

## AIは医療分野にどのような変化をもたらすか

株式会社Preferred Networks 取締役副社長 岡野原 大輔  
Preferred Networks America, Inc. 最高執行責任者 大田 信行

### 現在起こりつつある、二つの大きな革命

今、二つの異なる業種、ITとバイオにおいて、大きな革命が起こりつつある。その一つが人工知能(AI)、特に深層学習(ディープラーニング)と呼ばれる技術の登場である。ディープラーニングは人間の脳を作っている神経経路の基本部位であるニューロンとシナプスの働きにヒントを得て作り上げられたニューラルネットと呼ばれる数学モデルである。そのニューラルネットを多層に、そして複雑な経路として積み上げ、大量のビッグデータを使い学習させることにより、これまで専門家による分析や匠の経験に頼っていた特徴抽出やパターン認識・分析を自動的に行うことが可能になってきた。今やディープラーニングは様々な分野で使われ始め、最近では自動運転技術や、囲碁の世界チャンピオンを打ち破った人工知能AlphaGoも、その中心技術にディープラーニングが使われている。この産業革命を引き起こすと言われているディープラーニング技術は、これまでの理論上の革新技术から我々の生活の基盤技術として実用化されていくであろう。もう一つの革命は、バイオテクノロジーの分野で起こっている、\$1000ゲノムシーケンサーとゲノム編集の登場である。一人一人のゲノムを読み取り、病気の原因を見つけ、ゲノム編集を使って疾患ゲノムを改変することにより治療を施すという、SFの世界でしかありえなかった未来が現実になる可能性が出てきた。この二つの異なる分野で、我々の生活を根本から変える科学技術が、人類の歴史上ほぼ同時

期に登場したことは驚くべきことである。そして今まさにその二つの革命を融合させて生命の仕組みを理解し、未来の医療へと繋げていく動きが世界中で始まっている。

### プリファードネットワークスの取り組み

こうした状況の中、弊社Preferred Networks(PFN)はディープラーニングを始めとする人工知能技術の開発と産業への実用化に向けて、3つの領域、交通システム、製造業、そしてバイオ・ヘルスケアに特化して事業展開を行っている。バイオ・ヘルスケアではがん領域に特化し、昨年11月には国立がん研究センター、産業技術総合研究所と協力し、「人工知能(AI)を活用した統合的がん医療システム開発プロジェクト」を開始した。(写真参照)



左から、筆者たち：Preferred Networks America, Inc. 最高執行責任者大田信行、Preferred Networks取締役副社長岡野原大輔、そして、国立がん研究センター研究所がん分子修飾制御学分野長(研究代表) 浜本隆二、国立がん研究センター研究所長間野博行、産業技術総合研究所人工知能研究センター長辻井潤一、産業技術総合研究所人工知能研究センター機械学習研究チーム長瀬々潤(敬称略)

このプロジェクトでは、「国立がん研究センターに蓄積されている膨大な罹患者の詳細な臨床情報やマルチオミックスデータ（ゲノム、エピゲノム、画像情報および血液など網羅的な生体分子情報）、さらに疫学データと文献情報を、AI技術を利用して統合的に解析し、日本人のがん罹患者個々人に最適化された医療（Precision Medicine）の提供を目指した革新的がん医療システム開発」[1]を目指している。

## なぜAIがバイオに必要なのか？

それでは、今なぜAIとバイオを結びつけることが必要なのだろうか？

簡単に言ってしまうえば、本来、生体メカニズムはとても複雑であり、人間がすべての生命現象を正確に把握するのが不可能であるということである。20世紀の中頃、ワトソン・クリックが遺伝子情報としてのDNAの3次構造を提案し、その直後、クリックがセントラルドグマと言われる、DNAからRNA、そしてタンパク質へと情報が伝わっていく生命を司る基本ドグマが提唱された。半世紀以上前に出された、この基本概念は未だに正しいのであるが、その後、様々なマイナー修正がなされ、ここ最近10年ほどでその全体図は大きく書き換えられた（図1）。

