

外出行動の事前情報を提供するための Web からのアドバイス文抽出

小澤俊介[†] 岡本昌之^{††} 長野伸一^{††} 長健太^{††} 松原茂樹[†]

[†] 名古屋大学大学院 情報科学研究科 〒 464-8601 名古屋市千種区不老町

^{††} (株) 東芝 研究開発センター 〒 212-8582 川崎市幸区小向東芝町 1

E-mail: [†]kozawa@el.itc.nagoya-u.ac.jp

あらまし 従来のユーザコンテキストを考慮した研究では、主に行動時の状況に応じた情報提供を行っており、行動前についてはあまり考慮されていない。しかし、登山や観光などの外出行動をする際には、必要な持ち物や目的地への行き方などの事前に必要な情報がある。こうした外出行動の事前情報は Web 上に散在しているため、事前情報だけを得るには手間がかかる。そこで本稿では、外出行動の事前情報を提供するため、Web から事前アドバイス文を獲得する手法を提案する。アドバイス文が否か、及び、事前に必要なアドバイス文が否かを分類学習を用いて判定することにより事前アドバイス文を獲得する。評価実験により提案手法の有効性を確認した。本稿ではまた、提案手法を実装した事前情報提供アプリを紹介する。

キーワード アドバイス, 事前情報, ユーザコンテキスト, テキストマイニング

Advice Extraction from Web for Providing Prior Information of Outdoor Activities

Shunsuke KOZAWA[†], Masayuki OKAMOTO^{††}, Shinichi NAGANO^{††}, Kenta CHO^{††}, and Shigeki MATSUBARA[†]

[†] Graduate School of Information Science, Nagoya University

Furo-cho, Chikusa-ku, Nagoya-shi, 464-8601, Japan

^{††} Corporate R&D Center, Toshiba Corporation

1 Komukai Toshiba-cho, Saiwai-ku, Kawasaki-shi, 212-8582, Japan

E-mail: [†]kozawa@el.itc.nagoya-u.ac.jp

Abstract Conventional context-aware recommendation systems do not provide information before taking user's action, although they provide information considering user's ongoing activity. However, the users want to know prior information such as how to go to their destination and necessary items, when they do outdoor activities such as climbing and sightseeing. It takes time to collect the prior information since it is not so easy to appropriately find them. This paper proposes a method for extracting prior advices from the web. The method first identifies whether a given sentence is an advice or not. Then the method identifies whether the sentence is a prior advice or not, if the sentence is identified as advice. We show availability of the proposed method through our experimentation. We also developed a system for providing prior information using the proposed method.

Key words advice, prior information, user's contexts, text mining

1. はじめに

近年、スマートフォンなどのモバイル機器の発達に伴い、屋外でネットにアクセスし、情報を得る機会が増加している。そのため、ユーザの状況を考慮したシステムに関する研究が盛んに行われており [1]、中でも、ユーザの状況に応じた情報推薦に

関する研究が注目されている [2] ~ [5]。しかし、これらの研究は主に状況の推定技術に主眼があり、ユーザに提供する情報としては、予め整備されたものや位置のみに基づくものが多く、提供する情報の中身については十分に考慮されていない。

登山や観光などの外出行動をする際には、必要な持ち物や目的地への行き方など、事前に知りたい情報がある。例えば、富

表 1 アドバイス文の例.

