

# 地域限定性スコアに基づく 位置情報付きコンテンツからの地域限定語句の抽出

奥 健太<sup>1,a)</sup> 西崎 剛司<sup>1,b)</sup> 服部 文夫<sup>1,c)</sup>

受付日 2012年3月20日, 採録日 2012年7月6日

**概要:** 本研究では, 語句の地域限定性に着目することで, 旅行先などの現地ならではの語句を自動的に抽出することを目的とする. 本稿では, 対象とする空間全体 (本稿では日本全国) に対し, 現地における出現頻度が相対的に高い語句を地域限定語句と定義している. 情報源としては, グルメ情報サイトやスポット情報サイトなどから収集される位置情報付きコンテンツに着目する. 指定された現地に存在するコンテンツから語句を抽出し, 語句の地域限定性を表す地域限定性スコアを算出する. 地域限定性スコアは, *IDF* および地域関連重みとよばれる尺度を組み合わせて算出される. また, グルメ情報サイトであるぐるなびから収集された 62,571 件のスポットデータを対象に評価実験を行った. 特徴語句抽出によく用いられる *IDF* をベースラインとして比較したとき, 提案の地域限定性スコアでは地域限定語句を効果的に抽出することが確認できた. また, 観光地や三大都市, その他の都市など対象地域の種別ごとに抽出語句の違いや傾向について分析を行った結果, 提案手法では特に観光地において有効な効果が得られることを確認した. さらに, 既存のグルメサイトのスポット情報を基に, 地域特有のスポットを推薦する機能を付加した提案システムを, 元のグルメサイトと比較した. その結果, 旅行時や出張時というシチュエーションにおいては, 提案システムが有効に働くことを示した.

**キーワード:** 情報推薦システム, 地理情報検索, 位置情報付きコンテンツ, 語句抽出

## Extraction of Region-restricted Phrases from Geotagged Contents Based on Region-restrictedness Score

KENTA OKU<sup>1,a)</sup> TAKASHI NISHIZAKI<sup>1,b)</sup> FUMIO HATTORI<sup>1,c)</sup>

Received: March 20, 2012, Accepted: July 6, 2012

**Abstract:** In this study, we aim at extracting region-restricted phrases in local areas by focusing on region-restrictedness of phrases. In this paper, we define a region-restrictedness phrase as a phrase that occurs in local areas frequently compared within the entire space (all over Japan in this paper). We focus on geotagged contents gathered from gourmet or spot information sites as information sources. Firstly, our method extracts phrases from the contents existing in designated local areas, and calculates region-restrictedness scores that denote region-restrictedness of phrases. The region-restrictedness score consists of *IDF* and regional weight. We also conducted experiments using 62,571 spot data from Gournavi that is one of the gourmet information sites. Compared with *IDF*, which is widely used for extracting feature phrases, experimental results showed that region-restricted phrases could be extracted effectively by the region-restrictedness score. We also analyzed differences and tendencies of extracted phrases from different local areas such as sightseeing areas, the three largest cities, and cities. The results showed that the score works effectively in the sightseeing areas particularly. In addition, we compared the proposed system adding recommendation method that recommends region-restricted spots into exiting gourmet sites based on spot data from the sites. The results showed that the system is more useful, which is one of the Japan restaurant guide services, in situations of travels and business trips.

**Keywords:** recommender systems, geographical search, geotagged contents, phrase extraction

<sup>1</sup> 立命館大学  
Ritsumeikan University, Kusatsu, Shiga 525-8577, Japan

a) oku@fc.ritsumei.ac.jp

b) cc008077@ed.ritsumei.ac.jp

c) fhattori@is.ritsumei.ac.jp

## 1. はじめに

特定のエリアの中から条件に合致したスポットを検索するローカルサーチがさかんに利用されている。ローカルサーチのサービスとして、Google マップ<sup>\*1</sup>や Yahoo!地図<sup>\*2</sup>、ドコイク?<sup>\*3</sup>などが一般公開されている。利用者が、住所や駅名などの場所および「居酒屋」や「ランチ」などのキーワードを条件として指定することにより、その場所から距離が近く、かつキーワードに合致するスポットが地図上に提示される。

しかし、利用者や状況によっては、距離の近さやキーワードに基づく検索だけでなく、その地域ならではのスポットを好む場合も考えられる。たとえば、ある利用者が三重県松阪市に観光旅行に訪れたというシチュエーションを考える。夕食をとるために携帯電話からローカルサーチサービスを利用して、松阪駅の近くにある飲食店を検索した。すると、松阪駅から近くにあるさまざまな飲食店が検索されたが、なかには全国チェーン店のように、どこでも利用できるようなものも含まれていた。しかし、この利用者にとってはせっかくの観光であるため、これらどこでも利用できるような店よりも、やはり対象地域である松阪ならではの料理（たとえば「松阪牛」）を提供している店が検索された方が、利用者にとっては魅力的であるといえる。

「松阪といえば松阪牛」というように、その土地の名物を容易に連想できる場合は、キーワード検索により検索可能であるが、有名でない土地に観光に行ったときには、何が名物であるかも分からない。また、観光情報誌や観光情報サイト、ご当地グルメサイトなど、多くの情報源が存在するものの、これらは人手で作成されるため、観光情報の整備や保守を行うためには大きなコストがかかる。特に、あまり知られていないような土地においては、情報が充実していないこともある。したがって、対象地域ならではの料理などに関する語句を Web から自動的に抽出し、それを提示することで、利用者に気付きを与える仕組みが必要である。

そこで、本研究では、語句の地域限定性に着目することで、旅行先などの対象地域ならではの語句を自動的に抽出することを目的とする。本稿では、対象とする空間全体（本稿では日本全国）に対し、対象地域における出現頻度が相対的に高い語句を地域限定性の高い語句（地域限定語句）と定義する。また、情報源としては Web 上の位置情報付きコンテンツを対象としており、特にグルメ情報サイトやスポット情報サイトなどから収集されるスポットデータに着目する。スポットデータには、スポット名および位置情報（経度・緯度もしくは住所）、テキスト情報（PR 文

やレビューなど）が含まれているものとする。提案手法では、対象地域を入力とし、その範囲において出現する地域限定語句を出力として利用者に提示する。まず、指定された対象地域に存在するスポットデータのテキスト情報から語句を抽出し、語句の地域限定性を表す地域限定性スコアを算出する。地域限定性スコアは、特徴語句抽出によく用いられる *IDF* [12] および本稿で定義する地域関連重みの組合せにより算出される。

*IDF* は文書検索においてよく用いられ、文書集合全体に対して多数の文書に共通して出現する語句には低い重みを与え、特定の文書にのみ限定して出現する語句には高い重みを与えるものである。たとえば、Web ページ全体を文書集合全体とした場合、「コース」という語句は 2 億件以上ものページに含まれるのに対し、「吉野本葛」という語句は Web 上では約 26 万件ほどである（Google による検索を行った場合）。この場合、「コース」に対しては *IDF* が低く、「吉野本葛」に対しては *IDF* が高く重み付けされる。なお、本稿では Web ページ全体を文書集合全体とした *IDF* を算出する。

ただし、*IDF* を考慮するだけでは、「店長オススメスペシャル」など、特に対象地域には関連しないが特異である語句も重みが高くなる。そこで、対象語句がどの程度対象地域に関連しているかを表す尺度である地域関連重みを導入する。具体的には、地域関連重みは、対象としているスポットから抽出された語句と、そのスポットの存在する地域の市区町村名とが関連する度合いであり、Web 上での対象語句と市区町村名との共起頻度から算出される。たとえば、対象地域を「三重県松阪市」としたとき、「松阪牛」は「松阪市」に強く関連するため、地域関連重みは高くなる。一方で、「大阪府大阪市」を対象地域としたときには、「松阪牛」は「大阪市」との関連が弱いので地域関連重みは低くなる。

これら *IDF* と地域関連重みを掛け合わせた地域限定性スコアを考慮することで、対象地域に関連しつつ特異な語句を抽出することができると考えている。最終的には、この地域限定性スコアが高い語句を地域限定語句として利用者に提示する。

図 1 は、提案手法を飲食店推薦に適用した地域限定グルメ推薦システムである。利用者が地図上で対象地域を指定することで、その対象地域ならではの語句がリストとして提示される。リスト中の語句を利用者が選択することで、その語句を含む飲食店データが地図上で強調表示される。本システムの詳細については 4 章で述べる。

提案の地域限定性スコアにより抽出目標とする語句は、たとえば、三重県松阪市における「松阪牛」や、三重県伊勢市における「伊勢海老」、奈良県奈良市における「吉野本葛」などといった、その地域特有の語句であり、このような語句のスコアを高くする。一方で、指定された対象地域

\*1 <http://maps.google.co.jp/> (2012/07)

\*2 <http://maps.loco.yahoo.co.jp/> (2012/07)

\*3 <http://www.doko.jp/> (2012/07)

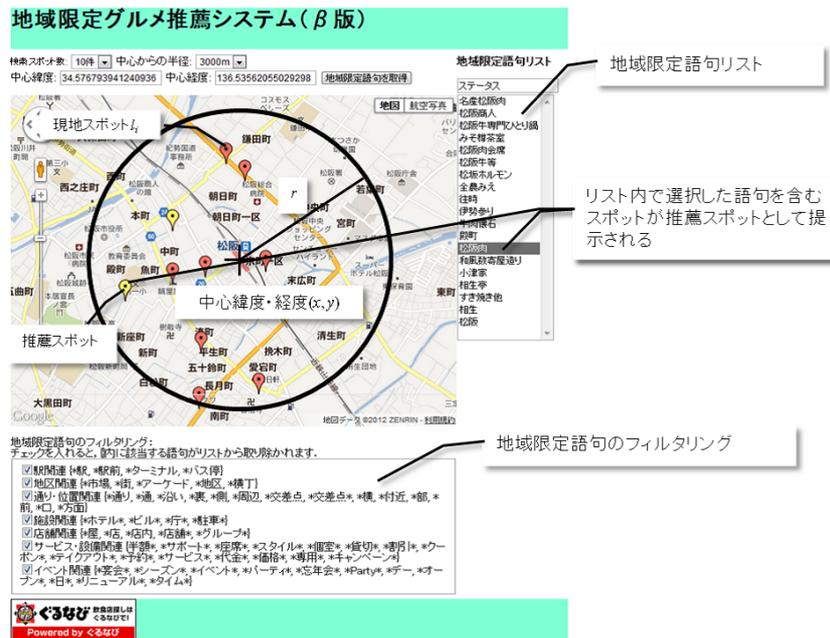


図 1 地域限定グルメ推薦システム  
Fig. 1 Region-restricted restaurant recommender system.

に現れるものの、どこにでもあるような語句は抽出対象とせず、スコアを低くする。

本研究の貢献は以下のとおりである。

- Web 上の位置情報付きコンテンツから地域限定語句を抽出するために、*IDF*、地域関連重みの各尺度を定義したうえで、これらを組み合わせた地域限定性スコアを提案した。
- ぐるなび<sup>\*4</sup>から収集された 62,571 件の飲食店データ (スポットデータ) を用いて、地域限定性スコアによる語句抽出性能について、定量分析および定性分析を行った。
- 実際の地域として、観光地 (「伊勢神宮」, 「松阪駅」, 「奈良公園」, 「近江八幡駅」, 「函館駅」), 三大都市 (「東京駅」, 「大阪駅」, 「名古屋駅」), 都市 (「札幌駅」, 「京都駅」, 「祇園」, 「博多駅」, 「広島駅」) をそれぞれ対象地域とした実験を行い、対象地域種別の違いによる抽出語句の傾向について分析を行った。
- 提案手法を飲食店推薦に適用した地域限定グルメ推薦システムを実装し、被験者実験によりその有用性評価を行った。実験では、既存のグルメサイトのスポット情報を基に、地域特有のスポットを推薦する機能を付加した提案システムを、元のグルメサイトと比較した。グルメサイトとして、国内の大手グルメサイトである、ぐるなびを選んだ。実験の結果、提案システムによる地域特有のスポットを推薦する機能を既存のグルメサイトに付加することで、有用性をより高めることができることを示した。

本稿は以下の構成となっている。2 章では関連研究を取り上げ、これらと本研究との違いについて述べる。3 章では提案手法である地域限定性スコアについて述べ、4 章で提案手法を実装した地域限定グルメ推薦システムについて説明する。5 章では実験結果から地域限定性スコアに関する定量分析および定性分析を行い、6 章で地域限定グルメ推薦システムを用いた被験者実験について述べる。最後に、7 章で本稿をまとめる。

## 2. 関連研究

本研究は、地理情報検索・推薦に関する研究および Web からの地域情報の抽出に関する研究に関連する。本章では、それぞれの関連研究について取り上げ、本研究との違いについて述べる。

### 2.1 地理情報検索・推薦

利用者の位置情報に基づく旅行情報提供システムは、古くから多くのものが提案されている [1], [2], [3]。しかし、これらは利用者の現在地から近くにあるスポットの情報を提供するものであり、その地域ならではのスポットを提供するというものではない。

奥ら [4] は、利用者の嗜好および状況を考慮した情報推薦システムを提案している。状況により変化する利用者の嗜好を SVM によりモデル化することで、嗜好および状況に応じた推薦を可能にしている。Kotera ら [5] は、利用者が地図上で行うカテゴリ選択および地図操作から利用者の興味度を推定することで、利用者の嗜好に合った地理情報を検索することを可能にしている。

