

氏名(本籍)	新出 尚之 (京都府)
学位の種類	博士(情報科学)
学位記番号	博論第140号
学位授与年月日	平成19年6月28日
学位授与の要件	学位規則第4条第2項該当 人間文化研究科
論文題目	BDI logic の sequent calculus による演繹体系と マルチエージェント環境への拡張
論文審査委員	(委員長) 教授 城 和 貴 教授 加古 富志雄 教授 落合 豊 行 准教授 高田 司 郎 (近畿大学理工学部)

## 論文内容の要旨

本論文では、人間やエージェントの自律的振る舞いを記述するための時相論理体系である BDI logic に対して、sequent calculus による演繹体系を与え、その健全・完全性を示している。また、Bratman の「意図の理論」において、合理的な意思決定が持つ性質として記述されている、心的状態の整合性公理を、著者の体系において扱えることを示している。さらに、その体系をマルチエージェント環境を対象として拡張し、相互信念などの扱いについて示している。近年では、自律エージェントシステムの研究において、エージェントに「意図」の概念を明示的に持たせ、その持続と更新によりエージェントの行為決定を制御する方式が有力視されている。自律エージェントは、人間と同様に、動的に変化する周囲の状況を認識し、目的の達成に向けた一貫的で合理的な行動を行えることが期待されるが、そもそも人間は、周囲の環境に関して全てを知ることはできず、また、それらから未来の予測に関する推論を全て行うこともできない。このような限られた知覚・知識(信念)や推論能力のもと、人間は、動的に変化する周囲の環境を認識しながら、自らの目的を達成するための計画を立て、それに基づいて行動を決定しようとする。このような人間の行為決定の過程において、「意図」という心的状態が重要な働きを持つとするのが、上述した Bratman の「意図の理論」である。それによると、人間は、上述した信念や推論能力の限界から、目的を達成するための行動の全てを一挙に決めることはできないため、まず部分的な計画を立て、その実行に際してその手段を随時階層的に決定(実践的推論)しながら行動する。このとき、計画の各部分に対して、それを現在ないし未来に「行

おう」と考える心的状態が「意図」と呼ばれる。ここで、行為の決定が合理的に行われるには、意図が、それ自身および信念や願望と整合性を持つことが要求される（心的状態の整合性）。また、計画を一度決めてそれを実行しようと決めた場合、それはある程度の慣性を持ち、計画の目標を達成するか、あるいはその達成をあきらめるなどの事象が起きるまで保持される。すなわち意図は、行為の一貫性を保つように働く。自律エージェントの実現においても、この「意図の理論」を適用し、エージェントに明示的な「意図」の概念を持たせて、その保持と更新によって、目的を達成するための合理的で一貫的な行動をとらせようとする手法が考えられる。

BDI モデル、および BDI アーキテクチャは、そのようなエージェントのモデルおよび基本設計として提案されているものである。BDI モデルは、エージェントの信念 (B)・願望 (D)・意図 (I) といった心的状態を明示的に表現し、「意図の理論」に基づく合理的行為を行うエージェントをモデル化しようとするものであり、その形式化に用いられる論理体系が BDI logic である。また、それらに基づくエージェントアーキテクチャが BDI アーキテクチャであり、これは、プランライブラリからその時の状況に応じて選択されたプランを意図として保持して、その意図の実現に向けた行為を選び、それを実行するという一連の行為のループを行うアーキテクチャである。プランライブラリやエージェントの内部状態、行為決定のメカニズムなどは、エージェントの信念・願望・意図およびそれらの時間的变化によって記述され、BDI logic を用いて形式化されることによって、行為過程や仕様などに関する厳密な議論に耐えうる点が特徴である。このため BDI モデルは、自律エージェントのモデル化として広く受け入れられており、これを基にして、多数の実装や BDI モデルに基づくエージェント記述言語、アーキテクチャの改良の提案や、また、他のエージェント制御手段との結合としては例えば学習の併用などをはじめ、多くの研究がある。BDI logic は、分岐時間時相命題論理の 1 つである CTL\* に、エージェントの心的状態を表す様相オペレータ BEL、DESIRE、INTEND を導入し、述語論理に拡張したものである。信念・願望・意図といった心的状態やその時間的变化を明示的に表現できるため、エージェントの意思決定の過程や満たすべき性質などを、BDI logic の論理式で記述することができ、それらの論理式の恒真性のチェックを行うことなどによって、「意図の理論」で述べられている心的状態の整合性をはじめ、エージェントの性質に関する形式的な検証が可能である。しかし、Rao らが BDI logic を提案した当初は、演繹体系が示されていなかったことから、検証はモデル論的に行うしかなく、また、自動検証はできないといった問題があり、形式化の利点を十分に生かせない欠点が指摘されていた。これに対し Rao らは、ベースを命題論理の CTL に制限した BDI logic について、Hilbert style の公理体系を構築し、その健全・完全性を示した。しかし、Hilbert style の公理体系であるため、証明可能性をチェックするには、可能な証明経路をほぼ全数探索することになり、現実的な応用はしにくい。このため、BDI logic で記述されたエージェントの性質を何らかの演繹に用いることが困難であるという問題点があった。

