パターン認識・メディア理解の基礎技術に関するOpen Idea

Open Ideas on Fundamental Research in Pattern Recognition and Media Understanding

安倍　満　木村昭悟　舩冨卓哉

Abstract

パターン認識・メディア理解分野においてこれから取り組むべき研究課題に関するopen ideaのうち，当該分野の基礎技術に関する項目を中心に報告する．パターン認識・メディア理解の処理の流れに沿って挙げられた項目を俯瞰したのち，その中から光情報の究極の観測，データの収集における本質的課題，認識の正答が一意ではない問題を取り上げ，それぞれにおいてこれから取り組むべき研究課題に関するopen ideaについて詳述する．

キーワード：パターン認識・メディア理解，グランドチャレンジ，open idea，基礎技術

1. パターン認識・メディア理解の基礎技術に関するopen ideaの曼荼羅

本企画で議論されたopen ideaの中から，本稿ではパターン認識・メディア理解の基礎技術に関する項目を中心に報告する．本稿では個々について全てを細かく記載することは避け，執筆担当者がそれぞれ一部を取り上げて説明することとするが，図1に列挙された項目を示しておく．基礎技術に大別された技術にも様々なレベルのものがあり，また相互に関連する項目もあった．挙げられた項目を整理する方法も幾つか議論されたが，ここではパターン認識・メディア理解の処理における流れに沿って「データの観測」，「データの収集」，「データへの正解ラベル付け」，「データ表現」，「扱う問題の定式化」，「認識処理」に整理し，列挙した．

安倍　満　正員　(株)デンソーアイティーラボラトリ研究開発グループ
E-mail　manbai@d-itlab.co.jp
木村昭悟　正員：シニア会員　日本電信電話株式会社NTTコミュニケーション科学基礎研究所
E-mail　akisato@ieee.org
舩冨卓哉　正員　奈良先端科学技術大学院大学先端科学技術研究科
E-mail　fuantomi@is.naist.jp
Mitsuru AMBAI, Member (Research & Development Group, Denso IT Laboratory, Inc., Tokyo, 150-0002 Japan), Akisato KIMURA, Senior Member (NTT Communication Science Laboratories, NIPPON TELEGRAPH AND TELEPHONE CORPORATION, Atsugi-shi, 243-0198 Japan), and Takuya FUNATOMI, Member (Graduate School of Science and Technology, Nara Institute of Science and Technology, Ikoma-shi, 630-0192 Japan).
電子情報通信学会誌　Vol.101 No.10 pp.996-1003 2018年10月

本稿で取り上げなかった項目については限られた文章から内容を推察するしかすべがないかと思うが，読者それぞれの視点から解釈して頂き，連鎖反応的に“open idea”の輪が広がるきっかけとなれば幸いである．以降では，執筆担当者の主観により幾つかの項目に絞って詳細を説明する．

2. 光情報の究極の観測に向けて

本章では，パターン認識・メディア理解の最初のステップである「データの観測」におけるopen ideaについて，一つの話題を紹介する．パターン認識・メディア理解技術が対象とする観測データには画像・音声だけでなく，例えば加速度センサの時系列データなど，様々な信号が含まれるが，特にここでは盛んに研究が行われている画像に着目する．

一般的に，いわゆる画像とは光の強度分布を二次元で記録したデータであるが，時間軸も含めた画像列，いわゆる映像が処理対象となることも多い．また，一概に画像といっても，モノクロ画像やカラー画像だけでなく，近赤外画像やマルチスペクトル画像などを対象とした研究もある．他にも，シーンを複数の視点から捉えた画像の集合である多視点画像も処理の対象となっている．このように，一概に画像といっても，様々な種類のデータが処理対象となっている．そもそもこれらの画像はいず

