

街 Blog からの体験抽出とその空間的提示手法の提案

倉島 健† 手塚 太郎† 田中 克己†

人間の行動は時間的・空間的要因によって規定されている。Blog の普及により、地域を実際に訪れた個人の情報発信が活発になり、さらには記述された日時が記録されているという、いわゆる街 Blog の特性によって、このような街の訪問者による体験や評価情報が得られるようになった。本研究においては、ある場所について書かれた個々のテキストから、人々の体験を、時間・空間・動作・対象属性間の相関ルールマイニングによって抽出する手法を提案する。そして、ユーザがそれらの属性を指定することで、抽出した体験を柔軟に検索・要約することのできるシステムの提案を行う。これにより、容易にその地域における人々の行動を把握することが可能となる。

Blog Map of Experiences: Extraction and Geographical Mapping of Visitor Experiences from Urban Blogs

TAKESHI KURASHIMA,† TARO TEZUKA†
and KATSUMI TANAKA†

The prevalence of Blogs has enabled observation of the personal experiences in a certain location and time. Such information were traditionally unavailable except indirectly through local newspapers and periodicals. This paper proposes a method of obtaining spatially-specific experiences of urban visitors, for example, visitors' activities at several sight-seeing spots and their evaluations, by extracting association rules from the contents of Blog articles. By geographical mapping of Blog articles, the proposed system enables users to observe visitors' real activities and impressions of their visiting places, which are often more diverse than the guidebooks and more trustworthy than the advertisements.

1. はじめに

人間の行動は時間的・空間的要因によって規定されている。春には多くの人が花を見に行くであろうし、地域特有の食べ物が存在すれば旅行者がそれを食べるであろう。Blog の普及により、地域を実際に体験した個人の情報発信が活発になり、さらには、記述された日時が記録されているという Blog の特性によって、このような人間の動きに関する情報が得られるようになった。

従来、このような人間の動きに関する情報は、地方紙や定期刊行物などから間接的ともいえる手法でしか得ることができず、そのすべてを把握することは不可能であった。Blog から直接的に得られる個人の体験は、その地域を訪れようと考えている潜在的な訪問者や、地域の流行に興味のあるマーケットアナリストにとって有用であるといえる。

現在、ある場所を実際に訪れて書かれた体験記のような Blog 文書を、位置情報に基づいてユーザが投稿するシステムが存在する¹⁾。しかし、それらのシステムが広く普及しているとは言いがたく、体験型 Blog の多くは Blog ホスティングサービス上で、単に一人の蓄積情報として扱われているのが現状である。さらに、Blog ホスティングサービスが提供するテキスト検索機能で「清水寺」などの地名をキーワードとした検索結果は大量であり、さらにはそのすべてが「清水寺」を実際に訪れて書いた体験記とは限らない。このように、ある地域に関して書かれた体験記 Blog が Web 上に分散して存在し、さらにはその量が膨大であるという現状においては、すべての Blog に目を通すことは困難であるといえる。その一部に目を通しただけでは偏った見解に陥ってしまいかねない。

本研究においては、ある場所について書かれた個々のテキストから、人々の体験を、時間・空間・動作・対象属性間の相関ルールマイニングによって抽出する手法を提案する。そして、本手法によって得られた体験を、それらの属性を指定することで検索・要約する

† 京都大学大学院情報学研究所
Graduate School of Informatics, Kyoto University

システムの提案を行う。これにより、容易にその地域における人々の行動を把握することが可能となる。

本稿の第2章では、本手法に関連する基本的事項や関連研究について述べ、3章では本手法の詳細について述べる。4章では、抽出した体験をユーザに提示するシステムの概要を述べ、5章ではシステムの操作、6章では実験的に体験を抽出した結果、7章ではまとめを述べる。

2. 関連研究

2.1 Blog マイニングに関する研究

奥村²⁾らは Blog を収集・監視し、収集した Blog をマイニングすることで得られた情報を閲覧することが出来る「blogWatcher」を提案している。Blog をマイニングするという点では本研究と共通しているが、blogWatcher は汎用的なシステムであり、話題という枠組みで抽出を行っているため、抽出されるキーワードは主に名詞である。本研究においては、抽出する語を具体的に体験とし、地域という単位で情報の統合を行っているという点で異なるといえる。

