

Environment Research and Technology Development Fund

環境研究総合推進費 終了研究成果報告書

4RF-2002 ビッグデータと機械学習を用いた国立公園の文化的サービス評価
(JPMEERF20204R02)

令和2年度～令和4年度

Evaluating Cultural Services in National Parks: Application of Machine Learning Methods on
Social Media Data

〈研究代表機関〉

国立研究開発法人 森林研究・整備機構 森林総合研究所

〈研究分担機関〉

新潟大学

○図表番号の付番方法について

「Ⅰ．成果の概要」の図表番号は「0．通し番号」としております。なお、「Ⅱ．成果の詳細」にて使用した図表を転用する場合には、転用元と同じ番号を付番しております。

「Ⅱ．成果の詳細」の図表番号は「サブテーマ番号．通し番号」としております。なお、異なるサブテーマから図表を転用する場合は、転用元と同じ図表番号としております。

令和5年5月

目次

I. 成果の概要	1
1. はじめに（研究背景等）	
2. 研究開発目的	
3. 研究目標	
4. 研究開発内容	
5. 研究成果	
5-1. 成果の概要	
5-2. 環境政策等への貢献	
5-3. 研究目標の達成状況	
6. 研究成果の発表状況	
6-1. 査読付き論文	
6-2. 知的財産権	
6-3. その他発表件数	
7. 国際共同研究等の状況	
8. 研究者略歴	
II. 成果の詳細	
II-1 ビッグデータと機械学習を用いた国立公園の文化的サービス評価 （国立研究開発法人 森林研究・整備機構 森林総合研究所）	16
要旨	
1. 研究開発目的	
2. 研究目標	
3. 研究開発内容	
4. 結果及び考察	
5. 研究目標の達成状況	
6. 引用文献	
III. 研究成果の発表状況の詳細	31
IV. 英文Abstract	33

I. 成果の概要

課題名 4RF-2002 ビッグデータと機械学習を用いた国立公園の文化的サービス評価

課題代表者名 小黒 芳生 (国立研究開発法人 森林研究・整備機構 森林総合研究所
森林植生研究領域 群落動態研究室 主任研究員)

重点課題 主：【重点課題⑭】生態系サービスの持続的な利用やシステム解明に関する研究・技術開発

副：【重点課題 】

行政要請研究テーマ（行政ニーズ） 非該当

研究実施期間 令和2年度～令和4年度

研究経費

18,000千円（合計額）

（各年度の内訳：令和2年度：6,000千円、令和3年度：6,000千円、令和4年度：6,000千円）

研究体制

（サブテーマ1）

生態系サービスの持続的な利用やシステム解明に関する研究・技術開発

（国立研究開発法人 森林研究・整備機構 森林総合研究所）

研究代表機関：国立研究開発法人 森林研究・整備機構 森林総合研究所

研究分担機関：新潟大学

他のサブテーマはない。

研究協力機関

研究協力機関はない。

本研究のキーワード 生態系サービス、文化的サービス、国立公園、レクリエーション、機械学習、ビッグデータ、ソーシャルメディア、SNS

1. はじめに（研究背景等）

国立公園は希少な生態系の保全だけではなく、文化的生態系サービスのひとつ、レクリエーションサービス供給の場としても重要である。日本のレクリエーションサービス評価は「生物多様性および生態系サービスの総合評価報告書（環境省）」の中で国立公園全体での利用者数がサービスの指標として用いられている。しかし、既存の手法では、広大な国立公園内の利用者の空間分布を把握することは困難であり、自然植生の存在やアクセス性といった国立公園内のどのような要因がサービス利用者を増やしているのかは未解明であった。一方、近年はレクリエーションサービスの評価に位置情報が付与されたソーシャルメディアへの投稿が用いられるようになってきており（Aiba et al. 2019など）、国立公園内での投稿数を指標とすることでサービス利用の空間分布や季節変化、利用を高める要因の解析が可能である。さらに、近年の機械学習技術の発展により、ソーシャルメディアに存在するビッグデータから投稿内容の感情や、居住地や年齢などの投稿者の属性など、位置情報以外にも様々な情報を推定する

ことが可能になってきている。これらの技術を応用することで、これまで解析が困難であった国立公園のレクリエーションサービスに対する利用者の印象や、レクリエーションサービスの需要-供給の関係を考慮した国立公園の自然資本としての価値の定量的な評価が可能である。

2. 研究開発目的

本研究は日本の国立公園を対象とし、ソーシャルメディアに存在するビッグデータと機械学習技術を活用してレクリエーションサービス利用者の空間分布とその季節変化、サービスに対する利用者の印象を明らかにし、さらに国立公園の自然資本としての価値を定量的に評価する。これらのデータを用い、自然植生などの自然的要因と都市からの距離などの社会的要因の相互作用によりどのようにレクリエーションサービスが生み出されるのかを解明することを目的とする。

3. 研究目標

全体目標	国立公園の供給するレクリエーションサービスの利用、サービスに対する利用者の印象、および国立公園の自然資本としての価値を決定する自然的・社会的要因を解明する。そして、国立公園のレクリエーションサービスと自然資本としての価値をより向上させるための管理・政策について提言を行う。
サブテーマ1	ビッグデータと機械学習を用いた国立公園の文化的サービス評価
サブテーマリーダー/所属機関	小黒芳生／国立研究開発法人 森林研究・整備機構 森林総合研究所
目標	ソーシャルメディア上に存在するビッグデータと機械学習を用いて、国立公園の供給するレクリエーションサービスの利用、サービスに対する利用者の印象、そして国立公園の自然資本としての価値の季節変化および空間分布を明らかにする。そしてサービスの利用・印象・自然資本としての価値を向上させる自然的・社会的要因の解明を行う。また、ソーシャルメディアの特性に由来するバイアスの推定および補正方法の検討を行う。

4. 研究開発内容

【データ基盤の整備】

Twitterはユーザー数も多く、位置情報が付加された一般ユーザーの投稿データを大規模かつ容易に取得可能である。よって、本研究は研究対象としてTwitterを用いた。解析に用いるデータセットは以下の手法で作成した。

・ ソーシャルメディア投稿の収集・整理

申請者らはTwitterへの投稿を自動取得するプログラムを開発し、2013年より継続的にデータを取得してきた。これを拡張し、国立公園の位置情報を持つ投稿を収集するシステムを開発し、2017年3月以降の国立公園での投稿データを整備した。さらに、投稿に含まれるメタデータを分析し、ロボットによる自動投稿を抽出・除去するシステムを構築した。

・ 地名位置情報データベースの構築

研究提案時にはこれらを解決するため、投稿本文に含まれる地名と地名位置情報データベースを用いて位置情報を推定する予定であった。しかし、1)2017年以降の全投稿を用いると解析に十分な量の投稿が集まったこと、2)Instagram連携によるTwitterへの投稿は携帯電話の電波が届かない場所の位置情

報が付加された投稿を含んでいたことから位置情報の推定が不要となった。構築したデータベースは位置情報の推定ではなく、公園内で行われるアクティビティ（スキー・海水浴など）の定量に用いた。

・ サービスへの印象と国立公園利用者の属性の取得

国立公園滞在時の感情を利用者の印象の指標として用いるため、IBM社が提供する機械学習によるテキスト分析サービス（Watson Natural Language Understanding）・日本語感情分析システムMLAsk（Ptaszynski et al. 2009）を利用して投稿の内容を分析し、滞在時の感情を推定した。また、ソーシャルメディア利用者の属性（性別、年齢、居住都道府県）のバイアスを定量するため、NTTデータ社のユーザープロファイルDB、マイボイスコム株式会社「Twitterの利用に関するアンケート調査」の結果を利用し、解析を行った。

・ 投稿に添付された写真を用いた利用者のアクティビティの推定

アドバイザーリーボード会合での議論を元に、当初の研究計画に加えて投稿に付加された画像と機械学習サービス（Google Cloud Vision）を組み合わせた解析で、国立公園で行われているアクティビティの定量を行った。

・ 投稿に付加された「いいね」を用いた文化的サービスの間接的利用の推定

当初の計画に追加して、Twitterに投稿された位置情報付きの自然写真の「いいね」獲得数を、その場所の間接利用の指標とし、国立公園の保護レベルの高いエリアは間接利用が多いという仮説を検証した。国立公園との比較対象として、都市公園と東京ディズニーリゾートに位置情報がある自然写真を解析対象に加えた。

【国立公園の供給する文化的サービスの推定】

上記手法で整備した国立公園内の位置情報を付加された投稿、地名位置情報データベース、投稿の感情、投稿者の属性データを用い、以下の手法でレクリエーションサービスの解析を行った。

・ 国立公園内の利用者の空間分布・季節変化の推定

国立公園内での投稿データを3次メッシュごとに集計し、利用者数の空間分布を推定した。これを投稿本文の感情データと結合し、利用者の印象の空間分布を推定した。また、これらを月別に集計することで季節変化を推定した。

・ 国立公園の自然資本としての価値評価

研究計画時には国立公園利用者の居住都道府県から国立公園までの移動費用を元に、トラベルコスト法を用いて国立公園の自然資本としての価値を計算する予定であった。しかし、トラベルコスト法により推定される自然資本としての価値は宿泊費用や食費など、現地で支払われる旅行コストに大きく依存することがわかった。よって、結果が仮定に依存する自然資本としての価値ではなく、移動コストそのものを経済的な価値の指標として用いた解析を行った。

・ サービス利用者数や利用者の印象、自然資本としての価値の国立公園間での違いを決める要因の解析

当初の計画では国立公園内の空間分布は考慮せずに国立公園ごとに利用者数・印象・経済的価値を集計する予定であったが、当初の想定よりも多くの投稿データが得られたため、全国立公園を対象に3次メッシュ単位でこれらの指標を集計し、各国立公園内の空間分布の定量を試みた。ソーシャルメディアデータの使用に伴うバイアスを考慮するため、モデルの説明変数には利用者の属性を加える予定であったが、投稿単位の解析で性別の影響が小さく、年齢の推定に不確実性が見られたため、モデルの不確実性を下げるため利用者属性は解析に用いずに、結果を解釈する際に利用者属性による影響を考慮することとした。

・ 国立公園内でサービス利用者数や利用者の印象、自然資本としての価値の空間分布を生み出す自然的・社会的要因の解明

サービス利用者数と印象、自然資本としての価値の空間分布を自然要因・社会要因で説明する機械学習モデルを作成した。このモデルを用いて国立公園のどのような自然的・社会的要因がレクリエーションサービスの利用や印象、自然資本としての価値を向上させるのかを明らかにした。

5. 研究成果

5-1. 成果の概要

・ソーシャルメディア情報の信頼性

○ 国立公園訪問者指標としての投稿データの信頼性

国立公園ごとに集計した国立公園内の位置情報を持った投稿を行った利用者数と自然公園等利用者数調から得た国立公園ごとの利用者数との関係を見ると、両者の絶対数には大きな違いが見られた一方、両者の間には強い正の相関が見られた（図-1.1）。この結果はTwitterへの投稿者数は国立公園利用者数の相対的な指標として利用可能であることを示している。また、得られたデータを用いることで、国立公園内の詳細な利用の空間分布（図-1.2）や時間変化（図-1.3）を取得可能なことがわかった。

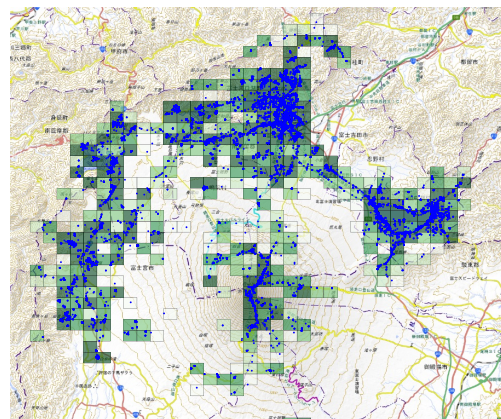
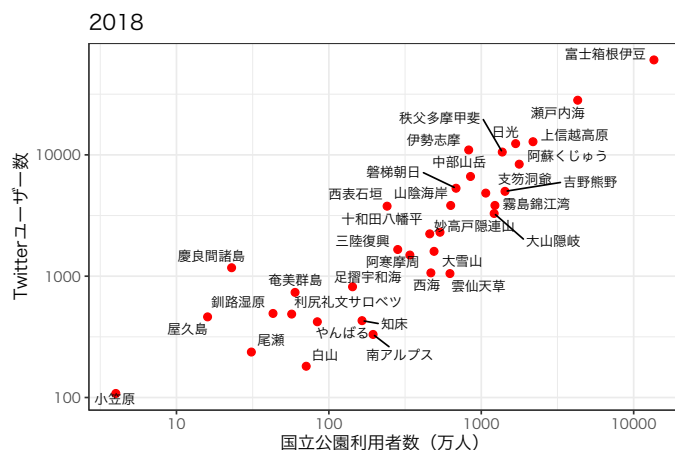


図-1.1：自然公園等利用者数調から得た国立公園ごとの利用者数と国立公園内の座標を持つ投稿を行ったTwitter投稿者数の関係。両者には強い相関があり、Twitter投稿者数を国立公園利用者数の相対的指標として用いることができることがわかる。

図-1.2：投稿の空間分布の例として富士箱根伊豆国立公園の富士山周辺を示した。青い点が投稿の座標を、背景色が3次メッシュ単位の投稿数を表し、濃色ほど投稿数が多い場所を示す。

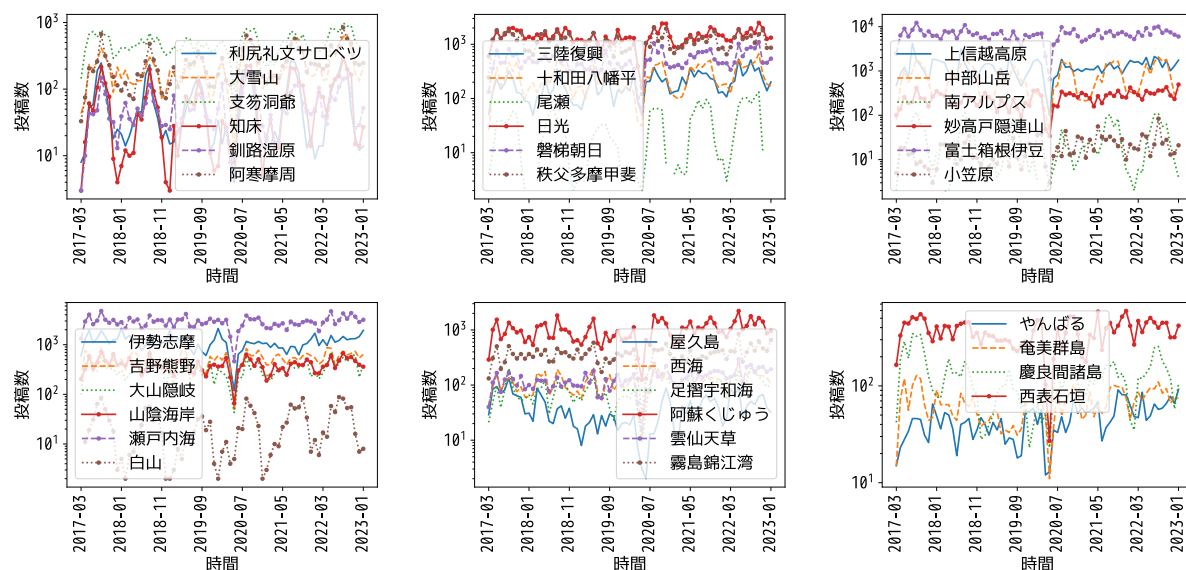


図-1.3：各国立公園の位置情報を持ったTwitterへの投稿数の月別変化

○ ソーシャルメディアデータ利用によるバイアスの影響の定量

解析の結果、Twitterに投稿された国立公園の位置情報を持った投稿には

- ・ 位置情報を付加された投稿は男性利用者のものが多いため、結果は男性の情報をより強く反映する

- ・ Twitter投稿者は若年層が多いため、若年層の情報をより強く反映する

というバイアスが生じていることがわかった。そこでこれらの属性が評価結果に与える影響を定量するため、利用者の移動コスト、感情スコアが投稿者の属性でどの程度説明できるのかを解析した。その結果、年齢・性別ともに移動コストへの影響は小さく、属性の偏りは移動コストの解析結果にあまり影響しないことがわかった（図-1.8左側）。

一方、感情スコアを同様にモデル化すると、高齢利用者および女性が若干ポジティブである、という結果が得られた（図-1.8右側）。しかし、モデルに投稿者の属性・投稿時間（曜日・月・コロナ前後）・訪問先国立公園・保護区分・添付画像ありなし・投稿者の個人差を加えてもモデルの R^2 は0.1程度、投稿者の個人差を考慮しない場合、 R^2 は0.01となり、ほとんどデータのばらつきを説明できなかった。よって感情に関してもソーシャルメディア情報の投稿者属性のバイアスは結果に大きな影響を与えないことがわかった。

さらに3次メッシュ単位で投稿数を集計し、全投稿数と性別・年齢ごとの投稿数の相関をみることで投稿数の分布へバイアスの影響を確認した。全投稿数と性別ごとの投稿数は相関が強かったが（図-1.9右）、年齢階層ごとの投稿数を見ると50代以上の投稿者による投稿数は全投稿数との相関が弱かった（図-1.9左）。この結果は投稿数を用いて利用頻度を推定すると、高齢者の傾向を過小評価する可能性を示している。

