様　式　Ｃ－１９、Ｆ－１９、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成 ２８ 年　 ６ 月 １６ 日現在

機関番号：１４７０１
研究種目：基盤研究(C)（一般）
研究期間：2012～2015
課題番号：２４５００１７４

研究課題名（和文）不誠実なエージェントに関する研究

研究課題名（英文）A study on dishonest agents

研究代表者
坂間　千秋（SAKAMA, Chiaki）

和歌山大学・システム工学部・教授

研究者番号：２０２７３８７３

交付決定額（研究期間全体）：（直接経費）　　4,100,000円

研究成果の概要（和文）：人間の不誠実な行動原理を理解するために、不誠実に振舞う人工知能エージェントに関する研究を行った。具体的成果は以下の通り。①ウソや欺瞞などの不誠実の概念を数理論理的に形式化し、不誠実な推論アルゴリズムを設計した。②エージェントが不誠実に振舞うようになるプロセスの帰納的学習モデルを構築した。③交渉や論争といった社会的状況で不誠実に振舞うエージェントのモデル化と実装を行った。④日常会話において話者の意図を推定する会話の含意と話者が聴き手を誤誘導するプロセスを形式化した。⑤エージェントが不誠実な行動を戦略的に選択するよう進化するプロセスをマルチエージェントシステムを使ってシミュレーション実験した。

研究成果の概要（英文）：The goal of this study is to understand dishonest behaviors of humans and to realize them by artificial agents. The results of this study are as follows. (1) We built a logical theory of dishonesty such as lies and deception. We developed algorithms for dishonest reasoning by agents. (2) We characterized processes of learning dishonest behaviors using inductive logic programming. (3) We developed formal models of negotiation and argumentation among agents who may behave dishonestly. We argued the possibility of distinguishing dishonest agents from others. (4) We formulated conversational implicature in human dialogues and showed how speakers could use it for misleading hearers. (5) We performed experiments to examine the effect of self-interested agents in cooperative agent societies. We developed multiagent systems and observed the process of evolving self-interested agents in resource bounded environments. We investigated conditions to suppress the increase of self-interested agents.

研究分野：人工知能

キーワード：不誠実推論　利己的エージェント　認識論理　議論フレームワーク　エージェント交渉　帰納推論　会話の含意　マルチエージェントシステム

１．研究開始当初の背景

「ウソ」をつく行為は人間の基本的行動原理であり、哲学の分野では「ウソとは何か？」という議論が過去数十年にわたって議論されている。一方、人間の知能を抽象化し、人間の行動を模擬する計算機を実現することを目的とした人工知能の分野において、「ウソ」に関する研究はこれまで殆ど行われていない。人工知能の分野で「ウソ」の研究が重要であると考えられる理由はいくつかある。第一に「ウソ」をつくという行為は人間特有の知的行為であり、知能と思考を要するものである。従って、「ウソ」をつく行動原理を探究することは人間知能を理解する上で重要である。第二に「ウソ」をつく思考プロセスを明らかにすることは、「ウソ」をつくコンピュータシステムを実現するステップになる。例えば、患者に真実を伝えない介護システムや、生徒に故意に誤答を提示し、生徒自身に対話的に誤りを発見させる教育システムなどへの応用が考えられる。第三に「ウソ」をつく行為は人間の社会的行動であり、不誠実なエージェントをモデル化することにより、インターネット社会で誤った情報を提供するエージェントを機械的に検出するための足掛かりとすることが考えられる。従来の人工知能研究においては、エージェントは真実もしくは正しいと信じる信念を知識ベースとして持ち、その信念の下で推論される結果に基づいて合理的に行動することが原則とされた。この仮定の下では、エージェントはいつも「誠実に」振舞う。しかし、人間は社会生活において自らの利益を確保するために事実を曲げたり不誠実に振舞ったりするため、いつも誠実に振舞うエージェントは人間のモデル化としては必ずしも適当であるとはいえない。こうした研究背景から、本研究ではこれまであまり注目されることがなかった「不誠実な(dishonest)」振舞いをするエージェントに関する研究を行う。

２．研究の目的

本研究では、不誠実に振舞うエージェントをモデル化し、エージェントが不誠実な行動をとる誘因、不誠実なエージェントの発見、エージェントが不誠実に振舞うようになる学習プロセスを明らかにする。具体的には、「ウソ」、「知ったふり」、「知らないふり」などの不誠実さの概念を数理論理的に形式化し、エージェントが行う不誠実な推論過程をアルゴリズム化する。次に、交渉や議論などの異なる社会的場面でエージェントが不誠実に振舞う状況を想定し、交渉エージェントや議論フレームワークの枠組で実現する。また、日常会話における会話の含意を使って話者が聴き手を誤誘導するプロセスを形式化する。さらに、エージェントが不誠実な行動を学習し戦略的に選択するよう進化するプロセスを、機械学習や進化計算の手法を用いてコンピュータ上に再現する。

