

8 感情分類を利用した観光体験自動分析システムの提案 Sightseeing Experience Analyze-System with "Wheel of Emotion"

2025Y13 小村 菖悟 (Omura Shogo)

担当教員 出口 利憲 (Deguchi Toshinori)、堀内 咲江 (Horiuchi Sakie)

1. 序論

1.1 背景

近年、インターネット上には口コミやレビューといった個人発信のテキストデータが膨大に蓄積されており、これらは観光分野において旅行者の行動意思決定に多大な影響を及ぼしている。特に、景勝地や天然記念物などの有力な観光資源には多様な属性を持つ旅行者が訪問するため、多種多様な主観的評価が集積する傾向にある。

こうして蓄積されたテキストデータから旅行者の潜在的な感情や詳細なニーズを精度高く抽出・構造化することは、観光資源の価値再定義や適切なマネジメントにおいて極めて重要な課題であるといえる。

1.2 目的

本研究は、観光資源に対して寄せられたテキストデータに自然言語処理を用いて詳細に分析し、特定の要素に対して旅行者が抱く感情の種別およびその強度を明らかにすることを目的とする。

具体的には、口コミに含まれる具体的な言及(単語の共起関係)をもとに、観光資源の各要素が持つ魅力や課題点を抽出し、それらを構造化されたニーズとして提示することを目指す。これにより、旅行者の感性に根ざした深い分析を実現し、観光資源の活性化に資するデータ駆動型のアプローチを確立する。

2. 研究手法

本研究では、高度な自然言語処理技術を用いた多次元感情分析と、単語の共起ネットワーク構築を組み合わせた解析手法を提案する。

2.1 感情分析モデルの構築

文章から取得された多次元ベクトルを入力として、8感情とその強度を出力する線形重回帰モデルを構築する。

ベクトル表現を獲得するための言語モデルとしては、Sentence-BERT (S-BERT)[1] の中でも、多言語対応の文強弱タスクに優れた paraphrase-

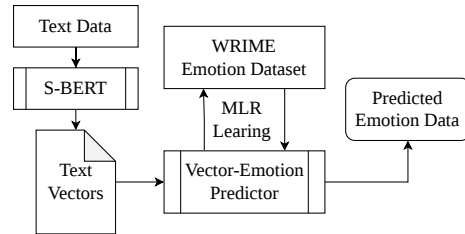


Fig. 1 Learning Flow of Vector-Emotion Predictor

multilingual-mpnet-base-v2 を、多次元的な感情推定を行うための学習データとしては、Robert Plutchik が提唱する 8 基本感情(喜び・悲しみ・期待・驚き・怒り・恐れ・嫌悪・信頼)とその主観的強度を付与した日本語データセット WRIME[2] を使用する。

WRIME に含まれる文章データ 43,200 個に対して、8 割を訓練用に、2 割を評価用として割り当てる。その訓練用データに対し S-BERT で 768 次元のベクトルを作成し、WRIME の感情強度データを教師データとして学習を行う。

Fig. 1 に、学習の流れをブロック図で示す。

2.2 解析アルゴリズムとニーズの構造化

解析の枠組みとして、馬場ら [3] の先行研究における手法を基盤としつつ、従来のポジネガ極性分類から 8 つの基本感情分析へと拡張・深化させる。具体的な手順は以下の通りである。

感情判定 構築した S-BERT + WRIME モデルにより、各対象テキストにおいて最も強度の高い感情成分を判定・分類する。

相関分析 形態素解析を経て抽出された単語群に対し、判定された感情クラスターごとに Jaccard 係数を用いて単語間の共起強度を算出する。

構造化と可視化 算出された結びつきの強さに基づき共起ネットワークを構築する。8 基本感情ごとにネットワークを生成・クラスター分類することで、「どの要素(単語)が、どのような感情と結びついているか」というユーザーの具体的なニーズを可視化する。

