マイマップによる類似マイマップ検索システムの提案と実装

金平 卓也 荒川 豊 安本 慶一

† 奈良先端科学技術大学院大学 〒 630-0101 奈良県生駒市高山町 8916-5 E-mail: †{kanehira.takuya.kj7,ara,yasumoto}@is.naist.jp

あらまし 近年、ユーザが独自に地理情報(POI: Point of Interest)を地図にまとめたマイマップの共有が行われている。しかし、各マップには適切かつ十分なメタ情報が含まれていないため、ユーザが目的のマップを見つけることが困難な状況にある。この問題に対して、本研究では POI 情報の補完手法と類似マイマップ検索システムの 2 つを提案する。提案システムでは、提案補完手法で補完された POI 情報の中でもカテゴリに着目し、マップ毎にカテゴリベクトルというマップの類似性を測るための特徴量を定義する。本稿では、マップのカテゴリベクトルの推定方法と推定結果を用いた類似マイマップ検索システムおよびこのシステムを用いた評価結果について述べる。

キーワード マイマップ, Point of Interest, カテゴリベクトル, 類似検索, 地図, 地理情報システム

1. はじめに

2005 年頃から、Google Maps や Open Street Map などの 地図サービスが普及している. これらのサービス開始当初の頃 は地図を閲覧できるだけであったが、PC やスマートフォンの 普及に伴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. また 2007 年には, ユーザ自身が地理情報 (POI: Point of Interest)を1つの地図にまとめたデジタルコンテン ツの一種であるマイマップを作成することが可能となった. こ こで、マイマップの例を図1に示す.このマップはユーザが地 理空間情報を視覚的かつ直感的に共有することができ,一般的 には Web ページに埋め込む形で共有されている. 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, 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の提 案[1][2]やマイマップに対して評価情報を付加することの有用 性の提言 [3] を行ってきた. 2014 年には, Google がマイマップ をフリーキーワードで検索することができる Maps Gallery と いう Web サービスを開始している. このサービスにより, 様々 な Web ページに埋め込まれて点在していたマイマップを一元 的に検索することが可能となった. しかし. 各マップに登録さ れている POI には適切かつ十分なメタ情報が含まれていない ため, ユーザが目的のマップを見つけることが困難な状況にあ る. 例として、東京の浅草の観光スポットのみ集めたマップを 見つけたいとする. 図 2 に Maps Gallery で「東京 浅草 観光」 と検索した際の結果を示す. また, 図3に検索結果から任意に 選択したマイマップに登録されている POI の詳細情報表示画 面を示す. 図2に示す検索結果は、ユーザが求めるマップを検 索できているように見える. しかし、観光スポット情報のみ集 めたマップもあれば、飲食店の情報を含むマップも混在してい る. よって、ユーザはマップを一つずつ選択し、登録されてい る情報を確認しながら適したマップを見つける必要がある. し かし、図3に示すように、登録されている POI に対してメタ 情報が含まれていないため、観光スポットであるかどうかすぐ に分からない状況にある.



図 1 大阪市の夜景スポットをまとめたマイマップ



図 2 Maps Gallery で「東京 浅草 観光」と検索した結果



図 3 マイマップに登録されている POI 詳細情報の例

このような背景から、本研究では既存のマイマップに登録されている POI の情報を補完する手法と類似マイマップ検索システムを提案する。提案する補完手法を用いることで、ユーザはマイマップに登録されている POI がどのようなジャンルのお店または場所であるか、理解することが容易となる。また、提案検索システムでは、ユーザが過去に作成したマイマップを入力として、任意の地域の他のユーザが作成した類似するマイマップを検索することが可能となる。つまり、ユーザの嗜好を考慮したマイマップの検索が可能となる。

提案する POI 情報補完手法はマイマップに登録されている POI と同じ POI を Foursquare の POI データベースから同定 する手法である. POI の同一性を測る指標としては新たに定 義した POI 類似距離を用いる. この POI 類似距離はそれぞれ 正規化した物理的距離と文字列の類似性を測る指標のレーベンシュタイン距離の総和で表し,値が小さいほど同一性が高いことを示す. POI を同定する際は,この類似距離をマイマップに 登録されている POI と POI データベースにある候補の POI のそれぞれの位置情報と名称を元に算出する. この手法の評価実験の結果,4万6996件中2万7645件を同定することが可能であり,同定精度は 58.82%であることを確認した.

