研 究 室 訪 問

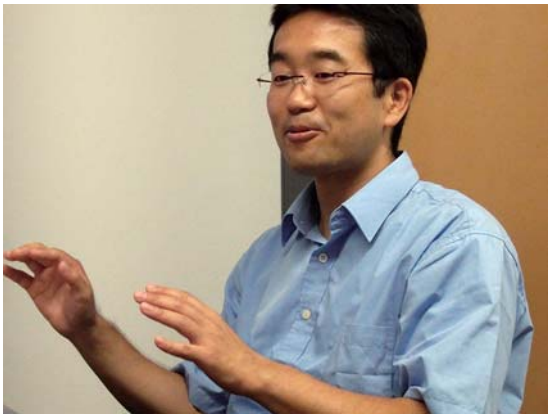
東京大学大学院　新領域創成科学研究科

杉山　将　教授　訪問記

（2007 年度受領者）

7 月 7 日木曜日，梅雨の晴れ間でとても暑い午後でした。当財団の阿草清滋選考委員長（南山大学大学院理工学研究科研究科長，京都大学学術情報メディアセンター客員教授），臼井支朗選考委員（理化学研究所神経情報基盤センター特別顧問，豊橋技術科学大学名誉教授），勅使川原正樹選考委員（オムロン株式会社）らとともに，東京大学の杉山将（まさし）教授の研究室を訪問しました。都会の喧騒からは隔離されているものの，暑いことに変わりはない本郷キャンパスとそこに聳え立つ安田講堂が我々を迎えてくれました。なお杉山先生は，文部科学省が進める AIP プロジェクトの実施を担う研究開発拠点として理化学研究所に設置される「革新知能統合研究センター」のセンター長にご就任され，その準備で大変ご多忙な中で我々の訪問を受け入れていただきました。

○（事務局）まずは，先生のご所属や研究室をご紹介して下さい。



組織が複雑なのですが，大学院は新領域創成科学研究科の複雑理工学専攻と情報理工学系研究科のコンピュータ科学専攻，学部は理学部情報科学科に所属しています。研究室は「杉山・佐藤研究室」ということで，講師の佐藤一誠先生といっしょに運営しています。またロケーションが柏キャンパスと本郷キャンパスに分かれており，特任助教，研究員，博士・修士の学生，その他支援・補佐のスタッフがいます。東京工業大学から 1 年半前に異動してきた関係で修士 2 年生はまだいないのですが，逆に東京工業大学時代の学生もいて，総勢 25 名程度の人員で研究を進めています。この中の 3 分の 1 程度が外国人です。また冬学期になると学部生も配属され，そのうちの何人かは来年修士に上がり，昨今の AI ブームにも乗って，今後ますます増えていきそうです。

○（事務局）先生の技術的な背景をご紹介して下さい。

私は子供のころからプログラミングが好きでプログラマを目指して大学は情報工学科に入学しました。学部の頃，ただ単にプログラムを書いていても面白くないなと思っていた時に，計算量理論の研究室でアルゴリズム解析の卒論を書き，それで理論に興味を持ちました。修士でニューラルネットの研究室に配属され能動学習の理論で論文を書き，博士論文で統計の勉強をしてモデル選択，その後共変量シフトなどを経てその流れで今に至っています。修士 1 年の時からテーマはあまり変わっていないですね。

○（事務局）研究内容はどういったものでしょうか？ また研究へのこだわりは？

研究内容は一言でいえば，“機械学習と統計的データ解析”であり，それにより“データの背後に潜む知識を自動的に学習”することを目標としています。具体的な研究テーマとしては，“学習アルゴリズムの開発”と“学習理論の構築”に分かれます。

“学習アルゴリズムの開発”は，人間がコンピュータの学習を手伝う“教師付き学習”，人間を介さずにコンピュータが学習する“教師なし学習”，エージェントが自発的に意思決定する“強化学習”などに取り組んでいます。また“学習理論の構築”は，これらの技術的背景として重要なもので，確率，統計，最適化などに基づき，機械学習の仕組みを数学的に解明することに取り組んでいます。

