



ソーシャルネットワークが 変える画像の認識・理解

木村昭悟 (日本電信電話 (株))

ソーシャルネットワークと画像

ソーシャルネットワーキングサービス (SNS) とは、インターネット上の交流を通じて社会的ネットワークを構築するサービスであり、より狭義には、人と人とのつながりを促進するコミュニティ型の会員制サービスのことを指す。この SNS 上で流通するメディアとしてのソーシャルメディア^{☆1}は、SNS 登場以前から普及していたオンラインコミュニケーションやブログの流れを汲むテキストメディアが、長きにわたりほぼそのすべてであり、現在も Twitter・Facebook・LINE などの主要な SNS において中心的な位置付けを占めている。

一方、SNS の普及と時を同じくして、インターネット上のサーバに不特定多数の利用者が投稿した画像・映像・音楽などのマルチメディアコンテンツを不特定多数の利用者で共有する、コンテンツ共有サイトの利用が拡大し、Flickr・YouTube・ニコニコ動画などの有力コンテンツ共有サイトは、コンテンツ流通の重要なインフラとしての地位を築いていった。SNS 普及当初、SNS 上の画像はもっぱらこれらのコンテンツ共有サイトとの連携によって流通するのが一般的であった。しかし、いわゆるスマートフォンの爆発的な普及により、一般ユーザが画像を撮影してアップロードすることの障壁が著しく低下した。すなわち、誰もがどこでも気軽に写真を撮影して、その場で写真をアップロードできるようになった。これにより、日々膨大な量の画像がコンテンツ共有サイトを介さず、直接 SNS に投稿され

るようになり、ソーシャルメディアにおける画像は、テキストと並ぶ重要な位置を占めるようになった。現在では、Twitter・Facebook などの主要 SNS が、Flickr・YouTube・ニコニコ動画などの主要なコンテンツ共有サイトと同等に大量の画像コンテンツを保有するようになり、その量は増加の一途を辿っている。また、Instagram のように SNS 上での画像をよりリッチにすることを意識したサービスや、Pinterest・Pixiv・Snapchat など画像に特化した SNS も普及し、SNS 上での画像の重要性は日に日に増している。

本稿では、画像流通で大きな役割を担うようになった SNS 上の画像コンテンツが、画像の認識や理解にとって非常に有用であることを示すとともに、この有用な資源をどのように活用すれば良いか、その考え方と方法について、いくつかの事例に基づいて概説する。

SNS 上の画像は何が違うか

前章に示したように、SNS は画像流通の中心的な位置を占めつつある。SNS 上の画像コンテンツは、他の画像資源と比較して、単に量が膨大になるだけではなく、画像の認識や理解のために有用となるさまざまな性質を持ち合わせている。本章では、SNS 上の画像コンテンツが画像の認識・理解に有用となる性質について論じる。

■ さまざまな付加情報とともにアップされる画像

デジタルカメラの普及は、従来のフィルムカメラでは取得できなかった重要な情報を、写真とともに取得し記録することを可能にした。それは、撮影条

^{☆1} 広義には、個人による情報発信やコミュニケーション、人の結びつきを利用した情報流通など、社会的な要素を含んだメディアを指すが、本稿ではより狭義の意味を採用する。

4 ソーシャルネットワークが変える画像の認識・理解

件に関するデータである。JPEG等の主要な画像保存形式には、EXIFと呼ばれるメタデータ保存のための規格が制定されており、撮影日時・撮影機器・撮影条件・GPS情報等をその中に記録することができる。現在流通するほとんどのデジタルカメラやスマートフォンでは、撮影時にこのEXIF情報を自動的に生成し、画像とともに記録するのが一般的である。Flickrなどの画像共有サイトでは、このEXIF情報が含まれる画像が大量にアップロードされていることを確認できる^{☆2}。

また、写真を撮影する主要なデバイスがデジタルカメラからスマートフォンに移行することにより、スマートフォンに搭載された各種センサの情報が同時に付与されるようになった。その代表的な情報が位置情報である。Wi-Fi電波強度から現在位置を正確に測定できるPlaceEngineにより、特にGPS衛星からの電波が届きにくい都市部や地下での位置情報の精度が大幅に向上した。Foursquareなどの位置情報専門SNSやTwitter・Facebookなどの主要SNSでは、この位置情報をキーとして商業施設や主要観光地などのスポットを検索することが可能となり、検索したスポット名や関連Webページを自動的に付与して、SNS上のコンテンツをよりリッチにすることができる。位置画像付き写真を専門に扱う写真共有サイトPanoramioでは、写真に付与された位置情報に基づき、写真を地図上に投影することで、地図を閲覧しながらそれぞれの場所での風景を同時に垣間見ることができる。