次に、提案する類似マイマップ検索システムは、マイマップ 毎にカテゴリベクトルという特徴を付加することにより, マイ マップの類似検索を実現する. 提案システムでは、まず初めに Web クローラを用いて Web 上から膨大なマイマップのデータ を収集する. 収集した各マイマップに登録されている POI は 提案 POI 情報補完手法により情報補完する. ここで,情報補完 した POI 情報の中でもカテゴリ情報に着目する. マイマップ に登録されている POI のカテゴリ情報をカテゴリ毎にカウン トし、カテゴリベクトルを生成する. 生成したカテゴリベクト ルはマイマップのデータと共に提案システム内のデータベース に保管する. 類似マイマップを検索する際に入力されるユーザ が過去に作成したマイマップに対しても同様にカテゴリベクト ルを生成する. このカテゴリベクトルは、マイマップがどのよ うなジャンルの POI をどれだけ登録しているかを示し、マイ マップの特徴に相当する. また, マイマップを作成したユーザ の嗜好が表現されていると考えられる. 提案システムで類似マ イマップを検索する際は、このカテゴリベクトルのコサイン類 似度に基づき,類似マイマップを提示する.システムの評価実 験では、実際にこのシステムを10人の被験者に使用してもら い、被験者が作成したマイマップと類似したマイマップを提示 できているか確認した. また, 類似していると判断するための カテゴリベクトルのコサイン類似度の閾値を確認した。その結 果,提案システムを用いることで類似マイマップが提示するこ とができ、類似しているコサイン類似度の閾値は75%であるこ とがわかった.

本稿の2章では、本研究と関連する POI 同定に関する研究 について述べる.3章では、提案 POI 情報補完手法とカテゴリベクトルによるマイマップの特徴付けについて述べる.4章では、提案システムである、類似マイマップ検索システムについて述べる.5章では、提案 POI 情報補完手法およびシステムの

評価実験の概要と実験結果およびその考察について述べる. 6 章では、本研究のまとめと今後の課題について述べる.

2. 関連研究

本研究で提案する類似マイマップ検索システムにおいて, 任 意地域の類似マイマップを提示するために我々が定義するマイ マップの類似性を測る指標のカテゴリベクトルが必要となる. カテゴリベクトルはマイマップに登録されている POI のカテ ゴリ情報より生成するが, 既存のマイマップに登録されている POI にはカテゴリ情報がない. そのため提案 POI 情報補完手 法を用いて情報補完しようと考えている. この提案手法では, マイマップに登録されている POIと Foursquare に登録されて いるカテゴリ付き POI を同定することにより情報補完を行う. つまり、本研究において POI 同定技術は重要である. POI 同 定に関する研究としては西田[4]らの,確率的訪問 POI 同定法 がある. これは、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ている. POI 同 定する際は同定対象の位置情報以外にユーザの SNS のチェック イン履歴から好みのカテゴリやそのカテゴリへの訪問回数の情 報が必要となる. しかし, 本研究で扱うマイマップからこれら の情報は得ることができないため, 彼らのモデルに適用できな い. そこで、本研究では登録されている POI の位置情報と名 称のみを用いた新たな POI 同定方法を提案し、POI 情報を補 完する.

次に、マイマップを推薦するに研究に関しては、我々が知る限り行われていない。任意の地域における POI の推薦システムに関しては様々な研究が行われている。加藤 [5] ら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 [6] ら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 [7] らは、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。本研究では複数の POI を 1 つにまとめたマイマップをユーザ嗜好を考慮して推薦したいと考えている。

3. 提案手法

本章では、マイマップに登録されている POI の情報を補完 する手法について、および、マイマップの類似性を測る指標と なるカテゴリベクトルについて述べる.

