

京都を訪れる外国人旅行客の行動モデル設計とエージェントベースシミュレーション

○滝頭匠 李皓（静岡大学）

Behavior model Design and Agent-based simulation of Foreign Travelers

* K. Taki and H. Lee (Shizuoka University)

概要— 訪日外国人旅行客の数は2016年に過去最高の2,403万人を記録し、増加傾向は今後も続くことが予想される一方、観光客が伸び悩んでいる地域も存在している。同時に地域の観光を活性化させるDMOなどの活動が注目され、旅行客の行動様式を理解することの重要性はますます高まってきている。我々は各観光地における外国人旅行客数の推移を推定するためのエージェントベースシミュレーションに向けて、本研究では、外国人向けの観光サイトのレビュー記事を言語解析することで、各観光地の評価を推定し、評価と外国人旅行客実績に合わせて、訪日観光客の行動モデルを提案する。

キーワード: インバウンド, テキストマイニング, エージェントベースシミュレーション

1. 背景と目的

訪日外国人旅行客の数は2016年に過去最高の2,403万人を記録した¹⁾。政府は訪日外国人の数を東京オリンピックの開催される2020年までに4,000万人にする目標を打ち出し²⁾、訪日外国人の増加傾向は今後も続くことが予測される。しかし、訪日外国人の増加は様々な問題を引き起こすことも懸念されている³⁾。交通渋滞や観光地の混雑、それに伴う、宿や公共交通機関のキャパシティの限界や、安全保障、制度整備の面においても対処すべき課題は多い。

また、DMOなどの活動に代表される観光地のマーケティング活動においても、外国人旅行客の行動を正しく理解し、適切な対処をすることの重要性はますます高まってきている⁴⁾。

観光行動をモデル化し、シミュレーションすることは有効な手法であり、既存研究も多く存在する。高橋ら⁵⁾は観光スポットの魅力度を考慮した観光行動モデルを作成し、入込客数を推定した。高橋らのモデルでは観光地の選択確率を、観光スポット間の時間距離とアンケート調査から得られた観光スポットの魅力度から推定している。しかし、観光地の魅力度をアンケート調査によって取得する方法は、分析対象を広げ、現実の多くのスポットを考慮したモデルを作成することが困難である。

また、鈴木ら⁶⁾は上記の行動モデルの魅力度に当たるパラメータを、人気度として定義し、口コミサイトのレビュー数から推定した。しかし、目的地決定には人気度だけではなく個人の嗜好を考慮した、より詳細な分析が必要であることを課題としている。近年はこれら定量化の難しい観光地の魅力度を、SNSの投稿やクチコミ情報などのテキストを分析することで推定する研究が盛んに行われている⁷⁾が、推定した魅力度を用

いて観光行動モデルを作成しシミュレーションを行う研究は少ない。

以上を踏まえ、本研究では、観光地における外国人旅行客の行動モデルのパラメータ推定を目的に、まず各観光地に対する英語のレビューに対しテキストマイニングを行い、外国人旅行客が認識する各観光地の特徴と、出身地域ごとの外国人旅行客の特徴を推定した上で、推定した観光地属性と観光客属性の類似度を推定した。

そして、分析対象を観光目的の訪問客が多い京都市に絞り、目的地決定に距離、人気度、観光客と観光地との類似度(嗜好)を考慮した、外国人旅行客の京都市滞在中の観光行動モデルを作成した。既存研究のモデルでは、観光目的地決定のみであったが、本研究では、移動時間、滞在時間、京都への訪問回数も考慮した、より現実に近い観光行動のエージェントモデルを作成した。作成したモデルでエージェントベースシミュレーションを行い、各観光地の年間集客数を推定した。シミュレーション結果と、京都市の外国人旅行客の実績データをもとにモデルを検証し、モデルの妥当性を確認した。

2. テキストマイニング

2.1. 分析対象データ

我々は「tripadvisor」⁸⁾の口コミを用いて分析を行った。tripadvisorとは、観光地の口コミ数1億件以上を掲載している世界最大級の旅行口コミサイトである。特徴としては、旅行を構成する重要な要素である宿泊施設、レストラン、観光地といったジャンルごとに口コミ評価がなされており、平成25年3月時点のサイト管理会社のプレリリースに寄れば、1分間に平均60件以上の新たな口コミ投稿が行われ、日々新しい情報が加わ

り、観光地情報と旅行者情報の最新傾向を入手できる。

tripadvisor には口コミ投稿者の出身地域の情報も記載されており、以上の、大量のデータ、リアルタイム性、地域性の3つの性質は、いずれもデータの性質としては、非常に有用であるため、我々は tripadvisor のデータを採用した。

我々は tripadvisor に掲載された各地域(市町村)の人気上位5スポットの中で、英語のレビューが1件以上投稿されているスポットを対象にデータを収集し、その結果を Table1 にまとめた。条件に当てはまる日本のスポットは、1,764の地域の、6,626のスポットの内、英語のレビューが1件以上投稿されている2,263のスポットとなった。また、総レビュー数は125,671件で、総抽出語は9,807,680語であり、その内、同一の単語を除いた異なり語数は86,271語となった。

Table 1 : 収集データ概要

レビューデータセット	
スポット数	2,263
レビュー数	125,671
総抽出語	9,807,680
異なり語数	86,271
文	766,998
段落	620,327

2.2. 分析手順

分析には、無償のテキストマイニングソフトである KHcoder⁹⁾¹⁰⁾を用いた。KHcoder は、操作性が高く、多くの分野の研究論文で利用されている¹¹⁾。

まず、収集したテキストデータは改行コードや、顔文字など、分析に不要な記号などが含まれているため、これらを削除するテキストのクリーニング作業を行った。日本語の場合は、テキストマイニングの前処理で、単語を切り分ける形態素解析を行う必要があるが、今回の分析対象は英文であり、予め単語と単語の区切りが明確に示されるため、解析精度は極めて高い。単語が品詞ごとに整理されたテキストデータを用いて特徴分析、対応分析を行う。

2.3. データ分析

まず全ての収集したデータについて分析をしていく。

(1)頻出語

収集したデータ全体の名詞、形容詞、動詞の頻出語を Table 2 に示す。

Table 2 : 全データ頻出語表

No.	名詞	出現回数	No.	形容詞	出現回数	No.	動詞	出現回数
1	place	46917	1	beautiful	24264	1	visit	37155
2	temple	35860	2	great	22219	2	walk	28530
3	view	31014	3	good	21377	3	make	17187
4	time	30146	4	nice	20064	4	enjoy	14353
5	day	24997	5	japanese	16874	5	look	12936
6	shop	24361	6	worth	12062	6	come	12109
7	area	24151	7	amazing	11029	7	recommend	10449
8	lot	22144	8	small	10881	8	want	9375
9	shrine	21659	9	interesting	10597	9	try	8957
10	people	21591	10	little	9551	10	spend	8829

頻出語について見ると、名詞、形容詞、動詞、ともに観光の一般的な用語が並んでいることがわかる。名詞については、”temple”, ”shrine”など日本の寺や神社を表す単語が上位にランクしており、これらの観光地のレビューの多さがわかる。

(2)対応分析結果

次に、対応分析を行った結果を示す。対応分析とは、2種類のカテゴリー変数に関して、各カテゴリーの頻度数を要素とするクロス表を作成し、各カテゴリーに数値得点を与えることによって、各カテゴリー変数の相対距離を数量的に表す分析手法である。数量化理論 III 類では、片方のカテゴリー変数がサンプルに置き換えられる場合が多いが、分析手法自体は対応分析と同等であるため、その詳細な説明は専門書に譲りたい。

Fig. 1 は、出現数上位 80 語と、口コミが 500 件以上ある観光地 45 スポットをカテゴリー変数とし、対応分析を行った結果である。テキスト別観光地の相対的距離と位置、同様に出現数上位語の位置を把握し、近傍に位置づけられた出現数上位語からテキスト別に各観光地のイメージを抽出することを目的として 2 次元によるプロットを行った。

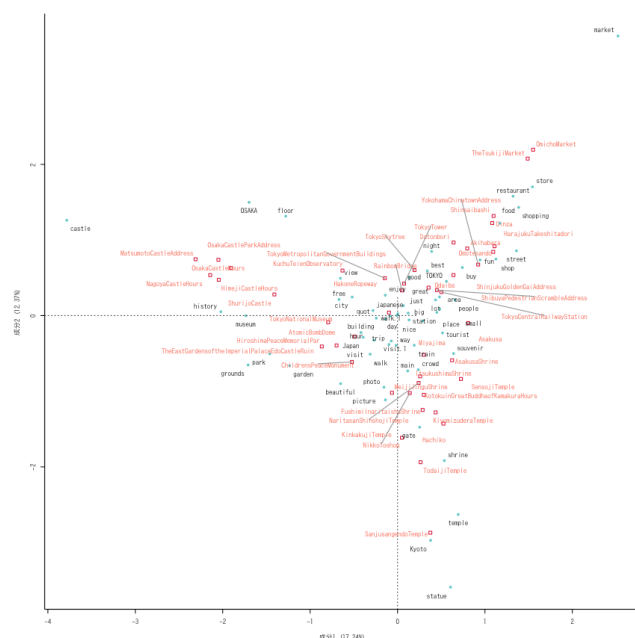


Fig. 1 : 全データ対応分析結果

成分1は城やミュージアムなどの文化的な面を表し、成分2は寺・神社などの伝統的な面を表すと考えられる。原点に近いほど平均的であることを意味しており、原点近くには観光に関する一般語である”visit”, ”work”, ”day”などがプロットされている。一方、原点から離れてプロットされた単語は特徴語として捉えることができ、同じように原点から離れた観光地の特徴を表しているものと考えられる。

