Vol.2021-AAC-17 No.2
2021/12/3

言語処理モデル mT5 と UD トークによる 盲ろう者向け自動要約筆記システム

市川涼介[1]　村田 勇樹[1]　堀江 則之[1]　巽 久行[1]

概要：視聴覚に障害をもつ盲ろう者への即時的な要約筆記を実現するためには，専門知識と技術を有する支援者が３人程度必要であり，事前準備や人材の確保など，利用者及び支援者双方の負担が大きい．点字による情報取得が可能な盲ろう者は単位時間あたりに取得できる情報量が少ないため，内容を崩さない範囲内で，可能な限り文章の冗長部分を削除する必要がある．本研究では，自然言語処理モデルである mT5 や UD トーク等を活用し，要約筆記に必須となる発話者タグも含め自動変更することのできる自動要約筆記システムを提案し ROUGE 指標などを用いて有効性の検証を行った．その結果，原文と比較して 2 割程度の文字列を圧縮することなどができ，一定の有効性が示された．発話者タグの自動変更は，メル周波数ケプストラム（MFCC）とサポートベクターマシン（SVM）を用いて実装した．SVM の精度は 9 割程度となったことから，MFCC 等を用いた発話者タグの自動変更は実現可能である．

キーワード：盲ろう者，要約筆記，点字，UD トーク，自然言語処理，深層学習

Automatic Summary Writing System for the Deafblind by mT5 Natural Language Processing Model and UDTalk

Ryosuke Ichikawa[†1]　Yuki Murata[†1]　Noriyuki Horie[†1]　Hisayuki Tatsumi[†1]

Abstract: In order to achieve immediate summary writing for deafblind individuals with audiovisual disabilities, about three supporters with specialized knowledge and skills are required for each deafblind individual. Such advance preparation and the securing of human resources impose a heavy burden on both the user side and the supporter side. Deafblind individuals who can only obtain information in Braille have a small amount of information that can be acquired per fixed time, so it is necessary to remove redundant parts of sentences as much as possible within the condition that key conversational content is not destroyed. In this research, we utilized an mT5 natural language processing model and UDTalk to create an automatic summary writing system that can change sentences including speaker tags, which are essential for summarization processing. We evaluated our system using the ROUGE index and others and found that it was able to compress about 20% of strings compared to the original text, indicating a certain level of effectiveness. We also implemented an automatic change of speaker tags by using the mel frequency cepstral coefficient (MFCC) and a support vector machine (SVM). Results showed that the SVM accuracy was about 90% and that an automatic change of speaker tags using MFCC was feasible.

Keywords: Deafblind, Summary writing, Braille, UD Talk, Natural language processing, Deep learning

1. はじめに

1.1 盲ろう者のコミュニケーション

盲ろう者とは，視覚と聴覚の両方に障害をもつ者を意味する．現在，日本全国には 23,000 人余りの盲ろう者が生活していると推定されている[1]．視聴覚障害を負った順序に応じて，先天盲ろう，盲ベース，ろうベース，などに大別される．また，視聴覚障害の状況に応じても，全盲ろう，盲難聴，弱視ろう，弱視難聴の４領域に区分されており，それぞれの障害領域に応じて，他者とのコミュニケーション方法も大きく異なる特徴が見られる．図１は，視聴覚障害の状況による盲ろう者の分類と主なコミュニケーション方法を示したものであり，多様化する盲ろう障害が見て取れる．

例えば全盲ろうの場合，活用できる感覚器官は視聴覚以外の触覚などに限定されるため，触手話や指点字，点字筆記が中心となる．また，盲難聴の場合，活用できる感覚器官は視覚以外の触覚や弱い聴覚などに限定される一方，前述した方法に加えて，補聴器等を活用した音声による会話などが可能な場合もある．

本研究で対象となる盲ろう者は，重度視覚障害を伴う盲ろう者，つまり，全盲ろうや盲難聴などを想定しているため，点字による情報取得が可能な者に限定する．よって，全盲ろうの状態にある盲ろう者であっても点字の利用が難しく，触手話などによるコミュニケーションが確立している場合には，本研究の対象から除外する．そして，視覚から活字情報を取得できる弱視ろうや弱視難聴の盲ろう者も同様に本研究の対象から除外する．活字情報を利用できる聴覚障害者や盲ろう者を対象とする要約筆記と点字情報を利用する盲ろう者向けの要約筆記では，提示する情報量が大きく異なるという点に留意する必要がある．

1 国立大学法人 筑波技術大学
Tsukuba University of Technology

ⓒ2021 Information Processing Society of Japan

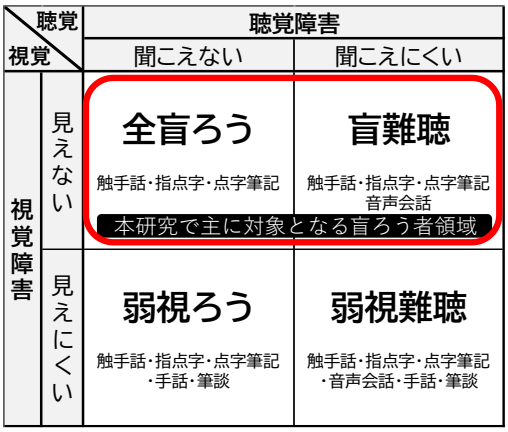

図 1　盲ろう者の分類
Figure 1　Classification of deafblind individual.