Kumar らは Blog 中のリンク構造について議論を行い、Blog コミュニティの拡大とコミュニティ間の変化を追跡するために *time graph* と *time-dense community tracking* を提案している³⁾⁴⁾。Bar-Ilan らは Blog 中のリンクと投稿を統計的に分析した。しかしながら、この分析は Blog 記事本文を対象としたものではない。また、Bar-Ilan は Blog が Blogger にとって興味深い Web サイトに対するリンクを持つという伝統的な側面に加えて、Blog 上における個人情報と自己表現がますます重要な側面になってきていると指摘している⁵⁾。

2.2 街 Blog に関する研究

上松らは Blog と、カメラや GPS といった機能が付加された携帯電話端末を用いて、各人が発信した情報を位置情報に基づいて整理し、新たな情報閲覧を可能にするシステム「場 log」を提案している。情報閲覧方法としては、指定した地点から近いデータを時系列順、距離の近い順で提示する「場所 Blog」と、データベースに登録された画像を、位置情報をもとに地図上に貼り付けて提示する「blog 地図」がある。また、場 log においては Blog ツールが生成する RSS に位置情報を付加するという方法で位置情報を取得している。場 log は空間という単位で Blog を投稿する仕組みを提案したものであるが、本研究はこのようにして収集された大量の Blog データから、マイニングによって情報集約を行おうというものである。また Hurst⁶⁾ら

は、Blog を自動的に GIS 上にマッピングし、複数の Blog ホスティングサービス間の相違点を発見する手法を提案している。

2.3 相関ルール分析

本章では、相関ルールについて説明する。相関ルール分析は、ある商品 A を買ったら他の商品 B も同時に買うというような暗黙的なルールをマイニングする手法である。相関ルール分析では、アイテムの集合であるトランザクションを対象にして、各トランザクションの中に出現するアイテム組に見られる共起パターンを発見する。相関ルールは以下のように表現できる。

$$\begin{array}{cc} X & Y \\ X: \text{条件部} & \\ Y: \text{結論部} & \end{array}$$

条件部と結論部は、複数のアイテムであってもよい。この相関ルールは、もし条件部のすべてのアイテムがトランザクション中に現れれば、そのトランザクション中には結論部のすべてのアイテムが現れるという意味となる。

次に、支持度と確信度という概念について説明する。データベース D 中においてアイテム集合 X と Y をともに含むトランザクションの全トランザクションに対する割合を支持度 (support value) といい、 $\text{sup}(X \ Y)$ と表記する。また、D 中の X を含むトランザクションのうち、Y を含むものの割合を確信度 (confidence value) といい、 $\text{conf}(X \ Y)$ と表記する。この支持度と確信度をもとにして、相関ルール分析を行う手法が Agrawal⁷⁾らが提案した Apriori アルゴリズムである。このアルゴリズムにおいては、ユーザが確信度 (最小確信度: minimum confidence value) と支持度 (最小サポート: minimum support value) の閾値を設定し、閾値以上の確信度と支持度をもつ相関ルールを重要な相関ルールとして抽出する。これにより、従来、大量のデータを対象としたとき、すべてのルールを導出するのは多くの計算時間が必要であったが、その問題を克服することが可能となった。

3. 体験抽出

本章では、人々の体験抽出手法について説明する。体験とは個々人が「実際に物事を行うこと」である。人々が何か体験をして Blog を記述する場合、必ず「何をしたか」という行為の内容を記述する。しかし、当然のことながら Blog 中に出現するすべての文がそのような人々の行為を意味するものではない。体験を通して得られた感想や知識、あるいはまったく関係のない話題について書かれた文も含まれている。Blog が

ら人々の体験を抽出するためには、行為を意味する文を抽出する必要があるといえる。

一般的に、文は次の3つの種類に分類することができる。

- する文（行為）
（例）紅葉を拝観する
- なる文（過程）
（例）イベントが延期になる
- である文（状態）
（例）桜が五分咲きである

人々の「体験」を示す表現は、この中の「する文（行為）」に該当すると考えることができる。これらの文はどのような動詞で終わっているかで自動的に分類することが可能である。そこで、本手法においては、動作動詞（見る、聞く、食べるなど）とサ変名詞に着目して体験抽出を行っている。サ変名詞は「サ変名詞＋する」のように用いられ、行為を示す動詞的な役割を果たす品詞である。以下にそのステップを示す。

