

GPS を用いた旅行経験に基づく観光地推薦手法の提案

久原 健佑[†] 灘本 明代[‡]

[†] 甲南大学理工学部情報システム工学科 〒658-0072 兵庫県神戸市東灘区岡本 8-9-1

[‡] 甲南大学知能情報学部 〒658-0072 兵庫県神戸市東灘区岡本 8-9-1

E-mail: [†] ss664025@center.konan-u.ac.jp, [‡] nadamoto@konan-u.ac.jp

あらまし 本研究では旅行先で素早く的確な情報を得るために、旅行経験に基づく観光地推薦手法を提案する。近年、GPS が携帯電話やカーナビゲーションシステムなど多くの物に搭載されている。しかしながら、主に位置情報から地図上に現在地を表示したり、ルート案内したりすることにしか利用されていない。そこで、本研究では GPS で取得した出発地から現在地までの所要時間や直線距離を用いて、ユーザの旅行経験を抽出する。そして、この旅行経験に基づいて、推奨観光地を Web から抽出し、ユーザに提示するシステムの提案を行う。

キーワード 時空間データ, GPS, 情報推薦, Web

Travel Experience-based Sightseeing Recommendation System by Using GPS Data(O)

Kensuke KYUHARA[†] Akiyo NADAMOTO[‡]

[†] Faculty of Science and Engineering 8-9-1 Okamoto, Higashinada-ku, Kobe, Hyogo, 658-8501 Japan

[‡] Dept. of Intelligence and Informatics 8-9-1 Okamoto, Higashinada-ku, Kobe, Hyogo, 658-8501 Japan

E-mail: [†] ss664025@center.konan-u.ac.jp, [‡] nadamoto@konan-u.ac.jp

Abstract We propose sightseeing spot recommendation technique based on trip experience. Our proposed system represents information which is based on user's travel experiments and distance of destination. Nowadays, many mobile equipments, such as cell-phones and car navigation systems, have GPS. We, however, use such equipments only for displaying the present location on a map, and guiding a route. In this paper, we propose the system which extracts user's travel experiments by using distance between present location and destination from GPS data. Then, it extracts the sightseeing spot which is based on user's experiment from the Web, and presents it.

Keyword Data between the space-time, GPS, Information recommendation, Web

1. はじめに

現在、GPS は携帯電話やカーナビゲーションシステムなどあらゆるものに搭載されている。その GPS の主な利用方法は地図上に現在地を表示したりルート案内をしたりすることである。最近では子供に GPS を搭載した携帯電話を持たせて親が子供の現在地を把握できるといったサービスもある[1]。しかし、この GPS を利用したサービスも現在地の位置情報を取得して地図上に表示しているだけである。また、携帯電話やモバイルノートパソコンの普及によりいつでもどこでもインターネットを利用でき、情報を簡単に得ることができるようになってきた。しかしながら、旅行先でそれらの機器を利用してインターネットで調べ物をしたり観光地マップやガイドマップを使ったりすることでたくさんの情報を得ることはできる。しかし、情報量が多すぎるといった問題点がある。その結果、旅行慣れして

いない人、旅行慣れしている人にかかわらず、これから効率がよく有意義で楽しい旅行にするために旅行者がそれぞれの確かな情報を素早く簡単に得ることは困難である。

そこで本研究では、GPS で取得したデータを解析することで抽出できる時間や緯度経度による位置情報に基づいた過去の行動履歴を利用し、ユーザが旅行先で素早く的確な情報を簡単に得ることを可能とするシステムを提案する。具体的には、ユーザの行動履歴である過去の GPS データから、所要時間、出発点から現在地の直線距離、旅行経験の 3 つの要素を抽出する。そして、これらのデータからユーザの行動パターンを分析し、8 つのパターンに分類する。この行動パターンとユーザが希望する観光地の種類に基づいたランドマークを Web ページから参照して最適であると思われるランドマークをユーザに提供する手法を提案する。

