

口コミを用いた嗜好抽出に基づくホテル推薦手法

田熊 浩二[†] 福原 楓 亀井 清華[†] 藤田 聡[†]

広島大学大学院工学研究科[†]

1. はじめに

宿泊するホテルを探す際にはホテル予約サイトを利用することが一般的である。ホテル予約サイトにある主な情報はホテルのホームページ、口コミ、口コミ投稿者による評価であり、ユーザはこれらの情報を参考にしてホテルを決める。ホテル予約サイトでは、ユーザが入力した金額や駐車場の有無等の条件を満たしているホテルの中から、総合評価の高い順にホテルのリストが表示されることが多い。ホテルの総合評価は実際にホテルに宿泊した人が投稿した評価を平均化したものであり、そこに各投稿者の好みは反映されていない。そのため、ある評価項目の評価が高いからといって、その項目にこだわりを持つユーザに合ったホテルであるとは限らない。

しかし、多くのユーザは一定以上の評価がされているホテルに泊まろうとするため、上位に表示されるホテルから選択する可能性が高い。そのため、表示順が上位のホテルほど口コミ投稿数も多くなる。すると、ユーザがどのホテルに泊まるか悩んだとき、本来ユーザの好みに合ったホテルを選択するための重要な情報である、自身の価値観に似た口コミや投稿者の評価を見つけることが、より困難になる。

そこで、本稿では投稿者による平均評価が高い順にホテルを推薦するのではなく、口コミから投稿者の嗜好を自動で抽出し、ユーザの嗜好と一致する投稿者の評価を重視した表示順にすることで、推薦精度の向上を図った。

2. 関連研究

ユーザに商品を推薦する際に、ユーザについての何かしらの情報を用いてユーザに合ったものを上位に表示する方法はいくつか存在する。例えば、佐賀ら[1]は、ユーザの購買履歴を用いてホテル間の関係を数値化することで、嗜好ネットワークを作成し、そのネットワークに照らし合わせることでユーザの嗜好に合ったホテルを推薦した。しかしこの手法では、ホテルを推薦する際にホテルとは関係ない一般的な商品の購買履歴が大量に必要となる。また、服部ら[2]は、ユーザの価値観に基づく推薦手法を提案した。具体的には、口コミ文から係り受け解析を行うことで、評価属性(映画だとストーリーなど)と評価語(すばらしいなど)の組を抽出し、その組に好評か不評かの極性をつけ、その結果とアイテムの評価値(星の数など)を組み合わせることで、ユーザにとって重要な項目とそうでない項目を計算する。しかしこの手法では、口コミ文によっては評価属性と評価語を抽出できないことや意味が真逆になってしまうという問題点がある。

そこで本研究では、口コミ文と評価といったホテル予約サイトに存在する情報のみを用いることで、ユーザの価値観に基づく推薦を行う手法を提案する。

3. 提案手法

本研究では、ホテルを評価する6つの評価項目(サービス、立地、部屋、設備・アメニティ、風呂、食事)を重視する順にランキングしたものを**嗜好**と呼ぶことにする。また、評価項目を以下では**カテゴリ**と呼ぶ。ここではホテル予約サイトにおけるホテル評価を各ユーザの嗜好にパーソナライズする手法を提案する。この手法は、口コミから口コミ投稿者の嗜好を抽出し、各ユーザと同じ嗜好を持つ投稿者が高く評価したホテルを上位に推薦するように評価値自体を変更する。図1に本システムの構成を示す。

投稿者の嗜好を抽出するにあたり、各投稿者は自身が口コミで言及している単語に関するカテゴリを重要視していると仮定する。そのために、どういった単語がどのカテゴリに関するものなのかの指標が必要となる。そこで、口コミコーパスから名詞を抽出し、重複ありで6つのカテゴリに手動で分類した。これを**嗜好辞書**と呼ぶ。

