

計測指向情報処理技術と情報処理指向計測技術の共進化

Co-evolution of Measurement Oriented Information Processing Technology and Information Processing Oriented Measurement Technology

鷲尾 隆

Takashi WASHIO

大阪大学産業科学研究所第1研究部門（情報・量子科学系） 知能推論研究分野
教授 博士（工学）

Professor, Department of Reasoning for Intelligence, Division of Information and Quantum Sciences,
The Institute of Scientific and Industrial Research,
Osaka University
Doctor of Engineering



IoT社会の到来を受け、極端な条件下で複雑な過程から成る原理を用いる計測技術の必要性が増している。一方、急速な発展を見ている機械学習・統計的推定による情報処理は、不完全で複雑な大量情報から高精度・高信頼な推定を行うことに長けている。このような背景から、計測を指向した情報処理と情報処理を指向した計測技術の体系的な共進化により、IoT社会のニーズを満たす新たな先端的計測技術を実現する可能性が拓けつつある。本稿では、計測を指向した情報処理の研究分野「計測インフォマティクス」で明らかとなった研究論点を幾つか述べ、さらにその原理を情報処理指向の先端的計測デバイス・装置の開発に生かした成果を紹介する。

With the advent of the IoT society, there is an increasing need for measurement techniques that use the principle of complex processes under extreme conditions. On the other hand, rapidly developing information processing using machine learning and statistical estimation is good at making highly accurate and reliable estimation from incomplete and complex large amounts of information. From such a background, the systematic co-evolution of measurement-oriented information processing and information processing-oriented measurement technology is opening up the possibility of realizing new advanced measurement technology that meets the needs of the IoT society. This paper describes some of the research issues that have become clear in the field of measurement-oriented information processing “measurement informatics”, and further introduces the outcomes of utilizing that principle in the development of advanced information processing-oriented measurement devices and equipment.

はじめに

現実世界から必要な情報をデジタル化して収集する計測技術は、情報通信技術や情報処理技術と並んで、今後のIoT社会が必要とするコア技術である。そこでは、従来型センサによる計測技術に留まらず、全く新しい対象を測る計測技術やより高頻度、高精度、高分解能、高信頼、頑健でかつコンパクト、低コストなセンシングを可能とする計測技術など、いわゆる先端的計測技術が多く必要とされている。

このようなニーズの下で、現在、種々の先端的な計測デバイス・装置の研究開発が行われている。これらの多くは、極限の低濃度物質、極限の微小な分子や生体、極限の解像度の画像や物質分布、極限の遠距離物体、極限の高頻度・短時間、極限のコンパクト性など、様々な極限条件での計測・情報処理を目指すものが多く、極端な条件下で複雑な過程から成る計測原理や計算原理を用いる。さらに、測定

対象も、時間・空間の両領域で大きな揺らぎを含み、かつ非常に高次元な自由度を有する複雑なものであることが多い。そのため、得られる測定情報は、大量であるが多くのノイズを含みかつ計測対象に関する不完全な情報しか含まないことが多い。従って、不完全な大量情報からの対象推定や、複数測定情報の統合、事前知識による補完など、十分な計測結果を得るために高度な推定処理が必要とされる。

一方、並行して、高度な数理理論や統計理論、計算理論に基づく機械学習・統計的推定やそれらアルゴリズム原理・技術が急速な発展を見せている。これらは、上記のような不完全な大量情報や複数測定情報、事前知識から高い精度や信頼性を有する推定を行うことに長けている。このような背景から、機械学習・統計的推定・アルゴリズムと先端的な計測デバイス・装置を融合し、最初に述べたIoT社会のニーズを満たす新たな先端的計測技術を実現する可能性

が拓けつつある^[1]。

しかしながらこのような融合研究開発は、これまで個別問題について散発的に行われて来ているに過ぎず、先端計測を指向する機械学習・統計的推定・アルゴリズムの原理や、これら情報処理を前提とした高性能計測実現を指向した新しい計測技術について、体系的な研究はあまり行われて来なかった。IoT社会における計測への厳しい要求を満たすためには、計測を指向した情報処理と情報処理を指向した計測技術の体系的な共進化が求められる。

