

広電沿線観光情報提示システムの構築

Construction of a System for Providing Travel Information

along Hiroden Streetcar Lines

石野亜耶¹ 難波英嗣¹ 竹澤寿幸¹

Aya Ishino¹, Hidetsugu Nanaba¹, and Toshiyuki Takezawa¹

¹ 広島市立大学大学院 情報科学研究科

¹ Graduate School of Information Sciences, Hiroshima City University

Abstract: In this paper, we propose a method for identifying Hiroshima Electric Railway (Hiroden) blogs in a blog database. Hiroden blogs are defined as travel journals that provide regional information along Hiroden streetcar stations. To investigate the effectiveness of our method, we conducted some experiments. From the experimental results, we obtained precision of 82.4% and recall of 64.5% in automatic identification of Hiroden blogs.

1. はじめに

2007年1月に「観光立国推進基本法」が施行され、2008年10月には国土交通省の外局として観光庁が設置されるなど、日本では「観光」を21世紀の基幹産業と位置付け、観光を支援する多様な取り組みが積極的に推進されている。現在、広島県では、NHK大河ドラマ「平清盛」に関するイベントが行われている。また、2013年4月からは、日本最大の菓子業界の展示会である全国菓子大博覧会(ひろしま菓子博2013)が開催されるなど、観光客を集める様々な取り組みが行われている。そこで、本研究では、広島の特徴のひとつである、広島電鉄の電車(広電)を使用した観光を支援するための枠組みの一つとして、広電の電停に関する旅行ブログ(電停ブログ)を収集し、路線図にマッピングし旅行者に提示する広電沿線観光情報提示システムの構築を行う。広電沿線観光情報システムを作成することで、ガイドブックに掲載されていない、地域に基づいた情報を発信することができると考えられる。また、近年ブログ著者の属性(性別、年齢、居住域など)を文体や記載内容から自動的に推定する研究が進んでいる[1、2、3]。このような技術を利用し、ブログ著者の属性と、システムの利用者の属性を照らし合わせることで、例えば「女性に人気のレストラン」や「若い人に人気の観光名所」など、利用者に適した観光情報を推薦することができると考えられる。

本論文の構成は以下の通りである。2節ではシステム動作例、3節では関連研究、4節では提案手法、

5節では実験結果と考察について述べ、6節で本稿をまとめる。

2. システム動作例

本研究で構築した広電沿線観光情報提示システムについて、その動作例を紹介する。図1は、広電沿線観光情報提示システムを、iPad上で動作させたときの画像である。図2は、広電沿線観光情報提示システムの画面である。広電の電停および主要な観光名所が描かれている。図2の路線図の一部をクリックすると、拡大路線図を表示することができる。図3は、図2の紙屋町エリア(図中①)をクリックした際の拡大路線図である。図3の電停をクリックすると、その電停に関する電停ブログのリンク集を閲覧することができる。図4は、図3の“原爆ドーム前”という電停をクリックした際に、閲覧することができる電停ブログのリンク集である。本研究では、広電沿線観光情報提示システムで提示する電停ブログを収集する手法を提案する。

