

# P06 大規模言語モデルを活用したレビュー分析による消費者の関心トピックと潜在ニーズ抽出

林 美里、伴 果純、山中 高夫（上智大学大学院）

**概要：**  
本研究では、大規模言語モデル（LLM）の自然言語処理能力を活用し、地域特性を考慮したレビュー分析手法を探索した。  
BERTopicを用いたトピックモデリングにより、ゴルフ場レビューから主要テーマを抽出し、消費者の関心トピックと潜在ニーズを明らかにした。  
本手法は、従来のLDAによる話題抽出よりも文脈を的確に反映でき、地域特性や利用者層別の特徴分析にも有効であることを確認した。

レビュー→前処理→BERTopic→トピック抽出→トピック命名→関心・ニーズ抽出

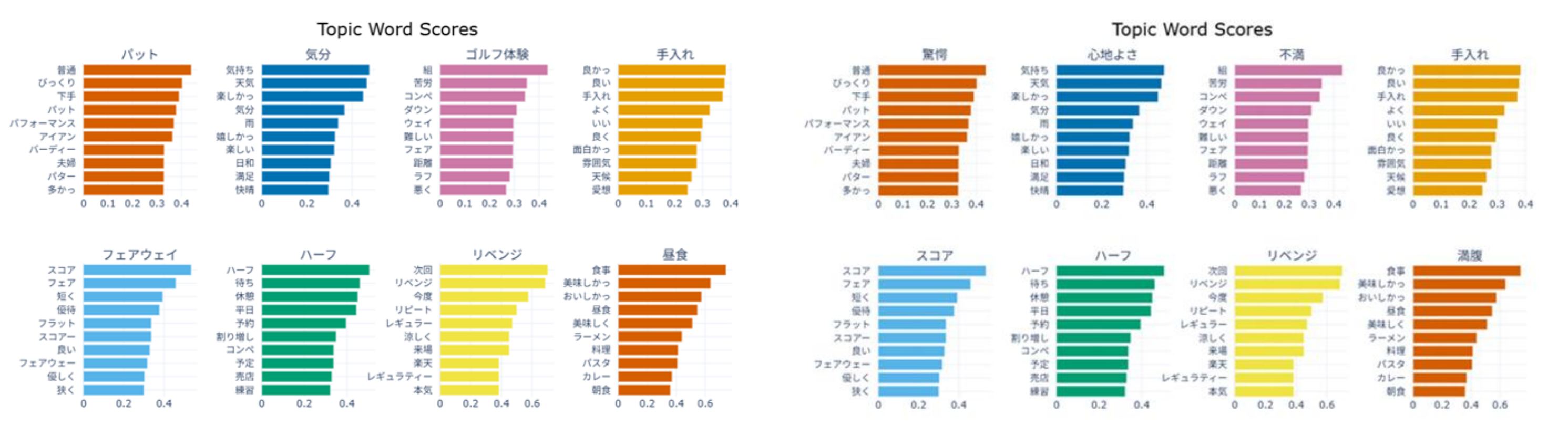
**研究の背景：**  
近年、インターネットの普及とスマートフォンの浸透により、消費者は商品やサービスに関する情報を容易に入手できるようになった。  
しかし、これまでの研究では特定地域や業態に絞った分析が少なく、地域特性を考慮した詳細な分析が求められていた。

**研究の目的：**  
大規模言語モデル（LLM）の自然言語処理能力を活用し、地域特性を考慮した詳細なレビュー分析の手法を探索する。  
この手法により、大規模データから消費者の関心トピックと潜在ニーズ抽出を効率的に把握することを目的とする。  
具体的には、自然を感じつつ心身のリフレッシュができるゴルフを題材として、若年層や女性といったノンゴルファーの「潜在ニーズ」を可視化し、Well-Being向上やノンゴルファー層拡大への示唆を得ることを目的とする[1]。

