

知的財産と汎用人工知能

Intellectual Property and Artificial General Intelligence

岡本義則^{1*}

Yoshinori Okamoto¹

¹ ユアサハラ法律特許事務所

¹ YUASA AND HARA

Abstract: This paper discusses a bottleneck of development of Artificial General Intelligence (AGI) and proposes a hypothesis that data acquisition issues can be the bottleneck of the AGI. To address the data acquisition issues, this paper proposes General Supervisor Database (GSD). Moreover, this paper introduces new intellectual property to realize the GSD and proposes General Supervisor Database by Intellectual Property (GSDIP).

1 はじめに

多層のニューラルネットワークは、たとえば、ネオコグニトロンなど、1980年代のいわゆる第2次AIブームの以前から存在していた[1]。

しかし、第2次AIブームの後に、AI研究は冬の時代を迎えた。近年の深層学習の発展等による第3次AIブームは、情報処理のモデルの進歩と共に、計算機の処理能力の飛躍的な向上や、限られた範囲ではあるが学習用の大量のデータが得られたことに支えられていると考えられる。

このように、(1) 情報処理のモデル、(2) 計算機の処理能力、(3) 学習のためのデータ、はいずれも人工知能の進展に必要と考えられる。

近年の人工知能(AI)の技術の発展は目覚ましいが、現在の人工知能は、基本的に特定の狭いタスクに向けられており、汎用人工知能からは遠く隔たっている。

汎用人工知能の研究において、(1) 情報処理のモデルについては、全脳アーキテクチャーなど、脳を参考にした情報処理のモデルが精力的に研究されている[2]。

また、(2) 計算機の処理能力については、消費エネルギーあたりの処理能力では脳に遠く及ばないものの、処理能力自体は毎年向上している。

しかし、(3) 学習のためのデータの問題については、目途が立っていない状況と考えられる。

データの中には、無料で公開されるものや、人工知能を用いた計測の自動化やIoT等の技術が進歩すれば自動的に獲得できる性質のものも存在する。

しかし、特定のタスクを解決する際に鍵となるようなデータは、経済的な価値を有する場合があります、そもそも公開されることが期待できない場合があります。また、

良質なデータの作成には、多くの人手と長い年月が必要となる場合があります、良質なデータ自体が存在しないことが多い。

汎用人工知能の研究が今後ある程度進展しても、特定のタスクで高い性能を出すためには、特定のタスクを解決する際に鍵となるようなデータが必要と考えられる。これは人間でも同様であり、秘密にされているノウハウを取得できなければ、情報処理能力が高くて、当該特定タスクを解決できない場合は十分に考えられる。特定の領域のタスクの解決には、特定の領域のデータがある方が有利であるのは、ノーフリーランチ定理にもみられるように、情報処理の性質から、やむを得ないことと考えられる。

そうすると、特定のタスクについてのデータは、人間であれ、人工知能であれ、高い性能を出すために重要と考えられる。

人間の場合には、特定のタスクについてノウハウを有する組織等に所属して、長い間の経験を実際に積むことで、実質的にデータを取得することが可能な場合がある。しかし、人工知能の場合、身体性の問題もあり、人間と同様なデータの獲得には困難があると考えられる。

2 汎用人工知能の律速段階仮説

汎用人工知能を実現する際のハードルとして、(1) 情報処理のモデル、(2) 計算機の処理能力、(3) 学習のためのデータの問題のうち、どのファクターが重要なのか、あるいは、別のファクターが重要なのかは現時点では明らかではない。

しかし、ある一つのファクターの研究の進展のスピードが遅ければ、他のファクターの研究が進展しても、ス

*連絡先: E-mail: ai@japan8.info

ピードの遅いファクターの進展のスピードで、汎用人工知能の進展のスピードが抑えられてしまうことが考えられる。この仮説を、汎用人工知能の律速段階仮説 (Bottleneck Hypothesis of AGI) と呼ぶことにする。

律速段階としては、今後の取り組みにもより、(1) から (3) 以外のものを含めて様々な可能性が考えうるが、ここでは、(3) 学習のためのデータの問題 (データを取得するための身体性の問題等を含む) が律速段階になり、汎用人工知能の進展のスピードを抑えてしまうという仮説を立てる。これを、「汎用人工知能のデータ律速段階仮説」(Data Bottleneck Hypothesis of AGI) と呼ぶことにする。

