

PLSA による観光レビューからの季節性トピック抽出の一手法

檜橋 昂[†] 上野 史^{††} 太田 学^{††}

[†] 岡山大学工学部情報系学科 〒700-8530 岡山市北区津島中 3-1-1

^{††} 岡山大学学術研究院自然科学学域 700-8530 岡山市北区津島中 3-1-1

E-mail: [†]pn9e9i4d@s.okayama-u.ac.jp, ^{††}{uwano, ohta}@okayama-u.ac.jp

あらまし 観光地のイベントや名物は、季節などの時間的要素の影響を受けやすく、変化しやすいものがある。そのため、観光客の間で話題とされている内容はその時期に応じて把握することが重要である。そこで本稿では、観光地のユーザレビューから季節性トピックを抽出する手法を提案する。具体的には、観光地の口コミを投稿時のタイムスタンプによって各季節に分割し、トピックモデルの Probabilistic Latent Semantic Analysis (PLSA) を用いて、季節ごとのレビューからトピックを抽出する。その後、各トピックの特徴ベクトルを求め、抽出したトピックを季節性トピックと非季節性トピックに分類する。実験では、岡山県の後楽園の観光レビューを用いてトピック抽出と季節性トピックの分類を行い、季節性トピックを取得した。取得した季節性トピックから、後楽園を訪れる観光客は、春には桜が満開な庭園の様子を楽しみ、夏にビアガーデンや花火などのイベントを話題にしていることが分かった。また、秋には、綺麗な紅葉の景色を見て、冬には丹頂鶴の放鳥や芝焼きなどの風物詩に関心があることが確認できた。

キーワード 観光, PLSA, トピック抽出

1 はじめに

観光業で地域活性化を目指すうえで、観光地の特徴やその価値を把握することは必要不可欠である。しかし、観光サービスを提供する旅行会社と観光客の間には、観光地の価値に関してギャップが生じる場合があり、観光客がその観光地に求めるニーズを把握することが重要である。従来はアンケートや口コミにより、それを把握していた。しかし、これらの情報は観光地によっては膨大となるため、その全てを精査し、人手でニーズを把握するには多くの時間やコストがかかる。また、観光地によって特徴は様々であるため、各観光地に求める観光客のニーズは変化する。そのため、観光客視点で各観光地に対する関心や話題をつかむことが重要となる。また、観光客視点の観光地に対する関心事や話題は季節などの時期によって変化するものも多い。そこで本稿では、観光客の視点で記述された観光レビューを用いてトピックを抽出し、抽出したトピックを季節性トピックと非季節性トピックに分類する手法を提案する。このようにして取得した季節性トピックを観光サービス提供側が見ることで、低コストで季節ごとの観光客の関心や話題を把握できる。提案手法では、旅行サイトのレビューをタイムスタンプで各季節に分割し、トピックモデルの Probabilistic Latent Semantic Analysis (PLSA) を用いて、季節ごとのレビューからトピックを抽出する。その後、各トピックの特徴ベクトルの類似度により、抽出したトピックから季節性トピックを分類により取得する。

本論文の構成は以下のとおりである。2 節では関連研究について述べる。3 節では、季節性トピックを取得する手法について述べ、4 節では各季節のレビューを用いたトピック抽出の実験について述べる。5 節では、抽出したトピックの分類実験について述べ、最後に 6 節でまとめる。

2 関連研究

2.1 口コミを用いた観光地の分析に関する研究

野守ら [1] は、日本全国の観光資源のユーザレビューデータを用いて、観光客のコメントが記されたレビューに PLSA を適用することで、話題とされている「歴史資料展示」、「動物園・水族館」などの 15 の観光テーマを抽出した。また、この研究では、本来の PLSA の入力である「文書」×「単語」の共起行列ではなく、「観光資源」×「係り受け表現」の共起行列を作成し、これに PLSA を適用する。ここで、係り受け表現とはレビューに出現する係り受け関係にある名詞と形容詞、または名詞と動詞の組み合わせを指す。この共起行列に PLSA を適用することで、観光資源と係り受け表現の背後にある観光テーマを抽出した。結果として、国内観光客が良く話題としてあげる 15 の観光テーマを抽出し、テーマごとに地域の特徴と観光客の特徴の定量的な分析を可能とした。

