

ソーシャルメディアからのタイプ別耳より情報の抽出手法の提案

服部 祐基[†] 灘本 明代^{††}

[†] 甲南大学大学院 自然科学研究科 〒658-8501 兵庫県神戸市東灘区岡本 8-9-1

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

E-mail: [†]mn124005@center.konan-u.ac.jp, ^{††}nadamoto@konan-u.ac.jp

あらまし ソーシャルメディア上には有益であり且つ特有な情報が多く存在するため、ソーシャルメディアからこのような情報を取得する事が重要となっている。しかし、ソーシャルメディア上には膨大な情報が存在し、様々な話題が混在しているため、有益な情報を取得することが困難である。そこでこれまで我々は、ユーザにとって有益な情報を「耳より情報」として、ソーシャルメディアからトピック毎にクラスタリングされた耳より情報を抽出するシステムの提案を行ってきた。本論文では、これら取得した耳より情報をタイプ別に分類することにより、ユーザが効率良く有益な情報を取得することを可能とする。具体的には、ソーシャルメディアからトピック毎に取得した耳より情報をさらに「提案・推薦」、「制止・抑制」、「現状・状況説明」、「可能・不可能」の4つのタイプに分類し、タイプ毎に提示することを行う。

キーワード ソーシャルメディア, SNS, 情報抽出, 経験マイニング, トピック推定

Extracting Categorized Tip Information from Social Media

Yuki HATTORI[†] and Akiyo NADAMOTO^{††}

[†] Graduate School of natural science graduate course, Konan University Graduate School

8-9-1 Okamoto Higashinada-ku, Kobe-shi, Hyogo, 658-8501 Japan

^{††} Dept.of Intelligene and Informatics, konan University

8-9-1 Okamoto Higashinada-ku, Kobe-shi, Hyogo, 658-8501 Japan

E-mail: [†]mn124005@center.konan-u.ac.jp, ^{††}nadamoto@konan-u.ac.jp

1. はじめに

SNS やマイクロブログなどのソーシャルメディアを用いて、ユーザ自身の経験談や感想等の情報の発信や、様々なユーザ同士での情報の交換が頻繁に行われている。そのためこれらのソーシャルメディア上には様々な情報が大量に存在しており、一般の Web ページに掲載されていないソーシャルメディア特有の情報が多く含まれている。例えば京都の紅葉のコミュニティでは、“東福寺に行って来ました。凄い人ですが、紅葉も見頃を迎えると、通天橋から見下ろす景色は、人ごみの中ならんでも価値ありの本当に凄く綺麗な景色です。”のようにユーザの経験談や体験談に基づくお得な情報が多く書き込まれている。そのためソーシャルメディアから情報を得る行為がとても重要となっている。ソーシャルメディアでは、あるテーマに興味のあるユーザたちによってコミュニティが形成されているため、あるテーマの情報を得る際には、そのコミュニティでの書き込みから情報を見つけることができる。しかしソーシャルメディア

上には、多種多様のコミュニティが存在し、またコミュニティ内に関連する複数のトピックが混在しているため、ユーザが欲しいと思うソーシャルメディア特有の情報を容易に取得することはとても困難である。そこで我々は、ユーザにとって有益な情報を「耳より情報」と呼び、ソーシャルメディア特有の「耳より情報」を抽出する手法を提案する。

本論文で提案する耳より情報は、ユーザの経験情報が含まれ、かつユーザにとって有益な情報と定義する。したがって本論文では経験情報が含まれ、かつユーザにとって有益なコメントを抽出する手法を提案する。ここで経験情報を用いる理由は、実際に経験した事が書かれているコメントはそうでないコメントよりも信用できると考えた為である。また多くのユーザが有益と感じた情報の中には共通のキーワードが含まれていると考え、これらキーワードが含まれているコメント群を抽出する手法を提案する。その際にソーシャルメディア上のコミュニティでは、コミュニティ内に関連する複数のトピックが混在している問題があった。そこで我々はコミュニティ内に混在している

トピックを推定し、そのトピック毎に耳より情報を抽出することを行う。

これまで我々は、ソーシャルメディア内のあるテーマに沿ったコミュニティから、トピック推定を行い、それぞれのトピックに沿った耳より情報を抽出してきた[1]。しかしこれらの情報の中には様々なタイプの耳より情報が存在していることがわかった。例えば「～した方が良い」や「オススメです」のように閲覧ユーザに対して「提案、推薦」をしている耳より情報や「～できますよ」や「～はできません」などの「可能、不可能」を表す耳より情報などがある。そこで本論文ではソーシャルメディアからトピック毎に取得した耳より情報をさらに「提案・推薦」、「制止・抑制」、「現状・状況説明」、「可能・不可能」の4つのタイプに分類し、タイプ毎に提示することを行う。膨大な情報から耳より情報をトピック毎に抽出し、タイプ毎に提示をすることにより、ユーザがより効率良く有益な情報を取得することが可能となる。本研究のシステムの流れを図1に示す。

