

観光の形態に基づいた旅行ブログエントリの自動分類

柴田 有基[†] 篠田 広人^{††} 難波 英嗣^{††} 石野 亜耶^{†††} 竹澤 寿幸^{††}

[†] 広島市立大学情報科学部 〒731-3194 広島県広島市安佐南区大塚東 3-4-1

^{††} 広島市立大学大学院情報科学研究科 〒731-3194 広島県広島市安佐南区大塚東 3-4-1

^{†††} 広島経済大学経済学部 〒731-0138 広島県広島市安佐南区祇園 5-37-1

E-mail: [†], ^{††} {shibata, shinoda, nanba, takezawa}@ls.info.hiroshima-cu.ac.jp, ^{†††} ay-ishino@hue.ac.jp

あらまし 本研究では、テキスト情報と画像情報、テキストにリンク付けされた Wikipedia の情報を解析し、旅行ブログエントリを観光の形態に基づいて自動分類する手法を提案する。近年、旅行者は従来の娯楽を追求するのみの観光だけではなく、様々な観光の形態が誕生し、現在もその多様化は進んでいる。例えば、健康回復や維持、増進につながる観光はヘルストourism、災害被災跡地や戦争跡地など、人類の死や悲しみを対象にした観光はダークtourismと呼ばれる。このような観光の形態を旅行ブログから自動判定することで、世界各地の観光地でどのような形態の観光が可能か調べることができる。また、特定の形態に基づいた観光地の推薦なども可能になると考えられる。機械学習を用いて旅行ブログエントリを6種類の観光の形態に自動分類する実験では、テキストと画像を考慮した分類で、精度 0.585、再現率 0.353 を得た。さらにリンク付けされた Wikipedia の情報を加えた分類で、精度 0.592、再現率 0.345 を得た。

キーワード tourism, 文書分類, 分散表現, 深層学習, Wikification

1. はじめに

近年、観光は従来の娯楽を追求するのみだけではなく、様々な形態が誕生し、現在もその多様化は進んでいる。例えば、健康回復や維持、増進につながる観光はヘルストourism、スポーツを体験または観戦することを目的とした観光はスポーツtourismと呼ばれる。旅行ブログにおいて、このような観光の形態の自動判定が実現すれば、世界各地の観光地でどのような形態の観光が可能か調べることができる。また、特定の形態に基づいた観光地の推薦や旅行計画も可能になると考えられる。そこで本研究では、9種類の観光の形態を定義し、機械学習を用いて旅行ブログエントリをこれらの観光の形態に自動分類する手法を提案する。

ある観光地における旅行者の情報を知るための従来の方法の一つに、旅行者に対して直接アンケートを実施する方法がある。アンケートでは、知りたい情報に関する質問を用意することで、欲しい結果が手に入りやすいというメリットがあるが、多くの時間やコストがかかるというデメリットも存在する。そこで、近年、TwitterなどのSNSやWeb上で公開されている旅行記、すなわち旅行ブログエントリを収集・分析するという方法が広まりつつある。特に、旅行ブログエントリには、各観光地における体験談や写真など、詳しくまとまった情報を持つものが多いことから、本研究では旅行ブログエントリを分類の対象に分析を行う。

本論文の構成は以下の通りである。2章では、本研究に関する関連研究について述べる。3章では、観光の形態の定義、自動分類の方針について述べる。4章では、実験内容とその結果、考察について述べる。5章で本論文をまとめ、6章で今後の課題について述べる。

2. 関連研究

本研究では、テキスト情報と画像情報、テキストに含まれる単語に関する Wikipedia の情報を用いて、旅行ブログエントリを観光の形態に基づいて自動分類するが、これまでも Tweet やブログエントリなどを対象とした分類や属性推定の研究は行われている。

テキストを用いた研究について、Gootら[1]は、ツイートのテキストを抽象的な内容に置き換えたもので性別予測を行う手法を提案している。これにより、語彙に依存しなくなることで言語横断な性別予測を可能にしている。

