

# 大規模言語モデルを用いた 評価グリッド法自動化のためのキーフレーズ生成

北野 泰成† 張 帆† 杉本 匡史† 橋本 翔†‡ 佐々木 香暖† 宮嶋 大輔† 長田 典子†

† 関西学院大学理工学部/感性価値創造インスティテュート 〒669-1330 兵庫県三田市学園上ヶ原 1 番

‡ 西南学院大学商学部 〒814-8511 福岡市早良区西新 6-2-92

E-mail: † { kitanott99, zhangfan, sugimoto.masashi, k-sasaki, daisukemiyajima, [nagata](mailto:nagata@kwansei.ac.jp) } @kwansei.ac.jp

**あらまし** ユーザーニーズを構造的に把握する方法として、評価グリッド法による個別インタビュー手法が知られているが、人的、時間的負荷が問題となる。本研究では評価グリッド法の自動化の一処理として、大規模言語モデルを用いたキーフレーズ生成手法を提案する。カテゴリ内にある文章の評価項目の含有率に応じて 2 種類の生成手法を使い分け、含有率が大きいカテゴリに対し大規模言語モデルを用いることでカテゴリ内外の単語情報をもとにキーフレーズを生成した。また、含有率が小さい残りのカテゴリは単語重要度に基づいてカテゴリ内からキーフレーズを選出した。既存のキーフレーズ抽出手法との結果を比較したところ、提案手法が高精度かつ適切な長さのキーフレーズを生成できることが確認された。

**キーワード** キーフレーズ生成, 評価グリッド法, 大規模言語モデル, GPT4

## Keyphrase generation for automation of the evaluation grid method using a large-scale language model

Taisei KITANO† Fan ZHANG† Masashi SUGIMOTO† Sho HASHIMOTO†‡  
Kanon SASAKI† Daisuke MIYAJIMA† Noriko NAGATA†

† School of Science and Technology / Kwansei Gakuin Institute of Kansei Value Creation, Kwansei Gakuin University 1  
Uegahara, Gakuen, Sanda-shi, Hyogo, 669-1330 Japan

‡ Faculty of Commerce, Seinan Gakuin University, Nishiara, Hukuoka-shi, Hukuoka, 814-8511 Japan

E-mail: † { kitanott99, zhangfan, sugimoto.masashi, k-sasaki, daisukemiyajima, [nagata](mailto:nagata@kwansei.ac.jp) } @kwansei.ac.jp

**Abstract** The individual interview method, by means of the evaluation grid method, is known as a technique for structurally understanding users' needs; however, it takes much time and effort to experiment and analyze. In this research, we aim to automate a step of the evaluation grid method by constructing a keyphrase-generation system using a large-scale language model. We employ two generation methods based on the number of sentences within clusters, utilizing a large language model to generate keyphrases using both intra- and extra-cluster word information. By comparing and analyzing the results of our proposed method with previous studies, we seek to gain insights into the effectiveness of our approach.

**Keywords** Keyphrase generation, Evaluation grid method, Large-scale language model, GPT4

## 1. はじめに

ユーザーニーズの多様化に伴い、個人の価値観や嗜好を具体的なデザインに反映する方法論が求められている。ユーザーニーズを発掘・抽出・把握する方法は、プロダクトやサービスにおけるパーソナライズや感性価値（経験価値、体験価値）向上に必須の技術であり、心理学的手法、マーケティングリサーチ手法、行動観察手法、感性評価手法、機械学習手法など、多様な分野で研究が進められている。従来、ユーザーニーズを把握する手法として、インタビュー調査や主観評価実験など心理学的データを用いた手法が有効とされてきた。この中でも代表的なインタビュー手法として評価グリッド法[1]がある。評価グリッド法では、インタ

