

観光レビューから季節特有な情報と季節によって変化する情報の抽出手法の提案

鳥山 実桜[†] 灘本 明代[†]

[†] 甲南大学知能情報学部知能情報学科 〒658-8501 兵庫県神戸市東灘区岡本 8-9-1

E-mail: [†]ts1871074@s.konan-u.ac.jp, ^{††}nadamoto@konan-u.ac.jp

あらまし 現在、インターネット上には様々な観光地を対象としたレビューがある（以下、観光地レビューと呼ぶ）。この、観光地レビューには実際に訪れた人の体験談や感想など観光地に関する様々な情報が書かれている。一方、観光地には期間限定の割引やイベントなど季節によって変化する情報が多数あり、これら変化する情報は観光地レビューには多数存在する。しかしながら観光地レビューは数が膨大であり、かつ様々な情報が混在していることから、このような変化する情報を取得することは困難である。そこで本論文では、これら変化する情報の内、季節に着目し季節特有の情報と季節ごとに変化する情報の2種類の情報を抽出する手法の提案を行う。具体的には、観光地レビューからトピックを抽出し、トピックごとにクラスタリングを行う。そのクラスタ内において季節特有の情報と季節により変化する情報を抽出する手法の提案を行う。

キーワード 観光情報, 情報抽出, レビュー

1 はじめに

トリップアドバイザー¹やフォートラベル²など観光地を対象としたレビューサイトが普及している。旅行を計画するとき、このような観光地のレビュー（以下、観光レビューと呼ぶ）から実際に旅行に行った人の現地で体験した情報を集めることができるため、旅行の計画において参考にする人も多い。一方、観光地には年末年始の無料サービスや夏に行われる祭りなど、期間限定で行われるサービスやイベントのように季節特有のイベントが存在する。さらに、同じ観光地の景色でも春には桜が咲き、秋には紅葉がきれいで、冬には雪化粧になるといったように、季節により移り変わる情報も多数存在する。しかしながら、レビューサイトには誰でも気軽にレビューを投稿できるため、1つの観光地に対してもレビュー数が膨大である。その中には「楽しかった」、「思い出に残ってよかった」など単に投稿者の主観的な思いが書かれていることも多くあるため、これら季節特有の情報や季節で変化する情報を取得することが困難である。そこで本研究では、レビューから季節特有の情報と季節ごとに変化する情報を抽出する手法の提案を行う。

本論文において、季節特有の情報とは1つの話題について記述しているレビューが特定の季節に偏っているものと定義する。さらに季節で変化する情報とは1つの話題に対して四季それぞれの季節のレビューで構成されていると定義する。そのため、本論文では話題ごとに季節分類を行い情報を抽出する。以下に、季節特有の情報と季節で変化する情報の抽出手順を示す。

- (1) ユーザは観光地の名前を入力する。
- (2) システムは観光レビューサイトからその観光地に関す

るレビューを取得する。

(3) (2)で取得したレビューに対してトピック抽出を行い抽出したトピックごとにクラスタリングを行う。

(4) 各クラスタ内でレビューを文単位で分割し、1文ずつ季節分類を行う。

(5) (4)で季節分類された各文をタイプ別に分類する。この時、タイプは「願望」、「評価」、「経験」、「その他」の4種類とする。また、レビュー中の1文につき1タイプとする。

(6) (5)で抽出した文のタイプの内「願望」以外の「評価」、「経験」、「その他」の季節の文を用いて季節のレビューを取得する。

(7) 各クラスごとに(6)で得た季節のレビュー数の比率を求め、季節特有の情報か季節ごとに変化する情報のどちらであるかの判別を行う。

本研究により、季節特有のレビューや季節で変化するレビューを取得することが可能となる。これにより、季節によって異なる情報を事前に得ることができ、より観光客の需要に合った観光地を容易に探す事が可能となる。また、観光地は季節によって訪れる観光客数に波がある場合も多い。季節で変化する情報や特有な情報は観光客が少ない時期のアピールポイントとなり、観光振興への手助けにも繋がると考える。

以下、2章でレビューや旅行に関する関連研究を述べ、3章でトピック抽出とクラスタリングについて、4章では季節分類及びタイプの分類、季節特有の情報か季節ごとに変化する情報かの判定方法について述べる。そして5章では実験と考察について、6章ではまとめと今後の課題について述べる。

