



卒業研究報告書

平成29年度

研究題目

観光者が持つ属性間の差異を用いた
観光地推薦の検討

指導教員 上野秀剛 准教授

氏名 志賀圭駿

平成30年2月16日 提出

奈良工業高等専門学校 情報工学科

観光者が持つ属性間の差異を用いた

観光地推薦の検討

上野研究室 志賀圭駿

特定地域の観光情報を提供するためのアプリケーション、ご当地観光アプリは膨大なコンテンツから構成されており、ユーザが自身の嗜好に合うコンテンツを発見するのは難しい。このような問題の改善方法として、ユーザの嗜好が反映されていると思われるデータ群に基づき嗜好を推測する推薦という技術がある。しかし、現在の推薦に広く用いられる協調フィルタリングは新規のユーザに対しての推薦が困難であり、1度限りの使用を想定しているご当地観光アプリへの適用は難しい。そこで著者は、推薦対象ユーザが入力する性別・年齢といった属性毎の嗜好から発生する、訪れるエリアの差異を用いての推薦を検討している。本研究ではユーザが持つ属性間の差異を用いた推薦が、ご当地観光アプリにおける推薦に有効であるかの検討を目的とする。その手段として、属性の違いによって訪れるエリアにどの程度の差異が発生するかを検証する。提案手法では属性が異なる2群のユーザに対して、緯度経度を小数点以下4桁で四捨五入した値を軸とするメッシュを訪れたユーザの割合を求め、差を取ることで属性間の差異を算出する。実験の結果、提案手法を元にした推薦は、単にある属性のユーザが頻繁に訪れる場所をその属性のユーザに対して推薦することに比べて、意外性のある推薦が可能である事が分かった。また、ユーザの入力する属性を増やせば、細分化した嗜好が顕在化し、よりパーソナライズされた推薦を行える可能性を示した。結論として、提案手法により算出した属性間の差異は、特に意外性の観点から1度限りの使用を想定したご当地観光アプリにおける推薦に有効であることが期待できる。

目次

1	はじめに	2
2	関連研究	4
2.1	GIS	4
2.2	推薦手法	4
2.3	ユーザ属性を用いた推薦	5
3	提案手法	6
4	実験	9
4.1	実験環境	9
4.2	提案手法の適用	11
5	結果と考察	12
5.1	可視化による分析	12
5.2	定量的な分析	17
5.2.1	分析方法	17
5.2.2	結果	18
5.3	まとめ	20
6	おわりに	21
	謝辞	22
	参考文献	23

1 はじめに

個人が所有する携帯端末の普及と共に，特定地域の観光情報を提供するためのアプリケーション，ご当地観光アプリが全国で多数配信されている [1]．ご当地観光アプリをはじめとする観光情報サービスは，観光スポット，飲食店，宿泊場所，口コミなど，多種多様なコンテンツから構成されている．しかし，観光情報サービスが有する膨大なコンテンツの中から，ユーザが自身の嗜好に合うコンテンツを発見するのは難しい．このような問題を改善する技術として推薦がある．推薦は，ユーザの嗜好が反映されていると思われるデータ群（たとえば商品の購入履歴や，視聴した音楽の履歴など）に基づいてユーザの嗜好に合うものを推測する．推薦はオンライン通販など，多量の情報からユーザにとって有用だと考えられる情報を絞り込む用途に広く用いられており，膨大なコンテンツを扱う観光情報サービスを実現するうえでも重要である [2]．

現在の推薦に広く用いられる手法のひとつに協調フィルタリングがある．これはあらかじめ蓄積された，ユーザの嗜好が反映されていると思われるデータ群からユーザの潜在的な嗜好を推測し，推薦を行う手法である．ご当地観光アプリにおけるユーザの嗜好が反映されていると思われるデータ群としてはユーザの移動履歴が挙げられる．しかし，一度限りの使用を想定しているものが多いご当地観光アプリはユーザの移動履歴を長期に渡って収集できない．よって，ご当地観光アプリには協調フィルタリングの適用が困難である．このような，新規ユーザに対する推薦が困難であることは cold-start 問題として知られている [8]．

著者は推薦対象ユーザの移動履歴を使用せず，ユーザが入力する性別・年齢といった属性と他のユーザが過去に残した移動履歴を用いての推薦を検討している．一般に店舗・観光地は主たる客層を持つことがあるため，属性が同じユーザがよく訪れると考えられる．例として特に若い女性を対象とした店舗や，特に遠方地からきた（すなわち，初来訪の可能性が高い）観光客に人気の高い観光地などが挙げられる．よって，属性が同じユーザはよく訪れる店舗・観光地が近いいため，似た嗜好を持っていると考えられる．対して，属性が異なるユーザはよく訪れる店舗・観光地が違うため，異なる嗜好を持っていると考えられる．

以上から，属性によってユーザをグルーピングすることで各属性を持つユーザの嗜好が顕在化し，グループ間で訪れる観光地や地域に差異が発生すると考えられる．著者はこの差異を用いて，ある属性に対して近い属性（例：若い男性に対して高齢の男性）がよく訪れる観光地を推薦すれば，ある程度パーソナライズされた観光地推薦が可能であると推測する．ここで同じ属性がよく訪れる観光地を推薦しないのは，ユーザにとって未知の可能性が高い観光地を推薦するためである．ご当地観光アプリの宣伝は基本的に推薦対象のエリア内において行われているため，ユーザは推薦対象のエリア内でアプリをインストールする可

能性が高い。つまり、ユーザがアプリを初めて起動する時点で既に観光は始まっており、観光地の事前調査も終わっているため、同じ属性がよく訪れる観光地は既知である可能性が高い。ユーザにとって既知の観光地は、ユーザが訪問するかどうかの判断が終わっているため、推薦する意味が無い。対して、ユーザにとって未知の観光地を推薦すれば、興味をもったユーザが訪れる可能性がある。なお、ユーザにとって未知であると考えられるアイテムの推薦は、一般に意外性のある推薦であるとされる[7]。また、ある属性に対して遠い属性（例：若い男性に対して高齢の女性）がよく訪れる観光地を推薦しないのは、嗜好が大きく異なり、未知の観光地であってもユーザの興味が薄いと考えたためである。

本研究では、以上の推薦手法がご当地観光アプリにおける推薦に有効であるかを検討する。その手段として、属性の違いによって訪れるエリアにどの程度の差異が発生するかを検証する。また、本研究では属性の異なるユーザ間で実際に発生する差異を、奈良県が配信するご当地観光アプリ「ならたん」において収集したデータを用いて検証する。

2 関連研究

2.1 GIS

Geographic Information System (GIS)とは地理的位置を手がかりに、位置に関する情報を持ったデータ(空間データ)を総合的に管理・加工し、視覚的に表示し、高度な分析や迅速な判断を可能にする技術である[3]. 近年はこのGISを題材とした研究が盛んに行われている.

佐藤らはユーザのスマートフォンに搭載されたGPS機能を用いた回遊調査から得たデータより、カーネル法を用いた時空間分析を行った[4]. 50%カーネル密度圏の重心を滞在地点とする視覚化手法を提案し、ユーザが持つ属性ごとに回遊の特徴が異なることを明らかにした. 原らはGPSロガーを用いた訪日個人旅行者の観光周遊行動調査とその分析を行った[5]. 結果として、観光庁などの統計調査では把握できなかった、訪日個人旅行者の滞在時間・訪問順序などの詳細を明らかにした.

これらの研究は取得したデータを視覚化することで目視による分析を行っているが、本研究ではデータを元にユーザが持つ男性、女性など、同時に持ち得ない属性間で発生するの差異を数値的な検証によって明らかにすることが目的である.

2.2 推薦手法

本稿における推薦とは、膨大な情報の中からユーザにとって有用だと考えられるアイテムをアルゴリズムによって選び出し、ユーザに提示する技術である. 既存の推薦手法の多くは協調フィルタリングと内容ベースフィルタリングに分類できる.