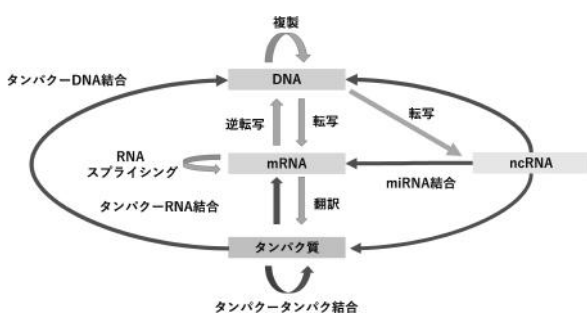


図1 セントラルドグマの新概念図

セントラルドグマの概念は、近年大きく書き換えられた。その複雑な生命の仕組みを理解すること、それら全体の流れや相互関係性を正しく理解していくためには、AI技術を利用せざるをえないであろう。淡線の流れ以上に、濃線の流れは今後新しい手法によりデータが爆発的に増えていくことが予想される。

以前は、ヒトゲノムのうち約30%がRNAに転写され、タンパク質へ翻訳されるのはわずか1～2%にすぎないと考えられており、ゲノムの98%はジャンクDNAと呼ばれていた。しかしながら、最近のシーケンサー技術の発展により、タンパク質をコーディングしていないが発現調整などの機能を持つnon-coding RNA (ncRNA) が大量に見つかり、タンパク質をコーディングするmRNA (約47%) よりもncRNA (約53%) がより多く存在していることが明らかになった。生命機能の中枢はタンパク質が担っていると考えられていた生命科学の常識を覆し、RNA新大陸 (new RNA world) とも呼ばれる新たな概念が生まれてきており、mRNAばかりでなくncRNAが生命を理解する上でも、DNAやタンパク質と同様に重要であることがわかってきている。さらには、mRNA-miRNA、protein-RNA、protein-protein、protein-DNAなどのインタラクションの様々な現象も理解する必要が出てきた。これらのセントラルドグマの新概念ばかりでなく、次世代シーケンサーや、その後の技術発展により低価格化が進み、\$1000ホールゲノムばかりでなく\$20ゲノムも言われ始め、シングルセルシーケンス、RNA発現解析、DNAチップなどの大量のゲノム関連データが指数関数的に増え続けている。しかしながら、遺伝子の改変 (genotype) がどのような表現型 (phenotype) と関連付けられるか、すなわち疾患などとの相関関係があるかを見いだすことは、上に述べた生命機能の根本的な複雑さ故、解明することは非常に難しい。これは、genotype-phenotype gapと呼ばれる生命科学の根本問題である。Precision Medicineと呼ばれる個別化医療を目指す場合、一人当たり約31億配列のゲノム、何十万と言われる生体物質、何千と言われる異なる種類の細胞、そして個人個人の違いを生み出しているSNP (Single Nucleotide Polymorphism) やSNV (Single Nucleotide Variation) を

考慮せねばならない。これまで主に全ゲノム関連解析 (GWAS) と呼ばれる手法により多くの疾患原因となるSNPが同定されてきたが、それらのSNPの多くは疾患感受性にそれほど大きな影響を与えていないことも明らかになっておりGWASの限界を示している。最近では、夢の技術であった、自由にゲノム編集できるCRIPER-Cas 9が登場し、ゲノム編集革命の真っ只中である。ゲノム編集技術を利用し、ハイスループットでゲノム上の全遺伝子の影響を研究することも始まっているが、生命科学における実験データはノイズも多く、大量のデータ以上に、本当に意義のあるデータを見だし、これらを人間が統合的に取り扱い、そこから有意義な知見 (特徴抽出) を得ることはまさに至難の技である。大量の生命科学の異なるデータを統合的に組み合わせ分析し、疾患の原因などを突き止めて、診断、治療、創薬の分野に応用し、我々のQOL (Quality of Life: 生活の質) の向上を進めるには、好むと好まざるに関わらず、AI技術を最大限利用せざるをえない時代になってきているのである。

