株式会社とめ研究所 知能情報関連懸賞論文

ニューラルネットワーク議論に関する研究

-議論の意味論を計算するニューラルネットワーク-

巻口航

新潟大学大学院 自然科学研究科

数理・情報電子工学専攻 博士前期課程２年

平成 21 年 3 月 26 日

概 要

近年，エージェントへの社会性実現の試みに，人工知能やソフトウェア科学の分野においても世界的に注目が集められている．とりわけ，矛盾した情報下におけるエージェント間の協調・衝突解決に対して，議論の適用が重要視されている．議論研究では，対話や討論の場でなされる論証を記号で形式化することによって，論理的に矛盾した論証同士の関係をもとに意味論や証明論を与え，議論の解決を数学的に規定できる枠組みを構築する．これらは数理論理学的に処理するのが一般的な試みであった．一方，形式化された議論の処理を，ニューラルネットワークで実現する新規的な試みが 2005 年に A. Garcez らによって提案された．Garcez らは数理論理学の推論機構とニューラルネットワークの学習能力を議論の枠組みの中で融合することを目標にしている．しかしながら，ニューラルネットワークによる計算結果が，従来議論研究で定義されてきた多くの意味論に対し，一致性が保証されていないという重大な問題点を残していた．すなわちニューラルネットワークが議論の意味論を扱えるのかという問いが未解決のままであった．本論文では，その疑問を解決すべく，議論研究において最も代表的とも言える P. Dung の定義した 4 つの議論の意味論（Stable Extension, Preferred Extension, Complete Extension, Grounded Extension）を計算するニューラルネットワークのモデルを提案する．さらに，そのモデルの計算の一致性に対する健全性を保証する定理と証明も与える．この提案によって，先に述べた重要な問いに対する答えを与え，connectionism と symbolism のそれぞれ異なった学問領域の融合可能性を示し，2 つの異なった学問の特性による相補的な計算モデルが構築可能になることに道が開けることを信じる．

目 次

第1章　はじめに

1.1　背景（記号議論とニューラルネットワーク議論）

近年，数理議論学とそのエージェント計算への応用に関する研究は，人工知能やソフトウェア研究への新しい有望な試みとして，現在世界的に注目され特にヨーロッパ連合の中で強く推進されている．現在，計算論的議論に関する四つの代表的な国際会議がそれぞれの会議の目的の住み分けを行いながら並行して開催されている：

ArgMAS: `http://homepages.inf.ed.ac.uk/irahwan/argmas/`

COMMA: `http://www.comma-conf.org/`

ArgNMR: `http://lia.deis.unibo.it/confs/ArgNMR/`

NMR: `http://www.kr.org/NMR/`

形式的議論フレームワーク $\mathcal{AF} =< AR, attacks >$ は，議論の参加者が提出する論証 (argument) の集合 AR と，それらの間の反論関係 $attacks$ からなる [7]．そこでは，論証が議論の結果として妥当あるいは社会的に合意可能な論証の集合を意味論的にあるいは証明論的に規定することに主要な関心が払われる．本論文では，このような記号のみによって議論の状態を判定しようとする研究方法を，記号主義的 AI にならって，記号 (主義的) 議論と呼ぶことにする．

2005 年，記号議論の従来的な方法ではなく，議論フレームワークをニューラルネットワークに変換し，並列処理と繰り返し計算によって，議論の結果としての論証の状態を決定しようとする興味深い示唆的な研究が Garcez らによって提唱された [6]．ここでは，Garcez らの方法を，記号主義的議論に対してニューラルネットワーク議論と呼ぶことにする．しかしながら残念なことに，Garcez らのニューラルネットワーク議論では，ニューラルネットワークによって計算された論証の状態が，これまで広く受け入れられてきた数学的意味論に基づく記号議論による論証の状態，すなわち議論の意味論と，どのように対応しているのかという非常に重大で重要な疑問には，言及も考察もなんら行われていないのが実状であった．すなわち，ニューラルネットワーク与えた論証の状態に基づき選び取られる正しい論証の集合と，記号議論において定義された議論の意味論により選び取られる正しい論証の集合の一致性が示されていない．本論文著者は過去の研究 [11] において，Garcez ら

の提唱したモデルが生成する結果（prevailing argument の集合）と記号議論の代表的な意味論である Grounded Extension に基づき得られる結果（justified argument の集合）が，概ね一致しながらも，厳密にはある議論において不一致性が生じることを数学的に考察した．Garcez らの研究では，ニューラルネットワーク議論で導入された 3 つの論証の状態 “prevailing”，“failure to prevail”，“defeated” は，それぞれが直感に基づいて述べられているだけであり，数学的に定義された概念とはなってない．したがって代表的な記号主義的議論の研究で定義されている論証の状態 “justified”, “defensible”，“overruled”[13] とは一致していない．別の言い方をするならば，Garcez らのニューラルネットワーク議論の重大な欠陥は，意味論を欠いていることにあるといってよい．このままでは古い論争 [8] が，ニューラルネットワーク議論と記号議論の間において再燃してくる恐れが出てくることになる．すなわち，ニューラルネットワークでは，議論という認知現象を記号議論のように厳密に捉え説明できないのではないかという懸念である．

1.2　目的

本論文では，ニューラルネットワーク議論の発想の新規性を活かしながら，Garcez らのニューラルネットワーク議論のこの欠陥を修復するために，記号議論の代表的な不動点意味論を含む４つの意味論（Stable Extension, Preferred Extension, Complete Extension, Grounded Extension）[7, 13] を取り上げ，それを計算する新しいニューラルネットワークの構成を与えることで，それが可能であることを示す．

本論文で提案するモデルが，ニューラルネットワーク議論の復権をもたらすのみならず，ニューラルネットワーク議論と記号議論の有効なハイブリッド型議論フレームワークへの道を拓くものと信じる．

1.3　論文の構成

本論文は，以下の章で構成される．

2 章： 準備として本論文で提案するニューラルネットワーク議論に必要な範囲内でニューラルネットワークの基礎知識を与える．

3 章： 本論文のニューラルネットワークが計算の対象とする議論の意味論を Dung の研究 [7] にならい導入する．

4 章： ３章で導入した議論の意味論を計算するための，ニューラルネットワークの生成アルゴリズムと，ニューラルネットワークに必要な計算手続きを定義し，その手続きが導く結果と記号議論の解釈が一致することの健全性の定理と証明を与える．

5章： 本研究の意義をまとめ，要検討課題について述べる．

第2章　ニューラルネットワークの基礎

この章では，本論文の展開に必要な範囲内でニューラルネットワークの基礎を準備する．なお，この導入は一部 [6] にならう．

人工ニューラルネットワーク（以下，論文内では単に「ニューラルネットワーク」と呼ぶ）は，以下のパラメータによって特徴付けられるユニット（ニューロン）を節点とする有向グラフである．時間 t においてニューラルネットワークのユニット i は，以下のパラメータを持つ：

- 入力ベクトル $I_i(t)$
- 入力ポテンシャル $U_i(t)$
- 活性化状態 $A_i(t)$
- 出力 $O_i(t)$

ネットワークの各ユニットは，方向と重みを持つ接続によって相互連結される．もしユニット i からユニット j への接続が存在するならば，この接続の重みを $W_{ij} \in \mathbb{R}$ と表す．時間 t におけるユニット i の入力ポテンシャルは $U_i(t) = \Sigma_j W_{ij} I_i(t)$，すなわちユニット i に向かう加重和により得られる (図 2.1 を参照)．時間 t におけるユニット i の活性化状態はユニット i の活性化関数を h_i としたとき，$A_i(t) = h_i(U_i(t))$ と定義される．通常 h_i は，線形関数，非線形 (ステップ) 関数，シグモイド関数のいずれかである．さらに，θ_i は，ユニット i の活性化関数の閾値として知られる．ユニットの出力値は，出力関数 $f_i(A_i(t))$ によって与えられる．通常 f_i は恒等関数である．

ニューラルネットワークのユニットは層状に体系化される．層の数を n としたとき，このネットワークを「n 層フィードフォワードネットワーク」と呼び，これは非巡回グラフである．このネットワークは，入力層，$n-2$ 個の中間層，出力層の間で接続され列を構成している．$n = 3$ のとき，このネットワークを「単隠れ層ネットワーク」と呼ぶ．本論文で扱うニューラルネットワークはこの形態であり（図 2.2 を参照），ある層に存在する各ユニットは，下の層に存在する各ユニットから接続される．これらの機構によって，フィードフォワードネットワークは関数 $\varphi : \mathbb{R}^r \to \mathbb{R}^s$($r$ はネットワークの入力層に存在するユニット数，s は出力層のユニット数) の計算を実現している．

