのに機械学習が盛んに用いられている。企業や研究機関が大規模なデータで長時間かけて機械学習した言語モデルなどはインターネットを通じて公開され、様々なタスクやデータに合わせて転移学習を施すことで従来より高性能なモデルを作成できるようになった。すなわち大規模データで事前学習されたモデルを他のタスクや他のデータ、他のドメインで再学習 (転移学習) するだけで大規模な学習の恩恵を受けたモデルを作ることができる [2], [3], [4]。しかし大規模なデータを前提とした設計をされているモデルに対し小規模なデータで

再学習を施すことは簡単ではなく、過学習が起き、大規模な学習の恩恵を失うことも多い。さらに深層学習の場合、学習に必要な不可欠な正解データ・ラベルデータが存在しないデータにドメイン適応させるのを教師なしドメイン適応とし、一般的に正解データ・ラベルデータの存在する教師あり学習より難易度が高い。

一方、本研究で取り扱う自動要約タスクでは抽象型要約 (生成型要約) の研究が進んでいる。抽象型要約は被要約文書の意味を要約器が解釈して新たに要約文書を生成する。そのため、出力される要約文書は 1 つの文書として完成度が高い。抽象型要約では、元文書の文や表現に捉われない柔軟な要約生成が可能である一方、他の深層学習ベースの手法と同じく、学習データである被要約文書と要約文書のペアを大量に必要とする。その結果として抽象型要約はニュース記事のように見出しがついた文章データなどの限られたドメインのデータを中心に研究が進んでいる。例えば CNN-Daily Mail (CNN-DM) コーパス [5], [6] は、英文書の自動要約において広く使われているニュース記事とその要約である 3 文もしくは 4 文の見出しがペアとなったデータセットである。しかし、このように抽象要約の正解とできる適切な見出しなどが付与された大規模なデータセットは種類が限られているため、これらの大規模データと合わないドメインのデータにはドメイン適応が必要になる。

¹ 東京農工大学 工学部 情報工学科

Department of Computer and Information Sciences, Faculty of Engineering, Tokyo University of Agriculture and Technology

^{†1} 東京農工大学大学院 工学研究院 先端情報科学部門

Presently with Division of Advanced Information Technology and Computer Science, Institute of Engineering, Tokyo University of Agriculture and Technology

本研究で注目する抽象型要約の教師なしドメイン適応では、擬似的な正解データを用意したりアノテーションを施したりしてから、再学習するのが一般的である。抽象要約のアノテーションを行うのはコストが高いため、抽出的な手法で擬似要約を作成し使用することが多い。例えば、被要約文書の最初の文から既定文字数分抜き出す手法や、軽コストな抽出要約の手法などを用いる。しかしこれらのような抽出的な手法で作成された擬似要約は、抽象要約によって作成される要約と性質が大きく異なる。そのため抽出的な擬似要約は抽象要約の再学習に悪影響を及ぼし、再学習前は抽象要約を出力できていたにも関わらず、再学習後では抽出的な要約を出力するようになる。これは再学習の際にモデルがドメイン適応先のデータに付与された抽出的な擬似要約の性質を学習してしまうからだと考えられる。そのため、擬似的な正解データとなる要約が抽象要約的なものかどうかを抽象要約の教師なしドメイン適応で重要である。

本研究では教師なしドメイン適応先としてレビューサイトに寄せられたレビューを用いる。ホテルやレストランを紹介する web サイトにはレビューが多く寄せられている。レビューサイトでは一般的に満足度などを数値で表すと共にユーザが書いたレビュー文を載せている。web サイトの視認性向上のために長文で書かれたレビューの規定文字数を超えている部分は省略されることが多く要約の必要性は高い。これらのレビューにはニュース記事の見出しのように要約として取り扱える部分がないため、再学習では擬似的な正解データとなる要約である擬似要約を作成する必要がある。擬似要約は何らかの抽出的な手法によって作成されるものが多く、被要約文の先頭に重要情報が集まっているとして、先頭部分を擬似要約とする手法などが用いられている。事前学習で用いるデータである CNN-DM のようなニュース記事は事実に基づき客観的な観点から文法に忠実な文章で書かれている。しかしレビューでは使われている語彙や表現もニュース記事と大きく異なり省略やスラング的表現が多い。またレビューはサービスなどを高評価するものと低評価するものの偏りがある。ドメイン拡張先のデータ数が少ない場合、これらのドメイン適応先の特徴は過学習の原因となる。そこで他の深層学習を用いるタスクで、小規模なデータに行うのと同様にデータ拡張を行い、データの偏りやデータの少なさに起因する過学習を低減させることが重要である。

