価値に駆動された
人の推論システムに関する研究

玉川大学大学院工学研究科　システム科学専攻

宮田　真宏

目次

第1章　　序論

人は思考する動物である．それが人を他の動物から大きく異ならせている．その思考の特徴は，言語や論理などに代表される離散的概念表現とその操作であろう．例えば名詞は一つの概念を表すことができ，我々は名詞を組み合わせて新しい概念を作ることができる．また，我々は単語をある規則に従って組み合わせ，意味を表現する文章を作ることができる．このような能力の背景には，さまざまな概念を単語で表現し，さらにその離散的な表現を操作する，シンボル処理の能力があると考えられている．そして論理は知的な思考の典型例であり，シンボル間の演算による意味の表現である．従来，人のこのような思考の能力は記号処理としてモデル化され，論理学や記号処理学により検証されてきた．多くの哲学者もまた「人の知能」とはこのような知であると長い間考えてきた．

しかし近代になって，脳についての科学が大きく進歩し，思考の実現の媒体としての脳の姿がだんだん明らかになってきた．脳は大量の，おそらくは性質の異なる神経細胞の集合体であり，その細胞数は 1,000 億個と言われている．それだけの神経細胞は脳内で大規模なニューラルネットを作り，個々の神経細胞の発火は二値の動作であるが数百個以上の集団になるとアナログ的に見える情報処理をしており，脳波のようなマクロな電気現象や，fMRI(functional magnetic resonance imaging)で観察される脳の多数の領野の活性化のような現象を示している．しかし，これだけの脳についての知識が集約されたとしても，脳に全体としての離散的思考が生まれるメカニズムは，謎のまま残っている．論理的な思考は意識とも大きく関わりのある現象であり，この謎に対するアプローチは，脳全体のシステム的な動作から入っていく必要があるであろう．その謎に対する本論文のアプローチは「価値と記憶を通しての論理的推論の実現」である．

本学位論文に関係する研究の当初の目的は人の感情に着目し，その脳内メカニズムをシミュレーションにより検証することで人の感情を理解することであった．その検証の為に人の感情の発生に関わる調査を行った．その結果として人の感情は従来，扁桃体が関わっていると考えられていたが，近年の研究成果より脳の様々な部位，および機能が関わることが示唆されていた．この調査の結果から本学位論文では，人の感情とはその場の状況として各感覚器官から入力された情報を脳部位ごとに処理し，そこで見いださ

れた特徴に対して価値を含めた計算をし, その結果として表出される現象であると考えた. そのため本学位論文の先行研究では脳部位ごとに見出される価値に着目した感情モデルを提案し, シミュレーションにより価値に駆動された感情の実現の可能性を検討した．しかし実装を考えた際，従来多くの推論手法では人の多岐多様な推論の説明ができないことが本研究の検証において問題となった. このことから本学位論文では人の感情に関わる価値計算のうち, 人の論理的な推論に着目しそのモデル化と機能の検証を試みた.

人の推論に関しては従来, 認知科学では直観的推論と論理的推論の 2 種類があるとされ，別々にモデル化されてきた．先行研究では，直観的推論はベイズ推論に代表される確率的な手法を用いることでモデル化されてきた．そして論理的推論は Tree 探索に代表されるシンボル的な手法によりモデル化されてきた [1]．一方，推論と脳部位とを対応付けた研究はあるが [2] [3], 脳の神経回路を考慮した推論の包括的なメカニズムについて言及したものは見つかっていない.

本研究ではまず, 人の推論過程は直観的推論と論理的推論に明確に分かれているのではなく, 一つの分散型ニューラルネットワークの動作モードの切り替えで実現されると考え，そのモデルを提案し，推論システムとして体系化した．その後，提案した推論システムの特性として, 統合パラメータを変更することによる推論システムの挙動の変化について検証した．最後に本推論システムの応用研究の一つとして，推論システムと強化学習とを組み合わせたシミュレーションを実施し，その効果について検証した．本論文は，上記のような経緯を踏まえて，以下のような章構成とした.

2 章では，本学位論文に関係する研究をするにあたり，当初の目的であった人の感情の発生メカニズムのモデル化について説明する．従来，感情のモデル化は様々な分野で行われてきたが，そのほとんどが感情を現象として扱っており，その発生メカニズムについては議論されてこなかった. 本学位論文では感情とは脳内における価値計算の結果をうけて表出される現象であると考え，モデル化を行った．そして機能ごとに価値計算の実装を考えた際，人の推論において，従来の Tree 探索に代表される手法では説明が困難であると考えた. このことから本学位論文では人の感情シミュレーションの前に人の推論についてモデル化し，その機能をシミュレーションする必要があると考え，その検証をした.

3 章ではまず，人の推論に関する先行研究を調査した．認知科学では従来，人の推論には無意識的で処理時間が短く，かつ確率的であるとされる直観的推論と，意識的で処理時間が長いとされる論理的推論の二種類があるとし，これらを別々のシステムとして切り分けてモデル化してきた．それに対し，従来の人工知能技術を用いた推論の基本モデルとして Tree 探索が挙げられる．Tree 探索とは個々の離散的な状態予測，およびその評価をする意識的かつシンボル的な，論理的推論の説明に用いられる方式である．しかし Tree 探索では，従来言われてきた人の直観的推論の説明ができない．また，我々は論理的推論とは別に，何かを知覚するとその影響に対する予測と評価を素早く行う直観的な推論過程も持っている．これは，感覚刺激からの自動的な連想による無意識的な予測と評価によると考える．本学位論文の先行研究でははじめ，粒子モデルを用いてこの直観的推論を実現してきた．しかし論理的推論については現状では実現できていない．その上で本章では，これら 2 つの推論機能を統一手法で説明可能とする推論の統合アーキテクチャを提案する．

4 章では，提案した推論の統合アーキテクチャの実現のために，この人の 2 つの推論過程を価値に駆動された連想記憶を用いた分散型のニューラルネットワークモデルにより説明することを試みる．連想記憶とは，記憶パターンを貯蔵し，部分的な記憶情報を基に必要な記憶を読み出す機能である．先行研究の神経回路による分散型連想記憶のモデルでは，複数個の記憶事項の記銘はそれらの相関行列の和（記憶行列）で表し，想起用の入力ベクトルと記憶行列の積を計算することで想起を再現するものであった．本研究では従来言われてきた人の 2 種類の推論に対し，別の処理システムとしてそれぞれをモデル化するのではなく，2 つの推論は 1 つの処理システムで実装され，その動作モードのスイッチングにより切り替えるという方式を採用している．そのため，直観的推論は連想記憶モデルの相互想起モデルを用いることにより，論理的推論は自己想起モデルを用いることによりそれぞれ実装した．

そして計算機シミュレーションにより，直観的推論では従来研究より言われている確率的な推論が実現できていることを確認した．さらに論理的推論では，従来研究より言われている，深い推論が実現できること，さらにその結果の解釈として従来の推論手法である Tree 探索の深さ優先探索のような結果が得られることを確認した．そしてその後，切り替えパラメータを変更することによる推論システムの特性の変化を計算機シミュレーションによって評価した．

5 章では，本研究にて実現した推論システムのシミュレーション環境における適用の可能性について検証する．我々は人の経験するすべての状態を円により表現した際，その円の一部の領域は，進化の過程おいて先天的に埋め込まれていると広く考え考えられている反射による領域や，過去の経験を基に強化学習などにより価値の割り振られている領域があると考える．このような場面において，推論とは未経験な領域から過去に経験した既に価値の割り振られている領域に対する状態空間の探索であると考える．このような条件を満たす環境下におけるシミュレーションによる効果を検証しようと考えると，他の学習器（強化学習など）と推論システムとの連携が必要になる．そこで本研究では，他の学習器との連携の可能性を確認するため，強化学習手法の一つである Q 学習との連携の可能性について検証を行った．さらに，我々が生活する上で見出す価値は必ずしも一つとは限らず，複数の価値を見出すことも大いにある．本章ではこのような複数の価値を同時に見出した際にも意思決定ができるかどうかのシミュレーションも行った．

6 章ではまとめとして，本研究で行ったシミュレーションの妥当性と一般性について議論する．そのうえで，脳のように複数の機能を組み合わせることで実現する行動決定モデルとしての本学位論文にて示したシミュレーションの位置づけについて議論する．その後は，2 章にて議論した人の感情を理解するという目的に対して本研究結果が示唆することについて議論する．そして最後に，いまだ未解決で謎の多い人の知能の理解のために本研究が示唆することについて議論する．

第2章　　背景

数年前から人工知能(AI)は第三次人工知能ブームと呼ばれ，世界中の注目を集めてきた．今後は現在の Amazon Echo などに代表される AI 技術を用いた製品がさらに多く社会に生み出されるだろう．その際に生み出される製品には，より人的な機能が求められると予想できる．その典型例は人との相互作用，つまり対人インタラクションであると考える．対人インタラクションとは，人を対象とした広義のコミュニケーションを指し，その実現にはコミュニケーションをする相手の意図や要望，ニーズを理解し，それに呼応した自身の認識の変更や意思の決定が必要と考えられる．

しかし，実際に人の行っているコミュニケーションの過程は先述した内容よりも更に多様かつ複雑である．それには，言語やサインを用いた明確な意図の伝達がある一方で，身振り手振りや行動による暗黙かつ曖昧な伝達まで，多様な相互作用が同時並列的に含まれている．このことから対人インタラクションは，これらの情報を同時進行的に読み取り，それに基づいて行動を柔軟に変更し，相手に働きかけることで相手の認識や行動を変えることを促し，それにより自己の目的達成をする課題であると考えることができる．この課題は人にとっては無意識的にでも行うことが可能であるため，比較的易しい課題であると考えられることが多い．しかし，実際にその内部処理を考えると決して単純なものではない．

このように複雑な対人インタラクションを理解するための鍵となる要素として動作や表情，会話中の間などが挙げられるが，本研究ではその中でも『感情』に注目した．感情は人だけでなく動物がコミュニケーション行動をする際に重要な機能を持つ現象である．しかし，これまでの感情についての研究の多くは，現象面からの感情の分類と解析であり，認知的なプロセス中に含まれる感情の発生やその計算論的役割について検討されているものは少ない．そこで本研究では，感情とは動物が行動する際の意思決定に用いられる価値計算システムの現れであると仮説を立て，その計算モデル化の可能性を検討する．

2.1 情動とインタラクションの関係

情動[1]に関しては心理学や生理学の分野でこれまで多くのモデルがあり [4] [5] [6]，コミュニケーション場面における感情の役割についても従来から多くの研究がなされてきた [7] [8]．対人インタラクションの場面では，阿部 [9]は保育士が操作するロボットと子ども間のインタラクションを対象に行動調査を行って分析し，インタラクション場面における感情の誘導がコミュニケーションの成功に重要であることを示している．また山田 [10]は実際の保育現場での観察を通して，保育者には子どもとの関わりあいの中から子どもの心を読み取る専門的な能力があるとした．そして，保育者が子どもの心の状態のパターンを記述し，その関係性を図として可視化するオントロジを作成することで子どもの心的状態を推定できると考え，心的状態推定モデルを提案した．

これらの研究はいずれも，相手の心的状態を推定することが人のコミュニケーションを理解する上で重要であることを示している．相手の心的状態は従来の言語や身振りによるコミュニケーションでは考慮されていない．他者の心的状態は明示的には計測できない変数であり，相手とのコミュニケーションのためには推定することが必要となる．そのような心的状態の推定は，技術的な困難が多いと考えられる．しかし人のコミュニケーションが相手の心的状態に依存して大きく変わることは自明な事実であり，人工知能やロボットのような擬人化エージェントによる対人インタラクションを考えるなら避けて通ることはできない．

ここでいう心的状態とは，注意の向き，認識の対象，理解の状態，心的状態の良し悪し，働きかけによる変化の予測など，極めて動的なものである．そして，人は他者の感情を知ることでインタラクションをより円滑に遂行できる．人のインタラクションを理解する上で感情のモデル化は重要である．

---

[1] 本研究において感情と情動は，その処理する内容の複雑さにより区別し，それぞれ，進化的に古く，かつ動物においても共通にあると考えられるもの(怒り，悲しみ，喜びなど)を情動とし，進化的に新しく，かつ人に特有であると考えられるもの(社会的価値，経済的価値など)を感情とする．しかし，本研究では情動と感情とでは機能的な観点では本質的な違いはないものとして扱う．そのため，感情と情動は本文中において明白な区別はせずに使用する．

2.2 感情に関する先行研究

2.2.1 認知科学の知見に基づく主なモデル

Ekman は，顔の表情は人の内面についての様々な現象の情報を含んでおり，顔の表情を分析することで，人格や精神病理学，そして人の初期の発達時の問題を分析することができると考えた．その中でも重要視したものが人の感情についてである．Ekman は顔の表情筋の緊張度を緻密に記述することで表情を 6 種類に分類する FACS モデル [5]を提案した．FACS モデルでは，感情の 6 種類の分類の内容として，恐れ，怒り，悲しみ，幸福，嫌悪，驚き，を挙げている．

それに対して Russel は第一軸に快－不快を，第二軸に眠気－覚醒で構成される軸を取り，感情はその軸上で表現できる空間に円環状に配置することが出来るとし，個別の感情状態間の関係を示した [4]．しかし，これらは感情を現象として捉えた記述モデルであり，脳や認知のメカニズム，さらにその計算論的意味について迫るものではない [11] [12]．また計算論的なモデルも提案されているが，まだ単純な段階にある [13].

それに対して戸田は，感情は適応的な行動選択システムであるとし，感情メカニズムのモデルとして人の比較的高度な感情を説明するアージ理論を提案した [14]．アージ理論では，人の複雑かつ多様な感情は基本的な情動と知的能力による推論により導きだされるものとして，多様な高次な感情の説明を試みている．その理論では，感情は自らがおかれている場の状況に価値を割り振って意思決定に至るまでの過程であると説明をしている．しかし戸田の理論は概念モデルにとどまっており，感情の具体的な処理過程については述べていない．

2.2.2 脳科学の知見に基づく主なモデル

脳科学においては，情動に関係する部位として扁桃体が挙げられてきた．ルドゥーは扁桃体が，人の知覚した情報が生体にとって報酬的であるか，罰的であるかに対して価値判断する機能があることに着目し，情動と扁桃体との関係を示した [6]．さらに，扁桃体と情動機能との関わりに着目した例として，Klüver らのクリューバー・ビューシー症候群が挙げられる．彼らは，扁桃体を切除したサルが恐怖をはじめとした情動反応をなくし、サルにとっては本来脅威刺激であるヘビやクモに対して恐怖反応を示さない行動の異常を観察した．このことから情動と脳との関係が示唆された [15].

しかし，情動に関わる脳部位は扁桃体だけではない．例えば，実際にヘビにかまれたことのない人でも，ヘビを見ると恐怖を感じるなど，我々の遺伝子に組み込まれていると考えられる要素がある．これに対して Koelsch らは人の感情のカルテットセオリーに関するモデルを提案した [16]．このモデルでは，感情を生み出す要素をより精密に分類して，身体の維持，安全の実現，愛着，経済的価値の 4 種類として，そのそれぞれに脳に対応する部位（脳幹(Brainstem)，間脳(Diencephalon)，海馬(Hippocampal formation)，前頭眼窩野(Orbitofrontal cortex)）があるとしている(図 2-1).

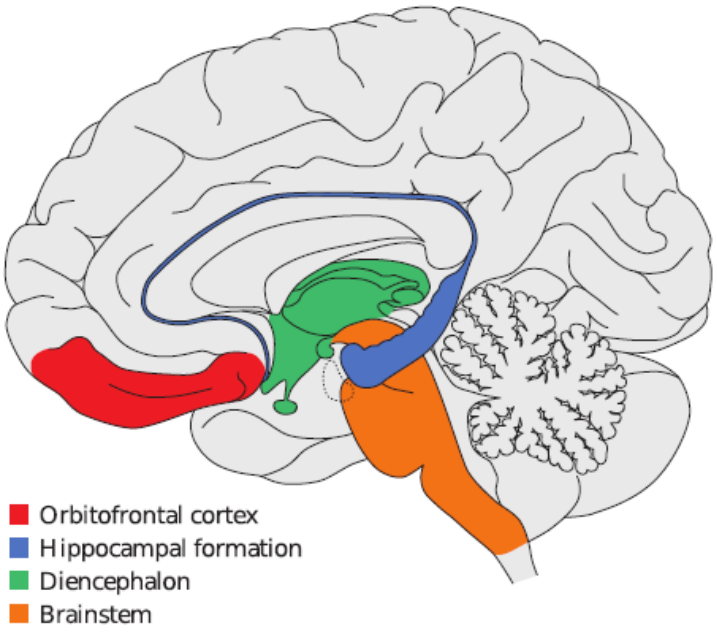

図 2-1　人の 4 つの感情システムと脳部位との関係

出典：S. Koelsch, The quartet theory of human emotions: An integrative and neurofunctional model (2015)，p.3

この 4 つの感情を生み出す要素の内，身体の維持とは，例えば食後の満足や空腹，腹痛などの基本的な身体状況のことを指す．通常，我々はこれを感情とは呼ばないが，空腹になれば不機嫌になるなどの経験などから考えても，これらの要素は感情の一部といっても間違いではないだろう．安全の要素には恐怖や闘争心などが挙げられ，これらは我々が普段使う感情に最も近いものである．愛着とは恐らくは記憶と感情が結び付いたもので，道具・環境・夫婦・共同体など多くの事例があり，さらに気に入ったものは多少のコストを払ってでも獲得・維持したいと考える現象である．経済的価値を感情に入れることについては，経済的価値は感情なのかという観点から考えると議論があろうが，例えばお金を得ると嬉しい，損をしたらがっかりするなど，経済的価値と感情の結び付きは大きい．一方で，著者の脳科学の知識からはカルテットセオリーが指摘した脳部位は狭いのではないかと考える．例えばルドゥーの例のように扁桃体では価値を正負に符号化するとされ，大脳基底核では強化学習により行動の価値を学習するとされる．さらに前頭葉内側部などは社会的行動に関わるとされる，などより多くの部位が感情には関係してくると考えられる．

2.3 動物に共通な感情

ヒトの脳が現在の形に至る過程についての仮説はいくつかある. MacLean はヒトの脳構造とその行動について, 生物の進化の過程と動物の原始的な本能から説明することを試み，3 種の脳の階層構造による「三位一体脳」仮説を提唱した [17]．ここでは，ヒト脳は爬虫類脳・旧哺乳類脳・新哺乳類脳の順番で進化し，進化のたびにその機能を複雑化かつ高度化してきたというものである. 三位一体脳のそれぞれに関連する脳部位とその機能は以下の通りである.

(1) 爬虫類脳 : 進化の過程で最も古く発生した脳部位であり，自律神経系の中枢である，脳幹と大脳辺縁系より成り立つとされている.
(2) 旧哺乳類脳 : 爬虫類脳の次に進化した脳部位であり，海馬や扁桃体などの大脳辺縁系から成り立つとされる．大脳辺縁系の出現により，個体の生命維持機能だけでなく，本能的な情動（快－不快など）や愛着などの機能が実現された.
(3) 新哺乳類脳 : 旧哺乳類脳に大脳新皮質の両半球が付加された. これにより，言語機能などの高次の情報処理が可能になった.

この中でも爬虫類脳は，脳を持つ動物(少なくとも爬虫類から哺乳類まで)に共通している脳部位として知られている．爬虫類脳があることにより，脳を持つ動物は自らの身体を守るために恐怖などの情動を持つことができるとされる．ルドゥー [6]は情動の一部である恐怖情動に注目し, 恐怖の認識や学習によって恐怖に対応する危険の認識が導出される過程について生理学的な視点から解明している. 動物はこの導出された恐怖の情動を受け，次の行動を決定している.

また，恐怖に限らず多くの喜び，怒りなどの基本的な情動はヒトを含む多くの動物の間で共通しており，進化の過程において変化が少なかったことが推測される．そう考えると，情動には進化の過程で維持されるべき存在理由があったはずである．それを明らかにするには，進化という現象を理解することが必要となると考える．しかしそれ以上に，現在の動物の持つ情動がどういう役割を果たしているか，抽象的なレベルで明らかにすることで人という生物の理解が進むであろう．なお，普段我々が一般に感情と呼ぶものは，情動に推論や学習などの知的情報処理が加わって実現されていると考える.

2.4 感情＝価値計算システム仮説

感情は意思決定に大きく影響する．また我々の感情表出は，他者の意思決定にも影響する．一方で現在，認知科学や脳科学の多くの研究では人の意思決定は報酬（価値）の計算に基づくとされている [18]. これより，本研究で我々は感情とは価値計算の結果の表出およびその副作用であり，感情の背後には脳の価値計算システムがあることを想定する．感情表出は注目対象に対する自己の評価を他者に伝えることで，無用な争いを避ける機能を持っている．私たち人は生まれた場所，環境などが異なっていても大きな意味では同様な食べ物を好んだりすることからも価値と感情表現の関係は多少の文化的な違いはあっても比較的固定であり，それゆえに価値伝達システムとして的確に機能できると考える．感情を理解するには，むしろ価値システムの理解が必要であろう．

従来の AI では知能の要素として，視覚や聴覚などの知覚・認識の感覚情報処理，それらを用いて予測や意思決定などを行う高次情報処理，さらにその結果を行動として出力する運動情報処理，すなわち心と言われるもののうちの知的情報処理の部分を考えてきた．それに対して本研究では，それらの背後にある報酬系，すなわち報酬に対応する価値計算の在り方について考える．我々の仮説である価値計算システムでは，新皮質の情報処理システムに感覚情報が入力され，その認識の結果は直ちに脳の異なる価値を計算する情報処理系に伝達されると想定する．我々の考える感情の価値とは Koelsch の提案した「生存」「安全」「愛着」「経済的価値」の 4 種を想定するが，それに対応する脳部位（脳幹・辺縁系・海馬系・前頭葉系）は彼らの考える部位よりも広く考える．これらの部位を統合すると，大脳皮質下の脳部位のかなりの部分を含んでいる(図 2-2).

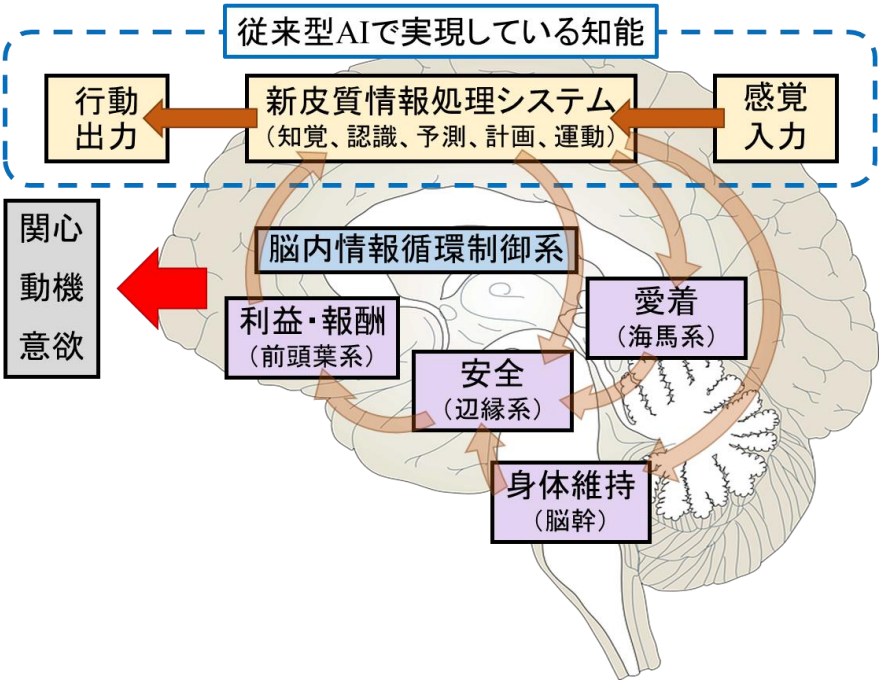

図 2-2　感情＝価値計算システム仮説の概要

本研究では，これらの脳処理系ごとに計算された価値が，前頭葉系で行われるとされてきた探索推論に重要な役割を果たす可能性を検討する．本研究で想定する推論は，新皮質において現状認識を表す神経興奮から次の時刻に起きうる事象またはその特徴群を予測し，それを即時に再認識して価値評価することで予測の分岐の枝刈りをし，より価値の高い予測状態を作りだすという基本サイクルを考える．そして，次のサイクルではその予測状態を起点としてより高い価値状態の探索を継続する山登り法的な過程を想定する．そのサイクルの媒体は，現時点では脳内の広域の脳波を想定しており，価値の高い予測を作り出した脳内領野を活性化(他を抑制)するメカニズムとの組み合わせを想定すると，高速かつ探索的な脳内の情報循環系路の制御が実現できよう．

2.5 感情系の Neural Network 的実装

前節の価値計算システムを実現する際，見出されるべきと考える価値と，それぞれの価値に対応すると考えられる脳部位と，従来の研究で実現されてきたそれに対応する Neural Network モデルとの関係を表 2-1 にまとめた．

表 2-1　価値と脳部位および従来の Neural Network モデルとの関係

見出される価値	脳部位	Neural Network モデル
身体維持	脳幹	センサによる直接検出およびその組み合わせによるパターン認識と固定価値
安全	間脳	感覚センサからの固定パターンおよび学習されたパターンの認識と価値判断
愛着	海馬	場面認識と価値の連合（エピソード）と，その一般化（強化学習）
強化学習	大脳基底核	報酬予測に基づく強化学習
利益・報酬	前頭葉	知覚された現在状況からの推論，Tree 探索や関数近似での状況・価値マッピング

身体維持と安全については，遺伝子で規定される身体部位に作用する部分であり，進化の過程で遺伝により埋め込まれている部分が固定であると考える．さらに，学習が含まれる部分も比較的容易なパターン認識で実現可能と著者は考える．ここでの一番の課題は，複雑な時系列を含む外界の事象や状況の認識であろう．

愛着とは，過去に経験した価値とそれに関わるエピソードの蓄積としての，特定の刺激に対する価値の習慣的(Habit 的)な想起により表出されるものであると考える．愛着の実現には，個々の場面での価値につながる事物／事象の認識・記憶と，その場面の特徴集合の一般化による価値想起の形成の二段階があると考えられる．前者はパターン認識として，後者は概念形成として実現できよう．そう考えると，愛着については基本的にはこれまでの機械学習の枠組みでアプローチが可能である．ここまでの計算理論については既に多くの研究 [19] [20]があり，実現にあたってはそのうちのどれを選択するか，場面や対象に応じたアルゴリズムの選択が課題となる．

しかし，利益・報酬という価値に対応して前頭葉で行うとされる Tree 探索のような推論のみが，脳内の神経回路の動作として自然な形であるかという点について疑問が残

る．人の脳では約 1000 億個の神経細胞の発火のみで推論を実現しているが，その神経細胞の繋がり（ネットワーク）が Tree 構造を表現しているという知見はない．このことから従来の Tree 探索はあくまで人が認識や説明をしやすくするための手法であると言える．

