トピックモデルを用いた訪日外国人周遊分析

インバウンド観光の実態把握と政策立案のため，国土交通省は，訪日外国人流動データ（以下FFデータ）を作成した．FFデータは周遊パターンや訪日目的などの多くの属性を記録しているが，その膨大な情報を効果的に政策立案に結びつけるには，先験知識によらず，代表的なパターンを抽出する手法の開発が必要である．本研究では，トピックモデルをFFデータに適用することで，訪日外国人の周遊特性をパターン抽出するとともに，経年分析した．その結果，年次間で共通するトピックのほか，関東や関西を周遊先に含む旅行が様々な地方を周遊するように経年変化する傾向が抽出され，同手法の有効性が示された．

キーワード　**機械学習，ビッグデータ，経年変化**

辰巳嘉大
TATSUMI, Yoshihiro
広島大学院工学研究科社会基盤環境工学専攻

塚井誠人
TSUKAI, Makoto
博士（工学）　広島大学院工学研究科准教授

1——はじめに

インバウンド観光は，経済，地域の活性化といった効果が期待される産業であり，わが国にとって重要な政策課題である．政府は訪日外国人増加に向けビジット・ジャパン事業[1]，訪日促進事業[2]などを実施しており，訪日外国人数は大きく増加している．訪日外国人数は2014年で約1300万人であったが，2016年には約2400万人となった．さらなる訪日外国人数増加策を検討するには，訪日外国人数の旅行特性の変化を経年的に把握する必要がある．

国土交通省は，上述の目的に対応するため訪日外国人流動データ[3]（Flows of Foreigners data：以下FFデータ）を作成した．このデータを解析すると，国籍，滞在日数などの各属性のほか，国内の周遊ルートについて分析できる．ただし2014～2017年分が公開されているFFデータの年次別サンプル数は約4.3～6.3万，総トリップ数は約17.9～22.8万にのぼる．つまりこのデータの分析において，膨大な属性の組み合わせや周遊パターンの全てを網羅することは，事実上不可能である．そこで分析者は，着目すべき属性の組み合わせや周遊特性に関する先験知識を活用するか，あるいはこの巨大なデータをマイニングする手法を開発して，分析に当たらなくてはならない．

本研究は，FFデータから先験知識によらず代表的な属性・周遊パターンを抽出するために，トピックモデルを活用したマイニング手法を提案する．トピックモデルへの入力データは，Bag of words（BOW）形式と呼ばれる．BOWは，カテゴリ変数および離散化した階級への各カテゴリへの帰属を表すダミー変数のカウントデータである．本研究では後述する手順によって，FFデータをBOW形式に変換する．さらに各年のFFデータにトピックモデルを適用して，それぞれ代表的なパターンを抽出したうえで，経年的な訪日外国人旅行特性の変化を把握する．

2——既往研究

本節では，訪日外国人とトピックモデルに関する研究を整理し，本研究の位置づけを示す．

訪日外国人に関する研究の多くは，訪日促進のための施策の提案を目的としている．田中[4]，はリピーター獲得のため顧客満足度を向上する施策を提案した．日比野ら[5]は，入込旅客数が減少している観光地を特定した．入込旅客数が減少している観光地でガイドツアー等の体験を取り入れる施策や，公園などを整備し観光地として再開発する施策を提案した．室谷[6]は，宿泊施設数などから観光地の魅力度を評価して，宿泊施設数が少ない観光地は，近隣の観光地と連携することによってその不足を補うなどの施策を提案した．早川ら[7]は，訪日外国人の日本における公共交通サービスに対する要望・意見を調査して，訪日外国人はJRパス，東京圏のフリー切符などのサービスには，基本的に満足していることを明らかにした．そのうえで多様な訪日外国人の要望に応えるために，のぞみ号も利用可能なJRパスの追加導入などの政策を提案した．栗原ら[8]はロジスティク回帰モデルを用いてインバウンド需要に影響を与える要因を分析した．分析の結果，韓国の海外旅行自

由化とアジアの経済成長がインバウンド需要の増加に大きな影響を与えていることを明らかにした．古屋[9]，岡本ら[10]はモデルを用いてアジア諸国における将来の国際旅行者数を予測した．アジア諸国の国際旅行は将来北米，ヨーロッパ方面への割合が増加することなどを予測した．櫻井ら[11]は訪日外国人が増加した場合の国内の生産，および雇用面の経済的影響を，産業連関分析によって分析した．ただしこの研究は，訪日外国人増加による経済影響の解明と関連施策の提案を目的としており，彼らの日本国内の周遊特性は解明されていない．

以下では，訪日外国人の周遊パターンや旅行特性の解明を目的とする研究を整理する．松井ら[12]は訪日外国人消費調査の個票データを用いて，個人属性別に訪問地と観光行動のパターンの違いをクロス集計によって分析した．その結果，近年増加している個人旅行者の観光行動は，都市部を中心として観光行動が多様化する傾向を明らかにした．ただしこの研究は，代表的な個人属性と周遊行動の組み合わせに関する分析に留まっている．

多属性（多変量）データが示す特徴を抽出する方法として，主成分分析，または因子分析があげられる．香川ら[13]はバス利用者の希望目的地に関する情報を収集したうえで，主成分分析を用いて複数の希望目的地に集約した．ただし主成分分析（あるいは因子分析でも）では多数変量からの次元の圧縮を固有値に基づいて行うため，抽出される特徴ベクトルは相互に直交する範囲に限定される（因子分析で事後的に斜交回転する場合も当初は固有ベクトルを抽出）．よって特徴ベクトルを相互に直交する範囲に限定する両手法では，寄与率の低い特徴ベクトルが抽出しにくいという課題がある．

菱田ら[14]はJNTO訪日外国人消費行動調査を用いて，中国，台湾，韓国，香港，シンガポールからの訪日旅行者の訪問地を，訪日経験別にクラスター分析した．特に中国について居住地域別に旅行特性が異なる可能性を踏まえて分析をしたところ，2010年には華南を除く中国全域で訪問地パターンが変化しており，訪問地選択が多様化していることを明らかにした．なおクラスター分析を目的地周遊問題に適用すると，各目的地を表す名義変数（最も単純には目的地別ダミー変数）が現れるため高次元（多変量）になりやすく，サンプル間の類似度が低くなる球面集中現象によって，適切なクラスタリングが困難となる「次元の呪い」が起こる[15]ことが知られている．古屋ら[16]は，訪日外国人消費動向調査を用いて，潜在クラスモデルによって24の訪問パターンを見出した．そのうえで，クラス別構成比率と主要国籍・地域，旅行形態，旅行時期，訪日回数などの各要因との関連性について，一般化χ^2乗検定によって明らかにした．潜在クラス分析は定式化を工夫することによって各サンプルが確率的に複数の潜在クラスに所属する分析も可能だが，潜在クラス数の設定には試行錯誤が必要である．

FFデータはサンプル数が多いため，データマイニング手法の活用が有効と考えられる．そこで以下では，本研究で適用するトピックモデルに関する研究を整理する．塚井ら[17]は同モデルによる討議分析の可能性について，Web上で公表されている各地の地域公共交通会議の討議録データへの適用を通じて検討した．分析の結果，各地域の課題に即したトピックと地域間で共通のトピックが得られ，モデルの有効性を確認した．塚井ら[18]は，地理情報データにトピックモデルを適用して，土地利用特性の抽出を試みた．各種用地面積や，人口・世帯，事業所など23属性を入力した．その結果，これら多属性が複合した空間的な集積の傾向が明らかとなった．また因子分析とトピックモデルとの比較を行い，後者からは1）より多くのトピックを得られること，2）トピック解釈の面で優れていること，を示した．なお一連の研究では，トピック数の設定やトピック解釈手順を明確にすべきことが指摘されている．川野ら[19]はトピックモデルと離散連続モデルを結合した新たな自由記述データの分析手法を提案した．この研究では，属性別の回答傾向の違いや，選択式設問の回答と自由回答中のトピックの対応が明確なことを確認して，同手法の有用性を示した．古屋ら[20]はGPSログデータを用いて，訪日外国人旅行者の訪問場所の組み合わせパターンを分類した．

既往研究の動向と課題についてまとめる．訪日促進施策に関する研究では，旅行者の満足度向上のための着地側の施策や，宿泊施設および交通機関などの課題に関連する周遊行動分析がみられた．またマクロな国際旅客動向に関する研究もおこなわれている．訪日外国人の周遊に関する研究では，訪問地や旅行者の発地に着目した分析が行われている．その一方で，既存の旅行者の周遊地や旅行者属性の分析手法の課題として，1）集計分析では（事実上）全属性の組み合わせを網羅した検討が難しいこと，2）主成分分析や因子分析では第一段階で抽出される特徴ベクトルは，相互に直交するという制約が課されること，3）クラスター分析は目的地などの膨大な多変量データには適用し難いこと，などの課題がある．一方で後述するトピックモデルはそれらの問題がない．さらに入力データを工夫することで，訪日外国人の訪日時期，訪日目的，訪日経験回数，利用交通機関分担率，宿泊数，滞在日数，訪日手配方法といった旅行特性の全体を包括的に捉えた分析ができる．トピックモデルに関する研究では，言語情報以外にも地理情報やGPSログなどへの適用が進んでいるが，FFデータのようなアンケートデータへの適用例は見られない．

本研究では，訪日外国人の都道府県間を跨ぐ周遊行動，訪日時期，訪日目的，訪日経験回数，利用交通機関分担率，

宿泊数，滞在日数，訪日手配方法といった旅行特性について，分析者による先験的な旅行特性の仮定が不要なデータマイニング型の分析手法として，トピックモデルを適用する手順を提案して，その有効性を確認する．トピックモデルの適用に当たっては，既往研究の課題を考慮して，トピック数の決定手順や，トピックの解釈手順を明確に示す．さらに提案した手順に基づいて，訪日外国人の周遊特性の経年変化を明らかにする．なおトピックモデルには，抽出されるトピックに関連する外部情報を，教師情報として用いるAuthor topic model[21]や，トピック間の相関を明示するCorrelated topic model[22]などの応用モデルも提案されている．しかし本研究では，まずこれらよりも，より基本的なトピックモデルの適用可能性を明らかにするため，最も基本的なLDAによるデータマイニングについて検討する．

3——トピックモデルの概要

トピックモデルにはいくつかのバリエーションが存在するが，最も基本的な考え方は，確率的トピック生成：Latent Dirichlet Allocation（LDA）モデルに基づいており，岩田[23]や佐藤・奥村[24]などが紹介している．なお以下は既往研究[25]に基づいており，新規性はない．LDAでは，1文書は複数の非観測潜在トピックスによって構成されると仮定して，その推定を行う．

