膨大なメディアデータの探索と活用～ビッグメディア時代のボトルネック解消に向けて～

柏野 邦夫（かしの くにお）
NTTコミュニケーション科学基礎研究所 特別研究員

動画や写真や音楽など，膨大な量のメディア情報を活用するためには，データを記述するデータ，つまりメタデータが必要です．しかし，膨大なメディアデータに対して，その内容を表すメタデータを人手で付与していくのは困難であり，メディア活用における新たなボトルネックになりつつあります．本稿では，メタデータ・ボトルネックの解消に向けて期待されるメディア認識・探索技術の最近の進歩について紹介します．

ビッグメディアとメディア認識

動画や写真や音楽など，膨大な量のメディア情報が蓄積され，流通する時代になりました．例えば，主要動画投稿サイトへの投稿量は１日当り数10万時間分，Web上で共有されている写真は累計で数千億枚にも及ぶといわれています．まさにビッグメディアの時代です．

このような膨大なメディア情報を活用するためには，データを記述するデータ，つまりメタデータが必要です．タイトル，投稿者，再生回数などは代表的なメタデータですが，内容の記述としては必ずしも十分ではありません．しかし，膨大なメディアデータに対して，内容に立ち入ったメタデータを人手で付与していくのは大変困難です．このように，従来のような記憶媒体や通信容量の制約に代わって，メタデータの付与の困難さが，メディア活用における新たなボトルネックになりつつあります．

このようなメタデータ・ボトルネックの解消に向けて期待されるのがメディア認識技術です（**図１**）．メディア認識技術は，音や画像や動画などのメディア情報を対象として，その中に何が存在するか，どのような出来事が起きているかといった，その内容を表す情報を自動的に抽出する技術です．人間の視覚や聴覚と同じような機能をコンピュータで実現することは今なお困難ですが，特定の目的や状況において，メディア認識技術が活用される場面は近年急速に増えています．例えばデジタルカメラの顔検出，自動販売機の人物属性の推定，自動車の運転支援，スマートフォンの文字認識などは，身近なメディア認識の例です．このように認識技術が幅広く使われるようになった要因の１つは，大量のメディア情報自体が認識技術の高度化に活用されるようになったことにあると考えられます．