筆者等は情報処理研究の立場から、計測を指向する機械学習・統計的推定・アルゴリズム原理の研究分野「計測インフォマティクス」の開拓に取り組んでいる。そこでは、先端的な計測デバイス・装置の研究開発に取り組む多くの研究者、技術者と共同し、情報処理を指向した計測技術開発のお手伝いもさせていただいている。本稿では、計測インフォマティクス研究分野で明らかとなって来ている情報処理研究の論点を幾つか述べ、さらにその研究から生まれた原理を情報処理指向の先端的計測デバイス・装置の開発に生かした成果を紹介する。

計測インフォマティクスを巡る状況と研究論点

機械学習・統計的推定・アルゴリズムの原理に関する近年の急速な発展を受けた計測インフォマティクス研究は、まだその緒に就いたばかりである。海外においても、個別課題に関する散発的な研究開発が行われるに留まっており、組織立った体系的な研究活動はない。その原因の1つとして、機械学習・統計的推定・アルゴリズムを含む殆どの人工知能研究が、各種サービスや金融、自然科学、工学を含むあらゆる分野に広く適用可能な汎用性を追求していることが挙げられる。その幅広い適用性が人工知能の大きな社会的インパクトの源泉であり評価されるべきことではあるが、逆に計測など個別分野に固有の問題設定は顧みられず、基礎研究段階から分野の間尺に合わないことが起こっている。もう1つの原因は、世界的にデータサイエンティストが極度に不足する中で、高度な情報処理と先端的な計測装置・デバイスの両方に精通する研究者・技術者が少なく、またそのような人材が育つ環境も未熟なことである。

このようにまだ草創期ではあるものの、計測インフォマティクス研究への取り組みを通じて、情報処理技術を計測に適用する上で、基礎研究レベルから体系的に考えることが必要な幾つかの論点が明らかになって来ている。以下、我々が把握している3つの論点を説明するが、この分野の発展に伴いさらに様々な課題が発見されて行くと予想される。

論点1：分析のための推定と計測のための推定^[1]

機械学習や統計的推定は、与えられた個々のデータに見られる何等かの規則性を明らかにする理論や技術である。これらは自然法則や現実世界の常識は顧みずに、たとえ偏って集められたデータからであっても、データに見られる規則性を導出することを目的としている。この性質は、与えられたデータの範囲では優れた推定を可能にするが、そこから逸脱する問題設定については誤った結果を与えることがある。

たとえば、京都の街中で数十メートル離れたところに犬か狼とおぼしき動物の影を見たとしよう。洛中に狼がいる可能性は低いので、犬だと推定するのが妥当である。標準的な機械学習や統計的推定においても、京都の街中で数十メートル離れたところから犬か狼のような影を撮影したデータを多数集めて高精度に学習すれば、たとえ狼に良く似ている画像を見ても犬だと推定するアルゴリズムが出来上がる。なぜなら、学習用データの中に減多に狼の画像が存在しないので、犬と推定すればほぼ常に正解するというデータの性質を含めて学習するからである。すなわち、個別の画像に基づく計測結果のみではなく、それに狼が減多にいないというデータ全体が示す規則性を加味した分析結果が得られる。これは、長年の経験に裏打ちされた医師が、患者のレントゲン画像の性質に加えて医師が持つ知見を加味して診断結果を導くのに似ている。解きたい問題が、計測装置・デバイスから得られる結果に過去のデータに裏打ちされた経験を加味する分析であるならば、標準的な機械学習や統計的推定の技術を適用すれば良い。

一方で、洛中で収集したデータで学習した推定装置をそのままヒマラヤの森に持ち込むと、犬か狼とおぼしき動物の影のほとんどが狼であるにもかかわらず、相変わらず犬と推定し続けるという困った状況に陥る。計測においてはこれが問題となる。計測では、過去の経験による分析が役立つ状況でも、環境に依らずに常に正しい推定を出力することが求められる。このためには、学習データ全体の性質や規則性は用いずに、個々の画像の特徴のみに基づいて予断なく犬か狼か推定するアルゴリズムを用いる必要がある。既存の機械学習や統計的推定の技術によってこれを得るには、偏りの無いバランスが取れた犬と狼の画像数から成る学習データを用いるなどの工夫が必要となる。しかし、推定したい情報が犬と狼のような単純な選択ではなく複雑な画像やスペクトルであれば、未知のあらゆる可能な画像やスペクトルも含めた学習データを偏り無く準備するという、実行不可能な要求に迫られることになる。従って、多くの複雑な先端的計測のためには、計測を指向する新しい学習や推定原理の研究開発が求められることになる。

