The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

速度耐性をもつ社会の実現にむけた基礎的考察

Fundamental consideration on future society with speed tolerances

山川 宏[*1*2]
Hiroshi Yamakawa

[*1] ㈱ドワンゴ
Dwango, Co. ltd.

[*2] NPO 法人全脳アーキテクチャ・イニシアティブ
The Whole Brain Architecture Initiative, non-profit organization

Technological progress is accelerating, so the relationship between people and organizations is becoming more complicated. Under these circumstances, in order to maintain social stability, it is necessary to adapt social rules more quickly than ever. For such adaptive control social decisions need to be made quickly, but resources that can be used and speeds required for social decisions are limited. Therefore, within a certain range, it is necessary to delegate the right to decide social rules to AI. In the AI responsible for this role, the way of designing the value function used for decision making is critically important.

1. はじめに

人工知能(AI)の高度化に伴い、それ自身およびそのネットワーク化が社会に与える影響が増大するため、それに関わる議論が盛んになりつつあり、国内では様々な省庁が関連する会議などが動きを早めてきた。国際的には米国 Future of Life Institute によるアシロマの原則[FLI 2017]、米国 IEEE が発行した Ethically Aligned Design [IEEE 2017]などが有名である。

こうした問題においては具体的な問題に対処することは重要である。一方で技術に牽引される形で変化速度が増大していることへの対処は重要であり、これをなおざりにすれば、最終的に個別問題の対処が追いつかないことで破綻する恐れがある。しかしながら、速度それ自体への対応についての議論は未だ不十分であると思われる。

本稿では、まず、現在において速度が速まっている理由を述べ、次に社会におけるルールの必要性について述べる。その上でルールづくりのための社会的決定に合意のための時間を要する点について考察する。この状況を解消するには主に緊急時において AI が社会的決定を代行する必要があることを述べる。そして AI が迅速に動作するためには、社会が共有する価値観の実装が必須であり、その設計に難しさがあることを述べる。さらに社会における上位の目標を設定できれば環境変化への対応性が高まるが、進化的に獲得された人々の価値観と乖離する可能性が高まることを示唆する。

2. なぜ速度が増しているのか –収穫加速の法則-

生命は、複製する個体という形により、長い時間を超えて存続してきた。個体として生存し、繁殖するために、自身の身体を含めて制御下にあるリソースの量や範囲を拡大しようとする傾向をもつ。ここではそうした傾向を生存志向と呼ぶことにする。少なくとも人間はこうした生存志向を欲求として自覚している。

人や動物などのエージェントは、外界をセンシングしアクションとして働きかけることで、自らの生存志向を実現しようとする。環境が複雑になれば、その環境内の様々な出来事について予測を行える知能を持つほうが効果的に生存志向を実現しうる。特に人などの動物においては、発達の期間を長くすることで、多様な環境へ適応する知能を進化させてきた。

連絡先：山川宏、所属： ドワンゴ ドワンゴ人工知能研究所、
NPO 法人全脳アーキテクチャ・イニシアティブ

人類社会は以上のように知能を用いて科学技術を発展させて繁栄し人口を増やし、指数関数的に変化のスピードを増していると言われている。特に一つの重要な発明は他の発明と結びつき、次の重要な発明の登場までの期間を短縮し、イノベーションの速度を加速する収穫加速の法則が知られている。最近は情報技術や AI 技術の進展がこの変化の主役となってきている。

本来、生命の存続にとっては、その周囲の環境が安定していることが望ましいが、逆に変化が助長されてきている。現代の資本主義社会で活動する企業というエージェントは社会や技術の変化から生じた新たな市場において経済的利益を確保し拡大しようとする。こうした傾向は、個人レベルでのキャリア形成においても同様に生じている。環境に変化があるとき、それまで日陰者であったプレイヤーが表舞台に躍り出ることは珍しくない。