D個の文書集合を考える．各文書dはN^d個の語から成り，文書dのn番目の語を$\{w^{n,d}\}_{n=1}^{N^d}$とする．それぞれの語は*1-of-V*表現$w_v^{n,d}\in\{e_v\}_{v=1}^{V}$で表す．*1-of-V*表現とは，$V$個の語彙に固有番号を割り振り，各文書の$n$番目に出現した語の語彙番号が$v$のとき，ベクトル$w^{n,d}$の$v$番目の要素を1，それ以外の要素を0とする表現である[26]．$w^{n,d}$は，文書全体では$N\times V$の要素を持つ．ただしNは全単語数であり，文書別単語数N^dの和である．

LDAでは，各語彙は潜在トピック$z^{n,d}\in\{e_k\}_{k=1}^{K}$に属すると仮定する．各文書は異なるトピック分布$\tilde{\theta}_d$を，また各トピック$k$はそれぞれ異なる語彙分布$\phi_k$を持つと考える．さらに各文書内で共起する語彙のまとまりを情報の基本単位と考える．他方で文書内の語彙の出現順序の情報は無視するため，FFデータの加工で示すような属性データの加工が成立する．潜在トピックあるいは共起語彙の出現確率は多項分布を用いて，式（1），（2）で表わされる．

$$P(z^{n,d}|\tilde{\theta}_d)=Multi_{k,1}(z^{n,d};\tilde{\theta}_d)$$
$$p\left(w_v^{n,d}\middle|z^{n,d},\phi_1,\dots,\phi_k\right) \quad (1)$$

$$=\prod_{k=1}^{k}\{Multi_{V,1}(w^{n,d};\phi_k)\}^{z_k^{n,d}} \quad (2)$$

多項分布のパラメータ$\tilde{\theta}_d$，ϕ_kの推定のため，その共役事前分布であるディリクレ分布を仮定する．これらは，式（3），（4）で表される．

$$p(\tilde{\theta}_d|\alpha)=Dir_K(\tilde{\theta}_d;\alpha) \quad (3)$$
$$p(\phi_k|\beta)=Dir_V(\phi_k;\beta) \quad (4)$$

各文書のトピック分布をD行K列の文書パラメータ$\Theta=(\tilde{\theta}_1,\dots,\tilde{\theta}_d)^t$，各トピックの語彙分布を$K$行$V$列のトピックパラメータ$\Phi=(\phi_1,\dots,\phi_k)^t$と定義する．なお右肩の添え字$t$は転置を表す．観測データ$W=[\{w^{n,d}\}_{n=1}^{N^d}]_{d=1}^{D}$と潜在変数$Z=[\{z^{n,d}\}_{n=1}^{N^d}]_{d=1}^{D}$の同時分布は，式（5）のように表される．

$$p(W,Z\,|\,\Theta.\Phi)$$
$$=\prod_{d=1}^{D}\sum_{n=1}^{N^d}p(w^{n,d}|z^{n,d},\phi_k)p\left(z^{n,d}\middle|\tilde{\theta}_m\right) \quad (5)$$
$$=\prod_{d=1}^{D}\prod_{n=1}^{N^d}\prod_{k=1}^{K}(\Theta_{d,k}\prod_{v=1}^{V}\phi_{k,v}^{w_v^{n,d}})^{z_k^{n,d}}$$

モデルの特徴を明らかにするため，潜在トピック$z^{n,d}$を消去して，W（N行V列）に関する周辺確率を求める．これは，式（6）で表される．さらに単語毎に*1-of-V*表現されたWを，式（7）によって定義される文書単位のBag-of-words（以下BOW）表現のデータMで書き改める．

$$p(W|\Theta,\Phi)$$
$$=\sum_{Z}p(W,Z|\Theta,\Phi)$$
$$=\prod_{d=1}^{D}\prod_{n=1}^{N^d}(\sum_{z^{n,d}\in\{e_k\}_{k=1}^{K}}\sum_{k=1}^{K}(\Theta_{d,k}\sum_{v=1}^{v}\Phi_{k,v}^{w_v^{n,d}})^{z_k^{n,d}} \quad (6)$$
$$=\prod_{d=1}^{D}\prod_{v=1}^{V}((\Theta\Phi)_{d,v})^{\sum_{n=1}^{N^d}w_v^{n,d}}$$

$$M=(m_1,\dots,m_D)^t,M_{d,v}=\sum_{n=1}^{N^d}w_v^{nd} \quad (7)$$

式（7）に示すように，この操作によって得られるMはD行V列となる．ここで，u_dは行列$U=(u_1,\dots,u_D)^t=\Theta\Phi$の第$d$行ベクトルである．すると式（7）は，データ$M$に関する確率分布として，式（8）に書き改められる．

$$p(M|\Theta,\Phi)=\prod_{d=1}^{D}N^d!\prod_{v=1}^{V}\frac{((\Theta\Phi)_{d,v})^{M_{d,v}}}{M_{d,v}!} \quad (8)$$

$$=\prod_{d=1}^{D}Mult\,i_{V,N^d}(m_d;u_d) \quad (9)$$

$$M\approx\Theta\Phi$$

式 (8) は，潜在パラメータ$\Theta\Phi$の積Uをハイパーパラメータとする，文書単位のBOWデータMの確率モデルである．この構造に文書別の単語数を乗じると，式 (9) が得られる．同式よりLDAは，トピック数Kをランクとする低ランク行列$\Theta\Phi$で，観測データMを近似する行列分解モデルとなっていることがわかる．繰り返しになるが，式 (3)，(4) よりD行K列のΘは文書別のトピック負荷を，K行V列のΦはトピック別の語彙負荷を，それぞれ表す行列である．トピックモデルのパラメータは，変分ベイズ法によって推定する．

トピックモデルの適合度は，perplexity (PPL) やトピックの解釈可能性に関するCohelenceなどの指標で評価される[24]．このうちPPLは尤度の関数であり，1単語当たりの期待予測語数を表す指標である．トピックモデルの対数尤度は，式 (8) の右辺に観測語を代入して得られる語彙観測確率の対数和であり，式 (10) より表される．

$$\begin{aligned} logL(M|\Theta,\Phi) &= \sum_d \sum_i log\ p(w_{di}|\Theta,\Phi) \\ &= \sum_d \sum_i \sum_k log\theta_{dk}\phi_{k,w_{di=v}} \end{aligned} \tag{10}$$

本研究では，PPLと同内容を表す対数尤度を用いる．トピックモデルのベースが多項分布であることに注意して，無情報 (初期) の選択確率を選択肢 (語彙) 数の逆数で与える．このとき，モデルの初期対数尤度は，

$$\begin{aligned} L(w_{di}|\Theta,\Phi_0) &= \sum_d \sum_i logp\,(w_{di}|\Theta,\Phi_0) \\ &= \sum_d \sum_i \sum_k log\frac{1}{K}\cdot\frac{1}{V} = -nlogV \end{aligned} \tag{11}$$

となり，トピック数Kとは無関係になる．本研究では，初期対数尤度と最終対数尤度の比で定義される尤度比を，Kの選択に用いる．

LDAの推定では，抽出トピック数Kを，モデル推定に先立って設定しなくてはならない．つまりモデル選択は，異なるKの設定下で尤度を参照しながら行うことになる．なお適切なトピック数Kの選択には尤度ばかりではなく，トピック解釈の容易性にも留意する必要がある．たとえば対数尤度が最大となるトピック数を選択しても，その中に類似度の高いトピック (の集合) が現れる場合は，トピック解釈に問題が生じる．そこでまず，トピックkとk'の類似度を定義しよう．

トピックkとk'が類似する場合は，トピックベクトル間の語彙の負荷パターンが類似する．トピックkとk'に対応する語彙負荷ベクトル (Φの第kおよびk'行を取り出した部分ベクトル) を，それぞれ$U_k, U_{k'}$とすると，それらのベクトル間のコサイン類似度$M_{kk'}$は，式 (12) で定義される．ここで式 (9) より行列Φの全要素は正のため，$M_{kk'}$の値は0〜1に限られる．つまり$M_{kk'}$の値が1に近い (ベクトルのなす角が小さい) ほど，トピックkとトピックk'の類似度は高く，0に近ければ全く類似しない．

$$M_{kk'} = \frac{U_k \cdot U_{k'}}{|U_k||U_{k'}|} \tag{12}$$

類似度$M_{kk'}$を全トピックペアについて算出して，類似／非類似の閾値を設定すれば，全トピックから類似トピック群を区別できる．なおLDAは固有値・固有ベクトルに基づく特徴ベクトルを抽出相互に直交せず，一部類似度の高いトピックが得られる場合がある．本研究では相互に異なる解釈ができるトピックのみが得られるように，類似と判定されたトピック群内のトピックの集約を行う．以下では最も簡単な集約の方法として，類似トピック群がえられたときは，単純にそれらのベクトル和から集約ベクトルを得る方法をとる．類似／非類似の閾値は，トピックベクトル間の成す角度が45° を基準とする．すなわち，類似とcos45° ≒0.7以上の類似度で，類似トピックと判定する．

以上の手順をまとめて示す．まず，1) 尤度比を用いて統計的な意味で適合するトピック数Kを算出する．次にトピックの解釈性を担保するため，2) 類似度$M_{kk'}$を全トピック間で算出し，3)$M_{kk'}$が閾値を超えるトピック群内のトピックは全てベクトル和によって集約する．なお上記の手順でトピックを集約した場合の集約トピックへの負荷は，トピック負荷行列Θについて，集約列の要素値の和を求めることによって算出できる．

4——トピックの抽出

4.1 FFデータの加工

本研究で用いるFFデータは，空海港から出国する外国人を対象としたサンプリング調査である．得られたサンプルを実際の訪日外国人流動量に拡大するため，国土交通省が，出入国管理統計[27]，国際航空旅客動態調査[28]，訪日外国人消費動向調査[29]で得た情報を基に拡大処理を行っている．各サンプルには，算出された拡大係数が，四半期，および年間のそれぞれについて，付されている．なお国内訪問地間の利用交通機関については，国際航空動態調査[28]で取得したOD別の交通機関分担率が，全データに適用されている．

FFデータの特徴は，入国海空港から国内訪問地，出国海空港までの一連のトリップチェーン情報が記録されていることである．データ内では，トリップごとに1行使用する形式で一連のトリップチェーンが記録されている．つまり1名の訪日外国人の周遊行動が3トリップから成る場合は，3