3.1 POI 情報補完手法

現在、マイマップに登録されている POI に対して適切かつ十分なメタ情報が含まれていないため、ユーザはどのようなジャンルの POI をまとめたマイマップか分からない状況にある。そこで、マイマップに登録されている POI の情報を基に情報補完する POI 情報補完手法を提案する。図 4 に提案 POI 情報補完手法の概念図を示す。提案手法は、マイマップに登録されている POI を数多くの POI 情報を保有している Foursquare の POI データベースから同定することにより、情報を補完する。

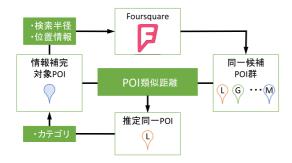


図 4 POI 情報補完手法の概念図

この際、マイマップに登録されている POI の位置情報(経緯 度)と名称(店名等)を用いる.なお,本研究で補完する情報 としては場所やお店のジャンルを示す POI のカテゴリである. 提案手法ではまず、Foursquare の POI 検索 API を用いて情報 補完対象 POI と同一な POI の候補群を取得する. この候補群 を同一候補 POI 群とする. POI 検索に必要な検索中心位置と しては情報補完対象 POI の位置情報を用い、検索半径 R とし ては 100[m] を用いる. この検索半径の値は, Foursquare の検 索誤差が約 70[m] ある [8] ことを考慮して決定した. 次に,情 報補完対象 POI と各同一候補 POI の間の POI 類似距離を算出 する、POI 類似距離は我々が提案する2つのPOIの同一性を 測る指標である. 詳細に関しては後述する. 情報補完対象 POI との POI 類似距離が最も小さい同一候補 POI を推定同一 POI とする. 最後に, Fousquare から得られる POI にはカテゴリ 情報が含まれているため同一と推定された POI のカテゴリを 情報補完対象 POI のカテゴリとして情報補完する.

POI 類似距離

POI 類似距離は 2 つの POI の類似性を測る指標であり,値が小さいほど類似性が高いことを示す.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.

$$S(P_A, P_B) = L(t_A, t_B) + \frac{D(p_A, p_B)}{100}$$
 (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 地点の経緯度を入力するとその 2 地点間の距離を求めることができるヒュベニの公式を用いている。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 $0 \sim 1$ 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 情報を補完できないとする.

3.2 カテゴリベクトルによるマップの特徴付け

本研究の目的として, ユーザが入力したマイマップに類似し

表 1 カテゴリおよび各カテゴリのサブカテゴリの数

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

図 5 京都 祇園 観光マップの KML 形式のデータの一部

たマイマップをユーザに対して提供することである。しかし、マイマップには類似性を測る指標がない。そこで、以下に示すカテゴリベクトル \vec{C} をマイマップの類似性を測る指標として提案する。

$$\vec{C} = [C_1, C_2, C_3, \dots, C_{746}] \qquad (0 \le C_i, 1 \le i \le 746)$$
 (2)

ここで、表1に本研究で用いるカテゴリ及び各カテゴリのサブカテゴリの数を示す。これは位置情報ソーシャルネットワークサービスの Foursquare が定めた 4 階層からなる階層カテゴリの内、3 階層までのカテゴリとサブカテゴリである。カテゴリベクトルの次元数は表1のサブカテゴリの総数と同値であり、746次元である。カテゴリベクトルの各成分、つまり各カテゴリの値はマイマップに登録されている全ての POI のカテゴリをカテゴリごとにカウントしたものである。よって、カテゴリベクトルはマイマップにどうのようなカテゴリの POI がどれだけ登録されているかを示し、この情報をマイマップに付加することにより、マップ同士の類似性を測ることを可能とする。なお、カテゴリベクトルを生成する際、マイマップに登録されている POI のカテゴリは事前に提案 POI 情報補完手法により情報補完されているものとする。

4. 類似マイマップ検索システム

本章では、初めに提案する類似マイマップ検索システムの概要について述べる.次に、システムアーキテクチャと3つの主要コンポーネントについて述べる.

4.1 システム概要

提案システムは、ユーザの嗜好に合う任意地域のマイマップ







a-1:京都 祇園 観光マップ

b-1:東京 浅草 観光マップ1



c-1:東京 浅草 観光マップ2



図 6 a:ユーザが作成したマイマップと別地域の b:類似したマイマップおよび c:類似していないマイマップと各マップのカテゴリ分布とカテゴリベクトル

を発見することに有用な検索システムである.提案システムでは、提案 POI 情報補完手法を用いてマイマップに登録されている POI の情報補完を行い、マイマップに対してカテゴリベクトルを付加することにより、検索地域に非依存でユーザが作成したマイマップに類似したマイマップの検索を実現する.つまり、提案システムに対してユーザが作成したマイマップと検索地域を入力すると、ユーザはそのマイマップのカテゴリベクトルに類似した指定地域の類似マイマップを得ることが可能となる.また、カテゴリと検索地域を指定することで、同様の結果を得ることも可能である.なお、提案システムにおいて、マイマップのデータ入力は KML 形式のファイルで扱う.KML形式のマイマップのデータ例を図5に示す.図5が示すようにKML 形式は XML 形式と同等と考えてよい.