本研究では、前処理として、耳より情報の要因となる共通のキーワードをユーザ実験により取得し、耳よりキーワードとして辞書を作成する。

以下第2章には関連研究を、第3章にはSNSからの耳より情報抽出手法について、第4章には耳より情報のタイプ分類を、第5章にはプロトタイプシステムと実験を、第6章にはまとめと今後の課題を述べる。

2. 関連研究

Web ページから意見や評判情報の抽出を行う研究が盛んに行われている。Liu ら[2]は、競合している製品の消費者の意見を分析、比較をするための新しいフレームワークを提案している。彼らの提案しているシステムは、消費者の意見から製品のさまざまな特徴の良いところ、悪いところをユーザが明確にわかるように可視化している。Hu ら[3]はある製品に対するレビューを要約する手法を提案している。彼らの提案する要約技術は従来のテキストに対する要約とは異なっており、製品のレビューがポジティブ、ネガティブの意見かどうか注目し、製品の特徴のみを取得することを行っている。小林ら[4]は、意見の主要部分を「対象、属性、評価」の3要素とみなし、評価対象表現、属性表現、評価表現の共起パターンを利用することで、これらの表現を効率的に収集し、対象名辞書、属性表現辞書、評価語辞書を作成している。藤村ら[5],[6]はWeb文書を肯定・否定に分類し、肯定・否定の意味合いの強さの順位付けを行っている。文を構成する上で主要な単語を用いて、文節のn-gramによる素性を提案し、Web文書から抽出した評判情報を肯定・否定に判定する分類器を構築している。また鈴木ら[7]はWeblogから評価表現を抽出し、肯定・否定の分類を行っている。評価表現には、「おいしい」などのように、この単語自体に肯定・否定の感情が含まれているものや、「大きい」などのように、文を考慮しなければ肯定・否定の感情がわからないものがある。このような評価表現をBlog文書中から抽出し、ラベル付きデータだけでなくラベル無しデータも学習に組み込んだSemi-Supervisedな学習手法により、肯定・否定に分類し

ている。これらに対し、本研究では実際の経験を基に書かれているコメントに注目し、その経験情報からユーザにとって有益である情報を抽出している点が異なる。安藤ら[8]は、楽天レビューにおいて強調文や「！」などの記号を用いることでインパクトを与えているレビューを「インパクトのある表現」と定義し、インパクトのある表現の分析を行っている。数多く存在するレビューの中からインパクトのあるレビューを抽出するために、インパクトという読み手の心を動かすような言語現象とは何であるかという要因を詳細に分析している。本研究ではレビューではなく、ユーザ同士が自由にやり取りができるSNSから耳より情報を抽出することを行なっている。特に経験情報を重要視し、経験情報から耳より情報を抽出している点で異なる。

またLDAを用いたトピックの推定の研究も多く行われている。小池ら[9]は、時系列のニュース記事集合と同じ話題のブログ記事集合に対してLDAを用いてトピックの同定を行なっている。そしてニュースとブログ間の話題の相関や時系列での話題の変遷を分析している。また本来LDAなどの潜在的意味解析手法では単語に対してトピックを割り当てられているため、語の関係について考慮できないという問題がある。そこで北島ら[10]は、係り受けの関係にある語をイベントとし、単語からイベントにトピックを割り当てるための潜在的トピック抽出手法を提案している。本研究はSNSのコメントからLDAを用いてトピックを推定し、トピック毎に耳より情報を抽出している為これらの研究とは異なる。

3. SNSからの耳より情報の抽出

これまで我々が提案してきたSNSからユーザにとって有益な情報である耳より情報を抽出する提案手法[1]の概要を説明する。我々のこれまでの提案では、まずユーザが選択したコミュニティからコメントを取得し、トピックの推定を行うことにより、各コメントをトピック毎にクラスタリングする。トピック推定には様々な手法があるが、堤田ら[11]は、データが疎な場合は、潜在的ディリクレ配分法(Latent Dirichlet Allocation, LDA)[12],[13]やグラフベースが優れていると述べている。そこで、本研究では、単語数の少ない(データが疎な)多数のコメントからトピックを推定しそのトピック毎にクラスタリングを行う点、及び一つのコメントが複数のクラスタに含まれる点より、トピックの推定にLDAを用いる。そしてトピック毎にクラスタリングされたコメントから経験情報を述べているコメントを取得する。次に予備実験により抽出した耳より情報の要因となるキーワードの辞書をあらかじめ作成し、その辞書を用いることにより経験情報から耳より情報の抽出を行う。