■表―1　本研究で用いた属性

属性	内容/カテゴリ
出国空海港	新千歳空港，旭川空港，函館空港，青森空港，仙台空港，秋田空港
	茨城空港，羽田空港，成田空港，新潟空港，富山空港，小松空港，静岡空港，中部空港
	関西空港，米子空港，岡山空港，広島空港，高松空港，松山空港，福岡空港
	福岡空港，佐賀空港，長崎空港，熊本空港，大分空港，宮崎空港，鹿児島空港
	那覇空港，石垣空港，北九州海港，博多海港，下関海港，厳原海港，その他空海港
国籍	韓国，台湾，香港，中国
	タイ，シンガポール，マレーシア，インドネシア，フィリピン，ベトナム
	インド，その他アジア
	イギリス，ドイツ，フランス，イタリア，スペイン，ロシア，その他ヨーロッパ
	アメリカ，カナダ，その他北アメリカ，南アメリカ
	オーストラリア，アフリカ，，その他オセアニア，無国籍
旅行目的	観光･レジャー，家族知人の訪問，業務，研修･学会等，留学，乗り継ぎ，その他，不明
旅行手配方法	団体旅行，個人旅行，不明
訪日回数	1回目，2回目，3回目，4回目，5回目，6～9回目，10～19回目，20回以上，不明
周遊出発地	47都道府県
	新千歳空港，旭川空港，函館空港，青森空港，仙台空港，秋田空港
	茨城空港，羽田空港，成田空港，新潟空港，富山空港，小松空港，静岡空港，中部空港
	関西空港，米子空港，岡山空港，広島空港，高松空港，松山空港，福岡空港
	福岡空港，佐賀空港，長崎空港，熊本空港，大分空港，宮崎空港，鹿児島空港
	那覇空港，石垣空港，北九州海港，博多海港，下関海港，厳原海港，その他空海港
周遊目的地	47都道府県
	新千歳空港，旭川空港，函館空港，青森空港，仙台空港，秋田空港
	茨城空港，羽田空港，成田空港，新潟空港，富山空港，小松空港，静岡空港，中部空港
	関西空港，米子空港，岡山空港，広島空港，高松空港，松山空港，福岡空港
	福岡空港，佐賀空港，長崎空港，熊本空港，大分空港，宮崎空港，鹿児島空港
	那覇空港，石垣空港，北九州海港，博多海港，下関海港，厳原海港，その他空海港
トリップ数	サンプルIDごとのトリップ数
季節	1～3月期，4～6月期，7～9月期，10～12月期
滞在日数	0日，1～2日，3日，4日，5日，6日，7日，8～10日
	11～14日，15～30日，31～90日，91～364日
利用交通機関	バス，鉄道，国内線飛行機，自動車
交通機関分担率	0～9％，10～19％，20～29％，30～39％，40～49％，50～59%
	60～69%，70～79%，80～89%，90～99%
拡大係数	サンプルIDごとの拡大係数

行にわたってデータが記録されている．トピックモデルの入力データは，文書Dごとに含まれる語彙Vの数をカウントした$D\times V$の行列データである．本研究では，文書＝旅行者，語彙＝旅行特性と考える．モデルへの入力データを得るには，オリジナルデータを行方向に訪日外国人数D，列方向に旅行特性Vをとった$D\times V$形式に加工する必要がある．このため，周遊目的地数に関する列を追加し，トリップチェーン内の訪問順に，訪問地1，訪問地2のようにコーディングした．この加工により，サンプル間の訪問地数が異なれば，記録される旅行特性数も異なるデータが得られる．なお訪問順序を考慮すると，考慮すべき属性の組み合わせが膨大になる．そこで本研究では簡単のため，訪問順序を考慮しないこととした．また利用交通機関分担率は，各トリップの拡大係数を合計して，トリップチェーン全体の値を算出した．

各個人のFFデータは，その旅行者の周遊全体で不変の個人属性と，周遊を構成するトリップごとに異なるトピック特性から成る．前者については，各変数のカテゴリカル値を表すダミー変数を作成して，それをBOWベクトルの要素とする．後者のうち，目的地別に得られる連続データである宿泊数は，まず階級値として離散化したうえで，目的地と統合して（たとえば，東京を目的地とするか否かのダミー変数と離散化した宿泊数を結合して），BOWベクトルを得る．FFデータの属性を表―1に示す．なお2014年のFFデータには，旅行手配方法，訪日経験回数の属性はない．以下の分析では，2014年から2016年の3年分のデータを用いる．

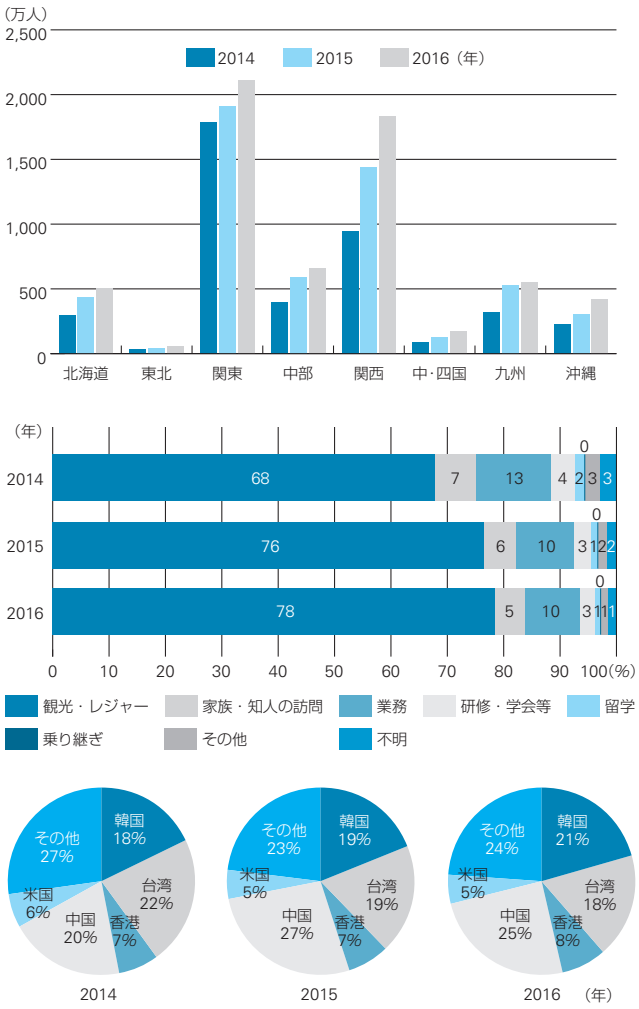
サンプル数の年次別セグメントの必要性を検討するため，まず全データをプールしてトピックモデルを適用した．しかし3年分のデータをプールすると，抽出トピック数は6に留まった．一方でFFデータを年次別に分割してそれぞれトピックモデルを適用したところ，各年次の抽出トピック数は6以上となったうえ，その中には複数地方を周遊先とするトピックが含まれていた．そこで以下では，3年次それぞれトピックモデルを適用した結果を報告する．また参考のため，図―1に周遊訪問地，訪日目的割合，国籍割合の単純集計の結果を示した．

4.2　トピック数の決定

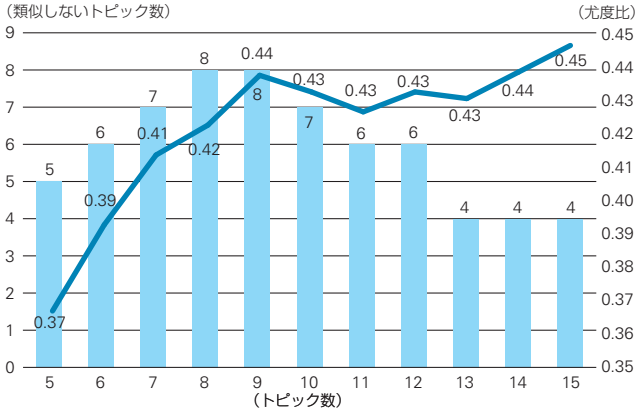
トピック数Kを設定してモデル推定を繰り返し，得られた尤度比を観察すると，いくつかのローカルピークを除いて，3年次ともK=60 を超えてもなお緩やかな増加傾向だった．そこで先述の通りコサイン類似度の閾値を0.7と設定して，トピック数Kを5から順次増やしながら非類似トピック数を記録した．その結果を図―1～3に示す，3年次とも，K=8～9あたりのローカルピークで非類似トピックが最大となった．一方K≧9~10では，尤度比はやや上昇する一方で，非類似トピック数は逆に減少した．そこで尤度比のローカルピーク以下のKで非類似トピック数が最大となるトピック数を採用したところ，それぞれ9（2014），8（2015），8（2016）となった．さらに，2014年はK=9の中に類似トピックが2つ含まれていたのでそれらを集約した．以上の結果，3年次とも8トピックを得た．

4.3　トピックの解釈・名付け

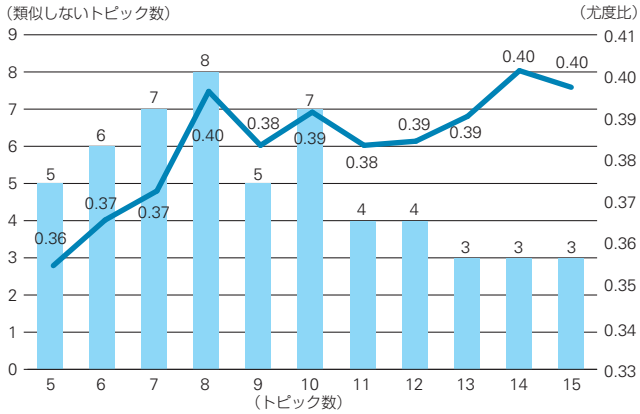
トピック―語彙行列Φの各行から得られるトピックは，



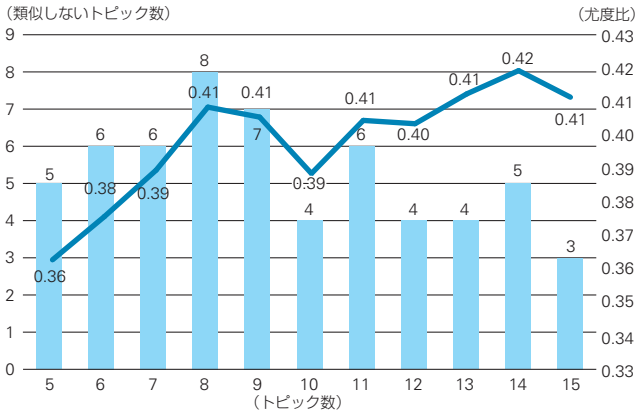
■図—1　単純集計結果



■図—2　尤度比と非類似トピック数 (2014年)



