

# レビュー情報を用いた LLM による観光地比較表生成

辻本 陵<sup>1</sup> 坪内 孝太<sup>2</sup> 山下 達雄<sup>2</sup>松田 裕貴<sup>3,1,4</sup> 諏訪 博彦<sup>1,4</sup> 大内 啓樹<sup>1,4</sup><sup>1</sup> 奈良先端科学技術大学院大学 <sup>2</sup> LINE ヤフー株式会社<sup>3</sup> 岡山大学 <sup>4</sup> 理化学研究所

{tsujimoto.ryo.tq0,h-suwa,hiroki.ouchi}@is.naist.jp

{ktsubouc,tayamash}@lycorp.co.jp yukimat@okayama-u.ac.jp

## 概要

Table Generation は、複数の要素を比較する際に有用な技術である。本研究では、観光地 (POI) のレビュー情報を基に観光地間の比較表を自動生成する新しい手法を提案する。レビューから属性-レビューペアを抽出し、それを属性ごとに整理した情報を LLMs に入力することで、客観的要素と主観的要素を含む比較表の生成を可能にした。評価実験では、無作為なレビュー入力やレビューを使用しない手法と比較して、提案手法が優れた比較表を生成できることを確認した。本研究は、観光地選定を支援する実用的なアプローチを提示し、Table Generation タスクに新たな視点を提供するものである。

## 1 はじめに

大規模言語モデル (LLMs) の進展により、自然言語データから構造化された情報を生成する技術が注目を集めている [1, 2]。中でも、表形式での情報提示 (Table Generation) はデータの比較や視覚的なわかりやすさに優れ、特に異なる要素を比較するタスクで重要な役割を果たす [3, 4]。

観光地 (Points of Interest, POIs) の選定において、ユーザーの投稿するレビューは重要な情報源であり、観光地の魅力や問題点を具体的に記録している [5]。しかし、膨大なレビューを比較・分析することは現実的ではなく、情報過多による負担が課題となっている [6]。この課題を解決するには、レビュー情報を効率的に要約し、観光地間の特徴や違いを明確に比較する手法が必要である。

これまでの研究では、テキストを表形式に変換する方法が提案されてきたが、情報の統合や推論を伴う複雑なシナリオには十分対応していない [7, 8]。さらに、観光地比較では固定された属性情報だけで

	東京都葛西臨海水族館	鴨川シーワールド
特徴	関東唯一のマグロ大水槽。	息をのむようなシャチのパフォーマンス。
施設	都心からアクセスが良く、手ごろな価格で楽しめる。	広大な敷地を活かした、自然豊かな環境が魅力。
体験	学びの場として、楽しく水族館や生き物を知ることができる。	海獣ショーが充実しており、ドキドキワクワクを楽しめる。

図 1 水族館を対象とした比較表の例。

なく、ユーザー体験に基づく柔軟な特徴や主観的な意見も重要である。本研究では、観光地に関するレビューから LLM を活用して共通点と違いを抽出し、表形式で比較するタスクを新たに設定する。

図 1<sup>1)</sup>に示すように、本研究ではレビューから属性-レビューペアを抽出し、LLM を用いて観光地間の比較表を自動生成するフレームワークを設計した。このアプローチにより、客観的要素に加え、ユーザー体験に基づく主観的要素を含む比較表の生成が可能となる。また、クラウドソーシングを活用した評価実験により、提案手法の有効性と他手法に対する優位性を確認した。

実験結果として、提案手法はレビューから要点を効率的に抽出し、観光地間の特徴を強調した比較表を生成する能力に優れていることを示した。本研究は、レビューを活用した観光地や商品の比較手法を提示し、Table Generation というタスクに新たな視点を提供するものである。

