映像の意味分析に基づくソーシャル写真サイトからの画像付与方式

西澤 真帆† 王 元元†† 河合 由起子††† 角谷 和俊†

†関西学院大学総合政策学部メディア情報学科 〒669-1337 兵庫県三田市学園 2 丁目 1 番地

††山口大学大学院理工学研究科 〒775-8611 山口県宇部市常盤台 2－16 - 1

†††京都産業大学コンピュータ理工学部 〒603-8555 京都市北区上賀茂本山

E-mail: †{dxx07386, sumiya}@kwansei.ac.jp, ††y.wang@yamaguchi-u.ac.jp,

†††kawai@cc.kyoto-su.ac.jp

あらまし 近年，Instagram などのソーシャル写真サイトの普及が進んでおりそこでは，画像とテキストタグによってコミュニケーションがとられている．しかし，ユーザのコミュニティ外である膨大な利用者の投稿から適切な画像を自力で検索することは困難であり十分に Instagram を活用しているとは言い難い．このことから，Instagram の利用者が増えれば増えるほど情報量は増えるためユーザが困惑する場面が増えるのである．そこで，本研究では映像を対象として映像にあった意味のある画像を Instagram から付与する手法を発見し，さらに，抽出してきた画像を対象の映像と同時に並列して閲覧することができるシステムを提案する．

キーワード ソーシャルネットワーク，画像推薦，テキスト分析，映像分析

1. はじめに

Instagram[1]などの写真共有サイトは年々ユーザ数が増加しており，膨大な写真が日々投稿され共有されている．それらの写真にはテキストとハッシュタグが付与されており，その中には写真の説明として，撮影地の名前やそこで何をしたのかといった感想など，ユーザの観点によって様々なものがつけられている．本研究において，Instagram などの写真共有サイトをソーシャル写真サイトと呼称しているのは，テキストとハッシュタグによってサイト内でユーザが様々な人と写真を通じてコミュニケーションをとることができるからである．それらのハッシュタグを参照すると，撮影画像に対する人物や建物，場所や雰囲気等，多くの情報を得ることが可能である．しかしながら，実際に Instagram からユーザにとって適切な画像を検索してくることは難しい．例えば，兵庫県三田市の情報が欲しいユーザが「三田」と検索した場合，兵庫県三田市の情報だけでなく，東京都三田といった全く異なる地域の画像も検索される．これは，検索キーワードが適切でないためと考えられる．そこで，本研究では，三田の意味的構造を得ることで，適切な検索キーワードを抽出する手法を提案する．この意味的構造を得る手法として，今回，映像コンテンツに着目した．映像は，画像が連続したもので，かつ，それら映像に対する字幕情報（クローズドキャプション）が付与されている．また，映像に対する意味的構造を取得して検索キーワードを生成することで，映像に合った適切な Instagram の画像を取得できることで，付加情報としてテキストだけでは伝わりきらない部分を画像で補うことができる．映像から兵庫県の三田市という意味的構造を得られることで，Instagram の検索の際にその意味的構造を反映することができれば，映像に合った画像のみを抽出してくることが可能なのである．

本研究では Instagram から意味のある画像を映像に付与することを目的とし，さらに抽出してきた画像を映像と並列して閲覧することができ画像によって映像の内容を補足するというシステムを提案する．今回提案するシステムを使用する環境としては，見逃し配信として提供されている番組コンテンツをパソコンやタブレットでユーザが視聴する場面を想定する．また，本研究では，最初の取り組みとして対象を地名と限定し，地名が多く出てくる旅番組を視聴対象とする．旅番組で訪れた地域のより多くの情報を Instagram の画像に付

[1] https://www.instagram.com

与されているハッシュタグを用いて映像を補足するシステムを構築する．

本論文の構成は以下の通りである．2 章ではシステム概要と関連研究について述べる．3 章では映像の意味分析に基づく画像抽出方式について説明する．4 章では本研究のまとめと今後の課題について述べる．

