

観光レビューを対象とした耳より情報提示手法の提案

阪井 奎伍[†] 灘本 明代^{††}

[†] 甲南大学 大学院 〒658-8501 兵庫県神戸市東灘区岡本 8-9-1

^{††} 甲南大学 〒658-8501 兵庫県神戸市東灘区岡本 8-9-1

E-mail: [†]m1524003@s.konan-u.ac.jp, ^{††}nadamoto@konan-u.ac.jp

あらまし 近年フォートラベルやトリップアドバイザーのような観光地のレビューサイトが普及している。これらレビューサイトには実際に観光地に行ったことのある人の経験に基づくレビューが多く書かれており、観光地の様々な情報が多数存在する。しかしながら、これらのレビューには感想などの不要な情報も存在しているため、その中からユーザにとって有用な情報を取得することは困難である。これまで我々は、観光地を対象としたレビューサイトから、ユーザが「参考になった」「知って得をした」と感じる文を耳より情報と呼び、この耳より情報を抽出する手法の提案を行ってきた。本論文ではこれら耳より情報を抽出する際の耳よりキーワードを再考すると共に、ユーザにわかりやすく耳より情報を提示する手法を提案する。

キーワード 観光情報, 情報抽出, レビュー

1. はじめに

近年、インターネット上には観光に関する様々な情報が多数存在している。そこで旅行者は旅行の計画を立てる際にインターネット上の観光地に関する情報を集めることが多数ある。しかしながら、検索エンジンを用いて観光地の情報を収集しようとしても、観光地の様々な公式サイトでは基本的な情報が多く、それ以上の情報を見つけることは困難である。また、Q&Aサイトでは質問に対してピンポイントな回答が多いため、旅行者が知りたいと思う情報を見つけることは困難な場合が多い。その結果、実際に観光地に訪れてから「知っていれば」と後悔する場合が多数ある。

一方でレビューサイトの普及に伴い、商品に対するレビューだけでなく観光地を対象とした様々なレビューが多数存在している。フォートラベル^(注1)やトリップアドバイザー^(注2)に代表されるレビューサイトには観光地での経験に基づくレビューが書かれており、公式サイトにあるような基本的な情報だけでなく様々なお得な情報が多数記載されている。しかしながら、有名な観光地ほどレビューの量は膨大であり、その中には有用な情報のみならず不要な情報も多数記載されている。そのためユーザにとって有用な情報を見つけることは困難である。そこで我々は、このような不要な情報を排除し、お得と思われる情報を自動でレビューサイトから収集できると便利であると考えた。本研究ではこのユーザがお得と感じる情報を耳より情報と呼び、耳より情報を自動で抽出し提示する手法の提案を行っている [1]。

レビューの中には観光地に関する基本情報以外の情報は多数あるが、その中でユーザにとって有用でない情報や、既知の情報は耳より情報ではないと思われる。本研究ではユーザにとっ

て「有用な情報」であり且つ「意外な情報」を耳より情報とする。実際にレビューを分析すると、「有用な情報」には共通のキーワードがある事がわかった。そこで、本研究ではこの有用な情報に含まれる共通のキーワードの事を「耳よりキーワード」と呼ぶ。そして、この耳よりキーワードを用いて有用な情報をレビューの中から抽出する。さらに意外な情報とはある事柄に対して公知ではないレアな情報であると考え、抽出した有用な情報を含むレビューをトピックごとにクラスタリングする。そのクラスタ内の中心ベクトルからある程度離れている所にある文を意外な情報とする。この抽出した情報が耳より情報となる。

本論文では、これまで提案してきた耳よりキーワードの再考を行うと共に、抽出した耳より情報を見やすくユーザに提示する為のインタフェースの提案を行う。

以下本論文では、2章でレビューや旅行に関する関連研究を述べ、3章で耳より情報抽出手法について、4章で耳より情報の提示手法について述べる。そして5章では実験について述べ、6章でまとめと今後の課題について述べる。