このシステムによりユーザは調べ物をして目的のものがみつからず時間だけを無駄にしまったり、目的の情報を探すだけで疲れてしまったりする機会を減らすことができる。そして、効率良く観光地を巡り、楽しい旅行ができるようになる。

以下、第2章に関連研究、第3章に提案手法、第4章に有名度、第5章にプロトタイプシステムについて述べ、第6章に評価実験、第7章にまとめと今後の課題について述べる。

2. 関連研究

Web からランドマークを抽出する手法とその利用に関して、すでに多くの研究が行われている。手塚ら[2]は人間の地理空間認知をグラフによってモデル化するという考え方に基づいてランドマークを2つに分類し、それぞれ Web 上に存在する大量の文書情報からランドマークを評価して抽出している。そして、表示中の地図の中心地点から最も近いランドマークとその Web ページを表示している。また、旅行など実空間における行動を決定する際に、Web ページを参照して適切な情報をユーザに提供する研究もすでに行われている。北山ら[3]は画像に関する情報に着目している。ユーザが取得する情報において、画像に関するトピックを Web 特性と地理的特性により分析し、流言の流布量モデルに合致するかどうかで信憑性判断を行なっている。戸田ら[4]はすでに研究されている地理情報検索技術（GIR: Geographic Information Retrieval）のランキングに一般的なテキスト検索のランキング手法として注目されているクエリ中のキーワードにおける文書中の近接性を導入して評価を行い、取得する情報の質を上げている。GPS データの利用法についての研究もすでに行われている。柳井[5]は GPS による位置情報を写真に CGM データのタグとして付加したジオタグ付き写真に注目し、大量のジオタグ付き画像データの新しい利用方法の1つとして、特定の物体やシーンに関する世界規模の文化的差異の発見を行うという新しい研究課題を提案している。このように様々な地理情報に関する研究があるが本研究では、GPS データに着目し、そこから様々な情報を抽出する。それを利用してランドマークを抽出し、その中でも観光地に限定してユーザに提供する。また、提供する観光地の決定も GPS データから抽出した要素をもとに行っている。

3. 提案手法

3.1 システムの概要

我々が提案するシステムは3つの処理に大きく分けることができる。1つ目は GPS データから、出発点から現在地までの直線距離・所要時間、そして旅行経験

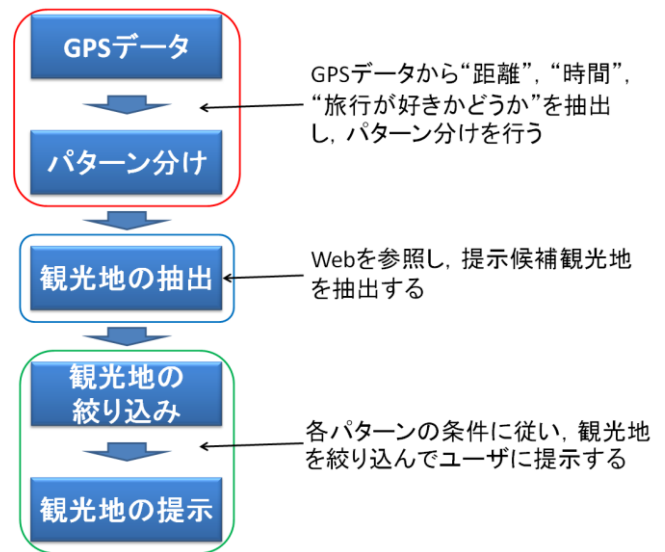


図1：簡単な処理の流れ

Fig1. A flow of simple processing

の指標として利用するユーザが旅行好きかどうかを判断するための同一フォルダ内にある GPS データの総直線距離の3つの要素を抽出する。2つ目は Web を参照し、現在地周辺のランドマークを抽出する。3つ目は GPS データから抽出した3つの要素をもとに分類したパターンの条件に従い、抽出したランドマークを5つに絞ってユーザに提示する。簡単なシステム中の処理の流れを以下と図1に示す。