大久保ら [2] はグローバルな観光競争環境で優位となるために、訪日外国人観光客からみた観光地イメージの把握を目的とした。そこで、海外版旅行ガイドブックと旅行口コミサイトを取り上げ、その内容を解析することで観光地イメージに関する有用な情報を抽出した。対応分析によって、旅行ガイドブックと口コミにある各観光地のイメージの特徴の一部を抽出することができた。また、観光地によっては、ガイドブックに記載されている情報とは異なる特徴を観光客がイメージしていることがわかった。出身国別の分析結果からは、出身国によって各観光地に対するイメージに差があることが明らかとなった。そのため、口コミに合わせたガイドブックの修正や、訪日外国人観光客を細分化した観光マーケティングを検討する必要があるこ

とがわかった。

2.2 時系列データからのトピック抽出に関する研究

松本ら [3] は、時系列テキストデータに対するトピックモデルの Dynamic Topic Model (DTM) [4] に代わるものとして、期間ごとに Latent Dirichlet Allocation (LDA) [5] を用いて取得したトピックに対して、同一期間内のトピックの集約と次期間のトピックへの推移関係の発見の手法を提案した。LDA を用いてトピックの特徴ベクトルを生成し、その類似性を確認することで、同期間のトピックの集約や次期間のトピックへの推移を表現した。結果として、トピック集約前の全 21 のトピックのうち、12 のトピックがトピックを集約した後のほうがコーヒレンスのスコアが上がるという結果になった。

高橋ら [6] は、時系列ニュースにおけるトピックのバーストを検出する方法を提案した。ここでいうバーストとは、世の中における特異な出来事に対応して、ある時期からその出来事に関連するニュース記事が急激に増加する現象を指す。提案した手法では、バースト解析にトピックモデルを適用することで、トピック単位でのバースト検出を可能にした。

2.3 PLSA

Probabilistic Latent Semantic Analysis (PLSA) は Hofman ら [7] が 1999 年に発表した次元圧縮の手法であり、トピックモデルの代表的な手法の一つである。元々、PLSA は、情報検索の分野で、膨大な文書データを分類するために開発された手法である。ここで、文書データとは、行に文書の情報を、列にその文書内の単語の出現頻度の情報を持つ「文書」×「単語」の共起行列を指す。これに、PLSA を適用して、例えば、文書 5,000 件×単語 10,000 語という高次元データが 5 個のトピックで表現できるとすると、高次元データは文書 5,000 件×トピック 5 個という低次元のデータに置き換えることが可能となる。PLSA の関連手法として、他には Latent Semantic Analysis (LSA) [8] や LDA が挙げられる。LSA は共起行列の特異値分解により、トピックを表す潜在変数を取得する次元圧縮手法である。また、LSA を確率的に処理したものが PLSA であり、LDA は PLSA をさらに拡張した手法である。文書における各トピックの生成確率が、PLSA では学習させた観測データのみから定義されるが、LDA では確率分布のディリクレ分布を仮定して定義される。PLSA では、学習データに過剰に適合することで他のデータの適合度が下がる過学習が生じやすく、また、新しい文書におけるトピックの生成確率は生成されないが、LDA は新しい文書におけるトピック推定を可能とする。

3 季節性トピックの抽出

3.1 観光レビュー

本稿では、じゃらん net¹の観光地の口コミである観光レビューを用いてトピックを抽出する。図 1 はじゃらん net のある観光地のユーザの実際の口コミである。この青枠で囲まれた箇所

