6R-2

クチコミから抽出した特徴語を利用する観光地検索支援

松本　敦志†　　杉本　徹‡

芝浦工業大学大学院 理工学研究科†

芝浦工業大学　工学部‡

1. はじめに

近年，旅行の計画を立てる際に観光地の情報を Web を用いて調べる人が増えており，ユーザが Web 上に自由に書き込んだクチコミ情報を参考にすることも多い．しかし，Web 上には数多くのクチコミが公開されており，それらに目を通すことはユーザにとって大きな手間となる．

本研究では観光地の特徴を表す言葉(特徴語)をクチコミから抽出し，ユーザに提示することにより観光地検索の支援を行う手法を提案する．ユーザは候補地を検索する際に特徴語を見ることにより数多くのクチコミを読む手間が減ることが期待できる．また，特徴語を提示することは候補地を絞り込む際の発想の支援にもなる．廣嶋ら[1]は地理情報検索の際のクエリ入力支援として特徴語の提示を提案し，提示する特徴語の抽出手法について研究を行った．廣嶋らは特徴語の候補を Wikipedia の見出し語に限定し，特徴語抽出の手法としてはポアソン確率を用いた．それに対し，本研究では抽出対象を任意の名詞として，4 種類の手法を用いて特徴語抽出を行い，どの手法が適しているのか検討を行った．また，抽出した特徴語を利用した検索支援システムを試作し，実験を通して特徴語提示の効果を検証する．

2. 特徴語の抽出

BIGLOBE 旅行(http://travel. biglobe. ne. jp/)にクチコミが存在した 3,751 か所の観光地についてそれぞれの観光地のクチコミを最大 200 件まとめたデータを 1 つの文書として，文書内に書かれている単語を解析してその観光地の特徴語を抽出する．本研究では単語の出現頻度に着目してそれぞれの単語がどれだけ特徴的かを表す特徴量の算出を行い，特徴量から特徴語の判断を行う．特徴量の算出には 4 種類の手法を用いて実験を行い，結果について検討を行う．

以降の説明において，総文書数を n，文書 i(1≦i≦n)における単語 w の出現回数を$tf_i(w)$，全文書における単語 w の出現回数を$tf(w) = \sum_{i=1}^{n} tf_i(w)$，w が出現する文書数を df(w)，文書 i の総単語数をm_i，全文書の総単語数を $M = \sum_{i=1}^{n} m_i$，文書 i の長さで正規化した単語の出現回数を$TF_i(w) = tf_i(w)/m_i$とする．

2. 1. TF-IDF

以下のように特徴量を定義する．

$$\text{単語 w の特徴量} = TF_i(w) * \log\frac{n}{df(w)}$$

2. 2. ATF(Average Term Frequency)

$TF_1(w)$から$TF_n(w)$までの合計を n で割った値を atf(w)とするとき,特徴量を以下のように定義する．

$$\text{単語 w の特徴量} = \frac{TFi(w)}{(atf(w))^{0.3}}$$

2. 3. ポアソン確率

廣嶋らが提案した算出手法を本研究に適用し特徴量を算出する．

$$\text{単語 w の特徴量} = \sum_{x=0}^{tf_i(w)-1} \frac{e^{-\lambda}\lambda^x}{x!}$$

ここでλ= $tf(w) * m_i/M$であり，文書 i における単語 w の出現回数の期待値を表す．この特徴量は文書 i における単語 w の出現回数が$tf_i(w)$未満である確率を表す．

2. 4. エントロピー

単語 w のエントロピー

$$H(w) = -\sum_{i=1}^{n} \frac{tf_i(w)}{tf(w)} \log\frac{tf_i(w)}{tf(w)}$$

を用いて特徴量を以下のように定義する．

$$\text{単語 w の特徴量} = TF_i(w) * \left(1 - \frac{H(w)}{\log n}\right)$$

2. 5. 抽出手法の比較実験

どの手法が妥当であるか比較実験を行うために清水寺と鳥取砂丘の文書に含まれる単語に対して各手法で特徴量を算出し，上位 30 語をそれぞれ特徴語として抽出した．実験の方法は 5 名の被験者に各手法で抽出した特徴語がその観光地の特徴になっていると思うかどうかを「思う，どちらとも言えない，思わない」の 3 段階で評価してもらった．この際に判断材料として

A Support Method for Sightseeing Spot Retrieval Using Feature Words Extracted from Word of Mouth

