

# 発話文口調変換タスクにおける 教師なしテキストスタイル変換手法の比較検討

## Comparison Study of Performance in Dialect Translation Task Using Unsupervised Text Style Transfer Methods

大原広嗣<sup>1\*</sup> 森辰則<sup>1</sup>  
Hirotsugu Ohara<sup>1</sup> Tatsunori Mori<sup>1</sup>

<sup>1</sup> 横浜国立大学大学院環境情報学府

<sup>1</sup> Graduate School of Environment and Information Sciences, Yokohama National University

**Abstract:** Appropriate control of the response dialect of a chatbot system can help the system achieve more natural and friendly conversations. We focused on unsupervised text style transfer and examined several methods, including a supervised method, for generating dialect translation models from text data with a specific dialect by deep learning, comparing the kind and amount of data used. Experimental results showed that the method of generating pseudo parallel data using back-translation showed high performance comparable to the supervised method, and Few-shot learning using ChatGPT is the most effective when the amount of data is limited. In addition, we showed that the Weighted Decoding approach proposed for conditional sentence generation can be applied to style conversion by combining it with a Denoising Autoencoder.

### 1 はじめに

対話システムは様々な話者の発話を含む大規模データで学習されることが多く、生成口調の一貫性に欠ける、または「です・ます」のような単調な口調に限定されるといった問題が生じることがある。自然な対話や親しみやすさを促進するため、対話システムの応答を意図した口調に制御する手法が重要となっている。

口調や個性など特定のスタイルで対話システムの応答を制御することを目的とした研究として、スタイル固有の対話データを用いてシステムを転移学習する手法 [1] [2] や、システム生成文のスタイルを変換するためのスタイル変換モデルを学習する手法 [3] [4] などが提案されている。しかし、前者の手法ではシステムの生成内容がスタイル固有のデータに影響される課題や、システムに使用される言語モデルが大規模化しており学習コストが高いといった課題が存在する。そのため本研究では後者の手法に焦点を当てた。

後者の手法はテキストスタイル変換と呼ばれ、テキストの意味内容を保ちつつ特定のスタイルに書き換えることを目的としている。スタイル変換において変換前後の対話データが利用可能な場合、教師ありの学習手法で変換モデルを生成することができるが、実課題におい

て対話データが利用できる場面は限られるため、目的のスタイルを持ったスタイル文のみを用いて変換モデルを学習する教師なしのスタイル変換手法が研究されている。教師なしテキストスタイル変換は、感情変換や形式変換といった分野で深層学習手法が盛んに研究されているが、日本語特有の課題である口調を扱った先行研究は少ない。また、口調を扱った先行研究の課題として、多くの手法で (i) 独自の非公開データセットを用いており、(ii) 主観的な評価を行っているため、手法ごとの相対的な比較が難しい現状がある。

以上を踏まえ、本研究では複数の教師なしスタイル変換手法について、利用できるデータ量や外部資源の利用などの条件を比較して、性能を定量的かつ相対的に検討する。そのために、一般に再現可能な対話の口調スタイルデータセットを作成し、教師あり手法をベースラインとして各手法の口調変換に対する性能の調査を行う。

### 2 関連研究

#### 2.1 教師なしテキストスタイル変換

多くの先行研究で、GAN や VAE などをベースとしたモデルを構築し、文の内容情報とスタイル情報を分離する手法が取られており、先行研究 [3] [4] も分離手法を用いている。しかし、これらの手法は固定サイズ

\*連絡先：横浜国立大学大学院環境情報学府  
〒 240-8501 神奈川県横浜市保土ヶ谷区常盤台 79-1  
E-mail: ridivarg@outlook.com

の潜在空間を前提とするため潜在表現が可変長である Transformer モデルと相性が悪く、さらに複数種類のスタイルデータの利用を前提としている。これらの手法の中でも、Prabhumoye ら [5] は“逆翻訳には文の意味を保ちつつスタイルを弱める働きがある”という特性に着目し、スタイル文を中間言語に翻訳後、元言語に逆翻訳した際の潜在表現を内容情報とみなし、内容情報から元のスタイル文を復元するデコーダを敵対的に学習する手法を提案している。

## 2.2 Weighted Decoding

Weighted Decoding(以下 WD と呼ぶ) は条件付き文生成で提案されたアプローチで、ある属性  $a$  で条件付けられた文  $X$  の生成  $P(X|a)$  を目的としている。WD では、ベイズの定理  $P(X|a) \propto P(X)P(a|X)$  により、 $P(X|a)$  が  $P(X)$  と  $P(a|X)$  に分解できることを利用する。Liu ら [6] は、望ましい属性  $a$  のデータで学習した言語モデル  $M^+$  と望ましくない属性  $\bar{a}$  のデータで学習した言語モデル  $M^-$  を導入し、2つの言語モデル  $M^+$  と  $M^-$  の対照性を利用して擬似的な  $P(a|X)$  を計算している。

## 2.3 本研究の位置づけ

上述のように、口調に関するスタイル変換研究は少なくどのような手法が有効であるかは分かっていない。また、特定の口調を持つスタイル文は収集が難しく、少資源であるという課題が存在する。そのため、本研究では利用可能なデータに基づいて様々な状況を設定し、各々の状況設定において口調変換に有効なスタイル変換手法の検討を行う。特に、単一のスタイルデータで学習可能な手法に焦点を当て、教師あり手法、逆翻訳による擬似対訳データ生成手法、WD をスタイル変換に応用する手法、ChatGPT<sup>1</sup>を用いた Few-shot 学習によるスタイル変換手法について比較検討を行う。

# 3 口調スタイル変換の定義とデータセット作成

## 3.1 口調スタイル変換の定義

本研究では口調スタイルを“特定の発話者を想起する特徴的な言葉遣い”と定義し、方言や役割語などを対象とする。また、一般的な話し言葉をノーマルスタイルとみなし、口調スタイル変換を“ノーマルスタイル文から特定の口調スタイル文への書き換え”とする。以降、変換前のノーマルスタイル文を原文、変換後の文を口調スタイル文、変換前後のペアの集合を対訳データと呼ぶ。

<sup>1</sup><https://openai.com/chatgpt>

## 3.2 口調スタイルデータセットの作成

日本語の口調制御に関する先行研究では、独自の非公開データセットを使用しているものが多い。本研究では公開データセットと口調変換ツールを活用し、一般に再現可能なデータセットを新たに作成した。

### 3.2.1 作成方法

特定の口調を持った文を作成するため、Web 上で公開されている口調変換ツール<sup>2 3 4</sup>を利用した。ツールに入力する原文データには JPersonaChat データセット [2] を用いた。これは話者が自身の特徴を表すプロフィール文を設定し、プロフィールに基づいたワーカー同士の会話を収集したデータセットである。多様な口調の文が含まれており、入力の原文に対してロバストな口調スタイル変換器の検討を行うことができると考える。

JPersonaChat の各文に正規化と分割の前処理を行い約 10 万文を得た。この原文をツールで変換し、口調スタイル文として「大阪弁」、「博多弁」、「お嬢様口調」、「廓口調」を作成した。以降、原文と口調スタイル文からなるデータセットを「口調スタイルデータセット」と呼ぶ。表 1 に「良い天気なので、わたしは買い物に行きます。」という原文に対する口調スタイル文を示す。

表 1: 例文に対して作成された各口調スタイル文。

	例文
原文	良い天気なので、わたしは買い物に行きます。
大阪弁	ええ天気やさかい、うちは買い物に行く。
博多弁	良か天気やけん、うちは買い物に行く。
お嬢様口調	良いお天気なので、わたくしは買い物に行きますわ。
廓口調	良い天気なので、わっちは買い物に行きんす。

### 3.2.2 口調スタイルデータセットの分析

生成した口調スタイル文の中には、ツールで変換されなかった文が一定数含まれていた。ツールで変換できた文を「変換文」、変換されなかった文を「無変換文」と呼ぶ。無変換文は (i) 変換せずとも目的の口調スタイルであるか、(ii) 変換すべきであるがツールの変換規則に当てはまらず変換できなかったか、のいずれかが原因だと考えられる。また、変換文の中には人目で見ても不自然な文が一定数含まれており、例えば「忙しいですが」という文をお嬢様口調に変換すると「忙しいですわが」という文が生成された。しかし、本研究では変換ツールが生成した文を目的の口調スタイルだとみなした。

<sup>2</sup><https://ojosama.jiro4989.com>

<sup>3</sup>[https://github.com/anmonite/expression\\_trans](https://github.com/anmonite/expression_trans)

<sup>4</sup><https://www.8toch.net/translate>

### 3.2.3 実験に用いるデータ

各口調スタイル変換手法に用いるデータとして、変換文 50000 文を学習データ、5000 文を評価データ、1000 文をテストデータとした。ただし、各手法がツールよりもロバストに変換できるかを検証するため、無変換文 1000 文をテスト用データに加え計 2000 文とした。

## 4 比較検討する口調変換手法

### 4.1 手法ごとの条件比較について

各手法の説明を行う前に、手法ごとの条件の違いを表 2 にまとめる。ここで、対訳データは口調スタイルデータセット内の対訳データを利用するかを表し、 $\Delta$ は手法内で擬似的な対訳データを生成することを示す。言語モデル数は各手法で学習する言語モデルの数を表している。また、WD 手法では翻訳器による逆翻訳データを利用し言語モデルを 3 つ学習する実験と、逆翻訳データを利用せず言語モデルを 2 つ学習する実験をそれぞれ試行した。詳細は節 4.4.3 の推論において説明する。

### 4.2 教師あり手法

以降で説明する教師なし手法に対するベースラインとして、対訳データを用いた教師あり学習によるスタイル変換を試行する。原文を入力として対訳の口調スタイル文が出力となるよう翻訳タスク形式で言語モデルを学習し、口調スタイル変換器を生成する。

節 5 の評価実験では、口調スタイルデータセットの対訳データを用いて T5 モデル [7] のファインチューニングを行った。学習データ数 {100, 500, 1000, 5000, 10000, 50000} でそれぞれ実験を行い、評価データ数は学習データ数の 10%、テストデータは同じものを用いた。

### 4.3 逆翻訳手法

逆翻訳手法では Prabhunoye ら [5] の研究に倣い、逆翻訳を用いて口調スタイル文からスタイル除去を試みる。口調スタイル文は方言や役割語を含む特性上、通常の翻訳器では適切に翻訳できない恐れがあるため、ロバストな翻訳器として Web 上で公開されている高性能な翻訳器を利用する。こうした翻訳器では潜在表現を利用できないケースがあるため、潜在表現を利用する先行研究とは異なり、逆翻訳で得られる翻訳文をノーマルスタイル文とみなして擬似的な対訳データの生成を行う。

評価実験では、口調スタイル文に逆翻訳を行うことで生成した疑似対訳データを用いて教師あり手法と同様の学習を行った。翻訳器の性能によるスタイル変換の

精度の違いを調べるため、逆翻訳に用いる翻訳器として DeepL 翻訳<sup>5</sup> と Hugging Face で公開されている NLLB 翻訳モデル<sup>6</sup> [8]、OPUS-MT 翻訳モデル [9] を利用した。また、逆翻訳の中間言語として英語を選択した。

### 4.4 Weighted Decoding 手法

WD 手法では、条件付き文生成で提案された Liu らの WD アプローチをスタイル変換に応用する手法を提案する。まず WD アプローチについて説明し、その後スタイル変換への応用手法について説明を行う。

#### 4.4.1 Weighted Decoding アプローチ

一般的な自然言語を学習した言語モデル  $M$  は、入力文の単語列  $X = \{x_1, \dots, x_n\}$  が与えられたとき、確率  $P(X)$  を自己回帰的に式 (1) で計算する:

$$P(X) = \prod_{t=1}^n P(x_t | x_{<t}) = \prod_{t=1}^n \text{softmax}(z_t) \quad (1)$$

ここで、 $z_t$  は言語モデルで計算される  $t$  番目の単語の logit である。特定のスタイル  $a$  を条件として文生成を行う場合、Liu らはベイズの定理を用いて  $P(X|a)$  を  $P(X)P(a|X)$  に分解し、望ましいスタイル  $a$  の文をモデル化する言語モデル  $M^+$  と望ましくないスタイル  $\bar{a}$  の文をモデル化する言語モデル  $M^-$  の確率を組み合わせ、擬似的な  $P(a|x_{\leq t})$  を式 (2) で計算している:

$$P(a|x_{\leq t}) = \left( \frac{P^+(x_t|x_{<t})}{P^-(x_t|x_{<t})} \right) \quad (2)$$

さらに、 $P(a|x_{\leq i})$  を言語モデル  $M$  の出力と掛け合わせ、生成文の  $t$  番目の単語の確率は式 (3) で計算される:

$$\tilde{P}(x_t|x_{<t}) \propto P(x_t|x_{<t}) \left( \frac{P^+(x_t|x_{<t})}{P^-(x_t|x_{<t})} \right) \quad (3)$$

#### 4.4.2 スタイル変換への応用

スタイル変換では、入力文  $I$  の意味を保ちつつスタイル  $a$  で条件付けるために、 $P(X|a, I)$  をモデル化する必要がある。そこで言語モデル  $M$  の代わりにノイズ除去オートエンコーダ (以下 DAE と呼ぶ) を導入し、式 (3) を式 (4) に書き換える。

$$\tilde{P}(x_t|x_{<t}) \propto P(x_t|x_{<t}, I) \left( \frac{P^+(x_t|x_{<t})}{P^-(x_t|x_{<t})} \right) \quad (4)$$

式 (4) において、DAE で計算される確率  $P(x_t|x_{<t}, I)$  は入力文  $I$  と近い内容の文の出力を試み、式 (2) の項は DAE の出力確率をスタイル  $a$  に近づける操作を行うと推測される。

<sup>5</sup><https://www.deepl.com/translator>

<sup>6</sup><https://huggingface.co/facebook/nllb-200-distilled-600M>

表 2: 各手法ごとの比較項目.

比較項目	教師あり手法	逆翻訳手法	WD 手法	ChatGPT 手法
対訳データ	○	△	×	×
学習データ数	100~50000	100~50000	50000	10~100
外部資源の利用	×	翻訳器	翻訳器/×	ChatGPT
言語モデル数	1	1	3/2	0

#### 4.4.3 各モデルの学習および推論方法

**DAE** 口調スタイル文に対してノイズを付与し、ノイズが付与された口調スタイル文から元の口調スタイル文を復元するように口調スタイルごとに T5 モデルの学習を行った。入力文に加えるノイズとして、入力文の単語を一定確率で削除し(本研究では削除確率を 0.1 とした)、その後入力文の単語列を無作為に並び替えた。

**言語モデル  $M^+ \cdot M^-$**  先行研究に倣い、言語モデル  $M^+ \cdot M^-$  として GPT[10] ベースのモデルを利用する。しかし、式 (4) を計算する際に全ての言語モデルの語彙が統一されている必要があるため、GPT モデルの Tokenizer と単語埋め込み層  $W_{GPT} \in \mathbb{R}^{V_{GPT} \times 1024}$  の代わりに T5 モデルの Tokenizer と単語埋め込み層  $W_{T5} \in \mathbb{R}^{V_{T5} \times 2048}$  を利用した。ここで  $V$  は各モデルの語彙サイズを示す。また、T5 モデルと GPT モデルの単語埋め込み層の出力サイズが異なるため、線形層  $W \in \mathbb{R}^{2048 \times 1024}$  を追加した。言語モデル  $M^+$  として GPT モデルを各口調スタイル文でファインチューニングしたものを用い、言語モデル  $M^-$  には逆翻訳手法で生成した DeepL 翻訳による逆翻訳文でファインチューニングしたものを用いた。

**推論** 推論では、式 (4) の計算を対数上で行い、DAE の重み付けとして、出力 logit に対して temperature( $T$ ) を用いて式 (5) のように確率分布の操作を行った。

$$P(x_t) = \frac{\exp(z_{t,i}/T)}{\sum_{j=1}^{|V|} \exp(z_{t,j}/T)} \quad (5)$$

ここで、 $i$  と  $j$  は語彙  $V$  における  $i$ (または  $j$ ) 番目の単語を表す。 $T$  が 1 より小さいほど  $P(x_t)$  における各単語の確率の差が大きくなり DAE の重みが大きくなる。また、言語モデル  $M^+$  と  $M^-$  の操作によって入力文  $I$  の内容から遠い単語が選択されることを防ぐため、DAE の logit  $z_t$  に対して top  $k$  サンプリングを行い、確率上位  $k$  個以外の単語の確率を  $1 \times 10^{-100}$  とした。

さらに追加の設定として、言語モデル  $M^-$  を用いず DAE と言語モデル  $M^+$  のみを用いた実験を行った。2 つの実験を区別するため、 $M^+$  と  $M^-$  を用いる実験を「 $M^+ \cdot M^-$  実験」、追加実験を「 $M^+$  only 実験」と表記する。

#### 4.5 ChatGPT 手法

大規模言語モデルはプロンプトによりモデルの応答スタイルを制御することが可能であり、特に ChatGPT はその能力が高いことが知られている。そこで、ChatGPT を用いた Few-shot での口調変換を試行する。

評価実験では、OpenAI API の gpt-4-1106-preview モデルを利用した。プロンプトで例示する口調スタイル文は学習データからランダムに {10,50,100} 文抽出し、各 shot 数を試行した。この際、モデルの出力トークン上限が 4k となっており、テストデータ全文の実験には多くの試行回数が必要であった。そのため、テストデータの変換文と無変換文を 100 文ずつ抽出した小規模データを用いて実験を行った。ChatGPT に与えたプロンプトを図 1 に示す。赤色部分には与えるデータに応じて適

```

あなたは「口調スタイル名」で喋り、以下のサンプル文で示すような口調で話します。
<サンプル文>
{1: サンプル文}
...
{n: サンプル文}

以下のテスト文について、「口調スタイル名」の文に書き換えてください。
ただし、文の内容はできる限り保ち、文の変更は必要最低限にしてください。書き換えた文のみをテキスト番号とともに出力してください。
出力フォーマットは以下の通りです:
<出力文>
1: [書き換えた文]
...
100: [書き換えた文]

<テスト文>
{1: テスト文}
...
{100: テスト文}

<出力文>
    
```

図 1: ChatGPT に入力するプロンプトのテンプレート.

切な文が挿入され、{ 口調スタイル名 } には例示する口調スタイル名が、{ サンプル文 } には例示する口調スタイル文が、{ テスト文 } にはテストデータの文が入る。

## 5 評価実験

本実験では, テストデータの原文に対して各手法で口調スタイル変換を行い, 以下の評価指標を用いて各手法の口調スタイル変換性能を調べる.

### 5.1 評価指標

テキストスタイル変換では (i) スタイル変換精度, (ii) 内容の保持, (iii) 文の自然さ, の3つの指標で評価されることが望まれる [11].

**スタイル変換精度** スタイル変換精度では, スタイル変換器の出力文が目的のスタイルであるかを測定する. 口調スタイル測定器として BERT[12] ベースの分類器を利用し, 分類結果の accuracy 値 (以下 ACC と呼ぶ) を算出した. 学習データの原文と口調スタイル文を用いて事前学習済みモデル<sup>7</sup>に対しバッチサイズ 100, 学習率  $5 \times 10^{-5}$ , epoch 数 3 でファインチューニングを行った.

**内容の保持** 内容の保持では, スタイル変換前後の文の内容が同じかを測定する. 口調スタイル変換前後の文に対し SentenceBERT[13] を用いて文ベクトルを生成し,  $\cos$  類似度 (以下 SIM と呼ぶ) を算出した. 学習データの対訳データを用いて事前学習済みモデル<sup>8</sup>にバッチサイズ 128, 学習率  $2 \times 10^{-5}$ , epoch 数 1 でファインチューニングを行い, 目的関数に MultipleNegatives-RankingLoss<sup>9</sup>を用いた.

**文の自然さ** 文の自然さでは, スタイル変換器の出力文が目的スタイル下で文法的に自然であるかを測定する. GPT ベースの言語モデルで perplexity 値 (以下 PPL と呼ぶ) を算出した. 学習データの口調スタイル文を用いて事前学習済みモデル<sup>10</sup>に対しバッチサイズ 200, 学習率  $5 \times 10^{-4}$ , epoch 数 10 で, LoRA 手法によりファインチューニングを行った.

#### 5.1.1 各評価モデルの検証

各評価モデルが口調スタイル変換の評価能力を有しているか検証するため, テストデータおよび2種類のダミーデータで評価を行った. 表3に評価結果を示す. input copy はテストデータの原文をそのまま出力文とした際の評価結果を示し, different pair はテストデー

表 3: テストデータおよびダミーデータによる評価モデルの検証結果.

	変換文			無変換文		
	ACC $\uparrow$	SIM $\uparrow$	PPL $\downarrow$	ACC $\uparrow$	SIM $\uparrow$	PPL $\downarrow$
テストデータ	0.9823	0.9496	12.5754	0.1550	1.0000	23.8534
input copy	0.0053	1.0000	36.6903	0.1550	1.0000	23.8534
different pair	0.5686	0.0217	18.2110	-	-	-

タの  $i$  番目の原文と  $i+1$  番目の口調スタイル文のペアの評価結果を示している. テストデータでは全ての指標で高い精度を示した一方で, input copy では ACC が 0.0053 と非常に低く, PPL もテストデータより高い値となった. また, different pair では対訳関係のない原文と口調スタイル文に対して, SIM が 0.0217 と類似度が非常に低いことを示しており, 各評価モデルが適切に評価できていることが分かる.

### 5.2 事前学習済みモデルと学習設定

本実験では, 事前学習済み言語モデルとして T5 モデル<sup>11</sup>と GPT モデル<sup>12</sup>を用いた. 各モデルのファインチューニングには LoRA[14] 手法を利用し, バッチサイズ 200, 学習率  $5 \times 10^{-4}$ , epoch 数は step 数が 10000 になるよう学習データ数に応じて設定し, early\_stopping を 3 とした. また, 推論には貪欲法を用いた.

### 5.3 実験結果

#### 5.3.1 教師あり手法

教師あり手法の評価結果を表4に示す. 「変換文」

表 4: 教師あり手法の評価結果.

データ数	変換文			無変換文		
	ACC $\uparrow$	SIM $\uparrow$	PPL $\downarrow$	ACC $\uparrow$	SIM $\uparrow$	PPL $\downarrow$
50000	0.9800	0.9505	12.4807	0.1955	0.9982	23.4522
10000	0.9783	0.9518	12.5632	0.2155	0.9976	23.7861
5000	0.9725	0.9529	12.8685	0.2400	0.9961	23.7418
1000	0.9188	0.9571	14.8272	0.2963	0.9920	23.7263
500	0.8903	0.9585	15.6864	0.3100	0.9900	24.4895
100	0.7610	0.9608	21.0600	0.2860	0.9866	24.7192

において3つの評価指標全てで非常に高い精度となり, 学習データ数を減らすと SIM はあまり変わらず ACC と PPL の精度のみが低下した. また「無変換文」では, SIM が非常に高くなっている一方, ACC と PPL の精度が大きく低下した. ただし, 学習データ数を減らしていくと, SIM と PPL の精度を大きく低下させることなく ACC が上昇することが確認できる.

<sup>7</sup><https://huggingface.co/line-corporation/line-distilbert-base-japanese>

<sup>8</sup><https://huggingface.co/sonoisa/sentence-bert-base-japanese-mean-tokens-v2>

<sup>9</sup>[https://www.sbert.net/docs/package\\_reference/losses.html](https://www.sbert.net/docs/package_reference/losses.html)

<sup>10</sup><https://huggingface.co/rinna/japanese-gpt-neox-3.6b>

<sup>11</sup><https://huggingface.co/retrieva-jp/t5-xl>

<sup>12</sup><https://huggingface.co/rinna/japanese-gpt-neox-small>

### 5.3.2 逆翻訳手法

各翻訳モデルの結果を表5に示す。「変換文」におい

表 5: 逆翻訳手法における各翻訳モデルの評価結果.

翻訳モデル	変換文			無変換文		
	ACC ↑	SIM ↑	PPL ↓	ACC ↑	SIM ↑	PPL ↓
DeepL	<b>0.9840</b>	<b>0.8826</b>	9.6935	0.7878	<b>0.8937</b>	12.3695
NLLB	0.9778	0.7948	8.3483	0.8565	0.7854	9.1317
OPUS-MT	0.9740	0.6195	<b>6.5307</b>	<b>0.8945</b>	0.5994	<b>6.8972</b>

て ACC は 0.98 前後, PPL は 10 以下となっており, 教師あり手法と同等以上の結果となった. SIM の精度は翻訳モデルによって大きく異なっており, 一番高い DeepL で 0.8826, 一番低い OPUS-MT で 0.6195 となった. また, 「無変換文」でも翻訳モデルによらず ACC と PPL が教師あり手法の結果を上回り, 教師あり手法に比べて「無変換文」でも変換できていることが分かる.

### 5.3.3 Weighted Decoding 手法

$M^+ \cdot M^-$  実験 WD 手法における「 $M^+ \cdot M^-$  実験」(節 4.4.3 参照) の評価結果を表 6 と表 7 に示す. ここで T は temperature の値, k は topk サンプリングの値である. 表 6 から, T の値が小さいほど DAE の出力の

表 6:  $M^+ \cdot M^-$  実験の T ごとの評価結果 (k=100).

T	変換文			無変換文		
	ACC ↑	SIM ↑	PPL ↓	ACC ↑	SIM ↑	PPL ↓
0.3	0.5618	<b>0.9592</b>	23.5437	0.3923	<b>0.9809</b>	<b>20.9174</b>
0.4	0.7023	0.9395	<b>22.8568</b>	0.4680	0.9707	21.4520
0.5	0.7855	0.8892	29.1263	0.5338	0.9109	31.4880
0.6	0.8525	0.8332	41.7806	0.6380	0.8302	50.1507
0.7	<b>0.8943</b>	0.7961	50.5259	<b>0.7070</b>	0.7848	60.4794

表 7:  $M^+ \cdot M^-$  実験の k ごとの評価結果 (T=0.5).

k	変換文			無変換文		
	ACC ↑	SIM ↑	PPL ↓	ACC ↑	SIM ↑	PPL ↓
1	0.1668	<b>0.9875</b>	30.2315	0.2168	<b>0.9885</b>	<b>23.1108</b>
10	0.7668	0.8957	<b>24.9705</b>	0.5190	0.9175	28.0899
100	0.7855	0.8892	29.1263	0.5338	0.9109	31.4880
1000	<b>0.7868</b>	0.8887	29.2136	<b>0.5345</b>	0.9105	31.4250
10000	<b>0.7868</b>	0.8887	29.2136	<b>0.5345</b>	0.9105	31.4250

重みが強く, SIM が高くなる代わりに ACC が低くなる事が分かる. また, 他手法では ACC と PPL の精度にはある程度の相関があったが, WD 手法では ACC の精度が上がっても PPL の精度が下がっている.

次に, 表 7 から k の値が小さいほど言語モデル  $M^+$  と  $M^-$  が操作できる語彙が減少し, SIM の精度が上がり ACC が下がることが分かる. ただし, T の値の操作ほど

各指標の精度に影響を与えていない. また, k=1 は DAE のみで出力を生成している状態を示し, 言語モデル  $M^+$  と  $M^-$  の操作を受けないため ACC が低くなった.

$M^+$  only 実験 表 8 に「 $M^+$  only 実験」(節 4.4.3 参照) の評価結果を示す. 言語モデル  $M^-$  を用いておら

表 8:  $M^+$  only 実験の T ごとの評価結果 (k=100).