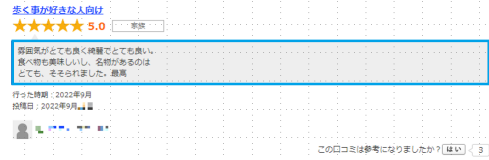


図 1 使用するじゃらん net のレビュー例

の口コミをレビューとして収集し、「行った時期」からこのレビューを四季に分割する。本稿では、3 月から 5 月を春、6 月から 8 月を夏、9 月から 11 月を秋、12 月から 2 月を冬と定義する。

3.2 共起行列の作成

形態素解析のツールである MeCab²と ipadic-neologd³の単語分かち書き辞書を用いて、取得したレビューごとに形態素解析をした。ipadic-neologd では、名詞が 14 項目に細かく分けられている。本稿では、その中の「非自立」、「代名詞」以外の名詞を PLSA に適用する「文書」×「単語」の共起行列の作成のために取得する。さらに、14 項目の中の「接尾」は「一本」の「本」のような接尾辞を指し、その単語のみではあまり意味を持たない。そのため、「接尾」と判定された単語はその直前の単語に連結する。また、「非自立」とは「言った矢先」の「矢先」のような単独では意味が理解しにくい名詞を指す。さらに、本稿ではトピック抽出を行う上で、ノイズとなる単語を除くなどのクリーニングを行う。例えば、レビューの形態素解析の結果、得られる名詞の中の数字は、全て 0 に変換する。この理由は、数字は観光地の特徴を探るうえで、重要性が低く、ノイズとなり得ると考えたためである。また、他にノイズとなり得る単語の出力を防ぐために以下のストップワードを定義し、これらを除く。

(1) slothlib⁴で公開されている日本語ストップワード (全 310 単語)

(2) 季節ごとのレビューにおいて出現回数が 1 回の単語

(1) の slothlib の日本語ストップワードの中には、「あちら」、「こちら」などの指示語から、単位を表す「円」、「個」などが含まれている。

3.3 季節性トピックの取得

本稿では、観光地の各季節のレビューから作成した共起行列に PLSA を適用し、まずトピックを抽出する。その抽出したトピックが季節性トピックであるか非季節性トピックであるかを分類して、季節性トピックを取得する。

3.4 PLSA によるトピック抽出

本稿では、PLSA の実装に PyPI⁵で公開されている外部ライブラリの plsa を用いる。また、トピックモデルでは事前に抽出するトピック数を決めておく必要があり、このトピック数に

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

2: MeCab <https://pypi.org/project/mecab-python3/>

3: ipadic-neologd <https://github.com/neologd/mecab-ipadic-neologd>

4: slothlib <https://ja.osdn.net/projects/slothlib/>

5: PyPI <https://pypi.org/project/plsa/>

よって結果が左右される。不適切なトピック数であれば、類似したトピックが多数抽出されたり、まとまりのないトピックが抽出されたりする。そこで本稿では、複数のトピック数を設定してトピックを抽出する。

3.5 季節性トピックかその他のトピックかの分類

3.4節の方法で抽出したトピックは、季節性トピックか非季節性トピックかに以下のように分類する。

(1) トピックの特徴ベクトルの算出

まず、PLSAによって抽出したトピックに現れる単語をそのトピックの構成単語とする。次に、Word2vec⁶により構成単語のうち生成確率上位20語の分散表現を取得し、その分散表現と構成単語の生成確率の積の平均をそのトピックの特徴ベクトルとする。

(2) コサイン類似度を用いた分類

1つの季節のトピックの特徴ベクトルとその他の季節の全てのトピックの特徴ベクトルとのコサイン類似度を算出する。その最大値が閾値 θ より小さければそのトピックを季節性トピックとする。よって、最大値が θ 以上であればそのトピックは季節性トピックではない。