協調フィルタリングはユーザの潜在的な嗜好を推測する手法である. 推薦対象ユーザが知らず、かつ推薦対象ユーザと類似した嗜好を持つユーザが興味のある情報を潜在的に興味があると判断し、推薦を行う. 篠田らは推薦対象のユーザと類似の行動パターンを持つ他のユーザの行動履歴を参照して協調フィルタリングを適用することで、ユーザが訪れたことの無い地域でも効果的な推薦を行っている[6]. しかし、協調フィルタリングの根本的な問題として、行動パターンの蓄積がない新規ユーザに対して適切な推薦を行うことが難しいというcold-start問題[8]があり、一度限りの使用が想定されたご当地観光アプリへの適用は難しい. 本研究の結果に基づく推薦では、ユーザがアプリの利用開始時に入力した属性情報を元に推薦することを想定しているため、新規ユーザに対するcold-start問題は起こり得ない.

内容ベースフィルタリングはユーザの嗜好と関連性の高い内容を持つコンテンツを推薦する手法である. 新規ユーザであっても嗜好情報が得られれば精度

の高い推薦が可能であるという利点があり，楽曲推薦[9]などで活用されている．しかし，サービス提供側が各コンテンツの性質を表す特徴ベクトルを設定しなければならないため，観光地の推薦に対して適用するには多大な労力を要する．本研究の結果に基づく推薦ではある属性に対して推薦するエリアを決定し，その後，推薦するエリア内に存在する観光地を推薦するという流れを想定している．よって，サービスの提供側の準備が必要なデータとしては観光地の緯度経度，名称，ご当地観光アプリとして最低限の詳細のみである．推薦を行うために特別なデータを用意する必要がなく，サービス提供側の労力が少ないため，手法の適用は比較的容易である．

2.3 ユーザ属性を用いた推薦

推薦を行うための方法としてユーザの属性情報を用いる手法が複数存在する[10, 11]．小野らはインタビュー形式，WEB形式でアンケートを取ることで，被験者の属性と状況，映画への評価からベイジアンネットモデルを作成した．また，それに基づいてユーザが自らの属性を入力するとおすすめの映画とその理由を提示する映画推薦システムを提案した[10]．精度評価の結果，ベースライン予測や協調フィルタリングに比べて良い予測精度を示した．本村らは嗜好やヘアスタイルのデータベースに基づくヘアスタイルの推薦において，ユーザの属性を考慮する協調フィルタリングがユーザの属性を考慮しない協調フィルタリングに比べて推薦精度が高く，有用であることを示した[11]．

これらの研究では，ユーザの属性に基づいて精度の高い推薦を行っている．しかし，これらの推薦システムは推薦を行うアイテムについて事前に大量のデータを取得せねばならず，サービス提供側は多大な労力を要する．また，これらの推薦システムは大規模なモデル，もしくはデータベースを構築しており，その内部にはユーザが入力する属性以外のデータが多数含まれている．よって，モデル，データベース内のデータを更新せずに使用した場合，流行りのヘアスタイルや新しい映画に対応できず，そのようなアイテムの推薦は難しくなると考えられる．これは観光地推薦においても同様であり，上記の研究で提案された手法では新しくできた観光地や今流行りの観光地を推薦することは難しいと考えられる．よって長期の利用を想定する際には再び調査が必要であり，継続的に使用する際のランニングコストが大きい．本研究の結果に基づく推薦ではユーザの位置情報と属性，観光地の緯度経度のみから推薦を行うことを想定している．ご当地観光アプリの初回起動時に，属性情報をユーザが入力すれば推薦に必要なデータが取得できるため，サービス提供側の労力は小さい．また，日時の新しい位置情報のみを用いることで，時流に沿った推薦が可能であると考えられる．

3 提案手法

本章ではユーザが入力した属性の違いによって訪れるエリアにどの程度の差異が発生するかを検証する手法を提案する．分析に必要なデータはユーザが入力した1種類以上の属性とご当地観光アプリの使用中に携帯端末から発信された位置情報とし，それぞれの位置情報はどのような属性のユーザが発信したのかが分かるものとする．

提案手法は属性が異なる2群のユーザに対して，ある地点を訪れたユーザの割合を求め，その差を取ることで属性間の差異を算出する．全ユーザの集合 $U = \{u_1, u_2, \dots, u_x, \dots, u_n\}$ において，各ユーザは属性の集合 $P = \{p_1, p_2, \dots, p_y, \dots, p_m\}$ を持つ．なお， n は全ユーザの人数， m はユーザが持つデータの数を示しており， u_x は U におけるいずれかのユーザ， p_y は P におけるいずれかの属性を表している．提案手法の概要を図3.1に示す．

図3.1では例として，比較する属性 (p_1) を性別として，男性 (M) と女性 (F) の2群 (U_M, U_F) を比較している．図3.1ではユーザが各メッシュに訪れた割合を格納したマップを男性と女性について作成し，2つのマップについて差を取ることで男性と女性の差異を算出している．具体的には図3.1の左側に示すように，2つのマップ内で緯度経度が同じメッシュについて，まずメッシュに訪れた男性の位置情報の数 $Gps(U_M, lon, lat)$ を男性の総数 $|U_M|$ で割る ($mesh(U_M, lon, lat)$)．そして同様に算出した $mesh(U_F, lon, lat)$ との差を取ることで差異を算出する．以下に提案手法の処理手順を述べる．

1. ユーザが携帯する端末から発信された位置情報の緯度経度を小数点以下4桁で四捨五入する（以後，本稿における位置情報とは特に指定がない限り四捨五入を行った後のものとする）．これは，提案手法を用いて属性間の差異を算出する際に，属性ごとのユーザが訪れた大まかなエリアを参照するためである．なお，奈良市における緯度経度0.001度はそれぞれ約79.9m，約111.3mである．
2. 1ユーザが2つ以上の同じ緯度経度を示す位置情報を残している場合，1つを残して他の位置情報を削除する．この操作は1ユーザの影響でそのユーザが属する属性全体の嗜好が歪むことや，ホテルなどの長期滞在を前提とした場所に位置情報が集中することを防止するために行う．
3. 比較を行うため，1つ以上の p_y を選択する．選択した p_y の値 v が同一な u_x の集合であるユーザ群 U_g は以下のように定義される．

$$U_g = \{u_x | p_y = v\} \quad (3.1)$$

ここで， g は選択した属性 p_y の値 v の組合わせを表す．例として属性 p_1 が性別，属性 p_2 が年齢の時，男性 (M) の高齢者 (H) であるユーザ群 U_{MH} は以下

マップ中の経度135.751,
緯度34.749に位置するメッシュ
($mesh(135.751, 34.749)$)