これは，単位時間あたりに取得することが可能な情報量に違いが見られるためである．牟田口らが 16 歳から 62 歳までの男女 36 人の点字熟達者を対象に実施した調査[2]によれば，1 分間あたりの点字読書量は平均 424.2 文字である．一方で，小林らが大学生 200 名を対象に実施した調査[3]によれば，1 分間あたりの読書量は平均 653 文字である．年齢的な差異は認められるものの 1 分間あたり 200 文字以上の開きがあることがわかる．発話内容を即時的に要約し，文字情報や点字情報に変換したものを利用者に届ける要約筆記では，正確で迅速な作業が求められる．なぜなら，発話内容が逐次更新される中で，要約筆記の利用者が発話者や他者と時間的，または，空間的な面でも状況を常に共有し，当事者として参画することが重要だからである．よって，要約筆記の利用者もまた即時的に情報を取得し，利活用できることが求められるのである．そのため，点字を利用する盲ろう者を対象とする要約筆記では，発話の内容を崩さない範囲内で可能な限り，文章の冗長部分を削除し，利用者に届ける情報の量を調整する必要がある．

1.2　点字利用者における現状と課題

著者らが所属する筑波技術大学保健科学部は，視覚に障害をもつ学生を専門に受け入れており，盲ベースの盲ろう者や重度の視覚障害を伴う盲ろう者も含まれる．盲ろうの学生が参加する授業（図 2 参照）では，専門の要約筆記者が 3 名程度同席し，逐次要約作業を行っている．

活字情報を利用する聴覚障害者や盲ろう者を対象とする要約筆記に比べ，発話の内容をさらに圧縮し，要約することが求められるため，点字を利用する盲ろう者向けの要約筆記では点字をはじめとする盲の世界の知識とより高度な技術が求められる．一方で，高度な知識と技術をもつ要約筆記者を安定的に確保することが難しい点や時間及び場所に限定されることなく利用できる支援ではないという点において，課題が残る．加えて，長時間の入力作業は要約筆記者に過度な負担を与える可能性があるため，複数のグループによる連携体制を整備する必要がある．

図 2　要約筆記を行う様子
Figure 2　Work scene of summary writing

1.3　本研究の目的

本研究では，重度の視覚障害を伴う盲ろう者や盲ベースの盲ろう者など，点字による情報取得が可能な盲ろう者向けの自動要約筆記システムを提案する．本提案システムは自然言語処理モデルの 1 つである mT5 と UD トーク等を活用することで，専門知識と高度な技術をもつ要約筆記者が同席できない場合など，時間や場所に制限されることなく，点字を利用する利用者に最適化された要約筆記を提供することができる．

本提案システムの有効性や安定性を評価するために，本提案システムの出力内容と専門の要約筆記者が要約した出力内容等を比較し，ROUGE をはじめとする様々な指標を用いて評価する．

2.　提案システムの概要

本提案システムは，“発話内容の音声認識”，“要約文への変換”，そして，“要約文の出力”という 3 つのプロセスにより構成される．以下では，各プロセスについて詳細に述べた後，要約筆記の出力の際に重要となる発話者判別のためのタグの役割についても述べる．

2.1　発話内容の音声認識

音声認識を可能にする API には Google が提供している“Speech-to-Text（以下，Google-API）”や IBM が提供している“Watson-API”など，数多くの選択肢がある．本研究では，シャムロック・レコード株式会社が提供している聴覚障害者向けのコミュニケーション支援アプリケーションの 1 つである“UD トーク”を利用する．

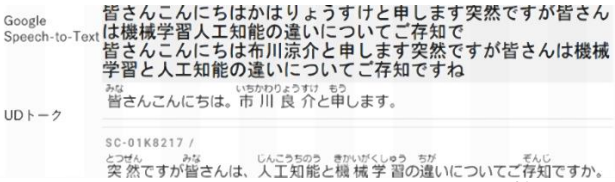

図 3　Google-API と UD トークの出力比較
Figure 3　Comparison of Google-API and UD-Talk output

図 3 は，同一マイクの環境下において，同様の発話内容を Google-API と UD トークを使用してテキスト化した結果をそれぞれ示している．

ⓒ2021 Information Processing Society of Japan

Google-API では，名前の認識や最終文章の終助詞に誤りがあることがわかる．また，句読点が打たれていないため，データの取り扱いが複雑になる点や音声認識結果が候補として複数出力されている点など，出力データの二次的利用の観点からも様々な課題が残る．一方，UD トークでは漢字の誤変換があるものの，それ以外は間違いがなく，正確にテキスト化されていることがわかる．なお，音声認識の誤変換については，本提案システムの最終的な出力形式が点字であるため，点訳の際に，全て平仮名表記となることから，同様の読み方をする漢字の場合には，誤変換が行われたとしても問題になる可能性は低い．そして，UD トークは，マルチプラットフォームの環境に対応しており，スマートフォンの OS（Android，iOS）に依存せずテキスト化を行うことが可能である．時間や場所に限定されることなく，点字を活用する盲ろう者に最適化した要約筆記を提供するために，本提案システムでは，UD トークを利用することが最良であるという結論に至った．