上述した分析のための推定と計測のための推定は、古くからそれぞれデータの分布に基づく「最大事後確率(MAP:

Maximum A posteriori Probability) 推定」とデータの分布を参照しない「最尤(ML: Maximum Likelihood)推定」と呼ばれている。深層学習を含め、現在研究開発されている多くの機械学習や統計的推定の原理やアルゴリズムは前者を行うものであり、計測のためには基礎研究レベルから作り直しが必要となる場合が多い。計測を目的とした最尤推定に基づく機械学習や統計的推定に関する研究は、計測インフォマティクスの重要な論点である。

論点2：ベイズ推定^[2,3]

計測や分析においては、上述したデータの分布が示す規則性以外に、我々が予め知っている推定対象の性質を反映させた推定を行う方が望ましい場合がある。たとえば、計測したいたんぱく質の分布が細胞の一部に局在しており、計測画像全体に広がることはないことを予め知っていれば、その事前知識を推定に反映することで、画像全体に広がる観測ノイズの影響を低減しつつ、より妥当なたんぱく質分布推定を行うことが可能となる。このような事前知識を反映する推定をベイズ推定という。

その中でも特に、上記例のように対象が計測結果や分析結果の一部に局在するはずだという知識を反映した推定はスパース推定と呼ばれ、多くの機械学習や統計的推定で用いられている。これ以外にも、たとえば対象が滑らかに存在するはずだ、幾つかの塊になって存在するはずだというような事前知識を反映するベイズ推定手法が提案されている。

しかしながら、我々の多様な事前知識を何でも容易に推定に導入できるわけではない。事前知識を推定に反映するためには、知識を数学的に定式化する必要があり、さらにその数式の容易な計算を可能にするアルゴリズムを発見する必要がある。任意の事前知識の数式化やその効率的計算アルゴリズムが存在する保証はなく、もし存在したとしてもそれらを見つけるには数理やアルゴリズム分野の職人技が要求される。計測や分析によく用いられる我々の事前知識にはどのようなものがあるか、そしてそれらを如何に幅広く容易にベイズ推定に導入するかは、計測インフォマティクスの重要な論点である。

論点3：センサ配置^[4-6]

多くの先端的計測では、極端な条件下において複雑な過程から成る計測原理が用いられる。特に複数のセンシングデバイスを組み合わせて配置することで、所望の計測結果や分析結果を得る場合が多い。この際、幾つのセンサをどう配置すれば、計測結果や分析結果の精度、ロバスト性を最大化できるかという問題は、コストや装置形状などの資源制約下においては極めて重要である。

こうしたセンサ数とその配置問題の多くは、数理的には

劣モジュール関数の最大化問題として定式化可能であることが知られている。例えばある建物において、幾つの監視カメラをどこに配置すれば、より広く建物内の面積を監視するシステムを構築できるかという問題を考える。この場合の監視面積は、カメラを設置可能なすべての候補地点からどの場所を何か所選択して設置するかによって決まる。すなわち監視面積は、設置地点集合を入力とする関数であり、集合関数と呼ばれる。1個のカメラのみでは監視できる面積は限られ、カメラを追加して最適に配置すると監視面積は広がる。しかし、カメラを増やしすぎるとカバー範囲が重複して、あまり監視面積は増加しなくなる。すなわち、監視面積は設置するカメラ台数に対して次第に増加しなくなる、いわゆる収穫逨減の法則に従う。一般に、収穫逨減の法則に従う集合関数を劣モジュール関数という。監視面積のような劣モジュール関数の出力を最大にするセンサ設置地点集合を見つける問題を、劣モジュール関数最大化問題と呼ぶ。