さらに，感情と推論との関係を考えると両者は切っても切れない関係にあると考えられる．感情とは現在状態のみにより喚起されるものではなく，過去の経験を思い出し，その情報を基に将来起こり得る状況を予測した際にも想起される．例えば，過去の経験として仲の良い友達 A 氏と遊びに行った際に楽しかった経験をしたとする．その友達の A 氏と次に遊びに行く約束をした際に，次に遊んだときを想像し，楽しみな感情になることがそれにあたる．以上のことを踏まえると感情と推論との間には，関係があると言える．

そのため本研究では，人の感情を理解するために我々人の価値を見出しその評価をし，意思決定までを制御する価値計算システムを実装し，評価する．しかしその前に一度，人の推論に着目して人の脳として在り得るニューラルネットワークモデルを提案し，その機能評価をすることが本価値計算システム仮説の検証に取り組む前に必要であると考えた．そのため次章以降では，研究背景である感情＝価値計算システム仮説を意識した上で人の推論機能についてモデルの提案，および検証を行う．

第3章　　人の推論のモデル的解明

日々の生活において，私たちの置かれている状況は刻々と変化し，私たちは日常的に新奇の状況に出会っている．そのような場合でも，私たちは自身の置かれている状況を把握し，その状況にあった意思選択(意思決定)ができる．このような意思選択問題に対しては人工知能(AI)，神経科学などの分野で人の行動や学習の研究が行われてきた．

意思決定に関しては，場面認識から行動に直結する回路が先天的に組み込まれた反射の要素があり，先行研究として過去の多数の試行錯誤した経験を蓄積・一般化する強化学習 [21]，過去の類似経験の個別記憶を基にするエピソード記憶の利用 [19]，などが提案されている．これらは基本的に過去に経験した場面における意思決定の手法であるが，新奇の場面で一般知識を用いて意思決定する手法に推論がある．

推論では，エージェント(自律的に行動する対象)はその場で選択可能な行動の後に起こるであろう状況を予測し評価するというサイクルを反復する．その手法として従来，その場の状況を受けて次に起こる状況を意識的に予測する論理的な推論と，瞬間的に予測する直観的な推論の 2 つがあるとされてきた．しかし，人の脳認知過程においてはその発生原理やメカニズムは解明されていない．本研究では，新奇場面での推論のこの 2 つの側面とそのメカニズムに焦点を当てる．

そこで本章ではまず，推論研究として行われてきた先行研究について紹介する．そして先行研究の，人の推論機能のモデルとしての立ち位置について議論する．

3.1 人の持つ思考の特性と謎

本研究の中心となる人の推論について説明をする前に，現時点で未解決であると言える「人の思考」の理解について著者の見解を述べる．人の思考とは，広義には人が持つ知的作用の全般を総称する言葉として用いられる．さらに狭義には概念の獲得や，判断，推理を行うことを指し，本研究で取り扱う人の推論もその一部であると考えられる．

従来研究より人の推論について現状分かっていることは，1 点目に多種多様な場面に適用することができ，非常に多様で強力であるということが挙げられる．そして 2 点目に推論の発生メカニズムについては未解明である，という点が挙げられるだろう．人の推論の先行研究として見られる典型的なモデル化の方法は，人の推論機能の一部に着目し，その機能を実現するものであった．つまり，人の推論機能の一部に対する近似であると言える．その例として挙げられるモデル化に用いられる手法の種類は大きく分けて 4 つある（表 3-1 参照）．

表 3-1　推論モデルの手法の種類とその特徴

推論モデル	特徴
論理推論	記号論理学，記号処理 etc.
ファジィ推論	あいまいさを許容する論理の拡張
直観的推論	確率モデル 例）Bayes 推論
Neural Network 学習	入出力情報の近似関数の獲得

この表 3-1 に示したように，従来より多くの研究分野において人の推論を対象とした研究が行われている．しかし，推論機能の近似だけでは人の多彩な推論の全体像を見出すことはできず，人の思考を理解することもできない．そこで本研究では，推論を人の脳活動の結果として表出されるものとして捉え，これをモデルにより解明することで人の推論メカニズムに対するアプローチができ，さらに統合的な推論を説明することができると考えた．本研究を進めることにより，脳のメカニズムとしての人の推論を知ることで，従来提案されてきた複数の近似理論を統合するための路になることを期待する．

3.2 意思決定のための価値推論

本研究で我々は，推論とは意思決定のための価値のある状態への経路探索であると考える．人の意思決定が価値の計算に基づくことは，行動経済学や認知科学により示されてきた [18] [22]．その際，例えば強化学習は期待報酬予測という形で価値を計算している．古典的な推論手法である Tree 探索は離散的な予測と評価の反復による価値の最大化を目的とする行動探索である．また，反射行動は不利益な事態を避けるための行動を生成すると考えると価値計算を含んでいる．すなわち，多様な意思決定アルゴリズムは価値の最大化という意味で共通の基盤を持っている．この全体像を表したものが図 3-1 で，意思決定における反射行動，強化学習，推論という手法が，価値計算という共通の基盤を持つことを示している [23].

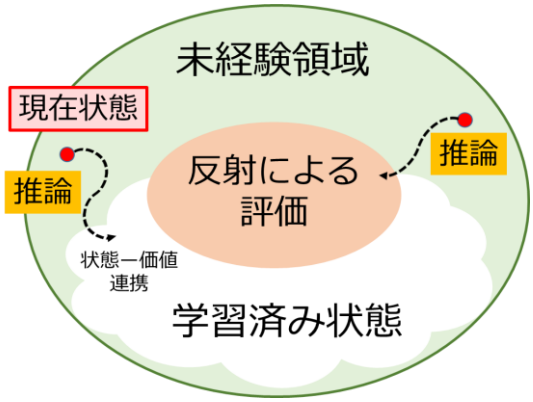

図 3-1　推論における価値探索の位置づけ

図 3-1 は，この世界で行動するエージェントが直面するであろう全ての状況を大きな円で表している．そのうちの一部は進化の過程で先天的に反射として埋め込まれており，また別の一部は過去の多くの経験から強化学習などの手法によりその状況に応じた行動学習がなされている．これらの状況内にいる間は個々の場面に応じた行動あるいは価値が割り振られており，エージェントは意思決定が可能である．しかし残りの部分は価値の割り振りのない新奇の状況であり，エージェントは推論などの方法で現在状態から価値の割り振られた状況までの行動経路を発見することで，現在状態における行動の価値を計算しなければならない．このように状態空間に価値を基準に学習アルゴリズムを配置すると，少なくとも反射，強化学習，推論は一続きとなり，それらの間の関係が明らかとなる．

3.3 認知科学における推論

推論の先行研究として，認知科学では人の推論には無意識的で自律的，さらに処理時間が短いとされる直観的推論(システム 1)と，意識的で処理時間が長いとされる論理的推論(システム 2)の二種類があるとされてきた(表 3-2) [24]．さらにそれぞれの推論の制御方法については，まずシステム 1 により必ずしも正確ではないが，ある程度の精度での推論を短時間で実行する．そして必要に応じて分析的過程であるシステム 2 で推論することで，意思決定がなされるという二重過程について議論されてきた [25].

表 3-2　推論の二重過程と二重システム仮説

直観的推論（システム 1）	論理的推論（システム 2）
作業記憶は使わず	作業記憶が必要
無意識的，自律的	意識的，メンタルシミュレーション
速い	遅い
バイアスに影響されやすい	規範的，公平
文脈依存	抽象的
確率的，分散的	論理的，シンボル的
暗黙知（経験的確率）を利用	明示的な知識を利用
推論が浅い	深い推論が可能
進化的に古い	進化的に新しい

出典：服部雅史，思考と推論: 理性・判断・意思決定の心理学 (2015)，p.174 を一部改変

私たちの日常の推論行動を考えても推論には二種類あることは容易にわかる．道を歩いていて自身の方にボールが飛んできた場合，ボールの軌道を予測し，自身に降りかかる危険度を判断してボールが当たらない方向に避ける．その際，我々は直観的な推論をしていると言えよう．それに対して，ミーティングなどの際に急いで書いた走り書きのメモを後日読み直す際，判断しにくい単語を前後の文脈から推定する過程は，論理的な推論であろう．前者を意識的に考えていては避けられないし，後者は意識的に前後を見直さなければ文脈を読み取ることはできない．

3.4 工学的な立場としての推論研究

我々人は，確率的かつ素早い直観的推論と，論理的で処理に時間のかかる論理的推論の 2 種類を持つのに対して，工学的な立場としての推論モデルは従来，Tree 探索に代表されるシンボル的な推論が提案されてきた [26]. シンボル的な推論では，命題に対して記号を割り振り，その記号の組み合わせにより真偽を判断するために状態を予測し，その状態を評価することで論理的推論としての機能を果たす．しかし，この処理に対応する脳の計算理論は現状では不明である．

一方で先述した通り，人は論理的推論とは別に，何かの認識時にその影響の予測，および評価を素早く行う直観的な推論過程も持っている．それは，感覚刺激からの自動的な連想による無意識的な予測と評価によると考えられる. そこで行われる予測は局所的でかつ多方向に並列的に進み, それを価値システムが評価したものの全体に対してこれまでは「直観」と呼んできたと考えられる [27].

このような推論の発生，および処理メカニズムを理解するために従来，動物実験の結果から推論をモデル化する研究がされてきた．Funamizu らは [2]，ラットに周りの状況をもとに推論しなければ報酬を得ることができない課題を設定し, その際の脳活動を記録し，その結果からラットの推論が動的ベイズ推定で説明できることを示した．

また Donoso らは [3]，人を実験の対象とし，課題自体に確率的な揺らぎを持たせた行動実験を行い, 人の行動選択パターンの遷移から人の推論をベイズ推定に当てはめることができることを示した．これらの研究では，比較的単純で過去の経験から行動を予測しやすい課題を用いていることから, 得られた結果は直観的推論に基づくものと考えられる．さらにこれらの研究では，その推論に対応する脳部位として前頭葉を挙げていることから，推論と前頭葉との関係が示唆される．しかし，論理的な推論が必要な課題を用いていないこと, および論理的推論との対応付けについては議論していないことから，推論に関する十分なモデルが得られているとは言い難い．さらに他の神経科学の知見においても, 実験の結果と論理的推論との関係を明確に関係づける知見はほとんど見られない.

他に工学的な手法では，強化学習手法を用いた Dyna-Q では処理を価値ベースとポリシーベースの 2 種類に分け，これらを組み合わせたモデルがある [21] [20]. この手法は一見，価値ベースの部分を直観的推論とし，ポリシーベースの部分を論理的推論とすることで推論の過程を統合しているように見える．しかし，これらの手法においてはポリ

シーベースの部分においてのみ予測をしており，強化学習の部分では経験を学習することに特化させているため予測をしていない．

さらに，これらの強化学習手法とは別に，ベイジアンネットは直観的推論に対応する活性伝播モデルの一つと考えられる [28]．ベイジアンネットでは離散状態をノードで表現し，ノード間の関係を事前に作りこむことで確率過程のパターンの計算を可能としている．しかし，この状態を表すノードは，人が事前に確率的な因果関係を考慮し作りこんでいるため，確率計算ではあるが論理的な関係の確率的な推論として捉えることが出来よう．それに対して Graves らは [29]，Deep Learning の手法を応用して人の論理的推論の再現を試みた．この手法では人の論理的推論の一つである三段論法を再現する内部手続きの獲得に成功しているが，家系図，路線図などの決まった枠組み内における論理的推論の再現のみを目的としており，人の直観的推論のメカニズムについての議論はない．また，認知アーキテクチャの一つである ACT-R を用いて記憶事項をグループ化し，そのグループ化した記憶を連想的に想起する際の潜時を示したものもある [30]が，人の二種類の推論に分けた議論はされていない．

このように従来のモデルでは，直観的推論と論理的推論とを区別し，さらに二種類の推論それぞれに対して脳内の発生メカニズムとして妥当な説明が可能な統合的なモデルは見つかっていない．そこで本研究では，この別々に考えられてきた二種類の推論を脳の連想記憶に基づく活性伝搬モデルとして統合できることを示し，さらに 2 つの推論を統合した意思決定過程のシミュレーションを行うことで，2 つの推論過程のそれぞれの特性について説明可能にし，脳の情報処理過程における推論の役割を含めたモデルを提示することを目指す．

3.5 直観的推論

3.5.1 直観的推論の特徴

直観的推論の特徴は表 3-2 で述べたように，無意識的に行われる，論理的推論に比べて推論にかかる時間が短い，などが知られている．

直観的推論は無意識であるため，局所特徴群の統合は起こらず，特徴の連想による予測と価値評価のサイクルも局所的で意識の有無に関わらず何らかの価値が計算されるという現象が生まれる．さらに，連想は多方向への分岐が起こり，神経興奮が同時並列的に多方向へ分散して確率的な予測となる (図 3-2)．ここで，世界についての知識は階層的認識の途中過程の時系列の「局所特徴 t＋行為 t→局所特徴 t+1」という連想を想定し，連想ネットワーク内に経験的に蓄積された確率モデルであるとする．

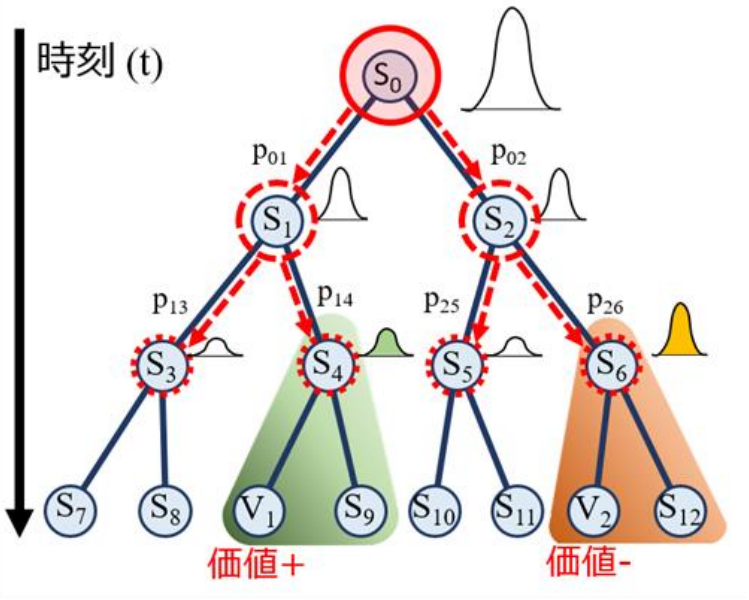

図 3-2　直観的推論の処理イメージ

例えば，物理法則・モノの操作・自己移動などによる知覚の時間変化を蓄積した predictive coding の一種と考える [32]．これらの機能の組み合わせと情報循環の動的制御により，確率的な状態予測と価値評価が神経興奮の伝搬として実現できよう．

3.5.2 粒子モデルによる直観的推論の例

本学位論文の研究の先行研究として著者らが行っていた研究として粒子モデルを用いた直観的推論の研究がある [23]. 粒子モデルによる手法は，多数の粒子を自身の置かれている状態空間に隣接する領域中に散布し，散布された状態空間中に埋め込まれている価値を探索することで意思決定する手法である.

現実世界の行動決定では，絶えず新しい状況に直面する．その際，新奇場面からリアルタイムに現在状態を基に推論し，価値の高い行動を選択する．そこで筆者が行った粒子モデルによる手法では，迷路世界で複数の欲求が次々と発生する動的なタスク場面を想定し，それに対する行動決定のための価値推論の過程をモデル化した.

モデルは「場所－価値連合」層と「確率並列探索」層の二層構造である．「場所－価値連合」層は空間地図を表現し，その一部には事前に強化学習により価値が付与されていると仮定した．(図 3-3)

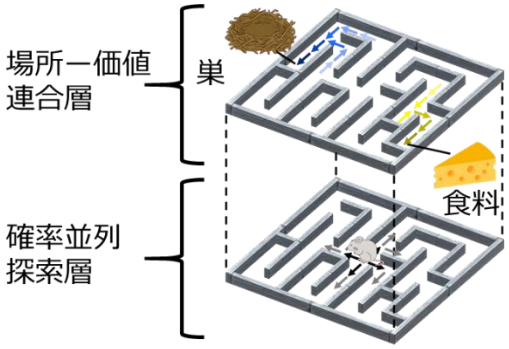

図 3-3　二層構造モデルの概要

「確率並列探索」層は，人の脳の神経細胞の興奮による連想的な伝搬を表し，事前の経験では価値が割り当てられていない新奇の場所からの連想により多方向への興奮伝搬が起きて，予測される価値が最大となる状態(ゴール地点)への価値の探索を行う．神経興奮はモンテカルロ法による粒子モデルとして表現し，「場所－価値連合」層が持つ地図知識による確率連想により，粒子を自身の置かれている環境に効率的に分布させる．この連想を反復することで，エージェントの内部での確率的な移動を表現する神経興奮が広がる．そして「場所－価値連合」層のもつ価値情報が，利用可能な状態への予測を表現する.

さらに，現実世界においては複数の欲求が同時に発生する動的な場面を想定した際，

エージェントは自身の内部により見出された複数の欲求と，現在状態にて見出されている価値の中から一つを選択して行動しなければならない．この問題を解決するために粒子モデルを用いた直観的推論ではメタシステムが必要であると考え実装を試みた．(図 3-4)

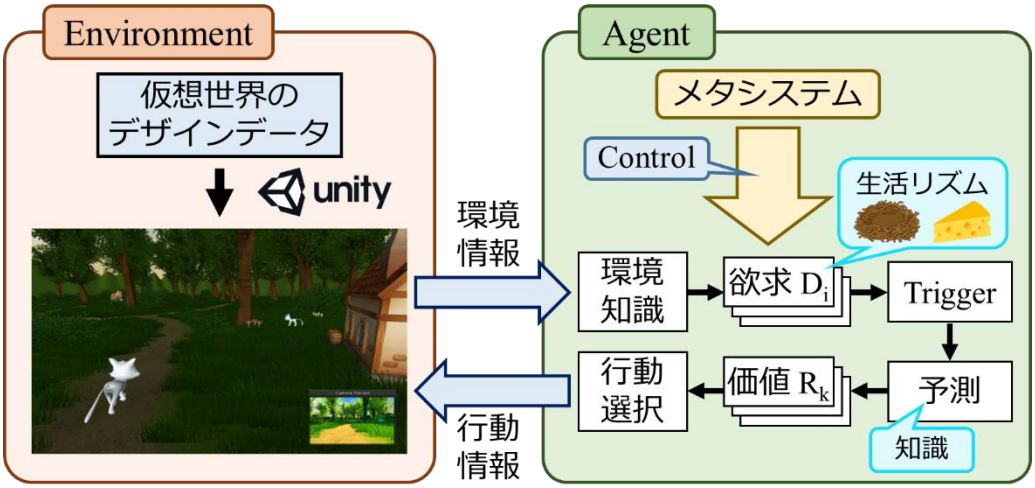

図 3-4　環境とエージェントとの処理関係

本粒子モデルを用いた直観的推論シミュレーションでは，研究の将来的な環境を考慮して，3D でのシミュレーションが可能となる環境である必要があると考え，ゲームエンジンの一つである Unity を用いてシミュレーション環境を構築した(図 3-4 左)．なお，粒子モデルを用いた直観的推論シミュレーションでは，粒子モデルを用いることにより直観的推論の実現が可能であることを確認するために，ゲームエンジンの特性であるリアルな環境とするのではなく，3D 環境をグリッドワールドに区切り使用した．仮想環境内にいるエージェントにはカメラが搭載されていることを想定し，環境からエージェントの内部には，カメラの画像が送られることを想定した(図 3-4 右)．そして環境より入力された環境情報に対して特徴量の抽出が行われ，現在エージェント自身がおかれている状態が把握できると考えた．さらにエージェントには，生活リズムにより徐々に変化する内部の状況を逐次的に判断し，その場で必要となる欲求 D_i を見出す．そして見出された欲求の内，行動価値を見出さなければならなくなる閾値を超えた欲求に対して過去の経験より得られて知識を基に推論し，欲求 D_i に対応する価値 R_k を探索する．そして探索により得られた価値 R_k と欲求 D_i の情報から次にとるべ

き行動を選択し，環境にいるエージェントに対してその行動情報を送り，シミュレーション環境中にいるエージェントが行動する．このサイクルを繰り返すことによりシミュレーションする．

粒子モデルを用いた直観的推論シミュレーションによる探索は，過去の経験より見出される事前確率の情報を含めて式(3.1)より結果を算出した．

$$argmax \sum_j D_i \, Pr(q|p) \, P_j^t R_k \;\; > \;\; Threshold \qquad (3.1)$$

この式(3.1)は欲求(i)と粒子数(j)の要素を含んでいる．さらに事前確率$Pr(q|p)$の p は現在状態を，q は現在状態の次の時刻の状態をそれぞれ表現する．なお，この事前確率は状態に達した際に事前確率が得られることを想定した．そしてその処理のイメージは過去の経験より算出された事前確率$Pr(q|p)$に従い，粒子P_jを近隣の領域に撒き，その状態空間に紐づいている価値R_kを取得する．なお粒子P_jには，自身の置かれている状況付近に多数撒き，その撒かれた先の価値情報を取得する以外の意味はない．そして，価値R_kは粒子モデルによる直観的推論課題では実数としているが，その値はその場の状況に応じて変更することが必要となる．これだけでも意思決定をすることは可能であるが，我々の住む実世界においてはエージェント自身の内部状態の表現の一つであると考える欲求D_i(粒子モデルによる直観的推論課題では 0 から 1 の範囲の実数により表現した)が意思決定に影響することは明白である．そのため粒子モデルを用いた直観的推論シミュレーションでは，事前確率や粒子，価値のみを用いて，これらを組み合わせて意思決定するのではなく，その推論する瞬間のエージェントの内部状態として推論対象となる欲求も意思決定をするための要素として考え計算式中に含んでいる．なお，計算上は欲求D_i毎にその価値の値を計算し，価値があると判断する閾値$Threshold$を超えたもののみを対象に，argmax 関数により値が最大となった価値の要素番号を取得し，意思決定する．

そして，この計算を粒子の見出した領域からの探索として考え，処理を繰り返し適用する(粒子 P_j^t の t の値を増やす)ことで，深い範囲に対しても推論することができる．

複数の報酬と，それぞれの報酬に紐づけられた価値領域，および複数の欲求に基づいた行動エージェントのシミュレーションを実現するために，計算機シミュレーションのタスクは迷路探索課題とした(図 3-5)．そして迷路中には 2 つの種類の異なる正の報酬

（赤色，青色のキューブ）を設置した．そして，迷路世界中の一部領域には事前の強化学習で期待報酬が学習されて価値が紐づけられていると仮定しているため，報酬 1（青色のキューブ）に紐づいている価値領域を水色の床で，報酬 2（赤色のキューブ）に紐づいている価値は薄いピンク色の床で表した．さらに，報酬 1 と報酬 2 の価値領域の競合している領域として黄緑色の領域を用意した．ただし，エージェントが行動する空間中に存在する報酬，および価値領域は正の報酬だけとは限らない．過去に経験した負の報酬につながる行動は避けるように行動決定をする必要がある．そのため地図上の紫色の領域を負の価値を見出す領域として設置した．ただし，灰色で示している床に対しても完全な未知の領域であるというわけではなく，エージェントは事前にランダムウォークなどにより迷路の地図に関わる情報は獲得済みであるとした．シミュレーションはゲームエンジン Unity による仮想環境と人工生命 LIS(Life in Silico)を組み合わせて実現した [33]．LIS とは 2016 年にドワンゴ人工知能研究所により開発された，ゲームエンジン Unity 上でエージェントが自らの意思決定により動作をすることができるプログラミング環境である．また，人工生命 LIS には DQN(Deep Q Network)が標準で搭載されている．しかし本粒子モデルによるシミュレーションではこの LIS の環境の内，仮想環境とエージェントとが Socket 通信し連携する機能のみに設定変更して使用している．

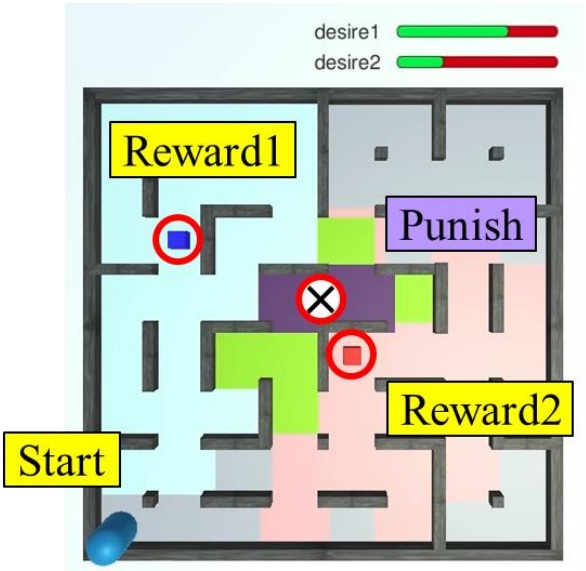

図 3-5　複数の報酬，価値領域，欲求を含んだ迷路地図

粒子モデルを用いた直観的推論シミュレーションでは，地図上の赤と青のキューブを少ない探索コストで得ることがエージェントの目的である．粒子モデルによる直観的推

論課題の 1 つ目としてエージェントは図 3-5 のスタート位置(左下)から推論と意思決定とを繰り返すことにより報酬とその価値領域の探索をした. 粒子モデルによる直観的推論では，エージェントの持つ内部欲求 D の値を評価し，探索が必要となる閾値以上となった欲求に対応する価値が見出すことができなければ, さらにもう一層深い範囲に対して探索をするようにした．なお, 粒子モデルを用いた直観的推論シミュレーションでは直観的推論における最大の探索範囲は 3 層までとした.

そして粒子モデルを用いた直観的推論手法による推論と推論回数の比較をするために，粒子モデルを用いた直観的推論シミュレーションでは Tree 探索の幅優先探索と結果を比較して 2 つの手法毎に異なる推論の特性を検証した．なお, Tree 探索には幅優先探索とは別に深さ優先探索と呼ばれる手法も存在している. しかし直観的推論の処理に近い推論を対象とすると，幅優先探索の方が近いこと，深さ優先探索では探索するルールにより結果が大きく異なることから，今回は比較対象とはしなかった.

シミュレーション時に意思決定のための探索のルールはスタート位置から探索をはじめ，探索した領域中に価値を抽出することができ，その結果を基に意思決定することができる状態に達した際に，意思決定することとした．その結果を示したものが図 3-6 である．なお，探索コストは粒子を周辺領域一度撒くことを 1 とし，探索する層が増えるごとに 1 ずつ増加するようにした．それに対して Tree 探索では，推論を開始する地点から隣接するノード 1 つを探索することを探索コストが 1 であるとした. 探索コストの計算は粒子モデルと Tree 探索とでは処理が異なることから厳密に等しいと言い切ることは難しい．しかし粒子モデルを用いた直観的推論では，それぞれの計算過程を考慮すると探索に対する概念単位で考えることで比較方法の意味が同じになると考え，このように定義した.