2.2 要約文への変換

要約文への変換は，形態素分析などを活用することで可能である．しかし，日本語の係り受けの複雑さや多様さなどから，想定する条件が膨大となることに加え，想定外の事象に対応できない．よって，利用用途が多岐にわたる本提案システムでは，形態素分析を活用することが困難である．

そこで，Google が 2020 年に発表した自然言語処理を実現可能とする機械学習モデル T5（Text-to-Text Transfer Transformer)を基盤に，101 の国地域で話されている言語に対応するために開発提案された事前学習済みモデル mT5（multilingual T5）を用いる[4]． mT5 は事前学習済みモデルとして Wikipedia などの大量の情報を学習した後，転移学習を行うことで，優れた精度を実現したモデルである．転移学習を行う mT5 は，その他のモデルと比較して最低限必要なデータセットの数が少量で済むという特徴がある．そのため，点字を利用する盲ろう者向け要約筆記など，専門性が高く，大量のデータセットを用意することが難しい場合には有効な手法である．

本提案システムでは，Google から公開されている事前学習済みモデルを使用し，点字を利用する盲ろう者向け要約筆記の学習データにてファインチューニングを行う．これにより，発話の内容から情報を圧縮し，点字を利用する盲ろう者向けの出力内容を生成する．一連の処理を経ることで，話し言葉から点字を利用する盲ろう者向けの要約内容，すなわち，読み言葉への変換を実現する．

2.3 要約文の出力と伝達

mT5 において要約された内容を利用者へ出力する．点字を利用する盲ろう者が普段から利活用しているデバイスやソフトウェアは個人により大きく異なる．そして，些細な環境の変化により，順応するまでに多大な時間と労力が必要になる場合も考えられる．そのため，本提案システムでは，要約内容を利用者の元へと出力するデバイスを唯一に限定せず，利用者が普段から利活用している環境に応じて，出力先を自由に変更することができる仕組みを実装する．具体的には，要約文の出力を汎用性の高いテキスト形式に統一することで，コードベースの出力先変更を可能にする．この仕組みにより，利用者は限りなく普段の環境に近い状態で要約文を確認することが可能となり，本提案システムのユーザビリティ向上につながると考えられる

出力方法の一例として，“LINE Messaging API（以下，LINE-API)”を利用した例を以下に示す．図 4 の例では，LINE-API を用いて BOT を作成し，要約筆記の内容をメッセージとして利用者へ伝達する．

図 4　LINE BOT を活用した出力

Figure 4　Output using LINE BOT

要約筆記では，提示された要約の内容が誰からの情報なのかを利用者へ明確に示す必要がある．そのため，“発話者/要約内容”という出力形式が広く利用されており，本提案システムにおいても同様の出力形式を採用する．

発話者をタグとして登録し，管理できる仕組みを実装することで，一度，発話者として登録した後は，再び変更がない限り同じ発話者タグが出力内容の先頭に追加される．そして，発話者タグを変更する際は，口頭で“発話者〇〇”と発言する．その際，発話者タグの誤変換を防止するために，形態素分析エンジンである“MeCab”を用いる．発言内容が人名であるかどうかを判断し，発話者タグの変更を行う（図 5 参照）．しかし，“東”や“林”などの名字は一般名詞と判断されることもあるため，例外リストを用意する必要がある．

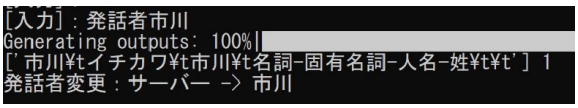

図 5　発話者切り替え時のログ出力画面

Figure 5　Log output screen when the speaker is switched

2.4 概念モデル

発話内容を UD トークの音声認識機能を利用してテキスト化を行う．その後，テキスト化された発話内容を取得するために UD トークが限定公開する専用 Web サイトから Python を使用してスクレイピングを行う．UD トークにて

ⓒ2021 Information Processing Society of Japan

公開された専用 Web サイトは，非同期の処理が行われており，通常のスクレイピング処理では，情報を取得することができないため，動的 Web サイトから情報を取得することが可能な Python ライブラリの selenium を用いて発話内容を逐次取得する．

点字を利用する盲ろう者向けの要約筆記専用に学習したファインチューニング済みモデルを本提案システムの要約エンジンとして構築する．UD トークの専用 Web サイトから取得した発話内容を構築した要約エンジンに受け渡し，要約文の生成を行う．要約エンジンの中核を担う mT5 はファインチューニング済みモデルのメモリ展開と解放，デコードを同一セッションにおいて処理するため，1 入力の処理に 1 分程度必要である．そこで，ファインチューニング済みモデルを PyTorch にコンパイルし，Simple-Transformers にマウントする手法を用いることで，一連の処理を 5 秒から 10 秒程度まで短縮する手法を採用した．