### ディープラーニングの登場と衝撃

2012年にジェフリー・ヒントン率いるトロント大学チームが、一般画像の認識率を競うコンテスト、ILSVRCにおいてディープラーニングという手法を使い、従来手法に比べて劇的な進歩を披露したことが機械学習の研究者らに衝撃を与えた。その後、このディープラーニングという手法は瞬く間に広がり、画像認識や音声認識などの分野に活用されてきたが、実は、その陰でディープラーニングは製薬業界を揺るがす大衝撃を与えていた。同年、製薬大手のメルク (Merck) がQSAR (定量的構造活性相関) のデータマイニングコンペティションを主催したのだが、同じくトロント大のグループがディープラーニングベースの手法を用いて勝利したのである。与えられたコンテ

スの課題は約15万という化合物 (フィンガープリントデータ) とその活性を調べた実験データを学習させて、その後、新規化合物がどのような疾患 (もしくは分子ターゲット) にどのような活性を持つかを予測することであった。このコンテストには、様々な製薬企業の創薬部門や大学のバイオインフォマティクスのグループが参加し競争したのであるが、トロント大グループは、そのような生物学、生化学、創薬関係の専門家をあえてチームに入れず、情報工学の人間だけで勝利したのである。そこには、ただコンテストで勝利するのではなく、ディープラーニングという手法を用いて、大量のデータからその分野の専門家がいなくても、機械が自動的にその特徴を学習、抽出することができることを示したいとの意図があったからであろう。彼らは優勝することにより、見事にこの仮説を実証することに成功した。さらに、このコンテストにおいて同時に、バイオ・ヘルスケアへ応用可能なディープラーニングが持つ大きな利点が浮き彫りになった。それはマルチモーダル、マルチタスクと呼ばれる特徴である。(図2)

マルチモーダルというのは、一つのニューラルネットワークに様々な異なるデータを取り込み学習でき

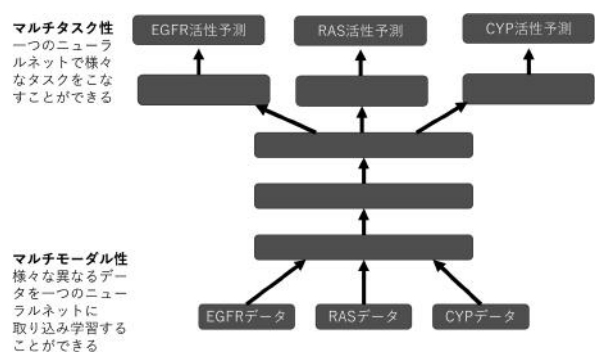


図2 ディープラーニングのマルチモーダル、マルチタスク性  
ディープラーニングは、一つのニューラルネットワークで様々なデータを取り込み (マルチモーダル)、様々なタスクを行えるように学習することが可能である。この図は化合物の活性化予測の簡単な例として示したものである。マルチモーダル、マルチタスクを行うためには、必ずしもこのアーキテクチャーである必要はない。様々な手法が提案されている。

