

図 1. 評価グリッド法自動化システムの全体図

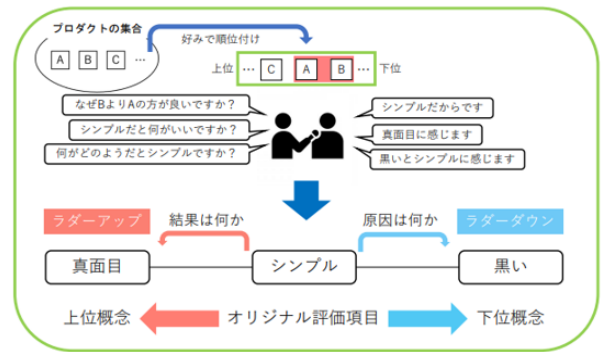


図 2. 評価グリッド法の概要

を目的とする。具体的には、文脈を考慮したベクトル化ができ短文にも適しているとされる BERTopic と感情分析を併用することによって手作業に近い分類を可能にすることを目指す。

## 2. 先行研究

### 2.1. 評価グリッド法に関する研究

評価グリッド法は讃井らが提案したインタビュー調査の一手法であり[1]、評価構造を明らかにすることにより、対象物の価値や印象、またそれらを得るための物理的な要因を明確にすることができる。

評価グリッド法の概要を図 2 に示す。手順は以下のとおりである。まず、比較対象となる複数の刺激を用意し、参加者にはそれらを嗜好度順で並べてもらう。インタビュアーは、参加者になぜそのように判断したか理由を尋ね、得られた評価項目に対しラダリングと呼ばれる方法でさらに評価項目を引き出していく。抽象的な価値判断や心理的な価値といった上位概念を引き出すラダーアップと、具体的かつ客観的な判断や物理的な状態といった下位概念を引き出すラダーダウンの質問を繰り返していく。最後に、得られた評価項目を階層構造として整理したものが評価構造である。我々は、ESV や、インタビューから可視化まですべてを自動化することを目的として大規模言語モデルを用いた自動化システムの開発を行っている[4,5]。この際に、複数人が回答した評価項目の表記ゆれや類似項目の統合が必要であり、手作業では人的・時間的負荷が高いことが課題である。

### 2.2. 文クラスタリングに関する研究

文や文書のクラスタリングにおいては、まず文を分散表現(ベクトル化)にし、そのベクトル化した文を、クラスタリング手法を用いて分類するのが大まかな流れである。文のベクトル化には TF-IDF や Doc2Vec などが、クラスタリング手法には K-means などが使われることが多いが、文が短いほどノイズやスパース性の問題から有効性を失うため困難な問題であるとされている。Salih ら[6]は、ベクトル化に TF-IDF、クラスタ

リングに K-means とウォード法を用いた際の比較をしているが、あらゆる種類のデータセットのクラスタリングに最適に機能する一般的なアルゴリズムはほとんどないと結論付けている。また近年では BERT など大規模言語モデルを用いた文脈化埋め込みが登場している。Sentence-BERT(SBERT)[7]は高精度の文章ベクトルを生成できるよう BERT をファインチューニングしたモデルであり、複数文書のクラスタリングをする場合や複数文書を時系列データとして入力する場合において BERT より優れている。しかし、評価グリッド法で収集する評価項目のように極度に短い文、かつ比較的小規模なデータセットのクラスタリングは困難であり、これらについての研究は十分に行われていない。

また、クラスタリングの一種であるトピックモデルは様々なデータに隠れた潜在的なトピックを効率的に抽出する上で有用な手法である。LDA などの従来のトピックモデルは、文章中に各単語が何回登場したかを数え作成した行列をベースとした手法であるため、単語同士のつながりが形成する文章の文脈や前後関係を必ずしも活用しきれないことが欠点としてあげられる。これに対して、BERTopic[8]はトピックの抽出に文章の文脈を考慮したトピックモデリングを行うフレームワークを提供しており、今までに文書分類やマーケティングに応用されている。Udup ら[9]は短文トピックモデリングにおける Gibbs Sampling for Dirichlet Multinomial Mixture (GSDMM)と BERTopic の比較を行っており、GSDMM に比べ BERTopic のほうが、コーヒーレンススコアが高いという結果を示している。

