

概要

近年、体験型観光が注目されている。鳥取県においては、農業と観光の連携がその一つであり、たとえばフルーツ狩りが挙げられる。しかし、フルーツ狩りそのものにかかる時間は長くはない。フルーツ狩りの前後の時間帯にも何かしらの関連する観光イベントがあれば、旅行のボリュームが増し、旅行者の満足度が高まると予想される。ゆえに、フルーツ狩りなどを基本テーマとして、それに関連する行動を多くの観光事例から分析することが必要となる。

ここで、旅日記のブログ記事に注目すると、個人の体験談や旅行の感想が多々記述されており、旅行者の行動分析の対象となるといえる。しかし、行動分析の際、体験談以外の記述、たとえば、ブログ友達への近況報告などは、不必要である。そこで、ブログ記事から体験情報のみを取り出すことができれば、行動分析が容易になると考えられる。

先行研究によると、観光開発の発想支援としてブログ文の自動抽出の方法が示されている [1]。しかし、その方法では、行動以外の情報を参照して観光地の分析を行っているため、本研究で焦点をあてている分析とは異なる。本研究では、行動の情報を得ること、その種類の把握、および、基本となるテーマとの前後関係を分析したい。そこで本研究では、まず、体験表現の抽出技術を用いることにする。次に、類似行動を分類するため、クラスタリング手法を導入する。最後に、クラスタの内容を概観するために、キーワード抽出手法も導入する。

以上について、入力テーマに関連する行動を分析する「行動推移分析支援システム」を開発した。その結果、関連行動を、適合率 0.47、再現率 0.69、F 値 0.56 という性能で抽出ができた。本システムの行動抽出を使用せずに、クラスタリングとキーワード抽出のみを使用した場合と比較すると、本抽出使用時の方が unnecessary な文が除去されており、分析が行いやすかった。したがって、観光事例から旅行者の行動を分析する有効な手法が開発できたといえる。

目次

第1章	はじめに	1
1.1	背景	1
1.2	関連研究	2
1.2.1	体験情報抽出	2
1.2.2	クラスタリング	2
1.2.3	キーワード抽出	3
1.3	本研究の目的	4
第2章	行動推移分析支援システム	5
2.1	構造	5
2.2	体験表現辞書	6
2.3	ブログ収集部	8
2.4	関連行動抽出部	9
2.5	クラスタリング部	10
第3章	実験	11
3.1	実験条件	11
3.2	実験手順	11
3.3	実験結果	12
3.3.1	体験表現文抽出の評価	12
3.3.2	関連行動の推移分析支援への評価	13
3.3.3	比較実験	16
第4章	おわりに	18

目 次

2.1	行動推移分析支援システムの構成	6
2.2	体験表現辞書の作成	7
2.3	辞書生成部	8
2.4	ブログ収集部	8
2.5	関連行動抽出部	9
2.6	クラスタリング部	10

表 目 次

2.1	体験表現抽出ルール	6
2.2	体験表現辞書の例	7
3.1	体験表現文の抽出性能	12
3.2	関連行動の抽出例	12
3.3	「ブルーベリー狩り」の前の関連行動のクラスタ	14
3.4	「ブルーベリー狩り」の後の関連行動のクラスタ	15
3.5	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	17

第1章 はじめに

1.1 背景

近年、体験型の観光が注目されている。体験型観光は、旅行者が旅行先で見たり聞いたりするだけではなく、旅行者が五感を通じて行動したり、人々と協同で行動したりするという観光である。たとえば、手打ち蕎麦体験やシーカヤック体験などがある。鳥取県では、昨年度、『農商工こらぼフォーラム～「農業」と「観光」の連携を考える～』が開催された。農商工こらぼフォーラムとは、都市住民のグリーンツーリズムや農業体験への関心が高まる中、観光農業という新しい視点を農業経営に加えることで、生産物の販路拡大や収益向上に繋げる可能性を考えるための取り組みで、「これからは農業と観光だ!」、「農業と観光の連携方法を考える」というテーマで講演がされた。講演の中では、「いちご狩り」や「びわ狩り」、「ブルーベリー狩り」などが話題に挙がっていた。

