

ピクトリアルマップの自動生成のための SNS上の投稿写真的位置情報と意味情報を考慮したPOI抽出

周 文トウ[†] 牛尼 剛聰^{††}

† 九州大学大学院芸術工学府 〒815-8540 福岡県福岡市南区塩原4-9-1

〒 815-8540 福岡県福岡市南区塩原 4-9-1

E-mail: †zhou.wentong.904@s.kyushu-u.ac.jp, ††ushiama@design.kyushu-u.ac.jp

あらまし ピクトリアルマップは地域の特徴や雰囲気を図を用いて表現する地図である。ピクトリアルマップ上に提示するユーザが関心を持っているスポットであるPOIは、従来の手法では、写真の低レベルの位置情報だけを用いて密度ベースのクラスタリング手法であるDBSCANを行って、密度が高いスポットとして抽出している。しかし、本研究では、より人間にとて意味のあるPOIを抽出するために、画像の内容を考慮したSemantic-DBSCAN手法を提案する。まず、画像認識で識別された画像のラベルとその確信度を基にエンベディング処理を行い、意味情報をもつ高次元の特徴ベクトルを得る。次に、距離と意味情報の両方から構成した潜在空間内にクラスタリングを行う。最後に、実データを用いた実験により、提案手法の有効性を検証する。

キーワード SNS, クラスタリング, 画像認識, 可視化

1 はじめに

ピクトリアルマップは、図やイラストを用いて、ある地域の特徴や雰囲気を分かりやすく表現する地図である。図1¹は人手で作成されたピクトリアルマップの例である。このようなピクトリアルマップを利用して、ユーザがその地域の面白さや魅力などを感じられるようになる。しかし、このようなピクトリアルマップを制作するためには、多くの時間と労力が必要である。また、人手で選ばれた地域の特徴的なものは、ユーザが実際に興味を持っているものとずれるかもしれない。そのため、ユーザの興味を見つけて地域の魅力にマッチングすることが必要である。



図 1 ピクトリアルマップの例: 大原野観光マップ

一方、近年、スマートフォンの普及やSNS(Social networking

service) の発展によって、ユーザーが目の前の風景を見た時、その瞬間を記録するため、写真に撮影し、SNS に投稿することが一般的に行われるようになった。SNS 上の投稿写真の量も急速に増加している。投稿写真は投稿者の興味や関心を反映されると言える。その中の位置情報付き写真はその地域の魅力を記録していると考える。そこで、膨大な位置情報付き投稿写真から地域の特徴や魅力を抽出し、抽出された特徴を表す写真を選択してマップ上にマッピングすることで、ピクトリアルマップを自動的に生成できると考えられる。

我々は、これまでにピクトリアルマップを自動的に生成する手法を提案している^[1]. しかし、この手法では、POIを抽出する時、SNSの画像の位置情報しか利用していなかったため、ユーザが興味を有するPOIを抽出できない場合があった. 本研究は、より人間にとて意味のあるPOIを抽出するために、意味情報と位置情報の両方を利用して、高次元の潜在的セマンティック空間を構築する. そして、潜在的セマンティック空間内でクラスタリングを行う.

本論文の構成は以下の通りである。第2章では関連研究を紹介する。第3章は提案手法について述べる。第4章は実データを用いた実験について述べる。第5章はまとめと今後の課題について述べる。

2 関連研究

2.1 SNS の投稿写真の意味情報を考慮した POI の自動抽出に関する研究

Chen ら^[2] は SNS 上の投稿写真の位置情報と投稿者自分が付いたタグの両方を利用してクラスタリングを行って POI を自動的に抽出する手法を提案している。 Hashimoto ら^[3] は画像認識 API で取得したラベル情報を基に、教師あり機械学習

1 : <https://gurutabi.gnavi.co.jp/area/kyoto/kyotoshi/oharano.html>

手法を用いて、最も代表的な地域の特徴タグを抽出する手法を提案している。滝本ら^[8,9]は画像認識 API で取得したラベル情報と確信度により重み付きタグ集合を構成し、構成したタグ集合間の類似度を考慮する SimDBSCAN を提案している。

本研究では、画像認識 API で取得したラベル情報と確信度を基に特徴ベクトルに変換する。そして、地理上の距離と特徴ベクトル間の距離の両方を考慮する Semantic-DBSCAN を提案する。

