

2.1.3 エージェント技術

(1) 研究開発領域の定義

自ら判断し行動する主体¹⁾をコンピューターシステムとして実現したものをエージェントと呼ぶ。広い意味では人工知能 (AI) そのものであるが、特に自律性・自発性・社会性・反射性といった特性がエージェントの特徴として取り上げられる。すなわち、「自分自身の動作の目標を設定して動作したり (自律性、自発性)、他のエージェントと協力して組織を構成して問題解決を実行したり (社会性)、種々の変化や変動を察知して適応的に動作したり (反射性) する処理体」²⁾がエージェントとみなされる。その自律的メカニズム (自律エージェント)、複数主体の協調 (マルチエージェントシステム)、人間とのインタラクション (インターフェースエージェント)、社会的活動・現象のシミュレーション (マルチエージェントシミュレーション) 等を実現しようとするのが、本研究開発領域である。

(2) キーワード

自律エージェント、マルチエージェントシステム、マルチエージェントシミュレーション、インターフェースエージェント、ヒューマンエージェントインタラクション、対話エージェント、対話システム、分散人工知能

(3) 研究開発領域の概要

[本領域の意義]

「2.1.1 知覚・運動系のAI技術」と「2.1.2 言語・知識系のAI技術」が、「見る」「聴く」「考える」「話す」「動かす」「学ぶ」といった、一人一人の人間の知能が持つある側面を実現することに重点を置いているのに対して、

エージェントの定義・特徴 ● エージェントは自ら判断し行動する主体をコンピューターシステムとして実現したもの ● 自分自身の動作の目標を設定して動作したり(自律性、自発性)、他のエージェントと協力して組織を構成して問題解決を実行したり(社会性)、種々の変化や変動を察知して適応的に動作したり(反射性)する特徴を持つ		政策的課題 ● 大規模コンピューティング基盤・データ基盤の整備・強化 ● 産学連携：実社会・実用現場でのデータ取得・検証 ● さまざまな分野の学際的な研究 ● マルチエージェント、対話エージェント、HAIの間の連携	
	マルチエージェントシステムおよびシミュレーション	対話エージェント	ヒューマンエージェントインタラクション (HAI)
領域の意義	● さまざまな社会活動・現象のモデル化とそれを用いたシミュレーション・予測 ● それを通して、社会活動・現象のより正確な理解や意思決定への活用	● 普段使っている言葉(自然言語)でコンピューターに指示や意図を伝えられる ● ハンズフリー・アイズフリーで使える	● 外見や応答の仕方も含めて、ユーザーそれぞれが心理的・認知的に受け入れやすいインターフェースエージェント設計 ● 自然性・親密性・効率性等の向上
研究開発の動向	● 1980年代は分散人工知能、1990年代にはマルチエージェント、2000年代は経済パラダイム志向、2010年代は社会実装志向(電子商取引、電力・交通等のシミュレーション) ● メカニズムデザイン：オークションやマッチング問題への適用、自動メカニズムデザインやオンラインメカニズムデザインへの発展 ● セキュリティゲームの社会適用 ● トップランク国際会議としてAAMAS、コンペティションAANC、国内ではJAWS等	● タスク・利用シーンを限定した実用化から、スマートフォンやスマートスピーカーの音声対話アシスタントへ広がり ● タスク指向型対話から非タスク指向対話(雑談)への広がり ● バイブライン構成から深層学習End-to-End構成へ ● 大規模データ学習による対話エージェント：Google Meena、Facebook Blender、OpenAI GPT-3 ● トップランク国際会議：SIGDIAL、ICMI、コンペティションAlexa Prize等	● 人間の心理・認知、自然言語だけでなくエージェントの外見・非言語情報まで総合的なインタラクションに着目した日本発の研究領域 ● 国内でHAIシンポジウム、日本発の国際会議HAI ● 仮想エージェント、ロボット、人間同士のコミュニケーションを研究 ● HAI設計論：エージェントのデザイン、擬人化と適応ギャップ、メディアの等式、ナッジ、意図スタンス
科学技術的課題	● 人間の判断・行動のモデル化の困難性、現実世界の多様性・不確実性 ● 正確な予測よりも問題・リスクの可能性と対策検討への活用、そのための条件を変えた反復による計算量の増大	● 文脈の一貫性、応答の多様性、外部知識の参照、キャラクターの一貫性、制御性の確保、評価方法・評価尺度 ● マルチモーダル情報統合・実世界連動	● 受け止め方の個人差を吸収する設計方法論・評価方法 ● キラアアプリケーションの探索 ● HAIの知見をマルチエージェントシミュレーションや対話エージェントに展開

図 2-1-5 領域俯瞰：エージェント技術

本節の「エージェント技術」は、自律的に行動するAIをベースとして、それが他者（他のAIや人間）や社会・環境とインタラクションするという側面にアプローチしている。それによって、例えば、以下のようなことを期待できる。