本研究では転移学習のうちドメイン適応に注目して、事前学習に用いるデータとはドメインの異なる小規模データにデータ拡張を施し、抽象型要約の教師なしドメイン適応での性能を向上させることを目的とする。大規模ニュース要約コーパス CNN-DM をソースドメイン、各 web レビューサイトのレビューをターゲットドメインとして、ターゲットドメインのデータを拡張し、より適切な擬似正解データ

の作成手法を提案する。

提案手法は類似データの作成によるデータ拡張手法である。類似データの作成では学習済みの言語モデルを用いて単語を文脈に合わせて適切に置き換えることで単語や表現の多様性を確保し、偏りを低減させる Word Level Rephrasing(WLRp) を提案する。WLRp を施した類似データによるデータ拡張を行う。

CNN-DM で事前学習した抽象要約モデルを擬似要約を付与したレビューデータで再学習する実験を行う。再学習ではデータ拡張前のデータとデータ拡張後のデータのそれぞれで抽象要約モデルの学習を行い、それぞれのモデルの性能を比較する。また WLRp による適切な類似データの作成数を決定するために、作成した類似データ数の異なるデータセットを3種類用意し、それぞれで学習した抽象要約モデルの性能を比較する。モデルの性能は抽象要約の代表的な定量評価手法である ROUGE による定量評価と、定性的な評価を行う。これらの評価によって WLRp による類似データの作成によるデータ拡張が抽象要約のドメイン適応に有効であるかを確認する。

2. 提案手法

2.1 Word Level Rephrasing

先行研究である文書分類用のデータ拡張手法 EDA では同義語で置換する際に同義語辞典から同義語をランダムに選択していた。しかし、同義語で置換した結果、文脈に合わない単語(意味的・文法的)や使われることが少ない単語を使うことになることがある。文脈に合わない単語を使用すればモデルが自然な文章を生成できなくなる。また使用頻度の低い単語はモデルの語彙を作成する際に選出されないため、未知語になってしまう。未知語がモデルの性能を下げる原因の1つになっているのは一般的である。

そこで同義語を同義語辞典からランダムに選択するのではなく、文脈に合わせて置換後も自然な文章を保てるように置換を行うのが理想的である。本研究では、同義語による置換に言語モデル BERT[3] を用いる。BERT は大規模なデータセットを用いた大規模な言語モデルであり、様々な NLP タスクで活用され発表当時、多くの State-of-the-art を更新した。その特徴としてモデルの学習時に Masked Language Modeling(MLM) と Next Sentence Prediction(NSP) を行っていることが挙げられる。MLM は文章中の単語の1つを MASK トークンに置き換え、その MASK トークンの予想をすることで言語モデルを学習させる手法であり、NSP は入力された2文が連続した文かどうかを判別するタスクである。この2つの学習を通して BERT は文章を双方向の文脈で理解している。また、この学習過程で MLM を行っているため、単語を双方向の文脈で考慮して置き換えることができる。置き換えたい単語を MASK トークンで置き換えて BERT に入力することで BERT はその MASK トーク

ンがどの単語だったのかを確率分布付きで出力するため、この確率分布の上位 N_{rep} 個から置換単語を選択する。しかしこの BERT だけで置換を行うと文法上は正しいが意味的に異なる文になってしまうことがあるため、同義語辞典を併用する。