Fig. 1 から、観光地は大きく3つのクラスターに分けることができる。1つ目はFig. 1の右上にまとまってプロットされている、築地市場や銀座などのショッピングやマーケット、食事などをするスポットのクラスターである、“store”、“street”などの単語と一緒にプロットされ、レストラン街やショッピング街を表していると考えられる。

2つ目はFig. 1の中央下にまとまってプロットされている、三十三間堂や東大寺などの寺や神社のスポットのクラスターである。“temple”、“Kyoto”などの単語と一緒にプロットされ、京都に代表される日本の寺や神社を表している。3つ目はFig. 1の左にまとまってプロットされている、大阪城や松本城など、城のクラスターである。“castle”、“museum”などの単語と一緒にプロットされ、城だけでなく、城の中の展示なども表していると考えられる。

このように、テキストマイニングにより、観光地の特徴を知ることが出来る。

2.4. 京都の観光スポットパラメータ推定

京都市は観光目的の訪問客が多くレビュー数も豊富である。また、京都市は京都観光総合調査¹²⁾と呼ばれる京都市の観光実績の調査を公開しており、その中で、25の観光地について、アンケート調査の結果から外国人旅行客が訪れた割合をまとめている。本研究では、この実績データをもとに妥当性の検証ができるため、京都観光総合調査の調査対象の25のスポットのうち、観光スポットとしてtripadvisorに投稿されている19のスポットのデータに対して分析を行った。Table 3に19のスポットの訪問実績を示す。ここで、アジアは実績データの中国、台湾、韓国、東南アジアの平均値の値であり、アフリカ、南アメリカ、中東は「その他」に分類される実績の値を用いた。

先行研究で、室谷¹³⁾は、京都市の魅力度をパラメータ表示した。室谷は、このパラメータをアンケート調査から推定したが、本研究では、テキストマイニングから取得したデータを元に推定を行う。

Table 3 : 外国人旅行客の訪問実績

順位	訪問地	africa	asia	europa	latin_america	middle_east	north_america	pacific
1	清水寺	68.00%	74.05%	52.40%	68.00%	68.00%	56.10%	50.70%
2	金閣寺	41.30%	50.80%	58.60%	41.30%	41.30%	49.30%	47.80%
3	二条城	48.90%	41.98%	58.00%	48.90%	48.90%	61.00%	53.70%
4	伏見稲荷大社	32.90%	37.75%	46.80%	32.90%	32.90%	47.10%	37.30%
5	京都御所	27.60%	20.43%	42.20%	27.60%	27.60%	47.10%	47.00%
6	ギオンコーナ	26.20%	26.40%	37.10%	26.20%	26.20%	33.60%	35.10%
7	銀閣寺	26.70%	23.25%	43.50%	26.70%	26.70%	31.40%	34.30%
8	錦市場	22.70%	26.33%	36.60%	22.70%	22.70%	32.30%	33.60%
9	八坂神社	20.90%	26.10%	22.00%	20.90%	20.90%	17.50%	12.70%
10	京都タワー	29.80%	26.68%	16.20%	29.80%	29.80%	20.60%	22.40%
11	南禅寺	11.60%	10.13%	26.50%	11.60%	11.60%	13.90%	17.20%
12	三十三間堂	12.00%	10.58%	17.90%	12.00%	12.00%	10.30%	8.20%
13	平安神宮	6.20%	12.95%	14.70%	6.20%	6.20%	12.10%	4.50%
14	高台寺	6.20%	8.80%	20.90%	6.20%	6.20%	12.60%	11.90%
15	東寺	7.10%	7.18%	18.50%	7.10%	7.10%	17.00%	6.70%
16	下鴨神社	8.00%	8.80%	9.90%	8.00%	8.00%	9.00%	6.70%
17	モンキーパーク	3.10%	7.18%	13.40%	3.10%	3.10%	12.60%	14.90%
18	東福寺	5.30%	7.28%	14.00%	5.30%	5.30%	8.50%	5.20%
19	西本願寺	4.90%	9.48%	10.10%	4.90%	4.90%	7.60%	3.00%

(1) クラスター分析

我々は、パラメータ推定のためにクラスター分析とdoc2vecの両方を用いた。比較検証の結果、クラスター分析の方が、精度が高いため、ここではdoc2vecに関連する説明を割愛する。

パラメータ表示までの流れは、まず、収集したデータをクラスター分析することで、似た特徴の単語をクラスターに分類する。その後、各クラスターにラベルをつけ、各観光地のレビュー文書にラベルの単語が出現する割合を算出する。そして、各クラスターの出現割合を標準化し、パラメータ表示をする。

レビュー数はスポットによって大きく異なる。例えば、京都の人気スポットである伏見稲荷大社の英語レビュー数は9,000件以上であるのに対し、レビュー数が100件に満たないスポットも存在する(Table 4)。レビュー数の差による影響をできる限り少なくした分析をするため、各スポットのレビューからランダムで100件ずつ抽出したデータセットを作成した。レビュー数が100件に満たない場合は、該当レビューを全数利用する。

Table 4 : 英語レビュー数

スポット名	英語レビュー数
清水寺	4,033
金閣寺	6,768
二条城	2,958
伏見稲荷大社	9,667
京都御所	850
ギオンコーナ	272
銀閣寺	1,637
錦市場	1,901
八坂神社	317
京都タワー	255
南禅寺	554
三十三間堂	1,281
平安神宮	284
高台寺	305
東寺	289
下鴨神社	67
嵐山モンキーパーク	1,476
東福寺	383
西本願寺	141

そして、KHcoderを用い、各スポットのレビューから100件ずつ抽出したデータセットの、出現回数が1,000回以上の名詞86語を対象にクラスター分析を行った。結果をTable 5に示す。ここで、クラスター分析では、ward法を用い、単語の距離にはjaccard係数を用いた。Jaccard係数とは0から1の数字で表され、単語が1に近づくほど関係性が強く、そのテキストの特徴を表す単語であることを示す係数である。

Table 5 : ラベル付きクラスター分析結果

クラスター	ラベル	単語
1	*shopping	food,souvenir,shop,restaurant,store,street,shopping
2	*transportation	station,train,bus,stop,trip,way,walk,hour,minute
3	*time・place	area,place,time,day,lot,people
4	*temple・shrine	visit,temple,shrine,building,garden
5	*mountain	ride,mountain,car,snow,weather
6	*island	bridge,island
7	*nature	hill,step,stair,park,season,road,tree,path,water,gate
8	*room	house,village,tea,room
9	*family	amp,family,fun,kid
10	*hotel	morning,night,hotel
11	*tourist	visitor,center,town,tourist,attraction
12	*experience	bit,thing,experience,quot,year,end,line,photo,spot,tour,site,point,picture,statue
13	*entrance	entrance,fee,ticket,yen
14	*castle	grounds,museum,castle,history
15	*tower	view,city,floor,tower

結果について取り上げてみると、まず、第1クラスターは”food”, ”souvenir”, ”shop”など買い物や食事を表す単語で構成されており、概して「ショッピング」を表すクラスターであると考えられる。第2クラスターは、”station”, ”train”, ”bus”など、交通機関を表す単語で構成されており、概して「交通機関」を表すクラスターであると考えられる。クラスター分析の結果、86の単語が15のクラスターに分類された。

(2)京都スポットパラメータ推定

京都の19スポットのデータを用いて、ラベル付けしたクラスターを構成する単語が出現した文書の割合を、スポットごとに算出した。ここで、文書の単位は1レビューを1文書とした。この割合を標準化した上で、平均値が0.5、標準偏差が0.1となるように正規化した。Table 6に結果を示す。