複数のエージェントが相互にインタラクションするマルチエージェントシステムの枠組みを用いて、さまざまな社会活動・現象をモデル化することができる。そして、そのモデルを用いたシミュレーションを通して、起こり得ることを予測したり、モデルの妥当性を検証したりすることで、そのモデルが対象とする社会活動・現象に対するより正確な理解や意思決定への活用が可能になる。昨今、機械学習技術を用いたビッグデータ解析が予測や意思決定に盛んに活用されるようになったが、社会活動・現象のような複雑な振る舞いをする対象に対しては限界がある。その原因としてミクロ・マクロループの存在が指摘されている³⁾。社会活動・現象におけるミクロ・マクロループとは、個人の行動の集積がマクロなレベルの社会全体の動きを生成し、さらに、社会全体の動きがミクロなレベルの個人の行動を変化させていくような循環を意味する。これが存在するため、ある期間のビッグデータから規則性を見いだしたとしても、将来の動きはその規則性から外れていってしまうということが起こる。この問題に対して、マルチエージェントシステムによるシミュレーションでは、構成論的アプローチを取ることで、よりダイナミックに動きを捉え、多面的な理解・予測をすることが可能になる。

また、インターフェースエージェントは、人間とのインタラクションやユーザーインターフェースにおいて、自然性・親密性・効率性等を高める。例えば、普段使っている言葉（自然言語）でコンピューターに指示や意図を伝えられるインターフェース（対話エージェント）ならば、階層的なメニューから探したり、特別のコマンド入力・操作方法をあれこれ覚えたりする必要がない。しかも、ハンズフリー・アイズフリーで使える。さらに、自然言語での対話に限らず、人間とインターフェースエージェントあるいは物理的身体を持つロボットとの関係を総合的に考えるヒューマンエージェントインタラクション（HAI）に関する知見を取り込むことで、外見や応答の仕方も含めて、ユーザーそれぞれが心理的・認知的に受け入れやすいインターフェースエージェントの設計が可能になる。

【研究開発の動向】

論文1) では、エージェント技術に関する研究を次のような4種類に分類している。

1つ目は、実世界において自律的に環境を観測し、判断し、行動することを可能にする計算モデルの研究である。知能ロボットが典型的な応用例であり、BDI（信念、願望、意図）モデル、強化学習、サブサンプシジョンアーキテクチャー等が研究されてきた。

2つ目は、多数のエージェントの協調や競争の計算モデルの研究である。電子商取引、電力マネジメント等に応用されている。黑板モデル、自動交渉、メカニズムデザイン等、マルチエージェントシステム研究として活発に取り組まれている。

3つ目は、人間（一人あるいはグループ）と言語的・非言語的な対話を行い、社会的役割を演じるインターフェースエージェントの研究である。自然言語による対話を実現する対話エージェントや、言語や擬人化のインターフェースに限ることなく、人間の心理・認知面を重視して総合的に人間とエージェントのインタラクションのあり方を考えるHAIの研究が進められている。

4つ目は、エージェントを用いたシミュレーションである。個々の行動主体を適切な粒度でモデル化し、そのインタラクションから生じる複雑な現象を観察する。

1つ目は2つ目の基礎にもなっており、4つ目は2つ目のマルチエージェントシステムを用いる。また、3つ目のうち対話エージェントは、自然言語処理（「2.1.2 言語・知識系のAI技術」で詳細記載）を基礎として

おり、HAIとは技術発展の流れが異なる。そこで、以下では、自律エージェントを含むマルチエージェントシステムおよびシミュレーション、対話エージェント、HAIに分けて、研究開発の動向を述べる。

① マルチエージェントシステムおよびシミュレーションの研究開発動向

初期のエージェントの研究分野は分散人工知能と呼ばれていた。1980年に米国で分散人工知能ワークショップDAI (Distributed Artificial Intelligence Workshop) の第1回が開催され、これは1994年までほぼ毎年開催された。当初は、分散した問題解決器の間で解くべき問題を共有し、協調しながら問題を解く分散協調問題解決の方法論に取り組みの中心があったが、その後、各問題解決器が自律性を備え、それぞれが独立の目標を持つようなケースへと取り組みが広がり、マルチエージェントシステムと呼ばれるようになる。1989年から欧州を中心にマルチエージェントに関するワークショップMAAMAW (European Workshop on Modelling Autonomous Agents in a Multi-Agent World) が開催されるようになった。1995年にはマルチエージェントシステム国際会議ICMAS (International Conference on Multi-Agent Systems)、1997年には自律エージェント国際会議AGENTS (International Conference on Autonomous Agents) が始まり、それらは2002年に統合されてAAMAS (International Conference on Autonomous Agents and Multi-Agent Systems) となり、この研究分野の中心的な国際会議となっている。日本国内でも1991年に日本ソフトウェア科学会に「マルチエージェントと協調計算」(MACC) 研究会が立ち上がった。このMACC研究会を中心に、2002年からJAWS (Joint Agent Workshop & Symposium) がスタートし、現在では、電子情報通信学会「人工知能と知識処理」研究専門委員会、情報処理学会「知能システム」研究会、人工知能学会「データ指向構成マイニングとシミュレーション」研究会、IEEE Computer Society Tokyo/Japan Joint Chapterとの協同による学会横断的なイベントとして運営されている。他にも主要な国際会議として、PRIMA (International Conference on Principles and Practice of Multi-Agent Systems)、IEEE ICA (International Conference on Agents)、PAAMS (International Conference on Practical Applications of Agents and Multi-Agent Systems) がある。PRIMAは当初の名称をPacific Rim International Workshop on Multi-Agentsとし、また、IEEE ICAはJAWSの国際セッションのコミュニティによって創設されたものであり、日本が中心的な貢献をして、アジア地域におけるエージェント研究の活発化につなげてきた面を持つ。