T	変換文			無変換文		
	ACC ↑	SIM ↑	PPL ↓	ACC ↑	SIM ↑	PPL ↓
0.8	0.7667	<b>0.8759</b>	15.1983	0.4400	<b>0.9181</b>	15.8812
0.9	0.8108	0.8569	14.4806	0.4935	0.8952	15.3171
1.0	0.8505	0.8379	13.9671	0.5670	0.8676	14.8000
1.1	0.9098	0.8026	13.7211	0.7028	0.8190	14.7624
1.2	<b>0.9313</b>	0.7771	<b>13.3558</b>	<b>0.7495</b>	0.7853	<b>14.1810</b>

ず言語モデル  $M^+$  による影響が強くなるため, T の値を高く設定して DAE とのバランスをとっている. また, ACC と PPL の精度に相関が出ており, 総合的な評価では  $M^+ \cdot M^-$  実験よりも高い変換精度となった.

### 5.4 口調スタイルごとの標準偏差

口調スタイルごとの評価結果に偏りが無いかわかるため, 各指標の標準偏差を算出した. 表 9 に標準偏差の値を示す. 逆翻訳手法では教師あり手法と同等の標

表 9: 各手法の評価値の標準偏差.

手法	変換文			無変換文		
	ACC	SIM	PPL	ACC	SIM	PPL
教師あり	0.0089	0.0159	1.5458	0.1250	0.0003	3.7138
逆翻訳	0.0117	0.0116	0.9766	0.0948	0.0106	1.9567
WD	0.0762	0.0266	2.3148	0.1588	0.0294	1.5969
ChatGPT	0.0658	0.0418	4.8774	0.0580	0.0418	10.9406

準偏差に収まっている一方で, WD 手法の ACC および ChatGPT 手法の全ての指標で標準偏差が大きくなっており, スタイルごとの性能に偏りがあることが分かる.

### 5.5 学習データ数による比較

データ数 50000 と 100 における各手法の評価結果について表 10 と表 11 にまとめる. ここで, 逆翻訳手法は DeepL 翻訳の結果を, WD 手法は「 $M^+$  only 実験」(T=1.1, k=100) の結果を記載している. 表 10 よりデータ数 50000 では逆翻訳手法が ACC と PPL で最も高い評価値となり, SIM は教師あり手法から 0.07 程度低い値で, 総合的に教師あり手法に匹敵する結果となった. また, 表 11 よりデータ数 100 では教師あり手法と逆翻訳手法で ACC の評価値が大幅に低下しており, ChatGPT 手法が総合的に最も高い性能となった.

表 10: 学習データ数 50000 における各手法の評価結果.

手法	変換文			無変換文		
	ACC	SIM	PPL	ACC	SIM	PPL
教師あり	0.9800	<b>0.9505</b>	12.4807	0.1955	<b>0.9982</b>	23.4522
逆翻訳	<b>0.9840</b>	0.8826	<b>9.6935</b>	<b>0.7878</b>	0.8937	<b>12.3695</b>
WD	0.9098	0.8026	13.7211	0.7028	0.8190	14.7624
ChatGPT	-	-	-	-	-	-

表 11: 学習データ数 100 における各手法の評価結果.

手法	変換文			無変換文		
	ACC	SIM	PPL	ACC	SIM	PPL
教師あり	0.7610	<b>0.9608</b>	<b>21.0600</b>	0.2860	<b>0.9866</b>	24.7192
逆翻訳	0.5903	0.9034	23.6577	0.5028	0.9376	<b>22.3865</b>
WD	-	-	-	-	-	-
ChatGPT	<b>0.9050</b>	0.8810	23.3524	<b>0.8675</b>	0.8589	29.8910

## 6 考察

### 6.1 無変換文に対する変換性能

表 4 から, 教師あり手法の「無変換文」では目的の口調スタイルでない入力文が変換されずに出力されている状態だと考えられる. セクション 3.2.3 において学習データに無変換文を含まなかったにも関わらず, 無変換を変換モデルが学習してしまっていることが伺える. これは, 変換文内に含まれる無変換箇所を変換モデルが過学習してしまったためだと考えられる. そのため, 学習データ数を減らすことで過学習を抑えることができ, ACC が上昇したのではないかと推察できる.

一方, 全ての教師なし手法で「無変換文」でも一定の ACC・PPL 精度が出ており, 教師あり手法よりもロバストに変換できていることが分かる. 対訳データを用いる逆翻訳手法でもロバストに変換ができていたのは, 逆翻訳で得られる文が多様性を持ち, 様々な変換パターンが学習データに含まれたためだと考えられる. 例として, 大阪弁の変換ツールは「いい天気だ」を「ええ天気や」に変換するが, 「いい天気です」は「ええ天気です」と「です」の部分を変換されず, 「～です」から「～や」への変換パターンが学習データに含まれない. 一方, 逆翻訳では「～や」が「～だ」や「～です」など様々な形に逆翻訳され, 多様なパターンを学習データに含むことができたのではないかと考えられる.

### 6.2 PPL が高く算出された手法とその原因

WD 手法では, 言語モデル  $M^+$  と  $M^-$  の対照性を利用して  $P(a|x)$  を計算しており, 言語モデル  $M^-$  を用いない場合, スタイルに関わらず言語モデル  $M^+$  の学習データに頻出する単語にバイアスがかかる恐れがある. しかし, 文の自然さの評価モデルも言語モデル  $M^+$  と  $M^-$  と同様の言語分布を持つデータで学習している

ため, PPL が低く算出される単語の生成を言語モデル  $M^-$  が遠ざけてしまっていると考えられる. そのため,  $T$  の値を大きくすると言語モデル  $M^-$  の影響が強くなり ACC の上昇とは反対に PPL 値が悪化し, 一方で言語モデル  $M^-$  を用いない「 $M^+$  only 実験」では ACC の精度と PPL の精度が共に上昇していると推察される.

ChatGPT 手法では, ChatGPT が事前学習したデータと本実験で扱った口調スタイルデータの言語分布に乖離があったため PPL が大きくなったと考えられる. 具体的には, セクション 3.2.2 で述べたように口調スタイル文には変換ツールの規則不備に起因する不自然な文が含まれており, 文法の自然さの評価モデルも変換ツールが生成した言語分布を学習している. ChatGPT は事前学習データに含まれる大阪弁や博多弁などの知識も生成に活用していると思われるが, 変換ツールが生成した口調スタイルがそうした一般的な方言・口調と乖離があったのではないかと考えられる.

### 6.3 各教師なし手法の詳細について

**逆翻訳手法** 表 10 より, データが十分に用意できる場合, 逆翻訳手法は他手法よりも高いスタイル変換性能を示しており, 逆翻訳が口調スタイル変換に有効であることが伺える. ただし, 翻訳モデルの性能によって文の内容保持性能が大きく左右されるため, 逆翻訳に用いる翻訳モデルの選択が重要となり, 翻訳が困難な口調スタイルを扱う場合は精度が低下する可能性がある. これは, ロバストな翻訳モデルでは口調スタイル文を適切に翻訳することができ逆翻訳文が擬似的な対訳文として上手く機能した一方で, 口調スタイル文を上手く翻訳できないモデルでは不完全な対訳データが生成され, 入力文との意味的類似性を考慮せず文生成を行う変換モデルが学習されるためだと考えられる.

**Weighted Decoding 手法** 表 10 から, WD 手法は他手法に僅かに及ばない結果となった. SIM の精度が低い原因として, DAE が削除された単語を復元するように訓練されているため, 元の文で省略されている主語や文末などの単語を補完して文を生成してしまっていることなどが考えられる. 一方, WD 手法では temperature と top k の値を操作することで生成を制御可能な利点がある. スタイル変換では変換精度と文の内容保持との間にトレードオフが存在するが, 上記の値を操作することでトレードオフのバランスを取ることができるため, 実課題における可用性が高いと考えられる.

**ChatGPT 手法** 表 11 より, データ数が非常に少数の場合は ChatGPT 手法が最も高い変換性能を示した. また, 表 10 が示すデータ数 50000 における各手法の結果

と比較しても競争力のある変換性能を発揮した。一方で、表9の結果より扱う口調スタイルごとの性能の偏りが他手法より大きいことも明らかとなった。これは、6.2で述べたように口調スタイル文と ChatGPT が事前学習した大阪弁や博多弁などの文の間に乖離が存在したからだと考えられる。実際に、他の教師なし手法では各スタイルの中でも比較的 ACC が高く出ている大阪弁について、ChatGPT 手法では非常に低い値となった。

## 7 おわりに

対話システムの口調制御に向け、方言や役割語を対象とした口調スタイル変換について、複数の教師なしテキストスタイル変換手法の比較検討を行った。実験の結果、全ての教師なし手法で教師あり手法よりもロバストな変換性能を示した。データ数が豊富な場合には、逆翻訳手法が教師あり手法に最も近い高い性能を示したが、翻訳器によって変換性能が左右されるため、扱う口調をロバストに翻訳できる翻訳器が重要となる。また、データ数が限られる場合には ChatGPT を用いた Few-shot 学習が最も高い変換性能を示したが、口調による性能差が大きいという欠点も確認された。本研究で提案した WD 手法は他手法にやや及ばなかったものの、従来の WD アプローチに DAE を組み合わせることでスタイル変換に活用可能だと判明し、パラメータ操作により変換精度と内容保持とのトレードオフを調整できる点で有用性を示すことができた。

本研究で扱わなかった口調スタイル変換として、口調スタイル文からノーマルスタイル文への変換や、口調スタイル文から別の口調スタイル文への変換などが挙げられる。これらの変換について今後の課題としたい。

## 参考文献

- [1] 赤間怜奈ほか。転移学習を用いた対話応答のスタイル制御。言語処理学会第23回年次大会発表論文集, pp. 338–341, 2017.
- [2] Hiroaki Sugiyama, et al. Empirical analysis of training strategies of transformer-based japanese chit-chat systems, 2021.
- [3] 谷川晃大ほか。変分オートエンコーダと注意機構を用いた発話文のキャラクター変換。人工知能学会全国大会論文集 第32回全国大会, p. 4G201, 2018.
- [4] 江崎拓哉ほか。Flow-base モデルを用いた文のスタイル変換。第12回データ工学と情報マネジメントに関するフォーラム, No. F7-1, 2020.
- [5] Shrimai Prabhunoye, et al. Style transfer through back-translation. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, pp. 866–876, 2018.
- [6] Alisa Liu, et al. DExperts: Decoding-time controlled text generation with experts and anti-experts. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, pp. 6691–6706, 2021.
- [7] Colin Raffel, et al. Exploring the limits of transfer learning with a unified text-to-text transformer. *The Journal of Machine Learning Research*, Vol. 21, No. 1, pp. 5485–5551, 2020.
- [8] NLLB Team, et al. No language left behind: Scaling human-centered machine translation. 2022.
- [9] Jörg Tiedemann and Santhosh Thottingal. OPUS-MT — Building open translation services for the World. In *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*, 2020.
- [10] Alec Radford, et al. Improving language understanding by generative pre-training. 2018.
- [11] Remi Mir, et al. Evaluating style transfer for text. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Vol. 1, pp. 495–504, 2019.
- [12] Jacob Devlin, et al. 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*, Vol. 1, pp. 4171–4186, 2019.
- [13] Nils Reimers, et al. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pp. 3982–3992, 2019.
- [14] Edward J Hu, et al. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*, 2022.

# ChatGPTを用いたストレスの解決を支援するチャットボットの開発

## Developing a Chatbot to Help Resolve Stress using ChatGPT

岡本 宗一郎<sup>1\*</sup> 砂山 渡<sup>2</sup> 服部 峻<sup>2</sup>  
Soichiro Okamoto<sup>1</sup> Wataru Sunayama<sup>2</sup> Shun Hattori<sup>2</sup>

<sup>1</sup> 滋賀県立大学大学院 工学研究科

<sup>1</sup> Graduate School of Engineering, The University of Shiga Prefecture

<sup>2</sup> 滋賀県立大学 工学部

<sup>2</sup> School of Engineering, The University of Shiga Prefecture

**Abstract:** According to a survey by the Ministry of Health, Labour and Welfare, the percentage of workers who feel stress related to their work and professional life continues to exceed half. Therefore, he states that it is important for each worker to provide self-care. In this study, we developed a chatbot that supports stress resolution by extracting stress comments based on stress evaluation items from conversational content using ChatGPT, and by suggesting solutions through conversation. The results of the experiment confirmed that chatbot-based conversations are effective in supporting stress resolution.

## 1 はじめに

厚生労働省による最近の「過労死等防止対策白書」[1]においては、仕事や職業生活に関することで強い不安、悩み、ストレスを感じている労働者の割合は半数を超える状況が続いている。そこで、労働者のメンタルヘルス不調を未然に防ぐために、企業側にストレスチェック制度の実施が義務付けられている。しかし、時間や金銭的な理由などでストレスチェックを受けた者の中で、医師による面接指導を受けている者は限られている。そのため、労働者一人ひとりがセルフケアを行うことが重要であるとしている。

セルフケアを促進していくためには、誰でも気軽に利用できる必要がある。スマートフォンの普及により、認知行動療法(CBT)などの心理療法をセルフで実施できるスマートフォンアプリなどの開発が進められている。また、LINEやFacebookなどのSNSが開発環境の提供を行っているため、チャットボットの導入が容易になりつつある。既存のチャットボットを用いたシステムでは、心理療法の知識やプログラムを提供するという観点でのアシストになる。スマートフォンとチャットボットを活用したメンタルヘルス講座を提案するシステムの研究[2]によると、チャットボットの導入はストレス

軽減やモチベーション維持の効果が示された。チャットボットはプログラムの提供や、利用意欲の向上目的として利用されている。そこで本研究は、チャットボット自身が軽度の、日常的なストレスを対象としたストレス解決を支援することを目指す。これにより、誰でも気軽にセルフケアができることを目的としている。

## 2 関連研究

### 2.1 SNS投稿からストレス分析を行う研究

本節では、SNSの投稿からストレス分析を行う関連研究について述べる。インターネットの普及によりSNSを利用するユーザが大幅に増加している現在、ユーザの感情的な反応をより多く取得することができるようになった。実際にTwitterの投稿内容を収集し、KH Coder Ver.3を用いて計量テキスト分析を行うことで新型コロナウイルス感染拡大に伴う不安やストレスの動向について調査する研究[3]がある。SNSデータを用いてストレスの実態の推移を把握できる可能性を示している。

他にも、自動的にストレスを検知するモデルの作成を行っている研究がある。畳み込みニューラルネットワーク(CNN)を使用して、自殺関連の投稿を検出する研究[4]や、テキストからうつ病を推測するLSTMモデルを構築する研究[5]などが挙げられる。これらの研究は、テキストから精神医学的ストレス要因を自動

\*連絡先：滋賀県立大学大学院工学研究科 電子システム工学専攻  
岡本 宗一郎  
〒522-8533 滋賀県彦根市八坂町2500  
E-mail: ov23sokamoto@ec.usp.ac.jp

的に抽出し、自殺や精神障害を未然に防ぐことを目的としている。しかし、これらの研究はテキスト情報からストレスを検知を行っているが、その解決までは支援していない。本研究では、ChatGPTを用いてストレス検出するとともに、ユーザのストレス解決を支援する。

## 2.2 メンタルヘルスケアにおけるチャットボットの利用に関する研究

本節では、メンタルヘルスケアにおけるチャットボットの利用に関する研究について述べる。1996年にWeizenbaumによってELIZA[6]というチャットボットが開発されて以降、次々と新しいチャットボットが開発されている。それに伴い、メンタルヘルスケアにおいても、チャットボットの導入がなされている。知覚制御療法(PCT)に基づいて作成されたMLYOというチャットボットを用いた研究がある[7]。この研究では、被験者に割り当てられたプログラムで現在経験している問題についての話し合いにおいて、MLYOは苦痛、うつ病、不安およびストレスの軽減の助けになるということがわかった。

Webベースの認知行動療法(CBT)アプリであるWoebotを用いた研究[8]では、CBTに基づいた自己啓発プログラムをチャットボットを介して提供することで、被験者の不安を大幅に軽減した。しかし、これらの研究ではチャットボットがストレス解決を行うのではなく、プログラムを提供する役割を担っている。そこで本研究では、ユーザと会話をを行い、解決策を提案するチャットボットの開発を行い、その有用性を評価する。

## 3 ストレスコメントの定義

### 3.1 ストレス反応について

本節では、本研究で扱うストレス反応について定義する。文部科学省によると、ストレス反応は、長時間ストレスの刺激を受けた場合や、強いストレスを受けた時に生じる生体反応であり、ストレスに対する生体の自然な適応反応とされている[9]。また、ストレス反応は以下の三つに分類される。

- 「心理的ストレス反応」  
情緒的反応として、不安、イライラ、恐怖、落ち込み、緊張、怒り、罪悪感、感情鈍麻、孤独感、疎外感、無気力などの感情が現れる。
- 「行動的ストレス反応」  
怒りの爆発、けんかなどの攻撃的行動、過激な行動、泣く、引きこもり、孤立、拒食・過食、幼児

返り、チック、吃音、ストレス場面からの回避行動などが現れる。

- 「身体的ストレス反応」  
動悸、異常な発熱、頭痛、腹痛、疲労感、食欲の減退、嘔吐、下痢、のぼせ、めまい、しびれ、睡眠障害、悪寒による震えなど、全身にわたる症状が現れる。

総務省の『情報通信政策研究』に記載されている調査[10]には、ネットアプリへの逃避の流れとして、日々生じるストレスによって共感による安心や現実的感覚の遮断が高まると考えられている。このことから、本研究では、主に「心理的ストレス反応」に該当するコメントをストレスコメントと定義する。

### 3.2 ストレス評価項目の設定

本節では、心理的ストレス反応を検出するために使用するストレス評価項目について定義する。本研究では、POMS(気分プロフィール)[11]という心理的ストレス尺度を用いる。POMSについて研究によると、多様な感情状態を呈するうつ病等において、その患者特有の感情状態を尺度化できるものと考えられるとしている。POMSで定義されている項目には、緊張-不安、抑うつ-落ち込み、怒り-敵意、活力、疲労、混乱の6項目がある。活力という項目は、他の項目とは他の5項目とは負の相関が認められる。そこで本研究においては、これらの項目の中で活力を除く5項目のいずれかに該当するコメントをストレスコメントと定義する。各ストレス項目の詳細を以下に示す。

- 「緊張-不安」  
緊張や不安の高まりが感じられる。神経の昂りや落ち着かないなどの特徴が見られる。
- 「抑うつ-落ち込み」  
自分に価値がない、希望が持てないと感じている。罪悪感があるなど自信が喪失している状態が見られる。
- 「怒り-敵意」  
怒りや他者への反感が感じられる。内心の腹立たしさや他人に意地悪したいなどの思いがある。
- 「疲労」  
疲労感があると感じられる。  
意欲や活力が低下している状態が見られる。
- 「混乱」  
当惑と認知効率の低さが感じられる。頭が混乱して考えがまとまらない状態が見られる。

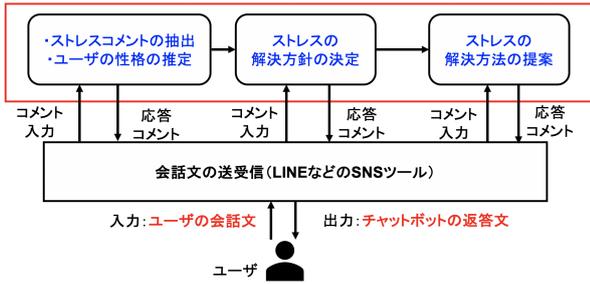


図 1: チャットボットを用いたストレス解決を支援する会話システムの構成

表 1: 平常時の会話生成プロンプト

プロンプト
「あくあ」というキャラクターになりきって、相手の話を聞くことに専念し、相手にたくさん話してもらえるように会話を続けてください。
「あくあ」の特徴
・性別：大人の女性
・口調：一人称は「わたし」を使う。語尾に「～だよ」や「～だよね」などを使う。
会話における制約条件
・自分なりの解決策などは提案しないでください。
・相手の悩みについて、優しく質問をしてください。
・文字数は 50 文字以内にしてください。

## 4 チャットボットを用いたストレス解決を支援する会話システム

本章では、チャットボットを用いたストレス解決を支援する会話システムについて述べる。図 1 にストレス解決を支援するチャットボットの構成を示す。

会話システムは大きくチャットボット (ChatGPT)[12] と LINE プラットフォーム [13] の 2 種類で構成されている。LINE プラットフォームでは、主にユーザの会話文の入力と、チャットボットの出力表示を行なっている。ユーザの会話文からストレスコメントの抽出と性格推定を行う。そして、ストレス解決方針を決定し、ユーザとの会話を元に解決方法の提案を行う。

### 4.1 チャットボットとの対話環境

本節では、チャットボットとの対話環境について述べる。システムの入力はユーザの会話文である。入力は LINE から行う。LINE プラットフォームを通して入力されたユーザの会話文をチャットボットへと入力する。Messaging API[14] を利用して、LINE プラットフォームとチャットボットとのデータ送受信を可能にしている。プラットフォームとして LINE を選んだ理由

表 2: ストレスコメントを抽出するプロンプト

プロンプト
あなたは精神医学に詳しい心理学者として、入力された文章が、以下の 6 つの項目のいずれかに当てはまるかを、「項目名」で回答してください。 いずれの項目にも当てはまらない場合は、「なし」と回答してください。
<項目名：定義>
「緊張と不安」：緊張や不安の高まりが感じられる。神経の昂りや落ち着かないなどの特徴が見られる。
「抑うつと落ち込み」：自分に価値がない、希望が持てないと感じている。罪悪感があるなど自信が喪失している状態が見られる。
「怒りと敵意」：怒りや他者への反感が感じられる。内心の腹立たしさや他人に意地悪したいなどの思いがある。
「疲労」：疲労感があると感じられる。意欲や活力が低下している状態が見られる。
「混乱」：当惑と認知効率の低さが感じられる。頭が混乱して考えがまとまらない状態が見られる。

は、誰でも簡単に利用できるからである。総務省が行ったソーシャルメディアの利用率調査結果 [15] によると、2022 年において全年代での LINE 利用率が 92.5 % である。このことから、誰でも簡単にチャットボットに相談できるようにするために、LINE を利用している。

システムの出力は、入力されたユーザの会話文に対する返答文である。チャットボットに生成された会話文が Messaging API を利用することで、LINE プラットホームに送られる。

### 4.2 平常時の会話応答

本節では、チャットボットの平常時に会話応答について述べる。表 1 に平常時の会話生成プロンプトを示す。ユーザにたくさん会話をしてもらうために、聞き手に徹するプロンプトにしている。

会話システムでは、ユーザはストレスを感じている時は積極的に相談してもらい、それ以外でも自由に会話をしてもらうことを想定している。ユーザが話やすいように、以下のようなキャラクター設定にしている。

- ・性別：大人の女性
- ・口調：一人称は「わたし」を使う。語尾に「～だよ」や「～だよね」などを使う。

### 4.3 ストレスコメントの抽出

本節では、ストレスコメントの抽出について述べる。会話システムでは、常に入力されたユーザの会話文か

表 3: ユーザの性格推定と解決の方針を決定するプロンプト

役割	プロンプト
性格推定	入力される「ユーザ」と「あくあ」の会話履歴を見て、ユーザの性格を簡条書きで出力してください。
解決の方針の決定	ユーザの性格とストレス状態を考慮し、最も適切と考えられるストレス低減方法を、次の2つの項目から一つ選んで「項目名」だけを回答してください。 項目の定義 ・「逃避的解決」: やりたい事を尋ね、リフレッシュの手助けをする ・「根本的解決」: 相手のストレスの原因解決の手助けをする

表 4: 解決方法を生成するプロンプト

役割	プロンプト
逃避的解決方法の生成	ユーザは以下に示すストレス状態と性格である。 入力される「ユーザ」と「あくあ」の会話履歴から、導かれるユーザがリラックスできる方法の内容だけを三つ、番号付きで簡単に回答してください。
根本的解決方法の生成	ユーザは以下に示すストレス状態と性格である。 入力される「ユーザ」と「あくあ」の会話履歴から、導かれる問題を解決する方法の内容だけを三つ、番号付きで簡単に回答してください。

らストレスコメントの判定を行なっている。ストレスコメントの定義は3章に示した通りである。ストレスコメントの抽出を行うプロンプトを表2に示す。

#### 4.4 ストレスの解決方針の決定

本節では、ストレスの解決方針の決定について述べる。ユーザの性格推定と解決の方針を決定するプロンプトを表3に示す。推定したユーザの性格とストレス状態を元に、解決方針の決定を行う。性格とストレス状態を考慮することで、利用するユーザごとに適した会話を行うようにする目的がある。ユーザの性格推定と解決方針の決定は、平常時の会話回数が5回以上且つ、ストレスコメントが1つ以上となったタイミングで行われる。

解決の方針は大きく二つある。

表 5: 質問を含んだ会話文の生成プロンプト

役割	プロンプト
逃避的解決のための質問	「あくあ」という対話相手のストレスを解消するための専門的知識を有するキャラクターになりきって、やりたい事を尋ね、ストレス解消方法を一緒に見つけ出す会話を続けてください。 「あくあ」の特徴 ・性別: 大人の女性 ・口調: 一人称は「わたし」を使う。語尾に「~だよ」や「~だよね」などを使う。 会話における制約条件 ・相手の気持ちに寄り添いながら、相手の趣味や、やりたいことを尋ねてください。 ・文字数は50文字以内にしてください。
根本的解決のための質問	「あくあ」という対話相手のストレスを解消するための専門的知識を有するキャラクターになりきって、ストレス原因の特定と理解を進める質問をし、一緒に解決策を見つけて出す会話を続けてください。 「あくあ」の特徴 ・性別: 大人の女性 ・口調: 一人称は「わたし」を使う。語尾に「~だよ」や「~だよね」などを使う。 会話における制約条件 ・相手の体調を気遣いながら、ストレス解決に向けた質問をしてください。 ・文字数は50文字以内にしてください。

- 逃避的解決: やりたい事を尋ね、リフレッシュの手助けをする
- 根本的解決: 相手のストレスの原因解決の手助けをする

この二種類から適切な解決方針をユーザの性格とストレス状態を元に決定する。

#### 4.5 ストレスの解決方法の提案

本節では、ストレスの解決方法の提案について述べる。チャットボットとユーザとの会話内容と推定されたユーザのストレス状態、性格を元に、3つの解決方法を提案する。解決方法を提案するために、ユーザにストレス解決方法を見つけて出すための質問を行う。質問を含んだ会話文の生成プロンプトを表5に示す。

表 6: 実験で扱った人物設定

人物	設定
人物 1	就職活動が上手くいっておらず、面接で連続して 10 社以上不採用になって落ち込んでいる 22 歳の情報学部の大学生。 IT 企業に興味を持っている。 自身の専門知識やスキルをアピールポイントとしてアピールしているが、なかなか企業からの内定を得ることが出来ないでいる。
人物 2	飲食店バイトでのミスで先輩に怒られ、落ち込んでいる 19 歳の大学生。 ラッシュ時に注文が殺到し、オーダーの取りこぼしというミスをしてしまった。 そのことで怒られてしまい、少し落ち込んでいる。
人物 3	テニスサークルでの人間関係で悩みを抱えている 18 歳大学生。 新しい人とのコミュニケーションに苦手意識を持っており、特に、先輩との関係を築けないでいる。

質問を伴う会話が 7 回以上になると、ユーザが入力した会話文が疑問文であるかの判定を行う。ユーザの会話文が疑問文でなかった場合、「私もいくつか解決策考えてみたけど、聞いてみる?聞きたい場合は「はい」、そうでなければ、「いいえ」を入力してね。」というメッセージを送信し、解決方法の提案を行う。解決策を生成するプロンプトを表 4 に示す。ユーザは提示された解決策の中から、納得できる解決策の一つを選ぶ。納得できる解決策がなければ、もう一度チャットボットとの会話に戻る。

## 5 ストレス解決支援の評価実験

本章では、チャットボットを用いたストレス解決を支援する会話システムの有用性を検証した実験について述べる。本実験では、チャットボットを用いた会話システムがユーザのストレス解決を支援できるかを検証した。実験参加者に三つの仮想的な人物になりきってもらい、チャットボットとの会話を通じて納得できるストレスの解決策を得ることができるかを評価する実験を行った。

