

行動タイプを利用した複数英語旅行ブログエントリの自動要約

光嶋 美宙[†] 難波 英嗣^{††} 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 4-5-1

^{††} 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 4-5-1

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

あらまし 旅行ブログエントリは、旅行者にとって貴重な情報源となる。一方で、旅行ブログエントリを公開する人の数も爆発的に増えており、すべてに目を通すのが非常に困難な状況になりつつある。そこで、本研究では、ある観光地や観光スポットに関する複数の旅行ブログエントリから画像付きのひとつの要約を作成する手法を提案する。提案手法では、ある観光地に関する英語で書かれた複数の旅行ブログエントリと行動タイプを入力とし、該当するエントリ集合の要約を出力する。提案手法の有効性を確認するために行った実験において、提案手法のひとつである LR+IMG 法がベースライン手法を上回ることを確認した。

キーワード 複数テキスト要約、旅行ブログ

1. はじめに

旅行ブログエントリとは、ある観光地を訪れた旅行者が旅先の経験を記録したいわゆる旅行記の一種である。こうした記録は、同じ地を訪れようとする別の旅行者にとって貴重な情報源となる。なぜならば、個々の旅行ブログエントリがそれぞれの旅行者の視点で書かれているため、同じ体験であっても、感じ方や捉え方は千差万別である旅行者の様々な視点を比べながら観光地あるいは観光スポットの情報が得られる点が一般的な観光ガイドブックと異なるからである。一方で、旅行ブログエントリを公開する人の数も爆発的に増えており、ある観光地に関するエントリすべてに目を通すのが非常に困難な状況になりつつある。そこで、本研究では、ある観光地や観光スポットに関する複数の旅行ブログエントリから画像付きのひとつの要約を作成する手法を提案する。

本論文の構成は以下の通りである。2 節では関連研究、3 節では複数英語旅行ブログの自動要約手法について述べる。4 節では実験、5 節で本論文をまとめる。

2. 関連研究

これまでに、情報源に旅行者が旅行中の経験を記録した“旅行ブログエントリ”を活用するための研究が行われている。石野らは、旅行ブログエントリをブログエントリ集合から機械学習を用いることで自動的に検出する手法を提案している[1]。また、Fujii らによって旅行者の目的に合った情報へのアクセスを支援する為、旅行ブログエントリを旅行者の行動タイプ(見る, 買う, 泊まる, 体験する, 食べる)を表 1 のように自動的に分類する手法が提案されている[2]。飯沼らはこれら既存システムの利便性を向上させる為、複数の旅行ブログを要約する研究を行っている[3]。基本と

表 1 旅行ブログエントリのタイプ[2]

観光タイプ	判定基準
見る	観光名所などの見て楽しめる物やイベントについての情報が記載されている
食べる	飲食に関する情報が記載されている
買う	お土産に関する情報が記載されている
体験する	〇〇体験やスキューバダイビングなど、自分の体を使って楽しめる物についての情報が記載されている
泊まる	宿泊に関する情報が記載されている

なる要約手法として LexRank を用いている点や 3 節で述べる旅行者の行動タイプを利用している点では飯沼の研究と共通するが、飯沼らが日本語で書かれたエントリを対象にしているのに対し、本研究では英語で書かれたものを対象にしている点、また、画像認識技術を用いている点が異なる。3 節では、画像認識技術を用いた複数ブログエントリの要約手法を提案する。

2.1. 旅行地の推薦および観光情報の提示

旅行地の推薦はテキスト情報や動画画像等も扱うマルチメディア分野で研究が盛んに行われている。特定の場所を推薦する場合タグ、代表画像、スニペット等「目的地をどのように表現するか」は、最終的に旅行者が旅行先を決定するための重要な手がかりとなる。