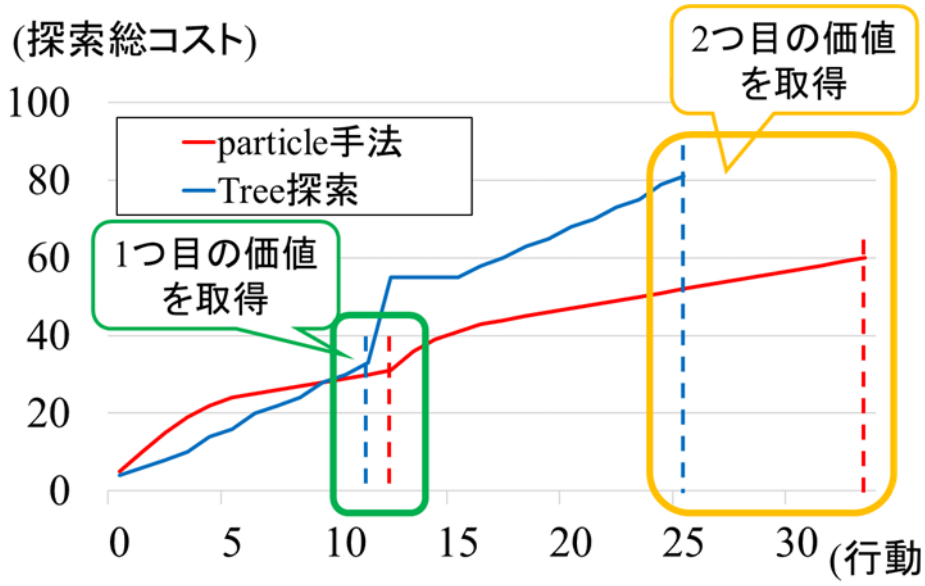

図 3-6　粒子モデルによる直観的推論結果と Tree 探索との探索総コスト比較

シミュレーションの結果，2 つの報酬を得るまでの行動数は，従来手法である Tree 探索を用いた方が 1 つ目，2 つ目共に少ない行動数で報酬を得る結果となった．しかし推論にかかった探索総コストでは，提案手法と Tree 探索とでは提案手法の方が，報酬の 1 つ目，2 つ目共に少ない探索コストで推論する結果となった．さらに Tree 探索は，現在状態からゴールまでの状態を場合によってはすべて探索する手法である．そのため，本学位論文で取り扱うような状態に紐づけられている価値を探索する手法ではない．そのためエージェントの探索は欲求状態を無視している．しかし結果として現在の位置からの移動回数の少ない順である青のキューブ，赤のキューブの順番で取得した．結果的に粒子モデルを用いた直観的推論シミュレーションではエージェントは各欲求が最大になる前に両方共の報酬を得ることができているが，欲求を考慮していない意思決定という意味では問題がある．

それに対して粒子モデルによる直観的推論手法では，内部欲求を基に状態空間に紐づけられている価値を探索し，その結果から意思決定する機能がある．そのためエージェントの得た報酬は，欲求の大きい順である赤いキューブ，青いキューブの順番となった．エージェントははじめ，赤の報酬と青の報酬の両方に対応する内部欲求があると判断され，その大小関係は赤の報酬に対応する内部欲求の方が高くなる．そのため，エージェ

ントの基本行動方針は，価値の高い赤の報酬に向かって行動することになる．しかしエージェントの初期位置から 4 ステップ目終了時までは探索範囲内に青の報酬に対応する価値しか抽出できない．そのため，エージェントは価値の小さい青の報酬に向かって行動する．その後，内部欲求の高い赤の報酬に対応する価値が抽出できた際には，赤の報酬に向かい行動選択をした．そのため，欲求に多寡に従った順番で行動する意思決定ができた．

さらに，推論特性を調査するためにエージェントのスタート位置を図 3-5 の迷路上の四隅として 4 つの条件として追加シミュレーションを実行した．シミュレーションを実施する際の推論の探索範囲の条件は①現在地点から探索し，最も近い範囲において価値領域を見出すことができたら 1 回行動し，その後再度推論と行動とを交互に繰り返しながらゴールを目指す方法，と②現在地点から報酬の位置が見いだせるまで深く推論し，報酬が見いだせたら 1 回行動し，再度報酬まで推論するという行為を繰り返しながら報酬の獲得を目指す方法の 2 種類を用意した．

①の方法では，4 つの条件の内，3 つの条件について提案手法の方が Tree 探索の探索総コストよりも小さい結果となった．提案手法ではすべての条件において内部欲求の高い赤いキューブ，青いキューブの順番で報酬を得ることができた．しかし左上の位置から開始した場合のみ，Tree 探索と比較して探索総コストが高い結果を得た．その理由として開始位置が左上の条件においては，開始位置から比較的近い範囲に赤いキューブに対応する価値を見出すことができず，探索範囲を広範囲にしなければならないため，探索総コストが 30 となり，Tree 探索の 12 よりも多い結果となった．しかし，他の 3 条件と合わせた 4 条件すべての探索総コストは Tree 探索に比べ，平均 22.5％削減となった (図 3-7).

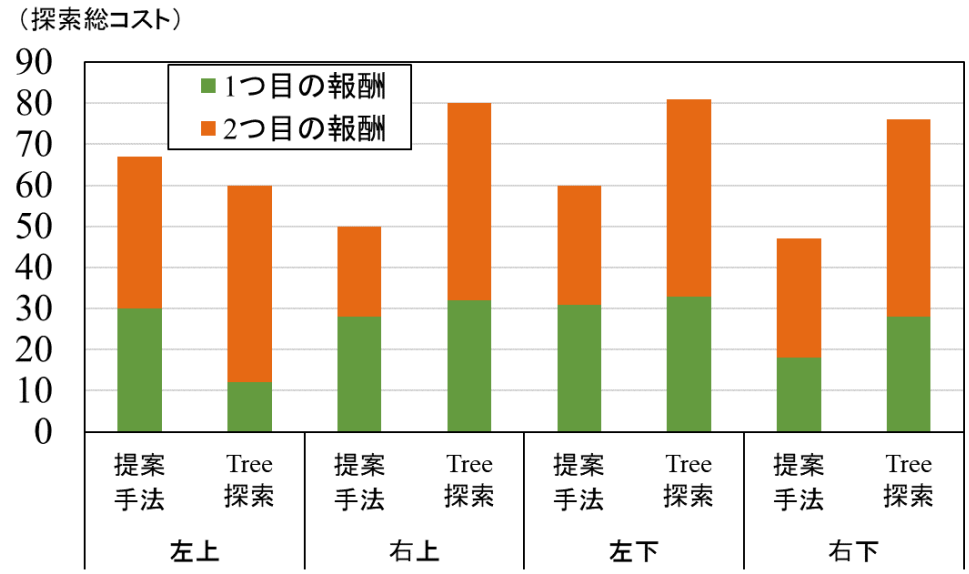

図 3-7　スタート地点を地図中の 4 隅とし，探索範囲を限定した場合の結果

提案した粒子モデルの特性を知る上で，探索範囲が狭いもののみではその特性を知り尽くすことはできない．そのため，②の探索時に現在位置から報酬の位置まで推論し，その結果に基づき行動することを繰り返す方法を用いたシミュレーションが必要であると考えた．その結果，すべての 4 条件すべてにおいて提案手法の方が Tree 探索よりも探索コストがかかる結果となった(図 3-8).

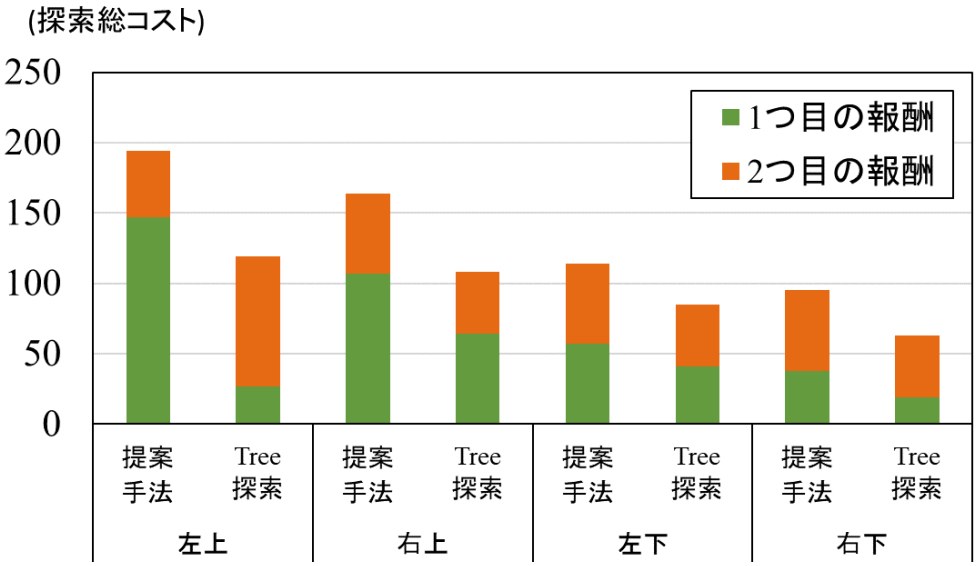

図 3-8　スタート地点を地図中の 4 隅とし，探索範囲を拡大した場合の結果

提案手法が Tree 探索よりも探索コストが高い結果となった原因として，粒子モデルを用いた直観的推論シミュレーションでは環境をグリッドに分けた迷路としている点が挙げられる．そのため，現在時点からの探索範囲に制限(最大でも上下左右の 4 種類)が発生する．そのため，Tree 探索は 1 つの条件辺りに探索する条件数が少なく計算パターンの少なく済む設定であったと考える．現実世界では粒子モデルを用いた直観的推論シミュレーションのような次の時刻として起こり得る状態が限定されていることはなく，むしろ取りうる行動条件も多岐にわたることが多い．そのような場合には Tree 探索を用いる場合では探索する条件数が次の時刻として起こり得る状態の組み合わせとその条件それぞれに対する行動数の積だけ増える．それに対して本粒子モデルを用いた手法では，1 度の探索で 1 層分一度に探索するため，探索条件が増えても計算回数は探索した層の数が上限となる為，状態数や行動に依存しない．

さらに，比較した探索手法の両方において探索範囲を狭くした条件よりもコストが多くかかる結果を得た．これは①の手法に対し，遠くの報酬の位置まで行動ごとに探索を繰り返す必要があるため，探索総コストが増加することは自明である．

なお，推論によりスタート位置から報酬の位置まで一気に推論し，その経路情報を作業記憶などで記憶しておき，スタート位置から報酬の位置までを一気に行動する方法もシミュレーションとしては考えられる．しかし，粒子モデルを用いた直観的推論ではこのシミュレーションについては実施していない．その理由は，推論をスタート位置のみの一度しか実行しない場合，人のようなその場の状況に応じた柔軟な行動をすることができないためである．

このように，粒子モデルを用いて事前確率に従った直観的推論の実現をすることができた．しかし，粒子モデルを用いた直観的推論シミュレーションでは人の深く，かつ意識的であると言われている論理的推論の創発方法については議論していない．

3.6 論理的推論

3.6.1 論理的推論の特徴

論理的推論の特徴は，表 3-2 で述べたように意識的に行われる，直観的推論に比べて推論にかかる時間が長い，などが知られている．それではこの論理的推論はどのようなモデル化が可能となるのだろうか. 本稿で考える論理的推論の処理過程は以下の三種類の機能要素からなるとする．

1) 価値の焦点化 : 直観的推論によって見出された，認識状態中に含まれる価値に焦点を当てる．これは図 3-9 の S_0 の位置であることを仮定した際に，直観的推論は過去の経験から見出される事前確率 p_{01} や p_{02} に従い S_1 および S_2 を見出す．価値の焦点化とは，この見出された S_1 および S_2 に焦点を当てることに相当する．
2) 価値の長期予測 : 現在認識している状態がこのまま続いた際に，その価値がどのように変化するかを予測する．これは，1)と同様の場面を考えた際に，長期推論すると S_1 と S_2 の価値関係の大小関係の変化に相当する．
3) 価値調節系 : 見出されている価値を状態空間に反映し，認識状態を更新する．これは例えば現在位置を S_1 とした際の推論として，S_4 に価値があることを推論した際に，その S_4 に見出されている価値の一部を S_1 に反映することに相当する．

これらの機能要素の処理を繰り返すことで，認識状態に関連した価値に対して将来的な価値の変化を予測することができる．さらにこれを反復することで，単一の価値が支配的になった際には，意思決定に用いることができる(図 3-9).

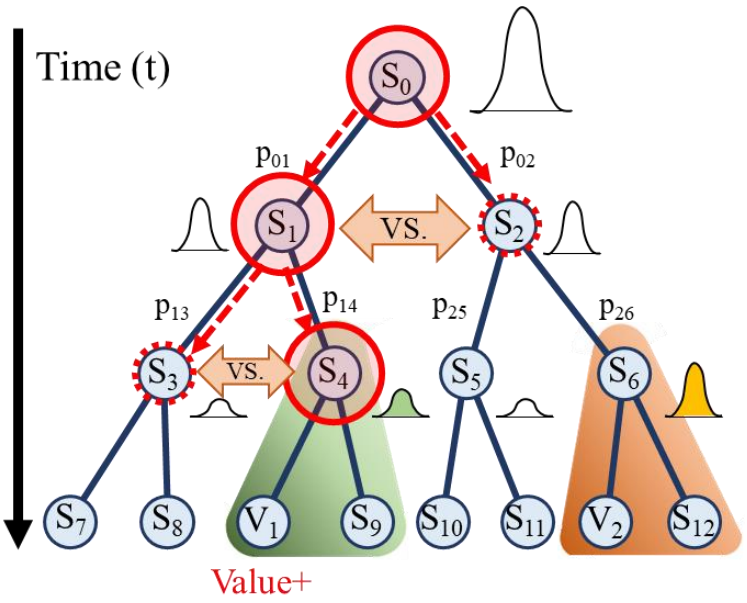

図 3-9　論理的推論の処理イメージ

3.6.2 Tree 探索を用いた論理的推論の例

従来研究の多くでは，論理的推論は Tree 探索により実現されてきた．Tree 探索とはそれぞれの状況をノード（節点）として表現し，そのノード間の繋がりを指定することで表現されることが多い．その例として図 3-10 のような〇×ゲームの戦況状態の記述が一般的によく知られている．〇×ゲームとは縦横 3×3 マスで構成された盤面に対して，プレイヤー2 名がプレイヤー毎に決められた記号（〇または×）を交互にマスに記載し，先に縦，横，斜めの何れか一列に自分の記号（〇，または×）を並べたプレイヤーが勝ちとなるゲームである．なお，本稿ではプレイヤー1 には〇が，プレイヤー2 には×が割り当てられているものとする．図 3-10 ではノードの根本にあたる場所を現在の盤面情報とした際の今後起こる状態を表現したものである．

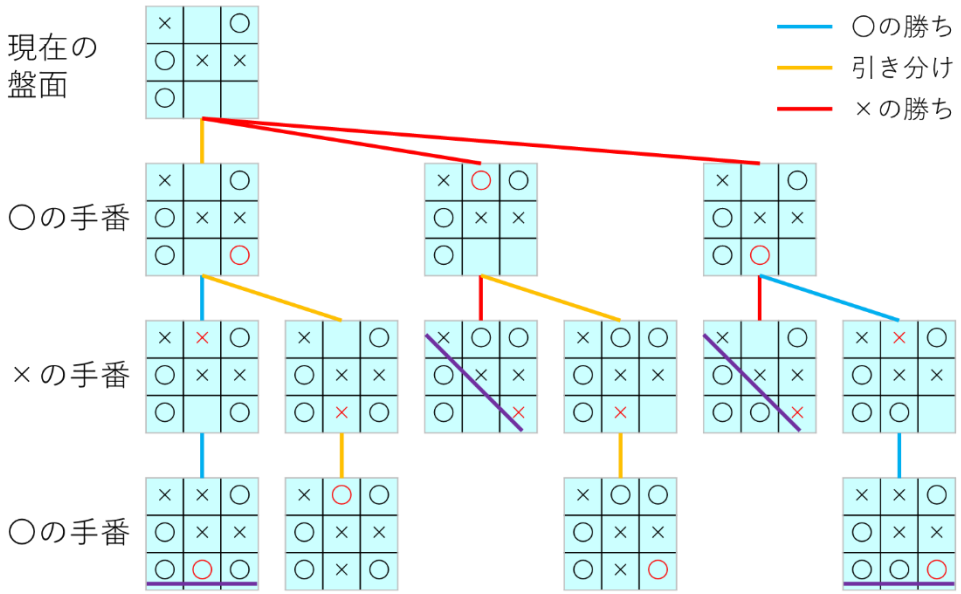

図 3-10　Tree 探索の例

さらに，Tree 探索の探索方針の代表例は 2 種類あり，それぞれ深さ優先探索と幅優先探索と呼ばれている．深さ優先探索とは，探索時に自身の置かれている現在状態からできる限り深い範囲を優先的に探索する探索手法である．探索の目的は最大の報酬を得ることである．この手法は，目的となる報酬が現在の位置から遠く，かつ探索した方向に存在する場合には探索回数が少なくて済むため，このような条件下においては有用である．しかし，報酬の位置が探索を始めた方向と異なっている場合，仮に報酬の位置が探索する地点に隣接する場合であっても探索コストが高くなるというデメリットも持っ

ている．

それに対して幅優先探索とは，現在地点からの距離が近い位置から順に探索する探索手法である．そのため，現在地点から近い距離の位置に報酬が存在する場合に有用となる手法である．しかし，報酬の位置が現在地点から遠い深い位置に存在していた場合には，探索総コストが大きくなるという特性がある．

このように Tree 探索を用いることで，人の論理的推論を説明することはできる．しかし，人が生活する上では論理的推論のように推論に時間をかけることのできる場面しかない訳ではない．生活していると急に自分の顔に目掛けてボールが飛んで来た時にそのボールが当たることを咄嗟に予測して避ける，車を運転している際に急に人が飛び出てくるなど，咄嗟に次に起こる出来事を予測し，その結果を基に次の行動を決定することを求められる場面が起こりうる．今回の例では，時間をかけて推論していると自身の顔にボールが当たり怪我をする，車でぶつかってしまったことにより相手を怪我させてしまう，などの状況が起こり得るだろう．このような状況を回避するための推論として直観的推論が存在することが考えられる．

さらに，人はこの推論という現象を脳の神経細胞の連続的な発火により創発している．先行研究において人の神経細胞一つ一つで状況を表現しているという研究は存在していない．さらに，脳科学の分野において，直観的推論は Tree 探索が発生するメカニズムの基となっているとする研究はない．このことからも Tree 探索のみでは人の推論現象全般を説明するには不足している部分があると考える．

3.7 直観的・論理的推論を含んだ統合アーキテクチャ

これまでに示してきたように，従来研究では人の推論を直観的推論と論理的推論の 2 種類に分け，別々のモデルとして人の推論と呼ばれる機能をモデル化してきた．しかしそれでは，従来と同様の人の推論として知られている機能の近似に終わってしまう．更に，著者らはこの人の推論システムは 2 つの別々の手法により実現されるのではなく，同一のメカニズムのスイッチングにより実現されることが必要であると考えた(図 3-11).

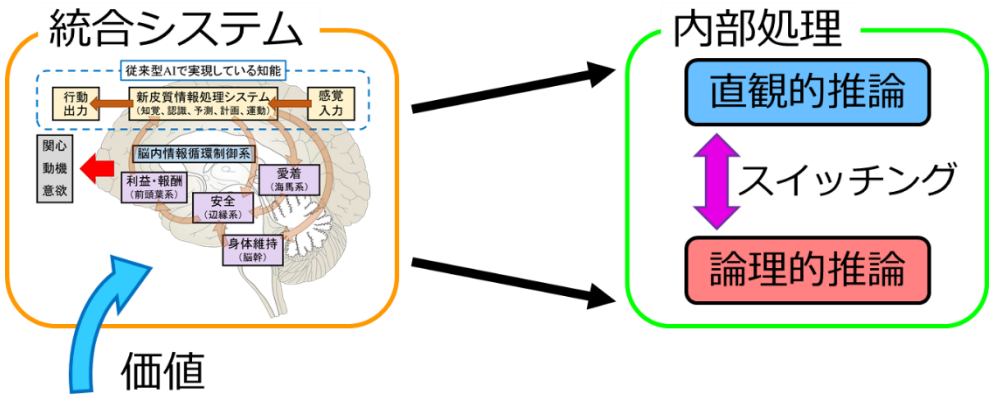

図 3-11　推論システムに求められること

この推論を含んだ統合システムの機能には，直観的推論と論理的推論の両方を実現できる機能があり，さらにこの 2 つの機能を任意のタイミングで切り替えることができることで人のような直観的推論と論理的推論の 2 つの振る舞いが表出する機能があると考える．さらにその統合システムに価値が入力された際には，その情報を推論機能においても処理することができることを想定する．これを踏まえて本研究の中心となる推論では，直観的推論と論理的推論の統合モデルとして図 3-12 のようなアーキテクチャを提案する．

本アーキテクチャでは，まず人の視覚や聴覚などの五感に相当する情報を感覚入力として入力する．入力された情報は感覚毎に特徴量を抽出し，その後基本的には入力された情報から次の時刻の状態空間を予測する．そして予測結果を個別価値計算システムに渡し，入力情報に対する Gain(利得)を計算し，現在状態からの予測結果に価値が見出されるか判断する．その結果は脳内情報循環制御系により再度新皮質システムに送られ，その結果に対応する行動を出力する．なお，この脳内情報循環制御系によ

り情報循環する処理は，本研究では人の持つ脳波によるものであると考える．これらの処理を実現することで，直観的推論で言われている推論にかかる時間が短くかつ並列的に行動を計算することに相当すると考える．そして直観的推論によって行動し，ある時ある感覚から見出された特徴に対して認識(意識)した際には，その特徴量から関連する情報がないか他の特徴量を連想する．その連想した特徴群を個別価値計算システムに渡し，入力情報に対する Gain を計算する．これらを繰り返し計算することで，次に起こる事象を直観的推論より正確に予測することができると考えた．

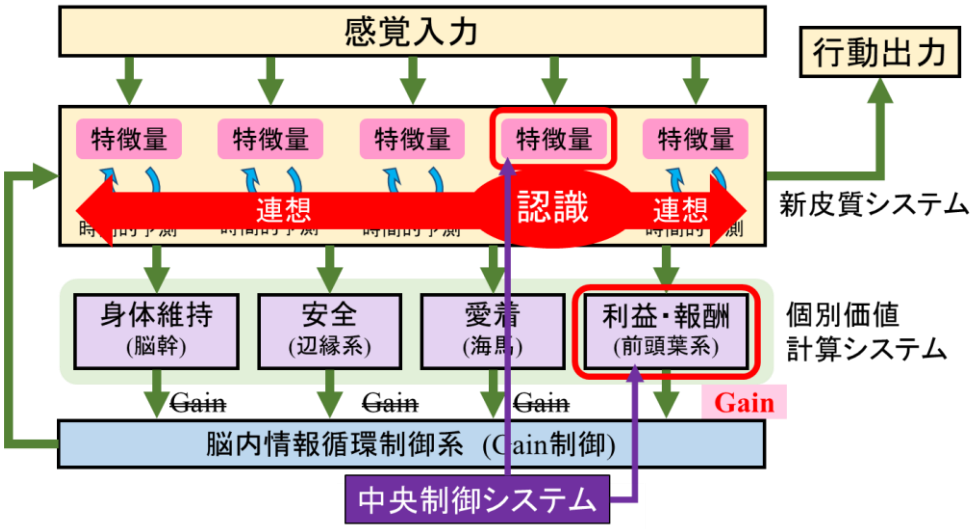

図 3-12　直観的推論と論理的推論の統合アーキテクチャ

また入力情報から見出された特徴を認識する，個別価値計算システム内の Gain を見出す部分は勝手に決まるわけではなく，自身の内部状態(欲求など)を基に選択されるだろう．これを実現するためには，中央制御システムが必要であると考える．中央制御システムでは主に内部状態の閾値を基に現在するべき行動なのかを判断する機能を担う必要がある．これらの機能を実現することで人のような多岐にわたる推論システムを説明することができると考えた．

第4章　　連想記憶モデルによる推論

4.1 連想記憶モデル

連想記憶とは，記憶パターンを分散的に貯蔵し，部分的な記憶情報を基に必要な記憶を読み出すことを言う．連想記憶手法を始めに提案した中野 [34]は連想記憶のイメージを図 4-1 のように一枚の絵で表現している．一般に記憶の過程には 3 つの段階があるとされ，それぞれ記銘，保持，想起の 3 段階がある．記銘とは，私たち人が外部から受けた何かしらの情報を記憶に取り入れることを指す．そして，保持とは記銘した内容をそのまま保存しておくことを指す．そして，想起とは保持により記憶として貯め込まれている情報を思い出すことを指す．ここでは，記銘するべき絵が記憶装置に提示されることで記憶装置に分散的に記憶(記銘)される．その後，この記憶装置にまた別の絵を提示すると，前の絵の情報を保持しながら新規の絵の情報を記銘する．

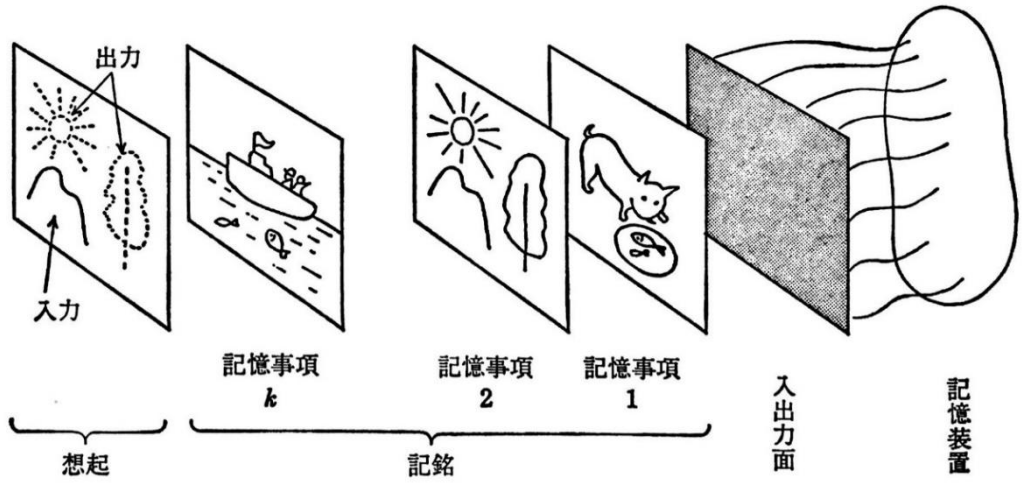

図 4-1　連想記憶のイメージ

出典：中野馨，アソシアトロン－連想記憶のモデルと知的情報処理－(1979) , 昭晃堂, p.14

このようにして記銘された絵を思い出す(想起)する場合には，想起したい内容の一部(図 4-1 で言えば山)を記憶装置に入力すると，その他に映っているもの(ここでは太陽と木)が想起される．

それでは，どのようにして連想記憶の機能を実現するか．連想記憶では，記憶内容は

一つの神経細胞の発火状態を一つのベクトルで表す．これを多数の神経細胞の組み合わせにより記憶内容を表現する．定式化する上では，以下のようにベクトル$\boldsymbol{x}^p$を用いて表現される(式(4.1)，ここで N は神経細胞数とし，一般には$\boldsymbol{x}^p = \{x_i^p \in \{1,-1\}: i = 1 \dots N, p = 1 \dots Q\}$となる)．

$$x^p = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix} \tag{4.1}$$

そして記憶を保持しておく記憶ネットワークWは，以下の式(4.2)により表現される．なお，記憶パターンの番号は p とした．

$$W = \sum_p x^p \frac{1}{N} x^{p\mathrm{T}} \tag{4.2}$$

本記憶ネットワークは，出力ベクトル$\boldsymbol{x}^p$ (N×1)と入力ベクトルの転置$\boldsymbol{x}^{p^T}$ (1×N)とを掛け合わせ，これを記憶パターンの総数分足し合わせるため，そのサイズは N×N となる．

さらにこの記憶ネットワークに埋め込む記憶パターンを作成する際，各記憶パターン間の情報の直交性が重要である．例えば，ここに 2 つの記憶パターンに対応するベクトル$\boldsymbol{x}^p$と$\boldsymbol{x}^q$があったとする．この 2 つのベクトルの要素が共に完全に同一である場合，これら 2 つのベクトルの間の内積を取り，そのベクトル長である N で割ることでその結果は 1 となる．それに対して 2 つのベクトルが$\boldsymbol{x}^p = \{x_i^p \in \{1,-1\}: i = 1 \dots N, p = 1 \dots Q\}$より構成されており，これらが異なっている場合，これら 2 つのベクトルの内積を取り，その結果をベクトルの次元 N で割ると結果はほぼ 0 になる(式(4.3))．

$$\frac{1}{N} x^{q\mathrm{T}} x^p \begin{cases} = 1 : p = q \\ \fallingdotseq 0 : p \neq q \end{cases} \tag{4.3}$$

そして，想起されたベクトル$\boldsymbol{x}^r$はこの記憶ネットワークに入力ベクトルを右からかけることで表現することができる(式(4.4))．この式の結果，式(4.4)の右辺のはじめの$\boldsymbol{x}^p$より後は 0 または 1 となる．そのため，想起されたベクトル$\boldsymbol{x}^r$では記銘時に入力元べ

クトルx^p共に記銘された想起ベクトルx^qの記憶パターンが結果として得られることになる．

$$x^r = Wx^p = \sum_q x^q \frac{1}{N} x^{p\mathrm{T}} x^p \tag{4.4}$$

そして，想起された状態ベクトルx^rの想起の強度は記憶ベクトルx^pのすべてとの間で相関をとることで確認することができる(式(4.5))．

$$correlation = x^{p\mathrm{T}} x^r \tag{4.5}$$

なお，連想記憶モデルには二種類あり，それぞれ相互想起型と自己想起型と呼ばれる．相互想起型は入力情報に対して別の記憶情報を想起する手法である．それに対して，自己想起型は入力情報の一部を入力すると入力情報の全体を想起するための手法である．この先ではこの二種類の連想記憶のモデルについて詳細を説明する．

4.1.1 相互想起

相互想起型の連想記憶は入力ベクトルから入力ベクトルとは別のパターンのベクトルを想起する際に用いる手法である. あらかじめ保持されている記憶ネットワークに入力ベクトルを与えて想起ベクトルを得るという意味で, 機械学習のパターン認識と同様のものとして考えることができる.

