

POI情報の補完によるマイマップ検索システム

荒川 豊^{1,2,a)} 金平 卓也¹ 安本 慶一¹

受付日 2018年5月7日, 採録日 2018年11月7日

概要: Google Maps や Open Street Map (OSM) では, 1つの地図に複数の地理情報 (POI: Point of Interest) を集めて独自の地図を作成可能にする機能が提供されている. この機能は, マイマップ (Google Maps) あるいは uMap (OSM) と呼ばれており, ニッチな情報を集めた地図を含めて, これまでに多種多様な地図が作成されている. しかしながら, これらの地図は, 含まれる情報が不十分であることが多く, キーワードを用いて, 目的に合致するマイマップを検索することが困難である. 本論文では, マイマップ検索の利便性を向上させるため, これまでに存在しなかったマイマップの検索法として, 含まれる POI のカテゴリ比率を特徴とした検索システムを提案する. また, 提案システムを実現するにあたり, 既存マイマップの問題である情報不足を補うため, 1) 不完全な POI 情報から POI 情報データベース内の POI を同定する手法と 2) POI 名称から属するカテゴリを推定する手法, から構成される POI 情報補完手法を提案する. インターネット上に公開されているマイマップ 5,526 件を収集し, その中に含まれる POI 46,996 件について, POI 情報の同定精度やカテゴリの推定精度を評価した結果, 58.82%を同定可能かつ 70.3%のカテゴリを正しく推定できることを明らかにした. さらに, プロトタイプシステムを実装し, 10名の被験者を用いて検索の利便性について評価したところ, 含まれる POI カテゴリ比率に基づく検索により, 所望するマイマップに近いマップを一定の精度で検索可能であることを確認した.

キーワード: マイマップ, 地理情報検索, POI (Point of Interest), 地理情報システム

A System for Searching MyMaps by Complementing Additional Information of POIs

YUTAKA ARAKAWA^{1,2,a)} TAKUYA KANEHIRA¹ KEIICHI YASUMOTO¹

Received: May 7, 2018, Accepted: November 7, 2018

Abstract: In Google Maps and Open Street Map (OSM), a function which allows users to create own map including necessary Point of Interests (POIs) is provided. This function is called My Maps in Google Maps and uMap in OSM. A variety of maps have been created so far including maps that gather niche information. However, it is difficult to search My Maps that meets the purpose using keywords because these maps often do not contain sufficient information. In this paper, to improve the accuracy and intuitiveness in searching My Maps, we propose a novel system for searching MyMaps where a category ratio is used as a feature of MyMaps. To realize the system, we propose methods to complement the insufficient information composed of 1) identifying a probable POI in POI database with incomplete POI information and 2) estimating a category name of POI from POI name. We collected 46,996 POIs included in 5,526 public MyMaps for evaluation. As a result, we achieved 58.82% accuracy in POI identification and 70.3% accuracy in POI category estimation. We also implemented the prototype system for evaluating the convenience of the search by ten subjects. The result shows our proposed system can retrieve MyMaps including desired POI sets by using a category ratio as a feature.

Keywords: My maps, Geographical Information Retrieval, POI (Point of Interest), GIS

¹ 奈良先端科学技術大学院大学
Nara Institute of Science and Technology, Ikoma, Nara 630-0192, Japan

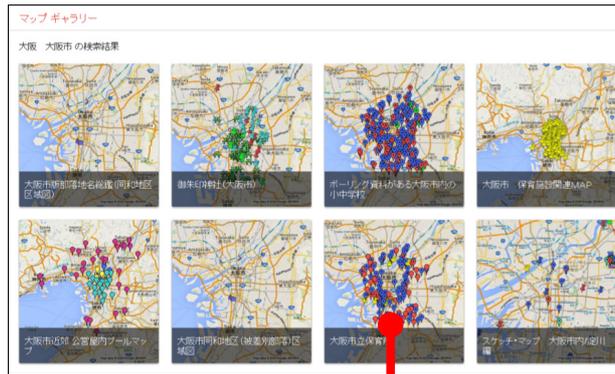
² JST さきがけ
Japan Science and Technology Agency, PRESTO, Chiyoda, Tokyo 102-0076, Japan

^{a)} ara@is.naist.jp

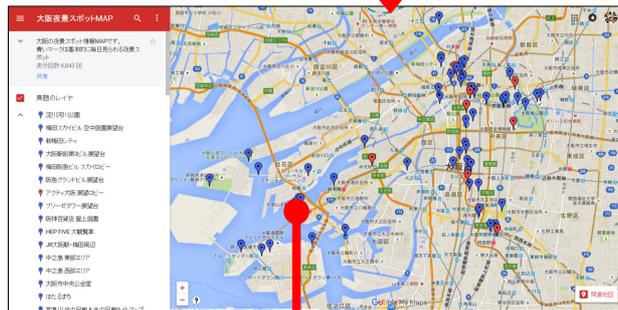
1. はじめに

Google Maps や Open Street Map (OSM) などの個人向け地図サービスは, 2005年に始まり, 今では欠かせないサービスとなっている. これらのサービス開始当初の頃は

(1) 検索結果



(2) マイマップ



(3) POI情報

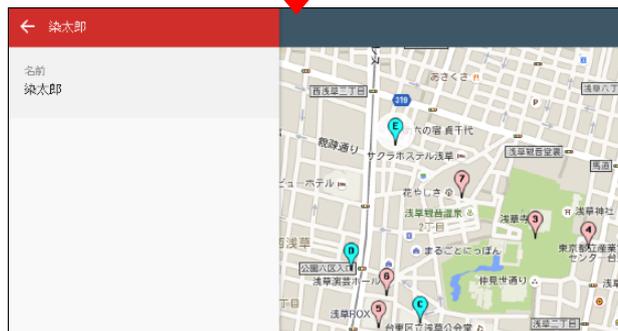


図 1 Maps Gallery 上のマイマップと POI の例

Fig. 1 Example of POI data and MyMaps in Maps Gallery.