3.1 トピック推定によるコメントのクラスタリング

本論文では耳より情報を抽出する最初の一步として、SNSを用いて耳より情報の抽出を行う。mixi^(注1)に代表されるようにSNSはあるテーマに沿ったコミュニティがあり、そのコミュニティ内には、サブテーマからなる複数のスレッドが存在する。そして1つのスレッドはそのコミュニティ内のユーザが投稿した

(注1): <http://mixi.jp/>

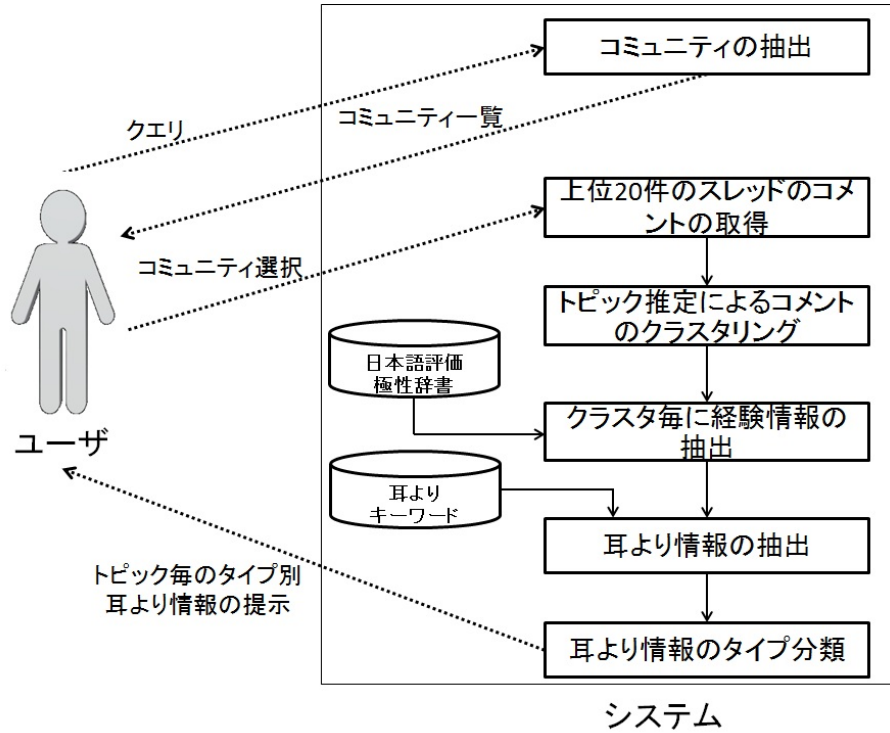


図1 システムフロー

複数のコメントから構成されている。ここで、一つのコミュニティ内で、スレッドが異なっても同一トピックに対して書いているコメントが多数あるため、本研究では、コミュニティ内の複数のスレッドのコメントを一つの文書集合と考え、トピック推定を行う。LDAでは、単語 w の集まりにより表現された文書集合と、トピック数 K から、各トピック z_k ($k = 1, \dots, K$) の単語 w の確率分布 $P(w | z_k)$ 、各文書 d のトピック z_k の確率分布 $P(z_k | d)$ ($d = 1, \dots, D$) を推定している。各文書は、トピック分布 θ を持っており、文書上の各単語が、 θ に従ってトピック z が選ばれる。そのトピック z の単語分布 ϕ に従って、単語 w 生成される。 D は文書数、 N_d は文書 d の単語の出現回数を表す。本研究では、文書集合の中の一つのコメントを一つの文書とみなし、LDAのハイパーパラメータである α と β は、それぞれ $\alpha = 0.5$ 、 $\beta = 0.1$ とする。さらに、トピック数 K は予備実験により、10として推定を行う。また文書に対してトピックを割り当てる場合、各文書 d のトピック z_k の確率分布 $P(z_k | d)$ ($d = 1, \dots, D$) がある閾値以上の場合にのみ行うが、これも実験により閾値を0.3とする。つまりは、 $P(z_k | d)$ の値が0.3以上の場合に、文書に対してトピックを与える。これにより文書をトピック毎にクラスタリングしていく。