監視エリアの広さに限らず多くの計測や分析問題において、精度やロバスト性など様々な性能指標が、劣モジュール関数であることが知られている。従ってこの分野において、劣モジュール関数最大化問題を解くことは重要である。しかしながら、最適なセンサ配置を決める上で、沢山の設置候補地点から監視面積を最大にする設置地点組み合わせ集合を選ぶには、膨大な組み合わせを試す必要がある。計算科学の分野では、厳密に劣モジュール関数を最大化する集合を探すことはNP困難問題という非常に計算の手間がかかる問題であることが知られており、実用的な速さで大規模な問題を高精度に解くアルゴリズムはまだ研究途上にある。このような実用的アルゴリズムの探求も、計測インフォマティクスの重要な論点である。

研究開発事例

ここでは、上述した計測インフォマティクスの中で、特に論点1の分析のための推定と計測のための推定を適用した先端的計測技術について、それぞれ情報処理研究者である筆者と計測技術研究者との共同研究事例を紹介する。

事例1：ナノポアによる極微粒子種別分析^[7-10]

この事例は、筆者が所属する大阪大学：産業科学研究所：産業科学ナノテクノロジーセンター：バイオナノテクノロジー研究分野において、ナノ計測技術研究に取り組む谷口正輝教授等との共同研究である。

ナノポアは、Figure 1に示すように電解質溶媒で満たされたナノからマイクロスケールのホールである。この上下に電極を配置して定電圧を印加すると、ナノポアを通じてイオン電流が流れる。そこに、電界や圧力などの作用によって上側から極微粒子を導き下側に向けてナノポアを通過さ