2. システム概要と関連研究

2.1 映像と写真の連動システム

本研究においては，Instagram の中でもハッシュタグに最も情報があると考えている。先程にも述べたが，Instagram の投稿には画像だけでなく，テキストやハッシュタグ，位置情報，投稿時間など様々な情報がある．しかし，ハッシュタグにはユーザによっては位置情報や時間情報，また気象情報など，ハッシュタグだけでほとんどの情報をカバーできるのである．

本研究では，ハッシュタグに注目し，Instagram から意味のある画像を抽出してくることを目的としている．抽出した画像を映像に付与することによって，映像だけでは知ることができない部分をテキストではなく画像という情報で視覚的に直接補足できるのではないかと考えた．

図 1 に提案システムの流れを示す．提案システムの手段としてまず，手法の対象となる映像シーンの分割をする．分割する基準としては映像内にでてくる地名を用いる．次に，分割されたシーンごとの地名を，そのシーンが何を説明しているのかを表すキーワードとする．そして，それらのキーワードの地理的関係性を，Wikipedia を用いてツリー構造で表し，映像の意味分析を行う．また，シーンごとの地名をハッシュタグとして Instagram で画像検索する際の対象とする．そして，抽出してきた画像を出力として画面に地図と共に表示する．また，提案システムは，画像だけではなく，映像の意味分析によって得られるシーンのキーワードに関連性があるものを関連タグとして画面に表示する．

2.2 関連研究

映像を対象として，地図とストリートビューで映像を補足する研究である[1]．この研究では，映像の字幕情報から地名の出現時間を抽出し，その地名の地理的関係を地図とストリートビューを用いて可視化することにより，ユーザに分かりやすく示している．また，Wang ら[2]は，映像の字幕情報から映像の話題語抽出に基づきシーンを検出し，シーンの話題性に基づくシーンの削除と，投稿映像，画像や地図を用いて新しいコンテンツを追加する映像視聴システムを提案している．本論文では映像の字幕情報を抽出し，映像の補足を目的としている点は同じだが，画像を用いて映像の地理情報を補足することによって簡単にその地域のイメージをしやすくすることができる．

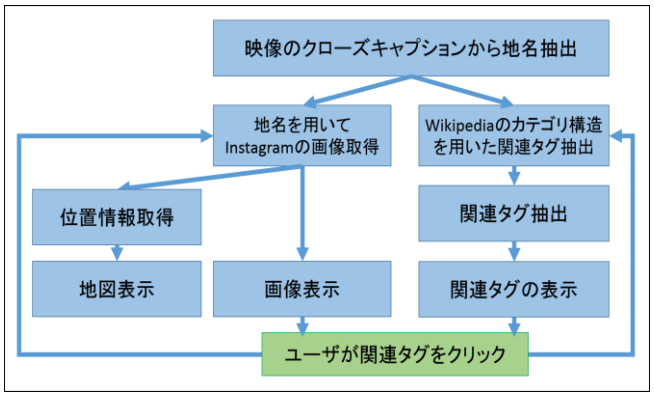

図 1　映像と写真の連動システムの概要図

異種メディアコンテンツの統合に関するものとしては，Ma ら[3]の研究があげられる．WebTelop は映像と Web コンテンツの連動を自動的に行い，情報の補完や統合を行うシステムである．本論文では，このような異種メディアコンテンツを同時に視聴できるようなシステムを提案する．

また，西脇ら[4]の研究は写真共有サイト Flickr の画像に付与されているタグや位置情報から写真をクラスタリングして穴場スポットの抽出を行っている．さらに，遠山ら[5]の研究は写真共有サイト Flickr からテキストタグの周期性を発見し，それに基づいた写真閲覧システムを提案しており，人間が意識できない周期で繰り返すイベントの発見が可能だと記している．これら研究から SNS におけるテキストタグからさまざまな情報が得られることがわかる．また，大崎ら[6]はテキストタグだけではユーザが求める画像を正しく検索できないとし，画像の色，テキスチャ，形状などから類似画像検索するとしている．さらに，松尾ら[7]は画像特徴に基づいたクラスタリング結果が，言語概念上の下位語による画像分類とどれだけ一致しているかという判定方法に言語のツリー構造を用いている．本研究では，画像の特徴を用いるのではなく