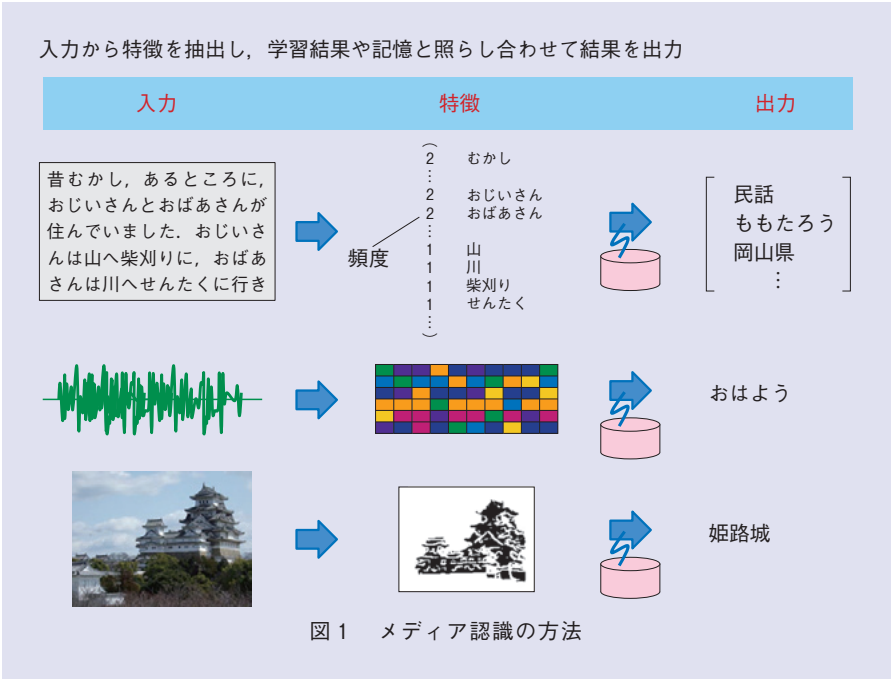

図１　メディア認識の方法

データの集積の活用

メディア認識の中心をなすパターン認識の技術は，入力パターン（メディア）から特徴データ（認識の基になる数値の組）を抽出する段階と，抽出した特徴データから対象を認識する段階の2段階からなります．特徴抽出では，異なる対象をよく区別できるような，認識に有効な特徴を抽出することが重要です．特徴抽出は，従来は経験的な要素が多く含まれる処理でしたが，近年では大量のメディアデータを解析して特徴を設計することが行われるようになりました．

また，認識段階にはいくつかの考え方がありますが，もっとも基本的なのは，学習用のデータに基づいて事前に特徴空間の中に境界を引いて，特徴空間をクラス（認識結果とする単位）ごとに分割しておき，入力された特徴データの属するクラスを識別する方法です．

もう1つの基本的な方法は，記憶と探索（メディア探索）に基づく方法です．これは数多くの事例を記憶・蓄積しておき，入力特徴と各事例の特徴を比較して，特徴空間での探索により該当する事例を見出して，そのラベルを認識結果とする方法です．

これらのいずれの方法でも，数多くの部分的な識別や照合の結果を集積することで，認識を高精度化できることが知られています．例えば，顔検出の代表的手法では，ある部分的な領域の中の画像が顔かどうかを単純な計算だけで識別する識別器が用いられますが，大量の顔データで学習された識別器を数多く用いることで，高い精度が実現されています．また，探索に基づく認識方法では，しばしば簡単化された局所特徴が用いられますが，多くの局所特徴の照合結果を集積することで，非常に高い精度で対象を認識できることが見出されています．

実データの活用

パターン認識の研究分野では，従来から実際の認識誤りなどを解析して精度向上に役立てることが行われてきましたが，近年，特に大量の実使用場面のデータを収集・分析し認識技術に反映させることで，認識技術の精度や性能が大幅に向上する例が知られるようになりました．音声認識や文字認識はその代表例といえるでしょう．

私たちの研究グループでも，継続的に実データの活用による技術向上を図っています．私たちは，約20年前から主に記憶と探索に基づくメディア認識技術の将来性に着目して研究を進めてきました．その研究成果の1つである音・映像の高速探索技術「ロバストメディア探索技術（RMS: Robust Media Search）」は，ある音や映像の断片（区間）に，既知の音や映像が含まれているかを高速に探索する技術です[(1)]．RMSでは，事前に検出したいコンテンツの特徴データベースをつくっておき，入力の特徴と比較して該当個所を検出します．特徴の比較により該当個所を検出すること自体は，信号の変化が少なく，かつ時間をかけても良い場合には，順次特徴データを照合することなどによっても実現できますが，信号の変化に耐える頑健性と，膨大なデータベースを細かくチェックしながら，どの部分が合致するかを素早く見出す高速性との両立には，特別な技術的工夫が必要となります．

RMSの研究においては，2008年に8カ月間にわたり，ネット上のコンテンツのクローリングと調査を専門に行う米国の企業と協力して，ネット上の実際のメディアコンテンツを特定する実証実験を行いました．比較的小規模な試行から始め，最終的には1日当り約100万分（4分間の動画に換算して25万ファイル，つまり当時世界中で1日に新規投稿される動画の全量に匹敵する規模の動画を1日で処理する処理容量）にまで，対象を拡大して実験を行いました．実験で得られる知見を同時進行で技術検討に反映しながら実験を進めたことで，この8カ月間で，実際的な状況における探索の精度を大きく向上させながら，同時に探索の速度を同一リソース当りおよそ35倍にまで高めることができました．

