2.1

2.1.5 意思決定・合意形成支援

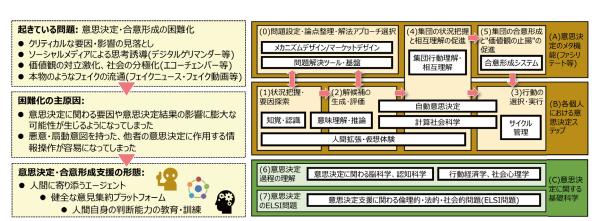
(1) 研究開発領域の定義

「意思決定」は、個人や集団がある目標を達成するために、考えられる複数の選択肢の中から一つを選択する行為である。その選択では個人の価値観がよりどころとなるが、集団の意思決定では、必ずしも関係者(メンバーやステークホルダー)全員の価値観が一致するとは限らない。関係者内で選択肢に関する意見が分かれたとき、その一致を図るプロセスが「合意形成」である。

情報爆発等による可能性の見落としやフェイクニュース等による悪意・扇動意図を持った思考誘導操作といった問題が顕在化し、意思決定ミスを起こしてしまうリスクが高まっている。このような問題・リスクを人工知能(AI)技術等の情報技術によって軽減し、個人・集団が主体性・納得感を持って意思決定できるように支援することを目指す研究開発領域である^{1),2)}。

(2) キーワード

意思決定、合意形成、意見集約、フェイクニュース、フェイク動画、デジタルゲリマンダー、インフォデミック、議論マイニング、マルチエージェント、自動交渉、メカニズムデザイン、計算社会科学、行動経済学、処方的分析



研究開発課題 (番号は右上図との対応)	これまでの取り組み	今後の課題・方向性
膨大な可能性の探索・評価(1)(2)	因果関係推論、マルチエージェントシミュレーション等	因果関係知識獲得、常識推論等
自動意思決定·自動交渉(0()2)(3)(5)	機械学習・最適化、自動交渉エージェント等	多様な価値観や不完全情報での意思決定等
大規模意見集約·合意形成 ⁽⁵⁾	討論型世論調査、ファシリテーションエージェント、メカニズムデザイン等	様々なタイプの合意形成問題への対応等
多様な価値観の把握・可視化(4)	言論マップ生成、議論マイニング、VR・ゲーミングによる追体験等	ディベートAI、法学AI等
フェイク対策(1)(6)(7)	ソーシャルネット分析、フェイク検出、ファクトチェック支援等	フェイク検出高度化、リテラシー教育等
意思決定に関する基礎科学(6)(7)	脳の意思決定メカニズム、行動経済学、認知バイアス等	ELSI・社会受容性、利用者から見た透明性等

関連動向

- AI・ビッグデータ技術による分析ビジネスは自動意思決定・意思決定支援を提供する処方的分析へ(Gartner)
- 意思決定問題は、古くから主に人文・社会科学の分野で取り組まれてきたが、今日、情報技術の関わりが増大
- AI(機械学習・最適化)が扱うのは評価関数が定まり解が1つ求まる問題、多様な価値観が混在する問題は新たな技術チャレンジ
- 欧米では、AIへの大型投資に加えて、学際的な意思決定研究が進んでいる(大規模意見集約、討論型世論調査、市民科学等)
- 米国では政治・ジャーナリズム分野でもAI技術との融合的な取り組みや議論が進展

図 2-1-8 領域俯瞰: 意思決定・合意形成支援

(3) 研究開発領域の概要

[本領域の意義]

我々は日々さまざまな場面で意思決定を行っている。クリティカルな場面での意思決定ミスは個人や集団の状況を悪化させ、その存続・生存さえも危うくする。例えば、企業の経営における意思決定ミスは、企業の業績悪化・競争力低下を招き、国の政策決定・制度設計における意思決定ミスは、国の経済停滞や国民の生活悪化にもつながる。また、個人の意思決定における判断スキル・熟慮の不足は、その個人の生活におけるさまざまなリスクを誘発するだけでなく、世論形成・投票等における集団浅慮という形で、社会の方向性さえも左右する。

情報技術が発展し、社会に浸透した今日、情報の拡散スピードが速く、膨大な情報があふれ、影響を及ぼし合う範囲が思わぬところまで広がっている。そのような意思決定の行為自体の難しさが増していることに加えて、意思決定の際のよりどころとなる価値観の多様化^{3), 4)} によって、合意形成の難しさも増している。さらには、価値観の対立から悪意・扇動意図を持った思考誘導の情報操作(フェイクニュース、フェイク動画等)まで行われるという問題も顕在化し^{5), 6)}、社会問題化している。2020年に世界を一変させたCOVID-19パンデミックでは、インフォデミック¹による社会混乱も発生した。

このような意思決定の困難化(意思決定ミスを起こすリスクの増大)という状況に対して、AI技術等の情報技術の活用は、問題のすべてを解決できるわけではなくとも、リスクを軽減し、状況を改善する手段になり得る。技術発展とともに急速に変化し、複雑化する社会環境にあっても、個人・集団が主体性・納得感を持って意思決定できるように、情報技術を用いて支援することが、上で述べたようなさまざまな意思決定場面で効果を生むと期待される。

また、集団での意思決定は、単に異なる意見の間の調整・交渉ではなく、多様な視点・考えからの集合知が期待でき、情報技術はその活性化・活用促進にも効果がある。

[研究開発の動向]

● 意思決定問題への取り組み

個人・集団の意思決定問題は古くから検討されてきた問題である。意思決定に関する先駆的な研究としては、1978年にノーベル経済学賞を受賞したHerbert A. Simonの取り組み^{7),8)} がよく知られている。Simonは意思決定プロセスを、(1)情報(Intelligence)活動、(2)設計(Design)活動、(3)選択(Choice)活動というステップで構成されるとした。(1)で意思決定に必要な情報を収集し、(2)で考えられる選択肢をあげ、(3)で選択肢を評価し、どれを選択するか決定する。これらのステップにおいて、必要な情報をすべて集めることができ、可能性のあるすべての選択肢をあげることができ、各選択肢を選んだときに起こり得るすべての可能性を列挙して評価することができるならば、合理的に最良の選択が可能になる。しかし、現実にはそのようなすべての可能性を考えて意思決定することはできず、人間が合理的な意思決定をしようとしても限界がある。このSimonが導入した「限定合理性」(Bounded Rationality)という概念は、意思決定に関する研究発展の基礎となった。Simonは、経営の本質は意思決定だと考え、

