6 考察

6.1 クロスモーダルメモリを取得するための手段としての情報の補完機能

本研究では、感覚運動統合学習の問題に時間遅延オートエンコーダーを適用し、ディープラーニングアルゴリズムの大幅な拡張性を実証した。我々は、クロスモーダルメモリ検索と実環境でのヒューマノイドロボットの次の適応行動の生成に関する実験結果を発表した。例えば、4.3 章での学習オブジェクト操作行動タスクの画像シーケンス検索実験においては、900 次元の画像特徴ベクトルのシーケンスは、300 次元の関節角度シーケンスの入力からのみ呼び出された。この結果は、3 倍の情報がオートエンコーダーの汎化能力により呼び出されたことを示している。

この強力な情報補完機能は、我々の提案した時間遅延オートエンコーダーの利点の一つである。オートエンコーダーの層の前半部分の汎化能力を利用して、特定のオブジェクト操作の挙動を表す高レベルの機能は部分的なモーダル入力からでも生成することができる。また、特徴ベクトルから元の入力を再構成することができるオートエンコーダーとして、予測出力は、再帰的に入力ノードにフィードバックすることができ、入力は、任意の不足しているモダリティ情報の代替として使用することができる。提案フレームワークでは、この再帰的な情報ループは、クロスモーダルメモリ検索性能の高レベルの安定性を可能にした。

層の数及びノード数は、メモリ容量やディープニューラルネットワークの汎化能力を説明するための重要な要素である。しかしながら、一般的に明確な説明は、ネットワーク構造とその学習能力との間の相関のために作られていない。このように、ニューラルネットワークの構造上の設計原理は、現時点では理論的基礎をほとんど持っていない。
これは、今後の検討のための重要な研究テーマであるかもしれない。

6.2 行動認識タスクでのロバスト性に寄与する三つの要素

動作認識評価に関する実験結果は、圧縮された一時的な機能はロバストな認識性能を可能にすることを実証した。異なる評価条件の認識率を比較することにより、我々は、以下の 3 つの要素は、行動認識タスクにおけるノイズ耐性に寄与することを示した。

（1）より高いレベルの機能の利用
（2）マルチモーダル情報の利用
（3）マルチモーダル行動認識における自己生成シーケンスの利用

以下では、提案フレームワークの内部機構に関連した三つの要因の機能に関する我々の見解を提示する。

6.2.1　より高いレベルの機能の利用

ルーらによる以前の研究ではそれは完全にラベル付けされていないデータを用いて、高レベルの概念のために選択的に働くニューロンをトレーニングすることが可能であること

を示した([19] 大規模な教師なし学習を使用した高レベルな特徴の構築)。実用的な結果として、彼らはラベル付けされていない YouTube のデータセットをディープニューラルネットワークを訓練することによって、このような猫と人体検出器ニューロンなどのクラス固有のニューロンを取得することに成功した。この結果、すなわち、意味のある特徴は、ラベル付けされていないデータと並行して自己組織化されることができる。つまり、特徴抽出機構としてオートエンコーダーを利用する利点を示している。同様の結果は、画像分類タスク([21] 畳み込みディープニューラルネットワークでの Imagenet 分類)と、音声認識タスク([20] 音声認識における音響モデリングのためのディープニューラルネットワーク)のネットワークを含む、関連研究で提示されている。これまでの研究のすべてを考慮すると、私たちの行動認識結果がディープニューラルネットワークが次第にデータに複雑な統計的構造を表現するために、非線形特徴検出器の多くの層を蓄積することにより顕著な汎化能力を持って、より高いレベルの機能を生成するという見解と一致するように見える([35] ドキュメントの認識を適用した勾配ベースの学習)。

6.2.2 マルチモーダル情報の利用

マルチモーダル配列から取得した情報量の観点からは、マルチモーダル時系列学習ネットワークは、単一モーダル時系列学習ネットワークよりもより正確な内部表現を生成する際に明確な利点がある。この事実は、4.6 章における（(2A) MTN+ IMG）のトレーニングデータセットからのノイズの多い関節角度入力と鮮明な画像入力における行動認識結果に示されている。これらの結果は、関節角度情報は情報価値がなくなった後でさえ、認識率の低下は、他の結果を上回るレベルに収束することを証明している。この場合には、明確な画像特徴入力が情報価値のない関節角度入力に対する動作のカテゴリを正しく表すために、より高いレベルの機能のための情報源として役立った。認識タスクのロバスト性に向けたマルチモーダルな学習の効果に関する現在の結果は、マルチモーダル音声認識タスクと同じであるとみなすことができる。例えば、マルチモーダル音声認識タスクの改善は、音と映像入力の組み合わせを利用している([24] マルチモーダルディープラーニング)。

