

位置情報サービスの利用状況を活用したPOI推薦手法

Point-of-Interest Recommendation using Location-based Service Usage

落合 桂一^{*1*2}

Keiichi Ochiai

深澤 佑介^{*1}

Yusuke Fukazawa

山田 渉^{*1}

Wataru Yamada

松尾 豊^{*2}

Yutaka Matsuo

^{*1}株式会社 NTT ドコモ

NTT DOCOMO, INC.

^{*2}東京大学

The University of Tokyo

With the widespread of smartphones which can easily obtain location using GPS, location-based service (LBS) become popular. Point-of-Interest (POI) recommendation is one of such services. In the existing works, geographical and temporal information is exploit to recommend POIs for users. However, in particular LBS, content distribution affects user's place of visit. In this paper, we propose a method to exploit LSB usage for POI recommendation. Our evaluation on twitter dataset shows that the proposed method outperforms the baseline method.

1. はじめに

スマートフォン上で GPS, Wi-Fi など位置情報の取得が容易に行えるようになったため、位置情報を利用したサービスやゲームが普及している。例えば 2016 年 7 月に Niantic 社がリリースした Pokémon Go^{*1} は世界中で利用されている。このような位置情報サービスの普及を背景に位置情報サービスの履歴を活用した POI (Point-of-interest) 推薦の研究が行われている。従来研究では、POI 推薦のためにチェックインの時間帯 [Yuan 13] やチェックイン間の距離 [Ye 11], ソーシャルネットワークの友人関係 [Ma 09] を活用した推薦が行われていた。しかしながら、位置情報サービスでは、その場に行くことでゲーム内のアイテムやポイントを手に入れるようになるものもあり、ユーザが利用しているサービスも訪問場所に影響を与えると考えられる。例えば、Pokémon Go では希少なキャラクターが特定の場所で出現しやすかったり、他のプレイヤーと戦う「ジム」やアイテムを手に入れる「ポケストップ」などのコンテンツが実世界の場所と関連付けられている。そのため、仮に公園に頻繁にチェックインしているユーザがいたとしても、公園で位置情報ゲームを行っているだけであり、公園自体に興味があるわけではない可能性もある。また、写真を共有するソーシャルメディアの Instagram^{*2} では、「インスタ映え」と言われるお洒落な写真が撮れるということが動機となり特定の施設を訪れるユーザも存在する^{*3}。つまり、利用しているサービスがユーザの訪問場所に影響を与えると考えられる。

そこで、本研究では位置情報サービスの利用状況を活用した POI 推薦手法を提案する。同一の位置情報サービスを利用しているユーザは類似する場所に訪問する可能性が高いと考えられる。そのため、予備実験としてチェックイン回数、チェックイン場所、チェックイン間の移動距離などを比較し、位置情報サービスのユーザの行動への影響を分析する。次に、位置情報サービスの利用がユーザの行動に与える影響に基づき、ユーザの利用サービスを考慮した POI 推薦手法を提案する。Foursquare^{*4} を通じて Twitter^{*5} に投稿されたツイートから

チェックイン履歴を収集し実験を行い、提案手法の有効性を確認した。

本研究の貢献は以下の通りである。

- 利用サービスによって行動に偏りがあることをクロスドメインの推薦に活用した POI 推薦手法を提案した。
- 実データを用いて利用サービスにより行動に偏りがあることを定量的に明らかにした。
- Pokémon Go ユーザを対象に提案手法による POI 推薦の有効性を示した。

2. 予備実験

本章では提案手法の予備実験として位置情報サービスを利用することによる行動の変化を分析する。本稿では Pokémon Go のユーザを対象として、サービス利用開始前後での行動の比較およびサービス利用者而非利用者での比較を行い、行動の違いを明らかにする。なお、本章の分析に用いたデータの詳細は 4 章で述べる。

2.1 定義

本研究ではユーザの行動を、一定期間内にユーザがチェックイン (訪問) した場所の集合であるチェックイン履歴とする。 U を全ユーザの集合、 L を全 POI の集合、 T を時間の集合 (例えば 1 日を 24 個に区切った時間帯)、それぞれの要素をユーザ $u \in U$, POI $l \in L$, 時刻 $t \in T$ とし、チェックインおよびチェックイン履歴を以下の通り定義する。