る能力のことである。マルチタスクというのは、一つのニューラルネットで（学習モデル獲得後）たくさんの異なる仕事を行うことのできる能力である。マルチモーダル、マルチタスクというディープラーニングの特徴をこのQSAR予測の例で説明する。例えば、EGFRをターゲットとするがん治療薬を探索するために、化合物とEGFRの活性化データを学習させるだけでなく、一見異なるRasの活性化データや様々なキナーゼの活性化データ等を同時に学習させると、驚くべきことにディープラーニングのEGFRの活性化予測の精度が、EGFRの活性化データ単独で学習した時よりも向上するのである。このようにマルチモーダルでニューラルネットを学習させたものに、新規化合物のデータを入れると、EGFRのみならず、様々なターゲットに対する活性化予測が同時にできるのである。これがディープラーニングの持つマルチタスク性である。このような性能向上がおきるのは、異なるタスクにおいても予測に有効な特徴が共通しており、異なるタスクの学習データも利用できるからである。さらには、多くのタスクに有効な特徴は重要だという事前知識も暗に利用できるからである。特にマルチタスクが利用可能な例としては、がん治療分子標的薬のキナーゼに対する活性化予測である。がんに有効なキナーゼ阻害剤は実は一つだけではなく、幾つかのキナーゼを同時に阻害することでがんに有効な場合が多いことが知られている。一方で、多くのキナーゼは、シグナル伝達経路のキープレイヤーとして大きな役割を果たしており、あまりにも多くの異なるキナーゼを阻害することは重篤な副作用を引き起こし、候補薬にならない。ディープラーニングのマルチタスク性を使えば、一つの候補薬について、様々なターゲットについての活性化状況、例えば副作用の原因の一つである、CYP450の活性を同時に予測することができる。より多くのデータが集まれば、このようなディープラーニングの創薬

への応用が可能なのである。ここで注意しておきたい点は、学習させるためにすべての化合物がすべてのターゲットへの活性化データを用意する必要がないことである。あるデータセットは、100の化合物を使った肺がんのターゲットの活性化データ、別のセットは、1万の化合物での糖尿病の活性化データなど、多くの化合物と多くのターゲットでデータが欠けている状況でも、ディープラーニングは学習し、自動的に特徴抽出して、欠けている活性化データを埋めることができる。この衝撃の後、メルクをはじめ、様々な欧米製薬企業がディープラーニングの凄さに気づき、すでに多くの創薬実証実験を始めている。弊社もすぐに2014年には東工大のスーパーコンピュータTSUBAMEを用いて実証実験を行い、大量の活性化データ（PubChemデータ）をディープラーニングにより蒸留分散学習することにより、トロント大グループ以上の予測精度を上げることに成功している [2]。

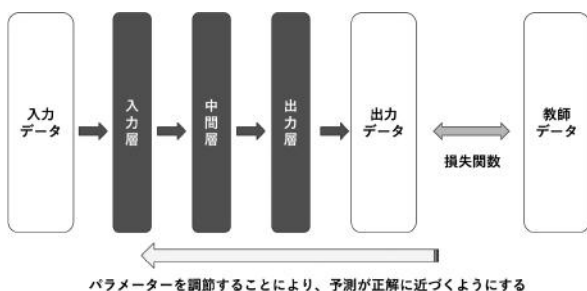
## ディープラーニングの進歩とバイオへの応用、そして展望

2012年以降、ディープラーニングは飛躍的な進歩を続け、その進化のスピードは凄まじく、毎日のように新規技術がどこかで生まれている。それらすべてのディープラーニング動向をここではカバーできないが、現在バイオ・ヘルスケアでよく使われている手法や、将来ゲノム分析などに有望と思われる新規手法を主に紹介したい。まず、ディープラーニングというのは、機械学習の技法の一つであり、その機械学習は大きく二つの学習法、教師あり学習（半教師ありも含む）と、教師なし学習に分類される。これまでバイオ・ヘルスケアにおいてよく用いられるのは上の活性化予測のような正解データを利用する、教師あり学習である。フィードフォワードニューラルネット（Feedforward Neural Networks）と呼ばれる、多層のレイヤーからなるニューラルネットに実験

データ（例：患者のゲノムデータ）を入力層に入れ、下層から順に計算をしながら結果を伝搬させ出力層で予測を行う（例：がんの種類を予測する）。学習の際には、出力値と正解データを比べ、予測と正解がどのようにずれているのかを調べ（損失関数と呼ばれるものを準備し）、このずれを小さくするようにニューラルネットが持つ大量のパラメータ（重みとバイアス）を少しずつ更新することによって（すなわち学習）、より予測精度が上がるようにするのである。（図3）

