

# 脳における双方向情報処理

講 師：川人 光男

レポーター：妻藤 公啓

平成 12 年 8 月 8 日

NISS2000 には凄く思い入れがありまして、何年か前からとくに銅谷先生ととくに日本の computational neuroscience の教育体制が整っていないということをずっと心配し続けています。それはどういうことかと申しますと、神経科学の分野で computational、計算論的なアプローチというのが世界的にも認められているし、予算もたくさん付くようになったのですが、今度はその予算で人を雇いたいと思ったときにじつは日本の中で computational neuroscience の教育をきちっと受けている人はいないわけですね。

翻ってアメリカを考えてみますと、例えばカルテク、MIT などの 7、8 つの大学院に computational neuroscience と名前が付いたようなプログラムがありまして、そこでシステムティックに教育されているわけです。また、Sloan Foundation という財団がアメリカにあるわけですが、そこで theoretical neurobiology という、主に元々物理や computer science を専攻していた graduate student やポスドクの人達を neuro scientist にシステムティックに変えるようなプログラムがあります。またヨーロッパでは、EU の Summer Course ( 今年イタリアのトリエステで行われる ) では 4 週間の長さに渡ってずっといろいろな講義と実習を続けるわけです。

このように欧米では computational neuroscience で研究をやっていけるようなレベルにまで大学院やサマーコース等で非常にしっかりと教育しているのですが、一方、日本の大学は非常に硬くて学科をどんどん新しく新設することはなかなかできないわけです。日本って悲しいねということもずっといい続けていたのですが、そんなことをいっていても始まらないので、自分たちで解決しなくてはならないわけです。そこで去年から、銅谷先生、塚田先生が中心となってこのサマースクールが始まったわけです。ことしは、神経回路学会、塚田先生の科研費、ほとんどの講師の先生が滞在費を自己負担したわけです。そのようなことを始めとして、多くの講師の方々が手弁当で作り上げている、そのような意味で非常に思い入れがあります。この中から computational neuroscience をこれから研究していく人が出て来ることを切に願っております。

このサマースクールの全体的なテーマは脳における双方向情報処理、予測と文脈生成ということです。日本の脳研究は「脳を守る」「脳を知る」「脳を創る」という三つの分野に分かれております。私の解釈では脳を創ることによって脳を知るまたは脳を作れるほどに脳を知ることが日本の脳研究のオリジナルなところであると思います。私の考えを述べますと、ロボットやコンピュータは進化はしているものの、人に比べたら劣っているわけです。人工知能やロボットの研究者があまり賢くないということもあるかも知れませんが、

神経科学、脳科学の人達が正しい方向で脳を分かっていないことが原因であるともいえます。物とか場所のことは分かるけれども、視覚や運動制御の事に付いては本当に私達には脳がどう働いているかということが分かっているならば、それを計算機のプログラムに載せてロボットやコンピュータを動かすことはそれほど原理的に難しい話ではないのです。結局、情報処理の本質的なところまでは分かっていないのです。科学としてやるには、物とか場所を研究した方がずっと信頼性が高いし論文もしっかり書けるわけですが、情報処理そのものに関して scientific な仕事をしようと思うと実は凄く大変なわけです。どうしても、研究者は物や場所に流れがちだから、脳を創るということを constraint におきながら脳を知るということは大切なポイントであります。ところが脳を創って試験管の中に浮かべていても脳の働きは分からないわけで、脳の機能を分かるためにはセンサー、体、環境などが大切になって来るわけです。そのためには人を作らなくてはならないということになりますが、それではタンパク質や DNA で人を創らなければならないかということ、それも技術的には無理と思われるので、私達にとっては、人間型のロボットを創って人工的な脳と組み合わせましようという「アトム計画」が目的になるわけです。もう少し硬くいきますと次のようになります。

### 計算論的神経科学

脳の機能を、その機能を脳と同じ方法で実現できる計算機のプログラム、あるいは人工的な機械を作れる程度に深く本質的に理解することを目指すアプローチを計算論的神経科学と呼ぶ。

ここで、「脳と同じ方法で」という部分を外してしまいますと、これは単なる人工知能やロボティクスの研究と同じになってしまいます。一方「作れる程度に」という部分を外してしまいますと、普通の神経科学の研究になってしまい、必ずしも情報処理の機構を理解するという所までいかなくなってしまいます。

我々の所のロボットがどのような事が出来るかというのをちょっと紹介致します。我々の所の人型のロボットに「カチャーシ」という沖縄の舞踊を踊らせています。必ずしも見まね学習の所に脳の原理がしっかりと入っているわけではないですけども、人に体に似たものをつくって脳の働きを知ろうとするアプローチを採っています。