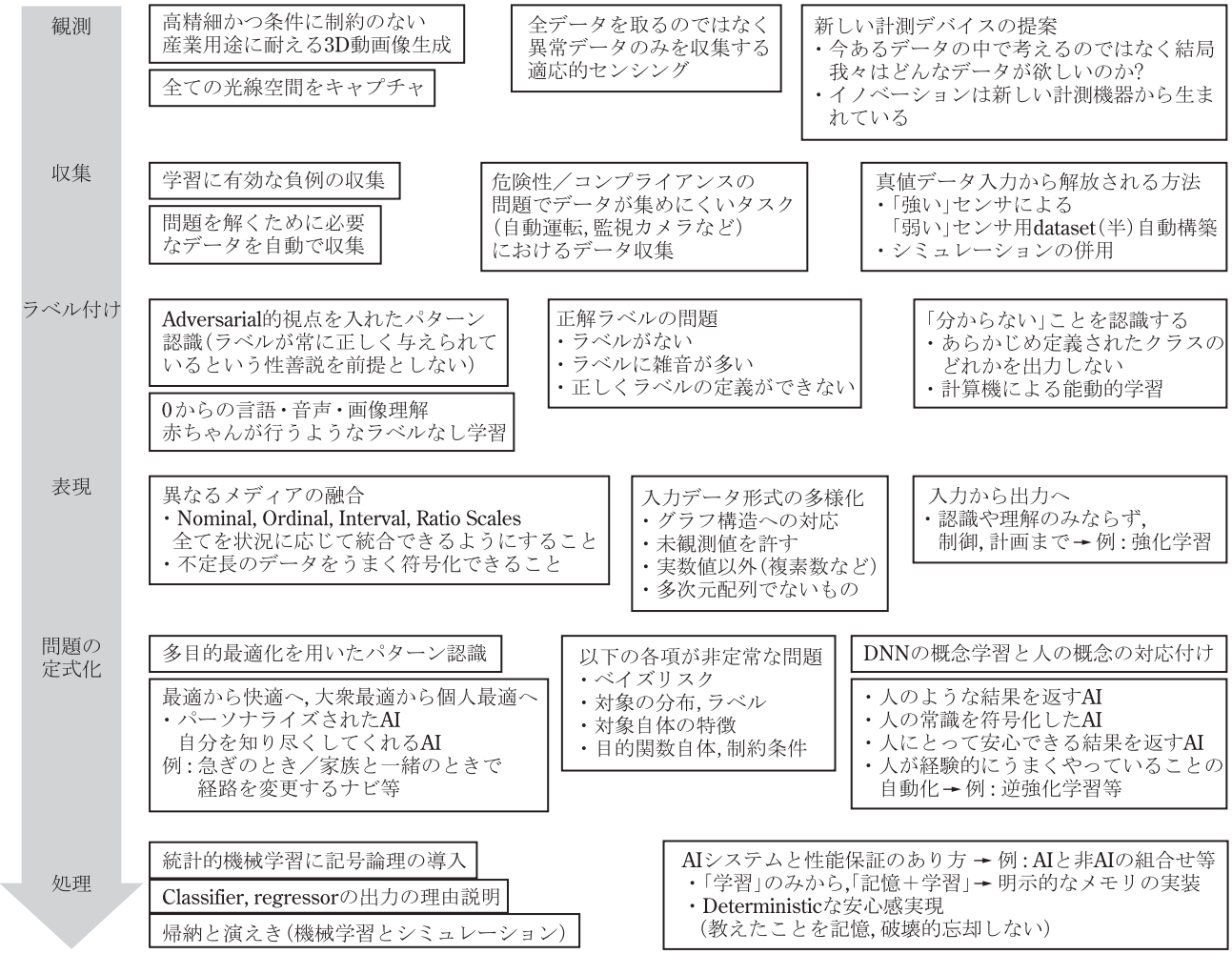

図1　パターン認識・メディア理解の基礎技術に関する open idea の曼荼羅

れも，シーンを飛び交う光の強度分布を記録したものと捉えることができ，Adelson ら[1]によって提唱された plenoptic function という概念で整理されている．

plenoptic function は空間を飛び交う光線の強度を表す七次元の関数 $P(\theta, \phi, \lambda, t, V_x, V_y, V_z)$ として定義されている．ここで (θ, ϕ) は球面座標系で表現した光線の角度，λ は光の波長，t は時刻，(V_x, V_y, V_z) は三次元空間の座標である．例えばピンホールカメラで二次元画像を取得する場合，ピンホールの位置が (V_x, V_y, V_z) に対応し，画像として得られる二次元の輝度分布はピンホール位置を通過する光線の角度 (θ, ϕ) についての強度分布に対応する．画像の輝度値は，カメラの分光感度特性に応じて λ について積分し，露光時間に応じて t について積分した値となる．映像であれば t，カラー画像やマルチスペクトル画像は λ，多視点画像であれば (V_x, V_y, V_z) のサンプリング数を増やしたものと捉えることができる．

人の視覚を模倣するセンサとしてカメラを考えるならば，λ, t についてのセンサの分解能はそれぞれ三つの錐体やその時間分解能に応じて設計することになる．実際，一般的なカメラが RGB の 3 チャネルを持ち，映像が約 30fps で記録される点などは，人の視覚に準じて設計されたと考えることもできる．また，特に近年の深層学習技術の進展に伴い，一般的なカメラで観測された画像を処理することで，様々なパターン認識やシーン理解が実現されていることも事実である．

しかし，光が元来持つ豊富な情報を生かすためには，必ずしも人の視覚を模した観測が適切ではなく，目的に応じた適切なサンプリングによって観測が行われること，つまり人の視覚機能にとらわれずに光を観測することが重要となる．実際，人とは異なる視覚系を有している生物はたくさん存在し，それらはその生態に合わせて適切な機能を獲得したと考えられる．