### 2.3. 感情分析に関する研究

感情の極性については感情分析が利用できる。感情分析とは与えられた情報を感情に分類しどのような意見なのかを分析するもので、評価グリッド法などと同じくマーケティングや意思決定の支援ができる。感情分析には主に 2 通りの方法があり、1 つ目は感情語辞書を使う方法、2 つ目は機械学習を活用する方法である。圓谷ら[10]は日本語学習済み BERT モデルをファインチューニングした感情分析 BERT を作成し、分散

表現の抽出に用いている。このような感情分析タスクでファインチューニングされたモデルは Huggingface Transformers で提供されているものも多くあり、日本語データセットで学習されたものや、3つの感情ラベル（ポジティブ・ネガティブ・ニュートラル）で学習されたものなどがある。

よって本研究では、評価グリッド法自動化を目的とし、BERTopic と感情分析との組み合わせによって手作業に近い分類を可能にする評価項目のカテゴリ生成に取り組む。

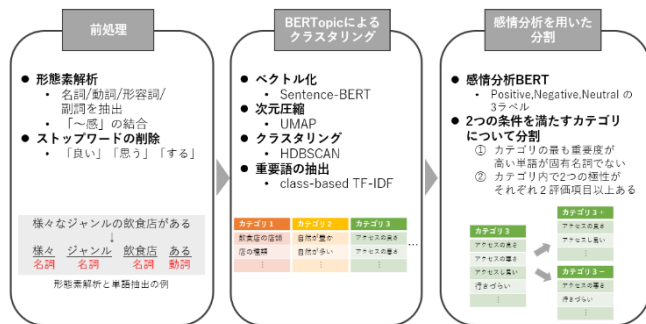


図 3. 提案手法の概要図

### 3. 提案手法

#### 3.1. 概要

本研究の評価項目をカテゴリ分けするための手法について説明する。概要を図3に示す。文章の文脈を考慮したトピックモデリングである BERTopic で文脈によるカテゴリ分類を行い、その後、感情分析によるカテゴリ分割を行うことで、評価項目を自動でカテゴリ分けする。

#### 3.2. 前処理

収集した評価項目は最初に前処理を行う。前処理は、記号の削除、文字列の正規化、重複項目の削除、形態素解析で単語単位に分割、必要な単語の抽出、ストップワードの削除の順で行う。形態素解析は日本語解析ツール MeCab を用い、システム辞書に mecab-ipadic-NEologd 辞書を用いる。評価項目の分類に必要な単語の抽出のため、助詞などの意味を持たない単語を除外する。抽出した具体的な品詞は、名詞（一般、固有名詞、サ変接続、形容動詞語幹）、動詞（自立）、形容詞（自立）、副詞で単語の活用形を原型に統一して抽出する。また「臨場感」など形態素解析によって「○○」と「感」に分割されてしまうものは「○○感」に統一する。「感」がつく複合語は身体感覚・心理感情・印象のいずれかであるため、「感」は上位的抽象的概念を表す体言であるとされているからである[11]。最後に、出現回数が多いが関連性が小さい単語「良い」、「思う」、「する」をストップワードとして削除する。

#### 3.3. BERTopic を用いたクラスタリング

前処理をした評価項目に対し、BERTopic を用いてクラスタリングを行う。まず、評価項目のベクトル化には SBERT を用いる。日本語事前学習済み BERT モデルをファインチューニングして作成された SBERT を使用する。次に、UMAP を用いて文章ベクトルの次元数を圧縮する。圧縮次元は BERTopic のデフォルト値である5次元とする。次に、HDBSCAN を用いてクラスタリングを行う。最小クラスタサイズは3とする。評価項目を手作業でカテゴリ分けする際には各カテゴリのサイズなどは考慮しないためである。最後に、各クラスタに対して class-based TF-IDF を用いて重要語の抽出を行う。これは、TF-IDF で行われている単語頻度の計算を、文書単位ではなくクラスタ単位で行うようにしたものである。なお、ここでクラスタリングした各クラスタを「カテゴリ」と呼ぶこととする。

#### 3.4. 感情分析を用いた分割

BERTopic でクラスタリングしたカテゴリに対し、感情分析によってさらに分割を行う。Huggingface Transformers で提供されている感情分析 BERT を用いて、各カテゴリ内の評価項目それぞれをポジティブ・ネガティブ・ニュートラルの3ラベルに分類する。次に、以下2つの条件を満たすカテゴリについて分割を行う。1つ目は、カテゴリの最も重要度が高い単語が固有名詞でないこと。固有名詞がキーワードとなってまとまっているカテゴリは分割する必要がないからである。2つ目は、カテゴリ内で2つの極性（ポジティブーネガティブ、もしくはネガティブーニュートラル）がそれぞれ2評価項目以上あること。これは評価項目が1つになるカテゴリはカテゴリとして機能しないためである。2つの極性で分割した際に3つ目の極性の評価項目が1つ残る場合はコサイン類似度が近いほうの極性に含めることとする。