文番号	文	アドバイスが否か	事前が否か
1	昭和 39 年に開通した富士スバルラインの終点に位置し、麓の国道 139 号線から約 1 時間。	1	1
2	七合目から八合目は本八合目も含めて距離が長く、傾斜も急なので大変キツイと思います。	1	0
3	登山にあたっては、落石の恐れがありますので、十分ご注意ください。	1	0
4	ほぼ毎年富士山を訪れているが、2009 年も 8 月 1 日に富士に登った。	0	-
5	年末年始に富士山に登って、初日の出を拝めるのでしょうか？	0	-
6	ゆっくり歩いて、長時間休まないのが、楽に登るコツですね。	1	0
7	積雪時は岩場の通行は 困難 になります。	1	0
8	上着は防寒用にウインドブレーカがお 薦め。	1	1
9	また、残雪や転石がある箇所がありますので、足元にもお 気をつけ ください。	1	0
10	私がお 薦め するルートがあります。	0	-
11	トレッキングシューズ、軽登山靴など、靴底がしっかりしたハイカットのものがお すすめ です。	1	1
12	砂や小石が入り にくく、下山時に足首をしっかり包むので 捻挫 の 危険 を減らしてくれます。	1	1
13	靴底の薄いスニーカーなどは滑り やすく、小石を踏んだときに穴があくことがあります。	1	1
14	ただし、お盆休み近辺は マイカー 規制で入れません。	1	1
15	万一、天候の急変により危険を感じたら最寄の山小屋に避難しましょう。	1	0
16	八合五勺から九合目、山頂と傾斜はますます急になってくるので足場に注意しながら 進ん でいきましょう。	1	0

士山で登山する際には、富士山への交通手段や必要な登山用品は出発する前に知りたい情報である。外出行動の場合、大きく分けて行動前と行動時の 2 つの状況があるものの、既存のユーザ状況に応じた情報推薦システムでは行動時の状況に応じた情報提供を目的としていることが多く、外出行動の事前情報についてはあまり考慮されていない。

しかし、こうした外出行動の事前情報は Web 上に散在しているため、既存の Web 検索ではこれらのみを獲得するのは困難である。また、事前情報を含む Web ページには行動時の情報も含まれていることが多いため、事前情報だけを得るには手間がかかる。テキストから情報抽出する従来研究として、体験談やトラブルを表す文の獲得が行われているが [6] ~ [10]、獲得した文がいつ、どんな場面で利用できるかは考慮されていないため、事前に必要な情報であるかは分からない。

そこで本論文では、外出行動の事前情報を提供するため、Web から事前アドバイスを獲得する手法を提案する。本手法は場所名と行動名をクエリとした Web 検索により得た HTML 文書中の各文に対して、アドバイスが否か、及び、事前に必要なアドバイスが否かを分類学習を用いて判定を行うことにより、事前アドバイスを獲得する。また、提案手法を実装した事前情報提供アプリについて紹介する。

2. アドバイス文の特徴分析

2.1 アドバイス文とは

アドバイス文とは、行動前または行動時の情報提供につながる文である。表 1 に富士山での登山に関するアドバイス文の例を示す。表 1 の 3 列目はアドバイスが否かを表し、アドバイス文である場合は 1、そうでない場合は 0 である。アドバイス文にはそれぞれ利用できる状況が存在する。行動によって多種多様な状況が想定できるが、本論文では様々な行動に共通した汎用的な状況を考慮するため、事前と行動時の 2 種類の状況を想定し、事前に必要な情報であるか否かによってこれらを区別する。表 1 の 4 列目は事前に必要か否かを表し、事前に必要な場合は 1、そうでない場合は 0 である。例えば、例文 1 は事前の

表 2 分析データの規模.

場所	行動	文数	アドバイス文数	事前のアドバイス文数
富士山	登山	2,581	899	638
穂高岳	登山	4,144	360	132
合計		6,725	1,259	770

アドバイス文であり、例文 2 は行動時のアドバイス文である。

2.2 分析データの構築

アドバイス文の特徴を分析するためのデータを構築した。まず、Yahoo! 検索 Web サービス^(注1)を利用して HTML テキストを収集する。クエリには「富士山 登山」と「穂高岳 登山」の 2 つのクエリを用いて、それぞれ上位 50 ページの HTML テキストを収集した。次に、各 HTML テキストから本文を抽出し、それを文分割した。最後に、各文に対して、アドバイスが否かの判定、及び、アドバイス文である場合、事前に必要なアドバイスが否かの判定を人手により行った。ただし、対象の 1 文のみで判断できない場合は、前後の文を参照して判定を行った。構築したデータの規模を表 2 に示す。

2.3 アドバイス文の特徴

アドバイス文の特徴を捉えるため、目視により分析データを調査した。分析データは登山ドメインに関するものになっているが、登山ドメインに依存しない汎用的な特徴を捉えるように分析した。その結果、以下の 5 つの特徴が見られた。