Wu らは観光情報を要約するシステムを提案しており、テキストや画像、動画などの異なるメディアを情報源としてクエリのカテゴリごとに選択する手法を提案している[4]。Hao らは、スニペットを要約として出力するモジュールや旅先の特徴を表すタグ等、ブログから地域特有の情報を発見する手法を提案している

[5]. また、ブログを情報源として用いる研究に安田らの研究があげられる。安田らは“食べ物”や“歴史”といった地理的範囲とトピックを入力として受け取り、対象範囲の情報を簡潔にまとめた文書を生成する要約手法を提案しており、ブログを要約対象として実験を行っている[6]。本研究では、旅行地の情報源に旅行ブログエントリを用いており、それらにはテキスト情報だけでなく旅行者が撮影した写真が数多く含まれている。ユーザに視覚的理解を促すのに画像は有用な情報源であるため、本研究では代表画像付きの要約生成を目標とする。

2.2. テキスト要約

テキスト要約とは重要な情報のみをテキスト情報から選択し、要点の迅速な把握を支援する技術である。膨大なテキスト情報が Web 上に蓄積されるようになり、必要な情報に効率的にユーザがアクセスすることを支援する技術が求められている状況で研究が活発になされてきた。

現在も多くの研究で用いられる重要文抽出手法は、重要なセグメント（文や段落）をテキスト中から抜き出し、適切に並べて出力することで要約とする手法である。文の重要度計算には、グラフベースの手法や文が持つ特徴から回帰モデルで重要度を計算する手法等が用いられてきた。

ここで、本研究で提案する手法のベースとなる、グラフベースの重要度計算手法 LexRank を以下に紹介する。LexRank は、テキスト要約に関する評価ワークショップ TAC 等多くのテキスト要約研究において、その有効性が確認されている代表的なテキスト要約手法のひとつである[7]。

LexRank は文のグラフ表現における固有ベクトル中心性の概念に基づき文の重要度を計算する。LexRank では、まず対象テキストに含まれる文間の類似度を計算して類似度グラフを作成する。次に、類似度が閾値以上ならば 1、それ以外は 0 を要素とする隣接行列を用意する。作成したグラフから、ノード（文） u の重要度は式 1 で求められる。これは PageRank と同様に、隣接行列に対してべき乗法を用いて主固有ベクトルを計算することで得られる。

$$p(u) = \frac{1-d}{N} + d \sum_{v \in \text{adj}[u]} \frac{p(v)}{\text{deg}(v)} \quad (1)$$

ここで、 N はノード数（文の数）、 d はダンピングファクタ[8]、 $\text{adj}[u]$ はノード u に隣接するノード集合、 $\text{deg}(v)$ はノード v の次数を表す。計算される重要度は他の多くの文と類似する文ほど高くなり、さらに、重

要度の高い文と類似する文の重要度も高くなる。また、Erkan らは隣接行列の各成分を類似するか否かの離散値ではなく、ノード間の重み（類似度）に設定した Continuous LexRank を提案している。(式 2)

$$p(u) = \frac{1-d}{N} + d \sum_{v \in \text{adj}[u]} \frac{\text{weight}(u,v)}{\sum_{z \in \text{adj}[v]} \text{weight}(z,v)} p(v) \quad (2)$$

飯沼らは、日本語で記述された複数の旅行ブログエントリから要約を作成する手法を提案している。飯沼らは、上述の LexRank を拡張し、ブログエントリのタイプを考慮しながら文と同様の仕組みで画像の重要度を算出することで、代表画像付きの要約を出力している。画像の重要度を計算する際、まず、各画像を色ヒストグラムと Bag of Visual Words でベクトル表現し、次に、画像間の類似度を測ることで、グラフを構築している。ここで、画像の重要度と文の重要度を別々に計算すると、抽出された画像と文の内容が一致しない可能性がある。そこで、飯沼らは文および画像をノードとしてひとつのグラフを作成し、文と画像の重要度を同時に計算している。なお、画像の前後に出現する文は被写体の説明をしている可能性が高いと仮定し、画像とその前後に出現した文の関係を表す隣接行列の成分は 1 に設定している。これにより重要な文に隣接する画像の重要度が高くなり、文、画像ともに重要かつ関連性の高いものが選ばれやすくなると考えられる。