4 トピック抽出実験

4.1 実験内容

ここでは、各季節のレビューから作成した共起行列に、3.4節で説明した方法でPLSAを適用してトピックを抽出する。なお、トピック数は5と10の場合で実験する。

4.2 データセット

本実験では、じゃらんnetから収集した岡山県岡山市にある後楽園の口コミのレビューを用いる。具体的には、2007年8月9日から2022年12月17日までの期間の後楽園のレビューを2048件収集した。この全レビューのうち、春のレビューが688件、夏のレビューが587件、秋のレビューが497件、冬のレビューが276件であった。

4.3 トピック抽出実験の結果

ここでは、後楽園の夏のレビューから抽出したトピックについてその結果を示す。表1、表2はそれぞれ、トピック数が5の場合のトピック抽出の結果とトピック数が10の場合の結果である。また、抽出したトピックに現れる単語を構成単語として、その構成単語の生成確率上位10語を見て、本稿の第一著者が主観で各トピックに名前を付けた。

二つの結果を比較すると、どちらも構成単語にまとまりのない「その他」トピックが一件ずつ抽出された。また、トピック数が10の場合、夏のイベントのトピックが2つ抽出された。一方、トピック数が10の場合は、トピック数が5の場合に得られなかった「ボランティアの説明」、「夏・食べ物」トピックが

表1 後楽園の夏のレビューから抽出したトピック (トピック数: 5)

トピック名	構成単語 (生成確率上位 10 語)
夏のイベント	幻想, ライトアップ, おすすめ, 庭園, イベント, ピアガーデン, 0, 夜間, 後楽園, 綺麗
その他	場所, 兼六園, 花火, 整備, 敷地, 岡山, 日本, 後楽園, 岡山城, 結婚式
自然・景観	季節, 写真, 花, 城, 綺麗, 手入れ, 自然, 散策, 植物, 桜
後楽園周辺のスポット	観光, 抹茶, 岡山城, 公園, 入園, 緑, セット, 後楽園, 友人, 花
ライトアップの様子	岡山駅, 雰囲気, 無料, 浴衣, 幻想的, ライトアップ, 0 時, 景色, 庭, 庭園

表2 後楽園の夏のレビューから抽出したトピック (トピック数: 10)

ラベル名	構成単語 (生成確率上位 10 語)
夏のイベント 1	幻想, ライトアップ, 浴衣, 庭園, 幻想的, イベント, おすすめ, ピアガーデン, 屋台, 開催
日本三大庭園・景観	兼六園, 公園, 日本, 整備, 借楽園, キレイ, スポット, 岡山, 敷地, 駐車場
雰囲気・景観	散歩, 庭, 雰囲気, オススメ, 岡山駅, 家族, 綺麗, 場所, 景色, 子供
ボランティアの説明	0 円, 0 分, 花, 四季, 大人, 入園, 岡山城, 説明, 緑, ボランティア
景観	日本庭園, 茶屋, 季節, 芝生, 三大, 花, 広大, 後楽園, 風景, 暑さ
その他	雨, 日陰, 手入れ, 庭園, 様子, 岡山城, 殿様, 対策, 岡山市内, 休憩
観光スポット	写真, 趣, 岡山, 日本三大名園, きれいな, 旅行, 整備, 観光スポット, きれいな, 庭園
夏・食べ物	セット, 大変, 抹茶, 真夏, 岡山城, 夏の夜, 売店, 購入, 気分, ソフトクリーム
夏のイベント 2	花火, 最高, お勧め, 綺麗, ライトアップ, イベント, 鯉, 花火大会, 食事, ピアガーデン
ガイドの案内	散策, 0 時間, 歴史, 友人, 仕事, ガイドさん, そば, 案内, 必要, 疲れ

抽出された。

表2の「夏のイベント1, 2」のトピックでは、構成単語として「ライトアップ」、「イベント」、「ピアガーデン」が取得された。「ライトアップ」は一年を通して行われているイベントではあるが、「花火」、「ピアガーデン」は夏の期間のみで開催されるため、これらのトピックは「夏のイベント」である。また、この2つのトピックの構成単語には同じ単語や類似する単語が取得されたことから抽出したトピックが重複して抽出されたといえる。また、「真夏」、「夏の夜」の単語や「抹茶」、「ソフトクリーム」といった単語が取得されたトピックのトピック名は「夏・食べ物」とした。