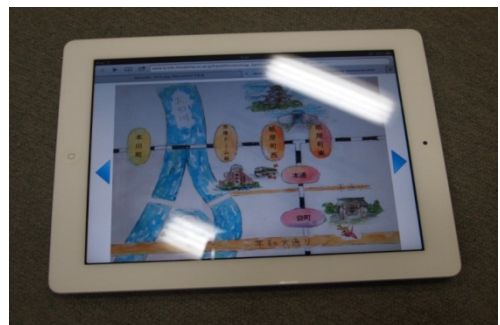


図1: 広電沿線観光情報提示システムの動作例

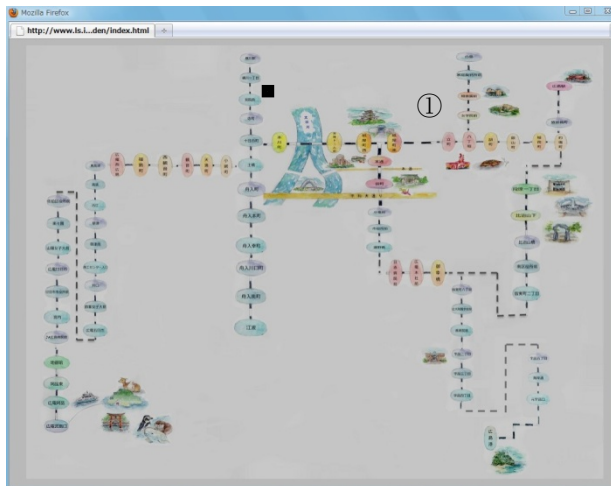


図 2: 広電沿線観光情報提示システムの路線図

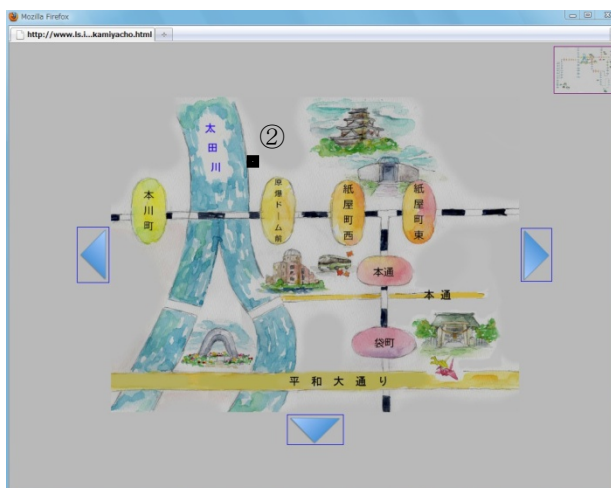


図 3: 紙屋町エリアの拡大路線図

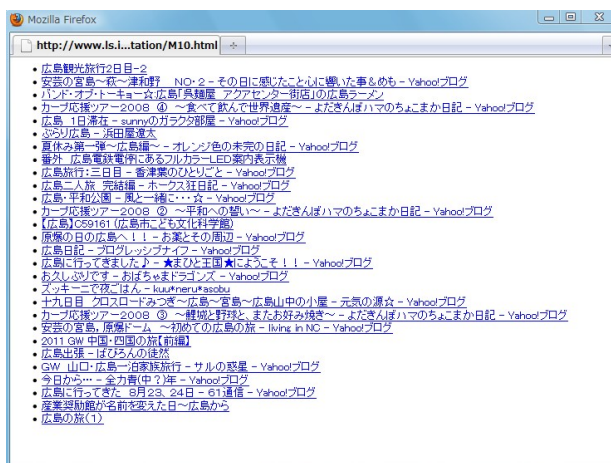


図 4: 電停ブログのリンク集

3. 関連研究

本研究では、広電を使用した観光を支援するための枠組みの一つとして、広電の電停に関する旅行ブログ（電停ブログ）を Web から収集する手法を提案する。本研究と同様に、Web から地域情報を自動収集する研究がある。大槻ら[4]は、地域情報ウェブディレクトリを自動編集するシステムを提案している。地域情報ウェブディレクトリは地域情報検索に利用される。大槻らは、地域情報として自治体が提供する地域情報サイトと、そのリンク先の地域サイトを対象としているが、本研究では、ブログを対象としている点で異なる。

本研究と同様に、ブログを情報源とし、地域情報を自動抽出する研究がある。岡本ら[5]は、一般のブログ検索エンジンを利用することで、地名を含むブログエントリを収集し、それらのブログエントリから、地域イベント情報を抽出する手法を提案している。本研究では、電停ブログを収集することを目的としているため、岡本らの研究とは異なる。

Web から観光情報を収集する研究として徳久ら[6]の研究がある。徳久らは、ブログから、観光開発のヒントとなる文を抽出する手法を提案している。石野ら[7]は、ブログデータベースから、機械学習を用いて旅行ブログを検出する手法を提案している。石野らは、“旅行”、“観光”、“ツアー”などの旅行ブログによく出現する単語の有無を素性に使用している。石野らは旅行ブログの収集を目的としているが、本研究では、電停ブログの収集を目的としている点で異なる。

旅行ブログやそのエントリを登録したポータルサイトとしては、“Travel Blog”¹、“旅行・観光ブログ村”²、“フォートラベル”³などがある。これらのポータルサイトでは、ブロガーが自身のブログを旅行ブログとして登録することで、旅行ブログの集積を行う。しかし、ブログ空間にはたくさんのブログが存在しており、このようなポータルサイトに登録されていない一般ブログの中にも電停ブログが多数存在する。電停ブログのように、ある地域に限定したブログは、有名観光地に関連するブログと比較すると、ブログの件数が少ないと考えられる。よって、本研究では、一般ブログを対象として、電停ブログかどうかを自動判定することで、より多くの電停ブログの収集を行う。

