

場所と旅行者の行動に焦点を当てた 複数旅行ブログエントリの自動要約

飯沼俊平^{†, ‡}, 難波英嗣[†], 竹澤寿幸[†]

[†]広島市立大学大学院情報科学研究科

^{††}現在, リクルートテクノロジーズ

あらまし 情報通信技術の発達により, 旅行に関する情報の収集方法が多様化している. 一般的な旅行ガイドブックと比べると, ブログなどのソーシャルメディアでは実際に観光地を訪れた旅行者の経験に関する情報を得ることができ, それらは訪問場所や宿泊施設を選択する際に大いに役立つ情報である. 本研究では, 画像を含む複数旅行ブログエントリの要約を自動生成する手法を提案する. 提案手法では, グラフベースの文の重要度計算手法である 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する. 提案手法の有効性を検証するために評価実験を行った結果, ベースラインよりも高い性能で要約を作成できることが確認できた. また, 旅行ブログエントリ閲覧システムを構築し, ユーザが指定した地理的範囲およびユーザの目的に合致するエントリの要約を動的に生成・提示する機能を実装した.

キーワード 旅行ブログ, マルチメディア要約, 観光情報処理

Automatic summarization of multiple travel blog entries focusing on travelers' behavior

Shumpei Inuma^{†, ‡}, Hidetsugu Nanba[†], Toshiyuki Takezawa[†]

[†]Graduate School of Information Sciences, Hiroshima City University

^{††}Presently with Recruit Technologies, Co., Ltd.

Abstract Evolution of information and communication technology makes it possible for collecting travel information in various ways. Social media including blogs is one of such useful information source when planning a trip. In this study, we propose a method for generating a summary from multiple travel blog entries containing images. Our method identifies important sentences as well as images using a graph-based approach, which takes account of travelers' behavior types. In order to confirm the effectiveness of our method, we conducted some experiments, and found that our method outperformed some baseline methods. We also implemented a system, which can generate summaries based on our method.

Keywords Travel Blog, Multimedia Summarization, Travel Information Processing

1. はじめに

旅行者が旅先の観光情報を収集するために利用する情報源の一つとして、旅行ガイドブックが挙げられる。一般的な旅行ガイドブックには、有名な観光名所、土産物、宿泊施設、飲食店など、観光に関連する基本的な情報が掲載されている。一方、ブログなどのソーシャルメディアでは、実際に観光地を訪れた旅行者の経験に関する情報を得ることができ、それらは訪問場所や宿泊施設を選択する際に、大いに役立つ情報である。

これまでに、旅行者が旅行中での経験を記した“旅行ブログエントリー”を情報源として活用するための研究が行われてきている。石野らは、機械学習を用いてブログエントリー集合から旅行ブログエントリーを自動検出する手法を提案している[1]。また、旅行者の目的に合致する情報へのアクセスを支援するために、エントリーを旅行者の行動タイプ(買う, 食べる, 体験する, 泊まる, 見る)に自動分類する手法が提

案されている[2]。上述の旅行ブログエントリーの検出・分類技術を利用して収集されたブログエントリーは、実際に広島 P2 ウォーカーで公開されている“ぶらり広島電停散歩 MAP”¹で使用されている。このシステムは地図上に旅行ブログエントリーをマッピングすることで、目的の場所に関する情報を地図上で検索できるようにしたシステムである(図1)。

本研究では、上述の既存システムの利便性を向上させるため、ユーザが指定した地理的範囲²および行動タイプの旅行ブログエントリーを要約して提示する機能を実装する。そのために、テキストだけでなく画像も対象とした要約手法を提案する。旅行者が旅行中での経験を記した旅行ブログエントリーを情報源に用いることにより、市販のガイドブックや公式の観光ポータルサイトとは異なる情報の入手が期待できる。画像付きの要約を提示することで、旅行者は目的地の特徴を容易に把握することができる。また、行動タイプを利用することで旅行者の目的に合った情報を提示することができる。



図1 ぶらり広島電停散歩 MAP

¹ <http://p2walker.jp/peace/ja/blog/>

² ここで述べる「地理的範囲にあるブログエントリー」とは、「ある緯度/経度の範囲内に存在するブログエントリー」を指す。具体的には、2節の図2において、画面内に表示されるブログエントリーを指す。



図 2 旅行ブログエントリ閲覧システム

本論文の構成は以下の通りである。2 節では実装したシステムの概要および動作例，3 節では関連研究，4 節では旅行ブログエントリの自動要約手法，5 節では評価実験について述べ，6 節で本稿をまとめる。

2. 旅行ブログエントリ閲覧システム

本節では，実装したシステムの概要および動作例について説明する。なお，システムで用いている旅行ブログエントリは，前節で言及した石野らの手法で自動検出され，藤井らの手法で自動分類されたものである。図 2 に本研究で開発した旅行ブログエン