Table 6 : 京都19スポット正規化パラメータ値

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
FushimiInariTaishaShrine	0.501	0.722	0.651	0.475	0.495	0.447	0.646	0.437	0.505	0.519	0.546	0.576	0.388	0.423	0.478
OkakuraTemple	0.468	0.477	0.341	0.572	0.477	0.469	0.530	0.444	0.437	0.444	0.430	0.397	0.429	0.492	0.493
GionCorner	0.467	0.544	0.501	0.258	0.412	0.450	0.353	0.871	0.470	0.495	0.692	0.612	0.757	0.418	0.440
HeianShrine	0.440	0.385	0.489	0.583	0.471	0.722	0.562	0.472	0.463	0.418	0.464	0.481	0.599	0.498	0.446
KinkakuTemple	0.445	0.422	0.491	0.555	0.496	0.459	0.462	0.461	0.470	0.451	0.629	0.537	0.634	0.478	0.468
KiyomizuderaTemple	0.435	0.542	0.581	0.558	0.519	0.448	0.560	0.465	0.540	0.503	0.621	0.503	0.443	0.460	0.603
KodaijiTemple	0.471	0.420	0.499	0.567	0.471	0.491	0.487	0.522	0.445	0.727	0.407	0.437	0.511	0.534	0.452
KyotoImperialPalace	0.427	0.501	0.582	0.503	0.456	0.466	0.502	0.505	0.505	0.444	0.474	0.624	0.454	0.638	0.458
KyotoTowerHours	0.547	0.600	0.440	0.382	0.508	0.450	0.374	0.449	0.518	0.724	0.539	0.373	0.623	0.412	0.914
MandokParkIwazayama	0.355	0.691	0.520	0.357	0.814	0.331	0.690	0.495	0.868	0.415	0.465	0.510	0.579	0.408	0.623
NanzenjiTemple	0.436	0.479	0.462	0.578	0.489	0.476	0.622	0.477	0.451	0.450	0.385	0.400	0.502	0.526	0.491
NijoCastle	0.434	0.457	0.423	0.498	0.443	0.457	0.446	0.629	0.474	0.433	0.364	0.468	0.457	0.839	0.548
NishiHonganji	0.456	0.546	0.410	0.564	0.421	0.467	0.487	0.493	0.468	0.478	0.500	0.478	0.452	0.474	0.460
NishikiMarket	0.834	0.392	0.568	0.324	0.436	0.445	0.343	0.494	0.541	0.449	0.394	0.408	0.323	0.409	0.408
SaigyoInariTemple	0.440	0.427	0.523	0.560	0.428	0.446	0.457	0.453	0.405	0.790	0.411	0.511	0.511	0.413	0.413
ShimogamoInari	0.421	0.565	0.539	0.574	0.585	0.540	0.561	0.454	0.554	0.448	0.659	0.474	0.491	0.495	0.438
TofukujiTemple	0.431	0.476	0.493	0.567	0.459	0.818	0.513	0.429	0.445	0.445	0.487	0.435	0.603	0.479	0.496
Toji	0.513	0.455	0.445	0.546	0.448	0.455	0.435	0.491	0.466	0.542	0.441	0.324	0.556	0.544	0.481
YasakaShrine	0.587	0.492	0.741	0.574	0.473	0.464	0.567	0.445	0.448	0.682	0.601	0.467	0.477	0.463	0.430

Table 6をもとに京都の19スポットのパラメータ推定をした。京都の19スポットのパラメータ推定の結果の一部をFig. 2に示す。

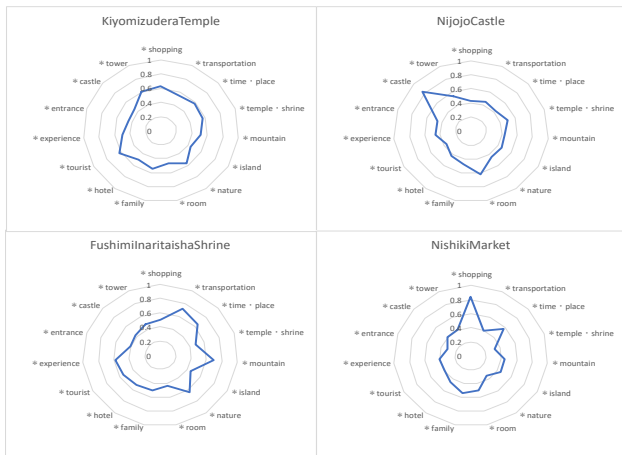


Fig. 2 : 京都のスポットパラメータ推定

パラメータ表示の結果、それぞれの特徴となるパラメータの値が大きくなっていることがわかる。例えば、錦市場では、”shopping”要素、京都タワーでは”tower”要素が大きくなっており、室内体験型スポットであるギオンコーナーでは、他のスポットと比較して”room”要素が大きくなっていることがわかる。一方、寺や神社

を表す”temple・shrine”要素は、全体的に寺や神社が多いことから、特別目立って大きい値を示すスポットは見られない。パラメータ表示をする事で、各スポットの特徴を表すと考えられる要素を視覚的にとらえることができた。

(3)地域パラメータ表示

tripadvisor に投稿されるレビューは、投稿者が出身地域を登録していれば、出身地域別に分類することができる。地域別にレビューを分析することで、その地域の観光客特徴を分析することができる。その際、レビュー数の差から出る誤差をなくすために、データは各地域からランダムで500件ずつレビューを抽出した。

まず、分類したレビューの数を地域別にカウントした結果をTable 7に示す。

Table 7 : 地域別レビュー数

地域	アジア	北アメリカ	ヨーロッパ	オセアニア	中東	南アメリカ	アフリカ
レビュー数	27,278	23,324	19,157	15,724	1,667	894	652

地域別のデータセットを用いて、京都のスポットと同様の手順で各地域についてもパラメータ表示も行った。各地域のラベル付けしたクラスターを構成する単語が、出現した文書の割合を算出し、標準化した上で、平均値が0.5、標準偏差が0.1となるように正規化した結果をTable 8に示す。

Table 8 : 世界7地域正規化パラメータ値

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
africa	0.441	0.516	0.379	0.581	0.524	0.504	0.482	0.497	0.436	0.477	0.470	0.560	0.421	0.470	0.522
asia	0.651	0.547	0.626	0.323	0.574	0.390	0.518	0.426	0.466	0.531	0.629	0.429	0.673	0.569	0.594
australia	0.423	0.434	0.475	0.381	0.457	0.390	0.526	0.351	0.376	0.509	0.522	0.422	0.456	0.526	0.401
latin america	0.356	0.355	0.644	0.297	0.300	0.447	0.373	0.443	0.428	0.299	0.355	0.541	0.279	0.396	0.424
middle east	0.580	0.405	0.491	0.407	0.499	0.676	0.373	0.354	0.593	0.509	0.397	0.360	0.463	0.421	0.385
north america	0.526	0.632	0.491	0.521	0.598	0.461	0.625	0.587	0.653	0.551	0.595	0.651	0.561	0.681	0.591

Table 8をもとに世界7地域のパラメータ推定をした結果の一部をFig. 3に示す。



Fig. 3 : 世界の地域パラメータ表示

各出身地域のパラメータ表示の結果は、スポットのパラメータ表示にほど大きな特徴は見られないが、アジアでは他の地域と比べて”shopping”要素の割合が高いことや、北アメリカは他の地域と比べて”castle”要素が大きいなど、それぞれの出身地域地域での観光客の特性を読み取ることができる。

2.5. 類似度の計算

行動モデルを設計する上で必要になる各観光客の各観光スポットの嗜好を推定するため、推定した京都の19スポットのパラメータ値(Table6)と、出身地域別のパラメータ値(Table8)を合わせることで、京都の19スポットと、各地域との類似度を推定した。具体的に、京都の19スポットのパラメータ値と、各地域のパラメータ値を、それぞれベクトルとして、二つのベクトルのcos類似度を計算することによって求めた。

以下にcos類似度の計算式を示す。以下にcos類似度の計算式を示す。

$$\cos(\vec{q}, \vec{d}) = \frac{\vec{q} \cdot \vec{d}}{|\vec{q}| |\vec{d}|} = \frac{\vec{q} \cdot \vec{d}}{|\vec{q}| \cdot |\vec{d}|} = \frac{\sum_{i=1}^{|V|} q_i d_i}{\sqrt{\sum_{i=1}^{|V|} q_i^2} \sqrt{\sum_{i=1}^{|V|} d_i^2}} \quad (2-1)$$

求めた類似度の結果を Table 9 に示す。

Table 9 : cos 類似度表

spot_name	Europe	Africa	MiddleEast	Asia	Pacific	NorthAmerica	LatinAmerica
KiyomizuderaTemple	0.485	0.616	0.310	0.376	0.368	0.678	0.647
KinkakujiTemple	0.808	0.271	0.741	0.594	0.564	0.281	0.447
NijoCastle	0.363	0.624	0.419	0.459	0.365	0.570	0.641
FushimiInariTaishaShrine	0.533	0.318	0.704	0.592	0.655	0.355	0.394
KyotoImperialPalace	0.652	0.407	0.565	0.684	0.434	0.389	0.409
GionCorner	0.401	0.704	0.438	0.351	0.363	0.670	0.537
GinkakujiTemple	0.498	0.407	0.632	0.466	0.561	0.391	0.626
NishikiMarket	0.510	0.409	0.415	0.705	0.325	0.647	0.403
YasakaShrine	0.420	0.747	0.539	0.290	0.511	0.535	0.461
KyotoTowerHours	0.377	0.647	0.281	0.266	0.508	0.765	0.583
NanzenjiTemple	0.636	0.351	0.722	0.596	0.455	0.373	0.491
SanjusangendoTemple	0.563	0.420	0.652	0.545	0.395	0.595	0.379
HeianShrine	0.819	0.212	0.676	0.644	0.435	0.343	0.578
KodaijiTemple	0.350	0.568	0.311	0.642	0.788	0.277	0.372
Toji	0.783	0.244	0.477	0.741	0.504	0.398	0.421
ShimogamoInn	0.490	0.504	0.483	0.386	0.429	0.638	0.603
MonkeyParkIwatayama	0.525	0.375	0.732	0.530	0.700	0.349	0.347
TofukujTemple	0.600	0.422	0.530	0.577	0.498	0.413	0.506
NishiHonganji	0.271	0.658	0.430	0.468	0.482	0.480	0.601