飯沼らの研究では、さらに、行動タイプ情報を利用した LexRank の拡張も行っている。飯沼らのシステムでは、ある地点に関する複数の旅行ブログエントリと行動タイプをシステムの入力とし、ユーザがタイプ“食べる”を指定した場合は、食べ物に関して言及している文の重要度を高く、タイプ“見る”を指定した場合は桜に関する文の重要度を高く設定するような Biased LexRank と呼ばれる手法[9]を用いている。この手法ではまず、文の“タイプらしさ”を算出し、次に、PageRank 値を計算する際、算出したスコアを各ノード（文）へのランダムジャンプ確率に設定する。すなわち、式(2)を次のように修正している。

$$p(u) = (1-d) \frac{\text{typeScore}(u)}{\sum_v \text{typeScore}(v)} + d \sum_{v \in \text{adj}[u]} \frac{\text{weight}(u,v)}{\sum_{z \in \text{adj}[v]} \text{weight}(z,v)} p(v) \quad (3)$$

ここで、 $typeScore(u)$ とは、ノード u のタイプらしさを表すもので、タイプに関する単語の情報利得と頻度から単語レベルで算出し、文に含まれる単語の“タイプらしさ”の最大値を文全体のスコアとしたものである。本研究でも、英語旅行ブログを対象に Biased LexRank を用いた拡張を行う。

3. 複数英語旅行ブログの自動要約

本節では、本研究で実装した自動要約機能について説明する。本研究で扱う旅行ブログエントリーは、Fujiiらの手法によって5種類のタイプ(見る, 買う, 泊まる, 体験する, 食べる)に分類されていると仮定する。システムはある観光地に関する英語で書かれた複数の旅行ブログエントリーと行動タイプを入力とし、該当するエントリー集合の要約を出力する。

3.1. LexRank による文と画像の重要度計算

2.2 節で説明したように LexRank および Continuous LexRank は文の類似度グラフを作成し、ノードの PageRank 値を計算することで文の重要度を計算している。ここで、ブログエントリー中の画像に関しては、Google Cloud Vision API¹を用いて行った画像認識の結果を用いる。このAPIを用いると、例えば図1の画像に対し、図2のような結果が得られる。



図1 Google Cloud Vision API で解析する画像のサンプル
(<https://www.travelblog.org/Europe/United-Kingdom/Scotland/Midlothian/Edinburgh/blog-210916.html>)

図2より、図1の画像は観光スポット(ランドマーク)“Alnwick Castle”に関するもので、緯度 55.415508, 経度 -1.706314 に位置すること、さらに、この画像には、“castle”, “historic site”, “building”, “stately home”というラベルが付与されていることがわかる。これらの観光スポット名やラベルのテキスト情報を画像の代わりに用いることで、画像とテキストの重要度が計算できる。飯沼らは画像周辺の文は画像を説明していると仮定し

<pre> "landmarkAnnotations": [{ "mid": "/m/011xs4", "description": "Alnwick Castle", "score": 0.3697571, "boundingPoly": { "vertices": [{ "x": 31, "y": 79 }, { "x": 149, "y": 79 }, { "x": 149, "y": 110 }, { "x": 31, "y": 110 }] }, "locations": [{ "latLng": { "latitude": 55.415508, "longitude": -1.706314 } }] }] </pre>	<pre> "labelAnnotations": [{ "mid": "/m/0d5gx", "description": "castle", "score": 0.8911375 }, { "mid": "/m/07yr8h", "description": "historic site", "score": 0.8786685 }, { "mid": "/m/0cgh4", "description": "building", "score": 0.8429155 }, { "mid": "/m/0sy84", "description": "stately home", "score": 0.66509914 }], </pre>
-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