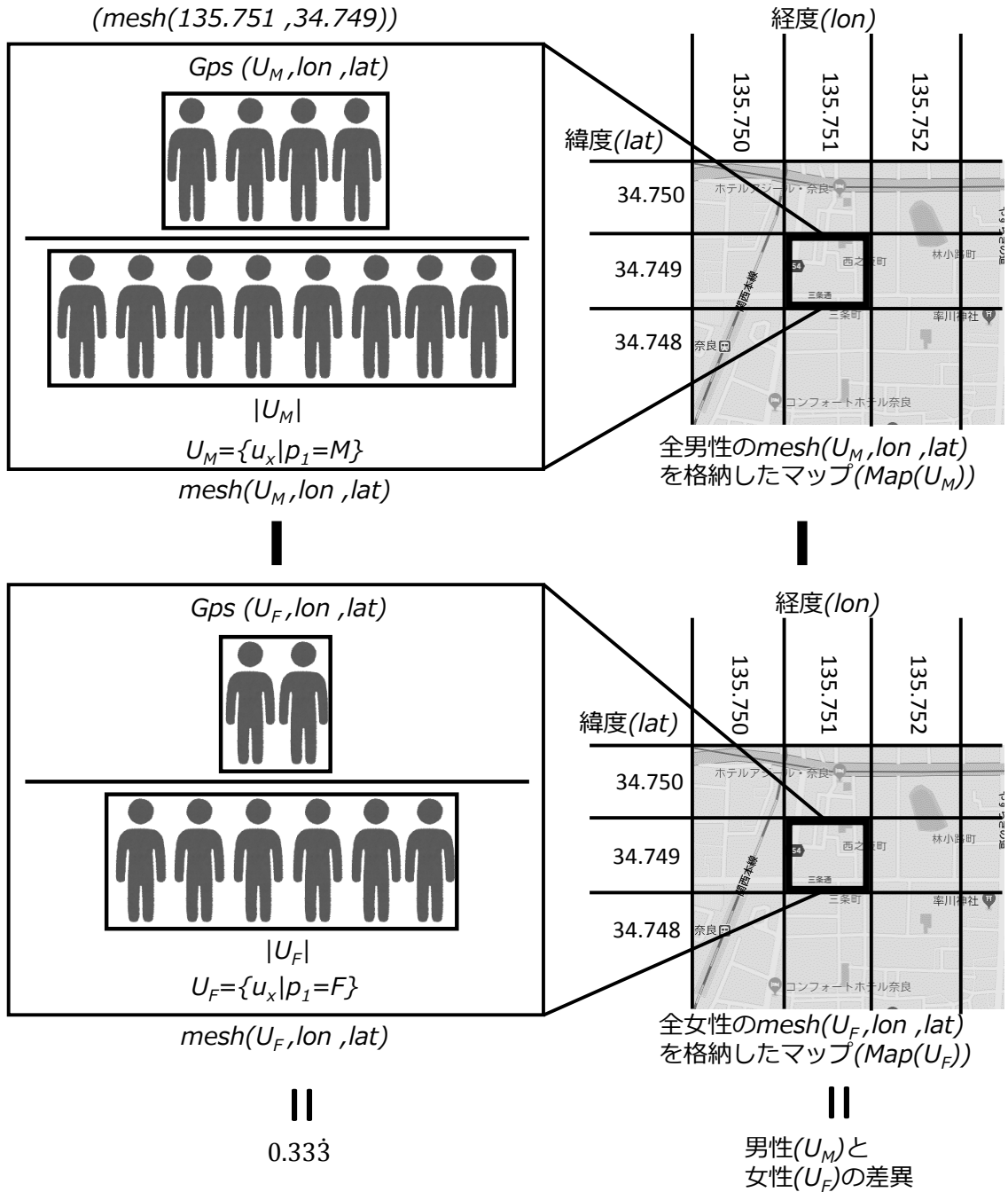


図 3.1 提案手法の概要

のように表される。

$$U_{MH} = \{u_x | p_1 = M, p_2 = H\} \quad (3.2)$$

なお $U_M = \{u_x | p_1 = M\}$ と $U_H = \{u_x | p_2 = H\}$ など、 y の値が異なる 2 群は理論上比較可能であるが、本稿では実施しない。 U_M と U_H を比較する場合、 U_{MH} は両属性に存在するため、結果として「 U_M 、 U_H 中で、 U_{MH} ではないユーザ」の差異が出力され、解釈が非常に難しくなるためである。

4. 選択した 2 群のユーザに対して、緯度経度を小数点以下 4 桁で四捨五入した値を軸とし、長方形のメッシュに区切った $Map(U_g)$ を作成する。各メッシュ

に存在する位置情報の数を $Gps(U_g, lon, lat)$ とする．ここで小数点以下4桁で四捨五入した値を使用するのは，メッシュがこれ以上に小さければユーザがよく訪れる場所を正確に検出できないおそれがあるためである．これはご当地観光アプリが移動履歴を送信する間隔として1分から5分程度を想定しており，ユーザの詳細な行動が把握できないことに起因する．

5. 2群それぞれの $Map(U_g)$ に含まれる全メッシュに対して，対応する p_y を持つユーザ数で除算を行い，各メッシュに U_g が訪れた割合 $mesh(U_g, lon, lat)$ を算出する．この値を用いることで，比較するユーザ数の偏りを考慮せずともよくなり，人数の偏りが大きいと思われる細分化された属性の比較も可能となる．

$$mesh(U_g, lon, lat) = \frac{Gps(U_g, lon, lat)}{|U_g|} \quad (3.3)$$

6. 2群のユーザから以下に定義する属性間の差異 $Diff(U_g, U_{g'})$ を算出する．

$$Diff(U_g, U_{g'}) = Map(U_g) - Map(U_{g'}) \quad (3.4)$$

属性間の差異が大きいメッシュは $Diff(U_g, U_{g'})$ の絶対値が大きくなる．また，値が正の場合は U_g が $U_{g'}$ に比べて頻繁に訪れるメッシュであり，値が負の場合はその逆である．

提案手法では，上記の処理により推薦対象のエリアにおいて2群のユーザが訪れる割合の差をとることで，属性の違いによって訪れるメッシュにどの程度の差異が発生するかを算出する．これにより，属性に基づいた推薦がご当地観光アプリにおける推薦にどの程度有効であるのかを明らかにできる．

4 実験

本研究ではご当地観光アプリ「ならたん」において収集したデータに提案手法を適用し、属性間の差異を検証する。

4.1 実験環境

ならたんは、2017年の1月から2月にかけて奈良県が実証実験を実施したご当地観光アプリであり、主に奈良市の観光情報が掲載されていた。2017年12月14日時点では既に正式版が配信されているが、掲載される観光情報の違い等から、本稿では実証実験時のデータのみを使用する。ユーザは自身が所有する携帯端末へアプリをインストールした後、利用開始時に自身の属性を入力し、観光を行う。その間、アプリが動作していれば、ユーザが所有する携帯端末から位置情報が5分毎に発信される。

表4.1に属性と属性値の変数を示す。ユーザが入力する属性は居住地($y=1$)、年齢($y=2$)、性別($y=3$)である。属性値について、本稿では居住地を近畿圏内In、近畿圏外Out、年齢を全ユーザの平均年齢以下Low、以上High、性別を男性Male、女性Femaleの2値として用いる。これは属性値を多値データとした場合、各属性の人数が極端に少なくなるためである。また、本研究ではユーザとして観光者を想定しているため、奈良県に住むユーザは分析の対象外とする。

アプリの総ダウンロード数は4053件である。本研究ではその内、居住都道府県が奈良県以外且つ3種類の属性を全て入力しており、位置情報が1点以上存在する378名のユーザを対象に分析を行う。また、本稿では3種類の属性(U_{ILM} など)では比較を実施しない。これは1ユーザ群あたりの人数が少なくなることや、解釈が難しいことが理由として挙げられる。表4.2に属性値ごとの人数を示す。表中の g には表4.1における属性値名の頭文字を用いている。なお、表中の x はドントケアである。

表 4.1 属性の変数

p_y	属性名	属性値名
p_1	居住地 : Kinki	近畿圏内 : In 近畿圏外 : Out
p_2	年齢 : Age	平均年齢以下 : Low 平均年齢以上 : High
p_3	性別 : Sex	男性 : Male 女性 : Female

表 4.2 属性値ごとの人数

U_g	Kinki	Age	Sex	人数(人)
U_I	In	x	x	242
U_O	Out	x	x	136
U_L	x	Low	x	162
U_H	x	High	x	216
U_M	x	x	Male	181
U_F	x	x	Famale	197
U_{IL}	In	Low	x	101
U_{IH}	In	High	x	141
U_{OL}	Out	Low	x	61
U_{OH}	Out	High	x	75
U_{IM}	In	x	Male	121
U_{IF}	In	x	Famale	121
U_{OM}	Out	x	Male	60
U_{OF}	Out	x	Famale	76
U_{LM}	x	Low	Male	70
U_{LF}	x	Low	Famale	92
U_{HM}	x	High	Male	111
U_{HF}	x	High	Famale	105
U	x	x	x	378

4.2 提案手法の適用

検証用プログラムの実装には統計解析向けのプログラミング言語「R」を使用する。まず、ユーザ毎に割り振られたIDと属性を持つデータと、IDとそのユーザが訪れた詳細な位置情報を持つデータをマージ操作によって紐付けする。この操作により、ユーザのID、詳細な位置情報、属性が記載されたデータを作成する。図4.1にその結果例を示す。図4.1から、R00421700000000のIDを持つユーザの属性は近畿圏外から訪れた平均年齢以上の男性であることと、そのユーザは9点の位置情報が示す場所に訪れていることが分かる。

次に、比較する2群を選択し、各メッシュにおける差異を算出する。図4.2にその結果例を示す。なお、図4.2において比較を行っているユーザ群は $U_M = \{u_x | p_1 = Male\}$, $U_F = \{u_x | p_1 = Female\}$ であり、 $Diff(U_M, U_F)$ の絶対値が高い正と負における5つの値を示している。図4.2の-11.4479625319012というDiffは、lat（緯度）とlon（経度）で示したメッシュにおいて、全女性ユーザ中の女性が訪れた割合は全男性ユーザ中の男性が訪れた割合に比べて-11.4479625319012(%)高いことを表している。

