SNS 投稿写真の Visual Concept に基づく時空間類似地域マイニング

陳　　ル† 川西　康友† 井手　一郎† 平山　高嗣†† 道満　恵介†††
出口　大輔†††† 村瀬　洋†

† 名古屋大学　大学院情報学研究科　〒 464–8601 愛知県名古屋市千種区不老町
†† 名古屋大学　未来社会創造機構　〒 464–8601 愛知県名古屋市千種区不老町
††† 中京大学　工学部　〒 470–0393 愛知県豊田市貝津町床立 101
†††† 名古屋大学　情報連携統括本部　〒 464–8601 愛知県名古屋市千種区不老町
E-mail: †chenl@murase.is.i.nagoya-u.ac.jp, kawanishi@i.nagoya-u.ac.jp, ide@i.nagoya-u.ac.jp,
murase@i.nagoya-u.ac.jp, ††takatsugu.hirayama@nagoya-u.jp, †††kdoman@sist.chukyo-u.ac.jp,
††††ddeguchi@nagoya-u.jp

あらまし　旅行を計画する際，満足のいく計画を立てるためには，旅先の雰囲気を予め把握することが重要である．ある地域の訪問予定時の雰囲気が，既に知っている地域のある時期の雰囲気と似ていることが分かれば，その地域の訪問時の雰囲気を直感的に把握することできる．これを実現するため，我々は旅行者の体験を反映した大量の SNS 投稿写真に着目した．具体的には，写真に付随する時空間情報を用いて写真をクラスタリングし「時期情報付き地域」を獲得し，写真の Visual Concept に基づいて時空間類似地域をマイニングすることで，時期を考慮した類似地域の組を獲得する手法を提案する．手法の有用性を示すため，Flickr に投稿された写真を用いて時空間類似地域マイニング実験を行った．また，被験者実験を通して，提案手法の妥当性を確認した．

キーワード　時期情報付き地域，時空間類似地域マイニング，SNS 投稿写真，Visual Concept

1　はじめに

休日の過ごし方として，旅行が人気を博している．日本交通公社（JTBF）の調査 [1] によると，2017 年の日本人旅行者数は 3 億 4 千万人に達し，ここ数年間横ばいに推移している．一方で 2017 年の訪日外国人旅行者は 2,869 万人に達し，2012 年以来増加している．

また，旅行中の行動の自由度が高いため，団体旅行より個人旅行を選択する旅行者の数が徐々に増加している．日本政府観光局（JNTO）の発表 [2] によると，2017 年 7〜9 月の個人旅行手配の割合は，2012 の 60.8%より 14.9 ポイント増加し，75.7%に達している．個人旅行では旅行者自身が旅行計画を立てる必要があり，また旅先には初めて訪れる地域を選ぶことが多い．そのため，旅行計画を円滑に進めるためには，事前に十分な情報収集を行う必要がある．

旅先に関する情報収集のために，ガイドブックやパンフレット、ウェブサイトなどで旅行案内が提供されている．特に近年はウェブサイトによる旅行案内が主流となっている．このような旅行案内では，観光地に関する情報が大量に存在するため，旅行者はキーワードや大まかなジャンルをもとに情報を絞り込み，比較しながら旅先を決める必要がある．旅先の種別やキーワードを明示的に指定できる場合，旅先を効率的に絞り込めるのに対し，言語で明示的に表現しにくい雰囲気を持つ旅先を調べたい場合，狙い通りのものを絞り込むことは難しい．

この問題に対し，雰囲気がわかっている既知の地域をもとに，その地域と雰囲気が似ている地域を抽出することができれば，明示的に表現しにくい雰囲気の場所であっても狙い通りの旅先を絞り込むことができると考えられる．これを実現するため，本研究では地域間の雰囲気の類似度について考える．

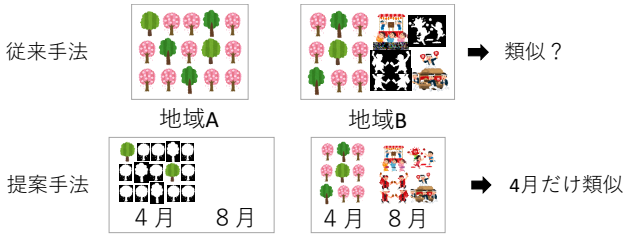


図 1: 従来手法と提案手法．矩角内はそれぞれ地域 A と B の視覚的内容である．従来手法では地域 A と B が類似地域と検出されたのに対し，提案手法では，地域 A と B は 4 月だけ類似するという結果が得られた．