提案手法では、嗜好辞書を元に、各投稿者の全ての口コミから名詞を抽出し、各口コミ文を6つのカテゴリに分類する。各投稿者について、カテゴリ毎の名詞の数をカウントし、数が多い順にカテゴリを順位付けし、その投稿者の嗜好とした。

ユーザには、6つの全カテゴリを順位付けし、嗜好を明示してもらう。そしてユーザの嗜好と投稿者の嗜好とのマッチングを行い、同じ嗜好を持つ投稿者のホテルに対する評価を50倍にし、嗜好が異なる投稿者の評価はそのままとすることで重みづけを行う。次に、一般的なホテル評価(0~5)と合わせるために、カテゴリごとに嗜好が一致した人数×50人分評価が増えたという扱いで、評価を人数で割り、平均評価を計算する。そして嗜好の*i*番目に重視するカテゴリの平均評価を(7-i)倍し、すべてのカテゴリの評価を足したものを**嗜好スコア**と呼ぶことにする。この嗜好スコアは一般的なホテル評価の総合評価にあたるものである。嗜好スコアの値域は0~105となる。

提案手法では、さらに、口コミの表示順についてもパーソナライズを行った。最新のものから時系列順に表示するのではなく、投稿者の嗜好とユーザの嗜好の類似度が高い物から順に表示するようにする。つまり、まず嗜好が同じものの中で新しいものから順に表示し、次に1つの順位が入れ替わっているようなものの中で新しいものから順に表示する。これにより、ユーザの求める情報に関する口コミが見つけやすくなると考えた。

4. 実験

提案手法を用いた推薦システムのプロトタイプを構築し、アンケートによる実験を行った。推薦対象ホテルは楽天トラベル[3]の広島市内にあるホテル・旅館のうち、口コミ数が100件以上のホテル72件とした。口コミは楽天トラベルに2009年10月から2012年6月の間に投稿さ

Hotel recommendation method based on preference extraction using reviews

[†] Graduate School of Engineering, Hiroshima University

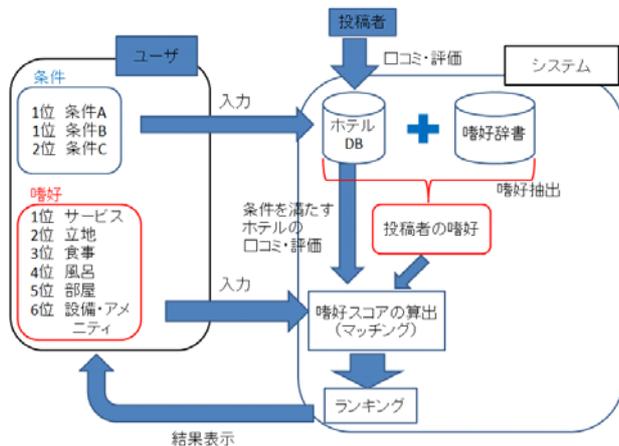


図1 システム構成

れたもののうち、対象ホテルに投稿した 12386 人が全国のホテルに投稿した 67465 件を対象とした。その内、対象ホテルへの口コミ数は 21742 件である。対象ホテルに投稿した投稿者一人当たりの平均口コミ数は 5.45 で、対象ホテル当たりの平均口コミ数は 301.97 である。

実験は被験者 30 人に対するアンケート方式で行い、提案手法、条件のみの各手法で推薦された各上位 30 件のホテルの中から、各手法で泊まりたいホテル上位 3 件を回答してもらった。この時、各被験者には、休日の旅行を想定した場合と出張・学会などを想定した場合のそれぞれについて回答してもらった。ただし、被験者はホテルを選択する際に必ずレビューを読んでから選択してもらった。このとき、被験者は推薦されたホテルの順位はわかるが、どの手法で推薦されているのかはわからない状態で選択してもらった。さらに各手法について、ホテルを選択する際にいくつのお口コミを読んだかも回答してもらった。

