

密度比推定に基づく統計的機械学習 Statistical Machine Learning based on Density Ratio Estimation

東京工業大学 計算工学専攻

杉山 将

●あらまし

インターネットやセンサーを通して膨大な量のデータが容易に入手できる「ビッグデータ時代」が到来しつつあります。近年、このような大量のデータからいかにして有用な知識を得るかが重要な研究課題となっており、データの統計的な性質を活用する統計的機械学習が有望な情報処理パラダイムの一つとして注目されています。本稿では、筆者らがテレコム先端技術研究支援センターの研究助成を受けて開発してきた統計的機械学習の基礎技術である「密度比推定」を紹介します。密度比推定とは確率密度関数の比をデータから推定する技術の総称であり、密度比の推定を通して、非定常環境適応学習、特徴選択、クラスタリング、パターン認識、条件付き確率推定、独立成分分析、異常値検出など、様々なデータ解析タスクを統一的かつ優れた精度で解決することができます。

●はじめに

機械学習の目的は、与えられたデータからその背後に潜む一般的な規則を自動的に獲得することです[1,2,3,4]。機械学習の技術は、コンピュータによる自然言語の理解、顔画像の認識、音声の識別、ロボットの制御、DNAの解析、脳機能の解明など、様々な分野に応用されています。

一口に機械学習と言っても、データ解析の目的によって、パターン認識、回帰、クラスタリング、異常検知、特徴選択など、様々なタスクが存在します。機械学習の最も汎用的なアプローチは、データを生成する確率分布を推定することです。なぜならば、データの生成分布を知ることはそのデータに関する全ての知識を得ることと本質的に等価だからです。しかし、データの生成分布の推定は統計学的に最も困難な問題として知られています。

一方、パターン認識や回帰などデータ解析の目的が具体的に定まれば、その目的を直接達成するアルゴリズムを開発することが最善のアプローチです。例えば、与えられたデータをそれが属するカテゴリに分類するパターン認識の問題では、各カテゴリに属するデータの生成確率をそれぞれ推定すれば、その確率に基づいて新しいデータが属するカテゴリを正しく予測できます。しかし、パターン認識のためには必ずしもデータの生成確率を推定する必要はなく、異なるカテゴリ間の境界線さえ求められれば十分です。この考えに基づいて、サポートベクトルマシン[5]とよばれる最先端のパターン認識技術では、異なるカテゴリ間の境界線を直接求めることにより高い認識性能を達成しています。

このようにしてパターン認識に対しては優れたアルゴリズムが開発されましたが、パターン認識以外にも様々なデータ解析のタスクが存在し、それぞれに対して優れたアルゴリズムを開発することは困難です。実際、サポートベクトルマシンの基礎理論は既に196

0年ころから研究されて始めており、半世紀にも及ぶ研究開発を経て、様々な実データ解析に応用されるようになりました。ビッグデータ時代には大量のデータを前に迅速に意思決定を行うことが望まれていますので、このような長期間に及ぶ基礎研究を行うことは現実的ではありません。一方、最も汎用性の高い生成分布推定アプローチでは、精度良くデータ解析を行うことはできません。

そこで筆者らは、生成分布推定アプローチとタスク特化アプローチの中間を考え、ある条件を満たすデータ解析タスクのクラスに対してアルゴリズムを開発するというアプローチを提案しました。具体的には、複数の確率分布が含まれるデータ解析タスクのうち、それらの確率分布そのものは必要なく、確率密度関数の比さえわかればデータ解析を行うことができるというクラスを考えました。このクラスには、非定常環境適応学習、特徴選択、クラスタリング、パターン認識、条件付き確率推定、独立成分分析、異常値検出など、様々なデータ解析タスクが含まれます。そして、この確率密度関数の比を、それぞれの確率密度関数を推定することなく直接推定するアルゴリズムを開発し、これら全てのデータ解析タスクを統一かつ優れた精度で解決することができるようになりました（図1）。

