



卒業研究報告書

平成29年度

研究題目

系列パターンマイニングによる観光アプリ
ユーザのスタンプラリー取得情報の分析

指導教員 上野秀剛 准教授

氏名 高橋主

平成30年1月11日 提出

奈良工業高等専門学校 情報工学科

系列パターンマイニングによる観光アプリ

ユーザのスタンプラリー取得情報の分析

上野研究室 高橋 主

近年の観光客のニーズは多様化・細分化しており、既存の観光スポットのみに限らず新規の観光スポットを開拓・情報発信することが必要となってきた。同時にIoTの普及が急速に進んでおり、観光地の情報を発信する様々な情報メディアが公開されている。しかし、情報媒体から得られる観光地の情報量は膨大であるため、その中から観光客が自分のニーズに合った観光地を素早く探し出すことは困難である。そのような背景から観光地の推薦システムの研究がさかんに行われている。本研究では、観光地の推薦を想定し観光行動の分析、推薦のための情報の抽出を行う。観光行動分析に広く用いられるGPS情報は経緯度と時間のみの情報であるため行動の目的を計るには別のプロセスが必要になる。そこで本研究では目的を類推して行動に区別をつける必要がないスタンプラリーの取得履歴を分析の対象にする。スタンプラリーは特典が存在することが多いため、ユーザが実際に推薦場所を訪れる期待は大きく、推薦の効果も高いと考えられる。また、様々な分野の推薦システムがユーザの属性・嗜好を考慮するように、スタンプラリーの取得履歴の傾向が参加者の属性によって異なる可能性を踏まえ、属性ごとに取得履歴を分類し分析を行う。分析手法には、複数のアイテム集合に頻出である順序を考慮した系列パターンと呼ばれるアイテムの組み合わせを列挙する系列パターンマイニングを用いる。これにより頻出なスタンプの組み合わせとその取得順がわかり、スタンプラリーの次点の推薦が可能である。実験では、頻出な取得順序に属性の有無による差異があるかを示すため、属性による分類を行わない分析も行い両者の結果を比較する。属性による分類を行った場合の結果に現れるパターンが、分類のない場合の結果にも出現する割合を求めることで検証する。実験の結果、分析で得られた各属性のパターンの大半が属性の分類を行わなかった場合の分析結果には現れなかった。この結果により、スタンプ取得履歴の系列パターンマイニングによる分析の結果は属性の有無によって差異が存在することを示した。属性の有無による差異を示したことにより、スタンプラリーの取得履歴の分析において属性による分類は重要であると考えられる。

目次

1	はじめに	1
2	関連研究	3
2.1	推薦システム	3
2.2	系列パターンマイニング	3
3	観光地推薦における順序性と属性	4
3.1	観光アプリ「ならたん」	4
3.2	順序性と属性の課題	5
3.3	アプローチ	6
4	実験	7
4.1	属性の分類	7
4.2	系列パターンマイニング	8
4.3	結果の比較	8
4.4	分析データ	9
4.5	予備分析	10
5	結果と考察	13
5.1	属性による分類の有無の比較	13
5.2	ステータスの違いによる結果の差異	15
5.3	2種類の出現割合の差	16
6	おわりに	17
	謝辞	18
	参考文献	19

1 はじめに

近年の観光客のニーズは多様化・細分化しており、既存の観光スポットのみに限らず新規の観光スポットを開拓・情報発信することが必要となってきた。総務省は観光客がスマートフォンやタブレットを用いて、地域のリアルタイムな情報を基に自らのニーズに沿った観光地を発見できるIoTシステムの導入を目指している[1]。これが実現できれば新たな観光需要の創出や観光客・観光消費の増加を狙えるとしている。

IoTを用いて観光客に情報を発信する手段には観光名所や飲食店などの情報を提供する観光アプリケーションがあり、様々な自治体によりご当地観光アプリケーションが開発されている[2][3]。また、新規観光スポットの情報発信の手段として有効であると考えられるものの1つにスタンプラリーがある。観光客の回遊性を向上させることから新規の観光スポットにおいてもスタンプラリーによる集客が見込まれる。さらにIoTの普及に伴ってスタンプラリーも電子化が進んでおり、手持ちのスマートフォンで誰でも気軽に参加できるというメリットがある。しかし、観光アプリケーションをはじめとする情報媒体から得られる観光地の情報は膨大であるため、その中から観光客が自分のニーズに合った観光地を素早く探し出すことは困難である。そのような背景から観光地の推薦システムの研究がさかんに行われている[4][5]。

本研究では、観光地の推薦を目的とした観光行動の分析、推薦のための情報の抽出を行う。観光地での行動の分析にはGPS情報が使われることが多いが、GPS情報は経緯度と時間のみの情報であるから行動の目的を計ることが容易でない。図1のように、例えばある一点に10分滞在している記録があっても、それが観光するために留まっているのか、ただ立ち止まって休憩しているのかを知ることが困難である。目的の区別をつけなければ、観光を目的としているユーザに対して何も観光するものがない休憩所なども推薦の対象に含まれてしまう。そのため観光客の目的に沿わない推薦を行ってしまう可能性がある。そこで本研究では分析対象にスタンプラリーの取得履歴を使用する。スタンプラリーの取得履歴であれば、スタンプを取得するという目的が全履歴に共通して含まれるため、履歴から目的を類推して行動に区別をつける必要がない。また、スタンプラリーには一定数以上取得することで懸賞に応募できるなどの特典が存在することが多い。そのため、GPSなどによる推薦と比較するとユーザが実際に推薦場所を訪れる期待は大きく、スタンプラリーの履歴を用いた推薦の効果は高いと考える。

分析手法には系列パターンマイニングを用いる。系列パターンマイニングは複数のアイテム集合で頻出なアイテムの組み合わせを列挙する。ただし、この組み合わせは順序性を考慮する。例えば、アイテム集合“A, B, C”, “A, D, B”, “B, A, E”が存在するとき、頻出なアイテムは“A”と“B”であるが、これらの組