その処理は以下の通りである．時刻 t における状態ベクトルは時刻 t+1 の状態ベクトルに対して相互にすべての要素間で全結合している(図 4-2). そのため, 時刻 t+1 の状態ベクトルが複数種類あった際, それぞれの結合強度により影響され想起された際の想起ベクトル中に含まれる要素が変化することになる.

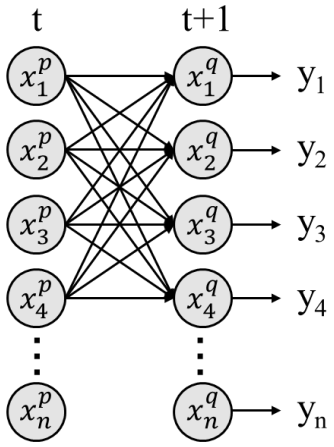

図 4-2　相互想起モデルの処理イメージ

その実現は，記憶事項はベクトル$\boldsymbol{x}^p$(p は記憶パターン番号，式(4.1))で表現することができる．その実現には，想起するためのトリガーとなる入力ベクトル$\boldsymbol{x}^p$と，想起する対象となる出力ベクトル$\boldsymbol{x}^q$を用意する．この際，ベクトルの次元は記憶する事象の数のみに依存する. 相互想起ネットワークW^mは式(4.2)を出力ベクトル$\boldsymbol{x}^q$と入力ベクトル$\boldsymbol{x}^p$にそれぞれ対応させ，式(4.6)より実現することができる.

$$W^m = \sum_q \sum_p x^q \frac{1}{N} x^{p\mathrm{T}} \tag{4.6}$$

さらにこの記憶ネットワークを作成する際には，式(4.3)に示したような各記憶パターン間の情報の直交性を持たせる必要がある．この直交性を持たせることで，式(4.6)により作成された記憶ネットワークに対して右側から現在状態ベクトル$\boldsymbol{x}^p$を掛けるこ

とで想起されるベクトルx^rは記銘時にx^q共に記銘されてベクトルx^qの種類，および強度の結果が合成された時刻 t+1 の情報を取得できることになる(式(4.7)).

$$x^r = Wx^p = \sum_p x^q \frac{1}{N} x^{p\mathrm{T}} x^p \tag{4.7}$$

4.1.2 自己想起

自己想起型の連想記憶は，記憶行列内にある記憶パターンの内，入力ベクトルの一部を与えるとそのベクトルの全体を想起させるものである．その例は，ある人を見た際，その人と過去にあった出来事を思い出すことが当たるだろう．

その処理は，入力ベクトル(時刻 t)と出力ベクトル(時刻 t+1)とを一致させ，出力ベクトル(時刻 t+1)を次の時刻の入力ベクトルとすることで徐々にその記憶ベクトルとの差をなくし収束する方向に想起強度が強くなる(図 4-3)．ここでは，時刻 t での入力ベクトルを$x_i = \{i = 1 \dots N\}$とし，自己想起ネットワークにより，自身の全体のベクトル情報を想起する．そして結果の状態ベクトル$y_i = \{i = 1 \dots N\}$が想起された内容として出力される．

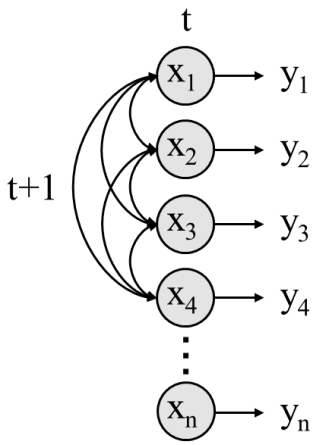

図 4-3　自己想起モデルの処理イメージ

その実現は，記憶事項はベクトルx^p(p は記憶パターン数，式(4.1))で表現され，入力ベクトル，出力ベクトルの両方にベクトルx^pを適応することで表現することができる．この際，ベクトルの次元は相互想起モデル同様に記憶する事象の数に依存する．自己想起ネットワークW^sは式(4.2)とすることでそのまま実現できる．

そして，式(4.4)の記憶ネットワークWを相互想起ネットワークW^sに変更し，ベクトルx^pをかけることで想起できる．この計算を 1 度しただけでは完全な情報を想起することはできないが，式(4.4)の出力結果として再度入力ベクトルとして計算することを繰り返すことで徐々に記憶ベクトルの情報と同一のものに収束していく．

4.2 連想記憶を用いた先行研究

アソシアトロン [35]とは 1972 年に中野により提案された人の脳の複雑でかつ優秀な情報処理を再現することを目指した手法である．狭義には，脳の仕組みを真似て構成した記憶装置のことを指す．人の脳の情報処理に見られる特徴の例として，記憶が連想形式となっており繋がっていることが挙げられる．アソシアトロンの基本原理は数学的には相関行列を応用することで実現できるものである．この処理を行うと人の神経回路網に似た分散的な記憶構造が現れる．この分散的な記憶構造がなされるため，記憶内容上の一部の情報が失われてもその記憶情報の全体が失われるわけではなく，一部が欠けただけの情報として抽出される．

さらに連想記憶手法の応用として，Sompolinsky [36]は時系列データと記憶パターンの周期を想起することのできるニューラルネットワークモデルを提案した．このモデルではニューロン間の結合を非対称とし，入力情報にノイズを加えたデータに対して時系列的な連想をシミュレーションにより研究した．その結果，動物の運動系における周期的なパターンに類似した動作パターンが発生することを示した(図 4-4).

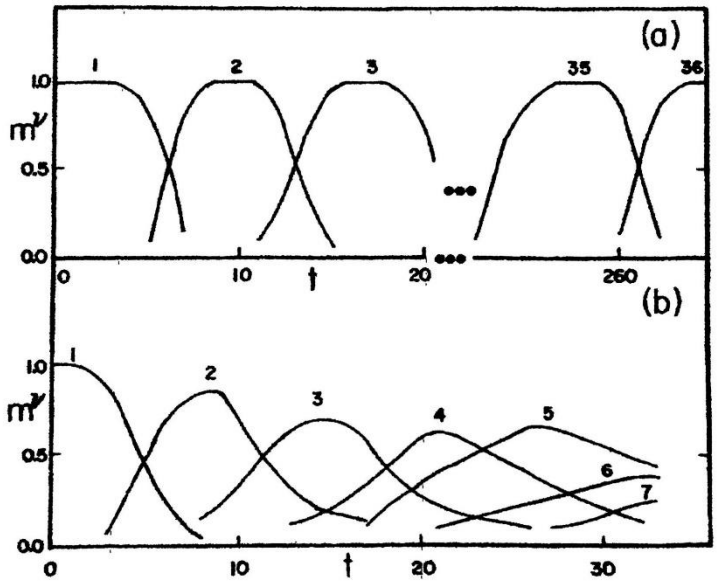

図 4-4　ホップフィールドモデルによる時系列連想の変化

出典：H. Sompolinsky，Temporal Association in Asymmetric Neural Networks, Physical review letters, (1986),p.2863

さらに，連想記憶を用いた応用的な研究として大森ら [37]は，人の知能の重要な特徴としてシンボルの使用を挙げており，動物の行動とそのシーケンシャルな処理に連想記憶を適用し，その概念操作の記号処理として注意に関して計算モデル化した．さらに脳を連想記憶と動的な注意システムの複合体であるとし，脳内における情報表現と動作に関する過去の経験からの記号処理と一致する記号的行動の出現する PATON モデルを提案した(図 4-5).

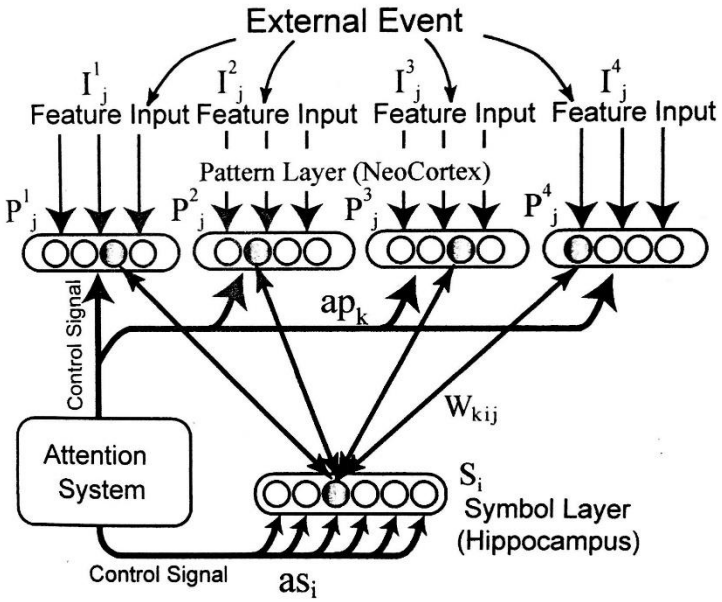

図 4-5　PATON の基本構造

出典：Takashi Omori et al，Emergence of symbolic behavior from brain like memory with dynamic attention., Neural Networks, (1999), p.1159

大森らは，主に連想記憶を用いた記号処理の計算理論を中心について述べており，さらに言語を用いた記憶探索の妥当性の検証にコンピュータシミュレーションと移動ロボットを用いた.

これらの例のように，連想記憶と人の記号処理についての研究はあるが，大森らも人の直観的推論の実現方法については議論していない．さらに連想記憶を用いた従来認知科学で言われてきた人の 2 つの推論に関して研究してきたものはなく，その実現方法についても検討が必要である.

4.3 連想記憶による推論システムの実現

4.3.1 推論システム実現の基本的アイデア

これまでに連想記憶の基本的な機能について説明して来た．本研究ではこれら 2 つの相互想起型と自己想起型の連想記憶とを組み合わせて人の推論システムの特性の再現を目指す．しかし従来の連想記憶モデルでは，想起元となる記憶ベクトルと想起される記憶ベクトルとの関係は一対一であった．しかし記憶情報は，記憶情報間の関係として必ずしも一対一の関係になるわけではない．我々は日々の生活において新奇の場面と直面し，過去の一部の類似情報を基に行動決定をしている．このことから直観的推論を実現する際，入力情報に対して複数状態の抽出が可能な特性が必要となる．

そして一般的に連想記憶の手法と，推論にかかる時間が短いとされる直観的推論と推論にかかる時間が長いとされる論理的推論とは別のものであると考えられることが多い．連想記憶による直観的推論と，論理的推論の速度的，および意識的な観点より，その処理イメージを図 4-6 に示す．図 4-6 は横軸に記憶する全事象を取り，縦軸に現在状態(想起の強度)を取ったイメージを示す．図中における紫色の丸は現在状態を表す．本学位論文で定義する直観的推論の状態変化は，青の矢印で示したように相互想起計算により，一度の計算で次の事象へと連続的に変化することで実現できるとした．そして，論理的推論は自己想起計算をする際，入力ベクトル中に含まれる記憶パターンの想起を誤差が最小となる状態に収束させるエネルギー関数により実現できると考えた(式(4.8)　α は 1 度に更新する重みを決めるパラメータ).

$$\frac{dx_i}{dt} = -\alpha \frac{\partial E}{\partial x_i} \tag{4.8}$$

エネルギー関数を用いることにより誤差が最小となり収束する記憶ベクトルのイメージは，図 4-6 の赤の矢印のように一つの記憶パターンは自己想起ネットワークを通すことにより，自身の中に含まれている想起強度が強い状態に徐々に収束することである．これを可視化すると図 4-6 のような構造をしており，その内容は最急降下法と同様の働きをするものである．

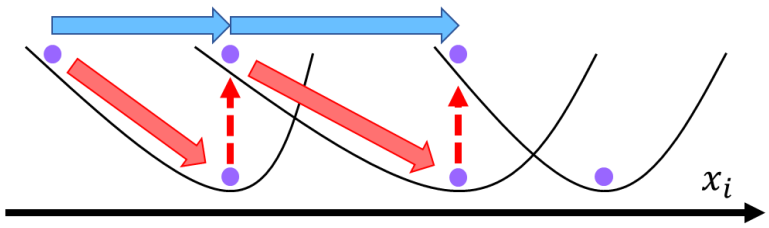

図 4-6　直観的推論と論理的推論の統合アーキテクチャ

本研究で目指す推論の型は図 4-7 に示す Tree 構造とする．記憶パターンは Tree の各ノードに対応し，探索は初期状態S_0から始まり，状態遷移は 2 分木が 3 階層だけ継続し，正の価値のあるノードV_1あるいは負の価値のあるノードV_2の周辺領域に推論が到達した場合に意思決定ができるとした．各ノード間には過去の経験により状態遷移確率Pr$(post\ \ state|pre\ \ state)$が割り当てられていることを想定しており，エージェントは探索時にこの確率を獲得しながら学習する．さらに現在状態から直観的推論を用いて状態遷移する際，その想起強度は対象の状態ベクトルに対応して記憶ネットワークに埋め込まれている想起ベクトルが合成ベクトルとして表現される．

一般の探索問題の研究では，より深い複雑な課題を扱うことが多いが，本節では初めにより基礎的な探索行動の創発を示すことを目指すため，単純なトイ問題を対象とする．

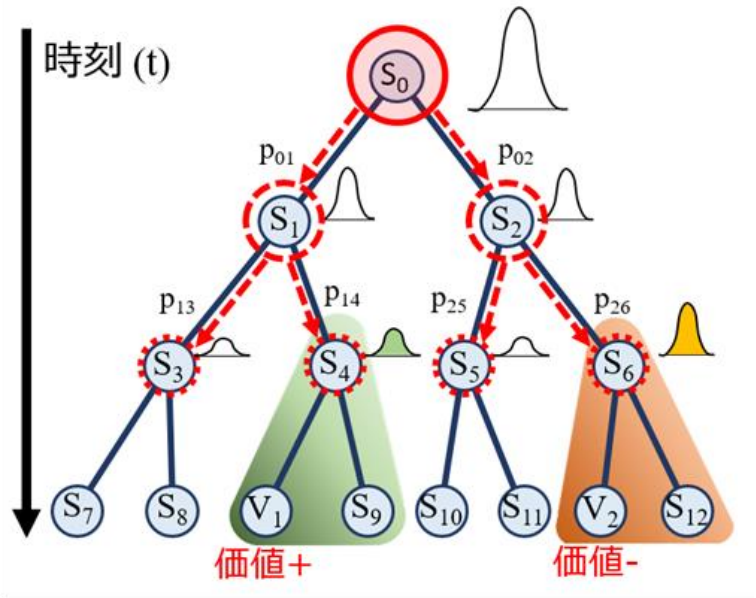

図 4-7　連想記憶を用いた推論に用いる 3 層構造 Tree

4.3.2 連想記憶を用いた推論アーキテクチャ

本研究で想定する連想的な推論システムの全体像は図 4-8 に示したように，連想計算層と価値認識層の二層からなる．その処理はまず，各感覚入力において特徴量として抽出された時刻 t の入力ベクトルx_tを受け取る．そして入力ベクトルの情報が相互想起，または自己想起のネットワークにその瞬間の想起パターンとして送られる．そして，各機能により見出された想起結果が出力ベクトルx_{t+1}に入る．その後出力ベクトルx_{t+1}はさらに価値認識層に送られ，その瞬間の身体維持機能などから見出される欲求などの情報に従いその場の状況に身体からニーズなどを基に価値を割り当て，これらの情報を組み合わせて用いることで状態評価，および意思決定につながる．

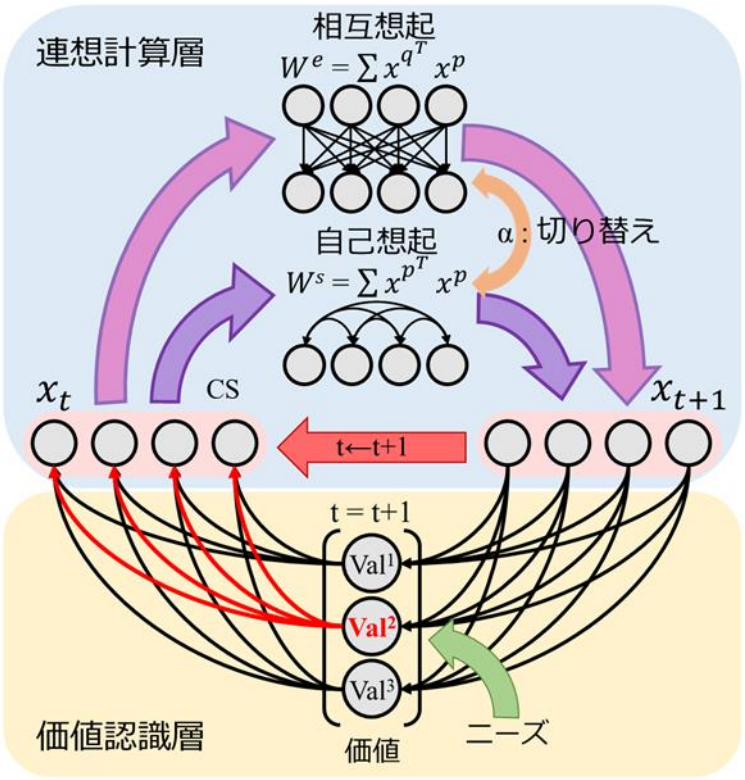

図 4-8　推論システムの全体像

直観的推論では，入力ベクトルx_tは相互想起用の記憶行列W^eにかけ合わせることで次の状態ベクトルx_{t+1}を得る．次の状態ベクトルx_{t+1}は次の時刻 t+1 の入力となる，直線的な連想の連鎖を想定する．しかし実際には一つの状態からの連想に分岐があることが多いため，複数の記憶パターンが混在した想起パターンが得られ，さらに次の連想で

分岐が広がるという，並列的な想起およびそれに伴う探索がなされる．直観的推論は処理が単純であるため，無意識的つまり自動的な過程で実現できるが，多くの記憶の混合パターンが想起される場合には個々の記憶パターンのゲインが小さくなるため，深い推論は実現できない．さらに次の状態ベクトルx_{t+1}に対して混合された記憶パターン中に含まれる状態空間中に対応する価値を付与することで，意思決定にも使用できる．

それに対して論理的推論では，直観的推論の結果として得られた次の状態ベクトルx_{t+1}中に，複数の記憶パターンが混合したパターンが想起された後，そのうちの一つの記憶パターンを選択的に想起する自己想起過程を追加することで，記憶パターンの混合を回避した深い推論を実現する．実際には次の状態ベクトルx_{t+1}に対し，現在状態から見出された価値を付与することで，状態ベクトルを変調する．これより，次の状態ベクトル中の混合パターンの割合を変化させる．その変化させた状態ベクトルに対して再度，自己想起過程を行うことを繰り返すことで，状態ベクトルの収束を促す．そのため，直観的推論に比べて処理回数が多く，結果を導くまでに時間がかかる．これが連想記憶の自己想起過程を用いた際に，従来言われている論理的推論には時間がかかることへの説明となる．

相互想起過程と自己想起過程の切り替えはモデルの中では単純なパラメータで実現でき，その動的スイッチングが一見して複雑な確率的処理と論理的処理の混合した過程を作り出す．以下，直観的推論と論理的推論の実現方法について説明する．

4.4 相互想起モデルによる直観的推論

4.4.1 直観的推論の計算方法

本稿で考える直観的推論は，現在状態から次に起こりうる記憶パターンの予測を連想的かつ並列に行うことで，次の時刻で起こりうる状態のみの評価を実現する．その予測のために，相互想起型の連想記憶モデルを用いる．

これを推論システムとして実装するために，記憶しておく状態ベクトルの要素は±1(興奮性/抑制性)からなるランダムベクトルとし，各々の状態ベクトルの次元は相互の直交性を確保するために十分に長くした(ベクトルの次元 N は本研究では 5,000 とした)(式(4.9))．なお，この状態ベクトルは式(4.1)と同様に縦ベクトルである．ここで p，q は記憶パターンの番号で，記憶パターン数は Q とする．結果として本研究では記憶パターン群は相互にほぼ直交するという条件を設けたことになる(式(4.10))．

$$\boldsymbol{x}^p = \left\{ x_i^p \in \{1, -1\} : i = 1 \dots N, p = 1 \dots Q \right\} \tag{4.9}$$

$$\frac{1}{N} x^{q\mathrm{T}} x^p \begin{cases} = 1 : p = q \\ \fallingdotseq 0 : p \neq q \end{cases} \tag{4.10}$$

相互想起による直観的推論は，以下の四つの機能要素からなるとする．これらの要素群は結果として，①図 4-8 の価値認識層の計算が連想計算層で得られた部分的な特徴に対して短いサイクルで価値を付与する，②外部から入力された状態ベクトルに対して連想計算層によって情報循環を積極的に制御する，という二つの機能を果たす．

(1) 現在状態 : 感覚入力や以前の状態から予測された状態認識を表現する各感覚入力からの特徴の集合である．特定の概念を想起しているときは，この状態が特定の記憶パターンと等しくなるが，直観的推論のための予測の場合には，この状態は複数の記憶パターンの線形和となる．

(2) 連想ネットワーク : 状態 S_t＋行為 A_t→状態 S_{t+1} という状態遷移を予測する機能を実現し，結果として世界についての知識を表わす．連想記憶モデルにおける連想ネットワークは，保持するべき記憶パターン群の相互の状態遷移とその確率を式(4.11)により連想行列W^eに埋め込む．その後，現在状態ベクトルx^r から次の状態ベクトルx^cへの想起は，式(4.10)の疑似直交条

件があるため，連想行列と現在状態ベクトルとの積を求めることで関連する記憶パターンの予測と合成の過程も含めて一度に計算できる．結果として式(4.12)で得られる状態ベクトルx^cは複数の記憶パターンに条件付き確率で重みづけした混合パターンとなる．

$$W^e = \sum_q \sum_p Pr(q|p) x^q \frac{1}{N} x^{p\mathrm{T}} \tag{4.11}$$

$$x^c = W^e x^r = \sum_q Pr(q|r) x^q \frac{1}{N} x^{p\mathrm{T}} x^r + \sum_q \sum_{p \neq r} Pr(q|p) x^q \frac{1}{N} x^{p\mathrm{T}} x^r$$

$$\fallingdotseq \sum_q Pr(q|r) x^q \tag{4.12}$$

(3) 価値評価系 : 式(4.12)で予測された状態ベクトルx^cの価値を評価する．本研究ではエージェントはx^cに含まれる記憶パターンそれぞれに対応した価値$\{R^q : q = 1 \ldots Q\}$を抽出し，その獲得または回避のための行動を決定する．

x^cからの価値抽出のため，価値連想行列W^{cR}を用意する(式(4.13))．W^{cR}とx^cの積よりx^cに含まれる記憶パターンx^qのそれぞれに対応した価値 RC を一度に計算できる(式(4.14))．意思決定は，この価値ベクトルのうち価値が最大となる要素 RI に注目し，その要素を取得することで行われる(式(4.15))．

$$W^{cR} = \frac{1}{N} \begin{bmatrix} R^1 x^{1T} \\ \vdots \\ R^q x^{qT} \\ \vdots \\ R^Q x^{QT} \end{bmatrix} \tag{4.13}$$

$$RC = \ W^{cR} x^c = \frac{1}{N} \begin{bmatrix} R^1 x^{1T} \\ \vdots \\ R^q x^{qT} \\ \vdots \\ R^Q x^{QT} \end{bmatrix} \ x^c \fallingdotseq \frac{1}{N} \begin{bmatrix} R^1 x^{1T} \\ \vdots \\ R^q x^{qT} \\ \vdots \\ R^Q x^{QT} \end{bmatrix} \ \sum_q Pr(q|r) x^q$$

$$\fallingdotseq \begin{bmatrix} R^1 Pr(1|r) \\ \vdots \\ R^q Pr(q|r) \\ \vdots \\ R^Q Pr(Q|r) \end{bmatrix} \tag{4.14}$$

$$RI = \underset{q}{\operatorname{argmax}} W^{cR} x^c = \underset{q}{\operatorname{argmax}} RC \tag{4.15}$$

(4) 脳内情報循環系：本モデルでは感覚入力－連想ネットワーク－価値評価系の間の情報循環をトップダウン的に制御するシステムを考える．循環のゲインを制御することで価値探索の機能を実現する．更に深い範囲への探索を行なう際には脳内情報循環系により式(4.12)を反復することで，より深い連想と探索を極めて単純な計算，かつ短時間で実現する．

なお，本研究で用いる状態ベクトルは探索する Tree の深い範囲での探索時にその探索の経路情報を状態ベクトル内に記銘していないため，深い探索で価値のある状態を発見してもそこに至る経路を想起・利用できない問題があった．そのため，価値ベクトルとは別に行動ベクトル$A = \{A^k : \ k = 1 \ldots M\}$を用意した．なお，行動ベクトル$y = \{y^k : \ k = 1 \ldots M\}$は式(4.14)同様に状態ベクトル x^c 中に含まれる行動ベクトルのみが活性化しており，その情報を計算に用いることで想起可能とする．まず探索のスタートの一層目の計算時に見出された行動ベクトル *A* を保持する．ついで探索が進み，結果として価値のある状態を見出した際には，保持しておいた行動ベクトル *A* のうち価値が最大となる行動要素を実行する．その計算は，式(4.14)の価値ベクトル RC^qを個々の要素が行動に対応する行動ベクトルA^Mに変更した式(4.16)を適用し，その行動確率が最大となる要素 *AI* を選択することで実現できる (式(4.17)) ．

$$AC = W^{cA} x^c = \frac{1}{N} \begin{bmatrix} A^1 y^{1^T} \\ \vdots \\ A^k y^{q^T} \\ \vdots \\ A^M y^{M^T} \end{bmatrix} x^c \fallingdotseq \begin{bmatrix} A^1 Pr(1|r) \\ \vdots \\ A^k Pr(k|r) \\ \vdots \\ A^M Pr(M|r) \end{bmatrix} \tag{4.16}$$

$$AI = \underset{k}{\operatorname{argmax}}\, W^{cA}x^{c} = \underset{k}{\operatorname{argmax}}\, AC \tag{4.17}$$

この行動情報の保持手法は脳科学の知見はないが，推論により行動を起こす際には何らかの形で経路情報を保持するシステムが必要である．さらに直観的推論においては，処理を高速にするためにより単純な方法で情報を保持しているであろうと考え，この手法とした．あえて言えば作業記憶に近いが，本稿ではその実装については考えない．

次にこの計算モデルを検証するシミュレーションを行った．まず，図 4-7 の Tree のノードのエピソードを表現する状態ベクトル群(S_i : i = 0…14)を事前に用意し，それに基づいて相互想起行列および自己想起行列を作った．ここで，エピソード間の遷移の条件付き確率は適当に決めた．その上で，現在状態 S_0 から過去の経験に基づいた状態ベクトル毎の状態遷移の強度を状態ベクトルx^cと記憶されているすべてのベクトルとの間の相関から計算した(図 4-9).