2. 関連研究

レビューに関する研究は種々ある。服部ら [2] は、mixi のコミュニティからイベントを対象とした耳より情報を抽出する手法を提案している。耳より情報を抽出する際のキーワードやキャッチフレーズを「提案・推薦」、「抑止抑制」、「現状状況説明」、「可能・不可能」の4つのタイプに分類し、タイプ別にキーワードを提案している。本研究では被験者実験より観光地のレビューを対象に耳より情報を人手で抽出し、「アドバイス」、「時間」、「金銭」、「便利」、「地理」、「その他」の6つに分類すると共に観光地を対象とした耳より情報を抽出するための耳よりキーワードを提案する。

小林ら [3] は、ユーザの意見が生まれた理由が商品の重要な特徴であると考え、「条件」、「理由」、「態度」、「対象」の要素を定義し、Web 上の商品レビューから理由の書かれている部分の

(注1) : フォートラベル <http://4travel.jp/>

(注2) : トリップアドバイザー <http://www.tripadvisor.jp/>

抽出手法を提案している。小林らの定義した4つの要素のうち「だと思う」といった「理由」や、「なので」といった「態度」は本研究で提案する耳よりキーワードと類似しているが、本研究では耳よりキーワードにより抽出した情報にクラスタリングを行い、類似度を用いている点で異なる。伊木ら[4]はECサイトにおけるレビューを、類似性、協調性、集中性、情報性の4つの指標を定義した上で信頼性を可視化する手法を提案している。レビューの信頼性は重要なものであるが、本研究では信頼性について考慮していないため異なる。

田村ら[5]は商品に対するレビューを対象にレビューテキスト中の単語の出現に対して、ユーザ・アイテムのトピックを仮定したトピックモデルを提案している。新里ら[6]は商品の使用感について記述された商品レビューをオノマトペを利用し抽出する手法を提案している。本研究では旅行を対象としている点及びオノマトペではなく耳よりキーワードを提案し用いている点で異なる。池田ら[7]はBlog記事からユーザが体験を記述する際に現れる表現を体験表現と定義し、体験情報の抽出手法を提案している。体験情報抽出では実際の経験談の抽出を行っている点で本研究と異なる。

また、旅行に関する研究も種々ある。小澤ら[8]は、観光地をクエリとしたWeb検索結果を分析し、有用な情報の特徴から機械学習を用いてアドバイス文か否か、外出行動前に参照しておきたい情報か否かを判定する手法を提案している。本研究では耳よりキーワード抽出の際に機械学習を用いている点で異なる。

遠藤ら[9]は、Web上に存在する膨大な観光情報に着目し、特定地域の有用な観光情報の自動抽出・融合を行うための特定地域に限定せず、低コストに対象地域の有用な観光キーワードを自動取得手法を提案している。観光地のキーワードを用いて有用な情報を抽出する点は類似しているが、観光地の情報をレビューに限定している点やキーワードの取得手法については本研究と異なる。藤井ら[10]は旅行ブログを対象に「買う」、「食べる」、「体験する」、「泊まる」、「見る」、「その他」の6種類のタイプへ自動的に分類し、タイプごとに地図上に提示する手法を提案している。観光情報を抽出する点は類似しているが、観光情報を分類分けし提示する手法に対し、本研究では有用な情報を抽出する点で異なる。

中嶋ら[11]は、旅行ブログから旅行先の地の名所を抽出し、各所付随情報として体験情報、評価表現、状態情報、名所の由来や歴史の4つを抽出する手法を提案している。体験情報は耳より情報と類似しているが、ブログの記事から抽出している点で本研究と異なる。安藤ら[12]は、楽天トラベルのレビューにおいて「良くも悪くもユーザの心をぐっと掴むような極端なレビュー」を集め、読み手の心に響く表現を「インパクトのある表現」と定義し、読み手の心を動かすような情報について分析している。本研究では旅行のレビューにおいて各々の文との類似度を用いて耳より情報を抽出している点で異なる。