2 関連研究

観光レビューに関する研究は数多くされている。野守[1]ら

1: トリップアドバイザー <https://www.tripadvisor.jp/>

2: フォートラベル <https://4travel.jp/>

は観光レビューを分析することで、地域観光振興への活用を検討している。分析結果を活用するために観光レビューに対してテキストマイニングとPLSAを適用することで観光テーマを抽出し、そのテーマを軸とした地域の特徴と観光客の特徴を出している。観光テーマを抽出する点は本論文のトピック抽出と類似しているが、本論文では、観光客の特徴ではなく、季節で変化する観光地の情報を抽出する点で異なる。また叶内ら[3]は抽象的な要求に対する根拠付きの推薦文を提示するため、根拠説明データセットとシステムを構築し評価している。本論文では、要求に対する根拠を示す情報だけではない情報を抽出する点で異なる。

渡邊ら[2]はTwitter上の観光地名が含まれていないツイートから観光地に関する感想の抽出手法を提案している。観光地に関する感想を含むツイートを形態素解析し、tf-idf値を用いて観光地の特徴語として抽出することで、観光地に関するツイートの特徴を明らかにしている。この研究では、Twitter上から観光地に対する感情・印象の情報と特徴を抽出しており、観光レビューサイトのレビューを対象にしている点と人の感情だけではない情報の抽出のという点で本論文とは異なる。新井ら[4]の研究では、観光ルートを推薦する手法の提案も行われている。Twitter上の観光体験に関わるツイートを4種類のカテゴリに分類し、そのツイートに含まれる観光情報を用いている。本論文では抽出した情報を元にした観光ルートの提示ではなく、季節特有な情報と季節ごとに変化する情報を抽出している点が異なる。

また、観光レビューサイトからだけではなくブログや他サイトから情報を抽出する研究も行われている。遠藤ら[5]は観光振興への一助を目的として、旅行ブログと地域サイトから観光キーワードを抽出し、その抽出した観光キーワードのtf-idf値を利用して関連性判断とカテゴリ化する手法の提案を行っている。本論文は、トピックモデルを用いてトピック抽出を行うことで観光地の情報を取得している点で異なる。

3 トピック抽出とクラスタリング

本研究では、季節特有の情報や季節ごとで変化する情報を抽出する際、観光レビューのトピックに着目してこれらの情報を抽出する。そのため、本研究ではトピックモデルを用いて観光レビューからトピックを抽出し、トピックごとにクラスタリングを行う。トピック抽出には、Bleiら[9]が提唱したトピックモデルLatent Dirichlet Allocation(LDA)を用いる。LDAとは、トピックモデルの代表的な手法で一文書に複数のトピックが含まれることを表現できる。具体的には、1つの文書に複数の潜在的トピックが存在することを仮定し、各トピック及びトピックが存在する単語を多項分布でモデル化する。そしてこのモデルのパラメータを推論アルゴリズムによって推定し、それを文書データに潜在的に存在する話題の分布とそれぞれの話題の単語の分布とする。[6]PythonライブラリGensimを用いて作成する。

観光レビューサイトじゃらん³とフォートラベル⁴から1つの観光地に対する観光レビューを収集する。1レビューを1文書として収集した観光レビューをLDAを用いてトピックを抽出し、そのトピックごとにクラスタリングを行う。本論文では観光地に関するトピックを抽出するため、LDAのモデル作成に用いる単語の品詞を名詞に限定する。単語の抽出をするための形態素解析器にはJUMAN++⁵を使用する。また、季節で変化する情報、季節特有の情報がそれぞれ観光地特有の情報である場合も考えられるため、対象となる観光地ごとにその観光レビューをLDAを用いてトピックを抽出と行そのトピックごとにクラスタリングを行う。抽出するクラスタ数は11から20とし、最適なクラスタ数は観光地ごとにグリッドサーチを用いて人手で決定する。

4 季節分類

季節分類はレビューを文単位で分割し、季節キーワードを用いて行う。文の分割には、「。」「!」などの文の終わりに使われる句点や記号、タブまたは改行が2行以上出現した場合にそこが文の終わりとして分割する。ここでいう季節キーワードとは、その季節を代表する花や動物の名前(春ならば桜など)、一般的な行事(夏休みなど)の名称でとしている。また、人手によって春夏秋冬を判定したレビューに対して形態素解析を行い、出現回数が多い季節を表現する一般的な単語も季節キーワードに決定した。季節キーワードの例を表1で示す。