A 文末機能語

アドバイス文は丁寧な文体で記述されていること(例文 3)が多く、過去の事実について記された文(例文 4)や疑問文(例文 5)には少ないという傾向があった。これらを捉えるため、文末の機能語に着目する。例文 3~5 の下線部のように、文末の機能語に着目することにより、丁寧な文であるかや疑問文であるかなどの特徴を捉える。

B 評価表現

アドバイス文には肯定の極性を持つ表現や否定の極性を持つ表

(注1): <http://developer.yahoo.co.jp/>

表 3 アドバイス文の特徴を表す手がかり表現.

クラス	手がかり表現
推薦	薦め, 勧める, 推奨, 方が良い, と良い, など
注意	気を付ける, 注意, 留意, 心掛ける, 用心する, など
準備	用意, 準備, 仕度, 持参, 所持, 携行, 備える, など
禁止	禁止, 禁物, 厳禁, 避ける, NG, ダメ, 違反, など
必要	必要, 必須, 必需, 不可欠, 要する, 欠かせない, など
計画	計画, 工程, スケジュール, ルート, プラン, など
混雑	渋滞, 混雑, 空く, 人混み, 満員, 空席, など
営業	営業, 開店, 開場, 閉店, 閉鎖, 開始, 終了, など
場合	場合, 際, とき, 状況, 時, に際して, 場面
緊急	万ー, 万が一, いざというとき, 非常時, もしも, など

現が含まれることが多いという傾向があった。例文 6 と 7 は肯定または否定の極性を持つ評価表現が含まれるアドバイス文であり、下線部は評価表現を表す。

C 手がかり表現

アドバイス文に含まれやすい表現を人手により収集した。470 表現を収集し、それらを意味ごとに 35 のクラスに分類した。収集した手がかり表現の一部を表 3 に示す。例文 8 と 9 は下線部に手がかり表現を含むアドバイス文である。

D 文末情報

評価表現や手がかり表現の出現位置によっても異なる傾向が見られた。例えば、例文 8 のアドバイス文では手がかり表現である「薦め」が文末にあるのに対し、アドバイス文ではない例文 10 では「薦め」が含まれているものの、文末ではない。このように、文の末尾の内容語が評価表現もしくは手がかり表現である場合、アドバイス文になりやすい傾向があった。

E 文脈情報

アドバイス文は Web ページ中でまとめて記述されることが多いため、対象文の前後の文にも評価表現や手がかり表現が含まれやすい傾向があった。例えば、例文 11 ~ 13 は同一の Web ページ中に連続して記述されていた文である。下線部は評価表現もしくは手がかり表現を表す。

2.4 状況に応じたアドバイス文の特徴

事前及び行動時のアドバイス文の特徴を捉えるため、目視により分析データを調査したところ、以下の 5 つの特徴が見られた。

A 手がかり表現

事前アドバイス文に含まれやすい 457 表現を人手により収集し、それらを意味ごとに 17 クラスに分類した。また同様に、行動時のアドバイス文に含まれやすい 109 表現を収集し、それらを意味ごとに 9 クラスに分類した。収集した手がかり表現の一部を表 4 に示す。表中の 1 列目の*が付与されたクラスは行動時を表しやすい手がかり表現である。例文 14 と 15 は下線部に手がかり表現を含むアドバイス文である。

B 動作動詞

アドバイス文に動作を表す動詞が含まれる場合、行動時のアドバイス文であることが多い傾向があった。例文 16 の行動時のアドバイス文では、動作動詞である「進む」が含まれている。

表 4 アドバイス文の状況を表す手がかり表現.

