

SNS 情報の協調フィルタリングに基づき 利用者の嗜好に適応する旅行計画支援システム

2018SE104 柳太助

指導教員：沢田篤史

1 はじめに

Web 上には利用者の嗜好に適応した旅行計画を作成できるツールがいくつか存在する。より良い旅行計画を立てるためには多様な状況に適応し、増減する魅力を反映できる旅行計画作成ツールが必要である。

これらのツールには、季節や時間、天候など観光地そのものの以外の環境（環境要因）が考慮されないという問題がある。多くの旅行計画推奨システムでは利用者の観光スタイルや観光地の嗜好といった人的要因に基づいて推奨が行われるので、環境要因は推奨に利用されにくい。

本研究の目的は、環境要因を加味した旅行計画を推奨できるシステムを実現することである。すなわち、特定の環境要因下における観光地の魅力の増減関係に着目し、それを推奨に利用する方法を確立することである。

訪問経験のない目的地に対する予測評価値を求めて推奨に利用するために、キーワードから算出した値（重み）を計算に含めた協調フィルタリングを用いる。得られた予測評価値の高い観光地を利用者に推奨する。環境要因を加味する仕組みとして、特定のキーワードを含む Twitter^{*1}のツイート（投稿）を取得し、投稿の内容に応じて処理をすることで推奨に活用する。

2 旅行計画推奨システムとその課題

旅行計画の推奨には天候や季節といった環境要因が加味されないという課題がある。観光地そのものや経路、交通手段以外の情報を旅行計画の推奨に利用できるツールは少ない。推奨システムで出力する旅行計画の多くは複数の目的地と目的地間の移動にスポットが当てられており、実際の旅行でどのように過ごすかが考慮されないで環境要因を加味した推奨が不可能である。

3 環境要因を加味した旅行計画推奨システム

3.1 システムの設計方針

本研究では、推奨システムとしてアイテムベースの協調フィルタリングを採用する。本研究で提案する手法では、利用者の評価と設定した環境の条件次第で推奨する観光地が変わるので、誰が利用しても似たような観光地が推奨されやすいという欠点を補うことができる。

観光地と環境要因に関するデータは SNS から取得する。中でも、本システムでは Twitter からデータを集める。利用者が多く、様々な観光地に関する投稿が高頻度でされ、データの取得が容易なので実験に活用するためのデータ

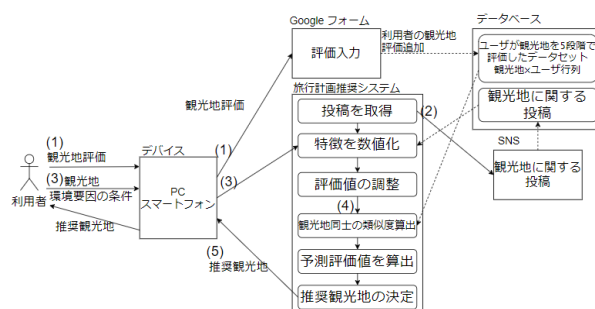


図1 システムの概要図

セットの作成に適している。

3.2 システムの概要

本システムの概要図を図1に示す。矢印に付与した番号は、以下に示す本システムの手法の番号を示している。また、破線の矢印は情報の移動を意味している。旅行計画推奨システムの中の類似度の算出、予測評価値の算出、推奨観光地の決定が協調フィルタリングに当たる。

本システムの推奨処理の手順は次の通りである。

- (1) 利用者は観光地を5段階で評価し、データセットに評価値を追加する
- (2) Twitter から投稿を取得し、データセットに追加する
- (3) 利用者は観光地名と環境要因を入力し、入力に応じて数値化した特徴を用いて評価値を調整する
- (4) 調整した評価値を基に類似度を算出し、予測評価値を算出する
- (5) 予測評価値を基に推奨観光地を出力する

3.3 観光地データセットの作成

愛知県観光レクリエーション利用者統計^{*2}を参考に愛知県内の観光地50か所を5段階で評価したデータセットを作成した。Google フォーム^{*3}を利用して各観光地を1から4（数字が大きい方が好き）および-1（訪問経験がない）の5段階で評価をするアンケートを作成し、大学生38人に協力してもらい、観光地と評価のデータセットを作成した。

3.4 アイテムベース型協調フィルタリングのアルゴリズム

アイテムベース型協調フィルタリングにおける観光地同士の類似度とは、好まれ方がどの程度似ているかという数値である。観光地同士の類似度は文献 [1] を参考に以下の

^{*2} <https://www.pref.aichi.jp/kanko/menu/toukei/recreation/2020sankou.pdf>

^{*3} https://www.google.com/intl/ja_jp/forms/about/

^{*1} <https://twitter.com>

表 1 ユーザ 24 の未訪問観光地に対する予測評価

観光地名	予測評価値
1. 豊川稲荷	3.233
2. 国営木曾三川公園 1 3 8 タワーパーク	2.848
3. 堀内公園	2.636
4. 道の駅藤川宿	2.602
5. 遍照院	2.583