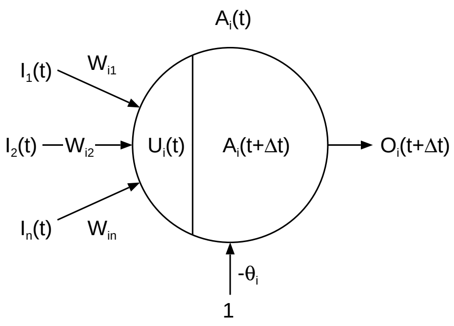

図 2.1: ニューラルネットワークのユニット

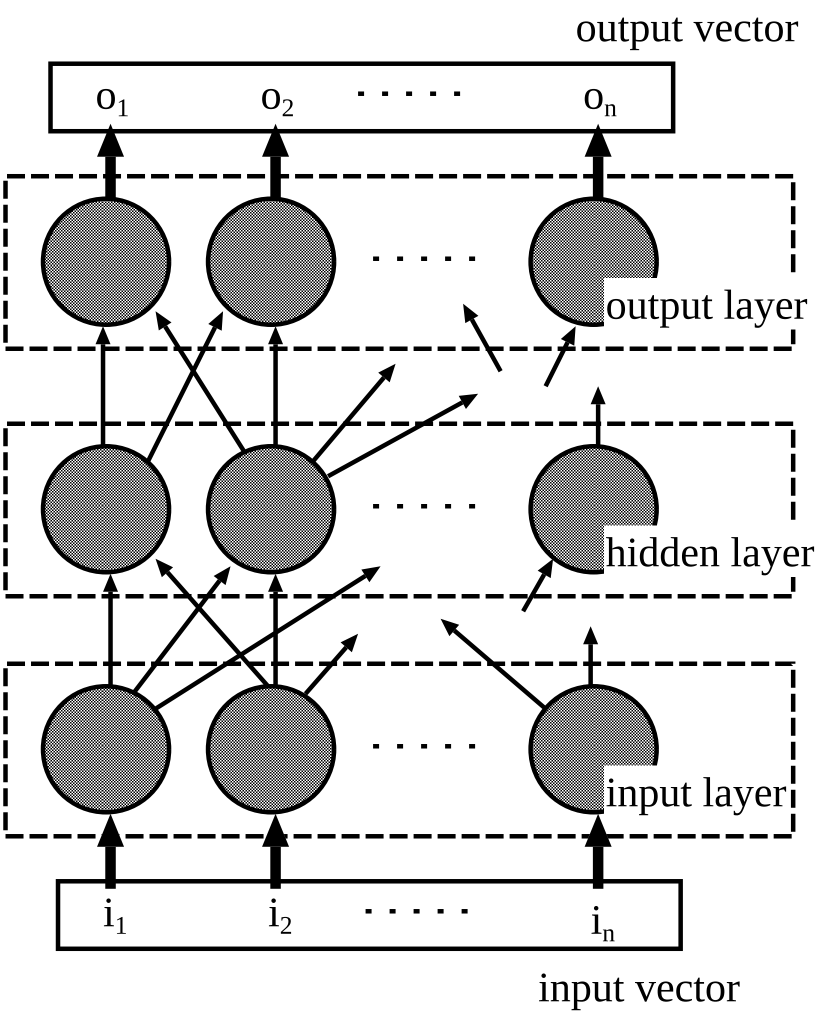

図 2.2: 単隠れ層ネットワークの例

単隠れ層ネットワークにおける，関数 φ の計算手順を以下に与える：

時間 t_1 入力ベクトルを入力層に与える．

時間 t_2 入力ベクトルが中間層に伝播し，中間層の各ユニットが入力ポテンシャルと活性化状態の値を計算する．

時間 t_3 中間層の各ユニットの活性化状態の値が出力層に伝播し，出力層の各ユニットが入力ポテンシャルと活性化状態の値を計算する．

時間 t_4 出力ベクトルが出力層から得られる．

ここまでが，本論文の展開に必要なニューラルネットワークの基礎知識である．

[6] では，ニューラルネットワークの学習アルゴリズムを導入することによって，ニューラルネットワーク特有の学習能力を，議論計算という枠組みの中で応用する方法を模索している．しかし，本論文では学習アルゴリズムによる接続重みや閾値変化を導入しない単純なニューラルネットワークのモデルにより，議論の意味論を扱う上で重要な関数の入出力を実現している．

第3章　議論の意味論

本章では，記号議論における意味論の定義を [7] にならい導入する．

3.1　議論フレームワーク

記号議論では，今着目している対話や討論における各発言・主張を論理的に導かれた「論証 (argument)」であると捉える．議論における論証は曖昧な情報（我々人間の持つ知識など）下より推論されるために，議論の場において論証同士が矛盾の関係，すなわち攻撃関係を持つ場合がある．例えば，a という命題と $a \rightarrow b$ という命題を前提として b を主張しようとする論証 $[a, a \rightarrow b]$ と，b と矛盾する命題である $\neg b$ を主張する論証 $[c, c \rightarrow \neg b]$（$\rightarrow$, $\neg$ は一般的に命題論理で用いられる含意，否定の記号と同じ意味である）が存在したとき，この２つの論証は互いに攻撃関係にあるという．以下論文内では，論証を A,B,$\ldots$ といったアルファベットによって置き換えて表記し，具体的に論証の内部がどのような命題の列になっているかは問わない．すなわち，論証 $[a, a \rightarrow b]$ を A，$[c, c \rightarrow \neg b]$ を B と置き換えると，A が B を攻撃し，B も A を攻撃していると規定できる（２つの論証の攻撃関係が常に上のような双方向であるとは限らず，一方的な攻撃関係を構築する場合もある．しかし，本論文では論証の中身の表現については問わないので，攻撃関係の構築の定義に関する詳細は省く）.

いかに議論の中で正しい論証を選び取るか，すなわち議論の意味論を規定するにあたって重要な概念が議論フレームワークである．本論文では，記号議論の研究で最も重要かつ影響力を持っている Dung の定義 [7] に基づき，議論フレームワークの定義を導入する．議論フレームワークは論証の集合と，論証間の攻撃関係の集合の二項組であるが，形式的には以下のように定義される．

定義 1.（議論フレームワーク [7]）
議論フレームワークは $\mathcal{AF} =< AR, attacks >$ と表される二項組である．ただし，AR は論証の集合であり，$attacks$ は AR 上の二項関係，すなわち $attacks \subseteq AR \times AR$ である．

上の例の場合，議論フレームワークは $\mathcal{AF} =< \{A, B\}, \{(A, B), (B, A)\} >$ と表現できる．以後 $attacks(A, B)$ を「論証 A が論証 B を攻撃する」という意味で用いる．

3.2　論証の無衝突性・受理可能性・許容可能性

この議論フレームワークをもとに議論の意味論を定義する．議論の意味論の定義における重要な概念として，論証の集合の無衝突性 (conflict-free) と受理可能性 (acceptability), 許容可能性 (admissibility) がある．それぞれの定義を以下に述べる．また以降の定義では，$\mathcal{AF} =< AR, attacks >$ を仮定する．

定義 2.（論証の集合の無衝突性 [7]）
論証の集合 $S \subseteq AR$ は無衝突である $\Leftrightarrow \forall A \in S, \forall B \in S(\neg attacks(A, B))$

定義 3.（論証の受理可能性と許容可能性 [7]）

1. $S \subseteq AR$ は $A \in AR$ を擁護する（$defends(S, A)$ または，A は S に受理可能である，とも読む） $\Leftrightarrow \forall B \in AR\ (attacks(B, A) \rightarrow attacks(S, B))$[1]

2. 無衝突な $S \subseteq AR$ は許容可能（*Admissible Set*）である $\Leftrightarrow \forall A \in AR\ (A \in S \rightarrow defends(S, A))$

さらに Admissible Set に関して，以下の定理が成り立つ．

定理 1.（$\emptyset$ の Admissible Set に対する所属性 [7]）
任意の $\mathcal{AF}$ に対して，$\emptyset$（空集合）は必ず $\mathcal{AF}$ の *Admissible Set* である．

この定理については [7] に述べられているが，自明であるので証明は省く．
以降では，$conflict\text{-}free(S)$ を「S は無衝突である」という意味で用いる．

3.3　議論における４つの意味論

上で導入した３つの概念をもとに議論の意味論を定義する．議論の意味論とは，直感的には「議論（対話）が終了した時点における，正しい論証の集合の選び方」である．我々人間が様々な議論の方法を持っているように，議論の意味論も様々に定義しうる．例えば，法廷の場で，検察官と被疑者の間で議論が行われ，裁判官は議論でなされた主張（論証）間の攻撃関係を考慮し，どの主張を最終的な判決に反映させるかを，法廷で用いるべき議論の意味論を考慮した上で判決を下す．法廷での議論の意味論は，正しい選び方をただ一つに定めるような（後に定義する Grounded Extension に基づく）定義がなされるが，日常的には複数の正しい論証の集合の選び方ができるような議論の意味論も存在する（例えば後に定義する Preferred Extension などの意味論がこれに該当する．このような意味論

[1] S が論証の集合（$S \subseteq AR$）かつ A が論証（$A \in AR$）のとき，$attacks(S, A) \Leftrightarrow \exists B \in S(attacks(B, A))$ と定義する．

のもとでは，上の例 $\mathcal{AF} =< \{A, B\}, \{(A, B), (B, A)\} >$ において，$\{A\}$ が正しいとも言えるし，$\{B\}$ が正しいとも言える）．ここでは，議論の意味論として Dung の定義を [7] より導入する．

定義 4. **(Preferred Extension [7])**
$S \subseteq AR$ は *Preferred Extension* である $\Leftrightarrow$ S は $\mathcal{AF}$ の許容可能な集合において，包含関係に関して極大なものである．

定義 5. **(Stable Extension [7])**
無衝突な集合 $S \subseteq AR$ は *Stable Extension* である $\Leftrightarrow \forall A \in AR(A \notin S \rightarrow attacks(S, A))$.

Grounded Extension の定義のために以下の特性関数 F_{AF} を導入する．

定義 6. **(特性関数 F_{AF} [7])**
特性関数 F_{AF} を以下のように定義する：

- F_{AF}: $2^{AR} \rightarrow 2^{AR}$
- $F_{AF}(S) = \{A \in AR \mid defends(S, A)\}$

F_{AF} に関して以下の命題が得られる．

命題 1. **(特性関数 F_{AF} の単調性 [7])**
F_{AF} は包含関係に関して単調である．

この命題の証明は [7] を参照されたい．命題 1 より，F_{AF} は包含関係に関して単調であるため，最小不動点を持つことが保証される [16]．F_{AF} の最小不動点により Grounded Extension を以下のように定義する．

定義 7. **(Grounded Extension [7])**
$S \subseteq AR$ は *Grounded Extension* である $\Leftrightarrow$ S は F_{AF} の最小不動点である．

この Grounded Extension を基礎として，H. Prakken らはこれを証明論的に判定する方法（対話的証明論）を提案し [13]，“justified”, “overruled”, “defensible” という論証の状態を定義した [14]．

定義 8. **(Complete Extension [7])**
無衝突な集合 $S \subseteq AR$ は *Complete Extension* である $\Leftrightarrow \forall A \in AR(A \in S \leftrightarrow defends(S, A))$

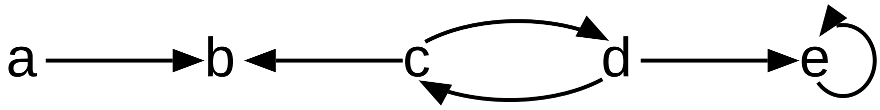

図 3.1: $\mathcal{AF}$ の例の図表現．各シンボルが論証を表し，矢印が攻撃関係を示す．

以上４つの議論の意味論を定義したが，以降 $\mathcal{AF}$ における Admissible Set，Preferred Extension，Stable Extension，Grounded Extension，Complete Extension をそれぞれ AS_{AF}，PE_{AF}，SE_{AF}，GE_{AF}，CE_{AF} と略記する．