ビューを通して対象刺激に対する評価理由を聴取し、その原因と結果を引き出すことで個人の評価構造を構築する手法である。抽象的な価値判断・心理的な価値から客観的な判断・物理的な状態までを階層的に整理し評価構造として可視化する。また、ラダリングと呼ばれる上位概念・下位概念を聞き出す対話手順を行うことで、効率的な評価構造の構築が可能である。この手法はその有用性から多くの研究開発に用いられているものの、人的・時間的な負荷が高いという問題がある。実施を支援するツールの開発なども試みられているが[2,3]、更なる効率化のための施策として、評価グリッドの作業全てを自動化するという試みが行われている[4]。この研究における自動化システムは図 1

に示す通り 4 つのサブシステムにより構成されている。大規模言語モデルによるインタビュー対話を通して得られた評価項目をクラスタリングした後、そのカテゴリごとに代表的な名前(キーフレーズ)を付与し、評価構造として可視化する。この中でも 3 番目の処理であるキーフレーズ生成は評価項目の集合ごとに最も適切なフレーズでタグ付けを行い、評価構造を簡潔化するという点で重要な役割を持つ。

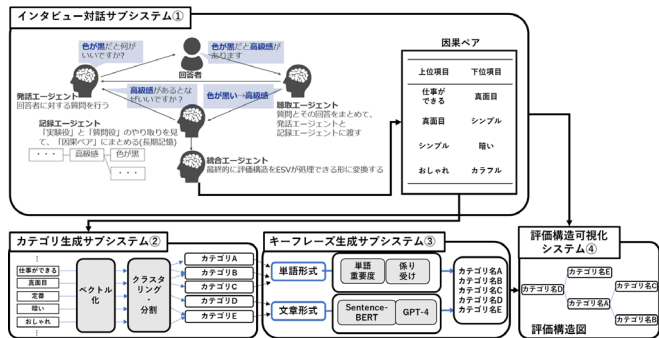


図 1 評価グリッド法自動化システム

また、この研究においてキーフレーズ生成の対象となるのはカテゴリと呼ばれる評価項目の集合である。評価項目は、主語述語を伴った文章形式のものや単語 1 つだけのものが混在している。しかし、キーフレーズ抽出・生成の既存手法では連続した文書を入力することが前提であり、本研究のデータのように集合内に文章や単語が両方含まれることが考慮されていない。

本研究では、評価グリッド法の完全自動化を目的とし、その一処理として、大規模言語モデルを用いて評価項目のカテゴリごとにキーフレーズを生成する手法を提案する。カテゴリ内の評価項目ごとに文章か否かの判定を行い、文章判定を受けた評価項目のカテゴリ内における含有率を算出する。この含有率に応じて 2 種類のキーフレーズ生成手法を使い分ける。含有率が大きいカテゴリは大規模言語モデルを用いてキーフレーズを生成し、残ったカテゴリに対しては単語重要度に基づきカテゴリ内からフレーズを選出する。

## 2. 関連研究

### 2.1. 評価グリッド法

評価グリッド法は、インタビューにより個人の持つ評価構造を階層的に表現する手法である。Kelly が臨床心理学の分野で開発したレポーター・グリッド法[5]を実務向けに改良した面接手法とされる。

図 2 に示す通り、評価グリッド法は「実験刺激に対する選好判断(インタビュー)」「ラダリング」「評価項目のカテゴリ化」の 3 つの処理に分類される。まず実験参加者に対して複数の実験刺激を嗜好度順に順

位付けしてもらい、その判断理由についてインタビューを行う。そこで得られた評価項目(回答)に対し、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を行う。その後、評価項目の中で意味が類似したものを一つのカテゴリとしてまとめ、カテゴリごとに名前を付与する。最後に、評価項目を階層構造として整理することで評価構造が構築される。

