

# 旅行ブログ中の画像と本文の対応付け

篠田 広人<sup>†</sup> 難波 英嗣<sup>‡</sup> 竹澤 寿幸<sup>‡</sup>

<sup>†</sup>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 4-5-1

<sup>‡</sup>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 4-5-1

E-mail: {shinoda, nanba, takezawa}@ls.info.hiroshima-cu.ac.jp

**あらまし** 本研究では、旅行者の感情という観点から観光スポットを推薦するシステムの構築を目指している。感情に基づく観光スポット推薦とは、例えば「驚き」の観光名所やレストランなど、感情を軸として旅行者に観光スポットを提示するものである。本研究では旅行ブログに着目し、旅行ブログ中のテキストと画像を対応付けた後、感情推定することで、感情に関する観光スポットを自動的に収集する。

**キーワード** 旅行ブログ, 感情推定, 観光スポット

## 1. はじめに

近年、訪日外国人旅行者は増加の一途をたどっており、来たる 2020 年の東京オリンピックに向けて、少なくとも今後数年はこの傾向が続くものと考えられる。訪日外国人旅行者は、旅先の経験を、ツイッターやブログなどのソーシャルメディア上で写真と共に紹介することも少なくない。こうした情報はリアルタイム性が高く、他の旅行者が旅行計画を立てる上で有用なことが多い。そこで本研究では、ソーシャルメディア、特に旅行ブログを情報源とした観光推薦システムの構築を目指す。

ソーシャルメディア上の情報を観光推薦に利用する研究やサービスは数多く存在するが、本研究では、旅行者の感情という観点から観光スポットを推薦するシステムの構築を目指す。

そこで、感情に基づく観光スポット推薦の例として、驚きや悲しみを軸とした観光スポット推薦例を紹介する。図 1.1 や図 1.2 はそれぞれカナダのトロント周辺の観光スポットであるナイアガラの滝と The Keg というレストランの巨大ステーキである。図 1.1 の斜体部分は「なんて美しい滝なんだ!」という意味になり、驚きを表す観光スポットであるといえる。また、図 1.2 の斜体で表している *amazing* も驚きを表す単語なので、驚きを表す観光スポットであるといえる。一方で、図 1.3 や図 1.4 はそれぞれ広島県広島市周辺の観光スポットである原爆ドームと旧日銀広島支店である。図 1.3 は斜体部分に「恐ろしさ」や「衝撃的な」という単語が含まれており、悲しみを表す観光スポットであるといえる。図 1.4 は斜体部分に「悲惨さ」を表す単語があるので、悲しみを表す観光スポットであるといえる。



*What a beautiful waterfall!*

図 1.1: 驚きを軸とした観光スポット推薦例 1



*It's amazing.*

図 1.2: 驚きを軸とした観光スポット推薦例 2



*A horrifying reminder of a shocking time in history, never to be repeated.*

図 1.3: 悲しみを軸とした観光スポット推薦例 1



*A place that feels the fear of the atomic bomb.*

図 1.4: 悲しみを軸とした観光スポット推薦例 2

感情という観点から実際の旅行ブログを眺めると、旅行ブログ中の画像から、旅行者本人の驚きや喜びの表情を読み取れることがある一方で、ブログ本文中に感情に関する言語表現が使われることもあることが分かる。そこで、まず旅行ブログ中の画像とブログ本文を対応付け、次に画像またはテキスト中の感情を推定することで、感情に関する観光スポットを自動的に収集する。本研究では、これら二つの手順のうち、第一の手順について述べる。

## 2. 関連研究

本研究の関連研究として、2.1 節では、観光情報推薦、2.2 節では、画像・テキストからの情報抽出、2.3 節では、文脈からの情報抽出をそれぞれ紹介する。

### 2.1. 観光情報推薦

田中[1]らは、近年、日本への外国人観光客が増えていること、英語が通じないことや英語で紹介できる人材の不足がみられることから、検索クエリとして観光スポットとその地域を入力として与え、英語を日常的に使用し、外国人の観光に有用な Twitter ユーザを検索するシステムを提案している。外国人観光客に有用な Twitter ユーザの判定のための手がかりの候補の選定において、有用なユーザの判定理由を踏まえて以下の4つの手がかりを設定している。