そもそも，地域の雰囲気は，その場所の風景や，そこにいる人たちの行動が自然に作り出すものである．地域によって気候が異なるうえ，同じ地域であっても，季節の変化によって雰囲気が変わることがある．そのため，旅行者が見たり体験したりすること，つまりその場所その時期に旅行者が興味を持つ対象は，旅行者が感じたその地域の雰囲気を反映している．そして，ある時期に多くの旅行者に共通する興味対象は，一般的なその地域の雰囲気を反映することができると考えられる．

近年，旅行者の旅先での興味対象は，インターネット上に投稿された写真から知ることができるようになった．JTB 広報

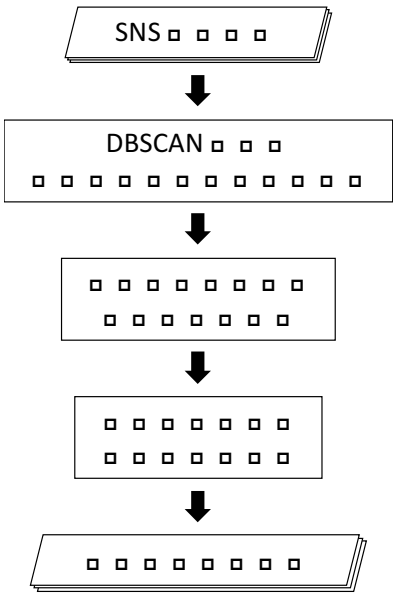


図 3: 時期情報付き地域の範囲決定手順

図 2: 提案手法の処理手順

室 [3] の調査によると，アルバムや旅行記などに保存することを目的として，71%の旅行者は旅先で積極的に写真を撮ることが報告されている．その内容の多くは，旅先の自然景観や歴史的建造物，現地の街並みや食べ物である．これらの写真を見れば，旅行時の旅先の雰囲気をおおむね把握することができると考えられる．また近年，スマートフォンの普及に伴い，ソーシャルネットワーキングサービス（SNS）の利用者数が急増し，写真を撮って共有する人の数も増えている．そのため，SNS から写真を大量に集めることで，撮影された地域で大勢の人の興味対象を知ることができ，その地域の雰囲気を知ることができると考えられる．

類似した地域を抽出する研究として，滝本ら [4] は SNS に投稿された写真を用いた地域間の類似度を算出する手法を提案している．しかし，この手法では写真を撮影した時期の情報が考慮されていないため，図 1 のように，時期によって雰囲気が変わる地域に対して，特定の時期しか雰囲気が似ていないことを表現できない．このことは，短期間しか滞在しない旅行者への支援には不十分である．

そこで我々は，写真に付随する時期情報を考慮し，時期を考慮した類似地域を提案することを目標に研究を進めてきた [5]．本発表では，SNS 投稿写真の時空間情報に基づいて範囲を決定した「時期情報付き地域」をもとに，各時期情報付き地域内で撮影された写真の画像内容に基づいて特徴を記述し，雰囲気が似ている地域をマイニングする手法を提案し，評価する．さらに，マイニング結果を可視化するインタフェースを作成する．

2 時期を考慮した類似地域マイニング

提案手法では，SNS に投稿された写真の時空間情報と画像内容に基づき，雰囲気が似ている地域をマイニングする．提案手法の処理手順を図 2 に示す．まず，SNS 投稿写真の時空間情報に基づくクラスタリングにより時期情報付き地域を決定する．次に，各時期情報付き地域の範囲内で撮られた多数の写真から画像内容を集計し，時期情報付き地域の特徴とする．最後に，特徴間の類似度を計算することにより雰囲気の類似した地域をマイニングする．