3.2 経験情報の取得

我々は経験情報を書いているコメントは、そのユーザが経験した結果から発言している可能性が高いと考える。実際の経験に基づいたコメントはそうでないコメントより、有益であると考えられるため、推定したトピックによってクラスタリングされたコメント群に対して経験情報が書かれているコメントを抽出する。本研究では、乾ら[14]の提案する経験マイニングの評価・感情、出来事、行為に着目し、行為+評価・感情または行

表1 行為単語辞書の例

動詞		名詞	
行く	買う	参戦	利用
見る	盛り上がる	参加	入場
できる	探す	乾杯	移動
聴く	並ぶ	行動	興奮
飲む	踊る	活動	休憩
食べる	着る	集合	撮影

為+出来事を示す単語のセットを含むコメントを耳より情報の候補となる経験情報として取得する。しかしながら、乾らの提案する行為を表す単語は一般的な行為であり、よりドメインに依存した行為表現の抽出が必要になる。そこで本研究では、対象となるドメイン毎に行為を示す単語を決定し辞書を作成する。この辞書を行為単語辞書と呼び、人手によりこれらを作成した。表1に我々の作成したイベントをドメインとした行為単語辞書の一部を例として示す。ここでは、miki上のイベントのコミュニティにおける290スレッド、30000コメントから行為を表す単語を抽出し、上記の手順によりストップワードを除いた動詞、名詞の上位50単語を行為を表す単語として取得した。その際「する」や「購入」等経験していないコメントにも頻繁に出現する単語をストップワードとした。行為単語辞書はあらかじめドメイン毎に作成しておく。

そして作成した行為単語辞書と乾らの提案する評価・感情と出来事を用いてコミュニティのコメントから経験情報を含むコメントを以下の手順で取得する。

- (1) 3.1節よりトピック毎にクラスタリングされたコメントを取得する。

表 2 耳よりキーワードの例

耳よりキーワード				
オススメ	の方がよい	いかが	すべき	是非どうぞ
断然	魅力	イチオシ	無難	鉄則
狙い目	穴場	便利	裏道	不便
ガラガラ	渋滞	混雑	名所	困難

(2) 取得されたコメントのうち、クエッションマークや「教えて」、「質問」などの質問文とされるフレーズを含んでいるコメントは、他のユーザに情報を求めているため耳より情報にならないと考え削除する。

(3) 残りのコメントのうち、行為単語辞書の単語が含まれているコメントを取得する。ここで名詞の場合には後ろに動詞がかかっているものを対象とする。得られたコメントは行為を表す単語を含んだコメントとなる。

(4) 行為を表す単語を含んだコメントから、評価・感情や出来事を表す単語が含まれているコメントを抽出する。

3.3 耳よりキーワードの抽出

実際の耳より情報のコメントには、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、京都の紅葉のイベントの場合「左京区の曼殊院門跡もオススメです。もしそちらに行くなら詩仙堂も立ち寄るといいと思います。こちらもお庭が絶景です。」のようなコメントがある。この場合、「オススメ」や「絶景」などとても魅力を感じるキーワードが含まれ、また「～する」といいなどの提案をしているキーワードも存在する。このようなキーワードを本論文では耳よりキーワードと呼ぶ。耳よりキーワードは、複数のユーザが耳よりと感じたキーワードであると考え、被験者実験により耳よりキーワードを決定する。具体的には、ユーザ実験により耳より情報であるコメントの収集を行い、そのコメント群から人手により耳よりキーワードを抽出する。以前我々はドメインをイベントに絞り、イベントのコミュニティから耳よりキーワードを抽出する実験を行った。表 2 に得られた耳よりキーワードの一部を示す。そして抽出された耳よりキーワードを用いて、3.2 節で得られた経験情報を含むコメントに対して、耳よりキーワードが含まれている場合には、そのコメントを耳より情報とし抽出する。

4. 耳より情報のタイプ分類

これまでの我々の提案により得られた耳より情報には、「車でではなく電車で移動した方が良いです。」や「清水寺の紅葉はオススメです。」のように閲覧ユーザに対して提案や推薦をしている耳より情報や、「高台寺の紅葉のライトアップはとても混雑している。」や「今週末が見頃です。」のように現状を説明している耳より情報等、様々なタイプの耳より情報が存在している。そのため、ユーザにとって欲しい耳より情報を即座に取得しにくいことがわかった。例えば、高台寺の紅葉のライトアップの混み具合のみを知りたい人にとっては、清水寺の紅葉はオススメですといったような推薦情報は耳より情報になり得ない場合がある。実際には、ユーザ個々に欲しい耳より情報が異なるため、耳より情報をパーソナライズすることが好ましいが、本論