類似度の値は、スポットや、地域によって大きく異なっており、それぞれの地域の特徴とスポットの特徴の類似度を定量的に表している。例えば、パラメータ表示の際にアジアは”shopping”要素の値が高くなっていたが、類似度についても”shopping”要素の値が大きい錦市場との類似度が0.7と高くなっている。

この類似度の結果は、3章のシミュレーションモデルの類似度の値として用いる。

3. シミュレーションモデル

3.1. 概要

テキストマイニングの結果を用いて、分析対象を観光目的の訪問客が多い京都市の代表的な19のスポットに絞り、目的地決定に距離、人気度、観光地との類似度(嗜好)を考慮した外国人旅行客の観光行動モデル

を作成した。既存研究のモデルでは、観光目的地決定のみであったが、本モデルでは、移動時間、滞在時間、京都への訪問回数も考慮した、より現実に近い観光行動モデルを作成した。このモデルをもとに、エージェントベースシミュレーションを行い、各観光地の年間客数を推定した。また、シミュレーション結果と、京都市の外国人旅行客の実績データをもとに、モデルの妥当性を検証した。

なお、このシミュレーションモデルは社会シミュレーション用言語であるSOARS¹⁴⁾¹⁵⁾を利用して作成した。SOARS言語はシミュレーション結果をアニメーションで確認することができるため、構築したモデルの直感的理解が容易である。我々は、行動モデルの理解を目的にしているため、SOARSを使用することとした。SOARSは、金谷ら¹⁶⁾による新型インフルエンザに代表される大規模感染症を対象とした社会シミュレーションの研究など実際の研究にも多く使用されている。

3.2. 観光行動モデル設計

先行研究において、鈴木ら(2015)は回遊者行動をLevyfightモデルに観光スポットの人気度と距離に依存したパラメータを導入し、モデル化を行った。Levyfightモデルとはランダムウォークの一種である。進行方向は常にランダムに選択され、稀に非常に長い距離の移動が行われることが特徴であり、ランダムウォークに比べ、効率の良い探索が行われることが示唆されている。本研究では、鈴木らのモデルを参考に、テキストマイニングで求めたスポット属性とエージェントの嗜好との類似度パラメータを加え、エージェントの目的地決定モデルを作成した。以下にモデルを示す。

地域別のエージェント集合を $I = \{i, \dots\}$ とし、エージェント i が未訪問の観光スポット集合を $S_i = \{s, t, v, m, \dots\}$ とし、スポット集合 S_i に含まれるスポットの数を $|S_i|$ とする。スポット s からスポット t への移動に必要な距離を $d(s, t)$ とする。この時、回遊者タイプがスポット s の後にスポット t を訪れる確率 $p_d(t|s)$ を次式で定義した。ここで重み w_d は0~1の値をとり、0に近づくにつれて、確率 $p_d(t|s)$ はランダム確率に近づく。

$$p_d(t|s) = \frac{\{d(s, t) \cdot w_d + |S_i|^{-1} \cdot (1 - w_d)\}^{-1}}{\sum_{v \in S_i} \{d(s, v) \cdot w_d + |S_i|^{-1} \cdot (1 - w_d)\}^{-1}} \quad (3-1)$$

また、各スポット t に対し、スポットの人気度 $f(t)$ に比例したスポット選択確率 $p_p(t|s)$ を以下のように定義した。ここで重み w_p は0~1の値をとり、0に近づくにつれて、確率 $p_p(t|s)$ はランダム確率に近づく。

$$p_p(t|s) = \frac{f(t) \cdot w_p + |S_i|^{-1} \cdot (1 - w_p)}{\sum_{v \in S_i} f(v) \cdot w_p + |S_i|^{-1} \cdot (1 - w_p)} \quad (3-2)$$

また、各スポット t に対し、スポット t とエージェント i との類似度 $f'(t,i)$ に比例したスポット選択確率 $p_{s_i}(t|s)$ を以下のように定義した。ここで重み w_{s_i} は0~1の値をとり、0に近づくにつれて、確率 $p_{s_i}(t|s)$ はランダム確率に近づく。

$$p_{s_i}(t|s) = \frac{f'(t,i) \cdot w_{s_i} + |S_i|^{-1} \cdot (1 - w_{s_i})}{\sum_{v \in S_i} f'(t,i) \cdot w_{s_i} + |S_i|^{-1} \cdot (1 - w_{s_i})} \quad (3-3)$$

本研究では、目的地決定モデルとしてエージェント i がスポット s の後にスポット t を訪れる確率 $p_{d,p,s_i}(t|s)$ を式(3-1), (3-2), (3-3)を用いて次式で定義する。

$$p_{d,p,s_i}(t|s) = \frac{p_d(t|s) \cdot p_p(t|s) \cdot p_{s_i}(t|s)}{\sum_{v \in S_i} \{p_d(t|s) \cdot p_p(t|s) \cdot p_{s_i}(t|s)\}} \quad (3-4)$$

また、パラメータの組合せからの7つの回遊者モデル、type0: 距離・人気度・類似度重視型、type1: 距離・人気度重視型、type2: 距離・類似度重視型、type3: 人気度・類似度重視型、type4: 人気度重視型、type5: 類似度重視型、type6: 距離重視型に分類し、目的地決定タイプ別に7シナリオを用いて分析をした。Table 10 にエージェント7シナリオを示す。

Table 10 : エージェント7シナリオ

	距離(w_d)	人気度(w_p)	類似度(w_s)
type0	1	1	1
type1	1	1	0
type2	1	0	1
type3	0	1	1
type4	0	1	0
type5	0	0	1
type6	1	0	0

3.3. エージェントの目的地決定パラメータ

エージェントが目的地を決定するのに必要なパラメータは、観光地間の距離と、観光地の人気度、そして、エージェントの出身地域ごとの各観光地スポットとの類似度である。観光地間の距離は、実距離を考慮し、グーグルマップのルート検索を利用して求めた。京都においては、移動手段として、外国人旅行者自ら自動車で移動するケースは少ないため、観光者は徒歩、バス、電車 のいずれかを利用すると仮定し、最も少ない時間距離を採用した。求めた時間距離の表を Table 11 に示す。

Table 11 : スポット間時間距離(分)

	BD.1	BD.2	BD.3	BD.4	BD.5	BD.6	BD.7	BD.8	BD.9	BD.10	BD.11	BD.12	BD.13	BD.14	BD.15	BD.16	BD.17	BD.18	BD.19	BD.20
BD.1 = 清水寺	0	65	40	35	45	20	35	30	15	30	35	25	25	15	45	45	60	35	40	35
BD.2 = 金閣寺	60	0	30	60	30	50	45	40	50	40	50	60	45	55	55	30	50	40	40	35
BD.3 = 二条城	40	30	0	35	25	25	40	20	30	20	25	35	25	35	35	40	40	20	20	20
BD.4 = 伏見稲荷大社	35	60	35	0	35	20	50	25	25	15	35	20	35	30	25	40	60	15	25	10
BD.5 = 京都御所	45	30	25	35	0	30	25	25	30	25	30	30	30	35	40	25	50	40	30	20
BD.6 = ギオンコーナー	20	50	25	20	30	0	30	15	5	25	25	15	20	10	40	30	50	25	30	25
BD.7 = 銀閣寺	35	45	40	50	25	30	0	35	25	40	20	35	20	35	60	30	65	40	50	40
BD.8 = 錦市場	30	40	30	25	25	15	35	0	15	15	30	25	25	20	30	30	45	25	25	15
BD.9 = 八坂神社	15	50	30	25	30	5	25	15	0	25	25	15	15	10	35	30	45	25	25	25
BD.10 = 京都タワー	30	40	30	15	25	25	40	15	25	0	30	10	30	25	20	35	45	20	10	35
BD.11 = 南禅寺	35	50	25	35	30	25	30	25	30	0	30	15	30	40	35	50	40	40	40	25
BD.12 = 三十三間堂	20	60	35	30	30	15	35	25	15	10	30	0	25	30	30	30	60	20	20	20
BD.13 = 平安神宮	25	45	25	35	30	20	20	25	15	30	15	25	0	20	50	30	50	40	40	30
BD.14 = 高台	15	55	35	30	35	10	35	20	10	25	30	20	20	0	40	35	60	30	35	30
BD.15 = 東寺	45	55	35	25	40	40	60	30	35	20	40	30	50	40	0	50	55	30	20	10
BD.16 = 下鴨神社	45	30	40	40	25	30	30	30	35	35	30	30	35	50	0	55	40	45	35	35
BD.17 = 嵐山モンキーパーク	60	50	40	60	50	50	65	45	45	45	50	60	50	60	55	55	0	60	40	35
BD.18 = 東福寺	35	60	40	15	40	25	40	25	25	20	40	20	40	30	30	40	60	0	35	15
BD.19 = 西本願寺	40	40	20	25	30	30	50	25	25	10	40	20	40	35	20	45	40	35	0	15
BD.20 = 京都市	35	35	20	10	20	25	40	15	25	5	25	20	30	30	10	35	35	15	15	0