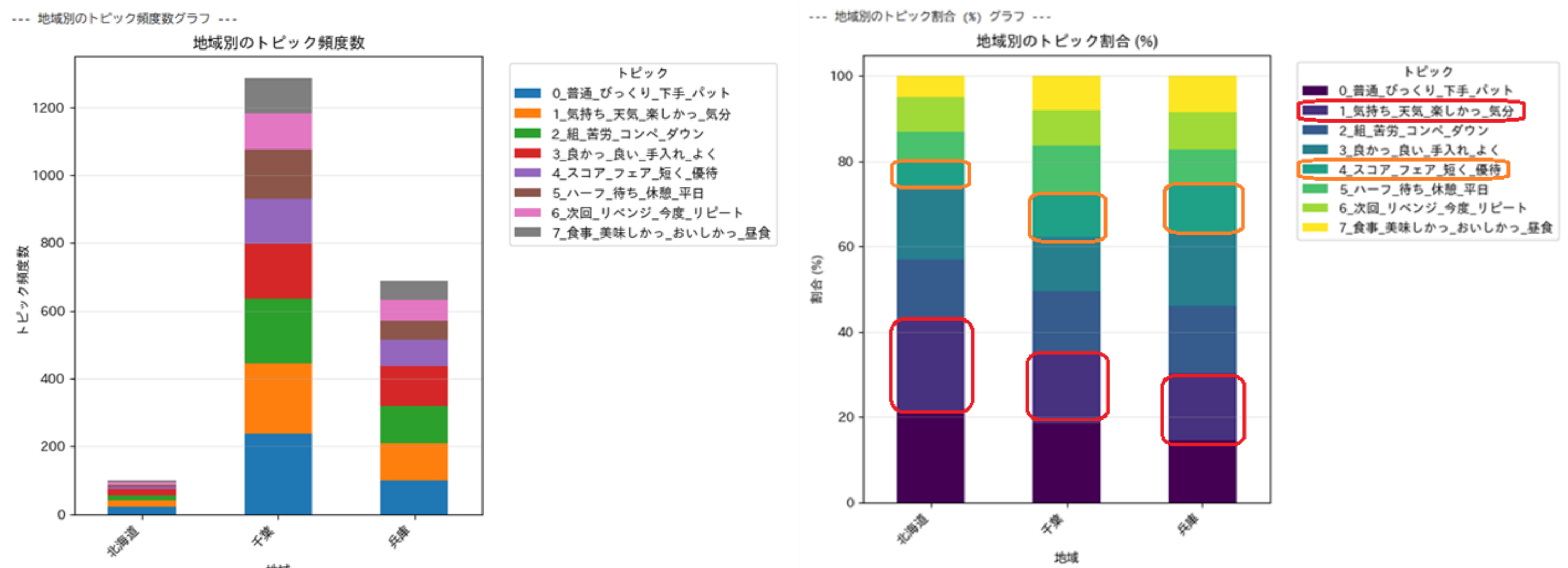
**先行研究：**  
- 先行研究1（NPS推定）  
池田ら[2]は宿泊施設レビューから低コストのNPS（顧客ロイヤルティ指標）を推定し、アンケートが不要で低コストのスコアリング手法を提案。  
- 先行研究2（難易度算出）  
富樫ら[3]はゴルフ施設レビューからコース難易度を自動算出する手法を提案。  
- 本研究では、これらの研究で確立された「レビュー定量化」アプローチを継承しつつ、LLMを活用して文脈理解と地域特性を考慮した潜在ニーズ抽出に挑戦する点に新規性がある。

**実装・分析手法：**  
- 対象データ：楽天GORA（1,669施設・約32万レビュー）のうちオススメタイプ：[1:女性]のレビューを対象[4]  
- 対象地域：北海道・千葉・兵庫  
- レビュー前処理：正規表現による文分割+fugashiで形態素解析  
- 埋め込みモデル：google/embeddinggemma-300m  
- トピック抽出：BERTopic（BERT埋め込み+クラスタリング）[5]  
- トピック命名：Qwen3-8b / Gemma3-12b により日本語アトリビュート生成[6]