本論文では、同じく命題論理の CTL をベースに制限した BDI logic に対してではあるが、sequent calculus による演繹体系を与え、その健全・完全性を示している。この体系は、命題論理の決定アルゴリズムとして知られる、Wang のアルゴリズムの拡張による決定アルゴリズムを持つため、全数探索によるよりも効率的な証明可能性のチェックが可能であり、BDI エージェントの性質に関する機械的な検証に向く長所を持ち、さらに、実行可能な仕様記述言語としての応用も考えられる。また、整合性公理の扱いについても示されている。また、Rao らの体系は、1つのエージェントの心的状態を表現するオペレータしか持たないため、単一のエージェントにしか適用できない。マルチエージェント環境におけるエージェントの社会的行為者としての振る舞いを形式的に扱うには、エージェントが他のエージェントの心的状態（に関する自分の信念）を明示的に思考に用いる状況を記述したり、また、エージェント間の協調などを行うには、相互信念について扱う能力も必要となる。本論文では、上述の体系に、複数のエージェントの心的状態を表現するオペレータを導入して拡張し、マルチエージェント環境に適用可能とした体系も与えている。

本論文は大きく3つに分かれ、まず第7章までが、心的状態の整合性を考慮しない BDI logic に関するものである。3章までで構文論および意味論を与え、4章で演繹体系を導入し、5・6章でこの演繹体系の健全性と完全性、7章で決定手続きについて述べてある。続いて10章までは、心的状態の整合性公理を考慮した体系に関するものである。8章で、心的状態の整合性公理を恒真とするよう意味論を変更した体系を導入し、9章でその体系に対する演繹体系とその健全性・完全性を示し、10章でその応用について述べてある。残りの部分がマルチエージェント環境への拡張に関するものである。11章で、マルチエージェント環境に拡張された構文論・意味論を導入し、12章で演繹体系とその健全性・完全性を与え、13章で応用について述べてある。第14章はまとめである。

## 論文審査の結果の要旨

本論文では、これまでに与えられていなかった、BDI logic の sequent calculus による演繹体系を与えた点が新規と認められる。従来 BDI logic の演繹体系は Rao らが与えた Hilbert style の公理化によるものしかなく、証明を探索する適切なアルゴリズムがなかった。そのため、BDI logic で記述されたエージェントの性質を何らかの演繹に用いることが困難であるという問題点があった。本論文で与えた sequent calculus による演繹体系は、Wang のアルゴリズムの拡張による決定アルゴリズムを持つため、BDI アーキテクチャの心的状態の整合性などの諸性質に関する自動演繹による証明や、仕様検証などへの応用が開ける他、将来的には実行可能な仕様記述言語としての応用が考えられる。また、マルチエージェントに拡張された BDI logic の演繹体系を与えた点も新規である。エージェント間のコミュニケーションや協調に関わる諸概念を BDI logic で扱うには、少なくとも、あるエージェントが他のエージェントの心的状態を明示的に記述したり、相互信念などについて扱ったりする拡張が必要である。そのような体系としては従来、Wooldridge らが扱っていたものはあるが、それらは演繹体系を持っていなかった。本論文では、上述の演繹体系をさらに、複数エージェントのそれぞれの心的状態や相互心的状態を扱えるように拡張している。これによって、マルチエージェント環境における、複数エージェントの相互心的状態、あるいはその時間変化について明示的に記述し、それに関する推論を行うことが可能となった。

本論文で提案した演繹体系は、決定手続きを持つため、推論エンジンとしての利用が可能である。特に複数のシナリオを実行するエージェントにおける整合性検証などの応用が考えられる。BDI アーキテクチャは、環境から得られる状況（信念）や意図に基づいて、実践的推論で選択した複数のシナリオ記述を並行実行するエージェントアーキテクチャである。個々のシナリオ内には、目標を達成するための計画が、合理的な行為列として、計画そのものの整合性を保つように記述されている。ところが、10.2節の例で述べられたような相矛盾する行為を含む異なるシナリオを並行に実行した場合、副作用として生じる信念の矛盾を検出する枠組みを構築して、いずれかのシナリオを中断する、または矛盾を修復するシナリオを起動する必要がある。本論文の演繹体系は、このような心的状態の整合性検証を行う推論エンジンへの応用が期待できる。ただし、効率的な実現方式などは今後の課題であろう。また、このような整合性検証システムが実現できれば、並行実行する複数シナリオのデバッグにも使用できる。さらに、マルチエージェントへの拡張によって、13.2の muddy children のように、現実に良く知られた問題への応用も可能となる。本来、sequent calculus による演繹体系では、前提の sequent が結論より小さくなることを利用して、Wang のアルゴリズムによる決定手続きを

与えるが、本論文では、sequent calculus による演繹体系に、証明のループの概念 (4.3) を導入し、ループの性質に依存して証明の可否を定義している。プロセス代数での sequent calculus など類似の手法を用いた例はあるが、BDI logic にこの手法を用いている点には独自性が認められる。

本論文では、演繹体系に決定手続きが与えられたことにより、BDI logic で形式的に記述された性質について、証明が存在するかどうかのチェックが可能となる。これによって、10章では、Bratman が「意図の論理」で取り上げた、非対象性テーゼや計画と信念との整合性などの性質について、実際に証明を行っている。特に、非対象性テーゼの後半部分の曖昧であった部分を、証明の有無による区別によってより厳密化した (10.1後半) 点や、両立しないという信念を持つ 2 つの計画も同時に達成しなければ支障はないことを形式的に示せた点などは評価できる。この他、Rao らによる公理化では導入されていなかった、8.1の (4) ～ (6) を整合性公理として導入できることも評価できる。