

汎用人工知能の標準問題としての人狼ゲーム

“Are you a Werewolf?” becomes a Standard Problem for General Artificial Intelligence.

篠田 孝祐*¹
Kosuke Shinoda鳥海 不二夫*²
Fujio Toriumi片上 大輔*³
Daisuke Katagami大澤 博隆*⁴
Takahiro Osawa稲葉 通将*⁵
Mitimasa Inaba*¹ 電気通信大学
UEC*² 東京大学
The University of Tokyo*³ 東京工芸大学
T.P.U*⁴ 筑波大学
University of Tsukuba*⁵ 広島市立大学
Hiroshima City University

We propose a standard problem “Are you a Werewolf?” for general artificial intelligence. This game is one of communication game around the table. We have thought this game is useful metrics for evaluating progress artificial intelligence.

1. はじめに

「ある村に、人間の姿に化けられる人喰い人狼が現れた。人狼は人間と同じ姿をしており、昼間には区別がつかず、夜になると村人たちを一人ずつ襲っていく。村人たちは疑心暗鬼になりながら、話し合いによって人狼と思われる人物を1人ずつ処刑していくことにした」・・・以上が、コミュニケーションゲームとして知られる人狼ゲームのカバーストーリーである。

この人狼と呼ばれるゲームがある。このゲームは、基本プレイヤ同士の話し合いにより進行し勝敗がきまる。この人狼ゲームは、言語のみで進行するゲームでありながら、全世界で楽しんでいるゲームであり、特に日本ではオンライン上で行う BBS 人狼が日々行われている。

本論文では、この人狼ゲームを、汎用人工知能の開発状況を評価するための標準問題として用いることを提案し、その可能性を議論する。

2. 人工知能における標準問題

標準問題とは、ある技術の性能などを評価するうえで、その良し悪しを比較可能とするために設ける問題のことである。人工知能が対象とする、現実存在する諸問題は非常に幅広く多様である。そのため、同じ領域を対象として課題を設けたとしても、必ずしも同様の環境を構築できるとは限らず再現性に乏しい。そこで、互いがそれぞれの問題を直接調べる代わりに、現実におけるさまざまな課題の中で、重要かつ一般的な問題を十分に表現でき、現実の問題へのアプローチとなる問題を標準問題としてきた。

これまで、人工知能の標準問題としてさまざまな問題が提案されてきている。代表的なものとしては、チェスや将棋・囲碁、追跡問題、囚人のジレンマ、RoboCup などがある。チェス・将棋・囲碁などはゲーム進行のスコアリングと探索の問題であり、追跡問題は共通の全体ゴールをもつ集団が全体の戦略と個々の戦術との関係を追及する問題である。そして、RoboCup では、ソフトウェアとしての人工知能技術とロボットというハードウェアを融合しサッカー・災害救助・日常環境という実社会への投入を前提として問題を提案している。これらに共通するのは、それぞれの課題に対して、共有すべき大きな目標を設定し、共通のフレームワーク上で互いの技術・知見を実装することで、利点欠点を比較可能となる。

3. 汎用人工知能における標準問題

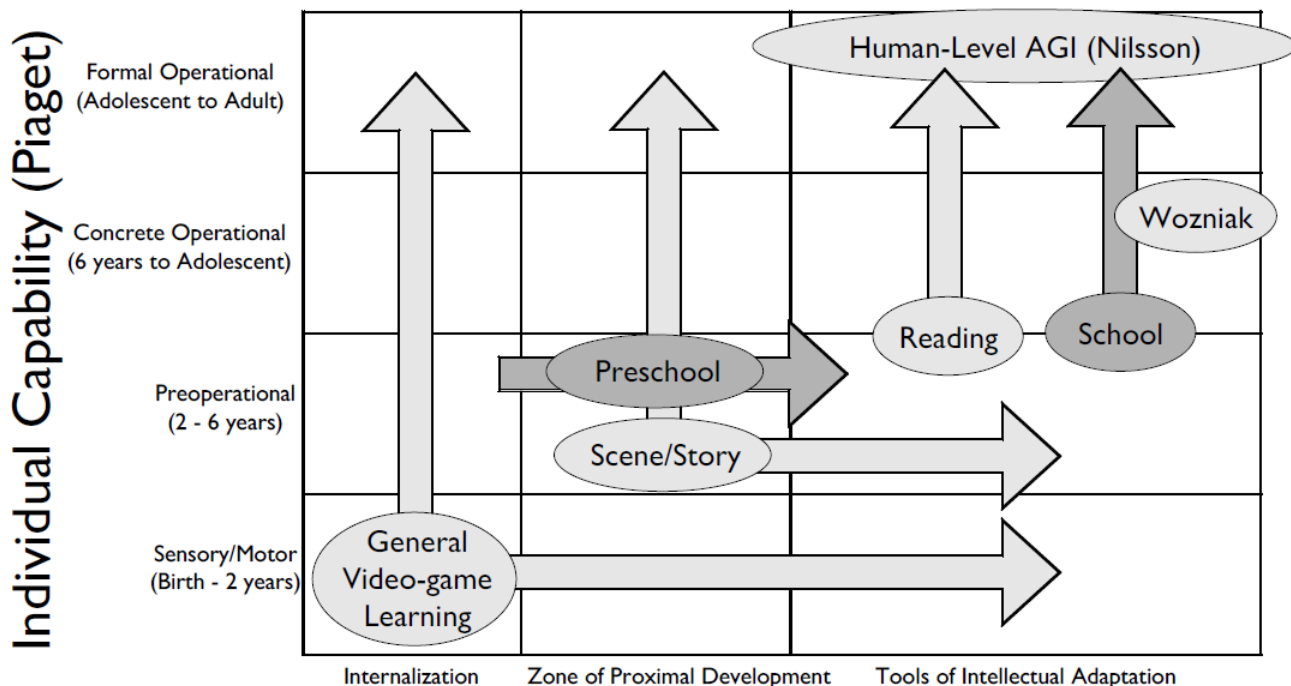
汎用人工知能とは、人間レベルの知能を実現することを目標とした人工知能研究のことである。この汎用人工知能の具体的な姿は人工知能や人間の汎用知能と同じく明確な定義は存在しないが、Adams らにより、汎用人工知能の実現にむけた課題や評価シナリオが整理されている[Adams2011]。汎用人工知能に関する詳細な説明は、本稿では割愛するが、Adams らは、汎用人工知能の評価のためのシナリオとして以下の 6 つのシナリオを提案している

