

観光サイトにおける閲覧目的に基づいた 旅行記概要の動的生成

三笠 弘貴¹ 奥野 拓²

概要：観光情報ポータルサイトや個人観光ブログには、各記事ごとに概要説明文章が存在する。記事を端的にまとめたものが概要文章であり、サイト利用者が記事を開覧するかどうかの判断材料となる。しかし、旅行記を扱うサイトにおいて記事投稿者の一連の旅行体験が一まとめに記載されることにより、一つの記事内に複数の観光トピックが含まれる場合がある。そのような場合において、記事中の観光トピック(名所、食事、宿泊施設、交通手段、等)について網羅的かつ簡潔に要約を行うことは困難である。本稿では旅行記を対象とし、観光トピックごとに文章分類を行い、動的に概要文章を生成する手法を提案する。手法の有効性評価として、従来の概要文章と動的に生成した概要文章が記載されたポータルサイトの比較を行った。評価の結果から、被験者の必要としなかったページの閲覧時間が削減され、従来の概要文章と比較しユーザの探していた情報が記載されるページを見つけられることが明らかとなった。

キーワード：文章要約, 文章分類, 概要文章, スニペット, ポータルサイト, 観光トピック

Summarizing Articles on the Sightseeing Web Sites Based on Users' Purpose of Browsing

Abstract: The personal travel blogs and tourist information portal sites which deal with tourism articles have a summary which tells its main points shortly. Users look for the articles with its elements such as a summary, title or thumbnails from large amount of articles during their information retrieval. However, since the travel experience of the series is recorded by the author, there are various tourism topics in a single article. In this case, it is difficult to summarize these various topics of the travel in an article. In this study, we propose the method which generate the summaries dynamically depending on the purpose of the site visitors on tourist information portal sites. In the experiment, examinees collected tourism information using two portal site which has summaries generated dynamically and conventional summaries. The result suggests that the examinees can find the articles which they are looking for with the portal site using proposal method by shortening the time of browsing the articles which they did not refer to.

Keywords: text summarization, text clustering, summary, snippet, portal site, topics of tourism

1. 記事要約技術による情報検索支援

1.1 背景

近年、インターネットの普及に伴い Web 上に膨大なテキスト情報が蓄積され、電子化された文書の流通がますます拡大している。そのような背景のなかで、蓄積された情報へアクセスするための入り口として情報ポータルサイト

の重要性が高まっている。サイト閲覧者が求める情報を効率的に収集することを支援するために様々な技術が生み出されている。中でも、文章中から重要な情報のみを拾い上げ要点を絞り込む自動要約技術の研究が盛んに行われている。

1.2 観光情報提供サービスにおける情報収集と活用の事例

情報ポータルサイトのようなサイト入り口となるサービスには多くの情報が蓄積されている。特に観光情報を扱うポータルサイトとして、国内の観光スポットを取り

¹ 公立はこだて未来大学大学院
Graduate School of Future University Hakodate
² 公立はこだて未来大学
Future University Hakodate

扱う「じゃらん」(<http://www.jalan.net>)や「るるぶ.com」(<http://www.rurubu.com>), 宿泊施設の情報を中心に扱う「楽天トラベル」(<http://travel.rakuten.co.jp>)など様々なサービスが提供されている。中でも旅行記を扱うブログポータルサイトには、投稿者の一連の観光体験が綴られた記事が大量に蓄積されている。奥村 [1] は、それらの記事には投稿者の意見やレビューなどのリアルタイム性の高い貴重な情報「生の声」が記載されており、評判分析やトレンド分析のようなブログマイニング技術がより活発に研究されるだろうと述べている。個人単位で観光ブログを活用する場面も多く、観光ブログポータルサイト上の記事の見出しや概要文章、写真(サムネイル)などの要素を手がかりに膨大な情報から必要な情報の取捨選択を行うこととなる。

1.3 旅行記の要約における課題と研究目的

旅行記には、観光に関する様々なトピック(観光名所、食事、宿泊施設、交通手段、等)が記載される。閲覧者は記事全体を見ずとも、端的に本文を要約した文章「概要文章」を読むことで、旅行記にどのような事が書かれているのか判断することができる。

一方、投稿者が概要文章として記事に含まれる多くの情報をまとめるには、書き手の文章技術が要されるため。また、ユーザの閲覧目的ごとに求める観光トピックは異なるため、あらかじめ用意された静的な概要文章を提示する場合、すべての閲覧者の要求を同時に満たすことは難しい。特に観光ブログの記事において、記事の投稿者は一連の旅行体験をひとまとまりの記事として投稿する 경우가多く、記事中に様々な観光に関するトピックが含まれる傾向にある。本研究では観光ブログを対象とし、閲覧目的に応じた動的な概要文章生成により情報発見を支援することを目指す。