また科学の分野でも，人の視覚機能をはるかに超えた能力で光を観測し，研究が進められている．例えば天文学の分野では，天体の速度計測や距離計測に光の波長の変化（赤方偏移）が利用されており，例えば SUBARU Prime Focus Spectrograph では約 2 Å の波長分解能で

宇宙を観測している(注1)．光学の分野では，一瞬で起こった出来事を 4.4 兆フレーム／秒(2)で記録する撮影法が提案されている．いずれも，λ，t において人の視覚をはるかに凌駕する分解能での観測を可能とするものであり，最先端の科学で求められている性能であると言えよう．

このような最先端科学における観測と比較すると，人の視覚を模した一般的なカメラで撮影した画像では，光が持つ豊富な情報が縮退しており，そこからできることは限られてしまう．短絡的に考えれば，plenoptic function で定義されている七つの変数それぞれに対して高分解能な観測ができれば，更に高度なパターン認識やシーン理解が可能になると考えられる．しかし，当然ながらそのような観測は容易ではない．光を受けてその強度を測るためのセンサは二次元平面であり，時間変化を加えるとプラス一次元の計三次元の計測が可能であるが，七次元の観測にはまだ不十分である．以上を踏まえ，次のような open idea を提唱する．

〈open idea〉
課題：
二 ＋ 一次元センサによる効率的な高次元光情報の取得の実現．
意義：
普通のカメラでは区別できないような細かい差異まで観測でき，より豊富なシーンの情報を取り出せる．
困難性：
光学設計，電子デバイス設計，情報処理アルゴリズム設計が三位一体となった研究開発が必要．

科学技術分野で利用されている観測機器では，二次元センサのそれぞれの次元を (θ, ϕ) に割り当てるのではなく，物理的な機構を工夫して波長 λ や時間 t に割り当てることによって，通常のカメラの能力をはるかに超えた分解能での光計測を実現している．例えば，走査式のハイパスペクトルカメラであれば，回折格子などを用いて波長軸 λ をセンサ平面の一次元方向にマッピングし，残り一次元はスリットを用いて θ にマッピングしている．このスリットを走査することによって ϕ を時間軸にマッピングすることで，多波長画像 (θ, ϕ, λ) を取得することができる．ただし，ϕ の獲得に時間軸を用いているため，動的なシーンの計測には不向きである．あるいは，カメラの性能を大きく超えた時間分解能を得るため，ストリークカメラでは入射光を光電子に変換し，時間変化する電圧によってこの軌道を曲げることによって，入射光がセンサに到着する時刻 t をセンサ平面の一次元方向にマッピングしている．

以上の例では，波長 λ や時間 t を (θ, ϕ) に変換することによって高分解能を達成していたが，観測可能な次元の数が増えたわけではない．これらと異なり，撮影した後からフォーカスを変更することができるカメラとして注目されたライトフィールドカメラ(3)は，マイクロレンズアレーを用いることによってセンサ平面の各軸に光情報の複数の次元 (θ, V_x)，(ϕ, V_y) をそれぞれマッピングしている．このカメラは，二次元平面のセンサで四次元の光情報を記録していると捉えることができ，これに倣った設計の工夫ができれば，高次元な光の情報をうまく二次元平面に展開することも可能であると考えられる．光情報が本来持つ七次元情報をできるだけそのままに二 ＋ 一次元へ展開できれば，パターン認識・理解技術の入力が持つ情報が豊富になり，シーンに関するより深い理解が実現できると期待される．

一方で，高次元な光情報を低次元なセンサに展開した場合，それぞれの次元の解像度は劇的に低下してしまう．センサの解像度自体を向上させればこの問題の解決につながるが，光エネルギーの総量は変わらないとすると，光情報を高次元に分解すればするほど，個々の観測が受けるエネルギーは微弱となり，雑音の影響が大きくなることも問題となる．そのため，高分解能な計測を究めるだけではなく，効率の良い計測が求められるだろう．この問題の解決には以下のような仮定が有効であろうと考えている．

まず，仮に七次元全てで高解像度に光情報を取得できたとしても，そのほとんどが冗長であると予想される．それは例えば，視点位置 (V_x, V_y, V_z) を変えても見える景色は劇的には変化しないし，分光分布はほとんどにおいて波長 λ に関し滑らかである．また，シーン自体の変化も時間的には滑らかであると自然に仮定されている．つまり，各次元において光情報はほとんど滑らかに変化していると考えられる．その中でも特に重要なのは，変化が特異な部分，つまり通常の画像におけるエッジ部分 $\left(\frac{\partial P}{\partial \theta}, \frac{\partial P}{\partial \phi}\right)$，視差の境界 $\frac{\partial P}{\partial X}$，分光分布が大きく変化する波長 $\frac{\partial P}{\partial \lambda}$，シーンが急激に変化する時刻 $\frac{\partial P}{\partial t}$ などであり(注2)，更に，そのような変化が特異な部分は疎であると仮定できるだろう．