具体的には以下の手順でユーザの旅行経験に基づく観光地情報を提示する。

- (1) システムはユーザの行動履歴である GPS データから現在地の緯度経度を抽出し、地図上に現在地を表示する。
- (2) ユーザはシステムが提示する7つの観光地の種類の中から1つを選択する。
- (3) システムは選択した観光地の種類と現在地の緯度経度から周辺のランドマークを抽出する。
- (4) GPS データから抽出した所要時間、出発地点から現在地までの直線距離、そして同一フォルダ内にある GPS データの総直線距離の3つの要素からユーザに提供するランドマークと地図の表示範囲を決める。
- (5) 最後に地図上に現在地を中心として適切なランドマークを5つユーザに提示する。それと同時に地図の右側にも地図上のランドマークと対応した観光地一覧を提示する。

3.2 選択させる観光地の種類について

ユーザに選択する観光地の種類は Wikipedia の日本

の観光地に関するカテゴリ[6]を参照し、その中から観光スポットとなるものを選ぶこととする。Wikipediaを利用した理由はそれぞれのカテゴリにおける具体的な観光地が示してあり、確認が容易だったためである。また、Wikipediaは多くのインターネットユーザによって構築されたものなので一般的な情報が掲載されていると考えたからである。本研究でユーザに選択させる観光地の種類は“動物園・植物園・牧場”，“博物館・美術館”，“寺・神社”，“公園”，“温泉”，“テーマパーク”，“その他の観光地”の7種類とする。

3.3 GPS から抽出するデータ

GPS データから抽出するのは所要時間、出発地から現在地までの直線距離、旅行経験の3つの要素である。

1 つ目の所要時間は出発地から現在地に到着する時間のことを示す。所要時間の抽出方法は、最初に位置情報を取得したデータと最後に位置情報を取得したデータから年月日と時間が記録されている部分を抽出して算出する。“年”，“月”，“日”，“時間”の順に比較し差分を取る。“年”は必要ない可能性が高いが、万が一のことを考えて処理に入れておく。所要時間の単位は分で求めている。

2 つ目の直線距離は出発地点から現在地までの移動距離ではなく、単純に点と点の直線距離のことを示す。直線距離の抽出方法は、最初に取得された緯度経度を出発地点、最後に取得された緯度経度を現在地と考えてそれぞれの緯度経度を抽出する。しかし、抽出した緯度経度は日本測地系なのでそれぞれ北緯に12秒加え、東経に12秒減じて世界測地系に変換する。そして、下記に示すヒュベニの距離計算式で算出する。

$$D = \sqrt{(M \cdot dP)^2 + (N \cdot \cos(P) \cdot dR)^2}$$

$$M = \frac{6334834}{\sqrt{(1 - 0.006674 \cdot \sin(P)^2)^3}}$$

$$N = \frac{6377397}{\sqrt{1 - 0.006674 \cdot \sin(P)^2}}$$

ここで、M は子午線曲率半径、N は卯酉線曲率半径、P は2点の平均緯度、dP は2点の緯度差、dR は2点の経度差を示す。所要時間は最初に位置情報を取得した時間と最後に取得した時間から算出する。