クラス	手がかり表現
交通	車, 電車, 高速道路, バス, 駅, 国道, など
準備	用意, 準備, 仕度, 持参, 所持, 携行, 備える, など
事前	事前, 予め, 前もって, 未然に, 先立って, など
天候	天気, 天候, 気温, 晴れ, 雨, 暑い, 寒い, など
不可能	不可, できない, 出来ません, られない, など
過去	た, だ
慎重*	慎重, きちんと, しっかり, ていねい, 丹念, など
コツ*	コツ, ポイント, 仕方, 方法, 対策, など
緊急*	万ー, 万が一, いざというとき, 非常時, もしも, など
可能*	可能, 出来る, できる, れる, られる, など

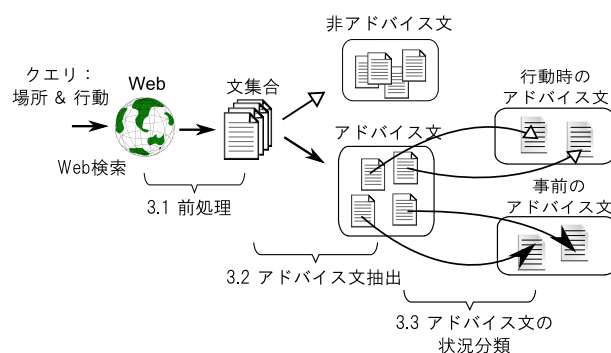


図 1 事前アドバイス文抽出の流れ

C 文末情報

末尾の内容語が事前を表しやすい手がかり表現である場合は事前アドバイス文、行動時を表しやすい手がかり表現もしくは動作動詞の場合は行動時のアドバイス文である傾向があった。

D 文脈情報

アドバイス文はまとめて記述される傾向があるため、事前アドバイス文の前後の文には事前を表しやすい手がかり表現、行動時のアドバイス文の前後の文には行動時を表しやすい手がかり表現や動作動詞が含まれやすい傾向があった。

3. 事前アドバイス文抽出

図 1 に事前アドバイス文の抽出の流れを示す。まず、場所と行動をクエリとして Web 検索を行い、タイトルに場所名を含む HTML テキストを獲得する。Web 検索には Yahoo! 検索 Web サービスを利用する。次に、取得した HTML テキストに対して前処理を行い、文集合を抽出する。そして、抽出した各文に対して、分類学習を用いてアドバイス文か否かの判定を行った後、アドバイス文と判定された文に対して、事前に必要なアドバイス文か否かの判定を行う。

3.1 前処理

HTML テキストから以下の手法により文を抽出する。まず、HTML タグの id 属性もしくは class 属性に以下のいずれかの文字列を含むタグ内の本文を抽出する。

content, entry, main

該当するタグが存在しない場合は、body タグ内の本文を抽出する。ただし、HTML タグの id 属性もしくは class 属性に以

下のいずれかの文字列を含むタグ内の本文は対象外とする。

head, foot, menu, copy, list, comment

次に、文分割をし、各文に対して Mecab [14] により形態素解析を行った後、以下の条件を満たす文を抽出する。ストップワードは Web 閲覧に関するアドバイスを除去するために用いる。

- 形態素数が 6 以上
- 動詞、副詞、形容詞、助動詞のいずれかを含む
- ストップワード (投稿, アカウント, ブラウザ, Adobe, JavaScript, スпам, コメント, など) を含まない

3.2 アドバイス文抽出

分類学習を用いて、アドバイス文か否かを判定する。分類学習の素性には 2.3 節の分析により得たアドバイス文の特徴を利用する。以下にアドバイス文抽出で用いる素性を示す。素性 a~e はそれぞれ 2.3 節で述べた特徴 A~E に対応する。

- 文末の助動詞が以下のいずれかであるか
「ぬ, ん, ない」「べし」「だ」「たい」「た」「う」「です」, 「ござる」「ます」「らしい」, その他
- 評価表現 (4 クラス, 13,590 表現) の各クラスの頻度
評価表現には日本語評価極性辞書 (用言編/体言編) [11], [12] を用いる。
- 手がかり表現 (35 クラス, 470 表現) の各クラスの頻度
- 末尾の内容語 (名詞, 動詞, 形容詞, 副詞) が評価表現もしくは手がかり表現の各クラスの表現であるか否か
ただし、末尾の内容語が以下のいずれかの場合は、その直前の内容語を調べる。
下さる, くださる, 思う, おもう, する, いる, くれる, できる, 出来る, おく, れる, られる, の, こと, もの
- 対象の前後 2 文における b~d
- 形態素長
- 各品詞の形態素長に対する割合