Step1: Blog の収集

Step2: Blog 文書の形態素解析

Step3: 関連ルールの導出

Step4: 要約の作成

次章以降でそれぞれのステップの詳細について述べる。

3.1 Blog の収集

Blog の収集は、Blog ホスティングサービスが提供するテキスト検索機能を使用して行う。また、Blog ホスティングサービスは検索結果を RSS 形式で配信するサービスを行っており、この機能を利用して検索結果を取得する。RSS とは Blog ホスティングサービスが提供する RDF rich Site Summary であり、メタデータでタイトル、本文などの情報が整理されている。地域に関連する地名、ランドマーク、建物名などは既存の GIS から抽出してリスト化しておく。そして、検索エンジンにこれらの語を定期的にクエリとして投げる。RSS における Blog 情報は、タイトル、日付、本文などから構成されるが、本研究においては、地名（クエリ）、タイトル、リンク、本文、日付情報の組み合わせをデータベースに蓄積する。以下にそのステップを示す。

Step1: 地域に関連する地名、ランドマーク、建物名を登録

Step2: 地名を検索語として検索エンジンを呼び出す

Step3: 検索結果を RSS 形式で取得

Step4: RSS を解析し、タイトル、リンク、本文、

日付情報を抽出

Step5: Blog データベースに格納

Step6: 1 から 5 のステップを定期的に行う

3.2 アイテムデータベースの作成

収集した Blog を解析し、関連ルールを抽出するトランザクション集合を作成する。RSS で得られる Blog の本文情報は、タイトル、あるいは Blog 本文中に出現する検索語の周辺テキストである。本手法においては、得られた周辺テキストを文節に区切り、さらに単語ごとに区切る。詳細については4章で述べる。

本研究においては、以下に示す要素に基づいて関連ルールを適用する。

（日付，地名，名詞，動詞，サ変名詞）

ここで、名詞は、代名詞、数、非自立、特殊、接尾、接続詞的、動詞非自立、サ変名詞を除いたものである。以後、本論文中で名詞と呼ぶのはこのような語とする。また、地名は検索エンジンにクエリとして投げる語である。各トランザクションを T とすると、 T はアイテムの集合である。本手法においては、ある文に含まれる個々の語がアイテムであり、その集合がトランザクションとなる。ひとつの文にひとつのトランザクションが対応する。トランザクションは以下のように記述することができる。

$$T = (date, g, n_1, ..n_n, v_1, ..v_m, s_1, ..s_l)$$

T はトランザクション、 g はクエリとした地名、 $date$ は文を含む Blog の日付である。 n, v, s はそれぞれ、文中に出現する名詞、動詞、サ変名詞である。

例えば、「京都駅」というクエリを検索エンジンに投げて、2005年5月に投稿され「京都駅でおみやげと切符を買う」という文章を含む Blog が得られたとする。この場合、地名は「京都駅」、名詞は「おみやげ」と「切符」、そして動詞が「買う」という語になり、この文に対応するトランザクションは以下になる。

京都駅でおみやげと切符を買う

$T = (05052005, \text{京都駅}, \text{おみやげ}, \text{切符}, \text{買う})$

嵐山で紅葉を見てから夕食を食べた

$T = (05052005, \text{嵐山}, \text{紅葉}, \text{夕食}, \text{見る}, \text{食べる})$

3.3 関連ルールの抽出

関連ルールの抽出には、前章で説明した Apriori アルゴリズムを用いる。実世界においては、その場所に特有の遊び方が存在する。また、同じ場所においても、ある時期に特有な遊び方や行為が存在する。本研究においてはこのような「ある時期にこの場所を訪れた場合に、このようなことをする」というルールを Blog から発見する。本研究においては、以下に示す3つの

タイプに分類される関連ルールに基づいて体験抽出を行う。

Type1:

[時間, 地名] [動作動詞]
 [時間, 地名] [サ変名詞]

Type2:

[時間, 地名] [動作動詞, 名詞]
 [時間, 地名] [動作動詞, サ変名詞]
 [時間, 地名] [サ変名詞, 名詞]

