プロジェクト報告書(最終) Project Final Report　　提出日(Date) 2025/01/21

Make Brain Project

プロジェクトリーダー：狩野政宗 / Masamune Karino

1. 背景

近年、人工知能は目覚ましい発展を遂げ、生活や社会に深く浸透しつつある。そして、様々な分野で新たな可能性を開き、急速に発展している。しかし、人工知能と生物の脳にはさまざまな違いが存在し、現在の人工知能では実現できていないことが多くある。例えば、ロボット制御では、生物の感覚的な動作や柔軟性を再現することは難しい。

本プロジェクトは、脳の情報処理を模倣した仕組みを取り入れた人工知能を開発し、現実問題に適用すること、また生物の脳の仕組みを理解することを目標とする。そして、生物は、触覚、聴覚、視覚などといったさまざまな感覚入力をもとに脳神経系で情報処理し、行動を出力している。この点に着目し、応用することをテーマとして、２つのグループが活動を行った。グループごとのテーマはそれぞれ、「触覚を持つ歩行ロボット」、「視覚を持つAIカー」である。加えて、脳で行われる思考に着目し、応用することをテーマとして、「声から予想されるキャラクターを出力するAI」というグループが活動を行った。

触覚を持つ歩行ロボット

生物は、周囲の環境に適応した歩行を行っている。中でも昆虫は人間や他の脊椎動物などと比較して、高度な脳の機能を持たないにもかかわらず、環境に適応した歩行を行っている。このことから、生物は歩行の際に周囲の感覚器官からのフィードバックを脳だけでなく、脳の下位の部分も利用して分散的に処理していると考えられる。

Owakiら(2017)は触覚を利用して分散的に処理を行い、歩行リズムを調整して平地での適応的な歩行を実現している[A-1]。以前のプロジェクト学習である脳を作るプロジェクトにおいて、石川ら(2021)はナナフシの歩行を神経回路を再現したプログラムによってコンピュータ上での3D物理シミュレーションとして実現していた[A-2]。

先行研究においては、シミュレーションや平面の歩行など、理想的な環境下での実験であった。これらの手法は、実際の運用環境を模擬するには限界がある。特に、不整地のような複雑な地形においては多くの課題が残されている。

声から予想されるキャラクターを出力するAI

ヒトはある音声を聴いたとき、その音声を発した人間の身体的特徴をある程度予想することができる。なぜなら、骨格・体格・人種などの視覚情報は、ピッチ・抑揚・各言語に固有のリズム等の発声者の身体的属性と関連があると考えられるためである。このような現象をモデル化した研究として、Speech2Face[B-4]がある。Speech2Faceは、音声からその話者の顔を予測するモデルであり、骨格や人種、性別といった身体的特徴を音声信号に基づいて推定することを目的としている。しかしながら、この推論の過程に類似した状況として、ヒトがアニメーションキャラクターの音声を聴いたときにそのアニメーションキャラクターの顔を予想できることに着目した顔画像の推論手法は依然として確立されていない。当グループが着目するアニメーションキャラクターは、その声を聴くだけでヒトが顔を予想できる点が特徴であり、これはヒトが既存の経験や文脈に基づいて推論を行っていることを示唆している。この現象を解明することは、アニメーションやゲームといった創作分野において、キャラクターの音声から外見を生成・調整するAI技術の開発につながる。たとえば、声優の音声からキャラクターの外見を自動生成する技術は、制作コストの削減やユーザーの没入感向上に寄与する可能性がある。

視覚を持つAIカー

近年、人工知能や自動運転技術の研究が進展し、交通安全向上への期待が高まっている。交通事故は深刻な社会問題である。日本における過去10年間のデータを通じて、交通事故件数は減少傾向の後再び増加傾向が見られ、死亡者数は年によって変動している。令和3年（2021年）から令和5年（2023年）の間で、交通事故件数は39,000件から37,000件へ減少し、令和5年には死亡者数が前年よりも68人増加して3,573人となっている。

本研究では、Raspberry Pi5を搭載したラジコンカーを用いて、画像認識と機械学習を活用した「自律走行システム」の開発に取り組んだ。自律走行は、交通事故の主な原因となるヒューマンエラーの低減や安全性の向上に寄与する技術として期待されている。本研究の目的は、人間の脳が担う運転操作を人工知能で再現し、交通規則を遵守しながら目的地までの最適な経路を自ら選択・走行できる自律運転の実現を目指すものである。

本研究は、公立はこだて未来大学の「Make Brain Project」を参考に進められた。先行研究では、Raspberry Pi4を搭載したラジコンカーで白線認識、障害物検知、道路標識認識を実現したが、精度向上や多様な環境へ