3 つ目の旅行経験を抽出するためにユーザの行動量に注目する。行動量を測るために現在所要時間、直線

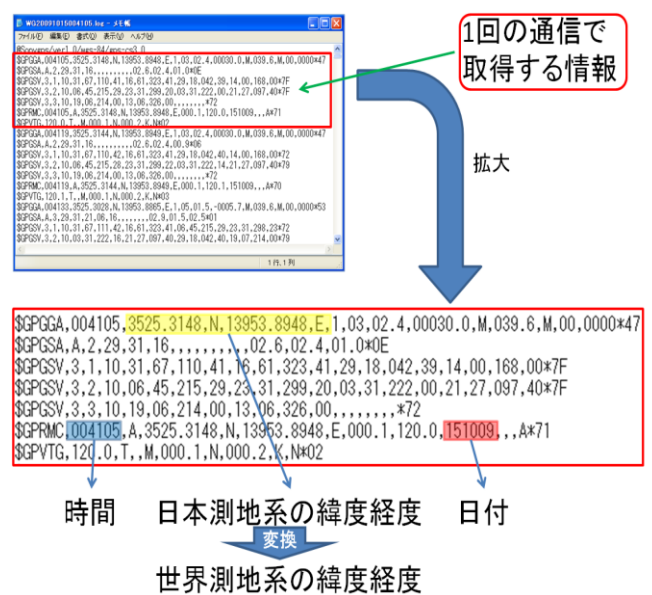


図2：GPS データ解析の流れ
Fig2. A flow of the GPS data analysis

距離の抽出の対象になっている GPS データと同一フォルダ内にある GPS データの総直線距離を求める。その具体的な抽出方法は同一フォルダ内にある各々の GPS データファイルに対して上記で示した直線距離を算出する方法で出発地点と現在地の直線距離を算出し、すべてを合算する。GPS データの解析は図2に示す部分に注目して行う。

3.4 提供する情報の種類

最初に、提供する情報のパターン分けについて述べる。本論文では出発点から現在地までの直線距離、所要時間、ユーザの旅行の好き具合（同一フォルダ内にある GPS データの総直線距離）の3つの要素から図3に示すように8つのパターンに分類する。

以下、距離の近い遠いはユーザの自宅からの距離の近さであり、時間のかかるかからないのはユーザの自宅からの所要時間とする。また、旅行好きかどうかはユーザの行動履歴の多さから判断するものとする。

パターン1：距離は近い・時間がかからない・旅行好きではない

基本情報 A を提供する。このパターンは現在地が住んでいる場所から近く、且つ時間がかからないので次の旅行先に選ぶ可能性が高いと考える。しかし、利用者は旅行好きではなく旅行そのものをあまりしないとえられる。そのためメジャーな観光地でも行ったことがない可能性が高いので、基本情報 A はメジャーな観光スポットを提示することとする。しかし、住んでいる地域から近いので旅行者に土地勘がある。ここで、全国的に有名なスポットは土地勘があれば場所や内容