■ 画像・付加情報・ユーザ相互の関係性

SNSへ画像を投稿し、その画像がSNSで消費されるというプロセスにより、画像・ユーザ・付加情報の相互の関連性が付与される。たとえば、誰がこの画像を投稿したのか、どの画像とともに投稿されたか、どんなテキストとともに投稿されたか、誰がこの画像を好んだか、SNS上でこの画像がどのよ

うに拡散されたか、その過程でどのようなコメントが付与されたか。このように、SNSでのコンテンツの生成・流通・消費のプロセスにより、SNSで流通したコンテンツしか持ち得ない特別かつ重要な補助情報が付加されることになる。

■ 画像はユーザがアップロード

ソーシャルメディア登場以前のメディアであるマスメディアとの対比として最も重要な点の1つに、メディアの閲覧者が同時に発信者としての資格を持ち、自身の責任で自由に情報を発信できる仕組みがある。また、写真を撮影するデバイスがデジタルカメラからスマートフォンに変化することで、誰もが写真を気軽に撮影し、その写真をすぐにアップロードできるようになった。すなわち、SNS上の画像は、SNSの仕組みやデバイスの発展によって急激に増加したユーザがアップロードした写真によって、その多くが構成されている。このことは、SNS上の画像が画像資源として以下の性質を持つことを示唆する。

- 量が非常に膨大であるとともに、多様性も非常に大きい。
 - 同じ時刻や同じ場所での出来事が、さまざまな視点や立場から撮影される可能性がある。
- 一方で、ユーザがアップロードした写真は必ずしも望ましい性質ばかりを持つわけではなく、たとえば以下のような不都合な性質もある。
- 撮影者のスキルや撮影機器の水準は必ずしも高いとは限らず、画像の質が必ずしも担保されなくなる。
 - 画像の付加情報の信頼性も必ずしも高いとは限らず、統制されたデータセットのラベル情報と同様に扱うことはできない。

SNS上の画像を認識・理解する

前章では、不特定多数のユーザがそれぞれ思い思いに画像をSNSにアップロードし、その画像にさまざまな付加情報が与えられる点について述べた。本章では、SNS上の画像が持つこのような性質を、

^{☆2} SNS上の画像はEXIF情報が含まれていないように見えるが、これはサイト側がプライバシーに配慮して削除しているためであり、アップロードの際にはEXIF情報が含まれている場合の方が一般的である。



図-1 位置情報が類似する画像群に付与されたテキストタグを集約すると、地域やランドマークの名称が主要タグとして浮かび上がる (Ahern, et al., 2007) ^{☆3}

画像の認識・理解にどのように活かしていくか、いくつかの事例に基づいて概説する。本稿では、特に付加情報と関係性に着目した興味深いアプローチについて、重点的に紹介する。

■「さまざまな付加情報」を用いた画像認識

付加情報として最も一般的に用いられるのは、位置情報である。位置情報が画像の認識・理解に有用であることを示す初期の試みの1つとして、2007年に Ahern らによって提案された World Explorer が知られている^{☆3}。この研究の基本的なアイデアはシンプルで、位置情報が類似する画像群に付与されたテキストタグを集約することで、特定の地域やランドマークの名称、その地域で人々の興味関心を集める対象などが、主要なタグとして浮かび上がる、というものである(図-1)。このプロセスにおいて、画像は位置情報とテキストタグをつなぐための単なる媒介に過ぎないが、さらに画像特徴量を利用することで、位置を指し示すテキストタグと、そのタグを端的に表現する代表的な画像を抽出することができる^{☆4}。

このように、画像における位置情報の重要性は広

く認知されており、競争型ワークショップ MediaEval^{☆5}では、その初年度である2010年から現在まで、与えられた画像が撮影された場所を推定する Placing タスクが設定されている。このタスクにおいては、他の付加情報を利用することで推定精度を向上させることも可能である。代表的な例の1つとして、Kalogerakis らは、画像に付与された時刻情報を利用

することで撮影場所の推定精度を向上させる方法を提案している^{☆6}。この手法では、時刻情報が付与された画像の系列を対象とした各画像の撮影場所の同時推定を問題として扱う。その画像の系列が1人の撮影者によって撮影されたことを仮定すると、ある限られた時間で移動できる距離には物理的な制約があるため、2枚の画像に付与された時刻情報の差分と位置情報の差分との間には非常に強い依存関係がある。この知見を利用し、時刻情報の差によって決定される事前確率を組み込んだ系列ラベリング問題を解くことで、撮影個所の推定精度を向上させる。類似するアイデアに基づく別のアプローチとして、Twitterのタイムラインを利用した Hauff らによる研究も興味深い^{☆7}。Twitterに投稿された画像の撮影場所を推定する問

^{☆3} Ahern, et al. : World Explorer : Visualizing Aggregate Data from Unstructured Text in Geo-referenced Collections, JCDL2007.