その後，利用者が普段から使い慣れているチャットシステムなどの環境に応じて，出力を行う．利用者は携帯用点字端末などで要約内容を確認することになる．

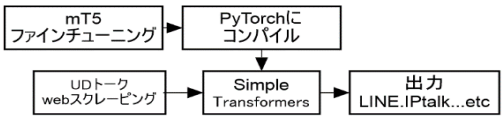

図 6　概念モデル

Figure 6　Conceptual model

3. データセットの構築

要約エンジンの構築に必要なデータセットは，国立国語研究所と情報通信研究機構（旧通信総合研究所），東京工業大学が共同開発を行い，提供している「日本語話し言葉コーパス（CSJ：Corpus of Spontaneous Japanese)」を原文として使用する．データセットの形式は Input を話し言葉，Target を読み言葉として設定する．学習データ，テストデータ，検証データをそれぞれ 8:1:1 に分割して，合計約 2,000 件のデータセットを作成する．また，1,000Step ごとに Check Point を設け，各層の重みを保存しながら，累計 5,000Step の学習を行う．

データセットの作成に当たっては，Input に CSJ 原文を使用し，Target には原文を要約したものを用いる．要約作業については，本学で聴覚障害者や盲ろう者を対象に要約筆記を行う特定非営利活動法人“PCY298”が発行しているマニュアル[6]に準拠する．本マニュアルには，“えー”などの無機能語の削除，念押し・挿入句・節（常套句）等の冗長部分を削除することや主語述語を明確にすることなどが細かく定義されている．

4. 評価指標

要約筆記内容をもとにファインチューニングを実施した学習モデル（以下，ファインチューニング済みモデル）を定量的に評価するために，Chin-Yew Lin 氏らがテキスト要約の評価指標として提唱した ROUGE-N 並びに，ROUGE-L[7]を用いる．Recall の式を以下に示す．

$$ROUGE_N = \frac{\sum_{S \in refs} \sum_{gram_n \in S} count_{match}(gram_n)}{\sum_{S \in refs} \sum_{gram_n \in S} count(gram_n)}$$

$$ROUGE_L = \frac{LSC(summary_{words}, refs_{words})}{refs_{words}}$$

ROUGE-N は n-gram を単位として示している．本研究では，原文と要約文の間で一致する割合を算出している．すなわち，ROUGE-1 は，uni-gram（1 単語）における一致度を算出しており，ROUGE-2 は bi-gram（2 単語）における一致度を算出している．また，ROUGE-L は一致する最大シーケンス（Longest Common Subsequence：LCS）の割合によってスコアを算出している．

4.1 評価指標の単位

ROUGE において使用する n-gram は，評価対象となる原文と要約文の文章について，MeCab を利用することで短単位に分割し評価を行う．短単位に分割する際に使用する辞書は，MeCab に標準搭載されている IPA 辞書を使用する．そして，データセットに利用している CSJ は，国立国語研究所が規定している UniDic を利用することで評価対象となる文章を短単位に分割していることになる．つまり，両者は異なる辞書により分割されており，短単位の単位が異なる．しかし，mT5 への Input は短単位ではなく，1 文や複数文章となるため，両者の辞書の違いによる単位の相違は問題にはならない．一般的に短単位を使用する利点として，単位が文脈から乖離し過ぎない点が挙げられる．点字を利用する盲ろう者への要約筆記では，主語述語などの語順も重要視されるため，文脈や語順を一定水準において，評価できる短単位が最良であるという結論に至った．図 7 には，最小単位から文節までの分割した例を示している．

文節	国立国語研究所の							
	↓ 文節を自立部と付属部に分けることで認定（トップダウン）							
長単位	国立国語研究所							の
	← 長単位を越える短単位は認定しない →							
短単位	国立		国語		研究		所	の
	↑ 最小単位の結合により認定（ボトムアップ）							
最小単位	国	立	国	語	研	究	所	の

図 7　最小単位から文節までの分割例　[8]より引用

Figure 7　Example of clause type by partition

4.2 日本語 ROUGE の困難さ

ROUGE の開発段階では，英文によるスコアを算出することが想定されていたため，日本語特有の句読点，特に読点によってスコアが大きく低下することがある．以下に，読点の有無における，ROUGE の差異を例として示す．なお，下線は ROUGE におけるマッチング箇所を示している．