例 1. ここでは例として，$AR = \{a, b, c, d, e\}$ かつ $attacks = \{$*(a,b), (c,b), (c,d), (d,c), (d,e), (e,e)*$\}$ であるような議論フレームワーク $\mathcal{AF} =< AR, attacks >$ を与える（図表現を図 *3.1* に与える）．各意味論に対して計算される論証の集合は以下の通りである（この例は *[3]* による）．$\mathcal{AF}$ に対し，各意味論による論証の集合は以下のように定まる．

- AS_{AF}*:* $\emptyset, \{a\}, \{c\}, \{d\}, \{a, c\}, \{a, d\}$
- CE_{AF}*:* $\{a\}, \{a, c\}, \{a, d\}$
- PE_{AF}*:* $\{a, c\}, \{a, d\}$
- SE_{AF}*:* $\{a, d\}$
- GE_{AF}*:* $\{a\}$

第4章　議論の意味論を計算するニューラルネットワーク

3章では，記号議論で定義される代表的な議論の意味論を導入した．これらの意味論に基づいて $\mathcal{AF}$ から正しい論証の集合を計算する手続きが，記号処理に基づく方法で様々に提案されている．例えば GE_{AF} を手続き的に計算する方法としては，Prakken らが [13] で以下のような提案をした：検察官と被疑者の役を与えられた2人が，議論の参加者として交互に論証の提出を行い，対話の木を構築し，対話が終了した時点の木の表現により議題に対する論証が GE_{AF} に含まれるかどうかを判定する．また，[17] では G. Vreeswijk らが，同様の2人の議論参加者による対話的な論証の提出を繰り返すことによって，証明論的に PE_{AF} に属する論証を判定する試みを提案している．他にも意味論により選ばれる論証の集合を，証明論的な手続きに依らずに判定する手続きが P. Besnard らによって提案されている [3][2].

上で挙げたような記号処理的手続きに依らない，全く新規な試みが Garcez らによって提案された [6]．彼らの手法では，ニューラルネットワークの計算により，Dung の定義による $\mathcal{AF}$ における正しい論証の集合を判定する[1]．しかし，ニューラルネットワークによって計算される正しい論証の集合が，どの意味論をもとにしているかについて明確な言及がなされていなかった．一見，ニューラルネットワークの入出力の反復計算が定義6の関数 F_{AF} の不動点を求めるかのように振る舞い，GE_{AF} を求める計算に対応するのではないかという予見が得られた．その予見を出発点として著者らは [11] にて，それらの一致性を数学的に検証し，Garcez らの提案モデルと F_{AF} の不動点を求める計算の不一致性を発表した．さらにニューラルネットワークの計算・構造と $\mathcal{AF}$ の関係をより明らかにすべく，Garcez らのモデルでは一致しなかった GE_{AF} を計算するニューラルネットワークによる新しい計算手続きを与えた [12]．著者らの ArgMAS2007 における発表 [11][12] において，主な関心の中心となったのは，GE_{AF} のみならず，数多く存在する他の意味論を計算可能なニューラルネットワークの構築は可能かどうか，という点であった．本論文では，この疑問に対する解答を提示することを重要な目標としている．

この章では，3章で定義した意味論全てを計算可能な新しいニューラルネットワーク議

[1]厳密には，Dung の $\mathcal{AF}$ を拡張した，Bench-Capon の拡張議論フレームワーク $\mathcal{VAF}$[1] を Garcez らは対象としているが，Dung の $\mathcal{AF}$ はより単純に定義されているため，Dung の $\mathcal{AF}$ 上でも計算が可能である．

論を提案する．まず，厳密な定義を始める前に，直感的アイデアを伝えるための例を導入する．

4.1 基本アイデアのトレース例

本論文で実現しているニューラルネットワーク議論は，任意の $\mathcal{AF}$ に対する $AS_{AF}, CE_{AF}, PE_{AF}, SE_{AF}, GE_{AF}$ を判定するような計算を行う．すなわち，ニューラルネットワーク議論に必要な入力は任意の $\mathcal{AF}$ であり，AS_{AF} と 4 つの意味論の集合 $CE_{AF}, PE_{AF}, SE_{AF}, GE_{AF}$ を出力とするような系を想定している．

今，図 4.1 を例に取り上げる．図 4.1 は $\mathcal{AF} = < \{i, j, a\}, \{(\ i, a\ ), (\ a, i\ ), (\ j, a\ )\} >$ の図表現である．

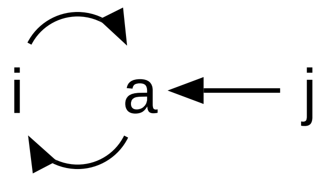

図 4.1: $\mathcal{AF}$ の例

3 章の定義に従えば，$AS_{AF}, CE_{AF}, PE_{AF}, SE_{AF}, GE_{AF}$ として以下のリストされた集合を得ることができる．

- AS_{AF}: $\emptyset, \{i\}, \{j\}, \{i, j\}$
- CE_{AF}: $\{i, j\}$
- PE_{AF}: $\{i, j\}$
- SE_{AF}: $\{i, j\}$
- GE_{AF}: $\{i, j\}$

今，$\{i\}$ が上の 5 つのいずれの集合に該当するのかをニューラルネットワーク議論で調べたい ($S_0 = \{i\}$ とする)．まず，計算を行うためのニューラルネットワークが必要である．図 4.2 がこの例で使用するニューラルネットワークであるが，これは $\mathcal{AF}$ から変換アルゴリズム（4.2.1 で導入）によって生成されるものである．具体的なニューラルネットワークの作り方，結線の意味，各ユニットの閾値の決定法などについては 4.2.1 以降で導入する．以下では，図 4.2 を用いて図 4.1 の AS_{AF} と 4 つの意味論の集合がどのように決定されるかを例を追って述べる．以下，図 4.2 のネットワークを $\mathcal{N}$ とする．

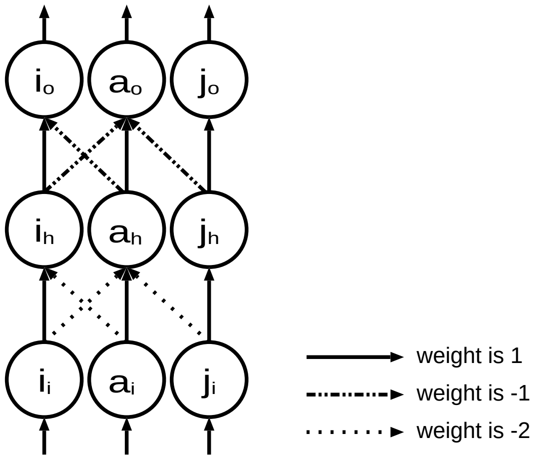

図 4.2: 図 4.1 から変換されたニューラルネットワーク

便宜上用いる記号の定義

例を追う前に，説明の便宜上用いる記号を以下のように定義する．今，$S \subseteq AR$ とする．

- S が擁護する論証の集合 $S' = \{A \in AR \mid defends(S, A)\}$
- S に攻撃される論証の集合 $S^+ = \{A \in AR \mid \exists B \in S\ (attacks(B, A))\}$

さらに上の記号に対して入出力の順番（タイムラウンド）を $\tau = k$（k は 0 以上の整数）として付随表記し,「k 回目の入出力計算で判定される集合」を意味する．例えば $S'_{\tau=2}$ ならば,「($\tau = 0$ を 0 回目として）2 回目の入出力の計算で S' と判定される集合」を意味する．

計算を始める前の事前知識

さらに，基本アイデアを捉える重要な概念を以下に列挙する：

- $\mathcal{N}$ は単体としては単隠れ層ネットワークであり，出力層と入力層を接続することで反復計算を行う．
- $\mathcal{N}$ は反復計算の中で出力が入力と一致した時点で，計算が安定状態にあるとする．
- $\mathcal{N}$ の入出力は，$1, 0, -1$ の 3 値を成分とするベクトルであるが，初期入力は任意の $S \subseteq AR$ を 4.2.2 で導入する定義に基づき，ベクトルに変換することで得られる．
- $\mathcal{N}$ の入出力ベクトルに現れるベクトル成分である $1, 0, -1$ の直感的意味はそれぞれ以下の通りである：
  - 入力ベクトル内の成分 1：判定の対象にしている論証の集合 S の要素

– 入力ベクトル内の成分 -1：S に攻撃される論証の集合 S^+ の要素

– 入力ベクトル内の成分 0：上の二つの集合のどちらにも属さない論証

– 出力ベクトル内の成分 1：S に擁護される論証の集合 S' の要素

– 出力ベクトル内の成分 -1：S' に攻撃される論証の集合 S'^+ の要素

– 出力ベクトル内の成分 0：上の二つの集合のどちらにも属さない論証

- 初期入力を与えてから安定状態に到達するまでの処理は，AR の任意の部分集合の内のひとつである S が，AS_{AF}, CE_{AF}, SE_{AF} のいずれに該当するかを判定している（この時点で全ての AS_{AF} と 4 つの意味論の集合を判定できるわけではない）.
- AR の任意の部分集合を S として，それら全てを初期入力として与えれば，AS_{AF} と 4 つの意味論の集合を全て判定できる.
- 反復計算の中で，$\mathcal{N}$ 単体が一回の入出力処理で行っているのは，入力を S, S^+（初期入力時は $S^+ = \emptyset$）として，出力を S', S'^+ として求めていることに相当する.
- PE_{AF} の判定は，$\mathcal{N}$ が AS_{AF} を全て求めていることが要求される．また，定義 4 は，極大な AS_{AF} を PE_{AF} の条件に必要としているが，極大な集合を抽出する操作の部分まで $\mathcal{N}$ が処理するわけではない.
- GE_{AF} を計算する場合は例外的に $S_0 = \emptyset$ としなければならない.

では，図 4.2 の $\mathcal{N}$ を用いて，図 4.1 の $\mathcal{AF}$ において，$S_0 = \{i\}$ を例に，この集合が AS_{AF}, CE_{AF}, SE_{AF} のいずれに該当するか計算を追ってゆく．随時図 4.3 を参照されたい.

初期入力ベクトル $S_{\tau=0}$

まず，$\mathcal{N}$ の計算開始のための初期入力ベクトルが必要である．このベクトルは，調べようとしている集合 $S_0 = \{i\}$ を変換することによって得られる．$S_0 = \{i\}$ をもとに，実際に入力されるベクトルは $[1, 0, 0]$ となるが，このベクトルの作り方に関しては 4.2.2 で定義を導入する.

以下，$\tau = 0$ の入出力計算において，$S_{\tau=0} = S_0$ として話を展開する.