被験者からの回答を以下の手法で評価した。各推薦手法で推薦されるホテルの順位が 1 位なら 30 ポイント、2 位なら 29 ポイント、3 位なら 28 ポイントというようにホテルにポイントをつける。そして 1 番泊まりたいホテルに選択されたホテルはポイントを 3 倍し、2 番目に泊まりたいホテルに選択されたホテルはポイントを 2 倍、3 番目に泊まりたいホテルに選択されたホテルはそのままのポイントで、その合計ポイントを各手法について計算し、比較評価を行った。

4.1 結果

表 1 に、全被験者の各手法の合計ポイントを示す。提案手法が条件のみの手法よりも約 180 ポイント多くなった。よって、ユーザがホテルを選択するときには、少なからず嗜好が影響していると考えられる。

表 2 に、想定した場合別のポイントを示す。休日等の場合では嗜好の影響が条件より大きく、学会等の場合では少し嗜好の影響が小さくなり、条件の影響が大きくなると考えられる。

表 3 に、全被験者によって選ばれたホテルの推薦順位の合計を示す。提案手法の方が、上位のホテルが選ばれやすかったことがわかる。よって、表示順位が下のホテルを選択するときにも嗜好が影響していることがわかる。

表 4 に、読まれた口コミ数の平均を示す。口コミに関しては表示順が上の十数件のみを参考にする傾向にある

表 1 選択されたホテルの合計ポイント

提案手法	条件のみ
9400	9221

表 2 選択されたホテルの場合別ポイント

休日・旅行の場合		学会・出張の場合	
提案手法	条件のみ	提案手法	条件のみ
4578	4348	4519	4558

表 3 順位の合計

提案手法	条件のみ
972	1069

表 4 平均既読口コミ数

提案手法	条件のみ
15.4	12.3

ことがわかる。読まれた平均口コミ数は少し提案手法が多くなったが、これは今回の実験時には、なぜその口コミを上位に示したかという根拠となる、投稿者との類似度を被験者に示していないことが一つの原因であると考えられる。その口コミの投稿者の嗜好がユーザの嗜好に似ていたとしても、その口コミ自体にその嗜好が十分に表れるかどうかは問うていないからである。

5. おわりに

本研究では、ユーザの条件だけでなく、嗜好をホテル推薦に反映させる手法を提案した。提案手法では、口コミから投稿者の嗜好を抽出し、ユーザの嗜好と一致した投稿者の評価を相対的に重視することで、ユーザの嗜好を反映した。また、ユーザと似ている嗜好の投稿者の口コミを上位に表示することでユーザの価値観に近い口コミを上位に表示した。

アンケートによる評価では、従来の手法に近い条件のみを反映した手法と比較して、提案手法はユーザの嗜好に合ったホテル推薦を行うことができていることを確認した。

口コミ表示順にも嗜好を反映させることでユーザに有益な口コミを上位に表示させようとしたが、実験では、読まれる口コミ数に関しては嗜好の影響があまりないことがわかった。

今後の課題としては、投稿者の嗜好をユーザに見せることによる効果について評価することを考えている。

謝辞

本研究では、楽天株式会社が提供し、国立情報学研究所が配布している「楽天公開データ」を利用させていただきました。ここに記して謝意を表します。

参考文献

- 1) 佐賀 亮介, 林 義浩, 辻 洋. 暗黙的評価を用いた嗜好推移に基づくホテル推薦手法, 情報処理学会全国大会講演論文集, 第 70 回平成 20 年(1):1-421-1-422, 2008-03-13.
- 2) 服部 俊一, 高間 康史, 価値観を利用した情報推薦についての検討, 人工知能学会情報編纂研究会第 4 回研究会, 2011
- 3) 楽天トラベル <http://travel.rakuten.co.jp/>
- 4) 飯島 正樹. 口コミの効果に関する研究, 日本経営工学会誌, 45(6):575-580, 1995-02-15.