6.2.3 マルチモーダル行動認識における自己生成シーケンスの利用

4.6 章に示される私たちの行動認識評価結果の中で最も注目すべき結果は、より高い認識性能は関節角度のための単一のモーダル入力と並行してマルチモーダルメモリによって実現されることである((2b)MTN+ IMG[仮想])。これは次のように結果を説明することができる。マルチモーダルメモリ利用して、マルチモーダル内部表現は、ノイズの多い関節角度入力からでも生成され、順次に伴う画像の特徴は、出力ノードから取り出される。画像特徴ベクトルは内部表現から呼び出されるので、その情報は、関節角度の観測結果に含まれたエラーから独立している。入力ノードに検索された画像特徴量をフィードバックすることで、この手順は、ノイズの多い関節角度シーケンスと並行して、ネットワークへの画像

特徴シーケンスを明示的に提供することによって、マルチモーダル認識処理に相当する内部表現を明確に導く。最近の神経心理学的研究では、メモリ障害の集団における改善しているメモリ内の自己参照型戦略のプラスの効果が([36] 自己想像効果：神経障害を有する記憶障害者の手がかり呼び出しの自己参照符号化戦略の利点)と([37] 神経障害を有する記憶障害者に認識記憶を強化する自己想像)によって報告されている。今後は、さらに私たちの現在の自己生成の仮想シーケンス機構が人間の認知過程で、このような心理的な現象に対応する方法の検討が興味深いものになるだろう。

6.3 感覚運動予測のための手段としてのクロスモーダルな因果関係のモデリング

本研究では、マルチモーダル情報を統合することにより、複数のモダリティの間で暗黙的な同期を抽出することができ、提案フレームワークを提示した。また、ベルリンギングタスクで検索された画像は、提案フレームワークが確定的に取得因果関係を反映したベルの画像を取得しただけでなく、他の鐘を識別できない場合でも、複数の可能性の中で候補を選択することによって、何らかの方法で代替情報を生成した。したがって、我々は提案されたメカニズムは、感覚運動状態の連続した結果を推測するロボットの予測機構として利用することができると信じている。認知科学の研究では、主体感は、動作と効果との因果関係の一般的な判断の産物であることが知られている。そして、実験結果は動作と推定される効果に関連している信号間に時間的な連続性やコンテンツの整合性がある時に主体感が生じていることを示している([38] 多感覚空間表現の神経基盤上の計算的視点)、([39] ベイズ多感覚統合とクロスモーダル空間のリンク)、([1] ロバストな知覚への感覚併合)。さらに、最近の研究は、主体感の発生に動作効果のグループ化の重要性を示している([4] クロスモーダルのグループ化に基づく動作効果の因果関係の知覚である主体感)。これら全ての研究において、予測と実際の感覚フィードバックとの間の時空間的適合性の評価は、代理店の間隔において重要な役割を果たしていると考えられる。私たちの現在の結果から、我々はクロスモーダルな因果関係のモデル化とその後のメモリ検索機能は感覚フィードバックの予測のための実用的な計算上フレームワークとして利用することができると考えている。したがって、我々は提案フレームワークは主体感のより深い理解を促進するために、今後の活動に利用することができると信じている。

6.4 時間遅延オートエンコーダによる時系列学習の機能的特性

6.4.1 私達の提案した時間遅延オートエンコーダと元の時間遅延ニューラルネットワークの違い

我々の研究で提案されている時系列の学習メカニズムは、時間窓内の時系列シーケンスの領域を切り取ることによって取得された時系列の一定の長さを入力する。このアプローチは、ラングらによる時間遅延ニューラルネットワークの研究からアイデアを継承している([29] 孤立単語認識のための時間遅延ニューラルネットワークアーキテクチャ) 。ここで

の違いは、入力と同一のベクトルは、我々の提案モデルでは入力用のベクトルが目標の出力を定義することであるのに対して、対象のラベルは、元のモデルでは対象のラベルが出力を定義することである。したがって、提案モデルの特徴の一つは、時系列の圧縮表現はオートエンコーダによる自己組織化であり、ネットワークは再帰的に入力ノードに出力をフィードバックすることで時系列を自己生成することができるということである。内部シーケンス生成の利点は、クロスモーダルメモリの取得および信頼性の低い関節角度観測でのロバストな行動認識機能を利用した適応的行動選択能力によって示された。