観光なのか休憩なのか 行動の目的の区別がつかない

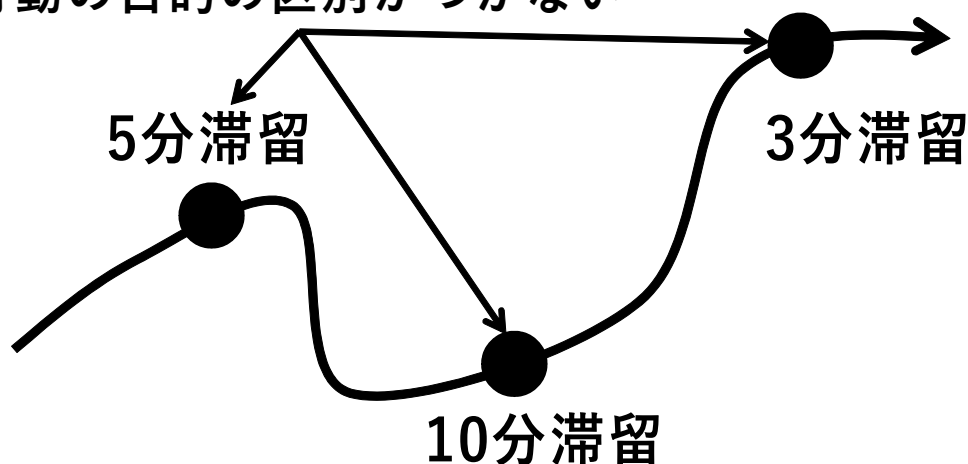


図1 GPSログの例

み合わせは“A, B”と“B, A”の2種類がある．順序性を考慮するというのはこの2つの組み合わせを別物として捉えるということである．この場合，組み合わせ“A, B”は全集合のうち2つの集合に現れているので頻出であり，“B, A”は1つの集合にしか現れていないので頻出ではない．系列パターンマイニングでスタンプ取得履歴を分析することで，頻出なスタンプの組み合わせとその取得順がわかり，スタンプラリーの次点の推薦が可能である．

また，推薦手法としてユーザの属性・嗜好を考慮した推薦を行うシステムの研究が行われている [6]．スタンプラリーの取得履歴の傾向が参加者の属性によって異なる可能性を考え，その属性ごとに取得履歴を分類し系列パターンマイニングによる分析を行う．ここでいう属性は性別，年齢層，居住地の3つのステータスからなる．それぞれのステータスは男女，老若，対象観光地付近に在住しているか否かの2種類の値のうち1つを持つ．これらのステータスの組み合わせ（例えば“観光地付近に住む若い男性”）を1つの属性とし，計8つの属性に参加者を分類する．

上記による推薦を想定し，各属性において頻出なスタンプの取得順序の抽出を行う．また，属性による分類を行わない場合の抽出も行い，両者の結果を比較することで頻出な取得順序が属性の有無によって差異があるかを示す．以下，2章では関連研究について述べ，3章で研究対象について説明する．実験の内容を4章にて述べ，5章でその結果と考察を示す．最後に6章で本研究のまとめと今後の課題を示す．

2 関連研究

本章では推薦システムおよび系列パターンマイニングに関する研究について述べる。

2.1 推薦システム

観光地を推薦するシステムの研究ではユーザの入力に対して、類似性の高い観光地を推薦する手法が使われている。Kanazawaらは、観光地の検索プロセスの効率を改善することを目的に、印象語などの分析に基づいて、行きたい場所のイメージから連想される行き先推薦システムを提案している[4]。嶋田らが提案したシステムでは、入力として具体的な1つの観光地を与えている[5]。これを観光者の嗜好として捉え、対象となる範囲の観光地との類似性を判断し、類似性の高い観光地を推薦する。類似度を測定する基準には集合知としてのWebを利用することで、複数の異なる情報源、異なる種類の情報を複合的に解釈し、類似度を算出している。これにより1つの情報源のみに依存しない推薦システムを構築している。これらの研究はユーザの入力から関連する観光地を推薦するシステムを提案しており、その点で本研究が想定するシステムと関連している。本研究で想定するシステムの入力はスタンプであり、推薦対象となる観光地の選出には観光地間の類似度ではなく、観光客の属性と過去の観光客の行動の統計を用いる。

2.2 系列パターンマイニング

系列パターンマイニングはWeb巡回ログなどの分析のほか、カルテの分析に用いられており医療の場面における医者意思決定支援システムの研究が行われている[8][9]。佐々木らの研究では系列パターンマイニングを行い、得られた結果を用いて医療行為の支援を行う方法に関して検討している[8]。患者毎のオーダログ中のシーケンスを解析することで医療行為の推薦や検証を行うことができるとしている。佐々木らの手法は医療行為の実施日の変動による影響を考慮するため、同じパターンとみなす医療行為の時間間隔に幅を持たせたマイニングを行っている。

Wrightらの研究では、系列パターンマイニングが、投薬間の時間的関係を特定し、患者に対して処方される可能性のある次の医薬品を正確に予測するのに有効であるか否かを検討している[9]。予測精度を測定する実験の結果、医薬品のおおまかな種類での予測では90.0%、個別の名称での予測では64.1%の精度を得ている。これらの研究は系列パターンマイニングを用いてデータログの分析を行っており、その結果を意思決定の支援に用いている点で本研究と類似している。本研究との相違点は、これらの研究が医療分野における医療行為の分析・予測を行っているのに対し、本研究では観光地の推薦を想定した観光行動の分析を行う。