全体の中でも特に変化が特異な部分だけを選択的に，かつ効率的に観測するには，スパースモデリングや圧縮センシングなどの情報処理技術と組み合わせるアプローチが有効であると予想している．光学設計，電子デバイス設計，情報処理アルゴリズム設計が三位一体となった研究開発が必要となるだろう．　　（執筆：舩冨卓哉）

（注 1）　http://pfs.ipmu.jp/research/parameters.html（平成 30 年 4 月 17 日取得）
（注 2）　文献（1）においても初期視覚における機能として同様の議論がされており，plenoptic structure と名付けられている．

3. データ収集の未来像

統計的機械学習に基づくパターン認識の研究は，学習とテストのためのデータを収集するところから始まる．大きくは教師なし学習と教師あり学習の二つに分類できるが，画像認識の実課題においては（特に製品として社会実装されているものについては），人手で真値を準備し，入力と出力を関係付ける写像を真値から学習する教師あり学習が適用されることが多い．ImageNet もまたクラスラベルの真値が付与されているデータセットであり，100 万枚を超える画像に対してラベルを付与されている．このデータセットにおいて研究者が識別タスクの性能を競い合った結果，畳込みニューラルネットワークの有効性の再発見につながった[4]．大規模にラベル付けされたデータセットがその後の研究コミュニティの方向性を決定付けたという意味で，印象深い出来事であったと言えよう．

3.1 データアノテーションの呪い

しかし，解くべきタスクが単純なクラス識別から物体検出・領域分割・オプティカルフロー推定などと高度になるについて，手作業による真値の入力が極めて困難になりつつある．例えば Xie ら[5]は，データ収集における本質的課題として「データアノテーションの呪い」を提唱している．図 2 は文献(5)からの引用であり，データ入力に要する時間と画像枚数の関係について，主要なデータセットごとにプロットしてある．ImageNet のように，画像にクラスラベルを付与するというタスクであれば，1 枚の画像に対して真値を付ける労力は高々 1 分程度で済むものの，1 ピクセルごとにクラスラベルを推論する semantic segmentation のような高度なタスクでは，学習データにも 1 ピクセルごとにクラスラベルの付与が必要であるため，1 枚当りのアノテーションの時間がはるかに増大していることが読み取れる．semantic segmentation に限らず，人体の姿勢推定や指の関節位置推定など，高度な認識タスクであるほどデータの準備が難しく，データ収集には工夫を凝らす必要が出てくる．機械学習でよく言われる本質的課題として次元の呪いというものがあるが，つまるところデータ収集自体も本質的課題として呪われているというわけである．なお，Xie らはデータアノテーションの呪いを突破するために，二次元画像に直接真値を入力するのではなく，三次元地図上にアノテーションをしておき，ここから擬似的に生成した真値を二次元画像に付与するというアプローチをとった．アノテーションの方法自体を工夫することで，データアノテーションの呪いの回避を狙ったわけである．

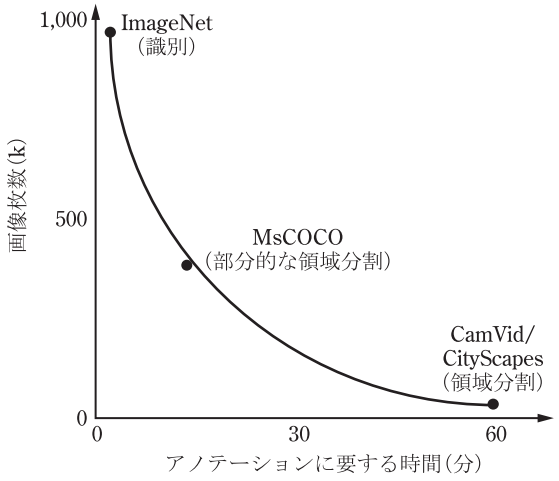

図 2 データアノテーションの呪い 文献(11)の図 1 を簡易化して引用した．タスクが複雑になるほど，データ入力に要する所要時間が長くなり，十分多くのデータを集めることが困難になる．

そして農業や介護などの社会課題を解くとなると，データアノテーションの難易度は更に高まるかもしれない．介護のように人と人とのコミュニケーションを扱う分野では，何をどのように定量化すべきかすら不明なことも多い．このようなタスクにおけるデータセットの準備は，技術的に興味深くもひどく悩ましい問題である．

3.2 コンプライアンスの壁

データ収集に立ちはだかる壁は，真値の入力に要する作業労力の問題だけではない．例えば，自動運転におけるデータ収集について考えてみよう．車線逸脱や近接車両の急接近など，危険な状況から回避行動をとるためには，そのような危険なシーンが学習データに多く含まれていることが理想である．しかしながら，もちろんそのような危険なシーンを意図して撮影するわけにはいかない．データ収集中に意図せず危険なシーンを撮影できてしまったとしても，そのようなシーンは膨大なデータセットの中のごく一部である．基本的に我々が入手できるデータは，安全な走行シーンが続く極めて偏りのあるデータセットでしかないのである．安全性に関するコンプライアンス遵守だけでなく，プライバシーやセキュリティに関しても同様のことが言えるだろう．多くのデータセットは，コンプライアンス遵守の観点から暗黙的にバイアスを受けていると考えるべきである．