このようにエージェント同士が競争する変化の激しい世界においては、効果的に知能を活用できるエージェントのほうが有利になりやすい。優れた知能を持てば、様々な変化に対して素早く適応しうるし、変化の兆候を早い段階で見つけ出しやすくなる。さらに、自らが変化を生み出すゲームチェンジャーになれば益々有利になる。こうした変化を好み生み出そうとする傾向が高まり、現代の社会において変化を加速する要因となっている。

3. 迅速に社会的ルールを制御する必要性

3.1 社会的ルールの必要性

環境の広さに対して同種エージェントが疎に存在するなら、そこで生ずる相互作用は少なく、個々のエージェントが対処すべき問題の対象は主に環境そのものや外敵である。

我々人類は科学技術を発展させて生産性を高め人口を増加させてきた。さらに文字、輪転機、インターネットなどの通信技術を発展させて知識の共有と蓄積を促進した。その結果として現代社会においては、多くの人間同士が高密度な相互作用をもつことで更に知能を高める状態に急速に変化してきている。こうなると、同種内での生存志向が衝突し競争が激化する。

そこで人類は、競争を制御しながら社会を安定させ、むしろ発展につなげるように社会的ルールを発展させた。なお本稿の議論において、社会的ルールとは、エージェント間の調整に関わる様々な仕組みを含む広義なものである。公平性・利他性などを含む倫理観としては、生来持っている性質と教育された性質を含む。また罰則を含む明文化／体系化された法律も含む。

さらには現代主流となっている貨幣経済・市場経済・資本主義・民主主義やその対立概念を含むルールである。

現在は、法制度の範囲が拡大しているが、これらは倫理的な価値観から導かれる最低限守るべき規範であると考えられる。少なくとも公平性などの倫理的概念は、類人猿にも存在するため、法制度以前に存在する倫理観があるだろう。また憲法学者の木村草太は法律とは「人が幸福に生きるためのルールを一般的、抽象的に定めたもの」とも述べている。

3.2 社会的ルールを制御する社会的決定

社会的ルールは社会的決定によって確定されるべきものであり、実際に法律であれば立法という形で制定される。

民主主義社会における社会的決定において公平性や納得性を確保する必要がある。今回は、図1に示すよう、社会的決定について急用度と重要度の観点から整理した。

重要度が高い事案は広く民主的な合意形成の上で決定されるべきである。その場合には最終合意案に対し、できるだけ多くの関係者から納得を得るには、議論を尽くすために時間を要する。しかし意見が対立する関係者間において根本的な利害や価値観の隔たりがあれば、優れた選択肢のいずれでであっても、関係者全員が納得することは難しく、社会的決定にいたるまでの時間は長くなりがちである。

この様に民主的な決定は時間と労力を要するため数時間程度で結論を出さざるを得ない案件は権利を移譲された担当チーム等が決定を行う。また急用度が低くても重要度が低い事案はやはり担当チーム等が決定を行うことでスループットをあげている。急用度が中程度か、もしくは、急用度も重要度も低かった場合には、権限を担当チームなどに移譲して決定をおこなうことが現実的である。

一方で、人間の認知速度を超えた状況で問題が起きた場合には、そもそも速すぎるために社会的決定を人間が行うことができないレイテンシの問題が発生する。こうした重要度と急用度が高い事態であっても、津波などのようにある程度想定が為され、事前のシミュレーションなどを参考に対策についての社会的決定を行うことで急用度を下げられる。

		重要度	
		低い	高い
急用度	低い（月以上）	担当チーム等に移譲	民主的な方法で対応
	中程度（数時間）	担当チーム等に移譲	担当チーム等に移譲
	高い（秒以下）	速過ぎて人による決定不能	速過ぎて人による決定不能

図1：速度に応じた人による社会的決定の分類案

いずれにしても技術によって社会的状況が速く変化すれば、それに伴い社会的ルールを制御するための社会的決定を迅速に行う必要がある。

3.3 社会的決定のオーバフロー（スループット問題）

急用度が中程度で重要度が高い場合は、本来は民主的に決定すべきであるので、この場合に権限が移譲されるチームは大きな責任を持つことになる。そのためチームは、少数精鋭の必要な専門性をもつメンバで構成される事になるだろう。