2.2 SNS の写真を利用した可視化に関する研究

SNS に投稿された膨大な写真データを分析・可視化することで、写真から反映された潜在的な内容に関する研究が活発に行われている。Redi ら^[10]は Instagram に投稿された写真的内容と写真的スタイルを分析して可視化することで、全世界有名な五つの大都市間の文化的類似点と相違点を明らかにしている。Manovich ら^[11]は Instagram に投稿する自撮り写真的内容を分析して可視化することで、世界中の 5 都市間のユーザの特性を明らかにしている。

本研究では、SNS の写真データを利用して、地域の特徴を反映されるピクトリアルマップを自動的に生成する手法を提案する。

3 提案手法

3.1 意味情報を考慮した POI の抽出

POI(Point of Interest) は、地理空間上で人々が興味を持つ場所を表す。SNS 上で話題になるスポットや、大勢の人が訪ね、チェックインする場所などが POI の候補となる。POI の付近では撮影・投稿する人数が多く、撮影した内容も重複しやすいと考えられる。そこで、本研究では、SNS 写真的投稿数が密集する、かつ同じオブジェクトを撮影する地点を抽出すると、ユーザの関心地点が発見できると考える。

3.1.1 画像認識による意味情報の取得

Hwang^[12] らは、画像の視覚属性 (visual attribute) を、物体カテゴリ間で共有される人間が理解可能な属性であると定義している。本研究では、画像認識技術を用いて、画像の視覚属性を識別する。近年、ディープラーニングを用いて高精度に画像認識を行うことができるようになった。そして、いくつかの代表的な画像認識アルゴリズムが WebAPI として利用できるようになっている。本研究では、代表的な画像認識 API の一つである Computer Vision API^[13] を利用して、対象となる全ての画像に関して画像認識を行う。Computer Vision API によって得られたタグ情報の例を、図 2 に示す。本研究では、タグと確信度の両方を利用する。

3.1.2 意味情報の特徴ベクトルの変換

前節で得られたタグと確信度を利用して意味情報の特徴ベクトルに変換するプロセスについて説明する。

いま対象となる地域の写真から取得したタグが合計 10 個 {"flower", "tree", "hill", "sky", "river", "temple", "building", "car", "tower", "moon"} であった場合を例に考え



図 2 Microsoft ComputerVision API による取得した結果の例

る。ある写真に Computer Vision API を適用した結果が {"flower": 0.85, "tree": 0.91, "sky": 0.75, "river": 0.65} となったとする。このとき、この写真のエンベディング結果は $x = (0.85, 0.91, 0, 0.75, 0.65, 0, 0, 0, 0, 0)$ となる。このようにすることで、画像の視覚属性をベクトル化することができ、ベクトル間の距離を計算することができる。

3.1.3 t-sne による特徴量の次元削減処理

Computer Vision API による取得可能なタグの種類は合計 3500 以上を超える。一枚の画像に対して、Computer Vision API による取得したタグを特徴ベクトルに変換した後、特徴ベクトルの長さは 3500 以上になる。そこで、本研究は、t-sne (t-Distributed Stochastic Neighbor Embedding) を用い、高次元の画像特徴ベクトルを 2 次元に次元圧縮する。

3.1.4 画像間の類似度の計算

本研究では、ユークリッド距離を用いて画像間の類似度を計算する。ユークリッド距離は、2 点間の距離を算出する方法の一つで、2 点が直線的にどれだけ離れているのかを示している。 x と y のユークリッド距離は以下の式を用いて求める。

$$d(x, y) = \sqrt{\sum_{i=1}^n |x_i^2 - y_i^2|} \quad (1)$$

$d(x, y)$ は 0 に近いほど、 x と y が類似する。

3.1.5 位置情報を用いた DBSCAN

本研究では、まず対象地域の全ての写真的位置情報を基づき、密度ベースのクラスタリング手法 DBSCAN^[14] (Density-based spatial clustering of applications with noise) を用いクラスタリングを行う。DBSCAN は、クラスタ間の距離の閾値 Eps とクラスタを構成する最小データ数の閾値 $MinPts$ との 2 つの閾値を持つ。ある点 x から、距離 Eps 内にある点集合を近傍 $N_{Epx}(x)$ と定義し、以下の接続関係を満たす時、同じクラスタに分類すると考えられる。