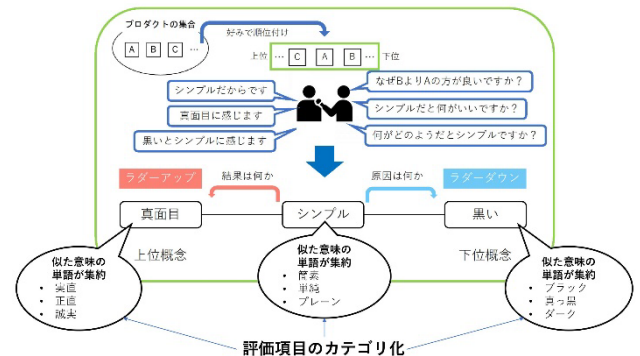


図 2 評価グリッド法の手順

この手法は、人間が持つ固有の理解や判断の仕組みにより、何を知覚しどのように評価を下しているかを把握することが可能である。また、手順が決められているため、インタビュアーにスキルや知識を要求しないこともメリットとして挙げられる。しかし、インタビューで得られた評価項目の構造化を手作業で行わなければならない、人的・時間的負荷が大きいため、大規模なデータの取得が困難である。このため評価グリッド法の全処理を自動化することで、作業をより効率的かつ低負荷で行う必要がある。評価グリッド法の工程の中でも評価項目のカテゴリ化に着目し、大規模言語モデルを用いてキーフレーズの生成を自動的に行う。

### 2.2. キーフレーズ抽出・生成

キーフレーズ抽出は自然言語処理における基本的なタスクであり、文書から最も網羅的かつ代表的な表現を識別および抽出する処理である。Tf-Idf や TextRank など多くの既存手法は、最初にキーフレーズ候補のリストを文書内から取得し、文書への重要度を数値化することでその候補をランク付けするという 2 つのステップを踏んでいる。また、言語モデルの発展に伴い、文書に現れないフレーズを再帰型ニューラルネットワーク(RNN)や chatGPT によって生成するキーフレーズ生成タスクを行った事例も確認されている[6][7]。しかし、これらの手法は入力文書であることが前提であり、文章や単語が混在した集合に対してキーフレーズを抽出・生成した例は確認されてい

い。本研究で扱う評価項目は文章と単語の両方が含まれており、カテゴリにも文章に偏ったものや単語に偏ったものがあるため、キーワードを獲得するための最適な手法がカテゴリ全体で共通していないことが考えられる。このため、文章・単語のカテゴリごとにキーワード生成を分ける必要がある。

### 3. 評価グリッド法のためのキーワード生成

#### 3.1. キーワード生成手法の概要

本研究では、評価グリッド法自動化のための一処理として、評価項目のカテゴリごとにキーワードを生成する手法を提案する。カテゴリは文章の評価項目が多いものと単語の評価項目が多いものに分かれる。カテゴリの構成要素が異なる場合、それに応じて出力されるキーワードの抽象度も変化することが望ましいため、キーワードの生成手法をカテゴリ内の構成要素に基づいて切り替える必要がある。このため、最初にカテゴリ内の評価項目ごとに文章か否かの判定を行う。文章判定の基準は、金谷[8]に基づき、1. 述語があるかどうか、2. 体言止めされていないかの2点とする。カテゴリ内における文章の評価項目の含有率を算出し、その値に応じて生成手法を切り替える。含有率が大きければ、最初に重要な評価項目を抜き出し、大規模言語モデルにより簡潔化する「文章ベースのキーワード生成」を行う。残ったカテゴリに関しては、単語重要度と係り受け解析を活用してキーワードを選出する「単語ベースのキーワード生成」を行う。なお、本研究では大規模言語モデルとして、transformer をベースとした OpenAI の生成モデル GPT4 を活用する。

#### 3.2. 文章ベースのキーワード生成

文章形式の評価項目がカテゴリ内に多く含まれる場合、出力されるキーワードも可能な限り情報が損なわれない文章形式であることが望ましい。このため、最初にカテゴリ全体の情報を最も表現している要素を EmbedRank アルゴリズム[9]によって抜き出し、GPT4[11]を用いて短く簡潔な表現に変換することでキーワードを生成する(図3)。

最初に、カテゴリ全体の情報を最も表現できる評価項目を EmbedRank アルゴリズムにより選出する。カテゴリの評価項目を全て連結した文字列をカテゴリ全体の情報と捉え、この文字列とカテゴリ内の全評価項目を分散表現に変換する。分散表現に変換したカテゴリと各評価項目のコサイン類似度を計算し、類似度最大の評価項目を抽出する。