文ではユーザにあった耳より情報を提供するはじめの一步として、耳より情報をタイプ別に分類しユーザに提供することを行う。そこで我々はこれらの特徴的なキーワードに注目し、耳より情報をタイプ別に分類することを行う。そこで、実際の耳より情報を分析した結果、耳より情報のタイプを「提案、推薦」、「制止・抑制」、「現状・状況説明」、「可能・不可能」の 4 つのタイプに分類できることがわかった。以下に 4 つのタイプの定義を述べる。

- 提案・推薦

「～した方がよい」や「～がオススメ」など自分が経験した情報をコメントの閲覧者に対して提案したり、推薦したりしているタイプである。

例：「橋の上から見る紅葉がオススメです。」

- 制止・抑制

「～しない方がよい」や「～はやめておくべき」など、コメントの閲覧者に対して行動を制止したり抑制したりしているタイプである。

例：「車はやめておいた方がよいです。」

- 現状・状況説明

「混雑している」や「見頃」など、ある時点でのイベントの現状や状況の説明をしているタイプである。

例：「連休だったので大混雑していました。」

- 可能・不可能

「～ができる」や「～は禁止」など可能や不可能を表すタイプである。

例：「近くに駐車場があるので、停めることができます。」

ここで分類を行う際、1 つの耳より情報のコメントの中には複数のタイプが含まれている場合が多いため、1 つのコメントを 1 つのタイプに分類するのではなく、文毎にタイプの分類を行う。それぞれのタイプを表す単語を抽出するために、予備実験を行った。クエリとして SNS のイベント関連のコミュニティ「PL 花火芸術」、「京都の紅葉」、「関西海水浴」、「ねぶた祭」、「なばなの里」、「札幌雪まつり」、「潮干狩り」、「東京ゲームショー」の 8 つのテーマ計 4193 のコメントを用い、被験者 3 名によりすべてのコミュニティのコメントを読んでもらい、「参考になった、知って得をした情報」を耳より情報として取得する。その際にコメントの中で特に耳より情報の要因となる文を抽出し、4 つのタイプ毎に分類してもらった。そしてタイプ毎に分類された文に対して、タイプを表す単語を人手により取得した。取得されたタイプを表す単語の例を表 3 に示す。これらのタイプを表す単語を用いて、これまでの我々の提案手法を用いて取得した耳より情報の中でタイプを表す単語が含まれている文に対して、4 つのタイプに分類する。

5. プロトタイプシステムと実験

5.1 プロトタイプシステム

提案手法を用いてプロトタイプシステムを作成した。開発言語として PHP と MySQL^(注2)を用い、ソーシャルメディアとし

(注2) : <http://dev.mysql.com/>

表 3 耳より情報のタイプ

タイプ	表記例
提案・推薦	～が良い, オススメ, ～すべき, 必須, 最適, どうぞ, 推奨する, ベスト, いかがですか, 好ましい, 望ましい
制止・抑制	～しない方が良い, やめておくべき, 避けて, 無茶, 控えて, 迷惑, 厳禁, 禁物
現状・状況説明	混雑, 見頃, 交通規制, 開放, 終了, 満車, 人出, 始まる, イマイチ, 状態, 例年並み, 遅れている, 待ち, 販売
可能・不可能	～できる, ～できない, 可能, 不可能, 大丈夫, 問題なく, 無理, 間に合う, 充分

て SNS である mixi を用いた。そしてトピックを推定するためのツールとして, JGibbLDA^(注3)を用いている。

プロトタイプシステムの流れを以下に示す。

(1) ユーザは耳より情報を取得したいテーマを入力する (図 2 (a))。

(2) システムは SNS からユーザが指定したテーマのコミュニティを取得し, これらを一覧としてユーザに提示する (図 2 (b))。

(3) ユーザは提示されたコミュニティから閲覧したいコミュニティを選択する。

(4) システムはユーザの指定したコミュニティから最新 20 件のスレッドを取得する。

(5) 取得された 20 件のスレッドからコメントを取得する。ここで取得されたコメントすべてに対しトピックを推定し, それらトピック毎にコメントをクラスタリングする。本研究では LDA を用いてトピックの推定及びトピック毎のコメントのクラスタリングを行う。