この作業を、ビッグデータを用いて何度も行うことで、ディープラーニングはデータの中に潜むパターンや特徴を自動抽出することができるのである。教師あり学習としてのディープラーニングは、様々な分類問題または二値問題、バイオにおける例では、診断予測、活性化予測、遺伝子発現や、結合予測などに使われる。分類問題ばかりでなく、回帰分析の数値予測などにも利用できる。一方、教師なし学習は、正解データがない場合である。正解データがなくとも、ニューラルネットワークは自己符号化器、生成モデルなどと組み合わせることでデータの適切な表現方法を獲得し、高次元の絡まり合ったデータを低次元の独立した因子群に変換できる、つまり認識をすることができる。

このようなディープラーニングをバイオの医療の世界で利用する際の第1番目の問題は、新NP問題と呼ばれる機械学習の根本問題がある。学習



**図3 ディープラーニングにおける学習のメカニズム**  
ディープラーニングの学習していく仕組みを教師あり学習の場合を例に簡単に模式図で示したもの。

の問題はデータを使って、パラメータを推定する問題と見なすことができる。線形方程式においてN個の変数を決定するにはN個の独立した方程式が必要なと同様に、学習においてもパラメータの数と同じオーダー数の学習データを必要とする。しかし、バイオをはじめとして世の中の多くの問題は観測数(N)よりパラメータ(P)の数が多き場合がある。この場合、従来の統計的手法や機械学習ではパラメータを決めることはできない。また、一つ決まったとしてもそれは偶然、学習データを説明できるだけであり、興味のある未知のデータを説明できないかもしれない。こうした現象は過学習と呼ばれる。従来はこの過学習を抑制するために、パラメータの数を減らすしかなく、強力なモデルを使うことができなかった。今ディープラーニングの世界では、過学習を防ぐためのさまざまな手法（正則化とよばれる）が編み出され非常に大量のパラメータを使ったとしても未知のデータをうまく扱えるようになっている。また、ディープラーニングは多くのパラメータを持ちながらも、多くの場合、最適解に近いモデルを構築できることが知られており、その理論上の理由をめぐって大いに研究が進められている。

バイオや医療の世界での二つ目の問題は、正解データが少ない場合が多いことである。ゲノムのデータが利用できたとしても、現実にはそのゲノムを持つ人の詳細な疾患データなどが得られないことが多い。特に、がんゲノムの研究を行う場合、個人情報保護の問題もあり、きちんとした疾患情報が紐付いているゲノムデータ数は少ないことがほとんどであろう。このような正解データが少ない場合は、ディープラーニングは学習できず無力だと思込んでいる医師や研究者が多く見られるが、実はこのような場合においてもディープラーニングは、ここ数年で大きな飛躍を遂げている。特に半教師あり学習と呼ばれる手法は、正解ラベルがないデータも利用して、わずか数%の正解ラ

ベル付きデータを利用するだけで、精度を飛躍的に向上させることに成功している [3]。また最近では、人間がただ一つのサンプルを学習するだけでも多くの分類や推論を行えることから、one-shot learningと呼ばれる手法 [4] も出てきており、このような正解ラベルがない医療データに対しても、精度を上げるために利用できる手法の開発が進んでいることを知っておいてほしい。一見するとどうしてこのようなことができるのか不思議に思われるかもしれないが、自然界のゲノムであれ、タンパク質、化合物など、理論上とりうる次元空間を全て占めているわけではなく、実はほんの狭い空間に閉じ込められていると言っている。例えば、ヒトゲノムの場合4つの塩基(ATGC)が続いているとして、 $4^n$  ( $n$ =ゲノムの長さ)の空間を全て利用しているわけではなく、あらゆる生命、そして人間に限れば、その利用空間は著しく偏っており、高次元空間のなかに低次元空間が埋め込まれていると考えられている。ディープラーニングでは、学習により、高次元空間からより適切な分析可能な低次元空間へと変換できる。それ故、より正確な予測が可能となる。