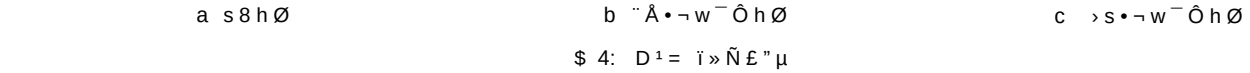
2.1 地域の範囲決定

一般に，地域とは，アジア太平洋地域，火山地域，商業地域等のように，地形が似通っている，同じ特徴をもっているなどの理由からひとまとめにされる土地のことで，階層的に定義されている．その中に，国家が円滑な国家機能を執行するために，領土を行政区画で細分化するものもある．同じ行政区画の中にも，住宅地域，工業地域など，地域の機能によって雰囲気が分かれている．本研究では，旅行者の過去の観光体験をもとに旅行支援を実現するため，大勢の人の興味を引く対象が存在する地理的・時期的に連続した範囲を時期情報付き地域と定義する．

本研究ではクラスタリングによりこの時期情報付き地域の範囲を決定する．その処理手順を図 3 に示す．まず，SNS 投稿写真の時空間密度分布に基づき，DBSCAN [6] を用いて写真を初期分割する．DBSCAN はクラスタ数を設定する必要がなく，一定以上の密度を持つ連続した空間に存在するサンプルを 1 つのクラスタとして検出し，密度が小さい空間に存在するサンプルをノイズとして除去する．具体的には，ある 2 サンプルに対して，互いの最大近傍半径 ε 内に含まれる点数がしきい値 $P_{\min}$ 以上のときに同一クラスタに属するとする．

その後，得られたクラスタのうち，時期方向に長いクラスタに対して，ある時期の特徴を強調するため，しきい値 $P_{\min}$ を δ ずつ増やし，初期クラスタリング結果を小さいクラスタに分割する．すべてのクラスタに属する写真の撮影期間がしきい値 T 以下の場合，この分割型クラスタリングを停止する．

続いて，地理的範囲が重なり，かつ撮影した時期が近いクラスタには同じような注目対象が存在する可能性があると考えられるため，凝集型クラスタリングを使用し，画像内容が類似するクラスタを統合する．画像内容の類似判定については，各クラスタの特徴量に対して正規化相互相関に基づいて類似度を算



$ 4: D¹= ï»Ñ£”µ

Z`|`VM‹ŽÍwÔùx¨Å`oM”qQ ... b”}›Ã”
w{ŠMq¨ÅSw‰ZMOx|f•g• 2.2 ... q2.3 ... p™
\b”}‡h|—:wâ”²w^UdØ`hø T’s”«å
µ»x|—:w t`Tµ¯›Ëh•oMsM•¬pK”qß
Q’•”hŠ|rgw0ÅŽq`o†^b”}

2.2 •¬w›ÃG\

²\`h‘Ot| SNS tdØ^•”ø x|qè`h•:t
SZ”qè w«è0Å›¯`oM”}fwhŠ|‰a•:p
G”t)B`h SNS dØø x|fw•¬›Z•”GNw
wµ¯0Å›Sé`|fwALfw•¬w “>›Séb”q
ßQ’•”}f\p|ŠZ€px¤«åµ»t b”ø wh
þº0›Ý Ý`|hþº0wìSüÍ›;Mofw•¬w “
>›G\b”}

hþº0qx|hþt ‡•”ÄÅw¹®$™¯• w
òQ~ ›¯b›ÃÉw\qpK”}ŠZ€px| Visual
Concept[7] ›;Mo SNS dØø whþº0›¯qb”} Visual Concept x|i••sr K”ú.tv’c|³”ïw
‘Os¿M¹®“æ›¯b\qUpV”}f‘x‡c| Visual
Concept Ý Ý+›;Mo¤ SNS dØø t ‡•”¤ Visual
Concept wìS›{Š|fwüÍ›hþº0›¯b›Ãqb
”}ŠZ€px|¤ø tSZ”ìSwÍ• 10 xw Visual
Concept w^›b;b”hŠ|f•ŽŽwìS› 0qb”}f
`o| 2.1 ... p~’•hÌ8ØCÇV•¬]qt|fwc“º
pqè^•hø T’¨Z`hhþº0›Bÿ`|fwÌ8Ø
CÇV•¬w›Ãqb”}