入力層の計算（$\tau = 0$）

今，入力層は初期入力ベクトル $[1, 0, 0]$ を与えられた．全ての入力ニューロン[2]の活性化関数は図 4.3 下段の $\mathcal{N}$ の入力層右部にあるような恒等関数である．活性化関数として恒等

[2]「入力層のユニット」を以下このように呼ぶ.「中間ニューロン」「出力ニューロン」についても同様である.

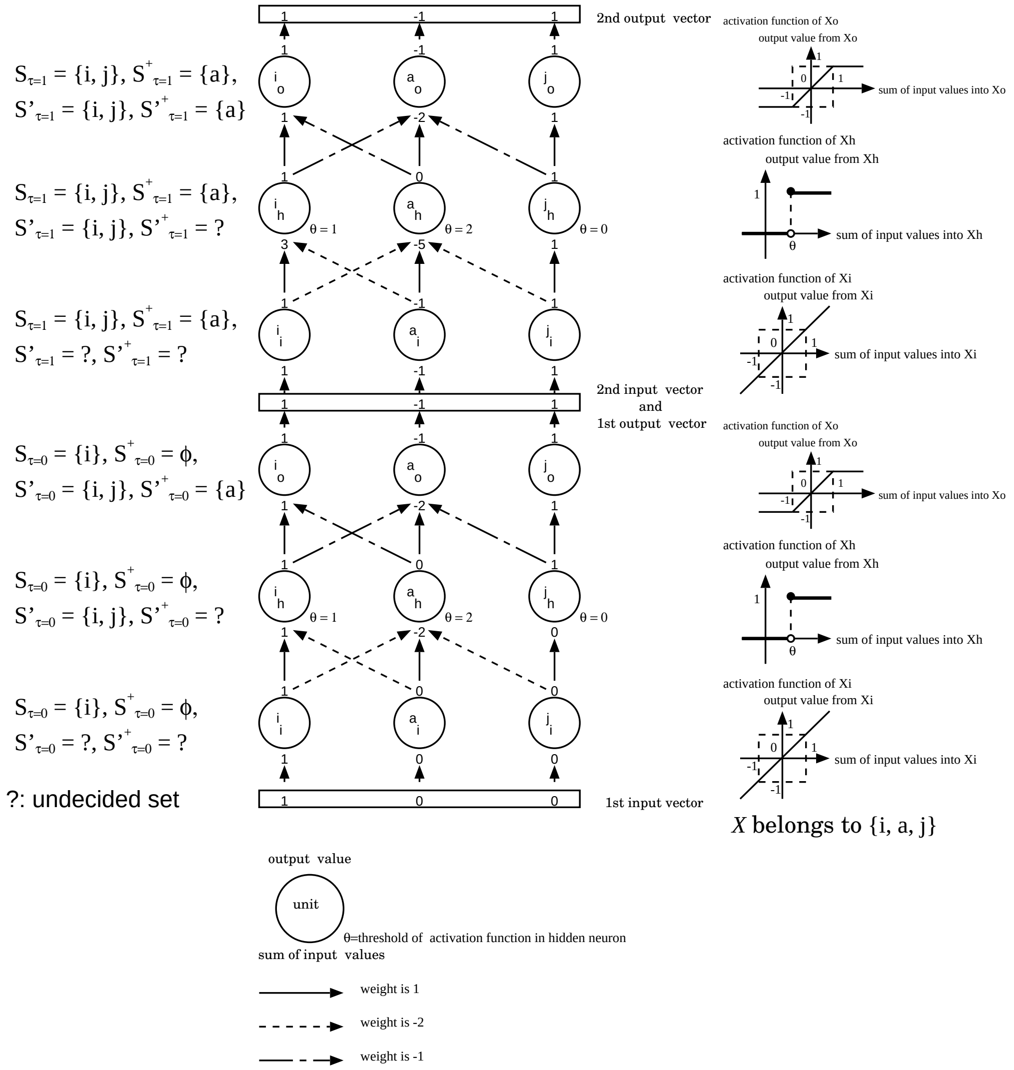

図 4.3: ニューラルネットワーク計算のトレース

関数を用いる理由は，入力層自体では計算を行わせず，入力層で与えられた入力ベクトルの情報を中間層の入力部に受け渡すためである．この活性化関数によって，入力層の出力部で得られるベクトル（[i_iの出力値, a_iの出力値, j_iの出力値]）は $[1, 0, 0]$ となる．

このベクトルをもとに，入力層は $S_{\tau=0} = \{i\}$ であることを判定し入力ニューロン i_i から中間ニューロン i_h へ重み 1 の結線を通して 1 を入力する．そして同時に論証 a が $S_{\tau=0} = \{i\}$ に攻撃されることを判定し，入力ニューロン i_i から中間ニューロン a_h へ重み -2 の結線を通して -2 が入力される．a_i, j_i の出力はそれぞれ 0 となるのでこれらのユニットが他の中間ニューロンに与える入力値は 0 となる．

よって，入力層はベクトル $[1, 0, 0]$ を与えられ，中間層の入力部にベクトル（[i_h の入力総和値（i_i からの入力値 $+a_i$ からの入力値），a_h の入力総和値（i_i からの入力値 $+a_i$ か

らの入力値 $+j_i$ からの入力値），j_h の入力総和値（j_i からの入力値）]）として $[1 \cdot 1 + 0 \cdot (-2), 1 \cdot (-2) + 0 \cdot 1 + 0 \cdot (-2), 0 \cdot 1] = [1, -2, 0]$ を与える．

中間層の計算（$\tau = 0$）

今，中間層は入力層からベクトル $[1, -2, 0]$ を与えられた．i_h, a_h, j_h の活性化関数は，閾値をそれぞれ $\theta_{i_h} = 1, \theta_{a_h} = 2, \theta_{j_h} = 0$ として（閾値の設定の方法の定義については，4.2.1 で述べる），図 4.3 下段の $\mathcal{N}$ の中間層右部のように定める．閾値をそれぞれこのように定める理由であるが，論証 X に対して X が自分自身を擁護（$defends(\{X\}, X)$）できる場合（図 4.1 の例では論証 i）X_h の閾値を 1，X が他の論証から攻撃されているが自分自身を擁護できない場合 X_h の閾値を 2 の自然数倍，X が他の論証から攻撃されていない場合 X_h の閾値を 0 と定め（このような論証の攻撃関係を 4.2 で uni-attacks, bi-attacks という言葉で形式的に定義する），それぞれ先のようなステップ関数を活性化関数を用いることによって，S に擁護される論証の集合に属する論証 X に対し X_h に 1 を出力させ，そうでない論証 Y に対し，Y_h に 0 を出力させ，結果としてこれら二種の論証を二値として判定できる．この活性化関数によって中間層の出力部で得られるベクトル（[i_hの出力値, a_hの出力値, j_hの出力値]）は $[1, 0, 1]$ となる．

このベクトルをもとに，中間層は i と j がそれぞれ $S_{\tau=0}$ に擁護される（$S'_{\tau=0} = \{i, j\}$）ことを判定し，中間ニューロン i_h から出力ニューロン i_o へ，また中間ニューロン j_h から出力ニューロン j_o へ，それぞれ重み 1 の結線を通して 1 を入力する．そして同時に $S'_{\tau=0} = \{i, j\}$ の要素である i, j に攻撃される論証が a である（$S'^{+}_{\tau=0} = \{a\}$）ことを出力層のユニットの活性化関数（後述）を利用して出力層に判定させるために，中間ニューロン i_h から出力ニューロン a_o へ，中間ニューロン j_h から出力ニューロン a_o へ，それぞれ重み -1 の結線を通して -1 が入力される．a_h の出力は 0 となるので，このユニットが他の出力ニューロンに与える入力値は 0 となる．

さらに，中間層のユニットの活性化関数と閾値の機能は，$S_{\tau=0}$ に属する論証から攻撃される論証 X に対する X_h の出力を 0 とする役割を持つ．すなわち 1 を出力している i_h, j_h に対し，$S'_{\tau=0}$ が無衝突であることを保証している．

よって，中間層がベクトル $[1, -2, 0]$ を与えられ，出力層の入力部に与えるベクトル（[i_oの入力総和値, a_oの入力総和値, j_oの入力総和値]）は $[1, -2, 1]$ となる．

出力層の計算（$\tau = 0$）

今，出力層は中間層からベクトル $[1, -2, 1]$ を与えられた．各出力ニューロンの活性化関数は，全て図 4.3 下段の $\mathcal{N}$ の出力層右部のように定める．このような活性化関数を定

める理由であるが，中間層に，入力総和値が正である出力ニューロン X_o の出力を 1 として $X \in S'_{\tau=0}$ と解釈させ，入力総和値が負である出力ニューロン Y_o の出力を -1 として $Y \in S'^{+}_{\tau=0}$ と解釈させ，入力総和値が 0 である出力ニューロン Z_o の出力を 0 として $Z \in S'^{+}_{\tau=0}$ でも $Z \in S'^{+}_{\tau=0}$ でもないと解釈させるためである．結果として，出力層の出力部で得られるベクトル（[i_oの出力値, a_oの出力値, j_oの出力値]）は $[1,-1,1]$ となる．

よって，この $\mathcal{N}$ の $[1,0,0]$ という入力層への入力ベクトルに対し，出力層から読み取られる出力ベクトルは $[1,-1,1]$ となる．

この出力ベクトルにより，出力層は論証 a を $S'_{\tau=0}$ に攻撃される論証であると判定していること（a_o が -1 を出力していることから）を意味する．出力層は $S'_{\tau=0} = \{i,j\}$ であることを，中間層からの入力によって判定している．

出力ベクトルを入力層へ入力（$\tau = 0 \to \tau = 1$）

今，図 4.2 の $\mathcal{N}$ 単体が $S = \{i\}$ を入力として，$S'_{\tau=0} = \{i,j\}$，$S'^{+}_{\tau=0} = \{a\}$ を計算した．

次に，$\mathcal{N}$ の出力層を同じ $\mathcal{N}$ の入力層に接続し（図 4.3 を参照），ここで得られた $S'_{\tau=0}, S'^{+}_{\tau=0}$ をそれぞれ次の入力の $S_{\tau=1}, S^{+}_{\tau=1}$ として入出力計算を繰り返す．

よって，出力層が入力層に与える入力ベクトルは $[1,-1,1]$ となる（$S_{\tau=1} = \{i,j\}, S^{+}_{\tau=1} = \{a\}$）．

