

観光地における行動の自動分析に向けて

山本 拓未*1 徳久 雅人*2 村田 真樹*2 村上 仁一*2

*1 鳥取大学 工学部 知能情報工学科

*2 鳥取大学 大学院 工学研究科 情報エレクトロニクス専攻

{s082066, tokuhisa, murata, murakami}@ike.tottori-u.ac.jp

1 はじめに

旅日記のブログには個人の体験談や旅行の感想などが多々記述されており、旅行者の行動分析の対象となっている。

池田らは、自発的な動作を表す動詞および名詞を用いて、ブログに体験情報が含まれているかを判別した [1],[2]。西原らは、ブログ内の語を手がかりに、場所・動作・対象物を画像で表現した [3]。石野らは、表層パターン手法と機械学習を用いて、土産物やホテル名に関する情報を抽出した [4]。高野らは、実体験を示す語等を用いて、ブログがイベントの体験表現を含むかを判別した [5]。しかし、いずれも行動の推移の分析には至っていない。

そこで、本研究では、ブログから「ブルーベリー狩り」をテーマとし、そのテーマに関連する行動を抽出する「行動推移分析支援システム」を開発する。

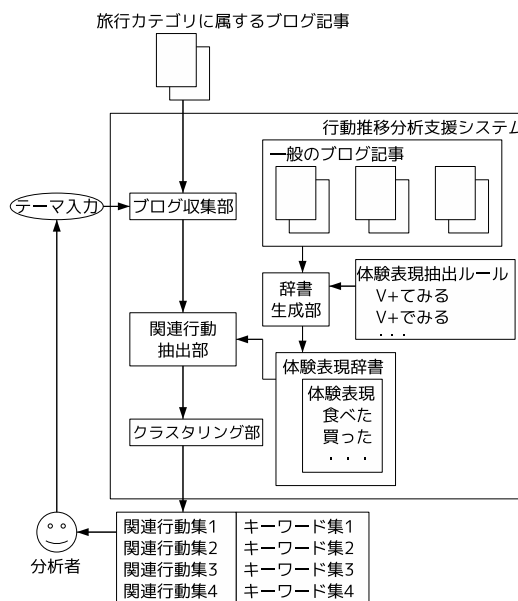


図 1: 行動推移分析支援システムの構成

2 行動推移分析支援システム

行動推移分析支援システムは、体験表現抽出ルール、辞書生成部、体験表現辞書、ブログ収集部、関連行動抽出部、および、クラスタリング部で構成する (図 1)。体験表現抽出ルールは、[1],[2]にならい、「～してみる」などの手掛かり語で構成する。辞書生成部は、一般のブログ記事に本ルールを照合し、得られた体験表現を辞書として登録する。ブログ収集部は、旅行カテゴリに属するブログ記事から、入力されたテーマに関する記事を収集する。関連行動抽出部は、体験表現辞書を該当テーマのブログ記事に照合し、適合文および前後一文を関連行動とみなして抽出する。最後に、関連行動の文のベクトルを求めてクラスタリングを行い、分析者に提示する。これを受けて、分析者は旅行者の行動を分析する。

2.1 辞書生成部

辞書生成部では、体験表現抽出ルールを用いて一般ブログから体験表現辞書を生成する。

体験表現抽出ルールは、文献 [1],[2] を参考にし、また、一般ブログを観察することで、表 1 のとおりとした。条件部は変数と字面で構成する。V は変数であり、「動詞」または「用言性名詞+する」に適合する。文と条件部が適合すると、3 通りの出力を行なう。この出力を「体験表現」として辞書に登録する。

表 1: 体験表現抽出ルール

条件	「V+てみる」「V+てみた」「V+てみます」「V+てみました」「V+たことがある」「V+たことがあった」のいずれかにマッチすること
出力	条件より抽出されたもの、V の過去形、「V+ました」

一般のブログ記事（5億4,700万文）を形態素解析したのから体験表現を抽出したところ、12,262件の表現を得た。表2にその一部を示す。たとえば、「行ってみました」と「V+てみました」が適合して出力されるものは、「行ってみました」、「行った」、「行きました」の3件である。