2. 観光情報ポータルサイトにおける記事要約手法

2.1 ポータルサイトにおける情報検索と各要素が与える影響

一般的なサイトの記事一覧ページ内には、各記事の概要文章やタイトル、サムネイル画像などが用意されており、それらの要素を頼りに情報検索を行う。図1は観光ブログを扱う旅行クチコミサイト「フォートラベル」(<http://4travel.jp>)の旅行記ページにおける概要文章と詳細文章の一例である。

このようなページ上の要素の役割について、Yesiladaら [2] や渡辺ら [3] はアイトラッキング技術を用いて分析している。記事タイトルが閲覧者の興味を引く一方で、スニペット(概要文章)は最終的に詳細文章を閲覧するかどうかを判断する役割を果たすと述べている。



図1 観光情報ポータルにおける概要文章と詳細文章の一例
Fig. 1 An Example of the Summary and Main Text in the Travel Articles

2.2 概要文章の利用目的と要件

概要文章を記述する際、詳細文章に含まれるトピックを網羅的に解説することと文章量の抑制はトレードオフの関係にある。概要文章は簡潔にまとめる必要がある以上、本文中のトピックすべてを含む概要文章を作成することは望ましいと言えない。奥村 [4] は、概要文章を利用目的ごとに「指示的」と「報知的」の2つに分類している。指示的な概要文章は、原文を読む前にそれを読むべき物かどうか個人の関心に合うかどうかを判断する目的で用いる要約の形式である。そのため、本文中の情報が網羅的に要約されているか、文章として読みやすく書かれているか、といったことは必ずしも必要ではない。原文がどのようなものなのか適切に判断できさえすれば十分だとしている。一方、報知的な概要文章は原文の代わりとして用いることを目的とした文章要約の方法である。そのため、指示的要約に比べ文章としての可読性や情報の網羅性が要求される。個別のサイトを原文とすると、情報ポータルサイトにおける各サイトの概要文章を指示的なものにするだけで、サイト閲覧するかどうかの判断材料になると考えられる。

2.3 重要文抽出手法を用いた概要文章の作成

概要文章の作成方法の中で最も一般的なものとして重要文抽出手法が挙げられる。重要文抽出手法とは、原文から重要な文を抽出しテキスト中での出現順に結合して概要文章とする手法である。概要文章は、一定の文章量に到達するまで重要度順に結合される。Paice [5] や奥村ら [6] は文の重要度の算出に、テキスト中の単語や単語間のつながり、文のタイトルなどの情報を用いる方法を提案している。重要文抽出を行った後に、各文を文節単位に解析し不要箇所を削除や簡潔な言い回しに置き換えたりする重要箇所抽出を併用する場合もある。本研究においても、文中の観光に関する単語を手がかりに文自体の重要度を算出する。

3. 提案手法

3.1 提案手法の概要

前章で述べたように、概要文章の作成において簡潔さを

維持したまま閲覧者の多様な閲覧目的を満たす概要文章を作成することは困難である。また観光ブログにおいては、記事中に一連の旅行体験がまとめて記録される傾向にあり、様々な観光に関するトピックが散在する傾向にある。そこで本研究では、観光ブログポータルサイトにおいて、閲覧者ごとの興味に応じて記事から動的に概要文章を生成することで、課題の解決を目指す。提案手法は次の3つの段階からなる。

- (1) 旅行記からの観光トピック抽出
- (2) 閲覧目的の指定
- (3) 閲覧目的に合わせ概要文章を生成

旅行記からの観光トピック抽出では、閲覧中の観光ブログサイトに対し、記事内の各文を観光トピックに分類する。閲覧目的の指定では、閲覧者自身がどのようなカテゴリに属する情報を知りたいか、予め用意された観光トピック一覧から選択する。閲覧目的に合わせ概要文章を生成するフェーズでは、指定された閲覧目的に基づいて、観光トピック抽出フェーズで分類された文の中から文を抜粋し提示する。