このようにエージェント研究のコミュニティが発展し、マルチエージェント間交渉、マルチエージェントシミュレーション、マルチエージェント学習、マルチエージェントプランニング、メカニズムデザイン、エージェント指向ソフトウェア工学等、マルチエージェントシステムをベースとした技術開発も広がりを見せている¹⁾。2000年～2010年頃には、マルチエージェントが経済パラダイムと相性がよいことが注目され、ゲーム理論や計算論的メカニズムデザインがオークションやマッチング問題に適用されて、実用的な成果を上げた。2010年代になると、社会活動・現象を扱い、実際に社会に実装することが強く目指されるようになってきた。電子商取引、電力マネジメント、交通シミュレーション、ワイヤレスセンサーネットワーク等への適用が行われている^{6), 7)}。このような社会活動・現象を扱う上で、マルチエージェントシステムの中に人間も含めてモ

1 マルチエージェントシステムやエージェントシミュレーションの技術開発や応用分野の広がりについては、人工知能学会誌や情報処理学会誌の特集号^{4), 5)}に掲載された一連の解説論文に詳しい。

デル化すること（例えば、人間にどうインセンティブを与えるか、人間の限定合理性²をどう考慮するか等）が検討されるようになった。また、マルチエージェントの研究コミュニティでは、国際会議と合わせて、コンペティションを通じた技術検証・改良も取り組まれてきた。特に注目されるものとしては、2000年代には電子商取引、オークション、サプライチェーンマネジメント等に関するTAC（Trading Agent Competition）が開催され、2010年からは自動交渉エージェントに関するANAC（Automated Negotiating Agents Competition）が開催されている。

なお、注目動向として、[新展開・技術トピックス]^①でメカニズムデザイン、[注目すべき国内外のプロジェクト]^①でセキュリティゲームを取り上げる。また、「2.1.5 意思決定・合意形成支援」にて自動交渉エージェント、「2.1.9 社会におけるAI」にてパーソナルAIエージェントを取り上げている。

② 対話エージェントの研究開発動向

音声対話技術を用いたインターフェースは、まず、タスクや利用シーンを限定した形で実用化が進んだ。このような限定によって、扱うべき語彙・文型・文脈を現実的な規模に抑えることができる。例えば、コンタクトセンターでの問い合わせ応答や、旅行会話の音声翻訳等が挙げられる。Nuance、IBM、NEC等から法人向けソリューションが提供されたほか、情報通信研究機構（NICT）で開発された個人旅行者向け多言語音声翻訳アプリVoiceTraは、スマートフォンアプリとして、民間企業や、スポーツイベントでの案内、病院診察での外国人対応等に使われている。

この間、スマートフォンが普及したことで、そのフロントエンドのインターフェースとして音声対話アシスタントが使われるようになった。2011年にリリースされたAppleのSiriや2012年にリリースされたNTTドコモの「しゃべってコンシェル」がその代表である。また、2014年にMicrosoftがWindowsのフロントエンドとして、対話インターフェースを持つCortanaをリリースした。同年にはAmazon Echo（アシスタントソフトウェアAlexa）が発売され、家庭向けの音声対話アシスタント内臓スピーカー（スマートスピーカーやAIスピーカーとも呼ばれる）という新しい製品形態が生まれた。同種の製品として、その後、2016年にGoogle Home（アシスタントソフトウェアGoogle Assistant）、2017年にLINE Wave（アシスタントソフトウェアClova）等が続いた。さらに、アシスタント型の対話（タスク指向型対話）よりも雑談型の対話（非タスク指向型対話）が中心の「チャットボット」（Chatbot）³と呼ばれるソフトウェアも利用が広がりつつある。企業にとっての新たな顧客接点として、SNSやメッセージアプリでチャットボットが会話の相手になるという使われ方もされている。

以上のような対話インターフェースは、厳密な意味での自律性を必ずしも備えてはいないが、知識ベースを備えたり、擬人化された外見や性格を持っていたり、ユーザーとの対話から学習したりといった知的な機能を持つことから「対話エージェント」とも呼ばれる。技術面に目を向けると、近年はニューラルネットワーク・深層学習の適用が進んでいる⁸⁾。従来、音声認識、言語理解、対話制御、応答生成、音声合成というモジュールをパイプライン接続する構成が取られていたが、各モジュールを深層学習でモデル化するという置き換えが進んだ。さらに、パイプライン構成を取らずに、深層学習によって入力文から応答文を直接出力

- 2 限定合理性（Bounded Rationality）は、ノーベル経済学賞を受賞したHerbert A. Simonが1947年に提唱した概念で、人間は合理的に行動しようとしても、認知能力の限界やさまざまな制約により、限られた合理性しか持ちえないことを意味する。
- 3 会話（Chat）とロボット（Bot）を組み合わせた造語。

するEnd-to-End構成も取られるようになってきた。最先端の対話エージェントでは、極めて膨大な量のテキストデータから学習する傾向が強まっている(具体的事例は[注目すべき国内外のプロジェクト]②に示す)。