文書分類に用いる手法として、Iyyer[2]らは、文書に含まれる単語を、Word2vecを用いてベクトルに変換し、それらを平均したものを分類に用いる手法 DAN(Deep Averaging Networks)を提案している。DANの概略図を図2.1に示す。この図では、“Predator is a masterpiece”のそれぞれの単語を赤い四角で表現されているベクトルに変換し、それらのベクトルを次元ごとに平均することで、緑の四角で表現されている新たなベクトルを生成する。この緑のベクトルをいくつかの層で処理し、最後のsoftmax層で判定する。この手法の特徴は、単純で理解しやすい上、

計算時間が短く、精度が高いことである。本研究では、分類器を構築するにあたって、この DAN を用いる。

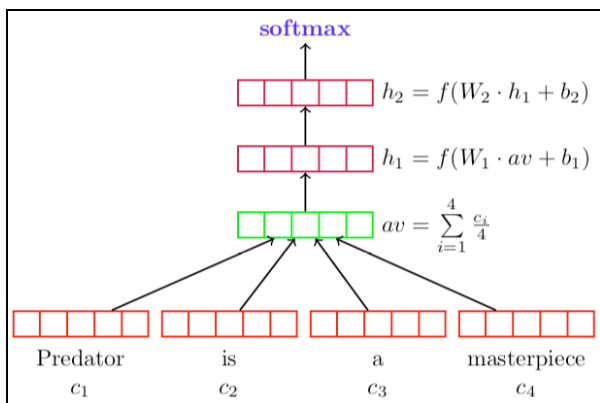


図 2.1: DAN の概略図(Iyyer[2]より引用)

画像を用いた研究について、鈴木ら[3]は、宿泊施設が宿泊予約サイトに掲載する自施設の写真を選定する際に参考となる知見を得るために、宿泊予約サイトに登録されている画像を Google Cloud Vision API¹を用いて分析している。なお、本研究も同様に Google Cloud Vision API を用いて、物体検出を行う。

テキストと画像を用いた研究について、三條ら[4]は、料理レシピの画像特徴とテキスト特徴から人気を予測するための深層学習アーキテクチャを提案している。テキストと画像を考慮する点や分散表現については DAN を用いていることなど類似点が多くあるが、本研究では画像認識技術による物体検出を利用している点に対し、三條らは、CNN(Convolutional Neural Network)を用いて特徴を得ているので異なると言える。また、テキストと画像に加えて本研究では、Wikipedia の情報を用いる点でも異なる。

観光に関するカテゴリ別の文書分類について、野沢ら[5]は、訪問ツイートを「観光」「ビジネス」「食事」「購買」の 4 つのクラスに自動分類する手法を提案している。また、藤井ら[6]は、旅行ブログエントリを「買う」「食べる」「体験」「泊まる」「見る」に分類する手法を提案している。これらの分類の観点は、旅行者の行動、つまり旅行者が何をやっているのかに基づいたものである。本研究で実施しようとしている「観光の形態」に着目した分類と関連性はあるものの、基本的には別の観点であると考えられる。また、藤井らの分類手法に加えて、本研究で提案する観光の形態に基づく分類が可能になれば、「カルチュラルツーリズム」で「食べる」に関する情報を調べる、といったように、よりきめ細かい検索が可能になると考えられる。

3. 観光の形態に基づいた旅行プランの自動生成

本研究では、旅行ブログエントリからブログ著者の観光の形態を自動判定することを目的とする。情報源として、旅行ブログエントリが登録されている TravelBlog²を利用

する。TravelBlog とは、旅行ブログエントリをメインとした海外の Web サイトであり、世界各国に関する旅行ブログエントリが写真とともに掲載されている。本研究では、英語で書かれた旅行ブログエントリのテキストと画像を対象に分析する。

3.1. 観光の形態の定義

観光の形態には厳密な定義はないため、使用者によって異なる解釈で用いられる場合がある。本研究では、それぞれの観光の形態に対して、過去の文献[7-11]を参考に独自に定義した。観光の形態とその定義、またその具体例を表 3.1 に示す。本研究では、表 3.1 の 9 種類の観光の形態に基づいて旅行者の観光の形態を明らかにするため、個々の英語で書かれた旅行ブログエントリを自動分類する。