図 2　映像と写真の連動システムの概要図

表 1　意味的関係の種類

空間的関係	包含関係
	並列関係
時間的関係	相対的関係
	絶対関係
概念的関係	is-a ~
	part of ~
	歴史的関係
	類似関係

映像の意味を分析し画像集合を絞ることによって，より正確な画像を推薦する．Kim[8]らは，1 つの画像からファセットと抽出する手法を提案しているが，本研究では，画像の意味的関係だけでなく，映像の構造にも着目している．.

3.　映像の意味分析に基づく画像抽出

3.1 映像シーンの分割

本節は，画像を付与する対象である映像の分割方法について述べる．システムでは，映像シーンの切り替えに付与する画像が自動的に変わっていくため，映像シーンの分割を行う．具体的には，Wikipedia を用いて映像に付与されているクローズドキャプションから地名を抽出する．映像の時系列に沿ってある地名からその後に出現する地名までの映像区間を 1 つのシーンとして分割する．例えば，地名 A→地名 B→地名 C の順で地名を抽出した場合，地名 B が字幕に出現するまでの映像区間を地名 A に関するシーン A とし，地名 C が字幕に出現するまでの映像区間を地名 B に関するシーン B として映像を分割する．また，ユーザインタフェースにクローズドキャプションから抽出された地名を中心とした地図を提示することで，ユーザが現在どの地域に関する映像なのかを簡単に理解することができる．図 2 の例では，「伊香保温泉」→「沼田町」→「猿ヶ京温泉」→「谷川岳」の順に 4 つのシーンに分割されている．このように映像から抽出した地名を Instagram からの画像検索の対象タグとして扱う．

3.2 ハッシュタグと映像における意味的関係

ユーザは Instagram に投稿する際，テキストだけでなくハッシュタグ（#）を付けることによって情報を拡散し，より多くの人に共有している．そのハッシュタグは 1 つだけではなく複数も付けるため，それらの間に意味的関係が存在すると考えられる．また，映像においても同じである．1 つの番組において，何も関係性がないシーンが続くという事はなく，各シーンにはなんらかの意味的関係が存在し映像が構成されているのである．本研究では，表 1 のように意味的関係は，空間的関係，時間的関係，概念的関係があると定義する．本論文は空間的関係の包含関係と並列関係だけに着目した．空間的関係とは地理的な領域関係であり，例えば，関西学院大学は神戸市にあり，神戸市は兵庫県の領域中に存在するように上位概念に下位概念が含まれている．図 2 での例では，「群馬県」と「沼田市」は包含関係であり，「伊香保温泉」と「猿ヶ京温泉」の両方は群馬県の領域中に存在するため，伊香保温泉と猿ヶ京温泉は並列関係である．

今回提案するシステムでは，これらの空間的関係を用いるのだが，映像から抽出した地名間の空間的関係を分析し，そこからユーザの知らない意外な関係性を抽出し推薦することが目的である．

3.3 Wikipedia を用いた関連タグ抽出

本研究では Wikipedia を用いて映像の意味を分析する．Wikipedia にはカテゴリページというものがあり，例えば「群馬県」のカテゴリページには 20 件の下位カテゴリと 12 ページの関連ページが含まれている．これを用いて映像の関係性を分析する．旅番組“いい旅・夢気分”を用いて構築し