^{☆4} Kennedy, et al. : Generating Diverse and Representative Image Search Results for Landmarks, WWW2008.

^{☆5} <http://www.multimediaeval.org>

^{☆6} Kalogerakis, et al. : Image Sequence Geolocation with Human Travel Priors, ICCV2009.

^{☆7} Hauff and Houben : Geo-location Estimation of Flickr Images : Social Web based Enrichment, ECIR2012.

4 ソーシャルネットワークが変える画像の認識・理解

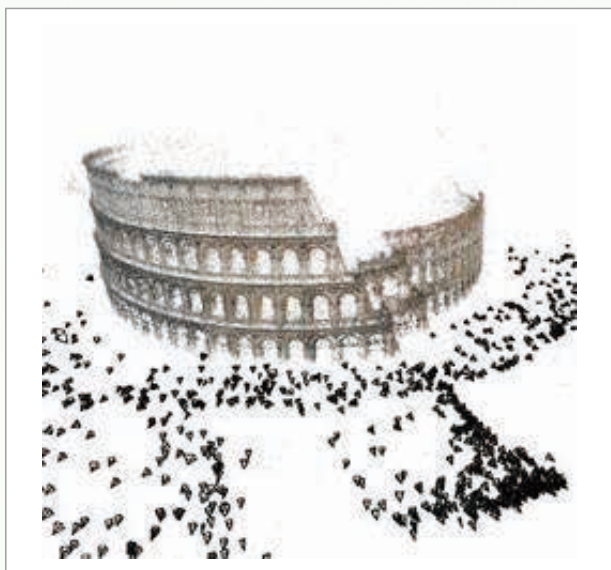


図-2 EXIF 情報から撮影条件に関する情報を活用することで、写真を撮影したカメラの位置の推定精度を向上できる。これは、Flickr 上の大量の画像からの街並みの 3 次元再構成を可能にする重要な要素技術である (Agarwal, et al. : Building Rome in a Day, Communication of the ACM (2011))

題において、画像を投稿したユーザがタイムライン上で自分の居場所を推定可能な情報を投稿していれば、投稿された画像がその場所の周辺で撮影されたであろうことを推測可能である。認識・理解に直接かかわる研究ではないが、EXIF 情報を利用した興味深い研究として、不特定多数のユーザによってアップロードされた Flickr 上の大量の画像から街並みの 3 次元モデルを生成する "Building Rome in a day" が有名である (図-2) ^{☆8}。3 次元再構成の基本技術である Structure from Motion (SfM) は、通常厳密に校正されたカメラ群を利用して撮影した複数の画像を合成することが一般的であったが、この研究では、機器も照明条件もバラバラではあるが、非常に大量に利用可能な Flickr 画像を利用した 3 次元モデル再構成を試みた。このとき、Flickr 画像に記録されている EXIF 情報から撮影条件に関する情報、特に焦点距離値や素子情報を活用することで、写真が撮影されたカメラの位置とそのパラメータの高精度な推定を実現している。

■「関係性」を用いた画像認識

先に述べたように、SNS への画像投稿と消費と

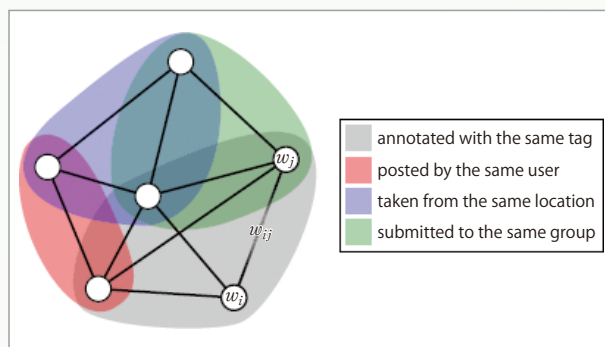


図-3 Flickr メタデータの共通度を画像間の関係を記述する特徴量として採用することで、画像分類の性能が向上する (McAuley, et al. 2012) ^{☆9}

いうプロセスにより、画像・ユーザ・付加情報の相互の関係性が付与される。この関連性も、付加情報同様にうまく活用すると、画像の認識や理解に大きく貢献することが知られている。