現在、観光の形態は独自に定義した 9 種類以外にも多く存在する。しかし、今回は分類をする上で有用であると思われる、かつある程度自動分類が可能であろうと思われる 9 種類とした。

3.2. 観光の形態に基づいた旅行ブログエントリの自動分類

本研究は、3.1 節で示した観光の形態に基づいて、旅行ブログエントリを自動分類する。本節では、3.2.1 節で自動分類の方針、3.2.2 節で機械学習を用いた旅行ブログエントリの自動分類について説明する。

3.2.1. 自動分類の方針

自動分類の基本的な方針として、各旅行ブログエントリ中のテキストと画像、テキストに含まれる単語に関する Wikipedia の情報を解析し、それらの結果を用いて「観光の形態」を自動分類する。まず、具体的なテキスト解析による分類の例を図 3.1 に示す。これは岐阜県の「白川郷」を訪れた外国人の方が書いたブログの一部である。ブログのテキストから“UNESCO Heritage Site”という単語が見られるため、これは「ヘリテージツーリズム」と考えられる。

The Shirakawa-Go village is not a very big place as I soon found out after bussing my way to the viewing point for an overview of the place.

Even though this is a **UNESCO Heritage Site**, I liked that the authorities have not overly exploited the site for tourist dollars.



ヘリテージツーリズム

図 3.1: テキスト解析による分類の例

ここで、テキストを対象にした場合、もしブログエントリ中に「スキー」という表現が存在しても「スキーをしたかったけどできなかった」という記述であれば、ブログ著者は実際にスキーをしているわけではない。これに対し、スキーの画像がブログエントリ中にあれば、ブログ著者は

¹ <https://cloud.google.com/vision/?hl=ja>

² <https://www.travelblog.org/>

表 3.1: 観光の形態の定義と具体例

観光の形態	定義	例
インフラ, ハードツーリズム[6]	近代的な建造物や娯楽施設を対象にした観光.	橋, ダム, テーマパーク, ショッピングモール, 水族館, 博物館, 動物園
ヘルスツーリズム[4]	心身を癒すことや散歩などの軽い運動を通して健康維持を目的とした観光.	宗教的巡礼, 温泉, ハイキング, トレッキング
ダークツーリズム[5]	災害被災跡地や戦争跡地など, 人類の死や悲しみを対象にした観光.	東北大震災, 福島原発事故, World Trade Center, 原爆ドーム
コンテンツツーリズム[7]	アニメや映画などの聖地巡礼や作品に関連した施設を対象にした観光.	千と千尋の神隠し, 坊ちゃん
スポーツツーリズム[4]	スポーツを体験または観戦することを目的とした観光.	MLB, プロ野球, サッカー
グリーンツーリズム[4]	自然と触れ合うことを目的とした観光.	農業(漁業)体験, フルーツ狩り, ピクニック
ヘリテージツーリズム[4]	世界遺産や歴史的な建築物を対象にした観光.	世界遺産, 国宝, 寺, 神社, 城
カルチュラルツーリズム[8]	それぞれの地域の生活や文化, 民族, 伝統などを対象にした観光.	着物体験, 神楽, 祭り, 初詣
スタディツーリズム[4]	留学や修学旅行など学習を目的とした観光.	留学, 修学旅行

確実にスキーをしていると判断できる。図 3.2 はスキーをしたと思われる観光客が書いたブログエントリに含まれる画像を Google Cloud Vision API で解析した例である。Google Cloud Vision API の解析結果を見ると、赤字に示す通りスキー関係の用語が並んでおり、この画像だけでスキーをしているとわかることから、この旅行ブログエントリは「スポーツツーリズム」であると判断できる。このように、ブログエントリを分類する際には、テキスト情報に加えて画像情報も分類の際に重要な情報になると考えられる。