$$y \in N_{Epx}(x) \quad (2)$$

$$|N_{Epx}(x)| \geq MinPts(\text{コア点}) \quad (3)$$

対象地域内に SNS の投稿写真的地理的密度分布が異なる。DBSCAN 手法では、密度によって適切に Eps と $MinPts$ を選択する必要がある。

3.1.6 意味情報と位置情報を考慮した Semantic-DBSCAN

前節で抽出した 1 つクラスタにおいて、複数のテーマの写真が集まっていると考える。例えば、ある観光スポット A を一つ

クラスタとして抽出する場合、スポット Aにおいて、人気のあるレストランや有名な建物、または自然風景など、複数の見どころが存在する可能性がある。そこで、それぞれ見どころを POI として抽出するため、意味情報と位置情報を考慮したクラスタリングを行う。

3.1.5 で利用した従来の二次元の DBSCAN はクラスタ間の距離の閾値 Eps とクラスタを構成する最小データ数の閾値 $MinPts$ という二つパラメータが必要となる。Derya ら^[15] は二次元の DBSCAN をベースに、三次元の ST-DBSCAN を提案している。本研究は三次元の ST-DBSCAN をベースに Semantic-DBSCAN を提案する。Semantic-DBSCAN は意味情報を考慮した Eps_2 を追加して三つパラメータ $Eps_1, Eps_2, MinPts$ を有する。 Eps_1 は二つ写真の位置情報から計算した地理的な距離の閾値であり、 Eps_2 は 3.1.3 で説明した二つ特徴ベクトル間のユークリッド距離の閾値であり、 $MinPts$ はクラスタを構成する最小データ数の閾値である。関数 $RetrieveNeighbors(o, D, Eps_1, Eps_2)$ を以下のように定義する。

$$\begin{aligned} & RetrieveNeighbors(o, D, Eps_1, Eps_2) \\ &= \{x \in D, dist_{geo}(o, x) \leq Eps_1, \\ & \quad dist_{sem}(o, x) \leq Eps_2\} \end{aligned} \quad (4)$$

ここで $dist_{geo}(o_1, o_2)$ は o_1 と o_2 の地理的な距離を表し、 $dist_{sem}(o_1, o_2)$ は o_1 と o_2 の意味的な距離を表す。Semantic-DBSCAN のアルゴリズムをアルゴリズム 1 に示す。

4 実験

4.1 データセット

提案手法の有効性を評価するための実験を行った。本実験では Flickr にアップロードされた画像から構成された写真のデータセット YFCC100M を使用した。YFCC100M^[16] は、世界中で撮影された約 1 億枚の写真及び動画を含んでおりその中の約 5,000 万件は位置情報付き写真である。本論文では、札幌市を対象に、写真の位置情報(経緯度)を用い、15,592 枚の札幌市の写真をフィルタリングした。そして、各写真に対して、Computer Vision API を用いて分析を行い、タグと確信度を求めた。15,592 枚の写真から合計 3,839 個種類のタグ情報を得た。

4.2 POI の抽出結果

3.1.5 で述べた位置情報を利用してクラスタリング結果を表 1 と図 3, 図 4 に示す。距離閾値を 0.5km にし、 $MinPts$ を 50 にすると、札幌市の郊外となる地域を多く抽出できた。一方、距離閾値を 0.1 km にし、 $MinPts$ を 150 にすると、札幌市の都心となる地域を抽出できた。そこで、2つクラスタリング結果の和集合から、合計 40 個のクラスタをスポット対象として得た。

4.3 Semantic-DBSCAN での抽出結果と考察

前節で抽出した 40 個の地点に対して、それぞれに Semantic-

アルゴリズム 1 Semantic-DBSCAN

```

Input:  $Eps_1, Eps_2, MinPts, D = \{o_1, o_2, \dots, o_n\}$ 
Output:  $C = \{C_1, C_2, \dots, C_k\}$ 
Cluster_Label = 0
for i=1 to n do
    if  $o_i$  is not in a cluster then
         $X = RetrieveNeighbors(o_i, D, Eps_1, Eps_2)$ 
        if  $|X| < MinPts$  then
            Mark  $o_i$  as noise
        else
            Cluster_Label = Cluster_Label + 1
            for j=1 to  $|X|$  do
                Mark all objects in  $X$  with current Cluster_Label
            end for
            Push(all objects in  $X$ )
            while not IsEmpty() do
                CurrentObj = Pop()
                 $Y = RetrieveNeighbors(CurrentObj, D, Eps_1, Eps_2)$ 

                if  $|Y| > MinPts$  then
                    for All objects o in  $Y$  do
                        if o is not marked as noise or it is not in a
                        cluster then
                            Mark o with current Cluster_Label
                            Push(o)
                        end if
                    end for
                end if
            end while
        end if
    end if
end for