関連して既に近年では法制定や法改正の増加さらにその文書量の増加する立法爆発と呼ばれる現象がおきている。社会にIT が普及した 2000 年前後に電子データや電子署名の法的有効性、公文書管理と情報公開、個人情報保護などの法整備が進められたことは立法爆発の主要な要因と考えられている。[榎並 2015]は、「ますます複雑化する社会、技術の進歩、家族関係の変容など、社会変化のスピードに法制度が追いつかず、官僚や法曹界など法律の専門家だけで内容のチェックをすることに限界がきていると考えられる」とのべている。

既に述べたように、今後は速い技術進展に伴い社会の変化が激化し、想定外の緊急事態に対して適切に社会ルールを設計したり変更したりする事態が増加しそうであるが、全ての重要度が高い案件に対して適切なチームを招集し全力で取り組み続けることは現実的には困難であろう。つまり処理すべき案件に対処しきれずオーバフローという課題が生ずる。これは処理量のスループットの問題である。

今後 AI がネットワーク化され、さらに自律的な判断能力を高めるにつれて、新たな社会的ルールを必要とする問題が増えるだろう。（以下のリストは例であり網羅的ではない）。

- 自動監視装置により、観察者は特定の他者のトレースすることで優位になりたいが、被観察者は監視を逃れて他者の影響力を避けたい。つまり両者間の生存志向同士の衝突することで、プライバシー保護の問題が顕在化する。
- 資本家は高度な AI を利用して作業を自動化したいが、これにより失業率が上昇するなら、人々の生活を守るための社会的ルール（ベーシック・インカムなど）が必要になる。
- 完全自動運転が導入されると効率的で安全な輸送が実現されるが、事故の責任帰属を扱うルールが必要となる。
- AI が環境から自律的にデータを収集し、さらに学習して能力を高めれば、より便利な道具となる。しかしそのデータに対してアクセスできる範囲をどのように設定するのか、さらに AI 自身の判断で生じた事故の責任を AI 自身に帰属させるか否かについて等の社会的ルールが必要になる。
- AI サイエンティストが出現すれば、技術開発を促進できるが、開発された新技術を制御したり、新種の事故の発生を制御する対策や損害賠償を分担したりするためのルールが必要になる。
- AI は殺傷兵器の能力を高めるために、非倫理的な利用を制限するための社会的ルールが必要になる。
- ある組織等が、トランスフォーマティブな AI 技術[1]を開発し独占した場合に、富・権力・知能がその組織等に集中して、社会に大きな格差が生まれうる。この状況を回避するための社会的ルールが必要になる。[GoodAI 2017]

多くの場合に、AI 自身は生存志向を持たないとしても、背後で AI に動作目的を与える人物には生存志向があるだろう。このために間接的な形で、生存志向の間で軋轢を生むことになる。

(1) 社会的決定を加速するための対策案

情報通信技術の発達により、意見集収や議論を行うことは容易になった。そこで広く民意を組み上げて政治に対する不満を減らせる可能性がある電子的な直接民主制(E-democracy)の可能性などが検討されてきている [山川 2006]。

また最近では AI 政治家を作るという議論もはじまっている。考えうる可能性として、今後は、多様な AI 政治家が出現するかもしれない。例えば中立的な AI 政治家は議論を促進できるだろうか。様々な AI 政治家は多様な発言を行うだろう。 バランスのとれた優れた選択肢、声の小さな多数派が望む選択肢、参考となる極端な選択肢などが議論の初期段階から表れ、議論全体として時間を短縮できる方法が見いだされるかもしれない。