地図を閲覧できるだけであったが、スマートフォンの普及にともない、ルート検索や場所検索などの機能も利用することが可能である。2007年には、Google Maps上で、ユーザ自身が複数の地理情報 (POI: Point of Interest) を1つの地図にまとめ、共有することを可能にする機能、すなわち、マイマップ機能の提供が始まった。OSMにおいても、uMAP^{*1}と呼ばれる同様の機能が提供されている。

図 1 (2) に示すようにマイマップは、ユーザの持つ地理情報を視覚的かつ直感的に1つの地図にまとめることができる。そして、この地図の URL を送ったり、Webサイトに埋め込むことにより、第三者に共有することが可能となっている。こうしたマイマップの利用形態としては、Webサイトに埋め込むか、地図のアドレス (URL: Uniform Resource Locator) をメールなどで共有するという形にな

る。たとえば、国際会議のホームページ内で、周辺のレストラン情報をビジュアルに掲載する手段として埋め込まれる。このレストランマップは、同じ会場で会議を行う別の会議にとっても有用な情報である。しかしながら、Webサイトに埋め込まれているため、Googleなどの一般的な検索手段では、マイマップを直接探すことはできない、という問題がある。

そこで、我々はいち早くこのマイマップに含まれる情報の有用性に着目し、2013年から、マイマップの作成や共有、横断検索を可能とするプラットフォームを提案してきた [1], [2], [3]。2014年には、マイマップ機能を提供するGoogleからも、一般公開されているマイマップをフリーキーワードで検索することができる Maps Gallery^{*2}というサービスがリリースされ、公開されているマイマップをキーワードで検索することが可能となった。

しかしながら、Maps Galleryを通じて、マイマップの検索に関する2つの問題が浮き彫りとなった。まず、個人によって作成されるマイマップは、含まれるPOIの情報が不十分かつ不正確であることである。もう1つの問題は、作成者の意図によって、地図名が似ていても、含まれる情報は異なるということである。観光スポットだけを集めたものもあれば、ホテルやレストランなどを含んだものもある。100件以上のPOIが含まれているものもあれば、数件しか含まれていないものもある。そのため、Maps GalleryやuMapで提供されているキーワード検索では、所望するマイマップの検索が困難であることが判明した。

例として、『大阪 大阪市』というキーワードに対する検索結果を図 1 (1) に示す。8つのマイマップが提示されているが、一見して、どのマイマップが所望する情報を含んでいるのか分からない。具体的にどのようなPOIが含まれているかを確認するには、1つずつ地図を開いて確認する必要がある。

しかも、各POIの情報は、図 1 (3) に示すように、POI名だけが書かれたような不完全なものも多く含まれている。これでは、レストランの名称なのか、観光スポットの名称なのかすら分からない。すでに、Googleは、Maps Galleryから、マイマップ検索機能を削除し、マイマップの作成と作成したマイマップの管理機能だけを提供するようになっているが、この問題が一因であると考えられる。

本論文は、この問題を解決し、既存のマイマップの中から所望のマイマップを検索可能にするため、マイマップに含まれるPOI群のカテゴリ比率を特徴として検索するシステムを提案する。キーワードに加えて、どのようなカテゴリを含んだマップを所望しているのかを入力することで、意図に沿ったマイマップを上位に提示できるようになると考えている。また、そのシステムを実現するため、マイマッ

^{*1} uMap: <https://umap.openstreetmap.fr/ja/>

^{*2} Google Maps Gallery <https://maps.google.com/gallery/>

プに含まれる POI 情報を補完する手法もあわせて提案する。POI 情報の補完は、1) 不完全な POI 情報から確からしい POI を推定する手法と 2) POI 名称からそのカテゴリ情報を推定する手法から構成される。前者は、Foursquare の POI 情報データベース^{*3}を用い、文字列の類似度と物理的距離の近さを加味した POI 類似距離に基いて推定する。後者は、Foursquare の POI 情報データベースにカテゴリが登録されている場合は、そのカテゴリ名を割り当て、登録されていない場合は、機械学習を用いて作成した回帰モデルから、POI が属するカテゴリを推定する。マイマップを検索クエリとするマイマップ検索システムでは、POI 情報の補完によって付与された各 POI のカテゴリ情報を、そのマイマップの特徴量と見なし、キーワードで指定された地域の中から、コサイン類似度によって算出される類似度の高いマイマップが検索結果として提示される。

インターネット上に公開されているマイマップを収集して評価した結果、POI 情報の補完については、収集したマイマップに含まれる POI 群 46,996 件中 27,645 件を同定する^{*4}ことが可能であり、同定精度は 58.82%であることを確認した。また、カテゴリ推定手法では、4 通りの特徴量を用いてサポートベクタ回帰により推定した結果、25 種類のカテゴリを平均 70.3%で推定できることを確認した。この結果は、既存研究 [11] より、カテゴリが 5 種類多く、推定精度は 14.4%高い。最後に、被験者 10 名により、プロトタイプとして構築したマイマップ検索システムを用いて、入力したマイマップと検索結果として提示されたマイマップが類似していると判断するコサイン類似度のしきい値を明らかにするとともに、システムの有用性について評価した。その結果、マイマップどうしが似ていると判断するコサイン類似度のしきい値は 75%であることと、マイマップを入力として類似した POI を含むマイマップを一定の精度で検索可能であることを示した。

2. 関連研究

2.1 マイマップ検索システム

マイマップは、Web サイトに埋め込んで利用する、あるいは、地図の URL をメールなどで共有するという使い方が想定されている。Google などの Web 検索によって、探している情報を含む Web サイトにたどり着き、そのなかで埋め込まれたマイマップを発見することはあるが、埋め込まれたマイマップを検索によって直接探すことはできない。そのため、マイマップを検索できるシステムは、マイマップ作成サービスを提供している Google および OSM

が提供する、Maps Gallery と uMap という 2 つのサービスだけに留まっている。

前者は、1 章で述べたように、多くの地図が登録されていたものの、POI 情報の不完全さから、提示される地図が意図に沿ったものでないことが多かった。そして、2016 年半ばに、検索機能が省かれている。

後者は、現在も検索サービスを提供している。しかし、2018 年 5 月現在、『東京』に対して提示されるマイマップは 1 件のみである。『tokyo』と入力した場合は、数十件提示されるが、『tokyo sightseeing』では 0 件となる。uMap の知名度が低いのか、提示される地図は、作りかけのものやまったく POI が登録されていないものも多い。

2.2 POI 推薦システム

複数の POI 情報を含んだマイマップを検索できるシステムは上述のとおりほとんど存在しないが、任意の地域における POI の推薦システムに関してはいくつかの研究が行われている。加藤ら [7] は飲食店の周辺相場関係の類似性を基に、任意の地域でユーザが選択した飲食店に類似した飲食店を推薦するシステムを提案している。また、Bao ら [8] はユーザのカテゴリ付き POI のチェックイン履歴からユーザ嗜好をモデル化し、他の地域のユーザの嗜好の類似性を基にその地域の POI を推薦するシステムを提案している。さらに、Yuan ら [9] は、ユーザの時間的嗜好変化を考慮した POI 推薦システムの提案している。

2.3 POI 名称の推定・補完手法

本論文で取り扱う静的な地図情報には適用できないが、訪問した POI を推定する手法として、Nishida ら [4] は、確率的訪問 POI 同定法を提案している。この研究は、ユーザの移動軌跡から得られる滞在地点に対して自動的にチェックインすることを目的とし、ユーザの SNS のチェックイン履歴から好みのカテゴリと、そのカテゴリへの訪問回数の情報を取得し、訪問したであろう POI を推定している。また、落合ら [5] は、ツイート 140 文字を分析し、静的な特徴語に加え、時間変化する特徴語を加味して、その発言地域を推定する手法を提案している。本論文で対象とするデータは、マイマップに登録された店名など数文字の情報しかなく、複数の特徴語を用いるような手法の適用は難しい。そのため、POI 情報データベースとマッチングするという単純な手法をとっている。