図2 Google Cloud Vision API で図1の画像を解析した結果の一部

て文を画像と結びつけることで画像とテキストの重要度を同時に計算する手法を提案したが、本研究では、画像認識の結果として得られた文字列をそのまま要約に用いるため、画像情報とテキスト情報を無理なく統合できるという点が飯沼らの手法と異なる。

3.2. 文間および画像間の類似度

文は tf 値を要素とするベクトルで表し、類似性尺度にはコサイン類似度を用いる。単語の重要度は、 $tf \cdot idf$ を用いるのが一般的であるが、本研究では tf を用いる。その理由は、まず、要約対象となる文書数が多くても10件程度であること、また、要約対象となる文書は、ある地域内のものとなっているため、その地域内のブログに頻出する重要語(例えば、EdinburghにおけるEdinburgh Castle)の重要度が下がってしまう可能性があるためである。

3.3. 行動タイプを用いた LexRank の拡張

旅行ブログには“見る”, “買う”等の旅行者の行動タイプが付与されているが、なかには複数の事柄に関

¹ <https://cloud.google.com/vision/>

して書かれたエントリも存在する。例えば、食べ物に関する記述と観光名所等の見て楽しむものに関する記述が混ざったブログエントリが要約対象であるとする。この時、「食べる」に関する要約を作成する場合には、「食べる」に関連する手がかり語をあらかじめ用意しておき、各文の PageRank 値を計算する際に、これらの手がかり語を含んだ文にランダムジャンプする確率を増やすことで、手がかり語を含んだ文を重要文として選択されやすくすることで、より観光客の目的と合った内容の文を選択することが可能となる。

本研究でも、飯沼らと同様に、式(3)を用いた Biased LexRank のアプローチを用いる。ただし、文の行動タイプらしさを示す $typeScore$ については、処理をより簡単にするため、Fujii らが行動タイプの分類に用いた手がかり語リストを用い、文 u 中にある行動タイプの手がかり語が出現すれば、 $typeScore(u)$ を 1 とする。表 2 に、手がかり語の一部を示す。

表 2 行動タイプを考慮した要約作成に用いた手がかり語の一部

行動タイプ	手がかり語
買う	buy, shop, market, sell, jewellery, shopping, cowboy, night-time, pair, shirt, rock, work, cattle, ad, interpret, charity, land, lane, bad, swear, stall, beach, deli, saunter, thrust, ware, latin, bazaar
見る	museum, history, see, build, park, several, site, temple, shrine, structure, display, visit, emperor, century, building, castle, meiji, bridge, die, war, city, sight, beautiful, place, destroy
経験する	climb, hike, mountain, dive, foot, water, rock, weather, cloud, summit, jump, gear, swim, morning, steep, scary, leg, hot, knee, descent, museum, trail, half, snorkel, sunrise
宿泊する	hotel, room, breakfast, night, ryokan, hostel, go, onsen, relax, dinner, shanty, windows, ceiba, legroom, brit, plumbing, honduras, coronation, taste, guest, bathroom, walk, check, restaurant, speaking
食べる	restaurant, food, dinner, sushi, delicious, grilled, dish, meal, noodle, hotel, eat, taste, drink, menu, street, chef, flavour, serve, sauce, breakfast, salad, lobby, store, mushroom, cook

3.4. 冗長性の削減

LexRank で計算したスコアの高い順に文を抽出すると、冗長性のある要約が作成される可能性がある。Radev らは、文中の情報の包含関係(CSIS) [10]に基づき、文をリランキングすることでこれを解決している。本研究では、文を重要度順に選択する際、要素間の類似度に閾値を設定しておき、類似度が閾値以上の要素がすでに選ばれているときは対象要素を要約に追加しないという処理を最後に行う。