- ① 写真、チェックイン付きツイートの頻度
- ② 写真、チェックイン付きツイートの語数
- ③ ツイートを行った場所の数
- ④ 日本の滞在期間

上記4つの手がかりすべてを用いたユーザランキングの提案手法と、各手がかり1つずつ除いた4種類のランキング手法とを比較し、手がかりの有効性を検証している。結果から、提案手法と比べて、写真、チェックイン付きツイートの頻度を除いた比較手法1とツイートを行った場所の数を除いた比較手法3が低いことが分かった。よって写真、チェックイン付きツイートの頻度やツイートを行った場所の数は、有用なユーザを判定する手がかりとして有効であると田中らは述べている。本研究と違う点は、Twitter を情報源に使うのではなく、旅行ブログを情報源とする点や外国人の観光に有用な Twitter ユーザを検索するのではなく画像に対する対応文を抽出する点で異なる。

王[2]らは、既存の写真共有サービスなどは旅行先に関する土地勘のないユーザが必要とする情報を得ることが難しいため、多くの人注目した風景の写真を、旅行を計画するユーザに対して分かりやすく表現した地図である観光マップを提案している。観光マップの有用性に関する実験の流れとして、被験者には京都旅行中と想定の中で、写真共有サイトと観光マップを用いて対象地域の風景を調べてもらい、観光マップが持

つ写真の風景カテゴリ分類機能を説明したうえで、その機能の有用性に関する5段階評価のアンケートを行っている。結果として、「1.役に立つ」、「2.どちらかという役に立つ」が全体の76%を占めたことから提案する観光マップが持つ写真の風景カテゴリ分類機能が、旅行先の情報収集に有用であると王らは述べている。本研究とは観光というテーマで共通しているが、王らはアンケートで有用性を図っているのに対し、本研究では精度・再現率で評価を図っている点で異なる。

### 2.2. 画像・テキストからの情報抽出

桂井[3]らは、画像特徴と感情の affective gap を低減させるため、画像・テキスト・感情語の潜在的な関連に基づく画像の感情分類手法を提案している。Web で公開されている画像の感情分類に関するデータセットは数百枚規模であり、疑似的な感情ラベルの画像タグの欠落やノイズがあるので、クラウドソーシングを通じてデータセットを構築している。感情分類に関するデータセットは Flickr と Instagram から収集している。各データセットの画像に対し、1枚の画像につき、3名のユーザからのアノテーションを得ている。各データセットから得られたポジティブ・ネガティブ画像の枚数を見てみると、ソーシャルメディアはポジティブと評価される傾向にある。本研究では、Flickr や Instagram ではなく旅行ブログを情報源とする点で異なる。また、画像から感情を抽出するのではなく、画像の解析結果やキャプションを利用する点で異なる。

倉島[4]らは、人間の経験を「状況」、「行動」、「主観」の3つの要素で成り立っていると定義している。それぞれの要素をさらに細分化し、以下に示す属性で人間の経験を表現している。

$E = \{\text{時間, 空間, 動作, 対象, 評価, 感情}\}$

例えば、“昨日、嵐山で紅葉を見ましたが、とてもきれいで感動しました。”という文章は{時間, 空間, 動作, 対象, 評価, 感情} = {昨日, 嵐山, 見る, 紅葉, きれい, 感動}のように経験情報で表現できる。なお、評価属性を除く、5属性について考えている。上記に示した5属性(時間, 空間, 動作, 対象, 感情)を大規模ブログデータから抽出し、得られた経験情報の中から、データマイニング分野で提案されているルールの興味深さ指標を用いた相関ルール抽出を行う手法、また、感情属性に基づき、経験が動作主にとって成功または失敗なのかを導き出す手法を述べている。特に、感情の抽出においては、感情を示す表現を人手で収集し、喜び、驚き、喜び、驚き、困惑、怒り、悲しみ、疲労、不安、不満の8カテゴリに分類している。本研究では、5属性を大規模ブログデータから抽出するのではなく、ブログの画像の解析結果や画像のキャプションを基に画像とその対応文を抽出している。