3. 耳より情報抽出手法

我々の提案する耳より情報抽出の手順を以下に示す。

(1) ユーザは耳より情報を取得したい観光地をクエリとして入力する。

(2) システムは、そのクエリを用いた検索結果のレビューを取得しこれらのレビューを文単位で分割する。

(3) (2)で抽出した文の中で耳よりキーワードを含む文を有用な情報として抽出する。

(4) (3)で抽出した有用な情報を、耳よりキーワード以外の名詞を用いてクラスタリングを行う。

(5) 各クラスタの中心ベクトルと各々のクラスタを構成する文の類似度を求める。

(6) 類似度がある閾値の幅内である文を有用な情報かつ人々にあまり知られていない意外な情報とし、これを耳より情報とする。

(7) 抽出した耳より情報を分かりやすくユーザに提示する。

ここで、観光地の各々のレビューは個人の感想や意見、経験談、そして耳より情報と様々な情報が混在している。本研究では耳より情報のみを抽出しユーザに提示するために、これらを分割する必要がある。実際のレビューを見てみると文単位で分割可能であることが分かる。例えば、「ふらっと訪れた伏見稲荷。頑張って登った奥社からの景色は最高でしたよ♪ただ上まで登るならば、相当な時間を見ておく方が良いでしょう。」という文では「。」と「♪」でそれぞれ一文に分割できる。そこで本研究では取得したレビューを全て文単位で分割し、文単位で耳より情報の抽出を行う。文の分割には、「。」、「.」、「!」、「!」、 「♪」、「☆」、「★」、「(笑)」、「w」、タブまたは改行が1つ以上出現した場合にそこが文の終わりとして分割する。またレビューの中には「小雨が降っていたので、足元も悪く、すべて躓いている方もいました。」という文と、これに続く「なので歩きやすい靴で行くのをお勧めします。」という文のように、2文以上で耳より情報と判断される場合も考えられるが、本論文ではこのような場合は考慮していない。

3.1 有用な情報の抽出

観光地のレビューに存在する耳より情報には、耳より情報の要因となるキーワードが含まれていると考えられる。例えば横浜中華街の「人気店だと複数店舗あるので、思っているよりも待たずに食べられます」というレビューでは、「待たずに」というキーワードが「時間」に関する有用な情報であると考えられる。また、「おすすめは、今人気の王府井の正宗生煎包という焼き小籠包と、これも定番ですが、皇朝の肉まんです」というレビューでは、「おすすめ」や「定番」というキーワードが相手に対する「アドバイス」という有用な情報であると考えられる。

このように耳より情報の要因となるキーワードを耳よりキーワードと呼び、これらを実験により決定し[1]、さらにこの実験結果を素性として機械学習により決定する。耳よりキーワードは様々な話題に関するキーワードがあり、実際のレビューを分析し、耳よりキーワードを「アドバイス」、「時間」、「金銭」、「便利」、「地理」、「その他」の6つに分類する。「アドバイス」は「～すべき」や「～推奨します」といった相手にアドバイスをしているキーワードを示す。「時間」は「昼間よりも夜の方が人が少なく待たずにすみません」という文のように、「待たずに」と言っ

表1 耳よりキーワード

アドバイス	推奨, 残念ですが, 不安定, 利便性, 要注意, の方がいい, ベスト, 良心的, 避けたほうが, もつてのほか, 禁止, 適して, 出来ず, 諦め
時間	往復, 長蛇の列, 長時間, 待たずに, 空いていて, 先着順, 整理券, ツアー時間, 無制限, ランチ
金銭	%off, 無料, 割引, 参加費, 請求, 方が安い
便利	足元, 飲み水, 必需品, レンタル, 着替え, 用意
地理	裏通り, 裏道, 脇道, トイレ, 地図, 交通機関, 施設, ロッカー, アクセス, 途中にある
その他	ベビーカー, 車椅子, 足の悪い, 接客態度