### 5.1 実験手順

本節では、実験手順について述べる。この実験では、実験参加者に表 6 に示す人物になりきってもらい、チャットボットと会話を会話をしてもらった。この実験は、14 人の男性を被験者とした。被験者の内訳は、社会人 1

人、大学院生 5 人、大学生 8 人である。実験手順を以下に示す。

1. 人物 1 になった前提で、「あくあ」(チャットボット)に現状を伝えた上で、落ち込んだ気持ちを解消するための会話をする。
2. 人物 2 になった前提で、「あくあ」(チャットボット)に現状を伝えた上で、落ち込んだ気持ちを解消するための会話をする。
3. 人物 3 になった前提で、「あくあ」(チャットボット)に現状を伝えた上で、よりよい人間関係を築くための会話をする。

会話を行ったあと、納得できる解決策が得られたかを 4 段階評価で回答してもらった。解決策に対する 4 段階評価の詳細を以下に示す。

- 4: 納得できる解決策が得られた
- 3: 少し納得できる解決策が得られた
- 2: あまり納得できる解決策が得られなかった
- 1: 納得できる解決策が得られなかった

また、実験後に各被験者にアンケートを行った。質問は以下に示す内容で自由記述で回答してもらった。

- 質問 1: 人間との会話に比べて良かったところを教えてください。
- 質問 2: 人間との会話に比べて悪かったところを教えてください。

### 5.2 各被験者の解決策に対する評価結果

本節では、チャットボットとの会話から得られた解決策に対する評価結果について述べる。表 7 に各被験者の解決策に対する評価結果を示す。各人物における解決策の評価平均が、3.8, 3.4, 3.6 であったことから、ストレス解決の支援に有効であったと考えられる。

### 5.3 抽出されたストレスコメント

本節では、実験において抽出されたストレスコメントについて述べる。対話から抽出されたストレスコメントの例を表 8 に示す。人物 1 では、就職活動に対する不安や不採用に対する落ち込みが抽出された。人物 2 では、ミスで落ち込んでいることや先輩に怒られたことが抽出された。人物 3 では、サークルに馴染めないことに対する不安などが抽出された。このことから、

表 7: 各被験者の解決策に対する評価結果

人物設定	A	B	C	D	E	F	G	H	I	J	K	L	M	N	平均
人物 1	4	3	4	3	4	4	4	3	4	4	4	4	4	4	3.8
人物 2	3	3	4	3	3	4	3	4	3	4	4	3	3	4	3.4
人物 3	3	3	4	2	3	4	4	3	4	4	4	4	4	4	3.6

表 8: 対話から抽出されたストレスコメントの例

ストレス項目	人物 1 (就活)	人物 2 (バイト)	人物 3 (サークル)
緊張-不安	<ul style="list-style-type: none"> <li>・追い詰められているから常に不安だよ</li> <li>・緊張してうまく話せなくて伝わっていない気がする</li> </ul>	<ul style="list-style-type: none"> <li>・注文が殺到してオーダーを取りこぼしちゃったんだよ</li> <li>・オーダーの取りこぼしのミスをしてしまいました</li> </ul>	<ul style="list-style-type: none"> <li>・コミュニケーションに苦手意識を持ってるんだ</li> <li>・人見知りで場に馴染めないんだ</li> </ul>
抑うつ-落ち込み	<ul style="list-style-type: none"> <li>・10 社不採用になってつらい</li> <li>・企業から内定がもらえずやる気を無くしてしまいました</li> </ul>	<ul style="list-style-type: none"> <li>・思い出してしまって何もやる気が出ない状態です</li> <li>・最近バイトのミスで怒られて落ち込んでるの</li> </ul>	なし
怒り-敵意	なし	<ul style="list-style-type: none"> <li>・お客様も先輩もめちゃくちゃ怒ってました</li> <li>・お客さんには謝って解決したけど先輩に怒られた</li> </ul>	<ul style="list-style-type: none"> <li>・先輩はかなり乱暴な性格です</li> <li>・先輩に痛い目を合わせたいです</li> </ul>
疲労	<ul style="list-style-type: none"> <li>・一日くらい完全に休んでまたがんばろかな</li> </ul>	<ul style="list-style-type: none"> <li>・そのまま終わるまで働いた</li> <li>・忙しかったから仕方ない</li> </ul>	<ul style="list-style-type: none"> <li>・話すのが面倒くさくなります</li> </ul>
混乱	<ul style="list-style-type: none"> <li>・質問されると頭が真っ白になる</li> <li>・今は考えたくない</li> </ul>	<ul style="list-style-type: none"> <li>・話しかけ方がわかんない</li> <li>・よく覚えてない</li> </ul>	<ul style="list-style-type: none"> <li>・自分が知らない話題が多かった</li> <li>・何を話せばいいのかわかりません</li> </ul>

設定した人物情報と近いストレスコメントが抽出できたことがわかる。

怒り-敵意の項目に着目すると、第三者の怒りの感情を検出していることがわかった。適切に抽出するには、文章の主語を判定する必要があると考えられる。

#### 5.4 各被験者が選択した解決策

本節では、各被験者が最終的に選択した解決策について述べる。表 9 に選択された解決策の例を示す。

人物 1 と人物 3 では、根本的解決に向けた具体的な活動方法が提示多く提示されていた。人物 1 においては、面接でどういった話をしているのか、練習状況はどうかといった就活状況について多くの会話がなされていた。人物 3 では、サークルメンバーとはどう接しているのかといった人間関係について会話がなされていた。そのため、各被験者の活動状況を聞き出したことにより、根本的解決に向けた具体的な活動方法が提案されたと考えられる。

人物 2 では、逃避的解決に向けた解決策が多く提示された。多くの被験者がバイトのミスで先輩に怒られ

たことに落ち込んでいることをチャットボットに伝えていた。それに対してチャットボットは気持ちをリフレッシュさせるために各被験者がやりたいことや、好きなことを聞き出す質問をしていた。そのため、ミスで落ち込んでいる気持ちを和らげるための方法が提示されたのだと考えられる。

本研究の会話システムは、会話で聞き出した内容を ChatGPT に入力し解決策の生成を行なっている。このことから、会話内容が反映された解決方法を提示できたことがわかった。

#### 5.5 アンケート結果

本節では、アンケート結果について述べる。人間との会話に比べて良かった点と悪かった点を尋ねたアンケート結果（一部抜粋）を表 10 に示す。質問 1 の人間と比べて良かった点では、人間と違って気軽に話せる、いつでも話を聞いてくれるという意見が多くあった。これらのことから、チャットボットには実際の人間に比べて気軽に話やすい利点があると考えられる。

質問 2 の人間と比べて悪かった点では、解決方法の

表 9: 被験者が最終的に選択した解決策の例

人物設定	解決策
人物 1 (就活)	<ul style="list-style-type: none"> <li>・ 模擬面接を行いフィードバックを得ることで、対策を立てる</li> <li>・ アピールポイントの見直し</li> <li>・ 面接時のコミュニケーションスキルの向上</li> <li>・ 模擬面接を受けることで、実際の面接に向けた準備と自信をつける</li> </ul>
人物 2 (バイト)	<ul style="list-style-type: none"> <li>・ 軽い運動や散歩をすることでリラックスを試みる</li> <li>・ J-POP やリズムある音楽を聞いてリラックスする</li> <li>・ 事前に忙しい時間帯に対応するための具体的なプランニングや優先順位の設定を行なう</li> <li>・ ゲームをプレイすることでリラックスし、気分転換を図る</li> </ul>
人物 3 (サークル)	<ul style="list-style-type: none"> <li>・ テニスサークルの活動やイベントへの参加を通じて、共通の経験を増やす</li> <li>・ 共通の趣味や関心事を見つけて話題にする</li> <li>・ テニスについての情報を集めて話の種にする</li> <li>・ 先輩の好きなことや興味があることについて、探りを入れる</li> </ul>

表 10: アンケート結果：人間との会話に比べて良かった点と悪かった点（一部抜粋）

質問 1：良かった点	<ul style="list-style-type: none"> <li>・ こちらの気持ちに常に寄り添ってくれるところ。</li> <li>・ ずっと聞いてくれるところ</li> <li>・ 人間相手ではないので会話に気を遣わなくてもよい点や悩みを相談しても恥ずかしくない点</li> <li>・ 人間に話すと怒られそうな内容でもしっかり聞いてくれる所</li> <li>・ すぐ返信してくれる，なんでも肯定してくれる</li> <li>・ 人じゃないと分かっているからこそ気軽に話せました</li> <li>・ 敬語とか対人だと気にするところが気にしなくてよいところ</li> </ul>
質問 2：悪かった点	<ul style="list-style-type: none"> <li>・ 状況の確認が機械的な受け答えだと感じた。</li> <li>・ 会話の道中で解決策が見つかったりするときは手間だと感じた。</li> <li>・ 解決策の提示のタイミングが不自然に感じた。</li> <li>・ たまに質問攻めされるところ</li> <li>・ 解決策を提示するまでに何度も同じような言葉で励まされる</li> <li>・ 友達や先輩に相談する時と比べて詰められている感が強い。</li> <li>・ 形式的な返答であると感じてしまった所</li> </ul>

提示のタイミングや、質問の仕方、返答が機械的であるという意見が多くあった。システムでは、解決策の提示を行う条件に一定回数以上の会話回数を設けていた。それにより、タイミングに違和感を感じたのだと考えられる。会話の流れから解決方法を提示するアルゴリズムを考える必要がある。

質問の仕方について、システムでは、質問の回数などの設定は行なっていなかった。そのため、質問の間隔を作ったり、ユーザがストレスを感じないような質問にする必要がある。

返答が機械的であると感じてしまう意見については、親しみやすい返答になるようにプロンプトを修正する必要がある。

ことでストレス解決を支援するチャットボットの開発を行った。

また、3人のストレスを抱えた人物を想定した会話を行ってもらい、ユーザのストレス解決支援に有用であるかを確認する評価実験を行った。この実験結果から、ストレス解決を支援する上で、開発したチャットボットの有用性が確認できた。

これを受けて今後の課題として、質問間隔や解決策提示のタイミングの調整、親しみやすい返答へのプロンプトの修正を行う。これにより、ストレス解決を支援を行うことを目標としていきたい。

## 参考文献

- [1] 職場におけるメンタルヘルス対策の状況：(URL)  
<https://www.mhlw.go.jp/content/11200000/000845811.pdf> (2023年11月10日確認)

## 6 おわりに

会話内容からストレス評価項目に基づいたストレスコメントを抽出して、会話を通して解決策を提示する

- [2] Kamita, Takeshi, et al.: A chatbot system for mental healthcare based on SAT counseling method, *Mobile Information Systems*, Vol.2019 (2019)
- [3] 四方田健二:新型コロナウイルス感染拡大に伴う不安やストレスの実態 Twitter 投稿内容の計量テキスト分析から, *体育学研究*, Vol.65, pp.757 – 774 (2020)
- [4] Du, J., Zhang, Y., Luo, J., Jia, Y., Wei, Q., Tao, C., Xu, H.: Extracting psychiatric stressors for suicide from social media using deep learning, *BMC medical informatics and decision making*, Vol.18, No.2, pp.77 – 87 (2018)
- [5] Amanat, A., Rizwan, M., Javed, A.R., Abdelhaq, M., Alsaqour, R., Pandya, S., Uddin, M.: Deep learning for depression detection from textual data, *Electronics*, Vol.11, No.5, 676 (2022)
- [6] Weizenbaum, J.: ELIZA: A computer program for the study of natural language communication between man and machine, *Comm. ACM*, Vol.9, No.1, pp.36–45 (1966)
- [7] Gaffney, H. et al.: Manage your life online (MYLO): A pilot trial of a conversational computer-based intervention for problem solving in a student sample, *Behavioural and Cognitive Psychotherapy*, Vol.42, No.6, pp.731 – 746 (2014)
- [8] Fitzpatrick, Kathleen Kara, Alison Darcy, and Molly Vierhile: Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial, *JMIR mental health*, Vol.4, No.2 (2017):
- [9] 第 2 章 心のケア各論:(URL)[https://www.mext.go.jp/a\\_menu/shotou/clarinet/002/003/010/003](https://www.mext.go.jp/a_menu/shotou/clarinet/002/003/010/003) (2023 年 11 月 10 日確認)
- [10] 大野志郎 : ネット逃避の現状—インターネットおよびアプリケーションの過剰使用者へのオンラインインタビュー調査より, *総務省学術雑誌『情報通信政策研究』*, Vol.2, No.2, pp.1 – 17 (2019)
- [11] 赤林朗, 横山和仁, 荒記俊一, 島田恭子:POMS(感情プロフィール検査)日本語版の臨床応用の検討, *心身医学*, Vol.31, No.7, pp.577 – 582 (1991)
- [12] OpenAI: (URL)<https://openai.com/> (2023 年 11 月 10 日確認)
- [13] LINE: (URL) <https://line.me/ja/> (2023 年 12 月 25 日確認)
- [14] LINE Developers: (URL) <https://developers.line.biz/ja/services/messaging-api/> (2024 年 1 月 28 日確認)
- [15] 令和 3 年度情報通信メディアの利用時間と情報行動に関する調査報告書 : (URL) [https://www.soumu.go.jp/main\\_content/000831290.pdf](https://www.soumu.go.jp/main_content/000831290.pdf) (2024 年 1 月 28 日確認)

# ChatGPT を用いた初対面のオンラインコミュニケーション における質問力向上支援システム

## A Support System for Improving Questioning Skills in Online Communication with Person Meeting for the First Time using ChatGPT

松田 駿宏<sup>1\*</sup> 砂山 渡<sup>1</sup> 服部 峻<sup>1</sup>  
Takahiro Matsuda<sup>1</sup> Wataru Sunayama<sup>1</sup> Shun Hattori<sup>1</sup>

<sup>1</sup> 滋賀県立大学 工学部

<sup>1</sup> School of Engineering, The University of Shiga Prefecture

**Abstract:** The term “questioning skills” is considered a necessary skill for effective communication. In order to acquire this skill, you need repetitive experience but there is not many opportunities to practice for acquiring this skill in real life. Therefore, in this study, we propose a system that enables users to practice for acquiring “questioning skills” by using ChatGPT, large language model that can imitate chat-based dialogues.

## 1 はじめに

コミュニケーションを円滑に、また効果的に行うためには「質問力」が不可欠と考えられ、「質問力」をうまく用いることができれば、個人やビジネスにおいて有利な状況をもたらす、また個人のスキルアップや新しい知見に繋げることもできる。したがって、「質問力」の向上を支援することは様々なメリットをもたらすこととなる。

しかしながら、個人でこの能力を効果的に養うことは難しい。なぜなら、基本的に質問はコミュニケーションを行う相手に対して行うものであり、かつ個人で「質問力」を向上させることのできるシステムや手段は見受けられないというのが現状としてあるためである。

また、近年のコミュニケーションでは、SNS 等の発達により、オンライン上でのチャットベースのコミュニケーションが盛んになっている。

そこで、本研究ではチャットベースでのコミュニケーションを模倣可能な大規模言語モデル ChatGPT を用いて、架空の人物を生成し、ユーザがこれと対話することで、ユーザの「質問力」の向上を図るための経験を積むことができるシステムを提案する。具体的には、ChatGPT により生成した人物像情報を持つ人物に対してユーザが質問を入力し、その質問に対して、ChatGPT を用いた評価やアドバイス、その人物からの回答を模

した文章を生成し、ユーザにフィードバックを行うことで、疑似的にコミュニケーションの経験を積むことのできる環境を提供し、質問力の向上を支援する。

## 2 関連研究

### 2.1 質問力についての研究

質問力を評価する手法を開発することを目的としている研究がある [1]。この研究では、結果的に、日頃から質問力向上の支援を行っていく必要性が説かれている。しかし、現状、日常的にそのような支援を行うことのできる仕組みやシステムは見受けられない。そこで、本研究では、日常的に質問力向上支援を行うことのできるシステムの提案を目指す。その上で、まずは「質問力」とは何かということについて定義する必要がある。

「質問力」の定義について、池田は「質問力」を「相手や自分、状況などを偏見をもたずに捉え、それを質問としてバイアスをかけずに表現できるような力」[2] であるとしている。一方、本研究の目的は「オンラインコミュニケーションにおける質問力向上支援」であり、対人のコミュニケーションが前提となる。その上で、前述の定義を解釈し、本研究に適合するように再構築し、「質問力」の定義を「相手の持つ情報を正確に把握し、それに対して適切に表現した質問を行うことで、さらなる情報を引き出す力」とした。

\*連絡先： 滋賀県立大学 工学部  
松田 駿宏  
〒 522-8533 滋賀県彦根市八坂町 2500  
E-mail: tz23tmatsuda@ec.usp.ac.jp

## 2.2 ChatGPT を用いたスキル獲得支援

2022年11月、アメリカ合衆国のAI研究開発企業の一つであるOpenAIは、自然言語処理のモデルであるGPTを搭載したチャットボットサービスChatGPTを公開した[3].

ChatGPTとは人工知能の一種で大規模言語、すなわち数十億単語以上のテキストデータを学習して、自然言語処理や対話を行うことができるシステムである[4]. ChatGPTにおいて特筆すべきはその学習するパラメータ数の膨大さにある。ChatGPTは生成AIの一種であるが、生成AIのニューラルネットワークは深層学習により学習が行われる。その点でChatGPTは先述の通り、大量のパラメータを学習することにより、システムの自由度が増し、処理能力が飛躍的に向上した[5]. これにより使用者はこれにプロンプトと呼ばれるテキスト入力を与えることで、恰も人間が対応したかのような返答を受け取ることができる。そして、使用者から見た大きな特徴は、ChatGPTからの返答が、文法や意味の一貫性のある程度保ちながら自然な形で出力され、また様々なタスクに対応できる汎用性を持つということである[6]. その有用性から、これを利用した様々なサービスが開発されている。

そのような、ChatGPTを用いて、公務員試験の受験対策支援を行うことを目的とする研究がある[7]. この研究の結果、「時事的な問題」について、適切に作問と解説が可能であることが示された。

このようにChatGPTを活用することで、人間側の能力を向上させようとする取り組みが存在しており、本研究も同様の観点により行ったものである。しかしながら、ChatGPT等、生成AIを用いて質問力を向上させるということを目的とした取り組みについては見受けられなかった。したがって、本研究に関して、新規性のあるものであると言える。

## 3 質問力向上支援システム

本章では、質問力向上支援システムについて述べる。質問力向上支援システムの構成を図1に示す。

システムの処理の流れとしては、まず、ユーザの会話相手となる人物像を生成する。その後、ユーザはその人物の情報を引き出すための質問を行う。システムは人物像情報を基にChatGPTにより質問に対する回答を生成する。その後、その質問に対する評価を独自で設定した「達成度」と「好感度」という指標をもとに、同じくChatGPTを用いて行い、最終的にそれらの判定結果や、会話のヒントとなる情報をユーザにフィードバックする。このやり取りを全ての人物の情報を引き出すまで、あるいは一定回数繰り返す。

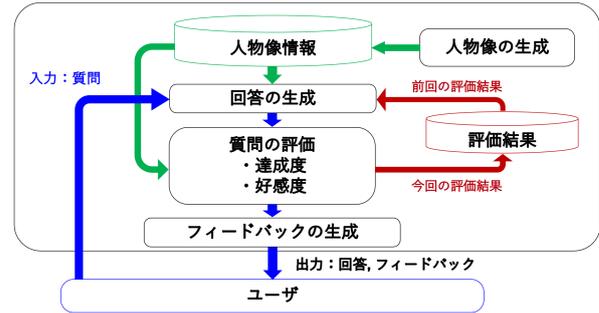


図 1: 質問力向上支援システムの構成

表 1: ChatGPT の API で用意されているプロンプトの役割

role	content
system	ChatGPT の回答の答え方に関する条件
assistant	ChatGPT の過去の回答
user	ChatGPT に与える質問

### 3.1 ChatGPT の利用

本節では、質問力向上支援システムにおける ChatGPT の利用方法について述べる。

ChatGPT には、プログラム上で ChatGPT の出力を呼び出すための API(Application Programming Interface) が用意されており、表 1 の役割 (role) と内容 (content) を組み合わせた入力 (プロンプト) を与えることで、出力としての回答を得ることができる。例えば、ChatGPT に単に質問するだけであれば、user の role を使って、content に質問を与えれば回答を得られる。また、ChatGPT にこれまでの質問や回答の履歴を元に回答をしてほしい場合には、assistant と user を使って過去の質問とその回答の系列を入力として渡した上で、質問の回答を求めることもできる。これらに加えて、ChatGPT の回答時に、どのような観点で回答してほしいか、どのような形式で出力してほしいか、などの答え方に関する条件を system の role に設定して渡しておくことで、期待する結果を得やすくすることができる。

### 3.2 人物像の生成

本節では、質問力向上支援システムにおける、ユーザの会話相手となる人物の人物像情報の生成手法について述べる。

本システムでは、ユーザの会話相手となる架空の人物像を生成する。人物像の情報は十分な量かつ一般的な種類として、「名前、年齢、性別、性格、職業、背景、目標、関心・趣味、悩み、外見」の 10 項目で定義され、基本的な情報からより詳細な情報まで最大 3 段階で深堀

表 2: 人物像の項目内の情報の深さの各内容

情報の深さ	内容
深度 1	基本的な情報 1 文
深度 2	普段の様子や一般的情報 1 文
深度 3	簡単なエピソード 2 文

表 3: 情報の解放の条件

情報の深度の変化	変化条件
1 から 2	「引き出せた深度 1 の人物像情報の項目の数が 4 個以上」かつ「好感度 4 以上」
2 から 3	「引き出せた深度 2 の人物像情報の項目の数が 4 個以上」かつ「好感度 8 以上」

して設定している。なお、各段階の内容をまとめたものを表 2 に示す。なお、各深度の基準は自己開示の深さを測定する自己開示尺度 [8] を参考にしつつ、本システムに合うように再構築した結果である。

### 3.3 質問の入力と回答の生成

本節では、質問力向上支援システムにおける、ユーザからの質問の入力と、それに対する回答の生成の手法について述べる。

本システムにおいて、ユーザはシステム側で生成した人物像の情報を持つ人物に対して質問を行う。3.2 節により人物像の情報が生成された後、質問を入力し、その後、人物の全ての情報を引き出すまで、あるいは一定回数質問を繰り返すまで、質問の入力を行う。

また、本システムでは、ユーザからの入力(質問)に対して、過去のやり取りも踏まえつつ、人物像の情報から適切な回答を ChatGPT により生成している。さらに、回答に用いる人物像の情報を諸条件により制限している。各種条件を満たした場合、より深い深度の情報を人物像の情報として含めることで、より多くの情報を引き出すことができるようにしている。例えば、システム起動直後は深度 1 の情報のみを人物像情報として取り入れ、諸条件をクリアした場合、人物像情報に深度 2 の情報を加えるようにしている。そこで、ここでは、より多くの情報を引き出せるようになることを「情報の解放」と呼ぶ。具体的には、表 3 の条件に従うものとする。

### 3.4 質問の評価：達成度と好感度

本節では、質問力向上支援システムで行われる、入力された質問に対する評価手法について述べる。ここで評価手法は「達成度の評価」、「好感度の判定」を指し、それぞれ本システムの目的に沿う形で設けられている。

表 4: 情報の深度ごとの付加される達成度の点数

情報の深度	一項目当たりの点数(点)
深度 1	5
深度 2	5.5
深度 3	6

表 5: 好感度の判定を行うプロンプト (一部抜粋)

これまでに行われた質問とそれに対する回答から、質問する人の質問の特徴を考え、質問する人に向けて、より良い質問ができるようなアドバイスを出力してください。「質問文」が丁寧さや自然さが感じられる好意的な文である場合は点数を 1、逆に「質問文」に粗雑さや乱暴さが感じられる敵対的な文である場合は点数を -1 とし、好意的でも敵対的でもどちらでもない判断できる場合は 0 を出力し、またその点数になった理由を出力してください。

#### 3.4.1 達成度の評価

本項では、質問力向上支援システムにおける、ユーザの質問が人物像情報を引き出したかどうか判定する手法について述べる。

質問力を評価する手法として QQTest という手法が開発、提案されている [1] が、これは文章を読み、それに対する質問を評価する手法であるため、本研究の方向性とは少々異なる。したがって、本システムに関しては新たに採点基準を設ける必要がある。そこで、本システムではなるべく定量的に評価を行うため、「最終的に、質問によりいくつ人物像情報を開示することができたか」という視点で点数化を行う。そこで、入力された質問とその回答により、新しくどの項目の情報を引き出すことができたかを判定する。

本システムでは、引き出された人物像の情報の量と深度により最終的な点数を算出している。ここで、付加される点数の基準をまとめたものを表 4 に示す。

3.2 節にて述べた通り、本システムでは、10 個の項目を持つ人物像の情報が生成されるが、引き出すことにより点数が付けられる項目は「職業、背景、目標、関心・趣味、悩み、外見」の 6 項目とし、その上で、表 4 のような基準で 100 点満点で点数を付けている。なお、表 4 に従い採点すると、全ての情報を引き出した場合 99 点となるため、その場合のみ例外で 100 点とし、それ以外は表 4 に従った採点を行っている。

#### 3.4.2 好感度の判定

本項では、質問力向上支援における、好感度の判定手法について述べる。

今回、対人におけるコミュニケーションが前提となっているため、現実的に考えれば、会話相手から情報を引き出す場合、会話の中で相手に不快感を抱かせないよ

表 6: アシストコメント生成用プロンプトと出力例 (一部抜粋)

<p>これまでに行われた質問と「すでに引き出せている情報」から質問相手からまだ得られていない人物像情報を引き出すために行ったら良い質問の仕方や質問する人がそのために目を向けるべきポイントについてのアドバイスを出力してください。</p>
<p>(生成された出力例) 悩みについてはその職業で生じやすい課題やストレスと、高齢者医療への貢献を目標としている点を結びつけて質問しましょう。打ち克とうとする障害に関する問いかけを行うことで、内面的な悩みに迫ることができません。</p>

表 7: 会話の総括生成用プロンプトと出力例 (一部抜粋)

<p>これまでに行われた質問とそれに対する回答から、質問する人の質問の特徴を考え、質問する人に向けて、より良い質問ができるようなアドバイスを出力してください。</p>
<p>(生成された出力例) アドバイス：質問の範囲をもっと広げるか、既に得られた情報に基づいてより深く掘り下げる新しい質問を試みてください。たとえば、医師としてのキャリアを深掘りするため、どのような困難に直面し克服したのか、または地域コミュニティとの関わりについて詳しく尋ねることができます。</p>

うにしつつ、相手の情報を引き出す必要が出てくる。そのため、本システムでは「好感度」という指標を導入している。好感度の判定を行うプロンプトを表 5 に示す。このプロンプトによって得られる結果を加算することで好感度の評価値とする。3.3 節により先述の情報の解放の条件として用いることで、より現実的なコミュニケーションを模した評価を可能としている。

### 3.5 フィードバックの生成

本節では、質問力向上支援システムにおいて行われる、ユーザへのフィードバックの生成手法について述べる。

本システムでは、ユーザの質問力向上に寄与するため、適切なフィードバックを行うことが求められる。そこで、本システムでは、システムとの会話中には「アシストコメント」を、会話終了後には「会話の総括」を生成する。

まず、本システムにおける「アシストコメント」とは、ユーザとシステムとの会話をアシストすることで情報を引き出すのに役立つ情報を表す。これは表 6 のプロンプトにより生成しており、一定回数の質問で、情報の解放が行われない場合にユーザに提供している。

次に、本システムにおける「会話の総括」とは、会話が終了された場合に今までの会話から ChatGPT により生成する、次の会話や現実での会話に有用な情報を

表 8: 達成度の評価値 (被験者平均)