3.3 評価のジレンマ

また別の観点として，パターン認識アルゴリズムの性能を評価するときにも，同様の困難さが生じる．例えばランド研究所のレポート[6]には，自動運転の安全性を統計的に証明することの難しさについて述べられている．Kalra らは，自動運転が人間による運転よりも安全であることを証明するためには，どの程度の距離を走行しなければならないかを，幾つかの大胆な仮定の下で概算した．その結論を引用すると，物損事故で 5,100 万マイ

ル，負傷を伴う事故で1億2,500万マイル，そして死亡事故で実に88億マイルの走行が必要とのことである．更に困ったことに，自動運転車が安全になればなるほど，その安全性評価が難しくなるというジレンマがある．無論，レポートで示された必要距離は市場投入前に試走できるデータ量を明らかに超えている．要するに単純に実データを大量に集めて評価をするというこれまでのアプローチとは，全く別の枠組みが必要ということを示唆しているわけである．

3.4 静的なデータセットから動的なデータセットへ

そこで，ここでは以下のようなopen ideaについて考察してみたい．

〈open idea〉
課題：
コンプライアンス上の問題によるデータ収集自体の困難さ，データの分布の偏り，真値入力作業の複雑化に伴う準備コストの増大は，データ収集の本質的課題と言える．
意義：
新しいデータセットの在り方が次の研究分野をけん引し得る．
困難性：
データ収集の本質的課題を解決できる「データセット」の在り方とは何か？

振り返ると，これまでのパターン認識分野において，データセットといえば入力と真値のペアが整理された状態で大量にストレージに保存されているものが主役であった．一方，強化学習の分野では行動（action）に応じて環境（environment）が変化し，観測（observation）が得られるというサイクルを繰り返す仮想環境がいわばデータセットとして活用されている．例えばよく知られているものとして，OpenAI gym[7]などがある．前述したような様々なジレンマを解消するためには，行動に応じて動的にデータ生成を行うような仮想環境が，これから先は積極的に活用されていくかもしれない．仮想環境であれば，危険な蛇行運転といったコンプライアンスに反する運転行動のデータを集めることも容易で，事実上無限のデータ収集を行うことができる．コンピュータグラフィックスを援用すれば，様々な真値を半自動的に生成することも可能であろう．

2018年3月に開催されたGTC2018において，NVIDIAが自動運転車の安全性テストを目的とした仮想環境DRIVE Constellationを開発したという発表があった．これはカメラ，ライダなどの自動運転車のセンサをシミュレートし，暴風雨・吹雪等の異常な天候を含む様々なテスト環境を膨大に生成するシステムという触れ込みで，明らかに前述の課題に対する意識が根底にあると思われる．また，ゲームエンジンとして幅広く浸透しているUnityは，強化学習のフレームワークとしてUnity Machine Learning Agents[8]を提供し始めている．ここにきてCGから機械学習へのアプローチが増えてきたことは興味深い．しかしながら，厳密にはCGと実画像はデータの生成過程が異なっており，CGで学習したモデルが実画像で良い汎化性能を達成できるかどうかは検討の余地がある．また逆に，CGによる仮想環境下でシステムを網羅的にテストしたとしても，それをもってして実世界における安全性を担保したことになるかどうかについては，まだまだ疑問が残るところである．とはいうものの，CGと実画像のギャップを埋めるための研究がここ数年の間に多数発表されており，新たな展開を見せそうな様相である．

CGとCVの研究者は，その境界領域としてimage based renderingなどといった任意視点生成技術を実現してきた．映画や放送の分野では既に実用化されてきた技術ではあるが，動的データセットたる仮想環境を実現する上で，任意視点生成技術は次の主要なトピックになるかもしれない．Bojarskiら[9]は，水平方向に3台のカメラを並べた車両を走行させ，車線逸脱（ドリフト）したときの画像を擬似的に生成することで，危険なシーンを学習するというユニークなアプローチを提案した．これを用いて画像からステアリングを直接回帰する自動運転システムを開発している．簡易的ではあるが，これは正にimage based renderingを機械学習に応用した好例であり，今後もこうした動きが広がるのではないだろうか．

静的なデータセットであるImageNetが深層学習の大きな潮流を生んだことを考えれば，仮想環境のような動的データセットが十分に整ったとき，次のブレイクスルーが生まれる可能性も少なくなかろうと思う．このような新しいデータセットの在り方から次のパターン認識の潮流が生まれる未来に，個人的には大いなる期待を寄せている．（執筆：安倍　満）