1 インフォデミック(Infodemic)は「情報の急速な伝染(Information Epidemic)」を短縮した造語で、正しい情報と不確かな情報が混じり合い、人々の不安や恐怖をあおる形で増幅・拡散され、信頼すべき情報が見つけにくくなるある種の混乱状態を意味する。

限定合理性を克服するための組織論も展開した。

そのように人間の判断・行動が必ずしも合理的になり得ず、心理・感情にも左右されるものであることを踏まえて、行動経済学が発展し、その中では意思決定に関わる興味深い知見が示されている。特に有名なのは、Simonの後、行動経済学の分野でノーベル経済学賞を受賞した二人、Daniel Kahneman(2002年受賞)とRichard H. Thaler(2017年受賞)の研究である。 Kahnemanは、直観的な「速い思考」のシステム1と論理的な「遅い思考」のシステム2というモデル 9)や、人間は利得面よりも損失面を過大に受け止めがちだといったプロスペクト理論 10)を提唱し、Thalerは、軽く押してやることで行動を促す「ナッジ」(Nudge)という考え方 11)を提唱した。

また、脳科学分野における脳の意思決定メカニズムの研究も進んでいる(詳細は「2.1.7 計算脳科学」を参照)。ドーパミン神経細胞の報酬予測誤差仮説等が見いだされ、モデルフリーシステムによる潜在的な意思決定と、モデルベースシステムによる顕在的な意思決定が協調および競合しつつ、人間の意思決定が動作していることがわかってきた¹²⁾。モデルフリーシステムは、事象と報酬との関係を直接経験に基づき確率的に結び付ける。モデルベースシステムは、事象と報酬との関係を内部モデルとして構築し、直接経験していないケースについても予測を可能にする。このような2通りのシステムはKahnemanのモデル(システム1・システム2)とも整合しており、意思決定が合理性だけによるものではないことの裏付けにもなる。

このような人文・社会科学分野や脳科学分野における意思決定に関する研究が、主に人間の側から掘り下げられてきた一方で、近年の情報技術の発展、Webやソーシャルメディアの普及は、意思決定を行う人間の環境を大きく変化させた。その結果、意思決定問題は新たな様相を呈するようになり、以前とは異なる困難さが生じている。今日、意思決定問題は情報技術との関わりが大きなものになっている。

2 意思決定問題の新たな様相・困難さ

本研究開発領域で取り組む意思決定・合意形成支援は、上で述べたような新たな様相・困難さに対処するために、情報技術を活用する。具体的な技術内容を説明する前に、新たに生じている困難さを示す事象 (問題)として顕著なものを4つ挙げる。

1つ目は、クリティカルな要因・影響の見落としの問題である。例えば、グローバル化したビジネス競争環境において、世界のあらゆる地域、思ってもいなかった業種から新たな競合が生まれ、想定していなかった法規制やソーシャルメディアで思わぬ切り口からの炎上も起こり得る。膨大な情報があふれ、社会がボーダーレス化した今日、意思決定に関連しそうな要因や意思決定結果の影響に膨大な可能性が生じ、人間の頭でそのあらゆる可能性をあらかじめ考えるのは極めて難しい。 Simon のいう限定合理性が極度に進み、問題として深刻化している状況である。

2つ目は、ソーシャルメディアによる思考誘導の問題である。 Web やソーシャルメディアを用いた情報発信・交流が広がり、それが人々の意思決定や世論形成に与える影響は無視できないものになっている $^{13), 14)$ 。 2016年の米国大統領選挙はその顕著な事例であり $^{5), 6)$ 、SNS(Social Networking Service)等のソーシャルメディアを用いた政治操作は「デジタルゲリマンダー」と呼ばれ $^{15)}$ 、フェイクニュースが社会問題化した。 SNS では、価値観が自分に近い相手としかつながらず、自分の価値観に沿った情報しか見ない、いわゆる「フィルターバブル」状態 $^{16)}$ に陥りやすいことも、SNSが思考誘導の道具になりやすい原因になっている。

3つ目は、価値観の対立激化、社会の分極化の問題である。集団の合意形成に難航し、対立が激化する傾向が強まっている。価値観の対立は古くから起こってきた事象だが、社会のボーダーレス化に伴う関係者

範囲の広がりや、SNSでの同調圧力やエコーチェンバー現象による意見同質集団の形成強化が、対立を強め、社会の分極化(Polarization)や政治的分断と言われる事態も引き起こされている $^{17),\,18)}$ 。

4つ目は、まるで本物のようなフェイク動画・画像の流通の問題である。前述のフェイクニュースは言葉 (SNSテキスト等)で伝達されるものが主であったが、深層学習を用いた敵対的生成ネットワーク (Generative Adversarial Network: GAN) 技術(GAN技術自体については「2.1.1 知覚・運動系の AI技術」を参照)によって、まるで本物のように見えるフェイク動画やフェイク画像が簡単に作れてしまうようになった(Deepfakes、Face2Face、FaceSwap等) (19)。特にフェイク動画は本物だと信じ込まれや すく、政治家や有名人の架空の発言・行為等を作るためにこれが悪用され、社会に流通すると、何が真実で何かフェイクか、真偽判断を見誤るリスクが増大し、さまざまな混乱が生まれると危惧される (17), 20), 21)。 さらに 2020 年には、まるで人間が書いたかのような自然なフェイク文章を生成することができる GPT-3 (22) というシステムも登場した。

以上の問題に見られるように、(1) 意思決定に関わる要因や意思決定結果の影響に、膨大な可能性が 生じるようになってしまったこと、(2) 悪意・扇動意図を持った、他者の意思決定に作用する情報操作が 容易になってしまったことが、意思決定の困難化の原因として顕著である。