4.1 タイプ分類

観光レビューを文ごとに季節分類する際、季節キーワードが含まれているにも関わらず実際に訪れた季節を表現していないレビューが存在する。例えば、12月に訪れた時の観光レビューに”次は春に行きたい”といった文が含まれている時、実際に訪れた季節は冬であるにもかかわらず「春」という単語が入っているため、春に分類されてしまう場合がある。このようにレビューに含まれる投稿者の願望は削除すべきと考える。反対に季節を表現しているレビューとは、例えば、4月のレビューに”桜がとてもきれいだっ”や8月に訪れたレビューに”お盆だったからか人が多くてあまりゆっくりできなかった”のような現地での実際の経験や現地に対する評価を表す文が含まれているレビュー、観光地の概要などが記述されているレビューのことである。この2つの文には、「桜」が含まれているため春に、「お盆」という単語が含まれているため夏に分類できる。以上のことより、レビューを文単位で分割しそれぞれの文を「願望」、「評価」、「経験」、「その他」の4タイプに分類し、その季節を表現する文を抽出する。以下に、「願望」、「評価」、「経験」、「その他」の4つのタイプの抽出手法を示す。また、文のタイプの種類とそのタイプの決定方法は、櫛部らの研究[6]を参考にしている。タイプ分類の手順を図1に示す。

3: じゃらん <https://www.jalan.net/kankou/>

4: フォートラベル <https://4travel.jp/>

5: JUMAN++ [https://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN %](https://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN%20)

表 1 季節キーワードの例

季節	キーワード
春 (42 語)	桜, 花見, gw, ゴールデンウィーク, 3 月, 4 月, 5 月, つつじ, あやめ, ひな祭り, 憲法記念日, 昭和の日, みどりの日, こどもの日
夏 (43 語)	ひまわり, 夏祭り, かき氷, アイス, 6 月, 7 月, 8 月, お盆, 七夕, 浴衣, 朝顔, 水かけ, 海の日, 山の日, 盆休み, お中元, セミ
秋 (48 語)	コスモス, 月見, 七五三, ハロウィーン, 9 月, 10 月, 11 月, 敬老の日, 秋分の日, スポーツの日, 文化の日, 勤労感謝の日, 体育の日
冬 (41 語)	雪, クリスマス, 年末, バレンタイン 12 月, 1 月, 2 月, 師走, 神無月, カイロ, タイツ, 大晦日, 新春, 新年, 成人の日, 建国記念の日

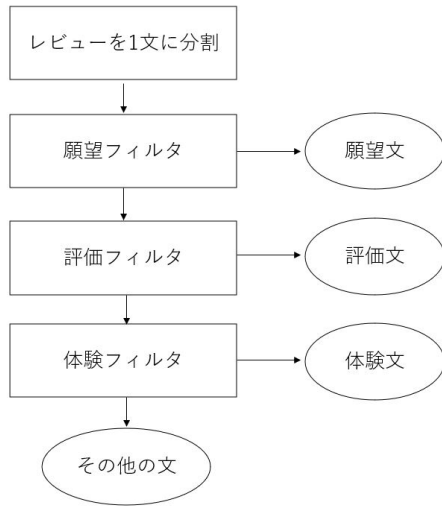


図 1 タイプ分類手順

● 希望タイプ

希望タイプとは「次は冬に行きたい」のような観光地に対する投稿主の希望を表す文を指す。「行きたい」、「参加したい」などの文末表現に着目し抽出する。具体的には希望を表現するルールを人手で判断して、その文に対し形態素解析を行い品詞の組み合わせルールを決定して抽出する。ここで決定した品詞の組み合わせルールとは、「参加してみたい」や「訪れたい」などのような自分自身がしたい言動を表現する「動詞＋て（接続助詞）＋みたい（形容動詞語幹）」、「動詞＋たい（助動詞）」、または、「取り入れてほしい」、「見学してみてほしい」のような自分以外の第三者への言動の希望を表現する「動詞＋て（接続助詞）＋ほしい（形容詞）」、「動詞＋て（接続助詞）＋みてほしい（形容詞）」である。また、「次回には」、「次の機会に」などのような先のことを示す表現を含む文も希望タイプと定義する。

● 評価タイプ

評価タイプとは、「夜景がきれいだった」のような観光地に対する評価表現がある文を示す。ここでは小林ら [7] の評価値表現辞書を用いる、具体的には評価表現辞書に当てはまる単語がある文を評価タイプとして抽出する。今回用いる評価表現辞書には「可憐」といった名詞、「美しい」といった形容詞などが収録されている。

● 経験タイプ

経験タイプとは、「無料で入れた」など投稿した人がその観光地現地においての実際の行動を表す文を示す。ここでは、池田ら [8] の提案する体験を記述する際に現れる体験表現を用いて