†Atsushi MATSUMOTO and ‡Toru SUGIMOTO

†Graduate School of Engineering and Science, Shibaura Institute of Technology

‡College of Engineering, Shibaura Institute of Technology

BIGLOBE 旅行に存在するクチコミを提示した．各被験者が回答した評価に 1 点，0 点，-1 点の得点を与えて，それぞれの特徴語に対して 5 名の被験者がつけた得点を合計した値を算出した．この値に特徴量の順位を考慮した値(1 位は 30，2 位は 29，3 位は 28，…)を乗算し，その特徴語の得点とした．各手法で抽出した特徴語の得点の合計を算出し，その合計値をその手法の得点として比較をした．清水寺に関して，最も高い得点はポアソン確率で 1,109，2 番目に高い得点は tf-idf で 1,070 であった．また鳥取砂丘に関して，最も高い得点はポアソン確率で 318，2 番目はエントロピーで 122 であった．比較の実験の結果，どちらの観光地もポアソン確率の得点が最も高い値となった．この結果から本研究では特徴語の抽出手法としてポアソン確率を用いる．例として表 2 にポアソン確率を用いて「清水寺」の特徴語を抽出した結果の上位 10 語を示す．

表 2:「清水寺」について抽出した特徴語の例

清水の舞台/京都市内/京都の紅葉スポット/ライトアップ/京都/音羽の滝/地主神社/紅葉の色/京都の夜景/京都タワー etc

3. 特徴語を利用した観光地検索

抽出した特徴語を 2 種類の方法で利用することにより観光地検索の支援を行う．

1 つ目の方法は絞込み検索の結果として表示される観光地名とともにその特徴語を提示することで、観光地のクチコミを読まなくても候補地の判断を行えることが期待できる．

2 つ目の方法は頻出する特徴語の提示である．観光地検索の際にカテゴリなどの絞込みの条件をユーザが指定した場合にその条件に合致する観光地で頻出する特徴語を提示する．この提示により新たな絞込み条件の追加など観光地を検索する上での発想の支援が期待できる．

これらの機能を取り入れた観光地検索システムの試作を行った．本システムでは一般的なクチコミサイトと同等の検索機能を再現するために，「カテゴリ」と「都道府県」による絞込みを行う機能を実装した．その上で観光地名とともにその特徴語を提示する機能と頻出する特徴語を提示する機能を追加した．図 1 にシステムの特徴語提示機能のスクリーンショットを示す．

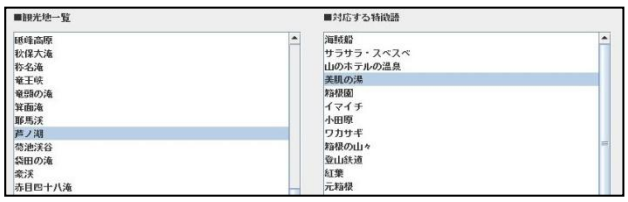

図 1：観光地に対する特徴語の提示

図 1 の左側のテキストボックスには検索結果として得られた観光地名が一覧表示されており，その観光地を選択することで右側のテキストボックスにその観光地の特徴語が特徴量の大きい順に提示されるようになっている．

4. 評価

観光地選択の際に特徴語が有用な情報になるかということを評価する．そのために 2 か所の観光地について特徴語を見てどちらに行きたいかを選択した場合とクチコミを読んだ上で選択をした場合で選択結果がどの程度一致するかということを調べた．具体的には 2 か所の観光地について，①観光地名，②特徴語，③クチコミという 3 つの情報を順に提示していきそれぞれの段階でどちらの観光地に行きたいと思うかを被験者に判断してもらい，その判断理由について自由に記述してもらった．観光地のペアを 5 組用意し，それぞれのペアに対する判断を 10 名の被験者に行ってもらった．

実験の結果，観光地名(①)のみでの判断とクチコミ(③)を読んだ上での判断の一致率が 58%だったのに対して，特徴語(②)を見ての判断とクチコミを読んだ上での判断の一致率は 76%であった．このことから，観光地の特徴語は観光地選択において有用な情報であると考えられる．一方，特徴語を見た後にクチコミを読むことで判断が変わったケースも 24%あった．この 24%に関する判断理由の記述を見ると「特徴語が思っていたことと違ったから」，「特徴語として提示されていなかった情報に興味を持ったから」などの記述が見られた．これは特徴語の抽出が単語単位であり文脈が考慮されていないことや特徴語抽出の際の抽出漏れに原因があると考えられる．

5. おわりに

本研究では観光地に関するクチコミから特徴的な言葉を抽出し，その言葉を利用することで観光地検索の支援を行うことを目標とし，特徴語の抽出と利用方法の提案，評価を行った．結果，観光地検索システムにおいて観光地の特徴語を提示することによりクチコミに代わる判断材料となる情報を提供できることが分かった．

参考文献

[1] 廣嶋伸章，安田宜仁，藤田尚樹，片岡良治，“地理情報検索におけるクエリ入力支援のための特徴語の提示”，第 26 回人工知能学会全国大会，2012