これら位置情報や嗜好、状況といった要素は、利用者が

\*4 <http://www.gnavi.co.jp/> (2012/07)

スポットを選択する際に重要な判断基準となりうるものであり、これらの要素を考慮した地理情報検索・推薦システムは、これまでも多く提案されている。我々は、スポットの地域限定性という要素もまたスポットを選択する際に重要な判断基準の1つとなりうるという考えのもと、地域限定性に着目した推薦システムの開発を行っている。提案システムの対象利用者層としては、地域限定的なスポットを好む利用者と想定している。本稿で提案する手法においては、利用者の嗜好や状況を直接的に取り入れたものではないが、将来的には、上記に示したような嗜好や状況を考慮した情報検索・推薦システムと連携することで、システム全体としての利用者満足度を向上させることができると考えている。

また、Tarumiら [6], [7] は、時空間限定情報を扱うシステム SpaceTag を提案している。SpaceTag は、時空間限定でアクセス可能な仮想オブジェクトであり、特定の場所、特定の時間でのみアクセスできるテキスト、画像、音声、プログラムなどの任意のオブジェクトである。ただし、SpaceTag は、企業や公的機関、一般利用者によって作成されるため、コンテンツの生成やメンテナンスを行うための人的コストが大きい。これに対し、本研究では実空間上の位置に関連付けられた膨大なスポット情報の中から自動的に地域に限定的な語句を抽出するものである。

## 2.2 Web からの地域情報の抽出

ブログから地域情報を抽出する研究として、岡本ら [8], [9] や石野ら [10] の研究がある。岡本ら [8], [9] は、「東京」や「秋葉原」などの地名をクエリとして検索されたブログエントリを情報源として、そこから地域イベント情報を抽出する手法を提案している。石野ら [10] は、旅行記が記述された旅行ブログエントリを検出し、その旅行ブログエントリから土産物情報や観光名所情報を抽出する手法を提案している。特に、ブログ集合から旅行ブログエントリを自動的に検出することで、観光情報の抽出精度を高めている。ただし、ブログから地域情報を抽出する際には、情報源となる対象地域に関連するブログや旅行ブログの検出精度に依存する。これらの研究に対し、本研究では、ブログではなく、スポット情報サイトから取得される、あらかじめ位置情報が付加されたテキストデータを対象としている。そのため、対象地域に関連するテキストデータの検出は正確に行うことができる。

手塚ら [11] は、ウェブページやオブジェクト（「紅葉」や「うどん」など）が持つ「地域性」を推定する手法を提案している。たとえば、「紅葉」が有名な場所を調べたいとき、「紅葉」というオブジェクト名を入力することで、最も「紅葉」と関連の深い地域を表示させる。つまり、オブジェクト名から関連する地域を取得するというものである。これに対し、本研究では、地域からその地域に限定的な語句

を抽出することを目指しており、入力と出力が異なる。特に、指定された地域に限定的な語句が利用者に提示されるため、利用者が自身で想起できないような語句を気付かせることができるという利点がある。

## 3. 提案手法

本研究では、グルメ情報サイトやスポット情報サイトなどから収集されるスポットデータから、地域限定語句を抽出する手法を提案する。まず本研究で扱うスポットの用語の定義について述べ、提案手法の概要を示す。その後、スポットからの語句抽出方法および抽出された語句の地域限定性スコアの算出方法について述べる。

### 3.1 スポットの定義

本稿では、飲食店や観光施設、娯楽施設など、実空間上の位置情報と関連付けられたオブジェクトをスポットと定義し、スポットに関する情報をスポットデータとよぶ。スポットデータは、ぐるなびや HOT PEPPER<sup>\*5</sup>、ドコイク?、楽天トラベル<sup>\*6</sup>などのグルメ情報サイトやスポット情報サイトなどから参照でき、各サイトが提供している API を用いて取得することができる。提案手法では、スポットデータとして、

- スポット名
- 位置情報（経度・緯度もしくは住所）
- テキスト情報（PR 文やレビューなど）

が得られることを前提とする。

### 3.2 提案手法の概要

1章で述べたように、本研究では、対象とする空間全体に対し、対象地域における出現頻度が相対的に高い語句を地域限定語句とし、その語句を抽出するための尺度として地域限定性スコアを提案する。地域限定性スコアでは、三重県松阪市における「松阪牛」や奈良県奈良市における「吉野本葛」などといった、他の地域に比べ相対的に出現頻度が高い語句、すなわち、その地域ならではの語句のスコアを高くする。

提案手法では、対象地域（中心経度・緯度  $(x, y)$  およびそれを中心とした半径  $r$  の円領域）および検索スポット数  $k$  を入力とし、その領域内にある地域限定語句を出力として提示する。ここで、検索スポット数  $k$  は、システムにより検索されるスポット数の上限であり、後述する方法で、最大  $k$  件のスポットを基に地域限定性スコアが算出される。

図 1 は、提案手法を実装したシステムのインタフェースであり、指定された対象地域における地域限定語句リストを提示する様子を示している。なお、本システムの詳細については、4章において説明する。提案手法の手順は以下

<sup>\*5</sup> <http://www.hotpepper.jp/index.html> (2012/07)

<sup>\*6</sup> <http://travel.rakuten.co.jp/> (2012/07)

のとおりである。

- (1) 利用者が対象地域  $(x, y, r)$  および検索スポット数  $k$  を指定する。
  - (2) 現地スポット集合  $L$  を取得する。
  - (3) 現地スポット  $l_i \in L$  に含まれるテキスト情報から語句  $w_j$  を抽出する。
  - (4) 抽出された語句  $w_j$  の地域限定性スコア  $s_j$  を算出する。
  - (5) 地域限定性スコア  $s_j$  を基準に語句のランキングを行い、上位の語句を提示する。
- 以下、各手順について述べる。

### 3.3 現地スポット集合からの語句抽出

対象地域  $(x, y, r)$  として指定された範囲内に存在するスポットを取得する。ここで、指定範囲内に存在するスポット数が  $k$  件を超える場合は、 $(x, y)$  から最寄りの  $k$  件を取得する。ただし、検索件数の上限を設けるよりも、より広い範囲のスポットを対象とする方が、提案手法の有用性の観点からは望ましいと考えられる。しかし、付録 A.1 に示すように、検索スポット数  $k$  を大きくすると、ユーザが地域限定語句を要求してから地域限定語句が出力されるまでの応答時間が長くなってしまふ。そのため、システムの有用性と応答時間とのトレードオフについて考慮する必要がある。4章で説明するプロトタイプシステムでは、ユーザに検索スポット数  $k$  を指定してもらうことで、多少時間がかかっても有用な情報が欲しい場合には  $k$  を大きくし、手早く情報が欲しい場合には  $k$  を小さくするなどの使い方ができるようにしている。

対象地域内で取得されたスポットを現地スポット  $l_i$  とよび、その集合を現地スポット集合  $L = \{l_1, l_2, \dots\}$  と表す。現地スポット集合  $L$  内の各スポット  $l_i$  のテキスト情報に含まれる語句を抽出する。語句抽出には、形態素解析器である茶筌\*7を用いる。抽出対象の品詞を表 1 のとおりとし

表 1 形態素解析による抽出対象の品詞

Table 1 Parts of speech extracted by morphological analysis.

抽出対象の品詞	例
名詞—一般	アナゴ, ウニ
名詞—固有名詞—一般	若草山, 春日山
名詞—固有名詞—人名—一般	紫式部
名詞—固有名詞—人名—姓	山岡
名詞—固有名詞—人名—名	鉄舟
名詞—固有名詞—組織	春日大社, 東大寺
名詞—固有名詞—地域—一般	伊江島, 松阪
名詞—固有名詞—地域—国	スペイン
名詞—サ変接続,	料理
名詞—接尾—一般,	焼き
未知語	ハウレン, ポジョレーヌーボー

\*7 <http://chasen.naist.jp/hiki/ChaSen/> (2012/07)

た。表 1 には、各品詞により抽出される語句の例を示している。ただし、「松阪牛」のような複合語に対応するため、連続する名詞や未知語は 1 つの語句として抽出している。

現地スポット集合  $L$  から抽出された語句の集合を、

$$W = \{w_1, w_2, \dots\} \quad (1)$$

と表す。

### 3.4 抽出語句の地域限定性スコアの算出

抽出された各語句  $w_j$  が、どの程度その地域に限定的なものであるかを調べるため、地域限定性スコアを定義する。地域限定性スコアは、以下の 2 種類の尺度の組合せにより表される。

(a) *IDF*

(b) 地域関連重み  $\omega$

以下、各尺度について述べる。

(a) *IDF*

Web 全体を通して出現しやすい語句はさほど重要ではない。このような重要でない語句の重みを下げるため、文書検索においてよく用いられる *IDF* [12] を用いる。語句  $w_j$  の *IDF* <sub>$j$</sub>  は次式で表される。

$$IDF_j = \log \frac{N}{n_j} \quad (2)$$

ここで、 $N$  は文書の総数である。本研究では Web 全体に対する *IDF* を求めるため、文献 [13] を参考に、Web 検索エンジンが持つ全文書数を仮定した  $N = 10^{10}$  を用いる。 $n_j$  は Web 上において語句  $w_j$  を含むページ数である。

(b) 地域関連重み

地域関連重み  $\omega_j$  は、対象語句がどの程度対象地域に関連しているかを表す尺度であり、具体的には、対象としているスポットから抽出された語句と、そのスポットの存在する地域の市区町村名とが関連する度合いである。たとえば、「松阪」における「松阪牛」のようにその地域との関連が強い語句には高い重みを与え、「店長オススメスペシャル」のようにその地域との関連が弱い語句には低い重みを与える。

この地域関連重みを考慮するために、Web 上での単語の共起頻度に基づいた単語類似度を表す指標である *WebPMI* [13] を用いる。単語  $p$  および  $q$  の *WebPMI* は次式で表される。

$$WebPMI(p, q) = \begin{cases} 0 & \text{if } H(p \cap q) \leq c \\ \log \frac{\frac{H(p \cap q)}{N}}{\frac{H(p)}{N} \frac{H(q)}{N}} & \text{otherwise} \end{cases} \quad (3)$$

ここで、 $H(p)$ 、 $H(q)$ 、 $H(p \cap q)$  は、それぞれ、「 $p$ 」、「 $q$ 」、「 $p$  AND  $q$ 」をクエリとして Web 検索を行ったときのヒット件数である。 $N$  は Web 検索エンジンが持つ全文書数であり、本研究では先述のとおり  $N = 10^{10}$  を用いる。また、

$c$ は低頻度語によるノイズを避けるために用いられる閾値であり、本研究では文献 [13] を参考に  $c=5$  とする。

語句  $w_j$  と現地スポット  $l_i$  の住所に含まれる市区町村名  $local$  との  $WebPMI(w_j, local)$  を求めることによって、語句  $w_j$  の地域関連重み  $\omega_j$  を算出する。つまり、 $\omega_j$  は次式のように定義される。

$$\omega_j = WebPMI(w_j, local) \quad (4)$$

ここで、現地スポット  $l_i$  の市区町村名の抽出には逆ジオコーディング API である Yahoo!リバーシジオコーダ<sup>\*8</sup>を用いる。現地スポット  $l_i$  の経度・緯度から逆ジオコーディングにより取得された市区町村名（「松阪市」や「千代田区」など）を  $local$  とする。なお、市区町村名としては、取得された住所において最初に表示される市区町村名を抽出する。たとえば、「東京都千代田区永田町」の場合「千代田区」が、「大阪府大阪市北区」の場合「大阪市」、「三重県松阪市殿町」の場合「松阪市」、「京都府相楽郡精華町」の場合「精華町」がそれぞれ抽出される。

最終的に、上記 (a), (b) を掛け合わせ、地域限定性スコアを算出する。語句  $w_j$  の地域限定性スコア  $s_j$  は次式により求められる。

$$s_j = IDF_j^{\alpha} \times \omega_j^{\beta} \quad (5)$$

$(0 \leq \alpha \leq 1, 0 \leq \beta \leq 1)$

ただし、 $IDF_j^*$ 、 $\omega_j^*$  は、それぞれ抽出語句集合  $W$  において、 $IDF_j$ 、 $\omega_j$  を  $[0, 1]$  の範囲に正規化したものである。具体的には、値  $x$  を次式で表される正規分布の累積分布関数を用いて  $[0, 1]$  に正規化している。

$$F(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt \quad (6)$$

ここで、 $\mu$ 、 $\sigma$  はそれぞれ、抽出語句集合  $W$  における各値の平均および標準偏差を表す。

また、式 (5) 中の  $\alpha$ 、 $\beta$  は、それぞれ、 $IDF_j$ 、 $\omega_j$  の重みを調整するためのパラメータである。5章では、これらのパラメータを変えたときの語句抽出性能の違いを分析する。

#### 4. プロトタイプシステムの実装

3章で説明した提案手法を実装したプロトタイプシステムとして地域限定グルメ推薦システム<sup>\*9</sup>の構築を行った。実装は、JSP、Java Servlet、JavaScript で行った。また、インタフェース部分の設計には HTML5 および CSS3 を用い、地図表示は、Google Maps API<sup>\*10</sup>を用いて行っている。データベース管理システムは MySQL を用い、ぐる

<sup>\*8</sup> <http://developer.yahoo.co.jp/webapi/map/openlocalplatform/v1/reversegeocoder.html> (2012/07)

<sup>\*9</sup> <http://lepus.okukenta.net/> (2012/07)