1. **General Video-game Learning**
特定のビデオゲームを対象にするのではなく、さまざまビデオゲームをプレイしながら学びクリアできるようにするという課題。主に視覚処理と入力操作の課題を内包したシナリオ
2. **Preschool Learning**
基本的な運動能力や絵を理解する能力を育てるシナリオ
3. **Reading Comprehension**
読解力をみにつけるためのシナリオ
4. **Story/Scene Comprehension**
文章を理解したうえで、場面における登場人物の関係性や物語の流れを読み解く力を学ぶシナリオ
5. **School Learning**
Preschool と関連あるシナリオではあるが、論理的な思考や社会的な関係性などをまなぶことを想定したシナリオ
6. **The Wozniak Test**
見知らぬ家を訪問しその家の主にコーヒーを給仕することを目標としたシナリオ。ある意味 RoboCup Home の課題に近い目標であるが、社会関係や他者の理解まで要求しているシナリオ

Adams らは、これらのシナリオをベースとして、必要とする環境とそこで行うべきタスクを設定することを求めている。また、彼らは、AGIの実現に必要なとされる能力の領域も提案している。具体的には、知覚・記憶。注意・社会インタラクション・プランニング・モチベーション・推論・運動作業・コミュニケーション・学習・情動・自己/他者モデル・構築/創造・定量化など 15 の能力エリアを設けている。

4. 標準問題としての人狼ゲーム

人狼ゲームは、アメリカのゲームメーカーLoony Labsが 2001年に発売されたパーティーゲーム「汝は人狼なりや」(1)及びそ



Social-Cultural Engagement (Vygotsky)

図1 人間の汎用知能の発達段階と想定されるシナリオの関係図([Adams2011]より)

の派生ゲームの総称である。多数の類似ゲームが世界中で市販され、世界中でプレイされている。日本においてもタブラの狼(2)やうそつき人狼(3)など多数のゲームが販売されている。この人狼ゲームは、コミュニケーションゲームの一種であり、人、もしくはエージェント間でのコミュニケーションのみで決定する。

人狼におけるゲームの特性には以下のようなものがある。

- コミュニケーションによるゲーム進行
これまでのゲームの多くは、環境の状況(他のエージェントの振舞も含む)をもとに行動を決定するという問題が中心であった。それに対して、人狼ゲームでは、そのアクションにコミュニケーションによる情報交換をもとにゲームを進行し勝敗を決する。もちろん、鳥海ら[鳥海 2014]が、ゲームの状況をもとにプレイヤーの行動を学習することでゲームが進行できることを示したように、会話による要素がなくてもある程度ゲームを実行することができる。だが、エージェント同士の対戦だけでなく、人間も交えてゲームを行うことを検討した場合には、コミュニケーションはゲームを楽しむ要素として非常に重要である。
- 情報の不完全性と非決定性
将棋や囲碁のような完全ターン性のゲームでは、互いの状況を完全に把握できる環境であるのに対して、人狼ではお互いの情報がある程度秘匿することでゲームが成立する。また、決定すべき行動に投票などあるため、個々の行動のみで場の状況が決定することはなく非決定的でもある。
- 信頼の構築・説得
情報の不確実性をベースにしているが、そこから情報を確定させていくためには、自己の持つ情報を開示するなどして説得する作業が求められる。ただし、プレイヤーには複数の陣営があるため、他の陣営から