表 2 「希望タイプ」の文

文
次は雪の金沢に行ってみたいです!
四季を楽しめる公園ですので、冬や秋にも行ってみたいです。
またゆっくり散策したいです。
景色はきれいでしたが、晴れてる時に見たかったです。

表 3 「評価」タイプの文

文
紅葉が見頃かなと思ったら所々だけでしたけど綺麗でした。
雪つりに雪が積もる様子が青空に映えて本当にきれいでした。
しかしそれが気にならないほど歩くのが苦痛ではないほどうつくしい。
綺麗なお庭には橋があり、鯉もおよいでおり見て楽しめました。

表 4 「経験」タイプの文

文
金沢と言えば兼六園という感じで家内と訪れた。
雪吊りと桜のちょうど間の時期に訪れたが、園内は賑わっていました。
今年の降雪の機会に訪れて来ました。
庭園の中のお店で団子を食べた。

表 5 「その他」タイプの文

文
紅葉する木もあまりなかったです。
一面、真っ白な雪景色でした。
夏場の開園時間はなんと朝 4 時です。
桜自体は園内にはあまりないけど、周辺にいっぱいあります。

抽出する。具体的には、参加してみたなど自己の試みを表す表現の「動詞＋て/で（接続助詞）＋みる（補助動詞）」、見たことがある等経験したことを表現する「動詞＋こと（名詞）＋が（格助詞）＋ある（動詞）」、食べた・行った等自身が行動したことを表現する「動詞＋た（動詞接続尾辞終止）」、使っていた等継続的な行為を表現する「動詞＋て/で（接続助詞）＋いる/いた（補助動詞）」などである。その他のルールについては、以上に述べたルールの応用であり、ですます調等の言い回しの語尾を細やかな変化に対応したものである。

● その他タイプ

その他タイプとは、上記 3 つのどのタイプにも分類できない文である。

4 つのタイプに分類した文を「希望」、「評価」、「経験」、「その他」の順に表 2, 表 3, 表 4, 表 5 に示す。分類した 4 種類のタイプの内、「希望」タイプは投稿主が未来に対する思いを表現することが多いと考えられる。そのため、「希望」タイプは季節分類・情報抽出に不要な文として季節分類の対象の文から「希望」タイプの文のみを取り除く。残りの「評価」、「経験」、「そ

赤：評価文 青：体験 黒：その他

季節	レビュー
春	ライトアップや無料日があります人が少なくゆっくり散策できますもうすぐ桜が咲きます 桜の季節が最高です。無料開放期間があります。
	今回はゆっくり回ろうと思って行きました。一番印象的なのは、ツツジでした。三重の地 元でよく見るツツジだけでなく、紫がかかったピンクや、赤、ピンクなど、兼六園の中だけ ではなく、道路沿いに咲いているツツジがとてもきれいでした。
夏	他の観光施設に近く、庭もなかなか見ごたえがありました。ただ、坂道があり、夏に訪れ るところではないかもしれません。
	8月に一人で金沢観光に行った時に訪れました。さすがに日本三大庭園だけあり素晴らしい 景観です。あとライトアップされる夜は素晴らしい景観でした。
秋	紅葉の中のライトアップを無料でやっていると知り、急遽訪問しました。夜間のライトア ップだと雰囲気はともよいのですが、紅葉のキレイさがわかりにくかったです。早朝無料 開放も利用させていただきました。
	何度か訪れています。とにかく広く、整備が行き届いていると思います。紅葉には少し早 かった時間でしたが、十分に楽しむことができました。
冬	とても美しく、素晴らしい。雪が降るがこんなに暖かい作業だとは思わなかった。紅葉も美 麗だった。
	雪つりなどで中継が多くなる金沢いちの観光名所です。庭園をめぐる水の利用と数々の 松の木が印象的な日本を代表する、和風庭園です。

図 2 ユーザーインターフェース

の他」タイプの文は投稿主が現地で感じたことや行動・事実が記述されているため、観光地に訪れた季節を判別するのに適当であると考え、この3つのタイプから季節特有な情報と季節ごとに変化する情報を抽出する。なお、文を「願望」以外の3つのタイプに分類するのは、ユーザに結果を表示をする際情報を見やすくするためである。図2には提案手法を用いて抽出したインタフェースを示す。図2は、わかりやすく表示するために「評価」「経験文」「その他」の文の色を分けている。