1) 図中のコメントは生成 AI が作成した架空のものである。実際のコメントは投稿者のプライバシーを考慮し非公開としている。

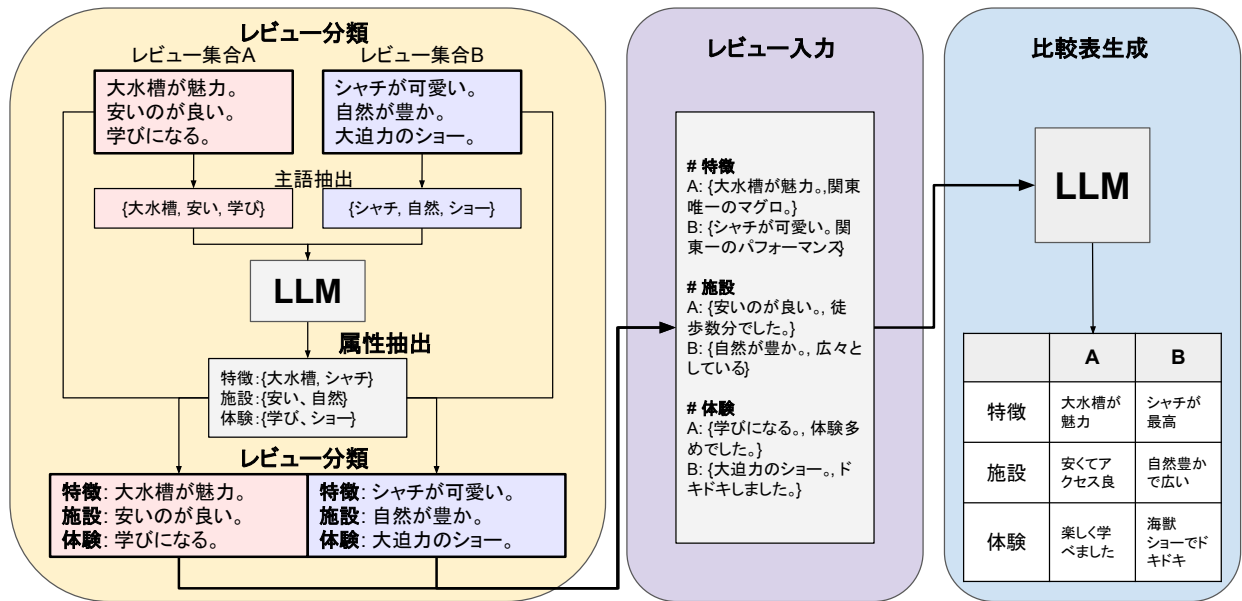


図2 提案手法の3ステップ: (1) レビュー分類, (2) レビュー入力, (3) 比較表生成

## 2 タスク設定

本研究ではタスクを以下のように定式化する。2つの観光地  $POI_A$  と  $POI_B$  およびそれぞれのレビュー集合  $R_A$  と  $R_B$  を入力として与える。各レビュー集合は観光地に関する多様な意見や特徴を含む。目標は、LLMを活用してこれらのレビューを解析・要約し、観光地  $POI_A$  と  $POI_B$  の違いや共通点を視覚的にわかりやすい表形式で生成することである。

## 3 関連研究

山田ら [9] は、転職サイトに寄せられたレビューから企業の特徴に相当する情報を抽出する手法を提案した。この手法では、特徴語と評価後の辞書を用いて企業情報を抽出しており、辞書の充実や適切な分類が重要であることを示唆している。

Furkanら [10] は、科学的リーダーボードの自動構築に関する研究を行い、手動で維持することが困難な状況を解決するための LLMs を活用したフレームワークを提案した。ここでは、さまざまな LLM がタスク、データセット、評価メトリックのように予め定められた要素に関する情報を認識することを示した。

これらの研究から、特徴を抽出する重要性和、情報抽出タスクにおける LLMs の有用性が示されている。本研究では、POI 比較表生成タスクにおける

LLMs を利用した情報抽出フレームワークの有効性を比較し、これらの手法の適用可能性を探る。

## 4 手法

提案手法「Categorized Review」は、図2に示すように、レビューを属性ごとに整理・構造化し、それを活用して観光地間の比較表を生成するものである。本手法は以下の2つのステップで構成される。