定義 (チェックイン) ユーザ u のチェックイン v_u はユーザ-POI-時刻の 3 つ組 (u, l, t) で表され、ユーザ u が POI l に時刻 t に訪問したことを示す。

定義 (チェックイン履歴) ユーザ u の i 番目のチェックインを $v_{u,i}$ とすると、ユーザ u のチェックイン履歴は $h_u = \{v_{u,1}, v_{u,2}, \dots, v_{u,n}\}$ と表される。これをユーザ u の行動であるチェックイン履歴と定義する。

2.2 チェックイン回数の比較

本節では位置情報サービスの利用有無がユーザのチェックイン回数 $|h_u|$ に影響を与えるかどうかを検証する。一般に協調フィルタリングではデータ密度が高いほど推薦性能が高いと言われている [Su 09]。チェックイン回数に着目したのは、提案手法では 3 章で述べるように協調フィルタリングによる推薦

連絡先: 落合 桂一, 株式会社 NTT ドコモ, ochiaike@nttdocomo.com

*1 <http://pokemongo.nianticclabs.com/>

*2 <https://www.instagram.com/>

*3 <https://retrip.jp/articles/61957/>

*4 <https://foursquare.com/>

*5 <https://twitter.com/>

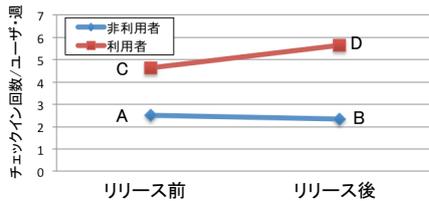


図 1: サービス利用者/非利用者の平均チェックイン回数の比較

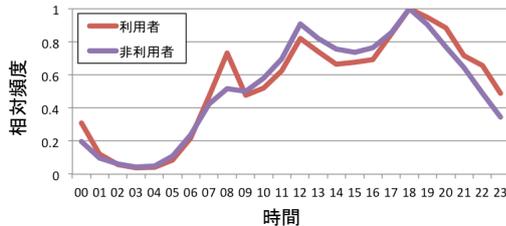


図 2: サービス利用者/非利用者のチェックイン回数推移の比較

を行っており、回数が多いほどデータ密度が高く推薦精度が高くなると考えられるためである。

サービスの利用がチェックイン回数の増加要因となるか検証するために差分の差分分析 [Althoff 16] によりチェックイン回数の差を求める。サービス利用者と非利用者のサービス提供開始前後でのユーザごとの週あたりの平均チェックイン回数を図 1 に示す。図 1 中の $(D - C) - (B - A)$ よりサービスを利用することで週あたり 1.177 回のチェックイン回数が増加したことがわかる。データ密度は利用者のみでユーザ-POI 行列を構成した場合 0.52%，全ユーザの場合 0.15% であり増加していた。

2.3 チェックインスポットの比較

本節ではサービス利用者のチェックイン場所 p に偏りがあるかを調べるため、サービス利用者とチェックインデータがある全ユーザにおけるチェックインランキングを調査する。表 1 に (a) 全ユーザを対象にチェックイン数を集計したランキング (b) サービス利用者内でチェックイン数を集計したランキング、および (c) サービス利用者内に特有のチェックイン場所を抽出するためにリフト値を計算したランキングの各上位 20 件を示す。チェックイン数は最大値で正規化し相対頻度とした。全ユーザでのランキングではさいたまスーパーアリーナなど大型イベント会場や主要な駅が多く入っている。それに対して、サービス利用者内でのランキングでは駅の他に公園が多く入っている。またリフト値を計算しサービス利用者内に特有のチェックインを抽出した結果でも公園が多く入っている。例えば錦糸公園、鶴舞公園などはサービスリリース当初に希少なキャラクターが入手できるというニュース^{*6} が出ておりサービスの影響で上位に入ったと考えられる。

2.4 チェックイン時間帯の比較

従来研究 [Yuan 13] で時間帯を POI 推薦に利用することが有効であると示されている。サービス利用者と非利用者のチェックイン時間帯を比較することでサービス利用有無と時間帯を組み合わせることの有効性を明らかにする。