3.2 観光トピックの定義とトピックに基づいた文章分類

観光ブログサイトや地方の観光情報サイトにおける記事のカテゴリを参考にし、旅行記に含まれると考えられる要素を調査した。次に示す5つの観光トピックとした。

- 食事
- 買い物・お土産
- 観光スポット
- 歴史
- 宿泊・交通

次の2つの文は、旅行クチコミサイト フォートラベルに登録されている旅行記から抜粋したものである。

- ホテルをチェックアウトし、函館駅前から市電で末広町へ。
- 函館山ロープウェイに乗りたいたいということだったので、行ってきました。

1文目には、筆者がホテルや市電を利用していることが記載されており、宿泊・交通トピックに属する文であると解釈することができる。2文目には函館山ロープウェイを利用したこと書かれているが、函館山ロープウェイが観光スポットであるか交通手段であるか判断することは難しい。このように2つ程度のトピックを含む文も存在するため、一意に分類出来ないことを考慮する必要がある。

本手法ではナイーブベイズ分類器を用いることで観光ブログにおける文を5つのトピックに分類する。ナイーブベイズ分類器とはベイズの定理を用いた教師あり型の分類器である。事後確率の上位2つまでを考慮することで、一文を最大で2つのカテゴリに分類することとした。

3.3 概要文章の個人適応

ユーザの検索目的から検索結果ページにおけるスニペット(概要文章)を提示する研究として、高見ら[7]の研究がある。閲覧者が入力した検索語の重要度を変化させることにより、各検索目的に合わせたスニペットを動的に生成する手法を提案している。検索サービスを提供しているGoogle(<https://www.google.co.jp>)では、検索結果としてスニペット(サイト内テキストの抜粋)を出力する。このスニペットはユーザが与えるクエリ(検索ワード)により、抜粋箇所を変化させ概要文章として提示する方法を用いている。これらの事例では文章と概要文章を一对一に限定せず、読み手の要求や状況に合わせて概要文章を動的に生成することで閲覧者に最適化した情報提示を目指している。本研究においても、閲覧者ごとの閲覧目的に応じて概要文章を動的に変化させることで、よりの確な検索結果を提示できると考えた。図2は、旅行記からの観光トピック抽出

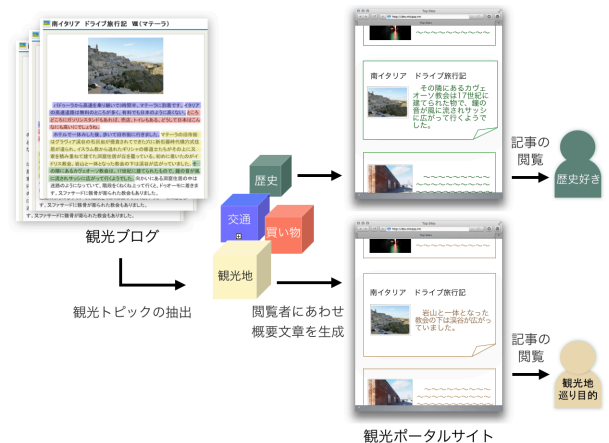


図2 観光トピックに基づく文章分類と閲覧者の目的に応じた概要文章の動的生成

Fig. 2 The Flow of Generating Summaries Based on the Purpose of Viewers and Text Clustering Based on Tourism Topics

によって分類された文章を用い、ユーザの閲覧目的に合わせて概要文章を生成する流れを示している。あらかじめ観光ブログ内の文をトピックごとに分類し、ユーザの閲覧目的に応じて概要文章を変化させる。

4. 観光ブログポータルサイトにおける記事のトピック潜在率の調査

4.1 調査概要と目的

観光ブログの記事において、記事の投稿者は一連の旅行体験をひとまとまりの記事として投稿する 경우가多く、記事中に様々な観光に関するトピックが散在する傾向にある。記事中の観光情報を網羅的かつ詳細に記述することは難しいため、概要文章中に出現しないトピック、つまり詳細文章中に潜在してしまうトピックが多く存在すると考え

られる。

本章では観光ブログポータルサイト フォートラベルを対象とし、記事本文に潜在するトピックを調査し分析を行った。それにより、的確に本文を要約している概要文章がどの程度存在するのか検証した。

4.1.1 トピック潜在率の定量化

概要文章と詳細文章におけるトピック数の比較を行うため、概要文章に含まれる観光トピック数と詳細文章に含まれる観光トピック数をそれぞれ t_s と t_d とし、トピック潜在率 T_p を次式のように定義した。

$$T_p = 1 - \frac{t_s}{t_d} \quad (1)$$

図3は旅行記の概要文章に含まれるトピック数と本文に含まれるトピック数の例である。概要文章に含まれるトピック数 t_s は「食事」と「観光スポット」で2となり、詳細文章に含まれるトピック数 t_d は「食事」「買い物・お土産」「観光スポット」「交通・宿泊」「歴史」で5となる。定義した潜在率の式にあてはめると、例として挙げた記事におけるトピックの潜在率は0.6となる。