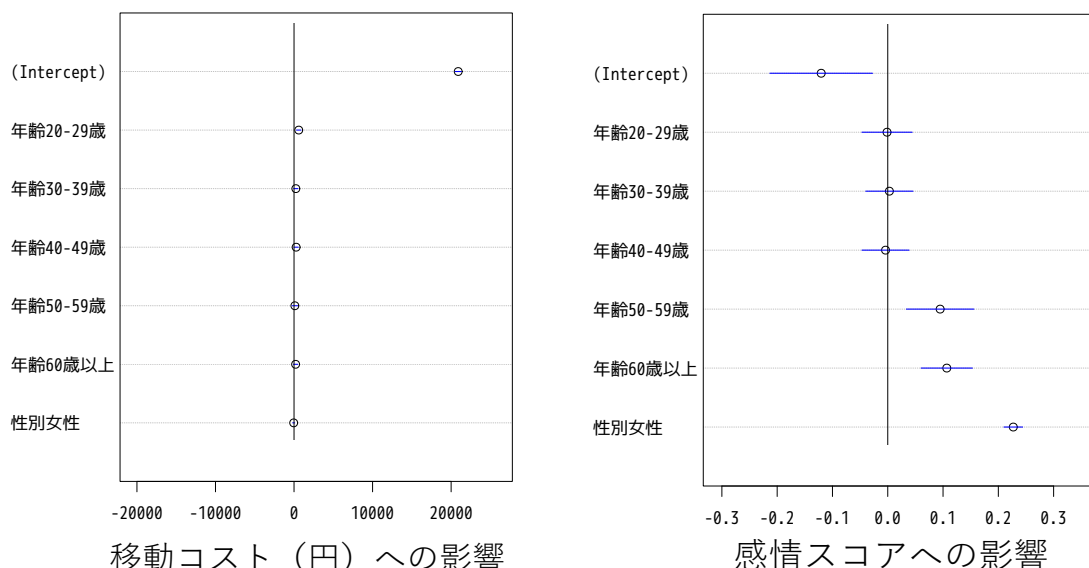


図-1.8：移動コスト・感情スコアに与える投稿者の年齢・性別の影響。感情スコアは値が高いほどポジティブな感情の投稿であること、値が低いほどネガティブな感情の投稿であることを示している。

Interceptはモデル全体の切片を表しており、10代男性の国立公園への移動コスト・感情スコアの平均を意味している。また、他の点は10代の投稿者と各年齢階層の投稿者の移動コスト・感情スコアの差、男性と女性の移動コスト・感情スコアの差を表している。点のエラーバーは推定値の信頼区間を表している。

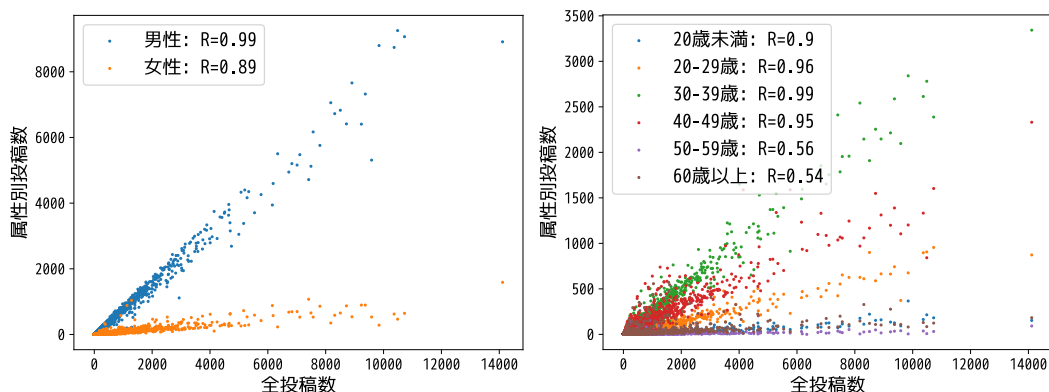


図-1.9：3次メッシュごとに集計したTwitterへの全投稿数と投稿者の性別・年齢階層ごとの投稿数の関係。凡例には相関係数も示した。

・国立公園訪問が利用者の感情に与える影響の解明と感情の季節変化

国立公園への訪問が利用者の感情に与える影響を調べるため、国立公園の位置情報を付加した投稿を行った投稿者による国立公園の位置情報が付加されていない投稿を収集し、感情スコアを比較した。その結果、IBM社のWatson Natural Language Understandingを用いた結果（図-1.10）では国立公園内の投稿はポジティブな感情を持つ投稿が多い一方、国立公園の位置情報が付加されていない投稿はネガティブな感情の投稿が多いことがわかった。また、日本語感情分析システムMLAskを用いた解析（図-1.11）でも国立公園の位置情報が付加された投稿は「哀」「厭」「怒」など、ネガティブな感情を持つ投稿が少ない一方、「喜」の感情を持つ投稿が多いことがわかった。これらの結果から国立公園への訪問は利用者の感情にポジティブな影響を与えていることがわかった。またどちらの手法を用いた結果を見ても感情を持った投稿の割合には大きな季節性は見られなかった。

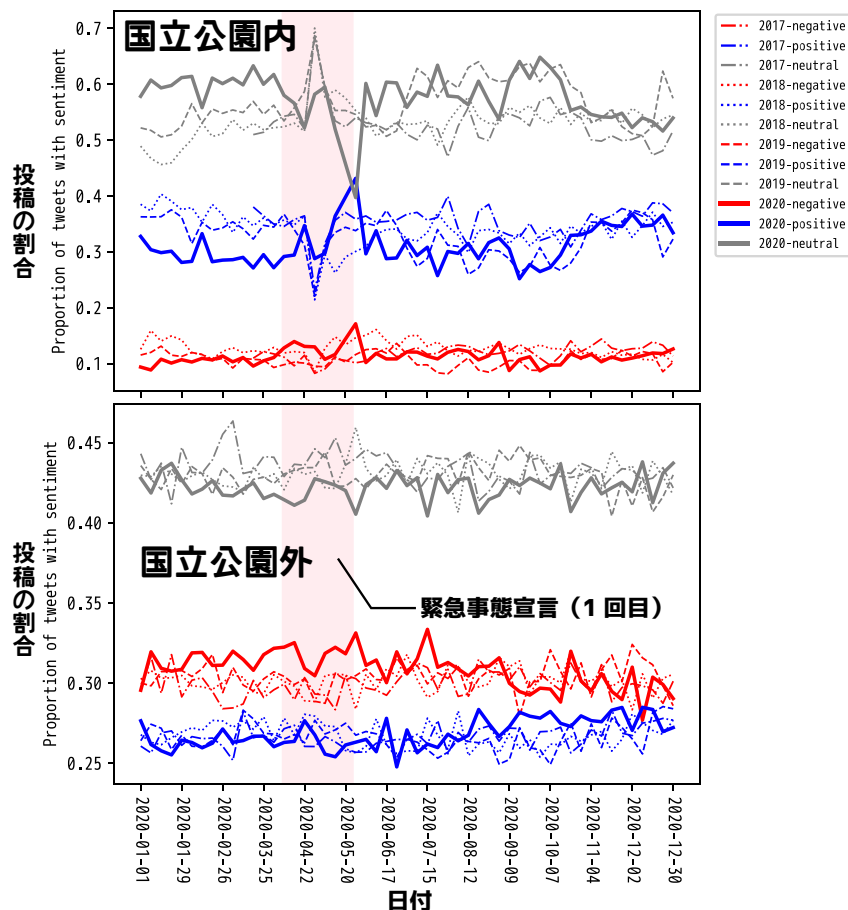


図-1.10：国立公園内の投稿と同じ投稿者が行った国立公園外の投稿の感情（赤：ネガティブ・青：ポ

ジティブ・灰：感情なし）。各線は各年の感情を含む投稿の割合の時間変化を示す。背景がピンク色の部分は2020年の新型コロナウイルス対応の1回目の緊急事態宣言期間で、この期間は国立公園内の投稿が少なく、国立公園内のデータの信頼性が低いことに注意が必要である。国立公園内はポジティブな感情を持った投稿が多く、ネガティブな感情を持った投稿が少ないこと、感情には大きな季節性が見られないことがわかる。

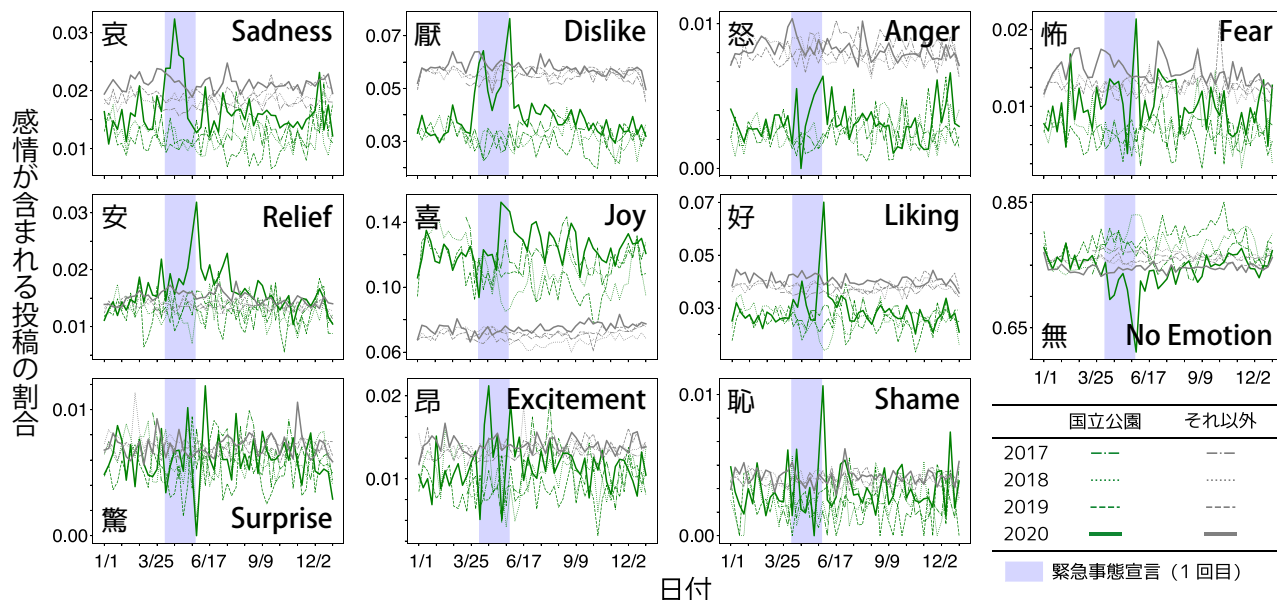


図-1.11：国立公園内の投稿と国立公園外の投稿の感情の違い。線の種類の違いは年の違いを表し、緑の線は国立公園内を灰色の線は国立公園外の投稿の感情の季節変化を表す。背景が青色の部分は2020年の新型コロナウイルス対応の1回目の緊急事態宣言期間で、この期間は国立公園内の投稿が少なく、国立公園内のデータの信頼性が低いことに注意が必要である。

・投稿された写真に基づくアクティビティの抽出とその季節性

Googleが提供する画像認識AIを利用し、国立公園内でTwitterに投稿された写真にどのような要素が写されていたのかを分析した。その結果、各国立公園で投稿される要素が大きく異なること、そして投稿される写真タグが季節により変化すること、さらに季節性は公園によって異なることが示された（図-1.12）。これらの写真タグに基づき8種類のアクティビティを抽出し、それぞれの国立公園でどのようなアクティビティに関連する写真が投稿されているのかを客観的に評価することができた（図-1.13）。さらに、アクティビティ別の写真投稿数に基づいたクラスター分析により、各国立公園を特徴づけるアクティビティを評価することができた（図-1.14）。

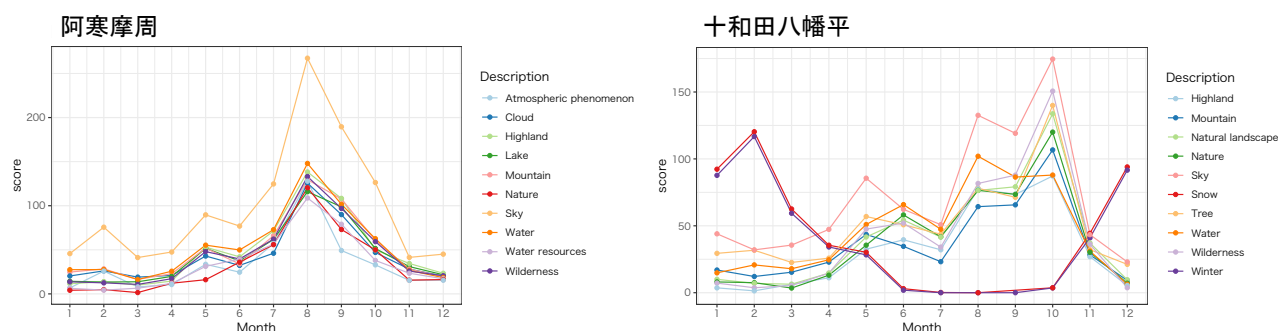


図-1.12：阿寒摩周国立公園および十和田八幡平国立公園でTwitterに投稿された写真に写されていた要素の月別変化。阿寒摩周国立公園では8月に写真投稿数のピークがあり、空や水、高原、湖、山などに関連する写真の投稿が多かった。十和田八幡平国立公園は2月と10月にそれぞれ写真投稿数のピークが見られ、2月には雪に関連する写真の投稿が多く、10月には樹木や原生自然、自然景観などに関連

する写真の投稿が多いことがわかる。

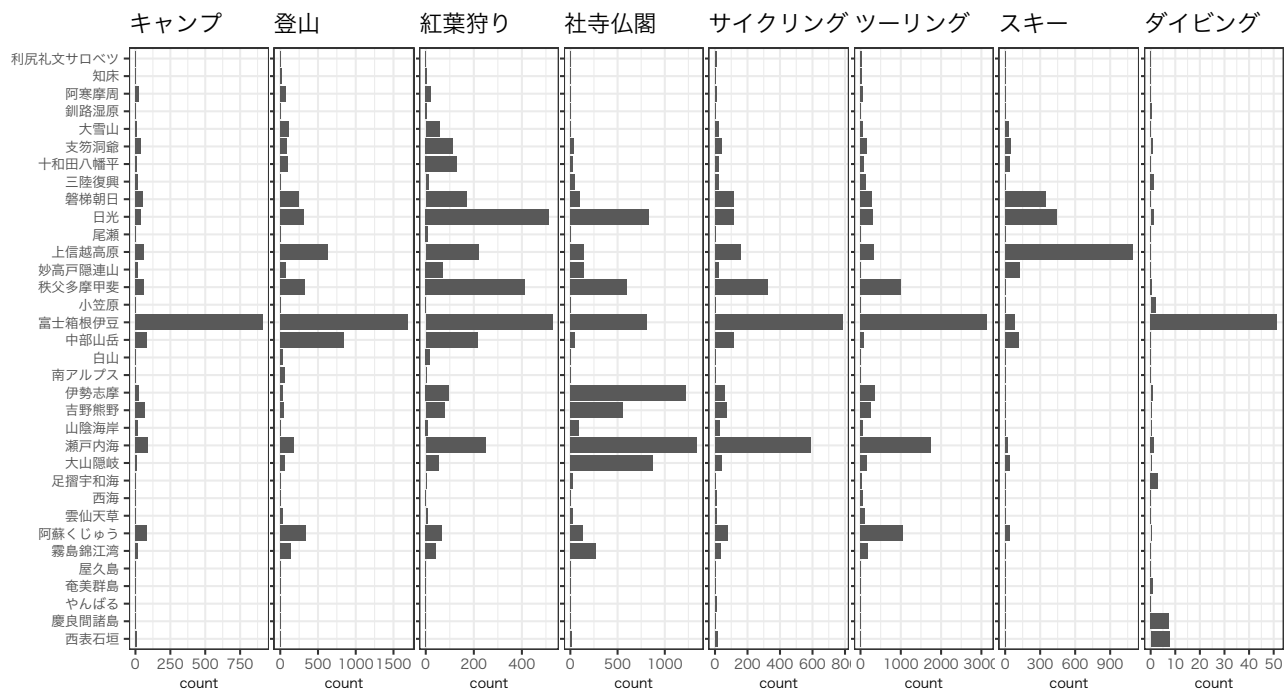


図-1.13：各国立公園における写真に写された要素から抽出したアクティビティ別の写真投稿数。

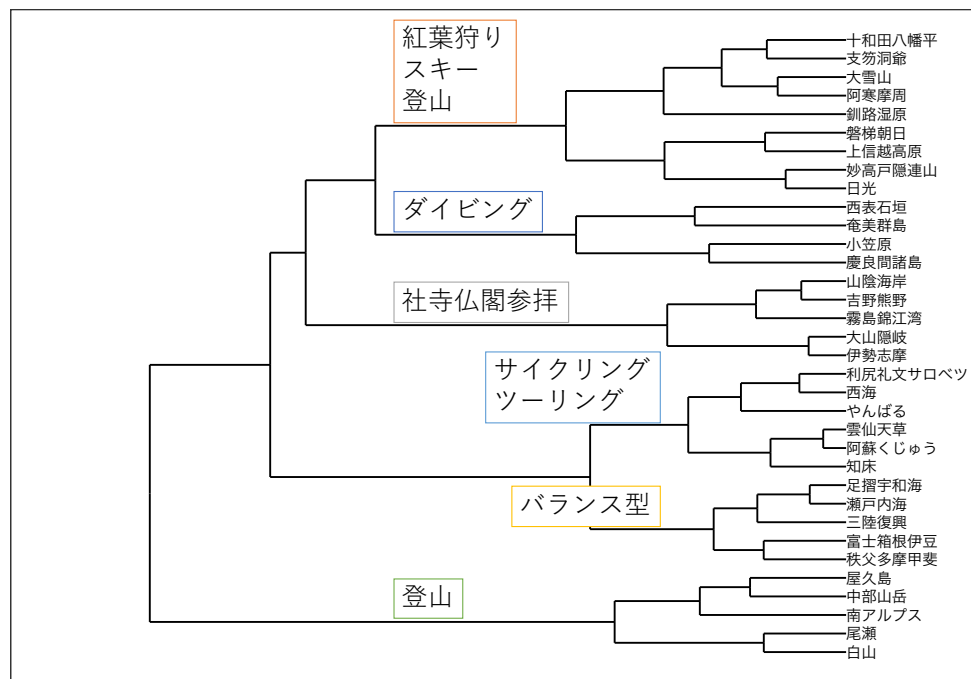


図-1.14：各国立公園のアクティビティ別の写真投稿数に基づいたクラスター分析。国立公園利用者が写真に投稿したアクティビティの種類が似通った国立公園が近いグループに分類されている。