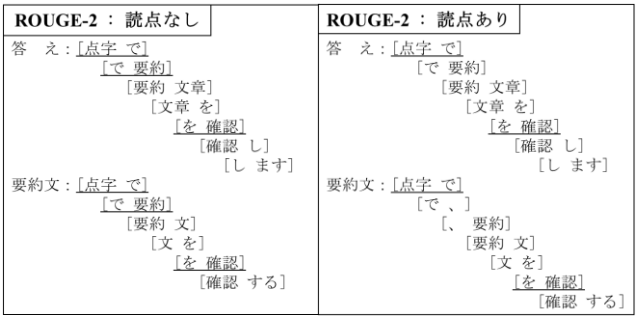

図 8　読点の有無における ROUGE-2 の例

Figure 8　Example of ROUGE-2 in the existence or not existence of comma

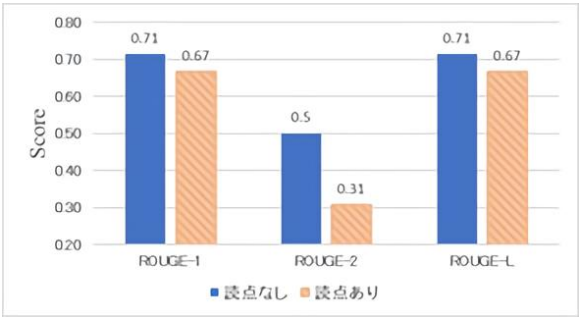

図 9　読点の有無による ROUGE のスコア

Figure 9　ROUGE score for the existence or not existence of comma

図 9 の ROUGE-2 に着目すると，“読点なし”のスコアが 0.5 であったのに対して，”読点あり”のスコアは 0.31 となり，大幅にスコアが減少していることがわかる．ROUGE-1 及び ROUGE-L においても同様に，減少していることがわかる．一般的に句読点を除外することによって，精度の向上や安定した評価が期待できるため，日本語 ROUGE の場合には句読点を除外することもある．しかし，点字を利用する盲ろう者への要約筆記を行う場合には，句読点の存在が点字への変換精度向上に影響を及ぼすことがある．よって，本提案システムでは，句読点を考慮した ROUGE を評価指標として算出している．

5. 結果

本提案システムを定量的に評価するために，“システムの精度”と“出力の有用性”，そして，“要約筆記者の出力内容との差異”という 3 つの評価項目を設定した．システムの精度では，mT5 に存在する各事前学習済みモデルの学習時の Loss 値と ROUGE を用いて，スコアを比較した．出力の有用性では，提案システムの情報の圧縮率，つまり，文字の削除率を算出し，既存の字幕規則に従った場合の圧縮率と比較した．要約筆記者の出力内容との差異では，提案システムの出力内容と要約筆記者の出力内容を比較した．なお，使用した要約筆記者の出力内容は，本学で盲ろう者などに要約筆記を行っている PCY298 から提供を受けたものである．

5.1 ファインチューニングモデルの Loss 値と ROUGE スコア

5,000Step 学習した mT5 を 1,000Step ごとに設定した Check Point の ROUGE スコアにより精度を評価する．mT5 には複数のパラメータでの，事前学習済みモデルが存在する（表 1 参照）．そのため，同じ学習データを用いて，各事前学習済みモデルのファインチューニングを行い，どのモデルが本研究に最適であるかを検討した．なお，XXL モデルについては，実行環境のスペック不足により検証することができなかった．

表 1　mT5 の各モデルとパラメータ数

Table4　Each model of mT5 and its number of parameters

Model	Parameter（千万）
Small	30
Base	58
Large	120
XL	370
XXL	1,300

表 2　ROUGE スコア最大値と Step 数

Table 2　Maximum value and number of steps for ROUGE

	Model	Step	Loss	ROUGE-1	ROUGE-2	ROUGE-L
	Small	3,000	0.00077	0.68161	0.55956	0.67935
	Base	1,000	0.00115	0.69210	0.56457	0.69084
	Large	1,000	0.00055	**0.70026**	0.58576	0.69900
★	XL	2,000	0.00020	**0.70770**	**0.59493**	**0.70612**

表 3　CheckPoint ごとの Loss 値

Table3　Loss value evaluated at Check Point

	Model	1000 Step	2000 Step	3000 Step	4000 Step	5000 Step
	Small	0.00169	0.00106	0.00077	0.00054	0.00038
	Base	0.00115	0.00037	0.00030	0.00021	0.00011
	Large	0.00055	0.00025	0.00013	0.00010	**0.00006**
★	XL	0.00020	0.00009	**0.00005**	**0.00002**	**0.00005**

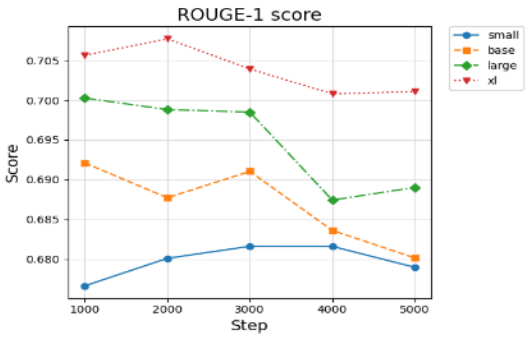

図 10　各 Step 数の ROUGE-1 スコア

Figure 10　ROUGE-1 score for each number of steps

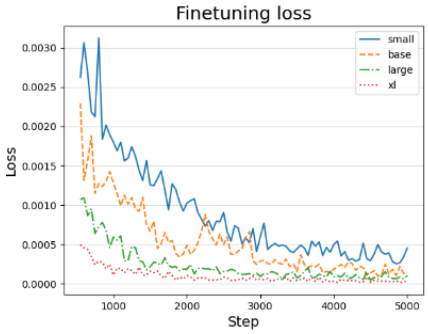

図 11　ファインチューニングの際の Loss 値
Figure 11　Loss value when fine tuning is applied

5.2 本提案システムと既存要約規則における削除率と出力の比較

作成した評価データをもとに，提案システムの文字削除率を一文字単位で算出した．原文をそのまま出力すると文字削除率はゼロを示す．数値が高いほど，原文から文字が削除され出力文字数が減少していることになる．図 12 には，本提案システムの削除率と“TED 日本語字幕（以下，TED 規則）”のマニュアルに則り要約された文章の削除率をそれぞれ示している．使用しているファインチューニング済みモデルは，ROUGE スコアが最も高かったことから，XL の 2,000Step 目を採用した．

