博　士　論　文

中国大陸観光客のレビューに基づくテキストマイニング分析を用いた
インバウンド観光発展のための傾向予測研究
A Trend Prediction Study by Text Mining Analysis Based on Reviews by the Mainland Chinese Tourists to Promote Inbound Tourism

2020 年 3 月

張　凱楽

目次

図目次

表目次

第 1 章　序論

1.1　問題意識

観光業は世界経済のバロメーターと言われ、経済の動向を予測する際に重要な指標であるとともに、世界の GDP においても重要な役割を果たしている。2018 年には、世界の観光総収入(国内観光収入と国際観光収入を含む)は 5.34 兆米ドルと世界全体の GDP の 6.1%に相当し、国際観光収入だけでも 1.59 兆米ドルである。また、経済の発展以外にも国際観光は、各国の人々どうしの文化交流を促進する重要な役割も担っている。[1][2]

観光業はサービス業のひとつとして、文化、交通、宿泊、政策、さらに IT 産業を統合した総合サービス業であり、観光業における様々なサービスの目的には観光客の満足度を向上させることが含まれる。そのため、ユーザーエクスペリエンス[1](user experience)は重要な要素と見なされている。

また、観光業を取り巻く大きな環境のひとつとして国レベルの観光政策があるが、観光事業者と自治体の観光行政部門は、国の観光政策については直接関与することができないため、各国政府の観光政策の転換や世界経済の景気変動に伴う地元観光業の変化や課題に直面しても、受動的に対応することしかできない。しかし、観光客のユーザーエクスペリエンスにおける満足度については、能動的にそれを向上させることを通じて地元観光業の発展を促進することが可能であることから、観光事業者と自治体の観光行政部門にとって直接的かつ実行可能な手段であると考えられる。

現在、観光業の発展において、中国大陸からのアウトバウンド観光の急速な発展に伴い、オンライン旅行代理店[2]（Online Travel Agency、以下「OTA」と略す）産業の業界内の台頭と、「OTA」企業が提供する観光レビューのプラットフォームの発展の 2 つが重要な役割

---

[1] ユーザーエクスペリエンス（UX）とは，「製品やサービスを利用・消費した時に得られる体験の総体」のことで，本来は IT 分野のヒューマンインタフェース研究から見出された概念である。本研究は，この UX を人とのリアルなインタフェースが不可欠な観光に当てはめることで，驚きや感動、上質なサービスを受けている事に対する充足感や事前期待を与える方法論の展開を目指している。

[2] オンライン旅行代理店とは、インターネット上だけで取引を行う旅行会社のこと。英語は Online Travel Agent であり、「OTA」と略す。店舗で営業を行っている旅行会社のオンライン販売は OTA とは呼ばない。国内外の宿泊や航空券などの手配旅行、宿泊と航空をセットにしたパッケージ、宿泊仲介、旅行保険などを取り扱うことが多い。24 時間いつでも商品・サービスを閲覧・検索でき、店舗へ出向く必要のない利便性が消費者の支持を得ている。

を担っている。

加えて、中国大陸観光客[3]を対象とした観光事業において、効率的・効果的な経営資源の配分は多くの観光事業者にとって経営的にも重要な関心事項となっている。例えば、中国大陸観光客向けに大きな市場潜在力を持つ観光体験（宿泊施設内の飲食体験など）や季節の特徴を生かした観光商品・サービスについて、より多くの企画や宣伝をすること、また、中国大陸観光客にとって潜在的なニーズが高い観光の時期や、繁忙期の現地サポートスタッフの増員などによる観光サービス品質の向上を限られた経営資源の中でどのように効率的・効果的に行えるか、ということである。

以上から、「OTA」Web サイトに投稿される観光レビューの分析を行い、中国大陸観光客の嗜好や傾向を把握・予測し、それを観光業務の現場に活用することは、今後の観光産業の発展を促進する大きな潜在力（特にユーザーの満足度を高めること）に大きく貢献できると考える。

[3] 中国大陸観光客とは、中国大陸からの観光客を意味する。中国大陸とは、香港、澳門、台湾を除く中国を指す。

1.2　研究背景

1.2.1　観光業と経済全般

前節 1.1 で述べたように、観光業が世界経済のバロメーターと呼ばれるのは、観光業の成長は世界経済の景気状況の影響を大きく受けるためである（図 1-1）。2008 年の金融危機以降、世界経済は停滞するか、わずかに減少を示した（WB：世界銀行と IMF：国際通貨基金の値は集計手法によりやや異なる）が、世界の観光業の成長率はマイナス 8.3%と大幅に下がった。また、2010~2011 年に世界経済が回復すると、世界の観光業の成長率はすぐに 17.8%増えた。図 1-1 では両者の大まかな傾向は一致しているものの、変動幅の差は大きく、世界経済の変化に対する観光業収入の感応度が高いことが分かる。[3][4][5]

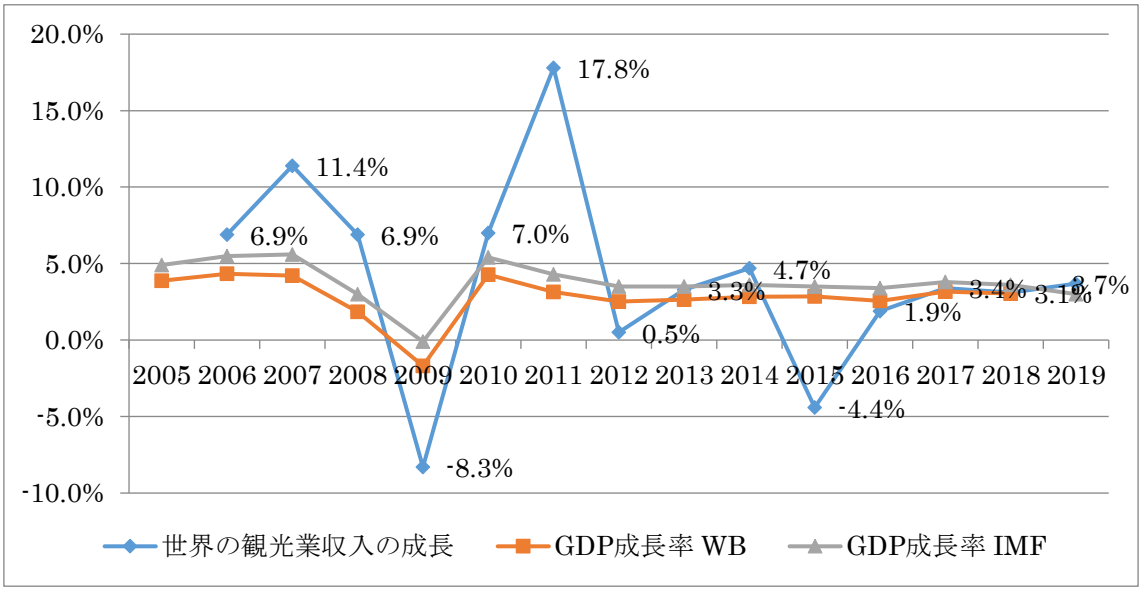


図 1-1　世界の観光業収入の成長と GDP 成長率 [2][3][4]

また、観光業収入は世界経済の動きに完全にリンクされているとも言えない。それは、まず観光業は確かに世界経済の影響を受けているものの、経済成長率のマイナスのインパクトは徐々に蓄積されるものであり、景気が少し上向きになると一気に解放されるため、必ずしも世界経済の動きとは正確には一致していない。次に、観光業自身が提供する商品・サービスや施設の品質も業界全体の収入に大きな影響を与えるからである。また、ある地域や国の経済発展とその地域や国の観光業の発展とは直接的な関連性が低い場合もある。その典

型的な例が日本である。図 1-2 に示した通り、2013~2018 年における日本の GDP 成長率は年平均わずか 1%となっているが、訪日外国人観光客の消費額は年平均 25%以上の伸び率を維持している。したがって、経済全体の大きな傾向に注目すると同時に、観光業が提供する製品やサービスの品質を向上させ、観光客の満足度を高めることはインバウンド観光産業にとって、より重要であると考えられる。

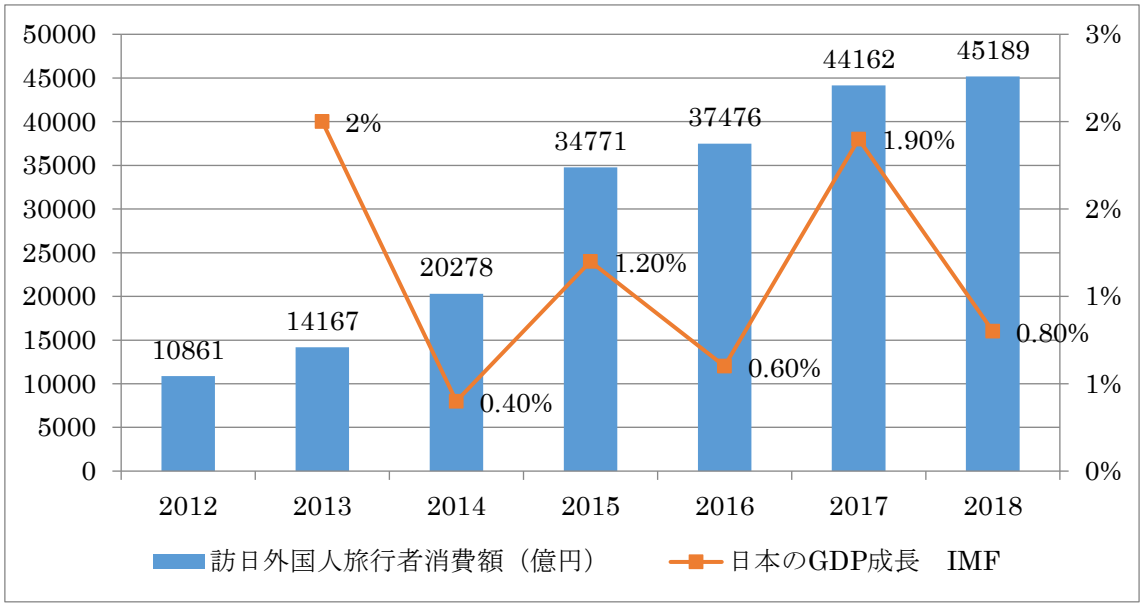


図 1-2　訪日外国人旅行者消費額の成長と日本の GDP 成長率 [5][6]

1.2.2　中国大陸からのアウトバウンド観光の発展と現状

中国大陸からのアウトバウンド観光は、中国経済の 1978 年から開始された改革開放以来の急速な成長と中国政府の国民出国政策の緩和に伴い、急速な成長を続けてきた。中国大陸からの出国人数は、1998 年の 842.56 万人から 2018 年の 16, 199 万人に増加し、現在、中国は世界の出国観光客数が首位であり、アウトバウンド観光の支出額の面でも 2770 億ドルと世界の 5 分の 1 を占め、これも首位である。さらに、タイ、日本、中国香港、中国澳門、中国台湾、ベトナム、シンガポール、インドネシア、ロシア、カンボジア、オーストラリア、オーストラリア、フィリピン等の中国近隣の国と地域におけるインバウンド観光客数もここ数年、中国大陸が首位を占めている。[7][8][9]

そして、2016 年の中国の出入国管理部門の報告によると中国国内の住民でパスポートを持っている人数は 1.2 億人と当時の人口の 9%にも満たなかったが、2027 年には 3 億人になると予想されており、中国大陸からの出国者数は今後も大幅に増加する見込みである。こ

のことから、中国大陸からのアウトバウンド観光には巨大な市場潜在力が存在していると言える。[10][11]

しかし、中国大陸からの観光客の観光消費における習慣と選択のパターンは独特の文化を形成している。「ニールセンとアリペイ」の調査によれば、図 1-3 に示される通り、インバウンド観光における消費傾向（ショッピング、宿泊、飲食など）の面で、中国大陸観光客と非中国大陸観光客とは明らかに異なる傾向を持っているため、中国大陸観光客の特性と習性は観光産業の発展等に影響を及ぼすことが指摘されている。[12][13][14]

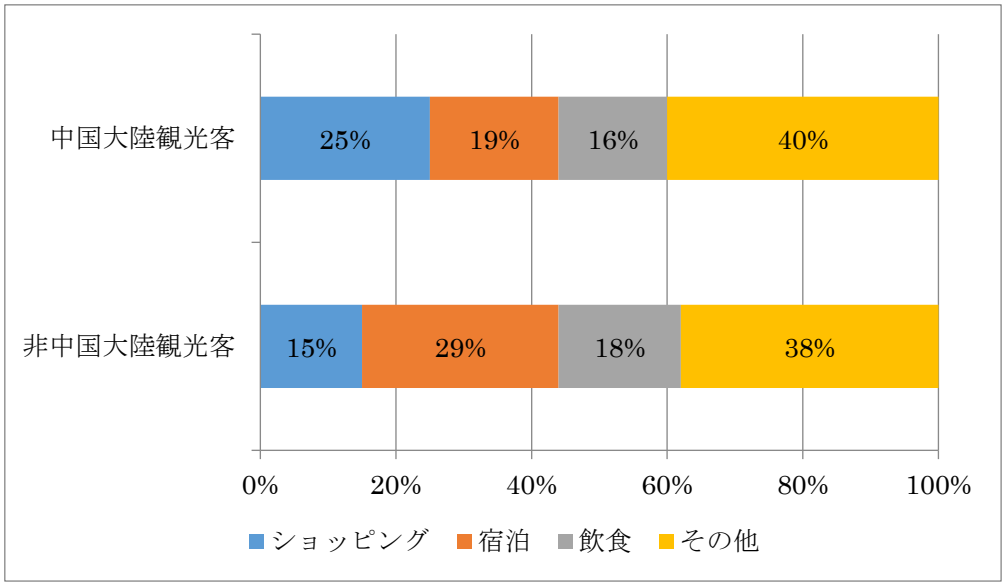


図 1-3　中国大陸と非中国大陸観光者が観光する際の消費配分

1.2.3　インターネット情報化による「OTA」業界の発展と観光レビューに関連する背景

オンライン旅行代理店（OTA）はインターネットの発展と普及に伴い、観光産業内で育まれた新しい産業発展の原動力であり、「OTA」業界はポジショニングサービス[4]、オンライン広告、オンライン予約と支払い、そしてオンラインレビューなどの機能を結合した新しい観光商品・サービスの販売モデルを提供する業態である。

世界の観光業と「OTA」業界の発展速度は図 1-4 に示した通り、世界の「OTA」業界が毎年 10%近くの成長率を維持している一方で、図 1-1 から読み取れるように、観光業全体の成長率は平均 6%しかない。しかし、中国大陸の「OTA」業界の発展速度に注目すると、図

[4] マーケティングにおける製品・サービスの他社との差別化を行うこと。

1-5 に示した通り、2014 年~2018 年成長率は平均 30%以上の増加を維持しており、将来的にも 15%の増加を維持することが読み取れる。この背景には、中国大陸のキャッシュレス決済が世界で最も発展していることもある。

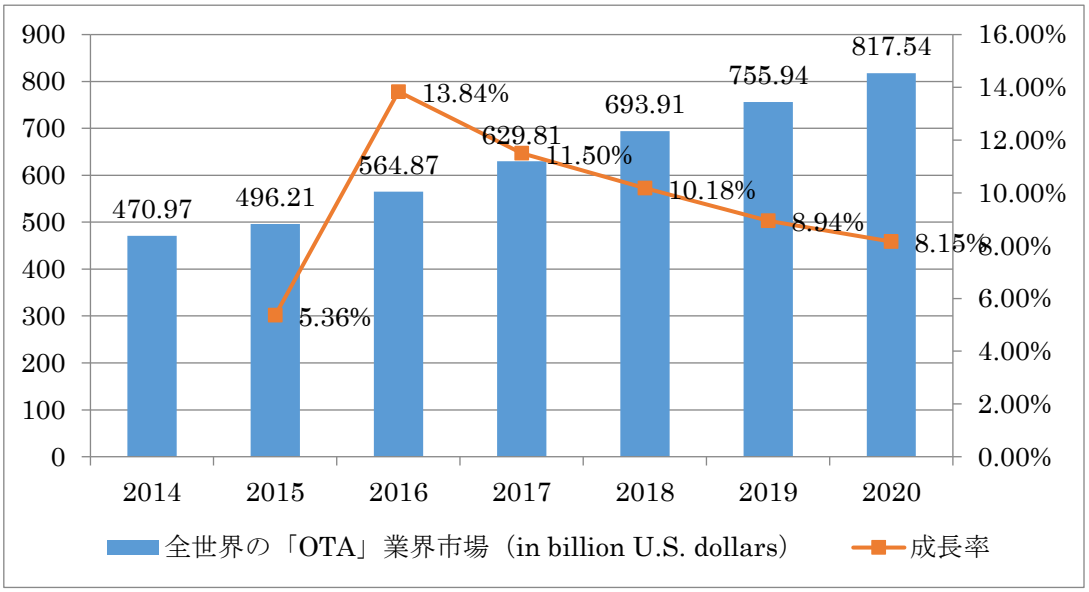


図 1-4　全世界の「OTA」業界市場[15]

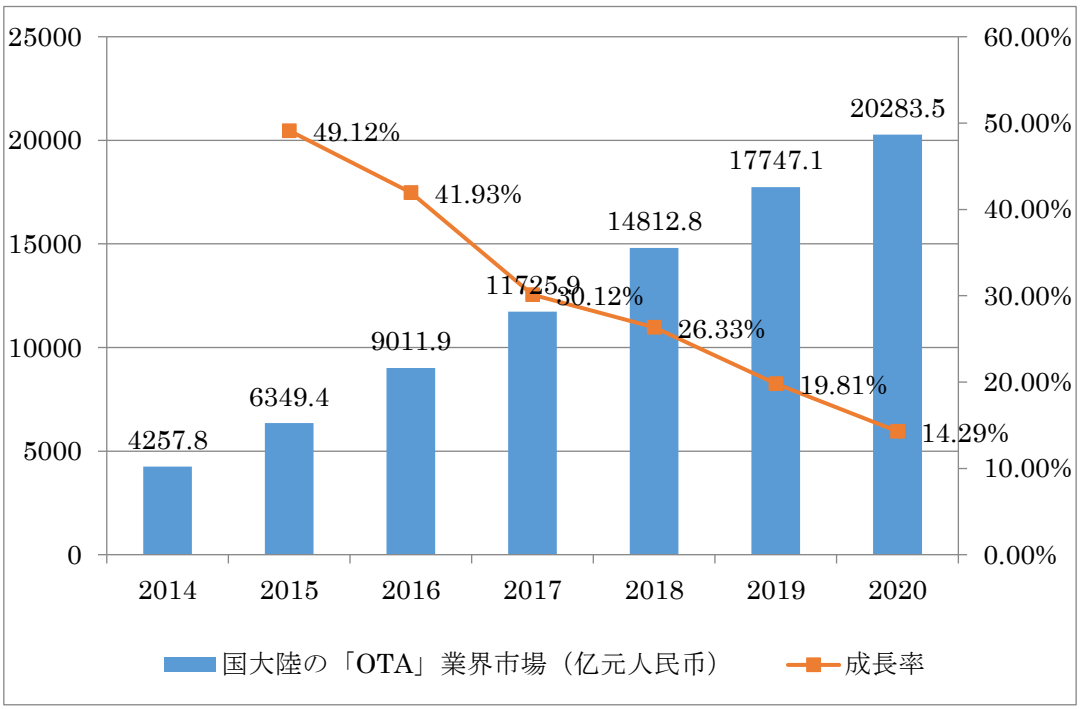


図 1-5　中国大陸の「OTA」業界市場[16][17]

また、ビッグデータの解析技術の発展に伴い、「OTA」業界の観光に関するデータの集約

機能は多くの新たな付加価値を生み出すことにも繋がっており、その中でも観光レビューに関するデータは観光産業の発展に貢献するような付加価値の高い情報が多く含まれている。そのため、本研究で分析を行うデータ群の基礎ともなっている。観光レビューの投稿機能は当初、観光消費者が観光事業者にフィードバックするプラットフォームとして作られたが、データの蓄積量の増加とユーザーの使用用途が多岐になっていくことに伴い、このプラットフォームは観光消費者のニーズを受け取めるツールや観光事業者の調査のプラットフォームとしてだけでなく、顧客が観光商品をオンラインで選択・予約する際にも多く利用されるようになったことから、観光業界全体にとっても非常に重要な役割を果たすようになった。

アンケートに調査[18]では、回答者の 95%がオンラインで宿泊施設を予約する際に観光レビューを参考にするという結果が出ており、今や観光レビューは観光情報を発信する個人の自己メディア（self-media）とも言え、「OTA」ユーザーの宿泊施設や観光商品の選択プロセスに大きな影響を及ぼすと考えられる。

また、オンラインの観光レビューは観光商品を「OTA」ユーザーに宣伝する側面を持つと同時に、観光事業者に対しても一定の牽制力を持つという側面も有している。このように、今後も観光業と情報技術との融合が深まるにつれて、観光レビューはさらに大きな役割を果たすことになる。

1.3　研究目的

前節 1.2 で述べたように、タイや日本、中国香港、中国澳門、中国台湾、ベトナム、シンガポール、インドネシア、ロシア、カンボジア、オーストラリア、フィリピン等中国の近隣の多くの国や地域を訪れる観光客数は中国大陸からのアウトバウンド観光が最大である。

加えて、中国大陸のアウトバウンド観光産業にはいまだ大きな潜在力があるため、中国大陸からの観光客はますます各国の観光産業が競合するターゲットになっている。同時に、中国大陸観光客は非常に明確な特性を持っているため、他国の観光事業者は交通機関、宿泊、飲食、消費などに関する中国大陸観光客の観光習慣や嗜好に関する理解を深くする必要があり、これらの特性を結び付けた形で観光商品・サービスを改善し、どのように中国大陸観光客の満足度を高めるかを考えることが重要になっている。

一方、「OTA」業界は、インターネットやキャッシュレス決済のグローバルな発展などと共に従来の物理的な店舗や電話などを用いた観光業界の形態からオンラインの形態に早い速度で変革してきている。「OTA」業界は、観光産業の国際的な発展を加速させただけでなく、オンラインの観光レビューのプラットフォームを通じて、多くの観光客に新しい観光消費のモデルを提供することになった。その主な役割は、将来の観光客の選択のために参考となる情報を提供すること、観光事業者にも観光商品・サービスの改善や新たな企画に関して直接的にアドバイスを提供することの 2 点である。

ただし、観光消費者個人あるいは単一観光事業者の個別の視点からは、これらの観光レビュー活用の成果は非常に限られ、観光業界全体や特定の観光地全体における観光事業の経営資源の調達や配分といった観点からの分析や提案をすることはできない。したがって、観光レビューに関する大量のデータを有効かつ効率的に統合・分析し、その結果を活用してターゲットとなる観光客の特性や嗜好に応じた観光商品・サービスを提供できるかどうかが、今後の観光業界の持続的発展に非常に重要である。

また、従来は観光客の特性や嗜好を理解するための手段はアンケート形式が多かったが、アンケート形式による調査・分析には以下に示す課題がある。

1) 観光客の特性と嗜好は継続的に変化しているため、その持続的な調査には多大なコストと時間が必要である。
2) 国毎の違いはもちろん、同じ国でも異なる観光地に行く中国大陸観光客はそれぞれが異なる特性と特有の嗜好を持っている。このような観光客群の差異を明らかにするためには異なる国と地域での調査を同時かつ網羅的に行う必要があるが、実施するのは非常に難しい。
3) アンケート調査は質問の対象を絞っているが、質問を設計する際には専門的な知識が求められ、特に地元の観光産業の状況を十分に理解する必要があるため、異なる地域毎の状況を全て理解した上で実施することが困難である。

4) 質問の順序や文章の書きぶりやトーンが回答者の回答内容に影響を与えるとともに、質問項目以外の内容を聞き出すことが限定的であることから、回答の自由度が限られている。

これに対して、「OTA」Web サイトに投稿される観光レビュー[5]は、上記のような課題を以下に示す理由から、解決できると考えられる。

1) 「OTA」Web サイトに投稿される観光レビューのデータは、すべて最新かつ無料で各社の Web サイトから収集することができる。
2) 異なる国や地域、さらに観光地の「OTA」Web サイトの観光レビューはそれぞれ既に分類された形でまとめられている。
3) 専門性の高い観点は不足しているが、観光客のレビューは全て現地の観光商品・サービスに関するユーザーの実際の体験に基づいた投稿であり、対象が明確で、様々な地域に関するデータが存在している。
4) 全ての観光レビューは観光消費者自身が自発的に書き込むため、アンケート調査のように質問者や質問内容に干渉されたり制限されることがない。

以上の理由から、本研究では「OTA」Web サイトの観光レビューのプラットフォームが所有する中国大陸観光客の観光レビューデータの分析において、データの特性を考慮し、ターゲットを絞った手法を利用して、「OTA」観光業界の特性・傾向と中国大陸観光客の特性・嗜好を分析する。その上で中国大陸観光客向けのインバウンド観光産業に注目し、訪問先の観光情報と特性に合わせて経営資源配分とサービスレベルを最適化し、中国大陸観光客の満足度を向上させると同時に、国を超えた文化交流と融合を促進し、受け入れ地域や国の観光産業を振興することを目的とする。

以下、この目的に基づき、第 2 章では本論文に関連する先行研究を 4 つの研究対象別に分けて考察し、第 3 章では観光レビューのデータ特性に基づき、それに適合する手法の検討を行った上で、「OTA」観光業界の特性・傾向と中国大陸観光客の特性・嗜好に関する分析・考察を展開する。次に、第 4 章では中国大陸観光客がインバウンド観光する際の情報の非対称性、つまり中国大陸からの観光客がインバウンド観光情報を完全に把握していないという特性に着目した分析と考察を行い、その過程において、未知情報を把握・予測するという目的にあった処理方法を検討する。さらに第 5 章では、分析対象とする観光レビューの範囲を広げ、相関分析を行う対象データを宿泊施設から観光地・スポットに拡大する。また第 3 章・4 章の内容の予測分析と傾向分析を統合した上で未知情報を活用し、観光業界の現場業務に反映することを容易にする適用方法を提示する。そして最後に第 6 章では、第 2 章から 5 章で考察した内容の結論と本研究の限界・今後の課題をまとめる。

[5] 「OTA」レビューとは、オンライン旅行代理店に記載している観光商品やサービスへの評価を意味する。

第 2 章　先行研究概要及び本研究への示唆

Parrinello[20]は 1993 年にそれまでの観光業を対象とした研究の成果をまとめ、産業革命後の社会に観光業がもたらした以下の 6 つの特徴と傾向を総括している。

1) 人々の自由時間が継続的に増加し、社会における休暇や旅行がより高い関心を持つようになってきている。
2) 生産活動の地方分散化及び第三次・第四次産業の緩やかに成長により、自由時間における祭事や文化的な活動、観光の重要性が増大する。
3) 交通手段の発達に伴うモビリティの向上により、連続的な空間移動が可能となり、社会活動や自分の存在意義に対するインセンティブや熱意の持ち方が重要になる。
4) 産業革命後の社会の特徴には、生態的な脅威の認識と、それに伴う「自然」の再発見及びありきたりな観光地ではない場所への観光や観光の形態に対するニーズの高まりがある。
5) 新しい形態の人と人との交流や地域社会での生活等の品質に対する社会ニーズがますます重視される。
6) 国際社会における生産活動の地域分散化などの見通し等を含め、情報通信サービスとテレビを介した情報を発信が重要になる。

このような、現代の観光産業発展の背景となる状況を踏まえ、前節 1.3 の研究目的で述べた通り、本研究は中国大陸で急速に成長している「OTA」業界を対象とし、訪問先の観光情報と特性に合わせた経営資源配分と観光商品・サービス最適化を通じて、中国大陸観光客の満足度を向上させ、人々の文化交流と融合を促進し、受け入れ現地の観光産業を振興することを目的とする研究である。

そのため、本章ではまず関連する先行研究を、「中国大陸からのアウトバウンド観光に関する研究」、「オンライン旅行代理店業に関する研究」、「観光事業の情報化に関する研究」、「日本で観光するインバウンド中国大陸観光客に関する研究」の 4 種類に研究対象で分別し、それぞれについての考察を展開する。

2.1　中国大陸からのアウトバウンド観光に関する研究

本節では、中国大陸からのアウトバウンド観光に関する研究を観光産業と観光客の 2 つの視点で分けて考察する。その理由は、この 2 つの視点は互いに関連性が高いが、それぞれの注目点がはっきり異なるからである。また、これまでの研究においては、中国大陸のアウトバウンド観光全体に着目した研究は極めて少なく、ほとんどの研究は特定の国か地域を観光する中国大陸の観光客を例として限定的な分析や考察を行い、それを中国大陸観光客の特性と中国大陸アウトバウンド観光産業発展の規則性としてまとめているが、実際、そのような結果は対象となった国や地域によって異なる。このように先行研究が示した異なる観光訪問先により生じた結果の多様性は、中国大陸アウトバウンド観光に関する研究において、分析対象とする事例サンプルの範囲を選定する際に、十分な考慮が必要であることを示した。

2.1.1　中国大陸のアウトバウンド観光産業に関する研究

Zhang と Heung[21]は、2001 年に中国大陸のアウトバウンド観光市場の発展は「リップル効果」（ripple effect）との一致性が良好であること示した。リップル効果とは、池に投げ込まれた石が波紋を広げるように、時間が経過するにつれて、アウトバウンド観光の市場発展の場が中国大陸から地理的に遠くなっていくということである。

このような市場発展のプロセスは時間の経過によって四段階に分けられる。第一段階は、国内観光の増加である。この段階ではアウトバウンド観光に触れてはいないが、アウトバウンド観光産業の発展には十分な国内観光市場の増加がポテンシャルとなる。第二段階は、1983 年から香港とマカオが中国大陸観光客の観光地となった段階である。第三段階は 1990 年からアジア近隣国を中心にアウトバウンド観光が始まった段階であり、特にシンガポール、マレーシアとタイへの観光ルートは当時、中国大陸で最も人気のコースとなった。そして、第四段階は 1999 年からニュージーランド、オーストラリア等の国々が国民の「approved destinations（出国認可先）」と承認された段階である。その後、中国はアウトバウンド観光客数が急速に伸び、世界でも首位となった。