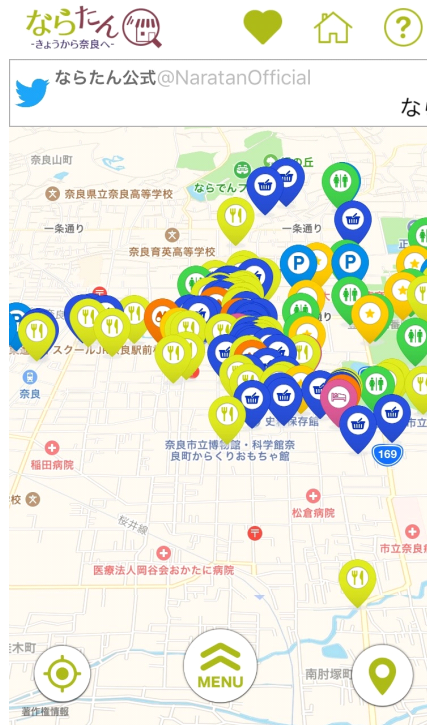


図2 「ならたん」のホーム画面

3 観光地推薦における順序性と属性

本研究での分析の対象は奈良県が開発・公開している観光アプリケーション「ならたん」[10]のユーザが取得するスタンプラリーの情報である。以下、「ならたん」のスタンプラリー機能を説明し、本研究の課題およびアプローチを述べる。

3.1 観光アプリ「ならたん」

奈良県ではIoTにより地域のデータを効率的に活用することで、奈良県経済の発展及び地域課題の解決への寄与を目的とした様々な取り組みが行われている。その一端でスマートフォン向け観光消費促進アプリケーションとして開発されたのが「ならたん」である。「ならたん」は奈良県各地の観光スポットや店舗などの情報が掲載されているほか、そのような場所を対象としたスタンプラリーの機能がある。「ならたん」のホーム画面（図2）は地図となっており観光スポットや店舗、スタンプラリーの位置などがプロットされている。

スタンプを取得するには、実際にスポットを訪れアプリ内の地図にプロットされているマーカーをタップする。取得判定にはGPS機能が用いられており、マーカーをタップした時点のユーザの位置情報が対象スポットの一定範囲内であればスタンプを取得できる。ユーザがスタンプを取得すると同時にアプリはどのユー

CC,	地点名1,	取得時間1,	地点名2,	取得時間2
R0042160000007500,	行基広場,	2017/01/28 14:42:26,	ならまちセンター,	2017/01/28 14:56:51
R0042170000000033,	夢CUBE,	2017/01/19 20:02:44,	三条通り,	2017/01/20 11:42:10
R0042170000000051,	行基広場,	2017/02/10 09:43:56		

図3 「ならたん」が記録するデータログの例

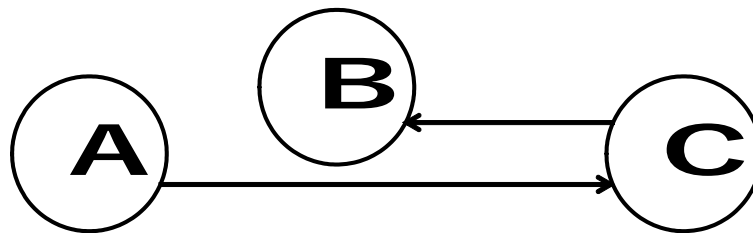


図4 非効率な回遊を与えてしまう推薦

ザがどのスタンプをいつ取得したか記録する。記録されたログはサーバ上に保存されており、全ユーザの履歴を取得することができる。スタンプラリーのログのほかにはアプリ起動中にGPS情報、アプリケーションの初回起動時にアンケートによってユーザの居住地・生年・性別・職業を記録している。図3にスタンプラリーの記録の一部を示す。スタンプラリーのログは、ユーザ固有の記号列である「CC」、1番目に取得したスタンプの地点名、取得した時刻、2番目に取得したスタンプの地点名、…と続く。

3.2 順序性と属性の課題

観光地の推薦を目的とする場合、観光地間の距離や地形などの地理的な要因が重要になる。そのため、観光地の推薦では順序性を考慮する必要がある。図4に順序性を考慮しない場合に起こりうる推薦の例を示す。観光地が“A”，“B”，“C”の順に地図上に並んで存在するとするとき、スタンプ“A”を取得したユーザには“B”，“C”の両方を推薦することが考えられる。しかし、ユーザの回遊における効率を考えると距離の近い“B”を推薦したのちに“C”を推薦すべきである。そのほか、距離は近くても川の対岸であり橋がない場合など観光地間に障害が存在することもある。この場合、距離は多少離れても橋付近にある地点を先に推薦した方が効率がよい。観光地回遊の効率を考えると推薦を行う地点の順序は非常に重要である。

また、観光行動には属性間の差異があると考えられる。観光地と一括りにいっても、歴史的な建造物や地域特有の店舗など様々なスポットが存在し、訪れる観光地の決定は個人の嗜好や属性が影響すると考えられる。そのため、より効果的な推薦

を行うにはユーザが興味を抱くような場所を推薦すべきであり、その指標に嗜好や属性が必要になる。

3.3 アプローチ

本研究では、3.2節のことを踏まえ、系列パターンマイニングによってスタンプラリーログを分析することで頻出なスタンプの取得順序をユーザの属性ごとに抽出する。系列パターンマイニングは1章で述べた通り、順序性を考慮した頻出パターンマイニングである。頻出パターンマイニングは複数のアイテム集合に共通して現れるアイテムを列挙するものであり、POSデータの分析によく使われる手法である。しかし、順序性は考慮されない。POSデータの分析を例にとると、商品“A”と商品“B”の組み合わせが頻出であることが分かれば、“A”の隣に“B”を陳列するなどして購買運動を促す戦略を立てることが可能となる。スタンプを1つのアイテムとみなせば1人のスタンプログは1つのアイテム集合として扱えるので、この手法は観光地の推薦の手法としても使用できる。