<sup>\*10</sup> <https://developers.google.com/maps/> (2012/07)

なび API により収集した飲食店データ 62,571 件をデータベースに格納している。アプリケーションサーバとしては Apache Tomcat を用いた。

1章で示した図 1 は、本システムのインタフェースである。本システムにおいては、利用者は以下の操作を行うことができる。

**地図操作** 地図上で移動操作やズームイン、ズームアウトなどの操作を行うことができ、任意の領域（中心経度・緯度  $(x, y)$ ）を表示させることが可能である。地図上には  $(x, y)$  から最寄りの  $k$  件の飲食店データが赤マーカーとして表示される。

**スポット情報の表示** 地図上に表示されたマーカーをクリックすることで、その飲食店の名前がポップアップウィンドウとして表示される。さらに、その飲食店名をクリックすることで、ぐるなびのサイトで飲食店情報の詳細が確認できる。

**入力情報の変更** 地図の上部にあるプルダウンメニューから、検索スポット件数  $k$  および中心からの半径  $r$  を変更することが可能である。

**地域限定語句の取得** 「地域限定語句を取得」ボタンをクリックすることで、対象地域  $(x, y, r)$  および検索スポット数  $k$  を入力としたときの地域限定語句が最大 20 件、リストとして提示される。

**地域限定スポットの提示** 利用者は提示された地域限定語句リストの中から興味のある語句を選択することで、地図上にその語句を含む飲食店データが推薦スポット（黄マーカー）として提示される。

**地域限定語句のフィルタリング指定** 駅や地区名など地域限定的ではあるが利用者にとって不要と思われる語句を、任意にフィルタリング指定することが可能である。フィルタリング指定された語句は地域限定語句リストから除外される。フィルタリング規則は駅関連や地区関連などカテゴリ分けされており、該当するチェックボックスにチェックを入れることでフィルタリングを指定することができる。

なお、現在登録しているフィルタリング規則については、付録 A.2 に示す。これらの規則は今回は著者の判断で登録したものであるが、将来的にはシステム運用時の利用者の利用状況などにに基づき、フィルタリング規則を追加、削除していく予定である。

#### 5. 評価実験

提案手法の有効性を評価するために、ぐるなびの飲食店データを用いた評価実験を行った。「伊勢神宮」や「奈良公園」などの実際の地域を想定し、提案手法による語句抽出性能について定量分析および定性分析を行った。また、観光地、三大都市、都市といった対象地域種別の違いやパラメータ  $\alpha$ 、 $\beta$  による語句抽出性能の傾向について分析を

表 2 実験に用いた対象地域  
Table 2 Local information used in experiments.

現地	現地種別	現地市区町村	経度 $x$	緯度 $y$	半径 $r$ (m)	検索スポット数 $k$	抽出語句数	正解データ数
伊勢神宮	観光地	伊勢市	136.702798	34.485031	3,000	20	243	24
松阪駅	観光地	松阪市	136.535514	34.577235	3,000	20	274	27
奈良公園	観光地	奈良市	135.840157	34.683437	3,000	20	315	27
近江八幡駅	観光地	近江八幡市	136.102577	35.123095	3,000	20	335	20
函館駅	観光地	函館市	140.726426	41.773785	3,000	20	331	41
東京駅	三大都市	千代田区	139.691844	35.681403	3,000	20	418	16
大阪駅	三大都市	大阪市	135.495389	34.702214	3,000	20	433	10
名古屋駅	三大都市	名古屋市	136.881619	35.170976	3,000	20	309	6
札幌駅	都市	札幌市	141.350737	43.068763	3,000	20	308	8
京都駅	都市	京都市	135.757838	34.982753	3,000	20	311	16
祇園	都市	京都市	135.780776	35.003725	3,000	20	427	71
博多駅	都市	福岡市	130.419459	33.589429	3,000	20	320	10
広島駅	都市	広島市	132.474782	34.397465	3,000	20	292	15

行った。さらに、4章で説明した地域限定グルメ推薦システムを用いた被験者実験を行った。以降、まず実験に用いたデータセットおよび正解データについて説明し、各実験結果について述べる。

### 5.1 データセット

提案手法は、飲食店や娯楽施設、観光施設などに対して適用しうる手法であるが、本実験では、スポットのジャンルとして飲食店を選択した。

スポットデータは、一般向けに公開されているグルメ情報サイトであるぐるなびから取得した。本実験では、ぐるなび API<sup>\*11</sup>により、各スポットデータの

- 飲食店名
- 位置情報 (経度・緯度)
- テキスト情報 (PR 文 (短) および PR 文 (長))

を取得した。取得したスポットデータの例は付録 A.3 に掲載している。

本実験では日本全国を対象に、62,571 件のスポットデータを取得した。また、表 2 に示す各地域をそれぞれ対象地域として設定した。特に、主な観光地として「伊勢神宮」、「松阪駅」、「奈良公園」、「近江八幡駅」、「函館駅」を選択した。また、三大都市として「東京駅」、「大阪駅」、「名古屋駅」を、三大都市以外で推定人口百万人以上である都市の中から、「札幌駅」、「京都駅」、「祇園」、「博多駅」、「広島駅」をそれぞれ選択した。各対象地域  $(x, y, r)$  は表 2 のとおりである。また、表には本実験において検索されたスポット数および語句数、後述する正解データ数をあわせて掲載している。

### 5.2 正解データ

提案手法の有効性を定量的に評価するために、あらかじめ各対象地域を対象に抽出された語句に対し、正解ラベルを付与した。正解ラベルの付与は被験者 10 名 (男性 8 名、女性 2 名) の意見に基づいて行った。被験者は 20 歳から 30 歳の学生および会社員であり、被験者の属性は付録 A.6 に掲載している。

具体的な正解ラベルの付与手順は以下のとおりである。

- 被験者に各対象地域において抽出された語句を提示した。
- 提示した語句の地域限定性について「この語句は対象地域に限定的なものであると思いますか」という質問を行い、被験者に 5 段階評価値 ( {5: 強くそう思う, 4: そう思う, 3: どちらともいえない, 2: そうは思わない, 1: まったくそうは思わない } ) で回答してもらった。
- 10 名の被験者において平均評価値が 4 以上の語句を正解データとし、それ以外の語句を不正解データとした。

たとえば、奈良公園を対象地域としたとき、地域限定語句として、「吉野本葛」や「東大寺」などが選ばれた。一方で、地域限定的でない語句として、「栗ご飯」や「鱧」などが選ばれた。これらは特にその地域ならではのものはなため、今回の手法の目的からは不正解データとして扱われる。

### 5.3 ベースライン手法との語句抽出性能の比較

3.4 節で定義した地域限定性スコアによる語句抽出の性能を評価するために、国内の主な観光地である、「伊勢神宮」、「松阪駅」、「奈良公園」、「近江八幡駅」、「函館駅」の 5 地点を対象地域として実験を行った。

ここでは、以下の尺度を比較対象として用いた。

\*11 <http://api.gnavi.co.jp/api/service.htm> (2012/07)

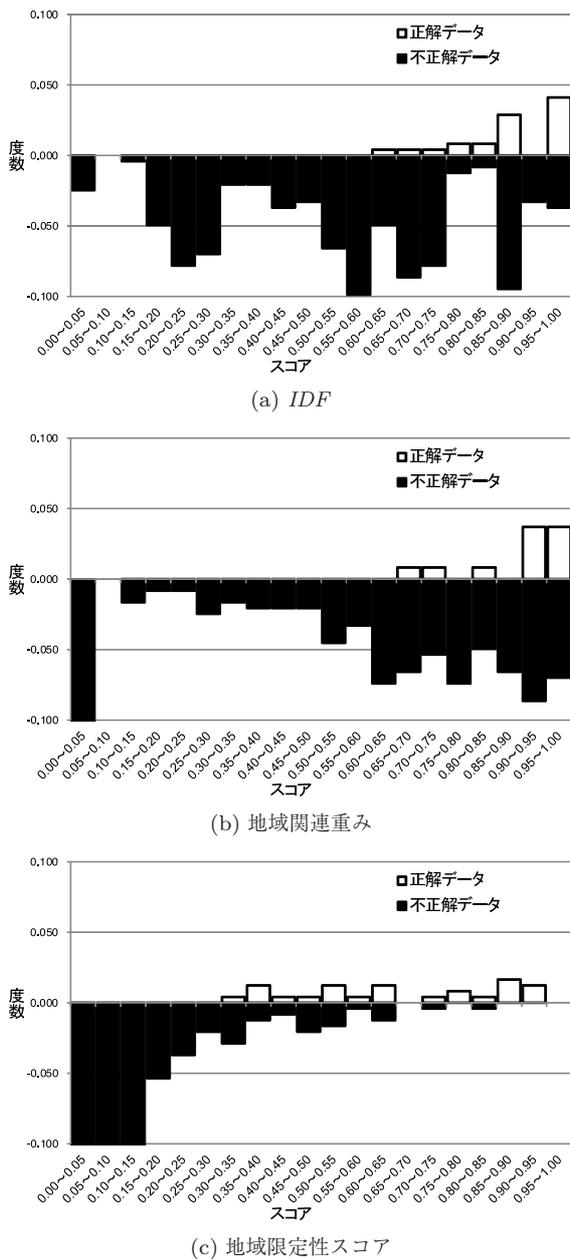


図 2 (a) IDF, (b) 地域関連重み, (c) 地域限定性スコアによる正解・不正解データの度数分布 (対象地域:「伊勢神宮」)

Fig. 2 Frequency distribution of relevant/irrelevant data by (a) IDF, (b) regional weight, and (c) region-restrictedness score (“Grand Shrine at Ise” as a local point).

(a) IDF (ベースライン手法)

(b) 地域関連重み  $\omega$

(c) 地域限定性スコア  $s$  (式 (5) のとおり. ただし,  $\alpha = \beta = 1$  としている.)

なお, (a) IDF は, 特徴語句抽出によく用いられる手法であり, 本実験のベースライン手法として用いる.

### 5.3.1 定量分析

まず, 各尺度の語句抽出性能を定量的に分析する. 各対象地域において抽出された語句に対し, 上記の各尺度によりスコア付けを行った. 階級数を 20 とし, 各階級における正解・不正解データの度数を調べた. 図 2 は, 「伊勢神宮」

を対象地域としたときの, 各尺度 (a)~(c) による正解・不正解データの度数分布を示している. ただし, グラフを見やすくするために, 不正解データの度数は負の方向に描いている. また, 度数は抽出された全語句数を 1 としたときの割合を示している.

(a) および (b) においては, スコア上位においても正解データと不正解データが混在していることが分かる. このように, (a) および (b) の尺度を単独で用いるだけでは, 正解データと不正解データを正確に分離できないことが分かる. 一方で, (c) においては, 地域限定性スコア  $s \geq 0.70$  において, 不正解データが 2 件混在しているものの, ほぼ正解データのみを抽出できていることが分かる. 下位にも正解データが含まれているが, 利用者には上位の語句を地域限定語句として提示することが有効であるといえる.

なお, 他の対象地域を対象としたときも同様の結果が得られた. 参考に, 付録 A.4 に, 他の対象地域を対象としたときの, 地域限定性スコアによる正解・不正解データの度数分布を示す.

### 5.3.2 定性分析

続いて, 各尺度 (a) および (b) により抽出できないような語句が, (c) 地域限定性スコアにより, どのように適切に抽出できているかを見るために, その語句抽出性能を定性的に分析する. ここでは, 「伊勢神宮」を対象地域としたときの結果から分析を行う. 表 3 には, 「伊勢神宮」を対象地域としたとき, 各尺度 (a)~(c) によるスコア上位 10 件の語句を示している. 各表には, その尺度によるスコア, 被験者評価に基づいて付与した地域限定性の区別 ( $\{0,1\}$ ), 各尺度によるランキング順位を示している. なお, 正解データ (地域限定性に “1” が与えられている語句) は灰色の塗りつぶしにより示している.

#### (a) IDF

「本場松阪肉」や「ブランド松阪肉」など, 特徴的な語句が上位に抽出されていることが分かる. しかしながら, 「業態本格焼酎」や「昭和得トクコース」など, 特に伊勢に限定的であるとはいえない語句も抽出されていることが分かる.

他の尺度による順位に着目すると, 「業態本格焼酎」や「昭和得トクコース」などは伊勢に関連するものではないことから, 地域関連重み  $\omega$  が低くなっている. 結果として地域限定性スコアによる順位が下がっていることが分かる.

#### (b) 地域関連重み $\omega$

「伊勢海老」や「地元伊勢志摩」など, 伊勢に関連する語句が抽出されていることが分かる. しかしながら, 「人気」や「海」など, 一般的な語句も抽出されていることが分かる. ただし, これら一般的な語句は, IDF が低くなり, 結果的に, 地域限定性スコア  $s$  において順位が下がっている.

#### (c) 地域限定性スコア $s$

「伊勢神宮外宮」や「伊勢海老」など, 伊勢に限定的な語

表 3 各尺度による抽出語句上位 10 件 (対象地域:「伊勢神宮」)

Table 3 Top 10 phrases extracted by each score ("Grand Shrine at Ise" as a local point).