3.3 アドバイス文の状況分類

分類学習を用いて、事前に必要なアドバイス文か否かを判定する。分類学習の素性には 2.4 節の分析により得た特徴を利用する。以下にアドバイス文の状況分類で用いる素性を示す。素性 a~d はそれぞれ 2.4 節で述べた特徴 A~D に対応する。

- 事前を表しやすい手がかり表現 (17 クラス, 457 表現) と行動時を表しやすい手がかり表現 (9 クラス, 109 表現) の各クラスの頻度
- 動作動詞 (2 クラス, 447 表現) の各クラスの動詞の頻度
動作動詞には動詞項構造シソーラス [13] の中から移動動作 (物理) に属する 76 表現と身体的動作に属する 371 表現を利用した。
- 末尾の内容語が手がかり表現もしくは動作動詞の各クラスの表現であるか否か
ただし、3.2 節の素性 d と同様に、一部の内容語についてはその直前の内容語を調べる。
- 対象の前後 2 文における a~c

4. 評価実験

本手法の性能及び各素性の妥当性を検証するため、評価実験を行った。

表 5 評価データの規模.

文数	アドバイス文数	事前のアドバイス文数
1,335	172	107

表 6 アドバイス文抽出の実験結果.

素性	精度	再現率	F 値
uni-gram (ベースライン)	49.4%	26.2%	34.2
提案手法 (素性 a~g)	61.2%	35.7%	45.1
提案手法 - 文末機能語 (素性 a)	56.4%	31.5%	40.5
提案手法 - 評価表現 (素性 b)	56.5%	36.3%	44.2
提案手法 - 手がかり表現 (素性 c)	0	0	0
提案手法 - 文末情報 (素性 d)	43.4%	25.6%	32.2
提案手法 - 文脈情報 (素性 e)	53.7%	34.5%	42.0

4.1 評価データ

実験のため、評価データを構築した。まず、Yahoo! 検索 Web サービスを用いて「高尾山 登山」をクエリとして Web 検索を行い、タイトルに「高尾山」を含む上位 20 ページを取得する。次に、各 Web ページに対して 3.1 節で述べた前処理を行い、文を抽出する。最後に、各文に対して、アドバイス文か否かの判定、及び、事前に必要なアドバイス文か否かの判定を人手により行った。評価データの規模を表 5 に示す。

4.2 アドバイス文抽出実験

学習データとして 2.2 節で構築した分析データを、テストデータとして評価データを、素性として 3.2 節で述べた素性を用いて分類学習による実験を行った。分類学習のモデルには Support Vector Machine (SVM) を、SVM の学習には LibSVM^(注2)を、カーネルは線形カーネルを利用した。評価指標には、精度 (正しく獲得されたアドバイス文数/システムがアドバイス文と判定した文数) と再現率 (正しく獲得されたアドバイス文数/評価データ中のアドバイス文数)、F 値を用いた。ベースラインとして、uni-gram を素性として用いる手法を設けた。

実験結果を表 6 に示す。ベースラインと比較すると、精度・再現率ともに提案手法が上回っていることが分かる。また、提案手法から 2.3 節で述べたアドバイス文の特徴に対応する素性 a~e をそれぞれ除去した場合、全てのケースにおいて F 値が低下していることから、各素性がアドバイス文抽出において妥当であるといえる。

4.3 アドバイス文の状況分類実験

学習データとして分析データを、テストデータとして評価データを、素性として 3.3 節で述べた素性を用いて分類学習による実験を行った。分類学習のモデルには SVM を、SVM の学習には LibSVM を、カーネルは線形カーネルを利用した。評価指標には、精度 (正しく獲得された事前アドバイス文数/システムが事前アドバイス文と判定した文数) と再現率 (正しく獲得された事前アドバイス文数/評価データ中の事前アドバイス文数)、F 値を用いた。ベースラインとして、uni-gram を素性として用いる手法を設けた。

実験結果を表 7 に示す。ベースラインと比較すると、精度・

(注2): <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

表 7 アドバイス文の状況分類の実験結果.