・地名位置情報データベースの構築とアウトドア関連スポットの定量

Foursquareに登録されていたアウトドア関連スポットは「山」が1498スポット、「展望所」が840スポット、「海岸」が785スポット、「温泉」が660スポット、「トレイル」が577スポットであった。抽出されたアウトドア関連スポット数やその小カテゴリーの内訳は国立公園ごとに大きく異なっており、各国立公園の自然景観の特徴がよく表れていた（図-1.15）。

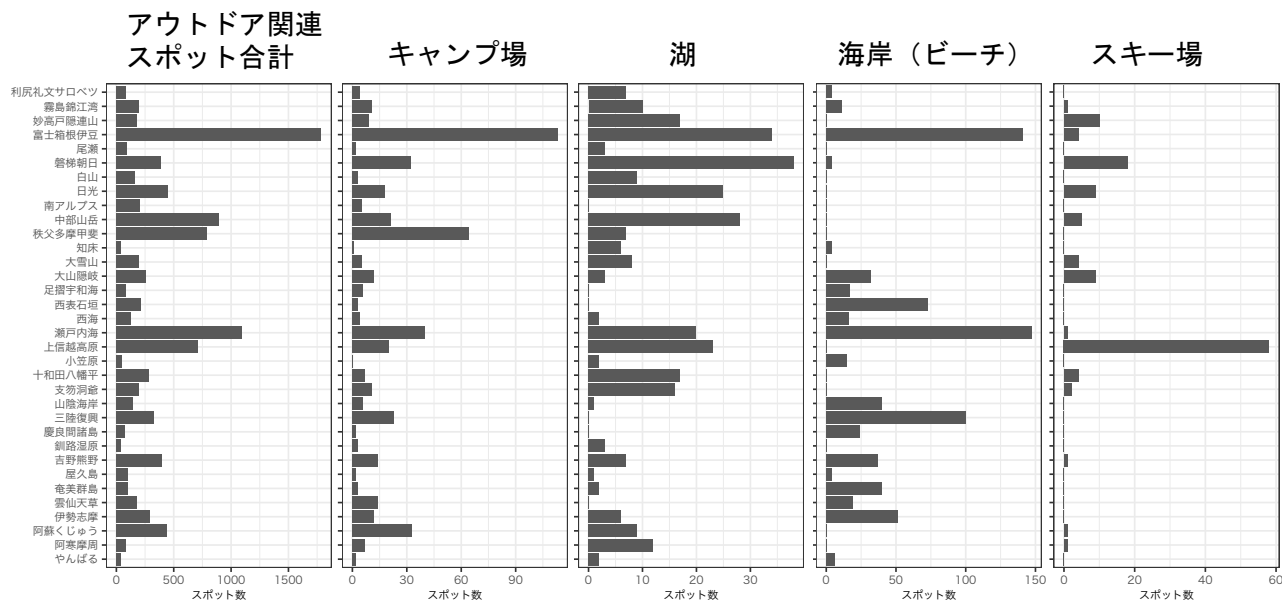


図-1.15 : Foursquareから抽出した国立公園ごとのスポット数。左から大カテゴリーが「アウトドア (Outdoor) 」で登録されていたスポットの合計数、大カテゴリー「アウトドア」に含まれる小カテゴリーの例として「キャンプ場 (Campground) 」、「湖 (Lake) 」、「海岸 (Beach) 」、「スキー場 (Ski Area) 」で登録されていたスポット数をそれぞれ示している。

・国立公園における保護レベルがソーシャルメディアを通じた間接利用に与える影響

国立公園の自然は本やテレビ、ソーシャルメディアに投稿された写真を通じて間接的にも利用されている。そこでTwitterに投稿された写真を見た人々が付加した「いいね」を指標にすることで、実際には訪問していない人々の国立公園生態系の間接利用を評価した。特に、保護レベルの高いエリアの位置情報が付与された自然写真は「いいね」獲得数が多い、つまり間接利用が増加することで国立公園の価値を高めているという仮説を検証した。国立公園により傾向にばらつきは見られたが、平均すると特別保護地区で投稿された自然写真は同じ投稿者が国立公園以外の場所で投稿した自然写真と比較して「いいね」獲得数の期待値は約1.2倍に増加した。普通地域と特別地域においても増加傾向は見られたが、増加率は約1.1倍と小さく、統計的にも有意ではなかった。比較対象とした都市公園は低下傾向、東京ディズニーリゾートは年により傾向が大きく異なるという結果となった(図-1.16)。以上の結果は、同じ自然写真であってもより保護レベルが高く原生的な自然が残されたエリアの写真ほどTwitter上で共感や感動を集めており間接利用が多いこと、都市公園のような人工的な自然景観は身近な訪問先ではあるがTwitter上での共感や感動は得られ難く間接利用が少ないことを示している。これらの結果は、国立公園の訪問者数の増加や利用者の利便性の向上を目的に安易に生態系に手を加えることは、ソーシャルメディアなどを通じた間接利用を減少させ、総合的には国立公園の価値の低下に繋がりがかねないこと、そして国立公園の利用と保護は必ずしも対立するものではないことを示している。

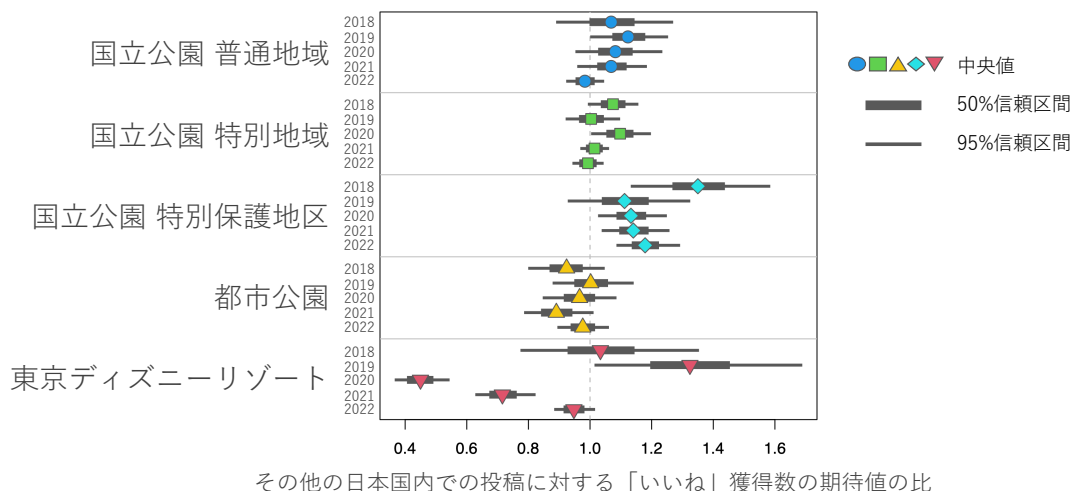


図-1.16：同じ投稿者が国立公園の各エリア、都市公園、東京ディズニーリゾート内の位置情報が付与された自然写真を投稿した場合と、その他の日本国内の位置情報が付与された自然写真を投稿した場合の「いいね」獲得数の期待値の比。95%信頼区間が1.0の点線から離れている場合に期待値が統計的に有意に増加（または減少）したことを示している。

・国立公園内でサービス利用者数や利用者の印象、自然資本としての価値の空間分布を生み出す自然的・社会的要因の解明

全国の国立公園を対象としたモデルでは、サービス利用者数はFoursquareに登録されていた観光スポット数や道路密度などの重要度が高く、観光スポットが存在しアクセス性も良い場所の利用者数が多いことがわかった（図-1.17-1, 2, 3）。また、重要度は高くはないものの原生植生率や特別地域率、特別保護地区が高い場所、つまり自然度が高いエリアほど利用者数が多かった（図-1.17-6, 13, 16）。

自然資本としての価値を合計移動コスト（各メッシュで投稿した人の訪問費用の合計）を対象に解析すると、観光スポットが多い場所や人口密度が高い地域で合計移動コストが高くなった一方（図-1.18-1, 2, 3）、人口密度が低い場所、沖縄や北海道（高緯度・低緯度）、保護地域の面積割合が高いほど合計移動コストが高いことがわかった（図-1.18-3, 4, 12, 13, 16）。これらの結果は利用者数や自然資本としての価値を維持するためには観光スポットの存在やアクセス性と高い自然度のバランスよい場所ほど遠くから多くの利用者が訪問することを示している。一方利用者の印象はモデルの予測力が非常に低く、利用者数や移動コストと同じ要因では説明できないことが明らかとなった。

各国立公園を対象としたモデルでは全国の国立公園を対象とした場合と同様の傾向が見られたが、それぞれの国立公園に特徴的な傾向も見られた。サービス利用者数については全国の国立公園を対象とした場合と同様に観光スポット数やアクセス性、特別地域率、特別保護地区率に加えて、湖や河川などの内水面率、標高や地形などの重要度が高い国立公園が多かった。合計移動コストは利用者数と同様の結果だったが、各メッシュへの平均移動コストをみると人口密集地から離れており、原生植生率や特別地域・特別保護地区率が高いほど平均移動コストが高い国立公園が多かった（図-1.19）。利用者の印象はほとんどの国立公園で予測力が非常に低かったが、一部の国立公園では予測可能だった。例えば南アルプス国立公園の尾根・山頂エリアや尾瀬国立公園の尾瀬ヶ原・尾瀬沼周辺などその国立公園を代表する景勝エリアで相対的に印象がポジティブとなる傾向が得られた（図-1.20）。

利用者数への影響

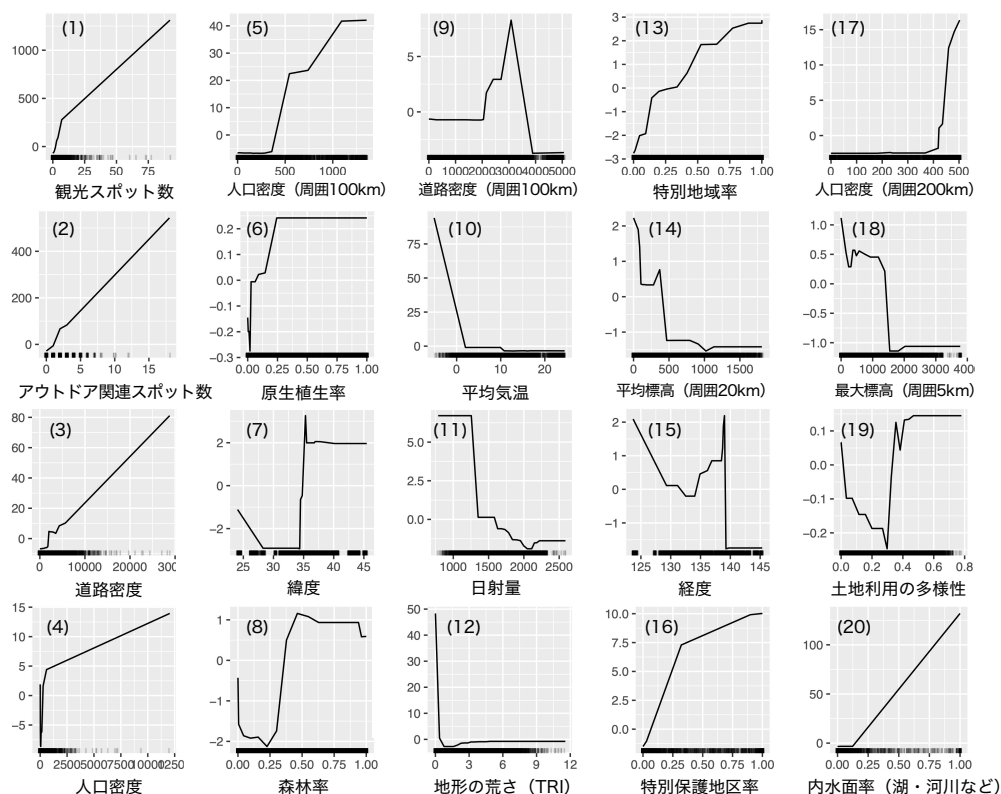


図-1.17：自然的・社会的要因がサービス利用者数に与える影響の機械学習による推定値。重要度の上位から20個の要因を示した。黒の曲線が右肩上がりの場合はその要因の値が大きい3次メッシュほどサービス利用者数が多い傾向があり、右肩下がりの場合は逆の傾向があることを意味する。

合計移動コストへの影響

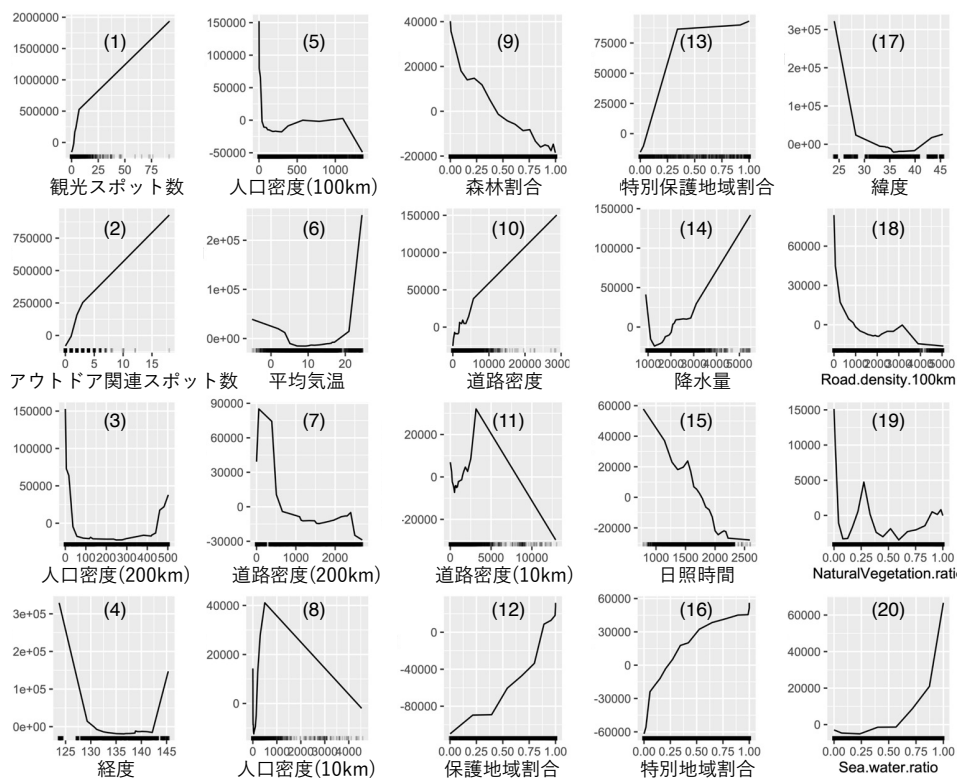
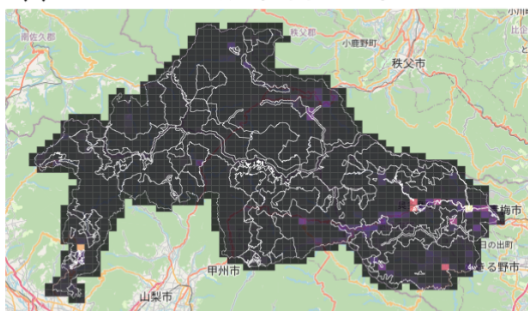