中国大陸のアウトバウンド観光に関する中国政府の政策や政治面についての研究と議論は多数ある。Tse[22]は、中国は中央集権的な社会主義国家であり、この体制はアウトバウンド観光の政策展開においても顕著であることを指摘した。同時に、中国大陸人の富が増加し続けることにより、アウトバウンド観光産業の重要性が増し、アウトバウンド観光産業も中国の外交や他国との交渉におけるソフトパワーとなり、政治と密接に結びつくようにな

った(Tse[23])。2013 年に Mak[24]は政治のイデオロギーと中国大陸のアウトバウンド観光との相互関係を研究し、イデオロギーが中国大陸のアウトバウンド観光発展において重要な役割を果たしているとみなしている。しかし、伝統的な観光そのものにおける要素と政治的な要素は共に観光産業の発展に影響を及ぼすものの、現在のアウトバウンド観光の発展状況には、政治的な制限や統制の影響は次第に緩められており、中国大陸国民そのもののニーズに、より合致した発展を遂げているものと考えられる。

また、観光の形態及びその変化は多くの研究で言及されている。例として、2007 年の Guo ら[25]の研究によると、観光会社が主催した団体ツアーは過去の 1993~2004 年の 10 年間で急伸し、また団体ツアーに参加する中国大陸観光客の特性が指摘された。例えば、団体観光客は、観光の支出と時間を考慮しながら、単一の国か地域ではなく、複数の国や地域への観光を選ぶ傾向が高いことが挙げられた。しかし、観光商品・サービスの改善と同時に、個人旅行への政府の制限も次第に緩められ（例としては、2013 年に香港とマカオへの個人観光の認可開始[26]）、入国手続きも次第に簡素化され、それまでの伝統的な観光形態も大きく変化している。

2017 年に邱[27]は、アウトバウンド観光における観光商品・サービスの選択肢及び観光客数の増加に伴い、ますます多くの人が既存の観光会社が提供する観光商品・サービスに満足せず、自分自身の嗜好に合わせて観光の目的地を選ぶなどする個人旅行を選ぶ人が徐々に多くなることを示した。中国観光研究院及び Ctrip（携程）が提供するデータを統合して分析した結果から、2018~2019 年の中国大陸からのアウトバウンド観光ツアーには団体ツアーを選ぶ人数と自由な個人旅行を選ぶ人数が同程度になったことが示され[28]、その中で中国大陸からの観光先で首位のアウトバウンド先である香港については個人旅行の割合が七割以上を占めていることが明らかになった。

また、2013 年に Arlt[29]は、近年の社会経済の著しい発展により「New Chinese Tourists（新・中国観光者）」が登場したと提唱した。このような新しいタイプの観光客のニーズを満たすために、観光事業者はソーシャルメディアや他の形式の口コミ等にも注意を払い、この新しいタイプの観光客に対する観光商品・サービスの魅力とプレステージを確保する必要があり、そのためにも情報産業との連携を絶えず深めることが非常に重要であることを示した。

2.1.2　中国大陸のアウトバウンド観光客に関する研究

前節 2.1.1 で述べた観光政策の側面の観光産業の発展への影響力に比べて、観光事業者の側面はそれほど決定的ではない。しかし、観光事業者は、政策的な側面には直接働きかけることができないため、自らの活動で改善可能な観光客の側面からの影響の方が実質的には

大きい。したがって、本研究では、観光客のニーズに応じて、観光客の満足度を高めることで産業の発展を促進することを目的のひとつとする。

観光客の視点については、ほとんどの研究では社会・文化的な側面を着目点としている。1999 年に、Mok と DeFranco[30]は、儒教、道教と仏教を根本とする中国の文化的価値観が中国大陸観光客の全体的な行為にどのような影響を与えているかについて研究を行った。結論としては、「権威への尊重」「相互依存」「メンツ」「集団的志向」「調和」「外因」（Respect for Authority、 Interdependence、 Face、 Group Orientation、 Harmony、 External Attribution）の 6 つの文化的価値観が中国大陸観光客の行為に影響を及ぼしているとした。Li ら[31]も、文化と経済の違いにより中国大陸観光客は特別な観光特性や嗜好などをもっているかもしれないが、西洋のマーケティング担当者がまだよく理解していないという点を指摘した。

また、多くの研究者が中国大陸観光客の特性を究明するために、中国大陸観光客と他国の観光客を比較している。例としては、2006 年に、Arlt[32]は中国大陸人観光客と西洋人観光客、及び中国大陸観光客と日本人観光客との間の行動の違いに関して考察した。2007 年に Kwek と Lee[33]は、中国大陸観光客とシンガポール華人観光客を対象としてクイーンズランドへのイメージに対する潜在的な違いを調査し、文化の違いによる商品のマーケティングへの影響を強調した。2013 年に、Cros と Liu[34]は、香港にいる中国大陸観光客のうち青年観光客を対象とする研究で、西洋観光客と比べると、一般的なオリエンタル文化よりも、香港地元の文化への興味が強いことを主張した。そして、これらの研究から中国大陸観光客は強い自己属性を持っているという結論が導かれた。

上述したように、観光客のニーズ、満足度、モチベーション等に関する大多数の研究は中国大陸から世界各地へのアウトバウンド観光客全体を対象とするのではなく、中国香港や中国マカオ、東南アジア、ヨーロッパ、オーストラリア、アメリカなどひとつの目的国や地域に限定された中国大陸観光客を対象としている。そして、これらの研究を通じて、目的地によって中国大陸観光客に影響する要素も異なることが明らかになった。Ye ら[35]の 2012 年の論文では中国大陸観光客が観光している際に持つ不満と現地で受けた差別に言及し、観光客に対して予想される差別と実際に知覚される差別からの影響をどうやって減らすかについて検討を行った。

2007 年に、Chow と Murphy[36]は専門家へのインタビューとアンケート調査により、中国大陸観光客はオーストラリアの観光では、「食事」、「観光」、「文化遺産」、「参加型イベント」、「娯楽とショッピング」をより好んでいることを示した。2011 年に、Corigliano[37]は、中国大陸観光客がイタリアを観光地として選ぶ動機に関して検討を行い、中国大陸観光客が「歴史と芸術」、「自然景観」、「生活と伝統」、「レジャーとスポーツ」に興味を持っていることを示唆した。以上のことから、中国大陸観光客のニーズをまとめて考慮するのではなく、訪問した国や地域、さらには訪問した景観地の要素まで細かく考慮することで、観光客のニーズを正確に把握し、満足度を高めることができると言える。

最後に、中国大陸観光客の時系列的な変化にも着目する必要がある。Zhang と Lam(1999)[38]は、香港への観光を例とし、観光客の動機を分析する手法を用いてプッシュ要因とプル要因に分けて分析を行った。香港を訪れる観光客の最も重要なプッシュ要因が「知識」、「名声」、「対人関係の強化」であり、最も重要なプル要因が「ハイテクイメージ」「支出」「アクセスのしやすさ」であるとした。しかし、2003 年に、Hsu と Lam[39]の研究によると、そのどちらの要因でもなく、観光が最も重要な動機であった。2011 年に、Li ら[40]は香港を訪れる中国大陸からの女性観光客の観光動機と行動モードをまとめて、4 つのプッシュ要因と 5 つのプル要因を確定した。プッシュ要因は、「知識と名声」、「社会関係の強化」、「休息とリラクゼーション」、「冒険と刺激」であり、プル要因は「現代的なイメージ」、「自然関係と景観地」、「安全と清潔」、「観光手配の利便性」、「ショッピング」である。プッシュ要因とプル要因には相応な関係があるが、明確な相違もあるため、時間的な要素も観光客の特性の重要な変化要素であることが示された。

2.2　オンライン旅行代理店業に関する研究

過去 20 年間、人々が情報通信技術（Information and Communication Technologies、以下 ICTs と略す）が観光業と観光客の行動に与える影響はますます大きくなってきている(Amaro と Duatre[41])。2008 年に、Buhalis と Law[42]は、インターネット技術が観光業の産業構造に影響を及ぼすと予測、特に「eTourism」の将来は観光消費者を中心とする技術に焦点を当てられることを示した。これらの技術は観光事業者とその顧客との動的な交流を促進するものであり、実際、このような技術の発展は観光客の行動を徐々に変えてきた。観光客はインターネットで情報を検索し、観光計画を立て、観光商品やサービスを購入する。また、これらのインターネット技術もより付加価値が高く、多用な提供形態に進化してきた。

そして、2005 年に、Jeong と Choi[43]は、有用で明瞭かつ完全な情報を提供することは観光消費者の強い行動意図に繋がると指摘した。この部分指摘に関して、本節では本研究との関連性を考慮した上で、主にオンライン旅行代理店のユーザーの観点と、そのユーザーが投稿する観光レビューの観点の 2 点から考察を展開する。

2.2.1　オンライン旅行代理店（OTA）の観光消費者に関して

観光消費者に関する統計データは最もよく研究されている要素のひとつである(Amaro と Duatre[41])。観光消費者としての「OTA」顧客の特性の分析については、まだ確たる結論もなく研究結果もまちまちである。例えば、観光消費者の年齢からみると、2004 年に、Wolfe ら[44]の研究は、若い観光消費者は「OTA」を利用して消費する可能性が高いことを示したが、2009 年の Moital ら[45]の研究では「e−commerce in travel（旅行における電子商取引）」の利用者は年齢、性別、経済的地位や教育レベルに大きな相違がないことを示した。また、2004 年に、Kim と Kim[46]の研究でもオンライン購入者と非オンライン購入者の収入レベルに違いがないことが示唆された。その一方で、Wolfe ら[44]は「OTA」を利用する観光者は高い収入水準にある可能性が高いと指摘した。

Amaro と Duatre[41]はこれらの人口統計的な変数に基づく分析結果相互矛盾の原因は「OTA」が提供する観光商品の購入者の人口構造変化にあると主張している。これはインターネットの普及により、「OTA」が提供する観光商品の購買が低所得層や教育水準の低い個人の間でも一般化しているからである。しかし、その相互矛盾の発生の背景にはサンプリング方法の違いや国の文化の違いなど他の要因による可能性もあるのではないかと考えられる。

このような状況を踏まえると、「OTA」の顧客に関する研究をさらに細分化し拡張する必

要がある。例えば、一部の研究では「OTA」顧客の観光に関する行動を分析し、「OTA」の利用を通じた観光消費行為は拡散的であることが示唆されている。具体的には、2001 年に Morrison ら[47]は、観光レビューの閲覧者をリピーターと一回限りの予約者を区切った分析モデルを構築し、特に「伝染性」という要素を考慮した。そして、他の人がインターネットで観光商品を購入したことを観光レビューの閲覧を通して知れば、その閲覧者もその商品を買う可能性が高くなると結論づけた。そのため、より多くの人により多くのオンライン Web サイトへの観光レビューに参加してもらうことは、「OTA」業界の発展にプラスの効果があると言える。

また、オンラインの観光レビューは「OTA」の Web サイトを通じて利用客が旅行の事前計画から、実際の観光時に、さらには観光後の感想や振り返りまで一貫してまとめて支援することができ、観光消費者の参加頻度と熱意やモチベーションを十分に高めることができる。

「OTA」利用のメリットに関する研究では、主に技術受容モデル「Technology Acceptance Model」の中の 2 つの要素：知覚された有用性と知覚された使用容易性（Perceived Usefulness と Perceived Ease of Use）に着目している。オンラインシッピングの優位性に関する研究においては、2005 年に Chang ら[48]は知覚された有用性と知覚された使用容易性が「OTA」が提供する観光商品の優位性として広く認識され、オンラインシッピングに顕著なプラス効果を与えていると指摘したが、同時にオンラインショッピングには商品そのもののリスクと取引中のリスクがあるとも指摘した。この点に関連して、2007 年には、Kamarulzaman[49]は従来の知覚された有用性と知覚された使用容易性に基づいて観光消費者の個別の特性に感知のリスク、信頼の要素を追加し、英国の観光消費者がオンラインで観光商品を購入する際の意向がどのような要素に影響されるか調査を行った。その結果として、知覚された有用性がオンラインでの観光商品購入と正の相関を持つ一方で、知覚された使用容易性はオンラインでの観光商品購入に直接の影響がないとした。このことから、先行文献ではオンラインでの観光商品購入における知覚された有用性の優位性が普遍的に認められていることが確認された。

2.2.2　オンライン旅行代理店レビューに関する研究

2012 年の Vlachos[50]の研究によると、国際観光者の約 87%がインターネットを利用して観光を計画し、そのうち 43%が他の観光者のレビューを読んだことがあるとしている。Stikky Media[51]も Web サイトの観光レビューやソーシャルメディアが観光者の予定に大きな影響を及ぼしているとし、そして 2012 年のデータの分析に基づき、「OTA」Web サイトの観光レビューを閲覧した人の半数以上が、レビュー内容に基づいて観光の計画を修正・

調整していると指摘した。このような観光客の行動に関して、2009 年に、Vermeulen と Seegers[52]はオンラインの観光レビューが観光商品・サービスの選択や観光客の行動に及ぼす影響を明らかにした。また、2017 年の張ら[53]の研究によると、観光に関するオンラインレビューは観光消費者の観光商品とサービスへの認識度を反映し、商品・サービスの提供側のマーケティング上の評判にも影響をおよぼす。観光事業に携わるマーケティング担当者は、オンラインの観光レビューモニタリングしながら活用することで観光商品・サービスの改善、マーケティングの成功率を向上させ、安定した観光ブランドを確立することができる。観光消費者にレビューを投稿してもらうためのプラットフォームを提供する「OTA」企業（アメリカの TripAdvisor や Booking.com、中国の Ctrip（携程）、日本の楽天トラベルなど）は、オンラインレビューの重要性を十分に認識、重要な資産と見なしており、更に観光事業者や観光地を管轄する自治体の観光行政部門の情報源にもなっている。

多くの研究者もオンラインの観光レビューとオンラインによる観光商品・サービスの購入行動との連動性の分析を通じて、オンラインの観光レビューの意義を議論している。2008 年に、Hu ら[54]は取引コスト理論と不確実性低減理論に基づく分析を通じて、オンライン観光消費者の観光レビューが取引の不確実性とコストを下げ、最終的な購入の意思決定を助けるという結論を提示した。一方、観光消費者はレビューの真偽を判別することができることから、観光レビューをその真偽性で仕分け・区別しており、レビュー内容を意図的に操作しようとする行動はほとんど発覚してしまうため、観光レビューの観光消費者の行動への影響に対しては独立性が強いと結論づけた。2006 年に Kumar と Benbasat[55]は、Web サイト上で有益なレビューを提供することで、Web サイトのユーザーどうしのコミュニケーションを促進することができると提唱した。言い換えれば、観光レビューは他の観光消費者からの Web サイトへのアクセスを引き付け、サイト上のプラットフォームに留まる時間が増加し、最終的には「OTA」企業の Web サイトでの観光商品・サービスの売上を増加させる。そして、更に観光消費者の観光レビューはその Web サイト自体の有用性や社会性を増加させることになる、このことも結果として「OTA」企業の売上にプラスの影響を与える。また、2016 年に、陳[56]は中国大陸観光者のオンライン観光レビューが観光商品・サービスの購買行動に与える影響に関する要素モデルを構築し、それぞれ観光オンラインの観光レビューWeb サイトと観光商品・サービスのプロバイダーにマーケティングとマネジメントに関する手法を提案することで、マネジメントの観点からオンラインの観光レビュー実質的な意義を定義した。

最後に、前節 2.2.1 で言及した知覚された有用性も多くのオンラインの観光レビューを考察する上で重要なポイントであるが、2015 年に、Liu と Park[57]は、観光レビュー投稿者の個人情報の開示レベルや、投稿内容の専門性の高さや評価といった特性に関する様々な文献を研究することで、オンライン観光レビューの知覚された有用性に影響を及ぼす要因を分析した。その要点は下記の 3 点である。

1)　レビュー投稿者の情報開示は、投稿されたレビューの知覚された有用性にポジティブ

な影響を与える。例として、2008 年に、Forman ら[58]は、オンラインの口コミと売上、投稿者コミュニティ内での認知度やレビュー評価システム及びオンラインコミュニティの規範の順守に関する研究の意義に関する考察を行った。Forman らは、オンラインコミュニティのメンバーが身分の情報を含むレビューに対して評価を受けることにより積極的であり、投稿者の身分開示の普及は、オンラインにおける商品の売上増加に繋がることを指摘した。

2) 専門知識レベルの高いレビュー投稿者は投稿レビューの知覚された有用性にポジティブな影響を与える。2008 年に、Cheung ら[59]の研究により、観光消費者はレビューが専門性と信頼性の高い個人によって投稿されたと判断すれば、レビューの知覚された有用性も高くなることが示唆された。

3) レビュー投稿者がそのサイトで高い評価を得ていることはそのレビューの知覚された有用性にポジティブな影響を与える。ただし、それは、潜在的な観光消費者の情報獲得の視点に基づくものである。実際に、熟練観光客であれ一般の人であれ、経験豊富なリピーターであれ、初めてその観光地を訪れる観光客であれ、自分なりの視点を持っているため、観光事業者か自治体の観光行政部門にとっても、その視点は有意義な情報である。

2.3　観光事業の情報化に関する研究

これまで述べてきたように、現在のオンラインショッピング環境の下では、オンラインの観光レビューは観光消費者の観光商品・サービスの購入意思を決定するための重要な情報源となっている。しかし、観光レビュー情報の急速な蓄積は、観光消費者にとって情報の処理と活用する際にこれまでにない課題に直面しており、テキストマイニング技術を研究することの意義とそれを実践することの価値がますます顕著になっている。

現在、オンライン観光レビューの情報をデータマイニングする研究は、主にオンライン観光レビューの情報抽出、感情分析、テキスト分類に関する技術及びビジネスへの応用に着目している。2012 年に、Aggarwal と Zhai[60]はテキストマイニングによる問題解決は今後ますます注目が高まるとし、このような研究方法として各種のテキストマイニングのアプリケーションを効率的に処理できる方法の開発とアルゴリズムを設計する必要があり、情報へのアクセスを支援するだけではなく、ユーザーが情報を分析・理解した上で意思決定することを支援すべきであると主張した。

広義のテキストマイニングの概念には、自然言語処理、情報検索、データマイニング、機械学習など様々な要素が含まれている。しかし、具体的に観光商品・サービスのレビューに対してテキストマイニング処理を行った研究は少なく、原因としては、観光商品・サービスは相対的形態が多用で商品・サービス毎に性を持つことが考えられる。

観光事業の情報化の観点から見ると、「OTA」Web サイトの観光レビューなどのネット上のレビューはインターネット時代の産物であり、その前身ともいえる「Word-of-mouth（口コミ）」は既に 1960 年代にマーケティングと風評伝播に関する専門研究に登場している。1967 年の Arndt[61]は口コミが顧客の行動に与える影響を研究した最初の研究者である。彼は口コミが人と人とが交わす商品・サービス及びマーケティングに関する対面のコミュニケーションであり、このコミュニケーションは非商業的で非公式的なものであるとした。また、このような口コミは、最初は口で直接伝えられていたことから、ラジオ・テレビといった放送メディアの時代から、インターネットの時代に至るまで、かなり長い時間をかけて進化した。そして、オンライン観光レビューが、「OTA」Web サイトでも最も人気のある交流プラットフォームになる直前は、BBS 掲示板とブログが、非常に重要な観光情報源として研究対象となっていた。

例えば、2007 年に、Wenger[62]は「OTA」企業の経営者として、「www.travelblog.org」に投稿されたオーストリア観光についてのブログ記事を分析し、ブロガーと実際の観光客との関連性を探り、ブログ投稿の内容が実際の観光行動に与える影響を分析した。また、ブログの価値決定要因を把握することで、観光市場に影響を及ぼすコアブロガーを特定することにも活用された。

同様に、王ら[63]はテキストマイニングの手法を用いて「Sina ブログ」における上海近郊

の観光地、朱家角に関する 2633 件のブログ記事を分析し、キーワードの共起分析法とロングテール理論の組み合わせを利用して観光客の観光後の感想を把握し、観光発展のための提案としてまとめた。しかし、観光レビューを実際に使用したテキストマイニングの研究は、主にテキストの分類に集中しており、そのレビュー対象も表現が明確でデータも収集しやすい宿泊施設に関するレビューに集中している。Hu ら[64]は、「K−modoids」クラスタリングアルゴリズムを用いて「OTA」Web サイトにおける宿泊施設に関するレビューの文章をクラスタリングした。その上で、レビュー投稿者の過去の履歴データを信頼性の根拠として、最終的に有用性の高いレビューに限定して分析することで宿泊施設の評価に反映し、最終事例分析の結果は、Hu らが選出した調査対象者グループの評価とほぼ合致することが証明された。

また、ソーシャルアプリケーション[6]（social application）の発展に伴い、その部分の観光情報は観光産業の分析にも使用されてきた。例えば、加藤と石川[65]は Twitter のデータを用いて外国人観光客の訪日観光における行動を分析した。その分析の目的は、投稿データを含むソーシャルアプリケーション情報を利用し、外国人観光客の行動と経路を追跡、観光における行動と人気スポットを観察することで観光客の特性を把握することであった。

[6] ソーシャルアプリケーションとは、SNS（ソーシャルネットワーキングサービス）などのコミュニティをプラットフォームとし、ユーザー同士の繋がりや交流関係を機能に活かした Web アプリケーションのことである。ソーシャルアプリケーションは、ユーザーが個人で利用する一般的なアプリケーションと異なり、コミュニティあるいはコミュニケーションと密接に結びついている。ソーシャルアプリケーションを通じて他のユーザーとコミュニケーションを図ったり、情報を共有したり、あるいは繋がりの情報を活かしたり、といった機能が主に提供される。

2.4　日本で観光するインバウンド中国大陸観光客に関する研究

本研究において最も重要と考える分析事例（第 4 章と第 5 章で考察を展開）が、訪日した中国大陸観光客を対象とした予測分析と観察であるため、本節では日本で観光するインバウンド中国大陸観光客に関する先行研究を概観する。日本を訪問先とする事例を選定する理由は、観光地の観光レビューのうち、日本語で投稿されたものは投稿者が自国（日本）の観光客であることがほぼ明確で、インバウンド中国大陸観光客の投稿レビューとは区別し比較しやすいという利便性が期待されるからである。これが主に英語で観光レビューが投稿されるような観光地の場合、投稿者の国籍を選別しチェックする必要があるが、実際には難しい。また、近年、中国大陸からのインバウンド訪日観光客が急増しており、他国への中国大陸観光客に対するサービス向上にも役立つと考えられるからである。但し、事例研究の効率性の観点から中国大陸からのインバウンド訪日観光産業を対象としたが、本研究の枠組み自体は特定国への外国人観光産業に限るものではない。

この部分の研究は英語文献が極めて少なく、ほとんどは日中両国の研究者の文献である。また、中国大陸は改革開放前、経済が現在の様には順調に成長しておらず、アウトバウンド観光の発展もいくつかの政策で制約されていた。そのため、中国は広大な国家ではあるが、中国大陸観光客の訪日観光に関する研究は学術界において重視されなかった。しかし、中国大陸の改革開放以来、日本を訪れる中国大陸観光客が年々増加し、この本日インバウンド観光の迅速な発展は研究者達の関心を引き起こし、訪日観光に関する研究は近年の日中の研究者の注目分野になっている。

2.4.1　中国大陸観光客の訪日観光産業に関する政策の現状と発展に関する研究

日本政府の観光入国促進政策の浸透に伴い、外国人向け観光業が急速に発展していると同時に、中国政府も出国制限政策を緩和した。このような両国の政策が結合された結果として、観光業が急増したことを示す研究論文も出されている。

2007 年に、清水[66]は中国大陸からの訪日観光客の観光形態の変化、中国国内での日本への観光ビザの取得しやすさの地域間格差の変化、日本国内の中国大陸から観光客の誘致に関する地域間の政策の違い、観光事業者の誘致や観光商品・サービスに係る経営方針の変化を観察することで、関連産業の発展について意見を述べた。これは、観光会社が主催したグループツアーを中心とした 21 世紀初頭の観光形態と一致している。

2010 年前後の同じ時期に中国大陸の観光客の訪問先国としてのアメリカと日本を比較すると、主な違いは日本を訪れる観光客の知覚リスクは自然災害であり、アメリカを訪れる観光客の知覚リスクは社会の安定と入出国手続きであるということであった(Lai ら[67]) (Chew と Jahari[68])。第一に、このような違いは両国の国内情勢に関係がある。第二に、インバウンド観光政策が異なることも重要な影響を与えている。入国ビザの手続き簡略化のような最も直接的な政策は、中国大陸の訪日観光客の市場を活性化し、観光消費の基盤を強固にした(滕[69])。

2015 年以来、中国は日本にインバウンド観光者を最も多く送り出す国になったが、2016 年に鄔[70]は観光業界と観光政策に関する研究をまとめて分析した。それによると、中国大陸の訪日観光客の増加の原動力の中心となるのが、日本の入国手続きの緩和と中国大陸のアウトバウンド観光促進政策であると指摘し、このような協業により大きな市場潜在力を生み出すには、中国経済の発展と日本政府が中国市場を重視することの両方が必要であるとした。

観光産業に関連する他の研究も、中国大陸観光客の全体的な観光消費の傾向と観光行動空間との関連に注目している(黄[71]) (金[72])。

2.4.2　中国大陸観光客の訪日観光の選択要因に関する研究

前節 2.4.1 で概観した研究では、市場の発展に与える主要な観光政策の影響をマクロの視点から分析した。それに加えて、中国大陸観光客の団体ツアーからフリープランへの転換、団体ツアーによる初めて訪日からリピーターへの転換により、観光客の訪日観光形態の選択と観光目的など他の要素も注目されてきた。

例えば 2012 年に、別々の研究者がそれぞれ訪日の中国大陸観光客の観光先及び中国での居住地の観点から異なる考察を行った。戴[73]は中国人観光客の観光先と目的についての特性と重要な影響要素をまとめたのに対して、菱田ら[74]は、居住地の異なる中国からの訪日観光客の違いをまとめたが、同じ中国大陸観光客であっても、本国での居住地の違いにより観光先や季節の選択、観光形態の選択が異なることを明らかにした。

訪日観光の選択には他にも、日本の地理的・地盤的特性のため、地震などの自然災害が発生しやすいという状況が、大きな影響を与えている。例えば、Wu[75]は 2011 年の東日本大震災がもたらした影響について深く分析した。

観光産業だけでなく、日中間のいくつかの歴史遺留問題と政治問題も多くの研究の注目点となっており、主に観光客の選択への影響を研究している。2011 年には東日本大震災のため日本への観光客数が激減し、2013 年にも今度は日中の政治摩擦紛争により観光客数が大幅に減少した。2015 年に、郭ら[76]は、尖閣諸島を巡る政治摩擦などの紛争のポイント

について、記述的統計手法と多次元の相互決定ツリーモデルを用いて分析し、日本を観光する中国大陸観光客の認識の相違を要約した。調査の結果、釣魚島事件後、ほとんどの中国大陸の人は日本への観光に対して慎重かつ傍観的な態度を持ち、初めて訪日する人はなおさらであることが分かった。しかし、2014 年以降に訪日観光客が急激に増加していることを考えると、その影響が継続していないことは明らかである。短い時間での影響は顕著であったが、その後の訪日客数の反転はそれまでの負の影響の大部分をオフセットした。

また、2016 年に、Ji ら[77]は中国大陸の訪日観光客数の変化を感情の観点で分析した。その分析により、政治の衝突はある程度観光客の不安を引き起こすが、観光業の発展を妨げることはないという結論が出された。論文は、交流が進むにつれ、観光客が日本人に対する歴史問題と政治的宣伝（プロパガンダ）による誤解を減らすことができるとしている。特に日本の先進技術や大衆文化の観察・体験により、中国大陸人が日本人に対する敵意を弱めることができる。したがって、中国大陸からの観光客が地元の日本人と直接交流をとる機会を増やし、観光客がより深い文化交流を体験できるよう観光関連事業者に提案し、観光業界に対するイデオロギーの影響を軽減し、観光をより純粋なものにすることができると考えられる。

第 3 章　中国大陸観光客のレビュー特性に関する考察

第 2 章の先行研究に関する考察で述べたとおり、オンラインにおける観光レビューを対象とした研究が少なく、研究対象のほとんどが観光に関する BBS 掲示板やブログに書き込まれたテキストである。また、今までの研究では、テキストのポジティブな要素とネガティブな要素だけに注目し、文脈に含まれる価値のある情報の多くが分析対象とはされてこなかった。

また、オンライン旅行代理店「OTA」Web サイトのユーザーによる観光レビューを対象とする研究は、その大半がクラスタリング分析と情報検索処理を行うに留まっている。本章では、ユーザーによる観光レビューのコンテンツをテキストマイニングし、観光情報と「OTA」Web サイトの観光レビューの特性や傾向を抽出するために適切な方法を選定するための考察を行うことを目的とする。また、「OTA」Web サイトの観光レビューを使うことで、観光客のフィードバックから得られた情報の価値を最大限に活用し、観光事業者と自治体の観光行政部門が観光客の現状とニーズを把握することを支援する枠組みを提示する。

具体的には、まず観光レビューが持つデータとしての特性及びその特性が生じる要因について分析してまとめた。その後、データの特性を把握する上で、適切な処理方法を選択するための考察を行い、最後に中国廬山地域の宿泊施設に対する顧客レビューを対象として分析・考察を行い結論を出した。

3.1　データの特性と生じる要因について

中国大陸観光客が「OTA」Web サイトにおけるユーザーの観光レビューを分析・考察するためには、まず分析対象となるデータの特性を考慮する必要がある。この特性が生じる要因は主に、中国大陸観光客によるものと観光商品によるものの 2 つである。