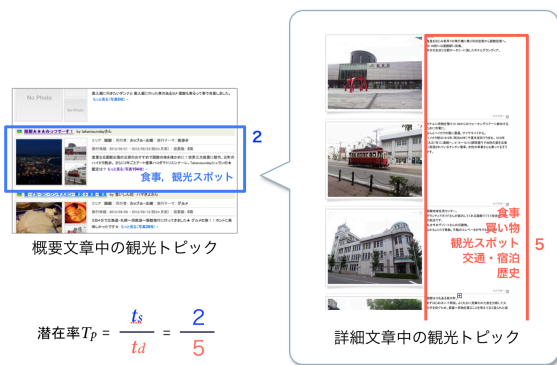


図3 概要文章と本文に含まれるトピック数 t_s と t_d の例

Fig. 3 A Example of The Number of Topics in the Main Text and Summary

4.1.2 分析対象とした旅行記について

調査対象として観光ブログポータルサイトに登録されている函館観光に関する記事を用いた。記事の中には、概要文章に旅行のあらすじを記し詳細文章には写真のみを記載するアルバムスタイルの記事が存在した。本調査ではそれらを除外した記事959件(約4万5000文)について、ナイーブベイズ分類器を利用し分類結果の分析を行った。アルバムスタイルの記事を除外するにあたり、詳細文章の文字数が概要文章の文字数よりも少ない記事をアルバムスタイルの記事と判断している。

4.1.3 函館観光に関するトピックを含む文章分類のための教師データの作成

記事中の文の分類については、ナイーブベイズ分類器を利用している。図4は函館市公式観光情報サイト「はこぶら」(<http://www.hakobura.jp>)に登録されている観光ス



図4 観光スポットページからカテゴリ情報と名詞句の抽出により作成された教師データ

Fig. 4 Training Data that is Created by Extracting Noun Phrase and Category from the Article of a Tourist Spots

ポットのページから教師データを作成する例である。「はこぶら」に登録されているスポットのメタデータにはジャンル情報(食べる・見る・遊ぶ・温泉・買う、など)が存在する。スポットごとのジャンル情報を利用し、定義した「食事」「買い物・お土産」「観光スポット」「歴史」「宿泊・交通」の5種類の観光トピックごとの教師データを作成した。

また歴史カテゴリについては、函館市が編纂を行う市史の電子版「函館市史デジタル版 通説編」(<http://www.city.hakodate.hokkaido.jp/soumu/hensan/hakodateshishi/>)を教師データとして利用した。

また、前処理として文章中の名詞以外の句が分類機を学習させる際のノイズになってしまうため形態素解析器 MeCab を利用し、前述の教師データに含まれる名詞句以外を除外した。形態素解析用の辞書として IPA 辞書に加え、函館に関する旅行記に対して正確に形態素解析を行うため、函館市観光情報サイトはこぶらの約800件の記事タイトルから、観光スポット名・店舗名、約800件を抽出し、解析用辞書とした。

表1 教師データとして利用した文章
Table 1 Texts which was Used as Training Data

カテゴリ	文章数	名詞句数
食事	5376	6077
買い物・お土産	3728	4462
観光スポット	3462	3952
歴史	6350	8516
宿泊・交通	2243	2785

表1は、教師データとして用いた文章数と、文章中に出現した名詞句の総数である。

4.2 調査結果

フォートラベルの函館旅行に関する記事959件について、ナイーブベイズ分類機を用いて分類した結果について述べる。表2は、概要文章と詳細文章のそれぞれにおける観光トピックごとの平均記載率を示したものである。概要文章では約2種類のトピックについて書かれているが、詳細文

章には約4種類のトピックを含む文章が記載されていることを示している。この結果から、フォートラベルにおける函館観光に関する記事の概要文章では、記事の詳細文章に含まれるトピックの5割しか説明できない事がわかる。

表2 概要文章と詳細文章における平均トピック数

Table 2 The Average Number of Topics in the Main Text and Summary

	概要文章	詳細文章
平均トピック数	1.86	3.72

表3は、旅行記の概要文章と詳細文章それぞれ存在する観光トピックの割合について示したものである。詳細文章における記載率については、交通に関しては概要文中でも触れられていることが多いが、歴史については、概要文章で触れられていない場合が多いことがわかる。また、買い物・お土産や宿泊・交通トピックに比べ、食事や歴史トピックの潜在率が高くなっていることがわかる。

