

# 観光スポット口コミデータ内の嗜好情報を利用した インバウンド向け観光スポット推薦手法

白井 里奈<sup>1</sup> 上原 宏<sup>2</sup> 鈴木 一哉<sup>3</sup>

秋田県立大学大学院システム科学技術研究科<sup>1</sup> 立正大学データサイエンス学部<sup>2</sup>

秋田県立大学システム科学技術学部<sup>3</sup>

## 1. はじめに

インバウンドの獲得は外貨収入増加に繋がることから、重要な課題の一つである。日本政府は観光業を GDP 成長戦略の柱と位置づけており、インバウンド需要拡大への目標値を幾つ掲げている [1]。本研究では、それらの課題からリピータ観光客獲得に着目する。リピータ獲得のためには、地方の観光資源のプロモーション強化が必要となるが、インバウンドの約 8 割がゴールデンルートと呼ばれる地域への訪問に偏っているという課題があり [1]、それに伴い Web サイト上の観光口コミにも偏りが生じている。

本研究では、インバウンドの嗜好を損なわない形で、日本人旅行者の口コミでインバウンドのデータ量問題を補い、最低 1 度以上訪日経験のあるインバウンドを想定した観光スポットを推薦手法の提案を行う。

## 2. 先行研究と本研究の位置付け

個人の嗜好情報を基にした観光スポット推薦に関する研究は広く行われており、SNS への投稿や GPS データを基にユーザーをプロファイリングする方法などが提案されている。山岸らは、Flickr に投稿されたデータを基に、Who, Where, When, What の四要素に着目した個人化観光システムの提案をしている [2]。

SNS 等から個人の嗜好を抽出する方法は、データの収集に時間と費用を要する 경우가多く、加えて geo-tag などのロケーションデータは有名な観光エリアのものに限定されている場合がある。Flickr 上のデータは比較的容易に収集でき、観光スポット推薦関連の研究で広く用いられるデータである。しかし、日本人ユーザーが少ないため、地方の観光資源も扱うことを目的としている本研究には適さない。

本研究では、先行研究と比べて、取得容易である旅行口コミサイト上の口コミ情報を使用し、ユーザーの口コミ内の潜在的な嗜好情報を個人の嗜好と捉えることで、データ量問題とその収集問題の回避を試みている点で異なる。加えて、提案手法は推薦の根拠情報を提供することが可能であるという点でも他の推薦手法と差別化される。

## 3. 解析データと提案手法

本研究で使用するデータは、世界最大級の旅行口コミサイト TripAdvisor.co.uk より収集した、日本人・外国人合わせて 71,646 件の口コミデータであり、データセットに内在する観光スポット数は 269 スポットである。

図 1 は観光スポット推薦過程を示しており、全 2 部で構成される。第 1 部はインバウンドに対し認知度の低い観光スポットも扱うことを可能とした拡張データセットの生成を示しており、第 2 部は第 1 部で生成した拡張データセットを基に、新規投稿口コミ（拡張データセットに含まれない口コミ）に内在する嗜好情報に応じた観光スポットの推薦過程を示している。

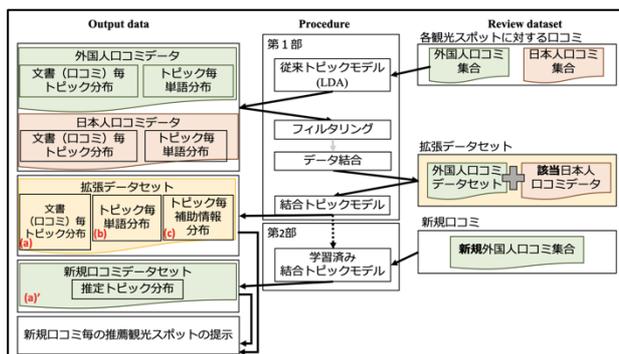


図 1. 提案手法の流れ

冒頭でも述べた通り、本研究はインバウンドには認知度の低い観光スポットを推薦することを目的としているため、日本の観光資源を全国横断的に扱う必要がある。しかし、インバウンドの口コミ数は日本人のものより少ないため、日本人の口コミを外国人データセットに結合す

Sightseeing spot recommendation approach for inbound tourists by using users' preferences information in posted sightseeing reviews

