

Wi-Fi パケットセンサーデータを用いた 散策型観光地における 観光スポットリコメンド手法の検討

花輪 圭祐¹・寺部 慎太郎²・柳沼 秀樹³・田中 皓介⁴

¹非会員 東京理科大学大学院 理工学研究科土木工学専攻 (〒278-8510 千葉県野田市山崎2641)

E-mail:7616100@ed.tus.ac.jp

²正会員 東京理科大学教授 理工学部土木工学科 (〒278-8510 千葉県野田市山崎2641)

E-mail:terabe@rsss.tus.ac.jp

³正会員 東京理科大学講師 理工学部土木工学科 (〒278-8510 千葉県野田市山崎2641)

E-mail: yaginuma@rs.tus.ac.jp

⁴正会員 東京理科大学嘱託助教 理工学部土木工学科 (〒278-8510 千葉県野田市山崎2641)

E-mail:tanaka.k@rs.tus.ac.jp

観光客のニーズは近年、多様化している。このため様々なニーズに応えることのできる観光情報提供の手法を検討する必要性がでてきた。本研究では、Wi-Fi プローブセンサーを用いて効率的に取得した観光客の周遊行動を元に観光客の特徴を推定し、観光客の好みに合った観光スポットを推定して情報提供を行うというものである。Wi-Fi プローブデータをそのまま使用した場合についてデータの観測数の偏りから、一部の観光地が強く推薦されてしまうため、複数箇所訪問した観光客のデータのみ絞ったリコメンドを実施した。その結果、精度が向上したことが確認された。さらに、観光客の各地点の訪問の有無を 0 と 1 で表現したダミー変数データを用いてリコメンドを行ったところ、Wi-Fi プローブデータを観測しにくい場所についてもリコメンドされるようになった。

Key Words : *Wi-Fi probe sensor, recommendation, Collaborative filtering, tourist information, Cosine similarity*

1. はじめに

観光客のニーズは近年、多様化している。そのため、様々なニーズに応える新たな観光情報提供手法を検討する必要性が生じている。WEB アンケート調査や SNS 等のロケーションデータを用いて、個々の観光客の興味を解析し、観光客に合わせた情報提供を行う手法が多く研究されているものの、Wi-Fi パケットセンサーデータの行動履歴を利用したリコメンドーションの研究は未だにない。

行動経路の収集手法について、GPS 等の従来の人流観測手法は、測位精度が低く屋内での測位が不可能なことから、隣接した施設を巡ったなどの狭い範囲の行動履歴を把握するためには、あまり適していない。これに対して、Wi-Fi パケットセンサーデータは詳細な観測範囲が設定できることから、隣接した施設の訪問を区別できるなど、狭い範囲の行動履歴を把握することに適している。また、Wi-Fi 搭載機器の所持率は高齢者を含めて、上昇を続けている。高齢者のスマートフォン所持率は H26 年に 30% 程度であったが、H30 年には 55% 程度まで増えている。高齢者のみならず、全年齢に

おいても、H26 年に 65% であったスマートフォン所持率も 80% 程度まで上昇した²⁾。また、スマートフォンだけでなくタブレット保有率も上昇し、取得データの信頼性は、向上しつつある。

ビッグデータを分析することで観光地を推薦する研究として、中嶋ら³⁾は Twitter や Foursquare, Instagram などの SNS のツイートから複数人の観光周遊行動を取得し、その数学的距離を用いて観光客の好みを推測した。ツイート中に含まれる、位置情報やキーワードから、食事、景観、行動の 3 パターンに関連する観光に関するツイートを収集し、そのツイートをもとに観光ルートの抽出を行った。人気の観光地においてはツイート数が多く、推薦されやすくなるという側面も確認された。