バイオ・ヘルスケアにおけるデータの形として、これまではゲノムデータに代表される静的なデータばかりが多かったが、今後は分析機器やセンサーの進化に伴い、より動的なデータ、すなわち時系列データの分析がより重要になってくると思われる。人は生まれながらに、がんを発病していることは稀であり、普通は年齢を重ねるごとに、変異が蓄積したりすることによって、最終的にがんになると考えられている。その意味で、どう体が変化しているか？健康体か？病気になっているかどうか？は、より時系列分析が必要となってくるであろう。ディープラーニングにおける時系列データを学習させる手法としては、伝統的にリカーレントニューラルネットワーク (Recurrent Neural Net (RNN)) という手法が用いられてきた。(図

4)

RNNはニューラルネットでの中間層からの出力をもう一度中間層に新たな入力データと一緒に戻し入れてあげることにより、いわば過去の入力データの記憶を溜め込むことで時系列の情報を処理させていく手法である。しかし、時系列の情報をうまく取り込むことが一筋縄でいかないことから(勾配消失問題)、LSTM (Long Short Term Memory) と呼ばれる手法がゲノム分析などでも、これまで多く使われてきた。しかしながら、このLSTMをもってしても、ゲノムサイズの情報をうまく取り込むことは不可能で、多くの論文では、イメージ分析に使われる空間的、局所的な関連情報をうまく吸い上げることのできる、畳み込みニューラルネットワーク (Convolutional Neural Net (CNN)) などを使い、LSTMなどと組み合わせながら、ゲノムの部分的問題に取り入れることが多かった。その意味で、全ゲノムを解析する標準的なディープラーニング手法はまだ確立されていない。また、多くの医療関係者や製薬企業、薬事関係者からよく聞かれる大きな疑問の一つに、たとえディープラーニングを利用して大きな成果を挙

#### RNNの仕組み

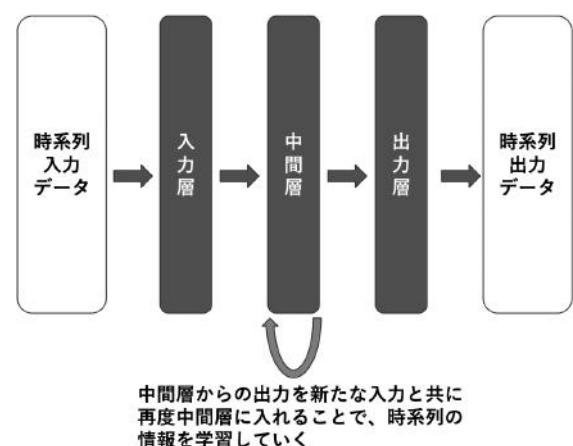


図4 RNN (リカーレントニューラルネット) の基本的な仕組み  
ゲノムデータのような、シーケンサーデータに代表される、時系列データや可変長のデータを扱う場合には、RNNが使われることが多い。しかしながら、全ゲノムを読み取り、分析するには理論的、技術的にまだ乗り越えなければならない問題が多い。

げることができたとしても、ニューラルネットの内部はブラックボックスであり、どうしてこの患者のゲノムから乳がんと判断したのか？など、人間が解釈可能な理由がうまく提示できないということが言われてきた。薬事法の認可や、PMDAなどからディープラーニング分析法が医療機器として認可・承認を得るためには、ある程度の科学的理由付けがないと難しいという問題もある。