筆者も複数の POI 情報データベースを一元的にまとめることで情報量を増やす手法を提案している [12]。この研究では、Twitter 社、Google 社、Foursquare 社など問い合わせ形式が異なる複数の POI 情報データベースを単一のクエリで一括検索できるシステムを提供し、得られた結果の中から、文字列の類似度に基づいて最も確からしい POI 名称を選出する手法を提案している。本論文では POI 情

^{*3} Foursquare Developer: <https://developer.foursquare.com/places-api>

^{*4} 同定とは、POI 情報データベース内を検索し、マイマップに登録された情報に対応する POI を選出することであり、収集した 46,996 件について人手で確認している。詳細は、5.1 節で述べる。

報データベースとしてカテゴリ情報を含む Foursquare 社を利用したが、登録されていない POI が多い場合は、複数の POI 情報データベースを統合して利用することも考えられる。

2.4 POI のカテゴリ推定手法

POI 名称からカテゴリを推定する手法については、Choi ら [11] の POI カテゴリ推定方法がある。この研究では、Yelp から得た 3,499 件の POI データをトレーニングデータとしてサポートベクターマシン (SVM) で機械学習し、別の 941 件の POI データをテストデータとして POI カテゴリの推定精度を評価している。POI データの内、POI 名称とカテゴリの関係を機械学習させ、20 種類の POI カテゴリを推定した結果、45.84%の精度とされている。機械学習における特徴抽出方法としては、POI の名称の 1 文字ずつを特徴量とする unigram が用いられている。

3. POI 情報の補完によるマイマップ検索システム

本論文では、キーワードに加えて、マイマップに含まれるカテゴリ比率を特徴として用いたマイマップ検索システムを提案する。提案システムについては、4 章において詳しく説明し、本章では、提案システムを実現するために提案する 2 つの要素技術について述べる。

3.1 POI 情報の補完手法

マイマップ内に含まれる不完全な POI 情報に対して、正式な名称やカテゴリなどの情報を付与することを、POI 情報の補完と呼ぶ。提案手法は、1) POI 情報データベースを用い、マイマップ内の POI 情報 (名称と位置) から、データベースに登録されている POI を同定する手法と、2) POI 名称からカテゴリを推定する手法から構成される。図 2

に、両者の関係を含めて、POI 情報の補完手法の流れを示す。

3.1.1 POI 情報データベースを用いた補完手法

提案手法では、マイマップに含まれる POI ごとに、その POI の位置を中心として半径 100 [m] 以内の POI 情報を POI 情報データベースから取得する。今回、POI 情報データベースとしては、チェックインサービスを展開する Foursquare 社のものを Application Programming Interface (API) を通じて利用する。この検索半径の値は、Foursquare の検索誤差の中央値が約 70 [m] である [10] ことを考慮して決定した。次に、得られた POI 群に含まれる各 POI に対して、同定対象 POI との POI 類似距離を算出し、最も距離が近いものを選出し、その情報 (名称、座標、カテゴリなど) を同定対象 POI の情報として補完する。

POI 類似距離

POI 情報データベースから類似した POI を抽出する際、文字列の類似性だけを考慮すると異なる場所にある似た POI が選ばれることがあり、逆に、物理的距離だけを考慮すると近隣の別の POI が選ばれる可能性がある。そのため、POI 類似距離という新しい指標を提案する。POI 類似距離とは、我々が独自に提案する指標であり、名称の類似度と物理的な距離の近さの両方を加味して、2 つの POI の同一性を評価する指標である。

ある POI P_A と類似性を測りたい POI P_B の 2 つの POI があるとき、POI 類似距離 $S(P_A, P_B)$ を以下のように定義する。

$$S(P_A, P_B) = L(t_A, t_B) + \frac{D(p_A, p_B)}{100} \quad (1)$$

ここで、 t_x は POI P_x の名称であり、 p_x は経緯度である。また、 $L(t_A, t_B)$ は 2 つの POI の名称のレーベンシュタイン距離であり、 $D(p_A, p_B)$ は 2 つの POI の物理的距離である。 $D(p_A, p_B)$ には 2 地点の経緯度を与えるとその 2 地点間の距離を求めることができるユークリッドの公式 [6] を用いている。分母の 100 は、POI 情報の探索範囲である半径 100 [m] を表しており、探索範囲のサイズによって変化するパラメータとなる。なお、 $0 \leq L(t_A, t_B) \leq 1$ であり、 $0 \leq D(p_A, p_B) \leq 100$ である。すなわち、 $0 \leq S(t_A, t_B, p_A, p_B) \leq 2$ である。ここで、レーベンシュタイン距離について説明する。レーベンシュタイン距離は 2 つの文字列の類似性を測る 0~1 の指標であり、類似しているほど低い数値となる。つまり、POI 類似距離は名称の類似性と物理的距離の近さを 1 つの指標で表現したものであり、数値が低いほど 2 つの POI が同一である可能性が高いと考えられる。なお、式 (1) の名称のレーベンシュタイン距離 $L(t_1, t_2)$ が 0.8 以上のときは類似性がないと判断し、POI 同定を行わない。つまり、POI 名称を補完できないとする。

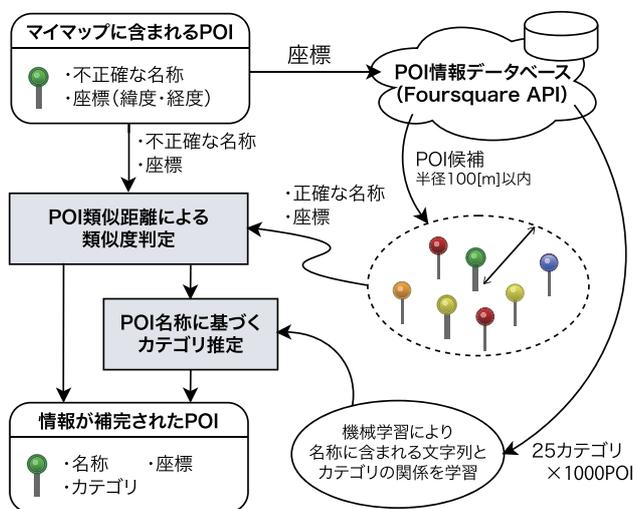


図 2 POI 情報の補完手法の流れ

Fig. 2 Procedure of complementing POI information.