素性	精度	再現率	F 値
uni-gram (ベースライン)	68.7%	76.7%	72.5
提案手法 (素性 a~d)	75.0%	81.3%	78.0
提案手法 - 手がかり表現 (素性 a)	62.0%	94.4%	74.8
提案手法 - 動作動詞 (素性 b)	73.1%	73.8%	73.5
提案手法 - 文末情報 (素性 c)	72.9%	80.4%	76.4
提案手法 - 文脈情報 (素性 d)	66.9%	75.7%	71.1

再現率ともに提案手法が上回っていることが分かる。また、提案手法から 2.4 節で述べた特徴に対応する素性 a~d をそれぞれ除去した場合、全てのケースにおいて F 値が低下していることから、各素性の妥当性が示された。

4.4 考察

アドバイス文抽出性能の低下原因について調査したところ、以下の原因が見られた。

- 目的の行動とは異なる行動のアドバイス文

誤って獲得してしまった 38 文のうち、42.1% にあたる 16 文には、対象とした「高尾山 登山」とは異なる行動に対する以下のようなアドバイス文が獲得されていた。

- 高尾山山頂よりも見晴らしが良いかも！

この誤りに関しては、アドバイス文が場所と行動に関連しているか否かを判定する必要がある。しかし、アドバイスを含むか否かという観点では、精度が 77.6% と高い性能を得られていた。

- 手がかり表現の不足

正しく獲得できなかったアドバイス文は 108 文あり、主な原因は手がかり表現の不足であった。108 文のうち、手がかり表現を含まない文が 31.5% の 34 文、1 つしか含まない文が 40.7% の 44 文あった。今後、手がかり表現を拡充する必要がある。

アドバイス文の状況分類性能の低下原因について調査したところ、以下の原因が見られた。

- 交通に関する手がかり表現

誤って事前アドバイス文と判定された 29 文のうち、27.6% にあたる 8 文には下記の例文の下線部のように、事前を表しやすい手がかり表現である交通に関する表現が含まれていた。

- 高尾山口 駅で、スタンプラリーなどをやっていけば、簡単なマップをもらえます。

これらを誤って獲得しないようにするためには、交通手段に関する記述であるかを判定する必要がある。そのため、交通を表す表現と「から」や「へ」などの起点や方向を表す助詞が共起しているかを考慮することにより対処できると考える。

- 動作を表さない動作動詞

誤って行動時と判定された事前アドバイス文は 20 文あり、それらのうち、35% にあたる 7 文には以下の例文の下線部のように、動作動詞が含まれていた。

- 道は比較的 歩き やすく整備されていますので、履き なれたスニーカーのような 歩き やすい靴でどうぞ。

上記の文の場合「歩く」や「履く」は、「歩きやすい」や「履きなれる」といった性質を表しており、動作を表していない。これらに対しては「やすい」や「にくい」などの性質を表しやす

表 8 SVM の確率推定値を用いた結果

	アドバイス文抽出	事前アドバイス文判定
0.9 以上	73.3% (11/15)	84.8% (39/46)
全て (0.5 以上)	61.2% (60/98)	75.0% (87/116)

表 9 他ドメインでの実験結果.

場所	行動	アドバイス文抽出	アドバイス文の状況分類
高尾山	登山	61.2% (60/98)	75.0% (87/116)
しながわ水族館	観る	54.3% (19/35)	75.0% (9/12)
東京ディズニーランド	遊ぶ	50.6% (39/77)	75.8% (22/29)

い形態素が動作動詞に後続しているかを考慮することにより対処できると考えている。

提案手法を用いて抽出したアドバイス文をユーザに提供することを想定した場合、システムにより提供できるアドバイス文数には限度がある。そのため、多くのアドバイス文をカバーするよりも高い精度でアドバイス文を獲得できることがより重要となる。そこで、よりアドバイス文らしい文のみを抽出した場合の精度について調査した。アドバイス文らしさには SVM が出力する確率推定値 (0.5~1.0) を用い、確率推定値が 0.9 以上のアドバイス文の精度を調査した。これは高いほどアドバイス文である確率が高いことを表す値である。アドバイス文の状況分類についても同様の調査を行った。結果を表 8 に示す。評価データ全体での精度に比べ、高い精度が得られていることが分かる。システムで利用する際、SVM の確率推定値を利用することにより、高い精度でアドバイス文を提供できるといえる。

