

ブログを利用した サイクリスト支援システムの構築

石野 亜耶[†], 難波 英嗣[‡], 竹澤 寿幸[‡]

[†] 広島経済大学 ビジネス情報学科

[‡] 広島市立大学大学院 情報科学研究科

[†] ay-ishino@hue.ac.jp

概要 : 本研究では, サイクリングを目的とした観光客に観光情報として提示するために, 過去にサイクリングを経験した観光客が記載したブログエントリ (サイクリストのブログエントリと呼ぶ) を, 機械学習を利用して自動抽出する手法を提案する. 機械学習には, ブログエントリに含まれるテキスト情報と画像情報を利用する. 実験の結果, 精度 0.78, 再現率 0.73 で, サイクリストのブログエントリを抽出することができた. また, 自動で抽出したサイクリストのブログエントリを地図にマッピングしてユーザに提示するサイクリスト支援システムを構築する.

Keywords : ブログ, サイクリスト, 観光支援

1. はじめに

近年, 日本各地の観光地でサイクリングを楽しむ観光客が増加している. 特に, しまなみ海道は, サイクリストの聖地とも呼ばれており, 国内外の観光客から特に人気があるサイクリングスポットである.

サイクリングスポットでサイクリングを楽しみたい観光客が, 観光情報を収集するために利用できる情報源としては, 地方公共団体が運営している Web サイトなどが挙げられる. しかし, このような Web サイトは, サイクリングイベントなどが開催される際に作成されるものが多く, 常時利用できるものは多くない.

そこで本研究では, サイクリングを目的とした観光客に提示する観光情報のための情報源として, 過去にサイクリングスポットを訪れサイクリングを経験した観光客が記載したブログエントリ (サイクリストのブログエントリと呼ぶ) に注目する. サイクリストのブログエントリには, ブログ著者が経験したサイクリングのルート, レストランや休憩所の情報, 感想などが記載されており, 有用な情報源であると考えられる. 本研究では, サイクリストのブログエントリを自動で抽出し, それらを地図上にマッピングすることで, ユーザがサイクリングに関する情報を閲覧できるサイクリスト支援システムを構築する. このシステムを利用することで, サイクリングをしながら立ち寄りや

すい観光スポットや, 他のサイクリストのサイクリングルートを探ることが可能である.

本論文の構成は以下の通りである. 2 節では, システムの動作例, 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. システムの動作例

本節では, 本研究で構築するサイクリスト支援システムの動作例を説明する. 本システムは, Google Maps API¹を利用している.

図 1 は, しまなみ海道を事例として構築したサイクリスト支援システムを, スマートフォンで動作させたときの起動画面である. 図 1 の「現在地を表示」ボタン (図中①) をクリックすると, 図 2 のように, スマートフォンから取得した位置情報を利用し, 現在地を示す場所に自転車のアイコンが表示される.

図 2 の赤いピン (図中②) をクリックすると, その場所に関連するサイクリストのブログエントリへのリンクが吹き出しに表示される. リンクをクリックすると, サイクリストのブログエントリの本文を閲覧することができる.

¹

<https://enterprise.google.com/intl/ja/maps/products/mapsapi.html>

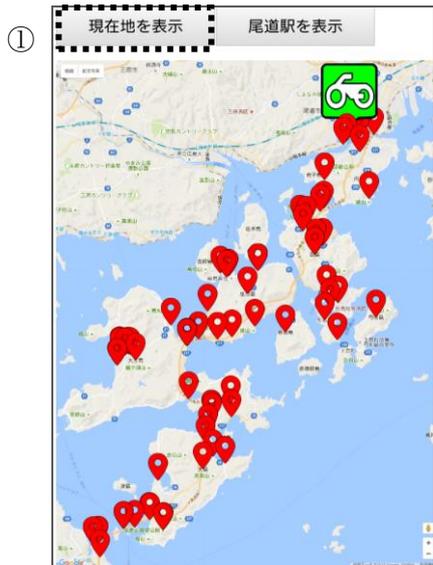


図1 サイクリスト支援システムの起動画面



図2 サイクリスト支援システムの動作例

図3は、サイクリスト支援システムで提示されるサイクリストのブログエントリの一例である。このサイクリストのブログエントリからは、サイクリングルートや立ち寄ったレストランの情報を得ることができる。また、ブログエントリに含まれる画像を閲覧すると、ブログ著者が訪れたレストランの前には、自転車用の駐輪場があり、サイクリストが立ち寄りやすいお店であることも確認できる。