トリ閲覧システム³を示す。地図上にエントリ集合を表示しており，画面下のボタンで，“見る”，“体験する”，“買う”，“食べる”，“泊まる”などの，旅行者の目的に沿ったエントリタイプ（表 1）を選択することができる。また，マーカーをクリックするとポップアップでエントリへのリンク付きのタイトルを表示する。ここまでの機能は，既存システム（ぶらり広島電停 MAP）とほぼ同等である。

次に，本研究で構築した自動要約システムを中心に，既存システムからの拡張点について説明する。拡張点は以下の通りである。

- リストビュー：表示範囲のエントリ一覧
- サマリービュー（図 3-5）：表示範囲の指定したタイプのエントリ集合の要約

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

画面左下の緑のボタンをクリックするとリストビューおよびサマリービューが表示され，上部のタブでビューを切り替えることができる。リストビューには画面中の表示されているすべてのエントリ一覧を表示する。既存システムでは地図上のマーカーをクリックすることでエントリのタイトルを表示することができたが，上述のリストビューのように，ある地点に関して書かれたエントリの一覧を表示するビューは実装されていなかった。

³ <http://www.ls.info.hiroshima-cu.ac.jp/blogMap/>



図3 広島県 宮島付近, タイプ“見る”の要約例



図4 広島県 宮島付近, タイプ“食べる”の要約例



図5 東京都 浅草周辺, タイプ“見る”の要約例

サマリービューでは、ユーザが指定した地理範囲、タイプのエントリ集合をトピックごとに自動要約した結果と、関連するエントリへのリンクを表示する。図3, 4は広島県の宮島付近でタイプ“見る”, “食べる”に該当するエントリ集合をそれぞれ自動要約した例である。同じ地点でも, “見る”を選択したときは「厳島神社」や「水中花火大会」, “食べる”を選択したときは「牡蠣祭り」のように, 選択したタイプに沿った要約を自動生成することができる。同様に, 図5は東京の浅草付近でタイプ“見る”を選択して要約した結果である。このように, 任意の場所に関しても要約を自動生成することができる。

既存システムでは, 目的の地域の情報を閲覧するために, 1. 地図上のマーカーをクリック, 2. ブログリンクをクリック, という操作を興味あるエントリの数だけ行う必要があった。サマリービューを導入することで, ユーザは画面遷移することなくエントリ集合の要約を閲覧できるので, 指定した地域の見所が発見しやすくなり, さらに, 個々のエントリ

へのリンクを示すことで, より詳細な情報へのアクセスが容易になる。また, エントリのタイプを利用することで, より旅行者の目的に合った要約を自動生成・提示することができる。

3. 関連研究

3.1 旅行地の推薦および観光情報の提示

旅行地の推薦は, テキスト情報に加えて動画像なども扱うマルチメディア分野で盛んに研究が行われている。特定の場所を推薦する際に, 「目的の場所をどう表現するか」(タグ, 代表画像, スニペットなど)は, 最終的に旅行者が旅行先を決定する際に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, クエリのカテゴリごとに, 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選択する手法を提案している[3]。Hao らは, 旅先の特徴を表すタグやスニペットを要約として出力するモジュール

ルなど、地域特有の情報をブログから発見する手法を提案している[4]。また、ブログを情報源に用いる研究として安田らの研究があげられる。安田らは、“歴史”や“食べ物”といったトピックと地理的範囲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作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[5]。本研究では、旅行地の情報源として旅行ブログエントリを対象としているが、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。画像はユーザに視覚的理解を促す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。

3.2 テキスト要約

テキスト要約はテキスト情報から重要な情報のみを選択し、要点の迅速な把握を支援する技術である。Web上に膨大なテキスト情報が蓄積されるようになり、ユーザが必要な情報に効率的にアクセスするのを支援する技術が求められている状況で活発に研究がなされてきた。

現在でも多くの研究で用いられている重要文抽出手法は、テキスト中から重要なセグメント（文や段落）を抜き出し、抜き出したセグメントを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、文が持つ特徴から回帰モデルで重要度を計算する手法やグラフベースの手法などが用いられてきた。

ここで、グラフベースの重要度計算手法であり、本研究で提案する手法のベースとなる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TAC⁴をはじめとする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[6]。

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いて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、ノードを文、類似度が閾値以上の文間をエッジとすることで、類似度グラフを作成する（図6）。

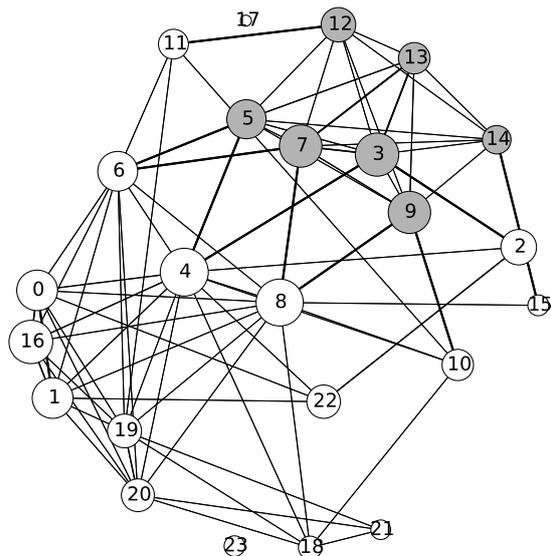


図6 類似度グラフの例

次に、類似度が閾値以上であれば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、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 はノードの数（文の数）、 d はダンピングファクタ[7]、 $\text{adj}[u]$ 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、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)において、 $\text{weight}(u,v)$ がノード間の重みを示している。