### 4.1 ステップ1: レビュー分類

従来の研究では、直接の係り受け関係に基づく単純なペア抽出によって属性-意見ペアを抽出する手法が提案されてきた。しかし、属性と意見が複数文節にまたがる場合や、重要な属性の被覆率が低いなどの課題が指摘されている [11]。

本研究ではレビューを属性ごとに分類し、情報を体系化するアプローチを採用した。このプロセスにより、レビュー情報の曖昧さを排除し、正確な比較表の生成を可能にする。

**属性抽出** レビューから属性を抽出するため、GinZA を用いて係り受け解析を行い、主語を取得する。この際、名詞複合語 (例: 大水槽) や修飾関係 (例: 大きな水槽) も含めて抽出するよう工夫している。抽出された主語は LLMs に入力され、共通属性や固有属性を JSON 形式で整理する。この JSON 形式は、後続のレビュー分類で使用される構造化

表1 アブレーション研究

	レビュー分類	レビュー入力
Categorized Review	○	○
Random Review	x	○
No Review	x	x

データを提供する。プロンプトの例は付録4に記載する。

**レビュー分類** 抽出した属性と名詞例を文埋め込みしてベクトル化し、レビューをコサイン類似度に基づいて分類する。固有表現をマスクしたレビューを用いることで、属性ごとの整理が正確に行える。この分類により、雑多な情報を属性ごとに体系化し、比較表生成の精度を向上させる。

## 4.2 ステップ2: レビュー入力

LLMsは事前学習によって一般的な知識を有しているが、特定のドメインにおける詳細な知識を正確に反映させるためには、適切な入力データが求められる[12]。

本研究では、ステップ1で体系化したレビューを基に、観光地間の比較表を生成する。この際、レビューは属性ごとに整理され、各観光地のレビューが体系的に入力データとしてまとめられている。この形式により、LLMsは属性ごとにレビューを明確に処理し、観光地間の共通点や違いを効率的に抽出することが可能となる。プロンプトの具体例は付録表5に記載する。この設計により、事前学習モデルの知識と整理されたレビュー情報を融合させた、高品質な比較表の生成を実現する。

## 4.3 アブレーション研究

本研究では、提案手法の有効性を検証するために、以下のアブレーション研究を実施した。詳細な設定は表1を参照されたい。

- **Random Review:** レビューを分類せずに、無作為に LLM に入力して比較表を生成。
- **No Review:** レビューを LLM に入力せず、事前知識のみで比較表を生成。

これらの手法と比較することで、提案手法が体系的なレビュー分類と整理を通じて生成結果にどのような改善をもたらすかを評価した。

表2 対象とした POI ペア

ペア ID	POI
0	ディズニーシー/ディズニーランド
1	下呂温泉/草津温泉
2	大阪城/姫路城
3	東京スカイツリー/東京タワー
4	銀閣寺/金閣寺
5	旭山動物園/円山動物園
6	上野動物園/葛西臨海水族園
7	キッザニア甲子園/USJ
8	マザー牧場/東京ドイツ村
9	八景島シーパラダイス/鴨川シーワールド

	旭山動物園	札幌市円山動物園
施設	広大で多様な動物展示	コンパクトでアクセス良好
展示物	この問題は「1点」を選択してください	定番の動物が多い
体験	迫力ある観察体験	親しみやすい間近の体験

図3 チェック設問における比較表例

## 5 実験

各手法の性能を比較するため、生成した比較表をクラウドソーシングを利用して評価した。

### 5.1 実験設定

**モデル** 主語からの属性抽出および比較表の生成には GPT-4o mini [13] を使用した。文埋め込みには、pkshatech/GluCoSE-base-ja モデルを利用した。固有表現抽出には、Yahoo!のテキスト解析 Web API 「固有表現抽出」<sup>2)</sup>を用いた。

**データ** データとして、「Yahoo!マップ」<sup>3)</sup>に登録された各施設のレビューを使用した。対象とした POI 施設数は 20 施設で、それぞれの施設に寄せられたレビューを新着順に 300 件収集した。データ収集日は 2024 年 11 月 18 日である。