本研究では、サイクリスト支援システムで提示するためのサイクリストのブログエントリを自動で抽出する手法を提案する。



図3 サイクリスト支援システムで提示されるサイクリストのブログエントリの例 (<http://blogs.yahoo.co.jp/datnadiary/34732024.html>)

3. 関連研究

本研究では、サイクリストのブログエントリを自動で抽出する手法を提案している。本研究と同様に、観光に関連するブログエントリを抽出する研究がある。Nanbaら[1]は、ブログデータベースから、旅行ブログエントリを自動で抽出する手法を提案している。Ishinoら[2]は、広電の電停に関する旅行ブログエントリを収集し、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している。Fujiiら[3]は、英語のブログサイトから収集した旅行ブログエントリを、旅行者の行動の種類により自動分類する手法を提案している。

また、ブログから観光情報を抽出するための研究もある。徳久ら[4]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。郡ら[5]は、ブログからユーザの行動時の代表的な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。

これらの研究は、ブログを利用した観光支援という点では類似しているが、本研究では特にサイクリストのブログエントリを利用する点で異なる。

4. サイクリストのブログエントリの自動抽出

本節では、サイクリスト支援システムで提示するサイクリストのブログエントリの自動抽出手法について説明する。サイクリストのブログエントリの収集手法は、以下の2つのステップに分けられる。(1)については4.1節、(2)については4.2節で説明する。

- (1) ブログエントリの収集
- (2) サイクリストのブログエントリの自動抽出

4.1 ブログエントリの収集

まず、「サイクリングスポット名 サイクリング」というキーワードをクエリとし、Yahoo!ブログ検索²で検索することで、サイクリストのブログエントリの候補となるブログエントリを収集する。サイクリングスポットとしては、以下に示す9箇所を使用する。また、検索条件は、ブログエントリの更新日時順とする。

- ・ 丘のまち美瑛
- ・ 山中湖
- ・ 琵琶湖
- ・ 天橋立
- ・ 淡路島
- ・ 中海
- ・ しまなみ海道
- ・ さざなみ海道
- ・ とびしま海道

4.2 サイクリストのブログエントリの自動抽出

本研究では、サイクリストが実際にサイクリングスポットを訪れ、サイクリングを体験したことが記載されているブログエントリを、サイクリストのブログエントリと判定する。しかし、4.1.節で収集したブログエントリには、「サイクリングスポット名」と「サイクリング」という単語は含まれるが、サイクリストのブログエントリではないものも含まれる。例として、いつかサイクリングスポットでサイクリングしたいという願望が記述されたブログエントリなどがある。

そこで本研究では、4.1 節で収集されたブログエントリに対し、サイクリストのブログエントリかどうかを、機械学習を用いて自動で判定する。機械学習の素性には、ブログエントリに含まれるテキスト情報と、画像情報を利用する。

まず、テキスト情報について説明する。テキスト情報としては、ブログエントリに含まれる単語の出現頻度を利用する。

次に、画像情報について説明する。サイクリストのブログエントリに掲載されている画像には、サイクリングに利用した自転車や、サイクリングの途中で撮影した景色が多く含まれているため、画像情報もサイクリストのブログエントリを判定する際の重要な手掛かりになると考えられる。そのため、画像に含まれる物体の情報を、機械学習の素性として使用する。