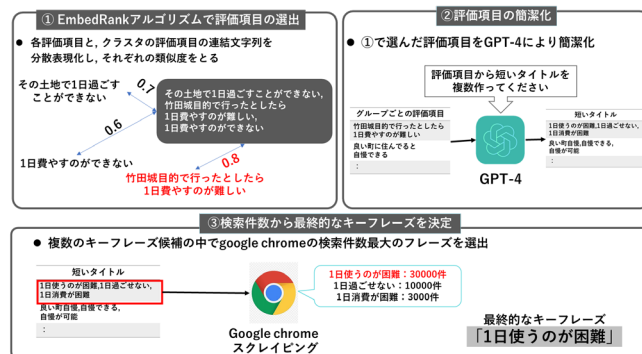


図3 文章ベースのキーワード生成手法

出力された評価項目は文章長が大きく評価構造に適した形式になっていない可能性があり、短く簡潔な表現に変換する必要がある。このため、GPT4[10]を用いて文章長の大きい評価項目を簡潔化する。出力された評価項目ごとに GPT4 に入力し、「入力の評価項目に対して短いタイトルを複数生成する」ようにプロンプトを与えることで簡潔化された表現を獲得する。

GPT4 に入力するプロンプトを図4に記載する。プロンプトは命令・制約・出力例・確認の4つから成る。命令パートでは選出された評価項目を入力し、評価項目の単語をもとに短いタイトルをカンマ区切りで複数作るように GPT4 に命令する。制約パートでは望まない出力結果を排除するために記述する。また、Chain of Thoughts[11]のプロンプト手法に基づき、モデルに問題を解かせる際にその思考過程も同時に学習させることで、タイトル生成時に段階的に試行しつつ結果を出力するよう要求する。出力例パートでは入力する評価項目とそれに対して出力が期待されるタイトル候補を提示する。最後の確認パートでは、命令パートで記述したプロンプトを再度説明し、間違った形式の結果が得られないようにする。

最後に、カテゴリごとにキーワードの候補から最適なキーワードを選出する。その表現が一般的に多用されるものか否かを選出基準とし、検索件数が多いほど一般的によく用いられる表現であると仮定する。キーワード候補をそれぞれブラウザ上で検索し、表示されたページに対して Web スクレイピングを行うことで検索件数を取得する。検索件数が最大となる表現を最終的なキーワードとする。検索件数を取得する対象ブラウザには Google Chrome を活用した。

```

<命令>
以下の<評価項目群>の評価項目ごとにそれぞれ
短いタイトルを作ってください。評価項目ごとに
複数タイトルの候補を生成し、半角カンマで区切
ってください。
<評価項目群>
{入力する文章}
</評価項目群>
</命令>

<制約>
丁寧語は使わない
文末を「～と思う」「～と思います」で締めない
ステップバイステップで考えること
:
</制約>

<例>
<評価項目>評価項目例 1</評価項目>
<回答>回答例 1</回答>
<評価項目>評価項目例 2</評価項目>
<回答>回答例 2</回答>....
</例>

```

図4 GPT4 に与えたプロンプト

### 3.3. 単語ベースのキーフレーズ生成

カテゴリに含まれる文章の割合が少ない場合、カテゴリ内の評価項目は単語、名詞句形式のものが多く考えられる。この場合は、GPT4 を用いてキーフレーズを生成するよりも、カテゴリ内で重要度の高い単語を抜き出した方がカテゴリの特徴を捉えたキーフレーズを獲得しやすいと考えられる。このため、カテゴリ内の評価項目を単語に分割し、Tf-Idf のスコアが上位の単語に係り受け解析を適用して関連の強い単語を付加することでキーフレーズを生成する(図5)。