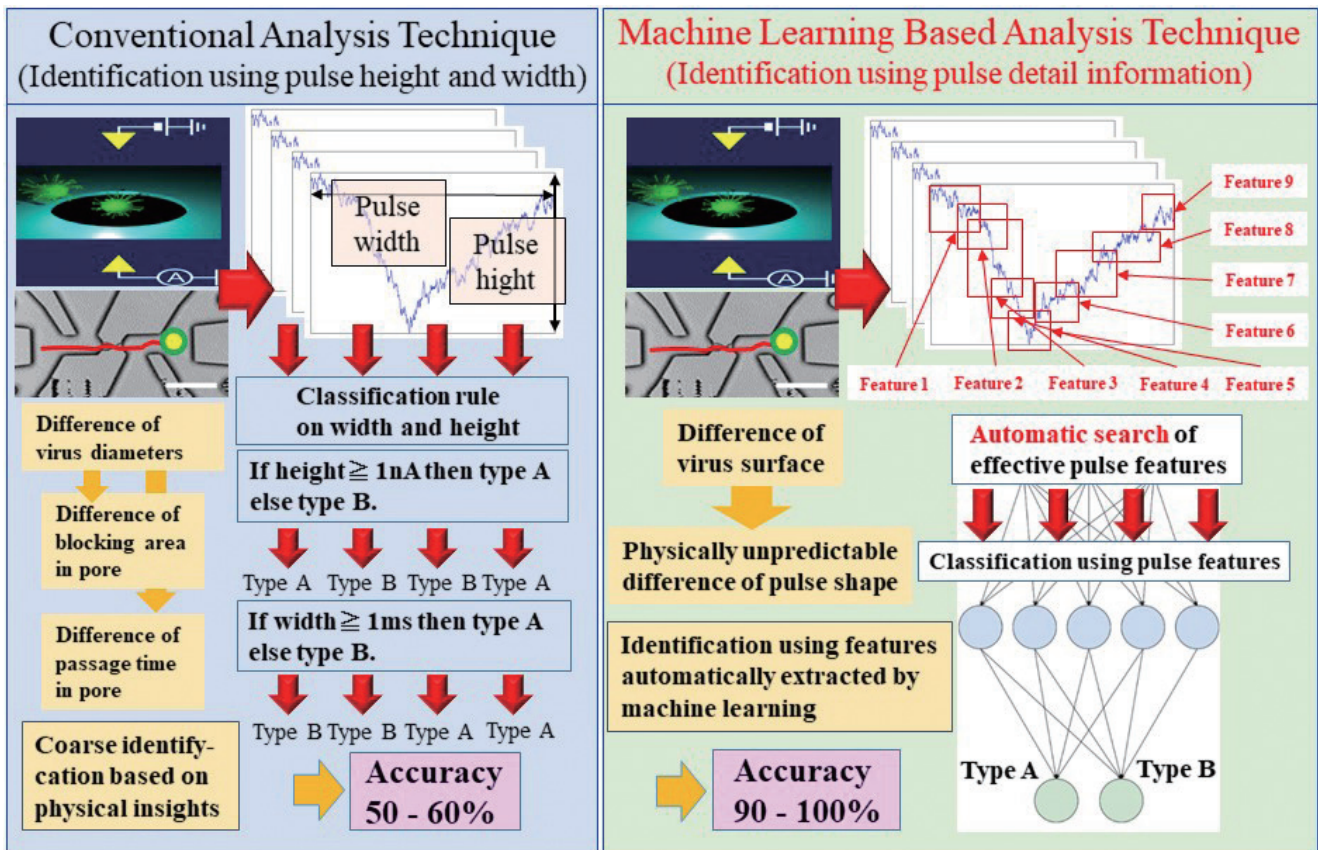


Figure 1 Difference between conventional nanopore pulse identification and nanopore identification using machine learning.

せると、通過時に粒子が穴を一時的に部分閉塞するためにイオン電流が下降し、減少電流パルスが得られる。このパルス形状には通過する極微粒子の大きさや通過速度のみならず、形状や表面状態など様々な情報が反映されている。そこで、ナノポアから出力されるパルス形状によって、個々に通過した極微粒子の種類や性質を分析できる。

従来のナノポア出力パルスによる極微粒子識別においては、Figure 1の左側に示すように各パルスの波高や波幅を基に識別を行っていた。粒子径が大きければナノポアの閉塞面積が大きいので、パルス波高は大きくなると想定される。粒子がナノポアを通過しやすい大きさや質量、表面状態を持てば通過速度が速いので、パルス幅は小さくなると想定される。従来はこのような物理的知見に基づき、パルスの波高や波幅の閾値判定によって、粒子種類を識別しようとした。しかし、同じ種類の粒子でも大きさや状態に個体差があり、必ずしも波高や波幅に種類の違いが明確に反映されるとは限らない。さらに計測される電流パルスにはノイズも多く載っている。そのため、かけ離れた大きさや性質の粒子同士を識別する場合を除いて、パルスの波高や波幅だけに頼っては十分な識別精度が得られないことが多かった。

そこで本共同研究ではFigure 1の右側に示すように、パルス波形の特徴を特徴量として指標化した。特徴量としては、従来の波高、波幅に加えて、短冊状に粗い時刻刻みでパル

ス波高を粗視化したもの、パルスの尖り具合や偏り具合を表す様々な指標、さらに機械学習技術によって自動生成した指標も用いた。識別対象とする極微粒子の種類毎にパルス波形を計測して特徴量化した学習用データを収集し、これに機械学習アルゴリズムを適用して、粒子種類を識別する分類器を得た。機械学習アルゴリズムには、対象とする粒子種類の識別に有効な特徴を自動選択するアルゴリズムを組み込み、粒子の個体差や計測ノイズにロバストな識別を行えるようにした。ここでは現実に各種の極微粒子が検体に現れる頻度に基づいて学習用データを準備したので、得られた分類器はパルス波形特徴のみならず各粒子種類の出現頻度も反映した識別を行う。すなわち、現実のデータの分布を考慮した分析による種類識別を行う。

この技術を大腸菌と枯草菌、表皮ブドウ球菌と黄色ブドウ球菌といった性質や形状が似通った細菌ペアの識別に適用したところ、従来技術では50~60%の識別精度しか得られないのに対し、90~100%の十分な識別精度を達成した。さらにA型、B型、A垂型の3タイプのインフルエンザウイルスの識別に適用したところ、従来技術では同じく50%以下の識別精度しか得られないのに対し、72%程度の識別精度を得た。個々のパルスでは十分な精度とは言えないが、別に複数回のパルス計測に基づいて識別するアルゴリズムを構築して検証したところ、99%以上の識別精度を達成した。