また、これは良くある例ですが、shape from shading といって、陰影から構造を復元する時に私たちの脳が特別な拘束条件を使っていることを示す一つの例です。我々が普段住んでいる所では上から光線が来ますので、上側が明るくて下側が暗い所は出っばって見えて、逆に上側が暗くて下側が明るいところはへこんで見えるということで、陰影から構造を復元できます。その時に照明光が上からくるといって、事前の知識を使ってある種の計算をしているわけです。

また、これも有名な例ですが、「aperture problem (窓枠問題)」と呼ばれる問題があります。回転している多面体のアニメーションを見ると、我々はそれが多面体がぐるぐる回転しているものであることが分かりますが、一方、我々の視覚野にある一つ一つの神経細胞の受領野は二つ以上の辺が同時にはいることはないくらい小さく、また辺に対して垂直な動きは分かるのですが、輪郭そのものが接線方向にどのように動いているかというこ

とは分からないわけです。しかしそれにも関わらず、我々は全体として立体的な動きを認識しているわけです。この問題に対しても、コンピュータビジョンの人達はいろいろなことを考えていまして、ここに挙げるのが唯一正しい理論というわけではありませんが、例えば Hildreth という女性研究者 ( David Marr の弟子 ) は次のように考えました。第一次視覚野のニューロンにしる、コンピュータビジョンのアルゴリズムにしる、観測できるのは輪郭に対して垂直方向の速度成分です。一方、知りたいのは真の速度ベクトルなのですが、輪郭に対して接線方向の成分は観測できませんので、真の速度ベクトルの解が一意に決まりません。そこで真の速度ベクトルを輪郭に対して垂直方向に射影したものと実際に観測される輪郭に対して垂直方向の速度との差を出来るだけ小さくするという条件に加えて、速度ベクトルの輪郭に対して接線方向の変化が出来るだけ小さくなるようにするという基準を設けます。すると、先ほどの多面体の動きを回復することが出来ますし、或は床屋のマークが上に上がる barber pole の錯視の例を説明することも出来ます。これらの「shape from shading の問題」や「structure from motion の問題」などは、基本的に標準正則化理論と呼ばれる方法で解かれています。

### 標準正則化理論

光学:

$Az = y$  データ  $y$  から  $z$  を解く ( 逆問題 )

画像から三次元世界の様子を推測

$|Az - y|^2 + \lambda |Pz|^2$  を最小にする  $z$  を求める。

画像データ拘束 世界のモデル、滑らかさ、先験的知識。

$P$ : 線形汎関数

網膜に入ってくる画像を  $y$  とします。視覚の問題で知りたいのは、真の速度ベクトルや奥行きなどの三次元世界を記述する変数  $z$  です。画像から変数  $z$  を知りたいのですが、実際に画像が生成される過程で何が起きているかを考えますと、面の奥行きや向きや速度にある optics ( 光学 ) の過程  $A$  を作用させる、それによって画像が出てくると考えることができます。ここで視覚の問題というのは、 $y$  から  $z$  を求めるという問題なのですが、これは実は逆問題になっていまして  $A^{-1}$  は一般には計算不可能です。その為に他の拘束条件を入れなくてはなりません。即ち、 $z$  が滑らかだとか、激しく変化しないなどの条件を入れて  $z$  を求めます。このような形でコンピュータビジョン或はコンピューテーショナルビジョンのアルゴリズムが理解されます。ここで注意すべきことは、 $Az = y$  には三次元世界から光学の画像をつくる過程のモデル ( 順光学モデル、internal model、forward-optics model などと呼ばれる ) が入っているわけですが、このようなモデルがないと視覚の問題が解けないということです。また、上で述べた標準正則化理論というのはベイズ推定の立場からは、最大事後確率推定として解釈する事が出来ます。

私は、京都大学の乾先生と一緒に 10 年くらい前に、双方向理論 ( bi-direction theory ) の視覚のモデルを考えました。三次元世界の中には、ある奥行きや物体の形などの表現があって、それから二次元画像データが作られます。これは画像の生成過程或は順 optics などとされます。それに対して、脳の中では二次元画像データから高いレベルでの、或