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

テキストと画像をリンクさせることができれば、テキストと同様の枠組みで画像の重要度を計算できると考え、本研究ではグラフベースの手法を採用した。前田らの研究[8]ではLexRankを画像の類似性グラフに適用していたが、本研究では画像と文を一つのグラフで扱う点

⁴ <http://www.nist.gov/tac/>

が異なる。また、我々の手法では LexRank を拡張した biased LexRank [9]のアプローチを用いることでエントリのタイプに焦点を当てた要約を実現する。

近年では、テキスト要約を組み合わせ最適化問題として定式化することが多くなっている[10]。例えば、指定した要約長を超えないように、重要度の和を最大化して冗長性を最小化するなど、同時に様々な制約を設定できるという特徴がある。本研究では画像も要約の対象としているが、画像から得られる特徴量はテキストから得られる特徴量と性質が異なるため、これらの手法を画像に適用することは困難である。このような理由から、我々はグラフベースの手法を採用した。

4. 旅行ブログエントリの自動要約

本研究で扱う旅行ブログエントリは、藤井らの手法により 5 種類のタイプに分類されていると仮定する。システムは、エントリタイプと地理的範囲を入力として、該当するエントリ集合の要約を出力する。なお、閲覧システムの表示領域を考慮して、1 トピックにつき画像 3 枚、3 文から 5 文程度の短い要約を目標とする。要約は大まかに次の手順で作成する。

1. エントリ集合をクラスタリング
2. クラスタごとに文と画像の重要度を計算
3. 重要度が高い順に文と画像を選択

エントリにはすでに“見る”や“食べる”といったタイプが付与されているが、ユーザが指定した範囲には、さらに複数のトピックが混在する可能性があるため、これらをグループ化する必要がある。本研究では、最遠隣法を用いた階層的クラスタリングを行い、クラスタ間の距離が閾値以下の時にクラスタを統合する方法をとる。なお、エントリは $tf*idf$ 値を要素とする文書ベクトルとして扱い、距離関数は $f(c_i, c_j) = 1 - \cos(c_i, c_j)$ を用いる。ここで、 $\cos(c_i, c_j)$ はコサイン類似度を表す。

4.1 LexRank による文と画像の重要度計算

3.2 節で説明したように LexRank (Continuous LexRank) は文の類似度グラフを用意して、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。同様に、エントリ集合に含まれる画像間の類似度を計算し、隣接行列を用意すれば、多くのエントリに出現する物体が写った代表画像を得ることができると考えられる。本研究では、文および画像をノードとしてグラフを作成し、文と画像の重要度を同時に計算する。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 に設定する。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た。まとめると、隣接行列の各成分は次のように決定する。

$$a_{i,j} = \begin{cases} \text{sim}(s_i, s_j) & (\text{type}(s_i) = \text{type}(s_j)) \\ 1 & (\text{type}(s_i) \neq \text{type}(s_j) \text{ and } |i - j| = 1) \\ 0 & (\text{otherwise}) \end{cases} \quad (3)$$

ここで、 s はエントリを構成する要素(文または画像)のシーケンスを表し、 s_i はその i 番目⁵の要素を示す。 $\text{type}(s_i)$ は要素 s_i の型(文または画像)、 sim は類似度関数を表している。なお、類似度関数は要素の型ごとに用意する。要素間の類似度計算に関しては次節で説明する。図 6 は実際に存在するエントリをもとに作成した類似度グラフである。白いノードは文、灰色のノードは画像を表している。また、ノードの大きさは 4.2 節で説明する式(4)を用いて計算した重要度を表している。

4.2 文間および画像間の類似度

文は、文中の単語(名詞、動詞、形容詞)の $tf*idf$ 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。画像は 2 種類のベクトルで表現し、それぞれのコサイン類似度を計算し、その平均値を画像間の類似度とする。利用するベクトル表現は、色ヒストグラムと Bag of Visual Words ベクトルである。

色ヒストグラムは、HSV 色空間を用いて H, S, V の値域をそれぞれ 10, 4, 4 分割することで画像中の各ピクセルを 160 色に減色させ、色ごとにピクセルの出現頻度を求めることでヒストグラムを計算する。Bag of Visual Words は画像から得られる複数の局所特徴をベクトル量子化してヒストグラムを化したものである[11]⁶。本研究では、まず、画像集合から SIFT 特徴量を、スケールを固定して格子状に抽出し、得られた特徴量を k-means 法によりクラスタリングする。次に、個々の画像から抽出した SIFT 特徴量を、クラスタリングにより得たセントロイドを用いてベ

