観光ビッグデータを活用した テキストマイニングによる 観光スポットの改善点抽出に関する分析

馬場　優大[1]・藤生　慎[2]・森崎　裕磨[3]

[1]学生会員　金沢大学　理工学域地球社会基盤学類（〒920-1192 石川県金沢市角間町）
E-mail: asmr4ub2ttt@stu.kanazawa-u.ac.jp

[2]正会員　金沢大学　融合研究域融合科学系（〒920-1192 石川県金沢市角間町）
E-mail: fujiu@se. kanazawa-u. ac. jp (Corresponding Author)

[3]正会員　金沢大学特任助教　融合研究域融合科学系（〒920-1192 石川県金沢市角間町）
E-mail: morisaki@staff. kanazawa-u. ac. jp

我が国において，人口減少と少子高齢化が進行する地方では，地域の機能をどのように維持し，活性化させていくかが課題となっている．こうした中，観光産業はこれらを解決する施策として注目されており，観光産業を発展させていくためには多様化する観光ニーズに基づいた施策の立案・実施が重要である．そこで本研究では，近年のインターネットの普及に伴い蓄積された観光ビッグデータの内，旅行情報サイトの観光スポットに関するロコミデータを収集し，深層学習を用いたセンチメント分析を行うことでロコミから観光客のネガティブな感情が伺える文の抽出し，抽出したネガティブな文でテキストマイニングを行うことで観光スポットの改善点の把握を行った．

Key Words: *tourism,big-data,deep learning,text mining,*

1. 本研究の背景と目的

(1) 本研究の背景

我が国において，人口減少と少子高齢化が進行する地方では，地域の機能をどのように維持し，活性化させていくかが課題となっている [1]．こうした中，観光産業はこれらを解決する施策の 1 つとして注目されており [2]，現在，地方自治体が積極的に取り組んでいる．この中でも，ブランド総合研究所が約 3 万人を対象に調査を実施する「都道府県観光意欲度ランキング」において，結果の確認できる 2017 年から 2021 年にかけて 47 都道府県中，毎年 10 位(最高 6 位)に選出されている [3]，国内観光客からの需要が高い地方である石川県においては，2016 年にほっと石川観光プラン 2016 が地方自治体によって策定され，北陸新幹線の金沢・敦賀開業や観光ニーズの多様化といった観光を取り巻く環境変化を見据えた施策が展開されている[4].

この観光産業をより効果的に発展させるためには，観光ニーズに基づいた施策を実施・立案することが重要である [5]．こうした中，近年のインターネットの普及に伴い旅行情報サイトやソーシャルネットワークサービス(以下 SNS と呼ぶ)と呼ばれる新しいコミュニケーションツールが発達し，人々が観光地での体験談をインターネット上に自ら発信するようになった．この人々の観光地での体験談が観光客特性や多様化する観光ニーズの把握に利用できる観光ビッグデータとして蓄積されており，実際に観光庁は多様化する観光ニーズの把握に，この観光ビッグデータを活用している [6]．さらに近年の言語処理技術の発展により，文章から伺える感情を自動的に推定する精度が飛躍的に上昇している [7]．これにより，インターネット上に発信された観光地での体験談であるテキスト情報から観光客の感情を推定することが可能となり，より詳細に多様化する観光ニーズの把握が可能となった.

(2) 本研究の目的

前節で述べた観光ビッグデータの内，旅行情報サイトに投稿された観光地に対するロコミは観光地での体験談そのものであり，この体験談には観光客が観光地に対して抱いた感情，観光地での行動が伺える．したがって，

本研究では旅行情報サイトに投稿されたロコミを用いて石川県の主要観光スポットである兼六園，近江町市場，ひがし茶屋街，金沢 21 世紀美術館の 4 箇所を対象として観光ニーズに含まれる各観光スポットの改善点の把握を行うことを目的とする．

2. 既往研究の整理と本研究の目的

本章では，旅行情報サイトに投稿された観光地に対するロコミデータを分析し，得られた 結果をもとに観光ニーズの把握を行っている既往研究の整理と，それらを踏まえた上での本研究 の位置付けを述べる．