Type3:

[時間, 地名, 動作動詞] [名詞]
 [時間, 地名, 動作動詞] [サ変名詞]
 [時間, 地名, サ変名詞] [名詞]

すべての関連ルールを作成するのではなく、上記のような品詞を条件部と結論部に持つルールのみを抽出する。このようにして得られた(動作動詞, 名詞)、(動作動詞, サ変名詞)(サ変名詞, 名詞)の組が体験である。また、一般的な動作動詞はストップワードとしている。例えば「行く」「来る」などの動詞である。これらは「京都に行く」などのように用いられるが、情報としてはその場所に Blogger が訪れたということだけであって、何をしたかという情報までは含んでいない。

Type3 においては、Type1 の関連ルール抽出後に得られた動詞、及びサ変名詞の対象となっている語を抽出している。これらの対象は一般的に名詞になる場合が多い。例えば「見る」という動詞に対応する「桜」という名詞である。また、本研究においては、サ変名詞も動作の対象となりうる品詞としている。これは例えば「紅葉を見る」というように、サ変名詞が「見る」という動作の対象となる場合もあるからである。

3.4 要約の作成

抽出した関連ルールは、ユーザからの入力、要求に応じて要約の形で提示を行う。その詳細については5章の中でシステムの使用例とともに説明する。

4. システム構成

前章までに説明したアルゴリズムで抽出した体験を提示するシステムについて述べる。システム処理の流れを以下に示す。

(1) Blog 収集

複数の Blog 検索エンジンに地名を投げて Blog を RSS 形式で収集する。本システムにおいては、goo ブログ⁸⁾、及び livedoor ブログ⁹⁾ を用いて実装を行った。また、検索語とする地名はゼンリンの住宅電子地図情報¹⁰⁾ から抽出したものであ

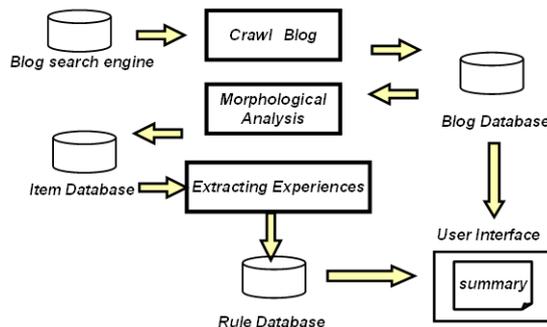


図 1 システムの構成

る。収集した Blog は Blog データベースに格納する。本システムにおけるデータベースはすべて MySQL¹¹⁾ を使用している。

(2) Blog 文書の形態素解析

Blog 文書のタイトル、及び本文から文節を抽出する。それぞれの文章を、日本語形態素解析プログラム茶筌¹²⁾ を用いて解析する。茶筌は文を語に区切り、それぞれの品詞を解析するプログラムである。得られた品詞に基づき、関連ルールを適用するトランザクションを作成する。トランザクションはアイテムデータベースに格納する。

(3) 関連ルールの抽出

アイテムデータベースのそれぞれのトランザクションに対して体験抽出を行う。関連ルールは過去1ヶ月、もしくは月ごとというように、ある程度多くのトランザクションが得られるような時間単位で抽出する。動作動詞の抽出は、日本語彙体系ソフト¹³⁾ のデータを利用して行う。抽出したルールはルールデータベースに格納する。

(4) ユーザインタフェース

ユーザは、地名、時間、動作、対象をクエリとして指定することで検索を行うことができる。ユーザ操作の詳細については次章で説明する。

5. システムの操作

5.1 ユーザの入力

抽出した関連ルールは、ユーザの要求に応じて要約の形で提示を行う。ユーザの入力は、4つの属性(地名、時間、動作、対象)のいずれかひとつ、もしくはその組み合わせである。検索フィールドに属性値をスペースで区切って入力する。時間については、関連ルールを月単位、あるいは「1ヶ月以内」というような単位で抽出しているため、リストから期間を選ぶ形式である。また、自然言語による検索も可能である。シス