3.1.2 POI 名称に基づくカテゴリ情報の補完手法

上記手法により、データベース内から類似 POI が得られた場合は、そのカテゴリを情報として補完することができる。しかしながら、マイマップは、様々なユーザがそれぞれの目的に応じて作成しているため、データベースに POI が含まれていないことも多い。そこで、POI の名称を手がかりに、カテゴリを推定する手法を提案する。この手法は、人であれば、『四川飯店』という文字列であれば『中華料理店』あるいは『レストラン』というカテゴリを推定可能であるということに着想を得ている。

提案手法では、カテゴリ情報を有する POI データを大量に収集し、機械学習によって、POI 名称に含まれる文字とカテゴリの関係を学習する。POI データの収集には、上記と同様に Foursquare 社の POI 情報データベースを用い、機械学習フレームワークには、Jubatus [13] を用いる。このフレームワークにはサポートベクターマシン (SVM) に比べて多クラス分類に向いているサポートベクター回帰 (SVR) のオンライン版である Passive Aggressive アルゴリズム [14], [15], [16] が採用されており、リアルタイムに機械学習することが可能である。

Foursquare の POI 情報データベースから収集した POI データは、25 種類のカテゴリについて各 1,000 件ずつ、計 250,000 件である。機械学習の事前処理として、得られた POI の名称から下記に示す 4 パターンの特徴量を抽出する。

- **origin** : POI の名称をそのまま特徴量として用いる。
- **unigram** : POI の名称の 1 文字ずつの特徴量として用いる。既存研究 [11] でも用いられている。
- **mecab** : POI の名称を形態素解析し、この結果得られる形態素を特徴量として用いる。形態素解析には MeCab^{*5}を用いる。
- **hybrid** : unigram の特徴量と mecab の特徴量の積を特徴量として用いる。たとえば POI の名称の文字数が 3 文字で、形態素解析の結果、形態素数が 2 の場合、 $3 \times 2 = 6$ つの特徴量があることになる。

なお、POI の特徴数は文字数によって変化するが、本研究では機械学習フレームワーク Jubatus に含まれる特徴ベクトルの次元数の違いを吸収する機能を用いて一元的に扱っている。

3.2 カテゴリベクトルによるマイマップの特徴付けと比較

本論文では、マイマップの検索クエリとして、キーワードに加えて、他のマイマップを利用可能にすることを提案する。これは、ユーザがある地域の観光スポットと周辺の飲食店を網羅した理想的なマイマップを保有している場合、そのマイマップを検索クエリとして別の地域を検索することができれば、ユーザが理想的と考える POI 情報を

表 1 カテゴリおよび各カテゴリのサブカテゴリの数

Table 1 The category list with the number of sub categories.

| カテゴリ | サブカテゴリ数 |
|-----------------------------|---------|
| Arts & Entertainment | 61 |
| College & University | 38 |
| Food | 238 |
| Nightlife Spot | 24 |
| Outdoors & Recreation | 88 |
| Professional & Other Places | 92 |
| Shop & Service | 155 |
| Travel & Transport | 50 |

バランスよく含んだマイマップを簡単に発見できる、という発想に基づいている。なお、地図を使った地図検索は、これまで調査した限り存在していない。

このシステムを実現するためには、各マイマップを何らかの値で特徴付け、その特徴を用いて類似度を算出する必要がある。本論文では、3.1.1 項および 3.1.2 項によって得られた各 POI のカテゴリ情報を特徴と捉え、マイマップに含まれるすべての POI のカテゴリを行列で示したものをカテゴリベクトル \vec{C} として下記のように定義する。

$$\vec{C} = [C_1, C_2, C_3, \dots, C_{746}]$$

$$(0 \leq C_i, 1 \leq i \leq 746) \quad (2)$$

カテゴリベクトルの次元数は 746 次元である。この数字は、POI 情報データベースの元である Foursquare 社によって独自に定義された、4 階層カテゴリのうち、3 階層目にあたるサブカテゴリの総数 746 と同数となっている。そのため、異なる POI 情報データベースを利用する場合は、次元数が変化する。表 1 は、具体的なカテゴリ (2 階層目) とサブカテゴリ (3 階層目) を示したものである。また、カテゴリベクトルの各成分、つまり各カテゴリの値は、当該マイマップに登録されているすべての POI を、カテゴリごとにカウントした数値とする。各マイマップのカテゴリベクトルは、そのマップに含まれる POI 群のカテゴリとその度合いを表す。そして、比較対象となる 2 つのマイマップのカテゴリベクトル間のコサイン類似度を計算することで、2 つのマイマップ間の類似度を測ることが可能になる。

4. プロトタイプシステムの実装

実装した検索システムのプロトタイプについて説明する。図 3 に示すように、プロトタイプシステムは、2 つのメインコンポーネントで構成されている。1 つは、図上部のオンライン特徴マッチングコンポーネントであり、もう 1 つはオフライン特徴付加コンポーネントである。オフライン特徴付加コンポーネントは、サーバ側で定期的に行われており、Maps Gallery などのマイマップサービスからマイマップを収集し、データベースに保存する機能を担う。

オンライン特徴マッチングコンポーネントは、ユーザが

*5 MeCab: <http://taku910.github.io/mecab/>

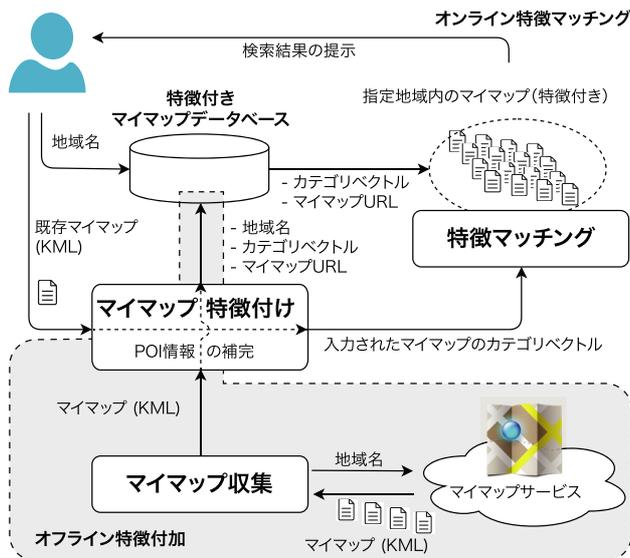


図 3 提案マイマップ検索システムのアーキテクチャ

Fig. 3 Architecture of proposed MyMap search system.