RMSは，現在いくつかのNTTグループ会社で実用に供されています．例えばNTTデータでは，上記の実証実験の形態に近い「ネットモニタリング」のほか，放送番組やネット配信コンテンツに使用されている音楽を自動的に全曲リスト化する「楽曲使用リスト作成」，スマートフォンで音や映像をとらえることでTV番組とネットコンテンツとの連動を可能にする「セカンドスクリーン」など，RMSを核とするさまざまなサービスを国内外の

40社以上に提供しています．

異種データ複合分析の活用

異種のデータを関連付けて解析することは通常のビッグデータ解析でもよく行われていますが，メディア認識の分野でも，メディアの特徴に着目するばかりではなく，それとは異質の外形的情報を用いて推定することが試みられるようになっています．

私たちもこの課題に取り組んでいます．前述のように，メタデータのボトルネックを解消するためにメディア認識技術を用いるとしても，そのメディア認識技術は，正しいクラスのラベルがつけられた大量の学習データを必要とします．その大量の学習データを準備することに大量の人手がかかったのでは，結局実現が困難なものになってしまうため，外形的情報の利用により自動的に学習データを準備することが重要と考えるからです．

私たちが着目した外形的情報は，そのデータの扱われ方，例えば検索のされ方やリンクのされ方などの情報です[(2),(3)]．写真共有サイトにおいて，ある写真とある写真がリンクされていれば，それらには関連する内容が写っている可能性があります．このような情報を集めることで，画像特徴を使わずに，実際に画像の内容や，画像内容の記述に有用な画像特徴を推定できることが確かめられつつあります．

ところで，テキスト処理では辞書が重要な役割を果たしていますが，メディア処理でも，同一性の解析技術や高速・頑健なメディア探索技術が発展すると，メディア認識に使えるメディア辞書の構築や活用が可能になるかもしれません．そのためのステップになると考えられるのがインスタンス探索の研究です．

インスタンス探索は，画像を入力して動画データベースを探索し，「同一の具体物」を検出することを指します．同一の具体物とは，同一人物，同一のロゴマーク，同一の建造物などです．例えば，**図2**のように，エッフェル塔の写真を元に，エッフェル塔の映っている動画の場面を探す，といったことです．これは従来の類似画像探索の機能とは異なります．類似画像探索では類似した画像を出力すれば良いのですが，インスタンス探索の場合には，エッフェル塔であれば昼の場面も夜の場面も検出すべき対象となる一方，たとえ見た目が類似していても東京タワーを検出すると不正解となります．このように，メディアどうしで同一性を持つ部分を検出できるようにすることは，より高度なメディア認識を実現するための基礎となる重要な機能と考えられます．

私たちも音や映像のインスタンス探索の研究に取り組んでいます．2013年に提案した手法（**図3**）は，それまでのRMSと同様に局所特徴どうしの適合度を評価して探索結果を導くものですが，用いる局所特徴の情報量が多い点や，適合度を適合確率として評価する点などが異なっています．この方法は，2013年の国際競争型ワークショップTRECVIDにおいて参加チーム中トップレベルの精度を達成するなど，有望な結果を得ています[(4)]（**図4**）．