図 2 にサービス利用者と非利用者の時間帯ごとのチェックイン回数 $|h_{*,t}|$ の推移を示す。 $h_{u,t}$ はユーザ u 、時刻 t のチェッ

*6 http://www.huffingtonpost.jp/2016/07/23/pokemon-twitter_n.11159780.html

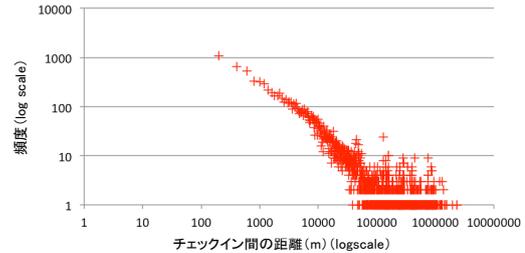


図 3: チェックイン間距離

クイン集合、* は全ユーザを示す。図ではチェックイン回数の最大値で正規化している。この図からサービス利用者と非利用者でチェックインの時間帯に差があること、特に朝 7、8 時と夜 20 時以降のチェックインがサービス利用者では多いことがわかる。

2.5 チェックインの地理的特性の比較

従来研究 [Gonzalez 08] で、移動距離が指数法則に従う人の行動特性が知られている。POI 推薦の従来研究 [Yuan 13, Ye 11] でも、この特性を活用して POI を推薦している。そこで、本研究でも検証の対象とするデータが同様の特性があるか検証する。図 3 にサービス利用者の連続するチェックイン間の距離の頻度を両軸とも対数を取ってプロットした結果を示す。この結果から、本研究で扱うサービス利用者のチェックイン履歴においても指数法則に従っていることがわかる。

3. POI 推薦手法

前節での分析より、サービス利用者と非利用者で行動に差があることがわかる。そこで、本研究では特定のサービスを利用しているユーザ同士で POI を推薦することで推薦性能を向上する手法を提案する。提案手法は最も基本的な推薦手法の 1 つであるユーザベース協調フィルタリングを拡張したものである。そのため、はじめにユーザベース協調フィルタリングを用いた POI 推薦について説明する。次に提案手法である特定のサービスの利用有無を用いた POI 推薦について述べる。

3.1 ユーザベース協調フィルタリング

ユーザベース協調フィルタリングでは、推薦対象のユーザと他のユーザの類似度を計算し、他ユーザの履歴をユーザ間の類似度で重み付けすることで推薦対象のアイテムの推薦スコアを計算する。各ユーザの訪問場所を示すユーザ-POI 行列 C の各要素 $c_{i,j}$ は、ユーザ $u_i \in U$ が POI $l_j \in L$ にチェックインしているときに、 $c_{i,j} = 1$ 、チェックインしていないときに $c_{i,j} = 0$ とする。ユーザ u_i に対する各 POI の推薦スコア $\hat{c}_{i,j}$ は以下の式で計算する。なお、表記は [Ye 11] に従っている。

$$\hat{c}_{i,j} = \frac{\sum_{u_k} w_{i,k} c_{k,j}}{\sum_{u_k} w_{i,k}} \quad (1)$$

$w_{i,k}$ はユーザ u_i と u_k の類似度を示し、従来研究と同様に次式で表されるコサイン類似度により計算する。

$$w_{i,k} = \frac{\sum_{l_j \in L} c_{i,j} c_{k,j}}{\sqrt{\sum_{l_j \in L} c_{i,j}^2} \sqrt{\sum_{l_j \in L} c_{k,j}^2}} \quad (2)$$