### 2.3. 文脈からの情報抽出

狩野[5]らは、どの観光地に需要があるかを分析する上で、Twitter を用いて旅行中ツイートコーパスの作成、および、作成したコーパスを用いて単一ラベルと一連ラベルを推定する機械学習モデルの構築と評価を行っている。単一ラベルはツイート単体で旅行中と判断されるもの、一連ラベルは前後のツイートを考慮して旅行中と判断されるものである。ラベルデータ数から一連ラベルがつけられたツイートは単一ラベルより約3倍多い。このことから従来のように単一ツイートから旅行中ツイートであるかどうかを判断するには、旅行情報に多くの見落としが存在していると狩野らは述べている。Bag of Words によるツイート素性を利用した SVM によって学習モデルを構築し、評価を行った結果、前後のツイートが素性として有益であると述べている。狩野らは、Twitter のツイートをデータの対象としているが、本研究では旅行ブログを対象に LSTM を使って機械学習を行う点で異なる。

石野[6]らは、旅行者が気軽に観光情報を発信する旅行ブログから観光情報を自動抽出する方法を提案している。この手法はまずは旅行ブログの抽出を行い、次に土産物情報・観光情報を抽出するという流れで構成されている。旅行ブログの抽出、土産物情報・観光情報を抽出する方法において共に機械学習の手法として用いるのは CRF である。

土産物情報・観光情報の抽出に関する実験に基づき、石野[6]らは旅行ブログが観光情報の抽出のための有益な情報源であると述べている。石野[6]らは CRF を用いて前後のブログエントリを考慮して旅行ブログの検出を行っているが、本研究では比較手法として CRF を用いる。また、ブログ中の本文だけでなく画像のキャプション、画像の解析結果の情報も考慮している点で異なる。

Gupta[7]らは、論文の要旨やタイトルから3つのカテゴリ (FOCUS, TECHNIQUE, DOMAIN) に該当する語句を自動的に抽出する手法を提案している。Gupta らの手法はパターンマッチに基づいており、例えば、動詞「propose」の直後に出現する直接目的語を「FOCUS」を表す語句として抽出している。本研究においては Gupta のように3つのカテゴリを自動抽出するのではなく、画像解析情報や画像のキャプションを基に画像に対応する文を自動抽出する点で異なる。

Tateisi[8]らは、論文に出現するモノとモノの意味関係を同定するための手法を提案している。まず、エンティティを示す語句に対して「TEAM」、「OBJECT」、「MEASURE」のいずれかのタグを付与し、エンティティ間に「PERFORM (動作主体)」や「CONDITION (実験条件)」など16種類の有向関係を付与したタグ付

けコーパスを用いて SVM による関係抽出器を作成し、その結果に基づいて論文中の文を解析している。本研究においては論文を情報源とした研究ではなく旅行ブログを情報源とした研究である点が異なる。また、本研究では手法として SVM を用いるのではなく LSTM を用いる点でも異なる。

### 3. 画像とブログ本文の対応付け

#### 3.1. システム構成

本研究の目的は旅行ブログをもとに、画像の解析を行い、その結果から、画像とその対応文の抽出を行うことが本研究の目的である。図 3.1 にシステム概要を示す。図 3.1 のシステムにおいては「画像解析」と「キャプション抽出」、そしてそれらと本文のマッチングを行うことで画像と対応文の抽出ができると考えられる。3.2 節で旅行ブログ中の画像解析について、3.3 節で画像と本文の対応付けについて説明する。