図 3.2: Google Cloud Vision API による画像解析結果の例
 人手で分類を行う際に根拠となる情報源は、上記で述べたテキストや画像のどちらかとなるが、旅行ブログエントリの内容によっては、テキストや画像から読み取ることができたモノに関する外部知識が必要になることがある。図 3.3 に外部知識が必要な例を示す。これは、奈良県の「東大寺」を訪れた外国人が書いたブログの一部である。このブログを観光の形態に分類する場合、「東大寺」が世界遺産のため、「ヘリテージツーリズム」となる。しかし、図

3.1 のようにテキスト中に「東大寺」が世界遺産であるという情報がないため、これを正しく「ヘリテージツーリズム」と判断するには、外部知識が必要になると考えられる。これについては、Google Cloud Natural Language API³を利用する。Google Cloud Natural Language API では、テキストを API 経由でクラウドに送ると、形態素解析、構文解析、固有表現抽出に加え、Wikification[12, 13]と呼ばれる、Wikipedia の関連エントリとのリンク付けも行われることから、リンク先の Wikipedia の該当ページから東大寺が世界遺産であるという情報が得られると推測される。そこで、本研究では、旅行ブログエントリに含まれるテキストと画像中の物体に加えて、外部知識としてテキストにリンク付けされた Wikipedia の情報を与えることで、より精度の高い分類の実現を目指す。



図 3.3: 外部知識が必要な例

3

<https://cloud.google.com/natural-language/?hl=ja>

3.2.2. 機械学習を用いた旅行ブログエントリの自動分類

本研究では、まず、分類対象のブログエントリに含まれる画像に対し、画像認識技術を用いて物体検出する。次に、テキストに含まれる単語に対し、リンク付けされた Wikipedia の abstract に含まれる単語を抽出する。なお、abstract には、Wikipedia の最初の段落の内容が記述されている。そして、物体検出の結果として得られた単語集合と Wikipedia の abstract から得られた単語集合、ブログエントリのテキストを入力とし、DAN を用いて分類器を構築する。分類器の概略図を図 3.4 に示す。本研究では、図 3.4 のように旅行ブログエントリ中のテキストと画像、Wikification によってリンク付けされた Wikipedia の該当ページの abstract に含まれる単語集合のそれぞれの入力ごとに処理をし、最後に統合することで分類を行う形式とする。

観光の形態には様々な種類があるが、地域によって過去の歴史や土地柄などの理由から、偏りがあると推測される。例えば、広島県廿日市市では世界遺産の「厳島神社」が有名であることから、観光の形態の一つである「ヘリテージツーリズム」が多くなると考えられる。こうした情報を明らかにすることにより、各訪問地の特徴を活かした観光施策を行うことができると思われる。また、世界遺産の「日光の社寺」がある栃木県日光市も同様の傾向になるようであれば、似た観光地同士で観光施策を参考にすることも期待される。

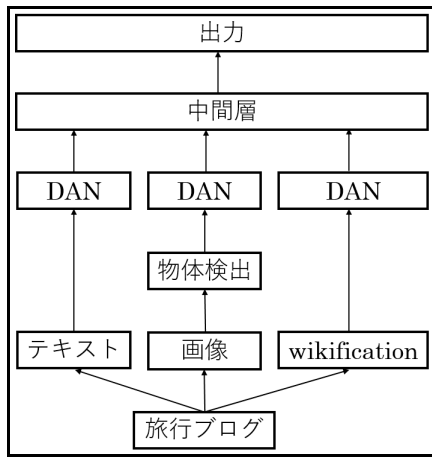


図 3.4: 分類器の概略図

4. 実験

本実験で使用する旅行ブログエントリでは、3 章で述べた TravelBlog のテキストデータと画像データを用いる。

4.1. 実験条件

【実験に用いるデータ】

2,017 件の旅行ブログエントリを実験に用いた。各ブログエントリに対し、観光の形態 9 種類のカテゴリに人手で分類し、これらを機械学習の際の訓練用および評価用データとして用いた。人手で分類した結果の内訳を表 4.1