検索を行う際に用いられるもので、入力されたキーワードとマイマップに基づき、データベースから類似度の高いマイマップを提示する。

4.1 オフライン特徴付加コンポーネント

このコンポーネントは、Maps Gallery などから取得したマイマップに対してカテゴリベクトルによる特徴付けを行い、その特徴付きマイマップデータを蓄積するものであり、マイマップ特徴付けコンポーネント、マイマップ収集コンポーネントの2つのサブコンポーネントで構成されている。

4.1.1 マイマップ収集コンポーネント

このコンポーネントは、Maps Gallery にアクセスし、地域名でキーワード検索を行い、検索結果として提示されたマイマップを収集するものである。検索地域は日本に限定し、検索キーワードは「都道府県名」と「都道府県名」「市区町村名」とした。収集するデータとしては、マイマップのデータ本体である KML ファイル、マイマップを Web ページで表示するための URL、マイマップのサムネイル画像の URL の3つである。

4.1.2 マイマップ特徴付けコンポーネント

このコンポーネントは、入力されたマイマップに対して、そのマップの特徴であるカテゴリベクトルを生成することである。まず、KML ファイル形式で入力されるマイマップのデータを解析し、含まれる POI 情報をすべて取得する。次に、取得した各 POI の情報を 3.1 節で提案した POI 情報補完手法を用いて、カテゴリ情報を補完し、カテゴリベクトルを生成する。

4.1.3 特徴付きマイマップデータベース

このデータベースは、オフライン特徴付加コンポーネントにより得られる特徴付きマイマップに関連するデータを

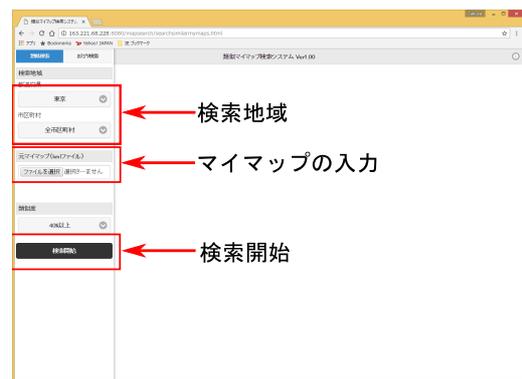


図 4 プロトタイプシステムの検索画面

Fig. 4 The search view in our prototype system.



図 5 プロトタイプシステムにおける検索結果画面

Fig. 5 The search result view in our prototype system.

蓄積する。データの構成としては、マイマップを Web ブラウザで表示するための URL、マイマップのサムネイル画像の URL、都道府県名、市区町村名、カテゴリベクトルの5つである。

4.2 オンライン特徴マッチングコンポーネント

このコンポーネントは、ユーザが指定する任意の地域におけるユーザが入力したマイマップに類似したマイマップを提供するものであり、マイマップ特徴付けコンポーネントと特徴マッチングコンポーネントの2つのサブコンポーネントから構成されている。

図 4 に、プロトタイプシステムの検索画面を示す。検索地域をテキストで指定するためのフォームと、既存マイマップを入力するフォームが設けられている。

入力されたテキストは、地域名に限定し、特徴付きマイマップデータベースから、その地域に関連するマイマップが抽出される。また、入力された既存マイマップは、オフライン特徴付加コンポーネントと共用のマイマップ特徴付けコンポーネントに入力され、POI 情報が補完された後、マイマップの特徴となるカテゴリベクトルが付与される。

4.2.1 特徴マッチングコンポーネント

キーワードに基づきデータベースから抽出されたマイマップ群に対して、入力されたマイマップとの類似度を算

出する。類似度は、双方のカテゴリベクトルのコサイン類似度を尺度とし、類似度が高い順にユーザに提示する。

図 5 は検索結果の画面である。入力に基づいて、特徴付きマイマップデータベースから抽出されたマイマップがタイトル状に表示される。各マイマップの左上には入力したマイマップとの類似度が表示される。また、地図をクリックすると登録されている POI を確認することができる。

5. 評価実験

提案手法の有効性を確認するため、POI 情報の補完手法、および、既存マイマップを入力クエリーとしたマイマップ検索システムの双方について評価を行った。それぞれの評価は独立して行い、前者は収集したデータを用いて補完性能を定量的に評価し、後者はプロトタイプシステムを用いてユーザによる主観評価を行った。

5.1 POI 情報データベースをの補完手法について

3.1.1 項で定義した POI 類似距離という指標に基づき、POI 情報データベースの中から類似度の高い POI を抽出し、POI 情報の補完を行う手法について、どの程度、情報を補完可能か検証した。ここでは、POI 情報データベース内の情報と対応付けすることを補完と呼ぶ。検証に用いたのは、収集したマイマップ内に含まれていた延べ 46,996 件の POI である。なお、POI の名称が住所のみのものや、名称に場所の意味を持たないもの（目印 1 やスポット 02 など）であった 4,259 件は除外している。本評価においては、対象となる 46,996 件の POI に関して、1 件ずつ人手で、提案アルゴリズムによって選出された POI と同一か否かを判定した。比較にあたっては、必要に応じて Google Maps の地図サービスやサーチエンジンを用い、以下の条件のいずれかを満たした場合、POI を同定できたと判定した。

条件 1 同定対象 POI と推定 POI の名称が完全一致

条件 2 同定対象 POI の名称を単語単位で順番を入れ替えると推定 POI の名称と一致

(例：「地域センター荳原第四」と「荳原第四地域センター」)

条件 3 同定対象 POI の名称が推定 POI の名称を内包（逆もあり）

(例：「品川区役所 荳原第四地域センター」と「荳原第四地域センター」)

条件 4 条件 2 と条件 3 を組み合わせた条件

(例：「品川区役所 地域センター荳原第四」と「荳原第四地域センター」)

条件 5 同定対象 POI の位置情報を用いて地図サービスで検索した際、推定 POI を確認できる場合

条件 6 サーチエンジンを用いて同定対象 POI と推定 POI の同一性を確認できる場合

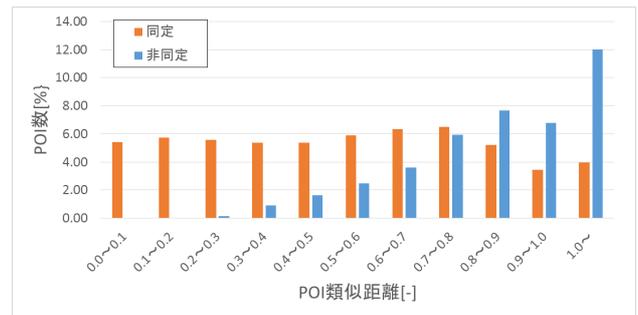


図 6 提案 POI 情報補完手法の POI 類似距離別の同定精度
Fig. 6 Identification accuracy against POI similarity distance.