鳥取県の農業に関連した観光開発には発想支援技術を導入する余地がある。すなわち、上述のフォーラムでは、梨園・いちご園・ブルーベリー園などの農業事業者や、旅行エージェントらが多くアイデアを必要としていた。とりわけ、フルーツ狩りに注目してみると、フルーツ狩りそのものは、一時間程度で終了することがあるため、フルーツ狩りだけでは、旅行としてのボリュームが不足すると予想される。より満足度の高い観光にするためには、フルーツ狩りの前後にも観光イベントを充実させる必要があるだろう。このように基本テーマとしてフルーツ狩りを想定すると、そのテーマに関連する体験的な行動を、観光事例から分析する必要があり、そうした分析を支援する情報処理技術に期待が寄せられるだろう。

ここで、観光事例としては、旅日記の書かれたブログ記事が挙げられる。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。しかし、行動分析を行なう際、体験談と感想以外の記述はノイズとなってしまうため不必要である。そのため、ブログから体験情報のみを取り出すことができれば、行動分析が容易になる。

先行研究によると、観光開発の発想支援としてブログ文の自動抽出の方法が示されて

いる [1]。しかし、その方法では、行動以外の情報を参照して観光地の分析を行っているため、本研究で焦点をあてている分析とは異なる。本研究では、行動の情報を得ること、その種類の把握、および、基本となるテーマとの前後関係を分析したい。そこで本研究では、まず、体験表現の抽出技術を用いることにする。そして、類似行動を分類するため、クラスタリング手法を導入する。さらに、クラスタの内容を概観するために、キーワード抽出手法も導入する。

1.2 関連研究

1.2.1 体験情報抽出

体験情報とは、実際に見た対象物や、何かを体験した際の感想などの、客観的あるいは主観的な情報である。ブログや日記の文章においては、筆者の体験した事柄が記述されている部分から得られることができる。具体的な体験情報としては、「行きました」、「食べたことがあります」などの筆者自身が行った表現や、「楽しめました」、「良かった」などの感想がある。

池田らは、自発的な動作を表す動詞および名詞など、体験を記述する際に現れる体験表現に着目した。ブログに体験情報が含まれているかを体験表現らしさ（スコア）でソートして提示する「体験情報抽出システム」を作成した [2],[3]。

西原らは、出来事の記述には場所・動作・対象物が用いられていることに着目した。出来事が記述されているブログ記事には動詞の過去形が含まれているとして、動詞の過去形を含んだブログ記事内の語を手がかりに、場所・動作・対象物を画像で表現した [4]。

石野らは、表層パターン手法と機械学習を用いて、観光情報リンク集の構築を行った。構築は、旅行ブログエントリから抽出した土産物情報・観光名所情報と、旅行ブログエントリ中に存在するリンクを分類・収集して行った [5]。

高野らは、ブログ内の文章に対して構造解析を行なうことによって体験情報を抽出した。場の雰囲気や感想などの、実体験を示す語に着目して、対象のブログがイベントの体験表現を含むかを判別した [6]。

1.2.2 クラスタリング

クラスタリングとは、幾つもの事例のうち、類似する複数の事例をまとめて1つのクラスタを形成し、また異なる事例から別のクラスタを形成するということを繰り返す、事

例集を分割することである。ここで、1つの事例が複数のクラスタに属してよいクラスタリングのことを、ソフトクラスタリングという。一方、1つのクラスタだけに限定することをハードクラスタリングという。

ハードクラスタリングとしてよく用いられる手法は、k-means法である。これは、得られるクラスタ数をあらかじめ指定しておき、クラスタに属する事例から重心を求め、全ての事例について、事例の属するクラスタの重心とその事例との距離の平均値が最小になるように、事例の属するクラスタを求めるものである。

k-means法の応用手法として、repeated bisection法がある。この方法は、まず、すべての事例を1つのクラスタに格納し、2分割のk-means法クラスタリングを繰り返し実行することで、事例集を分割するものである。この手法は、k-means法などと比較して、高速に実行でき、また精度も良好である。

