1L-03

多様化した旅行者のニーズを考慮した検索タグの決定手法の提案

宮井　和輝 †　奥野　拓 †
公立はこだて未来大学 †

1　はじめに

近年，旅行者のニーズが多様化している．しかし，従来の観光情報サイトには膨大な量の観光情報があるため，そこから旅行者が求める観光スポットを探すことは容易ではない．

従来の観光情報サイトで観光スポットを検索する方法の１つとしてタグ検索がある．しかし，一般的に人の手によるタグ付けが行われているため，タグの数に限度があり，旅行者は多様化した旅行者のニーズに応じた検索をすることが困難である．

そこで本研究では，観光スポットの説明文から名詞を抽出し，多様化した旅行者のニーズを考慮したタグを自動的に決定することを目的とする．

2　従来の観光情報サイトにおける検索方法の問題点

従来の観光情報サイトにおける観光スポットの説明文には，観光スポットの情報が詳細に説明されており，旅行者のニーズを満たすような質の高い情報が記載されている．しかし，観光情報が階層的に分類されているため，説明文を読むには，大量のリンクを辿る必要があるという問題がある．

検索する負担を軽減する方法として「幼児向け」のような検索タグ（以下，タグとする）を選択して検索する方法がある．この方法では，目的に応じたタグを選択することで，ユーザが求める情報を容易に検索することができる．しかし，一般的に人の手によるタグ付けが行われているため，網羅的にタグ付けを行うには限界がある．

3　関連研究

観光情報から有用な情報を抽出する研究として，松本らは観光情報サイトのクチコミから観光スポットの特徴を表す言葉を抽出する手法を提案している [1]．この研究では観光スポットの情報を抽出するが,「映画に関する観光スポットを探したい」のような多様化した旅行者のニーズに応じた検索が困難である．

そこで本研究では，旅行者のニーズを満たすような情報が記載されている観光スポットの説明文から多様化した旅行者のニーズを満たす情報を抽出する．

4　検索タグの決定手法

本研究では，従来の観光情報サイトに記載されているすべての観光スポットの説明文から名詞を抽出し，これらの中で旅行者のニーズに該当する名詞をタグとする．旅行者のニーズは，旅行業者および団体による観光ニーズに関する調査の結果を基に３種類のニーズを定義する [2][3]．旅行者のニーズとタグの例を表１に示す．

名詞を抽出する際，観光情報サイトには表記ゆれや「日本最古」,「日本最初」のような類似した意味の名詞が多数存在するため，類似した意味の名詞が複数抽出されてしまう．そこで，階層クラスタ分析を行い，意味的に関連のある名詞をまとめ，定義した旅行者のニーズに該当する名詞を抽出する．

以下のような手順で定義したニーズに該当する名詞を抽出し，タグを決定する．

1. すべての観光スポットの説明文に対して，形態素解析を行い，複合名詞を抽出する．
2. 抽出した名詞をベクトル化する．
3. ベクトル間の類似度を算出する．
4. 算出した類似度を基に階層クラスタ分析により樹形図を作成する．
5. それぞれのクラスタに対して，クラスタの特徴を表すラベル（名詞）を付与する．
6. 付与したラベルから定義したニーズに該当するものを抽出し，それをタグとする．

A Proposal of a Decision Method of Search Tags Considering Diversified Needs of Travellers
†Kazuki MIYAI, Taku OKUNO
†Future University Hakodate

表 1: 旅行者のニーズとタグの例

旅行者のニーズ	タグの例
地域ならではのもの	イカ，ラーメン，海鮮丼，温泉，夜景，教会，土方歳三，石川啄木
旅行者にとって関心のあるテーマ	映画，登山，絶景，鉄道，市電，健康，美術，おしゃれ
旅行者の状況・状態によるニーズ	春，秋，ペット，学生，夫婦，家族連れ，ワンコイン，修学旅行

手順5において，階層クラスタ分析を行った後，それぞれのクラスタが何を表しているかを判断する必要がある．そこで，出現頻度が最も高い名詞がクラスタの特徴を表していると判断し，名詞の出現頻度を求める．ラベリング方法は次のとおりである．まず，階層クラスタ分析を行った名詞に再び形態素解析を行い，複合名詞を分割する．次に，分割した名詞の出現頻度を調べ，出現頻度が最も高い名詞をラベルとする．

5 「地域ならではのもの」に該当するタグの抽出方法

「地域ならではのもの」として地域の特徴語を抽出する．そして，地域特徴語をまとめたリストを作成し，4章の手順でタグを抽出する．この地域特徴語リストは4章の手順6における「定義したニーズ」のことを指す．地域特徴語リストの作成方法は以下のとおりである．

1. 検索エンジンを用いて「函館 観光」のような地域名と観光というクエリで検索する．
2. 検索結果から得られたウェブページに対して，形態素解析を行い，名詞を抽出する．
3. 4章で定義した手順2〜5により階層クラスタ分析を行い，ラベリングを行う．
4. 付与したラベルをまとめた地域特徴語リストを作成する．

6 予備実験

「地域ならではのもの」に該当するタグ抽出方法の有効性を検証するために，予備実験を行った．実験では，函館公式観光情報サイトはこぶらで公開されているデータを利用した．形態素解析器にはMeCabを用い，名詞のベクトル化にはword2vec[4]を用いた．word2vecの学習データにはWikipediaのデータから作成したコーパスを用いた．また，ベクトル間の類似度にはユークリッド距離を用い，クラスタリング手法にはWard法を用いた．

本手法では抽出した複合名詞がコーパスに登録されていない場合がある．そこで，登録されていない名詞は，再び形態素解析を行い，各名詞のベクトルを求め，句ベクトルを生成した．句ベクトルの生成では，各名詞のベクトルを加算する方法を用いた．

地域特徴語リストは，FindTravel，RETRIP，ぐるたびの3つのサイトを用いて作成した．

はこぶらから抽出した名詞に対して，階層クラスタ分析を行い，4435語のラベルを付与した．また，地域特徴語リストでは2643語の単語を抽出できた．この2つを文字列一致で比較した結果，609個のタグが抽出できた．

実験の結果,「夜景」,「イカ」,「重要文化財」のような函館ならではのものを抽出することができた．しかし，大半が「料理」,「年月」,「カウンター」などの不適切な名詞であった．この原因として不適切な名詞をラベリングしていることが考えられる．今後は，ラベリングをした後，不適切な名詞を除去する．

7 まとめ

本研究では，旅行者のニーズを考慮したタグを自動的に決定する手法を提案した．今後は定義した残り2つのニーズに該当するタグの抽出方法を検討した後，提案手法における有効性の評価を行う．

参考文献

[1] 松本敦志，杉本徹，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：情報処理学会全国大会講演論文集，第75回，pp.307-308(2013).

[2] 日本交通公社：〔テーマ1：観光の経済効果研究の今〕【1】地域と観光客、それぞれのニーズ，http://www.jtb.or.jp/research/column-now-01.

[3] 日本旅行業協会：第1部 国内旅行の現状と課題認識，https://www.jata-net.or.jp/membership/info-japan/research/03_1st.html.

[4] Mikolov, T., et al. : Distributed representations of words and phrases and their compositionality. In Proc. NIPS, pp.3111-3119(2013).