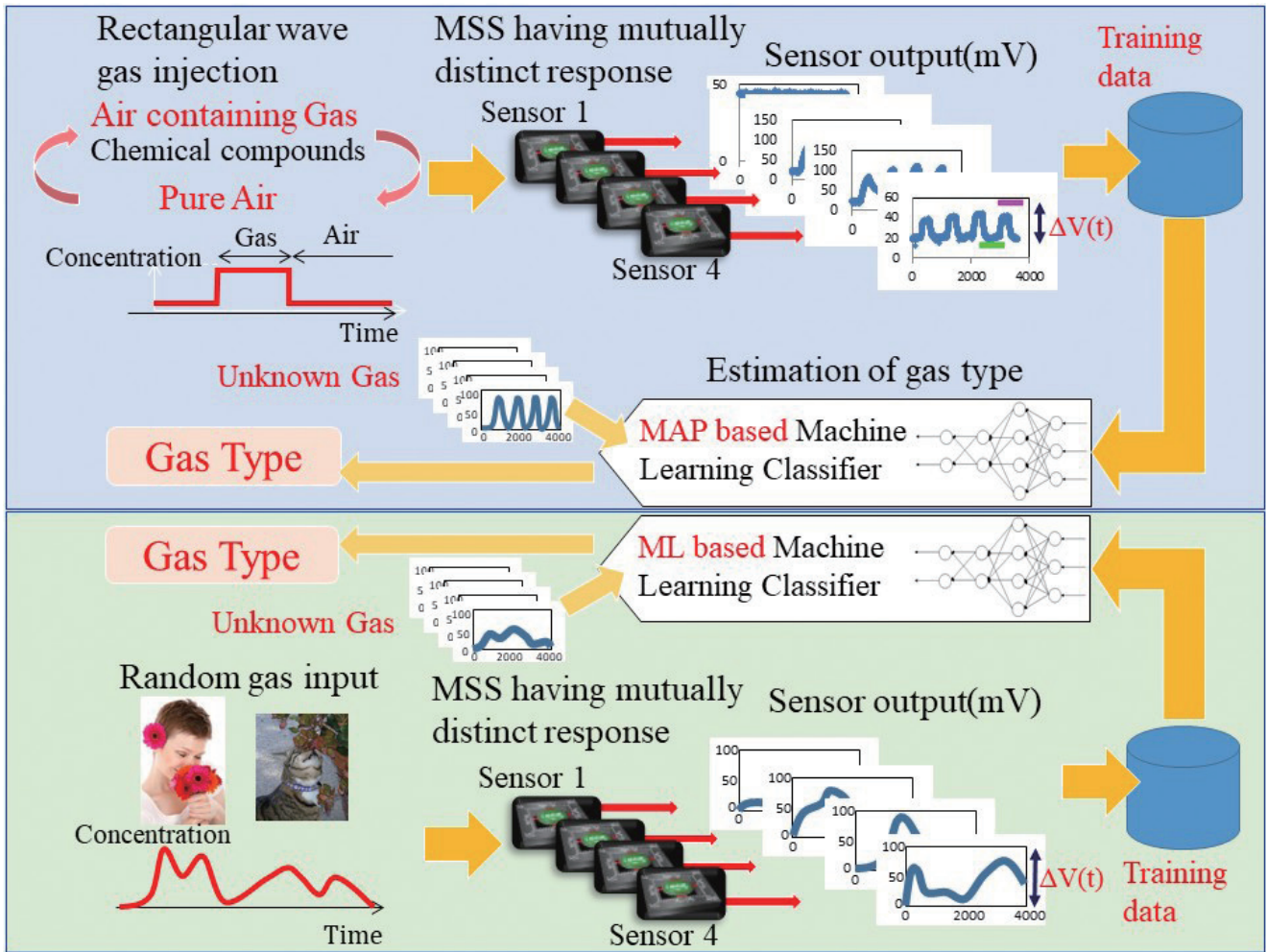


Figure 2 Difference between olfactory sensor by conventional machine learning and olfactory sensor by novel machine learning performing maximum likelihood estimation.