(a) IDF

抽出語句	スコア	各尺度による順位			
		地域限定性	IDF	地域関連重み	地域限定性スコア
和具漁港直送	1.000	1	1	24	5
本場松阪肉	0.998	1	2	33	8
昭和鍋三昧コース	0.997	0	3	47	10
ブランド松阪肉	0.993	1	4	34	9
業態本格焼酎	0.992	0	5	73	16
昭和得トクコース	0.990	0	6	49	11
釜飯膳	0.978	0	7	67	13
伊勢神宮外宮	0.977	1	8	12	2
伊勢海老	0.977	1	9	1	1
志摩和具	0.975	1	10	9	3

(b) 地域関連重み

抽出語句	スコア	各尺度による順位			
		地域限定性	IDF	地域関連重み	地域限定性スコア
伊勢海老	0.992	1	9	1	1
人気	0.990	0	68	2	144
海	0.990	0	75	3	102
企業	0.989	0	76	4	37
車	0.988	0	90	5	58
地元伊勢志摩	0.988	1	12	6	4
木	0.987	0	88	7	119
ダシ	0.987	0	57	8	47
志摩和具	0.986	1	10	9	3
名前入り	0.986	0	91	10	92

(c) 地域限定性スコア

抽出語句	スコア	各尺度による順位			
		地域限定性	IDF	地域関連重み	地域限定性スコア
伊勢海老	0.972	1	9	1	1
伊勢神宮外宮	0.965	1	8	12	2
志摩和具	0.965	1	10	9	3
地元伊勢志摩	0.964	1	12	6	4
和具漁港直送	0.963	1	1	24	5
外宮	0.953	1	14	15	6
内宮	0.936	1	19	19	7
本場松阪肉	0.935	1	2	33	8
ブランド松阪肉	0.931	1	4	34	9
昭和鍋三昧コース	0.920	0	3	47	10

表 4 対象地域ごとの抽出語句上位 20 件 (対象地域種別:観光地)

Table 4 Top-20 phrases extracted in each local point (sightseeing areas as a local type).

(a) 伊勢神宮				(b) 松阪駅				(c) 奈良公園			
順位	抽出語句	地域限定性	地域限定性スコア	順位	抽出語句	地域限定性	地域限定性スコア	順位	抽出語句	地域限定性	地域限定性スコア
1	伊勢海老	1	0.972	1	名産松阪肉	1	0.988	1	伊 ZAKAYA	0	0.955
2	伊勢神宮外宮	1	0.965	2	松阪牛専門ひとり鍋	1	0.974	2	東大寺北角	1	0.928
3	志摩和具	1	0.965	3	松阪商人	1	0.954	3	倉院展	1	0.840
4	地元伊勢志摩	1	0.964	4	松阪肉会席	1	0.950	4	猿沢池	1	0.838
5	和具漁港直送	1	0.963	5	松阪牛ステーキ	1	0.908	5	貸し衣装	0	0.777
6	外宮	1	0.953	6	松阪牛ステーキ	1	0.906	6	奈良商工会議	1	0.766
7	内宮	1	0.936	7	松坂ホルモン	1	0.888	7	吉野本葛	1	0.764
8	本場松阪肉	1	0.935	8	みそ博茶室	1	0.832	8	有明柳川	1	0.758
9	ブランド松阪肉	1	0.931	9	松阪牛	1	0.829	9	正園子提ゆかり	1	0.758
10	昭和鍋三昧コース	0	0.920	10	往時	0	0.826	10	大小お座敷	0	0.745
11	昭和得トクコース	0	0.907	11	全農みえ	1	0.787	11	鴨雄鶏ガラ鍋	0	0.740
12	ダシ	0	0.853	12	松阪肉	1	0.771	12	天下造品!	0	0.732
13	釜飯膳	0	0.848	13	上がり席	0	0.757	13	町家風	0	0.729
14	スペシャルカレー	0	0.844	14	牛肉横石	0	0.737	14	大仏焼	1	0.716
15	鮎	1	0.843	15	宴席	0	0.734	15	本場大和	1	0.713
16	業態本格焼酎	0	0.840	16	伊勢参り	1	0.733	16	三つわらび御膳	1	0.712
17	宇治山田	1	0.834	17	相生	1	0.731	17	興福寺	1	0.683
18	あわび	1	0.820	18	殿町	1	0.712	18	苜蓿餅	1	0.674
19	アワビ	1	0.816	19	牛	0	0.702	19	html 吉野本葛	1	0.661
20	久居	1	0.803	20	松阪	1	0.698	20	奈良公園散策	1	0.660

(d) 近江八幡駅				(e) 函館駅			
順位	抽出語句	地域限定性	地域限定性スコア	順位	抽出語句	地域限定性スコア	地域限定性
1	牛舎直送	1	0.967	1	函館近海	1	0.971
2	八幡海苔	1	0.913	2	函館名物イカソーメン・イカ刺	1	0.969
3	丁字麩	1	0.882	3	地ビール工場併設	0	0.969
4	近江ゆず地	1	0.871	4	五稜郭タワー向い	1	0.966
5	近江しゃも	1	0.843	5	函館西波止場	1	0.959
6	近江八幡料理	1	0.832	6	函館名産	1	0.958
7	JR 近江八幡	1	0.827	7	函館ならではの	1	0.931
8	鮎寿司	1	0.808	8	函館朝市	1	0.922
9	森本養鶏場	1	0.788	9	函館湾	1	0.903
10	将軍家	0	0.685	10	活イカ刺	1	0.898
11	盛り上り	0	0.681	11	函館名物	1	0.866
12	近江牛料理	1	0.659	12	JR 函館	1	0.856
13	近江八幡	1	0.659	13	道産料理	1	0.844
14	鮎寿司	1	0.616	14	保冷トラック	0	0.834
15	関西寿し	1	0.609	15	産地漁港	1	0.782
16	バームクーヘン豚	1	0.606	16	札幌すずき	1	0.781
17	元禄年間	0	0.594	17	ダシ	0	0.777
18	秋近江	1	0.583	18	地ビール	0	0.772
19	スペシャルカレー	0	0.546	19	寿し	0	0.772
20	最高	0	0.540	20	函館味	1	0.750

句が上位に抽出されていることが分かる。

以上のように、各尺度 (a) および (b) を単独に用いるだけでは適切に抽出できなかった語句が、それぞれを組み合わせた地域限定性スコアを用いることで適切に抽出できているということが確認できる。

#### 5.4 対象地域種別の違いによる語句抽出性能の傾向分析

表 2 に示した、観光地および三大都市、都市の各対象地域を対象としたときの抽出語句を、それぞれ、表 4、表 5、表 6 に示す。各表には、地域限定性スコアによるランキング上位 20 件の語句を掲載している。また、表 3 と同様に、正解データは灰色の塗りつぶしにより示している。

##### (a) 観光地

まず、これまでの実験結果で示したように、「伊勢神宮」や「函館駅」など観光地を対象としたとき、地域限定性スコアではおおむねランキング上位に正解データを抽出できていることが確認できる。「伊勢神宮」における「伊勢海老」や、「函館駅」における「函館名物イカソーメン・イカ刺」、「奈良公園」における「吉野本葛」など、といったように、各地域ならではの語句が抽出されている。

また、「伊勢神宮」における「伊勢神宮外宮」や「奈良公園」における「東大寺北角」などといったランドマークを表す語句が抽出されている点は、その地域の特性が現れて

表 5 対象地域ごとの抽出語句上位 20 件 (対象地域種別:三大都市)  
**Table 5** Top-20 phrases extracted in each local point (the three largest cities as a local type).

(a) 東京駅

順位	抽出語句	地域限定性	地域限定性スコア
1	ピモン冬	0	0.920
2	八重洲地下街産地直送	1	0.864
3	ろフレックひつ	0	0.833
4	鉄板ピストロ RICOLO	1	0.810
5	椿壽	1	0.753
6	ボジョ	0	0.694
7	八重洲地下街雨	1	0.691
8	浜田漁港	0	0.671
9	サバディーニイタリアン	0	0.669
10	珊瑚シェウマイ	0	0.655
11	親子飯茶漬け!	0	0.600
12	牛たん炭焼利久	0	0.542
13	サバディーニ	0	0.535
14	糸魚川漁港	0	0.514
15	塩釜漁港	0	0.492
16	ミニ亀ゼリー	0	0.492
17	エキオカ	1	0.491
18	自家製サンタリア	0	0.465
19	蟹すき鍋	0	0.462
20	赤城和牛	0	0.453

(b) 大阪駅

順位	抽出語句	地域限定性	地域限定性スコア
1	パティシエ特製ラウンドケーキ	0	0.992
2	好評くる	0	0.903
3	NEWOPEN	0	0.886
4	ラザニヤ	0	0.884
5	JR 大阪三越伊勢丹	1	0.822
6	ロドス用	0	0.809
7	やすみ	0	0.798
8	酒類	0	0.796
9	各線梅田	1	0.789
10	筑面ビール	0	0.785
11	各種お祝い	0	0.782
12	点心師	0	0.782
13	アニバーサリープラン	0	0.780
14	極み	0	0.773
15	パール	0	0.770
16	モダンいもすじ	1	0.768
17	団体席	0	0.767
18	グランド OPEN	0	0.763
19	甘み	0	0.758
20	大阪キタ	1	0.749

(c) 名古屋駅

順位	抽出語句	地域限定性	地域限定性スコア
1	名駅タワーズ	1	0.925
2	うどんすき	0	0.924
3	ピッツァ	0	0.862
4	装い	0	0.858
5	自家製生	0	0.848
6	JR セントラルタワーズ	1	0.814
7	出汁	0	0.807
8	名古屋名物きし	1	0.804
9	女子会	0	0.784
10	WEB 限定	0	0.781
11	コク	0	0.773
12	製法	0	0.772
13	季節会席	0	0.769
14	化学調味	0	0.730
15	細メン	0	0.726
16	最高	0	0.723
17	忠亭	0	0.723
18	焼き物	0	0.718
19	ぼんざい	0	0.716
20	彩り	0	0.711

サイトなど調べたりしてみたが、これらの語句を探し出すことは困難であった。このように、提案手法では、あまり知られていないような名物なども抽出することができるため、利用者に対して新たな気付きを与えるという効果も期待できる。

一方で、奈良では、「茶粥」や「柿の葉寿司」など有名であるが、提案手法ではこれらの語句は抽出できなかった。これは単純に今回対象としたスポットのテキスト情報 (PR 文) の中には、これらの語句が含まれていなかったことが原因である。対策としては、PR 文だけでなくスポットの詳細ページなどを対象テキスト情報に含めたり、情報源となるスポット情報サイトを拡張したりすることなどがあげられるが、これらについては別途検討課題とする。

(b) 三大都市

三大都市である、東京、大阪、名古屋を対象に実験を行った。対象地域としては、「東京駅」、「大阪駅」、「名古屋駅」を選んだ。これらを選定した理由としては、各都市の都道府県庁所在地名を駅名としているため、各都市の代表駅として連想しやすいと考えたからである。

表 5 に示すように、「名古屋駅」における「名古屋名物きし」を除いて、特にその地域ならではのものが抽出されたとはいえない。

これらの地域では、日本全国や各国の料理を扱っている店が多く、その地域ならではの料理を扱っている店は相対的に少ないといえる。実際に、表 2 に示すとおり、「東京駅」、「大阪駅」、「名古屋駅」を対象地域としたとき、抽出語句数に対する正解データ数は、それぞれ 418 件中 16 件、433 件中 10 件、309 件中 6 件であった。ただし、本来、不正解データについては地域限定性スコアが低くなるのが望ましいが、「ピモン冬」、「パティシエ特製ラウンドケーキ」、「うどんすき」などのように、不正解データに対して高い地域限定性スコアが与えられている。この点に関する考察については、次の (c) 都市の結果とあわせて後述する。

(c) 都市

都市部である、札幌、京都、広島、博多を対象に実験を行った。対象地域としては、「札幌駅」、「京都駅」、「祇園」、「広島駅」、「博多駅」を選んだ。

有効に抽出できた例としては、「京都駅」における「京湯葉そば」や「賀茂茄子」、「広島駅」における「広島かき」や「広島名物」などがあげられる。しかしながら、「札幌駅」、「京都駅」、「博多駅」、「広島駅」において、全体的にはその地域ならではのものを有効に抽出できたとはいえない。これらは先述した「東京駅」や「大阪駅」を対象にしたときと同様の理由によるものであるといえ、(b) 三大都市の結果とあわせて後述する。

一方で、同じ京都であっても、「祇園」を対象地域としたときには、上位 20 件中 15 件の正解データが抽出された。従来のご当地グルメサイトなどは、都道府県単位やある一

いるといえる。他にも、「奈良公園」における「正岡子規ゆかり」や「松阪駅」における「松阪商人」のような語句が抽出されている。このように、料理に関する語句だけでなく、その地域ゆかりの人物やランドマークなどが抽出されている点から、このような語句を利用者に提示することも有用であるといえる。

さらに、「松阪牛」や「伊勢海老」のように有名な特産物であれば、容易に検索することができるが、たとえば、「近江八幡駅」の「バームクーヘン豚」などは、あまり知られていないということもあり、その知識がない場合には検索することが困難である。実際に、「滋賀名物」や「近江八幡名物」などでキーワード検索を行ったり、ご当地グルメ

表 6 対象地域ごとの抽出語句上位 20 件 (対象地域種別: 都市)

Table 6 Top-20 phrases extracted in each local point (cities as a local type).