(1) 既往研究の整理

野守ら [8]は，旅行ロコミサイト「フォートラベル」における国内旅行者の全国の観光地に対するロコミデータのうち，タイトルとコメント本文を紐づけたテキストデータから，名詞と形容詞，形容動詞及び名詞と動詞などの係り受け表現を使用データとして抽出し，確率的潜在意味解析(PLSA Probabilistic Latent Semantic Analysis)と呼ばれる文書分類の次元縮約手法を用いた分析を行っている．その結果，観光地と係り受け表現の背後に潜在する観光テーマ(動物園や水族館で家族と楽しむなど)を抽出し，そのテーマ内で出現した観光地名(沖縄美ら海水族館)や係り受け表現の特徴(大人＋楽しむ＋できる)を確率的に示すことに成功している．杉本ら [9]は，事前に 8 種類の感情(喜び，悲しみ，受容，嫌悪，恐れ，怒り，驚き，期待)と，それに類する表現(「喜び」に対して嬉しい，楽しい，良いなど)とを紐づけた感情語辞書を作成し，観光 Web サイトの「トリップアドバイザー」に投稿されたロコミに対して上述の感情語辞書のパターンマッチングを行うことで，都道府県別に各種感情表現の出現する度合いについて解析している．その結果，都道府県によって感情表現の出現度合いの傾向にはあまり差がないことが示唆されているが，異なるロコミサイトを使用して同様の分析を行った場合，感情表現の出現度合いに差があることが示されている．竹岡ら [10]は，旅行情報サイトのじゃらん net 上に掲載されている水族館に対するロコミからテキストマイニングによって抽出された消費者の体験と，施設の来場者数，延床面積などの外形的データとの相関関係を分析している．その結果，飼育種類数の多い水族館ほど「面白い」と投稿するロコミが多い傾向を示すなど，消費者の体験の価値を高める要素を持つ施設の特徴を明らかにしている．後藤ら [11]は，じゃらん net に投稿されたロコミから，観光客の観光テーマ(観光動機や目的)や観光スタイル(観光地でどのように過ごすか)の特徴を次元縮約クラスタリング手法の一つである線形判別分析(LDA:Linear Discriminant Analysis)を用いて抽出し，さらにそれらと選択観光地の関係性について因果探索手法のベイジアンネットワーク 7)を適用している．その結果，例えば函館山を訪問した人は観光テーマとして「夜景」を持つ確率が高く，また「夜景」をテーマにする人は「計画的」な観光スタイルを有する確率が高いことを示し，個人の観光テーマや観光スタイルに即した観光地を推薦出来ることを証明している．

(2) 既往研究を踏まえた上での本研究の位置付け

既往研究では，旅行情報サイトに投稿されたロコミのテキストデータに対し，施設の外形的データを紐づけた分析や，感情語辞書のパターンマッチングを行った分析等，様々な手法で観光地評価が行われている．しかし，いずれの既往研究においても，ロコミに対して文章の内容がポジティブ，ニュートラル，ネガティブのいずれかを判定といった感情極性推定を行い，その内のネガティブな感情が伺える文を抽出し，これを用いて観光スポットの改善点の把握を行っている既往研究は存在しない．これが本研究の新規性である．

3. ロコミデータの概要

(1) 利用する旅行情報サイトの概要

本研究では，日本観光復興協会が公開している 2020 年，2021 年における日本国内の観光関連サイト閲覧数ランキング [12]において，2 年連続で PC・スマートフォンからの閲覧数が 1 位であった，じゃらん net[13]をデータの収集先として利用した．じゃらん net は株式会社リクルートが管理，運営する宿泊予約サイトであり，20,000 軒以上に上る宿泊施設の情報が掲載されている．また同サイトには観光スポットに関する情報も掲載されており，その地点数は 2023 年 1 月 21 日時点で 107,679 箇所とされている．

(2) ロコミデータの収集方法

本研究ではロコミの収集にウェブスクレイピングを利用した．ウェブスクレイピングとはウェブサイトから情報を自動的に抽出する技術の一つである．今回は HTML 取得に urllib，HTML の解析に BeautifulSoup という Python のライブ ラリを用いてウェブスクレイピングを行った．

(3) 本研究において収集したロコミデータの概要

前節で述べたスクレイピングを行うことによって，2010 年 9 月 1 日から 2022 年 8 月 31 日 の 12 年間に本研究で分析対象とする 4 つの観光スポットに対して投稿された，ロコミデータを収集しデータベース化した．収集した各観光スポットに対して投稿されたロコミ件数は兼六

園は 3704 件，近江町市場は 2752 件，ひがし茶屋街は 2265 件，金沢 21 世紀美術館は 1946 件であった．このロコミデータを基に石川県の主要観光スポット 4 つの改善点抽出を行った．

4. ロコミの感情極性推定モデルの構築とネガティブな感情が伺える文の抽出

本研究では収集した大量のロコミから，ネガティブな感情を自動抽出する機械学習モデルを構築し，構築したモデルを用いて収集した各観光スポットに対して投稿されたロコミからネガティブな感情が伺える文を抽出を行う．

(1) 感情極性推定モデルの構築

本研究では，構築する感情極性推定モデルに用いる訓練データのラベルである，ポジティブ，ネガティブ，ニュートラルの三種類の感情を扱う．感情極性推定モデルの種類においては，2018 年に Google により提案された自然言語における深層学習のモデルである BERT を用いる[14]．BERT の特徴として，従来の深層学習モデルと比べて学習に必要な訓練データ数が少なくて 済むこと，任意の単語の前後両方の文脈を考慮して学習できることが挙げられる．BERT は 感情極性辞書による手法と比べて深層学習による手法が苦手な点を克服しており，長所を強化している．村田[15]は，BERT と感情極性辞書を用いる手法による，センチメント分析の 精度を比較し，訓練データがあれば，センチメント分析には BERT を利用することを推奨し ている．これらの理由より，筆者らは BERT を用いてテキストの感情極性推定モデルを構築 し，ロコミからネガティブな文の抽出を行う．