表 2: 体験表現辞書の例

動詞型	行った, 行きました, 食べた, 食べました, 作った, 作りました, 聞いた, 聞きました, 立ち寄った, 立ち寄りました…
用言性名詞型	観光した, 観光しました, 体験した, 体験しました, 作成した, 作成しました, 料理した, 料理しました, 冷凍した, 冷凍しました, …

2.2 ブログ収集部

ブログ収集部では、分析者が入力したテーマを検索キーとして、一般の検索サイトを通じてブログ記事を検索する。

ここで、「ブルーベリー狩り」をテーマに検索すると、「ブルーベリー狩り」を体験していない記事もヒットする。たとえば、「試験が終わったら、ブルーベリー狩りに行きたいなあ。」という文にヒットする場合である。したがって、ブログ記事のタイトルにテーマが含まれることを条件に、記事の絞り込みを行なうという選択肢を用意している。

収集した記事は、文の境界を自動判別しながら記事ごとに格納する。

2.3 関連行動抽出部

関連行動抽出部では、ブログ記事から、体験表現を含む部分を検索する。

ブログ記事の1文ずつを切り出したのでは、体験の様子が想像しにくい。そこで、体験表現を含む文（以降、体験表現文と呼ぶ）、および、その前後文の3文を1セットとして抽出する。こうして得られたセットを、テーマに関連する行動の情報とみなし、「関連行動」と呼ぶこととする。

関連行動は、テーマの記述の前に出現したものと、テーマの記述の後に出現したものとに分けて格納する。これにより、テーマの前に行なう行動か、後に行なう行動かを区別する。

たとえば、「ブルーベリー狩り」をテーマとするとき、「旅行者がブルーベリー狩りの前にどんなことをしているのか、ブルーベリー狩りの後にどんなことをしているのか」ということを分析できると予想している。

2.4 クラスタリング部

クラスタリング部では、関連行動の集合を幾つかに分割するとともに、そのテーマ「ブルーベリー狩り」ブログ記事に概要を表現する。

多くの関連行動を読むことは、分析者に負担である。そこで、まずは、類似する関連行動をいくつかのクラスタに自動分割する。本稿では、Bayon [6] を用いる。Bayon は、K-means 法、または、Repeated Bisection 法でクラスタリングを行なう。次に、得られたクラスタの概要を把握するためにキーワードの抽出を行なう。本稿では、KeyGraph [7] を用いる。KeyGraph は、文書の主張とその関連語を抽出することができる。論文などの文書は、重要語の繰り返しがあることと、重要語にともなう説明文があるため、KeyGraph は有効に動作する。本稿ではクラスタに対して KeyGraph を試みるため、KeyGraph 本来の用途と少し異なるのだが、簡便であり、類似研究 [8] で使われていることから、試行する。

クラスタリング部は、得られたクラスタごとに、KeyGraph で抽出されたキーワードを表示する。分析者は、クラスタに含まれる各文とキーワードを参考に、分析を行なう。

3 実験

体験表現文の抽出およびクラスタリング結果について、評価実験を行なう。

3.1 実験手順

手順 1 実験の対象となるブログを収集する。本実験では、ameba ブログのブログ検索を用いる。テーマは、「ブルーベリー狩り」とする。

手順 2 収集したブログから、評価用の正解データを作成する。人手により抽出すべき体験表現文に注釈を付与する。

手順 3 収集したブログから体験表現文を抽出する。そして、関連行動を得るために前後文を追加する。

手順 4 抽出した関連行動の集合に、クラスタリングを行う。

手順 5 各クラスタの文集合より、KeyGraph でキーワードを抽出する。

手順 6 各クラスタの文集合とキーワードより、人手により関連行動の概要を把握する。

手順2と手順3の比較により、体験表現文の抽出についての性能が評価できる。手順6の結果より、関連行動の推移分析に向けての性能を、分析者の感想として評価する。

3.2 実験条件

クラスタリングの対象は関連行動である。関連行動のベクトル化は、含まれる文の名詞の *uni-gram*、動詞の見出し語の *uni-gram*、類語大辞典(講談社)[9]の名詞・動詞の意味コードを素性に用いて行う。