1 Satona Shirai, Graduate School of Systems Science and Technology, Akita Prefectural University

2 Hiroshi Uehara, Faculty of Data Science, Risho University

3 Kazuya Suzuki, Faculty of Systems Science and Technology, Akita Prefectural University

ることでデータ量問題の解決を図る。しかし、外国人と日本人では日本の観光地に対する嗜好が異なるため、両者のデータを単純結合してしまうと、日本人特有の嗜好が外国人にとってノイズとなってしまふ可能性がある。そこで、両者を結合する前に、トピックモデルで広く用いられている手法である latent Dirichlet allocation (LDA) [3]を使用し、両者データセットからそれぞれ嗜好情報を取り出す。その後、両者の嗜好情報を比較し、共通する嗜好を持つ日本人口コミのみをデータセット内から抽出し(図1:フィルタリング)、外国人データセットに結合することで拡張データセットを生成する。この手順を踏むことによって、外国人の嗜好情報を保持したまま、推薦対象観光スポットの拡大を図ることが可能となる。

第1部の最終工程として、上記のステップで生成した拡張データセットをLDAの拡張版である結合トピックモデル[4]で解析することで、各口コミの嗜好トピック情報(a)、各嗜好トピックの解釈情報となる単語分布(b)、各嗜好トピックに対応する観光スポット情報(c)をそれぞれ確率分布で出力できる。ここで出力された確率分布は第2部で使用する。尚、上記で述べた分布(a)(b)(c)は、それぞれ図1で示しているものに対応する。

第2部は、新規投稿口コミ(拡張データセットに含まれない口コミ)から嗜好情報を取り出し、取り出した嗜好情報に対する観光スポット情報を推薦する過程を示す。

まず、第1部で学習された結合トピックモデルのパラメータを使用することで、新規口コミの嗜好トピック分布(a)'を推定できる。推定された嗜好トピック分布を参照し、確率値が高い嗜好トピックに関連する観光スポット分布(c)、根拠情報分布(b)がそれぞれ、推薦観光スポット(c)とその根拠情報(b)として提供される。

#### 4. 実験結果と評価

第2章で示した提案手法に対し、新規投稿口コミ3種類(根津美術館、姫路城、森美術館)を用い定性的評価を行った結果が、表1である。以下の2点において、本手法の独創性が結果に反映されている。

- ・ 根津美術館と森美術館はどちらも美術館であるが、新規投稿口コミの内容によって推薦スポットのカテゴリが異なっている点
- ・ 下線部分の観光スポットは外国人のみのデータセットには含まれないため、拡張データセットの利点が反映されている点

加えて、本研究で用いている結合トピックモデル

は各嗜好トピックを表す単語分布を生成するので、推薦の根拠情報として該当嗜好トピックの単語分布を提示できる。

以上の点から、本研究は訪日外国人にマイナーな観光スポットも推薦対象に含んだ推薦に有効である。

表1: 評価用新規投稿口コミに対する観光スポット推薦推定結果

評価用レビュー	レビュー内容	上位5推薦観光スポット
Nezu Museum	Garden, Exhibition, Café, Shop	Kobe Nunobiki Herb Gardens, Tokyo Metropolitan Government Building Observation Decks, Ritsurin Garden, Okochi Sanso Garden, Koko-en
Himeji Castle	Structure and Information of Caslte	Kumamoto Castle, Okayama Castle, Matsumoto Castle, Osaka Castle Park, <u>Hirosaki Castle</u>
Mori Art Museum	Exhibition, Entry fee	<u>Towada Art Center</u> , <u>Nara National Museum</u> , <u>Sagawa Art Museum</u> , <u>Chishaku-in</u> , Mori Art Museum

#### 5. おわりに

本研究は、取得容易であるデータを元に、新規口コミ内の潜在嗜好に対して、インバウンド向けの多様な観光スポットを含む推薦の手法を提案した。提案手法は、定性的評価により潜在嗜好に基づいた推薦が可能かつインバウンドの訪問が少ないマイナー観光スポットの提示が可能であることが確認された。今後の課題として、評価用データセットの拡大と実証実験による実用性の検討を定量的に評価する必要がある。

#### 参考文献

- [1]. 明日の日本を支える観光ビジョン構想会議, '明日の日本を支える観光ビジョン -世界が訪れたくなる日本へ-', 観光庁, <https://www.mlit.go.jp/common/001126598.pdf>
- [2]. 山岸立, 馬強, ユーザー体験に基づく個人か観光推薦システムの提案, In DEIM Forum 2020 J3-3, 2020
- [3]. Blei, David M., Andrew Y. Ng, and Michael I. Jordan. "Latent dirichlet allocation." *the Journal of machine Learning research* 3 (2003): 993-1022.
- [4]. Mimno, David, et al. "Polylingual topic models." *Proceedings of the 2009 conference on empirical methods in natural language processing*. 2009.