4.2 季節のレビューを取得

分割したレビューの1文が4.1節で季節の文として抽出される場合そのレビューを季節のレビューとして取得する。図3に季節のレビューを取得した例を示す。「昨年のお花見の時期に友達に連れられて、一日中お花見しました。」という文は提案手法によって春の文として抽出される。その春の文はレビュー「昨年のお花見の時期に友達に連れられて、一日中お花見しました。兼六園の中も金沢城のお堀の周りも満開の桜がとても綺麗でした。夜は夜でライトアップされた桜が幻想的で、今まで見た桜の中でもより一層綺麗でした。」の1文であるためそのレビューを春のレビューとして取得する。また、「今年は8月14日15日に無料開放されていました。」は提案手法によって夏の文として抽出される。その夏の文はレビュー「今年は8月14日15日に無料開放されていました。混んでるかなと思ったら、出入り自由なのでみんなふらっと入ってふらっと出て行く感じで、全然混んでなかったです。」の1文であるためそのレビューを夏のレビューとして取得する。各クラスごとに春夏秋冬、それぞれの季節のレビューを取得する。

4.3 季節で変化する情報の判別

4.2節で取得した季節のレビューを用いて各クラスが特定の季節の情報であるのか、季節ごとで変化する情報であるのかの判別を行う。具体的な判別方法を図4に示す。

各クラスごとに春夏秋冬それぞれのレビュー数の比率を求める。レビュー数の比率がそれぞれの季節で15%以上あれば季節ごとで変化する情報、レビュー数の比率が15%未満の季節が1つでもあれば比率が高い季節の季節特有な情報であると判別する。図4ではクラス0は四季それぞれレビュー数の比率が

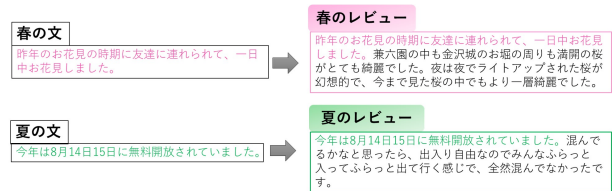


図 3 季節のレビューの取得例

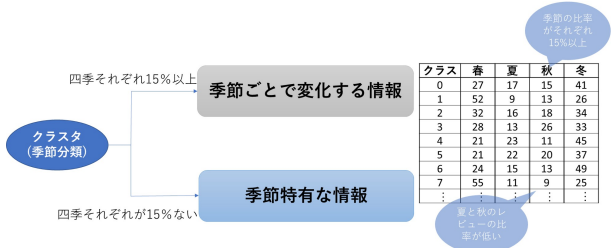


図 4 情報の判別方法

15%以上であるため季節で変化する情報、クラス7は夏と秋のレビュー数の比率が15%未満であるため季節特有な情報(クラス7の場合は春特有な情報)として判別する。

5 実験と考察

本論文の提案手法の有用性を確かめるために抽出したトピックに対して目視による実験と適合率を求める2つの実験を行う。また、提案手法によって季節で変化する情報が抽出できているかの考察も行う。

5.1 目視によるトピック評価実験

レビューからトピック抽出が上手くできているのかを示すために目視による実験を行った。

実験要件

提案手法を用いて、観光レビューのトピック抽出を行う。対象とした観光地は「兼六園」、「六甲山」、「美ら海水族館」の3つで、じゃらんとフォートラベルからそれぞれ7091件、2751件、8318件収集した観光レビューを用いた。クラス数数はそれぞれグリッドサーチによって16, 12, 13とした。

結果

兼六園、六甲山、美ら海水族館の観光レビューを対象としたトピック抽出の結果の一部を兼六園、六甲山、美ら海水族館の順に表6、表7、表8に示す。

考察

トピック抽出の3つの表より兼六園や美ら海水族館はトピック分類が上手くできた結果と思われるのに対して、六甲山は上手く分類できていないと結果となっている。六甲山の観光レビューから抽出したトピックでは、「夏」や「家族連れ」のように同じ単語が出てきており、LDAのモデルの学習が十分にできていないといえる。これは、LDAが1文書に含まれる単語数が少ない文章に対しては学習が困難という指摘があることから、[9]収集した六甲山の観光レビュー数が兼六園や美ら海水族館の4分の1と少なく、六甲山の1つの観光レビュー当たりの

表 6 兼六園の抽出したトピック

クラス	トピック
0	0.026* 灯籠 + 0.016* 龍石 + 0.014* 藩主
1	0.066* ライトアップ + 0.052* 桜 + 0.033* 夜
5	0.056* 無料 + 0.044* 早朝 + 0.022* 開放
12	0.040* 紅葉 + 0.035* 時期 + 0.024* 手入れ

