

# 心理学，神経科学，機械学習における注意

Grace W. Lindsay

## 要約

注意は、限られた計算資源を柔軟にコントロールする重要な能力である。神経科学や心理学の分野では、意識、警戒、注意、実行制御、学習など、他の多くのトピックと関連して研究されている。また、最近では、機械学習のいくつかの領域でも応用されている。生物学的注意研究と、人工ニューラルネットワークを強化するツールとしての注意の使用との関係は、必ずしも明確ではない。このレビューでは、まず、神経科学や心理学の文献で注意がどのように概念化されているかを概観する。次に、機械学習における注意の使用例をいくつか取り上げ、生物学的に対応するものがある場合はそれを示す。最後に、複雑で統合的なシステムを構築するために、人工的な注意力を生物学からさらに引き出す方法について検討する。

## 1. はじめに

注意は、公に広く議論され、科学的にも広く研究されているトピックである。心理学、神経科学、そして最近では機械学習を含む複数の分野で、様々な定義がなされている (Chun et al.2011; Cho et al. 2015)。ウィリアム・ジェームズは実験心理学の黎明期にこう書いた:「注意とは何かは、誰もが知っている。注意とは、同時に可能と思われる複数の対象物や思考の流れのうち、1つを明確かつ鮮明な形で心に刻むことである」。ジェームズがこの言葉を残して以来、このプロセスをより正確に定義し、定量化するために多くの試みがなされており、また、このプロセスを生み出す根本的な精神的・神経的構造も明らかにされている。しかし、ひとつの概念として語られるものを研究するために、さまざまな実験的アプローチや概念が氾濫していることが、研究者の反発を招いている。概念に対するより進化に基づいたアプローチを主張する最近の論文のタイトルにもあったように、「注意とは何かは誰にもわからない」のである (Hommel et al. 2019)

注意は、明確で統一された概念とは程遠いものである。しかし、多くの曖昧な定義があり、時には相反する定義があるにもかかわらず、脳や、最近ではAIシステムの情報処理にとって非常に重要であることが明らかになっている、注意の中核となる性質がある。注意とは、限られた計算機資源を柔軟に制御することである。なぜ注意の資源が限られているのか、どのように制御するのがベストなのかは、ユースケースによって異なる。しかし、情報の流れを動的に変化させたり、ルーティングしたりする能力は、あらゆるシステムの適応性にとって明らかに有益である。

注意が脳内で多くの役割を果たしていることを考えれば、人工ニューラルネットワークに注意が加わるのは当然のことと言える。人工ニューラルネットワークは、ニューロンの基本的な入出力機能を模倣して設計された、個々のユニットからなる並列処理システムであり、現在、機械学習や人工知能 (AI) の分野で主流となっているモデルである。当初は注目されずに構築されていたが、現在ではネットワークの表現や構造を動的に再構成するためのさまざまな仕組みが追加されている。

続く第2節では、神経科学と心理学における「注意」という言葉のさまざまな使用法と、他の一般的な神経科学のトピックとの関連性について概説する。全体を通して、限られた資源をコントロールする方法としての注意の概念が強調される。行動学的研究は、注意の能力と限界を示すために使用され、神経メカニズムは、これらの行動的効果が現れる物理的手段を示す。第3節では、機械学習における注意の研究状況を要約し、AIの注意と生物学的な注意との関係がある場合にはそれを示す。第4節では、生物学的注意から得られた知見が人工的な注意に影響を与える追加的な方法を紹介する。

本概説論文の第一の目的は、AIや機械学習の分野の研究者に、神経科学や心理学において注意がどのように概念化され、研究されているかを理解してもらい、実りあるインスピレーションを得られるようにすることである。第二の目的は、生物学的注意を研究している研究者に、これらのプロセスがAIシステムでどのように運用されているかを伝えることである。これにより、生物学的知見の機能的意味合いについての考え方に影響を与える可能性があるだろう。

## 2. 神経科学と心理学における注意

注意の科学的研究は心理学から始まった。慎重に行動実験を行うことで、さまざまな状況下での注意の傾向や能力を正確に示すことができる。認知科学や認知心理学では、こうした観察結果をもとに、どのような心のプロセスがそのような行動パターンを生み出すのかをモデル化することを目指している。これまでに、さまざまな基礎的メカニズムを想定した多くの単語モデルや計算モデルが作られてきた (Driver, 2001; Borji and Itti, 2012)。

人間以外の霊長類における単一細胞の神経生理学の影響や、EEG、fMRI、MEGなどの非侵襲的な脳活動のモニタリング手段により、基礎となる神経プロセスの直接観察が可能になった。これにより、注意に関連した神経反応の特定の特徴を再現できる神経回路の計算モデルが構築された (Shipp, 2004)。

以下では、注意に関するいくつかの異なるクラスについて、行動的および神経的な知見を説明する。

### 2.1. 覚醒，アラートネス，警戒，としての注意

最も一般的な形では、注意とはすなわち、周囲の環境に関与するための全体的な警戒心や能力のレベルであると言える。このようにして、注意は覚醒や睡眠・覚醒スペクトルと相互作用する。心理学における「Vigilance 警戒」は、注意を持続する能力を意味しており、これも関連している。なお、これらの言葉は同じ意味で使われているが、異なるニッチな文献ではより具体的に使われていることもある (Oken et al, 2006)。

睡眠覚醒サイクルのさまざまな段階、睡眠不足、鎮静剤の服用中の被験者を研究することで、このような注意の形態がどのように変化し、その結果どのような行動が起こるのかを知ることができる。スクリーン上の特定の領域にボールを置くなど、持続的な注意力を必要とする反復的な課題を被験者に与えることで、研究者たちは、脳波信号の変化と関連して、眠気を感じている患者の成績が長時間低下することを観察した (Makeig et al, 2000)。しかし、課題をより魅力的にすることで、眠気や鎮静状態でも成績を向上させることができる方法がある。課題を実行することで得られる報酬を増やす、目新しさや不規則性を加える、ストレスを導入するなどである (Oken et al., 2006)。したがって、一般的な注意には限られた蓄えがあり、平凡な課題や十分な報酬が得られない課題の場合には展開されないが、より有望な課題や興味深い課題の場合には呼び出せるようである。

興味深いことに、覚醒度が高ければ良いというわけではない。Yerkes-Dodson曲線 (図1B) は逆U字型で、十分に難易度の高い課題に対する覚醒度の関数としてパフォーマンスを表している。覚醒度が低いとパフォーマンスは低下し、中程度になると良好になり、高くなると再び低下する。オリジナルの研究では、マウスに電気ショックを与えて覚醒度を変化させたが、この結果は他の測定方法でも繰り返されている (Diamond, 2005)。アデロールやカフェインなどの精神刺激剤が、ある人にはある用量で集中力を高める働きをするが、他の人には有害になる理由を説明できるかもしれない (Wood et al., 2014)。

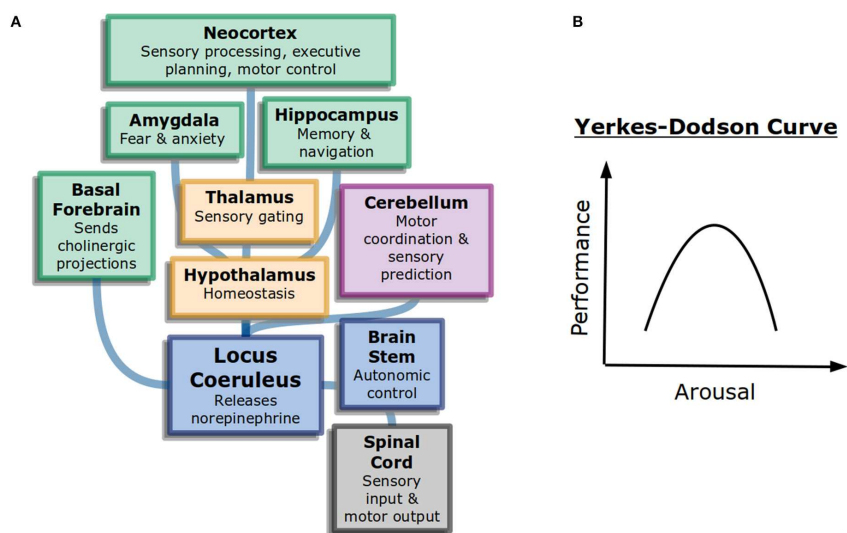


図1 青斑核の細胞は、ノルエピネフリン (ノルアドレナリンとしても知られている) を、他の神経調節系を含む、さまざまな機能を持つ脳の多くの部分に放出する。これが全体的な覚醒に寄与する (Samuels and Szabadi, 2008)。ここでの色は、前脳(緑)、間脳(黄)、脳幹(青)といった脳の各部位を表している。(B) ヤーケス・ドッドソン曲線は、覚醒度と課題遂行能力の非線形関係を表している。

睡眠・覚醒サイクルの基盤となる神経回路は、主に脳幹にある (Coenen, 1998)。これらの回路は、視床と先天性皮質への情報の流れを制御する。さらに、神経調節系は全般的な注意力の制御に大きな役割を果たしている。ノルエピネフリン、アセチルコリン、ドーパミンは、それぞれ、警戒心、重要な情報への方向付け、注意の実行制御に影響を与えると考えられている (Posner, 2008)。神経調節物質の解剖学的構造は、その機能とも一致している。例えば、ノルエピネフリンを放出する神経細胞は、細胞体は脳幹にあるが、脳全体に非常に広く分布しており、情報処理を広範囲に制御することができる (図1A)。

## 2.2. 感覚的注意

全体的な覚醒度に加えて、覚醒している被験者は特定の感覚入力に対して選択的に注意を展開することができる。特定の感覚システムの中で注意を研究することで、刺激と注意の位置の両方を厳密に制御することができる。一般的に、このタイプの注意を調べるには、使用する課題がかなり難しいものである必要がある。例えば、変化検出課題では、2つの刺激の間の検出されるべき違いは非常に小さいかもしれない。より一般的には、課題の難しさは、刺激を短時間提示するか、弱く提示することで達成できる。

システム神経科学や心理学における注意研究の大部分は、特に視覚的注意を中心としている (Kanwisher and Wojciulik, 2000)。これらの分野では、他の感覚システムよりも視覚処理の研究を重視するという一般的な傾向 (Hutmacher, 2019) と、霊長類の脳において視覚が支配的な役割を果たしていることを反映していると考えられる。さらに、注意のより一般的で認知的な側面を扱う研究でも、視覚刺激が頻繁に使用されている。

視覚的注意は、空間的注意と特徴に基づく注意に大別される。

### 2.2.1. 視覚的空間的注意

サッカードとは1秒間に数回行われる小さくて速い眼球運動のことである。眼窩は網膜上で最も高い解像度を持つため、眼窩をどこに配置するかは、限られた計算機資源をどこに配置するかという選択に他ならない。このように、眼球運動は注意の所在を示す。注意の移動が外に向かって見えることから「顕在的視覚注意」と呼ばれている。

研究者たちは、異なる画像を提示したときの眼球運動を追跡することで、自動的に注意を引きつける画像パターンを特定してきた。このようなパターンは、エッジの向き、空間周波数、色の対比、強度、動きなどによって定義される (Itti and Koch, 2001)。注意を引く画像領域は「顕著」とみなされ、

「ボトムアップ」方式で計算される。つまり、意識的な処理や努力を必要とせず、視覚システムに組み込まれた特徴検出器の結果であると考えられる。そのため、顕著性(顕在性)は非常に高速に計算される。さらに、どの領域が顕著なのか、特に最初の数回のサッケードで識別された領域については、異なる被験者でも一致する傾向がある (Tatler et al, 2005)。

