

多言語旅行ブログエントリを用いた観光情報提示システム

Travel Information System Using Multilingual Travel Blog Entries

藤井一輝 難波英嗣 竹澤寿幸
Kazuki Fujii Hidetsugu Nanba Toshiyuki Takezawa

広島市立大学大学院情報科学研究科
Graduate School of Information Sciences, Hiroshima City University

1. はじめに

インターネット上には、多くの観光情報が存在しており、その1つとして旅行ブログエントリがある。旅行ブログエントリは、ブロガーが旅行先で実際に体験した体験談や感想などが記載されており、国を問わず多数存在している。そのため、そのような情報は旅行計画を行う際、有益な情報源と考えられる。

本研究では、日本語または英語で記述された旅行ブログエントリを用いて、観光情報提示システムの構築を目指す。このシステムでは、旅行ブログエントリを地図上に提示することにより、視覚的にそれらの情報を提示して、観光地の知識を持たない旅行者でも、観光スポットやその周辺の情報を入手することが可能となる。しかし、地図上の旅行ブログエントリを閲覧すると、宿泊施設や美味しいランチのお店についての旅行ブログエントリが全て混在して提示されてしまう。このため、旅行者がランチのお店について情報を得たい場合、多くの旅行ブログエントリの中から情報を見つけ出す必要がある。そこで、本研究では旅行者の目的に沿って旅行ブログエントリを提示する。

本研究では、旅行ブログエントリを観光の主な目的となる「買う」、「食べる」、「体験する」、「泊まる」、「見る」の5種類のタイプに自動的に分類する手法を提案する。旅行ブログエントリをタイプごとに地図上に提示することにより、旅行者は目的に沿った旅行ブログエントリへ容易にアクセスすることができる(図1)。



図1 システムの動作例

2. 旅行ブログエントリのタイプ分類

本研究では、情報利得とドメイン適応を用いることにより、旅行ブログエントリのタイプ分類を行う。

まず、情報利得(IG: Information Gain)を用いた手法について説明する(IG手法)。情報利得とは、単語のエントロピーの減少量を求め、単語の有無のみを素性とした時、文書を分類するのに適した単語を求めることができる。情報利得

により、収集された各タイプのキーワードを素性とし、機械学習を用いて自動的に日本語、または英語で記述された旅行ブログエントリのタイプ分類を行う。次に、情報利得とドメイン適応(DA: Domain Adaptation)を用いた手法を提案する(IG+DA手法)。本研究では、日本語旅行ブログエントリを日本語ドメイン、英語旅行ブログエントリを英語ドメインとして、ドメイン適応を行う。具体的に、IG手法により収集した日本語ドメインの素性を用いてモデルを獲得し、獲得したモデルを英語ドメインに適応することによって英語で記述された旅行ブログエントリのタイプ分類を行う。その際、英語で記述された旅行ブログエントリを日本語に翻訳する。

3. 実験

実験では、人手によりタイプ判定を行った日本、または英語旅行ブログエントリをそれぞれ1,000件、660件を用いた。機械学習にはSVMを用い、2分割交差検定を行った。評価尺度には、精度・再現率を使用した。旅行ブログエントリは日々作成され膨大に存在するため、本研究では再現率よりも精度を重視した。また、提案手法の有効性を示すため、旅行ブログエントリに出現する全単語を素性した場合をベースライン手法とした。5種類のタイプをそれぞれ2値分類した際の各タイプの平均結果を表1に示す。

表1 タイプへ自動分類した際の平均結果

手法	日本語ブログ		英語ブログ	
	Base	IG	Base	IG+DA
Precision	0.479	0.637	0.148	0.597
Recall	0.197	0.462	0.702	0.337

表1より、各タイプを平均した結果において、提案手法は比較手法に比べて高い精度を得ることができた。以上の結果から、提案手法の有効性を示せたと言える。

4. おわりに

本研究では、日本語の旅行ブログエントリには情報利得を用い、英語の旅行ブログエントリには情報利得とドメイン適応を用いて、各ブログエントリを「買う」、「食べる」、「体験する」、「泊まる」、「見る」の5種類のタイプへ自動分類する手法を提案した。提案手法により分類された旅行ブログエントリを地図上へマッピングし、旅行者の目的に沿った情報を提示するためのシステムを開発した。

外部発表情報

藤井一輝, 石野亜耶, 藤原泰士, 前田剛, 難波英嗣, 竹澤寿幸, “多言語旅行ブログエントリを用いた観光情報提示システム”第6回データ工学と情報マネジメントに関するフォーラム (DEIM2014), 2014.