この仮説に対しては、(1) と (2) が進展すれば、言語をある程度理解できるようになり、(3) データの問題が解決するという考え方もありうる。しかし、言語理解を可能にするためには、(3) データの問題についての取り組みが必要と思われる。また、仮に言語化された知識の獲得が可能となっても、言語化された知識を用いて特定のタスクを解決する場合、特定の問題領域について、人間のエキスパートのレベルにすら達するとは限らない。

人間においても、特定の分野のエキスパートとなるには、本を読むなど言語化された知識を獲得するだけでなく、特定の分野についてノウハウを有する組織等に所属して長い間の経験を実際に積むことで、実質的にデータを取得することが重要と考えられる。

言語化された知識のように、人間の意識に上る情報は、無意識下の情報が欠落しており、すべての情報を含んでいない。大量の教師付データ (なお、本稿ではラベル付データを含むより広い概念として、教師あり学習に用いることのできるデータを教師付データと呼ぶ) には十分な情報が含まれている場合があり、現在の AI の技術でも、大量のデータを用いた機械学習により、特定タスクにおいて人間を超える性能が実現される事例が見られる。

人工知能におけるモデル獲得の壁は非常に高いが、大量のデータを用いて学習をすること、特に、大量の教師付データを用いることにより、モデル獲得の壁が、ほんの一部にせよ破れたことは、深層学習のもたらした成果であると考えられる。

汎用人工知能の場合、単に特定の領域の問題を解くよりも、多くの能力が必要となる [4]。しかし、汎用人工知能においてもデータは重要であり、汎用人工知能は、特定領域におけるデータを不要とするものではないと考えられる [5]。

むしろ、汎用人工知能の場合、ある特定領域におけるデータだけではなく、理想的には、すべての領域におけるデータが必要となるため、データの必要性は、専用の人工知能よりも一層高くなると考えられる。

3 汎用教師データベース

現在は、特定の領域に絞っても、大量の教師付データを集めるのは困難と考えられており、どうすれば大量の教師付データを用いずに問題を解決できるかという方向に、人工知能の研究の方向が動いているように思われる。

しかし、特定の領域のタスクについて問題を解決する場合、大量の教師付データを用いずに方法で、十分な性能が出るのかは明らかではない。人工知能研究の難しさとして、困難な道を迂回しても、その先に道が拓けているのかは不明と考えられる。

それにもかかわらず、研究の方向が、どうすれば大量の教師付データを用いずに問題を解決できるかという方向に動いているのは、大量の教師付データを入手することが、非常に困難であると認識されているからと思われる。

しかし、大量の教師付データを入手することの困難を前提として、それを迂回しようとする研究の主流に対して、正面から大量の教師付データを入手する手法を検討するのも、一つの研究の方向性と考えられる。

本稿では、大量の教師付データの収集の問題を解決するためのコンセプトとして、汎用教師データベース (General Supervisor Database (GSD)) を提案する。

汎用教師データベース (GSD) は、広範な領域、理想的には、あらゆる領域における教師付データを集めたデータベースである。特定タスクを解決するために必要な教師付データも含まれている。また、一般的な知識に関係するデータを含めて、あらゆる種類の教師付データが含まれている。

汎用教師データベース (GSD) に集められた AI 学習用データには、学習の有効性が確認されている教師付データが含まれており、学習に有効な良質なデータが集められている。

このように、汎用教師データベース (GSD) は、あらゆる領域における良質な教師付データを集めたものであり、研究者を含む一般の人に利用が開放され、大規模な商用利用を除いて無料で利用できるデータベースとして構成することができる。

問題は、汎用教師データベース (GSD) をどのようにして実現するかである。国がデータ入力者を大量に雇用するなど、実現の方法は複数考えられるが、一般には特定の領域における多数の教師付データを集めるのさえも困難と考えられており、あらゆる分野における良質な教師付データを大量に集めることは、容易ではない。特に、特定タスクを解決するために必要な良質な教師付データは、経済的価値を有し、限られた人だけが作成可能で、しかも作成に莫大な労力がかかる場合があり、無償で作成され、公開されることは期待できない。

そこで、AI学習用データに対する新しい知的財産権を創出し、その出願により形成されるデータにより、汎用教師データベース（GSD）を作成することを提案する。このように、知的財産についての出願により形成される汎用教師データベース（GSD）を、知的財産による汎用教師データベース（General Supervisor Database by Intellectual Property（GSDIP））と名付ける。