表 7 六甲山の抽出したトピック

クラス	トピック
0	0.034* 夏 + 0.019* 標高 + 0.016* ドライブ
6	0.028* 夏 + 0.018* 空気 + 0.013* 箱詰め
9	0.024* 家族連れ + 0.017* 何度 + 0.014* 激
11	0.022* 家族連れ + 0.020* 有名 + 0.020* 夜中

表 8 美ら海水族館の抽出したトピック

クラス	トピック
0	0.035* ジンベエザメ + 0.029* 迫力 + 0.024* 海
6	0.034* チケット + 0.030* 夕方 + 0.024* 割引
11	0.028* 子供 + 0.025* 混雑 + 0.016* 平日
12	0.041* イルカ + 0.036* 無料 + 0.036* ショー

単語数も兼六園と美ら海水族館の単語数と比較すると少なかったためトピック分類の良い結果が得られなかったと考えられる。また、六甲山のレビューには兼六園及び美ら海水族館のようなイベントやキャンペーンなどの情報が少ないことから上手く分類できなかったと考えられる。

兼六園のクラス 1 のレビューには 3 つの単語からトピックが「夜間のライトアップ」のことで推定でき、表 9 に示すレビューのように桜の時期のライトアップや雪の時期のライトアップについて記述されているレビューが多くクラスタリングされており、兼六園のクラスタリングは上手くできたといえる。美ら海水族館では夕方 4 時以降にチケットの割引を行っており、その情報は「チケット」「夕方」「割引」の 3 つの単語群があることからクラス 6 でトピック抽出ができています。クラス 6 には表 10 に示すような割引に関するレビューが多くクラスタリングされていることから美ら海水族館のトピック抽出が上手くできているといえる。

また、兼六園の「灯籠」、「龍石」、「藩主」の歴史や有名スポットであったり、美ら海水族館の目玉であるジンベエザメに関したトピックが抽出できていることから観光地特有の情報を抽出できていると考えられる。反対に六甲山にはそういった六甲山特有の情報が少ないことからトピック抽出が上手くできなかったと考えられる。今回の考察により、観光レビュー数が少ないもしくは、1 つのレビューに含まれる単語数が少ない観光地、観光地特有の特徴が顕著ではない観光地のトピック抽出手法の改善が必要である点が確認できた。

5.2 定量的な評価実験

季節分類が上手くできているのかを確かめるために適合率を求める実験を行った。提案手法を用いて季節分類されたレビューを見てもらい、そのレビューがどの季節について記述されているのかをクラウドワーカー 10 人に判定してもらった。

なお、レビューの季節を判定する基準には「観光地に訪れた季節」、「季節に関連した記述がある」の 2 つを指定した。正解データは 6 人以上が判定した季節とし適合率を求める。

実験要件

対象の観光地は 5.1 節の目視による実験でトピック分類がうまくできた「兼六園」、「美ら海水族館」の 2 つである。それぞれの観光地に対して提案手法を用い、グリッドサーチによってクラスタ数を兼六園は 16、美ら海水族館は 13 とした。本実験では、兼六園はクラス 1 とクラス 5、美ら海水族館はクラス 6 とクラス 12 のそれぞれ 2 クラスを対象に実験を行った。なお、実験に用いた 2 クラスはそれぞれ各観光地の季節特有な情報、季節ごとで変化する情報である。

実験結果と考察

「兼六園」、「美ら海水族館」の 2 クラスの適合率の結果をそれぞれ表 11 と表 12 に示す。

兼六園は特に適合率が高い春と秋はうまく季節分類ができたといえる。しかし、クラス 1 の夏の適合率が低くなっている。これは夏の季節キーワードに「アイスクリーム」を定義しており、兼六園の名物である「金箔アイスクリーム(ソフトクリーム)」について記述されたレビューが提案手法では夏のレビューとして分類されたためと考えられる。また、冬の季節キーワードに「雪」を定義しているため、提案手法では「雪吊り」についての記述があるレビューが冬に多く分類されている。しかし、兼六園では雪吊りは準備が秋から行われており、実験では冬と判定されなかったレビューが多かったため冬の適合率が低くなったと考えられる。したがって、兼六園を対象とした実験結果から、提案手法には観光地ごとで適切な季節のキーワードの定義が必要であることがわかる。

美ら海水族館の実験結果からクラス 6、クラス 12 ともに夏の適合率が他の季節と比べ一番低いことがわかる。これは「炎天下」や「日焼け」などを夏のキーワードに定義しており、気温が高い沖縄では夏以外の季節でも同じ語句が使用されているため適合率が低くなったと考えられる。したがって、美ら海水族館を対象とした実験結果から、季節キーワードは観光地の気温に依存しない単語の定義が必要であると考えられる。