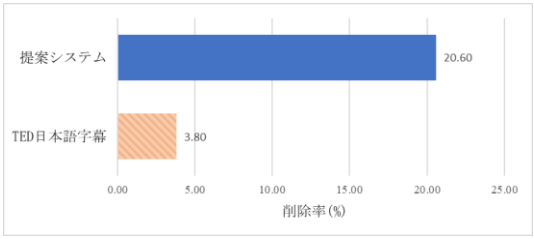

図 12　提案システムと TED 規則適用時の文字削除率
Figure 12　Character deletion rate of proposed system and its value when the TED rule is applied

続いて，原文と本提案システムの出力の内容をもとに ROUGE スコアを算出した．原文と要約文にどの程度，文章的な差異があり，原文の情報がどの程度，損失しているかを評価した．スコアが高いほど，要約文の中に原文の情報が保持されていることを示している．比較として図 13 に，TED 規則に基づく要約文と原文のスコアを示す．

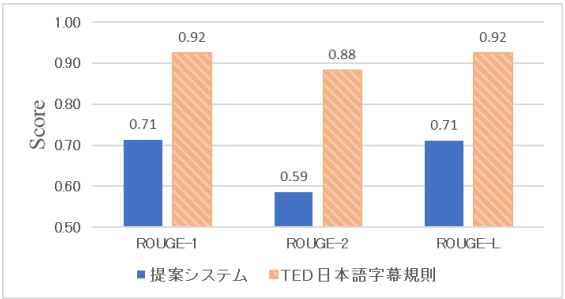

図 13　原文と各要約文の ROUGE スコア
Figure 13　ROUGE scores for the original text and each summary text

5.3 本提案システムと通常要約筆記における削除率と出力の比較

同じ原文をもとに，本提案システムの出力文章と専門の要約筆記者が要約した文章を比較した．原文には，本学の講義において教員や学生が実際に発話した内容を無機能語が含めテキスト化を行った．そして，人の癖による発話内容の偏りを防ぐため，4 人分の発話内容を格納した．図 14・15 は，ROUGE スコアと削除率を，前述にて作成した原文と本提案システム及び通常要約筆記から算出する．

図 16 は，通常要約筆記の出力と本提案システムの出力との ROUGE スコアを算出したものである．スコアが高いほど，両者の出力に含まれる情報量が等しいといえる．

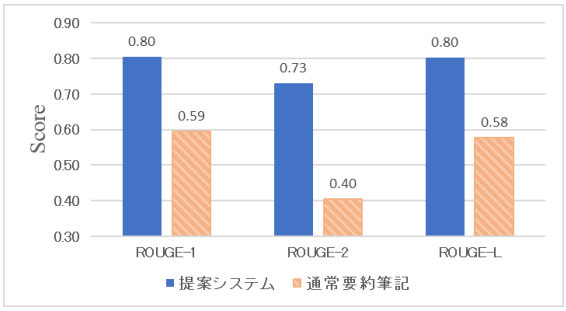

図 14　原文と各要約文の ROUGE スコア
Figure 14　ROUGE scores for the original text and each summary text

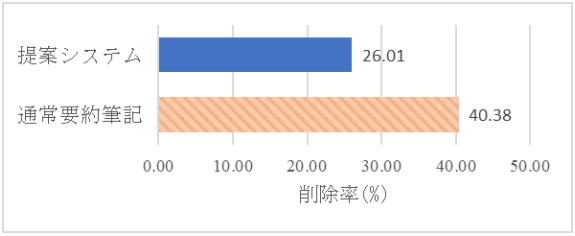

図 15　提案システムと通常要約筆記の文字削除率
Figure 15　Character deletion rate of proposed system and general writing summary

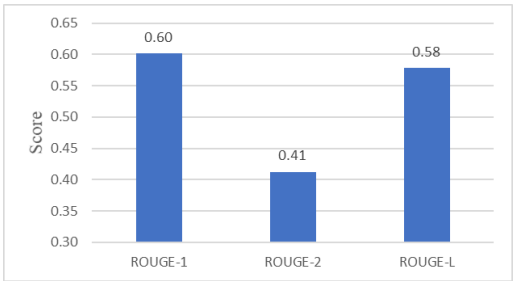

図 16　提案システムと通常要約文の ROUGE スコア
Figure 16　ROUGE scores of the proposed system and general writing summary

6. 考察

6.1 ファインチューニングモデルの Loss 値と ROUGE スコア

図 11 に示した 5,000Step の学習における Loss 最小値は，XL の 4,842Step 目の 1.05E-5 であった．また，ROUGE の

各指標での最大スコアは 2,000Step 目の XL であり，ROUGE-1 においては 0.707 と非常に高いスコアとなった．表 2 並びに図 10 から，ROUGE のスコアはパラメータ数に大きく依存していることがわかる．これは，参考文献[5]で述べられている点と合致する．しかし，Large と XL のスコア差は 0.01～0.001 と非常に小さく，パラメータ数の増大と計算コストを考えた際に効率が良い方は Large であると考察する．

