

時間の影響度を考慮した Point of interest 推薦の提案

小島 直己 高木友博

明治大学大学院理工学研究科

〒214-8571 神奈川県川崎市多摩区東三田 1-1-1

E-mail: {n-kojima, takagi}@cs.meiji.ac.jp

概要 Foursquare や Facebook places などのような位置情報を用いたサービスが一般化し、ユーザーに対し興味のあるスポット (Point of interest : POI) を推薦するサービスの必要性が高まっている。本研究では、各時間の POI への訪問確率とユーザーの興味との関係を分析し、その関係の強さによって推薦スコアを調整することで時間やユーザーの興味を反映し、最適なスポットを推薦するための POI 推薦システムを提案する。

キーワード 推薦システム、協調フィルタリング

1 はじめに

Foursquare や Facebook places などのような位置情報を用いたサービスが一般化したことにより、膨大なユーザーの各スポットへの訪問履歴データが利用可能になった。そのようなデータを情報推薦技術に適用し、ユーザーに興味のありそうなスポット (Point of interest : POI) を推薦することは非常に有用である。本研究では特にユーザーの POI への訪問確率が時間による影響を受けているということを想定し、ユーザーにとって興味があり、かつユーザーにとって時間的に最適な POI を推薦することを目的としている。

2 提案手法

2.1 User based collaborative filtering

提案手法は推薦システムの中でも代表的な手法である User based collaborative filtering (以降, "UBCF") をベースとしている。UBCF で推薦する場合、まず POI 訪問履歴から推薦を行うユーザー (以降, "対象ユーザー") と他のユーザーとの類似度を計算する。次に、他のユーザーの類似度と POI 訪問履歴との重み付けを行い、POI へスコア付けをする。そして、スコアが高い POI 順に対象ユーザーへ推薦を行う。具体的な計算式は以下の通りである。

$$w_{u,v} = \frac{\sum_l c_{u,l} c_{v,l}}{\sqrt{\sum_l c_{u,l}^2} \sqrt{\sum_l c_{v,l}^2}} \quad (1)$$

$$\hat{c}_{u,l} = \frac{\sum_v w_{u,v} c_{v,l}}{\sum_v w_{u,v}} \quad (2)$$

$c_{u,l}$ はユーザー u が POI l に訪問した場合は 1、していない場合は 0 をとる変数である。 $w_{u,v}$ はユーザー u, v 間の類似度であり、今回はコサイン類似度を用いた。 $\hat{c}_{u,l}$

はユーザー u の POI l に対する予測スコアである。

2.2 時間を考慮した UBCF

本研究ではユーザーが POI に訪問した時間も考慮したユーザー間類似度を計算し、またユーザーが現在おかれている時間に適した推薦を行うために、UBCF の計算式を以下のように拡張した。

$$w_{u,v} = \frac{\sum_{t=1}^T \sum_{l=1}^L c_{u,t,l} c_{v,t,l}}{\sqrt{\sum_{t=1}^T \sum_{l=1}^L c_{u,t,l}^2} \sqrt{\sum_{t=1}^T \sum_{l=1}^L c_{v,t,l}^2}} \quad (4)$$

$$\hat{c}_{u,t,l} = \frac{\sum_v w_{u,v} c_{v,t,l}}{\sum_v w_{u,v}} \quad (5)$$

$c_{u,t,l}$ はユーザー u が時間 t のときに POI l に訪問したかを表す変数で $\hat{c}_{u,t,l}$ はその予測スコアである。

2.3 スムージング

2.1 のような時間を考慮しない通常の UBCF の場合、使用する全ユーザー POI 訪問履歴は User×POI の行列で表される。一方の 2.2 のように時間を考慮した場合、User×POI×Time の 3 階テンソルで表され、2.1 のときと比べてデータが疎になりやすくなる。UBCF では一般的に使用するデータが疎である場合、ユーザー間類似度が正確に測れない、推薦可能な POI が少なくなるなどといった問題点がある。