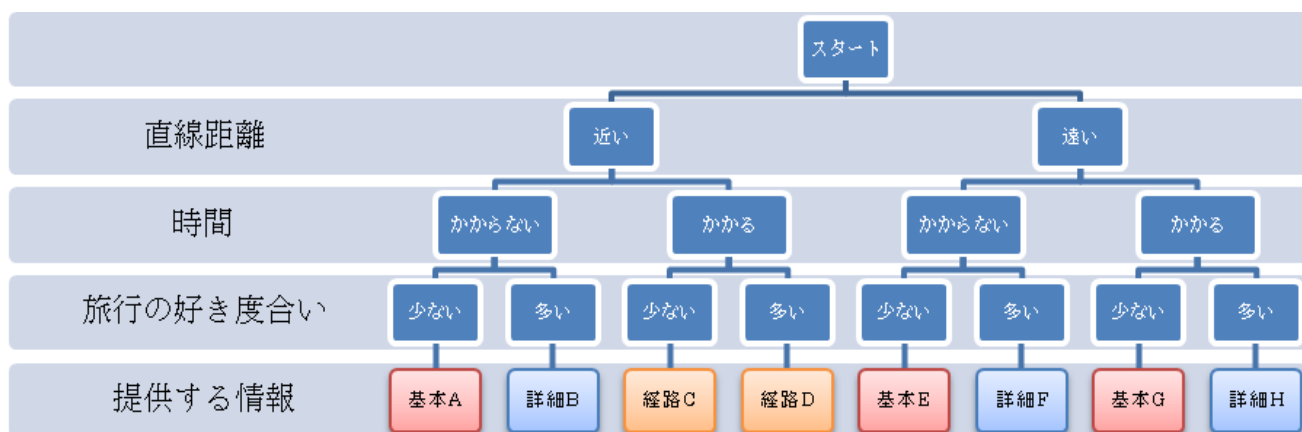


図 3：パターン分け

Fig3. Infotmation pattern

がわかる可能性が高いのでこのパターンに該当するユーザには余分な情報であると判断して提示しない。

パターン 2：近い・時間がかからない・旅行好き
詳細情報 B を提供する。このパターンは目的地が近くて時間がかからなく、且つ時間がかからないので次の旅行先を選ぶ可能性が高いと考える。さらに、旅行が好きのため現在地周辺を次の旅行先に再び選び、何度も訪れる可能性が高いと考えられる。これらの条件から、旅行慣れしていると考え、メジャーな観光地は下調べをしている可能性が高い。そのため、詳細情報 B はマイナーな観光地を提示する。また、何度も訪れたときのことを考え、広範囲で表示して情報量を少し多めにする。

パターン 3：近い・距離のわりには時間がかかる・旅行好きではない

経路途中情報 C を提供する。このパターンは距離が近いので車や電車などの経路の途中で融通が利きやすい交通機関を利用すると考える。また、ユーザは旅行好きでないことから、あまり観光地を知らない可能性が高いと考える。そのため、経路途中情報 C は旅行の行きと同じ道を帰りも通ることを想定し、帰りの道中を有意義にするため経路途中周辺のメジャーな観光地を提示する。

パターン 4：近い・距離のわりには時間がかかる・旅行好き

経路途中情報 D を提供する。このパターンもパターン 3 と同様に距離が近いので車や電車などの経路の途中で融通が利きやすい交通機関を利用すると考える。しかし、パターン 3 と違って履歴が多くユーザは旅行好きであると考えられる。そのことから旅行慣れしている可能性が高くある程度の観光地は知っているか下調べを行っていると考え。そのため、旅行の行きと同

じ道を帰りも通ることを想定し、帰りの道中を有意義にするため経路途中情報 D はパターン 3 とは違い、経路途中周辺のマイナーな観光地を提示する。

パターン 5：遠い・距離のわりには時間がかからない・旅行好きではない

基本情報 E を提供する。このパターンは遠いわりには時間がかからないので何度も訪れやすい場所ではある可能性が高い。しかし、旅行好きではないことから再び訪れる可能性は低いと考えられる。これらの条件からユーザは旅行慣れしておらず、的確な下調べをしていない可能性が高い。そのため、詳細情報 E はメジャーな観光地を提示する。このパターンではパターン 1 と違いユーザに土地勘がないことから全国的に知られている有名な観光地でも場所がわからない可能性が高いので提示する。

パターン 6：遠い・距離のわりには時間がかからない・旅行好き

詳細情報 F を提供する。このパターンは遠いわりには時間がかからないので何度も訪れやすい場所ではある可能性が高い。さらに、ユーザは旅行好きであることから旅行慣れしており、メジャーな観光地は下調べをすると考え。そのため、詳細情報 F はマイナーな観光地を提示する。

パターン 7：遠い・時間がかかる・旅行好きではない

基本情報 G を提供する。このパターンは遠くて、且つ時間がかかることから今後の旅行先として選ぶ可能性が低くなる。また、旅行好きではないことからさらに再び訪れる可能性が低くなる。これらの条件からユーザは旅行慣れしておらず、的確な下調べをしていない可能性が高い。そのため、基本情報 G はメジャーな観光地を提示する。また、再び訪れる可能性がとても低

表 1：観光地提示情報
Table1. Sightseeing spot presentation information

	基本情報 A	詳細情報 B	基本情報 E	詳細情報 F	基本情報 G	詳細情報 H
提示候補の 観光地抽出数	約 50 件	約 50 件	約 25 件	約 25 件	約 25 件	約 25 件
提示観光地の 決定方法	有名度値 10000 以下 のうち上位 5 件	下位 5 件	上位 5 件	下位 5 件	上位 5 件	有名度値 10 以上 のうち 下位 5 件
地図の 表示範囲	標準	広い	広い	標準	狭い	狭い