樽井⁴⁾は協調フィルタリングとコンテンツ分析を利用して、全国規模での観光地の推薦を行う手法のプロトタイプを作成した。複数の観光地に対して、観光地の特徴ダミー (自然、保養、レジャー、文化、歴史遺産、グルメ) を設定し、これと観光地を行列とした観光地特徴ベクトルを設定した。さらにアンケートから



図-1 小布施町 Wi-Fi センサー設置地点

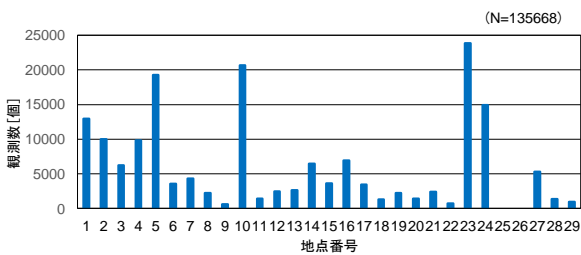


図-2 各地点の MAC アドレス観測数

得られた観光客の行動記録利用者履歴ベクトルと設定して、これと先ほどの観光地特徴ベクトルを掛け合わせて正規化を実施し、利用者特徴ベクトルを導いた。この行列に相関係数を用いて、利用者同士の数学的距離を導いて観光地を推薦した。

これらを踏まえつつ、本研究では Wi-Fi プローブデータを用いて取得できた行動履歴を元に、コサイン類似度を用いたリコmendを行うと、どのような結果が取得可能か、またどのような特徴が生じるか検討した。

2. データの概要

リコmend手法を検討するにあたっては、小布施町において本研究室および小布施町の協力で実施した Wi-Fi パケットセンサーによる観光周遊調査のデータを使用した。

本調査の全 29 地点に Wi-Fi パケットセンサーと呼ばれるデータ収集のための機器を設置し (図-1)、センサー周辺に存在するスマートフォンなどから自動的にプローブ要求の通信データを取得した。調査対象が観光客であることから、一般的に観光客は土日祝日に集中することや小布施町の繁忙期が秋であることを踏まえ、2017年10月21日(土)~22日(日)の2日間に調査を行った。ここで得られたデータのうち、観光客以外のデータをクリーニングするため、複数日で観測されていないデータを抽出した。今回は2日間のうち、データクリーニング済みの10月21日(土)のデータを

用いる。

各地点における MAC アドレスの観測数は図-2の通りである。地点による偏りは比較的大きく、数か所で局地的にカウント数が上昇することがわかる。

3. リコmendの方法⁵⁾

本研究では、リコmendを行うにあたって、コサイン距離を用いた協調フィルタリングのアルゴリズムを用いた。協調フィルタリングでは、次の4段階を経て、それぞれの観光スポットが各人にどれほど適しているのかの値を出力するものである。

まず第1段階は、個人ごとに各地点に対する評価を定め、その評価をベクトルとする。第2段階は、その複数のベクトル同士について、(1)の式でコサイン距離を算出し、リコmendのターゲットにするベクトルに、コサイン距離の近いベクトルを一定数を選び出す。第3段階は、コサイン距離と各人の地点ごとの観測回数を掛け合わせ、最後に、地点ごとに類似度に基づいて加重平均を求め、その値を地点ごとの平均類似度とする。最後に平均類似度が大きい順に、ターゲットへリコmend結果として順位を表示する。なお今回の標本では、センサー地点番号 No.12, No.22, No.25, No.26 で十分な数が取得できなかったため、リコmendの評価対象としていない。

$$\cos_dist(x, y) = \frac{\vec{x} \cdot \vec{y}}{|\vec{x}| |\vec{y}|} = \frac{\sum_{i=1}^{|\vec{v}|} x_i y_i}{\sqrt{\sum_{i=1}^{|\vec{v}|} x_i^2} \sqrt{\sum_{i=1}^{|\vec{v}|} y_i^2}} \quad (1)$$

$\cos_dist(x, y)$: コサイン距離 x : リコmendターゲット
 y : 蓄積したデータのうちの一人 x_n, y_n : n 番目の映画への評価