4.4 考察

本稿では、後楽園の夏のレビューからトピック数を5または10として、トピックを抽出した結果を示した。どちらのトピック数でトピック抽出を行った場合も、構成単語にまとまりのない「その他」トピックが抽出された。しかし、トピック数が10の

6: Word2vec <https://radimrehurek.com/gensim/models/word2vec.html>

場合は、トピック数が5の場合で実験を行ったときには抽出されなかったトピックが抽出された。また、他の季節のレビューから抽出したトピックにおいても同様の事例が確認できた。例えば、春のレビューから抽出されたトピック10件のうち、「その他」トピックは1件であり、重複したトピックは0組であった。

また、トピックの構成単語には、「キレイ」、「綺麗」のような表記ゆれがあった。トピックの重複を防ぐためにもこの表記ゆれへの対処が必要である。

本稿では、第一著者がトピックの構成単語の生成確率上位10件を見て、トピック名を付けた。しかし、主観によってトピック名を付けることは必ずしも望ましくはないため、自動でトピック名を付与する方法を検討する必要がある。

5 季節性トピック抽出のための分類実験

5.1 実験内容

ここでは、4節でトピック数を10としたときに抽出したトピックを3.5節で説明した手法により分類して、季節性トピックを取得する。そして、著者がトピックの生成確率上位10語の単語を見て、そのトピックに観光地特有な季節性があるか、観光地によらない普遍的な季節性があるか、季節性がないかを判断し、分類結果と比較する。ここで、特有な季節性とは、ある季節のみで行われるその観光地特有のイベントや名物のトピックを指す。例えば、岡山県倉敷市の倉敷美観地区には春に春宵祭りが開催されるが、この「春宵祭り」は特有な季節性トピックである。また、普遍的な季節性とは、ある季節のみで行われるその観光地特有のものでない普遍的なイベントや名物のトピックを指す。例えば、お花見トピックや紅葉トピックがこれに該当する。

なお、提案手法はトピックを季節性トピックか非季節性トピックかに分類する。すなわち、特有な季節性トピックか普遍的な季節性トピックであるかについては区別しない。

5.2 季節性トピックの分類結果

表3には、四季のレビューから抽出した全てのトピックの40件について季節性があるかを判断した件数と、そのトピックに分類を行った結果の件数をまとめる。

閾値 θ が0.4の場合は2件が季節性トピックと分類されたが、この2件は誤分類である。また、閾値 θ が0.5の場合は17件が季節性トピックと分類されたが、正しく分類されたのは7件であり、10件は誤分類である。

また、再現率と適合率を表3から閾値ごとに以下の式で算出する。なお、再現率は、季節性があるとしたトピックのうち、実際に季節性トピックと分類された割合を表す。適合率は季節

表3 季節性トピックの分類結果とその判定

		分類結果			
		$\theta=0.4$		$\theta=0.5$	
		季節性	非季節性	季節性	非季節性
正解	季節性	0件(TP)	9件(FN)	7件(TP)	2件(FN)
	非季節性	2件(FP)	29件(TN)	10件(FP)	21件(TN)