いことから 1 度の旅行で消化できる情報量に抑えるために表示する範囲を狭くする。

パターン 8：遠い・時間がかかる・旅行好き

詳細情報 H を提供する。ここのパターンはパターン 7 と同様に遠くて、且つ時間がかかることから今後の旅行先として選ぶ可能性が低くなる。しかし、ユーザは旅行好きで旅行慣れしているのである程度下調べをすることが考えられる。そのため、詳細情報 H はマイナーな観光地を提示する。また、再び訪れる可能性がとても低いことから 1 度の旅行で消化できる情報量に抑えるために表示する範囲を狭くし、あまりにも細かすぎる情報は提示しない。

ここで、直線距離は閾値 α (メートル)、所要時間は閾値 β (分)、旅行好きかどうかを判断するための同一フォルダ内にある GPS データの総直線距離の大小判断は閾値 γ (メートル) を基準とする。本稿では $\alpha=140000\text{m}$ 、 $\beta_1=90$ 分、 $\beta_2=200$ 分、 $\gamma=5000000\text{m}$ 、とする。本稿ではパターン 3 とパターン 4 については実現せず、今後の課題とする。

上記の 8 つのパターンにおいて本稿では、距離の近い遠いはある一定の距離を基準に考える。例えば、出発地を兵庫県神戸市とすると、パターン 1 とパターン 2 場合、現在地は直線距離が約 61000 メートル (6.1 キロ) で車を利用すると所要時間が 60 分かかる京都になる。その場合、例えばユーザが観光地の種類に寺・神社を選択したとするとパターン 1 では東本願寺を提示し、パターン 2 では盛光院を提示する。パターン 3 とパターン 4 の場合、目的地は直線距離が 130000 メートル (130 キロ) で車を利用すると所要時間は約 210 分 (3 時間 30 分) かかる鳥取県鳥取市 (目的地は鳥取砂丘) になる。その場合、例えばユーザが観光地の種類に公園を選択したとするとパターン 3 では道中の西は

りま天文台公園を提示し、パターン 4 では道中のパレンティンパーク作東を提示する。パターン 5 とパターン 6 の場合、現在地は直線距離が約 430000 メートル (430 キロ) で新幹線を利用すると所要時間は約 180 分 (3 時間) かかる東京になる。例えばユーザが観光地の種類に博物館・美術館を選択したとするとパターン 5 では東京国立近代美術館を提示し、パターン 6 ではわだつみのこえ記念館を提示する。パターン 7 とパターン 8 の場合、現在地は直線距離が約 1190000 メートル (1190 キロ) で飛行機を利用すると所要時間は約 210 分 (3 時間 30 分) かかる沖縄県那覇市になる。この場合、ユーザが観光地の種類に寺・神社を選択したとするとパターン 7 では護国神社を提示し、パターン 8 では神徳寺を提示する。

3.5 提示する情報について

Yahoo!JAPAN ローカルサーチ API[7]を使って「観光地の種類 緯度 経度」をクエリとして検索を行い、提示候補のランドマークを約 25 件または約 50 件抽出する。そのランドマークをすべて Yahoo!JAPAN ブログ検索[8]で検索し、そのブログ検索ヒット数を抽出する。そして、すべてのランドマークの有名度を測る。最後に有名度地を比較し、ユーザに提示するランドマーク 5 件をパターン分けの条件によって決定する。ランドマーク 5 件は 8 つのパターンの条件によって変化する。Yahoo!JAPAN ブログ検索を利用する理由は多くの人の感覚を有名度判断に取り入れたいと考えたからである。表 1 にパターンごとに違う提示候補のランドマーク抽出数、そのランドマークを 5 つに絞る条件、システムの起動時を規準とした地図の表示範囲を示す。