顕著な領域は、画像をどのように見るかという特定の指示を与えられていない「自由観察」の状況で調べることができる。特定の課題が与えられると、ボトムアップ注意とトップダウン注意の相互作用が明らかになる。例えば、配列の中の特定の視覚目標にサッケードするように指示された場合、被験者は、代わりに特に顕著な妨害刺激に誤ってサッケードすることがある (van Zoest and Donk, 2005)。より一般的には、複雑な自然画像を見ているときに、高レベルの課題 (例えば、人物の年齢を評価したり、社会経済的地位を推測したりするような課題) を与えられた場合、課題の指示がサッケードのパターンに大きな影響を与える可能性がある。さらに、被験者がサンドイッチを作るような実世界の課題を実行する際の自然な視線移動のパターンは、基本的な認知プロセスに関する洞察を与えてくれる (Hayhoe and Ballard, 2005)。

被験者は、複数回のサッケードを連続して行う必要がある場合、最近見た場所には戻らない傾向があり、その場所で何か関連することが起こっても反応が鈍くなることがある。この現象は「戻り抑制」として知られている (Itti and Koch, 2001)。このような行動は、視覚システムに、元々最も顕著であると考えられた画像領域を利用するだけでなく、他の領域も探索するように促す。また、サッケードを生成するシステムには、一種の記憶が必要である。これは、最近見た場所の表現を短期的に抑制することで実現されると考えられている。

眼球運動は視覚的注意を制御する有効な手段であるが、唯一の選択肢ではない。「隠れた」空間的注意とは、焦点の位置を明らかに変化させることなく、異なる空間的位置の処理を強調する方法である。一般に、隠れた空間注意の研究では、被験者は課題中、中心点を固定しなければならない。このとき、被験者は、自分の視覚課題に関連した刺激が現れる可能性の高い周辺視野の場所に隠密に注意を払うように促される。例えば、方位識別課題では、空間的な手がかりが与えられた後、手がかりとなった場所に方位のある格子が点滅し、被験者はその方位を示す必要がある。無効手がかり試行 (手がかりのない場所に刺激が現れる場合) では、被験者は有効手がかり試行 (または手がかりのない試行) よりも成績が悪くなる (Anton-Erxleben and Carrasco, 2013)。これは、隠れた空間的注意が、柔軟に展開できる限られた資源であり、視覚情報の処理を助けることを示している。

隠れた空間的注意は、特定の領域が他の領域を犠牲にしてさなる処理のために選択されるという意味で、選択的である。これは、注意の「スポットライト」と呼ばれている。重要なのは、過剰な注意とは対照的、隠れた注意では、視覚システムへの入力と同じでも、その入力の処理は柔軟に選択されるということである。

隠れた空間的注意は、ボトムアップの顕著性によっても影響を受ける。無関係だが顕著な物体が、課題に関連する刺激がある場所で点滅した場合、無関係な刺激によって引き寄せられた外因性の空間的注意は、課題に関連する刺激に適用され、成績に利益をもたらす可能性がある。しかし、無関係な場所でフラッシュされた場合は、助けにはならず、成績に悪影響を与える可能性がある (Berger et al, 2005)。ボトムアップ/外因性注意は時間経過が早く、妨害刺激が現れてから 80~130 ms の間、隠れた注意に影響を与える (Anton-Erxleben and Carrasco, 2013)。

注意理論の中には、隠れた空間的注意は顕在的注意を導くために存在するというものがある。特に、運動前注意理論では、サッケードを計画する神経回路と隠蔽された空間的注意を制御する神経回路は同じであると考えられている (Rizzolatti et al, 1987)。眼球運動の制御には、前頭眼野 (FEF) が関与していることが知られている。FEF のニューロンを、眼球運動を誘発するには低すぎるレベルで刺激すると、隠れた注意に似た効果が得られることが示されている (Moore et al, 2003)。このように、隠れた注意は、あからさまに見るべき場所を決めるための手段であると考えられる。目の動きは、秘密にしておいた方がよい知識や意図に関する情報を伝えるので、隠密に注意を払う能力は、社会的な種にも役立つかもしれない (Klein et al, 2009)。

隠れた空間注意の神経相関を研究するために、研究者は、注意の手がかりの違いだけに基づいて (刺激のボトムアップの特徴の違いではなく) 神経活動のどの側面が異なるかを特定する。注意が記録されたニューロンの受容野に向かって手がかり付けされる試行では、神経活動の多くの変化が観察されている (Noudoost et al, 2010; Maunsell, 2015)。一般的に報告されている所見は、発火率の増加で、典型的には 20~30% の増加である (Mitchell et al, 2007)。しかし、変化の正確な大きさは、調査する皮質領域によって異なり、後の領域ほど強い変化を示す (Luck et al, 1997; Noudoost et al, 2010)。注意は、神経発火のばらつきにも影響を与えることが知られている。特に Fano Factor で測定される試行間の変動性を減少させ、ニューロン対間のノイズ相関を減少させる。さらに、注意はニューロンの電気生理学的特性にも影響を与え、パースト的に発火する可能性を減少させ、個々の活動電位の高さを減少させることがわかっている (Anderson et al, 2013)。

一般に、注意に伴う変化は、注意した刺激を表すニューロンの信号対雑音比を増加させると考えられているが、脳領域間のコミュニケーションにも影響を与える可能性がある。この目的のためには、注意が神経の同期に及ぼす影響が重要である。視覚野では、注意によって、ガンマ帯 (30~70Hz) のスパイクのコヒーレンスが高まることが示されている (Fries et al, 2008)。ニューロンのグループが同期して発火すると、下流の共有領域に影響を与える能力が高まる。さらに、注意が領域間のコミュニケーションを直接調整している可能性もある。2 つの視覚野の間の同期活動は、コミュニケーションが活発になっていることを示すことがある。例えば、注意は V1 野と V4 野の参加刺激を表すニューロン間の同期を高めることが示されている (Bosman et al, 2012)。この領域間の同期の制御は、視床枕によって行われているようだ (Saalmann et al, 2012)。

注意が視覚経路のニューロンにどのような影響を与えるかを調べるだけでなく、トップダウンの注意の源を探る研究も行われている (Noudoost et al, 2010; Miller and Buschman, 2014)。ボトムアップの注意の処理は、外側頭頂内野 (LIP) で生成される顕著性地図に集約されるようである。この領域の細胞は、課題とは無関係だが重要な妨害物を含む、重要な刺激が受容野にあるときに反応する。一方、FEF のような前頭前野は、空間的注意のトップダウン制御に必要な信号を格納しているようで、妨害にはあまり反応しない。

感覚的注意の神経相関に関する研究の多くは大脳皮質を中心に行われているが、皮質下の領域も注意制御や成績向上に強い役割を果たしているようである。特に、上丘は隠れた、および顕在的空間的注意を助け、この領域の不活性化は注意を損なう可能性がある (Krauzlis et al, 2013)。また、上述したように、上丘は、特に大脳皮質に対するゲーティング効果に関して、注意に役割を果たしている (Zhou et al, 2016)。

## 2.2.2. 視覚的特徴への注意

特徴注意は、隠れた選択的注意の一種である。特徴注意では、特定の場所に注意を向けるのではなく、特定の色、形、方向などの視覚的特徴に注意を向けるように試行ごとに指示される。課題の目的は、手がかりとなる特徴が画面上に存在するかどうかを検出したり、その特徴の別の性質を読み取ったりすることである (例えば「正方形の色は何か?」への答えは、まず正方形に注意が向くはずである)。このように、参加した特徴についての有効な手がかりは、成績を向上させます。例えば、ある特定の方向に注意を向けると、その方向の微弱なグレーディングを他の方向よりもよく検出することができた (Rossi and Paradiso, 1995)。全体的な課題 (例: 方向性のある縞模様の検出) は変わらないが、特定の指示 (90° の縞模様の検出 vs. 60° vs. 30°) は、個々

の試行ごとに、あるいは場合によってはブロック毎にキューイングされる。試行毎のキューに成功したことで、この形式の注意は速いタイムスケールで柔軟に展開できることがわかった。

視覚探索課題も、特徴に基づく注意を活性化すると考えられている(図2)。この課題では、スクリーン上に配列された刺激が表示され、被験者は手がかりとなる刺激の位置を目の動きで示す必要がある(頻繁に目を動かす)。被験者は通常、タスク中にサックードをしながら手がかりとなる刺激を探すことができるので、このタスクは、隠れた特徴に基づく注意と顕在的な注意を組み合わせたものである。実際、サックード選択に関する領域であるFEFでは、トップダウンの特徴に基づく注意のシグナルが見つっています(Zhou and Desimone, 2011)。例えば、複数の黒い図形の中に1つだけ赤い図形があるとすぐに注目されるように、ある種の特徴はポップアウト効果を生み出すことがあるため、視覚探索課題ではボトムアップ的な注意が働くことになり、課題によってはそれを抑制する必要がある(Wolfe and Horowitz, 2004)。

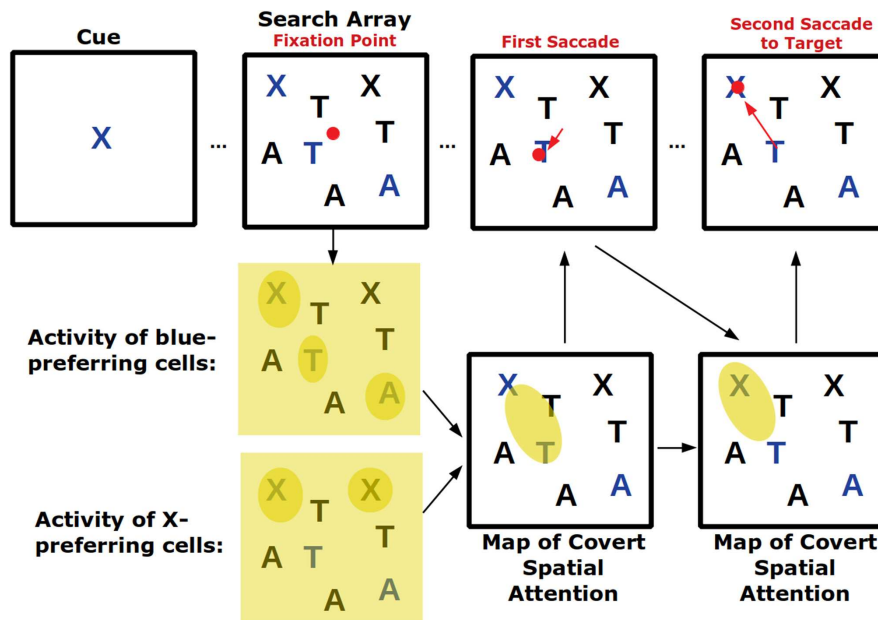


図2 視覚探索課題では、さまざまな形で視覚的な注意が働く。上列は、視覚探索課題の進行を示している。まず、視覚探索の対象を示す手がかり(この場合青いX)が表示される。次に、多くの非標的を含む探索配列が現れる。青やXの形を表す細胞にトップダウンの特徴的な注意を向けると、視覚野全体で発火が増加するが、青やXが実際に存在する場所で発火が最も強くなる。これらの神経反応は、サックード前に視覚空間を探索するために使用できる隠れた空間注意地図を生成する役割を果たす。最初のサックードで顕在意識がシフトした後、隠れた意識マップが作り直される。最後に、ターゲットの位置を確認し、サックードに成功する。もし、視覚的配列にポップアウトした刺激(例えば緑のO)が含まれていたなら、ボトムアップ方式で隠れた空間的注意を捉え、さらに誤ったサックードを引き起こしたかもしれない。