4. 実験

4.1. 実験設定

Fujii らの分類手法で行動タイプを自動付与した旅行ブログエントリを用いる。なお、旅行ブログエントリとして適切であること、また、タイプが適切に付与されていることを人手で確認済みのエントリを用いる。表 3 に示す世界各地 11 地点に関して地点ごとに平均 6 件のエントリを選択し、実験に用いた。

表 3 実験に用いるトピック, 行動タイプ, 旅行ブログエントリ

トピック番号	トピック(地点)	行動タイプ	旅行ブログエントリ数
1	Edinburgh	見る	5
2	Innsbruck	経験する	5
3	Canary Island	経験する	6
4	Barcelona	見る	7
5	Chile	見る	7
6	Hiroshima	見る	5
7	Hiroshima	食べる	3
8	Valencia	食べる	4
9	Venice	経験する	5
10	Nara	見る	5
11	Mt. Fuji	経験する	4

以下の式で定義される評価尺度 ROUGE[11]を用い、提案手法の有効性を検証した。

$$ROUGE(C, R) = \frac{\sum_{e \in n\text{-gram}(C)} \text{Count}_{\text{clip}}(e)}{\sum_{e \in n\text{-gram}(R)} \text{Count}(e)}$$

ここで、 $n\text{-gram}(C)$ は、システム要約に含まれる N グラム、 $n\text{-gram}(R)$ は、参照要約に含まれる N グラム集合を示す。Count(e)は、ある N グラムの出現頻度を数える関数であり、Count_clip(e)は、システム要約に含まれる N グラムのシステム要約における出現頻度 "Count($e \in n\text{-gram}(C)$)" と参照要約における出現頻度 "Count($e \in n\text{-gram}(R)$)" の小さいほうの値を採用する。Lin らは、 N を 1~4 まで変化させ、マニュアル評価結果との相関を調べた結果、 $N=1, 2$ が最も高い相関であったと報告している。今回の実験では $N=1$ を用いる。

比較手法は以下の通りである。いずれの手法でも、抽出する文字数と画像の数は正解と同じになるように調整した。

- LR 法 (baseline) : 文と画像から構築した類似度グラフに対して LexRank を適用する手法。
- cLR 法 (baseline) : 文と画像から構築した類似度グラフに対して cLexRank を適用する手法。

- LR+IMG 法：隣接する文と画像をリンクさせることで、文と画像の類似度グラフを統合したものに LexRank を適用する手法。
- cLR+IMG 法：隣接する文と画像をリンクさせることで、文と画像の類似度グラフを統合したものに cLexRank を適用する手法。
- LR+TYPE 法：タイプ情報を利用した LexRank により、文の重要度を計算する手法。
- cLR+TYPE 法：タイプ情報を利用した cLexRank により、文の重要度を計算する手法。

4.2. 実験結果

実験結果を表 4 に示す。この結果から、行動タイプを考慮するとベースライン手法よりも評価値が下回ったが、画像認識を用いると LexRank を改善できることが確認できた。

表 4 ROUGE-1 を用いた各要約手法の評価結果

手法	ROUGE-1
LR(baseline)	0.043
cLR(baseline)	0.044
LR+TYPE	0.039
cLR+TYPE	0.039
LR+IMG	0.054
cLR+IMG	0.050

4.3. 考察

全体的に、行動タイプを考慮すると、ベースライン手法(LR および cLR)を下回るといった結果になった。原因のひとつは、日本語ブログエントリーと比べ、英語ブログエントリーは手がかり語だけでタイプらしさを捉えるのが難しいということが挙げられる。実際、旅行ブログエントリーの行動タイプの分類では、日本語を対象にした石野らの手法では再現率 0.510、精度 0.659 であるのに対し、ほぼ同じアプローチをとっている英語を対象にした Fujii らの手法では、もっとも分類性能の高い IG+MT 手法でも再現率 0.336、精度 0.597 にとどまっていることにも関係すると思われる。この問題を解決するには、単単語ではなく、フレーズを手がかり語として用いるなどの方法が考えられる。