3.1.1　中国大陸観光客の特性

中国大陸観光客の特性については、先行研究でも言及されているが、これらの内容を踏まえて本研究では次の 3 つの観点から考察を展開する。

1)　経済発展の背景

中国大陸の観光業の発展は中国政府が 1978 年に開始された改革開放の発展に伴い経済が安定的に向上した後に急成長したものであり、アウトバウンド観光もある程度まで経済が発展した後に、関連政策の緩和などに伴いようやく一般民衆にも利用できる商品・サービスになった。その結果として、中国大陸観光客数と観光客全体を構成する社会的階級の内訳が変化している。[78]

2)　中国大陸観光客の社会・文化的要素

第 2 章で述べたとおり、「権威への尊重」、「相互依存」、「メンツ」、「集団的志向」、「調和」、「外因」という 6 大要素が中国大陸観光客の消費に大きな影響を及ぼすと考えられる。また、第 2 章の比較研究においては、中国大陸観光客を他のアジア諸国の観光客と比較し、その特性と属性を定義した。また、一部分の先行研究から他国の中国系観光客には中国大陸観光客とは明らかな違いがあることも示された。そして、これらの結果から、中国大陸観光客に独自性があることが示唆された。そのため、単純に東方文化あるいは華人文化という枠で社会・文化の影響要素を一括して論ずることはできないため、本論文では、中国大陸の伝統文化と実際の国の情勢が中国大陸観光客に影響する重要な要素であるとして論を進める。特に中国大陸の情勢変化に伴う観光客の行動への影響は明らかに時代背景を反映しており、今後もその中国の情勢変化を観察し続ける必要がある。[30][32][33]

3)　中国大陸観光客の観光消費行動

第 1 章の図 1-3 に示したとおり、観光による訪問先でのショッピングについて、中国大陸観光客の消費は他国の観光客より遥かに多いことは明らかである。これは社会・文化的な背景に強く起因しており、「人情社会」という風習の下で、数多くの観光客が親戚にお土産を買うことが観光消費行動における重要な部分となっている。このような背景で注目すべきことは、中国大陸観光客の消費行動における嗜好に関する研究において、観光商品・サー

ビスの品質がますます重視され、より内容の充実した観光形態が普通の定型の団体ツアー観光の形態にとって代わっていく傾向にあることが示された点である。したがって、このような観光形態に関する特性も変化しつつあることが分かる。[79]

3.1.2　観光商品の特性

観光商品は商品ではあるが、実際には一般的に店頭で販売されているような商品より多様ではるかに多くの内容が含まれている。観光客が観光商品を評価する際も、観光商品そのものだけではなく、それに関連する要素を総合的に考慮することが多い。そのため、観光商品の評価は単純にその商品自体の良し悪しに限らず、関連のサービスと施設、さらに観光した年や季節といった時期的な関連要素も商品の評価に影響を及ぼす。そのため、一般的な商品との違いを考慮し、次の 3 つの特性としてとりまとめた。

1)　集積性（integration）

観光商品の販売には一定の観光エリアが必要であり、単一の観光地や単一の宿泊施設では観光客全体の消費や需要に応えることが困難であるため、ほとんどの観光商品は複数の観光地を訪問する形で販売されている。その組み合わせは、同質の観光地だけを周遊するのでなく、特徴の異なる複数の観光地の組み合わせも含まれる。

2)　関連性（relevance）

観光商品は観光客の食、宿泊、交通、遊び、ショッピング、娯楽など多方面の需要を満足させる必要があるが、このような側面は、お互いに独立ではなく強い連関を持ち、特にある個別の側面を評価する際には、関連する他の側面に対する満足度にも強い影響を及ぼすことを考慮しなければならない。これは「OTA」Web サイトにおけるユーザーの観光レビューでもよく現れる状況である。

3)　一意性（uniqueness）

観光商品は一般の製品とは異なり、時間帯が異なることだけでも商品に対する観光客の体験と体感に大きな差異が生じる。前述した季節、天気などの要因を除き、同じ観光スポットであったとしても、同時にその場で観光消費行動を行っているその他の観光客の行動が観光商品を通じた体験に大きな影響を及ぼす可能性がある。

3.2　データ分析手法の選択

観光レビューのデータを具体的に分析を行うためには分析対象となるデータのうちテキストデータを統合する必要がある。したがって、データ分析手法としては、自然言語処理に代表されるテキストマイニング手法が第一に考えられる。しかし、テキストデータが中国大陸観光客の「OTA」Web サイトのユーザーによる観光レビューであることを考慮すると、従来の通常の用語を対象とした自然言語処理では限界があるため、これらの特性に合わせた調整が必要である。先行研究では特性の要因分析において、時間とともに変化する属性が多く指摘されているため、本章では時間的な傾向分析に焦点を当てて観光レビューに関わるテキストデータを分析対象とした。また、適切な可視化処理を施すことで、観光事業者や自治体の観光行政部門のようなデータ分析を専門としない人々でも直観的かつ効果的に情報を利用することが可能となった。

3.2.1　自然言語処理

3.2.1.1　前処理

自然言語データ処理モデルの基礎は「単語」と「ドキュメント」の共起行列を構築するものである。しかし、分析目的への要求、元データでは直接行列を構築できないなどの問題により、重複データの削除や繁体字や簡体字の中国語としての統一など、単純ではあるもののいくつかの前処理を行う必要がある。その前処理で最も中心となる作業は、形態素解析[7](Morphological Analysis)である。これは中国語データが文字の特性上、英語と異なり（自然言語処理は英語に基づいて開発された手法である）、単語と単語の間に分割できるスペースがないことを考慮する必要があるためである。例えば 12 文字の中国語文「我是名古屋工业大学的学生(私は名古屋工業大学の学生です)」のように、中国語の基礎知識を持っている人は文の区切り方を理解できるが、コンピュータが直接その作業を行うことはできない。ただし、図 3-1 で例示した Rwordseg というパッケージ・ソフトウェアなど、既存のツールやソフトウェアを使用することで、前処理を行うことができる。

[7] 形態素解析とは，自然言語処理分野で主に事前処理として用いられる手法であり，対象となる言語の文法や単語の品詞情報をもとに，文章を形態素(単語が意味を持つ最小の単位)に分解する解析を指す．本研究は主に単語と単語の分割、すなわち、セグメンテーションを利用した。

図 3-1　Rwordseg を使用した形態素解析の例

Rwordseg はオープンソース・フリーソフトウェアの統計解析向けの R 言語で開発された中国語を対象とする形態素解析ツールである。このツールの最大の利点は、カスタム辞書に新しい単語を追加できるということである。本研究の解析では、観光専門用語や観光地の地名に注目する必要があるが、従来のセグメンテーションリソースには含まれていない。そのため、これらの新たに必要な用語や単語を独自に追加する必要があるが、Rwordseg を利用することでこの要求に対応できる。また、Rwordseg のもうひとつの利点は大量のテキストデータのセグメンテーション処理に用いることができるという点である。[80]

3.2.1.2　共起行列の構築

・単語の選択

単語の選択においては、その単語が頻出していることと実際の意味を含んでいる概念語[8]（notional word）であることが重要な選択条件となる。それに加え、「观光」（観光）と「旅游」（旅行）といったような、2 つ以上の類義語が組み合わされることも考慮しなければならない。このような点を考慮しながら本節では以下のような考察を行った。

まず最初に、後続分析の精度を高めるために、中国語類語辞書を利用して単語の類義語を組み合わせながら改めて再統計処理を試みたが、最終的な分析結果は有効ではなかった。これは、観光レビューの投稿者が微妙な表現でコメントを書くことで、一般的な中国語の用語・単語定義だけでは観光レビューに含まれる具体的な意味の全てを抽出することができないことが多発したからである。例えば、最初に分析を試みたマウンテンビューサンプル(「OTA」企業 Ctrip（携程）」が 2013 年に実施した廬山観光の評価で収集されデータ)では、「台阶」と「楼梯」という同義語は、中国語で「人々が低い標高と高い標高の場所を移動するのを助ける道具」という意味を表している。しかし、実際の観光レビューの表現を見ると、

[8] 概念語とは、概念的な単語間の文法的関係を単に表す関係語とは対照的に、人または物、行為、または質を示す語。

「台阶」のほとんどは観光客が屋外で使用する階段といった標高移動の道具であるが、「楼梯」の半分は構造物内部の階段を指しており、意味が明らかに異なる傾向が確認された。したがって、観光客のレビューにおける細かな表現の違いを十分に考慮するためには、観光レビューに含まれる細かな意味の差異を可能な限りに把握する必要がり、本章の研究では全体を通じて通常の中国語の意味特性を同義語として統合せずに、既存の単語をそのまま保持した上で解析処理を行うこととした。

・ドキュメントの選択

ドキュメントの部分の解析では煩雑な処理は比較的少なく、前処理でも簡単なスクリーニングを実施するに留めた。ただし、センテンスを単位とするのか、レビュー全体を単位とするかの選択をする必要があった。共起行列を構成した場合、各レビューの長さには差があるものの、下記の理由から、最終的にはセンテンスではなくレビューを単位として解析を行うこととした。

1) レビュー中の文脈間のつながりを維持する必要があるため。
2) 可能な限り個人を単位として表現を認識する必要があるため。
3) レビューの書き方が文法的に正しくない場合や、句読点の使い方も文法通りでない場合が多いため。

以上の理由により、ドキュメントとしての解析の対象はセンテンス単位ではなく、1 つのレビューを単位としたが、これに伴い、重複や統計した単語の出現頻度が 2 回以下の短文のレビューの削除を考慮する必要が生じた。

3.2.1.3 傾向分析

傾向分析を行うには、観光レビューのテキストを時間で区分する必要があるため、テキストデータ全体の構成を考慮しつつ、異なる時間区分における時間による変化の傾向を抽出する必要がある。まず、共起行列に基づいて時間で区分された区間ごとに特徴的な用語・単語を抽出し、対応分析の手法により区間ごとの差異を最も特徴的に表しているその区間の代表的な用語・単語を抽出する必要がある。その後、潜在的意味解析（Latent Semantic Analysis）手法を用いて単語間の関連度を分析し、最後に多次元尺度構成法を用いて示し、各区間の特徴を示す語群を集計し可視化することで、結果のまとめと考察行うこととした。

3.2.2 潜在意味解析（Latent Semantic Analysis）

観光レビューにおける潜在意味の処理は、本研究における分析の中核となる計算処理で

ある。対応分析を行った後の特徴語（抽出の手法は次節の事例で説明する。）のみに着目しすぎると、単語と単語の区間が孤立しすぎてしまう傾向があるため、より広い意味区間[9](Semantic space)の単語群を知ることで区間表現の認識に柔軟性を持たせる必要がある。そこで、下記の理由により、潜在意味解析（以下「LSA」と略す）をこの部分の研究に利用した。

LSA は、意味研究における新しい派生分野である。従来の意味論では、単語と単語の意味および単語と類義語、反義語との関係が研究されていた。LSA は単語間の関係を解析するが、これは単語を使用する文脈の条件に基づいており、辞書の定義によるものではない。つまり、ドキュメント中の個々の単語は多次元空間中のひとつの点と見なすことができる。

LSA は知識の獲得と表示のための計算理論と方法であり、テキストに含まれる単語と単語の間に関係があり、そこに潜在的な意味構造が存在することを分析の出発点とする。このような潜在的な意味構造は、テキスト中の単語の文脈における使用パターンに隠れている。そこで、統計計算の手法を用いて、この潜在的な意味構造を見つけるために、大量のテキストを分析する。これは、特定のセマンティックコーディング[10]を必要とする代わりに、文脈中の単語と単語の関係のみに依存するためである。そして、単語間の相関から単語とテキストの意味構造の表現へと変換することで、テキストベクトル[11]を容易に作成することができる。

LSA は通常、単語間の相関性とテキスト間の相関性について研究に用いられる。代表的なケースとしては、ソースデータとしてグロリアのアメリカの学術百科事典を分析した上で大量の単語の類似度を作成し、TOEFL の語彙テストに応用したところ、LSA を用いた計算結果による正解率は 65%であり、非英語圏の受験生の正解率とほとんど変わらないことが分かったというケースがある。[81]

LSA のメリットについては様々であるが、本研究でこの手法を採用した理由は、以下の点である。

まず、LSA は k 次元の語義空間を抽出できることが挙げられる。大部分の情報を保持しながら k<<v(原次元)とすることで、元の空間ベクトルの代わりに低次元の見出し語と文書ベクトルを用いることできるようになり、大量のテキストベースの分析を効率的に扱うことができる。

次に、LSA のコア処理は特異値分解(Singular Value Decomposition、以下は「SVD」と略す)過程であることが挙げられる。「SVD」の利点は、本章の議論の中心となる誘導メカニ

---

9 意味区間単語とは、の持つ概念の表現手法のひとつに，意味空間モデル（semantic space model）がある．このモデルでは，多次元空間上に単語を配置し，任意の単語間の意味的距離を空間上の距離を用いて表現する．

10 セマンティックコーディングとは、それにおいてそれの音またはビジョンと対比されるように何か（言葉、フレーズ、写真、イベント 何 ）の意味がエンコードされるエンコーディングの具体的なタイプである。研究は、私達が、よりよいメモリーを、当社が、それと店に、意味的なエンコーディングを使うのを意味していることで結び付ける物に抱くことを示唆する。

11 テキストベクトルとは、テキストを基づく、単語とドキュメントを作成した意味ベクトル区間。

ズムのパターンを提供できること、異なる次元を区別する代表的な方法であること、ひとりの人間が長年の経験の中で遭遇しうる量やタイプに近いデータにも適用できることの 3 つである。言い換えると、人間の思考をシミュレーションする数学的なプロセスでもある。

加えて、従来の自然言語処理プロセスや人工知能プログラムとは異なり、完全に自動化されていることも LSA の重要な利点である。この自動化とは、LSA が人間の関与を必要とせず、あらかじめ言語学的あるいは知覚的類似性知識(人間が構築した辞書、知識ベース、意味ネットワーク、文法、構文解析器などを使用せず、インプットは未処理テキスト列のみである)を持つ必要がないことである。これは、通常の数学学習で使用されている手法に基づいて適切な低次元に圧縮し、自然言語処理の他の基本的な手法と組み合わせることでオブジェクトやテキストの内容を効果的に提示することを可能にしている。大量のテキストを分析することにより、LSA は分類や予測などの人間の知識獲得能力を自動的にシミュレーションすることができる。観光レビューは分析対象として専門性が高いため、従来のテキストマイニング手法を用いれば専門用語群を構築するには時間がかかるが、LSA の自主的な学習能力を用いることで分析を効率的に行うことができる。

次に特異値の分解方法を下記に示す。

任意の行列$X(m \times n)$（Xのランクを r とする)は 2 つの直交行列と 1 つの対角行列の積に分解できる。

$$X = TSD^T \quad (1)$$

$S_{r\times r} = diag(\sigma_1, \sigma_2, \cdots \sigma_r)$は対角行列であり、その対角元素である$\sigma_1, \sigma_2, \cdots \sigma_r$は特異値分解と言われる。具体的な数値は$XX^H$ か$X^H X$ の固有値の平方根に等しい（実行列$XX^H = XX^T$）。

$T_{m\times r} = (t_1, t_2, \cdots t_r)$ も$D_{m\times r} = (d_1, d_2, \cdots d_r)$ も直行行列である。そのうち、$t_1, t_2, \cdots t_r$ と$d_1, d_2, \cdots d_r$ が左特異ベクトルと右特異ベクトルである。

特異値は行列中の重要な情報に対応することに加え、重要性と特異値の大きさは正の相関を持っている。SVD のメリットは、簡単な方法により、原始行列を次元の大幅縮小した近似行列に投影できることである。LSA は図 3-2 に示すように、SVD のもとで最大の k 個の特異値(k < r)を維持しながら、他の小さい特異値は無視し、k は低次元空間の次元数とする。そして特異値の分解量の逆計算を行い、原始行列の近似行列を得た。 [82][83][84]

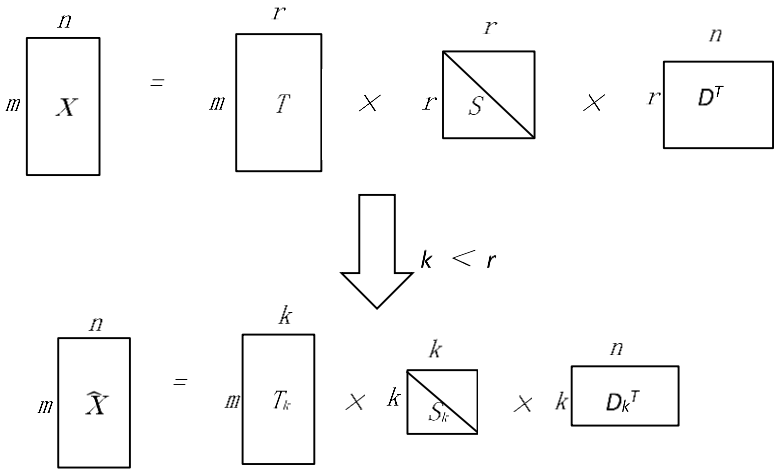


図 3-2　特異値分解と次元削減

次に、頻出語と観光レビューのドキュメントからなる共起行列の特異値分解処理を Matlab を用いて行った。処理の例を図 3-3 に示す。

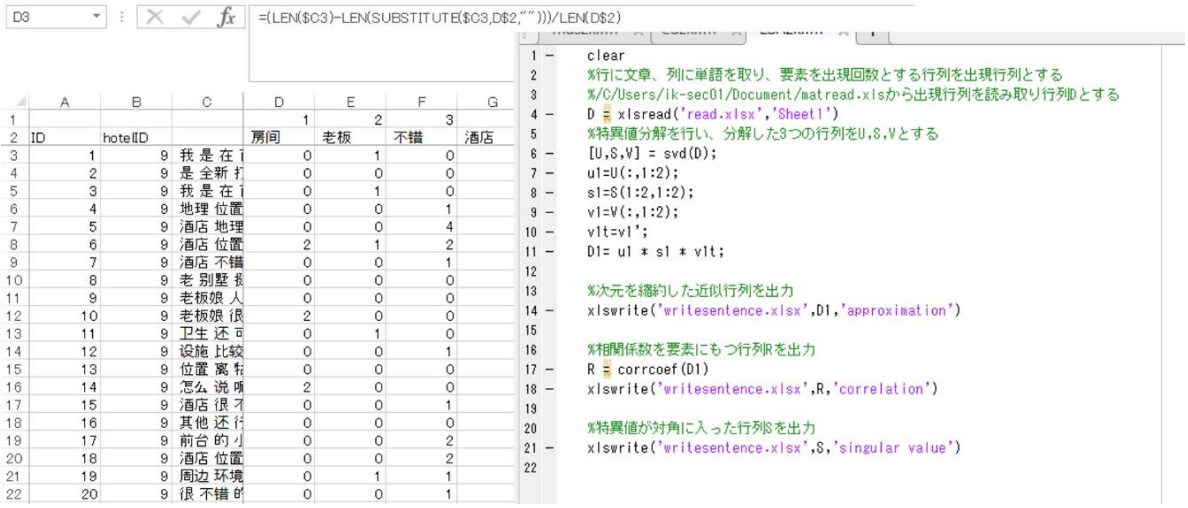


図 3-3　Matlab による共起行列と特異値分解

3.2.3　多次元尺度構成法（Multi Dimensional Scaling）による可視化処理

本章では自然言語処理手法を用いることにより膨大なテキストデータを処理することで有意義な結果得ることに加えて、非言語処理の研究への活用や観光事業の経営者が今後の発展のための手法として活用し、更に一般の観光消費者にも直観的な結果から観光に関する情報を把握することが可能となることを目指した。

このようなことを実現する方法としては、一般的にはMDSが有効な手法であると考えられている。この手法は1950年代に生み出されし、異なる対象間の類似的なデータの分析を通じて、多次元のベクトルからその隠れた内部ルールを発見するのが主な目的であった。数学的な観点からすると、主に1組の高次元のベクトルを1つの低次元空間に射影する過程である。低次元空間における図形を描くことにより、異なる対象の間の類似性を可視化することができ、異なるテーマの中でも対象となるデータを容易に分類できるようになる。

本研究で扱った多次元スケール手法の基礎データは、従来のLSAの結果に由来するものであるため、それ自体が高次元からの帰納結果であり、多次元スケールの低次元空間配置の考え方との整合性を保っている。また、年次によってデータ量が異なり、結果をまとめて比較を容易にするために、LSA の関連度の結果も標準化して二次元の平面に射影する処理を行った。したがって、MDSを採用することは妥当であると考えられる。[85][86]

3.3　中国廬山の宿泊施設の顧客レビューを対象とする分析

3.3.1　廬山の概要及び選定理由

中国の中部に位置する江西省九江市南郊は、ユネスコの世界文化遺産として登録されている。その中でも廬山は中国でも伝統的な避暑地として有名な観光地のひとつである。このような基本的な条件に加えて、観光レビューの分析対象として廬山を選択する理由を以下に挙げる。

1)　観光産業の充実

廬山は中国でも原住民がまだ常住している唯一の山景観光地である。原住民は平野への移住もしているが、2017 年の統計では常住人口は 2 万人以上を維持しており、その大部分が観光関連産業に従事している。そのため、廬山は観光地としての良好な産業基盤を持っている。それに加え、以前から「OTA」企業とも強い連携をしており、「OTA」Web サイトの観光レビューの事例も豊富であること。

2)　データ収集の容易性

廬山における観光業の中でも特に宿泊施設が地域的に集中しており、データ収集が容易である。また廬山山脈全体の範囲は広いものの、中心的な観光資源は山上にある「牯嶺鎮」であり、山上の宿泊施設の多くは「牯嶺鎮」とその周辺にあるため、データの収集が容易で便利かつ正確なうえ、廬山全体をほぼ網羅するデータを集めることが可能であること。

3)　季節による気候の変化

廬山観光の特徴は季節による景観や観光客数の変化が顕著である。廬山は 1 年間の季節による気候変化に伴って風景が大きく異なり、夏期が廬山観光の繁盛期である。このような条件は傾向分析の観点から理想的な条件であること。

図 3-4　廬山の位置

3.3.2　データ収集

収集するデータは、「OTA」業界の中国大陸市場シェア 1 位の企業 Ctrip（携程）の Web サイトに投稿された廬山の宿泊施設に関する観光レビューを収集したもので、対象データは廬山にある全 164 宿泊施設の 2013 年 5 月から 2016 年 6 月までの計 14, 950 件のレビューとした。そして、分析の精度を高めるため、次の 2 つの側面からまずデータのスクリーニングを実施した。

1)　データの信頼性の確保

実際に中国の「OTA」企業の最大手にヒアリングを実施したところにより、宿泊施設に対するレビューのほとんどは真実で信頼性の高い情報であり、偽（やらせ等）のレビューはほぼ存在しない。悪意のあるレビューの発生は、多くの場合、「OTA」Web サイトに新しく登録された宿泊施設に対して発生していることから、虚偽情報の干渉を最小限に止めるため、2014 年の時点でレビューが存在する宿泊施設を対象とし、レビューの選択範囲を 2015~2016 年に限定した。

2)　データの正確性の確保

大量の観光レビューに含まれるテキストを閲覧し分析した結果、テキストに含まれる用語の違い生じる大きな要因のひとつが観光者の消費レベルであることが確認された。その

違いにより生じる誤差を考慮して、宿泊施設の階層[12]により分析対象を検討したところ、最終的には「旅館」を分析対象とすることとした。

最終的に、抽出された対象データを廬山観光における季節変化の要因を考慮し、半年を時間区分の単位とする、A、B、C3 つの区間に分けた。それぞれ対応する期間と観光レビュー数は、2015 年 1~6 月が 275 件のレビュー、2015 年 7~12 月が 531 件のレビュー、2016 年 1~6 月が 154 件のレビューとなり、合計 1060 件のレビューを本手法による分析処理の対象とした。

3.3.3　全体分析

前節 3.2.1 に述べた通り、データ処理プロセスに Rwordseg を形態素解析ツールとして用いることで、独立して存在している観光地特有の固有単語を追加し、意味を含む単語の頻度を統計的に組み合わせることで、共起行列を構成する単語の中から 127 個の頻出語を抽出することができた。これに基づき、前節 3.3.2 で設定した A、B、C3 つの異なる時間区間を考慮し、KH Coder[13]を介して全体的なデータについての対応分析を行い、次の図 3-5 に示すような 2 次元のプロットグラフを得た。このグラフの結果から、それぞれの区間で示された中心には近く、同時に他の区間の中心からは遠い特徴的な単語を明確に読み取ることがでる。そして、最終的に、この中から各区間の上位 10 位の特徴を表す単語を抽出することで、各地区の特徴の解釈を試みる。[87]

これらの抽出された特徴語は以下の通りである（各単語に対応する元となる中国語を括弧に示した）:

・A:「エアコン（空调）」、「観光地スポット（景点）」、「楽しい（愉快）」、「入浴（洗澡）」、「宿泊地（住在）」、「観光地（景区）」、「店主（主人）」、「玲姐（玲姐）」、「お湯（热水）」、「内装（装修）」

・B:「登る（爬）」、「見る（看到）」、「遠い（远）」、「フロント（前台）」、「宿泊施設（住宿）」、「食べる（吃）」、「おいしい（好吃）」、「料理（饭菜）」、「情熱（热情）」、「遊び（游玩）」

・C:「知っている（知道）」、「Ctrip（携程）」、「時間（时间）」、「電気毛布（电热毯）」、「内部（里面）」、「楽しい（愉快）」、「問題（问题）」、「感じ（觉得）」、「車（车）」、「廬山（庐山）」

しかし、これらの単語のみに頼るだけでは、区間の明確な特性と他の区間との変化を十分に把握するための情報としては不足するため、特徴単語の関連単語を更に抽出することで

[12]宿泊施設の階層とは、「OTA」の Website が宿泊施設のレベルによる分類です、例えばホテル、旅館、民宿、ユースホステルなどです。

[13] KH Coder とは、テキスト型（文章型）データを統計的に分析するためのフリーソフトウェアです。アンケートの自由記述・インタビュー記録・新聞記事など、さまざまな社会調査データを分析するために制作しました。

考察を進める。

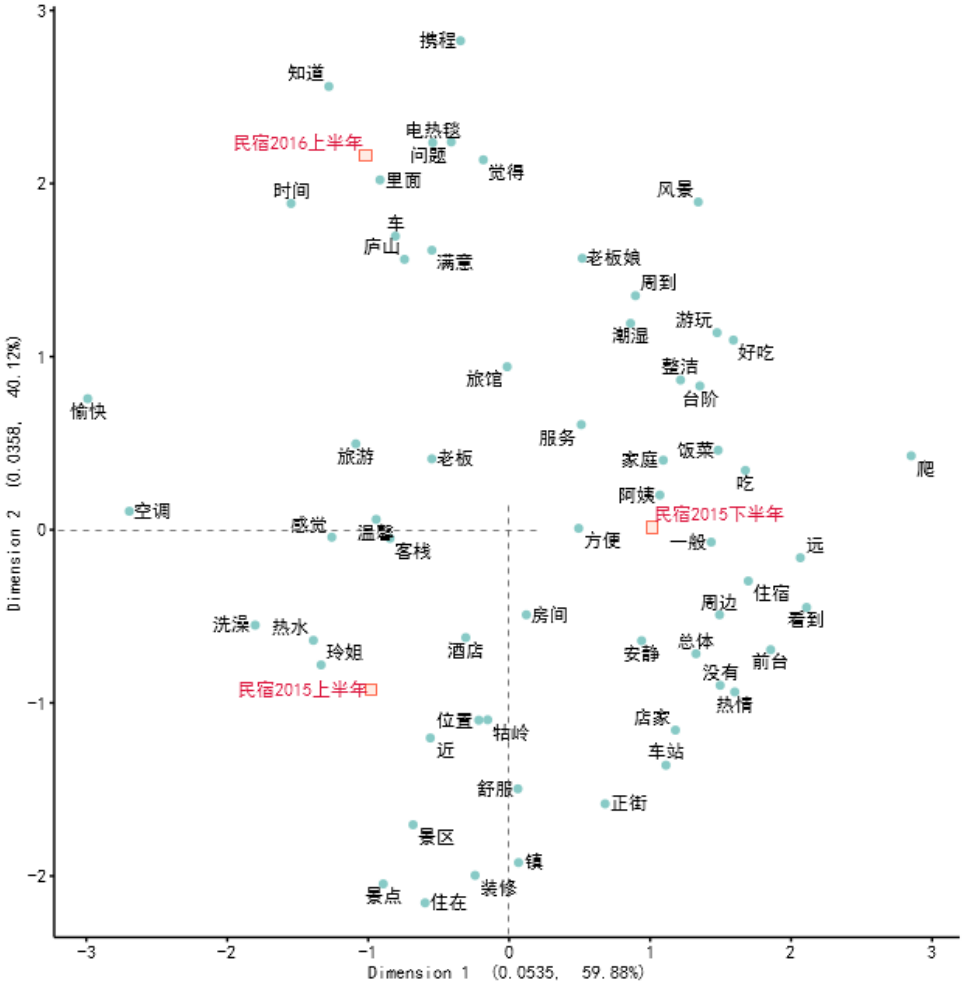


図 3-5　対応分析の結果

具体的には 3 つの区間のそれぞれのレビュー対象データを用いて、統計分析により抽出された 127 の高頻度単語を組み合わせ、それぞれの共起行列を形成し、LSA による分析を行った。前節 3.3.3 で考察した LSA の事例分析の精度を考慮すると、次元の特異値の合計が全特異値の 35%を占めるという実験結果が最も望ましと考えられるため[81]、A、B、C3 区間における高次元から低次元への次元削減にもその数値を適用したところ、最終的な次元数は、それぞれ 15、18、12 となった。その後、LSA の類似性データに基づいて、各単語間のユークリッド距離を計算し、MDS の手法を用いて、3 つの区間の 2 次元平面散布図をプロットした。結果は図 3-6、3-7、3-8 の通りである。