3.2 利用サービスに基づく協調フィルタリング

本節では提案手法である利用サービスに基づく協調フィルタリングについて説明する。2.2 節の分析より利用サービスで

表 1: チェックインスポットランキング

順位	(a) 全ユーザでのランキング		(b) サービス利用者内でのランキング		(c) 利用者内でのリフト値によるランキング	
	スポット名	相対頻度	スポット名	相対頻度	スポット名	正規化したリフト値
1	秋葉原駅	1	秋葉原駅	1	目黒川	1
2	東京駅	0.7907	錦糸公園	0.8095	小山内裏公園	0.9774
3	ヨドバシカメラマルチメディア Akiba	0.7885	渋谷駅	0.7381	江古田の森公園	0.7959
4	さいたまスーパーアリーナ	0.5725	東京駅	0.7143	山下公園	0.7143
5	名古屋駅	0.5392	不忍池	0.6905	木場公園	0.6842
6	幕張メッセ	0.4386	ヨドバシカメラ マルチメディア Akiba	0.6667	寛永寺 不忍池 弁天堂	0.646
7	渋谷駅	0.4223	世田谷公園	0.619	大阪城	0.5714
8	池袋駅	0.3617	代々木公園	0.5476	不忍池	0.4896
9	ヨドバシカメラ マルチメディア梅田	0.3536	名古屋駅	0.5238	円山公園	0.4398
10	アキバ ソフマップ 1号店	0.3158	池袋駅	0.4762	新宿御苑 新宿門	0.4127
11	大阪駅	0.3143	扇町公園	0.4524	ポケモンスター 東京駅店	0.3824
12	鈴鹿サーキット	0.3055	鶴舞公園	0.4286	白川公園	0.3714
13	アニメイト 秋葉原店	0.3047	品川駅	0.381	扇町公園	0.3676
14	仙台駅	0.304	横浜駅	0.3333	宇都宮城址公園	0.3439
15	AKIBA カルチャーズ劇場	0.267	大阪駅	0.2857	八坂神社	0.3439
16	横浜駅	0.2618	さいたまスーパーアリーナ	0.2857	イトーヨーカドー 錦町店	0.3316
17	品川駅	0.2611	レインボーブリッジ	0.2619	赤羽駅 3-4 番線ホーム	0.3202
18	京都駅	0.2589	池袋西口公園	0.2619	世田谷公園	0.3037
19	京セラドーム大阪	0.253	アキバ ソフマップ 1号店	0.2619	錦糸公園	0.2883
20	道の駅 針 T・R・S (針テラス)	0.2441	SMILE GARDEN	0.2381	代々木公園	0.2637

ユーザを区切ることでユーザ-POI 行列のデータ密度が高くなることが期待される。また 2.3 節の分析から、同一の位置情報サービスを利用しているユーザは類似する場所に訪問する可能性が高いと考えられる。そこで、ユーザベース協調フィルタリングの計算を行うユーザを、サービス利用者のみを対象に計算する。具体的にはユーザ u_i に対する各 POI の推薦スコア $\hat{c}_{i,j}$ は以下の式で計算する。

$$\hat{c}_{i,j}^{(s)} = \frac{\sum_{u_k \in U_S} w_{i,k} C_{k,j}}{\sum_{u_k \in U_S} w_{i,k}} \quad (3)$$

U_S は特定のサービス利用者の集合を示す。 $w_{i,k}$ は式 (2) により計算する。

3.3 統合モデル

ここまで述べた 2 つの協調フィルタリングの推薦スコアを統合し、最終的な推薦スコアを計算する方法について述べる。本研究では、ユーザベース協調フィルタリングによる推薦スコアと利用サービスに基づく協調フィルタリングの推薦スコアの線形結合により最終的なスコアを計算する。そのため、まず各推薦スコアを次式により正規化する。

$$\tilde{c}_{i,j} = \frac{\hat{c}_{i,j}}{\max_{l_j \in L-L_i} (\hat{c}_{i,j})} \quad (4)$$

$$\tilde{c}_{i,j}^{(s)} = \frac{\hat{c}_{i,j}^{(s)}}{\max_{l_j \in L-L_i} (\hat{c}_{i,j}^{(s)})} \quad (5)$$

ここで L_i はユーザ u_i がチェックインしたことがある POI の集合を示す。そして、未訪問の POI に対する推薦スコアは以下の式で計算する。

$$c_{i,j} = (1 - \alpha) \times \tilde{c}_{i,j} + \alpha \times \tilde{c}_{i,j}^{(s)} \quad (6)$$

2.4 節、2.5 節で検証した通り、時間帯や 1 つ前のチェックインからの距離も推薦精度に影響を与えるため、従来研究 [Ye 11, Yuan 13] と同様にそれらの要素も考慮することで、より高精度な推薦が可能であると考えられる。本研究では、サービス利用の有無を活用した POI 推薦の有効性を確認するため式 (6) の通り基本的なユーザベース協調フィルタリングのスコアとサービス利用者のみでの推薦スコアの線形結合としている。

