In Laboratory Now

# 研究室訪問 3

# より賢いコンピューターへ

## 杉山 将 研究室～計算工学専攻

杉山 将 准教授

機械学習と呼ばれる技術は、様々な形で私たちの生活を支えている。平仮名の漢字変換、インターネット検索、指紋認証、ゲームの AI、例を挙げればきりがない。

杉山研究室ではそれぞれの応用例に対して個別に取り組むことよりも、それらに共通する課題の解決に力をそそいでいる。

その中でも『密度比』という考え方は、応用例の多くに共通する課題を解決できる。今回の取材では密度比とそのさまざまな応用例について話を伺った。

## あ 機械学習のもたらすもの

コンピューターは人間にはできないさまざまなことができる。高速な計算や大量のデータの蓄積はその代表例だ。だが一方で、人間にとって簡単なことがなかなかできない場合がある。

例えば、コンピューターに迷惑メールかどうかを判定させる事を考えて欲しい。単純に「特定の単語が入っていたら迷惑メール」というルールを作ったとしよう。このルールを用いれば多くの迷惑メールに対応できそうに見える。だが、迷惑メールに使われる単語の種類はさまざまであるため、その単語を登録するだけでも一苦労なうえ、そのうち今まで想定していなかった単語が出現する可能性もある。また、必要なメールでその単語が出てくることもあるため、その場合は迷惑メール扱いしないような例外処理を行わなければならない。こうして、正しく迷惑メールを判別するためにルールを増やそうとすると、その数はやがて膨大なものになってしまい、人手で作成するのは難しくなるのだ。

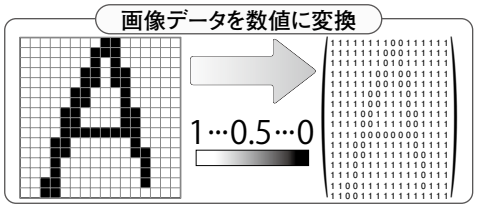


図 1　画像を数値のデータ

そこで登場するのが機械学習である。機械学習とは、コンピューターが手元にあるデータの中から、規則性や関連性などの知見を得る手法だ。ここではコンピューターに適当な数の迷惑メールを渡して、その特徴を学習させる。そうすることで、コンピューター自身が迷惑メールを識別するルールを作ることができ、人が大量のルールを作る必要が無くなるのだ。

では、コンピューターがデータの特徴を学習し、問題を解決する過程を詳しく見ていくために『A と B を見分ける手書き文字認識』の例を考えよう。手書き文字認識とは、コンピューターが人間の書いた文字を識別することだ。

この問題では、入力は『A という文字』か『B という文字』を表す手書き文字が描かれた白黒の画像で、サイズは縦 16 ×横 16 画素とする。これをコンピューター上では 16 × 16 の行列で表す。

行列の各要素はそれぞれの画素の色に対応した0から1までの実数の値を取り、画素の色が黒に近いほど値は0に近くなり、白に近いほど値が1に近づく(**図1**)。この行列を受取り、コンピューターはそれがAかBかを判断して答えなければならない。しかし、人間の書いた文字は、同じ文字でもさまざまな形になることがこの問題を困難にしている。

機械学習のアプローチではこの問題を解くために、まず色々な人にAを書いてもらい、そのサンプルをもとにしてAの特徴を『学習』する。特徴を学習するのには色々な手法があるが簡単な例として以下のようなものがある。

各文字のサンプル画像行列の平均をそれぞれ取り、平均的なAとBの画像行列を作成する。こうしてA、Bの特徴を学習した後、それらと判定したい文字の画像行列を比較し差が小さい方の文字を答える(**図2**)。

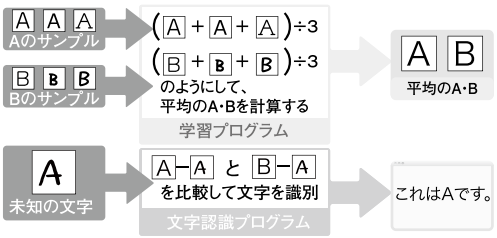


**図2　画像認識の機械学習**

以上のようにコンピューターに学習する機能を持たせることで、人間が細かいルールを与えなくともコンピューターが色々な事を判断できるようになる。

# 確率分布と確率密度関数

ここでは、機械学習にとって重要ないくつかの概念を説明する。手書き文字認識の例での出力は『それはAの文字です』『それはBの文字です』といったように断定的なものだった。ここで、歪んだ形の6面サイコロを考えてみる。何度もサイコロを振ると、その結果からサイコロの特徴を推定することができる。その特徴は『4の値が一番出ます』といったように断定的に表すことも可能であるが、『1が出る確率は1/12、2が出る確率は1/6……』のように確率で表した方がより多くの情報を得ることができる(**図3A**)。このように、それぞれの事象がどの程度起こるかを示したものを確率分布といい、機械学習においても断定的な値より、確率分布で結果を出力したいケースがよくある。