4. リコmendの実行

はじめに、調査で収集したデータを、端末ごとに固有なアドレスである MAC アドレスで整理する。MAC アドレスがどの地点で何回観測されたかを集計し、「MAC アドレス×地点」の行列で整理した。続いてリコmendされるターゲットデータを作成する。先ほどの行列からランダムに1行分のデータを抽出し、そこから、ある地点 A の観測回数を削除し、その人が地点 A を訪問しなかったと仮定する。そして、この仮定をしたターゲットデータにふさわしい観光スポットの推薦を実施し、削除した地点 A が含まれるかどうかを確認する。

リコmendをするにあたって比較検討するために標本群を二通り用意した。無加工の行動履歴 X と 3 か所以上の訪問がある行動履歴 Y の 2 種類の標本群である。

これらの標本群を用いてそれぞれ同じターゲットデータに対してリコメンドし、比較した。

一例として、地点 No.1 での観測回数を削除し、リコメンドのターゲットとした実行例を示す。結果は表-1 の通りとなった。削除した地点は 1 位としてリコメンドされた。図-2 より、No.1 では観測数が多いことも順位を上げる要因となっていると考えられる。

また、削除する地点を変更し、地点 No.3 「北斎館」(美術館)での観測記録を削除したものをリコメンドのターゲットとした。結果は表-2 の通りである。この結果より、標本群を複数の地点で観測された MAC アドレスで構成することで、精度が向上することが分かった。しかし、順位に関しては No.3 「北斎館」の記録を削除した場合に、順位が下がってしまった。記録数が No.3 より多い別の他の地点 No.23 と No.24 の影響を受けてしまったと考えられる。

5. ダミー変数データの使用

第 4 章では、MAC アドレスが各地点で観測された回数をもとに推薦を行ったが、図-2 で示される観測回数の多寡に大きく影響を受けていることが伺われた。そこで、観測回数の代わりにその地点で観測されたかどうかを示す 1 と 0 を用いたダミー変数データを使用することにより、推薦結果にどのような差異が生じるかを考察する。

無加工の行動履歴を利用して推薦した表-1 比較して、表-5 の観測結果で示されるダミー変数データを使用した結果では得点の散らばりが減少した。特に注目すべき所としては、No.1 「日本のあかり博物館」は中心街から奥まった所で、観測回数が少ないという特徴を持っているにもかかわらず表-5 のようにリコメンドされた。つまり観測されにくい場所においても推薦された。

6. リコメンド結果の精度評価

リコメンドの精度の評価は再現率を用いて実施した。すべてのターゲットに対してリコメンド結果に仮に削除した地点が含まれている確率を指す。100 人分のデータを抽出し、これに対し推薦するものとする。本研究では、無加工の行動履歴 X と 3 か所以上の訪問がある行動履歴 Y の 2 種類の標本群を用いてリコメンドし、比較した。結果はそれぞれ、図-3、図-4 で示す。

表-1 No.1 を削除したターゲットへの結果例

順位	得点	番号	地点名
1	0.419	1	桜井甘精堂・茶蔵 (さくら)
2	0.033	7	森の駐車場
3	0.032	4	小布施堂本店
4	0.024	5	竹風堂小布施本店
5	0.009	27	桜井甘精堂駐車場
6	0.008	15	サンクゼールワイナリー小布施

表-2 No.3 を削除したターゲットへの結果例

順位	得点	番号	地点名
1	0.156	4	小布施堂本店
2	0.098	5	竹風堂小布施本店
3	0.069	27	桜井甘精堂駐車場
4	0.009	3	北斎館

表-3 No.1 を削除したターゲットへの結果例
(3 地点以上訪問歴あり, 7 位以下は省略)