	提案システム	比較システム
人物像 1	77.1	82.4
人物像 2	61.6	72.6
人物像 3	78.0	68.4

表す。なお、そのアドバイスを生成する際に ChatGPT に与えるプロンプトは表 7 のようになっている。

## 4 質問力向上支援システムの評価実験

本章では、質問力向上支援システムの評価実験について述べる。

本実験の目的は、質問力向上支援システムを使用したオンライン上のコミュニケーションにおいて、システムが与えるアシスト機能の有無が被験者のコミュニケーションに及ぼす影響を評価し、アシスト機能の有効性についての洞察を得ることである。

### 4.1 実験の手順

本節では、質問力向上支援システムの評価実験の手順について述べる。

本実験では、14 名の大学生、大学院生を被験者として、7 名に提案システムをそのまま、7 名に「アシストコメント」と「会話の総括」のフィードバックがない比較システムを利用してもらい、3 名の架空の人物の情報を引き出してもらった。この 3 名は事前に本システムを用いて生成していた人物から選んだものであり、年齢、性別、職業等の情報ができるだけ幅広くなるような 3 名を採用した。

### 4.2 実験結果と考察

本節では、検証実験の結果と考察について述べる。

#### 4.2.1 各グループの比較

本項では、各 7 名のグループの実験結果を比較した結果とそれに対する考察について述べる。

表 8 に各グループ、各人物像の達成度の評価値を平均した結果をまとめた。

表 8 より、各状況において、平均 60 点以上取れていることが分かる。このことから、システム上で質問により人物の情報を引き出すことが可能であり、質問力向上支援システムとしての基盤は完成できていると考えら

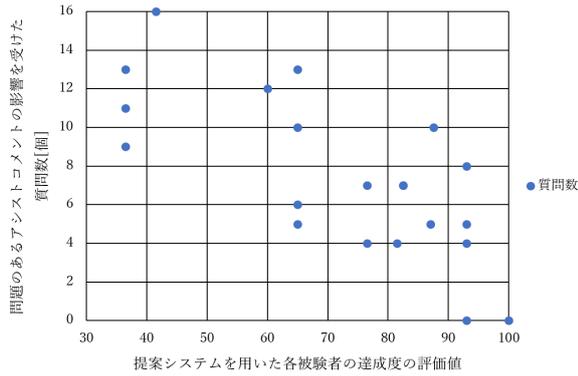


図 2: 問題のあるアシストコメントの影響を受けた質問と評価値の関係

れる。一方、提案システムの結果と比較システムの結果を比較すると、人物像 3 については、提案システムの方が点数が高くなったが、人物像 1, 2 の場合は比較システムの方が点数が低くなったことが読み取れる。本システムにおいて、本来、フィードバックがある方が、点数取得には有利と考えられる。したがって、各システムの点数の差にばらつきがあるということは、フィードバックの効果が必ずしも上がらなかったということが分かる。フィードバックの効果が必ずしも上がらなかった原因としては、システムからのフィードバックが点数取得に不適切な方向にユーザーを誘導してしまったことが挙げられる。特に、人物像の情報の中に存在しない内容をフィードバックしてしまうケースが多く、そのようなフィードバックが反映されたと考えられる質問の数と提案システムを用いた各被験者の達成度の評価値の関係を表したグラフを図 2 に示す。

図 2 からわかる通り、これらの関係は負の相関関係があり、実際に相関係数を求めると、 $-0.73$  となる。これにより、人物像の情報の中に存在しない内容のフィードバックが点数取得の妨げとなった主な原因として考えられる。

#### 4.2.2 人物像間の評価値のばらつき

本項では、表 8 によりみられた人物像間の評価値のばらつきについて結果をまとめ、考察を行う。

図 3 に各人物像における情報の項目(職業、背景等)をそれぞれ提案システムを用いた被験者が何名引き出すことができたかを示すグラフを示す。なお、各項目の識別方法として、例えば、深さ 1 の職業についての情報は「職業 1」、深さ 3 の外見についての情報は「外見 3」という様に命名している。

図 3 より、項目によっては、人物像によって引き出せた人数に大きくばらつきがあることがわかる。そこで、各項目について、引き出せた人数が最大であった人物像

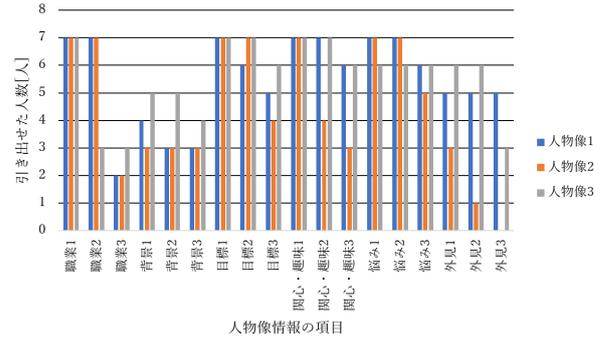


図 3: 各人物像における情報の項目を引き出すことのできた提案システムを用いた被験者の人数

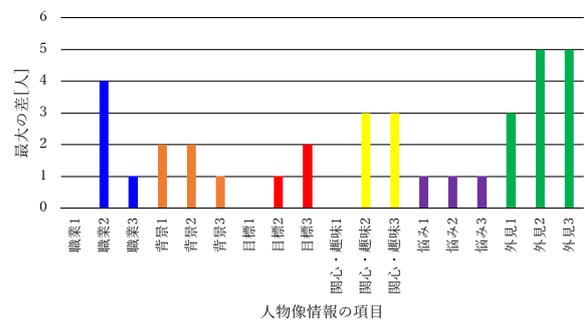


図 4: 各項目における引き出せた人数が最大であった人物像の際の人数と引き出せた人数が最小であった人物像の際の人数の差

の際の人数から、引き出せた人数が最小であった人物像の際の人数を引いた数について示したグラフを図 4 に示す。この数について具体的に説明すると、例えば、図 3 の外見 1 に注目すると、人物像 3 の本項目を引き出せた人数が 6 人で最大であり、人物像 2 の本項目を引き出せた人数が 3 人で最小のため、求める数はその差を取り、3 となる。

図 4 より、特に、職業 2、外見 2、外見 3 の人物像間のばらつきが大きいということが分かる。この要因としては、引き出せない項目の内容は具体性が欠けており、アシストコメントによる補助は基本的に具体的な内容を引き出すように要求するものが多かったことが考えられる。したがって、人物像によってばらつきが大きい項目は人物像によっては具体性の欠けた内容になっているものとそうでないものに分かれたためであると言える。これの根本的な問題としては、やはりフィードバックの内容が偏っていたことが挙げられる。

## 5 おわりに

本研究では、ユーザの質問力を向上させるための、質問力向上支援システムを構築した。

また、本研究のシステムを使用したオンライン上のコミュニケーションにおいて、システムが与えるフィードバックの有無が被験者のコミュニケーションに及ぼす影響を評価するための実験を行った。結果としては、各情報において達成度の評価値が平均 60 点以上となっており、質問力向上支援システムとしての基盤は完成できていると考えられる。一方、本システム上でのフィードバックについて様々な改善点が浮き彫りとなった。具体的には、「フィードバックが人物像の情報に存在しない情報について言及してしまう」、「人物像の情報が具体性に欠けている場合、そこについてフィードバックでフォローできていない」、「フィードバックによる情報供給が過多であり、質問のポイントを絞れない」などの改善点が挙げられた。このような改善点から、本システムにおける適切なフィードバックには、「まだ引き出すことのできていない情報を引き出しやすくすることのできる内容であること」、「実際のコミュニケーションにおいて有用な内容であること」、「分量が過多にならず、質問のポイントが絞れること」の三点が求められることが分かった。

今後の課題としては、今回の実験を受けての本システムの改良と、本システム上での達成度の評価値と外部指標の相関関係について調査することで、本システムにおける評価値の妥当性について検討することが挙げられる。

## 参考文献

- [1] 池田文人, 岩間徳兼, 成田秀夫: QQTest の開発と試行 - 探求学習の成果を測定するための指標づくりを目指して -, 大学入試研究ジャーナル, Vol.32, pp.35 - 42, (2022)
- [2] 池田文人: どのように質問力を評価するか?: メタモデルに基づいた評価の信頼性に関する検証, 高等教育ジャーナル: 高等教育と生涯学習, Vol.25, pp.55 - 61, (2018)
- [3] ChatGPT :  
(URL)<https://chat.openai.com/> (2024 年 1 月 29 日確認)
- [4] 矢守恭子, 杉山雅英: 生成 AI のサポートによるコード作成, 情報学研究, Vol.32, pp.1 - 3, (2023)
- [5] 平田隆幸: ChatGPT と工学部教育 - 生成 AI によって大学教育は変わるのか -, 福井大学 学術研究院工学系部門 研究報告, Vol.72, pp.17 - 25, (2023)
- [6] 松田翔太: ChatGPT を活用した情報教育の提案, 第一工科大学研究報告, Vol.35, pp.37 - 40, (2022)
- [7] 古矢一翔, 林浩一: 公務員試験対策のための対話型生成 AI の活用, 情報教育シンポジウム, pp.204 - 209, (2023)
- [8] 丹羽空, 丸野俊一: 自己開示の深さを測定する尺度の開発, パーソナリティ研究, Vol.18, No.3, pp.196 - 209, (2010)

# コンテキストに適した英語表現学習支援のための 映像作品の字幕データのセンチメント分析と可視化

## Visualizing Sentiment Analysis on Subtitles in Dramas for Learning English Conversation Complying with Context

北畑 由莉香<sup>1</sup> 山西 良典<sup>1\*</sup>  
Yurika Kitahata<sup>1</sup> Ryosuke Yamanishi<sup>1</sup>

<sup>1</sup> 関西大学総合情報学部

<sup>1</sup> Faculty of Informatics, Kansai University

**Abstract:** This paper describes the sentiment analysis for English subtitles of dramas, aimed at enhancing the learning of conversational expressions within their contextual use. Traditional educational materials may not always encompass expressions and vocabulary encountered in daily interactions among friends, including informal and colloquial language ( occasionally, even profanities) which can be essential in making friendships. Focusing specifically on drama subtitles, we conducted a detailed sentiment analysis to visualize the context and construct of sentences where specific words are used. This paper discusses the efficacy of such visualizations in facilitating the acquisition of conversational language skills.

### 1 はじめに

急速に進む国際化やインターネットの普及に影響され、言語・文化が異なる人とのコミュニケーションの機会は増加傾向にある。特に、国際コミュニケーションの手段として世界的に使用され、影響力を拡げている英語 [Quirk 85] を用いたコミュニケーション能力の獲得は、非ネイティブ話者にとって共通の課題である。生涯にわたって様々な場面で必要とされるその能力を身に着けるため、英語能力の獲得への需要はますます高まるものと考えられる。

Computer Assisted Language Learning (CALL) は、コンピュータによる言語習得の支援を目指した研究分野 [Graham 01] であり、様々な技術や事例が報告されている [Hughes 23, Hamad 17]。英語能力の習得を支援するために開発された CALL システムはこれまでも数多く存在するが、それらの研究は単語学習 [内田 15] や構文学習 [阿部 21]、発音学習 [葉田 01]、リスニング学習 [Winke 10, 九鬼 19] を支援し、語彙や文法規則、リスニング能力などの基礎能力の習得と定着を支援するものが多い。しかし、英語でのコミュニケーションの実現場には、これらの基礎的な知識や能力だけでなく、「状況に応じて適切な表現を選択する能力」が求められる。この能力は、主に「聞く」「読む」「話す」「書

く」などの実践的コミュニケーションを繰り返し経験することによって身につけられる [山西 23] が、実践的コミュニケーションを経験するための課題も多い。非ネイティブ話者が経験していく中では、自身がコミュニケーションの中で用いた表現 (単語、熟語、フレーズ) が、その「場面」「状況」「背景」「対象」などと照らし合わせて適切か否かを学習者自身からは判断不可能である。例えば、コミュニケーションを取る相手が家族の場合、友人の場合、目上の人物の場合では、同内容であっても異なる表現を用いて会話することが容易に想定される。また、同一の相手であっても、空間 (例えば、フォーマル/インフォーマルの違い) の違いによる言葉の使い分けや文脈次第で大きく意味が変化する言葉を学ぶ必要がある。既存の学習教材の多くは、文法的に正しいとされる表現や広く一般的に使用可能な表現を学ぶうえでは有用である。一方で、口語で頻出するコミュニケーションの場面や状況、背景、対象に影響されて意味や印象が変化する表現は少なく、様々な文脈に合わせた表現を学ぶ手段が少ない現状にある。

多種多様なパーソナリティの話者が背景設定・関係性・場面状況・文脈の中でコミュニケーションを行う様子を観測可能なエンタテインメントとして映像作品 (映画・TV番組) の存在 [Birulés-Muntané 16, 植松 04] がある。映像作品で出現するセリフは、文脈に適應した言語表現の学習用教材としての活用 [Shan 21, 九鬼 19, 吉野 03, 湯舟 15] が考えられる。映像作品では字幕セ

\*連絡先： 関西大学総合情報学部  
〒 569-1095 大阪府高槻市霊仙寺町 2-1-1  
E-mail: {k940098,ryama}@kansai-u.ac.jp

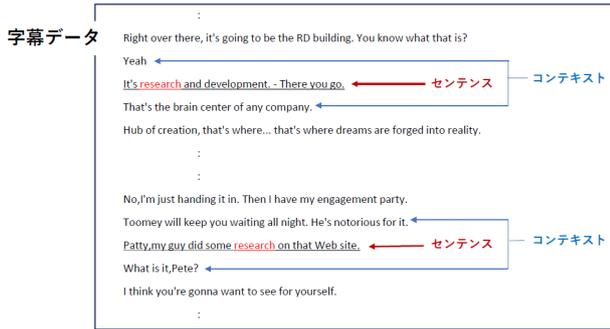


図 1: 分析センテンスと分析コンテキストの例

表 1: センチメント分析モデルの感情評価の精度確認の結果

		Model の出力ラベル		
		positive	neutral	negative
主観ラベル	positive	77	45	3
	neutral	30	287	46
	negative	1	41	109

リフをテキスト化した字幕が用意されていることが多い。この字幕を分析することで、任意の単語が「どのような文脈のどのような発話の中で使用されているのか」を可視化可能であると考えた。

本稿では、「字幕に登場する単語」と「その単語を含む発話」、「発話が行われた文脈」と「文脈と発話から感じ取れる感情」の4点に着目する。映像作品の英語字幕に対してセンチメント分析を適用して可視化可能なシステムを構築した。各感情の文脈での単語出現頻度の多寡を視認可能にすることで、文脈を考慮した英単語の学習支援の可能性を検討する。

## 2 分析手法と可視化方法

任意の単語が「どのような文脈のどのような発話の中で使用されているのか」を分析するために、まず文脈と発話を定義する。任意の単語が含まれる一文を発話文 (i.e., 分析センテンス) とし、センテンスの前後それぞれ1文を発話が行われた文脈 (i.e., 分析コンテキスト) とする。例えば、図1に示すような字幕データ (発表年:2007, imdbId:1003353, 字幕ファイル ID:3140984, 行:272-278, 506-510) で ‘research’ について分析する場合を考える。文中に ‘research’ が含まれている、3行目 “It’s research and development. - There you go.”, 8行目 “Patty, my guy did some research on that Web site.” の2文が分析センテンスとなる。このとき、2行目 “Yeah” と4行目 “That’s the brain center of any company.” は3行目に対応した分析コンテキスト、7行目 “Toomey will keep you waiting all night. He’s

表 2: 感情ラベルごとのセンチメント分析モデルの評価指標値 (小数点第二位以下四捨五入)

	再現率 (%)	適合率 (%)	F 値 (%)
positive	61.6	71.3	66.1
neutral	79.1	76.9	78.0
negative	72.2	69.0	70.6

notorious for it.” と9行目 “What is it, Pete?” は8行目の分析コンテキストとなる。

### 2.1 センチメント分析

分析センテンスと分析コンテキストそれぞれにセンチメント分析を行って、それぞれの極性を得る。得られる分析センテンスと分析コンテキストそれぞれの極性を組み合わせによって、単語の使用傾向を表現可能にする。

Loureiro らによって作成された、約12,400万の tweet によって英語のセンチメント分析用に調整された RoBERTa モデル [Loureiro 22] (以降センチメント分析モデルと呼称)<sup>1</sup> を使用する。センチメント分析モデルによって、入力されたテキストに対して positive, neutral, negative の3種類の感情それぞれの確信度が出力される。本稿では、確信度が最も高く出力された感情ラベルを入力テキストの感情として扱う。

### 2.2 センチメント分析モデルの精度確認

tweet によって学習・構築されたセンチメント分析モデルによって字幕のセンチメント分析が可能であるかについて事前検証を行った。既存研究 [徐 22] を参考にし、2022年8月時点の Netflix で「オリジナル言語」<sup>2</sup> が「英語」である映像作品の字幕から1,000会話文を用意した。この会話文集合から、ランダムに639文を抜き出して評価用のデータセットを作成した。評価用データセットに対して、第1著者の主観評価で各文が3種の感情のうちどの感情を示すかラベル付けを行った。一方で、これらの文に対してセンチメント分析モデルを用いた極性評価も行った。

表1に、主観とセンチメント分析モデルの評価結果の比較を示す。表2に、感情ラベルごとのセンチメント分析モデルの評価指標値 (再現率, 適合率, F 値) を示す。文献 [Loureiro 22] では、TweetEval ベンチマーク [Barbieri 20] を用いたモデルの精度評価が行われており、センチメント分析タスクに対し、マクロ平均再現率が73.7%であった。ここで、クラス数と*i*番目の

<sup>1</sup><https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment-latest> (2023年12月1日確認)

<sup>2</sup>その映像作品が制作された言語



図 2: 単語入力画面

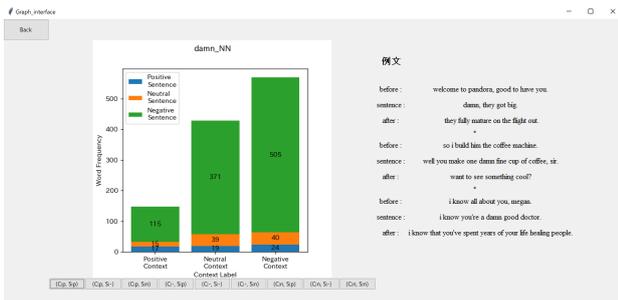


図 3: グラフ・用例提示画面

クラスの再現率をそれぞれを  $n$  と  $(x_{Recall})_i$  とするとマクロ平均再現率は下式で求められる。

$$x_i = \frac{1}{n} \sum_{i=1}^n (x_{Recall})_i. \quad (1)$$

表 2 と数式 1 から、本稿でのマクロ平均再現率は 70.9% となり、文献 [Loureiro 22] と同程度の値が得られた。このことから、本稿で用いるセンチメント分析モデルは、字幕の会話文に対しても妥当な感情推定結果を出力可能であることが示唆された。

## 2.3 可視化方法

提案システムは、図 2 に示す「単語入力画面」と図 3 に示す「グラフ・用例提示画面」の 2 画面から構成される。ユーザの入力や操作に基づいて 2 画面を切り替え、単語入力画面でユーザが入力した単語のセンチメント分析の結果を示すグラフと用例を提示する。

単語入力画面(図 2)では、画面上部中央の入力フォームに英単語を入力し、入力フォーム右下の「check」ボタンを押下する。フォームへ入力された英単語(品詞別)のセンチメント分析結果のデータがシステム内に存在するか走査する。走査の結果、結果データが存在した品詞に対応するボタンのみが有効化される。ユーザが、有効化された任意の品詞ボタンを押下することで、グラフ・用例提示画面に遷移する。

単語入力画面で有効化された任意の品詞ボタンを押下することで、グラフ・用例提示画面(図 3)に遷移する。グラフ・用例提示画面ではまず、押下された品詞の単語が登場するセンテンスとそのコンテキストに対するセンチメント分析結果を可視化したグラフ(図 3 中左)とグラフ下部に 9 つのボタンが表示される。ここで、9 つのボタンはコンテキスト(C)とセンテンス(S)の極性(positive:p, neutral:-, negative:n)組み合わせに対応する。ユーザがボタンを押下すると、ボタンが示す極性の組み合わせに対応した用例がランダムに 3 件選出され、表示される。

## 3 分析対象とする字幕データ

本稿では、Lison らが構築した、OpenSubtitle コーパス [Lison 16, Tiedemann 16] を対象として分析する。OpenSubtitles<sup>3</sup>は、ユーザーによって作成・投稿された映画や TV 番組の字幕データを集積し、60 以上の言語、300 万以上の字幕データから成る大規模な字幕データベースを提供する Web サイトである。OpenSubtitles コーパスは、OpenSubtitles で提供されている字幕データに対して、字幕の誤りの訂正、文章の整理などが処理されている。さらに、識別子などのメタデータを付与した 62 言語の字幕データを、任意の 2 言語の文単位での対訳データの形式で得ることができるパラレルコーパスとなっている。

OpenSubtitles コーパスはデータを xces 形式や tmx 形式等で提供しており、本稿では HuggingFace 上<sup>4</sup>の英語字幕と日本語字幕の対訳データ 810,000 件が含まれるファイルを対象とする。表 3 に、データセットの一部を例として示す。以降に示す字幕データは、全てこのファイルから引用する。

表 3 に示すように、データセットは ID、メタ情報、英語字幕文と日本語字幕文の対訳の 3 つの情報で構成される。このうち、英語字幕文と日本語字幕文の対訳から英語字幕文を抽出した。次に、英語字幕文に含まれるすべての英単語の種類・数の把握するため、英語字幕文の分かち書きと各単語の品詞の取得を行った。同一単語であっても、異なる品詞タグが付与されたものは別単語として扱い、単語と品詞のペアが重複を許さないようにしたところ 80,333 種類の英単語が得られた。このうち、アルファベット以外の文字列(数値、記号、英語以外の言語の文字)が含まれておらず、出現回数が上位 10% より多い単語を抽出した。さらに、これらの単語に表 4 に示す 11 種の品詞で絞り込みを行って 8,713 単語を分析対象および可視化用の単語として準備した。

<sup>3</sup><https://www.opensubtitles.org/>(2023 年 12 月 1 日確認)

<sup>4</sup>[https://huggingface.co/datasets/open\\_subtitles](https://huggingface.co/datasets/open_subtitles)(2023 年 12 月 1 日確認)

表 3: OpenSubtitles データセットの一部. ID は字幕固有の番号, メタ情報には公開年 (year), IMDb 識別子 (imdbId), 字幕ファイルの ID(subtitleId, 英語と日本語それぞれ), 訳文がそれぞれ字幕ファイルの何行目か (sentenceIds, 英語と日本語それぞれ) の情報を含む. 対訳は英語と日本語文の対訳.

ID	メタ情報	対訳
0	"{'year': 1920, 'imdbId': 10323, 'subtitleId': {'en': 5777654, 'ja': 5996100}, 'sentenceIds': {'en': array([1], dtype=uint32), 'ja': array([1], dtype=uint32)}}"	"{'en': '""THE CABINET OF DR. CALIGARI""', 'ja': 'カリガリ博士の小屋'"}
:	:	:
102	"{'year': 1924, 'imdbId': 15324, 'subtitleId': {'en': 5530518, 'ja': 5620547}, 'sentenceIds': {'en': array([2, 3], dtype=uint32), 'ja': array([2], dtype=uint32)}}"	"{'en': '""Don't try to do two things at once and expect to do justice to both. This is the story of a boy who tried it.""', 'ja': '""二兎を追う者は一兎をも得ず""'"}
103	"{'year': 1924, 'imdbId': 15324, 'subtitleId': {'en': 5530518, 'ja': 5620547}, 'sentenceIds': {'en': array([4], dtype=uint32), 'ja': array([3, 4, 5], dtype=uint32)}}"	"{'en': 'While employed as a moving picture operator is a small town... theater he was also studying to be a detective.', 'ja': 'しかし この物語の主人公はあえてそれに挑戦した 彼は町の映画館に勤める 映写技師であり また同時に 探偵修業中の 身でもあった'"}

表 4: 使用する品詞タグ (11 種) と対応する品詞名の一覧

品詞タグ	対応する品詞名
NN	名詞
NNS	名詞の複数形
JJ	形容詞
JJR	形容詞の比較級
JJS	形容詞の最上級
VB	動詞の原形
VBD	動詞の過去形
VBG	動名詞または現在分詞
VCN	動詞の過去分詞形
VBP	動詞の三人称単数以外の現在形
VBZ	動詞の三人称単数の現在形

## 4 字幕データの分析と可視化

3 節で用意した単語集合に対して, 単語のセンテンスとそのコンテキストに対しセンチメント分析を行った. 紙面の都合上, 興味深い結果が得られた damn(動詞の原形 (VB)) と incredible(形容詞 (JJ)) の可視化結果を示して考察する. 図 4 と図 5 に, それぞれ damn(動詞の原形 (VB)) と incredible(形容詞 (JJ)) が含まれるセンテンスとコンテキストのセンチメント分析の可視化結果を示す. 以下, コンテキストとセンテンスの組み合わせを示すために, 提案システムでのボタンの表示に合わせた (C:p, S:p) のような表記も用いる.

### 4.1 damn(動詞の原形 (VB)) のセンチメント分析結果

'damn' は, 一般的には「永遠に罰する、地獄に落とす、酷評する、けなす、ののしる、のろう、ちくしょう!」といった意味をもち<sup>5</sup>, negative な極性の単語と考えられる. 同図から, 'damn' は negative な会話の流れ

<sup>5</sup><https://ejje.weblio.jp/content/damn>(2023 年 11 月 24 日閲覧)

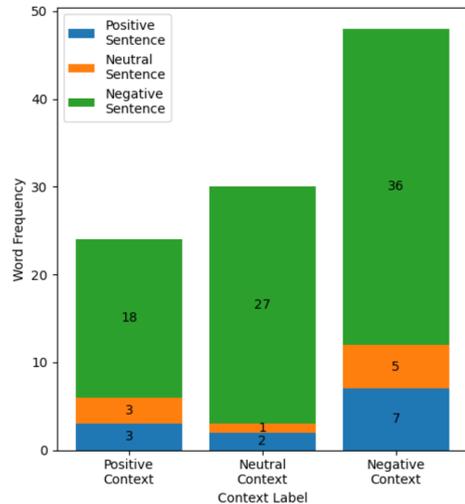


図 4: damn の用法の可視化結果

で登場することが多く, negative な発話で用いられる. 一方で, 'damn' は negative なコンテキストやセンテンスのみで用いられるわけではなく, 'damn' という単語そのものが持つ, negative なイメージと離れた positive なセンテンスやコンテキストでも用いられる.

表 5 にそれぞれ, damn が含まれるセンテンスで, コンテキストがそれぞれ positive, neutral, negative であるものの例を示す. 表中, センテンスの極性が negative((C:p, S:n), (C:-, S:n), (C:n, S:n)), もしくは極性の組み合わせが (C:n, S:-) となっている例では, 'damn' は主に非難や罵りの意味で用いられている. 一方で, コンテキストが positive, 極性の組み合わせが (C:-, S:p), (C:-, S:-), (C:n, S:p) となっている例では, 'damn' は後続する言葉を強調する目的で用いられており, 強い肯定や, 賛辞の強調を示している.