ここで、図6に3つのマイマップおよび各マップのカテゴリ 分布,カテゴリベクトル $\vec{C_x}$ を示す.図6のa-1 はあるユーザ が作成した京都の祇園の観光マップを示し、b-1 と c-1 は他の ユーザが作成した東京の浅草の観光マップを示す。a-1 と b-1 は寺院と商店が立ち並ぶ道に関する POI を多くまとめたマッ プであり、c-1 は飲食店に関する POI を多くまとめたマップで ある. a-2~c-2 は各マイマップに登録されている POI のカテ ゴリを示しており、図中の $Ci(i = 1, 2, \dots, 7)$ はカテゴリを 示す. 同数字であれば、その POI は同じカテゴリであること を意味する. また、下方には各マイマップのカテゴリベクトル \vec{C}_x を示す. この図から分かるように、カテゴリベクトルの値 はマイマップに登録されている全ての POI のカテゴリをカテ ゴリごとにカウントしたものである. 図6のマイマップとカテ ゴリベクトルを用いて提案システムの動作例について述べる. あるユーザが東京の浅草の観光マップを探したいと考える. こ のユーザが提案システムに対して, 自身が作成したマイマップ (a-1)と検索地域(東京 浅草)を入力すると、提案システムは 提案 POI 情報補完手法により入力されたマイマップに登録さ れている POI の情報補完を行い、このマイマップのカテゴリ ベクトル (\vec{C}_a) を生成する. その後, 指定地域に該当するマイ マップ(b–1 と c–1)のカテゴリベクトル($\vec{C_b}$ と $\vec{C_c}$)との類似

度($0\sim100\%$ の間の数値)を算出し,75%を超えた場合はユーザへ類似マイマップとして提供する.なお,カテゴリベクトルの類似度の算出に関しては 4.2 章で述べる.また,閾値は評価実験の結果より決定した.今回の例の場合, \vec{C}_a と \vec{C}_b のの類似度は 87%であり,b-1 のマイマップはユーザへ類似マイマップとして提供する.一方, \vec{C}_a と \vec{C}_c の類似度は 47%であり,c-1 のマイマップはユーザへ提供しない.

4.2 システムアーキテクチャ

提案類似マイマップ検索システムのアーキテクチャを図7に示す。図7に示すように、提案システムは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はMaps Gallery にある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、特徴付きマイマップのデータを蓄積する。オンライン特徴マッチングコンポーネントはユーザが入力したマイマップに特徴付けを行い、類似するマイマップを蓄積した特徴付きマイマップから見つけ出しユーザへ提供する。この2つのメインコンポーネントと特徴付きマイマップデータベースについて以下に述べる。また、各メインコンポーネントでも用いる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、特徴マッチングコンポーネントの3つのサブコンポーネントについても述べる。

オンライン特徴マッチングコンポーネント

このメインコンポーネントの目的は、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 マップを提供することである.この目的を達成するために,こ のコンポーネントは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. 各サブコンポーネントについては後述する. このコンポーネントではまず初めに、マイマップ特徴付けコン ポーネントにユーザからマイマップのデータが入力される. マ イマップ特徴付けコンポーネントにマイマップのデータが入力 されると、そのマップの特徴であるカテゴリベクトルを出力す る. 出力されたカテゴリベクトルはユーザからの検索地域名と 共に特徴マッチングコンポーネントに入力される. 特徴マッチ ングコンポーネントは特徴付きマイマップデータベースの中か ら入力された特徴に類似するマイマップを検索し, 複数の類似 マイマップをユーザへ提供する. この検索の際, 検索地域名を 元に地域でフィルタリングする. なお, このメインコンポーネ ントはユーザから類似マイマップの検索が開始された際に実行 される.