たような時間に関してお得と思われるキーワードを示す。「金銭」は「帰ろうとしたとき、休憩所があり、無料でお茶が飲みました」という文のように、「無料」と言ったような金銭に関してお得と思われるキーワードを示す。「便利」は「自販機などはないから飲み水は下から用意して持って行ったほうが良い」という文のように、「用意」といったような知っておいた方が良いキーワードを示す。「地理」は「有名な場所は観光客が多いのでベビーカーや車椅子を使わない場合は公共の交通機関の利用を強く推奨します」という文のように、「交通機関」といったような観光地または周辺の地理に関してお得と思われるキーワードを示す。「その他」は「一部に接客態度に疑問の残る店がないわけではないが、総じて味に関しては期待を裏切らない」といった文のように、「接客態度」といったような他の5つの分類に当てはまらないがお得と思われるキーワードを示す。耳よりキーワードとその分類の例を表1に示す。

また、これらの耳よりキーワードは実験より耳より情報と判断された文の中から耳より情報の要因となったキーワードとして作成したため、比較的汎用性のあるキーワードであると考えられる。しかし、レビューの中にはその観光地特有のキーワードがあると考えられる。例えば「段差が多いので、階段が辛い人にはちょっと大変かもしれません」という文は、しっかり整備されているようなアミューズメントパーク等では見られないレビューであるが、山の周辺にあるような神社等のレビューではよく見られる文である。そこで我々は、表1の耳よりキーワードの出現頻度を素性とし、Support Vector Machine(SVM) [13]を用いて対象の観光地に適した耳よりキーワードか否かを判定し、耳よりキーワードと判定したキーワードを新たに観光地に適した耳よりキーワードとする。

具体的には、ある観光地に対して耳より情報と判断された文を手で抽出し正解データとする。この正解データを対象に、表1の耳よりキーワードとそのTF-IDF値を素性とした学習データを作成する。次に、耳より情報を抽出する対象の観光地のレビューのすべての文を対象にjuman [14]により形態素解析を行う。そして表1の耳よりキーワードの名詞、動詞、形容詞とそのTF-IDF値を素性として、SVMを用いて形態素解析結果が耳よりキーワードであるか否かを判定する。ここで耳よりキーワードと判定したキーワードを対象の観光地に適した耳よりキーワードとする。

観光地のレビューにおいて、表1の耳よりキーワードもしくは観光地に適した耳よりキーワードを含む文を有用な情報とする。耳より情報の候補として有用な情報の抽出を行う。

3.2 意外情報

1つの話題の中でも多くの人が知っている内容とそうではない内容があり、多くの人が知らない内容はその話題の中ではあまり知られていないレアな情報であると考えられる。そこで本研究では、このようなレアな情報はユーザにとって意外な情報である場合がありそれが耳より情報となると考え、1つの話題の中であまりレビューがされていない内容である文をその話題の中の意外情報とする。また、3.1節で抽出した有用な情報には様々な話題が存在しているため、話題毎に分類し意外情報の抽出を行う。つまりは、有用な情報を話題毎にクラスタリングを行い、そのクラスタの中心ベクトルからある程度離れている範囲にある文を耳より情報とする。

具体的には3.1節で抽出した耳より情報の候補となる有用な情報すべての文を対象に、juman [14]により形態素解析を行い名詞を抽出する。この時、単に名詞だけを抽出すると、「商売繁盛」のような2つの名詞が連続して1つの単語となる複合名詞が「商売」と「繁盛」に分かれてしまうが、クラスタリングを行う際に「商売繁盛」という1つの語として扱うために、名詞が連続して出現した場合に限り連続している名詞同士を連結し、1つの語として抽出する。また、有用な情報抽出に用いた耳よりキーワードには名詞も含まれているため、耳よりキーワードを除く名詞を対象としてクラスタリングを行う。クラスタリングの手法は種々あるが、単文にある程度適していると考えられる [15]Repeated Bisection [16]を用いてクラスタリングを行う。

3.2.1 Repeated Bisection

本研究ではRepeated Bisectionを用いてクラスタリングを行うが、Repeated Bisectionについて簡単に述べる。Repeated Bisectionはクラスタリングツールbayon [17]やCLUTO [18]で使用されているクラスタリング手法であり、K-means法を $k=2$ で $n-1$ 回繰り返して n 個のクラスタを得る。すべてのデータを1つのクラスタに格納し、以下の手順を繰り返し、クラスタを2分割していき、クラスタリングを行う。