4. 評価実験

4.1 データセットおよび実験設定

チェックインデータには 2016 年 7 月 1 日～8 月 12 日の Foursquare のデータを利用した。データは Foursquare を通じて Twitter に投稿されたツイートから取得した。Foursquare を投稿元とするツイートは本文が「I'm at 施設名称 in 市, 県 <https://~>」や「ユーザ記述文 (@ 施設名称 in 市, 県) <https://~>」となる。ここから施設情報を抽出した。ツイートは Twitter API ^{*7} により取得した。以上の処理により 943,472 件, 79,221 ユーザのチェックインデータを収集した。その中で、7 月 22 日～8 月 12 日の期間で「ポケモン」「ポケストップ」などサービスを利用していると判断できるキーワードを含むツイートを 3 件以上投稿しているユーザを Pokémon Go のサービス利用者、キーワードを含む投稿が 0 件のユーザをサービス非利用者とした。前述の条件により投稿を抽出しているため、Pokémon Go と Foursquare を同時に利用している必要はない。そのため Pokémon Go 利用者と判定されたユーザのチェックインの中でも Pokémon Go と関連のない場所も含まれる。Pokémon Go の利用者は 1,055 ユーザ、非利用者は 46,644 ユーザであった。2 章で行った分析には以上のフィルタリングを行ったデータを用いた。3 章の POI 推薦の評価を行うためのデータは、先行研究 [Yuan 13] に従い前述のデータに対して、5 人以上がチェックインしている POI、且つチェックインした POI が 5 箇所以上存在するユーザにフィルタリングして実験を行った。この条件でのユーザ数は、Pokémon Go の利用者が 386 ユーザ、非利用者は 11,500 ユーザであった。

POI 推薦の評価実験では各ユーザの 20 % のチェックイン POI を正解データとして削除し、残り 80 % のチェックインデータから推薦を行った。推薦は推薦スコアが高い上位 N 件を利用する。実験では N=1, 5, 10, 20 とした。

4.2 評価指標

評価指標には従来研究で用いられている precision@N, recall@N および accuracy@N を用いる。precision@N は推薦された POI のうち正解データ含まれる POI の割合、recall@N

*7 <https://dev.twitter.com/rest/public>

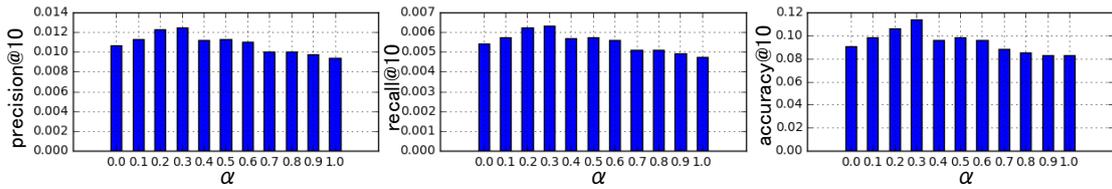


図 4: パラメータチューニング

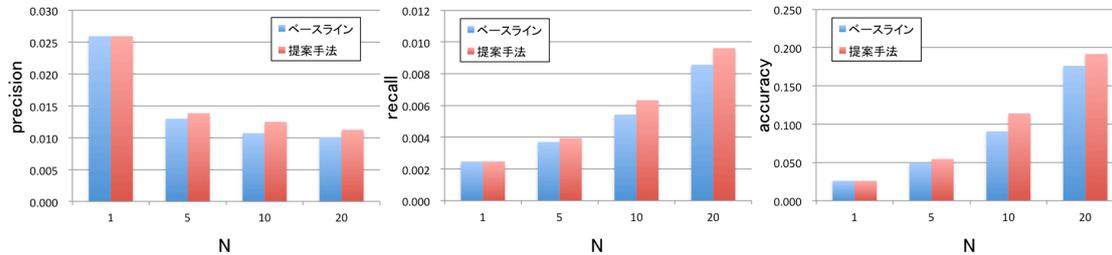


図 5: 性能評価結果

は正解データのうち推薦された POI に含まれる割合, $accuracy@N$ は評価対象ユーザのうち, 推薦された上位 N 件の POI のうち少なくとも 1 件の POI が正解データに含まれるユーザの割合である.