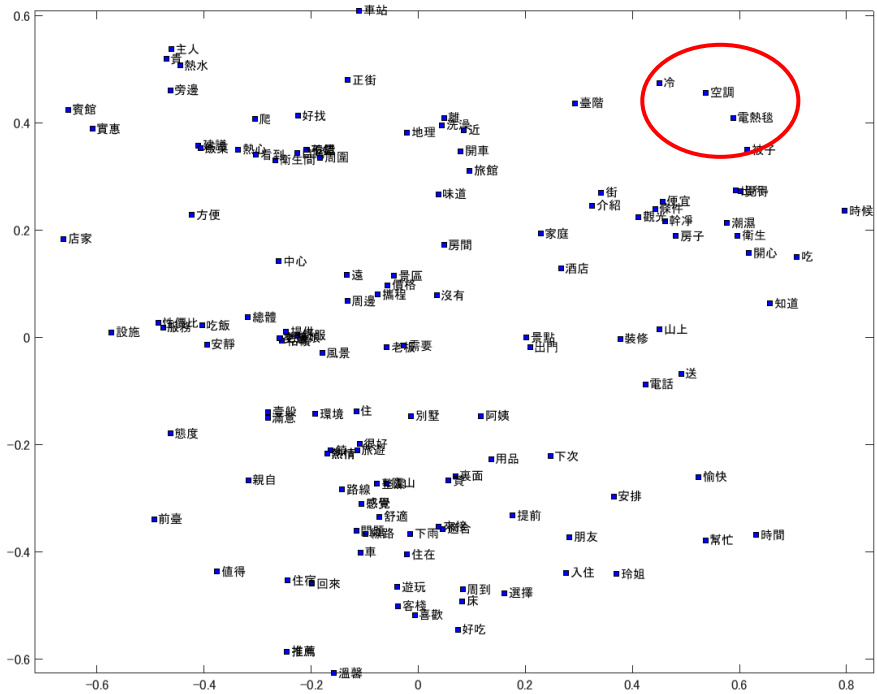


図 3-6　散布図-期間 A

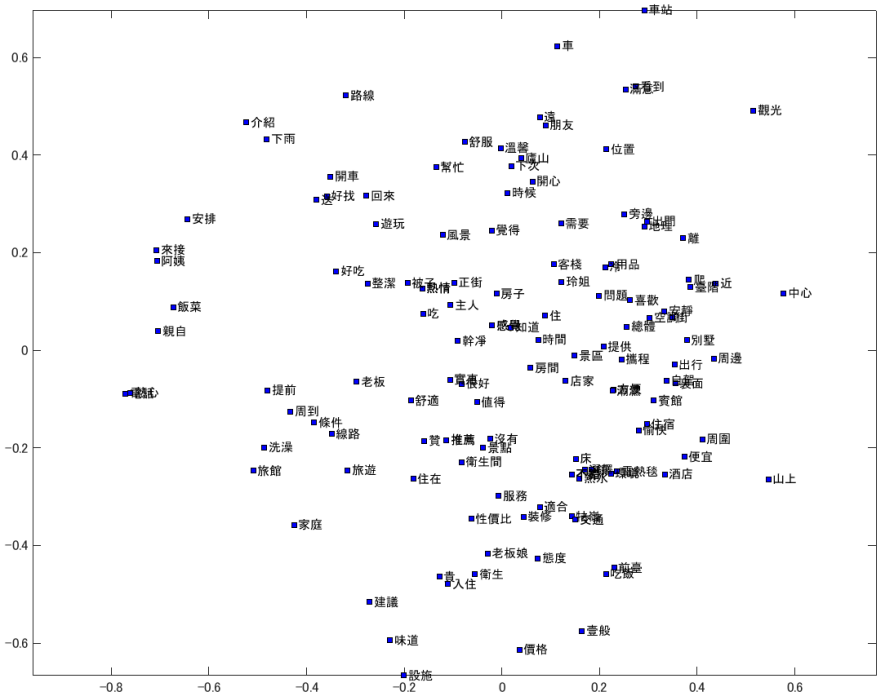


図 3-7　散布図-期間 B

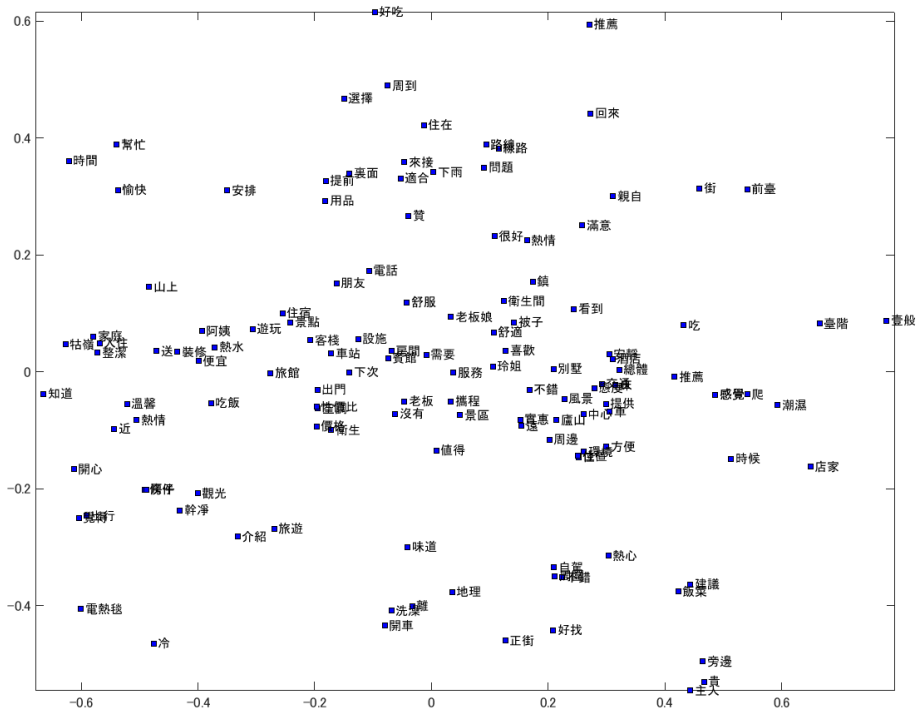


図 3-8　散布図-期間 C

3.3.4　区間傾向の変化に関する検討

図 3-6、3-7、3-8 では各区間に対応する 10 個の特徴語に基づき、LSA による分析に基づき特徴語とそれに強い関連性をもつ単語を統合した。例えば、区間 A の頻出語のひとつ「エアコン（空调）」は、「電気毛布（电热毯）」、「寒い（冷）」、「掛け布団（被子」）の 3 つにつながる。(図 3-6 のスパン A の赤色付け円内に示した)これらの強い関連性をもつ単語は、コアとなる特徴語とともに区間 A の単語特性群を形成している。区間 B と区間 C も同様で、その結果を最終的に整理したものを表 3-1 に示す。

表 3-1　セマンティック特徴語グループ

	1	2	3	4	5	6	7	8	9	10
区間 A	空调	景点	愉快	洗澡	住在	景区	主人	玲姐	热水	装修
2015 1-6 月	电热毯	出门	帮忙	离	下雨	价格	贵	入住	贵	山上
	冷		时间	近		携程	热水	朋友	主人	电话
	被子		安排	开车					旁边	
				地理						
区間 B	爬	看到	远	前台	住宿	吃	好吃	饭菜	热情	游玩
2015 7-12 月	台阶	满意	朋友	吃饭	愉快	热情	整洁	亲自	被子	回来
	近				宾馆	主人		阿姨	吃	
								来接	正街	
									主人	
区間 C	知道	携程	时间	电热毯	里面	愉快	问题	觉得	车	庐山
2016 1-6 月	牯岭	景区	帮忙	冷	提前	帮忙	路线	出行	提供	周边
		服务	愉快	觉得	用品	时间	线路			风景
				出行						中心

以上の結果を踏まえ、本研究では以下の 3 つの側面から傾向の変化について考察を行った。

1)　全体の需要中心の期間変化

A 区間に対する関心の大部分は、「入浴（洗澡）」、「お湯（热水）」、「エアコン（空调）」などの設備に関する単語及び「内装（装修）」などハードウェアに関連があることが分かる。そして B 区間は、観光客の注目点が「フロント（前台）」、「食べる（吃）」、「料理（饭菜）」、「おいしい（好吃）」、「食事（吃饭）」に徐々に変化している。C 区間は、観光客は「ルート（路线）」、「旅行（出行）」、「ヘルプ（帮忙）」、「観光地（景区）」などのツアーの形態やガイドのサービスにかかわる単語が強調されている。

このことから、期間よる観光客の需要の変化は、最初の内部環境から飲食サービスへ、そして観光へのニーズへと変化していることが確認された。観光客は宿泊施設が提供するサービスや宿泊産業務を多元化し、単純な宿泊条件面での要求から全体的なサービスの質を求めるように変化している。したがって、旅館に対して、旅行客の需要に最大限にマッチする観光サービス体系を供給することが、観光客の満足度をさらに向上させるために必要で

ある。

2)　観光形式

各区分の単語特徴グループには、例えばA区間では「ドライブ（开车）」、B区間では「登る（爬）」、C区間では「車（车）」という単語がある。しかし、C区間の「車（车）」という単語の意味は非常に広く、単独では表現の意味を正確に判断できない。そのため、KH Coderのネットワーク分析結果により、語彙共起の観点から、ここでの「車（车）」は実際には観光遊覧用のバスや旅館が所有する送迎車を指すことが多いことが判明した。[87]

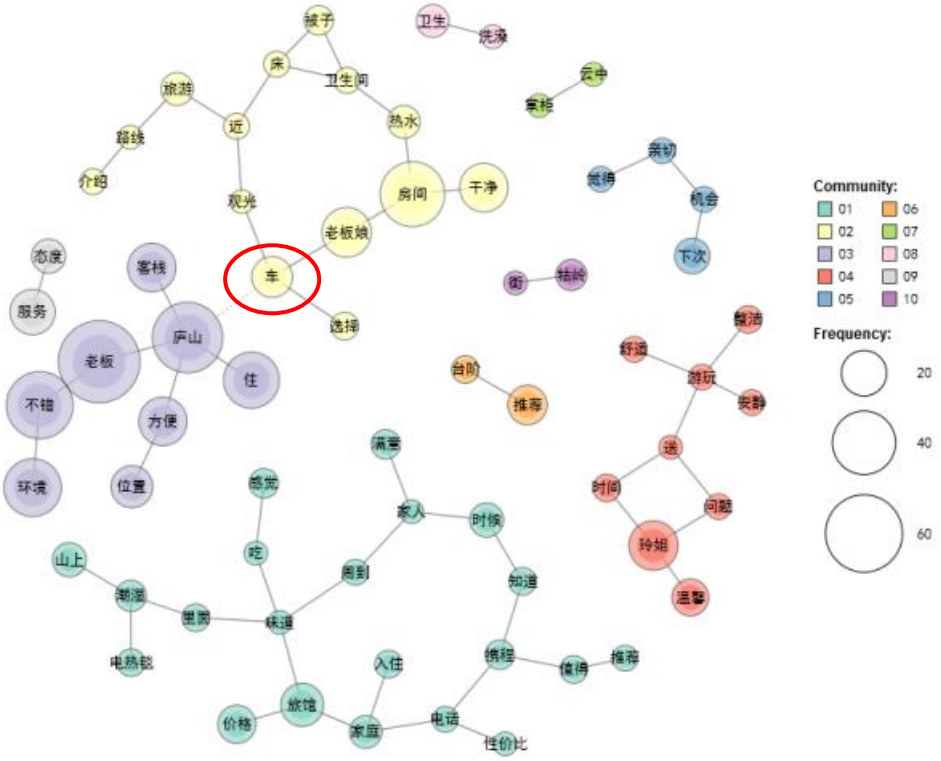


図 3-9　区間Cのネットワークマップ

したがって、B区間がA区間、C区間と異なる点は、徒歩による観光が強調されているという点である。B 区間のレビューが繁忙期の夏季に明らかに集中しているためにこのような移動パターンに適しているのに対し、A区間、C区間は冬季におけるレビューが多く、観光客が徒歩以外の移動手段を利用する必要があるからである。また、移動手段の面でも、観光客は自分が運転することよりも公共交通機関を用いることが多かった。

したがって、これらの情報は、観光事業者に対して観光バスの時刻表を渡すなどの適切なサービスを提供する必要性があることを裏付けている。また、観光客自身が運転しないツアーの割合の増加によって、公共交通サービスの質を向上させる必要があるため、これは山や景勝地の観光の質を向上するための重要な要素である。

3)　閑散期と繁盛期の間隔の違い

廬山の観光における観光業界の戦略や地元自治体の政策改善するためには、時間の経過に伴い変化する、閑散期と繁盛期の差異を考察することが重要である。

まず、A 区間、C 区間については同様な部分が多く、B 区間は A 区間、C 区間と大きく異なる。これは、年次が異なる場合においても、同じ季節の均質性が非常に高いことを示している。

また、ユーザーエクスペリエンスの部分に関して、B 区間が代表する繁盛期の重点は飲食店にあり、特に「おばさん（阿姨）」、「店主（主人）」、「自ら（亲自）」などのサービス関係の語彙が強調されている。更に、A 区間、C 区間が代表する閑散期の多くは直接的な評価であり、例えば「楽しい（愉快）」、「高い（贵）」、「寒い（冷）」、「ヘルプ（帮忙）」などである。そのため、繁盛期における飲食サービスやスタッフの態度は観光消費者のレビューに大きな影響を及ぼし、閑散期における観光消費者は、観光体験を全体的に評価する傾向にあることが分かる。

このような結果から、観光行政部門は観光事業者に季節ごとに異なる経営資源の配分をとるようアドバイスすべきであると考えられ、例えば繁盛期では特にレストランの管理に注目し、閑散期では観光客の個人的なニーズに注目する必要がある。

第 4 章　中国大陸からのインバウンド旅行の満足度を高めるための予測分析

第 3 章における中国大陸からの観光客のレビューに関する考察と分析処理を通じて、中国語による観光レビューのデータ分析には従来の自然言語処理手法ではなく、特殊な処理を加えることの必要性を確認した。したがって、インバウンド観光のオンラインレビューを分析する際には、国内観光と国際観光との差異を十分に考慮しなければならない。そのため、本章では観光レビューのデータ分析には従来のテキストマイニング手法では不適切と考え、中国大陸からインバウンド観光のレビューデータの分析に適した新たな分析手法を提案し、中国大陸観光客のレビューデータを地元訪問客のレビューデータと統合して不完全行列を構築した。そして、行列の分解手法を用いて分析・予測を行い、最後に「箱根の宿泊施設レビューをサンプルとした中国大陸からの訪日インバウンド観光客の未取得情報」を事例として予測・分析を行い、結論を導いた。

4.1　中国大陸の国内観光と海外観光の差別によって、アウトバウンド観光について研究の意義

第 3 章では、中国大陸観光客の「OTA」レビューのテキストマイニング分析を通じて中国大陸観光客の特性をまとめた、その上で自然言語処理方法を観光レビュー分析の対象となる文章にどのように適用できるのか考察した。ただし、第 3 章における分析は、国内観光と海外観光との差異を考慮せずにまずは行ったため、本章ではその差異を考慮した上で更に分析を行い、その結果を考察する。

まず、中国大陸観光客のアウトバウンド観光が発展した背景についての考察を行う。中国は、世界で最も海外への観光客が多い国に成長し、そのうち近年で急伸の傾向が認識されるのは訪日観光である。日本の観光庁の統計によると、2018 年の訪日観光客の総数は 3100 万人を超えており、その中でも特に注目すべきなのは、全体の 4 分の 1 以上を中国大陸からの観光客が占めているということである。図 4–1 に示した通り、近年、中国大陸からの観光客が増加しなかった年は東日本大震災の影響を受けた 2011 年と、中国大陸内の政治的反日感情が高まった 2013 年だけである。それ以外の年は中国大陸からの観光客は増加しており、特に 2015 年には 2 倍以上の成長が見られた。その後、2015 年から現在に至るまで、中国大陸からの観光客が訪日者数で最も多い国・地域となっている。ここ数年、中国大陸観光客のインバウンド観光の伸び率は中国国内観光の伸び率をはるかに上回るといえる。[6][88]

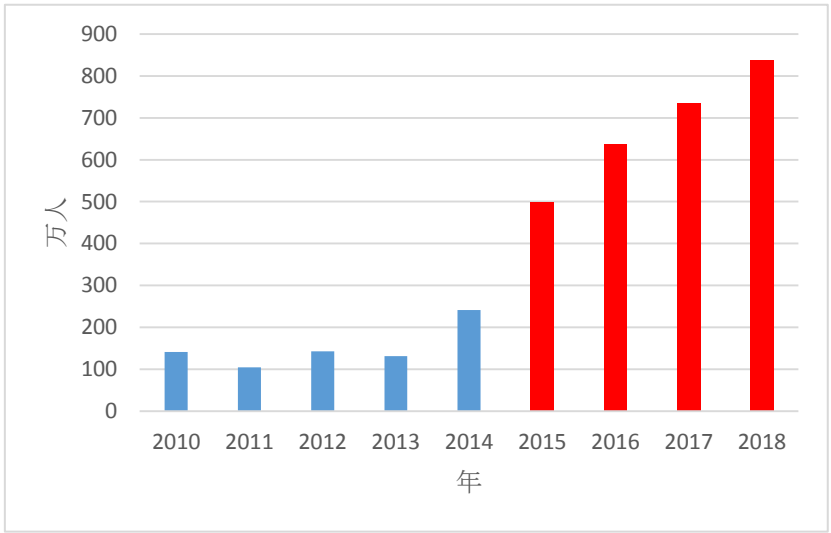


図 4-1　訪日インバウンド中国大陸観光客の数

次に、中国大陸観光客のイデオロギー的な側面を考慮すると、国際観光は国内観光と異な

り、観光商品・サービスなどそのものの要因に加えて、国の観光産業に関連する政策や政治に関する要因も産業の発展に影響を及ぼす可能性が高い。これは、第 2 章の先行研究で述べたように、多くの論文でも議論されている。その政策や政治の側面が与える影響は著しいが、この部分に関する観光関連業者の対応は明らかに受動的にならざるを得ない。また本研究の対象は観光産業であるが、本研究では政策や政治に関する要因は考慮しない。しかし、観光事業者にとって常にフォローしなければならない事項である。実際に観光商品・サービスの企画・開発時には、国家間の相互理解の促進や、文化交流の習慣が異なることによる誤解の排除、地域間の平和と安定と発展の促進等に配慮する必要がある。

そして最後に、本研究で観光客のレビュー情報を中心にデータを分析を行い、観光客の観光需要と関連する情報の把握することを本章の目的とする重要性について述べる。

観光客の基本的な状況を理解するためには、まず、観光業の観光形態の考察が必要である。そこで、中国大陸からの訪日インバウンド観光客を例に、観光客の観光形態の割合について分析と考察を行った。

1) 日本の観光庁の統計データによると、訪日する中国大陸からの観光客のうち初回の訪問者数の割合は 2010 年の約 80%から 2017 年には約 60%に減少した。
2) 移動の形式の割合は、グループによるツアー観光が 2010 年の 80%から 2017 年には 40%未満に減少した。

以上の結果から、

1) 中国大陸からの観光客初回の訪問者の半分以上を占める、初回の訪問者は引き続き重要な構成要素である。これらの観光客にとって、日本も含めた海外地域の情報の多くは中国を出発するまでに国内で取得したものであるため、得られた情報は常に最新とは限らず、また、訪日後も日本で直接入手できない情報もある。
2) 個人旅行を選択する人が増えていることは、中国大陸観光客がグループによるツアー観光の制約から抜け出し、日本も含めた海外でのインデプストラベル[14]（in-depth travel）を求めていることを示している。

[14] インデェプストラベルとは、より深い経験・体験ができるような個別旅行。

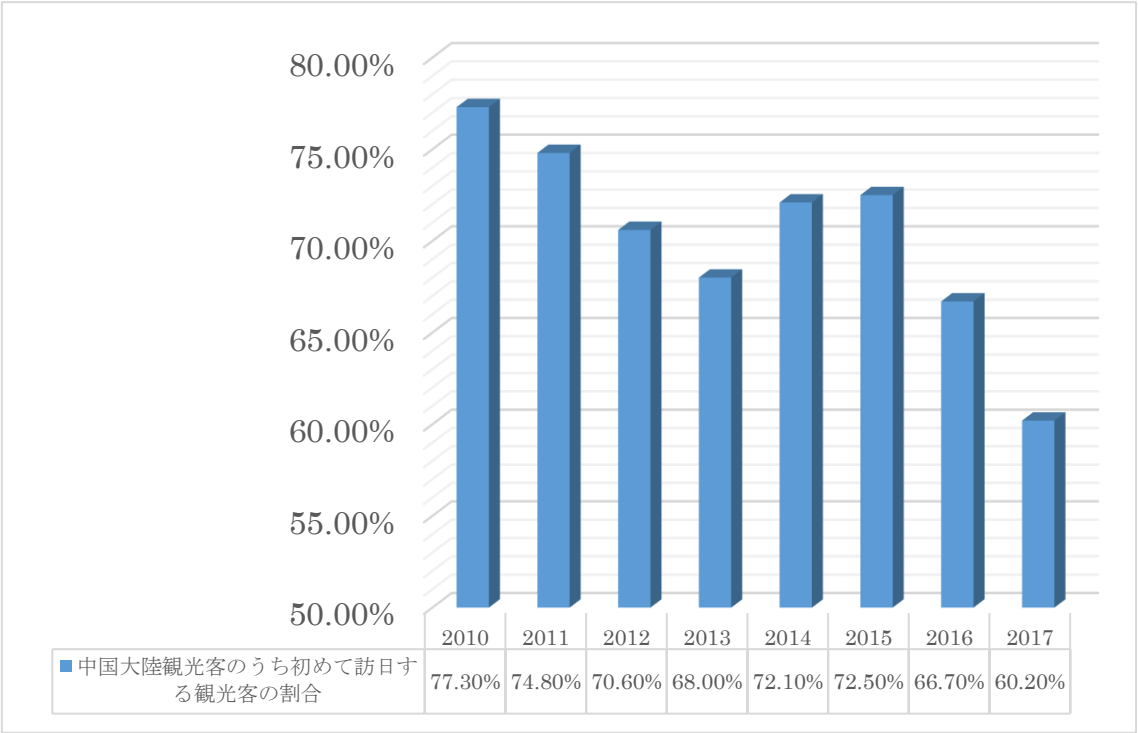


図 4-2　中国大陸観光客のうち初めて訪日する観光客の割合

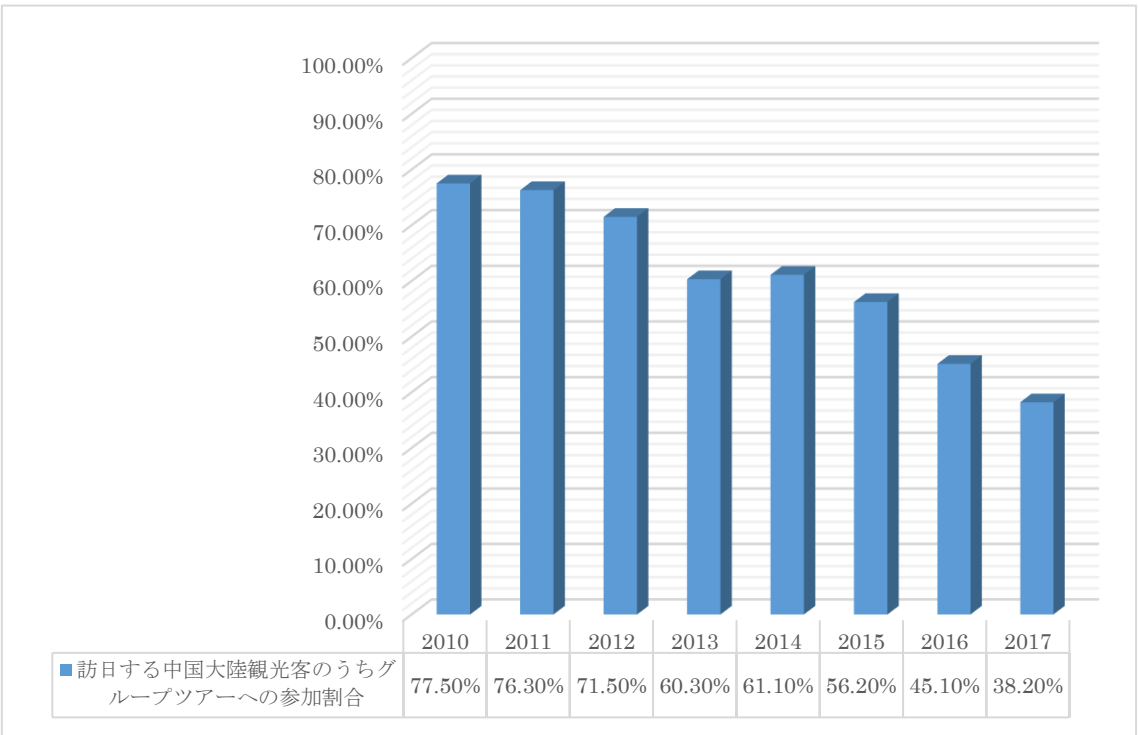


図 4-3　訪日する中国大陸観光客のうちグループツアーへの参加割合

以上のことから、中国大陸観光客は中国国内の旅行についての情報を十分に把握することが可能ではあるが、アウトバウンド観光の情報に関して把握することは不十分であるため、これらの未知の部分があることもアウトバウンド観光の魅力の一部として捉える必要がある。

以上より、未知の情報から中国大陸観光客に適する部分を把握し、そこにより多くの観光資源を投入することで関連する観光商品・サービスの品質を向上させ、中国大陸観光客にこれらの情報をわかりやすく把握してもらうと同時に、観光客の満足度を向上させ、最終的に観光客からのフィードバックにより観光産業の発展を推進する仕組みが必要であることを明らかになった。[6][66][70]

4.2　自然言語処理と不完全行列を用いた予測方法についての考察

4.2.1　対照とする基準の組み合わせの選択

前節 3.2.1 で述べたように、共起行列は本研究におけるテキストデータの分析の基礎である。そこで、前節 4.1 で行った中国大陸からのアウトバウンド観光者のレビュー情報の分析に基づき、次の表 4-1 に示すように、海外旅行における中国大陸観光客の観光レビューのみを含むテキストデータを、既知の情報区域と未知の情報区域の 2 つに分けて共起行列を構成した。

表 4-1　2 つの情報区域をもつ共起行列

		既知の情報				未知の情報
		W1	W2	W3	W4…… ……	
中国大陸観光客のレビュー	C1	3	0	0	1 …… ……	? ? ? ……
	C2	0	0	0	2 …… ……	? ? ? ……
	C3	0	2	0	1 …… ……	? ? ? ……
	C4	0	1	1	0 …… ……	? ? ? ……
	……	……	……		…… ……	…… ……
	……	……	……		…… ……	…… ……

このような未知の情報は、中国大陸観光客の観光レビューの分析だけでは直接入手できないため、未知の情報の探索と分析を行うためには対照の組み合わせが必要である。その組み合わせには本研究の未知情報を把握・予測するという目的と処理方法を考慮すると、以下の条件を満たす必要がある。

1) 対照の組み合わせは、同じ期間、同じ観光エリア、そして同じタイプの観光商品・サービスに関するレビューであること。(観光レビューのサンプル収集の範囲を限定するため)

2) 指定した観光エリアの観光商品・サービスを十分に体験するために、その観光地を初め

て訪れる中国大陸観光客にとって未知の情報となる部分が既知の情報をうまく補完することができる組み合わせであること。(観光レビューに関する情報の包括性を確保するため)

3) 地元の文化的特徴を十分に活かしながら、中国大陸観光客の文化交流に関するニーズを満たすコンテンツを示唆できる組み合わせであること。(文化交流を観光産業の発展に反映させるため)

4) 予測対象と類似な観光嗜好を示す組み合わせであること。(予測の正確さを高めるため)

5) 組み合わせに関するデータサンプルの取得が簡単かつ量も十分であること。(検証レベルを確保するため)

そこで、分析対象となる観光レビューが入手可能な観光客のグループを、以下の 3 種類のグループに分類して考察を進める。

グループ A：中国系の熟練観光客
グループ B：中国系以外の外国人観光客
グループ C：地元の観光客

表 4-2　3 つの異なる分類グループの検討

	1: 同じ観光商品の体験あり	2: 観光商品の十分な経験あり	3: 地元の文化的特徴を理解済	4: 類似の観光嗜好を有する	5: サンプルデータ取得の容易さ
A: 中国系の熟練観光客	適正	良い	悪い	最も良い	不適正
B: 中国系以外の外国人観光客	適正	悪い	悪い	悪い	適正
C: 地元の観光客	適正	最も良い	最も良い	普通	適正

グループ A はグループ B と C に比べてより条件を満たしている面はあるものの、実際に中国系の熟練観光客を「OTA」レビューの中の情報や投稿者の ID からからスクリーニング

して識別することが難しく、分析・予測に必要な十分なサンプル量が得られないことから排除される。

そして、グループ C は日本在住の地元観光客であるため、条件 4 の観点からは中国大陸観光客の観光習慣との類似性は高くはないが、国内観光であったとしても訪問先の食文化や生活風習などの文化習慣への同調意識を持っているという観点からすると条件 3 を十分に満たしており、条件 4 に欠ける部分をある程度は補完できると考えられるため、最終的にグループ C が未知情報を予測するための分析対象とした。

4.2.2 宿泊施設の「OTA」レビューに関するデータ収集

中国大陸観光客の分析において、宿泊施設等に関する「OTA」Web サイトに書き込まれる観光レビューは、下記の理由により、データ分析をする上で最も理想的な有効かつ効率的に観光客からのフィードバックを得ることができるプラットフォームである。