(6) トピック毎にクラスタリングされたコメント群から経験情報を取得する。

(7) 経験情報から耳よりキーワードが用いられているコメントを耳より情報として取得する。

(8) 取得された耳より情報に対してタイプ毎の分類を行い, ユーザに提示する (図 2 (c))。

図 2 にシステムの画面図を示す。図 2 (c) 上部耳より情報のトピック表示部では, ユーザの選択したコミュニティ名に合わせて, トピック推定を行った結果である耳より情報のトピックを表す単語を掲載している。その際各トピックに対する単語の確率分布 $P(w | z_k)$ の上位 20 件を出力しているが, この値が大きいトピックは重要な単語であると考え, 重要な単語程文字サイズを大きく表示している。大きさは 3 段階にしており, 0.1 以上の場合に最大, ついで 0.1 未満 0.05 以上, 次に 0.05 未満を最小に表示している。また図 2 (c) 下部の耳より情報表示部では抽出された耳より情報に対して, タイプ毎に色で分類を行っている。提案・推薦をしている文には橙色に, 制止・抑制をしている文には青色, 現状・状況説明をしている文には緑色, 可能・不可能を表す文には黄色に分類し, 可視化することでどのタイプの耳より情報かを判断しやすくしている。

5.2 実 験

本提案手法の有効性を示すために実験を行った。以前我々は適合率のみを求めていたため, 本論文では, 正解データを作成することにより適合率と再現率, F-値を求めた。クエリとして

SNS のイベント関連のコミュニティ「PL 花火芸術」,「京都の紅葉」,「関西海水浴」,「ねぶた祭」,「なばなの里」,「札幌雪まつり」,「潮干狩り」,「東京ゲームショウ」の 8 つのテーマ, 計 4193 のコメントを用いた。予め被験者によりすべてのコミュニティのコメントを読んでもらい,「参考になった, 知って得をした情報」を耳より情報として取得し, 被験者 3 名のうち 2 名以上により耳より情報と判定されたコメントを耳より情報の正解データとして取得する。

考察

表 4 に各コミュニティにおける適合率, 再現率, F 値を, 表 5 には各被験者毎の適合率, 再現率, F 値を示す。表 4 より平均適合率が 68%, 平均再現率が 31% となった。「PL 花火芸術」,「京都の紅葉」,「関西海水浴」,「なばなの里」,「札幌雪まつり」,「潮干狩り」においては適合率が 60% 以上と良い結果となったが,「ねぶた祭」と「東京ゲームショウ」においては適合率が 60% 以下と悪い結果となった。悪い結果の原因として, イベントに対する優待席やお店に関する宣伝をしているコメントが多く含まれていたことが挙げられる。このようなコメントには,「参加する」や「飲む」,「楽しい」,「盛り上がる」などの経験を表す単語が含まれ,「プレゼント」や「割引」,「絶景」などの耳よりキーワードが含まれていたため, 宣伝のコメントも耳より情報として抽出されてしまっていた。このような宣伝のコメントに対しては, ユーザは耳より情報であると判断しないため, 適合率が大きく下がった結果となった。また再現率においてはすべての結果で 50% を下回る結果となった。原因として本研究で提案している耳より情報は, 経験情報が含まれユーザにとって有益な情報としている。そのためシステムにより得られた耳より情報は経験情報を含んでいる情報となるが, 実験で取得した正解データのコメントには,「～は絶景でオススメです」のように耳よりキーワードのみを含み, 経験情報を含んでいないコメントも多く存在していた。このようなコメントは経験情報を含んでいないためシステムは抽出できず, 再現率を下げた結果となった。経験に基づいて書かれたコメントは, そうではないコメントよりも信用できると考え, 耳より情報として経験情報を用いたが, ユーザは情報に対して信用性があるかどうかはあまり重要視していないことがわかる。しかしながら経験情報を含まない耳より情報は, 経験情報を含む耳より情報より信用性は劣るため, 経験情報を耳より情報とする点は必要だと考えられる。また表 5 より, 被験者によって大幅に適合率, 再現率, F 値の値が異なることがわかった。原因として被験者によってクエリとなるイベントに対する知識量や興味の度合いが異なるため, 耳より情報に違いが生じたと考えられる。そのため今後はよりユーザに特化した耳より情報の抽出手法を提案する必要