３．研究の方法

本研究では、不誠実なエージェントに関する知識の表現、推論、認知、学習という４つの観点から研究を行い、それぞれの基礎理論の構築とマルチエージェントシステムを使った実装実験を行う。具体的には不誠実なエージェントのオントロジーの整備と行動原理の論理的形式化、不誠実な推論アルゴリズムの構築、さまざまな社会的場面における不誠実な振舞いのモデル化、不誠実なエージェントが発現し進化するプロセスのシミュレーション実験を行う。これらの研究成果を踏まえて人間の不誠実な行動原理を理解し、より「人間らしい」人工知能を設計するための基盤技術に貢献するとともに、悪意のあるエージェントを機械的に発見するシステムの実現に向けてのステップとする。

４．研究成果

本研究期間中に行った研究テーマとその成果は以下の通りである。

(1) 不誠実の概念の数理論理的形式化

本研究では、ウソ、知ったふり、知らないふり、といった異なる種類の不誠実な行動を認識論理(epistemic logic)を用いて形式化し、系統的に比較を行った。また、話者の心的状態によって定義されるウソに対して、聴き手の心的状態の変化によって決まる欺瞞(deception)を動的認識論理を使って形式化し、さまざまな種類の欺瞞を論理的に形式化した。

(2) 不誠実な行動の学習モデル

本研究では、人間が不誠実に振舞うようになるプロセスをモデル化するために、ある目的を達成するために作り話をでっち上げる作業を、事例を説明するために仮説を構成する帰納的学習と対比させ、帰納論理プログラミングを使って不誠実な推論を行う手法を示した。さらにエージェントが不誠実な態度を行動ルールとして獲得するプロセスを定式化した。

(3) 不誠実なエージェントによる交渉

本研究では、エージェント間交渉における不誠実な振舞いのモデル化と実装を行った。交渉の過程で人間は自らの利益を高めるために真実とは異なる言明を行うことがある。そこで交渉におけるエージェントの不誠実な振舞いを仮説論理プログラムを使ってモデル化し、交渉のプロセスを解集合プログラミングシステムの上で実装した。

(4) 不誠実なエージェントによる論争

本研究では、論争において偽りの事実に基づく主張や真偽が不明な事実に基づく主張を展開するエージェントをモデル化するために、抽象的議論フレームワークに基づく「論争ゲーム」を導入した。また２人論争ゲーム

において、プレイヤーがゲームに勝つための条件や戦略、さらに相手の不誠実な主張を見破る方策を検討した。

(5) 会話の含意と誤誘導
本研究では、日常会話において話者の意図を推定する「会話の含意」を認識論理を使って定式化し、アブダクションを用いて話者の意図を推定する方法との比較を行った。また、聴き手の推測を逆手にとって話者が聴き手を欺く(misleading)プロセスの形式化を行い、対話システムで人間を欺くチューリングテストへの適用事例についても述べた。

(6) 不誠実なエージェントの創発と進化
本研究では、不誠実な行動が創発する状況を観察するために、資源が限られた環境下でエージェントが非協力的に振舞うようになるプロセスをマルチエージェントシステムを使って実現した。また、複数のエージェントが協同作業を行う環境において、自らの利益を最大化するために利己的に振る舞うエージェントを導入し、エージェント社会への影響についてシミュレーション実験を行った。

５．主な発表論文等
（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 3 件）
①Chiaki Sakama and Katsumi Inoue.
Abduction, Conversational Implicature and Misleading in Human Dialogue. Logic Journal of the IGPL, DOI:10.1093/jigpal/jzw027 **(印刷中)**, 2016. **(査読有)**

②Chiaki Sakama, Martin Caminada and Andreas Herzig.
A Formal Account of Dishonesty. Logic Journal of the IGPL, vol.23(2), pages 259-294, 2015. DOI: 10.1093/jigpal/jzu043 **(査読有)**

③ Tran Cao Son, Enrico Pontelli, Ngoc-Hieu Nguyen and Chiaki Sakama.
Formalizing Negotiations Using Logic Programming. ACM Transactions on Computational Logic, vol.15(2), Article No.12, 2014. DOI: 10.1145/2526270 **(査読有)**

〔学会発表〕（計 12 件）
① Ryuki Shimoji and Chiaki Sakama.
Multiagent Collaborative Search with Self-Interested Agents. IEEE/WIC/ACM International Conference on Intelligent Agent Technology (IAT'15), Singapore, December 9, 2015.