- (1) 全クラスタ中から最もまとまりの悪いクラスタを1つ選択する。
- (2) クラスタの中からランダムに2つの要素を選択しそれぞれを格納したクラスタを作成する。
- (3) 元のクラスタ内の全ての要素に対し、ランダムに選択した要素との類似度を比較する。
- (4) 類似度を比較した結果、より類似度高いクラスタに要素を格納する。
- (5) クラスタ間で要素の移動を行い、クラスタ内で類似度をそれぞれ比較し直す。
- (6) (5)を移動できる要素がなくなるまで繰り返す。

3.2.2 意外情報の抽出

有用な情報を対象にRepeated Bisectionを用いてクラスタリングを行った結果の各クラスタの中心ベクトルはそのクラスタを代表するトピックであるため、このトピックの中心に近

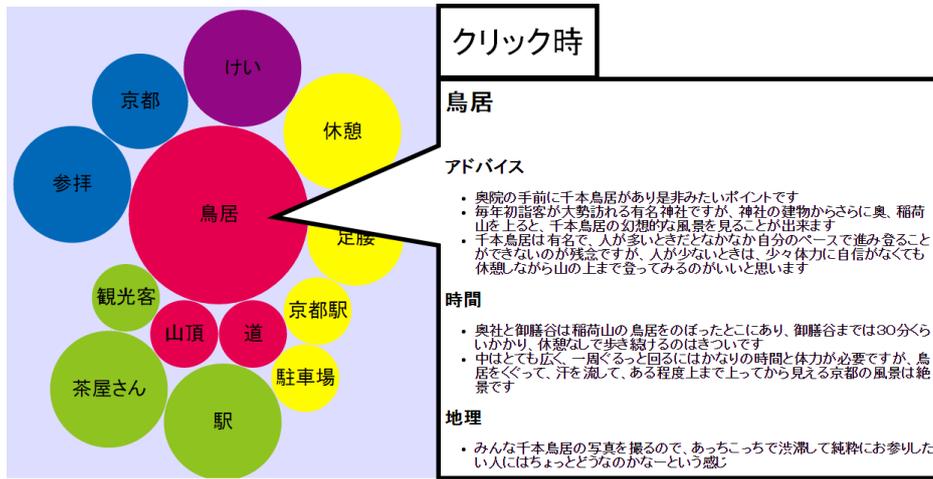


図1 インタフェース

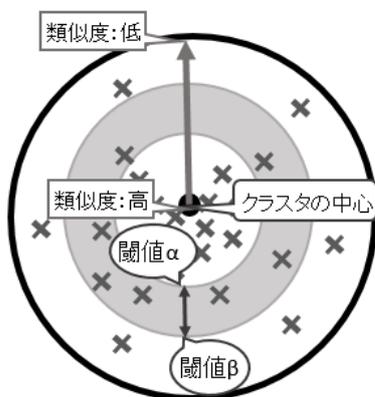


図2 クラスタ内の概要

いほど多くの方がレビューをしてい公知な内容であると考えられる。また、トピックの中心から遠くなるほど多くの方がレビューしていないレアな内容や、トピックとの関係が薄い内容になると考えられる。そこで我々は、トピックの中心からある程度離れている文は多くの方があまりレビューしていないレアな内容であり、あまり知られていない内容であれば意外な情報になると考え、トピックの中心からある程度離れているが離れすぎではない範囲にある文を意外情報とする。つまりは、図1に示すようにクラスタの中心からある程度離れた距離、閾値 α と β の間にある文が意外情報であるとする。このように求めた文は、有用な情報かつ意外情報であるので耳より情報とする。意外情報抽出の際の閾値 α 、 β は実験によりそれぞれ $\alpha=0.55$ 、 $\beta=0.6$ [1]とする。

4. 提示手法

3章で抽出した耳より情報をユーザにわかりやすく提示するために、耳よりキーワードの分類と意外情報の抽出の際の話題を用いる。提示手法について伏見稲荷大社を例として説明を述べる。意外情報抽出の際に、話題毎にクラスタリングした結果の各クラスタの中心ベクトルを構成する特徴語の特徴量が最も高い語をそのクラスタの話題として抽出する。ユーザはこの話