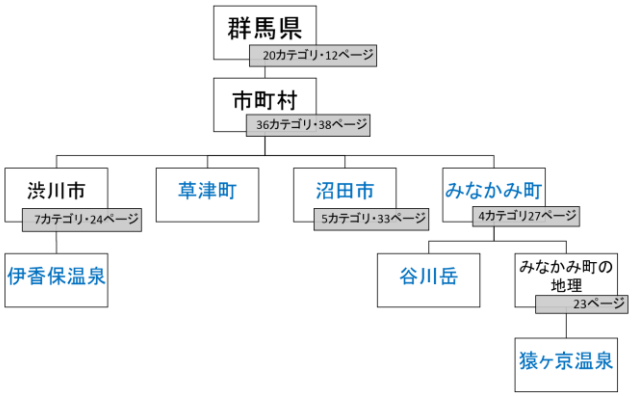

図 3　番組 “いい旅・夢気分” でのツリー構造

たツリー構造を図 3 に示す．青文字は実際にクローズドキャプションから抽出できた地名である．ツリー構造図から分かるように「沼田市」，「草津町」，「みなかみ町」は並列関係にあたり，みなかみ町と谷川岳は包含関係にあたる．このカテゴリページを用いてツリー構造を構築することによって，映像の中の地名間の空間的関係性を判定することができる．

しかしながら，ツリー構造の末端が数多くあることと，1 つの地名に対してさまざまなツリー構造を作成することができるという問題点がある．そこで，本研究ではカテゴリページにおいて 5 ページ以下しか情報が記載されていないものはツリー構造には含まないとした．また，1 つの地名に対してツリー構造が複数できるということに対して．例えば，「伊香保温泉」という地名は図 3 のツリー構造の他に，図 4 のようなツリー構造も作成することができる．これにより，本研究では，対象地名としている 1 つ上の上位概念がもつカテゴリ数が多いものでツリー構造を作成する．「伊香温泉」の場合，図 3 においては「渋川市」，図 4 においては「群馬県の温泉」があてはまる．そして，7 カテゴリ・24 ページを含む「渋川市」と 53 ページを含む「群馬県の温泉」を比較して，より多くのカテゴリを含んでいる「渋川市」でツリー構造を構築していくという手法である．

3.4 Instagram からの画像抽出

提案システムでは映像を入力とし，Instagram から検索してきた画像を出力としている．Instagram から適切な画像を検索するために，本研究ではユーザ情報，位置情報，投稿時間などのデータがある中でハッシュタグと位置情報を用いて検索

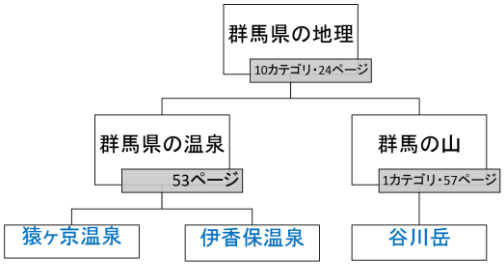

図 4　「伊香保温泉」に関する 別のツリー構造

を行う．

3.1 節で分割したシーンに対する地名を地名タグとして Instagram 全体の画像を検索する．図 2 の例では，最初のシーンに「伊香保温泉」という地名がクローズドキャプションに出現し，これを「＃伊香保温泉」として Instagram を検索する．実際，Instagram で「＃伊香保温泉」を検索した結果は 14,328 件の投稿があった．これらの画像を提案システムのインタフェースに提示する．

3.5 関連タグの抽出方法

まず，関連タグの定義を述べる．関連タグとは 3.1 節で説明した地名タグに関連するタグのことである．つまり，その地名に関係しているが，映像では紹介されていない情報を 3.4 節で作成したツリー構造から分析し Instagram から抽出してユーザに推薦するということである．