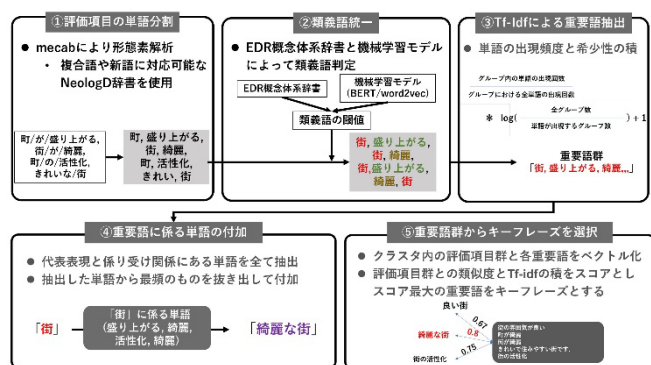


図5 単語ベースのキーフレーズ生成手法

最初に、形態素解析を行いカテゴリ内の全評価項目を単語の集合に分割する。解析器には MeCab を使

用し、対応する辞書として固有名詞や新語に対応可能な NeologD 辞書を採用する。全評価項目を単語に分割し、意味を持たない単語を除去するために単語の品詞を名詞・動詞・形容詞・副詞の4つに限定する。

次に、単語のカテゴリに類義語があれば出現頻度の大きい方に統一する。EDR 概念体系辞書から抽出した類義語・非類義語のペアから類義語判定の閾値を自動的に決定し、カテゴリ内の単語ペアの中で閾値を超えるものがあれば類義語として統一する。

類義語統一を行ったカテゴリの単語それぞれに対して Tf-Idf を適用し、単語集合の中から重要度の高い単語(重要語)を複数選出する。Tf-Idf はある文書に含まれる単語の出現頻度(TF)と、単語の希少性(IDF)の積により求められる。カテゴリ C の単語 w における Tf-Idf のスコア s は式(1)のように表される。

式 1 Tf-Idf のスコア

$$s = (C \text{ における } w \text{ の出現回数} / C \text{ における全単語の出現回数}) * (\log(\text{全カテゴリ数} / w \text{ が出現するカテゴリ数}) + 1) \dots (1)$$

その後、係り受け解析を活用することで重要語に関連する単語を抽出し、適切な形式で重要語に付与する。単語の抽出には単語間の関係をタグで表現できる日本語の形態素解析器 Ginza を用いる。カテゴリ内の評価項目を Ginza で解析し、重要語と係り受け関係にある単語を全て抽出する。抽出した単語のうち大部分を占める単語があれば重要語に付加する。

最後に、重要語群から最もキーフレーズとして適切な言葉を選択する。キーフレーズの選出にあたり、教師データを必要とせずに文書内で重要なキーフレーズを選び出す EmbedRank アルゴリズムを応用する。各重要語とカテゴリ内の評価項目群を分散表現化し、評価項目群の分散表現と重要語ごとのコサイン類似度を算出する。全重要語の評価項目群とのコサイン類似度と Tf-Idf の値の積を重要度スコアとし、最も値の大きいものをキーフレーズとする。

### 4. キーフレーズの生成結果と考察

2 種類のキーフレーズ生成手法を用いて、複数の評価項目カテゴリからキーフレーズを生成し、手作業で割り振ったキーフレーズと比較した結果を示す。また、これらの結果に対して、評価項目ごとのデータの観点、キーフレーズの手法ごとの観点から考察を行う。

表 1 キーフレーズ生成における各手法の正解率

	提案手法	文章ベースのみ	単語ベースのみ	Tf-Idf	TextRank	YAKE	EmbedRank
オンラインライブ	61.0	56.1	53.7	56.1	53.5	17.1	75.6
朝来・三宮の観光	63.6	55.5	55.5	40.9	46.4	30.9	83.6
ネクタイ	80.0	73.9	70.4	62.6	70.4	47.8	75.7

#### 4.1. 検証において用いたデータ

本検証において用いるデータは評価グリッド法によって抽出したクラスタリング済みの評価項目データである。意味の類似した評価項目がまとまって 1 つのカテゴリを形成しており、各カテゴリには手作業でキーフレーズが付与されている。評価項目データは、オンラインライブに関する評価データ(41 カテゴリ)、兵庫県朝来市・三宮の観光に関する評価データ(110 カテゴリ)、ネクタイの印象評価データ (115 カテゴリ)の全 3 種類である。