(a) 札幌駅				(b) 京都駅				(c) 祇園			
順位	抽出語句	地域限定性	地域限定性スコア	順位	抽出語句	地域限定性	地域限定性スコア	順位	抽出語句	地域限定性	地域限定性スコア
1	キウジーズ...	0	0.819	1	房 PURUDA	0	0.827	1	麩餅点心	1	0.993
2	近郊農家直送	0	0.750	2	本店 PURUDA	0	0.823	2	東大谷祖廟	1	0.953
3	LION ビザ	0	0.687	3	東九条上殿田	1	0.819	3	一流科学	0	0.841
4	あいそ	0	0.662	4	烏丸東寺道	1	0.809	4	京舞鑑賞	1	0.805
5	ミクニならでは	0	0.646	5	木津屋橋下ル...	1	0.801	5	高台寺茶寮	1	0.769
6	三國清	0	0.614	6	ステーキ定食	0	0.773	6	円山良葛原	1	0.739
7	札幌パセオ	1	0.561	7	下京	1	0.740	7	村井吾兵衛	1	0.732
8	札幌ステア...	1	0.554	8	しんそば	1	0.729	8	八坂神社	1	0.714
9	ミクニフレンチ	0	0.543	9	エンブレム京紋	0	0.723	9	北政所高台院	1	0.707
10	ステラプレイス	0	0.541	10	生ビール	0	0.714	10	北海道道北玄蕃麦	0	0.702
11	只今札	0	0.538	11	賀茂加子	1	0.713	11	往時	0	0.676
12	魚屋直営	0	0.535	12	南御徒歩	0	0.703	12	桃山びと	1	0.669
13	or 博多スープ	0	0.512	13	月替りランチ...	0	0.677	13	祇園円山	1	0.669
14	道内各地	1	0.502	14	粟そば	0	0.676	14	高台寺羽柴特製	1	0.668
15	化学調味	0	0.496	15	落ち着き	0	0.668	15	高台寺アイトアップ	1	0.668
16	遺産	1	0.467	16	京湯葉そば	1	0.661	16	石塀小路	1	0.662
17	無煙コースター	0	0.463	17	近鉄京都	1	0.648	17	本格京ゆ	1	0.660
18	鍋 or 刺身盛り	0	0.460	18	ブルダ	0	0.645	18	近江牛すきや	0	0.654
19	鶏持	0	0.444	19	京風味たっぶり	1	0.631	19	ふところ	0	0.641
20	茶壺本格焼酎	0	0.433	20	京大和	1	0.625	20	東山散策	1	0.628

(d) 博多駅				(e) 広島駅			
順位	抽出語句	地域限定性	地域限定性スコア	順位	抽出語句	地域限定性	地域限定性スコア
1	最上級白味噌ベース	0	0.933	1	広島大黒神島	1	0.916
2	ぶりっぶりっのモツ	0	0.894	2	ブレンナン広島	1	0.915
3	おやま	0	0.888	3	江田島沖	1	0.890
4	杜ひろば	1	0.883	4	広島名産	1	0.876
5	元祖博多豚兵衛	1	0.836	5	お寺	0	0.869
6	食材ひとつひとつ	0	0.816	6	イオンモール府中	1	0.867
7	アミューザ博多	1	0.808	7	生ビール	0	0.866
8	ひつどん甘味 JR 博多	1	0.800	8	ダン	0	0.850
9	鉄板バル	0	0.800	9	食材	0	0.836
10	岡垣	0	0.781	10	広島名物	1	0.830
11	リンガーハット自慢	0	0.769	11	広島かき	1	0.818
12	金目鯛柚子胡椒	0	0.762	12	コク	0	0.813
13	ちゃんぽん	0	0.758	13	各種ワイン	0	0.811
14	最大行事♪	0	0.736	14	広島ならではの	1	0.798
15	リンガースパイ	0	0.733	15	カキ	1	0.783
16	JR 博多シティ	1	0.730	16	たつ最大	0	0.782
17	コンパ	0	0.725	17	ピッツァ	0	0.781
18	仙台炭火焼	0	0.720	18	賀茂台地	1	0.778
19	特典消費	0	0.713	19	マノマッジョ	0	0.775
20	ボリュームたっぶり	0	0.711	20	エビスビール	0	0.765

定の規模を持ったエリア単位での名物を紹介するものが多い。これに対し、提案手法では、「京都駅」や「祇園」のように対象地域をピンポイントで指定することで、その対象地域の情報を取得できるという利点がある。

以上のように、提案手法による語句抽出性能について、(a) 観光地、(b) 三大都市、(c) 都市という対象地域種別ごとに傾向分析を行った。結果、提案手法は「伊勢神宮」や「奈良公園」といった観光地に適用することで、特に効果的に働くということが確認できた。また、「伊勢神宮」と「松阪駅」、「京都駅」と「祇園」のように、近傍するエリアにおいても、ピンポイントで対象地域を指定することで、その対象地域の情報を取得できるという利点も確認できた。

ただし、三大都市および都市においては、不正解データに対しても高い地域限定性スコアが与えられている。この点について観光地と比較しながら考察する。地域限定性スコアは、式(5)のとおり、 $IDF$  と地域関連重みを掛け合わせることで算出される。このうち  $IDF$  は対象地域に依存しないため、ここでは三大都市において各上位 20 件に抽出された不正解データに対して、地域関連重みがどのように算出されるかに着目する。

まず、表 5 に示すように「東京駅」、「大阪駅」、「名古屋駅」を対象地域としたとき各上位 20 件に抽出された不正解データ(「ビモン冬」、「パティシエ特製ラウンドケーキ」、

「うどんすき」など)を対象に、表 2 に示した各対象地域の市区町村名との地域関連重みを算出した。表 7 に得られた地域関連重み(正規化前)を対象地域ごとに平均したものを示す。すると、比較的、観光地における平均地域関連重みに比べ、三大都市、都市における平均地域関連重みの方が高くなる傾向があるといえる。このことを客観的に示すために、目安ではあるが、本実験で用いたデータセットから各対象地域の市区町村に存在する飲食店データ数を調べた。その結果を表 7 に記載している。ここで、平均地域関連重みと市区町村に存在する飲食店データ数との相関係数を算出すると 0.675 となり、中程度の相関があることが確認できた。

今回、データセットとしてぐるなびを対象としたことから、抽出語句の上位には、「ラザーニャ」や「酒類」など飲食店に関連する語句が多く抽出される。このような語句は Web 上で飲食店の紹介ページに現れることが多く、通常飲食店の住所とともに掲載されることが多いといえる。こういったことから、Web ページ上での対象語句と対象地域の市区町村名との共起頻度に基づいて地域関連重みを算出する場合、その対象地域の市区町村における飲食店の絶対数が多いほど、対象語句と共起する頻度は必然的に高くなる。実際に、Google による検索を行った場合、たとえば「パティシエ特製ラウンドケーキ」と「大阪市」が共起するページ数は約 33,000 件である一方で、「伊勢市」と共起す

表 7 三大都市において上位 20 件に抽出された不正解データと各対象地域との地域関連重み  
 Table 7 Regional weight of irrelevant data in top-20 phrases in the three largest cities for each local point.

対象地域	対象地域種別	市区町村	平均地域関連重み	対象地域に存在する飲食店データ数
伊勢神宮	観光地	伊勢市	6.568	28
松阪駅	観光地	松阪市	5.832	35
奈良公園	観光地	奈良市	6.916	245
近江八幡駅	観光地	近江八幡市	5.654	28
函館駅	観光地	函館市	6.397	61
東京駅	三大都市	千代田区	7.181	1929
大阪駅	三大都市	大阪市	7.887	5859
名古屋駅	三大都市	名古屋市	7.265	1912
札幌駅	都市	札幌市	6.968	1356
京都駅, 祇園	都市	京都市	7.828	1728
博多駅	都市	福岡市	7.548	1303
広島駅	都市	広島市	7.515	644

るページ数は約 5,000 件程度であった。三大都市や都市部など飲食店の絶対数が多い対象地域ほど、飲食店に関連する語句の地域関連重みが高くなる傾向にあり、結果として地域限定性スコアが高くなっているといえる。

以上のことをふまえ、今後は、対象地域の飲食店の絶対数が多い場合などには、地域関連重みに加えて別の尺度を追加したりするなどの検討が必要であると考え。なお、実際には、Web 上にはぐるなびだけでなく Hotpepper や食ベログなどの飲食店情報、さらには飲食店以外にもさまざまな店舗や施設情報があるため、これらのような外部の情報も含めて考慮に入れる必要がある。

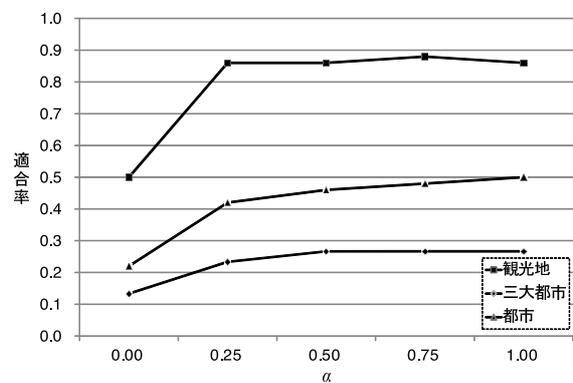
ただし、飲食店の絶対数が多い都市においても、祇園のように元々の正解データ数が多い場合には、相対的に正解データに対する地域限定性スコアが高くなり、提案手法により地域限定語句が有効に抽出できているといえる。

### 5.5 パラメータ $\alpha$ , $\beta$ の感度分析

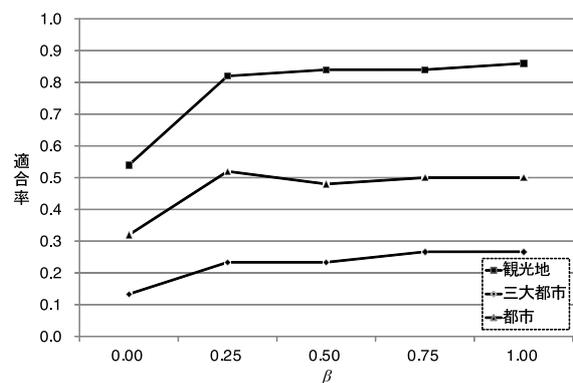
3.4 節に示した式 (5) のパラメータ  $\alpha$ ,  $\beta$  の感度分析を行った。ここでは、各パラメータの値を {0.00, 0.25, 0.50, 0.75, 1.00} に変化させたときの語句抽出性能の違いをみる。ここで語句抽出性能としては、抽出された上位 10 件の語句に含まれる正解データを適合データとしたときの適合率および再現率を用いた。各対象地域における全正解データ数は、表 2 に示すとおりである。

図 3 および図 4 に、(a)  $\alpha$  を変化させたとき ( $\beta = 1$  とする)、(b)  $\beta$  を変化させたとき ( $\alpha = 1$  とする)、それぞれの適合率および再現率の変化を示している。図 3、図 4 は、対象地域種別ごとに示しており、各対象地域における適合率、再現率の平均をとっている。

図 3(a)、図 4(a) に示す結果について、対応のある  $t$ -検定を行った結果、観光地における適合率について、 $\alpha = 0.00$



(a)  $\alpha$  を変化させたときの適合率



(b)  $\beta$  を変化させたときの適合率

図 3 対象地域種別ごとのパラメータ  $\alpha$ ,  $\beta$  を変化させたときの適合率の違い

Fig. 3 Difference of precision by parameters  $\alpha$ ,  $\beta$  in each local type.

と  $\alpha = \{0.25, 0.50, 0.75, 1.00\}$  との差、三大都市における適合率について、 $\alpha = 0.00$  と  $\alpha = \{0.50, 0.75, 1.00\}$  との差、都市における適合率について、 $\alpha = 0.00$  と  $\alpha = \{0.50, 0.75, 1.00\}$  との差に、それぞれ有意水準 5% で有意差がみられた。また、観光地における再現率について、 $\alpha = 0.00$

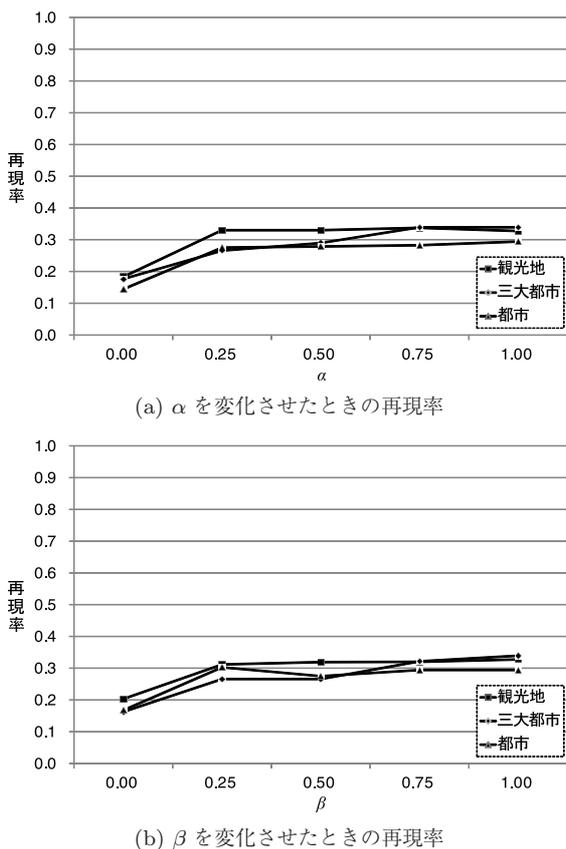


図 4 対象地域種別ごとのパラメータ  $\alpha, \beta$  を変化させたときの再現率の違い

Fig. 4 Difference of recall by parameters  $\alpha, \beta$  in each local type.