\\p|•¬› w›Ã›§Ðb”hŠ|~’•hÌ8ØC
ÇV•¬w›Ãt0`o| TF-IDF[8] ›;MoOˆÇZ›b
”} TF-IDF x|{t ‡•”o wOAS›°Ab” O
w1mpK”}ŠZ€px|¤ Visual Concept › 1mwo
q`|¤Ì8ØCÇV•¬› 1mw{q`oßQ”}rwÌ
8ØCÇV•¬p‹‘Xq•” Visual Concept x|‘“°`
$s›Ã›ËmqßQ’•”hŠ|ìS›ÿXb”}°M|›
wÌ8ØCÇV•¬p`Tq•sM Visual Concept x|•
¬w›Q›¯bqßQ’•”hŠ|ìS›ôXb”}\•t‘
“|¤Ì8ØCÇV•¬w›ÃG\U~’•”}

2.3 ¨Å•¬Ú Çï¬

¨Å`hÌ8ØCÇV•¬›Ú Çï¬b” Oq`o|L
Š’ [4] q‰7t|b,owÌ8ØCÇV•¬wÊˆù~dt
mMo¨ÅS›-‰`|¨ÅSU`VM‹ ŽÍwÊ›¬Rb
”MO›q”}\\p|¨ÅSx¤›ÃG\ wYF=ì“ì
›;M”}é.$tx|Ú™w 2•¬ $r_a;r_b$ w¨ÅStm
Mo|2.2 ... p~’•h›ÃÕ«Äç $\hat{V}_a;\hat{V}_b$ t‘“|ÍÜw‘
Ot‰Zb”}

$$S_{r_a;r_b} = \frac{\sum_{n=1}^{N} (\hat{V}_a(n) - \overline{\hat{V}_a})(\hat{V}_b(n) - \overline{\hat{V}_b})}{\sqrt{\sum_{n=1}^{N} (\hat{V}_a(n) - \overline{\hat{V}_a})^2 \sum_{n=1}^{N} (\hat{V}_b(n) - \overline{\hat{V}_b})^2}} \quad (1)$$

\\p| N x›ÃÕ«ÄçwAÉ:pK“| $\hat{V}_a(n);\hat{V}_b(n)$ xÌ
8ØCÇV•¬ $r_a;r_b$ w›ÃÕ«ÄçwAÉpK”}²\`
h‘Ot|¨ÅS $S_{r_a;r_b} \geqq$ qs”Ì8ØCÇV•¬ $r_a;r_b$ w
Ê›¨Å•¬qb”}

2.4 ALwD¹=

Ú Çï¬`hAL›üsb”hŠtD¹= ï»Ñ£”µ
›^R`h}\w ï»Ñ£”µx|¨Å•¬›¯ÔpV”°
M| ›s•¬›¯Ôb”;ó‹ b”}\\p| ›s•¬
qx|„Tb,ow•¬qw¨ÅSU`VM‹Ž<p„Tt¨
Åb”•¬UsM‘Os•¬w\qpK”}

ï»Ñ£”µwhØ«›$ 4tÔb} ï»Ñ£”µw7
Íæpx| Visual Concept w ¨|¯Ô`hMÚ Çï¬AL
q`VM‹›ƒ b”\qUpV”}\\p|¯Ô`hMÚ
Çï¬ALqx¨Å•¬q ›s•¬w 2pK”}¨Å•¬
›¯Ôb”Ôù|â”²U¬R`h 1mw•¬qw¨ÅSUƒ
`h`VM‹ŽÍpK”•¬›¯Ôb”}°M| ›s•¬
›¯Ôb”Ôùtx| wb,ow•¬qw¨ÅSU`VM‹
Ž<w•¬›¯Ôb”}

7s|$ 4atÔb‘Ot|(Íw•$t 2.1 ... p~’•h
Ì8ØCÇV•¬Ub,o¯Ô^•oM”}•$Íw•”Ú”
§wíx|f•g•w«åµ»t ‡•”ø wqèÌ8›¯

4.2 時期分割の妥当性評価

本節では，SNS 投稿写真を時期情報付き地域に分割する妥当性について評価する．時期に分割した地域間の類似判定と，1 年間を通した地域間の類似判定の違いを評価した．

4.2.1 実験条件

比較対象として，1 年間を通した地域の類似度を算出するために，現在の時期情報付き地域を元に，同じ地理的範囲内で撮影された写真をすべて統合した．

地域間の類似度計算方法は時期分割する場合と同様に，写真の画像内容に基づいて地域の特徴を記述した後，正規化相互相関を用いて算出した．また，類似すると判定するしきい値は 0.80 とした．