■図—3　尤度比と非類似トピック数 (2015年)



■図—4　尤度比と非類似トピック数 (2016年)

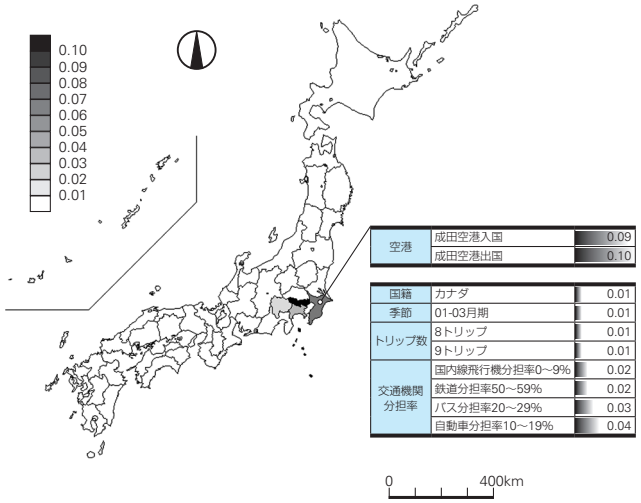
以下の手順で命名する. 2014年は14, 2015年は15, 2016年は16のように, トピック名の末尾にそれぞれ年号の識別記号を付けた. 次に, 観測された全てのBOW属性数 (トピックモデルでは総語数) に占める構成比率が大きいトピックから順に, 1, 2, 3の番号を付したそのうえで, 周遊地方が明確な場合にはその地方名, 他は都道府県名を付けた. ただし, 東京都, 京都府などを周遊するトピックはゴールデンルートと名付けた. 周遊地方や都道府県が不明のトピックは, 旅行目的を付した. たとえば, 2014年に構成比率が最も高いトピックは関東を周遊しており, 構成比率が2番目のトピックも関東を周遊していた. そこでそれぞれ, 14.1.関東1, 14.2関東2と名付けた.

5——訪日外国人旅行特性と経年変化

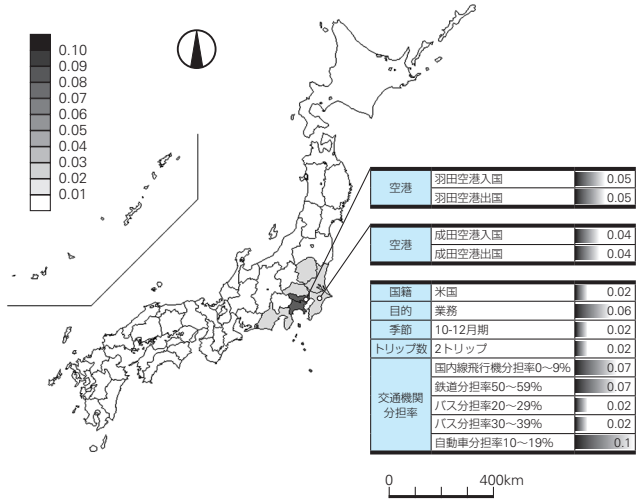
5.1 時点間で共通のトピック別旅行特性

命名したトピック名を図—5～28に示す. 各トピックの特徴を明らかにするために, 特に2014年について, トピック別に寄与度が上位の属性に着目する. なおトピックごとに帰属サンプル数が異なるため, トピック別の属性寄与度は, トピック全体で1になるように基準化して算出した.

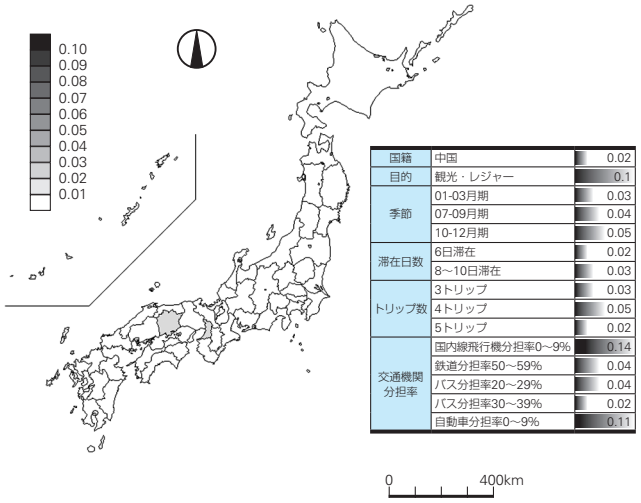
図—12, 20, 28より北海道トピックは, 主に北海道では過ごしやすい7−9月期に来訪する台湾や中国の団体旅行が多くを占めている. 5日滞在, 5トリップが上位である. 北海道以外にも北陸, 四国, 九州の属性も出現するが, それらの寄与は0.03未満にとどまり, 北海道の寄与 (0.2～0.5程度) が10倍以上卓越している. バス分担率60～69%, 70～79%が示すように, 移動手段は主にバスと考えられる. 出入国空港としてはともに新千歳空港の寄与が高いが, 入国のみ函館空港が現れている. すなわち新千歳ばかりでなく, 函館空港から入国し, 新千歳空港から出国するト