評価結果と考察

比較評価の結果、提案する POI 情報補完手法によって 46,996 件中 27,645 件の POI を同定することができ、その精度は 58.82% となった。

図 6 に POI 類似距離別の同定 POI 数と非同定 POI 数を示す。横軸に POI 類似距離を示し、縦軸に POI 数を示す。この図の同定 POI 数の合計は同定精度と同値である。この図より、POI 類似距離が 0.7 以下を示す場合、同定対象 POI の 8 割以上に対して、データベース内の POI と同定できることを確認した。特に、POI 類似距離が 0.3 以下となる同定対象 POI に関しては、ほぼすべて（約 9 割 9 分）の POI が同定できることが分かった。一方、POI 類似距離が 0.7 より大きい値となる場合は、同定対象 POI の 4 割弱の POI しか同定できないことを確認した。なお、POI が同定できない要因としては、同定対象 POI に対応する POI データが Foursquare 社の POI 情報データベースに存在しないためであると考えられる。

次に、POI 類似距離を算出する式 (1) の構成要素のうち、POI 名称のレーベンシュタイン距離と物理的距離が、それぞれ POI の同定にどの程度寄与しているか確認する。

表 2 は、横にレーベンシュタイン距離、縦に物理的距離として、同定率を示したものである。レーベンシュタイン距離と物理的距離は、ともに 0 から 1 の間の値をとり、数値が小さいほど距離が近いことを意味している。この表から、レーベンシュタイン距離が 0.2 以下の場合、物理的距離に依らず 9 割以上同定できていることが分かる。しかしながら、レーベンシュタイン距離が大きい場合、物理的距離によっては同定率が大幅に低下するため、両者を加味した POI 類似距離を用いることが有効であるといえる。また、Foursquare の検索誤差の中央値は 70 [m] とされているため、物理的距離（ユーザが登録した位置と POI 情報データベース内の位置）は 0.5 以上であることも多い。こうした POI に対しても、文字列の近さから対応する POI を同定することができており、この点からも両者を加味した POI 類似距離を用いることが有効であるといえる。

表 2 レーベンシュタイン距離および物理的距離に対する同定率分布
 Table 2 Distribution of identified POIs ratio against POI similarity distance and physical distance.

| 同定率 [%] | | レーベンシュタイン距離 | | | | | | | |
|---------|-----------|-------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| | | 0.0 ~ 0.1 | 0.1 ~ 0.2 | 0.2 ~ 0.3 | 0.3 ~ 0.4 | 0.4 ~ 0.5 | 0.5 ~ 0.6 | 0.6 ~ 0.7 | 0.7 ~ 0.8 |
| 物理的距離 | 0.0 ~ 0.1 | 99.97 | 98.35 | 64.36 | 69.47 | 73.36 | 68.74 | 54.66 | 43.41 |
| | 0.1 ~ 0.2 | 99.91 | 97.67 | 56.68 | 56.00 | 58.25 | 48.51 | 35.55 | 24.79 |
| | 0.2 ~ 0.3 | 100.00 | 97.46 | 60.43 | 57.34 | 55.15 | 47.56 | 31.88 | 23.68 |
| | 0.3 ~ 0.4 | 99.91 | 96.86 | 63.04 | 52.52 | 51.31 | 44.30 | 28.94 | 21.33 |
| | 0.4 ~ 0.5 | 100.00 | 96.35 | 54.25 | 47.84 | 53.07 | 43.66 | 28.60 | 14.26 |
| | 0.5 ~ 0.6 | 100.00 | 95.93 | 55.38 | 45.75 | 48.15 | 39.15 | 22.67 | 15.54 |
| | 0.6 ~ 0.7 | 100.00 | 91.30 | 55.13 | 48.15 | 40.48 | 38.61 | 21.27 | 9.52 |
| | 0.7 ~ 0.8 | 99.40 | 97.83 | 45.24 | 42.86 | 44.80 | 27.37 | 11.56 | 8.72 |
| | 0.8 ~ 0.9 | 99.17 | 91.30 | 41.67 | 35.82 | 38.20 | 20.45 | 16.67 | 9.43 |
| | 0.9 ~ 1.0 | 98.68 | 92.86 | 34.62 | 35.48 | 24.07 | 27.78 | 9.09 | 5.50 |

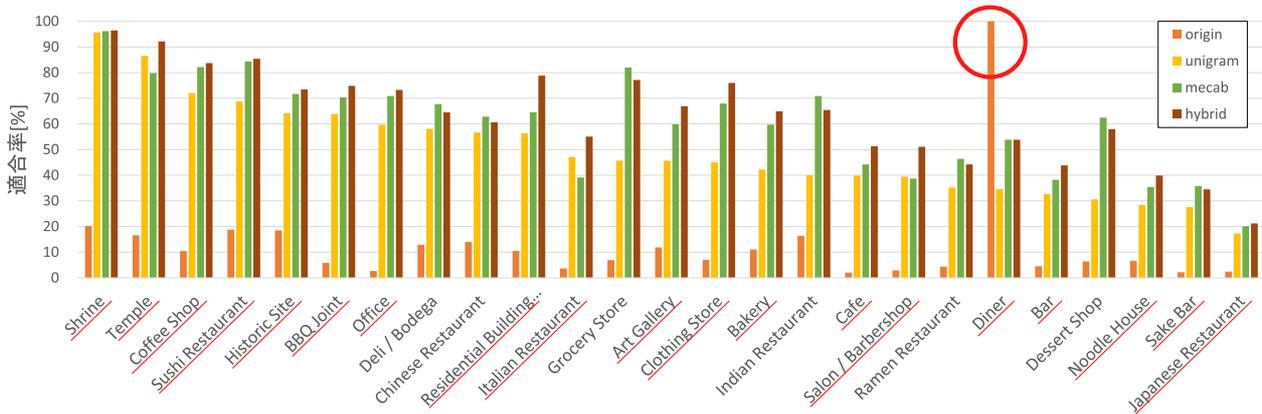


図 7 各特徴抽出方法におけるカテゴリ推定精度 (カテゴリ別)
 Fig. 7 Accuracy of category estimation for each category.

5.2 POI 名称に基づくカテゴリ情報の補完方法について

提案する各特徴抽出方法を用いた機械学習により得られた分類器のカテゴリ推定精度を評価するために、様々なマイマップに登録されている 9,200 件の POI データをテストデータとして使用する。9,200 件の POI のカテゴリ別の件数を表 3 に示す。なお、これらの POI のカテゴリ情報は POI 同定法の同定精度を測った際に得られたものである。これらの POI の名称を各分類器に入力して、正しいカテゴリをどれくらい出力するか確認し、カテゴリ推定精度を評価する。

次に、図 7 に各特徴抽出における各カテゴリの推定精度を示す。origin に関しては、図中の赤丸のようにほとんどの POI のカテゴリが Diner と推定され、分類できないことが分かる。図中の赤線が引かれたカテゴリは hybrid が最も良い推定結果であることを示し、unigram と mecab の両方の特徴抽出における良い部分をうまく組み合わせることができていると考えられる。

表 3 テストデータのカテゴリ別の POI 件数

Table 3 The number of POIs including in each category.

| カテゴリ | POI 件数 | カテゴリ | POI 件数 |
|----------------------|--------|--------------------|--------|
| Shrine | 2,060 | Dessert Shop | 157 |
| Temple | 767 | BBQ Joint | 155 |
| Ramen Restaurant | 647 | Bakery | 154 |
| Cafe | 647 | Italian Restaurant | 138 |
| Office | 605 | Salon / Barbershop | 137 |
| Residential Building | 553 | Art Gallery | 127 |
| Japanese Restaurant | 543 | Clothing Store | 100 |
| Coffee Shop | 443 | Sushi Restaurant | 96 |
| Grocery Store | 350 | Bar | 89 |
| Historic Site | 346 | Diner | 78 |
| Noodle House | 331 | Deli / Bodega | 62 |
| Chinese Restaurant | 328 | Indian Restaurant | 55 |
| Sake Bar | 232 | | |