しかし頻出パターンマイニングを用いた場合、図4の問題が起こりうる。“A”を取得したユーザには“B”の次に“C”、“C”の次に“B”のように2通りの推薦が可能である。2通りの推薦のうちどちらが効率的に回遊できる順序であるかは別のプロセスで距離を測るなどしない限り判定できない。一方で系列パターンマイニングの結果は順序性を持つため、“B”の次に“C”と、“C”の次に“B”は別の組み合わせとして扱われる。このため、多数のユーザが非効率な順序でスタンプを取得していない限りは、より効率的なパターンが高頻出となり別プロセスを用いてどちらが効率的であるかの判定をする必要はない。以上のことから本研究では系列パターンマイニングを分析手法として用いる。

また、スタンプラリーの取得にも属性間の差異があると仮定すると、属性による分類を行った系列パターンマイニングの結果、その属性に特有または比較的高頻出なパターンが抽出される。しかし、属性による分類を行わなければ各属性に共通なパターンばかりが上位となり、各属性に特有なパターンは埋もれてしまう。系列パターンマイニングを用いた推薦では抽出結果を最も頻出なものから順にランク付けし、その上位を推薦に用いることがほとんどである。このことから、ある属性では高頻出で有益であるはずのパターンが推薦に用いられないことが危惧される。

属性ごとに頻出なパターンを抽出したのち、属性による分類を行わないで系列パターンマイニングを行い、両者の結果を比較する。これにより、実際に各属性で特有なパターンが属性による分類を行わない場合の結果において埋もれてしまっているかを検証する。

表1 属性の構成

属性	居住地	生年	性別
IYM	県内	1974年以降	男
OYM	県外		
IOM	県内	1973年以前	
OOM	県外		
IYF	県内	1974年以降	女
OYF	県外		
IOF	県内	1973年以前	
OOF	県外		
ALL	県内外	全年齢	男女

4 実験

異なる属性ごとに系列パターンマイニングを用いて頻出なスタンプの取得順序を抽出する。まず、分析に用いるデータであるスタンプラリーログをそのユーザの属性によって分類を行う。次に、ログを系列パターンマイニングを行うのに適した形に整理し、系列パターンマイニングを適用させて頻出なスタンプの取得順序を抽出する。続いて、ログに対して属性による分類を行わず、すべてのログに対して同様に系列パターンマイニングを行う。その後、両者の結果を比較し、属性による分類の有無によってどの程度差異があるか検証する。

4.1 属性の分類

属性の分類について説明する。1章でも述べたように、観光行動に影響を与えると考えられる居住地、生年、性別の3つの要素から8つの属性を定義する。各ステータスはそれぞれ2種類の変動しない状態を持つ。本来、生年は10代、20代、30代…、のように細かく区切ることが望まれるが、今回は細かく区切られるほどログの数は十分でなかった。そのため、ログの数が2等分される1974年以降と1973年以前で区切ることにした。各ステータスの2種類の状態とは居住地は奈良県内であるか奈良県外であるか、生年は1974年以降か1973年以前か、性別は男性か女性かである。

これら3つの要素を組み合わせることで定義した属性を表1に示す。各属性は大文字アルファベット3文字で表す。“ALL”を除いて、1文字目は“Inside”か“Outside”，2文字目は“Young”か“Old”，3文字目は“Male”か“Female”の頭文字であり、そのままログのステータスを示している。“ALL”は属性による分類を行わない場合を示している。以降、本稿では属性による分類を行わない場合も含め、それぞれの属性を表1に示した通りに呼称する。

tid,	time,	item
R0042160000007500,	1,	行基広場
R0042160000007500,	2,	ならまちセンター
R0042170000000033,	1,	夢CUBE
R0042170000000033,	2,	三条通り
R0042170000000051,	1,	行基広場

図5 整形後のログ

4.2 系列パターンマイニング

系列パターンマイニングで分析するにあたって NYSOL[11] のパッケージであるデータマイニングパッケージ「TAKE」¹を用いる。NYSOLとはビッグデータ解析に関する様々な研究成果に基づいて構築されたソフトウェアツールの総称である。使用されているアルゴリズムはLCMアルゴリズム[7]である。

このツールで分析するにはまずログをツールが扱える形に直す必要がある。図3のログを整形した形が図5である。「tid」はアイテム集合をあらわすIDであり、ここではユーザとアイテム集合がただ1つに決まる「CC」を用いている。「item」はアイテム集合の要素である取得されたスタンプであり、「time」は「item」がアイテム集合のうちの何番目か、つまり「item」のスタンプを何番目に取得したかを指す。

ログを整形した後、前述の通りに系列パターンマイニングを行う。系列パターンマイニングで見つかる系列パターンはサポート値と呼ばれる指標で評価される。サポート値とは各パターンの出現頻度のことであり、この値が大きいほど多くのアイテム集合に共通することを示す。サポート値はそのパターンを含むアイテム集合の数をアイテム集合の総数で割ることで求められる。即ちサポート値は全アイテム集合のうちパターンが含まれる集合の割合をあらわす。この値を基準に降順でパターンを列挙する。

しかし、系列パターンマイニングによって得られるパターンの数は分析するデータの数が多いほど膨大な数が見つかるため、列挙するパターンに制限を設ける必要がある。今回の分析では、サポート値の下限を0.05、サポート数（パターンが含まれるアイテム集合の数）の下限を3に設定し分析する。また列挙するパターンの最小サイズ（1つのパターンに含まれるアイテムの最小数）は3とする。