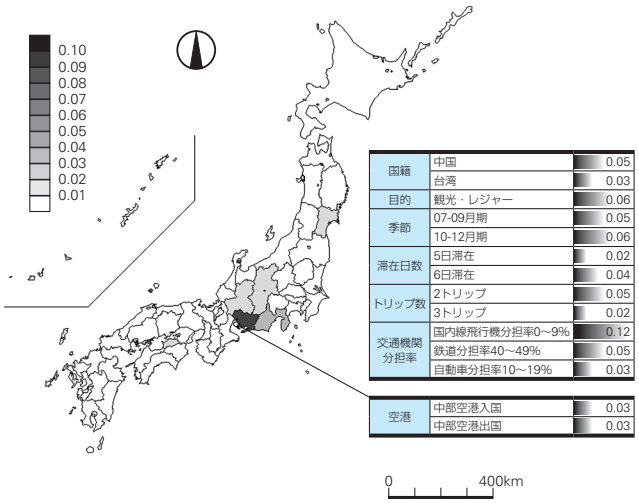
■図―5　14.1.関東1（年号. 構成比率. 周遊先）



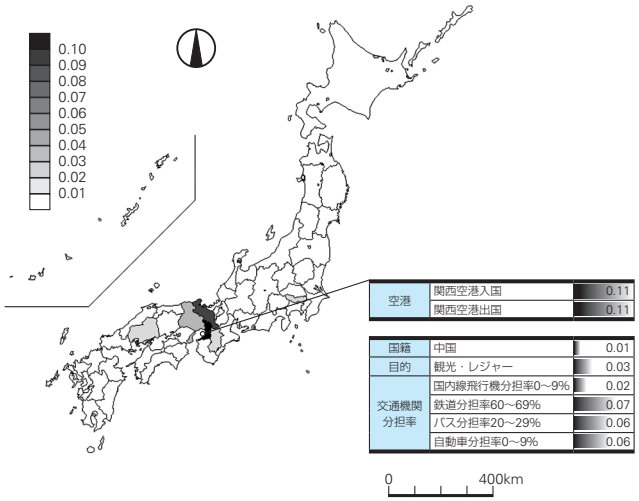
■図―6　14.2.関東2（年号. 構成比率. 周遊先）



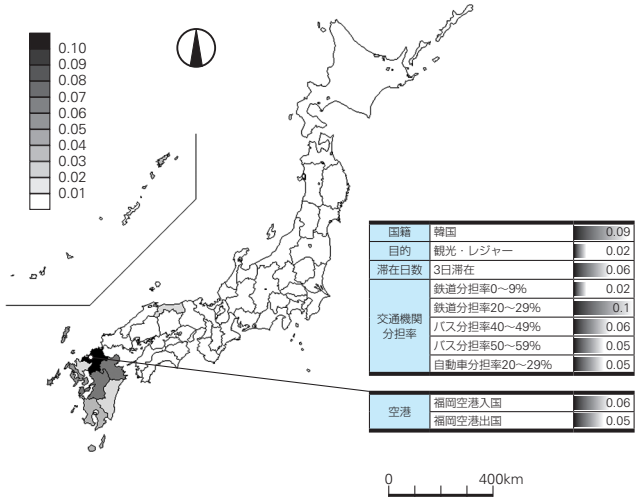
■図―7　14.3.観光（年号. 構成比率. 訪日目的）



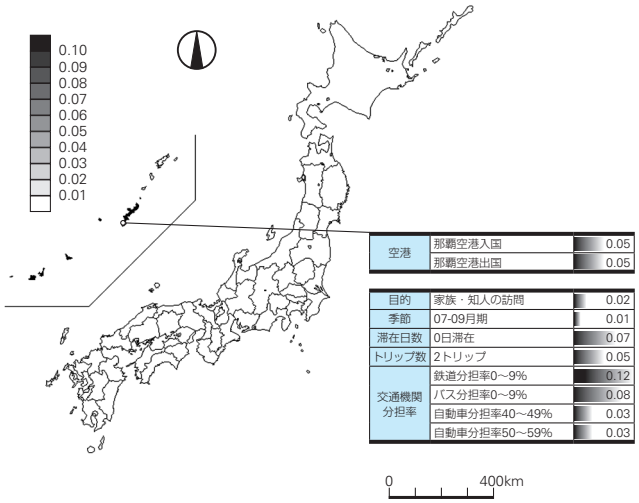
■図―8　14.4.中部（年号. 構成比率. 周遊先）



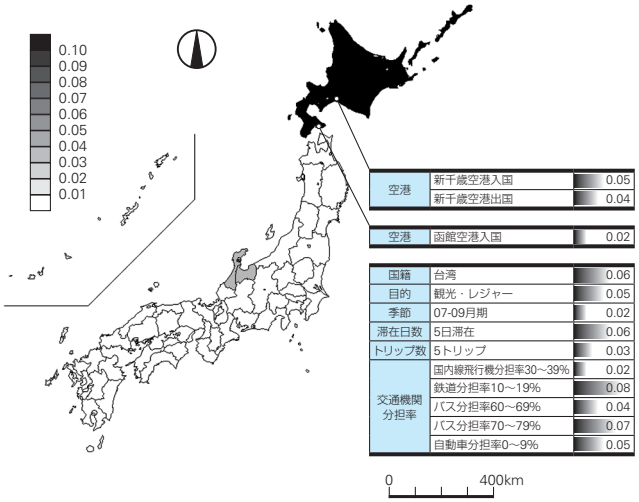
■図―9　14.5.関西（年号. 構成比率. 周遊先）



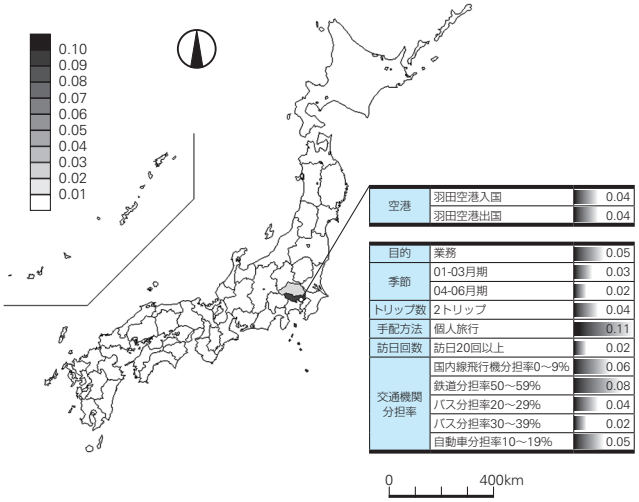
■図―10　14.6.九州（年号. 構成比率. 周遊先）



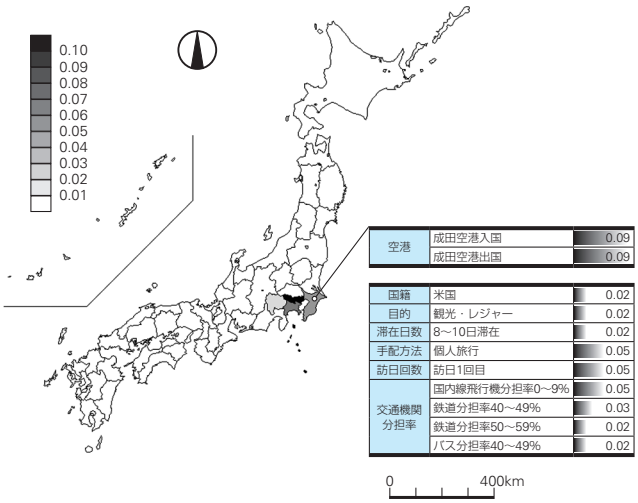
■図―11　14.7.沖縄（年号. 構成比率. 周遊先）



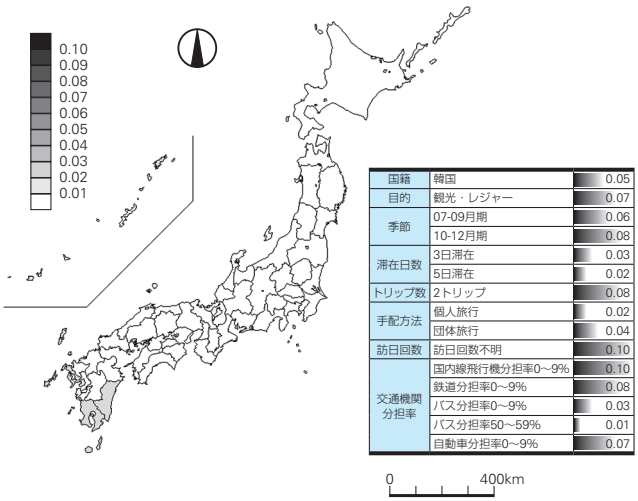
■図―12　14.8.北海道（年号. 構成比率. 周遊先）



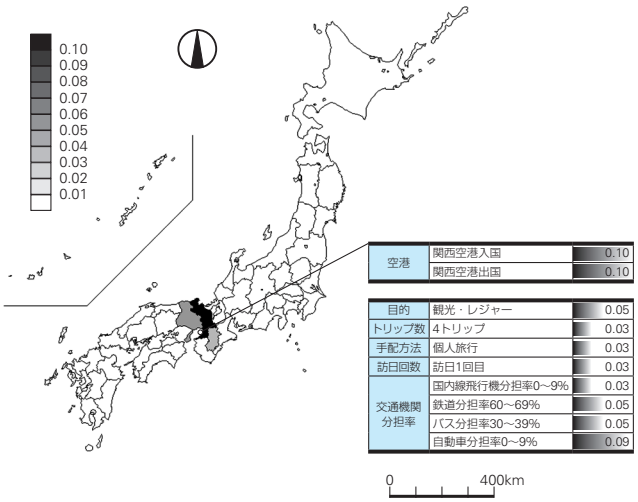
■図―13　15.1.関東1（年号. 構成比率. 周遊先）



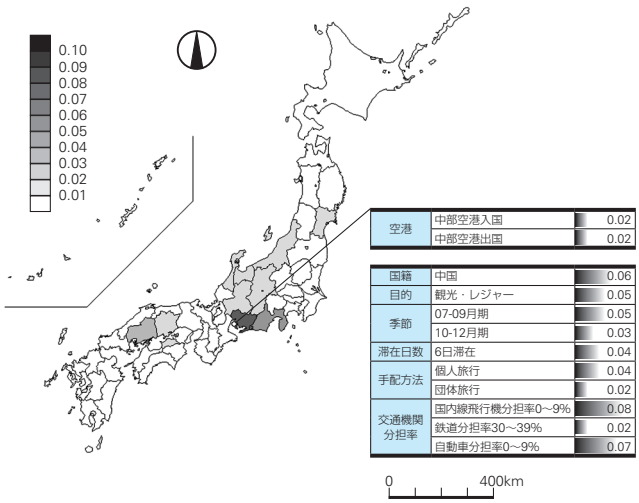
■図―14　15.2.関東2（年号. 構成比率. 周遊先）



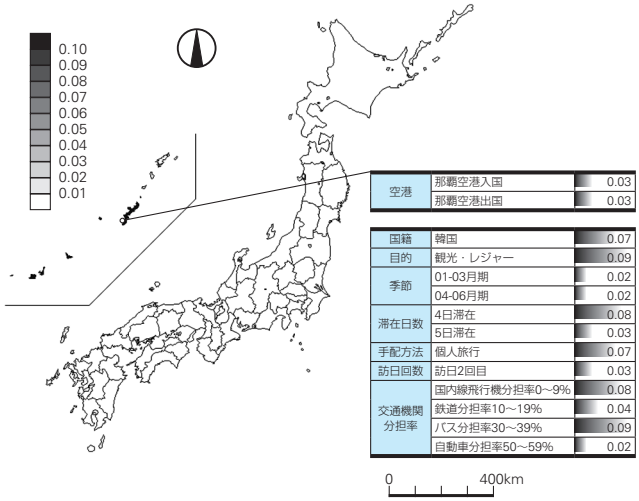
■図―15　15.3.観光（年号. 構成比率. 訪日目的）



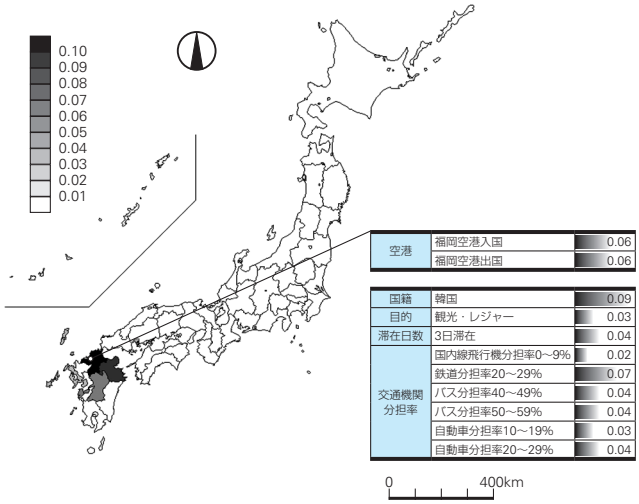
■図―16　15.4.関西（年号. 構成比率. 周遊先）



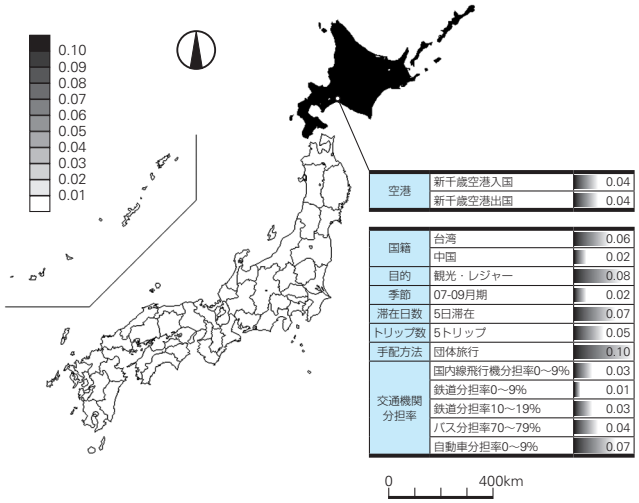
■図―17　15.5.中部（年号. 構成比率. 周遊先）



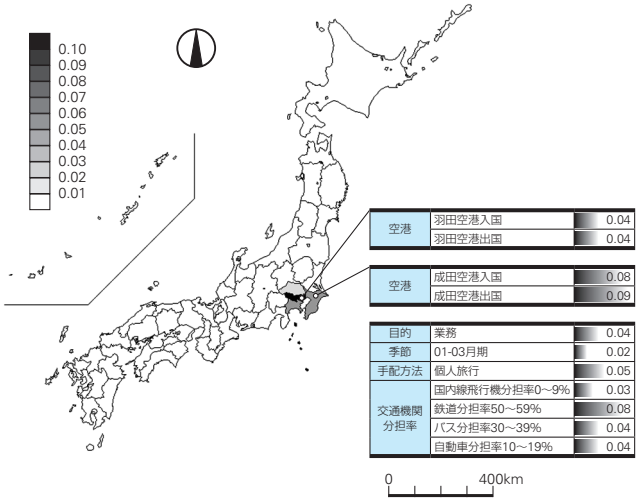
■図―18　15.6.沖縄（年号. 構成比率. 周遊先）



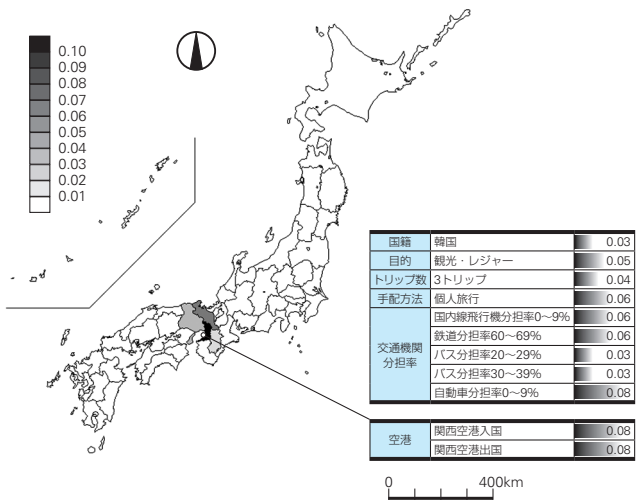
■図―19　15.7.九州（年号. 構成比率. 周遊先）



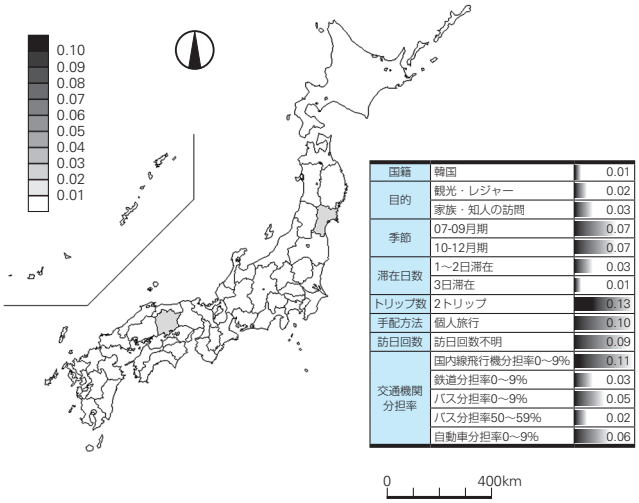
■図―20　15.8.北海道（年号. 構成比率. 周遊先）



■図―21　16.1.関東（年号. 構成比率. 周遊先）

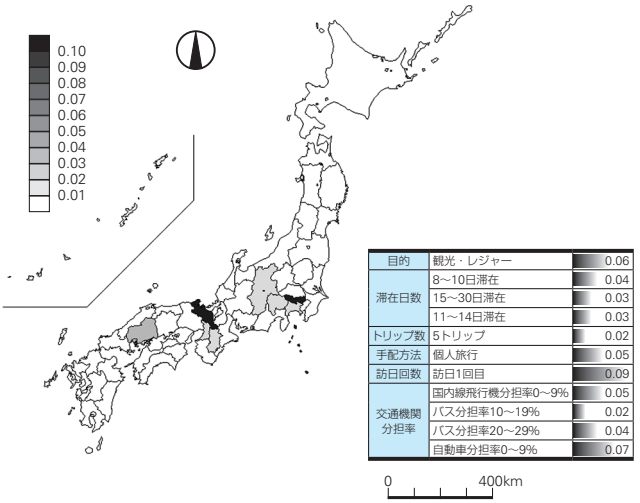


■図―22　16.2.関西（年号. 構成比率. 周遊先）



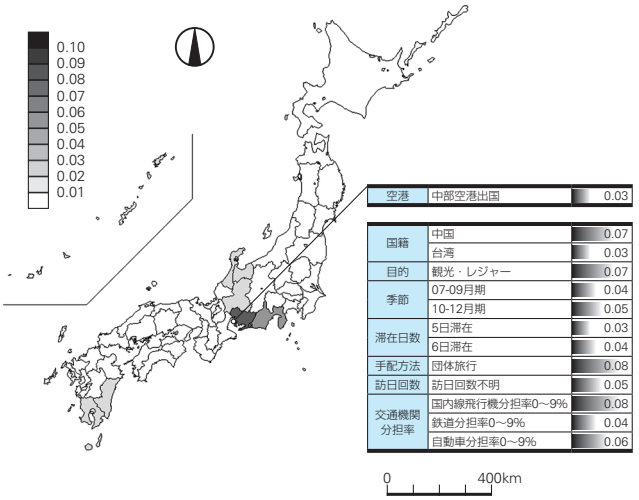
国籍	韓国	0.01
目的	観光・レジャー	0.02
	家族・知人の訪問	0.03
季節	07-09月期	0.07
	10-12月期	0.07
滞在日数	1〜2日滞在	0.03
	3日滞在	0.01
トリップ数	2トリップ	0.13
手配方法	個人旅行	0.10
訪日回数	訪日回数不明	0.09
交通機関分担率	国内線飛行機分担率0〜9%	0.11
	鉄道分担率0〜9%	0.03
	バス分担率0〜9%	0.05
	バス分担率50〜59%	0.02
	自動車分担率0〜9%	0.06

■図—23　16.3.観光（年号. 構成比率. 訪日目的）



目的	観光・レジャー	0.06
滞在日数	8〜10日滞在	0.04
	15〜30日滞在	0.03
	11〜14日滞在	0.03
トリップ数	5トリップ	0.02
手配方法	個人旅行	0.05
訪日回数	訪日1回目	0.09
交通機関分担率	国内線飛行機分担率0〜9%	0.05
	バス分担率10〜19%	0.02
	バス分担率20〜29%	0.04
	自動車分担率0〜9%	0.07

■図—24　16.4.ゴールデンルート（年号. 構成比率. 周遊先）



空港	中部空港出国	0.03

国籍	中国	0.07
	台湾	0.03
目的	観光・レジャー	0.07
季節	07-09月期	0.04
	10-12月期	0.05
滞在日数	5日滞在	0.03
	6日滞在	0.04
手配方法	団体旅行	0.08
訪日回数	訪日回数不明	0.05
交通機関分担率	国内線飛行機分担率0〜9%	0.08
	鉄道分担率0〜9%	0.04
	自動車分担率0〜9%	0.06

■図—25　16.5.中部（年号. 構成比率. 周遊先）

空港	那覇空港入国	0.03
	那覇空港出国	0.03

国籍	韓国	0.07
	台湾	0.02