は低いレベルでの三次元世界の画像を復元しなくてはならないわけです。また、非常に良く知られている事は、大脳皮質には色々な種類の領野があるのですが、その間に順方向と逆方向の結合があります。画像のデータは一次視覚野から高次視覚野に送られる後ろから前に流れる経路とは別に前から後ろに流れる back-projection と呼ばれる流れもあります。これに関してはほとんど同時期にいろいろな人が殆んど同じようなことを述べたのですが (例えば Mumford など)、私たちが考えた事は、一次視覚野から高次視覚野には逆 optics の過程がありまして、高次視覚野から一次視覚野には順 optics の過程があります。これはどういう事であるかといいますと、高次視覚野で表現されている表現から画像を復元するようなプロセスが back-projection で形成されていて、このように高次の予測から復元された画像データと入ってくるものとの誤差をとってそのエラーの部分だけを上に上げてやることで視覚の問題が非常に速く、かつ確実に解けるといわけです。

渡辺先生、Rao 先生の講義はこのような流れと非常に関係があります。高次視覚野から低次の視覚野に戻ってくる back projection というのは一般にはトップダウン (例えば attention, consciousness) による制御と考えられます。昔はトップダウンによる制御の影響というのではないと考えられていましたが、最近では例えば第一次視覚野まで戻ってくるということが分かってきました。

しかし、このようなコンピューショナルビジョンのいくつかの例題を解決しても、運動認知に関してはまだまだ困難な問題があります。例えば、我々がヒトが歩いているビデオを見たときに、どのような関節運動が行われているかということはすぐに分かるのですが、これをコンピュータビジョンのアルゴリズムでやろうとすると非常に大変です。ここで紹介する方法は我々の研究室の Ales Ude さんがやっている方法ですが、それは、関節角やカメラのパラメータの情報から画像を直接再現するようなある意味で順光学的なモデルを持っていて、それとカメラからの画像をマッチングすることによって EM アルゴリズムと robust statistics を用いて関節角を推定するといったことができています。しかし、このアルゴリズムを用いても例えば腕をひねるような動きが入ったダンスは追跡不可能になるといった問題があります。私たちはヒトの動きを視覚的に理解する為に、私たちがどのようにそういった運動を作り出しているかという、ある種の内部モデル (順モデル) をもっと多く使っていると思われれます。特に明日の講義の中では、順モデルと逆モデルという話が運動制御の流れでたくさん出てくると思います。

### 順モデルと逆モデル

- 運動制御の対象の入力は運動指令、出力は運動の軌道
- 順モデルの入力は運動指令のコピー、出力は運動の軌道、もしくは感覚フィードバックの推定：入出力が順方向
- 逆モデルの入力は望ましい運動軌道、出力は前向きの運動指令：入出力が運動制御対象と比べて逆方向

順モデルと逆モデルというのは、運動制御だけでなく視覚の世界にもすでに出てきまし

た。運動制御の世界では順モデルと逆モデルがそれぞれどのようなようになるかといいますと、運動制御の対象というのは例えば目とか手足などですが、順モデルの入力というのは運動司令のコピーで、出力は運動の軌道もしくは、感覚のフィードバックの推定になります。この場合は入出力の方向が運動制御の対象と同じなので順方向といいます。一方逆モデルの入力は望ましい運動軌道で、出力は前向きの運動司令です。入出力は運動司令と比べて逆方向になります。そういうわけで、順モデル、逆モデルというのは、モデル化するものと入出力の方向が同じかどうかということで同じであれば順モデル、逆であれば逆モデルといいます。先ほど、なぜ視覚の世界で内部モデルというものが出てくるかということを書きましたが、運動制御の世界でなぜ内部モデルが必要かという議論は、少なくとも私たちのグループではフィードバック制御だけで制御が行えるかどうかという問題意識から出てきたものです。これに関しては、片山先生の講義のなかで詳しく述べられます。フィードバック制御というのは、工学的には、実現された軌道を何らかのセンサーで計って目標軌道と比較して制御する方法です。ところが、もし逆モデルというものが存在しますと前向き制御ができましてフィードバックに頼らなくてもすむわけです。五味先生の実験では、P F Mという人間と同じくらいに速く動くマニピュランダムを使うことによって人の腕の硬さ（剛性 stiffness、粘性 viscosity）が計られました。それは、フィードバック制御で考えるとゲインの部分の部分を計っていることになります。その結果、ゲインが非常に低いことから、これだけでは速くて滑らかな運動制御ができないことがわかり、それゆえ内部モデルが必要であるという議論がなされるわけです。内部モデルに関しては、私や今水先生や北沢先生の講義で触れますが、内部モデルのかなりの部分は小脳にあるだろうということが分かってきました。以上の話より、小脳の古い部分に関しては運動制御に関わる内部モデルがあるということが分かるのですが、最近の神経化学の研究で分かってきたことは小脳や大脳基底核などが今までは運動制御だけに関わると思われてきましたが、実は言語や意識などといった高次機能にも関わることが分かってきました。それは実は解剖学的な結論を見ると、全然不思議でも何でもありません。実際、小脳の新しい部分は大脳皮質の46野や9野など frontal-cortex の作業記憶に関わるどころ繋がっています。小脳の新しい部分が高次機能に関わっているということは、分かってみるとぜんぜん不思議ではありません。さらに最近では、小脳の中の順モデルの異常で分裂病までもが説明できるのではないかということが言われるようになってきました。