3 意思決定・合意形成支援のための技術群

図 2-1-8の右上部に、個人・集団の意思決定プロセス(合意形成を含む)に対応させて、関連する技術群を示した。Simonの 3 ステップに相当する(B)各個人における意思決定ステップを中心に、(A)意思決定のメタ機能と(C)意思決定に関する基礎科学を上下に配置した 3 層構造で技術群を整理した 1, 2, 以下、これらを 6 つの技術群に分けて、取り組みの現状と今後の方向性について述べる 21, 23, 23, 23, 23, 23, 24, 23, 24, 23, 24, 23, 24, 25, 25, 25, 25, 26, 26, 27, 28, 29, 21, 29,

a.膨大な可能性の探索・評価

上記**②**に示した原因への対策としてまず求められるのは、意思決定に関わる要因や意思決定結果の影響における膨大な可能性を探索し、それらの組み合わせの中から目的に合うものを評価して絞り込む技術である。自然言語処理による因果関係推論 $^{24)}$ やマルチエージェントシミュレーションによる影響予測 $^{25)}$ 等の研究開発が進められており、具体的なシステムとして「なぜ?」「どうなる?」等の因果関係に関する質問応答を扱うことができる情報通信研究機構(National Institute of Information and Com-munications Technology:NICT)のWISDOM $X^{26)}$ が知られている。しかし、さまざまな分野・文脈で推論が行えるようにするには、常識を含め推論に必要な知識の獲得や、推論が成立する前提条件の精緻化等、取り組まなくてはならない技術課題がまだ多く残されている。

b. 自動意思決定・自動交渉

米国 Gartner 社は、データ分析の発展を記述的分析(Descriptive:何が起きたか)、診断的分析(Diagnostic:なぜそれが起きたか)、予測的分析(Predictive:これから何が起きるか)、処方的分析(Prescriptive:何をすべきか)という4段階で自動化が進むとし、4段階目の処方的分析は「意思決定支援」と「自動意思決定」という2通りがあるとしている 27)。この段階が進むほど、データ分析の顧客価値が高く、ビジネス上の競争も処方的分析へと進みつつある。「自動意思決定」はデータ分析の結果に基づき、何をすべきかというアクションまで自動決定するものであり、「意思決定支援」はアクションの候補を人間に提示し、どんなアクションを実行するかは最終的に人間が決定するものである。一見すると、意思決定支援より

も自動意思決定の方が、より発展したものであるかのように思えるが、現状、意思決定問題の性質が異なると考えるのが適切である。すなわち、コスト、精度、速度、売上等のような明確な指標(いわば価値観に相当)が定められ、それを評価関数・効用関数として合理的に解が一つ定められる意思決定問題は、機械学習・最適化等のAI技術を用いて「自動意思決定」が可能になる。それに対して、さまざまな価値観が混在している状況下、あるいは、価値観が不確かな状況下での意思決定問題は、最終決定に人間が関わる「意思決定支援」の形が基本になる。これに関しては、人間参加型(Human-in-the-Loop)のAI・機械学習が考えられている²⁸⁾。

自動意思決定には、強化学習や予測型意思決定最適化等、機械学習・最適化技術をベースとした方式 が開発・適用されている。強化学習(Reinforcement Learning)²⁹⁾は、学習主体が、ある状態で、あ る行動を実行すると、ある報酬が得られるタイプの問題を扱う機械学習アルゴリズムである。将来的により 多くの報酬が得られるように行動を選択する意思決定方策を、行動選択と報酬の受け取りを重ねながら学 習していく。囲碁で世界トッププロに勝利したGoogle DeepMindのAlphaGo³⁰⁾ で使われたことがよく 知られている。強化学習が適するのは、大量に試行錯誤することが可能な類の意思決定問題である。一方、 古典的なオペレーションズリサーチ(Operations Re-search:OR)で扱われているような類の意思決定 問題(例えば大規模システムの運用計画や小売業の商品価格設定戦略等)は、意思決定で失敗したときの ダメージが大きく、大量の試行錯誤は難しい。このような類の問題を、機械学習からの大量の予測出力(予 測が当たるかは確率的)に基づくOR問題とみなした新しいアプローチが予測型意思決定最適化31)である。 さらに、集団の意思決定を、異なる価値観(効用関数)を持ったエージェント間の交渉として定式化し た自動交渉技術の研究(詳細は[新展開・技術トピックス] ②を参照)も注目される。また、各個人がそ れぞれの効用関数を持つときに、集団として望ましい意思決定がなされるようにルールやプロトコルを定め るメカニズムデザイン^{32), 33)} のアプローチも盛んに研究されている。セキュリティーゲームとしての定式化 (警備 vs. 攻撃の確率・計算論的な問題設定)や自動メカニズムデザイン(大規模な制約最適化問題として の求解) 等が検討されている³⁴⁾ (これらを含むマルチエージェントシステムやメカニズムデザインの研究動 向については「2.1.3 エージェント技術」「2.3.4 メカニズムデザイン」を参照)。

c. 大規模意見集約・合意形成

上述の自動交渉は異なる価値観をもつ者の間の勝負という面があり、集団の意見集約・合意形成を目的とするならば、建設的な議論の進め方や相手への共感による価値観の変化といった面、および、そこでのファシリテーターの役割³⁵⁾ が重要なものになる。

集団の意見集約・合意形成のために情報技術を活用するシステムは、古くはグループウェアやCSCW (Computer Supported Cooperative Work) の研究分野での取り組みが見られる。例えば、Issue (課題、論点)をベースに木構造の表現でまとめるファシリテーション技法であるIBIS (Issue Based Information Systems) 法をグラフィカルに実現したgIBISという意思決定支援ツール 36 がよく知られている。一方、政治学の分野では、あるテーマについて回答を得る前に回答者にグループ討論をしてもらう討論型世論調査(Deliberative Poll) 37 が、熟議に基づく民主主義の方法論として有効だと認識されるようになった。

近年は集合知の収集・活用の学際的研究が進んでおり、米国マサチューセッツ工科大学(MIT)に2006年に設立されたMIT Center for Collective Intelligence(集合知研究センター:CCI)が注目される。インターネットを使った大規模な議論を、その論理構造の可視化によって支援するシステム Deliberatorium ³⁸⁾ や、