その代表的な例として、メタデータの共通性に着目した関係性を画像分類に活用する McAuley による研究がある ^{☆9}。画像認識用のベンチマークデータセットに含まれる画像の多くは、実は Flickr からの転用である。この研究では、ベンチマーク内の画像を Flickr で探し出し、Flickr 上で獲得可能な関係性に関する情報がベンチマークの性能にどの程度貢献するかについての検討を行った。具体的には、EXIF 情報・テキストタグ・説明文・画像閲覧数・ユーザ情報・グループ情報など、Flickr で取得可能なさまざまなメタデータを各画像ごとに回収し、メタデータの共通度を画像間の関係を記述する特徴量とした。このとき、画像をノード、関係性を記述する特徴量をエッジにそれぞれ対応付けたグラフ構造 (図-3) を考え、各画像にラベルを付与するかどうかの 2 値分類問題を、このグラフ構造のグラフカットの問題として定式化した。この定式化により、共通のメタデータを持つ画像の対が類似するラベルを持つという自然な仮定を導入した画像のラベリングを可能とし、本研究で対象としたいいずれのベンチマ

^{☆8} Agarwal, et al. : Building Rome in a Day, ICCV2009, <http://grail.cs.washington.edu/rome/>

^{☆9} McAuley, et al. : Image Labeling on a Network : Using Social-network Metadata for Image Classification, ECCV2012.

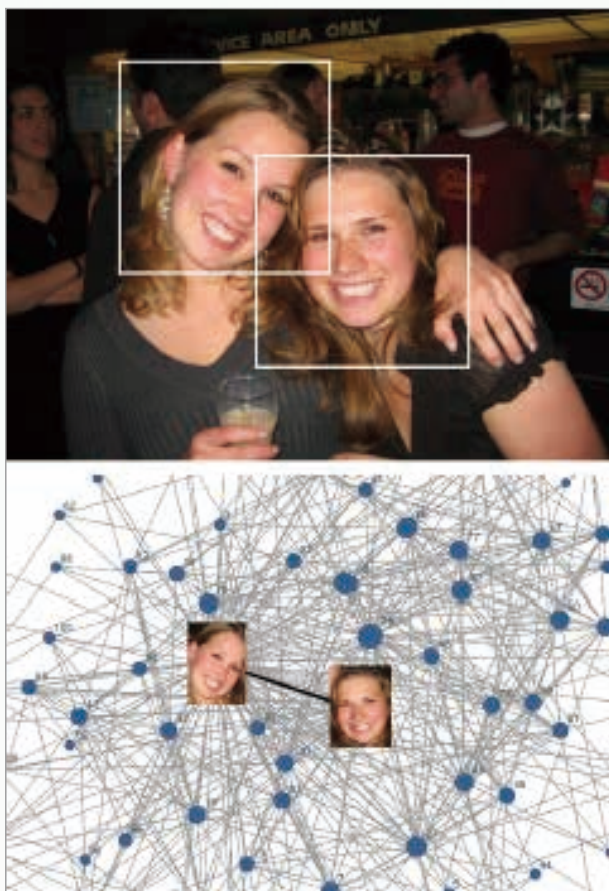


図-4 1つの写真に同時に写る2人は何らかの交友関係にあると考えるのが自然。SNS上の交友関係を利用すれば、顔認識の性能を大きく向上させることができる (Stone, et al. 2008) ^{☆10}

ークにおいても有意に良い性能を得た。

SNS上の画像に含まれる顔の認識にSNSの交友関係を利用する、Stoneらによる2008年の論文も、関係性を効果的に利用した代表的な例の1つであろう^{☆10}。比較的少数の顔が含まれる画像を見たとき、その人々が何らかの交友関係にあると考えるのが自然だろう。この研究では、SNS上の交友関係を利用することで、1つの写真に同時に写る可能性の高い人々を絞り込み、それにより顔認識の精度を向上できることを示した(図-4)。具体的には、画像の中から顔を検出し、検出されたすべての顔に同時にラベルを付与する同時ラベリングとして、顔認識の問題を定式化する。このとき、それぞれの顔が誰であるのがもっともらしいかを示す尤度に加え、2つの顔が誰と誰であるのがもっともらしいかをSNS上の友人関係の有無で決定することにより、個別に顔の認識を行うよりも有意に良い精度で認識