に示す。なお、一つの旅行ブログエントリには複数の観光の形態が付与されることや逆に一つも付与されないこともありうるという設定にしているため、9 つの件数の総和が 2,017 より小さくなっている。また、表 4.1 からわかるように観光の形態によっては、ブログ件数に偏りがある。そこで本実験では、比較的ブログ件数が多い表 4.1 中の観光の形態 1, 2, 5, 6, 7, 8 のデータで分類実験を行う。

【実験条件】

単語の分散表現には、Google News を対象に Word2vec を用いて獲得された 300 次元のモデル⁴を用いる。画像認識には、3.2.2 節で説明した Google Cloud Vision API を用いる。DAN では、入力データの前処理としてあらかじめ stopword を削除したものをを用いる。

【評価尺度】

実験では、5 分割交差検定を行い、評価尺度には、精度・再現率・F 値を用いる。また、これらを算出するにあたって、観光の形態ごとのデータ件数の偏りを考慮するため、micro 平均を採用する。また、本研究では、多クラス分類を行う方法として、それぞれの観光形態での 2 値分類を拡張することで実現する。

表 4.1: 人手で分類した結果の内訳

観光の形態	件数
1. インフラ, ハードツーリズム	168
2. ヘルスツーリズム	125
3. ダークツーリズム	19
4. コンテンツツーリズム	15
5. スポーツツーリズム	57
6. グリーンツーリズム	450
7. ヘリテージツーリズム	193
8. カルチュラルツーリズム	48
9. スタディツーリズム	2
判定したブログエントリの総数	2,017

【比較手法】

本実験では、以下に示す 2 種類の提案手法と 6 種類のベースライン手法で実験を行った。なお、入力データについて、“(txt)”はテキストに含まれる単語集合を、“(img)”は画像から物体検出の結果として得られた単語集合を、“(wiki)”はテキスト中の単語に関する Wikipedia の abstract に含まれる単語集合を、“(txt+img)”は“(txt)”の入力データに“(img)”の入力データを加えたものを、“(txt+img+wiki)”は“(txt+img)”の入力データに“(wiki)”の入力データを加えたものを意味する。

提案手法

- DAN(txt+img) : 2 つの入力データそれぞれに対し、Word2vec を用いて 300 次元のベクトルに変換し、DAN を適用した分類器で分類を行う。パラメータについて、epoch 数は 100 とする。

⁴ GoogleNews-vectors-negative300.bin.gz

- DAN(txt+img+wiki) : 3つの入力データそれぞれに対し, Word2vecを用いて300次元のベクトルに変換し, DANを適用した分類器で分類を行う. パラメータについて, epoch数は100とする.

ベースライン手法

- fastText(txt) : fastText[14]とは, Facebookが提供する単語のベクトル化とテキスト分類をサポートした機械学習のライブラリである. 計算時間が短い上, 精度が高いという特徴がある. パラメータについては, Word2vecの次元数は300, epoch数は100, 単語はuni-gramとする.
- SVM(txt) : カーネル関数については, RBFカーネルを用いる. また, ソフトマージンのパラメータCについては, C=10とする.
- ランダムフォレスト(txt) : 入力データ中の異なるデータで決定木を複数活用し, 多数決をとることで分類を行う手法. パラメータについて, 決定木の数は10とする.
- DAN(txt) : テキストに含まれる単語集合に対し, Word2vecを用いて300次元のベクトルに変換し, DANを適用した分類器で分類を行う. パラメータについて, epoch数は100とする.
- SVM(img) : カーネル関数には, RBFカーネルを用いる. また, ソフトマージンのパラメータCについては, C=100とする.
- DAN(img) : 物体検出の結果として得られた単語集合に対し, Word2vecを用いて300次元のベクトルに変換し, DANを適用した分類器で分類を行う. パラメータについて, epoch数は100とする.
- DAN(wiki) : テキスト中の単語に関するWikipediaのabstractに含まれる単語集合に対し, Word2vecを用いて300次元のベクトルに変換し, DANを適用した分類器で分類を行う. パラメータについて, epoch数は100とする.