ID,	lat,	lon,	Kinki,	Age,	Sex
R00421700000000,	34.685,	135.833,	Out,	Low,	Male
R00421700000000,	34.681,	135.819,	Out,	Low,	Male
R00421700000000,	34.680,	135.820,	Out,	Low,	Male
R00421700000000,	34.681,	135.820,	Out,	Low,	Male
R00421700000000,	34.682,	135.825,	Out,	Low,	Male
R00421700000000,	34.682,	135.820,	Out,	Low,	Male
R00421700000000,	34.683,	135.818,	Out,	Low,	Male
R00421700000000,	34.679,	135.819,	Out,	Low,	Male
R00421700000000,	34.682,	135.823,	Out,	Low,	Male

図 4.1 属性と位置情報の紐付け

lat,	lon,	Diff
34.684,	135.828,	-11.44796253190120
34.680,	135.829,	-10.50845556272260
34.677,	135.831,	-10.16630675603670
34.678,	135.829,	- 9.28569425358275
34.678,	135.831,	- 8.18072187789214
~~~~~	~~~~~	~~~~~
34.680,	135.820,	2.80730291387385
34.683,	135.829,	2.86900187901394
34.685,	135.825,	3.31491712707182
34.679,	135.820,	3.40466107636649
34.681,	135.820,	4.73399332529377

図 4.2 2群の差異

## 5 結果と考察

### 5.1 可視化による分析

データの可視化を図るため、 $Diff(U_g, U_{g'})$ の絶対値が高い正負それぞれの上位30件の *mesh* を比較結果として地図上にプロットした。なお、本稿では主に奈良市街を観光する観光客について属性間の差異を検証するため、近隣にある駅などの今回の分析とは関係性が薄いと考えられるデータは削除した。また、地図上のプロットから有用性の高い情報を読み取ることが困難であるため、複数属性の差異 ( $U_{IL}$  と  $U_{IH}$  の差異など) について本章では取り扱わない。

近畿圏内のユーザ群  $U_I$  (●) と近畿圏外のユーザ群  $U_O$  (▲) の比較結果を図5.1に示す。なお、今後示す地図中における縮尺、範囲はすべて同一である。

図5.1より、 $U_I$  は  $U_O$  に比べて油坂 (歓楽街) 付近に多く滞留している。また、転害門や東大寺といった駅から離れた観光地においても滞留している。一方で  $U_O$  は全体的に  $U_I$  と比べて多くの *mesh* に滞留しており、特に春日大社付近や奈良公園などの駅から比較的近い観光地や駅周辺、ショッピング街といった場所に多く滞留している。

図5.2に  $U_I$  (左側) と  $U_O$  (右側) について  $Gps(U_g, lon, lat)$  が大きい上位30の *mesh* を示

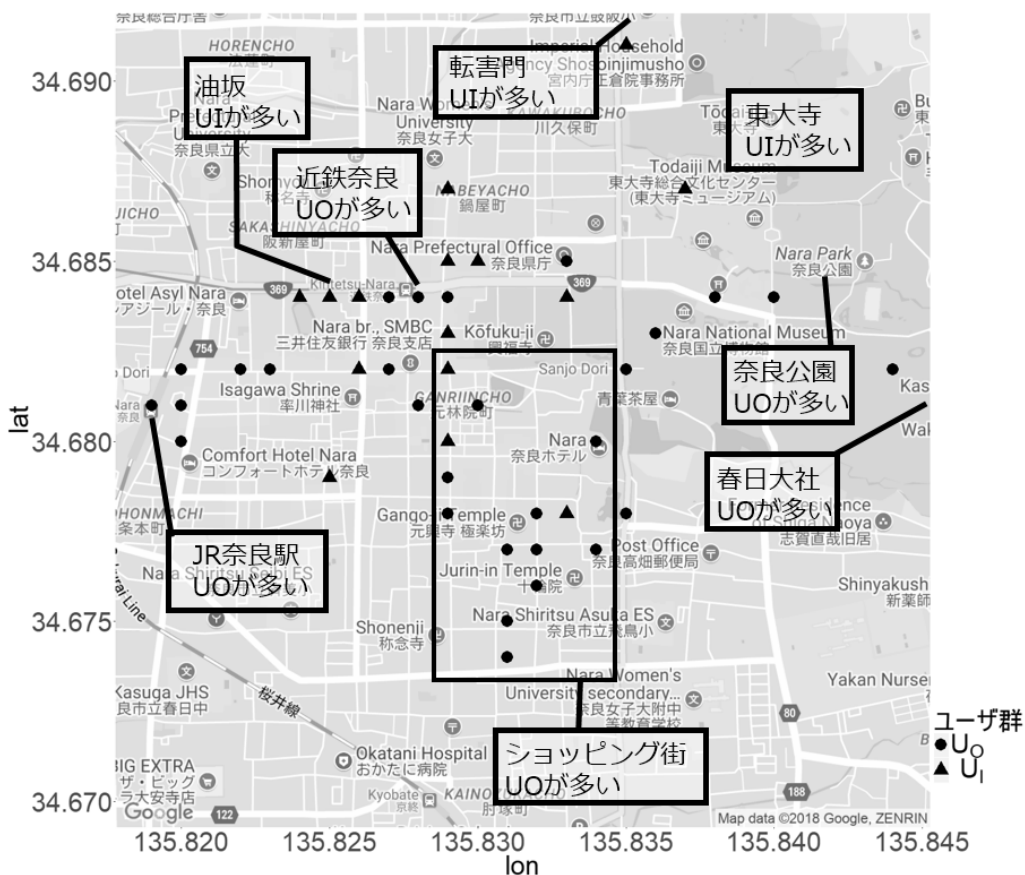


図 5.1  $U_I$  (▲) と  $U_O$  (●) の比較結果

す. 図5.2より,  $U_I$ と $U_O$ の $Gps(U_g, lon, lat)$ が大きい $mesh$ はどちらも駅周辺やショッピング街, 奈良公園付近などに多く, 差異はあまり見られない. これは, 多くのユーザが訪れる場所は類似しており, ユーザの居住地とは関係が薄いことを示している. 対して図5.1では $U_I$ と $U_O$ の差異が明確に現れている. よって, 提案手法は居住地の違いによってどの程度の差異が発生するか分析するのに有効であると考えられる. これらの結果から, 提案手法を用いることで $U_I, U_O$ の嗜好が顕在化したと考えられる. このことから,  $U_I$ と $U_O$ の差異を元にした推薦は, 単に $U_I$ や $U_O$ がよく訪れる場所をその属性のユーザに推薦する方法に比べて, ユーザにとって未知であると考えられるアイテムの推薦が行えると考えられる. このような推薦は, 一般に意外性のある推薦であるとされる.