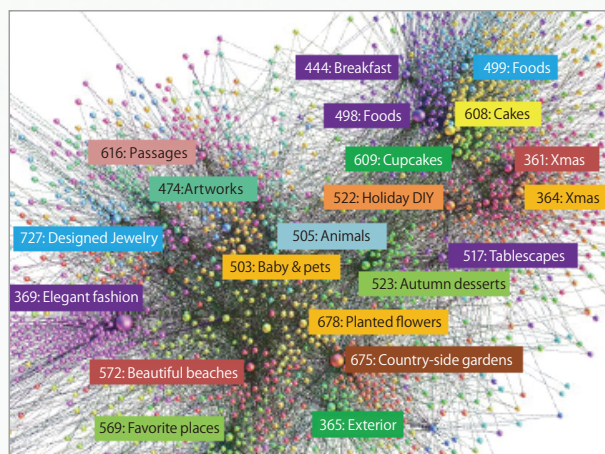


図-5 数多くの画像を共有する画像コレクションの対は類似したテーマを持つことが期待されることを考慮すると、SNS上での画像コンテンツの拡散の様子も、画像の認識・理解に有用な情報となり得る (Kimura, et al. 2013) ^{☆11}

を行うことができる。

さらには、SNS上での画像コンテンツの拡散の様子も、画像の認識・理解に有用な情報となり得る^{☆11}。この研究では、画像に特化したSNSの1つであるPinterestを対象としている。Pinterestのユーザは、*board*と呼ばれる画像コレクションをいくつか所有し、Webにある好みの画像を自分が所有するいずれかのboardに画像を*pin* (画像へのリンクを作成)する。また、他のユーザが持つboardを閲覧して、好みの画像を*repin* (画像へのリンクをコピー)することもできる。多くのユーザは、自分自身や他のユーザが好みの画像を見つけやすいように、各boardに特定のテーマを紐付けて分類する。各boardが特定のテーマを持ち、かつ画像が*repin*を介して複数のboardを伝播することを考慮すると、伝播を通じて数多くの画像を共有したboardの対は類似したテーマを持つことが期待される。この知見を利用すると、boardをノード、共有する画像の数をエッジとして持つグラフ(図-5参照)を構築し、そのグラフで相互連結の強いboardの集合を見つけることで、類似したテーマを持つ画像コレクシ

^{☆10} Stone, et al. : Autotagging Facebook : Social Network Context Improves Photo Annotation, IJCV2008.

^{☆11} Kimura, et al. : Image Context Discovery from Socially Curated Contents, ACM MM 2013.

4 ソーシャルネットワークが変える画像の認識・理解

ョンを検出することが可能となる。この例においても、画像コレクションの類似性を獲得するプロセスにおいて、画像は単なる媒介に過ぎないが、さらに画像特徴量を利用することにより、この類似性を表現する画像特徴量の傾向を学習することができ、これにより画像分類の性能が向上することが知られている^{☆12}。

特有の性質 + 洞察 = 価値ある情報

ここまでで、SNS上の画像が持つ性質を活かして画像を認識・理解するいくつかの例について挙げてきた。

ここまでの事例を見直してみると、その多くが、単にSNS上の画像が持つ性質をそのまま利用するだけではなく、その裏にある物理的な制約や人間の行動パターンに関する洞察が加わることにより、SNS上の画像が持つ性質が、画像の認識や理解に大きな貢献をもたらす価値ある情報に変化することが見てとれる。Kalogerakisらの撮影スポット推定に関する論文や、KimuraらのSNS上の画像伝播を用いた論文などの例からもうかがい知れるように、画像が生成・共有・消費されるプロセスに人

間の行動が深く関与している場合には、特にこの洞察が非常に大きな鍵を握ることになる。自分であれば、SNS上で画像をアップするときはどうするか、どんな画像であれば共有したいと思うか、SNSが変わると自分の行動も変わるのか、まずは自分をテストケースとして考えると、洞察のきっかけがつかめるかもしれない。さらに、さまざまな付加情報や関係性、もしくはSNSの特性を深く理解して活用することができれば、これまでに挙げた事例をはるかに凌駕する興味深い取り組みが数多く産まれる可能性が十分にあるだろう。

また、SNS上で流通するコンテンツは画像だけに限るものではなく、映像や音楽なども同様に流通しており、その量も爆発的に増加している。画像で検討されてきたさまざまなアプローチのうちいくつかは、映像や音楽を認識し理解するためにも利用可能であろうと考えられる。本稿がそのようなより発展的な研究のアイデアの源泉となれば、幸甚である。

(2015年3月31日受付)

木村昭悟 akisato@iecee.org

NTTコミュニケーション科学基礎研究所メディア情報研究部メディア認識研究グループ主任研究員。2000年東京工業大学大学院理工学研究科修士課程修了、同年、日本電信電話(株)入社。博士(工学)。入社以降、パターン認識、メディア理解、データマイニングに関する研究開発に従事。

^{☆12} Alvarez, M. et al. : Exploiting Socially-generated Side Information in Dimensionality Reduction, IWSAM2013.