モデルには，現在日本語のセンチメント分析おける事前学習モ デルとして，広く使われている，東北大乾研が日本語 Wikipedia により事前学習を行なった モデル[16]を扱える，huggingface 社が提供している BertForSequense-Clasification[17]を使用した．訓練データには，文に感情極性がラベル付けされており，BERT の訓練データとして扱える オープンデータである，じゃらん net に投稿された日本語の口コミ特性として最も類似して いると考えられる Japanese Realistic Textual Entailment Corpus[18]を利用した．このデータはじ ゃらん net に投稿された日本語の口コミに感情極性ラベルを付与しており，オープンデータの中で本研究が対象とするロコミデータと特性が最も類似していると考えられる．収集さ れているデータは，クチコミデータから抽出した 5553 件の文をアノテーション作業者がポ ジティブ，ネガティブ，ニュートラルの三つのうちいずれかの判定ラベルを付与したものである．ポジティブが 1，ネガティブが 1，ニュポートラルが 0 に対応 している．文がどの判定ラベルに対応するかはアノテーションを行った 3 名による多数決 によって決定している．データセットのラベル別に集計した結果を**図-1** に示す．

図-1 からラベル毎のデータ数に偏りがあることがわかる．したがって，深層学習モデルの 訓練データとして，ラベル毎のデータ数に著しい偏りがある場合，推定結果に悪影響を及ぼす可能性がある．本研究では訓練データのラベル数をデータ数が最も少ないネガティブとラベル付けされている文数に合わせて各ラベル数を統一した．それらを 3:1 の割合で分割した 1,839 件の文を学習データ，615 件の文を訓練データとしてモデルの構築に利用した．

モデルの性能を評価するために，混合行列を作成し，ネガティブ，ポジティブ，ニュート ラルの 3 つの感情極性について，モデルが正しく分類した件数，誤って分類した件数を集計する．作成した混合行列を**図-2** に示す．**図-2** から，ネガティブな文に対してモデルが正しくネガティブと予測した文数は 177 文，誤ってニュートラルと予測した文数は 23 文，誤ってポジティブと予測した文は 5 文，ニュートラルな文に対して，モデルが誤ってネガティブと予測した文数は 24 文，誤ってポジテ ィブと予測した文数は 9 文であった．また本研究では，ロコミからネガティブな文を抽出することを目的としてセンチメント分析を行うモデルを構築したといった理由から，ポジテ ィブ，ネガティブ，ニュートラルの内，ネガティブな感情が伺える文に着目し，精度の評価を行った．具体的には，正解ラベルがネガティブである文に対して，モデルが感情極性がネガティブと予測した文の数を真陽性(TP：True-Positive)，正解ラベルがネガティブ以外である文に対して，モデルが感情極性がネガティブ以外と推定した文の数を真陰性(TN：True-Negative)，正解ラベルがネガティブ以外である文に対して，モデルが感情極性がネガティブと予測 した文の数を偽陽性(FP：False-positive)，正解ラベルがネガティブである文に対して，モデルが感情極性がネガティブ以外と推定した文の数を偽陰性 (FN：False-Negative)を集 計した．それらをもとに各モデルにおける分類性能の評価指標として，再現率（Recall），適 合率（Precision）及び F 値（F1-score）を算出した．各評価指標の算出方法を式(1)，式(2)， 式(3)に示す．

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{1}$$

$$Precision = \frac{TN}{\mathrm{TN} + \mathrm{FP}} \quad (2)$$

$$F1 - score = \frac{2 \times Precision \times Recall}{precision + Recall} \quad (3)$$

本章における適合率（Precision）は，モデルがネガティブと予測した文のうち，実際の感 情極性がネガティブとラベル付けされていた文の割合を示している．また再現率（Recall） は，感情極性がネガティブとラベル付けされている文の内，モデルがネガティブであると正 しく予測した割合を示している．F 値（F1-score）は再現率と適合率の 2 値のバランスを示す数値である．構築したモデルの再現率，適合率，F 値を表-1 に示す．適合率は 0.84，再現率は 0.86，F 値は 0.85 であった．これは，モデルがネガティブと推定した文の内，正解ラ ベルがネガティブだった割合が約 84%であり，正解ラベルがネガティブな文の内，モデル がネガティブと推定した文の割合が約 86%であることがわかる．本研究ではこのモデルを用いて口コミからネガティブな文の抽出を行った．

(2) ネガティブな感情が伺える文の抽出

本研究では，兼六園，近江町市場，ひがし茶屋街，金沢 21 世紀美術館の 4 つの観光スポ ットに対して投稿された口コミを文に分割し，構築したモデルを用いてセンチメント分析 を行い，ネガティブな感情が伺えると判定された文を抽出した．各観光スポットに対する口 コミの件数，分割した文数，モデルがネガティブな感情が伺えると判定された文の数を表-2 に示す．兼六園において，口コミ件数は 3,704 件，文数は 10,673 文，その内モデルによ ってネガティブな感情が伺えると判定された文は 1,701 文となった．近江町市場において， 口コミ件数は 2,752 件，文数は 7,985 文，その内モデルによってネガティブな感情が伺える と判定された文は 1,474 文となった，ひがし茶屋街において，口コミ件数は 2265 件，文数 は 6,363 文，その内モデルによってネガティブな感情が伺えると判定された文は 1,062 文と なった．金沢 21 世紀美術館において，口コミ件数は 1,946 件，文数は 5926 文，その内モデ ルによってネガティブな感情が伺えると判定された文は 1,647 文となった．