表4 適合率と再現率

	再現率	適合率
$\theta=0.4$	0	0
$\theta=0.5$	0.778	0.412

表5 季節性トピックの分類結果

季節	トピック名	構成単語の一部	季節性か	分類結果 (閾値)	
				(0.4)	(0.5)
春	日本三大庭園・春の景観	兼六園, 偕楽園, 桜, 綺麗, 満開	○ (特有)	×	○
	桜・お花見	弁当, お花見, 芝生, 綺麗, 岡山城	○ (普遍)	×	×
	後楽園の感想	雰囲気, 価値, 便利, 桜, 有名	×	×	×
夏	夏のイベント1	幻想, 浴衣, 幻想的 ライトアップ	○ (特有)	×	○
	夏のイベント2	花火, ライトアップ, ビアガーデン	○ (特有)	×	○
	夏・食べ物	真夏, 抹茶, ソフトクリーム	○ (特有)	×	○
秋	秋の後楽園の様子1	紅葉, 綺麗, 小川, 鶴, 鯉	○ (特有)	×	○
	秋の後楽園の様子2	紅葉, 池, 庭園, 茶畑, 岡山城	○ (特有)	×	○
	歴史	藩主, 城, 綱, 池田, 政	×	○	○
冬	放鳥・芝焼き	芝, 芝焼き, 鶴, 放鳥	○ (特有)	×	○
	冬の景観	紅葉, 梅, 広大, 木々, キレイ	○ (特有)	×	×
	その他	運動, 公園, 元日, ボート, 一日	×	×	○

性トピックに分類されたトピックのうち、実際にそのトピックに季節性があるとしたものの割合である。その結果を表4で示す。

$$\text{再現率} = \frac{TP}{TP + FN} \quad (1)$$

$$\text{適合率} = \frac{TP}{TP + FP} \quad (2)$$

この結果、閾値 θ が0.4の場合は季節性トピックを全く分類できなかったことがわかった。また、閾値 θ が0.5の場合は、再現率からは、季節性があると判断したトピックの多くを季節性トピックとして正しく分類できたことがわかった。適合率からは、季節性トピックであると分類したトピックには誤分類が多いことがわかった。

表5には、各季節のレビューから抽出したトピックの分類結果とその構成単語の一部を示す。「日本三大庭園・春の景観」、「桜・お花見」トピックは、それぞれ構成単語に「桜, 綺麗, 満開」、「弁当, お花見, 芝生」の単語があることから春の季節性トピックである。また、「桜・お花見」トピックは後楽園に限ら

ず、普遍的なイベントであることから普遍的な季節性トピックである。分類結果は「日本三大庭園・春の景観」トピックは季節性トピック、「桜・お花見」トピックは非季節性トピックとなった。「放鳥・芝焼き」トピックの構成単語の「芝焼き、鶴、放鳥」は冬の時期に後楽園で開催される芝焼きと丹頂鶴の放鳥のイベントに関するものである。このトピックは季節性トピックとして正しく分類された。

5.3 考察

本実験では、4節で抽出したトピックを分類して季節性トピックを抽出した。結果として、著者が季節性トピックと判断した9件のトピックのうち、閾値を0.5とした場合、7件のトピックを抽出できた。しかし、「桜・お花見」トピックと「冬の景観」トピックは非季節性トピックに分類された。この二つのトピックは、構成単語にそれぞれ「芝生、綺麗、岡山城」、「広大、木々、キレイ、敷地」などが含まれる。これらの単語は、後楽園のレビューに一年を通して多く確認された。そのため、他のトピックの構成単語にも類似した単語が取得されていることから季節性トピックに分類されなかったと考える。

また、著者が季節性がないと判断した31件のトピックのうち、閾値を0.5とした場合、10件のトピックが季節性トピックに誤って分類された。この10件のトピックのうち3件が構成単語にまとまりのない「その他」トピックであった。これらの構成単語には、他のトピックにあまり取得されていない単語が含まれていた。そのため、誤分類されたと考える。また、秋のレビューから抽出された「歴史」トピックは閾値が0.4の場合でも季節性トピックに誤分類された。この「歴史」トピックは秋と冬のレビューからのみ抽出されたため、季節性はないが季節性トピックに分類された。

また、本稿では、抽出したトピックに第一著者の主観で観光地特有な季節性があるか、観光地問わず普遍的な季節性があるか、季節性がないのかを判断した。ただし、分類では、特有な季節性トピックか普遍的な季節性トピックかについては区別していない。観光地特有な季節性トピックが取得できれば、その観光地のある時期にしか体験することのできない希少性の高いイベント情報が期待できる。また、普遍的な季節性トピックが取得できれば、紅葉などの季節性のイベントや名物を持つ観光地をその季節に応じて推薦する観光スポット推薦の研究への展開が期待できる。そのため、特有な季節性トピック、普遍的な季節性トピックに関して分析することは重要である。