順位	得点	番号	地点名
1	3.185	1	桜井甘精堂・茶蔵 (さくら)
2	1.623	5	竹風堂小布施本店
3	1.190	4	小布施堂本店
4	0.853	27	桜井甘精堂駐車場
5	0.727	15	サンクゼールワイナリー小布施
6	0.530	23	JA ながの農産物直売所

表-4 No.3 を削除したターゲットへの結果例
(3 地点以上訪問歴あり, 7 位以下は省略)

順位	得点	番号	地点名
1	1.735	4	小布施堂本店
2	1.351	5	竹風堂小布施本店
3	0.417	27	桜井甘精堂駐車場
4	0.189	24	道の駅オアシスおぶせ
5	0.187	23	JA ながの農産物直売所
6	0.118	3	北斎館

表-5 ダミー変数データを使用したターゲットへの結果例 (No.1 を削除)

順位	得点	番号	地点名
1	0.027	15	サンクゼールワイナリー小布施
2	0.009	1	桜井甘精堂・茶蔵 (さくら)
3	0.009	5	竹風堂小布施本店
4	0.009	6	日本のあかり博物館
5	0.009	4	小布施堂本店

表-6 ダミー変数データを使用したターゲットへの結果例 (No.3 を削除, 6 位以下は省略)

順位	得点	番号	地点名
1	0.055	4	小布施堂本店
2	0.053	5	竹風堂小布施本店
3	0.032	15	サンクゼールワイナリー小布施
4	0.032	3	北斎館
5	0.022	6	日本のあかり博物館

		地点番号																													
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	
含まれる確率	上位 x 位置	1	67	41	37	9	37	8	12	6	9	87	0	0	1	11	2	1	83	6	6	0	58	0	0	1	0	0	3	9	0
		2	74	43	40	24	42	16	15	7	7	90	2	0	1	16	13	1	83	6	9	3	58	0	1	1	0	0	31	17	1
		3	78	43	41	35	46	17	19	11	10	90	9	0	2	19	18	2	83	6	10	3	58	0	1	1	0	0	31	17	1
		4	78	44	45	38	48	18	19	17	10	90	14	0	2	20	26	3	83	6	11	64	58	0	2	1	0	0	46	24	1
		5	79	45	45	39	53	19	22	19	11	90	15	0	2	20	37	3	83	6	16	64	58	0	3	3	0	0	51	28	1
		6	80	46	46	39	53	19	23	26	12	91	18	0	3	20	39	5	84	6	21	65	58	0	3	5	0	0	54	34	1
		7	80	46	46	40	53	20	25	30	12	91	20	0	3	20	39	5	84	6	21	65	58	0	4	5	0	0	60	35	1
		8	80	46	46	40	53	20	25	31	12	91	20	0	4	21	40	5	84	6	21	65	58	0	4	6	0	0	62	36	1
		9	80	46	46	40	53	20	25	32	12	91	21	0	4	21	40	5	84	6	21	65	58	0	4	6	0	0	62	37	1
		10	80	46	46	40	53	20	26	32	12	91	21	0	4	21	40	5	84	6	21	65	58	0	4	6	0	0	62	37	1

割合 ■ <-0% □ 100%-> ■
図-3 Xの精度評価(無加工, N=5000)

		地点番号																													
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	
含まれる確率	上位 x 位置	1	63	93	71	14	71	85	40	1	0	89	3	0	1	69	14	3	81	90	7	0	54	0	86	3	0	0	19	2	8
		2	68	94	74	32	89	52	51	4	1	96	7	0	4	73	29	6	82	90	12	0	58	0	89	3	0	0	34	13	10
		3	73	98	95	42	98	92	58	6	3	96	9	0	7	78	44	6	82	91	15	0	73	0	91	6	0	0	49	25	16
		4	85	98	97	50	95	64	12	4	98	12	0	14	84	64	9	83	91	21	0	74	0	94	88	0	0	72	35	20	
		5	89	99	98	58	97	67	14	11	99	16	0	21	90	64	12	84	91	24	0	74	0	94	89	0	0	78	37	20	
		6	92	100	99	57	99	98	71	19	105	22	0	26	91	95	15	84	91	30	0	74	0	94	89	0	0	79	41	20	
		7	93	100	100	58	99	99	78	27	111	22	0	26	91	95	15	84	91	34	0	74	0	95	89	0	0	86	42	24	
		8	99	100	100	58	99	99	79	37	23	100	79	0	49	93	90	32	86	92	39	1	74	0	97	92	0	0	89	45	26
		9	99	100	100	59	99	99	82	39	27	100	85	0	54	94	90	34	87	93	43	64	74	0	97	92	0	0	96	48	26
		10	99	100	100	59	99	99	86	41	34	100	86	0	56	94	90	38	89	93	45	64	74	0	98	96	0	0	100	62	29