観光地の人気度には、鈴木らのモデルと同様に、tripadvisor に投稿されているレビューの数を使用した。ただし、外国人旅行者のレビュー数だけに絞るため、日本語以外のレビュー数を使用した。11月30日時点の、各スポット日本語以外のレビュー数の表を Table 12 に示す。

Table 12 : スポット別日本語以外のレビュー数

スポット名	日本語以外のレビュー数
清水寺	7,056
金閣寺	11,567
二条城	4,471
伏見稲荷大社	15,131
京都御所	1,309
ギオンコーナー	380
銀閣寺	2,903
錦市場	2,918
八坂神社	640
京都タワー	519
南禅寺	877
三十三間堂	2,127
平安神宮	463
高台	546
東寺	493
下鴨神社	152
嵐山モンキーパーク	1,842
東福寺	625
西本願寺	291

類似度にはクラスター分析の結果から求めた類似度 (Table 9) を用いる。

モデル作成にあたり、初めて京都を訪れる人と、2回以上訪れたことのあるリピータを区別した。そして、2回以上訪れたことのあるリピータは前回の訪問で5つのスポットに訪れたと仮定し、前回訪れスポットには訪れない設定をした。リピータ率とはこのリピータが全体の何%いるかを表す割合で、京都観光総合調査の、「京都への来訪回数」を参考に、本モデルでは初めて来訪する約78%を除いて22%とした。(Table 13)

Table 13 : 京都のリピータの割合

訪問回数	割合
初めて	77.8%
2回目	14.3%
3回目	3.3%
4回目	1.3%
5回目以上	3.0%
無回答	0.3%
合計	100.0%

各スポットでの滞在時間は、kyotrip¹⁷⁾に掲載されているものを参考に作成した。Kyotrip とは、京都の観光地を中心に観光情報を掲載している web サイトであり、滞在時間などの詳しい情報も掲載されている。kyotrip に掲載されていない、ギオンコーナー、錦市場などの観光地については個人のブログ¹⁸⁾¹⁹⁾²⁰⁾²¹⁾などを参考に作成した。各スポットの滞在時間の表を Table 14 に示す。

Table 14 : 各スポット滞在時間

スポット名	滞在時間(分)
清水寺	45
金閣寺	60
二条城	90
伏見稲荷大社	120
京都御所	60
ギオンコナー	50
銀閣寺	45
錦市場	90
八坂神社	20
京都タワー	30
南禅寺	90
三十三間堂	40
平安神宮	40
高台	30
東寺	60
下鴨神社	40
嵐山モンキーパーク	40
東福寺	45
西本願寺	30

3.4. SOARS モデル

SOARS モデルにおける行動主体を示すエージェントは世界の地域ごとに7種類で、それぞれのエージェントの数は1,000とし、共通のルールを適用した。ルールにはエージェントの行動ルールが設定され、ステージごとに条件式と行動が順に記述されている。実際のSOARSシミュレーターの画面をFig. 4に、ステージをTable15に示す。

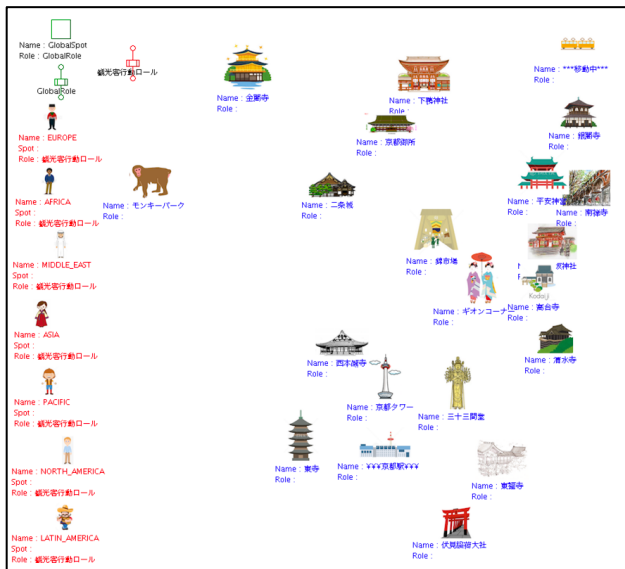


Fig. 4 : シミュレーション画面

Table 15 : SOARS モデルステージ表

ステージ	ステージ名	ステージ説明
初期ステージ	モデルタイプ決定	エージェントが目的地決定の時に用いるルールのタイプを決定する
	エージェント初期実数設定	エージェントの初期設定に必要なパラメータを設定する
	リビート率設定	リビート率を設定する
メインステージ	移動可能判定&目的地決定	エージェントが移動可能かどうか判定し、移動可能ならば目的地を決定する
	移動時間の設定	移動時間を決定する
	移動終了	移動時間を超えていれば、移動完了とする
	目的地到着	移動が完了したら、目的地到着とする
	入込客カウント	目的地の入込客にカウントする

本モデルではシミュレーションの開始時刻はシミュレーション内の時間8:00で、終了時刻は17:00とした。

また、シミュレーション実行間隔をシミュレーション内の時間で5分に設定した。これはシミュレーション開始時に初期ステージの処理が行われてから、シミュレーション内の時間で5分ごとにメインステージが繰り返されることを意味する。エージェントの行動フローチャートをFig. 5に示す。

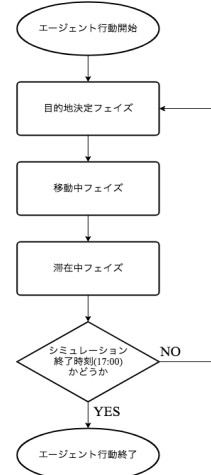


Fig. 5 : エージェント行動フローチャート

各エージェントは京都駅をスタート地点として、8:00になると行動を開始する。

目的地決定フェイズでは、各エージェントは、式3-4で定義した目的地決定モデルに基づいて、目的地を決定する。

目的地決定後は、移動中フェイズに移る。移動中フェイズでは、決定した目的地によって、移動時間を設定し、エージェントは移動中スポットに移動する。

移動終了時間に達すると、目的地までの移動が完了したとみなし、滞在中フェイズに移る。エージェントは目的地の観光スポットに移動し、観光スポットごとの滞在時間をもとに、スポットでの滞在時間を決定する。

そして、スポットでの滞在時間が経過すると、終了時刻の17:00に達しているか判定し、達していないならば、次の目的地に向けて、目的地決定フェイズに移る。終了時刻の17:00に達していればそこでエージェントの行動は終了となる。

エージェントの観光行動の例をFig.6に示す。

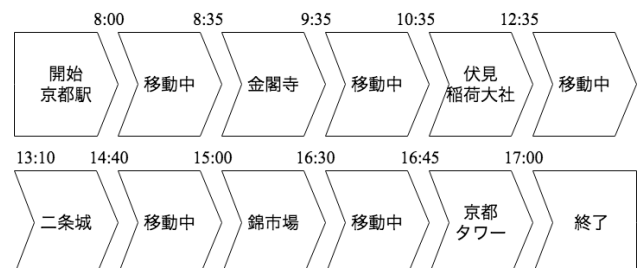


Fig. 6 : エージェント観光行動例

4. シミュレーション結果・分析

以下にシミュレーションを行なった結果を示す。また、タブ別シミュレーション結果では目的地決定の際に重視するパラメータによって、7つの回遊者モデル(Table 10)を用いて分析をした結果を示す。

4.1. シナリオ別シミュレーション結果

各地域における実績データとの誤差の、絶対値の平均を、type 別、地域別に算出した。結果を Table 16 に示す。

Table 16：出身地域の実績データとの誤差

	type0	type1	type2	type3	type4	type5	type6
africa	11.98%	12.83%	20.42%	12.93%	13.25%	17.22%	22.19%
asia	11.01%	12.03%	19.66%	11.93%	13.01%	17.53%	20.74%
europa	8.59%	8.68%	17.21%	8.86%	10.29%	14.69%	17.73%
latin_america	11.91%	12.94%	19.61%	12.18%	12.85%	17.78%	21.44%
middle_east	12.08%	12.70%	20.45%	12.88%	13.31%	17.34%	21.45%
north_america	9.91%	10.03%	20.19%	10.83%	10.59%	17.50%	19.68%
pacific	11.55%	11.62%	19.08%	11.74%	11.87%	17.36%	20.11%
誤差平均	11.00%	11.55%	19.52%	11.62%	12.17%	17.06%	20.48%