② Chiaki Sakama.
A Formal Account of Deception. AAAI Fall 2015 Symposium on Deceptive and Counter-Deceptive Machines, Arlington, Virginia, USA, November 12, 2015.

③ Martin Caminada and Chiaki Sakama.
On the Issue of Argumentation and Informedness. 2nd International Workshop on Argument for Agreement and Assurance (AAA 2015), Kanagawa, November 17, 2015.

④ Chiaki Sakama and Tjitze Rienstra.
Representing Argumentation Frameworks in Answer Set Programming. 1st International Workshop on Argumentation and Logic Programming (ArgLP 2015), Cork, Ireland, August 31, 2015.

⑤ Chiaki Sakama and Katsumi Inoue.
Abduction, Conversational Implicature and Misleading. 7th International Conference on Model-based Reasoning in Scientific and Technology (MBR'15), Sestri Levante, Italy, June 25, 2015.

⑥ Chiaki Sakama.
Counterfactual Reasoning in Argumentation Frameworks. 5th International Conference on Computational Models of Argument (COMMA 2014), Pitlochry, Scotland, September 9, 2014.

⑦ Naoki Yamada and Chiaki Sakama.
Evolution of Self-Interested Agents: An Experimental Study. 7th Multi-Disciplinary International Workshop on Artificial Intelligence (MIWAI'13), Krabi, Thailand, December 10, 2013.

⑧ Chiaki Sakama.
Abduction in Argumentation Frameworks and its Use in Debate Games. 2nd International Workshop on Argument for Agreement and Assurance (AAA 2013), Kanagawa, November 28, 2013.

⑨ Chiaki Sakama.
Debate Games in Logic Programming. 27th Workshop on Logic Programming (WLP), Kiel, Germany, September 13, 2013.

⑩ Chiaki Sakama.
Learning Dishonesty. 22nd International Conference on Inductive Logic Programming (ILP 2012), Dubrovnik, September 18, 2012.

⑪ Chiaki Sakama.
Dishonest Arguments in Debate Games.
4th International Conference on Computational Models of Argument (COMMA 2012), Vienna, Austria, September 12, 2012.

⑫ Chiaki Sakama.
A Formal Model of Dishonest Communication. Workshop on Formal Models of Communication, Opole, Poland, August 6-10, 2012.

〔図書〕（計 8 件）
①Martin Caminada and Chiaki Sakama, New Frontiers in Artificial Intelligence, Lecture Notes in Artificial Intelligence, Springer-Verlag, **印刷中**, 2016.

② Chiaki Sakama.
Computational Models of Argument, Frontiers in Artificial Intelligence and Applications, IOS Press, pages 385-396, 2014.

③ Naoki Yamada and Chiaki Sakama.
Multi-disciplinary Trends in Artificial Intelligence, Lecture Notes in Artificial Intelligence 8271, Springer-Verlag, pages 329-340, 2013.

④ Chiaki Sakama. Declarative Programming and Knowledge Management, Lecture Notes in Artificial Intelligence 8439, Springer-Verlag, pages 185-201, 2013.

⑤ Chiaki Sakama. New Frontiers in Artificial Intelligence, Lecture Notes in Artificial Intelligence 8417, Springer-Verlag, pages 285-303, 2013.

⑥ Chiaki Sakama. Inductive Logic Programming, Lecture Notes in Artificial Intelligence 7842, Springer-Verlag, pages 225-240, 2013.

⑦ Chiaki Sakama. Computational Models of Argument, Frontiers in Artificial Intelligence and Applications, IOS Press, pages 177-184, 2012.

⑧ Katsumi Inoue, Chiaki Sakama and Lena Wiese.
Applications of Declarative Programming and Knowledge Management, Lecture Notes in Artificial Intelligence 7773, Springer-Verlag, pages 134-151, 2012.

〔その他〕
研究成果は以下のホームページで公開中。
http://www.wakayama-u.ac.jp/~sakama

６．研究組織

(1)研究代表者
坂間　千秋（SAKAMA, Chiaki）
和歌山大学・システム工学部・教授
研究者番号：20273873

(2)研究分担者
（　　）

研究者番号：

(3)連携研究者
（　　）

研究者番号：