視覚系における特徴に基づく注意の神経効果は、一般的に空間的注意のそれと似ている。例えば、注意された特徴を表すニューロンは発火率が増加し、非常に異なる特徴を表すニューロンは発火率が抑制される(Treue and Trujillo, 1999)。しかし、空間的注意とは対照的に、特徴に基づく注意は空間的に大域的である。つまり、特定の特征に注意を向けると、視覚空間のあらゆる場所でその特徴を表すニューロンの活動が調整される(Saenz et al., 2002)。空間的注意と特徴的な注意のもう一つの違いは、トップダウンの注意源が視覚系の正しいニューロンをどのように標的にするかという問題である。網膜トピック地図では、近くの細胞が近くの空間的な位置を表しているため、空間的なターゲティングは簡単だが、好ましい視覚的特徴に応じて細胞がきれいに整理されているわけではない。

空間的注意と特徴的な注意の効果は相加的であると考えられる(Hayden and Gallant, 2009)。さらに、特徴注意も空間注意も、視覚野の分割正規化を実行する局所神経回路に作用することで効果を生み出すと考えられている(Reynolds and Heeger, 2009)。モデリングの研究では、トップダウン接続がこれらの回路の細胞に対象となるシナプス入力を提供すると仮定することで、選択的注意の神経効果の多くを捉えることができることが示されている(Lindsay et al., 2019)。しかし、神経調節物質であるアセチルコリンの効果に依存するモデルでも、注意の神経相関を再現することができる(Sajedin et al., 2019)。

トップダウンの特徴に基づく注意の潜在的な源は、持続的な活動が出席した特徴をエンコードする前頭前野に見出されている(Bichot et al., 2015; Paneri and Gregoriou, 2017)。腹部前頭葉領域を非活動化すると、探索課題の成績が低下する。前頭前野からは、上位の視覚野が下位の視覚野に入力を送るという逆階層的な方法で注意信号が伝わると考えられている(Ahissar and Hochstein, 2000)。

特徴注意と密接に関連する話題として、物体注意がある。この場合、注意は視覚刺激の前に抽象的な特徴に展開されるのではなく、視覚シーンの中の特定の物体に適用される(Chen, 2012)。視覚階層の活動の最初のフィードフォワード経路は、背景との明確で顕著な違いを持つ対象物があれば、事前に対象物を背景から分離して、視野内に並列に配置することができる。背景との違いが顕著であれば、対象物を背景から並列に分離することができる。しかし、より混雑した複雑な視覚シーンでは、異なる物体を識別するために、再帰的で連続的な処理が必要になる(Lamme and Roelfsema, 2000)。連続的な処理とは、限られた注意力を画像内のある場所から別の場所に移動させることである。連続処理は、限られた注意資源を画像内のある場所から別の場所に移動させることであり、隠蔽的または顕在的な空間的注意の移動という形で行われる(Buschman and Miller, 2009)。図形と地面の分離や物体の識別には、視覚系の再帰的な接続、すなわち、同じ視野の近くのニューロンからの水平方向の接続と、より高い視野のニューロンからのフィードバック接続が役立つ。脳がどのようにして低レベルの特徴をグループ化し、一貫した物体識別を行うのかという問題は、1世紀近くわたって研究されてきた。グループ化には、特に新規の物体や複雑な物体に対して注意が必要であると考えられている(Roelfsema and Houtkamp, 2011)。このことは、複数の特徴の組み合わせによって定義される物体を探す必要がある視覚探索課題において、特に重要であると考えられる。

神経学的には、対象ベースの注意効果は、対象の部分が精神的にトレースされることで、空間にゆっくりと広がっていく(Roelfsema et al., 1998)。対象物の外側に注意を移すことは、同じ距離にある対象物の内側に注意を移すことよりも大きなコストがかかるようである(Brown and Denney, 2007)。ま



た、視覚対象に注意が向けられると、その対象のさまざまな特徴に対する特徴ベースの注意が視野内で活性化されると考えられている (O'Craven et al, 1999)。

もう1つの注意は、特徴的な次元全体に注意を向けることで、特徴的な注意と呼ばれることがある。例えば、ストループ検査では、色の名前が異なる色のインクで書かれており、被験者はその言葉を読むか、インクの色を言う必要がある。ここでは、あらかじめ特定の特徴に注意を向けることはできず、単語や色という次元にのみ注意が向けられる。神経学的には、次元の切り替えは視覚経路の感覚コーディングに影響を与え、前頭葉で制御されているようである (Liu et al, 2003)。

### 2.2.3. 視覚的注意の計算モデル

視覚的注意は、注意の神経科学において最も研究されているテーマの一つであり、注意がどのように機能するかについての多くの計算モデルを刺激してきた。一般に、これらのモデルは、様々な神経生理学的知見を統合して、注意の行動上の影響がどのように生じるかを説明するのに役立つ (Heinke and Humphreys, 2005)。

顕著性を計算するためのいくつかの計算モデルが考案されている (Itti and Koch, 2001)。これらのモデルは、低レベルの視覚特徴検出器 (通常、視覚系のもと同じように設計されている) を用いて、画像固有の顕著性地図を作成し、同じ画像に対する人間のサッケードパターンを予測することができる。また、情報理論的な第一原理に基づいて顕著性を計算する別のアプローチも検討されており、ある種の視覚探索行動を説明することができた (Bruce and Tsotsos, 2009)。

注意の行動と神経の相関関係は、注意がボトムアップであってもトップダウンであっても同様のものがある。注意のバイアス競争モデルでは、刺激は神経反応を支配するために互いに競争する (Desimone, 1998)。注意 (ボトムアップまたはトップダウン) は、この競争を注意の対象である刺激に向けて偏らせることで機能する。偏り競争モデルは、単に直感を導くための「言葉モデル」として使われることもあるが、これを明示的に計算機上で具体化したものも作られている。トップダウンバイアスと、水平結合を介した局所的な競合を含む視覚経路の階層モデルは、注意の複数の神経効果を再現することができた (Deco and Rolls, 2004)。

同様の原理でスパイクニューロンを用いたモデルも実装された (Deco and Rolls, 2005)。

同様のモデルは、上述のストループ検査のような属性命名課題を扱うために明示的に構築されている。例えば、選択的注意モデル (Selective Attention Model, SLAM) は、感覚符号化モジュールと運動出力モジュールの両方に局所的な競争があり、ストループ検査のような簡単な検査と難易度の高い検査における応答時間の既知の特性を模倣することができる (Phaf et al., 1990)。

視覚は、ベイズ推論の問題として組み立てられ、モデル化されている (Lee and Mumford, 2003)。この文脈では、注意は、推論がより困難な状況下で、主に事前知識を調整することによって不確実性を解決するのに役立つ (Rao, 2005)。例えば、Chikkerur ら (2010) では、空間的注意は物体の同一性に関する不確実性を低減するように機能し、特徴的注意は空間的不確実性を低減するように機能している。これらの原理は、注意の行動的特徴と神経的特徴の両方を捉えることができ、生物学的にインスパイアされた神経モデルに実装することができる。

注意の特徴類似性利得モデル (FSGM) は、特徴領域と空間領域の両方に適用できる、トップダウン注意の神経効果の説明である (Treue and Trujillo, 1999)。このモデルでは、注意によってニューロンの反応がどのように変化するかは、そのニューロンのチューニングに依存するとしている。チューニングとは、ニューロンが異なる刺激に対してどのように反応するかを示すもので、FSGM によれば、例えば青という色を好む (つまり強く反応する) ニューロンは、トップダウンで青に注目するとその活動が増強されることになる。また、FSGM では、好まない刺激に注意を向けると発火が減少し、増加しても減少しても、活動は注意によって乗算されるとされている。当初は計算モデルとして定義されていなかったが、その後、このような形の神経変調は、困難な視覚課題の成績を高めるのに有効であることがモデリングによって示されている (Lindsay and Miller, 2018)。

他のモデルでは、注意をネットワークを介した情報の動的なルーティングとして概念化している。この形式の注意の実装は、Selective Attention for Identification Model (SAIM) (Heinke and Humphreys, 2003) に見られる。ここでは、注意は網膜から「注意の焦点」とみなされる表現に情報をルーティングする。現在の課題に応じて、網膜表現の異なる部分が注意の焦点にマッピングされる。

### 2.2.4. 他の感覚モダリティにおける注意

聴覚における選択的注意の必要性を示す有名な例として「カクテルパーティー問題」がある。複数の話者や他の雑音で混雑した部屋の中で、一人の話者の音声に集中することが難しいという問題である (Bronkhorst, 2015)。この問題を解決するためには、ピッチなどの音声の低レベルの特徴を用いて、どの聴覚情報をさらに言語処理に回すかを決定する「早期」選択が必要だと考えられている。興味深いことに、選択的聴覚注意は、聴覚処理の最も初期のレベルである蝸牛でさえも神経活動を制御する能力を持っている (Fritz et al, 2007)。

体性感覚においても、空間的注意と特徴的注意が研究されている。体のさまざまな場所でタップを期待する手がかりを与えられた被験者は、その手がかりが有効であった場合、感覚をよりよく検出することができる。しかし、これらの効果は、視覚系での効果に比べて弱いようである (Johansen-Berg and Lloyd, 2000)。検出課題において、被験者が指上の刺激の向きを手がかりにした場合、反応時間が速くなる (Schweisfurth et al, 2014)。

キューイングされた味を検出する能力をテストした研究では、有効にキューイングされた味は、無効にキューイングされた味よりも低濃度で検出できることが示された (Marks and Wheeler, 1998)。これは、特徴に基づく視覚的注意で見られる行動効果を模倣したものである。嗅覚の特徴に対する注意については、あまり詳しく調べられていないが、視覚的に誘導された香りに対する期待は、その検出を助けることがある (Gottfried and Dolan, 2003; Keller, 2011)。

また、複数の感覚信号の統合を必要とする課題を実行するために、注意を複数のモダリティに分散させることもできる。一般的に、一致する複数の感覚信号を使用することは、単一のモダリティのみに頼る場合と比較して、物体の検出を助ける。興味深いことに、他の領域からの信号が同じように有効であっても、人間は視覚領域にバイアスをかけている可能性を示唆する研究もある (Spence, 2009)。具体的には、手がかりの空間的な位置を特定する必要がある課題では、視覚領域が最も優位になるようである (Bertelson and Aschersleben, 1998)。腹話術では、ダミーの口が動いているという視覚的な手がかりが、声源の真の位置に関する聴覚的な証拠よりも優先されることがよくわかる。また、視覚的証拠は、例えば、ゴム腕錯覚のように、触覚的証拠よりも優先されることがある (Botvinick and Cohen, 1998)。

感覚処理のクロスモダルな性質のもう一つの効果は、あるモダリティでの注意の手がかりが、別のモダリティでの注意の方向付けを引き起こすことである (Spence and Driver, 2004)。一般に、手がかりのないモダリティでの注意効果は弱くなる。このようなクロスモダルな相互作用は、内因性 (トップダウン) の注意と外因性 (ボトムアップ) の注意の両方の文脈で起こる可能性がある。

## 2.3. 注意と実行制御

複数の課題が同時に競合する場合、どの課題をいつ実行するかを決定する中央管理者が必要となる。さらに、課題をどのように実行するのがベストなのかは、履歴や文脈によって異なる。感覚入力と過去の知識を組み合わせ、効率的なタスクの選択と実行という仕事のために、複数のシステムを調整することが実行制御の役割であり、この制御は通常、前頭前野と関連している (Miller and Buschman, 2014)。前述したように、トップダウンの視覚的注意の源もまた、前頭葉領域に位置している。注意は、実行制御の出力として合理的に考えることができる。したがって、実行制御システムは、注意の対象を選択し、それを実行するシステムに伝えなければならない。前述の逆階層理論によると、上位領域が入力を得た領域に信号を送り、その信号が下位の領域に送られる、ということになる (Ahissar and Hochstein, 2000)。つまり、それぞれのポイントで、注意の指示を、対象となる領域にとって意味のある表現に変換する必要がある。このようなプロセスを経て、実行制御領域のハイレベルな目標は、例えば初期の感覚処理などの非常に具体的な変化をもたらすことができる。