### 5.2 評価方法

「Yahoo!クラウドソーシング」<sup>4)</sup>を利用し、2種類のタスクを通じて評価を行った。

**評価指標** 以下の2つの利用状況を考慮したシナリオを設定した。

- 友達への旅行先の推薦時
- 自分の旅行先の計画時

また、以下の4つの比較表の評価基準を設定した。

- 2) <https://developer.yahoo.co.jp/webapi/jlp/ner/v1/>
- 3) <https://map.yahoo.co.jp/>
- 4) <https://crowdsourcing.yahoo.co.jp/>

表3 スコア付けタスクの評価結果（平均点）

シナリオ	評価観点	Categorized Review	Random Review	No Review
友達への旅行先の推薦	内容の正しさ	3.85	3.47	2.98
	比較軸の妥当性	3.82	3.54	3.08
	情報の豊富さ	3.91	3.45	3.07
	総合的な良さ	3.80	3.48	3.09
自分の旅行先の計画	内容の正しさ	3.77	3.54	2.82
	比較軸の妥当性	3.80	3.59	3.15
	情報の豊富さ	3.84	3.40	3.15
	総合的な良さ	3.89	3.61	3.15

- 内容の正しさ
- 比較軸の妥当性
- 情報の豊富さ
- 総合的な良さ

これらを組み合わせ、計8種類の質問を作成した。

**タスク** 以下の2種類のタスクを実施した。

- **スコア付けタスク**: 比較表を提示し、1点から5点のリッカート尺度で評価を依頼する。
- **比較タスク**: 2種類の比較表を提示し、どちらが優れているかを選択してもらう。

各タスクでは無作為な回答を防ぐため、図3に示すようなチェック設問を導入した。各タスクは、10問の設問と1問のチェック設問で構成され、各設問を100回繰り返し出題した。8種類の質問形式と3種類の比較表を組み合わせた結果、各タスクあたりの総サンプル数は24,000件となった。

### 5.3 結果

表3は、スコア付けタスクにおける各質問の平均点を示している。付録表6は、比較タスクにおける各質問の Elo レートを示す。どのタスクおよび質問においても、提案手法は他の手法よりも一貫して優れた結果を示した。

### 5.4 分析

**レビューの体系化による影響** スコア付けタスクの結果に対して、Games-Howell [14] の多重検定比較法を適用した。分析の結果、Categorized Review のスコアが最も高く、他の条件との差は統計的に有意であった ( $p < 0.05$ )。一方で、No Review は最も低いスコアであった。この結果は、LLM が元々有している知識だけから比較表を生成するのは難しいことを示す。また、Random Reviews は中間的なスコアを示し、レビューの属性に応じた体系的な入力がスコア向上に寄与することを示唆する結果となった。

これにより、レビューを使用しない場合よりも使用した場合、さらに体系的に整理して入力した場合のほうがスコアが向上することが示された。

**事例ベース分析** スコア付けタスクでは、ペアに関係なく提案手法が他の手法よりも一貫して優れた結果を示した。一方、比較タスクでは「横浜・八景島シーパラダイス & 鴨川シーワールド」のペアでのみ、提案手法のスコアが Random Review による比較表に劣っていた。この結果については、提案手法が体験属性において親子連れと家族連れを比較していたのに対し、Random Review では家族訪問とデートを比較するなど、同じ属性内での対称性が保たれていたことが要因であると考えられる。

**回答者ベース分析** 比較タスクでは、No Review で生成された比較表に対して全く点数を付けない回答者が一定数存在した。この傾向は、ユーザーがその場所を訪問した経験があるかどうかに関連している可能性が考えられる。今後のクラウドソーシング調査では、回答者の訪問経験などの属性情報も収集し、さらなる分析を行うことが求められる。