本研究では、クラスタリングに、bayon[7]を用いる。bayonはrepeated bisection法を採用しているクラスタリングツールであり、シンプルな構成で、かつ大規模なデータでも実用的なスピードで実行できることを目標に作成されている。

1.2.3 キーワード抽出

キーワード抽出とは、文章を構成する要素一つ一つの登場頻度、互いの繋がりや数などを計算して、重要な役割を果たしている要素を分類・抽出することである。文書に含まれる単語の特徴度は、TF-IDFやBM25により算出されることがある。しかし、観光分析のための行動のキーワードを得るためには、「何かを、どうにかする」というように、対象と行動の組が得られる方が分かりやすいと予想した。1単語ごとの特徴度よりも単語の共起に着目する抽出法が妥当と予想した。

ここで、意見の分析において、KeyGraph[8]を用いた研究事例がある[9]。KeyGraphでは、文書における、単語の分類、および、単語間の関係について「土台」、「屋根」、「柱」という概念を導入した。これらの概念は、文書中の文単位での単語の共起の仕方により定義される。KeyGraphは、これらの概念をグラフ（辺とノード）で表示することで、さらに、分かりやすく文書の構造を表現することに使われる。

ただし、KeyGraphに基づく分析とグラフの表示は、ライセンスの下で使用するソフトウェアを必要とする[10]。本研究では、文献[8]に示された計算式に基づきグラフ（辺とノード）の生成を実装し、グラフの可視化は、graphvizライブラリのdotコマンド[11]を使用する。

1.3 本研究の目的

本研究では、まず、あるテーマに関連する行動をブログ記事から抽出する「行動推移分析支援システム」を構築する。システム構築では、前述の体験表現の抽出法、クラスタリング手法、および、キーワード抽出法を実装する。

次に、本システムを用いた実験では、「ブルーベリー狩り」をテーマとして取り上げて、関連する行動の分析を行い、分析支援の可能性を検討する。検討の方法は、体験表現の抽出を行う場合と行わない場合のそれぞれで、クラスタリングとキーワード抽出を行い、得られた結果を比較するというものである。クラスタリング結果から人手でクラスタの行動を要約し、その結果が、抽出されたキーワードと類似していれば、分析支援に効果がありうると判定する。

第2章 行動推移分析支援システム

本章では、まず、行動推移分析支援システムの構造を概説し、次に、個々の処理部の詳細および実装を説明する。

2.1 構造

行動推移分析支援システムは、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 2.1 にシステムの構成を示す。

体験表現辞書は、体験表現の集合である。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(これは「体験表現文」である)および前後一文を「関連行動」とみなして抽出する。「関連行動」は、本文中でテーマの記述された文との位置関係により、前か後かに分類する。次に、関連行動を文ベクトル化によりクラスタリングを行なう。最後に、KeyGraph[8]を用いて関連行動のキーワードを抽出する。