また、過去の情報を活用したり、現在の目標を維持したりするためには、ワーキングメモリが必要であることから、実行制御とワーキングメモリは相互に関連しています。さらに、ワーキングメモリは、前頭前野の持続的な活動として認識されることが多い。実行制御、ワーキングメモリ、注意の三者間の関係の結果として、課題にとって望ましくない場合でも、ワーキングメモリの内容が注意に影響を与えることがある (Soto et al, 2008)。例えば、ある物体をワーキングメモリに保持しながら、同時に別の物体を視覚的に検索しなければならない場合、検索配列に記憶された物体が存在すると、検索に悪影響を及ぼす可能性がある (Soto et al, 2005)。このことは、ワーキングメモリが注意の実行制御を妨害する可能性を示唆している。しかし、注意の実行制御には、ワーキングメモリだけでは妨害されない要素がまだあるようである。このことは、被験者が記憶した項目を報告する必要があると思っているのに、代わりに出席した項目の検索配列を見せられると、視覚探索の成績がさらに低下するという研究に見られる (Olivers and Eimer, 2011)。このことは、ワーキングメモリ内のすべての対象が注意に何らかの影響を及ぼす可能性がある一方で、実行制御者はどの対象が最も影響を及ぼすかを選択できることを示唆している。

感覚モダリティの中で注意を柔軟にコントロールするだけでなく、モダリティ間で注意をシフトさせることもできる。行動実験によると、感覚モダリティ内の2つの異なる課題の間で注意を切り替える (例えば、視覚的な対象の位置を特定することから、それを識別することへ)、あるいは感覚モダリティ間で注意を切り替える (聴覚的課題から視覚的課題へ) と、計算コストがかかることがわかっている (Pashler, 2000)。このコストは通常、課題を切り替えた直後の試行と、同じ課題を繰り返している試行とで、成績がどの程度低下するかで測定される。興味深いことに、モダリティ内での課題の切り替えは、モダリティ間の切り替えよりも大きなコストがかかるようです (Murray et al., 2009)。同様の結果は、応答のモードを切り替えた場合にも見られ (例えば、ボタン押し対言語報告)、これは感覚処理に特有のものではないことが示唆されている (Arrington et al, 2003)。このような知見は、モダリティ内での切り替えには、同じ神経回路の再構成が必要であり、単に異なる感覚系の回路を動かせるよりも困難であることに起因すると思われる。これは、異なる感覚システムの回路を単に動かすよりも難しいことである。効率的な実行制御者は、注意を移すことを決定する際に、このようなコストを認識し、理想的には最小化しようとする必要がある。スイッチコストは訓練によって低減できることが示されている (Gopher, 1996)。

注意の実行制御に関する最後の疑問は、学習によって注意がどのように変化するかということである。眼球運動の研究によると、検索された項目は、新規のものよりも馴染みのある設定でより迅速に検出できることが示されており、以前に学習した連想が顕在的な注意を導くことが示唆されている (Chun and Jiang, 1998)。このような利点は、海馬に依存していると考えられている (Aly and Turk-Browne, 2017)。しかし一般的には、注意を向ける方法の学習は、注意過程の他の側面ほど研究されていない。いくつかの研究では、被験者が無関係な課題情報を抑制する能力を高めることができ、その抑制の一般性はトレーニング手順に依存することが示されている (Kelley and Yantis, 2009)。注意学習の神経相関を見ると、イメージングの結果から、学習に伴う神経の変化は感覚経路そのものではなく、むしろ注意制御に関連する領域で起こることが示唆されている (Kelley and Yantis, 2010)。研究は必ずしも容易ではありませんが、乳幼児期や小児期における注意システムの発達は、注意がどのようにして学習されるのかについて、さらなる手がかりを与えてくれるかもしれません (Reynolds and Romano, 2016)。

## 2.4. 注意と記憶

注意と記憶には、さまざまな相互作用が考えられる。例えば、記憶の容量が限られているのであれば、脳が何を記憶に入れるかを選択することは理にかなっている。このようにして、全情報のサブセットを動的に選択する注意の能力は、記憶システムのニーズによくマッチしている。逆に言えば、特定の記憶を思い出すことは、限られた資源をどのように使うかという選択である。したがって、記憶の符号化と検索の両方が注意に頼ることができる。

記憶の符号化における注意の役割はかなり強いようである (Aly and Turk-Browne, 2017)。情報が適切に記憶に符号化されるためには、その情報が注意の対象であることが最適である。

被験者に単語のリストを記憶させながら、同時に注意を分散させる二次的な課題に従事させると、後でそれらの単語を意識的に思い出す能力が損なわれる (ただし、その単語を馴染みのあるものとして認識する能力はそれほど影響を受けない; Gardiner and Parkin, 1990)。二次課題の難易度を上げると、左腹下前頭回と前海馬における記憶符号化に関連する活動パターンが弱まり、背外側前頭前野と上頭頂領域における二次課題情報の表現が増加することが、イメージング研究で示されています (Uncapher and Rugg, 2005)。したがって、符号化という課題に置かれた限られた神経処理能力がなければ、記憶は損なわれます。注意もまた、空間的に定義された記憶の符号化に関与しており、場所細胞の表現を安定化させるようである (Muzzio et al, 2009)。

暗黙の統計的学習は、注意によってもバイアスがかかる。例えば、Turk-Browne ら (2005) では、被験者は赤と緑の図形からなる刺激の流れを見ていた。課題は、注意した色図形が2回連続して現れたときに検出することであった。被験者には知られていなかったが、刺激の流れには統計的な規則性があり、3つの図形が近接して出現する可能性があった。被験者は、3つの形のセットを2つ見せられたとき、1つは実際に共存する3つの形、もう1つは同じ色の形をランダムに選択したもので、実際の3つの形の方が親しみやすいと認識した。しかし、被験者は、参加した色の3つの図形である場合にのみ、本物の3つの図形をよりよく知っていると認識した。

しかし、意識的な注意を払わなくても、学習が行われることがある。例えば、Watanabe (2003) では、患者は視野の中心にある文字を検出する課題に取り組み、背景には閾値以下のコントラストでランダムドットの動きが表示されていた。この動きは、現在表示されている文字と相関のある方向に10%のコヒーレンスを持っていた。この課題を学習する前後に、被験者は閾値以上の方向分類課題を行った。この課題を学習した後、方向分類は、対象となる

文字に関連する方向に対してのみ向上した。これは、ターゲットによって活性化された報酬関連信号が、刺激の非注目要素についての学習につながったことを示唆している。

多くの行動研究では、記憶の想起にどの程度の注意が必要かが検討されている。例えば、記憶した単語のリストを思い出すと同時に、カードを並べ替えるなどの二次的な作業をさせることで、記憶の想起が、作業と同じ限られた注意資源から得られるかどうかを調べることができる。このような研究の中には、注意を必要とする課題を同時に行うと記憶の回復が阻害されるというものがあり、記憶の回復が注意に依存した過程であることを示唆している。しかし、正確な所見は、使用する記憶課題と非記憶課題の詳細に依存している (Lozito and Mulligan, 2006)。

もし、記憶の検索が共有された注意資源から得られるものではないとしても、ある瞬間に他の記憶よりも鮮明に検索される記憶が選択されることは明らかである。そのためには、選択のプロセスが必要である。神経イメージングの結果を見ると、注意のトップダウン的な配分とボトムアップ的な捕捉を司る頭頂部の脳領域が、記憶想起の際にも同様の役割を果たしていることが示唆されている (Wagner et al, 2005; Ciaramelli et al, 2008)。

記憶検索の研究は通常、中長期記憶を対象としているが、ワーキングメモリ内の項目に注意を向けるメカニズムも提案されている (Manohar et al.2019)。それは、ワーキングメモリの2つの異なるメカニズムに依存している。非注意項目のためのシナプスの痕跡と、注意項目のための持続的な活動である。

記憶の中には、感覚処理の流れそのものの中で自動的に行われるものもある。プライミングとは、ある時点での刺激の存在が、その後の刺激の処理や解釈に影響を与えるという心理学の有名な現象である。例えば、「school」という単語よりも「hospital」という単語の方が「doctor」という単語よりも早く認識される場合がある。このように、プライミングには、過去の刺激が現在の刺激に影響を与えるための暗黙的な記憶が必要である。概念的プライミングや意味的プライミングに関するいくつかの研究では、プライミング効果が生じるためには、最初の刺激に対する注意が必要であることが示されている (Ballesteros and Mayas, 2015)。これは、一般的に記憶のエンコーディングには注意が必要であるという知見と同じである。

ほとんどのプライミングはポジティブなもので、ある時点で刺激が存在すると、その刺激や関連する刺激の検出や処理が後から行われる可能性が高くなることを意味する。このように、プライミングはボトムアップの注意を偏らせるものと考えることができる。しかし、トップダウンの注意がネガティブプライミングを引き起こすこともある。ネガティブプライミングでは、前回の試行で妨害刺激として機能していた刺激が、今回の試行で注意の対象となると、成績が低下する (Frings et al, 2015)。これは、現在のターゲットとなる刺激に対して、妨害刺激抑制のメカニズムがまだ活性化されているホールドオーバー効果に起因すると考えられる。

適応は暗黙の記憶の一形態とも考えられる。ここでは、同じ刺激に繰り返しさらされると、神経の反応が低下する。繰り返しの反応が減少することで、刺激の変化がより顕著になる。注意は、注意を受けた刺激に対する神経応答を増加させることで、適応の効果を打ち消す (Pestilli et al, 2007; Anton-Erxleben et al, 2013)。このように、プライミングでも適応でも、トップダウンの注意は、ボトムアップの注意を導いている可能性のある下位レベルで起こる自動プロセスを克服することができます。

## 3. 機械学習における注意

人工的な注意という概念は、現在の人工ニューラルネットワークの復活以前にも出てきているが、現在よく使われているのは、ANNを中心としたものである (Mancas et al, 2016)。人工ニューラルネットワークにおけるアテンションメカニズムの使用は、脳における注意の必要性が明らかになったように、ニューラルシステムをより柔軟にする手段として生まれた。機械学習における注意メカニズムは、1つの訓練された人工ニューラルネットワークが複数課題や、長さ、大きさ、または構造が変化する入力を持つ課題で優れた性能を発揮することを可能にする。機械学習における注意の精神は、確かに心理学から着想を得ているが、その実装は、後述するように、生物学的な注意について知られていることとは必ずしも一致しない。

もともと ANN 用に開発された注意の形では、注意機構は、エンコーダ-デコーダの枠組みで、系列モデルの文脈で機能していた (Cho et al, 2015; Chaudhari et al, 2019)。具体的には、入力系列がエンコーダー (リカレントニューラルネットワークと思われる) を通過し、デコーダー (これもリカレントニューラルネットワークと思われる) の仕事は別の系列を出力することになる。エンコーダとデコーダをつなぐのは、注意機構である。

一般的に、符号化器の出力は、入力系列の各要素に対して1つずつのベクトルのセットです。注意は、これらのベクトルのうち、どのベクトルを使って出力を生成すべきかを決定するのに役立つ。出力系列は1要素ずつ動的に生成されるため、注意は各時点で異なる符号化されたベクトルを動的にハイライトすることができる。これにより、符号化器は入力系列の最も関連性の高い部分を柔軟に利用することができる。