全ユーザ中で平均年齢以下のユーザ群 $U_L(\blacktriangle)$ と全ユーザ中で平均年齢以上のユーザ群 $U_H(\bullet)$ の比較結果を図5.3に示す. 図5.3より,  $U_L$ は $U_H$ に比べて東大寺, 春日大社, 奈良公園など, 知名度の高い観光地やショッピング街に多く滞留している. 一方で,  $U_H$ は $U_L$ と比べてJR奈良駅周辺やショッピング街に多く滞留している. しかし, ショッピング街においては $U_L$ と $U_H$ の多く滞留している $mesh$ が混在しており, ショッピング街の中でも好みの店舗が違う, もしくは明確な差が無いと推測できる.

図5.4に $U_L$ (左側)と $U_H$ (右側)について $Gps(U_g, lon, lat)$ が大きい上位30の $mesh$ を示す. 図5.4より,  $U_L$ と $U_H$ の $Gps(U_g, lon, lat)$ が大きい $mesh$ に大きな差異は見られない. 対して図5.3では $U_L$ と $U_H$ の差異が明確に現れている. よって,  $U_L$ と $U_H$ の差異を元にした推薦は, 単に $U_L$ や $U_H$ がよく訪れる場所をその属性のユーザに推薦する方法に比べて, 意外性のある推薦であると考えられる.

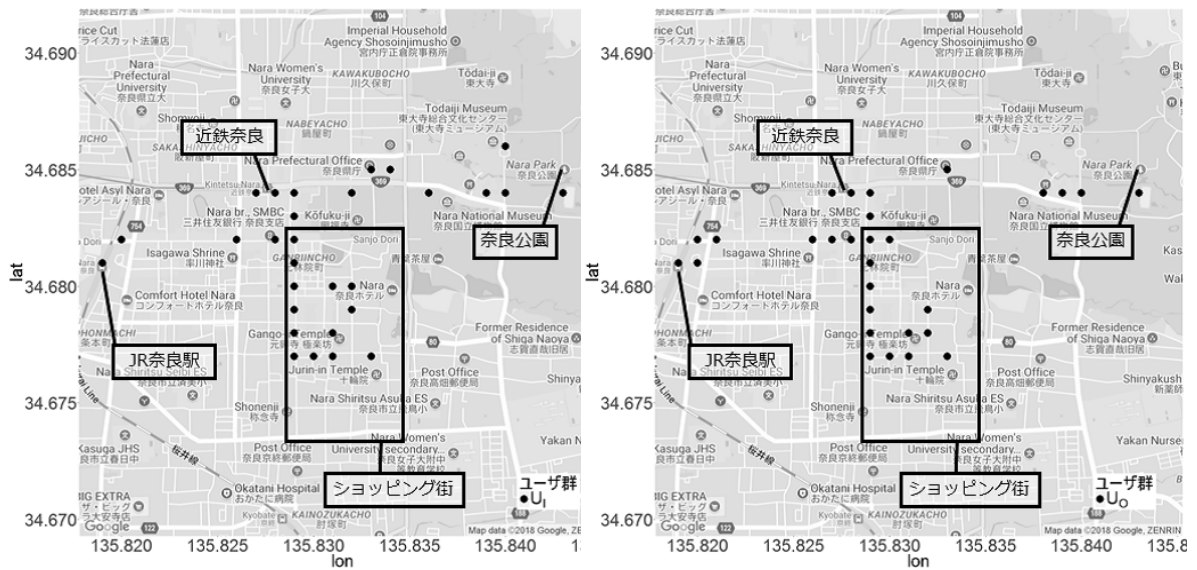


図 5.2  $U_I$ (左側)と $U_O$ (右側)について $Gps(U_g, lon, lat)$ が大きい $mesh$

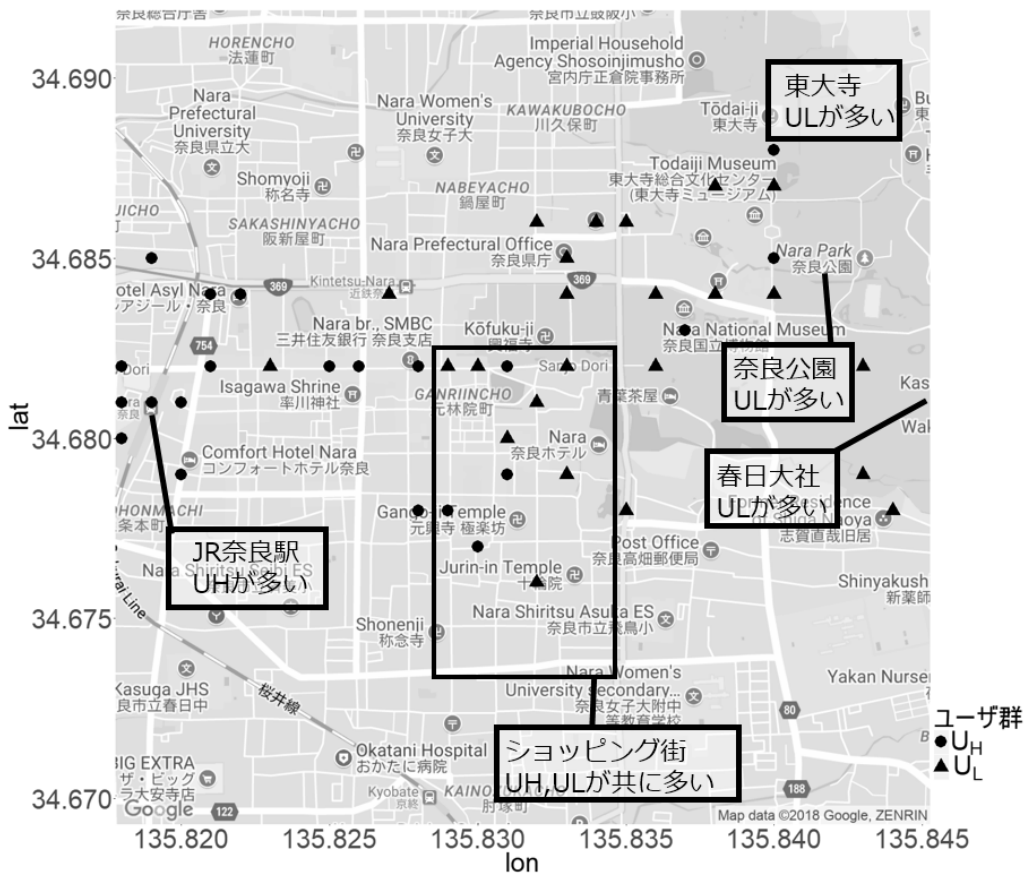


図 5.3  $U_L(\blacktriangle)$  と  $U_H(\bullet)$  の差異

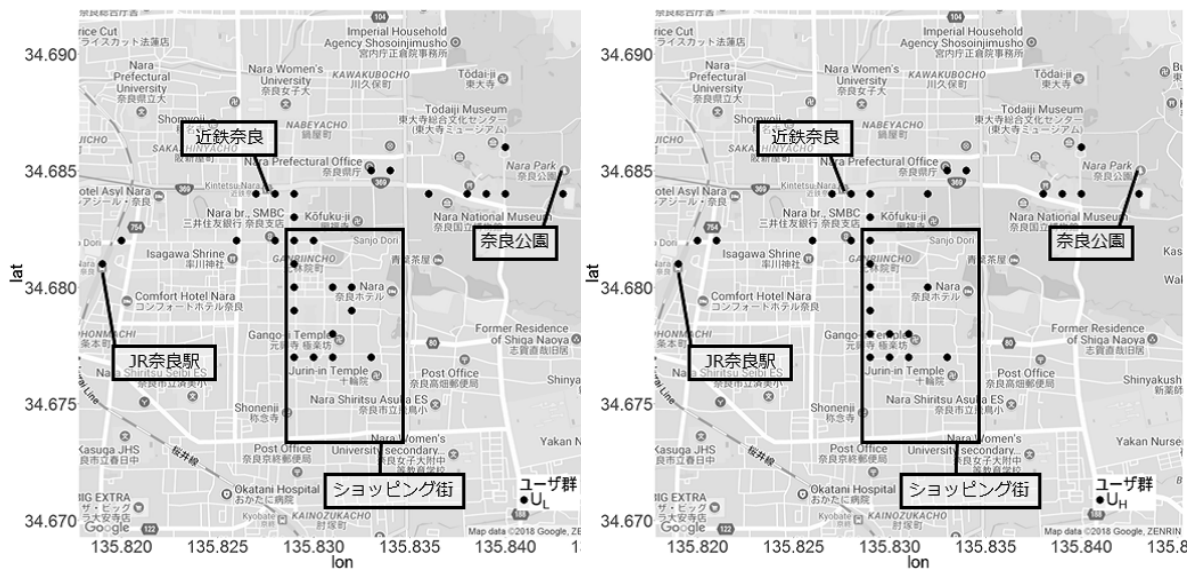


図 5.4  $U_L$ (左側) と  $U_H$ (右側) について  $Gps(U_g, lon, lat)$  が大きい mesh



男性のユーザ群 $U_M(\blacktriangle)$ と女性のユーザ群 $U_F(\bullet)$ の比較結果を図5.5に示す。図5.5より、 $U_M$ は $U_F$ に比べて駅周辺の大通りや転害門付近に多く滞留している。なお、図中の大通りに見られる点は、車移動によるものだと考えられる。 $U_F$ は $U_M$ と比べて東大寺や奈良公園などの有名な観光地やショッピング街において多く滞留している。

図5.6に $U_M$ (左側)と $U_F$ (右側)について $Gps(U_g, lon, lat)$ が大きい上位30のmeshを示す。図5.6より、 $U_M$ と $U_F$ の $Gps(U_g, lon, lat)$ が大きいmeshに大きな差異は見られない。対して図5.5では $U_M$ と $U_F$ の差異が明確に現れている。よって、 $U_M$ と $U_F$ の差異を元にした推薦は、単に $U_M$ や $U_F$ がよく訪れる場所をその属性のユーザに推薦する方法に比べて、意外性のある推薦であると考えられる。

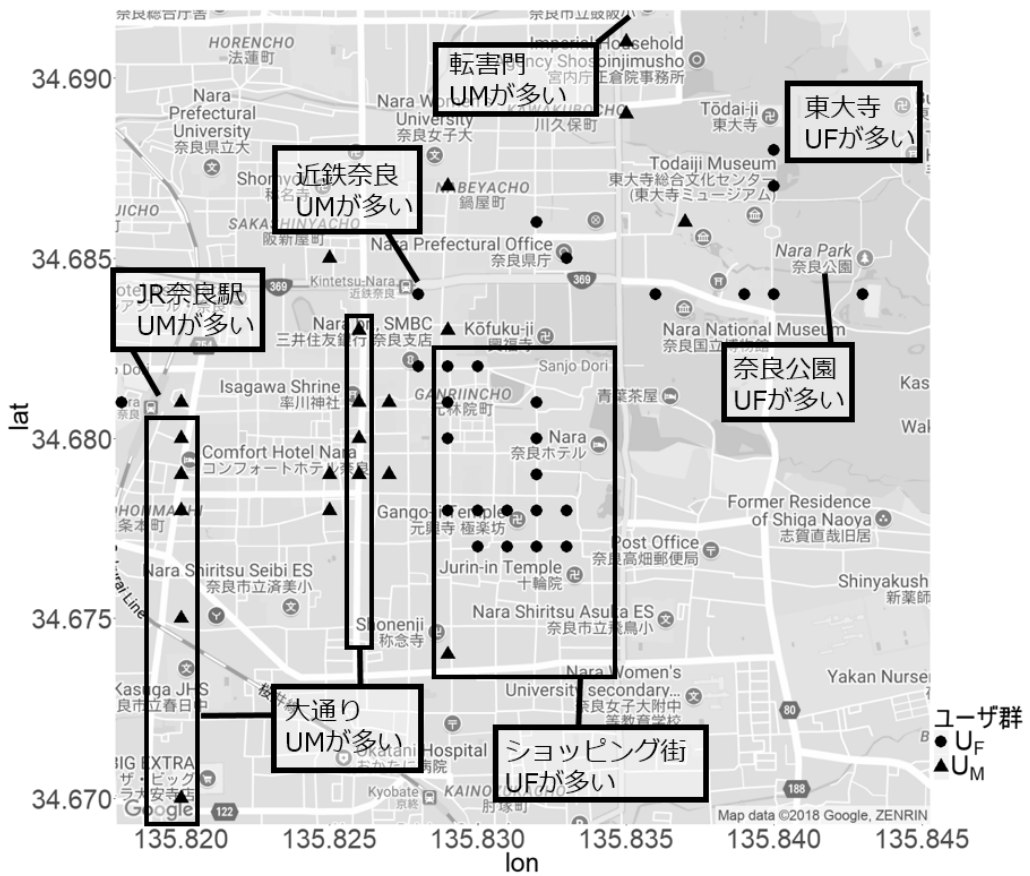


図 5.5  $U_M(\blacktriangle)$  と  $U_F(\bullet)$  の差異

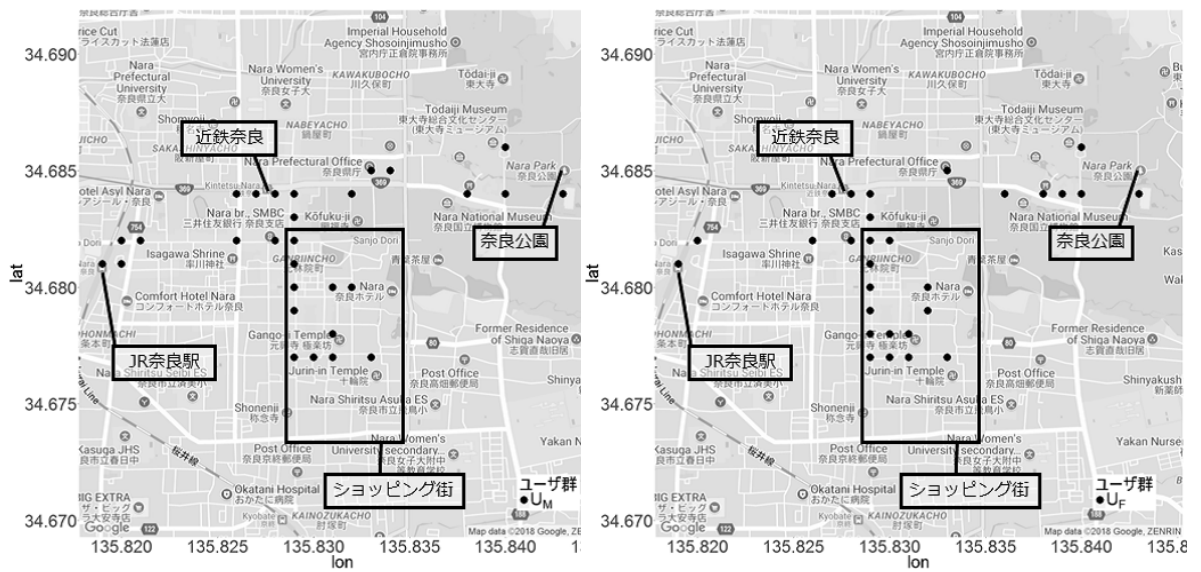


図 5.6  $U_M$ (左側) と  $U_F$ (右側) について  $Gps(U_g, lon, lat)$  が大きい mesh

## 5.2 定量的な分析

### 5.2.1 分析方法

本研究にて行うのは、属性間においてどの程度の差異が発生するかを明らかにすることである。しかし、 $Diff(U_g, U_{g'})$ を図5.1に示した方法で分析した場合、ある地点において $U_g$ もしくは $U_{g'}$ がもう一方の属性に比べて高い割合で訪れている以上のことは読み取れない。よって、比較した属性間に全体としてどの程度の差異があるのかを示す指標が必要である。また、ある1つの $Diff(U_g, U_{g'})$ を元にした指標を使用した場合、その値がある属性のユーザが偶然多く通り過ぎたなどの理由で発生した外れ値である可能性がある。よって本稿では $Diff(U_g, U_{g'})$ の絶対値が大きい正負の $a$ 番目から $b$ 番目までの絶対値を加算し、加算結果における正負の組数である $b-a+1$ で除算を行う。この操作により $Diff(U_g, U_{g'})$ の平均をとった値を指標 $Diff_{a-b}(U_g, U_{g'})$ とする。例として、図4.2の $Diff(U_M, U_F)$ について1番目から2番目までの値から指標を算出した結果を示す。なお、表示可能な文字数の関係上、式中に示す値は小数点以下5桁までとする。