地球温暖化問題を取り上げて、解決プランを協議するシステム The Climate CoLab 等のプロジェクトを進めている。さらに、CCIのトップである Thomas W. Malone は、2018 年の著書 $^{39)}$ で、人間の集合知に AI との協働を含めた Superminds の方向性を示した。そのような方向で実際に伊藤孝行研究室 2 では、議論構造の可視化に加えて、エージェント技術によるファシリテーター機能を導入した大規模合意形成支援システム D-Agree を開発し、それを用いた社会実験にも取り組んでいる (詳細は[注目すべき国内外のプロジェクト] ②に記載)。

d.多様な価値観の把握・可視化

多様な価値観が混在する状況下での意思決定・合意形成に向けては、その状況や価値観の違いを可視化する技術が有効である。賛成・反対の各立場から意見と根拠を対比する言論マップ生成 40 、主張・事実等への言明とその間の関係(根拠・支持、反論・批判等)を推定する議論マイニング(Argumentation Mining) 41 、議題に対して賛成・反対の立場でディベートを展開するシステム(IBMの「Project Debater」、日立の「ディベートAIJ 42)等)が研究開発されている。より応用をフォーカスし、論理構築・推論を深める研究として法学 AI 43 , 44)もある。これらは自然言語処理技術を用いた手法だが、集団の相互理解促進のためにはVR(Virtual Reality)技術やゲーミング手法を用いて相手の立場を追体験させるアプローチも効果がある。

Project Debaterは、2018年6月に米国サンフランシスコで開催されたイベントWatson Westにて、イスラエルの2016年度ディベートチャンピオンとライブ対戦³し、「政府支援の宇宙探査を実施すべきか否か」という議題で勝利して話題となった。ニュース記事や学術論文を3億件収集・構造化して用いており、2011年に米国のクイズ番組Jeopardy!で人間のチャンピオンに勝利したIBM Watson⁴⁵⁾の自然言語処理に加えて、ナレッジグラフや議論マイニング等の技術が組み合わせて実現されたものと考えられる。

e.フェイク対策

ソーシャルメディア上での情報伝播の傾向や、そこで起きている炎上、フェイクニュース、エコーチェンバー、二極化等の現象を把握・分析すること $^{5)$, $^{13)}$, $^{14)}$, $^{18)}$, $^{20)}$, $^{46)}$, $^{47)}$ は、フェイク対策のための基礎的研究となる。フェイクニュースへの対抗としては、発信された情報が客観的事実に基づくものなのかを調査し、その情報の正確さを評価・公表するファクトチェックという取り組みが立ち上がっている $^{5)}$, $^{48)}$ 。ファクトチェックを行う団体として比較的早期に立ち上がった米国のSnopes(1994年~)やPolitiFact(2007年~)がよく知られている。この動きは世界的に広がっており、日本では2017年にFactCheck Initiative Japan(FIJ)が発足した。2015年には、こういった活動をつなぐInternational Fact-Checking Network(IFCN)も組織されている。しかし、大量に発信される情報を迅速にチェックするには人手では限界があり、コンピューター処理によってフェイクニュースの検出を効率化する試みが進められている(FIJでの取り組み 49)や2016年から始まった競技会 Fake News Challenge等)。また、フェイク動画・フェイク画像・フェイク音声等の判定については、オリジナルの動画・画像・音声から改ざんされていないか、当

- 2 2020年9月まで名古屋工業大学、10月から京都大学。
- 3 ライブ対戦の進行は、対戦する両者が議題の肯定派と否定派に分かれ、まず4分間ずつ主張を述べ、次に4分間ずつ相手の主張に対する反論を述べ、最後に2分間ずつまとめを述べるという形で行われ、その勝敗は聴衆の支持数で決まる。ディベートの議題は直前に与えられ、その場で相手の主張も踏まえつつ、自分の主張を組み立てることになる。

事者が実際に発話や行動をしていない虚偽の動画・画像・音声ではないか、といったことを動画・画像・音声の特徴分析によって判定することも行われている。フェイク検出技術の詳細は、[新展開・技術トピックス] ①で述べる。ただし、フェイクの検出技術とフェイクの作成法は往々にしていたちごっこになるため、技術開発だけでなく、メディアリテラシーの教育・訓練や、表現・言論の自由を損なわないように配慮しつつ法律・ルールの整備による対策も進めることが必要である^{5),20)}。

f. 意思決定に関する基礎科学

情報技術によって人間の意思決定を支援するにあたって、そもそも人間の意思決定とはどういうものか、どうあるべきかを理解しておくことは重要である。既に言及した通り、行動経済学や脳科学の分野(「2.1.7 計算脳科学」を参照)で意思決定プロセスのモデルやメカニズムが研究されてきた。また、社会心理学・認知科学等の分野で研究されている確証バイアスを含む認知バイアス⁵⁰⁾ も意思決定に大きく関わる。加えて、人間の意思決定・合意形成を支援する機能がELSI(Ethical, Legal and Social Is-sues:倫理的・法的・社会的課題)の視点から適切であるかについても常に考えておかねばならない。

(4) 注目動向

[新展開・技術トピックス]

フェイク検出技術

フェイクニュース検出は次の4つの面から試みられている⁴⁷⁾。1つ目は知識ベース検出方式で、従来の人手によるファクトチェックを強化するように、クラウドソーシング的な仕組みを使って専門家集団に検証してもらったり、あらかじめ蓄積された知識ベースと自動照合したりする取り組みがある。2つ目はスタイルベース検出方式で、誤解を生みやすい見出し表現や欺くことを意図したような言葉使い等に着目する。3つ目は伝播ベース検出方式で、情報拡散のパターン(情報伝播のグラフ構造やスピード等)に着目する。例えば、フェイクニュースは通常ニュースよりも速く遠くまで伝わる傾向があることが知られている。4つ目は情報源ベース検出方式で、ニュースの出典・情報源の信頼性やその拡散者の関係等から判断する。社会環境・文脈等によって真偽の捉え方が変わるし、科学的発見によって真理の理解が変わることもあり、真偽が定められない言説も多いため、最終的には人間による判断が不可欠だが、上に示したような技術は怪しいニュース・情報を迅速に絞り込むのに有効である。また、1件のニュース単独で真偽判定するよりも、複数の情報の間の関係比較や整合性判断、および、複数の視点からのチェックを行う方がより確かな判断が可能になる。