目的	観光・レジャー	0.10
季節	01-03月期	0.02
	04-06月期	0.02
滞在日数	4日滞在	0.08
	5日滞在	0.03
手配方法	個人旅行	0.07
訪日回数	訪日2回目	0.03
交通機関分担率	国内線飛行機分担率0〜9%	0.08
	鉄道分担率0〜9%	0.04
	バス分担率20〜29%	0.06

■図—26　16.6.沖縄（年号. 構成比率. 周遊先）

国籍	韓国	0.09
目的	観光・レジャー	0.03
滞在日数	3日滞在	0.05
手配方法	個人旅行	0.02
交通機関分担率	自動車分担率10〜19%	0.03
	鉄道分担率20〜29%	0.06
	バス分担率40〜49%	0.04
	バス分担率50〜59%	0.04
	自動車分担率20〜29%	0.04

空港	福岡空港入国	0.06
	福岡空港出国	0.06

■図—27　16.7.九州（年号. 構成比率. 周遊先）

空港	新千歳空港入国	0.06
	新千歳空港出国	0.05
空港	函館空港出国	0.02

国籍	台湾	0.03
目的	観光・レジャー	0.04
滞在日数	5日滞在	0.03
トリップ数	5トリップ	0.03
手配方法	団体旅行	0.05
交通機関分担率	鉄道分担率10〜19%	0.06
	バス分担率60〜69%	0.05
	バス分担率70〜79%	0.02
	自動車分担率10〜19%	0.05

■図—28　16.8.北海道（年号. 構成比率. 周遊先）

リップやその逆のトリップが一部に現れる.

図―5, 6, 13, 14, 21, 24より関東を周遊するトピックには, 業務目的の旅行（図―6, 13, 21）と観光目的の旅行（図―5, 14, 24）がある. 前者は成田空港だけでなく, 羽田空港も利用している. 目的地は主に東京となっており, その他の目的地は現れない. また2トリップ, 訪日20回以上の属性が出現しており, 来日頻度が高い短日程の旅行である. 来訪時期は, 特定の季節によらず, 季節変動は少ない. また個人旅行が多く, 鉄道の分担率が高いことも特徴である. 一方で観光目的の旅行では, 成田空港のみが利用されており, 個人旅行で欧米からの旅行が多いという特徴がある. また目的地が広域に広がっており, 滞在日数が長く, トリップ数も多いという特徴もみられる.

図―8, 17, 25より中部トピックは, 中国, 台湾の団体旅行によって占められている. 主に中部空港を利用しているが, その他の空港の寄与も比較的大きいため, 直接中部地方から出入国しているわけではない. 滞在日数は5日滞在, 6日滞在が上位だが, トリップ数は2～3程度である. すなわち中部地方を中心に1か所に1～2日程度滞在する傾向がみられる. トピックが観測される時期は, 7-9月期, 10-12月期など, 1年の後半に多い.

図―9, 16, 22より関西トピックは, 韓国や中国の個人旅行が多くを占めている. 主に観光目的であり, 関西空港を利用している. 初訪日の旅行者が多いが, 滞在日数や季節の属性は出現していない. すなわち, 滞在日数は旅行者によってばらついており, 需要の季節変化は少ない特徴がみられる. 交通手段に関しては, 主に鉄道とバスを利用している.

図―10, 19, 27より九州トピックは, 韓国からの観光目的の旅行が多くを占めている. 主に福岡空港を利用しており, 福岡周辺の県を周遊する一方で, 福岡から離れた宮崎, 鹿児島県への周遊は少ない. 滞在日数が3日と短いことも, 周遊先が比較的狭い傾向と合致している. バス分担率, 自動車分担率が高い点も特徴である.

図―11, 18, 26沖縄トピックでは, 韓国や台湾からの個人旅行が多くを占めている. ただし沖縄県の寄与がほかの目的地よりも際立って高く（0.1~0.2）, また4, 5日滞在の寄与が高いことから, 主に沖縄県内に長期間滞在することがわかる. 那覇空港を出入国に利用し, 国内線飛行機分担率も低いことも主に沖縄県内を周遊する傾向を示している. 交通手段は, 主に自動車である.

図―24に示すトピックは, ゴールデンルートと名付けた. このトピックは, 関東だけでなく様々な地方を周遊している. 主に東京と京都が目的地として現れるほか, 海外にも知られた観光名所がある都道府県を周遊している. なお空港の属性が上位に現れなかったため, 出入国空港は広域に分散していると思われる. また滞在日数は8日以上30日までの属性が現れる一方で, トリップ数は5となっているため, それぞれの訪問地での宿泊数は1.6～6であり, 各目的地に比較的長く滞在する旅行が多いと思われる.

図―7, 15, 23より, 観光トピックでは訪問地属性は上位に現れない一方で, 非訪問地属性の寄与が大きい. つまりこのトピックは, 訪問地以外の代表的な特徴を示していると考えられる.