便利で豊かな社会の実現に向けて

本稿では，大量のメディアデータの活用によってメディア認識技術が進化

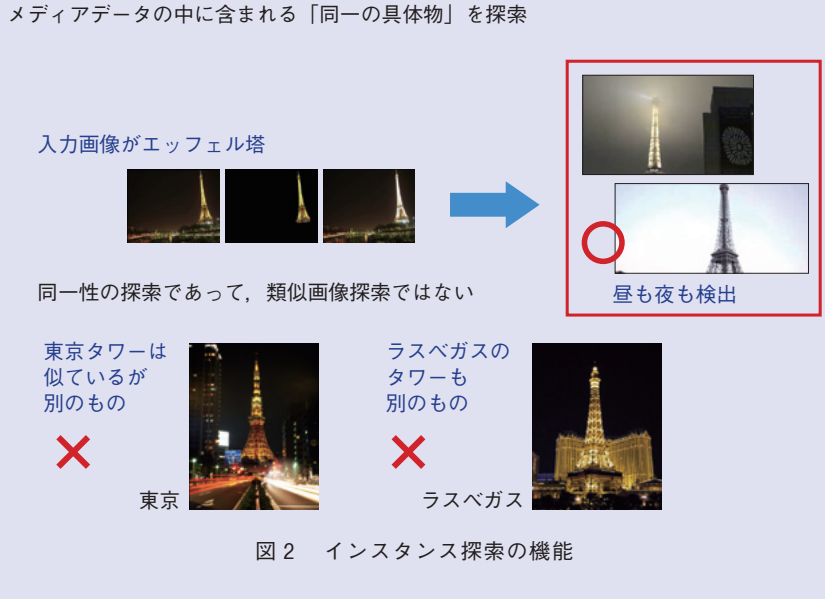

図2　インスタンス探索の機能

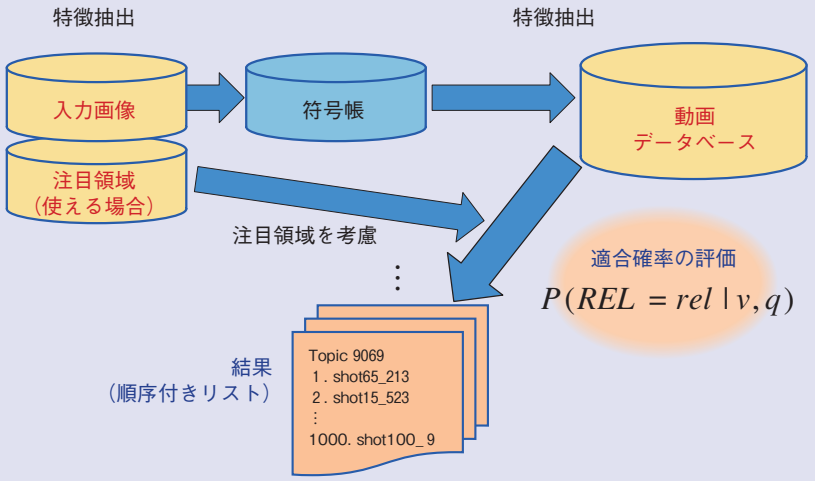

図3　インスタンス探索の仕組み

図4　提案法の効果

しつつあり，そのメディア認識技術の進化によってビッグメディアをさらに活用できるようになることを述べました．メディア認識技術は，実世界と情報世界の橋渡しという，今後のICT活用において大変重要な役割を担っていますので，身近なところで私たちの暮らしに役立ちながら，さらに進化を続けていくと考えられます．私たちも，安心・安全，便利で豊かな社会の実現に向けて，研究コミュニティにおいてはもとより，実際のフィールドや応用領域の方々とも密に連携しながら，さらなるCo-Innovation に取り組んでいきます．

■参考文献
⑴ 柏野・向井・大塚・永野・泉谷・木村・黒住・大和：“高速メディア探索，”NTT技術ジャーナル，Vol.19，No.6，pp.29-32，2007.
⑵ 木村・数原・高橋・横山：“画像検索でのユーザ行動を利用した大規模画像アノテーション，”信学論(D)，Vol.J96-D，No.8，pp.1711-1723，2013.
⑶ A. Kimura，K. Ishiguro，M. Yamada，A. M. Alvarez，K. Kataoka，and K.Murasaki：“Image Context Discovery from Socially Curated Contents，”ACM Multimedia，pp.565-568，2013.
⑷ M. Murata，H. Nagano，K. Kashino，and S. Sato：“NTT Communication Science Laboratories and National Institute of Informatics at TRECVID 2013 Instance Search Task，”TRECVID 2013 Workshop paper，2013.

◆問い合わせ先
NTTコミュニケーション科学基礎研究所
企画担当
TEL　0774-93-5020
FAX　0774-93-5015
E-mail　cs-liaison@lab.ntt.co.jp