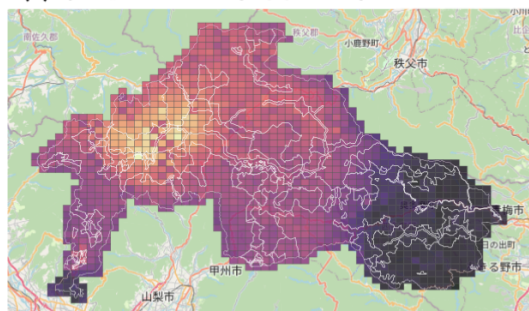
図-1.18：自然的・社会的要因が合計移動コストに与える影響の機械学習による推定値。重要度の上位から20個の要因を示した。図の番号は重要度の順位を示す。黒の曲線が右肩上がりの場合はその要因の値が大きい3次メッシュほど合計移動コストが高い傾向があり、右肩下がりの場合は逆の傾向があることを意味する。

移秩父多摩甲斐国立公園

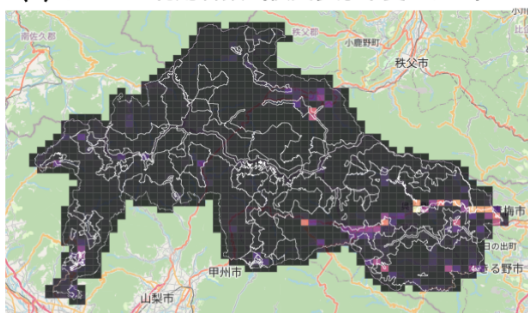
(a) サービス利用者数（全国モデル）



(c) 平均移動コスト（全国モデル）



(b) サービス利用者数（秩父多摩甲斐モデル）



(d) 平均移動コスト（秩父多摩甲斐モデル）

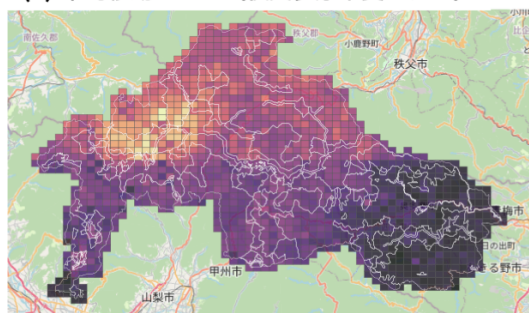
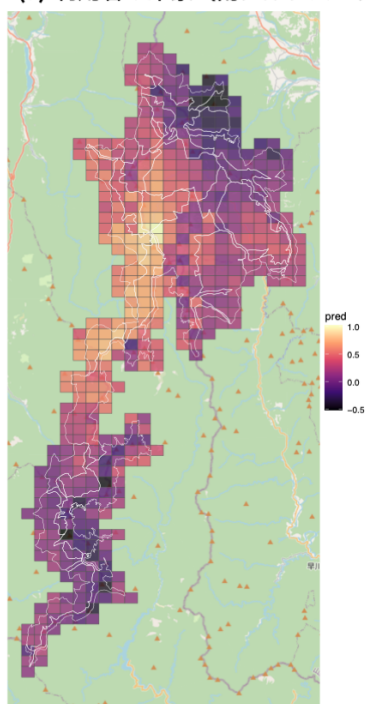


図-1.19：機械学習によるサービス利用者数、平均移動コストの推定値の空間分布（秩父多摩甲斐国立公園の例）。全国を対象としたモデル(a, c)と各国立公園を対象としたモデル(b, d)をそれぞれ示している。両モデルの予測結果は同様に、利用者数は観光スポットの空間分布に大きく依存しており、平均移動コストは大都市圏からの距離が遠く原生的な自然が多く残されているエリアで高くなっている。

(a) 利用者の印象（南アルプスモデル）



(b) 利用者の印象（尾瀬モデル）

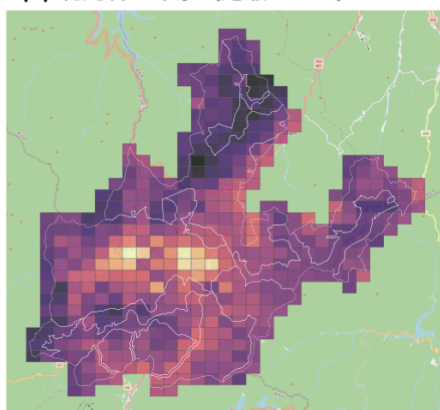


図-1.20：機械学習による利用者の印象の推定値の空間分布（南アルプス国立公園、尾瀬国立公園の例）。色が明るい場所は暗い場所と比較して相対的にポジティブな投稿が多いことを示している。

○ 環境政策への提言

本研究の結果、その結果、観光スポットが多い場所や人口密度が高い地域で合計移動コストが高くなった一方（図-1.18(1)～(3)）、人口密度が低い場所、沖縄や北海道（高緯度・低緯度）、保護地域の面積割合が高いほど合計移動コストが高いことがわかった（図-1.18(3)(4)(12)(13)(16)）。これらの結果は自然資本としての価値の決定要因として観光スポットの存在やアクセス性が最も重要ではあるが、これらが同様の条件であれば手付かずの自然が残されているほどより遠くから多くの利用者が訪問することを示している。よって国立公園の価値を維持するためには保全と利用施設の整備のバランスが重要であると考えられる。

また本研究の結果、同一国立公園内でも特別保護区のように保護レベルが高い場所でSNSを介した間接利用が多いことが示された（図-1.16）。この結果は訪問による直接利用だけではなく、間接的な国立公園利用にも生態系の保全が必要であることを示している。また、国立公園の広報の際には特別保護区内の自然度の高い地域の写真を用いることが効果的であると考えられる。

さらに国立公園の利用は利用者の感情にポジティブな影響を与えることが示された（図-1.10, 11）。この結果は、国立公園は経済的な指標だけでは評価しきれない価値を有していることを示している。

5-2. 環境政策等への貢献

<行政等が既に活用した成果>

特に記載すべき事項はない。

<行政等が活用することが見込まれる成果>

・位置情報を持ったTwitterへの投稿により国立公園の詳細な利用動向が把握可能である

各国立公園におけるTwitterへの年間の投稿数は既存の統計による年間利用者数と相関が強く、各国立公園の利用者数の指標として利用可能であることがわかった。また、投稿数の空間分布や季節変化から国立公園内の時間的・空間的に高解像度な利用動向を把握できること、投稿された内容や付加された写真から利用者の印象や活動内容などの詳細な利用実態が把握できることがわかった。さらに、投稿者の分析を行うことで国立公園利用者の年齢層や居住都道府県などもある程度推定が可能であった。このような従来のアンケート調査などでは把握が難しかった広域的・高解像度な利用動態情報は、今後の公園計画の見直しの際などに活用が見込まれる。ただし、投稿数の空間分布は50歳未満の比較的若い層の利用動態を強く反映している可能性が示されたため、この点には留意が必要である。

・国立公園は利用者の感情にポジティブな影響を与える

国立公園におけるTwitterへの投稿は国立公園外での投稿と比べてポジティブな内容の投稿割合が多かった。この結果は、国立公園の訪問が利用者の感情にポジティブな影響を与えたことを示している。これまで森林浴などの自然への訪問が人々の健康に与える様々な影響が分析されてきたが、自然公園への訪問そのものが人々の感情に与える影響について示した事例はほとんどないと思われる。この結果は国立公園が観光産業への影響などの経済的価値だけでは表現できない価値を有していることを示しており、例えば環境省ウェブサイト「データで見る国立公園の健康効果とは？」などの国立公園の持つ価値のPR資料等への活用が可能である。

・国立公園の特別保護地区はソーシャルメディアを介した間接利用が多い

特別保護地区など、国立公園の保護レベルが高いエリアで撮影され、投稿された自然写真は他のTwitter利用者から高く評価されやすいことが示された。この結果は実際には訪問していない人々もTwitterに投稿された写真を見ることで間接的に特別保護地区の生態系から文化的サービスを楽しんでいること、そして生態系の保全がこのような間接利用による価値を高めていることを示している。テレビやソーシャルメディアなどを介した生態系の間接利用を評価した事例はこれまでにほとんど無く、本研究結果は国立公園の訪問者を対象とした分析だけでは過小評価されていた新たな価値を示すものであ

る。このような新たな生態系の評価指標は生物多様性総合評価などに活用することが可能である。

・国立公園の経済的な価値を維持するためには保全と利用のバランスが重要

国立公園の利用者数や経済的価値は展望台などの観光スポットが多い場所や人口密集エリアから近くアクセスが良い地域で高かった一方、特別地域や特別保護地区の面積割合が高く原生的な自然がより多く残された場所でも高いことがわかった。特に、生態系を保全し原生的な自然を守ることはより広い地域から訪問者を集客する上で重要であることが示された。このように観光スポットなどの施設整備と生態系の保全の両方が国立公園の価値の維持に重要であることをデータから客観的に示した事例は少なく、公園計画の見直しや自然公園の新規指定・拡張を行う際に考慮されることが期待される。

5－3．研究目標の達成状況

全体目標	目標の達成状況
国立公園の供給するレクリエーションサービスの利用、サービスに対する利用者の印象、および国立公園の自然資本としての価値を決定する自然的・社会的要因を解明する。そして、国立公園のレクリエーションサービスと自然資本としての価値をより向上させるための管理・政策について提言を行う。	<u>目標を上回る成果をあげた。</u> 国立公園の供給するレクリエーションサービスの利用頻度・経済的価値の指標としての利用者の移動コストの空間情報を作成し、ユーザー属性・自然的要因・社会的要因との関係を解明した。一方、印象の指標としての感情は同様の解析を行った結果、利用頻度や移動コストと同じ要因では説明ができないことを解明した。また結果を基に国立公園の価値を維持するためには施設整備と保護の両方が重要であるとの提言を行った。 また、当初研究計画にはなかった投稿に添付されている画像や観光スポット情報からスキーや紅葉など、各国立公園で積極的に行われているアクティビティを抽出する技術の開発を行った。さらに当初の研究計画にはなかった文化的サービスのインターネット経由での間接利用の解析を行い、保護レベルの高い場所ほどサービスの間接利用が多いことを明らかにした。 以上のとおり当初計画の研究目標を達成した上で、さらに当初計画にはなかった技術開発を積極的に行ったため、目標を上回る成果を上げたと判断した。

サブテーマ1 目標	目標の達成状況
ソーシャルメディア上に存在するビッグデータと機械学習を用いて、国立公園の供給するレクリエーションサービスの利用、サービスに対する利用者の印象、そして国立公園の自然資本としての価値の季節変化および空間分布を明らかにする。そしてサービスの利用・印象・自然資本としての価	<u>目標を上回る成果をあげた。</u> 国立公園の供給するレクリエーションサービスの利用頻度・経済的価値の指標としての利用者の移動コストの空間分布・季節変化情報を作成し、ユーザー属性・自然的要因・社会的要因との関係を解明した。一方、印象の指標としての感情は同様

値を向上させる自然的・社会的要因の解明を行う。また、ソーシャルメディアの特性に由来するバイアスの推定および補正方法の検討を行う。	の解析を行った結果、利用頻度や移動コストと同じ要因では説明ができないことを解明した。また、ソーシャルメディアに由来するバイアスの影響を定量し補正方法を検討したが、推定の不確実性のため、現時点では補正を行わないことが適切と判断した。 また、当初研究計画にはなかった投稿に添付されている画像や観光スポット情報からスキーや紅葉など、各国立公園で積極的に行われているアクティビティを抽出する技術の開発を行った。さらに当初の研究計画にはなかった文化的サービスのインターネット経由での間接利用の解析を行い、保護レベルの高い場所ほどサービスの間接利用が多いことを明らかにした。 以上のとおり当初計画の研究目標を達成した上で、さらに当初計画にはなかった技術開発を積極的行ったため、目標を上回る成果を上げたと判断した。
---	---

6. 研究成果の発表状況

6-1. 査読付き論文

<件数>

0件

<主な査読付き論文>

特に記載すべき事項はない

6-2. 知的財産権

特に記載すべき事項はない

6-3. その他発表件数

査読付き論文に準ずる成果発表	0件
その他誌上発表（査読なし）	0件
口頭発表（学会等）	10件
「国民との科学・技術対話」の実施	2件
マスコミ等への公表・報道等	0件
本研究費の研究成果による受賞	1件
その他の成果発表	0件

7. 国際共同研究等の状況

特に記載すべき事項はない。

8. 研究者略歴

研究代表者

小黒 芳生

東北大学大学院生命科学研究科卒業、生命科学博士、現在、森林総合研究所主任研究員

研究分担者

柴田 嶺

東北大学大学院生命科学研究科卒業、生命科学博士、現在、新潟大学農学部助教

Ⅱ．成果の詳細

Ⅱ－１ ビッグデータと機械学習を用いた国立公園の文化的サービス評価

国立研究開発法人 森林研究・整備機構 森林総合研究所

森林植生研究領域 群落動態室

小黒 芳生

研究分担者

新潟大学

自然科学系（農学部）

柴田 嶺

〔要旨〕

自然的要因・社会的要因の相互作用により、どのようにレクリエーションサービスが生み出されるのかを解明するため、日本の国立公園を対象として、ソーシャルメディアを用いた研究を行った。Twitterに投稿された国立公園の位置情報を持つ投稿を取得し、利用頻度の指標として投稿数を、利用者の印象の指標として投稿の感情を、自然資本としての価値の指標として移動コストを計算し、その空間分布を説明する自然的・社会的要因を解析した。また、ソーシャルメディア利用によるバイアスの影響を解析した。さらに、当初計画に加えてサービス間接利用の解析、投稿に添付された画像を用いたアクティビティの解析も行った。

解析の結果、国立公園ごとのTwitterへの投稿数は既存統計の結果とよく相関し、投稿数を利用頻度の指標として用いることが可能であることがわかった。Twitterは若年層の利用者が多く、年齢により投稿の空間分布が異なる可能性が示されたため、Twitterを用いた解析結果は若年層の行動を反映しがちであることがわかった。一方、性別や居住都道府県のバイアスの影響は見られなかった。

国立公園利用は利用者の感情にポジティブな影響があることがわかった。また投稿された写真を用いることで各国立公園に特徴的なアクティビティを把握できることがわかった。さらに特別保護地区で撮影された写真はソーシャルメディアを通じた間接的な利用が多いことがわかった。国立公園の利用頻度や自然資本としての価値はアクセス性が高く人口密集地から近い場所ほど高い一方、自然度が高い場所も高く、両者のバランスが重要であることが示唆された。これらの結果は当初計画を上回ると自己評価した。

1. 研究開発目的

本研究は日本の国立公園を対象とし、ソーシャルメディアに存在するビッグデータと機械学習技術を活用してレクリエーションサービス利用者の空間分布とその季節変化、サービスに対する利用者の印象を明らかにし、さらに国立公園の自然資本としての価値を定量的に評価する。これらのデータを用い、自然植生などの自然的要因と都市からの距離などの社会的要因の相互作用によりどのようにレクリエーションサービスが生み出されるのかを解明することを目的とする。

2. 研究目標

ソーシャルメディア上に存在するビッグデータと機械学習を用いて、国立公園の供給するレクリエーションサービスの利用、サービスに対する利用者の印象、そして国立公園の自然資本としての価値の季節変化および空間分布を明らかにする。そしてサービスの利用・印象・自然資本としての価値を向上させる自然的・社会的要因の解明を行う。また、ソーシャルメディアの特性に由来するバイアスの推定および補正方法の検討を行う。

3. 研究開発内容

【データ基盤の整備】

日本国内ではTwitter、Instagram、Facebookなどのソーシャルメディアが用いられているが、

InstagramやFacebookは自動的データ取得に用いるApplication Programming Interface (API)という機能を廃止したため、一般ユーザーのデータの収集が行えない。一方、Twitterはユーザー数も多くAPIにより一般ユーザーのデータを大規模かつ容易に取得可能である。また、投稿の一部には位置情報データが含まれており、国立公園に関連する投稿を収集することも容易である。よって、本研究は研究対象としてTwitterを用いた。解析に用いるデータセットは以下の手法で作成した。

・ ソーシャルメディア投稿の収集・整理

申請者らはTwitterへの投稿を自動取得するプログラムを開発し、2013年より継続的にデータを取得してきた。これを拡張し、国立公園の位置情報を持つ投稿を収集するシステムを開発した。また、蓄積されたデータ・新規に取得したデータを用いて2017年3月以降の国立公園での投稿データを整備した。

収集した投稿には解析の妨げになる様々なノイズ（ロボットによる自動投稿など）が含まれるため、解析を行う前にこのようなノイズを除去する必要がある。この問題を解決するため、投稿に含まれるメタデータを分析し、ロボットによる自動投稿を抽出・除去するシステムを構築した。

・ 地名位置情報データベースの構築

Twitterに占める位置情報が付加された投稿の割合は全体の0.18%程度と低く、また携帯の電波が届かず投稿ができない地域の情報は収集できない。研究提案時にはこれらを解決するため、投稿本文に含まれる地名と地名位置情報データベースを用いて位置情報を推定する予定であった。この目的のため、店や観光スポットなどの位置情報を提供するソーシャルメディアであるFoursquareに登録されているアウトドア関連スポットを収集し、地名位置情報データベースを構築した。しかし、1)2017年以降の全投稿を用いると解析に十分な量の投稿が集まったこと、2)Instagram連携によるTwitterへの投稿は携帯電話の電波が届かない場所の位置情報が付加された投稿を含んでいたことから、取得したデータで本研究の目標は達成できる見込みとなり、位置情報の推定が不要となった。構築したデータベースは位置情報の推定ではなく、公園内で行われるアクティビティ（スキー・海水浴など）の定量に用いた。