対話エージェント技術に関する国際的な研究発表・議論の場としては、SIGDIAL (Annual Meeting of the Special Interest Group on Discourse and Dialogue) やICMI (ACM International Conference on Multimodal Interaction) があるとともに、ACL (Annual Meetings of the Association for Computational Linguistics) をはじめとする自然言語処理分野の国際会議や、NeurIPS (Conference on Neural Information Processing Systems) をはじめとする機械学習分野の国際会議で発表されている。また、共通タスクを設定して方式や性能を比較するコンペティション型のワークショップも活発に開催されている。近年は、非タスク指向型対話に関するコンペティション(Amazonが開催するAlexa Prize⁹⁾等)が注目されている。

③ ヒューマンエージェントインタラクション(HAI)の研究開発動向

HAIは人間とエージェントとのインタラクションに重点を置いている^{10), 11), 12)}。マルチエージェントシステム研究と比べると、人間の心理・認知といった面からインタラクションを捉えている。対話エージェント研究と比べると、自然言語だけでなく、エージェントの外見や動き等の非言語情報や人間の心理・認知等も含めた総合的なインタラクションを扱っている。このような立場からHAIという新たな研究分野が日本発で立ち上がり、RO-MAN (IEEE International Symposium on Robot and Human Interactive Communication)、IROS (IEEE/RSJ International Conference on Intelligent Robots and Systems)、人工知能学会全国大会等でのオーガナイズドセッションやワークショップを経て、2006年に第1回のHAIシンポジウムが開催された。2013年からは国際会議HAI (International Conference on Human-Agent Interaction) も開催されている。

HAI研究において、人間とインタラクションするエージェントとして、主に3つの形態が考えられている。1つ目は仮想的なエージェントである。その分かりやすい例は、人間の外見を持つ擬人化エージェントであるが、ユーザーや目的によっては、必ずしも擬人化されたエージェントが最良のインターフェースであるとは限らない。2つ目は物理的な身体を持つエージェントであるロボットであり、センサーによって置かれた環境を把握し、アクチュエーターを介した動作によって環境に変化を及ぼすことも可能である。3つ目は人間である。HAI研究では、人間同士のコミュニケーションも研究対象の一つになっている。人間同士のコミュニケーションを解析・理解することは、1つ目や2つ目の形態をデザインする上で参考になるとともに、エージェントを介した人間同士のコミュニケーションCMC (Computer-Mediated Communication) の改善支援にHAI研究を役立てることにもつながる。

HAIや対話エージェントに関する研究は、AIだけでなく、認知科学、心理学、哲学、ヒューマンインターフェース、ロボティクス等がクロスする学際的研究分野である。対話エージェントが既に実用の域にあるのと比較すると、HAIはまだ新しい研究分野だが、AI技術がさまざまなアプリケーション・サービスでさまざまな人間(ユーザー)に利用されていくことを考えると、HAI研究に基づくインターフェースエージェントの設計論は今後ますます重要になっていくと考えられる([新展開・技術トピックス]②を参照)。

(4) 注目動向

[新展開・技術トピックス]

① メカニズムデザイン^{1), 6), 13), 14)}

マルチエージェントシステムにおいて、その系が複数の利己的なエージェントで構成される場合、系の設計者は個々のエージェントの行動を直接制御することはできない。その代わりに、エージェント間の関係、つまりゲームのルールを設計することで、間接的にエージェントを制御しようというのが、メカニズムデザインの考え方である。そのようなことを可能にするルールを設計する上で、社会的余剰 (Social Surplus) の最大化や誘因両立性 (Incentive Compatibility) の充足といった点が考慮される。社会的余剰とは、参加者全員の効用 (うれしさ) の総和であり、資源割り当ての良さを表す指標である。誘因両立性は、どの参加者にとっても、真の評価値を正直に申告することが最適戦略になるという性質である。

オークション問題を例にすると、単純には、各入札者に他の入札者の行動を観察せずに一度だけ入札させ、最高額の入札をした者を勝者とし、その者の入札額を支払わせる方式 (第一価格秘密入札方式) が考えられる。しかし、この方式では他者の評価値を推測しつつ自分の評価値を決めるため、真の評価値の正直申告が最善策とはならず、社会的余剰の最大化にも失敗する。これに対して、最高額の入札をした者を勝者とするのは同じだが、その勝者が支払う金額は2番目に高い入札者の金額とするという第二価格秘密入札 (Vickrey 入札) 方式が考案された。この方式を用いれば、正直申告が最善策となり (誘因両立性を満たし)、社会的余剰が最大化される。

このようなメカニズムデザインの考え方は、社会経済・電子商取引と相性が良く、特にオークション問題やマッチング問題で実問題への応用が進んでいる。第二価格秘密入札方式が Google 等の検索エンジンにおける検索連動広告で使われていることはよく知られている。そのような応用と並行して、扱う問題の前提・制約条件を広げたアルゴリズムの拡張・発展も進んでいる。複数の財が同時に販売され、財の組み合わせに対する入札が認められた「組み合わせオークション」、エージェントの数や関係がダイナミックに変化する環境下で逐次的な決定を行う「オンラインメカニズムデザイン」、エージェントの数や性質が与えられた条件下で、制約最適化問題としてルールを自動的に導出する「自動メカニズムデザイン」等が挙げられる。また、メカニズムデザインの研究業績により、Eric Maskin、Alvin Roth、Paul Milgromらが立て続けにノーベル経済学賞を受賞した。なお、メカニズムデザインの詳しい説明や具体事例等は「2.3.4 メカニズムデザイン」に記載している。