オフライン特徴付加コンポーネント

このメインコンポーネントの目的は、Maps Gallery 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ことである。この目的を達成するために、このコンポーネントは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。ここで、マイマップ特徴付けコンポーネントで共2つのメインコンポーネントで共

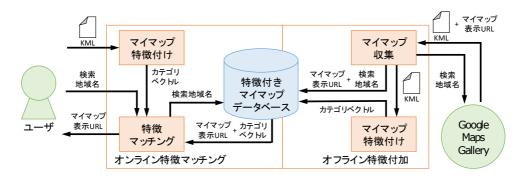


図 7 提案類似マイマップ検索システムのアーキテクチャ

通利用されており、提案システムにおいて重要なコンポーネントであることが分かる。各サブコンポーネントの詳細については後述する。このコンポーネントはまず初めに、マイマップ収集コンポーネントが、定期的かつ自動的に Maps Gallery から任意地域のマイマップを収集する。収集したマップはマイマップ特徴付けコンポーネントへ入力され、出力されるそのマップの特徴であるカテゴリベクトルと共に特徴付きマイマップデータとしてデータベース蓄積する。

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.

マイマップ特徴付けコンポーネント

このサブコンポーネントの目的は、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。このコンポーネントでは、まず初めに 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、登録されている POI 情報を取得する。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補完し、カテゴリベクトルを生成する。カテゴリベクトルの生成方法に関しては3.2 節で既に述べているため省略する。

 \mathbf{s}

マイマップ収集コンポーネント

このサブコンポーネントの目的は、地域別にマイマップに関するデータを収集することである。収集方法としては、Maps Gallery で地域名でキーワード検索を行い、検索結果ページをWeb クローリングする。なお、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"」である。例として図8に検索キーワードが「大阪大阪市」のときの検索結果ページを示す。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。KML ファイルはマイマップ特徴付けコンポーネントへ出力する。各 URL と検索地域名はマイマップ特徴付けコンポーネントから出力されるカテゴリベクトルと共に特徴付きマイマップデータベースに蓄積する。

特徴マッチングコンポーネント

このサブコンポーネントの目的は、入力される検索地域名とカテゴリベクトルを基に、特徴付きマイマップデータベースから類似したカテゴリベクトルを持ち、なおかつ検索地域に該当するマイマップをユーザへ提供することである。このコンポーネントではまず、特徴付きマイマップデータベースから、ユーザによって与えられた検索地域に該当するマイマップのカテゴリベクトルを取得する。得られた各カテゴリベクトルと与えられたカテゴリベクトルのコサイン類似度を算出し、類似性を測る。提案システムではコサイン類似度が75%以上のとき、類似していると判定する。そして、カテゴリベクトルが類似していると判定されたマイマップの情報をユーザへ提供する。提供するマイマップの情報としては、Webブラウザで表示するためのマイマップのURL、マイマップのサムネイル画像URL、マップのコサイン類似度である。

5. 評価実験

本章では、提案 POI 情報補完手法の同定精度と類似マイマップ検索システムのそれぞれの評価方法と結果について述べる.

5.1 POI 情報補完手法の同定精度

評価方法

提案 POI 情報補完手法により、どの程度情報補完できるかを確認するために、POI の同定精度を評価する.評価で用いる同定対象 POI のデータは、実際のマイマップに登録されている 4万 6996 件の POI データである. なお、4259 件の 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等)は除外している.評価方法としては、同定対象 POI と推定同一 POI のそれぞれの情報を目視で比較し、同定できているかを確認する.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て比較を行う. 具体的には以下の条件のいずれかを満たした場合、同定できていると判定する.