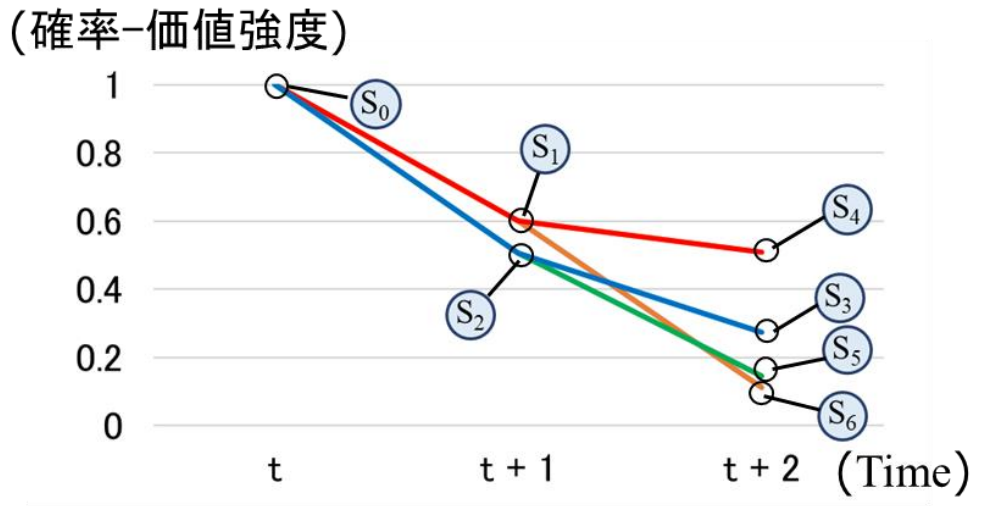

図 4-9　直観的推論の処理結果の例

図 4-9 では現在状態 S_0 から一層先(t+1)の探索時に取得された状態ベクトルx^c (=$W^e x^0$)と価値ベクトルW^{cR}との積 *RI* に意思決定するための価値が割り振られておらず，価値の抽出ができなかったことを想定した．現在状態 S_0 から一層目(t+1)を連想したが行動に至る価値を抽出できない場合，さらにもう一層先(t+2)の状態を計算する必要がある．その際は現在状態 S_0 の一層目(t+1)の状態ベクトルx^c (=$W^e x^0$)に対して再度式(4.12)を適用し，t+2 の状態ベクトル$W^e(W^e x^0)$を算出する．そして，これを意思決定することのできる高い価値を抽出するまで繰り返す．これにより確率的かつ並列

的な探索ができ，シミュレーションにあった任意の閾値を超えた際に行動選択が可能となる結果を得た．ただし，価値のある状態を見出せないままに連想を反復すると，個々の記憶パターンとの間の相関により算出される相関値(想起強度)が極めて小さくなるため，反復回数を多くすることは困難である．これが，直観的推論での推論が浅くなることの説明となる．

4.5 自己想起モデルによる論理的推論

4.5.1 論理的推論の計算方法

論理的推論の実装は直観的推論の計算過程で求められた価値に注目し，現在状態ベクトルと価値ベクトルとの和を取ることにより，現在状態に価値を含んだ現在状態価値ベクトルを作成する．そしてこの現在状態価値ベクトルと自己想起行列とをかけることで，次の時刻の現在状態ベクトルを更新する．この計算を，現在状態ベクトルが焦点化した価値に対応する状態ベクトルに収束するまで繰り返す．以下はその詳細である．

(1) 価値の焦点化 : 直観的推論により予測された現在状態ベクトルx^cに価値が割り当てられた状態ベクトルが含まれているときには，そのうち式(4.14)で検出された価値が閾値以上で最大の記憶ベクトル要素 RC^{RI} を選択する．

(2) 自己想起行列 ： 現在状態ベクトルx^cの想起強度を強化するため，式(4.11)の p，q を p=q とした，自己相関行列 W^s を用意する(式(4.18)).

$$W^s = \frac{1}{N}\sum_p x^p x^{p\mathrm{T}} \tag{4.18}$$

(3) 状態ベクトルの変調：まず，時刻 t=0 において，現在状態ベクトルx^cを意識にのぼったベクトルとしてx_0^{tmp}に代入する(式(4.19)). 次いで，x_0^{tmp}内の価値要素 RC^{RI} に対応する記憶ベクトル成分を，RC^{RI} に比例して強化する(式 (4.20)). 状態空間x_0^{tmp}中に複数の価値が競合している場合は，価値の間で相互抑制をかけることにより，最も価値の高い要素への収束を促す(式(4.21)) ． この計算と正規化(式(4.22))の反復により，現在状態ベクトルx^cはその成分の中で価値の最も高い状態ベクトルに収束していく．この計算サイクルを 1 度することを時刻 t=1 とし，計算サイクルを増やすごとに t の値が 1 ずつ増えていく．

$$x_0^{tmp} = x^c \tag{4.19}$$

$$x_{t+1}^{tmp} = W^s\left(x_t^{tmp} + RC^{RI}x^{RI}\right) \tag{4.20}$$

$$x_{t+1}^{tmp} = x_t^{tmp} - \beta(ONE - I)x_t^{tmp} \tag{4.21}$$

$$x_{t+1}^{c} = \frac{x_{t+1}^{tmp}}{\left\| x_{t+1}^{tmp} \right\|} \tag{4.22}$$

ここで, ONE は要素がすべて 1 の行列であり, βはその更新する割合を示す値である.

以上の処理により, 推論の各段階で現在状態ベクトルは特定の状態ベクトル成分に収束し, 次の直観的推論の出発点となる. 本研究ではその収束状態が意識化されたシンボル状態であるとする.

論理的推論を実施するにあたり上記の(1)～(3)を繰り返し実施することで, 通常は想起強度がほぼ 1 に収束することが自己想起の特性上知られている. しかし以下の 2 つの条件のうち, どちらかを満たすと論理的推論の想起強度が 1 に収束しないことがある. ①想起された混合状態ベクトル内に複数の記憶情報があり, 想起強度をとった際にその差がほぼない. ②記憶情報に対応する価値情報に差がない. これらの内, どちらかが満たされると各ベクトル同士が釣り合ってしまい想起強度が収束しないことがある. これを回避するために, 本研究では相互抑制を用いた. 相互抑制とは, 自身の影響度の一部を他者に渡すことにより他者の影響度を抑制することを, 見出されている要素すべてに実施する機能である(図 4-10). これを関係するすべての状態空間に適応することにより, 想起強度が元から大きかった状態ベクトルは相手からの抑制があまりされず強化され, 想起強度が小さい状態ベクトルは自身より大きい状態ベクトルから抑制を受けるため想起強度が小さくなる.

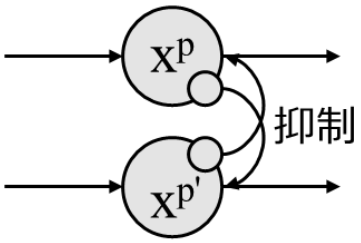

図 4-10　相互抑制の処理イメージ

さらに相互抑制を用いる際, 自分以外の状態に対して影響を及ぼす自身の力の割合について検討する必要がある.

そのため相互抑制の効果検証シミュレーションでは, 二分探索木を一層分だけ用意し, 状態空間および見いだされる価値の差がほぼ 0 となる状態を作った. この条件のもと,

根元のノードとなる位置からの推論結果が図 4-11 の赤色と緑色で表現されたグラフである．相互抑制を取り入れない場合はこのように見いだされる価値に差がないと想起強度が釣り合ってしまい，想起強度が 1 に収束しない．そして相互抑制する際の抑制度合 a を 0.3，0.5，0.7 とした結果をグラフの上下により示す．グラフを見るとわかるように抑制度合 a を大きくすることで，収束するまでの計算サイクル数が短くなる．なお相互抑制の効果検証シミュレーションでは，相互抑制の効果をより分かりやすくするために 1 ステップごとに想起強度の影響度 a を変更しながら抑制度合を変更するのではなく，相互抑制の計算は論理的推論の想起強度が 1 に収束しないと判断した 20 ステップ目の時点で一度だけ計算する形としている．

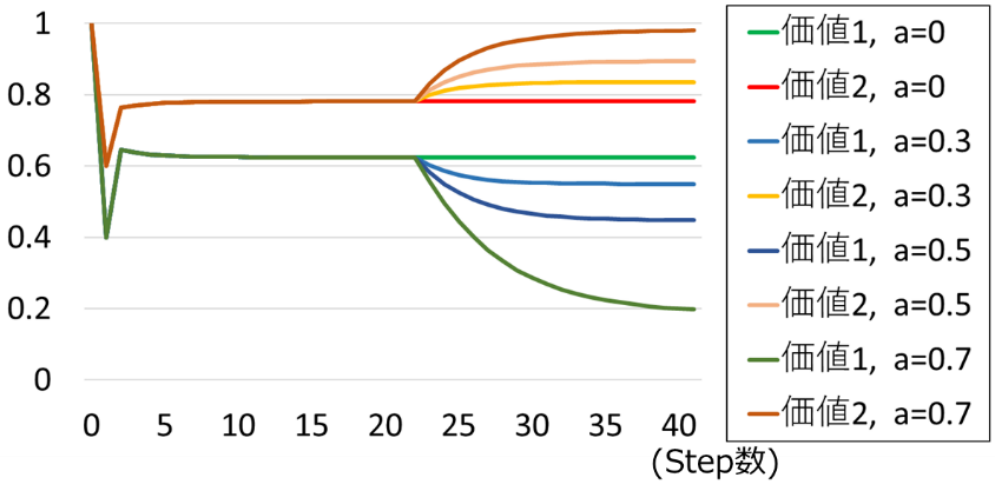

図 4-11　相互抑制の抑制強度による影響

本学位論文にて提案した論理的推論の基本動作を確認するため，図 4-7 の Tree 構造を対象としたシミュレーションを行った．ここでは最初(S_0からS_1，S_2の探索)の推論の際に，一層目では状態S_1，S_2に対応する価値が見つからないため，過去の経験より算出された事前確率に基づいた相互抑制により状態ベクトルがS_1に収束する．さらに，その次のS_1からの探索で二層目の状態S_3，S_4を連想で想起する．本シミュレーションでは，そのうちS_3では価値が見いだされないが，S_4では価値が見出されると想定した(図 4-12).

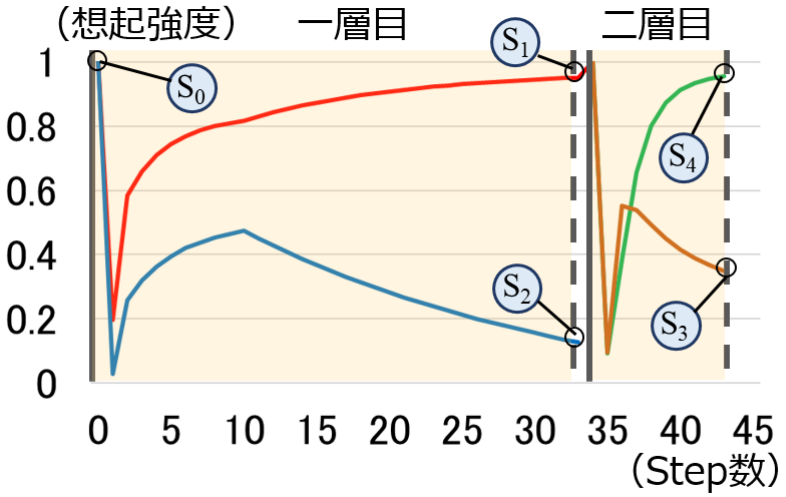

図 4-12　論理的推論のシミュレーション結果

結果は一層目では直観的推論と同様の行動選択が実現されるが，それを見出すまでの時間が論理的推論のほうが長いことが示された(図 4-12 の Step10～33 の区間)．次いで二層目では，直観的推論のみ，すなわち過去の経験からはS_3に対応する価値に影響された状態ベクトルの方が強い想起強度となる．しかし，実際にはS_4に対応する価値に影響される状態ベクトルの方がその瞬間の状況では大きい価値を持っていた．そして時間を経るごとに状態ベクトル中のS_4に対応する価値に影響された状態ベクトルの強度が強くなっていき，最終的にはS_4の想起強度がほぼ 1 に収束した．それに対し，S_3の想起強度は 0.4 以下と弱くなった(図 4-12 Step36～42)．なお，この自己想起による論理的推論では，想起強度は完全には 1 に収束しない．それ原因として状態ベクトルx_0^{tmp}は直観的推論により算出されるがその際，完全に直交しているわけではなく，ほぼ直交しているという条件が入っていた(式(4.10))．そのため，この状態ベクトルx_0^{tmp}には状態ベクトルx_0^{tmp}以外の要素も直交しきれなかった値の分だけ含まれるためである．なお，本学位論文内のシミュレーションすべてにおいて，特に断りがない場合は相関強度が 0.95 を超えた際には相関強度が 1 に収束したとしている．この相関強度が 1 に近づいた状態ベクトルを意思決定に，または次の推論の初期ベクトルとすることで，論理的推論と同等の時間はかかるが価値に駆動された推論が実現されることが示された．

4.5.2 記号的推論の現れ

さらに, 推論過程を横軸に想起強度, 縦軸に時刻を取りグラフ化したものが図 4-13(左)である. このグラフより自己想起を用いた論理的推論を 1 に収束したものから再度推論を開始するという連続的な推論の解釈は, 時間変化に伴いはじめに示した過去の経験のみに影響される直観的推論の直後に，その後に得られる価値を評価・予測してその最大化を行う論理的推論の過程が続き，その反復による順次的な推論の振る舞いとなった. またこの結果を探索の元となった Tree 状に示したものが図 4-13(右)である. この結果から，Tree 探索の深さ優先探索と同等の推論が行なわれていることが示された.

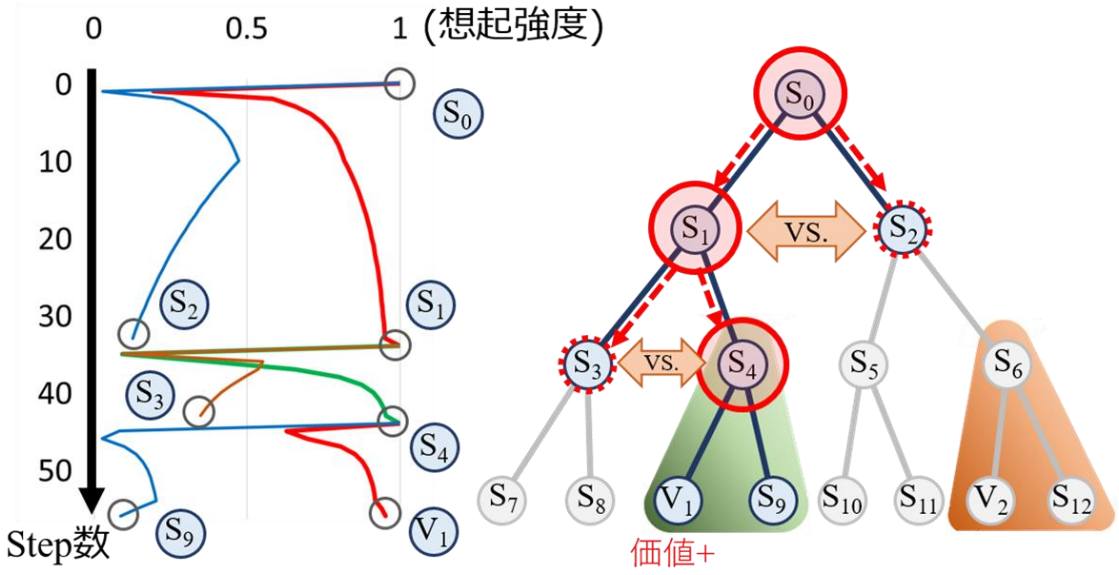

図 4-13　S_0 からの連続的な論理的推論の解釈

4.6 直観的推論および論理的推論の統合

4.6.1 推論システムを統合するメカニズム

4.4 節，4.5 節では，連想記憶の相互想起の機能を用いた直観的推論で言われている確率的かつ計算回数の少ない推論の定式化，およびシミュレーション，および自己想起の機能を用いた意識的かつ直観的推論と比較して推論に時間のかかる論理的推論の定式化，およびシミュレーションをしてきた．本項では人の推論システムは二つの別々のシステムとして存在するわけではなく，一つの推論システムの動作モードの切り替えることができると考えており，その概念的なメカニズムが図 4-14 である．

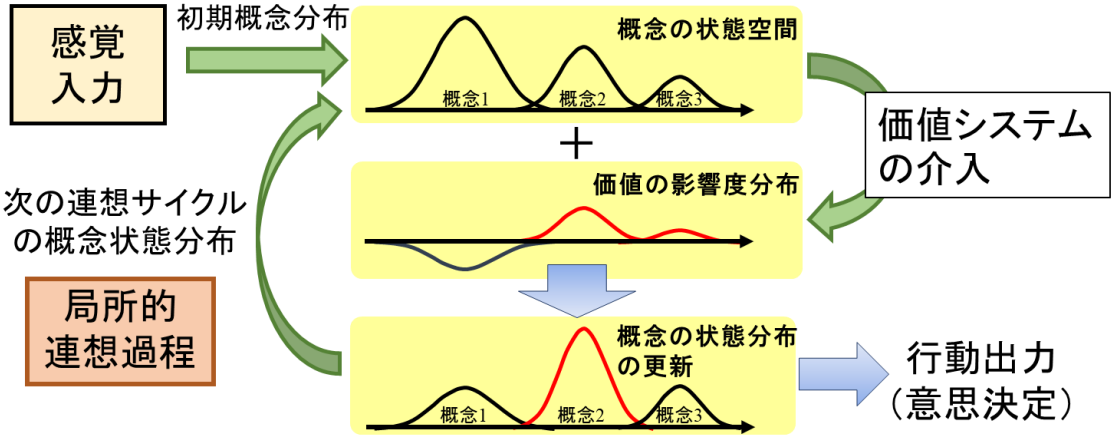

図 4-14　直観的推論と論理的推論の統合メカニズム

本項では，エージェントに搭載されているセンサからの情報を感覚入力として入力されることを想定する．この入力された感覚入力を基にニューラルネット(例えば Deep Learning)などを用いて特徴量を抽出する．このニューラルネットなどにより抽出される特徴量は入力層に近い部分では，様々な情報に反応する必要があるニューロンが多いため，抽象度が低いものとなっていることが知られている．しかし処理をする層が上がるにつれ，入力情報中に含まれる特異的な情報に限定されて発火することも知られている．そのため抽象度が向上しその情報に含まれる概念レベルの情報を抽出することができると考える．これを推論する際の初期の概念状態分布とする．そして，この概念状態分布に対して，価値システムが現在の内部状態などを考慮して介入することで，見出されている概念状態分布に対応する価値の影響度分布を作成する．この概念分布と価値の影

響度分布を足し合わせることで，概念状態分布を更新する．直観的推論では，この更新された状態分布に対して相互想起のネットワークにかけ，さらに次の時刻の概念状態分布を作成する．そしてこの次の時刻の概念の状態分布に対して再度価値分布の付与を行う，といったサイクルを少ない回数分繰り返し，行動決定(意思決定)することができる状態になったと判断した際に，行動出力することで実現できることを想定している．そして論理的推論では，概念分布に価値の影響度分布を足し合わせた更新後の概念状態分布の結果中に含まれる価値状態に意識をし，その状態に対して自己想起のネットワークにかけることで次の時刻の自身の概念状態分布を作成する．そして作成した概念状態分布を評価し，一つの概念に収束していなければ再度この価値の影響度分布を足し合わせ評価するというサイクルを繰り返す．そして一つの概念に収束した際には意思決定し，行動を出力する．

このように直観的推論と論理的推論は想起するネットワークは異なっていたとしても，計算方法はほぼ同じであり，かつその予測精度と意識・無意識の違いの処理として説明することができると考えられる．

4.6.2 連想記憶を用いた推論システムの統合

本稿でいう論理的推論は直観的推論に自己想起処理の計算過程を加えたものであり，そのスイッチングにより，二つの推論過程の統合は可能である．そこで 2 つの推論過程の切り替えの方法として，次式に示す連想過程の統合式を採用した．

$$x_{t+1} = \alpha \left(\sum_q Pr\left(x_{t+1}^q \middle| x_t^p\right) W_{pq}^e\right) x_t^c + (1-\alpha) \sum_q W_{pq}^S x_t^c \tag{4.23}$$

この式の右辺第一項は，直観的推論を実施する相互想起による連想項である．この処理はスイッチングのためのパラメータ α を 1 にすることで直観的推論を単独で実施できる．ここでは過去のエピソードを表す記憶ベクトルからイベントごとの連想行列W^eを作成し，入力ベクトルx_t^rから想起されるエピソードを連想的に探索する．この項から想起されるベクトルは，入力ベクトルとの繋がっている状態の内，条件付き確率が高い状態ほど強い強度で想起される．すなわち想起された状態ベクトルは過去の経験(連想行列の蓄積)に基づきその強さが決まる．この連想を反復することで，短時間かつ広い範囲に対して，確率に基づいた並列的な探索が可能となる．そして探索中に価値のある記憶パターンの要素を発見したときは，その価値の大きさを評価して意思決定する．

式(4.23)右辺の第二項は論理的推論の項である．論理的推論を単独で実施する際には，α の値を 0 とする．ここでは第一項で見出された価値に焦点を当て，その価値を最大化させる記憶想起の反復計算を行う．すなわち，入力ベクトルに含まれる価値のある成分を強化して，想起によって特定の価値が支配的になるまで自己想起型の連想計算を反復する．そして，収束した状態に対応した行動選択をする．

4.6.3 パラメータ α を整数(0,1)とした場合の振る舞い

パラメータ α を整数とした動作確認シミュレーションでは，提案したモデルの内，直観的推論と論理的推論とを統合した式(4.23)による推論の特性を確認する．なお，パラメータ α を整数とした動作確認シミュレーションでは，エージェントの直前の位置はワーキングメモリに記憶されていることを想定し，その行動には inhibition of return がかかることで，元の場所には戻らないとした．すなわち，予測された状態ベクトルに直前にいた場所情報が含まれていたとしてもそれは選択肢から除外した．

パラメータ α を整数とした動作確認シミュレーションでは式(4.23)のパラメータ α は，直観的推論を必ず実行させるために初期値を α=1 とした．論理的推論への切り替えは，直観的推論により得た現在状態ベクトル x^c に対して式(4.14)を適用し，複数の価値が同時に見出されて競合した際に α=0 とし，論理的推論による長期的な価値予測に基づく推論をするようにした．

各推論をする上での条件は以下のようにした．直観的推論は探索領域を広げると見いだされる推論結果の想起強度は急激に小さくなる．そのため，深い探索をすると見出すべき状態ベクトルの想起強度が小さくなりすぎるため，他の状態ベクトルとの違いが見いだせなくなる．そのためパラメータ α を整数とした動作確認シミュレーションでは，直観的推論における探索範囲は 3 層を上限とした．そして，論理的推論では状態ベクトルは自己想起を繰り返すことでほぼ 1 には収束するが完全に 1 に収束しないことがある．そのため，本研究では状態ベクトルの想起強度を取った際，想起強度が 0.95 を上回った際に論理的推論による状態ベクトルの収束が終わっているとした．

本迷路課題の遂行によりエージェントは，地図上の位置に対応する入力ベクトルを受け取り，その中に見出された価値の関係により直観的推論と論理的推論を切り替えることにより，それぞれの状態に応じた柔軟な行動が表出すると想定した．さらに論理的推論の実行には複数の価値の競合が必要となる．そのため，地図中には複数の報酬源(報酬 A は大きな報酬(例えば 0.9)，報酬 B は小さな報酬(例えば 0.6))を配置した．なお，それぞれの報酬の多寡は推論結果に影響は及ぼさない一方で，状態遷移の事前確率 Pr(q|p)（式(4.23)を参照）と価値との積が閾値を超えないと競合と認識されないため，その点においては注意が必要である．

地図上の各位置に割り振られた価値の値については以下のようにした．まず地図世界そのものは既知とした．しかし報酬源は比較的最近に存在を知ったため，地図のすべて

の場所には価値が割り振られておらず，報酬源の近くの限られた領域のみに，報酬源からの距離に応じた割引率(本稿では 1 ステップごとに 0.9 倍)で減衰しながら割り振られているとした(図 4-15).