この「密度比推定」の技術的な詳細は専門書[6]および解説記事[7]をご覧くださいことにし、本稿では密度比推定によってどのようなデータ解析が行なえるかを概観します。

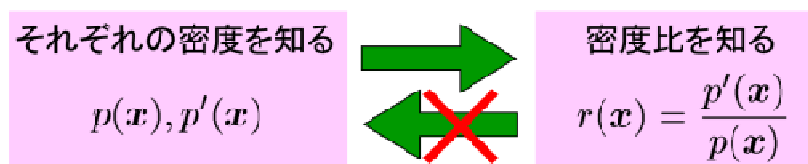


図1: 密度比推定. それぞれの密度を推定するよりも密度比を推定するほうが容易です。

●非定常環境適応学習

入力と出力が対になったデータの背後に潜んでいる入出力関係を推定する問題を教師付き学習とよびます（図2）。この名称は、入力が生徒の質問、出力が教師の答えに例えられることによります。未知の入出力関係を学習することができれば、学習に用いていない新しい入力に対する出力を予測できるようになります。未知の状況に対して一般化できるということから、これは汎化能力とよべます。この汎化能力の獲得こそが、教師付き学習の目的です。

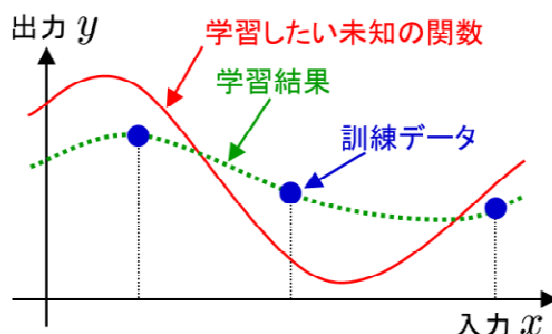


図2: 教師付き学習. 入力と出力が対になった訓練データから、その背後に潜んでいる入出力関係を推定します。

汎化能力の獲得を理論的に保証するために、学習に用いる訓練データと将来予測を行いたいテストデータが同じ規則に基づいて生成されているという条件が一般的に仮定されます。しかし、近年の機械学習の多くの応用分野では、この基本的な仮定が成り立ちません。例えば、脳波解析では脳の振る舞いが時間と共に変化するため、訓練データとテストデータの傾向が異なります。音声や画像の認識では訓練データとテストデータを収集する環境が一般に異なります。また、ロボット制御ではロボットの行動規則が学習と共に更新されるため、結果としてデータの生成規則が変化します。

一方、訓練データとテストデータが全く別の規則に基づいて生成されると、訓練データからテストデータの情報を予測することは原理的に不可能です。従って、訓練データとテストデータをつなぐ何らかの仮定が必ず必要となります。共変量シフトは、そのような仮定の一つです[8]。共変量とは入力データの別称であり、共変量シフトとは、入力データの生成規則が訓練時とテスト時で変化するが、入出力関係は変化しないという状況を指します。以下では、ブレイン・コンピュータ・インターフェースを例に、共変量シフトに対処するための適応学習技術を紹介します。

ブレイン・コンピュータ・インターフェースとは、脳波によって計算機に意志を伝える技術であり、手足を動かすことのできない人でもコンピュータを操作できるようにするための重要な技術です(図3)。ここでは、脳波でマウスカーソルを左右に動かすタスクを考えることにしましょう。学習の目標は、脳波パターンとその脳波によって伝えようとしている意志(左か右か)が対になった訓練データをもとに、脳波パターンと意志との関係を学習することです。これにより、将来与えられる脳波パターンによって示唆される意志を正しく予測できるようになります。ただし脳の非定常性のため、学習用の脳波パターンと将来与えられる脳波パターンは一般に異なる確率分布に従います。