さらには，これらのテーマを通じて研究スタイルとしてこだわりたいことが３つあります。一つ目は，テキスト，音声，画像，ロボット，医療，生命など幅広い実データを対象としている中で，特定の応用に依存しない汎用性を常に意識して開発したいと思っています。

二つ目は，理論的な裏付けと実用性を両立したいと考えています。理論のための理論，理論のない応用は研究の対象から排除します。一方で研究者や学生には，確率，統計，最適化などの高度な数理知識を要求するとともに，実用化のために最低限のプログラミングスキルも必須としています。

三つ目は国際性を重視しています。日本はこの領域ではやや遅れているので，海外のレベルやスピードについていくために，研究室内のミーティングやセミナーは全て英語で行い，論文も英語で執筆し，そしてトップレベルの国際会議での論文発表を研究者や学生の目標としています。海外からの留学生が多いのもそのためです。前任の東京工業大学では，博士課程の学生が全て留学生だったこともありました。

○（事務局）先生のご研究における代表的な理論的アプローチをご紹介して下さい。

機械学習には様々なタスクがありますが，最も汎用的なアプローチは“生成的アプローチ”とよばれるものです。これは，データを生成する規則である確率分布を推定すれば，あらゆる機械学習タスクが解決できるというもので，各クラスの確率分布からパターン認識することです。ところが，確率分布の推定は容易でなく，必ずしも高い学習精度が得られるとは限りません。

これに対して，確率分布の推定を行わずに各タスクを直接解く“識別的アプローチ”とよばれる方法もあります。例えば，1960 年代にアイデアが提案され，1990 年代後半にアルゴリズムが確立された手法である“サポートベクトルマシン”では，各クラスのデータ生成分布を推定せずに，パターン認識に必要な決定境界のみを学習します。このように各タスクに特化したアルゴリズムを開発した方が良い性能が出るのですが，様々なタスクに対して個別に研究開発を行うことになり，アルゴリズム考案，理論的性能評価，高速かつメモリ効率の良い実装，エンジニアの技術習得などの面で多大な労力を要します。

そこで我々はこれらの中間的なアプローチとして，全てではなくあるクラスのタスク群に対して，抽象的な量の推定によってデータ解析を行うことに着目しました。抽象的な量としては，確率密度比，確率密度差，距離，情報量，確率密度微分などが考えられます。そして，複数の確率分布を含む多くのタスクを解くためには，それぞれの確率分布を知る必要はなく，抽象的な量の中でも確率密度関数の比，つまり確率密度比が分かれば十分であるという知見に至りました。

バプニックが 1998 年に「ある問題を解くとき，それより一般的な問題を途中段階で解くべきではない」という原理を提唱しました。定理ではないので反例は存在するのですが，これを我々の知見にあてはめると，「確率密度比を推定している時に，個々の確率密度を求めるべきではない」ということになります。

実際に，この確率密度比を推定する方法は，密度比の二乗誤差を最小にする密度比モデルを学習する“最小二乗密度比適合”という問題に落とし込むことができ，多次元の場合では多様体次元削

減効果が得られ，MATLAB ではわずか 3 行のプログラムで書けるものでした。その結果，タスクに特化した“識別的アプローチ”よりもはるかに汎用性が高く，必要な個々の確率分布を求める“生成的アプローチ”よりもはるかに精度の高い方法として多くのタスクに適用できました。しかも，パラメトリックモデルにおいてもノンパラメトリックモデルにおいても，データが多くなると最適な収束率を達成していることが証明できました。

○ **（選考委員）確かにこの理論は非常にすっきりしているとともに，最適性を担保していて素晴らしいですね。具体的な応用事例にはどんなものがあるのでしょうか。**

はい，この確率密度比推定法という理論を抽象化していくつかの機械学習アルゴリズムに展開することにより，様々な事例に応用できます。《以降，詳細な解析手法とともに応用事例をご説明いただきましたが，ここでは詳細を割愛して，アルゴリズムと応用事例の列挙に留めます。》