表3 フォートラベルの記事における観光トピック記載率

Table 3 The Recording Rate of Tourism Topics on articles of the Fortravel

	食事	買い物・お土産	観光スポット	歴史	宿泊・交通
概要文章	7.7%	70.0%	27.7%	5.2%	75.4%
詳細文章	66.2%	88.0%	82.4%	37.3%	97.5%
トピックの潜在率 T_p	88.3%	20.5%	66.3%	86.0%	22.7%

図5は、対象とした959件の旅行記について、トピックの潜在率 T_p ごとの記事数を示したものである。トピックの潜在率 T_p が0%である記事が131件あることに對し、トピックの潜在率 T_p が50%である記事は177件、トピックの潜在率 T_p が80%である記事は210件あるという結果となった。

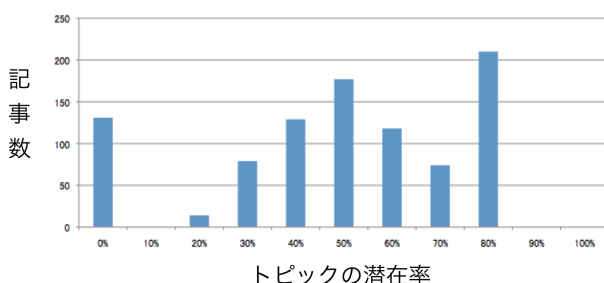


図5 観光トピックの潜在率

Fig. 5 The Potential Rate of Tourism Topics on Articles

表4は調査対象とした959件の記事における、観光トピックの潜在率 T_p が0%である記事と、 T_p が0%を越え

る記事の割合を表したものである。表より、約86%の記事については潜在率 T_p が0%を越える値であることがわかる。

表4 潜在率ごとの記事の割合

Table 4 The Percentage of Articles in Each Potential Rate

	潜在率ごとの記事の割合
潜在率 T_p が0%の記事	14.06%
潜在率 T_p が0%を越える記事	85.94%

4.3 考察

前節の調査結果から、全記事中の約86%の記事については、詳細文章に含まれるトピックより概要文章に含まれるトピックが少ないため、概要文章から詳細文章中のすべてのトピックを推察することは難しいと考えられる。これらの結果から、フォートラベルのような観光ブログポータルサイトにおける旅行記の概要文章から、詳細文章に存在するトピックを知ることは難しく、詳細文章を閲覧するかどうかの判断指標として十分なものであるとは言えない。

また、食事トピックや歴史トピックの潜在率 T_p が他のトピックに比べ高くなっているという調査結果から、それらの観光トピックについて知りたい記事閲覧者に対して、詳細文章に潜在しているトピックを取り出し要約文章を生成することで、記事閲覧の判断を下す際の有効な判断材料になると考えられる。

5. 提案手法の評価

5.1 評価の概要

閲覧者の閲覧目的に合わせ概要文章を変化させることで被験者の情報収集にどのような影響を与えるか調査を行った。被験者には函館観光を計画している人になりきってもらうために、事前に課題を与えてから観光に関する情報の収集を行ってもらった。情報収集に用いるツールは、本評価のために開発したブログポータルサイトのみとした。ポータルサイトの詳細については次節で詳細に述べる。試行開始から終わりまでの画面上の操作を記録し、評価終了後に閲覧行動の分析を行った。

5.1.1 評価課題とアンケート

情報収集の課題は、行きたいと思った場所や食べ物、利用したいと思った施設などの名前挙げ、興味を持ちきっかけとなった旅行記(または記事中の一文)を回答するものである。「興味を持った物」「参考にした旅行記の名前」「参考にした理由」の3つを對にして回答して貰った。また課題に取り組んだ後、ポータルサイト内の概要文章についてのアンケートに回答してもらった。アンケートは以下に記載した6問からなっており、5段階の尺度で回答する。

Q1 提示された概要文章が記事本文を閲覧するかどうか

の判断材料となったか。

- Q2** 提示された概要文章を読むことで記事本文の内容を推察(予想)することができたか。
- Q3** 提示された読むことで期待通りの記事(興味のある情報)に出会えたか。
- Q4** 提示された概要文章中に自身の興味のある情報が記載されていたか。
- Q5** 提示された概要文章中に興味を惹く情報が記載されていたか。
- Q6** 提示された概要文章から記事投稿者の意見や感想が読み取れたか。