飯沼らは、予備実験において cLexRank が良い結果を得られたため、cLexRank を実験に用いている。これに対し、本研究では LexRank と cLexRank の両方を用いて拡張を行ったが、全体的に、ほとんど差がないという結果になった。

飯沼らの実験では、ROUGE-1 による評価で 0.3 を越える値が得られていた。これに対し、本研究では一番

良い結果が得られた LR+IMG でも 0.054 程度である。データが異なるので直接的な比較はできないものの、全体的にかなり精度が低い傾向にある。原因はいくつか考えられるが、ひとつは、日本語と英語のブログエントリーの書き方の違いによるものではないかと思われる。日本語ブログエントリーは 1 エントリーの長さが英語と比べると比較的短い。このため、同程度の長さの要約を出力するタスクであっても、英語ブログエントリーを対象にした場合の方が、要約率が小さく、日本語と比べるとタスクとしても難しい。

5. おわりに

本研究では、ある観光地や観光スポットに関する複数の旅行ブログエントリーから画像付きのひとつの要約を作成する手法を提案した。代表的な要約手法のひとつである LexRank および cLexRank を、行動タイプを考慮して拡張した。さらに、Google Cloud Vision API による画像認識技術を用いた手法を提案した。提案手法の有効性を確認するために行った実験において、提案手法のひとつである LR+IMG 法がベースライン手法を上回ることを確認した。

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。

参考文献

- [1] 石野亜耶, 難波英嗣, 竹澤寿幸, “旅行ブログエントリーからの観光情報の自動抽出”, 『知能と情報』, Vol.22, No.6, pp. 667-679, 2010.
- [2] K. Fujii, H. Nanba, T. Takezawa, T. Takezawa, A. Ishino, M. Okumura, and Y. Kurata, “Travellers’ Behaviour Analysis Based on Automatically Identified Attributes from Travel Blog Entries,” Proc. of Workshop of Artificial Intelligence for Tourism, PRICAI 2016, 2016.
- [3] 飯沼俊平, 難波英嗣, 竹澤寿幸, “行動タイプを利用した複数旅行ブログエントリーの自動要約”, 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.
- [4] X. Wu, J. Li, and S.-Y. Neo, “Personalized Multimedia Web Summarizer for Tourist,” Proc. of World Wide Web Conference, 2008.
- [5] Q. Hao, R. Cai, C. Wang, R. Xiao, J.-M. Yang, Y. Pang, and L. Zhang, “Equip Tourists with Knowledge Mined from Travelogues,” Proc. of World Wide Web Conference, 2010.
- [6] 安田宜仁, 西野正彬, 片岡良治, “地理範囲とトピックに応じた動的要約生成,” 第 26 回人工知能学会全国大会, 2012.
- [7] G. Erkan and D. R. Radev, “LexRank: Graph-based Lexical Centrality as Saliency in Text Summarization,” Journal of Artificial Intelligence Research, pp. 457-479, 2004.

- [8] S. Brin and L. Page, "The Anatomy of a Large-scale Hypertextual Web Search Engine," *Computer Networks and ISDN Systems*, pp. 107-117, 1998.
- [9] J. Otterbacher, G. Erkan, D. R. Radev, "Biased LexRank: Passage Retrieval Using Random Walks with Question-based Priors," *Information Processing & Management*, Vol. 45, Issue 1, pp. 42-54, 2009.
- [10] D. R. Radev, H. Jing, and M. Budzikowska, "Centroid-based Summarization of Multiple Documents: Sentence Extraction, Utility-based Evaluation, and User Studies," *Proc. of the NAACL-ANLP Workshop on Automatic summarization*, Vol. 4, pp. 21-30, 2000.
- [11] C. Y. Lin, "ROUGE: a Package for Automatic Evaluation of Summaries," *Proc. of Workshop on Text Summarization Branches Out*, pp. 74-81, 2004.