と  $\alpha = \{0.25, 0.50, 0.75, 1.00\}$  との差, 三大都市における再現率について,  $\alpha = 0.00$  と  $\alpha = \{0.75, 1.00\}$  との差, それぞれ同様の有意差がみられた. この結果から, いずれの対象地域種別においても, 適合率, 再現率の観点から,  $\alpha \geq 0.25$  とすることが望ましいと考えられる. また, 特に  $0.25 \leq \alpha \leq 1.00$  においては有意差がみられなかったことから, 前節までの実験において  $\alpha = 1$  を採用していた点については妥当であったといえる.

図 3(b), 図 4(b) は, 地域関連重み  $\omega$  の尺度の重みを表すパラメータ値を変化させたときの適合率, 再現率の変化を示したものである. 同様に, 対応のある  $t$ -検定を行った結果, 観光地における適合率について,  $\beta = 0.00$  と  $\beta = \{0.25, 0.50, 0.75, 1.00\}$  との差, 都市における適合率について,  $\beta = 0.00$  と  $\beta = \{0.25, 0.50, 0.75, 1.00\}$  との差に, それぞれ有意水準 5% で有意差がみられた. また, 観光地における再現率について,  $\beta = 0.00$  と  $\beta = \{0.25, 0.50, 0.75, 1.00\}$  との差, それぞれ同様の有意差がみられた. この結果から, いずれの対象地域種別においても, 適合率, 再現率の観点から,  $\beta \geq 0.25$  とすることが望ましいと考えられる. また, 先述の  $\alpha$  のケースと同様, 特に  $0.25 \leq \beta \leq 1.00$  においては有意差がみられなかったことから, 前節までの実験において  $\beta = 1$  を採用していた点については妥当であった

といえる.

以上のように, 今回の実験結果からでは, 観光地, 三大都市, 都市という対象地域種別の違いによらず,  $\alpha = \beta = 1$  と固定してシステム設計を行うことは問題ではないと考えられる.

### 5.6 IDF の閾値を設定したときの語句抽出性能の比較

5.3.2 項において, 式 (5) により表される地域限定性スコアを用いることの有用性について確認した. ただし, 式 (5) では,  $IDF$  と地域関連重み  $\omega$  を掛け合わせることで, 地域限定性スコアを算出していたが, 表 3 から, 単純に適切な  $IDF$  の閾値を設定することで抽出語句のフィルタリングを行う方法も考えられる.

そこで, 本節では次式のように  $IDF$  の閾値  $\theta$  を設定した地域限定性スコアを定義し, これを用いた実験を行う.

$$s_j = \begin{cases} \omega_j & IDF \geq \theta \\ 0 & IDF < \theta \end{cases} \quad (7)$$

式 (7) において, 閾値  $\theta$  を変化させたとき, 語句抽出性能にどのような影響が現れるか分析を行った.

ここでは, 5.5 節と同様に, 上位 10 件の適合率・再現率を語句抽出性能として用いた. 図 5 は対象地域種別ごとに,  $\theta = \{0.0, 0.1, \dots, 0.9\}$  と変化させたときの適合率・再現率を示している. 図 5 より, 観光地においては,  $\theta = 0.8$  のとき適合率, 再現率ともに最大となった. 都市においては,  $\theta = \{0.8, 0.9\}$  のとき適合率が,  $\theta = 0.9$  のとき再現率が, それぞれ最大となった. 三大都市においては,  $\theta = 0.4$  のとき適合率が,  $\theta = 0.9$  のとき再現率が, それぞれ最大となった. このように適切な  $IDF$  の閾値を設定することにより, 語句抽出性能を向上させることが期待できる.

式 (7) を用いた際に得られた適合率・再現率を, 式 (5) を用いた際の適合率・再現率と比較した. 図 5 には, 各対象地域において, 式 (5) ( $\alpha = \beta = 1$ ) を用いた際の適合率・再現率を破線で示している. ■で示した破線, ◆で示した破線, ▲で示した破線は, それぞれ観光地, 三大都市, 都市を対象地域としたときの適合率・再現率を示している. 観光地において  $\theta = 0.8$  としたときに, 式 (5) による適合率・再現率にほぼ同等であったことを除いては, いずれも式 (5) による適合率・再現率を下回った.

実際に, 表 4 に示した語句のうち, 閾値  $\theta = 0.8$  としたときに  $s_j = 0$  となった語句は, 「伊勢湾 ( $IDF = 0.788$ )」, 「松坂 ( $IDF = 0.578$ )」, 「奈良公園散策 ( $IDF = 0.699$ )」, 「近江八幡 ( $IDF = 0.678$ )」, 「近江牛料理 ( $IDF = 0.745$ )」, 「秋近江 ( $IDF = 0.790$ )」などであった. このように閾値  $\theta$  の設定の仕方によっては, 式 (5) では上位に抽出された語句が, 抽出されないという可能性が生じる. そのため, 式 (7) を用いる際には, 閾値  $\theta$  の選び方が課題となる. 一方で, 式 (5) を用いる際には, 閾値の設定は不要である.

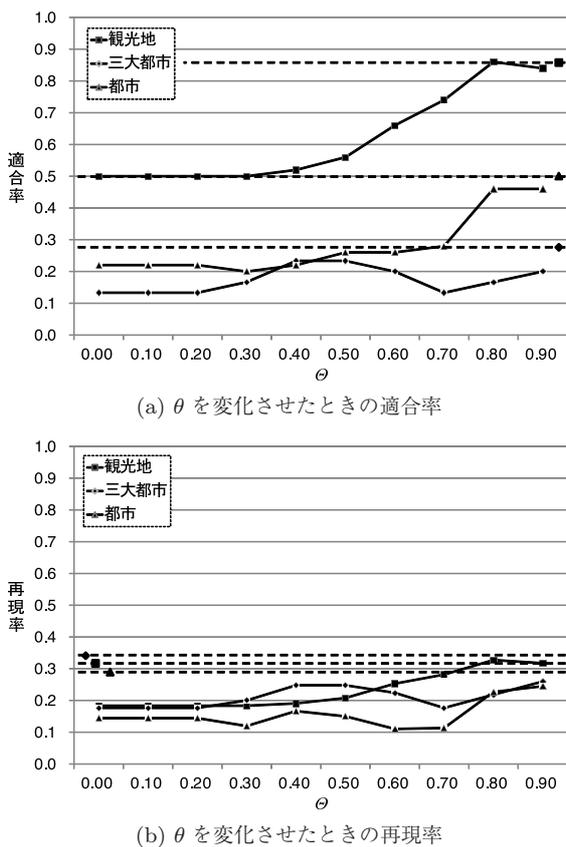


図 5 対象地域種別ごとの閾値  $\theta$  を変化させたときの適合率・再現率の違い

Fig. 5 Difference of precision and recall by threshold  $\theta$  in each local type.

また 5.5 節で示したように、対象地域によらず  $\alpha = \beta = 1$  と固定しても、特に問題ではないことも確認している。そこで、本稿では、閾値  $\theta$  を設定しなくても高い語句抽出性能を得られた点、実質パラメータ調整が不要である点をふまえ、式 (5) に基づく地域限定性スコアを採用する。

## 6. 被験者実験による有用性評価

4 章で説明したプロトタイプシステムである地域限定グルメ推薦システム（以降、提案システム）を用いて提案手法の有用性を評価した。ここでは、既存のグルメサイトのスポット情報を基に、地域特有のスポットを推薦する機能を付加した提案システムを、元のグルメサイトと比較した。グルメサイトとして、国内の大手グルメサイトである、ぐるなびを選んだ。被験者は、5.2 節で述べた被験者と同一であり、詳細な被験者属性は、付録 A.6 に示している。

### 6.1 実験手順

提案システムの操作方法を被験者に教示した後、以下の手順で実験を行った。

- i 次のタスクの中から、被験者に遂行したいタスクを選んでもらった。  
タスク A: 「XXXX に旅行または出張に来ました。各

システム（地域限定グルメ推薦システム、ぐるなび）を利用して、行きたいと思った飲食店の候補を 1-3 件ずつ探してください。」

タスク B: 「今日は休日で、自宅付近で外食したいと思います。各システム（地域限定グルメ推薦システム、ぐるなび）を利用して、行きたいと思った飲食店の候補を 1-3 件ずつ探してください。」

- ii 被験者に、対象地域として想定する「都道府県」および「市区町村」を選択してもらった。ここで選択された対象地域はタスク A 中の XXXX の部分に相当する。タスク B を選択した場合は、被験者自身の居住地を選択してもらった。
- iii 被験者に、地域限定グルメ推薦システムおよびぐるなびの両方を利用して、上記タスクを遂行してもらった。ここで、システムを利用する順序は被験者任意としている。
- iv 上記タスク遂行中に見つけた飲食店（最大 6 件）について、1 位から 6 位の順序を付けてもらった。

以上の手順を 1 タスク、1 対象地域ごとに行った。希望する被験者には、引き続きタスクまたは対象地域、もしくはその両方を変えて、同様の手順を行ってもらった。ここで、実行するタスク数および対象地域数は、被験者任意とした。実験終了後、被験者には、被験者属性および提案システムに関するアンケートに回答してもらった。

### 6.2 ぐるなびとの比較

10 名の被験者から合計 15 回（タスク A: 8 回、タスク B: 7 回）の試行の回答が得られた。表 8 に、15 回の試行の回答を示す。タスク A の対象地域として「三重県松阪市」や「奈良県生駒市」など、タスク B の対象地域として「滋賀県近江八幡市」や「兵庫県西宮市」などが選択された。たとえば、被験者 b は、タスク A において回答しており、提案システムにより見つけられた 3 件の飲食店に対しては、それぞれ 1 位、2 位、4 位の順位を与えている。一方で、ぐるなびにより見つけられた 3 件の飲食店に対しては、それぞれ 3 位、5 位、6 位の順位を与えている。また、表には、各被験者のシステムの利用順序を示している。たとえば、被験者 b は、本実験においては、ぐるなびを利用した後、提案システムを利用している。

表 8 のとおり、タスク A において、提案システムの方がぐるなびに比べ平均順位が上位であることが分かる。ここで、同一被験者により与えられた、提案システムに対する平均順位とぐるなびに対する平均順位の組を対応のあるデータと見なしたとき、対応のある  $t$ -検定を行った結果、両者には有意水準 1% で差があることを確認した。一方で、タスク B においては平均順位に有意差はみられなかった。

以上の結果より、提案システムは旅行時や出張時の飲食店選びにおいて、有用性があるといえる。このように提案

表 8 タスクごとの各システムにより見つけられた飲食店に対する順位  
Table 8 Ranking of restaurants found by each system in each task.

(a) タスク A

被験者	タスク	提案システム (I) により見つけられた飲食店に対する順位 (1~6 位)				ぐるなび (II) により見つけられた飲食店に対する順位 (1~6 位)				システムの利用順序
		飲食店 1	飲食店 2	飲食店 3	平均順位	飲食店 1	飲食店 2	飲食店 3	平均順位	
b	A	1	2	4	2.3	3	5	6	4.7	II → I
e	A	1	2	5	2.7	4	6	3	4.3	II → I
e	A	2	1	6	3.0	5	3	4	4.0	I → II
g	A	1	3	5	3.0	6	2	4	4.0	II → I
h	A	1	3	-	2.0	2	4	-	3.0	I → II
g	A	1	3	4	2.7	2	6	5	4.3	I → II
i	A	1	-	-	1.0	2	-	-	2.0	I → II
c	A	1	-	-	1.0	2	-	-	2.0	II → I
平均					2.2				3.5	

(b) タスク B

被験者	タスク	提案システム (I) により見つけられた飲食店に対する順位 (1~6 位)				ぐるなび (II) により見つけられた飲食店に対する順位 (1~6 位)				システムの利用順序
		飲食店 1	飲食店 2	飲食店 3	平均順位	飲食店 1	飲食店 2	飲食店 3	平均順位	
d	B	1	5	4	3.3	3	2	6	3.7	I → II
a	B	6	4	1	3.7	5	2	3	3.3	I → II
c	B	1	-	-	1.0	2	-	-	2.0	I → II
e	B	1	4	2	2.3	3	5	6	4.7	II → I
h	B	2	3	-	2.5	1	4	-	2.5	II → I
j	B	1	5	6	4.0	3	2	4	3.0	II → I
f	B	1	6	5	4.0	3	2	4	3.0	I → II
平均					3.0				3.2	

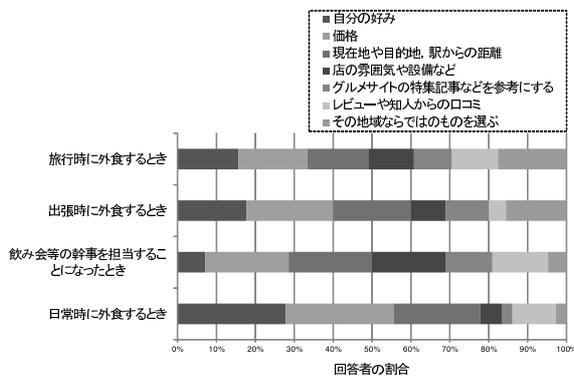


図 6 状況別の飲食店選択基準に関する調査

Fig. 6 Survey of restaurant selection criteria in different situation.