5.1.2 評価手順

評価は函館に4年以上在住する4名の被験者に対し行った。評価の際、被験者が利用したブログポータルサイトはパターンAとパターンB、パターンCとパターンDの4種類である。それぞれのサイトには、事前に観光ブログサイトフォートラベルからスクレイピングした約700件の函館観光に関する記事が登録されている。表5は、各パターンのサイトに登録された記事と機能を表したものである。

表 5 評価用ポータルサイトに登録された記事と利用可能な機能
Table 5 The Functions of the Portal Sites for the Evaluation

ポータルサイトのパターン	登録された記事	機能
パターン A	2010 年から 2014 年の函館旅行記	記事のキーワード検索
パターン B	2009 年以前の函館旅行記	記事のキーワード検索 概要文章の動的生成
パターン C	2010 年から 2014 年の函館旅行記	記事のキーワード検索 概要文章の動的生成
パターン D	2009 年以前の函館旅行記	記事のキーワード検索

ブログポータルサイトパターンBとパターンCについては、ユーザが選択したパターンの概要文章を生成し表示する機能(概要文章の動的生成機能)が備わっている。生成可能な概要文章は「食事・買い物」「観光体験」「交通・宿泊」の3パターンである。本文中の文を各パターンに分類し、一定の文字量に達するまで抜粋した文を連結している。連結された文と文の間には「◆」の記号が表記される。

なお、評価順序や記事内容による影響を考慮し、被験者2名に対しブログポータルサイトパターンAとパターンB、残りの2名に対してブログポータルサイトパターンCとパターンDを利用してもらった。

ブログポータルサイトを利用して観光情報収集した後、課題Aと課題B、アンケートの順で回答してもらった。ブログポータルサイトパターンAとパターンBを用いた被験者2名には次のような手順で実験を行った。

- (1) ブログポータルサイト A のみを 10 分間利用する。
- (2) 課題 A に回答する。(3分程度)
- (3) ブログポータルサイト B のみを 10 分間利用する。
- (4) 課題 B に回答する。(3分程度)
- (5) アンケートに回答する。(10分程度)

またブログポータルサイトパターンCとパターンDを用いた被験者2名にも同様の手順で、アンケート1-Cと1-D、アンケート2に回答してもらった。

5.1.3 評価用ブログポータルサイトの機能説明



図 6 評価用観光ブログポータルサイトパターン A

Fig. 6 Tourism Blog Portal Site Pattern A Used for the Evaluation



図 7 評価用観光ブログポータルサイトパターン B

Fig. 7 Tourism Blog Portal Site Pattern B Used for the Evaluation

図6は評価実験に用いたブログポータルサイトパターンAである。被験者は、検索フォームに任意のキーワードを入力することで旅行記を絞り込むことができる。また、各記事の詳細が知りたい場合「もっと見る」リンクをクリックすることで、記事本文を閲覧することができる。ブログポータルサイトパターンDも同様の機能を備えている。

図7はブログポータルサイトパターンBである。ブログポータルサイトパターンBには、ブログポータルサイトパターンAの機能に加えて、サイト右上部のプルダウンメニューから記事の概要文章のパターンを指定し切り替えることができるようになっている。パターンCのポータルサ

イトも同様の機能を備えている。

表 6 は、観光ブログ「トワイライトエクスプレスで冬の北海道へ4」

(<http://4travel.jp/travelogue/10844999>) における各パターンの概要文章をあげた表である。表内の「原文」は記事投稿者が記載した概要文章を示しており、ポータルサイトパターン A とパターン D 利用時に提示される文章である。「食事・買い物」パターンの文章では、タパスや白豚、生ハム、もも肉のステーキなどの単語が出現しており、食事に関する文章が抽出されていることがわかる。

表 6 観光ブログにおける各概要文章のパターンの一例

Table 6 An Example of Patterns of Summary in the Tourism Blog

要約のパターン	概要文章
原文	この日の夜は吹雪いていましたが、吹雪の中の夜景は幻想的でした。
食事・買い物	まずはタパスの9種盛りをオーダーしました ◆ 道産の白豚で作られた生ハム ◆ もも肉のステーキ
体験観光	函館に着いた時には雪のなかった函館の街も、一晩で真っ白になっていました ◆ 今回は、下りが大幅に遅延して、トワイライトが函館特発になる事も想定して、最終日を函館泊にしました ◆ やはり函館は、夜が魅力的な街です ◆ この日の夜は吹雪いていましたが、吹雪の中の夜景は幻想的でした ◆ 波止場に泊まってあった船も、ライトアップされとても綺麗でした ◆ ホテルに行く時には駅から歩いて行きましたが、この日は雪でスーツケースをひいて歩く事が出来ないくらい雪が積もっていたので、タクシーで駅に向かいました
交通・宿泊	JR 北海道のホームページで下りのトワイライトが定刻で札幌に向かっているのを確認後、帰りのトワイライトに始発の札幌から乗りたかったのでスーパー北斗で札幌に戻ります ◆ 行きが臨時北斗だったので、帰りのスーパー北斗は、とても早く感じました ◆ 札幌到着後、お決まりの立ち食いうどんで昼食を済ませ、上りのトワイライトで帰路に就きます ◆ チェックイン後、夕食場所まで夜景を楽しみながら散策しました ◆ 館内は、レトロで洒落た作りでした ◆ 赤レンガにあるクリスマスツリー