1 トランスフォーマティブな AI 技術としては、自己再帰的な改良能力をもつ汎用人工知能などが想定されている。

法律についていえば、 その数が増大して複雑に絡み合ってくることで、ソフトウェアでいわれるスパゲティ・プログラムのように、理解が難しいものになると同時に、様々な問題を引き起こす要因ともなりうる。こうした状況へ対処するためには立法支援に AI を導入することも考えられる。

3.4 秒単位の社会的決定（レイテンシ問題）

秒単位より速い判断を要求される急用度が高い場合は、人間の判断速度を上回り、人間が対応することは不可能である。つまりここでは判断のレイテンシが問題となる。

想定しうる仮想的な事例としては、高度に知的に制御されているある地域のライフライン・インフラに対する未知のサイバー・テロへの対処などがある。有効な防御方法が見つからないため、政府の権限により、致命的な被害拡大を避けるために、数秒間の間に一部の市民には甚大な被害を生じる制御の切り替えなどなどの対策を行うといった、重大な社会的決定を行わざるを得ないような状況である。こうした社会的決定が政府で行われた場合に、上記インフラに関与する民間で所有される AI も新たなに決定されたルールを迅速に反映する事になる。

ただし予めこうした事態が想定されて、対応策について社会的決定がなされていれば、決定済みのルールに基づいて AI が対応すればよいので、この問題は生じない。

現在においても、AI が取引を行う株式市場のように人間の認知速度を超えた状況で問題が起きた場合に、これに対処する新たなルールを秒単位で設定するような事態を想定するとこれは人間にはできない。また将来、自動運転車や自律的に飛行するドローンなどが多数に行き交うような状況において、予期しない大規模なトラブルがおきたような場合にも、その進行が速いために人間の判断能力では対応しきれないかもしれない。AI が社会に浸透する中で、効率化が進む一方で、物事が人間の認知能力を超えることが増えていると思われる。

本来は、民主的に決定すべき重要案件について、時間不足で判断を行えないことは大きな問題である。

4. 社会的決定の AI への移譲

AI を開発することは、開発する組織等からみれば生存志向の点からみて有益であるし、しかも人々に対して多くの便益を与えるのも事実である。 そして AI 技術は世界中の至る所で比較的簡単に開発を進めることができる。このため AI 技術の進展に対してブレーキを踏むことは困難である。今後とも AI が発展しそれがネットワーク化する状況が続くとすれば、社会全体はブレーキのないままにアクセルを踏み続ける車のように技術上の制限による最高速度まで加速するだろう。

社会的ルールの制御を人や人の組織による社会的決定のみに頼ることは、レイテンシとスループットの両面においては速度不足を招く恐れがある。それに起因して、政情不安定、失政、無政府状態、紛争の勃発といった問題を生ずるかもしれない。そこで社会ルールを素早く制御できる社会的決定の仕組みが必須となる。

4.1 社会的決定の AI への移譲

以上述べてきたような、社会の変化速度に起因する危険の回避には、社会ルールの策定などを含む社会的決定において、適切に AI に権限を移譲する以外に選択肢がないように思われる。なぜなら AI による社会的決定は、以下でのべる価値関数の問題を解決すれば、人間による決定速度を遥かに上回れるので、少なくとも緊急時の安全装置として有効であろう。既に我々の社会において多くの社会的決定は代議員などに移譲しており、移譲自体は目新しいものではない。

こうして AI が緊急時に行う社会的決定の結果は、時に人々からみて不適切と捉えられるであろう。そうであっても、人が行う社会的決定が遅れることにより大惨事になるよりは暫定的な緊急対応が必要である。これは車に例えればアンチロックブレーキシステムのように、人の社会的決定では間に合わない制御を行うものだが、この場合はブレーキというよりも崖から落ちないように AI がステアリング操作するようなものである。

なおレイテンシの問題により AI が社会ルールを緊急決定した場合には、あとからで人々がルールの確認や見直しを行うべきであろう。一方でスループットの問題を解決するために AI を利用した場合は、そもそも人間側の判断するリソースの不足がもんだいであるので、人々が実際に確認を行うことは難しく、いつでも検証可能な状態に判断結果を公開するなどの対策をとることになるであろう。

