

ホテル顧客満足度向上のためのLLMを用いたレビューデータ分析

河野吉宏, 池田貴康, 富樫健太, 中塚晶仁, 小川和浩, 荒川健, 浪岡保男 (東京都立産業技術大学院大学)

背景

- 顧客満足度向上のために、レビューに含まれる内容から顧客の注目点を明確化し、優先すべき改善箇所を具体化する。
- レビュー分析は依然として負担が大きい課題である。しかし、大規模言語モデル(LLM)の進展により自然言語処理(NLP)の活用が進展している。一方で、実運用においては以下の課題が存在する。
 - 大規模な計算リソースの必要性
 - ハルシネーションによる不正確な出力のリスク

目的

- 軽量のローカルLLMを活用し、低コストかつ高精度なレビュー解析を実現する。
- レビュー本文と評価スコアを関連付けることでハルシネーションを抑え、顧客満足度の具体的な改善ポイントを特定する仕組みを提案する。
- データ駆動型のアプローチを用いて、ホテル運営の改善方針を提案する。

提案手法

データセット

楽天トラベル・レビューデータ
・ユーザー投稿本文
・0-5段階の評価項目

楽天トラベルの点数基準
5点大変満足, 4点満足, 3点普通, 2点やや不満, 1点不満
(該当サービスがない項目については評価不能が選択可能)

レビューの例

投稿本文

繁華街に近く、部屋も素晴らしいビジネスホテルです。チェックイン混雑時、フロントでの並び方が不明瞭で後から来た方が先にチェックインする場面が数多く見受けられました。

評価項目

項目	スコア
立地	5
部屋	4
食事	0
風呂	5
サービス	3
設備・アメニティ	4

LLM 投稿本文の分析

レビュー投稿本文をLLMで解析し各評価項目別に**キーワードの抽出**と**ポジティブ/ネガティブ/ニュートラルに分類**

```
{  
  "location_score": 1,  
  "location_keywords": ["繁華街に近く"],  
  "room_score": 1,  
  "room_keywords": ["部屋も素晴らしい"],  
  ...  
  "service_score": -1,  
  "service_keywords": ["フロントでの並び方が不明瞭",  
    "後から来た方が先にチェックイン"]  
}
```

評価項目の分析

レビュー評価スコアを**ポジティブ/ネガティブ/ニュートラルに分類**

評価者間の基準の違いを統一するため、レビュー評価スコアを Z-score normalization で標準化し以下に分類

- ポジティブ (Zスコア > 0.5)
- ネガティブ (Zスコア < -0.5)
- ニュートラル (-0.5 ≤ Zスコア ≤ 0.5)

項目	ポジネガ判定
立地	1
部屋	0
...	
サービス	-1

解析結果の精度検証

LLMによる投稿本文の感情分析結果を評価項目の感情分析結果と照合し、**一致した項目を「正解」としてラベル付け**

項目	ポジネガ	キーワード
立地	1	繁華街に近く
サービス	-1	フロントでの並び方が不明瞭 後から来た方が先にチェックイン

LLM トピックス生成

正解と判定された結果をもとに、各評価項目に対してNMF (非負値行列因子分解) [1]を適用し、LLMで**擬似的にトピックモデリングを再現**。評価項目ごとの改善点や顧客満足度向上のための課題を抽出

実験結果

本実験では、楽天トラベルの特定宿泊施設レビュー1175件を使用し、Gemma2 9Bという軽量LLMを採用。ハルシネーションを抑制するため、レビュー本文と評価スコアを関連付けるフィードバックループを構築した。

解析結果の精度検証

本評価では、LLMを用いてレビュー本文の感情分析結果と評価項目の感情判定を照合し、一致率を検証した。

結果：LLMがレビュー本文から評価項目ごとの感情を正確に分析できることが確認され、軽量のローカルLLMでも実用性が高いことが示された。

column	precision	f1_score	accuracy	correct_count
location_score	0.945205	0.784091	0.849802	215
room_score	0.949772	0.912281	0.867550	262
meal_score	0.970588	0.910345	0.881098	289
bath_score	0.933333	0.922353	0.889262	265
service_score	0.974522	0.866856	0.846906	260
facility_score	0.800000	0.675325	0.746193	147

facility_scoreの一致率が低い理由として、以下の点が挙げられます。

- 特に「駐車場が無料」など、駐車場に関連するワードで評価の相違が見られる
- 宿泊者特典 (QUOカードプレゼント) が本項目に含まれて判断されている
- 「マッサージ機を導入してほしい」など、要望が本項目に含まれて判断されている

トピックス生成

NMFを用いて評価項目ごとのトピックモデリングを行い、レビュー内容を適切に反映しているかを検証した。

結果：トピック生成により、各評価項目の改善点や要素を明確化し、改善施策の提案が可能であることが示された。

トピック1: 立地条件

- ポジティブ(60%): 繁華街に近い, 駅から近い, 便利な場所, 立地が良い
- ネガティブ(40%): 駅からは遠い, 駅からの距離, 遠すぎる, 徒歩8分

トピック2: アクセス・駐車場

- ポジティブ(50%): 立体駐車場, 車利用に便利, 高速が近く, 駐車場の前
- ネガティブ(50%): 駅からのアクセスが悪い, 地下道の長さ, 出口12番, タクシー利用

トピック3: 周辺環境

- ポジティブ(65%): コンビニ近い, ラーメン屋さんがある, 飲食物が多い, 飲食店が多い
- ネガティブ(35%): 客引きが寄って来る, 治安が悪い, 飲み屋街, 場末の飲み屋街, 女性の場合

まとめ

主な結果

本研究では、軽量のローカルLLMを活用し、高精度なレビュー解析とハルシネーション抑制の両立を実現した。

- レビュー単位で評価項目ごとに感情とキーワードを抽出することで、トークンを抑え高精度に解析
- ハルシネーション抑制のため、レビュー本文と評価スコアを関連付け、解析結果の信頼性を向上
- 実験結果により、軽量LLMにおいても感情判定の一致率が高いことが確認され、改善方針の具体化が可能であることを示した

今後の展望

- 最新のプロンプトエンジニアリング技術を活用し、軽量LLMによるレビュー解析と評価スコアの最適化を図る。
- NMF分析の精度および信頼性を向上させ、具体的な改善提案に結びつけることを目指す

謝辞

本研究では、国立情報学研究所の情報学研究データリポジトリ(IDR)によるデータセット提供サービスを通じて、楽天グループ株式会社様からご提供いただいた「楽天データセット」(https://rit.rakuten.com/data_release/)を利用させていただきました。ここに記して謝意を表します。

参考文献

[1] Lee, D.D. and Seung, H.S. Learning the Parts of Objects by Non-negative Matrix Factorization. Nature 401, 788-791 (1999)