事例2：MSSセンサによる超口バスト嗅覚計測^[11, 12]

この事例は、国立研究開発法人物質・材料研究機構：国際ナノアーキテククス研究拠点：ナノシステム分野：ナノメカニカルセンシンググループにおいて、嗅覚センサ技術研究に取り組む吉川元起グループリーダー等との共同研究である。

MSS嗅覚センサは、Figure 2に示すように複数の異なるMSSセンサから成る。各センサは、空気中に含まれる微量のガス化合物分子の吸脱着により動的に伸縮する薄膜が生じる応力を、電圧に変換するデバイスである。従って、空气中で同一の濃度変化に従う同一ガス化合物分子に暴露されても、薄膜の性質が異なるセンサからは異なる電圧波形が出力される。MSS嗅覚センサは、異なる性質の薄膜を実装することで互いに異なる応答特性を持たせたMSSセンサで構成されており、これらの出力電圧波形の組み合わせパターンに基づいてニオイの元である空气中的の微量のガス化合物分子を識別する。MSS嗅覚センサのセンサ部はMEMS技術により作られており、全体で数ミリ角程度にコンパクト化されている。

従来の嗅覚センサでは、Figure 2の上段に示すように入

にマスフローコントローラやポンプを配し、入力される空気流量やその中のガス化合物分子濃度を矩形波状に制御してセンサ部に導入し、繰り返し矩形波入力に対する各センサ応答電圧波形を計測する。入力波形が完全に制御されているので、各センサの出力電圧波形はガス化合物分子に対する各薄膜の吸脱着の特性を表す。この組み合わせはガス化合物分子の種類毎に異なるので、予め収集したデータを用いて通常の機械学習アルゴリズムによって最大事後確率(MAP)推定を行う分類器を学習する。未知のガス化合物分子の計測電圧波形が与えられると、この分類器によってその種類を推定する。これは論点1で述べた分析のための推定を行う分類器であり、矩形波入力に対する電圧波形組み合わせ入力に対して最も高い精度を与えるように学習されている。京都市内のデータで訓練した分類器をヒマラヤの森に持って行くと役に立たないように、それ以外の波形入力条件に関しては、電圧波形やその組み合わせ分布が全く異なるので推定精度は著しく劣化する。そのため入力の流量制御を欠くことはできず、センサ部は数ミリ角程度のコンパクトなものであるにもかかわらず、装置全体としては大きくなってしま

これに対して本共同研究ではFigure 2の下段に示すよう

に、新たに開発した最尤(ML)推定を行う分類器を得る機械学習アルゴリズムを開発した。詳細は省くが、MSS嗅覚センサから出力される電圧波形組み合わせから、入力波形に依存せずガス化合物分子の種類にのみ依存する特徴量を抽出する学習を行うことで、入力される空気流量やその中のガス化合物分子濃度の波形の影響を受けないでガス化合物分子の種類を推定する分類器を学習した。これは論点1で述べた計測のための推定を行う分類器であり、これによってマスフローコントローラやポンプによる入力の流量制御を不要とした。その結果、数ミリ角程度のMSS嗅覚センサ本体部のみでニオイ識別を行うセンサを開発した。

このMSS嗅覚センサを、エタノール、水、ヘプタン、エチルアセテートと言った単一のガス化合物を入れたビーカーの上部に手でかざして軽く振ったところ、99.6%の精度で種類を識別できた。また、ローズマリー、レッドチリペッパー、ガーリックといった芳香性の混合化合物を入れたビーカーについて同様な計測を行った結果、同じく89%の識別精度を得た。これに対して、MSS嗅覚センサ本体のみを用いて従来の機械学習で識別を試みたところ、実用に耐えない非常に低い精度しか得られなかった。このように、計測問題に適した機械学習アルゴリズムの開発によって、デバイスや装置の開発では達成できない超コンパクトで超ロバストな計測を実現することができた。