5. テキストマイニングによる観光スポットの改善点抽出

(1) 共起ネットワークの構築による観光スポットに対するネガティブな感情の抽出

本研究では，KHCoder というテキストマイニングツールを用いて，5 章で構築した感情極性推定モデルによって抽出されたネガティブな口コミに含ま れる名詞，形容詞，形容動詞，動詞から単語の出現数と単語同士の関連性の強さを示す値である Jaccard 係数を利用し， 共起ネットワークを構築，構築した共起ネットワークにおいてサブグラフ検出(modularity) を行うことで，単語同士が比較的に強く結びついている部分を自動的に検出してク

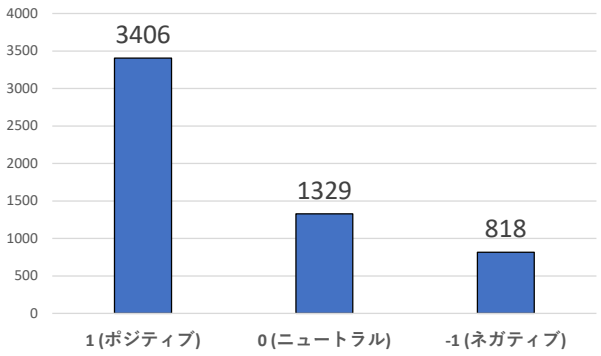

図-1 訓練データのラベル別集計結果

		予測値		
		ネガティブ	ニュートラル	ポジティブ
正解値	ネガティブ	177	23	5
	ニュートラル	24	153	28
	ポジティブ	9	16	180

図-2 テストデータの正解値とモデルの予測値で作成した混合行列

表-1 モデルの推定精度

推定精度の評価指標	値
再現率(Recall)	0.86
適合率(Precision)	0.84
F値(F1-score)	0.85

表-2 訓練データのラベル別集計結果

観光スポット名	文数	抽出したネガティブな文数
兼六園	10673	1701
近江町市場	7985	1474
ひがし茶屋街	6363	1062
金沢21世紀美術館	5926	1647

ラスター分けを行い，クラスターごとに色分け，ナンバリングを行った．なお，Jaccard 係数とは任意の 2 つの単語において，どちらか 1 つの単語を含む文の数に対する，2 つの単語を含む文の割合である． Jaccard 係数の定義としては，任意の単語 2 つにおいて，どちらか一方の単語を含む文数に 占める，2 つの単語を含む文数の割合である．この任意の単語 2 つにおける Jaccard 係数が 大きいほど，単語同士の関連性が強く，小さいほど単語同士の関連性が弱い．任意の単語 x を含む文の集合を X，任意の単語 y を含む文の集合を Y としたとき，x と y の jaccard 係数 を式(4)に示す．

$$jaccard\,係数 = \frac{X \cap Y}{X \cup Y} \tag{4}$$

また，共起ネットワークの各クラスターを構成する単語の出現回数を集計した．各観光スポットに対して投稿されたロコミから抽出されたネガティブな文で構築し，サブグラフ検出(modularity)によってクラスター分け及びナンバリング，色分けを行ったものをそれぞれ**図-3**，**図-4**，**図-5**，**図-6** に示す．

(2) 共起ネットワークのクラスターの解釈

共起ネットワークの各クラスターを構成する単語から伺える内容の感情（ネガティブ，ポジティブ，ニュートラルのいずれか）を筆者らが推定し，ネガティブと推定したクラスターに対して，単語から伺えるネガティブな感情の詳細な内容を筆者らが解釈した．各観光スポットのネガティブな文で構築した共起ネットワークのクラスターの解釈をそれぞれ**表-3**，**表-4**，**表-5**，**表-6** に示す．

兼六園について，クラスター 2 は「庭園が広く，園内を歩いて回るのが大変，特にベビーカーでは砂利道や坂が大変」といった移動に関するネガティブな感情が解釈できた．クラスター3 は「写真を撮っている人や外国人が多くて混雑している」といった人の混雑に関するネガティブな感情が解釈できた．クラスター4 は「雪や雪吊りを見ることを期待していたが，見ることが出来なくて残念，雨が降っていて残念」といった，天候に対するネガティブな感情が解釈できた．クラスター7 は「駐車場やバスが混んでいる」といった交通の混雑に関するネガティブな感情が解釈できた．クラスター8 から伺える内容の感情極性がネガティブなものであったが何に対するネガティブな感情であるのかは不明であった．

近江町市場について，クラスター1 は「店が閉まるのが早い，お店や美味しい海鮮丼を食べるのに並ぶ」といった店，閉店時間，行列に関するネガティブな感情が解釈できる．クラスター4 は「お土産を買うのを迷う」といった店に対するネガティブな感情が解釈できる．クラスター5 は「海鮮丼を購入するのに行列が出来る」といった店，行列に対するネガティブな感情が解釈できる．クラスター6 は「飲食店に列ができて混雑している」といった店，行列，混雑に対するネガティブな感情が解釈できる．クラスター7 は「駐車場が狭い」といった駐車場に対するネガティブな感情が解釈できる．クラスター9 は「歩くのが大変で混む」といった移動，混雑に対するネガティブな感情が解釈できる．クラスター10 は「店の値段が高い」といった，店の値段に対するネガティブな感情が解釈できる．