表 4 各特徴抽出方法における平均カテゴリ推定精度 (総合)

Table 4 Mean accuracy of category estimation in total.

| 機械学習アルゴリズム | SVR | | | | SVM* |
|------------|--------|---------|-------|--------|---------|
| | origin | unigram | mecab | hybrid | unigram |
| 特徴抽出方法 | | | | | |
| 推定精度 [%] | 11.54 | 60.28 | 67.04 | 70.27 | 45.84 |

*既存研究 [11] の推定精度

評価結果と考察

各特徴抽出における平均カテゴリ推定精度を表 4 に示す. origin に関しては, 表から分かるようにカテゴリを推定することは不可能であることが分かる. unigram に関しては Choi らの既存研究の推定精度を 14.44 ポイント上回っている. これは SVR ベースの機械学習に変更したことによる影響であると考えられる. mecab は unigram より推定精度が 6.76 ポイント上回り, hybrid に関しては約 10 ポイントも高い結果を得ることができている.

5.3 類似マイマップ検索システムの評価

提案システムが, 入力された既存マイマップと地域名に応じて, 類似した POI データを含むマイマップを提示できることを評価するために, 作成したプロトタイプシステムを用いてユーザ評価を行った.

本評価実験は 10 人の被験者に実施してもらった. 評価方法としては, まず被験者にこのプロトタイプシステムを使用して, 2つの既存マイマップを用いてマイマップの検索をしてもらう. 既存マイマップは, 全被験者共通でこちらが指定したマイマップ (京都祇園周辺のカフェをまとめたもの) と, 被験者自身に実験前に自由なテーマに基づいて作成してもらったマイマップである. 次に, 各マイマップの検索結果を基に以下のアンケートに回答してもらい, 各マイマップの類似マイマップを提示することができているか評価を行う.

アンケート 1-1 指定マイマップに類似したマイマップは検索結果の内, どのくらいありますか? (類似度別)

アンケート 1-2 指定マイマップで検索した際, あなたが思う類似しているマイマップの類似度は何%以上ですか?

アンケート 2-1 オリジナルマイマップのマップタイトルを記述してください.

アンケート 2-2 オリジナルマイマップは主にどの地域の POI についてまとめていますか?

アンケート 3-1 オリジナルマイマップに類似したマイマップは検索結果の内, どのくらいありますか? (類似度別)

アンケート 3-2 オリジナルマイマップで検索した際, あなたが思う類似しているマイマップの類似度は何%以上ですか?

なお, 検索地域に関しては「東京全域」に固定して行い, 評価実験の前に東京全域にある 5,526 件のマイマップを取

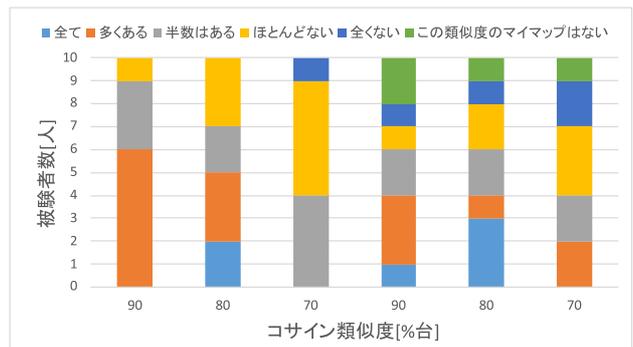


図 8 指定マイマップを用いて類似検索したときの各コサイン類似度における類似マイマップの量

Fig. 8 The number of selected similar MyMaps in case using the assigned MyMap.

集し, マイマップの特徴を示すカテゴリベクトルをそれぞれ付加している. また, 本評価実験は, 類似マイマップとして提示するための入力マイマップとの類似度合を示すコサイン類似度の閾値を選定する目的もある. 全被験者共通のマイマップを用意した理由としては被験者ごとに類似マイマップと判定するコサイン類似度の閾値がどれくらい異なるかを確認するためである.

評価結果と考察

アンケート 1-1 の結果を図 8 に示す. また, アンケート 1-2 の結果は 70% という中央値を得た. この図より, コサイン類似度が 80% 以上のとき, 提示したマイマップの半数以上は類似マイマップがあるという回答を得ることができたことが分かる. この結果とアンケート 1-2 の結果を考慮して, 指定マイマップにおける類似マイマップを提示するコサイン類似の閾値は 70~80% にあると考えられる.

次に, アンケート 2-1 の結果, 「ごはんマップ@なら」や「服に関するマップ」, 「好みのラーメン屋」など, 様々なカテゴリのマップを用いて類似マイマップ検索が行われたことが分かった. また, アンケート 2-2 の結果, 奈良や東京, 大阪などの 5 つの道府県の地域に関するマップが類似検索に用いられたことが分かった.

次に, アンケート 3-1 の結果を図 9 に示す. また, アンケート 3-2 の結果は 75% という中央値を得た. アンケート 2-1 の結果とこの図より, マイマップにまとめる POI のカテゴリの違いによって類似マイマップと提示される量が異なるということが分かる. さらに, コサイン類似度が 80% 以上のとき, 提示したマイマップのすべてが類似マイマップであるという回答をいくつか得ることができたことが分かる. この結果とアンケート 3-2 の結果を考慮して, 任意のマイマップにおいても類似マイマップを提示するコサイン類似度の閾値は 70~80% にあると考えられる.

以上の評価結果と考察から提案類似マイマップ検索システムを用いることで, マイマップを入力としてそのマップに類似したマイマップを提示できることが分かった. ま

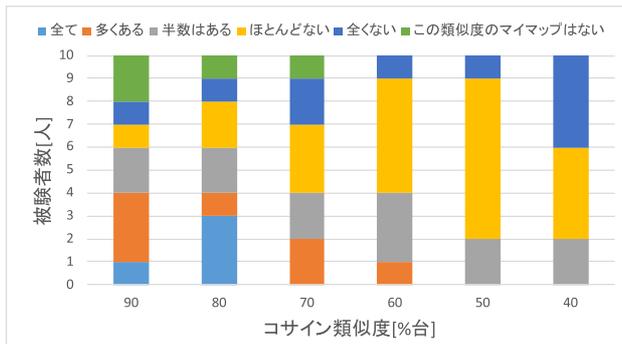


図 9 オリジナルマイマップを用いて類似検索したときの各コサイン類似度における類似マイマップの量

Fig. 9 The number of selected similar MyMaps in case using the user-original MyMap.