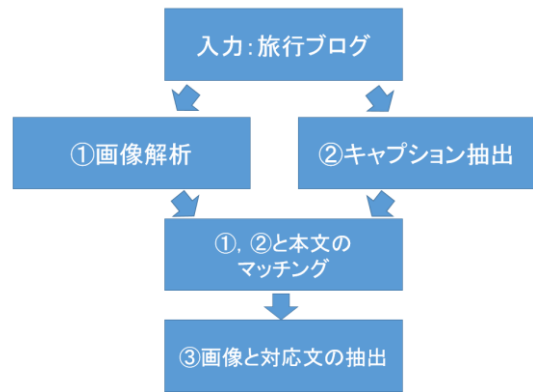


図 3.1: システム概要

#### 3.2. 旅行ブログ中の画像解析

・画像に写っている観光スポットの検出

ここでは、画像に写っている観光スポットの検出について述べる。TravelBlog 中から、観光地などを写した画像に着目し、Google Cloud Vision API を使って解析すると図 3.2 のようになる。図 3.2 から、「amusement park」、「resort」が図 3.2 の画像と一致していると考えられる一方で、「amusement ride」や「games」といったように図 3.2 の画像にそぐわないものもあった。

"description": "amusement park",  
"description": "park",  
"description": "outdoor recreation",  
"description": "recreation",  
"description": "games",  
"description": "amusement ride",  
"description": "resort",  
"description": "inflatable",

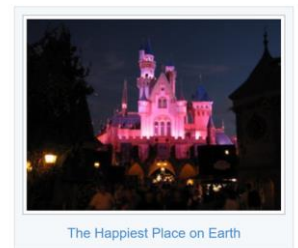


図 3.2: カテゴリ検出の解析結果

・画像中の人物の感情推定

ここでは、画像中の人物の感情推定について述べる。今度は、人物が写った画像に着目し、Google Cloud Vision APIで解析した結果を図3.3に示す。図3.3にあるように、喜びの度合いを表す「joyLikelihood」が「VERY\_LIKELY」、つまり喜んでいる可能性が非常に高いことが分かる一方、悲しみ、怒り、驚きを表す「sorrowLikelihood」、「angerLikelihood」、「surpriseLikelihood」が「VERY\_UNLIKELY」であることから悲しんでいる、怒っている、驚いている可能性は低いことが分かる。

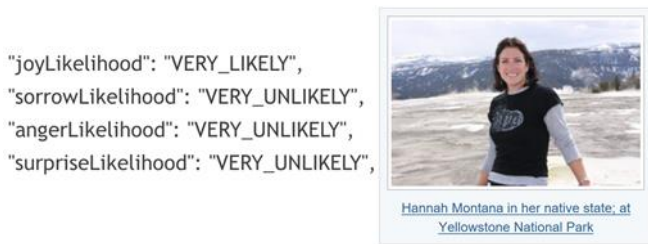


図 3.3: 感情推定の解析結果

3.3. ブログ中の画像と本文の対応付け

3.2 節の Google Cloud Vision API の表情認識とカテゴリ検出の解析結果を用いて提案システムの動作例を説明する。

・画像のカテゴリ検出結果が本文中の単語と一致した場合

まずは画像のカテゴリ検出結果が本文中の単語と一致した場合について説明する。図3.4に解析結果を示す。図3.4から右に画像があり、左にカテゴリ検出の解析結果を示している。図3.4から下の文が解析結果に対応する本文中の文で、解析結果の Moraine lake と本文中の Moraine lake が一致していることが分かる。



And this was the day when we met the one we were always looked for, Sunrise at Moraine lake the bear.