郡ら[8]は、ブログからユーザの行動時の代表的な

¹ <http://www.travelblog.org/>

² <http://travel.blogmura.com/>

³ <http://4travel.jp/>

経路とその文脈を抽出し、それらを地図上にマッピングすることにより、集約して提示するシステムを提案している。また、Davidov[9]は、Web から交通手段や経路の地理的なネットワークを見つける手法を提案している。Ishino ら[10]は、旅行ブログから、機械学習を用いて、旅行者の行動経路を抽出する手法を提案している。これらの研究と、本研究で構築した広電沿線観光情報提示システムを組み合わせることで、旅行者に最適なモデルルートを推薦することができると思われる。

寺西ら[11]は、観光情報雑誌に、旅行ブログや、Yahoo!知恵袋付けることで、より網羅性の高い観光情報を提示する手法を提案している。本研究では、電停ブログを収集し路線図にマッピングし提示するシステムを構築することを目的としている。今後は、Yahoo!知恵袋など他のコンテンツも、広電沿線観光情報提示システムで提示できるよう改良していく予定である。

4. 広電沿線観光情報提示システムの構築

本節では、広電沿線観光情報提示システムで提示する電停ブログの収集手法について説明を行う。電停ブログの収集手法は、以下の2つのステップに分かれている。(1)については4.1節、(2)については4.2節で説明を行う。

- (1) ブログの収集
- (2) 電停ブログの判定

4.1. ブログの収集

電停ブログの収集のためには、各電停に関連する情報が記載されたブログが必要になる。そこで本研究では、各電停の名称(78件)をクエリとしてYahoo!検索(ブログ)で検索を行い、ブログの収集を行った。その結果、1,748件のブログが収集された。

4.2. 電停ブログの判定

本研究では、広電の電停に関するブログや、ブログ著者が広電の電停で下車、観光を行ったブログを電停ブログと定義する。

4.1節で収集したブログには、広島電鉄の電停と同一の地名や他の交通機関の駅名に関する情報が記載されているブログや、観光に関連しないブログが含まれる。本研究では、4.1節で収集されたブログに対し、電停ブログかどうかを、機械学習を用いて自動判定する。

図5は、人手で電停ブログであると判定されたブログの一例である。図5に示すブログのように、電

停ブログには、広電の電停名や、“市電”、“電停”などの広電に関連する単語が頻出する傾向がある。ブログ著者が観光を行った際には、“観光”、“散策”などの単語がよく使われる。また、撮影した写真を掲載する傾向がある。よって本研究では、機械学習に以下の素性を使用することで、電停ブログの自動判定を行う。

- 電停名の出現頻度 (78件)
- 広島電鉄関連の単語(広島電鉄、広電、市電、電停など)の出現頻度 (5件)
- 広島電鉄の電停に関連しない単語(JR、新幹線、フェリーなど)の出現頻度 (6件)
- 旅行関連の単語(観光、散策、撮影など)の出現頻度 (9件)
- 写真の有無

ココに車を停めて後の移動は市内電車**広電市電**と宮島行の船が一日中乗り放題の一日乗車乗船券を買ってまずは**原爆ドーム前**へ
昨今の原発問題もあって、そこらじゅうで署名活動してました
.....(略).....
周辺のテキヤを満喫し**広電**で宮島口まで移動

図5: 電停ブログの一例

5. 実験

本研究で行った実験について説明する。

データセット

実験用データには、4.1節で収集したYahoo ブログ1,748件に対し、人手で電停ブログかどうかの判定を行った結果を用いる。人手で電停ブログの判定を行った結果を表1に示す。