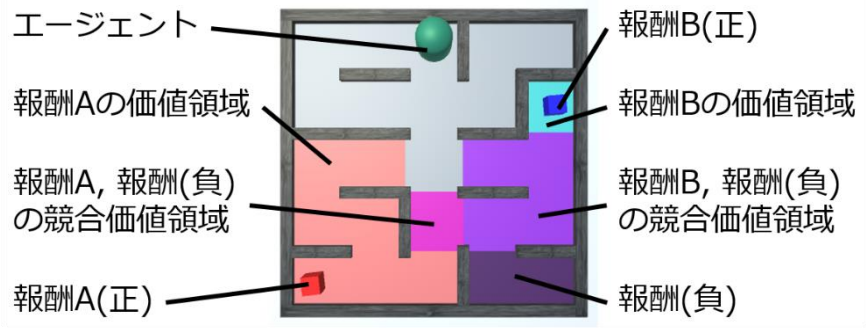

図 4-15　迷路探索用の地図，および価値配置

また，連想行列W^eに含まれる条件付き確率 Pr(q|p) （式(4.23)を参照）は，原則として過去に経験した地図上の移動回数に比例すると考えられる．ここでは，エージェントが事前にランダムウォークによって条件付確率を獲得すると想定しているため基本的にはほぼ等確率になるが，競合を想定している場所(図 4-15 の地図上の中央においては，報酬 A と B の比率が 6:4 になるようにした)，および報酬に向かう方向に多少の偏りを与えた．これは，直観的推論の事前確率に従う推論結果と論理的推論の推論結果が異なるようにするためである．図 4-15 の中央の位置からは，価値は上下からはほぼ 0，左右からは報酬 A および報酬 B からの価値を見出すことができるが，その関係は報酬 A の方が大きくなるようにした．状態ベクトルの次元は N=5,000 の±1(興奮性/抑制性)からなるランダムベクトルとし，地図上の各位置に対応させた．

この環境中においてエージェントは，地図中の位置に対応する入力ベクトルの情報を基に直観的推論と論理的推論とを柔軟に切り替えることで，推論行動を変化させることによる地図世界のナビゲーション課題の解決を試みた．エージェントの目的は，地図中の任意の位置から試行錯誤なしで報酬の位置にたどり着くことである．エージェントは，地図上の現在位置から行動を開始し，過去に報酬を得た経験回数は多いが報酬量の小さい青のキューブ，または過去に経験回数は少ないが報酬量の大きい赤のキューブを得ることである．シミュレーションはゲームエンジン Unity を用いて仮想環境を築き，その中にエージェントを置いて，行動探索を行なった [33].

シミュレーションの結果，エージェントはスタート位置（緑色のカプセルの配置され

ている位置）から推論を開始し，図 4-16 左の水色の経路をたどり赤色の報酬量の大きいキューブを得た．さらにエージェントの推論結果をグラフ化したものが図 4-16 右である．エージェントはシミュレーションを開始したスタート位置における推論行動として，スタート位置から直観的推論を用いて探索する．その結果 1 層目および 2 層目では，意思決定に使用できるだけの価値を見出すことができなかった．しかし 3 層目の推論では青のキューブに対応する価値を見出すことができたため，下方向に進むことで報酬が得られることを予測し，エージェントは下方向へ移動し，十字路になっている位置へ移動した．十字路の位置における推論は，スタート位置の推論同様に直観的推論を用いた 1 層目の探索では意思決定するための価値を見出すことができなかったが，2 層目の探索において 1 ステップ目と同様に下方向へ向かうことの価値を見出すことができたため，下方向へ行動し迷路上の中央まで移動した．迷路中の中央からの推論では，エージェントの現在位置から直観的推論を 1 層分だけ実行した際に，現在位置から左方向に赤のキューブに対応する価値を見出し，さらに現在位置から右方向に青のキューブに対応する価値を見出す．そのため，論理的推論の実行条件である価値の競合が起こる．そのためエージェントは式(4.23)の統合パラメータ α の値を 1 から 0 に切り替えることで推論行動を論理的推論へと切り替えた．論理的推論を開始してから 2 ステップ目までは青色のキューブに対する価値の方が強い推論強度となっていたが，3 ステップ目以降では赤色のキューブに対応する価値の方が大きくなった．そして，論理的推論を初めて約 30 ステップ目に左方向を示す状態ベクトルの想起強度が 0.95 を上回ったため論理的推論が完了したと判断し，左方向へと行動をした．その次のピンク色の位置に入った位置における推論では，エージェントは推論開始時に直観的推論を実施するため統合パラメータ α を 1 に変更した．この位置では左方向にしか価値を見出すことができないため，エージェントは直観的推論により見出された左方向へと行動する．その次の位置におけるエージェントの推論は，自身の右の位置，および下の位置において価値を見出すことができるが，先述したように inhibition of return により自身の直前にいた位置の状態ベクトルは無視されるため，エージェントの推論結果中に含まれる状態ベクトルは下方向に関係するもののみに絞られる．そのため，エージェントは下の方向へと移動した．これ以降の行動は inhibition of return を用いたことにより，直前にいた場所の状態が見いだされることがなく，報酬を得る方向へと直観的推論を用いるのみで進むことができた．

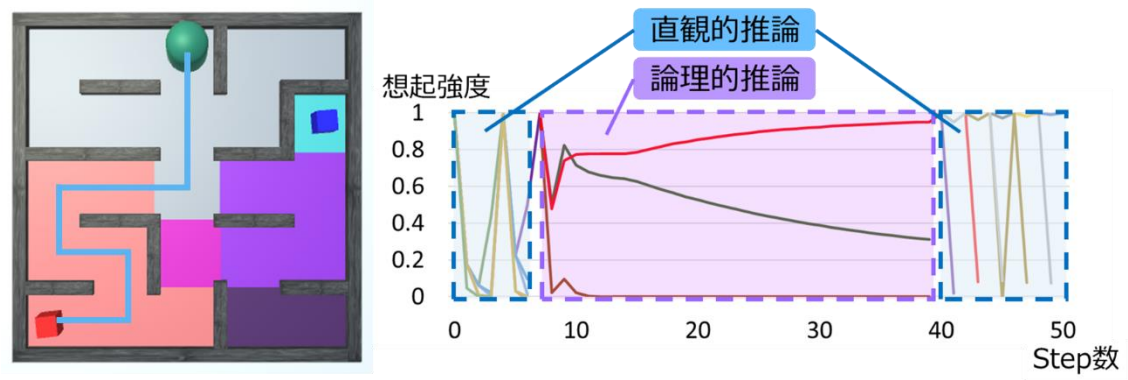

図 4-16　エージェントの通った経路，および推論結果

パラメータαを整数とした動作確認シミュレーションより，入力ベクトルから直観的推論により推論し，その結果得られる次の時刻における合成ベクトル中に含まれる状態ベクトルに対応する価値が競合するかどうかの結果を見て直観的推論，および論理的推論を柔軟に切り替えることができる結果を得た．この結果は従来言われてきた人の推論特性の一つである，推論システムを柔軟に切り替える二重過程システムに対応すると考える．

4.6.4 パラメータ α を 0 から 1 までの間の実数とした場合の振る舞い

本項では，提案した推論システムの特性を知るために提案した推論モデルを統合した式(4.23)の内，パラメータ α を 0 から 1 までの間の実数にした際に推論結果に与える影響について検証する．本研究ではその特性上，直観的推論と論理的推論を同時に処理する場合，次の時刻についてのみを対象とするのではなく，さらに先の時刻の状態を連想的に想起することを予想した．

パラメータ α を実数とした動作確認シミュレーションでは図 4-17 の形をした 3 層からなる二分探索木とした．図中の各ノードは結果のグラフの色と対応付けするために着色しているが，その色には特別な意味はない．各ノード間の経験数は枝の部分に添え字として書いてあるものとした．さらに各ノードにより見出される価値は，ノードの上に赤字で書かれているものとした．パラメータ α を実数とした動作確認シミュレーションは S9 にて大きな価値が見出すことができるとし，そこに向かって経験数，および価値が見出されるように各パラメータを手動で設定した．

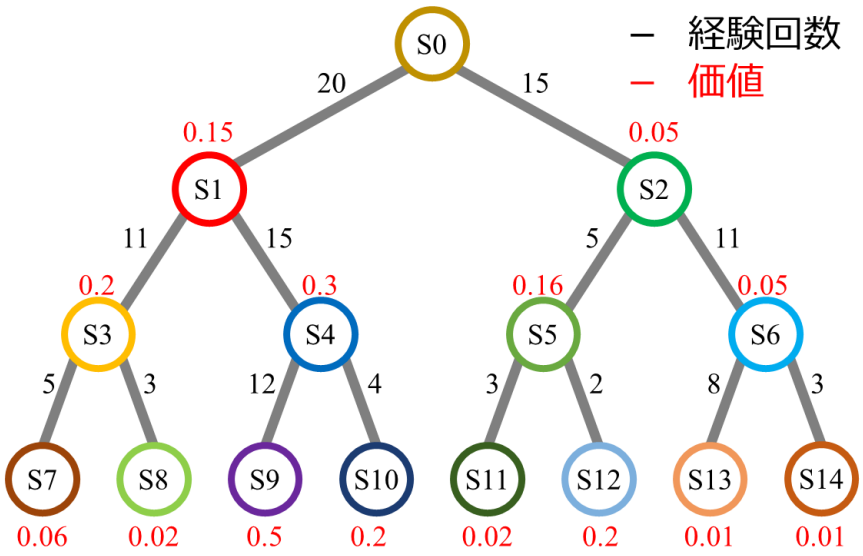

図 4-17　シミュレーションに用いた Tree 構造，およびパラメータ

なお直観的推論により想起された混合状態ベクトル中に含まれる複数の状態ベクトルの判断は，記憶されている状態ベクトルすべてとの間で相関を取り，その想起強度が 0.1 を超えた場合に想起されていると判断し，価値ベクトルを付与した．状態ベクトルの次元は N=3,000 の±1(興奮性/抑制性)からなるランダムベクトルとし，Tree 上の各位

置に対応させた.

しかしαの値を 0 から 1 までの間の実数としてシミュレーションする際，直観的推論と論理的推論との間には, 処理が終わるまでの時間的なスケールが異なるという推論の特性の違いがある．前述したように，直観的推論では次の事象を 1 回の計算で推論することで結果を見出すが, 論理的推論ではその何倍もの時間をかけて状態ベクトル中に含まれる情報を 1 つに収束させる. つまり直観的推論と論理的推論の処理時間には差が生じる．そしてこの二つの推論が同じ事象について処理をしていると捉えられるのは，論理的推論は数サイクル分が限界であると考えた. そのためパラメータαを実数とした動作確認シミュレーションでは, 論理的推論の処理をこれまでに示した状態ベクトルが 1 に収束するまでではなく，ここでは 5step 分のみとした.

さらに，直観的推論と論理的推論の処理に影響力はパラメータαの値に依存する．パラメータαを実数とした動作確認シミュレーションでは二つの推論の内，先の時刻を次々に予測しようとする直観的推論よりも次の時刻のみに着目し，確実に時刻 t の情報を強化する論理的推論とでは, 推論が早く進むことよりもプロセスを考慮し着実に推論していく論理的推論の方が優先されると考えた．そのため，論理的推論の影響度の大きくするためにαの値は 0.1 とした.

シミュレーションの手順はまず，現在状態であるノード S0 から直観的推論により弁財状態ベクトルである$\boldsymbol{x^t}$を基に推論を開始する．ここでは，従来の直観的推論と同様に次の時刻(t+1)の状態ベクトル (ここでは S1 および S2 の混合ベクトル$\boldsymbol{x^{t+1}}$)を想起する(式(4.24)).

$$x^{t+1} = 0.1(\sum_j W_{ij}^S x^t) + 0.9\left(\sum_j Pr\left(x_j^{t+1}\middle|x_i^t\right)W_{ij}^e\right)x^t \tag{4.24}$$

その後，式(4.24)により見出された次の時刻(t+1)の状態ベクトル$\boldsymbol{x^{t+1}}$を対象に，統合パラメータαを実数(本稿では 0.1)にして論理的推論を 5step 分実施し，その結果を 0.9 倍した．その後，式(4.24)により見出されていた次の時刻(t+1)の状態ベクトル$\boldsymbol{x^{t+1}}$の結果を 0.1 倍にし，論理的推論結果と足し合わせることで次の時刻(t+1)の状態ベクトルx^{t+1}とした．その結果が図 4-18 の 1 サイクル目の結果である.

その後は 1 サイクル目の結果として得られた状態ベクトル$\boldsymbol{x^{t+1}}$に対して, 再度式(4.24)の右辺第一項の直観的推論の処理を適用し，さらに次の時刻(t+2)予測を試みる．そして

その結果に対して変数αの値である 0.1 倍にする．同様に論理的推論においても状態ベクトルx^{t+1}に対して，現在状態が続いたらどちらの価値が支配的になるかを再度式(4.24)の右辺第二項の論理的推論の処理を 5step 分適用することで算出し，その結果を 0.9 倍にする．これらを足し合わせることで，さらに 2 サイクル目の結果とする．

この処理を繰り返し，4 サイクル目までは，現在状態 S0 から隣接する S1，S2 の想起強度が強く，かつこのまま継続すると統合パラメータαを整数とした際の論理的推論のように 1 に収束するかのような推論結果を得た．しかし 5 サイクル目の結果として，急に S1 に接続されている S4 の想起強度が最大となる結果を得た．この 4 サイクル目の結果を解釈する上で確認すべき点は，先に述べた 4 サイクル目まで S1 の想起強度が徐々に強くなることではなく，統合変数αを 0.1 として加えられていた直観的推論の計算より徐々に強化されている S4 である (1〜4 サイクル目の青色の線)．S4 ははじめ，S1 を含んだ状態ベクトルに対して直観的推論をすることにより想起されるが，その想起強度は価値計算をするほど強くなかった．しかし，サイクルを繰り返し 4 サイクル目の計算の終了時，S4 の状態ベクトルは状態ベクトル中に価値計算の対象であると判断する閾値の想起強度 0.1 を超えた．そのため，5 サイクル目の論理的推論における価値計算を含んだ状態ベクトルの強化は 4 サイクル目までの S1 と S2 だけでなく，S4 に対しても同様に行われた．さらに，価値は基本的に先の状態の方が値を大きくしていることから S4 の想起強度が急激に増加したと解釈することができる．さらに，その次の直観的推論では，これまでの S1 および S2 に関連する状態ベクトル(S3〜S6)だけでなく，S4 に関連する状態ベクトル(S9，S10)についても想起の対象となることになる．そのためサイクルを続け，6 サイクル目の終了時には 4 サイクル目の終了時と同様に，今度は S9 に対応する状態ベクトルが価値計算対象となる閾値を超えた．その後は S4 の時と同様に S9 に対応する状態ベクトルの想起が強化され，想起強度が高くなっていく結果を得た．

上記のように本シミュレーションではパラメータを実数としてサイクルを繰り返すことで，想起される状態ベクトルが徐々に価値の大きい状態に遷移する結果を得た．それでは，この結果はどのような解釈になるであろうか．この結果の一つの解釈を図 4-19 に示す．

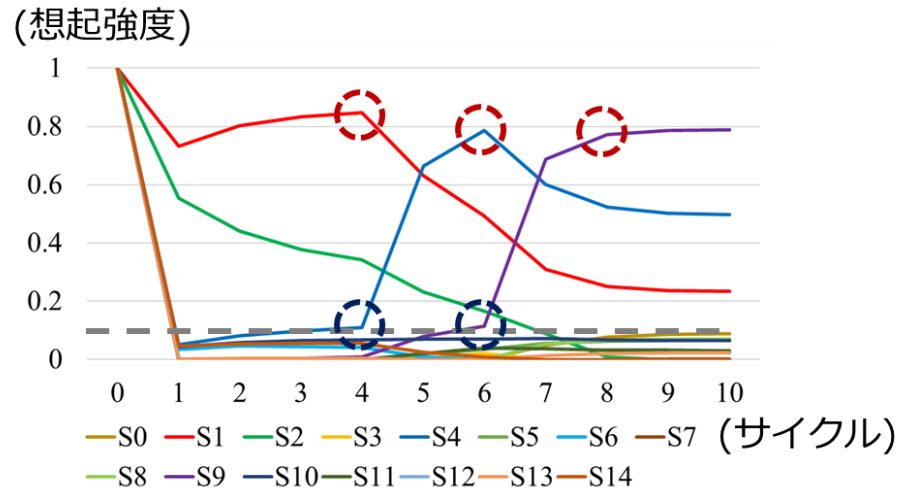

図 4-18　本シミュレーション条件におけるサイクルごとの各状態ベクトルの想起強度

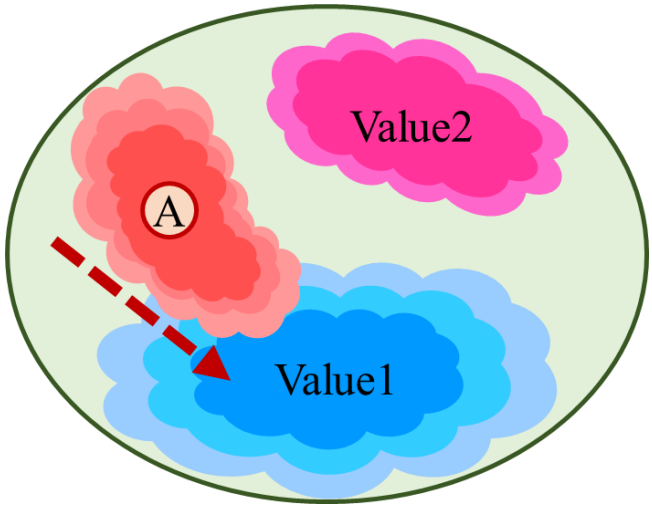

図 4-19　複数の価値分布と推論サイクルとの間の関係の解釈例

この図中の緑色の円はエージェントが経験するすべての状況とした．そしてそのうちの一部は，強化学習などにより既に価値が割り振られている．この世界中においてエージェントは現在地点から推論により価値のある領域を探索する．そしてパラメータ α を 0 とした際の論理的推論では意識的に一つの状態ベクトルに収束するように働く．それに対してパラメータ α 0 から 1 までの間の実数とした場合には，意識的に状態ベクトルを収束させている論理的推論の結果とは異なり，意識にはのぼらないが価値の強い状態に向かって徐々に推論が変化していくことに相当すると考える．

4.6.5 各推論システムの基本特性比較

本シミュレーションでは，提案したモデルの内，直観的推論をシステム 1(4.4.1 項), 論理的推論をシステム 2(4.5.1 項)とし, これらを統合した式(4.23)による推論をシステム 3(4.6.2 項)として，各システムの特性を確認する.

シミュレーション条件は 4.6.3 項と同様の地図，価値の割り振り，報酬量とした．そして各推論システムの条件は以下のようにした.

1) 直観的推論で探索することのできる探索範囲は 3 層までとする
2) 論理的推論によって 20 ステップ分計算した際に推論結果が収束しない場合には推論結果が収束しないと判断し, 相互抑制をかけることにより収束を促す
3) 推論の統合システムのパラメータ α の切り替え条件は, 推論する際の初期値は α を 1 とし，直観的推論を必ず実施するようにした．そして，直観的推論の結果として得られた次の状態ベクトル $\boldsymbol{x}^c$ 中に価値が複数含まれていた際にパラメータ α を 0 とし, 論理的推論に切り替わるようにした.

   なお本手法では，直観的推論において深い推論をした際，その行動状態を保持しているわけではない．そのため，直観的推論の 2 層目以降の結果として価値が含まれていた場合, 直観的推論の 1 層目の結果として得られた過去の経験に基づく行動数が最大となる方向に行動する(式(4.17)).

エージェントの推論結果に基づく探索経路を図 4-20 に示す．探索経路は 2 つのパターンに分けられた．システム 1 による意思決定では，報酬量の少ない青のキューブを得る行動をとり，システム 2，およびシステム 3 による意思決定では，報酬量の大きい赤のキューブを得る行動をとった．さらに，それぞれのシステムによる意思決定の過程を見ると，その計算処理は 4 フェーズに分けられた(図 4-20 破線部内フェーズ 1～4).

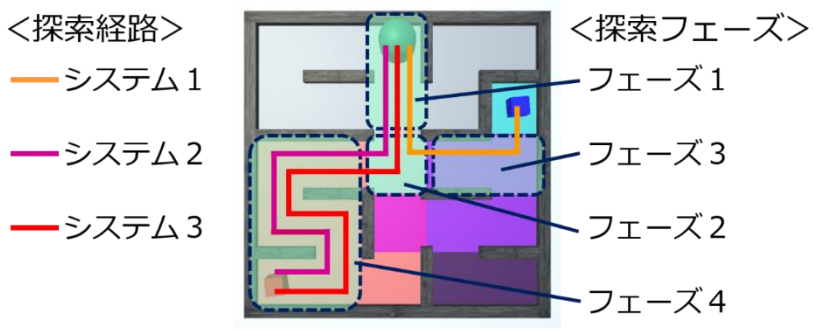

図 4-20　迷路探索用の地図，および価値配置

表 4-1 に 3 つのシステムそれぞれの各フェーズ内における平均探索ステップ数を示した．まず，システム 1 では，探索時に通過したすべてのフェーズにおいて平均探索ステップは 3 システムの内で最小であった一方で，エージェントが得たのは小さい報酬であった．これは過去に多くの回数を経験していた青のキューブに到達する経路への移動確率が高く，多くの価値が予測されたことによる．次にシステム 2 による推論では，探索にかかる平均探索ステップはすべてのフェーズにおいて最大であった一方で，エージェントは報酬量の大きい赤のキューブを得た．さらにシステム 2 のフェーズ 1 においては，事前の行動確率が一定であること，価値が見いだせない領域であることの 2 点により推論が 1 に収束しない．そのため，推論回数の上限時(本シミュレーションでは 100 回とした)における状態ベクトルに対して式(4.15)を適応することで行動決定する結果を得た．このことから，価値による変調を行いながら推論をするシステム 2 を用いることで，探索ステップはかかるが，確実に報酬量の多い行動を取ることができることが示唆された．

表 4-1　推論システムごとの探索ステップ数比較

(平均探索ステップ数／action)

探索フェーズ	システム 1	システム 2	システム 3
フェーズ１	2.5	37	2.5
フェーズ２	1	65	58
フェーズ３	1	―	―
フェーズ４	―	7.8	1

それに対して，システム 3 による推論では，探索にかかる平均探索ステップはフェーズ 1 と 4 では最小であったが，フェーズ 2 に対してはシステム 2 に近い大きいものであった．そして，エージェントは報酬量の大きい赤のキューブを得た．これは，フェーズ 1 ではスタート地点の近傍領域では価値のある状態を見出せなかったが，連想を繰り返すことで遠くに価値のある状態を見出すことができ，適切な行動ができたことによる．一方でフェーズ 2 では，青のキューブに対する価値と赤のキューブに対する価値の両方が同時に見出されて競合が起こり，システム 2 が機能して価値の高い赤のキューブに向かう行動を選択できたと考えられる．そしてフェーズ 4 では赤のキューブに対する価値のみがあるため，システム 1 のみで意思決定できたことによる．

これらのシミュレーションの結果から，提案したシステム 1 から 3 までのすべての推論システムにおいて推論特性に応じたシミュレーション結果を得ることができた．

第5章　　迷路課題による統合推論システムの検証

これまでに従来 2 つあるとされ別々にモデル化されてきた人の推論の統合モデルについてその手法を提案し，連想記憶により人の推論を表現することができるかの検証をしてきた．その結果，Evans ら [24]により言われてきた直観的推論の早く確率的な計算をする機能は相互想起型の連想記憶を用いることで，論理的推論の時間はかかるが意識にのぼる推論をする機能は自己想起型の連想記憶を用いることでそれぞれ表現できることを示した．

本章では，本研究で考えている図 5-1(詳細説明は図 3-1 を参照)に示したような環境において本研究で提案した推論システムを適用することで，予測状態空間での価値を最大化する意思決定のための状態空間探索としての機能を果たすことができることの検証を行う．

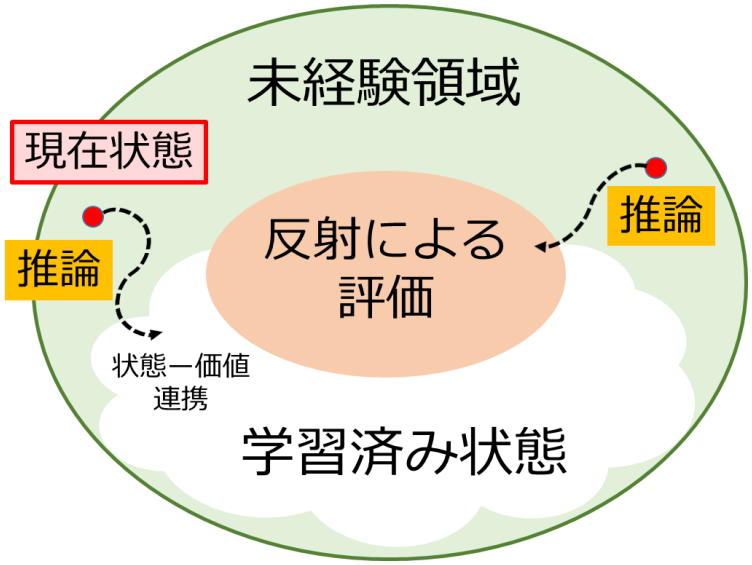

図 5-1　推論による価値探索の位置づけ

そのためには，第 4 章にて示した手動にて迷路探索中の報酬や価値を割り当てるなどの方法では不十分であると考える．なぜならば現実世界を考えた際，我々人は何かを学習した際にそれに至るまでの学習の情報をすべて教わるわけではない．我々は何か行動をし，その結果として報酬を得た際，ただ報酬を得た事実だけを学ぶのではなく，自らが報酬を得るまでに経験した行動情報も一緒に学習する．このようにしない