割合 ■ <-0% □ 100%-> ■
図-4 Yの精度評価(3か所以上で観測, N=5000)

		地点番号																													
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	
含まれる確率	上位 x 位置	1	73	48	17	26	41	3	5	1	0	96	1	0	10	20	31	0	77	6	0	61	57	0	0	0	0	19	4	0	
		2	79	48	21	37	47	11	14	7	0	91	4	0	13	27	39	0	77	6	0	61	57	0	0	1	0	42	30	1	
		3	83	48	21	38	50	17	24	7	0	92	9	0	15	28	45	0	77	6	3	61	57	0	1	0	0	64	32	1	
		4	83	49	22	41	57	23	32	10	0	92	17	0	21	29	46	4	79	1	8	61	57	50	2	0	0	69	41	0	
		5	83	49	22	42	57	23	34	13	3	92	19	0	26	29	47	5	84	5	9	61	57	50	2	0	0	79	45	1	
		6	83	49	22	42	57	23	36	14	8	92	27	0	36	29	47	6	84	5	16	61	57	50	2	3	0	0	72	46	3
		7	83	49	22	42	57	23	40	17	15	92	32	0	42	29	48	9	86	6	28	61	57	50	3	4	0	0	75	46	6
		8	83	49	22	42	57	23	43	25	23	92	34	0	47	29	48	11	87	8	31	61	57	50	3	6	0	0	73	46	6
		9	83	49	22	42	57	23	43	26	27	92	34	0	48	29	48	17	89	9	36	61	57	50	3	6	0	0	73	46	14
		10	83	49	22	42	57	23	43	29	32	92	34	0	48	29	48	20	89	9	36	61	57	50	4	7	0	0	73	46	20

割合 ■ <-0% □ 100%-> ■
図-5 Zの精度評価(ダミーデータ, N=5000)

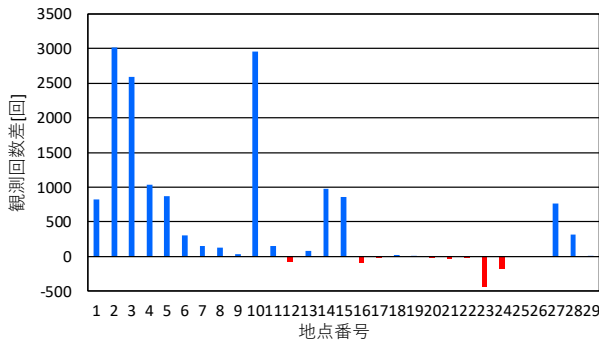


図-6 XとYの地点ごとの標本数の差 (X-Y)