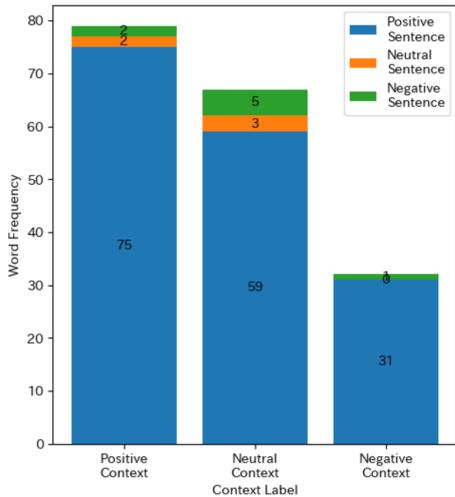


図 5: incredible の単語特徴可視化結果

negative なセンテンスやコンテキストに登場した場合, ‘damn’ は非難や罵りなどの negative な意味で用いられており, ‘damn’ という単語が一般的にもつ negative なイメージと一致する. 一方で, positive あるいは neutral なセンテンスやコンテキストで用いられる場合, ‘damn’ は強い肯定や賛辞の強調として用いられており, 一般的なイメージとは乖離した用法であると考えられる. 辞書の意味とは異なる用法をシステムの出力結果から一瞥して理解可能であると考えられる.

## 4.2 incredible(形容詞 (JJ)) のセンチメント分析結果

‘incredible’ は, 一般的には「信じられない, 信用できない」と「驚くべき, 非常な, 途方もない」といった意味をもつ単語<sup>6</sup>である. 図 5 より, ‘incredible’ は positive なコンテキスト, もしくはコンテキストを問わず positive なセンテンスで登場することが多く, neutral, negative なセンテンスへの登場は極端に少ない.

表 6 に, incredible が含まれるセンテンスで, コンテキストの極性がそれぞれ positive, neutral, negative であるものの一例を示す. 同表から, ‘incredible’ という単語が示す 2 種類の意味(「信じられない, 信用できない」と「驚くべき, 非常な, 途方もない」)は, コンテキストやセンテンスの極性によらず, どちらの極性でも用いられることがわかる. 例えば表 6 中, (C:p, S:p) のセンテンス “you did it darling, that was incredible.” と, (C:n, S:n) のセンテンスである “incredible. so sorry.” で, incredible は「信じられない」を示す言葉として用

いられている. これらの 2 例文で “incredible” は異なるニュアンスを示している. (C:p, S:p) の例では, 「(予想していたより良い結果を得られ)信じられない」という positive な反応である一方, (C:n, S:n) の例では「(期待外れの出来事が起き)信じられない」という negative な反応として ‘incredible’ が使われている. 同一の単語であるにもかかわらず, 異なる意味・感情として捉えられる龍としてセンテンス中の他の表現による影響が考えられる. (C:p, S:p) の例文中には ‘darling’ という親しい人物への呼びかけの言葉があり, コンテキストには “nice job sweetheart.” という恋人を褒める positive な文が含まれている. 一方で, (C:n, S:n) の例文中には, “so sorry” という謝罪の言葉が含まれている. これらのことから, ‘incredible’ は使用されるコンテキスト・センテンス中に登場する表現の極性に影響を受ける単語であることが理解できる.

以上の考察をまとめると, ‘incredible’ は登場するコンテキスト・センテンス中に登場する表現の極性に影響を受け, 示す意味の極性を変化させる単語である. また, ‘incredible’ は, positive なコンテキストやコンテキストを問わず positive なセンテンスで用いられることが多く, 一般的に positive な意味合いで用いられる単語であることがわかる.

## 5 おわりに

本研究の最終目標は, 映像作品の英語字幕を英語学習に活用することで, 英語表現学習における既存の課題を解決することにある. 本稿では, 「字幕に登場する単語」と「その単語を含む発話」, 「発話が行われた文脈」と「文脈と発話から感じ取れる感情」の 4 点に着目し, 映像作品の字幕に登場する任意の単語が「どのような文脈のどのような発話の中で使用されているのか」を可視化した. 出現傾向を示す積み立てグラフと具体的な用例を合わせて提示することで, 任意の単語の主な使われ方, 登場頻度は少ないが特徴的な使われ方の双方を理解することができる可能性が示唆された. 今後は, 日本語の対訳データを利用したコンテキストに応じた単語の使い分けを分析するほか, 提案システムによる学習効果の検証も行う.

## 謝辞

本稿では, 分析および表 3, 表 5, 表 6 での例示に OpenSubtitle コーパス [Lison 16, Tiedemann 16] のデータ ([https://huggingface.co/datasets/open\\_subtitles](https://huggingface.co/datasets/open_subtitles)) を引用させて頂いた. 本稿の執筆にあたって, 立命館大学情報理工学部の Jeremy White 教授に助言を頂いた. 記して謝意を表す.

<sup>6</sup><https://ejje.weblio.jp/content/incredible>(2023 年 11 月 24 日閲覧)

## 参考文献

- [Barbieri 20] Barbieri, F., et al.: TweetEval: Unified Benchmark and Comparative Evaluation for Tweet Classification, in *Findings of the Association for Computational Linguistics: EMNLP 2020*, pp. 1644 – 1650 (2020)
- [Birulés-Muntané 16] Birulés-Muntané, J. and Soto-Faraco, S.: Watching Subtitled Films Can Help Learning Foreign Languages, *PLOS ONE*, Vol. 11, No. 6, pp. 1 – 10 (2016)
- [Graham 01] Graham, D.: CALL (computer assisted language learning), <https://web-archive.southampton.ac.uk/www.llas.ac.uk/resources/gpg/61.html> (2001)
- [Hamad 17] Hamad, M.: Using WhatsApp to Enhance Students’ Learning of English Language “Experience to Share”, *Higher Education Studies*, Vol. 7, p. 74 (2017)
- [Hughes 23] Hughes, L.: Effects of Dynamically Computer-Mediated Communication and ‘Pseudo’ Communication on L2 Learning, *CALL-EJ*, Vol. 24, pp. 106–137 (2023)
- [Lison 16] Lison, P. and Tiedemann, J.: OpenSubtitles2016: Extracting Large Parallel Corpora from Movie and TV Subtitles, in *Proc. the Tenth Intl’ Conf. on Language Resources and Evaluation*, pp. 923 – 929 (2016)
- [Loureiro 22] Loureiro, D., et al.: TimeLMs: Diachronic Language Models from Twitter, *CoRR*, Vol. abs/2202.03829, (2022)
- [Quirk 85] Quirk, R. and H.G.Widdowson, : *English in the World: Teaching and Learning the Language and Literatures*, Cambridge University Press in association with The British Council (1985)
- [Shan 21] Shan, J., et al.: アニメの部分会話の難易度の推定による日本語のリスニング練習支援, *芸術科学会論文誌*, Vol. 20, No. 2, pp. 108–119 (2021)
- [Tiedemann 16] Tiedemann, J.: Finding Alternative Translations in a Large Corpus of Movie Subtitle, in *Proc. the Tenth Intl’ Conf. on Language Resources and Evaluation*, pp. 3518 – 3522 (2016)
- [Winke 10] Winke, P., et al.: The Effects of Captioning Videos Used for Foreign Language Listening Activities, *Language Learning and Technology*, Vol. 14, (2010)
- [阿部 21] 阿部 真由美, 向後 千春: 英語多聴学習での拡散的好奇心による学習行動と学習意欲への影響, *日本教育工学会論文誌*, Vol. 45, No. 1, pp. 127 – 134 (2021)
- [植松 04] 植松 茂男: DVD 映画教材利用時の英語字幕が英語学習に与える影響について, *メディア教育研究*, Vol. 1, No. 1, pp. 107 – 114 (2004)
- [内田 15] 内田 千尋, 橋田 光代: TypeSinger: 歌うように“弾く”歌詞タイピングゲーム, *エンタテインメントコンピューティングシンポジウム 2015 論文集*, Vol. 2015, pp. 247–250 (2015)
- [九鬼 19] 九鬼 慧太 他: 映画を用いた英単語学習における話者追従字幕がコンテンツの記憶保持に与える影響, *電子情報通信学会技術研究報告*, Vol. 119, No. 190, pp. 79 – 84 (2019)
- [徐 22] 徐 敏徹: ;研究ノート;OTT サービスを利用したパラレルコーパスの構築方法, *京都大学言語学研究*, Vol. 41, (2022)
- [葉田 01] 葉田 善章 他: Viclle: 会話映像の添削による語学学習支援システム, *情報処理学会論文誌*, Vol. 42, No. 6, pp. 1412 – 1423 (2001)
- [山西 23] 山西 良典 他: 英会話はハートがすべてなんで!? キャンユースピークイングリッシュ?, *エンタテインメントコンピューティングシンポジウム 2023 論文集*, pp. 201 – 207 (2023)
- [湯舟 15] 湯舟 英一, 田淵 龍二: 映画場面検索サイト Seleaf を利用した授業の学習効果, *Language education & technology*, No. 52, pp. 389 – 410 (2015)
- [吉野 03] 吉野 志保: 英語学習に効果的な字幕提示タイミングの検討 (;特集; 第二言語学習とその支援に関する教育工学研究), *日本教育工学雑誌*, Vol. 27, No. 3, pp. 237 – 246 (2003)

表 5: damn の含まれるセンテンスの一例

(C: コンテキストの極性, S: センテンスの極性)	センテンス (括弧内はデータセットから抽出した, センテンスに対応している日本語訳)	センテンスと対応するコンテキスト (上: センテンスの前の文, 下: センテンスの後の文)
(C:p, S:p)	feels so warm, i can't help it. i want you. <b>damn</b> . (温かい体だ 君が欲しくてたまらない)	don't you know i couldn't lose you? (君が大切だから...) i love the way you smell. i love you, and i always have. i want to tell you... (ずっと好きだったと 打ち明けたいよ)
(C:p, S:-)	<b>damn</b> the meaning, i'll memorize it. (意味はいいんだ 丸暗記するから...)	our director has unceasingly served... 'served' means... (学長は ここで奇跡を成し遂げた この言葉の意味は...) chatur was the introductory speaker at the teachers' day function. (チャートルは 先生の日の スピーチに選ばれた)
(C:p, S:n)	stay in there, <b>damn it!</b> (そこから出てくるな!)	don't you dare follow us! (2度と後を追ってくるな!) thanks again for the money, clothes and everything. (お金から服から 何から何までありがとう)
(C:-, S:p)	you're <b>damn right</b> you're glad. (あなたの心遣いが嬉しいわ)	well, i'm glad i'm not working for you. (嬉しいよ 君が僕のために働いてなくて) so, do i have to smack you in the mouth, lieutenant? (それで 僕は口移しで 薬を飲ませればいいのか?)
(C:-, S:-)	doc, what's going on here, <b>damn it!</b> everybody. y'all (博士, 何が起こってる?)	30 minutes to decontamination. (汚染浄化まで 30分) get your stuff and let's go! (みんな急ぐんだ さあ!)
(C:-, S:n)	see how polacks have worked on me? <b>damn the basterds!</b> (俺は地主どもにやられて こんな有様だからな)	i've sent her for some home brew. (酒を買いに出したんだ) you'd better tell me. (聞きたい事があるんだけど)
(C:n, S:p)	it'd <b>damn well better</b> . (一見させるさ)	can it copy dvds? (見れるか?) that's not how it opens! (ここはだめだ ティビ)
(C:n, S:-)	i'm a cop, <b>damn it</b> . i'll show you my identification. (私は警察だ ID も持っている)	why should i trust you? you could be the maniac. (信用できないわ あなたが異常者かもしれない) here. look, can you see it? (ほら 見えるか?)
(C:n, S:n)	<b>damn</b> the day i met you, you <b>basterd!</b> (あんたと結婚したのが 間違いよ)	go! and may you rot there! (いいわゾーンに行行って 死になさい) god himself has cursed you by giving you this child! (だから呪われた子供が 生まれてー)

表 6: incredible の含まれるセンテンスの一例

(C: コンテキストの極性, S: センテンスの極性)	センテンス (括弧内はデータセットから抽出した, センテンスに対応している日本語訳)	センテンスと対応するコンテキスト (上: センテンスの前の文, 下: センテンスの後の文)
(C:p, S:p)	you did it darling. that was <b>incredible</b> . (良くやったわ 信じられない)	katniss everdeen, the girl on fire! (カットニス・エヴァディーン ”炎の少女”でした！) nice job sweetheart. (良くやったな)
(C:p, S:-)	<b>incredible?</b> what is? ("ものすごいこと"って どのようなこと?)	what would you do, if i became incredible? (そしたらどうする? あたしが ものすごいことになっちゃったら) you know, everything! (だから”何もかも”だよ)
(C:p, S:n)	i don't like surprises. i have an <b>incredible</b> ... (いやよ 驚くなんて)	what? no. (何?) trust me, you're gonna love this. (信じて きっと気に入るよ)
(C:-, S:p)	being with him was an <b>incredible</b> adventure. (彼という事は 素晴らしい冒険だった)	he saw things between lines of text that nobody else could. (書物の行間を 読む人よ 誰も出来ないわ) so you did love him? (彼を愛していたと?)
(C:-, S:-)	temperatures are rising with <b>incredible</b> velocity in hot zones across the globe. (全地上に渡って 地温が 急上昇しています)	it's not just yellowstone, sir. (イエローストーンだけではないのです) are you absolutely sure about this, adrian? (確信を持っているのか? エイドリアン)
(C:-, S:n)	what's <b>incredible</b> , though, is how spectacularly ignorant he is (「驚くべき事に特定の分野に ついては全く無知なのだ」)	! sherlock sees through everything and everyone in seconds. (「シャーロックは一瞬で どこの誰だかを見抜くが」) about some things. hang on, i didn't mean that... (待ってくれ 変な意味じゃ...)
(C:n, S:p)	what is <b>incredible</b> is i just finished the mazethis morning, and i'm not gonnateach her how to run it until hour from now. (何がすごいかって、今朝この迷路を 完成したばかり なんだよ 数時間後に道を教える 予定だったんだ)	i'm sorry. how is a rat running through a bloody maze so incredible? (迷路をねずみが抜けれるのが なぜそんなにすごいって言うんだ?) s so so you you sent herto the future. (君がこいつを未来へ 送ったって言うのか?)
(C:n, S:-)	なし	なし
(C:n, S:n)	<b>incredible</b> . so sorry. (信じられん ー 申し訳ございません)	this is not a flee market. (ここは蚤の市じゃない) i have to make a phone call. excuse me. (ちょっと電話してくる、失礼)

# ChatGPTを用いた系統的学習の復習支援システム

## Review Support System for Systematic Learning using ChatGPT

中村 和哉<sup>1\*</sup> 砂山 渡<sup>1</sup> 服部 峻<sup>1</sup>  
Kazuya Nakamura<sup>1</sup> Wataru Sunayama<sup>1</sup> Shun Hattori<sup>1</sup>

<sup>1</sup> 滋賀県立大学 工学部

<sup>1</sup> School of Engineering, The University of Shiga Prefecture

**Abstract:** In recent years, the use of ChatGPT has been gaining momentum, and its application to education is being considered. In fields that require systematic learning that consists of multiple units, it is difficult for students to learn new content if they forget what they have learned. In fields that require systematic learning consisting of multiple units, students may not be able to learn new content if they forget what they have learned once. In this study, therefore, we have developed a system that allows students to obtain explanations of keywords that they do not fully understand, as well as to suggest units to review. The system also suggests units that should be reviewed.

## 1 はじめに

近年、社会の情報化が進んでいる中で ChatGPT が急激に世の中に普及し始めている。そこで「教育」に着目して、授業やインターネットで学んだ内容の中で、十分に理解できなかった箇所をなくしたいという課題を ChatGPT で解決できないかと考えた。

そのための解決策として、わからない箇所を ChatGPT に質問して教えてもらえるようにすること、また、質問した内容から推測される、復習が必要な単元を ChatGPT に提示してもらうことを考えた。

本研究では、上記の2点を実現する ChatGPT を用いたシステムを提案する。

ChatGPT を用いたシステム構築によって、学習分野を既存の教育分野に限ることなく、就職や結婚といった、多角的な情報を学んだ上で実施していくことが必要な分野を含めて学習可能にする。加えて、複数の単元からなる分野を学ぶ際に、理解が十分でなく欠けている知識があると、学習や理解が進まないといった事態が起こるため、復習が必要な単元を推定し、ユーザに提示することで系統的な学習を支援する。

また本研究で復習支援としているのは、最初に学ぶ際には、学習用の専門的な教材を利用して学んでもらうことを前提として、提案システムはあくまでも補助的な教材として利用してもらうことを想定していることによる。

## 2 関連研究

e-Learning 学習履歴を用いたドロップアウト兆候者早期抽出手法の提案と検証を行った研究 [1] がある。この研究では e-Learning の学習履歴と成績からドロップアウト兆候者を早期に発見し、対象者に関しては補習を行うことでプログラミングの授業におけるドロップアウト兆候者を支援している。この研究においては、支援する内容をあらかじめ人が用意しておく必要があるが、本研究では ChatGPT を用いることで、支援システムを容易に構築できる環境を提供する。

実際に SDGs の講義に ChatGPT を導入してみて SDGs 教育・ESD において大学生がどのように ChatGPT を用いることが効果的な学習につながるのかを明らかにすることを目的にした研究 [2] がある。講義で ChatGPT について説明した上で、グループワーク時の議論、講義後の個人課題時に ChatGPT を利用するか否かを選択させて、ChatGPT が学習の質にどう影響するか調べている。この研究では、SDGs を学ぶ上で発生する議論や、SDGs を学んだ後の課題の支援に ChatGPT を活用しているが、本研究では何かを学んでいるときに学習そのものを支援する。

また、復習支援に関する研究 [3] においては、デジタル教科書と課題管理システムの学習ログを活用したプログラミング学習支援システムを提案している。この研究においても、共通の学習項目のタグが付与された教科書と課題をあらかじめ人手で用意する必要があったが、本研究では復習すべき項目の推定を ChatGPT の API[4] を用いることで自動化する。

\*連絡先： 滋賀県立大学 工学部  
中村 和哉  
〒 522-8533 滋賀県彦根市八坂町 2500  
E-mail: on23knakamura@ec.usp.ac.jp

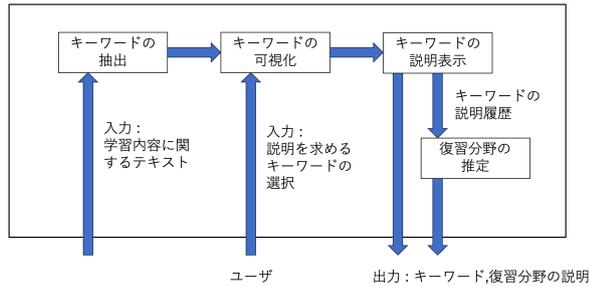


図 1: 系統的学習の復習支援システムの流れ

### 3 系統的学習の復習支援システム

本章では系統的学習の復習支援システムについて、学習内容における疑問点をどのように解消し、また、どのようにして復習支援をするのかについて述べる。なお本研究における「系統的学習」とは、学習分野において、学ぶべき項目(単元)が複数あり、それらを順序よく学習することを指す。

#### 3.1 系統的学習の復習支援システムの概要

系統的学習の復習支援システムの流れを図1に示す。まず学習内容を表すテキストが入力として与えられる。その中から学習内容を理解する上で重要になるキーワードをハイライトして可視化する。ハイライトされたキーワードをユーザが選択するとそのキーワードの説明が表示される。さらに、その説明が理解できなかった場合は、質問したキーワードの履歴から復習すべき単元を推定して出力する。

#### 3.2 入力：学習内容に関するテキスト

入力として与える学習内容に関するテキストは、授業の音声テキスト化したものや、インターネット上にある学習内容を記述したテキストとする。

#### 3.3 キーワードの抽出

入力されたテキストに含まれる「学習内容を理解する上で重要になる単語」をキーワードとして抽出する。抽出対象とするキーワードは、英語や数学のような教育分野の学習においては、教科書の用語集などの単語をキーワードとして用いることができる。また、任意に設定する学習分野の学習においては、ChatGPTに「〇〇(分野や単元名を与える)を行う上で重要な要素やポイントを△個教えてください」と質問することで、キーワードを取得することができる。

三平方の定理は、直角三角形の辺の長さの関係に関する定理だ。

図 2: キーワードの可視化の例

表 1: キーワードの説明を生成するプロンプトの例

プロンプト
あなたは中学校3年生の数学の先生です。三平方の定理について数学が苦手な中学三年生に簡潔に説明してください。ただし回答には中学3年生で習う範囲までの知識を使って回答し、それが不可能な時は「回答できません」と出力して下さい。また、回答は質問された単語の意味や定義など必要最低限のものにして下さい

例えば、結婚活動を学習分野とする場合、「結婚活動をする際に重要な要素やポイントを10個教えてください」とChatGPTに質問する。結果として、「共通の価値観、コミュニケーション能力、相手への尊重、信頼関係の構築、金銭面の合意、感謝の意識、共通の目標と計画、愛情表現、プロポーズの計画」というキーワードが得られる。

#### 3.4 キーワードの可視化

入力テキスト中のキーワードをハイライトして可視化する。図2にキーワードの可視化の例を示す。キーワードは前節で定義したように、「学習内容を理解する上で重要になる単語」であるため、この単語の意味がわからない場合に、説明を確認しやすくすることを意図している。

#### 3.5 キーワードの説明表示

前節で可視化されたキーワードを、マウスでクリックすることで、その説明をChatGPTを利用して表示させる。説明を表示させるプロンプトの例を、表1に示す。

また、条件を与えて説明を再生成するプロンプトを、表2に示す。これらのプロンプトは、後述するインタフェースに設けられたボタンを押すことで機能する。

#### 3.6 復習が推奨される分野の推定

ユーザが説明を求めたキーワード集合から、復習が推奨される単元を推定する。復習を推奨する単元を推定するために、単元間の学習順序を表す木構造を生成し

表 2: 条件を与えて説明を再生成するプロンプト

追加条件	プロンプト
もう少し易しく	(キーワード) について前回の回答をもう少し易しくしてください。前回の回答: (前回生成した説明)
もう少し詳しく	(キーワード) について前回の回答をもう少し詳しくしてください。前回の回答: (前回生成した説明)
具体例	(キーワード) の具体例を示してください。
応用例	(キーワード) の応用例を示してください。

表 3: 学習順序を尋ねるプロンプト

プロンプト
○○(学習分野)を学習する際に、(単元 A) と (単元 B) の 2 つの単元で、事前に学んでおかないともう片方が理解できないような単元はどちらですか。どちらを先に学んでも問題ない場合はどちらを先に学んでも良いと回答して下さい。

た上で、説明が求められた複数のキーワードの親ノードに与える単元を復習が推奨される単元として出力する。ツリー構造は、以下の手順で生成する。

1. 学習分野内の単元集合に対して、全ての単元のペアについて、表 3 の学習順序を尋ねるプロンプトを ChatGPT に与えてその回答を集める。
2. 順序が発生したペアについて、有向リンクを生成する。
3. 単元 A から単元 B に到達可能なパスが複数存在する場合、もっとも長いパスのみを残し、短いパスが含むリンクを削除する。

この手順で作成された学習分野「英語」のツリー構造を図 3 に示す。

### 3.7 学習支援インタフェース

本システムの学習支援インタフェースを図 4 に示す。

- ① 入力した学習内容テキストを表示する領域。キーワードをクリックして説明表示が可能。
- ② 質問した単語や復習内容の説明を表示する領域。
- ③ キーワード以外の単語を質問する入力フォーム
- ④ 質問履歴を表示する領域。

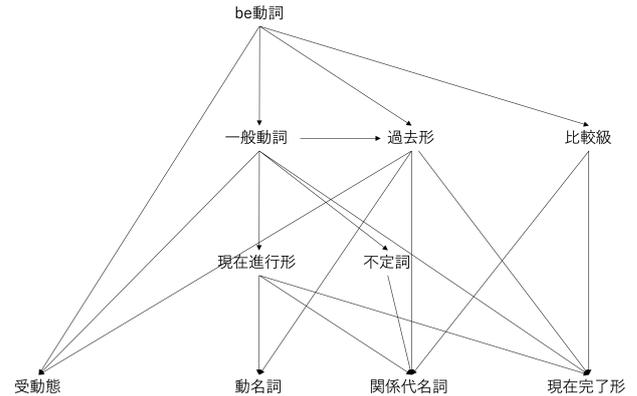


図 3: 学習分野「英語」のツリー構造

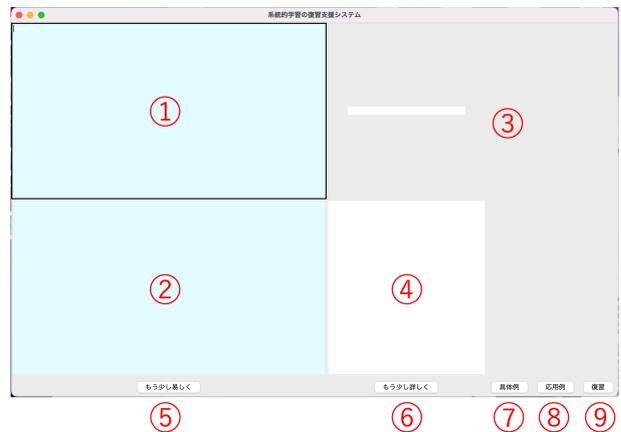


図 4: システムのインタフェース

- ⑤ 「もう少し易しく」ボタン
- ⑥ 「もう少し詳しく」ボタン
- ⑦ 「具体例」ボタン
- ⑧ 「応用例」ボタン
- ⑨ 「復習」ボタン

例えば、「自動車の運転」を学んでいる人が、キーワード「駐車の手順」をマウスでクリックすると、図 5 のように説明が 2 の領域に表示される。そして、「復習」ボタンを押すと、図 6 のように復習が推奨される単元の説明が同領域に表示される。

## 4 系統的学習の復習支援システムの評価実験

本章では、提案する系統的学習の復習支援システムが、学習を支援する効果があるかを確認するために行った実験について述べる。

質問：駐車の手順

回答：駐車手順は以下の通りです。

1. 駐車スペースを見つけます。スペースがあることを確認し、周囲の車や障害物に注意して近づきます。
2. ウィンカーを出します。駐車スペースの前に車を進める前に、方向指示器（ウィンカー）を出し、駐車を他のドライバーに示します。
3. 車をスペースに進めます。スピードを緩めながら、両側の車やスペースの適切な位置に注意しながら進みます。
4. 車体の位置を調整します。スペースに進んだ後、適切な位置に車体を調整します。スペースの中央に車を配置することを目指しましょう。
5. 車を止めます。ブレーキを踏んで車を停止させます。必要に応じて、パーキングブレーキを使用します。
6. エンジン进行を切ります。停止位置でエンジンを停止させます。

以上が基本的な駐車手順です。安全確認を忘れずにを行い、周囲の車や歩行者に配慮しながら駐車を行ってください。

図 5: 「駐車の手順」についての説明表示

復習したら良い内容：ブレーキの操作  
 回答：ブレーキの操作は、自動車や自転車などの乗り物の速度を減速または停止するために行う操作です。ブレーキを踏むことで、車輪にかかる摩擦や圧力が増加し、乗り物の運動エネルギーを減少させます。これにより、乗り物の速度を制御することができます。ブレーキの操作は安全運転に欠かせないものであり、必要な場面や状況に応じて適切に行う必要があります。

復習したら良い内容：標識の理解  
 回答：標識の理解とは、特定の情報やメッセージを伝えるために設置された標識の内容や意味を正確に理解することです。標識は道路や公共の場などでよく見られるもので、交通ルールや安全に関する指示、施設の案内、警告、禁止事項などの情報を提供します。標識を正しく理解することは、交通事故や危険な状況を避けるために非常に重要です。標識の形状、色、図案、テキストなどの要素が使用されており、それらを適切に解釈することで、標識が示す意味や目的を正確に把握することができます。

復習したら良い内容：アクセルの操作  
 回答：アクセルの操作には以下の意味があります。

1. アクセルを踏み込む：車の速度を上げるためにアクセルペダルを踏み込みます。踏み込む量に応じてエンジン出力が増え、車の速度が上昇します。
2. アクセルを離す：アクセルペダルから足を離し、エンジンの出力を抑えます。これにより車の速度を下げることができます。
3. アクセルを踏む深さを調整する：アクセルペダルを踏む深さを調整することで、車の速度を微調整することができます。アクセルを踏む深さを増やせば加速し、減らせば減速します。
4. アクセルを完全に踏む：アクセルペダルを最も深く踏み込み、エンジンの最大出力を引き出します。これは、