4. 有名度

本研究では多くの人に感覚を取り入れるために Yahoo!ブログ検索を利用している。しかし、ブログは

表 2：有名度における要素

Table2. An element in the famous degree

都道府県人口(人)	δ	該当都道府県数
1000 万以上	0.6	1
700 万以上 1000 万未満	0.7	4
400 万以上 700 万未満	0.8	4
100 万以上 400 万未満	0.9	31
100 万以下	1.0	7

実際にその場に訪れた人によって書かれることが多いので人口の多い地域の観光地は訪問者数が必然的に多くなる．その結果，ブログのヒット数が増えてしまうと考えた．

そこで各都道府県の人口に対応して Yahoo! ブログ検索の検索ヒット数に重みづけを行い，有名度値を求める．有名度値 F は下記の式で算出する．

$$F = \text{Yahoo! ブログ検索ヒット数} \times \delta$$

本稿では重みづけに使う δ の値をどのような基準で決定したかを表 2 に示す．

5. プロトタイプシステム

プログラミング言語は Ruby を使い，ユーザインターフェースの構築には Ruby on Rails を使った．現在地が兵庫県神戸市にある三宮駅のときのシステム起動時とユーザが観光地の種類に“寺・神社”を選択したときの推奨観光地提示時をそれぞれ図 4 と図 5 に示す．

6. 評価実験と考察

6.1 評価実験

被験者 5 人に対して前提条件を設定した 2 つのパターンを 2 回ずつ行った．また，観光地の種類については毎回被験者各自に選択させた．

パターン 1

前提：ユーザは“現在地と出発地が近く”，“時間がかからない”，“旅行が好きではない”とする．

パターン 2

前提：ユーザは“現在地と出発地が遠く”，“距離のわりには時間がかからない”，“旅行が好き”とする．

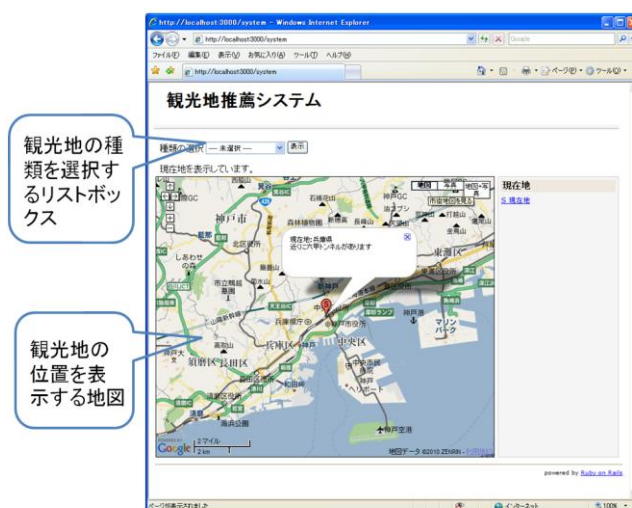


図 4：システム起動時

Fig4. At the time of system start

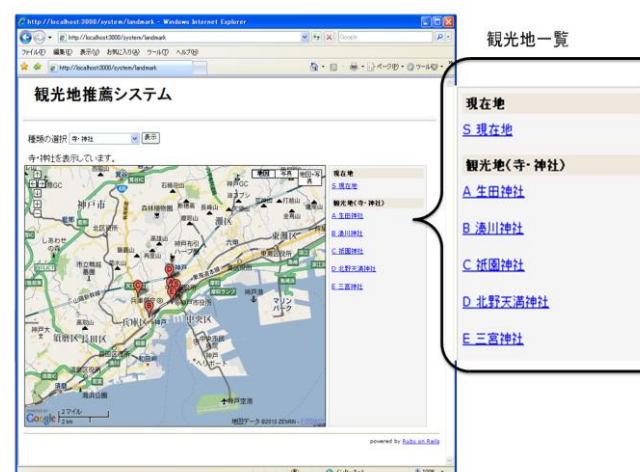


図 5：観光地表示時

Fig5. At the time of sightseeing spot indication

評価項目は観光参考度と適切観光地数の 2 つで行った．実験結果は表 3，表 4 に示す．

また，システム全体に対する評価も行った．評価項目はユーザに選択してもらう観光地種類の選択肢について，ユーザインターフェースのレイアウトについて，システムを使ってみて全体的な満足度についての 3 つで行った．実験結果を表 5 に示す．