入力層の計算（$\tau = 1$）

今，入力層は出力層から入力ベクトル $[1,-1,1]$ を与えられた．入力層の出力部で得られるベクトルは $[1,-1,1]$ となる．このベクトルをもとに，入力層は $S_{\tau=1} = \{i,j\}$ であることを判定し，入力ニューロン i_i から中間ニューロン i_h へ，入力ニューロン j_i から中間ニューロン j_h へそれぞれ重み 1 の結線を通して 1 が入力される．そして同時に $S_{\tau=1} = \{i,j\}$ に a が攻撃されていることを判定し，入力ニューロン i_i から中間ニューロン a_h へ，入力ニューロン j_i から中間ニューロン a_h へそれぞれ重み -2 の結線を通して -2 が入力される．さらに，すでに入力層で与えられたベクトルから $S^{+}_{\tau=1} = \{a\}$ であることを判定し，入力ニューロン a_i から中間ニューロン a_h へ重み 1 の結線を通して -1 が入力される．また，$S^{+}_{\tau=1} = \{a\}$ に論証 i が攻撃されている（$attacks(S^{+}_{\tau=1}, i)$）ことを判定して，入力ニューロン a_i から中間ニューロン i_h へ重み -2 の結線を通して 2 が入力される．

よって，入力層がベクトル $[1,-1,1]$ を与えられ，中間層の入力部に与えるベクトルは $[3,-5,1]$ となる．

中間層の計算（$\tau = 1$）

今，中間層は入力層からベクトル $[3, -5, 1]$ を与えられた．i_h, a_h, j_h の活性化関数の閾値は $\tau = 0$ の入出力時と同じであり，活性化関数は図 4.3 上段の $\mathcal{N}$ の中間層右部のように定める．よって，中間層の出力部で得られるベクトルは $[1, 0, 1]$ となる．

このベクトルをもとに，中間層は i と j がそれぞれ S に擁護される（$S'_{\tau=1} = \{i, j\}$）ことを判定し，中間ニューロン i_h から出力ニューロン i_o へ，また中間ニューロン j_h から出力ニューロン j_o へ，それぞれ重み 1 の結線を通して 1 を入力する．そして同時に $S'_{\tau=1} = \{i, j\}$ に攻撃される論証が a である（$S'^{+}_{\tau=1} = \{a\}$）ことを出力層に判定させるために，中間ニューロン i_h から出力ニューロン a_o へ，中間ニューロン j_h から出力ニューロン a_o へそれぞれ重み -1 の結線を通して -1 を入力する．a_h の出力は 0 となるので，このユニットが他の出力ニューロンに与える入力値は 0 となる．

よって，中間層がベクトル $[3, -5, 1]$ を与えられ，出力層の入力部に与えるベクトルは $[1, -2, 1]$ となる．

また，$\tau = 0$ と同様，この時点で $S'_{\tau=1}$ が無衝突であることが保証されている．

出力層の計算（$\tau = 1$）

今，出力層は中間層からベクトル $[1, -2, 1]$ を与えられた．各出力ニューロンの活性化関数は，全て図 4.3 上段の $\mathcal{N}$ の出力層右部のように定める．よって，出力層の出力部で得られるベクトル（$[i_o$の出力値$, a_o$の出力値$, j_o$の出力値$]$）は $[1, -1, 1]$ となる．

よって，この $\mathcal{N}$ が $[1, -1, 1]$ という入力層への入力ベクトルに対し，出力層から読み取られる出力ベクトルは $[1, -1, 1]$ となる．

この出力ベクトルは，出力層が論証 a を $S'_{\tau=1}$ に攻撃される論証であると判定している（a_o が -1 を出力していることから）ことに相当する．

今，図 4.2 の $\mathcal{N}$ 単体が $S_{\tau=1} = \{i, j\}, S^{+}_{\tau=1} = \{a\}$ を入力として，$S'_{\tau=1} = \{i, j\}$，$S'^{+}_{\tau=1} = \{a\}$ を計算した．

ここで，この 2 回目（$\tau = 1$）の入出力計算において入力ベクトルと出力ベクトルが一致していることに着目されたい．これは $\mathcal{N}$ が，ある関数の不動点を求めたことに相当し，$\mathcal{N}$ 上での計算はここで停止する（以降，同様に入出力を繰り返しても得られる出力ベクトルは常に同じである）．

AS_{AF} と 4 つの意味論の判定

これまでの $\mathcal{N}$ の計算によって，$S_{\tau=0} = \{i\}$ に対して $S'_{\tau=0} = \{i, j\}, S'^{+}_{\tau=0} = \{a\}$，$S_{\tau=1} = \{i, j\}$ に対して $S'_{\tau=1} = \{i, j\}, S'^{+}_{\tau=1} = \{a\}$ であることが求められた．また，$\{i\}, \{i, j\}$ が無

衝突であるということも判定されている．

S が AS_{AF} であるための条件は，3 章の定義に従えば $conflict\text{-}free(S)\&\forall X(X \in S \to X \in S')$ であり，明らかに $\{i\}, \{i,j\}$ がこの条件を満足させており，$\{i\}, \{i,j\}$ が AS_{AF} であるということがわかる．

さらに CE_{AF} であるための条件は，$conflict\text{-}free(S)\&\forall X(X \in S \leftrightarrow X \in S')$ であり，$\{i\}$ がこの条件を満足させず，$\{i,j\}$ がこの条件を満足させることから，$\{i\}$ が CE_{AF} ではなく，$\{i,j\}$ が CE_{AF} であることがわかる．

加えて SE_{AF} であるための条件は，$\forall X(X \notin S \leftrightarrow attacks(S,X))$ であり，$S_{\tau=1} = \{i,j\}$ に対して $S'_{\tau=1} = \{i,j\}, S'^{+}_{\tau=1} = \{a\}$ であることから，$\{i,j\}$ が SE_{AF} であることを判定できる．

よって，AR の部分集合全てを入力 S として計算すれば，AS_{AF} が全て求まるので，その中から極大なものを選び取ることにより PE_{AF} がわかる．

GE_{AF} については，例外的に，$S = \emptyset$ を初期入力集合として入力ベクトルに変換し，安定状態における $S'_{\tau=n}$（n は安定状態におけるタイムラウンド）をその $\mathcal{AF}$ における GE_{AF} として得ることができる（ある論証の集合 $S \subseteq AR$ を入力ベクトルに変換して，S が GE_{AF} に該当するかどうかを判定するのではない点に注意されたい）．

ニューラルネットワーク議論の直感的説明のまとめ

この節では，任意の $\mathcal{AF}$ より変換された $\mathcal{N}$ に，ある集合 S を入力とし，S' と S'^{+} の計算を反復的に行い，いかに $AS_{AF}, CE_{AF}, PE_{AF}, SE_{AF}, GE_{AF}$ の判定を行うかという直感的説明を与えた．

では，4.2 以降で，$\mathcal{AF}$ を $\mathcal{N}$ に変換するためのアルゴリズムや，入出力のための計算アルゴリズム，$AS_{AF}, CE_{AF}, PE_{AF}, SE_{AF}, GE_{AF}$ の判定アルゴリズムの厳密な定義を導入する．

4.2　議論の意味論を計算するニューラルネットワーク

この節では，議論の意味論を計算するニューラルネットワークの構成の仕方を定義する．具体的には，議論フレームワークをニューラルネットワークに変換するアルゴリズムを与え，そのニューラルネットワークが行うべき計算の方法を定義する．この変換アルゴリズムと計算の方法は [6] の “neural argumentation algorithm” よりヒントを得ているが，意味論を計算するという今回の目的に併せて変更を加えている．

変換アルゴリズムを導入する前に，表記の便宜上以下に 2 つの新しい論証間の関係を定義する．

定義 9. (**Uni-attacks**)
A と B を論証とする ($A \in AR$ かつ $B \in AR$) とき，$\mathit{uni\text{-}attacks}(A, B) \Leftrightarrow (A, B) \in \mathit{attacks}$ & $(B, A) \notin \mathit{attacks}$．ただし A と B が一致するような場合も，$\mathit{uni\text{-}attacks}(A, A)$ であるとする．

定義 10. (**Bi-attacks**)
A と B を論証とする ($A \in AR$ かつ $B \in AR$) とき，$\mathit{bi\text{-}attacks}(A, B) \Leftrightarrow (A, B) \in \mathit{attacks}$ & $(B, A) \in \mathit{attacks}$．ただし A と B は同一の論証ではない（すなわち A と B が一致するような場合，$\mathit{bi\text{-}attacks}(A, A)$ ではない点に注意）．

直感的には，$\mathit{uni\text{-}attacks}(A, B)$ を「A は B を一方的に攻撃する」，$\mathit{bi\text{-}attacks}(A, B)$ を「A と B は互いに攻撃し合う関係にある」と読める．

例 2. 図 *3.1* の議論フレームワークでは，論証間の関係 *uni-attacks, bi-attacks* として以下が得られる：

- $\mathit{uni\text{-}attacks} = \{(a, b),\ (c, b),\ (d, e),\ (e, e)\}$
- $\mathit{bi\text{-}attacks} = \{(c, d), (d, c)\}$

4.2.1 変換アルゴリズム

以下のアルゴリズムでは，対象とする議論フレームワーク $\mathcal{AF}$ から変換されたニューラルネットワークを $\mathcal{N}$ と表記する．

1. $AR = \{\alpha_1, \cdots, \alpha_n\}$ であるような議論フレームワーク $\mathcal{AF} = < AR, \mathit{attacks} >$ を与える．

2. AR の中の各論証 $\alpha_1, \ldots, \alpha_n$ に対して，ニューラルネットワーク $\mathcal{N}$ の入力層を構築し，入力層の中の k 番目 $(1 \leq k \leq n)$ のユニット α_{ki} が AR の中の論証 α_k に対応するように $\mathcal{N}$ へユニットを追加する．同様に $\mathcal{N}$ の中間層と出力層を構築し，中間層のユニット α_{kh} と出力層のユニット α_{ko} が AR の中の論証 α_k に対応するように $\mathcal{N}$ へユニットを追加する．

3. AR の中の各論証 $\alpha_k (1 \leq k \leq n)$ に対して以下を行う：

   (a) 入力ニューロン α_{ki} から中間ニューロン α_{kh} へ接続を行い，接続重み $W_{\alpha_{ki}\alpha_{kh}} = 1$ を設定する；