4. 認識処理再考

パターン認識における最も基本的な問題設定は，与えられたパターン$\boldsymbol{x}$に対して何らかの認識結果$\boldsymbol{y}$を出力する，というものである．認識問題をクラス分類で定義できる場合には，出力は離散値のクラスラベルである．また，認識が何らかの予測を伴う場合には，出力は実数値若しくは実数ベクトルとなる．この他，画像から物体領域を抽出するセグメンテーション，画像を入力して説明文を出力する画像説明文生成，画像と属性情報を入力して画像を属性に従って変換する画像変換など，様々なタスクがこの枠組みに含まれる．

標準的なパターン認識の問題設定では，この入出力関係 f を決定的若しくは確率的な関数でモデル化する．すなわち，入力が同一であれば出力も確率的な変動 ε を除けば同一であることを仮定する．

$$\boldsymbol{y}=f(\boldsymbol{x})+\varepsilon,\quad \varepsilon\sim p(\varepsilon).$$

実用上の多くの場面において，この仮定は十分にリーズナブルであり，各種の機械学習・最適化技術の恩恵により，入出力関係 f が極めて複雑であってもその関係を表現できるようになっている．

しかし，現実世界におけるパターン認識の問題は，入出力関係が必ずしも関数として表現できない，すなわち，入力が同一であっても出力が同一とは限らない，「正解不定の認識問題」も多数存在する．

〈open idea〉
課題：
認識の正答が一意ではない問題をどのように扱うか．
意義：
個人・文脈・環境に合わせて適切な応答を返せる，より人間に寄り添う認識システムの構築が可能になる．
困難性：
個人・文脈・環境をどのように programmable な形に表現するか，個人・文脈・環境をどのように sensing するか，あるいはそれらの影響をどのようにして小さくするか．

その端的な例として，正解が状況によって異なる「パーソナライズド認識」と，正解がどこにもない「究極の選択マシン」を考える．

パーソナライズド認識とは，唯一無二の正解が存在することを仮定する従来の認識問題とは異なり，個人・文脈・環境によって正解が異なる認識問題とする．先行研究[10]では，物体認識のタスクにおいて，最も詳細な物体名称ではなく，万人に理解できると思われる一般的な物体名称を認識結果とする（図 3）ことにより，パーソナライズド認識の複雑な問題を回避しようとしている．しかし，実際には，認識結果を受け取る人の知識や趣味し好によって，どの程度詳細な物体名称を返すのが最良であるのかは大きく変わり得る．

究極の選択マシンとは，考え得るどの選択肢を選んでも同等の結果になる，正解がどこにもない認識問題であるとする．Dumon による記事[11]では，自動運転における事故の例が挙げられている（図 4）．この例では，1 から 3 のいずれの選択肢を選択しても，自分若しくは他人

図 3　パーソナライズド認識の考え方（文献（8）から引用，英文は和訳して表示）

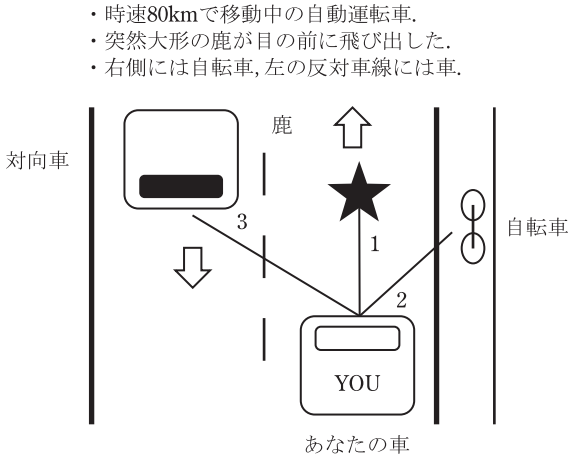

図 4　究極の選択マシンの一例（文献（3）から引用，一部改変）

が高い確率で死ぬことになるが，それを回避する手段はなく，何らかの判断をしなければならない状況である．この問題は，本質的に正解がないわけではなく，ある評価尺度，例えば「自分が死ぬ確率を最小にする」などの基準をその場で選択することができれば，その基準においての正解はただ一つ（図 4 の例では 2 を選択）に定まる．すなわち，この問題の本質的な困難さは，想定し得る評価尺度が複数存在し，どの評価尺度が実際に使われるのかを認識システムが事前に把握できない点にある．

これらの例はいずれも，認識対象は同一であるにもかかわらず，認識結果の評価基準が様々な要因によって大きく変化するため，最適な認識結果が不定になる問題であると考えることができる．

これに加え，認識問題の正解自体がそもそも不定となる場合も多く存在する．例えば，感情・意図など人間の内部状態を予測する認識問題の場合には，内部状態の正解を定義することが極めて困難であり，内部状態との相関が強いとされている計測量・統計量を用いた別の認識問題の出力を正解として採用せざるを得ない場合も多い．人間の内部状態の予測に関する議論は，後述の本小特集3の1.を参照されたい．