本研究では，Wikipedia を用いて作成したツリー構造において，関連タグとして対象にしている情報の並列関係にあたる情報が最も関連性をもっているのではないかと考え，その部分をユーザに推薦したいと思う．関連タグの抽出手法としては，3.4 節で説明した，映像の分析によって構築されたツリー構造を利用し．映像内で紹介されていないカテゴリまたはページの部分を取り出す．例えば，図 2 のように映像が「伊香保温泉」のシーンである場合，関連タグが「＃敷島温泉」，「＃水沢うどん」，「＃小野子山」，「＃渋川へそ祭り」になる． 図 3 を示すように，この例の場合，「伊香保温泉」の 1 個上の上位概念は「渋川市」となる．「渋川市」は 7 カテゴリ・24 ページを下位概念として含んでおり，「伊香保温泉」はそのうちの 1 つのページにすぎない．そこで今回は映像の中で紹介されていない，残りの 23 ページの情報，つまり「伊香保温泉」と並列関係にあたる情報を推薦するということである．（図 5）

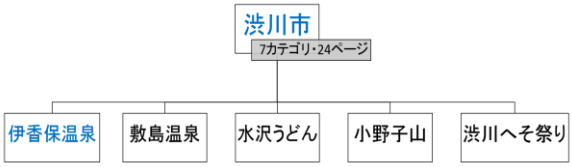

図 5　“伊香保温泉”の関連タグの構造

しかし，23 ページ全てを推薦することは困難なため，ランダムで 5 件を表示する．

さらに，提案システムとして関連タグをクリックしたら，新たな情報が表示されるというように，ユーザにとって受動的なだけでなく能動的に動くシステムである．関連タグをクリックすることによって，集合体を絞ることができ，よりユーザは有益な情報が得ることができる．今回提案する手法は，関連タグをクリックしたときには，主タグと関連タグの両方に関係している画像を表示し，さらに，最初に提示した関連タグに対しても新たに関連タグを推薦するというものである．例えば，「伊香保温泉」の関連タグの 1 つである「＃水沢うどん」をクリックした場合，表示する画像の抽出方法としては，主タグの「＃伊香保温泉」と関連タグの「＃水沢うどん」が，「＃伊香保温泉」「＃水沢うどん」 というような形で 2 つのタグが同時につけられている画像集合と，映像分析によって作成されたツリー構造での「伊香保温泉」の 1 つ上の上位概念「渋川市」の領域内で投稿された「＃水沢うどん」とタグがつけられている画像集合の 2 つを抽出する．次に，「＃水沢うどん」の関連タグの抽出方法として，映像分析において作成したツリーでの上位概念，この例の場合は「渋川市」以外の上位概念を Wikipedia から取り出し，その上位概念を親として新たにツリー構造を作成し，そこで「水沢うどん」と並列関係にあたる情報を推薦する．「水沢うどん」の上位概念としては「渋川市」以外にも，「関東地方の麺料理」と「群馬県の食文化」の 2 つ存在していることが Wikipedia からわかる．この 2 つのうち，より多くのページ数を持っている上位概念を用いてツリー構造を作成する．「関東の麺料理」は 15 ページ，「群馬県の食文化」には 29 ページ含まれているため，この場合は“群馬県の食文化”をツリー構造の親として，「水沢うどん」以外の残り 28 ページの部分を関連タグとして推薦する．また，関連タグをクリックしたとき表示する画像の変動と同時に地図も変動する．

図 6　提案システムのインタフェース

3.6 ユーザインタフェース

提案システムのインタフェースを図 6 に示す．これは「伊香保温泉」のシーンでの実行例である．画面左上には対象となる映像を配置している．そして右下には地図を配置している．本研究では空間的関係に焦点を絞って行っているため，地図を表示することによって映像がどの地域について放送しているかがユーザにとって簡単に理解することができる．赤いピンはシーンの地名を表している．青いピンは Instagram の投稿に付与されている位置情報を表している．シーンが変わると自動的に地図も次の地点へピンが移動するが，前のシーンの記録も記憶させておき，視聴後に見直すことができるようにする．右上には「＃伊香保温泉」で検索した結果，該当する画像を表示する．表示の仕方は，人気順または新着順かどちらかをユーザが選択できるようにする．人気順とは，Instagram の機能の一つである「いいね」の多い順とする．画面左下には 3.5 節で説明した関連タグのうち上位 5 つを表示する．上位 5 つとする判定の仕方は，その関連タグが Wikipedia で持つページ数を基準として判定を行うとする．シーンの移り変わりによって自動的に提示する画像，地図，関連タグも変わるようにする．そして，ユーザが関連タグの中から気になったハッシュタグがある場合，そのハッシュタグをクリックすることによって，提示する画像と関連タグも同時に変更される．また，映像視聴後に画像や関連タグの情報について詳しく見られるように，画像と関連タグに「いいね」ボタンを設置し，視聴後にユーザ自身が「いいね」した情報を閲覧