以下にその置換手順を示す。

- (1) 置換したい単語 w を MASK トークンに置き換えて学習済み BERT に入力し、尤度付き置き換え候補単語群を得る。
- (2) 同義語辞典から単語 w の同義語群を取得する。
- (3) 得られた単語群の両方に存在する単語のうち (1) で得られた尤度が最大なもので置換する。ただし (1) と (2) で共通する単語がない場合、置換は行われぬ。

図 1 で WLRp の適用例を示す。この適用例では “Frankly” に対して、BERT の出力から得られた候補語群と類義語辞典から得られた候補語群で重複している “honestly” で置換を行っている。また (3) で置換可能な語数に対して確率 $\alpha \neq 1$ で置換を行うことで、データ 1 件に対してそれぞれ異なるいくつかの拡張データを n_{aug} 件作成することができる。 $\alpha = 1$ の場合、作成される拡張データの件数は元のデータ数と同じかそれ以下である。しかし文書を通じて全てのトークンで置換が行われぬ可能性があるため、データの拡張件数は保証されない。文書内の全ての単語で置換が起きぬ可能性は大変低く、また既存手法である EDA の SR でも同様に置換が発生しぬ可能性は存在する。そこで評価実験を行う際には拡張後のデータ件数についても比較を行う。

置換を行う間、BERT のパラメタの更新は行われぬ。今回の実験では $\alpha = 1.0, 0.5, 0.1$ の 3 通りでそれぞれ 1 件のデータに対して、 $n_{aug} = 1, 2, 4$ 件のデータを作成する。表 1 に置換確率 α と作成するデータ件数についてまとめる。

表 1: WLRp でデータ 1 件ごとに作成するデータ件数

α	n_{aug}
0.1	4 件
0.5	2 件
1.0	1 件

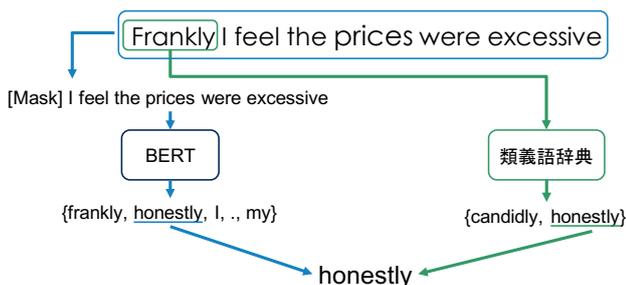


図 1: Word Level Rephrasing の適用例

2.2 学習方法

ソースドメインである大規模な教師ありデータによる事前学習と、ターゲットドメインである小規模な教師なしデータによる再学習に分かれる。事前学習で抽象要約タスクの特徴を学習し、再学習でターゲットドメインの特徴を学習する。事前学習終了時にモデルの全ての中間層の状態を引き継いで、擬似要約付きのターゲットドメインのデータで再学習を行う。以上の流れを図 2 に示す。事前学習と再学習で使用するモデル・ハイパーパラメタは共通であるが、学習がハイパーパラメタに含まれる最大エポック数に達する前に早期終了する場合がある。1 エポックの学習終了ごとのバリデーションを行う。バリデーションを行った際の損失が $N_{patience}$ エポック連続で減少しなかった場合、モデルが改善されなかったとみなし学習を早期終了する。また事前学習と再学習で使用する語彙は統一する。ソース・ターゲットの両ドメインで登場する全ての単語で登場回数の多い単語を語彙とする。

3. 実験

3.1 実験設定

提案した手法とベースラインとなる手法の実験を行う。擬似要約は先頭文要約を用いる手法で作成する。提案手法の実験は WLRp によるデータ拡張をそれぞれ異なる α と n_{aug} で施したデータセット 3 種類で行う。従来手法としてデータ拡張をしないものと EDA によるデータ拡張を施したものの 2 種類を行う。以下の表 2 にデータ拡張手法拡張後のデータ件数を示す。データ件数は拡張前のものと比して何倍かで示している。