   (b) 中間ニューロン α_{kh} から出力ニューロン α_{ko} へ接続を行い，接続重み $W_{\alpha_{kh}\alpha_{ko}} = 1$ を設定する．

4. 各攻撃関係 $(\alpha_k, \alpha_l) \in attacks$ に対して以下を行う：

   (a) 中間ニューロン α_{kh} から出力ニューロン α_{lo} へ接続を行い，接続重み $W'_{\alpha_{kh}\alpha_{lo}} = -1$ を設定する；

   (b) 入力ニューロン α_{ki} から中間ニューロン α_{lh} へ接続を行い，接続重み $W''_{\alpha_{ki}\alpha_{lh}} = -2$ を設定する．

5. 各論証 α_k に対し, $uni\text{-}attacks(\alpha_l, \alpha_k)$ を満たすような論証 $\alpha_l \in AR$（α_k を一方的に攻撃するような α_l）の数を n_u，$bi\text{-}attacks(\alpha_m, \alpha_k)$ を満たすような論証 $\alpha_m \in AR$（α_k と互いに攻撃し合うような α_m）の数を n_b とする．

   (a) $n_u = 0$ かつ $n_b = 0$ を満たすとき，中間ニューロン α_{kh} の活性化関数の閾値を $\theta_{\alpha_{kh}} = 0$ に設定する；

   (b) $n_u = 0$ かつ $n_b \neq 0$ を満たすとき，α_{kh} の活性化関数の閾値を $\theta_{\alpha_{kh}} = 1$ に設定する；

   (c) $n_u \neq 0$ を満たすとき，α_{kh} の活性化関数の閾値を $\theta_{\alpha_{kh}} = 2n_u$ に設定する．

6. $\mathcal{N}$ の各中間ニューロンの活性化関数として以下のような $h(x)$ を設定する：$x \geq \theta$ のとき $h(x) = 1$，$x < \theta$ のとき $h(x) = 0$.

7. $\mathcal{N}$ の各入力ニューロンの活性化関数として $g(x) = x$（恒等関数）を設定する．

8. $\mathcal{N}$ の各出力ニューロンの活性化関数として以下のような $f(x)$ を設定する：$x \geq 1$ のとき $f(x) = 1$，$1 > x > -1$ のとき $f(x) = x$，$x \leq -1$ のとき $f(x) = -1$.

なお，$\mathcal{N}$ の全てのユニットの出力関数は恒等関数である．

例 3. 図 *4.4* に，この変換アルゴリズムにより図 *3.1* の $\mathcal{AF}$ から変換された $\mathcal{N}$ を与える．図 *4.4* の $\mathcal{N}$ では，各中間ニューロンの活性化関数の閾値は $\theta_{a_h} = 0$，$\theta_{b_h} = 4$，$\theta_{c_h} = 1$，$\theta_{d_h} = 1$，$\theta_{e_h} = 4$ である．

4.2.2 意味論を計算するための計算規則

次に 4.2.1 で定義した変換アルゴリズムによって得られたニューラルネットワークが行う計算規則を定義する．以下の定義では，$\mathcal{AF} =< AR, attacks >$ が議論フレームワーク，$\mathcal{N}$ が $\mathcal{AF}$ より変換アルゴリズムによって得られたニューラルネットワークを意味するも

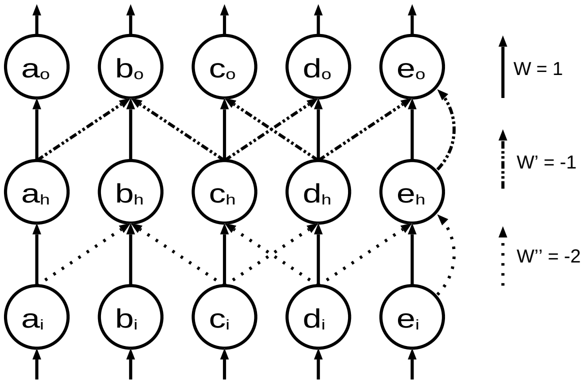

図 4.4: 図 3.1 の議論フレームワークから変換アルゴリズムにより得られるニューラルネットワーク

のとする．まず，$\mathcal{N}$ の反復計算の理解を容易にするためにタイムラウンドの概念を以下に定義する．

定義 11. **（タイムラウンド）**
$\mathcal{N}$ の計算において，$\mathcal{N}$ の入力層に入力ベクトルを与えてから出力層より出力ベクトルが出力されるまでの区間をタイムランドと呼び，τ の記号によってこれを表す．初期値は 0 であり，前のタイムラウンドの出力ベクトルが $\mathcal{N}$ の入力ベクトルに与えられる．出力層から入力ベクトルが与えられる毎にこの値をインクリメントする．

$\mathcal{N}$ は $S \subseteq AR$ がどの意味論による集合と一致するかを判定する．計算の開始のために，まずこの S を $\mathcal{N}$ が受け取るべき初期入力ベクトルとして，以下の定義に従い変換する．

定義 12. **（$S \subseteq AR$ に対する初期入力ベクトル）**
$S \subseteq AR$ を $\mathcal{N}$ の初期入力ベクトルとして以下のように変換を行う．

- 各 $a_k \in S$ に対して，$\mathcal{N}$ の入力ニューロン a_{ki} に 1 を入力する．
- 各 $a_l \notin S$ に対して，$\mathcal{N}$ の入力ニューロン a_{li} に 0 を入力する．

例 4. 今 $AR = \{A, B, C, D, E\}$ と $S = \{B, D, E\}$ を仮定する．この時，定義 *12* に基づき $\mathcal{N}$ に入力されるべき入力ベクトルは $[0, 1, 0, 1, 1]$ である．ただし，$\mathcal{N}$ の入力層の各入力ニューロンは，左から A_i, B_i, C_i, D_i, E_i の順番で配置されているものとする．

定義 13. **（$\mathcal{N}$ の安定状態）**
$\mathcal{N}$ に入力された入力ベクトルと $\mathcal{N}$ から出力された出力ベクトルが同じ τ のもとで一致していた場合，$\mathcal{N}$ の計算が安定状態にあるという．

以下の定義では，$\mathcal{N}$ の計算結果に対して，$\mathcal{AF}$ の各意味論への解釈を与える．

定義 14.（$\mathcal{N}$ における Admissible Set）
各 $a \in S$ に対して，$\mathcal{N}$ が定義 *12* により S より変換された入力ベクトルを入力され，同一タイムラウンドにおいて $\mathcal{N}$ の出力ニューロン a_o が *1* を出力したとき，$S \subseteq AR$ は $\mathcal{N}$ における *Admissible Set* である．

上の定義は以下のような解釈に置き換えることができる：定義 12 により S から変換した初期入力ベクトルを $[i_1, i_2, \ldots, i_n]$（n は AR に含まれる要素（論証）数，または $\mathcal{N}$ の入力ニューロンの数である），安定状態での入出力ベクトルを $[c_1, c_2, \ldots, c_n]$ とした時，$\forall i_k(1 \leq k \leq n)(i_k = 1 \rightarrow c_k = 1) \Leftrightarrow S$ は $\mathcal{N}$ における Admissible Set である．

ただし，定義 14 に従うと，$\mathcal{N}$ は $\emptyset$ を $\mathcal{N}$ における Admissible Set であると判定しない．しかし定理 1 より，任意の $\mathcal{AF}$ は $\emptyset$ を常に AS_{AF} として持つことが保証されている．故に，$S = \emptyset$ の時に限り，この定理を用いて $\mathcal{N}$ は計算に依らず $S = \emptyset$ を $\mathcal{N}$ における Admissible Set であるとする．

以下,「$\mathcal{N}$ における Admissible Set」を AS_N と表記する．

例 5. 図 *4.4* において，$S_1 = \{a\}$ に対する初期入力ベクトル $[1, 0, 0, 0, 0]$ を $\mathcal{N}$ に与え，$\tau = 0$ で出力ベクトル $[1, -1, 0, 0, 0]$ が得られる（以下では $\mathcal{N}$ のこのような入出力の写像 "$\mapsto_{\tau=n}$" を，$[1, 0, 0, 0, 0] \mapsto_{\tau=0} [1, -1, 0, 0, 0]$ のように用い,「$\tau = 0$ で，$\mathcal{N}$ が入力ベクトル $[1, 0, 0, 0, 0]$ に対する出力ベクトル $[1, -1, 0, 0, 0]$ を出力する」を意味するものとして用いる）．定義 *14* により，$S_1 = \{a\}$ は AS_N である．なぜなら，$\tau = 0$ で a_i が入力値 1 を与えられ，この τ で a_o が 1 を出力しているからである．すなわち，入力値 1 を与えられた任意の入力ニューロンに対応する出力ニューロンが同一タイムラウンドで 1 を出力している．

一方，$S_2 = \{a, c, d\}$ に対する入力ベクトル $[1, 0, 1, 1, 0]$ に対して，$\mathcal{N}$ は入出力 $[1, 0, 1, 1, 0] \mapsto_{\tau=0} [1, -1, 0, 0, 0]$ を行う．よって，$\tau = 0$ で c_o と d_o が 1 を出力しないので，$S_2 = \{a, c, d\}$ は AS_N ではない．

定義 15.（$\mathcal{N}$ における Complete Extension）
以下の 2 つの条件を満たすとき，$S \subseteq AR$ は $\mathcal{N}$ における *Complete Extension* である：

- 各 $a \in S$ に対し，$\tau = 0$ で S に対する入力ベクトルを $\mathcal{N}$ に与えたとき，出力ニューロン a_o が $\mathcal{N}$ の安定状態において 1 を出力する．

- $\tau = 0$ で入力ニューロン a_i が 1 を与えられたとき，$a \in S$ であるような出力ニューロン a_o のみが，$\mathcal{N}$ の安定状態において 1 を出力する．

言い換えると，S の要素に対応する出力ニューロンのみが $\mathcal{N}$ の安定状態で 1 を出力する．

定義 15 は以下のような解釈を与える：定義 12 により S から変換した初期入力ベクトルを $[i_1, i_2, \ldots, i_n]$，安定状態での入出力ベクトルを $[c_1, c_2, \ldots, c_n]$ とした時，$\forall i_k (1 \leq k \leq n)(i_k = 1 \leftrightarrow c_k = 1) \Leftrightarrow S$ は $\mathcal{N}$ における Complete Extension である．