## 小脳内の順モデルの異常：分裂病 ( Frith, Blakemore, Wolpert )

- 自分でくすぐってもくすぐたくない。
- 小脳の活動で、体性感覚野の活動が抑制される。  
( Nature Neurosci., 1998; NeuroImage 1999 )
- 分裂病の患者は自分でくすぐってもくすぐりたい。
- 幻聴などの多くの症状が、順モデルによる運動指令からの感覚フィードバックの予測の異常とみなせる。  
( Brain Res. Rev., 2000 )

私が共同研究している Daniel Wolpert という computational neuroscience の研究者のグループが次のようなことを主張しています。我々は、自分で自分をくすぐってもくすぐったくはありません。これはどういうことかといいますと、小脳で自分が発している運動司令から感覚フィードバックを予測してそのフィードバックを第一次体制感覚野で差っぴいて、体制感覚野の活動が抑制されてくすぐたくないということが pet や fMRI を用いた実験で分かってきました。ところが、分裂病の患者さんは自分でくすぐってもくすぐりたいと感ずります。これは順モデルによる運動司令からの感覚フィードバックの予測の異常とみなせるといった説が出されるようになってきました。これはどういうことかといいますと、私たちが動き回っているときに私たちが受け取っている感覚情報のほとんどすべては運動によって引き起こされたものです。例えば、聴覚系に入ってくる音響フィードバックは自分自身がしゃべっている声であるし、視覚世界のスリップは自分自身が動き回った結果であるし、足の裏から入ってくる体制感覚情報は自分自身が動いているから入ってくるのです。このように、自分が受け取る感覚情報のかなりの部分は自分自身の運動によって引き起こされるものであります。それを誰かほかのエージェントが引き起こして自分の感覚フィードバックが変わるということとしっかり区別する必要があります。そうしないと、本当に大切な情報を落としてしまう危険があります。そういったことが普通にできているのが正常な人で、できていないのが分裂病の患者さんだということです。運動制御のモデルとして目標としているものから、逆モデルを使って motor command を出してある動きがあって、センサーから感覚情報が戻ってくるのですが、逆モデルが運動司令のコピー（遠心性コピーなどと呼ばれる）を使ってそれから順モデルを使うとどんな状態が出てくるかとかどんな感覚フィードバックが出てくるかといったことが予測できるわけですね。その予測と実際に入ってくる感覚フィードバックとを差っぴいた結果キャンセルされればそれは自分の運動で引き起こした感覚フィードバックの変化ですからほとんど注意する必要がなくなります。ところが、ここでキャンセルできないものというのは何かほかのエージェントが起こしたものですからそれは注意しなくてはならない。という構造になっています。もしここで順モデルが壊れると何が起こるかといいますと、自分が引き起こした運動を全部ほかの人が引き起こしたように思われるわけです。私が聴音器官に送る運動司令からどんな聴覚フィードバックがあるはずだという順モデルが壊れたとしますと、い

ま聞いている声が全部他人の声に聞こえるわけです。実際に分裂病の患者さんが auditory ... を起こしているときにはじつは自分でしゃべっているわけです。それは内言といいまして、喉のあたりの筋電を計ると筋肉の活動は認められるのですが、聞いているものが神の声、第三者の声、あるいはおじいさんが私に催眠術をかけて私の体を操っているということを使うわけです。これは結局意識の問題なんです。自分の運動を、自分が起こしたと意識できるかどうか、即ち motor control の consciousness に関わるわけで、どれが自分が随意的に起こした現象でどれがそうでないかを区別する、自分と他人を区別することが順モデルを使った予測で左右されるわけです。私たちは最近、順モデルや逆モデルの多重あるいは階層的な構造というのは人の知性にとっても本質的でコミュニケーションの問題や言語の問題などもこういったところから理解できるだろうと思っています。これに関しては、春野先生の講義で詳しく取り上げます。