$$\begin{aligned} Diff_{1-2}(U_M, U_F) &= \frac{(|-11.44796| + |4.73399|) + (|-10.50846| + |3.40466|)}{2-1+1} \\ &= 15.04758 \end{aligned} \tag{5.1}$$

この操作を行うことにより、属性間の差異を1つの値で表現する事が可能となり、またその値は複数の $Diff(U_g, U_{g'})$ を元に生成しているため外れ値の影響を受けにくい。そのため、比較した属性間に全体としてどの程度の差異が発生したのかを簡易に検証できる。

以上の分析を行う際に、比較する属性数( $Diff(U_M, U_F)$ なら1つ、 $Diff(U_{ML}, U_{FH})$ なら2つなど)が多い場合により細分化された嗜好が顕在化し、よく訪れる場所が大きく異なるために $Diff_{1-30}(U_g, U_{g'})$ が大きくなると考えられる。そのため、比較する属性数が違う $Diff_{1-30}(U_g, U_{g'})$ は同時に扱うべきでない可能性がある。よって、比較する属性数の違いによって $Diff_{1-30}(U_g, U_{g'})$ の値に有意差があるかを確認する。図5.7に比較する属性数ごとの $Diff_{1-30}(U_g, U_{g'})$ について実施したWelchのt検定の結果を示す。なお、検定には $Diff_{1-30}(U_g, U_{g'})$ の算出に用いた $Diff_1(U_g, U_{g'})$ から $Diff_{30}(U_g, U_{g'})$ の全ての値を使用した。これは $Diff_{1-30}(U_g, U_{g'})$ を用いて検定を行った場合、分散が極端に小さくなってしまい、正しい結果が得られないためである。

図5.7は比較する属性数が多ければ $Diff_{1-30}(U_g, U_{g'})$ が大きい傾向があることを示している。検定の結果、比較する属性数が1種類と2種類である場合には有意差( $p < 0.01$ )が見られた。この結果より、属性数が違う $Diff_{1-30}(U_g, U_{g'})$ の値は、結果を考察する上で同時に扱うべきではない事が分かる。よって、分析結果の考察は比較する属性数毎に行う。

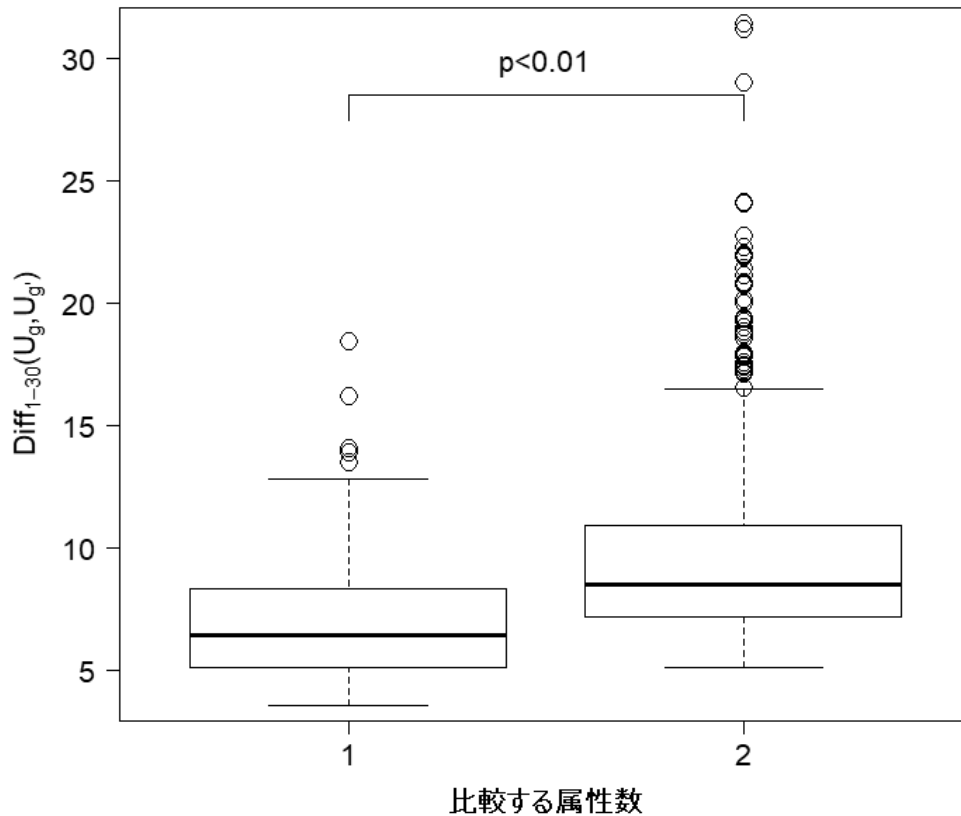


図 5.7 比較する属性数ごとの  $Diff_{1-30}(U_g, U_{g'})$

### 5.2.2 結果

表 5.1 に比較する属性数が 1 つである場合の  $Diff_{1-30}(U_g, U_{g'})$  を示す。

表 5.1 属性数が 1 種類の  $Diff_{1-30}(U_g, U_{g'})$

比較した属性	属性間の差異
$Diff_{1-30}(U_M, U_F)$	8.207
$Diff_{1-30}(U_I, U_O)$	7.563
$Diff_{1-30}(U_L, U_H)$	5.761

表 5.1 は性別による属性間の差異が特に大きく、次いで居住地、年齢の順に差異が小さいことを示す。多重比較 (Ryan 法) の結果、 $Diff_{1-30}(U_L, U_H)$  と  $Diff_{1-30}(U_M, U_F)$  の間には  $p < 0.01$ 、 $Diff_{1-30}(U_L, U_H)$  と  $Diff_{1-30}(U_I, U_O)$  の間には  $p < 0.05$  の有意差が見られた。なお、多重比較に用いるデータとしては 5.2.1 節と同じく  $Diff_1(U_g, U_{g'})$  から  $Diff_{30}(U_g, U_{g'})$  の全ての値を使用した。多重比較の結果から、属性の中でも差異が比較的大きい属性と小さい属性があることが分かる。よって、提案手法に基づいた推薦を実際に運用する際には、推薦に用いる属性の選択が必要とされる。具体例として、図 5.5 より  $U_M$  にショッピング街の店舗を推薦した場合には、 $Diff_{1-30}(U_g, U_{g'})$  が比較的大きいためユーザーにとって未知である可能性が高いと考えられる。対して、図 5.3 より  $U_H$  に春日大社付近の観光地を推薦した場合に