4.3 結果の比較

属性による分類を行った場合の結果と行わなかった場合の結果を比較する手順を以下に示す。手順4で一定順位以内のみで比較するのは、属性によって抽出され

¹宇野毅明氏を中心に開発されたデータマイニングソフトウェアを拡張し利用しやすくしたコマンド群

表2 各属性のログ数

属性	ログの数
IYM	26
IYF	53
IOM	38
IOF	45
OYM	38
OYF	51
OOM	47
OOF	38
ALL	336

るパターンの数が異なる可能性があり、単純な比較が難しくなるからである。

1. ALLを含む各属性内で、列挙したパターンをサポート値によってランク付ける。
2. ALL以外の各属性における一定順位以内のパターンがALLにおいても一定順位以内に出現するか判定する。
3. その出現する割合を計算し、各属性の上位にあるパターンのどの程度がALLでは上位に現れなくなってしまうかを示す。
4. 同様にALLの一定順位以内のパターンが各属性の一定順位以内に現れる割合を計算し、ALLの上位パターンが各属性のどの程度の上位パターンを含んでいるかを示す。

4.4 分析データ

分析に用いるデータは「ならたん」の試験運用期間（2017年1月1日～2月28日）中に取得されたスタンプラリー取得ログである。この期間にスタンプを1つ以上取得したユーザの総数は397人である。ここから属性を構成する居住地、生年、性別の3つステータスのうち1つ以上欠如しているユーザや生年と職業の辻褄が合わないなどのユーザのログを除外し、計336人のログを分析に用いる。表2に各属性のログ数を示す。

同期間中にスタンプスポットとして登録されていた20か所を表3に示す。これらのスポットは、観光地のほか店舗や博物館等の施設となっている。以降、「行基広場」および「行基広場」で取得されたスタンプを“A”というように表3に示した記号でスポットとスタンプを指す。なお、記号の振り方は全データにおいて最も取得されている地点から順にA, B, C…としている。

表3 スタンプのスポット

記号	スポット名
A	行基広場
B	夢CUBE
C	三条通り
D	にぎわいの家
E	総合観光案内所
F	ならまちセンター
G	夢風ひろば
H	おもちゃ館
I	東大寺ミュージアム
J	油坂
K	ならまち工房
L	I・RA・KA
M	きたまち案内所
N	鹿の舟
O	入江泰吉旧居
P	春日大社二ノ鳥居
Q	転害門観光案内所
R	若草山
S	志賀直哉旧居
T	奈良市写真美術館

4.5 予備分析

分析の前に、属性間にスタンプラリーの取得の傾向に差異が実際に存在するかを検証する。はじめに、各属性のスタンプの取得数と各ステータスにおける平均をそれぞれ表4, 5に示す。この表から1人当たりのスタンプの取得数は、県外在住のユーザ（平均6.16個）よりも県内に在住しているユーザ（平均6.35個）の方が多い。生年で比べるとより高齢なユーザ（平均6.35個）の方が若齢なユーザ（6.16個）より多く取得している。特に性別を見ると男女での差は1人当たり1個近くの差（男性平均6.66個，女性平均5.84個）があり，男性の方がスタンプを多く集める傾向があるといえる。

次に、各属性のユーザのうちそのスタンプを取得したユーザの割合を各スタンプスポットについて計算し、そのグラフを3つのステータスのうち1つのみが異なる属性同士を比較する形で示す。図6, 図7, 図8がそれぞれ居住地のみが異なる属性同士の比較，生年のみが異なる属性同士の比較，性別のみが異なる属性同士の比較である。グラフの横軸は各スタンプスポットを指し，縦軸がその地点での取得割合をあらわす。例えば図6(a)では，若齢であることと男性であることは同じで居住地のみが違うIYMとOYMの各地における取得割合を比較している。図

表4 各属性のスタンプ取得数

属性	人数	総取得数	1人当たり
IYM	26	192	7.38
IYF	53	279	5.26
IOM	38	257	6.76
IOF	45	269	5.98
OYM	38	198	5.21
OYF	51	345	6.76
OOM	47	343	7.30
OOF	38	204	5.37
ALL	336	2087	6.21

表5 各ステータスにおける平均取得数

属性	人数	総取得数	1人当たり
県内	40.5	249.3	6.35
県外	43.5	272.5	6.16
若齢	42.0	253.5	6.16
高齢	42.0	268.3	6.35
男性	37.3	247.5	6.66
女性	46.8	274.3	5.84

中のアスタリスクは差が10%以上開いている箇所を示している。

居住地のみ異なる属性同士の比較において、若齢の男性（図6(a)）ではユーザ全体の取得数が低い9か所でいずれも県内のほうが高い。高齢の男性（図6(b)）の比較においては、全体の取得数が高い7か所で県外、スポット“M”では県内のほうが10%以上高くなっている。若齢の女性（図6(c)）では8か所で県外が、1か所では県内が10%以上高くなっており、特に“K”では県外の方が30%以上高い。高齢の女性（図6(d)）でも10か所において10%以上の差が見られ、全体の取得数が高い2か所と低い4か所では県内、中腹の4つのスポットでは県外のほうが高い。生年が異なる属性同士の比較では、県内男性(図7(a))の5か所、県外女性（図7(d)）の10か所のいずれも若齢のユーザのほうが10%以上高い。反対に県外男性(図7(b))の10か所、県内女性（図7(c)）の3か所では、いずれも高齢のユーザのほうが10%以上高くなっている。性別の比較では、県内若齢者(図8(a))の11か所、県外高齢者（図8(d)）の9か所のいずれも男性ユーザのほうが10%以上高い。県外若齢者(図8(b))では6か所で女性、1か所では男性のほうが10%以上高い。県内高齢者(図8(c))においては、全体の取得数が最も多い“A”で女性のほうが、3か所では男性のほうが10%以上高い。