5.3 トピックが季節で変化する情報であるかの考察

提案手法を用いて季節で変化する情報を抽出できたのか考察を行った。表 13 と表 14 に「兼六園」、「美ら海水族館」の 2 クラスの正解データから季節ごとのレビュー数の比率を求めた結果を示す。

表 11 から、提案手法によってクラス 1 は春特有な情報、クラス 5 は季節ごとで変化する情報として抽出されたことがわかる。また、表 13 からは本論文の季節で変化する情報の定義と季節ごとの正解データのレビュー数の比率により、クラス 1 は春特有な情報、クラス 5 は季節ごとで変化する情報であることがわかる。クラス 1 の春のレビューには夜桜のライトアップについての記述されたレビューが多く、提案手法を用いて春特有な情報を抽出できたといえる。また、クラス 5 はそれぞれの季節のどの期間に「無料開放」があるについての記述されたレ

表 9 兼六園のトピック 1 のレビュー

レビュー文
<p>昨年のお花見の時期に友達に連れられて、一日中お花見しました。兼六園の中も金沢城のお堀の周りも満開の桜がとても綺麗でした。夜は夜でライトアップされた桜が幻想的で、今まで見た桜の中でもより一層綺麗でした。屋台の食べ物もいいですが、やっぱり、お弁当が最高です!一年に一度のお花見の時期。ぜひ兼六園へ足を運んで見てください。その美しさは圧巻です!</p> <p>兼六園の桜は綺麗だと思いました。夜のライトアップを見れなかったのが残念でしたが、桜を見れてよかったです。</p> <p>金沢に旅行来たら必ず訪れるといえる金沢の一大観光地、兼六園の夜間ライトアップに行ってきました。昼間は入場料がかかりますが、夜間ライトアップは入場無料で見学することが出来ました。雪とライトアップのコラボを見にきたのに、この日は 18 時ごろまで雨と気温が高かったため雪がなかったのが残念でした。しかしそのためか人は少なく、落ち着いて庭園を回ることが出来ました。昼間の見学もいいですが、夜間もおすすめです。雪が積もっている中、ライトアップをしていると聞いて夜に行きました。雪吊りされた樹木がとっても綺麗でした。</p>

表 10 美ら海水族館のトピック 6 のレビュー

レビュー文
<p>那覇空港からは遠いですが、沖縄旅行の定番スポットと言えばやはり、美ら海水族館です。入口を入ってすぐに、ヒトデやナマコとふれあえるコーナーもあります。巨大な水槽の中を悠々と泳ぐジンベイザメやマンタ、魚の群れなどが見れて時間を忘れてしまうくらい幻想的な光景です。午後 16 時以降は入館料が割引になりお得です。</p> <p>何度も行っているのですが今回は 16 時からの割引チケットで入りました。とにかく昼間暑いので海洋博公園内を歩くのも一苦労です。</p> <p>最初 p2 に駐車したのですが係員に水族館に近い駐車場を尋ねると p7 だと教えられて移動。3 割引きの 4 時からチケットを購入して入場。人気があるだけのことはあり人も多いです。いろいろな展示を見ましたが特にジンベイザメの大きさには驚きました。</p>

表 11 兼六園の適合率

クラス	トピック	季節	レビュー数	正解数	適合率
1	ライトアップ 桜 夜	春	304	276	0.908
		夏	51	33	0.647
		秋	74	68	0.919
		冬	152	113	0.743
5	無料 早朝 開放	春	36	29	0.806
		夏	39	32	0.821
		秋	35	31	0.886
		冬	64	51	0.797

表 12 美ら海水族館の適合率

クラス	トピック	季節	レビュー数	正解数	適合率
6	チケット 夕方 割引	春	22	16	0.727
		夏	67	48	0.716
		秋	21	19	0.905
		冬	21	17	0.810
5	イルカ 無料 ショー	春	18	15	0.833
		夏	88	48	0.545
		秋	14	10	0.714
		冬	30	24	0.800

表 13 兼六園の正解データをから求めた季節のレビュー数比率

クラス	トピック	季節	正解数	レビュー数比率 [%]
1	ライトアップ 桜 夜	春	276	56
		夏	33	7
		秋	68	14
		冬	113	23
5	無料 早朝 開放	春	29	20
		夏	32	22
		秋	31	22
		冬	51	36

表 14 美ら海水族館の正解データをから求めた季節のレビュー数比率

クラス	トピック	季節	正解数	レビュー数比率 [%]
6	チケット 夕方 割引	春	16	15
		夏	48	49
		秋	19	10
		冬	17	25
5	イルカ 無料 ショー	春	15	16
		夏	48	48
		秋	10	19
		冬	24	17