ひがし茶屋街について，クラスター2 は「お店が閉まるのが早く入れなくて残念」といった店，閉店時間に関してのネガティブな感情が解釈できる．クラスター3 は「車で行ったが駐車場が少なく探した」といった，駐車場に対してのネガティブな感情が解釈できる．クラスター4 は「金沢に訪れた雨で大変」といった天候に対するネガティブな感情が解釈できる．クラスター8 は「写真を撮る人や観光客が多くて残念」といった人の混雑に対するネガティブな感情が解釈できる．クラスター9 は感情極性がネガティブと分類できるがネガティブな感情の対象が解釈困難であった．

金沢 21 世紀美術館について，クラスター1 が「有名なプールの展示は人が多くて見れず残念」といった期待外れといったネガティブな感情が解釈できる．クラスター2 は「チケットを購入するのに行列ができていて入れない」といった行列に対するネガティブな感情が解釈できる．クラスター3 は「作品に近づいたり，触れるとスタッフが注意してくる」といったスタッフに対するネガティブな感情が解釈できる．クラスター6 は「スタッフの感じや対応が悪い」といったスタッフに対するネガティブな感情が解釈できる．クラスター10 は「トイレが少ない」といった設備に対するネガティブな感情が解釈できる．

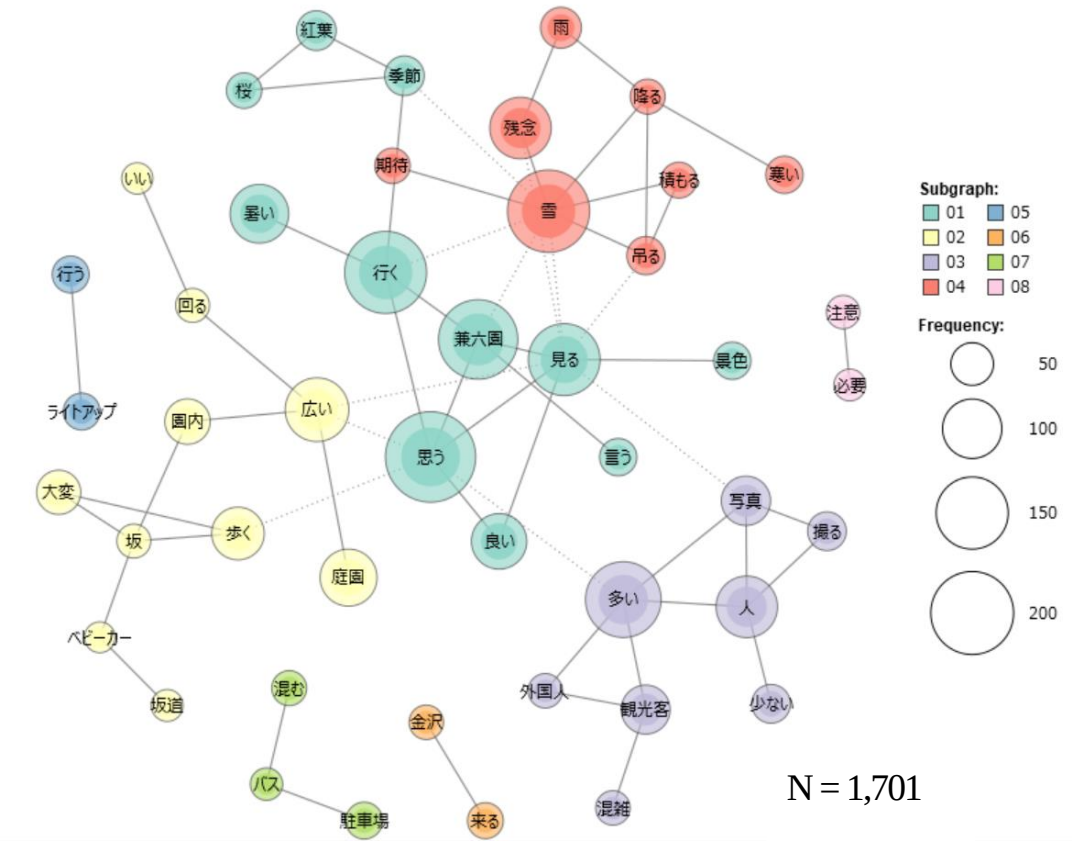

図-3　兼六園の共起ネットワーク

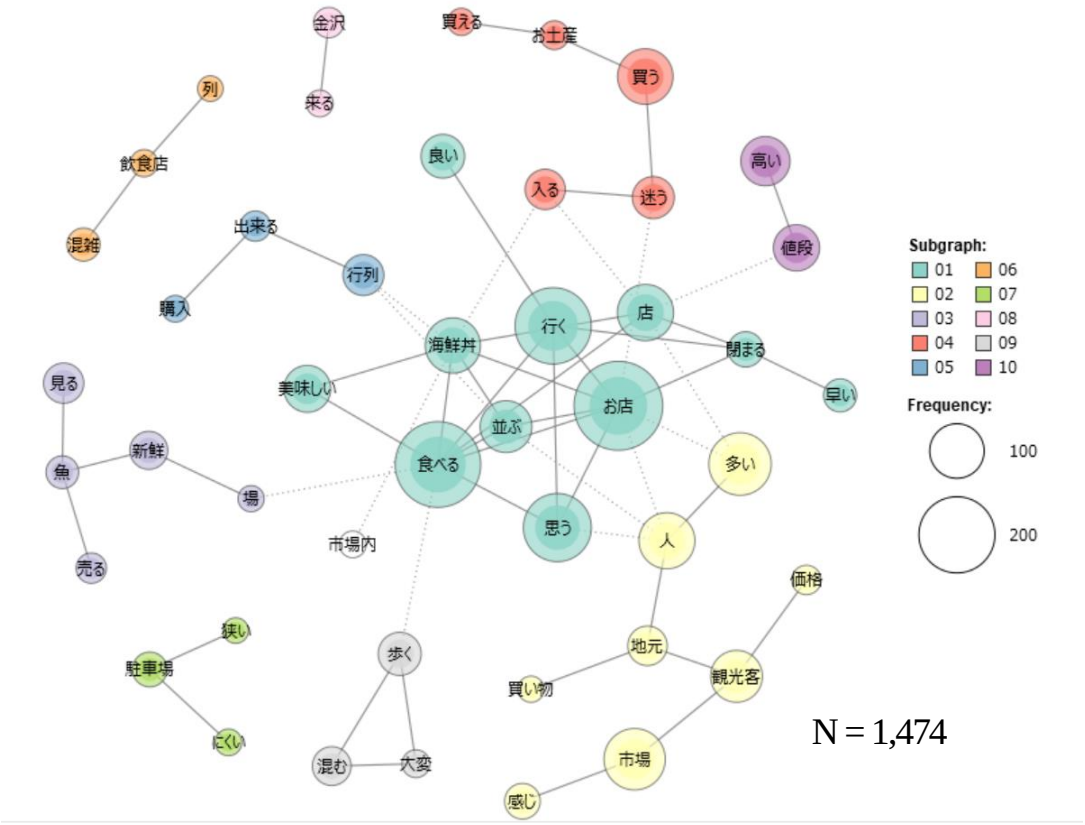

図-4　近江町市場の共起ネットワーク

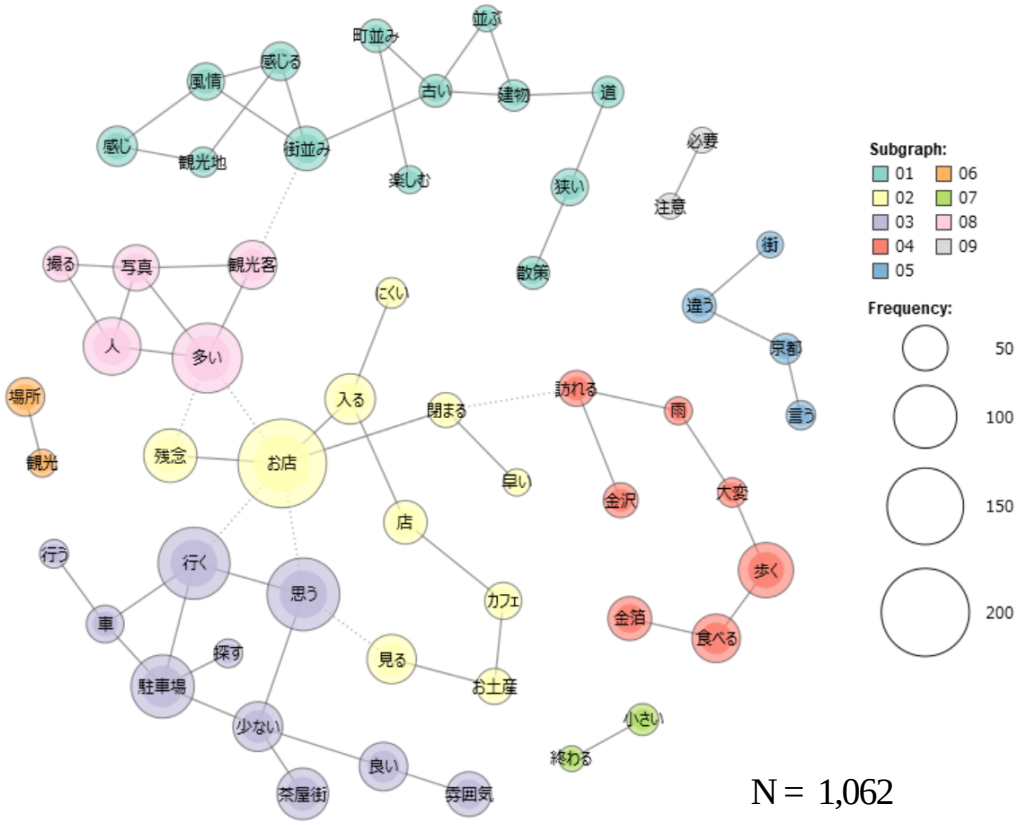

図-5　ひがし茶屋街の共起ネットワーク

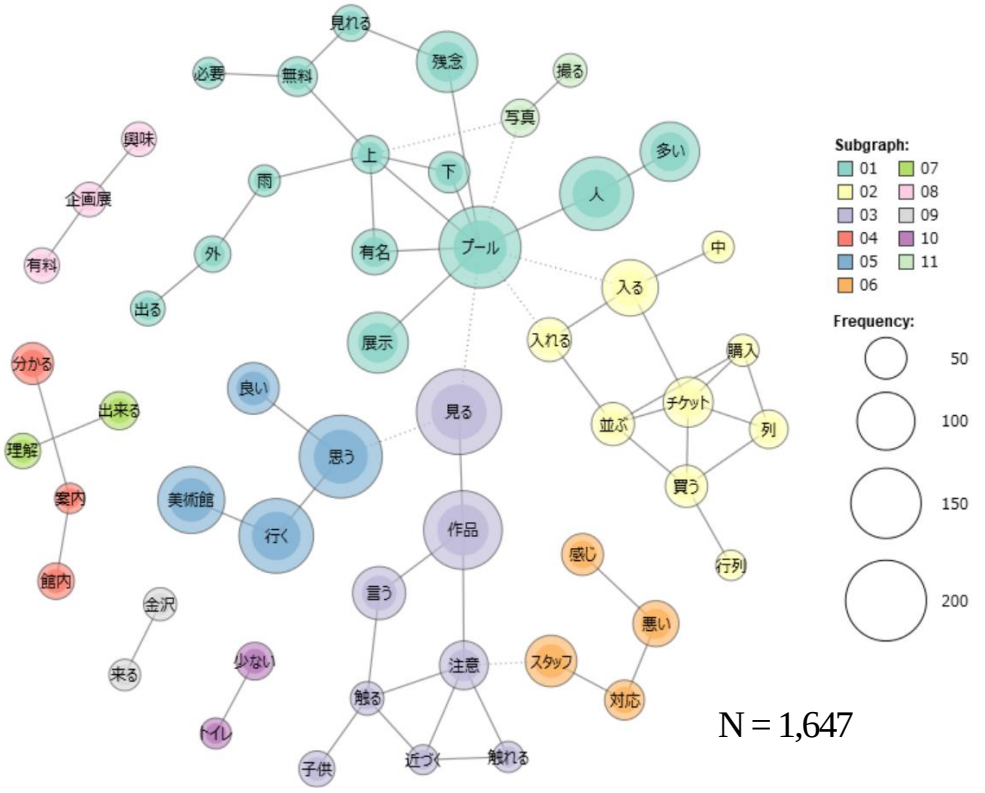

図-6　金沢 21 世紀美術館の共起ネットワーク

表-3　兼六園のクラスターの解釈

クラスター番号	伺える内容	感情極性
1	不明	-
2	庭園が広いので園内を歩いて回るのが大変, 特にベビーカーでは砂利道や坂が大変	ネガティブ
3	写真を撮っている人や外国人が多くて混雑している	ネガティブ
4	雪や雪吊りを見ることを期待していたけど見れない，雨が降っていて残念	ネガティブ
5	ライトアップを行う	ニュートラル
6	金沢に来る	ニュートラル
7	駐車場やバスが混んでいる	ネガティブ