② HAI 設計論

HAI 研究の主要な目標の一つは、人間とインタラクションするエージェントの設計論である。HAI は比較的新しい研究分野であり、この設計論が確立されたとはまだ言えないが、設計において考慮すべき事項や指針が徐々に見出されつつある^{10), 11), 12), 15)}。そのいくつかを以下に示す。

- a. エージェントのデザイン：外見の選択 (人型・動物型等、人型の場合も性別・性格等、実写風・アニメ風等のどれを用いるか、親しみやすさと知性の印象が重要)、表出される情報の表現の選択 (自然言語による発話、非言語な表情、ジェスチャー等のどれを用いるか、タスクの種類や環境にも依存)、機能レベル (学習機能やユーザーの状態・意図の理解機能等) を適切に選択する必要がある。
- b. 擬人化と適応ギャップ：擬人化エージェントはユーザーがその機能や能力を推測しやすい反面、ユーザーが期待した機能・能力と実際との間に生じるギャップがリスクとなる。
- c. メディアの等式 (Media Equation)：メディアの世界と現実を無意識のうちに同一視してしまうとい

う人間の心理が、エージェントに対しても起こり得る。

- d. ナッジ (Nudge) : ナッジは、もともと「ひじで人を軽く突く」という意味があり、人々の行動を強制的に変えるのではなく、少しの刺激を与えることで自発的に望ましい行動を選択するように促すという、行動経済学で提唱された概念である。ナッジは悪用されるリスクがあり、倫理面の配慮や適正な選択構造の設計が求められる。
- e. 意図スタンス : 哲学者 Daniel Dennett は、人間が対象の振る舞いを理解し予測する際に、物理スタンス (物理現象として振る舞いを理解)、設計スタンス (設計仕様から振る舞いを理解)、意図スタンス (振る舞いの目的・意図から理解) という3つの心的姿勢を使い分けるとしている。HAI は上記 a ~ d のような点をコントロールしながら、意図スタンスにも踏み込んだ設計を行う。

[注目すべき国内外のプロジェクト]

① セキュリティーゲームの社会適用 (Milind Tambe の Teamcore)

Milind Tambe らのチーム (Teamcore) ⁴ は、マルチエージェント技術・ゲーム理論に基づく「AI for social good」のための研究開発に取り組んでいる。特に有名なのは、セキュリティゲームとしての定式化に基づくロサンゼルス空港の警備計画アプリケーション ARMOR の成功である。セキュリティゲームは、テロリスト等の攻撃者から空港等の重要施設を守るために、適切な警備員配置を決定する警備計画問題である ^{1), 6), 13), 14), 16)}。警備側・攻撃側ともに人員リソースは有限であり、かつ、互いに相手の動きを読み合い、最も効果的な人員配置を行おうとする。これはゲーム理論におけるシュタッケルベルグ (Stackelberg) ゲームとして定式化でき、均衡を計算することで、効果的な警備員配置の計画が可能になるというものである。この研究成果をきっかけとして、セキュリティゲームはマルチエージェントシステム研究におけるホットトピックとなり、理論展開と具体的応用が広く発展した。

Teamcore では、上述のような治安問題の他にも、次のような問題に取り組んでいる。公衆衛生問題では、HIV リスク行動削減、結核予防、母子保健等に対して、ソーシャルワーカーや公衆衛生のリソースに限りがある条件下で、社会的ネットワークの効果を最大化する取り組みが進められている。保全問題では、絶滅の危機に瀕した野生動物を保護するためのレンジャー配置計画をセキュリティゲーム (グリーンセキュリティゲーム) として扱うシステム PAWS が開発され、世界 100 か所以上の国立公園で使われている。他にも、サイバーセキュリティ、森林保護、ウイルスに対する薬設計、ソフトウェアコードテスト、交通システム等への適用も考えられている。このようなさまざまな社会的課題に対して実用的な成果を展開していることは注目される。

② 大規模データ学習による対話エージェント (Google、Facebook、OpenAI)

自然言語処理の分野では、トランスフォーマー (Transformer) という深層学習モデルで、大量のテキストデータによる自己教師あり学習を行うアプローチが顕著な成果を上げている (詳細は「2.1.2 言語・知識系の AI 技術」参照)。対話エージェントにおいても、このアプローチが取り入れられ、2020 年には Google、Facebook、OpenAI から立て続けに、それぞれが開発した最先端システムの成果が発表された。

4 以前は南カリフォルニア大学情報科学研究所 (USC/ISI : University of Southern California/Information Sciences Institute)、現在はハーバード大学に属している。

まず2020年1月にGoogleがチャットボットMeena(ミーナ)の研究成果を発表した¹⁷⁾。Meenaは、ソーシャルメディア上での会話データ341ギガバイトから、ニューラルネットワークのパラメーター26億個を学習した。人間らしい会話であるかを示す指標としてSSA (Sensibleness and Specificity Average: 整合度・特有度平均値⁵⁾)を用いて評価した結果、Meenaは79%で、それまで最高レベルのチャットボットとみなされていたMitsukuの56%を上回り、人間のスコア86%に迫った。Meenaは会話の中にジョークも盛り込むという。