4.5 他ドメインへの適用

異なるドメインの外出行動の場合、知りたい事前情報のタイプが異なることが想定される。例えば、水族館や遊園地に行く場合、ショーやイベントなどの事前情報が知りたいと考えられる。そこで、異なるタイプの事前情報を獲得できるかを検証するため、「しながわ水族館 観る」と「東京ディズニーランド 遊ぶ」の 2 つの外出行動に対して、提案手法を適用した。

アドバイス文の獲得精度を表 9 に示す。「高尾山 登山」の結果と比較すると、同程度の精度でアドバイス文を獲得できていることが分かる。以下に獲得例を示す。1 つ目は「しながわ水族館 観る」に対するショーの混雑に関するアドバイスであり、2 つ目は「東京ディズニーランド 遊ぶ」に対する上手にアトラクションに乗る準備に関するアドバイスである。

- ショーの時間通りに行かれると混雑が予想され見れない場合があります。

- 全 44 アトラクションの内、「必ず乗る」「できれば乗る」「乗らない」の 3 つに分類してください。

このように、異なるドメインに対して、登山では現れないような事前情報が獲得できていることから、提案手法は他ドメインへも適用可能であるといえる。

5. 事前アドバイス文提供アプリ

提案手法を用いて、事前アドバイス文提供アプリを試作した。本システムは場所名と行動名を入力としてユーザから受け取る。

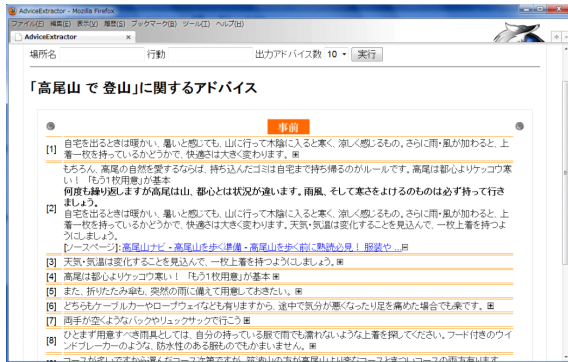


図 2 事前アドバイス文提供アプリ

入力された場所名と行動名をクエリとして用いた Web 検索の結果から場所名をタイトルに含む上位 20 ページを取得し、それらから提案手法により抽出した事前アドバイス文を提供する。

場所名を高尾山、行動名を登山として検索した結果を図 2 に示す。図 2 では、「高尾山 登山」に関する事前アドバイス文を 10 文提供している。各アドバイス文の末尾には「+」ボタンがあり、これを押すことにより前後 2 文を参照することができる。また、事前アドバイス文を抽出した Web ページの URL を辿ることができるようになっている。

システムの実装には、Perl を用いた。Web 検索には Yahoo! 検索 Web サービスを利用した。

6. 関連研究

Web からアドバイス文と類似する情報を抽出する研究がいくつか行われている。De Saeger らはトラブルを表す表現などを用いて Web からトラブルを表す文を抽出する手法を提案している [10]。Yin らは手がかり表現などを用いて Web からノウハウを獲得する手法を提案している [15]。また、乾らや Park らは Web から個人の経験を獲得する手法を提案している [7], [9]。しかし、これらの研究では抽出対象の情報であるか否かのみに着目しており、獲得した情報がいつ、どんな場面で利用できるかは考慮されていない。

ブログから地名・ランドマークに関する体験を抽出する手法が提案されている [6]。位置情報は考慮されているものの、いつ利用できる情報かは考慮されていない。また、倉島らはブログから経験情報を獲得する手法を提案している [8]。経験情報は状況と行動、主観から構成されており、このうち、状況は時間と空間として定義されている。時間にはブログの RSS 配信時間、空間には固有表現抽出技術によって抽出した地名及び組織名を用いている。しかし、これらはブログのユーザが経験した状況であるため、情報提供できる状況を表しているとは言えない。