5.2 結果

フォートラベルの函館旅行に関する記事 959 件について、ナイーブベイズ分類機を用いて分類した結果について述べる。記事投稿者が記載した概要文章を用いたポータルサイト（ポータルサイトパターン A とパターン D）と動的に生成された概要文章を用いたポータルサイト（パターン B とパターン C）比較する。図 8 は被験者 4 名のアンケート結果の平均をとったグラフである。ブログポータルサイ

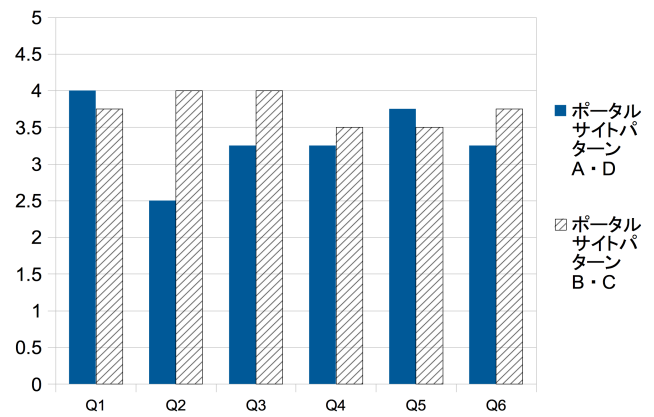


図 8 被験者 4 名のアンケート結果

Fig. 8 The Survey Results of Four Examinees

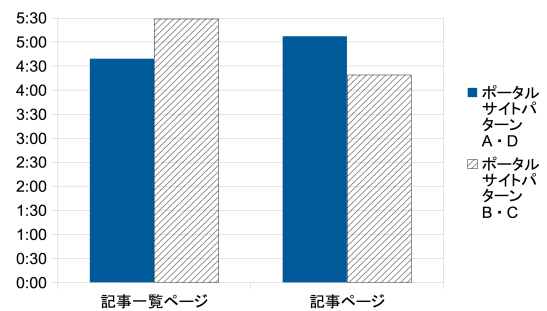


図 9 記事一覧と記事ページにおける合計滞在時間（被験者 4 名の平均値）

Fig. 9 The Total Time Spent in the Article Page and Article List Page

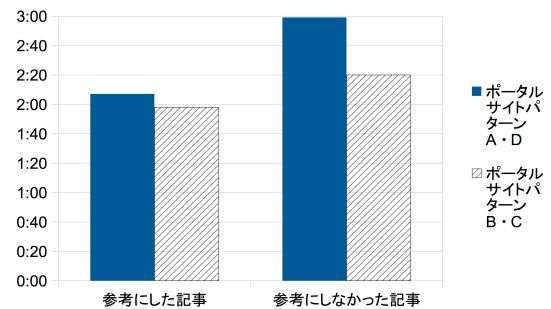


図 10 「参考にした記事」と「参考にしなかった記事」の合計滞在時間（被験者 4 名の平均値）

Fig. 10 The Total Time Spent in the Article Referred and did not Referred

トパターン A とパターン D と比較して、ブログポータルサイトパターン B とパターン C は、読むことで本文の内容を推察できる文章であるというアンケート結果となった。また、読むことで期待通りの記事に出会えたか、という設問に対してもパターン B とパターン C のサイトが優位な結果になった。

図 9 は記事一覧ページと記事ページにおける合計滞在時間を表したグラフである。ポータルサイトパターン B とパターン C について、記事一覧ページの滞在時間が伸び、一