同定条件 1: 同定対象 POI と推定同一 POI の名称が完全一致 同定条件 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同一 POI の名称と一致

(例:「<u>地域センター</u>荏原第四」と「荏原第四<u>地域センター</u>」) **同定条件 3**: 同定対象 POI の名称が推定同一 POI の名称を内包(逆もあり)



図 8 提案 POI 情報補完手法の POI 類似距離別の同定精度



図 9 各 POI 類似距離における POI 名称のレーベンシュタイン距離 の同定数分布



図 10 各 POI 類似距離における POI 名称のレーベンシュタイン距離 の非同定数分布

(例:「品川区役所 <u>荏原第四地域センター</u>」と「<u>荏原第四地域セ</u> $\underline{\nu}$ $\underline{\nu}$ $\underline{\nu}$ $\underline{\nu}$)

同定条件 4:条件2と条件3を組み合わせた条件

(例:「品川区役所 <u>地域センター</u>荏原第四」と「荏原第四<u>地域センター</u>」)

同定条件 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,推定同一 POI を確認できる場合

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

評価結果と考察

比較評価の結果,提案 POI 情報補完手法によって 4万 6996 件中 2万 7645 件の POI を同定することができ、58.82%の同定精度を得ることができた。ここで、図 9 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 以下の場合、ほとんどの同定対象 POI が同定することが分かる。推定同一 POI との POI 類似距離が 0.7 以下の同定対象 POI に関しては8割以上の POI が同定できることを確認した。特に、POI 類似距離が 0.3 以下の同定対象 POI に関してはほぼ全て(約 9割 9分)の POI が同定できることが分かった。一方で、



図 11 各 POI 類似距離における物理的距離の同定数分布



図 12 各 POI 類似距離における POI 名称の物理的距離の非同定数 分布

POI 類似距離が 0.7 より大きい場合は、同定できない POI の 方が多くなっていることが図 9 から分かる.推定同一 POI と の POI 類似距離が 0.7 より大きい同定対象 POI に関しては、4 割弱の POI しか同定できないことを確認した.つまり、POI 類似距離が 0.7 より大きい場合は同定することが困難であると 考えられる.なお、POI が同定できない要因としては、同定対象 POI と同一の POI データが Foursquare の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 の構成要素である、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 同定にどのように寄与しているか確認する。図 10 に POI 類似距離別かつ POI 名称のレーベンシュタイン距離別の同定 POI 数を示し、図 11 に非同定 POI 数を示す。横軸と縦軸は図 9 と同様であり、POI 名称のレーベンシュタイン距離は色分けをして示す。図 10 より、POI 類似距離が 0.7 以下のとき、POI 名称のレーベンシュタイン距離は 0.6 以下が多いことが分かる。また、図 11 より、POI 類似距離が 0.7 より大きいとき、POI 名称のレーベンシュタイン距離は 0.6 より大きいとき、POI 名称のレーベンシュタイン距離は 0.6 より大きいものが多いことが分かる。

次に、図 12 に POI 類似距離別かつ物理的距離別の同定 POI 数を示し、図 13 に非同定 POI 数を示す。横軸と縦軸は図 9 と同様であり、物理的距離は色分けして示す。図 12 より、同定できる POI の大半が物理的距離が 0.3 以下であることが分かる。これは、同定できる POI の内の 8 割強に該当する。全体の割合からすると 45.59%に当たる。なお、図 13 から分かるように、同定できない POI の大半も物理的距離が 0.3 以下であることに注意する必要があり、これは同定できない POI の内の 6 割に該当する。全体の割合からすると 25.44%に当たる。つまり、同定対象 POI と同一候補 POI の物理的距離が 0.3(30[m]) 以下であれば、その候補の POI が 45.59/(45.59 + 25.44) = 66.08%の確率で同一 POI であると考えられる。

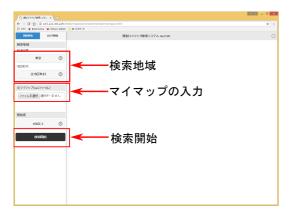


図 13 提案類似マイマップ検索システムのプロトタイプ



図 14 提案類似マイマップ検索システムのプロトタイプにおける検 - 家例

以上のことから、POI 類似距離を用いて POI を同定する際、同定対象 POI と推定同一 POI の名称のレーベンシュタイン距離が 0.6 以下であり、物理的距離が 0.3 以下の場合、多くの POI を同定できると考えられる。つまり、物理的距離より POI 名称のレーベンシュタイン距離の方が POI 同定には大きく寄与していると考えられる。