6.4.2　時系列の学習ネットワークの内部表現の特性

時系列の学習ネットワークは、累積複数の位相ごとの時間的なセグメントを記憶することにより、長い時系列の動態を事実上モデル化している。したがって、ワンショットの入力から生成された特徴ベクトルは、シーケンスの時間的な位相を表す。この現象は、それらが、図 9 のように閉じたループ形状を形成した場合を観察することによってベルリンギングタスクの特徴ベクトルのプロットから確認することができる。同じ現象は第 2 のタスクから二つの別個のライン上の特徴ベクトルプロットの逆数遷移は図 16 に示す各々の左右の腕運動パターンに対応することで、確認することができる。

6.4.3　時間遅延オートエンコーダが処理するコンテキスト情報の長さ

入力時間セグメントの長さは、時系列の学習ネットワークによって処理されたコンテキスト情報の長さを定義する。したがって、原理的には、時間的なセグメントよりも長いコンテキスト情報は考慮されない。周期的なニューラルネットワーク([40] ヘッセ行列のない最適化での周期的ニューラルネットワークの学習)のような他の時系列学習機構と比較すると、これは根本的な違いである。提案フレームワークでは、我々のタスクの設定においてロボットの動作の実行が長いコンテキストの状況を把握する必要がなかったため、コンテキスト表現のこの制限があるにもかかわらず、我々の実験では正常に働いていた。例えば、オブジェクト操作およびベルリンギング動作のために、コンテキスト情報のほとんどは、環境中に埋め込まれている（例えば、ロボットアームの姿勢、ボールの位置等)。このように、コンテキストの内部的な神経表現は、このタスクを達成するために必要ではなかった。

7.　結論

本研究では、視覚、聴覚、および運動を含む時系列のマルチモーダル統合学習を可能にし、ディープニューラルネットワークのフレームワークを提案した。提案フレームワークの性能は、実環境で人型ロボットを利用して二つのタスクで評価した。タスクは、オブジェクト操作とベルリンギングのタスクから構成されている。我々の結果は有意な次元での大量のトレーニングデータを扱うことにおいて、提案フレームワークの拡張性を実証した。取得した感覚運動統合モデルの 3 つのアプリケーションを示した。まず、クロスモーダル

記憶検索を実現した。ディープオートエンコーダの汎化能力を活用し、提案フレームワークは、画像と動作間の双方向の時間的シーケンスを取得することに成功した。第二に、ロバストな行動認識は、教師あり行動の分類学習への入力として取得したマルチモーダル機能を利用することにより実現した。第三に、マルチモーダルな因果関係のモデル化が実現した。我々の実験結果は、提案フレームワークは、ロボットのベルリンギング行動から色、ピッチ、ベルの位置と対応するベルリンギングの動きとの間の同期性をモデル化し、それらの相対関係を記憶できることを実証した。

実環境におけるオブジェクト操作行動の実時間遷移の結果はまた、生の画像データを利用するための我々の現在のアプローチは、依然として照明条件の急激な変化を取り扱うために十分に安定ではないことを明らかにした。今後の研究は、より多様なデータセットで訓練された畳み込みネットワークの導入を介して、ディープネットワークの汎化能力の可能性を引き出すことにより、画像認識機能のロバスト性を向上させることを含んでいる。現在のベルリンギングタスクのように、我々は 2 つだけのベル位置での音と動きから画像検索性能を評価した。今後の研究では、より多くの種類のベル位置を使用して、ベルリンギングの行動において私達のシステムを訓練することによってベル位置の一般化表現をモデル化することも面白いかもしれません。もう一つの重要な課題は、異なる感覚源の相対的な信頼性を考慮に入れることによって、動的に複数の感覚モダリティを組み合わせることである。もし信頼性に依存する統合が、我々のフレームワークの中で達成されているならば、より高いレベルの特徴量は内部表現によるモダリティを意図的に抑制することによって取得される可能性がある。これは、よりロバストな行動認識性能をもたらす可能性がある。

補足

Imagnet：自然言語処理の分野で有名な WordNet のオントロジーに従って、各単語（今のところ名詞のみ）に対応する画像を収集したもの。