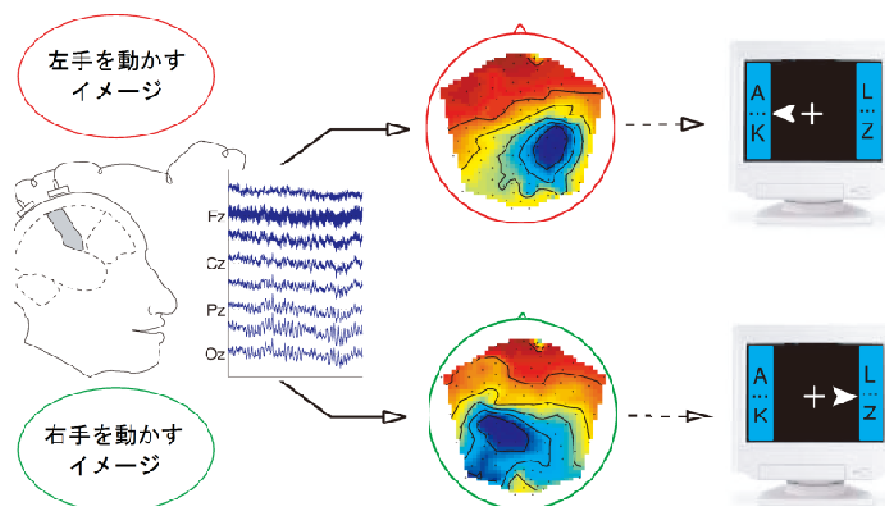


図3: ブレイン・コンピュータ・インターフェース. 脳波でコンピュータを直接操作します。

共変量シフト適応の考え方を図4に示します。訓練用の脳波パターンとテスト用の脳波パターンの確率密度関数の比によって与えられる「重要度」に従って、学習規準を重み付

けします。密度比推定によってこの重要度を推定することによって、共変量シフトに起因する非定常性に対して適応的に学習を行なえます。

共変量シフト適応技術は、上述したブレイン・コンピュータ・インターフェースにおける非定常環境適応以外にも、ロボット制御における標本再利用や最適データ収集、自然言語処理におけるドメイン適合、顔画像からの年齢予測における照明環境適合、加速度センサーからの行動識別におけるユーザ適合、話者識別における声質適合、半導体露光装置の位置合わせなど、様々な実問題に応用されています。

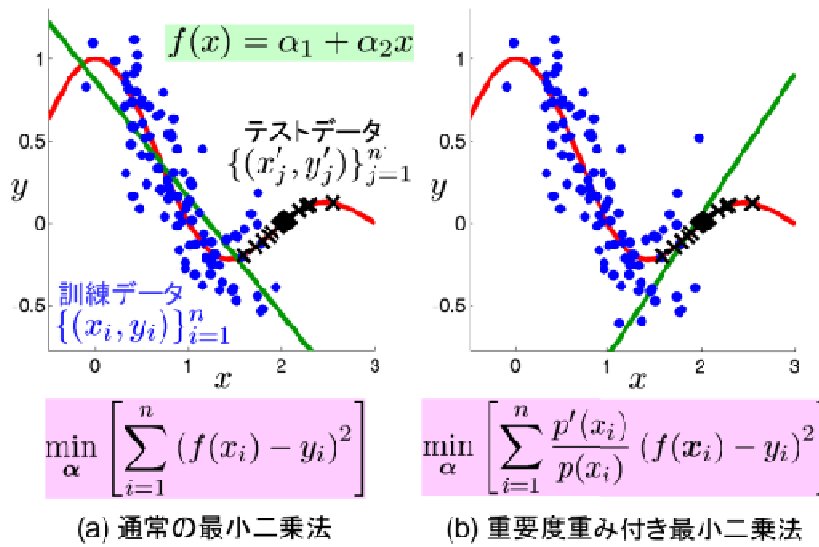


図 4: 重要度重み付き最小二乗法による共変量シフト適応. (a)通常の最小二乗法ではモデルを訓練データに適合させるため、テストデータが訓練データと異なる確率分布に従う場合はテストデータをうまく予想できません. (b)テストデータに近い訓練データに強い重みをつけることにより、テスト出力に適合させることができます.