画像の物体検知には、Cloud Vision API³を利用する。例えば、図4のようなサイクリングの様子

を撮影した画像をCloud Vision APIで解析すると、「cycling」、「road bicycle」などのラベルを得ることができる。



図4 サイクリングの様子を撮影した画像

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911件に対して、人手でサイクリストのブログエントリかどうかを判定した結果を用いる。人手でサイクリストのブログエントリの判定を行った結果を表1に示す。

表1 サイクリストのブログエントリの手手での判定結果

| サイクリストのブログエントリである (件) | サイクリストのブログエントリでない (件) | 合計 (件) |
|--------------------------|--------------------------|-----------|
| 619 | 292 | 911 |

比較手法

提案手法の有効性を確認するために、以下の手法により実験を行う。

- ・ 比較手法：テキスト情報（単語の出現頻度）のみを素性として利用する。
- ・ 提案手法：テキスト情報（単語の出現頻度）と画像情報（画像に含まれる物体のラベル）を素性として利用する。

機械学習と評価尺度

サイクリストのブログエントリの判定にはSVMを利用した。SVMには、機械学習ライブラ

² <http://blogs.yahoo.co.jp/>

³ <https://cloud.google.com/vision/docs/?hl=ja>

りの scikit-learn⁴を用いた。2 次の多項式カーネルを使用し、5 分割交差検定を行った。評価尺度として、精度・再現率を用いた。

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、提案手法では、精度は 0.11 ポイント、再現率は 0.02 ポイント改善することができた。よって、テキスト情報に加え、画像情報を利用する提案手法の有効性が確認できたといえる。

表 2 サイクリストのブログエントリの自動抽出結果

| | 精度 | 再現率 |
|------|------|------|
| 提案手法 | 0.78 | 0.73 |
| 比較手法 | 0.67 | 0.71 |

サイクリスト支援システムの有用性評価

本研究で構築したサイクリスト支援システムを、大学生 8 名にしまなみ海道でサイクリングを行う際に利用してもらい、使用感についてアンケート調査を行った。その結果、以下のような意見が得られた。サイクリストのブログエントリを閲覧できる点は、高評価であった。

- ・ 他の Web サイトと比較すると、サイクリスト支援システムでは、実際にサイクリングで訪れた人のブログエントリが閲覧できるため、どのような場所かがわかりやすかった。
- ・ サイクリスト支援システムは、サイクリングの途中で、休憩場所や寄り道先を探すのに役立った。

6. おわりに

本研究では、サイクリストのブログエントリを、機械学習により、テキスト情報と画像情報を利用することで自動で抽出する手法を提案した。また、提案手法の有効性を確認するために実験を行った。実験の結果、精度 0.78、再現率 0.73 という結果を得ることができ、テキスト情報に加え画像情報を利用する提案手法の有効性を確認することができた。また、自動で収集したサイクリストブログを地図にマッピングすることでサイクリスト支援システムを構築した。

今後の課題としては、ジオコーディングの自動化や、多言語ブログエントリへの拡張が挙げられる。

謝辞

本研究の一部は、総務省による戦略的情報通信研究開発推進制度 (SCOPE) の支援を受けて行われた。

サイクリスト支援システムの構築にご協力頂いた、石野ゼミの小川郁生さん、川上拓哉さん、藤村雄偉さん、梅本優毅さん、藏本佑佳さん、藤井翼さんに感謝します。

参考文献

- [1] Nanba, H., Taguma, H., Ozaki, T., Kobayashi, D., Ishino, A. and Takezawa, T.: Automatic Compilation of Travel Information from Automatically Identified Travel Blogs, In Proceedings of Joint Conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing, pp.205-208, 2009.
- [2] Ishino, A., Nanba, H. and Takezawa, T.: Construction of a System for Providing Travel Information along Hiroden Streetcar Lines, In Proceedings of the 3rd IIAI International Conference on e-Services and Knowledge Management, 2012.
- [3] Fujii, K., Nanba, H., 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.
- [4] 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.
- [5] 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.

⁴ <http://scikit-learn.org/stable/index.html>