⁵ ここで述べる i 番目とは、HTML より、テキストを文分割し画像タグも分割した断片における i 番目のことを指す。

⁶ Bag of Visual Words に関する詳細な説明は、文献[12]などを参考にされたい。

クトル量子化ヒストグラムを作成する。

4.3 タイプ情報を利用した LexRank の拡張

旅行ブログエントリには“見る”や“食べる”など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、桜などの見て楽しむ物に関する記述が混在するエントリが要約対象であるとする。この時、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したい。このように、ユーザの指定したタイプと文の内容の関連度を考慮するために、biased LexRank のアプローチを導入する。ここではまず、文の“タイプらしさ”を算出する。そして、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)を次のように修正する。

$$p(u) = (1 - d) \frac{\text{typeScore}(u)}{\sum_v \text{typeScore}(v)} + d \sum_{v \in \text{adj}[u]} \frac{\text{weight}(u, v)}{\sum_{z \in \text{adj}[v]} \text{weight}(z, v)} p(v) \quad (4)$$

本研究では、上述の“タイプらしさ”(typeScore(u))を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typeScore(u)とする。単語のタイプらしさは次のように算出する。

$$\text{typeScore}(w) = IG(w, t) \log \left(1 + \frac{\text{count}(w, D_t)}{\text{count}(w, D)} \right) \quad (5)$$

ここで、 $IG(w, t)$ は「単語 w が含まれているか否か」という情報が、エントリをタイプ t とそれ以外のタイプに分類する時のエントロピーをどのくらい減少させるかを表している⁷。ある単語がエントロピーを大幅に減少させるということは、それだけその単語はエントリをタイプ t に分類するかどうかに関与する有効な単語であると考えられる。 $\text{count}(w, D)$ はエントリ集合 D （要約対象のエントリではなく、コーパスに含まれるすべてのエントリ集合）のうち、単語 w を含むエントリ数を表す。また、 D_t はタイプ t に該

当するエントリ集合である。

式(5)の \log の底は e とする。式(5)において、 $\text{count}(w, D_t) \leq \text{count}(w, D)$ が常に成り立つので、 $\log(1 + \text{count}(w, D_t)/\text{count}(w, D))$ の値域は、 $\log(1)$ 以上 $\log(2)$ 以下、すなわち 0 以上 $\log_e 2 = 0.693$ 以下となる。 w は type t にのみ頻出する場合、 $\text{count}(w, D_t)/\text{count}(w, D)$ の値が 1 に近くなり、逆に w があらゆるタイプのエントリに出現する場合は、 $\text{count}(w, D_t)$ よりも $\text{count}(w, D)$ の値が大きくなり、 $\text{count}(w, D_t)/\text{count}(w, D)$ の値が 0 に近くなる。なお、 $\log(1 + \text{count}(w, D_t)/\text{count}(w, D))$ 単語 w のタイプらしさを表す上での係数であり、単語ごとの係数の大小関係のみが重要であるため、上述の上限値が 0.693 であることそのものについては特に意味はない。

4.4 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) [13]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(文または画像)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。

5. 評価実験

提案手法の有効性を調べるため、2種類の実験を行った。実験1では、要約そのものの品質を評価した。実験2では、旅行者が旅先の情報をブログ集合から得るといった状況において、ある範囲内にあるブログエントリを検索し、そのすべてを読むという状況と比べ、検索提案手法で作成した要約がどの程度有効であるか評価した。実験1については5.1節、実験2については5.2節でそれぞれ述べる。

5.1 要約の品質の評価(実験1)

5.1.1 実験設定

石野らの手法で収集し、藤井らの分類手法で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。日本全国20地点に関して地点ごとに平均10件のエントリを選択し、タイプごとに画像3枚、5文程度の正解要約を合計47件人手で作成し、ROUGE [15]による評価と被験者2人による主観評価を行い、提案手法の有効性を検証した。

比較手法は以下の通りである。いずれの手法でも、抽

⁷情報利得に関するより詳細な説明は、例えば、書籍[13]の4.6.2節を参照されたい。

出す文字数と画像の数は正解と同じになるように調整した。

- **Lead 法 (baseline)**: 要約対象のすべてのエントリの先頭から文と画像を抜き出す手法。
- **LR 法 (baseline)**: 文と画像, それぞれの類似度グラフに対して **LexRank** を適用する手法. (式 2)
- **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. (式 2, 式 3)
- **LR+TYPE 法**: タイプ情報を利用した **LexRank** (式 4) により, 文の重要度を計算する手法。
- **LR+IMG+TYPE 法**: 上述の **LR+IMG 法** で作成した文と画像の類似度グラフに対し, **LR+TYPE 法** を適用した手法。