注意機構の具体的な仕事は、符号化されたベクトル ( $v^i$ ) のそれぞれに1つずつ、一連のスカラー重み付け  $\alpha_t^i$  を生成することである。各ステップ  $t$  において、注意メカニズム ( $\phi$ ) は、デコーダの前回の隠れた状態 ( $h_{t-1}$ ) と符号化されたベクトルに関する情報を取り込み、正規化されていない重み付けを生成する。

$$\hat{\alpha}_t = \phi(h_{t-1}, v)$$
(1)

注意は限られた資源であるため、これらの重み付けは相対的な重要性を表す必要がある。 $\alpha$  値の合計が1になるように、正規化されていない重み付けをソフトマックスにかける。

$$\hat{\alpha}_t^i = \frac{\exp(\alpha_t^i)}{\sum_j \exp(\alpha_t^j)}$$
(2)

これらの注意値は、符号化されたベクトルをスケーリングして、符号化器が条件付けできる単一の文脈ベクトルを作成する。

$$c_t = \sum_j \hat{\alpha}_t^j v^j$$
(3)

このような注意の仕方は、完全に微分可能なので、単純な勾配降下法でネットワーク全体をエンド・ツー・エンドで訓練することができる。

このような AI 的な注意は、反復的な再重み付けの一形態である。具体的には、前処理された入力の異なるコンポーネントを、出力生成に必要なに応じて動的にハイライトする。これにより、生物学的な注意と同様に、柔軟で状況に依存したものとなっている。そのため、本質的に動的なものでもある。

系列モデリングにはすでに時間的な要素が含まれているが、この形式の注意は静的な入力と出力にも適用することができ(後に画像処理の文脈で説明します)、その結果、モデルにダイナミクスが導入される。

注意を払わない従来の符号化と復号化の枠組みでは、符号化器は、入力の長さや特徴に依存しない固定長のベクトルを生成し、復号化の過程では静的なものとなっていた。このため、長い配列や複雑な構造を持つ配列は、短い配列や単純な配列と同じ次元で表現されることになり、復号化は復号化過程で入力のさまざまな部分に質問することができなかった。しかし、入力を入力配列と同じ長さのベクトルの集合として符号化することで、復号化の各時点で関連する入力配列の部分に選択的に注目することが可能になる。ここでも、脳における注意の解釈と同様に、AIシステムにおける注意は、限られた資源を柔軟に活用する方法として役立つ。復号化は入力のすべてを条件とすることはできないので、ある時点でボトルネックを導入しなければならない。注意のないシステムでは、固定長の符号化ベクトルがボトルネックになっていた。注意機構が追加されると、復号化が入力のどの部分に注目するかを決定する際に、ボトルネック(文脈ベクトルの形)が動的に生成されるため、復号化はより大きくなる。

このような注意機構を人工システムに追加する動機は、もちろんその性能を向上させるためである。しかし、注意のもう一つの利点は、解釈可能性であると言われている。復号化の過程で、入力のどの部分に注意が向けられているか(すなわち、どの $\alpha^i$ 値が高いか)を特定することで、復号化器がなぜそのような出力を出したかを理解できるかもしれない。ただし、注意の出力を解釈する際には、期待通りにモデルの振る舞いを説明できない場合もあるため、注意が必要である(Jain and Wallace, 2019; Wiegrefe and Pinter, 2019)。

以下の節で、この一般的な注意の概念の具体的な応用例と、この枠組みにうまく収まらない例を取り上げる。また、生物学との類似性にも注目する。

### 3.1. 自然言語処理における注意

以上のように、系列を処理するモデルには、注意機構が頻繁に追加されている。自然言語処理(NLP)は、系列モデリングの最も一般的な応用分野の一つである。また、機械学習における注意のオリジナルの領域ではなく、生物学との共通点も少ないが、NLPは注意の最も一般的な応用分野の1つでもある(Galassi et al, 2019)。

人工ニューラルネットワークにおけるこのような形の注意の初期の応用例として、翻訳課題があった(Bahdanau et al, 2014; 図3)。この仕事では、リカレントニューラルネットワークが入力文を、文中の各単語に1つずつ対応する「注釈」ベクトルのセットとして符号化する。出力であるターゲット言語の文は、リカレントニューラルネットワークによって一度に1つの単語ずつ生成される。生成された各単語の確率は、それまでに生成された単語、リカレントニューラルネットワークの隠れ状態、および注意機構によって生成された文脈ベクトルの関数である。ここで、注意メカニズムとは、出力ネットワークの隠れ状態と、現在の注釈ベクトルを取り込み、すべての注釈ベクトルに対する重み付けを行う小さなフィードフォワードニューラルネットワークである。

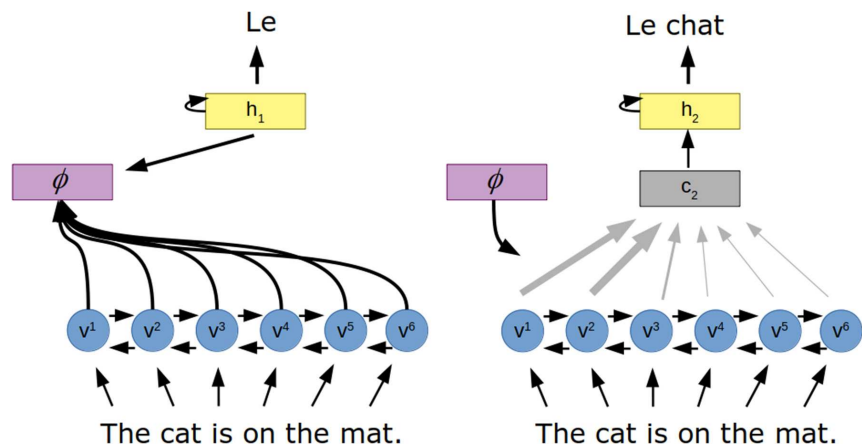


図3 ニューラル機械翻訳への注意

翻訳されるべき文は、リカレントニューラルネットワークを介して一連のベクトル( $v$ )に符号化される。注意機構( $\phi$ )は、デコーダの隠れた状態( $h$ )とこれらのベクトルを用いて、エンコードされたベクトルをどのように組み合わせて文脈ベクトル( $c$ )を生成すべきかを決定する。このベクトルは、デコーダの次の隠れた状態に影響を与え、したがって翻訳文の次の単語に影響を与える。

このようにして文中のすべての単語の情報をブレンドすることで、ネットワークは出力する単語を生成する際に、前の部分と後の部分のどちらかを利用することができる。これは、標準的な語順が異なる言語間の翻訳に特に有効である。著者らは、入力文の中で注意が適用された場所を視覚化することで、注意がこの問題に役立つことを確認した。

この最初のアプリケーション以来、言語翻訳のための注意ネットワークの多くのバリエーションが開発されてきた。Firatら(2016)では、1つの言語だけでなく、複数の言語のペア間の翻訳に使用できるように、注意機構が適応された。Luongら(2015)では、著者らは、すべての入力単語に一度にアクセスする能力が必要かどうかを判断するために、異なる注意の構造を探っている。また、Chengら(2016)では、より柔軟に文の表現を作るために、文の符号化と復号化を行うリカレントニューラルネットワークに注意機構を追加した。

2017年、影響力のある「Attention is All You Need」という論文では、機械翻訳にまったく異なるスタイルのアーキテクチャが活用された(Vaswani et al, 2017)。このモデルは再帰性を持たないので、学習がよりシンプルになる。代わりに、文中の単語が並列に符号化され、これらの符号化によってキー表現とクエリ表現が生成され、これらが組み合わされて注意の重み付けが作られる。これらの重み付けは、単語の符号化自体をスケールリングして、モデルの次の層、つまり「自己注意」として知られる過程を作る。

この過程は繰り返し行われ、最終的には自己回帰復号化と相互作用する。自己回帰復号化は、コード化された入力(標準的な注意の形)と以前に生成された出力に柔軟に焦点を合わせることができる注意機構も備えている。この新しい注意機構に与えられた名前であるTransformerは、これまでの多くのモデルを凌駕し、すぐに機械翻訳だけでなく他のタスクの標準となった(Devlin et al, 2018)。



興味深いことに、自己注意は、もともと機械翻訳に使われていたリカレント注意モデルに比べて、生物学的注意との共通点が少ない。第一に、脳は言語処理や注意選択などの逐次処理課題において必然的に再帰性に依存しているのに対し、自己注意は再帰性とダイナミクスの役割を減らしている。第二に、自己注意は単語間の水平方向の相互作用を提供する。これにより、符号化された文の中の単語は周囲の文脈の中で処理されるが、この機構には、復号化器のニーズによって駆動される明らかなトップダウンの要素は含まれていない。実際、自己注意は、ある状況下では、画像処理で頻繁に使用される標準的なフィードフォワード計算であるコンボリューションを単純に実装することが示されている (Andreoli, 2019; Cordonnier et al, 2019)。このように、自己注意は、限られたリソースに基づいて課題固有の注意のような選択を行うよりも、良い符号化を作成することが重要である。時間的課題の文脈では、心理学における最も近い類似物はプライミングかもしれない。プライミングは、先行刺激に基づいて後続刺激の符号化を変化させるからである。もちろん、機械学習エンジニアの直接的な目標は、脳を再現することではなく、課題をうまくこなせるように簡単に訓練できるネットワークを作ることである。このような制約があるため、機械学習が大きく進歩しても、必ずしも脳に近いモデルができるとは限らない。

人間の言語処理における注意の研究は、神経科学の他の分野ほど大規模ではないが、読書中の眼球運動を追跡する研究がいくつか行われている (Myachykov and Posner, 2005)。彼らは、現在読んでいる文章を明確にするために、特に代名詞の先行詞を見つけるという文脈で、文章の前の部分を振り返ることを発見した。このような注意のシフトは、現在の処理要求に最も関連する過去の情報を示している。

## 3.2. 視覚課題のための注意

神経科学や心理学と同様に、機械学習の研究の大部分は視覚的な課題で行われている。コンピュータビジョンの注意喚起ツールの1つである「顕著性マップ」は、エッジ、色、奥行きなどの低レベルの視覚的特徴に基づいて、画像内のどの領域が最も顕著であるか、またその領域が周囲とどのように異なるかを識別するものである (Itti and Koch, 2001)。このようにして、人や動物の「ボトムアップ」の注意によってどの領域が捉えられるかを示すのが、顕著性 saliency マップである。コンピュータ科学者は、画像処理パイプラインの一部として、さらなる処理のために領域を識別するために、顕著性地図を使用している。

より近年では、コンピュータビジョンモデルは深層学習に支配されている。そして、2012年のImageNetチャレンジで成功して以来 (Russakovsky et al, 2015)、畳み込みニューラルネットワークは、機械学習における視覚課題のデフォルトアーキテクチャとなっている。

畳み込みニューラルネットワークのアーキテクチャは、ほ乳類の視覚システムに大まかに基づいている (Lindsay, 2020)。各層では、フィルターのバンクが下層の活動に適用される (第1層では画像)。これにより、神経活動の  $H \times W \times C$  のテンソルが作成される。チャンネル数  $C$  は適用されたフィルターの数に等しく、 $H$  と  $W$  はフィルターを適用した結果の2次元特徴地図の高さと幅を表す。

畳み込みニューラルネットワークにおける注意は、分類、セグメンテーション、画像を利用した自然言語処理など、さまざまな課題成績を向上させるために用いられている。また、これらの注意過程は、神経科学の文献と同様に、空間的な注意と特徴に基づく注意に分けられる。

### 3.2.1. 空間的注意

NLP 課題の注意に使われる構造を基に、視覚的注意が画像キャプションに適用されている。Xu et al. (2015)では、符号化モデルは畳み込みニューラルネットワークである。注意機構は、第4の畳み込み層での活動に作用する。キャプションの各単語が生成されると、画像表現の空間的な位置に渡る重み付けの異なるパターンが作られる。このようにして、キャプション生成時の注意は、翻訳課題における符号化された単語ベクトルのセットを、符号化された画像位置のセットに置き換える。このモデルは、高い重みを持つ場所を視覚化することで、キャプションを生成している現在の単語に最も関連する対象に注意するように見える。