そこで提案手法では訪問履歴データを正規分布に基づいたスムージングによって補間し、データをある程度密にしてから 2.2 の手法を用いて推薦を行う。

正規分布は以下の式で表されるように平均値 μ と分散値 σ^2 をパラメータにもち、平均値 μ の付近に集積するようなデータの分布を表した確率分布である。

$$N(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad (6)$$

μ のとき最大値をとり、 σ^2 の変化によって分布の形が変化する。この正規分布の特性を活かして、ユーザー u

の POI に対する各時間の訪問履歴を実際に訪れた時間 t を中心にするように正規分布の形でスムージングする。

また、分散値 σ^2 を POI ごとに変化させることで各 POI の時間の影響度に適したスムージングを行うことが出来る。例えば昼食、夕食時など特定の時間のみユーザーの訪問が集中するレストランのような POI に関しては分散値を低く設定することで訪問が集中している時間とその周辺時間のみをスムージングすることができる。一方、どの時間帯にも平均してユーザーが訪問する空港のような POI に関しては分散値を高く設定することでどの時間帯にも平均的にスムージングできる。

提案手法では POI の各時間の訪問確率分布から POI の時間の影響度を算出し、その値をもとに分散値を設定する。具体的には POI における時間の影響度 s_l を各時間での POI への訪問確率の分布 $P_l(t)$ とどの時間帯にも平均的に訪問したときの一様分布 $P_u(t)$ とのジェンセン・シャノン情報量(JSD)を $[0, 1]$ で正規化した値とする(式(7))。このとき s_l は特定の時間に訪問が集中するような時間の影響度の高い POI に関して 0 に近い値となり、どの時間にも平均的に訪問されるような時間の影響度の低い POI には 1 に近い値となる。したがって、 s_l を正規分布の分散値 σ^2 の値とすることで POI ごとに時間の影響度に適したスムージングを行うことができる。

以上をまとめると、正規分布によるスムージングは以下の式で表される。

$$s_l = e^{-JSD(P_l(t), P_u(t))} \quad (7)$$

$$\tilde{c}_{u,t,l} = \max_{t' \in T_{u,l}} (N(\mu = t', \sigma^2 = s_l)) \quad (8)$$

3 実験

3.1 データセット

実験では、シンガポールで 2010/08 ~ 2011/07 にかけて収集されたユーザーの POI の訪問履歴をデータセットとして用いた。データの規模は以下の通りである。

表 1 データセットのサイズ

総ログ数	ユーザー数	POI 数
194,108	2,321	5,596

データセットの 25% を正解データ、残りを訓練データとし、4 セットの交差検定を行った。

3.2 評価指標

実験では Precision, Recall による評価を行った。本システムでは各ユーザーに対して、各時間において推薦するため以下の式で評価値を求めた。

$$precision(t) = \frac{\sum_{u \in U} P_{u,t} \cap C_{u,t}}{\sum_{u \in U} P_{u,t}} \quad (9)$$

$$recall(t) = \frac{\sum_{u \in U} P_{u,t} \cap C_{u,t}}{\sum_{u \in U} C_{u,t}} \quad (10)$$

$$Precision = \frac{1}{T} \sum_{t \in T} precision(t) \quad (11)$$

$$Recall = \frac{1}{T} \sum_{t \in T} recall(t) \quad (12)$$

$P_{u,t}$ は時間 t のときにユーザー u に推薦された POI 集合、 $C_{u,t}$ は正解データの中で時間 t のときにユーザー u が訪問した POI 集合である。各時間での評価値を計算し(式(9),(10))、それらの平均値をそのシステムの Precision, Recall の値とする(式(11), (12))。

3.3 比較手法

実験ではスムージングを行わずに時間分割して UBCF を行った手法(以降、“ベース手法”), 時間同士の類似度からスムージングを行う既存手法[1]との精度比較を行った。

3.4 結果

ユーザーに推薦する POI 数を 5 件, 10 件, 20 件で Precision, Recall 値を比較したときの結果が以下の通りである。実験結果から提案手法が比較手法2つに比べ Precision, Recall ともに上回っていることがわかる。

表 2 Precision 比較

	ベース手法	既存手法	提案手法
5	0.0268	0.0329	0.0406
10	0.0186	0.0240	0.0286
20	0.0126	0.0168	0.0195

表 3 Recall 比較

	ベース手法	既存手法	提案手法
5	0.0825	0.0992	0.1259
10	0.1137	0.1439	0.1752
20	0.1520	0.2010	0.2363

4 おわりに

本稿では、UBCF に POI への訪問時間とスムージングの要素を入れ、ユーザーに対し興味のある、かつ時間に最適な POI を推薦する手法を提案し、その有用性を示した。今後、POI のカテゴリー情報を用いて、各時間でどのようなカテゴリーの POI が推薦されているかの定性的な評価を行う予定である。

参考文献

- [1] Yuan, Quan, et al. "Time-aware Point-of-interest Recommendation." Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval. ACM, 2013.