システムによる地域特有のスポットを推薦する機能を既存のグルメサイトに付加することで、有用性をより高めることができるといえる。本結果と関連する調査として、被験者に状況別の飲食店選択基準に関する調査を行った。その結果を図 6 に示す。図 6 のとおり、旅行時や出張時においては、日常時に比べ、「その地域ならではのものを選ぶ」割合が高くなっている。この結果は、先の表 8 の結果と一致しているといえる。提案システムの位置付けとしては、このような地域ならではのものを選びたいという利用者の要求に対応するものである。

なお、旅行時や出張時においても、利用者の好みや価格、距離などを基準に選ぶという利用者は多いが、このような利用者の要求に対応したサービスやシステムはすでに多く存在する。今後、これら既存のサービスやシステムと連携させることで、システム全体としての利用者満足度向上につながると期待できる。

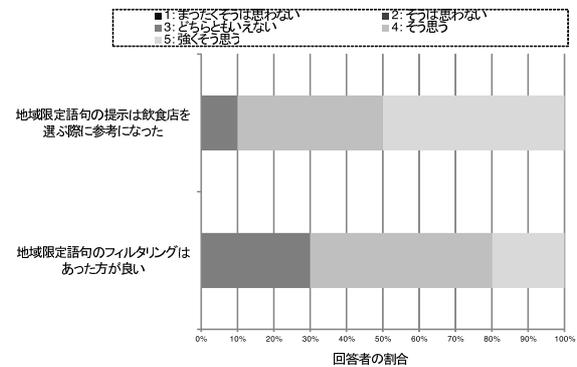


図 7 地域限定グルメ推薦システムに関する調査

Fig. 7 Survey of region-restricted restaurant recommender system.

### 6.3 提案システムに関するアンケート結果

提案システムに関して、被験者にアンケート調査を行った。その結果を、図 7 に示す。図 7 より、利用者に地域限定語句を提示することは有用であるといえる。また、今回のシステムでは利用者がフィルタリング指定を行うことで提示する語句をフィルタリングすることを可能としたが、この機能についても「あった方がよい」と回答する被験者が多かった。ただし、被験者からは、「フィルタリング指定を用いても『オープンです』『全額贈呈』などの直接店を選ぶ際の情報としては役に立たない語句がリストに表示されてしまった」といった指摘があった。今後、フィルタリング規則について検討していくことで、利用者満足度の向上を目指す必要がある。

また、被験者による自由記述からは、「ぐるなびなどでは下位に表示されるものが上位に表示されることもあり、『隠れた名店』を探しやすい」や「タスク B において、地元で知らなかった泉州バーガーについて知られたのでよかった」という回答もみられた。これらの点から、既存のグル

メサイトに提案システムの機能を付加することにより、利用者が知らなかった語句についても、提示することで気付きを与えることができていると考えられる。一方で、対象地域によっては、抽出されてほしい語句が抽出されないという回答もあった。原因としては、そもそもぐるなびに登録されていなかったことなどがあげられる。この点については、他のグルメサイトや位置情報付きコンテンツにもデータ収集対象を拡大することで、対応していく必要があると考えている。

その他、自由記述により得られた回答については、付録A.7に掲載している。

## 7. おわりに

本稿では、地域限定語句を抽出するための尺度として、地域限定性スコアを提案した。地域限定性スコアは、*IDF* および地域関連重みの尺度の組合せにより定義される。評価実験では、ぐるなびから収集された62,571件のスポットデータを用いて、地域限定性スコアによる語句抽出性能について、定量分析および定性分析を行った。その結果、ベースライン手法である *IDF* に比べ、提案の地域限定性スコアではランキング上位に正解データが抽出されるなど有用性を示した。また、実際の地域として、観光地、三大都市、都市を対象地域とした実験を行い、対象地域による抽出語句の違いや傾向について分析を行った。結果、提案手法は特に観光地を対象としたときに、より有効に働くということを示した。

また、既存のグルメサイトのスポット情報を基に、地域特有のスポットを推薦する機能を付加した提案システムを、元のグルメサイトと比較した。その結果、旅行時や出張時というシチュエーションにおいては、提案システムが有効に働くことを示した。さらに、被験者に提案システムに関するアンケート調査を行った結果、「「隠れた名店」を探しやすい」や「地元で知らなかった泉州バーガーについて知られたのでよかった」などの回答が得られた。この回答から、提案システムにより、利用者が知らないような地域限定語句に気付かせることができているといえる。

これまで観光情報に関しては、観光情報誌や観光情報サイト、ご当地グルメサイトなど、多くの情報源から入手可能ではあるが、これらはコンテンツを人手で作成する必要があるため、観光情報の整備および保守を行うためには大きなコストがかかる。これらと比べたとき、提案手法では、対象地域が指定されたとき、そこから自動的に地域限定語句を抽出できるため、次のような利点がある。

- Web から収集するスポットデータをつねに更新しておくことで、つねに最新の情報を抽出することができる。
- 「近江八幡駅」における「バームクーヘン豚」などのように、あまり知られていないような情報も抽出可能であり、このような語句を提示することで利用者にとっ

て新たな気付きを与えることができる。一方で、観光情報誌などでは知名度の低いような情報は掲載されないことが多い。また、Web 検索を用いても、「近江八幡名物」など容易に思い付くキーワードによる検索では、発見することは困難である。

- 「京都駅」と「祇園」といったように比較的近いエリアにおいても、それぞれの周辺のスポットデータのみを参照するため、対象としている対象地域ピンポイントの地域限定語句の抽出が可能である。一方で、観光情報誌などでは、都道府県単位やある一定の規模を持ったエリア単位など、作成者により定義されたエリアでの名物を紹介するものが多い。

今回は情報源としてぐるなびを、テキスト情報としてスポットのPR文を対象としたが、今後は、テキスト情報としてユーザレビューを選定するなど、情報源を拡張することで、より多角的な分析を行う。また、利用者の地元地域を考慮することで、地元にはなく対象地域にはあるような相対的な限定性に着目したり、利用者の嗜好を考慮したりするなど、提案手法の個人化について検討する。

**謝辞** 本研究の一部は、文部科学省科学研究費補助金若手研究(B)「コンテキスト限定価値を考慮した情報推薦方式」(研究代表者:奥健太, 課題番号23700132)による。ここに記して謝意を表します。

## 参考文献

- [1] Abowd, G.D., Atkeson, C.G., Hong, J., Long, S., Kooper, R. and Pinkerton, M.: Cyberguide: A mobile context-aware tour guide, *Wireless Networks - Special issue: mobile computing and networking: selected papers from MobiCom '96*, Vol.3, No.5, pp.421-433 (1997).
- [2] Sumi, Y., Etani, T., Fels, S., Simonet, N., Kobayashi, K. and Mase, K.: C-MAP: Building a Context-Aware Mobile Assistant for Exhibition Tours, *Community Computing and Support Systems, Social Interaction in Networked Communities*, Vol.1519, pp.137-154 (1998).
- [3] Cheverst, K., Davies, N., Mitchell, K. and Friday, A.: Experiences of Developing and Deploying a Context-Aware Tourist Guide: The GUIDE Project, *The 6th Annual International Conference on Mobile Computing and Networking (Mobi-Com2000)*, pp.20-31 (2000).
- [4] 奥健太, 中島伸介, 宮崎純, 植村俊亮: 状況依存型ユーザ嗜好モデリングに基づく Context-Aware 情報推薦システム, *情報処理学会論文誌: データベース*, Vol.48, No.SIG 11 (TOD34), pp.162-176 (2007).
- [5] Kotera, R., Kitayama, D., Oku, K. and Sumiya, K.: Geographical Recommendation Method using User's Interest Model based on Map Operation and Category Selection, *Proc. 5th International Conference on Ubiquitous Information Management and Communication (ICUIMC2011)*, pp.3-12 (2011).
- [6] Tarumi, H., Morishita, K. and Kambayashi, Y.: Public applications of spacetag and their impacts, *Digital cities: technologies, experiences and future perspectives, Lecture Notes in Computer Science (Stateof- the Art Survey)*, Vol.1765, pp.350-363 (2000).
- [7] 森下健, 中尾恵, 垂水浩幸, 上林弥彦: 時空間限定

オブジェクトシステム SpaceTag: プロトタイプシステムの設計と実装, 情報処理学会論文誌, Vol.41, No.10, pp.2689-2697 (2000).

- [8] Okamoto, M. and Kikuchi, M.: Discovering Volatile Events in Your Neighborhood: Local-Area Topic Extraction from Blog Entries, *Asia Information Retrieval Symposium*, LNCS 5839, pp.181-192 (2009).
- [9] 岡本昌之, 菊池匡晃: 3. ブログからの地域イベント情報抽出 (特集) 社会に浸透する新たなコンピュータ/ネットワークの世界, 情報処理, Vol.51, No.1, pp.14-17 (2010).
- [10] 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, 知能と情報 (日本知能情報フェジ学会誌), Vol.22, No.6, pp.667-679 (2010).
- [11] 手塚太郎, 近藤浩之, 田中克己: 混合ガウス分布を用いたウェブコンテンツの地域性推定とオブジェクトレベルローカルサーチ, 情報処理学会論文誌: データベース (TOD), Vol.1, No.1, pp.13-25 (2008).
- [12] 北 研二, 津田和彦, 獅々堀正幹: 情報検索アルゴリズム, 共立出版 (2002).
- [13] Bollegala, D., Matsuo, Y. and Ishizuka, M.: Measuring semantic similarity between words using web search engines, *WWW '07: Proc. 16th International Conference on World Wide Web* (2007).

## 付 録

### A.1 検索スポット数と応答時間との関係

表 2 に示した 13 件の対象地域において, それぞれ検索スポット数  $k$  を {10, 20, 30, 50, 100} と変えたときの応答時間を計測した. 実行環境として, Windows 7 Professional 64 bit OS (Intel Core 2 Duo CPU, 6.00 GB メモリ) を用い, ブラウザとして Google Chrome 19.0 を用いた. 図 A.1 は, 各条件における応答時間を 10 回ずつ計測した平均を示している. ただし, 「伊勢神宮」, 「松阪駅」, 「近江八幡駅」, 「函館駅」においては, 対象地域に存在するスポット数がそれぞれ, 25 件, 26 件, 26 件, 43 件であったため, 「伊勢神宮」, 「松阪駅」, 「近江八幡駅」においては 20 件までの, 「函館駅」においては 30 件までの応答時間を計測している. この結果, 検索スポット数  $k$  が多いほど応答時間がかかることが確認でき, 特に  $k = 100$  のとき, 4.5 秒以上の応答時間がかかることが確認された.

### A.2 地域限定語句リストのフィルタリング

地域限定語句のフィルタリング規則を表 A.1 に示す. 表にはフィルタリング規則を正規表現を用いて表している. ここで, 「\*」は「任意の 0 文字以上の文字列」を意味する. フィルタリング指定がされている場合は, 地域限定語句リストを提示する際に規則に該当する語句がリストから除外される.

### A.3 実験で取得したスポットデータの例

実験で取得したスポットデータの例を表 A.2 に示す.

### A.4 他の対象地域における地域限定性スコアによる正解・不正解データの度数分布

「松阪駅」, 「奈良公園」, 「近江八幡駅」, 「函館駅」を対象地域としたときの, 地域限定性スコアによる正解・不正解データの度数分布を図 A.2 に示す. 図のとおり, スコア上位においては, 不正解データが数件混在しているものの, ほぼ正解データのみを抽出できていることが分かる.

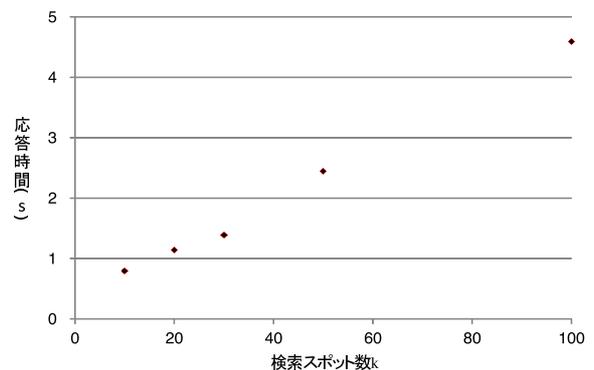


図 A.1 検索スポット数と応答時間との関係  
Fig. A.1 Relationship between the number of search spots and response time.

表 A.1 地域限定語句のフィルタリング規則

Table A.1 Filtering rules for region-restricted phrases.

カテゴリ	フィルタリング規則
駅関連	{*駅, *駅前, *ターミナル, *バス停}
地区関連	{*市場, *街, *アーケード, *地区, *横丁}
通り・位置関連	{*通り, *通, *沿い, *裏, *側, *周辺, *交差点, *交差点*, *横, *付近, *部, *前, *口, *方面}
施設関連	{*ホテル*, *ビル*, *庁*, *駐車*}
店舗関連	{*屋, *店, *店内, *店舗*, *グループ*}
サービス・設備関連	{半額*, *サポート*, *座席*, *スタイル*, *個室*, *貸切*, *割引*, *クーポン*, *テイクアウト*, *予約*, *サービス*, *代金*, *価格*, *専用*, *キャンペーン*}
イベント関連	{*宴会*, *シーズン*, *イベント*, *パーティ*, *忘年会*, *Party*, *デー, *オープン*, *日, *リニューアル*, *タイム*}

表 A.2 実験で取得したスポットデータの例  
Table A.2 An example of spot data used in experiments.