味方の陣営の情報のエントロピーを意識して活動する必要がある。そのためにも、信頼を獲得し説得する作業が重要である。

- 他者の意図推定
説得するためには、他者の意図の推定を行う必要がある。
- 他者の中の自己意図推定
上記の他者の意図の推定と同時に、他者にある自己意図を推定する必要がある。この意図推定は自分の発話の信頼性を向上させるのに重要な要素となる。

我々は、この人狼ゲームを、ソフトウェアエージェントを用いたゲームとして、そして将来的には、人間を交えたゲームとしておこなえるようプロジェクトを立ち上げる準備をしている。

人狼ゲームをソフトウェアエージェントにより対戦するゲームとしたうえで、人工知能という観点から見たときに、それを実現するには、エージェントやインタラクション・認知科学・心理学、ロボット工学など多様な領域にまたがる。Adams らの挙げた能力エリアでいうと、ソフトウェアエージェントとして実装するだけでも、記憶・社会的インタラクション・プランニング・推論、コミュニケーション・学習・自己/他者モデル・定量化など多岐にわたる能力が必要となる。

これに、人間を交えた実世界でのゲームを行うとした場合には、知覚器官や運動能力なども加わる。視覚情報、自然言語処理、ロボットの実装などの基本的なところはもちろん、高度な会話能力や表情、社会性など、人間として高度な能力エリアの大半をカバーする課題と言える。おそらく、標準問題として唯一欠けているのは、人狼ゲームとして閉じた世界でのゲーム対戦であるために、環境のオープン性が求められない。しかしながら、「対戦相手に勝つ」ではなく、「対戦相手とゲームを楽しむ」という観点からゲームプレイを汎用人工知能に求めるのであれば、ゲームに閉じた行動では楽しむことは難しいと考える。

知覚器官 視覚 嗅覚 触覚 味覚 聴覚 クロスモデル	記憶 作業 エピソード 印象 意味的 聴覚 手続的	注意 視覚的 聴覚的 社会的 振舞的 コミュニケーション ジェスチャー	社会インタラクション コミュニケーション 妥当な 社会的影響 協力的 競争的 関係性 学習 物まね 強化学習 対話的 メディア思考 競争的 思考的	プランニング 戦術的 戦略的 物理的 社会的 情動 受動的 抑制的 コントロール 理解 共感 自己移入 定量化 目に見える要素の数え上げ 単純な道具の計測 小さい数の数値計算	モチベーション サブゴール生成 感情に基づく 持続的な 利他的な 自己・他者モデル 自己認識 他者認識 関係性 自己抑制 心の理論 共感 自己移入
推論 帰納的 演繹的 仮説的 物理的 因果 連想的	運動・作業 物理的能力 道具の活用 ナビゲーション 建物 / 創造 オブジェクトによる物理的構築 新規コンセプトの形成 言語考察 社会的組織	言語 音楽 絵による表現 競争的 思考的			

図 2 AGI 能力エリア図([Adams2012]より、筆者により翻訳)

まあ、汎用人工知能という観点からすこし離れるが、このようなコミュニケーションゲームを使用することで、大学生の人間教育に有用であるという示唆を得ている。したがって、汎用人工知能を社会に役立てるまへの教育システムとして、人狼ゲームを知育ゲームとして活用するのに有用な可能性も高いと考える。

我々は、この人狼ゲームを General Video-Game か School Learning のシナリオの一部として用いられるよう、人狼ゲームに含まれる人工知能的な課題を整理する予定である[稲葉 2012, 稲葉 2011, 大澤 2013, 鳥海 2014]。

5. まとめ

本論文では、汎用人工知能の標準問題として人狼ゲームの検討を始めた。今後は、このゲームをベースとした競技会などを設けることにより、普及を図る予定である。

なお、我々の人狼プロジェクトにご興味を持たれた方は、プロジェクト HP (<http://www.aiwolf.org/>) をご覧いただければ幸いです。

参考文献

- [稲葉 2012] 稲葉通将, 鳥海不二夫, 高橋健一: 人狼ゲームデータの統計的分析, ゲームプログラミングワークショップ論文集, no.6 pp.144-147, 2012.
- [稲葉 2013] 稲葉通将, 大島菜央実, 鳥海不二夫, 高橋健一, “雑談ばかりしていると殺される-人狼 BBS におけるプレイヤーの発言傾向と意思決定・勝敗の分析-,” JAWS2013
- [大澤 2013] 大澤博隆, “コミュニケーションゲーム「人狼」におけるエージェント同士の会話プロトコルのモデル化, HAI シンポジウム 2013, pp.122-130
- [鳥海 2014] 鳥海不二夫, 梶原健吾, 稲葉通将, 大澤博隆, 片上大輔, 篠田孝祐, 西野順二, “人工知能は人狼の夢をみるか-人狼知能プロジェクト” デジタルゲーム学会 2014
- [Adams2011] Sam S. Adams et al.: Mapping the Landscape of Human-Level Artificial General Intelligence, AAAI, 2011.