また、動画・画像・音声等がオリジナルから改ざんされていないか、当事者が実際に発話や行動をしていない虚偽の動画・画像・音声ではないか、といったことを動画・画像・音声の特徴分析、ニューラルネットワーク・機械学習を用いて判定したり、改ざんの箇所や方法を特定したりといった技術が開発されている^{21),51),52)}。例えば、不自然なまばたきの仕方、不自然な頭部の動きや目の色、映像から読み取れる人間の脈拍数、映像のピクセル強度のわずかな変化、照明や影などの物理的特性の不自然さ、日付・時刻・場所と天気の整合等が手掛かりになる。フェイクの検出技術とフェイクの作成法はいたちごっこだとも言えるが、人間の目・耳では見分けがつかないレベルのフェイク動画・画像・音声が作れてしまう事態において、コンピューターによる分析は不可欠である。さらに、動画・画像・音声の内容解析とは別に、ブロックチェーンを使って履歴を管理することで、改ざんが入り込むことを防ぐという方法もある。

2 自動交渉

自動交渉は、それぞれの効用関数(いわばそれぞれの価値観)を持った複数の知的エージェント(AIシ ステム)が相対する状況において、一定の交渉プロトコルに従ってうまく合意案を見つける技術である(マ ルチエージェントシステムの考え方がベースにある: 「2.1.3 エージェント技術」を参照)。ある問題について、 複数のステークホルダーの間で対立したり、協調しようとしたりするとき、交渉プロトコルや効用関数を定 めてエージェントに代行させて自動交渉を行うならば、人間同士が交渉するよりも、その条件で考えられる 最適な合意点に高速に到達できると期待される。効用関数の複雑さやプライバシーの扱い、合意形成でき るか、目的関数を達成できるか等の保証、交渉の論点数やエージェント数のスケーラビリティーの要件等、 いろいろな問題設定で検討されている。2010年からは毎年、国際自動交渉エージェント競技会ANAC (Automated Negotiating Agents Competition) が開催されており⁵³⁾、これを共通の場として技術発 展が進んできた。交渉理論を踏まえたモデルベースの方式のほか、機械学習ベースの方式も検討されている。 自動交渉技術のアプリケーション開発も進められている。国内では、産業競争力懇談会(COCN)にお いて、「人工知能間の交渉・協調・連携」が2016年度・2017年度に調査・検討テーマに取り上げられ、 NECをリーダー企業として20前後の企業・組織が協力して提言をまとめた54)。その中では、製造バリュー チェーン、交通・人流、電力・水、自動運転車・移動体という4つの産業分野を中心に交渉・協調・連携 技術のユースケースも示された。その後、具体的な開発フェーズに入り、新エネルギー・産業技術総合開 発機構(NEDO)のファンドによる「ロボット・ドローンが活躍する省エネルギー社会の実現プロジェクト」 (2017年度~2021年度)では、上記技術を用い、多数の移動体(ドローン)の安全飛行を可能にする衝 突回避技術と運航管理システムの研究開発が行われた。2019年10月に福島ロボットテストフィールドにて、 同一空域で複数事業者の運航管理システム相互接続試験(ドローン事業者29者が参加)に成功した。また、 内閣府の戦略的イノベーション創造プログラム(SIP)「ビッグデータ・AIを活用したサイバー空間基盤技術」 (2018年度~2022年度)では、「AI間連携によるバリューチェーンの効率化・柔軟化」をテーマとし、 受発注会社間での商取引条件の調整を支援・自動化するAI間連携基盤技術の開発が進められている。海 外では、Pactum 社が開発した、取引条件交渉から契約文書の取り交わしまで一気通貫のソリューション を提供する自動交渉AIが注目される。ウォルマートが試験採用中で、自動交渉AIはロングテールに該当す るサプライヤーに対して、オンラインチャットボットで交渉を行う。その結果に基づいて必要な後段処理も 自動実行し、交渉結果から学習もする。

[注目すべき国内外のプロジェクト]

1 DARPA の Cognitive Security 関連プロジェクト

フェイクニュースに見られるように、悪意を持ったオンラインやオフラインでの誘導・干渉によって人々の 思考や行動に影響を与える問題は、Cognitive Securityと呼ばれる分野の中心課題の一つである。米国 国防高等研究計画局(DARPA)では、これが国家安全に関わる問題だと考え、以下のようなさまざまな 研究開発プロジェクトを推進している。

- ・Media Forensics (MediFor):画像・動画の改ざん検知
- · Brandies: 個人情報・プライバシーが目的外に使われないように管理
- · Semantic Forensics (SemaFor):画像・動画の意味的不整合・フェイクの検知
- ・Active Interpretation of Disparate Alternatives (AIDA): 状況を理解し、アクションするため、さまざまな情報源の間の矛盾・整合を踏まえながら仮説を生成

- ・Active Social Engineering Defense (ASED): 人間の心理的な隙や行動のミスにつけ込んで個人が 持つ秘密情報を入手する攻撃(ソーシャルエンジニアリング)の検知・防御
- ・Harnessing Autonomy for Counering Cyberadversary Systems(HACCS):悪意のあるボットネットワークや大規模マルウェアに対抗する自律ソフトウェアエージェント

2 大規模合意形成支援システム D-Agree

[研究開発の動向] ③ cで述べた大規模意見集約・合意形成に取り組む注目プロジェクトとして、伊藤孝行研究室で開発された大規模合意形成システム D-Agree 55)が挙げられる。 D-Agree はクラウドベースで、場所・時間の制約なしに多人数がオンライン参加でき、多人数からさまざまな意見を集めつつ、建設的な議論・合意形成が進められるように、エージェント技術・自然言語処理技術・機械学習技術等を用いて、(a)議論内容を抽出・分析・構造化して可視化する機能、(b)炎上防止フィルター機能、(c)自動ファシリテーション機能等を提供している。伊藤研究室で2010年から2015年頃に開発された同種のシステム Collagree 56)では、(a)(b)は提供されていたが、(c)ではなく人間が行うファシリテーションの補助機能が提供されていた。 D-Agree は (c)の自動ファシリテーションを行うエージェントを導入したことが最大の特長である。Collagree、D-Agreeともに名古屋市のタウンミーティング等で社会実験に適用されてきたが、D-Agree は海外(アフガニスタンのカブール市等)にも展開されている。名古屋工業大学の大学発ベンチャー育成の仕組みを使ってAgreeBit社も起業された。