季節性があると判断した9件のトピックのうち、8件が特有な季節性トピックであり、1件が普遍的な季節性トピックだった。これらの結果から、後楽園を訪れる観光客の間では、他の観光地にあるような普遍的なイベントは話題にあがっておらず、後楽園特有のイベントや名物に対して関心があることがわかった。例えば、春は桜が満開の後楽園の景観を楽しむことができ、夏は庭園のライトアップや花火、ピアガーデンなどのイベントが話題に上がっていることがわかった。また、秋には後楽園の紅葉の景色、冬には丹頂鶴の放鳥や芝焼きの風物詩に関心があるといえる。

6 まとめ

本稿では、PLSAを用いて各季節の観光レビューからトピックを抽出した。そして、各トピックの構成単語の分散表現に生成確率を乗じてトピックの特徴ベクトルを算出し、その類似度から季節性トピックか非季節性トピックに分類した。

各季節に分割したレビューからトピック抽出を行う実験では、岡山県の後楽園の四季のレビューに対して、PLSAを適用しトピック数を5と10にしてトピックを抽出した。その結果、いずれの季節においても、トピック数が10の場合に、トピック数が5の場合では得られなかったトピックを抽出することができた。

また、トピック数が5と10のどちらの場合においても、構成単語にある程度まとまりのあるトピックが抽出されたが、一方で、まとまりのないトピックも抽出された。さらに、構成単語に重複がある類似したトピックもいくつか抽出された。例えば、夏のレビューから抽出された「夏のイベント」に関するトピックは2件あった。

抽出したトピックが季節性トピックであるか非季節性トピックかの分類実験では、「観光地特有な季節性トピック」、「普遍的な季節性トピック」としていた9件のトピックのうち、7件のトピックを正しく季節性トピックに分類できた。しかし、「非季節性トピック」の約三分の一が、誤って季節性トピックに分類された。

本稿では「観光地特有な季節性トピック」と「普遍的な季節性トピック」の分類は行っていないため、今後はそれらの分類方法についても検討したい。また、トピック数を自動で決定する方法についても考えたい。

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, サージンロジー論文誌, 2-2 (2019), 1-12.
- [2] 大久保立樹, 室町泰徳, 旅行ガイドブックと口コミの言語解析による訪日外国人の観光地イメージに関する研究, 公益社団法人日本都市計画学会 都市計画論文集, 49-3 (2014), 573-578.
- [3] 松本直彰, 湯本高行, 山本岳洋, 大島裕明, 時間的推移を考慮した報道トピックの抽出と再構成, 情報処理学会研究報告, IFAT-139-1 (2020), 1-6.
- [4] David. M. B, and John. D. L, Dynamic Topic Models, *Proceedings of the 23rd international conference on Machine learning*, (2006), 113-120.
- [5] David. M. B, Andrew. N, and Michael. J, Latent Dirichlet Allocation, *Journal of Machine Learning Research*, 3 (2003), 993-1022.
- [6] 高橋佑介, 横本大輔, 宇津呂武仁, 吉岡真治, 河田容英, 神門典子, 福原知宏, 中川裕志, 清田陽司, 時系列トピックモデルにおけるバーストの同定, DEIM2012, F5-5 (2012).
- [7] Thomas. H, Probabilistic latent semantic analysis, *Proceedings of the 15th Conference Uncertainty in Artificial Intelligence*, (1999), 289-296.
- [8] Scott. D, Susan. T. D, George. W. F, Thomas. K. L and Richard. H, Indexing by Latent Semantic Analysis, *Journal of the American Society for Information Science*, 41-6 (1990), 391-407.