は,  $Diff_{1-30}(U_g, U_{g'})$  が比較的小さいためにユーザにとって未知である可能性が低いと考えられる. よって, ユーザにとって未知のスポットを推薦したい場合には  $U_M$  と  $U_F$  の差異を用いた推薦を選択すれば良い.

表5.2に比較する属性数が2つである場合の  $Diff_{1-30}(U_g, U_{g'})$  を示す.

1章において, 推薦対象の属性に対して遠い属性がよく訪れる観光地の推薦は, ユーザの嗜好が大きく異なるために未知の観光地であってもユーザの興味が薄いと想定している. しかし, これは実際の推薦を行わなければ検証できず, 実際はユーザの嗜好が大きく異なれば未知の観光地を多く推薦できるために, よりパーソナライズされた推薦を行える可能性もある. よって, ここでは異なる  $p_y$  が多ければ  $Diff_{1-30}(U_g, U_{g'})$  が大きくなるのかを検証する. 表5.2より, 図5.8に異なる  $p_y$  が1種類である場合と2種類である場合について実施したWelchのt検定の結果を示す. なお, 検定に用いるデータとしては5.2.1節と同じく  $Diff_1(U_g, U_{g'})$  から  $Diff_{30}(U_g, U_{g'})$  の全ての値を使用した.

図5.8は異なる  $p_y$  が2種類である場合, 1種類である場合に比べて  $Diff_{1-30}(U_g, U_{g'})$  が大きいことを示している. 検定の結果から, 2群には有意差 ( $p < 0.05$ ) が見られた. よって, 比較する属性数が2種類である場合, 異なる  $p_y$  が多いほど  $Diff_{1-30}(U_g, U_{g'})$  が大きいことが分かった. このことから, 異なる  $p_y$  が多いユーザ群は嗜好が大きい

表5.2 属性数が2種類の  $Diff_{1-30}(U_g, U_{g'})$

比較した属性	属性間の差異
$Diff_{1-30}(U_{OM}, U_{OF})$	11.728
$Diff_{1-30}(U_{IF}, U_{OM})$	11.491
$Diff_{1-30}(U_{IM}, U_{OF})$	10.968
$Diff_{1-30}(U_{IM}, U_{OM})$	10.866
$Diff_{1-30}(U_{LM}, U_{HF})$	10.460
$Diff_{1-30}(U_{IH}, U_{OH})$	10.391
$Diff_{1-30}(U_{LF}, U_{HM})$	9.907
$Diff_{1-30}(U_{LM}, U_{LF})$	9.863
$Diff_{1-30}(U_{HM}, U_{HF})$	9.674
$Diff_{1-30}(U_{IM}, U_{IF})$	9.477
$Diff_{1-30}(U_{IH}, U_{OL})$	9.412
$Diff_{1-30}(U_{IL}, U_{OL})$	9.350
$Diff_{1-30}(U_{IL}, U_{OH})$	9.276
$Diff_{1-30}(U_{OL}, U_{OH})$	9.270
$Diff_{1-30}(U_{IF}, U_{OF})$	9.137
$Diff_{1-30}(U_{LM}, U_{HM})$	8.046
$Diff_{1-30}(U_{LF}, U_{HF})$	8.009
$Diff_{1-30}(U_{IL}, U_{IH})$	7.587

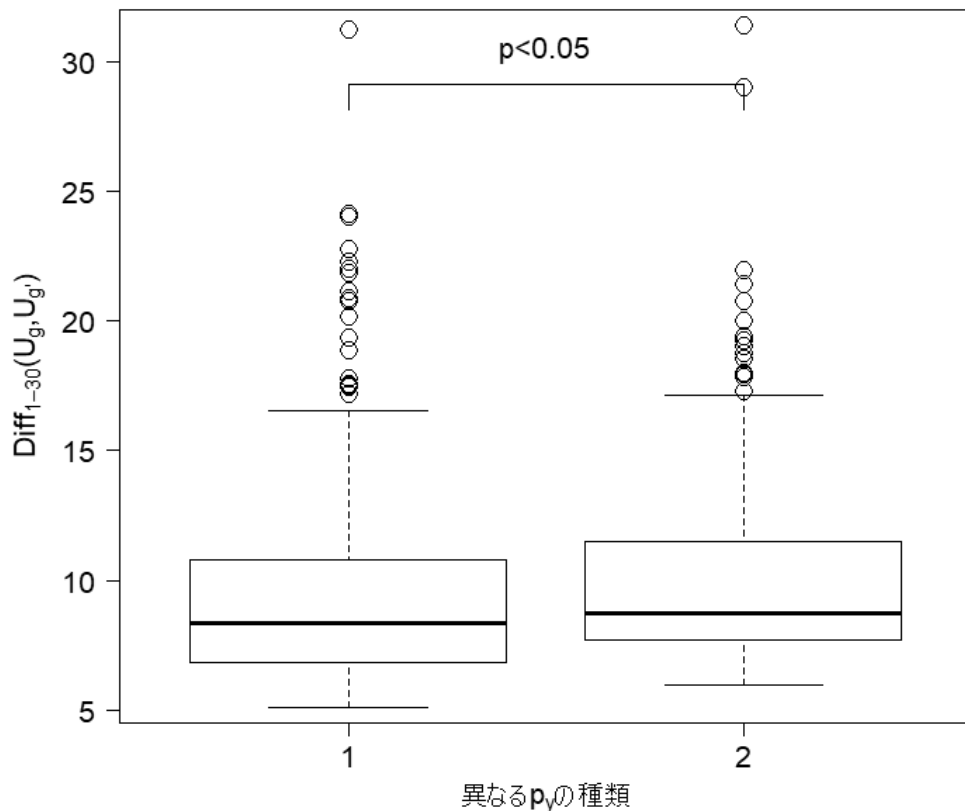


図 5.8 異なる  $p_y$  の種類ごとの  $Diff_{1-30}(U_g, U_{g'})$

く異なると考えられる．よって，ユーザの入力する属性を増やせば細分化した嗜好が顕在化し，よりパーソナライズされた推薦を行える可能性がある．

### 5.3 まとめ

5.1節より，提案手法を元にした推薦は，単にある属性のユーザが頻繁に訪れる場所をその属性のユーザに対して推薦することに比べて意外性のある推薦が可能であることが分かった．5.2節より，ユーザの入力する属性を増やせば，細分化した嗜好が顕在化し，よりパーソナライズされた推薦を行える可能性があることが分かった．以上の結果から，提案手法により算出した属性間の差異は，特に意外性の観点から1度限りの使用を想定したご当地観光アプリにおける推薦に有効であると考えられる．