これらの問題についても、最近大きな可能性を秘めた新手法が提案されてきている。LSTMのような複雑なユニットではなく、解析しやすく、理解しやすく、制御しやすい非常に単純な計算要素の組み合わせで同じ計算能力を獲得できるようになっている。例えば近年提唱されたQRNN(Quasi Recurrent Neural Net) [5] はCNNを使いながらも、擬似的にRNNの時系列データを分析できる能力を持てるように作られたニューラルネットといえる。(図5)

この手法では、CNNを時系列なデータに取り込み、時系列にプーリング伝搬させることで幾つかのRNNが抱える問題点を回避しながら、大規模並列処理を可能にしている。この手法、または将来その発展系がゲノム分析などに有望な理由は、ゲノムというビッグデータの様々な部分はその時の環境や様々な細胞、時系列な経過をもとに、多様

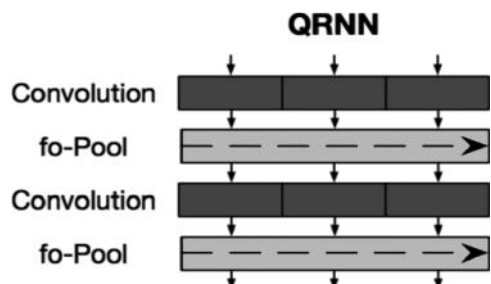


図5 QRNN (quasi recurrent neural net) の仕組み  
QRNNでは、時系列データを畳み込み(CNN)(濃帯)処理した上で、時系列のつながりをプーリングと呼ばれる手法(淡帯の部分)で時系列情報や全ゲノムデータなどをうまく学習することが可能になるかもしれない。この形のネットワークは、大規模分散並列処理も可能となる。<https://arxiv.org/abs/1611.01576>

Figure 1

に変化する状況をうまく捉えることができるかもしれないことである。その時々での発現状況や抑制機能をフィルターやゲートとして捉え、そこからマルチオミックス的に組み合わせることにより、DNA⇒RNA⇒Proteinの流れをニューラルネットとして記述し組み立てることも考えられる。このQRNNの場合、それぞれの時系列の畳み込みの結果を独立として捉えているので、どの遺伝子がどの時点でどのような影響を及ぼしているかを見るのが以前よりも分かりやすくなり、いつどの時点でゲノムの発現などが疾患に影響しているかなど、人間が解釈できる分析が可能となりうる。このように、今後も人間がより解釈しやすい、科学的に仮説を証明するためゲノム編集技術による実証実験を簡単にデザインすることができるディープラーニング手法が開発されてくると思われる。また、弊社ではこのような新しいディープラーニング技術を、柔軟に人間の直感で分かりやすく開発できるディープラーニングのオープンソースプラットフォーム「Chainer(チェイナー)」[6]を開発提供しており、現在、世界中のディープラーニングのトップ研究者たちが使い始め、新しいディープラーニング技術を活発に開発している。手前味噌であるが、上に紹介したQRNNも、海外のグループがこのChainerを使い開発したものである。

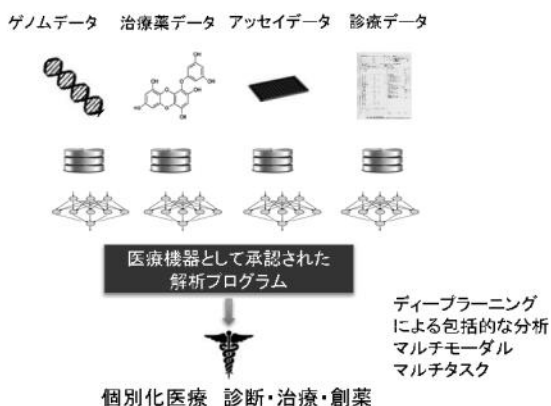
ディープラーニングは、様々なデータから自動的に特徴を抽出する、人類が得た初めての手法だと言える。今後、我々は大量のゲノムデータ、医療データから、これまでは不可能であった生命の仕組みを理解する知見が大量に得られるようになるであろう。将来的には、そこから得られた生命の仕組みを、これまでのすでに得られている生命の知識、例えばKEGGに代表されるシグナル伝達経路などとうまく組み合わせ自動構築していくシステムを作る必要があると考えられる。分野は異なるが、画像処理の分野では、Visual Genome