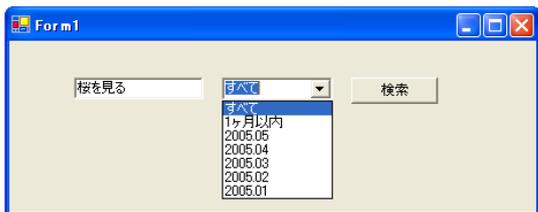


図 2 システムの検索画面

テムは自然言語によるクエリを、形態素解析ソフトを用いて解析し、助詞などの品詞の語を切り捨て、単語を抽出する。次章で、その典型的な使用例について述べる。

5.2 典型的な利用例

抽出した関連ルールは、ユーザの要求に応じて要約の形で提示を行う。典型的な要求例として以下のようなものが挙げられる。

Case1: クエリ = 空間, 時間 検索結果 = 動作, 対象

現在地をクエリとすることで、その場所で行うことができる体験を調べたいという場合にこのような検索が想定される。また、ユーザが旅行などで訪れる場所と期間が決定しており、その場所でどんな体験を行うことができるのかを調べたいという状況でもこのような検索は有効である。人間の体験は、四季という要因によって規定される場合が多く、一年前、同じ場所で人気のあった体験は、再び人気が出る可能性が高いからである。

(例1) クエリ: 京都 期間: 7月

(例2) クエリ: 嵐山 期間: すべて

システムはユーザの入力に対してルールデータベースの検索結果をもとに体験を提示する。この場合、表1のような検索結果となり(名詞, 動詞)(名詞, サ変名詞)(サ変名詞, 動詞)という組で構成される体験は、動作(動詞またはサ変名詞)に基づいて整理されてユーザに提示される。また、ユーザが興味をひかれた体験を指定すれば、システムは「空間」「時間」属性と選択した体験をもとに Blog データベース、もしくは Blog 検索エンジンにアクセスし、該当する Blog 文書を収集してくる。これにより、体験の詳細やその

表 1 空間, 時間をクエリとした場合の検索結果

Location name :	Date :
[rank] Verb1 :	(conf([Location name, Date Verb])
Object1 :	(conf(Location name, Date, Verb Noun)
Object2 :	(conf(Location name, Date, Verb Noun)
[rank] Verb2 :	(conf([Location name, Date Verb])
Object1 :	(conf(Location name, Date, Verb Noun)
Object2 :	(conf(Location name, Date, Verb Noun)

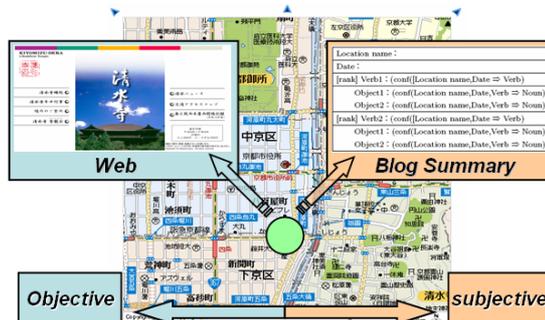


図 3 地域情報検索システムとの融合

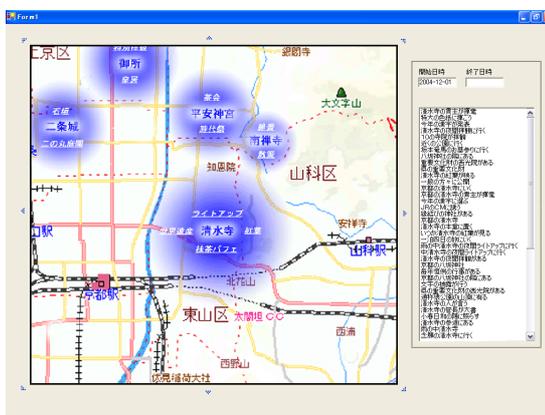


図 4 地図を用いた体験の空間的提示

感想を読むことが可能である。

Case2: クエリ = 時間, 動作, 対象 検索結果 = 空間

ユーザの興味のある動作とその対象が決まっており、それを行う場所を調べたいという状況のもとでこのような検索が想定される。以下、検索例を示す。

(例) クエリ: 桜を見る 期間: 4月

システムの検索結果は、それぞれの関連ルールの頻度に基づいて地名をリスト表示する。

Case3: クエリ = 空間, 時間, 動作 検索結果 = 対象

ユーザが、ある特定の時間にある場所において、したい動作も決定しているが、その対象を何にするかを決めたい。そのような場合にこのような検索が想定される。以下、検索例を示す。