●確率分布比較

データ集合に含まれる異常値を見つける問題を、異常値検出とよびます。このような入力データだけからの機械学習問題は、前述の教師付き学習と対比して教師なし学習とよばれます。一般に教師なし学習は、教師付き学習と比べてデータ解析の目的があいまいです。異常値検出も例外ではなく、どのようなデータを異常とみなすかを決めないと主観的な議論に陥ってしまいます。しかし異常には様々なパターンが存在し得るため、あらかじめ異常とは何かを厳密に定義することは困難です。

そこで、逆に正常とは何かを定義し、正常でないものを異常とみなすことにします。具体的には、異常を発見したいデータ集合以外に、正常データの集合が与えられると仮定し、もとのデータ集合のうち正常データから外れたものを異常値とみなすことにします。この考え方は、異常値を検出したいデータ集合の確率密度関数と正常データの確率密度関数の比を推定し、この比の値が 1 から大きく離れたデータを異常値とみなすことにより実現できます (図 5)。このような方式に基づく異常値検出は、光学部品の異常検出やローン顧客の審査などに応用されています。

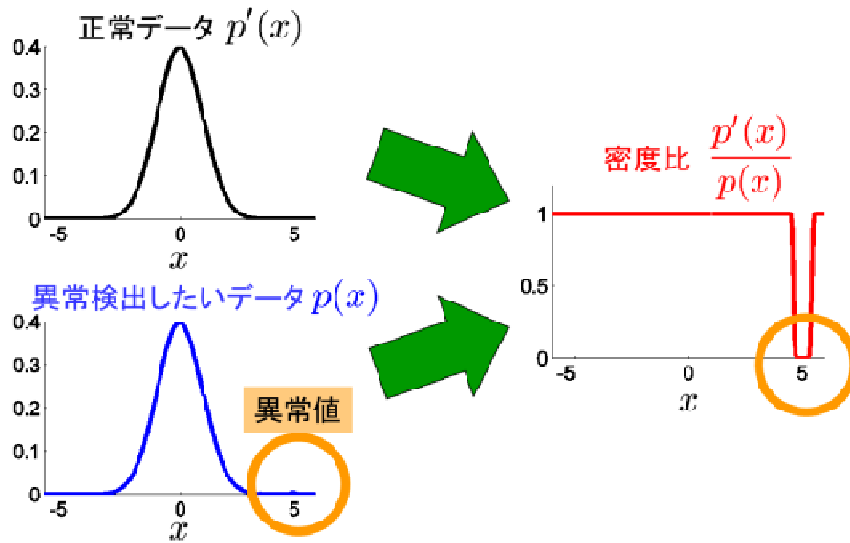


図 5: 密度比に基づく異常値検出. データ中の異常値を直接検出するのは困難だが (左下), 正常データの密度 (左上) との比を取ることで, 異常値が強調され検出が容易になります (右).

異常値検出は二つの確率分布の一点を比較することに対応しますが, 二つの確率分布の全体を比較することも重要です. これは, 二つのデータ集合が同じ確率分布から生成されたかどうかを判定する問題に対応し, 二標本検定とよべます. 二標本検定は, 例えばカルバック距離やピアソン距離など, 二つの確率分布間の距離がある閾値より大きいかどうかを判定することにより実現できます. 二つの確率分布間の距離は, 密度比推定により精度良く推定できます. 二標本検定は, 共変量シフトが起こっているかどうかの判定や, 異なる状況で採取されたデータを合併して処理して良いかどうかの判定などに用いることができます.

また, 過去の時系列データと現在の時系列データが従う確率分布間の距離を推定することにより, 時系列の傾向の変化検出を行うこともできます (図 6). このような変化検出手法は, 生体信号からの状態推定, 画像中の注目領域の抽出, 動画やツイッターからのイベント抽出などに応用されています.