```

表 1 DBSCAN による札幌市のクラスタリング結果

パラメータ	spot の個数
$Eps = 0.5\text{km}, MinPts = 50$	32
$Eps = 0.1\text{km}, MinPts = 150$	11

DBSCAN を用いて POI の抽出を行った。その中で、三つスポットを例として結果と観察を述べる。

図 5 と図 6 は藻岩山のふもとの近くのエリアの POI の例を示す。表 2 は使用したパラメータと抽出した POI の個数に示す。図 5 は、パラメータを $Eps_1 = 200m, Eps_2 = 75, MinPts = 10$ に設定して得られた POI の例 A である。P4 は神社の写真をものであり、P3 は画像のメインの部分は黒色となる写真を集めたが、公園の噴水の写真も含まれている。P2 は花の写真と空をメインにした写真の両方が含まれている。P1 は冬の雪の写真を集めたものや冬のバスの写真、緑色の電車の写真を集めたもの、または赤色のバスの写真の集めたものが含まれている。そこで、パラメータの Eps_2 を 75 から 20 に変えてクラスタリングを行った結果を図 6 に示す。前例における複数テーマの写真が集めた P1 は、6 つの POI を分けることができ、P11 は冬の雪の写真を集めたもの、冬のバスの写真、緑

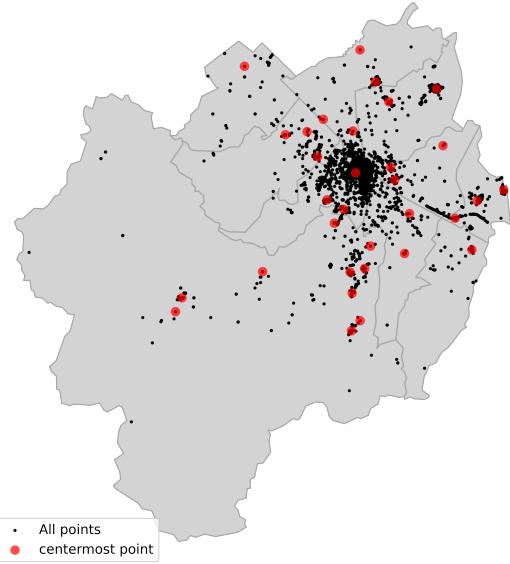


図 3 札幌市における抽出した 32 個クラスタの結果

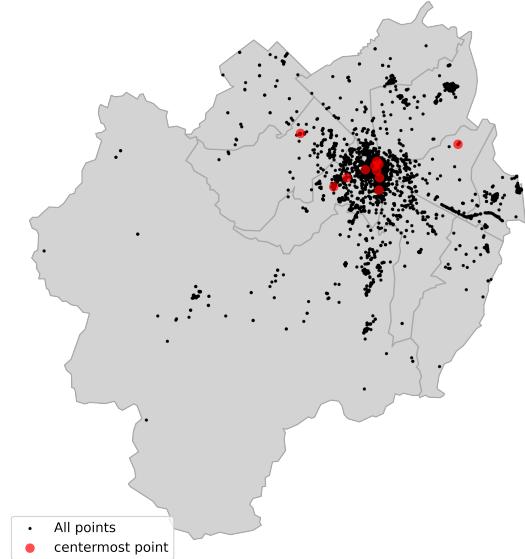


図 4 札幌市における抽出した 11 個クラスタの結果

色の電車の写真であり、P1 は赤色のバスの写真を集めたものである。また、P2 は冬の雪の写真だが、窓が主となるので別の POI として分けられ、駅の写真や電車内部の写真、または電車の細部の写真なども別の POI として分けられた。