最も取得されることが多いスポットはどの属性でも“A”であるが、2番目以降

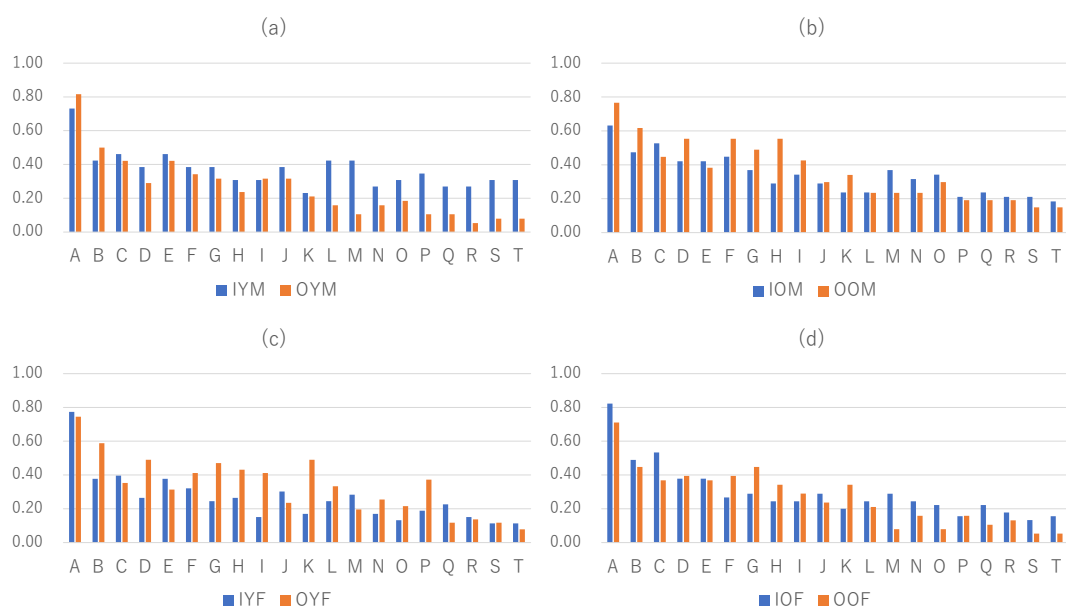


図6 居住地のみ異なる属性同士の比較

は各属性によって異なっている。多くのスポットにおいてステータス1つの違いで10%以上の差が開いており、中には30%以上となる場合もあった。また、図6~図8(a)に注目すると、全体の取得数が低いマイナーなスポットでIYM(県内の若い男性)のほうが高いことが共通している。これは県内の若い男性のマイナーなスポットに対する興味を示唆している。そのほか、3つの図中すべてでOOM(県外の高齢な男性)は全体の取得数が多いスポットで高くなっており、メジャーなスポットを集中してまわっていることが考えられる。反対に、IYF(県内の若い女性)は3つの比較の中で10%以上高くなる箇所は1つのみであり、他属性に比べてスタンプラリーへの参加具合が低いといえる。以上のことから、スタンプラリーの取得の傾向において属性による差異は存在することがいえる。

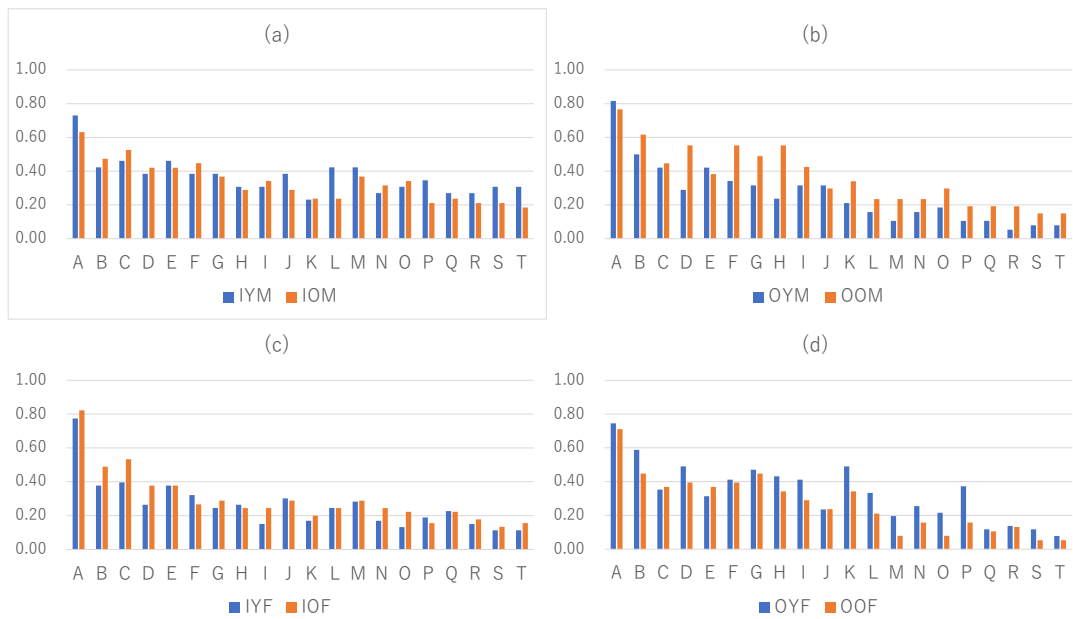


図7 生年のみ異なる属性同士の比較

5 結果と考察

5.1 属性による分類の有無の比較

系列パターンマイニングによるスタンプの取得パターンの抽出を行った結果、各属性で列挙されたパターンの数は表6の2列目のようになった。その後、各属性においてサポート値の高い順にランク付けを行った。パターンが最も多く列挙されたのがOOMで4067パターン、最も少なかったのがOOFであり114パターンとなった。それに対してALLのパターン数は394であり、予想した通り各属性のパターン数は大幅に違った。パターン数の違いによる出現割合の差異を軽減するため、上

表6 各属性において抽出されたパターンの数

属性	抽出パターン数	100位以内のパターン数
IYM	1220	134
IYF	1030	177
IOM	2456	376
IOF	580	580
OYM	487	182
OYF	1476	163
OOM	4067	110
OOF	114	114
ALL	394	125