このスタイルの注意は、空間的な位置に対する視覚的な特徴の重み付けされた組み合わせを生成するため、「ソフト」と呼ばれている (図4B)。「ハード」な注意は、他のすべてを犠牲にして復号化に渡す1つの空間位置を選択する代替形態である (図4A)。Xuら (2015)では、どの場所がこのハード注意を受けるべきかを決めるために、各空間場所に対して生成された注意の重みを確率として扱った。これらの確率に応じて1つの場所が選択される。この確率的な要素をネットワークに加えることで、訓練はより難しくなるが、ソフトな注意よりもやや良好なパフォーマンスが得られることがわかった。

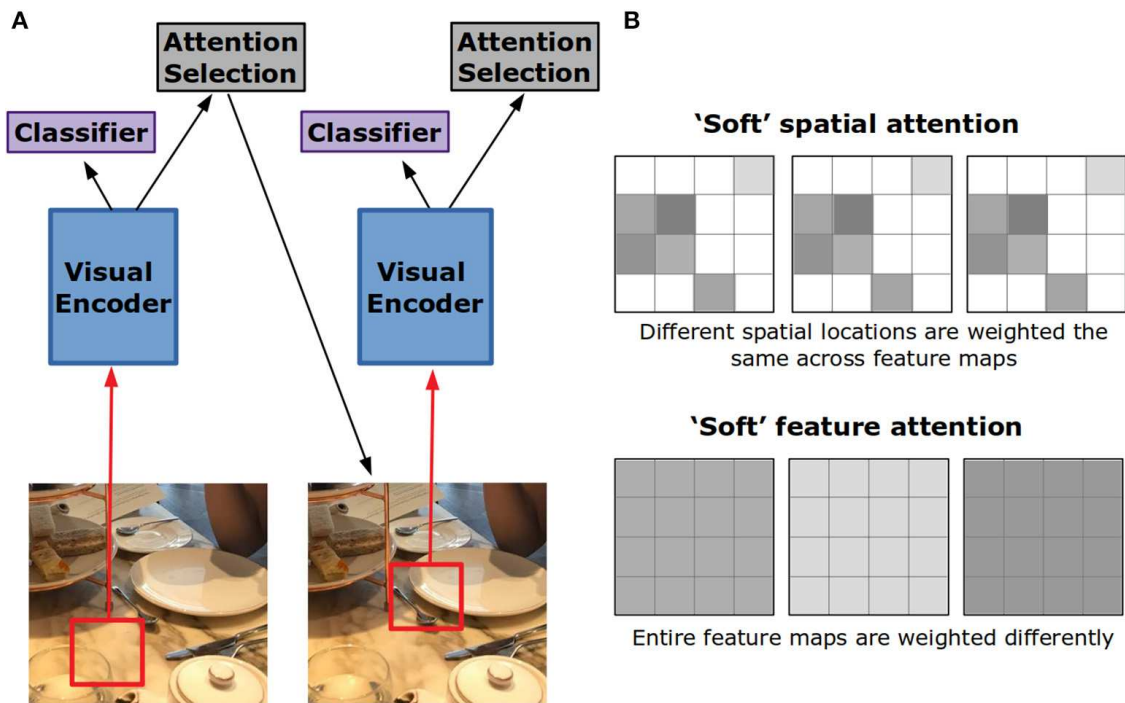


図4 人工ニューラルネットワークにおけるハードな視覚的注意とソフトな視覚的注意。(A) ハードアテンションでは、ネットワークは画像全体のごく一部からしか入力を得られない。この部分は、注意選択機構を介して、ネットワークによって反復的に選択される。入力が中心窩にあれば、ネットワークは低解像度の周辺部を使ってこの選択を導くことができる。(B) 畳み込みニューラルネットワークの特徴地図は、下層にフィルタを適用することで作成される活性化の2次元グリッドである。ソフトな空間注意では、このグリッド上の異なる位置が異なる重みを持つ。ソフトな特徴注意では、異なる特徴地図が異なるように重み付けされる。

2014年の研究では、強化学習を用いてハードな注意ネットワークを訓練し、厳しい条件下で物体認識を行うことができた (Mnih et al, 2014)。このモデルの中核となるのはリカレントニューラルネットワークで、ネットワークが行う複数の「チラ見 glimpses」で取り込んだ情報を記録するとともに、次のチラ見の位置を出力する。チラ見するたび、ネットワークは画像の小さなパッチから中心窩のような入力 (中心部は高解像度、周辺部は低解像度で表現される) を受け取る。ネットワークは、これらのチラ見から得られた情報を統合して、画像内の物体を見つけ出し、分類しなければならない。これは上述のハードな注意と似ているが、ここでの場所の選択は、画像のどの部分を次にサンプリングするかを決定する (上のケースでは、すでに処理された画像のどの場所を復号化に渡すかを決定していた)。これにより、ネットワークは画像のすべてを処理する必要がなくなり、計算機資源を節約することができる。また、画像内に複数の物体が存在し、ネットワークがそれぞれを分類しなければならない場合にも有効である (Ba et al 2014)。最近の研究では、事前訓練のステップを追加することで、複雑な画像に適用されるハードな注意の成績が向上することが示されている (Elsayed et al, 2019)。

生物学的な注意と人工的な注意の対応は、多くの点で、視覚的な空間的注意の場合に最も強く現れる。例えば、画像のさまざまな場所を順次サンプリングして処理するハードな注意は、サッケードの過程を再現したもので、神経科学や心理学の文献にある「オバートな視覚的注意」に似ている。ソフトな注意は、ネットワークへの入力に変化を与えることなく、ネットワークの画像表現の異なる領域を動的に再重み付けするという点では、隠れた空間的注意に似ている。また、ソフトな注意の適用方法は、特定の場所にあるすべてのユニットの活動を乗算的にスケーリングすることであるため、隠蔽された空間的注意に関する神経学的な知見を再現している。

ソフトな空間的注意は、視覚的な質問と回答 (Chen et al, 2015; Xu and Saenko, 2016; Yang et al, 2016) や、動画でのアクション認識 (Sharma et al, 2015) など、他の課題にも使用されている。また、ハードな注意は、実体セグメンテーション (Ren and Zemel, 2017) や、異なるレベルの画像解像度を用いて適用した場合のきめ細かな分類 (Fu et al, 2017) にも用いられている。

### 3.2.2. 特徴への注意

ソフトな空間的注意の場合、重みは画像表現の異なる空間的位置で異なるが、その位置にあるすべての特徴チャンネルでは同じである。つまり、異なる視覚的特徴を表すネットワーク内のユニットの活動は、画像空間内の同じ場所を表していれば、すべて同じように変更されることになる。特徴的注意は、個々の特徴地図を動的に再重み付けすることで、特徴の処理を空間的に大域的に変化させることを可能にする。

Stollenga et al. (2014) では、畳み込みニューラルネットワークに特徴ベースの注目メカニズムを搭載している。画像が標準的なフィードフォワードアーキテクチャを通過した後、ネットワークの活動は、異なる層の異なる特徴地図をどのように重み付けすべきかを決定するポリシーに渡される。この再重み付けは、異なるネットワーク活動をもたらし、それが異なる再重み付けにつながる。ネットワークがいくつかのタイムステップで実行された後、最終層の活性が画像内の対象を分類するために使用される。重み付けの値を決定するポリシーは強化学習によって学習され、事前に学習された任意の畳み込みニューラルネットワークに追加することができる。

Chen et al. (2017) のモデルは、特徴的な注意と空間的な注意を組み合わせ、画像のキャプション付けを支援する。畳み込みネットワークのフィードフォワードパスの活性は、以前に生成された単語とともに注意機構に渡され、CNNの各層で異なるチャンネルに対する注意の重み付けを作成する。これらの重みは活性値をスケーリングするために使用され、次に別の注意機構が空間的重みを生成するために同じ手順を行う。空間と特徴の両方の注意重みが生成され、各時点でネットワークに適用される。

De Vriesら（2017）のモデルでは、質問の内容を利用して、視覚的な質問と回答の課題のために CNN が画像を処理する方法を制御する。具体的には、言語埋め込みネットワークのアクティビティを多層パーセプトロンを通して、CNN の各チャンネルのバッチ正規化のための加算パラメータを生成する。この手順は条件付き一括正規化と呼ばれ、質問に依存した特徴の注意の形として機能する。

動的な特徴の再重み付けの異なる形が「squeeze-and-excitation」ネットワークに現れる (Hu et al, 2018)。このアーキテクチャでは、異なるチャンネルに適用される重み付けは、同じ層の他のチャンネルの活動の非線形関数である。前述の「自己注意」と同様に、これは、重み付けがネットワークの後半の活動の関数であり、かつ/または出力生成器のニーズによって偏っている、より「トップダウン」なアプローチとは根本的に異なる。生物学的には、このような相互作用は、分割正規化などの計算を行うことが知られている視野内の水平結合に最もよく似ている (Carandini and Heeger, 2012)。

特徴に基づく注意の生物学的研究では、被験者は通常、特定の視覚的特徴に注意を向けたり検索したりするように合図される。この方法では、注意すべき特徴は事前に知られており、特定の副次課題に関連している (例えば、一般的な形状検出課題の所定の試行で特定の形状を検出する)。これは、上記の人工的な特徴注意の例とは異なり、特定の画像に関する知識が得られる前に、外部の手掛かりがネットワーク処理を偏らせることはない。むしろ、特徴の再重み付けは画像自体の機能であり、一定の課題におけるネットワークの成績を向上させることを目的としている (これは、前述の人工的な空間的注意の形態の場合にも当てはまる)。

生物学的注意の研究にキューパラダイムを用いる理由は、実験者が注意をどこに置くかをコントロールできる (つまり、知ることができる) からである。しかし、明示的な手がかりがなくても、私たちの脳は常にどこに注意を置くかを決定していることは明らかである。これらは、視覚系への局所적および長距離的なフィードバック接続によって媒介されていると考えられる (Wyatte et al, 2014)。したがって、生物学的特徴注意の研究と AI システムでの使用とでは課題構造が異なるが、この違いは表面的なものに過ぎないかもしれない。基本的に、AI システムは、トップダウンの情報を手がかりの形で与えられるのではなく、フィードフォワードの画像情報を使って、トップダウンの注意信号を内部で生成している。

とはいえ、人工的なシステムの中には、外部からキューされた特徴に注意を払うことができるものもある。例えば、Cao et al. (2015) のネットワークでは、カテゴリに対する事前処理を設定することで、特定のカテゴリのローカライズがうまくいくようになっている。Wang ら (2014) のネットワークは、量み込みではないが、特定の物体カテゴリの検出を偏らせる手段を持っている。また Lindsay and Miller (2018) では、手がかり付き物体検出課題中の生物学的特徴の注意のいくつかの性能と神経的側面を CNN を用いて再現した。Luo ら (2020) では、CNN で手がかり付き注意の形式を使うことのコストとメリットが検討された。

上述したように、活動の乗法的スケーリングを使用することは、生物学的な視覚的注意から得られたある知見と一致している。さらに、特徴地図全体を同じスカラー値で変調することは、特徴注意が視覚系において空間的に大域的に作用するという前述の知見と一致する。