図 6: 復習が推奨される単元の説明表示

#### 4.1 実験手順

12名の大学生の被験者に、表4に示す5つの学習分野について、「説明対象とした単元」についての説明文を読んでもらい、提案システムを利用してもらった上で、以下の2つの課題を行ってもらった。

1. 「説明対象とした単元」についてあなたが理解した内容を100字以内で回答してください
2. 「特定の単元」を理解していない相手に「説明対象とした単元」について説明する内容を100字以内で回答してください

5つの学習分野のうち、「英語」と「プログラミング」は、教育系の分野として用意した。また、「結婚」「自動車の運転」「ダイエット」は、教育系以外の分野で、実施に一定の手順を要する分野として用意した。表5に各学習分野が含む単元を、図7にシステムによって構築された学習分野「ダイエット」における学習順序を表すツリー構造を示す。

#### 4.2 実験結果

表6に単元「カロリーの消費」の説明文の回答例を示す。このように得られた説明に文に対して、あらか

表 4: 実験に用いた学習分野と「説明対象とした単元」および「特定の単元」

学習分野	説明対象とした単元	特定の単元
英語	現在完了形	一般動詞
プログラミング	配列変数	変数
結婚	信頼関係の構築	自己分析
自動車の運転	駐車の手順	ブレーキの操作
ダイエット	カロリーの消費	適切な運動

表 5: 実験に用いた学習分野における単元

学習分野	単元
英語	be 動詞, 一般動詞, 現在進行形, 過去形, 不定詞, 動名詞, 比較級, 受動態, 現在完了形, 関係代名詞
プログラミング	標準入力, 標準出力, 変数, 条件分岐, 繰り返し, 配列変数, 多重ループ, main 関数, 引数, 戻り値
結婚	コミュニケーション能力, 自己分析, 身だしなみ, 家族との関係, 相手の分析, 金銭面の合意, 信頼関係の構築, プロポーズの計画, 共感力, 愛情表現
自動車の運転	交通法規の理解, アクセルの操作, ブレーキの操作, 注意力, 標識の理解, 駐車の手順, 車線変更の手順, 予測力, 速度管理, 車両の点検
ダイエット	目標設定, 栄養バランス, 摂取カロリー, 食事のタイミング, カロリーの消費, 適切な運動, 外食時の工夫, 適切な食量, ストレス管理, 食事記録の作成

じめ用意した表7に示す各単元の説明に必要と考えられるキーワード（類義語を含む）を、全て含む説明文を作成した人数を表8に示す。また、

12名の全被験者に対して、多くの単元において過半数の人が必要な単語をすべて含む説明文を作成できており、一定の効果があったことが確認できる。単元「配列変数」の説明においては、被験者が全員プログラミングを学習しており、普段からプログラミングを行っていたことから、十分な説明を行うことができたと考えられる。一方で、単元「駐車の手順」と「信頼関係の構築」においては他の学習分野に比べて効果が高くなかった。これは、自動車の運転や結婚に向けた活動は、知識のみで実行できるものではなく経験を必要としたこと、また特に結婚においては、まだ活動の経験がなかったことなどが影響したためと考えられる。しかし、そのような経験が少ない分野においても、約半数の人が説明文を作成できたことには意味があると考えられる。

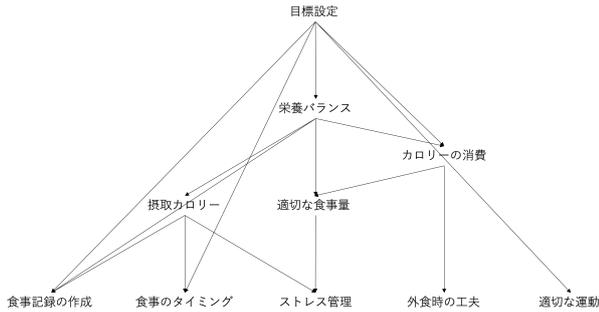


図 7: 学習分野「ダイエット」における学習順序を表すツリー構造

表 6: 単元「カロリーの消費」の説明文の回答例

課題	作成された説明文
自分が理解した内容	カロリーの消費とは、摂取した食物をエネルギーとして利用する過程で消費されるカロリーの量のこと。基礎代謝と運動によってカロリーを消費することができ、消費カロリーを増やすことでダイエットができる。
相手に説明する内容	ダイエットにおけるカロリーの消費は適切な運動を意識することが大切である。ここで適切な運動とは、有酸素運動や筋力トレーニングのことであり、体重を減らすだけでなく、筋肉をつけたり、基礎代謝向上にもつながる。

## 5 結言

本研究では、ChatGPTを用いた系統的学習の復習支援システムを提案し、その効果を実験により検証した。提案システムでは、既存の学習分野はもとより、系統的な学習が想定される任意の学習分野においても、その復習を促すことができ、実験により一定の効果がある可能性を確認した。

今後の課題としては、現在手動で行っている、学習分野における単元やキーワードの設定の自動化が挙げられ、最終的に学習分野のみを与えるだけで、その分野の系統的な復習を可能にするシステムとしていきたい。それにより、学習教材を提供する人の調整が不要となるため、学習者が任意の学習分野を選択して、自由に学習（復習）していくことができるようになると思われる。

本研究では、復習支援を念頭に置いたシステムとしたが、結婚活動のような任意の学習分野を設定する際には、必ずしも最初の学習用の教材が存在しないことも考えられる。そのため、そのような分野の学習においては、最初の学習においても活用できるような教材を作成するにも、本システムが活用できる可能性もあ

表 7: 各単元の説明に必要と考えられるキーワード

単元	自分の理解	相手への説明
現在完了形	過去分詞, 現在, 過去	動作, 状態, 現在, 過去
配列変数	連続的, 格納	型, データ, まとめる
信頼関係の構築	尊重, コミュニケーション	理解, 自分
駐車の手順	安全確認, ブレーキ	安全確認, ペダル, 踏む
カロリーの消費	運動, 代謝	有酸素運動, 筋力トレーニング

表 8: 各単元の説明に必要な単語をすべて含む説明文を作成した人数

単元	現在完了形	配列変数	信頼関係構築	駐車手順	カロリーの消費
自分が理解した内容	7	12	7	7	9
相手に説明する内容	8	9	6	4	11

ると考えている。

## 参考文献

- [1] 高岡詠子, 大澤佑至, 吉田淳一: e-Learning 学習履歴を用いたドロップアウト兆候者早期抽出手法の提案, 検証および今後の可能性, 情報処理学会論文誌, Vol.52, No.12, pp.3080 – 3095 (2011)
- [2] 平本督太郎, 亀田樹, 島田高行: SDGs 教育・ESD における大学生の ChatGPT を用いた主体的な学習方法に関する基礎的研究, Beyond SDGs イノベーション研究, Vol.4, No.6, pp.1 – 14 (2023)
- [3] 落合初音, 田中昂文: デジタル教科書と課題管理システムの学習ログを活用したプログラミング学習支援システム, 電子情報通信学会技術研究報告 (KBSE), Vol.123, No.124, pp.76 – 81 (2023)
- [4] ChatGPT API: (URL) <https://openai.com/blog/introducing-chatgpt-and-whisper-apis> (2024年2月27日確認)

# 視線計測を利用して音楽から想像された色の推定

## Estimation of Imagined Color from Musics using Eye-tracking

Shan Junjie<sup>1\*</sup> 西澤 太次郎<sup>2</sup> 西原 陽子<sup>2 †</sup>  
Junjie Shan<sup>1</sup> Taijiro Nishizawa<sup>2</sup> Yoko Nishihara<sup>2</sup>

<sup>1</sup> 立命館グローバル・イノベーション研究機構

<sup>1</sup> Ritsumeikan Global Innovation Research Organization, Ritsumeikan University

<sup>2</sup> 立命館大学情報理工学部

<sup>2</sup> College of Information Science and Engineering, Ritsumeikan University

**Abstract:** 音楽はジャンルにより人間に与える印象が異なる。その印象から想像される色も異なると考えられる。本研究では視線計測装置を用いて音楽の印象から想像される色を推定する手法を提案する。ユーザに音楽を聴かせながら、ランダムに生成された色パネルを提示し、ユーザが見た色を視線計測で記録する。記録した色をクラスタリングし、頻度の上位3位までのクラスタの色を表示する。ロック、バラード、ポップスのジャンルで男声ボーカルと女声ボーカルの合計6つのジャンルの楽曲を3曲ずつ、合計18曲を用意し、10名の被験者で評価実験を行った。実験の結果、音楽のジャンル毎に想像された色の違いとその特徴が確認された。

## 1 はじめに

音楽はジャンルにより人間に与える印象が異なる。その印象から想像される色も異なると考えられる。異なる音楽ジャンルに対応する想像色を解明することは、沢山の実際的な応用につながる。例えば、音楽ライブでは、様々な色の照明や映像を使い、音楽に合った雰囲気づくりをして観客を楽しませる工夫がなされる。その際、音楽の印象に合う色を知ることは、音楽を用いた雰囲気づくりを効果的に行うために必要である。Webページやゲームに音楽を用いる際にも、音楽の印象に合う色を使用することで、伝えたい雰囲気をより強く伝えることができ、音を流さずに閲覧やプレイするユーザに対しても音楽で伝えられない雰囲気を色で補うことができる。

しかし、そのためには音階やリズムの構成といった音楽理論の専門知識と、周波数や音圧といった音声コンテンツに関する複雑な信号処理が必要になることが多い [1]。それでも、結果は直感的で比較しやすいものにはならないことも多い。

そこで、本研究では、視線計測装置を利用して音楽から想像された色を計測するシステムを開発して、そのシステムを用いて音楽のジャンルに合った想像色を推定する手法を提案する。提案手法では、視線計測装置を利用して、音楽を聴きながらユーザが想像した色を

記録し、記録された色のデータはクラスタリング分析により、異なる音楽ジャンルと想像色のマッチングを実現する。

本研究では、主に以下の三つの内容を達成した：

1. 視線計測装置を用いてユーザに見られた色を記録するシステムを開発し、記録した色をクラスタに分けて可視化するインタフェースを実装した。
2. 記録された色データをクラスタリングするときに適切なパラメータを実験で確定した。
3. 開発した計測システムを利用して、代表的な三つの音楽ジャンル、ロック、バラード、ポップスで男声ボーカルと女声ボーカルの楽曲から、ユーザの感覚に合った想像色を評価実験で測定した。

## 2 関連研究

### 2.1 音楽の感覚とのマッチング

2002年に岩井らは、被験者にピアノの異なる鍵を聴かせて感じる色を選ばせることにより、調性・音色と色彩の間の非言語的なマッピング (non-verbal mapping) を調査した [2]。一方、若田らは2018年に心理学のSD法を使って、形容詞をもとに色、香り、音楽の種類の違いによる被験者への因子負荷量を調べ、それぞれの相関関係のマッチングや分類を行った [3]。2017年にItoh

\*連絡先：立命館グローバル・イノベーション研究機構  
〒525-8577 滋賀県草津市野路東1-1-1  
E-mail: shan@fc.ritsumei.ac.jp

† E-mail: nisihara@fc.ritsumei.ac.jp

らは、音楽の七つの音である「ド」「レ」「ミ」「ファ」「ソ」「ラ」「シ」が、虹色の七つの色である「赤」「橙」「黄」「緑」「水色」「青」「紫」との対応関係を15名の被験者からの共感覚で調査した [1]。Itohらの研究に基づいて、2021年、天野は音楽から色コスチュームの映像を生成する手法を提案した [4]。天野は五線譜に基づく「音階の螺旋表現」や「音色」、「音量」などの情報を用いて、入力された音楽から自動的に立体的な虹色の画像を合成し、プロジェクターでコスチュームの映像を人に投影する。赤井らは2014年に、26種類の楽器からの音を被験者に聴かせてイメージされた色を選ばせる実験を行い、音色と色彩の関係について考察した [5]。

色彩以外の内容について、大山らは2007年に、任意の音楽と画像を入力し、画像の印象に合わせて音楽を自動的にアレンジする手法を提案した [6]。大山らは画像の色分布から各色のリズムパターンを算出し、楽曲のリズムとベースを差し替えることにより、楽曲を自動的にアレンジする。合志らは2008年に、音楽の音圧や周波数を考慮した音楽の感性を自動推定する手法と音楽に合わせて照明を自動的に演出するシステムを開発した [7]。

本研究では、既存研究に多く使われた音楽理論の専門知識に触れず、音楽の感覚と色のマッチングをより直接的に行う。このために視線計測装置を用い、人が感じる音楽と色のマッチングを取得する。

## 2.2 視線計測応用に関する研究

視線計測装置はユーザの注目場所や視線に関連する情報が取れるため、様々な場面や領域で研究を行っている。従来の研究では、ユーザーインターフェース (UI) の領域に焦点を当てていた。生田目ら (2011年) や松延 (2014年) は、視線計測を用いて、アクセシビリティとユーザビリティを両立したウェブサイトのデザインに関して調べた [8,9]。杉邑らは2013年に、視線計測を利用して、ユーザーインターフェースに対する「慣れる程度」を計測する手法を提案した [10]。同じく2013年に、石井らは縦スクロールでの表示方法の違いにより、文章を読む際の視線の動きの相違点を視線計測で判明した [11]。

近年になると、視線計測装置はより豊かな研究シナリオで活躍している。藤脇ら (2021年) や坂口 (2023年) は、障害者や高齢者などの安全な生活をサポートするために、視線計測を用いて電動車いすの搭乗や直線単路を自動車で走行する時の被験者の視線を計測し、被験者が安全確認のために注視している場所や走行中の注目する特徴をそれぞれ分析した [12,13]。江草らは2021年に、視線計測装置を用いて、漫画表現に

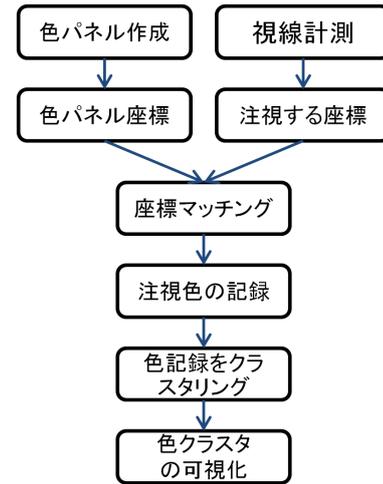


図 1: 視線計測で注視する色の記録及び可視化の流れ

よる解説を鑑賞する際の視線移動の測定及びキャラクターや吹き出しの効果を検証していた [14]。山田らは2021年に、水平作業台に搭載するディスプレイで情報を表示する時に、作業台にある遮蔽物を避けて情報を表示するため、視線計測装置を用いて作業者が注視している場所の推定手法を提案した [15]。

本研究では、ユーザの注目する色を計測するために視線計測装置を用いる。より即時的にユーザが注目した色を反映できるため、視線計測を用いてユーザに見られた色を記録するシステムを提案し、刺激を与えたときにそれに対応して注目する色を精度高く計測するため、専用のインターフェースを開発する。また、記録された注目色の集合から、刺激に対応し想像された色を推定するためにクラスタリングの技術を用いる。

## 3 提案手法

### 3.1 システムの流れ

図1に提案した視線計測装置を用いてユーザに注目された色の記録と可視化するシステムの流れを示す。本研究では、作成された色パネルにある各色のスクリーン座標を取得し、ユーザの視線座標とマッチングすることにより、ユーザに注視される色を記録する。

### 3.2 色パネルの作成

本研究では、4×4のパネルにランダムに生成される色を付けることにより、16色のパネルを作成する。

表1に色パネルのセッティング情報を示す。色パネルを表示するウィンドウのサイズは縦横各800ピクセルである。各色ブロックの間の距離は20ピクセルと設置する。毎回ランダムに生成される16色の種類は出来

表 1: 色パネルのセッティング情報.

項目	数値
縦	800 px.
横	800 px.
パネルサイズ	4 × 4
各色ブロック間の距離	20 px.
各生成色間の RGB 閾値	150
更新間隔	2.5 sec.

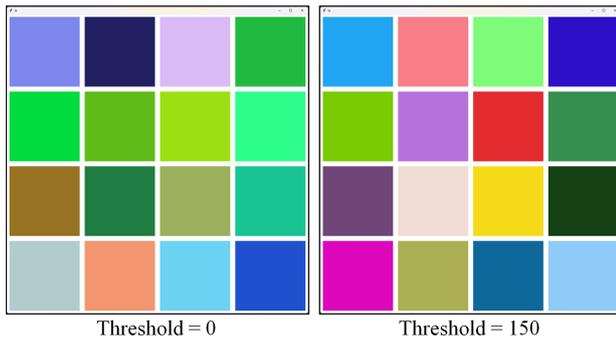


図 2: 異なる RGB 閾値の設定で生成した色パネル

るだけ豊かになるため、生成される各 RGB 値の色の間に 150 の閾値 (Thre) が設定される (Eq.1). この閾値により、ランダムに生成された全部の 16 色に対して、任意の 2 つの RGB 値の差の合計 (マンハッタン距離) をコントロールしている。

$$|R_i - R_j| + |G_i - G_j| + |B_i - B_j| \geq Thre; \quad (1)$$

$(i, j \in \text{Generated 16 Colors})$

図 2 は、作成された色パネルの様子と、RGB 閾値が 0 と 150 の場合でランダムに生成された色パネルの区別を示す。色パネルは 2.5 秒の間隔で更新され、そのうち 0.5 秒は毎回更新後にユーザの視線に与える反応時間で (表 2)、2 秒はユーザの視線に対する記録時間である。

### 3.3 視線計測から注視色の抽出

本研究では Tobii Pro Nano<sup>1</sup> を利用して視線計測を行う。この装置をモニタの下部に設置し (図 5)、モニタに向いているユーザの左目と右目のスクリーン内の座標をそれぞれ計測する。提案手法では、計測されたユーザの左目の座標と右目の座標の中点をユーザの注視点として算出する。ユーザの注視点がある色ブロックの範囲にある時、そのブロックのその時点での色の RGB 値を抽出し、注視色リストに記録する。

<sup>1</sup><https://connect.tobii.com/s/nano-get-started>

表 2: 視線計測のセッティング.

項目	数値
記録間隔	100 ms.
注視に対する視線移動閾値	50 px.
色パネル更新に与える反応時間	500 ms.

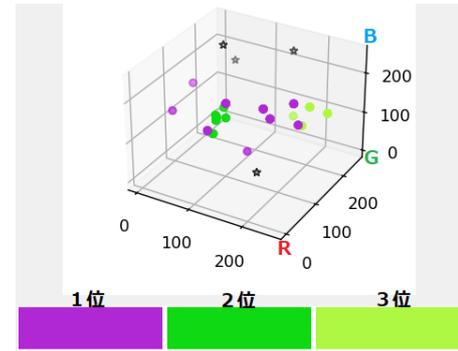


図 3: 推定するユーザの注視色の可視化

表 2 に視線計測から注視色の抽出に対するセッティングを示す。記録間隔は 0.1 秒 (100ms) で、一秒間に最大 10 回の注視色を記録する。ユーザは色パネルから気になる色を探し出す時に、視線が色パネル上に移動する必要がある。視線が移動途中で注視点の通過する色ブロックは記録されないように、視線の移動閾値を 50 ピクセルに設置している。毎回記録を行うたびに、ユーザの注視点の座標が前回の記録時の注視点座標から 50 ピクセル以上離れている場合は、ユーザの視線が移動中と見なし、その時点の注視点の色ブロックの RGB 値を記録しない。色パネルが更新されたとき、最初の 0.5 秒 (500ms) 間はユーザに与える反応時間として、色の記録は発生しない。

### 3.4 注視された色種類の推定と可視化

色パネルはランダムに色を生成するため、ユーザが同じ種類の色を注視したとしても、記録された色の RGB 値は必ずしも同じとは限らない。たとえ人の目で同じ色に見えていても、RGB 成分の各数値に僅かな差が生じることがある。そこで、ユーザに注視された色種類を推定するため、記録された色の RGB 値リストを用いて、クラスタリングを行う。

提案手法では、DBSCAN 法を用いて記録された色リストを複数のクラスタに分ける。分けたクラスタの数は、ユーザに注視された色の種類数と見なし、各クラスタにある記録数は、ユーザがその種類の色を注視した頻度と考える。そのため、より多くの記録数を含む

クラスタが、より頻繁にユーザに注視された色の種類である。本研究では、各クラスタを、ユーザに注視された色の種類として、そのクラスタに含まれる色記録のRGB値の平均を、その種類の代表色として推定する。

提案システムは、各クラスタが含む記録数の順番により、上位三つのクラスタに対して、色記録のRGB値の3D空間分布と平均色を可視化する。図3は一つの可視化のサンプルを示す。下部の色バーは左から1位、2位、3位の順で上位三つのクラスタの平均色を展示し、上部は各クラスタが3DのRGB空間内での分布を示している。黒い星マークの所はDBSCANで検出したノイズの記録点を示す。

## 4 評価実験

### 4.1 計測パラメータを決める予備実験

提案システムを用いて音楽からユーザの想像色を推定する際に、以下の三つのパラメータを設定する必要がある：

1. 計測時間：色パネルはランダム生成なので、計測時間が長くなると、記録データ数が増える一方で、ノイズも増える可能性がある。ノイズが沢山ある場合はDBSCANのクラスタリング精度に影響する。
2. DBSCANの近傍の密度を示す閾値 `min_samples`：一つの色種類に対して、少なくとも何個の色記録が必要である事を示す。
3. DBSCANの半径の距離 `eps`：RGB成分の各数値にある差を同じ種類の色と見なす範囲を示す。

上記三つのパラメータのうち、2番目の `min_samples` は5に設定した。注視色の記録間隔は0.1秒で設定しているため、一つ注視された色の種類と見なすには、少なくとも0.5秒以上（記録数5回以上）ユーザに注視された必要があると考えられる。1番目の計測時間と3番目の `eps` は、予備実験により適切なパラメータ値を決める。

#### 4.1.1 予備実験の手順

適切な計測パラメータを判断する予備実験の手順は以下となる：

1. 色探しのテーマを用意する。
2. 被験者に色探しテーマを与える。
3. 提案システムを使って、被験者に色パネルからテーマに合う色を目で探させる。

表 3: 予備実験で与える色探しテーマ。

決まった色	自由に想像する色
ボス下の色	自分のイメージカラー
草原の色	好きな野菜の色
Facebook ロゴの色	染めてみたい髪色
ピカチュウの色	好きなマカロンの色
青空の色	好きなかき氷の色
巨峰の色	好きな金平糖の色

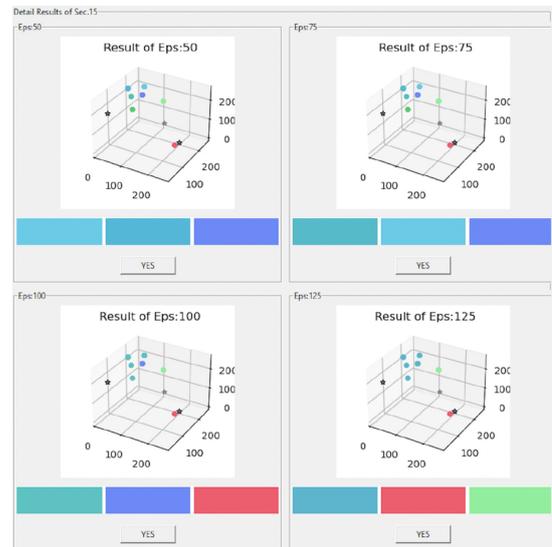


図 4: 計測時間と eps の組合せの可視化結果サンプル

4. 計測時間と eps が異なる色記録のクラスタリングの可視化結果を被験者に提示する。
5. 被験者は可視化結果から与えたテーマの色と合った結果を投票する。
6. 色探しのテーマを変えて、用意したテーマを終わるまで手順2～手順5を繰り返す。

用意した色探しテーマは表3に示す。「決まった色」と「自由に想像する色」の二つのグループに分けて、それぞれ六つのテーマ、合計12個のテーマを用意した。計測時間は10秒、15秒、20秒、25秒、epsは50、75、100、125の各四つずつに設定し、各計測時間と各epsの組合せで可視化する合計16種類の結果を被験者に提示する。図4に、計測時間が15秒で、4種類のepsでの可視化結果のサンプルを示す。被験者が自分に与えたテーマに合わせる可視化結果であると思った組合せの下の「YES」ボタンを押して、組合せを投票する。

#### 4.1.2 予備実験の結果

予備実験は7人の被験者で実施した。表4は、全12個の色探しテーマに対して、7人の被験者の各計測時間とepsの組合せへの投票結果を示す。本研究は音楽から

表 4: 全 12 個のテーマにおいて被験者が各組合せへの投票結果.

eps	計測時間 (秒)				合計
	10	15	20	25	
50	<u>55</u>	54	46	52	<b>207</b>
75	<u>55</u>	48	46	49	198
100	52	49	44	50	195
125	52	51	49	48	200
合計	<b>214</b>	202	185	199	800

表 5: 6 個の「自由に想像する色」のテーマにおいて被験者が各組合せへの投票結果.

eps	計測時間 (秒)				合計
	10	15	20	25	
50	<u>27</u>	21	20	26	94
75	<u>27</u>	22	22	26	97
100	24	22	19	22	89
125	21	21	19	19	80
合計	99	88	80	93	360

想像される色を計測することが目標であるため、「自由に想像する色」のグループにある六つのテーマにおける各組合せの可視化結果への投票結果も統計した (表 5 に示す).

全 12 個のテーマに対する投票結果では, eps が 50 の場合, 他の eps 値より多くの投票数 (207) が集まって, 計測時間が 10 秒の時に集まった投票数 (214) が, 他の計測時間より多くなった. 一方, 「自由に想像する色」のテーマに対する投票結果は, eps が 50, 計測時間が 10 秒の場合 (27) と eps が 75, 計測時間が 10 秒の場合 (27) での投票数は, 他の組合せより沢山の投票を集まった.

予備実験の結果から見ると, 計測秒数が短い方, eps が小さい方の組合せでの計測・可視化結果に対して, 被験者のイメージした色に合う傾向がある事を判明した.

## 4.2 音楽から想像された色の推定実験

本実験の目的は, 提案システムを用いて, 音楽のジャンルごとにイメージされる色を推定し, その推定結果を分析することである.

### 4.2.1 実験手順

実験の手順は以下の通りである:

1. ジャンル別の音楽を用意する.
2. 被験者に用意した音楽を聴かせる.

表 6: 用意したジャンル別の音楽リスト.

	ロック	バラード	ポップス
男声	KILLING ME Remember Jagerbomb	8 Letters One Call Away Too Good At Goodbyes	Lucky Strike Runaway Baby Say Somethin'
女声	Feel A Thing Lonely Girl Despondency	Lose You To Love Me Thinking About You Rise Up	Black Magic No Excuses Into You



図 5: 実験風景

3. 音楽を聴かせ始めてから 5 秒後に色パネルを表示し, 被験者の視線を計測する.
4. 被験者に色パネルの中から, 聴いている音楽から想像される色を目で探させる.
5. 提案システムにより, 被験者が注視した色を記録する.
6. 聴かせる音楽を変えて, 手順 2~手順 5 を繰り返す.

表 6 に実験で使用したジャンル別の音楽を示す. 被験者に聴かせる音楽はロック, バラード, ポップスの三つのジャンルごとに男声ボーカルと女声ボーカルの二種類で楽曲を 3 曲ずつ, 合計 18 曲を用意し, 6 つのグループに分けた. さらに, 言語からの影響を抑えるために, 用意した楽曲は全部英語の楽曲から選んだ. 各音楽に対する実験時間は 30 秒である (5 秒の導入 + 25 秒の計測).