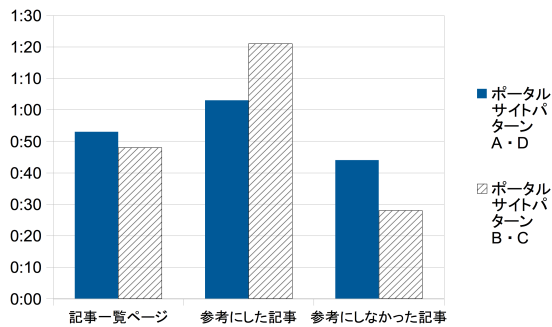


図 11 1 ページの平均滞在時間 (被験者 4 名の平均値)

Fig. 11 The Average of Time which Spent on a Page

覧ページの滞在時間が約 1 分短縮される結果となった。図 10 は記事ページにおける滞在時間を「参考にした記事」と「参考にしなかった記事」の 2 つに分けたものである。「参考にした記事」は、被験者が課題に回答する（興味のある物を挙げる）うえで参考にした旅行記の滞在時間合計を示している。「参考にしなかった記事」は、記事本文を閲覧したが課題の回答でとりあげられなかった記事の滞在時間を示している。ポータルサイトパターン B とパターン C について、「参考にした記事」の滞在時間はほぼ変化していないが、「参考にしなかった記事」の滞在時間は約 40 秒短縮される結果となった。図 11 は 1 ページの平均滞在時間を表したグラフである。パターン A とパターン D に関して、参考にした記事（アンケートに回答する上で参考にした旅行記）の滞在時間が 20 秒近く長い結果となった。記事一覧ページの閲覧と比較して、参考にした記事の閲覧時間が約 1.5 倍の閲覧時間となった。

5.3 考察

ポータルサイトパターン B とパターン C において、参考にしなかった記事の滞在時間のみが短縮されたことから、被験者が必要としなかったページの滞在時間自体が減ったと考えられる。記事一覧ページについては、滞在時間が増加したことから多くの記事概要の閲覧に時間を費やすことができたと考えられる。また参考にした記事（被験者が興味をもった物を挙げる上で参考にした旅行記）にかかる滞在時間が 20 秒長くなっていることから、従来の概要文章による記事を判断材料とするよりもユーザの探していた情報にたどり着くことができたといえる。

6. まとめと今後の展望

6.1 まとめ

本研究では旅行記に含まれる文を対象とし、観光トピックを基に文章分類を行う手法を提案した。旅行記の本文に潜在するトピックを調査し分析を行うことで、的確に本文を要約している概要文章がどの程度存在するのか検証を行った。調査結果より、一部の観光トピックについては詳

細文章内に観光トピックが潜在してしまうことが明らかとなった。概要文章生成手法の有効性評価の実験では、被験者に投稿者が記述した概要文章を用いたポータルサイトと観光トピックごとに生成した概要文章を用いたポータルサイトを交互に利用してもらい、閲覧行動がどのように変化したのか調査を行った。評価の結果から、被験者の必要としなかったページの閲覧時間が削減され、従来の概要文章と比較しユーザの探していた情報が記載されるページを見つけることができるとわかった。

6.2 今後の展望

トピックごとに動的に生成された概要文章により、閲覧者が必要とする記事や情報の発見を支援することが出来ると考えられる。今後は旅行記だけではなく、観光以外の分野の記事を管理するポータルサイトへ本手法を適用し、どのような効果が得られるか検証する必要がある。

特にブログのようにトピックが散在しがちな文章や、概要文章と本文で記載されるトピックに偏りがある文章において、本手法を活用することで効果があると考えられる。

参考文献

- [1] 奥村学, プログマイニング技術の最新動向, 電子情報通信学会誌 91(12), 1054-1059, 2008-12-01
- [2] 渡辺奈夕子, 岡本昌之, 菊池匡晃, 飯田貴之, 服部正典, Web 検索結果の推薦における提示項目が印象に与える影響, 情報処理学会, P61-68, 2009.3.13
- [3] Yesilada, Y., Jay, C., Stevens, R. et al.: Validating the use and role of visual elements of web pages in navigation with an eye-tracking study, Proc. WWW' 08, ACM, pp.11-20, 2008
- [4] 奥村学, 難波英嗣 (2005) 『テキスト自動要約 (知の科学)』 オーム社.
- [5] C.D.Paice, Constructing literature abstracts by computer: techniques and prospects. Information Processing and Management, Vol. 26, No. 1, pp. 171-186, 1990
- [6] 奥村学, 自然言語による情報アクセス技術 特集 3 テキスト自動要約, 情報処理 Vol.45, No.6, pp.574-579, 2004.6
- [7] 高見真也, 田中克己, 検索目的に基づくスニペットの動的再生成によるウェブ検索結果の個人適応化, 日本データベース学会, P33-36, 2007.9