4.2 合意形成結果を近似するための価値観数

AI による高速な社会的決定を行う際に、明らかにボトルネックとなる技術がある。それは、民意の総体として選択肢に対する評価を高速に近似する価値関数である。なぜならある案件について合意形成に到達するには、単に決定するだけでなく、関係者に納得してもらう必要がある。

当面考えうる二つの方法を以下で述べる。有力な可能性はこの二つを適切に組合せて利用することであろう。

(1) 民意の高速な推定

今後、日常的に個人を観測することで個人の価値観を推定することが可能になるだろう。たとえば、個人の表情や振る舞いから情動や感情を推定し、そこから価値観を推定する方法である。さらに踏み込んで個人に対して様々な体験を仮想的に提示することで暗黙化された価値観を表出させることも可能であろう。

こうして得られた個人に関わる観測データから逆強化学習や認知シミュレーションなどを利用して、内在する価値観をモデル化する。こうしたモデル化が実現されれば、未知の事象に対する価値推定も可能になる。多くの個人についての価値を推定できれば、その累計により、ある事案にたいする民意を推定できる。

推定した民意を足がかりに社会的な決定を行うことで、現状を維持する保守的な傾向を高速な価値観の近似関数に組み込める点で有用である。しかしながら、個人の価値観を推定することはプライバシーの問題を伴うため、その弊害を避けるための技術的工夫が必要になるであろう。また大きな変化によって生じる非日常的な状況に対する価値判断は、しばしばその場に遭遇しなければ想像できないため価値観の予測は困難であろう。

(2) より上位の社会的目標の設定

二者間の意見対立時に、より上位の目標において共通点を見出すことは有効である。一般的に、より上位における抽象度の高い目標ほど合意が得られやすく、その目標を実現する手段としの選択肢もしくは副目標の範囲が広がる。つまり我々の社会で共有しうるより上位の社会的目標が設定できれば、権限を移譲された AI は状況に応じたより柔軟な対応が可能になる。

一方で、わたし達が直感的に行っている価値判断は、日常的な状況に適応している可能性が高いかもしれない。そうだとした場合には非日常的な状況でも適切な決定を行うためには、社会におけるより上位の目標が設定されているほうが有益かもしれない。

しかしながら何を上位の目標として設定するかは非常に大な問題である。グローバル化が進んだ現代において種としての

「人類の生存」を副目標とすることは、比較的受け入れやすいと思われる。これは、[山川 2017]の中で述べた、「万人の幸福」と「人類の存続」の間のトレードオフの緩和を AI がおこなうというアイディアにも通じている。

5. 考察: 直感的に理解し難い上位の社会的目標

ここでは AI から離れて人間の価値観について考察する。

人間がもち自覚している直観的な価値観は自己保存、リソースを確保(食欲など)、繁殖(性欲)などの欲求や、さらには倫理観などを含む。そこからみるとの直観的感覚からすれば「人類の生存」は理解しがたいように思える。

人間のもつ基本的な欲求などは、おそらくは「人類の存続」という上位の目標から導きうる副目標ではないかと考える。しかし通常我々はその繋がりを認識していない。こうした副目標は進化の環境に適応した形で設定されているだろうが、他の目標との関係性は保持されていない。

一方で具体的な副目標は実世界での素早い反応を可能とする点で優れている。上位目標から手段としての副目標を導き出するには、上位目標を達成する多数の選択肢を作成し、さらに評価する必要がある。このため仕事などの日常シーンにおいても、多くの人々はわかりやすく自身が貢献しやすい具体的な目的に魅力を感じやすい。

