

表情から推定した興味を考慮する観光案内システム

A dynamic tourist information system adjusting user interests estimated from facial expressions

横 勇紀

Yuki Yoko

広島市立大学大学院 言語音声メディア工学研究グループ

Language and Speech Media Engineering Group, Graduate School of Hiroshima City University

概要 新型コロナウイルスの収束から旅行者の数が増加傾向にあり、観光案内システムの需要が高まっている。しかし、従来の観光案内システムは旅行者に対して一方的な案内を行うものが多く、旅行者の興味を考慮した案内を行うことができていない。そこで本研究では、旅行者の表情から、話題に対する興味の有無を推定し、推定結果に基づいて案内内容を変更する観光案内システムを提案する。

1 はじめに

新型コロナウイルスの感染拡大によって大きく減少した旅行者数も 2024 年時点ではぼコロナ前と同程度まで回復しており[1]、人手不足から観光案内システムの需要が高まっている。旅行者が満足する観光案内を行うには、旅行者の反応から興味のもっている名所やジャンルを推測し、相手に合わせた説明をする必要がある。しかし、従来の観光案内システムは旅行者に対して一方的に情報を提示するものが多く、旅行者の興味を考慮した案内が行えていない。その点を考慮し、タッチパネル型の対話型観光案内システムを導入している観光地もあるが、タッチパネルのため観光客に求める操作が多く、使ってみようとした人も対話のテンポの悪さから途中で投げ出してしまふ人が多い[2]。また、三重県伊勢市では、旅行者のニーズに応える観光案内を行うためリモート観光案内システムを導入した[3]が、案内は人手で行っているため、稼働時間や導入できるシステム数に制限がある点が課題となっている。

本研究では、旅行者の表情から興味を推定し、推定した結果に基づいて案内内容を柔軟に変更する観光案内システムを提案する。本システムは、旅行者に対して観光地に関する話題を順に説明していき、その中で興味があるような反応を示したと判断した話題に関しては詳細な説明を行う。逆に明らかに興味が無さそうな反応であれば予定していた説明を端折り、次の話題に移行する。なお、本稿で説明する興味推定処理では、旅行者の表情から興味があるか否かの二値判定を行う。

2 提案手法

2.1 興味を考慮する観光案内システムについて

旅行者の興味を正確に推定するには、カメラに映る旅行者の表情、視線、頭部の動きといったマルチモーダルな特徴量が必要となる。そのため、映像解析によって得られた特徴量を機械学習器によって処理し、興味判定を行う。

提案手法の具体的な構成を図 1 に示す。まず、観光案内システムが旅行者に対して、観光地に関する説明を行う。そして、説明を聞いている旅行者の表情を Web カメラで撮

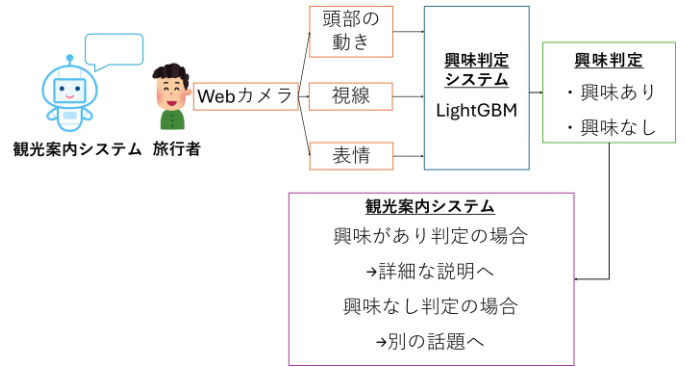


図 1. 提案手法の構成

影し、1 秒毎に画像として切り取る。切り取った画像に対して OpenFace[4]を用いて解析し、頭部の動きと視線、表情の特徴量を抽出する。その特徴量に基づき学習済みモデルが興味あり・興味なしの 2 クラス分類を行う。そしてその分類結果から、興味ありと判定された話題に関しては詳細な説明を行う。また、興味なしと判定された話題に関しては、次の話題に関する説明を行う。

2.2 マルチモーダル情報からの興味度推定

2.2.1 興味度推定に用いる実験データ

本研究では、興味有り・無し の 2 クラス分類を行うため、実験データとして、興味の有る様子と興味の無い様子の動画が収録されているマルチモーダルコーパス Hazumi[5]を使用した。Hazumi では Wizard-of-Oz 方式で動作するエージェントと実験参加者が対話を行っており、実験参加者にとって興味の有る話題と無い話題が半数ずつになるように設定されている。また、実験参加者の発話毎に 6 名または 3 名の第三者が興味度のアノテーションを行っており、興味有り (o)、不明 (t)、興味無し (x) の 3 段階で付与されている。本研究では、過半数のアノテーターが興味有り と付与したデータに興味有りのラベルを付け、過半数のアノテーターが興味無し と付与したデータに興味無しのラベル付け、学習データとして使用した。そのため、興味有り と付与した評価者が 3 名であり、無し と付与した評価者が 3 名だった場合は、学習データとして使用しない。