### 3.3. マルチタスク注意

マルチタスク学習は、機械学習の中でも特に難しいテーマである。例えば、物体分類、エッジ検出、顕著な領域の特定を行う CNN のように、1 つのネットワークに複数の異なる課題を要求された場合、個々の課題を実行するために必要な重みが互いに矛盾してしまうため、学習が困難になる。ひとつの方法は、課題固有のパラメータセットを用意して、共有ネットワークの活動を課題ごとに異なるように調整することである。常にそうと呼ばれるわけではないが、ネットワークの機能を柔軟に変化させることから、これは合理的には注意の一形態と考えられる。

Maninis ら (2019) では、共有されたフィードフォワードネットワークを複数課題のすべてで学習させる一方で、課題に特化したスキップ接続やスクイーズ・励起ブロックは、その特定の課題でのみこの活動を変調するように学習させている。これにより、ネットワークは、全課題に共通する処理を共有することで利益を得つつ、それぞれの課題に多少特化することができる。

Rebuffi ら (2017) では、同様の手順で、複数の異なる画像ドメインで分類を行うネットワークを作成した。ここでは、入力画像からドメインを特定し、課題固有のパラメータセットを実行時に自動的に選択することができた。

Zhao ら (2018) では、同じ画像をネットワークに渡して、異なる次元 (例えば、写真の人物が笑っているかどうか、若いか古いかな) に沿って分類することができる。これらの異なる分類を実行するために、特徴チャンネルの課題固有の再重み付けが使用される。

Strezoski ら (2019) のモデルでは、ハードな特徴注意の一形態と解釈できるものを用いて、課題ごとに異なる情報をルーティングしている。特徴チャンネル上の 2 値化マスクは、課題ごとにランダムに選ばれる。これらのマスクは、全課題での訓練中とランタイムに課題固有の方法で適用される。このネットワークでは、課題固有の注意パラメータは学習されない。なぜなら、これらのマスクは事前に決定され、訓練中に固定されるからである。代わりに、ネットワークは、異なる課題を実行するために、結果として生じる異なる情報経路を使用することを学習する。

最近の研究では、課題固有のパラメータの概念は完全に取り払われた (Levi and Ullman, 2020)。その代わりに、フィードフォワード CNN の活性化を課題入力と組み合わせて第 2 の CNN に通し、調整用ウェイトのフルセットを生成する。これらの重みは、元のネットワークの活動をユニット固有の方法で調整する (これにより、空間と特徴の両方に注意を払うことができる)。その結果、複数の視覚課題に柔軟に対応できる単一のフィードフォワード結合荷重セットが得られる。

同じ入力を異なる課題に応じて異なるように処理する場合、これらのネットワークは本質的に、特徴的な注意に依存したモード内課題スイッチングの一形態を実施していることになる。この点では、先に述べたストループ検査に最も似ているかもしれない。

### 3.4. 記憶への注意

深層ニューラルネットワークは、明示的な記憶を持たない傾向にあり、そのため記憶への注意は研究されていない。しかし、Neural Turing Machine は、外部メモリストアを含むハイブリッドなニューラルアーキテクチャである (Graves et al, 2014)。ネットワークは、訓練を通じて、これらの記憶保持と効果的に相互作用する方法を学習し、記憶された系列のソートや反復などの課題を実行する。この相互作用を容易にすることが、注意の一形態である。記憶は、ベクトルの集合として保存されている。この記憶から情報を取り出すために、ネットワークは各ベクトルに重みを生成し、記憶の加重和を計算する。この重みを決定するために、リカレントニューラルネットワーク (外部から課題に関連した入力を受け取る) は、ベクトルを出力し、記憶はこのベクトルとの類似性に応じて重み付けされる。このようにして、各時点で、ネットワークは文脈に関連した記憶にアクセスすることができる。

前述したように、脳がどのようにして注目すべき記憶を選択し、それに注目するのは完全に解明されていない。このモデルでは、類似性計量尺度を使用しているため、神経科学の文献にある連想記憶モデルと同様に、生成された活性ベクトルとの重なりに基づいて記憶が検索されることになる。これは、後者の問題、つまり記憶への注意が脳内でどのように実行されるのかという問題に対する機構を提供するものである。このモデルでは、生成された活性ベクトルが、どのような記憶に注意を向けるかを制御しており、生物学との関係はあまり明確ではない。

## 4. 人工的な注意と生物学的な注意の間の将来の相互作用のアイデア

これまで述べてきたように、生物学からのヒントを得て、AIニューラルネットワークではすでにいくつかの注意の例が生まれている(図5にまとめた)。このような注意メカニズムを追加することで、これらのシステムの性能はかなり向上したが、まだまだ不十分な点も多く、さらなるインスピレーションを得る機会も存在している。近い将来、このようなインスピレーションは、現在存在する特殊な人工システムを少しずつ改良していく形で得られるだろう。しかし、脳にインスパイアされたAIの真の目的は、より統合された、多くの課題に柔軟に対応できる多目的エージェントを実現することである。

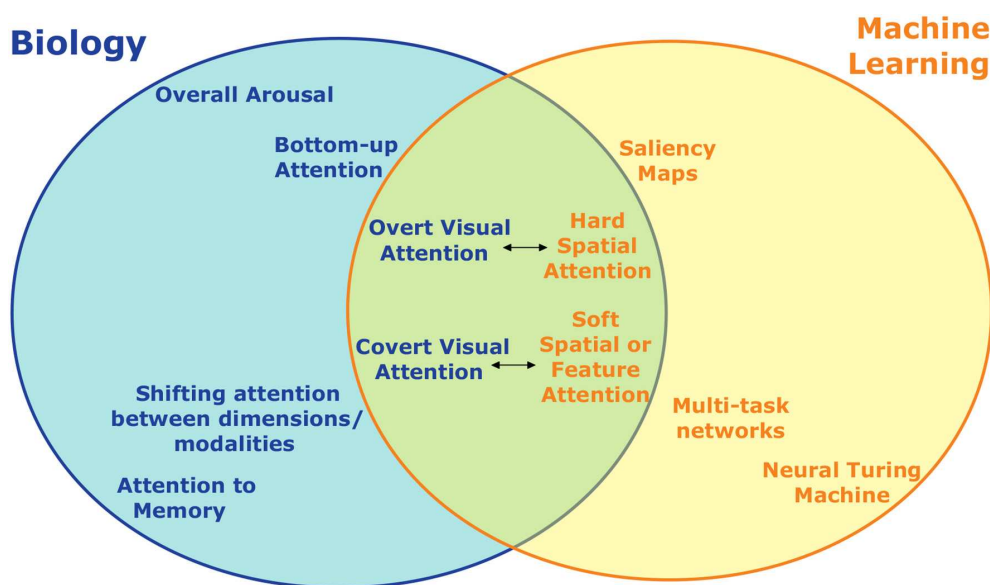


図5 神経科学・心理学と機械学習で研究されているさまざまな注意の種類とその関連性を不完全ながらまとめたもの。左は生物学的に研究されている注意の区分、右は人工知能や機械学習のために開発された注意の区分である。同じ水平方向の位置にあるトピックは、ある程度の類似性があり、その距離が類似性の近さを示す。例えば、視覚的注意の形態は、生物学と機械学習の間で最も重複しており、最も直接的に比較することができる。全体的な覚醒度など、注意の形態によっては、明らかに人工的な類似性がないものもある。

### 4.1. パフォーマンスを向上させるには

脳における注意の働きの研究には、表裏一体ともいえる2つの要素がある。

1つ目は、注意がどのようにして成績を向上させるのか、つまり、注意に関連する神経の変化がどのようにして脳を課題遂行能力に優れたものにするのかという問題である。2つ目は、注意がどのようにして、また、なぜそのように展開されるのか、つまり、ある特定の項目や課題が注意の対象として選択され、他のものは選択されないのはどのような要因によるのか、という問題である。

神経科学者たちは、前者の疑問について多くの時間を費やして研究してきた。しかし、これらの知見をAI神経システムに適用するのは、大部分が簡単ではないかもしれない。活動の乗法的スケールリングは、生物学的にも人工的にも現れるものであり、注意を実装するための有効な手段である。しかし、脳内で観察される注意の効果の多くは、主に、ノイズの多いスパイク状のニューロンが伝える信号を増加させる手段として意味をなしている。これには、ニューロン間の同期の増加や、発火のばらつきが減少などが含まれる。このような変化を示すアナログがディープニューラルネットワークになれば、そこからヒントを得ることは難しい。さらに、ニューラルネットワークの学習手順は、明確に定義された課題成績を向上させるために必要な活動の変化を自動的に決定することができるため、生物学的変化から得られる教訓はあまり意味がないかもしれない。

一方で、注意が活動電位の高さやバースト性、正確なスパイク時間などのスパイク特有の特徴に影響を与えるという観察結果は、スパイクングネットワークの有用性を示していると考えられる。具体的には、スパイクングモデルでは、注意が制御できる自由度が高いため、注意がより大きな、あるいはより微かな影響を与える可能性がある。

注意の解剖学的構造を観察することは、AIシステムのアーキテクチャを設計する人々に有益な洞察を与えるかもしれない。例えば、視覚的注意は、V4などの後期視覚領域でより強く活動を調節するように見える(Noudoost et al, 2010)。だが、聴覚的な注意は、処理経路のかなり早い段階で活動を調節することができる。このように、注意がどのレベルで作用するかは、アーキテクチャの変数として関連している可能性がある。このような観点から、最近の研究では、Transformerモデルの初期層から自己注意を取り除くことで、特定の自然言語処理課題での成績が向上し、また、言語処理中の人間のfMRI信号の予測にも適したモデルになることが示されている(Toneva and Wehbe, 2019)。

ある感覚モダリティで注意を引くと、別のモダリティでも同じ物体や場所に注意が展開されるというクロスモーダルなキューの存在は、異なる感覚システム間の直接的な相互作用をある程度示していると考えられる。機械学習におけるマルチモーダルモデルの多くは、完全に別々の処理経路を使用し、最終的にのみ結合されるが、異なる入力経路間の水平方向の接続を可能にすることで、それらの処理を調整することができる。



また、注意は、通常感覚処理で行われるような適応と相互作用する。一般に、ニューラルネットワークモデルには適応のメカニズムがない。つまり、同じ入力を複数の時間ステップで与えられた場合、ニューロンは活動を低下させる手段を持たない。適応は、変化や異常を目立たせるのに有効であることを考えると、適応させることは有用であると考えられる。適応のあるモデルでは、注意機構は、繰り返される刺激が重要であると判断された場合、適応したニューロンを再活性化するように働くはずである。

最後に、注意の形態の中には、同じシステムに対して複数の働きをするものがあるようである。例えば、視覚的注意は次のような働きをすると考えられている: (1) 大脳皮質の視覚ニューロンの活動を調節することで、その感度を高める。(2) 大脳皮質下の活動を変化させて、感覚情報が異なる形で読み出されるようにする(Birman and Gardner, 2019; Sreenivasan and Sridharan, 2019)。このようにして、注意は、脳の異なる部分にある2つの異なるメカニズムを使って、その効果を生み出す。注意がモデルアーキテクチャの複数の構成要素を補完的に調節できるようにすることで、より強固で効果的な影響を与えることができるかもしれない。

## 4.2. 注意をどのように配置するか

複雑で統合された人工知能を作り出すためには、注意をどのように展開するかという問題の方が重要であると考えられる。流れてくる刺激の中から関連する情報を選択したり、取り組むべき最適な課題を選んだり、あるいは何かに取り組むかどうかを決定するには、エージェントが自分の状態、環境、ニーズを統合的に理解している必要がある。