4.3 パラメータチューニング

提案手法ではユーザベース協調フィルタリングと利用サービスに基づく協調フィルタリングの推薦スコアを統合するためのパラメータ α がある. 図 4 に $N=10$ とした場合のパラメータを 0.0 から 1.0 まで 0.1 刻みで変化させたときの評価指標の変化を示す. $\alpha = 0.3$ が最も性能が高かった. そのため, 次節の評価では $\alpha = 0.3$ として評価を行った.

4.4 評価結果

ユーザベース協調フィルタリングをベースラインとして, 提案手法による推薦の効果を評価した. 評価結果を図 5 に示す. N を変化させた場合, $N=5$ 以上において提案手法が従来手法より高精度に推薦を行うことができた. 各手法の推薦結果に対して符号検定を行ったところ, $N=10$ の場合は $p = 0.038 < 0.05$ となり有意な差があることを確認した.

5. 関連研究

POI 推薦に関する従来研究では, 地理的な距離を活用する手法 [Ye 11], POI ごとにチェックインされる時間帯を活用した手法 [Yuan 13] などがある. これらの研究ではベースにはユーザベース協調フィルタリングを利用し, 各特徴を追加することで推薦精度の向上を図っている. 本研究のようにサービスの利用状況を POI 推薦に活用している研究ではない.

Althoff らの研究 [Althoff 16] ではリストバンド型の活動量計と Bing の検索ログを組み合わせて, Pokémon Go ユーザの活動量がサービスの利用前後で増加することを報告している. 平均で 1000 歩歩数が増加し, 各ユーザの平均寿命が 41.4 日増えると試算している.

6. おわりに

本稿では, 位置情報サービスがユーザの行動に与える影響を分析し, その影響を活用した POI 推薦手法を提案した. チェックイン回数, チェックイン場所, チェックイン間の移動距離などを分析し行動への影響を評価した. 次に位置情報サービスの利用者に対する POI 推薦手法を提案し, Foursquare のチェッ

クインデータを利用して評価を行った. 実験では提案手法が統計的に有意な差があることを確認した.

今後の課題として, チェックイン行動に影響を与える位置情報サービスを自動で特定する方法が考えられる. 本稿では Pokémon Go を対象として検証を行ったが, 他のサービスやアプリへ適用可能か検証をする. また, POI 推薦のスコア計算の部分で, 2.4 節および 2.5 節で分析したチェックインの時間帯や地理的な距離を考慮していないため, 従来研究 [Yuan 13, Ye 11] と同様にそれらの手法と提案手法を統合した POI 推薦を行い評価する.

参考文献

- [Althoff 16] Althoff, T., White, R. W., and Horvitz, E.: Influence of Pokémon Go on physical activity: Study and implications, *Journal of Medical Internet Research*, Vol. 18, No. 12 (2016)
- [Gonzalez 08] Gonzalez, M. C., Hidalgo, C. A., and Barabasi, A.-L.: Understanding individual human mobility patterns, *Nature*, Vol. 453, No. 7196, pp. 779–782 (2008)
- [Ma 09] Ma, H., King, I., and Lyu, M. R.: Learning to Recommend with Social Trust Ensemble, in *Proceedings of the 32nd ACM SIGIR '09*, pp. 203–210 (2009)
- [Su 09] Su, X. and Khoshgoftaar, T. M.: A survey of collaborative filtering techniques, *Advances in artificial intelligence*, Vol. 2009, p. 4 (2009)
- [Ye 11] Ye, M., Yin, P., Lee, W.-C., and Lee, D.-L.: Exploiting Geographical Influence for Collaborative Point-of-interest Recommendation, in *Proceedings of the 34th SIGIR '11*, pp. 325–334 (2011)
- [Yuan 13] Yuan, Q., Cong, G., Ma, Z., Sun, A., and Thalmann, N. M.: Time-aware Point-of-interest Recommendation, in *Proceedings of the 36th SIGIR '13*, pp. 363–372 (2013)