の適応が課題とされた。本研究では、これらの課題の克服を目指し、より高度な自律走行技術の開発に挑戦した。

2. 課題の設定と到達目標

触覚を持つ歩行ロボット

昆虫は歩行をするとき感覚器官からのフィードバックを脳だけでなく、脳よりも下位の神経システムで分散的に処理していると考えられる。そのひとつとして、歩行するときの足先の触覚の働きを調べた。そして、この触覚が不整地で適応的な歩行に役立てられていると考えた。

不整地を歩行する際の問題として、地面の凹凸によって体が大きく揺れてしまうことがあげられる。そして、足先のフィードバックを用いてこの揺れを減少させ、安定した歩行をさせることを目標とした。

このことを調べるために、足先のフィードバックを用いた２つの数理モデルを作成した。そして、その数理モデルを実装した、それぞれの足を自律分散的に制御できるロボットを作成した。

これらの２つの数理モデルを実装した歩行と、フィードバックを用いない制御則における揺れの大きさを比較した。

声から予想されるキャラクターを出力するAI

アニメーションキャラクターの音声を聞くだけで、ヒトはそのキャラクターの顔を予想することができる。そこで本グループではこの点に着目し、この予想の際に行われる情報処理を再現できる人工知能を開発することを試みた。これにより、ヒトが音声からアニメーションキャラクターの顔をどのように予想しているのかを理解することを目指した。具体的には、アニメーションキャラクターの音声と顔画像の特徴を学習し、入力された音声の特徴に基づいて予想されるアニメーションキャラクターの顔画像を出力する人工知能を開発することを目標とした。

視覚を持つAIカー

本研究では、交通事故削減に向けた自動運転技術の基礎研究として、画像認識技術を活用した自律走行車両の開発を目指す。この課題に対し、以下の3つの具体的な到達目標を設定し、これらの目標達成を通じて、人間の脳モデルの理解を深めるとともに、機械学習と画像認識技術の実践的な活用方法を習得することを目指す。

1. 白線認識による自律走行の実現
   車両に搭載したカメラモジュールを用いて走行路の白線を認識し、それに従って自律走行できるシステムを開発する。
2. 効率的な経路探索の実装
   幅優先探索アルゴリズムを用いて、目的地までの最適な経路を探索・走行できるシステムを実装する。
3. 交通標識認識システムの構築
   カスケード分類機を用いた機械学習により、一時停止などの交通標識を認識し、適切な制動を行えるシステムを構築する。

3. 課題解決のプロセスとその結果

触覚を持つ歩行ロボット

触覚によるフィードバックを用いた制御則と用いない制御則を作成し、ロボットに実装した。そして、それぞれの制御足を用いて歩行した際の揺れを比較した。

まず、触覚によるフィードバックを用いない歩行を制御則Aとした。そして、触覚によるフィードバックを用いて不整地を歩くときの体の揺れを減少させるための2つの数理モデルを考案し、制御則B、制御則Cとした。

制御測Bは足先に受けた力に応じて脚をまわす大きさを調整するモデルである。足先に受ける力が小さい時は脚をさらに伸ばそうとする。逆に、足先に受ける力が大きい時は脚を引き寄せようとする。このように脚に受けた力の大きさによって脚を回す動きを調整することによって歩行を調整する。

制御則Cは足先に受けた力に応じて脚を引き寄せるモデルである。ロボットの関節はサーボモーターによって角度が固定されており、柔軟な動きを実現できない。生物の脚がもつばねのような柔軟な動きを再現する。制御則Bとは異なり、脚がついたことを認識した瞬間に脚を引き寄せる動きをする。

ロボットの関節にはサーボモーターを搭載して脚を動かせるようにした。そして、ロボットの足先には感圧センサを搭載して、フィードバック情報を得られるようにした。

制御則A、制御則B、制御則Cを適用した歩行の際の揺れの大きさをそれぞれ計測した。簡単な不整地の例として、1cmの段差をすべての脚が乗り上げるまで歩行させ、加速度センサを用いて、鉛直方向と水平方向の揺れをそれぞれ計測した。そして、計測結果の分散の平均をそれぞれの制御則について比較を行った。

制御則Aと比較して、制御則Bは鉛直方向と水平方向の揺れはどちらもあまり変化がなかった。一方、制御則Cは制御則Aと比較して、鉛直方向と水平方向の揺れはどちらも小さかった。

声から予想されるキャラクターを出力するAI

本プロジェクトでは先述の目標を達成するために、データセットAnim400k［B-1］を用いることとした。このデータセットには、日本語と英語の両言語で合計425,000以上のアニメーションビデオクリップが含まれており、各作品にはジャンル、テーマ、評価、キャラクタープロフィール、アニメーションスタイルなどのメ

タデータが存在する。しかし、このメタデータにおいてキャラクターの顔画像と音声は一対一対応していない。そのため、顔画像から検出したキャラクターと音声から検出したキャラクターの一対一対応を事前に作成した後、その対応を元に音声から顔画像を推論するプログラムを作成した。