### 4. 実データによる提案手法の検証

#### 4.1. データの概要

実際に評価グリッド法を用いて収集した3ドメインの評価項目を用いた。1つ目はオンラインライブについて「ライブ感」を最上位概念としたラダーダウンで収集したデータで、クラウドソーシングによって実施した。2つ目は観光地である兵庫県の三宮と朝来について魅力的なところと魅力的でないところを聴取したデータである。3つ目はネクタイの好みに関する実験で、従来の評価グリッド法で実施したデータである。それぞれ研究者の手作業により複数のカテゴリに分類、カテゴリ名を付与した。各データの評価項目数と平均単語数、手作業時のカテゴリ数を表1に示す。平均単語数は前処理後の単語数で計算したものである。

表 1. データセットの詳細

評価項目データセット	評価項目数	平均単語数	手作業でのカテゴリ数
オンラインライブ	681	9.8	41
三宮・朝来	772	2.9	110
ネクタイ	471	1.1	115

表 2. 各手法でのカテゴリ数

	手作業でのカテゴリ数	G-means	スペクトラルクラスタリング	本手法 (BERTopic)
オンラインライブ	41	57	65	64
三宮・朝来	110	70	90	99
ネクタイ	115	31	31	72

## 4.2. カテゴリ生成の結果・考察

### 4.2.1. クラスタリング

3手法を用いてカテゴリ生成の結果を比較する。最もよく利用される TF-IDF と G-means, TF-IDF と連結性を重視したスペクトラルクラスタリング, 本手法 (BERTopic) の3手法である。G-means は K-means を拡張したもので K-means のパラメータであったクラスター数を自動で決定するアルゴリズムである。G-means においてはクラスタリングした際に要素が1つになるクラスター (カテゴリ) が発生したが, 評価グリッド法において評価項目が1つのカテゴリは意味をなさず手作業においてもそのようなカテゴリは作成しないため, 除外することとする。スペクトラルクラスタリングはクラスター数の自動決定はされないため, 予め G-means でクラスター数を予測しそのクラスター数を指定することでクラスタリングを行うこととする。

まず, それぞれのカテゴリ数は表3のとおりである。どのデータでも G-means より BERTopic のほうが, カ

テゴリ数が多くなるという結果になった。評価グリッド法における評価項目のカテゴリ化ではカテゴリ数が多くなる傾向にあり, また分割するよりも統合するほうが人的・時間的負荷が少ないことから多くなるほうが良いと考えられる。三宮・朝来データとネクタイデータにおいては手作業の正解データと比べると減っているものの BERTopic で大きく改善できていると言える。

次に, 各クラスタリング結果において, 正解データと比較することで精度の評価を行った。カテゴリの解釈可否結果を正解データのカテゴリと比較し, (i) 概ね一致, (ii) 正解データの複数 (2~3) のカテゴリが統合, (iii) 正解データのカテゴリが分割されたカテゴリ, (iv) 正解データにはない新規カテゴリとして使用できる, (v) 解釈不可 (4つ以上のカテゴリが含まれているなど) の5つに分類した。また, これらは研究者3名による合議で決定した。なお, BERTopic のみ自動的にノイズカテゴリが1つ生成されるが, それは

表 3. 各データのクラスタリング結果

#### a. オンラインライブデータ

	G-means		スペクトラルクラスタリング		BERTopic	
	カテゴリ数の割合 [%]	評価項目数の割合 [%]	カテゴリ数の割合 [%]	評価項目数の割合 [%]	カテゴリ数の割合 [%]	評価項目数の割合 [%]
(i) 概ね一致	33.3	26.4	55.4	36.6	51.6	42.1
(ii) 複数 (2~3) のカテゴリが統合	45.6	52.7	21.5	14.8	25.0	31.8
(iii) 分割されたカテゴリ	0.0	0.0	6.2	3.1	9.4	4.6
(iv) 新規カテゴリ	1.8	1.5	4.6	1.6	7.8	4.7
(v) 解釈不可	19.3	19.4	12.3	43.9	6.3	16.9