以上の結果が示すように, トピックモデルでは, 訪問地属性と訪日時期, 訪日目的, 訪日経験回数, 利用交通機関分担率, 宿泊数, 滞在日数, 訪日手配方法などの旅行者属性が柔軟に組み合わされて, 各年次の特徴を表すトピックが抽出できていると思われる.

5.2 訪日外国人旅行特性の変化

算出されたトピックの経年変化を捉えるために, 2014年と2015年, 2015年と2016年, 2014年と2016年のトピックベクトル間の類似度を算出した. その結果を表―2に示す. トピックの合成基準と同様に, 0.7を閾値として, その値を超えるトピックペアを年次間で同一のトピックとみなして, その経年変化を確認する. なお観光トピックには周遊先が含まれないため, 類似度が閾値を超えるペアがあっても, 考察対象からは除外した.

北海道トピックは, 3年次とも1トピックずつ推定されている. これらはいずれも類似度が0.9を超えており, 経年変化は小さいと考えられる. 周遊先と旅行者属性の組み合わせの変化も少ない. 北海道以外の周遊先も年により出現しているが, その寄与は北海道に比べて際立って低い. すなわちどの年も, 北海道内を周遊していると考えられる.

関東トピックは, 2014年と2015年で2トピック, 2016年では1トピック推定されている. 年次間でトピックが類似するグループは（14.1.関東2, 15.2.関東2, 16.1.関東）と,（14.2.関東1, 15.1.関東1）であった. 前述したように前者は業務目的の旅行である. 訪日目的, 旅行手配方法, 利用交通機関分担率などの変化は少ない. 年次間の主な違いは成田空港の利用の有無である. 2014年と2015年は羽田空港だけでなく, 成田空港を利用している. 成田空港を利用する旅行の周遊先は東京周辺にも及んでいた. 業務旅行の周遊先は, 短日程のため, 出入国海空港に左右されやすいと思われる. 後者では, 周遊先, 出入国空港, 利用交通機関分担率の変化は少ない. また, 国籍, トリップ数, 滞在日数などは上位に出現しないため, それぞれ多様なままと考えられる. 2016年のみ, 関東への観光目的の周遊トピックの代わりに, ゴールデンルートトピックが出現している. この結果は, 関東地方を目的地に含む観光トピックが, より広域を周遊するようになったことを示しており, FFデータの基礎的な特性報告とも合致している.

■表―2　トピックベクトル間コサイン類似度

	14.1. 関東１	14.2. 関東２	14.3. 観光	14.4. 中部	14.5. 関西	14.6. 九州	14.7. 沖縄	14.8. 北海道
14.1.関東１	1							
14.2.関東２	0.40	1						
14.3.観光	0.10	0.42	1					
14.4.中部	0.07	0.46	0.69	1				
14.5.関西	0.04	0.06	0.26	0.09	1			
14.6.九州	0.01	0.07	0.11	0.17	0.01	1		
14.7.沖縄	0.03	0.23	0.43	0.52	0.05	0.17	1	
14.8.北海道	0.01	0.03	0.19	0.14	0.05	0.02	0.18	1

	14.1. 関東１	14.2. 関東２	14.3. 観光	14.4. 中部	14.5. 関西	14.6. 九州	14.7. 沖縄	14.8. 北海道		15.1. 関東1	15.2. 関東2	15.3. 観光	15.4. 関西	15.5. 中部	15.6. 沖縄	15.7. 九州	15.8. 北海道
15.1.関東1	0.26	0.70	0.39	0.36	0.08	0.06	0.21	0.01	15.1.関東1	1							
15.2.関東2	0.92	0.44	0.18	0.14	0.04	0.02	0.08	0.02	15.2.関東2	0.30	1						
15.3.観光	0.03	0.25	0.67	0.61	0.13	0.24	0.47	0.16	15.3.観光	0.10	0.29	1					
15.4.関西	0.02	0.09	0.40	0.17	0.91	0.02	0.17	0.09	15.4.関西	0.10	0.15	0.23	1				
15.5.中部	0.06	0.22	0.64	0.72	0.14	0.10	0.27	0.17	15.5.中部	0.11	0.18	0.53	0.23	1			
15.6.沖縄	0.04	0.23	0.52	0.48	0.07	0.22	0.80	0.15	15.6.沖縄	0.15	0.40	0.42	0.24	0.30	1		
15.7.九州	0.02	0.12	0.15	0.17	0.02	0.92	0.13	0.03	15.7.九州	0.04	0.10	0.16	0.05	0.12	0.23	1	
15.8.北海道	0.02	0.05	0.30	0.20	0.07	0.03	0.18	0.93	15.8.北海道	0.04	0.04	0.26	0.13	0.31	0.18	0.05	1

	14.1. 関東１	14.2. 関東２	14.3. 観光	14.4. 中部	14.5. 関西	14.6. 九州	14.7. 沖縄	14.8. 北海道		15.1. 関東1	15.2. 関東2	15.3. 観光	15.4. 関西	15.5. 中部	15.6. 沖縄	15.7. 九州	15.8. 北海道		16.1. 関東	16.2. 関西	16.3. 観光	16.4. ゴールデンルート	16.5. 中部	16.6. 沖縄	16.7. 九州	16.8. 北海道
16.1.関東	0.84	0.70	0.19	0.15	0.02	0.02	0.11	0.01	16.1.関東	0.59	0.88	0.08	0.09	0.07	0.18	0.05	0.02	16.1.関東	1							
16.2.関西	0.04	0.16	0.54	0.31	0.84	0.08	0.25	0.11	16.2.関西	0.30	0.13	0.36	0.93	0.31	0.42	0.11	0.16	16.2.関西	0.12	1						
16.3.観光	0.03	0.30	0.53	0.58	0.09	0.15	0.41	0.09	16.3.観光	0.52	0.12	0.85	0.20	0.44	0.38	0.09	0.15	16.3.観光	0.15	0.32	1					
16.4. ゴールデンルート	0.38	0.29	0.61	0.34	0.39	0.04	0.22	0.13	16.4. ゴールデンルート	0.42	0.52	0.38	0.49	0.43	0.40	0.10	0.20	16.4. ゴールデンルート	0.39	0.51	0.37	1				
16.5.中部	0.05	0.22	0.65	0.71	0.12	0.10	0.35	0.19	16.5.中部	0.17	0.10	0.72	0.22	0.90	0.35	0.11	0.36	16.5.中部	0.06	0.31	0.52	0.37	1			
16.6.沖縄	0.06	0.20	0.52	0.43	0.12	0.22	0.70	0.13	16.6.沖縄	0.37	0.11	0.46	0.17	0.30	0.86	0.22	0.17	16.6.沖縄	0.11	0.37	0.36	0.42	0.38	1		
16.7.九州	0.02	0.14	0.17	0.19	0.02	0.92	0.17	0.04	16.7.九州	0.14	0.04	0.20	0.05	0.12	0.31	0.97	0.05	16.7.九州	0.06	0.14	0.13	0.11	0.10	0.28	1	
16.8.北海道	0.02	0.09	0.09	0.11	0.01	0.02	0.11	0.93	16.8.北海道	0.05	0.01	0.08	0.03	0.12	0.10	0.05	0.92	16.8.北海道	0.03	0.04	0.04	0.06	0.14	0.08	0.05	1

中部トピックは，3年次ともそれぞれ1トピックが算出された．それぞれの周遊先に着目すると，いずれも愛知県，静岡県，岐阜県が含まれている．国籍，季節，訪日目的，滞在日数，旅行手配方法などの旅行者属性の変化は小さい一方で，周遊先は年次間でばらつきが大きい．中部トピックは団体旅行の寄与が高いため旅行者属性の変化は少ないが，団体旅行の企画内容によって周遊先が異なるためと考えられる．

関西トピックは，3年とも類似度が高い．詳細に比較すると，訪日目的，出入国海空港，旅行手配方法，利用交通機関分担率，訪日回数などの経年変化は小さい．また周遊先に大阪，京都，兵庫が含まれる点には変化がみられない．2014年は東京や広島が含まれていたほか，2015年は4トリップであった旅行回数が，2016年では3トリップとなった．これらの傾向から関西トピックは，他の地方を周遊する旅行から，関西のみを周遊する旅行へと変化したと考えられる．

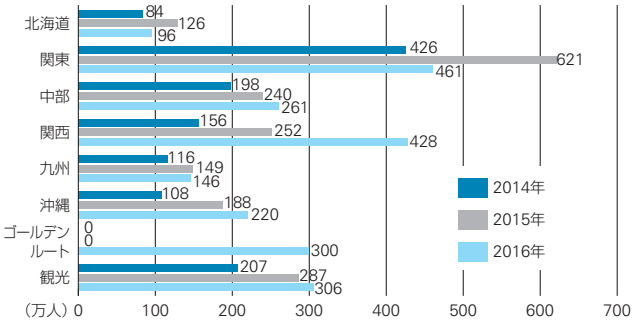
九州トピックは，北海道トピックと同様に，どの年次間も類似度が0.9を超えており，経年変化が少ない．各年次を構成する属性を詳細に比較しても，国籍，滞在日数，訪日目的，出入国海空港は変化していない．なお周遊先は2014年は九州全体だったが，年々福岡以外の寄与が減少している．すなわち，九州トピックは九州広域の周遊から福岡県にのみの周遊に変化している．

沖縄トピックは，出入国海空港，利用交通機関分担率などの変化が小さい．しかし2014年から2015年にかけて訪問目的や滞在日数などが変化した．2014年では，旅行目的は家族知人の訪問だったが，2015年，2016年では観光・レジャーが多くなっている．また2014年と比較すると，滞在日数が増加している．

5.3　トピック構成比率と人数の経年変化

前節までに算出したトピックは，観測した旅行者数の比率に基づいて抽出している．一方でFFデータには，出国空海港の訪日外国人国籍比率に合わせて算出した旅行者レコード別の拡大係数が付されている．そこで本節では，旅行者―トピック行列Θから得られる旅行者別のトピック構成比率に拡大係数を乗じたうえで，トピック別に人数和を求める手順で，トピック別の旅行者数を算出した．図―29にその結果を示す．

北海道と九州，および関東は，2015年に大きく観光客数が伸びている．一方，中部，関西，沖縄と，特定の地域が上位の寄与リストに現れない観光は，経年的に総旅行者数が増加している．これらの中では，特に関西の伸びが著しい結果となった．2016年には複数地域を周遊するゴールデンルートトピックが現れており，その人数は，関東，関西，観光に次いで4番目に多く，北海道，中部，九州，沖縄を上回っている．