できるようにする．

4. おわりに

本論文では，映像分析に基づくソーシャル写真サイトからの画像付与方式を提案している．映像に対して，SNS を組み込む研究は現在たくさん行われているが，Instagram と映像の融合という研究は多くは見られなく，従来の研究とは異なるものであるといえる．また，本システムでは映像内容に合った意味のある画像を Instagram から抽出するだけでなく，映像の意味分析も行うことにより関連タグの表示も提案している．これによってユーザにとって広がりのある情報を映像という 1 つのコンテンツから自動的に得ることができ，ユーザが自分で調べるという負担を減らすことができる．

今後の課題として提案システムを実装し，システムの有効性を検証する予定である．また，本研究では映像から地名を抽出しシーンを分割しているが，この手法は完全な地名にしか適用できない．しかし，実際の映像には不完全な地名も多く存在している．今後は不完全な地名もシーンを表すキーワードとして抽出することができる手法について検討を行う必要がある．さらに，空間的関係のみだけでなくさまざまな意味的関係に対応できるような手法も検討する予定である．

謝　辞

本研究の一部は，JSPS科研費26280042の助成を受けたものである．ここに記して謝意を表す．

参　考　文　献

[1] Y. Wang, D. Kitayama, Y. Kawai, and K. Sumiya, “Automatic street view system synchronized with TV program using geographical metadata from closed captions,” in Proc. of the 2014 International Working Conference on Advanced Visual Interfaces (AVI2014), pp. 383-384, 2014.

[2] Y. Wang, Y. Kawai, K. Sumiya, Y. Ishikawa, “An Automatic Video Reinforcing System based on Popularity Rating of Scenes and Level of Detail Controlling,” in Proc. of the 2015 IEEE International Symposium on Multimedia (ISM 2015), pp. 529-534, 2015.

[3] Q. Ma and K. Tanaka, “WebTelop: dynamic TV-content augmentation by using web pages,” in Proc. of IEEE International Conference on Multimedia & Expo (ICME2003), Vol.2, pp.173-176, 2003.

[4] 西脇達也，北山大輔, “写真共有サイトを用いた穴場スポットの抽出,” 第 7 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 Forum 2015), P4-5, 2015.

[5] 遠山由自，廣田雅春，石川博，横山昌平, “ソーシャルメディア上に投影された情報の偏在性及び遍在性の可視化,” 第 7 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 Forum 2015), P4-5, 2015.

[6] 大崎慎一郎，宮田高道，小林亜樹，酒井善則，”Web 画像検索のためのキーワード特徴の抽出と合成によるクエリ画像生成” 映像情報メディア学会誌 Vol.64, No.11, pp.1628~1638, 2010.

[7] 松尾賢治，川野悠，大島裕明，田中克己，”下位語を利用した単語概念が持つ視覚的多様性の数値化”, 画像の認識・理解シンポジウム(MIRU2011)論文集 2011, 401-408, 2011.

[8] Eunggyo Kim, Takehiro Yamamoto, Katsumi Tanaka: Computing Tag-Diversity for Social Image Search, Proc. of the 16th International Conference on Asia-Pacific Digital Libraries (ICADL 2014), Springer, Lecture Notes in Computer Science, Vol.8839, pp. 328-335, 2014.