例 A と例 B の結果に対する考察を以下に述べる。

(1) Eps_2 は大きく設定すると、潜在空間上における密度が大きいクラスタが合併されてしまう可能性がある（例 A の P3, P2, P1）。一方、 Eps_2 は小さく設定する場合、密度が小さいクラスタが似たようなクラスタに分けられる可能性がある（例 B の P2 と P3, 例 A の P5 と P6）。

(2) 都市の写真や自然風景の写真など幅広いテーマのものが含まれるエリアでは、特徴ベクトルに変換した後の潜在空間上におけるテーマごとに特徴ベクトル間の密度に差が大きいので、

表 2 Semantic-DBSCAN による藻岩山のふもとの近くのエリアの POI の結果

パラメータ	POI の個数
$Eps_1 = 200m, Eps_2 = 75, MinPts = 10$	4
$Eps_1 = 200m, Eps_2 = 20, MinPts = 10$	15

表 3 Semantic-DBSCAN による都心のエリアの POI の結果

パラメータ	POI の個数
$Eps_1 = 200m, Eps_2 = 40, MinPts = 3$	4
$Eps_1 = 200m, Eps_2 = 10, MinPts = 3$	18

同じパラメータを使う場合、上手くクラスタ分けられない場合がある（例 A の P4 は分けられるが、P3, P2 などは分けられない）。

さらに、都心となるエリアの結果の例を示す。図 7 と図 8 は POI の例で、表 3 は使用したパラメータと抽出した POI の個数を示す。図 7 は、パラメータは $Eps_1 = 200m, Eps_2 = 40, MinPts = 3$ に設定して得られた POI の例 C である。P1 は桜の写真を集めたものである。P4 は会議会場の写真と将棋大会の写真を集めたもので、二つイベントの写真が含まれる。P3 は食事の写真をクラスタとして分けられるが、花畠の写真も同じクラスタとして抽出した。そこで、パラメータの Eps_2 を 10 に変えてクラスタリングを行った（図 8）。例 D の P18 は将棋大会の写真以外、ほか文字の要素が主となる写真は同じクラスタとして抽出した。しかし、P19 では会議会場の写真が抽出したが、一部の食事の写真も同じクラスタに合併される。花畠の写真は似たようなクラスタ P7, P8 に分けられる。

図 9 は「札幌芸術の森」のエリアの POI の例である。パラメータは $Eps_1 = 200m, Eps_2 = 20, MinPts = 3$ に設定した。P5, P7, P8 は「野外美術館」のアート作品ごとに分けられたが、どのクラスタでも秋の風景の写真が含まれている。P9 は空をメインにした写真の集めたもので、複数の地点を含まれている。空のような一般的な風景の要素をメインにした写真は同じクラスタに属する可能性が高いことが分かった。

5まとめと今後の展望

本論文では、ピクトリアルマップの自動生成のため、SNS 上の投稿写真の位置情報と意味情報の両方を考慮した Semantic-DBSCAN 手法を提案した。実験結果により、提案手法の有効性を確認できた。今後は適切な意味情報の特徴ベクトルの変換手法を変えて比較実験を行う。

謝辞 本研究は JSPS 科研費 19H04219 の助成を受けたものです。

文 献

- [1] Wentong Zhou, Taketoshi Ushijima, “Automatic Generation of Pictorial Maps from Photos on Social Media