8	注意が必要	ネガティブ

表-4　近江町市場のクラスターの解釈

クラスター番号	伺える内容	感情極性
1	狭い道を散策したり,古い建物が並ぶ街並みの風情を感じて楽しむ	ポジティブ
2	お店が閉まるのが早くて入れなくて残念	ネガティブ
3	車で行ったが駐車場が少なく探した	ネガティブ
4	金沢に訪れたが雨で大変	ネガティブ
5	街が京都と違うと言う	ニュートラル
6	観光場所	ニュートラル
7	不明	-
8	写真を撮っている人や観光客が多くて残念	ネガティブ
9	注意が必要	ネガティブ

表-5　ひがし茶屋街のクラスターの解釈

クラスター番号	解釈	感情極性
1	店が閉まるのが早い，お店や美味しい海鮮丼を食べるのに並ぶ	ネガティブ
2	観光客や地元の買い物をしている人が多い	ニュートラル
3	新鮮な魚が売られていたり見れる場	ニュートラル
4	お土産を買うのを迷う	ネガティブ
5	海鮮丼を購入するのに行列が出来る	ネガティブ
6	飲食店に列ができて混雑している	ネガティブ
7	駐車場が狭い	ネガティブ
8	金沢に来る	ニュートラル
9	歩くのが大変で混む	ネガティブ
10	店の値段が高い	ネガティブ