### 心の理論 ( Theory of Mind )

サリーとアンが部屋の中にいます。  
サリーがボールを籠の中に入れました。  
サリーが部屋から出ていきます。  
サリーがいない間にアンはボールを籠から箱に移します。  
サリーが帰ってきてボールを探すのですが  
サリーはどこを探すでしょうか。

最近、心の理論 ( theory of mind ) と呼ばれる理論が霊長類学や発達心理学においてはやっています。上に挙げた問題に対して、8歳以上くらいの子供であればサリーは「籠の中」を探すと正しく答えます。しかし、4歳以下の子供や10歳以上で知能指数はかなり高くても自閉症の子供だと「箱の中」を探すと答えます。物理的にはボールは箱の中にあるので「箱の中」というのは物理的には真実です。しかし、サリーはアンがボールを移すところは見えていないわけですから、アンの心の状態を推測することができれば「籠の中」と正しく答えることができるのです。心の理論というのは、物理学者が物理世界に対して理論を立てるようにすべての一般人も他人の心の動きに関してある理論を持っていて、そういった理論を持っているからこそ、他人の心の状態を推定できたり、それをコントロールしたりしてコミュニケーションができるのです。そういう能力が人間のコミュニケーションにとって必須であるということで theory of mind と呼ばれます。霊長類学者はチンパンジーでもこういった能力を持っていると主張しているし、人を取り扱っている発達心理学者は人間にしかない、チンパンジーに関しては実験室でやると証明できないなどといった論争になっております。こういった能力も実は順モデルとか逆モデルと言った視覚や運動制御の非常に基本的な計算論的な考え方をベースにして、他人の脳の構造、ありようをほかのエージェントが獲得できるかどうかという問題として捕らえることができるということです。今回のコースでは、全体では非常に基本的である視覚情報処理から運動制御まで予測、あるいは順モデルと逆モデルを使って双方向に情報処理をするという考え方が非常に大切です。双方向情報処理の中では、トップダウンの上のレベルから下のレベルに向かって情報処理をする順モデルが文脈生成をしているというように考えられます。このよ

うな概念はついこの間までは neuro science のなかでは純粋に理論的で、生物学的バックグラウンドを持たないものと思われていましたが、いろいろな研究によってそういったものが確かに存在する、あるいはその神経機構はどんなものか、どうやって学習されるのかといったことが分かってきました。それだけでなく、より高次へ高次へとコミュニケーションや言語とか意識とか分裂病とか同じような枠組みの中で分かるかもしれないいうようになってきています。

#### 質問と回答

Q1: NHK のテレビで小脳を壊した子供が運動ができるという話がありましたが、小脳以外でも小脳の役割を果たしている部分があるのではないかということについてどのようにお考えですか。

A1: 損傷の実験というのは非常に難しく、破壊のされ方にもよるところが大きいです。あるイギリスの若手数学者で、IQ が 115 程度で数学の賞をもらって将来が有望な人がいましたが、亡くなった後に脳を解剖したら、彼が実は水頭症で大脳皮質がほとんどなかったということです。言いたいことは、どこかが壊れていてかなり長い間生きていていろいろなことができますということが分かったからといって壊れた部分には正常な人にはその機能が無いという証拠にはほとんどなりません。

Q2: 先ほどのサリーの話では運動制御で言うところの体制感覚のフィードバックは何に当たるのですか。

A2: それは非常に難しい問題です。運動制御の場合は比較的モデルを作るのが簡単で、運動制御の運動司令は脳の中にあ驕フですね。それから体制感覚のフィードバックもあるわけですね。すると順モデルによって運動司令からどんな体制感覚のフィードバックが獲られるかということ予測できます。教師あり学習の枠組みの中で内部モデルが作れるわけです。ところが他人の心のモデルを作るとなると観察できるものは基本的には他人の運動パターンだけなんです。たとえば、コミュニケーションでは、しゃべろうが身振り手振りだろうが、顔の表情の変化だろうが、要するに他人の作った運動パターンを視覚、聴覚、体制感覚で捕らえるわけです。それでどうやって他人の心の中のモデルを作るかというのは、基本的にはできそうに無い問題に思えるのですが、拘束条件がありまして、例えば銅谷先生と私の脳がぜんぜん違うものであったらこれは絶対に不可能です。火星人の心の理論というのはなかなか作れないのですが、どこか共通なものがあるとすれば自分の脳と他人の脳が似ていると思うと、それを拘束条件にして他人が作り出している運動司令を推定することができる。他人の脳の内部構造を推定することがある程度できるのです。そういったことが起きているのであろうと私たちは考えています。