

観光レビューの季節分類のための季節キーワード抽出の検討

Examination of seasonal keyword extraction for season classification of tourism reviews

檜橋 昂

Subaru Narahashi

岡山大学 太田研究室

Ohta Laboratory, Okayama University

概要 観光レビューサイトの膨大なレビューから特定の季節のレビューを抽出することは一般的に困難である。そのため、本稿では、口コミを季節ごとに分類するための季節キーワードについて検討する。具体的には、fastTextを用いて単語の分散表現により季節キーワードの抽出を行う。また、その季節キーワードが含まれているレビューをその季節として分類する実験を行った。

1 はじめに

旅行の際に、観光地を決める手段の一つとして、じゃらんnet¹などの観光レビューサイトの口コミを参考にする人は多い。また、実際に観光地を決めるには、観光地までの距離や食事、景観などのいくつかの要因が考えられる。本研究では、その中でも季節ごとの観光地の特色や期間限定のイベントに関する情報に着目する。しかし、不特定多数の人が投稿するレビューは自由記述が多いため、その中から特定の観光地の季節ごとの特色に言及したレビューを探し出すのは一般的に困難である。そのため本研究では、季節分類をするための季節キーワードの抽出について検討する。また、これらのキーワードが含まれているか、否かで観光レビューの分類をする。

観光レビューの季節分類を行っている研究として、鳥山らは観光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した[1]。[1]では、まず、レビューを文に分割して、それぞれの文を「願望」、「評価」、「経験」、「その他」の4タイプに分類する。そして、「願望」以外の3タイプの文に対して、季節キーワードを用いて季節分類を行う。この季節キーワードは、各季節を代表する花や一般的な行事の名前で、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、出現回数が多い単語を用いた。これらの季節キーワードの例を以下に示す。

- 春(42語)：桜、3月、緑の日
- 夏(43語)：ひまわり、夏祭り
- 秋(48語)：月見、敬老の日
- 冬(41語)：雪、カイロ

本稿では、まず2節で季節分類を行うレビューについて説明する。3節では、本稿で行う季節キーワードの抽出とそれを用いた季節分類の実験について説明し、4節で季節キーワードの抽出結果と、それらを用いたレビューの季節分類の結果について示す。最後に、5節でまとめる。

2 季節分類を行うレビュー

本研究では、じゃらんnetの口コミのレビューで、岡山県の「倉敷美観地区」の「総合評価」、「タイトル」、「本文」、「行った時期」、「投稿日」の項目を収集した。収集件数は3,143件(2022年7月5日時点)であった。

次に、季節の特色のあるレビューであるか、季節に関係なく、その観光地に共通する内容であるかを著者が人手で判定して正解データを作成した。まず、それぞれの季節を、春は3月から5月、夏は6月から8月、秋は9月から11月、冬は12月から2月とする。それから、収集したデータのうち、「タイトル」と「本文」のみを読んでレビューに関係のある季節を「春」、「夏」、「秋」、「冬」、「その他」の5種類に重複を許してラベル付けを行った。表1に、正解データの例を示す。

表1:レビューとその季節の例

レビュー	季節
美観地区は、川の流れのようにゆったりとした落ち着いた町でした。ちょうどよく天領夏祭りがあり、最高でした。	夏
春や夏に来ると、川沿いに青々とした木々や桜の花が見えてきれいだと思います。	春, 夏
桃太郎博物館に行きました。最初は怪しい館だなと思いましたが、行ってみると1番面白い場所でした。ちくわ笛の演奏がシュールで素敵です!	その他

また、3,143件のレビューの季節の内訳は以下の通りとなった。

春:(152件), 夏:(98件), 秋:(29件), 冬:(42件), 春, 夏:(5件), 春, 秋:(10件), 春, 冬:(5件), 夏, 秋:(2件), 夏, 冬:(2件), 春, 夏, 冬:(1件), 春, 夏, 秋:(1件), 夏, 秋, 冬:(1件), 春, 夏, 秋, 冬:(11件), その他:(2784件)

3 季節キーワードの抽出とレビューの季節分類

3.1 季節キーワードの抽出

本研究では、レビューを季節分類する際に、季節キーワードを用いる。季節キーワードとして「春」、「夏」、「秋」、「冬」のそれぞれの単語に類似した単語を収集する。

そのため本稿では、fastText²の日本語学習済みモデルを用いて、類似する単語を取得する。

fastTextはWord2vecのライブラリをもとに、単語をベクトル化することで単語間の距離を計算して言葉の処理を可能にする自然言語処理向けアルゴリズムである。

本稿では、鳥山らの収集した季節キーワードを利用して新たな季節キーワードを抽出する。ここで、鳥山らのレビューの季節分類に用いる季節キーワードの一部を表2に示す。

¹ <https://www.jalan.net/>

² <https://fasttext.cc/>

表 2:鳥山らの季節キーワード[1]