属性		値の例
スポット名	飲食店名	割烹 大喜
位置情報	経度・緯度	136.713167, 34.489014
テキスト情報	PR 文 (短)	伊勢志摩の旬の素材を使った郷土料理が自慢。近鉄宇治山田駅前で交通至便。予約なしでも可。宮内庁御用達。
	PR 文 (長)	創業 60 年の老舗。近鉄宇治山田駅前で交通に便利。伊勢志摩の旬の食材、伊勢海老・鮑・サザエ等を使った郷土料理が自慢。他寿司・幕の内・丼物も気軽に味わえる。カウンターお一人様から宴会 70 名様まで可能。会席料理は 5,000 円から。他にも 1,500 円から天ぷら定食や和定食などもある。英語のメニューも完備。外人客も多い。伊勢神宮外宮へは徒歩 10 分、内宮へは車で 15 分の好立地にある。宮内庁・神宮司庁御用達の老舗。

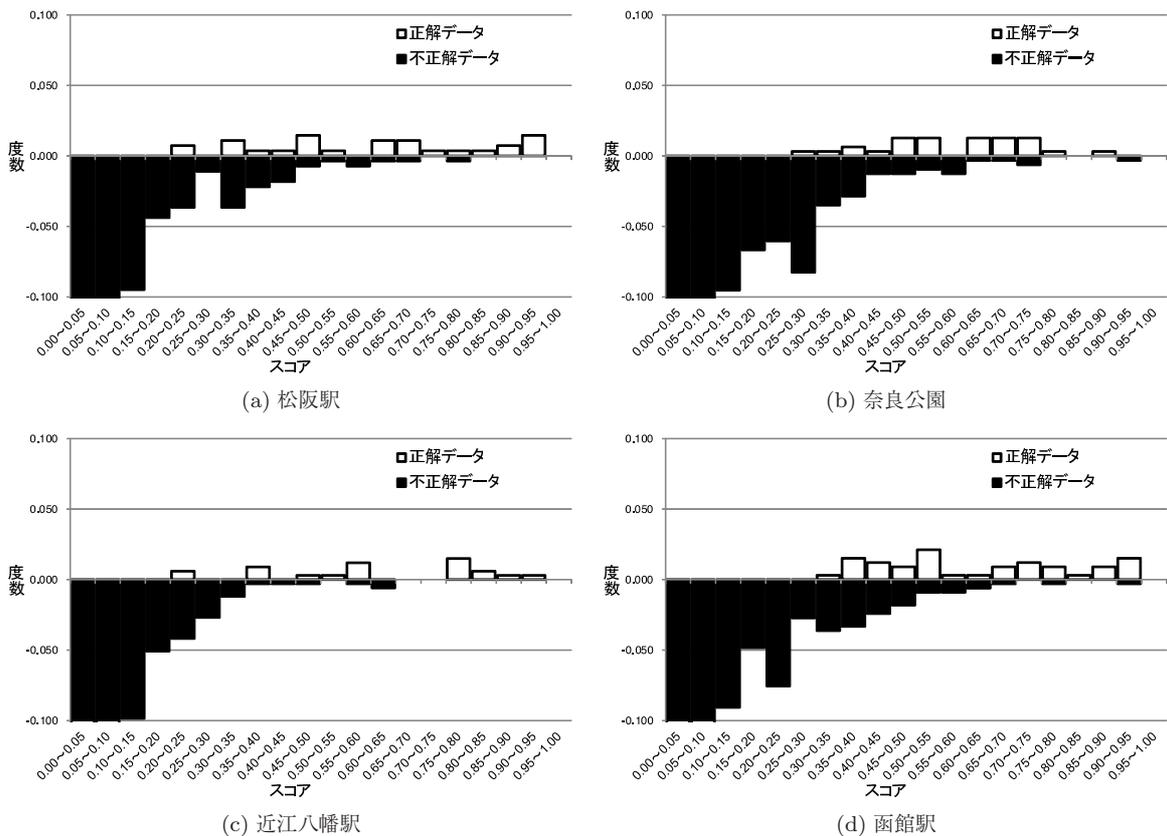


図 A.2 他の対象地域における地域限定性スコアによる正解・不正解データの度数分布 (対象地域:「松阪駅」,「奈良公園」,「近江八幡駅」,「函館駅」)

Fig. A.2 Frequency distribution of relevant/irrelevant data by region-restrictedness score in other local points (“Matsusaka Sta.,” “Nara Park,” “Ohmi Hachiman Sta.,” and “Hakodate Sta.” as local points).

### A.5 境界地域を対象地域としたときの地域限定性スコアの比較

境界地域を対象地域としたときの地域限定性スコアのの違いを確認するため、「奈良市」と「木津川市」との境界で地域限定語句を取得した。対象地域の中心経度・緯度  $(x, y)$  を、それぞれ奈良市側:  $(135.812104, 34.716899)$ , 木津川市側:  $(135.8122428, 34.71728518)$  とし、中心からの半径  $r$  を 3,000 m とした。また検索スポット件数  $k$  は 100 件としている。これらの地域は、両対象地域周辺に 20 件以上の現

地スポットが取得できること、両対象地域に存在する現地スポット数が同等であること、また取得される現地スポット集合が両対象地域で同一であることをふまえて選定した。なお、取得された現地スポット数は 24 件であり、そのうち奈良市側に存在する現地スポット数は 12 件、木津川市側に存在する現地スポット数も 12 件であった。

表 A.3 は、各対象地域で抽出された上位 10 件の語句を示し、奈良市側、木津川市側で取得したときの地域限定性スコアを掲載している。このように、境界地域で地域限定語句を取得した場合にも、地域限定性スコアに与える影響

はないことを確認した。

### A.6 評価実験における被験者属性

正解データの作成および提案システムの評価実験においては、被験者 10 名（男性 8 名，女性 2 名）に協力してもらった。被験者は 20 歳から 30 歳の学生および会社員である。

被験者の料理ジャンルに対する嗜好の分布は図 A.3 のとおりである。この料理ジャンルはぐるなびに登録されているものを参考にした。

また、被験者のインターネット利用頻度およびグルメサイト利用頻度，外出頻度は，図 A.4 のとおりである。インターネットは全被験者が定期的に利用しているが，グルメサイトはごくまれに（これまでに数回程度）といった被験者が多かった。また，外出頻度については，適度に（週

表 A.3 境界地域を対象地域としたときの地域限定性スコアの比較（対象地域：「奈良市」，「木津川市」）

Table A.3 Comparison between region-restricted scores in boundary areas (“Nara-city” and “Kizugawa-city” as local areas).

抽出語句	地域限定性スコア	
	奈良市側	木津川市側
ガーデンモール木津川向かい	1.000	1.000
木津 IC	0.973	0.973
アンナプルナ	0.911	0.911
東大寺北角	0.897	0.897
高の原	0.889	0.889
大和西大寺	0.820	0.820
籠松明	0.799	0.799
食材ひとつひとつ	0.796	0.796
かごたいまつ	0.793	0.793
正岡子規ゆかり	0.784	0.784

に 1-3 回程度）という回答が多かった。

### A.7 自由記述による被験者からの回答

提案システムの評価実験において，被験者に選択された対象地域は表 A.4 のとおりである。

タスク A については，「旅行や出張などで行ったことがある」，「住んでいたことがある」，「あまり知らない」地域

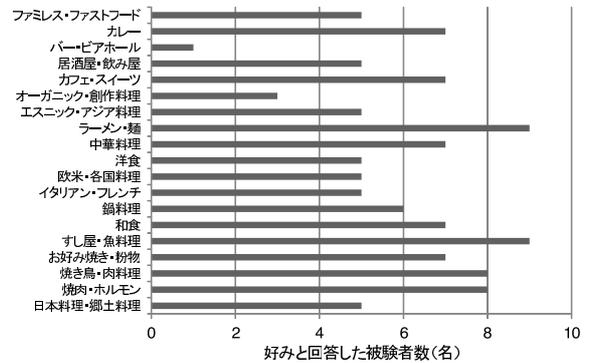


図 A.3 被験者の料理ジャンルに対する嗜好  
Fig. A.3 Subjects' tastes for cuisine genres.

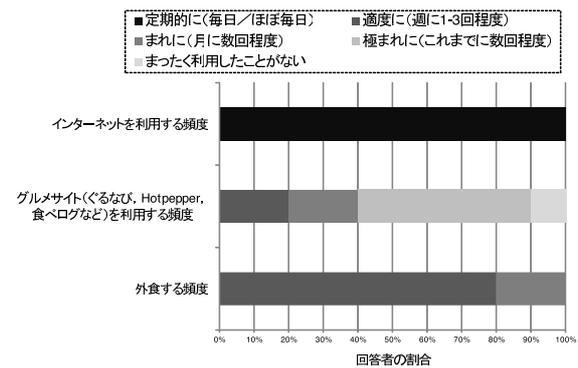


図 A.4 被験者の経験に関する調査  
Fig. A.4 Survey of subjects' experiences.

表 A.4 被験者に選択された対象地域

Table A.4 Local information selected by subjects.

タスク	現地名	現地における被験者の経験	地域限定語句リストの中で興味を引いた語句
A	三重県松阪市	旅行や出張などで行ったことがある	大和高原野菜研究，大柳生，山地鶏 漁師盛り，おかみ特製，名物溶岩焼き 国産讃岐牛，讃岐うどんならではの ひつうどん，ぷりぷりっのモツ，揚げたてさつま揚げ
A	奈良県生駒市	住んでいたことがある	
A	大阪府大阪市	旅行や出張などで行ったことがある	
A	福岡県福岡市	あまり知らない	
A	岡山県岡山市	住んでいたことがある	
A	香川県高松市	旅行や出張などで行ったことがある	
A	福岡県福岡市	旅行や出張などで行ったことがある	
A	広島県広島市	あまり知らない	
B	滋賀県近江八幡市	現在住んでいる	バームクーヘン豚
B	滋賀県草津市	現在住んでいる	近江鶏，三崎まぐろ
B	兵庫県西宮市	現在住んでいる	粉もんダイニング 鮭タルタル，揚げたてさつま揚げ ほうとう，鳥もつ，地ワイン，ヌーボー 泉州バーガー
B	滋賀県草津市	現在住んでいる	
B	福岡県福岡市	住んでいたことがある	
B	山梨県甲府市	住んでいたことがある	
B	大阪府泉佐野市	住んでいたことがある	

表 A-5 地域限定グルメ推薦システムに対する被験者からのフィードバック  
 Table A-5 Subjects' feedback for region-restricted restaurant recommender system.

利点/欠点	自由記述
利点	<ul style="list-style-type: none"> <li>・ぐるなびなどでは下位に表示されるものが上位に表示されることもあり、「隠れた名店」を探しやすい</li> <li>・タスク B において，地元で知らなかった泉州バーガーについて知れたのでよかった</li> <li>・地図で指定できるので，場所がわかりやすい</li> </ul>
欠点 (改善要望)	<ul style="list-style-type: none"> <li>・地域によって処理速度が遅いところがあった</li> <li>・フィルタリング指定を用いても「オープンです」「全額贈呈」などの直接店を選ぶ際の情報としては役に立たない語句がリストに表示されてしまった</li> <li>・主な地元グルメがない場合は，ありきたりのものが多く表示されてしまう</li> <li>・居酒屋が多い</li> <li>・赤マークで提示された店ももう少し情報が欲しかった (店のジャンルや予算など)</li> <li>・「和食」, 「洋食」, 「イタリアン」などの個別のジャンルに絞って検索できる機能があれば良い</li> <li>・価格の上限設定ができると良い</li> <li>・香川や福岡にはぐるなびに登録されていないが，地元ではとても有名でおいしい讃岐うどん屋やとんこつラーメン屋が数多くあると思うが，そのような飲食店も表示することができればさらに良い推薦システムになるのではないかと思った</li> <li>・長野県白馬村については，両システムともに抽出語句が 0 件であった</li> </ul>

がまんべんなく選択されている。地域限定語句リストの中で興味を引いた語句としては、「国産讃岐牛」や「近江鶏」, 「泉州バーガー」など料理に関する語句が特に選択されていることが分かる。

表 A-5 には，提案システムに対する被験者からのフィードバックを掲載している。得られたフィードバックを基に，システムのインタフェースとして，提示する飲食店情報の詳細化や，ジャンル絞り込み，価格などの条件指定などの機能を拡張していく。また，処理速度の改善やフィルタリング精度の向上，対象スポットデータの拡張を行うことで，利用者満足度の向上を目指す。



奥 健太 (正会員)

立命館大学情報理工学部助教。博士(工学)。2004年大阪市立大学工学部土木工学科卒業。2006年奈良先端科学技術大学院大学情報科学研究科博士前期課程修了。2009年同大学院同研究科博士後期課程修了。同年より現職。

主に情報推薦の研究に従事。電子情報通信学会，日本データベース学会，ACM，ヒューマンインタフェース学会，土木学会，地理情報システム学会，観光情報学会各会員。



西崎 剛司

1988年生。2011年立命館大学情報理工学部情報コミュニケーション学科卒業。同年同大学大学院理工学研究科博士前期課程入学，現在に至る。日本データベース学会学生会員。



服部 文夫 (正会員)

立命館大学情報理工学部教授。博士(情報学)。1973年早稲田大学理工学部電子情報通信学科卒業。1975年同大学大学院理工学部研究科修士課程修了。同年日本電信電話公社(現NTT)入社。2004年より現職。データベース，エキスパートシステム，エージェント通信等の研究に従事。電子情報通信学会，人工知能学会，AAAI，ACM等各会員。

(担当編集委員 鎌原 淳三)