(5) 科学技術的課題

まず、[研究開発の動向] の項に挙げたような意思決定・合意形成支援に関わるさまざまな要素技術の研究開発は、それぞれさらなる研究開発が必要である。それに加えて、それらの要素技術を活用・統合して、個人・集団の意思決定・合意形成を支援する機能・サービスを、具体的なシステムとして実現する取り組みが求められる。例えば、以下に示すような支援形態 ① ② ③ が考えられる 1),2)。ただし、これらは一例であり、どのような機能・サービスが支援形態として有効かということ自体が研究課題である。

● 人間に寄り添うエージェント

意思決定の困難化の原因(1)として、意思決定に関わる要因や意思決定結果の影響に膨大な可能性が 生じるようになってしまったことを挙げた。この問題に対して、膨大な可能性を探索することはコンピューターが得意なタスクであることから、例えば、人間(個人や集団)に寄り添うエージェント型の支援形態が 考えられる。近年、スマートフォンの音声対話インターフェースや家庭に置かれたスマートスピーカー等、 エージェントインターフェースのアプリケーションが広がりつつある。その将来的な拡張機能として、個人 向けエージェントならば、個人の意思決定場面で膨大な可能性の中から適切な選択肢をアドバイスしてくれ たり、集団向けエージェントならば、議論・交渉の状況をモニタリングしていて、有望な合意点や新たな観 点を提示してくれたりといった支援機能が考えられる。

2 健全な意見集約プラットフォーム

意思決定の困難化の原因(2)として挙げたのは、悪意・扇動意図を持った、他者の意思決定に作用する情報操作が容易になってしまったことである。情報技術による対策には限界があり、多面的に取り組む必要がある問題だが、Web・SNS等の情報システムをプラットフォームとして用いた意見集約(意見の発信・

収集・投票等)のケースは、情報技術による対策が考えられる。健全な意見集約を行えるように、「悪意・ 扇動意図を持った、他者の意思決定に作用する情報操作」を検出・防止する機能やメカニズムが組み込ま れたプラットフォームであってほしい。今後期待される対策機能としては、例えば、フェイクニュースやデマ 等を検出・排除する機能、ニュース等の一次情報や意見の根拠を追跡・確認する機能、声の大きい意見だ けでなく公平に意見を集める機能、ある意見について異なる立場の意見も併記・比較する機能、投票や意 見集約プロセスにおいて不正直な申告や裏工作の効果がなくなる(正直申告が最良となる)メカニズム等 が考えられる。

人間自身の判断能力の教育・訓練

意思決定の困難化の2つの原因に対する直接的な支援形態は上記①②の通りだが、もう一つ考えておくべき支援形態として、人間自身の判断力を高めるような教育・訓練がある。なぜならば、上に述べたような支援機能によって意思決定・合意形成が楽なものになると、人間は支援機能に頼り、自分自身で考えなくなっていき、人間の判断能力が低下してしまうと懸念されるからである。また、フィルターバブルに陥りやすいか否かには個人差があるものの、誰しもある程度の確証バイアスを持つことは避けられない。そこで、意思決定・合意形成支援の一環で、人間自身の判断能力の教育・訓練という側面も考えておきたい。その実現形態としては、判断能力の教育・訓練用ツールのような直接的な実装だけでなく、①や②に示したような支援機能を利用する過程で、人間自身が深く考えるように促したり、より多面的な検討を促したりするといった組み合わせ型の実装も考えられる。

(6) その他の課題

● ELSI および社会受容性に配慮した研究開発

人間の意思決定・合意形成を支援する機能が、倫理的・法的・社会的な視点(ELSI面)から適切であるかを常に考えておかねばならない。例えば、人間の支援機能を意図したものが、思考誘導や検閲(表現・言論の自由の制限)と受け止められてしまう可能性もある。逆に、フェイク問題のようなケースに対しては法的規制をかけてしまえばよいのではないかという意見を聞くことがあるが、絶対的な真偽が定まらない言説は非常に多く、法的規制が強く働くと表現・言論の自由が制限されるリスクが高まることに注意を要する²¹⁾。この点を踏まえて、法的規制の検討は、極めて慎重に行う必要がある。その一方、人間のメディアリテラシーを高める教育が重要である。

利用者から見た透明性を確保し、社会受容性に配慮した技術開発が求められる。そのためには、実社会の具体的な問題に適用して社会からのフィードバックを受けるプロセスを、短いサイクルで回しながら判断・ 改良していくのがよいと考える。

2 分野横断の研究開発体制・推進施策

本研究開発領域は、AI技術等の情報技術だけでなく、計算社会科学、脳科学、認知科学、心理学、経済学、政治学、社会学、法学、倫理学等が重なる学際的な領域であり、分野横断の研究開発体制・推進施策が必要である。そのためには、初期段階から分野横断で研究者を共通の問題意識・ビジョンのもとに束ねる研究開発マネジメントが望ましい。現状、本研究開発領域の個々の技術課題・要素技術に関わる研究者は多いものの、研究者それぞれの取り組みは全体の問題意識に対してまだ断片的なものにとどまっている感が強く、分野横断の連携・統合による骨太化が求められる。その際、情報技術側で扱いやすい形の問

題にしてしまうとか、人文・社会科学側から結果に対して駄目出しするとかではなく、具体的な問題に対する定式化において双方がコミットすべきである。実社会への適用において発生するさまざまな制約事項を、アルゴリズム・原理のレベルで扱うのか、運用上の制約(法規制等)の形で扱うのかによって、技術的なアプローチは変わってくる。

3 国・社会の Cognitive Security に対する意識向上

米国が国家安全保障の観点から重要な研究開発領域と位置付けて投資しているのに対して、日本ではその意識が弱い。日本は米国の事例ほど、フェイク問題や社会分断が深刻化していないため、国・社会の危機感が薄いように思われるが、民主主義を揺るがし得る、社会の方向性を左右し得る、国・組織・個人に対する新しいサイバー攻撃になり得る、といった国・社会にとっての大きなリスクが生じることに備えておくべきである。フェイクを含む Cognitive Security を脅かす攻撃に対する防御技術を育てておくことや、人々のメディアリテラシーを高めるための教育や啓蒙施策等を進めることを通して、健全な社会的意思決定・集合知を育てる意識・環境が、安全で信頼できる社会を発展させていくために極めて重要である。