図 3.4: 解析結果（画像のカテゴリ検出結果が本文中の単語と一致した場合）

・画像に対する対応文の自動判定

本研究では、ブログ中の一連の文がある画像に関するものであるかどうかの判定を行った。画像に対応する文には「1」ラベル、それ以外の文は「0」ラベ

ルを付与した。

表3.1にラベルの付与例を示す。また図3.5に画像の表情認識結果を示す。図3.5の右側には枠で囲ったキャプションに Dinner という単語があり、表3.1の一番目の文、二番目の文にも下線で示した dinner という単語がある。よって画像と対応している文と考えられるので、ラベルを1としている。また、表3.1の三番目の文には、図3.5の表情認識の結果、斜体で示すように喜びの度合いが非常に高い。よって、表3.1の三番目の文で驚きや喜びを表す amazing やキャプションの Dinner に関連する単語である food があるので、ラベルを1としている。表3.1の四番目の文は図3.5の画像と関係のない文なのでラベルを0としている。

ブログ中の一連の文がある画像に関するものであるかどうかを判断するタスクを、本研究では系列ラベリング問題として解く。石野[9]らは、ある Twitter ユーザが投稿した一連のツイートから、旅行計画中および旅行中のものを、畳み込みニューラルネットおよび Long short-term memory (LSTM)を用いて抽出する手法を提案している。本研究では、石野らの手法に基づき、画像に対する対応文であるかを LSTM で抽出する。

表 3.1: ラベルの付与例

ブログ本文	ラベル
Today we were real tourists and with video camera in tow went up and had <u>dinner</u> on the Canadian National tower.	1
We got given a voucher for Christmas to have a swish <u>dinner</u> at the 360 degree revolving restaurant.	1
It was <i>amazing</i> both the view and the food!	1
Sunday we will head down with the relies to Niagara Falls.	0

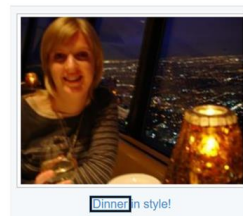


図 3.5: 画像の表情認識結果

4. 実験

本章では、提案手法の有効性を示すために行った実験とその結果、考察について述べる。4.1 節で実験設定について、4.2 節で実験結果について、そして 4.3 節で考察について述べる。

#### 4.1. 実験設定

##### 実験データ

本実験では、旅行ブログサイトの一つである TravelBlog のデータを実験に用いる。TravelBlog は日本だけでなく世界中の人々が観光地の様子や感想を述べている。表 4.1 に TravelBlog のブログエントリ数、画像数、Google Cloud Vision API で解析済みの画像数を示している。また、Google Cloud Vision API で解析した感情推定結果を表 4.2 に示す。表 4.2 に示してあるように喜びの度合いを表す「joyLikelihood」が多くの割合を占めていることが分かった。

表 4.1: TravelBlog ののブログエントリ数, 画像数, Google Cloud Vision API で解析済みの画像数

ブログエントリ数	478,034
画像数	1,224,684
Google Cloud Vision API で解析済みの画像数	933,997

表 4.2: Google Cloud Vision API で解析した感情推定結果

joyLikelihood	60067
sorrowLikelihood	39
angerLikelihood	82
surpriseLikelihood	387

##### 正解データ

正解データは、TravelBlog から、TreeTagger を用いて英文を形態素解析している。ブログのリンク、画像番号、文番号から構成されており、文番号は画像番号に対応する文として表している。表 4.1 にある画像数 1,224,684 画像に対し、任意に選んだ画像数 256 画像を正解データとしており、画像と関連する文を手で対応付けたものを訓練および評価に用いる。その 256 画像に対し、対応文の数は全 896 文あり、1 画像当たり平均 3.5 文対応付けられている。

##### 実験条件

本実験では、5 分割交差検定により評価を行った。ブログ中の各文を、まず 100 次元の実数ベクトルに変換し、次に畳み込み関数を適用して 300 次元のベクトルにする。このベクトルに対し、各次元の最大値を計算した結果として得られる 300 次元のベクトルを LSTM 関数に入力する。LSTM 関数の出力に線形関数を適用し、最も値の大きい次元に対応するラベルを予測ラベルとする。

##### 評価尺度

評価尺度は以下に示す精度と再現率、F 値を用いて実験を行った。

$$\text{精度} = \frac{\text{提案手法によって抽出された正しい対応文の数}}{\text{提案手法によって抽出された対応文の数}}$$

$$\text{再現率} = \frac{\text{提案手法によって抽出された正しい対応文の数}}{\text{すべての対応文の数}}$$

$$\text{F 値} = \frac{2 \cdot \text{精度} \cdot \text{再現率}}{\text{精度} + \text{再現率}}$$