おわりに

IoT社会の進展に伴い、計測技術への要求はますます高度化し、それに伴って先端計測にはより高度な情報処理と計測デバイス・装置技術の融合が求められていくと思われる。そのためには、情報処理と計測デバイス・装置の基礎原理から、それぞれ計測と情報処理を前提とした体系的な研究開発が重要になる。本稿では、情報処理サイドから計測指向情報処理を体系的に研究することを目的とした計測インフォマティクスとその幾つか重要な論点を述べ、さらにその一部の成果を適用した先端計測技術研究開発の実例を紹介した。計測インフォマティクス研究の進展に伴い、今後様々な先端計測分野での成果が期待されると共に、さらなる研究論点が見いだされて行くと思われ。

筆者等は、世界的にも遅れている計測インフォマティクス研究の促進、研究者・技術者間の情報共有と人材育成を目指して、2018年度から(一社)人工知能学会の下に「計測インフォマティクス研究会」^[13]を設立し、年2回の研究会を開催している。人工知能学会非会員でも自由に参加・聴講可能なので、興味のある方の参加を期したい。

参考文献

- [1] T. Washio, G. Imamura and G. Yoshikawa. "Machine learning independent of population distributions for measurement". Proc. DSAA2017: 4th IEEE International Conference on Data Science and Advanced Analytics, Tokyo (2017), DOI: 10.1109/DSAA.2017.28.
- [2] Y. Nakanishi-Ohno, T. Obuchi, M. Okada and Y. Kabashima. "Sparse approximation based on a random overcomplete basis". J. Statistical Mechanics: Theory and Experiment(2016) 063302.
- [3] T. Obuchi, Y. Nakanishi-Ohno, M. Okada and Y. Kabashima. "Statistical mechanical analysis of sparse linear regression as a variable selection problem". J. Statistical Mechanics: Theory and Experiment(2018)103401, 1-41.
- [4] Y. Kawahara, K. Nagano, K. Tsuda and J. Bilmes. "Submodularity cuts and applications". Advances in Neural Information Processing Systems (2009)22, 916-924(Proc. NIPS2009).
- [5] Yoshinobu Kawahara and Takashi Washio. "Prismatic algorithm for discrete D.C. programming problem". Advances in Neural Information Processing Systems (2011) 24, 2106-2114(Proc. NIPS2011).
- [6] 河原吉伸, 永野清仁. "劣モジュラ最適化と機械学習". 機械学習プロフェッショナルシリーズ, 講談社サイエンティフィック (2015).
- [7] M. Tsutsui, Y. He, K. Yokota, A. Arima, S. Hongo, M. Taniguchi, T. Washio and T. Kawai. "Particle trajectory-dependent ionic current blockade in low-aspect-ratio pores". ACS Nano, American Chemical Society(2016)10[1], 803-809.
- [8] M. Tsutsui, T. Yoshida, K. Yokota, H. Yasaki, T. Yasui, A. Arima, W. Tonomura, K. Nagashima, T. Yanagida, N. Kaji, M. Taniguchi, T. Washio, Y. Baba and T. Kawai. "Discriminating single-bacterial shape using low-aspect-ratio pores". Scientific Reports(2017)7(1)17371.
- [9] M. Tsutsui, M. Tanaka, T. Marui, K. Yokota, T. Yoshida, A. Arima, W. Tonomura, M. Taniguchi, T. Washio, M. Okochi and T. Kawai. "Identification of individual bacterial cells through the intermolecular interactions with peptide-functionalized solid-state pores". Analytical Chemistry(2018) 90, 1511-1515.
- [10] A. Arima, M. Tsutsui, I. H. Harlisa, T. Yoshida, M. Tanaka, K. Yokota, W. Tonomura, M. Taniguchi, M. Okochi, T. Washio and T. Kawai. "Selective detections of single viruses using solid-state nanopores". Scientific Reports, (2018)8, 16305.
- [11] G. Imamura, G. Yoshikawa and T. Washio. "Development of machine learning models for gas identification based on transfer functions". Proc. 17th International Meeting on Chemical Sensors, Vienna(2018)AR1.1.
- [12] G. Imamura, K. Shiba, G. Yoshikawa and T. Washio. "Free-hand gas identification based on transfer function ratios without gas flow control". Scientific Reports(2019)9, 9768.
- [13] <http://www.ar.sanken.osaka-u.ac.jp/SIG-MEI/>