この支援を受けて、分析者は旅行者の行動を分析する。

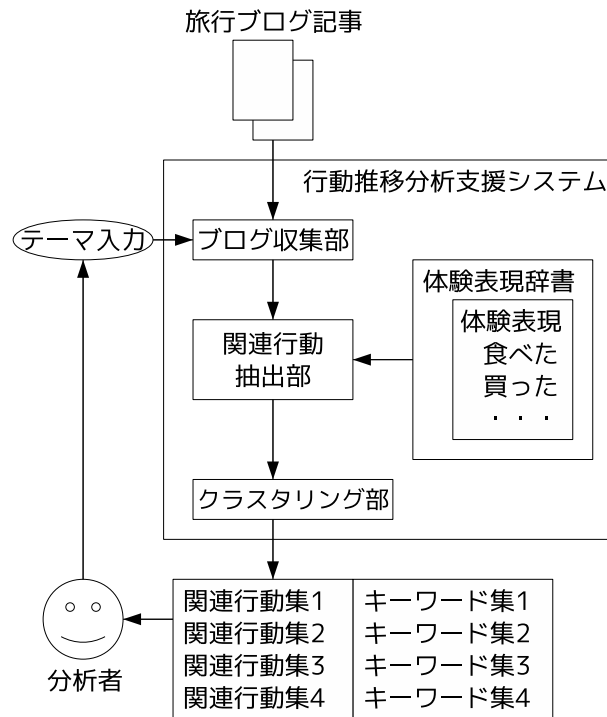


図 2.1: 行動推移分析支援システムの構成

2.2 体験表現辞書

体験表現辞書は、体験表現抽出ルールを用いて、一般ブログから自動生成する。図 2.2 に体験表現辞書生成の流れを示す。

体験表現抽出ルールは、文献 [2],[3] を参考にし、また、一般ブログを観察することで、表 2.1 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 2.1: 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」

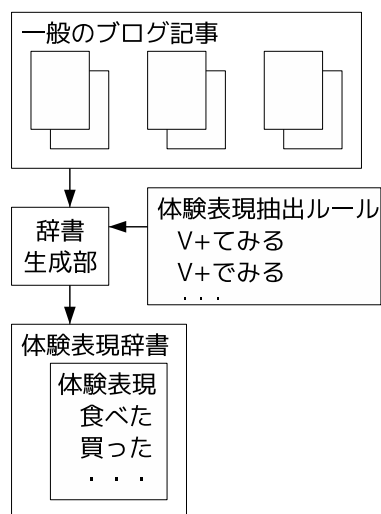


図 2.2: 体験表現辞書の作成

一般のブログ記事（5億4,700万文）を形態素解析したもの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 2.2 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 2.2: 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, …

まず、一般のブログ記事を形態素解析する。解析された一般ブログ記事と、体験表現ルールとのマッチングにより、マッチした語を抽出する。マッチした語の末尾を、動詞の過去形や「ですます調」に変形し、複数の体験表現を作成する。作成した体験表現を辞書型に格納し、体験表現辞書としてファイル出力する。

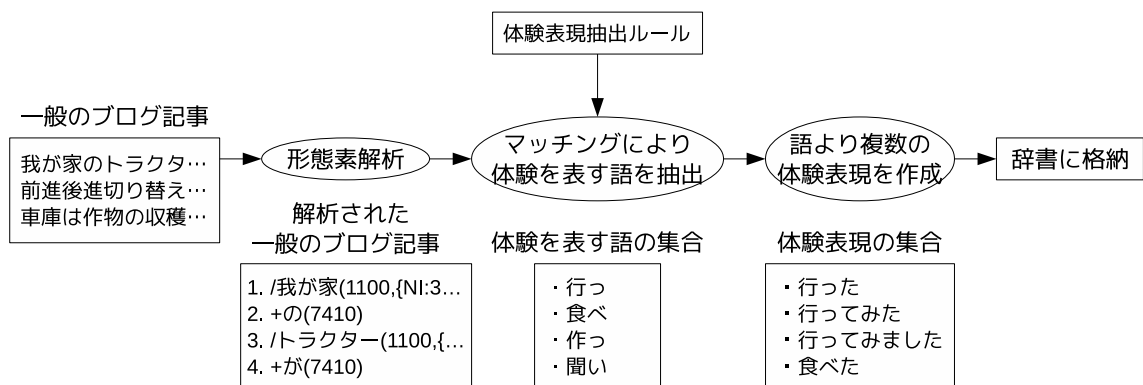


図 2.3: 辞書生成部

2.3 ブログ収集部

ブログ収集部（図 2.4）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。

分析者より入力されたテーマを用いて、ブログ検索サイトの検索を行なう。ブログ検索の結果より、テーマに関するブログ記事の URL を取得し、URL リストを作成する。その後、取得した URL のブログ記事を解析して、記事の本文部分のみを行ごとに取得する。URL リストの全てのブログ記事本文を格納した出力データを作成する。