ここまでの議論を踏まえ，上記に示した「正解不定の認識問題」の困難さをより明確にするため，この問題を以下のような最適化問題として単純化する．ここで，f, g 及び h はそれぞれ，入力 $\boldsymbol{x}$ から認識結果 $\hat{\boldsymbol{y}}$ に変換する認識関数，観測困難な真の正解 $\boldsymbol{y}$ を観測可能な形 $\tilde{\boldsymbol{y}}$ に変換するラベリング関数，認識関数の出力 $\hat{\boldsymbol{y}}$ を正解 $\tilde{\boldsymbol{y}}$ と比較する評価関数であり，$\boldsymbol{\theta}_*$ は各関数のパラメータ，$\boldsymbol{z}_*$ は未知の環境変数，ε_* は観測雑音である．

$$
\begin{aligned}
&\hat{\boldsymbol{\theta}}_f = \min_{\boldsymbol{\theta}_f} h(\tilde{\boldsymbol{y}}, \hat{\boldsymbol{y}}; \boldsymbol{z}_h, \boldsymbol{\theta}_h),\\
&\hat{\boldsymbol{y}} = f(\boldsymbol{x}; \boldsymbol{z}_f, \boldsymbol{\theta}_f) + \varepsilon_f,\\
&\tilde{\boldsymbol{y}} = g(\boldsymbol{y}; \boldsymbol{z}_g, \boldsymbol{\theta}_g) + \varepsilon_g.
\end{aligned}
$$

パーソナライズド認識や究極の選択マシンは評価関数 $g(\cdot)$ の環境変数 $\boldsymbol{z}_g$ の影響が大きい例であり，人間の内部状態を予測する問題はラベリング関数 $g(\cdot)$ の形状あるいは環境変数 $\boldsymbol{z}_g$ が未知である事例である．この定式化によれば，正解不定の認識問題の困難さは，仮に各関数の形状が既知であったとしても，環境変数 $\boldsymbol{z}_*$ が未知，かつ多くの実問題において観測困難であるため，常に最適な出力を返す認識関数 f の設計若しくはそのパラメータ $\boldsymbol{\theta}_f$ の推定は極めて困難である，という点にある．

この困難さの解決のためには，少なくとも以下の二つの課題を解決する必要がある．

1. 未知かつ観測困難な実現象をどのようにして環境変数 $\boldsymbol{z}_*$ として表現するか．
2. 未知かつ観測困難な環境変数 $\boldsymbol{z}_*$ の影響をどのようにして小さくするか．

数理的な観点に立つとこれら二つの課題は表裏一体であり，環境変数の影響が小さくなるようなモデル化を導入できる環境変数の表現形式を選択する必要がある．例えば，環境変数 $\boldsymbol{z}_*$ を支配する何らかの確率密度関数を仮定して，環境変数について各関数を周辺化することで環境変数を消去するアプローチが考えられる．パーソナライズド認識の例では，環境変数として対象物体に対する知識水準を採用して，これを実数スカラ値で表現し，この環境変数の分布として正規分布を採用すると，この周辺化操作による結果として，平均的な知識水準を持つ人に合わせた認識結果を返す認識器が得られる．この例のように，非常に単純化されたモデルを利用すれば，幾つかの限定された状況においては，正解不定の認識問題の解法が見つかる可能性がある．

しかしながら，環境変数の予測若しくはモデル化は一般には容易ではない．なぜなら，多くの場合，環境変数は人間の内部状態に強く関連しているためである．パーソナライズド認識や究極の選択マシンの例では認識結果を評価する人間の知識水準・選好基準，感情・意図認識などでは正に人間の内部状態そのものを予測しようとしている．

環境変数が認識システムに関わる人間の内部状態と強く関連することに着目し，その人間との間で何らかのインタラクションを許容すると，環境変数の影響を小さくできる可能性がある．パーソナライズド認識の例では，認識結果を受け取る人間との対話が可能であれば，その対話を通じて知識水準や趣味し好を絞り込むことができ，認識結果を受け取る人間の意図に沿った認識結果を出力できるようになるだろう．

インタラクションの導入を許容すると，正解不定の認識問題を強化学習の一種として捉えることもできるかもしれない．すなわち，評価関数 $h(\cdot)$ の出力を報酬として，認識関数 $f(\cdot)$ のパラメータ $\boldsymbol{\theta}_f$ を何らかのインタラクションによって最適化する，という可能性である．強化学習を用いるアプローチでは，環境変数が時間的あるいはインタラクションによって変動し得る場合を自然に包含している．