#### 4.2. 評価の方法

被験者 3 名に対し、評価項目データごとに手作業と提案手法のキーフレーズを比較してもらい、その類似度を 5 段階で評価してもらう。全カテゴリ数のうち類似度の合計が 15 点満点中 12 点以上のカテゴリの割合をキーフレーズ生成の正解率として算出する。今回はキーフレーズの比較対象として 4 つの先行研究 (Tf-Idf, TextRank, YAKE, EmbedRank)の手法により抽出されたキーフレーズの正解率も提示する。

#### 4.3. キーフレーズ生成結果

各評価項目データの正解率を表 1 に記載する。(2 つの生成手法の切り替えの閾値を、文章形式の評価項目の含有率 65%とした)提案手法におけるネクタイデータの正解率が 80%と最高であり、オンラインライブデータの正解率は 61%で最低となった。

#### 4.4. 考察

##### 4.4.1. 評価項目データごとの結果比較

評価項目 1 つあたりの平均単語数をデータごとに算出した結果、オンラインライブは 7.78 語、朝来・三宮は 6.03 語、ネクタイは 2.88 語だった。提案手法はオンラインライブにおける正解率が最も低く、ネクタイが最大であるため、評価項目の単語数が大きいほど精度が下がる傾向がみられた。単語数の多い評価項目のカテゴリは手作業においてより具体的なキーフレーズが割り振られる傾向にあるため、カテゴリによっては手作業と比較して抽象的なキーフレーズが生成され、全体的な精度が下ると推察された。

##### 4.4.2. 提案手法の有効性検証

提案手法の有効性を確認するため、文章ベースの手法

のみ、並びに単語ベースの手法のみでキーフレーズを生成した場合の正解率を提案手法と比較した。その結

果、提案手法は全データでどちらの場合よりも正解率が高くなった。全カテゴリに対し文章ベースの生成手法のみ適用する場合、カテゴリによっては、重要語が GPT4 により不必要な変換を受けたことで意味が異なるものになってしまった可能性がある。一方で提案手法はカテゴリ内の評価項目が文章であるかを考慮し、単語が多いカテゴリに対しては重要度に基づいてキーフレーズを選出するため、余分な変換が行われず適切なキーフレーズが得られたと考えられる。

また、単語ベースの手法のみを用いる際は単語がキーフレーズとなる場合が多いため、キーフレーズの抽象度が高くなる場合が多い。このため、文章の評価項目が多いクラスタにおいて、手作業で得られたキーフレーズとの類似度が下がったと考えられる。提案手法では、文章の評価項目が多いクラスタに対して GPT4 を使い詳細なキーフレーズを生成できるため、正解率が比較的大きくなったと推察する。

##### 4.4.3. 先行研究との比較

提案手法と 4 つの先行研究について、正答率を比較した。表 1 に示す通り、全データを通して提案手法は Tf-Idf, TextRank, YAKE の正答率を上回っていた。これらの先行研究の手法は予め文書を単語に分割し重要度順にランク付けしたうえでキーフレーズを抽出するため、1 つの単語のみ、もしくは似た意味の単語の羅列が抽出されるなど、抽象的なキーフレーズが多く見られた。手作業で割り振られたキーフレーズには修飾語を伴った具体的なものも多いため、キーフレーズ間の類似度が下がり、全体的に正答率に影響したと考えられる。一方、提案手法は文章ベース・単語ベースという 2 種類の生成手法を使い分けているため、文章の評価項目が多いカテゴリ・単語の評価項目が多いカテゴリそれぞれに対応することができる。これにより、キーフレーズの形式が柔軟に変化し、先行研究より正答率が上がったと考えられる(表 2)。