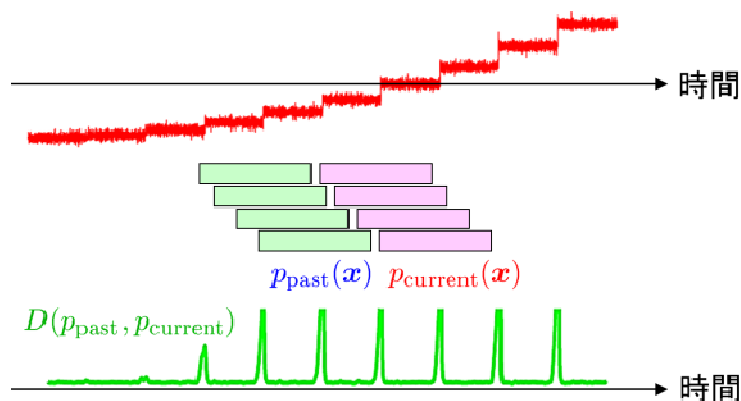


図 6: 密度比に基づく変化検知. 過去の時系列データと現在の時系列データの確率分布間の距離を推定することにより, 時系列の傾向の変化を捉えられます.

●相互情報量推定

入出力データが与えられた時, 入力と出力に依存性があるかどうかを判定することによって, 様々なデータ解析を行なえます. 例えば, 入力ベクトルの一部の要素が出力と独立であることがわかると, そのような要素は教師付き学習においては無視することができます. これは, 出力の予測に役立つ入力変数ベクトルの部分集合を求めることに対応し, 特徴選択とよばれます. 特徴選択によりデータの解釈性が高まるため, 例えば遺伝子や脳波の解析に応用することができます. 一方, 出力の予測の精度を向上させるために, 入力ベクトルを低次元表現に変換することを特徴抽出とよびます. 特徴抽出は, 出力との依存性が最大の低次元表現を求める事により実現できます.

入力データだけが与えられる場合でも, それらと最も依存性が高い出力ラベルを求めることにより, データのクラスタリングを行なえます. 他にも依存性の推定により, ブラインド信号源分離, 異ドメイン間オブジェクト適合, 独立性検定, 因果解析など様々なデータ解析を行うことができます (図 7).

二つの確率変数間の依存性は, それらの同時確率密度から周辺確率密度の積までの距離によって見積ることができます. 例えば, カルバック距離を用いた相互情報量や, ピアソン距離を用いた二乗損失相互情報量がよく用いられます. これらの情報量は, 密度比推定により精度良く推定できます.

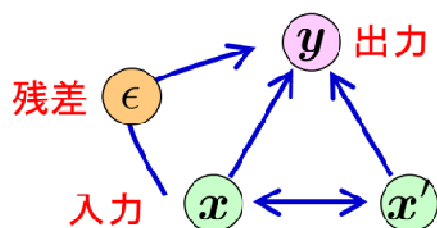


図 7: 相互情報量に基づく独立性判定. 入出力間の独立性判定により, 特徴選択, 特徴抽出, クラスタリングが行なえます. 同様に, 入力間の独立性判定により, 独立成分分析やオブジェクト適合が行なえます. また, 入力と残差との間の独立性を判定することにより, 因果推論が行なえます.

●条件付き確率推定

回帰とよばれる教師付き学習では, 連続値をとる出力変数の, 入力を与えられたもとの条件付き期待値を推定します. しかし, 出力の条件付き分布が多峰性や非対称性を持つときは, 回帰分析では十分な情報が得られないため, 条件付き確率密度そのものを推定することが重要となります (図 8(a)). このような条件付き密度の推定は, データの可視化や移動ロボットの状態遷移確率などに応用できます.