しかし副目標が固定化されていると環境変化に対応する柔軟性を欠き、複数の副目標に拮抗や矛盾を生じた際に上位目標い上がって問題を解決することができない。

ここで仮に、上位目標としての「人類の存続」と人間の直観的な価値観が衝突するシナリオを考えてみる。 何らかの困難な環境変化が発生したときに、人類を遺伝子から再現する技術が完成していれば、一旦は全人間を殺傷してから、時間をおいて人類を復活させるのが現実的かもしれない。だが人間にとって、この判断は俄には受け入れがたい。ところが視点を広げ昆虫の世界に目を移せば珍しいことではない。例えば 1 化性の昆虫であるミドリシジミは年間の半年以上は卵で過ごしている。つまり人類全体が卵の時期を過ごすことを許容できるか否かである。

人間が持って生まれた固定的な価値観を乗り越えて、柔軟に社会的な決定を行えるか否かは大きなチャレンジである。もしかすると人間の価値観をより上位なものに昇華させるために AI などの技術を利用できるのかもしれない。先の例で言えば、人々が個体としての死を受け入れることによって人類の未来が開けることを納得することを促進する技術である。人間の選好を操作することは現在でもマーケティングという形で日常的に行われている。そう考えると今後さらに感情を操作するコンピューティングが進歩すれば技術的には遠いものではないかもしれない。

6. おわりに

我々の社会は大規模に組織化されており、それを安定的に持続するためには、状況に応じた社会ルールが必要である。技術進展により社会の変化が速まれば、立法等による社会的ルールの制御を速める必要があり、それを支える社会的決定も迅速化される必要がある。スループットが問題となるケースは既に立法爆発などとして顕在化している。レイテンシが問題となる場合として、たとえば、緊密に相互作用する AI のネットワークに支えられるようなインフラは、意図的もしくは偶発的に生じたトラブルに対して、数秒単位の短時間で対応する必要があり、対処の遅れは被害の拡大を招くかもしれない。

そこで社会的決定の一部を AI に移譲する可能性を議論した。そうなるとこうした AI を制御するために事案ごとに生ずる選択肢に対して高速に評価を行うための目的関数をいかにつくるかが重要な課題となる。とりうる選択肢としては、事案ごとの市民の総意を何らかの方法で高速に推定する方法、および、人類の存続などのより上位の目的を設定する、もしくは、それらを組合せたものになるであろう。

現在でも、「人では対処しきれないから AI に判断を任せたほうが、安全かつ効率的ですよ」というセールストークは、よく見られるものである。しかし今回検討したように社会的決定の権限を AI に委譲することは、明らかに大きなリスクを伴うことになるだろう。今後の技術進展により、重要性が高くとも速度的に人間が対処しきれないような案件が表れてくると想定するなら、決定の権限を AI に移譲したことで回避できるリスクを勘案した上で、必要最小限の範囲で AI への委譲を行う必要がある。このケースの中で特に AI の決定が人間の直観的な価値観から乖離する可能性ある場合は問題が大きく、何らかの対策が望まれる。

参考文献

[FLI 2017] Future of Life Institute, アシロマの原則, , 2017. https://futureoflife.org/ai-principles-japanese/

[IEEE 2017] IEEE, Ethically Aligned Design, v2, 2017. https://ethicsinaction.ieee.org/

[GoodAI 2017] GoodAI and AI Roadmap Institute, Report from the AI Race Avoidance Workshop, 2017. https://goo.gl/VJHjRU

[榎並 2015] 榎並 利博, 立法爆発と法律のオープン化 第 1 回 立法爆発の実態と専門家の限界, 法と経済のジャーナル, 2015/09/18. https://goo.gl/ei6lbq

[山川 2003] 山川宏, "集合的決定におけるマルチエージェント強化学習 - E-Democracy の制度設計利用を中心に -", 信学技報, NC2002-124, pp. 43-48, 2003.

[山川 2009]山川宏. 多段委任投票の公正化を促進する有力投票者推薦. 人工知能学会論文誌,Vol. 24, No. 1, pp. 170-177, 2009.

[山川 2017] 山川宏, ドワンゴ人工知能研究所の所長の山川宏とのインタビュー、Future of Life Institute, 2017. https://goo.gl/rtKsU5