1) レビューに関する情報には特定の訪問場所と投稿時間に関する情報が含まれているため、詳細な追跡が可能である。したがって、様々な条件を変えながらレビューを簡単に検索および抽出することができる。
2) 宿泊施設に関するレビューは実際に宿泊施設を利用した観光客のみが投稿することができ、「OTA」企業側では削除することができない。そのため、それらのレビューは観光客が実際に経験した事実であり信頼性が高い。
3) インタビューなど回答者と質問者が相互に影響を及ぼす環境で調査する方法に比べて、「OTA」レビューの投稿過程には何ら外部からの影響もなく、観光客単独の自主的なものである。
4) 投稿されるレビューの長さは限られており、観光客が最も気になる部分のみに焦点を当てることができる。また、個々のレビューがひとつのテキストデータとなっており、共起行列の計算に直接利用することができる。
5) サンプル量が十分で、かつ前節 2.2.2 で述べた通り観光レビューシステムが成熟しており、「OTA」業界は急速に発展している。その中でも宿泊施設に関するレビューのシステムは観光レビューの中でも初めて導入されて仕組みだと言われ、今では広範に使われている。また、このような状況は中国語のレビューであってもその他言語のレビューであっても同様である。したがって、この宿泊施設に関する観光レビューのデータに基づく解析手法は広く世の中に受け入れられている。

4.2.3　不完全な共起行列の構築

4.2.3.1　前処理

第 3 章の研究と同様に、共起行列の構築には分析対象データの適切な前処理が必要であり、無意味なレビュー、重複したレビュー、短すぎるレビューなどを排除し、中国語の繁体字と簡体字の統一も処理を行った。もちろん、核となるのは形態素解析であり、中国語の部分の形態素解析は変わらないが、中国大陸観光客の訪問国によっては、その国の「対照組み合わせ」言語の形態素解析が必要となるケースがある。例えば、訪日観光産業の事例を分析した結果、日本の観光地を訪れる日本人は日本語でレビューを書くことがほとんどであったため、中国語のデータには Rwordseg、日本語のデータには Mecab を、前処理に使用する。以下、前処理に用いる 2 つのソフトウェアの特徴と有効性を示す。

中国語を対象とするソフトウェア Rwordseg のメリットとしては、カスタム辞書を追加できる機能があることである。中国大陸観光客のアウトバウンド観光に関わる地名や海外専門語には、例えば「芦ノ湖（芦之湖)」や「食べ放題（放题)」といった単語が多いが、従来からある Rwordseg 以外の形態素解析ソフトウェアの辞典にはこのような単語が含まれていない。したがって、これらの一般的ではないアウトバウンド観光に関わる専門用語をカスタム辞書に追加する必要がある場合などは、Rwordseg のカスタム辞書機能が有効である。[80]

日本語の Mecab は形態素解析を行うと同時に、単語を自動的に品詞分解することができる機能（品詞タグ付け機能）を有している。そのため、後続の分析作業では主に名詞を集中的に分析できるため、研究の効率化を図ることが可能となる。[89]

4.2.3.2　単語とドキュメント

前節 4.2.1 の考察から、中国大陸からのインバウンド観光客のレビューに比べて、地元観光客のレビューはより広範囲の観光要素に関わる項目をカバーしている情報源であると考えられる。中国語で書かれたレビューは分析の主対象であるが、実際に観光レビューのドキュメントは、中国大陸観光客のレビューと地元観光客のレビューで構成されている。単語の選択には、「対照組み合わせ」のうちでも、訪問国・地域の地元観光者のレビューで頻出する名詞を使用する。

第 3 章の単一言語における観光レビュー分析で使用した共起行列とは異なり、この研究では 2 つの異なる言語を分析対象としているため（第 3 章の処理の文章が全部中国語であったが、本章では「予測組み合わせ」と「対照組み合わせ」で中国語と外国語としての日本語の 2 つ言語の文章がある)、中国語と日本語の同じ意味の単語を統合的に認識する必要が

あるため、翻訳辞書を作成する。その際、中国語と日本語の 1 対 1 の対照は完全にはできないため、同義語をグループとして組み合わせて分類することを本研究での処理方法とする。最後に、分析対象となる観光レビューの文章は「OTA」業界に関連していることを考慮すると、観光産業における業界用語などの慣用句とネットワークで使用されるスラング（隠語、略語、俗語など）を標準辞書には無かった頻出語をカスタム辞書に追加するなど、可能な限り参照できる状況にする必要がある。

4.2.3.3　不完全行列

次に不完全行列を生成するが、まず、「対照組み合わせ」の地元観光客のレビュー（非中国語）に基づいて、中国語のレビューの高頻度単語を W とする。一般的には、中国大陸観光客が気にしない内容として W'とみなされることが多い。しかし、実際にこの理由を除き、中国大陸からのインバウンド観光客にとって、文化の差異による不明な情報が多く、通常の訪問先での観光レビューの文章よりも単語の出現頻度が低くなる可能性が高い。行列の観点からみると、スコアなしまたは低スコアを正確に区別することはできない。そこで、この部分を統一してコア分析対象として再設定し、W'の単語は中国語のレビューの共起区域と記号「－」とする。以上のプロセスを経て構成された「対照組み合わせ」のび不完全行列を下記の表 4-3 に示す。

表 4-3　不完全行列

		日本語の頻度名詞	
		中国語の高頻度名詞単語	中国語の非高頻度名詞単語
		W1 W2 W3 W4 …… …… …… ……	W'1 W'2 W'3 …… ……
予測組み合わせ（中国語のレビュー）	CN-C1	0 2 0 1…… …… …… ……	－ － － …
	CN-C2	1 0 0 1…… …… …… ……	－ － － …
	CN-C3	2 0 0 0…… …… …… ……	－ － － …
	CN-C4	0 2 3 0…… …… …… ……	－ － － …
	……	…… …… …… …… …… ……	－ － － …
	……	…… …… …… …… …… ……	… … … … …
対照組み合わせ（日本語のレビュー）	JP-C1	0 0 0 1…… …… …… ……	0 0 1 ……
	JP-C2	1 2 0 0…… …… …… ……	1 2 0 ……
	JP-C3	0 1 0 1…… …… …… ……	0 3 0 ……
	JP-C4	2 1 0 1…… …… …… ……	2 0 1 ……
	……	…… …… …… …… …… ……	…… …… ……
	……	…… …… …… …… …… ……	…… …… ……
	……	…… …… …… …… …… ……	…… …… ……

次に、このような不完全行列を具体的に処理し、計算する方法について検討を行う。

4.2.4　行列の分解手法

本研究では、「OTA」レビューを分析し、大量のテキストデータを自動的に結合する方法として、テキストマイニングのための自然言語処理を使用する。通常、自然言語処理では、まずテキストデータについて完全な共起行列、つまり観光レビュー（ドキュメント）における頻出語の出現回数に基づいた行列を構築する必要がある。しかし、本研究の目的は、中国大陸観光客のレビューには明示的に表記されていない潜在的な観光需要情報を予測することであるが、中国大陸観光客が具体的に意識しない場合、それが観光レビューとしてテキストで表現される頻度は高くなく、文化的背景の違いにより情報が埋もれてしまう可能性が高い。また、観光レビューに関わる内容全体を含む共起行列が表す情報をスコアリングシステムで評価した場合、「低スコア」と「スコアなし」の単語を効果的に区別することができない。

そこで、本研究ではこの部分を予測し、判断することを目的として、中国大陸観光客のレビューの共起行列を不完全な行列として生成し、行列分解はこの不完全な行列の中で欠落した値を予測する効果的な処理方法を提案することとした。以下、その具体的な方法について考察を進める。

4.2.4.1　行列分解と潜在因子モデル

通常の行列分解では、固有値分解(Eigenvalue Decomposition、EVD)と特異値分解(Singular Value Decomposition、SVD)を行うことが一般的である。特に後者はテキストマイニングの関連処理として適用されることが多く、第 3 章における分析の主要な方法としても採用した。しかし、この方法には下記のような限界がある。

1)　行列が不完全である場合には、従来の特異値分解は不確定である。SVD は分解されるべき行列に空白があると処理できない。
2)　一部の既知項のみを扱う場合には、過剰適合が起こりやすい。

そこで、本論文では、潜在因子モデルの手法を用いて不完全行列の行列補完を実現するために、この手法を応用した推薦システムを採用することとした。

アルゴリズム的には、user の関心と item の特徴を潜在因子(latent factor)によって関連づけて解釈する。行列因子分解モデルは、user と item の両方を次元 k の共同潜在因子空間にマッピングし、その空間における user−item の相互作用を内積としてモデル化する。

たとえば図 4-4 のように、元のスコアリング行列 $R(3\times4)$に対して、3 つの潜在的特徴が

あると仮定し、行列 R(3×4)をユーザー特徴行列 P(3×3)と物特徴行列 Q(3×4)に分解する。user1 の item1 に対する評価点を考察すると、user1 が 3 種類の潜在的特徴 class1、class2、class3 に対する関心度はそれぞれ $P11$、$P12$、$P13$ であり、これら 3 種類の潜在的特徴と item1 との関連度はそれぞれ $Q11$、$Q21$、$Q31$ である。

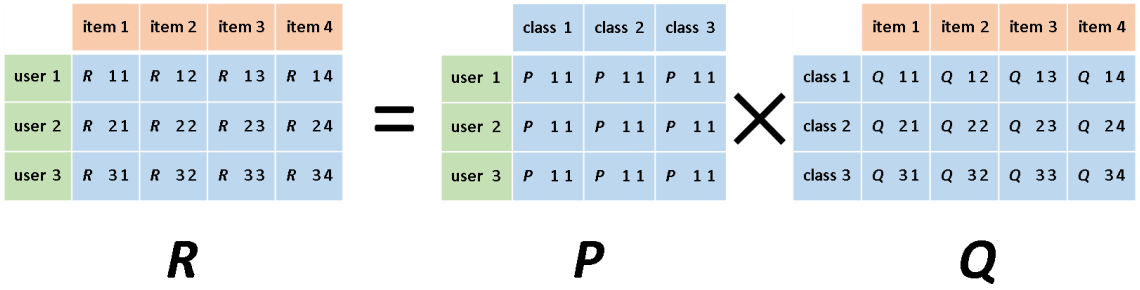


図 4-4　潜在因子モデルの例

4.2.4.2　行列分解と潜在因子モデル手法の本研究への採用に関する考察・まとめ

本研究で用いる行列分解と潜在因子モデル手法の特徴を下記に示す。

全てのアイテムiと潜在的な特性ベクトル$Q_i \in R^k$に関連付けられる。

全てのユーザーuと潜在的な特性ベクトル$P_u \in R^k$に関連付けられる。

与えられるアイテムiに対して、これらの潜在的な特徴はそのアイテムがその属性におけるポジティブまたはネガティブの程度を評価する。

与えられるユーザーuに対して、これらの潜在的な特徴は、ユーザーが対応する要素にどの程度興味を持つか、それがポジティブなのかネガティブなにかを評価する。

ユーザーuのアイテムiに対する最終的な関心度の評価スコアは、それぞれの潜在的な特徴の次元におけるuがiに対する関心度の和であり、ここでuがiに対する関心度は、現在の潜在的な特徴に対するuの関心度にiと現在の潜在的な特徴に関する関心度を掛けて表される。特性ベクトルの内積行列は$\sum_{k=1}^{K} P_{u,k} Q_{k,i}$で表すことができる。

そのうち$\hat{r}_{ui} = Q_i^T P_u$はユーザーuがアイテムiに対する評価スコアに近似する。

したがって、それぞれの潜在的な特徴ベクトルであるQ_iとP_uを計算すると、潜在的な未知情報を簡単に評価および予測でき、潜在的な特徴ベクトルの内積によって直接利用できる。

しかし、より適切な潜在的な特徴ベクトルを取得するために、実際の行列と潜在的な特徴内積行列の違いを最小限に抑える方法が必要である。

そこで、以下の公式を用いる。

$$\min_{Q^*,P^*} \sum_{(u,i)\in G} (r_{ui} - Q_i^T P_u)^2 + \lambda(||Q_i||^2 + ||P_u||^2) \quad (2)$$

そのうち、Gは行列Rにある既知アイテムの（u,i）組合せである。

この方法に用いる数式は 2 の部分に分かれており、前半部分$(r_{ui} - Q_i^T P_u)^2$の主要な役割は、潜在的な情報における予測評価スコアを既存の評価スコアとできるだけ一致させる数式である。これは、予測値の正確さをより向上させるものであり、この部分が小さければ小さいほど、潜在的な特徴を表す内積結果が実際のスコアサンプルに近くなる。

そして、もうひとつの後半部分は、学習パラメータQ_iとP_uの正規化を行うことで、観測データ過剰適合を回避する数式である。定数 lamda は、正規化の程度を制御し、潜在的な特徴手法への改善であると考えても良い。最後に、予測値の結果を利用して、この不完全な行列をマトリックス補完（matrix completion）することができる。

以上のように、行列分解と潜在因子モデル手法を用いるのは、主に以下の 2 つの事由に基づく。

1) 個々の観光レビューは、潜在因子で非常に良好な反応を示しているが、本研究で予測の対象とする中国大陸観光客のレビューは、予測部分を行列によって計算することができる。但し、その際には適切な対照項目を選択する必要がある。
2) 現在、上記の式を参照して、次節における実際のデータ計算、特に勾配降下(gradient descent)手法の使用、および元のデータと比較する損失額(loss value)の監視が可能になり、結果の精度と計算上のエラーが出た場合の再計算を行う部分へのアクセシビリティの高さを向上させることができる。[90][91][92]

4.2.5　予測計算の実施

本節では、行列分解と潜在因子モデルの手法を適用した予測計算を具体的に実施するために、Python ソフトウェアを用いて関連するプログラムを作成した。

図 4-5 の Python プログラムによると、numpy と pandas の 2 のパッケージが最初に呼び出された。そして、前節 4.2.4.2 で示した式(2)の誤差を最小限に抑えるために、確率的勾配降下法(stochastic gradient descent)の計算方法を用いた。

```
In [ ]: import numpy as np
        import pandas as pd
```

```
In [ ]: def matrixFactorization(R, K, steps=10, gamma=0.001,lamda=0.02):

            N=len(R.index)
            M=len(R.columns)
            P=pd.DataFrame(np.random.rand(N,K),index=R.index)
            Q=pd.DataFrame(np.random.rand(M,K),index=R.columns)

            for step in range(steps):
                for i in R.index:
                    for j in R.columns:
                        if R.loc[i,j]>=0:
                            eij=R.loc[i,j]-np.dot(P.loc[i],Q.loc[j])
                            P.loc[i]=P.loc[i]+gamma*(eij*Q.loc[j]-lamda*P.loc[i])
                            Q.loc[j]=Q.loc[j]+gamma*(eij*P.loc[i]-lamda*Q.loc[j])
                e=0
                for i in R.index:
                    for j in R.columns:
                        if R.loc[i,j]>=0:

                            e= e + pow(R.loc[i,j]-np.dot(P.loc[i],Q.loc[j]),2)+lamda*(pow(np.linalg.norm(P.loc[i]),2)+pow(np.linalg.norm(Q.loc[j]),2))
                if e<0.001:
                    break
                print(step)
                print(e)
            return P,Q
```

```
In [ ]: dataFiletest='/Users/Zhang/2019/a9/new_a9_b10.xlsx'
        xls = pd.ExcelFile(dataFiletest)
        new_a9_b10 = xls.parse('Sheet1')

        (P,Q)=matrixFactorization(new_a9_b10.iloc[:90,:109],K=2,gamma=0.005,lamda=0.005, steps=1000)

        P.to_csv("P9-10-k2-st1000.csv",index=False,sep=',')
        Q.to_csv("Q9-10-k2-st1000.csv",index=False,sep=',')
```

図 4-5　Python のプログラム

以下が、本計算で用いた勾配降下法の概要である。

勾配ベクトルは幾何的な意味で関数変化が最も速く増加する場所である。具体的には、関数f(x,y)について、点(x_0, y_0) における勾配ベクトルの方向が$(\frac{\partial f}{\partial x_0}, \frac{\partial f}{\partial y_0})^T$ の方向は、f(x,y)の増加が最も大きい所である。言い換えれば、勾配ベクトルの方向に沿って、関数の最大値を見つけることができる。逆に、勾配ベクトルの逆方向、すなわち$-(\frac{\partial f}{\partial x_0}, \frac{\partial f}{\partial y_0})^T$ の方向に沿って勾配が最も早く減少する、つまり関数の最小値も簡単に見つけることができる。

機械学習アルゴリズムでは、損失関数（Loss function）を最小化する際に、勾配降下法により反復的に解を求めることで、最小化された損失関数を得ることができる。

その際、潜在因子の考え方に基づき、関数を構築するには以下のパラメータが必要となる：

R:　関連する不完全行列の処理

K:　latent factor の次元

Steps:　計算ステップの数

Gamma:　勾配変動係数

Lamda:　正規化係数

プログラム関数 matrixFactorization の第 1 部分では、不完全行列の行と列の数に基づいて、初期レビューの潜在的特徴 P と単語の潜在的特徴 Q をランダムに生成する。

プログラム関数 matrixFactorization の第 2 部分では、勾配降下法を使用し、勾配の方向に応じて継続的に減少させる損失関数を“e”と記す。また特定のレビューj にある特定の単語 i の損失値(loss value)は、プログラムでは以下の図 4-6 ように記述する。“e”は、あらゆる既知項の損失値の和である。

```
pow(R.loc[i,j]-np.dot(P.loc[i],Q.loc[j]),2)+lamda*(pow(np.linalg.norm(P.loc[i]),2)+pow(np.linalg.norm(Q.loc[j]),2))
```

図 4-6　Python のプログラムの損失関数

そして最終的に適切な P と Q が特定され、各ステップが継続的に計算されると、全部の損失値の和“e”の値が表示される。[93]

4.3　箱根の宿泊施設レビューをサンプルとした中国大陸からの訪日インバウンド観光客の未取得情報の予測分析

前節 4.2 で述べた予測計算方法に基づき、本節では日本の代表的な観光地である箱根を分析事例として取り上げ、予測計算に基づく考察を行った。箱根のような日本の景勝地を選ぶ主な理由は、英語の観光レビューしか投稿されないような観光地のデータからは投稿者の身元と国籍を遡ることが難しいからである。つまり、これらのレビューが海外からの訪日客なのか、地元の観光客であるかどうかを直接判断できず、「対照組み合わせ」を選択することが難しい。一方、箱根のように英語と日本語の観光レビューが混在し、日本語のレビューのほとんどが日本人観光客によって書かれていると考えられる観光地区の場合は、中国語と日本語の「対照組み合わせ」を選択することが容易である。また、日本の国内観光産業は非常に良好であり、日本の観光客による優れた経験・体験に基づく観光レビューは、中国大陸観光客の満足度を高めるためにも利用する価値がある。

4.3.1　サンプルの収集と分析単語の選出

1) 日本と中国の観光客が旅行を検討する際に、それぞれ自国の「OTA」Web サイトを主に使用していることを考慮し、日中でそれぞれ市場でのシェアが最も高いオンライン旅行会社である楽天トラベルと Ctrip（携程）のユーザー評価システムからデータサンプルを収集した。収集時期は 2018 年 4 月から 2019 年 3 月までの 1 年間の観光レビューを分析対象とし、レビューの収集を行うには「八爪鱼采集器」というソフトウェアを用いた。
2) 単語数が非常に少ない(10 バイト未満)レビューを削除することで、日本によるレビュー(「対照組み合わせ」)の頻出名詞を統計した。同義語を 1 つにまとめた後、最終的に 109 個の候補語を共起行列の統計用語として使用した。
3) この 109 個の単語を中国語に翻訳した単語と組み合わせ、中国語のレビューのそれぞれの頻度をカウントし、頻度の高い 39 個の単語を未知の情報部分として予測分析した。
4) その語、特定の単語の出現回数が計算結果に与える誤差の影響を減らすために、共起行列の出現回数を 0 回出現したものを 0、1 回出現したものを 1、2 と 3 回出現したものを 2、4 回以上出現したものをまとめて 3 とすることで、共起行列における単語の出現回数を分類した。まとめた結果を使うことで、観光レビューに出現した単語に潜在的に含まれる観光客がある物への関心を示す度合いを表示することが可能になった。

5) 現在の演算能力を考慮すると、最終的に解析したデータのレビューサンプルは、この年の日中における観光レビューから、予測サンプルとして中国語のレビューのうち、最も出現した単語を含む 300 件のレビューと、最も頻度の高い単語を含む 600 個の日本語レビューからなる対照サンプルのセットから形成されたものとなった。

4.3.2 グループ化による分析計算

前節 4.3.1 で選択したデータから抽出した分析処理に使用するサンプルは 109×900 の大きな行列であるが、このような行列を単独に分析すれば、使用する分析処理に用いる計算資源が限られている場合は大変非効率的であり、また、一度勾配降下法で実行されると、ローカルミニマム(local minimum)に大きな誤差が生じる可能性が高い。そこで、Python プログラムを効果的に実行し、既存の研究機器を最大限に活用するために、本研究では大きな行列を複数のグループに分割し、異なるデバイスに分散して計算分析を行った。

具体的なグループ化は以下の通りである。

まず 300 件の中国語のレビューを 10 個のグループに均等に分割し、それぞれを a1−a10 と記し、各グループに対して 30 件のレビューを作成する。次に、600 件の日本語レビューを上記と同様に 10 個のグループに均等に分割し、それぞれを b1−b10 と記し、各グループで 60 件のレビューからなった。各中国語グループには、日本語グループと対になるように小さな行列を新たに形成し、a1−b1、a1−b2、... a1−b10, a2−b1、a2−b2、... a2−b10, ...、a10−b10、のように合計 100 個の小さい行列とする。これにより、複数の行列が並列に操作され、ひとつの処理デバイスに限定されず、複数のデバイスで同時に操作できる。

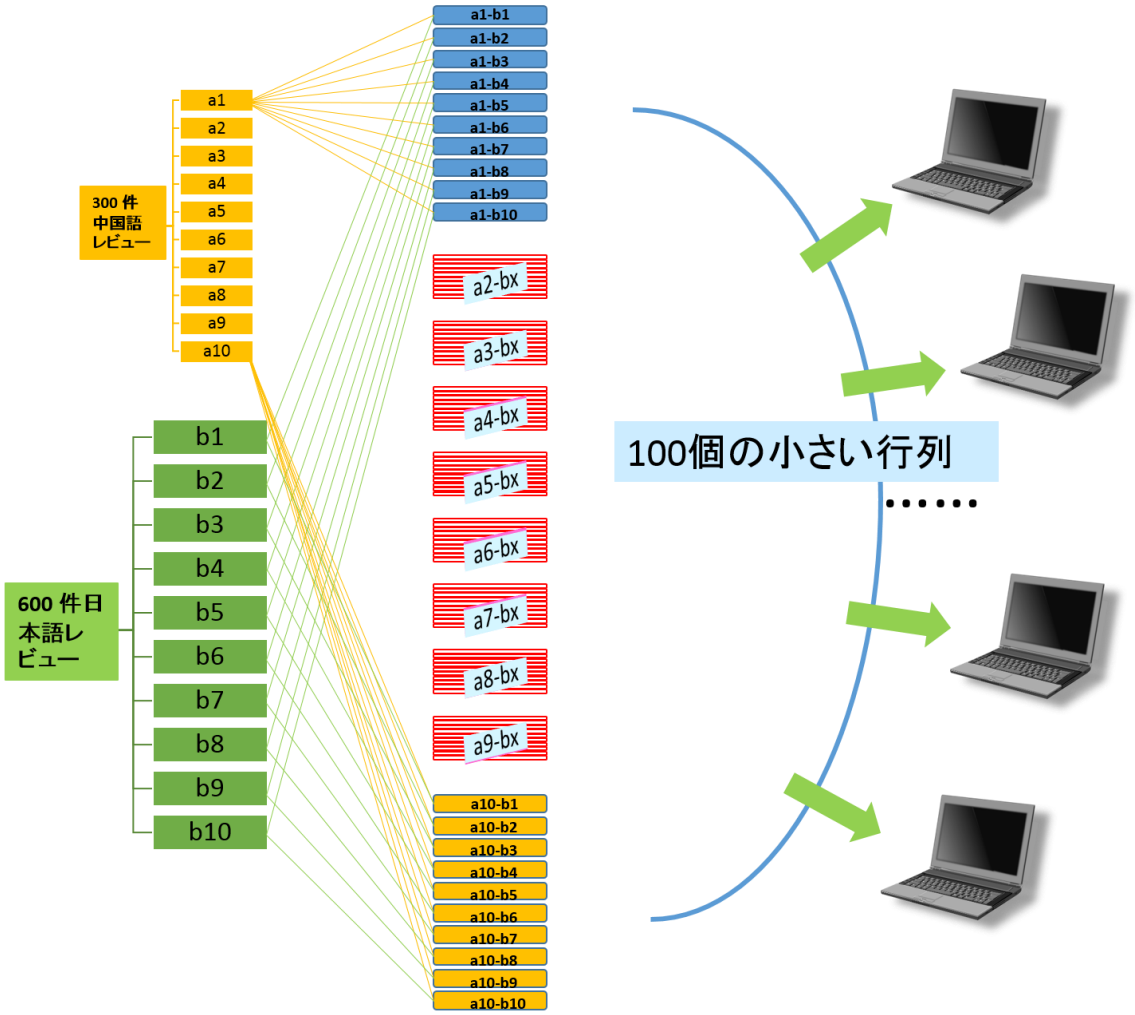


図 4-7　グループ分け手法の概略図

計算処理の容易さ、効率的な分析に加えて、グループ化のもうひとつのメリットは、計算中にローカルミニマムが出現した場合でも、計算と初期ランダム値を再選択することが容易であり、最終的なエラー値を観察する際、異常に大きな結果が突然現れても計算全体を再実行する必要がないことである。この小さな行列を再計算するだけで、計算における解析結果の安定性が保証される。

このような処理により、各中国語のレビューから 10 組の異なる行列が得られる。例えば、a1 の中にあるレビューに対して、a1−b1、a1−b2、.....a1−b10 の 10 セットで 10 回予測され、その後、これらの 10 の予測結果を平均計算し、最終的な予測結果を得た。

4.3.3　事例の分析結果

このような一連の処理を経た結果、最終的に、日本人観光客のレビューにおいて出現回数が多い 39 件の実用語について、中国大陸からの観光客の関心度を予測し、その順位を表 4-4 に示した。

表 4-4　訪日中国大陸観光客の予測関心度の順位

単語	单词	predicted degree
バイキング	自助	113.4709277
デザート	甜点	99.61116816
車	开车	75.00245357
階段	楼梯	68.97082571
眺め	眺望	60.50150455
掃除	打扫	58.29989874
ご飯	米饭	47.46552326
コーヒー	咖啡	47.24309167
テーブル	桌子	45.16785961
娘	女儿	43.84809163
男性	男生	39.33489705
放題	吃到饱	37.38131509
確認	确认	35.60457779
母	妈妈	35.56466946
説明	说明	34.05660503
廊下	走廊	33.99460413
駐車	停车	33.6946386
雰囲気	气氛	31.41187489
外国	外国	30.30807086
事前	事前	25.98472705
タオル	毛巾	23.33057675
館内	内部	22.96987861
両親	父母	22.00615944
アイス	冰激凌	21.63119526
メニュー	菜单	21.00362454
妻	老婆	20.40678892
夫婦	夫妻	18.58886263
笑顔	笑容	17.39486751
席	座位	13.78848248
思い出	回忆	11.69230035
源泉	源泉	10.86136368
シャワー	淋浴	9.011975649
誕生	生日	6.581941896
大人	成人	6.276046214
床	地板	4.803066697
洗面	洗脸池	2.35885824
冷蔵庫	冰箱	-6.036662906
本館	本馆	-11.46957273
記念	纪念	-21.92699735

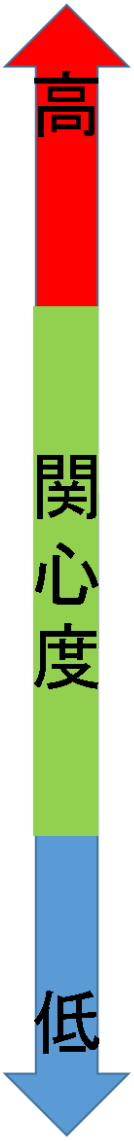


表 4-4 の結果より、以下の結論が導出される：

1)　「記念（纪念）」、「誕生（生日）」、「思い出（回忆）」などの単語に対しては、あまり関

心がない

これは、中国大陸観光客とは明らかに違い、箱根への日本人観光客の多くは記念日や誕生日などを祝うために箱根を訪れるため、記念日に実施される観光計画やパッケージ活動は、中国大陸観光客にはあまり適さないと推測される。

2) 「冷蔵庫（冰箱）」、「洗面（洗脸台）」、「床（地板）」などの単語の注目度が低く、「車（开车）」や「眺め（眺望）」などの注目度が高い

前者は宿泊施設の屋内環境機器に関する単語であり、後者は宿泊施設の外環境を説明する単語である。特に中国大陸からの観光客は初めて日本を訪れる観光客が多いため、なるべく多くの観光スポットを楽しむことが最優先である。また、日本人観光客が時間に余裕を持って観光するのに対し、中国大陸観光客は観光中の移動を重視する傾向があるため、中国大陸観光客は屋内の部分に特に関心がない。したがって、宿泊施設経営者は中国大陸観光客に対して、近くの観光スポットについての観光案内や支援を強化し、専門的で便利なツアーを提供できると、観光客のさらなる満足度を高められ、より評価の高い観光レビューや口コミを増加することができる。

3) 「ご飯（米饭）」、「コーヒー（咖啡）」、「デザート（甜点）」、「バイキング（自助）」、「放題（吃到饱）」などの食事関連の単語の注目度が高い

食事関連の単語は、もともと中国語のレビューにおける頻出語ではなく、中国の文化と関係があることが要因であると考えられる。まず、中国大陸観光客は米の品質について高い要求をしていない。次に、コーヒーの人気も日本ほど高くなく、食事後のデザートの食事習慣は、西洋料理の食事後の活動に限定されることが多い。更に、食べ放題のような飲食販売形式は、中国国内のレストランではめったにない。しかし、このような単語は、高い関心度の予測スコアをもっているので、関連観光事業者は中国大陸観光客にこれらの側面の飲食サービスを推奨するだけでなく、文化の違いに着目すべきであり、基本的なルールの説明と導入をすることで、中国大陸観光客の観光体験を改善することができる。