表2 データセット

観光地	データ数
伏見稲荷大社	2337 文
京都水族館	962 文
水木しげるロード	905 文
嵐山	769 文
横浜中華街	2022 文

題を選択することで、その話題における耳より情報を閲覧することができる。

提示手法の話題選択と話題選択後を図1に示す。各話題がどのような内容の耳より情報が含まれているかや、その話題における耳より情報の数が直感的に分かるように、クラスタ毎に各話題をラベルとし円で示す。クラスタの色は耳よりキーワードの6つの分類を示す。アドバイスは赤、時間は緑、金銭は橙、便利は紫、地理は黄、その他は青とする。色の決定はクラスタの耳よりキーワードの内最も多く含まれている分類とする。表示円の大きさは、クラスタ内に含まれる耳より情報の数を算出し、その数が多いほど円のサイズを大きくする。

ユーザは各話題、色、円のサイズから興味のある話題を選択する。選択した話題に含まれる耳より情報を、耳よりキーワードの分類毎に各文に含まれる耳よりキーワードの数により昇順で提示する。

以下にシステムの流れを示す。

- (1) ユーザは目的の観光地名を入力する。
- (2) システムは入力された観光地のレビューを取得し、有用な情報であり且つ意外情報である文を耳より情報として抽出する。
- (3) システムで抽出された耳より情報を話題毎に分けて提示する。
- (4) ユーザは提示された話題をクリックすることで選択した話題の耳より情報を閲覧する。

5. 実験

我々の提案した耳より情報の抽出手法は、基本情報を提示せ

ず耳より情報のみを提示することを行う。これは基本情報を知らないその観光地に行ったことのない人よりも、基本情報をすでに知っているその観光地に行ったことのある人の方が、我々の提案手法は効果的であると考えられる。そこで、観光地に行ったことのある人となない人各々に対して実験を行い、我々の提案手法がいずれの人々に有用であるかを計る。このとき、観光地へ行く旅行者はその観光地に興味があると考えられるため、被験者は行ったことのある人ない人両方とも観光地に対して興味がある人を対象とする。また、SVMを用いた観光地に適した耳よりキーワードの抽出の再現率を測るための実験を行った。

5.1 訪問経験の有無

5.1.1 実験条件

実験条件を以下と表2に示す。データセットは無作為に抽出した表2に示す有名な観光地である伏見稲荷大社、京都水族館、水木しげるロード、嵐山、横浜中華街の5つを用いた。被験者は各々の観光地での訪問経験有り3名、訪問経験なし3名である。

実験手順

(1) 各々の観光地に対して興味があり訪問経験もある3名、興味があり訪問経験はない3名の被験者6名が観光地のレビューの文を読み、耳より情報であると感じた文とそうでない文の2択で判断する。

(2) 興味あり訪問経験ありの3名の被験者の内2名以上が耳より情報と判断した文を訪問経験ありの耳より情報、同様に興味あり訪問経験なしで3名の被験者の内2名以上が耳より情報と判断した文を訪問経験なしの耳より情報とする。

(3) これらを用いて訪問経験の有無による違いの分析を行う。

5.1.2 結果と考察

表3に訪問経験の有無別の耳より情報の割合を示す。表3より京都水族館以外の観光地では、訪問経験がある人の方が耳より情報となるレビューが多いことがわかった。これは、実際に訪問したことがあるためレビューの内容を理解しやすく、それを次の訪問で活かすイメージをしやすいのではないかと考えられる。また、京都水族館のレビューでは「2012年にできたばかりで、館内はとてもきれいです」という文のように、公式サイトには載っていないが、一度でも訪問経験があれば知っているような観光地の説明等の情報が多くあったため、訪問経験のない場合の耳より情報が多くなったと考えられる。

このように、訪問経験の有無によって耳より情報と判断する文は変化することがわかった。そして我々の提案手法はレアな情報を耳より情報としているため、訪問経験ありの場合に適していると考えられる。