3.3 実験結果

3.3.1 体験表現文抽出の評価

テーマ「ブルーベリー狩り」の記事は、60件(1,216文)得られた。体験表現文は、正解データとしては154文が含まれていた。本システムにより、体験表現文は229文が得られた。

正解データと本システムの比較を表3に示す。ここで、ベースラインとは、全ての文を体験表現文であると判別した場合である。

表3: 体験表現文の抽出性能

手法	適合率	再現率	F値	(一致数; 出力数)
ベースライン	0.127	1.000	0.225	(154; 1,216)
提案手法	0.467	0.695	0.559	(107; 229)

※ 正解データ数は154

表4に、関連行動(体験表現文+前後文)の抽出結果のうち、正解とするものと不正解とするものをそれぞれ示す。正解文は、ブルーベリー狩りというテーマに関連した行動となっており、不正解文は、ブルーベリー狩りの日とは別の日であったり、コメントへの返信であったりと、テーマに関連のない行動となっている。

表4: 関連行動の抽出例

正解	<ul style="list-style-type: none"> 帰宅後市販のスポンジケーキに生クリームをホイップしてブルーベリーをたっぷりデコレーションして(^w^)紫が涼しげなブルーベリーケーキを作りました。 この旗が目印ブルーベリーにも、沢山の種類があって、それぞれ味が違うので試食をして気に入った品種を好きなだけ摘み取ります♪
不正解	<ul style="list-style-type: none"> が活けてあり、注文した飲み物の茶碗やカップは自分で選べます。私が選んだのは、この季節ならではの(ギリギリ!?)のほうぎの平茶碗にしました 要は、タンスは信じるなってことですね。で、ストーリーのほうはさっき説明したとおりなんですが、主人公たちは、しょっぱなからはぐれちゃうんですよ。

3.3.2 関連行動の推移分析支援への評価

関連行動は、テーマの記述の前と後に分類される。本実験の結果、「ブルーベリー狩り」の前に行なう行動として、23件が得られ、「ブルーベリー狩り」の後に行なう行動として、206件が得られた。本実験で得たブログ記事には、ブルーベリー狩りの後の体験談が多く記述されているといえる。

repeated bisection法でオプション(--idf -n 4)を指定し、「ブルーベリー狩り」の前後でそれぞれ、4つクラスタに分割した。表5に「ブルーベリー狩り」の前の関連行動のクラスタを、表6に「ブルーベリー狩り」の後の関連行動のクラスタをそれぞれ示す。例文は各クラスタの一部であり、各3つずつ関連行動を表示している。キーワードは、KeyGraphを読みとる際に注目した一部分である。またクラスタを手で分析した結果を添える。

表5: 「ブルーベリー狩り」の前の関連行動のクラスタ

#	例文/主なキーワード/クラスタの人手分析結果
1	<ul style="list-style-type: none"> 今日は剣道で朝練がありました♪8時~3時間弱でしたが、ミッチリ搾られました(*≧▽≦*) 今日は横浜にある森農園さんに遊びにきましたー(o^^o)♪ HELLO!今日は朝からRiが旦那実家に遊びに行ったよなので、お迎えまでママタイム~ /今日 /ブログを書いた当日に行っていたこと
2	<ul style="list-style-type: none"> 本当に久しぶりに、本職...いや本道の温泉&食のブログを書こうと思うよ。こちらに伺ったのは7月だったわ。 えー！と思ったらね...無農薬のブルーベリーを食して、本当に何にもないけどそれがいい! 大好きだったおじさんの一回忌でした。にこちゃんが生まれて1週間程で亡くなったのにこちゃんのお誕生日が来るとおじさんの命日が来ます。 /なし /不明
3	<ul style="list-style-type: none"> この時期、窓を開けたときの秋のかおり好きですさてさて、溜まりまくった8月ネタをチョコチョコ更新していきます☆ 将来は、紫陽花農家になりたい私です。手作りのドライ紫陽花にしてみました。作り方は、自己流ですが簡単でした。 先日UPしました ブルーベリー狩りとBQQの続きです。 /なし /ブログの出だし文句など
4	<ul style="list-style-type: none"> あつという間に9月になっちゃいました(*^_^*)ちよつと前に久々天神に行くともうすっかり秋モード。 あー！私長い夏休みが終わってしまいました。昨日、鹿児島から大阪へ帰ってきました。 12月~4月中旬は冬季休業です車をどんどんこ進めて行くと、空もどんどんこ晴れてきて、現地に着いたときには綺麗な青空が広がっていたわ。 /月、終わり /時期のこと