と人はいつまでたってもランダムに行動し続けてしまう．このような行動原理では，人は現時点までに絶滅しているだろう．

そこで本章では，迷路探索課題をグリッドワールドとして取り扱うことで，学習済みの領域と未学習の領域それぞれを構築し易く，これらが混在している領域を作ることで図 5-1 の状態を再現し，検証を行った．

5.1 シミュレーション環境

本節以降のシミュレーションでは迷路探索課題を作成するために，これまでシミュレーションにて用いてきた環境(Unity)とエージェント(Python)とをソケット通信により接続する環境を用いなかった．その理由として以下 3 点が挙げられる．

1) Unity と Python との間では時間が同期されておらず，使用するマシンに依存して計算時間が異なることを避けるため
2) 迷路課題を作成するにあたり，地図情報が共有されていないため，地図情報をグリッドワールドに落とし込む際に Unity で作成した地図に関する環境情報を Python 側でも同様に作成することが必要となるため
3) 本節以降のシミュレーションは，トイモデルによる検証を考えており，物理法則を含んだ環境でなくとも実現可能なシミュレーションであるため

これらの問題を解決するために本章以降のシミュレーションでは以下のような環境とした．環境はプログラミング言語 Python のゲームを作成するために開発されたライブラリである Pygame を用いて作成することとした．エージェントの内部処理は従来通り Python を用いて作成することとした．そしてこれらのプログラミング言語 Python により共通化された環境とエージェントのシステムを統合化した．これらを実現することにより，環境とエージェントの処理時間は同期し，迷路の環境地図も冗長にならない環境が実現できると考えた．

そして我々が生活する際，その瞬間に必要な欲求を知り，欲求に従った意思決定をする．そのためには自身の置かれている状況の把握，そしてその内部状態として必要となったものを判断し，探索のコントロールをする機能が必要になると考える．そのために本研究ではこのコントロールをするシステムの存在は，推論などの実処理をする機能の上位概念であると考え，メタシステムと名付けた．

これらを含めた本研究で考える環境とエージェントとの関係，およびエージェント内において処理される内容を概念図として表したものが図 5-2 である．本シミュレーションでは，探索する迷路内にいるエージェントには RGB カメラなどの外部センサの有無を考慮しない．なぜならば特徴量は，外部センサによって取得した情報から

DeepLearning などを用いて抽出されたものであると考えており，本研究の本質の対象外であると考えたためである．そのため，本シミュレーションでは，環境からの情報として地図上の位置を特定することのできる感覚情報(視覚や聴覚など)が入力され，その後入力された感覚入力から抽出された特徴量が入力されることを想定する．そしてエージェントではその特徴量の抽出と同時に，メタシステムの働きによりエージェント自身の内部欲求を自己評価し，現時点で，必要な価値は何なのかを判断する必要がある．この機能が存在しないとエージェントは現在の状態から見出された最大の価値のある方向にひたすらに進み続けることとなる．そしてメタシステムの働きによって見出された欲求などの内部状態の状態をトリガーとすることで，エージェントは推論すべき価値を見出すことができる．その見出された価値情報に従いエージェントは推論により自身の選択した行動ごとの次の状態を予測することができ，さらに大きな価値が見いだされる行動が何なのかを探索することが可能となる．

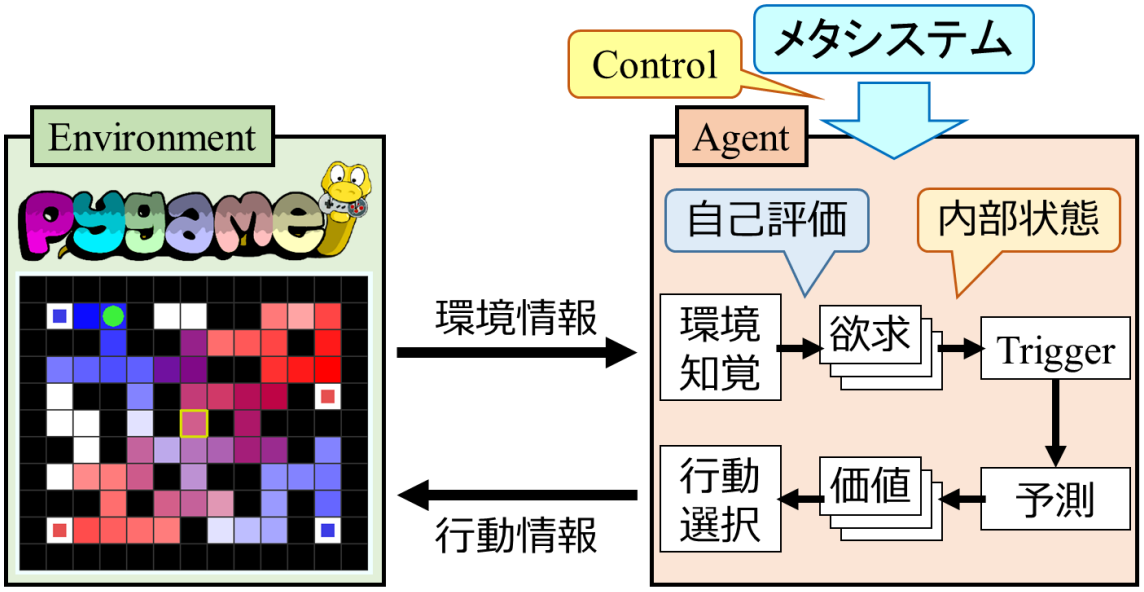

図 5-2　強化学習との連携に用いたシミュレーション環境

5.2 強化学習－推論の統合

本節ではエージェントが経験した内容を学習するための手法として, 強化学習を用いた．強化学習とは環境情報を観測し，その結果をもとに行動した際の価値の最大化(＝報酬の最大化)を目指す手法である．本節では，この強化学習手法の中でも最も基本的な手法の一つである Q 学習と本研究にて提案した推論システムとを統合することで，強化学習と推論システムとの統合の可能性を検証した.

強化学習と推論システムとの統合を考えた際，図 5-1 で示した未経験領域，学習済みの状態, 反射による評価のそれぞれの領域における強化学習と推論システムの働きについて議論する必要がある．従来の強化学習手法の基本 [21]は，未経験領域ではランダム行動をし，学習済みの状態に入った際は強化学習の学習結果に従い行動をし，最後に報酬を得る．そしてその経路上の情報を学習していくというものがほとんどであった．しかし，この手法では問題があると考える．その理由として，この手法ではエージェントは現在時点で必要な価値を見出し行動することはできるが, 学習済みの領域中において複数の価値が見いだされた場合でも, 強化学習ではその特性上価値の最大化を目指すことからその場の状況において価値の大きく見出されているものに対して行動し続けることになる. このような強化学習のみを用いた手法では人のようなその場の状況に応じた柔軟な行動は表出されない．人は何か目的に向かって行動をしていたとしても，途中で他の情報が入ってきたことにより行動が変化することはしばしば起こる. このような場合, 入力された情報を逐次評価することをしなければ柔軟な行動は表出することはできない．人のような柔軟な行動をするためには，4 章にて提案した推論システムなどによりその状態を逐次評価し，その結果に従い行動することが必要である．これらことから本研究では, エージェントの現在状態が変更されるたびに推論システムによる予測を行うこととした.

5.2.1 経験の加算よる事前確率の作成による推論手法の検証

本システムをリアルタイムシステムとして実現することを考えた際, 問題となるのは式(4.11)の相互想起ネットワークを作成する際の事前確率Pr(q|p)の作成方法である．本研究ではこの事前確率を作成する方法として, エージェントがランダム行動によって状態空間を経験した際, 相互想起ネットワークにその経験をする度に加えていくことで経験をネットワークに蓄積し，その経験数を別途記憶させた．そして相互想起ネットワークから現在状態ベクトルを基に想起する際に, 記憶させておいた入力の状態ベクトルに対応する経験数で割ることにより正しく想起計算が可能となることを想定した．

連想ネットワークは, エージェントの行動ごとに元々いた場所 p と現在いる場所 q との間の連想行列を作成し, 元々の連想ネットワークに足し合わせることにより更新した(式(5.1))．さらに，その状態ごとの経験数 N を別途記憶情報として記憶した．

$$W^{e\prime} = W^{e\prime} + x^q x^{p\mathrm{T}} \tag{5.1}$$

そして, 現在状態ベクトルx^rから想起ベクトルx^cを算出する際に状態の経験数 N で割ることで正しく想起できると考え，式を変形した．

$$\begin{aligned} x^c = W^{e\prime}\frac{1}{N}x^r &= \sum_q Pr(q|r)x^q x^{p\mathrm{T}}\frac{1}{N}x^r + \sum_q\sum_{p\neq r} Pr(q|p)x^q x^{p\mathrm{T}}\frac{1}{N}x^r \\ &\fallingdotseq \sum_q Pr(q|r)x^q \end{aligned} \tag{5.2}$$

強化学習と推論システムの統合シミュレーションは図 5-3 のような迷路課題を用いて検証した．推論方法は式(4.23)のパラメータ α の値は整数(0 または 1)とし，推論の開始時に必ず直観的推論を実施し, その探索において価値のある状態が複数見つかった際に論理的推論に切り替えるようにした．そして論理的推論では，価値の差がほとんどないなどの理由から状態ベクトルが収束しないことがある(詳細は 4.5.1 を参照)．そこで本経験の加算による事前確率算出シミュレーションでは, 相互抑制は論理的推論を始めてから 7 ステップ目経過時に収束しないと判断し，実行することとした．

シミュレーション条件は探索する迷路中には価値の大きさが異なる二つの報酬が設置されている(図 5-3 左の左上の報酬量を少なく，右下にある赤い四角の報酬量を多く

した). そして迷路内の情報は，エージェントが事前に獲得することを想定し，ランダムウォークによって経験させ，地図上のすべての位置，および方向に対する経験数がおおよそ均等になるようにした．なお，このランダムウォーク中は報酬の位置に到達しても報酬は得られないこととした．その後は Q 学習によって報酬の位置を学習させた．図 5-3 左中の経路が赤色の位置は価値の大小を示しており濃い方から薄い方になるにつれて, 価値が大から小への移り変わりを表現している. なお, その強化学習の経験により，左上の報酬量の小さい報酬の経験と右下の報酬量の大きい報酬の経験数の比は 2：1 となるようにした．このような条件とした際に，エージェントは地図上の緑の丸の位置から推論を用いて価値の大きい報酬を得ることである．

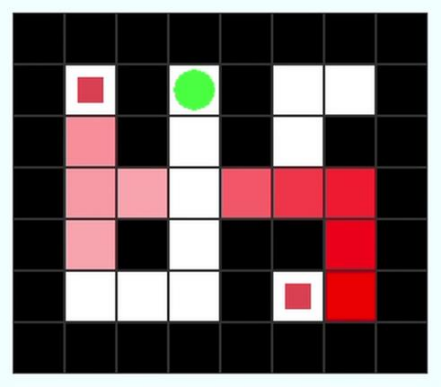

図 5-3　強化学習との連携に用いた迷路探索課題の地図

エージェントの推論結果を図 5-4 に示した．エージェントは初め，現在地点から直観的推論を用いて推論を開始した．しかし，エージェントが直観的推論を用いて 2 層目以降の深い推論をした際に，その計算結果が理論上の結果と大きく異なる結果となった．これは事前確率Pr(q|p)を相互想起ネットワークに作成する際，一つ前の時刻の状態ベクトルと現在状態ベクトルから記憶行列を作成し, 相互想起ネットワークに足し合わせる．そして記憶の元となった状態の経験数を別に記憶させておくというものであった．この方法の内, 現在状態から直観的推論を用いて想起する際に記憶させておいた経験数で割る方法をとったことに問題がある. この方法では直観的推論によって 1 層目を計算する際には問題は生じないため問題なく推論することができる. しかし問題は 2 層目以降である．2 層目として想起されるベクトルが 1 つに限定することができる場合は同様に問題なく計算することができるが, 想起されるベクトルが複数個存在する場合に問題が生じる．それは相互想起ネットワークにより想起されるベクトルが複数ある場合，そ

れぞれの経験数が異なる為 2 層目以降の相互想起をする際に割る経験数が異なってしまうことにある．その結果，2 層目以降の想起結果が正しく算出することができないことが問題となる．本シミュレーションはこの問題を理解したうえで，エージェントは直観的推論を 1 層のみ実行することとした．そして，1 層目にて行動価値が見出すことができない際には過去の経験をもとに softmax 法により行動決定する方法とした．

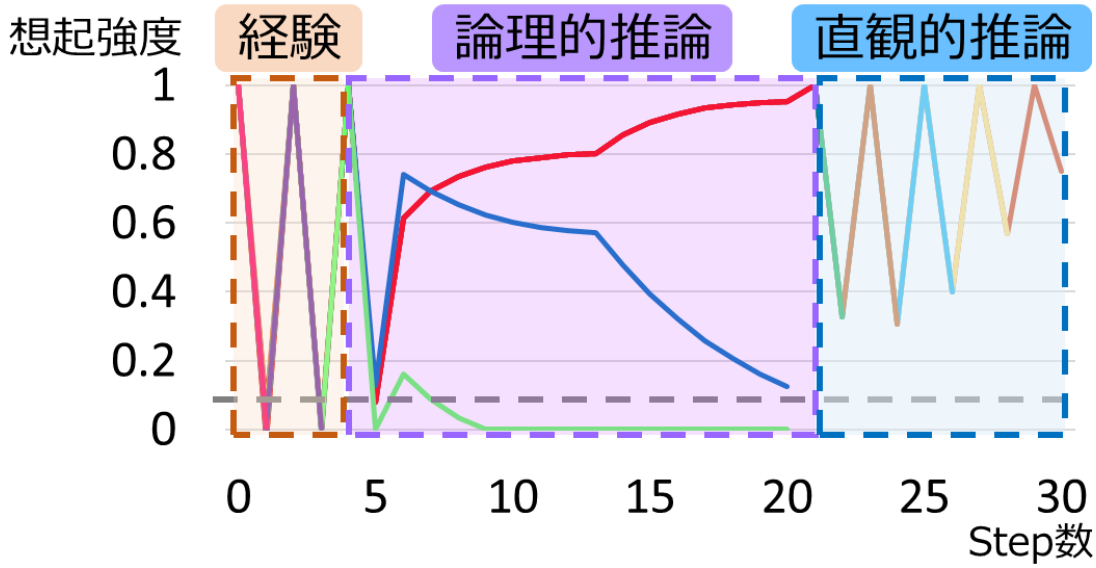

図 5-4　強化学習との連携によるシミュレーション結果

エージェントは初めの 2 ステップでは現在地点から推論した対象の領域において価値を見出すことができない．そのため，1 ステップ目では過去の経験，および連想記憶の対象となる地点が 1 カ所に限定されるため下方向に向かい行動する．そして 2 ステップ目の推論では，同様に直観的推論の結果として推論結果の位置に価値を見出すことができない．しかし，推論結果はエージェントが元居た位置，およびさらに下方向に進む方法の 2 種類が混合ベクトルとして想起される．しかし，エージェントには inhibition of return がかかっているため，元の位置に戻る行動は無視される．そのためエージェントはさらに下方向へ進む行動を選択する． 3 ステップ目は，地図上の中心に近く十字路になっている位置における推論である．この位置においては，下方向へ進む状態ベクトル(黄緑色)，左方向へ進む状態ベクトル(青色)，右方向へ進む状態ベクトル(赤色)の 3 つのパターンが直観的推論において想起されている(図 5-4 中の 5 ステップ目)．つまり，この状態は左右方向に関する状態ベクトルが競合したと言える．そのためこの時点で統

合パラメータ α を 0 にして論理的推論に切り替えて推論を続行した. その結果が 6 ステップ目から 20 ステップ目までである．その際 6 ステップ目の時点で推論結果として想起強度が弱かった下の方向への状態ベクトルは，3 ステップで 0 に収束した．それに対して競合していた左方向, および右方向の状態ベクトルは論理的推論の 7 ステップ目までは左方向の状態ベクトルの想起強度が高かったが，8 ステップ目以降では右方向の状態ベクトルの想起強度が高くなった. さらに本シミュレーションでは論理的推論繰り返し実行しているが 7 ステップ目までに収束しなかった際に相互抑制をするようにしていた. そのため論理的推論に切り替わってから 7 ステップ目(13 ステップ目)以降では相互抑制が働き，推論結果を 1 に収束することを促し，実際に右方向を示す状態ベクトルに収束した．そしてその次の行動として右に進んだ後は価値競合しておらず，さらに inhibition of return がかかっているため，元に戻る方向も想起されない．そのため直観的推論を用いることのみで意思決定することができ, エージェントは行動し, 報酬を得た. その際の特徴は，強化学習を用いて価値を割り振っているため，想起強度が報酬のある位置から離れるにつれて徐々に小さくなることが挙げられる. これは強化学習を用いて価値を伝播する際に，価値伝播の割合を状態空間ごとに 0.9 倍としているためこのような結果となったと言える．想起強度の割合が必ず価値の 0.9 倍とならない理由は，地図上の各位置において経験数が異なるため元々の想起強度が異なることが考えられる.

これらの結果から, 強化学習と本研究にて提案した推論システムとを統合することにより, 報酬や価値の配置を手動にて行った結果と同様の推論結果を得られることを示唆する結果を得た．しかし，本項で問題になったように，直観的推論では探索する深さは 1 層とは限らない. 深い層に対しての推論にならずとも, 複数層への探索は必要となる. そのため次項 5.2.2 では本経験の加算による事前確率算出シミュレーションにて 1 層しか正しく計算することができず問題となった相互想起ネットワークの作成手法について他の手法を検討し，その検証をする.

5.3 複数種類の価値による推論行動の切り替え

前節では図 5-1 のような人の経験する領域と価値との関係を想定し，強化学習と 4 章にて提案した推論システムとの統合の可能性が示唆される結果，およびリアルタイムでの推論システムの使用方法について示した．本節では実世界により近い環境でシミュレーションすることとし，エージェントに内部欲求の要素を加えた．さらに，実世界中では現在状態において複数の種類の価値が含まれている状態は多々存在する．このことから環境中に種類の異なる価値を複数設置することを想定した．価値領域が複数存在し，さらにこれらが競合することを想定すると図 5-1 は，図 5-7 のように拡張して捉えられる．

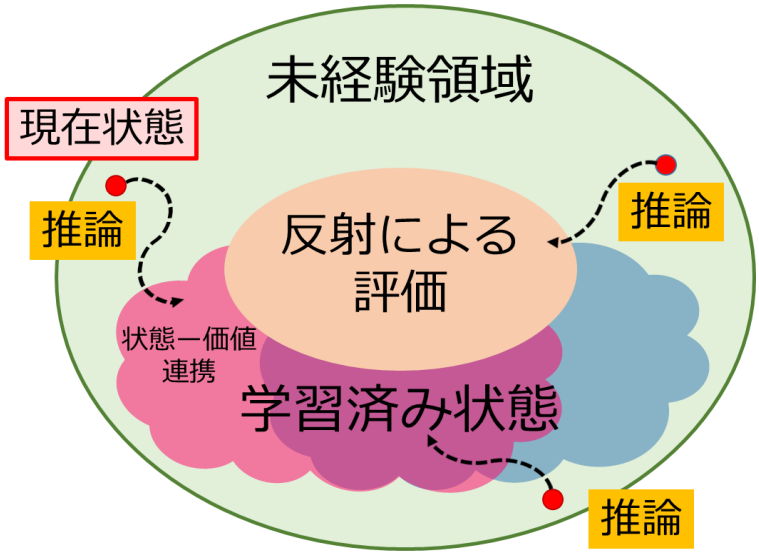

図 5-5　推論による複数学習領域の探索

図中の強化学習などで学習した価値領域は，単一の価値要素のみで構成されるのではなく，複数の報酬に対する価値領域があり，さらにこれらの領域はオーバーラップすると考えられる．

この環境においてエージェントが推論することができることを確認するために，本研究では，エージェントは事前に図 5-6 のような迷路環境中で個別の報酬(青の四角，赤の四角)ごとに報酬の位置を事前学習させ，その価値領域を取得(図 5-6 下段がそれぞれの報酬，およびその価値に対応)させた．それぞれの結果を組み合わせた図中の左上の地図では，探索する迷路の全体像，およびエージェントの位置，報酬の位置と報酬

に対応する価値マップをオーバーラップさせて表示している．これにより，エージェントが現在どの位置においてどのような方向に行動したかを可視化することができる．

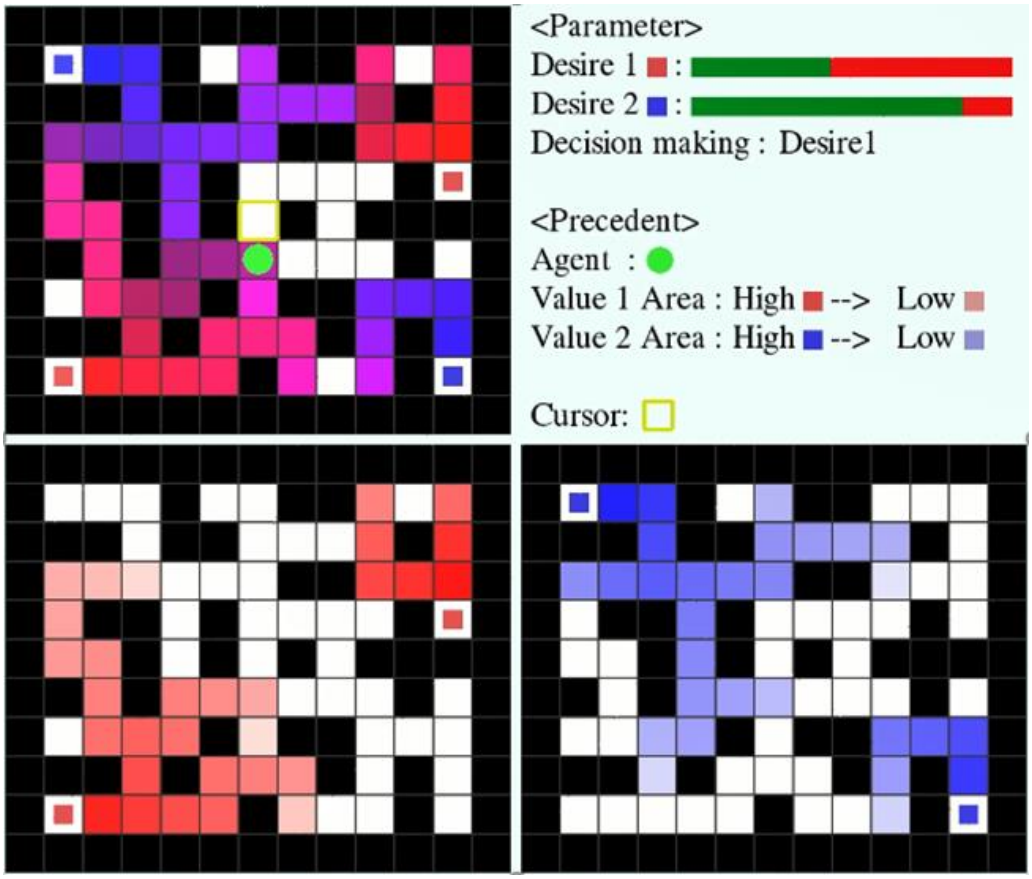

図 5-6　複数の価値領域を考慮したシミュレーション

そして図中の右上では，エージェントの内部で持つ欲求の内部パラメータ，およびエージェントの意思決定に用いられた内部状態を可視化している．この欲求の内部パラメータは緑色のバー，および赤色のバーで表現されており，緑色のバーと赤色のバーの総和を取ったものが，内部欲求パラメータの最大値となる．緑色のバーではエージェントの内部欲求の現在の値を示しており，エージェントが行動するたびに徐々に減少する仕組みとした．そしてエージェントが現在の内部状態に従い柔軟な行動をするためには，エージェントの内部状態を評価する必要がある．本研究ではこのエージェントの内部状態を各欲求の値に応じて重みづけをし，その結果を評価をする方法としてエージェントの内部欲求をそれぞれ最大が 1 になるように正規化(式(5.3))しその結果に対して softmax 法(式(5.4))を用いることで決定した．

$$Desire(i) = \frac{Desire(i)}{Max\ \ Desire\ \ Value} \tag{5.3}$$

$$Decision\ \ making(i) = \frac{exp\left(\frac{Desire(i)}{T}\right)}{\sum_j^n exp\left(\frac{Desire(j)}{T}\right)} \tag{5.4}$$

式(5.3)では，徐々に時刻経過により徐々に減っていく欲求の最大値を Max Desire Valueとし，欲求毎に現在の欲求の値から割ることで，それぞれの欲求の値とした.

式(5.4)の Softmax 法とは正または負のデータを確率情報に変換するために用いられる計算方式であり，その結果の合計を 1 になるようにする特性がある．その方法はまず式(3.3)の分子において入力情報毎のx_i(ここでは，入力情報を欲求の種類数とし，これらに探索に用いられた粒子, および粒子が探索により状態空間より抽出した欲求毎の価値の積)を計算することで，入力された情報を正の値へと変換する．そして分母では，入力情報 (見出されている全欲求)の総和を計算する．そして入力情報毎に算出した分母の入力情報の総和で割ることで，全体を 1 にする確率計算をすることが可能となる．ここでは説明の簡単化のために Softmax 関数の制約緩和理論のパラメータである温度パラメータ T を 1 に固定し，内部状態(欲求)を 2 つ(欲求 1，2)に限定する．そして欲求 1，2 それぞれの現在状態の値がそれぞれ 0.25 と 0.65 として入力として Softmax の式に入力した．その結果，出力としてそれぞれ約 0.4 と約 0.6 を得た．この約 0.4，0.6 がそれぞれの選択確率となる．このように Softmax を用いることで今回のような入力の合計が 1 にならない(今回の入力はそれぞれ 0.25，0.65 なので合計は 0.9 となる)場合でもそれぞれの選択確率を常に 1 にした状態で評価することができる.

また，強化学習には学習状況を高速化する方法が存在する．その手法の一つに報酬を受け取った際に報酬を得るまでに経験した情報を一度に学習する Profit Sharing などの方法がある [38]が，本研究ではこの手法を取り入れていない．その理由は本シミュレーションの目的は学習を早くすることではなく，あくまで強化学習と推論システムとを統合することで，内部欲求に応じたエージェントの柔軟な行動が観察できるか，ということに主眼を置いているからである．そのため，本研究で用いた Q 学習の学習方法は現在地点の隣に価値のある領域が存在した場合に現在地点にのみ価値を割り振る一番基本となる手法を使用した.