実験で使用するシステムは Python3 を用いて実装し、深層学習のモデル実装には Pytorch と Pytorch-lightning を用いる。英文のトークナイザーとして NLTK を使用する。実験で用いる言語モデル BERT は huggngface のライブラリ transformers で公開されている学習済みモデルを使用する。使用する BERT は base-uncased であり、大文字小文字の区別をしない wikipedia のデータで学習されたものである。

ドメイン適応の再 train loss には抽象要約モデルの全ての層 (単語埋め込み層, Encoder と Decoder の中間層) で学習

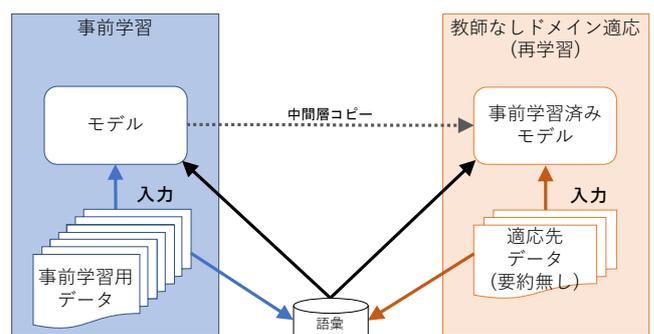


図 2: 事前学習と再学習の流れ

表 2: 実験の概要

提案手法	
データ拡張手法	データ件数
WLRp($\alpha = 0.1$)	5 倍
WLRp($\alpha = 0.5$)	3 倍
WLRp($\alpha = 1.0$)	2 倍
比較手法	
データ拡張手法	データ件数
なし	1 倍
EDA($\alpha = 0.1$)	5 倍

を行う。通常ではドメイン適応などの再学習を行う際に、一部の層だけを再学習するが多いが、本研究では事前学習で用いるデータとドメイン適応先のデータの性質の乖離が大きいと判断し、単語埋め込み層と中間層の両方で再学習を行う。

抽象要約モデルとして Pointer-Generator Networks を使用する。類義語辞典には NLTK の WordNet を用いる。実験は ROUGE-1, ROUGE-2, ROUGE-L による定量評価と定性評価を行う。実験に使用するデータについて説明する。抽象要約の事前学習に CNN-DM を用いる。ドメイン適応先のデータにはホテルのレビューデータを用いる。CNN-DM のデータ件数は 311971 件であり、レビューデータのデータ件数は 11470 件である。またテキストは全て小文字に統一する。

全ての実験で共通するパラメータについて説明する。まず抽象要約モデルのパラメータについて説明する。隠れ層の次元数は 512、単語埋め込み層は 256 とする。Pointer-Generator Networks の元論文の設定にならって、最適化アルゴリズムには Adagrad を用いて学習率を 0.15、アキュムレーターの初期値を 0.1 とする。事前学習とドメイン適応での再学習において語彙は統一する。2つのデータセットの出現単語の出現頻度上位 80000 単語を語彙に収録する。テスト時には Beam search を行う。この時 Beam 幅は 4 とする。

被要約文の最大長を 400 トークンとし、擬似要約文の最小長を被要約文長の 15% とする。被要約文の先頭から最小長を超えるように、文単位で被要約文から文を抽出して擬似要約を作成する。要約文の最大長を 100 トークンとする。超えた場合には切り捨てる。アノテーションによって作成された正解データも同様の長さである。

抽象要約の事前学習で用いる学習用データ 311971 件のうち 160 件を validation 用のデータとして、残りのデータを訓練用のデータとする。訓練用のデータで学習後、validation 用のデータを入力し、損失が改善されなかった場合にモデルの学習を早期終了するようにした。最大 epoch 数を 35 として学習を始め、34epoch 終了時に学習が早期終了した。また coverage はドメイン適応後に行うため、ここでは使用