表5を見ると、全体的にブログの出だしに書く出来事が多かった。特に、クラスタ1と4には時期等のことが書かれており、当日に何を行っていたかが分析できる。

表6を見ると、多少は異なる関連行動が分類されていたが、全てのクラスタにおいて特定の偏りが見られ、クラスタリング前より容易に分析が行えた。また、KeyGraphの結果から注目したキーワードも、分析結果と類似したキーワードが得られた。

表 6: 「ブルーベリー狩り」の後の関連行動のクラスタ

#	例文/主なキーワード/クラスタの人手分析結果
1	<ul style="list-style-type: none"> ・帰って一日中、目の下にいたもんでグッタリしてたけど少しお試してジャム作ってみました♪なかなか上手くてくれたのでヨーグルトとかパンにつけて食べたいと思います☆ ・大きなタッパ4つ分帰ってから早速旦那様に料理はまかせ、鶏肉のブルーベリー煮込みとワインで乾杯しました。 ・自分で採ったブルーベリーを買い上げたわ。結構な量があったから「余ったらジャムにしようかしら (*´O´*)」 ／ブルーベリー, 作る, 食べる, 採る, ケーキ, ジャム ／収穫したこと, 何かを作った (料理した) こと
2	<ul style="list-style-type: none"> ・途中で、松崎がアイス買ってきてくれたり、はなちゃんっていう猫探し始めたり、楽しすぎたあハナに会えなかったけど、カメのあーちゃんには会えました。 ・だからなるべく熟してる実を採るほうが良いそうです。KOUちゃんも暑い中 ベビーカーで付き合ってくれました というか付き合わせた 爆 ・朝もぎの枝付き枝豆が1束300円で売っていて購入！その場で枝からもいっていいんだそうで、みんなでポチポチと採りました ／買う ／何かを購入したこと
3	<ul style="list-style-type: none"> ・食べ放題…さて初めてのブルーベリーはかなり気に入ったようものすごい量をものすごい勢いで食い荒らしていましたたっさんのブルーベリーを食べたのに ・日当たりのいい場所のブルーベリーは、いい色になっていて、甘くて本当に美味しかったわ。食べた食べた。 ・甘ずばっくって 美味しかった (-▽-) バクバク食べながら 二人で1.2Kg採ったよ。 ／ブルーベリー, 食べる ／何かを食べた, また食べた感想のこと
4	<ul style="list-style-type: none"> ・少し前になりますが、ブルーベリー狩りに、京丹後にあじわいの郷の手前にある、ファームガーデン空詩士 (ソラシド) さん到着しました 広大な敷地に、すごい数のブルーベリーの木 ・お抹茶が頂きます店内はカウンターとテーブル席で落ち着いた雰囲気 優しい雰囲気の店主の女性が迎えて下さりやさしい時間が過ぎそう ・お昼には、田上農園さんの駐車場をかりて、ホルモンや鶏チャンを囲みました☆皆でワイワイ楽しみました (*≧▽≦*) ／ブルーベリー, 農園 ／農園, その他の施設のこと

4 考察

関連行動抽出を行わずに、テーマ「ブルーベリー狩り」ブログ記事の全文をクラスタリングした場合での分析を試みた。表 7 にその結果を示す。

クラスタリングはできているが、体験したこと以外の記述が多く、さらに、一文単位であるため具体的な行動が理解しづらかった。そのため、分析に時間がかかってしまった。

したがって、本システムを用いる方が、明らかに短時間でブログ筆者の行動を分析できた。

5 おわりに

行動推移分析支援システムを使用して行動分析を行った結果、注目するテーマの前後の行動について、ある程度まとまりを見ることができた。今後はテーマを増やして、観光分析に活用しながら、システムを改良する。