5.2 類似マイマップ検索システムの評価 評 価 方 法

提案類似マイマップ検索システムが任意の地域におけるユーザが入力したマイマップと類似したマイマップを提示できることを評価するために、図 14 に示すプロトタイプシステムを構築した。図 15 にプロトタイプシステムで類似マイマップを検索した際の検索結果例を示す。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 つのマイマップの類似マイマップを検索してもらう。1 つは全被験者共通でこちらが指定したマイマップであり、京都祇園周辺のカフェをまとめたマップである。もう 1 つは被験者自身が作成するオリジナルマイマップであり、自由なテーマで POI をまとめている。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

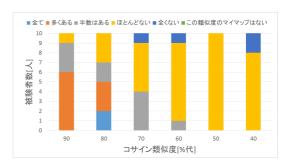


図 15 指定マイマップを用いて類似検索したときの各コサイン類似度 における類似マイマップの量

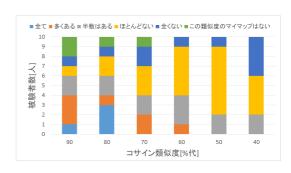


図 16 オリジナルマイマップを用いて類似検索したときの各コサイン 類似度における類似マイマップの量

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」に固定して行う.このため、評価実験の前に東京全域にある5526件のマイマップを収集し、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毎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 16 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 17 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全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わかった。また、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

ユーザが目的のマイマップを見つけることが容易になるように、POI情報補完手法と類似マイマップ検索システムの2つを提案した.

提案 POI 情報補完手法では、マイマップに登録されている POI の名称と位置情報を用いて Foursqure の POI データベースから POI を同定することにより POI の情報補完を行うことを提案した。また、POI を同定する際は POI 類似距離という、同定対象 POI と同一候補 POI の名称のレーベンシュタイン距離と物理的距離を用いた同定指標を提案した。評価実験の結果、実際のマイマップに登録されている POI データ 4万 6996 件の内、2万 7645 件の POI を同定できることが分かった。つまり、提案 POI 情報補完手法の同定精度が 58.82%であることが分かった。

提案類似マイマップ検索システムでは、カテゴリベクトルというマイマップの類似性を測る指標を新たに定義した.このカテゴリベクトルにより、マイマップにどのようなカテゴリのPOIがどれだけ登録されているか分かるようになった.また、カテゴリベクトル同士のコサイン類似度を計算することによりマイマップの類似性を測ることが可能となった.提案類似マイマップシステムのプロトタイプシステムを構築し、評価実験を行った結果、任意のマイマップを入力するとそのマイマップに類似したマイマップを提示することが可能であることがわかっ

た. また,類似したマイマップであると提示する際のコサイン類似度の閾値は75%程度であることが分かった.

今後の展望としては、提案類似マイマップ検索システムでより多くの類似マイマップを提示できるようにするために、提案POI 情報補完手法の同定精度を向上させる、またはこの方法を補助するような機能を追加していきたいと考えている。具体的には、Foursquare から得られる膨大な POI 名称とカテゴリの関係を機械学習を用いてモデルを構築し、POI の名称情報のみでその POI のカテゴリを推定する機能である。

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 ヒューマンプローブ研究会,2014.
- [2] Takuya Kanehira, Yutaka Arakawa, Keiichi Yasumoto and Takeshi Wada, "Curap: Curating geo-related information on a map", In Proceedings on IEEE International Conference on Consumer Electronics (ICCE 2016), pp. 325 – 326, 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] Kyosuke Nishida, Hiroyuki Toda, Takeshi Kurashima and Yoshihiko Suhara, "Probabilistic identification of visited point-of-interest for personalized automatic check-in", In Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing(UbiComp 2014), pp. 631–642, 2014.
- [5] 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.
- [6] Jie Bao, Yu Zheng and Mohamed F. Mokbel, "Location-based and preference-aware recommendation using sparse geo-social networking data", In Proceedings of the 20th International Conference on Advances in Geographic Information Systems (SIGSPATIAL 2012), pp. 199–208, 2012.
- [7] Quan Yuan, Gao Cong, Zongyang Ma, Aixin Sun and Nadia Magnenat Thalmann, "Time-aware point-of-interest recommendation", In Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval(SIGIR 2013), pp. 363–372, 2013.
- [8] Blake Shaw, Jon Shea, Siddhartha Sinha and Andrew Hogue, "Learning to rank for spatiotemporal search", In Proceedings of the Sixth ACM International Conference on Web Search and Data Mining(WSDM 2013), pp. 717–726, 2013.