#### b. 三宮・朝来データ

	G-means		スペクトラルクラスタリング		BERTopic	
	カテゴリ数の割合 [%]	評価項目数の割合 [%]	カテゴリ数の割合 [%]	評価項目数の割合 [%]	カテゴリ数の割合 [%]	評価項目数の割合 [%]
(i) 概ね一致	12.9	8.4	53.3	32.6	61.6	47.7
(ii) 複数 (2~3) のカテゴリが統合	50.0	46.8	22.2	19.8	22.2	28.4
(iii) 分割されたカテゴリ	1.4	0.6	6.7	3.2	5.1	4.4
(iv) 新規カテゴリ	0.0	0.0	7.8	3.4	4.0	3.0
(v) 解釈不可	35.7	44.2	10.0	40.9	7.1	16.6

#### c. ネクタイデータ

	G-means		スペクトラルクラスタリング		BERTopic	
	カテゴリ数の割合 [%]	評価項目数の割合 [%]	カテゴリ数の割合 [%]	評価項目数の割合 [%]	カテゴリ数の割合 [%]	評価項目数の割合 [%]
(i) 概ね一致	9.7	4.9	58.1	22.1	59.7	50.3
(ii) 複数 (2~3) のカテゴリが統合	25.8	15.5	25.8	12.5	26.4	22.7
(iii) 分割されたカテゴリ	0.0	0.0	0.0	0.0	0.0	0.0
(iv) 新規カテゴリ	0.0	0.0	3.2	0.8	1.4	0.8
(v) 解釈不可	64.5	79.6	12.9	64.5	12.5	26.1



表 4. 三宮・朝来データのクラスタリング結果の一例

	手作業	クラスタリング (カテゴリ化)
(i)	治安が悪い 治安が悪い所がある 治安が良くなさそう 夜の治安が悪い時がある 裏道に一步入った所が治安が悪そう 一步細い道に入ると治安が良くない 治安が良いとは言えない	治安が悪い所がある 治安が良くなさそう 夜の治安が悪い時がある 裏道に一步入った所が治安が悪そう 一步細い道に入ると治安が良くない 治安が良いとは言えない
	新鮮な気分 新鮮 新鮮な気持ち 新鮮に感じた 新鮮な気持ちになる	新鮮な気分 新鮮 新鮮な気持ち 新鮮に感じた
(ii)	アクセスの良さ 交通の便が良い 乗り換えが便利 いろんな電車が集まっている JR 阪急阪神などいろんな行き方がある 行きやすい	交通の便がいい 交通の便が良い 交通手段がたくさんある
	アクセスの悪さ 行くのに時間がかかる 行くのが不便 交通の便が悪い電車が無い	交通の便があまりよくない 交通の便が悪い
	飽きしてしまう すぐ飽きてしまうかもしれない 飽きが来る 同じ町並みだと飽きが来る もう一回行こうと思わない	すぐ飽きてしまうかもしれない 飽きが来る 何回に行っても飽きない
	飽き来ない 何回に行っても飽きない 飽きない いつ来ても飽きない	飽きない いつ来ても飽きない
(iii)	食の楽しみがある 食べ物おいしい ちょっとご飯を食べて帰ることができる おいしいご飯が食べれそう いろんな食文化を楽しむことができる 洋食屋さんがいっぱいある おいしい洋食屋さんがある	ちょっとご飯を食べて帰ることができる おいしいご飯が食べれそう お米や食べ物おいしい おいしいご飯を食べるのが大事なこと 洋食 昔からあるお洋食屋さん 洋食屋さんがいっぱいある おいしい洋食屋さんがある