LexRank の計算は, Python で自作したコードを用いた。なお, 各種パラメータは次のように設定した。エントリを階層的クラスタリングする際のクラスタの距離の閾値を 0.9 (4 節), **Bag of Visual Words** ベクトルを計算する際は **SIFT** 特徴量のスケールを 16 ピクセルに固定して 8 ピクセルごとにサンプリングを行い, **k-means** 法はクラスタ数を 1,000 に設定した (4.2 節)。 **SIFT** 特徴量の抽出には, **OpenCV** (<http://opencv.jp/>) を用いた。 **LexRank** を計算する際のダンピングファクタは 0.85 に設定し (式 2), 冗長性を削減するためのコサイン類似度の閾値は文の場合 0.9, 画像の場合 0.5 と設定した (4.4 節)。 4.3 節で説明した単語のタイプらしさは, 日本全国の要約対象外のエントリも含めた 1,836 件の旅行ブログエントリ (買う: 147 件, 見る: 1,145 件, 体験する: 119 件, 泊まる: 38 件, 食べる: 693 件) を用いて算出したものを用いた。

システムの出力した要約は, テキスト部分は **ROUGE** による自動評価, 画像に関しては精度・再現率を用いて評価する。正解画像は, 要約作成者にブログエントリ集合から代表画像を選択してもらい, それらに評価に用いる。

ROUGE-N は正解要約とシステムの出力の両方に共通して含まれる **N-gram**⁸ の数を, 正解要約中の **N-gram** の数で割った数値 (再現率) である。すなわち, 正解に含まれる情報をどれだけ網羅できているかを自

⁸ 本実験では, 単語単位で, また単語はすべて原形にして **ROUGE-N** の計算を行った。なお, ここで述べる **N-gram** とは, 単語の連鎖確率ではなく, 単語の連鎖そのものを指す。

動評価するための指標である。 **ROUGE** は, テキスト要約に関する評価ワークショップ **TAC** でも採用されているテキスト要約分野の代表的な自動評価指標のひとつである。なお, 今回は **N=1 (ROUGE-1)** および **N=2 (ROUGE-2)** で評価を行った。

ROUGE に加えて, 以下の観点から人手による主観評価も行った。

1. 要約元のブログエントリの内容が適切にまとめられているか?
2. 要約に含まれている画像とテキストの内容は違っていないか?

人手による評価では, 原文を読んだ上で要約を読み, 評価するように被験者に指示した。観点 1 は, 人手による正解要約を比較手法に加えた 6 種類の要約を読み比べ, 1~6 位の順位を付けた。この評価方法は **NTCIR-3** のテキスト要約タスクでも用いられている [16]。なお, 要約がほぼ同じ場合は同一順位にすることも許可している。観点 2 に関しては, 要約結果を 5 点満点で評価した。

5.1.2 実験結果

表 2 に **ROUGE** による評価結果を示す。 **t** 検定を行ったところ, **LR+TYPE 法** と **LR+IMG+TYPE 法** はベースライン手法のひとつである **LR 法** と比べ, 有意水準 5% で有意に差があることが確認できた。なお, **ROUGE-1** および **ROUGE-2** による判定結果の **p** 値を, 表 3 および表 4 にそれぞれ示す。⁹ また, 表 5 に画像の選択精度・再現率 (全タイプの平均) を示す。顕著な改善はみられないが, **LR+IMG+TYPE 法** が最も有効であることが確認できた。

表 6, 7 に人手による主観評価の結果を示す。表 6 は複数の手法で生成した要約を被験者 2 人に良い順に並べてもらった時の平均順位を表している。表 7 はテキストと画像の内容が合っているかを 5 点満点で評価した結果である。どちらの観点でも, 文と画像を一つのグラフで扱った **LR+IMG 法** が最も良い評価値を得ていることが分かる。また, **t** 検定を行ったところ, **Lead 法** と提案手法 (**LR+IMG 法**, **LR+TYPE 法**, **LR+IMG+TYPE 法**) の評価値に有意水準 5% で有意差を確認することができた。

⁹ 検定には Python のモジュール

“[scipy.stats.ttest_rel](https://docs.scipy.org/doc/scipy-0.14.0/reference/generated/scipy.stats.ttest_rel.html)” を用い, 対応のある **t** 検定を行った。

https://docs.scipy.org/doc/scipy-0.14.0/reference/generated/scipy.stats.ttest_rel.html

表 2 ROUGE による評価結果

	ROUGE-1	ROUGE-2
Lead 法 (baseline)	0.318	0.222
LR 法 (baseline)	0.316	0.207
LR+IMG 法	0.331	0.227
LR+TYPE 法	0.345	0.240
LR+IMG+TYPE 法	0.340	0.237

表 5 代表画像の選択精度および再現率

	精度	再現率
Lead 法 (baseline)	0.351	0.359
LR 法 (baseline)	0.341	0.341
LR+IMG 法	0.338	0.342
LR+TYPE 法	0.351	0.359
LR+IMG+TYPE 法	0.372	0.367

表 3 ROUGE-1 による評価における p 値

		Baseline	
		Lead	LR
提案手法	LR+IMG	0.3993	0.1380
	LR+TYPE	0.0791	0.0105
	LR+IMG+TYPE	0.1325	0.0110

(背景がグレーのものは有意差なし)

表 6 要約の内容に関する主観評価

	順位の平均値
正解要約	1.28
Lead 法 (baseline)	4.01
LR 法 (baseline)	3.09
LR+IMG 法	2.85
LR+TYPE 法	3.22
LR+IMG+TYPE 法	2.99