(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	0	7	マルチエージェントシステムの分野で、オークション・マッチングの理論 研究やインセンティブメカニズムの研究が多い。
	応用研究・開発	0	7	大規模合意形成支援システム等で先端的な取り組みや、AI間の交渉・協調・連携に関するCOCNの取り組みが進展している。
米国	基礎研究	0	7	MIT CCIのDeliberatoriumやThe ClimateCoLab、Stanford Universityの討論型世論調査をはじめ、学際的な基礎研究が根付いている。 AI・マルチエージェントシステムの分野で、メカニズムデザイン、オークションやマッチングの理論研究が広く行われている。
	応用研究・開発	0	7	上記基礎研究がそのまま応用研究やベンチャーによる産業化につながる傾向が強い。国および企業によるAI分野への大型投資が行われている(Facebookの自動交渉エージェント等)。
欧州	基礎研究	0	7	Imperial College London、Oxford University、Delft University of Technology 等、自動交渉の基礎研究が強く、論理的なアプローチによる自動交渉の研究もおこなわれている。
	応用研究・開発	0	7	市民からの意見集約や合意形成のためのシステム・応用に盛んに取り組まれている。自動交渉の応用ソフトウェア(電力売買等)への取り組みも見られる。
中国	基礎研究	0	7	Hong Kong Baptist Universityのメカニズムデザインや自動交渉の基礎理論研究をはじめ、取り組みが活発になってきている。
	応用研究・開発	Δ	\rightarrow	顕著な活動は見当たらない。
韓国	基礎研究	Δ	\rightarrow	顕著な活動は見当たらない。
	応用研究・開発	Δ	\rightarrow	顕著な活動は見当たらない。

(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

〇: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) 科学技術振興機構 研究開発戦略センター,「戦略プロポーザル:複雑社会における意思決定・合意形成を支える情報科学技術」,CRDS-FY2017-SP-03 (2018年12月).
- 2) 福島俊一,「複雑社会における意思決定・合意形成支援の技術開発動向」,『人工知能』(人工知能学会誌) 34巻2号(2019年3月), pp. 131-138.
- 3) Edmond Awad, et al., "The Moral Machine experiment", *Nature* Vol. 563 (24 October 2018), pp. 59-64. DOI: 10.1038/s41586-018-0637-6
- 4) 亀田達也, 『モラルの起源-実験社会科学からの問い』(岩波書店,2017年).
- 5) 笹原和俊,『フェイクニュースを科学する一拡散するデマ,陰謀論,プロパガンダのしくみ一』(化学同人,2018年).
- 6) 湯淺墾道,「米大統領選におけるソーシャルメディア干渉疑惑」,『情報処理』(情報処理学会誌) 58巻 12号(2017年12月), pp. 1066-1067.
- 7) Herbert A. Simon, *Administrative behavior: a study of decision-making processes in administrative organization* (Macmillan, 1947). (邦訳: 二村敏子・桑田耕太郎・高尾義明・西脇 暢子・高柳美香訳,『新版 経営行動: 経営組織における意思決定過程の研究』,ダイヤモンド社,2009).
- 8) Herbert A. Simon, *The Sciences of the Artificial* (The MIT Press, 1969). (邦訳: 稲葉元吉・吉原英樹訳、『システムの科学 第3版』、パーソナルメディア、1999年)
- 9) Kahneman, D., *Thinking, Fast and Slow, Farrar* (Straus and Giroux, 2011). (邦訳:村井章子訳, 『ファスト&スロー:あなたの意思はどのように決まるか?』,早川書房,2014年)
- 10) Daniel Kahneman and Amos Tversky, "Prospect Theory: An Analysis of Decision under Risk", *Econometrica* Vol. 47, No. 2 (March 1979), pp. 263-291.
- 11) Richard H. Thaler and Cass R. Sunstein, *Nudge: Improving Decisions About Health, Wealth, and Happiness* (Yale University Press, 2008). (邦訳:遠藤真美訳,『実践 行動経済学』,日経 BP 社,2009年)
- 12) 坂上雅道・山本愛実,「意思決定の脳メカニズム―顕在的判断と潜在的判断―」,『科学哲学』(日本科学哲学会誌) 42-2号(2009年).
- 13) 遠藤薫、『ソーシャルメディアと〈世論〉形成』(東京電機大学出版局,2016年).
- 14) 田中辰雄・山口真一, 『ネット炎上の研究』(勁草書房,2016年).
- 15) 金子格・須川賢洋(編),「小特集 ディジタルゲリマンダとは何か一選挙区割政策からフェイクニュースまで」, 『情報処理』(情報処理学会誌)58巻12号(2017年12月), pp. 1068-1088.