一方, 出力がカテゴリ値を取るとき条件付き確率はカテゴリの事後確率を表すため, これを最大にするカテゴリを選ぶことによってパターン認識を行なえます (図 8(b)). このパターン認識法には, カテゴリの予測だけでなく予測の信頼度も同時に得られるという特徴があり, 顔画像からの年齢予測や加速度センサーからの行動識別などに応用されています.

条件付き確率は，その定義から確率密度比の形で表すことができるため，密度比推定によって精度よく推定できます。

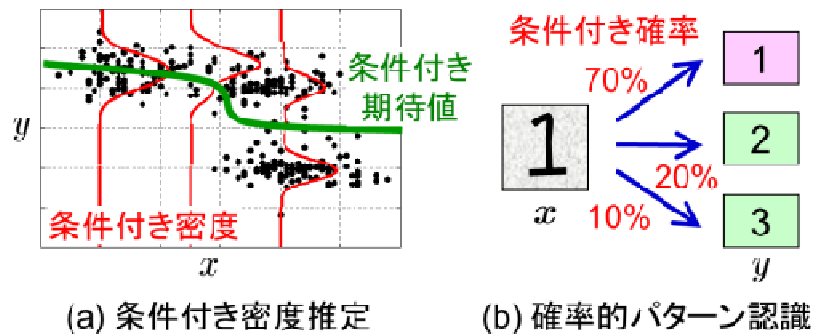


図 8: 条件付き確率推定. (a)出力変数が連続値を取るときは条件付き確率密度の推定に対応し，条件付き期待値を推定する回帰分析の一般化になっています．条件付き確率密度の推定は，出力の条件付き確率分布が多峰性や非対称性を持つときに有用です．(b)出力変数がカテゴリ値を取るときは確率的パターン認識とよばれ，カテゴリの予測だけでなく予測したカテゴリの信頼度も同時に得ることができます．

●まとめ

本稿では，筆者らがテレコム先端技術研究支援センターの研究助成を受けて開発した密度比推定に基づく統計的機械学習技術の概要を紹介しました．密度比推定によって様々な機械学習タスクを統一的に解決できるため，密度比推定の精度や計算効率を更に向上させることにより様々な機械学習アルゴリズムの性能を一挙に改善できます．今後は，密度比推定の基礎技術を更に発展させていくとともに，密度比推定により解決できる新たなデータ解析タスクを開拓し，それらの機械学習技術を様々な実世界問題の解決に活用していくことが期待されます．

密度比推定に関する論文やソフトウェアが，著者のホームページ

<http://sugiyama-www.cs.titech.ac.jp/~sugi/>

からダウンロードできます．興味を持って下さった方は，そちらも合わせてご覧いただけましたら幸いです．

●参考文献

- [1] 元田 浩, 栗田 多喜夫, 樋口 知之, 松本 裕治, 村田 昇 (編) . パターン認識と機械学習 (上) : ベイズ理論による統計的予測, 丸善出版, 2007.
- [2] 元田 浩, 栗田 多喜夫, 樋口 知之, 松本 裕治, 村田 昇 (編) . パターン認識と機械学習 (下) : ベイズ理論による統計的予測, 丸善出版, 2008.
- [3] 杉山 将. 統計的機械学習 : 生成モデルに基づくパターン認識, オーム社, 2009.
- [4] 杉山 将. イラストで学ぶ機械学習:最小二乗法による識別モデル学習を中心に, 講談社, 2013.
- [5] 赤穂 昭太郎. カーネル多変量解析 : 非線形データ解析の新しい展開, 岩波書店, 2008.
- [6] Sugiyama, M., Suzuki, T., & Kanamori, T. Density Ratio Estimation in Machine Learning, Cambridge University Press, 2012.
- [7] 杉山 将. 機械学習入門. オペレーションズ・リサーチ, vol.57, no.7, pp.353-359, 2012.

[8] Sugiyama, M. & Kawanabe, M. Machine Learning in Non-Stationary Environments: Introduction to Covariate Shift Adaptation, MIT Press, 2012.