実験に参加した被験者数は 10 人であり, 全員は日本人の大学生であった. 図 5 は提案システムを用いて音楽を聴きながら視線を計測する実験の様子を示す. 実

表 7: 被験者, 音楽のジャンルごとに計測された色記録のクラスタ数. (計測時間は 25 秒, eps は 50)

被験者 ID	音楽ジャンルごとのクラスタ数					
	男声 ロック	女声 ロック	男声 バラード	女声 バラード	男声 ポップス	女声 ポップス
ID1	15	15	19	24	15	18
ID2	22	9	15	15	17	6
ID3	5	9	11	16	15	12
ID4	7	10	19	18	12	12
ID5	14	19	19	22	12	9
ID6	25	16	17	21	15	16
ID7	26	13	14	25	17	10
ID8	7	14	12	21	10	15
ID9	9	14	10	18	21	14
ID10	7	18	8	23	12	11
平均	13.7	13.7	14.4	<b>20.3</b>	14.6	<b>12.3</b>

表 8: 被験者, 音楽のジャンルごとに計測された色記録から検出したノイズ数. (計測時間は 25 秒, eps は 50)

被験者 ID	音楽ジャンルごとのノイズ数					
	男声 ロック	女声 ロック	男声 バラード	女声 バラード	男声 ポップス	女声 ポップス
ID1	60	50	44	68	48	39
ID2	41	46	63	50	64	46
ID3	28	16	27	33	37	12
ID4	18	32	37	39	17	21
ID5	39	49	51	40	35	42
ID6	18	34	39	30	27	40
ID7	81	49	57	74	77	45
ID8	13	24	42	36	21	32
ID9	58	22	25	58	52	29
ID10	31	21	28	42	29	17
平均	38.7	34.3	41.3	<b>47.0</b>	40.7	<b>32.3</b>

験で使用したモニタのサイズは 24 インチ, 解像度は 1920 × 1080 ピクセル, リフレッシュレートは 60Hz であった.

#### 4.2.2 実験結果

表 7 は, 音楽のジャンル, 被験者ごとに, 25 秒の計測時間で記録された注視色リストに対するクラスタ数の統計結果を示す. 表 8 は, 音楽のジャンル, 被験者ごとに, 25 秒の計測時間で記録された注視色リストに対して DBSCAN で検出したノイズ数の統計結果を示す. 図 6 は, 計測時間が 10 秒, eps が 50 の場合で 10 人の被験者の全ての記録に対してクラスタリングを行い, 上位 3 つのクラスタにある色記録の RGB 値の 3D 空間分布を可視化した結果である.

女声バラードの楽曲で注視された色のクラスタ (色の種類) の平均数は最も高く (20.3), 女声ポップスの楽曲で注視された色クラスタの平均数は最も低かった (12.3). 同じ傾向で, 被験者の女声バラードの楽曲での色記録から検出されたノイズの色は最も多かった (47.0), 女声ポップスの楽曲での色記録から検出されたノイズは最も少なかった (32.3).

## 5 考察

被験者の注視色記録をクラスタリングし可視化した結果により, 音楽ジャンル毎に想像される色の違いとその特徴が確認された.

計測時間 10 秒での色クラスタの分布結果 (図 6) からみると, 男声ボーカルのロック曲 (①) では, 濃い赤色が多く記録されており, 女声ボーカルのロック曲 (④) では, 明るい赤色やオレンジ色が多く記録されていた. 男声ボーカルのポップス曲 (③) では, 黄色や黄緑色の所に多く偏っており, 女声ボーカルのポップス曲 (⑥)

では, 主にピンク色に集中していた. しかし, バラード曲の場合, 男声ボーカル (②) は水色や緑色の所に集中しているが, 女声ボーカル (⑤) は緑色やオレンジ色のような空間上に離れている色がそれぞれ多く記録されていた. この結果により, 男声ボーカルのロック曲 (①), 女声ボーカルのロック曲 (④), 男声ボーカルのポップス曲 (③), 女声ボーカルのポップス曲 (⑥), 男声ボーカルのバラード曲 (②) の 5 つの場合 (女声ボーカルのバラード曲 (⑤) 以外) において, 音楽からイメージされる色は人によらず共通している傾向があると考えられる.

被験者に注視されたジャンルごとの色記録から推定された色クラスタ数と検出したノイズ数の結果も同じ傾向を反映している. 音楽のジャンルごとに計測された色記録の平均クラスタ数により (表 7), 男声ロック (13.7), 女声ロック (13.7), 男声バラード (14.4), 男声ポップス (14.6), 女声ポップス (12.3) で注視された色の平均種類数は同じ程度であったが, 女声バラード (20.3) では多くなった. 女声バラードでの計測記録から検出された平均ノイズ数 (47.0) も他のジャンルより多かった. これらの結果により, 被験者が女声バラードのジャンルから想像された色の種類は集中しておらず, 多数の色を思い浮かべている傾向があると考えられる.

一方, 女声ポップスで計測された色の平均クラスタ数と検出されたノイズ数両方も, 全ての音楽ジャンルの中で最も少なかった (12.3 と 32.3). 可視化された色クラスタの分布グラフから見ると, 被験者が女声ポップスで注視した色の記録は, ほぼ一つの色種類 (ピンク) だけに集中していた. これらの結果から, 可視化した音楽から想像された色の推定が, 計測したユーザに注視された色の記録と概ね一致していることがわかった.

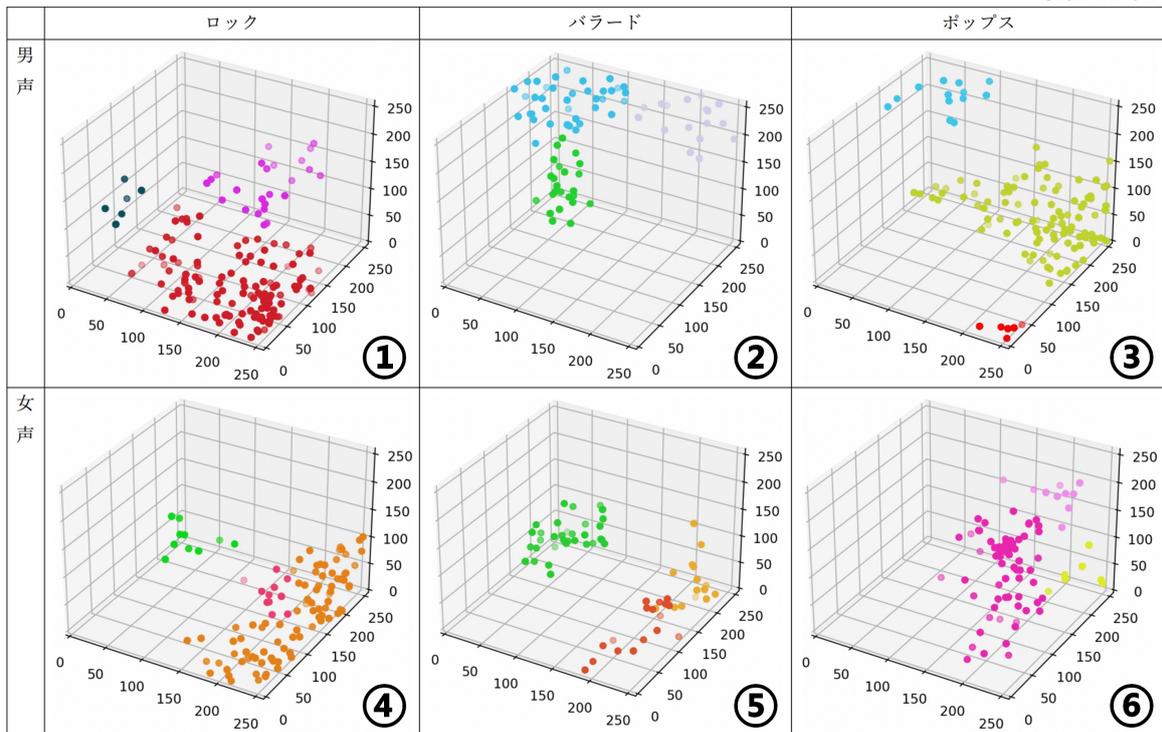


図 6: 被験者全員の注視色の記録から推定される音楽ジャンル, ボーカルごとの上位 3 つの色クラスタの分布の可視化 (計測時間 10 秒, eps は 50).

## 6 結論

本研究では視線計測装置を用いて音楽の印象から想像される色を推定する手法を提案した。提案手法では、視線計測でランダムに生成された色パネルからユーザに注視された色の記録とクラスタリングにより可視化するシステムを構築した。構築したシステムを用いて、音楽から想像される色を推定する実験を行った。6つの音楽ジャンルごとに3曲ずつ用意し、10名の被験者で計測実験を評価した結果、音楽は人に与えた印象からの想像色がジャンルにより分布と特徴が異なるのである事を確認した。6つの音楽ジャンルの中、5つのジャンルで被験者から得られたイメージ色は一致しており、1つのジャンルではイメージ色は一致せず、様々な色が得られた。

今後の課題として、様々なプロフィールの被験者を増やし、音楽から想像された色が年齢層、文化背景での違いや関連性を探索する予定である。

## 謝辞

本研究は、立命館グローバル・イノベーション研究機構、第4期拠点形成型 R-GIRO 研究プログラム「「心の距離メータ」を用いたフィジカル・サイバー空間における人間関係構築技術の開発」の支援を受けて実施していた。感謝の意を表します。

## 参考文献

- [1] K. Itoh, H. Sakata, I. L. Kwee and T. Nakada.: Musical pitch classes have rainbow hues in pitch class-color synesthesia, *Scientific Reports*, Vol. 7, No. 17781 (2017).
- [2] 岩井大輔, 長田典子, 津田学, 和氣早苗, 井口征士, etc: 音と色のノンバーバルマッピング-色聴保持者のマッピング抽出とその応用, 情報処理学会研究報告音楽情報科学 (*MUS*), Vol. 2002, No. 100, pp. 97-104 (2002).
- [3] 若田忠之, 森谷春花, 齋藤美穂: 色, 香り, 音楽に共通する印象次元の検討 2, *日本色彩学会誌*, Vol. 42, No. 3, pp. 96-99, (2018).
- [4] 天野憲樹: 音楽の立体的な色彩化とウェアラブル化の試み, *芸術科学会論文誌*, Vol. 20, No. 2, pp. 72-81 (2021).
- [5] 赤井良行, 李昇姫: 音色からイメージされる色彩の寒暖と音色構造の関係, *日本感性工学会論文誌*, Vol. 13, No. 1, pp. 221-228 (2014).
- [6] 大山喜冴, 伊藤貴之: Diva: 画像の印象に合わせた音楽自動アレンジの一手法の提案. *芸術科学会論文誌*, Vol. 6, No. 3, pp. 126-135 (2007).

- [7] 合志和洋, 清田公保, 三好正純, 古賀広昭: 音楽の印象に合わせた照明表現システムの研究開発, 熊本電波工業高等専門学校研究紀要, Vol. 35 (2008).
- [8] 生田目美紀, 北島宗雄: ユーザビリティとアクセシビリティを両立させるリンク表現の検討: 視線計測によるウェブデザインに関する研究, デザイン学研究, Vol. 58, No. 2, pp. 105–112 (2011).
- [9] 松延拓生: ウェブサイトのユーザビリティ評価のための視線分析方法, ヒューマンインタフェース学会論文誌, Vol. 16, No. 4, pp. 293–302 (2014).
- [10] 杉邑洋樹, 上野秀剛, etc: 視線情報に基づいたユーザインタフェースへの慣れの定量化, 研究報告ヒューマンコンピュータインタラクション (HCI), Vol. 2013, No. 2, pp. 1–8 (2013).
- [11] 石井亮登, 森田ひろみ: 縦スクロール表示された文章の快適な読み速度と眼球運動, 情報処理学会論文誌, Vol. 54, No. 6, pp. 1784–1793 (2013).
- [12] 藤脇蓮, 木室義彦: 傾斜歩道における構造化モデリング法を用いた車いす搭乗者の視線分析, 電気・情報関係学会九州支部連合大会委員会 2021 年度電気・情報関係学会九州支部連合大会 (第 74 回連合大会) 講演論文集, pp. 102–102 (2021).
- [13] 坂口雄介: 高齢者と若年者の直線単路走行時の注視特徴, 現代社会文化研究, Vol. 76, pp. 81–91 (2023).
- [14] 江草遼平, 青木良太, 楠房子, 稲垣成哲: 科学系博物館におけるマンガ表現による解説法のデザイン: キャラクターと吹き出しの導入による読み取りプロセスの変化に関する事例的検討, 一般社団法人日本科学教育学会・日本科学教育学会年会論文集, Vol. 45, pp. 399–402 (2021).
- [15] 山田孟, 長野真大, 中嶋良介, 仲田知弘, 松野省吾, 岡本一志, 山田哲男, 杉正夫: 水平作業台ディスプレイにおける作業者の頭部位置移動に対応した注視点推定システムの提案, 公益社団法人精密工学会 2021 年度精密工学会春季大会・精密工学会学術講演会講演論文集, pp. 161–162 (2021).

# ニュースに対するコメントが 閲覧者の意見形成に与える影響に関する予備調査

## The Influences of Online News' Comments on Viewer Opinion Formation

安尾 萌\*                      林 沙也加                      松下 光範  
Megumi Yasuo                      Sayaka Hayashi                      Mitsunori Matsushita

関西大学大学院総合情報学研究科  
Graduate School of Informatics, Kansai University

**Abstract:** 本研究の目的は、web ニュースを介した世論形成において SNS 上に投稿された意見やニュースサイトに投稿されるコメントが、ニュース閲覧者の意見に及ぼす影響を明らかにすることである。ソーシャルメディアを通じてニュースを収集する閲覧者は、しばしばニュースに関する他者の感想やコメントを読んだ上で、そのニュースに対する印象を形成する。そのため、他者のコメントによって web 上に形成される世論は、閲覧者の意見形成に影響を及ぼす可能性がある。本稿では、閲覧者が閲覧するコメントの多寡に着目し、ニュースに対するコメントを対象とした予備調査を行った。

### 1 はじめに

インターネットの普及は、ニュースへの接触行動を大きく変容させている。以前は、新聞やテレビ、ラジオといったマスメディアの報道によりニュースを受動的に知ることが多かったが、Web を介してニュースにアクセスできるようになった現在では、従来のマスメディア経由ではないニュース認知機会が増加し、興味をもったニュースの関連情報を能動的に辿る行為も容易になっている。特に SNS の発展は Web ニュース記事の共有や拡散に大きな影響を与えており、これらを通じてニュースを知る人も多い。SNS はニュースの共有・拡散だけでなく、ニュースに対する意見表明や議論を行うことも可能である。そのため、世論を形成、あるいは観測する新たな場として期待する声も存在する [10]。ソーシャルメディアを介してニュースを知る利用者は、そのニュースを拡散したユーザのニュースに対する反応を見た上で情報を取得する。例えば X (以下、旧 Twitter と表記) では、ある話題が利用者間の一部で一定数共有されると、共有された話題に関する語が抽出され、「トレンドワード」として表示される。このトレンドワード機能は、トレンドワードを含む投稿をワンクリックで検索、表示す

る機能であり、利用者はこの機能によって現在サービス上で注目されている話題を円滑に知ることができる。利用者のコメントを経由してニュースの詳細を知る利用者は、大元の記事を把握する前に、入手経路となった利用者のコメントに記述される意見や感想を閲覧することになる。そのため、このような経路でニュースを知る利用者は、大元の情報源を閲覧する前に、閲覧した利用者のコメントによって先入観を形成する。このような先入観は、大元の情報源から閲覧者が受ける印象を変容させ、ニュースに対する意見形成時に認知バイアスとして働くことが懸念される。こうした背景の下、本研究では、あるニュースに対して付与された意見が、他者の意見形成に及ぼす影響を明らかにすることを目指す。この課題を解決するにあたり、本稿では閲覧するコメントの量と、コメントの発信者の背景情報（過去の発信傾向やプリファレンス）の閲覧量に着目し、この2点の間にトレードオフが存在するという仮説のもと、コメントの受信環境の差によって形成される意見の差を明らかにするための予備調査を行った。

### 2 関連研究

ニュース記事に対するオンライン上のコメントの影響を分析する研究は、主に英語圏や韓国語圏などで広く行われている。Lee らの研究では、ニュース記事に付与されるオンライン上のコメントの論調や引用のされ方に

\* 連絡先：関西大学総合情報学部  
〒569-1095 大阪府高槻市霊仙寺町 2-1-1  
E-mail: k290993@kansai-u.ac.jp

よって、ニュースおよびコメントの閲覧者がニュースの論調をより友好的であると認識することを報告している [5]。またこの研究では、努力を要する認知活動を楽しむ心理的傾向である「認知欲求 [1]」についての調査を行い、より高い認知欲求を持つ閲覧者が、コメントから世論を推測しようとする傾向があることを示している。Houston の研究 [4] では、オンライン上の政治的なコメントが閲覧者に対して与える影響についての実験を行っている。この研究では、コメントが閲覧者に与える影響として、ニュース記事などの情報が閲覧者自身よりも他者に対してより大きな影響を与える、と考える心理的傾向である「第三者効果 [2]」に焦点を当てている。この実験では、ある政党を熱烈に支持するコメントが、閲覧者に対してより強い第三者効果を与えるという仮説のもと、2008 年のアメリカ合衆国大統領選挙に関する記事を対象に、2 人の大統領候補者についての支持・不支持による統制をとってコメントを閲覧させる実験を行った。その結果、ユーザコメントが大統領候補者のメディア上の描写に影響を与え、コメント群の政治的態度に関する認識にも影響することを報告している。また、Elidersらは、異なる文化圏においてユーザコメントが世論認識および閲覧者に与える影響についての調査を行い、韓国とドイツを対象に調査を行い、両国ともに集団主義的なグループにおいて「世論が自身の見解と概ね合致する」と認識する閲覧者が多く存在することを報告している [3]。これらの研究報告からは、ニュースに対するコメントがコメントの閲覧者に対してニュースの認識に影響を与えることを示している。

日本においてニュース記事へのコメントがもたらす影響に焦点を当てた研究は限られるが、ニュース記事について投稿されたコメントから、報道対象に対する人々の興味関心を分析する研究はいくつか試みられている (e.g., [11], [9])。時事問題についてのソーシャルメディア上の反応を分析する試みは、計算社会科学の分野で広く行われている (e.g., [7], [8])。特に、東日本大震災や熊本地震などの大規模災害や、新型コロナウイルスに代表される世界的課題を対象とした調査が報告されている。これらの分析対象としては、旧 Twitter をはじめとした大規模なソーシャルプラットフォームからの情報をもとに分析されてきた。一方、日本におけるオンライン上で政治的な議論に参加するユーザについて調査した報告 [6] では、ニュース記事に対してコメントを付与する閲覧者は、全体のうちのわずかしかないことを報告している。この調査では、オンライン上で積極的にニュース記事をシェアしたり、コメントを残す回答者は調査対象者のうち 6% にとどまることを明らかにしている。

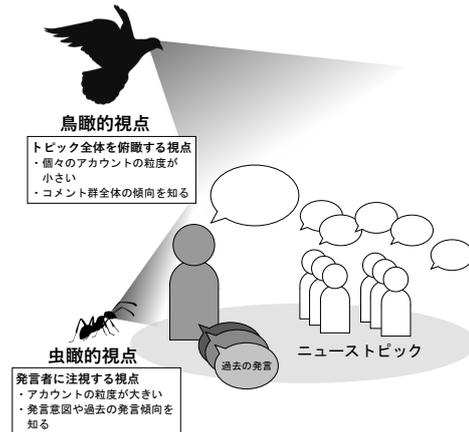


図 1: 視点の差異によるコメントの受信環境の変化

以上の従来研究からは、ニュースに対して意見を投稿するユーザがわずかであるにもかかわらず、これらの意見が閲覧者の印象に様々な影響を与えることや、世論を分析しようと試みる閲覧者はこのような web 上の意見を重視して世論分析を試みる傾向があることが明らかになっている。また、大量のコメントを利用して世論の分析を試みる研究からは、ソーシャルメディアが世論の形成、分析の対象として期待されていることが窺える。一方、従来研究では閲覧者の心理的傾向や属性情報を考慮した調査は存在するものの、閲覧者の情報受信環境に着目したものは限られている。そこで本稿では、ニュースにコメントを付与するアカウントを観察する視点に着目し、鳥瞰的・虫瞰的視点からコメントを観察した際に形成される意見にどのような差があるかを観察する。

### 3 鳥瞰的・虫瞰的視点からの情報アクセスと印象形成

本研究では、ニュースに対するコメントの閲覧者がコメントから受ける影響として、閲覧するコメントの総量と、コメントの発信者に関する背景情報の量が、意見形成に影響すると考えた。1 章で述べたように、発信者の背景情報とは、コメントを投稿した発信者のプロフィール情報や選好、過去の投稿などを含む情報が該当する。

あるニュースに対するコメントの集合を閲覧する場合、その最小単位は 1 人のユーザが投稿したコメントである。コメントの閲覧者は、そのニュースに関わるコメント全体を鳥瞰的に広く閲覧するか、各コメントに着目して虫瞰的に閲覧するかを決定することで、コメントからそれぞれ異なる情報を獲得することができる。本研究で想定するニュースコメントの受信環境の概念図を図 3 に示す。多数のユーザを対象にした鳥瞰的な閲覧を行

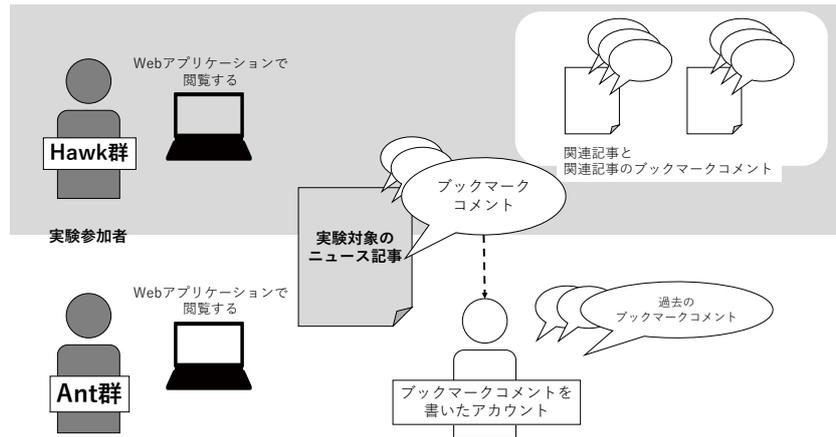


図 2: 各群ごとの情報の受信環境

う場合、閲覧者は情報群を「どのような趣旨の情報の集  
合か」という観点で捉える。この視点は、コメント群の  
傾向について鳥瞰的に眺めるものとして捉えることが  
できる。鳥瞰的な情報受信環境にいるユーザは、例えば  
あるトピックについて複数のニュース記事や複数のソー  
シャルメディアを通じてコメントを広く閲覧する状況が  
想定される。この視点で情報を観察する場合、個々の情  
報の詳細は削ぎ落とされる。一方、少数のユーザに焦点  
を当てた虫瞰的な閲覧を行う場合は、コメントの詳細な  
内容や、発信された背景を含めて詳細に目を通すこと  
ができる。虫瞰的な情報受信環境は、例えば特定のコミュ  
ニティに所属し、少数のユーザと長期的に関係を保っ  
ている状況が想定される。この視点で情報を観察する場  
合、観察対象となる情報の数が限られる。コメントの受  
信環境が意見形成に及ぼす影響を明らかにすることは、  
ニュースに対するコメントの閲覧者の意見が、他者のコ  
メントからどのような影響を受けているかを考慮した意  
見の分析を可能にし、よりバイアスのかからない情報の  
取得や意見の発信を促進するシステムの設計に貢献する  
ことが期待できる。

## 4 予備調査

ニュース記事に付与されたコメントによる意見形成の  
影響を明らかにするため、閲覧するコメント数、および  
コメントを投稿した発信者の背景情報に着目した実験  
を検討した。ニュース記事のコメントを対象とした実験  
を実施するにあたり、対象とするデータソースの選定と  
ニュース記事およびコメントの収集を行った。この実験  
では、web ページのソーシャルブックマークサービス  
である「はてなブックマーク」から、対象となるトピッ

クを選定し、記事を収集した。はてなブックマークは、  
ユーザ数 1,214 万人の「はてな」が提供するサービス  
の一つである\*1。このサービスの利用者は主に 30 代から  
40 代の男性で構成されており、これらの属性を持つユー  
ザによるコメントの獲得が想定される。調査対象とする  
トピックは、(1) 賛成意見と反対意見が混交する、(2) 多  
数の観点に基づく意見が獲得できる、の 2 点が期待でき  
ることを選定基準とし、「ウクライナ侵攻」および「画像  
生成 AI」の 2 つのトピックを対象とした。

### 4.1 実験デザイン

各群ごとの情報の受信環境を図 2 に示す。まず実験  
参加者を、虫瞰的視点で観察する群（以下、Ant 群と表  
記）、および鳥瞰的視点で観察する群（以下、Hawk 群と  
表記）の二群に分け、各群に対してニュース記事および  
記事に付与されたコメント（以下、ブックマークコメン  
トと表記）を提示する。その後、各群ごとに「記述内容  
に共感できたコメント」「記述内容に共感できなかった  
コメント」としてブックマークコメントを各 1 件選定さ  
せ、その後「自身の意見」を自由記述で回答させる。こ  
の時、Ant 群には実験刺激として、実験対象のニュース  
記事 1 件、およびそれに付与されたコメントをランダ  
ムに提示する。この群の参加者は、コメントを付与したア  
カウントが過去に投稿したコメント、およびそのコメン  
トの対象となった記事名を最大 30 件閲覧することが  
できる。回答時に、「選択したユーザのコメント」につい  
てのアンケート項目を設けることで、選定されるアカ  
ウントの過去のコメントの閲覧を促した。一方、Hawk 群

\*1 はてなメディアガイド 2024 年 1-3 月版, <https://hatena.co.jp/ads> (2024/2/26 確認)。

表 1: 付与するラベル名と付与基準

ラベル名	付与基準
noise	記事内容に無関係の内容、質問文の複製、誹謗中傷など
indifference	「興味が無い」「どうでもいい」という趣旨の意見
distrust	「信用できない」「疑わしい」という趣旨の意見
trust	「信用できる」という趣旨の意見
accept	「納得できる」「当然の結果である」という趣旨の意見
reject	「納得できない」「受け入れ難い」という趣旨の意見
P_impression	「嬉しい」「楽しい」などのポジティブな感想
N_impression	「悲しい」「腹が立つ」などのネガティブな感想
conflict	「複雑だ」「なんともいえない」など葛藤、混乱した感想
suggestion	「～すべきだ」「～のほうがいい」など提案を含む意見
criticism	記述内容、記述対象に関連する批判・非難の意見
consideration	記事内容を踏まえた考察を述べた意見

には実験刺激として、実験対象のニュース記事 1 件、およびそれに付与されたコメントに加え、関連記事として他に 2 件のニュース記事、およびそれらに付与されたコメントが同時に提示される。この群の参加者は関連記事および関連記事のコメントを読んだ上で回答を行う。調査は web アプリケーションを実装して行い、実験協力者は Yahoo!クラウドソーシングを通じて収集した。

#### 4.2 ケース 1: ウクライナ侵攻に関するニュース記事を対象とした分析

このケースでは、ウクライナ進行に関するニュース記事（以下、ロシア記事と表記）を対象として意見の収集と分析を行った。対象としたニュース記事は、NHK によって報道された「侵攻継続と和平交渉で意見が二分 ロシア 世論調査」\*2である。この記事について付与されたコメントを はてなブックマークから収集した結果、34 件のコメントが獲得された。Hawk 群に提示する関連記事は、はてなブックマーク上で関連記事として表示される記事から、ブックマーク数の多い順に選出した。選出した記事は、「ロシア石油大手、会長が「重病」経て死去と発表 病院から転落死か」\*3および「ロシア、行き場失う天然ガスを焼却か 環境への影響懸念」\*4である。調査協力者は 313 人であり、そのうち Ant 群は 158 人、Hawk 群が 155 人であった。