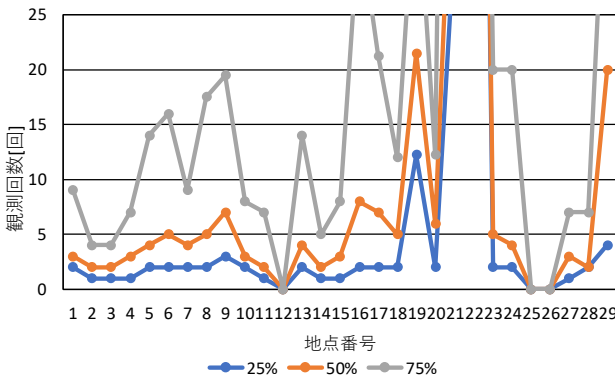


図-7 Yの標本群の観測回数の四分位数

Xを用いてリコメンドを行う場合は、多くの地点で、0~40%程度の評価になった。

ところが、標本群 Y では、No.1~No.7 や、No.14、~No.24などで無加工のものよりも精度が上昇した。これら要因としては、それぞれの標本群の訪れた地点の違いと、多くのプローブ要求を発信したことによるデータの偏りの可能性がある。標本群 X と Y の各地点ごとの観測標本数の差によって、地点番号 No.1~No.7 まで大きく精度が改善していると考えられる。これは、図-6のように標本群 X よりも Y が、1 地点当たりの観測回数が増加しているためである。さらに図-7 に示される通り、標本数が 0 であった地点を除いて、観測回数の四分位範囲は広い範囲をとっていることが分かった。以上の理由から、先に挙げた理由が精度上昇の一因であるといえる。

またダミー変数データ Z を利用した場合の精度は、図-5 のように無加工に比べて若干の改善がみられた。なお、図-2 の各地点の MAC アドレス観測回数グラフに比較して、おおよそそれぞれの地点の精度は MAC アドレス観測数に一致しているようにみえる。このことから、コサイン類似度を用いたリコメンドには各地点の観測数の影響を無視せず、MAC アドレスの観測数に合わせた縮尺係数を導入することで、ある地点が極端にリコメンドされにくくなってしまいうを防ぐことができると期待される。

7. おわりに

本研究では、Wi-Fi パケットセンサーデータを蓄積し、協調フィルタリングを用いて、異なる標本群を元にリコメンドを実施し有効であるかを検討した。その結果、リコメンドの元となるデータに 3 か所以上で観測された MAC アドレスのみを使用すること、さらに MAC アドレスの観測数が少ない地点においてはダミー変数データを使用することすることがリコメンドの精度を高める結果に寄与することが明らかになった。

参考文献

- 1) 寺部慎太郎, 一井啓介, 柳沼秀樹, 小野瑞樹, 田中皓介, 康楠: Wi-Fi パケットセンサーを用いた歩行者行動・観光客周遊行動研究の包括的レビューとそれを踏まえた分析例示, 土木学会論文集 D3 (土木計画学) Vol.75, No.5 [土木計画学研究・論文集 36 卷], pp.1669-1679, 2019
- 2) 平成 30 年 通信利用動向調査報告書 (世帯編, 総務省)
- 3) 中嶋勇人, 新妻弘崇, 太田学: 位置情報付きツイートを利用した観光ルート推薦, 情報処理学会研究報告, Vol.2013-DBS-158, No.28, 2013
- 4) 樽井勇之: 協調フィルタリングとコンテンツ分析を利用した

観光地推薦手法の検討, 上武大学経営情報学部紀要第 36 号,

p.1-14, 2011

(??.?受付)

5) 片田陽介: フリーライブラリで学ぶ 機械学習入門, 第 5 章

レコメンデーション入門, p146-p166, 2017

EXAMINING RECOMMENDATION SYSTEM FOR TOURIST SPOTS BY WI-FI SCANNER DATA

Keisuke HANAWA, Shintaro TERABE, Hideki YAGINUMA, Kosuke TANAKA

These days, tourists are visiting more various tourism spots than before. Therefore, we need a new method of providing information of tourist spots. In this study, we examined recommendation system for tourism spots by Wi-Fi scanner data. We ran recommendation system by Collaborative Filtering and evaluated the result by recall-proportion. We confirmed that result of recommendation improved by sample groups who visited many places. It is caused by difference of number of samples in the group, and observations of prove-requests.