・ サービスへの印象と国立公園利用者の属性の取得

国立公園滞在時の感情を利用者の印象の指標として用いるため、IBM社が提供する機械学習によるテキスト分析サービス（Watson Natural Language Understanding）・日本語感情分析システムMLAsk（Ptaszynski et al. 2009）を利用して国立公園内での投稿の内容を分析し、滞在時の感情を推定した。また、ソーシャルメディア利用者の属性のバイアスを定量するため、NTTデータ社のユーザープロフィールDBを用いて国立公園の位置情報を持った投稿を行った投稿者の属性（性別、年齢、居住都道府県）を推定した。また、ユーザープロフィールDBの信頼性を確認するため、マイボイスコム株式会社「Twitterの利用に関するアンケート調査」の結果を入手し、ユーザープロフィールDBの解析結果と比較した。

・ 投稿に添付された写真を用いた利用者のアクティビティの推定

アドバイザーボード会合での議論を元に、当初の研究計画に加えて投稿に付加された画像を用いた解析も行った。Twitterへの投稿には利用者が国立公園内で撮影した写真が添付されていることがある。この写真を機械学習サービス（Google Cloud Vision）で解析することで、国立公園で行われているアクティビティの定量を行った。

・ 投稿に付加された「いいね」を用いた文化的サービスの間接的利用の推定

国立公園の豊かな生態系はソーシャルメディアを通じて多くの人々に間接利用されている可能性がある。そこで、当初の計画に追加して国立公園の間接利用の評価を行なった。ソーシャルメディアを通じた間接利用により好意や共感、感動といったポジティブな影響を受けた人々の多くは、その投稿に対して「いいね」することで気持ちを表明する。そこで、Twitterに投稿された位置情報付きの写真の「いいね」獲得数を、その場所の間接利用の指標とし、国立公園の保護レベルの高いエリアは間接利用が多いという仮説を検証した。写真は機械学習サービス（Google Cloud Vision）で解析し分類することで、自然に関する写真の投稿に限定した。国立公園内の投稿は、地種区分に応じて3段階の保護レベルに区分した。国立公園との比較対象として、都市公園と東京ディズニーリゾートに位置情報がある自然写真を解析対象に加えた。解析は同一の投稿者が国立公園・都市公園・東京ディズニーリゾートそれ

ぞれの内と外に位置情報がある自然写真を投稿した場合の「いいね」獲得数の期待値の変化率を階層ベイズモデルにより推定した。

【国立公園の供給する文化的サービスの推定】

上記手法で整備した国立公園内の位置情報を付加された投稿、地名位置情報データベース、投稿の感情、投稿者の属性データを用い、以下の手法でレクリエーションサービスの解析を行った。

・国立公園内の利用者の空間分布・季節変化の推定

国立公園内での投稿データを3次メッシュごとに集計し、利用者数の空間分布を推定した。これを投稿本文の感情データと結合し、利用者の印象の空間分布を推定した。また、これらを月別に集計することで季節変化を推定した。

・国立公園の自然資本としての価値評価

研究計画時には国立公園利用者の居住都道府県から国立公園までの移動費用を元に、トラベルコスト法を用いて国立公園の自然資本としての価値を計算する予定であった。しかし、トラベルコスト法により推定される自然資本としての価値は宿泊費用や食費など、現地で支払われる旅行コストに大きく依存することがわかった。また、Twitterから得られた移動データも各都道府県から国立公園内への移動であり、出発都道府県内の移動を考慮することもできなかった。よって、結果が仮定に依存し不確実性が高い自然資本としての価値ではなく、移動コストそのものを経済的な価値の指標として用いた解析を行うことにした。各都道府県から各国立公園までの移動費用の計算には国土交通省が公表しているOD別交通サービス水準を利用した。

・サービス利用者数や利用者の印象、自然資本としての価値の国立公園間での違いを決める要因の解析

当初の計画では国立公園内の空間分布は考慮せずに国立公園ごとに利用者数・印象・経済的価値を集計する予定であったが、当初の想定よりも多くの投稿データが得られ詳細な解析が可能となったため、全国立公園を対象に3次メッシュ単位でこれらの指標を集計し、各国立公園内の空間分布の定量を試みた。得られたデータを用い、全国立公園を対象に利用者の属性（性別・年齢・居住地など）、自然的・社会的要因との関係を解析するための機械学習モデルを構築した。ソーシャルメディアデータの使用に伴うバイアスを考慮するため、モデルの説明変数には利用者の属性を加える予定であったが、投稿単位の解析で性別の影響が小さいこと、年齢により投稿数の空間分布パターンが異なることが示されたが、50歳以上の利用者の推定が不確実であったこと、そして50歳以上の利用者が少ないことが投稿数で見られた空間分布パターンに影響を与えた可能性もあったことから、モデルの不確実性を下げるため利用者属性は解析に用いずに、結果を解釈する際に利用者属性による影響を考慮することとした。

・国立公園内でサービス利用者数や利用者の印象、自然資本としての価値の空間分布を生み出す自然的・社会的要因の解明

サービス利用者数と印象、自然資本としての価値の空間分布を自然要因・社会要因で説明する機械学習モデルを作成した。このモデルを用いて国立公園のどのような自然的・社会的要因がレクリエーションサービスの利用や印象、自然資本としての価値を向上させるのかを明らかにした。

4. 結果及び考察

【サブテーマ1】ビッグデータと機械学習を用いた国立公園の文化的サービス評価

・ソーシャルメディア情報の信頼性

○ 国立公園訪問者指標としての投稿データの信頼性

国立公園ごとに集計した国立公園内の位置情報を持った投稿を行った利用者数と国立公園等利用者数調から得た国立公園ごとの利用者数との関係を見ると、両者の絶対数には大きな違いが見られた一方、両者の間には強い正の相関が見られた（図-1.1）。この結果はTwitterへの投稿者数は国立公園利用者数の相対的な指標として利用可能であることを示している。また、得られたデータを用いることで、国立公園内の詳細な利用の空間分布（図-1.2）や時間変化（図-1.3）を得ることができることがわかった。

NTTデータ社の手法は高齢ユーザーを検出しにくい傾向がある可能性が考えられる。

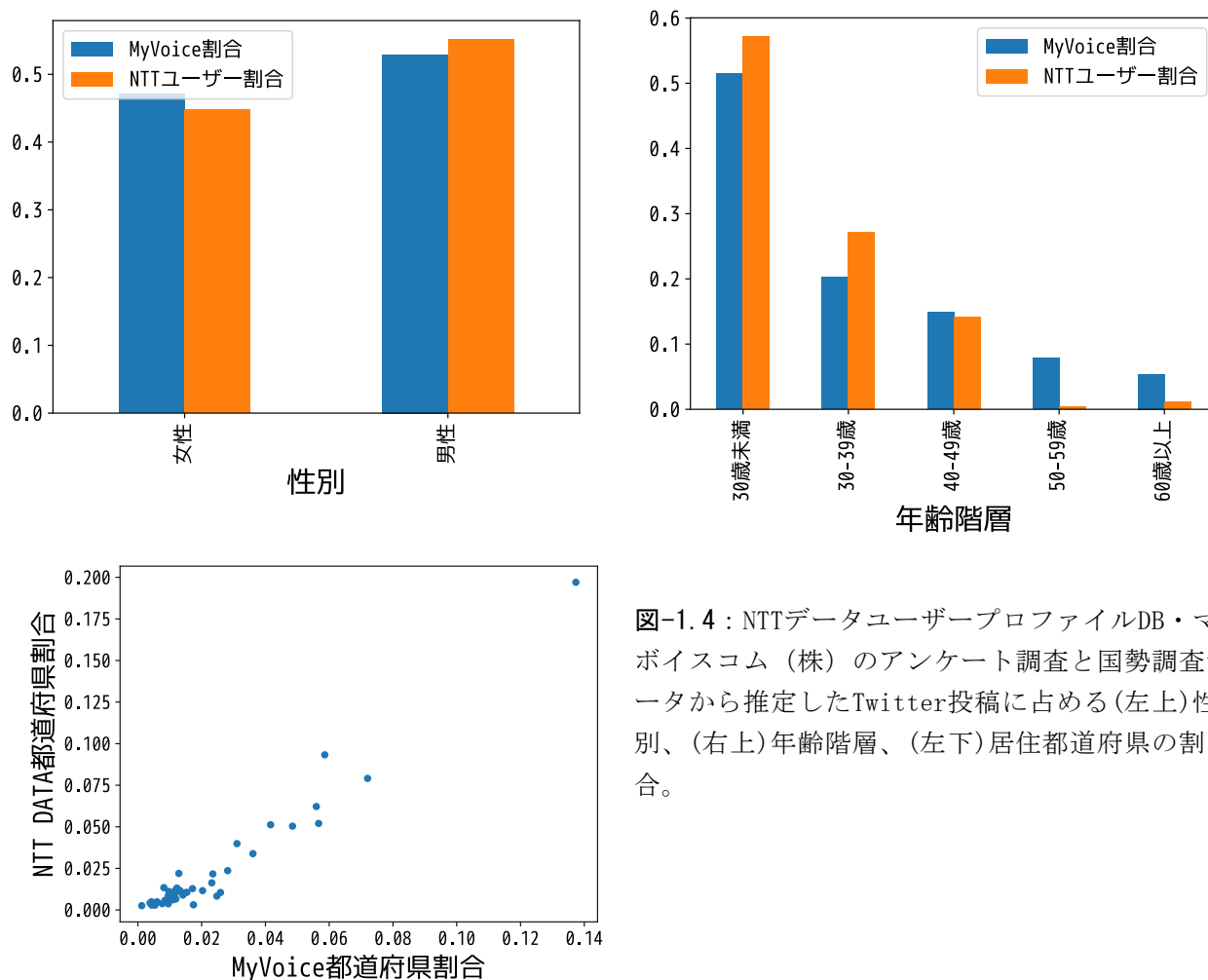


図-1.4：NTTデータユーザープロフィールDB・マイボイスコム（株）のアンケート調査と国勢調査データから推定したTwitter投稿に占める（左上）性別、（右上）年齢階層、（左下）居住都道府県の割合。

○ ソーシャルメディアデータ利用によるバイアスの定量

ソーシャルメディア情報を使った解析には特有のデータのバイアスが生じるが、実際どのようなバイアスが生じているのか定量された事例は少ない。そこで、(1)ランダムに抽出した投稿者、(2)位置情報付加した投稿を行った投稿者、(3)国立公園の位置情報を付加した投稿を行った投稿者の属性を、NTTデータユーザープロフィールDBを用いて推定した。

推定された性別割合（図-1.5）を見ると、解析不能（投稿者が削除されたなどの理由で属性が取得できなかった投稿者）・不明（解析は可能だったが性別が推定不能だった投稿者）の割合はランダムに抽出した投稿者では4割程度であったが、位置情報付加した投稿を行った投稿者では8割程度の投稿者の性別が推定できた。また、ランダムに抽出した投稿者は男性の割合が若干高いが、位置情報付加した投稿を行った投稿者・国立公園投稿を行った投稿者は男性の割合が圧倒的に高いことがわかった。一方、2015年にイギリスで行われた先行研究(Sloan et al. 2015)は位置情報を付加する投稿者に性差はないという結果を報告している。本研究との結果の違いは国による社会状況の違いや調査時期の違いによるプライバシー認識の変化が影響していると考えられる。

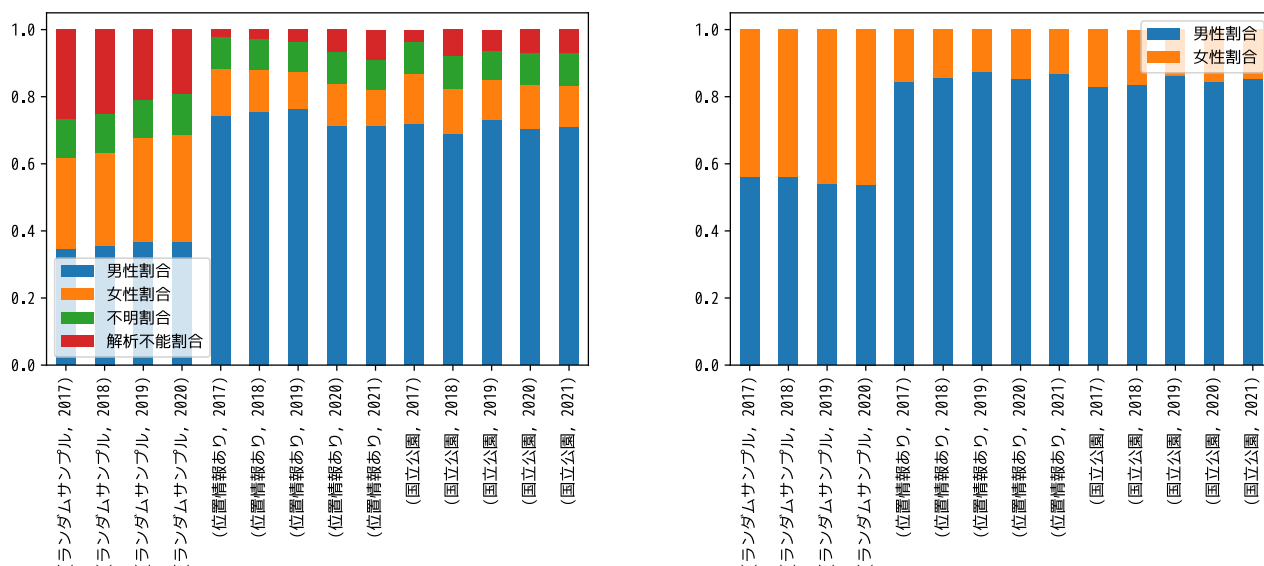


図-1.5：推定された(1)ランダムに抽出した投稿者、(2)位置情報付加した投稿を行った投稿者、(3)国立公園の位置情報を付加した投稿を行った投稿者の性別割合。左側のパネルは推定不能・不明を含み、右側は推定が行えた投稿者だけの比率を示している。データは年別に示した。

推定された年齢の割合（図-1.6）を見ると解析不能・不明の割合は4割程度であり、位置情報のありなしによって大きな違いは見られなかった。ランダムに抽出した投稿者では30代以下が多かったが、それと比較すると位置情報を付加する投稿者は若干40代以上の割合が高くなっていた。Zenlyなど位置情報共有専用のアプリは主に若年層に使われることが知られており、得られた結果は位置情報を共有するソーシャルメディアが年齢層によって異なることを反映しているのかもしれない。その一方、位置情報が付加された投稿者だけを見ても、依然として日本の人口分布よりも人口構成は若年層に偏った分布をしていた。よって、この結果からもTwitterを対象に用いた解析結果は若年層の傾向をより反映しがちであることがわかった。

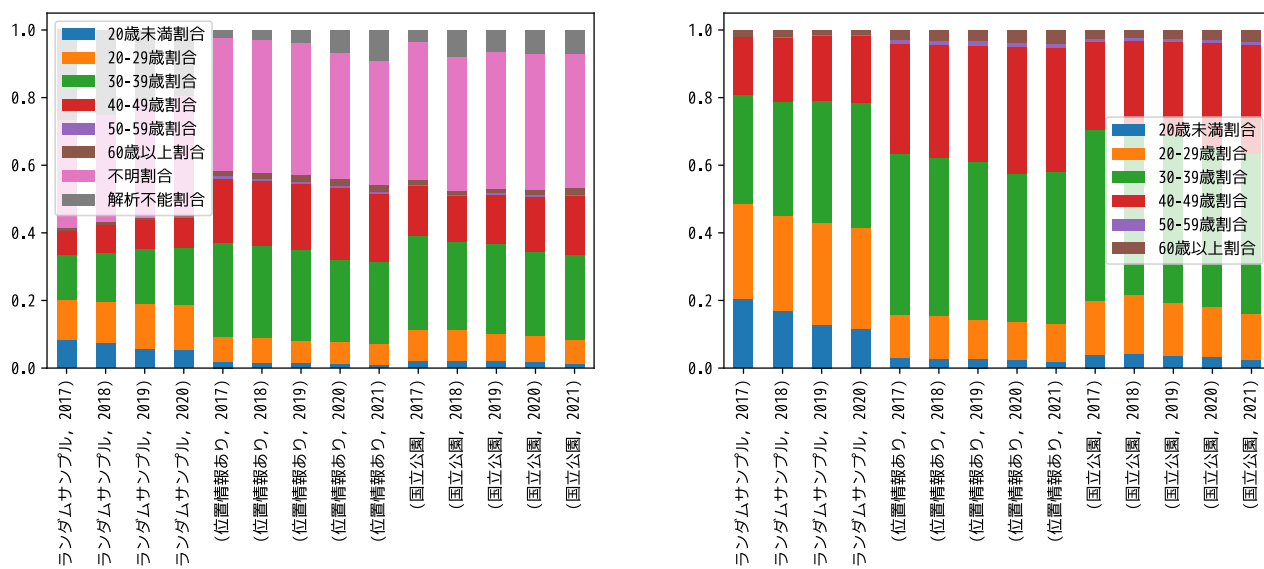


図-1.6：推定された(1)ランダムに抽出した投稿者、(2)位置情報付加した投稿を行った投稿者、(3)国立公園の位置情報を付加した投稿を行った投稿者の年齢割合。左側のパネルは推定不能・不明を含み、右側は推定が行えた投稿者だけの比率を示している。