## 6 まとめ

本研究では、LLMs を用いてレビューから POI の比較表を生成する手法を提案した。クラウドソーシングを活用した評価実験により、提案手法が生成する比較表の有効性を確認した。分析の結果、単にレビューを入力するだけでなく、属性に基づき体系的にレビューを整理・入力することでスコアが向上することが明らかになった。一方で、取り扱った事例の少なさが結果に影響を与えた可能性も示唆された。今後の課題として、取り扱うペア数の増加や生成する比較表の多様化が挙げられる。さらに、データ処理手法やプロンプト設計の改善により、より整合性の取れた比較表を生成できるようにすることが求められる。



## 参考文献

- [1] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners, 2020.
- [2] Aakanksha Chowdhery, Sharan Narang, Jacob Devlin, Maarten Bosma, Gaurav Mishra, Adam Roberts, Paul Barham, Hyung Won Chung, Charles Sutton, Sebastian Gehrmann, Parker Schuh, Kensen Shi, Sasha Tsvyashchenko, Joshua Maynez, Abhishek Rao, Parker Barnes, Yi Tay, Noam Shazeer, Vinodkumar Prabhakaran, Emily Reif, Nan Du, Ben Hutchinson, Reiner Pope, James Bradbury, Jacob Austin, Michael Isard, Guy Gur-Ari, Pengcheng Yin, Toju Duke, Anselm Levskaya, Sanjay Ghemawat, Sunipa Dev, Henryk Michalewski, Xavier Garcia, Vedant Misra, Kevin Robinson, Liam Fedus, Denny Zhou, Daphne Ippolito, David Luan, Hyeontaek Lim, Barret Zoph, Alexander Spiridonov, Ryan Sepassi, David Dohan, Shivani Agrawal, Mark Omernick, Andrew M. Dai, Thanumalayan Sankaranarayanan Pillai, Marie Pellat, Aitor Lewkowycz, Erica Moreira, Rewon Child, Oleksandr Polozov, Katherine Lee, Zongwei Zhou, Xuezhi Wang, Brennan Saeta, Mark Diaz, Orhan Firat, Michele Catasta, Jason Wei, Kathy Meier-Hellstern, Douglas Eck, Jeff Dean, Slav Petrov, and Noah Fiedel. Palm: Scaling language modeling with pathways, 2022.
- [3] Yanai Elazar, Akshita Bhagia, Ian Magnusson, Abhilasha Ravichander, Dustin Schwenk, Alane Suhr, Pete Walsh, Dirk Groeneveld, Luca Soldaini, Sameer Singh, Hanna Hajishirzi, Noah A. Smith, and Jesse Dodge. What’s in my big data?, 2024.
- [4] Xi Fang, Weijie Xu, Fiona Anting Tan, Jiani Zhang, Ziqing Hu, Yanjun Qi, Scott Nickleach, Diego Socolinsky, Srinivasan Sengamedu, and Christos Faloutsos. Large language models (llms) on tabular data: Prediction, generation, and understanding – a survey, 2024.
- [5] 広田すみれ, 高橋聖奈. レストランクチコミサイトにおける評価の数や質と意思決定の関係. 東京都市大学横浜キャンパス情報メディアジャーナル = Journal of information studies, No. 15, pp. 32–36, 04 2014.
- [6] Cheol Park and Thae Min Lee. Information direction, website reputation and ewom effect: A moderating role of product type. **Journal of Business Research**, Vol. 62, No. 1, pp. 61–67, 2009.
- [7] Sam Wiseman, Stuart Shieber, and Alexander Rush. Challenges in data-to-document generation. In Martha Palmer, Rebecca Hwa, and Sebastian Riedel, editors, **Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing**, pp. 2253–2263, Copenhagen, Denmark, September 2017. Association for Computational Linguistics.
- [8] Jekaterina Novikova, Ondřej Dušek, and Verena Rieser. The E2E dataset: New challenges for end-to-end generation. In Kristiina Jokinen, Manfred Stede, David DeVault, and Annie Louis, editors, **Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue**, pp. 201–206, Saarbrücken, Germany, August 2017. Association for Computational Linguistics.
- [9] 菱田隆彰, 炭竈桂輔, 遠藤正隆. 口コミを用いた企業の特徴の抽出方法の提案. 愛知工業大学総合技術研究所研究報告, No. 17, pp. 97–100, 09 2015.
- [10] Furkan Şahinuç, Thy Thy Tran, Yulia Grishina, Yufang Hou, Bei Chen, and Iryna Gurevych. Efficient performance tracking: Leveraging large language models for automated construction of scientific leaderboards, 2024.
- [11] 中野裕介, 湯本高行, 新居学, 佐藤邦弘. 商品レビュー要約のための属性-意見ペア抽出, 11 2014.
- [12] Yevgeni Berkovitch, Oren Glickman, Amit Somech, and Tomer Wolfson. Generating tables from the parametric knowledge of language models, 2024.
- [13] Gpt-4o system card.
- [14] Paul A. Games and John F. Howell. Pairwise multiple comparison procedures with unequal  $n$ 's and/or variances: A monte carlo study. **Journal of Educational Statistics**, Vol. 1, pp. 113 – 125, 1976.