図 10 では，4,000Step 以降の ROUGE スコアが低下傾向を示している．また，表 2 は 3,000Step から 4,000Step にかけて Loss 値が総じて減少していることを示している．図以上の 2 点のことから，ROUGE スコアの減少傾向は過学習によるものと推察する．

4 章で述べた通り日本語 ROUGE は句読点によっても，スコアが大きく変動することがある．そのため，0.6 から 0.7 というスコアは，要約筆記者が要約した文章に非常に近いと判断できる．

6.2 本提案システムと既存要約規則における削除率と出力の比較

図 12 の文字削除率では，TED 規則と比べ 5 倍以上の削除率であることがわかる．つまり，本提案システムは原文に対して 2 割の冗長部分を削除し，8 割程度に圧縮した上で要約内容を利用者に伝えていることがわかる．

しかし，図 13 の原文との比較では，TED 規則の ROUGE スコアが非常に高く，提案システムのスコアが TED 規則と比較し平均 0.24 下回っている．つまり，TED 規則の場合，原文の情報を十全に近い形で保持しているといえる．

一方，本提案システムでは，情報の圧縮の過程で原文の冗長部分を省略しているため，ROUGE-1 におけるスコアが，0.72 に留まった．

6.3 本提案システムと通常要約筆記における削除率と出力の比較

図 14 では本学講義から作成した原文を用いて，本提案システムと通常要約筆記の 2 種類の出力を ROUGE にて比較した．その結果，本提案システムの ROUGE-1 スコアでは， 0.8 と原文の情報がある程度保持されていることがわかる．しかし，通常要約筆記の ROUGE-1 スコアでは，0.59 と低く本提案システムのスコアと乖離が見られた．これは，要約筆記を行う際に即時性が求められるため，長い単語や文章を短く入力しやすい単語などに，置き換え要約を行っているためであると考察する．講義をもとに作成し 5.3 で使用したデータでは，“育むことを”という原文を“教育することを”と要約されていた．選ぶ単語は要約筆記者の判断によるため，一意ではない．そのため，単語自体を置き換えた通常要約筆記と原文の単語をある程度流用している本提案システムの 2 者間では ROUGE のスコアは 0.6 程度にとどまっている（図 16 参照）．次に図 15 の削除率に着目する．本提案システムでは，26%程度にとどまったものの，通常要約筆記では，40%程度の文章を省略していることがわかる．2 者間の削除率の違いも前述した単語の置き換えによるものであると考察する．

7. 提案システムの改良

本提案システムを試験的に使用したところ，発話者タグの変え忘れが発生した．また，会話中に発話者タグを変更するというのは，発話者に対し一定の負担をかけていることが分かった．

そこで，現在手動で変更している発話者タグを，自動で変更できるよう提案システムを改良する．具体的には，メル周波数ケプストラム係数（MFCC）並びにサポートベクターマシン（SVM）を用いて，本提案システムの利用頻度が高い人物の肉声の特徴量を抽出し学習する．前述にて学習したモデルを提案システムの発話者判別機として構成する．これにより，学習した人物に限定されるものの，発話者タグを自動で変更することが可能である．

発話者タグの自動変更の実装にあたり，マイク入力可能なパソコンを 1 台用意する．そのため，本提案システムを使用する場合，利用者の付近にパソコンとスマートフォンの 2 台の端末を置く必要がある．UD トークが音声認識を開始した際に，要約エンジンが動作している Server から TCP 通信を用いて，利用者の付近に置かれているパソコンにフラグを送る．フラグを受信した後，パソコンのマイクを使用し 2 秒間音声を取得し，発話者の判別を行う．その後，特定した人物名を，発話者タグとして TCP 通信を用いて，パソコンから Server 側に送信する．最終的に，Server 側で，要約エンジンの出力と受信した発話者タグの 2 つを合成し，LINE などの媒体へ出力を行う．

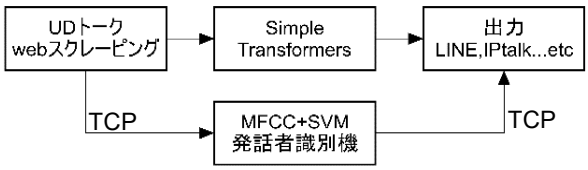

図 17　改良後の概念モデル

Figure 17　Improved conceptual model

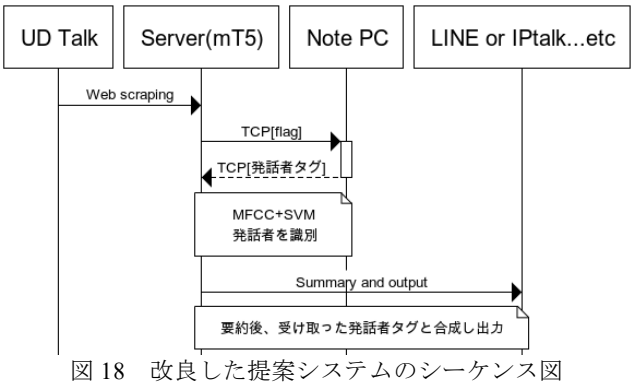

図 18　改良した提案システムのシーケンス図

Figure 18　Sequence diagram of the improved proposed system

8. 発話者識別機の実装と評価

7 章で述べた発話者識別機を MFCC と SVM を用いて実装し精度を評価する．一般的に音声識別として用いられる MFCC の次元は，直行成分である 1 次元目を除く 13 次元までである[9]．そのため，本提案システムにおいても自前述した次元数で実装を行う．SVM は，MFCC で取得した特徴量を分類するために使用する．