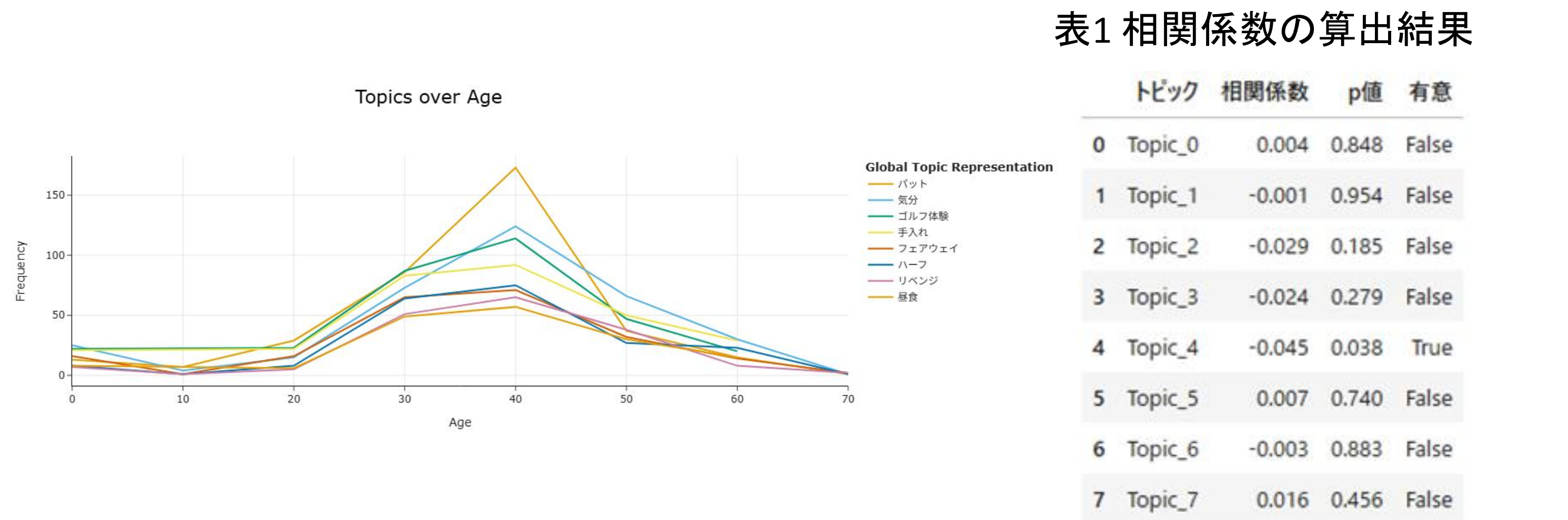
**分析結果：**  
・主要トピックの抽出結果  
- トピック例：「気分、心地よさ」「手入れ」「フェアウェイ、スコア」「ハーフ」など  
- Qwen3, Gemma3ともに類似のトピック名を抽出。  
- Qwen3は短く直感的な日本語表現を生成した。  
- Gemma3は感情表現を多く含む傾向があった。  
- トピック名の可読性・解釈性の観点から、Qwen3を採用。