ビューが多く、季節ごとで変化する情報を抽出ができたといえる。したがって、兼六園のような季節によって景色が変化する、または季節ごとで異なるイベントがある観光地に対しては、提案手法を用いて季節で変化する情報を抽出できると考えられる。

表 12 からは、提案手法によってクラス 12 は夏特有な情報、クラス 6 は季節ごとで変化する情報として抽出されたことがわかる。また、表 14 から本論文の季節で変化する情報の定義と正解データを用いた季節ごとのレビュー数の比率により、クラス 6 は季節ごとで変化する情報、クラス 12 は夏特有な情報であることがわかる。しかし、美ら海水族館ではイルカショーは一年通して無料で行われているため、クラス 12 は夏特有な情

報ではなく、提案手法を用いて季節ごとで変化する情報を抽出できていないといえる。また、クラス 6 の夕方からあるチケットの割引は季節によって割引率の変化や時間の変化がないため、季節ごとで変化する情報ではなく、提案手法を用いて夏特有な情報を抽出できていないといえる。したがって、美ら海水族館のような季節によって景色の大きな変化がない、または季節ごとで異なるイベントがない観光地に対しては、提案手法を用いて季節で変化する情報を抽出できるとはいえないと考えられる。

6 まとめと今後の課題

本論文では、季節で変化する情報の定義を行い、それらを抽

出する手法の提案を行った。まず、観光レビューサイトから収集した観光レビューから LDA を用いてトピック抽出をし、その抽出したトピックでクラスタリングを行う。そして、季節分類に不要な文の削除を行うために、レビューを「願望」、「評価」、「経験」、「その他」の 4 種類のタイプに分類し、そのうちの「願望」タイプ以外の「評価」、「経験」、「その他」に分類された文に対して季節キーワードを用いてレビューを季節分類する手法の提案をした。実験結果から季節による変化が大きい観光地に対しては季節で変化する情報を抽出できるが、季節による変化が小さい観光地に対しては季節で変化する情報を抽出できるとはいえないことがわかった。

今後の課題として、提案手法で抽出できなかった情報を抽出するためにトピック抽出に Tweet のような短い文章に対して LDA より一貫性の高いトピック抽出ができる BTM(Biterm topic model) を用いることを検討する必要がある。また、季節キーワードによって季節分類がうまくいかなかったレビューに対して、季節分類の精度を上げるためにも文のタイプを利用した季節分類手法の提案が挙げられる。

謝 辞

論文の一部は JSPS 科研費 19H04218, 19H04221, 20K12085, 及び私学助成金 (大学間連携研究助成金) の助成によるものである。ここに記して謹んで感謝の意を表する。

文 献

- [1] 野守 耕爾, 神津 友武, ” 観光に関するユーザーレビューデータを用いた観光客の話題分析と地域観光振興への活用の検討”, サービスロジー論文誌 2 巻 2 号 2019
- [2] 渡邊 小百合, 吉野 孝, ” 観光地名なしツイートからの観光地に関する感想の抽出手法”, 情報処理学会論文誌 Vol59 No.1 43-51(2018)
- [3] 叶内 晨, 根石 将人, 林部 裕太, 岡崎 直観, ” 旅行情報サイトのレビューを用いた抽象的な要求に対する根拠付き推薦文の生成”, 言語処理学会 第 26 回年次大会 発表論文集 (2020 年 3 月)
- [4] 新井 晃平, 新妻 弘崇, 太田 学 ” Twitter を利用した観光ルート推薦の一手法”, DEIM Forum 2015 G7-6
- [5] 遠藤 雅樹, 大野 成義, ” 地域サイト及びブログからの観光情報の自動抽出と融合”, DEIM Forum 2013 F9-2
- [6] 櫛部 健汰, 伊藤 恵, ” 観光地に関するソーシャルメディア投稿の分類手法の提案”, 日本ソフトウェア科学会第 38 回大会 (2021 年度) 講演論文集.
- [7] 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊, ” 意見抽出のための評価表現の収集”, 自然言語処理, Vol. 12, No. 3(2005), pp. 203-222.
- [8] 池田 佳代, 田邊 勝義, 奥田 英範, 奥 雅博, ” Blog からの体験情報抽出”, 情報処理学会論文誌 (2008)
- [9] D. M. Blei, A. Y. Ng, and M. I. Jordan, ” Latent Dirichlet Allocation, ”, Journal of Machine Learning Research, 3:993-1022, 2003.