シミュレーションの結果、type0(距離・人気度・類似度重視型)が誤差の平均値が 11.00%となり、もっとも実績データとの誤差が少なくなっていることが分かる。次に実績データとの誤差が少ないのは、type1(距離・人気度重視型)で、誤差の平均値は 11.55%となった。そして、誤差が少ない順に type3(人気度・類似度重視型)、type4(人気度重視型)、type5(類似度重視型)、type2(距離・類似度重視型)と続き、もっとも実績データとの誤差が大きかったのは type6(距離重視型)となった。

もっとも誤差の値が小さい type0 について、スポット別、地域別の誤差を Table 17 に示す。

Table 17：type0_実績データとの誤差

type0	africa	asia	europa	latin_america	middle_east	north_america	pacific	絶対値平均
清水寺	-3.10%	-6.65%	12.20%	-1.90%	-2.30%	7.20%	15.50%	6.98%
金閣寺	24.00%	14.30%	6.10%	21.80%	22.20%	13.40%	17.40%	17.60%
二条城	11.20%	19.73%	4.60%	13.00%	14.10%	3.50%	7.30%	10.49%
伏見稲荷大社	40.60%	35.95%	27.40%	30.90%	11.50%	28.30%	38.30%	36.04%
京都御所	-2.30%	5.68%	-19.20%	-2.00%	-4.90%	-22.60%	-23.60%	11.47%
ギオンコーナー	-11.30%	-12.90%	-22.90%	-12.80%	-10.10%	-19.90%	-20.10%	15.71%
龍泉寺	16.00%	20.85%	-1.00%	18.30%	16.60%	13.00%	9.40%	13.59%
錦市場	22.60%	14.68%	10.00%	22.20%	22.80%	12.30%	12.40%	16.71%
八坂神社	-0.70%	-6.70%	-2.90%	-0.90%	-2.30%	1.60%	7.40%	3.21%
京都タワー	-8.10%	-1.48%	6.20%	-6.80%	-8.20%	2.80%	1.80%	5.07%
南禅寺	10.40%	13.68%	-3.40%	9.60%	8.70%	6.40%	2.60%	7.77%
三十三間堂	23.90%	22.53%	15.40%	25.50%	23.30%	24.00%	24.90%	22.69%
平安神宮	9.90%	0.95%	1.10%	6.80%	10.00%	3.50%	9.60%	5.98%
高台寺	6.40%	6.00%	-10.30%	6.20%	6.10%	-0.70%	1.00%	5.24%
東寺	3.20%	2.83%	-8.00%	4.20%	3.50%	-8.10%	2.40%	4.60%
円覚寺	2.90%	-4.30%	-3.80%	-3.90%	-3.90%	-4.60%	-3.40%	4.11%
モンキーパーク	14.70%	10.13%	4.70%	15.50%	13.90%	3.30%	3.70%	9.42%
東福寺	10.20%	8.23%	0.90%	9.80%	11.90%	9.00%	10.50%	8.65%
西本願寺	6.70%	1.63%	1.20%	4.80%	2.80%	3.10%	8.30%	4.08%
絶対値平均	11.98%	11.01%	8.59%	11.91%	12.08%	9.91%	11.55%	11.00%

Type0 の各スポットの誤差を見ると、伏見稲荷大社の誤差が最も大きくなっている。伏見稲荷大社はシミュレーション結果が実績データの値を大きく上回ることによって誤差が生じている。金閣寺、三十三間堂、嵐山モンキーパークでも同様の傾向が見られる。逆に、ギオンコーナーや京都タワーはシミュレーション結果が実績データの値を下回ることによって誤差が生じていることがわかる。

4.2. パラメータチューニング

初期の目的地決定モデルでは、距離・人気度・類似度の重みを 0 か 1 で決定している。これらの重みを適切に調整することで、より実績データとの誤差が少なくしていくことが可能であると考えられる。また、このようなチューニング作業を行うことで、影響の大きい要因の特定が可能である。

Table 18：チューニング結果

		類似度=0	類似度=0.25	類似度=0.5	類似度=0.75	類似度=1
距離=0	人気度=0	17.408%	17.208%	17.098%	16.970%	17.005%
距離=0	人気度=0.25	13.816%	13.477%	13.222%	12.970%	12.815%
距離=0	人気度=0.5	12.334%	11.943%	11.644%	11.435%	11.240%
距離=0	人気度=0.75	11.510%	11.274%	11.053%	10.890%	10.811%
距離=0	人気度=1	12.119%	11.899%	11.690%	11.632%	11.485%
距離=0.25	人気度=0	20.323%	19.932%	19.822%	19.649%	19.436%
距離=0.25	人気度=0.25	16.305%	16.006%	15.765%	15.533%	15.382%
距離=0.25	人気度=0.5	13.305%	13.012%	12.721%	12.557%	12.500%
距離=0.25	人気度=0.75	11.389%	11.139%	11.038%	10.959%	10.902%
距離=0.25	人気度=1	11.529%	11.432%	11.196%	11.159%	10.974%
距離=0.5	人気度=0	20.320%	20.054%	19.832%	19.643%	19.528%
距離=0.5	人気度=0.25	16.315%	16.017%	15.786%	15.545%	15.335%
距離=0.5	人気度=0.5	13.323%	13.037%	12.795%	12.585%	12.524%
距離=0.5	人気度=0.75	11.353%	11.179%	11.001%	10.940%	10.908%
距離=0.5	人気度=1	11.542%	11.373%	11.259%	11.070%	11.052%
距離=0.75	人気度=0	20.273%	20.100%	19.776%	19.643%	19.515%
距離=0.75	人気度=0.25	16.290%	16.047%	15.765%	15.548%	15.365%
距離=0.75	人気度=0.5	13.355%	12.989%	12.772%	12.611%	12.463%
距離=0.75	人気度=0.75	11.392%	11.146%	11.015%	10.986%	10.968%
距離=0.75	人気度=1	11.527%	11.407%	11.236%	11.140%	11.067%
距離=1	人気度=0	20.247%	20.041%	19.819%	19.614%	19.498%
距離=1	人気度=0.25	16.266%	16.025%	15.789%	15.541%	15.465%
距離=1	人気度=0.5	13.340%	13.013%	12.824%	12.624%	12.479%
距離=1	人気度=0.75	11.333%	11.156%	11.053%	10.938%	10.907%
距離=1	人気度=1	11.503%	11.429%	11.272%	11.084%	11.046%

我々はまず、距離・人気度・類似度のそれぞれの重みを 0 から 1 まで 0.25 ずつ変化させ、最も誤差が少なくなる重みの組み合わせを探索した。パラメータチューニング結果を Table 18 に示す。ここで、黄色の枠で囲んだセルは上位 30%の値であり、赤色の枠で囲んだセルは上位 10%の値である。

チューニングの結果、最も実績データとの誤差が少なくなったパラメータの組み合わせは、距離=0、人気度=0.75、類似度=1 であり、誤差の平均値は 10.794%となった。Table 18 をみると、まず、距離についてはほとんど結果に影響を与えないことがわかる。人気度は、0.75 に近づくにつれて、誤差の値が減少していくが、人気度=1 になると、また少し誤差が大きくなる。このことから、人気度は 0.75 が最適なパラメータであることがわかる。類似度についても、1 に近づくにつれて、誤差が減少していくため、類似度=1 が最適なパラメータであることがわかった。

ここで、人気度は 0.75 に近づくにつれて誤差が減少していくが、1 になると誤差が大きくなることから、最適なパラメータは 0.75 付近に存在することがわかる。そこで、より最適な人気度を推定するため、距離を 0、類似度を 1 に固定し、人気度のみを 0.75 付近で 0.05 ずつ変化させる、チューニング作業を行った。結果を Table 19 に示す。

Table 19：人気度チューニング結果

		類似度=1
距離=0	人気度=0.55	11.065%
距離=0	人気度=0.60	10.990%
距離=0	人気度=0.65	10.910%
距離=0	人気度=0.70	10.845%
距離=0	人気度=0.75	10.848%
距離=0	人気度=0.80	10.786%
距離=0	人気度=0.85	10.815%
距離=0	人気度=0.90	10.949%
距離=0	人気度=0.95	11.030%

人気度のチューニングの結果、0.80 が最適な値であることが示された。

4.3. スポット別チューニング結果

また、チューニングした際に誤差が最小になる各パラメータの重みの組み合わせはスポットごとに異なる。スポットごとの最小の誤差と、その時の各パラメータの重みの組み合わせを Table 20 に示す。

結果を見るとそれぞれ最適なパラメータを設定することで、誤差の平均が 5.452% になった。また全体のチューニング結果では、距離=0、人気度=0.75、類似度=1 が最適なパラメータの組み合わせであったが、スポットごとに最適なパラメータは大きく異なることがわかる。