今、機械学習において気象レーダー画像から降水量を予測する問題を考える。「ちょうど15 mm雨が降る確率」はどれぐらいだろうか。「降水量15 mm」といっても厳密に考えれば15.0001 mmなどわずかに差があり15 mmではない。このように連続的な値を取る事象に対しては単純に確率を求められない。

ここで登場するのが確率密度である。確率密度とは連続的な値を取る事象のそれぞれの起こりやすさのことである。降水量がxの時の確率密度を返す確率密度関数をp(x)とすると、降水量が10 mm以上15 mm以下になる確率は[10, 15]の区間においてp(x)を積分する事で得ることができる(**図3B**)。このように確率密度関数をある区間に渡って積分する事でその区間の事象が起こる確率となる。この確率密度を使えば、各降水量ごとの起こりやすさを表すことができる。

これらの前提知識をもとに、杉山先生が取り組んでいる問題を見ていこう。

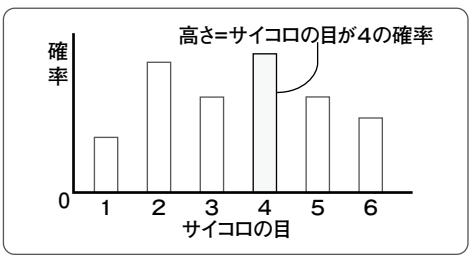


**図3A　サイコロの出る目の分布**

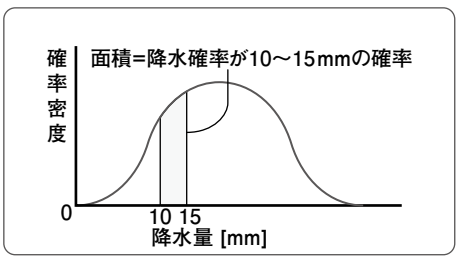


**図3B　降水量に関するグラフ(確率密度関数の例)**

# 密度比による異常値検出

異常値検出とは、データの集合から異常なデータを検出することである。以下のような例を考える。Aの手書き文字の集合を用いた学習から得た知見をもとに、調べたい手書き文字の集合に含まれたA以外の文字（異常値）を検出したい。以下、Aの手書き文字を集めた集合を訓練集合と呼び、調べたい手書き文字の集合をテスト集合と呼ぶこととする。

まず、訓練集合からAの文字がどのような形をとりやすいかという確率分布を求め、これをある画像データ x に関する確率密度関数 p(x) で表す（**図４左**）。この求めた p(x) において、p(x) の値が極端に小さいような x は、『Aの文字が滅多に取らない形をした画像データ』ひいては『異常なAの文字の画像データ』と言える。

このようにして確率密度関数から異常値を検出できることがわかる。ただ、この手法には問題がある。それは行列やベクトルである x の要素の数が多かったり、確率密度関数の形が複雑であった場合、p(x) を求めるのが困難であることだ。

この問題を解決するため、先生は密度比に注目した。密度比とは二つの確率密度関数の比のことである。ここでは、先程述べた訓練集合から求めた確率密度関数 p(x) ともう一つ、テスト集合から同じように求めた確率密度関数 q(x) の二つの比を考える。q(x) では、テスト集合に異なる文字が混じっており、その異なる文字もAが取りうる形だと認識されるため、本来ほとんどAに成りえないような画像データ x の q(x) の値が大きくなっている（**図４中央**）。よって確率密度比 p(x)/q(x) を考えると異常値においては飛びぬけて小さい値になっているのがわかる（**図４右**）。

以上のように密度比からも異常値が検出できることがわかる。しかし、p(x) と q(x) を一度求めて、そこから密度比 p(x)/q(x) を求めるのでは、依然困難であることに変わりはない。

先生はその困難さを回避するために、二つの確率密度関数を別個に求めることなく、直接密度比を求められる KLIEP（カルバック・ライブラー重要度推定法）という手法を考案した。

KLIEP では、KL 情報量という二つの確率密度関数の差を表す尺度を用いる。詳細は省くが、確率密度関数 p(x) に関して真の関数と予測した関数との間の KL 情報量が小さくなるように密度比を調整することで、p(x), q(x) をそれぞれ別個に求めることなく、直接密度比を推定することができる。これによって比較的容易に異常値を検出することが可能になる。

この異常値検出にはさまざまな応用例が考えられる。例えば生産ラインの不良品検出、パケット解析によるネットワークシステムへの不正侵入検出などが可能になる。また、異常値とは都合の悪い値のみを指すわけでは無く、たくさんのブログの文章における異常値の検出によって新しい話題の発生を検出することもできる。

