

訪日外国人旅行者を対象とした地域情報マイニング技術の研究開発 (152308003) Development of a Regional Information Mining Technology for Inbound Tourists

研究代表者

難波英嗣 広島市立大学大学院情報科学研究科

Hidetsugu Nanba Graduate School of Information Sciences, Hiroshima City University

研究分担者

竹澤寿幸[†] 奥村学^{††} 倉田陽平^{†††} 石野亜耶^{††††}

Toshiyuki Takezawa[†] Manabu Okumura^{††} Yohei Kurata^{†††} Aya Ishino^{††††}

[†]広島市立大学 ^{††}東京工業大学 ^{†††}首都大学東京 ^{††††}広島経済大学

[†]Hiroshima City University ^{††}Tokyo Institute of Technology ^{†††}Tokyo Metropolitan University

^{††††}Hiroshima University of Economics

研究期間 平成 27 年度～平成 28 年度

概要

近年、訪日外国人旅行者数が増加している一方で、必要な情報を提供するインフラの整備が十分でないという問題があった。そこで、本研究課題では、訪日外国人旅行者を対象にした情報インフラを、知能情報技術を用いて整備することで、開かれた真の国際都市の形成実現を目指した。本研究では、旅行ブログから、(a) 文化や習慣の違いによるトラブルを避けたり、より快適に旅行したりするためのノウハウ情報、(b) 訪れる地域に固有の魅力に関する情報を抽出し、訪日外国人旅行者に提供するシステムを構築し、実験によりその有効性を確認した。

1. まえがき

近年、訪日外国人旅行者数が増加している一方で、必要な情報を提供するインフラの整備が十分でないという問題があった。そこで、本研究課題では、訪日外国人旅行者を対象にした情報インフラを、知能情報技術を用いて整備することで、開かれた真の国際都市の形成実現を目指す。特に、旅行ブログから、(a) 文化や習慣の違いによるトラブルを避けたり、より快適に旅行したりするためのノウハウ情報、(b) 訪れる地域に固有の魅力に関する情報を抽出し、訪日外国人旅行者に提供することを目的とする。

2. 研究開発内容及び成果

本研究開発では、以下のモジュール群を開発した。

- (a) 文化や習慣の違いによるトラブルを避けたり、より快適に旅行したりするためのノウハウ情報の抽出と分析
 - (a-1) ノウハウ情報の集約(地域別)
 - (a-2) ノウハウ情報の集約(ブログ著者の属性別)
 - (a-3) 日本語を対象にした旅行ブログの収集と地域カテゴリへの分類
 - (a-4) 日本語旅行ブログから抽出したノウハウ情報の集約および英語ノウハウとの比較
- (b) 訪れる地域に固有の魅力に関する情報の抽出
 - (b-1) 類似する地域のグループ化
 - (b-2) 地域に固有の魅力の抽出

以下に、各項目の開発状況について述べる。

(a-1, 2) ノウハウ情報の集約(地域別 & ブログ著者の属性別)

代表的な英語旅行ブログサイトのひとつである TravelBlog (<http://traveblog.org>)の全データを対象に、申請者らが開発したシステムを用いて、ノウハウ情報を含んだ 75、730 件の旅行ブログを自動抽出した。この中で日本に関するものは 1,363 件存在した。研究補助員数名で、

これらのブログをすべてチェックし、ノウハウブログデータベースを構築した。各ブログには、もともと TravelBlog で分類されている地域カテゴリが付与されているが、さらに、申請者らが構築したシステムを用いてブログ著者の属性情報(性別、使用言語、旅行者の行動の種類)を付与することで、ノウハウ情報を地域カテゴリおよび著者の属性から検索できるようになった。このデータを使い、例えば、「広島を訪れたフランス語圏の男性のノウハウ情報」といった、きめ細かなノウハウ情報の検索が可能となっている。

(a-3) 日本語を対象にした旅行ブログの収集と地域カテゴリへの分類

TravelBlog の英語旅行ブログには地域カテゴリが人手で付与されているが、申請者らが過去に開発した技術を用いて自動収集した日本語旅行ブログには、こうした地域カテゴリが付与されていないため、特定の地域の旅行ブログだけを閲覧する、といったことができなかった。そこで、ブログ中で記載されている場所を正確に判定する 2 種類の技術の開発に取り組んだ。

第一の技術は、ブログ中で言及されている特定の場所について緯度経度を自動的に付与する、いわゆるジオコーディングと呼ばれる技術を開発した。従来のジオコーディング技術と異なる点は、従来手法では、例えば「広島に旅行した」という文において、「広島」のような広い地域を示す単語に対しても緯度経度を付与するため、位置情報の厳密さに欠けるという問題があった。この問題に対し、本研究では、例えばブログ著者が訪れたレストランやイベント会場などの場所を示す単語列を抽出することにより、より厳密なジオコーディングが可能となった。

第二の技術は、ブログ中の画像から画像認識技術を用いて場所(緯度経度)を推定し、その推定結果をブログの場所と見なす方法である。画像認識技術には Google Cloud Vision API を利用し、実験により、約 90%の精度で、全体の約 1 割の画像の場所を推定することができることがわかった。