Table 20：スポット別チューニング結果

スポット名	最小誤差	距離	人気度	類似度
清水寺	7.424%	0	1	1
金閣寺	4.804%	0.25	0.5	1
二条城	4.764%	0.75	0.5	1
伏見稲荷大社	4.800%	0.25	0	0.5
京都御所	8.756%	0	0.25	0
ギオンコーナー	4.096%	0	1	0.25
銀閣寺	4.899%	1	0	0
錦市場	5.191%	0	0.25	1
八坂神社	2.990%	0.25	1	0
京都タワー	4.295%	0	0.25	0
南禅寺	6.612%	0	1	0
三十三間堂	13.831%	0	0.5	1
平安神宮	3.967%	0	1	0.25
高台	3.957%	0	1	1
東寺	4.321%	0.75	1	1
下鴨神社	3.510%	0	1	1
嵐山モンキーパーク	6.029%	1	0.5	1
東福寺	6.896%	0	1	0
西本願寺	2.451%	0	1	0
平均	5.452%	0.224	0.671	0.526

4.4. 出身地域別チューニング結果

チューニングした際に誤差が最小になる各パラメータの重みの組み合わせは地域ごとにも異なる。地域ごとの最小の誤差と、その時の各パラメータの重みの組み合わせを Table 21 に示す。

Table 21：地域別チューニング結果

地域	最小誤差	距離	人気度	類似度
africa	11.878%	1	0.75	1
asia	10.817%	1	0.75	1
europa	7.584%	0	0.75	0.75
latin america	11.859%	0.25	0.75	0.75
middle east	11.852%	1	0.75	0.75
north america	9.327%	0	0.75	1
pacific	11.145%	0	0.75	1
平均	10.638%	0.464	0.750	0.893

それぞれ最適なパラメータを設定することで、誤差の平均が 10.638% になり、全体のチューニング結果である 10.794% と比べてあまり差がないことがわかった。また全体のチューニング結果では、距離=0、人気度=0.75、類似度=1 が最適なパラメータの組み合わせであったが、人気度と類似度に関しては、ほとんど同様の結果となった。一方、距離については、考慮した方がいい地域と考慮しない方がいい地域に別れた。

5. 考察

5.1. シナリオ別シミュレーション結果考察

シナリオ別シミュレーション結果では、type0(距離・人気度・類似度重視型)の結果がもっとも実績データとの誤差が少なくなることが示された。この結果から、実績データとの差を少なくするには、従来の人気度や距離だけではなく、観光客と観光地の類似度も考慮する必要があることが明らかになった。

5.2. パラメータチューニング結果考察

チューニングの結果について見ると、距離は考慮せず、類似度は考慮した方が全体の結果はよくなることが示された。人気度については、0.75 に近づくにつれて誤差の値が減少していくが、人気度=1 になると、誤差が大きくなった。そこで、さらにチューニングをした結果、最適なパラメータは 0.8 であることがわかった。よって、距離=0、人気度=0.8、類似度=1 が最適なパラメータの値となり、誤差の平均値は 10.786% となった。シナリオ別のシミュレーション結果と同様に類似度を考慮することで全体の結果に改善が見られることが明らかになった。

5.3. スポット別チューニング結果考察

スポットごとのチューニングの結果を見ると、スポットごとに最適なパラメータの値は大きく異なることがわかる。

まず、距離について、全体の平均では距離を考慮しない方が、誤差が小さくなるという結果が出たが、スポットごとに見ると、二条城、銀閣寺、東寺、嵐山モンキーパークでは、距離を考慮した方がいいという結果となった。これらのスポットは、比較的京都の外周

に立地しているため、観光客の意思決定に距離要素が影響する可能性がある。そのため、スポット間の距離が比較的近い場合は距離による影響は小さく、逆に、スポット間の距離が遠い場合は距離を考慮する必要があるということが考えられる。

人気度は、ほとんどのスポットで考慮した方がいいという結果になった。しかし、伏見稲荷大社では考慮しない方が、誤差が小さくなっている。伏見稲荷大社は、レビューの数が15,131件と全スポット中最も多く、type0の誤差の結果においても、シミュレーション結果が実績値を大きく上回ること、全スポット中最も誤差が大きくなっている。

一方で、ギオンコーナーについても、人気度を考慮しない方が、誤差が少なくなるという結果が出ている。ギオンコーナーは実績データ上位のスポットであるが、レビュー数が380件と少なく、type0の誤差の結果においても、シミュレーション結果が実績データを下回ること、大きな誤差を生じている。これらの結果から、レビューの多すぎる場所や、逆に少なすぎる場所では、人気度がマイナスに働くことから、人気の過大評価や過小評価を緩和するために、レビュー数の平方根を用いたり、一定の範囲でクラス分けをしたりするなど、調整を行うことで、精度を高められることが考えられる。

類似度は、全体の平均では考慮した方がいいという結果になったが、スポットによって、考慮した方がいい場所としない方がいい場所が別れている。金閣寺や、二条城、三十三間堂など、京都のスポットの中でも一般的な知名度が高く、特徴的なスポットにおいて、類似度を考慮する必要があると考えられる。逆に、類似度が低い時に最小誤差になるスポットは、京都御所、銀閣寺、八坂神社、京都タワー、南禅寺、平安神宮、東福寺、西本願寺の8つであった。これらのスポットは、レビュー数の平均値が553件であり、これは19スポットの平均値である1,760件に比べると、少ないことがわかる。そのため、類似度を算出するにはレビュー数が十分でなかった可能性が考えられる。

また、これらのスポットを最小誤差のパラメータ値ではなく全体的にみると、類似度を加えたことで、誤差の平均値が増えたものは、銀閣寺、南禅寺、東福寺、西本願寺の4つのスポットであった。これらのスポットは、パラメータ表示の際に用いた15のパラメータ値の標準偏差が、4スポットを除いた標準偏差の平均値0.096に比べて、0.064と、小さい傾向であることがわかった。つまり、これらの4スポットでは、特徴の差異が大きいとは言えず、類似度の効果が小さいと考えられる。

5.4. 地域別チューニング結果考察

出身地域ごとのチューニングの結果を見ると、人気

度は全体の平均である人気度=0.75に近い値となった。全ての地域で考慮した方がいいという結果となった。このことから、類似度をモデルに組み込むことで、出身地域で異なる観光客の嗜好の違いを調整し、誤差を小さくしていると考えられることができる。

距離についてみると、ヨーロッパ、北アメリカ、オセアニアが距離を考慮しない方が、誤差が小さくなる結果になった。最もスポット間の距離の平均値が大きい嵐山モンキーパークの入込客の実績値をみると、全体平均は8.6%の訪問率であるのに対し、ヨーロッパは13.40%、オセアニアは14.90%、北アメリカは、14.90%と大きく平均値を上回る実績値となっている。これらの地域の観光客は、距離をよりも他の要素を重視する傾向があり、それがパラメータの値にも現れていると考えることができる。

6. まとめと今後の課題

本研究では、まず観光地における外国人旅行客の行動モデルのパラメータ推定を目的に、旅行情報サイトにある各観光地に対する英語のレビューに対しテキストマイニングを行い、外国人旅行客が認識する各観光地の特徴と、出身地域ごとの外国人旅行客の特徴を推定した。推定した特徴をもとに、パラメータを求め、出身地域で分類した外国人旅行客のレビューと、各観光地のレビューとの類似度を推定した。

そして、分析対象を観光目的の訪問客が多い京都市に絞り、目的地決定に距離、人気度、観光客と観光地との類似度(嗜好)を考慮した、外国人旅行客の京都市滞在中の観光行動モデルを作成した。既存研究のモデルでは、観光目的地決定のみであったが、本研究では、移動時間、滞在時間、京都への訪問回数も考慮し、より現実に近い観光行動モデルを作成した。

これらのモデルをもとにエージェントベースシミュレーションを行い、各観光地の年間集客数を推定した。シミュレーション結果と、京都市の外国人旅行客の実績データをもとにモデルを検証した結果、従来の観光行動モデルでは考慮されていなかった旅行者嗜好と観光スポットとの類似度を観光行動モデルに組み込むことで、より実態に近いモデルの作成が可能であり、その際の類似度は、旅行口コミサイトに投稿されるレビューのテキストマイニングを行うことで推定することが可能であることを明らかにした。

次に、実績データとの誤差が少なるようにパラメータのチューニング作業を行った結果、距離=0、人気度=0.8、類似度=1.0のモデルが最も実績データとの差が小さくなるという結果になった。

またスポット別、地域別のチューニング結果から、最適なパラメータの組み合わせはスポットごとに異なり、最適なパラメータを調整することで、より実績デ

ータとの誤差が少なくすることが可能であることを示した。そして、それぞれのパラメータ値には、考慮すべきスポットやしないスポットが存在し、違いを考察すると、パラメータごとにそれぞれの特性があることが明らかになった。

距離は遠いスポットにおいては考慮すべきだが、スポット間の距離が一定以内に収まるのであれば、考慮しない方が実績値との誤差が小さくなった。人気度は、レビュー数が多すぎるスポットや、逆に少なすぎるスポットでは、考慮しない方が実績値との誤差が小さくなったことから、調整が必要であることが明らかになった。類似度に関しては、レビュー数が少ないスポットや、特徴があまりないスポットでは考慮しない方が実績値との誤差が小さくなったことから、レビュー数を一定以上に増やし、特徴のある地域に限って適用することが望ましいことが明らかになった。

以上のことから、パラメータごとにそれぞれの特性があり、シミュレーションモデルを設計する際には、随時パラメータの特徴にあった調整をすることが必要であるということも明らかになった。