KLIEP を用いることで可能になった密度比推定の適用範囲は異常値検出に留まらない。他にも次のような問題がある。

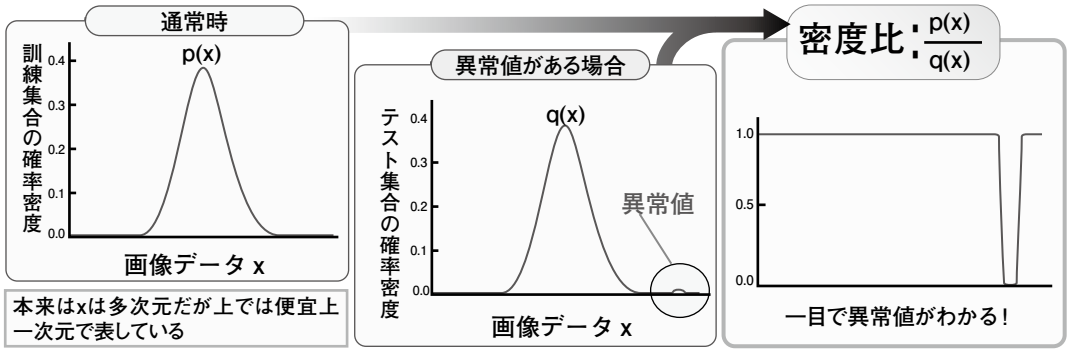


**図４　確率密度関数と密度比**

# 非定常環境下での学習

機械学習のポピュラーな課題の一つに直線フィッティング問題がある。この問題は未知の関数 f(x) に従う有限個のデータから f(x) がどのような値を取るかを推定する問題である。

まず、適当な数の (x, y) の対が与えられる。ここで y は以下のように表される。

$$y = f(x) + ノイズ$$

ノイズが現れるのは実験における観測誤差の様に、何らかの要因により理論値からずれるという現実を表している。

具体例を見ていこう。f(x) を sin(x)/x とし、そこからノイズの乗ったいくつかのサンプルを生成する（**図５左**）。今、直線で f(x) を近似しそれを g(x) とする。g(x) は以下のように表される。

$$g(x) = ax + b$$

後は、a と b を適切に決めればよい。この a と b を決めるには多くの場合、全てのサンプルにおける誤差の２乗 $(g(x)-y)^2$ の合計を最小にするように決める。こうして求めた g(x) を用いて、新しい x に対する f(x) の値を推定する（**図５左**）。

ここまでが通常の直線フィッティングの問題である。ここからさらに難しい非定常環境下、すなわち環境が一定でない状況における直線フィッティングを考える。

先ほどの例では、暗黙のうちに学習のために与えられた (x, y) の集合に含まれる x と、予測したい新しい x の分布が同じだと仮定していたが、非定常環境下では学習データの分布と予測したいデータの分布が変化する（**図５右**）。これは、例えば株価の予測を行う際などに生じる問題で、予測をする時点では学習に使用できるデータは過去のものしか無いが、予測する対象は未来のデータなので過去とはデータの分布が変化している。

この問題にも密度比を用いることができる。ここで考える密度比は学習データにおいてどのような x が出やすいかを表す確率密度関数 q(x) と、予測するデータにおいてどのような x が出やすいかを表す確率密度関数 p(x) との比、p(x)/q(x) だ。

先ほどは g(x) の a, b を決定するのに $(g(x) - y)^2$ の合計を最小にするような a, b を選択したが、今回は $(p(x)/q(x)) \cdot (g(x)-y)^2$ を最小にするように a、b を選ぶ。p(x)/q(x) は予測データにおいて出やすい x で大きくなるため、予測するデータにとって関連深い学習データが重視されるようになり、学習データ付近の予測は外れてしまうが、代わりに予測したいデータ付近での予測の精度が上昇するのである（**図５右**）。

この非定常環境下での学習は、先に述べた株価の予測の他、感情や体調に左右される脳波によってデバイスを操作するブレインコンピューターインターフェースなどに応用することができる。先生はそれらにも取り組んでいる。

ここまで杉山研の密度比に関する研究を紹介した。今回は取り上げなかったが、密度比推定は他の色々な問題にも適用可能である。先生は密度比推定をより正確にするとともに、それらの応用についても研究されている。この密度比の話からもわかるように一つの課題をクリアすることで多くの課題が解決することがある。先生はまず密度比を直接推定する手法を思いつき、そこから異常値検出などさまざまな応用へ適用していったそうだ。このように杉山研では多くの課題の解決に結びつくような研究を日夜行っている。

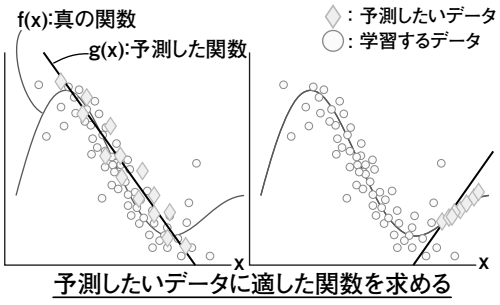


**図５　直線フィッティング**

---

取材では大変わかりやすく研究を説明して頂き、知的好奇心が刺激されました。お忙しい中、取材のみならず記事を丹念に読んでくださった杉山先生にお礼申し上げます。　　（老木　智章）