次にFacebookが4月にチャットボットBlender(ブレンダー)の研究成果¹⁸⁾とオープンソース化を発表した。Blenderは掲示板サイトRedditの会話データ15億件から最大94億パラメーター(Meenaの3.6倍の規模)を学習した。Facebookは独自の設定によるMeenaとの比較実験に基づきBlenderの優位性を主張している。また、2020年11月に日本国内で開催された対話システムライブコンペティション3において、NTTコミュニケーション科学基礎研究所の日本語版Blenderが優勝した。

さらに、6月にはOpenAIからGPT-3が発表された¹⁹⁾(Microsoftがその独占ライセンスを取得)。GPT-3は3千億件のテキストデータから1750億個のパラメーターを学習した言語モデルである。GPT-3自体はチャットボットではないが、いくつかの単語や文を入力すると、対話例も含む膨大なテキストからの学習結果に基づき、その先に続くものを予測して提示することができるので、対話文も生成可能である。

(5) 科学技術的課題

① マルチエージェントシステムおよびシミュレーションの技術課題

近年、マルチエージェントシステムによるモデル化やそれを用いたシミュレーションは社会活動・現象への適用が進み、さまざまな社会的問題解決に効果を示しつつある。既に述べたようにメカニズムデザインに基づくオークションやマッチングのルール設計、セキュリティゲームとしての定式化による警備・保全・公衆衛生等の資源割り当て計画のほか、マルチエージェントシミュレーションに基づく電力マネジメント、交通マネジメント、ワイヤレスセンサーネットワーク制御、災害対策・避難計画等にも活用されている。

このようなマルチエージェント社会シミュレーション²⁰⁾への期待が高まるが、いくつかの技術的な課題も残されている。まず、シミュレーションに用いるエージェントのモデルの限界である。社会活動・現象に人間の行動が含まれる場合、その個々の人間の判断・行動を完全にモデル化することは難しい。そのため、シミュレーションから導かれた最適計画は、現実においても最適なものであるとは限らない。むしろ、さまざまな条件でシミュレーションを繰り返すことで、どんなことが起こり得るのか、重大なリスクの可能性と対策を検討するという使い方が有効なのかもしれない。人間の判断・行動を含め、現実世界における多様性・不確実性をエージェントモデルにおいてどのように扱っていくのかは、引き続き検討を要する。次に、計算量の増大も問題になる。社会活動・現象をモデル化する際のマルチエージェントシステムの規模の増大と、条件を変えたシミュレーションの繰り返しによって、計算量・計算時間が増大することも、現実の問題に広く適用していく上で課題となる。

5 Sensiblenessは話の筋が通っているかを意味し、日本語では整合度や分別度と訳される。Specificityは「それはいいね」のようなどんな話題にも当てはまるような応答ではなく、いまの文脈に固有の応答であるかを意味し、日本語では特有度、特異度、詳細度といった訳が使われている。

② 対話エージェントの技術課題

タスク指向型対話・非タスク指向型対話ともに深層学習の適用が進み、対話における話題の広がりや対話の自然さが改善されてきたが、以下のような技術課題が残されている⁸⁾。

- a. 文脈の一貫性：複数ターンにわたって対話を続ける際に、内容・用語等が首尾一貫したものであることが求められる。過去の会話を考慮するモデル等が検討されている。
- b. 多様性：「はい」「そうですね」等のありきたりな応答ではなく、より多様な応答文を生成できることが望ましい。前述のBlender等、大規模言語モデルを用いたシステムでは改善が見られる。
- c. 外部知識：外部知識を参照する機構を導入することで、よりリッチな情報をもとにした対話が可能になる。外部情報として利用するデータの種類や参照機構等が検討されている。
- d. キャラクターの一貫性：対話エージェントの性格には一貫性が求められる。End-to-End 応答文生成の際に、話者IDや話者プロフィールを考慮する手法等が検討されている。
- e. 制御性：特定の意図や感情を反映させた発話をさせたいケースがある。そのような意図・感情に基づいて応答文出力を制御する機構が検討されている。また、オープンな会話からの学習によっては不適切な対話をしかねないという、ある種の脆弱性問題²¹⁾に対しても、制御性の確保が求められている。
- f. 評価尺度：タスク指向型対話は各モジュールの性能やタスク達成率・タスク達成時間等で評価されるが、非タスク指向型対話の評価方法・評価尺度は大きな課題である。前述のSSAをはじめ、意味的類似度・多様性尺度や、これまで機械翻訳・要約・情報検索等で用いられてきた評価尺度の応用等、さまざまな検討が進められている。

また、対話という面を広げて、音声対話だけでなく、視覚情報等も含めたマルチモーダル情報を統合して扱うという方向への取り組みも進められている。カメラ映像として得られる実世界の状況やそこに存在する人や物に対する行為・操作を、音声対話と連動させるような試みが行われている（例えば「それを赤い箱に片づけて」や「缶コーヒーを買ってきて」とロボットに指示する等、「2.1.2 言語・知識系のAI技術」参照）。