表 4 ROUGE-2 による評価における p 値

		Baseline	
		Lead	LR
提案手法	LR+IMG	0.7964	0.0643
	LR+TYPE	0.2454	0.0039
	LR+IMG+TYPE	0.7685	0.0381

(背景がグレーのものは有意差なし)

表 7 テキストと画像の内容の一致度に関する主観評価

	評価値の平均値
正解要約	4.33
Lead 法 (baseline)	2.80
LR 法 (baseline)	3.09
LR+IMG 法	3.12
LR+TYPE 法	2.96
LR+IMG+TYPE 法	3.05

次に、要約のタイプ別に LexRank (LR)と提案手法を比較する。表 8 は人手による内容評価(順位付け)結果をもとに、LR 法と提案手法の優劣関係をタイプ別にカウントした結果である。同様に、人手によるテキストと画像の内容一致に関する評価結果をもとに、手法の優劣関係を表 11 にまとめた。なお、表の中で背景が灰色の部分には提案手法が優勢の箇所である。表 8 から、画像と文の類似度グラフを結合した手法(LR+IMG)がどのタイプでも LR よりも優勢であることが分かる。しかし、LR+TYPE 法はどのタイプに関しても LR と比べて劣勢である。テキストと画像の内容一致に関する評価(表 9)では、タイプ“見る”に関しては LR+TYPE 法を含めたすべての提案手法が優勢であるが、“食べる”に関してはどの手法も LR 法と同等または劣っていることが分かる。同様に、要約対象エントリに含まれるテキストの量と画像の量のバランスと、要約性能の関係を調査した。表 10 は、エントリに含まれる“画像あたりの文字数”別に

LR 法と提案手法の優劣関係をカウントしたものである。表 10 より、100 文字未満の場合は LR+IMG 法のみが優勢であるが、文字数が 100 文字以上の場合であれば LR+IMG 法と LR+IMG+TYPE 法が優勢であることが分かる。

表 8 LexRank と提案手法のタイプ別の比較 (人手による内容評価)

手法の比較	見る	食べる	買, 体験, 泊	全タイプ
LR 法 < LR+IMG 法	12	12	4	28
LR 法 > LR+IMG 法	7	6	3	16
LR 法 = LR+IMG 法	23	12	15	50
LR 法 < LR+TYPE 法	7	9	2	18
LR 法 > LR+TYPE 法	14	9	6	29
LR 法 = LR+TYPE 法	21	12	14	47
LR 法 < LR+IMG+TYPE 法	13	15	7	35
LR 法 > LR+IMG+TYPE 法	13	11	6	30
LR 法 = LR+IMG+TYPE 法	16	4	9	29

表 9 LexRank と提案手法のタイプ別の比較 (人手によるテキストと画像の内容一致度評価)

手法の比較	見る	食べる	買, 体験, 泊	全タイプ
LR 法 < LR+IMG 法	8	10	4	22
LR 法 > LR+IMG 法	5	10	3	18
LR 法 = LR+IMG 法	29	10	15	54
LR 法 < LR+TYPE 法	6	6	2	14
LR 法 > LR+TYPE 法	4	11	7	22
LR 法 = LR+TYPE 法	32	13	13	58
LR 法 < LR+IMG+TYPE 法	12	9	6	27
LR 法 > LR+IMG+TYPE 法	9	12	6	27
LR 法 = LR+IMG+TYPE 法	21	9	10	40

5.1.3 考察

実験の結果, 対象とする要約タイプやエントリに含まれる画像あたりの文字数によって最適な手法が異なることが分かった. 例えば, 文と画像の類似度グラフを結合する手法(LR+IMG)はどのタイプでも LR 法より性能が良いため, 最適な手法であると言える. 文字数別に比較すると, 画像あたりの文字数が 100 文字以上であれば, LR+IMG 法よりも LR+IMG+TYPE 法の方が優れている事例が多いため, タイプ情報を用いた後の方が適していると言える. 画像あたりのテキスト量が多い時に LR+IMG+TYPE 法の性能が良いのは, 提案手法ではテキストのタイプ情報を利用しているためであると考えられる.

次に, 実際に閲覧システム(図 2-5)を使用して明らかになった提案手法の問題点について述べる. まず, 本研究で用いている旅行ブログエントリには複数のタイプが付与されたエントリも多く存在する. 特に, タイプごとの情報量(文, 画像の数)の偏りが大きいと, タイプに沿った要約がうまくできない傾向にある. たとえば, 宮島では牡蠣祭り(タイプ“食べる”)が毎年行われているが, 参加者の中には, 同時に厳島神社(タイプ“見

る”)や, 宮島に到着するまでにフェリーに乗ったことなども記述する人が多い. 本研究では, この問題を解決するために LR+TYPE 法および LR+IMG+TYPE 法を提案したが, LexRank の特性上, 多くのエントリで言及される事柄の重要度が高くなりやすいため, “食べる”の要約に, 船や風景画像が混入することがある. 実験では, PageRank 値を計算するダンピングファクタは 0.85 に固定したが, これをエントリに含まれるタイプ情報のばらつきなどによって調整することで, この問題は解決できると考えている.