シミュレーション条件は以下の通りである．

(1) 赤の四角の報酬，青の四角の報酬に対する学習は別々に事前学習を実施することにより行う．
(2) エージェントのランダムウォークによる状態空間の探索は報酬の位置に 10 回到達するまで(報酬は与えないため学習行われない)とする．
(3) 強化学習による事前学習は，ランダムウォークの実施し報酬の位置に 10 回到達する(報酬を与え，価値を伝播させ学習させる)までとする
(4) 赤の四角の報酬，青の四角の報酬は別の種類の報酬であると考え，内部欲求において判断された価値とは異なる価値は推論の対象外とする．
(5) 内部欲求の評価はエージェントの行動ごとに毎回行うこととする．

この条件に従ったエージェントの行動シミュレーションを行った結果，エージェントは内部欲求に従った，その場で内部欲求に従った価値判断をする柔軟な行動を示した．つまり，エージェントは必ずしも現在の位置から近い報酬を得るために行動するのではなく，報酬の位置が遠くてもその報酬に対応している価値領域の状態に従い行動し，報酬を得る結果となった．

なお，本来であれば本シミュレーションの結果を評価することが必要であるが，本研究では現状その評価はしていない．その理由は本課題によって評価したい内容は人のような柔軟な行動をすることができるエージェントの行動であるのに対し，その評価をする基準が現時点で存在しないことが挙げられる．この柔軟な行動をするエージェントの行動評価をする方法については，今後の課題とする．

第6章　　まとめ

本論文では，大きく以下の点について調査およびモデル化をし，その結果として以下のような新規性を見出すことができた.

(1) 神経科学や他の研究分野の知見を統合し, 感情の価値計算システム仮説の提案した.
(2) 分散型連想記憶の相互想起モデル, 自己想起モデルを用いた人の推論システムのモデル化した. そして提案した推論モデルの論理的推論の解釈として，従来の Tree 探索の深さ優先探索のような結果を得た.
(3) 提案した推論モデルの迷路課題を用いた実用性を検証し, 強化学習などと連携可能なこと, 複数の価値がある環境において推論が可能であることを示唆する結果を得た.

以下では，本学位論文のまとめとして先に述べた内容に着目し，それぞれについてまとめるとともに，最後に本論文の成果が知能についてどのように貢献するかという点について述べる.

6.1 シミュレーションの妥当性と一般性

本研究では 4.6.5 項にて, 連想記憶モデルを用いた直観的推論(システム 1)と論理的推論(システム 2), およびシステム 1 とシステム 2 とを組み合わせた推論(システム 3)のアーキテクチャを提案し，簡単な迷路探索課題を解きその効果を検証した.

計算機シミュレーションの結果，システム 1 では，連想記憶モデルに過去の経験に基づく確率的かつ，分散的なニューラルネットワーク手法を取り入れることで，ステップ数の短い推論を実現できた(表 4-1 システム 1 列). さらに，システム 2 では，状態ベクトルに価値情報を付与し，その結果の自己想起を繰り返して収束を待つことで，シンボル的な推論が可能となり，収束までにステップ数の多くかかる，すなわち遅い推論が実現できた(表 4-1 システム 2 列). さらに，このシステム 1 とシステム 2 とを組み合わせたシステム 3 では，推論システムを動的に切り替えることで，システム 1，2 の単独で

は説明できない推論の二重過程モデルの内部過程を示唆する結果を得た(表 4-1 システム 3 列).

しかし，シミュレーションのデザイン，つまり環境とモデルのパラメータ設定の妥当性については検証が必要である．図 4-15 の地図は一般の行動決定課題に比較してサイズが小さい．これは，システム 1 における記憶パターンの強度が連想の反復により急激に減衰し，連想の深さ，および連想範囲を広げることにより連想の対象となる状態をこれ以上多くしても解が得られないためである．それに対して，システム 2 はより深い推論が可能であり，図 4-16 の報酬 A までも探索は到達している．システム 1 とシステム 2 の特性の違いは明確であり，論理的推論による深い探索の可能性の評価という課題の目的は達成できたと言える．

報酬 A, B の設定値や事前の強化学習による周辺への価値の広がりの設定は，システム 2，あるいはシステム 3 の基本的な動作への影響は少ないと考えられる．これは，想起ベクトルが価値最大の記憶パターンに収束するのは連想記憶への価値認識層からのフィードバックによるもので, そこでの価値の多寡は収束までのステップ数には影響するものの，収束条件には影響が少ないことによる．

本研究手法の応用課題として図 5-1 のような状態空間と価値との関連を想定し，強化学習(本研究では Q 学習)と連携させたシミュレーションを行った．その結果として強化学習により付与された価値領域に対して推論が可能となること, およびリアルタイムでの相互想起ネットワークを作成することができることを示唆する結果を得た．さらに本研究では図 5-1 の拡張として，それぞれの学習は個別に行う条件があるが，複数の種類の報酬が迷路中に存在することを想定した際, エージェントの現在位置から近い報酬を必ず得るのではなく, エージェントの内部欲求の判断に従った柔軟な行動を示す結果を得た．

また，記憶パターン間の連想の強度を決める条件付き確率 $Pr(q|p)$ にはベイズ確率としての条件以外には制約はない. その意味で本シミュレーションの結果はある程度の一般的に用いられる課題(迷路課題や人を対象とした心理実験など)に対して適用が可能であると考えられる．

6.2 脳のように動作する行動決定モデルとしての位置づけ

本研究では人の意思決定は価値に駆動されるという行動経済学や認知科学の心理レベルの知見からスタートした．本モデルを脳のモデルとしてみたとき，連想記憶，直交性，パラメータ切り替えなどの妥当性について議論する必要があろう．

脳は多くの領野が相互結合したネットワークであり，それらをまとめると大規模な連想記憶システムともいえる．その場合，状態ベクトルをいくつかの区間に分けて領野に対応させると，連想行列がそのまま脳内結合を表すことになり，そこに相互想起の非対称結合と自己想起の対象結合のコネクションが重なっているという解釈になる．脳という幅広い性質を持つ部位を一つの連想記憶として扱うことを正しいということは現段階では難しい．しかし，論理的推論という現象の創発にはまずは本研究のような荒いレベル間でのモデルが必要であると考える．

本モデルの重要かつ議論のある仮説は，状態ベクトル群が相互にほぼ直交する条件であろう．神経科学では神経興奮の直交性は確認されていない．CNN のような階層型ネットワークでは，その下位層では情報表現の抽象度が低く興奮パターンは入力に依存した局所的相関があると考えられる．しかし，上位層になるにつれて情報表現の抽象度が上がって概念化されることで，ベクトルの直交性は向上すると考えられる．ベクトルのスパース性が変わっても，この性質は不変であろう．これより，階層ネットワークのより上位の階層で本モデルの示す推論が現れるなら，離散的な推論が意識されやすい現象と矛盾はない．

連想記憶のパラメータ α の切り替えはシステム 3 の実現の鍵である．シミュレーションでは α の値を切り替えるルールを決めることで実装した．実際の脳で α に相当する機能がどのように実現されているかは不明であるが，脳内の信号の流れを切り替えるシステムが脳にあることは、人の fMRI 研究からも確実である．例えば Baddeley の中央実行系はまさに本モデルのように脳内の信号の流れや領域の活動を制御すると考えられるシステムであるが [39]，その実際の脳過程や働きについてはいまだ未解明である．

6.3 本研究結果が感情研究に対して示唆するもの

本研究では，今後求められるであろうAIを用いたロボットなどには，人のコミュニケーションを理解する機能が必要であると考えた．このコミュニケーションを理解するために必要となるものとして人の感情が挙げられると考え，『感情＝価値計算システム仮説』を提案した．そしてこの仮説は，人の感情を脳の一部のみにより想起されるものではなく，脳全体として価値計算した結果として表出される行動であると考えたものである．

それに対して本研究の主軸は人の持つ推論であり，一見内容が離れており関連がないように見える．しかし，本研究では人の推論を『予測状態空間での価値を最大化する意思決定のための状態空間探索』として考えている．本シミュレーションの解釈として感情と推論の立ち位置は，推論により予測状態空間中に含まれる価値を最大化して次の時刻の状態を予測し，その結果を用いて意思決定した結果として見いだされる行動結果を感情として考えることができると考える．そして人が生み出すことのできる行動パターンは顔の表情や腕や足などの各部位を用いた曲げや捻りだけでなく，これらを組み合わせることができ非常に多彩である．それに対して本研究で行ったシミュレーションはグリッドワールドを用いた迷路探索課題であるため，エージェントの行動パターンは最大でも上下左右に限定される．そのため，本研究結果から直接感情の創発に繋がる結果は得られていない．

しかし本研究の仮説として考えている，感情とは価値計算した結果として表出される行動である，という点については本研究で提案した推論システムを用いることで，推論から意思決定につながるまでの計算ができることを示す結果を得た．現時点では人のような身体を持ったエージェントを用意し，人のような行動をとらせることはしていないが，この先の研究の方向性としてこのような研究を進めていく方向もあるように思う．

6.4 本モデルが知能について示唆するもの

実用的な機械学習の立場からも，推論のモデル化は重要である．現在，意思決定の主流と考えられる強化学習は，試行錯誤的な探索により強力な行動学習を実現しているが，行動学習に多数の試行を必要とする，報酬が変わると再学習が必要など，現実世界の問題に対して実用的とは言えない性質がある．それに対して推論は，新奇場面でも対象世界の法則についての知識があれば，その場での内的探索による意思決定が可能であり，強化学習と補完的な性質を持っている．ただ，これまで論理的推論はシンボル的な手法による実装が主流であり，ニューラルネットが実現する分散型の表象に対する汎用的な推論は困難であった．本モデルはそれに対する一つの方向であろう．階層的事物認識ネットワークの途中の階層での情報表現は，直観的・論理的の両方の推論の実現の可能性がある．

意思決定手法の一つに，エピソード記憶の利用がある．本モデルでは各場面の表現としてほぼ直交した記憶パターンを用いており，個別の事象を記憶するエピソード記憶とは相性がよい．エピソード記憶による意思決定，推論によるエピソード類似場面での意思決定，強化学習によるエピソードの一般化，という意思決定のモデルの統合もまたありうる方向であろう．

知覚と認知との関係について Barsalou [40]は知覚的シンボルシステムを提案し，入力層から上位層に向かうにつれて単純な特徴パターンから概念の表現パターンに変化していくとした．この点においては本研究において仮定した階層ネットワークの上位層の直交性と共通する見方である．さらに概念に基づく行動決定では，見いだされた概念の流動的な組み合わせによる行動決定をシミュレートし，最適と判断された行動を実際に行うとしている．本研究で用いた記憶パターンには流動性はないが，概念を表す多くの記憶ベクトルの状態空間に価値を付加してベイズ推論する直観的な意思決定過程は，その計算過程の候補となる可能性があろう．

本研究で提示した論理的な推論の計算は，事象の意識化ともいえる過程を含んでいる．シンボルの創発に関しては，意識化することにより思考を一つに絞ると考える Global Workspace Theory とも関係があろうが，本稿ではこの点については深く議論しない [41]．ただ，論理的な推論は意識とのかかわりがあることは否定できず，この点については今後も検討を続けていく必要がある．

謝辞

本研究を行うにあたり，玉川大学工学部の大森 隆司 教授には指導教員，および主査として常日頃からあたたかくご指導，ご鞭撻を賜りました．大森教授には，私が玉川大学大学院 工学研究科に入学した際より気をかけていただき，研究だけでなく様々な方面に対して助言をいただきました．心より深く感謝致します．また，ドワンゴ人工知能研究所の山川 宏 氏，玉川大学工学部の相原 威 教授，および玉川大学量子情報科学研究所の加藤 研太郎 教授には，副査として厳しく，かつ前向きな研究議論をしていただきました．ここに深い感謝の意を表します．そして玉川大学工学部の佐々木 寛 教授には，私が玉川大学 工学部 知能情報システム学科に入学したころから玉川大学大学院工学研究科の修士課程に至るまで，長期間にわたりご指導いただきました．深く感謝いたします．また，玉川大学大学院工学研究科の諸先生方には，私が玉川大学大学院の修士課程の頃から気にかけていただき，様々なご指導や励ましの言葉をいただきました．ここに深く感謝致します．

また，本研究を進めるにあたり日頃から研究議論だけでなく，励ましの言葉をいただきました玉川大学脳科学研究所の研究員である山田 徹志 氏に深い感謝の意を表します．また本研究を進めるにあたり研究議論，および研究補助してくださった大森研究室の学生諸君に感謝致します．

なお，本研究で用いたエピソード記憶を用いるというアイデアのきっかけは，全脳アーキテクチャ・イニシアティブ主催の第 3 回全脳アーキテクチャ・ハッカソン「目覚めよ海馬！：汎用人工知能プロトタイプにむけた海馬モデルの組み込み」に参加したことによります．また，エピソード記憶を用いた更なる研究をすることができたのは公益財団法人　科学技術融合振興財団により補助金の助成を受けたことによります．ここに深く感謝致します．

参考文献

[1] S.J.Russell et al., エージェントアプローチ人工知能 第 2 版, 共立出版, 2008.

[2] Akihiro Funamizu et al., “Neural substrate of dynamic Bayesian inference in the cerebral cortex,” nature neuroscience, Vol. 19, pp. 1682-1689, 2016.

[3] Maël Donoso et al., “Foundations of human reasoning in the prefrontal cortex,” Science, Vol. 344, No. 6191, pp. 1481-1486, 2017.

[4] J. A. Russell, “A circumplex model of affect,” Journal of personality and social psychology, Vol.39, pp.1161-1178, 1980.

[5] P. Ekman et al., “What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS),” Series in Affective Science, Oxford University Press, 1997.

[6] J. LeDoux, “The Emotional Brain: The Mysterious Underpinnings of Emotional Life,” Simon & Schuster, 1998.

[7] D. Derksa et al., “The role of emotion in computermediated communication: A review,” Computers in Human Behavior, Vol.24, No.3, pp.766-785, 2008.

[8] A. Ilyasova, “Emotional competencies: Connecting to the emotive side of engineering and communication,” IPCC, pp.1-5, 2015.

[9] 阿部香澄 他, “子供と遊ぶロボット : 心的状態の推定に基づいた行動決定モデルの適用,” 日本ロボット学会誌, Vol.31, No.3, pp.263-274, 2013.

[10] 山田徹志 他, “「保育の質」の定量化に向けた子どもとロボットの関わり－子どもの心的状態推定へのアプローチー,” 第 33 回日本認知科学会大会, OS13-4, 2016.

[11] O.D.Chernavskaya et al., “An architecture of the cognitive system with account for emotional component,” Biologically inspired cognitive architectures, Vol.12, pp.144-154, 2015.

[12] A.V. Samsonovich, “Emotional biologically inspired cognitive architecture,” Biologically inspired cognitive architectures, Vol.6, pp.109-125, 2013.

[13] J. Vallverdú, “A cognitive architecture for the implementation of emotions in computing systems,” Biologically inspired cognitive architectures, Vol.15, pp.34-40, 2016.

[14] 戸田正直, 感情, 東京大学出版会, 1992.

[15] 野村 理朗, “情動, DOI：10.14931/bsd.3050,” [オンライン]. Available: https://bsd.neuroinf.jp/wiki/%E6%83%85%E5%8B%95.

[16] S. Koelsch et al., “The quartet theory of human emotions: An integrative and neurofunctional model,” Physics of Life Reviews, Vol. 13, pp. 1-27., 2015.

[17] MacLean PD, “A triune concept of the brain and behaviour,” Toronto: University of toronto press, 1973.

[18] 大竹文雄　他, 脳の中の経済学, ディスカヴァー携書, 2012.

[19] 栢沼晋太郎　他, “エピソード記憶と価値の連合した行動決定アルゴリズムの評価,” 人工知能学会全国大会, 2L2-OS-6a-04, 2018.

[20] Greg Wayne et al., “Unsupervised Predictive Memory in a Goal-Directed Agent,” arXiv:1803.10760, 2018.

[21] Richard Sutton, “Integrated Architectures for Learning, Planning, and Reacting Based on Approximating Dynamic Programming, Appeared in Proceedings of the Seventh Int. Conf. on Machine Learning, pp. 216-224,” 1990.

[22] 信原幸弘, 情動の哲学入門 価値・道徳・生きる意味, 勁草書房, 2017.

[23] Masahiro Miyata et al., “Modeling emotion and inference as a value calculation system,” BICA2017, Vol. 123, pp. 295-301, 2017.

[24] Jonathan St　et al., “How many dual-process theories do we need? One, two, or many?,” Oxford Scholarship Online, 2009.

[25] 服部雅史, 思考と推論: 理性・判断・意思決定の心理学, 北大路書房, 2015.

[26] Russel, Norvig et al., エージェントアプローチ人工知能第 2 版, 共立出版, 2008.

[27] 大森隆司　他, “粒子モデルと価値評価系による直観的推論の計算アーキテクチャ,” 日本神経回路学会全国大会, pp. 55-56, 2017.

[28] Yu J et al., “Advances to Bayesian network inference for generating causal networks from observational biological data.,” Oxford University Press 2004, Vol.20, No.18, pp.3594-3603, 2004.

[29] Alex Graves et al., “Hybrid computing using a neural network with dynamic external memory,” Nature, Vol. 538, pp. 471-476, 2016.

[30] John Anderson et al., “A Production System Theory of Serial Memory,” Psychological review, Vol.104, No.4, pp.728-748, 1997.

[31] 宮田真宏　他, “感情の価値計算システム仮説にもとづく前頭葉推論モデルの検証,” 人工知能学会大会, 3K1, 2017.

[32] W. Lotter et al., “Deep Predictive Coding Networks for Video Prediction and Unsupervised Learning,” arXiv:1605.08104, cs.LG, 2016.

[33] LIS(Life In Silico), “学習環境シミュレータ Life in Silico(LIS)上で 150 人が「AI 開発の民主化」に向け第一歩,” WBAI, [オンライン]. Available: http://wba-initiative.org/1036/.

[34] 中野馨, アソシアトロン－連想記憶のモデルと知的情報処理－, 昭昇堂, 1979.

[35] Kaoru Nakano, “Associatron-A Model of Associative Memory,” IEEE Transactions on systems, man, and cybernetics, Vol.SMC-2, pp.380-388, 1972.

[36] Haim Sompolinsky, “Temporal Association in Asymmetric Neural Networks,” Physical review letters, Vol. 57, No. 22, pp. 2861-2864, 1986.

[37] Takashi Omori et al., “Emergence of symbolic behavior from brain like memory with dynamic attention,” Neural networks, Vol.12, No.7-8, pp.1157-1172, 1999.

[38] Xiao Huang et al., “Novelty and Reinforcement Learning in the Value System of Developmental Robots,” Computer Science and Engineering Department Michigan State University, 2002.

[39] A D Baddeley et al., “Working Memory,” Psychology of Learning and

Motivation, Vol 8, pp.47-89, 1974.

[40] L. W. Barsalou, “Perceptual symbol systems,” Behavioral and Brain Sciences, Vol. 22, pp. 577-660, 1999.

[41] Sid K, “ Levels of processing during non-conscious perception: a critical review of visual masking,” Mental processes in the human brain, Vol. 362, No. 1481, pp. 857-875, 2007.

研究業績

学術論文

1. 宮田真宏，大森隆司: 価値に駆動された連想記憶に基づく人の推論過程の統合モデルの提案，日本知能情報ファジィ学会，2018

国際会議

1. Muhammad Attamimi, Masahiro Miyata, Tetsuji Yamada, Takashi Omori, Ryoma Hida: Attention Estimation for Child-Robot Interaction, pp.267-270, HAI2016, 2016
   DOI: http://dx.doi.org/10.1145/2974804.2980510
   →ポスター発表は自ら実施
2. Takashi Omori, Masahiro Miyata: Modeling of Emotion as a Value Calculation System, pp.308-315, ICONIP 2016, 2016
   DOI: https://doi.org/10.1007/978-3-319-46687-3_34
3. Masahiro Miyata, Takashi Omori: Modeling emotion and inference as a value calculation system, BICA2017, Vol.123, pp.295-301, 2017
   DOI: https://doi.org/10.1016/j.procs.2018.01.046
   →BICA RESEARCH PRIZE 受賞
4. Ryoma Hida, Tetsuji Yamada, Masahiro Miyata, Takashi Omori: Development of human behavior observation system for mental state estimation, 2017 International Workshop on Smart Info-Media Systems in Asia, SS3-1, pp.158-161, 2017
5. Ryoma Hida, Tetsuji Yamada, Masahiro Miyata, Takashi Omori: Development of Interest estimation Tool for effective HAI, pp.483-486, HAI2017, 2017
   DOI: https://doi.org/10.1145/3125739.3132597
6. Masahiro Miyata, Takashi Omori: Emergence of symbolic inference based on value-driven intuitive inference via associative memory, pp.370-375, BICA2018, Vol.145, 2018
   DOI: https://doi.org/10.1016/j.procs.2018.11.087

国内会議

1. 宮田真宏, 相原威, 佐々木寛: 事象関連電位を用いた大脳優位半球の推定, 信学技報, vol. 115, no. 513, MBE2015-131, pp.159-162, 2016
2. Masahiro Miyata , Takeshi Aihara , Hiroshi Sasaki: Study on non-invasive estimation of the language lateralization, 第 39 回日本神経科学大会, 2016
3. 宮田真宏, 大森隆司: 感情の価値システムとしてのモデル化の試み, 第 33 回日本認知科学会大会, O3-1, 2016
4. 山田徹志, アッタミミ・ムハンマド, ジャン・ビン, 宮田真宏, 中村友昭, 大森隆司, 長井隆行, 岡夏樹, 西村拓一: 「保育の質」の定量化に向けた子どもとロボットの関わり －子どもの心的状態推定へのアプローチー, 第 33 回日本認知科学会大会, OS13-4, 2016
5. 肥田 竜馬, 山田 徹志, 宮田 真宏, 大森 隆司, 長井 隆行, 岡 夏樹: ロボットから紐解く保育士の対人インタラクション技能の定量化, HAI シンポジウム 2016, G-9, 2016
6. 宮田真宏, 肥田竜馬, 山田徹志, 張斌, 中村友昭, 大森隆司: 『保育の質』の定量的分析に向けた半自動アノテーションツールの開発, 第 17 回計測自動制御学会 システムインテグレーション部門講演会, pp.2366-2369, SI2016, 2016
7. 山田徹志, 宮田真宏, 肥田竜馬, 大森隆司: 子どもの主体的な行動を通した保育の質の客観化手法の検討 -AI を用いた子どもの行動計測と心的状態推定-, 日本発達心理学会第 28 回大会, P4-4, pp.342, 2017
8. 宮田真宏, 大森隆司: 感情の価値計算システム仮説にもとづく強化学習による脳幹モデルの検証, 信学技報, vol. 116, no. 521, NC2016-64, pp.1-6, 2017
9. 肥田竜馬, 山田徹志, 張斌, 宮田真宏, 石川久悟, 根岸諒平, 大森隆司, 中村友昭, 長井隆行, 岡夏樹: 保育の質の定量化のための人間行動センシングと解析ツールの開発, 第 31 回人工知能学会大会, 2H3-OS-35a-5, 2017
10. 宮田真宏, 大森隆司: 感情の価値計算システム仮説にもとづく前頭葉推論モデルの検証, 第 31 回人工知能学会大会, 3K1-OS-06a-2, 2017
11. 山田徹志, 肥田竜馬, 宮田真宏, 大森隆司, 中村友昭, 長井隆行, 岡夏樹: 子どもの関心の推定を通した保育の質の客観化の試み, 日本教育工学会 第 33 回全国大会, 3a-101-04, pp.775-776, 2017
12. 大森隆司, 宮田真宏: 粒子モデルと価値評価系による直観的推論の計算アーキテクチャ, 第 27 回 日本神経回路学会 全国大会, 2017
13. 肥田竜馬, 山田徹志, 宮田真宏, 大森隆司: 対人インタラクションのための人の心的状態推定システムの研究, HAI シンポジウム 2017, P-9, 2017
14. 宮田真宏, 大森隆司: 連想記憶モデルに基づく人のシンボル的推論のモデル化, 第 8 回 人工知能学会 汎用人工知能研究会, SIG-AGI-008-02, 2018
15. 堤優奈, 栢沼晋太郎, 川添紗奈, 宮田真宏, 大森隆司: エピソード記憶と価値を紐づけた海馬モデルによる行動学習の分析, 第 8 回 人工知能学会 汎用人工知能研究会, SIG-AGI-008-03, 2018
16. 山田徹志, 肥田竜馬, 宮田真宏, 大森隆司: AI による保育研究支援システム開発に向けた予備的調査, 第 32 回人工知能学会大会, 1O3-OS-15b-03, 2018

17. 栢沼晋太郎, 川添紗奈, 堤優奈, 宮田真宏, 大森隆司: エピソード記憶と価値の連合した行動決定アルゴリズムの評価, 第 32 回人工知能学会大会, 2L2-OS-6a-04, 2018
18. 宮田真宏, 大森隆司: 価値に駆動される連想記憶によるシンボル的推論の検証, 第 32 回人工知能学会大会, 2L3-OS-6b-02, 2018
19. 宮田真宏, 大森隆司: 価値に駆動された連想記憶に基づく人の推論過程の統合, 第 34 回ファジィシステムシンポジウム, MC3-1, 2018
    →FSS2018 ポスター・デモセッション 優秀発表賞 受賞
20. 浅利恭美, 山田徹志, 宮田真宏, 大森隆司: 子どもの関心推定のための行動センシングシステムの開発, 日本教育工学会 第 34 回全国大会, 2a-B203-02, pp.679-680, 2018
21. 山田徹志, 浅利恭美, 宮田真宏, 中村友昭, 長井隆行, 岡夏樹, 大森隆司: AI により子どもの発達・教育研究を支援する分析手法の検討- 子どもの位置・向き情報による関心の推定 -, 日本教育工学会 第 34 回全国大会, P1a-C102-01, pp.51-52, 2018
22. 宮田真宏, 大森隆司: 価値に駆動された連想記憶に基づく人の推論システムの機能的検証, 信学技報, vol. 118, no. 470, NC2018-73, pp.157-162, 2019
23. 青栁燎, 山田徹志, 騰川裕太, 浅利恭美, 宮田真宏, 大森隆司: 保育活動映像からの子どもの関心推定の試み, 日本教育工学会 研究会(JSET19-1), B-15, pp.185-191, 2019

その他

1. 宮田真宏, 早川博章, 川添紗奈, 栢沼晋太郎, 堤優奈: エピソード記憶による行動選択, 第 3 回全脳アーキテクチャ・ハッカソン「目覚めよ海馬！:汎用人工知能プロトタイプにむけた海馬モデルの組み込み」, 2017
   →優秀賞 受賞
2. 川添紗奈, 栢沼晋太郎, 小島和弥, 宮田真宏: 価値を用いた行動決定モデルは汎用的な課題に適用できるか, 第 4 回全脳アーキテクチャ・ハッカソン「AI にまなざしを」, 2018
   →優秀賞 受賞