た、類似したマイマップであると判定するコサイン類似度の閾値は 75%程度であると考えられる。ただし、コサイン類似度が 75%以上であっても提示したマイマップの半数以上が類似していない結果の場合がある。より多くの類似マイマップを提示するには様々なカテゴリのマイマップに対応する必要があるため、提案 POI 情報補完手法の同定精度を向上させる必要があると考えられる。

6. おわりに

本論文では、これまでに作成された膨大なマイマップの中から、ユーザの目的にあったマイマップを検索可能にするため、マイマップに含まれる POI 群のカテゴリ比率を特徴として利用するマイマップ検索システムと、そのシステムを実現するための POI 情報補完手法を提案、実装した。

POI 情報の補完手法に関しては、実際のマイマップに登録されていた POI データ 46,996 件を用いた評価実験の結果、58.82%に相当する 27,645 件の POI を Foursquare 社の POI 情報データベース内に登録されている POI と同定できることを明らかにした。また、同定できなかった POI に対して、その名称に含まれる文字からカテゴリを推定する手法は、従来方式よりも 14.4%高い、70.3%の精度で推定可能であることを示した。ただし、この場合のカテゴリ数は 25 種類に限定しており、マップを特徴付けるために用いるカテゴリ数 746 種類よりも大幅に少ない。カテゴリ数が少ない場合、マップに対する特徴量の多様性が低下するため、名称からのカテゴリ推定を類似マイマップ検索に利用するためには改善が必要である。

最後に、プロトタイプを実装し、10 名の被験者によるユーザ評価実験を行った結果、カテゴリ比率のコサイン類似度が 75%以上で類似したマイマップといえることと、提案システムによって所望の情報を含んだマイマップを一定の精度で検索できることが分かった。

謝辞 本論文の一部は、JSPS 科研費 15H05708 および 15KK0011 の助成を受けたものである。

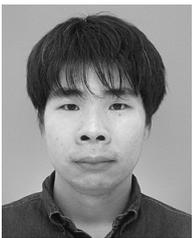
参考文献

- [1] 荒川 豊, 金平卓也, 安本慶一, 和田 健: User Generated Maps: 地理情報データ作成・共有システム, 電子情報通信学会ヒューマンプロブ研究会 (2014).
- [2] Kanehira, T., Arakawa, Y., Yasumoto, K. and Wada, T.: Curap: Curating geo-related information on a map, *Proc. IEEE International Conference on Consumer Electronics (ICCE 2016)*, pp.325-326 (Jan. 2016).
- [3] 金平卓也, 荒川 豊, 安本慶一, 和田 健: 評価機能付きパーソナル地理情報表示システムの提案と実装, 情報処理学会モバイルコンピューティングとユビキタス通信研究会 (MBL), Vol.2013-MBL-69, pp.1-6 (2013).
- [4] Nishida, K., Toda, H., Kurashima, T. and Suhara, Y.: Probabilistic identification of visited point-of-interest for personalized automatic check-in, *Proc. 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp 2014)*, pp.631-642 (Sep. 2014).
- [5] 落合桂一, 鳥居大祐: 時間変化する特徴語によるマイクロブログ地名曖昧性解消, 情報処理学会論文誌データベース (TOD), Vol.2, No.7, pp.51-60 (2014).
- [6] 三浦英俊: 緯度経度を用いた 3 つの距離計算方法, オペレーションズ・リサーチ, 経営の科学, No.60, pp.701-705 (2015).
- [7] 加藤 誠, 大島裕明, 小山 聡, 田中克己: アナロジーに基づく地理情報検索, 全国大会講演論文集, Vol.72, pp.125-126 (2010).
- [8] Bao, J., Zheng, Y. and Mokbel, M.F.: Location-based and preference-aware recommendation using sparse geo-social networking data, *Proc. 20th International Conference on Advances in Geographic Information Systems (SIGSPATIAL 2012)*, pp.199-208 (Nov. 2012).
- [9] Yuan, Q., Cong, G. Ma, Z., Sun, A. and Thalmann, N.M.: Time-aware point-of-interest recommendation, *Proc. 36th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2013)*, pp.363-372 (July 2013).
- [10] Shaw, B., Shea, J., Sinha, S. and Hogue, A.: Learning to rank for spatiotemporal search, *Proc. 6th ACM International Conference on Web Search and Data Mining (WSDM 2013)*, pp.717-726 (Feb. 2013).
- [11] Choi, S.J., Park, S.-B. and Kim, K.-Y.: Estimating Category of Pois Using Contextual Information, *Journal of Indian Journal of Science and Technology*, Vol.8, No.S7, pp.718-723 (2015).
- [12] 荒川 豊: Place API の統合, 情報処理学会研究報告, モバイルコンピューティングとユビキタス通信 (MBL), pp.1-6 (2013).
- [13] Shohei, H., Tokui, S. and Oda, S.: Jubatus: An open source platform for distributed online machine learning, NIPS 2013 Workshop on Big Learning (2013).
- [14] Crammer, K., Dekel, O., Shalev-Shwartz, S. and Singer, Y.: Online passive-aggressive algorithms, *Proc. 16th Annual Conference on Neural Information Processing Systems (NIPS)*, pp.1229-1236 (2003).
- [15] Crammer, K. and Singer, Y.: Ultraconservative online algorithms for multiclass problems, *Journal of Machine Learning Research*, Vol.3, pp.951-991 (Jan. 2003).
- [16] Crammer, K., Dekel, O., Keshet, J., Shalev-Shwartz, S. and Singer, Y.: Online passive-aggressive algorithms, *Journal of Machine Learning Research*, Vol.7, pp.551-585 (Mar. 2006).



荒川 豊 (正会員)

1977年生。2001年慶應義塾大学理工学部情報工学科卒業。2003年同大学大学院修士課程修了。2006年同大学院博士課程修了。博士(工学)。2006年同大学大学院特別研究助手(2007年より助教に変更)。2009年3月九州大学大学院システム情報科学研究院助教。2013年3月より奈良先端科学技術大学院大学准教授。主として、ユビキタスコンピューティングと行動変容に関する研究に従事。APCC 2008 Best Paper Award (2008), 情報処理学会山下記念研究賞(2011), 安藤博記念学術奨励賞(2011), ICMU2014 Best Poster Award (2014), 情報処理学会長尾真記念特別賞(2014), UbiComp/ISWC Best Demo Award (2016), IPSJ/IEEE-CS Young Computer Researcher Award (2018)等各賞受賞。ACM, IEEE, IEEE-CS, IEEE-ComSoc, 電子情報通信学会, 各会員



金平 卓也

2013年大阪府立大学工業高等専門学校総合工学システム学科卒業。2015年同校総合工学システム専攻科卒業。2017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。同年株式会社いい生活入社。自社クラウドサービス向けのAPI開発に従事。



安本 慶一 (正会員)

1991年大阪大学基礎工学部情報工学科卒業。1995年同大学大学院博士後期課程退学後、滋賀大学経済学部助手。2002年奈良先端科学技術大学院大学情報科学研究科助教授, 2011年より同研究科教授。博士(工学)。ユビキタスコンピューティングに関する研究に従事。ACM, IEEE 各会員。