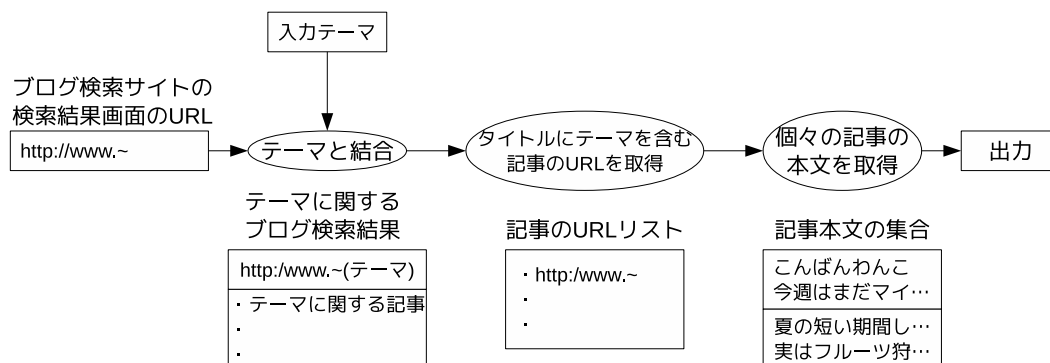


図 2.4: ブログ収集部

2.4 関連行動抽出部

関連行動抽出部（図 2.5）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．

ブログ収集部より受けとったデータと，体験表現辞書とのマッチングを行い，体験表現文を取得する．取得した体験表現文に，前後一文を結合して，関連行動を作成する．関連行動は，ブログ記事中のテーマの記述より前から取得したか，後から取得したかで，別の出力データを作成する．

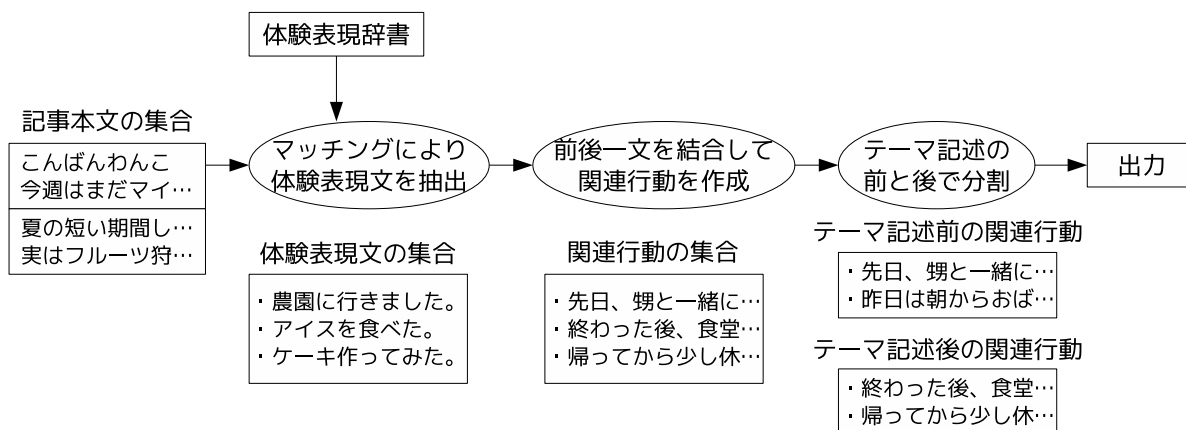


図 2.5: 関連行動抽出部

2.5 クラスタリング部

クラスタリング部（図 2.6）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 [7] を用いる。bayon 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 [8] 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 [9] で使われていることから、試行する。