表 2 提案手法と先行研究の手法 (Tf-Idf, TextRank, YAKE) との比較

カテゴリ内の評価項目	提案手法	Tf-idf	TextRank	YAKE	EmbedRank
<ul style="list-style-type: none"> <li>観客目録のカメラ構図。</li> <li>観客席からの視点からの映像。</li> <li>固定の観客席の視点から画面を固定する事。</li> <li>斜す角度を観客目録にする。</li> <li>カメラを定点にする。</li> <li>定点からの視点での撮影...</li> <li>観客がいる音が入っていること。</li> <li>拍手などの観客の音。</li> <li>演奏後に観客からの拍手のリアクションが起こった事。</li> <li>観客の音が聞こえる。</li> <li>ライブ特有の観客の拍手が収録されていること...</li> </ul>	座の視点の撮影	視点	カメラ 構図 観客 席	による 撮影 によって	観客目録のカメラ構図
<ul style="list-style-type: none"> <li>次また来ようと思わない。</li> <li>何回も行きたい感情が湧かない。</li> <li>何回も行きたいと思わない。</li> <li>次も来ようと思わない。</li> <li>臨場感のある音の広がりが必要。</li> <li>ステージ上の音の広がりを中心に保つこと。</li> <li>ホール場での音の広がり方聴ね返りの分数を捉えるのが必要。</li> </ul>	何回も行きたいと思わない	思う	感情	(結果なし)	何回も行きたい感情が湧かない
<ul style="list-style-type: none"> <li>臨場感のある音の広がりが必要。</li> <li>ステージ上の音の広がりを中心に保つこと。</li> <li>ホール場での音の広がり方聴ね返りの分数を捉えるのが必要。</li> </ul>	音の広がり安定	広がり	ことホール場	広がり	ステージ上の音の広がりを中心に保つこと。

また、EmbedRank で得られたキーワードの抽出精度はネクタイ以外の 2 つのデータで提案手法を大きく上回った。ここでの EmbedRank は分散表現に基づいて評価項目をそのまま抜き出すため、カテゴリによっては極端に文章長の長いキーワードが抽出される。オンラインライブと朝来三宮観光のデータは評価項目に含まれる単語数が多いため、手作業でも単語数が多いキーワードが得られている。このため、キーワードの詳細さという点で EmbedRank との相性がよく、提案手法より精度が高くなったと推察する。しかし、評価項目に一切の処理を加えずキーワードとして抽出したことで、EmbedRank の結果の中には文章長が極端に大きいキーワードが多く含まれる。このため、キーワードの妥当性も考慮すると EmbedRank の出力結果は最良とは言い難い。反面、提案手法では GPT4 による文章の簡潔化によりカテゴリにない単語を用いて 1 フレーズに含まれる単語数を抑えているため、手作業で割り振られたキーワードとの類似度は落ちるものの、適切な長さのフレーズが得られている(表 3)。