4) 家族関係の単語では、「娘（女儿）」は「両親（父母）」、「妻（老婆）」、「母親（母亲）」よりもはるかに高いと予測されている

主な理由として次の 2 つの側面が考えられる：

a.初めて日本を訪れる中国大陸観光客が多く、日本についてまだ多くを理解していないことも含めて日本に対する新鮮な感覚を持つことや、子供と一緒に観光することが多いため、業界は中国大陸観光客に対して丁寧なサービスや観光案内を提供する必要がある。

b.年上の両親と一緒に観光するよりも、子供と一緒に旅行する方が中国大陸観光客に受け入れてもらいやすいため、親子向けの旅行プログラムをより多く提供することや、子供に関するサービスや施設を増やしたり充実させたりすることを観光関係事業者に対応してもらう必要がある。

第 5 章　中国大陸からのインバウンド観光発展のための傾向予測と実例

本章の内容は、3 章・4 章の内容の統合と拡張である。第 3 章で考察した傾向分析、そして、第 4 章で考察した予測分析を活用し、分析対象とする観光レビューの種類の範囲をさらに拡大するため、本章では相関分析を行う対象データを前述の宿泊施設ではなく、観光地・スポットとした。このように選択を行う目的は、研究方法の適用性をさらに検証し、本研究の結果を実際の観光業務に活用することを容易にするためである。

そこで本章では、まず観光地・スポットに対するレビューを処理する際に、データの局限性を修正・統合する必要があることを示す。次に、予測分析と傾向分析を統合する際に、各時間区分の区切りと予測分析の整合性を十分に配慮する必要があることを示す。そして、最後に、「箱根の美術館訪問タイプ観光スポットのオンラインレビューに基づく観光目的と嗜好傾向の予測変化分析」を事例として考察した結論に基づき、未知情報を活用して、観光業界の現場業務への反映を容易にするため適用方法を提示した。

5.1　本章分析の意義

5.1.1　データ選択の多様化

これまでの事例分析では、宿泊施設に関する観光レビューを基に中国大陸観光客の傾向に関する考察と関心のある項目について予測分析を行ってきた。宿泊施設のレビューを用いた理由としては、「OTA」レビューの中で最も数が多く、内容項目も最も広範囲に及ぶため補助的な情報が多く含まれるテキストデータであるためである。宿泊施設レビューは観光者の宿泊予約や投宿といった実際の観光消費に繋がり、「OTA」の Web サイトを閲覧する観光客は宿泊施設を探す際に、必ずオンラインのレビューを参照するため、レビューを見ながら宿泊施設を予約する習慣がついているとも言える。

一方、宿泊施設に関するレビューと比較すると、観光地に関するレビュー数は少ない。これは観光の計画時に何を優先して考えるかという視点に関係があると考えられる。ほとんどの観光客はまず宿泊施設を事前に予約するが、観光地そのものが大混雑するようなリスクを考慮する人はほとんどいない。その理由としては、観光客は当日の天気や疲れ具合など常に変化している状況に応じて、その日のスケジュールの選択にある程度の柔軟性を持ち、自由に変更したいからであり、観光地のチケットの事前ネット予約などは、それほど重要ではない、と考えられるからである。

したがって、観光地のチケットやサービス等を観光客に「OTA」の Web サイトを通じてより多く事前予約などで利用してもらうためには、観光客を行列に並んで切符を買うことから解放したり、複数の観光スポット、各種イベント、飲食店・土産店などを、より多くの組み合わせで予約販売する形式などによる魅力的な割引を提案することで事前予約による観光消費をさらに促進すべきである。

このようなオンラインによる事前予約の促進方法は次の 3 種類に大別される。

1)　宿泊施設と観光施設との連携による割引・特典の提供

宿泊施設の宿泊費を連泊すると割引するプランを作成したり、オンラインで直接購入すると割引やポイントなどの特典を提供することや、宿泊施設を観光地の観光施設やイベントのチケットと同時に予約すると割引やお土産引き換えなどの特典を提供することが例として挙げられる。

図 5-1　観光スポット白鹿洞書院の「入場券＋宿泊」の特割セット

(Web サイト：https://piao.ctrip.com/dest/t2290.html より)

2)　観光地限定商品・サービスや記念品の提供

観光地で、独自の限定商品やサービスを提供する。例えば、オンライン予約限定のドリンクデザートセットをチケットと併せて販売したり、記念品をプレゼントする特典などを予約時に提供。

図 5-2　「OTA」の Web サイト限定プラン

(Web サイト：
https://www.jalan.net/kankou/spt_14382cc3300033812/activity/l000023A52/?showplan=spot_detail_kankou&asobiKbn=1 より)

3)　観光地におけるオプション商品・サービスの提供

観光プラン策定のためにサイトを利用する際に、必要な情報を提供するのと同時に、観光地のオプション商品・サービスを販売する。例えば、家族・団体チケットの購入、駅や空港からの専用送迎車や専用ツアーガイド通訳の手配などである。

図 5-3　入場チケットや器具レンタルのチケットを含む日帰りプラン

（Web サイト：https://activity.ctrip-ttd.hk/ottd-activity/dest/t20590802.html より）

このように多様な販売形式の発展に伴い、観光地に関する「OTA」レビューも豊富に書き込まれるようになり、相乗的に観光客の関心も高まっていく。

更に、観光地に直接関連するレビューは、観光地の発展との関係がより密接となり、観光商品・サービスを提供する側も、この部分の内容を参照しながら産業を発展させることで、より高い効果が得られる。そこで、本章では分析対象として「OTA」Web サイトに投稿された観光地に関するレビューを用いてテキストマイニング分析を行い考察を展開する。

5.1.2　産業発展傾向の予測

「OTA」Web サイトを通じてオンラインで提供される観光商品は、他のオンラインショッピングの商品などと明らかに異なり、商品の多様性、一意性が非常に明白であるが、オンラインショッピングで販売される一般の商品に関するレビューの多くは、ある特定の商品に対する評価であり、不良品の誤差を除けば商品そのものには差がない。しかし、「OTA」Web サイト経由でオンライン提供される観光商品の内容は商品ごとに大きな違いがある。

宿泊施設を例とすると、同じタイプの宿泊施設であっても、宿泊フロアの高さや部屋の向きの違いは顧客に異なる影響を及ぼし、また同じ部屋であっても、季節や年、天候の変化により著しい差異が発生する可能性がある。さらに、観光客が宿泊中に経験・体感する状況が個々に異なることに加え、観光商品に対する評価に影響を及ぼす要素が多いため、観光消費者の単純に「良い」「悪い」といったポジティブかネガティブかの感情表現を直接考慮したり、商品そのものの満足度だけを分析したりすることは、基準判断としては不十分である。

また、第 3 章で中国大陸観光客に関する分析と考察を行った通り、中国大陸の観光産業の発展が急速であるため、観光するタイミングによって顧客の需要と関心の注目点が変化しており、年間でも閑散期と繁忙期で区分すると明らかな差異が認識された。

その第 3 章での観光レビューのテキスト分析に基づき、第 4 章では中国大陸からの訪日インバウンド観光客の満足度を高めるために、中国大陸観光客の観光レビューと日本在住の地元観光客の観光レビューを組み合わせて、中国大陸観光客のレビューであまり言及されない単語があることの原因について考察を行った。中国大陸からのインバウンド観光客の属性や構成を考えると、初めて海外に行く人の割合が大きく、さらに中国文化と他の文化の差が大きいため、アウトバウンド観光に関する情報収集が不十分で基本的に事前情報が不足している。したがって、一部の単語が言及されていないからといって、中国大陸観光客が関心を持っていないとは考えるのは正しくないため、本章では、行列の分解手法を用いてその関心分野の予測を行うこととした。

時間の経過に伴う観光客の関心分野やレベルも変化するという観光レビューの特性も考慮し、本章では異なる年のテキストデータを用いて観光に関する関心の予測を行うことにより、中国大陸観光客が海外の観光地に関する未知の情報の中から優先的に薦めるべきアイテムの傾向を抽出することを試みた。それを地元の観光産業の経営や観光政策に反映することは、中国大陸観光客へのサービス向上に大きな役割を果たせることができると考える。

5.2　データ構造と手法

5.2.1　観光スポットレビューの統合

第 4 章で述べたように、「OTA」Web サイトにおける観光レビューは通常、実際の観光消費に伴って投稿されるため、同じ観光地内の複数の観光スポットのチケットなどをセットにして販売された観光商品・サービスに関するレビューは、特定の観光スポットに対するコメントの抽出が難しいため分析対象から除き、範囲を限定しやすい単一の観光スポットに関するレビューを分析の対象とした。

一方で、観光スポットに関するレビューのデータ量が十分かどうか判断する必要があるため、「OTA」の Web サイトにあるデータ量を調べたところ、単一の観光スポットだけでは不足している地域もあることが判明した。このことから、観光地で、そこに対するレビューだけで十分に分析できる可能性は低い。したがって、地域によっては複数の観光スポットのレビューをある程度統合する必要があると考えられる。

観光スポットの様々な特性を考慮すると、たとえ同じ観光エリアの観光スポットであっても、それぞれの観光資源のセールスポイント、観光客の観光目的及び観光客の構成は、例えば自然景観型スポットと人文観賞型スポットや団体ツアーと個人旅行のように大きく異なることに着目する必要がある。

そこで、本研究では、観光地の中でも代表的な観光スポットを特定・抽出し、この部分に関する観光レビューをひとつのまとまりとして統合した上で分析・考察を行った。具体的には、同じタイプの観光スポットを訪れる観光客は類似な趣味や観光目的を持ち、分析結果の考察でも、その指向性が確認できた。例えば、寺社仏閣といった宗教関係の観光スポットの統合、遊園地類の観光スポットの統合、自然景観を観光資源とする観光スポットの統合などである。以下、類似する観光スポットに関するレビューが統合された異なる分類の分析結果をまとめた上で、観光地全体の分析まで考察を展開する。

5.2.2　時間区分によるレビューの検討

観光スポットの予測分析を行うための時間区分に関しては、以下の 3 点の理由から年単位で検討することとする。

1)　分析計算の観点から考えると、予測結果は正確な数値ではなく傾向として表される。そ

のため、年間の閑散期と繁忙期とで区分してしまうと、両者のコア単語（頻出語）の差異は非常に大きくなり、その結果として行列計算の処理対象となる単語の構成が異なってしまうため、それぞれの共起行列の作成と処理が必要になり、予測分析に必要な処理が必要以上に複雑、かつ大量になってしまう。したがって、観光スポットの予測分析の効率性の観点から、年単位の時間区分で頻出語を統合する手法が最適であると考えられる。

2) 異なるタイプの観光スポットでは、閑散期と繁忙期に受ける影響も大きく異なる。例えば、室内観光スポットと屋外観光スポットでは季節や天候から受ける影響は全く異なる。また、同じ観光地の違う観光スポットでも閑散期と繁忙期の時期も異なる。そのため様々な観光スポットをまとめ、観光地全体を分析するためには、このような差異を平準化できる年単位で区分することが最適である。

3) 異なる観光地において本手法を適用する場合、年単位で区分する手法は最も簡単、有効かつこれまでの研究や統計でも用いられてきた伝統的で手法でもあり、実際に閑散期と繁忙期の状況を把握できない観光地に対してもこの年単位での区分方法を利用して分析することが可能となる。

5.2.3 予測サンプル・参考サンプルを用いた予測分析

第 4 章で述べたように、効率的な予測計算を行うために最も重要な前処理は不完全な共起行列を作ることであり、予測計算を実現するための基礎でもある。この共起行列は単語とドキュメントの 2 つからなる。

1) 単語

具体的な分析計算の対象となる単語は第 4 章で考察した分析手法に従い、以下の手順で前処理を行う。本研究では顧客が関心を持つ度合いを予測し、分析結果を実際に有意なものとするため、名詞のみを分析対象の単語とした。また、「対照組み合わせ」のレビューにおける頻出語をコア単語と定義する。そして、分析の結果が長期にわたっても有効となるよう、頻出語の抽出対象期間は年ごとではなくレビュー期間の全体を対象とする。そして最後に、コア単語を確定した上で、「予測組み合わせ」の単語と対照する。その際、同じ頻出語を既知情報と定義し、出現頻度が明らかに「対照組み合わせ」より低いか、全くない単語を予測情報と定義する。

ここで、「予測組み合わせ」に含まれる単語は中国語であるが、中国大陸観光客のアウトバウンド観光の訪問先の地元観光客の観光レビューは、ほとんどの国と地域で中国語は使われていないため、頻出語の翻訳した上で「対照組み合わせ」を作成する必要がある。また、

翻訳処理の過程で、一対一対応が高精度ではできないため、同義語を統合した上での処理も必要不可欠である。さらに、この翻訳後の「対照組み合わせ」を作成する際、単語の意味対照だけでなく、既存の辞書にはまだ含まれていないや観光に関する専門用語やインターネット上で使われるスラング（俗語・隠語・略語など）を考慮する必要もあり、前処理する際にこれらの用語・単語をカスタム辞書に追加設定する。

2)　ドキュメント

観光レビューに含まれるテキストに関しては第 4 章の分析と同様に、テキストデータを「予測組み合わせ」と「対照組み合わせ」に分ける。前節 5.2.1 と前節 5.2.2 で考察してレビューの統合・時間区分に基づき、「予測組み合わせ」は中国大陸観光客のある観光地での代表的な観光スポットに対する観光レビューの組み合わせであり、これらを収集・統合した後に時間による区分を行う。また「対照組み合わせ」は予測対象となる観光地の地元観光客がよく利用する「OTA」Web サイトの観光スポットに関するレビューの組み合わせであり、上記と同様に収集・統合した後に時間によって区分を行う。

以上に基づき、時間による区分を行い、下記の表 5-1 に示した異なる年の「単語―ドキュメント」の不完全な共起行列を形成した。

表 5-1　「単語―ドキュメント」の不完全な共起行列

		日本語の頻度名詞									
		中国語の高頻度名詞単語							中国語の非高頻度名詞単語		
		W1	W2	W3	W4	.....	.....	.....	.....	W'1	W'2	W'3	.....	.....
予測区分（中国語のレビュー）	CN-C1	0	2	0	1.....	.....	.....	.....		—	—	—	...	
	CN-C2	1	0	0	1.....	.....	.....	.....		—	—	—	...	
	CN-C3	2	0	0	0.....	.....	.....	.....		—	—	—	...	
	CN-C4	0	2	3	0.....	.....	.....	.....		—	—	—	...	
	.....	.....	.....	.....		.....	.....	.....		—	—	—	...	
	.....	.....	.....	.....		.....	.....	.....		...	...	...	...	...
対照区分（日本語のレビュー）	JP-C1	0	0	0	1.....	.....	.....	.....		0	0	1	.....	
	JP-C2	1	2	0	0.....	.....	.....	.....		1	2	0	.....	
	JP-C3	0	1	0	1.....	.....	.....	.....		0	3	0	.....	
	JP-C4	2	1	0	1.....	.....	.....	.....		2	0	1	.....	
	.....	.....	.....	.....		.....	.....	.....		.....	.....	.....		
	.....	.....	.....	.....		.....	.....	.....		.....	.....	.....		
	.....	.....	.....	.....		.....	.....	.....		.....	.....	.....		

この不完全な共起行列は、潜在因子モデル手法を用いた不完全な行列のマトリックス補完（matrix completion）によって予測「単語」の数値を計算した結果である。本研究では、これらの計算結果数値の大小関係に基づき、観光レビューに出現した単語の中から中国大陸観光客が潜在的に持つ関心の度合いを表示することを可能にした。

そして、各時間区分の状況を個別に注目し、その区間の特徴を抽出した上で区間相互の比較を行う。観光レビューのデータが時間の経過と共に変化することを考慮すると、関心度に係る傾向の変化を把握することが、比較分析においては最も重要な点である。最終的な具体的な結果の検討は数値の変化そのものではなく、相対的なランキング上の変化の傾向に注目することになるが、その背景については、後述の事例分析で詳しく考察を行う。

5.3　箱根の美術館訪問タイプ観光スポットのオンラインレビューに基づく観光目的と嗜好傾向の予測変化分析の事例

本節では日本有数の観光地、箱根における美術館訪問タイプの観光に関するオンラインレビューを予測変化分析の事例として選んだが、それは、以下の利点を考慮したためである。

1) 中国大陸からの訪日インバウンド観光は急増しており、中国大陸観光客の満足度を向上させる取り組みにおいて、日本は他国にとっても参考となるモデルケースになっていること。
2) その中でも、箱根は区域区分を明確にすることが可能な単独の観光スポットとして訪問する中国大陸観光客の数、そして日本在住の地元観光客の数、さらにオンラインレビュー数も十分であること。
3) 「対照組み合わせ」の常用言語からみると、英語で投稿されたレビューは書き手の身元と国籍を遡ることができないことに対して、日本語で投稿されたレビューの書き手は、ほぼ日本人で地元の観光客であると判断できること。
4) 第 4 章では箱根の宿泊施設に関する観光レビューをデータ分析の対象としたが、予測変化分析でも箱根地区の観光スポットに関する観光レビューを対象として分析を行うことで、同一の観光地を複数の視点から分析することができること。
5) 美術館訪問タイプの観光スポットは箱根の観光産業において、非常に重要かつ主要な観光パターンであること。

5.3.1　観光レビュー・データの収集

中国大陸観光客の観光レビューを対象とする「予測組み合わせ」について、現在、中国大陸のオンライン観光市場でシェアが最も高い「OTA」企業の Ctrip（携程）のユーザー評価システムからデータサンプルを収集した。そして、「対照組み合わせ」については、中国大陸観光客の訪問先の日本の地元企業であるため分析対象となる観光レビュー数がやや多いが、本研究ではネットワークフロー（観光に関する情報提供、各種予約、旅行代金決済、口コミの投稿など）が最も多い jalan.net（日本の旅行関連サイトで最も取引量の多い企業：https://www.similarweb.com/ja/top-websites/category/travel-and-tourism）に投稿された観光レビューをデータの収集元とする。この jalan.net は日本の「OTA」企業ランキングの上位にある「OTA」企業 Web サイトの中でも観光スポットの網羅性も高く、かつデータ量も多いサイトでもある。

表 5-2　日本「OTA」企業トラフィックデータ順位

（資料：オンライン旅行業界の国際会議「WIT Japan」でシミラーウェブとフォーカスライトのデータによるトラフィックデータ（2018 年 5 月）、プレゼン資料より）

日本「OTA」会社トラフィックデータランキング		
ランキング	「OTA」会社名	ネットワークフローの2018年5月3日に総訪問数（M）
1	じゃらんnet	40.44
2	楽天トラベル	28.34
3	トリップアドバイザー	25.70
4	全日本空輸	16.19
5	Japan Airlines	14.91
6	@4travel.jp	11.57
7	LINEトラベルjp	9.04
8	RETRIP	8.64
9	エイチ・アイ・エス	8.60
10	一休.com	7.56

このように、観光レビューを収集する「OTA」企業サイトを選定した後、サイト上で箱根エリアで最も人気のある 25 個の観光スポットを抽出し、更に、そのうちの美術館訪問タイプの観光スポットを選択した。

具体的には、「彫刻の森美術館」「箱根ガラスの森美術館」「ポーラ美術館」「箱根美術館」「箱根芦ノ湖成川美術館」「箱根ラリック美術館」「箱根武士の里美術館」の 7 つの美術館について、jalan.net にあるそれぞれの紹介ページにアクセスし、レビューとその関連情報を収集した。

jalan.net のサイトには、個別の観光レビューごとに多くの付加情報が含まれており、訪問・投稿時間や顧客の基本情報などが含まれているため、下記の図 5-4 に示すように、本研究では今後のさらなる研究のために全ての情報を併せて収集した。

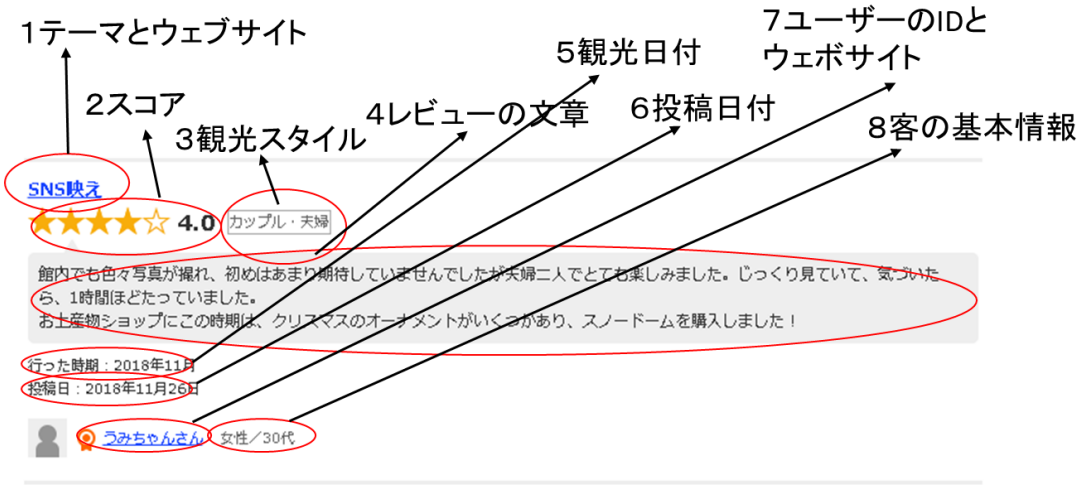


図 5-4 「jalan.net」のサイト

一方、jalan.net と比べると、Ctrip（携程）の観光スポットのレビュー内容は、レビューテキスト自体を除くと非常に乏しく、評価点とユーザーID、投稿日の 3 つの付加情報しか提供していない。

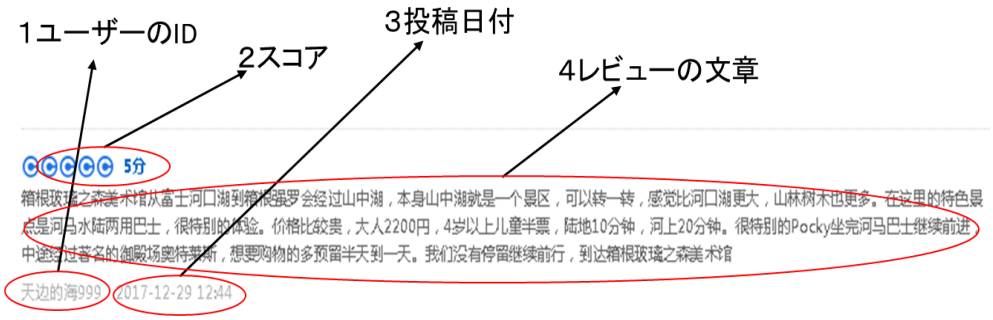


図 5-5 「C-trip」（携程）のサイト

ただし、本研究はレビュー内容のテキストマイニングを中心とする研究であり、時間のデータがあれば傾向の分析を検討するには十分であると判断した。収集の手法は前述した 2 つの事例と同様に、「八爪魚采集器」というソフトウェアを用い、収集したデータは 7 つの美術館スポットに関する 2016 年 4 月 1 日から 2019 年 3 月 31 日までの全てのレビューである。

5.3.2　不完全な共起行列の構築

1)　単語の選択

前節 5.3.1 で述べた通り、まず jalan.net における 3 年間の観光レビューから頻出語を集計し、原則として出現数が 20 回以上の名詞（コア単語）を抽出した。また、出現数が 5 回以上 20 回未満の単語も抽出し、コア単語の意味に類似する単語についてはコア単語との統合を行った。その結果として、同義語を統合した 108 個の頻出語が得られた。また、これらの単語は Ctrip（携程）の中国語による観光レビューでも頻出語であるかどうかを判断するために、中国語の観光レビューについても同様の分析計算を行った。

その結果、108 個の頻出語のうちの 62 個の単語は、中国語による観光レビューにおいても頻出語となっていた。そこで、この 62 個単語の情報を既知情報、そして、残りの 46 個の単語を予測情報として、本研究の分析で取り扱うこととした。この予測情報とする 46 個の単語は以下の通りである（日本語に対応する中国語をその単語の後の括弧中に記す）。

「子供（孩子）」、「レストラン（餐厅）」、「雰囲気（气氛）」、「カフェ（咖啡）」、「土産（特产）」、「家族（家庭）」、「遊歩道（散步）」、「眺め（眺望）」、「お茶（茶）」、「女性（女性）」、「夏（夏）」、「紫陽花（紫阳花）」、「コンサート（音乐会）」、「平日（工作日）」、「ランチ（午餐）」、「クリスマス（圣诞）」、「ラリック（拉利克）」、「車（开车）」、「小学生（小学生）」、「カップル（情侣）」、「池（池子）」、「イルミネーション（夜景灯光）」、「オリエント（东方）」、「途中（路上）」、「アスレチック（体育运动）」、「階段（台阶）」、「外国（国外）」、「夫婦（夫妇）」、「娘（女儿）」、「雪（雪）」、「歌（歌）」、「トイレ（厕所）」、「ベビーカー（婴儿车）」、「シャボン（泡泡）」、「ススキ（狗尾草）」、「仮面（面具）」、「空気（空气）」、「スタッフ（工作人员）」、「コーナー（角落）」、「傘（雨伞）」、「冬（冬）」、「ジャム（果酱）」、「夜（晚上）」、「パスタ（意大利面）」、「苔（苔）」、「夕方（傍晚）」

2)　ドキュメントの選択

3 年間の観光レビューを用いて、分析可能な共起行列を構築するために、観光レビューのドキュメント内のテキストに対して絞り込み作業を行った。

まず、重複するレビューを削除し、次に、「対照組み合わせ」のレビューにおける出現頻度が 2 回以下のデータと「予測組み合わせ」の出現頻度が 1 回以下のレビューを取り除いた。この 2 つの作業で絞り込んだ結果、「予測組み合わせ」のうち中国語による観光レビューは 311 件となった。それを時間を遡る順で 1 年単位で時間区分し、それぞれに A 組（2018.4~2019.3）76 個、B 組（2017.4~2018.3）79 個、C 組（2016.4~2017.3）76 個と名付け分類した。一方、「対照組み合わせ」の日本語による観光レビューは全部で 2761 個であったが、「予測組み合わせ」と同様の時間区分を行い、それぞれに X 組（2018.4~2019.3）529 個、Y 組（2017.4~2018.3）812 個、Z 組（2016.4~2017.3）1420 個と名付け分類した。

上記のように分類した「予測組み合わせ」と「対照組み合わせ」からひとつずつ組み合わせて予測分析対象を構成し、A−X、A−Y、A−Z、B−X、B−Y、B−Z、C−X、C−Y、C−Z の 9 通りの組合せを生成した。

3)　共起行列における単語の出現回数の再評価

データの収集と統計分析処理により絞り込んだ単語とドキュメントに基づき共起行列の作成を行った後、特定の単語の出現回数が計算結果に与える誤差の影響（ひとつの単語の過剰な出現で計測誤差が過大になることなど）軽減するために、第 4 章と同様にすべての共起回数の分類を行った。

具体的には、共起行列に 0 回出現したものを 0、1 回出現したものを 1、2 回もしくは 3 回出現したものを 2、4 回以上出現したものをまとめて 3 とすることで、共起行列における単語の出現回数を分類した。このようにとりまとめた結果により、観光レビューに出現した単語に潜在的に含まれる、観光客が関心を示す対象を関心度と併せて表示することが可能になる。

5.3.3　グループ化を用いた計算分析

本研究で用いた計算分析用のハードウェア構成の分析能力は、非常に大規模な不完全行列の補完作業を処理するのには不十分であるものの、既存の研究機器を最大限に活用するため、第 4 章で説明したように研究室にある複数台の PC を接続し、グループ化することで計算を分散する形で処理を行った。そして、分析対象のグループ化を合理化することが出来、効率的かつ正確な予測結果を得ることができた。

本事例では単語数が 108 個であることから、計算では単一の行列が 108×100 程度の不完全行列になすが、計算効率を確保するためには、予測の割合を 10~15%とすることが妥当であると考えた。そこで、予測のドキュメント数と対照のドキュメント数との割合を約 3:7 とした。したがって、（30×46）/（108×100）=12.8%であれば妥当と考える。

そして、「予測組み合わせ」を更に A 組を a1~a3 の 3 組、B 組を b1~b3 の 3 組、C 組を c1~c5 の 5 組に区分した。「対照組み合わせ」も同様に、X 組を x1~x7 の 7 組、Y 組を y1~y10 の 10 組、Z 組を z1~z18 の 18 組に区分した。その各組見合わせの具体的なレビュー数を下記の表 5-3 に示す。

表 5-3　各組み合わせのレビュー数量

期間	時間区分（予測組）	計算区分	各レビュー数	時間区分（対照組）	計算区分	各レビュー数
2018.4 ～ 2019.3	A組	a1	25	X組	x1	72
					x2	72
		a2	25		x3	72
					x4	73
		a3	26		…	73
					x7	73
2017.4 ～ 2018.3	B組	b1	26	Y組	y1	77
					…	77
		b2	26		y4	77
					y5	78
					…	78
		b3	27		y10	78
2016.4 ～ 2017.3	C組	c1	31	Z組	z1	76
		…	31		…	76
		c4	31		z17	76
		c5	32		z18	77

11 個の「予測組み合わせ」をそれぞれ 35 個の「対照組み合わせ」にひとつずつ対応させ、11×35=385 個の計算組合せを作成する。この組見合わせ全体の中で最も小さい行列が（25+72）×108 であり、最も大きい行列が（32+78）×108 である。これで予測割合も適正な範囲内に収まり、選出された潜在因子数は 5 個となった。