図-1.7に推定された居住都道府県ごとの投稿者の割合を示した。ランダムに抽出した投稿者の9割は居住都道府県が推定不可能もしくは不明だったが、位置情報を付加した投稿を行った投稿者は4割程度で

居住都道府県が推定可能だった（図-1.7左側）。NTTデータ社のシステムの詳細は公開されていないが、このような結果が得られた背景としては、NTTデータ社のシステムで内部的に位置情報が利用されている、もしくは位置情報を公開する投稿者ほど居住地に関する投稿も行いやすい傾向があるといった理由が考えられる。一方、居住都道府県が推定可能であった投稿者のデータ（図-1.7右側）を見ると、3つの投稿者の抽出基準にかかわらず、各都道府県の割合はほぼ一定だった。この結果は居住都道府県の推定にはデータ抽出過程によるバイアスがあまり生じていないことを示していると考えられる。

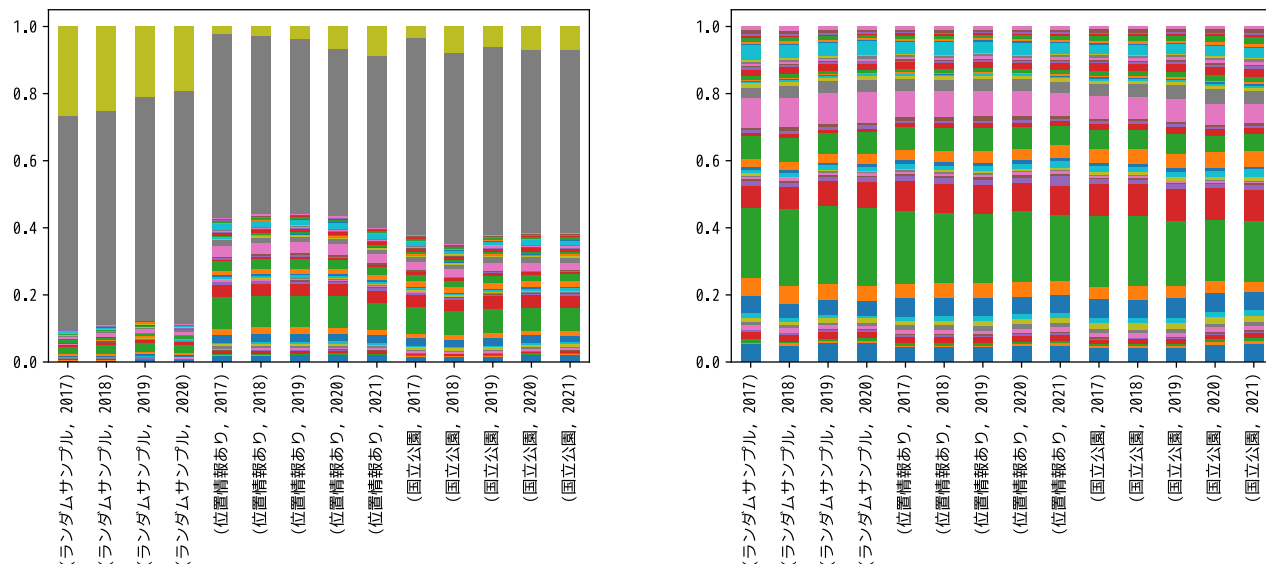


図-1.7：推定された(1)ランダムに抽出した投稿者、(2)位置情報付加した投稿を行った投稿者、(3)国立公園の位置情報を付加した投稿を行った投稿者の居住都道府県の割合。左側のパネルは推定不能・不明を含んでいる。識別が難しいため凡例は省略したが、最上段の黄色の部分推定不能を、二段目の灰色部分は居住都道府県が不明なユーザーの割合を示している。一方、右側のパネルには居住都道府県の推定が行えた投稿者だけの比率を示している。

○ ソーシャルメディアデータ利用によるバイアスの影響の定量

上記の解析結果でTwitterに投稿された位置情報を持った投稿を用いて国立公園の文化的サービス評価を行うと、居住都道府県の割合には代表性が期待できる一方、

- ・ 位置情報を付加した投稿を行う投稿者は男性が多いため、結果は男性の情報をより強く反映する
- ・ Twitter投稿者は若年層が多いため、若年層の情報をより強く反映する

といったバイアスが生じていることがわかった。そこでこれらの投稿者属性が評価結果に与える影響を定量するため、本研究で経済的価値の指標として用いた利用者の移動コスト、感情スコアが投稿者の属性でどの程度説明できるのかを統計モデルを用いて解析した。

移動コストへの年齢・性別の影響を図-1.8左側に示した。居住都道府県は移動コストに影響することが明らかであるため、解析に加えなかった。解析の結果、年齢・性別ともに移動コストへの影響は小さく、Twitterデータに見られる属性の偏りは移動コストを解析する結果にあまり影響しないことがわかった。また、投稿者の属性・投稿時間（曜日・月・コロナ前後）・位置情報が付与された国立公園・保護区分・添付画像ありなし・投稿者の個人差も含めて行った解析ではモデルの R^2 は0.76と高い値となり、利用者の移動コストのばらつきはこれらの変数、特に位置情報が付与された国立公園で説明できることがわかった。よって、ソーシャルメディアから推定した移動コストを用いた解析は上記の性別や年齢層によるバイアスの影響を受けにくいことがわかった。

一方、感情スコアを同様に解析すると、高齢利用者および女性が若干ポジティブであるという結果が得られた（図-1.8右側）。しかし、移動コストと同様の変数を用いた解析でもモデルの R^2 は0.1程度と低い値となり、さらに投稿者の個人差を考慮しない場合、 R^2 は0.01となり、ほとんどデータのばらつきを説明できなかった。よって感情に関してもソーシャルメディア情報の投稿者属性のバイアスは結果

に大きな影響を与えないことがわかった。

3次メッシュ単位で投稿数を集計し、全投稿数と性別・年齢ごとの投稿数の相関を見ることで投稿数の分布へバイアスの影響を確認した。全投稿数と性別ごとの投稿数は相関が強かったが（図-1.9右）、年齢階層ごとの投稿数を見ると50代以上の投稿者による投稿数は全投稿数との相関が弱かった（図-1.9左）。この結果は投稿数を用いて利用頻度を推定すると、高齢者の傾向を過小評価する可能性を示している。

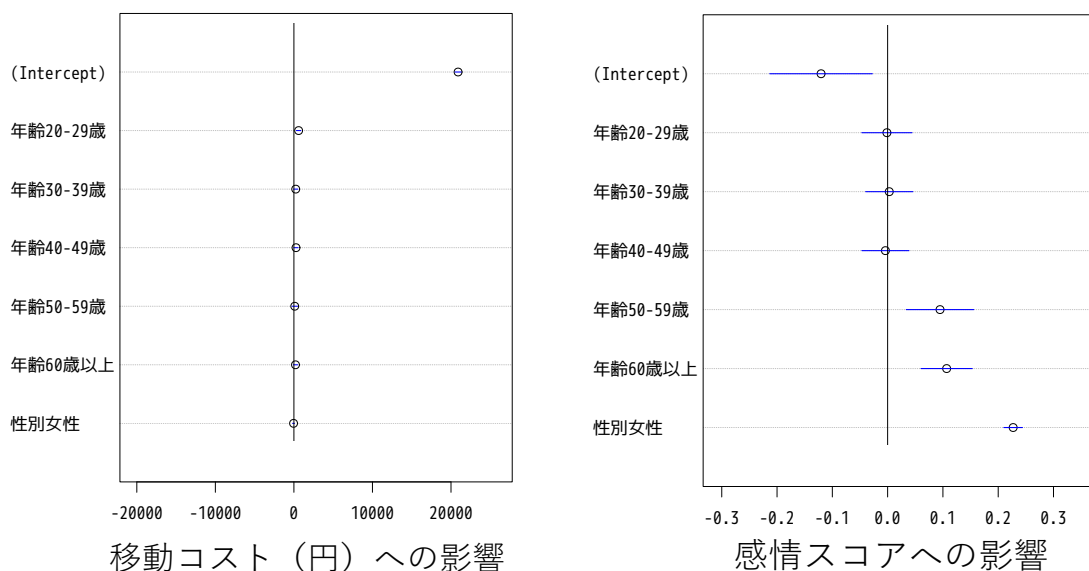


図-1.8：移動コスト・感情スコアに与える投稿者の年齢・性別の影響。感情スコアは値が高いほどポジティブな感情の投稿であること、値が低いほどネガティブな感情の投稿であることを示す。Interceptはモデル全体の切片を表しており、10代男性の国立公園への移動コスト・感情スコアの平均を意味する。また、他の点は10代の投稿者と各年齢階層の投稿者の移動コスト・感情スコアの差、男性と女性の移動コスト・感情スコアの差を表し、点のエラーバーは推定値の信頼区間を表す。

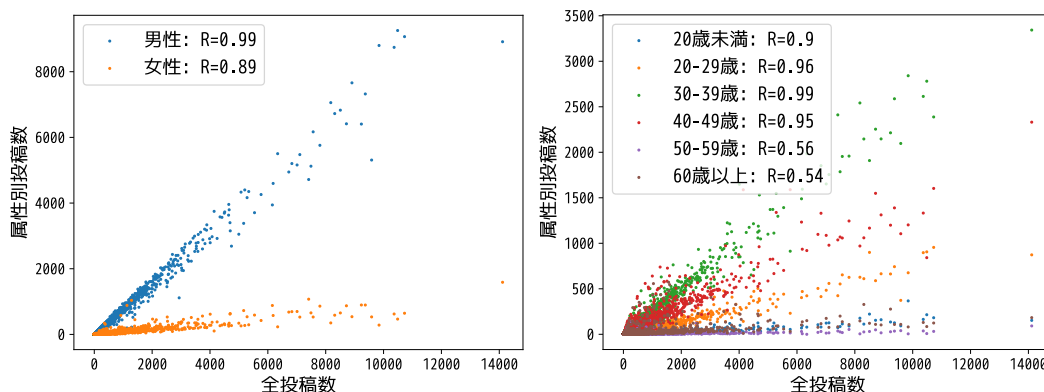


図-1.9：3次メッシュごとに集計したTwitterへの全投稿数と投稿者の性別・年齢階層ごとの投稿数の関係。凡例には相関係数も示した。

・国立公園訪問が利用者の感情に与える影響の解明と感情の季節変化

国立公園への訪問そのものが利用者の感情に与える影響を調べるため、国立公園の位置情報を付加した投稿を行った投稿者による国立公園の位置情報が付加されていない投稿を収集し、感情スコアを比較した。その結果、IBM社のWatson Natural Language Understandingを用いた結果（図1-10）では国立公園内の投稿はポジティブな感情を持つ投稿が多い一方、国立公園の位置情報が付加されていない投稿はネガティブな感情の投稿が多いことがわかった。また、日本語感情分析システムMLAskを用いた解析（図1-11）でも国立公園の位置情報が付加された投稿は「哀」「厭」「怒」など、ネガティブな感情を

持つ投稿が少ない一方、「喜」の感情を持つ投稿が多いことがわかった。これらの結果から国立公園への訪問は利用者の感情にポジティブな影響を与えていることがわかった。その一方、国立公園では「好」の感情が少なくなっていた。MLAskで「好」の感情に分類されている単語を見ると、「可愛い」「大好き」など、その多くは人に対する好意を表す単語であった。国立公園では利用者の注意・関心は人ではなく自然的な対象に向けられるため、このような結果が得られたのかもしれない。またどちらの手法を用いた結果を見ても感情を持った投稿の割合に大きな季節性は見られなかった。

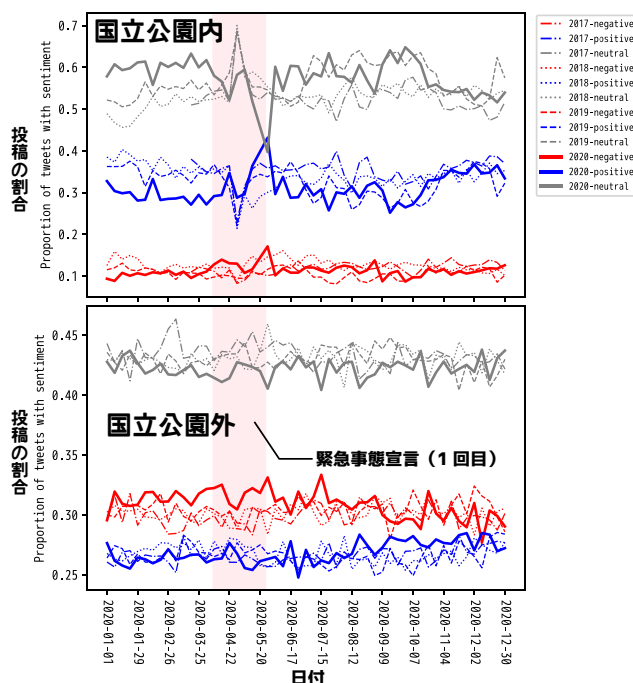


図1-10：国立公園内の投稿と同じ投稿者が行った国立公園外の投稿の感情（赤：ネガティブ・青：ポジティブ・灰：感情なし）。各線は各年の感情を含む投稿の割合の時間変化を示す。背景がピンクの部分は2020年の新型コロナウイルス対応の1回目の緊急事態宣言期間で、この期間は国立公園内の投稿が少なく、国立公園内データの信頼性が低いことに注意が必要である。国立公園内はポジティブな感情を持った投稿が多く、ネガティブな感情を持った投稿が少ないこと、感情には大きな季節性が見られないことがわかる。

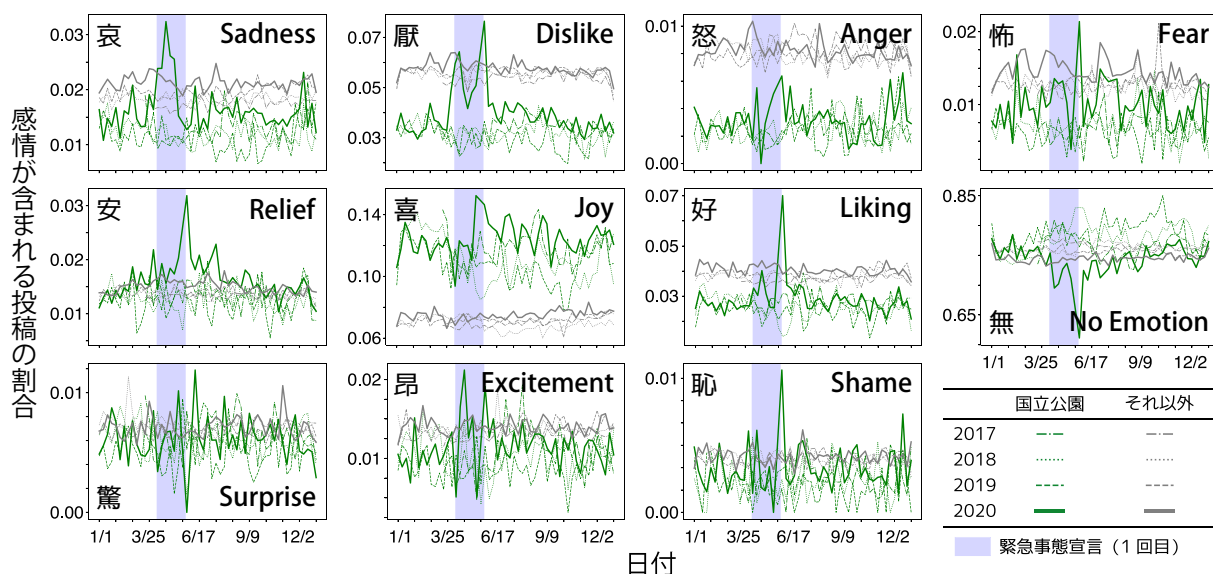


図-1.11：国立公園内の投稿と国立公園外の投稿の感情の違い。線の違いは年の違いを表し、緑の線は国立公園内を灰色の線は国立公園外の投稿の感情の季節変化を表す。背景が青色の部分は2020年の新型コロナウイルス対応の1回目の緊急事態宣言期間で、この期間は国立公園内の投稿が少なく、国立公園

内のデータの信頼性が低いことに注意が必要である。

・投稿された写真に基づくアクティビティの抽出とその季節性

Googleが提供する画像認識AIを利用し、国立公園内の投稿に付加された写真にどのような要素が写されていたのかを分析した。その結果、写真に写される要素が国立公園により大きく異なること、そして写される要素が季節により変化すること、さらに季節性は公園によって異なることが示された（図-1.12）。これらの写真に写された要素に基づき8種類のアクティビティを抽出することで、それぞれの国立公園でどのようなアクティビティに関連する写真が投稿されているのかを客観的に評価することができた（図-1.13）。さらに、アクティビティ別の写真投稿数に基づいたクラスター分析により、各国立公園を特徴づけるアクティビティを評価することができた（図-1.14）。