③ HAI の技術課題

HAIはユーザーの心理・認知特性と関わりが深く、受け止め方の個人差が大きいことが、設計論の確立という面では難しさがあると考えられる。この課題に対して、従来のHCI（Human-Computer Interaction）の方法論への批判的観点から、多くのHAI研究では大規模な参加者実験による一般性の評価を行うことで対応している。一方、HAIの重要課題として、基礎的な観点から、ケーススタディー研究が多く、統一的な設計論が確立されていないこと、工学的な観点からは、インパクトのあるキラーアプリケーションが見えていないこと等が挙げられる。AI技術がさまざまなアプリケーション・サービスで幅広く利用されていくことを考えると、インターフェースエージェントの設計におけるHAIの側面は今後ますます重要なものになっていく。[新展開・技術トピックス] ②で述べたような設計上の留意点や指針に関する基礎的な知見を、今後も積み上げていくことが必要であろう。

また、HAIの考え方が効果的に働く応用分野として、医療、介護、福祉、教育、ビデオゲーム等が考えられる¹⁰⁾。このような応用を通して、設計論や要素技術を検証していくとともに、新たな技術課題や知見が生まれてくるものと思われる。さらに、既に実用化が進んでいるマルチエージェントシミュレーションや対話エージェントに対して、HAI研究から得られた知見を加えることで、新たな価値を生み出すという方向の展開が期待できる。

（6）その他の課題

① 大規模コンピューティング基盤・データ基盤の整備・強化

〔注目すべき国内外のプロジェクト〕②に記載したように、最先端の対話エージェントの言語モデルは、膨大な量のテキストデータから学習したものである。この規模の学習処理を実行できるのは、OpenAI、Facebook、Googleのような極めて大規模な計算資源を保有する組織だからこそである。また、マルチエージェント社会シミュレーションも、今後、エージェント数やシミュレーション条件のバリエーションが増えると、計算量が組み合わさ的に増大する。このようにエージェント研究には大規模なコンピューティング基盤やデータ基盤が必要になっており、これはもはや大学の一研究室で確保できる規模のものではなくなってきている。産業技術総合研究所のABCIや理化学研究所の富岳のように、国内の研究機関が共同利用できる大規模コンピューティング基盤・データ基盤の継続的な整備・強化が必要である。

② 産学連携および分野横断の研究開発推進

マルチエージェント社会シミュレーションや対話エージェントの研究開発では、実社会・実用現場でのデータ獲得や検証が不可欠であり、産学連携はそのための手段として、今後もいっそう推進されるであろう。また、HAIや対話に関する研究は、AI、ヒューマンインターフェース、ロボティクス、認知科学、心理学、哲学等、さまざまな分野の学際的な研究分野であり、分野横断の研究コミュニティとして発展してきている。さらに、現状では、マルチエージェントシステム／シミュレーションと対話エージェントとHAIとの間の連携はまだあまり強くないが、今後、これらの間の技術統合のための連携も生まれていくであろう。ここで述べた種類の分野横断の研究開発の推進体制の重要性が高まっていくと考える。

（7）国際比較

国・地域	フェーズ	現状	トレンド	各国の状況、評価の際に参考にした根拠など
日本	基礎研究	○	→	HAIは日本発の研究分野である。自動交渉エージェントに関するANACでも日本の貢献が見られる。対話分野の国際会議では一定数の論文が採択されているが、エージェント基礎研究全般として研究者層や国際学会での存在感はまだ弱い。
	応用研究・開発	○	↗	オンデマンド交通・避難計画等の交通・人流シミュレーションでの実用化が見られる。エージェント研究者も関わっている人狼ゲームやロボカップ等のコンペティションも日本が先導している。
米国	基礎研究	◎	→	他のAI分野と同様に研究者層が厚く、エージェント分野のAAMASや対話分野のSIGDIALやICMIの採択論文数も圧倒的1位である。
	応用研究・開発	◎	→	セキュリティゲームでの実用化をはじめ、エージェント技術の先端的な応用が進んでいる。大規模学習による対話エージェントでは圧倒的優位性を示しており、Facebook、Amazon、Microsoft等からの論文発表も多い。
欧州	基礎研究	◎	→	米国に次いで研究者層が厚い。特に英国の存在感がある（AAMASの採択論文数では米国に次いで英国が2位をキープしている）。欧州はマルチモーダル対話も比較的強い。
	応用研究・開発	○	→	基礎研究に比べると、企業から目立った研究成果の発表は見られない。

2.1

俯瞰区分と研究開発領域
人工知能・ビッグデータ

中国	基礎研究	○	↗	AAMASの採択論文数は徐々に増加している。自然言語処理・対話関連はACL・EMNLPで米国と並んで二強。
	応用研究・開発	○	↗	AI分野全体では米中二強の状況であるが、対話分野（タスク指向・非タスク指向）で企業（Tencent、Alibaba等）の取り組み・発表が活発であることを除いて、エージェント分野全般での中国の存在感はまだそれほどではない。しかし、中国のAI応用は急速に拡大しており、エージェント応用研究面も伸びつつある。
韓国	基礎研究	△	→	AAMASの採択件数もごくわずかである。SIGDIAL・ICMIでも目立った活動は見られない。
	応用研究・開発	△	→	目立った活動は見られない。
イスラエル	基礎研究	◎	→	バルイラン大学等はメカニズムデザインやゲーム理論等の基礎研究でトップレベルである。
	応用研究・開発	△	→	企業から目立った研究成果の発表は見られない。