具体的な分析計算は前述の通り複数のコンピュータで分散して行い、誤差が大きくならないようにひとつずつ勾配降下法で計算した損失値の和をチェックした上で時間区分によってとりまとめ、予測値を収集し並べ替えた上で比較を行った。

5.3.4　時間分区による分析

時間区分によるとりまとめのプロセスは、「予測組み合わせ（A、B、C 組）」と「対照組み合わせ（X、Y、Z 組）」のそれぞれのデータを時間区分によって統合することである。例えば A−X 組の場合、その中の a1 組は x1−x7 に対応する 7 つの計算結果からなるため、予測される部分は 7 つの異なる 46×25 行列となり、これら 7 つの行列を加算して 7 で割ると、「対照組み合わせ」のうち X 組の a1 組に対する予測結果となる。a2、a3 組も同様に計算することで X 組の a2、a3 組に対する予測結果が得られ、a1、a2、a3 の結果を統合したものは、A 組全体が X 組を対照とする予測をまとめた結果と見なされる。

帰納予測結果を得るためには下記の 3 手法が候補となる

「予測組み合わせ」と「対照組み合わせ」が同じ年のデータだけで考察を行う。具体的には、A−X、B−Y、C−Z の 3 組の結果を用いる。

手法1)　「予測組み合わせ」と「対照組み合わせ」を同じ年のデータだけで考察を行う。具体的には、A−X、B−Y、C−Z の 3 組の結果を用いる。

手法2)　各「予測組み合わせ」を、それぞれ 3 つの異なる時間帯の「対照組み合わせ」に対応させる。

手法3)　いずれの「予測組み合わせ」も 3 つの異なる時間帯の「対照組み合わせ」に対応させるが、「予測組み合わせ」と「対照組み合わせ」が同じ年の場合、その結果を加重する。具体的には、同じ年の「予測組み合わせ」と「対照組み合わせ」の結果を強調するため、2 倍の加重とする。

そして、これら 3 つの手法の優劣を比較すると、

手法 1 については計算データが同じ年の組み合わせであり、その対応性は最も強いが、A−Y 組、A−Z 組といった異なる時間区分を組み合わせた分析は利用できないという欠点がある。次に手法 2 については、すべてのデータを利用し、結果として数値的に最も数学的な一貫性が高いものの、「対照組み合わせ」の時間変化による影響を考慮することができない。そして、手法 3 は、すべてのデータを利用するのと同時に、「対照組み合わせ」の同一時間における影響力の強さを強調することができるものの、結果の考察において具体的な数字ではなく単語の順位でしか結果を示すことができないという欠点もある。

以上の比較に基づき、本研究では数学的な値の一貫性と比べて傾向の考察の正確さがより重要なことから手法 2 は採用せずに、本事例では手法 1 と手法 3 を組み合わせた手法で結果の考察を行うこととした。

5.3.5　事例の分析結果と考察

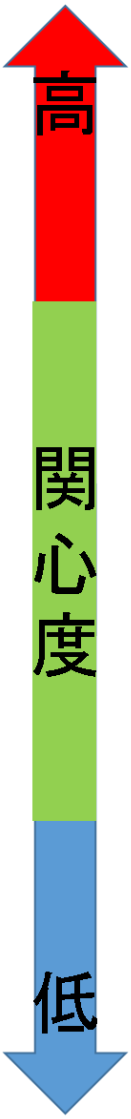
5.3.5.1　3 年間の観光レビュー全体のデータに基づく予測

まず、第 4 章で考察して選択した手法を用いて、3 年間の観光レビュー全体のデータについて予測分析を行った。

まず、傾向と時間区分は考慮せず、計算の手法として前節 5.3.3 で述べたように計算価値を等価にするため「予測組み合わせ」を 11 等分、「対照組み合わせ」を 35 等分した。そして、計 385 個の予測行列にそれぞれ 46 個の予測単語の予測値を重ね合わせた結果を以下の表 5-4 に示す。

表 5-4　3 年間の観光レビュー全体のデータに基づく予測結果

	単語	予測合計
1	子供	372.084074
2	レストラン	122.3933817
3	雰囲気	110.031716
4	カフェ	92.07002811
5	遊歩道	77.30961133
6	家族	73.35530554
7	土産	60.80250817
8	眺め	53.89506908
9	コンサート	41.61329688
10	お茶	40.15605021
11	夏	36.31603212
12	女性	35.04425254
13	平日	33.8557898
14	ラリック	33.4766459
15	ランチ	30.61089412
16	車	29.32421026
17	クリスマス	29.25788238
18	紫陽花	26.92430067
19	カップル	26.62038845
20	アスレチック	24.23654366
21	イルミネーション	23.71644082
22	小学生	23.62254329
23	オリエント	22.88154667
24	外国	22.84553355
25	コーナー	22.32831082
26	娘	22.12719574
27	途中	21.67216833
28	階段	21.15111302
29	ベビーカー	20.1673092
30	池	19.99467693
31	仮面	19.836626
32	夫婦	19.50017503
33	空気	19.49116119
34	雪	19.44163399
35	シャボン	18.15804789
36	スタッフ	17.85984377
37	冬	17.73190297
38	トイレ	16.41698186
39	ジャム	16.35490987
40	歌	16.22932817
41	ススキ	15.24404552
42	傘	13.96575339
43	夜	13.72751879
44	苔	11.63547073
45	パスタ	11.23200848
46	夕方	9.974771921



この結果を踏まえ、本節では下記の予測の情報に関わる 5 つの側面から考察を行った。

1)　観光パターンについて

「子供（孩子）」、「家族（家庭）」の予測値が非常に高いのに比べ、「夫婦（夫妇）」、「カップル（情侣）」などの単語は関心度としては比較的下位にある。この関心度の違いから、箱根の美術館訪問タイプの対象となる観光スポットについては、夫婦旅行やカップル旅行よりも家族旅行や親子旅行などを薦めたほうが効果的であることがわかる。そして、このような結果に基づき、美術館運営関連者に向けて、家族旅行や親子旅行に割引した入館料チケットやギフトショップでの購入割引券、記念品などをセットしたプランを企画・宣伝すること

で中国大陸観光客をより誘致することを推奨できる。

2) 季節について

関心度の違いから旅行をする季節として夏の優先順位は、秋と冬よりもはるかに高いことが確認された。これは雪が降る日の観光を推薦する度合いが低いことだけでなく、植物への関心度が高いことも反映されている。例えば、「秋ならススキ、夏ならアジサイ」といったレビューがあったが、実際には表 5-4 の予測単語の関心度の順位からすると、「紫陽花（紫阳花）」は 18 位で、「ススキ（狗尾草）」は 41 位であることから、アジサイが象徴する夏への潜在的な関心は、ススキが象徴する秋より、はるかに高いことがわかる。したがって、夏に中国大陸観光客を獲得することで、より多くの利益をもたらす可能性がある。

3) 施設サービスついて

「レストラン（餐厅）」、「カフェ（咖啡）」、「喫茶店（茶馆）」などが「トイレ（厕所）」、「スタッフ（工作人员）」よりはるかに関心度の順位が高いことが確認された。この背景には、美術館を訪れる中国大陸観光客にとって「日本のトイレは機能が豊富である」または「スタッフのサービスが素晴らしい」といった施設やサービスの品質よりも、レストランやカフェ、喫茶店などの観光消費場所そのものに対するニーズが極めて高いと考えられる。したがって、この点から、観光関連事業者は中国大陸観光客を対象に、より具体的な消費に関するサービスを提供することを提案することが重要であることがわかる。例えば、詳細な中国語メニューを用意したり、中国の「OTA」企業と連携して限定の飲食セットメニューを提供したり、特別なイベントを行うこと等でより多くの中国大陸観光客を引き付けることができる。

4) 音楽について

「コンサート（音乐会）」、「歌（歌）」の関心度の順位から見ると、ふたつとも同じ音楽の分野ではあるが、現場により近い「コンサート」の方が関心度の順位も高く、中国大陸観光客向け観光商品として潜在的なターゲットとして有効であり、BGM 音楽を流すだけなどの行為には特に中国大陸観光客は魅力を感じることはないことがわかる。

5) 旅程について

「眺め（眺望）」、「遊歩道（散步）」等の観光における具体的な活動かつ昼間の活動の単語の関心度の予測値が高いことに比べ、「夕方（傍晚）」や「夜（晚上）」などの遅い時間帯に関する単語の関心度の予測値が低い。この背景には、中国大陸観光客が好む旅程と密接な関係があると考えられる。

具体的には、中国のアウトバウンド観光紹介のサイトで書かれた箱根地域のお薦めコースは、ほとんどのサイトで日帰りツアーを推奨している。加えて中国大陸観光客の旅行中の移動はほぼ電車やバスなどの公共交通機関であるため、観光地周辺の都市に戻って宿泊するには、観光地との往復に十分な時間が必要となるため、夕方や夜間は中国大陸観光客が観光に充てる時間帯にはならない。中国大陸観光客に中には箱根に泊まり、温泉など宿泊施設での体験をしながら美味しい食事を楽しむことが目的の観光客もいるがそれは一部である

ため、観光関連事業者は昼間の活働に合わせて旅程を組むことで、より効率的に中国大陸観光客の満足度を高めることができる。

5.3.5.2　傾向分析の結果考察

前節 5.3.4 で述べた通り、最終的に手法 1 と手法 3 の組み合わせで結果について傾向分析の結果考察を行った結果を本節で考察する。なお、本事例に使われるデータの各組み合わせのレビュー数は比較的少なかったため、まず同じ年の「予測組み合わせ」と「対照組み合わせ」だけを考慮する（手法 1）のみで傾向分析を行い、下記の表 5-5 に示されるような最終的な結果が得られた。

表 5-5　手法 1 の予測結果

	A-単語	予測合計
1	子供	32.67721
2	レストラン	11.57792
3	雰囲気	10.33581
4	カフェ	9.778363
5	土産	8.887407
6	お茶	7.011596
7	家族	6.969111
8	遊歩道	5.634068
9	女性	4.933328
10	眺め	4.868614
11	紫陽花	4.578681
12	夏	3.831681
13	小学生	3.729514
14	ラリック	3.487293
15	オリエント	3.33587
16	池	3.300365
17	クリスマス	3.183304
18	車	2.99211
19	ランチ	2.861856
20	途中	2.821068
21	コンサート	2.732642
22	平日	2.65282
23	夫婦	2.648328
24	カップル	2.625163
25	雪	2.621891
26	アスレチック	2.544003
27	歌	2.482346
28	階段	2.439566
29	イルミネーション	2.370025
30	トイレ	2.187426
31	ススキ	2.125173
32	スタッフ	2.116035
33	仮面	1.982953
34	シャボン	1.979879
35	娘	1.975782
36	傘	1.819983
37	空気	1.780915
38	ペピーカー	1.731242
39	夜	1.722684
40	外国	1.458666
41	コーナー	1.313065
42	ジャム	1.273219
43	昔	1.191532
44	冬	1.094774
45	パスタ	0.672901
46	夕方	0.583902

	B-単語	予測合計
1	子供	25.79987
2	レストラン	11.30323
3	雰囲気	9.663537
4	カフェ	8.089125
5	遊歩道	7.311572
6	家族	5.423671
7	コンサート	5.438459
8	土産	4.074659
9	眺め	3.910052
10	平日	3.827166
11	夏	3.745844
12	車	3.247416
13	ランチ	3.210701
14	娘	2.876575
15	イルミネーション	2.563248
16	外国	2.224574
17	スタッフ	2.19875
18	女性	2.187397
19	途中	2.089434
20	お茶	2.076914
21	カップル	2.058206
22	ラリック	1.919525
23	小学生	1.918961
24	冬	1.917466
25	ジャム	1.852237
26	シャボン	1.66131
27	夫婦	1.660398
28	雪	1.658773
29	ペピーカー	1.65388
30	紫陽花	1.604247
31	コーナー	1.564108
32	トイレ	1.555244
33	夜	1.507705
34	アスレチック	1.489773
35	ススキ	1.431365
36	クリスマス	1.221984
37	空気	1.156739
38	階段	1.145539
39	仮面	1.128271
40	オリエント	1.10513
41	傘	1.100034
42	池	1.096747
43	パスタ	0.994784
44	夕方	0.921322
45	歌	0.853774
46	昔	0.85049

	C-単語	予測合計
1	子供	59.16316
2	レストラン	19.261
3	雰囲気	15.69316
4	カフェ	11.71329
5	眺め	10.87216
6	遊歩道	10.53916
7	家族	10.09841
8	土産	7.086453
9	クリスマス	6.477949
10	カップル	5.414361
11	ラリック	5.399231
12	ランチ	5.149047
13	アスレチック	5.080659
14	コーナー	4.78376
15	平日	4.448727
16	夏	4.353333
17	外国	4.30409
18	イルミネーション	4.091057
19	空気	4.010172
20	コンサート	3.845224
21	女性	3.722585
22	階段	3.52241
23	ペピーカー	3.079126
24	途中	3.073942
25	シャボン	3.0441
26	お茶	3.013742
27	オリエント	2.946687
28	冬	2.865567
29	池	2.852111
30	昔	2.813725
31	パスタ	2.79699
32	ジャム	2.78938
33	仮面	2.666975
34	歌	2.59163
35	紫陽花	2.516055
36	雪	2.471599
37	ススキ	2.342584
38	トイレ	2.322726
39	夜	2.111885
40	スタッフ	1.960211
41	車	1.936634
42	小学生	1.797381
43	娘	1.73328
44	夫婦	1.722398
45	夕方	1.54963
46	傘	0.849779

予測結果の点数が高いことは関心度が高いことを示し、それを代表する赤色の背景の予測結果の値は差が大きく比較的区別しやすいが、一方で、中国大陸観光客があまり関心を持っていない、関心度の点数が低い青色の背景の予測結果の値は相互にあまり大きな差がなかった。これは、特定の個別レビューが全体の点数に影響を与え、結果に誤差が生じさせてしまったのではないかと考えられる。

そのため、手法 3 を用いた結果についても並行して考察を行うため、前節 5.3.4 で述べたように、同じ年の結果に 2 倍の加重をかけて実施した。その結果は以下の通りである。

表 5-6　手法 3 の予測結果

	A-単語	予測合計
1	子供	66.59734
2	レストラン	23.26278
3	雰囲気	20.20734
4	カフェ	17.75929
5	土産	14.6166
6	家族	13.83857
7	遊歩道	12.46598
8	眺め	10.90594
9	お茶	9.595135
10	女性	8.064959
11	夏	8.009866
12	紫陽花	7.857505
13	コンサート	7.615835
14	平日	6.64084
15	ランチ	6.472151
16	クリスマス	6.314137
17	ラリック	6.047409
18	車	5.832886
19	小学生	5.5722
20	カップル	5.533541
21	池	5.280748
22	イルミネーション	5.255006
23	オリエント	5.107465
24	途中	5.096942
25	アスレチック	4.913436
26	階段	4.620612
27	外国	4.525779
28	夫婦	4.39812
29	娘	4.396074
30	雪	4.380772
31	歌	4.257596
32	トイレ	4.21724
33	ペピーカー	4.158098
34	シャボン	4.109488
35	ススキ	4.002455
36	仮面	3.856667
37	空気	3.81443
38	スタッフ	3.776713
39	コーナー	3.711765
40	傘	3.61917
41	冬	3.501027
42	ジャム	3.402253
43	夜	3.326881
44	パスタ	2.504122
45	苔	2.293471
46	夕方	2.010989

	B-単語	予測合計
1	子供	51.9793
2	レストラン	19.42857
3	雰囲気	17.49089
4	カフェ	14.88795
5	遊歩道	11.95059
6	家族	11.24928
7	土産	8.585694
8	眺め	8.326252
9	コンサート	8.081065
10	お茶	5.952814
11	夏	5.94951
12	平日	5.74567
13	ランチ	5.00854
14	車	4.896856
15	ラリック	4.639293
16	女性	4.601202
17	クリスマス	4.132343
18	カップル	3.931046
19	娘	3.858906
20	イルミネーション	3.794933
21	外国	3.633962
22	途中	3.554955
23	アスレチック	3.250753
24	紫陽花	3.237334
25	小学生	3.22841
26	スタッフ	3.201385
27	雪	3.138636
28	オリエント	3.107573
29	階段	2.954438
30	冬	2.851077
31	ペピーカー	2.848596
32	コーナー	2.827432
33	夫婦	2.750472
34	ジャム	2.750191
35	仮面	2.657971
36	シャボン	2.605127
37	空気	2.603946
38	トイレ	2.536967
39	池	2.361133
40	ススキ	2.349725
41	夜	2.127936
42	歌	2.006006
43	傘	1.733419
44	パスタ	1.650296
45	苔	1.565374
46	夕方	1.50626

	C-単語	予測合計
1	子供	126.2856
2	レストラン	39.57641
3	雰囲気	35.16388
4	カフェ	28.17816
5	遊歩道	25.98013
6	家族	23.3354
7	眺め	17.54075
8	土産	17.22322
9	ラリック	11.45465
10	コンサート	11.11791
11	お茶	10.5812
12	女性	10.27762
13	夏	10.16407
14	平日	10.00574
15	クリスマス	9.62408
16	ランチ	9.435557
17	カップル	8.894471
18	アスレチック	8.5113
19	コーナー	8.455425
20	車	8.020443
21	イルミネーション	7.320446
22	外国	7.256691
23	オリエント	6.91958
24	空気	6.801118
25	小学生	6.733589
26	紫陽花	6.716803
27	階段	6.554263
28	ペピーカー	6.309085
29	仮面	6.292774
30	途中	6.17641
31	娘	6.101436
32	池	5.980069
33	シャボン	5.707053
34	夫婦	5.617057
35	雪	5.577541
36	冬	5.452751
37	スタッフ	5.089322
38	ジャム	4.982429
39	歌	4.814938
40	トイレ	4.486982
41	苔	4.386764
42	ススキ	4.219404
43	夜	4.080079
44	パスタ	3.693925
45	傘	3.515186
46	夕方	2.997563

この手法 3 を用いた予測結果においては関心度が低いことを示す青色の背景の予測結果

の値は差が多くなり区別を比較的簡単に行うことができ、手法 3 は手法 1 より関心度が低い単語に対する関心度の値の範囲が比較的広く、優れているといえる。ただし、逆に関心度が高い部分については手法 1 に比べると単語の関心度の順位に変化がなく、時間経過に伴う中国大陸観光客の関心度の高い部分の変化傾向を示すことは難しい。

したがって、関心度の高い単語について有効な予測結果が得られる手法 1 と、関心度の低い単語について有効な予測結果が得られる手法 3 を用いた分析から得られた 2 つの結果を組み合わせることで、中国大陸観光客の関心度の変化傾向を網羅的に説明することができると考えられる。

また、全体的には、3 つの時間区分における中国大陸観光客の関心度に関する予測のは良好な一貫性を保っており、特に関心度に関する予測の点数が高い単語について、8 つの単語が 3 年間共通して上位 10 位以内を維持している。これは、箱根にある美術館は、展示内容の変動が少ないため、これは、箱根にある美術館は、展示内容の変動が少ないが、逆に観光客はそれをいつも同じ安定した関心度でとらえていると考えられる。

しかし、細部にはいくつかの変化があり、時間区分の 2018 年 4 月から 2019 年 3 月を対象とした A 組に関する分析結果について、さらに考察を展開すると、以下のような変化が確認された。

1) 「女性（女性)」という単語の関心度の予測が 3 年連続で高まった。関心度の高さが C 組（時間区分:2016.4~2017.3）の 21 位から B 組（時間区分:2017.4~2018.3）の 18 位、さらに A 組では 9 位に上り、女性を対象とした展示内容や観光商品・サービスが、より中国大陸観光客に好まれることが明らかになった。このような潜在的な関心度の高まりは、新たな観光消費者のニーズを満たすためには、性別に応じた個人サービス（特に女性向け）を提供することが効果的になりつつあることを示している。

2) 「お茶（茶)」の関心度の予測が時間の経過と共に上昇し、関心度の順位が C 組の下位半分から B 組の 20 位、さらに A 組では 6 位に上った。このことから、お茶を、単純に飲み物と捉える場合でも、お土産として位置づけ販売する場合でも、いずれにしても観光事業者は提供の仕方を重視すべきであることを示している。例えば、お茶関連の商品には中国語の説明を付け加えたり、広報や宣伝も強化することは、中国大陸観光客の満足度にプラスの効果をもたらすと考えられる。

3) 「夏 (夏)」という単語は関心度の高さにおいて、比較的高い順位を常に保っているが、「紫陽花（紫阳花)」は A 組だけで潜在的に関心のある高点数の単語になった。これと同様に、「池（池子)」という単語も A 組における点数が他の 2 つ時間区分 C、B より上回っている。このことから、屋外景観にかかわる要素は近年、中国大陸観光客の関心を集めつつあることが明らかになり、この部分を強化することも中国大陸観光客に影響を及ぼす重要な要素になっていることが認識された。

4) 「小学生（小学生)」と「ベビーカー（婴儿车)」との関心度における順位の逆転が確認された。具体的には、「ベビーカー（婴儿车)」は C 組における順位がまだ上位にあっ

たのが、A 組になると下から 9 番目に下がっている。一方、「小学生（小学生）」は C 組の下から 5 番目から A 組の 13 位に上っている。これは観光客属性の変化をある程度示しており、今後、中国大陸観光客の構成の多くが、乳幼児を連れた家族構成から、ベビーカーなどは不要なもう少し大きい子供を持つ家族構成に変わっていき、観光関連事業者はこの潜在的な傾向の変化に合わせたサービスを提供しなければならないことが認識できる。

5) それぞれの組み合わせの関心度の下位半分に注目すると、手法 3 で得られた予測結果では、「冬（冬）」と「雪（雪）」については A 組と C 組ではほぼ同様の順位であったが、B 組での順位は少々上っている。これは、その年の冬の天気に直接関係していると推測されたため、A、B、C の 3 つの期間の箱根の冬の気候を調べ、箱根に最も近い気温観測点である小田原における 12 月、1 月と 2 月の 3 か月の月間平均気温データを調査し考察を行った。結果は以下の図 5-6 と表 5-7 に示している。

図 5-6　箱根周辺の気温観測所

（国土交通省の気象庁、Web サイト：
https://www.data.jma.go.jp/obd/stats/etrn/select/prefecture.php?prec_no=46&block_no=0390&year=2016&month=&day=&view=より）

表 5-7　小田原における冬季の気温（2016 年~2019 年）

（出典：国土交通省の気象庁）

期間	12月の日平均気温（℃）	1月の日平均気温（℃）	2月の日平均気温（℃）
2018.4～2019.3	8.2	5.1	7.3
2017.4～2018.3	6.1	4.4	5.1
2016.4～2017.3	8.6	5.2	6.4

表 5-7 で示した通り、B の期間の平均気温は A と C の期間に比べて 2 度ほど低くなっており、A と C の冬は B の冬よりも暖冬であったことが分かる。したがって、気温が低いほど、中国大陸観光客が感じる冬の特徴は他の期間に比べてより顕著にあらわれることから、冬や雪への関心度が高くなったと考えられる。このことは、データ予測結果の妥当性を論理的に証明していると考えられる。

5.4　現場業務への反映を容易にするための未知情報の適用方法

中国大陸からの観光客の既知部分に対する分析結果は、第 3 章の分析結果で既に示した。その分析結果を得るために可視化処理を試み、平面図(図 3-6、3-7、3-8)と共起ネットワーク図(図 3-9)で、観光客のレビューの内容や構造等を効率的かつ直観的に表すことができた。

一方、本研究の中心となる議論は未知部分を予測することであり、「予測組み合わせ」に対する分析結果は上記のような直観的な構造図を構成し可視化することはできない。そのため、どのようにして、分析結果を観光事業者や自治体の観光行政部門における現場業務に容易に反映することができるのか検討することが非常に重要な議論となる。

本研究では、第 4 章と本章の事例検討の部分で分析対象としたのは、いずれも未知の予測情報であり、これらの情報に関する考察は、個別の事例分析を通じて様々な特性や予測を見い出しし、それに基づいた提案を行いより模式化された方法も開発した。このような手法や枠組みを観光事業者や関連する自治体の観光行政部門の現場業務に容易に反映するためには、未知の情報を適切に分類するための標準化された処理方法を確立することが大変重要だと考える。

その際、用いられる分類法については、具体的に次の 3 つの側面が考慮されるべきである。

1) 予測分析における分類法に関連する有効な既往研究や手法がないため、本研究で得られた事例を通じた予測結果は既存の唯一の実用的な参照可能なデータであること。
2) 観光レビューの全般的な分析について先行研究で提示されたいくつの分類法は、主に観光消費者の単純に「良い」「悪い」といったポジティブかネガティブかの感情表現しか考慮されていないが、本論文の第 3 章で検討したそれ以外の文脈的な情報に関連する側面の区分も分類の参照とする必要があること。
3) 観光産業に関する特性、いわゆる食事、宿泊、移動、観光、ショッピング、娯楽といったものは基本的な要素であるが、本研究における観光レビュー分析の枠組みを通じて、これらの要素にどのような重点配分で経営資源を投下し観光商品・サービスの開発や観光行政の立案をするのかが非常に重要であること。

以上の観点から、本研究は未知情報の現場業務への反映分野を以下の 5 つに大別し、観光事業者や自治体の観光行政部門が予測分析の結果の要点を容易に把握し、本研究の結果を基にさらなる深い理解を進め、現場業務に活用する方法を提案する。

1) 観光商品・サービス企画

観光商品・サービスに関わる単語は、直接販売された観光商品そのものに加えて、テーマ、ブランド、や観光商品に合わせて販売された観光サービスに係る項目の単語も参考にすることで、ビジネスに直接活用できる可能性が最も高い。このような観光商品・サービスに関

する予測値は、中国大陸の観光客に優先的に推奨する商品・サービスの提供方法や、商品をより目立つような配置を検討する際に観光事業者にとって有益となる。また、予測結果で関心度が高い商品に関わる単語は、今後、中国大陸の観光消費者に好まれる可能性が高いことを意味する。例えば、代表的な単語としては、「コーヒー」「デザート」「茶」などが挙げられる。

2)　環境・施設・サービスの整備

環境、施設やサービスに関する単語は、主に観光レビューの対象となる投稿者が主観的に求める付加価値に関するものであり、多くの場合、実際の観光利用者が投稿する評価対象の中でも一般的な要素である。環境、施設やサービスを中心に、ハードウェア整備の観点から見ると、迅速に大きな改築や修正を行うことは容易ではないが、長期的なメンテナンスや追加開発の観点から、関連する単語への関心度の増加が明らかであれば、その要素をさらに深く理解しその後の変化をモニタリングしつつ、実際に中国大陸観光客がどのような環境や施設に関心を持ち得るのかを観察することで、より実効性の高い対応策を打つことができる。

一方、予測分析において出現頻度の低い単語でも、提供側が観光客に供給したいとサービスと考えるのであれば、適切な改善を行うことで、中国大陸観光客により多くの関心を持たせ、より高い満足度を持ってもらえることも考えられる。例えば、代表的な単語としては、「冷蔵庫」「トイレ」「階段」などが挙げられる。

3)　観光地外の地域との連携

観光地外に関わる単語は、観光レビュー対象となる主体の範囲外の要素についてであり、宿泊施設や観光地そのもの以外の観光に関わる記述である。しかし、このような単語は、宿泊施設や観光地と外部とのつながりを表しており、観光地をより広く捉えた観点からのコメントである。

そのため、本来的にはその対象となる地域や近隣観光地などが主体となって検討すべき内容である。しかし、観光地もこのような地域外への関心を把握することで、観光者が必要とする送迎サービスや交通手段に関わる情報を提供することができる。また、観光地全体の発展状況をより広い範囲から把握することができるため、この観点からすると地元自治体の観光行政部門にとってより広範囲で行政立案ができるといった観点からも重要な意義をもっている。例えば、代表的な単語として、「運転」「途中」「遊歩道」などが挙げられる。

4)　観光対象となる時間、季節、気候の配慮

観光に関する時間、季節、気候に関する単語は、中国大陸観光客の関心が時間、季節、気候などによって変化することを示したものであり、本研究の第 3 章と第 5 章でも傾向変化が与える影響を分析したように、観光客は時間とともに異なる体験をしたいというニーズ

があること示している。また、これらの単語の分析結果は、観光客の時間配分をどのように計画すれば効果的なのかを関係者に示すことにも活用できる。

例としては、中国大陸観光客が夜間や夕方などの単語にあまり関心を持たないため、中国語を話せるスタッフをその時間帯にもより多く配置して、その時間帯の観光を充実させることも検討することができる。また、これらの単語で異なる気候による観光客の関心の変化にも注意を払うことができ、このような情報を把握することで観光地全体の観光事業者は季節により経営資源をどのように配分すべきかを効率的に判断することができる。代表的な単語として、「夜」「冬」「クリスマス」などが挙げられる。

5) 観光客の構成

観光客の構成に関わる単語は、レビューの投稿者自分から、同伴者や出会った他の観光客などにも広範囲に含んでいる。観光レビューには観光客の身分・属性などについても述べられていることが多いが、観光のパターンを例に考えてみると、親との観光は、夫婦観光や親子観光とは異なるニーズがあり、この部分にも注目することで観光方式の変化を追跡・分析することができるため、観光事業者にとってより観光客のニーズにマッチした商品・サービスを提案することができる。代表的な単語としては、「夫婦」「親」「子ども」「小学生」など挙げられる。

以上のような 5 つの分野への導入を行えば、観光事業者又や自治体の観光行政部門はそれぞれの状況から適切な分析結果や方法を選択して、現場業務に反映することが可能になると考える。

第 6 章　全体考察

6.1　結論

6.1.1　従来の中国大陸アウトバウンド観光産業に関する研究との比較

本研究は現存するオンライン観光レビューに欠損している情報の補完をすることで、より付加価値の高い需要予測が可能になるという点で新規性があり、インターネット通信技術とデータ分析・予測技術とを融合し、従来の観光産業の経営やサービスの効率を改善し、観光客の満足度を高め、その結果として観光産業の発展をより効率よく促進させる新しいモデルを提供した。そのモデルは、主に観光産業のマクロ面に着目していた従来の研究に対して以下に示す優位性を持つと考える。