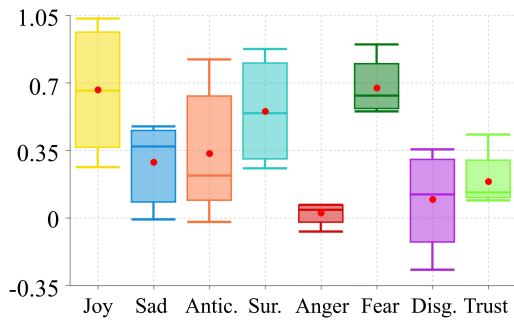


Fig. 2 Distribution Map for Predicted Emotion

3. 初期解析結果と考察

提案手法の有用性を確認するため、実際の口コミデータを用いた感情推定の初期検証を行った。本稿では、歴史的観光資源の1つである熊本城の解析結果を代表例として取り上げ、考察を行う。

3.1 感情分布の可視化

収集した5件の口コミテキストに対し、S-BERT + WRIME を用いた感情推定モデルを適用し、8基本感情の強度ベクトルを算出した。Fig. 2 に、熊本城に対する各感情スコアの分布を箱ひげ図で示す。

3.2 解析結果の考察

Fig. 1 の感情分布から、本手法がテキスト内の複雑な文脈を的確に捉え、二元的な極性判定では表現できない重層的な感情を定量化できていることが確認された。主な特徴は以下の2点である。

歴史・災害的背景に対する恐れを表出 一般的に星評価の高い観光地であれば、喜び (Joy) が突出するのは自明であるが、本解析では恐れ (Fear) のスコアが全体的に高く、かつ分布のばらつきが小さい (中央値が約 0.55 付近で安定している) という特異な結果が得られた。入力テキストを定性的に確認すると、「崩れた石垣という震災の爪痕」や「西南戦争の攻撃を耐え抜いた」といった記述が多く見受けられた。すなわち、モデルが単なる施設の不備に対する不満ではなく、自然災害の脅威や歴史的建造物の被害に対する畏怖や哀惜といった念を「恐れ」として正しく抽出していることが分かる。

驚きと期待の多様性 驚き (Surp.) や期待 (Antic.) の感情スコアは、比較的高い値を示しつつも、最大値と最小値の幅が大きく、訪問者によって感情の揺れ幅が異なることが示唆された。「天守閣が立派に復旧できていたことへの感動 (Surp.) や

「復旧工事が終わった際の再訪への思い (Antic.)」など、復興プロセスという動的な状態にある観光資源ならではの、多様な来訪者の視点が数値として現れていると言える。

3.3 小括

以上の初期解析から、従来の星評価や、単純な感情分析では「高評価」として一括りにされてしまうデータの中にも、実際には畏怖 (恐れ) や復興への驚きといった重層的な感性構造が存在することが確認された。本手法を用いることで、旅行者が観光資源のどの文脈に対し、どのような深い感情を抱いているかを定量的に評価可能であることが実証された。

4. まとめと今後の展望

本中間報告により、多次元感情推定モデルが観光資源の持つ複雑な文脈を抽出可能であることが確認された。今後は以下の手順でニーズの構造化を進める。

データセットの拡大 解析対象を他の観光資源 (自然景観、商業施設など) へ拡大し、リソースごとの感情分布の差異を比較・検証する。

共起ネットワークの構築 抽出された感情クラスター (例: 恐れの高スコアテキスト群) ごとに、形態素解析および Jaccard 係数を用いて単語間の共起ネットワークを生成する。

ニーズの構造化 ネットワークのクラスター分析を行い、「どの要素が特定の感情を牽引しているのか」を視覚的にマッピングすることで、エビデンスに基づく観光資源の価値再定義と改善策の提示を目指す。

参考文献

- [1]Nils Reimers, Iryna Gurevych: "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks", Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP 2019), pp. 3982-3992, 2019.
- [2]梶原 智之, Chenhui Chu, 武村 紀子, 中島 悠太, 長原 一: "主観感情と客観感情の強度推定のための日本語データセット", 言語処理学会第 27 回年次大会発表論文集, pp. 523-527, 2021.
- [3]馬場 優大, 藤生 慎, 森崎 裕磨: "旅行情報サイトに投稿された口コミデータを用いた観光地の改善点抽出システムの提案", AI・データサイエンス論文集, 4 巻, 3 号, pp. 942-951, 2023.