5.2 SVMを用いて追加した耳よりキーワードによる有用な情報抽出の精度

本論文では、これまで我々が提案してきた実験により決定した耳よりキーワード[1]に加えてSVMを用いて耳よりキーワードを新たに追加した。そこで本提案手法による耳よりキーワードの有用性を示す実験を行った。

表3 訪問経験の有無による耳より情報の割合

観光地	データ数	訪問経験有	訪問経験無
伏見稲荷大社	2337文	9.1%	1.9%
京都水族館	962文	1.2%	6%
水木しげるロード	905文	5.9%	2.7%
嵐山	769文	6.8%	2.9%
横浜中華街	2022文	8.4%	3.5%

表4 SVMにより抽出した耳よりキーワード例

	京都水族館	水木しげるロード	横浜中華街
アドバイス	目玉, 作れる	期待, 出会える	詐欺, 入れる
時間	うじゃうじゃ	活気付いて	ランチタイム
金銭	コスト, 田植え	予算, 提供	試食, 押し売り
便利	重宝, 難点	はがき, 持参	ガイドブック
地理	満員, ベンチ	マップ, フェリー	東京メトロ
その他	トレーナー	接待, 放映	勧誘, バイクン

表5 有用な情報抽出の再現率

観光地	再現率
京都水族館	75%
水木しげるロード	74%
横浜中華街	82%

5.2.1 実験条件

5.1節の実験結果から観光地に対して、興味があり訪問経験もある3名の内2名以上が耳より情報と判断した文を正解データとする。SVMの学習にはlibSVMを利用し、素性は伏見稲荷大社と嵐山のレビューの文を対象とし、表1の耳よりキーワードとそのTF-IDF値を正例、耳よりキーワード以外の名詞、動詞、形容詞とそのTF-IDF値を負例とした。このとき、データ数の大きい順に負例を正例のデータ数に合わせた。カーネルは線形カーネルを用いた。

次に、京都水族館、水木しげるロード、横浜中華街の各々のレビューを形態素解析し3つの観光地それぞれの名詞、動詞、形容詞とそのTF-IDF値を用いてSVMによりキーワードを抽出した。そのキーワードを表1の耳よりキーワードに追加し、評価データに用いた3つの観光地のレビューから有用な情報の抽出を行う。

5.2.2 結果と考察

3つの観光地各々において耳よりキーワードと判定されたキーワードと、そのキーワードを用いて抽出できた有用な情報の再現率をそれぞれ表4と表5に示す。表4に示すように、新しく追加した耳よりキーワードの中には、必ずしもその観光地特有のキーワードであるとは限らないキーワードも多く含まれている。しかしながら、観光地特有ではないものの、「詐欺」や「押し売り」といった表1にはなかった耳より情報に含まれるキーワードを抽出することができた。

表4のキーワードを追加した有用な情報の抽出では表5に示すように、70%以上の再現率で抽出することが出来た。これにより、SVMにより耳よりキーワードを追加する手法は有効であると考えられる。

6. まとめと今後の課題

我々はこれまで、観光地のレビューサイトから耳より情報を含む文を抽出する手法を提案してきた。本論文では抽出した耳より情報を話題毎に分け、耳よりキーワードにより「アドバイス」、「時間」、「金銭」、「便利」、「地理」、「その他」の6つに分類し、それぞれの話題に含まれる耳より情報によって6つに色分けすることで感覚的に分かりやすく提示する手法を提案した。また、耳より情報の中には観光地特有のキーワードを含む場合があると考え、機械学習を用いて観光地に適した耳よりキーワードの抽出手法を提案した。我々の提案する手法を用いることで、ユーザは効率的に新たな知識を得て、旅行の計画を立てる際の参考になると考えられる。

今後の課題として、ユーザインタフェースにおいて話題選択ではそれぞれの話題の座標は特に考慮していないので、ユーザにとってより分かりやすくするための座標計算をする必要がある。さらに、耳よりキーワードや各クラス内の耳より情報の数だけを用いて提示しているため、他の手法を検討していきたい。また提示手法の有用性を図るために、耳より情報をそのまま提示する場合と色分けや話題の円のサイズの変化を用いて提示する場合との比較を行う必要がある。現時点ではユーザが観光地名を入力するだけで耳より情報を提示することはできていないので、観光地名を入力して耳より情報を提示できるようにし、その上でシステムの処理速度等を考える必要がある。また、機械学習により追加した耳よりキーワードでの抽出では高い再現率を出すことができたが、その観光地特有の耳より情報に含まれるキーワードの抽出はできていないため改善する必要がある。