(註1) フェーズ

基礎研究：大学・国研などでの基礎研究の範囲

応用研究・開発：技術開発（プロトタイプの開発含む）の範囲

(註2) 現状 ※日本の現状を基準にした評価ではなく、CRDSの調査・見解による評価

◎：特に顕著な活動・成果が見えている

○：顕著な活動・成果が見えている

△：顕著な活動・成果が見えていない

×：特筆すべき活動・成果が見えていない

(註3) トレンド ※ここ1～2年の研究開発水準の変化

↗：上昇傾向、→：現状維持、↘：下降傾向

参考文献

- 1) 長尾確・大沢英一・伊藤孝行, 「エージェント・マルチエージェントの過去と現在」, 『人工知能』(人工知能学会誌) 35巻3号(2020年5月), pp. 430-443.
- 2) 木下哲男・他, 「分散協調とエージェント」, 電子情報通信学会知識ベース「知識の森」7群7編. http://ieice-hbkb.org/portal/doc_673.html (accessed 2021-01-10)
- 3) 和泉潔, 「ビッグデータとエージェントシミュレーション」, 『情報処理』(情報処理学会誌) 55巻6号(2014年6月), pp. 549-556.
- 4) 服部宏充・栗原聡(編), 「特集: エージェント」, 『人工知能』(人工知能学会誌) 28巻3号(2013年5月), pp. 358-423.
- 5) 青木健児・浅井達哉(編), 「特集: マルチエージェントシミュレーション」, 『情報処理』(情報処理学会誌) 55巻6号(2014年6月), pp. 528-590.
- 6) 伊藤孝行・他, 「未来の社会システムを支えるマルチエージェントシステム研究(1) —経済パラダイム, 交渉エージェント, 交通マネジメント—」, 『人工知能』(人工知能学会誌) 28巻3号(2013年5月), pp. 360-369.
- 7) 伊藤孝行・他, 「未来の社会システムを支えるマルチエージェントシステム研究(2) —電力システムおよびワイヤレスセンサネットワークへの応用—」, 『人工知能』(人工知能学会誌) 28巻3号(2013年5月), pp. 370-379.
- 8) 東中竜一郎・増村亮, 「対話システムにおける深層学習の適用」, 『人工知能』(人工知能学会誌) 34巻

4号 (2019年7月), pp. 460-466.

- 9) Raefer Gabriel, et al., “Further Advances in Open Domain Dialog Systems in the Third Alexa Prize Socialbot Grand Challenge”, *Proceedings of the Alexa Prize Socialbot Grand Challenge 3* (2020) .
- 10) 大澤博隆, 「ヒューマンエージェントインタラクションの研究動向」, 『人工知能』(人工知能学会誌) 28巻3号 (2013年5月), pp. 405-411.
- 11) 山田誠二・小野哲雄, 『マインドインタラクション: AI学者が考える《ココロ》のエージェント』(近代科学社, 2019年) .
- 12) 山田誠二 (編), 「特集: 進化するHAI: ヒューマンエージェントインタラクション」, 『人工知能』(人工知能学会誌) 24巻6号 (2009年11月), pp. 809-884.
- 13) 松原繁夫, 「マルチエージェントシステムにおける経済学的アプローチ」, 『計測と制御』(計測自動制御学会誌) 55巻11号 (2016年11月), pp. 948-953.
- 14) 岩崎敦・東藤大樹, 「ゲーム理論・メカニズムデザインに関する研究動向」, 『人工知能』(人工知能学会誌) 28巻3号 (2013年5月), pp. 389-396.
- 15) 山田誠二・寺田和憲・小林一樹, 「人を動かすHAI デザインの認知的アプローチ」, 『人工知能』(人工知能学会誌) 28巻2号 (2013年3月), pp. 256-263.
- 16) Milind Tambe, *Security and Game Theory: Algorithms, Deployed Systems, Lessons Learned* (Cambridge Press, 2011) .
- 17) Daniel Adiwardana, et al., “Towards a Human-like Open-Domain Chatbot”, arXiv : 2001.09977 (2020) .
- 18) Stephen Roller, et al., “Recipes for building an open-domain chatbot”, arXiv : 2004.13637 (2020) .
- 19) Tom Brown, et al., “Language Models are Few-Shot Learners”, *Proceedings of the 34th Conference on Neural Information Processing Systems* (NeurIPS 2020; December 6-12, 2020) .
- 20) 野田五十樹, 「マルチエージェント社会シミュレーションが浮き彫りにする緊急時避難の課題」, 『学術の動向』 23巻3号 (2018年3月), pp. 42-47.
- 21) Mark Molloy, “Microsoft ‘deeply sorry’ after AI becomes ‘Hitler-loving sex robot’”, The Telegraph, 26 March 2016. <https://www.telegraph.co.uk/technology/2016/03/26/microsoft-deeply-sorry-after-ai-becomes-hitler-loving-sex-robot/> (accessed 2021-01-10)

2.1

俯瞰区分と研究開発領域
人工知能・ビッグデータ