(注3) : <http://jgibblda.sourceforge.net/>

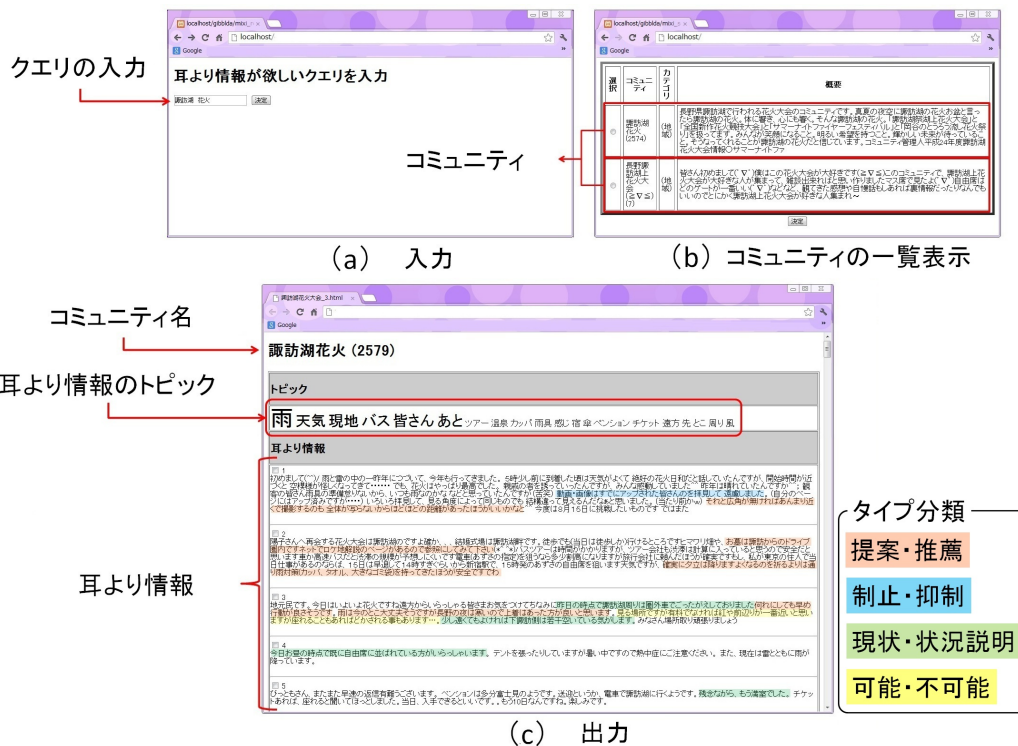


図 2 プロトタイプシステム

表 4 本提案手法の有効性

コミュニティ	適合率	再現率	F-値
PL 花火	75%	41%	0.53
京都紅葉	82%	33%	0.47
関西海水浴	64%	30%	0.40
ねぶた祭	44%	33%	0.38
なばなの里	72%	32%	0.45
雪まつり	74%	33%	0.46
潮干狩り	81%	25%	0.38
東京ゲームショー	51%	18%	0.26
平均	68%	31%	0.42

表 6 取得された耳より情報の例

PL 花火 芸術	トピック	狭山、駅、大阪、富田林、電車
	耳より情報	行きは近鉄・富田林駅、帰りは南海・大阪狭山市駅を利用すれば、駅に並ばずに電車に乗れます。富田林駅の一つ手前の喜志駅周辺も観覧スポットで、この駅なら富田林駅ほど混みません。阿倍野橋行きではなく、若干遠回りの河内長野経由で難波方面に向かうのがスムーズですよ。
諏訪湖花火 大会	トピック	花火、水上、写真、湖上、動画
	耳より情報	風向きは運ですが、最近の傾向だと、岸に向かって吹くことが多い気がします。と言うことと、人でごった返す岸よりも遊覧船の方が楽ですし、距離の関係もそんなに変わらないと思いますよ。むしろ安全に観れる絶好のポジション。遊覧船は打ち上げ場所の真ん中（初島あたり？）に陣取ってますから、障害物なしの大パノラマで観れます。船の中の場所取りはわかりませんが、毎年湖上で見ますが、僕自身なんの不満も抱きません。

がある。本システムで得られた耳より情報の例を表 6 に示す。

6. まとめと今後の課題

本研究では以前我々が提案した耳より情報の抽出手法から得られた耳より情報に対して、タイプ分類を行う手法の提案を行った。耳より情報の中には、提案を促すものや現状を知らせるものなど様々なタイプが存在しているため、本研究では耳より情報を「提案・推薦」、「制止・抑制」、「現状・状況説明」、「可能・不可能」の 4 つのタイプに分類した。タイプ分類を行うことにより、ユーザは数多くある耳より情報に対して、効率的に自分の欲しい耳より情報を見つけることが可能となる。今後の課題は以下の通りである。