旅行ブログエントリには観光情報として有用な情報以外にも, 個人的な事柄に関する記述も多く含まれるため, それらを要約から除外することも重要な課題である. エントリ数が多いクラスタに関しては, 内容の共通部分を重要箇所として検出できるが, エントリが少ないクラスタに関してはその検出が難しくなるため, 観光には関係のない情報が要約に含まれる可能性も高くなる. LR+TYPE 法では, タイプに関する単語の情報利得値を用いて“タイプらしさ”を表現した. 同様に, 旅行ブログエントリと一般的

表 10 “画像あたりの文字数”別 LexRank と提案手法の比較

手法の比較	100 文字未満	100 文字以上
LR 法 < LR+IMG 法	18	10
LR 法 > LR+IMG 法	10	6
LR 法 = LR+IMG 法	30	20
LR 法 < LR+TYPE 法	11	7
LR 法 > LR+TYPE 法	20	9
LR 法 = LR+TYPE 法	27	20
LR 法 < LR+IMG+TYPE 法	21	14
LR 法 > LR+IMG+TYPE 法	22	8
LR 法 = LR+IMG+TYPE 法	15	14

なブログエントリから、単語や表現の“旅行ブログらしさ”を算出し、重要度計算の際に利用すれば、上述のようなノイズを軽減できると考えられる。

5.2 観光情報の収集課題における要約の有用性の評価(実験 2)

5.2.1 実験設定

表 11 に示す 5 地点において、表中のそれぞれの観光タイプに関する情報をブログ集合から得る、という状況を想定し、実際に情報を得るのに、どのくらい時間がかかったのか、また、得られた情報の質について、被験者による実験を行った。

実験では、被験者に以下の 3 種類のブログエントリ集合を与え、それぞれをすべて読むのにかかった時間を算出した。

- (集合 1)ある緯度/経度の範囲にあるすべてのブログエントリ
- (集合 2)集合 1 のうち、特定のタイプのもの
- (集合 3)集合 2 から提案手法により自動作成された要約

さらに、提案手法により自動作成された要約の質についても調査した。実験 2 では、実験 1 のような正解データは作成していないが、被験者は各課題の地域のブログすべてを読むことになるため、正解データがなくても自動作成された要約の品質については判断できると考え、ひとつの課題が終わるごとに被験者にインタビューを行った。実験では、2 名の被験者が 1 課題につき 1 名ずつ評価した。

表 11 観光情報の収集課題における要約の有用性の評価(実験 2)に用いた課題

場所	観光タイプ	全ブログエントリ数	当該観光タイプのブログエントリ数
厳島神社	見る	63	50
柴又	買う	26	8
上野	見る	12	4
浅草	見る	102	77
高野山	見る	11	10
広島市(繁華街)	食べる	37	36

5.2.2 実験結果

実験結果を表 12 に示す。まず、全ブログエントリを読むよりも、特定のタイプのエントリだけ読む方がかかる時間は短い。ただし、観光タイプだけでどの程度まで目的のブログを絞りこめるのかは地域による。例えば、高野山では寺社を見るのが観光の主目的であり、表 11 からわかるとおり、この地域で「見る」以外のタイプのブログはほとんど存在しない。このような場所では、観光タイプはほとんど絞り込みの役には立たない。一方で、上野や柴又のように観光スポットと飲食店が混在している地域には、複数のタイプのブログエントリが存在するので、観光タイプによる絞り込みが有効に機能する。

表 12 観光情報の収集課題における要約の有用性の評価(実験 2)に用いた課題の実験結果(単位: 分)

場所	観光タイプ	全ブログエントリを読むのにかかった時間	当該観光タイプのブログエントリを読むのにかかった時間	要約を読むのにかかった時間
厳島神社	見る	40	30	2
柴又	買う	12	6	1
上野	見る	14	3	1
浅草	見る	143	110	2
高野山	見る	27	26	1
広島	食べる	36	36	1

提案手法で生成された要約については、被験者はすべての課題について1~2分で読むことができた。集合1~3を読む速度は、被験者が集合1~3をどの順番で読むのかにも影響されると思われるが、その可能性を考慮しても、読む速度という点において要約がある程度効率的であることがわかる。

次に被験者へのインタビューによる要約の品質に関して報告する。インタビューの結果、どの課題についても、概ね妥当な要約が作成されているということが分かった。被験者の報告によると、類似した内容のブログエントリを何度も読まなければならない、その作業に時間がかかるなどのことであった。一方で、提案手法は4.4節でも説明したとおり、冗長性を削減する処理を行っているため、同じ内容を何度も読まなければならないという状況にはならない。