4.2. 実験結果と考察

4.1節で説明したそれぞれのベースライン手法と提案手法による実験結果を表4.2に示す. これより, 精度では画像を考慮したSVMで最も高い値0.760を得ている. しかし, 再現率が0.165と極端に低いため, 取りこぼしているものが多いことがわかる. 再現率については, テキストを考慮したDANで最も高い値0.398を得ている. F値については, 提案手法のテキストと画像を考慮したDANで最も高い値0.440を得た.

表 4.2: 観光の形態に基づくブログエントリの分類結果

分類手法	精度	再現率	F 値
DAN(txt+img)(提案手法)	0.585	0.353	0.440
DAN(txt+img+wiki) (提案手法)	0.592	0.345	0.436
fastText(txt)	0.409	0.275	0.329
SVM(txt)	0.636	0.282	0.391
ランダムフォレスト(txt)	0.579	0.135	0.219
DAN(txt)	0.472	0.398	0.432
SVM(img)	0.760	0.165	0.271
DAN(img)	0.497	0.264	0.345
DAN(wiki)	0.314	0.243	0.274

ところで, それぞれのDANに注目すると, 再現率ではテキストのみを考慮したほうが高くなったが, テキストと画像, Wikipediaの3つの入力データから, 用いるデータの数を増やしていくことで徐々に精度が上がっていき, 3つの入力データ全てを考慮した分類が最も高い精度であることがわかる. そこで, 実際にDAN(txt)では分類できなかったが, DAN(txt+img)では分類できた例を図4.1に示す. これは北京の「天壇」という世界遺産を訪れた旅行者のブログの一部である. 世界遺産を訪れていることから, 「ヘリテージツーリズム」であるが, テキストのみでは世界遺産という情報もないためか正しく分類されなかった. しかし, 画像を含めた場合, 画像に対する物体検出の結果から“temple”をはじめ, 多くの世界遺産と関連があると思われる単語が並んでいる. よって, それが根拠となり, このブログエントリは「ヘリテージツーリズム」と判定されたと考えられる.

- title: The Temple of Heaven
(補足: 北京の「天壇」という**世界遺産**)
- text: The Temple was completed in 1420...
- 画像解析結果: town, **temple**, tourism, **place of worship**, plaza, tours, **shrine**



The Temple of Heaven

図 4.1: 画像解析結果を加えることで正しく分類できたブログエントリの例

次に, DAN(txt+img)では分類できなかったが, DAN(txt+img+wiki)で分類が成功したブログの一部を図4.2に示す. ブログのタイトルから, カリフォルニアやラ

スベガスを観光し、その内容の中で、赤字で示すように“Disneyland”を訪れていることがわかる。“Disneyland”はテーマパークであることから、表 3.1 の例にもあるように「インフラ、ハードツーリズム」と言える。しかし、このブログエントリに含まれる総単語数 909 に対し、“Disneyland”の単語は図 4.1 中にもある 2 回しか出現しておらず、“Disneyland”の内容が記述されている箇所の単語数は 62 と少なかった。また、画像解析の結果からもテーマパークを示唆するような単語が含まれなかった。そこで、Google Cloud Natural Language API による Wikification の結果を見てみると、赤字で示す“Disneyland”やそのアトラクションの一つである“It's_a_Small_World”が含まれることから、根拠となる情報が増えたことで正しく分類できたと考えられる。

- title: Californication, Vegas Baby & the Long Road to Canada
- Text: The next day we fulfilled one of Ursula's childhood dreams and whisked ourselves off to **Disneyland**. We gained another fellow road tripper too: a guy called Alex (yep, another one!) who we met in the LA hostel. **Disneyland** was fabulous.
- 画像解析結果:
black man, night, male, muscle, darkness, human, human body, arm
- Wikification:
Los_Angeles, Las_Vegas, Beverly_Hills, California, Sunset_Boulevard, **Disneyland**, BET_Awards, BET, Hollywood, Carl_Lewis, Vancouver, **It's_a_Small_World**, Las_Vegas_Strip