(例) クエリ: 京都駅, 食べる 期間: すべて

5.3 システムの発展的な利用

地域情報検索における検索対象は店舗・寺社情報、位置情報、イベント情報などである。これらの情報は、カテゴリ検索やキーワード検索、地図検索をすることができる。特に、地図を用いた検索は空間という尺度で情報を選択することができ、地域情報検索に特徴的なシステムである¹⁴⁾。しかし、このようなシステム

で収集することができるのは、店舗が作成したホームページなどの情報である。これらの情報は、あくまで情報提供者側からの広告情報であり、信頼性に乏しい。また、ユーザは位置、カテゴリ、キーワードなどの属性を用いて情報を絞り込むことができても、そのようにして得られた複数の店舗からひとつの店舗を選択する決め手がない。そこで、地域情報検索における情報の絞り込みの新たな尺度として、街の消費者としての人々の動きを取り入れた地域情報検索が考えられる(図3)。人々の動きには流行が反映されている。また、個人の多様な行動の中には、新たな街の楽しみ方の発見があると考えられる。さらに、興味のある体験について書かれている Blog の中身をチェックし、その体験を実際に行った人々の感想を得ることも可能であり、複数の情報の絞り込みに利用することもできる。

6. 抽出結果

表 2 体験の抽出結果 地名:清水寺, 期間:期間:2005.01.01 ~ 2005.06.10

対象	動作	出現数
客	観光	60
桜	見る	45
人	思う	36
日記	書く	34
紅葉	見る	30
月	行う	29
写真	見る	26
人	見る	25

表 3 動作の抽出結果 地名:清水寺, 期間:期間:2005.01.01 ~ 2005.06.10

動作	出現数	確信度 (%)
見る	517	2.4
思う	482	2.2
旅行	442	2.0
観光	436	2.0
行う	424	2.0
言う	308	1.4
歩く	286	1.3
旅	256	1.2
書く	243	1.1
食べる	222	1.0

goo ブログ、及び livedoor から Blog を収集し、実験的に体験の抽出を行った。検索語は「清水寺」、収集期間は 2005.01.01 から 2005.06.10 である。収集した Blog 数は 360 件である。表 3 の確信度は、ルール: [2005.01.01 ~ 2005.06.10, 清水寺] [動作], 表 4 の確信度はルール: [2005.01.01 ~ 2005.06.10, 清水寺, 見

表 4 動作:「見る」の対象となる名詞・サ変名詞 地名:清水寺, 期間:2005.01.01 ~ 2005.06.10

名詞・サ変名詞	出現数	確信度 (%)
桜	45	8.7
紅葉	30	5.8
写真	26	5.0
人	25	4.8
ライトアップ	16	3.1
金閣寺	15	3.0
夜桜	15	3.0
月	14	2.7
雪	14	2.7

表 5 動作:「食べる」の対象となる名詞・サ変名詞 地名:清水寺, 期間:2005.01.01 ~ 2005.06.10

名詞・サ変名詞	出現数	確信度 (%)
ご飯	20	9.0
湯豆腐	14	6.3
店	14	6.3
パフェ	12	5.4
抹茶	11	5.0
ごはん	10	4.5
友達	10	4.5
都路	8	3.6

る] [対象], 表 5 の確信度はルール: [2005.01.01 ~ 2005.06.10, 清水寺, 食べる] [対象] のものである。

6.1 抽出結果の分析と今後の課題

表 2 の実験結果を見ると、「桜, 見る」「紅葉, 見る」のように時間という属性に特徴的な体験が一定の精度で抽出できていることがわかる。また、表 4, 表 5 に示すように、「見る」「食べる」のような動詞に関しては高い精度で語が抽出できていることがわかる。しかし、以下に示すような問題点があげられる。

- (1) 「客, 観光」のように動作とその対象という関係にない組も抽出されている(表 2)
- (2) 「言う」「思う」「書く」のような文章を書く際に一般的に利用されている表現が抽出されている(表 3)
- (3) 「旅行」「旅」のような同義語が含まれている(表 3)