しない。

抽象要約の再学習では、最大 35epoch 学習を行うが、事前学習同様に早期終了できる。その後、coverage loss で 3epoch 学習を行う。学習に要したエポック数は表 3 にまとめた。テストには、レビューデータの一部 200 件にアノテーションが施されているため、これを用いる。テストデータはモデルの学習には使用しない。validation 用のデータはドメイン適応先の学習データのうち 10 バッチ分にあたる 160 件を使用する。これはデータ拡張によってデータ数が増加しても変更しない。再学習の際には事前学習よりも少なく半分以下のエポック数で学習が終了していることがわかる。

表 3: ドメイン適応に要したエポック数 (Coverage で学習したエポック数を除く)

データ拡張手法	データ数	エポック数
拡張なし	11470	9
EDA($\alpha = 0.1$)	57350	9
WLRp($\alpha = 0.1$)	57350	7
WLRp($\alpha = 0.5$)	34410	11
WLRp($\alpha = 1.0$)	22940	16

3.2 実験結果

教師なしドメイン適応の定量評価

レビューデータ 200 件に人手による要約を付与したアノテーションデータを正解要約として、生成された要約と ROUGE-1, ROUGE-2, ROUGE-L の値を計算し表 4 にまとめた。データ拡張手法ごとの F 値を示している。先頭文要約は被要約文書の先頭部分を擬似要約としたものである。ROUGE-1, ROUGE-2, ROUGE-L の全てで置換確率 $\alpha = 0.1$ の WLRp を用いてデータ拡張をした場合が最もモデルの精度が良かった。

表 4: 先頭文要約を用いた教師なしドメイン適応の ROUGE-1, 2, L 結果

データ拡張手法	ROUGE-1	ROUGE-2	ROUGE-L
拡張なし	0.178	0.053	0.148
EDA($\alpha = 0.1$)	0.161	0.040	0.135
WLRp($\alpha = 0.1$)	0.190	0.060	0.157
WLRp($\alpha = 0.5$)	0.174	0.046	0.146
WLRp($\alpha = 1.0$)	0.173	0.049	0.145

生成された要約文の比較

生成された要約文の例を被要約文とアノテーターの作成した要約文と共に示す。以下は被要約文にあたるレビュー文とアノテーターの作成した要約である。以下に示されている文章はそれぞれ空白の使い方が本来の英文のそれとは

異なっているが、これは本研究で用いた NLTK の Tokenizer による分割の結果を示しているためである。また大文字は全て小文字に統一されている。

アノテーターはレビューの感情を 1 文目にまとめ、2 文目に清潔さなど観点をあげた上で評価をまとめている。

Review(original)

this is a new hotel in an easy access location . the front desk staff was very friendly , courteous and helpful . the hotel is very clean and modern as are the rooms . our room was clean and very comfortable . this is one of the nicest holiday inn 's that we have stayed in . breakfast was great with lots of choices . do n't believe the one negative review entitled , " terrible " . this hotel is way beyond expectations for cleanliness , comfort and security .

アノテーターの作成した要約

one of the nicest holiday inn 's we have stayed in . way beyond expectations for cleanliness , comfort , and security .

次に以上の被要約文に対して生成された文章を示す。データ拡張の手法と擬似要約の種類ごとに以下に生成された要約をまとめた。

データ拡張:なし, 擬似要約:先頭文要約

this is a new hotel staff friendly , courteous .

データ拡張を行わない場合、先頭文要約で学習したモデルの生成した要約は被要約文の冒頭 2 文を要約したものになっているが、文法的に不自然な形になっている。

データ拡張:EDA, 擬似要約:先頭文要約

this is a new hotel in an easy this is a negative review this is . the front desk staff was very friendly , courteous .

EDA によるデータ拡張を施した場合、先頭文要約で学習したモデルの生成した要約は被要約文の冒頭部分と似ているが、不自然な単語の挿入などが見られる。

データ拡張:WLRp($\alpha = 0.1$), 擬似要約:先頭文要約

this is a new hotel in an easy access location . the front desk staff friendly this is very friendly , courteous .

データ拡張:WLRp($\alpha = 0.5$), 擬似要約:先頭文要約

this is a new hotel in an friendly , courteous .