解釈不可として計算した。

各データセットにおける、比較結果を表 3 に示す。オンラインイブデータでは、(i) の正解データのカテゴリ名と概ね一致しているカテゴリはスペクトラルクラスタリングが最も多いが、評価項目数でみると BERTopic が最も多い結果となった。(ii) の複数のカテゴリが統合しているカテゴリは、カテゴリ数・評価項目数ともにスペクトラルクラスタリングが最も少なく、ほか 2 つの手法より正確なクラスタリングができているように見える。一方で、(v) の解釈不可であった評価項目数はスペクトラルクラスタリングが最も多く、BERTopic で最も少なくなったためより多くの評価項目を分類できているのは BERTopic である。(iii) 分割されたカテゴリは BERTopic で増加していたが、これは手作業より詳細な分類ができると言える。三宮・朝来のデータセットやネクタイのデータセットでも、同様の結果となっている。ネクタイデータでは、(i) や (v) から G-means とスペクトラルクラスタリングに比べ BERTopic で大幅に向上したことがわかるが、これはカテゴリ数が原因であると考えられる。カテゴリ数が少ないと 1 カテゴリあたりに含まれる評価項目が増え解釈が難しくなるからである。ネクタイデータは 1 評価項目あたりの単語数が少ないため、TF-IDF でベクトル化する際に計算している単語の重要度が分散してしまう。そのため G-means ではカテゴリ数が減ってしまうのではないかと考えられる。またこのデータは正

解データのカテゴリ数が多く、1 カテゴリあたりの評価項目数も少ない、というように正解データが細かく分類されているため、(iii) 分割されたカテゴリは出現しなかったと考えられる。

次に、生成されたカテゴリの意味的な考察を行う。三宮・朝来データのクラスタリング結果例を表 4 に示す。適切なクラスタリングができている例として (i) の「治安が悪い」などがあり、手作業の正解データと同様のカテゴリになっている。(ii) 複数のカテゴリが統合しているカテゴリは、「アクセスの良さ」と「アクセスの悪さ」などポジティブ・ネガティブが統合されているものが多かった。これは、評価項目収集の際にポジティブ要因とネガティブ要因それぞれラダリングを行い、両方の評価項目が含まれていたのが要因の一つと考える。(iii) 分割されたカテゴリの例としては「食の楽しみがある」があり、「おいしいご飯」と「洋食屋」という解釈ができるようなカテゴリに分かれている。これは上記でも述べたように、手作業より詳細な分類ができると言え、また分割するよりも統合するほうが人的・時間的負荷が少ないことから良い分割であると考えられる。

#### 4.2.2. 感情分析による分割

ここではクラスタリング後のカテゴリに対し感情分析による分割を行った際の結果と考察を述べる。表 5 は三宮・朝来データセットにおいて BERTopic でクラスタリングしたものと、そこから感情分析を用いて分割対象となったカテゴリに対し分割を行ったものの比較結果である。三宮・朝来データのみならずどのデータセットでも (ii) が減少し (i), (iii), (v) が微増する結果となった。特に、ポジティブ・ネガティブの対になったカテゴリが多く含まれていたこのデータでは、(ii) の減少が顕著である。表 6 に三宮・朝来データにおける分割されたカテゴリの一例を示す。極性での分割ができている例が「乗り換え」と「交通の便」である。これらの例ではニュートラル・ネガティブもしくはポジティブ・ネガティブで分割されており、分割後のカテゴリをみても的確に分割できていると言える。3 つ目の「人が多い」の例ではニュートラル・ネガティブで分割されており、ネガティブカテゴリは 2 つの評価項目が異なる意味でありカテゴリとして機能しないため解釈不可となる。しかしニュートラルカ

表 5. 三宮・朝来データで感情分析による分割を行った結果

	感情分析なし		感情分析あり	
	カテゴリ数の割合 [%]	評価項目数の割合 [%]	カテゴリ数の割合 [%]	評価項目数の割合 [%]
(i) 概ね一致	61.6	47.7	60.0	51.0
(ii) 複数 (2~3) のカテゴリが統合	22.2	28.4	10.0	16.6
(iii) 分割されたカテゴリ	5.1	4.4	12.5	8.8
(iv) 新規カテゴリ	4.0	3.0	5.0	3.2
(v) 解釈不可	7.1	16.6	13.3	20.3