AI学習用データに関して、新しい知的財産権を創出することで出願を促し、また、出願に対して審査制を導入することで、出願されるデータを審査し、良質な教師付データを収集することが考えられる。このような知的財産権の例として、筆者は、AI学習用データ権（仮称：データ特許）の制度を提案している [3]。

データの保護については、現行の知的財産制度での保護が十分でないことから、現在、不正競争防止法の改正による保護が検討されている。しかし、不正競争防止法による保護をAI学習用データの保護の中核に据えた場合、AI学習用データについては、複数の企業間など、データが公開される範囲が限定的なものとなり、広く一般の人々（AI研究者を含む）にはデータを公開しない方向にインセンティブが与えられることになりうる。

不正競争防止法によるデータの保護をAI学習用データの保護の中核に据えた場合、AI学習用データの作成と公開の十分なインセンティブとならず、今後の人工知能の発展の速度が遅くなるおそれがある。

しかし、AI学習用データは、人工知能の発展において重要なファクターとなりうるため、できる限り多くの良質な学習用データを集め、それを広範な人々が利用できるようにすることが重要と考えられる。

そうすると、AI学習用データについては、その作成を奨励し、作成されたデータを広く一般に公開することを促進するために、不正競争防止法による保護を中心とするのではなく、新しい知的財産権の付与が必要となる。

AI学習用データについて、新しい知的財産権による保護を与えることで、多くの良質な学習用データが作成される。それを広範な人々が利用できるようにするために、AI学習用データ権については、学術利用、研究のための利用、非営利の利用等は無料とし、大規模な営利利用についてのみ有料とすることが考えられる。

そして、AI学習用データ権の出願にあたっては、公衆からできる限り多くのデータを集めるように制度設計をすることが重要となる。たとえば、出願料は無料ないし低額に設定し、有用なデータの出願には、データの利用実績に応じて、出願奨励金が得られるようにすることが考えられる。

また、AI学習用データについては、出願書類にデータの使い方の詳細を記載し、審査官が実際に学習がで

きる有用なデータであるかを審査することにより、データの質を確保することが考えられる。

そして、汎用教師データベース（GSD）は、法整備をして、一般の人がアクセスしてあらゆる分野のデータを検索して利用できるようにし、多数の人が多数の異なる日時に作成したデータを、容易に集めて利用できるようにすることが重要となる。

たとえば、ある分野において、10年間において、年間10万人の人が、平均して1万個のデータセットを出願している場合、多数の人が多数の異なる日時に作成したデータを併用することにより、100億個のデータセットを利用できることになる。

画像認識の研究においては、ある程度の量のラベル付きの画像データが作成されている [6]。しかし、様々な分野の多数の人々がデータの作成に長期間関与しなければ、あらゆる分野において莫大な量の教師付データを作成することは困難と考えられる。

知的財産による汎用教師データベース（GSDIP）においては、出願は毎年行われるため、時間が経過するにつれて、データが集積されていく。

また、新しい知的財産権を用いた知的財産による汎用教師データベース（GSDIP）においては、知的財産権のライセンス収入や出願報奨金による収入により、継続的にデータの作成をする人や企業の参入を期待することができる。

このように、知的財産による汎用教師データベース（GSDIP）により、人工知能の研究者等がデータを整備するのに比べて、はるかに多くの多様な教師付データを整備することが可能となる。

このようなAI学習用データ権を新設し、あらゆる分野においてデータの出願を促し、教師付データを蓄積することにより、汎用教師データベース（GSD）を実現することが考えられる。

4 考察

現在、データの保護については、不正競争防止法の改正による保護が検討されている。これは、データの保護についての革新的な取り組みであり、一般的なデータの保護法制としては重要な一歩と考えられる。

しかし、人工知能の性能を高度化するという視点からは、AI学習用データについては、一般のデータの保護の問題とは異なる考慮が必要である。

残念ながら、AI学習用データについては、一般のデータの保護の問題とは異なる取り扱いが必要であるという認識が、知的財産コミュニティの中で浸透しているとはいえない。

汎用人工知能研究会を初めとする人工知能のコミュニティが、できる限り多くの良質な学習用データを集

めることが人工知能の発展に重要であるという意見を述べることにより、A I学習用データについては、一般的なデータの保護の問題とは異なる取り扱いを別途検討すべきという意見が、知的財産コミュニティに浸透することが期待される。