(visualgenome.org) というプロジェクトがすでに始まっている。このVisual Genomeとは、大量の画像データから人や物の関係性を導き出し、それを大量の知識データとして蓄え、あたかも生まれた赤ちゃんが目を通して様々な情報を知識として脳に蓄えて、その上で言語能力を獲得するように、画像と言語をつなげるデータベースを構築する壮大なプロジェクトである。近い将来、バイオ・ヘルスケアにおいても同様に、シン・ゲノムプロジェクトとも言える、生命の仕組みを理解するために、大量のゲノムデータから科学知識を自動抽出し、自動構築していく国家プロジェクトが必要になると考える。

### AIを活用した統合的がん医療システムの開発と個別化医療への挑戦

国立がん研究センターとの「人工知能 (AI) を活用した統合的がん医療システム開発プロジェクト」においては、ディープラーニングの特徴である、マルチモーダルを最大限に利用することにより、膨大な罹患者の詳細な臨床情報やマルチオミックスデータを組み合わせて、包括的な医療統合システムを構築する。そしてディープラーニングの持つマルチタスク性により、診断、治療、創

#### 人工知能(AI)を活用した統合的がん医療システム開発プロジェクト



**図6** 人工知能 (AI) を活用した統合的がん医療システム開発プロジェクト  
PFNは国立がん研究センターと協力して、ディープラーニングを利用することにより、様々な医療データをマルチモーダルの取り込み、診断、治療、創薬の三つの異なるタスクを行える、医療機器として承認される解析プログラムを開発する予定である。

薬のそれぞれ異なるタスクを行えるようにしていく計画である。(図6)

すでに我々は国立がん研究センターと協力して、血液一滴から様々ながんを早期予測するプロジェクトに取り組んでおり、これまで約5000症例のデータを使い、99%以上の精度で様々ながんの早期予測が可能であることを実証している。このディープラーニングを用いた画期的がん診断法は今後、世界で標準がん診断法として利用され、世界中の多くの人々をがんで苦しむことから未然に防ぐことが可能となるであろう。今後は、臨床試験を行い、2-3年後の実用化に向けて着実に進めていきたい。しかしながら、がんのメカニズムを知り、最終ゴールである真のPrecision Medicine実現のためには、ディープラーニングなどの最新AI技術を取り入れながらも、物理学、数学、生物学、生化学、遺伝学や免疫学に至るまで、これまでの人類が蓄積した英知を組み合わせる必要があると考えている。この道のりは限りなく遠く、一歩先すら簡単には見えない、困難を伴うものと予想されるが、我々は一歩ずつ着実に研究開発を続け、がん医療の質の向上を推進していきたい。

#### 参考資料:

1. 国立がん研究センター  
[http://www.ncc.go.jp/jp/information/pr\\_release/press\\_release\\_20161129.html](http://www.ncc.go.jp/jp/information/pr_release/press_release_20161129.html)
2. GTC Technology Conference 2015 Silicon Valley,  
<http://on-demand.gputechconf.com/gtc/2015/presentation/S5813-Nobuyuki-Ota.pdf>
3. Kingma, Diederik P., et al. "Semi-supervised learning with deep generative models." Advances in Neural Information Processing Systems. 2014
4. Salakhutdinov, R., et al. "Learning with Hierarchical-Deep Models, IEEE Transactions on Pattern Analysis and Machine Intelligence, 35, 8, 1958-1971 2013.
5. Bradbury, J., et al. "Quasi-Recurrent Neural Networks" <https://arxiv.org/abs/1611.01576>
6. ディープラーニングのオープンソースプラットフォーム「Chainer (チェイナー)」  
<http://chainer.org/>