- 耳より情報として抽出されたコメントの中には、イベントに関する広告、宣伝やコミュニティのテーマと関係のないコメントなどが多くふくまれていた。そのため今後はこのようなコメントを削除する必要がある。

- 実験により、ユーザによって耳より情報が異なることがわかった。よってパーソナライズされた耳より情報を抽出する手法を提案していく。

- 本研究では耳より情報をタイプ毎に分類を行ったが、今後は良い耳より情報や悪い耳より情報などの評価を考慮した耳より情報の分類を行う。

表 5 被験者毎の有効性

コミュニティ	A			B			C		
	適合率	再現率	F-値	適合率	再現率	F-値	適合率	再現率	F-値
PL 花火	76%	36%	0.49	42%	57%	0.48	79%	36%	0.49
京都紅葉	77%	25%	0.37	75%	31%	0.44	41%	24%	0.30
関西海水浴	73%	27%	0.39	50%	37%	0.43	54%	23%	0.33
ねぶた祭	49%	19%	0.27	19%	44%	0.26	54%	19%	0.28
なばなの里	76%	24%	0.37	33%	44%	0.38	77%	32%	0.45
雪まつり	75%	25%	0.38	32%	22%	0.26	70%	24%	0.36
潮干狩り	80%	21%	0.34	11%	10%	0.10	75%	15%	0.25
東京ゲームショウ	49%	16%	0.24	32%	25%	0.28	55%	15%	0.23

文 献

- [1] 服部祐基, 灘本明代, トピック推定に基づくソーシャルメディアからの耳より情報抽出手法の提案, 第 5 回 Web とデータベースに関するフォーラム, 2012.
- [2] B. Liu, M. Hu, and J. Cheng. Opinion observer: Analyzing and comparing opinions on the web. In *Proceedings of the 14th International Conference on World Wide Web*, pp. 342–351, 2005.
- [3] M. Hu and B. Liu. Mining and summarizing customer reviews. In *Proceedings of ACM-KDD*, pp. 168–177, 2004.
- [4] 小林のぞみ, 乾健太郎, 松本祐治, 立石健二, 福島俊一, 意見抽出のための評価表現の収集, 自然言語処理, Vol. 12, pp. 203–222, 2005.
- [5] 藤村 滋, 豊田 正史, 喜連川 優, 電子掲示板からの評判表現および評判情報の抽出, 第 18 回人工知能学会全国大会, pp. 4, 2004.
- [6] 藤村 滋, 豊田 正史, 喜連川 優, 文の構造を考慮した評判抽出手法, 電子情報通信学会 第 16 回データ工学ワークショップ 論文集, 6C-i8, 2005.
- [7] 鈴木泰裕, 高村大也, 奥村学, Weblog を対象とした評価表現抽出, 人工知能学会研究会資料, A401-02, 2004.
- [8] 安藤 まや, 石崎 俊, 楽天トラベルレビューにおけるインパクトのある表現の分析, 楽天研究開発シンポジウム, 2011.
- [9] 小池 大地, 横本 大輔, 牧田 健作, 鈴木 浩子, 宇津呂 武仁, 河田 容英, 吉岡 真治, 神門 典子, 福原 知宏, 中川 裕志, 清田 陽司, 関 洋平, ニュース・ブログにおける話題の相関と変遷の分析 -震災に関する話題を例題として-, 第 4 回データ工学と情報マネジメントに関するフォーラム 論文集, 2012.
- [10] 北島 理沙, 小林 一郎, 文書内の事象を対象にした潜在的ディリクレ配分法による要約, 第 3 回データ工学と情報マネジメントに関するフォーラム 論文集, 2011.
- [11] 堤田 恭太, 中辻 真, 内山 俊郎, 戸田 浩之, 内山 匡, アクセスログを用いたクロスドメイン環境における情報推薦, 情報処理学会研究報告, 2012-DBS-154, No.4, pp. 8.
- [12] D. M. Blei, A. Y. Ng, M. I. Jordan. Latent Dirichlet Allocation. In *Proceedings of Journal of Machine Learning Research.*, volume 3, pp. 993–1022, 2003.
- [13] T. Griffiths, M. Steyvers. Finding scientific topics. In *Proceedings of the National Academy of Sciences.*, 2004.
- [14] 乾健太郎, 原一夫, 経験マイニング: Web テキストからの個人の経験の抽出と分類, 言語処理学会第 14 回年次大会論文集, pp. 1077-1080, 2008.