生物学的注意から影響を受ける最も直接的な方法は、それを直接模倣することである。例えば、スキャンパスモデルは、何年も前から顕著性の研究に存在している。これらは、人間が画像を見ているときに行う一連の固視を予測しようとするものである(Borji and Itti, 2019)。注意を訓練するためのより直接的なアプローチは、Linsley ら (2018) で用いられた。ここでは、被験者に物体分類に最も関連する画像の領域をラベル付けさせることで、人間のトップダウン注意の大規模なデータセットを収集した。この方法で作成された課題固有の顕著性地図は、物体認識を主課題とする深層畳み込みニューラルネットワークの注意を訓練するために使用された。その結果、この方法で中間層の活動に影響を与えると、成績が向上することがわかった。また、Zagoruyko and Komodakis (2016) では、教師の顕著性地図を学習する別の方法が示されている。

課題と人間の視覚領域から収集した神経データを組み合わせた訓練も、CNN の性能向上に役立っている(Fong et al, 2018)。特に注意課題で収集した神経データを使えば、注意モデルの訓練に役立つ可能性がある。このような転送は、他の課題も行うことができる。例えば、読書中の眼球運動を追跡することで、NLP モデルに情報を与えることができる。これまでのところ、眼球運動は品詞タグ付けモデルの訓練に使われている(Barrett et al, 2016)。興味深いことに、乳児は周りの大人が何に注意を向けているかに気を配ることから学ぶことができ、より広くエージェント間で注意を調整することは、社会的な種において非常に有用であると考えられる。したがって、他者の注意は、注意がどのように導かれるかに影響を与えるはずである。共同注意を調整する試みは、注意システムに統合する必要があるだろう(Kaplan and Hafner, 2006; Klein et al, 2009)。

興味深いことに、乳児は周りの大人が何に注意を向けているかに注意を向けることから学ぶことができ、より広範にエージェント間で注意を調整することは、社会的種族において非常に有用であると考えられる。したがって、他者の注意は、注意の導き方に影響を与えるはずである。共同注意を調整する試みは、注意システムに組み込まれる必要がある(Kaplan and Hafner, 2006; Klein et al, 2009)。活動は、いくつかの可能な目標のうち、いつでもどの目標を達成すべきか、したがってどこに注意を置くべきかを柔軟に決定する必要があるだろう。この問題は、強化学習、特に副次課題の選択を伴う階層型強化学習の問題と密接に関連している。このような決定は、予想される正または負の結果に基づいて行う必要がある。実際、注意と報酬の間には密接な関係があり、以前に報酬を得た刺激は、もはや報酬を得られない状況でも注意を引きつける(Camara et al, 2013)。人間がどのかのだいをいつ実行するかをどのように選択するかについて理解を深めれば、人間の行動をマルチタスク AI の設計に反映させることができる。

そのためには、異なる過程を制御する脳の限られた能力の配分は、その制御の期待値に基づいて行われるという、Shenhav ら (2013) が提唱した理論が役に立つかもしれない。この枠組みでは、背側前帯状皮質は、制御の認知的コストを含む多様な情報を統合して、制御の期待値を計算し、その結果、注意などの過程を指示する役割を担っている。複雑な課題における人間の実行制御を理解するための別のアプローチとして、逆強化学習がある。この方法は最近、人間が使用する報酬関数とポリシーを決定するために、視覚探索中の眼球運動のデータセットに適用された(Zelinsky et al, 2020)。

生物学的な注意を促進するが、人工的な注意システムではおそらく十分に表現されていない追加的な要因は、好奇心である(Gottlieb et al, 2013)。生物学的には、新規の刺激、混乱した刺激、驚くべき刺激が注意を引くことがあり、下側頭葉と末梢皮質は、馴染みのある入力に対する反応を減少させる適応メカニズムを介して、新規の視覚的状況を通知すると考えられている。状態の値の推定の一部に新規性を含める強化学習アルゴリズムは、このような探索を促すことができます(Jaegle et al, 2019)。しかし、異なる状況で驚きや新しさを具体的にどのように計算するかは、必ずしも明確ではない。生物学的注意に関するこれまでの研究では、ベイズ的に驚きや情報収集の観点から注意選択を理解しており、こうしたフレーミングは AI システムにも有用であると考えられる(Itti and Baldi, 2006; Mirza et al, 2019)。

注意の選択における最後の問題は、葛藤をどのように解決するかである。覚醒、ボトムアップ、トップダウンなど、脳には複数の注意の形態があるが、適切な注意の所在に関する葛藤はどのように解決されるのだろうか。視覚系を見てみると、これらの複数のシステムが対象とする局所回路がこの課題を担っているように見える。これらの回路は、神経調節のための入力とトップダウンの信号を受け取り、それらを活動を促すボトムアップの入力と統合しなければならない。水平方向の接続がこの競争を仲介し、勝者と勝者の関係を築くメカニズムを利用することができる。これは、AI システムのアーキテクチャでも真似ることができる。

## 4.3. 注意と学習

注意は、何が記憶に入るかを決定する役割を果たし、学習を導く。注意を備えた AI システムの多くは、訓練中に注意機構を含んでいる。このようにして、注意機構は基本アーキテクチャと一緒に学習される。しかし、Neural Turing Machine を除いては、注意機構が機能した後、モデルは学習を続けない。したがって、これらのシステムでは、学習や記憶を制御する注意の能力はまだ明示的に考慮されていない。

注意は、入力中の関連する成分や関係性に学習を向けることで、データを効率的に利用するのに役立つだろう。例えば、顕著性地図は、様々なコンピュータビジョン課題の前処理の一部として使用されている(Lee et al, 2004; Wolf et al, 2007; Bai and Wang, 2014)。本質的に顕著な領域のみに後続の処理を集中させることで、無関係な領域への無駄な処理を防ぐことができ、また、ネットワーク学習の観点からは、これらの領域へのオーバーフィッティング

を防ぐことができる。しかし、このように顕著性マップを利用するには、問題に応じた顕著性の定義が必要である。人間のボトムアップの注意を引く画像の特徴を利用することは、コンピュータビジョンの問題ではうまくいっているが、他のモダリティの人間のデータを見ることも有用である。

これに関連して、乳幼児の研究では、顔などの関連する刺激に注意を向けるための事前知識を持っていることが示唆されている。このような事前知識を用いることで、重要な刺激をどのように処理するか、また、関連する特徴にどのように注意を向ければよいか、という両方の学習を起動させることができる (Johnson, 2001)。

トップダウンの注意は、データのどの部分を処理するかを決めることに加えて、処理中にネットワークのどの要素を最も動かせるべきかを選択することと考えることができる。ネットワークのどの部分が最も強く働いているかで学習が行われるので、これも注意が学習を導く手段の一つである。任意の入力に応じて更新されるパラメータの数を制限することは、ドロップアウトやバッチ正規化の使用に見られるように、正則化の効果的な形態である。注意は、どのユニットを使うかをランダムに選択するのではなく、この課題の成績にも役立つユニットを選択するように制約されている。そのため、より課題に特化した形での正則化となる。

このように、注意は、ネットワークが既に学習した他の課題成績を乱さないようにしながら、特定の課題でより良い成績を発揮するようにネットワークをアップデートすることを目的とする継続的学習に特に役立つ可能性がある。関連する概念である条件付き計算は、最近、継続的学習の問題に適用されている (Lin et al, 2019)。条件付き計算では、ネットワークのパラメータは現在の入力の関数であり (そのため、注意によって行われるタイプの変調の極端な形と考えることができる)、ネットワークを効率的な継続学習のために最適化するには、ネットワークの効率的な継続学習のためにネットワークを最適化するには、異なる入力間の干渉の量を制御する必要がある。より一般的には、注意は、望ましくないシナプスの変化を防ぐための手段であると考えられるのがよいだろう。

注意と学習もループしている。具体的には、注意は世界について何が学習されるかを導き、内部世界モデルは注意を導くために使用される。この相互依存関係は、近年、人間の学習や意思決定の説明に成功している認知ベイジアン推論モデルも取り入れた強化学習フレームワークの観点から形式化されている (Radulescu et al.2019)。大脳基底核と前頭前野の相互接続は、強化学習と注意選択の相互作用を支えていると考えられている。

より抽象的なレベルでは、脳のアーキテクチャに注意が存在するだけで、表現学習に影響を与えることができる。意識の大域的ワークスペース理論では、どの瞬間にも、脳の活動から選択された限られた量の情報がワーキングメモリに入り、さらなる共同処理に利用できるとしている (Baars, 2005)。これに触発されて、機械学習における「意識先行」は、基礎となる高次元の状態表現に適用される注意から生じる低次元表現を持つ神経ネットワークアーキテクチャを強調している (Bengio, 2017)。この低次元表現は、将来の状態を要約したり予測したりするのに使えるような、抽象的なレベルで世界を効率的に表現する必要がある。この注意を媒介としたボトルネックの存在は、行動を導き、予測を行うために柔軟に組み合わせることができるように、すべてのレベルで分離された表現を促すというトリクルダウン効果をもたらす。

楽器の演奏など、多くの複雑なスキルの学習には、意識的な注意が必要である。しかし、いったん学習が完了すると、これらの処理は自動的に行われるようになり、注意力が解放されて他のことに集中できるようになる可能性がある (Treisman et al, 1992)。この変換のメカニズムは完全には解明されていないが、タスクの負担を別の、おそらくより低いより反射的な脳領域に移動させることに依存しているようである限り、注意によって異なる働きをすることができる複数の冗長な経路を持つことは、人工システムにとって有益であると考えられる (Poldrack et al, 2005)。

## 4.4. 注意の限界 バグか機能か？

生物学的注意は完全には機能しない。前述したように、異なる種類の注意を切り替えるときにパフォーマンスが低下したり、ピークパフォーマンスに達するためには覚醒レベルがちょうど良くなければならなかったり、トップダウンの注意が無関係だが顕著な刺激によって中断されたりする。AI システムに注意を移す際の問題は、これらの制限を避けるべきバグなのか、それとも取り入れるべき機能なのかということです。

気が散りやすいというのは、一般的には、注意力のバグというよりも、むしろ特徴のように思われる。課題に集中しようとしても、生命を脅かす可能性のある環境の変化を意識し、気が散ってしまうことは有益なことである。問題となるのは、脅威でもなく、関連性のある情報でもない入力に対して、エージェントが過度に注意をそらす場合です。したがって、人工的なシステムでは、トップダウンの注意力の強さのバランスをとり、予期しないが情報を提供する刺激の処理を可能にする必要がある。例えば、注意の瞬きとは、ターゲットとディストラクタの流れの中で、第1のターゲットの後に第2のターゲットがすぐに現れた場合に、被験者がそのターゲットを見落としてしまう現象のことである (Shapiro et al, 1997)。これは成績を低下させるが、脳が最初のターゲットを処理して行動する時間を与えるためには必要なことかもしれない。このようにして、フォロースルーを確実にするために気が散ることを防ぐ。

人工的なものであれ、生物学的なものであれ、どのようなエージェントであっても、そのエネルギー資源にはある程度の限界がある。そのため、いつ社会に出るのか、それとも睡眠などの省エネ状態に入るのかを慎重に判断することが常に重要になる。多くの動物では、睡眠はスケジュールに沿って行われるが、前述したように、注意を要する状況では睡眠が遅れたり中断されたりすることもある。そのため、いつ睡眠状態に入るかは、起きていることで何が得られるかという費用対効果の分析に基づいて決定しなければならない。睡眠には記憶の定着など、省エネ以外にも重要な動きがあることが知られており、その判断は複雑なものになるかもしれない。AI システムがこの判断を下すためには、現在の状態と将来の要求を統合的に理解する必要がある。

## 5. 結論

注意は、心理学、神経科学、人工知能にまたがる大きくて複雑なテーマである。この名前で研究されているテーマの多くは、その機構に重複はないが、限られた資源を柔軟にコントロールするという中核的なテーマを共有している。柔軟性や資源の賢明な利用に関する一般的な知見は、人工知能の開発に役立つし、特定の感覚モダリティや課題に注意を向けるための最適な方法に関する具体的な知見も得られる。