謝辞 本研究は、科学研究費補助金 (若手研究 (B) :22700100) のもとで行いました。

表 7: 関連行動抽出を行わずにクラスタリングのみを行ったクラスタ

#	例文/主なキーワード/クラスタの人手分析結果
1	<ul style="list-style-type: none"> ・でツイッターもしていますのでフォローお願いします ・袋にたくさんとって持ち帰りにしました ／ブルーベリー ／何かをすること, したこと, したかったことなど
2	<ul style="list-style-type: none"> ・iPhone からの投稿 ・めっちゃ楽しかった ／携帯, 投稿, 楽しい, おいしい ／投稿した端末, 感じたこと
3	<ul style="list-style-type: none"> ・こんなにいらない…(笑) ・私に似たのね (@°▽°@)ノ ／なし ／顔文字, 括弧を使った表現
4	<ul style="list-style-type: none"> ・9/15 (木) 新ペアーレ四日市にて 13:00~授業「パンダ」(見学 OK) ・定休日 水・木 (4.5.9.10月) 無休 (6月) 水 (7.8月) ／月 ／日時, 時間, その他の数字
5	<ul style="list-style-type: none"> ・今日は、心地良い秋晴れですね ・今日は姫路から来られた家族と一緒に。 ／今日, 行く ／ブログを書いた当日のこと
6	<ul style="list-style-type: none"> ・ブルーベリー狩りとはとにかく楽しかった～! ・ブルーベリーって、ブドウくらいの大きさになるんだ。。 ／ブルーベリー, 狩り, 食べる, 行く, 楽しい ／ブルーベリー狩り, ブルーベリーのこと
7	<ul style="list-style-type: none"> ・あぐらをかいて食べるなんてもってのほか ・園内では、試食程度に食べるのは OK ですが食べ放題ではありません ／食べる, 思う, おいしい ／食べる, 食べたこと
8	<ul style="list-style-type: none"> ・」という情報は経験者の方のブログ等から知っていたのですが、本当にきつかったです ・昨年はその場で食べるだけだったけど、今回はお土産用にひとバック分摘んだんだ ／ブルーベリー, 食べる ／不明

参考文献

- [1] 池田佳代, 田邊勝義, 奥田英範: “体験表現を手がかりにしたブログの体験情報の抽出”, 電子情報通信学会, データ工学ワークショップ 2007(DEWS2007), A8-1, 2007.
- [2] 池田佳代, 田邊勝義, 奥田英範, 奥雅博: “ブログからの体験情報抽出”, 情報処理学会論文誌, Vol.49, No.2, pp.838-847, 2008.
- [3] 西原陽子, 佐藤圭太, 砂山渡: “出来事画像表現によるブログからの体験談獲得支援”, 日本知能情報ファジィ学会誌, Vol.20, No.5, pp.757-767, 2008.
- [4] 石野亜耶, 難波英嗣, 竹澤寿幸: “旅行ブログエントリーからの観光情報の自動抽出”, 日本知能情報ファジィ学会誌, Vol.22, No.6, pp.667-679, 2010.
- [5] 高野太希, 井上潮: “文章構造に基づいたブログからの体験情報抽出方法”, 電子情報通信学会, データ工学と情報マネジメントに関するフォーラム 2011(DEIM2011), A4-2, 2011.
- [6] Bayon - a simple and fast clustering tool - Google Project Hosting
<http://code.google.com/p/Bayon/>
- [7] 大澤幸生, ベンソン ネルス E, 谷内田正彦: KeyGraph : 語の共起グラフの分割・統合によるキーワード抽出, 電子情報通信学会論文誌, D-1, Vol.82, No.2, pp.391-400, 1999.
- [8] 串間宗夫, モリス ノア, 荒木賢二, 鈴木斎王: 口蹄疫問題についての新聞報道テキストデータマイニング, 電子情報通信学会技術研究報告, 思考と言語, Vol.111, No.170, pp.7-12, 2011.
- [9] 柴田武, 山田進: 類語大辞典, 講談社, 2002.