謝辞

本論文の一部はJSPS 科研費 26330347 及び、私学助成金（大学間連携研究補助金）の助成によるものです。ここに記して謝意を表します。

文 献

- [1] 阪井奎伍, 灘本明代, “観光を対象としたレビューからの耳より情報抽出”, 研究報告データベースシステム (DBS), 13, pp.1-6, 2015.
- [2] Yuki Hattori and Akiyo Nadamoto “Tip Information from Social Media based on Topic Detection” International Journal of Web Information Systems, Vol. 9, No. 1, pp. 83-94, 2013.
- [3] 小林 大祐, 井上 潮, “Web 上のレビュー情報からユーザが重要視する製品の特徴を抽出する手法の提案”, DEIM Forum 2009 C6-4, 2009.
- [4] 伊木 惇, 亀井 清華, 藤田 聡. (2014). “レビューを対象とした信頼性判断支援システムの提案” 情報処理学会論文誌, 55(11), 2461-2475.
- [5] 田村 一樹, 吉川 大弘, 古橋 武. (2015). “評点付きレビュー文書を対象としたトピックモデルの構築に関する検討” 情報処理学会論文誌, 56(3), 1013-1027.
- [6] 新里 圭司, 益子 宗, 関根 聡. (2015). “オノマトベを利用した商品の使用感の自動抽出” 情報処理学会論文誌, 56(4), 1305-1316.
- [7] 池田 佳代, 田邊 勝義, 奥田 英範, 奥 雅博. (2008). “Blog からの体験情報抽出” 情報処理学会論文誌, 49(2), 838-847.
- [8] 小澤俊介, 岡本昌之, 長野伸一, 長健太, 松原茂樹. (2012). “外

- 出行動前のユーザへの情報提供を目的とした Web からのアドバイス文抽出” 情報処理学会論文誌, 53(1), 105-116.
- [9] 遠藤 雅樹, 横山 昌平, 大野 成義, 石川 博, “特定地域に限定しない観光キーワードの自動抽出”, DEIM Forum 2014 E9-2, 2014.
 - [10] 藤井 一輝, 石野 亜耶, 藤原 泰士, 前田 剛, 難波 英嗣, 竹澤 寿幸, “多言語旅行ブログエントリを用いた観光情報提示システム”, DEIM Forum 2014 P4-1, 2014.
 - [11] 中嶋 勇人, 太田 学, “旅行ブログからの名所とその付随情報の抽出”, DEIM Forum 2013 B8-4, 2013.
 - [12] 安藤 まや, 石崎 俊, “インパクトの視点に基づく WEB 上のユーザレビューの分析”, 言語処理学会第 18 回年次大会, pp.731-734, 2012.
 - [13] C. Cortes, V. Vapnik, “Support-Vector Networks” Machine Learning, Vol. 20, pp. 273-297, 1995.
 - [14] 京都大学 大学院情報学研究所 知能情報学専攻知能メディア講座言語メディア分野:日本語構文解析システム JUMAN : <http://nlp.ist.i.kyotou.ac.jp/index.php?JUMA>
 - [15] 花井俊介, 灘本明代, “酷似レシピ抽出のためのクラスタリング手法の提案”, DEIM Forum 2014 F8-6, 2014.
 - [16] Ying Zhao and George Karypis. Comparison of agglomerative and partitional document clustering algorithms. Technical report, Department of Computer Science, University of Minnesota, Minneapolis, MN 55455, 2002.
 - [17] Bayon - a simple and fast clustering tool - Google Project Hosting <http://code.google.com/p/Bayon/>
 - [18] CLUTO - Software for Clustering High-Dimensional Datasets <http://glaros.dtc.umn.edu/gkhome/cluto/cluto/overview>