6.2 考察

評価実験ではメジャーな観光地を提示したときよりマイナーな観光地を提示したときの方が良い結果が得られた．それは全国的に有名とされる観光地がどこでもたくさんあるわけではなく，ほとんどの地域でメジャーな観光地が提示範囲内にあまり多く存在しないことが原因だと考えられる．また，ユーザインターフェ

表 3：観光地参考度

Table3. Sightseeing spot reference degree

	とても参考になる	まあまあ参考になる	あまり参考にならない	全く参考にならない
パターン1	2	7	1	0
パターン2	5	4	0	1

表 4：適切観光地数

Table4. The number of the appropriate sightseeing spots

	5つ	4つ	3つ	2つ	1つ以下
パターン1	2	0	6	0	2
パターン2	5	0	1	1	3

表 5：システム全体の評価

Table5. Evaluation of the whole system

	とても満足した	まあまあ満足した	どちらでもない	あまり満足しなかった	全く満足しなかった
選択肢について	1	4	0	0	0
レイアウトについて	0	3	2	0	0
満足度	0	4	1	0	0

ースに関しては比較的良い結果が得られた。

7. まとめと今後の課題

本研究では GPS データに着目し、旅行経験を測るために GPS データから所要時間、出発地から現在地までの直線距離、好きかどうかを判断するための同一フォルダ内の総直線距離の 3 つの要素を抽出した。そしてパターン分けを行った。次に、Web を参照して提示候補観光地を抽出した。その提示候補観光地の有名度を Yahoo!JAPAN ブログ検索と都道府県人口を利用して測って比較した。最後にパターンの条件に合った推奨観光地をユーザに提示するシステムの提案について述べた。

今後の課題として、観光地の表記ゆれによる同一観光地の複数表示の解消や本稿では実現できなかった経路途中の情報を提供するパターン 3 とパターン 4 の実現が求められる。また、GPS を搭載した携帯電話に対応させたり、提示した観光地の付加情報を導入したりすることでより便利で役に立つシステムになる。

謝 辞

本研究の一部は、平成 21 年度科研費特定領域研究域「コミュニティ型コンテンツのコンテンツホール検索に関する研究」(課題番号：21013044, 代表：灘本明代)及び甲南大学平生太郎基金科学研究奨励助成金による。ここに記して謝意を表します。

文 献

- [1] “株式会社 NTT docomo のイマドコサーチ”
<http://www.nttdocomo.co.jp/service/safety/imadoco/index.html>
- [2] 手塚太郎, 田中克己, “ウェブからのランドマーク抽出に基づくクエリフリーな地域情報閲覧” 日本データベース学会 Letters Vol.4
- [3] 北山大輔, 李龍, 角谷和俊, “地理的特性と Web 特性に基づくトピックの曖昧度分析” 社団法人情報処理学会
- [4] 戸田浩之, 安田宜仁, 松浦由美子, 片岡良治, “地理情報と内容情報の近接性を考慮した地理情報検索”
- [5] 柳井啓司, “Web 上のジオタグ付き画像を用いた世界各地の文化的差異の発見”
- [6] “Wikipedia の日本の観光地に関するカテゴリ”,
<http://ja.wikipedia.org/wiki/Category:日本の観光地>
- [7] “Yahoo!JAPAN ローカルサーチ API”,
<http://developer.yahoo.co.jp/webapi/map/localsearch/v1/localsearch.html>
- [8] “Yahoo!JAPAN ブログ検索”,
<http://blog-search.yahoo.co.jp/>