関連行動抽出部より受けとったデータを、クラスタリングのためにベクトル化する。クラスタリング時のベクトル化では、素性として、名詞・動詞の単語 uni-gram, および、意味属性コード [12] を用いる。次に、ベクトル化された関連行動集を bayon を使用して、8 つにクラスタリングする。作成された 8 つのクラスタに対して KeyGraph を使用し、各クラスタごとのネットワーク図を作成する。最後に、クラスタリングされた関連行動集とネットワーク図を出力する。

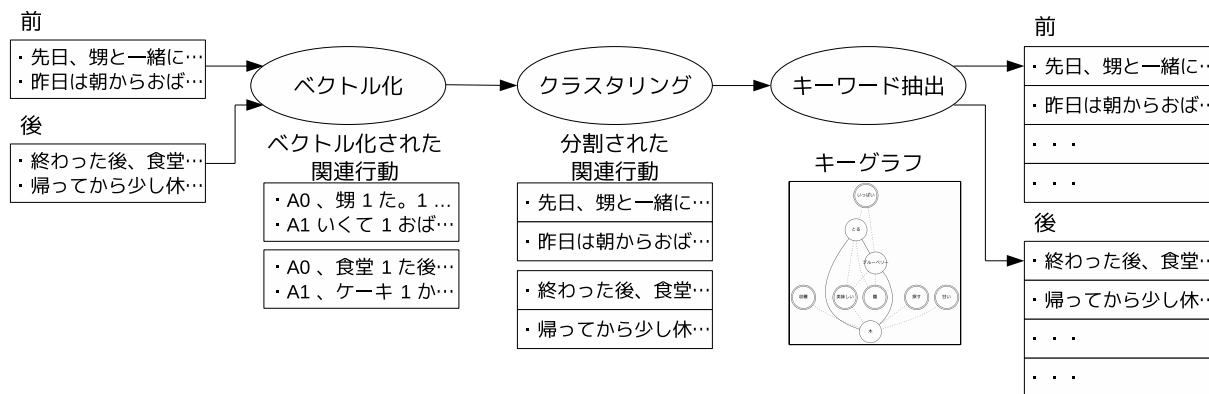


図 2.6: クラスタリング部

第3章 実験

「ブルーベリー狩り」の観光開発を想定して、ブルーベリー狩りの前後での行動を分析する。

3.1 実験条件

テーマは「ブルーベリー狩り」とする。ブログ検索サイトは「Ameba ブログ」を使用する。評価実験のために、人手で抽出すべき体験表現文に正解フラグを入れる。体験表現辞書 8,089 語で実験を行なう。キーワード抽出には、KeyGraph を用いる。

3.2 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なう。

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

手順6 各クラスター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。

3.3 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、107文が正解と一致した。

正解データと本システムの比較を表3.1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 3.1: 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3.2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 3.2: 関連行動の抽出例

正解	<ul style="list-style-type: none">・帰宅后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分だけ摘み取ります♪
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(ギリギリ!?)のほうずき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けど、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した．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．

「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．「ブルーベリー狩り」の前の関連行動のクラスタを表3.3に，「ブルーベリー狩り」の後の関連行動のクラスタを表3.4にそれぞれ示す．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，クラスタを人手で分析した結果を添える．

表 3.3: 「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～ <p>／今日 ／ブログを書いた当日に行っていたこと</p>
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のよ。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で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。 <p>／なし ／不明</p>
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きてきて、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。 <p>／なし ／ブログの出だし文句など</p>
4	<ul style="list-style-type: none"> ・あっという間に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どこ進めて行くと、空もどどこ晴れてきて、現地に着いたときには綺麗な青空が広がっていたわ。 <p>／月、終わり ／時期のこと</p>

表 3.3 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 3.4: 「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，日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♪なかなか上手くできたのでヨーグルトとかパンにつけて食べたいと思いまーす☆ ・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，楽しすぎでし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बीカーで付き合ってくれました といつか付き合わせたい 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で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く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。 ・甘ずぱくって 美味しかった (-^▽^-) パクパク食べながら 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土(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 で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せそう ・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