意外性のある推薦は推薦した観光地がユーザにとって未知である可能性が高いというメリットがある．意外性があるとされる推薦手法である協調フィルタリングは，cold-start問題からご当地観光アプリでの適用が難しい．本研究の結果に基づく推薦では推薦対象ユーザの入力が自信の属性のみであるため，新規ユーザに対するcold-start問題が発生しないという優位性がある．

## 6 おわりに

本稿ではご当地観光アプリ「ならたん」において収集データに基づいて、属性の異なるユーザ間で発生する差異を検証した。提案手法は属性が異なる2群のユーザに対して、緯度経度を小数点以下4桁で四捨五入した値を軸とするメッシュを訪れたユーザの割合を求め、差を取ることで属性間の差異を算出する。実験結果の分析から、提案手法により算出した属性間の差異は、特に意外性の観点から1度限りの使用を想定したご当地観光アプリにおける推薦に有効である可能性を示した。意外性のある推薦は推薦した観光地がユーザにとって未知である可能性が高いというメリットがある。

今後の課題として、本研究で算出した差異の実際の推薦における活用法や、実際の推薦における精度評価の検討が挙げられる。また、ほとんど差異が発生しない等の推薦には不向きな属性が、手動の分析以外では特定できない点が挙げられる。ユーザの嗜好は時間経過と共に変化する可能性があるため、現状ではシステムの始動後に人が定期的に手を加えなければならない。解決策としてはこの分析の自動化が挙げられる。自動化の手法としては5.2説で算出した  $Diff_{1-30}(U_g, U_{g'})$  を用いる方法が考えられる。自動化に成功すれば、サービス提供者の労力が減るというメリットがある。本研究では属性間の差異のみを算出したが、属性間の類似度を算出し、両者の関係性を検証することで、新たな観点からより有効な推薦手法を提案できる可能性もある。

## 謝辞

本論文の執筆および研究をすすめるにあたって，多くの方々に協力していただきました．この場を借りてお礼を申し上げます．指導教員である上野秀剛准教授には研究へのアドバイスや論文の添削など，多忙の中様々なお指導を頂きました．心より感謝を申し上げます．内田眞司准教授や，査読教員である松尾賢一教授にはご意見・ご指摘を頂きました．心より感謝を申し上げます．



## 参考文献

- [1] 倉田陽平, 青木美岬, 相尚寿, “日本国内のご当地観光アプリの概要把握”, 観光情報学会第12回全国大会論文集 (2015).
- [2] 松原仁, 山本雅人, 川村秀憲, 鈴木恵二, “観光情報学—情報処理技術の観光への適用”, システム／制御／情報, Vol.57, No.8, pp.316-322 (2013) .
- [3] GISとは・・・ | 国土地理院, <http://www.gsi.go.jp/GIS/whatisgis.html>, 2017/12/13 参照 .
- [4] 佐藤貴大, 円山琢也, “カーネル密度推定法を応用したスマホ型回遊調査データの時空間分析”, 都市計画論文集, Vol.51, No.2, pp.192-199 (2016) .
- [5] 原辰徳, 矢部直人, 青山和浩, 倉田陽平, 村山慶太, 大泉和也, 嶋田敏, “サービス工学は観光立国に貢献できるか?-GPSロガーを用いた訪日旅行者の行動調査とその活かし方-”, デジタルプラクティス, Vol.3, No.4, pp.262-271 (2012) .
- [6] 篠田裕之, 竹内亨, 寺西裕一, 春本要, 下條真司, “行動履歴に基づく協調フィルタリングによる行動ナビゲーション手法”, 情報処理学会研究報告マルチメディア通信と分散処理, Vol.2007, No.91, pp.87-92 (2007) .
- [7] 神島敏弘, “推薦システムのアルゴリズム(1)”, 人工知能学会誌, Vol.22, No.6, pp.826-837 (2007) .
- [8] 神島敏弘, “推薦システムのアルゴリズム(2)”, 人工知能学会誌, Vol.23, No.1, pp.89-103 (2008) .
- [9] 吉井和佳, 後藤真孝, “音楽推薦システム”, 情報処理, Vol.50, No.8, pp.751-755 (2009) .
- [10] 小野智弘, 黒川茂莉, 本村陽一, 麻生英樹, “ユーザ嗜好の個人差と状況差を考慮した映画推薦システムの実現と評価”, 情報処理学会論文誌, Vol.49, No.1, pp.130-140 (2008) .
- [11] 本村佳菜子, 波多野賢治, “「人の属性」ごとに異なる嗜好傾向を考慮した推薦システムの提案”, DEIM Forum 2010 (2010) .