表-6　金沢 21 世紀美術館のクラスターの解釈

クラスター番号	伺える内容	感情極性
1	有名なプールの展示が人が多くて見れなくて残念	ネガティブ
2	チケットを購入するのに行列ができていて入れない	ネガティブ
3	作品に近づいたり，触れるとスタッフが注意してくる	ネガティブ
4	館内，園内がわかる	ニュートラル
5	美術館に行くといいと思う	ポジティブ
6	スタッフの感じや対応が悪い	ネガティブ
7	理解できる	ニュートラル
8	有料の企画展に興味がある	ニュートラル
9	金沢に来る	ニュートラル
10	トイレが少ない	ネガティブ
11	写真を撮る	ニュートラル

7. まとめと今後の課題

本研究では，観光ビッグデータの内，旅行情報サイトに投稿された口コミを収集し，それを用いてセンチメント分析及び共起ネットワークを構築することで，観光スポットの改善点抽出を行った．その結果，兼六園については，移動，人の混雑，天候，交通の混雑に関する改善点，近江町市場については閉店時間，店，行列，駐車場，値段に関する改善点，ひがし茶屋街については店，閉店時間，駐車場，天候，人の混雑に関する改善点，金沢 21 世紀美術館は期待外れ，行列，スタッフ，設備に対する改善点が抽出できた．

今後の課題として，6 章で構築した，共起ネットワークのクラスターにおいて，感情極性がネガティブと分類したが対象が解釈困難なクラスターについて，分析を行いどのようなネガティブな感情が発生しているのかを把握する必要性がある．

REFERENCES

1) 高齢化の現状と将来像｜令和 4 年版高齢社会白書（全体版） - 内閣府 (cao.go.jp) https://www8.cao.go.jp/kourei/whitepaper/w-2022/html/zen-bun/s1_1_1.html,
2) 笠木 秀樹：地域資源と観光 観光振興による地方創生, 兵庫教育大学地理学研究室研究報告 2019 年,24 巻 p.53-66
3) ブランド総合研究所 https://www.tiiki.jp/index.php（最終閲覧日：, 2023 年 1 月 21 日）
4) ほっと石川観光プラン https://www.pref.ishikawa.lg.jp/kankou/documents/plan2016.pdf,
5) 観光地域づくり法人(DMO)による観光地域マーケティングガイドブック http://www.mlit.go.jp/kankocho/content/001580600.pdf,
6) 平成２７年度 ICT を活用した訪日外国人観光動態調査 事業実施報告書（概要） https://www.mlit.go.jp/common/001158956.pdf,
7) BERT（バート）とは？次世代の自然言語処理の凄さやできること・書籍を紹介 | AI 専門ニュースメディア AINOW https://ainow.ai/2019/05/08/166723/
8) 野守 耕爾，神津 友武，「口コミデータに PLSA を適用した観光客目線による観光地分析」，人工知能学会全国大会論文集，第 29 号，pp.1-4 (2015)
9) 杉本 祐介，水野 忠則，「口コミに含まれる感情語を利用した観光地分類の検討」，マルチメディア・分散・協調とモバイル(DICOMO2014)シンポジウム, pp.1345-1350 (2014)
10) 竹岡志郎，「機械学習を活用したテキストマイニング-外形的データを併用することによる特徴分析-」，経営学論集第 89 集 自由論題，pp.1-7(2018)
11) 後藤孝輔，大野高裕，川中孝章，枝川義邦，「口コミデータを用いた観光スタイルと観光行動の関係分析」，第 67 回日本経営システム学会全国研究発表大会(2021)
12) 【調査リリース】2021 年観光関連サイト閲覧者数ランキング https://www.nihon-kankou.or.jp/home/userfiles/files/autoupload/2022/02/1643812044.pdf
13) 宿・ホテル予約 - 旅行ならじゃらん net (jalan.net) https://www.jalan.net/,（最終閲覧日：2023 年 1 月 21 日）
14) Devlin, Jacob, et al. “Bert: Pre-training of deep bidirectional transformers

for language understanding." arXiv preprint arXiv:1810.04805, 2018.（最終閲覧日：2023年1月25日）
15) 村田　龍也:BERTによる感情分析，http://nalab.mind.meiji.ac.jp/2020/2021-murata.pdf（最終閲覧日：2023年1月25日）
16) https://github.com/cl-tohoku/bert-japanese（最終閲覧日：2023年1月25日）
17) https://github.com/huggingface/transformers（最終閲覧日：2023年1月25日）
18) https://github.com/megagonlabs/jrte-corpus（最終閲覧日：2023年1月25日）

USING BIG DATA FOR TOURISM TEXT MINING ANALYSIS ON EXTRACTION OF IMPROVEMENT POINTS OF SIGHTSEEING SPOTS

Yuta BABA, Makoto FUJIU and Yuma MORISAKI

In Japan, local regions with declining populations, falling birthrates, and an aging population face the challenge of how to maintain and revitalize regional functions. In order to develop the tourism industry, it is important to formulate and implement policies based on diversified tourism needs. In this study, using the recent spread of the Internet, big data on tourism that has been accumulated in recent years, this study collects word-of-mouth data on tourist spots from travel information websites and performs sentiment analysis using deep learning to extract sentences from the word-of-mouth data that suggest negative sentiments of tourists, and extracts Text mining was then conducted on the negative sentiments to identify points for improvement of tourist attractions.