1)　観光客ニーズの尊重

本手法の分析対象となるデータは、よりオープンな公開情報から取得しており、ヒアリングやアンケートを用いた調査方法などのような調査者の設計による影響がない。

2)　観光産業の特性への考慮

観光客のオンラインレビューの傾向を分析・把握することで、1978 年の改革開放以来の中国大陸のアウトバウンド観光産業における急速な発展と大きな変化における特性と、観光業界で急速に発展する「OTA」を利用した観光消費者の特性も明確に反映できる。

3)　観光業界の今後の発展に関する予見

本研究の目的は将来の観光サービスに関わる予測と方向性を提供することであり、業界の現状を把握するのに役立つだけでなく、今後の観光業界の発展に求められる方向性を提供することができる。

4)　研究の継続性

本研究は従来の観光産業研究と比べると、データの収集から、分析、検討、考察までの一連のプロセスを含んでいる点で実効性が高く、また使用するデータも最新のものをインターネット経由で無料で収集できることから、持続可能な研究の枠組みと言える。今後、PC の普及や性能や演算能力の向上に伴い、ますます多くの中国大陸からのインバウンド観光客が「OTA」レビューのプラットフォームに投稿・参照するなど関与度合を高めることで、

得られる情報も更に豊富になり、それによって観光客の需要の把握も更に正確になる。

6.1.2　事例サンプルの分析結果のまとめ

1)　本研究で用いた手法の他地域への適用性

第 4 章と第 5 章における訪日インバウンド観光への分析は箱根を事例として分析を進めたが、本研究の手法は箱根観光エリアに固有の特性に左右されるものではなく、そのまま別の観光地にも適用することが可能である。特に、第 5 章の最後で提案した結果を分類検討するための処理方法は、分析処理に必要なデータの収集からスクリーニング、最後の結果の分類化と分析の考察までの一連の流れを模式化して行うことができ、観光事業者や自治体の観光行政部門における分析の効率を向上させることができる。また、中国大陸観光客に既に人気のインバウンド観光地では既存の観光のレビュー数も十分であり、今後も更に増加することが見込まれるが、「OTA」業界の発展に伴い、他の多くの観光地でも中国大陸観光客のレビューがますます普及していくと考えられる。したがって、本研究で提案した分析手法もより多くの観光エリアで適用することが可能になる。

2)　中国大陸観光客の特性

本研究の第 3 章と第 5 章の傾向の変化に関する分析結果は、中国大陸における観光産業の特性と様々な年代の中国大陸からのインバウンド観光客の特性に関する分析と考察を通じて、時間の変化が観光客の需要に大きな影響を及ぼすことを明らかにした。

時間変化に伴う影響は観光産業自身の傾向の変化から生まれるが、一方で中国大陸観光客層の構成の変化とも関わっており、団体ツアーと個人旅行の割合の変化、初回海外観光客とリピーターの割合の変化などが関連している。また、第 4 章と第 5 章の事例データ収集による「予測組み合わせ」と「対照組み合わせ」での分析の結果で得られた頻出語の両者間の大きな差異（「予測組み合わせ」の頻出語のうち「対照組み合わせ」でも頻出語なのは半数前後）から、中国大陸観光客と地元観光客の間には明らかな相違があるという現状が確認できた。

3)　レビューの対象の違いによるデータの有効性の差異

本研究では「OTA」の Web サイトにおける観光ビューにおける対象の異なるタイプのレビューを利用してデータ分析を試みた。第 3 章と第 4 章で展開した分析では宿泊施設を対象としたレビューデータ、第 5 章では観光スポット類を対象としたレビューデータを用いた。両者とも予測面ではサンプル地域の観光に対して有意義な結果を示しているが、下記の点で観光スポット類を対象としたデータより宿泊施設を対象としたデータの方が可用性、

信頼性、付加価値性観点から優れていると考える。

まず、観光客の関与程度から見ると、宿泊施設のレビューは観光スポット類よりはるかに高く、収集可能なデータ数も多い。次に、データの信頼性から見ると、宿泊施設のレビューは「OTA」企業によって厳格に規制されており、投稿者が実際の利用者でなければならないことに対して、観光スポット類のレビューのほとんどはオンラインユーザーの自発的なメッセージである。そして、中国大陸観光客の海外におけるレビューのデータ量から見ると、宿泊施設を対象としたレビューは情報量が十分であり、時間、観光の種類さらに観光事業者からの返信メッセージなど付属の情報も同時に収集することができるため、内容的にも豊富であり、今後研究の対象を広げる際にも利用しやすい。

4)　予測分析の評価

中国大陸観光客のインバウンド観光における需要予測の分析は本研究の中心であり、第 4 章と第 5 章ではこの分析を行い、中国大陸観光客がインバウンド観光する際の情報の非対称性、つまり中国大陸からの訪日客がインバウンド観光情報を完全に把握していないという特性に着目した分析と考察を展開した。同時に、観光客は地元の文化をより多く知りたいという需要もあるため、適切な「対照組み合わせ」を選択することで中国大陸観光客の潜在的な需要を明確にし、これらの「対照組み合わせ」によって観光全体に関する情報をより良く網羅することができるため、中国大陸観光客の需要をさらに満たすことができる。

また、事例分析の結果から中国大陸観光客の観光目的と嗜好など情報が得られ、予測結果を「観光商品・サービス企画」「環境・施設・サービスの整備」「観光地外の地域との連携」「観光対象となる時間、季節、気候の配慮」「観光客の構成」の 5 種類に分類・考察して、観光事業者が中国大陸観光客のまだ認知していない要素の中で、どれを中国大陸観光客に優先的に推奨するべきか、どのようなサービス側面が重視されるべきかを理解し、把握するのに有効であることが認められた。このような手法を用いることで、観光産業においても限られた経営資源を合理的に配分を行うことにより、中国大陸観光客の満足度をより効果的に高めることができる。

5)　傾向分析の評価

第 3 章と第 5 章では観光レビューの傾向に関する分析を行ったが、傾向の時間区分と分析処理の手法によって差異が生じた。第 3 章では閑散期と繁忙期を考慮し、第 5 章では時間区分をより広げ、1 年を単位とした。分析処理の手法に関しては、第 3 章は中国国内の観光地を対象とする分析であるため、単純にレビューの変化を考慮するのみであったのに対して、第 5 章では、予測手法を用いることで訪日インバウンド観光レビューの年による傾向変化についての考察を行った。特に、第 3 章での傾向変化分析結果から、観光客のレビューは閑散期と繁忙期から受ける影響が非常に強いことが示された。しかし、異なる観光地において本手法を適用する場合、年単位で区分する手法は最も簡単、有効かつこれまでの研究

や統計でも用いられてきた手法でもあり、且つ、実際に閑散期と繁忙期の状況を把握できないなどの理由も考慮し、予測分析するためには、年単位で考察する方がより有効であると結論づけた。

6.1.3　結語

本研究はまず第 1 章で、現代の観光産業と中国大陸からのアウトバウンド観光の発展の背景となっている、中国大陸で急成長している「OTA」業界を対象とし、訪問先の観光情報と特性に合わせた経営資源配分と観光商品・サービスの最適化を通じて、中国大陸観光客の満足度を向上させ、人々の文化交流と融合を促進し、受け入れ現地の観光産業を振興する目的を示した。

第 2 章では、本研究に関連する先行研究を、「中国大陸からのアウトバウンド観光に関する研究」、「オンライン旅行代理店業に関する研究」、「観光事業の情報化に関する研究」、「日本で観光するインバウンド中国大陸観光客に関する研究」の 4 種類に研究対象で分別し、それぞれについての考察を行った。

第 3 章では、観光レビューが持つデータとしての特性及びその特性が生じる要因について分析した上で、宿泊施設に対する中国大陸観光客の顧客レビューを対象として分析・考察を行った。その過程において、観光レビューのデータ分析手法としてテキストマイニング処理のモデル手法の有効性を示した。そして、傾向変化の分析結果から、観光客のレビューは時間区分から受ける影響が非常に強いことを明らかにした。

第 4 章では、インバウンド観光のオンラインレビューを分析する際には、国内観光と国際観光との差異を十分に考慮しなければならないことから、中国大陸観光客がインバウンド観光する際の情報の非対称性、つまり中国大陸からの観光客がインバウンド観光情報を完全に把握していないという特性に着目した分析と考察を行った。そして、その結果を用いて観光産業で限られた経営資源を合理的に配分を行うことにより、中国大陸観光客の満足度をより効果的に高めることができることを示した。

第 5 章では、第 3 章・4 章の内容を統合した上で更なる考察を展開した。具体的には、第 3 章で考察した傾向分析に加えて第 4 章で考察した予測分析を活用し、分析対象とする観光レビューの種類の範囲を拡大するため、相関分析を行う対象データを宿泊施設から観光地・スポットに拡張し分析・考察を展開した。そして、その結果に基づき、研究方法の適用性をさらに検証し、未知情報を活用して、観光業界の現場業務への反映を容易にするため適用方法を提示した。

このように、本論文では第 2 章から 5 章を通じて、中国大陸のアウトバウンド観光の急速な発展と「OTA」業界の観光産業における台頭及び「OTA」Web サイトに投稿される観

光レビューの成長を背景として、その有効な分析と予測の方法の検討と考察を展開した。また、テキストマイニングを行う際、中国大陸観光客のインバウンド観光レビューの 2 つの主要な特性である「観光産業発展に伴う観光ニーズの急速な変化」と「未知情報の欠損があること」とを結合して実施した。そして、この 2 つの特性を考慮し、中国大陸観光客の潜在的な観光目的と嗜好を予測することで、更に観光事業における経営資源を合理的に効率よく配分することができることを示した。

その上で、サービス品質の向上を行い、中国大陸観光客のより良い満足度を実現し、訪問先の観光産業を発展させるため、本研究ではテキストデータに基づく有効な一連の分析手法を最後に提示した。具体的には、まず分析処理に必要なデータを収集し、次に、分析サポートの「対照組み合わせ」を選別、そして、データの前処理、時間分区とコア単語を選び、更に、データから共起行列に変換、その後、処理の共起行列が不完全な特性を着目して補完作業の手法行列分解して、未知情報の予測を行うという流れである。この分析手法によって得られた予測結果から、観光事業者や自治体の観光行政部門はニーズに合わせた情報を選択することが可能となった。

以上から、本研究の成果は、中国大陸観光客のインバウンド観光の今後の発展に大きく寄与することができると確信する。

6.2　本研究の限界と今後の課題

6.2.1　研究範囲の拡張

これまでの研究で扱われていた事例の範囲は主に観光地、つまり観光資源が集中している観光エリアであった。一方、本研究では単一観光地に焦点を当てすることで、より高い精度で中国大陸観光客の観光目的と嗜好傾向を予測できるため、観光地の観光産業における情報収集や将来計画、経営資源の配分と投入、政策の立案と天気等により具体的に役立つと考えられる。ただし、より広い地域の観光産業や全国的な観光政策に適用する場合は、さらに広範囲に渡るデータを収集・分析する必要がある。

この主な要因は、本研究で用いた観光レビュー文章の制約であり、レビュー投稿者の言及した内容は主に個別具体的な点で、ほとんどが単一の観光商品または小さな範囲内の観光産業を対象としていたためである。

したがって、今後は以下の 3 点に留意して研究を進める必要がある。

1)　より幅広い宿泊施設と観光スポットを対象とする

分析サンプルの抽出範囲を改善する。例えば、第 3 章で扱ったサンプルは全て旅館を対象としたレビューであり、得られた分析結果自体は有効であったが、後続の研究ではより幅広い宿泊施設（ホテル、民宿、ユースホステルなど）や観光スポットのレビューを対象として更に深く分析を行いながら、分析方法の適応性を確認する必要がある。

2)　より地理的に広い範囲に広げる

特定の国や地域における複数の代表的な観光エリアの観光レビューのデータを収集し、その観光需要の予測結果と傾向分析の結果を統合し、より広範囲の地域にまたがる観光産業戦略や政府・自治体の政策への提言を試みる。

3)　他地域からの観光客も対象とする

本研究を通じて収集・予測分析を行った観光レビューは中国大陸観光客から投稿されたものが中心であった。より深くインバウンド観光客の現状・特性・潜在ニーズを把握するためには、他の地域から観光客の観光レビューも同じ分析方法を試みながら統合する必要がある。

6.2.2　現場業務への適用

本研究では観光地に対する実際の観光レビューデータを用いて様々な考察を展開したが、

今後は観光産業の現場業務に実際に適用して改善を行いながら継続的に観察することで、実際の効果を検証する必要がある。そのため、以下の二者と協力し研究の実効性をより高める必要がある。

1) 「OTA」企業と協力し、観光地への初回訪問者やリピーターといった観光客の付属情報を研究対象として追加する。観光客の詳しいプロファイル情報を更に深く理解することで、より詳しい分析・考察を行うことが可能となり、その結果、「OTA」企業も観光客のニーズに合致した満足度の高いサービスを提供することに活用することができる。
2) 観光地における官民の関係機関と協力し、データ分析により得られた結果を観光行政や観光事業者の経営者自身の経験とも統合、連携して改善のための対策を提案・実行し、観光客の満足度の向上効果を協業しながら継続的に観察する。また、観光産業の発展を支援する観点から本研究で提案した手法の改善点を継続的に抽出、検証、評価し、最適化するために実行する。そして、本研究の前節 5.4 で試みた分類による分析結果の検討は観光事業者や自治体の観光行政部門の現場業務における具体的な取り組みの基礎となるはずである。

謝辞

本論文を作成するにあたり、数多くの方々の御支援、御協力を頂きました。心から感謝申し上げます。

指導教員である名古屋工業大学大学院社会工学専攻　渡辺研司教授と青山友美助教に感謝いたします。渡辺教授には、ゼミ発表で御意見を頂き、研究を勧めていく上での貴重な示唆や、熱意のこもった御指導を承りました。また、論文執筆、ジャーナル投稿では、論文や発表資料の構成、内容について御意見を頂き、研究内容を正確にわかりやすく人に伝えることの難しさや重要性を習得させていただきました。研究活動以外の留学生活においてもお心添え頂きました。ここに感謝の意を表します。

副査の中出康一教授、荒川雅裕教授には、博士論文執筆で貴重な御指摘、御助言をいただきました。ここに感謝の意を表します。

また、越島一郎教授には、ご定年まで研究での貴重な御指摘、ゼミ発表等をご教授頂き、ここに感謝の意を表します。

筑波大学の永井明彦先生には、最も困難な留学の最初段階に、大変お世話になり、ここに感謝の意を表します。

渡辺・青山研究室の青木翔吾君、山田脩嗣君、桑田健壮君、望月暁史君、黒柳利之君、佐藤篤至君、土屋友彦君、山田和樹君に同じ研究室の仲間として何度も研究や生活や日本語も助けられました。そして、渡辺・青山研究室で有意義な楽しい時間を共に過ごすことが出来ましたことに心よりお礼を申し上げます。

名古屋工業大学大学院の後輩の唐子楗君には、日本語や生活に御協力いたあき、ここに感謝の意を表します。

四日市市の澤田克彦さん、澤田啓子さん及び御家族には、長い日本の生活で、家族のように大変お世話になりました。ここに感謝の意を表します。

最後に自分の日々の生活を支えてくれた両親、張岩さんと余燕さん、そして婚約者の陳楽君さんに心より深く感謝致します。愛しています。

参考文献

[1] 宋 瑞：Report on World Tourism Economy Trends 2018 （世界旅游经济趋势报告 2018），世界旅游城市联合会；中国社会科学院旅游研究中心，（2018）

[2] 宋 宇， 宋 瑞， 李宝春等：Report on World Tourism Economy Trends 2019 （世界旅游经济趋势报告 2019）， 世界旅游城市联合会；中国社会科学院旅游研究中心，（2019）

[3] THE WORLD BANK：GDP growth. 2005~2018 https://data.worldbank.org.cn/indicator/ny.gdp.mktp.kd.zg?end=2018&start=2005

[4] International Monetary Fund： Real GDP growth of world. 2005~2019 https://www.imf.org/external/datamapper/NGDP_RPCH@WEO/OEMDC/ADVEC/WEOWORLD

[5] International Monetary Fund： Real GDP growth of Japan. 2013~2018 https://www.imf.org/external/datamapper/NGDP_RPCH@WEO/OEMDC/ADVEC/WEOWORLD/JPN

[6] Japan Tourism Agency 日本国土交通省観光庁：訪日外国人の消費動向―訪日外国人消費動向調査結果及び分析 平成 24 年~平成 30 年.（2012~2018）

[7] 中国旅游研究院， 携程旅游大数据联合实验室：2018 年中国游客出境游大数据报告.（2018）

[8] 中国国家统计局：中国统计年鉴. http://www.stats.gov.cn/tjsj/ndsj/

[9] 中国国家统计局：2018 年国民经济和社会发展统计公报. http://www.stats.gov.cn/tjsj/zxfb/201902/t20190228_1651265.html

[10] 中国公安部：国务院关于出境入境管理法执行情况的报告。 （2016）http://www.npc.gov.cn/zgrdw/npc/cwhhy/12jcwh/2016-11/07/content_2001415.htm

[11] UNWTO World Tourism Organization： International Tourism Highlights 2019 Edition.

[12] Nielsen: OUTBOUND CHINESE TOURISM AND CONSUMPTION TRENDS 2017 Survey.

[13] UNWTO World Tourism Organization：World Tourism Barometer 2017. (2018)

[14] UNWTO World Tourism Organization：World Tourism Barometer 2018. (2019)

[15] Worldwide Digital Travel Sales: eMarketer's Estimates for 2016–2021. (2017)

[16] 艾瑞咨询: 2017 年中国在线旅游度假行业研究报告.（2017）

[17] 艾瑞咨询: 2018 年中国在线旅游度假行业研究报告.（2018）

[18]Margaret Ady, Donna Quadri-Felitti: Consumer Research Identifies How To

Present Travel Review Content For More Bookings. TrustYou. (2015)
[19] 马蜂窝旅游网： 2018 中国出境自由行大数据报告. （2018）
[20] GL Parrinello: Motivation and anticipation in post-industrial tourism. Annals of Tourism Research, (1993)
[21] Hanqing Zhang, Vincent C.S. Heung: The emergence of the mainland Chinese outbound travel market and its implications for tourism marketing. Journal of Vacation Marketing, (2001)
[22] Tony S.M. Tse: A Review of Chinese Outbound Tourism Research and the Way Forward. Journal of China Tourism Research, (2015)
[23] Tony S.M. Tse: Chinese Outbound Tourism as a Form of Diplomacy. Tourism Planning & Development, (2013)
[24] Barry Mak: The Influence of Political Ideology on the Outbound Tourism in China. Journal of China Tourism Research, (2013)
[25] Yingzhi Guo, Samuel Seongseop Kim, Dallen J. Timothy: Development Characteristics and Implications of Mainland Chinese Outbound Tourism. Asia Pacific Journal of Tourism Research, (2007)
[26] Individual Visit Scheme (港澳個人遊): https://en.wikipedia.org/wiki/Individual_Visit_Scheme
[27] 邱 曼绮: 我国出境旅游的现状与发展思路. 旅游管理研究, （2017）
[28] 环球旅讯: 2018-2019 年出境游报告：新跟团游兴起，半自助游增长超六成. https://www.traveldaily.cn/article/130937 , (2019)
[29] Wolfgang Georg Arlt: The Second Wave of Chinese Outbound Tourism. Tourism Planning & Development, (2013)
[30] Connie Mok, Agnes L. Defranco: Chinese Cultural Values: Their Implications for Travel and Tourism Marketing. Journal of Travel & Tourism Marketing, (1999)
[31] Xiang (Robert) Li, Chengting La, Rich Harrill, Sheryl Kline, Liangyan Wang: When east meets west: An exploratory study on Chinese outbound tourists' travel expectations. Tourism Management, (2011)
[32] Wolfgang Arlt: China's Outbound Tourism (Contemporary Geographies of Leisure, Tourism and Mobility). https://books.google.co.jp/books?id=Nqh-AgAAQBAJ&dq=China%E2%80%99s+outbound+tourism.+Oxfordshire:+Routledge.&hl=zh-CN (2006)
[33] Anna Kwek, Young-Sook Lee: Intra-Cultural Variance of Chinese Tourists in Destination Image Project: Case of Queensland, Australia. Journal of Hospitality & Leisure Marketing, (2008)
[34] Hilary du Cros, Liu Jingya: Chinese Youth Tourists Views on Local Culture.

Tourism Planning & Development, (2013)
[35] Ben Haobin Ye, Hanqin Qiu Zhang, Peter P.Yuen: An Empirical Study of Anticipated and Perceived Discrimination of Mainland Chinese Tourists in Hong Kong: The Role of Intercultural Competence. Journal of China Tourism Research, (2012)
[36] Ivy Chow, Peter Murphy: Travel Activity Preferences of Chinese Outbound Tourists for Overseas Destinations. Journal of Hospitality & Leisure Marketing, (2007)
[37] Magda Antonioli Corigliano: The Outbound Chinese Tourism to Italy: The New Graduates' Generation. Journal of China Tourism Research, (2011)
[38] Zhang Qiu Hanqin, Terry Lam: An analysis of Mainland Chinese visitors' motivations to visit Hong Kong. Tourism Management, (1999)
[39] Cathy H.C.Hsu, Terry Lam: Mainland Chinese Travelers' Motivations and Barriers of visiting Hong Kong. Journal of academy of business and economics, (2003)
[40] Mimi Li, Tong Wen, Ariel Leung: An Exploratory Study of the Travel Motivation of Chinese Female Outbound Tourists. Journal of China Tourism Research, (2011)
[41] Suzanne Amaro, aulo Duarte: Online travel purchasing: A literature review. Journal of Travel & Tourism Marketing, (2013)
[42] Dimitrios Buhalis, Rob Law: Progress in information technology and tourism management: 20 years on and 10 years after the Internet—The state of eTourism research. Tourism management, (2008)
[43] Miyoung Jeong, Jiyoung Choi: Effects of Picture Presentations on Customers' Behavioral Intentions on the Web. Journal of Travel & Tourism Marketing, (2004)
[44] Kara Wolfe, Cathy H.C. Hsu, Soo K.Kang: Buyer Characteristics Among Users of Various Travel Intermediaries. Journal of Travel & Tourism Marketing,(2004)
[45] Miguel Moital, Roger Vaughan, Jonathan Edwards: Buyer Characteristics AUsing involvement for segmenting the adoption of e-commerce in travelmong Users of Various Travel Intermediaries. The Service Industries Journal, (2009)
[46] Woo Gon Kim, Dong Jin Kim: Factors affecting online hotel reservation intention between online and non-online customers. International Journal of Hospitality Management, (2004)
[47] Alastair M.Morrison, Su Jing, Joseph T.O'Leary, Liping A. Cai: Predicting Usage of the Internet for Travel Bookings: An Exploratory Study. Information Technology & Tourism, (2001)
[48] Man Kit Chang, Waiman Cheung, Vincent S. Lai: Literature derived reference models for the adoption of online shopping. (2005)

[49] Yusniza Kamarulzaman: Adoption of travel e-shopping in the UK. International Journal of Retail & Distribution Management, (2007)
[50] George Vlachos: ONLINE TRAVEL STATISTICS 2012. INTERNET, TRAVEL & TOURISM, https://infographicsmania.com/online-travel-statistics-2012/ (2012)
[51] Stikky Media: 2012 & 2013 Social Media and Tourism Industry Statistics. internet, social media marketing for tourism | Marketing, Social Media, Tourism, (2014)
[52] Ivar E. Vermeulen, Daphne Seegers: Tried and tested: The impact of online hotel reviews on consumer consideration. Tourism Management, (2009)
[53] 张 补宏,周 旋, 广 新菊: A Review on the Research of Domestic and Foreign Tourism Online Reviews(国内外旅游在线评论研究综述). Geography and Geo-Information Science(地理与地理信息科学), (2017)
[54] Nan Hu, Ling Liu, Jie Jennifer Zhang: Do online reviews affect product sales? The role of reviewer characteristics and temporal effects. Information Technology and management, (2008)
[55] Nanda Kumar, Izak Benbasat: Research Note: The Influence of Recommendations and Consumer Reviews on Evaluationsof Websites. Information Systems Research, (2006)
[56] 陈 新华: 旅游者在线评论撰写行为的影响因素研究. Enterprise Economy(企业经济), (2016)
[57] Zhiwei Liu, Sangwon Park: What makes a useful online review? Implication for travel product websites. Tourism Management, (2015)
[58] Chris Forman, Anindya Ghose, Batia Wiesenfeld: Examining the Relationship Between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets. Information Systems Research, (2008)
[59] Christy M.K. Cheung, Matthew K.O. Lee, Neil Rabjohn: The impact of electronic word-of-mouth - The adoption of online opinions in online customer communities. Internet Research, (2008)
[60] Charu C. Aggarwal, ChengXiang Zhai: AN INTRODUCTION TO TEXT MINING. Mining Text Data, (2012)
[61] Johan Arndt: Role of Product-Related Conversations in the Diffusion of a New Product. Journal of Marketing Research, (1967)
[62] Anita Wenger: Analysis of travel bloggers' characteristics and their communication about Austria as a tourism destination. Journal of Vacation Marketing, (2008)
[63] 王 媛, 许 鑫, 冯 学钢, 吴 文智:Research on Tourists' Percieved Image of Ancient Town Using Web Text Mining Methods: A Case Study of Zhujiajiao(基于文本挖掘的古镇旅游形象感知研究—以朱家角为例). Tourism Science(旅游科学), (2013)

[64] Ya-Han Hu, Yen-Liang Chen, Hui-Ling Chou: Opinion mining from online hotel reviews –A text summarization approach. Information Processing & Management, (2017)
[65] 加藤 大受, 石川 博: Twitter データを活用した訪日外国人の観光行動分析の実現について. システム/制御/情報, (2019)
[66] 清水 伊織: 中国人の訪日旅行の形態とその変化. 地理学論集, (2007)
[67] Chengting Lai, Xiang (Robert) Li, Rich Harrill: Chinese outbound tourists' perceived constraints to visiting the United States. Tourism Management, (2013)
[68] Elaine Yin Teng Chew, Siti Aqilah Jahari: Destination image as a mediator between perceived risks and revisit intention: A case of post-disaster Japan. Tourism Management, (2014)
[69] 滕 鑑: 中国の海外旅行需要とその拡大要因について. 岡山大学経済学会雑誌, (2010)
[70] 鄔 雅瓊: 中国観光客の訪日行動と日中両国の観光政策. 北海商科大学学術研究会, (2016)
[71] 黄 愛珍: 訪日中国人観光客の旅行とインバウンド消費の動向. 静岡大学人文社会科学部アジア研究センター, (2017)
[72] 金 玉実: 日本における中国人旅行者行動の空間的特徴. 地理学評論, (2009)
[73] 戴 二彪: 訪日中国人観光客の旅行先分布構造と影響要因. 北九州発アジア情報, 国際東アジア研究センター, (2011)
[74] 菱田 のぞみ, 日比野 直彦, 森地 茂: 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 (土木計画学), (2012)
[75] Lihui Wu, Haruo Hayashi: The impact of the great east Japan earthquake on inbound tourism demand in Japan. 地域安全学会論文集, (2013)
[76] 郭 英之, 陈 芸, 黄 剑锋, 苏 勇: Travel Intentions of Chinese Residents to Japan Based on A Multidimensional Interactive Decision Tree Model (基于多维交互决策树模型的赴日旅游意愿研究). Tourism Tribune(旅游学刊), (2015)
[77] Mingjie Ji, Mimi Li, Cathy H. C. Hsu: Emotional Encounters of Chinese Tourists to Japan. Journal of Travel & Tourism Marketing, (2016)
[78] 夏 杰长, 徐 金海：中国旅游业改革开放 40 年：回顾与展望（Reform and Opening up of Tourism in China from 1978 to 2017:Retrospects and Prospects）经济与管理研究, （2018）
[79] Yanjun Xie, Miao Li: Development of China's Outbound Tourism and the Characteristics of Its Tourist Flow. Journal of China Tourism Research, （2009）
[80] Jian Li: Use of Rwordseg (Rwordseg 使用说明). http://jianl.org/cn/R/Rwordseg.html
[81] Thomas K Landauer, Peter W. Foltz, Darrell Laham: An Introduction to Latent Semantic Analysis. Discourse Processes, (1998)

[82] 特異値分解,
https://ja.wikipedia.org/wiki/%E7%89%B9%E7%95%B0%E5%80%A4%E5%88%86%E8%A7%A3
[83] Natural Language Processing（NLP） of Latent semantic analysis. (潜在语义分析)
https://zhuanlan.zhihu.com/p/48454667
[84] TK Landauer, ST Dumais: A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. Psychological review, (1997)
[85] Forrest,W.Young: Multidimensional Scaling History, Theory, and Applications. (1987)
[86] 多次元尺度構成法,
https://ja.wikipedia.org/wiki/%E5%A4%9A%E6%AC%A1%E5%85%83%E5%B0%BA%E5%BA%A6%E6%A7%8B%E6%88%90%E6%B3%95
[87] 樋口 耕一：社会調査のための計量テキスト分析―内容分析の継承と発展を目指して, (2014)
[88] Japan National Tourism Organization 日本政府観光局：訪日外客数（2010~2018）
[89] MeCab: Yet Another Part-of-Speech and Morphological Analyzer.
https://taku910.github.io/mecab/
[90] Koren, Yehuda, Robert Bell, Chris Volinsky: Matrix factorization techniques for recommender systems. Computerpp. 30-37, vol. 42 (2009).
[91] Thomas Hofmann: Probabilistic latent semantic analysis. Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence, (1999)
[92] Paul Covington, Jay Adams, Emre Sargin: Deep Neural Networks for YouTube Recommendations. Proceedings of the 10th ACM Conference on Recommender Systems Pages 191-198, (2016)
[93] L Bottou: Large-Scale Machine Learning with Stochastic Gradient Descent. Proceedings of COMPSTAT'2010, Springer (2010)