今回の手法においては、係り受けを考慮していない。一文の中の語と語の相関関係のみを考慮しているが、これでは名詞と動詞の組が「動作」と「対象」の関係にあるかどうかまでの判断は難しい。今後の対応策として、名詞と動詞の組が格助詞「を」でつながる割合を Web 検索エンジンを利用して計算するなどして、体験のフィルタリングを行っていかうと考えている。また、文章における一般的な表現が抽出されてしまうという問題に対しては、複数地名間における語の出現頻度の相違と、同じ場所での違う期間における語の出

現頻度の相違を観察し、フィルタリングしていく方法が考えられる。この方法は、時間的・空間的要因に規定される体験を抽出する際にも有効な方法であると考えられ、今後、実験をするなどして有効性を検証していきたい。最後に同義語の問題であるが、これは同義語辞書を使うなどして対応していきたい。同義語をクラスタリングすることで、ルール抽出の精度も上がると予想できる。

7. ま と め

本稿においては、ある場所について書かれた個々のテキストから、人々の体験を、時間・空間・動作・対象属性間の相関ルールマイニングによって抽出する手法を提案した。そして、本手法によって得られた体験を、それらの属性を指定することで検索・要約するシステムを、その利用例と共に説明した。さらに、本システムの今後の展開と暫定的な実験を行った結果を示した。今後、現在判明している問題点を分析し、定量的な評価実験をすすめながら今回提案したアルゴリズムを修正、改良していく予定である。

謝 辞

本研究の一部は《知的資産》文部科学省科学技術振興費プロジェクト「異メディア・アーカイブの横断的検索・統合ソフトウェア開発」(代表:田中克己)および、平成17年度科研費特定領域研究(2)「Webの意味構造発見に基づく新しいWeb検索サービス方式に関する研究」(代表:田中克己)および、21世紀COEプログラム「知識社会基盤構築のための情報学拠点形成」によるものです。ここに記して謝意を表すものとします。

参 考 文 献

- 1) 上松 大輝, 沼 晃介, 徳永 徹郎, 大向一輝, 武田 英明: 場 log:Blog 環境における位置情報利用の提案, 第6回人工知能学会セマンティック Web とオントロジー研究会, 2004.
- 2) 奥村 学, 南野 朋之, 藤木 稔明, 鈴木泰裕: blog ページの自動収集と監視に基づくテキストマイニング, 第6回人工知能学会セマンティック Web とオントロジー研究会, 2004.
- 3) R. Kumar, J. Novak, P. Raghavan, and A. Tomkins, On the bursty evolution of blogspace, Proceedings of the 12th International World Wide Web Conference, pp. 568-576, 2003.
- 4) R. Kumar, J. Novak, P. Raghavan, and A. Tomkins, Structure and evolution

- of blogspace, Communications of the ACM, 47(12) pp. 35-39, 2004.
- 5) Judit Bar-Ilan, An Outsider's View On 'Topic-oriented' Blogging, Proceedings of the Alternate Papers Track of the 13th International World Wide Web Conference, pp. 28-34, 2004.
- 6) Matt Hurst, GIS and the Blogosphere, WWW 2005, 2nd Annual Workshop on the BloggingEcosystem: Aggregation, Analysis and Dynamics, China, Japan, 2005.
- 7) R. Agrawal and R. Srikant, Fast algorithms for mining association rules in large databases, Proceedings of the 20th Intl. Conf. on Very Large Data Bases, pp. 487-499, 1994.
- 8) goo ブログ, <http://blog.goo.ne.jp>
- 9) livedoor ブログ, <http://blog.livedoor.com/>
- 10) Zenrin Co.,Ltd, <http://www.zenrin.co.jp/>
- 11) MySQL, <http://www.mysql.com/>
- 12) 奈良先端科学技術大学松本研究室, 茶筌ホームページ <http://chasen.aist-nara.ac.jp/index.html>
- 13) 池原悟, 宮崎正弘, 白井諭, 横尾昭男, 中岩浩巳, 小倉健太郎, 大山芳史, 林良彦: 日本語語彙体系, 岩波書店.
- 14) ジオリンク京都, <http://www.digitalcity.gr.jp/>