これらに対し、本研究では、情報提供できる状況を考慮するため、アドバイス文であるか否かだけでなく、事前に必要な情報であるか否かも判定することによりアドバイス文を獲得する。

7. おわりに

本研究では、外出行動の事前情報を提供するため、Web から事前アドバイス文を獲得する手法を提案した。また、提案手法

を実装した事前アドバイス文提供アプリについて述べた。

今後の課題としては、アドバイス文が対象の場所及び行動と関連しているかを考慮したアドバイス文抽出手法の考案が挙げられる。また、事前と行動時の 2 種類だけでなく、より詳細な状況を考慮することが挙げられる。

文 献

- [1] Baldauf, M., Dustdar, S. and Rosenberg, F., "A Survey on Context-Aware Systems," International Journal of Ad Hoc and Ubiquitous Computing, Vol. 2, No. 4, pp. 263-277, 2007.
- [2] Cheverst, K., Davies, N., Mitchell, K. and Friday, A., "Experiences of Developing and Deploying a Context-Aware Tourist Guide: the GUIDE Project," Proc. of the 6th Annual International Conference on Mobile Computing and Networking, pp. 20-31, 2000.
- [3] Van Setten, M., Pokraev, S. and Koolwaaij, J., "Context-Aware Recommendations in the Mobile Tourist Application COMPASS," International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems, pp. 235-244, 2004.
- [4] Oku, K., Nakajima, S., Miyazaki, J., Uemura, S. and Kato, H., "A Recommendation Method Considering Users' Time Series Contexts," Proc. of the 3rd International Conference on Ubiquitous Information Management and Communication, pp. 306-312, 2008.
- [5] Zheng, V. W., Zheng, Y., Xie, X. and Yang, Q., "Collaborative Location and Activity Recommendations with GPS History Data," Proc. of the 19th International Conference on World Wide Web, pp. 1029-1038, 2010.
- [6] Kurashima, T., Tezuka, T. and Tanaka, K., "Blog Map of Experiences: Extracting and Geographically Mapping Visitor Experiences from Urban Blogs," Proc. of the 6th International Conference on Web Information Systems Engineering, pp. 496-503, 2005.
- [7] Inui, K., Abe, S., Hara, K., Morita, H., Sao, C., Eguchi, M., Sumida, A., Murakami, K. and Matsuyoshi, S., "Experience Mining: Building a Large-Scale Database of Personal Experiences and Opinions from Web Documents," Proc. of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, pp. 314-321, 2008.
- [8] Kurashima, T., Fujimura K. and Okuda, H., "Discovering Association Rules on Experiences from Large-Scale Blog Entries," Advances in Information Retrieval, Proc. of the 31st European Conference on Information Retrieval, pp. 546-553, 2009.
- [9] Park, K.C. and Jeong, Y. and Myaeng, S.H., "Detecting experiences from weblogs," Proc. of the 48th Annual Meeting of the Association for Computational Linguistics, pp. 1464-1472, 2010.
- [10] De Saeger, S., Torisawa, K. and Kazama, J., "Looking for Trouble," Proc. of the 22nd International Conference on Computational Linguistics, pp. 185-192, 2008.
- [11] 小林のぞみ, 乾健太郎, 松本裕治, 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, No.2, pp. 203-222, 2005.
- [12] 東山昌彦, 乾健太郎, 松本裕治, "述語の選択好性に着目した名詞評価極性の獲得," 言語処理学会第 14 回年次大会論文集, pp. 584-587, 2008.
- [13] 竹内孔一, 乾健太郎, 竹内奈央, 藤田篤, "意味の包含関係に基づく動詞項構造の細分類," 言語処理学会第 14 回年次大会論文集, pp. 1037-1040, 2008.
- [14] 工藤拓, "Mecab: Yet Another Part-of-Speech and Morphological Analyzer," <http://mecab.sourceforge.jp/>
- [15] Yin, L. and Power, R., "Adapting the Naive Bayes classifier to Rank Procedural Texts," Lecture Notes in Computer Science, 3936, pp. 179-190, 2006.