表 6. 三宮・朝来データにおける分割結果の例

分割前		分割後	
評価項目	label	評価項目	label
乗り換えが便利	Positive	乗り換えが便利	Positive
いろんな交通機関があるバス電車など	Neutral	いろんな交通機関があるバス電車など	Neutral
乗り換えがしやすい	Neutral	乗り換えがしやすい	Neutral
乗り換えで使うことが多い	Neutral	乗り換えで使うことが多い	Neutral
乗り換えの移動時に周辺が混雑してしまう	Negative	乗り換えの移動時に周辺が混雑してしまう	Negative
乗り換え移動時間がかかる	Negative	乗り換え移動時間がかかる	Negative
交通の便がいい	Positive	交通の便がいい	Positive
交通の便が良い	Positive	交通の便が良い	Positive
交通手段がたくさんある	Neutral	交通手段がたくさんある	Neutral
交通の便があまりよくない	Negative	交通の便があまりよくない	Negative
交通が整っていない	Negative	交通が整っていない	Negative
人もたくさんいて賑やか	Positive	人もたくさんいて賑やか	Positive
人が多い	Neutral	人が多い	Neutral
人が多い所	Neutral	人が多い所	Neutral
人も多い	Neutral	人も多い	Neutral
人が多すぎる	Negative	いろんな人がいる	Neutral
いろんな人がいる	Neutral	人が多すぎる	Negative
人があまりいない	Negative	人があまりいない	Negative

カテゴリのほうはノイズが除去されて高品質なカテゴリになったとも考えられる。また、前節でも述べたような重複するカテゴリが増加してしまうことや、ノイズとして除去されたものが解釈不可として増加してしまうため、結合処理や解釈不可の評価項目に対する再度のクラスタリングなど追加処理が必要になると考える。

## 5. まとめ

本研究では、評価グリッド法の自動化を目的として類似項目の統合が必要である評価項目のカテゴリ生成を行った。前処理を行った評価項目に対し、文脈を考慮したベクトル化ができ短文にも適しているとされるBERTopicと感情分析を組み合わせた手法を提案した。3つのドメインのデータセットを用いて従来手法との比較を行ったところ、提案手法は評価項目のカテゴリ化に適しており、また、ポジティブ・ネガティブの分割により手作業に近い分類が可能になった。このことから提案手法が評価項目のカテゴリ化に有効であることが示された。今後の課題として、重複するカテゴリが生成されていることや、解釈不可のカテゴリに対する処理が挙げられ、より精度の高い評価グリッド法の自動化に近づくと考える。

## 文 献

[1] 讚井純一郎, 乾正雄, “レパートリー グリッド発展手法による住環境評価構造の抽出—認知心理学に基づく住環境評価に関する研究 (1)—,” 日本建築学会計画系論文, no.367, pp.15-22, 1986.  
 [2] 土田義郎, 小酒祐貴, “評価グリッド法支援ツールの開発と応用,” 日本建築学会技術報告集, 14,

no.27, pp.205-208, 2008.  
 [3] 大谷俊太, 橋本翔, 杉本匡史, 長田典子, “単語の係り受け関係に基づく印象評価構造の自動構築,” 第17回日本感性工学会春季大会, 2D2-04, 2022.  
 [4] 宮嶋大輔, 張帆, 杉本匡史, 佐々木香暖, 北野泰成, 橋本翔, 長田典子, “大規模言語モデルを用いた評価グリッド法に基づくインタビュー対話システム,” 信学技報, vol.123, no.180, MVE2023-20, pp.33-38.  
 [5] M. Sugimoto, Y. Yagi, and N. Nagata, “How different tourist sites evoke different emotions: Investigation focusing on the urban and rural sites in Japan,” Human-Computer Interaction. HCII 2023. Lecture Notes in Computer Science, vol.14012, pp. 332–343, Springer, Cham, 2023.  
 [6] N. M. Salih and K. Jacksi, “Semantic document clustering using k-means algorithm and ward’s method,” in 2020 International Conference on Advanced Science and Engineering (ICOASE), pp. 1–6, IEEE, 2020.  
 [7] N. Reimers and I. Gurevych, “Sentence-bert: Sentence embeddings using siamese bert-networks,” arXiv preprint arXiv:1908.10084, 2019.  
 [8] M. Grootendorst, “Bertopic: Neural topic modeling with a class-based tf-idf procedure,” arXiv preprint arXiv:2203.05794, 2022.  
 [9] A. Udupa, K. Adarsh, A. Aravinda, N. H. Godihal, and N. Kayarvizhy, “An exploratory analysis of gsdmm and bertopic on short text topic modelling,” in 2022 Fourth International Conference on Cognitive Computing and Information Processing (CCIP), pp. 1–9, IEEE, 2022.  
 [10] 圓谷顯信, 高橋宏和, 安達由洋. “BERTによる日本語文の感情分析と話題分析,” 第84回全国大会講演論文集 2022, no.1, pp.37-38, 2022.  
 [11] 曾睿. “語構成から文構成へ: 形態素「-感」と自立語「感」との関わりから,” 国語学研究, no.56, pp.142-155, 2017.