更には，入力パターン $\boldsymbol{x}$ に別の補助情報を付け加えることによって，評価関数の環境変数 $\boldsymbol{z}_g$ の影響を小さくするアプローチも考えられる．このとき，認識関数の環境変数 $\boldsymbol{z}_f$ が補助情報の役目を果たすことになる．パーソナライズド認識の例で具体的に説明すると，単に画像の物体名を返すのではなく，「一般的な名称としては…です，より詳細には…と呼ばれます」などのように，解を制御する条件を認識側で準備しておき，その条件と解を同時に提示する，という戦略が対応する．

ここまでの議論から一端が見えるように，パターン認識の問題は人間同士のコミュニケーションと結び付く点が少なくない．同じ情報を提示したとしても，受け取る側がどのように評価するかは大きく変わり得るし，その評価ができるだけ変動しないようにするためには，文脈を注意深く設計しなければならない(注3)．深層学習の発展によって通り一遍のことができるようになったと思われているパターン認識が，次の一歩を踏み出すとすれば，コミュニケーションとしてのパターン認識の再考になるのかもしれない．（執筆：木村昭悟）

（注3） http://blog.szk.cc/2016/10/08/does-it-need-a-sympathy/に，映像CMを題材とした例が紹介されている．

文　　献

（1） E.H. Adelson and J.R. Bergen, “The plenoptic function and the elements of early vi-sion,” In Computational Models of Visual Processing, pp. 3-20, MIT Press, 1991.

（2） K. Nakagawa, A. Iwasaki, Y. Oishi, R. Horisaki, A. Tsukamoto, A. Nakamura, K. Hirosawa, H. Liao, T. Ushida, K. Goda, F. Kannari, and I. Sakuma, “Sequentially timed all-optical mapping photography（STAMP）,” Nature Photonics, vol. 8, no. 9, pp. 695-700, DOI : 10.1038/nphoton. 2014.163, 2014.

（3） R. Ng, M. Levoy, G. Duval, M. Horowitz, and P. Hanrahan, “Light field photography with a hand-held plenoptic camera,” Technical Report CSTR, 2005.

（4） O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg, and L. Fei-Fei, “ImageNet large scale visual recognition Challenge,” Int. J. Comput. Vis.（IJCV）, vol. 115, no. 3, pp. 211-252, DOI : 10.1007/s11263-015-0816-y, 2015.

（5） J. Xie, M. Kiefel, M.-T. Sun, and A. Geiger, “Semantic instance annotation of street scenes by 3D to 2D label transfer,” The IEEE Conference on Computer Vision and Pattern Recognition（CVPR）, pp. 3688-3697, 2016.

（6） N. Kalra and S.M. Paddock, “Driving to safety : how many miles of driving would it take to demonstrate autonomous vehicle reliability?,” Tech. rep. RAND Corporation, 2016, https://www.rand.org/pubs/research_reports/RR1478.html

（7） OpenAI Gym, https://gym.openai.com/

（8） Unity Machine Learning Agents, https://unity3d.com/jp/machine-learning

（9） M. Bojarski, D.D. Testa, D. Dworakowski, B. Firner, B. Flepp, P. Goyal, L.D. Jackel, M. Monfort, U. Muller, J. Zhang, X. Zhang, J. Zhao, and K. Zieba, “End to end learning for self-driving cars,” arXiv : 1604.07316, 2016.

（10） V. Ordonez, J. Deng, Y. Choi, A.C. Berg, and T.L. Berg, “From large scale image categorization to entry-level categories,” 2013 IEEE International Conference on Computer Vision, pp. 2768-2775, Dec. DOI : 10.1109/ICCV. 2013.344, 2013.

（11） O. Dumon, “Why technology in our society still requires the human empathy component,” 2016, https://www.huffingtonpost.com/olivier-dumon/why-technology-in-our-society_b_9187018.html

（平成 30 年 5 月 16 日受付　平成 30 年 6 月 4 日最終受付）

安倍　満（正員）

2002 慶大・理工・情報卒．2007 慶大大学院理工学研究科後期博士課程了．博士（工学）．同年，(株)デンソーアイティーラボラトリに入社．現在，同社研究開発グループシニアリサーチャ．パターン認識，機械学習に関する研究に従事．

木村　昭悟（正員：シニア会員）

1998 東工大・工・電気電子卒，2000 同大学院理工学研究科修士課程了．同年，日本電信電話株式会社に入社．2007 東工大大学院理工学研究科博士課程了，博士（工学）．2017～2018 University of Cambridge 客員研究員．現在，NTT コミュニケーション科学基礎研究所企画担当主任研究員．パターン認識，データマイニング，機械学習に関する研究開発に従事．

舩冨　卓哉（正員）

2002 京大・工・情報卒．2004，2007 に京大大学院情報学研究科修士，博士課程をそれぞれ了．2006 学振特別研究員 DC2．2007 京大学術情報メディアセンター助教．2014 Stanford University 客員助教．2015 から奈良先端大准教授．博士（情報学）．三次元モデル，Computational photography に関する研究に従事．