2.2.2 映像特徴量

OpenFace は顔のランドマーク座標、視線、頭部、AcutionUnit (AU) の強度を検出することができるオープンソースツールである。視線情報に関しては、目のランドマーク 56 点の 2 次元および 3 次元座標出力し、両目の視線を 3 次元視線ベクトルとして推定する。また、頭部の角度を Pitch, Yaw, Roll の 3 成分に変換して推定することも可能である。本研究では、OpenFace で取得できる顔と目のラン

ドマーク（626次元）、視線（8次元）、頭部の動き（40次元）、AU（35次元）の計709次元の特徴量を使用する。

2.2.3 機械学習実験

本稿では、LSTM、SVM、LightGBMの3種類のモデルで機械学習実験を行った。LSTMはバッチサイズを64に設定し、エポック数10、中間層を2層、最適化関数をAdamと設定し、学習率を0.001とした。SVMはカーネル関数をrbfカーネルとし、C値を1.0、 γ 値を0.1と設定した。LightGBMは、木の深さを7にし、最大葉枚数を70、学習率0.05と設定した。なお、どの機械学習モデルもtrain:test=8:2として実験を行い、機械学習器がテストデータを興味あり・なしの2クラスで分類した結果と実際にラベル付けされた結果を比べ、幾つ正解しているかを確認した。実験結果を表1に示す。本実験結果より、興味判定システムには、最も正解率の高かったLightGBMを採用する。

表 1. 興味推定実験の結果

機械学習モデル	LSTM	SVM	LightGBM
正解率	0.61	0.94	0.98

2.3 興味度による案内内容の分岐チャート

本研究では、宮島に訪れた旅行者に向けて案内するシステムを想定し、案内内容の分岐チャートを事前に作成する。分岐チャートの例を図2に示す。図2の分岐チャートでは、「厳島神社」、「揚げ紅葉」、「宮島水族館」の3つの話題について大まかな説明をする。それぞれの話題を説明する時間は15秒とし、各話題について説明している間の聞き手の様子から興味ありと判定された話題については、その話題に関する歴史や行き方など詳しい説明を行う。

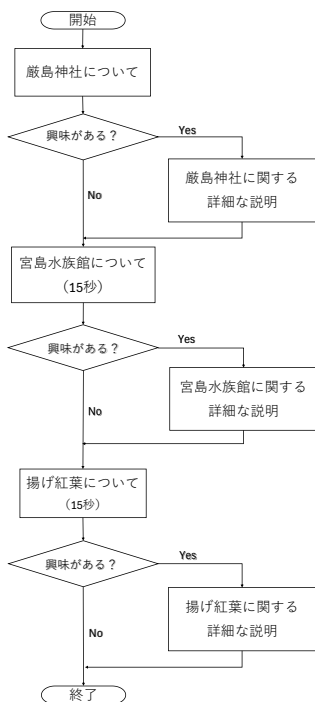


図 2. 興味度による案内内容の分岐チャートの例

3 今後の予定について

本研究では、旅行者の表情から興味推定し、推定結果から案内する内容を変える観光案内システムの提案を行った。

今後は、観光案内システムが話す内容と評価内容を具体的にし、印象評価実験を行っていく予定である。

さらに、今回の研究では興味ありと興味なしの2クラス分類だったが、必要に応じて分類するクラスを増やしていくことに加え、システムが話す内容や分岐を増やすことによって、より柔軟な案内が行えるよう改善する予定である。

4 参考文献

- [1] 国土交通省，令和6年度観光白書，(<https://www.mlit.go.jp/statistics/file000008.html>，2024年6月30日アクセス)
- [2] 香山他，“ユーザの顔向きに基づく推薦を行う音声対話型観光案内システムの構築”，人工知能学会第二種研究会資料SIG-SAI-010-04，2010.
- [3] 伊勢市観光情報，2023，伊勢市リモート案内システム (<https://www.city.ise.mie.jp/kankou/keikaku/1014118.html> 2024年6月29日アクセス)
- [4] T. Baltrusaitis, et al., “OpenFace 2.0: Facial Behavior Analysis Toolkit”, in Proc. International Conference on Automatic Face and Gesture Recognition (FG), IEEE Computer Society, pp.59–66, 2018.
- [5] 駒谷他，“マルチモーダル対話コーパス Hazumi 公開と生体信号を含む新規データ収集”，人工知能学会研究会資料SIG-SLUD-C002-35, pp.170-177, 2020.