図 4.2: Wikipedia の情報を加えることで正しく分類できたブログエントリの例

5. おわりに

本研究では、ブログエントリ中のテキストと画像、テキストにリンク付けされた Wikipedia の情報を考慮し、旅行ブログエントリに対して、定義した 9 種類中の 6 種類の観光の形態に自動分類する手法を提案した。画像に対しては、Google Cloud Vision API を用いて、画像中に含まれる物体を検出し、その結果を分類の素性に用いた。また、Wikipedia の情報に対しては、Google Cloud Vision API の解析結果を用いてリンク付けを行い、リンク先の Wikipedia の abstract に含まれる単語集合を分類の素性に用いた。実験の結果、提案手法のテキストと画像を考慮した DAN について、精度ではベースライン手法の画像を考慮した SVM に 0.175 ポイント劣るものの、F 値では最も高い値 0.440 が得られた。また、ベースライン手法、提案手法に用いたそれぞれの DAN に注目すると、提案手法のテキストと画像、テキスト中の単語に関する Wikipedia の情報を考慮した DAN では、テキストのみを考慮した DAN に比べて精度が 0.120 ポイント上昇した。また、画像のみを考慮した DAN と比較すると、精度が 0.095 ポイント上昇した。これらのことから、複数の入力データを考慮した

提案手法の有効性が示されたと考えられる。

6. 今後の課題

本研究では、機械学習を用いて、旅行ブログエントリを対象に観光の形態に基づいた自動分類を行った。分類器には DAN を用い、素性には人手で分類した正解ラベルとテキスト、画像に対しての物体検出の結果、テキストからリンク付けされた Wikipedia の情報を使用した。DAN については、分散表現で表された各単語の平均を文書のベクトルとして分類に用いるという単純な構造であるため、今後語順や各単語の出現回数などを考慮することで更なる精度向上が期待できる。

参考文献

- [1] Rob van der Goot, Nikola Ljubesic, Ian Matroos, Malvina Nissim, and Barbara Plank, “Bleaching Text: Abstract Features for Cross-lingual Gender Prediction,” in Proceedings of the Association for Computational Linguistics(2018).
- [2] Mohit Iyer, Varun Manjunatha, Jordan Boyd-Grader, and Hal Daume, “Deep Unordered Composition Rivals Syntactic Methods for Text Classification,” in Proceedings of the Association for Computational Linguistics (2015).
- [3] 鈴木祥平, 湯舟佑樹, 宮坂涼, 倉田陽平, “宿泊予約サイトにおいて夜間写真の果たす役割とは,” 観光情報学会第 15 回研究発表会講演論文集, 17-20(2017).
- [4] 三條智史, 桂井麻里衣, “画像特徴とテキスト特徴に基づく料理レシピの人気予測,” 第 10 回データ工学と情報マネジメントに関するフォーラム(2018).
- [5] 野沢悠哉, 遠藤雅樹, 江原遥, 廣田雅春, 横山雅平, 石川博, “マイクロブログを用いたユーザの訪問目的と動向の推定,” 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(2016).
- [6] 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, 83-96 (2017).
- [7] 山下晋司, “観光学キーワード,” 有斐閣(2011).
- [8] 井出明, “ダークツーリズム拡張—近代の再構築,” 美術出版社(2018).
- [9] 藤井千賀子, “インフラツーリズムガイド 2018,” 芸文社(2018).
- [10] 岡本亮輔, “聖地巡礼,” 中央公論新社(2015).
- [11] 中島真美, “カルチュラル・ツーリズムによる「住民参加型開発」の有効性,” 第 8 回観光に関する学術研究論文入選論文集, 1-17(2002).
- [12] Davaajav Jargalsaikhan, 岡崎直観, 松田耕史, 乾健太郎, “日本語 Wikification コーパスの構築に向けて,” 言語処理学会第 22 回年次大会, 793-796(2016).
- [13] Rada Mihalcea and Andras Csomai, “Wikify!:

Linking documents to encyclopedic knowledge,” in Proc. of CIKM, 233–242(2007).

- [14] Armand Joulin, Eduard Grave, Piotr Bojanowski, and Tomas Mikolov, “Bag of Tricks for Efficient Text Classification,” arXiv:1607.01759v3 [cs.CL](2016).