以下,「$\mathcal{N}$ における Complete Extension」を CE_N と表記する．

例 6. 図 *4.4* の $\mathcal{N}$ において，$S_1 = \{a\}$ に対する入力ベクトル $[1,0,0,0,0]$ を与えたとき，$\mathcal{N}$ は入出力計算は $[1,0,0,0,0] \mapsto_{\tau=0} [1,-1,0,0,0] \mapsto_{\tau=1} [1,-1,0,0,0]$ となる（$\tau = 1$ で $\mathcal{N}$ は安定状態になる）．結果として，$S_1 = \{a\}$ は CE_N であると判定される．また，$S_3 = \{d\}$ に対する入力ベクトル $[0,0,0,1,0]$ に対する計算は $[0,0,0,1,0] \mapsto_{\tau=0} [1,-1,-1,1,-1] \mapsto_{\tau=1} [1,-1,-1,1,-1]$ である．故に，$S_3 = \{d\}$ は CE_N ではない．なぜなら，a_o を例に取ると，安定状態で 1 を出力しているが，$\tau = 0$ で a_i が 0 を入力されており，定義 *15* の 2 番目の条件を満足しないためである．

定義 16. （$\mathcal{N}$ における Preferred Extension）
$\mathcal{N}$ における *Preferred Extension* は定義 *15* に従い判定される AS_N の中で，包含関係に関して極大なものである．

以下,「$\mathcal{N}$ における Preferred Extension」を PE_N と表記する．

例 7. 図 *4.4* の $\mathcal{N}$ を例に取ると，AR の任意の部分集合を入力として計算した結果，AS_N として $\emptyset$, $\{a\}$, $\{c\}$, $\{d\}$, $\{a,c\}$, $\{a,d\}$ を得る．この中で包含関係に関して極大となるのは，$\{a,c\}$ と $\{a,d\}$ であるので，この 2 つが PE_N である．

定義 17. （$\mathcal{N}$ における Stable Extension）
$S_s = \{\ a \in AR \ \mid\ \mathcal{N}$ の収束状態において a_o が 1 を出力する $\}$，$S_s^+ = \{\ b \in AR \ \mid\ \mathcal{N}$ の収束状態において b_o が -1 を出力する $\}$ とし，任意の $S_i \subseteq AR$ に対する入力ベクトルを $\mathcal{N}$ に与えたとき，$S_i = S_s$ かつ $S_s \cup S_s^+ = AR$ を満たすとき，S_i は $\mathcal{N}$ における *Stable Extension* である．

定義 17 の別表現を以下のように与える：定義 12 により $S \subseteq AR$ から変換された初期入力ベクトルを $[i_1, i_2, \ldots, i_n]$，安定状態での入出力ベクトルを $[c_1, c_2, \ldots, c_n]$ とした時，$\forall i_k (1 \leq k \leq n)(i_k = 1 \rightarrow c_k = 1) \& \forall i_l (1 \leq l \leq n)(i_l = 0 \rightarrow c_l = -1) \Leftrightarrow S$ は $\mathcal{N}$ において Stable Extension である．

以下,「$\mathcal{N}$ における Stable Extension」を SE_N と表記する．

例 8. 今 $S_4 = \{a,c\}$ と $S_5 = \{a,d\}$ がそれぞれ SE_N であるかどうかを調べる．$AR(=\{a,b,c,d,e\}) \neq S_{4s}(=\{a,c\}) \cup S_{4s}^+(=\{b,d\})$ であることから，$S_4 = \{a,c\}$ は SE_N ではない．一方で，$AR(=\{a,b,c,d,e\}) = S_{5s}(=\{a,d\}) \cup S_{5s}^+(=\{b,c,e\})$ であるために，$S_5 = \{a,d\}$ は SE_N であると判定できる．

定義 18. **（$\mathcal{N}$ における Grounded Extension）**
$\tau = 0$ で，$\mathcal{N}$ の全ての入力ニューロンに 0 を入力したとき（これは $S_i = \emptyset$ を入力ベクトルに変換したことに等しい），$S_g = \{\ a \in AR \ \mid\ \mathcal{N}$ の収束状態において出力ニューロン a_o が 1 を出力する $\}$ と定義される S_g は $\mathcal{N}$ における *Grounded Extension* である．

以下,「$\mathcal{N}$ における Grounded Extension」を GE_N と表記する．

例 9. 図 *4.4* における GE_N を判定しよう．$\tau = 0$ で $\mathcal{N}$ にベクトル $[0,0,0,0,0]$ を与えたときの入出力計算は，$[0,0,0,0,0] \mapsto_{\tau=0} [1,-1,0,0,0] \mapsto_{\tau=1} [1,-1,0,0,0]$ となる． 安定状態で a_o が 1 を出力しているので，$\mathcal{N}$ は GE_N として $S_g = \{a\}$ を判定している．

著者らは [11] において，既に GE_N の計算方法を提案しているが，そちらでは GE_N の計算のためだけに特別な活性化関数の閾値の変化の条件をタイムラウンド毎に $\mathcal{N}$ へ導入しており，用いている $\mathcal{AF}$ から $\mathcal{N}$ への変換アルゴリズムも安定状態までに $\mathcal{N}$ が不要なタイムラウンド数を要するモデルである．本論文で，4.2.1 において提案される変換アルゴリズムは，他の意味論の計算も行え，安定状態に達するまでの入出力の計算の回数もより改善している．

4.3 $\mathcal{N}$ の計算の健全性

この章で導入した変換アルゴリズムにより得られる $\mathcal{N}$ による計算が，3 章で定義した意味論との一致性に関して健全な計算を行うという定理を以下に与える．

定理 2. **（$\mathcal{N}$ の計算の健全性）**

1. $\mathcal{N}$ において *Admissible Set* ならば，$\mathcal{AF}$ において *Admissible Set* である．

2. $\mathcal{N}$ において *Complete Extension* ならば，$\mathcal{AF}$ において *Complete Extension* である．

3. $\mathcal{N}$ において *Preferred Extension* ならば，$\mathcal{AF}$ において *Preferred Extension* である．

4. $\mathcal{N}$ において *Stable Extension* ならば，$\mathcal{AF}$ において *Stable Extension* である．

5. $\mathcal{N}$ において *Grounded Extension* ならば，$\mathcal{AF}$ において *Grounded Extension* である．

この定理に対する証明を以下に与える．

Proof. 以下の証明では，便宜のため以下のように集合を定義する：

- $S_{\tau=n} = \{\ a \in AR \mid \tau = n(n \geq 0)$ にて $\mathcal{N}$ の入力ニューロン a_i は入力として 1 を与えられる $\}$
- $S'_{\tau=n} = \{\ a \in AR \mid \tau = n(n \geq 0)$ にて $\mathcal{N}$ の出力ニューロン a_o は 1 を出力する $\}$（出力層で得られた出力ベクトルを次のタイムラウンドの $\mathcal{N}$ の入力ベクトルとして与えるので $S'_{\tau=n} = S_{\tau=n+1}$ である）
- $S'^{+}_{\tau=n} = \{\ a \in AR \mid \tau = n(n \geq 0)$ にて $\mathcal{N}$ の出力ニューロン a_o は -1 を出力する $\}$

1. 定理 2 の 1 の証明：$S'_{\tau=0}$ は，AR 内の任意の論証に攻撃されない ($n_u = 0$ かつ $n_b = 0$ であるような) 論証か，$S_{\tau=0}$ に含まれかつ自分自身を擁護できる ($n_u = 0$ かつ $n_b \neq 0$ であるような) 論証のいずれか要素として含む． すなわち結果として，$S'_{\tau=0}$ の中の全ての要素は $S_{\tau=0}$ に擁護されるような論証であり（$\mathcal{AF}$ 内で攻撃されない論証は「$\emptyset$ に擁護される」という点に注意），$S'_{\tau=0}$ は無衝突である．また，$S'_{\tau=1}$ の中の全ての要素は $S_{\tau=1}$ に擁護されるような論証である．なぜなら $S'_{\tau=1} = S_{\tau=1}(= S'_{\tau=0}) \cup \{\ a \in AR \mid a$ は $\mathcal{AF}$ において $S_{\tau=1}(= S'_{\tau=0})$ に擁護される $\}$ かつ $S'_{\tau=1}$ が無衝突であるからである．故に，$S'_{\tau=n(n\geq1)} = \{\ a \in AR \mid a$ は $S_{\tau=n}$に擁護される $\}$ かつ $S'_{\tau=n}$ が無衝突であることが成立する．結果として，もし $S_{\tau=n(n\geq0)} = S'_{\tau=n}$ (すなわち $\mathcal{N}$ は $\tau = n$ で安定状態である) ならば $S_{\tau=n(n\geq0)}$ は $S_{\tau=0}$ の中の論証を全て要素として含むような *Admissible Set* である．これにより定理 2 の 1 が成り立つ．
2. 定理 2 の 2 の証明：安定状態（$S_{\tau=n(n\geq0)} = S'_{\tau=n}$）において，$S'_{\tau=n}$ が無衝突であり $S'_{\tau=n}$ の中の全ての要素が $S_{\tau=n}$ に擁護されるような論証である．これは,「もし $S_{\tau=n}$ に属するような論証があれば，その論証は $S'_{\tau=n}$ に擁護される」かつ「$S'_{\tau=n}$ に擁護されるような論証があれば，その論証は $S_{\tau=n}$ に属する」を意味する．これにより定理 2 の 2 が成り立つ．
3. 定理 2 の 3 の証明：定理 2 の 1 が成り立つことより定理 2 の 3 の成立は自明である．
4. 定理 2 の 4 の証明：定義 17 より，$S_{\tau=n(n\geq0)} = S'_{\tau=n}$ かつ $S'^{+}_{\tau=n} = AR \backslash S'_{\tau=n}$ である．よって $S'^{+}_{\tau=n}$ の全ての要素は $\mathcal{AF}$ において $S'_{\tau=n}$ に属さず，かつ $S'_{\tau=n}$ の中の論証に攻撃されるような論証である．これは $S'_{\tau=n}$ が $\mathcal{AF}$ における SE_{AF} と一致することを意味する．よって定理 2 の 4 が成り立つ．
5. 定理 2 の 5 の証明：$S_{\tau=0} = \emptyset$ のとき $S'_{\tau=0}$ の中の全ての要素は $\emptyset$ に擁護されるような論証の集合（すなわち $\mathcal{AF}$ において，攻撃されないような論証の集合）である．