## A プロンプト

表4 レビュー主語群から属性抽出するためのプロンプト

```
# Instruction
1. 与えられた名詞セットから、各 POI の共通要素と固有要素を識別してください。2. 共通要素は、両方の POI に
関連する名詞を含むカテゴリとして分類してください。3. 固有要素は、それぞれの POI にのみ関連する名詞を含
むカテゴリとして分類してください。4. 共通要素は、5つ考えてください。5. 共通要素のカテゴリ名を考えてくだ
さい。6. 各カテゴリには、5つ程度の名詞を含めてください。7. 結果を json 形式で出力してください。

# Input Example
POI: POI A
名詞: [名詞 1, 名詞 2, 名詞 3, ..., 名詞 m]

POI: POI B
名詞: [名詞 1, 名詞 2, 名詞 3, ..., 名詞 n]

# Output Example
“json
{ "共通要素": { "カテゴリ 1": ["名詞 1", "名詞 2", "名詞 3", "名詞 4", "名詞 5"], "カテゴリ 2": ["名詞 6", "名詞 7", "名
詞 8", "名詞 9", "名詞 10"], ... }, "固有要素": { "POI A": { "特有": ["名詞 11", "名詞 12", "名詞 13", "名詞 14", "名詞
15"] }, "POI B": { "特有": ["名詞 16", "名詞 17", "名詞 18", "名詞 19", "名詞 20"] } } }
“
```

表5 体系的な情報から比較表生成するためのプロンプト

```
# Instruction
1. 各 POI の特徴を、カテゴリごとに簡潔かつ一貫した文体で説明してください (30~40 文字程度で)。2. 各カテ
ゴリで POI 同士の特徴が対照的かつ同じ軸で比較できる形で1文にまとめてください。3. 比較が難しい場合、カテ
ゴリ自体を再検討し、同じ軸で評価できるように調整してください。4. 結果を json 形式で構造化し、リスト形式で
はなく文章形式で説明してください。

# Input Example
カテゴリ: 施設- A 美術館- 広い- 新しい- B 美術館- 狭い- レトロな雰囲気...

# Output Example
“json
{ "施設": { "A 美術館": "広くて新しい", "B 美術館": "狭いがレトロ" }, "展示物": { "A 美術館": "絵画が多い", "B 美
術館": "彫刻が多い" } ... }
“
```

## B 比較タスク

表6 比較タスクの評価結果 (Elo レート, 初期値 1500, K Factor=4)

シナリオ	評価観点	Categorized Review	Random Review	No Review
友達への旅行先の推薦	内容の正しさ	1657	1490	1351
	比較軸の妥当性	1662	1506	1331
	情報の豊富さ	1701	1512	1285
	総合的な良さ	1669	1514	1316
自分の旅行先の計画	内容の正しさ	1625	1520	1353
	比較軸の妥当性	1623	1522	1354
	情報の豊富さ	1705	1527	1266
	総合的な良さ	1690	1502	1307