顔画像と音声の一対一対応の作成手順は大きく4つに分けられる。まず、ビデオクリップから顔画像にトリミングなどの前処理を施した。次にアニメビデオクリップの各フレームからアニメ顔検出を行った。そして、キャラクターの顔画像と音声を別々に特徴量抽出した。顔画像の特徴量抽出ではDeepFace［B-2］ライブラリを、音声の特徴量抽出ではspkrec-ecapa-voxceleb［B-3］を使用した。次に各特徴量をクラスタリングし、顔画像のクラスターと音声のクラスターを作成した。顔画像・音声のクラスターの形状が球体に近くなるように階層的クラスタリング手法を選択した。最後に、顔画像のクラスターと音声のクラスターを最大マッチングを行い、対応付けを行った。

推論プログラム内には5つの処理がある。まず、入力された音声データに対して、先ほどと同じように特徴量抽出を行う。次に、その特徴量と、事前に計算された音声のクラスターの重心を元にそのクラスターに重みを付ける。続いて、事前に作成された顔画像クラスターと音声クラスターの一対一対応を利用して、重み付き音声のクラスターから重み付き顔画像のクラスターへの変換を行う。その後、顔画像のクラスターの重心に重みを適用し、音声から予想される顔画像の特徴量を求める。最後に、その埋め込みベクトルに近い埋め込みベクトルを持つ顔画像を検索・表示する。

結果については、音声から画像群を出力することに成功した。このモデルに対する定量的な評価手法は未確立であるため、今回は主に目視による定性的な評価を行った。約半数は入力音声から想定される顔として妥当であった。しかし、人ではないキャラクターの画像も含まれており、これについては妥当性の判断が困難であった。さらに、出力画像群にはほぼ同一の画像が複数存在していることが判明した。また、異なる音声入力を与えたにもかかわらず、同一の画像群が出力されるという問題も確認された。

視覚を持つAIカー

1. ソフトウェア

機械による認知から判断までのプロセスを実装するにあたって最も気を使ったのは、認知させた情報と実際の挙動に不整合を起こさないように条件分岐を調整することである。今回のプログラムでは、カメラから取得したフレームを用いた白線認識と標識認識、グリッドマップの近似を利用したルーティング、およびそれらに係る動作についてが記述されている。プログラムの制作途中、白線検知のシステムで認識した通行可能領域に対して正確に車が追従しなかったり、ルーティングシステムの仕様上、開始地点と終着点を主導で入力する必要があるが入力とは違う挙動を行ったりと、それぞれの機能で想定された情報の扱い方が他の機能にうまく引き継がれないという事態が多発した。

このような事態に対し、コードのカプセル化や変数のグローバル/プライベート化を多用することで対処した。引き継ぐべきデータの扱いを分かりやすく関数化することで引継ぎがスムーズになり、事後のデバッグもスムーズに行うことができた。

2. ハードウェア

白線を正確に認識するため、カメラには道路以外の背景が極力映らないように調整する必要があった。また、標識を認識するにはカメラが少し左方向を向くことが求められた。この目的を達成するため、Adobe Illustrator、レーザーカッターを用いてそれぞれ異なる役割を担う2台のカメラの土台を設計した。その結果柔軟かつ効率的なカメラ配置を実現した。

4. 今後の課題

触覚を持つ歩行ロボット

制御測Cで最も揺れを減少させることができた。このことから、足先の触覚によるフィードバックを脳よりも下位の神経システムで処理することで、不整地での歩行における体の揺れを減少させることができると考えられる。

評価・検証の方法について、不整地での歩行を目標としたが、行った検証では段差に乗り上げるのみの動作であった。より正確な検証をするためには、実際に昆虫が歩行しているような複雑な環境を用意する必要がある。また、その場合は複雑な環境を歩行し、センサの値を正しく取得できるロボットを用いる必要がある。

昆虫などの複数の脚を持つ生物は各脚を動かすリズムが、歩行速度や体の重さによって変化することが知られている。このことから、歩行のリズムを調整することによる歩行の改善をする数理モデルについても検討したい。

また、本プロジェクトでは脳よりも下位の神経システムに着目したが、昆虫の歩行における脳の役割についても調べる必要がある。

声から予想されるキャラクターを出力するAI

このプロジェクトでは、データセットのアノテーションファイルから話者が単一のものであるとわかる音声・動画クリップを使用し、話者と動画のフレームに映っているキャラクターが同一のものであるという仮定のもとクラスタリングを行った。そのため、フレームに映っているキャラクターと実際の話者に相違がある場合に、マッチングのノイズになっていた可能性がある。