表 3 提案手法と EmbedRank との比較

カテゴリ内の評価項目	提案手法	Tf-idf	TextRank	YAKE	EmbedRank
<ul style="list-style-type: none"> <li>竹田城の近くに着る所が少ない、あっても営業していない。</li> <li>食べ物、名産物が分かるものがない。</li> <li>馴染み深い食べ物、名産物がない。</li> <li>カフェ、飲食店がパッと見られてなかった...</li> <li>飲食店などから声を掛けられる、美容院などの勧誘が多い。</li> <li>居酒屋さんのキャッチがすごい。</li> <li>夜になるとキャッチが多い。</li> <li>店のキャッチが多い...</li> <li>今住んでいるところそういう街並みがないので、訪問時に観光している感覚がある。</li> <li>観光地が明に感じている。</li> <li>観光地に来たという感じを受ける。</li> <li>臨場感のある音の広がりが必要。</li> <li>ステージ上の音の広がりを中心に保つこと。</li> <li>ホール場での音の広がり方聴ね返りの分数を捉えるのが必要。</li> </ul>	飲食店の開店	食べる	少ない いろんな お店	食べる という 楽しみ	竹田城の近くに着る所が少ない、あっても営業していない
<ul style="list-style-type: none"> <li>今住んでいるところそういう街並みがないので、訪問時に観光している感覚がある。</li> <li>観光地が明に感じている。</li> <li>観光地に来たという感じを受ける。</li> <li>臨場感のある音の広がりが必要。</li> <li>ステージ上の音の広がりを中心に保つこと。</li> <li>ホール場での音の広がり方聴ね返りの分数を捉えるのが必要。</li> </ul>	多い勧誘	キャッチ	多い人	キャッチがすごい	飲食店などから声を掛けられる、美容院などの勧誘が多い
<ul style="list-style-type: none"> <li>今住んでいるところそういう街並みがないので、訪問時に観光している感覚がある。</li> <li>観光地が明に感じている。</li> <li>観光地に来たという感じを受ける。</li> <li>臨場感のある音の広がりが必要。</li> <li>ステージ上の音の広がりを中心に保つこと。</li> <li>ホール場での音の広がり方聴ね返りの分数を捉えるのが必要。</li> </ul>	現地感覚を楽しむ旅	来る	感じいつも	そういう街並み	今住んでいるところそういう街並みがないので、訪問時に観光している感覚がある
<ul style="list-style-type: none"> <li>臨場感のある音の広がりが必要。</li> <li>ステージ上の音の広がりを中心に保つこと。</li> <li>ホール場での音の広がり方聴ね返りの分数を捉えるのが必要。</li> </ul>	音の広がり安定	広がり	ことホール場	広がり	ステージ上の音の広がりを中心に保つこと。

## 5. まとめ

本研究では、人的・時間的コストを伴わない評価

グリッド法の自動化を目的とし、その 1 プロセスとして、評価項目のカテゴリごとにキーワードを生成する手法を提案した。単一の生成手法のみで得られたキーワードの精度を提案手法と比較したところ、2 種類の生成手法を使ったことによる有効性を確認できた。また、出力結果を先行研究と比較したところ、精度の高さと適切な長さを両立したキーワードが得られ、手法の有用性が示された。

## 文献

- [1] 讚井純一郎, 乾正雄, “レポートリー グリッド発展手法による住環境評価構造の抽出ー認知心理学に基づく住環境評価に関する研究 (1)ー,” 日本建築学会計画系論文, no.367, pp.15-22, 1986.
- [2] 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.
- [3] 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. HCI 2023. Lecture Notes in Computer Science, vol.14012, pp. 332-343, Springer, Cham, 2023.
- [4] 宮嶋大輔, 張帆, 杉本匡史, 佐々木香暖, 北野泰成, 橋本翔, and 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] G. A. Kelly, “The psychology of personal constructs,” WW Norton & Company, 1963.
- [6] R. Meng, S. Zhao, S. Han, D. He, P. Brusilovsky, and Y. Chi, “Deep keyphrase generation,” arXiv preprint arXiv:1704.06879, 2017.
- [7] M. Song, H. Jiang, S. Shi, S. Yao, S. Lu, Y. Feng, H. Liu, and L. Jing, “Is chatgpt a good keyphrase generator? a preliminary study,” arXiv preprint arXiv:2303.13001, 2023.
- [8] 金谷武洋, 日本語に主語はいらない 百年の誤謬を正す, 講談社, 2002.
- [9] K. Bennani-Smires, C. Musat, A. Hossmann, M. Baeriswyl, and M. Jaggi, “Simple unsupervised keyphrase extraction using sentence embeddings,” in Proceedings of the 22nd Conference on Computational Natural Language Learning, CoNLL 2018, pp. 221-229, Association for Computational Linguistics, 2018.
- [10] J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. L. Aleman, D. Almeida, J. Altschmidt, S. Altman, S. Anadkat, et al., “GPT4 technical report,” arXiv preprint arXiv:2303.08774, 2023.
- [11] J. Wei, X. Wang, D. Schuurmans, M. Bosma, F. Xia, E. Chi, Q. V. Le, D. Zhou, et al., “Chain-of-thought prompting elicits reasoning in large language models,” Advances in Neural Information Processing Systems, vol. 35, pp. 24824-24837, 2022.