データ拡張:WLRp($\alpha = 1.0$), 擬似要約:先頭文要約

this is a new hotel this is very friendly , courteous .

WLRp でデータ拡張を施し先頭文要約で学習したモデル

の生成した要約は全て被要約文の冒頭 2 文を踏まえたものになっている。特に置換確率 $\alpha = 0.5, 1.0$ では被要約文の冒頭 2 文を要約したものになっている。

3.3 考察

定量評価と生成された要約文の関係

4 から WLRp($\alpha = 0.1$) でデータ拡張を施し、被要約文の先頭部分を擬似要約としたものが全体的に ROUGE のスコアが高かった。一方でデータ拡張の有無や手法に関わらず先頭文要約を擬似要約とした場合、モデルの生成する文は被要約文の冒頭を抽出した部分が多く存在した。元文書の先頭部分をそのまま抽出すれば、損失が小さくなるためこのような形に収束したと考えられる。大規模な抽象要約データで学習した抽象要約タスクの特徴をモデルが再学習時に失ってしまう結果となった。学習済みモデルを用いて学習したタスクの特徴を忘れないようにしながら、ドメイン適応先のデータの特徴だけを学習する必要がある。そのため、モデルの中間層全体ではなく Embedding 層など一部の層だけを再学習する、ドメイン適応先のデータに事前学習のデータを混合する、学習を今回のそれより早期に終了するなどの対策が考えられる。またテスト時に用いた Beam-Search は出力する単語の決定に尤度の大きさをを用いる。モデルが単語の生成よりも単語の抽出に特化したということは Pointer-Generator Networks における単語の生成確率 P_{gen} が小さくなったということであるため、Beam-Search において単語を生成する結果が少なくなってしまう。Pointer-Generator Networks の目的関数である対数尤度と抽象要約の良さは一致しないため、抽象要約を正解データとしてモデルの学習ができている場合のみ Beam-Search をする際に対数尤度で生成文を比較できる。そのため、抽出要約を擬似要約として学習した場合には抽象要約を定量的に比較できる手法を用いて Beam-Search を行うか、もしくは他の手法でモデルの出力する確率分布から単語を選択する必要がある。

3.3.1 単語の置き換えと生成された要約文の関係

表 4 から EDA による文脈を無視した類義語の挿入や置換よりも WLRp による双方向の文脈を考慮した単語の置換の方が、定量評価において結果が優れていた。文脈に沿わない単語の削除や入れ替えや挿入などは学習におけるノイズとしてモデルに悪影響を与えるが、適切な単語の置き換えによって表現の偏りが低減され、少ない学習データに対して過学習を抑えることがわかった。ただし擬似要約に抽出型要約を用いたことでモデルが単語の生成よりも単語の抽出に特化したため、テスト結果全体の中でその影響は小さかった。擬似要約において 1 単語を対象にした置き換えだけでなく、いくつかの単語の集まりであるフレーズを抽象的なものに置換することで、抽象的なフレーズの生成につながると考えられる。

4. まとめ

本論文では、抽象型要約の教師なしドメイン適応に注目して、レビューサイトにおける投稿文書の抽象型要約を実現することを目的とし、小規模レビューコーパスからレビュー向け要約を実現するためのデータ拡張手法を提案した。提案手法では双方向の文脈を考慮した単語の置換アルゴリズムである WLRp を定義した。提案手法を用いて小規模レビューコーパスをデータ拡張して教師なしドメイン適応の実験及び評価を行った。提案した WLRp を利用したデータ拡張を行って学習したモデルは、既存のデータ拡張手法や拡張を行わなかった場合と比べて ROUGE スコアにおいて高いスコアを示した。文章内の単語の不適切な置換を無くすことで、文章にノイズを発生させずに学習用のデータ件数を増やすことに成功した。被要約文章の先頭部分を用いた擬似要約は ROUGE で高いスコアを記録したものの、抽出要約の傾向が強く見られた。抽象要約の教師なしドメイン適応に用いる擬似要約の作成には改善の余地が見られる結果となった。