実験で得られた回答のうち、3 章で述べた実験参加者の「自身の意見」（以下、実験者コメントと表記）に対し、12 種類のラベルを付与した。付与するラベル名と付与基準を 1 に示す。ラベルの付与は 3 名の付与者によ

\*2<https://www3.nhk.or.jp/news/html/20220902/k10013799531000.html> (2024/2/27 確認)

\*3<https://www.afpb.com/articles/-/3421844>(2024/2/27 確認)

\*4[https://www.nikkei.com/article/DGXZQGN2704ZOX20C22A8000000/\(2024/2/27 確認\)](https://www.nikkei.com/article/DGXZQGN2704ZOX20C22A8000000/(2024/2/27 確認))

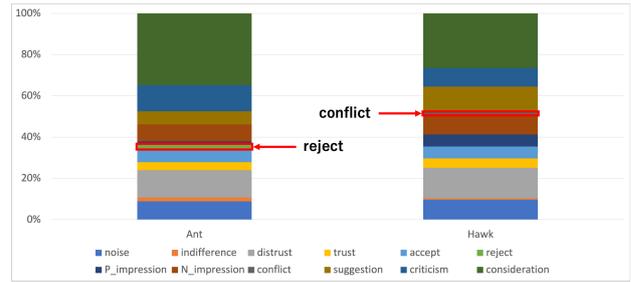


図 3: ラベルの構成比 (ロシア記事)

表 2: ブックマークコメントを選択した数 (ロシア記事)

選択されたユーザ	Ant	選択されたユーザ	Hawk
user_1	20	user_2	19
user_2	15	user_1	14
user_3	13	user_3	13
user_4	10	user_4	9
user_5	9	user_6	8
user_6	8	user_7	8
user_7	7	user_5	7
user_8	6	user_11	6
user_9	6	user_8	5
user_10	6	user_12	5

て実施し、うち 2 名以上が選定したものを各コメントのラベルとして決定した。3 名がいずれも異なるラベルを付与していた場合は、筆頭著者の判断によってラベルを決定した。ラベルの構成比を図 3 に示す。実験者コメントの傾向や Ant 群と Hawk 群のコメント内容の差異を観察した。獲得したコメントのラベルの構成比を観察した結果、reject は Ant 群のみに現れ、conflict は Hawk 群のみに現れた。

回答者から多くの賛同を得たブックマークコメントとその要因を調査するために選択されたブックマークコメントの割合を集計した。集計結果を表 2 に示す。その結果、Ant 群と Hawk 群ともに上位 3 位は user\_1, user\_2, user\_3 であった。これらのコメントはいずれも文字数の多いコメントであったため、ブックマークコメントの文字数と選択数の相関について調査した。ブックマークコメントの文字数と選択数について図 4, 5 に示す。このデータについて相関係数を算出したところ、Ant 群の相関係数は 0.510, Hawk 群の相関係数は 0.488 となり、いずれも中程度の正の相関が見られた。

#### 4.3 ケース 2: 画像生成 AI に関するニュース記事を対象とした分析

4.2 節と同様のケースが他のトピックでも確認されるかを明らかにするため、異なるジャンルのニュース記

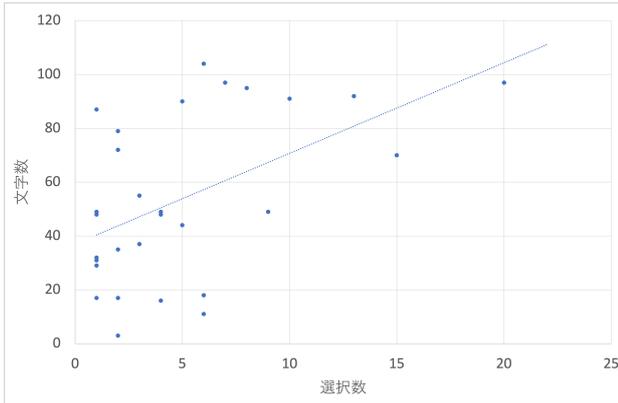


図 4: ロシア記事について文字数と選択数の相関関係 (Ant 群)

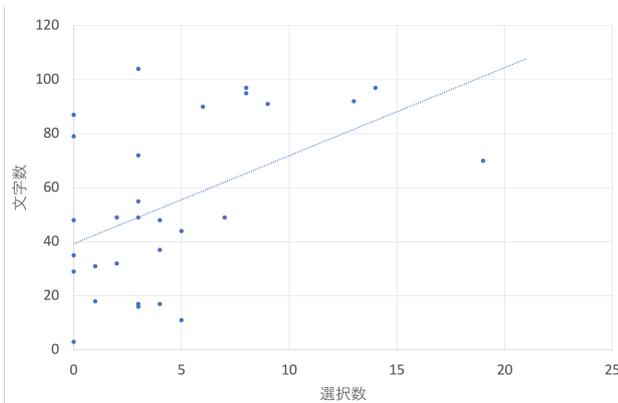


図 5: ロシア記事について文字数と選択数の相関関係 (Hawk 群)

事として、画像生成 AI に関するニュース記事（以下、AI 記事と表記）を対象に調査を行った。実験対象とした記事は「画像生成 AI 「Midjourney」の描いた絵が美術品評会で 1 位を取ってしまい人間のアーティストが激怒」\*5であり、獲得されたコメント数は 338 件であった。Hawk 群に提示する関連記事は「AI 画家「midjourney」に早速の活用例 Twitter 小説「ニンジャスレイヤー」の挿絵作りで活躍」\*6および「絵柄の特徴を学びイラストを生成する AI 「mimic」が登場 “自作発言” や “他人の絵の悪用”などを心配する声も」\*7である。調査協力者は 300 人であり、そのうち Ant 群が 151 人、Hawk 群が 149 人であった。

得られた回答のうち、回答者のコメントについては

\*5<https://gigazine.net/news/20220901-midjourney-win-fine-arts-competition/> (2024/2/27 確認)

\*6<https://www.itmedia.co.jp/news/articles/2208/05/news103.html> (2024/2/27 確認)

\*7<https://nlab.itmedia.co.jp/nl/articles/2208/29/news164.html> (2024/2/27 確認)

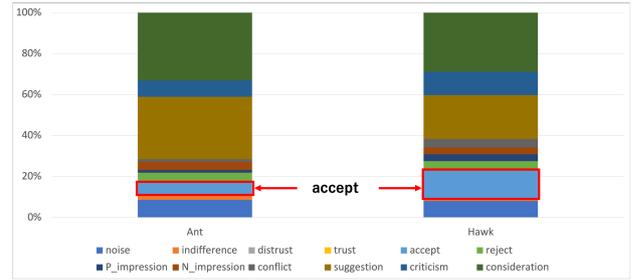


図 6: ラベルの構成比 (AI 記事)

表 3: ブックマークコメントを選択した数 (AI 記事)

選択されたユーザ	Ant	選択されたユーザ	Hawk
user_1	7	user_1	4
user_2	3	user_11	4
user_3	3	user_12	4
user_4	3	user_13	3
user_5	3	user_14	3
user_6	3	user_15	3
user_7	2	user_2	2
user_8	2	user_7	2
user_9	2	user_8	2
user_10	2	user_9	2

4.2 節と同様に表 1 の付与基準に基づいてラベルを付与した。Ant 群と Hawk 群に付与したラベルの構成比の特徴として、Hawk 群が Ant 群より accept が約 4.5 倍多いという点が挙げられる (図 6)。

実験者に選択されたブックマークコメントの割合を算出するにあたり、AI 記事ではブックマークコメントの数が多くブックマークコメントを選択した実験者にばらつきが生じたため、上位 50 コメントを算出した。Ant 群の上位 3 位は user\_1, user\_2, user\_3, Hawk 群は user\_1, user\_11, user\_12 が同率 1 位であった (表 3)。Ant 群は Hawk 群より user\_1 を選択した実験者数が約 2 倍であった。Hawk 群はブックマークコメントを選択した数に大きな差はなかった。

## 5 議論

4.2 節におけるコメント長と選出率の結果から、群に関わらずより長いコメントが共感できるコメントとして選出されやすい傾向が確認された。この結果は、本調査を行う際にコメント長によるデータ数の絞り込みが可能であることを示唆している。ソーシャルメディアに投稿されるコメントは、最投稿数や高評価数など、複数の観点で評価を行う仕組みが存在する。コメント長と選定数が正の相関にあることは、閲覧者が共感するかどうかという観点でデータを観察した際のデータ選定の一助と

なる。4.3 節におけるコメント長と選出率の結果は、選択先が分散する結果となった。これは獲得されたブックマークコメントの多さに起因すると考えられる。そのため、本調査ではテキスト分類技術などを通して、類似するコメント群から代表するコメントを選出するなど、コメント数を絞る工夫が必要となる。

ラベルの構成比に関しては、4.2 節、4.3 節のいずれのケースにおいても Ant 群と Hawk 群で現れやすいラベルに差が観察された。この結果は、本手法によって提示されるニュース記事のコメントが、意見形成に影響を及ぼしている可能性がある。そのため、得られた回答の意味的な分類を通して、提示されたコメントの影響について精査する必要がある。

## 6 おわりに

本研究は、web ニュースを介した世論形成において、ソーシャルメディア上に投稿されたコメントが、コメントの閲覧者に及ぼす影響を明らかにするため、閲覧するコメントの量とコメントを投稿する発信者の背景情報に着目した調査を行った。調査対象のニュース記事に対して付与されたコメントを獲得し、それぞれ異なる視点からコメントを提示する実験を実施した。実験では、(1) 獲得されたコメント群と関連記事、および関連記事に付与されたコメントを観察する群と、(2) 獲得されたコメント群と、コメントを付与したアカウントが過去に投稿したコメントを観察する群、の 2 種類に分けた。予備調査の結果、実験参加者が「共感したコメント」として選出したコメントは、コメントの文字数と正の相関があった。また、実験参加者がニュースについて記述したコメントを人手で分類した結果、群間で現れやすいラベルに差があることが示唆された。今後は、テキスト分類技術などを通して実験対象とするコメント数の絞り込みを行い、本調査を行う。また、得られた回答の意味的な分類を通して、提示されるコメントの影響について精査する。

## 謝辞

本研究の一部は JST RISTEX (課題番号 JP-MJRS23L2) の支援を受けた。記して謝意を表す。

## 参考文献

[1] Cacioppo, J. and Petty, R.: The Need for Cognition, *Journal of Personality and Social Psychology*, 42, 116–131, DOI: 10.1037/0022-3514.42.1.116

- (1982).
- [2] DAVISON, W. P.: The Third-Person Effect in Communication, *Public Opinion Quarterly*, 47(1), 1–15, DOI: 10.1086/268763 (1983).
- [3] Eilders, C. and Porten-Cheé, P.: Effects of on-line user comments on public opinion perception, personal opinion, and willingness to speak out: A cross-cultural comparison between Germany and South Korea, *Journal of Information Technology & Politics*, 20(3), 323–337, DOI: 10.1080/19331681.2022.2103766 (2023).
- [4] Houston, J. B., Hansen, G. J. and Nisbett, G. S.: Influence of User Comments on Perceptions of Media Bias and Third-Person Effect in Online News, *Electronic News*, 5(2), 79–92, DOI: 10.1177/1931243111407618 (2011).
- [5] Lee, E.-J., Jang, Y. J. and Chung, M.: When and How User Comments Affect News Readers' Personal Opinion: Perceived Public Opinion and Perceived News Position as Mediators, *Digital Journalism*, 9(1), 42–63, DOI: 10.1080/21670811.2020.1837638 (2021).
- [6] Newman, N., Fletcher, R., Eddy, K., Robertson, C. T. and Nielsen, R. K.: Reuters Institute digital news report 2023, Reuters Institute for the Study of Journalism, DOI: 10.60625/risj-p6es-hb13 (2023).
- [7] Toriumi, F., Sakaki, T., Kobayashi, T. and Yoshida, M.: Anti-vaccine rabbit hole leads to political representation: the case of Twitter in Japan, *Journal of Computational Social Science*, DOI: 10.1007/s42001-023-00241-8 (2024).
- [8] 浅谷公威, 鳥海不二夫, 大橋弘忠: コミュニティ間における多面性と意見形成, *人工知能学会論文誌*, 30(5), 658–666, DOI: 10.1527/tjsai.30.658 (2015).
- [9] 宇野毅明, 武富有香, 小林亮太, 橋本隆子, 久保山哲二, 申吉浩: 多様性の解析を用いたニュース記事に対するコメント集合の分析, *じんもんこん 2022 論文集*, 207–212 (2022).
- [10] 大谷卓史: SNS は世論を製造するか?, *情報管理*, 57(6), 420–422, DOI: 10.1241/johokanri.57.420 (2014).
- [11] 田島佳征, 畔柳昭雄: 「海離れ」記事への書き込みコメントから見た海水浴に対する要因把握, *環境情報科学論文集*, 35 245–249, DOI: 10.11492/ceispapers.ceis35.0\_245 (2021).

# ニュースの内容とコメントの類似度が 投稿するコメントに与える影響の分析

## Effect Analysis of the Similarity of News and Posted Comments on New Comments

西原陽子<sup>1\*</sup> Sheng Wang<sup>2</sup> Junjie Shan<sup>3</sup>  
Yoko Nishihara<sup>1</sup> Sheng Wang<sup>2</sup> Junjie Shan<sup>3</sup>

<sup>1</sup> 立命館大学情報理工学部

<sup>1</sup> College of Information Science and Engineering, Ritsumeikan University

<sup>2</sup> 立命館大学大学院情報理工学研究科

<sup>2</sup> Graduate School of Information Science and Engineering, Ritsumeikan University

<sup>3</sup> 立命館グローバルイノベーション研究機構

<sup>3</sup> Ritsumeikan Global and Innovation Research Organization

**Abstract:** Comments on news sites play an important role in understanding the news. On the other hand, many comments are not necessarily highly relevant to the news. Comments with low relevance may hinder the understanding of the news. We hypothesized that new comments would be more relevant to the news if the comments that had already been posted showed high relevance to the news. In this paper, we conducted an experiment in which we evaluated the relevance between news and comments in terms of similarity, and clarified the effect of the similarity of comments on newly posted comments. In the experiment, we used 20 articles in four categories from Yahoo! News. We used Japanese-RoBERTa to obtain embedding vectors for the news articles and their comments. The similarity between the news embedding and each comment's embedding was calculated using the cosine similarity. The participants were presented with comments with high similarity and comments with low similarity, and asked to write a new comment after reading the news and the comments. The experimental results showed that the similarity of the new comment was affected by the similarity between the news and the presented comments.

## 1 はじめに

オンラインニュースサイトやソーシャルネットワークワーキングサイト (SNS) では、ニュースや個人の投稿に対しコメントを書き込むことが一般的に行われている。コメントには個人の意見が含まれており、ニュースや投稿に対し他者が持つ意見を容易に知ることができるようになっており、ニュースの理解に役立っている。Stroudらの研究により、米国に住む人の55.3%の人がニュースに対してコメントを読み書きする、また24.6%の人がコメントは書かないが読むことが明らかになっている [5]。ニュースの理解において、コメントの重要性が高くなっている。

一方で書き込まれるコメントの中には、ニュースとの関連が高いものもあれば、高くないものも含まれて

いる。関連が高いコメントはニュースに対する示唆を与え、より深い考察や議論へと発展させる可能性を持つ。反対に関連が高くないコメントはニュースを読む人を混乱させ、他の人の意見を誤解させるなどの影響を与える可能性がある。関連が高いコメントと高くないコメントが混在していると、ニュースやコメントを読む上で支障が生じると考えられる。

フィルタリングやランキングにより、関連が高くないコメントを上位に表示しない手法が存在するが、ニュースに対して書かれるコメントは、関連が高くないものは少ないほうがニュースの理解には望ましい。関連が高くないコメントがあまり書かれられないようにするにはどうすればよいか。著者らは、既に書かれたコメント欄に関連が高いコメントが表示されていれば、新しく書かれるコメントも関連が高くなるのではないかと考えた。そこで本論文では、ニュースとコメントの関連

\*立命館大学情報理工学部, 滋賀県草津市野路東 1-1-1,  
nishihara@fc.ritsumeikan.ac.jp

度を類似度で評価し、提示される既存コメントの類似度が新たに投稿するコメントに与える影響を実験により明らかにする。

## 1.1 既存研究

関連が高いコメントと高くないコメントを分類する研究は既存研究で行われている。例えば、MozafariらはSNS上のニュース記事に対するコメントの類似度を評価することで、関連が高いコメントと関連が低いコメントを分類する手法を提案している [4]。手法を用いた評価実験の結果から、関連が高いコメントは客観的でニュース内容に関する事が書かれている事が多く、高くないコメントは主観的でニュース内容に関する事が書かれていない事が多い事が明らかになっている。また、Kolhatkarらはニュースに対するコメントを建設的なものとそれ以外に分類する手法を提案している [3]。彼らの研究で建設的なコメントは、記事と関連し、かつ感情的な反応を誘発しない明確な議論となるものである。手法を用いた評価実験において、コメントの建設性と毒性（汚い言葉遣い・攻撃的なコメント・ヘイトスピーチなど）との関連も分析しており、分析の結果、建設性が高いからといって毒性が低いわけではなく、建設的なコメントと非建設的なコメントの間で毒性の差は見られないことを明らかにしている。このように、関連が高いコメントと高くないコメントを分類する研究は存在するが、関連が高いコメントや低いコメントを提示したときに、新たに書かれるコメントのニュースとの関連がどうなるかを調べた研究は少ない。本論文では同一ニュースに対し、類似度が異なるコメント群を提示することで、新たに書かれるコメントの類似度がどのようになるかを実験で明らかにする。

## 2 分析手法

本論文では、ニュースとコメントの類似度を算出し、類似度が高いコメント群と低いコメント群を作成して実験を行い、実験結果を分析する。はじめに、ニュースとコメントの類似度を算出する方法について説明する。

図1に、類似度が高いコメント群と低いコメント群を作成する流れを示す。ニュースの記事とニュースに付与されたコメント集合を入力すると、それぞれの埋め込み表現を獲得する。ニュース記事の埋め込み表現と各コメントの埋め込み表現の類似度を算出し、類似度が高い順にコメントを並び替える。コメントの類似度の閾値以上のものを類似度が高いコメント群、閾値以下のものを類似度が低いコメント群とする。

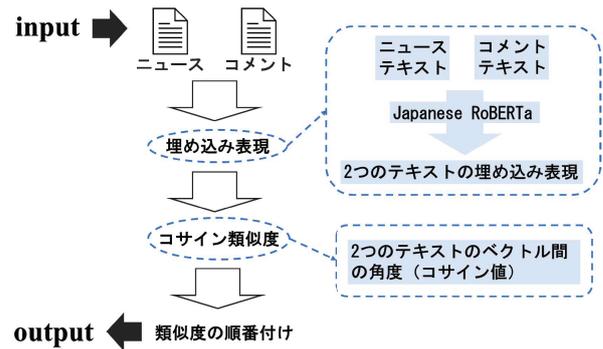


図1: 類似度が高いコメント群と低いコメント群を作成する流れ。

## 2.1 埋め込み表現を獲得するモデルの選定

本研究ではニュース記事とコメントの関連度を類似度で評価する。このため関連度を類似度で評価できる適切なモデルを選定するために予備実験を行った。

実験でははじめに、ニュースとコメントの関連度について被験者アンケートを取った。複数の埋め込み表現のモデルを用意し、ニュースとコメントの埋め込み表現を獲得し、類似度を算出した。最後に関連度と類似度の決定係数を算出した。

被験者アンケートで使用したニュースはYahoo!ニュースに掲載された3つのニュースであった。ニュースのジャンルは国際であった。各ニュースに付与されたコメントを無作為に90件ずつ抽出してアンケートに用いた。被験者は全部で30名で、20代の大学生であった。被験者にはニュースとコメントを読んでもらい、コメントのニュースとの関連の有無の程度を4段階で評価してもらった。4段階評価の0が全く関連しない、3が強く関連するとした。1つのコメントを5名の被験者に評価してもらい、5名の評価値の平均を正解の関連度とした。

用意した埋め込み表現のモデルは2つあり、日本語のWikipediaで訓練したDoc2Vec [2]とJapanese-RoBERTa [1]の2つであった。類似度は2種類の方法で評価し、コサイン類似度と内積の2つであった。

実験結果を説明する。表1に各モデルと各類似度計算方法での決定係数を示す。決定係数が最も高いのは、Japanese-RoBERTaとコサイン類似度を用いた場合であった。このため本論文ではJapanese-RoBERTaとコサイン類似度でニュースとコメントの関連度を評価する。

表 1: モデルと類似度計算方法ごとの決定係数.

モデル	類似度計算方法	決定係数
Wikipedia-Doc2vec	コサイン	0.0997
	内積	0.1345
Japanese-RoBERTa	コサイン	0.1608
	内積	0.0852

### 3 実験

同一ニュースに対して類似度の異なるコメント群を提示することにより、新たに投稿されるコメントの類似度がどのようになるかを実験により調べた。

#### 3.1 実験目的

以下の2点を明らかにすることを目的とした。

1. ニュースに対して新たにコメントが投稿されるとき、既に書かれたコメントの内容を参考にしているか。
  2. 既に書かれたコメントの内容を参考にするとき、新たに投稿されるコメントと既に書かれたコメントの類似度は高くなるか。
1. については、新たに書かれるコメントが提示するコメントの影響を受けるかを調べるためには、被験者が提示されたコメントを参考にしているかどうかを知る必要があると考えたため、実験の目的にいった。2. については、1. が成り立つ場合に新たに書かれるコメントが提示されるコメントの影響をどの程度受けるかを知るため、実験の目的に入れた。

#### 3.2 実験手順

実験は以下の手順で行った。

1. 先に示した分析手法を用い、ニュースと各コメントの類似度を算出し、類似度が高い群と低い群に分ける。それぞれの群から20件のコメントを抽出する。
2. 実験者はニュースと20件のコメントを提示する。被験者は、ニュースについての新たなコメントを書く。
3. 実験者は得られたコメントと提示コメントの類似度を評価する。新しく得られたコメントは、提示コメントの影響を受けるかを評価する。類似度が閾値以上であるコメントを抽出する。

4. 新しく得られたコメントの中から、類似度が閾値以上であるコメントを抽出する。実験者は抽出されたコメントとニュースの類似度を算出する。

手順の詳細を説明する。手順1において、用いたニュースはYahoo!ニュースの「国際」「国内」「エンタメ」「スポーツ」の4つのカテゴリに属する、それぞれ5件であり、合計20件であった。2023年の秋に行われた実験実施時期から3ヶ月以内に発表され、既に投稿されたコメントが500件以上あるニュースを用いた。ニュースと各コメントの類似度を算出し、類似度が0.8以上であったものを類似度が高いとし、類似度が0.6以下であったものを類似度が低いとした。類似度がそれぞれ閾値を満たすコメントの中からランダムに20件ずつコメントを抽出した。提示されたコメントの平均文字数は約90文字であった。

手順2において、実験者はYahoo!クラウドソーシングで実験を行った。実験者は1つのニュースと20件のコメントを提示し、被験者にニュースとコメントを読んだ上で自分の新たなコメントを書いてくださいと指示した。被験者はニュースについての新たなコメントを書き、5つのニュースで先の作業を繰り返した。実験に参加した被験者の総数は600名であった。ニュースやコメントを読んだことを確認するため、被験者にはニュース内容に関するクイズを解いてもらい、クイズに正解したコメントのみを収集した。さらに、提示コメントと同じコメント、文字数が少ないコメント(20文字未満)、意味のない文字列の羅列となっているコメントも適切に書かれなかったと判断し、分析の対象外とした。

手順3において、得られたコメントと提示されたコメントの類似度を評価する。類似度が閾値以上であれば、提示されたコメントを参考にして書かれたと評価する。最後に手順4において、書かれたコメントとニュースの類似度を評価する。

#### 3.3 実験結果

初めに、分析対象となったコメント数を示す。得られたコメントから分析対象外のコメントを除いた結果、類似度が高いコメントを提示した群では1,420件、類似度が低いコメントを提示した群では1,404件のコメントが得られた。

続いて、得られたコメントとニュースの類似度を算出したところ、類似度が高いコメントを提示した群では平均0.609、類似度が低いコメントを提示した群では平均0.608となった。2つの群の間で類似度の差は見られなかった。

新たに得られたコメントと提示したコメントの類似度の平均値を算出し、平均値を閾値としてコメントを

抽出することで提示コメントを参考にして書かれたコメントを抽出した。抽出されたコメントとニュースの類似度を算出したところ、類似度が高いコメントを提示した群では平均 0.694、類似度が低いコメントを提示した群では平均 0.610 となった。提示コメントを参考にしてコメントを書いた場合、類似度が高いコメントを提示する方が書かれるコメントの類似度が高くなった ( $p < 0.05$ )。

### 3.4 考察

初めに得られたコメントとニュースとの類似度について考察する。類似度が高いコメントを提示した群と類似度が低いコメントを提示した群を比較したところ、コメントとニュースの類似度に差は見られなかった。この結果は、被験者はニュースとコメントを読んだ上で自分のコメントを書くように指示されても、既に投稿されたコメントの影響は余り受けない被験者が一定以上いるということを示している。

続いて、提示コメントを参考にして書かれたと思われるコメントとニュースとの類似度について考察する。2つの群を比較したところ、類似度が高いコメントを提示した群の方が新たに書かれるコメントとニュースの類似度が高くなった。この結果は、自分のコメントを書くときに既に投稿されたコメントを参考にする場合は、ニュースに関連するコメントを書くことが多いということを示している。

以上より、新たな投稿がコメントされるときに、既に投稿されたコメントを参考にできるようにすることで、ニュースと関連するコメントが投稿される可能性が示された。

## 4 おわりに

本論文ではニュースとコメントの関連度を類似度で評価し、提示される既存コメントの類似度が新たに投稿されるコメントに与える影響を明らかにする実験を行った。類似度は Japanese-RoBERTa による埋め込み表現のコサイン類似度で評価し、Yahoo!ニュースの4つのカテゴリに属する20件の記事を選んで実験に用いた。実験では被験者に類似度の高いコメントと低いコメントを提示し、ニュースとコメントを読んだ上で新たなコメントを書いてもらった。実験の結果、コメントを新たに書く場合に既に書かれたコメントの影響を受けないことが多いと分かった。そして、提示されたコメントを参考にして新たなコメントを書く場合、提示されたコメントのニュースとの類似度に影響を受け、高いコメントを提示された群は新たなコメントもニュースとの類似度が高くなることが分かった。

## 謝辞

本研究の一部は科研費(22H03708)と立命館グローバルイノベーション研究機構の支援を受けて行われました。記して謝意を申し上げます。

## 参考文献

- [1] 趙天雨, 沢田慶: 日本語自然言語処理における事前学習モデルの公開, 第93回言語・音声理解と対話処理研究会, pp. 169-170 (2021)
- [2] Andrew M. Dai and Christopher Olah and Quoc V. Le: Document Embedding with Paragraph Vectors, arXiv, 1507.07998 (2015)
- [3] Varada Kolhatkar and Maite Taboada. Constructive Language in News Comments. In Proceedings of the first workshop on abusive language online, pp. 11-17 (2017)
- [4] M. Mozafari, R. Farahbakhsh and N. Crespi: Content Similarity Analysis of Written Comments under Posts in Social Media, 2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS), pp. 158-165 (2019)
- [5] Natalie Jomini Stroud, Emily Van Duyn, and Cynthia Peacock: News Commenters and News Comment Readers. Engaging News Project, pp. 1-21 (2016)