そして、A I学習用データに関する新しい知的財産制度が作られ、A I学習用データの出願と公開が長期的になされることにより、汎用教師データベース（GSD）を実現し、充実させることが重要と考えられる。

汎用教師データベース（GSD）が実現することにより、研究者は、あらゆる分野における大規模な教師付データを無料で利用できるようになる。

たとえば、ある領域における人工知能の研究の際に、当該領域における大量の良質な教師付データを得ることができるようになる。また、一般的な知識に関係する大量の教師付データなど、汎用人工知能の研究に必要なデータも得ることができる。さらに、あらゆる分野の教師付データは、学習だけではなく、汎用人工知能の性能評価のために用いることもできる。このように、汎用人工知能の各種の研究において、汎用教師データベース（GSD）を用いることができる。

また、汎用教師データベース（GSD）は、特定の領域における専用の人工知能においても、その性能を飛躍的に高めることができる。汎用人工知能の研究の進展により、隣接領域のデータからある程度の類推が効くようになれば、データの有用性はさらに高まることが期待される。

良質な教師付データの作成には莫大な人手が必要であり、その収集には長い時間がかかるため、できる限り早い取り組みが必要となる。時間をかけて、良質な教師付データを収集した汎用教師データベース（GSD）を充実させていくことにより、汎用人工知能を含む人工知能の研究に役立つと考えられる。

5 おわりに

本稿では、汎用人工知能の律速段階仮説を提案し、データの問題が律速段階になりうるという仮説を立てた。そして、データの問題を解決する手段として、あらゆる分野の教師付データを集めたデータベースである汎用教師データベース（GSD）を提案した。

汎用教師データベース（GSD）を具体的に実現する方法として、A I学習用データについての新しい知的財産権を用いることを提案し、知的財産による汎用教師データベース（GSDIP）を提案した。

データを作成する場合、人工知能の研究者の人数には限りがあり、また、研究者はデータ作成の作業だけに従事できるわけではない。様々な分野の特殊なデータを作成可能な人々を含めて、莫大な数の人々がデー

タの作成に長期間関与しなければ、あらゆる分野における莫大な量の教師付データを作成することは困難と考えられる。

A I学習用データについての新しい知的財産権を用いて、汎用教師データベース（GSD）を実現することは、人工知能の研究の促進に大きく寄与すると考えられる。

現在、データ保護法制が検討されているが、A I学習用データについて、一般のデータの保護の問題とは異なる考慮が必要であるという認識は浸透していない。

汎用人工知能研究会を初めとする人工知能のコミュニティが、A I学習用データについては、できる限り多くの良質な学習用データを集めることが人工知能の発展に重要であるという意見を述べることにより、A I学習用データについては新しい知的財産制度を別途検討すべきという意見が、知的財産コミュニティに浸透することが期待される。

このように、人工知能のコミュニティと知的財産コミュニティが協力して問題を検討することで、人工知能の潜在的な可能性を十分に発揮させる知的財産制度の実現が可能となると考えられる。

本稿が、人工知能と知的財産の分野の境界領域において、今後の議論の一助となれば幸いである。

参考文献

- [1] 福島邦彦: 位置ずれに影響されないパターン認識機構の神経回路モデル-ネオコグニトロン, 電子情報通信学会論文誌, Vol.J62-A, No.10, pp.658-665 (1979)
- [2] 山川宏, 市瀬龍太郎, 井上智洋: 汎用人工知能が技術的特異点を巻き起こす, 電子情報通信学会誌, Vol.98, No.3, pp.238-243 (2014)
- [3] 岡本義則: 人工知能(AI)の学習用データに関する知的財産の保護, パテント, Vol.70, No.10, pp.91-96 (2017)
- [4] 市瀬龍太郎: 認知アーキテクチャ評価手法の一考察, 第1回汎用人工知能研究会, No. SIG-AGI-001-12. JSAI (2015)
- [5] 山川宏: 一般事前知識についての考察 -究極のAGIを目指して-, 第1回汎用人工知能研究会, No. SIG-AGI-001-11. JSAI (2015)
- [6] 中山英樹: 画像解析関連コンペティションの潮流, 電子情報通信学会誌, Vol.100, No.5, pp. 373-380 (2017)