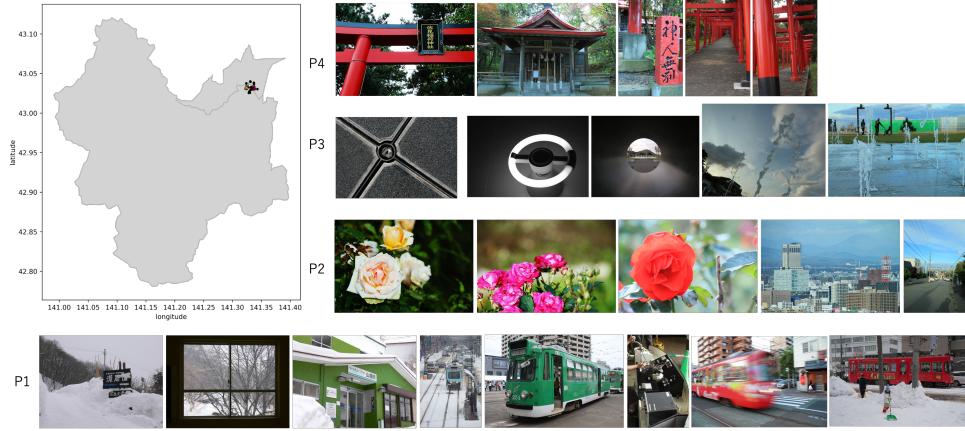


図 5 藻岩山のふもの近くにおける抽出したPOIの例 A

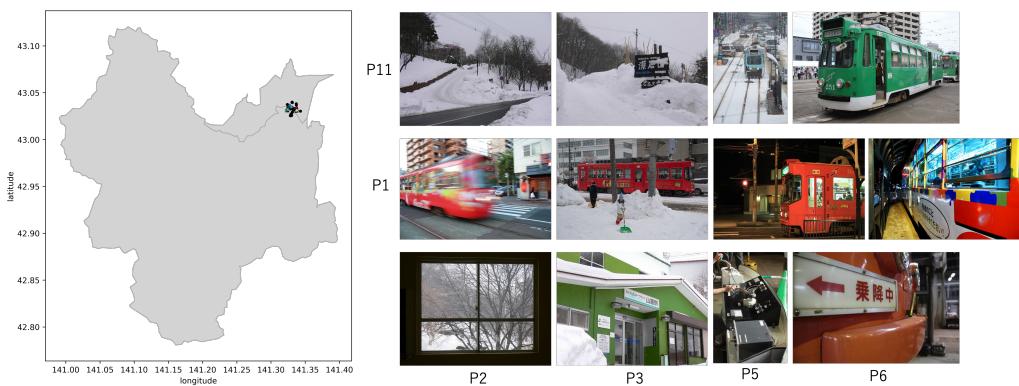


図 6 藻岩山のふもの近くにおける抽出したPOIの例 B



図 7 札幌市の都心における抽出したPOIの例 C

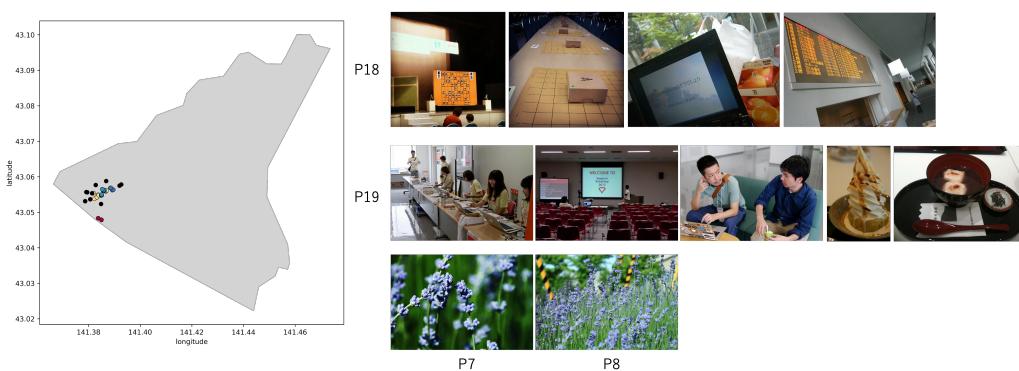


図 8 札幌市の都心における抽出したPOIの例 D

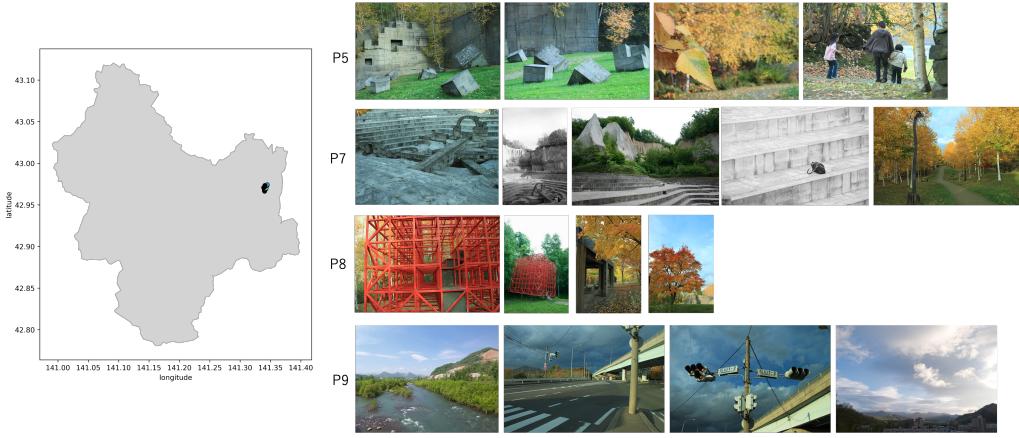


図 9 「札幌芸術の森」における抽出した POI の例 E

- to Represent Regional Features,” 2020 14th International Conference on Ubiquitous Information Management and Communication (IMCOM), pp. 1–5, 2020.
- [2] Wei-Chao Chen, Agathe Battestini, Natasha Gelfand, Vidya Setlur, “Visual summaries of popular landmarks from community photo collections,” 2009 Conference Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers, pp. 1248–1255, 2009.
- [3] Hashimoto Kohjiro, Miyosawa Tadashi, Miyabe Mai, Tsuchiya Takeshi, Ozaki Takeshi, Hirose Hiroo, “Sightseeing Hot Spots Analysis by Using SNS’s Photos and Taking Location Information Based on Image-to-Tag Method,” International Journal of Machine Learning and Computing, Vol. 10, No. 5, pp. 624–629, 2020.
- [4] Martin Ester, Hans-Peter Kriegel, Jorg Sander, Xiaowei Xu, “A density-Based Algorithm for Discovering Clusters,” KDD-96, pp. 226–231, 1996.
- [5] Slava Kisilevich, Florian Mansmann, and Daniel Keim, “P-DBSCAN: a density based clustering algorithm for exploration and analysis of attractive areas using collections of geo-tagged photos,” In Proceedings of the 1st International Conference and Exhibition on Computing for Geospatial Research & Application, No. 38, pp. 1–4, 2010.
- [6] 中原, 石川, 廣田, “ジオタグ付き写真を用いた高度を考慮したホットスポットの分析,” 第 10 回データ工学と情報マネジメントに関するフォーラム, DEIM2018, 2018.