季節	季節キーワードの例
春(42語)	桜, ゴールデンウィーク, 3月, 4月, 5月, つつじ, ひな祭り, 昭和の日, みどりの日
夏(43語)	ひまわり, かき氷, 6月, 7月, 8月, お盆, 七夕, 夏祭り
秋(48語)	コスモス, 月見, 9月, 10月, 11月, 敬老の日, 秋分の日, 七五三
冬(41語)	雪, クリスマス, 12月, 1月, 2月, カイロ, 大晦日, 新年

抽出方法としては、まず、鳥山らの季節キーワードのうち、下記の八つを参照単語とする。

表 3:季節キーワードの抽出に用いる参照単語

季節	参照単語
春	桜, ゴールデンウィーク
夏	かき氷, 七夕
秋	コスモス, ハロウィン
冬	クリスマス, 大晦日

次に、各参照単語からそれが属する季節を表す語を引いて、他の季節を足したベクトルを求める。この分散表現を季節キーワード候補を抽出するために用いる。例えば、参照単語である「桜」に対応する秋の季節キーワードは、「桜」-「春」+「秋」のように計算して求める。

すなわち、「桜」-「春」+「秋」を計算したベクトルと学習済みの全ての単語のベクトルとのコサイン類似度を求める。その類似度の上位3件の単語をこの場合は秋の季節キーワード候補として抽出する。

3.2 季節キーワードを用いたレビューの分類

3.1節で抽出した季節キーワードを用いてレビューの季節分類を行う。分類方法は、季節キーワードが含まれているレビューをその季節のレビューとする。ただし、同義語は対応していない。例えば、春の季節キーワードの一つの「ゴールデンウィーク」は「GW」のような言い回しも一般的にされるが、これは無視し、「ゴールデンウィーク」と記述のあるレビューのみ分類する。

4 季節キーワードの抽出方法とレビュー分類の結果

4.1 季節キーワードの抽出結果

本稿では、抽出した単語のうち上位3件を季節キーワード候補とする。また、そこから同義語は取り除き、抽出した季節キーワード候補が、複数の季節と重複した場合は一番コサイン類似度が高い季節にのみ振り分ける。

3.1節で説明した抽出方法により、得られた結果を表4に示す。表4について、桜に関しては秋と冬において、「紅葉」や「吹雪」のような期待するキーワードを抽出することができた。しかし、かき氷に関しては、冬に夏を連想させるアイスクリームやソフトクリーム抽出され、その季節に適当でない季節キーワードとなった。

また、参照単語の「コスモス」では、「ウルトラマン」や「ルナモード」が出力された。これらは、いずれもアニ

メ「ウルトラマンコスモス」に関連する単語である。よって、同音異義語があるとこのような問題が生じる。

また、表4の季節キーワードと表3の参照単語に春夏秋冬の各季節を合わせて、季節分類に利用するキーワードを定めた。これらを表5にまとめる。

表 4:抽出した季節キーワード

参照単語 (属する季節)	季節	季節キーワード(コサイン類似度)
桜 (春)	夏	並木(0.484), 木の下(0.481)
	秋	紅葉(0.609), 八重桜(0.573), 枝垂れ桜(0.546)
	冬	吹雪(0.528), 千本(0.510), ぴかぴかよう(0.492)
ゴールデンウィーク (春)	夏	夏休み(0.653), 連休(0.586), 海の日(0.537)
	秋	お盆(0.654), 週末(0.537)
	冬	年末年始(0.593), 春休み(0.525), お正月(0.501)
かき氷 (夏)	春	パンケーキ(0.496)
	秋	ジェラート(0.573), パフェ(0.519), たこ焼き(0.509)
	冬	アイスクリーム(0.601), ソフトクリーム(0.565), ぜんざい(0.531)
七夕 (夏)	春	ひな祭り(0.555), 爛漫(0.476), 桜花(0.475)
	秋	サンデーセイラ(0.521), 端午(0.514), 十五夜(0.513)
	冬	クリスマス(0.537), 冬至(0.504), 節分(0.503)
ハロウィン (秋)	春	イースター(0.560)
	夏	仮装(0.540)
	冬	クリスマス(0.638), サンタクロース(0.525),
コスモス (秋)	春	チューリップ(0.507), パンジー(0.486), ポピー(0.485)
	夏	ひまわり(0.501), ルナモード(0.469), ウルトラマン(0.422)
	冬	スノードロップ(0.426), ポインセチア(0.405), サボテン(0.403)
クリスマス (冬)	春	ひと足早い(0.590), クリスマスイヴイヴ(0.556)
	夏	クリスマスイブ(0.612), バレンタイン(0.568), パースデー(0.549)
	秋	ハロウィン(0.606)
大晦日 (冬)	春	元日(0.590), 年明け(0.562), 年の瀬(0.557)
	夏	大みそか(0.668), 年越し(0.540)
	秋	年末(0.628), 年の暮れ(0.522)