実際に作成された要約には、市販のガイドブックや公式の観光ポータルサイトとは異なる情報が含まれていた。例えば、巖島神社の要約では、満潮時と干潮時の両方の大鳥居の写真が要約に含まれていた。満潮時の大鳥居の写真は有名であるが、干潮時には鳥居の下まで歩いて行くことができる、ということが要約から分かる。また、毎年夏に開催される宮島水中花火大会に関するブログエントリが要約の中で大きな割合をしめていた。これは、宮島水中花火大会に関して記述されたブログエントリが非常に多かったためである。各ブログエントリには執筆日時が記載されているので、例えば、ユーザが夏に本システムを利用する時には、夏に開催されるイベントを重点的に要約に含めるといったユーザのコンテキストに応じた要約作成も、将来的には可能になると思われる。このような柔軟性が、おそらく市販のガイドブックなどとの一番大きな違いになると思われる。

6. おわりに

本研究では、旅行ブログエントリ閲覧システムを構築し、ユーザが指定した地理的範囲、タイプのエントリを自動要約して提示する機能を実装した。グラフベースの文の重要度計算手法であるLexRankを拡張し、文と画像に適用することで画像付きの要約を作成する手法を提案した。また、タイプ情報を利用することで旅行者の目的に合った要約手法を提案し、その有効性を実験により示した。

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] 藤井一輝, 石野亜耶, 藤原泰士, 前田剛, 難波英嗣, 竹澤寿幸, “多言語旅行ブログエントリを用いた観光情報提示システム,” 第6回データ工学と情報マネジメントに関するフォーラム, 2014.
- [3] X. Wu, J. Li, and S.-Y. Neo, “Personalized Multimedia Web Summarizer for Tourist,” Proc. of World Wide Web Conference, 2008.
- [4] 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.
- [5] 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.
- [6] G. Erkan and D. R. Radev, “LexRank: Graph-based Lexical Centrality as Salience in Text Summarization,” Journal of Artificial Intelligence Research, pp. 457-479, 2004.
- [7] S. Brin and L. Page, “The Anatomy of a Large-scale Hypertextual Web Search Engine,” Computer Networks and ISDN Systems, pp. 107-117, 1998.
- [8] 前田剛, 難波英嗣, 竹澤寿幸, “場所に焦点を当てた複数旅行ブログの自動要約,” 第7回データ工学と情報マネジメントに関するフォーラム, 2015.
- [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] H. Takamura and M. Okumura, “Text Summarization Model Based on Maximum Coverage Problem and its Variant,” Proc. of the 12th Conference of the European Chapter of the Association for Computational Linguistics (EACL '09), pp. 781-789, 2009.
- [11] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray, “Visual Categorization with Bags of Keypoints,” Proc. of ECCV International Workshop on Statistical Learning in Computer Vision, pp. 1-22, 2004.
- [12] 柳井啓司, “物体認識技術の進歩,” 日本ロボット学会誌, Vol. 28, No. 8, pp. 257-260, 2010.
- [13] 高村大也, “言語処理のための機械学習入門,” コロナ社, 2010.
- [14] 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.
- [15] C. Y. Lin, “ROUGE: a Package for Automatic Evaluation of Summaries,” Proc. of Workshop on Text Summarization Branches Out, pp. 74-81, 2004.
- [16] T. Fukushima, M. Okumura, and H. Nanba, “Text Summarization Challenge 2 / Text Summarization Evaluation at NTCIR Workshop 3,” in Working Notes of the 3rd NTCIR Workshop Meeting, PART V, pp. 1-7, 2002.

謝辞

本研究は、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。



飯沼 俊平 (非会員)

1990年生. 2014年 広島市立大学情報科学部知能工学科卒業.

2016年 広島市立大学大学院情報科学研究科知能工学専攻 博士前期課程修了. 2016年 株式会社リクルートホールディングスに入社, リクルートテクノロジー

ーズに配属. 現在に至る. 検索基盤ユニットで, 検索 UX 向上のための分析, 検索ロジック開発, アーキテクチャ設計に従事.



難波 英嗣 (正会員)

1972年生. 1996年東京理科大学理工学部電気工学科卒業. 2001年北陸先端科学技術大学院大学

情報科学研究科博士後期課程修了. 博士(情報科学). 2001年日本学術振興会特別研究員. 2002年

東京工業大学精密工学研究所助手. 同年広島市立大学情報科学部講師. 2010年広島市立大学大学院情報科学研究科准教授. 現在に至る. テキストマイニング, 情報検索, テキスト要約に関する研究に従事. 観光情報学会, 情報処理学会, 人工知能学会, 言語処理学会会員.



竹澤 寿幸 (正会員)

1961年生. 1984年早稲田大学理工学部電気工学科卒業. 1989年

早稲田大学大学院理工学研究科博士後期課程修了. 工学博士.

1987年早稲田大学情報科学研究教育センター助手. 1989年(株)

ATR 自動翻訳電話研究所研究員. 音声対話翻訳の研究開発に従事. 2007年より広島市立大学大学院情報科学研究科教授. 現在に至る. 音声対話や観光情報学の研究と教育に従事. 観光情報学会, 電子情報通信学会, 情報処理学会, 人工知能学会, 日本音響学会, 言語処理学会会員.