よって $S'_{\tau=0}$ は $F_{AF}(\emptyset)$ と一致する．さらに $\tau = n (n \geq 1)$ のとき，$S'_{\tau=n}$ の中の全ての論証は $S_{\tau=n}$ に擁護される．故に $F_{AF}(S_{\tau=n}) = S'_{\tau=n}$ が成り立つ．安定状態（$S_{\tau=n(n\geq 0)} = S'_{\tau=n}$）のとき，$F_{AF}(S_{\tau=n})$ は F_{AF} の最小不動点と一致する．以上より，定理 2 の 5 が成り立つ．

□

第5章　おわりに

5.1　まとめ

ニューラルネットワークの記号処理能力を議論の領域において検討しようとする研究は，Garcez らの試みを除き未だ知られていない．本論文では，記号議論の代表的な意味論をを計算する新しいニューラルネットワークの機構を提案し，両者の計算の一致性の証明を与えた．

ニューラルネットワークが記号処理をいかに厳密に捉え説明可能かという点については，現在も様々な研究領域で盛んに論じられているが [9]，未だその能力については限定的または否定的な見解も散在し，多くの分野でニューラルネットワークの能力を疑問視する声も上がっている．このような状況にあって，本論文の結果は，人工ニューラルネットワークが議論という認知現象を捉えることが可能であることを示す有力な証左であるといってよい．一方，ニューラルネットワーク議論を記号議論における対話的証明論 [14] のもとで，記号による対話形式に変換しようとする方法は，著者らの他の論文で与えられた [12]．これはニューラルネットワークによる計算の結果を記号で説明することを可能にするものであるため，しばしば指摘されるニューラルネットワークの説明能力の欠如を補うことになる．したがって，この結果と本論文の結果によって，数学的に厳密な connectionism と symbolism の融合の一つの形を，議論の世界で発現させることができたと言える．

5.2　検討課題

5.2.1　他の意味論に対する計算能力の可能性

本論文では，主に代表的な４つの意味論 Stable Extension, Preferred Extension, Complete Extension, Grounded Extension に焦点を当て，それらに対するニューラルネットワークの計算能力の可能性を論じたが，著者らは [10] にて，M. Caminada によるもう一つの議論の意味論 Semi-stable Extension[5] も今回扱った意味論に加えて，本論文と同じ変換アルゴリズムで生成されたニューラルネットワークにより判定ができることを示している．

しかし，3.3 で述べたように，議論の意味論は無限に定義しうるものである．議論の研究では他にもまだ多くの意味論の定義が成されている [4]．これらの意味論についてもニューラルネットワークが扱える手続きが存在するのか，また意味論のどういった特性をニューラルネットが一般的に扱えると言えるのか，検討の余地が残されている．

本論文の中で対象とした意味論の定義は,「無衝突性」「受理可能性」「許容可能性」といった概念を土台としているものが多い．さらに具体的には主に以下の二点の性質により定義されているといえる：

- 論証の集合に対する所属性
- 論証の集合の上の二項関係

故に，著者はこの 2 点に基づき定義される意味論ならば，本論文で導入したようなニューラルネットワークにより判定可能な能力を有するのではないかという期待を得ている．

5.2.2 学習アルゴリズムの適用可能性

本論文で用いたニューラルネットワークは，ニューラルネットワーク一般に適用されるような学習アルゴリズム（例えばバックプロパゲーションなど）を用いることを意識していない．なぜなら，意味論を計算するという目的においては単隠れ層ネットワークの反復的な入出力計算によって十分に達成できるからである．

しかし，ニューラルネットワークに学習という概念を持ち込むことで，どのような展開が可能になるか興味が尽きないところである．Garcez らはこの学習能力を従来的な記号議論の研究に適用することを提案していたが，本来記号議論にない概念を持ち込んで決定した論証の状態をどう解釈すべきか，という点に検討の余地を残している．

ただし，はじめから特定の意味論を意識したニューラルネットワークのみを用いるのではなく，選び取るべき正しい論証の集合を教師あり学習によってニューラルネットワークのパラメータを調節してゆくことによって，既になされた議論の結果を経験的に学習してニューラルネットワークが自発的に意味論を与えるようなモデルも考えられるのではないだろうか．しかし，一般的な学習アルゴリズムの適用を意識すると，ニューラルネットワークの形態を常に固定する，すなわち変換前の議論フレームワークを固定しなければならず，常に同じ論証数と攻撃関係のもとでしか学習が行えないといった問題が浮かび上がる．そこで，変換アルゴリズムによって生成されたニューラルネットワークを対象に学習を適用するのではなく，変換アルゴリズム自体の重みや閾値設定といったパラメータを対象に学習を適用することによって学習の適用可能性が生まれるのではないかという予見もある．

5.2.3 対話的証明論とニューラルネットワーク議論の関係

本論文では詳細には述べなかったが，議論の研究には対話的証明論によってある意味論のもと正しい論証を選び取る手続きが存在する．Prakken らは [13] において Grounded Extension に基づく論証の状態を決定する対話的手続きを提案し，Vreeswijk らは [17] にて Preferred Extension に属する論証を対話的に判定する試みを提案している．

著者らは [12] において，ニューラルネットワークにより Grounded Extension と一致する対話的証明論の抽出が可能であることを述べている．これは議論フレームワークと，変換アルゴリズムにより生成されたニューラルネットワーク間で，論証の集合と論証間の攻撃関係が損なわれていないことによるもので，ニューラルネットワークの形態は変わってはいるが本論文で導入した変換アルゴリズムから生成されたニューラルネットワークに対しても [12] の対話抽出アルゴリズムが適用可能である．

他の意味論においては，Preferred Extension を除いて対話的証明論がまだ定義されていない．Preferred Extension に関しては，[12] と同様の議論フレームワークと変換されたニューラルネットワークの等価性に着目した方法により，[17] と一致する対話的証明論が構築可能なのは明らかである．対話的証明論が定義されていない他の意味論については，本論文で導入したニューラルネットワークによる意味論の判定の中で，判定するまでの計算過程（ユニットの活性の伝播）が存在することから，これをヒントに対話的証明論と同様の，意味論の解釈と一致する結果を導く対話的手続きを与えられる可能性はある．すなわち，ニューラルネットワークが，記号議論の世界でまだ定義されていない対話的証明論の与え方を示唆できるという期待がある．

謝辞

本研究を進めるにあたって，新潟大学 自然科学研究科 澤村一准教授には，様々な場面で多大なる御鞭撻と熱心な御指導，貴重な場での発表の機会を賜りました．深く感謝申し上げます．

また，日頃よりセミナー等を通じて多くの助言や支援をして下さった新潟大学 澤村・萩原研究室の皆様に心より感謝いたします．

最後に，本研究に関してコメントを寄せて下さった Artur S. d'Avila Garcez 氏をはじめ，学会などの場面で非常に有意義な疑問を投げ掛け，助言を下さった査読者や学会参加者の方々に感謝の意を表します．

関連図書

[1] T.J.M. Bench-Capon. Persuasion in practical argument using value-based argumentation frameworks. *Journal of Logic and Computation*, 13(3):429–448, 2003.

[2] Philippe Besnard and Sylvie Doutre. Characterization of semantics for argument systems. In Didier Dubois, Christopher A. Welty, and Mary-Anne Williams, editors, *KR*, pages 183–193. AAAI Press, 2004.

[3] Philippe Besnard and Sylvie Doutre. Checking the acceptability of a set of arguments. In James P. Delgrande and Torsten Schaub, editors, *NMR*, pages 59–64, 2004.

[4] M. Caminada. Comparing Two Unique Extension Semantics for Formal Argumentation: Ideal and Eager. In *19th Belgian–Dutch Conference on Artificial Intelligence(BNAIC–2007)*. BNAIC, 2007.

[5] Martin Caminada. Semi-stable semantics. In Paul E. Dunne and Trevor J. M. Bench-Capon, editors, *COMMA*, volume 144 of *Frontiers in Artificial Intelligence and Applications*, pages 121–130. IOS Press, 2006.

[6] Artur S. d'Avila Garcez, Dov M. Gabbay, and Luis C. Lamb. Value-based argumentation frameworks as neural-symbolic learning systems. *Journal of Logic and Computation*, 15(6):1041–1058, 2005.

[7] P.M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logics programming and n-person games. *Artificial Intelligence*, 77:321–357, 1995.

[8] J. A. Fodor and Z. W. Pylyshyn. Connectionism and cognitive architecture. *A critical analysis, Cognition*, 28(1-2):3–71, 1988.

[9] Arun Jagota, Tony Plate, Lokendra Shastri, and Ron Sun. Connectionist symbol processing: Dead or alive? *Neural Computing Surveys*, 2:1–40, 1999.

[10] Wataru Makiguchi. Studies on Neural Network Argumentation. Master's thesis, Graduate School of Science and Technology, Niigata University, Niigata, Japan, January 2009. `http://www.cs.ie.niigata-u.ac.jp/~makiguti/masterpaper_makiguchi.pdf`.

[11] Wataru Makiguchi and Hajime Sawamura. A Hybrid Argumentation of Symbolic and Neural Net Argumentation (Part I). In Rahwan et al. [15], pages 197–215.

[12] Wataru Makiguchi and Hajime Sawamura. A Hybrid Argumentation of Symbolic and Neural Net Argumentation (Part II). In Rahwan et al. [15], pages 216–233.

[13] H. Prakken and G. Sartor. Argument-based extended logic programming with defeasible priorities. *J. of Applied Non-Classical Logics*, 7(1):25–75, 1997.

[14] H. Prakken and G. Vreeswijk. Logical systems for defeasible argumentation. In *In D. Gabbay and F. Guenther, editors, Handbook of Philosophical Logic*, pages 219–318. Kluwer, 2002.

[15] Iyad Rahwan, Simon Parsons, and Chris Reed, editors. *Argumentation in Multi-Agent Systems, 4th International Workshop, ArgMAS 2007, Honolulu, HI, USA, May 15, 2007, Revised Selected and Invited Papers*, volume 4946 of *Lecture Notes in Computer Science*. Springer, 2008.

[16] A Tarski. A lattice-theoretical fixpoint theorem and its application. *Pacific Journal of Mathmatics*, 5:285–309, 1955.

[17] Gerard Vreeswijk and Henry Prakken. Credulous and sceptical argument games for preferred semantics. In Manuel Ojeda-Aciego, Inman P. de Guzmán, Gerhard Brewka, and Luís Moniz Pereira, editors, *JELIA*, volume 1919 of *Lecture Notes in Computer Science*, pages 239–253. Springer, 2000.