表1: 電停ブログの人手での判定結果

電停ブログ (件)	その他 (件)	合計 (件)
568	1,180	1,748

比較手法

提案手法の有効性を確認するため、4.1節で収集したブログ1,748件を、全て電停ブログとして判定した場合を比較実験とした。

機械学習と評価尺度

電停ブログの判定の機械学習にはTinySVMを用いた。2次の多項式カーネルを使用し、4分割交差検

定を行った。評価尺度として、精度・再現率・F 値を用いた。

実験結果と考察

実験結果を表 2 に示す。表 2 の実験結果より、比較手法に比べ、提案手法が精度・F 値ともに高い数値を記録した。よって提案手法の有効性を示せたといえる。

表 2: 電停ブログの自動判定結果

	精度(%)	再現率(%)	F 値(%)
提案手法	82.4	64.5	72.4
比較手法	32.5	100.0	49.1

提案手法では、精度に比べ、再現率が低い結果であった。本論文では、再現率低下の原因について考察を行う。

再現率の低下の原因は、手掛かり語の不足であった。本研究では、電停ブログの判定を、4.2 節で示した手掛かり語を用いて機械学習により行った。使用した手掛かり語は、広電関連の単語が大部分を占めている。しかし、電停ブログには、電停で下車した後、観光や、食事した状況が詳しく記述される場合がある。この場合、本研究で使用した、広電関連の手掛かり語が、あまり出現しない傾向がある。この問題を解決するためには、収集した電停ブログを解析し、各電停の電停ブログによく出現する観光名所や、レストラン名、土産物の名前などを収集し、手掛かり語として追加することが考えられる。

6. まとめ

本研究では、電停ブログを収集する手法を提案した。電停ブログの収集手法は、(1)ブログの収集、(2)電停ブログの判定の 2 つのステップに分かれる。電停ブログの判定では、精度 82.4%、再現率 64.5%を得られており、提案手法の有効性を示すことができた。また、収集した電停ブログを路線図にマッピングし、旅行者に提示する広電沿線観光情報提示システムの構築を行った。

今後の課題としては、収集した電停ブログを、“観光”や“食事”などの観点で分類し、旅行者が電停ブログを効率的に閲覧することができるようにすることが挙げられる。また、電停に関連する Yahoo!知恵袋や、ニュースなど、様々なコンテンツを自動で収集することで、より網羅性の高い広電沿線観光情報提示システムを構築することが考えられる。

参考文献

- [1] Yasuda, N., Hirao, T., Suzuki, J., and Isozaki, H.: Identifying Bloggers' Residential Areas, Proceedings of AAAI Spring Symposium on Computational Approaches for Analyzing Weblogs, pp.231-236, (2006)
- [2] Ikeda, D., Takamura, H., and Okumura, M.: Semi-supervised Learning for Blog Classification, Proceedings of the 23rd AAAI Conference on Artificial Intelligence, pp.1156-1161, (2008)
- [3] Schler, J., Koppel, M., Argamon, S., and Pennebaker, J.: Effects of Age and Gender on Blogging, Proceedings of AAAI Symposium on Computational Approaches for Analyzing Weblogs, pp.199-205, (2006)
- [4] 大槻 洋輔, 佐藤 理史: 地域情報ウェブディレクトリの自動編集, 情報処理学会論文誌, Vol.42, No.9, pp.2310-2318, (2001)
- [5] 岡本 昌之, 菊池 匡晃: ブログからの地域イベント情報抽出, 情報処理, Vol.51, No. 1, pp.14-17, (2010)
- [6] 徳久 雅人, 奥村 秀人, 村田 真樹: 観光開発支援のためのブログ記事からの評判分析, 観光と情報, Vol.7, No.1, pp.85-98, (2011)
- [7] 石野 亜耶, 難波 英嗣, 竹澤 寿幸: 旅行ブログからの観光情報の自動抽出, 日本知能情報ファジィ学会誌, Vol.22, No.6, pp.667-679, (2010)
- [8] 郡 宏志, 服部 峻, 手塚 太郎, 田島 敬史, 田中 克己: ブログからのビジターの代表的な経路とそのコンテキスト抽出, 情報処理学会研究報告データベースシステム研究会, Vol.2006, No.78, pp.35-42, (2006)
- [9] Davidov, D.: Geo-mining: Discovery of Road and Transport Networks Using Directional Patterns, Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, pp.267-175, (2009)
- [10] Ishino, A., Nanba, H. and Takezawa, T.: Automatic Compilation of an Online Travel Portal from Automatically Extracted Travel Blog Entries, Proceedings of the 18th international Conference on Information Technology and Travel & Tourism (ENTER2011), (2011)
- [11] 寺西 拓也, 野村 達二, 平山 智子, 石野 亜耶, 難波 英嗣, 竹澤 寿幸: 観光ガイドブックへの旅行ブログエントリーと質問応答コンテンツの対応付け, 言語処理学会 第 18 回年次大会, (2012)