表 2 環境要因を加味した場合のユーザ 24 の未訪問観光地に対する予測評価 (1)

観光地名	予測評価値
1. 豊川稲荷	2.953
2. 国営木曾三川公園 1 3 8 タワーパーク	2.883
3. 道の駅藤川宿	2.737
4. 堀内公園	2.640
5. デンパーク	2.583

式を用いて算出する。 $RB_{i,j}$ はアイテム i とアイテム j の両方を評価したユーザのセットを示している。 r_{ui} はユーザ u のアイテム i への評価、 \bar{r}_u はユーザ u の平均評価である。

$$itemSim(i, j) = \frac{\sum_{u \in RB_{i,j}} (r_{ui} - \bar{r}_u)(r_{uj} - \bar{r}_u)}{\sqrt{\sum_{u \in RB_{i,j}} (r_{ui} - \bar{r}_u)^2} \sqrt{\sum_{u \in RB_{i,j}} (r_{uj} - \bar{r}_u)^2}}$$

環境の条件による魅力の増減関係を判別するために、観光地名と環境要因を文中に含む Twitter のツイートを取得する。取得したテキストを感情分析し、テキスト全体に対する極性の割合を重みとする。算出されたポジティブ度・ネガティブ度に応じて、評価値を調整する。感情判定用の辞書として日本語評価極性辞書 [2] を利用する。

訪問経験のない観光地に対する予測評価値は文献 [1] を参考に以下の式を用いて算出する。 $ratedItems(u)$ は評価されたアイテムの全体を示している。 r_{ui} はユーザ u のアイテム i への評価である。

$$pred(u, i) = \frac{\sum_{j \in ratedItems(u)} itemSim(i, j) \cdot r_{uj}}{\sum_{j \in ratedItems(u)} itemSim(i, j)}$$

4 プロトタイプシステムを用いた評価

4.1 システムの実装

予測評価値の算出には類似度の上位 5 つの観光地の評価値を用いる。評価値の調整は以下に示す 2 種類の手法で行う。

- (1) ポジティブ度とネガティブ度の大小関係に応じて評価値を +1 または -1 する手法
- (2) 評価値を + (ポジティブ度 - ネガティブ度) * 定数とする手法

4.2 実験

環境要因を加味していない通常の協調フィルタリングでの推奨結果を示す。予測評価値をまとめた表を表 1 に示す。

表 3 環境要因を加味した場合のユーザ 24 の未訪問観光地に対する予測評価 (2)

観光地名	予測評価値
1. 豊川稲荷	3.146
2. 国営木曾三川公園 1 3 8 タワーパーク	2.884
3. 道の駅藤川宿	2.740
4. 堀内公園	2.641
5. デンパーク	2.577

環境要因を加味しない推奨では、ユーザ 24 に対する豊川稲荷の予測評価値は 3.233 と算出された。ユーザ 24 が評価した 28 か所の観光地における評価の平均値は 2.679 であり、平均値と比較すると少し高い評価であった。

手法 (1) で算出される予測評価値をまとめた表を表 2 に示す。(1) の手法で計算を行った結果、豊川稲荷の予測評価値は 2.953 と出力された。環境要因を加味していない場合と比べ、予測評価値は 0.28 減少した。

手法 (2) で算出される予測評価値をまとめた表を表 3 に示す。(2) の手法で計算を行った結果、豊川稲荷の予測評価値は 3.146 と出力された。環境要因を加味していない場合と比べ、予測評価値は 0.087 減少した。

5 考察

本研究で提案するシステムでは、特定の条件のもと魅力が変動するという仕組みを観光地の評価に反映させて推奨を行った。その結果、指定した条件に合わせて異なる推奨観光地の出力をすることができた。

実験の結果、豊川稲荷は朝という状況では魅力が増大するにもかかわらず、手法 (1) (2) とともに予測評価値が減少した。予測評価値の算出には類似度が上位 5 か所の観光地の評価値を利用したので、ユーザが低く評価した観光地が偶然類似度の高い観光地であった場合予測評価値は減少する。

6 おわりに

SNS 情報と利用者の嗜好情報を協調フィルタリングするシステムを構築することで、環境要因を加味した場合の推奨観光地の出力が可能となる。本研究で実装したプロトタイプを用いた利用者実験では、環境要因に適応した観光地を推奨できた点で一定の成果はあったものの、実用的な旅行計画の推奨という観点からは必ずしも十分ではない結果となった。

参考文献

- [1] J. Ben Schafer, et al., “Collaborative Filtering Recommender Systems”, The Adaptive Web, pp. 291-324, 2007.
- [2] H. Masahiko, et al., “Learning Sentiment of Nouns from Selectional Preferences of Verbs and Adjectives”, Proceedings of the 14th Annual Meeting of the Association for Natural Language Processing, pp. 584-587, 2008.