学習データは発話者が同一のパソコンマイクを用いて肉声を録音したデータを 2 秒ごとに分割し使用する．前述したデータ形式のものを，1 人あたり 100 件程度 3 名分用意した．データセットの音声を 2 秒ごとに分割する理由は，本提案システムの実使用場面で発話者判別機に渡される音声ファイルが 2 秒であるためである．つまり，実使用場面に近いデータを想定し学習を行う．用意したデータを，学習データ，評価データにそれぞれ 8:2 に分割する．その際，3 人のデータがそれぞれ均等になるよう調整しデータセットを作成する．

8.1 結果と考察

学習した SVM の精度を評価するため，8 章で作成した評価データを使用する．

表 4　SVM の精度
Table4　Accuracy of SVM

	利用者 A	利用者 B	利用者 C	全体
精度	0. 88	0. 88	0. 94	0. 90

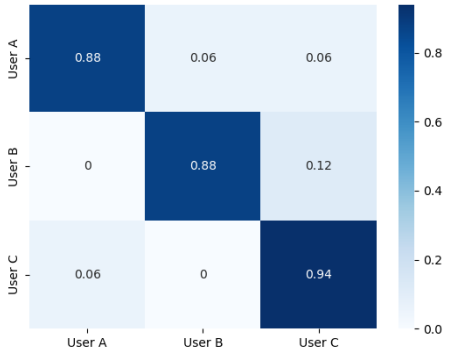

図 19　SVM の精度[詳細]

Figure 19　Accuracy of SVM [detail]

表 4 並びに図 19 では，発話者判別機として学習した SVM の評価を行った結果を示した．3 名発話者別の精度では，0.88 から 0.94 と高いスコアを示した．つまり，2 秒間という短い時間での音声においても，ある程度の判別が可能であるといえる．

このことから，合計 300 件程度のデータセットから学習した MFCC と SVM を用いることで，発話者判別及び発話者タグの自動変更が可能である．

9. おわりに

本研究では，重度の視覚障害を伴う盲ろう者や盲ベースの盲ろう者など，点字による情報取得が可能な盲ろう者向けの自動要約筆記システムを提案した．システムの精度と出力の評価，出力の有用性では，学習時に過学習と思われる現象が観測されたものの，一定の有用性が見られ，実用に足る精度を示した．このことから，mT5 を使用することによって，点字を利用する盲ろう者向け自動要約筆記の実現が可能であるといえる．

しかし，通常要約筆記との比較では，原文の文字削除率に 14%程度差があり，本提案システムに改善の余地があることが判明した．これは，要約筆記者が行う単語の置き換えによる差であり，現在，本提案システムでは再現しきれていない点である．使用するデータセットなどを更に要約筆記者の出力に近づけることで，前述問題の改善を今後の課題とする．

MFCC を用いた発話者タグの自動変更では，一定の精度を示し実現可能であると判断できた．しかし，MFCC の歴史は古く，2021 年現在では，MFCC と比べ精度の高いモデルは数多く存在する．今後，高い精度を示しているモデルを使用し，要約筆記に重要とされる発話者タグの誤変更を減らしていきたい．

謝辞　本研究では UD トーク開発者である青木秀仁氏からアプリケーションの提供を受けた．深く謝意を表する．

参考文献

[1] NPO 法人 東京盲ろう者友の会, http://www.tokyo-db.or.jp/?page_id=106 (参照 2021-11-11).
[2] 牟田口. “点字読み熟達者の読速度に関する研究”, 特殊教育学研究, 2012, vol. 50, no. 4, p. 343-352.
[3] 小林, 川嶋. “日本語文章の読み速度の個人差をもたらす眼球運動”, 映像情報メディア学会誌, 2018, vol. 72, no. 10, p. J154-J159.
[4] Linting Xue, et al.. “mT5: A massively multilingual pre-trained text-to-text transformer”, 2021, Proc. Conf. North Amer. Chapter of the Association for Computational Linguistics: Human Language Technologies, p.483–498.
[5] Colin Raffel, et al.. “Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer”, J. Machine Learning Research, 2020, vol. 21, p.1-67.
[6] NPO 法人 PCY298. “PC 通訳基礎知識”, 2020 年版.
[7] Chin-Yew Lin. “ROUGE: A Package for Automatic Evaluation of Summaries”, Proc. Workshop on Text Summarization Branches Out, 2004, p. 25-26.
[8] UniDic 国語研短単位自動解析用辞書, https://ccd.ninjal.ac.jp/unidic/glossary (参照 2021-11-11)
[9] 中尾, 松本. “MFCC を用いた Support Vector Machine によるテキスト独立型話者認識”, 電子情報通信学会, 情報・システムソサイエティ特別企画, 2015, p. 35.

ⓒ2021 Information Processing Society of Japan