4.2.2 結果および分析

時期分割をする／しない場合を比較したところ，5,031 組のうち，3,156 組は時期分割により類似しない地域の組から類似地域の組に変わった．その結果の一部を表 2 に示す．

表 2 に示す通り，類似判定結果が変わった地域の組の多数は大都市の組である．大都市は多種多様な機能を持ち，単一の観光体験にとどまることはほぼないため，通年の特徴は現れにくく，類似度が低くなる．一方，時期に分割することにより，各時期の特徴が明確になったため，地域間の類似度が高くなったと考えられる．

図 8 にその一例を示す．この場所 8a における 1 年間を通した写真 8b には，雪や食べ物，祭りなどが混在しているため，特徴を把握することが難しい．一方，時期で分割することにより，夏には祭りがあり 8e，冬には雪が降る 8c ことが把握でき，また春に食べ物に関する写真が多い 8d ため，地域内に飲食店が多く存在することも分かった．このことから，写真を時期に分割することで，特定の時期の特徴が明確になったことが確認できた．

5 むすび

本発表では，SNS に投稿された写真の時空間情報と画像内容を用いて，類似地域をマイニングする手法を検討した．

従来の SNS 投稿写真に付随する位置情報を用いた手法に対し，時期情報を加えてクラスタリングを行い，ある時期のみ類似している地域の組をマイニングする手法を提案した．

実際に，Flickr に投稿された日本国内の写真を用いて，類似地域マイニング実験を行った．提案手法では，花に関するイベントがある地域の組やスポーツの開催地の組など，特定の時期だけ類似する地域を抽出することができた．

また，被験者実験で人間の感覚と比べ，約 88.5%は人間の感覚と一致することで，提案手法の妥当性を確認した．同じ地理的範囲内で撮影された通年の写真と比べることで，時期分割の妥当性を確認した．

今後の課題として，ユーザの行動情報を取り入れること，Visual Concept の潜在意味解析，本研究の具体的活用方法の検討などがあげられる．

謝辞

本研究の一部は，科学研究費補助金による．

文献

[1] 日本交通公社，旅行年報 2018，https://www.jtb.or.jp/wp-content/uploads/2018/10/Annual-Report-all-2018.pdf [2018/12/28 参照]

[2] 日本政府観光局，旅行動態の変化の状況，https://www.kantei.go.jp/jp/singi/kanko_vision/kankotf_dai16/siryou2.pdf [2018/12/28 参照]

[3] JTB 広報室，旅行先で写真を撮りますか？，https://www.jtb.co.jp/myjtb/tabiq/pdf/20110726.pdf [2018/12/28 参照]

[4] H. Takimoto, M. Philippe, Y. Kawanishi, I. Ide, T. Hirayama, K. Doman, D. Deguchi, and H. Murase, Detection of similar geo-regions based on visual concepts in social photos, Advances in Multimedia Information Processing—PCM 2017, 18th Pacific-Rim Conference on Multimedia, Revised Selected Papers Part I, Lecture Notes in Computer Science, Vol.10735, pp.497–504, May 2018

[5] 陳ル，川西康友，井手一郎，平山高嗣，道満恵介，出口大輔，村瀬洋，SNS 投稿写真の時空間クラスタリングに基づく類似地域マイニングの検討，信学技報，MVE2018-34, Oct. 2018.

[6] M. Ester, H.P. Kriegel, J. Sander, and X. Xu, A density-based algorithm for discovering clusters in large spatial databases with noise, Proc. 2nd Int. Conf. on Knowledge Discovery and Data Mining, pp.226–231, Aug. 1996.

[7] T. Deselaers and A. Hanbury, Evaluating systems for multilingual and multimodal information access, Proc. 9th Workshop of the Cross-Language Evaluation Forum, pp.531–538, Sept. 2008.

[8] G. Salton and M.J. McGill, Introduction to modern information retrieval, McGraw-Hill, 1986.

[9] 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, Comm. ACM, Vol.59, No.2, pp.64–73, Feb. 2016.

[10] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva and A. Torralba, Places: A 10 million image database for scene recognition, IEEE Trans. Pattern Anal. Mach. Intell., Vol.40, No.6, pp.1452–1464, July 2017.