- [7] Olga Yanenko, “Introducing Social Distance to St-DBSCAN,” In Proceedings of the 22nd AGILE Conference 2019, Limassol, Cyprus, pp. 17–20, 2019.
- [8] 瓢本 祥章, 石川 佳治, 杉浦 健人, 脇田 佑希子, “ジオタグ付き写真を用いた意味的な移動軌跡の分析,” 第 15 回情報科学技術フォーラム (FIT 2016), pp. 7C-3, 2016.
- [9] 瓢本 祥章, 杉浦 健人, 石川 佳治, “ジオタグ付き写真の被写体を考慮した意味的な移動軌跡の分析,” 第 9 回データ工学と情報マネジメントに関するフォーラム (DEIM 2017), pp. H7-5, 2017.
- [10] Miriam Redi, Damon Crockett, Lev Manovich, Simon Osindero, “What Makes Photo Cultures Different,” Proceedings of the 24th ACM international conference on Multimedia, pp. 287–291, 2016.
- [11] L. Manovich, M. Stefaner, M. Yazdani, D. Baur, D. Goddemeyer, A. Tifentale, N. Hochman, and J. Chow, Selfiecity: <http://selfiecity.net/#intro>
- [12] Sung Ju Hwang, Fei Sha, Kristen Grauman, “Sharing Features Between Objects and Their Attributes,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR), pp. 1761–1768, 2011.
- [13] Computer Vision API:
<https://azure.microsoft.com/ja-jp/products/cognitive-services/computer-vision/>
- [14] Martin Ester, Hans-Peter Kriegel, Jorg Sander, Xiaowei Xu, “A density-Based Algorithm for Discovering Clusters,” KDD-96, pp. 226–231, 1996.
- [15] Derya Birant, Alp Kut, “ST-DBSCAN: An algorithm for clustering spatial-temporal data,” Data & Knowledge Engineering, Vol. 60, pp. 208–221, 2007.
- [16] B. Thomee, D.A. Shamma, G. Friedland, B. Elizalde, K.Ni, D. Poland, D. Borth, and L. J. Li, “YFCC100M: The new data in multimedia research,” Communications of the ACM, Vol.59, No.2, pp. 64–73, 2016.