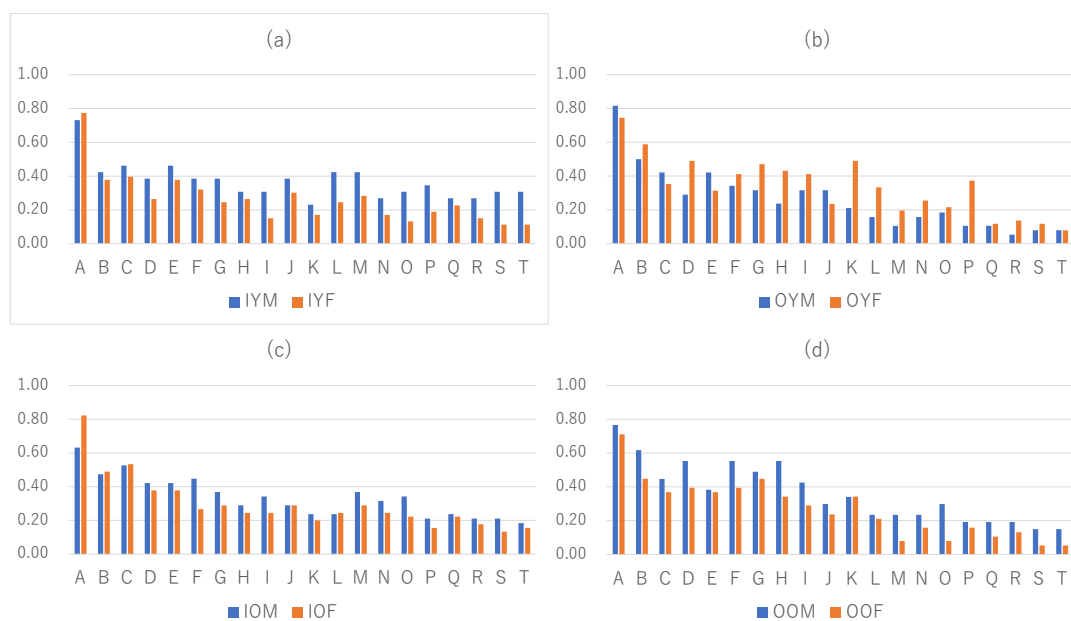


図8 性別のみ異なる属性同士の比較

表7 各属性のパターンのALLにおける出現割合

属性	ALLにおける各属性の出現割合	各属性におけるALLの出現割合
IYM	0.104	0.112
IYF	0.288	0.408
IOM	0.120	0.360
IOF	0.084	0.392
OYM	0.220	0.320
OYF	0.405	0.528
OOM	0.518	0.456
OOF	0.447	0.408
平均	0.273	0.373

位100位以内のパターンのみで比較を行った。各属性における100位以内のパターンの数は表6の3列目に示す通りである。

各属性の上位100位以内のパターンがALLにおいても100位以内に出現した割合を表7の2列目に、ALLの上位100位以内のパターンが各属性の100位以内に出現した割合を3列目に示す。ALLにおける各属性の出現割合において、OOMを除いて各属性のパターンがALLの100位以内に出現した割合は半分以下であった。特にIYM, IOM, IOFのパターンは20%すら出現しておらず (IYM: 0.104, IOM: 0.120, IOF: 0.084), 全属性の平均も27.3%と3割を超えていない。つまり、全属性の半数近くまたはそれ以上のパターンがALLのパターンの中に出現していない。

この結果から、本来各属性において頻出であるパターンのお大半が、属性による分類を行わずに系列パターンマイニングによる分析を行った場合では頻出にな

表8 出現割合のステータスごとの平均

	ALLにおける平均出現割合	ALLの平均出現割合
県内平均	0.149	0.318
県外平均	0.398	0.428
若齢平均	0.254	0.342
高齢平均	0.292	0.404
男性平均	0.241	0.312
女性平均	0.306	0.434
総平均	0.273	0.373

らないことがわかる。したがって、属性による分類を行わない系列パターンマイニングを用いた推薦では、本来は特定の属性にとって有益であるはずのパターンが推薦されないと考えられる。また、各属性におけるALLの出現割合においてはOYF以外の全属性で50%を下回った。このことから、ALLの半数近くのパターンは各属性において頻出ではないパターンであるといえ、当該パターンを推薦しても各属性のユーザが興味を示すかは未知である。以上より、スタンプ取得履歴の系列パターンマイニングによる分析の結果は属性の有無によって差異が存在するといえる。つまりは、スタンプラリーの取得履歴の分析において属性による分類は重要であるといえる。

5.2 ステータスの違いによる結果の差異

ステータスの違いによって出現割合に差があるのかを見るため、表7に示した結果の各ステータス毎の平均値を表8に示す。ALLにおける平均出現割合は、各ステータスに該当する属性のパターンのALLにおける出現割合を平均したものを指す。またALLの平均出現割合とは、各ステータスに該当する属性におけるALLのパターンの出現割合を平均したものを指す。

この結果を見ると、ALLにおける出現割合とALLの出現割合のどちらにおいても、県内(0.149)より県外(0.398)、若齢(0.254)より高齢(0.292)、男性(0.241)より女性(0.306)のほうが高い。これは県内に在住するユーザより県外在住のユーザ、若齢なユーザより高齢なユーザ、男性ユーザより女性ユーザのほう全体に共通して頻出なパターンの順に取得することが多いことを示唆している。この結果より、ステータスの違いがスタンプ取得の傾向に影響を与えていると考えられ、スタンプラリーを用いた推薦に際してはユーザ属性として居住地、年齢、性別の3つのステータスを考慮すべきだといえる。