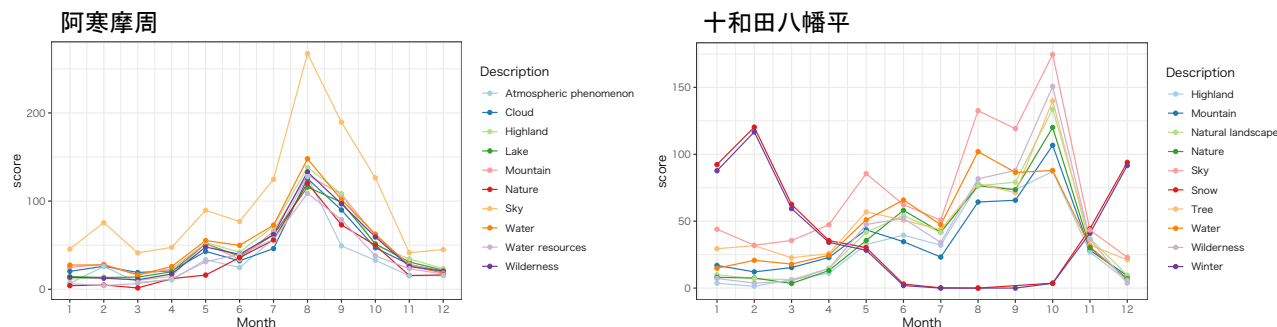


図-1.12：阿寒摩周国立公園および十和田八幡平国立公園でTwitterに投稿された写真に写されていた要素の月別変化。阿寒摩周国立公園では8月に写真投稿数のピークがあり、空や水、高原、湖、山などに関連する写真の投稿が多かった。十和田八幡平国立公園は2月と10月にそれぞれ写真投稿数のピークが見られ、2月には雪に関連する写真の投稿が多く、10月には樹木や原生自然、自然景観などに関連する写真の投稿が多いことがわかる。

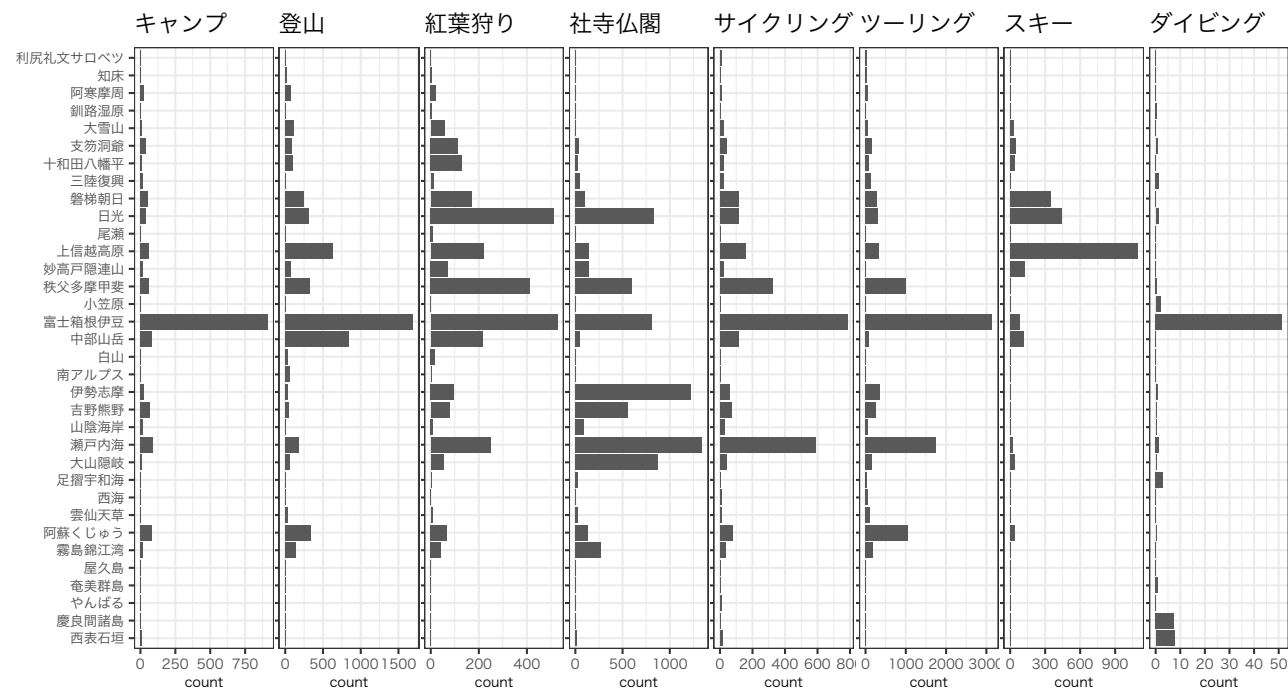


図-1.13：各国立公園における写真に写された要素から抽出したアクティビティ別の写真投稿数。

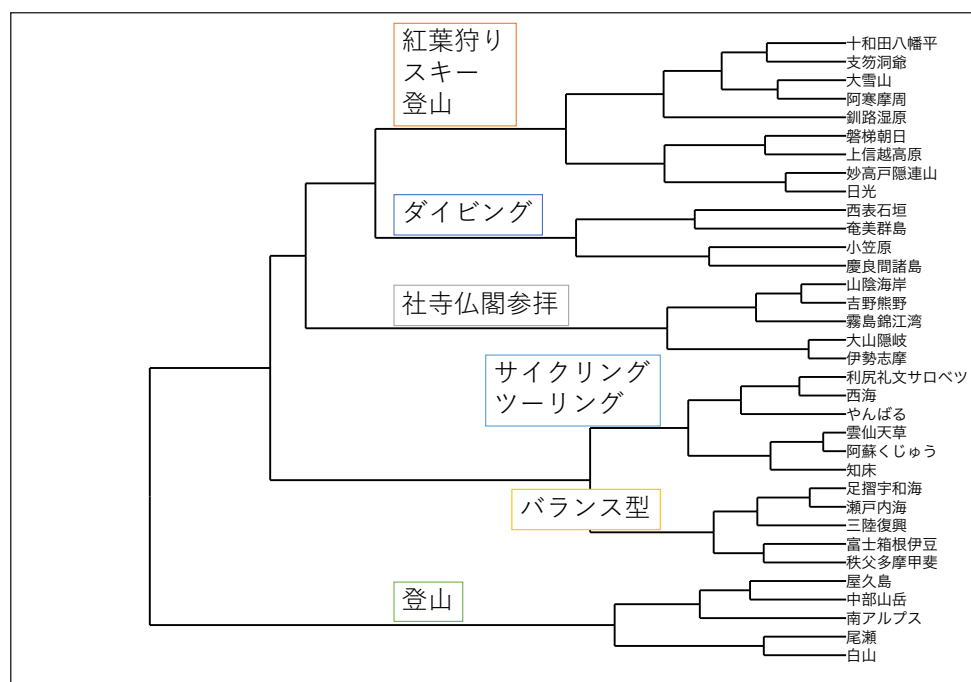


図-1.14：各国立公園のアクティビティ別の写真投稿数に基づいたクラスター分析。国立公園利用者が写真に投稿したアクティビティの種類が似通った国立公園が近いグループに分類されている。

・地名位置情報データベースの構築とアウトドア関連スポットの定量

Foursquareに登録されていたアウトドア関連スポットのうち、国立公園内に位置する計8384スポットについて詳細な情報を取得し、データベース化した。データベースに登録したアウトドア関連スポットはさらに細かく小カテゴリーが割り振られており、それぞれ多い順に「山」が1498スポット、「展望所」が840スポット、「海岸」が785スポット、「温泉」が660スポット、「トレイル」が577スポットであった。抽出されたアウトドア関連スポット数やその小カテゴリーの内訳は国立公園ごとに大きく異なり、各国立公園の自然景観の特徴がよく表れていた（図-1.15）。

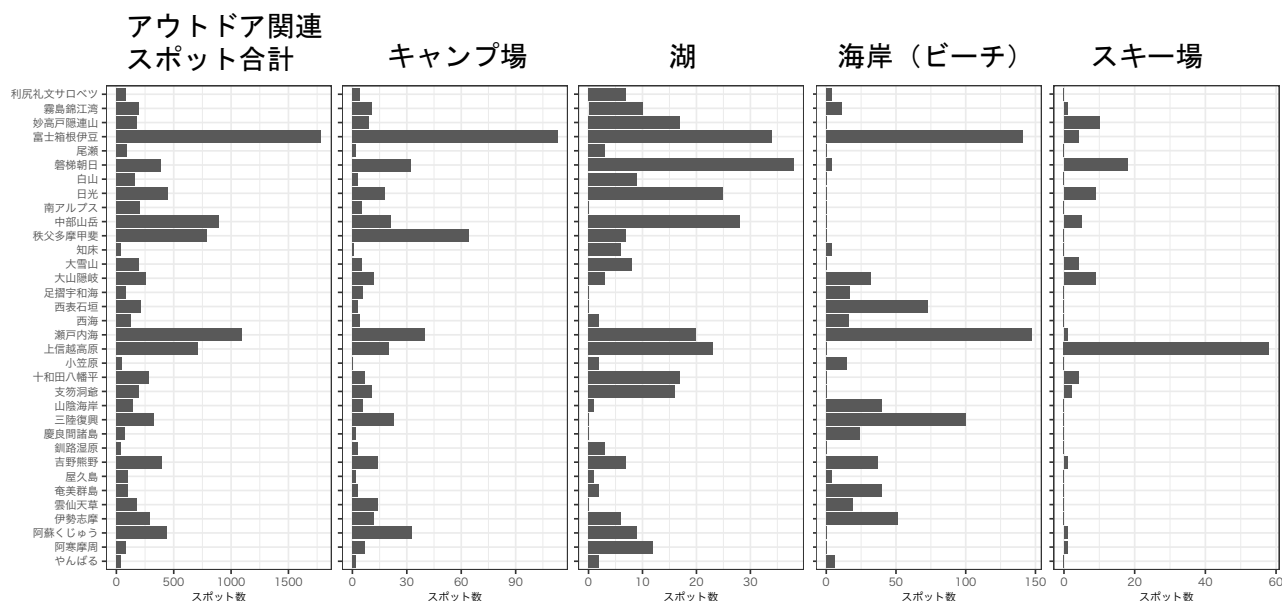


図-1.15：Foursquareから抽出した国立公園ごとのスポット数。左から大カテゴリーが「アウトドア（Outdoor）」で登録されていたスポットの合計数、大カテゴリー「アウトドア」に含まれる小カテゴリーの例として「キャンプ場（Campground）」、「湖（Lake）」、「海岸（Beach）」、「スキー場（Ski Area）」で登録されていたスポット数をそれぞれ示している。

・国立公園における保護レベルがソーシャルメディアを通じた間接利用に与える影響

国立公園の自然は訪問による直接利用に加えて、本やテレビ、ソーシャルメディアに投稿された写真を通じて訪問していない人々にも間接的に利用されている。特に、特別保護地区などの保護レベルが高いエリアは原生林や山頂などアクセスの悪い場所が多く、また利用調整地区などの利用者数の制限を行なっている場所も含まれる。よって、直接利用のみを指標とした評価は国立公園の価値を過小評価する恐れがある。そこでTwitterに投稿された写真を見た人々が付加した「いいね」を指標にすることで、実際には訪問していない人々の国立公園生態系の間接利用を評価した。国立公園により傾向にばらつきは見られたが、平均すると特別保護地区で投稿された自然写真は同じ投稿者が国立公園以外の場所で投稿した自然写真と比較して「いいね」獲得数の期待値は約1.2倍に増加した。普通地域と特別地域においても増加傾向は見られたが、増加率は約1.1倍と小さく、統計的にも有意ではなかった。比較対象とした都市公園の「いいね」獲得数は低下傾向、東京ディズニーリゾートは年により傾向が大きく異なるという結果となった（図-1.16）。以上の結果は、同じ自然写真であってもより保護レベルが高く原生的自然が残されたエリアの写真ほどTwitter上で共感や感動を集めており間接利用が多いこと、都市公園のような人工的な自然景観は身近な訪問先ではあるがTwitter上での共感や感動は得られ難く間接利用が少ないことを示している。これらの結果は、国立公園の訪問者数の増加や利用者の利便性の向上を目的に安易に生態系に手を加えることは、ソーシャルメディアなどを通じた間接利用を減少させ、総合的には国立公園の価値の低下に繋がりがかねないこと、そして国立公園の利用と保護は必ずしも対立するものではないことを示している。

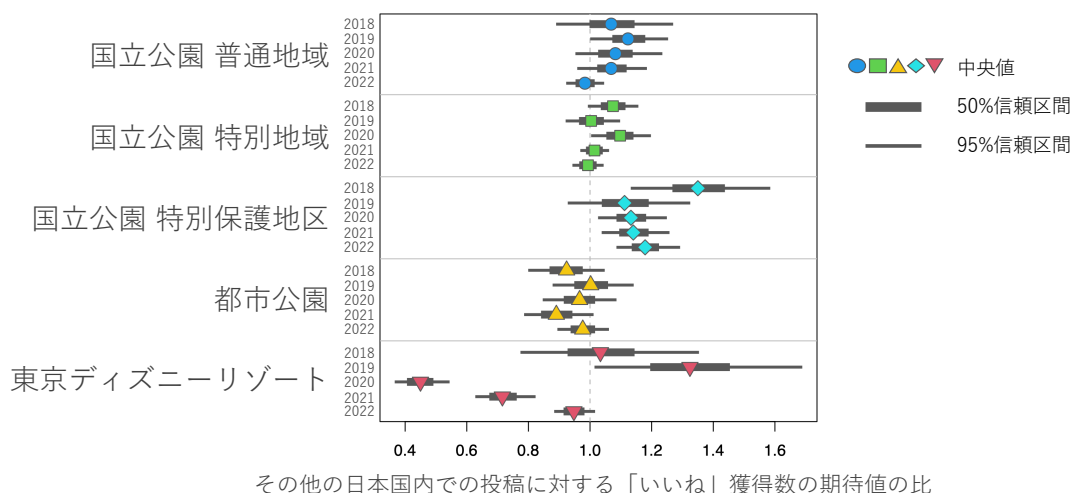


図-1.16：同じ投稿者が国立公園の各エリア、都市公園、東京ディズニーリゾート内の位置情報が付与された自然写真を投稿した場合と、その他の日本国内の位置情報が付与された自然写真を投稿した場合の「いいね」獲得数の期待値の比。95%信頼区間が1.0の点線から離れている場合に期待値が統計的に有意に増加（または減少）したことを示している。

・国立公園内でサービス利用者数や利用者の印象、自然資本としての価値の空間分布を生み出す自然的・社会的要因の解明

サービス利用者数と印象、自然資本としての価値の空間分布を生み出す要因を調べるため、全国の国立公園・各国立公園の3次メッシュを対象とした機械学習モデルをそれぞれ作成した。全国の国立公園を対象としたモデルでは、サービス利用者数はFoursquareに登録されていた観光スポット数や道路密度などの重要度が高く、観光スポットが存在しアクセス性も良い場所の利用者数が多いことがわかった（図-1.17-1, 2, 3）。また、重要度は高くはないものの原生植生率や特別地域率、特別保護地区が高い場所、つまり自然度が高いエリアほど利用者数が多かった（図-1.17-6, 13, 16）。

自然資本としての価値を合計移動コスト（各メッシュで投稿を行った人の訪問にかかる移動費用の合計）を対象に解析すると、観光スポットが多い場所や人口密度が高い地域で合計移動コストが高くな

った一方（図-1.18-1, 2, 3）、人口密度が低い場所、沖縄や北海道（高緯度・低緯度）、保護地域の面積割合が高いほど合計移動コストが高いことがわかった（図-1.18-3, 4, 12, 13, 16）。これらの結果は利用者数や自然資本としての価値の決定要因として観光スポットの存在やアクセス性が最も重要ではあるが、これらが同様の条件であれば手付かずの自然が残されているほど遠くから多くの利用者が訪問することを示している。一方、利用者の印象はモデルの予測力が非常に低く、利用者数や移動コストと同じ要因では説明できないことが明らかとなった。

各国立公園を対象としたモデルでは全国の国立公園を対象とした場合と同様の傾向が見られたが、それぞれの国立公園に特徴的な傾向も見られた。サービス利用者数については全国の国立公園を対象とした場合と同様に観光スポット数やアクセス性、特別地域率、特別保護地区率に加えて、湖や河川などの内水面率、標高や地形などの重要度が高い国立公園が多かった。合計移動コストは利用者数と同様の結果だったが、各メッシュへの平均移動コストをみると人口密集地から離れており、原生植生率や特別地域・特別保護地区率が高いほど平均移動コストが高い国立公園が多かった（図-1.19）。利用者の印象はほとんどの国立公園で予測力が非常に低かったが、一部の国立公園では予測可能だった。例えば南アルプス国立公園の尾根・山頂エリアや尾瀬国立公園の尾瀬ヶ原・尾瀬沼周辺などその国立公園を代表する景勝エリアで相対的に印象がポジティブとなる傾向が得られた（図-1.20）。