##### 比較手法

本研究では、比較手法として条件付き確率場(CRF)を用いた。

#### 4.2. 実験結果

LSTM と CRF を用いた実験結果を表 4.3 に示す。ベースラインより提案手法の方が F 値で 0.203 ポイント上回ったことにより、有効性が確認できた。

表 4.3: 実験結果

	精度	再現率	F 値
LSTM	0.452	0.229	0.304
CRF (baseline)	0.058	0.385	0.101

#### 4.3. 考察

4.2 節の実験結果について、考察を行う。4.1 節で提案手法である LSTM と比較手法である CRF の 2 種類で比較する実験を行った。表 4.3 にもあるように LSTM の方が、精度が高く、CRF の方が、再現率が高いことが分かる。LSTM にしても CRF にしても再現率が低い原因のひとつに、画像解析の結果得られた言語表現(ラベル)とブログ中の表現が一致しない点が挙げられる。

#### 5. おわりに

本研究では、旅行ブログ中の画像と本文を対応付ける手法を提案した。そこで機械学習の提案手法として LSTM, 比較手法として CRF を用いた。表 4.3 より、CRF より LSTM の方が高い精度を取っていたことが分かるが、十分な結果を得ることができなかった。今後、精度、再現率を向上させるために、さらなる改善が必要である。

#### 6. 今後の課題

本研究ではソーシャルメディア、特に旅行ブログを情報源とした観光推薦システムの構築を目指す中で、まずは画像とブログ本文の対応付けを行った。今後は、例えば WordNet のような同義語辞書を利用し、物体検出の結果のラベルとブログ中の表現の一致率を上げることで、再現率の向上を目指す。また、今回は Google Cloud Vision API の物体検出とカテゴリ検出を用いたが、Google Cloud Vision API の表情解析機能を用いるとさらに精度の向上が期待できると考えられる。

## 参 考 文 献

- [1] 田中 匠, 関 洋平(2016). 外国人観光客の相談相手となりうる Twitter ユーザの検索. 観光と情報, 第 12 卷, 第 1 号, pp.75-84.
- [2] 王佳な, 野田 雅文, 高橋 友和, 出口 大輔, 井手 一郎, 村瀬 洋(2011). Web 上の大量の写真に対する画像分類による観光マップの作成. 情報処理学会論文誌 pp.3588-3592.
- [3] 桂井 麻里衣, 佐藤 真一(2016). 画像・テキスト・感情の潜在的な相関に基づく画像の感情分類. DEIM Forum 2016.
- [4] 倉島 健, 藤村 考, 奥田 英範(2009). 大規模テキストからの経験マイニング. 電子情報通信学会論文誌 D Vol. J92-D, No.3, pp.301-310
- [5] 狩野 竜示, 谷口 元樹, 根本 啓一, 大西 健司, 大熊 智子(2017). 文脈情報を考慮した旅行中ツイートの判別. 言語処理学会 第 23 回年次大会 発表論文集.
- [6] 石野 亜耶, 難波 英嗣, 竹澤 寿幸(2010). 旅行ブログエントリ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(6), pp.667-679.
- [7] Sonal Gupta, Christopher D.Manning (2011). Analyzing the Dynamics of Research by Extracting Key Aspects of Scientific Papers. CoNLL, pp.98-108.
- [8] Yuka Tateisi, Yo Shidahara, Yusuke Miyao, Akiko Aizawa (2014). Annotation of Computer Science Papers for Semantic Relation Extraction. Proceedings of the Ninth International Conference on Language Resources and Evaluation, pp.1423-1429.
- [9] 石野 亜耶, 難波 英嗣, 竹澤 寿幸(2017). Twitter を利用した旅行計画者の行動分析. 観光情報学会 第 16 回研究発表会.