学習アルゴリズム	汎用的手法	具体的な応用事例
重点サンプリング	共変量シフト，ドメイン適応，多タスク学習	顔画像からの年齢予測，話者認識，テキスト分割，ブレインマシンインタフェース
確率分布間距離推定	二標本検定問題，外れ値検出，変化点検知	製鉄プロセスの異常診断，光学部品の品質検査，ローン顧客の審査，画像中の注目領域抽出，動画からのイベント検出，ツイッターデータ解析
相互情報量推定	独立性検定，変数選択，独立成分分析，次元削減，因果推定，クラスタリング，オブジェクト適合	遺伝子解析，画像認識，音響認識，モーションキャプチャデータの適合，半導体の内部検査，医療画像の位置合わせ，写真の自動レイアウト
条件付き確率推定	可視化，状態遷移推定，確率的パターン認識	ヒューマノイドロボットの制御，加速度データからの行動認識

このように密度比は最小二乗法で精度よく，効率よく推定できるので，多くの学習アルゴリズムが実は“最小二乗密度比適合”で解けることを実証し，上述のようにブレインマシンインタフェース，ロボット制御，音声認識，画像認識，自然言語処理，バイオインフォマティクス，データマイニングなどに応用してきました。一つの理論から多くの論文や特許が生まれ，多くの企業と応用研究を積み重ねてきました。

○ **（事務局）先生は 2007 年度に当財団の研究助成を受けられていますが，それはどのような研究だったのでしょうか？**

当時は，今述べてきた研究と並行して，次元削減の研究をやっていました。「次元の呪い」という言葉がありますが，高次元空間では通常の幾何学的な直感が成り立たないので非常に扱いにくいことから，データの特徴を保持しつつ次元を減らしたいというのが動機です。その頃に貴財団から「複雑な構造を持つ高次元データに対する機械学習」という課題で助成していただきました。

教師なし次元削減としては，元のデータを保存する古典的な“主成分分析（PCA）”と，元のクラス構造を保存する“局所保存射影（LPP）”が，教師付き次元削減としては，元のクラス間の分離性を保存する“フィッシャー判別分析（FDA）”がよく知られています。この助成課題では，FDA に LPP を組み合わせ，類似したデータを自動的にグルーピングする“局所フィッシャー判別

分析（LFDA）”を開発し，その有効性を甲状腺疾患データの可視化によって示しました。この課題を達成した後，LFDA と PCA を組み合わせることにより“半教師付き局所フィッシャー判別分析（SELF）”を開発しました。

LFDA の成果で，2010 年人工知能学会 DMSM 研究会の研究会優秀賞をいただきました。また LFDA の国際会議論文とジャーナル論文は，合わせて 800 回以上引用されており，インパクトの高いものでした。

余談になりますがこの LFDA の論文はどちらかというと小ネタなのですが，理論に関する 50 ページの大作かつ自信作よりも，引用数が多いのです。そんなものなのですね。

○ **（選考委員）基礎理論の論文は数式も多くて，どうしても大作になり難解になりますね。応用している側にとっては，簡単な方がとっつきやすいということでしょうね。ところで，それらの成果がどのように今の研究につながったのでしょうか？**

この時に確立した次元削減の手法は，それと並行して進めていた確率密度比の推定においても生きました。各確率密度から生成されたデータを LFDA で分けるように次元削減すると，実は密度比推定の意味でも有用な次元削減になっていることを示したのです。このアルゴリズムで 2011 年に日本神経回路学会の論文賞をいただきました。