また、パラメータの性質に合わせ、観光客のモデルを設計すると、逆に観光客の行動パターンが見えてくる。例えば、京都を訪れる観光客が考える距離は、一定時間内ならば多少の差は気にせず、人気に関しても、過剰に人気なスポットや逆に人気が非常に少ないスポットよりも、一般的な人気のスポット間の人気差を意識することがあり、類似度に関しても、特徴の判別が難しいスポット間では嗜好に基づく選別が行われていない可能性があることが明らかになった。

本研究で行ったシミュレーションは DMO などにおいて、地域の観光スポットを分析する際にも、活用が見込まれる。例えば、本研究ではシミュレーションから、各スポットの入込客を推定し、実績データとの誤差を比較したが、誤差にはシミュレーション結果が実績値を上回ることによって誤差が生じる場合と、逆にシミュレーション結果が実績値を下回ることによって誤差が生じる場合がある。この時、シミュレーション結果が実績値を上回ることによって誤差が生じる場合は、口コミ評価から推定した結果では、より多くの入込客が見込めるはずであるが、実際には推定結果よりも入込客が少ないことになる。

本研究のシミュレーション結果では、伏見稲荷大社や、三十三間堂で、シミュレーション結果が実績値を上回って誤差が生じている。伏見稲荷大社は、tripadvisor の京都のスポットランキングで 1 位であるが、実績値では 4 位であり、三十三間堂はスポットランキングで 4 位であるが、実績値では 12 位である。このように、ネット上での口コミ評価は高いが実績値が伴っていないスポットなどは、入込客を増やす可能性が高く、ポテンシャルのあるスポットであり、例えば、

ネット以外のメディアなどの露出を増やすなどの PR 活動や、マーケティング活動に注力することで大きな成果が見込まれることが考えられる。

逆に、シミュレーション結果が実績値を下回ることによって誤差が生じる場合は、口コミ評価から推定した結果では、入込客はあまり見込めないはずであるが、実際には推定結果よりも入込客が多いことになる。本研究のシミュレーション結果では、ギオンコーナーや、京都タワーで、シミュレーション結果が実績値を下回って誤差が生じている。ギオンコーナーは、tripadvisor の京都のスポットランキングで 169 位であるが、実績値では 6 位であり、京都タワーはスポットランキングで 57 位であるが、実績値では 10 位である。

これらのスポットは、観光客の確保に成功しているスポットであり、マーケティングの際に参考にすべきスポットでもある。ギオンコーナーでは、ほかのスポットとは異なり、日本の誇る伝統芸能を約 1 時間で体感できる体験型の観光スポットであり、外国人旅行者にとっては、一度は訪れたい魅力的なスポットとなっている。京都タワーは、ほとんどの神社や寺などが夕方方には拝観受付が終了するのに対し、21 時まで営業をしており、食事をすることもできる。このように夜間まで営業し、消費活動を喚起することで、さらなる顧客満足度と経済を向上させようという動きは、ナイトタイムエコノミーと呼ばれ、今後日本の、インバウンド消費額拡大に向け注目されている。このように実績値がシミュレーション結果を上回るスポットからは、観光客の確保に向けたマーケティング手法や戦略を学ぶことができると考えられる。例えば、京都の寺院の、夜間拝観の頻度を増やすことや、地元の飲食業との夜間限定のコラボレーション企画などの可能性を検討する余地がある。

また、観光スポットの問題を「人気」、「交通」、「特徴」に分類し、解決すべき課題を分析することも出来る。例えば、人気度が比較的低いスポットでは、「インスタ映え」などの SNS の拡散によるマーケティング戦略などを用いて、認知度を高めることで、入込客の増加が見込まれる。また、京都外周にある観光スポットでは、バスの運行ルートを整備し、交通アクセスをよくすることで、入込客の増加が見込まれる。また、特徴が比較的に少ないスポットは、夜間でも観光できる制度を整えたり、その場所でしかできない体験ができるスポットにしたりするなど、周りのスポットとの差別化を図ることで、入込客の増加が見込まれる。このように、それぞれのスポットの問題を「交通」、「人気」、「特徴」から分析することで、その場所に合った適切な問題解決を行うことが可能であると考えられる。

今後の課題として、本研究では口コミのテキストマイニングを行うことで、観光スポットと観光客の類似

度を推定したが、口コミだけでなく、SNS、ブログ、ガイドブックなどの様々なメディアについての分析を行うことでより正確なパラメータの推定が行えることが考えられる。また、本研究では英語の口コミのみを分析したが、近隣諸国の言語である中国語、韓国語についても分析が必要になると考えられる。さらには、テキストだけでなく、画像や動画などを複合的に分析することも有効であることが考えられるため、それらについても今後検討が必要である。

また、本研究では実績データの存在する京都市に絞ってシミュレーションを行ったが、都市型観光地の東京や大阪、地方の観光地などの様々な観光地においても、本モデルの応用が可能かどうか、さらなる検証を行う必要がある。ただし、現状では実績データの存在する地域は限られており、本研究のように限られた範囲での分析になることが考えられる。さらなる検証のためには、地方自治体や国が保有するデータの活用が必要である。このようなデータの活用にはプライバシーの問題など、多くの問題が存在するが、官民データ活用推進基本法²²⁾が制定されるなど、環境の整備は着々と進められてきている。それに伴って、さらなる分析が可能になるため、今後の研究の発展が期待される。

外国人旅行者は今後も増加することが見込まれるが、2,000万人を突破した2017年において、すでに宿泊施設や、公共交通機関のキャパシティの限界など、様々な問題が浮き彫りとなっている。2020年までに政府の目標とする4,000万人を突破するとなると、問題はさらに深刻化することは明らかである。これらの問題に対し、外国人旅行者の動向を的確に理解し、適切な対応が求められる。また、東京オリンピックを控えた現在の好調な観光市場を一過性のものにしないためにも、継続的な分析や研究が必要である。観光立国日本に向け、今後の研究の進展が期待される。

7. 参考文献

- 1) JNTO 日本政府観光局, <https://www.jnto.go.jp/jpn/>
- 2) 観光庁, 明日の日本を支える観光ビジョン, http://www.mlit.go.jp/kankocho/topics01_000205.html
- 3) President online 「訪日客は迷惑」京都を悩ます"観光公害", <http://president.jp/articles/-/22664>
- 4) Lew, M. and McKercher, B. : Modeling Tourist Movement : A Local Destination Analysis, *Annals of Tourism Research*, 33(2), 403/423, (2006)
- 5) 高橋, 五十嵐 : 観光スポットの魅力度を考慮した観光行動分析と入込み客数の予測, *土木計画学研究論文集 No.8*, 233/240, (1990)
- 6) 鈴木優伽, 伏見卓恭, 斎藤和巳, 風間一洋, 回遊行動モデルに基づく重要観光スポット抽出法, *情報処理学会第 77 回全国大会* 2-259/2-260, (2015)

- 7) 大久保立樹, 室町泰徳, 旅行ガイドブックと口コミの言語解析による訪日外国人の観光地イメージに関する研究, *都市計画論文集* 49(3), 573/578, (2014)
- 8) Tripadvisor, <https://www.tripadvisor.jp>
- 9) 樋口耕一 : テキスト型データの計量的分析—2つのアプローチの峻別と統合—, *理論と方法*, vol.19, No.1, 101/115, (2004)
- 10) 樋口耕一 : KH Coder Index Page , <http://khc.sourceforge.net>
- 11) KHcoder を用いた研究事例, <http://khc.sourceforge.net/bib.html>
- 12) 室谷正裕, 観光地の魅力度評価-魅力ある国内観光地の整備に向けて-, *運輸政策研究* 1(1), 14/24, (1998)
- 13) 京都観光総合調査 H28,
<<http://www.city.kyoto.lg.jp/sankan/cmsfiles/contents/0000222/222031/28chousa.pdf>>
- 14) 田沼英樹, 出口弘 : エージェントベース社会シミュレーション言語 SOARS の開発, *電子情報通信学会論文誌 D*, Vol.J90-D, No.9, 2415/2422, (2007)
- 15) SOARS Project, <http://www.soars.jp/>
- 16) 金谷泰宏 : 新型インフルエンザに対するパンデミック対策プログラムとプロジェクト分析, *オペレーションズ・リサーチ : 経営の科学*, 53(12), 667/671, (2008)
- 17) kyotrip
<<https://kyotrip.jp/all-list/>>
- 18) ギオンコーナー公式サイト
<<http://www.kyoto-gioncorner.com/global/ja.html>>
- 19) ブログ : 錦市場へ行こう !
<<http://www.kuraryoko.com/kyonisiki.html>>
- 20) 京都タワーへ行ってみよう !
<<http://www.kuraryoko.com/kyototower.html>>
- 21) オフィシャルサイト京都観光ナビ
<<https://kanko.city.kyoto.lg.jp/detail.php?InforKind-Code=1&ManageCode=9000001>>
- 22) 官民データ活用推進基本法
<<http://www.kantei.go.jp/jp/singi/it2/senmon/dai9/siryoul.pdf>>