- 16) Eli Pariser, *The Filter Bubble: What the Internet is Hiding from You* (Elyse Cheney Literary Associates, 2011). (邦訳:井口耕二訳,「閉じこもるインターネット」,早川書房,2012年)
- 17)「Politics and Technology テクノロジーは民主主義の敵か?」,『MITテクノロジーレビューSpecial Issue』Vol. 11(2018年).
- 18) 田中辰雄・浜屋敏, 「ネットは社会を分断するのか-パネルデータからの考察-」, 『富士通総研 経済研究所 研究レポート』No. 462 (2018年).
- 19) Ruben Tolosana, et al., "DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection", arXiv: 2001.00179 (2020).
- 20) 「特集:フェイクニュース」, 『DIAMONDハーバード・ビジネス・レビュー』2019年1月号, pp. 16-82.
- 21) 科学技術振興機構 研究開発戦略センター,「公開ワークショップ報告書: 意思決定のための情報科学 ~情報氾濫・フェイク・分断に立ち向かうことは可能か~」,CRDS-FY2019-WR-02 (2020年2月).
- 22) 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) .
- 23) 科学技術振興機構 研究開発戦略センター,「科学技術未来戦略ワークショップ報告書 複雑社会における意思決定・合意形成を支える情報科学技術」,CRDS-FY2017-WR-05 (2017年10月).
- 24) 井之上直也,「言語データからの知識獲得と言語処理への応用」,『人工知能』(人工知能学会誌) 33巻 3号(2018年5月), pp.337-344.
- 25) 服部宏充・栗原聡,「エージェント研究におけるシミュレーション」,『人工知能』(人工知能学会誌) 28 巻 3号(2013年5月), pp. 412-417.
- 26) 水野淳太・他,「大規模情報分析システム WISDOM X, DISAANA, D-SUMM」,『言語処理学会第 23 回年次大会発表論文集』(2017年), pp. 1077-1080.
- 27) Yannick de Jong, "Levels of Data Analytics", *IThappens.nu* (20 March 2019) . http://www.ithappens.nu/levels-of-data-analytics/ (accessed 2020-11-08)
- 28) 鹿島久嗣・小山聡・馬場雪乃、『ヒューマンコンピュテーションとクラウドソーシング』(講談社,2016年).
- 29) 牧野貴樹・澁谷長史・白川真一(編著),他19名共著,『これからの強化学習』(森北出版,2016年).
- 30) 大槻知史(著)・三宅陽一郎(監修),『最強囲碁 AI アルファ碁 解体新書(増補改訂版)』(翔泳社, 2018年).
- 31) 藤巻遼平・他, 「予測から意思決定へ ~予測型意思決定最適化~」, 『NEC 技報』69巻1号(2016年), pp. 64-67.
- 32) 横尾真・他,「メカニズムデザイン(基礎編)」,『コンピュータソフトウェア』(日本ソフトウェア科学会誌) 29巻4号(2012年11月), pp. 15-31.
- 33) 横尾真・他,「メカニズムデザイン(応用編)」,『コンピュータソフトウェア』(日本ソフトウェア科学会誌) 30巻1号(2013年2月), pp. 34-52.
- 34) 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.
- 35) 桑子敏雄,『社会的合意形成のプロジェクトマネジメント』(コロナ社,2016年).
- 36) Jeff Conklin and Michael L. Begeman, "gIBIS: a hypertext tool for exploratory policy discussion", *Proceedings of the 1988 ACM conference on Computer-supported cooperative work*

- (CSCW'88; Portland, USA, 26-28 September 1988), pp. 140-152. DOI: 10.1145/62266.62278
- 37) James S. Fishkin, When the People Speak: Deliberative Democracy and Public Consultation (Oxford University Press, 2011). (邦訳:岩木貴子訳,曽根泰教監修,『人々の声が響き合うとき: 熟議空間と民主主義』,早川書房,2011年)
- 38) Mark Klein, "Enabling Large-Scale Deliberation Using Attention-Mediation Metrics", Computer Supported Cooperative Work Vol. 21, No. 4-5 (2012), pp. 449-473. DOI: 10.2139/ssrn.1837707
- 39) Thomas W. Malone, *Superminds: The Surprising Power of People and Computers Thinking Together* (Little, Brown and Company, 2018) .
- 40) 水野淳太・他、「言論マップ生成技術の現状と課題」、『言語処理学会第17回年次大会発表論文集』(2011年), pp. 49-52.
- 41) 岡崎直観,「自然言語処理による議論マイニング」,『人工知能学会全国大会(第32回)』(2018年) 1D2-OS-28a-a (OS-28招待講演). https://www.slideshare.net/naoakiokazaki/ss-100603788 (accessed 2021-02-13)
- 42) 柳井孝介・他,「AIの基礎研究:ディベート人工知能」,『日立評論』98巻4号(2016年4月), pp. 61-64.
- 43) 佐藤健, 「論理に基づく人工知能の法学への応用」, 『コンピュータソフトウェア』(日本ソフトウェア科学会誌) 27巻3号(2010年7月), pp. 36-44. DOI: 10.11309/jssst.27.3_36
- 44) 日本学術会議公開シンポジウム「AIによる法学へのアプローチ」(2019年1月24日). http://research.nii.ac.jp/~ksatoh/ai-law-symposium/(accessed 2021-02-13)
- 45) "Special Issue: This is Watson", *IBM Journal of Research and Development* Vol. 56 issue 3-4 (May-June 2012).
- 46) Robert M. Bond, et al., "A 61-million-person experiment in social influence and political mobilization", *Nature* Vol. 489 (12 September 2012), pp. 295-298. DOI: 10.1038/nature11421
- 47) Xinyi Zhou and Reza Zafarani, "A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities", *ACM Computing Surveys* Vol. 53, No. 5 (September 2020), Article No. 109. DOI: 10.1145/3395046
- 48) 立岩陽一郎・楊井人文、『ファクトチェックとは何か』(岩波書店,2018年).
- 49) Tsubasa Tagami, et al., "Suspicious News Detection Using Micro Blog Text", *Proceedings* of the 32nd Pacific Asia Conference on Language, Information and Computation (PACLIC 32; Hong Kong, 1-3 December 2018).
- 50) 鈴木宏昭、『認知バイアス:心に潜むふしぎな働き』(講談社,2020年).
- 51) Darius Afchar, et al., "MesoNet: a Compact Facial Video Forgery Detection Network", Proceedings of IEEE International Workshop on Information Forensics and Security (WIFS 2018; Hong Kong, 11-13 December 2018) . DOI: 10.1109/WIFS.2018.8630761
- 52) Huy H. Nguyen, Junichi Yamagishi and Isao Echizen, "Capsule-forensics: Using Capsule Networks to Detect Forged Images and Videos", *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing* (ICASSP 2019; Brighton, 12-17 May 2019). DOI:

10.1109/ICASSP.2019.8682602

- 53) 藤田桂英・森顕之・伊藤孝行、「ANAC: Automated Negotiating Agents Competition(国際自動 交渉エージェント競技会)」、『人工知能』(人工知能学会誌)31巻2号(2016年3月)、pp. 237-247.
- 54) 産業競争力懇談会 (COCN),「人工知能間の交渉・協調・連携」、「産業競争力懇談会 2017 年度プロジェクト最終報告」(2018年2月21日).
- 55) Takayuki Ito, et al., "D-Agree: Crowd Discussion Support System Based on Automated Facilitation Agent", *Proceedings of the 34th AAAI Conference on Artificial Intelligence* (AAAI-20; New York, 7-12 February 2020), pp. 13614-13615. DOI: 10.1609/aaai.v34i09.7094
- 56) 伊藤孝行・他、「エージェント技術に基づく大規模合意形成支援システムの創成 一自動ファシリテーションエージェントの実現に向けて一」、『人工知能』(人工知能学会誌) 32巻5号(2017年9月), pp. 739-747.