5.3 2種類の出現割合の差

表7に示した出現割合の結果において、OOMとOOFを除くすべての属性で、ALLにおける各属性のパターンの出現割合より各属性におけるALLのパターンの出現割合のほうが高い結果となった。これは各属性で列挙されたパターン数の影響であると考えられる。ALLでの出現割合と各属性での出現割合のどちらにおいても100位以内に出現するパターンの数は一定であるため、2つの割合の差異はALLと各属性のパターン数の違いのみによって決まる。割合の計算式からいうと、分子はどちらも同じ値であり、分母はALLでの出現割合については各属性のパターン数、各属性での出現割合においてはALLのパターン数となる。いま、表6の100位以内のパターン数を見るとOOMとOOFのパターン数(OOM:110個, OOF:114個)のみがALLのパターン数(125個)を下回っている。一方で他すべての属性のパターン数はALLを上回っている。つまり、OOMとOOFのみ各属性における出現割合の分母に比べALLにおける出現割合の分母が小さくなっている。したがって、ALLよりパターン数の少ないOOMとOOFのALLにおける出現割合(OOM:0.518, OOF:0.447)は各属性における出現割合(OOM:0.456, OOF:0.408)より大きくなった。逆にOOM, OOF以外の属性はALLよりパターン数が多いので各属性における出現割合のほうが大きくなった。このことから属性によって出現割合の最大値が1のものと1未満のものがあるとわかる。本実験での比較方法で各属性の比較をより厳密に行うには各属性のパターン数が等しくなる必要がある。

なお、ALLにおけるIYMの出現割合とIYMにおけるALLの出現割合だけ両方が非常に小さい値となっている。これはIYMの分析に用いたログ数が少なく、より個人の特性に引張られた結果、特有のパターンがより多く抽出された影響が大きいと考える。個人の特性によりすぎないパターンを抽出するには十分な数のログを用意する必要がある。

6 おわりに

本研究では、観光地の推薦を想定し、系列パターンマイニングを用いることでユーザの属性ごとにスタンプラリーにおいて頻出なスタンプの取得順序の抽出を行った。実験では属性による分類の有無が抽出結果に与える影響を検証するため両者の結果の比較を行った。比較方法の1つとして各属性において頻出なパターンが属性の分類のない場合の結果に出現するかを判定した。

その結果、すべての属性の半数近くまたはそれ以上のパターンが属性の分類のない場合のパターンの中に出現しないことがわかった。また、属性の分類を行わない場合のパターンが各属性のパターンの中に出現するかを判定すると、ほとんどの属性で半数を下回った。これは、属性分類のない場合のパターンの半分以上は各属性の視点からでは頻出でないパターンであることを示している。これらの結果から、スタンプ取得履歴の系列パターンマイニングによる分析の結果は属性の有無によって差異が存在することを示した。同時に、ステータスの違いによる出現する割合の差の比較を行った。この結果では、県内に在住するユーザより県外在住のユーザ、若齢なユーザより高齢なユーザ、男性ユーザより女性ユーザのほうが全体で頻出なパターンを多く含むことを示唆している。以上のことを踏まえると、スタンプラリーを用いた系列パターンマイニングによる観光地推薦を考える場合、ユーザの属性による分類を行う必要があるといえる。

本研究の実験で用いたデータは属性をより細かく分類するほどの数を用意できなかったため、個人のスタンプの取得における特性が結果に影響を与える場合があった。そのほか生年の分類では2世代でしか分類できず、各世代の特性として扱うには信ぴょう性が低いと思われる。今後、より多くのデータを用いた分析が必要である。また、各属性における抽出パターン自体の比較は本研究では行っていないため、属性の分類を伴う推薦を行うにはこれらの結果が属性ごとにどのように違っているかを示す必要がある。その後、本研究で抽出した結果を用いた推薦方法の提案を目指したい。

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のご協力をいただきましたこと、この場をお借りしてお礼を申し上げます。ありがとうございました。

指導教員である上野秀剛准教授にはお忙しい中、数回に渡る論文の添削を丁寧に行なっていただき、的確なご指摘をいただきました。ありがとうございます。

中間発表会では、松尾賢一教授に的確なご指摘をいただき、その後の資料の作成について非常に参考になる意見をいただきました。ありがとうございます。

奈良県の各関係者様には貴重なデータのご提供をいただきました。ありがとうございます。

また、同研究室の先輩、同輩、クラスメイトの皆様には研究を進めるにあたって様々な助言や励ましを頂き、相談もさせていただきました。ありがとうございます。

参考文献

- [1] 地域IoT実装推進ロードマップ, 総務省, 平成28年12月9日.
- [2] 札幌いんふお, <http://www.hbc.co.jp/rocket/sapporoinfo> (参照: 2017-12-12).
- [3] 東京ハンディガイド, <http://www.metro.tokyo.jp/INET/OSHIRASE/2015/12/20pcm200.htm> (参照: 2017-12-12).
- [4] Yuya Kanazawa, Yosuke Hidaka and Katsuhiko Ogawa: Destination Retrieval System using an Association Retrieval Method, International Journal of Future Computer and Communication, Vol.2, No.3, pp.169-173, 2013.
- [5] 嶋田和孝, 上原尚, 遠藤勉: 集合知に基づく観光地推薦システムの構築, 観光情報学会誌, Vol.10, No.1, pp.113-124, 2014.
- [6] 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.
- [7] Takeaki Uno, Tatsuya Asai, Yuzo Uchida, Hiroki Arimura: An Efficient Algorithm for Enumerating Closed Patterns in Transaction Databases, Discovery Science, LNAI 3245, pp.16-31, 2004.
- [8] 佐々木夢, 荒堀喜貴, 串間宗夫, 荒木賢二, 横田治夫: 電子カルテシステムのオーダログ解析による医療行為の支援, DBSJ Japanese Journal, Vol.14-J, No.10, 2016
- [9] Aileen P. Wright, Adam T. Wright, Allison B. McCoy, Dean F. Sittig: The use of sequential pattern mining to predict next prescribed medications, Journal of Biomedical Informatics Vol.53, pp.73-80, 2015
- [10] ならたん, <https://www.naratan.com/> (参照: 2018-01-06)
- [11] NYSOL, <http://www.nysol.jp/> (参照: 2018-01-06)