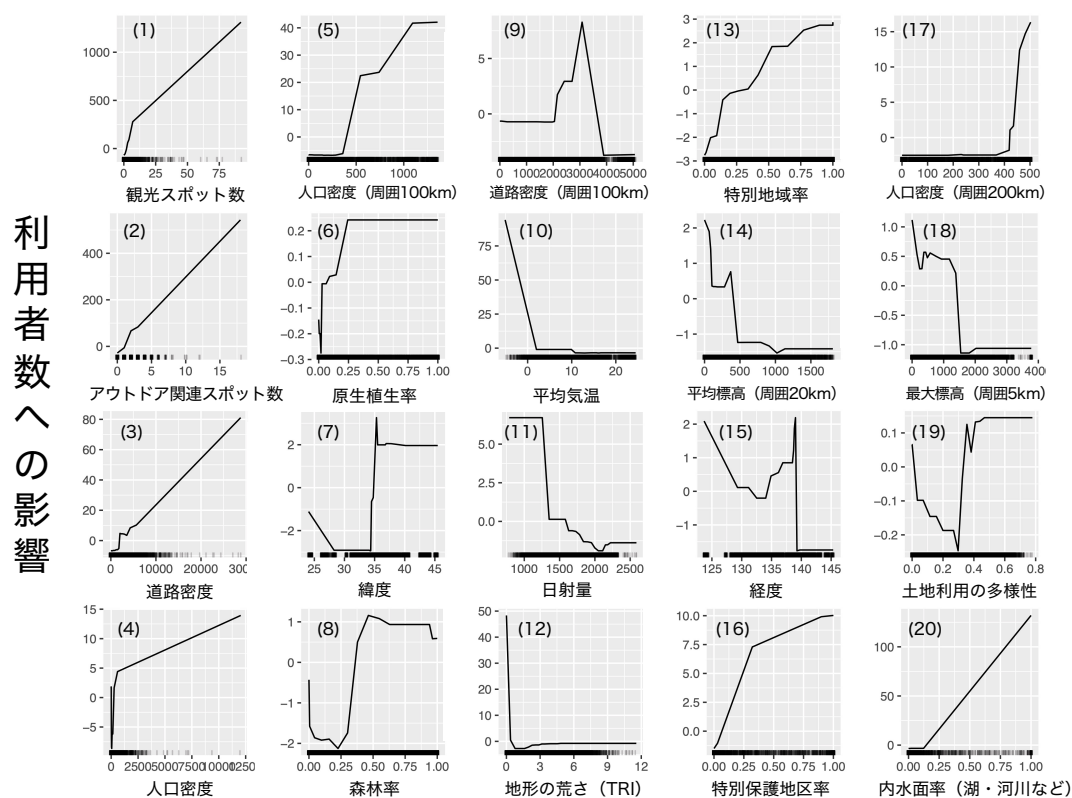
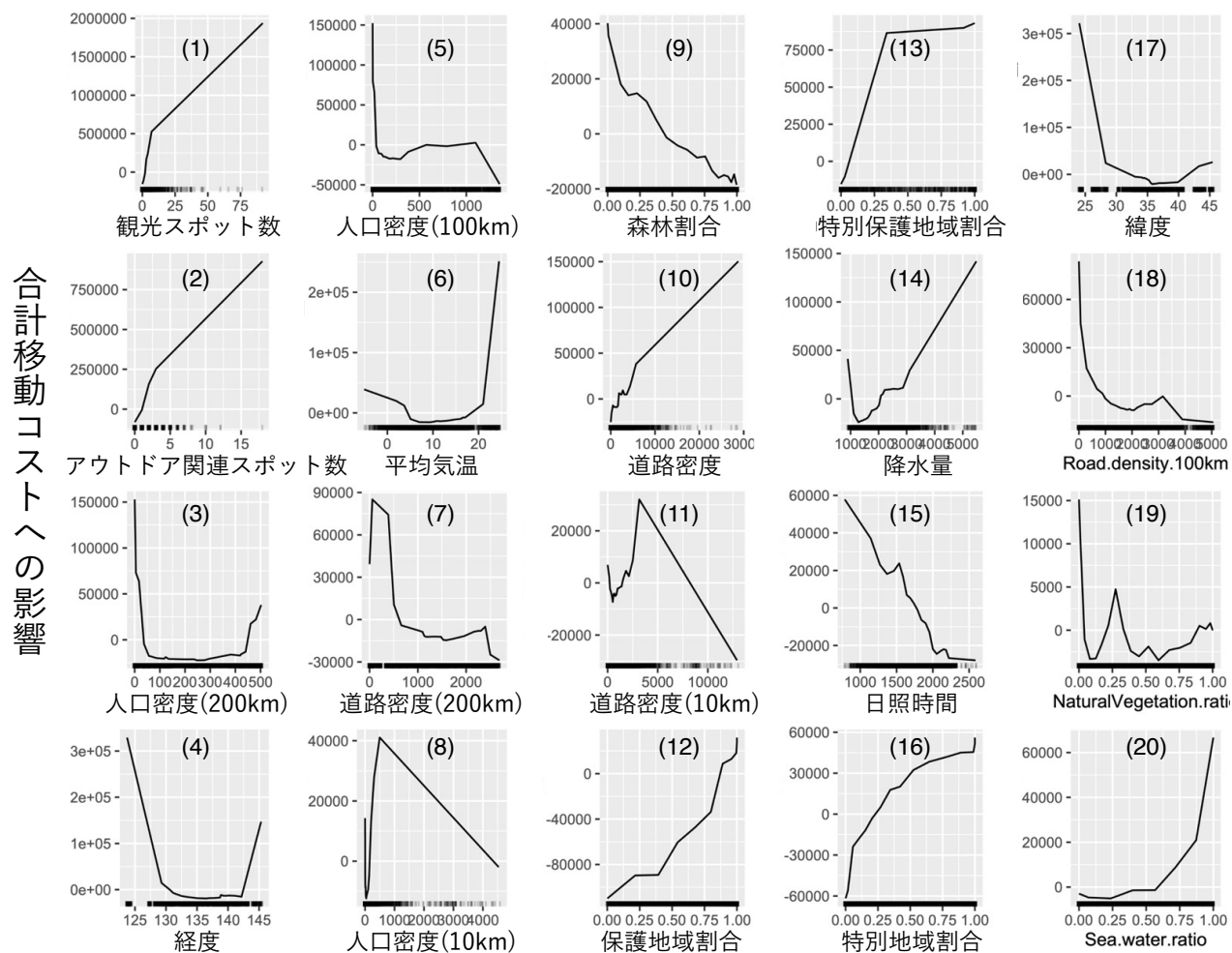
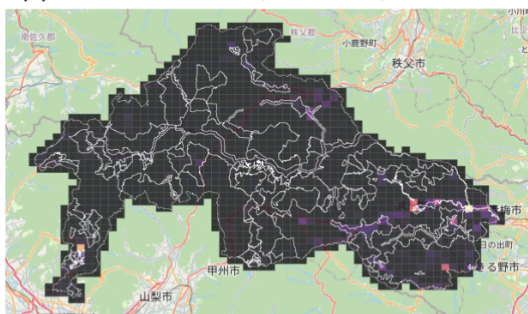


図-1.17：自然的・社会的要因がサービス利用者数に与える影響の機械学習による推定値。重要度の上位から20個の要因を示した。黒の曲線が右肩上がりの場合はその要因の値が大きい3次メッシュほどサービス利用者数が多い傾向があり、右肩下がりの場合には逆の傾向があることを意味する。

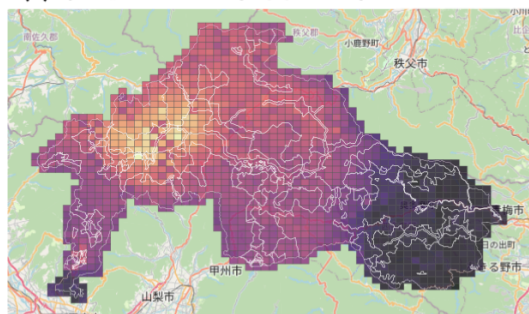


移秩父多摩甲斐国立公園

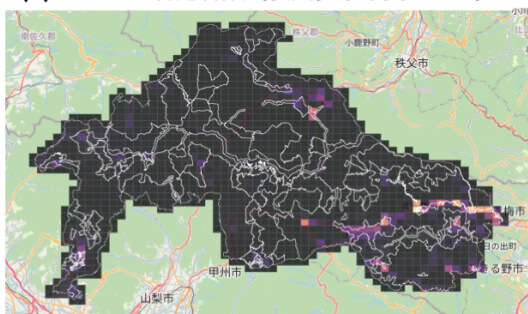
(a) サービス利用者数（全国モデル）



(c) 平均移動コスト（全国モデル）



(b) サービス利用者数（秩父多摩甲斐モデル）



(d) 平均移動コスト（秩父多摩甲斐モデル）

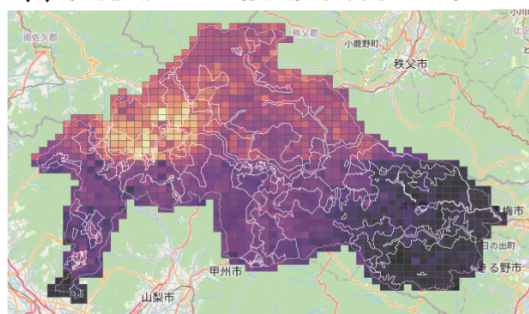
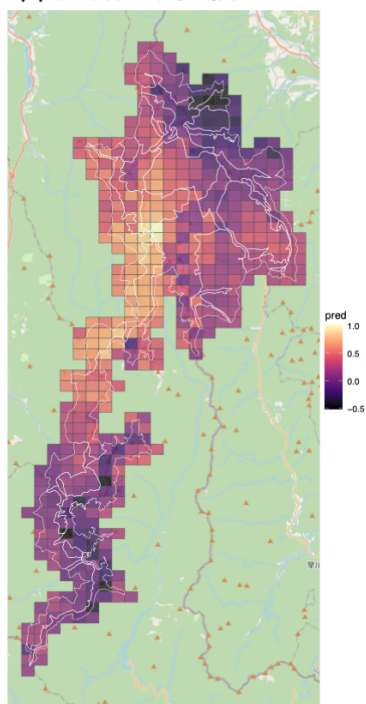


図-1.19：機械学習によるサービス利用者数、平均移動コストの推定値の空間分布（秩父多摩甲斐国立公園の例）。全国を対象としたモデル(a, c)と各国立公園を対象としたモデル(b, d)をそれぞれ示している。両モデルの予測結果は同様に、利用者数は観光スポットの空間分布に大きく依存しており、平均移動コストは大都市圏からの距離が遠く原生的な自然が多く残されているエリアで高くなっている。

(a) 利用者の印象（南アルプスモデル）



(b) 利用者の印象（尾瀬モデル）

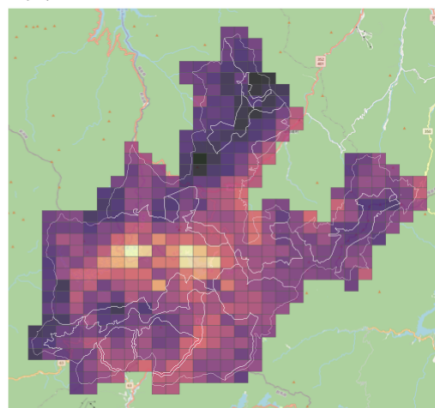


図-1.20：機械学習による利用者の印象の推定値の空間分布（南アルプス国立公園、尾瀬国立公園の例）。色が明るい場所は暗い場所と比較して相対的にポジティブな投稿が多いことを示している。

5. 研究目標の達成状況

目標を上回る成果をあげた。

国立公園の供給するレクリエーションサービスの利用頻度・経済的価値の指標としての利用者の移動コストの空間情報を作成し、利用者属性・自然的要因・社会的要因との関係を解明した。一方、印象の指標としての感情は同様の解析を行った結果、利用頻度や移動コストと同じ要因では説明ができないことを解明した。

また、当初研究計画にはなかった投稿に添付されている画像や観光スポット情報からスキーや紅葉など、各国立公園で積極的に行われているアクティビティを抽出する技術の開発を行った。さらに当初の研究計画にはなかった文化的サービスのインターネット経由での間接利用の解析を行い、保護レベルの高い場所ほどサービスの間接利用が多いことを明らかにした。

以上のとおり当初計画の研究目標を達成した上で、さらに当初計画にはなかった技術開発を積極的に行ったため、目標を上回る成果を上げたと判断した。

6. 引用文献

1. Ptaszynski, M., et al. (2009). A System for Affect Analysis of Utterances in Japanese Supported with Web Mining. 知能と情報 21(2): 194-213.
2. Sloan, L. and J. Morgan (2015). Who Tweets with Their Location? Understanding the Relationship between Demographic Characteristics and the Use of Geoservices and Geotagging on Twitter. PLoS ONE 10(11)

Ⅲ. 研究成果の発表状況の詳細

(1) 誌上発表

<査読付き論文>

【サブテーマ1】

特に記載すべき事項はない。

<査読付き論文に準ずる成果発表>

特に記載すべき事項はない。

<その他誌上発表（査読なし）>

【サブテーマ1】

特に記載すべき事項はない。

(2) 口頭発表（学会等）

【サブテーマ1】

1. 柴田嶺・小黒芳生・饗庭正寛・中静透（2021）ソーシャルメディア分析による野外レクリエーションサービスの全国評価．日本生態学会第68回大会（2021.03，オンライン開催）
2. 柴田嶺・小黒芳生（2021）ソーシャルメディア分析による日本の各国立公園の魅力の評価．第132回日本森林学会大会（2021.03，オンライン開催）
3. 小黒芳生・柴田嶺（2021）ころなういるす vs 生態系サービス ～レクリエーションの逆襲～．日本生態学会第68回大会（2021.03，オンライン開催）
4. 饗庭正寛・小黒芳生・柴田嶺：第69回日本生態学会大会，福岡市（2022.3，オンライン），陸域生態系サービスの全国評価とシナリオ分析
5. Shibata, R. & Oguro, M., 7th International Symposium on Strategies for Sustainability in Food Production, Agriculture and the Environment, Niigata (2021.12, Online), The Impact of Nature Conservation on Human Well-being -Do Photos Taken in National Parks Get More 'Likes' on SNS?
6. 柴田嶺・小黒芳生：第133回日本森林学会大会，山形市（2022.3，オンライン），国立公園の利用は人々の幸福度を高めるのか？～SNS分析による検証～
7. 大庭準・柴田嶺：第133回日本森林学会大会，山形市（2022.3，オンライン），佐渡島におけるSNSを用いた文化的サービスの経済的評価
8. 大庭準・柴田嶺・小黒芳生：第134回日本森林学会大会，鳥取市（2023.3，オンライン），トラベルコスト法とSNS分析を用いた文化的生態系サービスの経済的評価
9. 柴田嶺・小黒芳生：日本生態学会第70回全国大会，仙台市（2023.3，オンライン），ソーシャルメディアの「いいね」獲得数を用いた国立公園の間接利用の評価
10. 鈴木海帆・富高まほろ・柴田嶺・小黒芳生・佐々木雄大：日本生態学会第70回全国大会，仙台市（2023.3，オンライン），日本の主要国立公園の観光スポットへの訪問満足度を左右する要因の検証

(3) 「国民との科学・技術対話」の実施

【サブテーマ1】

- 1) 小黒芳生、「生物多様性ってなあに」、学都「仙台・宮城」サイエンスデイ 2021（オンライン）にて4回講演（聴講者約30人）

- 2) 小黒芳生、「生物多様性ってななに」、学都「仙台・宮城」サイエンスデイ 2022 (仙台市) にて4回講演 (聴講者約120人)

(4) マスコミ等への公表・報道等

【サブテーマ1】

特に記載すべき事項はない。

(5) 本研究費の研究成果による受賞

【サブテーマ1】

- 1) Shibata, R. & Oguro, M., Excellent Poster Presentation Award, 7th International Symposium on Strategies for Sustainability in Food Production, Agriculture and the Environment, Niigata (2021.12, Online), The Impact of Nature Conservation on Human Well-being -Do Photos Taken in National Parks Get More ‘Likes’ on SNS?

(6) その他の成果発表

特に記載すべき事項はない。

IV. 英文Abstract

Evaluating Cultural Services in National Parks: Application of Machine Learning Methods on Social Media Data

Principal Investigator: Michio OGURO, Rei SHIBATA

Institution: Forestry and Forest Products Research Institute, Tsukuba, JAPAN

Tel: +81-(0)29-829-8223 / Fax: +81-(0)29-874-3720

E-mail: mog@ffpri.affrc.go.jp

Cooperated by: Niigata University

[Abstract]

Key Words: Ecosystem services, Cultural services, National park, Recreation, Machine learning, Big data, Social media, SNS

To investigate the interactive effects of natural and social factors on recreational services, we conducted a study using georeferenced social media posts on the national parks in Japan. By analyzing Twitter posts, we estimated the spatial distributions of frequency of use, user impressions, economic values and investigated factors affecting them. We also analyzed the effects of demographic biases brought by social media data on the indicators noted above.

The results showed that the number of posts for each national park correlated well with the existing usage statistics, indicating number of posts can be used as an indicator of usage frequency. On the other hand, the age structure of Twitter users was skewed toward younger generations, indicating results obtained in this study tended to reflect activity of younger generations.

We found that use of the national parks brought positive emotions on the park users. Also, we found that by using nature images attached to Twitter posts, we can estimate activities done by the users. Moreover, we found that pictures taken in the areas with highest protection level were indirectly used more frequently compared to the pictures taken in other areas or other kind of parks. Finally, we found that frequency of use and economic values were higher in areas with higher accessibility and higher number of tourist destinations whereas they were also higher in more natural locations. These results indicate importance of balancing use and conservation of natural areas to keep values of the national parks.