表 3.4 を見ると、何かを作った (料理した) ことや何かを購入したこと、また農園とその他施設のことなど、全てのクラスタにおいて特定の偏りが見られた。

「ブルーベリー狩り」の後の関連行動のクラスタには、分析価値のある関連行動が分類されていた。また、この分析は、KeyGraph で生成したキーワードと類似していた。このことは、KeyGraph の出力を分析することで、容易にクラスタの内容が分析できることを意味する。ゆえに、関連行動の推移分析支援に有効な結果が得られているといえる。

3.3.3 比較実験

比較のために、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 3.5 にその結果を示す。

この結果、クラスタリングはできているものの、体験したこと以外の不必要な記述が多く、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。このことから、関連行動抽出を行った場合の方が、関連行動抽出を行わない場合よりも分析の手間が軽減できることが分かる。しかし、関連行動の抽出において、テーマに関連していない行動も少量ながら含まれている。これは、同一ブログ記事中に複数の出来事が記述されている際に、異なる出来事でも関連する行動として抽出してしまうためであり、この問題を解決するためには、ブログ記事中の話題を変えるときの変換を判定したり、また、日時を表す表現を判定するようにすることが考えられる。

以上より本システムは、改善の余地があるものの、分析支援効果があるといえる。

表 3.5: 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"> ・ こんなにいない…(笑) ・ 私に似たのねゞ (@°▽°@)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, 心地良い秋晴れですね ・ 今日, 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
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 ・ 去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パック分摘んだんだ ／ブルーベリー, 食べる ／不明

第4章 おわりに

近年、体験型観光の関心が高まっていることに注目し、鳥取県の農業に関連した観光開発の発想支援のニーズがあると考えた。そこで本研究では、フルーツ狩りに着目して、その前後に観光イベントや娯楽を充実させることで、より満足度の高い観光になるのではないかと予想した。「ブルーベリー狩り」をテーマと想定して、そのテーマに関連する体験的な行動を、観光事例から分析する「行動推移分析支援システム」を開発した。

関連行動を、適合率 0.47, 再現率 0.69, F 値 0.56 という性能で抽出ができた。この結果は、関連行動抽出を行わない場合と比較して、不必要な文が除去されており、分析する際の手間を軽減できた。また、クラスタリング手法及びキーワード抽出を用いて類似行動を分類した。この結果、分類結果から分析が行えるようになり、分析する際の手間を軽減できた。

今後は、関連行動の抽出精度を向上させるとともに、他のテーマでも実験を試み、どのような分析結果が得られるのかを調査していきたい。

謝辞

本研究を進めるにあたり，数々の御助言を頂きました鳥取大学大学院工学研究科情報エレクトロニクス専攻知能情報高額コースの村田真樹教授に心から御礼申し上げます。本研究を進めるにあたり，御指導を頂きました村上仁一准教授に心から御礼申し上げます。徳久雅人講師には，終始に渡り研究の進め方や本論文の書き方など，細部にわたる御指導を頂きました。ここに深く感謝いたします。

その他様々な場面で御助言を頂いた計算機工学講座C研究室の皆様に感謝の意を表します。

参考文献

- [1] 謝花博:“能動学習の利用による観光分析精度の向上”, 鳥取大学卒業論文発表会, 2012.
- [2] 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.
- [3] 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.
- [4] 西原陽子, 佐藤圭太, 砂山渡:“出来事の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.
- [5] 石野亜耶, 難波英嗣, 竹澤寿幸:“旅行ブログエントリ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.
- [6] 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.
- [7] bayon - a simple and fast clustering tool - Google Project Hosting
<http://code.google.com/p/bayon/>
- [8] 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-I, Vol.82, No.2, pp.391-400, 1999.
- [9] 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.
- [10] KeyGraph
<http://www2.kke.co.jp/keygraph/>

[11] Graphviz - Graph Visualization Software
<http://www.graphviz.org/>

[12] 柴田武, 山田進: 類語大辞典, 講談社, 2002.