表 5:季節分類に用いる季節キーワード

季節	季節キーワード
春	春, 桜, ゴールデンウィーク, パンケーキ, ひな祭り, 爛漫, 桜花, イースター, チューリップ, パンジー, ポピー, ひと足早い, クリスマスイヴイヴ, 元日, 年明け, 年の瀬
夏	夏, かき氷, 七夕, 並木, 木の下, 夏休み, 連休, 海の日, 仮装, ひまわり, ルナモード, ウルトラマン, クリスマスイブ, バレンタイン, パースデー, 大晦日, 年越し
秋	秋, コスモス, ハロウィン, 紅葉, 八重桜, 枝垂れ桜, ジェラート, パフェ, たこ焼き, サンデーセイラ, 端午, 十五夜, 年末, 年の暮れ
冬	冬, クリスマス, 大晦日, 吹雪, 千本, ぴかんぴょう, 年末年始, 春休み, お正月, アイスクリーム, ソフトクリーム, ぜんざい, クリスマス, 冬至, クリスマス, サンタクロース, スノードロップ, ポインセチア, サボテン

4.2 レビューの季節分類の結果

季節キーワードが含まれるかどうかで「倉敷美観地区」のレビューの季節分類を行った。まず、各季節キーワードのレビュー件数を表 6 に示す。なお、「春, 夏」のレビューには、表 5 の春の季節キーワードと夏のそれが両方含まれる。

表 6:季節キーワードを含むレビュー件数

季節	レビュー件数
春	117 件
夏	65 件
秋	145 件
冬	69 件
春, 夏	6 件

次に、レビューを人手で判別した正解データと比較する。分類した正誤判定の結果を表 7 から表 10 に示す。また、表 11 に「春, 夏」に分類されたレビューの内訳を示す。

マルチラベルは「春, 秋」などの季節とする。それから、()内の数字は正解の季節を含むマルチベクトルとし、その数字は内数とする。

表 7:「春」に分類されたレビューの内訳

正	誤				
春	夏	秋	冬	その他	マルチラベル
105 件 (13 件)	0 件	0 件	1 件	11 件	0 件

表 8:季節分類の「夏」の正誤判定の内訳

正	誤				
夏	春	秋	冬	その他	マルチラベル
16 件 (0 件)	8 件	1 件	2 件	37 件	1 件

表 9:季節分類の「秋」の正誤判定の内訳

正	誤				
秋	春	夏	冬	その他	マルチラベル
39 件 (13 件)	2 件	19 件	4 件	70 件	11 件

表 10:季節分類の「冬」の正誤判定の内訳

正	誤				
冬	春	夏	秋	その他	マルチラベル
33 件 (13 件)	6 件	3 件	1 件	26 件	0 件

表 11:季節分類の「春, 夏」の内訳

	春	春, 秋	その他
春, 夏	4 件	1 件	1 件

季節分類の結果では、春は正解のレビューが多かった。正解レビューの 105 件のうち、104 件が「春」, 「桜」, 「ゴールデンウィーク」のいずれかの季節キーワードが含まれるレビューであった。しかし、これらの季節キーワードは参照単語と季節名であり、抽出した季節キーワードを含むレビューは「ひな祭り」の一件しかなかった。

一方、秋に関しては正解と判定された 39 件のレビューのうち 29 件に「紅葉」が含まれていた。これは 3.1 節の方法で抽出した季節キーワードである。

今後の課題としては、本稿で判定することができなかった同義語や表記ゆれへの対応が必要である。また、本稿では、季節キーワードは名詞がほとんどであったが、「寒い」などの形容詞や動詞などの活用のある単語についての対応も必要である。

5 まとめ

本稿では、観光サイトのレビューから季節分類を行う際の季節キーワードの抽出を行った。まず、じゃらん net で倉敷美観地区において、人手で季節の判別を行った。それから、季節キーワードの抽出を fastText で分散表現によって、鳥山らの定義した季節キーワードを利用して行った。また、抽出した季節キーワードを用いて、それらが含まれているかどうかで倉敷美観地区のレビューの季節分類を行った。

今後の課題としては、表記ゆれの対応が必要である。また、動詞、形容詞の季節キーワードを扱う場合は、様々な活用形にも対応させる必要がある。

6 参考文献

[1] 鳥山実桜, 灘本明代, “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,” DEIM Forum 2022, 2022, B43-1, 2022.