(a-4) 日本語旅行ブログから抽出したノウハウ情報の集約および英語ノウハウとの比較

TripAdvisor のクチコミ情報を対象にしたノウハウ情報の抽出および日英ノウハウ情報の比較に関する研究を実施した。TripAdvisor は、世界最大の観光ポータルサイトのひとつで、ある観光スポットを訪れた人々が様々な言語で投稿するクチコミ情報を調べることができる。この投稿のうち、日本語と英語で記述されたものからノウハウ情報を抽出するシステムを開発した。このシステムを使うことで、例えばしまなみ海道サイクリングを訪れた観光客の日本語と英語のクチコミ情報の中から、以下のようなノウハウ情報を抽出することが可能になっている。

- 橋の上は風切音がすごくて会話がわりと難しい。
- レンタル自転車へ行ったところ、アシスト付きは当日返却のみだった。
- 電動自転車・ママチャリ・タンデム自転車など選べる。
- 帽子などをかぶる時には風が結構あるので飛ばされないように注意。

(b-1) 類似する地域のグループ化

旅行ブログには、ワイナリー見学、スキー、サイクリングなど、旅行者の体験について記述されたものがある。旅行者が何らかの体験ができる観光スポットを「体験型スポット」と呼ぶことにする。本研究では、体験型スポットについて書かれた旅行ブログを自動検出し、さらに、類似体験が可能な地域のブログを並べて表示することで、複数の類似地域が比較可能なシステムを構築した。

(b-2) 地域に固有の魅力の抽出

上記(a-3)で述べたブログの位置推定技術を用いて緯度経度情報が付与されたブログを対象に、ある地域に関する複数のブログを要約するシステムを開発した。図3に本研究で開発した旅行ブログエントリ閲覧システムを示す。地図上にエントリ集合を表示しており、画面下のボタンで、“見る”、“体験する”、“買う”、“食べる”、“泊まる”などの、旅行者の目的に沿ったエントリタイプを選択することができる。また、マーカーをクリックするとポップアップでエントリへのリンク付きのタイトルを表示する。



図1 旅行ブログエントリ閲覧システム

次に、本研究で構築した自動要約システム¹を中心に、既存システムからの拡張点について説明する。画面左下の緑のボタンをクリックすると、画面上に表示されているブログを自動要約し、ポップアップウィンドウ内に、その結果が表示される。図2は広島県の宮島付近でタイプ“食べる”に該当するエントリ集合をそれぞれ自動要約した例で

ある。これは、毎年冬に開催される牡蠣まつりに関する要約であり、このように、選択したタイプに沿った要約を自動生成することができる。



○先日、宮島の牡蠣まつりに行ってきました。👉

○かきの浜焼き体験とは、殻つきの牡蠣を、炭火を使って自分で焼いて食べるイベントです。👉

○牡蠣(かき)は最高です(・ω・;)「もみじまんじゅう」やあの「まんじゅう」も送っちゃうかも!👉

図2 広島県 宮島付近、タイプ“食べる”の要約例

3. 今後の研究開発成果の展開及び波及効果創出への取り組み

本研究開発の技術を用い、中国語や韓国語で記述されたブログに対し、5種類の観光タイプ：“見る”、“体験する”、“買う”、“食べる”、“泊まる”および緯度経度情報を推定するシステムを開発中である。また、訪日外国人旅行者向けに英中韓の旅行ブログ情報を提供するサービスの2017年内公開に向け、現在、国内の観光関連企業と調整中である。

4. むすび

本研究課題では、日本語および英語で記述された旅行ブログから旅のノウハウや地域の魅力を自動抽出するシステムを構築し、実験によりその有効性を確認した。

【誌上发表リスト】

- [1] 飯沼俊平、難波英嗣、竹澤寿幸、“場所と旅行者の行動に焦点を当てた複数旅行ブログエントリの自動要約”、観光と情報、Vol.13 No.1 2017.
- [2] 藤井一輝、難波英嗣、竹澤寿幸、石野亜耶、奥村学、倉田洋平、“旅行者の行動分析のための旅行ブログエントリの属性推定” 観光と情報、Vol.13 No.1 2017.
- [3] Fujii, K., Nanba, H., Takezawa T., Takezawa, T., Ishino, A., Okumura, M., and Kurata, Y. “Travellers’ Behaviour Analysis Based on Automatically Identified Attributes from Travel Blog Entries”, In Proceedings of Workshop of Artificial Intelligence for Tourism, PRICAI 2016, 2016

【受賞リスト】

- [1] 飯沼俊平、第8回データ工学と情報マネジメントに関するフォーラム(DEIM 2016) 学生プレゼンテーション賞、“行動タイプを利用した複数旅行ブログエントリの自動要約”、2016

【本研究開発課題を掲載したホームページ】

<http://nlp.indsys.chuo-u.ac.jp/research/scope/>

¹ <http://www.ls.info.hiroshima-cu.ac.jp/blogMap>