■図―29　旅行トピック別の旅行者数

関東では2016年に訪日外国人数が減少しているが，これは複数地方を周遊するゴールデンルートトピックが出現したためと考えられる．ゴールデンルートでは東京，京都の目的地寄与が高い．他方で，同期間中に関東を旅行する訪日外国人数は減っていないことを考え合わせると，関東のみを周遊する訪日外国人数は減少したが，他地域を含む広域周遊先に関東が含まれるように変化したと考えられる．

6――おわりに

本研究では，FFデータを用いて訪日外国人の周遊に関する特性を経年分析した．分析手法として，元来は文章解析に用いられるトピックモデルを応用した．トピックモデルを適用するため，FFデータに含まれる連続変数の離散化処理を行うことによって，BOW表現を得る手法を提案した．トピックモデルを用いた既往研究で課題としてあげられていたトピック数の決定手順や，トピックの解釈手順については，以下の工夫を行うことにした．すなわち，尤度比とトピック間の類似度を指標としてトピック数の候補を選定し，各トピック数から得られる分析結果に基づいて，その解釈性に配慮したトピック数を採用した．その結果，膨大な訪日外国人の訪問パターンや訪問時期，利用交通機関等の組み合わせから，特徴的なパターンを抽出することができた．抽出されたパターンは，結果に対する先験知識を用いることなく得られたものである．つまり，トピックモデルを適用することによって，膨大なサンプル数からなる多属性データの特徴を，簡便に抽出できることが示された．

抽出した旅行トピックの特性や旅行者数に基づいて，訪日外国人の周遊特性と，その経年変化の特徴が明らかとなった．北海道トピックを構成する旅行特性の経年変化は小さかった．関東トピックでは，関東を業務目的の旅行は出入国海空港の違いにより周遊先が異なっていた．また，観光目的で関東の都県を周遊する旅行は，関東のみならず様々な地方を周遊するようにトピック特性が変化してい

る．中部トピックは旅行者属性の変化は少ないが，中部地方内の目的地が変化している．関西，九州トピックでは，旅行者属性は変化していないが，周遊先は九州全域から福岡周辺に縮小している．沖縄トピックでは，周遊先は変化していないが，旅行目的が知人の訪問から観光へと変化するとともに，滞在日数が増加している．

以上の結果に基づいて，インバウンド政策への示唆について，簡単にまとめておく．本研究で得られた訪日外国人の周遊特性は，基本的に国土交通省の報告[30]で言及された内容と合致していた．北海道や沖縄などは，それぞれ地域内のみを周遊する観光客が一定数存在している．域内の観光地を発掘してそれらの地点をスムーズに周遊できる観光ルートの整備の一方で，同一地点に長期滞在できる地域観光素材の開発を並行して行う必要がある．関東トピックの経年変化は，周遊先の広域化を示している．これらの旅行者の多くが個人旅行のリピーターと考えられるため，彼らのニーズに応えられる都市間交通網の整備が必要である．他方で，関西や九州，特に福岡は，特定地域の特定時期に観光客が集中するオーバーツーリズムが懸念される．これらの地域の観光地としての魅力を損なわないような観光期のシフトと，周遊先の分散が必要と考えられる．

今後の研究課題を以下にまとめる．本研究では，2014年，2015年，2016年ごとにデータを分割してトピックモデルを適用した．ただし，データをあらかじめ季節，出国海空港別にセグメントすれば，異なる結果が得られると思われる．本研究のような経年比較ばかりでなく，異なるデータセグメントから得られる結果を比較することによって，統計的にロバストな結果が得られる可能性がある．また，類似判定の基準として用いたコサイン類似度の閾値の設定についても検討を行う必要がある．

FFデータは，都道府県単位での周遊のみ記録されているが，旅行商品や観光地域づくり戦略の策定に対しては，空間的な集計単位が粗すぎる．今後は，周遊先の空間的な集計単位を細かくするための補助データの活用方法などを検討する必要がある．

なおトピックモデルには，あらかじめ取得できる外部情報を半教師データとして用いるAuthor topic modelなどの応用モデルも存在する．さらに近年研究の蓄積が著しい，深層学習や強化学習などを組み合わせた特徴抽出手法の適用可能性についても，検討する必要がある．

参考文献

1) 日本政府観光局（JNTO）[2003]，“ビジット・ジャパン事業について”，https://www.jnto.go.jp/jpn/projects/promotion/vj/index.html，2019/6/13.
2) 観光庁 [2011]，“訪日旅行促進事業について”，http://www.mlit.go.jp/kankocho/shisaku/kokusai/vjc.html，2019/6/13.
3) 国土交通省 [2014]，“FF-Data（訪日外国人流動データ）”，http://www.mlit.go.jp/sogoseisaku/soukou/sogoseisaku_soukou_fr_000022.html，2019/6/13.
4) 田中賢二 [2007]，“外国人観光客の訪日促進に関する研究－国際観光の現状の分析と安定的な旅行者の獲得を中心として－”，「運輸政策研究」，Vol.10，No.1，pp.11～21.
5) 日比野直彦・早川信二・森地茂・金兌奎 [2009]，“観光地の特性と入込客数の時系列変化に関する基礎的研究”，「運輸政策研究」，Vol.11，No.4，pp.30～36.
6) 室谷正裕 [1998]，“観光地の魅力度評価－魅力ある国内観光地の整備に向けて－”，「運輸政策研究」，Vol.1，No.1，pp.14～24.
7) 早川信二・奥山忠裕・室井寿明・Michelle P. Pernia・毛塚宏・藤崎耕一 [2010]，“訪日外客の公共交通に対する選好の定量分析－成田空港アンケート調査によるWTP推計とコンジョイント分析－”，「運輸政策研究」，Vol.13，No.3，pp.4～14.
8) 栗原剛・岡本直久 [2010]，“インバウンド需要に影響を与える政策および外的要因の考察”，「土木計画学研究・論文集」，Vol.27，No.1，pp.147～155.
9) 古屋秀樹 [2013]，“国外旅行者数を用いたアジア諸国の相対的魅力度推定－目的地選択率による逆解析手法の適用－”，「運輸政策研究」，Vol.15，No.4，pp.41～49.
10) 岡本直久・栗原剛 [2007]，“アジア諸国における将来の国際旅行に関する考察”，「運輸政策研究」，Vol.10，No.3，pp.2～10.
11) 櫻井賢一郎・細江宣裕 [2005]，“北海道観光振興計画はアド・バルーンか？”，「運輸政策研究」，Vol.8，No.1，pp.2～10.
12) 松井裕樹・日比野直彦・森地茂・家田仁 [2016]，“訪日外国人旅行者の個人行動データを用いた訪問地および観光行動に着目した観光行動分析”，「土木学会論文集D3」，Vol.72，No.5，pp.533～546.
13) 香川喬之・桑野将司・福山敬・谷本圭志・川村尚生・菅原一孔 [2016]，“バス経路探索履歴データを用いた移動希望特性の分析”，「交通工学論文集」，Vol.2，No.2，pp.115～124.
14) 菱田のぞみ・日比野直彦・森地茂 [2012]，“訪問地選択の多様性に着目した訪日中国人旅行者の居住地別観光行動の時系列分析”，「土木学会論文集D3」，Vol.68，No.5，pp.667～677.
15) 石井健一郎・上田修功・前田英作・村瀬洋 [1998]，『わかりやすいパターン認識』，オーム社.
16) 古屋秀樹・劉瑜娟 [2016]，“潜在クラス分析を用いた訪日外国人旅行者の訪問パターンの分析”，「土木学会論文集D3」，Vol.72，No.5，pp.571～583.
17) 塚井誠人・椎野創介 [2016]，“討議録に対するトピックモデルの適用”，「土木学会論文集D3」，Vol.72，No.5，pp.341～352.
18) 塚井誠人・塚野裕太 [2018]，“トピックモデルによる詳細地理情報分析”，「土木学会論文集D3」，Vol.74，No.2，pp.111～124.
19) 川野倫輝・佐藤嘉洋・円山琢也 [2018]，“トピックモデルと離散連続モデルを用いた自由記述の量的分析”，「土木学会論文集D3」，Vol.74，No.5，pp.277～284.
20) 古屋秀樹・岡本直久・野津直樹 [2018]，“GPSログモデルを用いた訪日外国人旅行者の訪問パターンの分析手法の開発”，「運輸政策研究」，Vol.20，pp.20～29.
21) Rosen-Zit，M. Griffiths，T. Steyvers，M. and Smyth P.［2004］，“*The author-topic model for authors and documents*”，UAI，pp.487-494.
22) Blei，D. M. and Lafferty，J. D.［2007］，“A correlated topic model of science”，*The Annals of Applied Statistics*，Vol. 1，No. 1，pp17~35.
23) 岩田具治 [2015]，『トピックモデル』，講談社
24) 佐藤一誠・奥村学 [2015]，『トピックモデルによる統計的潜在意味解析』，コロナ社
25) 中島伸一 [2016]，『変分ベイズ学習』，講談社
26) 坪井祐太・海野裕也・鈴木潤 [2017]，深層学習による自然言語処理』，講談社
27) 法務省 [2006]，“出入国管理統計”，http://www.moj.go.jp/housei/toukei/housei05_00016.html，2019/6/13.
28) 国土交通省 [2005]，“航空局実施の統計調査”，http://www.mlit.go.jp/koku/koku_tk6_000001.htm，2019/6/13.
29) 観光庁 [2010]，“訪日外国人消費動向調”，http://www.mlit.go.jp/kankocho/siryou/toukei/syouhityousa.html，2019/6/13.
30) 国土交通省 [2019]，“FF-Data（訪日外国人流動データ）の概要と利用例”，http://www.mlit.go.jp/common/001290684.pdf，2020/1/6.

（原稿受付2019年9月19日，受理2020年3月17日）

A Tour of Analysis of Foreign Visitors to Japan by Applying Topic Model

By Yoshihiro TATSUMI and Makoto TSUKAI

Inbound policy making requires to identify the tour trip characteristics. The exogenous data aggregation would be, however, difficult because of too many possibilities in the combination of tour, or trip attributes. In order to find some significant and representative tour or trip patterns in foreign visitors to Japan, an analytical tool to overcome the conventional problems should be developed. This study applies a topic model, which can efficiently analyze the latent patterns in the Bag-of-words data set, to the data in Flow of Foreigners visiting to Japan, and clarifies the change of tour trip patterns.

Key Words : ***machine learning, big data, longitudinal change.***