また、特徴量とクラスター間の距離を基に、距離の逆数を重みとして適用したが、より適切な重み付け関

数を導入する余地がある。一方で、一対一対応の作成には成功しており、これを基にGANなど他の推論手法を探求することができる。

さらに、クラスタリング結果について定量的な評価が可能な部分については、例えばクラスタリング精度や推論の正確性を評価するために適切な指標を設定する必要がある。具体的には、クラスタリングにおいてはクラスタ純度や正解率を用い、推論結果の評価には特徴量の類似度やキャラクター識別の一致率などを採用することが考えられる。これにより、手法の有効性を客観的かつ再現可能な形で示すことができる。

視覚を持つAIカー

今後の課題点は3つある。1つ目は出力される映像に遅延があったことだ。プログラムコードの処理負荷が高いことや、Wi-Fiの電波が届きにくい場所での実験が原因と考えられる。特に、画像処理アルゴリズムの計算負荷がラグの発生に寄与した可能性が高い。

2つ目はライン走行の精度が低いことだ。車両がラインの中心ではなく端を走行する問題が観察された。この問題の原因として、逆透視変換を用いてラインの角度をリアルタイムで計算するアルゴリズムが挙げられる。この処理では、カメラ映像にラインが1本しか映っていない場合と、2本映っている場合とで、サーボモーターの制御が適切に分岐できていなかった。特に、一方のラインがカメラの視野外に出ると、残った1本のラインの角度だけを基準にサーボが制御される結果、車体がラインに対して平行な位置で端を走行する状態となった。カメラの画角を広げてラインが常に2本映るよう調整する方法も検討したが、その場合、交差点や対向車線のラインが映り込むため、制御が困難になる問題が新たに生じた。

3つ目はモバイルバッテリーによる電力供給ができなかったことだ。モバイルバッテリーを用いた電力供給が正常に機能せず、動作に必要な電力を十分に供給できない問題が発生した。この原因として、コードが重くなり、実行時の消費電力が増加したことが挙げられる。モバイルバッテリーの出力が不足したことで、システム全体の安定動作が妨げられたと推測される。

最後に

触覚を持つ歩行ロボット、声から予想されるキャラクターを出力するAI、視覚を持つAIカーの3つのグループが、それぞれ独自の課題に対して解決策を提案し、一定の成果を得ることができた。しかしながら、実環境での運用を見据えた場合、各システムにはまだ多くの課題が残され、解決するためには、シミュレーションと実環境での検証を繰り返し、システムの頑健性を高めていく必要がある。今後は、生物学や神経科学の知見をより積極的に取り入れ、生物に学ぶことで、より高度な人工知能システムの開発につなげていくことが重要である。この研究活動を通して、人間の脳を再現するような高度な人工知能を目指すためには、日常的に感覚的に行われている行動を論理的に理解する必要がある。触覚、聴覚、視覚といった感覚の機能をより深く模倣するために、生物学や神経科学の知見を積極的に取り入れ、それを技術的に応用することが重要である。

参考文献

[A-1] Dai Owaki, Masashi Goda, Sakito Miyazawa, Akio Ishiguro(2017), A Minimal Model Describing Hexapedal Interlimb Coordination: The Tegotae-Based Approach. Frontiers in Neurorobotics, 11, Article 29. https://doi.org/10.3389/fnbot.2017.00029

[A-2] 石川慶孝,大渡健太,井上祐貴,高良拓馬(2021), 神経回路モデルによる歩行シミュレーション. https://www.fun.ac.jp/wp-content/uploads/2022/01/document20_A.pdf

[B-1] Kevin Cai, Chonghua Liu, and David M. Chan (2024), Anim-400k: A large-scale dataset for automated end-to-end dubbing of video.

[B-2] Sefik Ilkin Serengil and Alper Ozpinar(2021), Hyperextended lightface: A facial attribute analysis framework.In 2021 International Conference on Engineering and Emerging Technologies (ICEET), pp. 1-4. IEEE.

[B-3] Mirco Ravanelli, Titouan Parcollet, Peter Plantinga, Aku Rouhe, Samuele Cornell, Loren Lugosch, Cem Subakan, Nauman Dawalatabad, Abdelwahab Heba, Jianyuan Zhong, Ju-Chieh Chou, Sung-LinYeh, Szu-Wei Fu, Chien-Feng Liao, Elena Rastorgueva, Fran¸cois Grondin, William Aris, Hwidong Na,Yan Gao, Renato De Mori, and Yoshua Bengio(2021), SpeechBrain: A general-purpose speech toolkit. arXiv:2106.04624.

[C-1] 警視庁, ″各種交通事故発生状況（令和5年度）, ″ 2023. [Online]. Available: https://www.keishicho.metro.tokyo.lg.jp/about_mpd/jokyo_tokei/tokei_jokyo/vta.html [Accessed: Dec. 20, 2024].