翻訳タスクや要約タスクなどの系列生成タスクにおいて、モデルの出力する確率分布から単語を選択する手法に強化学習を利用する手法 [7], [8], [9] や、GAN を用いた文章生成を行う手法 [10] などが提案されている。強化学習や GAN を利用することで、正解となる要約が利用できない場合においても、学習に必要な時間は増加するものの効果的な学習を行うことが期待できる。教師なし学習に適したモデルを利用することを検討し研究を進める必要がある。

また 1 文を文意を保ったまま別の形で言い換える paraphrasing タスクを行うモデル [11], [12], [13], [14] が提案されている。これらの paraphrasing モデルを利用して、文意を保ったまま被要約文を別の形に言い換えることでデータ件数を増加させる手法を提案し、実現することが展望として挙げられる。また被要約文だけでなく抽出的に作成した擬似要約を言い換えることでより抽象的な擬似要約の作成が期待できる。

参考文献

[1] Y. Wu, Mike Schuster, Z. Chen, Quoc V. Le, Mohammad Norouzi, Wolfgang Macherey, M. Krikun, Yuan Cao, Q. Gao, Klaus Macherey, Jeff Klingner, Apurva Shah, M. Johnson, X. Liu, L. Kaiser, S. Gouws, Y. Kato, Taku Kudo, H. Kazawa, K. Stevens, G. Kurian, Nishant Patil, W. Wang, C. Young, J. Smith, Jason Riesa, Alex Rudnick, Oriol Vinyals, G. S. Corrado, Macduff Hughes, and J. Dean. Google's neural machine translation system: Bridging the gap between human and machine translation. *ArXiv*, abs/1609.08144, 2016.

[2] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21(140):1–67, 2020.

[3] 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)*, pages 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.

[4] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9, 2019.

[5] Karl Moritz Hermann, Tomáš Kočiský, Edward Grefenstette, Lasse Espeholt, Will Kay, Mustafa Suleyman, and Phil Blunsom. Teaching machines to read and comprehend. *arXiv preprint arXiv:1506.03340*, 2015.

[6] Ramesh Nallapati, Bowen Zhou, Caglar Gulcehre, Bing Xiang, et al. Abstractive text summarization using sequence-to-sequence rnns and beyond. *arXiv preprint arXiv:1602.06023*, 2016.

[7] Yaser Keneshloo, Tian Shi, Naren Ramakrishnan, and Chandan K Reddy. Deep reinforcement learning for sequence-to-sequence models. *IEEE transactions on neural networks and learning systems*, 31(7):2469–2489, 2019.

[8] Mohammad Norouzi, Samy Bengio, Navdeep Jaitly, Mike Schuster, Yonghui Wu, Dale Schuurmans, et al. Reward augmented maximum likelihood for neural structured prediction. *Advances In Neural Information Processing Systems*, 29:1723–1731, 2016.

[9] Wojciech Kryściński, Romain Paulus, Caiming Xiong, and Richard Socher. Improving abstraction in text summarization. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1808–1817, Brussels, Belgium, October–November 2018. Association for Computational Linguistics.

[10] Lantao Yu, Weinan Zhang, Jun Wang, and Yong Yu. Seqgan: Sequence generative adversarial nets with policy gradient. In *Proceedings of the AAAI conference on artificial intelligence*, volume 31, 2017.

[11] AB Siddique, Samet Oymak, and Vagelis Hristidis. Unsupervised paraphrasing via deep reinforcement learning. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1800–1809, 2020.

[12] Chaitra Hegde and Shrikumar Patil. Unsupervised paraphrase generation using pre-trained language models. *arXiv preprint arXiv:2006.05477*, 2020.

[13] Zichao Li, Xin Jiang, Lifeng Shang, and Qun Liu. Decomposable neural paraphrase generation. *arXiv preprint arXiv:1906.09741*, 2019.

[14] Zichao Li, Xin Jiang, Lifeng Shang, and Hang Li. Paraphrase generation with deep reinforcement learning. *arXiv preprint arXiv:1711.00279*, 2017.