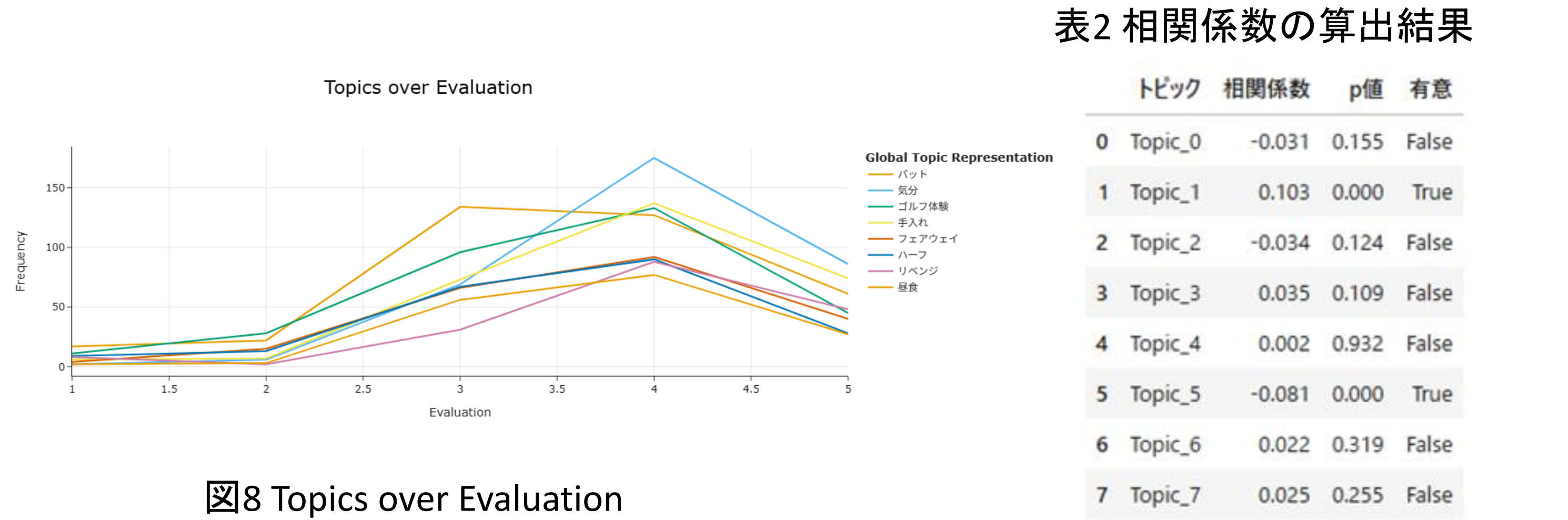
・地域別比較  
- 北海道：「トピック1（気分、心地よさ）」など、心理的関心に関連するトピックが多い。  
- 千葉・兵庫：「トピック4（フェアウェイ、スコア）」など、実的関心に関連するトピックが多い。



・年齢との相関分析  
- 「トピック4（フェアウェイ、スコア）」は弱い負の相関 (r=-0.045, p<0.05)



・総合評価：（1:悪い-5:良い）との相関分析  
- 「トピック1（気分、心地よさ）」は弱い正の相関 (r=0.103, p<0.05)  
- 「トピック5（ハーフ）」は弱い負の相関 (r=-0.081, p<0.05)



**まとめ：**  
- LLMを用いたBERTopic分析により、レビューから潜在的関心を抽出可能。  
- 地域・年齢などの属性別特徴を明確化。  
- 消費者中心の新たなゴルフ体験設計へ示唆。

**考察・展望：**  
- 従来の評価軸（スコア・距離）に加え、「感情的満足度」指標の導入が有効。  
- 今後は感情スコアリングとの統合分析やプレイヤー属性（経験年数等）の反映、NPS推定等の指標化とカスタマージャーニー連携に発展予定。  
- 本手法は、宿泊・観光・飲食などの地域別マーケティング戦略の立案に資する。

**参考文献：**  
[1] 吉川雄大. ゴルフ産業の現状とその可能性. 2023.  
[2] 池田貴康他. 宿泊施設レビューからの顧客ロイヤルティ向上のためのNPS推定. P11, IDR ユーザーフォーラム2024.  
[3] 富樫健太他. ゴルフ施設レビューデータからのコース難易度自動算出手法. P10, IDR ユーザーフォーラム2024.  
[4] 楽天グループ株式会社 (2010): 楽天GORAデータ. 国立情報学研究所情報学研究データリポジトリ. (データセット). <https://doi.org/10.32130/idr.2.3>  
[5] 千島雄太. BERTopicを使ったトピック分類：初学者向けチュートリアル. 2025.  
[6] 秋山明日香他. BERTopicを用いた漫画のあらすじに対する特性表現の検討. Webインテリジェンスとインタラクション研究会, P125-P128, 2024.

**謝辞：**  
本研究では、国立情報学研究所のIDRデータセット提供サービスにより楽天グループ株式会社から提供を受けた「楽天データセット」 ([https://rit.rakuten.com/data\\_release/](https://rit.rakuten.com/data_release/)) を利用した。