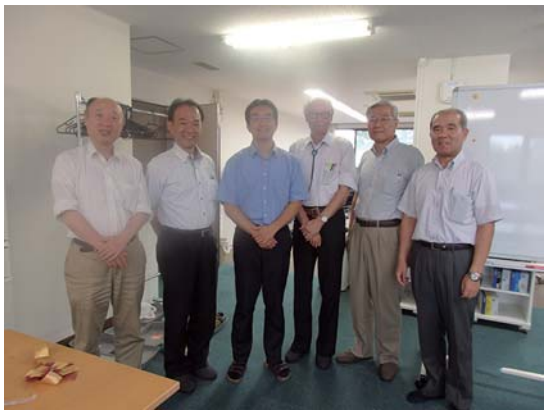
○ **（選考委員）ところで，AI（人工知能）ブームは，コンピュータや通信技術の進歩とともに過去に 20 年から 30 年の周期で到来しています。最初は 1960 年代，2 回目が 1980 年代でエキスパートシステムという言葉がその代表でした。今回はディープラーニング（深層学習）に牽引されて 3 回目のブームとなっています。今の AI ブームについてどのように見ておられますか？**

ディープラーニングは，若い研究者と企業が中心になってアグレッシブにやって，役立つ事例がたくさん出ています。したがって否定するものではありませんが，高次元のデータをそのまま力づくで解析していて，それ以外の問題に対してはまた別のアプローチが必要といったように，汎用性や理論的背景が希薄なように感じます。あんな複雑なモデルでやらなくてももっと簡単なモデルでできる，実はこっちの方が近道だったというようなことを示せればいいと思います。

また，AI という言葉の定義や範囲があいまいで，セキュリティ，IoT，ロボット，脳情報など諸々を AI に含めよう，もしくは AI ブームにあやかろうという傾向があるように見えます。

○ **（選考委員）理化学研究所の「革新知能統合研究センター」センター長として日本の AI をどのように牽引して行こうとお考えでしょうか？**

日本はこの領域に弱いと言われています。世界に通用する人材も少ないです。だからこそ，限られた予算と人材で，選択と集中により，基礎理論を強化して汎用的な手法を確立すべきと考えています。



左から，田中理事，勅使川原選考委員，
杉山先生，臼井選考委員，阿草選考委員長，
菅原常務理事

特に「革新知能統合研究センター」は，もちろん内外との調整は必要ですが，基礎理論の追求と構築，そして汎用的な手法の開発に注力しつつ，他の省庁，研究機関，企業と連携して，応用面で成果が出るような目標設定とそれに向かった運営ができればと思っています。また人材面では，海外から研究者をリクルートする必要もあり，実際に何人かとは交渉中ですが，給与格差の問題もあるので，日本人の若手研究者を育成しながら進め

ていくことになると思います。

○（選考委員）是非とも基礎理論に軸足を置いて，応用でより多くの成果が出ることを期待します。

○（事務局）最後に当財団へのご要望がありましたらお願いします。

私の場合もそうだったのですが，若手研究者の基礎理論研究を支援していただきたいと思います。どうしても，見た目の華やかな応用研究を行なっている方に研究費が集中しがちで，基礎理論研究者の予算獲得機会は限定的です。ぜひ若手の基礎理論研究者も手厚く助成して頂ければと思います．私自身も若いころにサポートしていただき，非常に助かりました。

杉山先生の確固たる研究スタンス（信念と言ってもいいでしょう）に触れ，軸足を基礎理論に置きながら長期的展望に立って，応用領域で成果を次々とあげていらっしゃることをお聞きして，訪問者一同大変感銘を受けました。また今や時の人となっていらっしゃることにも納得しました。そして当財団の 9 年前の助成が，少なからず現在の先生の研究につながっていることも大変うれしく思いました。

記事中の賞以外にも，最近では 2011 年長尾真記念特別賞（情報処理学会），2012 年船井学術賞（財団法人船井情報科学振興財団），2014 年文部科学大臣表彰若手科学者賞と次々と受賞されており，いかに先生の研究が高く評価されているかが分かります。

また先生からいただいた当財団へのご要望については，研究助成の募集案内で「若手研究者による萌芽的な基礎研究活動も歓迎します」と言っていますので，今後も当財団としてその部分にはこだわっていきたいと思います。大変お忙しい中，ありがとうございました。

（リポータ：理事　田中敏文）